Azimuthal Angle Asymmetries of Hadrons and Jets in Deep Inelastic Scattering at HERA

Eduardo Rodrigues

Department of Physics

H. H. Wills Physics Laboratory

July 2002

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Doctor of Philosophy in the Faculty of Science

Word count: 50000

Abstract

The analyses presented in this thesis are based on the data sample recorded with the ZEUS detector during the HERA data taking period $1996-97$. The total integrated luminosity amounts to $34\ \mathrm{pb}^{-1}$ and corresponds to positron-proton collisions at a centre-of-mass energy $\sqrt{s} \approx 300\ \mathrm{GeV}$.

A detailed study was performed of the azimuthal angle correlations of hadrons and jets in deep inelastic scattering at HERA with the ZEUS detector.

The azimuthal angle distributions of hadrons (measured around the direction of the exchanged boson) and the $\langle cos\phi \rangle$ and $\langle cos2\phi \rangle$ moments of these distributions have been measured in the hadronic centre-of-mass frame. The evolution of these moments with the hadron's transverse momenta was confronted to the theoretical leading order QCD calculations implemented in a program containing a simple model to account for the transverse momenta of the intrinsic partons within the proton and the hadronisation transverse momentum. The data were also compared with the leading order Monte Carlo event generators Lepto and Ariadne. Both the QCD and Monte Carlo predictions were found to be in general qualitative agreement with the data. In particular, the $\langle cos2\phi \rangle$ moment was measured for the first time and shown to be to a large extent due to the perturbative QCD contribution.

A similar analysis was performed in the Breit frame with dijet events. The limitations of using a neural network to discriminate quark- from gluon-initiated jets and classify the events as QCD-Compton or BGF are shown. The predicted $\cos 2\phi$ dependence of the dijet cross section is clearly observed in the data. Leading order Monte Carlo and next-to-leading order QCD calculations were found to be in good agreement with the data.

Suma

As análises apresentadas nesta tese são baseadas em dados experimentais obtidos com o detector ZEUS durante o período de recolha de dados $1996-97$ do acelerador HERA. A luminosidade integrada total eleva-se a $34\ \mathrm{pb}^{-1}$ e corresponde a colisões de positrões-protões com uma energia no centro de massa de $\sqrt{s} \approx 300\ \mathrm{GeV}$.

Foi realizado com o detector ZEUS um estudo das correlações azimutais angulares dos hadrões e dos jactos em difusão inelástica profunda com o detector ZEUS.

Foram medidas no sistema do centro de massa hadrónico as distribuições azimutais angulares dos hadrões (medidas à volta da direcção do bosão virtual) e os momentos $\langle cos\phi \rangle$ e $\langle cos2\phi \rangle$ dessas mesmas distribuições. A evolução destes momentos com os momentos transversais dos hadrões foi confrontada com cálculos teóricos de QCD implementados num programa contendo um modelo simples para a descrição dos momentos transversais dos partões intrínsecos do protão e do momento transversal de hadronização . Os dados foram também comparados com os geradores Monte Carlo de eventos Lepto e Ariadne. Tanto os prognósticos da QCD como os de Monte Carlo mostraram estar de uma maneira geral em conformidade com os dados. Foi medido pela primeira vez o momento $\langle cos2\phi \rangle$, que revelou ser devido, em grande parte, à contribuição perturbativa de QCD.

Foi realizada no sistema de Breit uma análise semelhante de eventos com dois jactos de hadrões. São mostradas as limitações devidas ao uso de uma rede neuronal artificial para distinguir os jactos provocados por quarks dos provocados por gluões e para classificar estes eventos como QCD-Compton ou BGF. Os dados mostram claramente a dependência em $\cos 2\phi$ da secção eficaz de eventos com dois jactos. Os cálculos de Monte Carlo e os de primeira ordem em QCD estão em boa conformidade com os dados.

Acknowledgements

I would like to take this opportunity to thank Dr. Nick Brook for his constant advice and expertise. The help given and availability always shown were very much appreciated. And many thanks for correcting (patiently) this thesis.

Several colleagues with whom I've shared offices and/or time contributed to my discovery of the british culture and to a good atmosphere at work: cheers all! I would also like to thank the members of the Hadronic Final State and QCD Group at DESY; in particular Oscar and Juan, for several fruitful and interesting discussions.

The many friends I made during the course of these years are to a large extent responsible for a great time. Special thanks go to Alex, Federica and Shannon, ... This time simply wouldn't have been so much fun without you!

A big thank you to my family for their "background support". They've always been there – can one ask for more? Also thank you to my mum for encouraging me to finish the thesis with the now famous question, "So, is it done?".

Special thanks to the Fundação para a Ciência e a Tecnologia for the financial support through the Ph.D. studentship PRAXIS XXI / BD / 13446 / 97. I also acknowledge financial funding from the CLRC during my long term attachment at the DESY laboratory, in Hamburg.

Esta tese foi suportada financeiramente pela bolsa de Doutoramento PRAXIS XXI / BD / 13446 / 97 da Fundação para a Ciência e a Tecnologia, no âmbito do Programa de Intervenção PRAXIS XXI: os meus sinceros agradecimentos.

Para a mamã e o chavinho

Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the Regulations of the University of Bristol. The work is original except where indicated by special reference in the text and no part of the dissertation has been submitted for any other degree. Any views expressed in the dissertation are those of the author and in no way represent those of the University of Bristol. The dissertation has not been presented to any other University for examination either in the United Kingdom or overseas.

Eduardo Rodrigues

Date: July 2002

Contents

List of Tables

List of Figures

Chapter 1

Motivation and Foreword

The HERA Collider at the DESY laboratory commenced taking data in 1992, and a considerable amount of studies has been performed since. These studies include investigations of topics such as the proton structure, photoproduction, diffraction, hadron multiplicity distributions, jet-related observables and production of heavy flavours.

The steady increase in statistics together with a progressively better understanding of the data have allowed both the ZEUS and H1 experiments to enter what is called a period of "precision measurements". The highly competitive determination of the strong coupling constant, as extracted from inclusive jet production cross sections in Deep Inelastic Scattering (DIS), undoubtedly corroborates the statement "Jet production observables are among the most sensitive probes of QCD at high energy colliders, where they are used for example to determine the strong coupling constant" [1].

Other detailed studies and tests of Quantum Chromodynamics (QCD) can be done at HERA. There has been a renewal of interest in the investigation of angular correlations between the scattered lepton and the produced hadrons of the DIS final state, both theoretically and experimentally.

In the late seventies, H. Georgi and H. D. Politzer proposed the analysis of angular correlations in semi-inclusive DIS as a clean means of testing perturbative QCD [2]. They argued that "the most interesting results of the perturbative analysis concern features which are totally absent in zeroth order, such as non-zero P_T and non-trivial

ϕ dependence." [2]. Although nothing can be pointed out against this argument, the same cannot be said with respect to their assumptions, namely the absence of non-zero transverse momentum P_T and non-trivial azimuthal angle, ϕ, dependence at the Quark Parton Model level – zeroth order in QCD. As shown immediately after by R. Cahn, the non-zero intrinsic transverse momenta of the partons confined within the proton give rise to an azimuthal dependence in the cross section [3, 4], spoiling the pure first order QCD effect predicted by Georgi and Politzer.

Fixed target experiments [5–8] have performed studies of the azimuthal angle distribution of the final state hadrons with respect to the lepton scattering plane in a low energy regime in which the above-mentioned non-perturbative effects dominate. Indeed, leading order QCD contributions to the observed asymmetries were not required to explain the data – although they could not be ruled out.

But non-perturbative effects are expected to play a diminishing role with increasing Q^2, Q being the virtuality of the probing boson in a DIS process. At the energies currently available with the HERA collider, the non-perturbative effects should be masked by the pure QCD effects, allowing for a clean and original test of the underlying theory. A first paper was published by the ZEUS Collaboration that shows a qualitative agreement between the predictions and the data [9].

My own investigation involved a detailed study of the azimuthal asymmetries in DIS at HERA, both with hadrons and with jets. This thesis will present in detail these analyses.

Other aspects of my Ph.D work will not be presented here; I briefly summarise them.

In my first year I concentrated on Monte Carlo simulation of deep inelastic events at HERA. I studied in detail one particular generator, Ariadne, with respect to DIS. The major study was a comparison of the "old" and latest versions of Ariadne. Theoretical developments had led the author to slightly modify the philosophy of the model implemented in the generator, and a detailed study of its consequences was necessary and essential. This research was done as part of the 1998/99 Workshop on Monte Carlo generators, held at DESY, in Hamburg, where I presented the work [10]. I then pursued

tuning the model parameters to achieve a better description of the HERA DIS data. The conclusions of the comparison of Ariadne to HERA data were then collected with similar work done for other event generators and published in the proceedings of the HERA Monte Carlo workshop [11].

In addition to the physics analyses, I've also been involved, together with a research assistant, in the setting up of a relational database for the ZEUS Central Tracking Detector (CTD). The database contains all the relevant information necessary to control and maintain the readout system of the CTD. It will be accessed via a Web-based interface. The project is basically composed of three parts: (1) the relational database (Microsoft Access), which stores all the relevant information about the readout system; (2) the web interface, that makes use of the Active Server Pages (ASP) technology which together with Structured Query Language (SQL) eases the access, manipulation and changes of the contents of the database: and (3) finally, the C++ code, that allows the modifications to be implemented by the user, and stored in the database, and links to the actual configuration files of the CTD readout system. I focussed on the first 2 aspects of the project.

The thesis is organised as follows. Chapter 2 briefly reviews the topics of DIS most relevant to the subject of the present thesis, emphasising aspects of semi-inclusive positron-proton scattering and the physics behind the origin of azimuthal asymmetries. The third chapter describes the main components of the ZEUS detector and the HERA accelerator. The next chapters detail the analyses performed with tracks and with jets, as well as some theoretical numerical calculations. The last chapter summarises the experimental results and the conclusions obtained from the latter.

Notations

Throughout this thesis all particles energies and momenta will be expressed in Giga electron-volts (GeV). The term "electron" will be used sometimes to refer to an electron or a positron interchangeably whenever no specification is required.

Chapter 2

Basics of Deep Inelastic Scattering

2.1 General Formalism of DIS

The physical process

$$e(k^\mu) + P(P^\mu) \rightarrow e'(k'^\mu) + X(P'^\mu) \quad , \tag{2.1}$$

depicted in figure 2.1, is referred to as *(inclusive) deep inelastic positron(electron)-proton scattering* (DIS) when the energy transferred during the interaction between the positron e^+ (electron e^-) and the proton P is much higher than the rest mass of the proton. Here P^μ, k^μ, k'^μ and P'^μ denote, respectively, the 4-momentum of the

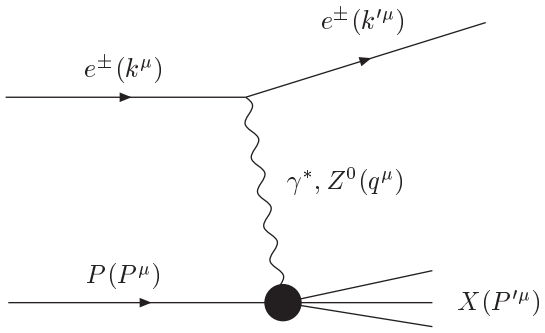

Figure 2.1: *Schematic diagram of a neutral current deep inelastic interaction of a positron e^+ (or electron, e^-) off a proton P.*

proton, the 4-momenta of the incoming and outgoing positrons (electrons), and the

4-momentum of the hadronic final state X. In the course of this hard scattering process the proton generally breaks up and gives rise to a complex hadronic final state.

2.1.1 Kinematics for Inclusive DIS

The kinematics of an electron-proton DIS event is studied by means of a series of relativistic invariants:

- the square of the centre-of-mass energy of the colliding electron and proton,

$$s = (k^\mu + P^\mu)^2 \approx 4E_e E_p, \tag{2.2}$$

  E_e and E_p being the electron and proton beam energies, respectively;

- the (negative) square 4-momentum transferred between the electron and the proton,

$$Q^2 = -q^2 = -(k^\mu - k'^\mu)^2 \qquad , \qquad Q^2 \in [0, s - m_p^2] \quad ; \tag{2.3}$$

- the Bjorken invariant x,

$$x = \frac{Q^2}{2P.q} \qquad , \qquad x \in [0, 1] \quad ; \tag{2.4}$$

- the Bjorken invariant y,

$$y = \frac{P.q}{P.k} = \frac{E_e - E'_e}{E_e} \qquad , \qquad y \in [0, 1] \quad , \tag{2.5}$$

  equal, in the proton rest frame, to the fractional electron energy loss (transferred to the hadronic system);

- the invariant mass of the hadronic system, or, equivalently, the squared centre-of-mass energy of the system exchanged boson-proton,

$$W^2 = (P^\mu + q^\mu)^2 = Q^2(\frac{1}{x} - 1) + m_p^2 \qquad , \qquad W^2 \in [m_p^2, s] \quad . \tag{2.6}$$

Bjorken x can be interpreted as the proton's fractional longitudinal momentum carried by the initial parton in the approximation of a rapidly moving proton.

A DIS process is governed by only 2 independent variables, often chosen to be x and Q^2. The variables x and y can be related to Q^2,

$$Q^2 \approx sxy \quad , \tag{2.7}$$

neglecting the electron and proton rest masses.

2.1.2 Kinematics for Semi-inclusive DIS

In a semi-inclusive DIS process (SIDIS), one hadron is observed out of the hadronic final state. The process is depicted as

$$e(k^\mu) + P(P^\mu) \rightarrow e'(k'^\mu) + h(P_h^\mu) + X(P'^\mu) \quad , \tag{2.8}$$

where h represents the observed hadron.

The complete kinematics is governed by 5 independent variables which define the 5-fold differential cross section. Physically, 3 extra variables are needed to specify the energy and angles of the detected hadron. The set of variables that completely specifies a SIDIS event are the usual DIS variables, the invariant z_h, the transverse momentum P_{Th} and the azimuthal angle ϕ_h of the hadron. The z_h invariant is defined as

$$z_h = \frac{P.P_h}{P.q} \quad , \tag{2.9}$$

P_h^μ being the 4-momentum of the hadron. It takes values in the range

$$0 \leq z_h \leq 1$$

and can be interpreted, in the Quark-Parton Model (QPM), as the fractional momentum of the parent parton carried by hadron h in the fragmentation process.

2.1.3 Basics of DIS

Both the neutral (NC) and charged current (CC) DIS events are described by the exchange of a virtual vector boson between the incoming electron and a parton from within the proton; its 4-momentum is simply

$$q^\mu = k^\mu - k'^\mu \qquad , \qquad q^2 < 0 \quad . \tag{2.10}$$

The virtual exchanged boson can be either a photon γ or a massive Z^0 in the neutral current case, whereas charged current interactions are mediated by a massive charged boson $W^{\pm}$.

In the naïve QPM, a hadron is considered as built up of non-interacting partons identified as the quarks. Within this framework, the hard scattering process gives rise to a "1+1" parton configuration: the outgoing struck parton and the proton remnant (denoted "+1"). The cross section for NC DIS is generally expressed in the form [12]

$$\frac{d^2\sigma}{dxdQ^2} = \frac{2\pi\alpha^2}{xQ^4}\Big[[1+(1-y)^2]F_2(x,Q^2) \pm [1-(1-y)^2]xF_3(x,Q^2) - y^2F_L(x,Q^2)\Big] \quad , \tag{2.11}$$

α being the electromagnetic coupling constant. $F_2(x,Q^2)$ is a structure function that has to be determined experimentally. It parameterises the unknown details of the interaction at the exchanged boson-proton vertex. In this equation, F_3 corresponds to the electroweak contribution coming from the Z^0 exchange; it is only relevant in the very high Q^2 region – due to the high mass of the Z^0 – and can most often be neglected. Also the contribution F_L coming from the longitudinal polarisation of the exchanged boson can be neglected, as it only becomes important in the high y region. These are good approximations in the kinematic region of interest to us in the experimental investigations. The equation then simplifies to

$$\frac{d^2\sigma}{dxdQ^2} = \frac{2\pi\alpha^2}{xQ^4}[1+(1-y)^2]F_2(x,Q^2) \quad . \tag{2.12}$$

In the QPM, the F_2 structure function can be interpreted [13] as the charge-weighted momentum density of the quarks and antiquarks (measured in units of the proton's momentum):

$$F_2(x,Q^2) = \sum e_q^2 x[q(x,Q^2) + \bar{q}(x,Q^2)] \quad , \tag{2.13}$$

where $q(x,Q^2)$ and $\bar{q}(x,Q^2)$ are the probabilities of finding a quark or antiquark carrying a proton's momentum fraction x, and e_q is the charge of quark q in units of the proton's electric charge. $q(x,Q^2)$ and $\bar{q}(x,Q^2)$ are the so-called parton distribution functions (PDFs).

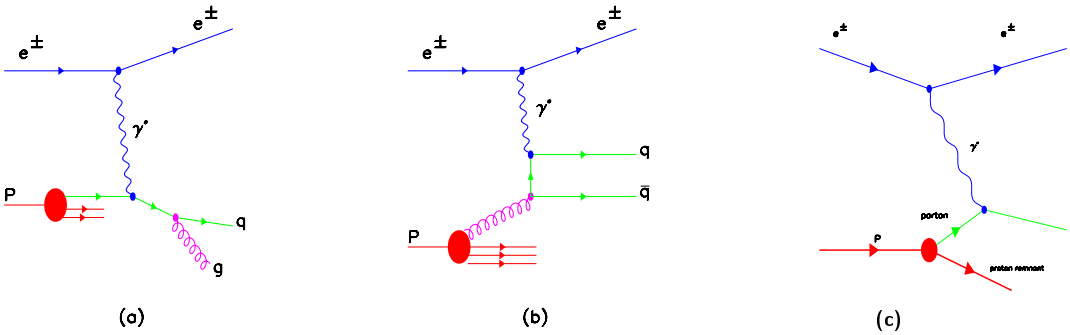

Figure 2.2: *Feynman diagrams for the (a) QCDC, (b) BGF and (c) QPM processes.*

QCD corrections to this naïve picture, which take into account the role of the mediating gluons, are important in the kinematic region studied at HERA. To order $\mathcal{O}(\alpha_s)$, the contributing diagrams are the QCD-Compton (QCDC) and the Boson-Gluon Fusion (BGF) processes; they correspond to a "2+1" parton configuration. All diagrams to order $\mathcal{O}(\alpha_s)$ are presented in figure 2.2.

Perturbative QCD causes corrections to the structure functions, which no longer exhibit scaling in Bjorken x. The Q^2 dependence of these scaling violations is expressed in the DGLAP evolution equations [14–16] of perturbative QCD.

2.2 Reference Frames

The HERA laboratory frame is not the most appropriate reference frame for physics studies as the scattering process involves the interaction of the exchanged boson with a parton from within the proton rather than the interaction of the incoming electron with the proton, and hence introduces a variable transverse momentum.

The 2 frames commonly used in studies of the DIS hadronic final state are the hadronic centre-of-mass frame (HCM) and the Breit frame (BF). In these frames the exchanged boson collides head-on with the struck quark; they are in fact related by a boost along the z-direction (in the QPM). We hereafter describe both in some detail, as they will be used in the analyses.

2.2.1 The Hadronic Centre-of-mass Frame

The azimuthal angle is defined in the hadronic centre-of-mass (HCM) frame in which the proton is collinear with the momentum transfer of the leptons; thus the proton interacts head-on with the exchanged virtual boson.

The velocity of the HCM frame with respect to the laboratory is given by

$$\vec{\beta} = \frac{\vec{P} + \vec{q}}{P^0 + q^0} \quad . \tag{2.14}$$

$P^\mu = (P^0, \vec{P})$ and $q^\mu = (q^0, \vec{q})$ refer to the proton and exchanged boson 4-momenta, respectively.

The boost to the HCM frame is then performed with the above $\vec{\beta}$ vector. Subsequently, the whole event is rotated for the $x-z$ plane to be defined as the lepton scattering plane, containing the incoming and outgoing leptons and the virtual boson, the latter pointing in the negative z-direction. The x component of the momenta of both leptons is positive.

The struck quark initially has a longitudinal momentum equal to a fraction x of the proton's momentum $P_z, p_z = xP_z$. The transverse component makes an azimuthal angle ϕ with the x-axis in the $x-y$ plane. Also both initial and final partons have the same ϕ angle, since the momentum transfer is along the z-axis.

2.2.2 The Breit Frame

The Breit frame (BF) [17] is also often chosen in studies of the hadronic final state. It is the best reference frame to separate the struck quark fragmentation from the behaviour of the incoming quark and the proton remnant, as it corresponds to the frame where there is a maximal separation between the incoming and the outgoing struck quarks.

The Breit frame (see figure 2.3) is defined as the frame in which the struck parton is back-scattered elastically, $|\vec{q}^{in}| = |\vec{q}^{out}|$; the incoming parton having longitudinal momentum $q_z^{in} = Q/2$ collides with the exchanged virtual boson and "bounces back" with momentum $q_z^{out} = -Q/2$ (in the QPM). Hence the virtual boson is purely spacelike:

$$q^{\mu(BF)} = (0, 0, 0, -2xP^{BF}) \quad , \tag{2.15}$$

P^{BF} being the proton momentum in the BF.

It is also easy to demonstrate that the velocity of the Breit frame with respect to the laboratory is given by

$$\vec{\beta} = \frac{\vec{q} + 2x\vec{P}}{q^0 + 2xP^0} \quad . \tag{2.16}$$

More details on the kinematics of the Breit frame can be found in [18].

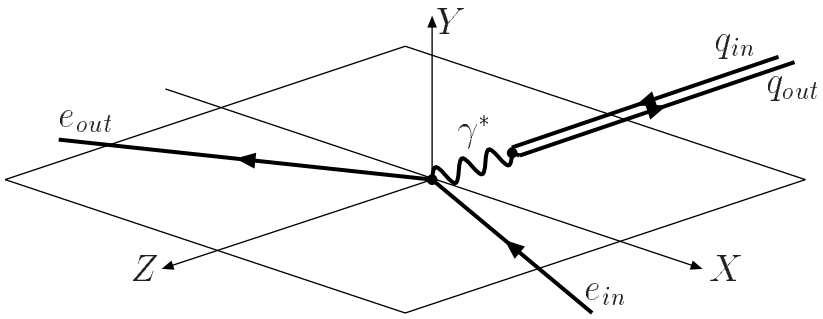

Figure 2.3: *Definition of the Breit frame.*

In an ideal case, all hadrons formed from the (outgoing) struck quark will be in the current fragmentation region of the BF defined as the hemisphere for $p_z^{BF} < 0$, while the fragments originating from the incoming parton and the proton remnant populate the target region, where $p_z^{BF} > 0$.

Note that the Breit frame is aligned with the HCM frame but boosted along a common z-direction. Hence the 2 reference frames are simply connected by a longitudinal boost, in which case the azimuthal angle is the same.

2.3 Azimuthal Distributions and Asymmetries

Azimuthal angle, ϕ, asymmetries of hadrons and jets in DIS are the main topic of interest in this thesis. This section outlines an explanation of the various properties that give rise to the non-trivial azimuthal angle dependence of the cross section, and presents the key results and predictions.

2.3.1 Non-perturbative Effects

Even within the framework of the QPM the consideration of an intrinsic parton transverse momentum gives rise to a ϕ dependence in the cross-section, resulting in an asymmetry around the momentum transfer direction. Indeed the cross section depends on the relative orientation of the lepton and quark planes[1]. This was first reported by Cahn [3, 4].

The "standard" partonic cross section reads

$$\frac{d^2\hat{\sigma}_i}{dxdy} = \frac{2\pi\alpha^2 e_i^2}{yQ^2}[1 + (1-y)^2] \tag{2.17}$$

(e_i is the charge of quark i in units of the proton's electric charge) from which one obtains, after some calculations, the standard expression for the differential DIS cross section given in equation 2.12.

Taking into account the intrinsic transverse momentum p_T of the initial state parton,

$$p = \left(\sqrt{p_T^2 + p_\parallel^2}, p_T\cos\phi, p_T\sin\phi, p_\parallel\right), \tag{2.18}$$

the cross section now exhibits a clear non-zero azimuthal angle dependence that adds to the previous expression:

$$\begin{aligned}\frac{d^2\hat{\sigma}_i}{dxdy} = \frac{2\pi\alpha^2 e_i^2}{yQ^2}\times \\ \Big[1 + (1-y)^2 + \frac{4p_T^2}{Q^2}(1-y) \\ -\frac{4p_T}{Q}\sqrt{1-y}(2-y)\cos\phi + \frac{4p_T^2}{Q^2}(1-y)\cos 2\phi\Big] \quad .\end{aligned} \tag{2.19}$$

Here ϕ is the azimuthal angle of the outgoing struck parton measured with respect to the lepton scattering plane.

The hadronic cross section corresponding to equation 2.12 is rather complicated in this case. One needs to convolute the partonic cross section with the parton distribution functions and the fragmentation functions. Moreover, the expression above has been given for a fixed value of p_T of the struck parton; a more realistic description should

[1] It is worth noticing that a parton plane can only be defined when the parton's intrinsic transverse momentum is not neglected.

model the transverse momentum distribution of the partons inside the proton. Further details are given in appendix A, which follows closely the papers by Chay, Ellis and Stirling [19, 20]. Cahn's parton level predictions are then modified by 2 factors:

1. even the fastest hadron will not carry all of the parton's momentum $\rightarrow p_T^{quark} \sim P_T^{hadron}/z_h$;

2. the strength of the azimuthal asymmetry will be reduced by the random transverse momentum added by the quark decay during fragmentation.

The parton level $\cos\phi$ and $\cos 2\phi$ moments of the distribution 2.19 are readily obtained by integration:

$$\langle\cos\phi\rangle = -\frac{2p_T}{Q}\frac{\sqrt{1-y}(2-y)}{1+(1-y)^2+\frac{4p_T^2}{Q^2}(1-y)} \quad , \tag{2.20}$$

$$\langle\cos 2\phi\rangle = \frac{2p_T^2}{Q^2}\frac{1-y}{1+(1-y)^2+\frac{4p_T^2}{Q^2}(1-y)} \quad . \tag{2.21}$$

They give a crude but useful estimation of the magnitude of the asymmetries and the sign: since the ratio p_T/Q is small for relatively high values of Q, the $\langle\cos 2\phi\rangle$ term is predicted to be smaller than the $\langle\cos\phi\rangle$ term, and both should rapidly decrease in magnitude when Q increases (for constant p_T). The signs of the moments are expected to be

$$\langle\cos\phi\rangle < 0 \qquad , \qquad \langle\cos 2\phi\rangle > 0 \quad . \tag{2.22}$$

2.3.2 Perturbative Effects

QCD is responsible for the perturbative effects. Two processes contribute to leading order QCD, the QCDC and BGF events depicted in figure 2.2. The $\mathcal{O}(\alpha_s)$ contribution to the cross section is given by

$$\frac{d^5\sigma}{dx\,dy\,dz\,dP_T\,d\phi} = A + B\cos\phi + C\cos 2\phi \quad . \tag{2.23}$$

Again, the ϕ dependence of the cross section is of the same form as in equation 2.19. This 5-fold differential cross section now depends on ϕ and 2 other variables needed to

specify the energy and angles of one of the outgoing partons. The latter 2 variables are defined covariantly as

$$x_p = \frac{Q^2}{2p.q} \tag{2.24}$$

and

$$z_p = \frac{p.p'}{p.q} \tag{2.25}$$

(p^μ and p'^μ are the 4-momenta of the incoming and outgoing struck partons). x_p is the partonic equivalent of the Bjorken x variable, and z_p can be related to the angular distribution of the 2 partons in their centre-of-mass frame.

The A, B and C coefficients in equation 2.23 are functions of x_p and z_p (and y). The full expressions are described in the papers by Chay, Ellis and Stirling [19, 20] and Méndez [21]; they are listed in appendix A for convenience and self-consistency.

Figure 2.4 shows the QCD predictions, as implemented in the Monte Carlo event generator LEPTO [22], for quarks and gluons in the case of the first order QCDC and BGF processes. One can conclude the following:

- For the QCDC process, the quark and the gluon have opposite distributions that cancel each other. The quark tends to be emitted in the opposite direction to the scattered electron whereas the gluon goes in the same direction. Moreover, the QCDC contribution is predominantly of the $\cos\phi$ type;

- For the BGF process, there is a perfect asymmetry between the quark and the antiquark, whose distributions can simply be added. Hence the only contribution is to the $\cos 2\phi$ moment.

In spite of the complexity of the expressions for the coefficients A, B and C in the ϕ distributions (equation 2.23), QCD predictions are simple: hadrons are preferentially produced in such a way as to give rise to a negative $\langle\cos\phi\rangle$ moment and a positive $\langle\cos 2\phi\rangle$ moment of the produced ϕ distribution with respect to the lepton scattering plane.

Whereas the non-perturbative contribution to the ϕ distribution moments falls off rapidly with Q, the leading order QCD corrections show a weak dependence on Q, and

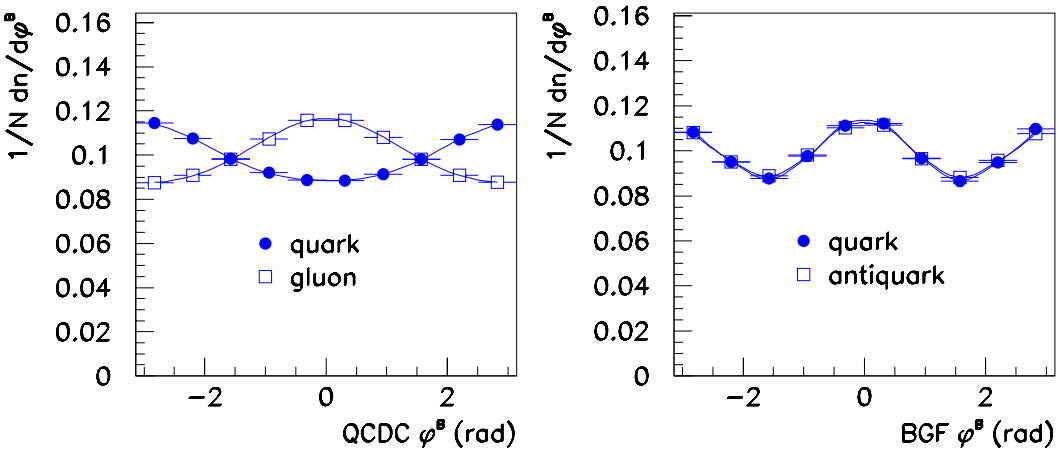

Figure 2.4: *ϕ distributions as predicted by QCD in the Breit frame for the outgoing partons in QCDC and BGF events. The distributions were obtained with the LEPTO Monte Carlo event generator.*

should therefore dominate in the high Q^2 region of the phase space. A hard cut on the transverse momentum of the hadrons of the final state should also suppress the non-perturbative contribution and allow for a cleaner study of the perturbative QCD behaviour. This is because the transverse momentum of the struck parton is limited to a few GeV, and the further contribution to a hadron's transverse momentum coming from the fragmentation process will fall off rapidly with increasing transverse momentum.

Next-to-leading order (NLO) QCD corrections to these predictions introduce additional $\sin\phi$ and $\sin 2\phi$ terms in the expression 2.23 of the cross section. But the latter are one to two orders of magnitude smaller than the $\cos\phi$ and $\cos 2\phi$ terms, and can safely be neglected.

2.3.3 Azimuthal Angle Dependence and Helicity Amplitudes

Remarkably, an azimuthal dependence of the cross section at $\mathcal{O}(\alpha_s)$ does not depend on the specific value of the spin of the interacting particles but is rather a consequence of their transverse momenta [3, 4][2]; an azimuthal asymmetry would still exist if, *e. g.*,

[2] "... the azimuthal dependence in vector exchange interactions is inevitable since the partons have transverse momentum as as consequence of being confined and such dependence certainly does not

the quarks were spin 0, although its characteristics would be different.

But the azimuthal angle dependence of the QCD contribution to the cross section, equation 2.23, is intimately related to the spin of the interacting particles, *i.e* with the spin of the exchanged boson and the struck quark. This is most easily seen decomposing the leading order QCD matrix elements in terms of the helicity amplitudes. In the hadronic centre-of-mass frame, the 3 polarisation vectors of the exchanged boson are

$$\epsilon^{\mu(\pm)} = \frac{1}{\sqrt{2}}(0, \pm 1, -i, 0) \tag{2.26}$$

and

$$\epsilon^{\mu(0)} = \frac{1}{\sqrt{2}}(0, 0, 0, 1) \tag{2.27}$$

$\epsilon^{\mu(+)}$, $\epsilon^{\mu(-)}$ and $\epsilon^{\mu(0)}$ refer, respectively, to a virtual boson polarisation of $\lambda = +1, -1$ and 0, the 2 transverse and the longitudinal components. The hadronic tensor $H^{\mu\nu}$ can always be decomposed in the basis of these polarisation vectors:

$$h_{\lambda\lambda'} = \epsilon_{\mu}^{*(\lambda)} H^{\mu\nu} \epsilon_{\nu}^{(\lambda')} \qquad (\lambda, \lambda' = +1, 0, -1) \tag{2.28}$$

The full calculation of equation 2.23 shows that [23]

1. coefficient A depends on the diagonal terms h_{00}, h_{++}, and h_{--};
2. coefficient B depends on the interference between the transverse and the logitudinal amplitudes h_{+0}, h_{0+}, h_{-0}, h_{0-};
3. coefficient C depends on the transverse interference amplitudes h_{+-} and h_{-+}.

require any special mechanism like gluon bremsstrahlung" [3, 4]

Chapter 3

The ZEUS Experiment at HERA

3.1 The HERA Accelerator

HERA, the Hadron Electron Ring Accelerator [24, 25], is an electron proton collider at the DESY laboratory, in Hamburg. Its facility presently houses four experiments: the two collider detectors H1 and ZEUS, and two dedicated experiments HERMES and HERA-B, to study the spin structure of the nucleon and CP violation with B mesons, respectively.

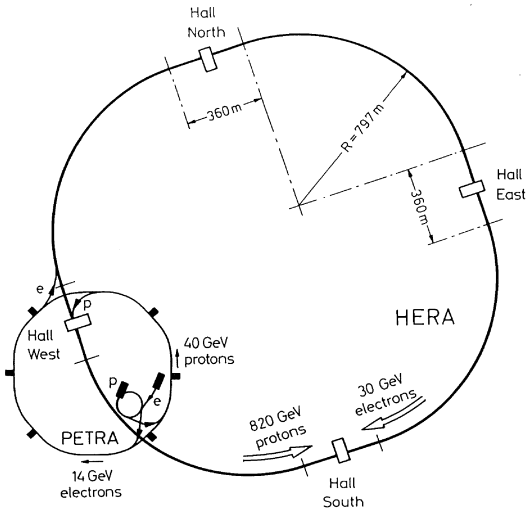

Figure 3.1: *Schematic layout of the HERA accelerator facility.*

The HERA tunnel is about 6.3 km in circumference. It was constructed underground,

at a depth of up to $25\ \mathrm{m}$, and occupies an area much larger than the DESY perimeter. The accelerator facility is composed of the bending sections separated by straight sections that accommodate the experimental halls. The layout of HERA is pictured in figure 3.1. The proton ring is mounted on top of the electron ring, except at the interaction points, where the beam bunches cross at $0°$ every $96\ \mathrm{ns}$. The transverse (longitudinal) size of the bunches is of the order of $50 - 200\ \mu\mathrm{m}$ (tens of centimeters) (see table 3.1).

HERA has provided positron-proton collisions since May 1992 at a centre-of-mass energy of about $\sqrt{s} \approx 300\ \mathrm{GeV}$, the positron and proton beam energies being of $26.7 - 27.5\ \mathrm{GeV}$ and $820\ \mathrm{GeV}$, respectively. Positrons were used instead of electrons in order to improve the lifetime of the lepton beam, since positrons, being positively charged, interact less with the residual positive ions in the beampipe. The proton beam energy was increased to $920\ \mathrm{GeV}$ in 1998.

Positrons and protons follow a series of acceleration steps before being injected into HERA with energies of $14\ \mathrm{GeV}$ and $40\ \mathrm{GeV}$, respectively, before ramping to full energy.

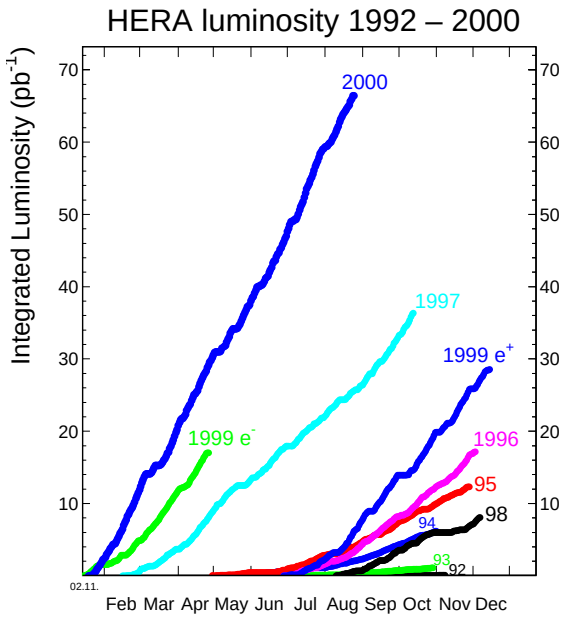

Figure 3.2: *The evolution over the years in integrated luminosity of the HERA accelerator.*

The design luminosity of HERA is $1.5 \times 10^{31}\ \mathrm{cm}^{-2}\ \mathrm{s}^{-1}$. Figure 3.2 gives a summary of

the evolution of the integrated luminosity delivered by HERA during the various running periods since ZEUS started taking data in May 1992. A steady increase of luminosity over the years can clearly be seen, demonstrating a good and improving performance of the accelerator.

The H1 experiment also started data taking in 1992, while HERMES only began in 1995. HERA-B began the operation of the detector in 1998.

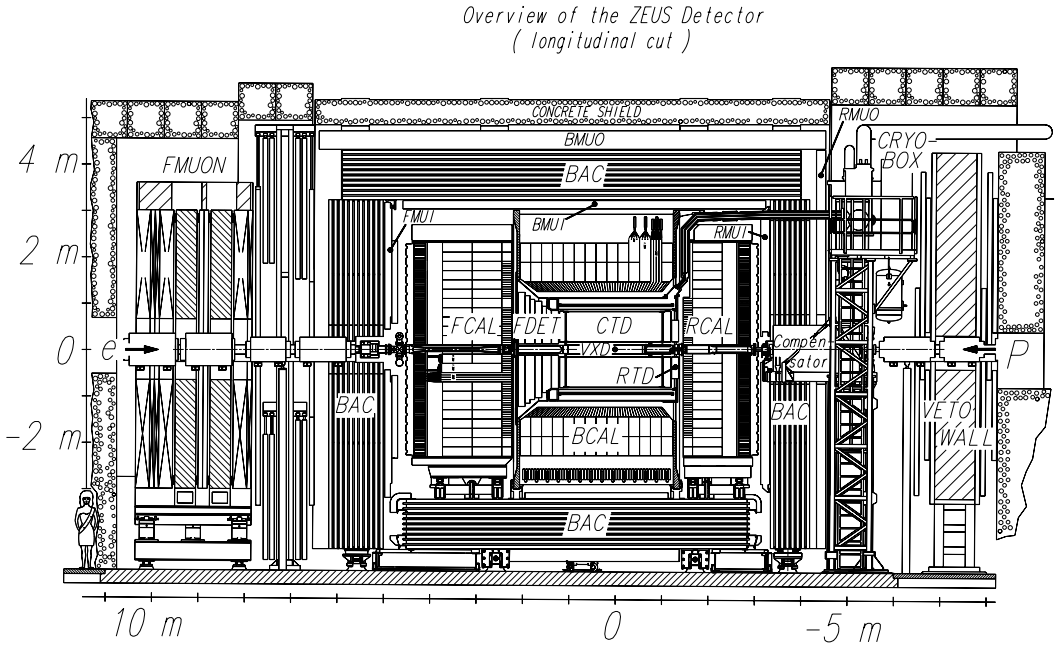

Figure 3.3: *Longitudinal view of the ZEUS detector.*

3.2 The ZEUS Detector

The ZEUS detector [26, 27] is housed in the south interaction region of the HERA accelerator facility (see figure 3.1). Longitudinal and transverse views of the detector are shown in figures 3.3 and 3.4, respectively.

Due to the large imbalance between the positron and the proton beam energies, the centre-of-mass of the colliding system is not at rest but moves in the proton direction. As a consequence, particles are produced mostly in the forward (proton) direction, and the design of the detector reflects that asymmetry.

The ZEUS detector components may be separated into several groups: the inner tracking chambers, the calorimeter system, the muon chambers, and a series of other dedicated smaller components.

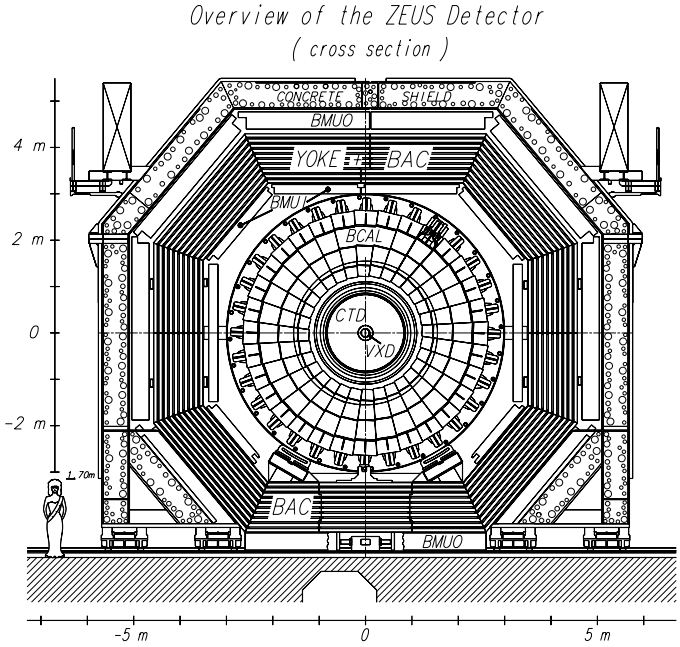

Figure 3.4: *Transverse view of the ZEUS detector.*

Closest to the beampipe is the tracking ensemble composed of a Central Tracking Detector (CTD), a Forward Detector (FDET) and a Rear Tracking Detector (RTD). The CTD is enclosed in a superconducting coil that provides a high magnetic field of 1.43 Tesla homogeneously in all but the most forward and rear regions. ZEUS also comprised a vertex detector (VXD) enclosed between the beampipe and the CTD; it was removed at the end of the 1995 running period due to serious radiation damages.

Outside the solenoid is the Uranium-scintillator Calorimeter (UCAL), also divided into three parts, barrel (BCAL), forward (FCAL) and rear (RCAL). Thick instrumented iron plates, the Backing Calorimeter (BAC), surround the UCAL and also function as a return yoke for the superconducting magnet.

Typical muons, of energy of a few GeV, normally pass through the calorimeter. Dedicated drift chambers effectively track high momentum muons. The muon chambers are divided into inner and outer detectors, with modules in each of the three regions.

The inner modules (BMUI, FMUI, RMUI) were installed inside the iron yoke, between the UCAL and the BAC. The outer modules (BMUO, FMUO, RMUO) were installed outside the iron yoke.

Other important components are the Veto Wall and the Luminosity Monitor (LUMI). The Veto Wall, situated about $7.5\ \mathrm{m}$ upstream of the interaction point, is composed of an iron wall with scintillator hodoscopes. It serves to reject background events and to protect the detector from the beam halo accompanying the proton bunches; in particular, it detects beam halo muons originating from decaying pions produced in the collisions of the protons with the residual gas in the beampipe.

The LUMI system [28] consists of two lead-scintillator calorimeters, an electron and a photon detector, situated about $35\ \mathrm{m}$ and $107\ \mathrm{m}$ to the rear of the nominal interaction point. They detect (in coincidence) the very low angle electron ($\theta_e \leq 6\ \mathrm{mrad}$) and photon ($\theta_\gamma \leq 0.5\ \mathrm{mrad}$) from the bremsstrahlung Bethe-Heitler process $ep \to ep\gamma$, for which the large cross section is known to a good precision, providing a means for the luminosity measurement. The electron detector is also used to tag very small angle electrons from photoproduction events, whereas the photon calorimeter helps identifying initial state radiation from the beam electrons.

The ZEUS coordinate system is a right-handed Cartesian system, with the z-axis pointing in the direction of the proton beam and the x-axis pointing towards the centre of the HERA accelerator. The origin is defined by the nominal interaction point.

A more detailed description of the main ZEUS components relevant to the analyses of this thesis follows in the next paragraphs.

3.2.1 The Tracking Detectors

The Central Tracking Detector (CTD) [29–31] is a cylindrical drift chamber designed to accurately detect and measure charged particle momenta. It has a length of about $2\ \mathrm{m}$, and an outer radius of about $80\ \mathrm{cm}$; it covers the angular range $15^\circ < \theta < 164^\circ$.

The chamber is filled with a mixture of argon, ethane, CO_2 and a trace of ethanol (less than 1%) optimised to provide a fast drift velocity around $50\ \mu\mathrm{m/ns}$, essential for

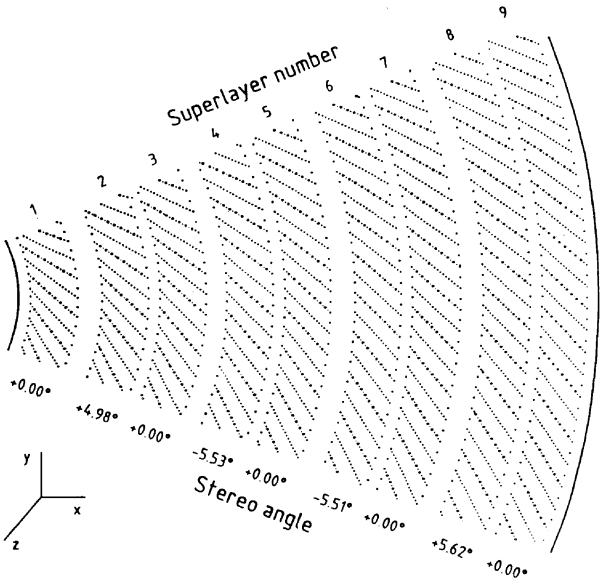

Figure 3.5: *An octant of the CTD with the details of the superlayers and the rows of sense wires (indicated by the large dots).*

a key role in triggering. It is composed of a total of 24192 wires, of which 4608 are sense wires that collect drifting electrons; the remaining are used to provide a uniform electric field. The chamber is partitioned into nine concentric "superlayers" (figure 3.5). And each superlayer is further divided in r–ϕ into cells, each comprising 8 (central) sense wires alternating with ground wires. The odd-numbered superlayers ("axial superlayers") have their wires parallel to the chamber's axis, whereas the other superlayers ("stereo superlayers") make an angle of about $\pm 5^\circ$, providing, in combination, a three-dimensional track reconstruction. The resolution of the CTD is $\frac{\sigma(p_T)}{p_T} = 0.0058 p_T \oplus 0.0065 \oplus 0.0014/p_T$ (p_T is expressed in $\mathrm{GeV/c}$) for long tracks traversing all superlayers [32].

The central row of eight sense wires of a cell is tilted by 45° with respect to the radius vector. This way, the electric field in the detector and the solenoid's magnetic field of 1.43 Tesla (the gas mixture is also a contributing factor) were adjusted such that the Lorentz angle is about 45°, in which case the electrons drift normally to the sense wires, i.e. purely in the ϕ direction. This tilt also makes it highly probable for a typical track from the interaction point to cross a plane of sense wires in each superlayer closely enough for the drift time of the first cell hit to be well below 96 ns, the beam

crossing interval. Hence this information can be used at the First Level Trigger (FLT) to correctly identify the bunch crossing number.

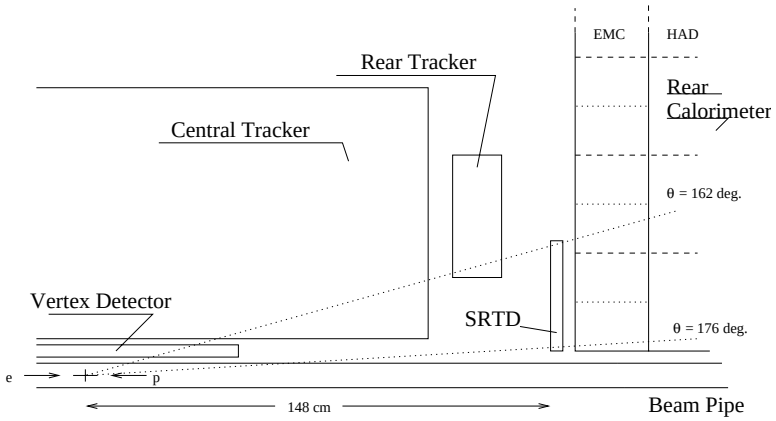

Figure 3.6: *Location of the SRTD in the ZEUS detector.*

The signals coming from the sense wires of the CTD are read out by two independent systems: the $r-\phi$ Flash Analogue-to-Digital Converter (FADC) and the Z-by-Timing (ZbyT) readout systems. Each of the sense wires provides hit information in the form of a pulse height and a drift time. The pulse height is digitised by a FADC. The ZbyT system is implemented in all sense wires of superlayer 1, and in the odd-numbered sense wires of superlayers 3 and 5. It provides a fast determination of the z-position – with a resolution around $4\ \mathrm{cm}$ – of a hit by measuring the difference in the arrival times of the signal at each end of the wire. This information is important input into the FLT. Hits from both systems are matched according to their drift times.

The CTD is not only physically at the centre of the detector but also occupies a central role: it is used to accurately determine the position of the event vertex, to detect charged tracks and measure their momenta, and to reconstruct secondary vertices and tracks from decaying products.

The Forward Detector (FDET) and the Rear Tracking Detector (RTD) extend the polar angle acceptance of the CTD. The FDET consists of three planar drift chambers (FTD) separated by two Transition Radiation Detectors (TRD). Its polar angle coverage spans from 28° to 7.5°. The RTD is composed of a single planar drift chamber covering, at the rear of the CTD, the polar angle range from 160° to 170°.

The Small Angle Rear Tracking Detector (SRTD) [33–35] was installed in 1994 between the RTD and the RCAL to improve both the tracks position and energy measurements in the rear region around the beampipe (figure 3.6). In particular, it detects pre-showering electrons, giving a way of correcting for the energy loss before reaching the calorimeter. Additionally, it provides fast timing information to the FLT and improves the rejection of background events from upstream proton interactions. The SRTD covers an area of $68\ \mathrm{cm} \times 68\ \mathrm{cm}$ (except for the beampipe hole) with a complete acceptance in the polar angle range between $167°$ and $174.5°$. It is composed of two overlapping layers of $5\ \mathrm{mm}$ thickness scintillator strips disposed horizontally and vertically (see figure 3.7). Light from the scintillator strips is directed through optical fibers to photomultipliers and then read out.

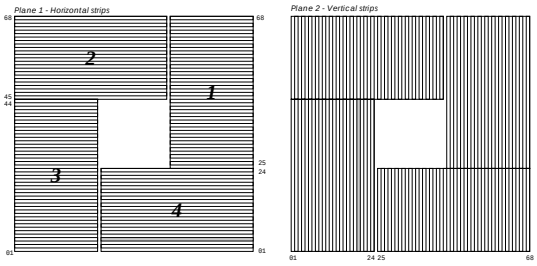

Figure 3.7: *The two layers of scintillator strips of the SRTD.*

3.2.2 The Uranium Calorimeter

The ZEUS calorimetry complements the information provided by the inner tracking chambers. It also measures high momentum tracks with better resolution than a tracking device and allows for the detection and position measurement of neutral particles. Furthermore, it plays a crucial role in the trigger chain: information from the calorimeter (CAL) is used from the first level trigger.

In a dense medium, electrons, photons and hadrons interact, producing showers of secondary particles. It is the energy of these secondary particles that is detected and measured. The ZEUS Uranium Calorimeter [36–38] is a compensating calorimeter: it was designed to respond as equally as possible to electromagnetic and hadronic energy

deposits. The precision in a calorimeter measurement is always limited by the unavoidable fluctuations in the shower development. Electromagnetic and hadronic showers behave differently; in particular, the energy fluctuations can be significantly greater for hadronic showers, and the ratio of electron-to-hadron response – for a given and same shower energy – is in general greater than one. This imbalance therefore needs to be compensated in order to obtain an equal response to both components. The ZEUS calorimeter achieves this compensation for the energy loss in hadronic showers due essentially to nuclear capture. It uses alternated $3.3\ \mathrm{mm}$ layers of depleted Uranium (^{238}U) with $2.6\ \mathrm{mm}$ layers of plastic scintillator whose thicknesses were optimised to achieve the above-mentioned ratio of 1.

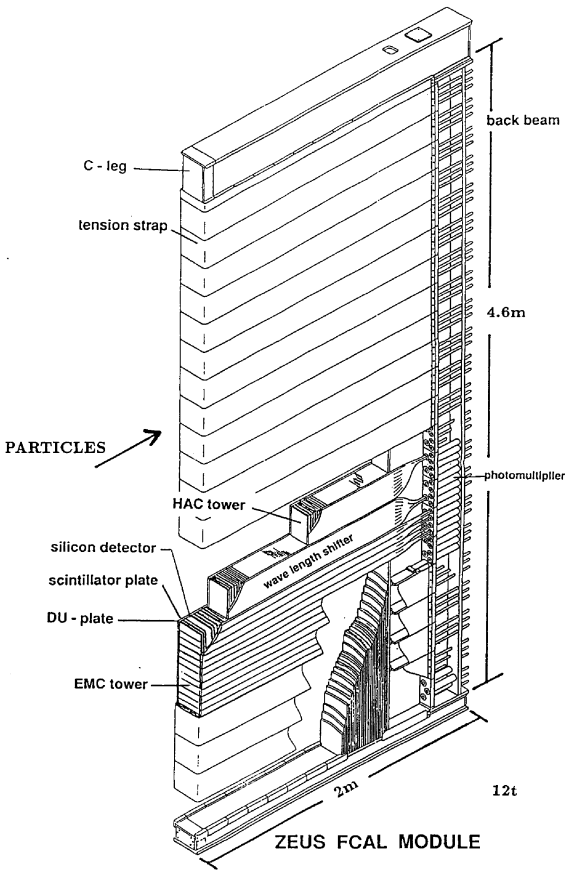

Figure 3.8: *Detail of a typical module of the forward calorimeter (FCAL).*

The UCAL is composed of a cylindrical barrel (BCAL), a forward (FCAL) and a rear (RCAL) regions. The three regions are divided into modules, further subdivided into

towers. An example of a typical module of the FCAL is displayed in figure 3.8. Each module comprises inner electromagnetic (EMC) and outer hadronic (HAC) sections.

Electrons and photons produce electromagnetic showers usually well contained in the EMC parts of the UCAL, whereas hadronic showers deposit relatively low amounts of energy in the EMCs, and travel farther into the HAC sections. The longitudinal segmentation of the CAL exploits this behaviour: the EMC sections have a thickness of 25 radiation lengths X_0; it is equivalent to 1 only interaction length λ_I for hadrons, so that hadrons deposit relatively little energy in the EMC. The HAC sections have a depth equivalent to 6 interaction lengths in the FCAL and BCAL and 3 in the RCAL. Electrons can therefore be distinguished from hadrons by the percentage of their total measured energy deposited in each section.

The UCAL provides a good hermetic calorimeter: the FCAL, BCAL and RCAL cover the polar angle ranges of $2.2°$ to $39.9°$, $36.7°$ to $129.1°$ and $128.1°$ to $176.5°$, respectively. The energy resolution measured from test beam data is $\sigma_E/E = 0.18/\sqrt{E}$ for electrons and $\sigma_E/E = 0.35/\sqrt{E}$ (E in GeV) for hadrons.

The light produced in the scintillator material is directed through wavelength-shifters and light guides to photomultipliers (PMTs), which convert and amplify the signal to be suitable for an electronic reading.

Two other components form the ZEUS calorimetry system, the Hadron-Electron Separator (HES) and the Backing Calorimeter (BAC). The HES improves the separation of the electron from the remaining of the final state and, in particular, from the other particles of a jet. This is achieved with the use of a higher degree of granularity than the one characterising the UCAL, allowing for a better comparison of electromagnetic and hadronic shower profiles. The HES consists of planar arrays of $3\ \mathrm{cm} \times 3\ \mathrm{cm}$ silicon diodes sandwiched between specific layers of the FCAL, BCAL and RCAL. The BAC completely surrounds the UCAL in order to measure energy leakage out of the latter. It also acts as a yoke for the ZEUS solenoid, returning the solenoid's magnetic flux. Its energy resolution is only of $\sigma_E/E = 1.1/\sqrt{E}$.

3.3 The Trigger and Data Acquisition System

The ZEUS trigger system [39] is an essential part of the detector. HERA bunches cross every 96 ns, and the total interaction rate is typically in the region of ~ 100 kHz, from which only a small fraction corresponds to interesting positron-proton collisions, that hence needs to be filtered from the "noise". This background comes mainly from beam-gas interactions between the proton beam and the residual molecules in the beampipe vacuum, from cosmic rays and from synchrotron radiation from the positron beam. Also the rate at which the selected events can be written to tape – about 3–5 Hz – imposes a severe rate reduction. The ZEUS trigger is therefore a complex data acquisition (DAQ) system designed to select interesting positron-proton events and to reduce the output down to an acceptable rate.

The trigger is divided into three levels that perform increasingly detailed event selection. Each detector component has its own trigger electronics which transfers the information to a global decision module. All information is combined at each trigger level to produce a decision.

The First Level Trigger (FLT) mainly rejects the bulk of background events. Each detector component uses a subset of its event data and sends preliminary information to the Global FLT (GFLT). The information is combined and a decision on whether to accept the event is issued 4.4 μs after the bunch crossing. Most importantly, a decision is made every 96 ns, so that the process remains deadtime free. As the 4.4 μs (corresponding to 46 bunch crossings) is large compared to the bunch crossing interval, the data needs to be pipelined in buffers. The output rate is reduced down to less than 1000 Hz, the maximum rate the Second Level Trigger (SLT) can handle.

The SLT has access to a larger amount of the full data, which is processed mostly on transputer networks, which perform iterative calculations. It can reject a greater amount of background events, reducing the rate to $20-100$ Hz.

Events accepted by the GSLT are passed on to the Event Builder (EVB); it receives all data from all detector components and produces an event record. The Third Level Trigger (TLT) then makes a final decision based on this information. It runs a fast subset

proton beam energy (GeV)	820	920
electron beam energy (GeV)	27.5	30
proton beam (mA)	105	140
electron beam (mA)	43	58
beam size σ_x/σ_y (μm)	200/54	118/32
luminosity ($\mathrm{cm}^{-2}\mathrm{s}^{-1}$)	1.5×10^{31}	7.4×10^{31}

Table 3.1: *Some HERA parameters as of 1997 compared to the goals of the upgrade.*

of the offline reconstruction program (ZEPHYR), reconstructing the event kinematics and classifying the accepted events into physics categories according to filters set by the various physics groups. The output rate is in the range of a few Hz.

The final stage of the reconstruction is done offline, typically a few days after data taking. The central tool for this full reconstruction is ZEPHYR, the ZEUS physics reconstruction program. It consists of a set of dedicated software packages, each of which refers to a specific detector component. During this process, the raw data as made available at the TLT level is combined with calibration information to produce calibrated physical quantities that are more suitable for physics analysis.

3.4 HERA and ZEUS Upgrades

The HERA accelerator and the collider experiments ZEUS and H1 have been upgraded during the long shutdown in the period 2000-2001. ZEUS has put a major effort into the design and installation of new tracking devices.

3.4.1 The HERA Accelerator Upgrade

It was the goal of the HERA machine upgrade to increase the luminosity by a factor of five and to provide longitudinal lepton beam polarisation to the two collider experiments ZEUS and H1. A total luminosity of $1\ \mathrm{fb}^{-1}$ is expected to be delivered during a six-year running period [40].

Some key parameters of HERA are listed in table 3.1 [40]. As can be seen from this

table, the significant increase in luminosity will be achieved by a better focusing of both beams close to the interaction region and by an increase in the beam currents.

The upgrade design aims at a longitudinal lepton beam polarisation of 70%.

3.4.2 The ZEUS Detector Upgrade

The HERA machine will offer the possibility to make a wealth of precise and new measurements in electron-proton scattering. Reference [41] reviews in detail the various physics motivations for the upgrade and the future potential physics program. The ZEUS upgrade reflects the main questions to be addressed in the future: precise determination of the structure functions and gluon density of the gluon in the proton, precise measurements of neutral and charged current cross sections, charm and beauty production, electroweak tests, jet physics and physics beyond the Standard Model (the list is by no means exhaustive).

These topics need, in general, good reconstruction of the scattered lepton and final state particles, and good heavy flavour tagging efficiency. All benefit from an increase in luminosity – to reduce the statistical fluctuations and errors – and others can only be addressed with large data samples.

A silicon micro-vertex detector (MVD) [42,43] has been installed between the beampipe and the CTD to improve tracking, and in particular to identify displaced (secondary) vertices and provide an efficient charm and beauty tag. The forward direction part of the ZEUS detector was complemented with the installation of planar straw tube trackers (STT) [44]. This tracking system will enhance the tracks reconstruction in the forward direction, where most particles travel.

Along with the luminosity increase, both the luminosity measurement and trigger systems will be upgraded. A precise luminosity measurement is essential if one aims at diminishing the systematic uncertainties on cross section measurements. On the other hand, a better and more efficient trigger is required in order to effectively select the most interesting ep scattering channels such as heavy flavour production and high E_T jet production.

Chapter 4

Monte Carlo Models and Detector Simulation

The description of the DIS hadronic final state is rather complex: the theoretical description of this process is perturbatively calculable at the partonic level, but the effects of the hadronic structure in the initial state and the hadronisation in the final state are not calculable from first principles. Moreover, analytical calculations become effectively impossible, specially when higher order corrections are needed. As particle production is statistical in nature, Monte Carlo (MC) programs provide a suitable probabilistic alternative to the analytical calculations; they simulate the random production of particles in a physics process according to the theory and phenomenological models.

Apart from the event generation, MCs are also essential in describing the data, namely the experimental distributions under study. The latter are smeared, distorted, etc. by the response, resolution and inefficiencies of the detector. It is necessary to correct for these effects before confronting the data to the theory. This process is often called unfolding. Furthermore, MC programs provide a crucial tool for the assessment of systematic errors of experimental measurements.

The full simulation of events in the ZEUS detector is performed in several steps: an event is first generated by an event generator such as LEPTO [22] or HERWIG [45], which results in a complete set of 4-vectors describing the final state particles. This

information is then fed into other MC programs, MOZART and ZGANA, which simulate the response of the detector and the trigger, respectively. This chapter will describe these 2 steps of the MC simulation, focusing on the MC programs used in the analyses.

4.1 Event Generation

A MC event generator is used to randomly produce physics processes according to the cross sections of the various contributing sub-processes. This procedure takes into account the matrix elements (ME) and higher order corrections of the scattering process, and hadronisation.

The generation of an event comprises several steps: the hard sub-process, as given by the matrix element (the leading order expression is used in LEPTO), the parton shower or QCD cascade, and the hadronisation. We briefly describe them it what follows.

4.1.1 Lepto

The LEPTO event generator [22] is used to simulate the hard partonic sub-processes according to the leading order (LO) matrix elements. The soft and collinear divergences in the cross sections are regulated with the introduction of lower and upper cuts on the z_p variable (*cf.* equation 2.25). Also the invariant mass squared $\hat{s}$ of the 2 outgoing partons is required to exceed a minimal value, $\hat{s}_{min}$. Below this region, the parton emissions are treated as parton showers, whose dynamics is governed by the DGLAP evolution equations.

The LEPTO event generator uses the leading logarithm approximation parton shower (PS) [46] approach for the higher order corrections of the fragmentation process. And it uses JETSET [47, 48] for the hadronisation phase (see section 4.1.4). Colour coherence is taken into account in soft gluon emission: the emissions are angular ordered, *i.e.* the emissions are at decreasing opening angles in the subsequent emissions.

4.1.2 Ariadne

Ariadne [49] can be used in conjunction with LEPTO (to generate the hard scattering) to treat the fragmentation in a different approach from the PS approach. It simulates the QCD parton cascade according to the Colour Dipole Model (CDM) [50]. Within the framework of the CDM, gluon emission is treated as radiation from a chain of independently radiating dipoles: the first dipole is formed by the point-like struck quark and the extended proton remnant. Two dipoles are hence formed, one between the gluon and the struck quark, and the other between the gluon and the proton remnant, and so forth.

Since the proton remnant is treated as an extended object, the coherence condition allows only for a fraction of this source to be involved in the gluon radiation. Also the probing photon is only capable of resolving the struck quark to a distance $\lambda \sim 1/Q$, and is therefore treated as an extended source. As a consequence gluon emissions are suppressed in the proton and photon directions.

We have used the latest version of Ariadne, validated in [11], which implements a high Q^2 modification developed to remove the previously observed suppression of radiation in the current region of the Breit frame at high Q^2 due to the extension of the proton remnant. I became involved in the study of the consequences resulting from these modifications made during the 1998/99 HERA Workshop on event generators, and provided a tuning of Ariadne to HERA DIS data [11].

The BGF process is introduced in Ariadne in a rather *ad hoc* way: the initial dipole between the struck quark and the proton remnant can either emit a gluon according to the ME, or an anti-partner is "emitted" according to the BGF matrix element.

4.1.3 DJANGOH

One-loop electroweak corrections and initial and final state QED radiation are incorporated in the HERACLES program [51,52]. The program is interfaced to LEPTO with the DJANGO program [53]. The set is hereafter referred to as DJANGOH.

4.1.4 The String Model of Hadronisation and JETSET

At the end of the parton shower evolution, the event consists of a collection of final state partons. These are coloured, and not observable. Hadrons are the observed final state particles. The transition from the partons to the hadrons is called hadronisation. Being a low-energy non-perturbative process, it needs to be modeled on the basis of the QCD properties.

Two distinct models of hadronisation are commonly used: the Lund String fragmentation model and the cluster model. The latter, incorporated in the HERWIG event generator, will not be discussed, as it is of no relevance to our analyses.

The Lund string model [54] is available in the JETSET MC program [47, 48]. In this model, the colour field that develops between 2 colour charges is stretched as they move apart, with a linear energy density, in accordance with the property of a linear confinement potential as given by lattice QCD calculations. As the distance between the 2 colour sources increases the field strength increases, until there is energy enough to produce another quark-antiquark pair. Thus more and smaller strings are produced in an iterative procedure until the energy available in each of them is below a (parameterised) production threshold. Each string gives rise to a hadron.

JETSET was used throughout the analyses to simulate the hadronisation process.

4.2 Detector Simulation

After MC simulation of the final state particles of a given physics process, one needs to take into account the effects introduced during detection and reconstruction. The behaviour and response of the ZEUS detector is simulated with the MOZART (Monte Carlo for ZEUS Analysis Reconstruction and Trigger) program, based on GEANT. GEANT is a software package with simulation tools for the description of the passage of particles in matter. MOZART fully simulates the interaction of the final state particles with the detector material as they propagate through from the interaction point. It includes our best knowledge of the geometry of the detector, the calibration, the intrinsic

resolutions of the various components, etc. The trigger system is also described in the detector simulation with ZGANA. The output of the simulation then comes in a form identical to the raw data. Hence the simulated events can be processed by the offline reconstruction software used for real data analysis, ZEPHYR.

4.3 Monte Carlo Samples

All the generated events were NC DIS events in the kinematic region defined by

$$0.007 < x < 0.3 \quad , \quad 0.13 < y < 1.0 \quad , \quad Q^2 > 100 \text{ GeV}^2 \quad .$$

The larger kinematic region compared to the analyses cuts actually applied stems from the necessity of avoiding the introduction of any bias when applying cuts in the MC studies. The CTEQ5D set of PDFs was used in all samples.

The Ariadne sample consisted of 200000 events (corresponding to an integrated luminosity of 108 pb^{-1}) for each of the 1996 and 1997 detector configurations and vertex distributions.

200000 events were also generated with LEPTO for 1996, and 400000 for 1997. But these extra 200000 events for 1997 were only used in the jets analysis, to provide a significant sample for the neural network studies.

4.3.1 Monte Carlo Programs and Azimuthal Asymmetries

It is important to remark that the available MC programs do not correctly implement the azimuthal angle asymmetries. In Lepto, an intrinsic parton transverse momentum distribution is part of the event simulation, but the dependence of the cross section on the azimuthal angle is not implemented.

Chapter 5

General Analyses Data Reconstruction and Selection

5.1 Outline

This chapter outlines the general procedure of data reconstruction and selection followed in the analyses. The methods employed for the reconstruction of the kinematics are first summarised. The general event reconstruction and selection is then described in detail; they are identical for both analyses. We also discuss the various corrections used during the reconstruction and analysis of each event. We finish with the exposition of the Monte Carlo description of the general event variables.

5.2 Reconstruction of the Kinematics

Different methods are available for the reconstruction of the event kinematics [55], which are used both during the offline reconstruction and in the analysis procedure. They are essential for an effective data selection of the true NC DIS events.

Due to a good hermiticity, the ZEUS detector allows for a simultaneous measurement of the kinematic variables from the scattered positron and from the hadronic final state. This redundancy of information can be used by defining several reconstruction methods

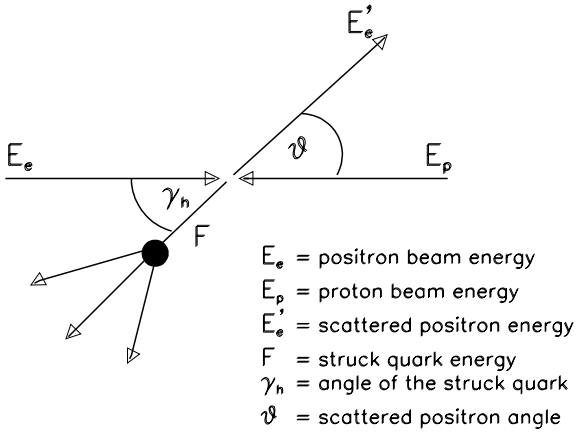

Figure 5.1: *Definitions of the kinematic variables of the incoming and outgoing positrons and the struck partons.*

for the variables x, y and Q^2; all use combinations of the following variables [55] (figure 5.1):

- the energy E'_e of the scattered positron;
- the polar angle θ_e of the scattered positron;
- the energy F of the struck quark;
- the polar angle γ_h of the struck quark.

The reconstruction of the angle γ_h is not straightforward. But this variable can in fact be easily and accurately reconstructed using the Jacquet-Blondel method (description follows in paragraph 5.2.2). We briefly describe the 3 methods used in the analyses.

5.2.1 Electron Method

The electron method reconstructs the kinematics solely from the energy E'_e and polar angle θ_e of the scattered positron:

$$y_e = 1 - \frac{E'_e}{2E_e}(1 - \cos\theta_e) \quad , \tag{5.1}$$

$$Q^2_e = 2E_e E'_e(1 + \cos\theta_e) \quad , \tag{5.2}$$

$$x_e = \frac{E_e}{E_p}\frac{E'_e(1 + \cos\theta_e)}{2E_e - E'_e(1 - \cos\theta_e)} \quad . \tag{5.3}$$

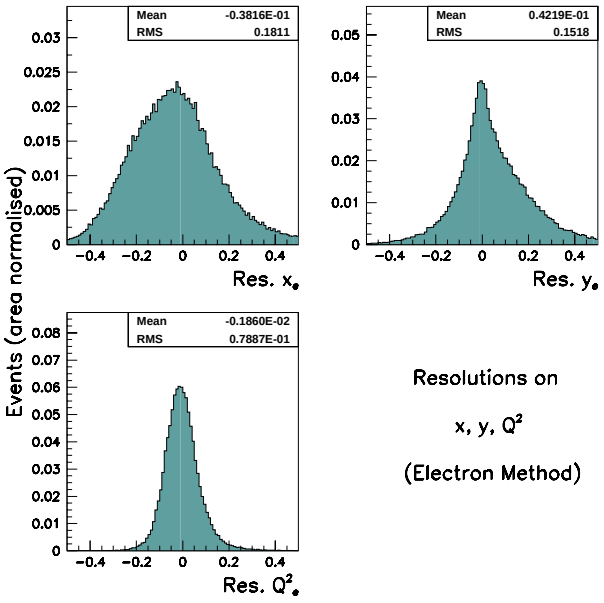

Figure 5.2: *Resolutions on the kinematic variables x, y and Q^2 as reconstructed with the electron method (kinematic range defined in section 5.6.2).*

The method gives mainly good results for Q^2 and at high y, when the scattered positron energy is significantly different from the incoming positron beam energy. At low y the resolution becomes poor and unstable, in particular for the determination of x, as can be concluded from the "theoretical" resolutions

$$\frac{\delta Q_e^2}{Q_e^2} = \frac{\delta E'_e}{E'_e} \oplus \tan\frac{\theta_e}{2}\delta\theta_e \quad , \tag{5.4}$$

$$\frac{\delta y_e}{y_e} = \left(1 - \frac{1}{y}\right)\frac{\delta E'_e}{E'_e} \oplus \left(\frac{1}{y} - 1\right)\cot\frac{\theta_e}{2}\delta\theta_e \tag{5.5}$$

and

$$\frac{\delta x_e}{x_e} = \frac{1}{y}\frac{\delta E'_e}{E'_e} \oplus \left[\tan\frac{\theta_e}{2} + \left(1 - \frac{1}{y}\right)\cot\frac{\theta_e}{2}\right]\delta\theta_e \tag{5.6}$$

($X \oplus Y$ stands for the sum in quadrature of X and Y). A precise estimation of the resolutions requires a detailed simulation of the detector response; the remarks above are hence only intended as a qualitative study of the accuracy of the method. After passing the data through the whole reconstruction chain, including the detector simulation and the event selection described in section 5.6, we obtained the resolutions pictured in figure 5.2 (the kinematic range is specified in section 5.6.2). They were calculated with the

Ariadne MC sample from the expression

$$\frac{x_e - x_{true}}{x_{true}} \tag{5.7}$$

(and *mutatis mutandis* for the other variables). Note that x_e, y_e and Q_e^2 were recalculated after a correction was made to the positron polar angle θ_e (*c.f.* section 5.3.2). As expected from the discussion above the resolution for Q_e^2 is rather good, but those for x_e and y_e are of the order of $15-18\%$.

5.2.2 Jacquet-Blondel Method

The energy F and polar angle γ_h of the struck quark can also be used instead of the positron variables to reconstruct the kinematics. From 4-momentum conservation it can be shown [55] that F and γ_h can be expressed, to a good approximation, as a function of the 4-momenta of the final state hadrons (of 4-momenta $(E_h, \vec{P}_h)$):

$$\begin{aligned} y_{JB} &= \frac{F(1-\cos\gamma_h)}{2E_e} \\ &= \frac{\sum_h (E_h - P_{zh})}{2E_e} \quad , \end{aligned} \tag{5.8}$$

$$Q^2_{JB} = \frac{(\sum_h P_{xh})^2 + (\sum_h P_{yh})^2}{1 - y_{JB}} \quad , \tag{5.9}$$

$$x_{JB} = \frac{Q^2_{JB}}{s y_{JB}} \tag{5.10}$$

and

$$\gamma_h = \frac{(\sum_h P_{xh})^2 + (\sum_h P_{yh})^2 - [\sum_h (E_h - P_{zh})]^2}{(\sum_h P_{xh})^2 + (\sum_h P_{yh})^2 + [\sum_h (E_h - P_{zh})]^2} \quad , \tag{5.11}$$

where $\sum_h$ stands for the sum over all detected hadrons in the final state. It is evaluated summing all but any energy deposits of the scattered positron in the calorimeter.

The Jacquet-Blondel method (JB method) [56] reconstructs the event kinematics from the measurements of the hadronic system, and is therefore used for the reconstruction of charged current events and events where the scattered positron escapes detection. It implicitly assumes a good containment of the hadronic system in the calorimeter. More

specifically, it assumes that the transverse momentum associated with the target remnant and the energy escaping through the rear beampipe are negligible (those particles lost in the rear direction – $z < 0$ – contribute double, since $E_h - P_{zh} \sim 2E_h$).

At low y, ($y \leq 0.04$), y_{JB} has a better resolution than y_e.

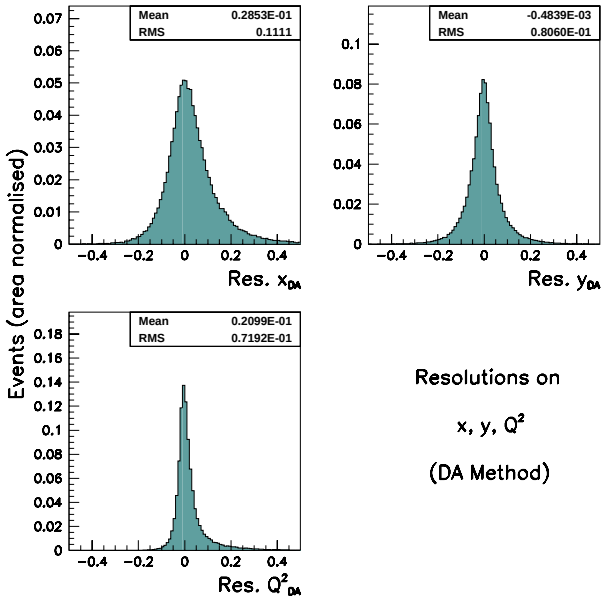

Figure 5.3: *Resolutions on the kinematic variables x, y and Q^2 as reconstructed with the DA method (kinematic range defined in section 5.6.2).*

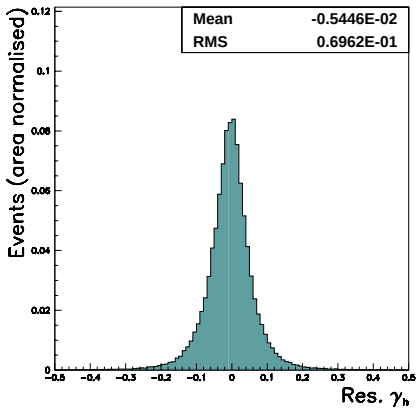

Figure 5.4: *Resolution on the reconstructed polar angle γ_h of the struck quark (kinematic range defined in section 5.6.2).*

5.2.3 Double Angle Method

This method is widely used in ZEUS for the reconstruction; it gives the best overall resolution in the variables x and Q^2 over most of the kinematic plane. It reconstructs the kinematic variables solely from the measured angle of the scattered positron and from the reconstructed polar angle of the struck quark:

$$y_{DA} = \frac{\sin\theta_e(1-\cos\gamma_h)}{\sin\theta_e + \sin\gamma_h - \sin(\theta_e+\gamma_h)} \quad , \tag{5.12}$$

$$Q^2_{DA} = 4E_e^2 \frac{\sin\gamma_h(1+\cos\theta_e)}{\sin\theta_e + \sin\gamma_h - \sin(\theta_e+\gamma_h)} \quad , \tag{5.13}$$

$$x_{DA} = \frac{E_e}{E_p}\frac{\sin\theta_e + \sin\gamma_h + \sin(\theta_e+\gamma_h)}{\sin\theta_e + \sin\gamma_h - \sin(\theta_e+\gamma_h)} \quad . \tag{5.14}$$

A clear advantage of the method resides in the fact that it is rather insensitive, in first approximation, to the overall scale of the energy measurements, *i.e.* to the overall energy scale of the calorimeter [55].

The resolutions on the kinematic variables reconstructed with the double angle (DA) method are summarised in figure 5.3. By comparison with figure 5.2, it is clear that the DA method shows better resolutions than the electron method for x, y and Q^2: the DA resolutions are, respectively, 11%, 8% and 7%, which are to be compared to the corresponding values of 18%, 15% and 8% obtained with the electron method. The good resolution on the reconstructed polar angle γ_h of the struck (equation 5.11) is given in figure 5.4; it is $\sim 7\%$.

For the above reasons the DA method was chosen for the reconstruction in the analyses.

5.3 Reconstruction of the Positron

5.3.1 Positron Finder

We have used the SINISTRA package [57] for finding and selecting the positron candidates. It is based on neural network architectures referred to as feed-forward multi-

layer perceptrons (refer, *e.g.* to [58] for an introduction to perceptrons and neural networks).

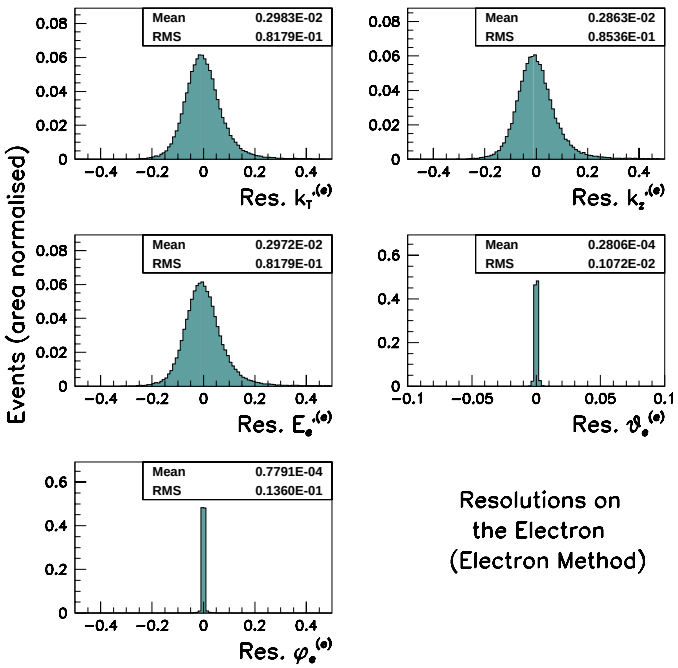

Figure 5.5: *Resolutions of the laboratory frame positron's 4-momentum as reconstructed with the electron method (denoted "(e)"). (k'_T refers to the positron's transverse momentum). The angular variables were corrected as explained in the text (kinematic range defined in section 5.6.2).*

The algorithm was trained on MC DIS events and uses information on the calorimeter energy deposits caused by the showering of particles. It distinguishes positrons from hadrons based on their respectively different shower profiles.

First, a clustering algorithm is run to find groups of energy deposits in the CAL. These clusters are interpreted as islands, with lower energy cells surrounding a rather central energetic deposit. It is required that at least 90% of the total cluster energy be inside the window, and that the energy outside be at most $1\ \mathrm{GeV}$ (a window being defined by the immediately surrounding cells of a given cell). These cuts reduce significantly the hadronic background with a corresponding loss of efficiency of the order of 1%.

Each cluster is then classified according to its (probable) origin: electromagnetic or hadronic. The SINISTRA package outputs a probability for a given energy cluster to

originate from a positron.

Positron candidates were selected with a SINISTRA probability above 0.9, which ensures a purity greater than 98% in the positron sample [57].

5.3.2 Momentum Reconstruction and Corrections

A direct measurement of the positron energy E'_e is obtained from the total energy of the calorimeter cells associated with it, as given by the SINISTRA positron finder. Also a first measurement of the positron angles θ_e and ϕ_e comes from the energy-weighted position of those same CAL energy deposits.

This CAL energy measurement has nevertheless the drawback that energy losses in the material in front of the CAL are responsible for an underestimation of the correct value. Corrections have been investigated at ZEUS for both the positron energy E'_e and polar angle θ_e. These standard correcting procedures make use of other detector components. Whenever available, preshowering energy detected in the SRTD or presampler is used to correct the positron energy. The angles are re-evaluated using other rather more precise information from the best of the available choices: (1^{st}) from tracking information, (2^{nd}) from the SRTD information, (3^{rd}) from the HES information, and (4^{th}) from the CAL information. The azimuthal angle ϕ_e of the positron was then calculated from the same information chosen for the reconstruction of θ_e.

The reconstruction of the positron 4-momentum from these corrected energy and angles is hereafter referred to as the electron method (denoted "(e)").

Alternatives exist for the reconstruction of the positron 4-momentum. The positron energy is commonly extracted in DIS studies from Q^2_{DA} (hereafter referred to as the DA method): using equations 5.2 and 5.13 it is easy to see that

$$\begin{aligned} E_e^{\prime(DA)} &= \frac{Q^2_{DA}}{2E_e(1+\cos\theta_e)} \\ &= 2E_e \frac{\sin\gamma_h}{\sin\theta_e + \sin\gamma_h - \sin(\theta_e+\gamma_h)} \quad . \end{aligned} \tag{5.15}$$

Figures 5.5 and 5.6 present the resolutions on the positron variables for the electron and the DA methods, respectively. For the purpose of a comparison, the resolutions

of the polar angles $\theta_e^{(DA)}$ and $\phi_e^{(DA)}$ refer to the uncorrected values calculated from CAL information only, whereas the variables $\theta_e^{(e)}$ and $\phi_e^{(e)}$ refer to the values corrected according to the prescription explained above. Both methods give good results: the DA method performs better in the reconstruction of the energy and 3-momentum, whereas the electron method gives excellent results for the reconstruction of the angles. As an example, the resolutions on the positron's transverse momentum k_T, longitudinal momentum k_z and energy E'_e are of the order of 7% and 8% for the DA and electron methods, respectively.

In our analyses the corrected values for the two angular variables, $\theta_e^{(e)}$ and $\phi_e^{(e)}$, were used in the previous equation for $E_e^{\prime(DA)}$ and also for the (central) 4-momentum determination, as this combination proves to be the best possible. In other words, the positron 4-momentum was reconstructed from $E_e^{\prime(DA)}$, $\theta_e^{(e)}$ and $\phi_e^{(e)}$.

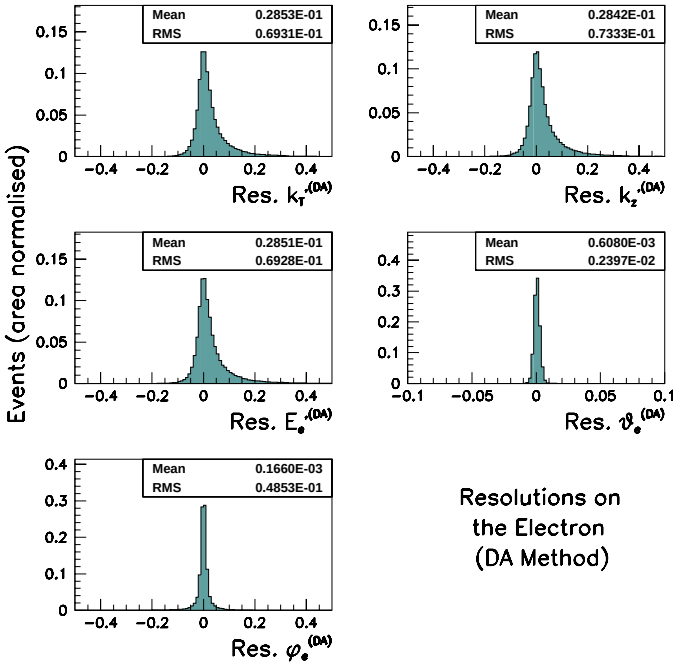

Figure 5.6: *Resolutions of the laboratory frame positron's momentum as reconstructed with the DA method (denoted "(DA)"). The angular variables displayed here were calculated solely from calorimeter information, as further explained in the text (kinematic range defined in section 5.6.2).*

5.4 Reconstruction of the Exchanged Boson

An accurate reconstruction of the exchanged boson is crucial to the present analyses; it is a prerequisite for a precise determination of the boost parameters. The photon's 4-momentum was calculated from

$$q^\mu = k^\mu - k'^{\mu(DA)} \quad . \tag{5.16}$$

Alternatively, as a cross-check, we also investigated the use of the electron method,

$$q^\mu = k^\mu - k'^{\mu(e)} \quad . \tag{5.17}$$

The sets of resolutions are shown in figures 5.7 and 5.8. They are of the order of 7% and 6% for the transverse momentum components for the electron and DA methods, respectively; but the boson's energy is clearly better reconstructed with the DA method compared to the electron method, the resolutions being 16% and 22%, respectively. Again these results strengthen our choice of the DA reconstruction method.

5.5 Reconstruction of the Boost Parameters

Be it a boost to the Breit or to the hadronic centre-of-mass (HCM) frame, the boosting parameters $\vec{\beta}$ are expressed as a function of the exchanged boson 4-momentum – usually a photon in our kinematic region – a good reconstruction of which is then essential. Experimentally the photon is reconstructed from the incoming and outgoing positron 4-vectors, and therefore using the DA method. The resolutions will be given in the following chapters in the descriptions of the analyses.

5.6 Event Selection

All the selection cuts described here were applied before any corrections were made to the data.

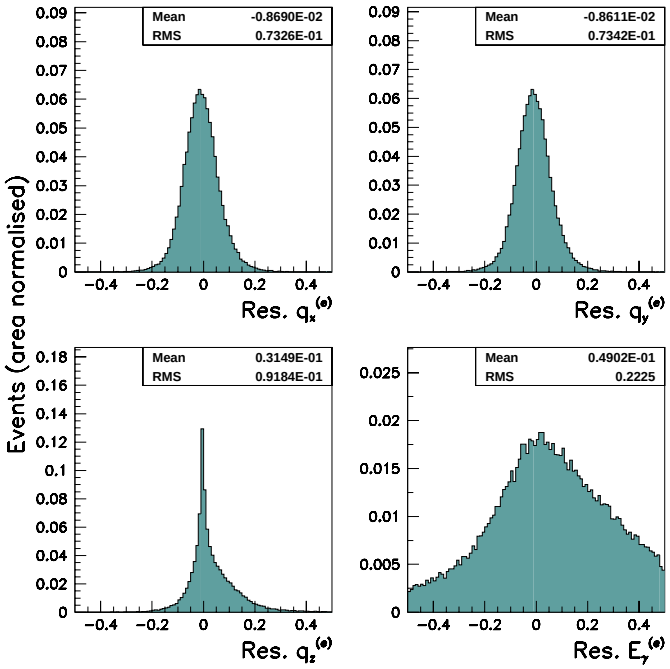

Figure 5.7: *Resolutions on the exchanged boson's 4-momentum as reconstructed with the electron method (kinematic range defined in section 5.6.2).*

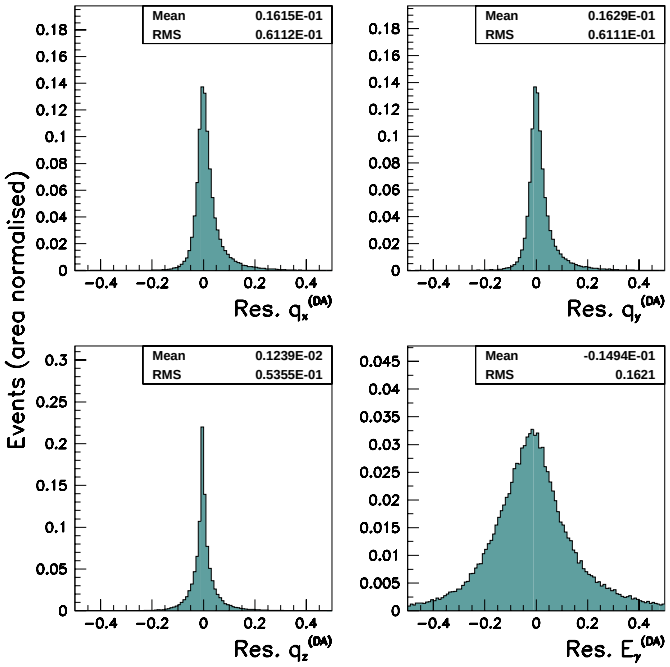

Figure 5.8: *Resolutions on the exchanged boson's 4-momentum as reconstructed with the DA method (kinematic range defined in section 5.6.2).*

5.6.1 Selection at Trigger Level

The selection at the trigger level is relatively loose, as only partial information is available. But during the full offline reconstruction the data is then labelled according to its type (NC DIS, D^* candidate, etc.) with so-called DST bits – Data Summary Tape bits.

Given the aim of the present analyses, we chose DST bit 12, which selects high Q^2 DIS events according to the following cuts, all applied before any corrections:

1. $Q^2_{DA} > 80 \text{ GeV}^2$;
2. $y_e < 0.95$;
3. Positron candidate found with $E'_e > 5 \text{ GeV}$;
4. $E - P_z + 2E_{\text{gam}} > 30 \text{ GeV}$ (E_{gam} is the energy deposited in the photon calorimeter of the LUMI detector).

5.6.2 Selection of the Phase Space

The analyses were performed in the region of phase space determined by

$$0.01 < x_{DA} < 0.1 \quad , \quad 0.2 < y_{DA} < 0.8 \quad . \tag{5.18}$$

It corresponds to the Q^2 range

$$181 < Q^2_{DA} < 7242 \text{ GeV}^2 \tag{5.19}$$

displayed in figure 5.9.

This region of the DIS phase space was selected with the goal of analysing a region where a clean study of the quark scattering and subsequent fragmentation can be performed: the region of relatively high x minimises the contribution of BGF events in the data sample and preferentially selects events originating from the scattering of the positron off a quark within the proton. The y selection ensures a high hadronic activity in the detector, a crucial requirement of any sensitive study of the hadronic final

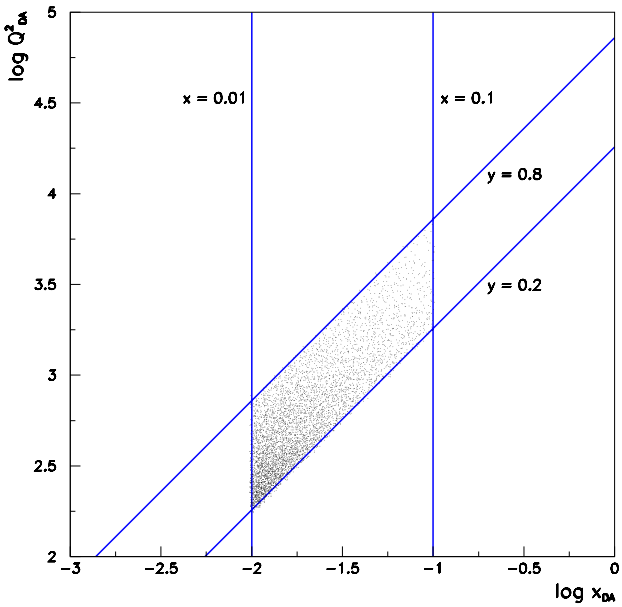

Figure 5.9: *Selected data sample in the kinematic phase space* (x, Q^2).

state. It also prevents the noise in the CAL from being relatively important compared to the energy deposit due to the hadronic final state. Furthermore, it also ensures that the hadrons be mainly produced in the central region of the detector where the tracks are well reconstructed within the region of acceptance of the CTD. This "central production", related to the γ_h angle, can be observed, for example, in figure 5.10.

5.6.3 Complete List of Selection Criteria

As mentioned previously (in section 3.3) the DIS events relevant to these analyses are overwhelmed by the bulk of "background" events – mainly from photoproduction. All the selections applied to the data aim at removing 3 categories of background: non-physics events, badly reconstructed events and non-DIS events (see table 5.1).

The various cuts are hereafter motivated and described in the next sections. The complete list is:

- **Cuts on the reconstructed kinematic variables:**

  1. $Q^2_{DA} > 100$ GeV^2

2. $y_e < 0.95$
3. $y_{JB} > 0.04$

- **Cuts on the positron:**
  1. A single positron found, with a matching vertexed track
  2. Positron energy $E'_e > 10$ GeV
  3. Positron position if in RCAL satisfying $|x_{elec}| > 16$ cm and $|y_{elec}| > 16$ cm (box cut)
- **Cuts on the vertex:**
  1. $-50 < z_{vtx} < 40$ cm
  2. $\sqrt{x^2_{vtx} + y^2_{vtx}} < 10$ cm
- **Other cleaning cuts:**
  1. $40 < (E - P_z)_{\text{total}} < 60$ GeV
  2. Good calorimeter timing
  3. Not a muon candidate
  4. Not a QED Compton event candidate
  5. Cut on luminosity monitor's gamma detector energy: $E_{\text{gam}} < 3$ GeV
  6. Cut on luminosity monitor's electron detector energy: $E_{\text{elec}} < 1$ GeV

Table 5.1 summarises the motivations behind the definition of the various selection criteria.

5.6.4 EVTAKE

The Data Quality Monitoring (DQM) is crucial in the whole data reconstruction chain. It is done both online and offline, and aims at identifying ranges of runs and events that had detector component problems, bad background conditions, etc. These

cut	non-physics background events	non-DIS background events		badly reconstructed events
		PHP	QED	
EVTAKE	×			×
$y_e < 0.95$		×		×
$y_{JB} > 0.04$				×
1 positron found E'_e box cut		×		
z_{vtx} r_{vtx}	×			×
$(E - P_z)_{\text{total}}$		×		×
$E_{\text{gam}} < 3$ GeV $E_{\text{elec}} < 1$ GeV		×	×	
QED rejection			×	
CAL timing	×			

Table 5.1: *Motivations behind the definition of the selection criteria. Each cut aims at removing specific types of events not relevant to the analyses (PHP stands for photoproduction).*

events are then flagged as inappropriate for data analysis, and recorded in the EVTAKE list. In our analyses we rejected all events not passing the EVTAKE filter.

5.6.5 Positron Selection

In its most simplistic form, a NC DIS event is characterised by the existence of a primary vertex and the presence of a scattered electron (positron).

The SINISTRA package was used to identify and select the positron (refer to section 5.3 for further details). The following selection cuts were dictated from studies of the performance of the positron finder, to give a reliable and efficient identification of the positron with a very low probability of misidentification.

The positron energy cut

$$E'_e > 10 \text{ GeV} \tag{5.20}$$

prevents the positron signature from being mimicked by EMC energy deposits due to the

decay of a π^0 to a pair of photons, $\pi^0 \to \gamma\gamma$, as the latter tend to have lower energies. It is to a great extent related to the behaviour and performance of the positron finder – the efficiency approaches 100% for positron energies above $10\ \mathrm{GeV}$.

It happens that more than one positron candidate is returned by SINISTRA; only events with one candidate were selected for analysis, for purity purposes.

The positron cluster was further requested to be matched to a CTD track originating from the event vertex: the distance of closest approach between the extrapolated vertexed track and the front face of the CAL was required to be less than $10\ \mathrm{cm}$. This cut is effective in rejecting high energy photon conversions and neutral pions, and improves the accuracy on the measurement of the angular variables of the scattered positron.

Also a box cut was applied on the position of the positron in the rear face of the CAL:

$$|x_{\mathrm{elec}}| > 16\ \mathrm{cm} \text{ and } |y_{\mathrm{elec}}| > 16\ \mathrm{cm}\,. \tag{5.21}$$

It prevents positrons scattered very close to the beampipe from being selected, since in such situations part of the showering energy may be lost. As a result both the positron's energy and position are badly measured.

The cuts just mentioned are imposed to achieve the best possible efficiency and purity. A further cut

$$y_e < 0.95 \tag{5.22}$$

is designed to suppress occurrences of a misidentified positron: events with a fake positron usually have very high y values as reconstructed with the electron method, as these misidentified positrons tend to be in the forward CAL.

5.6.6 Photoproduction Rejection

Due to the $1/Q^4$ behaviour of the ep cross-section, photoproduction is the main source of background physics processes to DIS. It corresponds to the scattering of a quasi-real photon, $Q^2 \approx 0$, off a proton.

Photoproduction events are characterised by the absence of a scattered electron detected in the calorimeter. Due to the non-detection of the electron, photoproduction

events tend to have a total $E - P_z$ value that is lower than the value for DIS. For a fully contained DIS event,

$$\sum_i (E_i - P_{z,i}) = (E_p - P_z) + (E_e - k_z) = 2E_e = 55 \text{ GeV}^2 \quad , \tag{5.23}$$

where the sum runs over all the final state particles (experimentally, the sum runs over all the calorimeter energy deposits). In case of initial state QED radiation the expression takes the form

$$\sum_i (E_i - P_{z,i}) + 2E_{\text{gam}} = 2E_e \quad , \tag{5.24}$$

E_{gam} being the energy measured in the photon calorimeter of the luminosity monitor. Note that final state radiation does not need any special calculation, since the accompanying photon will in general be very close to the electron, and will therefore be correctly taken into account in the sum over all CAL energy cells associated with the electron.

Events were required to be in the range

$$40 < (E - P_z)_{\text{total}} < 60 \text{ GeV} \quad , \tag{5.25}$$

to take into account the energy resolution of the CAL.

5.6.7 QED Compton Rejection

QED Compton events have a characteristic topology: the existence of 2 calorimeter energy deposits, back-to-back in azimuthal angle, one of which is matched to the electron track, the other corresponding to the photon.

They were removed offline during the analyses, using a dedicated algorithm that takes into account the characteristic topology mentioned above. Elastic QED Compton events were required to have at least 90% of the total CAL energy detected in the EMC section. Moreover, at most 3 GeV were allowed not to be associated with the 2 energy clusters. Also both clusters were required to be back-to-back (allowing for some flexibility as a result of finite resolutions). Inelastic QED Compton events have a rather more complicated topology because of the presence of a hadronic system.

5.6.8 Well-contained Vertex

A well-contained primary vertex is crucial for a correct reconstruction of the kinematics. Demanding a vertex also reduces considerably the background, which chiefly has 5 origins: beam-gas collisions between the proton bunches and the residual molecules of the beampipe vacuum, interactions of out-of-focus protons with the beamline components, synchrotron radiation from the electron beam, cosmic rays and halo muons coming from the decay of charged pions created in beam-related collisions. Fortunately, these undesired processes have characteristic signatures that allow for an effective identification and subsequent removal from the data sample. Cosmic rays, for instance, tend to have bad calorimeter timing; proton beam-gas interactions tend to have a fictitious upstream vertex, etc.

We required a reconstructed vertex in the range

$$-50 < z_{vtx} < 40 \text{ cm} \tag{5.26}$$

and well contained in the beampipe,

$$\sqrt{x_{vtx}^2 + y_{vtx}^2} < 10 \text{ cm} \quad . \tag{5.27}$$

The measured vertex distribution can be decomposed in a combination of a main peak and a series of secondary and smaller peaks – so-called satellite peaks – on each side of the main vertex peak. The asymmetric z-coordinate cut on the vertex position removes contributions from these satellite peaks.

5.6.9 Hadronic Activity

A significant hadronic activity is obviously needed in order for the hadronic final state to be well measured. Also the low-level noise coming from the uranium in the calorimeter needs to be suppressed to achieve this goal. At low y_{JB}, the hadronic jet is very close to the FCAL beampipe (*cf.* equation 5.8), in which case the signal from the residual CAL uranium makes a non-negligible contribution to y_{JB}, and hence distorts

its determination. As the related reconstruction of γ_h is essential in our analyses, such distortions are to be avoided. A cut

$$y_{\mathrm{JB}} > 0.04 \tag{5.28}$$

was applied to achieve this goal.

5.6.10 Other Cleaning Cuts

Other general cuts include calorimeter timing cuts. The excellent CAL timing resolution – better that 1 ns – provides an effective way of distinguishing between genuine ep collisions and beam-induced noise. Time in the various sections of the CAL is offset with respect to the HERA clock, so that the time in the various parts of the CAL is ≈ 0 for particles coming from a collision at the nominal interaction point. The application of cuts on the timing of signals detected in the various parts of the CAL effectively rejects background events. Furthermore, the pattern of the CAL energy deposits are also characteristic for the main sources of background, and are used to elimitate the latter. Cosmic muons, for instance, will be characterised by a non-zero absolute difference in the timings in the upper and lower parts of the barrel CAL.

Two cuts were especially introduced with the purpose of rejecting the bulk of initial state QED radiative events; they impose cuts on the amount of energy deposited in the calorimeters composing the luminosity detector:

$$E_{\mathrm{gam}} < 3 \text{ GeV} \quad , \quad E_{\mathrm{elec}} < 1 \text{ GeV} \quad . \tag{5.29}$$

These cuts reflect the fact that the bremsstrahlung photon of an initial state QED radiative event propagates at a very small angle and is finally detected in the photon calorimeter of the luminosity detector (*cf.* section 3.2 for details on this detector component).

5.7 Data Sample

The analyses presented in the following chapters are based on the data sample recorded with the ZEUS detector during the HERA data taking periods of 1996 and 1997. The total integrated luminosity amounts to ≈ 34 pb^{-1}, from which only a small sub-sample was selected for the analyses: of the ≈ 10.8 (27.9) pb^{-1} of data on tape for 1996 (1997) that passed the EVTAKE filter, only 3872 (10032) events were selected for the analyses after all the selection criteria discussed above. (Note that further cuts were applied, for example, to jets in the analyses of jets).

5.8 MC Description of the General Event Variables

This section will present the MC description of the data. Ariadne has been chosen as the primary MC sample for the determination of the bin-by-bin correction factors. The data will be compared to the Ariadne distributions. Lepto is also in reasonable agreement with the data, but Ariadne is better overall. This statement is even more true in the description of the tracking variables, which will be presented in the next chapter.

5.8.1 Kinematic Variables

There is a general good agreement between data and Ariadne for the distributions of the kinematic variables (figure 5.10). Only the distribution of y_e presents some clear distortions in the lower region around 0.3, well outside the region of application of the selection criterion (*cf.* paragraphs 5.6.1 and 5.6.5).

5.8.2 Positron Variables

The positron variables reconstructed with the DA method are well described by Ariadne in both the laboratory and the HCM frames, as can be inferred from figures 5.11

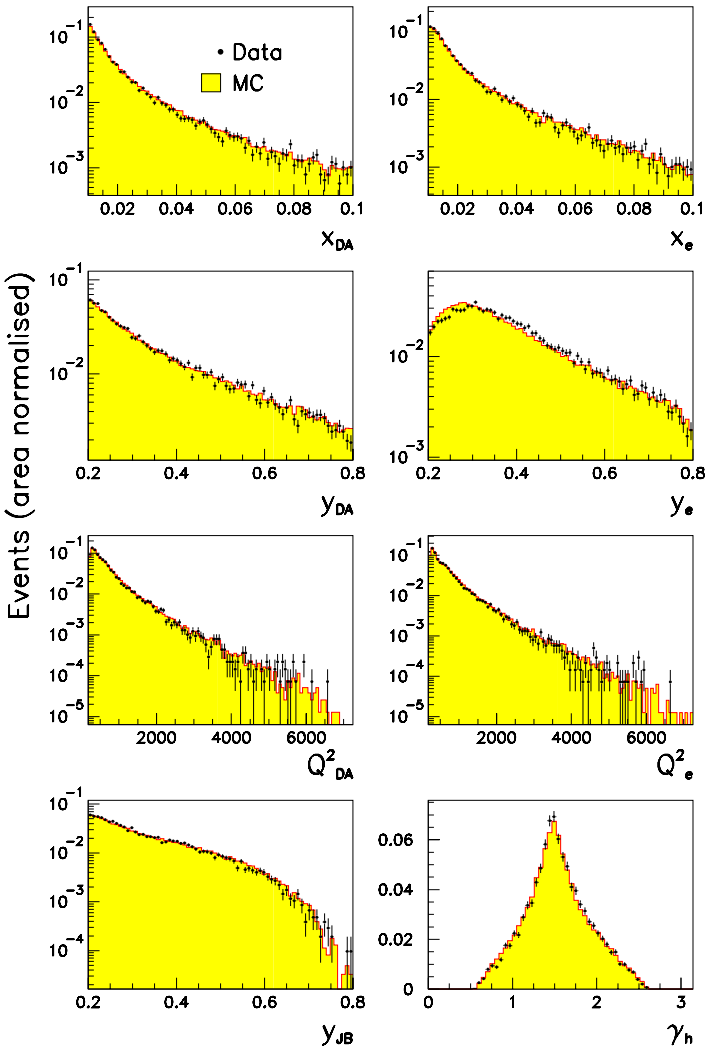

Figure 5.10: *Comparisons between the data and Ariadne for the kinematic variables in the laboratory frame.*

and 5.12. Not shown are the ϕ_e angle and the y-component of the positron momentum in the HCM frame, that are zero by definition. The same variables as reconstructed with the electron method are also well described (figure 5.11), except for a slight shift in the reconstructed positron energy; although the DA method provides a marginally better agreement between data and MC. Again the MC description of the data favours the choice of the DA reconstruction method.

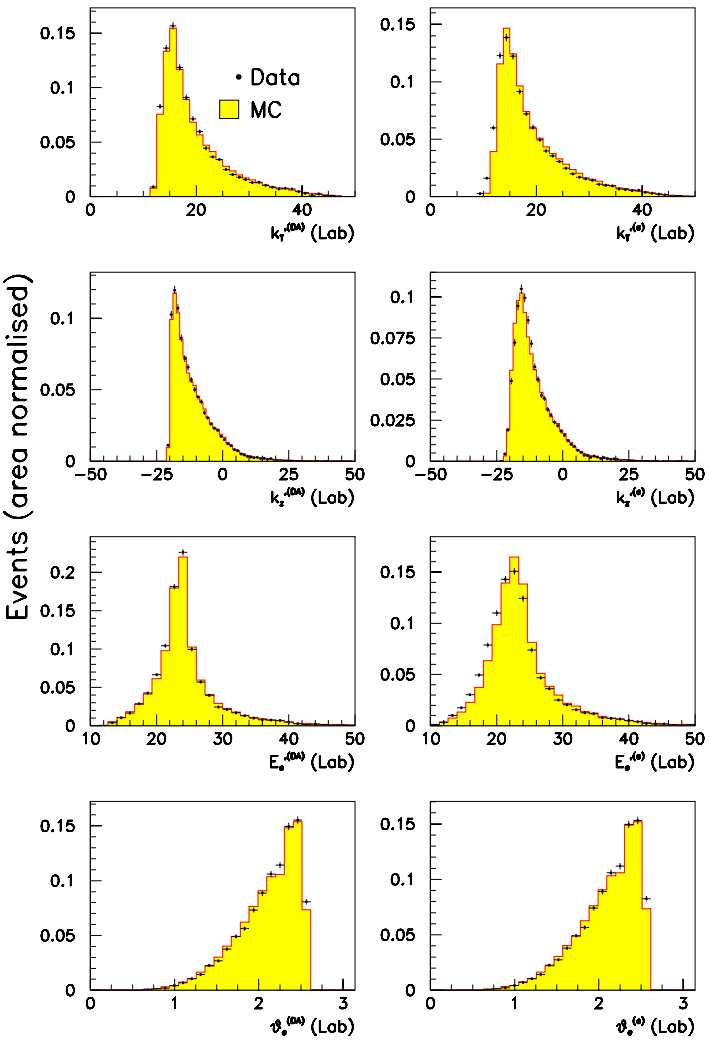

Figure 5.11: *Comparisons between the data and Ariadne for the positron variables in the laboratory frame as reconstructed with the DA and electron methods.*

5.8.3 Exchanged Boson Variables

The 4-momentum of the exchanged boson is also well described by both the reconstruction methods (figure 5.13). Here again, the DA method performs marginally better than the electron method.

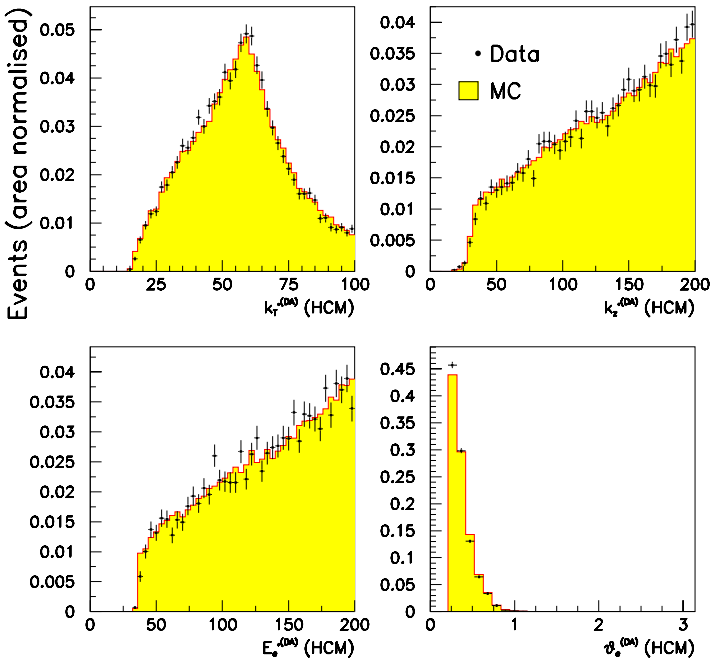

Figure 5.12: *Comparisons between the data and Ariadne for the positron variables in the HCM frame as reconstructed with the DA method.*

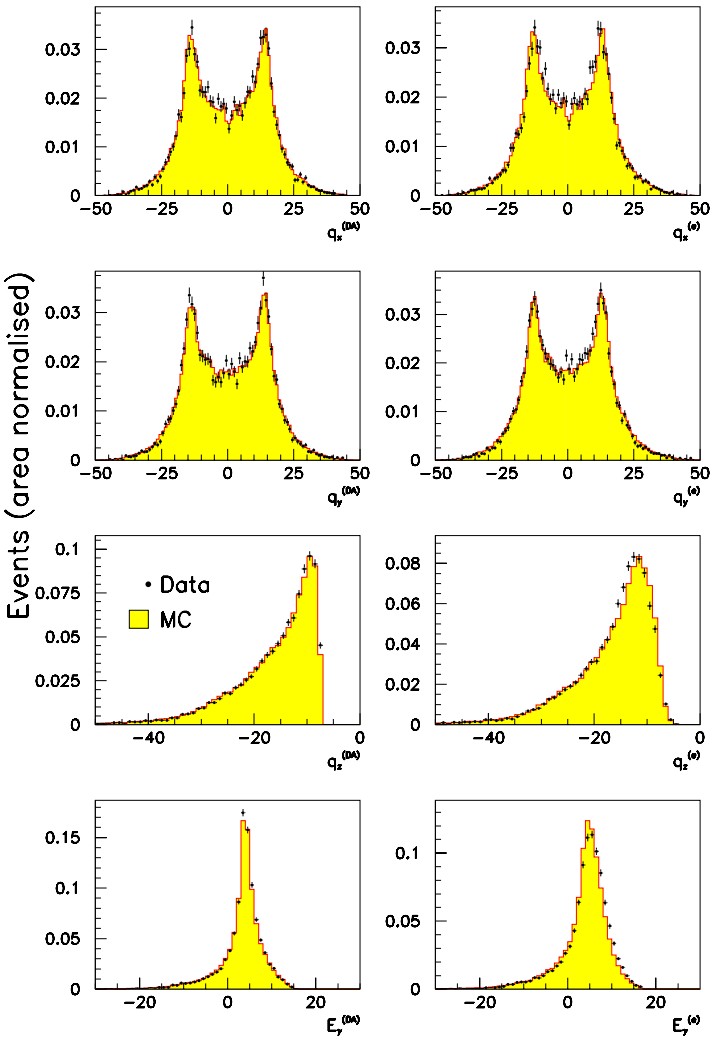

Figure 5.13: *Comparisons between the data and Ariadne for the exchanged boson variables in the laboratory frame as reconstructed with the DA and electron methods.*

Chapter 6

Analysis of Hadron Tracks

6.1 Outline

This chapter describes in detail the analysis on the azimuthal angle, ϕ, asymmetries of charged hadrons in the HCM frame.

The general event selection was described in detail in the last chapter. Here we focus on aspects relevant to the analysis of the charged hadron tracks – the boost to the HCM frame and the reconstruction and selection of the charged tracks. We first address these main aspects before presenting the corrections implemented. Finally we summarise the results for the azimuthal angle distributions and their moments.

6.2 Reconstruction of the Boost to the HCM Frame

The boost to the HCM frame required a careful treatment, as a bad reconstruction of the boost parameters can significantly distort the shape of the ϕ distributions. A correct reconstruction of the boost strongly depends on the determination of the positron's 4-momentum, and in particular on the reconstruction of its azimuthal angle. A good resolution around 6% and 0.2% was obtained for the transverse and the longitudinal components of the boost 3-vector $\vec{\beta}$ (figure 6.1).

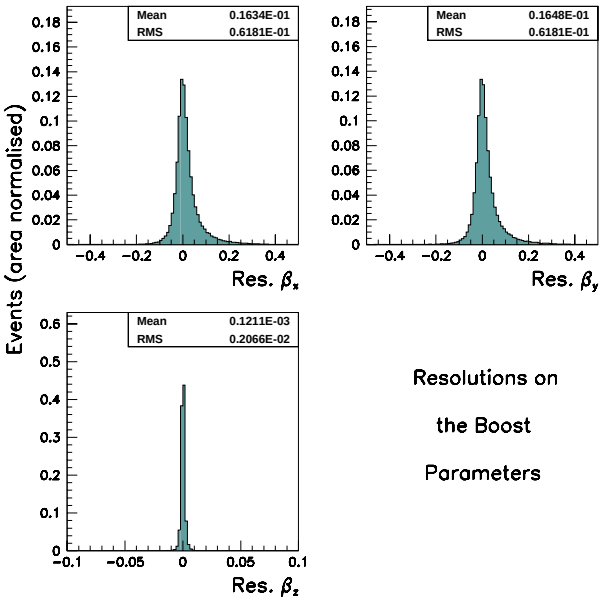

Figure 6.1: *Resolutions on the 3-vector $\vec{\beta}$ of the boost to the HCM frame (kinematic range defined in section 5.6.2).*

6.3 Reconstruction and Selection of Tracks

6.3.1 The VCTRAK Package

The reconstruction of tracks from the tracking chambers information is provided by the VCTRAK package [59].

The tracking algorithm runs twice over the set of tracking data – chamber hits. During the first pass only the tracks whose extrapolation to the innermost superlayer is possible are accepted. Then, shorter tracks that make it down to the first superlayer are found, and so forth, until tracks spanning one single superlayer have been found and reconstructed.

During the second step, those tracks with no hits in the innermost superlayer(s) are reconstructed if they span at least two axial superlayers and at least one intermediate stereo superlayer.

The reconstructed tracks are then subjected to a fit using a 5-parameter helix curve. The fitting based on the Kalman filter technique is done outwards starting at the vicinity

of the innermost measured hits, and taking into account the propagation through the magnetic field.

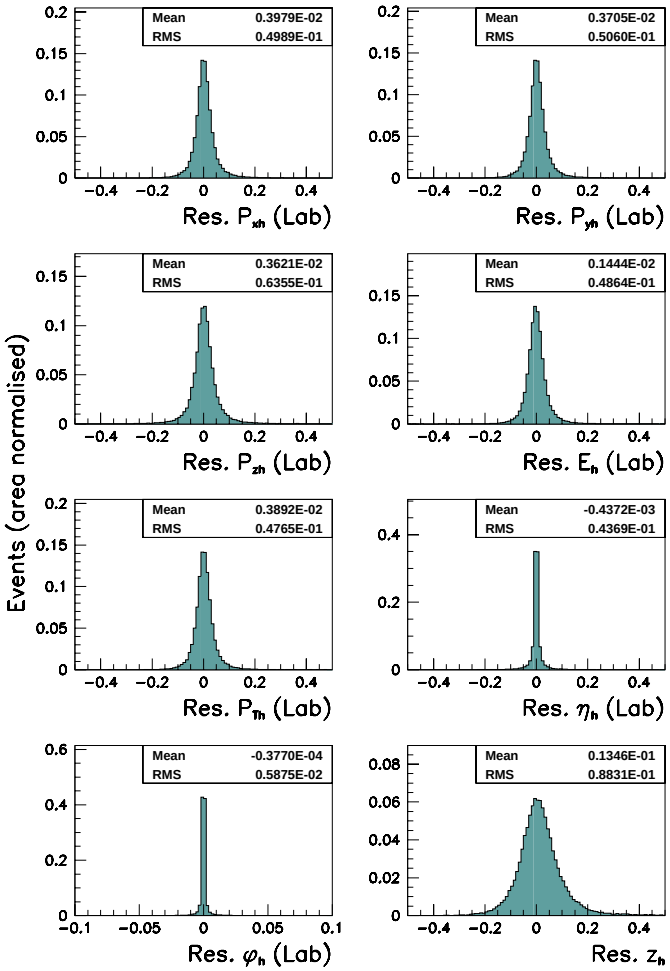

Figure 6.2: *Resolutions on the tracking variables in the laboratory frame.*

6.3.2 Tracks Selection

Tracks were selected according to the following cuts:

- track assigned to the primary vertex, not matched to the positron;
- $P_{Th}^{(lab)} > 150$ MeV;
- $\mid \eta_h^{(lab)} \mid < 1.75$;

- $z_h > 0.2$.

This set of cuts in the laboratory frame ensures the selection of only well-measured tracks; it reflects both the range of the CTD acceptance and the limitations of the VCTRAK reconstruction algorithm: the $\eta_h^{(lab)}$ cut selects tracks inside a region of the CTD where its behaviour is best understood, whereas the $P_{Th}^{(lab)}$ cut is needed to avoid very curved tracks that make the fitting procedure unreliable (the resolution of the CTD degrades at low transverse momentum, *cf.* section 3.2.1).

The z_h cut (the z_h variable was defined in section 2.1.2) mainly selects so-called leading particles. Due to the quark fragmentation functions being harder than the gluon fragmentation functions, this cut preferentially selects leading charged particles coming from the outgoing struck quark.

In what follows all track variables with no superscript will refer to quantities measured in the HCM frame. On occasion an explicit superscript will be introduced to avoid confusion.

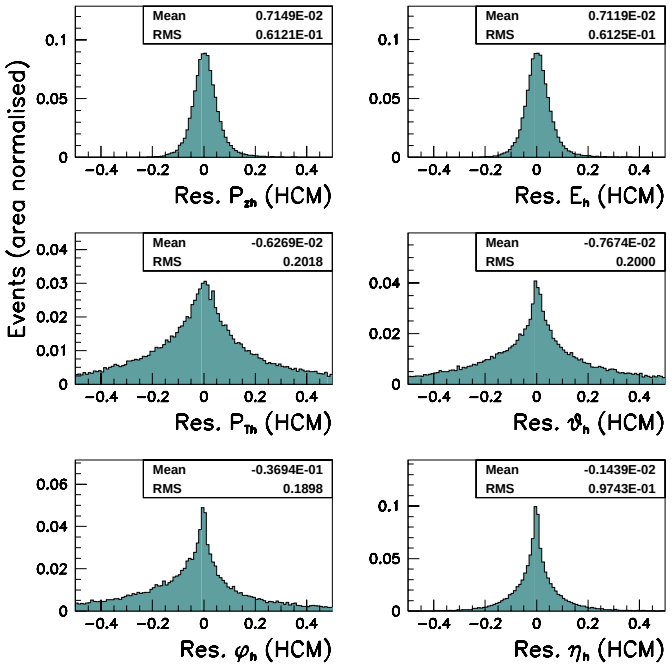

Figure 6.3: *Resolutions on the tracking variables in the HCM frame.*

6.3.3 Tracks Reconstruction

As described above, tracks are found and fit – in order to obtain their parameters such as momentum and angles – using the VCTRAK package. The tracks were reconstructed in the CTD with no particle identification. Those tracks passing the 3 selection criteria listed above, all applied in the laboratory frame, were then boosted to the HCM frame. They were assigned a pion mass for the purpose of performing the boost. It was checked, using MC studies, that this assignment is indeed correct for about 70% of the tracks that pass those 3 cuts used in the analysis, and that no systematic bias is introduced by this assumption. This is, in fact, the usual procedure in studies of charged particle production in DIS (*e.g.* in [9]).

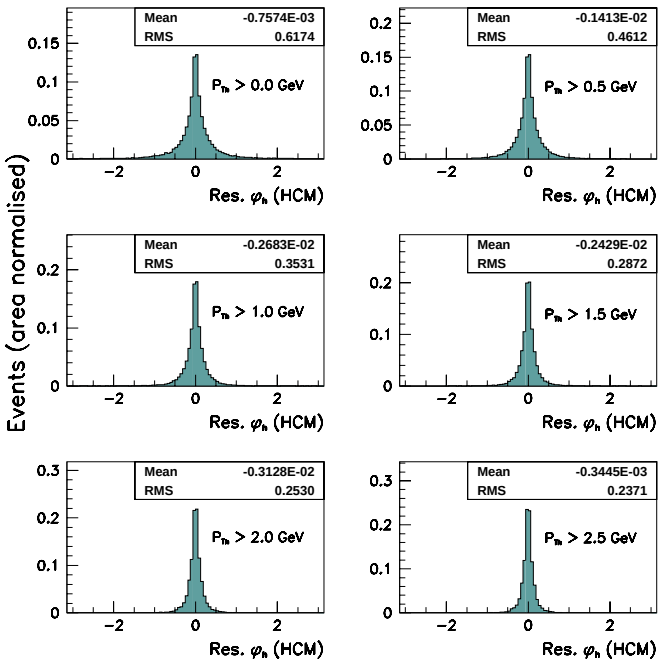

Figure 6.4: *Evolution with the transverse momentum P_{Th} cut (in the HCM) of the tracks of the resolution (defined as $\phi_{h,rec} - \phi_{h,gen}$) on the ϕ_h angle of the tracks.*

The assessment of the resolutions on the tracking variables requires the matching between the generated (true) particles and the reconstructed (observed) particles. This matching was performed in the following way: in the MC, one can relate a track to its

corresponding hits in the CTD. It is therefore possible to match a reconstructed to a generated track selecting the generated track that has more CTD hits in common with the given reconstructed track from the set of those sharing hits with it. The reconstructed (generated) tracks were boosted with the reconstructed (true) components of the boost.

The resolutions of the tracks variables in the laboratory frame are displayed in figure 6.2. The 4-momenta of the charged tracks are rather well measured, the resolutions being $\approx 5\%$. Also the azimuthal angle $\phi_h^{(lab)}$ and the z_h variable exhibit good resolutions of the order of 1% and 9%, respectively.

Figure 6.3 summarises the resolutions obtained for the most relevant tracking variables in the HCM frame. As expected, the resolutions degrade; this is particularly true for the transverse components of a track's momentum and hence for the ϕ_h angle.

Note that the resolution on the azimuthal angle of the tracks also depends on their transverse momenta: it improves as the transverse momentum P_{Th} of the tracks increases: figure 6.4 presents the evolution with P_{Th} of the resolution of ϕ_h (defined as $\phi_{h,rec} - \phi_{h,gen}$) in the HCM frame.

A good reconstruction of the tracks variables in the HCM mainly depends on the correct reconstruction of the tracks in the laboratory frame and on the correct determination of the boost parameters. We studied these 2 contributions separately.

In order to study the effect of the tracks reconstruction on the resolution of the tracks in the HCM frame, it is useful to boost the reconstructed tracks with the true components of the boost. On the other hand, the effect of the boost reconstruction on the resolution of the tracks in the HCM frame is investigated boosting the generated (true) tracks with the reconstructed components of the boost. The results are presented in figure 6.5 for ϕ_h. The distribution in 6.5(a) shows the resolution on the azimuthal angle ϕ_h of the hadrons in the HCM frame. The evolution of this resolution with the transverse momentum in the HCM, P_{Th}, is pictured in plot (b). The following profile plots, (c) and (d), present, respectively, the evolution with P_{Th} of the resolution on ϕ_h when performing the boost of the reconstructed tracks with the true components of the boost, and the boost of the generated tracks with the reconstructed components of the

boost. It is clear from these profile plots that the main contribution to the resolution of ϕ_h comes from the reconstruction of the boost rather than from the reconstruction of the tracks, which is, in fact, rather accurate.

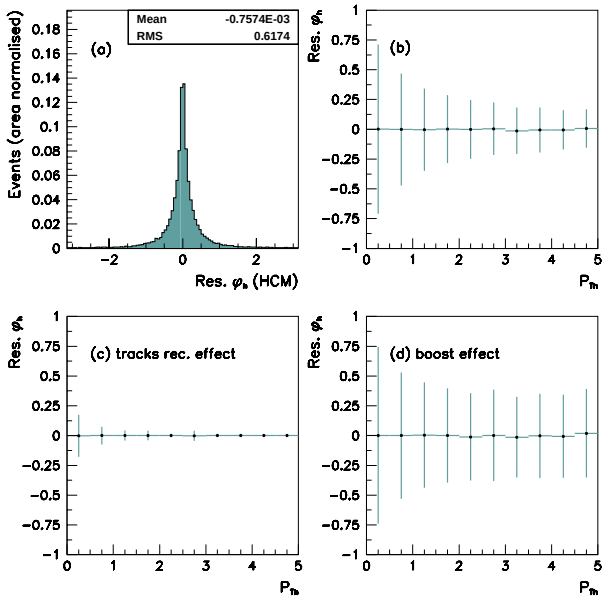

Figure 6.5: *Studies of the factors affecting the reconstruction of the azimuthal angle of the tracks in the HCM frame: (a) resolution on ϕ_h; (b) profile plot of the evolution with the transverse momentum P_{Th} of the tracks (in the HCM) of the ϕ_h resolution; (c) same as (b), but obtained boosting the reconstructed tracks with the true components of the boost, to assess the effect of the reconstruction of the tracks on the resolution of ϕ_h; (d) same as (b), but obtained boosting the generated tracks with the reconstructed components of the boost, to assess the effect of the reconstruction of the boost on the resolution of ϕ_h. Note that the error bars in plots (b)-(d) give the spread of the distribution and not the error on the mean. Refer to the text for further details and explanations.*

6.4 MC Description of the Data

This paragraph describes the comparisons between the data and the MC. Ariadne has been used throughout the analysis as default. Note that the comparisons refer to the complete 1996 and 1997 data set. The individual distributions for each year are not shown here separately, but the agreement between data and MC is very similar to

the one presented here in the next figures. All histograms have been normalised to the number of events.

The general event variables such as x, Q^2, etc. have been analysed in the previous chapter. Here we focus on aspects relevant to the present tracks analysis.

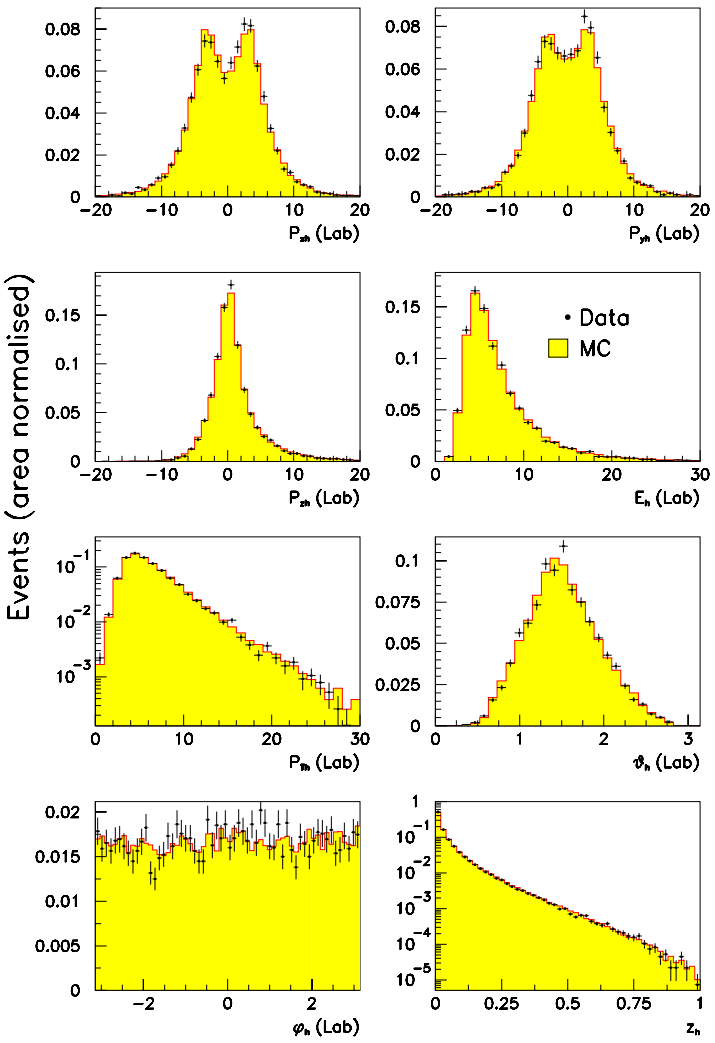

Figure 6.6: *Ariadne description of the tracking variables in the laboratory frame.*

6.4.1 Track Variables

Figures 6.6 and 6.7 show the rather good Ariadne modelling of the tracking variables in both the laboratory and the HCM frames. Some discrepancies can be seen in the

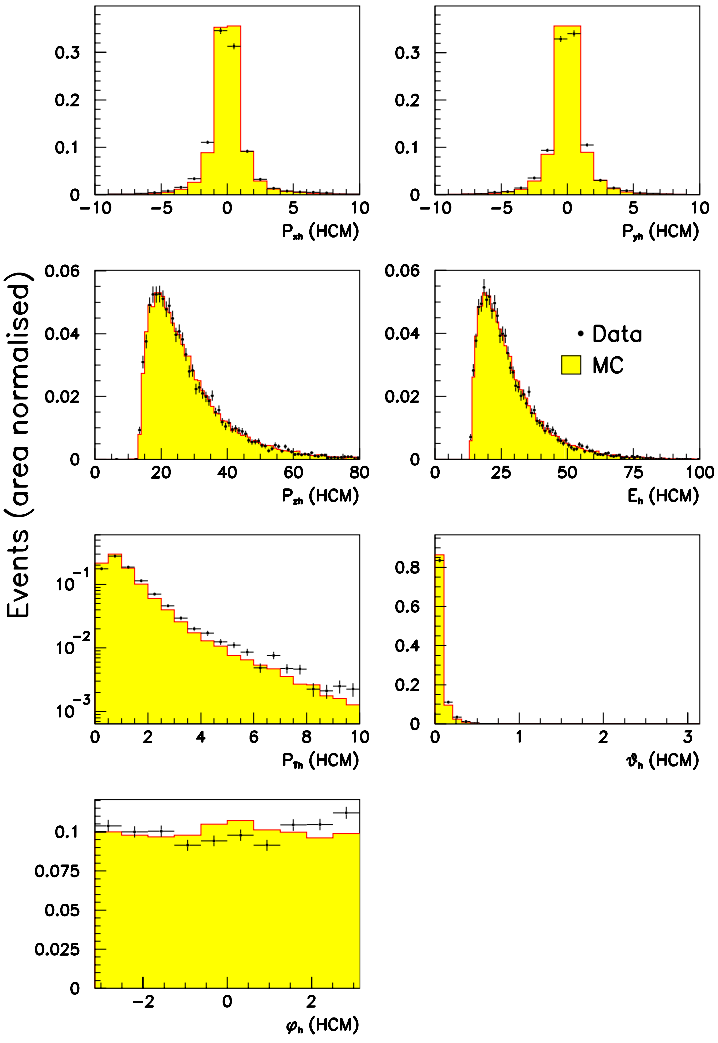

Figure 6.7: *Ariadne description of the tracking variables in the HCM.*

distributions in the HCM frame of the transverse momentum and the azimuthal angle ϕ_h of the tracks. The poor description of ϕ_h is a known feature of the MC [60].

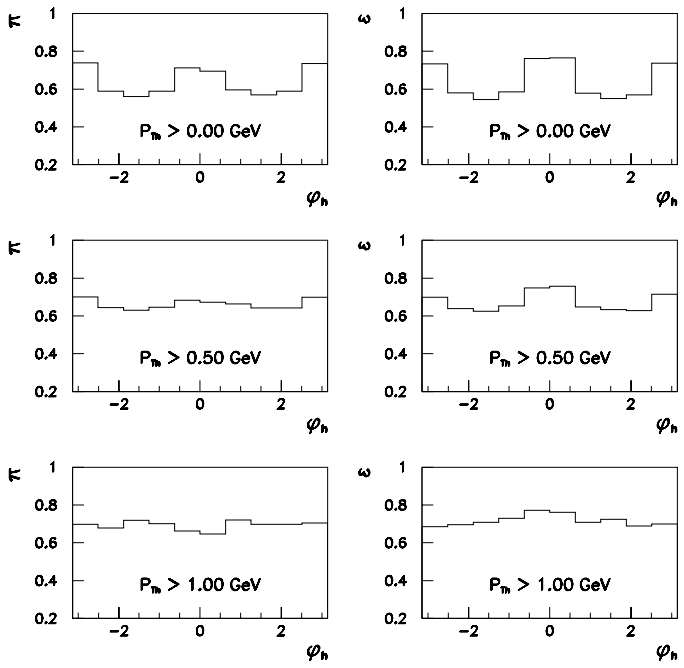

Figure 6.8: *Purities and efficiencies for the ϕ_h distributions of the tracks, for various cuts on their transverse momenta.*

6.5 The Raw ϕ_h Distributions

The azimuthal angle ϕ_h of a hadron is measured in the HCM frame with respect to the lepton scattering plane. If $\vec{k}_T$ and $\vec{P}_{Th}$ represent the transverse momenta (perpendicular to $\vec{q}$) of the incoming lepton and the observed hadron, respectively, then the azimuthal angle ϕ_h is simply the azimuthal angle between $\vec{k}_T$ and $\vec{P}_{Th}$. In the HCM, it corresponds to the azimuthal angle of the hadron measured from the x-axis in the $x-y$ plane.

The corrections to the raw distributions will be presented in the next paragraph. The raw ϕ_h distributions were first binned in accordance with the average resolution on the measured ϕ_h angle, of the order of 0.6 radians (*cf.* figure 6.4); therefore we chose to use 10 bins.

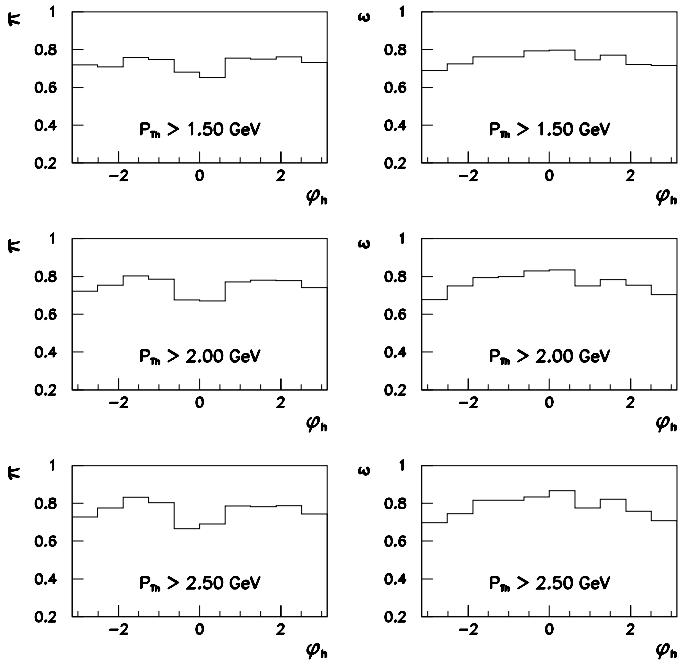

Figure 6.9: *Purities and efficiencies for the ϕ_h distributions of the tracks, for various cuts on their transverse momenta.*

The choice of the binning was also checked in terms of the purities and efficiencies, defined as

$$\pi(i) = \frac{\#\text{ of tracks generated and reconstructed in bin } i}{\#\text{ of tracks reconstructed in bin } i} \tag{6.1}$$

and

$$\epsilon(i) = \frac{\#\text{ of tracks generated and reconstructed in bin } i}{\#\text{ of tracks generated in bin } i} \quad , \tag{6.2}$$

respectively. They are shown in figures 6.8 and 6.9. Both purities and efficiencies are, on average, in the range $60 - 85\%$, which provides support in our choice of the binning in the azimuthal angle.

The observed non-trivial dependence on ϕ_h is an artefact of the P_{Th} resolution in the HCM. Events with a reconstructed P_{Th} smaller (greater) than the generated P_{Th} tend to correspond to true values of $\phi_h \approx 0, \pm\pi$ ($\phi_h \approx \pm\pi/2$). At low P_{Th} more reconstructed tracks have $P_{Th}^{rec} > P_{Th}^{gen}$, hence the dips at $\phi_h \approx \pm\pi/2$. At high P_{Th} the P_{Th} resolution tends to be slightly shifted towards $P_{Th}^{rec} < P_{Th}^{gen}$, which gives rise to dips around $\phi_h \approx 0, \pm\pi$.

The raw ϕ_h distributions as a function of the minimum P_{Th} of the tracks in the HCM frame are displayed in figure 6.10. The solid lines represent the result of a fit to the curve

$$\frac{d\sigma}{d\phi_h} = A + B\cos\phi_h + C\cos 2\phi_h + D\sin\phi_h \quad , \tag{6.3}$$

from which the moments of the ϕ_h distributions will be extracted; they are given by

$$\langle\cos\phi_h\rangle = B/2A \quad , \quad \langle\cos 2\phi_h\rangle = C/2A \quad , \quad \langle\sin\phi_h\rangle = D/2A \quad . \tag{6.4}$$

More details will be given in section 6.9.2 after the necessary corrections to the data have been applied.

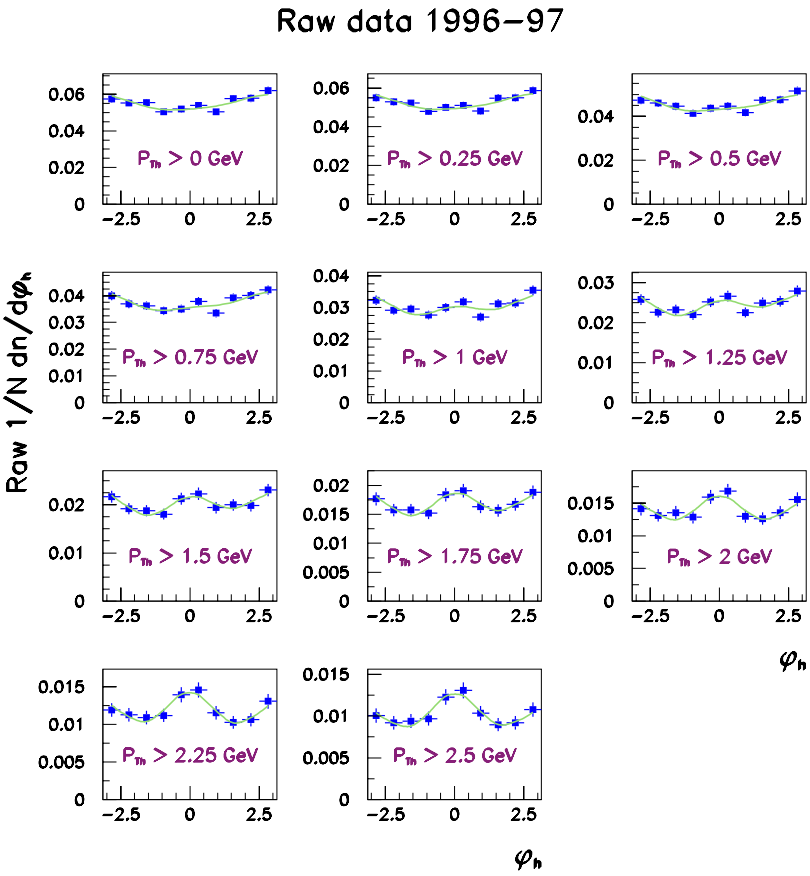

Figure 6.10: *The uncorrected ϕ_h distributions (raw data) in the HCM frame as a function of the minimal transverse momentum P_{Th} of the tracks. The lines are the result of a fit, and are merely to guide the eye (refer to the text for further details).*

6.6 Correcting Procedures

6.6.1 The Bin-by-bin Method

All measured distributions are distorted by the detector acceptance and inefficiencies, and effects introduced during reconstruction; they do not directly reflect the underlying theory. The process of unfolding is hence necessary to correct the data before any comparisons with theory and models can be envisaged. We chose for this analysis the bin-by-bin unfolding method.

The unfolding method relies on a complete Monte Carlo simulation of the DIS processes and takes into account the full simulation of the detector and the reconstruction procedure: the same reconstruction, selection and analysis was performed on this MC data as on real data. The distributions to be corrected are produced with both the generated (true) and the reconstructed (measured) information. In general they are related through a non-diagonal matrix, which incorporates all the above-mentioned effects, including migrations and correlations between the bins. The bin-by-bin unfolding method is based on the assumption that the correlations are negligible; it only considers the net migrations from or into bins. Hence the reconstructed and generated distributions are related by a simple (bin-by-bin) correction factor,

$$\Phi_i^{gen} = C_f(i)\Phi_i^{rec} \quad , \tag{6.5}$$

Φ_i^{rec} and Φ_i^{gen} being the reconstructed and generated distributions at bin i, respectively, and $C_f(i)$ the correction factor for bin i. These factors are given by the ratio, in the MC, of the generated to the reconstructed distributions.

The simulation has to describe well the data for the method to be viable. We used the Ariadne MC to perform the corrections, as it proved to give the best overall description of the data.

The raw distributions corrected using this bin-by-bin method will be presented after performing the systematic checks, the content of the next section.

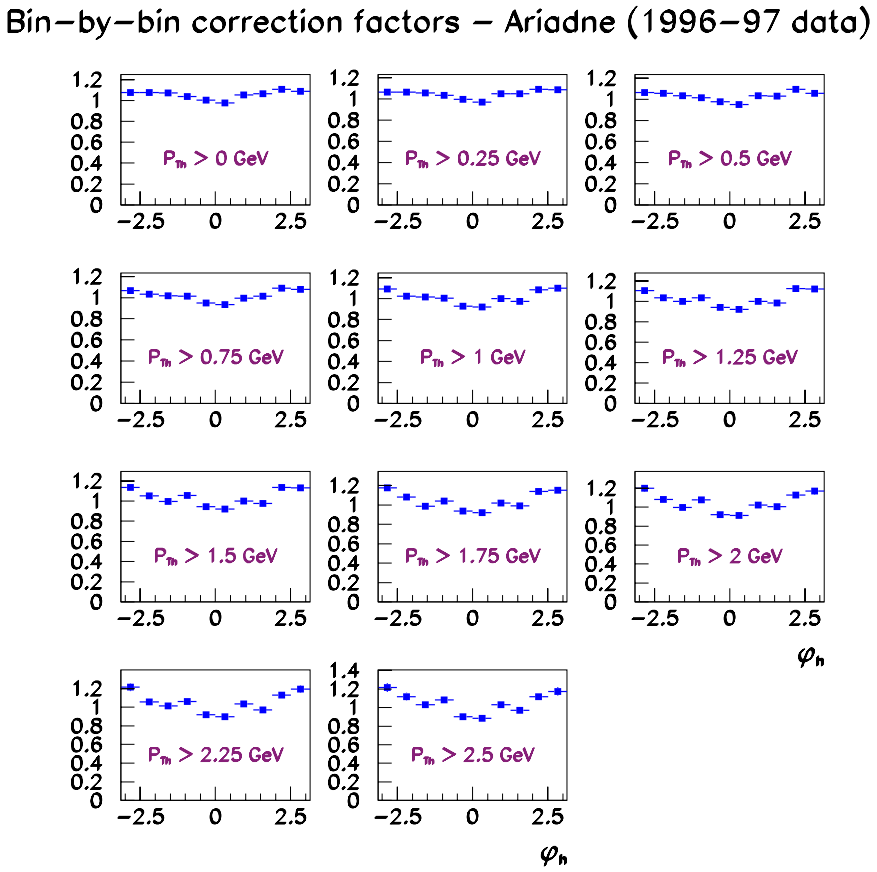

Figure 6.11: *Bin-by-bin correction factors for the ϕ_h distributions of the tracks.*

6.6.2 The Correction Factors

The correction factors for each of the ϕ_h distributions are defined from equation 6.5 as

$$C_f(\phi_h) = \frac{1}{N_{gen}} \left(\frac{dn}{d\phi_h} \right)_{gen} \Big/ \frac{1}{N_{rec}} \left(\frac{dn}{d\phi_h} \right)_{rec} \tag{6.6}$$

where N_{gen} and N_{rec} represent the number of generated and reconstructed events in the MC generated, $(dn/d\phi_h)_{gen}$, and reconstructed, $(dn/d\phi_h)_{rec}$, distributions.

The errors on the correction factors were carefully calculated taking into account the fact that the reconstructed and generated distributions entering the definition of the correction factors are not statistically independent (the bin in which a particle is reconstructed depends on the bin it was generated in). The correct prescription is detailed in [61].

As shown in figure 6.11, the bin-by-bin correction factors are not flat, but exhibit some ϕ_h dependence, which becomes more prominent as the P_{Th} cut on the tracks increases. They are in the (reasonable) range $0.9-1.2$.

6.7 Systematic Checks

Studies of systematic effects are essential to tackle the problem of determining the sensitivity of the results to the selection criteria and reconstruction methods.

The systematic checks can be grouped in 3 classes: (1) systematics due to the event reconstruction and selection, (2) systematics due to the tracks reconstruction and selection and (3) systematics due to the MC model used for correcting the data.

A series of systematic checks were performed on the (central) results:

1. lower the low cut on x by 1σ
   (1σ represents the resolution – RMS of the distribution – on the variable);
2. raise the high cut on x by 1σ;
3. lower the low cut on y by 1σ;
4. raise the high cut on y by 1σ;
5. lower the low z_h cut by 1σ;
6. apply no $\eta_h^{(lab)}$ cut to the tracks;
7. rescale the energy of the hadronic final state (measured in the calorimeter) in the MC by $+2\%/+2\%/+3\%$ in the F/B/RCAL;
8. rescale the energy of the hadronic final state (measured in the calorimeter) in the MC by $-2\%/-2\%/-3\%$ in the F/B/RCAL;
9. calculate the boost parameters using the angular variables as obtained from the CAL information instead of the "best" information (as described in section 5.3.2);

10. apply no energy corrections to correct for energy losses in the active material between the interaction point and the CAL;

11. use the non-vertexed VCTRHL tracks for the analysis instead of the vertexed VCTPAR tracks;

12. add an η_{max} cut to further remove diffractive events;

13. use the LEPTO MC program to correct the data.

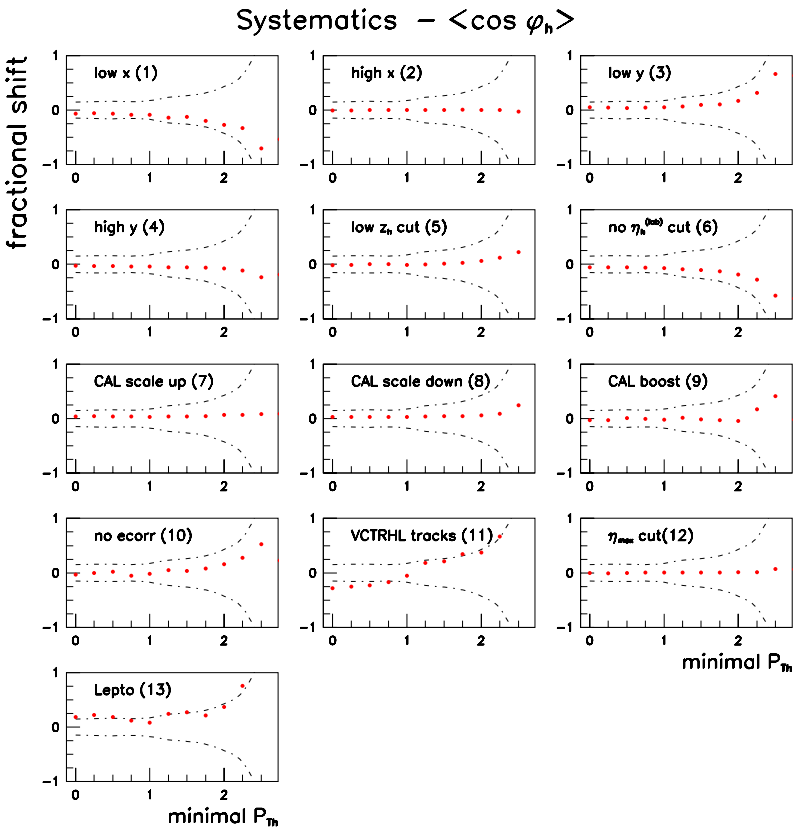

Figure 6.12: *Fractional shifts from the central values of the* $\cos\phi_h$ *moment for the various systematic checks. The fractional shifts are defined as* $\left(1/N\ dn/d\phi_h|_{syst} - 1/N\ dn/d\phi_h|_{central}\right) / \left(1/N\ dn/d\phi_h|_{central}\right)$ *. The dashed lines show the* (1σ) *errors on the moments as given by the fit to the central* ϕ_h *distributions (described in section 6.9.2).*

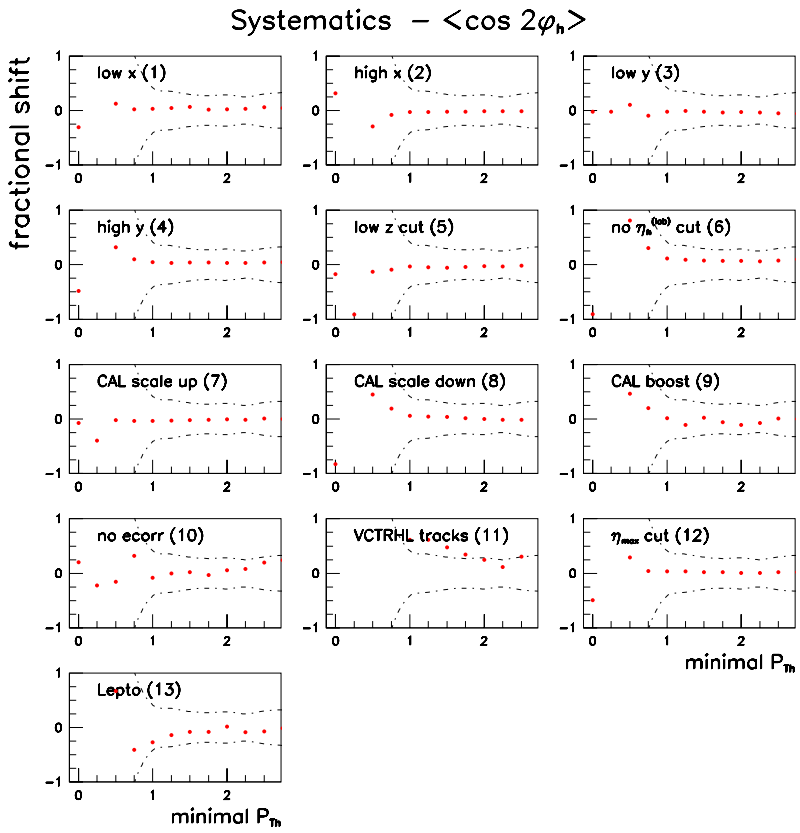

Figure 6.13: *Fractional shifts from the central values of the* $\cos 2\phi_h$ *moment for the various systematic checks. The dashed lines show the* (1σ) *errors on the moments as given by the fit to the central* ϕ_h *distributions (described in section 6.9.2).*

6.7.1 Systematics due to the Event Selection and Reconstruction

The first 4 items refer to standard systematic checks that take into account the finite resolution on the variables used to select the kinematic region under study. As discussed in section 5.6.2 we chose our kinematic phase space in the x and y variables. They were reconstructed with the DA method and then corrected back to hadron level, as described above. Due to finite resolutions, the reconstructed and true values are not quite the same; these systematic checks investigate how the measured distributions are influenced by these differences.

6.7.2 Systematics due to the Tracks Selection

Entry 5 allows to take into account the possible bias in the final distributions caused by the selection criterion applied to the tracks. Entry 6 investigated the possible bias introduced by the cleaning cut on the pseudorapidity of the tracks in the laboratory frame.

6.7.3 Calorimeter Energy Scale Uncertainty

Checks 7 and 8 reflect our uncertainty in the absolute energy scale of the calorimeter. The numbers quoted above were obtained in a thorough study of transverse momentum imbalance in high Q^2 DIS events [62].

6.7.4 Using Non-vertex Constrained Tracks

The analysis was also repeated with the non-vertexed tracks given in the VCTRHL table, to check the dependence of the results on the track fitting procedure during reconstruction and vertex assignment (systematic check 11).

These VCTRHL tracks were required to originate in the first superlayer and span at least 3 superlayers, and to have a distance of closest approach to the interaction point

$D_H < 1\,\mathrm{mm}$. The specific cut on the D_H variable was motivated by the distribution of this variable for all vertexed tracks; it roughly corresponds to the mean of the distribution of D_H for the selected vertexed charged tracks used in the central analysis, the distribution being rather steep.

6.7.5 Influence of Diffractive Events

Diffractive events are recognised from standard DIS events by the large rapidity gap (in the forward direction) with no hadronic activity. Imposing the cut 12, $\eta_{max} > 1.5$, η_{max} being the pseudorapidity η of the most forward energy deposit in the ZEUS calorimeter, provides a simple way to eliminate potential diffractive events from the DIS sample.

6.7.6 Choice of the Correcting Monte Carlo

Finally, Lepto was used to determine the bin-by-bin correction factors and hence to verify the dependence of the results on the MC modelling of the hadronic final state.

6.7.7 Combined Effects

The fractional shifts (with respect to the central values) on the fitted $\cos\phi_h$ and $\cos 2\phi_h$ moments (further details on the fitting procedure will be given in section 6.9.2) of the azimuthal angle distributions are displayed in figures 6.12 and 6.13, respectively. Most of the systematics have a small effect, being rather small compared to the errors on the fitted (central values of the) moments.

The $\cos\phi_h$ moment appears to be more sensitive to the effect of the systematics than the $\cos 2\phi_h$ moment. The errors on the $\cos 2\phi_h$ fitted values seem to be rather large at low P_{Th}; this is just due to the fact that this moment is very close to zero in this region. All systematics are in fact small for this ϕ_h moment. The VCTRHL and Lepto systematic checks represent the dominant contribution to the final systematic error on the $\cos\phi_h$ moment.

The various contributions to the total systematic error were then separately added in

quadrature for the positive and negative parts; the errors in the corrected distributions (discussed in section 6.9.2) are hence asymmetric.

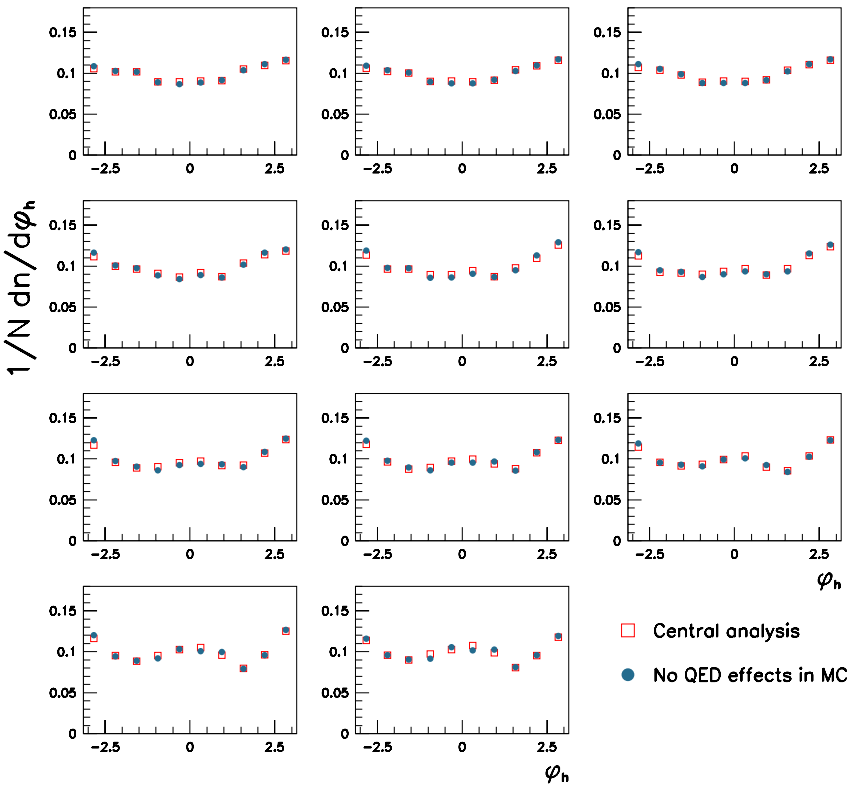

Figure 6.14: *Effects on the ϕ_h distributions due to the presence of QED radiative events: the ϕ_h distributions obtained correcting the data with an Ariadne sample with no QED radiation are superposed to those of the central analysis.*

6.8 Effects of QED Radiation

So far we haven't assessed the possible distortion effects caused by the presence of QED radiative events in the selected sample. These can be investigated if one corrects the data with a MC sample with no QED radiation present. We used again Ariadne. Figure 6.14 presents the results: the ϕ_h distributions obtained correcting the data with this "non-radiative" sample have been superposed to those obtained in the central analysis. The effects are reasonably small; they lead to the conclusion that potential distortions

in the ϕ_h distributions due to QED radiation are under control, and do not significantly affect our results.

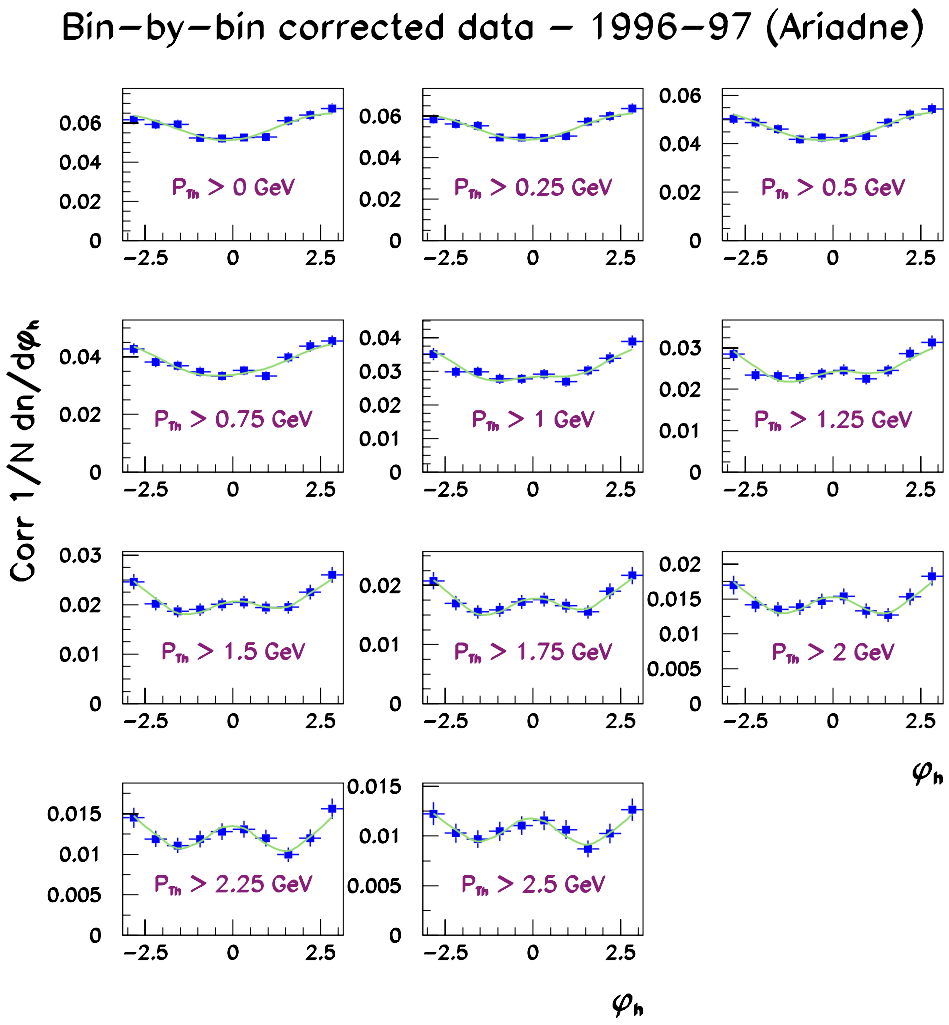

Figure 6.15: *Bin-by-bin corrected ϕ_h distributions in the HCM frame as a function of the minimum transverse momentum – P_{Th} cut – of the tracks. The solid lines show the result of the fit of the distributions to the theoretical form of equation 6.7. The Ariadne MC was used to correct the data.*

6.9 Results

6.9.1 Corrected Azimuthal Angle Distributions of Hadrons

The bin-by-bin corrected ϕ_h distributions of charged tracks in the HCM frame are shown in figure 6.15. They were obtained multiplying the experimental distributions Φ^{data} by the correction factors C_f, $\Phi^{corr} = C_f \Phi^{data}$.

Note that no systematics are included in these plots. The error bars take into account the statistical errors and the errors on the correction factors (added in quadrature).

The distributions exhibit the trends predicted by QCD: at low transverse momentum P_{Th} the distributions are mainly governed by the negative $\cos\phi_h$ contribution. As P_{Th} increases the $\cos 2\phi_h$ term comes into play.

6.9.2 Moments of the Distributions

Comparisons with QCD are most easily made analysing the moments of the distributions in figure 6.15. The moments of these distributions were obtained by a simple fit to the theoretically motivated parametric curve

$$\frac{d\sigma}{d\phi_h} = A + B\cos\phi_h + C\cos 2\phi_h + D\sin\phi_h \quad . \tag{6.7}$$

They are given by

$$\langle\cos\phi_h\rangle = B/2A \quad , \quad \langle\cos 2\phi_h\rangle = C/2A \quad , \quad \langle\sin\phi_h\rangle = D/2A \quad . \tag{6.8}$$

This equation 6.7 has the form given in equations 2.19 and 2.23 for the QCD-predicted non-perturbative and perturbative contributions to the ϕ_h distributions, respectively.

The $\sin\phi_h$ term was included in order to check its agreement with a null value as predicted by QCD: this term arises from one-loop contributions to the parton level scattering amplitude, and is time-reversal-odd (T-odd). It is only present in parity-violating weak interactions and in purely electromagnetic interactions with a polarised lepton beam [63, 64]. This moment measures the left-right asymmetry of hadron production

with respect to the lepton scattering plane. The $\sin\phi_h$ moment was found to be zero in most of the bins, within statistical errors.

The evolution of these moments with the transverse momentum P_{Th} cut of the tracks is displayed in figure 6.16. The statistical errors were obtained from the fit to the data. The $\langle\cos\phi_h\rangle$ moment is negative, in the order -0.05 in all the kinematic region studied, and presents a minimum around $1\ \mathrm{GeV}$. The $\langle\cos 2\phi_h\rangle$ moment is positive (within systematic errors) in all the kinematic region, and increases as the P_{Th} cut on the tracks increases. Its magnitude, of the order of 0.02, is smaller than the magnitude of the $\langle\cos\phi_h\rangle$ moment, as predicted from the QCD equations.

These results will be compared to the Monte Carlo and the theoretical calculations in chapter 11. The theoretical calculations will first be presented and discussed in chapter 10.

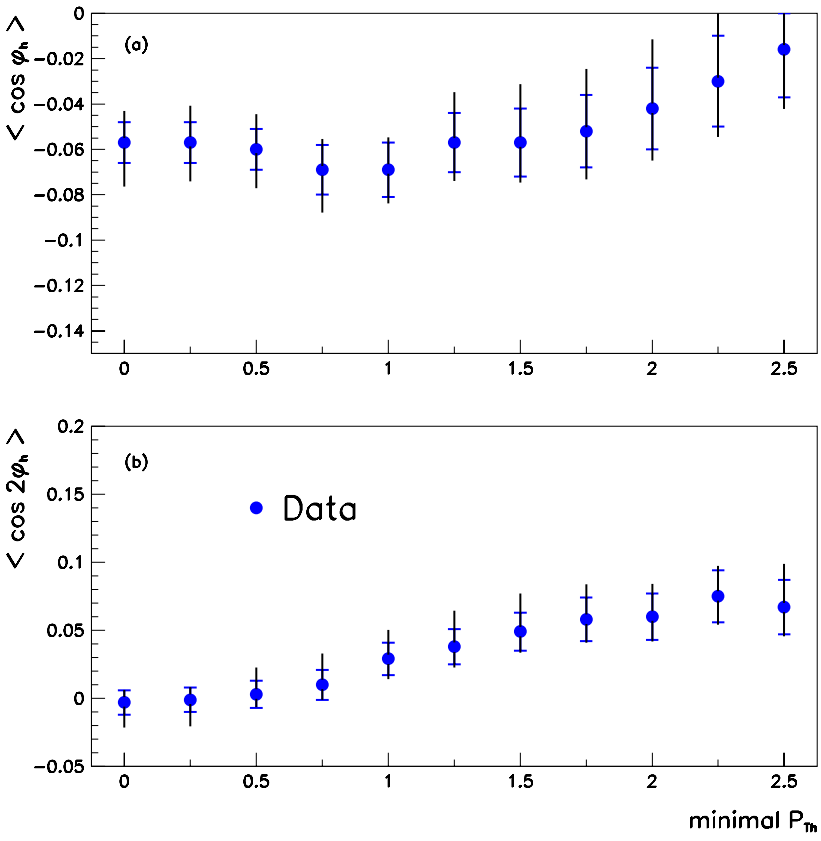

Figure 6.16: *Evolution with the minimal transverse momentum P_{Th} of the tracks of the mean of the moments of the ϕ_h distributions in the HCM frame: (a) evolution for $\langle \cos \phi_h \rangle$; (b)evolution for $\langle \cos 2\phi_h \rangle$. The inner and outer bars represent the statistical, and the statistical and systematic errors added in quadrature, respectively.*

Chapter 7

Analyses of Jets

7.1 Outline

The study of the azimuthal asymmetries of hadrons was detailed in the last chapter. We now pursue with the description of the analyses with jets. These analyses aim at a detailed study of the azimuthal asymmetries in separated samples of quark- and gluon-initiated jets.

The next chapter will first present the results obtained for the $\cos 2\phi$ moment of the jet azimuthal angle distribution in dijet events, when no jet flavour identification is performed.

Chapter 9 describes the techniques used to differentiate the flavour of the jets.

Here we first present the general aspects of the analyses: presentation of the jet finder, reconstruction and selection of the jets in the Breit frame, MC description of the jet data, reconstruction of the boost.

The Lepto MC was used throughout these analyses, as opposed to the analysis of hadrons, where Ariadne was chosen. The reason of this choice will be explained in the section 9.6.

7.2 Reconstruction of the Boost to the Breit Frame

As explained in section 2.2.2, the Breit frame was chosen for the jets analyses rather than the HCM frame. As with the analysis of tracks, a good reconstruction of the boost is crucial to avoid the introduction of distortions in the ϕ distributions. Reasonable resolutions around 6% and 16% were obtained for the transverse and the longitudinal components of the boost 3-vector $\vec{\beta}$ (*cf.* figure 7.1).

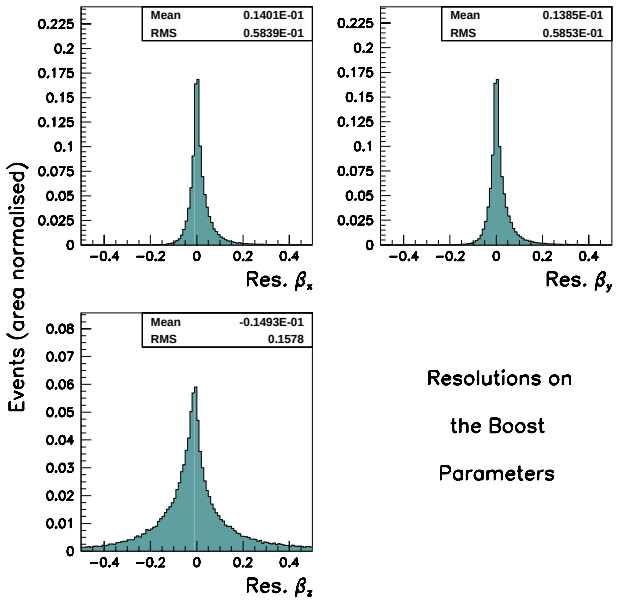

Figure 7.1: *Resolutions on the 3-vector $\vec{\beta}$ of the boost to the Breit frame, as obtained with the Lepto MC sample.*

7.3 Reconstruction and Selection of Jets

In what follows we will often refer to parton, hadron and detector level jets. A parton level jet is composed of the partons produced in the initial scattering and parton shower processes. The generated hadrons, after hadronisation, make up the hadron level jets. At the hadron level all charged and neutral particles with a lifetime greater than 3.10^{-10} sec were considered stable and used in the clustering. The detector level jets refer to the clustering of the detected particles, reconstructed as massless particles from

the CAL energy deposits. In the MC samples, these energy deposits were reconstructed after the simulation of the full ZEUS detector response.

Whenever the type of the input to the jet finder does not need to be specified, the partons, hadrons or CAL cells will be referred to as particles, or objects.

A jet's 4-momentum will be described by its transverse momentum E_T^{jet}, pseudorapidity η^{jet} and azimuthal angle ϕ^{jet}:

$$P^{\mu,jet} = (E_T^{jet}\cosh\eta^{jet}, E_T^{jet}\cos\phi^{jet}, E_T^{jet}\sin\phi^{jet}, E_T^{jet}\sinh\eta^{jet}) \quad . \tag{7.1}$$

These relations are exact for a massless jet.

7.3.1 The KTCLUS Jet Algorithm

Empirically, hadrons produced in high energy collisions tend to be confined to certain regions of the detector: they form jets. Quantitative comparisons between jet data and theory require the formal definition of a jet algorithm. Jet algorithms are demanded to be applicable to:

1. partons as generated in NLO programs such as DISENT;
2. final state hadrons as generated by LO MC programs such as LEPTO;
3. experimentally measured quantities, such as calorimeter cells.

Several possible definitions of jet algorithms are found in the literature. Due to the very nature of DIS, further requirements are imposed on the algorithms used for lepton-nucleon interactions. As the hard scattering process takes place between the electron and a parton from within the nucleon, the longitudinal momentum of the latter will vary on an event-by-event basis. Jet algorithms should not depend on this effect; thus DIS analyses use algorithms which are invariant with respect to a longitudinal boost (in the direction of the proton). Such algorithms cluster the particles based solely on their transverse energy and differences in pseudorapidity and azimuthal angle.

We used the k_T jet clustering algorithm proposed in [65,66] and implemented in the inclusive mode in the KTCLUS package [67] to find the jets in the data and MC samples.

This iterative algorithm clusters the particles based on the longitudinally boost-invariant quantities E_T, $\Delta\eta$ and $\Delta\phi$ (differences in η and ϕ). All particles are considered massless by setting $E = |\vec{p}|$.

The procedure starts with the full list of particles and an empty list of jets. Then:

1. for each particle i and for each pair of particles (i, j) the distances d_i and d_{ij},

$$d_i = E_{T,i}^2 R_0^2 \quad , \tag{7.2}$$

$$d_{ij} = \min(E_{T,i}^2, E_{T,j}^2) R_{ij}^2 \quad , \tag{7.3}$$

are calculated, where

$$R_{ij}^2 = \Delta\eta_{ij}^2 + \Delta\phi_{ij}^2 \tag{7.4}$$

is the radius distance in the (η, ϕ) plane, $\Delta\eta_{ij}^2$ and $\Delta\phi_{ij}^2$ being the differences in η and ϕ of the pair (i, j), respectively, and R_0^2 is a free parameter of the algorithm commonly set to unity;

2. the smallest of all the d_i and d_{ij} is called d_{min}. If d_{min} belongs to the d_i set then the corresponding particle is removed from the list of particles and added to the list of jets;

3. if d_{min} belongs to the d_{ij} set then the 2 corresponding particles are merged together into a pseudoparticle according to a specified recombination scheme (see description below), the 2 particles are replaced in the particles list by the merged pesudoparticle;

4. the procedure is repeated until there are no particles left, *i.e.* until all $d_i < \min(d_{ij})$. The remaining momenta are the (proto)jets.

The algorithm generally returns a large number of (proto)jets, many of which are single soft particles. These are then removed from the sample with a minimum cut (threshold) on the transverse energy in order to obtain the final list of jets.

Because of the nature of the algorithm, particles with $R_{ij}^2 < 1$ are merged; so the final jets are separated by distances $R > 1$. But, given the dynamic nature of the

merging, it might still happen that particles not belonging to a given jet be at a distance $R < 1$ to that jet's axis; conversely, particles of a jet may be at a distance $R > 1$ to the jet's axis.

The recombination of the particles was performed according to the E_T- recombination scheme, in which particles with relative transverse momentum are combined into jets: the momenta of particles i and j are merged into a particle k according to the definitions

$$E_{T,k} = E_{T,i} + E_{T,j} \quad , \tag{7.5}$$

$$\eta_k = \frac{E_{T,i}\eta_i + E_{T,j}\eta_j}{E_{T,i} + E_{T,j}} \tag{7.6}$$

and

$$\phi_k = \frac{E_{T,i}\phi_i + E_{T,j}\phi_j}{E_{T,i} + E_{T,j}} \quad . \tag{7.7}$$

This way, the final reconstructed jets are massless and their momenta follow the Snowmass convention [68]

$$E_T^{jet} = \sum_i E_{T,i} \quad , \quad \eta^{jet} = \frac{\sum_i E_{T,i}\eta_i}{\sum_i E_{T,i}} \quad , \quad \phi^{jet} = \frac{\sum_i E_{T,i}\phi_i}{\sum_i E_{T,i}} \quad . \tag{7.8}$$

The sums run over all particles i clustered into the jet.

7.3.2 Jets Selection

The jet selection expresses the need to ensure a good detection and reconstruction of the jets in the detector, and our choice was also driven by the absolute necessity of not to apply any cuts in the laboratory frame, to avoid the introduction of non-desirable distortions in the ϕ distributions. As demonstrated in [23], the use of jet cuts in the laboratory frame introduces a "kinematic component" that strongly distorts the authentic ϕ distributions.

Because of the detector acceptance and the trigger selection, jet studies at ZEUS always reject jets with a pseudorapidity of $\eta_{lab}^{jet} > 2.5$ in the laboratory frame. It was also shown in the past that a great deal of the jets found in the backward region, $\eta_{lab}^{jet} < -2$, tend to be fake jets, often originating in the misidentification of high energy photons.

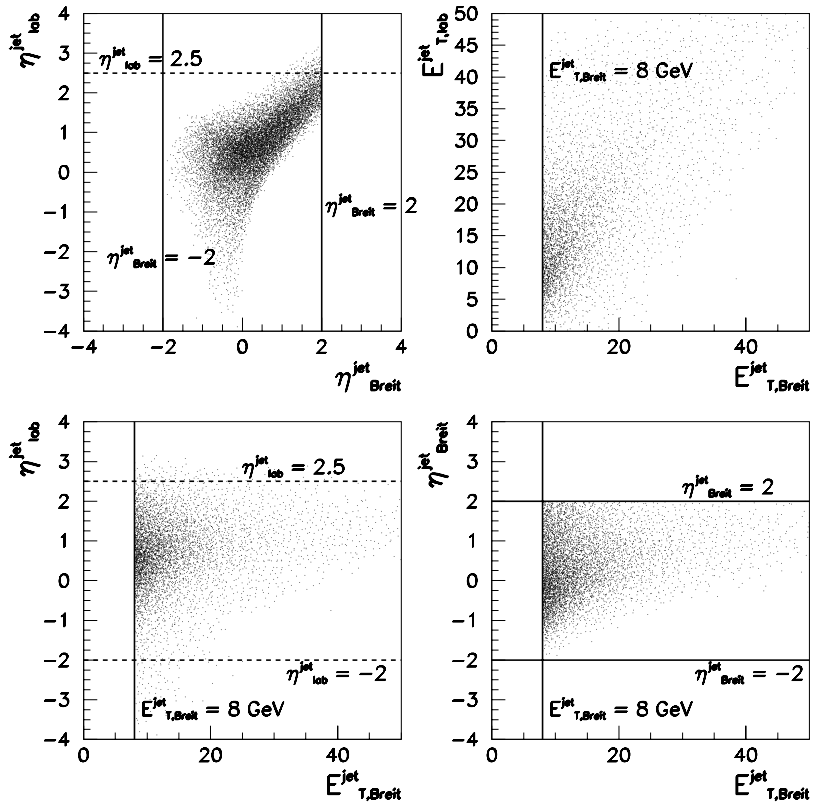

Figure 7.2: *Distributions of the generated jets at hadron level. The full lines show the cuts applied in the preliminary studies of the jets reconstruction, whereas the dotted lines merely serve to guide the eye, and were not used.*

The following set of jet cuts was optimised to remove most of the undesired events from the final data sample:

$$E_T^{jet} > 8 \text{ GeV} \quad , \tag{7.9}$$

$$-2 < \eta^{jet} < 2 \quad . \tag{7.10}$$

It is important to stress again that these jet cuts have been applied to all jets, exclusively in the Breit frame, and after the jet energy corrections had been performed.

All jets at detector level were first pre-selected with a loose cut of $E_T^{jet} > 4$ GeV in the Breit frame. The necessity for such a cut will be discussed in the following 2 sections on the methods used to reconstruct and correct the jets.

In what follows all jet variables with no subscript will refer to quantities measured in

the Breit frame. On occasion an explicit subscript will be introduced to avoid confusion.

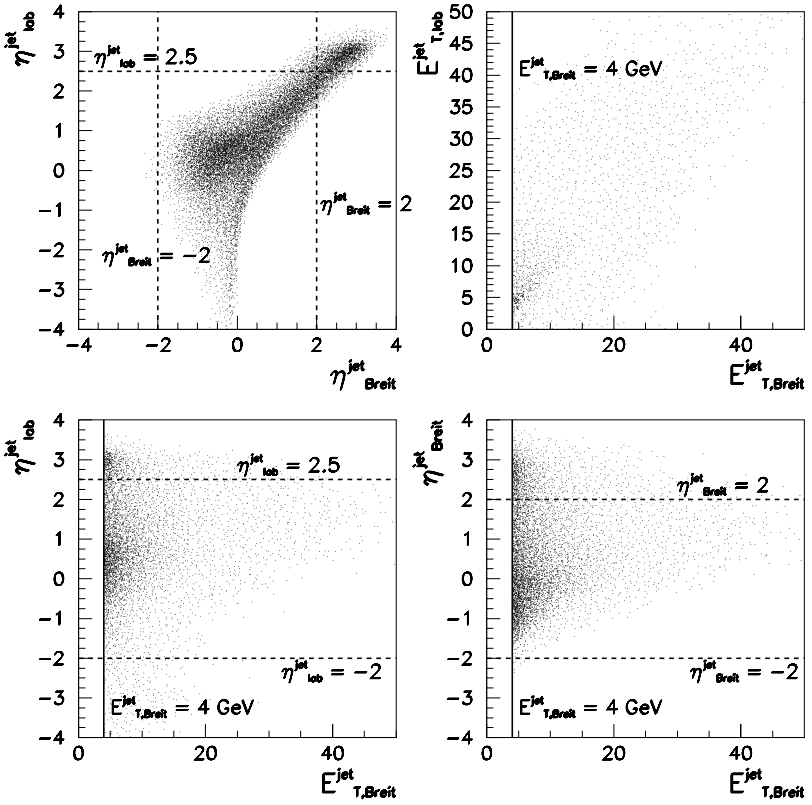

Figure 7.3: *Distributions of the reconstructed jets at detector level before application of the jet energy corrections. Thence loose cuts were used here (refer to the text for details.) The full lines show the cuts applied, whereas the dotted lines merely serve to guide the eye, and were not used.*

7.3.3 Jets Reconstruction

The objects (partons, hadrons or CAL cells) were first boosted to the Breit frame, and then clustered into jets with the KTCLUS jet algorithm. Parton and hadron level jets were boosted with the true components of the boost whereas the CAL cells were boosted using the reconstructed boost vector.

As a preliminary to the analyses of jets several steps were necessary in order to obtain the final corrected sample of jets: the "quality" of the reconstruction of the jets was

first studied comparing the jets found at detector level ("cal jets") with the jets found at hadron level ("had jets") in MC samples of generated events. In parallel, the set of jet cuts needed to be optimised.

The selected jets at both levels were matched according to the following principle: for the pair of a *cal* and a *had* jet the distance in the (η, ϕ) plane,

$$d(\eta, \phi) = [(\eta^{jet}_{cal} - \eta^{jet}_{had})^2 + (\phi^{jet}_{cal} - \phi^{jet}_{had})^2]^{1/2} \quad , \tag{7.11}$$

is defined. The sum of the $d(\eta, \phi)$ of each pair of a *cal* and a *had* jet is evaluated, and all possible combinations investigated. The combination that provides the minimum sum of the $d(\eta, \phi)$ gives the best matching between the jets at the two levels. (Note that jets are inevitably unmatched if the number of jets at each level differs.)

These reconstruction studies are typically done with a set of cuts as loose as practicable. We used our kinematic cuts and general DIS cleaning cuts. The cuts used in the final analyses, relations 7.9 and 7.10, were applied to the *had* jets; the *cal* jets had only to pass an $E^{jet}_T > 4$ GeV cut – all in the Breit frame.

As mentioned in the last paragraph, the (laboratory) most backward ($\eta^{jet}_{lab} < -2$) and forward ($\eta^{jet}_{lab} > 2.5$) regions should be avoided. Figures 7.2 and 7.3 show the correlation between the jets transverse energy and pseudorapidity in the Breit and the laboratory frames, after the cuts just stated, for the *had* and *cal* jets, respectively. A comparison makes clear the effect of the cuts:

- the cut $\eta^{jet}_{Breit} < 2$ cut effectively removes the bulk of the events with jets in the very forward region ($\eta^{jet}_{lab} > 2.5$) of the ZEUS detector;
- the $E^{jet}_{T,Breit}$ cut removes low $E^{jet}_{T,lab}$ jets, as well as jets in the very forward and backward regions, $-2 > \eta^{jet}_{lab} > 2.5$;
- the cut $\eta^{jet}_{Breit} > -2$ has a secondary impact. It mainly rejects low $E^{jet}_{T,Breit}$ jets.

With these 2 sets of cuts for *had* and *cal* jets, resolution and scatter plots were produced for the jet variables in the Breit and laboratory frames. Jets at the 2 levels were considered matched only if the (η, ϕ) distance, equation 7.11, verified $d(\eta, \phi) < 1$.

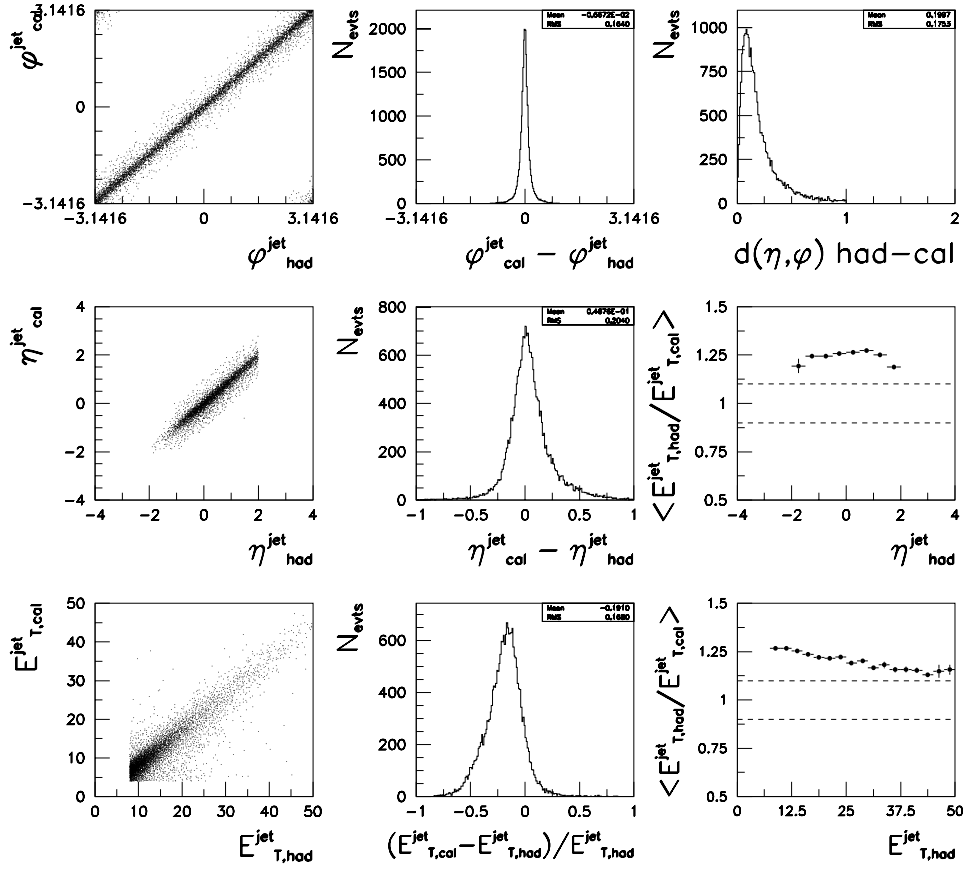

Figure 7.4: *Histograms and resolution plots showing the quality of the match of jets between the hadron and the detector levels, before any jet energy corrections were applied. All jet variables are in the Breit frame. The dotted lines show a ratio of* 0.9 *and* 1.1*; they merely serve to guide the eye.*

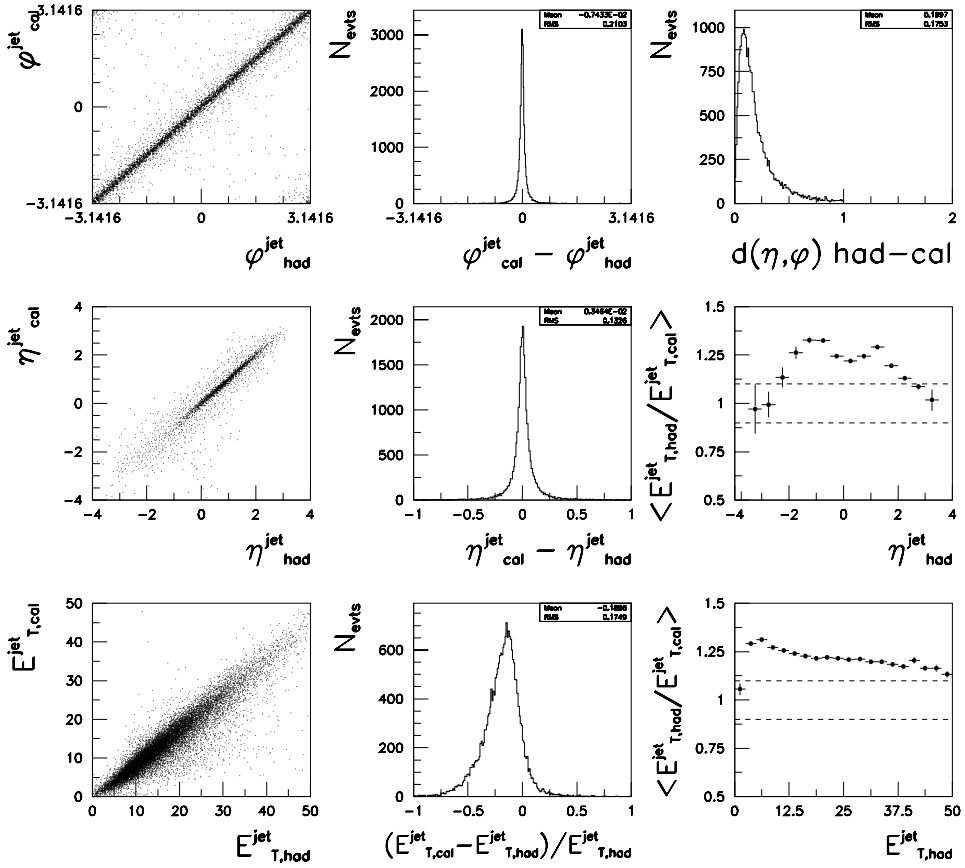

Figure 7.5: *Histograms and resolution plots showing the quality of the match of jets between the hadron and the detector levels, before any corrections were applied. All jet variables are in the laboratory frame, i.e. after boosting back from the Breit frame. The dotted lines show a ratio of* 0.9 *and* 1.1*; they merely serve to guide the eye.*

Figure 7.4 presents the variables in the Breit frame: there is an excellent correlation in the angular variables η^{jet} and ϕ^{jet}, the average difference (between the 2 levels) in these variables $\Delta\eta$ and $\Delta\phi$ being of 0.05 and -0.007, respectively. On the other hand, the transverse energy of the jets exhibits a significant shift, corresponding to an energy loss that needs to be corrected: the difference in the E_T^{jet} of the *had* and *cal* jets is offset by about 19%, and the spread of the distribution is of the order of 17%.

The same plots were also produced for the pairs of matched jets after performing a boost back to the laboratory frame (figure 7.5). Again the angular variables are very well correlated, whereas the E_T^{jet} "resolution" is again offset from 0 by $\sim 19\%$.

7.3.4 Jets Energy Corrections

In principle, the "raw" variables $E_{T,cal}^{jet}$, ϕ_{cal}^{jet} and η_{cal}^{jet} need to be corrected for detector effects. In the previous studies on the reconstruction of the jets it was shown that the distributions $\eta_{cal}^{jet} - \eta_{had}^{jet}$ and $\phi_{cal}^{jet} - \phi_{had}^{jet}$ do not present any significant systematic shift from the origin, and that the resolutions for the variables η_{cal}^{jet} and ϕ_{cal}^{jet} are about 0.20 and 0.16, respectively (see figure 7.4). No corrections to these variables are therefore needed; in the analyses it was assumed that $\eta_{cal}^{jet} \approx \eta_{had}^{jet} \approx \eta_{corr}^{jet}$ and $\phi_{cal}^{jet} \approx \phi_{had}^{jet} \approx \phi_{corr}^{jet}$.

On the other hand, the transverse energy of the *cal* jets at detector level presents a clear shift from zero, corresponding to an energy loss, that needs to be corrected. The jet energy correction, applied on a jet-by-jet basis, corrects, on average, the reconstructed E_T^{jet} for variations in the hadronic response of the CAL, for noise originating in the uranium, for energy losses in the material in front of the CAL and for detector non-uniformities. As expected, no ϕ dependence was found. But an η dependence is expected, and observed (*cf.* the profile plot of the evolution in η_{had}^{jet} of the ratio $\langle E_{T,had}^{jet}/E_{T,cal}^{jet}\rangle$ in the laboratory frame, in figure 7.5).

In general we can write the relation of the measured to the "true" energy as follows:

$$E_{T,true}^{jet} = C(E_{T,cal}^{jet}, \eta_{cal}^{jet}) \times E_{T,cal}^{jet} \quad , \tag{7.12}$$

$C(E_{T,cal}^{jet}, \eta^{cal})$ being the appropriate correction factor, which will be a function of $E_{T,cal}^{jet}$

η^{jet} bin #	η^{jet} range
1	[-4.00 , -2.00]
2	[-2.00 , -1.50]
3	[-1.50 , -1.25]
4	[-1.25 , -1.00]
5	[-1.00 , -0.75]
6	[-0.75 , -0.50]
7	[-0.50 , -0.25]
8	[-0.25 , 0.00]
9	[0.00 , 0.25]
10	[0.25 , 0.50]
11	[0.50 , 0.75]
12	[0.75 , 1.00]
13	[1.00 , 1.25]
14	[1.25 , 1.50]
15	[1.50 , 1.75]
16	[1.75 , 2.00]
17	[2.00 , 2.50]
18	[2.50 , 4.00]

Table 7.1: *Definition of the η^{jet} bins used in the jet energy corrections.*

in bins of η^{jet}_{cal}. The value of the resolution on η^{jet}_{cal} indicates the adequate minimum bin size for the η^{jet}_{cal} bins used in the jet energy corrections: we chose a bin of size $\Delta\eta_{bins} = 0.25$ for most but the very forward and backward regions; all bins are defined in table 7.1.

The jet energy correction factors are determined in the following way. First, a MC event sample is selected with the loose set of cuts given above. We have used the final jet selection cuts at hadron level, but loosened the jet cuts at detector level: for our sample of corrected $E^{jet}_T > 8$ GeV jets a threshold at 4 GeV was used at detector level and applied before correcting the data and the MC. Such a low cut at detector level is necessary to avoid biases in the final sample of corrected jets because of the energy losses. In practice the loosened cuts are determined from the resolution in E^{jet}_T: given a resolution in E^{jet}_T of $\sim 20\%$, the cut at 4 GeV stands at $\sim 2\sigma$ from the final 8 GeV cut. Cuts on the angular variables of the *cal* jets should not affect the results, but were avoided.

The selected jets at hadron and detector levels were then matched in the Breit frame and the matched pairs boosted back to the laboratory frame. The result of this matching gives the correspondence for the E_T^{jet} and the η^{jet} between levels.

Profile plots were produced of $\langle E_{T,cal}^{jet} \rangle$ as a function of $E_{T,had}^{jet}$ in the various η_{cal}^{jet} bins (all in the laboratory frame). As the relation turns out to be rather linear, we adopted the parameterisation

$$\langle E_{T,cal}^{jet} \rangle = a + b E_{T,had}^{jet} \quad . \tag{7.13}$$

The fitting of the profile plots, presented in figures 7.6 and 7.7, yielded the functions of η_{lab}^{jet}, a and b. The values are gathered in table 7.2.

Given the $E_{T,cal}^{jet}$ of a reconstructed jet its corrected value in the laboratory frame is then

$$E_{T,corr}^{jet} = (E_{T,cal}^{jet} - a)/b \quad . \tag{7.14}$$

The correction factors were obtained from the jets boosted back into the laboratory frame. But, in fact, they can be applied directly to the jets momenta in the Breit frame, since the boost is linear in the jets momenta.

Using the correction procedure based on the fit of the form in equation 7.13, we then compared again the distributions of the jet variables in the Breit and back in the laboratory frames, correcting the *cal* jets and applying the final set of cuts, equations 7.9 and 7.10, at both levels. The comparisons are displayed in figures 7.8 and 7.9. Jets at the 2 levels were again considered matched only if the (η, ϕ) distance, equation 7.11, verified $d(\eta, \phi) < 1$. Compared to the results before the usage of the corrections there is a clear improvement. The resolutions on the angular variables remained very similar, as well as the resolution on the transverse energy of the jets. But the latter is now roughly around 0 – a slight shift still remains – and is stable with respect to the full range of the jet variables. The same conclusions can be drawn for the comparisons made back in the laboratory frame (figure 7.9).

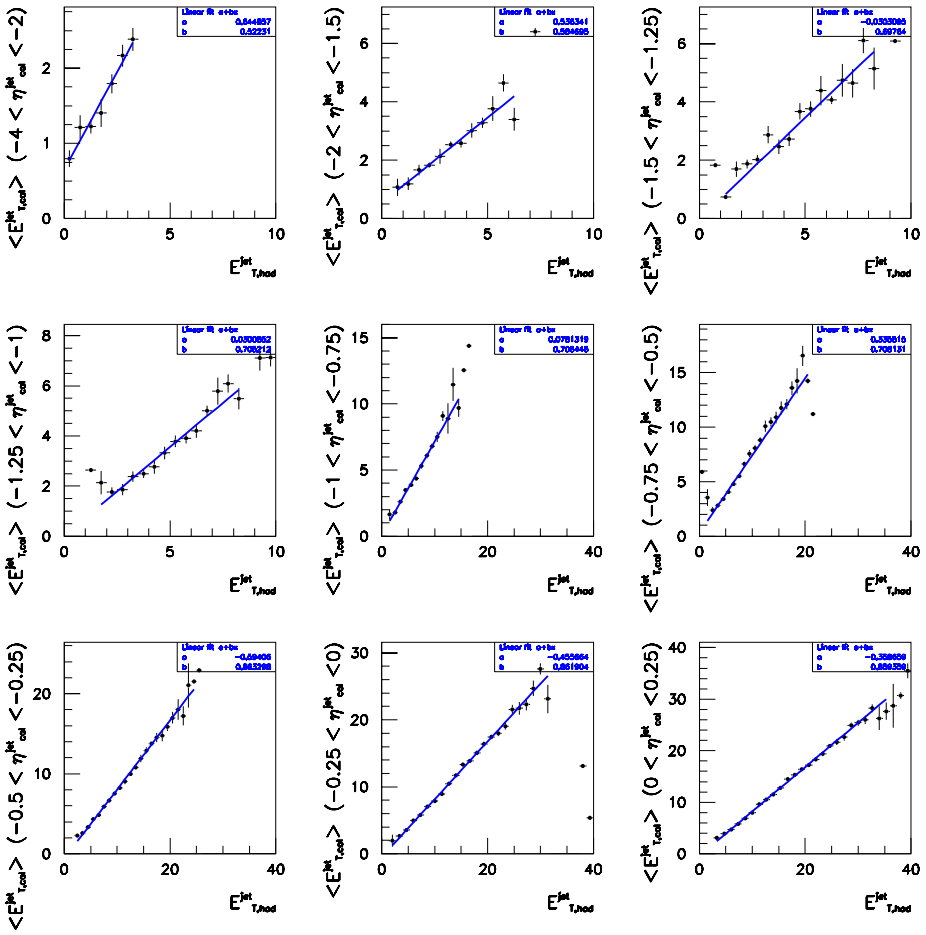

Figure 7.6: *Profile plots of $\langle E^{jet}_{T,cal} \rangle$ as a function of $E^{jet}_{T,had}$ in the various η^{jet}_{cal} bins used to determine the jet energy correction factors. The solid (blue) lines are the result of the linear fit $\langle E^{jet}_{T,cal} \rangle = a + bE^{jet}_{T,had}$.*

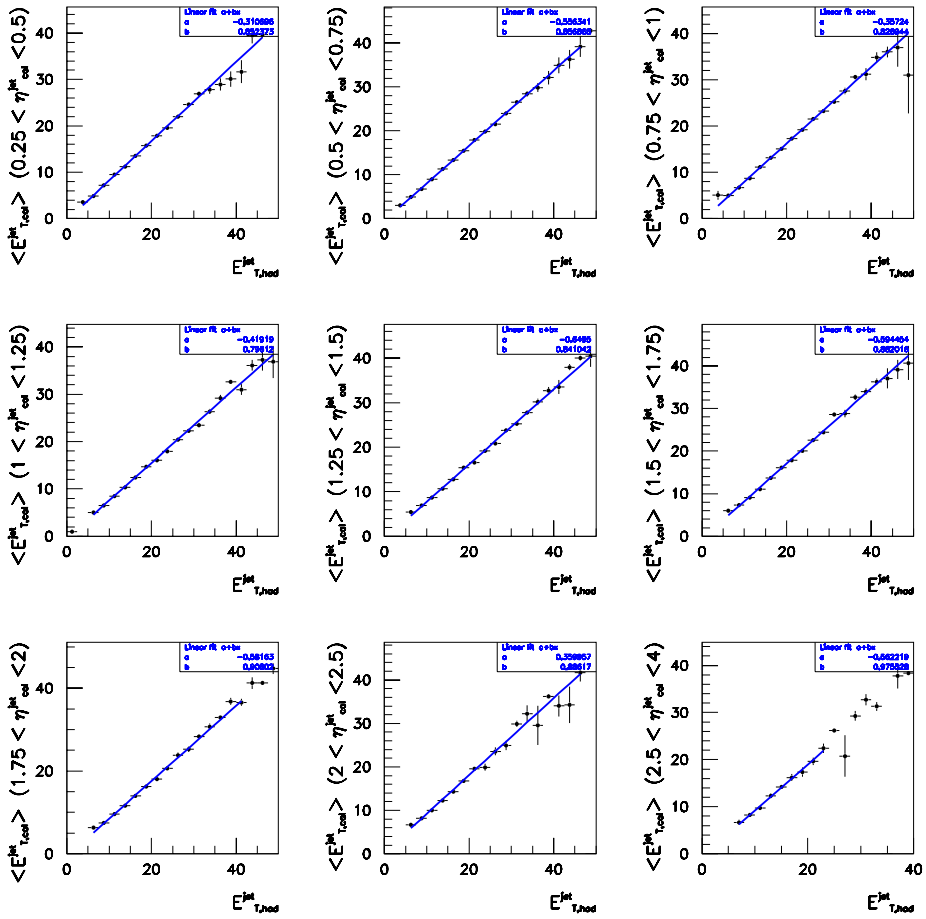

Figure 7.7: *Profile plots of* $\langle E^{jet}_{T,cal} \rangle$ *as a function of* $E^{jet}_{T,had}$ *in the various* η^{jet}_{cal} *bins used to determine the jet energy correction factors. The solid (blue) lines are the result of the linear fit* $\langle E^{jet}_{T,cal} \rangle = a + bE^{jet}_{T,had}$.

η^{jet} bin #	a	b
1	0.6449	0.5223
2	0.5363	0.5847
3	−0.0303	0.6976
4	0.0301	0.7052
5	0.0781	0.7084
6	0.3355	0.7081
7	−0.5941	0.8633
8	−0.4557	0.8619
9	−0.3587	0.8594
10	−0.3107	0.8524
11	−0.5563	0.8569
12	−0.3572	0.8289
13	−0.4192	0.7961
14	−0.6495	0.8410
15	−0.5945	0.8820
16	−0.5816	0.9080
17	0.3600	0.8862
18	−0.5622	0.9755

Table 7.2: *The jet energy correction factors as given by fit* $\langle E_{T,cal}^{jet} \rangle = a + bE_{T,had}^{jet}$ *(refer to the text for further details).*

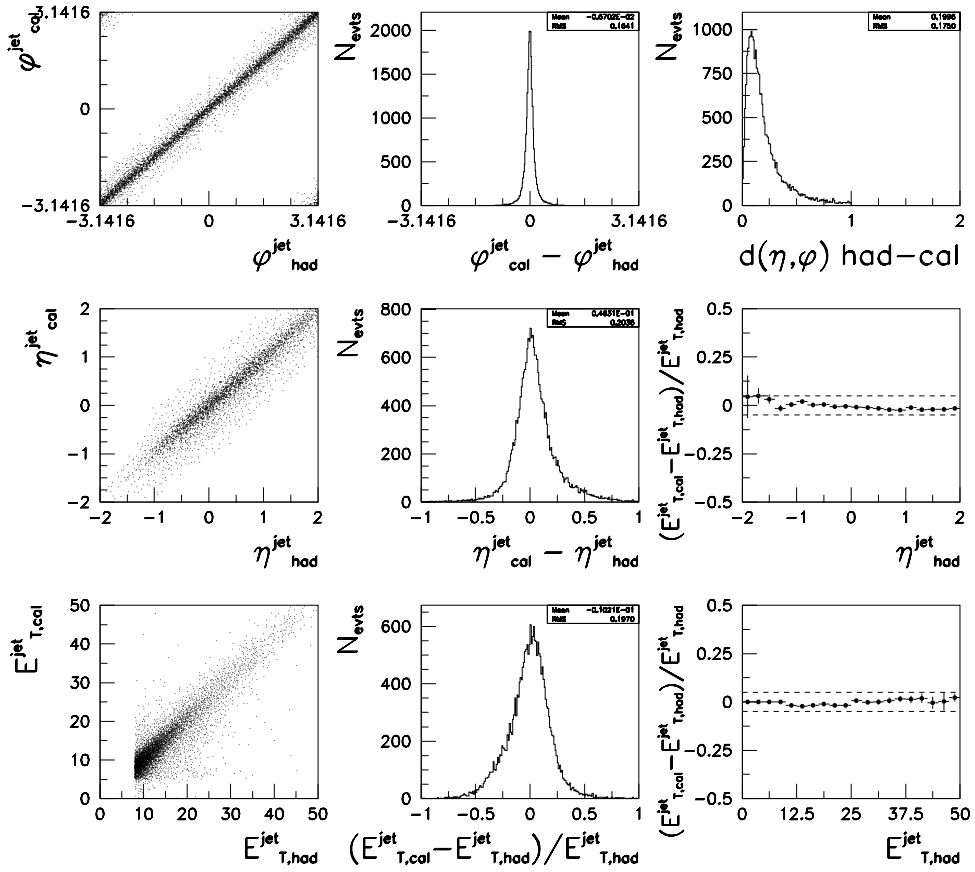

Figure 7.8: *Histograms and resolution plots showing the quality of the match of jets between the hadron and the detector levels, after application of the jet energy corrections. All jet quantities are in the Breit frame. The dotted lines show a resolution band of 5%; they merely serve to guide the eye.*

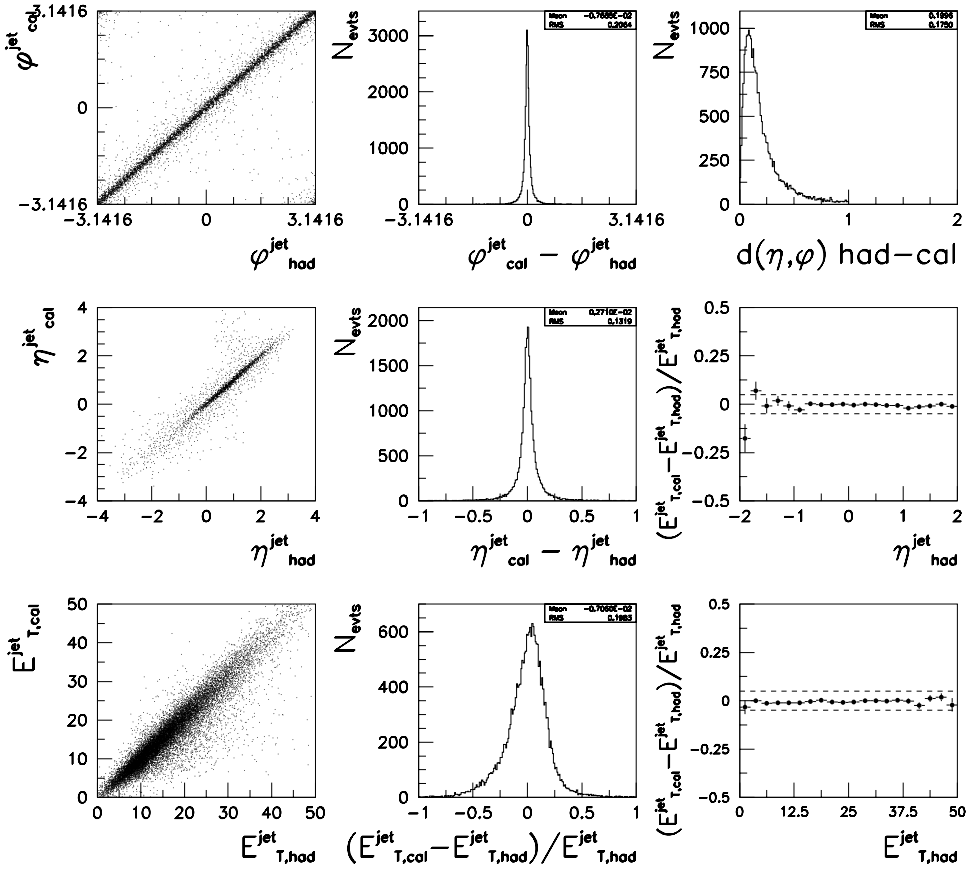

Figure 7.9: *Histograms and resolution plots showing the quality of the match of jets between the hadron and the detector levels, after application of the jet energy corrections. All jet quantities are in the laboratory frame, i.e. after boosting back from the Breit frame. The dotted lines show a resolution band of 5%; they merely serve to guide the eye.*

7.4 MC Description of the Data

In what follows the detector level jets have always been corrected using the procedure explained in the last section.

7.4.1 General Event Variables

The agreement between the MC and the data was presented in chapter 5 for the general event variables – the kinematic, positron and exchanged boson variables – but those distributions were obtained with the Ariadne sample. Nevertheless, we will not present here the same plots obtained with the Lepto sample; the agreement is very similar, although Ariadne tends to provide a slightly better overall description of these event variables.

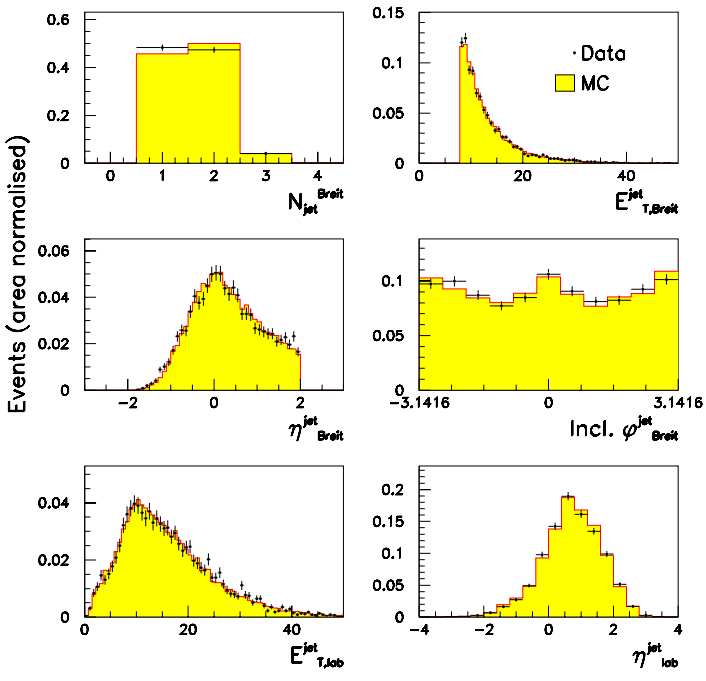

Figure 7.10: *Lepto description of the general jet variables. "Breit" and "lab" refer to the variables in the Breit frame and after boosting back to the laboratory frame, respectively.*

7.4.2 Jet Variables

Figure 7.10 shows the level of agreement between the data and Lepto for the relevant jet variables. The E_T^{jet} and η^{jet} distributions are presented in both the Breit frame and after the jets had been boosted back to the laboratory frame. Lepto provides a reasonable description of the jet data. In particular, the E_T^{jet} and η^{jet} distributions, used to select the jets, are rather well described. Also the ϕ^{jet} distribution of the inclusive sample of jets is reasonably well described by Lepto.

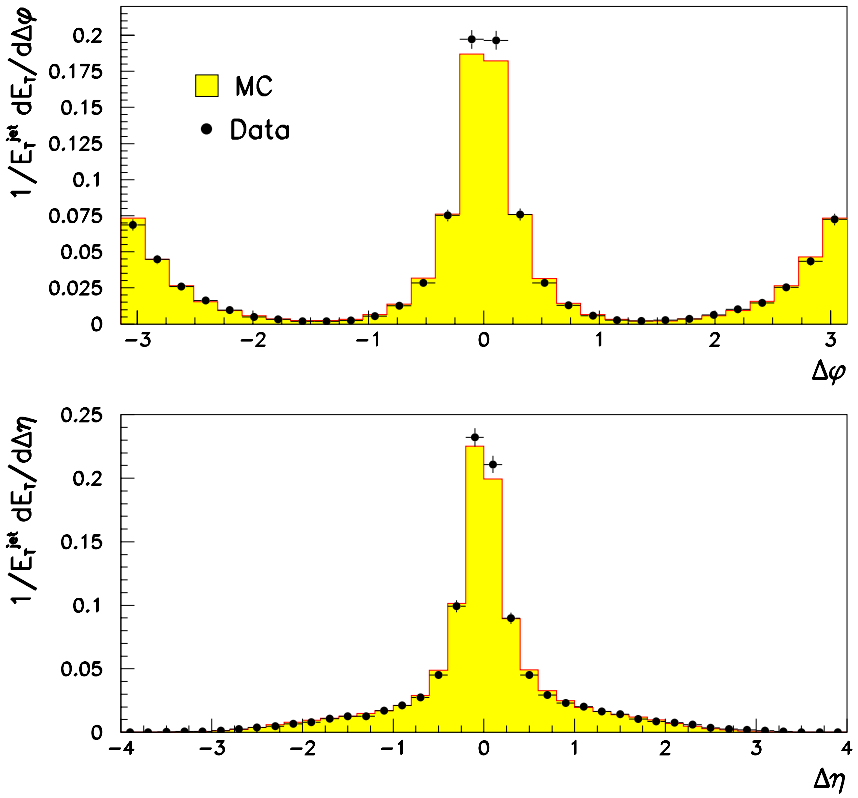

Figure 7.11: *Lepto description of the jet E_T^{jet} flows (E_T^{jet} profiles) in the Breit frame. The differences in ϕ and η, $\Delta\phi$ and $\Delta\eta$, were calculated with respect to the jet axis. The histograms were normalised to a unit area.*

Chapter 9 will study in detail the internal structure of the jets. Thence we here present the comparison between Lepto and data for general distributions and variables that study detailed aspects of the internal distribution inside a jet of the particles and

of the transverse energy. Jet substructure-sensitive variables will be studied in depth in chapter 9. Figures 7.11 to 7.16 show various studies of the jet transverse energy flows in the Breit frame,

$$\frac{1}{E_T^{jet}}\frac{dE_T}{d(\Delta\eta)} \tag{7.15}$$

and

$$\frac{1}{E_T^{jet}}\frac{dE_T}{d(\Delta\phi)} \quad . \tag{7.16}$$

In the figures, $\Delta\eta$ and $\Delta\phi$ are, respectively, the differences in η and ϕ of the objects, measured with respect to a jet axis (either the containing jet or the highest E_T jet of the event).

Lepto provides a fair overall description of these data (figures 7.11 and 7.12). It tends to perform better in the forward rather than in the backward region (see figure 7.12).

The lowest E_T jets are also marginally better described than the highest E_T jet of the event. This can be concluded from a comparison of figures 7.11-7.12 with figures 7.13-7.16, which separately show the inclusive sample, denoted "all jets", and the only second, third, ..., jets, denoted "jets 2,3,... wrt jet 1", in different η^{jet} regions. The jets were ordered with decreasing E_T, starting with "jet 1" for the highest E_T jet. Figures 7.13-7.14 and 7.15- 7.16 show the distributions for real data and at hadron level, respectively. These distributions are the result of plotting the E_T flows of the second, third, etc. jets with respect to the highest E_T jet in the event. The $1/E_T^{jet} dE_T/d\Delta\phi$ distributions show, in particular, the good balance in transverse momentum.

Figure 7.17 looks at the jet internal structure from another perspective; it shows the distribution of the distance

$$d(\eta,\phi) = [(\eta^{obj} - \eta^{jet})^2 + (\phi^{obj} - \phi^{jet})^2]^{1/2} \quad , \tag{7.17}$$

of the objects (hadrons or CAL cells) of the jet to the jet's axis. Here again Lepto is in reasonable agreement with the data.

We therefore conclude that the Lepto MC program provides a general reasonable description of the jets and their internal structure, and gives us confidence on the MC studies to be detailed in chapter 9. In particular, it was important to show here evidence

of a good agreement between the data and MC for the jet azimuthal angle distribution and the jet flows in azimuthal angle.

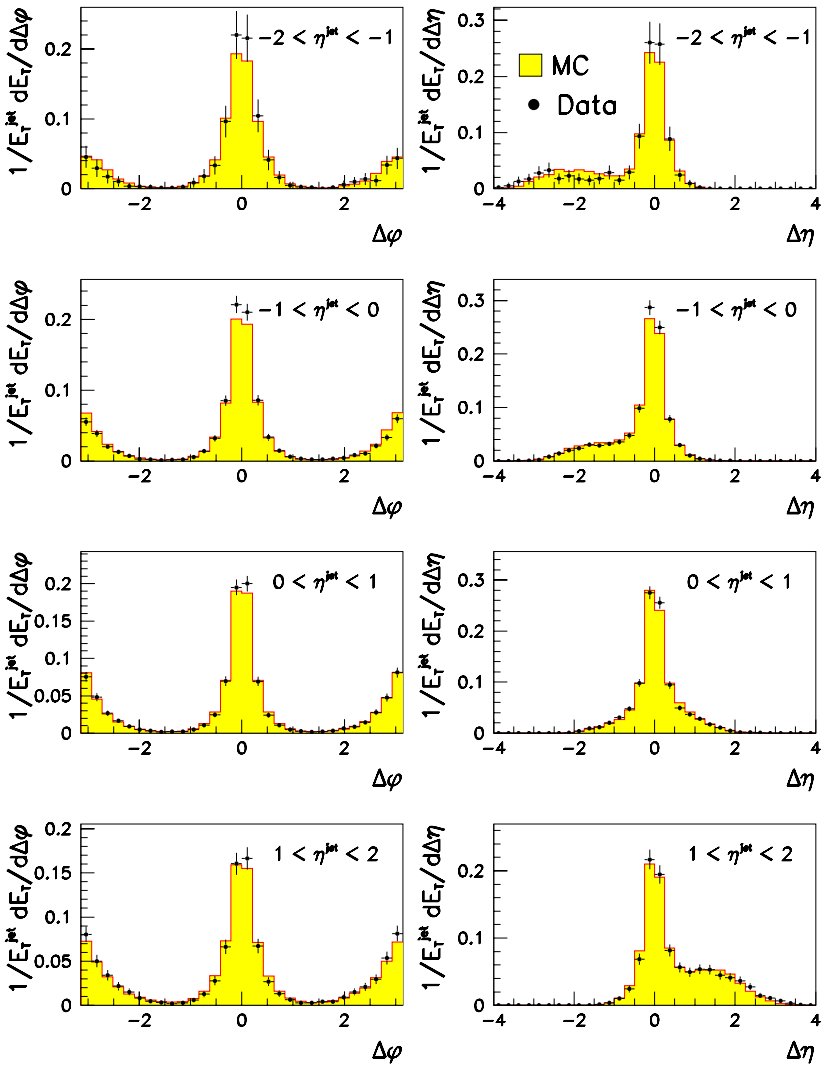

Figure 7.12: *Lepto description of the jet E_T^{jet} flows (E_T^{jet} profiles) in the Breit frame, for several η^{jet} regions. The differences in ϕ and η, $\Delta\phi$ and $\Delta\eta$, were calculated with respect to the jet axis. The histograms were normalised to a unit area.*

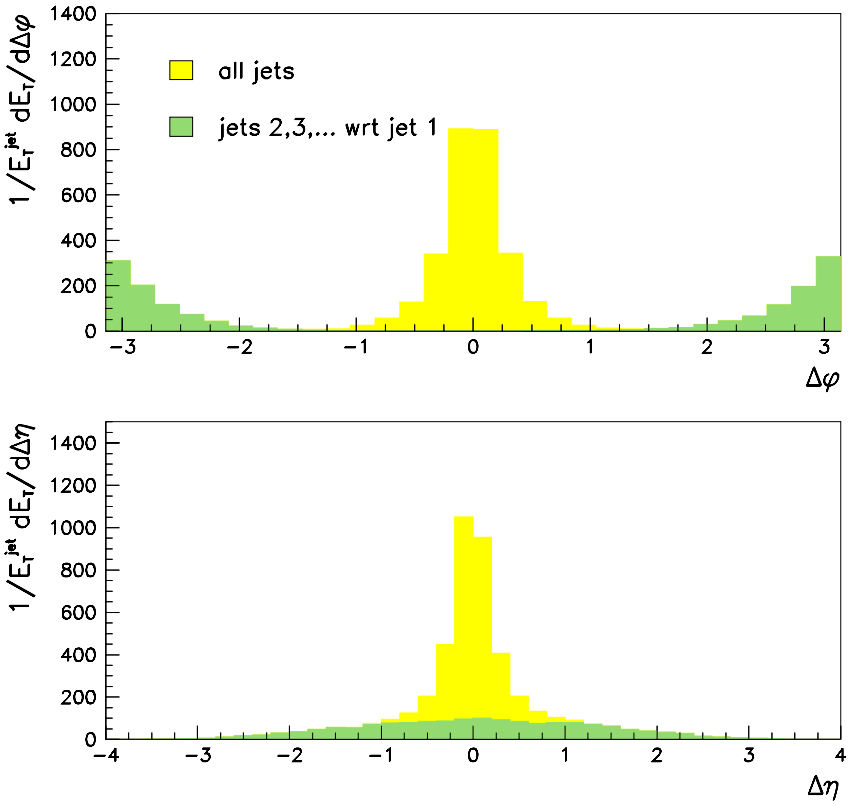

Figure 7.13: *Jet E_T flows in the Breit frame for real data. The distributions for the inclusive sample, denoted "all jets", and only for the second, third, ..., jets, denoted "jets 2,3,... wrt jet 1", are superposed. The differences in ϕ and η, $\Delta\phi$ and $\Delta\eta$, were calculated with respect to the axis of the highest E_T jet.*

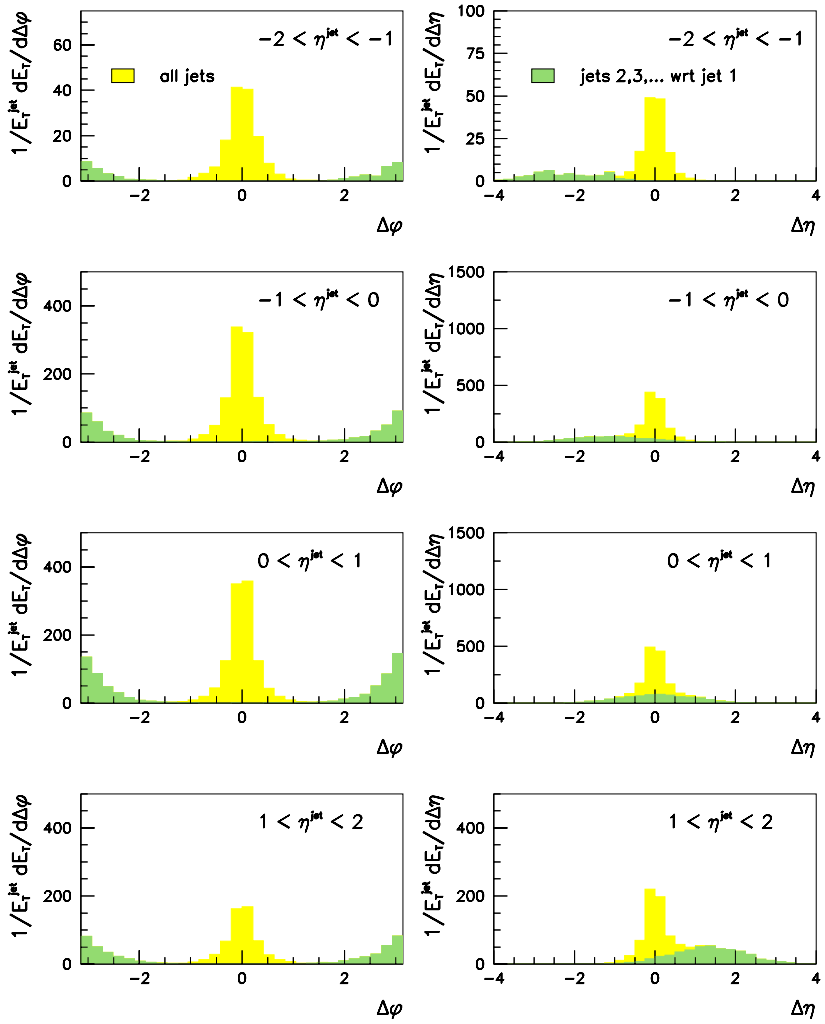

Figure 7.14: *Jet E_T flows in the Breit frame for real data, in several η^{jet} regions. The distributions for the inclusive sample, denoted "all jets", and only for the second, third, ..., jets, denoted "jets 2,3,... wrt jet 1", are superposed. The differences in ϕ and η, $\Delta\phi$ and $\Delta\eta$, were calculated with respect to the axis of the highest E_T jet.*

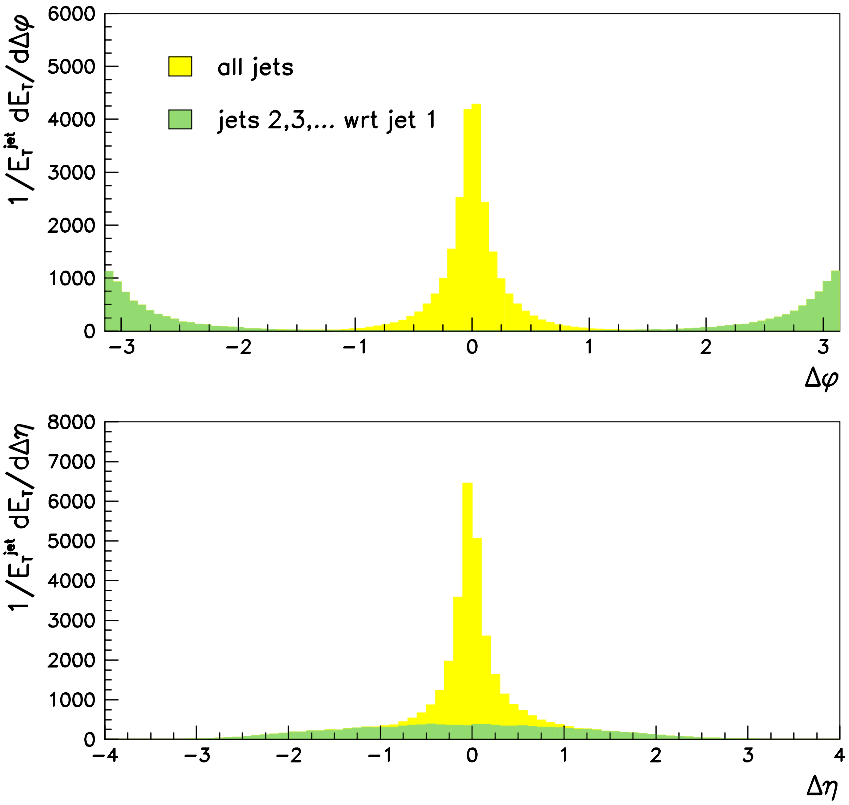

Figure 7.15: *Jet E_T flows in the Breit frame for hadron level jets. The distributions for the inclusive sample, denoted "all jets", and only for the second, third, ..., jets, denoted "jets 2,3,... wrt jet 1", are superposed. The differences in ϕ and η, $\Delta\phi$ and $\Delta\eta$, were calculated with respect to the axis of the highest E_T jet.*

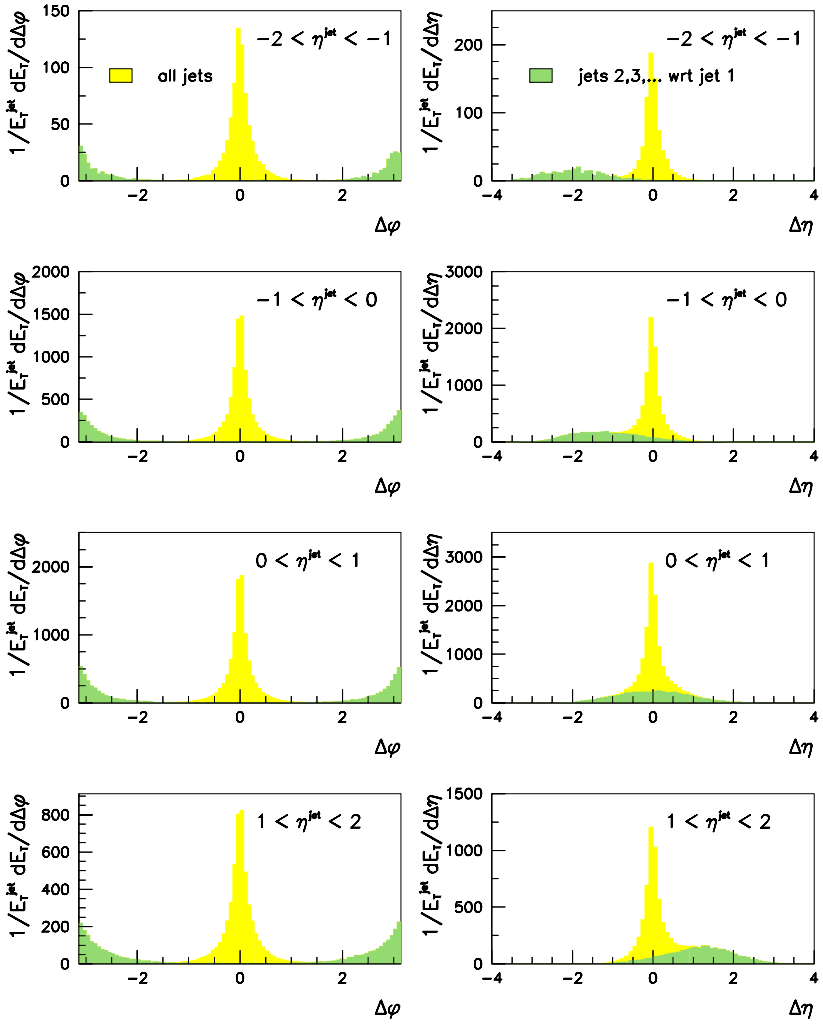

Figure 7.16: *Jet E_T flows in the Breit frame for hadron level jets, in several η^{jet} regions. The distributions for the inclusive sample, denoted "all jets", and only for the second, third, ..., jets, denoted "jets 2,3,... wrt jet 1", are superposed. The differences in ϕ and η, $\Delta\phi$ and $\Delta\eta$, were calculated with respect to the axis of the highest E_T jet.*

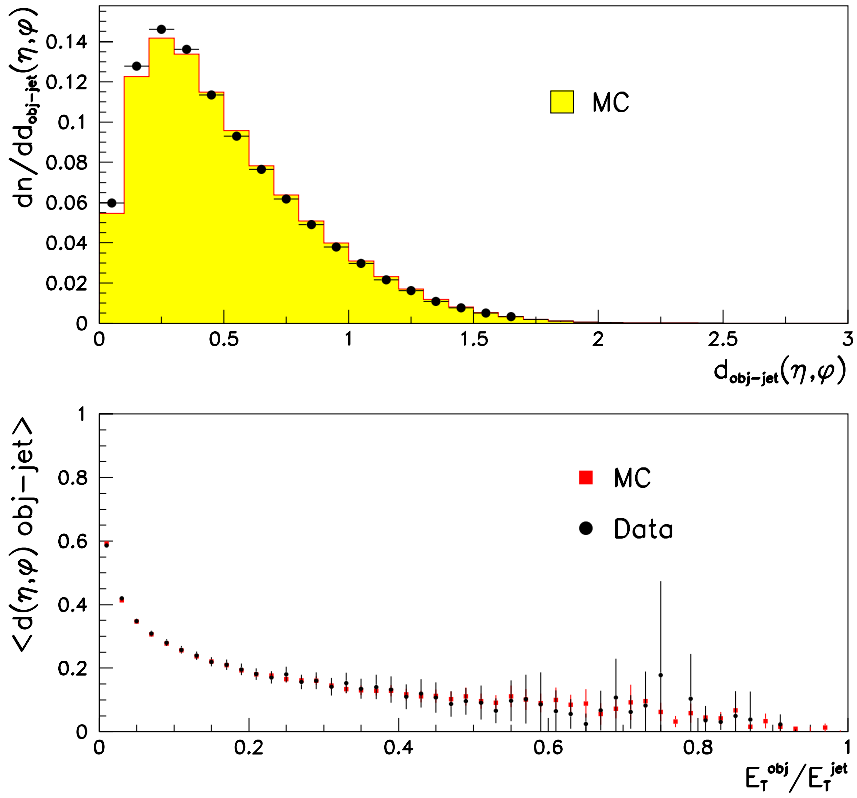

Figure 7.17: *Lepto description of the internal structure of the jets in the Breit frame: distribution of the distance of the objects of the jet to the jet's axis. The definition of the distance is given in equation 7.17.*

Chapter 8

Dijets Azimuthal Asymmetries

8.1 Outline

We have shown in the previous chapter that Lepto provides a good general description of jets in the Breit frame. This chapter investigates the azimuthal angular distribution of jets in dijet events in the Breit frame.

We are, again, probing QCD in the hadronic final state in DIS. In order to remove as much as feasible the possible sources of bias, we will analyse solely dijet events, for which the correspondence between the outgoing partons and the 2 selected jets is most clear.

As a prelude to the core of the analysis we will start by further motivating the non-inclusion of 1-jet events in the final sample, which would have given rise to an inclusive sample.

8.2 Effects of ISR on Jet Azimuthal Distributions

Distortions in the azimuthal angle distributions have mainly 2 sources: a bad reconstruction of the boost and the presence of radiated hard photons in the sample. It is imperative to select a kinematic region and a set of cuts that minimise these distortions.

A study was performed in our chosen (x, y) kinematic phase space (defined in sec-

tion 5.6.2) to investigate the effect of radiative events in the ϕ^{jet} distribution of our sample. We are interested in initial-state radiation (ISR), since the final-state photons will in general not cause problems as they tend to be detected inside the calorimeter cluster associated with the positron.

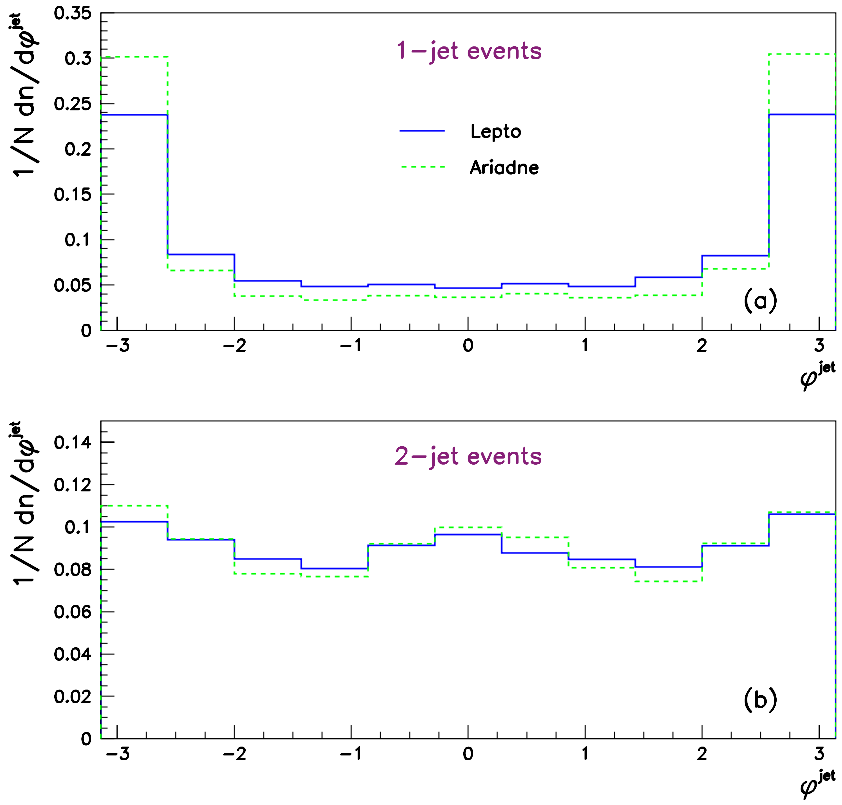

Figure 8.1: *ϕ^{jet} distributions for (a) 1-jet and (b) 2-jet events at hadron level obtained with Lepto and Ariadne.*

Azimuthal angle distributions were plotted in the Breit frame at both hadron and detector levels. We also plotted these same distributions at detector level using the true components of the boost vector in order to better assess the influence of the reconstruction of the boost on the distortions obtained.

Figures 8.1 and 8.2 present the ϕ^{jet} distributions for 1-jet and 2-jet events separately at hadron and detector levels, respectively. These distributions were obtained both with Lepto and Ariadne, which agree rather well (taking into account the statistical

fluctuations of the samples).

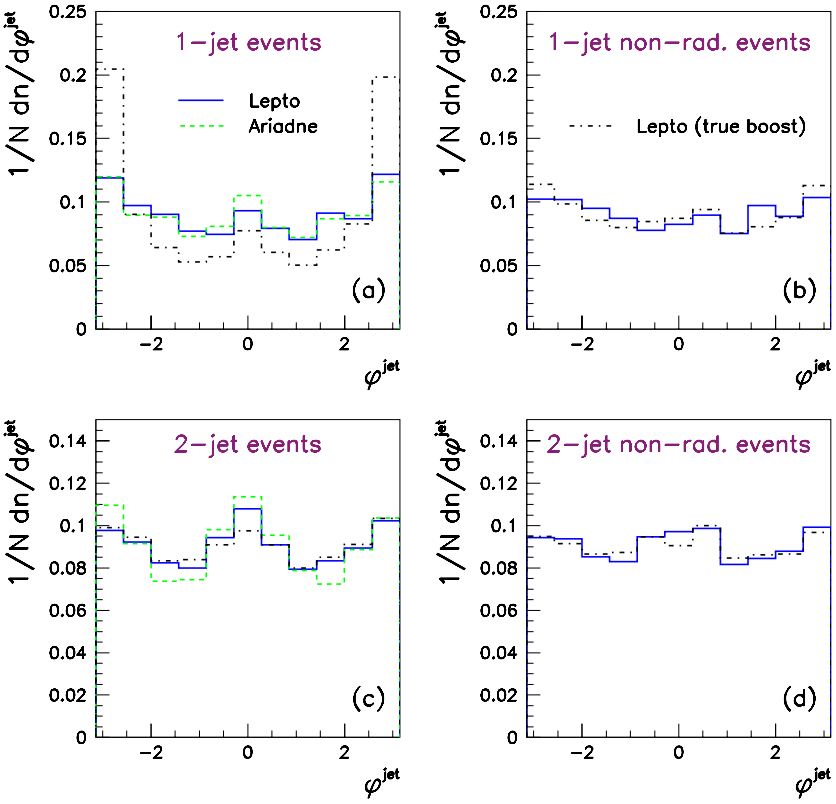

Figure 8.2: *ϕ^{jet} distributions at detector level obtained with Lepto and Ariadne: (a) 1-jet events, (b) 1-jet non-radiative events, (c) 2-jet events and (d) 2-jet non-radiative events. Distributions were also obtained with Lepto using the true components of the boost vector to boost the jets at detector level "(true boost)".*

At both levels the ϕ^{jet} distributions for 2-jet events show evidence for a $\cos 2\phi$-dominated shape. At detector level, one can also notice that the shapes of the distributions do not seem to be affected significantly either by the presence of ISR radiation or by the reconstruction of the boost (figures 8.2(c) and (d)). This statement is not valid for 1-jet events, though. When radiation is present, the ϕ^{jet} distributions have large peaks in the regions $\phi^{jet} \approx \pm\pi$. The effect is remarkably attenuated at detector level; only when the true boost is performed can one verify the significant distortion due to the presence of radiation (figure 8.2(a)). This leads us to the conclusion that the

reconstruction of 1-jet events is largely affected by ISR radiated photons.

A comparison between the distributions obtained at hadron and detector levels is plotted in figure 8.3, which shows the ϕ^{jet} distributions for 1-jet and 2-jet events obtained with Lepto and using the true components of the boost vector to transform the jets at detector level ("true boost"). Using the true boost on the reconstructed jets one can clearly see the effect of the reconstruction of the boost on the ϕ^{jet} distribution. As previously noted, the shape obtained at hadron level is visible also at detector level when the true boost is used. One can also see that 1-jet events tend to be badly reconstructed in the presence of ISR photons, whereas this effect seems to be absent in 2-jet events. As a consequence, the 1-jet distributions would result in large correction factors in the regions $\phi^{jet} \approx \pm\pi$ (figure 8.3(a)).

From what has been presented one can conclude that:

- the effect (distortions) of ISR events in the ϕ^{jet} distributions is most pronounced in 1-jet events, whereas it tends to be smeared in 2-jet events;
- the reconstruction of the boost is the dominant effect responsible for the distortions of the ϕ^{jet} distributions of 1-jet events, and to a much smaller extent in 2-jet events;
- the ϕ-jet distributions of 2-jet events do not seem to be affected to a large extent either by the presence of ISR radiation or by the reconstruction of the boost.

A last check on the reconstruction was done with a Lepto sample where the ϕ distribution of the outgoing partons was forced to a flat distribution ("flat-ϕ sample"). In such a sample the effects of radiated photons are most clearly seen, since one expects the distributions to be flat when no radiation is allowed (figures 8.4 and 8.5). As foreseen, all the ϕ^{jet} distributions for non-radiative events are flat at hadron and detector levels, apart from inevitable statistical fluctuations and small distortions due to possibly badly reconstructed events. The distributions at detector level are particularly interesting, as they show evidence of a good reconstruction of the boost and the absence of effects and cuts distorting them.

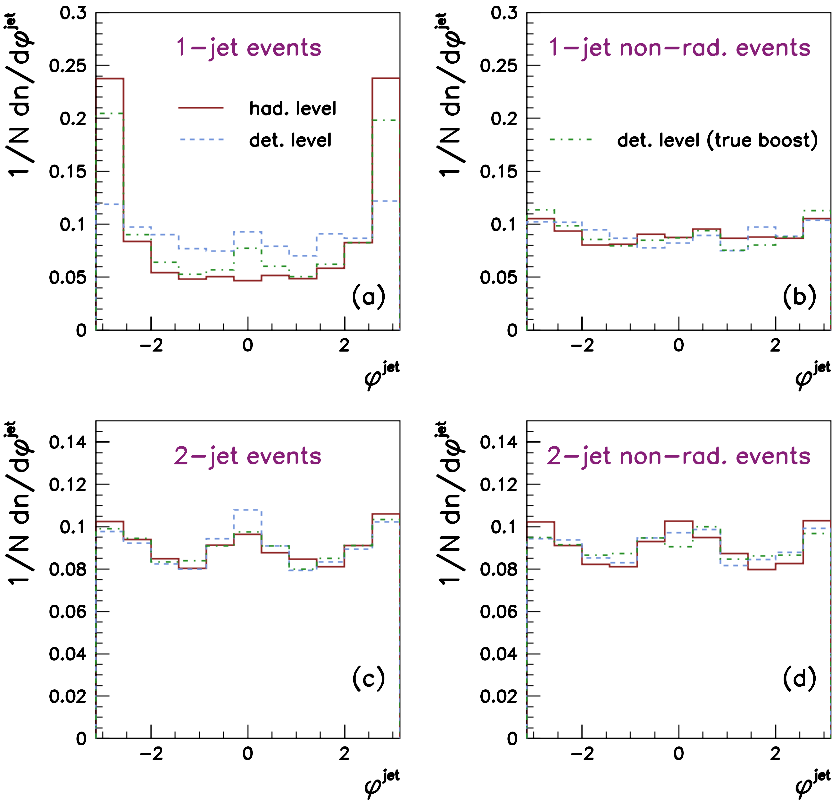

Figure 8.3: *Comparisons between the ϕ^{jet} distributions for 1-jet and 2-jet events obtained with Lepto at hadron and detector levels. Distributions were also obtained with Lepto using the true components of the boost vector to boost the jets at detector level "(true boost)".*

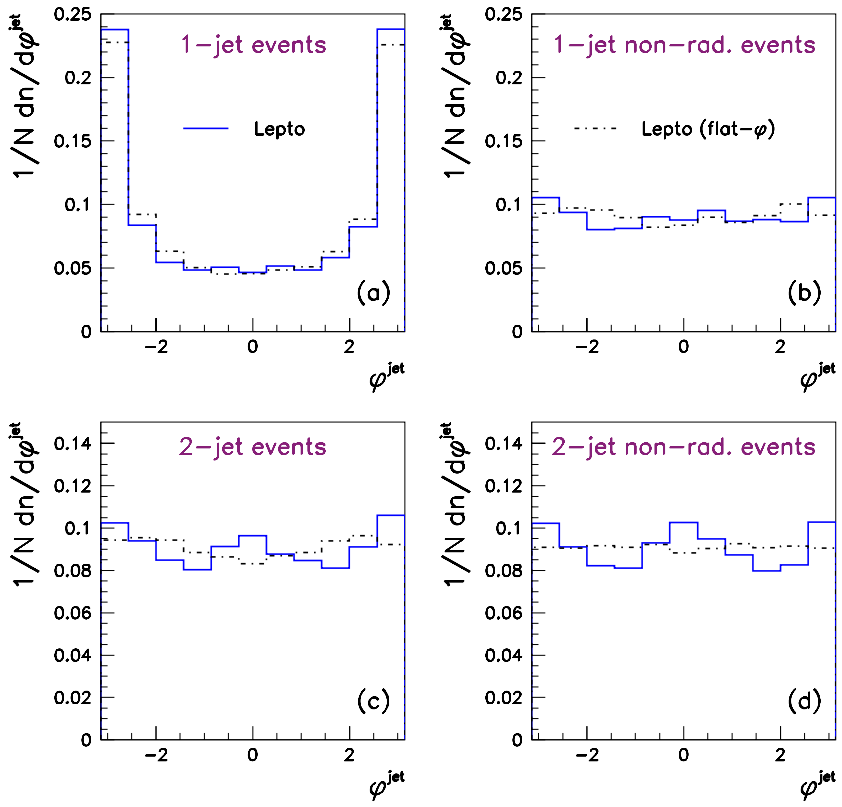

Figure 8.4: *ϕ^{jet} distributions for (a) 1-jet and (b) 2-jet events at hadron level obtained with Lepto and with the "flat-ϕ" Lepto.*

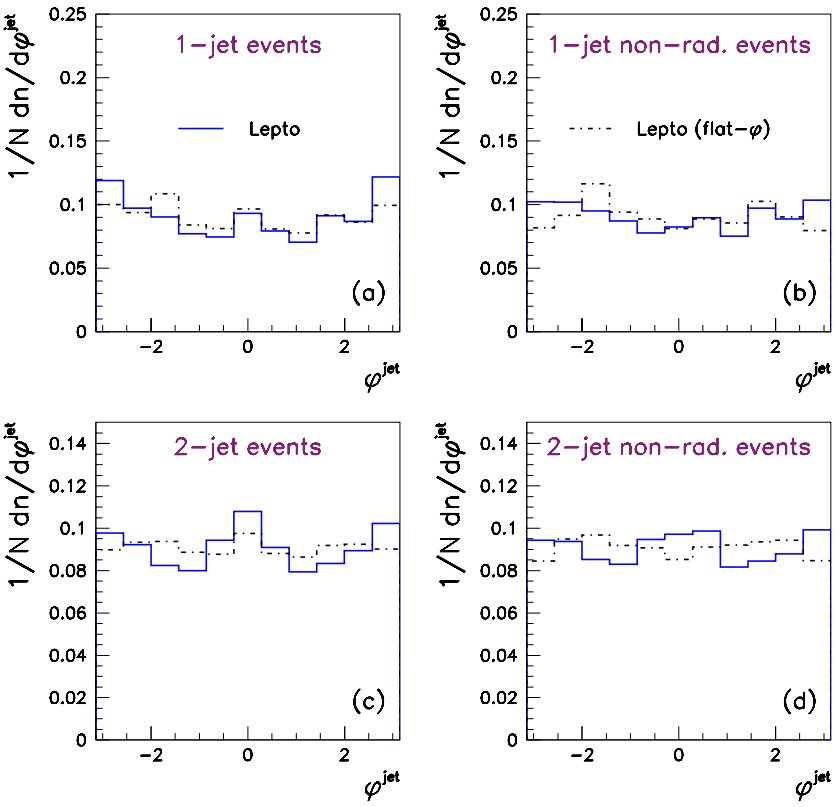

Figure 8.5: *ϕ^{jet} distributions at detector level obtained with Lepto and with the "flat-ϕ" Lepto: (a) 1-jet events, (b) 1-jet non-radiative events, (c) 2-jet events and (d) 2-jet non-radiative events.*

Again, the inclusion of radiation has a net effect in 1-jet events, a situation that does not seem to happen in the case of 2-jet events.

All the above studies strengthen our choice in the analysis of solely dijet events.

8.3 The Raw Distributions

In this analysis we chose to bin the ϕ^{jet} distribution in 11 bins. This binning is motivated by the good ϕ^{jet} resolution and the high purities and efficiencies obtained. The latter are defined as

$$\pi(i) = \frac{\#\text{ of jets generated and reconstructed in bin } i}{\#\text{ of jets reconstructed in bin } i} \tag{8.1}$$

and

$$\epsilon(i) = \frac{\#\text{ of jets generated and reconstructed in bin } i}{\#\text{ of jets generated in bin } i} \quad , \tag{8.2}$$

respectively.

They were calculated from Monte Carlo as follows: the denominators were obtained with a simple independent count in azimuthal angle of the number of selected jets in dijet events at both generated and reconstructed levels. The numerator required more care. All events with 2 reconstructed jets and at least 1 generated jet were considered, *i.e.* dijets were only required at reconstructed level. The jets were then matched between levels by minimising the sum of the distances in $\eta-\phi$ for the 1 or 2 pairs of jets. Note that no cut (usually $d(\eta,\phi) < 1$) on the distance of the matched jets was applied. The numerator was incremented for each pair of matched jets in the same ϕ^{jet} bin.

Figure 8.6 shows the results. Both purities and efficiencies are rather high, in the range $80-90\%$, and are independent of ϕ^{jet}, which provides support in our choice of the binning in the azimuthal angle.

The raw (uncorrected) ϕ^{jet} distribution of 2-jet events is displayed in figure 8.7. It is rather remarkable that these uncorrected data clearly show a distribution dominated by an expected $\cos 2\phi$ term.

The discussion of the applied corrections is presented in the following section.

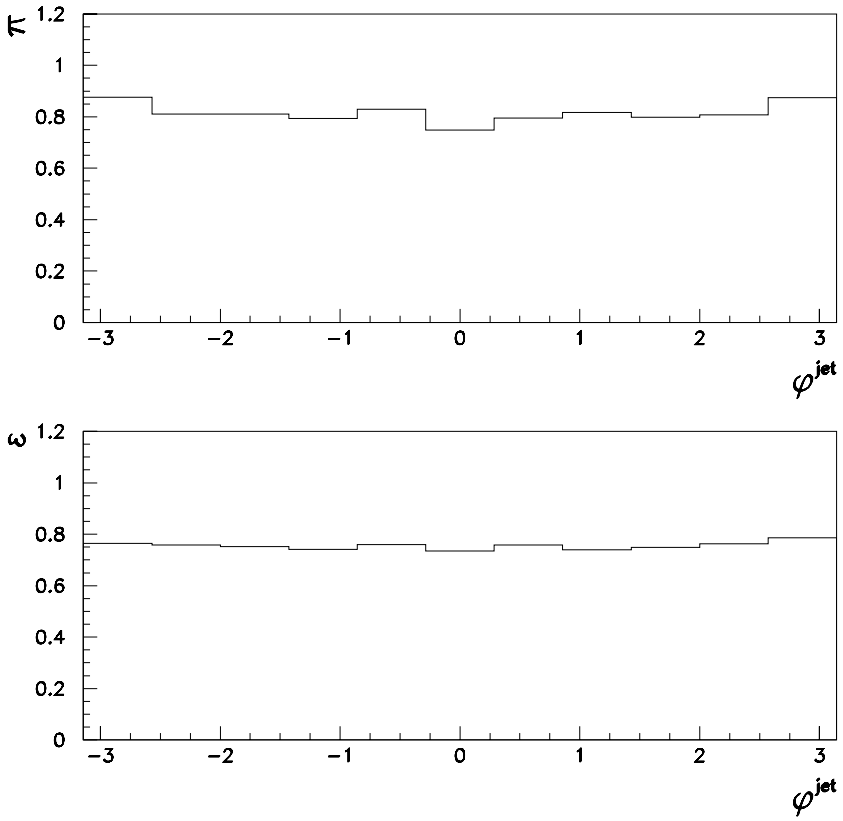

Figure 8.6: *Purity and efficiency obtained with Lepto for the ϕ^{jet} distribution of dijet events.*

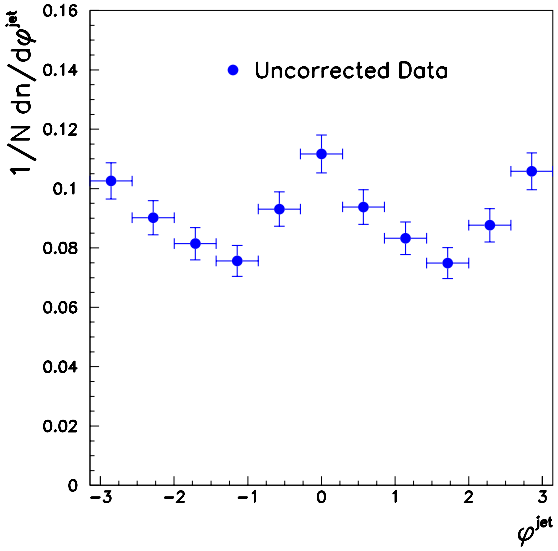

Figure 8.7: *Uncorrected ϕ^{jet} distribution for dijet events.*

8.4 Bin-by-bin Corrections

The data were corrected for detector effects using the bin-by-bin procedure. Using the Lepto sample, the ϕ^{jet} distributions at hadron and detector levels were plotted, and their ratio calculated (*cf.* section 6.6.1 for a detailed discussion of the method). The bin-by-bin correction factor is given by the ratio. For each ϕ^{jet} bin i the corrected distribution is then given by

$$\phi^{jet}_{corr}(i) = \left.\frac{\phi^{jet}_{had}(i)}{\phi^{jet}_{cal}(i)}\right|_{MC} \times \left.\phi^{jet}_{uncorr}(i)\right|_{Data} \quad . \tag{8.3}$$

The bin-by-bin correction factors are shown in figure 8.8. They are reasonably flat and not larger than 10%. The dip around $\phi^{jet} \approx 0$ can be related to the presence of ISR events in the final sample (*cf.* figure 8.3(c)).

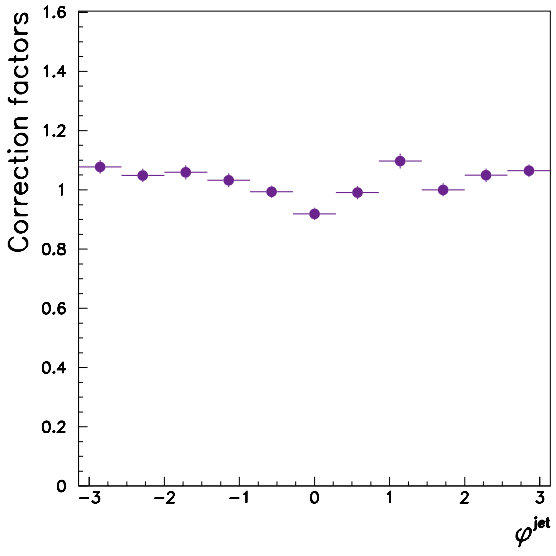

Figure 8.8: *Bin-by-bin correction factors for the ϕ^{jet} distribution of dijet events, as given by Lepto. The error bars represent the statistical uncertainty of the MC sample.*

8.5 Systematic Checks

Most of the systematic checks are common to the analysis of tracks. We briefly recall those which were already explained, and further detail the systematics specific to this analysis. The list of systematic checks that are common to both analyses are:

1. lower the low cut on x by 1σ;
2. raise the high cut on x by 1σ;
3. lower the low cut on y by 1σ;
4. raise the high cut on y by 1σ;
5. calculate the boost parameters using the angular variables as obtained from the CAL information instead of the "best" information (as described in section 5.3.2);
6. apply no energy corrections to correct for energy losses in the active material between the interaction point and the CAL;
7. add an η_{max} cut to further remove diffractive events;
8. use the "alternative" MC program to correct the data; here Ariadne was used for the check since Lepto was used as default.

Several checks – explained below – are directly related to the jets and were therefore absent in the analysis of charged particles. We further investigated the effects of the jet selection cuts on the dijets ϕ^{jet} distribution:

1. lower the low cut on the E_T^{jet} by 1.6 GeV (1σ);
2. lower the low cut on the η^{jet} by 0.2 (1σ);
3. raise the high cut on the η^{jet} by 0.2;
4. rescale the energy of the reconstructed jets in the MC by $+3\%$ ("jet energy scale uncertainty");

5. rescale the energy of the reconstructed jets in the MC by -3%;

The fractional shifts from the central values

$$\left(1/N\ dn/d\phi^{jet}|_{\rm syst} - 1/N\ dn/d\phi^{jet}|_{\rm central}\right)/1/N\ dn/d\phi^{jet}|_{\rm central}$$

are presented in figure 8.9.

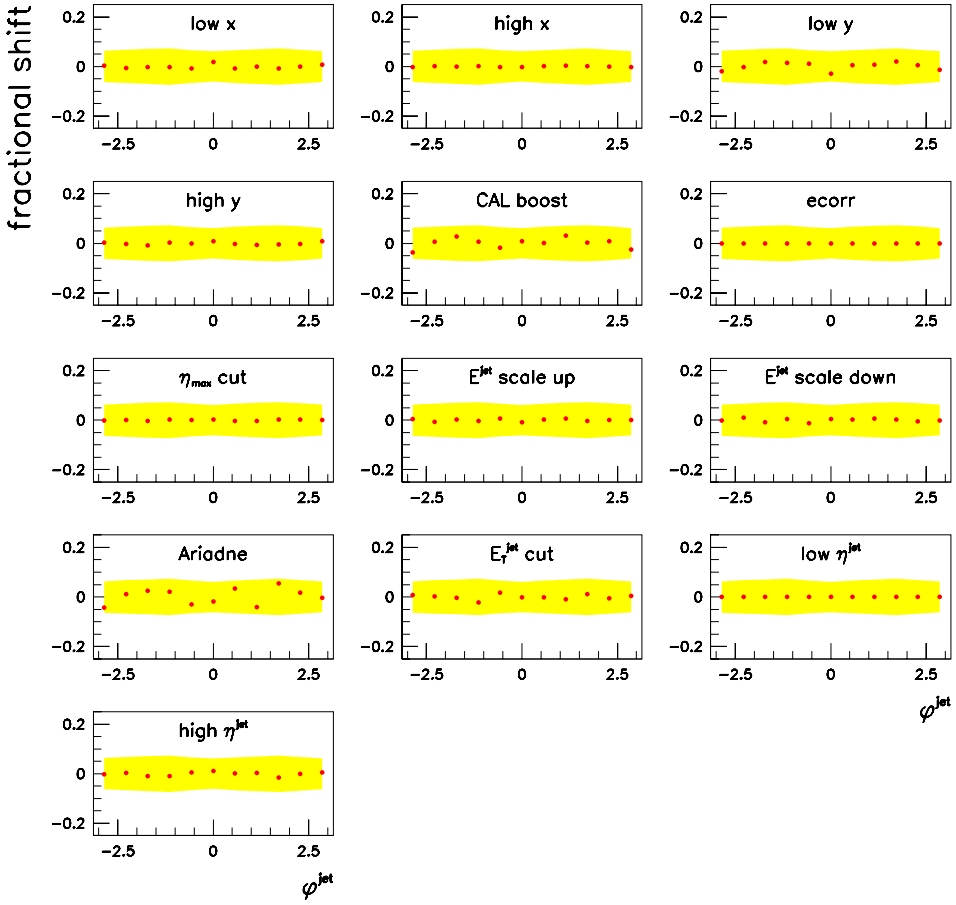

Figure 8.9: *Fractional shifts from the central values of the bin-by-bin corrected ϕ^{jet} distribution for dijet events for the various systematic checks. The (yellow) shaded areas show the size of the statistical errors on the corrected distribution.*

The systematic checks on the E_T^{jet} and η^{jet} of the jets are important, to verify that the results are stable with respect to the jets selection. They also take into account the fact that the jet variables used in the selection have a finite resolution.

The uncertainty in the energy scale of the jets is closely related to the CAL energy scale uncertainty. In this analysis we decided to take a conservative estimate of the jet energy scale uncertainty: we assumed it is 3%, which corresponds to the worse CAL energy scale uncertainty – in the RCAL. The energy of the jets at detector level in the MC was scaled up and down by 3% before any jet energy corrections or selection cuts were applied.

As a general conclusion, the ϕ^{jet} distribution is rather insensitive to most of the systematic checks, *i.e.*, it is rather stable. Note that we are here interested solely in the shape of the distribution. The analysis is well controlled, since all the systematic effects are well within the statistical (experimental) errors. The most significant systematic uncertainty comes from the MC used to correct the data.

8.6 Effects of QED Radiation

The data needs to be corrected for the effects of QED radiation in order to make a meaningful comparison with next-to-leading order QCD calculations. In all that follows the dijet azimuthal angle distribution has always been corrected for both detector and QED effects. We used Lepto to calculate the QED radiation correction factors, which turn out to be small, less than 10%.

8.7 Dijets Azimuthal Angle Distribution

Figure 8.10 presents the ϕ^{jet} distribution for dijet events after correction for detector and QED effects, as explained above. The uncertainty due to the systematics was also included: the inner and outer bars represent the statistical, and the statistical and systematic errors added in quadrature, respectively.

As with the analysis of tracks, this ϕ^{jet} distribution was then fitted with a function of the form given in equation 6.7. We obtained the following results:

- $\langle \cos \phi^{jet} \rangle = -0.001 \pm 0.009$;

- $\langle \cos 2\phi^{jet} \rangle = 0.084 \pm 0.014$;
- $\langle \sin \phi^{jet} \rangle = -0.002 \pm 0.011$.

The $\sin \phi$ term was again included in order to check its agreement with a null value as predicted from QCD, a prediction which is verified. As expected, the shape is given by the $\cos 2\phi$ term, and the $\cos \phi$ contribution is found to be zero.

This distribution will be compared to the LO and NLO Monte Carlo programs in chapter 10.

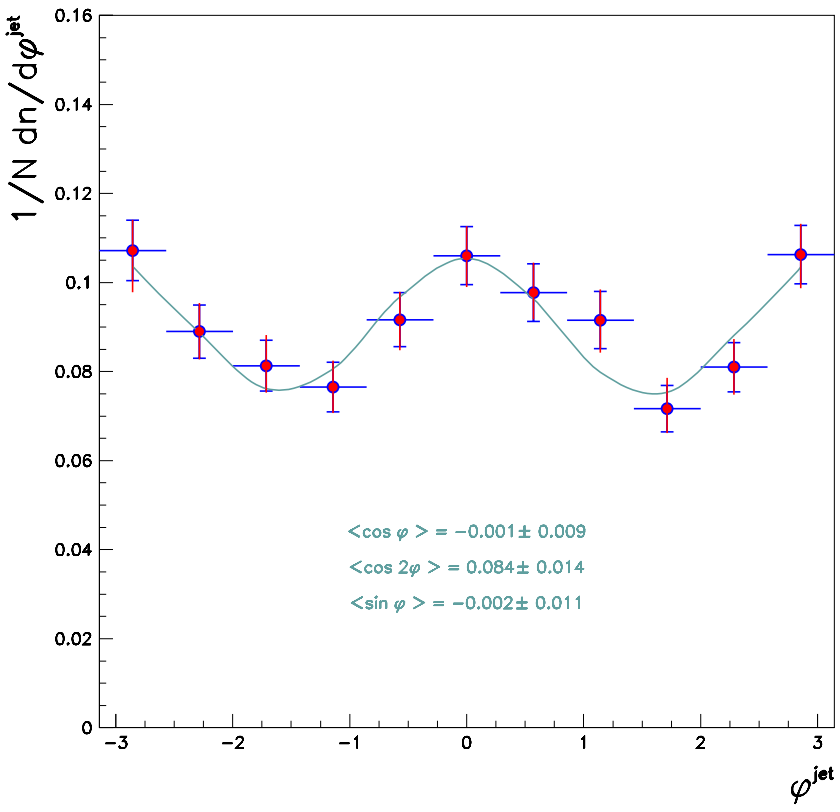

Figure 8.10: *Bin-by-bin corrected ϕ^{jet} distributions for dijet events. The inner and outer error bars represent the statistical and the statistical plus systematic errors added in quadrature, respectively. The solid line shows the result of the fit of the distribution to the theoretical form of equation 6.7.*

Chapter 9

Neural Network Approach to Jet and QCD Event Type Tagging

9.1 Introduction

The previous chapter presented the analysis without identification of the type of the jet, quark- or gluon-initiated. In this second jets analysis we will focus on an attempt to study in detail the azimuthal asymmetries in separated samples of quark- and gluon-initiated jets in QCDC and BGF events.

To achieve this aim it is necessary to identify the flavour of the jets and the QCD type of the event. These two identification and separation tasks are far from being straightforward. In the kinematic region of interest to us, there is no clear separation of quark and gluon jets based on their kinematic variables and both QCDC and BGF events present similar characteristics in terms of the underlying reconstructed parent parton variables $\hat{s}$ and x_p. Only the η^{jet} distribution shows differences: the gluon jets tend to correspond to the most forward jet.

Both types of jets show general similar characteristics, but QCD predicts gluon jets to be slightly broader on average, to exhibit a softer transverse momentum distribution of its constituents, and to have a higher multiplicity than quark-initiated jets. The differences are expected on the basis of the different quark-gluon and gluon self-couplings. The

study of the internal structure of jets therefore potentially provides a way of distinguishing between quark and gluon jets.

Several variables sensitive to the internal structure of jets were studied. Due to the complexity of the task, that involves a pattern recognition problem in a multi-dimensional space – of the discriminating variables – we opted for a neural network approach to tackle the identification of the type of the jets.

Assuming the jet clustering algorithm gives a correct representation of the underlying partonic structure of the event (see discussion in section 9.6.1), the jets found should relate closely to the QCD event type – QCDC or BGF. All these ideas will be fully exploited and investigated in this chapter.

We first describe the basics of multi-layer perceptrons, a type of artificial neural network often used in high energy physics for classification problems. We then present the MLPfit package [69], which we used, and the procedure of jet type identification based on several discriminative variables. Finally we explain the procedure taken to tag the QCD event type, expose the limitations of the method and explain in detail the results.

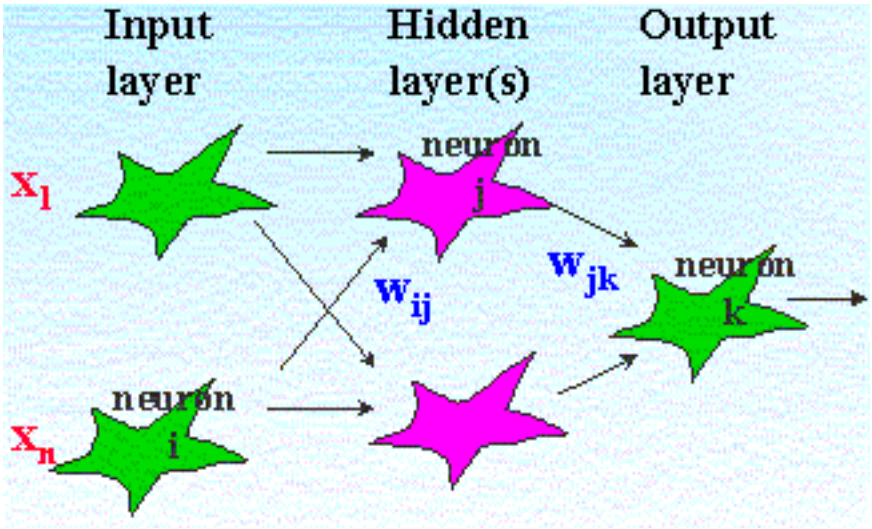

Figure 9.1: *A schematic diagram of a multi-layer perceptron (MLP).*

9.2 Artificial Neural Networks

Artificial Neural Networks (ANN) [58] are nowadays being used frequently in High Energy Physics. The range of applications are as varied as pattern recognition and classification problems, online triggering and tracks reconstruction. The classification tasks include particle identification, event classification and search for new physics[1].

ANNs are strongly inspired by studies of the human brain and are the outcome of models of its functionality. The basic building block of an ANN is therefore called the "neuron", or node. It is an elementary unit, possessing inputs and outputs. A NN can be seen in a most general way as a set of processing units that communicate by sending signals to each other over a (usually large) number of weighted connections.

There are several types of NNs. They are categorised in 2 main classes depending on the way the data is processed by the units. In a feed-forward network the data is processed in a strict forward way, from the input neurons to the output neurons; no feedback is allowed. In a recurrent network the information can be fed back. The typical example of a feed-forward network is the multi-layer perceptron (MLP), which is the most popular and widely used – at least in High Energy Physics.

9.2.1 The Multi-layer Perceptron

A multi-layer perceptron (MLP) generally consists of three layers of neurons. A schematic drawing of a MLP with one, so-called, hidden layer, is pictured in figure 9.1. The first layer, the input layer, merely feeds the input (discriminating) variables $x_1, \ldots, x_n$ into the first array of processing units; the number of input neurons always equals the number of discriminating variables. Each neuron of the second layer, the so-called hidden layer, receives the output (equal to the input) from all the neurons of the input layer and redistributes its output, after processing, to the neurons of the output layer. Sometimes it is useful to use a second hidden layer as it facilitates the learning stage.

[1] *e.g.* the ACAT series of international workshops on Advanced Computing and Analysis Techniques in Physics Research devote part of their program to these issues.

The input to a (hidden) neuron is calculated from a threshold, weighted sum of the output O_i of each of the neurons i of the preceding layer (the input variables x_i in the case of a single hidden layer):

$$y_j = \omega_{j0} + \sum_{i=1}^{n} \omega_{ij} O_i \quad . \tag{9.1}$$

n is the number of neurons in the previous layer, and ω_{ij} is the weight of the connection between neuron j and the neuron i of the preceding layer. ω_{j0} is a threshold – also called offset – of neuron j.

In general, the state of a neuron is evolved by means of an activation function f that changes the state O_j of the neuron to a new state depending on the neuron's inputs and its connections, $O_j \rightarrow f(O_i, \omega_{ij})$. The output of neuron j is obtained applying the activation (or threshold, transfer) function f:

$$O_j = f(y_j) = f\left(\omega_{j0} + \sum_{i=1}^{n} \omega_{ij} O_i\right) \quad . \tag{9.2}$$

The most commonly used activation function is the "sigmoid" function:

$$f(x) = \frac{1}{1 + e^{-\lambda x}} \tag{9.3}$$

depicted in figure 9.2 for various values of the λ parameter.

The most simple MLP comprises one single hidden layer. Assuming that the activation function of the hidden layer is the sigmoid and that the output neurons have a linear output ($O_k = y_k$), then this simple MLP effectively computes a linear combination of sigmoids. Such an architecture is already most useful due to 2 theorems (quoted from the user guides of the MLPfit software package MLPfit [69]):

1. a linear combination of sigmoids can approximate any continuous function of one or more variables [70];

2. trained with a desired answer of 1 for the signal and 0 for the background, the approximated function of the input vector $\vec{x}$ is the probability of signal knowing $\vec{x}$ [71].

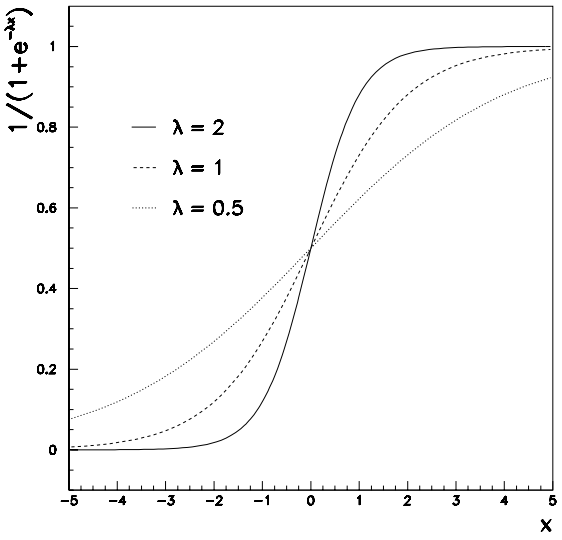

Figure 9.2: *Graph of the sigmoid function.*

These 2 theorems constitute the mathematical justification of MLPs, and are largely responsible for the popularity of MLPs in scientific applications such as classification and approximation problems.

For the purpose of applications, a network needs to be configured or trained to respond in an adequate way to its inputs. A learning rule is needed to determine how the connection weights need to be updated as the training evolves. The technique of supervised learning relevant to MLPs will be described in the next paragraph. Then we will apply these concepts to the context of the analysis of jets.

9.2.2 Training the Network – Supervised Learning

A MLP-based analysis consists of three steps: (1) the supervised learning, (2) a test of the network generalisation, and (3) the final use of the NN function on real examples.

Each event presented to the MLP consists of a set of discriminating variables defining a pattern; and each pattern belongs to one of two classes, one being the signal and the other the background. The signal (background) is usually identified with the output 1 (0). In the discussion that follows we will consider the rather more general case of several output neurons.

The technique of supervised learning consists in finding the set of weights ω_{ij} that give, for a pattern p, a NN output (vector) value $\vec{O}^{(p)}$ as close as possible to the known (true) value $\vec{T}^{(p)}$: the hidden nodes have the task of correlating and building up an "internal representation" of the patterns to be learned. Training the network corresponds to changing the weights ω_{ij} such that a given input pattern $\vec{x}^{(p)} = (x_1, \ldots, x_n)$ gives rise to an output value $\vec{O}^{(p)}$ that equals the desired output or target value $\vec{T}^{(p)}$. Mathematically, the learning translates into the minimisation of the "energy" function defined as

$$\begin{aligned} E &= \frac{1}{2}\sum_p \sum_k (O_k^{(p)} - T_k^{(p)})^2 \\ &= \frac{1}{2}\sum_p e_p \quad . \end{aligned} \tag{9.4}$$

The sums run over all patterns p and the output neurons k (usually a single one).

The weights are initialised with random values. As the learning procedure evolves (the evolution is here represented by the time variable t), all weights are updated according to the simple rule

$$\omega_{ij}(t+1) = \omega_{ij}(t) + \Delta\omega_{ij}(t+1) \quad . \tag{9.5}$$

The updates $\Delta\omega_{ij}(t+1)$ are calculated by back-propagation (explained hereafter) of the errors, and take the general form

$$\Delta\omega_{ij}(t+1) = -\eta\frac{\partial e_p}{\partial \omega_{ij}} + \alpha\Delta\omega_{ij}(t) \quad , \tag{9.6}$$

where η is the learning parameter and α is a momentum term. The learning parameter sets the importance the changes in the energy function will have on the updates of the weights, whereas the momentum term will weight the importance the last update will have on the latter. Both improve the convergence of the method [58].

The changes on the weights are given by (from equation 9.6)

$$\Delta\omega_{jk}(t) = -\eta\delta_k O_j + \alpha\Delta\omega_{jk}(t-1) \tag{9.7}$$

for a hidden neuron j to an output neuron k, with

$$\delta_k = (O_k - T_k) f'(\sum_k \omega_{jk} O_j) \quad . \tag{9.8}$$

f' is the derivative of the activation function f. For the input i to hidden j neurons, equation 9.6 can be shown [58] to be

$$\Delta\omega_{ij}(t) = -\eta \left(\sum_k \omega_{jk}\delta_k\right) f'\left(\sum_i \omega_{ij}x_i\right) x_i + \alpha\Delta\omega_{ij}(t-1) \quad . \tag{9.9}$$

In the above equation the indexes i and k run, as usual, over the input and output nodes, respectively.

Several learning methods exist which differ in the approach used to minimise the energy function 9.4. Of the multitude of possible approaches the most common and simple is the one just presented, based on formula 9.6. The basic idea of this algorithm is rather simple: the new weights are updated at each step with an error that is calculated to minimise the total energy. And the errors in the connections are back-propagated, starting from the output layer, up to the connections between the input and the hidden layer. This is the reason why this learning algorithm was given the name of back-propagation learning rule.

9.3 The NN Set-up – the MLPfit Package

The software package MLPfit [69] was chosen as the tool to set up, train, and use, multi-layer perceptrons. It allows the creation and usage of 1- or 2-hidden layer feed-forward MLPs.

By default, MLPfit uses a linear transfer function of the output neuron(s), and the sigmoid function for the hidden layer(s). We have also used the default learning method in MLPfit: the performant BFGS algorithm (see the MLPfit user guide [69] for details) which employs more sophisticated numerical methods of minimisation (of the energy, equation 9.4) than the traditional back-propagation algorithm described above. The latter was described for the purpose of an intuitive description of the problems at hand.

In the applications that follow the NN was trained on a subset of the Lepto sample and then cross-checked with a sample never presented before to the network and with the total sample. As the results were virtually the same we decided to present the final MC results obtained with the full sample, to have access to larger statistics.

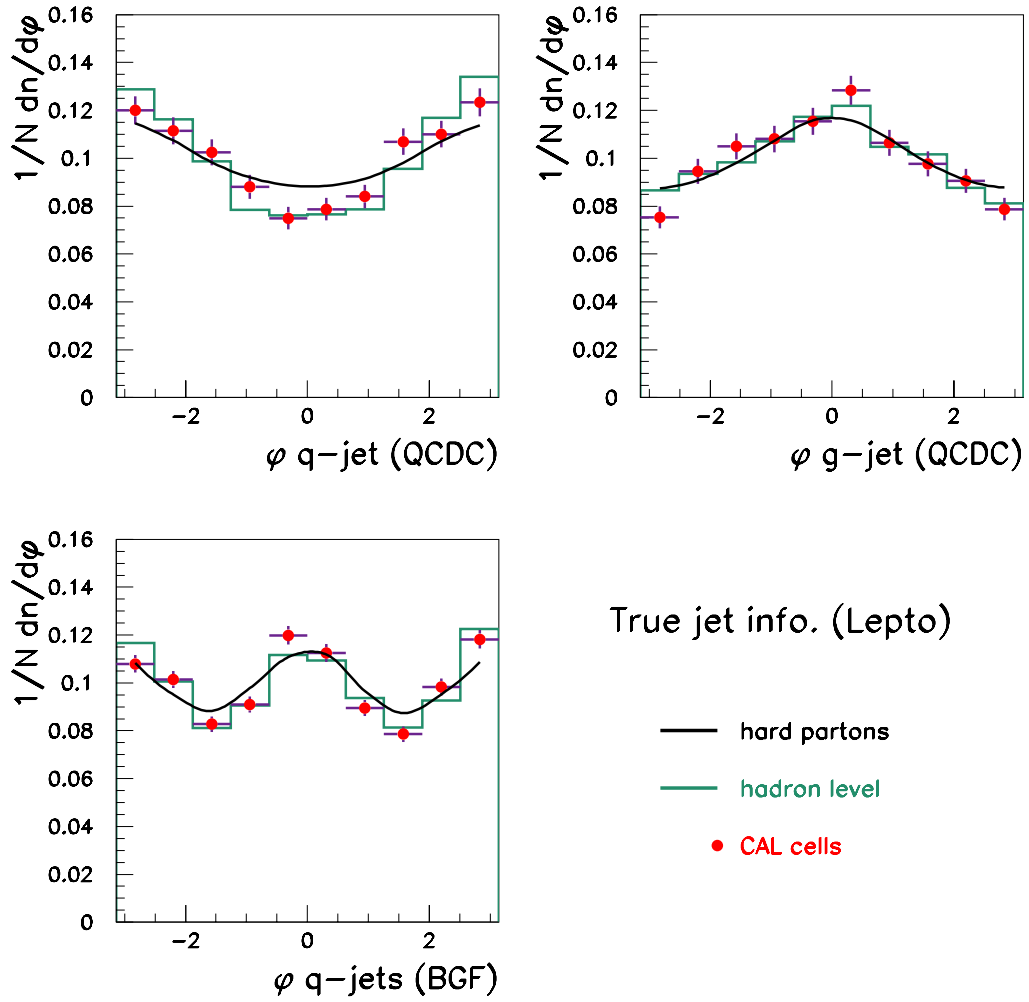

Figure 9.3: *The azimuthal angle distributions at parton, hadron and detector (using CAL cells) levels obtained with Lepto. The "true" MC information as explained in section 9.4.1 was used to identify and separate jets. But the QCD event type was then tagged as explained in section 9.6.*

9.4 Separation of Quark and Gluon Jets

Given the nature of the analysis, it was not obvious from the outset that the goal to measure the ϕ distributions in separate quark and gluon jets in QCDC and BGF events could be achieved. But as figure 9.3 proves, our proposed approach for the separation of the various contributions gives the expected results provided the jet type identification is correct: the figure shows the ϕ distributions for separated quark and gluon jets in tagged QCDC and BGF events at the parton, hadron and detector levels. In what follows we will describe in great detail the approach used to obtain these distributions.

9.4.1 Quark and Gluon Jet Identification in MC Programs

The first step in this analysis was the "building" of samples of quark- and gluon-initiated jets. The identification of quark and gluon jets in DIS MC programs such as Lepto and Ariadne is not a straightforward task. The assignment of the jet type (or flavour) to even parton level jets is not unique.

Both these programs use the Lund string hadronisation model implemented in JETSET, whose event record does not provide a direct way of assigning the jet's flavour, as the decay history is based on the strings in the hadronisation process. One has therefore to rely on a suitable flavour identification algorithm for jets.

The usual picture of a jet sustains the idea that it originates from the fragmentation and hadronisation of a parton – a quark or a gluon. Hence the identification of the type of a jet should be ideally done relating the jet to a hard parton (a *hard parton* refers to a parton coming directly from the hard scattering vertex). Unfortunately, not all available MC programs give us the possibility to do so. Lepto keeps in its event record the hard parton(s), but Ariadne's event record looses track of it (them) when the colour dipole cascade is applied. This is the important reason why we decided to use Lepto in this analysis rather than Ariadne.

It is straightforward to select the hard parton(s) in Lepto. The jet type identification can then be accomplished by simply relating the jet to a hard parton: the hadron or

detector level jet is identified as a quark or gluon jet according to the flavour of the hard parton closest to its axis. We use the usual definition of distance in the $\eta - \phi$ plane (as in equation 7.11).

It might happen that there are more selected jets than hard partons: the jet algorithm might return two selected jets in a QPM event (very seldom), or the event is a 3- or 4-jet event, etc. Hence, there is the possibility to either identify the same number of, say, highest E_T jets, as there are of hard partons, performing the matching simultaneously between the two sets, or simply decide to perform the matching in $\eta - \phi$ separately for each jet. The latter option was used in these studies.

The performance and reliability of this jet type identification algorithm will not be described here. We detailed these studies for this and 3 other approaches in a ZEUS note [72]. The investigations showed this method is most reliable and efficient.

9.4.2 Quark and Gluon Jet Discrimination in Data Samples

As mentioned in the introduction section of this chapter, the separation of quark- from gluon-initiated jets in data samples is a rather complex task. In e^+e^- interactions, efficient and successful methods have been derived to tackle this problem in the analysis of 3-jet events. The situation is, unfortunately, rather more complicated in DIS.

We decided to discriminate the jets based on information of their internal structure. A relevant set of variables sensitive to the internal structure of jets was first determined. These discriminating (tagging) variables have then been used as input to a MLP, that was trained on a sub-sample of our total Lepto sample. The following sections detail the various steps: definition and MC description of the discriminating variables, characteristics of the MLP architectures, and results of the NN studies. Note that all these studies were performed at both hadron and detector levels.

Several quantities have been used in the past to study the internal structure of jets in DIS. The HERA collider experiments ZEUS and H1 have published a series of results on jet shapes, subjet multiplicities [73–76].

Three types of structure-sensitive discriminating variables were used in the context

of the present investigation:

- The integrated jet shape;
- The subjet multiplicity;
- Fodor moments.

Also the pseudorapidity η^{jet} of the jets and other general jet variables were considered.

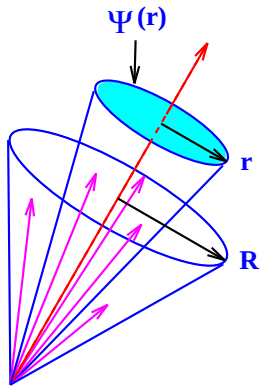

Figure 9.4: *Graphical interpretation of the integrated jet shape* $\Psi(r)$.

9.4.3 The Integrated Jet Shape

The shape of the E_T distribution inside a jet reflects its substructure, and, in particular, is sensitive to multi-parton emission from the primary (parent) parton. The radial distribution of E_T is affected by both the hard and the soft processes.

The mean integrated jet shape is defined as the average (over the jets in the event) fraction of a jet's transverse energy that lies inside a cone of radius r in the $\eta-\phi$ plane concentric with the jet's axes:

$$\langle \Psi(r) \rangle = \frac{1}{N_{jets}} \sum_{\text{jets}} \frac{E_T(r)}{E_T(R=1.0)} \quad . \tag{9.10}$$

The sum runs over all jets in the event (N_{jets}). The calculation is normalised to the jet's transverse energy inside a cone of unity radius $R = 1$ in the $\eta-\phi$ plane. Figure 9.4 gives a pictorial definition of $\Psi(r)$.

Here, jets need to be classified separately; hence we will use the integrated jet shape *for each jet*:

$$\Psi(r) = \frac{E_T(r)}{E_T(R = 1.0)} \quad . \tag{9.11}$$

9.4.4 The Subjet Multiplicity

Stated in the most naïve way, subjets are jet-like structures inside a jet, resolved at a resolution scale which is a fraction of the jet's transverse energy. Within the framework of the k_T jet clustering algorithm, a precise dynamical definition of the number of subjets inside a jet can be given: after the jet clustering algorithm has been applied the algorithm is again run on each jet. For each jet the only particles belonging to it are considered, and the clustering procedure repeated until, for all (i, j) pairs, the d_{ij}

$$d_{ij} = \min\left(E_{T,i}^2, E_{T,j}^2\right) \cdot \left[(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2\right]^{1/2} \tag{9.12}$$

verify

$$d_{ij} > y_{cut} \cdot (E_T^{jet})^2 \quad . \tag{9.13}$$

The remaining particles are called subjets.

The y_{cut} variable is a resolution parameter. Usual values are in the range $10^{-4}-10^{-1}$. At very low y_{cut} the resolution is such that one is simply counting the number of particles inside the jet (for $y_{cut} \to 0$ every particle in a jet is a subjet). Conversely, as $y_{cut} \to 1$, the resolution power decreases and the number of subjets tends to 1, $n_{sbjs} \to 1$.

9.4.5 The Fodor Moments

The Fodor moments of order (m, n) are defined as follows (refer, *e.g.*, to [77]):

$$F_{mn} = \sum_i \left(\frac{p_{\perp,i}}{E^{jet}}\right)^m (\eta_{\perp,i})^n \quad . \tag{9.14}$$

The sum runs over all particles i belonging to the jet, and the transverse momentum $p_{\perp,i}$ and pseudorapidity $\eta_{\perp,i}$ are calculated with respect to the jet's axis.

A general interpretation of the Fodor moments is somewhat difficult. Some have nevertheless obvious meanings: F_{00} is the multiplicity, whereas F_{10} gives the E^{jet}-normalised sum of the relative transverse momenta of the particles in the jet.

9.4.6 MC Description of the Tagging variables

Figures 9.5 to 9.8 present the MC description of the tagging variables defined previously in this chapter as provided by Lepto. It can be seen that Lepto generally describes the overall shapes of the distributions, but fails in some of the details. This is particularly true for the Fodor moments: various distributions of Fodor moments are poorly described in detail by Lepto. But one can conclude from the comparisons for the evolution with η^{jet} of the mean values (figure 9.8) that Lepto nevertheless gives in general a reasonably correct average value over the entire η^{jet} range under study.

9.4.7 Tagging Variables Distributions for Quark and Gluon Jets

This section presents and discusses the distributions of the set of tagging variables investigated: figures 9.9 to 9.12 and figures 9.13 to 9.16 display the distributions of some of the discriminating variables, for separated samples of quark and gluon jets, at hadron and detector level, respectively. In these MC studies the jet flavour was determined as described in section 9.4.1.

Of the jet kinematic variables, the distribution of the pseudorapidity η^{jet} is clearly different for quark and gluon jets: as expected, gluon jets tend to be produced more forward than the quark jets (figures 9.9 and 9.13).

All the differences between the 2 types of jets expected on the basis of QCD (listed in the introduction section of this chapter) can be observed in several of the distributions presented here (*e.g.* at hadron level):

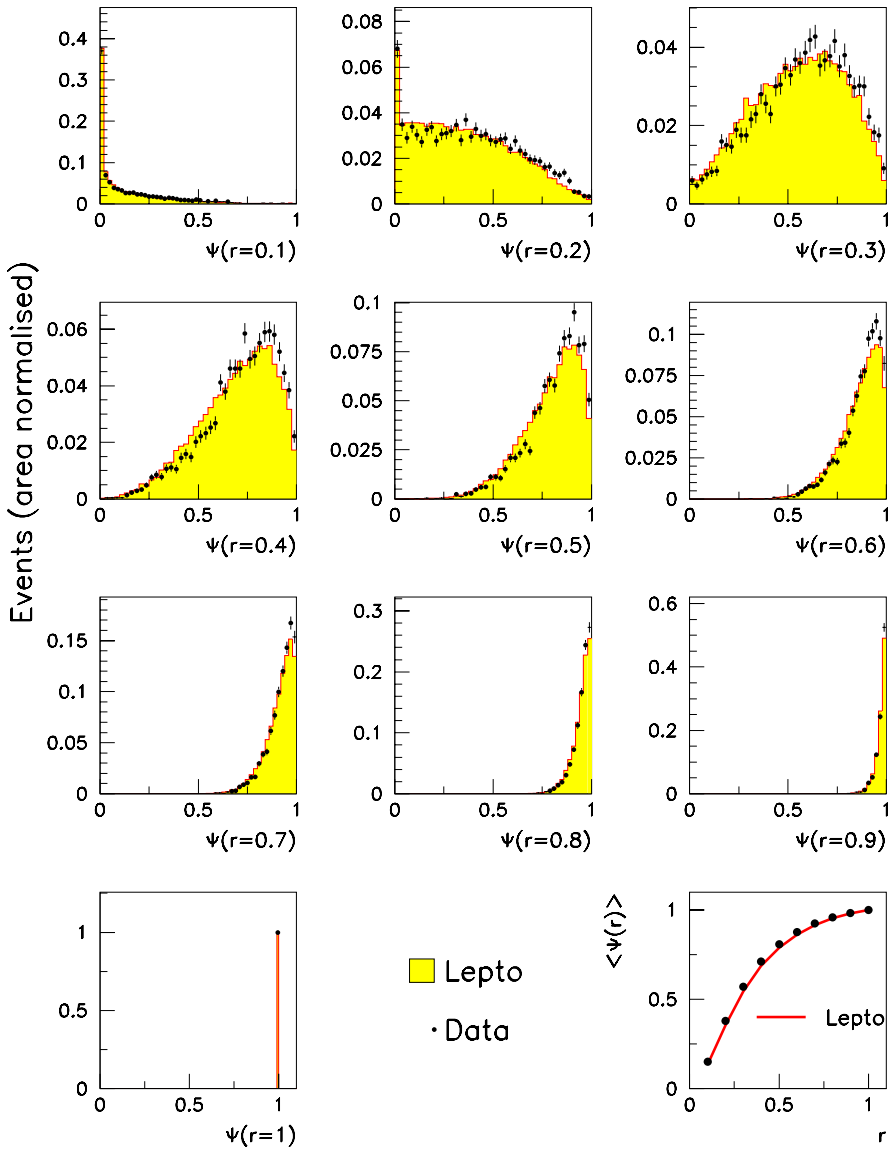

Figure 9.5: *Differential distributions and mean values of the integrated jet shape $\Psi(r)$: comparisons between data and Lepto.*

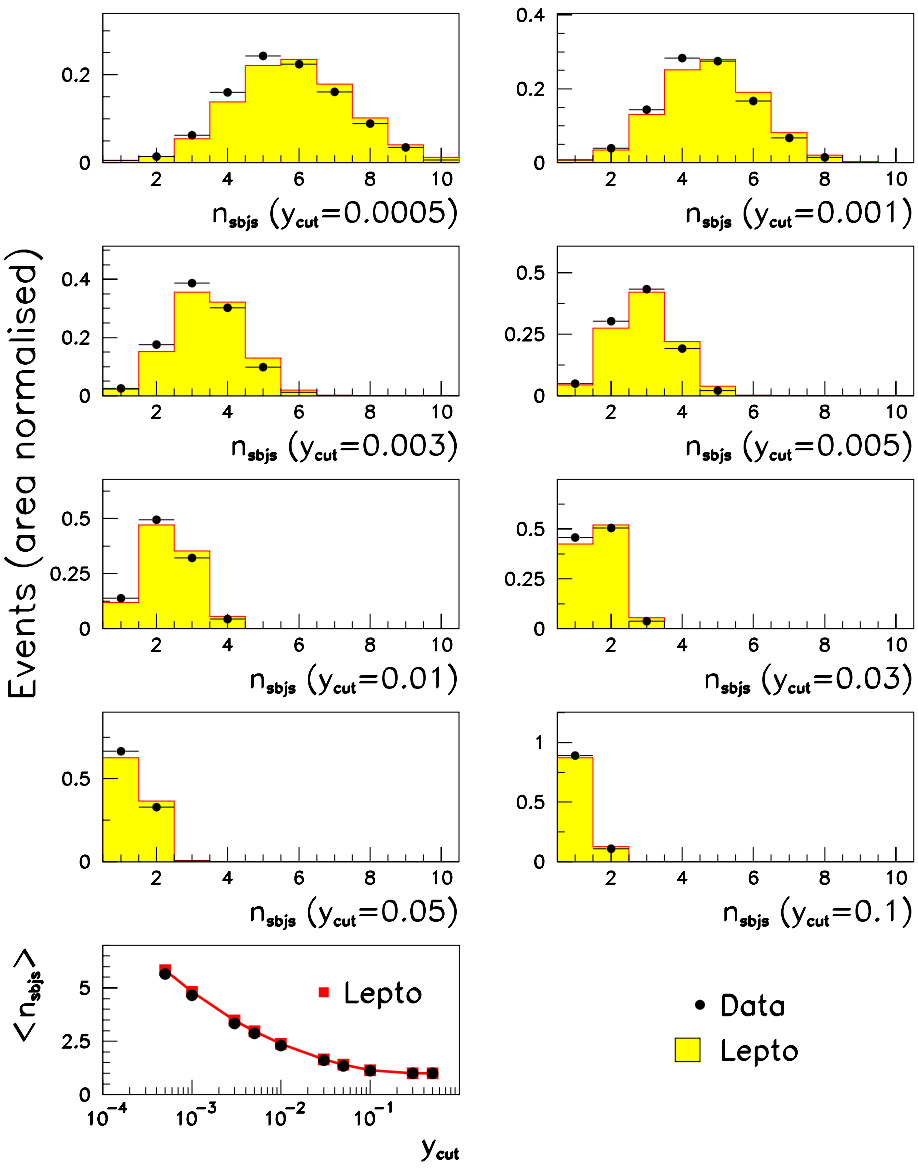

Figure 9.6: *Differential distributions and mean values of the subjet multiplicities: comparisons between data and Lepto.*

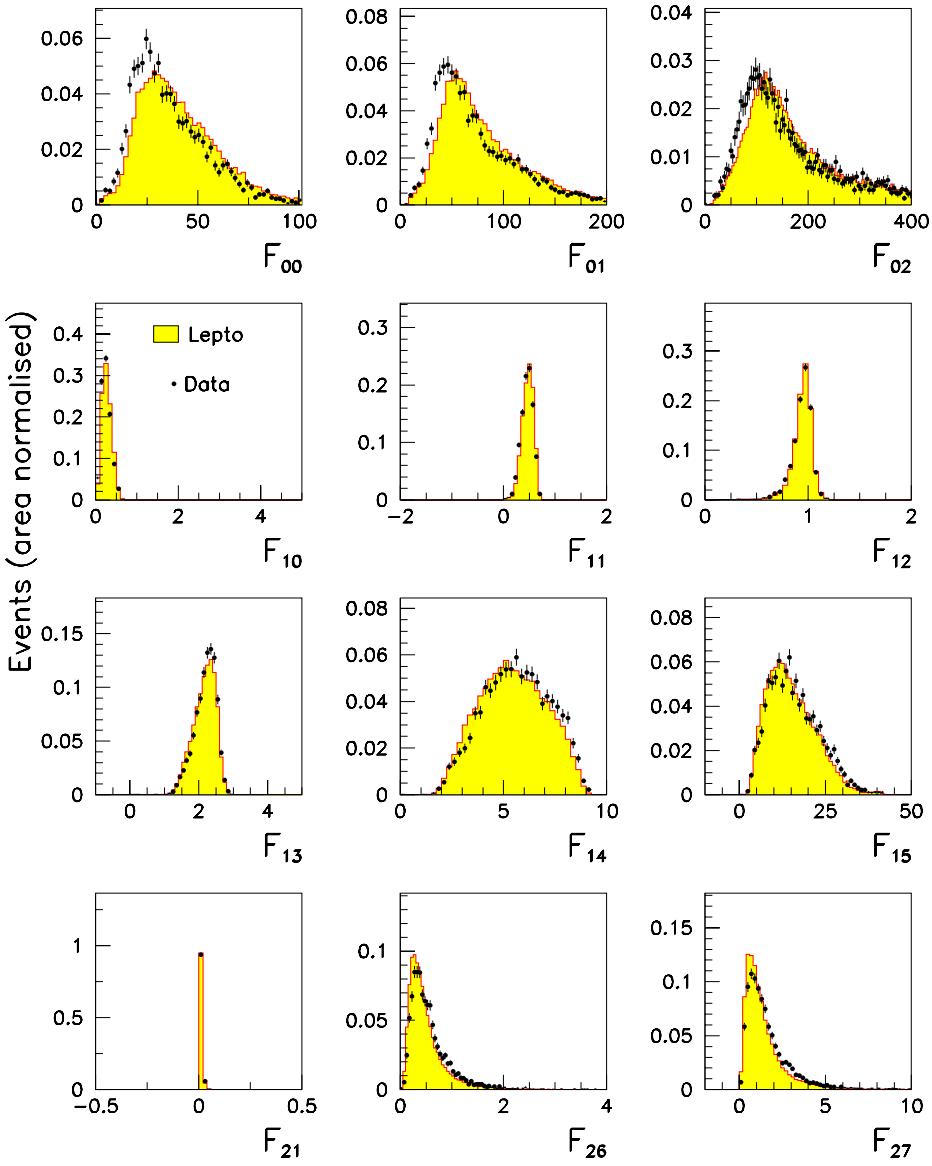

Figure 9.7: *Differential distributions of the Fodor moments: comparisons between data and Lepto.*

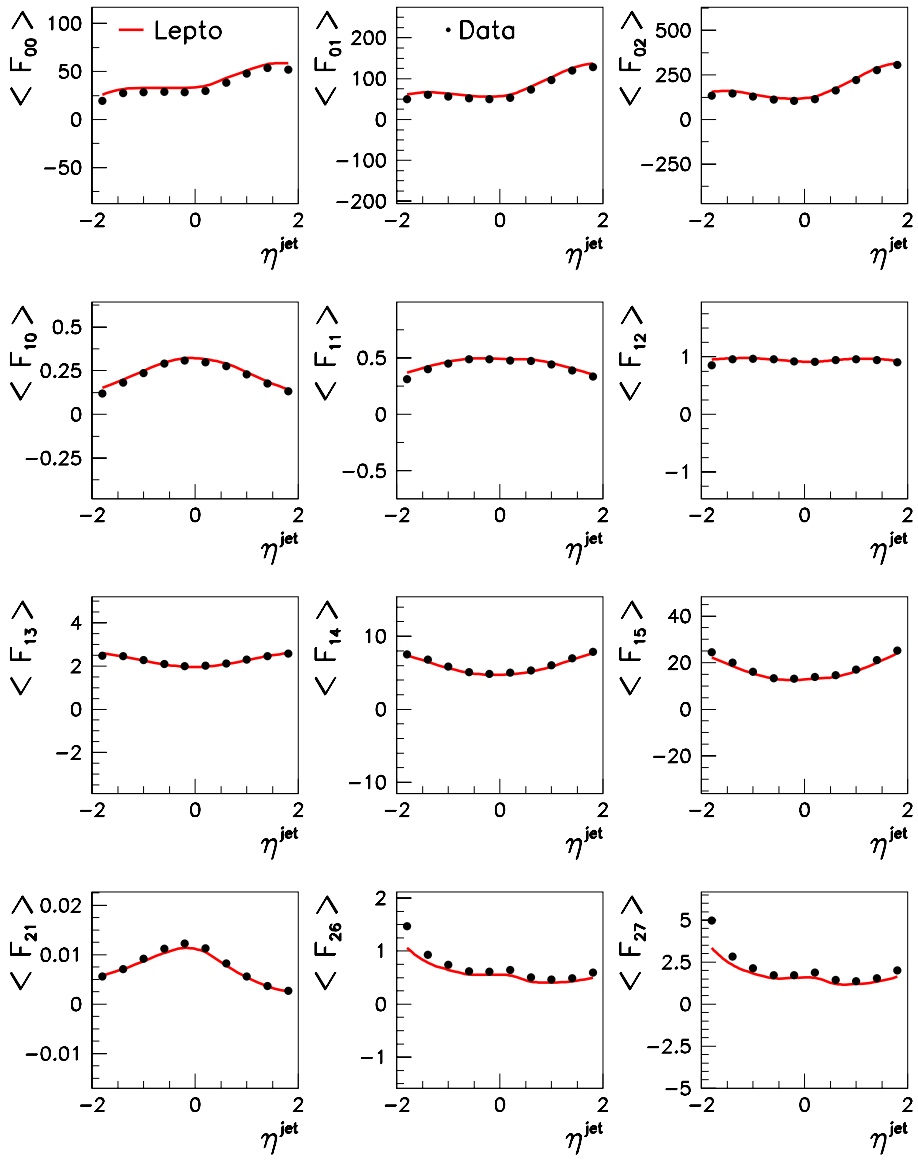

Figure 9.8: *Comparisons between data and Lepto for the description of the evolution with η^{jet} of (the mean value of) the Fodor moments. The solid lines show the results for Lepto.*

- $E^{lead\,part}/E^{jet}$, the distribution of the fractional energy of the leading particle in the jet (figure 9.9): the gluon fragmentation function is expected to be softer that the quark fragmentation function; therefore, this quark jet distribution peaks at larger values for quark than for gluon jets;

- $\Psi(r)$ distribution (figure 9.10): gluon jets are broader than quark jets, since the fraction of the jet's transverse energy inside a cone of radius r is always larger for quark than for gluon jets;

- F_{00} distribution: gluon jets have – on average – a larger multiplicity than quark jets (figure 9.11).

It is also clear from figure 9.10 that gluon jets have a higher subjet multiplicity than quark jets over the entire range of the resolution parameter y_{cut} (except for the high limit, where $n_{sbjs} \to 1$).

The Fodor moments (figures 9.11 and 9.12) are sensitive to both the distribution inside a jet of the E_T of the particles, but also to their position relative to the jet's axis. Again, differences for quark and gluon jets are visible.

The same conclusions can be drawn from the corresponding distributions at detector level. As a general rule, all variables sensitive to the internal structure of jets show differences for quark and gluon jets; and these differences are always more pronounced at hadron than at detector level.

9.5 Jet Tagging NN Training and Studies

A significant number of different MLP architectures was investigated: MLPs with 1 or 2 hidden layers, different number of neurons in the hidden layer(s), and different combinations and number of input discriminating variables.

These studies were all performed separately for hadron and detector level jets for the following reason: when using a NN it is imperative that the distributions of the tagging variables used in the training be reasonably well described by the MC, so that the use

of the NN output function reflects the data. Unfortunately, most of the variables used have different distributions and mean values for hadron and detector level jets. Hence the studies had to be made separately for both levels.

A sub-set of the total MC sample of jets was used during the training of the NN.

All the input variables X were normalised according to the simple rule

$$X \rightarrow \frac{X - <X>}{\sigma_X} \quad ,$$

$<X>$ and σ_X representing, respectively, the mean and the variance of the X distribution, in order for all the variables to vary within similar ranges.

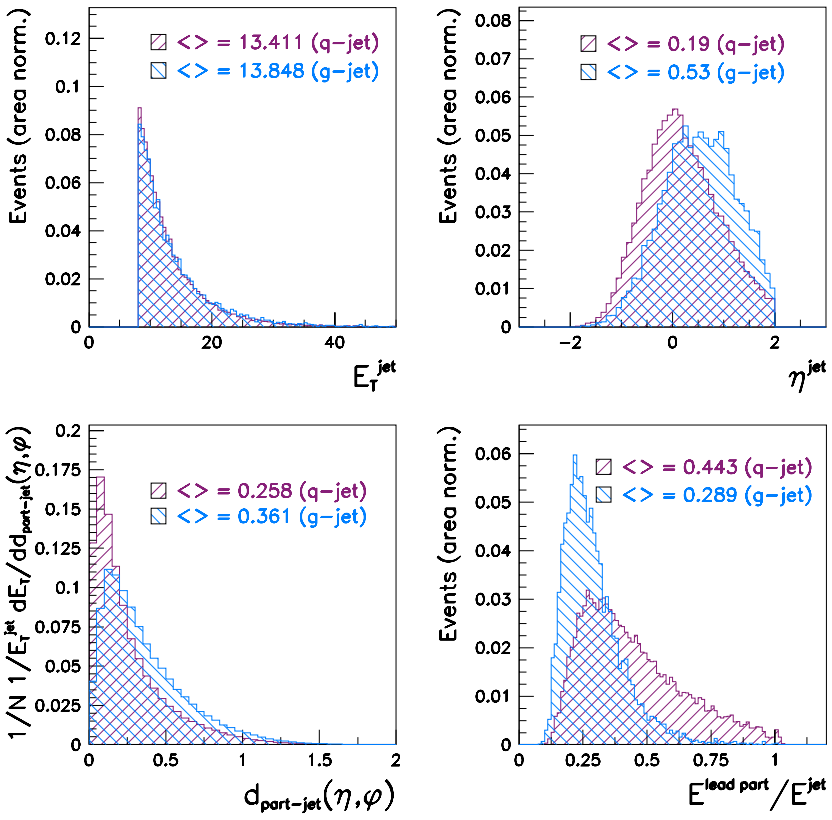

Figure 9.9: *Some discriminating variables investigated at hadron level: quark and gluon jet distributions of E_T^{jet}, η^{jet}, E_T differential distribution in the $\eta - \phi$ distance of particles within a jet, and distribution of the fractional energy of the leading particle in the jet ($E^{lead\,part}/E^{jet}$). All variables were computed in the Breit frame.*

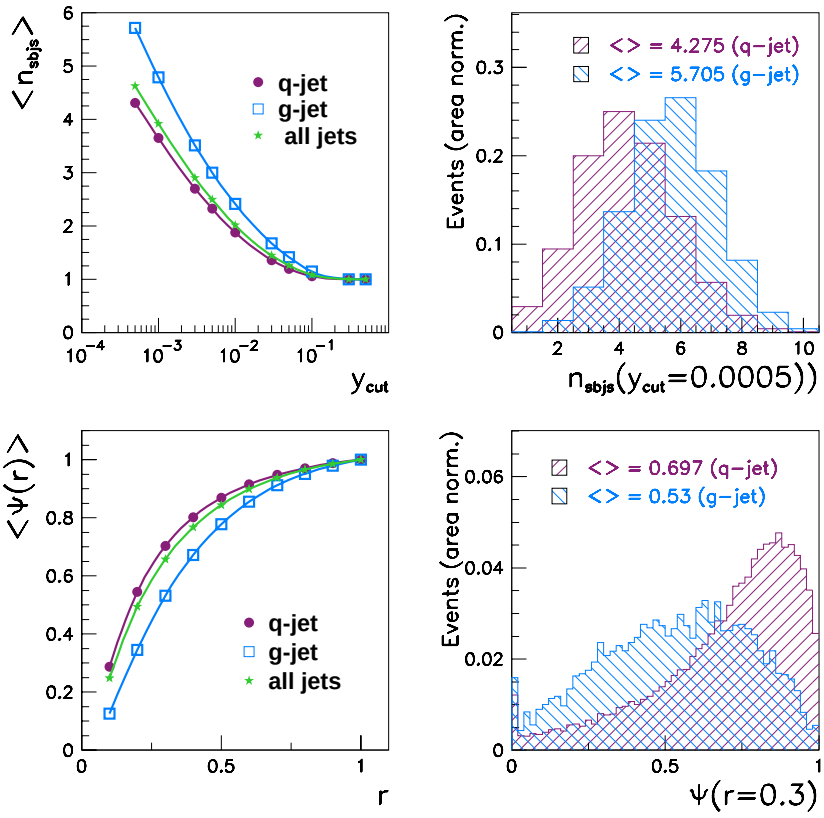

Figure 9.10: *The subjet multiplicity n_{sbjs} and integrated jet shape $\Psi(r)$ for quark and gluon jets at hadron level: means and differential distributions for a specific value of y_{cut} and r, respectively.*

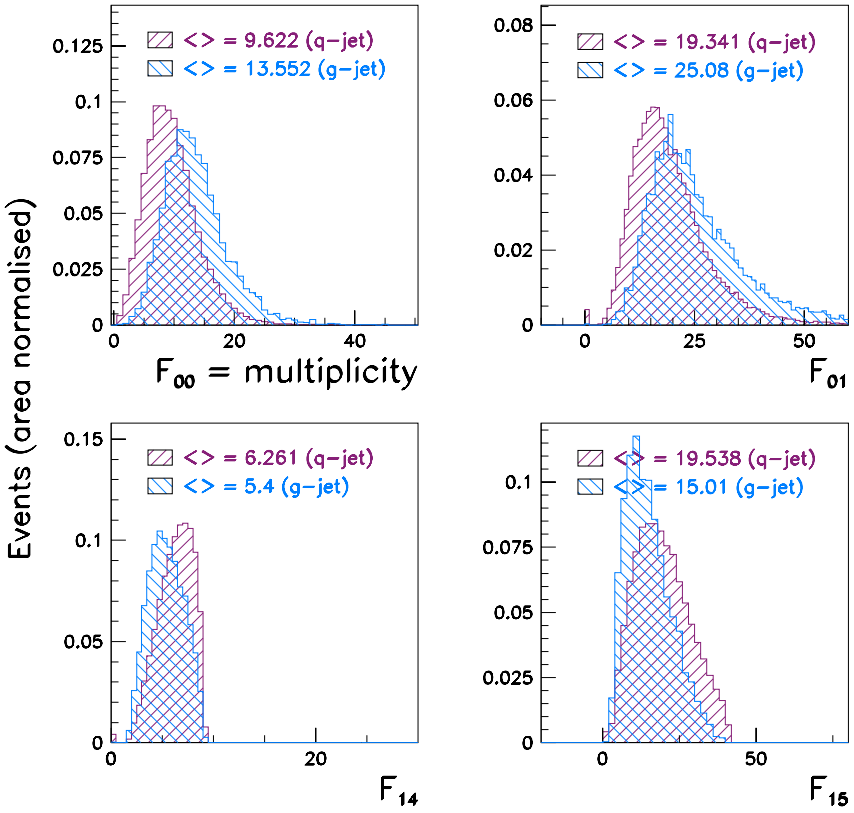

Figure 9.11: *Differential distributions of some Fodor moments for quark and gluon jets at hadron level.*

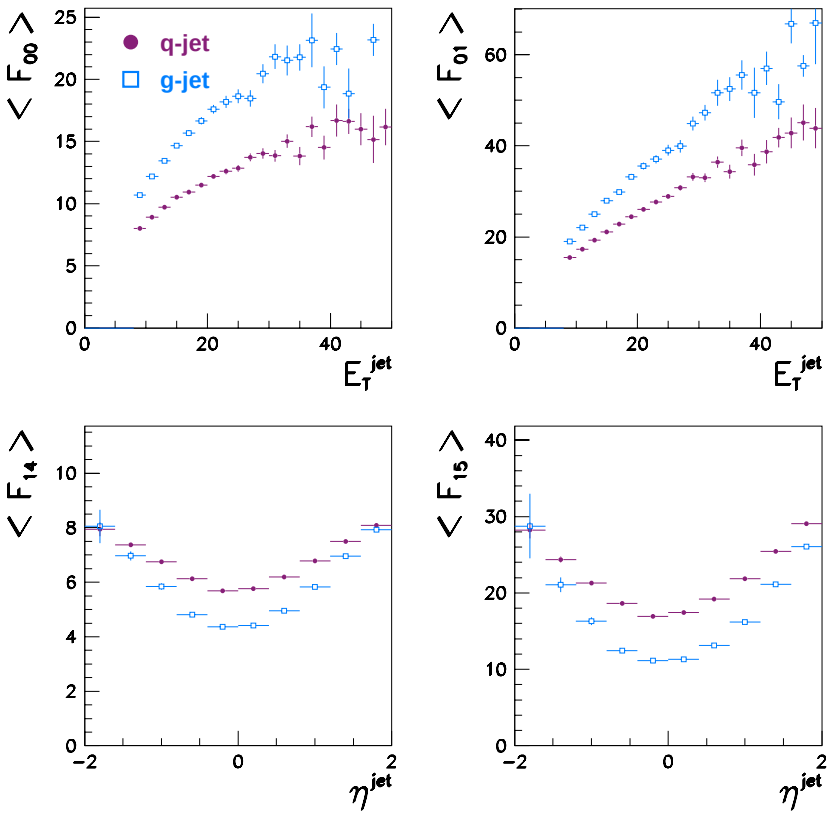

Figure 9.12: *Evolution with η^{jet} of the mean of some Fodor moments for quark and gluon jets at hadron level.*

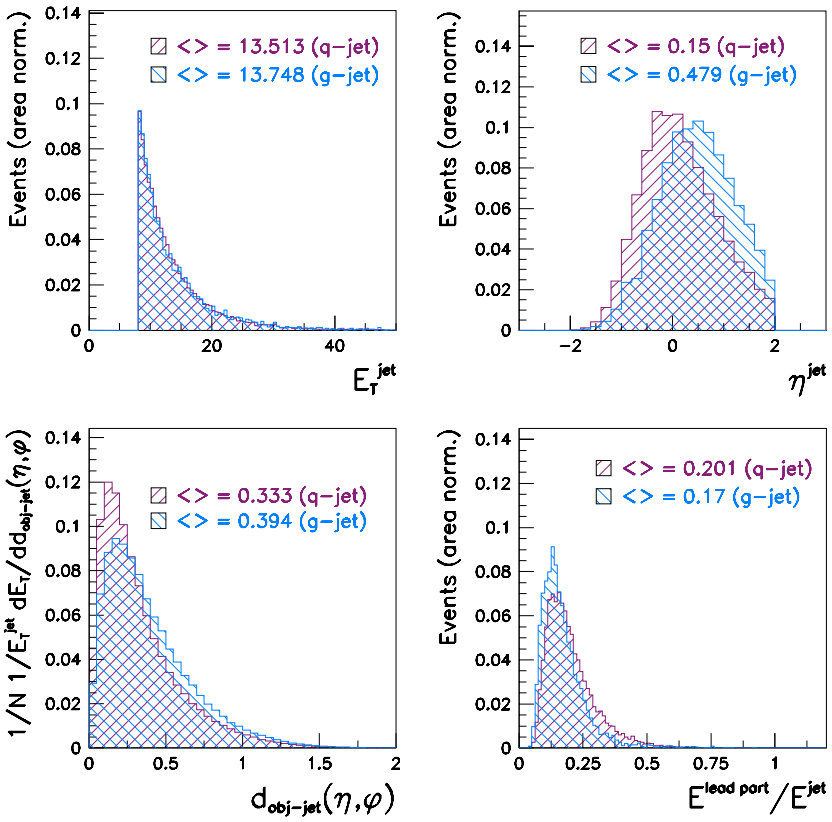

Figure 9.13: *Some discriminating variables investigated at detector level: quark and gluon jet distributions of E_T^{jet}, η^{jet}, E_T differential distribution in the $\eta - \phi$ distance of particles within a jet, and distribution of the fractional energy of the leading particle in the jet ($E^{lead\,part}/E^{jet}$). All variables were computed in the Breit frame.*

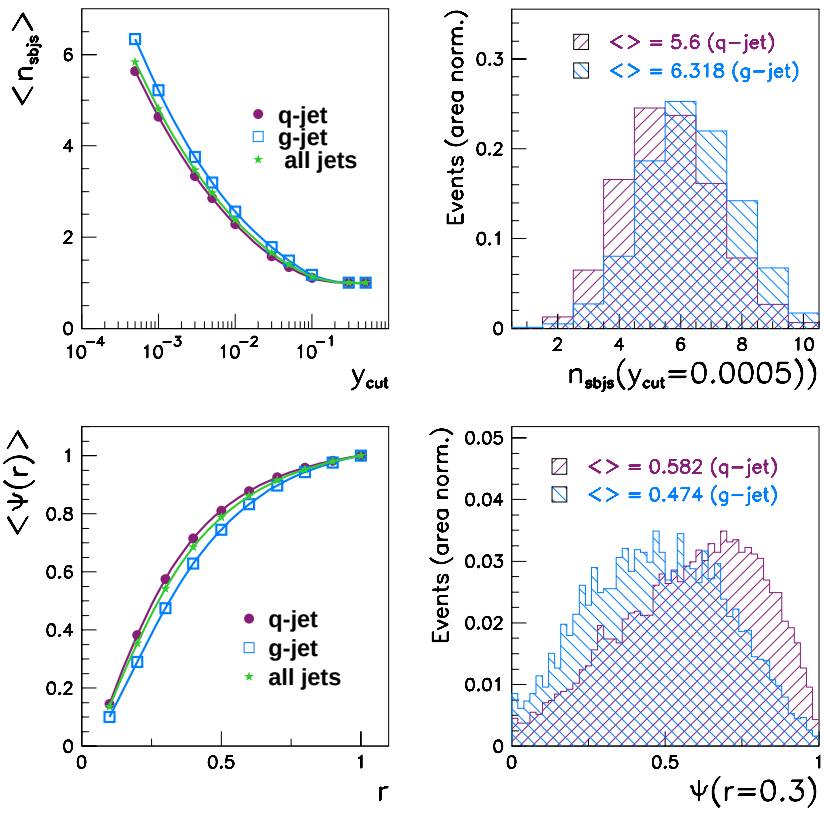

Figure 9.14: *The subjet multiplicity n_{sbjs} and integrated jet shape $\Psi(r)$ for quark and gluon jets at detector level: means and differential distributions for a specific value of y_{cut} and r, respectively.*

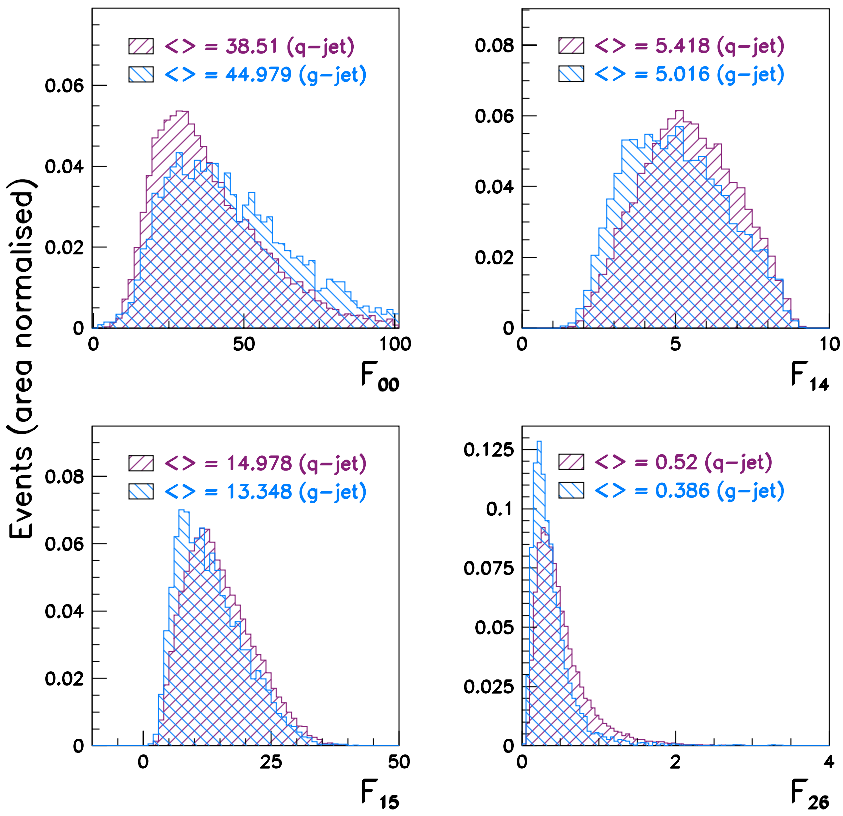

Figure 9.15: *Differential distributions of some Fodor moments for quark and gluon jets at detector level.*

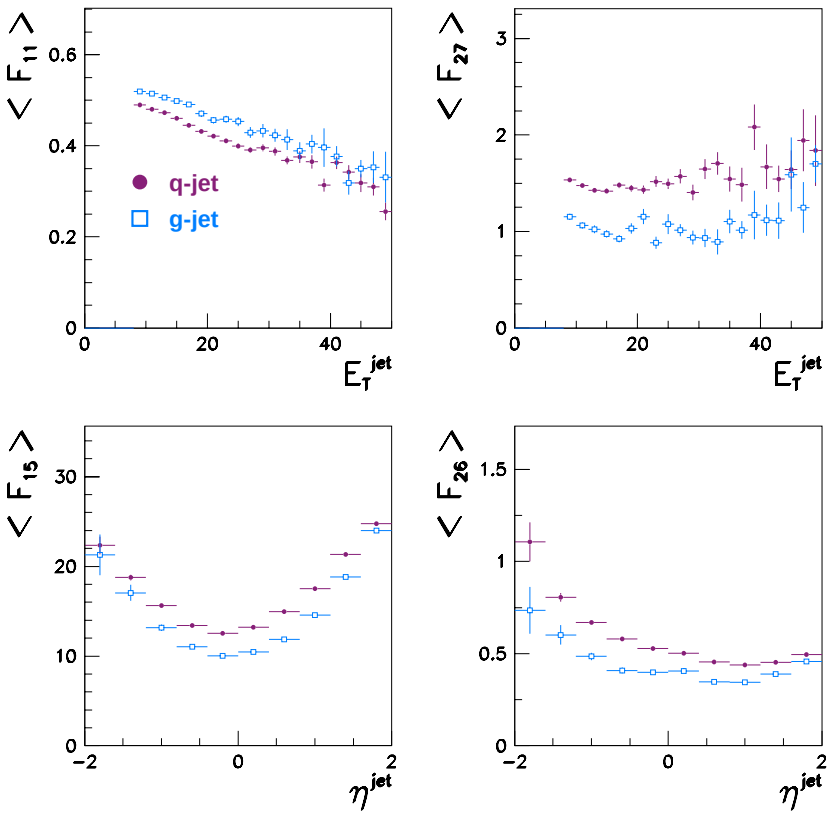

Figure 9.16: *Evolution with η^{jet} of the mean of some Fodor moments for quark and gluon jets at detector level.*

As expected, a small number of discriminating variables was unable to provide a satisfactory jet type discrimination. To determine the relative importance of the various input variables we defined the connection strength to an input neuron i as

$$SW_i = \frac{\sum_{j=1}^{N_{hid1}} |\omega_{ij}|}{\sum_{i=1}^{N_{in}} \sum_{j=1}^{N_{hid1}} |\omega_{ij}|} \quad , \tag{9.15}$$

where N_{hid1} is the number of neurons in the first hidden layer and the ω_{ij} is the weight of the connection of input neuron i to neuron j of the first hidden layer.

Given the "true" quark and gluon jets obtained as explained in section 9.4.1, we defined the purity of the NN tagging as

$$\pi_{q(g)}^{NN} = \frac{\#\ q(g) \text{ jets tagged as } q(g) \text{ by the NN}}{\#\ \text{jets tagged as } q(g) \text{ by the NN}} \quad . \tag{9.16}$$

The efficiency of the NN tagging is defined as

$$\epsilon_{q(g)}^{NN} = \frac{\#\ q(g) \text{ jets tagged as } q(g) \text{ by the NN}}{\#\ q(g) \text{ jets}} \quad . \tag{9.17}$$

9.5.1 Neural Network for Hadron Level Jets

A large number of possible NN topologies and numbers and combinations of input variables was studied. The NN was trained to output 1 (0) for quark (gluon) jets. We opted for the NN described below; other combinations were obtained that resulted in similar or worse results.

All the histograms and results presented for the hadron level jets in this paragraph were obtained with a MLP with 9, 4, 2 and 1 neurons (9421 topology) in the input, first hidden, second hidden and output layers, respectively. The following discriminating variables compose the NN: η^{jet}, $\Psi(r = 0.3)$, $\Psi(r = 0.4)$, $n_{sbjs}(y_{cut} = 0.0005)$, $n_{sbjs}(y_{cut}{=}0.001)$, F_{00}, F_{01}, F_{14} and F_{15}.

Figure 9.17 displays the MLP error function (defined in equation 9.4) obtained during the training of the NN as a function of the number of epochs; an epoch is defined as a run over all patterns in a sample. Two curves are shown: the "learning curve" and the "test curve". The most interesting curve is the test curve: during the training, a

second sample was used to cross-check (test) the performance of the NN. This standard procedure is most useful in avoiding the phenomenon of overfitting during the learning phase, which corresponds to the situation where the test curve flattens off and starts to increase, showing the NN is in reality "unlearning".

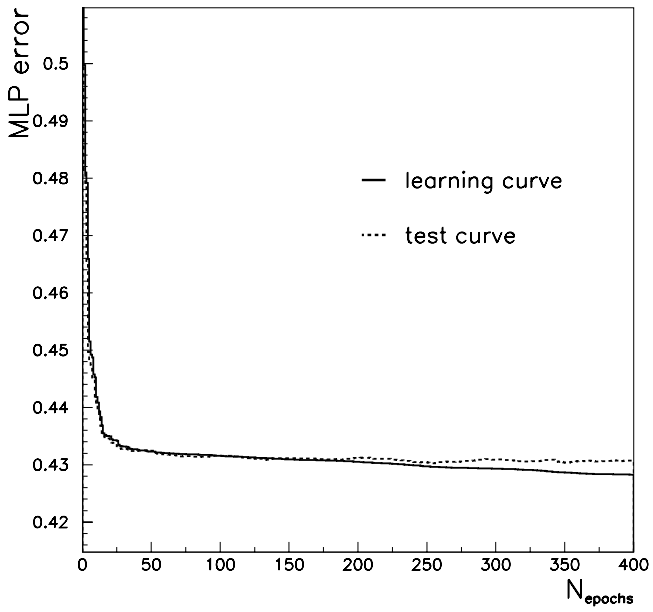

Figure 9.17: *The evolution of the MLP error function during the NN training of hadron level jets as a function of the number of epochs.*

At the end of the training, the discrimination power of the NN can be visualised in a histogram of the NN output for quark and gluon jets. Figure 9.18 shows such an histogram for the entire sample used in the training (denoted "all sample" in the y-axis of figure 9.18) and a sample of equal numbers of each flavour of jets (denoted "$50/50\%$ q/g jets"). As expected, the distributions for quark- and gluon-tagged jets peak towards 1 and 0, respectively; each of the distributions has a large tail towards the other "flavour region", showing the discrimination is not always straightforward.

The connection strengths SW_i defined in equation 9.15, corresponding to the various input variables, helped identifying the "best" discriminating variables for this NN topology: the most important variables were found to be η^{jet} ($SW \approx 19.6\%$), F_{14} ($SW \approx 15.4\%$) and the integrated jet shape $\Psi(r=0.3)$ ($SW \approx 12.6\%$).

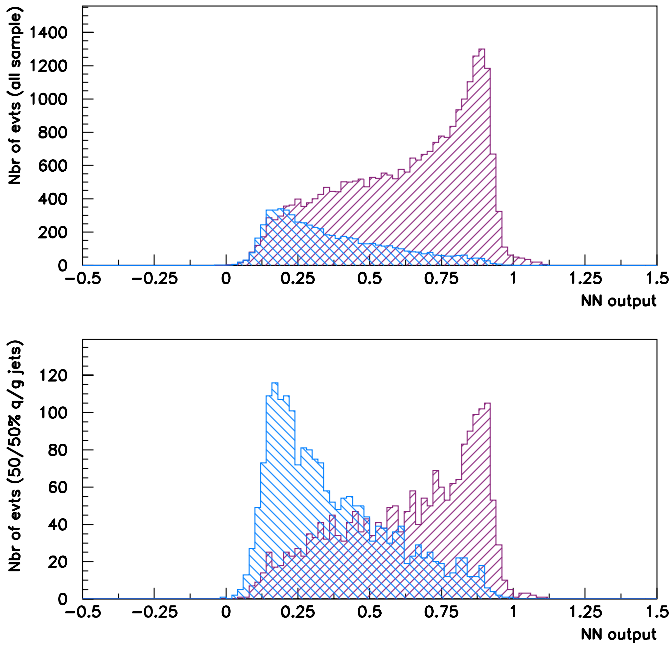

Figure 9.18: *NN output for hadron level quark and gluon jets at the end of the NN training.*

The output function of the trained NN was then used in an unpreviously seen MC sample[2]. In fact, we applied the NN to 2 different samples: the "correct mixture" sample refers to the (standard) sample with the proportion of QCDC and BGF events as obtained in our kinematic range with our jet cuts, *i.e.*, it corresponds to the total Lepto MC sample. Another sample, denoted "$50/50\%$ mixture", was used to assess the performance of the network when confronted with a sample of equal number of quark and gluon jets; in practice we used the MC sample of exclusively QCDC events to simulate the latter situation.

Figure 9.19 displays the NN output obtained with both these samples for the hadron level quark and gluon jets. The performance of a NN is assessed with a sample of equal number of each type of object to discriminate – the jet flavour in our "$50/50\%$ mixture" sample: it can be seen that the NN performs rather well.

The "correct mixture" sample is useful to exemplify the problem that will be encountered when using the NN with data. Due to the kinematic region and jet selection criteria

[2]As stated earlier in the chapter, we also added to the sample the one used to train the NN, for statistics purposes. This was checked to have an insignificant impact on the results.

used, MC studies show that the selected jets in our analysis come from a rather similar proportion of QCDC and BGF events. Hence we expect a ratio of quark:gluon jets $\approx$ 3:1 on the basis of the jet content in QCDC and BGF processes. In other words, the gluon jet identification will always be more challenging given the smaller relative number of gluon jets in the sample.

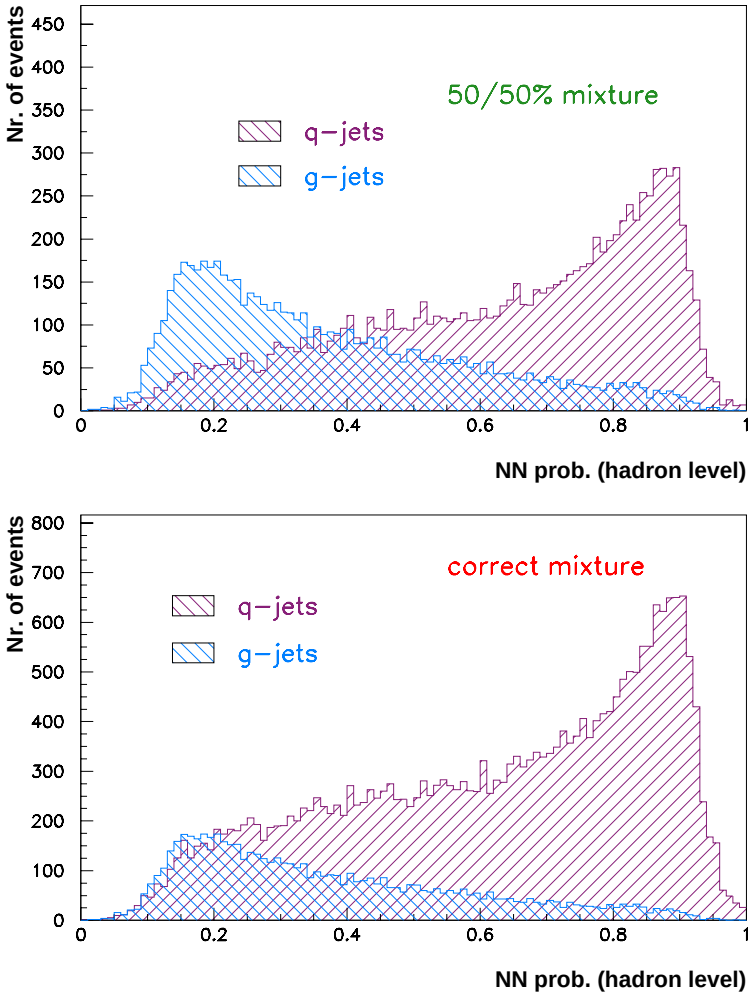

Figure 9.19: *NN output for hadron level quark and gluon jets when applied to the "correct mixture" and the "50/50% mixture" MC samples (refer to the text for details).*

The evolution of the quark- and gluon-jet purities and efficiencies (equations 9.16 and 9.17, respectively) with the cut on the NN output was also investigated. As can

be seen from figure 9.20, the expected evolution is observed: the efficiency for quark (gluon) jet identification decreases (increases) with an increasing cut on the NN output.

The tagging purities also follow an expected pattern. The quark-tagging purity is smaller in the "$50/50\%$ mixture" sample compared to the "correct mixture" sample due to the different ratio of quark to gluon jets in the samples. Conversely, the gluon-tagging purity is smaller in the "correct mixture" sample as a consequence of the dominance of quark jets in the sample.

It is important to notice that the purity of gluon-jet tagging is slightly misleading in the "correct mixture" sample: in such a case, it is clear that most of the quark jets misidentified as gluon jets, which decrease the purity, are coming from BGF events, that contain a pair of quark jets.

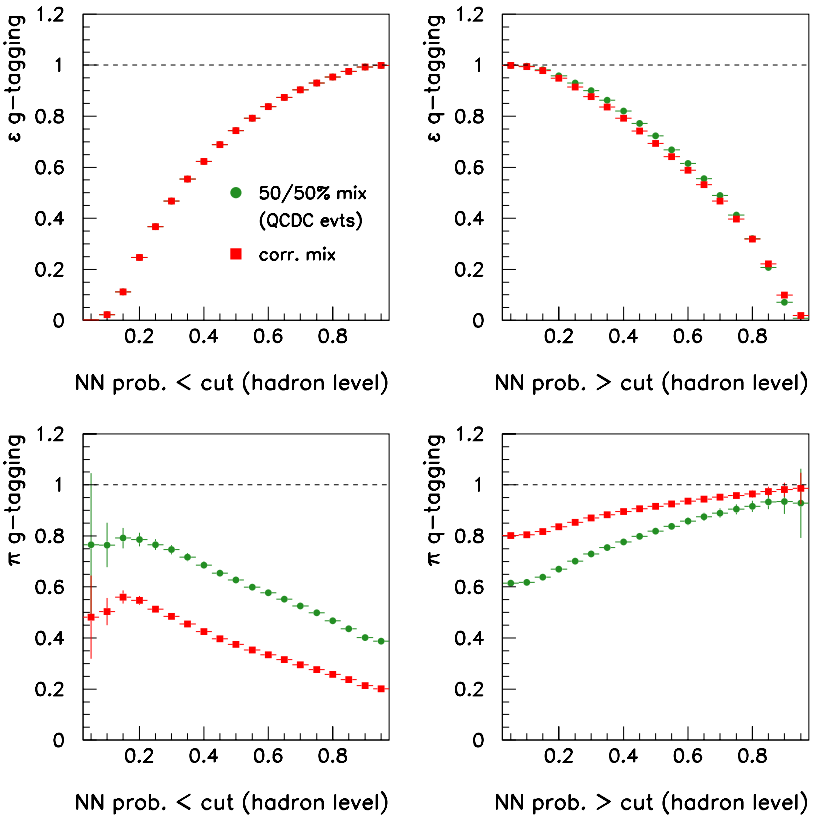

Figure 9.20: *Purities and efficiencies at hadron level for the NN quark- and gluon-jet tagging as a function of the cut on the output of the NN.*

9.5.2 Neural Network for Detector Level Jets

An identical procedure was followed to optimise and then train and use a NN for detector level jets. Again several combinations of NN topologies and input discriminating variables were tested. A MLP with a 9721 topology and the discriminating variables η^{jet}, $\Psi(r=0.3)$, $\Psi(r=0.4)$, $n_{sbjs}(y_{cut}=0.0005)$, $n_{sbjs}(y_{cut}=0.001)$, F_{11}, F_{15}, F_{26} and F_{27} was found to perform best.

For this NN the best discriminating variables were found to be η^{jet} $(SW \approx 23.7\%)$, the integrated jet shapes $\Psi(r=0.3)$ and $\Psi(r=0.4)$ $(SW \approx 12.2\%)$, and F_{15} $(SW \approx 11.4\%)$.

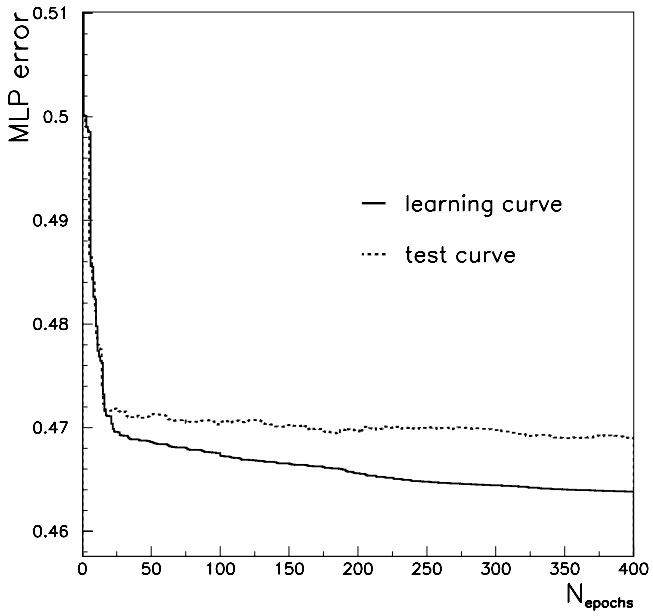

Figure 9.21: *The evolution of the MLP error function during the NN training of detector level jets as a function of the number of epochs.*

As expected from the distributions of the tagging variables presented earlier in this chapter, the discrimination of quark- from gluon-initiated jets is rather more evolved at detector level. Comparing figures 9.17 and 9.21, one notices that the MLP error is larger for detector level jets than it was previously for hadron level jets.

A rather more difficult jet flavour discrimination can also be inferred from figure 9.22.

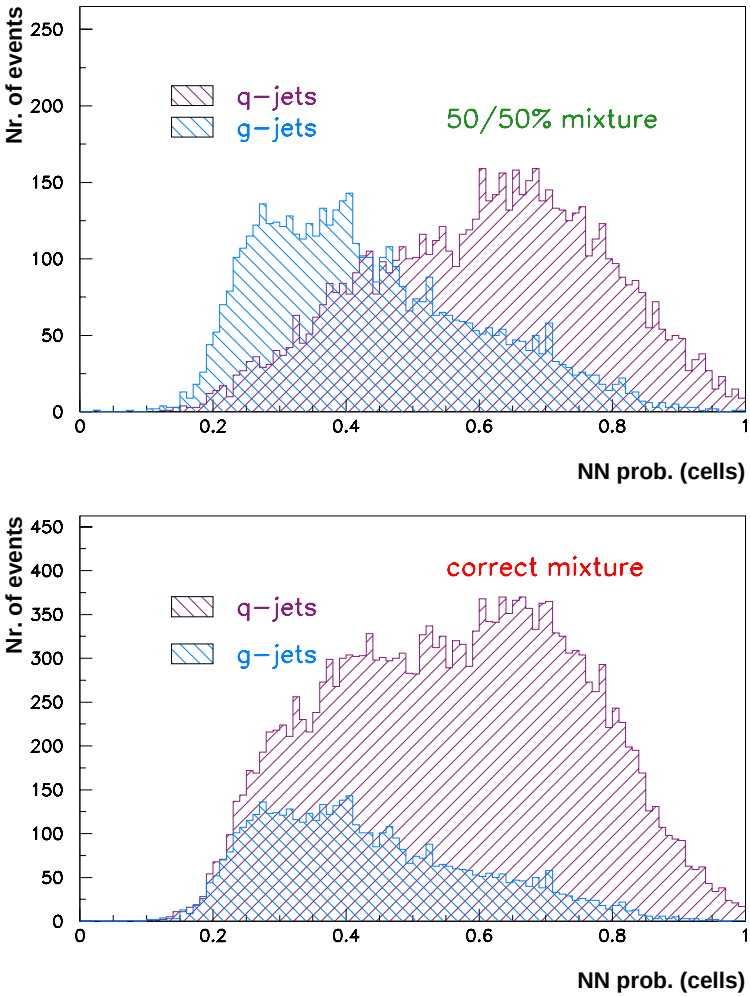

Figure 9.22: *NN output for detector level quark and gluon jets when applied to the "correct mixture" and the "50/50% mixture" MC samples (refer to the text for details).*

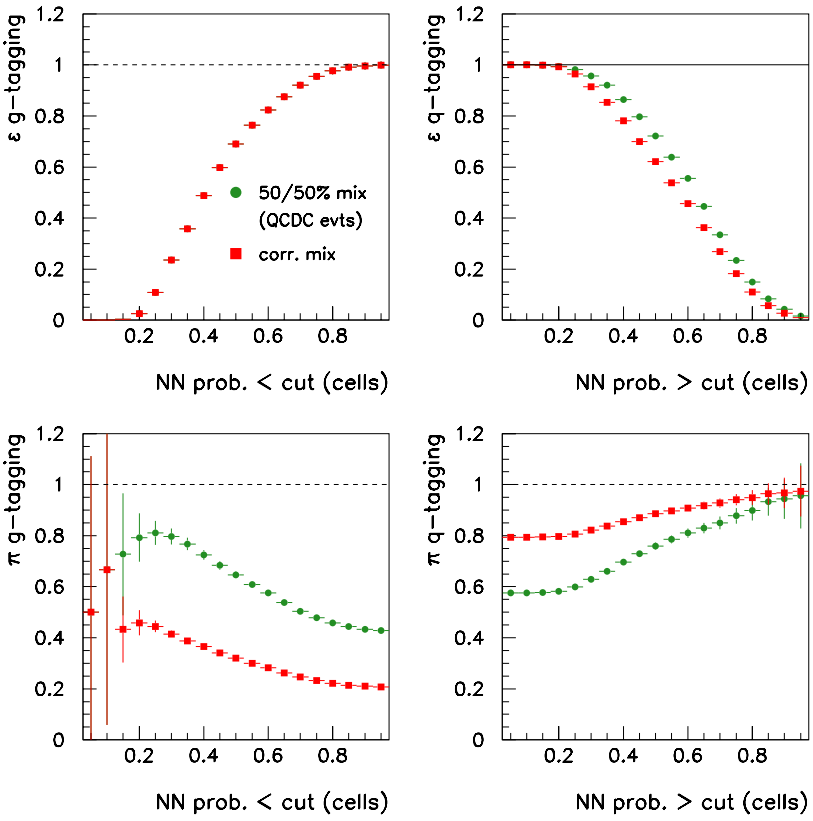

Figure 9.23: *Purities and efficiencies at detector level for the NN quark- and gluon-jet tagging as a function of the cut on the output of the NN.*

The gluon jet identification, in particular, is not clear, even in the "$50/50\%$ mixture" sample: the NN output does not give an output smaller than ≈ 0.2.

The tagging purities and efficiencies for detector level jets are collected in figure 9.23, which shows the evolution of the latter with the cut on the NN output. Again the NN is seen to perform worse than with hadron level jets; in particular for gluon jets.

9.6 QCDC versus BGF Processes

As stated in the introduction of this chapter, it was our original aim to study in detail the azimuthal asymmetries in separated samples of quark- and gluon-initiated jets in QCDC and BGF multi-jet events. Having developed a method to discriminate the flavour of the jets one still needs to tag the QCD event type – QCDC or BGF.

All the MC studies of the jet analyses have been performed with Lepto. The main reason of this choice follows. For the purposes of this analysis, we need a MC that generates QCDC and BGF processes according to the QCD matrix elements, and only Lepto generates events with the correct mixture; Ariadne includes the BGF process in a rather *ad hoc* approach, and does not seem to be an appropriate choice for these investigations.

9.6.1 Tagging Approach

In our kinematic region of interest a separation of QCDC from BGF events is not possible based on the kinematic variables of the final state partons; they are rather similar, on average. We hereafter describe our tagging approach.

Our approach assumes a good correspondence between the jets and the QCD type of the event; it is based on the idea that the flavours of all the jets in a given event should relate, to some extent, to the type of the first order QCD process of the generated event. It is our ansatz, but it does seem to be the most straightforward and logical approach. In fact, we investigated this issue in a ZEUS note [72], and the study supports this approach if and only if the jet flavour tagging is sufficiently accurate.

The procedure will be the following: the selected jets in a multi-jet event (n-jet event, $n \geq 2$) are tagged with the NN. The QCD event type is then tagged according to the type of the 2 highest E_T jets:

- event tagged QCDC if the 2 highest E_T jets are a quark- and a gluon-jet;
- event tagged BGF if the 2 highest E_T jets are quark-jets.

Note again that only events with at least 2 selected jets were considered in this attempt to separate the azimuthal angle distributions of quark- and gluon-initiated jets in QCDC and BGF multi-jet events. We motivated in the last chapter the exclusion of 1-jet events from an analysis of jet azimuthal asymmetries. In the context of this QCD event type tagging, the method requires at least 2 selected jets in the event.

The performance of this algorithm will be quantified by means of the efficiency of the event type tagging,

$$\epsilon^i_{\text{tag}} = \frac{\#\ i \text{ events tagged as } i}{\#\ i \text{ events}} \quad , \tag{9.18}$$

and the purity of the tagged samples,

$$\pi^i_{\text{tag}} = \frac{\#\ i \text{ events tagged as } i}{\#\ \text{events tagged as } i} \quad , \tag{9.19}$$

i being a QCDC or BGF multi-jet event.

In fact, the efficiency of the QCD event type tagging will dictate the amount of statistics that will be selected for the analysis. As expected, a good balance between efficiency and purity needs to be found.

Using our "correct mixture" sample we obtained the results summarised in table 9.1. Again the hadron level NN is seen to perform better that the detector level NN. The tagging efficiencies at hadron (detector) level are of the order 20% $(3-9\%)$, whereas the purities are in the range $60-80\%$ $(55-60\%)$. In particular, the purities at detector level turn out to be relatively low.

Table 9.1 shows the results of the event type tagging based on the NN jet flavour tagging. Given our original aim of separating the flavour of the jets and the QCD event

Performance of the QCD event type tagging based on the NN tagging		
	NN (hadron level)	NN (detector level)
$\pi_{\rm tag}^{QCDC}$	62.4	55.1
$\pi_{\rm tag}^{BGF}$	79.8	60.6
$\epsilon_{\rm tag}^{QCDC}$	23.9	9.3
$\epsilon_{\rm tag}^{BGF}$	18.0	3.0

Table 9.1: *Results on the performance of the NN at hadron and detector levels in multi-jet events. The definitions of the efficiencies and purities $\pi_{\rm tag}^{i}$ and $\epsilon_{\rm tag}^{i}$ (i =QCDC, BGF) are given in 9.18 and 9.19, respectively.*

type in data, it is also interesting to determine how the trained NN responds "as a whole" for the 2 types of events under consideration. We therefore investigated the tagging of the NN when faced with (true) QCDC and BGF multi-jet events. Table 9.2 presents the outcome at hadron level. On average, the jet flavour is rather well determined for QCDC events, with the 2 jets being identified as a quark and a gluon jets. But the NN also tends to identify BGF events as being formed of a quark and a gluon jet rather than a pair of quark-initiated jets.

q-/g-jet type identification at hadron level in true QCDC and BGF 2-jet events		
		NN
QCDC	% gg−jets	16.0
	% qg−jets	70.2
	% qq−jets	13.8
BGF	% gg−jets	7.0
	% qg−jets	38.6
	% qq−jets	54.4

Table 9.2: *NN quark and gluon jet assignment in true QCDC and BGF 2-jet events at hadron level.*

The degradation of the discriminating power of the NN at detector level is striking when one compares the results summarised in tables 9.2 and 9.3. The jet content of QCDC events is still rather well determined, but the quark-jet misidentification in BGF

q-/g-jet type identification at detector level in true QCDC and BGF 2-jet events		
		NN
QCDC	% gg–jets	11.7
	% qg–jets	72.3
	% qq–jets	16.0
BGF	% gg–jets	8.5
	% qg–jets	64.5
	% qq–jets	27.0

Table 9.3: *NN quark and gluon jet assignment in true QCDC and BGF 2-jet events at detector level.*

events is dominating: at detector level, most of the BGF events ($\approx 65\%$) are tagged as a pair of quark-gluon jets, the signature of a QCDC event.

These 2 tables, tables 9.2 and 9.3, explain rather well why the results for the event type tagging purities presented in table 9.1 are significantly lower at detector compared to hadron level.

Given the performance obtained, we are forced to conclude that our approach cannot at present be applied with confidence: the NN jet flavour tagging is not performant enough for the event type tagging to be reasonably correct. Still, we discuss in the following section the azimuthal angle distributions obtained with this tagging procedure, to assess most clearly the influence of the mistagging on the distributions. In section 9.8 of this chapter we will present an alternative NN at detector level, which focusses solely on QCDC events.

9.7 Tagged MC Distributions

Using the trained NNs to tag the jets flavours and the event type tagging procedure described above we plotted the distributions separately for quark and gluon jets in QCDC and BGF events. Figure 9.24 shows the ϕ^{jet} distributions obtained at hadron level with the approach described in this chapter for both the MC samples used. Jets were tagged at both levels as quark-initiated (gluon-initiated) if the NN output was > 0.7 (< 0.3).

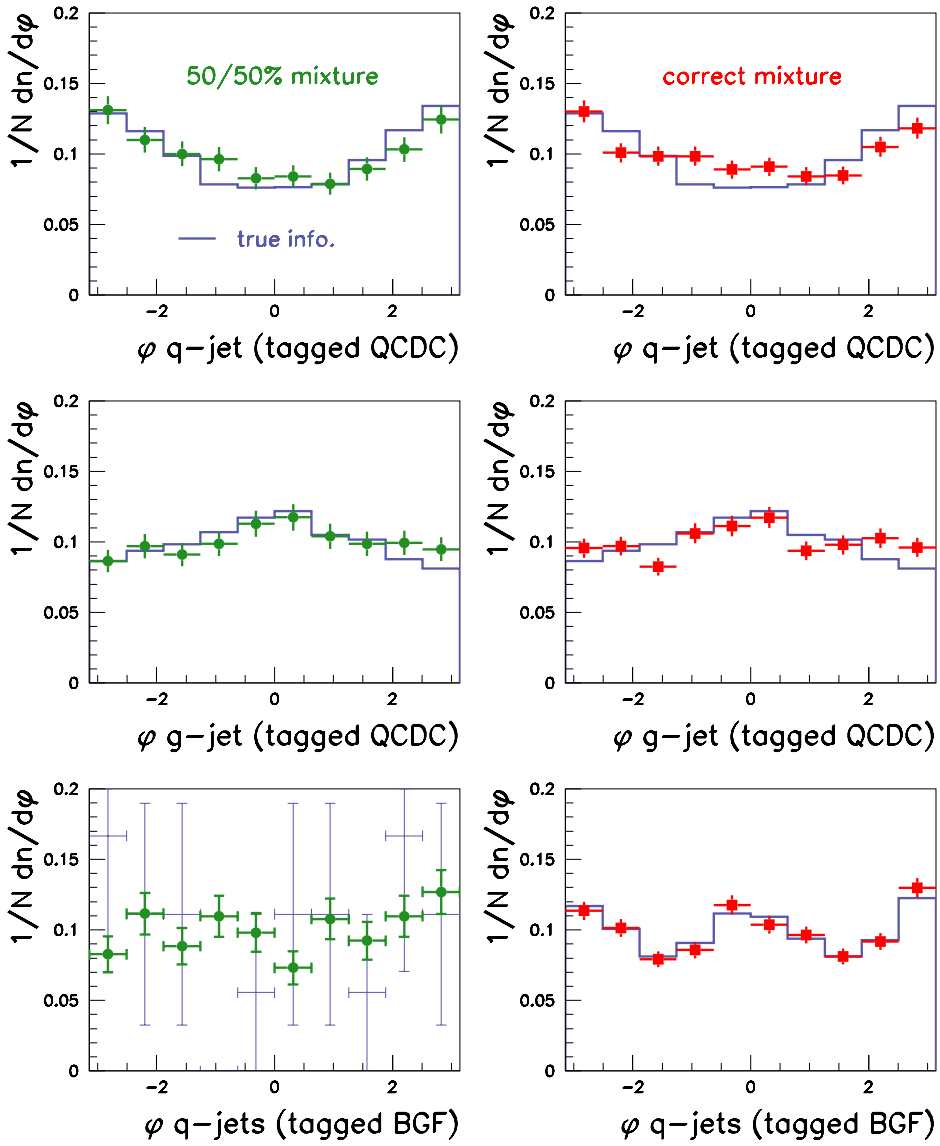

Figure 9.24: *Azimuthal asymmetries at hadron level for the tagged quark- and gluon-initiated jets in tagged QCDC and BGF multi-jet events (refer to the text for details). The solid (blue) histogram refers to the distributions obtained with the "true" MC information as explained in section 9.4.1.*

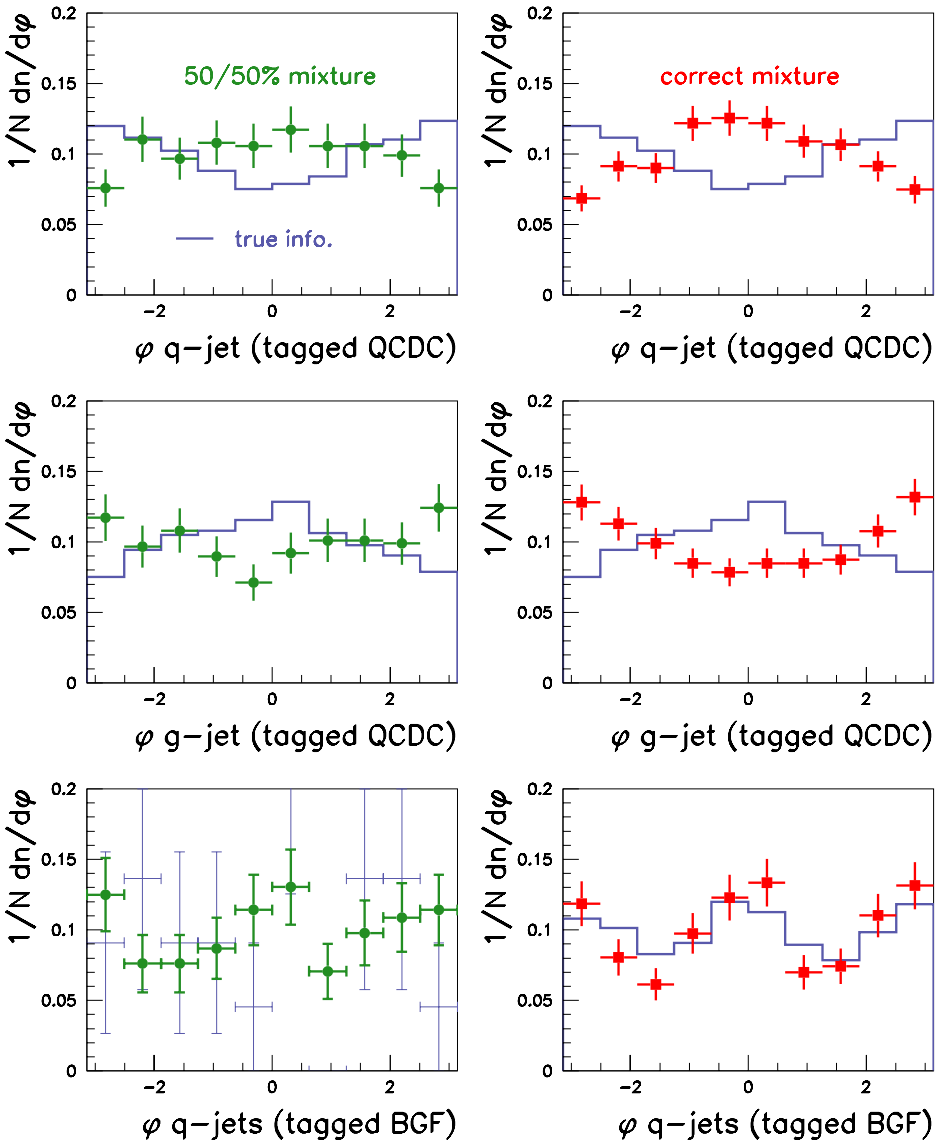

Figure 9.25: *Azimuthal asymmetries at detector level for the tagged quark- and gluon-initiated jets in tagged QCDC and BGF multi-jet events (refer to the text for details). The solid (blue) histogram refers to the distributions obtained with the "true" MC information as explained in section 9.4.1.*

The hadron level NN performs rather well in general, and is able to disentangle the jet flavour and consequently identify the QCDC event type with a reasonable efficiency and purity. For the "$50/50\%$ mixture" sample, the purities for quark and gluon jet tagging are rather high, approximately 89% and 75%, respectively. Both the quark and gluon ϕ^{jet} distributions in QCDC events are seen to be in very good agreement with the distributions obtained with the true MC information. And the ϕ^{jet} distribution of quark jets in BGF events (points with inner and thick error bars) is seen to be flat, as expected; it reflects those true QCDC events misidentified as being composed of 2 quark jets, and hence tagged as BGF in our procedure. Note that this same reason explains the rather large error bars (thin bars) in the BGF distribution of true quarks: the statistics are very low since the number of events mistagged as BGF, containing 2 "true" quark jets, is very low (see section 9.4.1 for the explanation of how our sample of "true" quark and gluon jets was obtained). The performance of the approach naturally degrades when applying it to the "correct mixture" sample. The purities for quark and gluon jet tagging are now 95% and 48.5%, respectively. Note that π_q^{NN} is now higher simply due to the higher ratio of quark to gluon jets in the sample. Here the BGF tagged distribution agrees very well with the true one, since the contribution from misidentified QCDC events is flat (the ϕ^{jet} distributions of quark and gluon jets cancel out in QCDC events). The contribution from misidentified BGF quark jets is clearly seen in both the quark and gluon ϕ^{jet} distributions.

Similarly, figure 9.25 presents the same distributions obtained with the NN trained on and for detector level jets. Again, jets were tagged as quark-initiated (gluon-initiated) if the NN output was > 0.7 (< 0.3). For the "$50/50\%$ mixture" sample, the purities for quark and gluon tagging are rather high, approximately 85% and 80%, respectively. But the ϕ^{jet} distributions are surprisingly wrong in shape. This last fact required further investigation, and can be understood as follows. For this "$50/50\%$ mixture" sample, composed uniquely of QCDC events, the true quark jets present a ϕ^{jet} distribution with a negative $\cos\phi^{jet}$ shape, with a dip at $\phi^{jet} \approx 0$. We studied the efficiency and purity of the jet flavour tagging as a function of ϕ^{jet}. For the quark jets in this sample we

found that $\epsilon_q^{NN}(\phi^{jet})$ is higher at $\phi^{jet} \approx 0$, whereas $\pi_q^{NN}(\phi^{jet})$ is lower at $\phi^{jet} \approx 0$. Consequently, a quark jet is more likely to be tagged by the NN if $\phi^{jet} \approx 0$, and gluon jets misidentified as quark jets also tend to accumulate around $\phi^{jet} = 0$. Hence the tagged QCDC quark jet distribution will tend to be enhanced at $\phi^{jet} \approx 0$. For the gluon jets the situation is precisely the complementary, fact that is again seen in figure 9.25. These observations also explain the non-flat pattern seen in the BGF ϕ^{jet} distribution. Only events with a misidentified gluon jet contribute to this histogram. The latter is enhanced (as observed) at $\phi^{jet} \approx \pm\pi$, since $\pi_g^{NN}(\phi^{jet})$ is lower at $\phi^{jet} \approx \pm\pi$; and the other jet in the event – the quark jet – will be separated by $\approx \pi$, *i.e.*, in the region around $\phi^{jet} = 0$. Hence the BGF shape observed. It is important to remark that the corresponding purities and efficiencies distributions $\pi_q^{NN}(\phi^{jet})$ and $\epsilon_{q(g)}^{NN}(\phi^{jet})$ were found to be very flat for the NN used for the hadron level jets; only $\pi_g^{NN}(\phi^{jet})$ presented a shape with lower purities in the region $\phi^{jet} \approx \pm\pi$. For the "correct mixture" sample the purities for quark and gluon jet tagging are approximately 93% and 41%, respectively. The tagging performance degrades as in the case with hadron level jets.

9.8 "QCDC" NN for Detector Level Jets

From the previous results and discussions it is clear that the most challenging task is the identification of the gluon jets. Also we are mainly interested in the experimental determination of the ϕ^{jet} distributions in QCDC events, since the jet distributions in BGF events can be added together (from QCD) and, in fact, constitute the only contribution to the "inclusive" distribution we presented in the previous chapter.

Here we present an alternative NN, focussing solely on QCDC events. We repeated the whole procedure of training of hadron and detector level jets with a sample of pure QCDC events. And used the (new) NN output functions both in the mixed sample and the QCDC sample, as before, using the event type tagging approach described in section 9.6.1.

The performance obtained at hadron level with the previous NN is good that we do

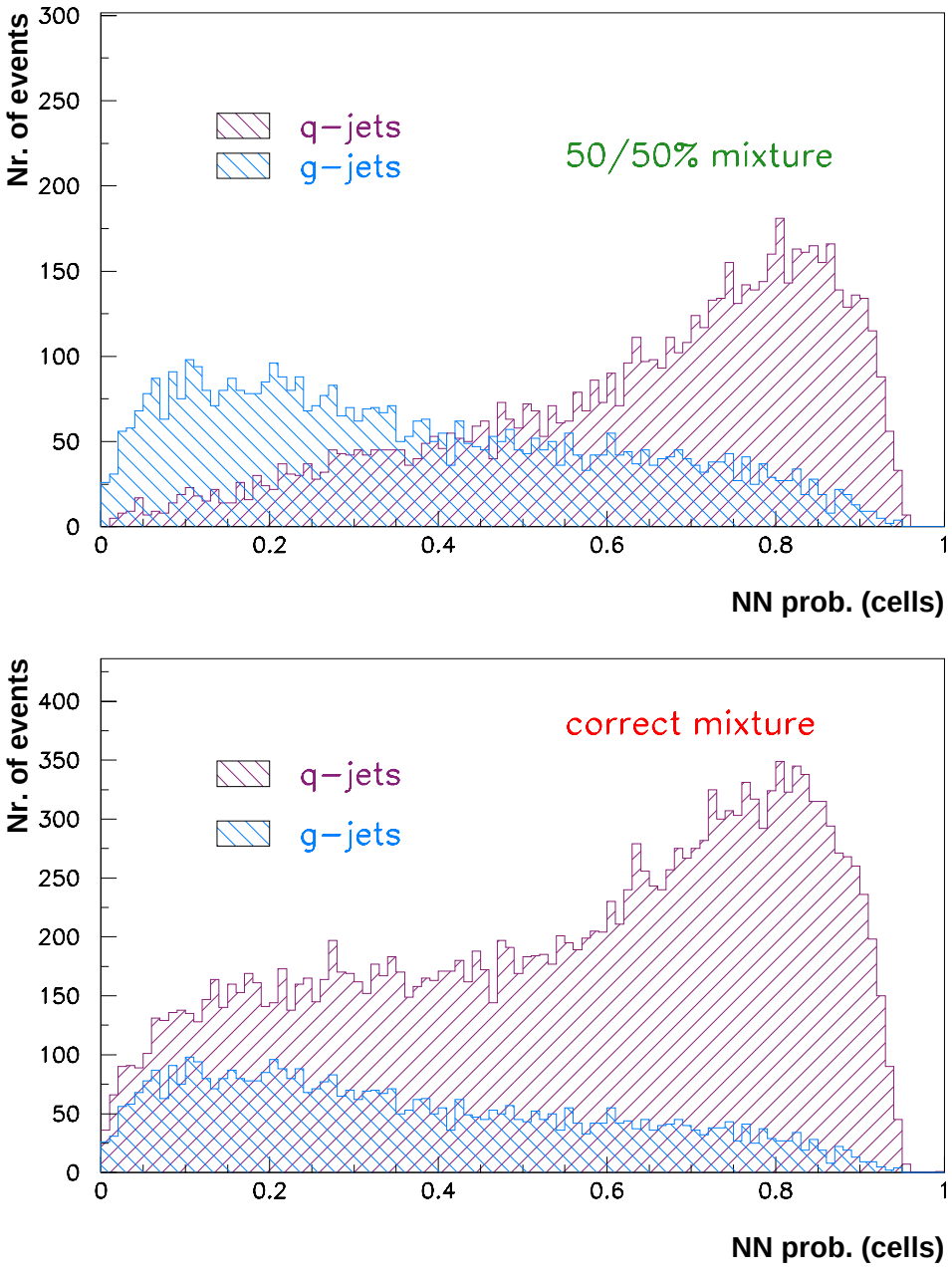

Figure 9.26: *"QCDC" NN output for detector level quark and gluon jets when applied to the "correct mixture" and the "50/50% mixture" MC samples (refer to the text for details).*

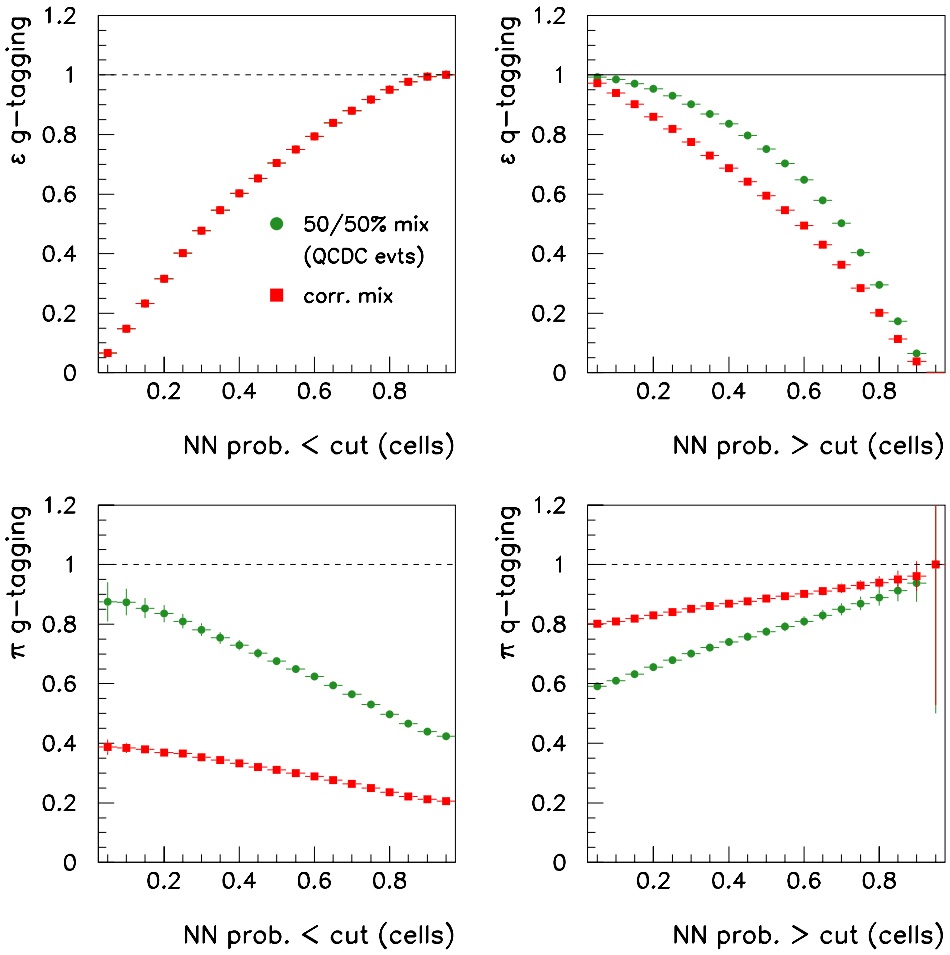

Figure 9.27: *Purities and efficiencies at detector level for the "QCDC" NN quark- and gluon-jet tagging as a function of the cut on the output of the NN.*

not present here the results obtained with the "QCDC" NN for hadron level jets. We are mainly interested in the detector level studies.

This "QCDC" NN was trained on a sample of only QCDC events containing detector level jets. A MLP topology 9721 with the discriminating variables η^{jet}, $\Psi(r = 0.3)$, $\Psi(r = 0.4)$, $n_{sbjs}(y_{cut} = 0.0005)$, $n_{sbjs}(y_{cut} = 0.001)$, F_{00}, F_{11}, F_{14}, and F_{27} – a set rather similar to the one used before at detector level – was found to perform best. The "distribution" of the connection weights SW_i is rather flat, although $\Psi(r = 0.3)$, F_{14} and F_{27} seem to get more weight in the NN.

At the end of the training procedure the MLP error function of the test curve was

Performance of the QCD event type tagging based on the "QCDC" NN tagging	
$\pi_{\rm tag}^{QCDC}$	53.3
$\pi_{\rm tag}^{BGF}$	50.2
$\epsilon_{\rm tag}^{QCDC}$	28.5
$\epsilon_{\rm tag}^{BGF}$	4.1

Table 9.4: *Results on the performance of the "QCDC" NN at detector level in multi-jet events. The definitions of the efficiencies and purities* $\pi_{\rm tag}^{i}$ *and* $\epsilon_{\rm tag}^{i}$ *(*i *=QCDC, BGF) are given in 9.18 and 9.19, respectively.*

found to approach a value around 0.43, almost as low as the value obtained previously with hadron level jets (*cf.* figure 9.17).

Figure 9.26 shows the NN output obtained with the "$50/50\%$ mixture" and the "correct mixture" samples. The shape of the NN output for the quark and gluon jets are now more similar to the ones obtained at hadron level. The same remark holds for figure 9.27, leading to the expectation that this network will perform better.

The performance of the NN can be assessed from tables 9.4 and 9.5. The former table indicates that the purities achieved with this "QCDC" network are similar to the ones obtained previously; but it performes worse than before in tagging BGF events, as expected. Also the efficiency for QCDC tagging increased considerably. Again these purities can be explained on the basis of the numbers presented in table 9.5, which shows how the jet flavours are tagged in true QCDC and BGF events. The jet content of QCDC events is now very well identified: about 77% of the QCDC events are identified as composed of a quark and a gluon jet. For comparison we kept the results obtained for BGF events. But these are of lesser importance, since we focussed all the training and therefore the NN on the discrimination of the jet flavour in QCDC events.

Figure 9.28 shows the ϕ^{jet} distributions, to be compared to the previous distributions (figure 9.25) also obtained at detector level. the improvement is clear. But further improvements on the purities are needed for the whole procedure to be applied with confidence to the data.

q-/g-jet type identification at detector level in true QCDC and BGF 2-jet events using the "QCDC" NN		
		NN
QCDC	% gg–jets	12.3
	% qg–jets	76.5
	% qq–jets	11.2
BGF	% gg–jets	19.6
	% qg–jets	69.0
	% qq–jets	11.4

Table 9.5: *"QCDC" NN quark and gluon jet assignment in true QCDC and BGF 2-jet events at detector level.*

9.9 Final Remarks

We detailed in this chapter several attempts to measure the jet azimuthal asymmetries in separated samples of quark- and gluon-initiated jets in QCDC and BGF events. We tackled this 2-fold problem is 2 steps.

First, we used neural network techniques to identify the flavour of the jets. We then used the jet flavour content, as tagged by the NN, to subsequently tag the QCD event type. As demonstrated, this method is valid provided the jet flavour identification is sufficiently accurate that the event type can be tagged with confidence. Unfortunately, the method performs well but suffers from large background.

As stated before, the main interest is in the quark and gluon jet distributions of QCDC events. At the end we presented an alternative "QCDC" network, which concentrated on QCDC events. It performs very well, but only provided we have a relatively pure sample of QCDC events.

One concludes that these methods can be used if another method is found to tag, or identify, the QCDC event type, in which case 2 independent and complementary pieces of information could be used to reduce considerably the effects of misidentifications.

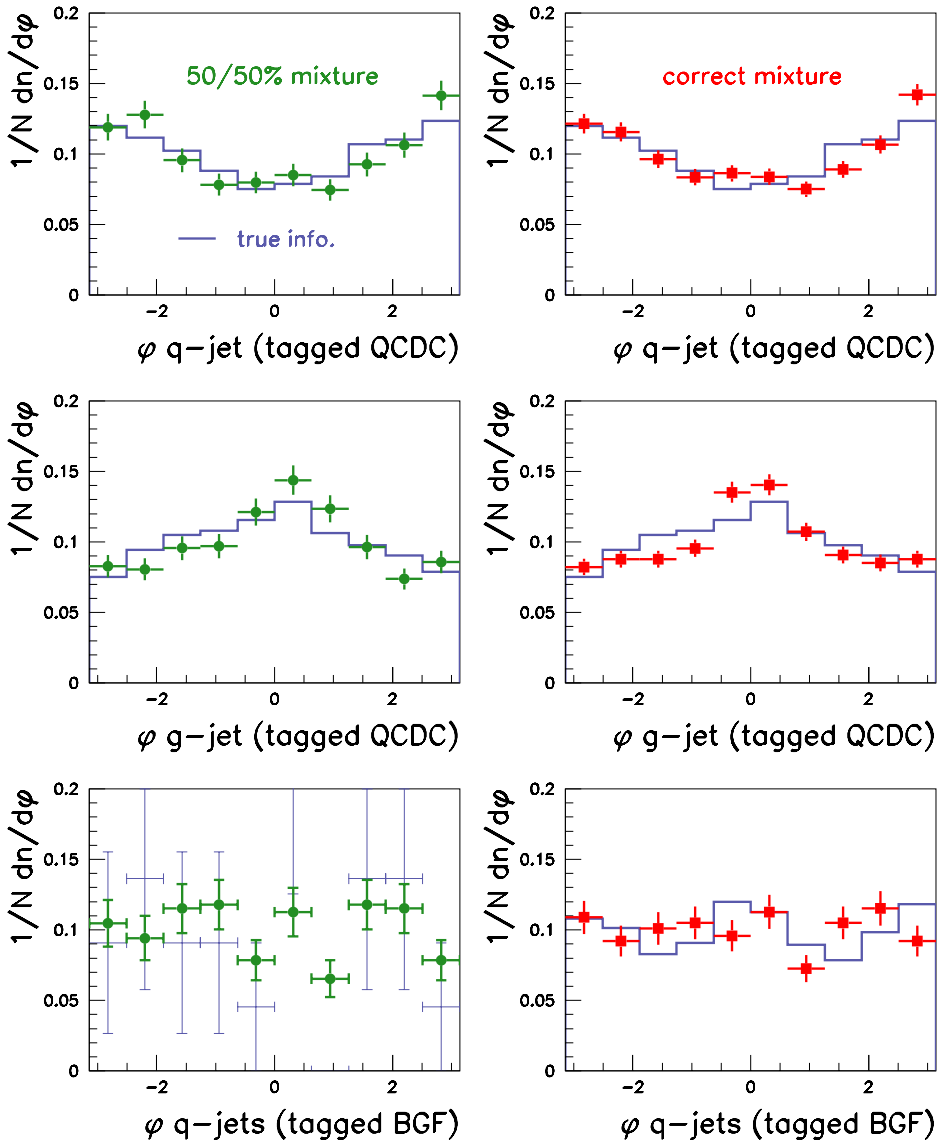

Figure 9.28: *Azimuthal asymmetries obtained with the "QCDC" NN at detector level for the tagged quark- and gluon-initiated jets in tagged QCDC and BGF multi-jet events (refer to the text for details). The solid (blue) histogram refers to the distributions obtained with the "true" MC information as explained in section 9.4.1.*

Chapter 10

Theoretical QCD Calculations

10.1 Outline

In the following chapter the data will be compared to LO and NLO MC programs. The LO programs are simply those used throughout the analyses – Lepto and Ariadne. We also tried to compared the data of both analyses with theoretical calculations. This chapter will focus uniquely on this latter aspect.

We will first present and discuss the QCD predictions based on numerical calculations of the LO QCD analytical expressions for the cross section of semi-inclusive DIS production of charged tracks. We coded the theoretical expressions in a program, AZICAL, which, unlike LO MC programs, models the non-perturbative part of the transverse momentum distributions due to the intrinsic parton k_T and the hadronisation momentum developed in the fragmentation process. These theoretical results will then be compared to the data in the next chapter.

Jet production can be compared to QCD calculations up to NLO with the help of MC programs. We introduce here the NLO program we will use later – DISENT – to compare with the jet azimuthal angle distributions in dijet events.

10.2 LO QCD Calculations for Hadrons

The results presented in this section refer to calculations made with the same conditions as in the analysis of charged tracks (chapter 6), *i. e.* in the same kinematic region and with the same selection cuts on the hadrons. They will be compared to the data in the next chapter. The CTEQ5L set of PDFs has been used unless otherwise stated.

10.2.1 The Program AZICAL

I developed AZICAL, a FORTRAN program that implements numerically all the non-perturbative and perturbative contributions to the SIDIS differential cross section; the explicit expressions are collected in appendix A. It is a flexible LO theoretical program that allows the computation of all relevant contributions. One can selectively calculate the differential azimuthal angle distribution $1/N\ dn/d\phi$ and the $\langle\cos\phi\rangle$ and $\langle\cos 2\phi\rangle$ moments for hadrons with a transverse momentum above a certain threshold (in the HCM frame). The relevant terms of the 5-fold differential cross section are numerically integrated with the help of the CERNLIB library routines [78].

A significant amount of parameters can be set up and defined (refer to appendix A for the definitions of certain parameters):

- the positron and proton beam energies;
- the kinematic range in x, y and Q^2 (consistency checks are performed and an effective range determined);
- the a and b parameters for the primordial parton k_T and fragmentation p_T Gaussian distributions, respectively;
- a range for the z_h of the selected hadrons;
- a range for the P_{hT} cuts on the transverse momenta of the hadrons, as well as the number of points to calculate within the range;
- the number of parton flavours interacting;

- choice of the PDFs (using the PDFLIB package [79]).

The a and b parameters defined in appendix A were introduced by Chay *et al.* [19, 20] to incorporate in the theoretical calculations an estimation of the non-perturbative contributions from the intrinsic k_T and hadronisation p_T, respectively. In their model, a and b parameterise the width of the non-perturbative part of the transverse momentum distributions, approximated by (factorised) Gaussians in the cross section expressions.

10.2.2 Influence of the Parton Distribution Functions

We ran the calculations with different PDFs to investigate the influence of a particular choice; we chose the LO CTEQ5L, CTEQ4L and GRV94LO PDFs for the fact that they have somewhat different gluon distributions and are frequently used to confront the data. The particular choice of the PDF has a negligible effect on the final ϕ distributions and moments. The differences are rather small compared to the numerical errors on the calculations and the experimental errors (as seen in figure 10.3).

10.2.3 Influence of the Transverse Momentum Parameters

The influence of the phenomenological transverse momentum parameters a and b was investigated. We varied these parameters by ± 0.1 around the "central" values of $a = b = 0.6$ GeV. These values were chosen here as reference, motivated by both previous theoretical fits to fixed target data [19,20] and the experimental data from fixed target experiments [6, 8]. The influence on the moments is insignificant compared to our experimental errors, in particular in the case of the $\langle \cos 2\phi \rangle$ moment (figure 10.1). This insensitivity to these non-perturbative transverse momentum parameters in our high energy region is strikingly different from the observations made in fixed target experiments, where, in fact, the measured distributions were fitted and used to empirically determine the average intrinsic parton transverse momentum.

The comparison of the variations of the 3 curves with our experimental errors led to the conclusion that no fine tuning of a and b will be relevant; these moments are not

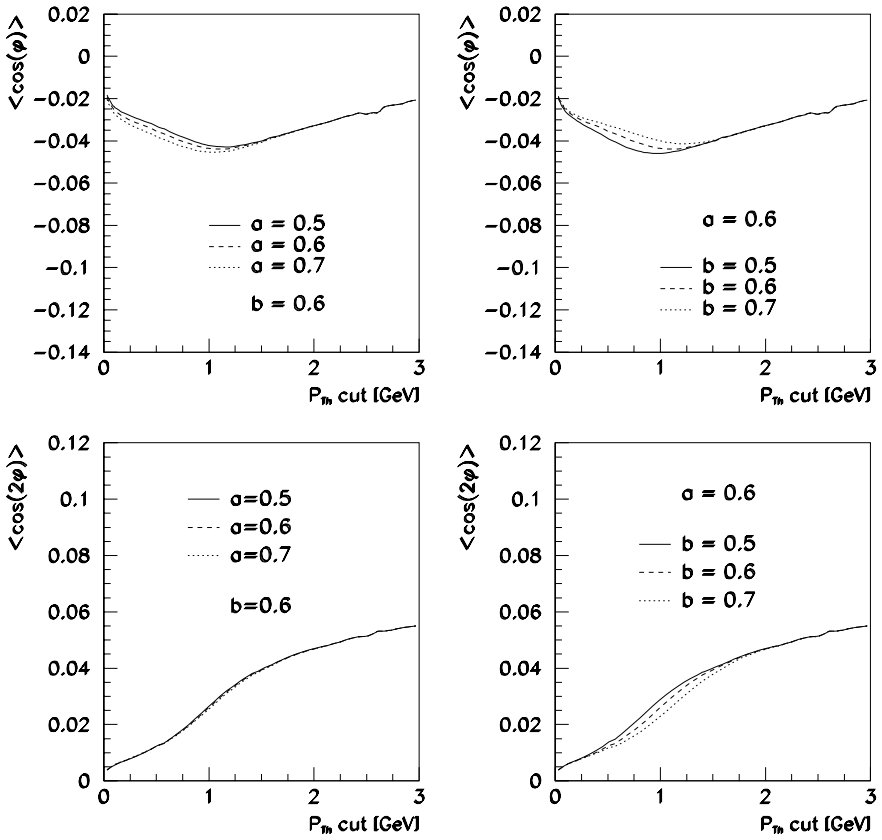

Figure 10.1: *Theoretical calculations obtained with the program AZICAL for the moments* $\langle \cos\phi \rangle$ *and* $\langle \cos 2\phi \rangle$ *as a function of the* P_{Th} *cut on the hadrons. The distributions show the influence of the intrinsic parton* k_T *(a parameter) and the hadronisation* p_T *(b parameter) on the moments (a and b are given in GeV).*

very sensitive to their specific value (varied within a reasonabe range) in the kinematic region investigated.

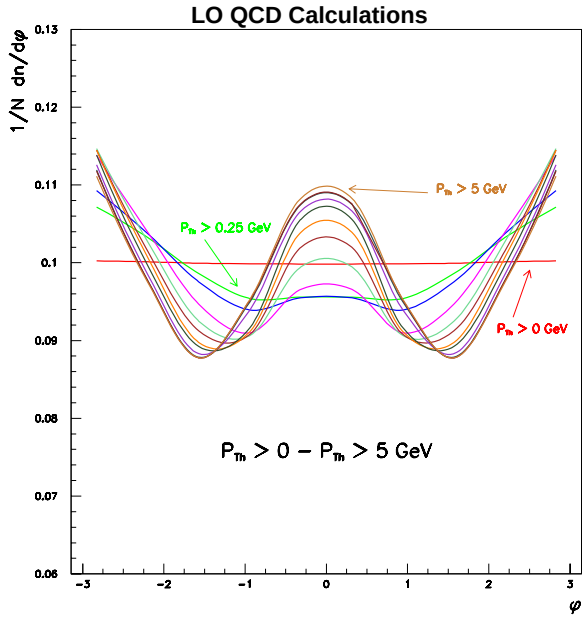

Figure 10.2: *Theoretical calculations obtained with the program AZICAL for the differential azimuthal angle distributions* $1/N\ dn/d\phi$ *as a function of the* P_{Th} *cut on the hadrons.*

10.2.4 ϕ Distributions and Moments

Figure 10.2 superposes the differential azimuthal angle distributions $1/N\ dn/d\phi$ for various values of the P_{Th} cut on the hadrons in the range from 0 to 5 GeV. The distribution for $P_{Th} > 0$ GeV is mostly flat – the moments are close to zero. But a clear negative $\langle \cos \phi \rangle$ and a positive $\langle \cos 2\phi \rangle$ appear as the P_{Th} cut increases. This evolution is best seen in figure 10.3: the LO QCD calculations predict the ϕ asymmetries to be of the order of a few percent. The $\langle \cos 2\phi \rangle$ moment increases with increasing P_{Th} cut and tends to flatten off at high P_{Th}. The $\langle \cos \phi \rangle$ moment is always negative, has a minimum in the region $P_{Th} \approx 1$ GeV, and approaches a value close to zero at high P_{Th}.

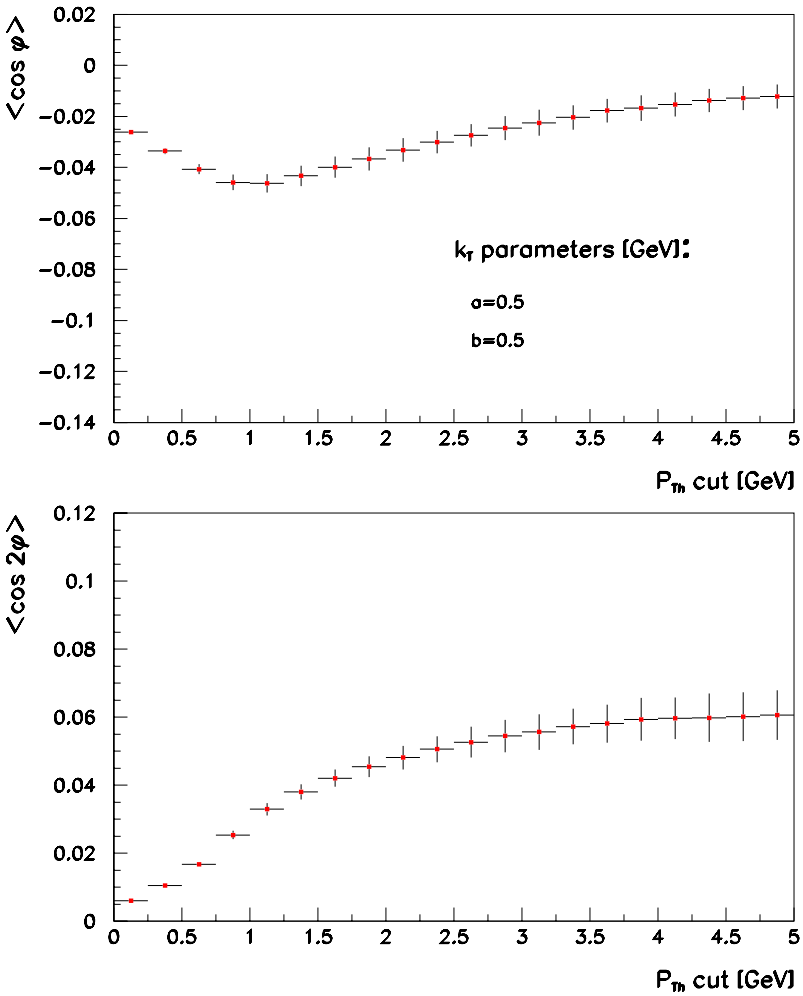

Figure 10.3: *Theoretical calculations obtained with the program AZICAL for the moments* $\langle \cos \phi \rangle$ *and* $\langle \cos 2\phi \rangle$ *as a function of the* P_{Th} *cut on the hadrons. The error bars represent the numerical uncertainty in the calculations.*

10.3 NLO QCD Calculations for Jets

Several MC programs are available to perform NLO QCD calculations. They differ mostly in the way the cancellation of the virtual and soft and collinear divergences is implemented numerically. The NLO MC program DISENT [80] is extensively used in ZEUS analyses for the comparison of jet cross sections to data. We also opted for DISENT.

10.3.1 Theoretical Uncertainties

As in all NLO MC programs, the matrix elements are calculated exactly to second order in the perturbative α_s expansion. And the running coupling constant α_s is evaluated at the renormalisation scale μ_R. The uncertainty due to the neglected higher order contributions is estimated in the so-called renormalisation-scale uncertainty. By convention, the latter is defined as the result of the variation of the renormalisation scale by a factor of 2. This uncertainty turns out to be the largest contribution to the total theoretical uncertainty.

The uncertainty on the factorisation scale also needs to be included in the total theoretical uncertainty. The factorisation scale μ_F is introduced in the calculations as the scale at which the divergencies are factorised: the divergencies coming from initial state radiation are "absorbed" in the definition of the parton densities, whereas the remaining virtual and soft and collinear divergencies are added together to cancel.

Other sources contribute to the theoretical uncertainty of NLO calculations: the errors on the parton distribution functions and the value of the strong coupling constant α_s. These contributions turn out to be negligible in comparison to the uncertainties on the scales, and were therefore neglected here.

10.3.2 Hadronisation Corrections

Unlike LO MC programs, DISENT does not incorporate any phenomenological model of hadronisation; the results are expressed at parton level. It is therefore imperative to

correct these NLO calculations for the effects of hadronisation before a comparison with the data can be made. Correction factors are estimated with LO MC programs. We used Lepto and Ariadne to calculate the correction factors from the ratio of the ϕ^{jet} distributions in dijet events at parton and hadron levels:

$$C_f(i) = \left. \frac{\Phi_i^{had}}{\Phi_i^{par}} \right|_{LO\ MC} , \tag{10.1}$$

Φ_i^{had} and Φ_i^{par} being the distributions for bin i at hadron and parton level, respectively, and $C_f(i)$ the correction factor for bin i. It is important to notice that such correction procedure implicitly assumes that the parton levels in LO and NLO programs are in agreement. We verified the applicability of this hadronisation correction comparing the parton level ϕ^{jet} distributions as given by DISENT, Lepto and Ariadne: the shape of the distributions were found to be in reasonably good agreement. The correction factors obtained were found to be rather flat for both Lepto and Ariadne, with fluctuations within 2%. Hence no hadronisation correction was applied to the DISENT calculation.

10.3.3 ϕ Distribution in Dijet Events

The central values of the DISENT calculations were obtained in the same kinematic range as the data and with the same set of cuts applied to the LO MC samples at hadron level for the analysis of dijet events (chapter 8). In addition, we chose the following settings for DISENT:

- factorisation scale $\mu_F = Q$;
- renormalisation scale $\mu_R = Q$;
- the CTEQ5D PDF.

The results obtained with DISENT are presented in figure 10.4. Also plotted are the theoretical uncertainties on the renormalisation and factorisation scales, both varied by a factor of 2 from their central values Q.

A fit to the distribution was performed: a non-negligible and positive $\langle \cos 2\phi \rangle$ moment is observed, whereas the $\langle \cos \phi \rangle$ and $\langle \sin \phi \rangle$ moments are zero.

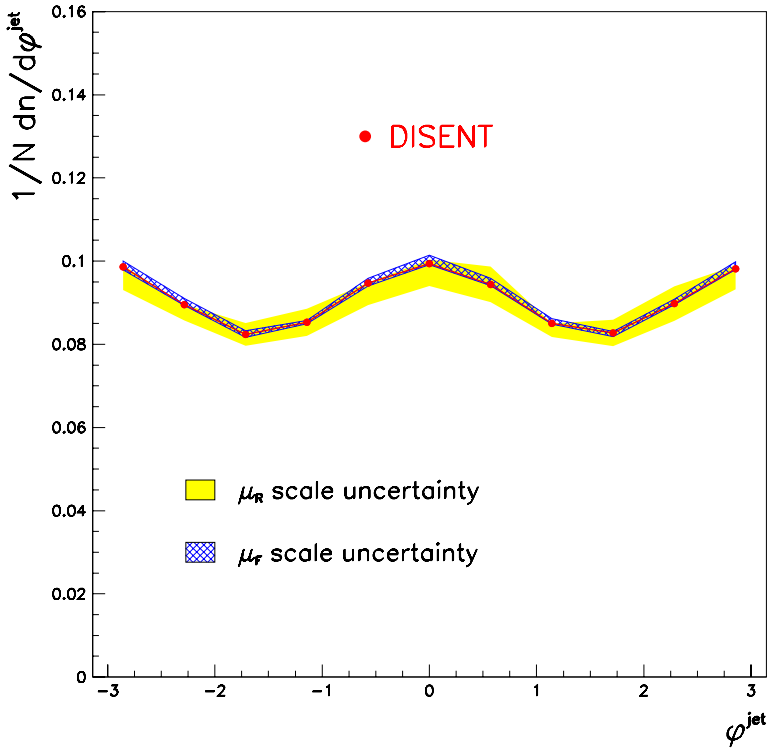

Figure 10.4: *NLO QCD prediction obtained with DISENT for the ϕ^{jet} distribution of dijet events. The solid (yellow) band represents the uncertainty on the renormalisation scale μ_R, calculated using $\mu_R = 2Q, Q/2$ instead of the central value $\mu_R = Q$. The hatched (blue) band are represents the uncertainty on the factorisation scale μ_F, calculated using $\mu_F = 2Q, Q/2$ instead of the central value $\mu_F = Q$.*

Chapter 11

Comparisons with MC and Theory and General Conclusions

This last chapter will summarise the experimental results of the analyses presented in this thesis and discuss comparisons of the latter with Monte Carlo programs and theoretical calculations.

11.1 Azimuthal Angle Distributions of Hadrons

11.1.1 Data and MC Models

Data were compared with the 2 MC generators used in the thesis: Lepto and Ariadne. The comparisons with data for the moments of the ϕ_h distributions of tracks in the HCM frame are presented in figure 11.1. The MC predictions were obtained at hadron level.

Both Ariadne and Lepto are in general qualitative agreement with the data: the $\langle \cos \phi_h \rangle$ moment is negative and its magnitude increases as the cut on the transverse momentum of the hadrons, P_{Th}^{cut}, increases, and has a turn at higher P_{Th}^{cut} values when it starts approaching zero. The $\langle \cos 2\phi_h \rangle$ moment is positive and increases with increasing P_{Th}^{cut}. But both MCs tend to underestimate the magnitude of the moments given the statistical errors: the magnitude of the $\langle \cos \phi_h \rangle$ moment is underestimated by the MCs

at low P_{Th}^{cut}, and the $\langle \cos 2\phi_h \rangle$ moment is underestimated at the higher P_{Th}^{cut} values (the data shows a steeper increase with P_{Th}^{cut} compared to the MC).

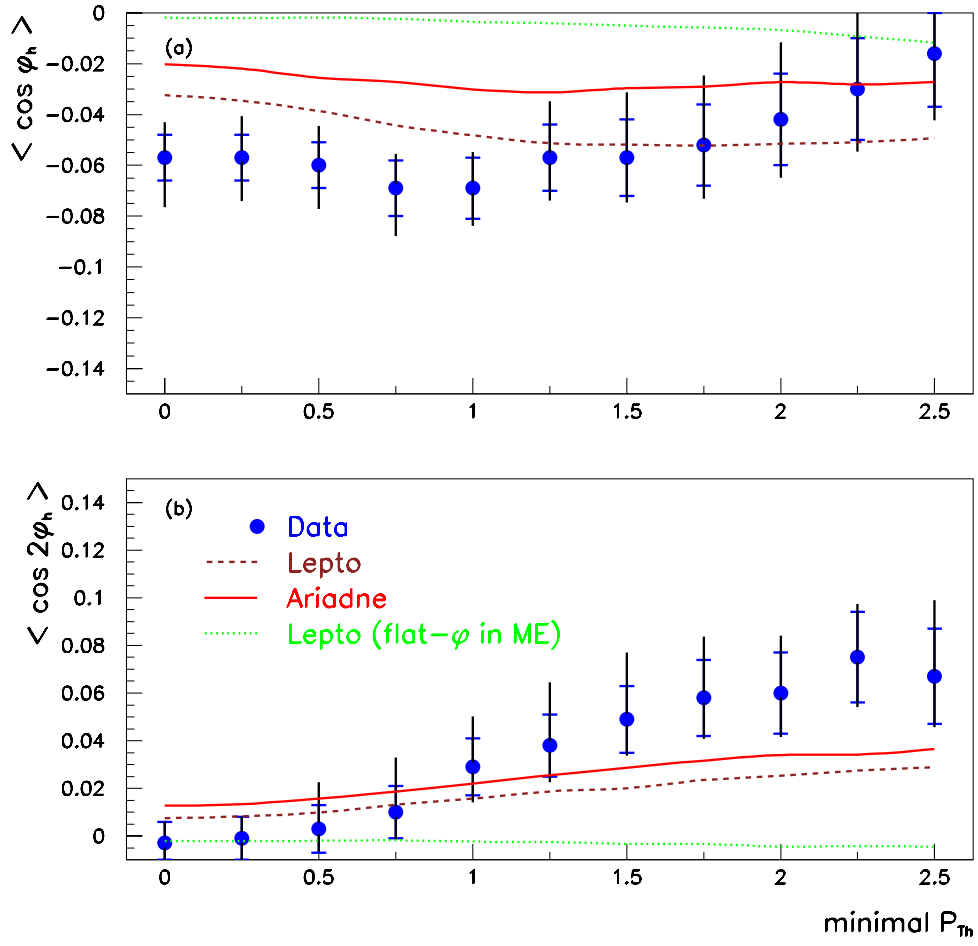

Figure 11.1: *Evolution with the minimal transverse momentum P_{Th} of the tracks of the $\langle \cos \phi_h \rangle$ and $\langle \cos 2\phi_h \rangle$ moments of the ϕ_h distributions: comparison of data with the LO MC event generators Lepto and Ariadne.*

Also shown are the predictions obtained with a modified Lepto sample denoted "flat-ϕ in ME" in figure 11.1. This sample is analogue to the "standard" Lepto sample, apart from the following: the azimuthal angle distribution of the matrix elements used to generate an event in Lepto was forced to be flat. This way the only contribution to a ϕ asymmetry must come from QCD effects in the hadronisation process. As can be

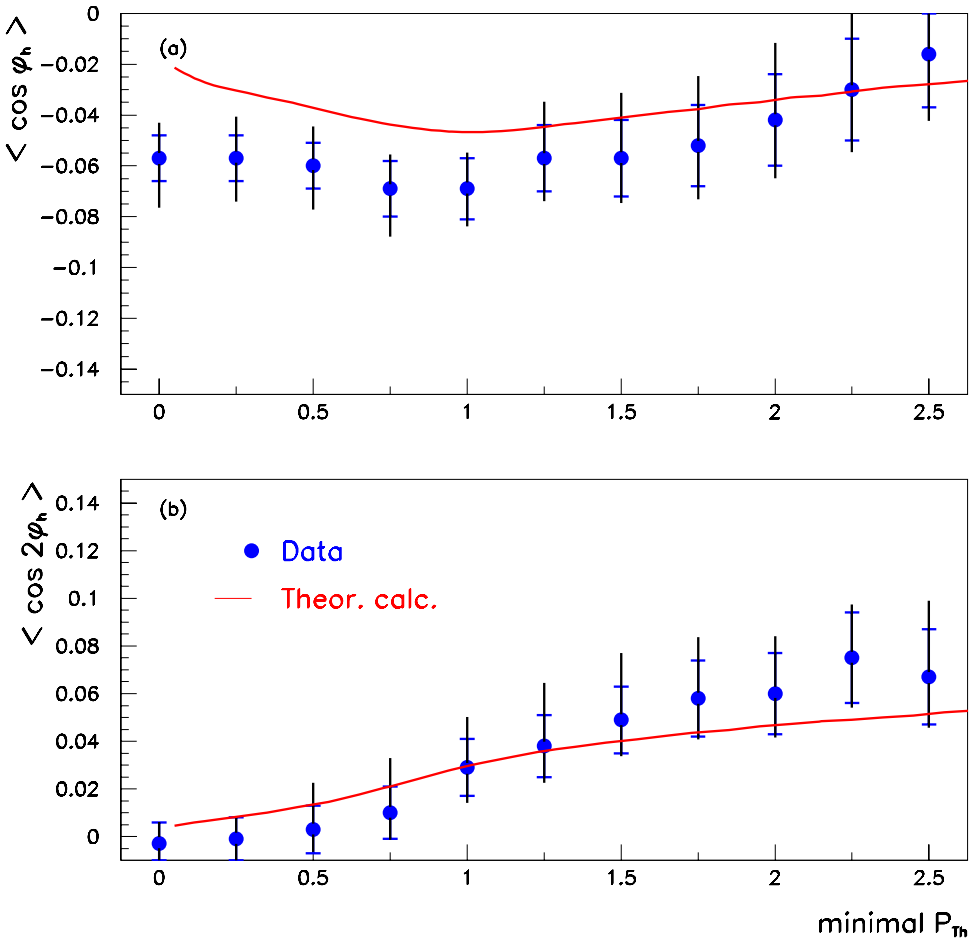

Figure 11.2: *Evolution with the minimal transverse momentum P_{Th} of the tracks of the $\langle \cos \phi_h \rangle$ and $\langle \cos 2\phi_h \rangle$ moments of the ϕ_h distributions: comparison of data with QCD theoretical calculations performed with the AZICAL program.*

seen in figure 11.1 the data clearly excludes such a situation: the flat-ϕ sample gives rise to moments that are consistent with zero regardless of the P_{Th} cut on the hadrons. A comparison between both samples hence leads to the conclusion that the observed asymmetries come mainly from the non-trivial ϕ dependence of the scattering process's matrix elements.

11.1.2 Data and Theory

The measured moments were also compared to the LO QCD theoretical calculations performed with the AZICAL program (presented in section 10). Figures 11.2(a) and 11.2(b) present the comparisons for the $\langle \cos \phi_h \rangle$ and $\langle \cos 2\phi_h \rangle$ moments, respectively. There is a qualitative agreement between data and theory, which improves as the transverse momentum cut on the hadrons increases. The $\langle \cos 2\phi_h \rangle$ moment is reasonably well described within the experimental errors, but the calculated magnitude of the $\langle \cos \phi_h \rangle$ moment is underestimated in the calculations at low P_{Th}.

We investigated with the AZICAL program the relative importance of the perturbative and non-perturbative contributions to the moments. The results are shown in figure 11.3. In general, the relative importance of the 2 contributions depends on the fragmentation k_T and hadronisation p_T (refer to section 10 for details of these quantities). For the $\langle \cos \phi_h \rangle$ moment the non-perturbative contribution is significant. But the non-perturbative contribution to the $\langle \cos 2\phi_h \rangle$ moment is negligible for a large range of reasonable fragmentation k_T and hadronisation p_T values. We therefore conclude that the experimental observation of a non-zero $\langle \cos 2\phi_h \rangle$ moment is evidence of the perturbative contribution, observed for the first time in DIS at HERA.

It is also worth noticing that the LO QCD calculations are in rather good agreement with the LO MC simulations. And indeed the theoretical calculations with AZICAL seem to reproduce the P_{Th} dependence of the moments marginally better than Lepto and Ariadne. This is particularly true for the $\langle \cos \phi_h \rangle$ moment, proving that both the intrinsic and hadronisation transverse momenta have a role in the azimuthal angle distributions.

11.2 Azimuthal Angle Distributions of Jets

11.2.1 Data and MC Models

The measured ϕ^{jet} distribution of dijet events is plotted in figure 11.4 together with the LO MC predictions of Lepto and Ariadne. Both MC programs are in rather good

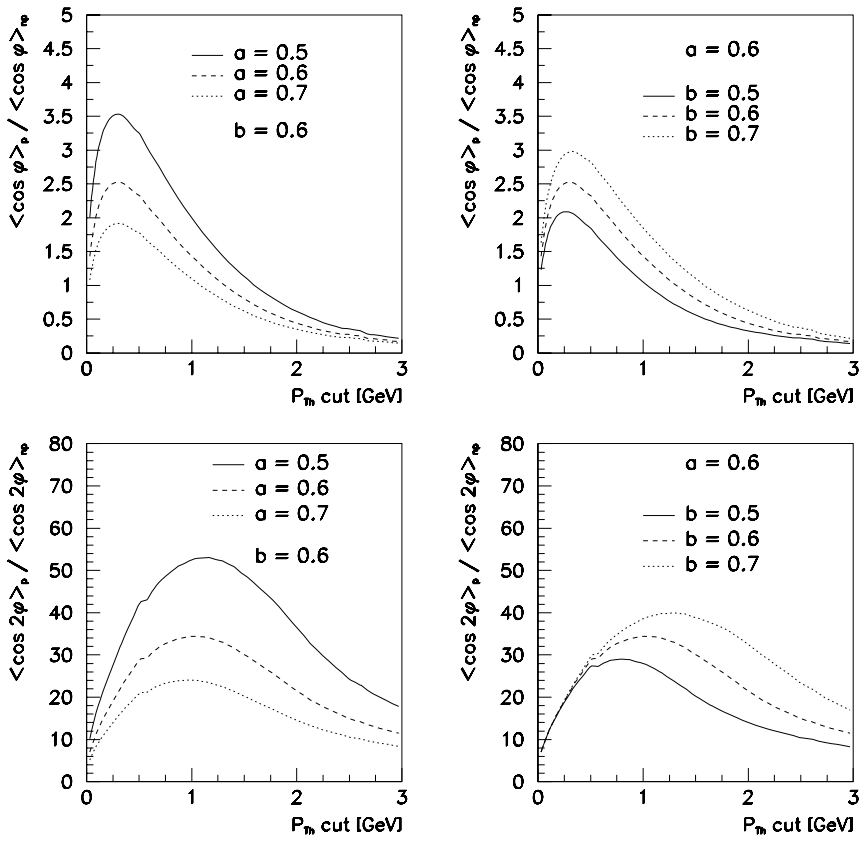

Figure 11.3: *Theoretical calculations obtained with the program AZICAL for the moments* $\langle \cos \phi \rangle$ *and* $\langle \cos 2\phi \rangle$ *as a function of the* P_{Th} *cut on the hadrons. The distributions show the ratio of the moments calculated from the non-perturbative (subscript "np") and perturbative (subscript "p") contributions separately, for various values of the intrinsic parton* k_T *(a parameter) and the hadronisation* p_T *(b parameter) on the moments (a and b are given in GeV).*

agreement with the data, and show the clear $\cos 2\phi^{jet}$ shape predicted by QCD.

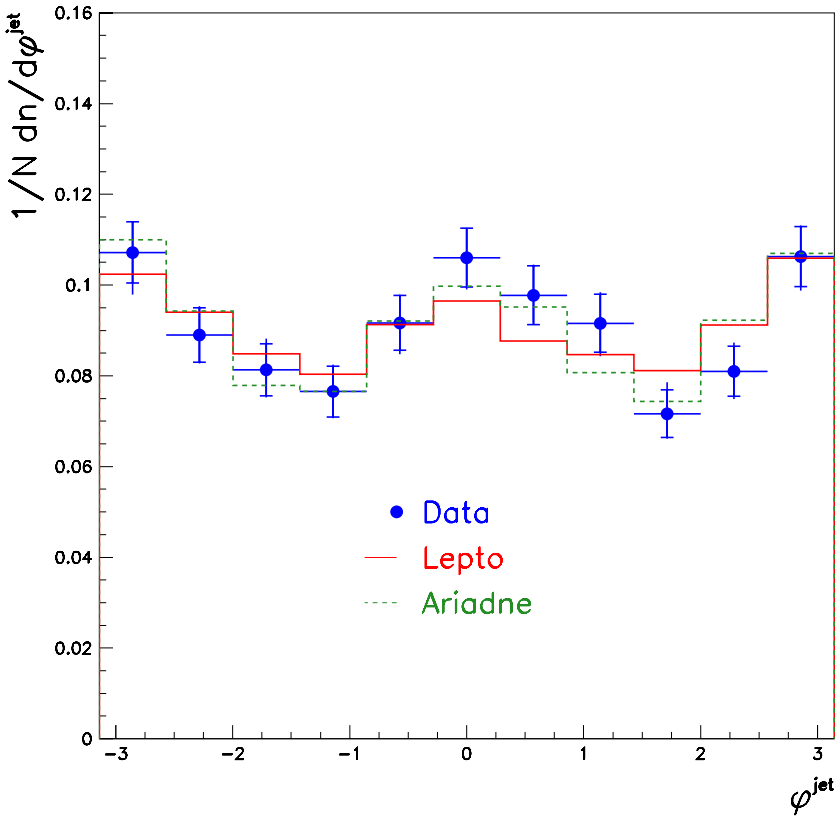

Figure 11.4: *Jet azimuthal angle distribution of dijet events: comparison of data with the LO MC event generators Lepto and Ariadne.*

11.2.2 Data and Theory

In figure 11.5 the measured jet azimuthal angle distribution of dijet events is compared with DISENT NLO calculations. The non-trivial azimuthal angle dependence is remarkably well established experimentally. The NLO ϕ^{jet} distribution is also in good agreement with the data, although the magnitude of the $\cos 2\phi^{jet}$ contribution seems to be slightly underestimated.

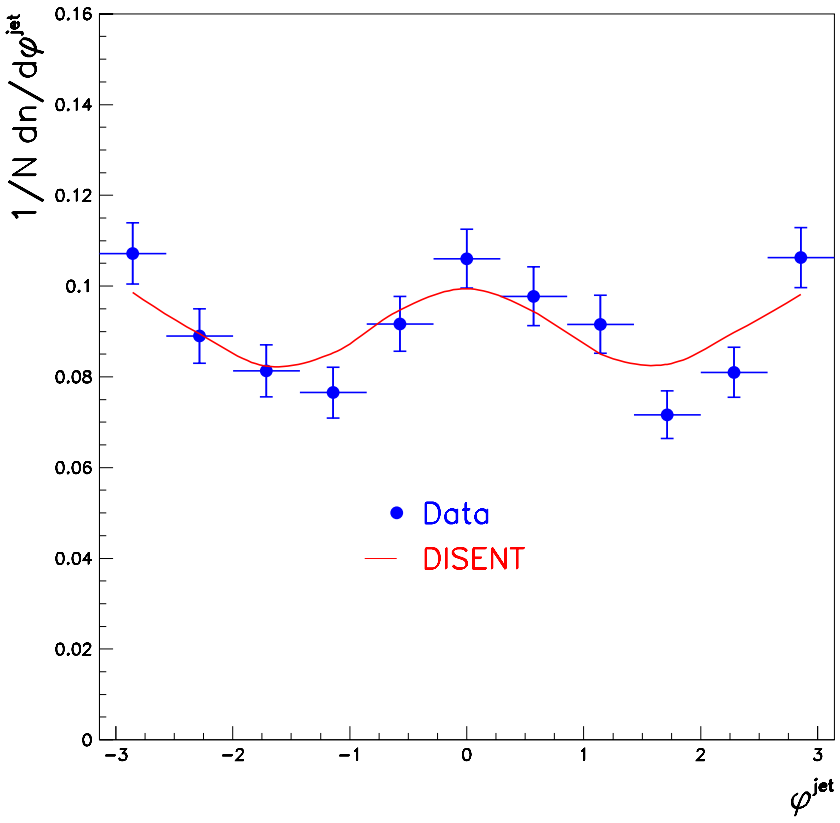

Figure 11.5: *Jet azimuthal angle distribution of dijet events: comparison of data with the NLO QCD program DISENT.*

11.3 General Conclusions

In this thesis we investigated in detail the azimuthal angle correlations of hadrons and jets in positron-proton deep inelastic scattering.

The distributions of hadrons in azimuthal angle ϕ_h (around the direction of the exchanged boson) were measured in the hadronic centre-of-mass frame. The $\langle cos\phi \rangle$ and $\langle cos2\phi \rangle$ moments were obtained from fits to these distributions. We studied the evolution of these moments with the hadrons transverse momenta.

Theoretical QCD calculations have been obtained with a program implementing all the leading order QCD matrix elements and a simple model to account for the transverse momentum of the intrinsic partons within the proton and the fragmentation transverse momentum. The data were also compared with the leading order MC event generators Lepto and Ariadne. Both the theory and the MC calculations are in general good agrement with the data, although the magnitude of the moments is underestimated, specially in the low transverse momentum region.

The $\langle cos2\phi \rangle$ was measured for the first time and shown to be to a large extent due to the perturbative QCD contribution.

A similar analysis was performed in the Breit frame with dijet events. The $\cos 2\phi$ dependence of the cross section predicted by QCD is clearly observed in the data. Comparisons were made with leading order Monte Carlo programs and NLO calculations. Both showed a good agreement with the data.

An attempt was also made to discriminate between quark- and gluon-initiated jets in QCDC and BGF events using artificial neural networks. The strengths and limitations of the method were described in length.

Appendix A

Differential Cross Sections for Semi-inclusive Leptoproduction

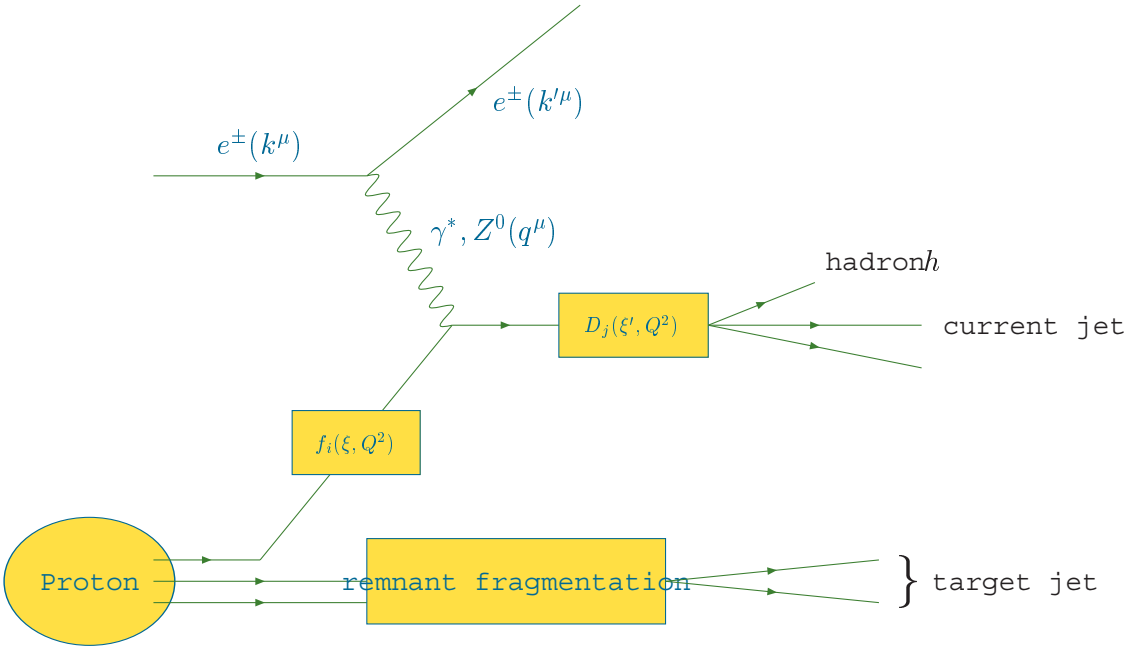

Figure A.1: *Schematic diagram of the factors intervening in the calculation of a semi-inclusive DIS process.*

The semi-inclusive DIS cross section for the process

$$e(k^\mu) + P(P^\mu) \to e'(k'^\mu) + h(P_h^\mu) + X(P'^\mu) \quad , \tag{A.1}$$

is calculated on the assumption that the hadron level expression is related to the parton level through an integration over the parton variables. The overall cross section is expressed as a convolution of (*cf.* figure A.1):

1. the parton distribution functions describing the partons in the initial state;
2. the fragmentation functions describing the distributions of the final state hadrons;
3. the parton hard scattering cross section.

Within the framework of the Quark Parton Model it is written as

$$\begin{aligned}\frac{d^5\sigma}{dx\,dy\,dz\,dP_{Th}\,d\phi} = \\ \sum_{i,j}\int dx_p\,dz_p\,dp_T\,d\xi\,d\xi'\,\delta(x-\xi x_p)\,\delta(z-\xi' z_p)\,\delta(P_{Th}-\xi' p_T) \\ \times f_i(\xi,Q^2)\frac{d^5\hat{\sigma}_{ij}}{dx_p\,dy\,dz_p\,dp_T\,d\phi}\,D_j(\xi',Q^2)\end{aligned} \tag{A.2}$$

The sum indexes i and j run over all partons – quarks, antiquarks and gluons – in the initial and final states,

$$e(k^\mu) + \text{parton } i(p^\mu) \to e'(k'^\mu) + \text{parton } j(p'^\mu) + X \quad . \tag{A.3}$$

The following variables are used to describe the cross section:

$$x_p = \frac{Q^2}{2p.q} = \frac{x}{\xi} \quad , \tag{A.4}$$

$$z_p = \frac{p.p'}{p.q} = \frac{z}{\xi'} \quad , \tag{A.5}$$

$$\phi \text{ (azimuthal angle of the outgoing hadron } h) \quad , \tag{A.6}$$

$$p_T \text{ (transverse momentum of the incoming parton)} \quad . \tag{A.7}$$

x_p is the partonic equivalent of the Bjorken x variable, and z_p can be related to the angular distribution of the 2 partons in their centre-of-mass frame:

$$z_p = \frac{1}{2}(1 - \cos\theta^*) \quad , \tag{A.8}$$

θ^* being the polar angle of one of the partons. $f_i(\xi, Q^2)$ is a parton distribution function; it gives , in the QPM, the probability for parton i to carry a fraction ξ of the proton's momentum: $p^\mu = \xi P^\mu$. $D_j(\xi', Q^2)$ is the fragmentation function which gives the probability for a hadron to carry a fraction ξ' of the momentum of parent parton j: $P_h^\mu = \xi' p^\mu$.

Within this framework, $x_p = z_p = 1$. The parton level cross section $\hat{\sigma}_{ij}$ is then

$$\frac{d^5\hat{\sigma}_{jk}}{dx_p\, dy\, dz_p\, dp_T\, d\phi} = \frac{\alpha^2}{y\, Q^2}[1 + (1-y)^2]e_i^2\delta_{ij}\delta(1-x_p)\delta(1-z_p)\delta(p_T) \quad , \tag{A.9}$$

which leads to the familiar cross section

$$\frac{d^2\sigma}{dx\, dy} = \frac{2\pi\alpha^2}{Q^4}[1 + (1-y)^2]\sum_i e_i^2 x f_i(x) \quad . \tag{A.10}$$

In what follows we will use the notations of Chay *et. al.* [19, 20] and Méndez [21]. All the expressions are here compiled for self-consistency, and can be grouped from the previously-mentioned authors and from [2].

In the expressions above, the intrinsic transverse momenta of the partons within the proton have been neglected. Also the transverse momenta of the hadrons produced in the fragmentation process were not taken into account. Most importantly, perturbative QCD (pQCD) corrections were not introduced.

Taking into account all these points, the 5-fold differential cross section for semi-inclusive deep inelastic scattering can then be written up to second order in the strong coupling constant, α_s, as

$$\begin{aligned}\frac{d^5\sigma}{dx\, dy\, dz\, dP_{Th}\, d\phi} &= \Big[(A^{(0)} + A^{(1)}) + (B^{(0)} + B^{(1)})\cos\phi + (C^{(0)} + C^{(1)})\cos 2\phi \\ &\quad + D^{(2)}\sin\phi + E^{(2)}\sin 2\phi\Big]. \end{aligned} \tag{A.11}$$

The superscripts on the As, Bs, etc. refer to the corresponding QCD $\mathcal{O}(\alpha_s)$ contributions. Each of the zeroth and first order α_s contributions are presented below in detail. The second order terms, $D^{(2)}$ and $E^{(2)}$, will not be described, as they are predicted to be zero in our experimental case. They are only present in parity-violating weak interactions and in purely electromagnetic interactions with a polarised lepton beam [63, 64]. They measure the left-right asymmetry of hadron production with respect to the lepton scattering plane. Theoretically, these terms arise from one-loop contributions to the parton level scattering amplitude, and are time-reversal-odd (T-odd). The expressions can be found in [63].

Some preliminary remarks are in order:

- All the As, Bs, etc. are functions of x, y, z and P_{Th};
- The "coupling constant" is given by
$$K = \begin{cases} \frac{\alpha^2}{Q^2} & \text{for NC events} \\ \frac{G_F^2 Q^2}{(2\pi)^2} & \text{for CC events} \end{cases}$$

A.1 Lowest-order Contributions

All lowest-order $\mathcal{O}(\alpha_s^0)$ formulae are given after integration over the momentum of the struck parton.

Chay *et. al.* [19,20] have introduced a simple model to take into account the parton's intrinsic transverse momentum and the hadronisation transverse momentum. These were assumed to follow a Gaussian distribution that factorises out of the parton distribution and fragmentation functions, respectively.

Experimentally one uses a cut $P_{Th} > P_c$ to select hadrons with a minimal transverse momentum. The integration over the relevant variables comprises the condition

$$\frac{Q^2 z^2 (1-x_p)(1-z_p)}{x_p z_p} \geq P_c^2 \quad , \tag{A.12}$$

equivalent to the range

$$0 \leq P_{hT} \leq \frac{Q^2 z (1-x)(1-z)}{x} \quad . \tag{A.13}$$

In fact the same condition applies to the integration of the higher order terms.

The QPM $\mathcal{O}(\alpha_s^0)$ contribution is (after integration)

$$\frac{d^4\sigma^{(0)}}{dx\,dy\,dz\,d\phi} = A^{(0)} + B^{(0)}\cos\phi + C^{(0)}\cos 2\phi \quad . \tag{A.14}$$

If one ignores the intrinsic transverse momentum of the struck parton, $p_T \equiv 0$:

$$\begin{aligned} A^{(0)} &= \tfrac{K}{y}\Big\{[1+(1-y)^2]H_1(x,z,Q^2) \pm [1-(1-y)^2]H_1'(x,z,Q^2)\Big\}\delta(P_{Th}) \\ &\quad \text{for } P_{Th} \equiv 0 \\ &= 0 \text{ for } P_{Th} > P_c \\ B^{(0)} &= 0 \\ C^{(0)} &= 0 \end{aligned}$$

If $p_T \neq 0$ and $P_{Th} \geq P_c$:

$$\begin{aligned} A^{(0)} &= \tfrac{K}{y} e^{-P_c^2/(b^2+a^2z^2)} \\ &\quad \times\Bigg\{\Big\{[1+(1-y)^2] + \tfrac{4(1-y)}{Q^2}\big[\tfrac{a^2b^2}{b^2+a^2z^2} + \big(\tfrac{a^2z}{b^2+a^2z^2}\big)^2(P_c^2+b^2+a^2z^2)\big]\Big\}H_1(x,z,Q^2) \\ &\quad \pm\Big\{[1-(1-y)^2] + \tfrac{4(1-y)}{Q^2}\big[\tfrac{a^2b^2}{b^2+a^2z^2} + \big(\tfrac{a^2z}{b^2+a^2z^2}\big)^2(P_c^2+b^2+a^2z^2)\big]\Big\}H_1'(x,z,Q^2)\Bigg\} \\ B^{(0)} &= -8K\frac{(2-y)\sqrt{1-y}}{y}\frac{a^2z}{Q\sqrt{b^2+a^2z^2}}\int_{\frac{P_c}{\sqrt{b^2+a^2z^2}}}^{\infty} u^2e^{-u^2}du \times H_1(x,z,Q^2) \\ C^{(0)} &= 4K\frac{1-y}{yQ^2}e^{-P_c^2/(b^2+a^2z^2)}\Big[\frac{a^2b^2}{b^2+a^2z^2} + \left(\frac{a^2z}{b^2+a^2z^2}\right)^2(P_c^2+b^2+a^2z^2)\Big]H_1(x,z,Q^2) \end{aligned}$$

The $H_i(x,z,Q^2)$ and $H_i'(x,z,Q^2)$ functions are collected at the end of this appendix. They are linear combinations of parton distribution and fragmentation functions.

The terms proportional to $H_i'(x,z,Q^2)$ have a $\pm$ sign, and come from the vector-axial interference terms of the scattering amplitude. The $+$ $(-)$ sign corresponds to neutrino (antineutrino) scattering. These terms do not contribute in the case of a purely electromagnetic interaction.

A.2 First-order Contributions

The $\mathcal{O}(\alpha_s)$ contribution to the cross section is given by

$$\frac{d^5\sigma^{(1)}}{dx\,dy\,dz\,dP_{Th}\,d\phi} = A^{(1)} + B^{(1)}\cos\phi + C^{(1)}\cos 2\phi \quad , \tag{A.15}$$

where

$$A^{(1)} = \alpha_s \frac{K}{y}\frac{1}{2\pi}\int_x^1 \frac{dx_p}{x_p}\int_z^1 \frac{dz_p}{z_p}\left[\sum_{i=1}^3 c_i A_i H_i(\tfrac{x}{x_p},\tfrac{z}{z_p},Q^2) \pm \sum_{i=1}^3 c_i A_i' H_i'(\tfrac{x}{x_p},\tfrac{z}{z_p},Q^2)\right] \times\delta\left(P_{Th} - \left[\tfrac{Q^2 z^2 (1-x_p)(1-z_p)}{x_p z_p}\right]^{1/2}\right)$$

with

$$\begin{aligned}
A_1 &= [1+(1-y)^2]\left[(1-x_p)(1-z_p) + \tfrac{1+x_p^2 z_p^2}{(1-x_p)(1-z_p)}\right] + 8(1-y)x_p z_p \\
&= [1+(1-y)^2]\tfrac{x_p^2+z_p^2}{(1-x_p)(1-z_p)} + 2y^2(1+x_p z_p) + 4(1-y)(1+3x_p z_p) \\
A_2 &= A_1(z_p \leftrightarrow 1-z_p) \\
A_3 &= [1+(1-y)^2][x_p^2+(1-x_p)^2]\tfrac{z_p^2+(1-z_p)^2}{z_p(1-z_p)} + 16(1-y)x_p(1-x_p) \\
A_1' &= [1-(1-y)^2]\left[2(x_p+z_p) + \tfrac{x_p^2+z_p^2}{(1-x_p)(1-z_p)}\right] \\
&= [1-(1-y)^2]\left[-(1-x_p)(1-z_p) + \tfrac{1+x_p^2 z_p^2}{(1-x_p)(1-z_p)}\right] \\
A_2' &= A_1'(z_p \leftrightarrow 1-z_p) \\
A_3' &= [1-(1-y)^2][x_p^2+(1-x_p)^2]\tfrac{2z_p-1}{z_p(1-z_p)} \\
&= y(2-y)[x_p^2+(1-x_p)^2]\tfrac{2z_p-1}{z_p(1-z_p)}
\end{aligned}$$

and

$$B^{(1)} = \alpha_s \frac{K}{y}\frac{1}{2\pi}\int_x^1 \frac{dx_p}{x_p}\int_z^1 \frac{dz_p}{z_p}\left[\sum_{i=1}^3 c_i B_i H_i(\tfrac{x}{x_p},\tfrac{z}{z_p},Q^2) \pm \sum_{i=1}^3 c_i B_i' H_i'(\tfrac{x}{x_p},\tfrac{z}{z_p},Q^2)\right] \times\delta\left(P_{Th} - \left[\tfrac{Q^2 z^2 (1-x_p)(1-z_p)}{x_p z_p}\right]^{1/2}\right)$$

with

$$\begin{aligned}
B_1 &= -4(2-y)\sqrt{1-y}\sqrt{\tfrac{x_p z_p}{(1-x_p)(1-z_p)}}\big[x_p z_p + (1-x_p)(1-z_p)\big] \\
B_2 &= -B_1(z_p \leftrightarrow 1-z_p) \\
B_3 &= -4(2-y)\sqrt{1-y}\sqrt{\tfrac{x_p(1-x_p)}{z_p(1-z_p)}}(1-2x_p)(1-2z_p) \\
B_1' &= -4y\sqrt{1-y}\sqrt{\tfrac{x_p z_p}{(1-x_p)(1-z_p)}}(x_p+z_p-1) \\
B_2' &= 4y\sqrt{1-y}\sqrt{\tfrac{x_p(1-z_p)}{(1-x_p)z_p}}(x_p-z_p) \\
&= -B_1'(z_p \leftrightarrow 1-z_p) \\
B_3' &= 4y\sqrt{1-y}\sqrt{\tfrac{x_p(1-x_p)}{z_p(1-z_p)}}(1-2x_p)
\end{aligned}$$

and

$$C^{(1)} = \alpha_s \frac{K}{y}\frac{1}{2\pi}\int_x^1 \frac{dx_p}{x_p}\int_z^1 \frac{dz_p}{z_p}\left[\sum_{i=1}^3 c_i C_i H_i(\tfrac{x}{x_p},\tfrac{z}{z_p},Q^2)\right] \times\delta\left(P_{Th} - \left[\tfrac{Q^2 z^2 (1-x_p)(1-z_p)}{x_p z_p}\right]^{1/2}\right)$$

with

$$
\begin{aligned}
C_1 &= 4(1-y)x_p z_p \\
C_2 &= 4(1-y)x_p(1-z_p) = C_1(z_p \leftrightarrow 1-z_p) \\
C_3 &= 8(1-y)x_p(1-x_p)
\end{aligned}
$$

Again the $\pm$ sign terms have the same meaning as above. The c_i coefficients are color factors: $c_1 = c_2 = 4/3$ and $c_3 = 1/2$.

The $H_i(x,z,Q^2)$ and $H_i'(x,z,Q^2)$ functions have the following expressions:

$$
\begin{aligned}
H_1(\xi,\xi',Q^2) &= \sum_q e_q^2 f_q(\xi,Q^2) D_q(\xi',Q^2) + \sum_{\bar{q}} e_{\bar{q}}^2 f_{\bar{q}}(\xi,Q^2) D_{\bar{q}}(\xi',Q^2) \\
&= \sum_{q,q'} |V_{q'q}|^2 f_q(\xi,Q^2) D_{q'}(\xi',Q^2) + \sum_{\bar{q},\bar{q}'} |V_{qq'}|^2 f_{\bar{q}}(\xi,Q^2) D_{\bar{q}'}(\xi',Q^2) \\
H_2(\xi,\xi',Q^2) &= \left[\sum_q e_q^2 f_q(\xi,Q^2) + \sum_{\bar{q}} e_{\bar{q}}^2 f_{\bar{q}}(\xi,Q^2) \right] D_g(\xi',Q^2) \\
&= \left[\sum_{q,q'} |V_{q'q}|^2 f_q(\xi,Q^2) + \sum_{\bar{q},\bar{q}'} |V_{qq'}|^2 f_{\bar{q}}(\xi,Q^2) \right] D_g(\xi',Q^2) \\
H_3(\xi,\xi',Q^2) &= f_g(\xi,Q^2) \left[\sum_q e_q^2 D_q(\xi,Q^2) + \sum_{\bar{q}} e_{\bar{q}}^2 D_{\bar{q}}(\xi,Q^2) \right] \\
&= f_g(\xi,Q^2) \left[\sum_{q,q'} |V_{q'q}|^2 D_q(\xi,Q^2) + \sum_{\bar{q},\bar{q}'} |V_{qq'}|^2 D_{\bar{q}}(\xi,Q^2) \right] \\
H_1'(\xi,\xi',Q^2) &= \sum_q e_q^2 f_q(\xi,Q^2) D_q(\xi',Q^2) - \sum_{\bar{q}} e_{\bar{q}}^2 f_{\bar{q}}(\xi,Q^2) D_{\bar{q}}(\xi',Q^2) \\
&= \sum_{q,q'} |V_{q'q}|^2 f_q(\xi,Q^2) D_{q'}(\xi',Q^2) - \sum_{\bar{q},\bar{q}'} |V_{qq'}|^2 f_{\bar{q}}(\xi,Q^2) D_{\bar{q}'}(\xi',Q^2) \\
H_2'(\xi,\xi',Q^2) &= \left[\sum_q e_q^2 f_q(\xi,Q^2) - \sum_{\bar{q}} e_{\bar{q}}^2 f_{\bar{q}}(\xi,Q^2) \right] D_g(\xi',Q^2) \\
&= \left[\sum_{q,q'} |V_{q'q}|^2 f_q(\xi,Q^2) - \sum_{\bar{q},\bar{q}'} |V_{qq'}|^2 f_{\bar{q}}(\xi,Q^2) \right] D_g(\xi',Q^2) \\
H_3'(\xi,\xi',Q^2) &= f_g(\xi,Q^2) \left[\sum_q e_q^2 D_q(\xi,Q^2) - \sum_{\bar{q}} e_{\bar{q}}^2 D_{\bar{q}}(\xi,Q^2) \right] \\
&= f_g(\xi,Q^2) \left[\sum_{q,q'} |V_{q'q}|^2 D_q(\xi,Q^2) - \sum_{\bar{q},\bar{q}'} |V_{qq'}|^2 D_{\bar{q}}(\xi,Q^2) \right]
\end{aligned}
$$

In all the H_i and H_i' expressions the first and second rows correspond, respectively, to the NC and the helicity $\lambda = +1$ CC case.

Bibliography

[1] T. Gehrmann. Two-loop jet physics: status and prospects. *Talk presented at the 9th International Workshop on "Deep Inelastic Scattering and QCD" (DIS2001), April 2001, Bologna, Italy*, 2001.

[2] H. Georgi and H. D. Politzer. Clean tests of quantum chromodynamics in μp scattering. *Phys. Rev. Lett.*, 40(1):3, 1978.

[3] R. N. Cahn. Azimuthal dependence in leptoproduction: a simple parton model calculation. *Phys. Lett. B*, 78(2,3):269, 1978.

[4] R. N. Cahn. Critique of parton-model calculations of azimuthal dependence in leptoproduction. *Phys. Rev. D*, 40(9):3107, 1989.

[5] J. J. Aubert *et al.* (EMC Collaboration). Measurement of hadronic azimuthal ditributions in deep inelastic muon proton scattering. *Phys. Lett.*, 130B:118, 1983.

[6] M. Arneodo *et al.* (EMC Collaboration). Measurement of hadron azimuthal ditributions in deep inelastic scattering. *Z. für Phys.*, C34:277, 1987.

[7] A. Mukherjee *et al.* Azimuthal energy flow in deep-inelastic neutrino scattering. *Phys. Rev. Lett.*, 60(11):991, 1988.

[8] M. R. Adams *et al.* (Fermilab E665 Collaboration). Perturbative QCD effects observed in 490 GeV deep-inelastic muon scattering. *Phys. Rev.*, D48(11):5057, 1993.

[9] J. Breitweg *et al.* (ZEUS Collaboration). Measurement of azimuthal asymmetries in deep inelastic scattering. *Phys. Lett.*, B481(2-4):199, 2000.

[10] E. Rodrigues. Traditional versus new treatment at high Q^2 for ariadne. *Talk presented at the Workshop on "Monte Carlo Generators for HERA Physics", 1st-5th February 1999, DESY, Hamburg*, 1999.

[11] N. H. Brook, T. Carli, E. Rodrigues, M. R. Sutton, N. Tobien, and M. Weber. A comparison of deep inelastic scattering Monte Carlo event generators to HERA data. In A.T. Doyle, G. Grindhammer, G. Ingelman, and H. Jung, editors,

Proceedings of the Workshop on Monte Carlo Generators for HERA Physics, DESY, Hamburg, Apr. 1998 - Feb. 1999.

[12] G. Ingelman and R. Rueckl. *Phys. Lett.*, B201:369, 1988.

[13] A. M. Cooper-Sarkar and R. C. E. Devenish. Structure functions of the nucleon and their interpretation. DESY 97-226, 1997.

[14] V. N. Gribov and L. N. Lipatov. *Sov. J. Nucl. Phys.*, 15:438, 1972.

[15] Yu. L. Dokshitzer. *Sov. Phys. JETP*, 46:641, 1977.

[16] G. Altarelli and G. Parisi. *Nucl. Phys.*, B126:297, 1977.

[17] R. P. Feynman. *Photon-Hadron Interactions.* Benjamin, 1972.

[18] B. Levtchenko. The Breit frame (primer). Technical Report ZEUS Note 94-099, 1994.

[19] J. Chay *et al.* Azimuthal asymmetry in lepton-proton scattering at high energies. *Phys. Lett. B*, 269:175, 1991.

[20] J. Chay *et al.* Azimuthal asymmetry in lepton-proton scattering at high energies. *Phys. Rev.*, D45:46, 1992.

[21] A. Méndez. QCD predictions for semi-inclusive and inclusive leptoproducion. *Nucl. Phys.*, B145:199, 1978.

[22] G. Ingelman *et al. Comp. Phys. Comm.*, 101:108, 1997.

[23] E. Mirkes and S. Willfahrt. Effects of jet azimuthal angular distributions in dijet production cross sections in DIS. *Phys. Lett.*, B(414):205, 1997.

[24] *HERA, A Proposal for a Large Electron-Proton Collising Beam Facility at DESY*, volume DESY HERA 81-10. 1981.

[25] R. Brinkman. HERA. DESY-HERA 88-03, 1988.

[26] ZEUS collaboration. The ZEUS detector technical proposal. Technical Report DESY, 1986.

[27] ZEUS collaboration. The ZEUS detector status report. Technical Report DESY, 1993.

[28] J. Andruszków *et al.* Technical Report DESY 92-066, 1992.

[29] N. Harnew *et al. Nucl. Instr. and Meth.*, A(279):290, 1989.

[30] B. Foster *et al. Nucl. Phys.*, B (Proc. Suppl.) 32:181, 1993.

[31] B. Foster *et al. Nucl. Instr. and Meth.*, A 338:254, 1994.

[32] R. Hall-Wilton *et al.* The CTD tracking resolution. Technical Report ZEUS Note 99-024, 1999.

[33] ZEUS SRTD Group. The small angle rear tracking detector of ZEUS. Technical Report DESY 97-157, Aug. 1997.

[34] M. Derrick *et al.* (ZEUS Collaboration). *Z. für Physik C*, 69:207, 1996.

[35] A. Bamberger. *Nucl. Instr. and Meth.*, A(401):63, 1997.

[36] A. Andresen *et al. Nucl. Instr. and Meth.*, A 309:101, 1991.

[37] A. Caldwell *et al. Nucl. Instr. and Meth.*, A 321:356, 1992.

[38] A. Bernstein *et al. Nucl. Instr. and Meth.*, A 336:23, 1993.

[39] W. H. Smith *et al.* The ZEUS trigger system. Technical Report ZEUS Note 89-084, 1989.

[40] U. Schneekloth. *Recent HERA Results and Future Prospects*, volume DESY 98-060. 1998.

[41] In R. Klanner Eds. G. Ingelman, A. de Roeck, editor, *Proceedings of the Workshop "Future Physics at HERA", Vols. 1 and 2, DESY 1996.*

[42] ZEUS Collaboration. A micro vertex detector for ZEUS. Technical Report DESY-PRC-97-01, 1997.

[43] A. Garfagnini. *Nucl. Instrum Meth.*, A435:34, 1999.

[44] ZEUS Collaboration. A straw tube tracker for ZEUS. Technical Report DESY-PRC-97-01, 1997.

[45] G. Marchesini *et al. Comp. Phys. Comm.*, 67:465, 1992.

[46] M. Bengtsson and T. Sjöstrand. *Z. Phys. C*, 37:465, 1998.

[47] T. Sjöstrand. *Comp. Phys. Comm.*, 39:347, 1986.

[48] T. Sjöstrand and M. Bengtsson. *Comp. Phys. Comm.*, 43:367, 1987.

[49] L. Lönnblad. *Comp. Phys. Comm.*, 71:15, 1992.

[50] L. Lönnblad. *Z. für Phys.*, C65:285, 1995.

[51] K. Kwiatkowski, H. Spiesberger, and H.-J. Möhring. *Z. für Physik C*, 50:165, 1991.

[52] K. Kwiatkowski, H. Spiesberger, and H.-J. Möhring. *Comp. Phys. Comm.*, 69:155, 1992.

[53] K. Charchuła, G.A. Schuler, and H. Spiesberger. *Comp. Phys. Comm.*, 81:381, 1994.

[54] B. Andersson *et al. Phys. Rep.*, 97:31, 1983.

[55] S. Bentvelsen *et al.* In *Proceedings of the 1991 Workshop on Physics at HERA, DESY*, volume 1, page 23, 1992.

[56] A. Blondel F. Jacquet. In *Proceedings of the study for an ep facility for Europe, DESY 79/48*, page 379, 1979.

[57] H. Abramowicz *et al.* Neural network based electron identification in the ZEUS calorimeter. Technical Report DESY 95-054, 1995.

[58] T. Jackson R. Beale, editor. *Neural Computing: an introduction.* Adam Hilger, 1990.

[59] G. F. Hartner *et al.* VCTRAK (3.07/04): Offline output information. Technical Report ZEUS Note 97-064, 1997.

[60] N. H. Brook and I. Skillicorn. Measurement of azimuthal asymmetries in deep inelastic scattering. Technical Report ZEUS Note 99-050, 1999.

[61] J. Repond. Calculation of errors on correction factors. Technical Report ZEUS Note 98-015, 1998.

[62] J. Breitweg *et al.* (ZEUS Collaboration). Measurement of high Q^2 neutral current e^+p deep inelastic scattering cross sections at HERA. *Eur. Phys. J.*, C 11(3):427, 1999.

[63] K. Hagiwara *et al. Phys. Rev.*, D27(1):84, 1983.

[64] M. Ahmed and T. Gehrmann. *Phys. Lett.*, B465:297, 1999.

[65] S. Catani *et al. Nucl. Phys. B*, 406:187, 1993.

[66] M. H. Seymour. *Nucl. Phys. B*, 513:269, 1998.

[67] S. D. Ellis and D. E. Soper. *Phys. Rev. D*, 48:3160, 1993.

[68] J. Huth *et al.* In Singapore Ed. E. L. Berger, World Scientific, editor, *Proceedings of the 1990 DPF Summer Study on High Energy Physics, Snowmass, Colorado*, page 134, 1992.

[69] B. Mansoulié J. Schwindling. MLPfit: a tool for designing and using multi-layer perceptrons, version 1.40, january 2000, http://home.cern.ch/ schwind/mlpfit.html.

[70] K. Hornik *et al.* Multilayer feedforward networks are universal approximators. *Neural Networks*, vol. 2:359, 1989.

[71] D. W. Ruck *et al.* The multilayer perceptron as an approximation to a Bayes optimal discriminant function. *IEEE Transactions on Neural Networks*, vol. 1(4):296, 1990.

[72] E. Rodrigues. Quark and gluon jet identification in DIS Monte Carlo programs. Technical Report ZEUS Note 2001-041, 2001.

[73] ZEUS Collaboration. Measurements of jet substructure in neutral current deep inelastic scattering and determination of α_s at HERA. *presented at the International Europhysics Conference on High Energy Physics 2001 (EPS HEP 2001), Budapest, Hungary, Contributed Abstract No. 641*, 2001.

[74] ZEUS Collaboration. Measurements of jet substructure in inclusive and charm-induced dijet photoproduction at HERA. *ibid., Contributed Abstract No. 640*, 2001.

[75] ZEUS Collaboration. Measurements of jet substructure in charged current deep inelastic scattering at HERA. *ibid., Contributed Abstract No. 646*, 2001.

[76] C. Adloff *et al.* (H1 Collaboration). Measurement of internal jet structure in di-jet production in deep inelastic scattering at HERA. *Nucl. Phys.*, B545:3, 1999.

[77] Z. Fodor. How to see the differences between quark and gluon jets. *Physics Review D*, 41(5):1726, 1990.

[78] CERNLIB program manual. Technical Report CERN Program Library Short Writeups, 1996.

[79] H. Plothow-Besch. *Comp. Phys. Comm.*, 75:396, 1993.

[80] S. Catani and M. H. Seymour. *Nucl. Phys.*, B485:291, 1997.