Probing the Parton Evolution in DIS at low x_{BJ} using Jet Observables

by

Michael Riveline
Department of Physics, McGill University
Montreal, Quebec
December 1998

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements of the degree of Doctor of Philosophy

Abstract

The advent of the ep collider HERA at the DESY research center has opened a wide kinematical window for the study of the Deep Inelastic Scattering (DIS) processes $e^+p \rightarrow e^+X$. Recent measurements of the structure function F_2 of the proton have confirmed the fact that the Bjorken scaling which arises from the naive Quark-Parton Model is broken in the limit where the scaling variable, x-Bjorken, is small. It also revealed that F_2 grows exponentially with x-Bjorken at low x. The standard picture of parton evolution, based on the DGLAP equations, reproduces this steep rise if it is used with an appropriate set of F_2 parametrizations. However these parametrizations include "ad-hoc" assumptions on the non-perturbative behaviour of the structure functions and depend on a large set of phenomenological parameters.

The steep rise can also be reproduced by using a different perturbative treatment for the parton evolution. This treatment, based on the BFKL evolution equations, differs from the standard DGLAP picture by the lack of ordering in the transverse energy of the emitted partons. Unlike DGLAP, the BFKL picture reproduces the steep rise of F_2 without assuming any particular shape of the structure function in the non-perturbative domain.

In order to gain an insight on parton dynamics in DIS at low x and discriminate between these two pictures, jet observables are used as they are expected to be closely related to the hard scattering and depend only slightly on the hadronization effects. Two jet observables are studied here in more details: the azimuthal correlation between the two leading-order jets and the cross section of the forward jet production. The measured cross sections are corrected for detector effects and compared to several DIS Monte Carlo models and next-to-leading order simulations over a wide kinematic range. The results are evaluated in the light of the BFKL and the DGLAP pictures. The experimental results are compared to other models of parton evolution as well, like the Colour Dipole Model (CDM), the Linked Dipole Chain (LDC) and the resolved photon model in DIS.

Résumé

L'arrivée du collisionneur positron-proton HERA au centre de recherche DESY a ouvert de nouveaux horizons dans l'étude des processus de diffusion inélastique profonde (DIP) $e^+p \to e^+X$. Les mesures récentes de la fonction de structure F_2 du proton ont confirmé le fait que l'échelle de Bjorken, conséquence du modèle naif quark-parton, est brisée pour les petites valeurs de la variable d'échelle x-Bjorken. Ces expériences ont également révélé que F_2 croît exponentiellement avec x dans la limite où x est petit. Le modèle standard de l'évolution partonique, basé sur les équations DGLAP, reproduit cette croissance rapide s'il est employé avec les paramétrisations modernes des fonctions de structure. Cependant, ces paramétrisations forment des hypothèses sur le comportement des fonctions de structure dans les domaines cinématiques qui ne sont pas calculables selon la théorie des perturbations et dépendent par conséquent d'un grand nombre de paramètres phénoménologiques.

On peut également reproduire la croissance "aigue" de F_2 en utilisant un autre traitement perturbatif de l'évolution partonique. Celui-ci, basé sur l'équation d'évolution BFKL, diffère du modèle standard DGLAP par l'absence d'arrangement dans l'énergie transverse des partons émis durant la collision. Contrairement à DGLAP, BFKL décrit la croissance de F_2 sans faire d'hypothèse sur la forme que cette fonction de structure doit adopter dans les domaines non perturbatifs.

Des observables basés sur les "jets" sont utilisés pour explorer la dynamique partonique dans la région cinématique caractérisée par les petites valeurs de x. Ces observables sont censés décrire de façon précise la dynamique partonique puisque les effets de l'hadronisation y sont faibles. Deux observables de jets sont étudiées ici: la corrélation entre les deux jets du premier ordre et la production de jets vers l'avant. Les sections efficaces sont mesurées pour ces deux observables et comparées aux prédictions de différentes simulations après correction des effets du détecteur. Les résultats sont discutés à la lumière des différents modèles DGLAP et BFKL. D'autres modèles sont pris en considération, comme le Colour Dipole Model (CDM), le Linked Dipole Chain (LDC) et le modèle de photon résolu en DIP.

Contents

Acknowledgments

I would like to express my gratitude to my supervisor Professor François Corriveau for his guidance and his support all along these four years of PhD. I am also specially grateful to Doctor Jutta Hartmann for her precious help and her patience which contributed greatly to the completion of this work, as well as Doctors Chris Matthews and Gabriele Zacek for having introduced me to the ZEUS environment. I also thank Professors Doug Stairs and David Hanna for their constant encouragements and moral support during these years of PhD.

I would like to thank my fellow analysers, whose professionalism and expertise in the field has made the work much easier: Doctors Mariusz and Maciej Przybycien, Doctors Sabine Wölfle, Steve Magill, Tobias Haas and David Simmons and to our theorists colleagues Mark Wüsthoff, Erwin Mirkes, Leif Lönnblad, Gunnar Ingelmann, Jochen Bartels and Gavin Salam for their helpful advices and interesting discussions. Thanks also to the physics coordinators Tony Doyle, Costas Foudas, Jose Repond and Laurel Sinclair for many fruitful discussions and enlightening ideas.

I am grateful to the members of the calorimeter group for their support: John Labs, Jit Ning Lim, Guy Cases, Pat Saull and Antonio Pellegrino.

I am also grateful to Doctor Söhngen and to the DESY administrative staff for having helped me solving many technical problems pertaining to my stay in Hamburg.

Many thanks to all the members of the Canadian Cluster of the ZEUS experiment, whose friendly behavior made life enjoyable and who provided a healthy work atmosphere in DESY. And also thanks to my Canadian colleagues in McGill.

Finally, my deepest gratitude to my parents and my sister for their constant support and their belief in me.

Chapter 1

Introduction

Particle physics is the science of the basic building blocks and the fundamental interactions of nature. Considerations of symmetry have helped scientists to describe the various forces and particles which have been recorded all along the history of science in the framework of the *Standard Model*. The goal of this model is to reproduce a maximum number of observable phenomena, from the properties of sub-atomic particles to cosmological considerations, with a reduced number of theoretical predicates (describing the most empirical properties with the least number of parameters is the general aim of all sciences). So far, four forces account for all the interactions which occur in nature: the gravitationnal force, the electromagnetic force, the weak and the strong forces. These forces act on the fundamental constituents of matter through a *gauge field* carried by *gauge bosons*, which are particles of various masses and integer spins. The strength of the coupling and the spin of the *gauge boson* define the interaction. For instance, gravitation is carried by "gravitons" of spin 2 and is the weakest of all forces, although its macroscopic properties are the most dramatic of all. The rest of the particles participating in the interactions and then building all the known matter of the universe, at least in the current shape of the Standard Model, have half-integer spin and are classified into two groups: the *leptons* and the *hadrons*. The first group is mostly characterized by its sensitivity towards the electro-weak interaction and its inability to interact through the strong force. Particles like electrons and neutrinos are part of this group. The second group, the hadrons, can interact through the strong force and are in fact bound states of fundamental particles called quarks. The proton and the neutron are examples of hadrons. Stable hadrons and leptons, namely protons, neutrons and electrons constitute the atom, the building block of the long evolution process giving rise to our familiar environment.

Of all the interactions listed in the Standard Model, the strong force has the largest coupling strength and its description is therefore most challenging. As a matter of fact, the stronger the force, the weaker is our ability to observe and characterize the properties of the particles sharing this interaction, as the well-known perturbative treatment which helps finding out the amplitude of the interaction mechanism is only valid in the limit where the particles can be considered as "free". The strong force is carried by the spin 1 gluons, which have the intriguing property of coupling to each other through an additional quantum number, the colour charge. Together with the incoming of the theory of colour, the concept of "quark" came out as the fundamental object taking place in the strong interaction mechanism. The strong force is then described through the exchange of coloured gluons between coloured quarks in the framework of Quantum Chromodynamics (QCD). The most fundamental property of this theory is that the quarks are never observed as free particles but always in a bound state. This property, called confinement, is at the basis of all known models aiming at describing the behavior of quarks and gluons. However, in the limit of large momentum transfer, these particles can be described as "free" and therefore perturbative calculations can be applied. The quarks and gluons then evolve towards the final state hadrons, first through a perturbative mechanism controlled by the so-called "evolution equations", then within a phase called "hadronization" and which is not calculable through perturbative theory. The first, perturbative part of this evolution is the object of this study and will be described in detail in chapter 2.

The experimental study of the properties of QCD is only achievable through the observation of the final state particles, i.e. after the hadronization phase. As this is not calculable, one needs models to reproduce it. These models can also implement the various perturbative schemes which can be probed, according to a specific set of approximations, and are usually given in the form of event generator of Monte Carlo simulation. This will be described in more detail in chapter 3.

As the size of the effects which are probed is extremely small, an optical microscope would not be capable of detecting them, even with a very good resolution. Probing the parton dynamics of QCD requires very large energies and specific processes, like Deep Inelastic Scattering (DIS), which consists of a high momentum transfer collision between a probe, which is often a lepton, and a hadronic target. Deep inelastic

processes are characterized by two variables, which parametrize the parton density of the proton: the momentum transfer of the collision called Q^2, which is also the inverse square of the momentum of the exchanged boson, and the so-called "scaling variable", x-Bjorken or x_{Bj}[1], which can be seen as the fraction of the momentum of the parton participating in the interaction with respect to the momentum of the hadron (in the frame where the hadron is moving very fast so that its mass can be neglected). Since the early 70's, many experiments investigated these DIS processes, thereby proving the composite nature of the proton, but none of them could probe such a wide phase space window as the HERA accelerator. The ZEUS detector which must satisfy the requirements set by the difficult kinematic region to be probed, and the HERA accelerator will be described in chapter 4. The event selection, to target a specific phase space region, will be described in chapter 5.

The quarks and gluons are not observable directly in the detectors, as for instance are the electrons and the muons. However, their properties can be probed through the use of phenomenological objects called jets, defined as collimated streams of particles, related to the fundamental constituents which are present during the perturbative phase of the collision. These objects will be defined in chapter 6, and two methods to use them in order to determine the properties of the parton dynamics at low x will be presented in chapters 7 and 8.

[1]In the rest of this work, we will follow the standard notation and refer to x_{Bj} simply as x unless otherwise specified

Chapter 2

Basics of Perturbative QCD: Various Models of Parton Evolution

One of the main challenges of the experimental study of the strong interaction theory, Quantum Chromodynamics (QCD), is to relate the properties of the basic constituents of the theory, the quarks and the gluons, to the properties of the observed particles, the hadrons, which are bound states of these fundamental elements. A major step towards a global understanding of the evolution of the fundamental partons towards final state hadrons was the discovery by Altarelli and Parisi in 1973 of asymptotic freedom [1] and the possibility to use the short range interactions to apply perturbative calculations to the hadronic system. One of the most important phenomenological successes of this theory was the accurate description of the deep inelastic scattering of a lepton off a hadron and in particular the explanation of the breaking of simple Bjorken scaling. A large number of experimental tests in fixed target as well as collider experiments already confirmed most of the predictions made in various kinematic limits. However the HERA collider, built from 1988 to 1992, can now push further down these limits and explore phase space regions which are at the boundaries of the range of validity of the standard descriptions of the parton evolution mechanism. These new regions, close to the so-called "Regge" limit, can still be described in a perturbative way, but with different tools, and with phenomenological implications which are expected to differ from the standard ones, as will be explained in this chapter.

The main question of this work is whether this new perturbative domain can be

reached within the limits of HERA and whether the conclusions drawn from the measurement stand firm with respect to non-perturbative physics.

2.1 Fundamental Properties of the Strong Interaction

Before entering into the details of QCD and Deep Inelastic Scattering (DIS), we shall briefly review the fundamental predicates on which the theory is based. Although DIS provided the strongest experimental evidence for the existence of quarks and gluons, these fundamental constituents were part of the theory of the strong interaction before, but their interpretation in terms of real particles, rather than theoretical concepts and part of the SU(3) theory, was uncertain. Moreover, the main properties on which QCD lies, like the concepts of colour and confinement, were inferred without a deep understanding of the structure of the hadrons. Today, thanks to new experimental evidence and much theoretical progress, both in perturbative and non-perturbative QCD, the nature of the partons and their evolution into the colour singlet states called hadrons is much better understood. On the other hand, the understanding of confinement, for instance, has not really improved since its introduction in the mid-60's.

In this section, we will review the concept of confinement, as well as its theoretical basis, the notion of colour. Finally, we will see how, in spite of that property, asymptotic freedom enables theorists to perform perturbative calculations on partonic mechanisms.

2.1.1 Hadron Spectroscopy: the Concept of Colour

The basis of our understanding of the particles involved in strong interactions (the hadrons) comes from the results of hadron spectroscopy. Without any assumption about their internal structure, the accounting of the observed hadrons suggested a classification based on their mass and quantum numbers. This classification, known as the "eight-fold way" was performed in 1961 independently by Gell-Mann and Ne'eman [2]. The hadrons are divided into two groups, according to whether their

	Mesons		Baryons	
J^P	0^-	1^-	$\frac{1}{2}^+$	$\frac{3}{2}^+$
	$K^+(494)$	$K^{*+}(892)$	p(938)	$\Delta^{++}(1232)$
	$K^0(498)$	$K^{*0}(896)$	n(940)	$\Delta^+(1232)$
	$\pi^+(140)$	$\rho^+(768)$	$\Lambda(1116)$	$\Delta^0(1232)$
	$\pi^0(135)$	$\rho^0(768)$	$\Sigma^+(1189)$	$\Delta^-(1232)$
	$\pi^-(140)$	$\rho^-(768)$	$\Sigma^0(1193)$	$\Sigma^{+*}(1383)$
	$\bar{K}^0(498)$	$\bar{K}^{*0}(896)$	$\Sigma^-(1197)$	$\Sigma^{0*}(1384)$
	$K^-(494)$	$K^{*-}(892)$	$\Xi^0(1315)$	$\Sigma^{-*}(1387)$
	$\eta(549)$	$\omega(782)$	$\Xi^-(1321)$	$\Xi^{0*}(1532)$
	$\eta'(958)$	$\Phi(1019)$		$\Xi^{-*}(1535)$
				$\Omega^-(1672)$

Table 2.1: List of the light hadrons, mesons and baryons in singlet, octet and decuplet groups. The mass in GeV/c^2 is given in brackets.

spin (i.e. the "internal" angular momentum of the particle at rest) is an integer (the mesons) or half-integer (the baryons). Each one of these groups is subdivided into two sub-groups, once again according to the spin of the particles (this is for the "light particles", without orbital angular momentum), J^P (where J is the spin and P, the parity quantum number of the hadron, noted $+$ or $-$ for $P = +1$ and -1 respectively): 0^-, 1^- for the mesons and $\frac{1}{2}^+$, $\frac{3}{2}^+$ for the baryons. The fundamental key to interpret this classification is to realize that mesons come as singlet or octet in the 0^- and 1^- groups while hadrons are seen as singlet, octet or decuplet (see table 2.1 for a listing of the light hadrons). The important point is that **1**, **8** and **10** are dimensional representations of the SU(3) group of three-dimensional unitary matrices. This observation has lead Gell-Mann and Zweig [3] in 1964 to suggest the idea of a fundamental triplet of SU(3) which could explain the formation of all known hadrons, using the following combination rules:

$$\text{mesons} : q\bar{q} \equiv \mathbf{3} \otimes \bar{\mathbf{3}} \equiv \mathbf{1} \oplus \mathbf{8}$$
$$\text{baryons} : qqq \equiv \mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} \equiv \mathbf{1} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{10}.$$

This fundamental triplet consists of three spin 1/2 particles: the quarks. These quarks come in three flavors up (u), down (d) and strange (s). The puzzling fact is that, unlike all other particles they have fractional charge: u has $+\frac{2}{3}$, d and s have $-\frac{1}{3}$. Besides, although the strong interaction Hamiltonian obeys naturally the SU(3) sym-

metry, the origin of this symmetry is unknown and therefore, the physical meaning of its constituents was put to doubt until the early 70's and the first DIS experiments; quarks were then merely considered as a mnemonic tool which helped building up the various hadronic wave functions.

The SU(3) theory has however a drawback as it predicts the existence of the Δ^{++} and Ω^{-}, made respectively of three u and three s quarks. Considering the quantum numbers of each one of the quarks (spin and flavour) and knowing that these quarks are fermions, this violates the Pauli uncertainty principle which states that two fermions can not occupy the same quantum state (in this case two of the u quarks must have the same spin and therefore all their quantum numbers are identical). To solve this puzzle, Greenberg introduced in 1964, the concept of colour charge and the colour wave function [4]. Colour is an additional quantum number carried by the quark. There are three basic colours: red (r), green (g) and blue (b), corresponding to the three colour states in which the quark can be found. To these three colours correspond three anti colours $\bar{r}$, $\bar{g}$ and $\bar{b}$ which react with the colour states the same way as charges react with anti-charges. The vector of the colour exchange mechanism is a spin-1 particle: the gluon (which is therefore the vector of the strong interaction, similar to the photon in Quantum Electrodynamics, QED). Gluons themselves carry a colour charge and can therefore couple to each other, unlike photons in QED.

An additional assumption arises from the concept of colour: the hadrons observed in nature do not carry a colour charge, which comes to say that the hadron are always found in a colour singlet state. This hypothesis leads to the so-called colour confinement [5] and restricts the number of quark systems to only $q\bar{q}$ and qqq. From this comes the name QCD, Quantum Chromodynamics, that is the part of quantum mechanics which studies coloured objects and the mechanisms of colour exchange.

2.1.2 Confinement and Asymptotic Freedom

We saw how the hypothesis of confinement of quarks arises from the concept of colour: only colour singlet objects can be observed as free particles in nature so quarks, which are in a colour octet state, can only exist in a confined state, that is as part of a combination. This translates into the shape of the potential of the strong interaction: in QED, the Coulomb potential decreases as 1/r to infinity (at leading order). By

contrast, in QCD, the hypothesis that the strength of the interaction rises at large distances generates a non-Coulomb potential which rises to infinity at large distances. The QCD potential can be written as [6]:

$$V(r) = -constant \cdot \alpha_s/r + V_1(r), \tag{2.1}$$

Here, α_s is the coupling parameter of the strong interaction, which yields the strength of the interaction (similar to α_{em} in QED); $V_1(r)$ is the non-Coulomb part of the potential. It is assumed that this property of the strong interaction arises from the fact that, contrary to QED, the vectors of the interaction, the gluons, are colour charged and can therefore couple to each other, increasing the effective coupling strength at long distance [5].

In this context, no perturbative calculation describing the behavior of quarks and gluons was possible, as it would require to expand the QCD Lagrangian in series of α_s, and the magnitude of the coupling term would be such that the series would diverge. The real birth of QCD happened in 1973, with the discovery of asymptotic freedom [1] which enabled perturbative calculations to be used to describe the quarks and the gluons. This is the property which states that the strength of the interaction (the magnitude of the coupling constant) decreases at very short distances (around 0.1 fm). In this limit, quarks and gluons can be treated as free particles. Some attempts of explanation for this effect exist: one of them deals with the fact that, similarly to the electron in QED, polarizing the vacuum and creating new electrical charges, the gluon creates in the surrounding vacuum new colour charges. The key point is that these colour charges couple to each other creating colour distributions which overlap at short distances. In this case the interaction strength is smaller than from point charges, so that the effective strength tends to zero and the colour charges can be considered as free particles. Another explanation considers the relative diamagnetic and paramagnetic contributions of the gluons, which couple to the magnetic field of the polarized vacuum through their spin [7].

2.1.3 Renormalization Group Equation and Running Coupling Constant

The two properties of confinement and asymptotic freedom reflect themselves in the behavior of the coupling strength of the strong interaction potential: α_s. They ac-

tually translate into the fact that α_s varies as a function of the scale of the process Q (α_s is said to "run" with the scale Q). This scale is defined as being the largest dimensional parameter of the process, so that all the dimensionless parameters are defined with respect to it.

This is however not the only type of scale which can be defined in quantum field theory: after having applied perturbative expansion to a large order, the perturbative series starts to diverge. In other words, at higher order, the series diverges. Because of the small wavelength of the particles participating in the interaction in this limit, this behavior is called ultra-violet divergence. In order to suppress these divergences, a cut-off has to be introduced in the form of a scale called the renormalization scale μ. As μ is an arbitrary parameter, any physical process of the process should be independent of them. It can be shown (see for instance [7]) that the scale dependence of a variable enters through the dependence of the coupling constant α_s on the scale Q. The renormalization group equation, which determines the running of the coupling constant is a consequence of that and can be written as [7]:

$$Q\frac{\partial \alpha_s}{\partial Q} = \beta(\alpha_s). \tag{2.2}$$

The $\beta(\alpha_s)$ is referred to as the "beta" function and can be calculated in a perturbative manner. By expanding it in perturbative series, the following equation is obtained:

$$Q\frac{\partial \alpha_s}{\partial Q} = -b_0\alpha_s^2 - b_1\alpha_s^3 - b_2\alpha_s^4 - \cdots, \tag{2.3}$$

where b_0, b_1, b_2, ... are known terms of the perturbative series (the perturbative expansion is commonly expressed as a function of β_i terms with $\beta_0 = 4\pi b_0$, $\beta_1 = 16\pi^2 b_1$, $\beta_2 = 64\pi^3 b_2$,...). If only the first term of this expansion is taken into account, the value of $\alpha_s(Q)$ can be expressed as a function of $\alpha_s(\mu)$, at a fixed reference scale (which can be for instance the renormalization scale):

$$\alpha_s(Q) = \frac{\alpha_s(\mu)}{1 + \alpha_s(\mu) b_0 ln(Q^2/\mu^2)}. \tag{2.4}$$

The coefficient b_0 is equal to $\frac{1}{4\pi}\left(11 - \frac{2}{3}n_f\right)$, were n_f is the number of quark flavours in the theory. If n_f is smaller than 17 (which is the case in the SU(3) theory), the coefficient b_0 is positive and therefore α_s decreases when Q becomes large and one recovers the property of asymptotic freedom. Should the number of flavours be larger

than 17, α_s would increase at large Q and QCD would not be an asymptotically free theory. By contrast, the coupling strength in QED can be expressed as:

$$\alpha_{em}(Q) = \frac{\alpha_{em}}{1 - \frac{\alpha_{em}}{12\pi} ln \frac{Q^2}{m_e^2}}. \tag{2.5}$$

There, the coefficient in the denominator is negative and $\alpha_{em}(Q)$ grows as Q grows and unlike QCD, QED is not an asymptotically free theory.

2.2 DIS: the Success of the DGLAP Equations

The best laboratory to study the theory of the strong interaction is the proton itself. Provided that small enough distances inside the proton can be probed, its fundamental constituents can even be studied. This is the concept behind deep inelastic scattering (DIS), whereby a high-energy lepton collides against a hadron target. Provided that the momentum of the exchanged boson, Q is larger than the inverse size of the hadron, the inner structure of the target can be revealed, and in particular the momentum distribution of its constituents. These momentum distributions can be parametrized in the form of structure functions and compared to theoretical predictions.

The first DIS experiments took place in SLAC (Stanford Linear Accelerator Center) in the late 1960's [8] and revealed a number of discoveries, in particular the invariance of the structure function of the proton with respect to the momentum transfer between the scattered electron and the target proton. This structure function was found to be only dependent on the momentum fraction of the struck constituent of the proton x. This "scale invariance" was interpreted as an evidence that the scattering involved point-like particles inside the proton: the partons. In 1969, following a model created by Feynman in 1968 (the parton model [9]), Bjorken and Paschos identified the partons with the quarks described in the previous chapter [10].

The breaking of this scale invariance, also called "Bjorken scaling", at low momentum transfer Q^2 was observed in the muon-proton scattering, BCDMS [11] and NMC [12] at CERN and then in ep scattering at HERA. It was the first evidence that the simple parton model was a crude approximation and that more complex diagrams (the NLO

corrections) had to be considered.

Several DIS experiments after the pioneering SLAC experiment took place in the CERN and FNAL institutes. These experiments could probe the structure function of the proton over an extended kinematic range, but none of them could match the HERA experiment, the world's first ep collider, where the center-of-mass energy of around 300 GeV enables to reach new kinematic ranges and probe new kinematic regimes.

2.2.1 The Quark-Parton Model (QPM)

Basic DIS Variables

The simplest and first order Feynman diagram to be considered in the deeply inelastic scattering of a lepton off a proton is shown in figure 2.1. There, the incoming lepton (of momentum k) exchanges a virtual boson (momentum q) with the target hadron (momentum p), which disintegrates into a number of particles. If k' is the momentum of the scattered lepton, p' the total momentum of the proton system after the collision, s the square of the total center of mass energy of the collision and t the squared energy transfer of the collision, the system is kinematically defined by the set of equations:

$$q^2 = (k - k')^2 = (p - p')^2 = t, \tag{2.6}$$

$$s = (k + p)^2 \simeq 4E_p E. \tag{2.7}$$

Here E is the energy of the incoming lepton. At leading order, the virtual boson scatters off one of the three valence quarks of the proton. In this picture, called the Quark-Parton Model, the total cross section of this process can be described by two independent variables. By convenience, Lorentz invariant parameters have been chosen:

$$Q^2 = -q^2, \tag{2.8}$$

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

Two other variables are usually introduced:

$$\nu = \frac{p \cdot q}{m_p}, \tag{2.10}$$

$$y = \frac{q \cdot p}{k \cdot p}. \tag{2.11}$$

If E' is the energy of the outgoing lepton in the rest frame of the proton, ν and y can be shown to be respectively the total and relative energy transfer between the lepton and the target hadron:

$$\nu = E' - E, \tag{2.12}$$

$$y = \frac{E - E'}{E}. \tag{2.13}$$

Using the Mandelstam variables, one can show that x, y and Q^2 are not independent but related to the square of the center of mass energy s by the relation:

$$Q^2 = sxy. \tag{2.14}$$

The important feature of the QPM is that the partons are supposed to act as free particles, without interaction. Besides, if the proton moves very fast in the center of mass frame, the transverse momentum of the partons inside the proton can be neglected with respect to their longitudinal momentum. In this context, the variable x can be shown to be equal to the momentum fraction of proton carried by the struck quark. In the experiment, the variables x, y and Q^2 can be reconstructed from the energies and angles of the scattered lepton and the hadronic system. Three reconstruction methods exist, which are described in the appendix A: the electron method, which relies only on the scattered electron (positron) information, the Jacquet-Blondel, which takes into account only the hadronic energy and average angle and the double angle method which combines both and relies on the angles of the scattered lepton and the hadronic system. According to the equation 2.8, Q^2 is the negative square of the momentum transfer and determines the scale of the process. As a matter of fact, the resolution power of the photon is given by its wavelength λ: the smaller λ, the smaller is the distance that the photon can probe within the proton. Through Heisenberg uncertainty principle, it can be shown through the De Broglie wavelength, that:

$$\lambda = \frac{1}{\mathbf{q}} = \frac{1}{\sqrt{(E - E')^2 + Q^2}} = \frac{1}{\sqrt{\nu^2 + Q^2}} \simeq \frac{2m_p x}{Q^2}. \tag{2.15}$$

This shows that the larger Q^2 is, the smaller is the wavelength of the photon and therefore the larger is the resolution power of the exchanged boson.

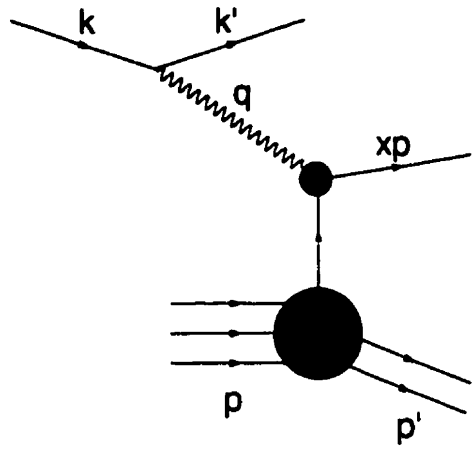

Figure 2.1: *Deep inelastic scattering between the virtual photon and the proton in the Quark Parton Model.*

DIS Cross Section and Structure Function

The cross section of the DIS process can be shown (for instance in [7] or [13]) to depend on two terms, one related to the electromagnetic vertex, calculable to all orders in Quantum Electrodynamics (QED) and another one, at the proton vertex, which can not be fully calculated in perturbative QCD (pQCD). In the case of a virtual photon exchange, the cross section can be parametrized the following way [1]:

$$\frac{d\sigma}{dxdQ^2} = \frac{4\pi\alpha_e^2}{xQ^2}[(1-y)F_2(x,Q^2) + y^2 x F_1(x,Q^2)], \tag{2.16}$$

where F_1 and F_2 are two independent structure functions, parametrizing the momentum distribution of the partons inside the proton. σ is sometimes written as a function of a longitudinal structure function, featuring the absorption by the proton of a longitudinally polarized virtual photon: $F_L = F_2 - 2xF_1$.

$$\frac{d\sigma}{dxdQ^2} = \frac{4\pi\alpha_e^2}{xQ^2}[(1-y+\frac{y^2}{2})F_2(x,Q^2) - \frac{y^2}{2}F_L(x,Q^2)]. \tag{2.17}$$

F_L is often neglected with respect to F_2. Indeed it vanishes in the QPM, due to the Callan-Gross relation [14] (which state that spin 1/2 quarks cannot absorb longitu-

[1] If Z or $W^{\pm}$ are involved in the process, a third terms appears which takes into account the amount of parity violation of the system.

dinally polarized photons):

$$F_2 \equiv 2xF_1. \tag{2.18}$$

The structure function F_2 cannot be calculated from the first principles, as it receives contributions from the long-range, non-perturbative part of the QCD Lagrangian, but it can be parametrized as a function of the contributions from the various quark flavours:

$$F_2 = x\sum_i e_i^2 f_i(x, Q^2), \tag{2.19}$$

where e_i is the charge of the quark of flavour i and f_i is the distribution of probability to find the quark i with the momentum fraction x. In the QPM, the partonic cross section $\hat{\sigma}$ can be similarly parametrized and a partonic structure function $\hat{F}_2$ can be defined. The amplitude of the collision: $e(k) + q(p_q) \rightarrow e(k') + q(p'_q)$ can be computed and $\hat{F}_2$ can be derived (see [7, 13]):

$$\hat{F}_2 = xe_q^2\delta(x - \epsilon), \tag{2.20}$$

where ϵ is the momentum fraction carried by the struck quark. This is conform to the idea that the photon collides against point like constituents of the proton (of fractional charge e_q) and therefore, at a given x, F_2 is independent of Q^2, which can then be written as:

$$F_2(x, Q^2) \rightarrow F_2(x). \tag{2.21}$$

The measurement of this structure function has been performed on an extended range of x and Q^2 by several experiments and in particular the ZEUS experiment [15] at HERA. The result of the F_2 measurement for different bins of x is shown in figure 2.2. The trend expected by the Bjorken scaling is seen at high Q^2 and high x, where the F_2 distribution is flat with respect to Q^2, in different bins of x. At lower x, however, systematic deviations from this scaling can be observed. A closer look at the plot would show that the amount of deviation from the Bjorken scaling is proportional to $\ln Q^2$. The scaling is said to be *broken* in logarithms of Q^2. One can also show that in the small-x limit, F_2 follows a power law as a function of x, such as: $F_2 \sim x^{-\lambda}$, where $\lambda \sim 0.3$. This behavior can not be explained within the limits of the QPM.

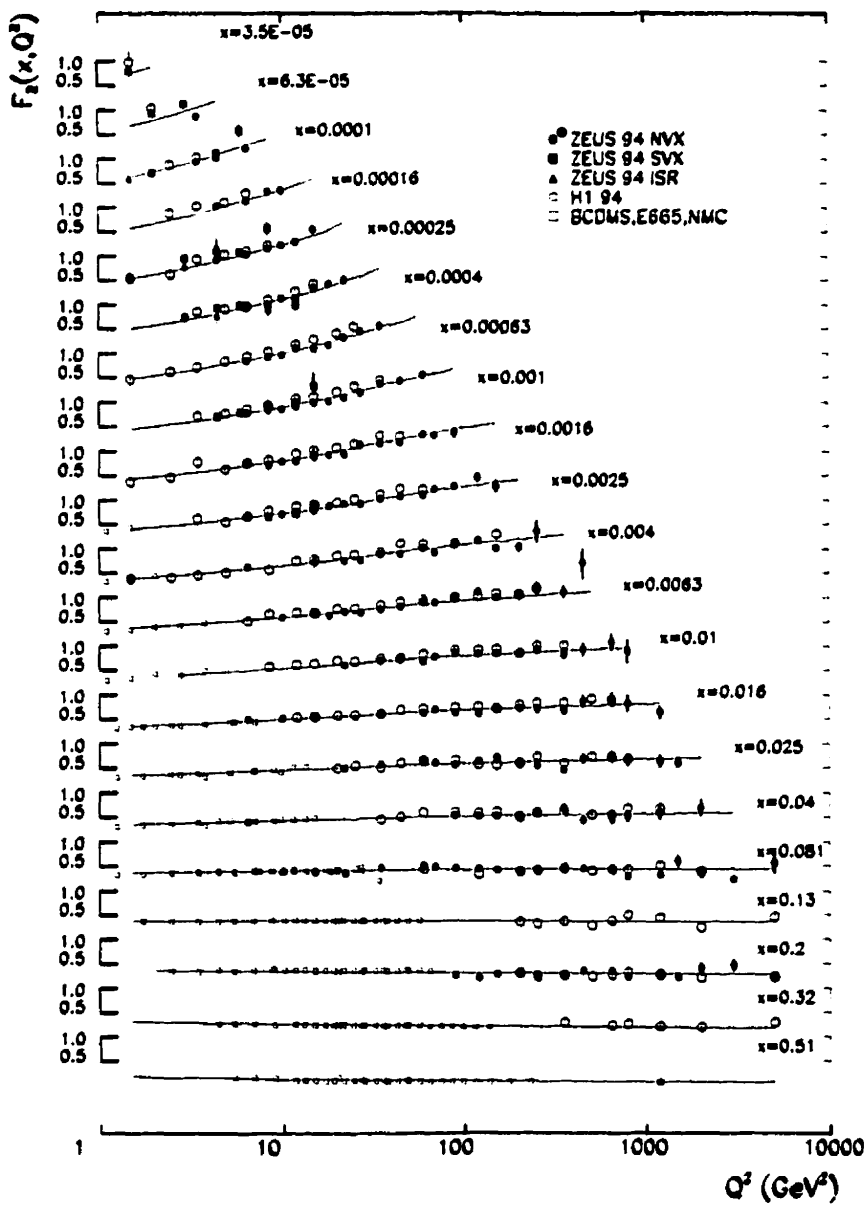

Figure 2.2: *Measurement of the structure function of the proton F_2 as a function of Q^2 in various bins of x. Together with the ZEUS measurement are presented H1, NMC, E665 and BCDMS measurements.*

2.2.2 Scaling Violation and Factorization Theorem

On top of the basic hypotheses of QCD, the QPM makes assumptions on the dynamic of the hadron collision, namely the absence of transverse momentum for the partons inside the proton and the absence of interactions between these partons. In the QPM, the proton is made of three valence quarks (it is said to be completely valence-like). This is however only true at high Q^2, where the longitudinal momentum of each one of the quarks within the proton is much larger than their transverse momentum (the proton is then said to be moving in the infinite momentum frame), so that their interaction can be neglected. At lower Q^2 and lower x, the quark can emit a gluon, which can in turn change the quark's transverse momentum. The photon can then couple directly to the emitted gluon.

Several additional Feynman diagrams have to be considered to take into account these high-order effects. The leading-order diagrams are shown in figure 2.3 and reflect the possible modes of coupling between the virtual photon and the partons within the proton; the photon can couple directly to the quark before (or after) the gluon emission: this diagram is called QCD Compton or QCDC (by reference to the "QED Compton" process where the photon couples to the lepton and is then re-emitted). The photon can also couple directly to the incoming gluon, giving rise to a quark-antiquark pair, through the Boson Gluon Fusion (BGF) mechanism. These diagrams show that at low x, the parton which is probed is not a fundamental constituent, but it has a structure related to its history (which can start within the proton itself). These fluctuations, which form the real history of the probed parton are large at small x; here x cannot refer any more to the momentum fraction of the original constituent, but to that of the propagator involved in the photon coupling. The smaller x is, the larger is the probability that the propagator parton arises from the fluctuation of another parton, rather than from the primordial constituent of the proton). Because Q^2 determines the resolution power of the process (remember that $\sqrt{Q^2}$ is proportional to the inverse wavelength of the virtual photon), as Q^2 becomes larger, so does the number of partons resulting from these fluctuations which can be observed. Therefore the effective density of partons within the proton, and hence the magnitude of the structure function F_2, will rise with the resolution power Q^2 at small x. This is the intuitive idea behind the concept of scaling violation.

In order to quantify this effect, the cross sections of each one of these processes, σ,

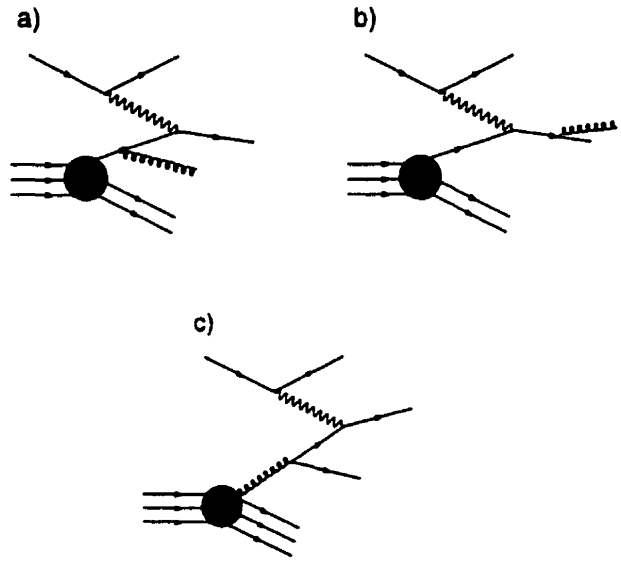

Figure 2.3: *Three modes of parton-photon coupling at leading order in α_s:a) and b) QCD Compton, c) Boson Gluon Fusion.*

have to be evaluated. But at this point, one faces two problems: first, the structure functions of the proton can not be calculated from pQCD, because the partons are confined within the proton and part of the observed cross section then comes from long-range, non-calculable processes. On the other hand, the calculation of the partonic cross section (when considering only the collision between the original parton and the virtual photon; for instance $\gamma^* + q \rightarrow g + q$ in the diagram 2.3 a)) leads to logarithmic divergences. Several techniques exist to get rid of these divergencies (see for instance [7, 13]), involving the use of cut-off or dimensional regularization (changing the number of space-time dimensions in order to regularize the divergent integral). These divergences occur when one of the partons of figure 2.3 is either soft ($k = 0$) or collinear to its parent parton ($k_T = 0$). In both cases the emitted parton has a long wavelength (this is why they are referred to as infra-red divergences) and corresponds to the long-range physics which has been described in the previous section, namely the processes where α_s is large and pQCD calculations cannot be applied any more.

The important result which has been obtained is that both problems, divergencies of the partonic cross section and impossibility to derive F_2 from the first principles, are

due to the same cause, the failure of applying perturbative QCD outside the scope of asymptotic freedom, that is to say, outside the limits of short-distance interactions. To solve this puzzle, one introduces a non-calculable ("bare") parton distribution, at a certain scale μ_F (the "factorization" scale): $q(x, \mu_F^2)$, which is related to the probability of finding a parton with momentum fraction x within the proton. This distribution is convoluted with the partonic cross section $\hat{\sigma}(x, \mu_F^2)$, according to the factorization formula:

$$\sigma_{DIS} = \sum_i q(x, \mu_F^2) \otimes \hat{\sigma}(x, \mu_F^2), \tag{2.22}$$

where i runs over the type of partons (gluons and various quarks flavours). The factorization theorem states that this formula is true for any kind of processes and that $q(x, \mu_F^2)$ is universal, i.e. it does not depend on the type of collision or on the diagram which is considered.

The introduction of the scale μ_F enables the solution of the divergencies the following way:

- If the transverse momentum k_T of the emitted parton is larger than the factorization scale μ_F (typically 1-2 GeV), it is included in the calculation of the partonic cross section $\hat{\sigma}$.
- if $k_T < \mu_F$, then the parton is "absorbed" into the bare parton distribution.

In practice, the exact determination of the parton distribution $q(x, \mu_F^2)$ requires the integration over a large number of diagrams, which can be summarized in the gluon ladder of figure 2.4. This involves integrations over the transverse momenta k_T of the emitted partons along the gluon ladder shown in figure 2.4, so that the real parton distribution is a combination of many distributions depending on the transverse momenta of the emitted partons:

$$q(x, \mu_F^2) = \int^{\mu_F^2} \frac{dk_{T,1}^2}{k_{T,1}^2} f(x, k_{T,1}) \int^{\mu_F^2} \frac{dk_{T,2}^2}{k_{T,2}^2} f(x, k_{T,2}) \cdots \int^{\mu_F^2} \frac{dk_{T,n}^2}{k_{T,n}^2} f(x, k_{T,n}). \tag{2.23}$$

Here $f(x, k_{T,n})$ is the parton distribution unintegrated over the gluon ladder (related to the probability to emit a parton with momentum fraction x and transverse momentum $k_{T,n}$).

By requiring the parton emissions to be strongly ordered, so that:

$$Q^2 \gg k_{T,n}^2 \gg \cdots \gg k_{T,2}^2 \gg k_{T,1}^2, \tag{2.24}$$

one can reorder the integrals and perform the integration over the n "rungs" of the gluon ladder (see [16]). Only in this case, equation 2.22 can be solved numerically while the full dependence over each of the transverse momenta of the emitted gluons is lost.

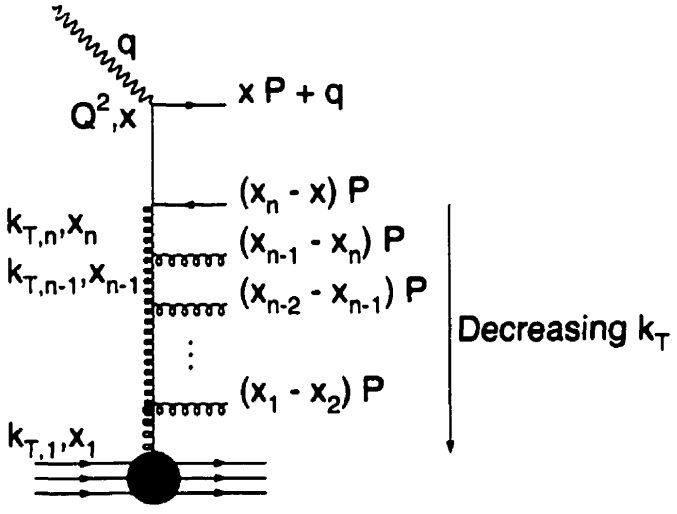

Figure 2.4: *Gluon ladder representing the sum of all the Feynman diagrams contributing to the DIS cross section. The momenta of all the gluon rungs are written to the right of the gluon line. From energy-momentum conservation, we obtain the relation* $x < x_n < \cdots < x_1$.

2.2.3 The DGLAP Equations and the Leading Log Approximation

From equations 2.22 and 2.16, one can in principle obtain an expression for the structure function of the proton F_2. But the determination of the exact value of F_2 is still bound to uncertainties: first, there is a non-perturbative, that is a non-calculable, part in this expression corresponding to the contributions from the long-distance part of the strong interaction to the total DIS cross section. Then, the expression involves an arbitrary scale μ_F, which delimits the relative contributions of perturbative and non-perturbative physics. Since this scale is not "physical", the total cross section should

not depend upon it. Finally, as we saw in the previous paragraph, some approximations are made which restrict the phase space available for the parton emissions in order to enable numerical estimations.

As in the case of the strong coupling constant α_s, whose scale independence requirement lead to the renormalization equation, a similar type of equation can be derived for the total DIS cross section (for convenience, the logarithmic derivative is taken):

$$\frac{d\sigma}{dln\mu_F} = \frac{d}{dln\mu_F}\sum_i q_i(x,\mu_F^2) \otimes \hat{\sigma}_i(x,\mu_F^2) = 0. \tag{2.25}$$

One can expand this equation:

$$\frac{d\sigma}{dln\mu_F} = \sum_i \left(\frac{d}{dln\mu_F} q_i(x,\mu_F^2) \otimes \hat{\sigma}_i(x,\mu_F^2) + q_i(x,\mu_F^2) \otimes \frac{d}{dln\mu_F}\hat{\sigma}_i(x,\mu_F^2) \right) = 0. \tag{2.26}$$

And it can be rewritten as [7]:

$$\frac{d}{dln\mu_F} q_i(x,\mu_F^2) = \frac{\alpha_s}{2\pi}\sum_j \int_x^1 \frac{d\epsilon}{\epsilon} P_{ij}(\frac{x}{\epsilon}, \alpha_s(\mu_F^2)) q_j(\epsilon,\mu_F^2). \tag{2.27}$$

The term P_{ij} is called the *splitting function* and corresponds to the probability for the parton i with momentum ϵP to branch to emit the parton j, with a fractional momentum x (smaller than ϵ, in order to conserve energy and momentum). There are four types of splitting functions: P_{qq}, P_{gq}, P_{qg} and P_{gg} (see figure 2.5). They can be expanded in perturbation series:

$$P_{ij}(\frac{x}{\epsilon}, \alpha_s(\mu_F^2)) = \delta_{ij} P_{ij}^0(\frac{x}{\epsilon}) + \frac{\alpha_s}{2\pi} P_{ij}^1(\frac{x}{\epsilon}) + \left(\frac{\alpha_s}{2\pi}\right)^2 P_{ij}^2(\frac{x}{\epsilon}) + \cdots \tag{2.28}$$

The leading order splitting functions have been derived (for instance in [16]) and are given below for the four type of branchings: $q \to qg$, $g \to gq$ and $g \to gg$ and $g \to q\bar{q}$

$$P_{qq}^0(x) = \frac{4}{3}\left(\frac{1+x^2}{1-x}\right), \tag{2.29}$$

$$P_{gq}^0(x) = \frac{4}{3}\left(\frac{1+(1-x)^2}{x}\right), \tag{2.30}$$

$$P_{qg}^0(x) = \frac{1}{2}(x^2+(1-x)^2), \tag{2.31}$$

$$P_{gg}^0(x) = 6\left(\frac{x}{1-x} + \frac{1-x}{x}\right) + x(1-x). \tag{2.32}$$

This leads to the famous leading order Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (or DGLAP) equations [17], where only the first order terms of the splitting function are taken into account:

$$\frac{dq_i(x,\mu_F^2)}{dln\mu_F} = \frac{\alpha_s}{2\pi}\sum_j \int_x^1 \frac{d\epsilon}{\epsilon} P_{ij}^0(\frac{x}{\epsilon}) q_j(\epsilon,\mu_F^2) \tag{2.33}$$

The solution to this equation enables to obtain the parton distribution at any scale

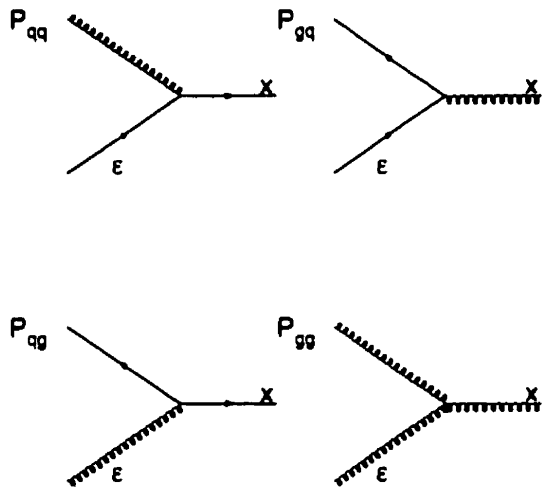

Figure 2.5: *Splitting functions corresponding to the probability, for a parton with momentum ϵP, to split into 2 partons, one having a momentum xP and the other $(\epsilon - x)P$.*

Q^2, if it is know at a given scale Q_0^2 ($Q_0^2 < Q^2$).

The DGLAP equations have an interpretation in terms of probability: by increasing the virtuality of the parton i by an amount of $ln\mu_F$, there is a certain probability to resolve its parent parton j, carrying a momentum fraction ϵ (that is $\frac{x}{\epsilon}$ larger momentum than i) and with a smaller transverse momentum (see figure 2.6).

In practice, all these approximations lead to retain, in the perturbative expansion of the parton density $q(x,\mu_F^2)$, only those terms with a leading $(\alpha_s ln Q^2)^n$. Therefore,

this method is called the "Leading Logarithm Approximation" or LLA(Q^2).

Thus, this picture predicts that each parton evolves independently of all the others in increasing virtuality and decreasing momentum fraction. As we will see in the following, these approximations are valid only in a certain kinematic range and one of the goals of the present study is to determine this range of validity.

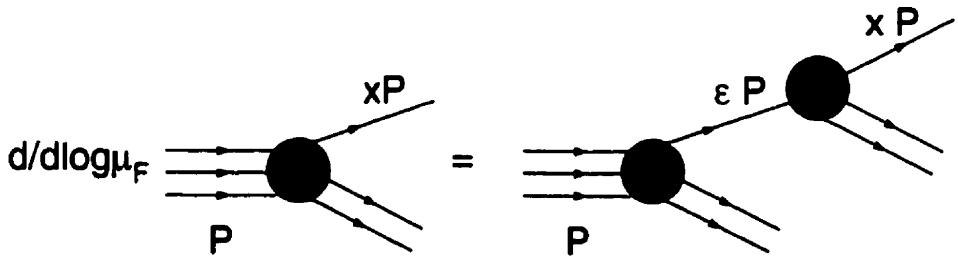

Figure 2.6: *Schematic representation of the DGLAP equations in terms of Feynman diagram: increasing the virtuality of the parton with momentum fraction x enables to resolve its parent parton with momentum fraction ϵ.*

2.3 The Low x Region: High Energy Limit and BFKL Equation

2.3.1 Non Linear Phenomena: Shadowing and Hot Spots

In the previous section, the partonic mechanism of DIS has been pictured as an independent evolution of a single parton: inside the proton, the parton "on-mass-shell" (this means that its invariant mass is equal to its rest mass) evolves towards the collision and the coupling with the photon by successively emitting daughter partons along a gluon ladder, losing each time a fraction of its total momentum. A look at the fundamental parameters of the evolution equation, namely the splitting functions (equations 2.28 to 2.31), shows that at low x, singularities appear:

$$P_{qq} \to \frac{4}{3},\ P_{qg} \to \frac{1}{2},\ P_{gq} \to \frac{4}{3}\frac{1}{x},\ P_{gg} \to 6\frac{1}{x}. \tag{2.34}$$

The branchings which give rise to a gluon are singular, leading to divergent gluon densities at low x. In [18], Mueller describes the dynamic of the low x processes in

the following terms: "At very small values of x, the number densities [of partons] obtained may become large enough that the quanta overlap spatially, in which case one expects scattering and annihilation to occur as well as evolution". In this context, the approximations which lead to a description of the parton dynamics using linear equations are not valid any more and the gluons can recombine and annihilate each other. At some point the gluon density might therefore saturate and evolution cannot occur any more. This process is known as shadowing [18, 19] and leads to a flattening of the structure functions at very low x. This happens when the transverse area occupied by the gluon[2] is comparable to the size of a nucleon πR^2 (R $\sim$ 1 fm). If the gluon density inside the proton occupies a transverse area of the size of a parton (R $\sim$ 0.4 fm), a different scenario is expected, which is called "Hot Spot" [20]. However all these effects should happen for low values of Q^2, where the uncertainty due to non-perturbative physics is expected to be large and the measurements of the shadowing and Hot Spot might be difficult to perform at HERA.

2.3.2 Double Leading Log Approximation and Description of the Low x Region by the Standard Structure Functions

In the previous section, we saw how, intuitively, the rise of the structure function at low x can be understood by taking into account the history of the parton involved in the coupling to the photon, that is, in practice, looking at higher order Feynman diagrams. This leads to a parton distribution which evolves as $(\alpha_s ln(Q^2/Q_0^2))^n$, where Q_0^2 is the starting scale, at which the parton distribution is known. In practice, this means that, in the perturbative expansion of the parton density in terms of lnQ^2 and $ln\frac{1}{x}$, only the terms with a leading lnQ^2 are retained, neglecting the terms involving $ln\frac{1}{x}$ and $lnQ^2 ln\frac{1}{x}$. In order to improve the description of the parton densities at low x and include the $ln\frac{1}{x}$ terms in the distribution, another approximation is performed when solving the DGLAP equations, which consists in retaining the most singular $ln\frac{1}{x}$ terms, when they are accompanied by lnQ^2. This corresponds to an additional requirement on top of the strong ordering of equation 2.24: the strong ordering in the momentum fraction of the emitted partons:

$$x \ll x_n \ll \cdots \ll x_1. \tag{2.35}$$

[2] If the photon probes the proton at a scale Q^2, it can resolve a transverse size of $1/Q$ and therefore, the transverse area goes as $xg(x,Q^2)\pi/Q^2$, where $g(x,Q^2)$ is the gluon density inside the proton

This method is called the Double Leading Log Approximation [21] (DLLA) as it is effectively summing up terms with large logarithms of Q^2 accompanied by large logarithms of $\frac{1}{x}$. It is only valid in the limit $x \to 0$ and $Q^2 \to \infty$. As the gluons dominate the parton densities within the proton in this low-x region, only the gluon distribution is relevant here. It takes the following shape:

$$g(x,Q^2) = \frac{1}{x} \exp\left\{\sqrt{\frac{12}{\pi b} \ln\left(\frac{\ln Q^2}{\ln Q_0^2}\right) \ln\left(\frac{1}{x}\right)}\right\}. \tag{2.36}$$

From this equation, we can see that, although they predict a Q^2 dependence for F_2, neither the Leading Logarithm Approximation, nor the DLLA have the kind of singular behavior in $x^{-\lambda}$ which can be inferred from the data.

However, the current parametrizations of F_2, using both leading and double-leading logarithm approximations, manage to describe the data. To achieve this agreement, the common procedure is to parametrize the structure function at a starting scale Q_0^2 (this gives the input parametrization, as a function of x) and then evolve it according to the DGLAP equations. The most popular and commonly used parametrizations use the following methods to reproduce the singular behavior of F_2 in the data:

- The CTEQ [22] (Coordinated Theoretical and Experimental Project on QCD) and the MRS [23] (Martin-Roberts-Stirling) groups use singular input, of the form $x^{-\lambda}$ at a large enough Q_0^2: $Q_0^2 = 2.56$ GeV^2 for CTEQ and $Q_0^2 = 4$ GeV^2 for the MRS parametrizations. Both parametrizations look like:

  $$xq_i(x,Q_0^2) = A_i x^{-\lambda_i}(1-x)^{\eta_i} P(x,i). \tag{2.37}$$

  Here i is the quark flavour or gluon. $P(x,i) = (1 + \epsilon\sqrt{x} + \gamma_i x)$ for MRS and $(1+\gamma_i x^{\epsilon_i})$ for CTEQ. The various parameters, A_i, γ_i, ϵ_i, λ_i, η_i must satisfy some basic requirements like flavour sum rules and other momentum sum rules (see for instance [24] for the various criteria). The other free parameters are derived from comparison to data at LEP or at fixed target experiments. The main problems with these methods is the big dependence on the various parameters: as the inputs are singular, a small change in the phenomenological parameters might result in dramatic discrepancies.

- The GRV (Glück-Reya-Vogt) group [25] uses flat input at very small starting scale ($Q_0^2 \sim 0.3$ GeV^2): in this picture, the proton is constituted by valence quarks at low Q^2, the gluons and sea quark being generated dynamically

through the DGLAP equations. The input parametrizations are actually the MRS "valence-like" parametrizations (at high Q_0^2) evolved backwards towards a smaller scale:

$$xq_i(x, Q_0^2) = A_i x^{\alpha_i}(1-x)^{\beta_i}. \tag{2.38}$$

Here again, i runs over the type of parton chosen and the various phenomenological parameters are derived from fits to the data. This method is much less sensitive to the value of the input parameters, but it uses pQCD (the DGLAP equations) at very small scales, where non-perturbative effects can be large and contributions from non-linear phenomena (see previous paragraph) can be important.

In 1994, the ZEUS collaboration measured the F_2 of the proton and compared it to the various parametrizations [15]. The result of this comparison is shown in figure 2.7. At small Q^2, the GRV parametrization lies a little bit above the data points, while the MRS and CTEQ distributions describe the data well (at Q^2 above their starting scales).

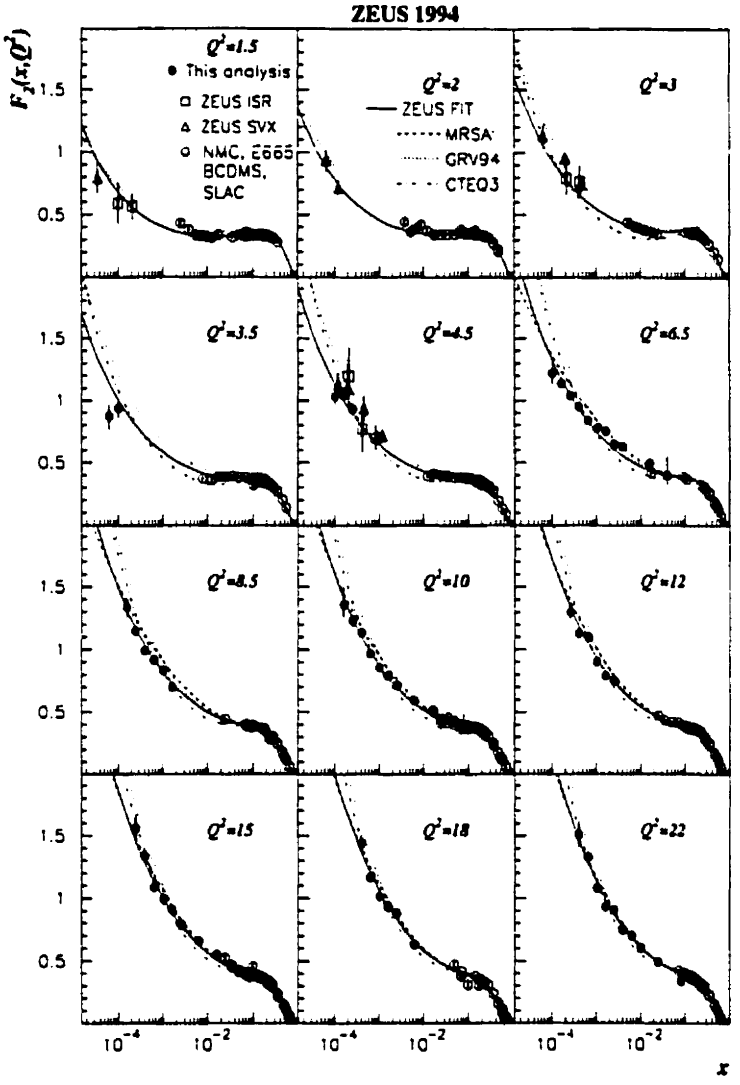

Figure 2.7: *Structure function F_2 of the proton as a function of x in various Q^2bins. The indexes "ISR" and "SVX" (respectively "Initial State Radiations" and "Shifted Vertex Runs") refer to specific methods to obtain F_2 at very low x and Q^2. Together with the ZEUS measurement are presented the results from NMC, E665, BCDMS and SLAC.*

2.3.3 A Perturbative Origin to the Steep Rise of F_2: the BFKL Equation

In the previous sections, the structure function of the proton was shown to rise at fixed x, as a function of Q^2, breaking this way the "Bjorken scaling" which arose from the QPM. This rise can be shown to follow a power law: $F_2 \sim x^{-\lambda}$ and it can be reproduced by a common set of parametrizations, using the standard DGLAP evolution picture, either by starting from a singular input at moderate Q^2(see previous section), or by chosing a flat input at very low Q^2. Although both methods agree with the data, they have drawbacks, mainly because they apply perturbative calculations in a region where non-perturbative and non-linear effects can be large. Moreover, we have seen in the section 2.2.3 that the splitting functions could be expanded in perturbative series involving terms of $\frac{1}{x}$ at high order (in particular the two splitting functions P_{gg} and P_{gq}). The DGLAP equations neglect those terms in $ln\frac{1}{x}$, if they do not come with a large logarithm of Q^2. However, at low x, these terms become large and can not be neglected any more and the DLLA might lead to missing some of these large logarithms of $\frac{1}{x}$, which do not come with a logarithm of Q^2.

We have seen that the LLA solution to the DGLAP equations was obtained by restricting the calculations to the phase space where the parton emissions were strongly ordered in k_T (see equation 2.24), and therefore uses the gluon density $xg(x,Q^2)$, integrated over k_T^2. In order to include the leading $ln\frac{1}{x}$ terms in the summation, a different technique has been developed, which is expressed as a function of the unintegrated gluon density defined as:

$$xg(x,Q^2) = \int_0^{Q^2} \frac{dk_T^2}{k_T^2} f(x,k_T^2). \tag{2.39}$$

The summation of the leading logarithms of $\frac{1}{x}$, taking into account the full dependence on the k_T of the emitted gluons is performed by the BFKL (Balitski-Fadin-Kuraev-Lipatov) equation [26]:

$$\frac{\partial f(x,k_T^2)}{\partial ln\frac{1}{x}} = \frac{3\alpha_s}{\pi} k_T^2 \int_0^{\infty} \frac{dk_T'^2}{k_T'^2} \left[\frac{f(x,k_T'^2) - f(x,k_T^2)}{|k_T'^2 - k_T^2|} + \frac{f(x,k_T^2)}{\sqrt{4k_T'^4 + k_T^4}} \right] \tag{2.40}$$

$$\frac{\partial f(x,k_T^2)}{\partial ln\frac{1}{x}} = \int d^2k_T' K_{LO}(\vec{k_T},\vec{k_T'}) f(x,k_T'^2). \tag{2.41}$$

As the equation only involves the unintegrated gluon density, it is no longer necessary to require strong ordering in the parton emissions and the phase space constrained

by equation 2.24 can be widened. However, because the equation sums only leading $ln\frac{1}{x}$ terms, a strong ordering on x's is required, similarly to the DLLA case:

$$x \ll x_n \ll \cdots \ll x_1. \tag{2.42}$$

The BFKL equation is only valid in the high energy limit, where $s \gg -t$ (using the Mandelstam variables s and t defined in section 2.2.1), that is, at fixed Q^2, in the low x limit. The derivation of this equation is quite difficult and is beyond the scope of the present work. It refers to the fundamentals of Regge theory, which has not been discussed here. For a detailed discussion on Regge theory, derivation and treatment of the BFKL equation, the reader can refer to [27].

The leading order term K_{LO} in equation 2.41 is the kernel of the Leading Order BFKL equation. The solution of equation 2.41 involves the determination of the eigenvalues of this kernel (it can be shown to be dominated by the largest eigenvalue of K_{LO} [28]).

The enticing feature of this evolution picture is that the integrated gluon distribution which is derived follows a power law as a function of x, reproducing the shape of the structure function F_2 as inferred through the experimental data, without having to make assumptions on the input parton distribution or to use pQCD at very low Q^2:

$$xg(x,Q^2) \sim f(Q^2)x^{-\lambda}. \tag{2.43}$$

Here, λ has been found to be of the form $\lambda = \frac{3\alpha_s}{\pi}4ln2 \simeq 0.5$. The rise of the structure function is therefore steeper than what is expected from the experimental data (where it has been found that F_2 was rising as $x^{-0.3}$), but this result relies on numerous approximations and is bound to some uncertainties, so that one can hope to improve this estimation.

First of all, because of the absence of k_T ordering, the parton emissions can diffuse in the non-perturbative region where hadronization uncertainties are large. A cut-off $k_{T,0}$ has to be introduced so that $Q_0 < k_{T,0} < k_T$. The result of the equation might therefore depend on this cut-off.

Next the result given in equation 2.43 has been obtained with a fixed α_s. Considering a running α_s might lead to a dramatic change of the results, as we just saw that the partons emitted might have very small k_T and therefore α_s can become very large.

In any case, a running α_s is expected to decrease the value of λ [29].

Unitarity corrections have to be envisaged as the parton distributions are singular for $x \to 0$. However considering the running α_s might reduce this effect as it is supposed to weight the distribution towards the infra-red (non-perturbative) limit.

Contrary to DGLAP, the BFKL equation does not require energy and momentum conservation at each branching. Implementing this requirement could lead to large discrepancies, in particular for the hadronic final states distributions (see [30]).

Finally, the most important changes to equation 2.43 might arise from the next-to-leading order (NLO) corrections to the BFKL equation. Much work has been done to implement them in the BFKL kernel [31] and the first estimations yield large and negative corrections. According to the authors, the BFKL equation at next-to-leading order would look like:

$$\frac{\partial f(x,k_T^2)}{\partial ln\frac{1}{x}} = \int d^2k'_T(\alpha_s K_{LO}(\vec{k_T},\vec{k'_T}) + \alpha_s^2 K_{NLO}(\vec{k_T},\vec{k'_T}))f(x,k'^2_T), \tag{2.44}$$

leading to the following solution:

$$xg(x,Q^2) \sim f(Q^2)x^{-\lambda+\Delta\lambda}. \tag{2.45}$$

Here, the exponent of x is $-\lambda + \Delta\lambda \sim 0.03$ [32]. At this stage, this number is only a rough estimate of the expected next-to-leading order corrections effects to the BFKL kernel and corresponds to the one-loop corrections only. It does not include effects of the running α_s and therefore the value of 0.03 should be considered very carefully. If this is true, this means that the BFKL picture is unable to describe the steep rise of F_2 and then the whole picture is wrong. However, as these NLO corrections are large, one might expect also large NNLO corrections (Next-to-Next-to-Leading Order), which could be on the other direction [32], increasing the slope of the predicted F_2. In this context, no solid conclusion could be drawn until one finds small high order corrections.

2.3.4 Attempts to Unify the DGLAP and the BFKL Pictures

Expanding the splitting function terms

One of the goals of the description of the parton evolution is to unify both pictures DGLAP and BFKL, as they are valid in complementary phase space regions: high x and high Q^2 for DGLAP, low x for BFKL. The validity range of each one of the model, as well as the kinematic boundaries which limit the use of perturbative calculations and linear equations, are shown in figure 2.8.

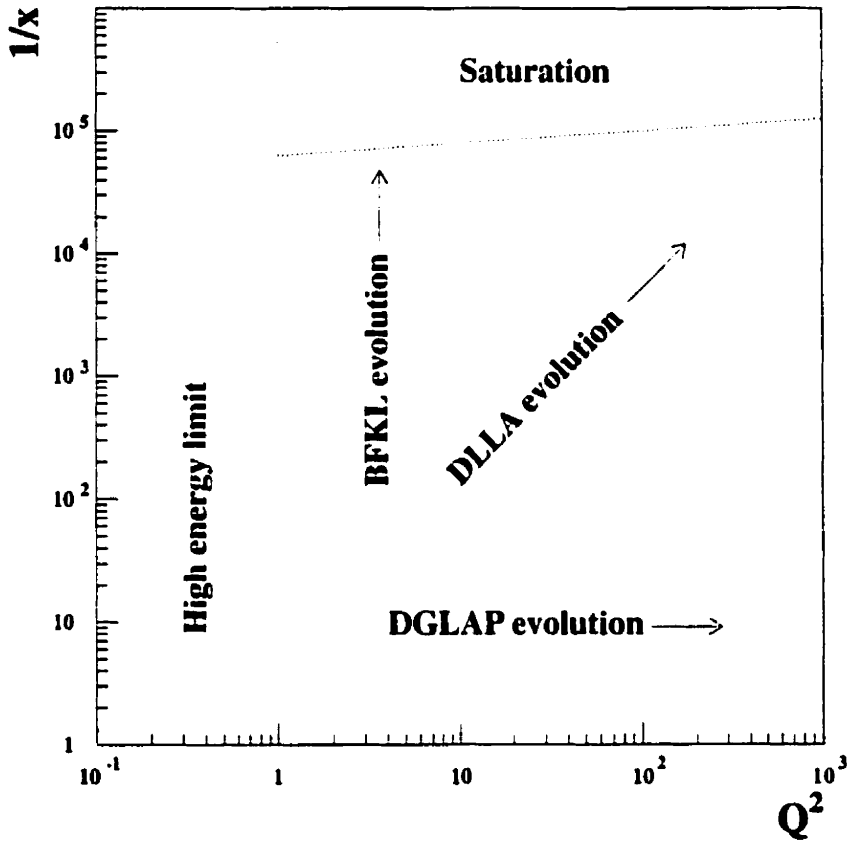

Figure 2.8: *Validity range for the three evolution equations: LLA(Q^2), DLLA and LLA($\frac{1}{x}$).*

To perform such a task, one might think about including in the DGLAP picture the higher order terms in $ln\frac{1}{x}$ which appear in the perturbative expansion of the splitting functions. The general shape of the splitting function $P(x)$ is (see [7, 24]):

$$xP(x,\alpha_s) = \sum_{n=1}^{\infty} \left(\frac{\alpha_s}{2\pi}\right)^n \left[\sum_{m=-\infty}^{n} A_m^n (ln\frac{1}{x})^{m-1}\right] \tag{2.46}$$

where the negative m terms are finite in the limit $x \to 0$. The various models which were part of the previous discussion correspond to the following approximations:

- $n = 1$: only leading logarithms of Q^2 are retained in the sum. This corresponds to the LLA(Q^2).
- $n = 2$: this is the so-called next-to-leading logarithm approximation (NLLA), where the first term in $ln\frac{1}{x}$ is taken into account.
- $n \geq 1, m = n$: this corresponds to the summations of the leading $ln\frac{1}{x}$ terms and therefore the LLA($\frac{1}{x}$). This can approximate the result of the BFKL equation.
- $n \geq 1, n = m$ *or* $n = m + 1$: it gives the next-to-leading logarithm correction to the $\frac{1}{x}$ sum. It is called NLLX($\frac{1}{x}$)
- $n = 1, m = 1$: this is the common point between the LLA($\frac{1}{x}$) and the LLA(Q^2) and corresponds to the DLLA.

The summary of these calculations is illustrated in figure 2.9.

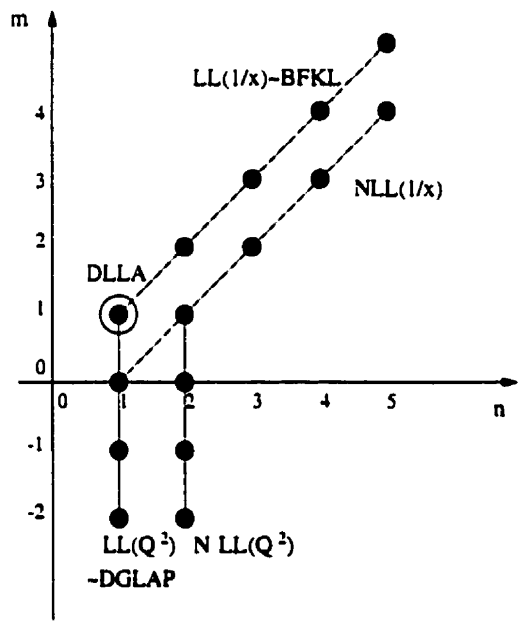

Figure 2.9: *The (n,m) plane defining the different kinds of summations, including the DGLAP and BFKL pictures.*

Angular ordering and the CCFM equation

A second attempt to match BFKL and DGLAP evolution pictures has been performed by the CCFM group [33] (Ciafaloni-Catani-Fiorani-Marchesini). Unlike the previous splitting function picture, the CCFM equation deals with an unintegrated gluon density (like the BFKL equation) and is based upon the idea that partons along the gluon ladder shown in figure 2.4 are emitted with an angular ordering, so that, if θ_i is the angle of the i^{th} emitted parton with respect to the original direction of the first gluon emitted in the ladder, $\theta_{i+1} > \theta_i$. The basis of this idea lies in the concept of colour coherence [34]. The theory of coherent emissions comes from QED: the starting point is the observation that a photon emitted from an e^+e^- pair could not resolve the internal structure of the pair if it is emitted with an angle larger than the angle between the two electrons. As a matter of fact, the transverse separation $\rho_\perp$ between the e^+e^- can be related to the emission angle of the photon through [16]:

$$\rho_\perp \approx \lambda_\perp \frac{\theta_{e^+e^-}}{\theta_{\gamma e}}. \tag{2.47}$$

Here $\lambda_\perp$ is the transverse wavelength of the emitted photon. Therefore, if $\theta_{\gamma e} \gg \theta_{e^+e^-}$, the wavelength of the photon is larger than the distance separating the electron-positron pair and the photon can not resolve each individual charge. It only resolves the total electric charge of the pair, which is zero in this case.

Similarly in QCD, the gluon emitted at large angle by a quark-antiquark pair can not resolve the internal structure of any one of the quarks or act as if it were emitted by the parent gluon (see figure 2.10). From this property follows that one can approximate the sequential emissions of partons along the gluon ladder as being ordered in angle. As the emission angle of the gluon is related to its energy E and transverse momentum k_T through $\theta = \frac{k_T}{E}$, the partons emitted in the high energy limit or at small x, where $E \gg k_T$ can be ordered in angle without being ordered in k_T, so the effects of the BFKL evolution can be approximated. At moderate x and Q^2, ordering in k_T is again implied and the result of the evolution should match the predictions based on the DGLAP equations.

Some work is currently made to match the CCFM predictions to the experimental data, but it has been found that the calculations are very much dependent on high order corrections, leading to large uncertainties in the description of the structure function of the proton [35, 36].

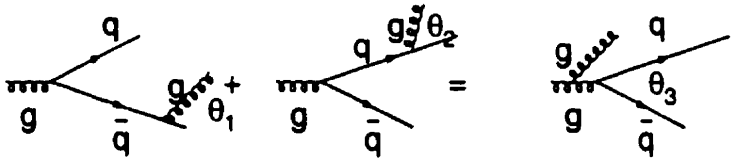

Figure 2.10: *Schematic representation of angular ordering in a QCD cascade due to colour coherence. The condition of colour coherence is:* $\theta_1 < \theta_3$ *and* $\theta_2 < \theta_3$.

2.4 Experimental Evidences for a BFKL-like Evolution Picture

The main purpose of the current study is to evaluate the validity range of the various pictures discussed above and in particular discriminate between the DGLAP and the BFKL evolution schemes. Since the early 90's, a great deal of work has been performed, especially at the *ep* collider HERA, to probe the low-x kinematic range with various observables.

We have seen that the structure function evolved with the DGLAP equations managed to reproduce quite well the steep rise of F_2 observed at HERA and in various fixed target experiments. To do that, the various parametrizations have to take into account the non-perturbative contributions to the structure functions F_2 either by chosing singular inputs to the DGLAP equations, or by starting the evolution mechanism at very low Q^2. It has then been concluded (for instance in [37]) that F_2 is too inclusive a quantity to be sensitive to the perturbative mechanisms which control the parton evolution. In order to probe the latter, it might be more suitable to use more exclusive quantities like the ones constructed from the hadronic final states, that is the hadronic outcome of the *ep* collision (further details on the definition of "hadronic final state" will be provided in chapter 6).

The starting point of all analyses aiming at determining the validity range of the various evolution pictures in DIS is the difference in the strong ordering requirements between the different evolution equations and in particular the absence of strong k_T ordering in the BFKL and CCFM pictures. Another important feature of the

BFKL/CCFM pictures is the steep rise of the gluon density at small x, which makes this kinematic region very sensitive to the parton evolution scheme.

The main quantities used up to now to probe the partonic mechanisms at small x are the transverse energy flow in the Hadronic Center of Mass (HCM) distribution, the transverse momentum p_T spectrum of charged particles, the forward π^0 meson production, the forward jet cross sections and the angular correlations between leading order jets in the HCM[3]. Although all these measurements can provide greater insight on small-x physics, they are bound to the same uncertainty as F_2, namely the hadronization and non-perturbative effects, although it is hoped that these are reduced in the case of hadronic final states.

2.4.1 Transverse Energy Flow in DIS

The first relevant observable to have been studied and measured at HERA is the transverse energy flow in the central rapidity region of the hadronic center of mass frame [4]. The studies have been performed mainly by the Durham group and explicit calculations were provided for comparison with the experimental data [39]. The idea is that in the BFKL picture, a larger amount of transverse energy than in the DGLAP scheme is expected in the region between the leading order partons (quark box in figure 2.4) and the proton remnant, which correspond to the central rapidity region in the HCM. The calculations have been performed for $x < 10^{-3}$ and have yielded a E_T distribution exhibiting a fairly flat plateau in rapidity with $E_T \approx 2$ GeV per unit of rapidity. The same calculations performed with the DGLAP equations lead to a much smaller amount of transverse energy in the rapidity plateau ($E_T < 0.5$ GeV per unit rapidity).

The transverse energy flow was first measured by the H1 collaboration in [40] using the 1993 data of the HERA ep collider (center of mass energy: $\sqrt{s} = 296$ GeV). The low-x range was probed by this analysis: $5 < Q^2 < 100$ GeV^2 and $10^{-4} < x < 10^{-2}$, and the quantity $\frac{1}{N}\frac{dE_T}{d\eta}$ was estimated in various bins of x and Q^2. A second measure-

[3] In the following introduction, we will concentrate on measurements performed at HERA and we won't speak about other attempts, like the one performed by the D0 collaboration at Tevatron, to study the production of a pair of jets at large rapidity interval [38]

[4] The rapidity is a longitudinally Lorentz invariant measure of the polar angle. In practice, the pseudorapidity η is used in the measurement. This quantity will be defined in chapter 4.

ment by the ZEUS collaboration [41] used the 1994 data sample of HERA with enhanced statistics and probed a similar kinematic range: $Q^2 > 10$ GeV^2 and $x < 10^{-3}$. Both measurements found trends which were compatible with a BFKL-like dynamics, namely an enhanced amount of transverse energy in the central region of rapidity in the HCM.

However there were large uncertainties with these results, mostly due to the hadronization effects: the data which is collected in the ZEUS and H1 detectors is made of stable hadrons, which live long enough to reach the main component of the detectors. At this stage, the long-range physics contributes significantly to the measured effect and the perturbative calculations, performed in the limit of short-range physics must be corrected for all the non-perturbative phenomena which enter into the physics process. As these effects are not calculable, one is bound to use models to estimate them. These simulations, called Monte Carlo models because they use random numbers in the process of event generation, will be discussed in details in the next chapter. In the transverse energy flow analysis, the DGLAP-based generator LEPTO [42] and the BFKL-like model ARIADNE [43] were used to estimate the "hadron level" energy flow. The non-perturbative effects included in the Monte Carlo model can be tuned and the transverse energy flow was found to be very sensitive to one of LEPTO's soft effect: the Soft Colour Interaction (SCI) (this effect will be discussed in some details in the next chapter). By varying the amount of SCI, the LEPTO model was able to reproduce the transverse energy flow in the central rapidity region (see [44]). The key point is that 60-80% of the energy flow predicted by this model is produced during the hadronization phase. By contrast, in a BFKL-like model like ARIADNE, only 30-40% of the energy flow is produced during the hadronization phase. The parton level predictions, based on pure BFKL calculations lie, significantly above both models predictions.

The measurement of the transverse energy flow is therefore unable to distinguish between the DGLAP and the BFKL pictures. Similarly to the F_2 case, the non-perturbative contributions to this process are too large to draw a conclusion. One might then think that the Monte Carlo models can be used to correct the data to the parton level, and then compare them directly with the analytical calculations. Unfortunately, these corrections vary considerably from one model to another and the systematic errors which arise from the model dependence are too large to allow a

statement.

2.4.2 Charged Particle Spectrum

Since the contribution from the non-perturbative phenomena is larger than the predicted effects of the perturbative evolution in the transverse energy flow analysis, some attempts have been performed to define observables which are less sensitive to hadronization.

In 1995, M.Kuhlen suggested the study of the hard tail of the transverse momentum p_T spectrum of the charged particles, as a way to discriminate between DGLAP and BFKL pictures [45] (in this context, the term "hard" refers to the large transverse momentum contributions which are supposed to arise from the perturbative part of the process). The measurement of the p_T spectrum of single particles, instead of the global transverse energy flow of an event, is expected to be closely related to the parton cascade process, which takes place before hadronization. In a scenario where this parton cascade is unordered in transverse momentum, more high p_T particles should be produced than in the DGLAP scheme, where the k_T-ordering suppresses this type of particles. The difference with the previous analysis is that in this case, the expected excess of high p_T particles is less likely to come from non-perturbative effects as these ones happen mainly at low p_T (the low transverse momentum contributions arising from non-perturbative processes are called "soft" effects).

The H1 collaboration measured this hard p_T tail in [46] using the 1994 data for $5 < Q^2 < 50$ GeV^2 and several x bins from $x = 0.0021$ to $x = 0.00016$. The charged particles were measured in the forward tracking chamber and were required to originate from the primary vertex. The result of this measurement is shown in figure 2.11. The two models LEPTO and HERWIG refer to DGLAP-based simulation, while ARIADNE is a model which incorporates a non-ordered parton shower (see next chapter). While the data agree with all the models (DGLAP and BFKL) at high x, the DGLAP-based models exhibit a "softer" p_T spectrum at low x, and ARIADNE, which does not require a strong ordering in the parton emissions agrees fairly well with the data in each x bin. These promising results have yet to wait for accurate theoretical calculations as input to the simulations.

2.4.3 Forward Pion Production

An alternative measurement to the forward jet cross section has been recently suggested and performed by the H1 collaboration: the study of forward going pions [47]. The basis of this measurement is the same than for the forward jet analysis: the most forward parton emitted in the gluon ladder displayed in figure 2.4 is followed through the entire evolution process and can be found as a single pion in the detector. The theoretical predictions for this analysis have been performed by the Durham group and can be found in [48].

The H1 analysis measured π^0 in the dominant channel: $\pi^0 \rightarrow \gamma\gamma$ and within the polar angle: $5^o < \theta_\pi < 25^o$. They were required to have an energy $E_\pi > 8$ GeV and transverse energy $E_{T,\pi} > 1$ GeV. The pion production rate as measured in the H1 detector was much larger than the predictions based on the DGLAP equations. As a complementary check, forward charged particles were also measured, using the forward tracker of the H1 detector and the results were found to be in good agreement with the π^0 production.

2.4.4 Forward Jets and Angular Correlation between Jets

A second class of study focuses on the observation of a single parton, emitted at the bottom of the gluon ladder shown in figure 2.4. This type of analysis, suggested by Mueller in the early 90's [49], is based on the idea that, due to the combined requirements of strong ordering in k_T and ordering in x_i (see figure 2.4 for the description of the variables), the cross section for a parton emitted at large x_i and large transverse momentum should be suppressed in the DGLAP-scheme, while it should be enhanced in the BFKL and CCFM pictures. In practice, obtaining large x_i partons means studying forward going jets, so the objects studied lie in the forward region of the detector.

Alternatively, one can study the angular correlation between the two partons at the top of the gluon ladder (see figure 2.4): if the subsequent parton emissions yield a small overall transverse energy, the two jets should be strongly back-to-back correlated in the hadronic center of mass frame. If the amount of transverse energy is large, like in the BFKL/CCFM pictures, then they must lose this back-to-back correlation.

As these two studies form the core of the present work, they will be described in more detail in chapter 6 and are not developed here.

2.5 From Theory to Experiment

In this chapter, we saw how the various models of parton evolution are supposed to describe the experimental data and in particular the behavior of the structure function of the proton F_2, within a set of approximations which are valid in some part of the kinematical domain: high x, high Q^2 for DGLAP, low x and moderate Q^2 for BFKL, both high and low x in the CCFM picture. Some experimental methods to discriminate between the various models have also been discussed and two of them will be studied more extensively later (chapters 7 and 8). Unfortunately, in most analyses and in particular in most QCD studies, the sole understanding of the perturbative theory is not sufficient to interpret the data. To achieve such an understanding, a more global picture, which also include the effects of hadronization, has to be considered. Such a picture is avalaible in the Monte Carlo simulation which is both an extension of the theory as it includes the results of the perturbative calculations as well as a model for the non-perturbative phase, and an experimental tool, as it is supposed to picture the distribution of particles as seen in the detector and can implement the response from the various detector components. The description of the various Monte Carlo models and in particular the implementation of the different models of parton evolution is discussed in the next chapter.

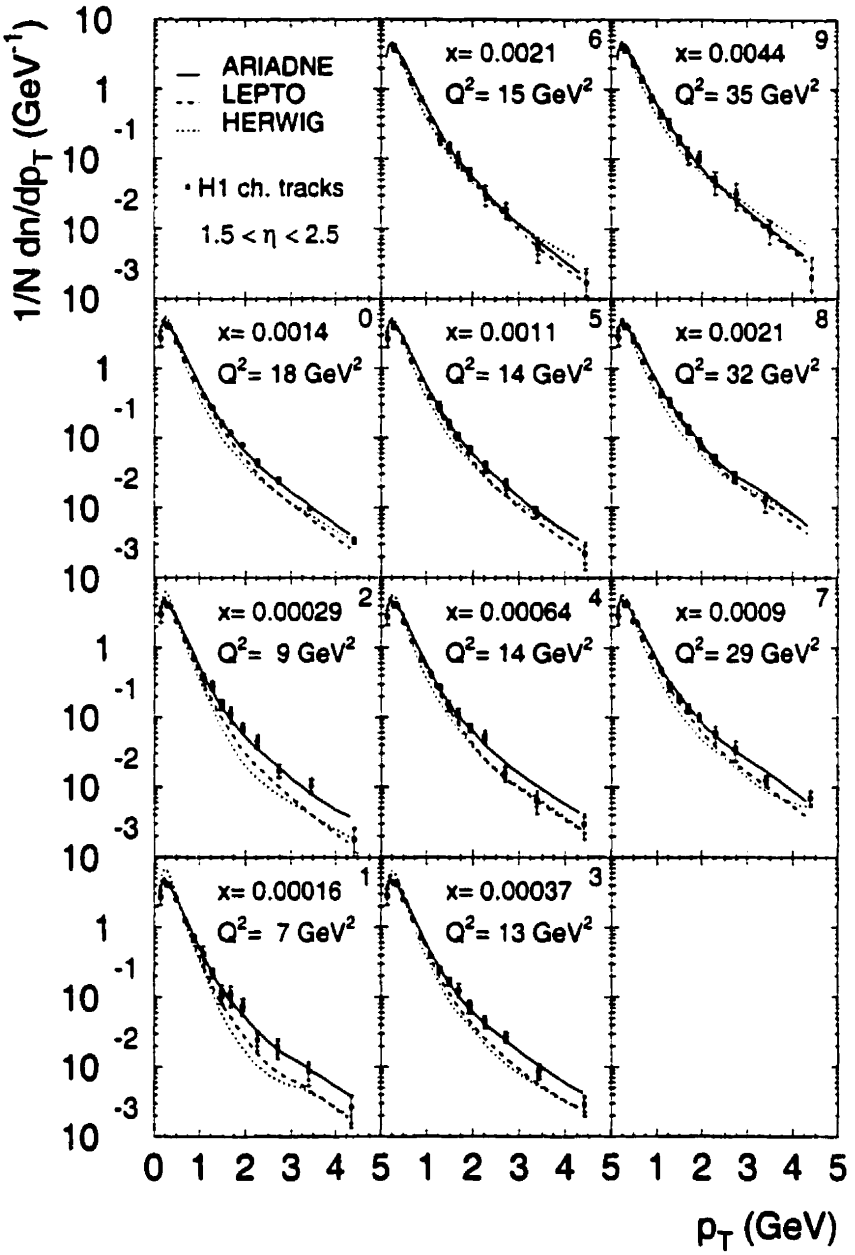

Figure 2.11: *Transverse momentum distribution for single charged particles in various bins of x and Q^2. The data are compared to two DGLAP-based models (LEPTO and HERWIG) and to ARIADNE, which does not implement strong k_T ordering for the parton emissions. The inner error bars are the statistical errors while the outer error bars represent the systematic uncertainties.*

Chapter 3

Event generators and Detector Simulation

The predictions provided by the various evolution pictures presented in the previous chapter are very useful from a theoretical side to understand the underlying parton dynamics, but they are very hard to compare to the data extracted from the *ep* collision at HERA. First, the data seen by the detector is made of hadrons and the theoretical calculations are mostly based on parton-level models (i.e. DGLAP and BFKL equations). The necessary step of hadronization is not analytically calculable and the uncertainty that it adds to the interpretation has to be taken into account. Moreover the theory does not take into consideration all the detector effects which can alter the results and generate additional uncertainties.

To include these two effects into the interpretation of the data, one has to rely on models, which are based on the various theory calculations and evolution pictures, but implement the additional effects. These models are useful for two reasons: they are used to correct the data for both detector and hadronization effects and they provide themselves a first theoretical approach on the physics which is being probed and can therefore be directly compared to the data. This last point is subject to polemic, as these models have a set of phenomenological parameters which must be tuned to some well-known distributions, in order to define their validity range. However it is not known for sure whether these parameters are universal and can be used in any kind of analysis, although the common understanding is that they should not be changed from one analysis to another. Another bone of contention is that, although all the models must have a similar response to the detector effects, the hadronization

step might be very different from one to another. In this case, it might be difficult to determine which model to compare the data to and what is the real contribution from hadronization to the effect to be seen or to the quantity to be measured. For instance, in the case of the transverse energy flow analysis (see chapter 2), two models were giving the correct description of the data, although both underlying partonic mechanism and non-perturbative effects were very different. In this case, the "compensating" effect of soft physics prevented any strong conclusion about the real nature of the parton evolution.

3.1 Basics of a Monte Carlo Simulation

The event generators which are used all along in this work and in most analyses in high energy physics have all a common statistical basis: that is the so-called Monte Carlo method. The purpose is to simulate a probability distribution, using the random behavior of the physical variables, as in games of chance, hence the name 'Monte Carlo'.

The principle of the method is the following: in order to describe the behavior of function f, for instance a differential cross section, which depends on one or several variables x, which can be for instance phase space variables, one uses the integral of the distribution $f(x)$ and interprets it as being an average. The normalization is taken care of at the end of the generation process. The integral of the distribution, within the phase space range under investigation (here defined by the lower limit x_1 and the upper limit x_2) can be written as:

$$I(x) = \int_{x_1}^{x_2} f(s)ds = (x_2 - x_1) < f(x) > \approx (x_2 - x)\frac{1}{N}\sum_{i=1}^{N} f(x_i) \equiv I_N. \qquad (3.1)$$

Here the value of the distribution f is known at the N rungs $x_1, x_2, \ldots, x_i$.

The integral I can therefore be approximated by a discrete sum I_N. This approach will be more accurate when the number of points in the sample will be large. The error can be estimated by the variance of the distribution.

If the various degrees of freedom x_i are independent, the Central Limit theorem gives the value of the variance V_N as a function of the number of integration points N and

the mean I of the distribution:

$$I \approx I_N \pm \sqrt{V_N/N} \tag{3.2}$$

Therefore the error on the true value goes as $1/\sqrt{N}$. This error is independent of the number of dimensions (i.e. number of phase space variables) and depends only on the number of samples one considers. In practice, the following method is used in the process of generating events: a given random number u is generated from a uniform distribution from 0 to 1 which corresponds to the integral I. For each u, the phase space variable x_k is obtained so that $I_N(x_{k-1}) < u < I_N(x_k)$. The probability distribution $f(x)$ is then derived through the relation: $f(x_k) = I_N(x_k) - I_N(x_{k-1})$. The method is illustrated in figure 3.1. One can see that the steeper the distribution is, the more points are needed for the sampling in order to describe accurately the probability distribution, so that in practice, one generates more events in the region where the function is rapidly growing.

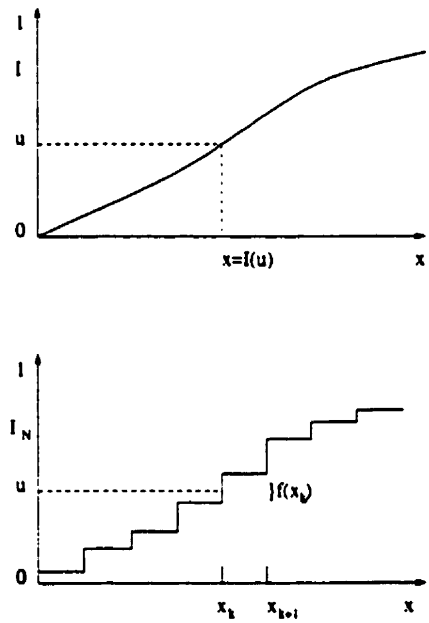

Figure 3.1: *Schematic representation of the sampling technic used in the Monte Carlo simulation, based on the choice of a random number generator to find the differential probability distribution* $f(x)$.

The phase space available to generate n particles in a Monte Carlo is defined by the

Lorentz invariant:

$$d(PS)_n = \prod_{i=1,n} \frac{d^3p_i}{(2\pi)^3(2E_i)(2\pi)^4} \delta^{(4)}\left(p_0 - \sum_{i=1}^{n} p_i\right). \tag{3.3}$$

Where p_0 is the momentum of the initial particle and p_i, the momentum of the i^{th} final state particle. The last term expresses energy and momentum conservation. In practice, the events are generated within a certain phase space, defined by cuts on the kinematic variables. These cuts are then the boundaries of the integral of the cross section. To save computer time, the phase space coordinates are generated once (in the first step of the simulation), on a uniform grid which depends on the boundary conditions which limit the kinematic region. The grid also depends on the input parton density which has to be given by the user and which is used in the subsequent parton evolution. The parton density sets the absolute normalization (total cross section) of the process. In the present work, the parametrization CTEQ4D [22] was used as it described the total event cross section best.

3.2 Models of Parton Evolution

Once a DIS event has been generated in the appropriate phase space, it must follow the usual steps of evolution until it reaches the stage of observable. The first part of this so-called QCD radiation process is to generate a perturbative parton level evolution according to the standard evolution equations like DGLAP, BFKL and other.

On top of the standard requirements proper to each evolution picture, the Monte Carlo simulation has to implement additional criteria to make the evolution realistic on a physics point of view and therefore describe the full final state observed in the experiment. These criteria are (usually) subleading corrections to the final cross section, like conservation of energy and momentum at each parton branching (which is implicit in the DGLAP picture but not in the BFKL equation) or radiation from quark (which is usually neglected in the CCFM picture) or heavy quark mass. All this makes the process of generating the parton evolution a tedious but necessary task.

3.2.1 The Matrix Elements-Parton Shower Model

Most of the current Monte Carlo simulations use the Matrix Elements -Parton Shower (MEPS) ansatz to generate the physics processes taking place at HERA. The idea is to use fixed-order perturbative calculations (up to now only at leading order) and then, simulate the higher-order processes by a "parton shower" based on a leading logarithm approximation, like the one obtained from the DGLAP equations. The schematic organization of the MEPS ansatz is illustrated in figure 3.2.

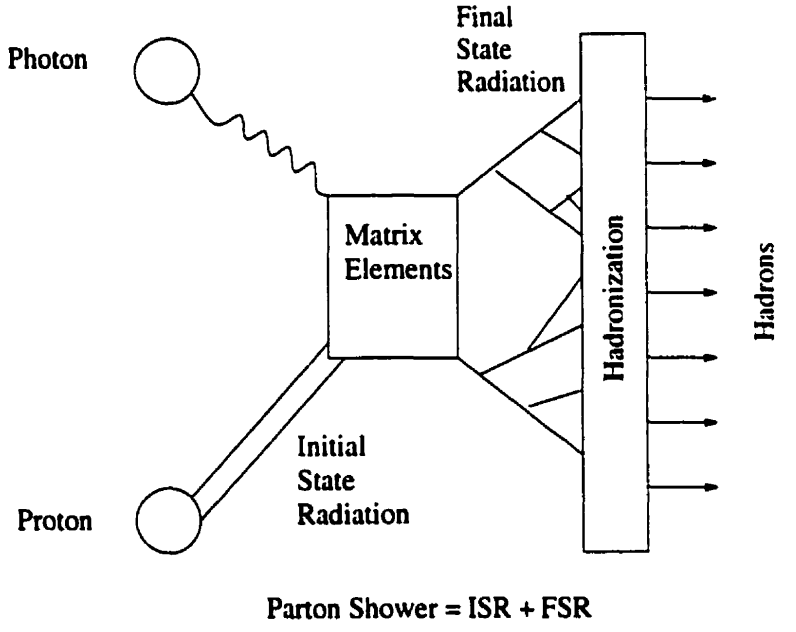

Figure 3.2: *Schematic representation of the Matrix Elements-Parton Shower ansatz. The parton shower is divided into initial state and final state radiation.*

Matrix Elements

The leading order process in DIS is the QPM: $\gamma q \rightarrow q$ shown in figure 2.1. In this process, there is no gluon vertex and the process is therefore $\mathcal{O}(\alpha_s^0)$. The two $\mathcal{O}(\alpha_s^1)$ (leading order QCD) processes are also included in the determination of the matrix elements: the BGF $\gamma g \rightarrow q\bar{q}$ and the QCDC $\gamma q \rightarrow qg$, both shown in figure 2.3. The determination of which process is to be used on an event-by-event basis is done through probabilities, given $P_q + P_{q\bar{q}} + P_{qg} = 1$. The total probability for a specific process to occur is given by the total cross section of this process within the kinematic region under consideration (and depends also upon the parton density and the choice

of cut-off).

As we saw in the section 2.2.2, the leading order QCD cross sections have logarithm divergencies which correspond to the case where the two partons are emitted collinearly or when they have very small energy ("infra-red" domain). To get rid of these divergencies, the Monte Carlo simulation uses the factorization property of the DIS cross section [13] and integrates the singularities into the parton densities, which are later evolved in the parton shower phase. In practice, the procedure is to impose a cut-off in the phase space. Usually, the cut-off is applied to the invariant mass distribution of the two leading-order partons $s_{ij} = (p_i + p_j)^2$. The cut-off is proportional to the invariant hadronic mass $W^2 = (P+q)^2$ (where P is the four momentum of the proton and q, the four momentum of the photon). The cut-off is expressed by the relation:

$$s_{ij} < y_{cut} W^2 \tag{3.4}$$

The cut-off y_{cut} is a tunable parameter in the package. Its effect is mainly to change the relative contributions of the three leading-order processes involved in the matrix elements (see [42] and [50] for a further discussion on the cut-off parameter and some alternative implementations in Monte Carlo and NLO calculations). So far, there is no NLO calculation implemented in the MEPS simulations. The main reason is that, because of the divergencies, it is quite difficult to interpret the NLO cross section in term of probabilities, as the differential cross section is not always positive definite [51] (which is mandatory in order to interpret the integral defined in the previous chapter as an average over a probability distribution). However, there are NLO predictions on the form of calculation package like MEPJET [52], DISENT [53] or DISASTER++ [54]. These calculations provide the full NLO corrections to the 1 and 2 partons emissions, including the virtual corrections.

Parton Shower

As the high-order calculations can not be included in the Monte Carlo simulation, these corrections are approximated by the parton shower processes which are based on the evolution equations discussed in chapter 2. The standard parton showers are ruled by the DGLAP equations [17], described in chapter 2. This picture is also called the "collinear limit" as it features partons evolving towards (or from) the photon coupling and gaining (or loosing) virtuality by successively emitting harder (softer) partons

(that is partons with larger (smaller) transverse momenta). As shown in figure 3.2, the parton shower is in practice split into two phase, which are indistinguishable from a theoretical point of view (Heisenberg uncertainty principle) but give a convenient picture of the DIS processes:

- the Initial State Radiation (ISR) pictures the evolution of a constituent of the proton, initially on mass-shell, which acquires increasing negative virtuality (this is called a space-like cascade) by emitting successively harder partons and loosing at each branching a fraction of its total momentum (from momentum-energy conservation). In the standard application of the collinear limit, this process is actually implemented backward (from the photon coupling towards the parton within the proton). The ISR corresponds to the evolution of the structure function F_2 of the proton.

- the Final State Radiation is comparable to the $e^+e^- \rightarrow q\bar{q}$ evolution picture. After the coupling with the photon, the off-shell parton (with a large space-like virtuality) reduces its mass by successive parton emissions. This is called a time-like evolution and it goes on until all emitted partons become on-mass-shell again (in practice, until α_s becomes too large to apply perturbative equations).

The branchings involved in a parton shower are defined by the splitting functions (see chapter 2) in the frame work of the DGLAP picture. In order to implement them in the simulation, a probability is defined keeping in mind that, if a parton branches in an interval $[q^2 + dq^2, q^2]$, then nothing must have happened between Q^2(at the start of the evolution) and q^2 (the virtuality of the parton before the branching). The situation is then analog to a radioactive decay: if $d\mathcal{P}$ is the probability that there is an emission between $q^2 + dq^2$ and q^2, and $\Delta_{q^2}(Q^2)$, the probability that there is no emission between $Q^2$and q^2, translates into:

$$\frac{d\Delta_{q^2}(Q^2)}{dq^2} = -\Delta_{q^2}(Q^2)\frac{d\mathcal{P}}{dq^2}. \tag{3.5}$$

Since the probability $\mathcal{P}_{ij}$ can be written as a function of one of the splitting functions P_{ij} defined in section 2.2.3, as

$$d\mathcal{P}_{ij} = \frac{\alpha_s}{2\pi}\frac{dq^2}{q^2}\int dx P_{ij}(x), \tag{3.6}$$

the solution of this equation gives the "Sudakov form factor" $\Delta_{q^2}(Q^2)$:

$$\Delta_{q^2}(Q^2) = exp\left(-\int_{q^2}^{Q^2}\frac{dk^2}{k^2}\int dx\frac{\alpha_s}{2\pi}P(x)\right) \tag{3.7}$$

The evolution process can therefore be calculated step by step, using the Sudakov form factor (a detailed description of the Monte Carlo method of evolution can be found in [7]). This is carried on until the partons reach a small enough virtuality Q_0^2 (on the order of 1-2 GeV^2, depending on the simulation). At this point, the evolution is stopped and the hadronization phase takes place.

Most of the modern Monte Carlo models use the MLLA [55] (modified leading logarithm approximation) instead of the standard LLA. This new approximation takes into account the interferences between ISR and FSR, so that the total contribution of the parton shower is not equal to the sum of the each ISR and FSR individual contributions. Moreover coherence effects are implemented in most of the modern simulations. All these effects improve a lot the description of the data, with respect of the basic picture of independent parton evolution.

The MEPS-based Monte Carlo used all along the present study are LEPTO 6.5 [42], RAPGAP 2.06 [56] and HERWIG 5.9 [57]. All these Monte Carlo simulations implement the DGLAP evolution picture and are therefore are expected to reproduce its validity range. Unlike LEPTO and RAPGAP which implement the strict k_T-ordering in the parton emissions, HERWIG implements a strict angular ordering in the FSR, while it has an energy and angular ordering in the ISR (in order to take into account coherence effects as accurately as possible) [57]. HERWIG also uses a different hadronization scheme based on the cluster model [58] (see below). RAPGAP implements, on top of the standard parton shower, a "backwards evolution" from the quark box to the photon side, which reproduces the effect of a resolved photon (this will be developed later in chapter 8).
Another Monte Carlo model, PYTHIA [59], uses the ME-PS method but it has been developed and checked more extensively for hadron-hadron collisions than for DIS, although a DIS mode exists. This is why this simulation is not used in this study.

3.2.2 The Colour Dipole Model

Another type of QCD radiation is implemented in the Colour Dipole Model (CDM) [60]. Unlike the previous ansatz, the CDM does not make explicit use of an evolution equation and Feynman diagram to describe the parton evolution process. Here parton pairs involved either in e^+e^- or in ep collisions are treated as independent dipoles, radiating other partons which are themselves part of a radiative dipole, as it is shown

in figure 3.3. In DIS, the radiations occur between the struck quark and the proton remnant, therefore all the QCD cascades can be treated as FSR. The total phase space for parton emission is limited in rapidity: $y < ln(W/P_T)$, where W is the total amount of energy radiated and P_T, the transverse momentum of the radiated parton. From the proton side, the phase space is further suppressed because the proton has a finite size and acts therefore as an extended antenna, so that the total phase space allowed for radiative emission looks like the one shown in figure 3.4.

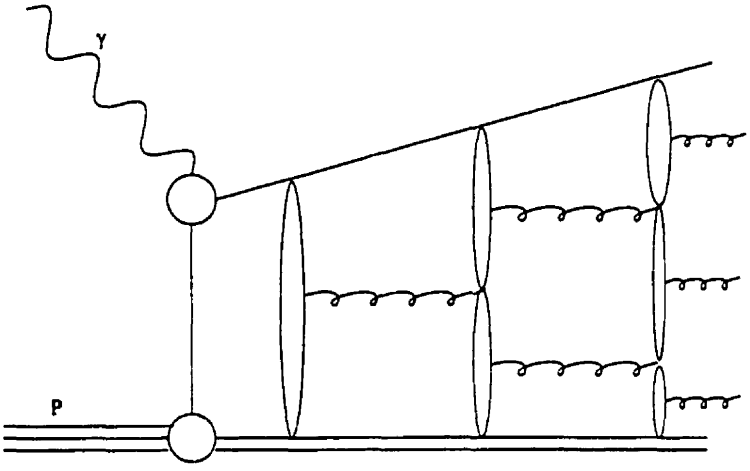

Figure 3.3: *The QCD cascade as implemented in the Colour Dipole Model. Each pair of partons is treated as independent radiating dipole.*

In the CDM, the boundary conditions which restrict the available phase space for evolution are relaxed, so that a parton radiation can occur at a scale higher than the scale of previous one (but the emissions are still ordered in rapidity, or, which is equivalent, in x). In this context, the CDM can be seen as a BFKL-like simulation, since it lacks strong ordering in the parton emissions. This feature should enhance the parton emissions at small x, exactly as it is expected in the BFKL scheme, although the CDM does not implement literally the BFKL equation.

The measurement of the di-jet cross section by the H1 collaboration [40, 61] confirmed that the CDM predicts a much larger rate than the MEPS-based models. Recently, it has been argued that this increase is due to an "unorthodox" suppression factor in the CDM, with respect to the MEPS models [62]. This factor is due to the absence of initial state radiation in the CDM, and the subsequent boundary conditions from the proton side. Instead, the boundary conditions arise from the fact that the proton

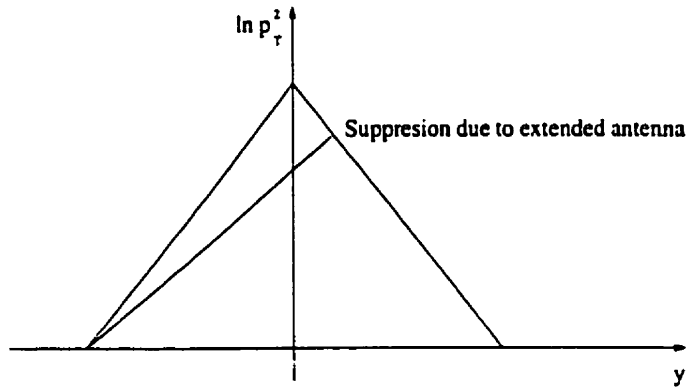

Figure 3.4: *Available phase space for parton radiation in the Colour Dipole Model. P_T is the transverse momentum of the emitted parton and y is its rapidity.*

emits radiations as an extended antenna. According to these authors, this creates an increase of hard parton emissions at small x. In any case, the physics implications of the CDM are not yet fully understood, and although it offers a convenient alternative to the DGLAP-based models (as it describes the total jet rate observed in the experiment), it is too premature to state on its correspondence with the BFKL picture.

In this study, the CDM is implemented in the ARIADNE package [43] (ARIADNE 4.08). Here, the LO matrix elements are taken from the LEPTO package, while the QCD cascade follows the CDM.

3.2.3 The Linked Dipole Chain

Recently, a new model has appeared which enables the implementation of the CCFM equation in a Monte Carlo simulation: the Linked Dipole Chain model (LDC) [63].

As it was shown in section 2.3.4, the CCFM picture interpolates between the DGLAP and the BFKL schemes, so that first estimations of a BFKL-type effect can be evaluated. This Monte Carlo simulation has more solid theoretical basis than the CDM, as it is based on an evolution equation and implements ISR.

The Linked Dipole Chain Monte Carlo (LDCMC) [64] uses the CCFM equation to evolve the structure function of the proton in the Initial State Radiation. On top of the angular ordering which restricts the phase space available for radiations, the LD-

CMC implements other requirements which are needed in a Monte Carlo simulation, like the conservation of energy and momentum at each branching, the quark density of the proton and the suppression factor for heavy quarks. The ISR creates a pattern of color charges which are then evolved into a final state cascade with the CDM itself.

The LDCMC (version 1.0) implements its own parametrization of the structure functions, which are described in [64]. It has been compared in [64] to several final state observables and its predictions lie between ARIADNE and LEPTO. As will be seen later (chapter 8), it is not yet understood why the LDC does not describe the data.

3.3 Hadronization Phase and Non-Perturbative Effects

The models presented earlier already differ in the perturbative treatment they apply to the parton cascade. In order to obtain the final state quantities, an additional step has to be taken, that is transforming the coloured partons into observable colour singlet hadrons. Unlike the parton cascades, the hadronization models are not based on perturbative calculations but are phenomenological pictures determined by a set of parameters which are often obtained by fitting experimental data. This hadronization phase decreases the predictive power of the Monte Carlo simulation, but is mandatory for a good description of the data.

3.3.1 String Fragmentation

The Lund string model [65] provides the most common picture for the Monte Carlo simulation. It predicts that a string is stretched between any quark-antiquark pairs created during the QCD radiation phase, as well as between quark-diquark (proton remnant) pair. The string creates a uniform QCD field on the order of 1 GeV/fm, which then creates $q\bar{q}$ pairs when the stored energy is large enough. The "break points" of the string are the color triplets q and $\bar{q}$, while the incident gluons, a color octet state, only produce "kinks" on the string, thus modifying its energy and momentum at the particular point where the gluon hits it, without breaking it (see figure 3.5a). These kinks lead to the rapidity (polar angle) distributions of the final

state particles in the detector. The attractive feature of this model is the independence of the final result on soft and collinear particles which can be created during the QCD cascade phase [66].

The string model is implemented in the LEPTO, ARIADNE, RAPGAP and LDC simulations via the JETSET package [59]. The various parameters are obtained through fits of the e^+e^- data.

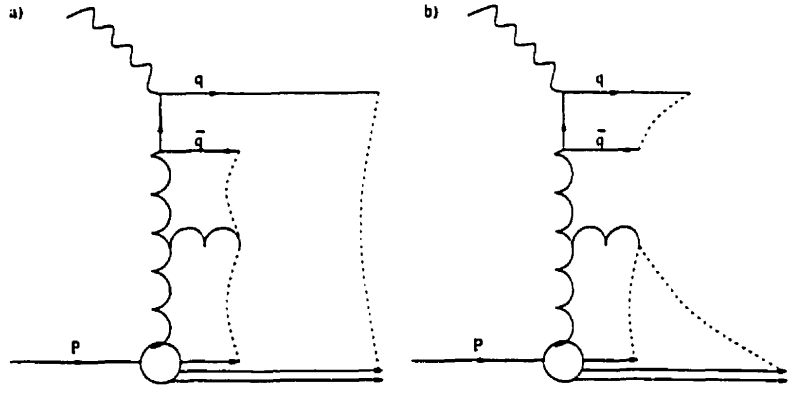

Figure 3.5: *a) Colour connection between partons in the standard string model. b) Soft colour interaction between parton pairs. No string is stretched between the quark box and the proton remnant creating a rapidity gap.*

3.3.2 Cluster Fragmentation

The cluster model [58] uses the property of "preconfinement" of colour to describe the hadronization phase: in the perturbative limit (high masses), the confinement of a partonic system is local, independent from the type of hard subprocess from Q^2. This leads to the property of Local Parton-Hadron Duality, meaning that the partons are converted into hadrons locally in phase space (this property will be discussed in chapter 6).

In the cluster model, each gluon which arises from the parton shower decays into a pair of quark-antiquark (or diquark-anti-diquark). Each parton of the pair is connected by a colour line to the neighboring parton (this is the planar approximation), thus forming a colour singlet cluster. This cluster follows the "preconfinement" property, that is, it is independent of the hard subprocess and the energy scale. The cluster then fragments isotropically (in its rest frame) into pairs of hadrons, following flavour

conservation.

This model is implemented in the HERWIG simulation and has the advantages of having less phenomenological parameters and of describing well the e^+e^- scattering. However, we will see in chapters 7 and 8 that it does not describe well all the kinematic variables of our DIS sample, in the kinematic range selected for this analysis.

3.3.3 Soft Colour Interaction

A third kind of non-perturbative effect has been implemented into the LEPTO model: the Soft Colour Interaction (SCI) [67]. In this model, similar to the Lund string model, all the partons created perturbatively after the parton shower phase interact softly between each other and with the colour medium of the proton through colour strings. These interactions modify the partons' colour but do not change their momenta. Unlike the strings from the Lund model, the soft colour connections can be formed between any pair of partons, including the gluons. This property affects the topology of the subsequent hadronization strings, thereby modifying the final state distributions. In figure 3.5, the pattern of a soft colour interaction is compared to the usual Lund string model picture. In the SCI case, the gluon can couple to the diquark (proton remnant) creating a colour singlet object. On the other hand, the leading order partons, from the quark box, are also connected by a soft string, leading to another colour singlet system. In between the two colour singlets, there is no colour connection, leading to a rapidity gap (no hadronic activity in a large rapidity interval). Historically, the SCI was introduced to explain the large fraction of DIS events with a large rapidity gap (around 10%) [68]. The theorists claim [69] that this model is adequate to describe the events observed at HERA.

A side effect of this model, presented in [69], is the increase of hadronic energy in the forward region of the detector (corresponding the central rapidity region in the HCM frame). This effect, apparently orthogonal to the previous one, is due to fluctuations created by the soft colour strings, leading to an increase of energy per unit of rapidity. In section 2.4.1, the excess of transverse energy in the central rapidity region of the HCM frame was interpreted as being one of the possible effects due to a new kinematic regime, namely the BFKL picture. The conventional Monte Carlo models, based on the DGLAP equations, were unable to describe correctly this amount.

The implementation of SCI in the LEPTO model enables to describe the distribution of transverse energy flow observed in the data, without having to require a new kinematic regime as it is shown in figure 3.6. The data collected by the H1 experiment [40, 70] are compared to the LEPTO model, implementing the SCI and a new sea quark treatment (SQT) [42]. These two non-perturbative phenomena manage to increase the relative amount predicted by the LEPTO model, so that this Monte Carlo simulation matches the rate observed in the data.

In LEPTO, the SCI are handled through a parameter adjustable by the user. It sets the probability for a parton to interact through SCI with another parton (or with the proton remnant). This parameter was tuned to reproduce the relative amount of rapidity gaps observed in the data and the amount of transverse energy observed at HERA. The value of this parameter used in this analysis was tuned to describe accurately the rate of events with a large rapidity gap observed at HERA [67].

3.4 Detector Simulation

After having been generated according to a specific model, the events need to pass through a detector simulation before being compared to the data, in order to take into account the response of the various components. The detector simulation is based on a fortran package MOZART [71], itself based on the GEANT [72] program used in most of the experiments in high-energy physics. They include all the information about geometry, calibration, position and energy resolution and other properties, as well as the beam gas effects on the various part of the detector (see chapter 4).

The triggers, which select "good" events as opposite as background (see chapter 5) were simulated by another fortran package: ZGANA [73]. Once the events have passed all the detector and trigger requirements, they are reconstructed (similarly for data and Monte Carlo simulation) using the ZEPHYR package [71]. At this stage, the events are called "reconstructed", by opposition to "generated" or "hadron level", before the stage of detector simulation. The comparison between reconstructed and generated events is used to correct the data for detector effects, evaluate the resolution for a specific kinematic variable and determine the quality of the reconstruction of the various kinematical variables as well as the amount of background in the data

Name	Hadronization scheme	Parton evolution model	Application
LEPTO	Lund string model	MEPS	DIS
ARIADNE	Lund string model	CDM	DIS
HERWIG	Cluster model	MEPS	DIS and photoproduction
PYTHIA	Lund string model	MEPS	photoproduction
LDCMC	Lund string model	LDC (CCFM and CDM)	DIS
RAPGAP	Lund string model	MEPS with resolved photon	DIS

Table 3.1: List of the Monte Carlo simulations with the corresponding parton evolution models and hadronization schemes.

sample. Finally, it is important to stress that the response of the detector component should not depend on the model used to generate the events. In practice, small model dependences are expected, as the components have a different response for different final state distributions and these model dependences must be taken into account in the calculation of the systematic errors. A summary of all the Monte Carlo models presented here is shown in table 3.1.

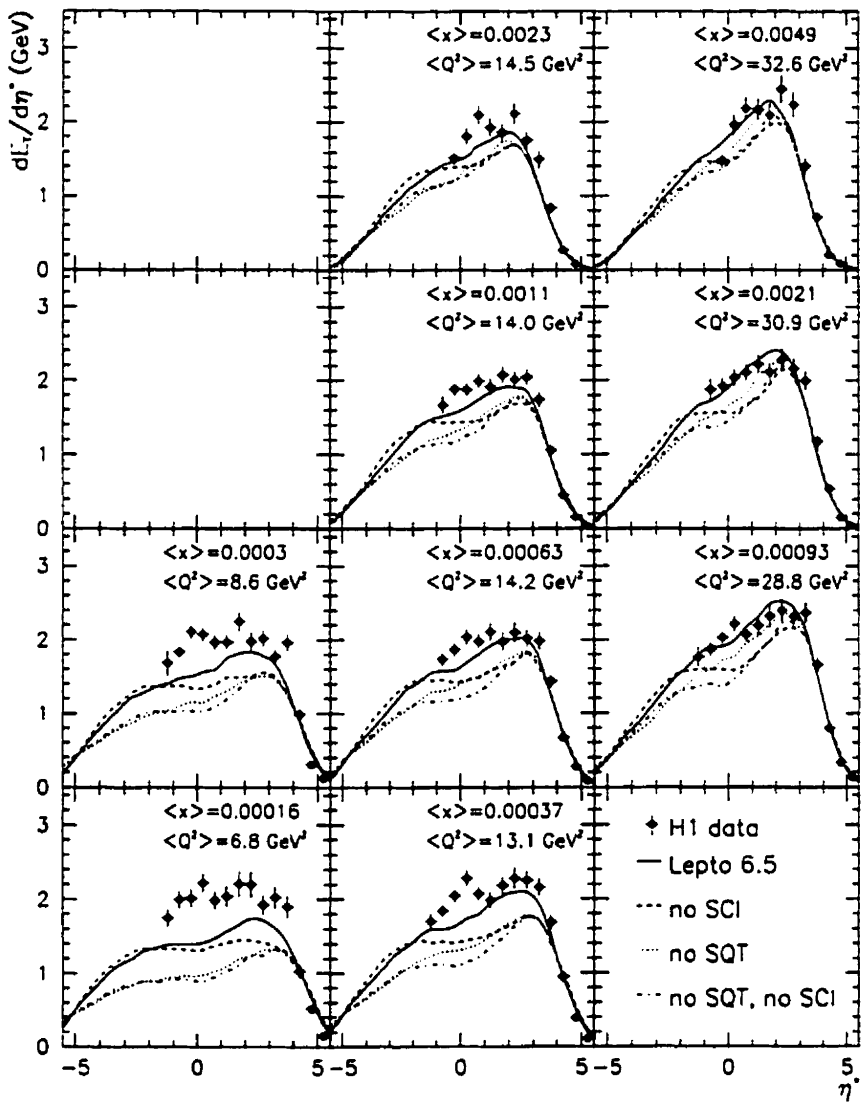

Figure 3.6: *Transverse energy flow versus pseudo-rapidity η^* defined in the hadronic center of mass frame in different bins of x and Q^2. The data come from the H1 analysis. The data are compared to the predictions of LEPTO 6.5, with and without SCI. The model has also been running with a new sea quark treatment (SQT).*

Chapter 4

Experimental Setup and Data Acquisition

4.1 The HERA Accelerator

The experimental observation and study of Deep Inelastic Scattering and related processes requires probing very small scales or distances. As the observed length is inversely proportional to the energy of the probe, looking at very small scales means using very high energies. This motivated the construction of the large accelerator HERA (*Hadron Elektron Ring Anlage*), which began in 1984 and was completed in November 1990. It is designed to accelerate electrons or positrons up to an energy of around 30 GeV and protons up to around 820 GeV (plans have been made to increase this value to around 920 GeV for the next running periods), yielding a center of mass energy $\sqrt{s} = 314$ GeV.

By reference to the fixed target experiments, the physics process going on at HERA is sometimes referred to as a collision between an electron (or positron) probe and a moving proton target. The traditional experimental studies on DIS were made on fixed target experiments like the SLAC-MIT (using the Stanford Linear Accelerator Center), BCDMS, EMC or NMC (in CERN), CCFR (in Fermilab) colliding lepton probes on fixed nuclei. The main advantage at HERA lies in the value of the center of mass energy, which is much larger than in the previous accelerators, allowing to probe much a wider kinematic range, going from very low x (below 10^{-5}) to very high Q^2 (10^4 GeV^2). The kinematic range covered by HERA and ZEUS is compared to the fixed target experiments in figure 4.1.

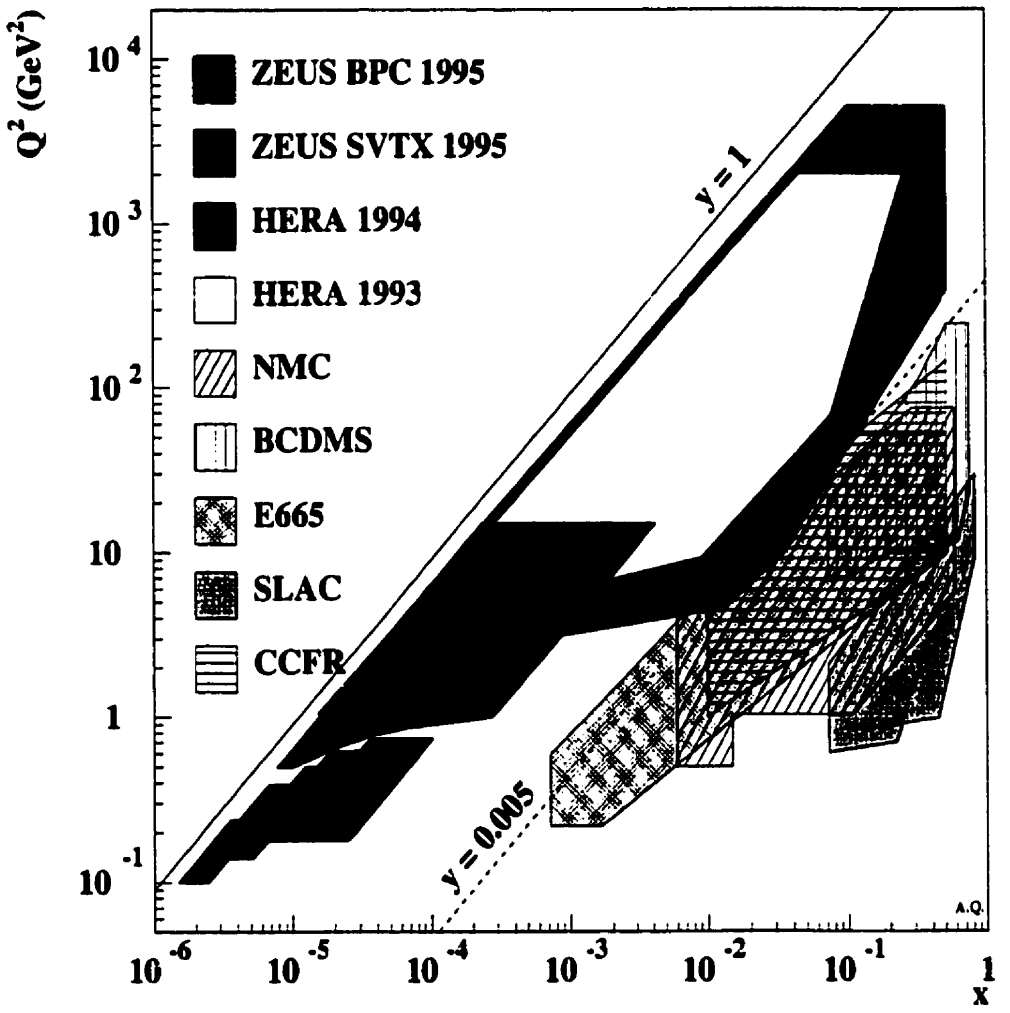

Figure 4.1: *Kinematic range covered by the HERA accelerator as a function of (x, Q^2), compared with several fixed target DIS experiments. Also shown is the kinematic range accessible in the ZEUS detector at HERA thanks to the beam pipe calorimeter (BPC) and the shifted vertex runs (SVTX).*

The two main experiments operating on the HERA accelerator are ZEUS and H1, located in the south and north hall respectively. They first took data in 1992, with an electron energy of 26.7 GeV and a proton energy of 820 GeV. Two other non-collider experiments, HERMES and HERA-B make also use the accelerator facilities to study respectively the spin structure of the nucleon and the $\mathcal{CP}$ violation in the $B^0\bar{B}^0$ system.

The HERA accelerator is schematically shown in figure 4.2. The electrons (positrons)

and the protons are accelerated in two separate rings using superconducting magnets and housed in a 6.34 km long tunnel.

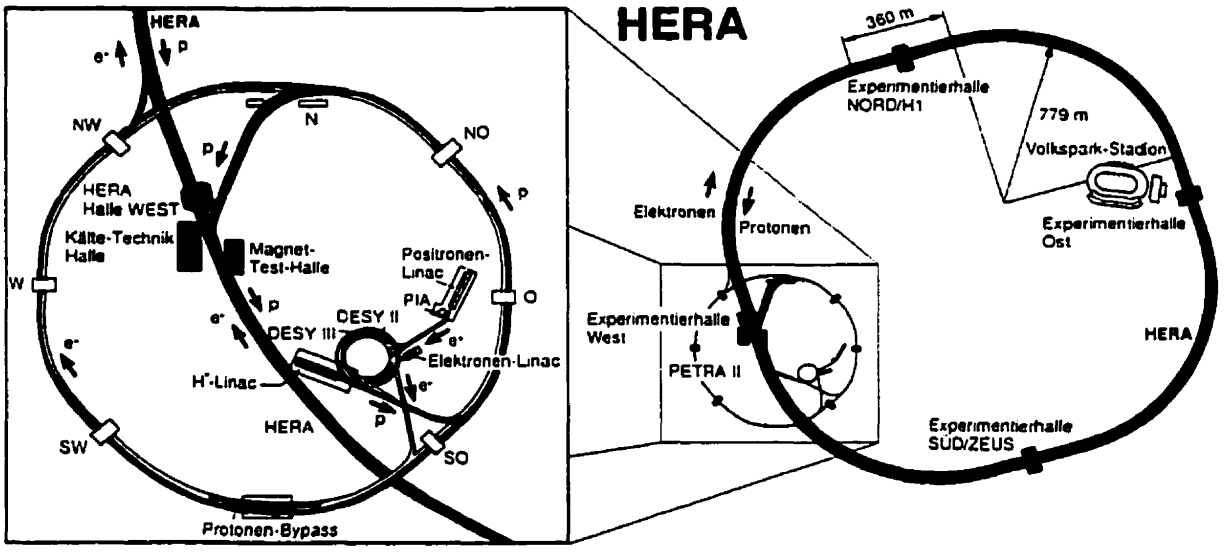

Figure 4.2: *Layout of the HERA ring and layout of the injection system.*

Starting from negatively charged hydrogen, H^-, brought to an energy of 50 GeV in the H^-, LINAC (linear accelerator), the protons are obtained by stripping the ions of their electrons before injecting them into the DESY III storage ring. At this point, there are eleven proton bunches separated by 96 ns. The protons are then accelerated to an energy of 7.5 GeV before being transferred to the PETRA (*Positronen Elektronen Tandem Anlage*), where seventy bunches are accumulated and accelerated to an energy of 40 GeV before being transferred to the HERA main ring. This is repeated until 210 bunches are accumulated. Finally the proton bunches are accelerated to an energy of 820 GeV.

The injection chain for the electrons (positrons) starts in the LINAC, where they are accelerated to an energy of 450 MeV. At this energy, they are stored in the positron intensity accumulator, PIA, where a single bunch of 60 mA is created. They are then transferred to the DESY II ring and accelerated to 7.5 GeV, before being transferred to the PETRA II ring where they are accelerated to 12 GeV. There 70 bunches are created, separated by 96 ns. Finally the electrons bunches are injected in the main HERA ring where up to 210 bunches can be accumulated (in practice, not all the electron and proton bunches are filled. There are some unpaired "pilot" bunches

generated to estimate the beam gas backgrounds). In HERA, the electron bunches are accelerated to an energy of 30 GeV.

The electron lifetime is very short ($\sim$ 2-3 hours), due to positively charged particles in the beam pipe which were attracted in the electron beam and scattered against them. This was solved by inverting the polarities of the magnets and using positrons instead of electrons (July 1994). The lifetime of the positron beam is much larger: in the order of 8 to 9 hours. Apart from charged current analyses, this change did not affect the physics process rates at HERA.

The luminosity obtained in HERA has been continuously increased since the first electron proton collision in 1992. The present analysis is based on 1995 data, when HERA delivered an integrated luminosity of 12.3 pb^{-1}. Out of this, ZEUS collected about 8 pb^{-1}, corresponding to the period during which the detector was operational and the beam conditions were good enough for data taking. After subtracting bad runs, due to faulty components or high beam gas background, we are left with 6.36 pb^{-1}. This is the luminosity used all along this study.

4.2 The ZEUS Detector

The exploration of finer structures and smaller scales in high energy physics leads to a growing degree of complexity in the design of the large detectors, together with larger collaborations and varieties of competences. The ZEUS detector, operating at HERA, represents a step forward with respect to the previous generation of detectors (like those operating at the Tevatron ring and the LEP collider), in particular in view of the size of the collaboration: more than 450 people coming from 51 institutes in 11 countries are participating in this effort. Because of its complexity, an exhaustive description of the detector would be too ambitious taking into account the scale of this work. This description can be found in [74]. ZEUS is a multi-purpose detector, covering most of the 4π solid angle around the collision point except for the small angle around the beam pipe. Because of the strong asymmetry in the collision due the boost caused by the difference of energies between the incoming proton and the positron, the detector is asymmetric: the forward direction, defined by the direction of the proton is larger and deeper than the rear region, defined by the direction of

the positron.

The coordinate system is orthogonal and right-handed: the center is taken as the nominal interaction point, the Z axis points towards the direction of the proton, Y points upwards and X towards the center of the HERA ring. The polar angle θ, defined with respect to the Z axis is therefore 0° for the proton beam and 180° for the positron beam. In most of high energy physics analyses, the measure of the polar angle is given by a relativistic invariant parameter (provided that the mass effects are neglected): the pseudorapidity η defined as: $\eta = -ln\ tan(\theta/2)$.

The various components of the ZEUS detector are pictured in figure 4.3: the inner part is made of a large wire chamber, the CTD (Central Tracking Detector) [75], surrounded by a superconducting solenoid generating a longitudinal field of 1.43 T. This detector yields information about the momenta of the charged particles and enables the collision vertex reconstruction. The FTD and RTD (forward and rear tracking detectors) are respectively in front and behind the CTD and provide the same kind of information for particles going outside the range of the CTD. Further behind the CTD, lies the SRTD [76] (Small Rear Tracking Device) which improves the detection of positrons scattered at small angles. These components are surrounded by the uranium scintillator calorimeter or UCAL [77], which measures the energy deposited by the particles passing the tracking devices. It is divided into three sections: the FCAL (forward calorimeter) in the forward direction, the BCAL (barrel calorimeter) in the central region around the threshold and the RCAL (rear calorimeter) in the rear region. A layer of scintillator tiles in front of the FCAL and RCAL, the presampler [78], improves the accuracy of the energy reconstruction by correcting for energy losses due to showers in the dead material in front of the calorimeter. The calorimeter is enclosed in a iron yoke made of several 7.3 cm thick iron plates which catches the return lines of the magnetic field and serves as a secondary absorber for the backing calorimeter (the BAC). This calorimeter measures the energy leakage of the main calorimeter. Finally, on the inner and outer sides of the yoke, limited streamer tubes are mounted outside the barrel part (BMUI, BMUO) and the rear part (RMUI, RMUO). They measure minimum ionizing particles traversing the calorimeter, mainly muons. They also allow to reject background events coming from cosmic rays and beam halos (results of the interaction between the protons and particles in the beam pipe). A forward spectrometer (FMUON) is installed in front of the FCAL

(in the positive Z direction) and allows in addition the precise determination of the muon tracks.

In the following, the components specifically used in this analysis will be shortly reviewed.

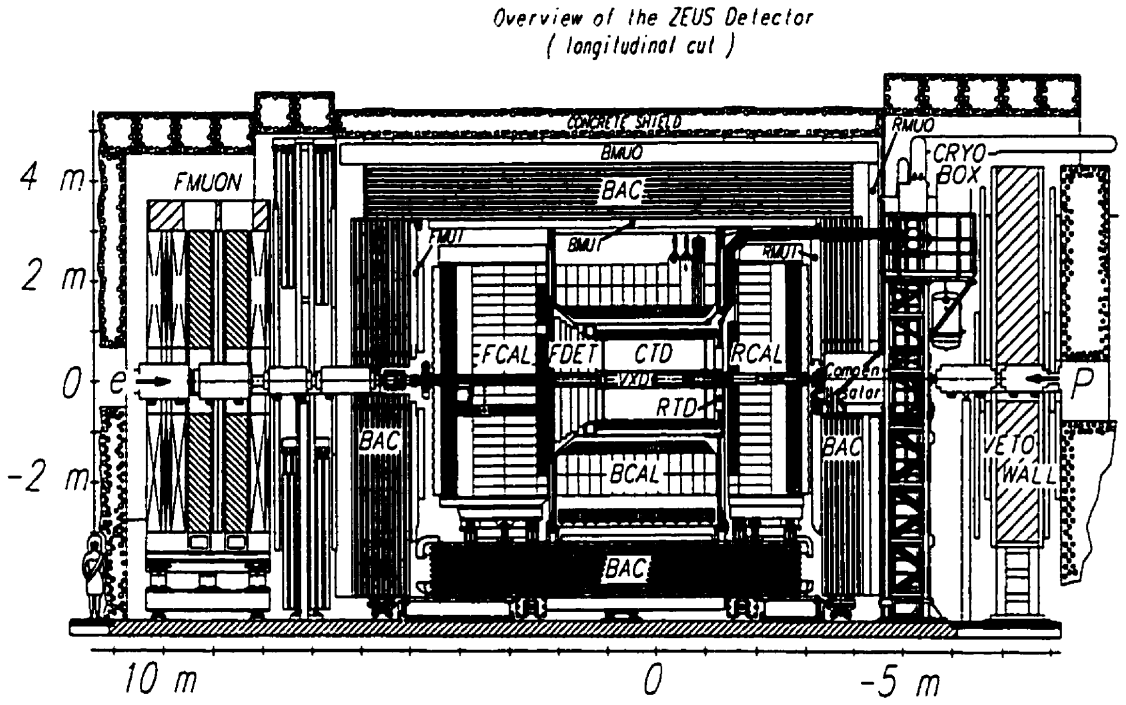

Figure 4.3: *Side view of the various components of the ZEUS detector.*

4.2.1 The Uranium Calorimeter (UCAL)

The energy reconstruction of an event observed by the ZEUS detector is performed by the uranium calorimeter. It is the main component of the detector as it determines whether an event is accepted or not and is thus at the bottom of the trigger chain which will be discussed in chapter 5. As mentioned above, the calorimeter is divided into rear, barrel and forward regions (RCAL, BCAL and FCAL) covering respectively a polar angle (pseudorapidity) region of: 128.1°(-0.72) to 176.5°(-3.49), 36.7°(1.1) to 129.1°(-0.74) and 2.2°(3.9) to 39.9°(1.0). 99.6% of the 4π solid angle around the interaction point is covered, leaving a hole of 20×20 cm^2 in the forward region and 20×12 cm^2 in the rear region to allow room for the beam pipe.

The calorimeter consists of alternating layers of absorber (3.3 mm of depleted uranium) and active material (2.6 mm thick plastic scintillator plates). Because of its stable radioactive properties, the uranium noise from the calorimeter is used to check the calibration of the data acquisition chain and the calibration of the photomultiplier tubes, which record the amount of energy deposition (known to 1-2% accuracy). The uranium calorimeter also provides compensation, i.e. the signal response for hadrons and electrons is equal ($e/h = 1.00 \pm 0.02$). This property improves the energy resolution of the hadrons: under test beam conditions, the resolution for the hadron jets has been found to be $\sigma_E/E = 0.35/\sqrt{E}$ (E is the energy of the particle in GeV). For the electron, the resolution was measured to be: $\sigma_E/E = 0.18/\sqrt{E}$.

Each of the calorimeter components is made up of modules of various lengths: there are 23 modules in the FCAL and the RCAL and 32 in the BCAL. The layout of an FCAL module is displayed in figure 4.4. These modules are further segmented into towers with a front surface area of 20×20 cm^2 in the FCAL and RCAL and 20×24 cm^2 in the BCAL.

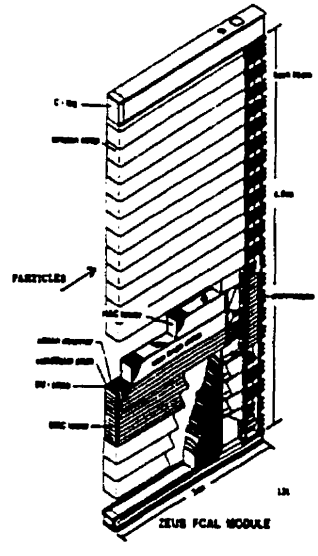

Figure 4.4: *Layout of an FCAL module.*

Each tower is further longitudinally subdivided into hadronic (HAC) and electromagnetic (EMC) sections: there is one EMC section in RCAL, BCAL and FCAL, two HAC sections in BCAL and FCAL and one HAC section in RCAL. Each section consists of one or more cells. One cell is divided into two photomultiplier tubes which read the signal given by the energy deposition in the calorimeter. The EMC sections

consist of four cells of 5×20 cm^2 in the FCAL and the BCAL and two cells of 10×20 cm^2 in the RCAL. The HAC section forms a single cell of size 20×20 cm^2. The BCAL EMC cells are projective to the interaction point. This information is particularly crucial for this analysis: as the measure of the polar angle is done in unit of pseudorapidity, the width of the cell in the (η, ϕ) plane will change accordingly to its position (here ϕ is the azimuthal angle). The crucial point is that η is a logarithmic quantity, so the cells in the forward region, in the (η, ϕ) plane are quite large. In figure 4.5, a cross section of the forward calorimeter is shown in the (η, ϕ) plane in one real event. For this example, the HAC cells only are displayed. The black circle defines a jet, which is now seen as a cluster of cells with a radius R=1 in the (η, ϕ) plane. The full dots represent the energy deposition in the cells which are associated with the jet in this event, while the small open dots represent the energy deposition in a cell which is not associated with the jet. The radius of the full and open dots is proportional to the amount of energy deposited in the cell (if the dot is not in the center of the cell, it means that the energy has been deposited in another HAC section or in a EMC cell). The increase in the size of the cell is very noticeable for $\eta > 2$. This results from a certain lack of accuracy in the position reconstruction, especially in the HAC sections, which are not segmented.

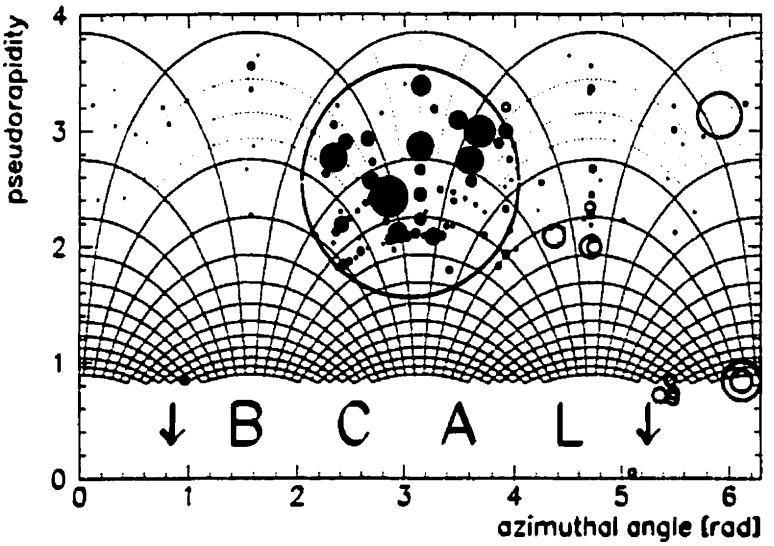

Figure 4.5: *Cross section of the HAC cells of the FCAL in the (η, ϕ) plane. The black dots represent the energy deposition in each cell. The black circle defines a cluster of cells with a radius of 1.*

The calorimeter has been designed to fully contain 95% of the hadronic (jet) energy

and 98% of the electromagnetic showers (in average for the highest energy particles produced). The parameter which defines the containment depth of the electromagnetic shower is the radiation length X_0. This is the average distance for an electron to lose 63% (1/e) of its energy (in the EMC cells, it is equal to 0.74 cm). The EMC cells have a radiation length of 25 X_0. The hadronic showers are given in units of interaction length λ. The absorption length in the three sections of the calorimeter is: 7.14 (FCAL), 4.92 (BCAL) and 3.99 (RCAL).

Besides energy and position information, the calorimeter provides a good time measurement: the time resolution for a calorimeter cell with an energy deposition above 3 GeV is $\sigma_t = 1.5/\sqrt{(E)} \oplus 0.5$ ns.

Every particle reaching the calorimeter has to cross a certain amount of inactive ("dead") material where it deposits some of its energy by interacting with the particles of the medium. Therefore the measured energy in the calorimeter is not equal to the true energy of the produced particles (the correction method in the case of hadron jets will be reviewed in chapter 7). In the front and rear calorimeters, a segmented scintillator array (the presampler detector) has been installed to correct for the energy loss of the positron: the presampler measures the multiplicity of particles created by the showering of the incoming positron in front of the F/RCAL and uses it to correct for energy losses on an event-by-event basis. The result of the presampler correction on a test sample of 25 GeV electrons is shown in figure 4.6 [78].

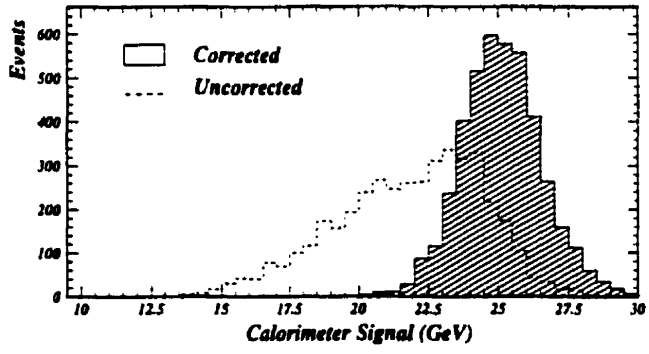

Figure 4.6: *Energy distributions for 25 GeV electrons before and after the presampler corrections.*

4.2.2 The SRTD

The small rear tracking detector is installed in front of the RCAL and cover a polar angle domain of 162° to 176°: a region mostly not covered by the other tracking detectors. Its goal is to improve the measurement of the positron energy and position, when it is scattered at low angles. This corresponds to the kinematic region of low Q^2 as $Q^2 = 4E_{e'}E_e cos^2(\theta/2)$, where $E_{e'}$ is the energy of the scattered positron, E_e, the energy of the incoming positron and θ, the angle of the scattered positron (the formula is given by the electron method, see appendix A). It is therefore particularly suitable for low-$Q^2$and low-x studies.

The SRTD consists of two layers of scintillator strips (the same material as for the ZEUS calorimeter has been chosen for the scintillators). It is divided in four quadrants in the XY plane of 24×44 cm^2. The scintillation light is transported to the photomultiplicator tubes via light guides. The scintillator strips have a fine segmentation, so that in average the position resolution of the positron is better than in the calorimeter (it is approximately 0.3 cm, while in the RCAL it is only 1 cm in average). The energy resolution of the positron in the SRTD is $\sigma = 26\%\sqrt{E_{e'}}$, where $E_{e'}$ is the energy of the positron in GeV.

4.2.3 The Central Tracking Detector

The central tracking detector is mostly used in this analysis to measure the value of the vertex position in this analysis. Generally, it enables the reconstruction of the four momentum of the charged particles tracks. The CTD consists in a cylindrical drift chamber of inner radius of 16.2 cm and outer radius of 85.0 cm and length 241 cm. It is made of 72 cylindrical layers arranged into 9 superlayers. The odd-numbered superlayers have axial wires (parallel to the beam line) while the even-numbered superlayers have stereo layers (tilted by $\pm 5°$ with respect to the beam line to provide a measurement of the Z position of the track). It is filled with a gas mixture of argon, CO2 and ethane.

The resolution of the CTD in 1995, in the (r, ϕ) plane was about 230 μm and 2 mm on the Z axis. The resolution on the transverse momentum is: $\sigma(p_t)/p_t = \sqrt{(0.005 \cdot p_t)^2 + (0.0016)^2}$, where p_t is in GeV/c.

4.2.4 The Luminosity Monitor

The luminosity is measured in ZEUS through the radiative emissions of real photons from the positron: $ep \rightarrow e'p\gamma$. This process, known as Bremsstrahlung, has been calculated by Bethe and Heitler [79] in 1934: the total cross section for a Bremsstrahlung radiation by a relativistic photon of energy E and producing a photon of energy E_γ is given by:

$$\frac{d\sigma}{dE_\gamma} = 4\alpha_{em} r_e^2 \frac{E'}{EE_\gamma}(E/E' + E'/E - 2/3)(ln\frac{4E_p EE'}{m_p m_e E_\gamma} - \frac{1}{2}) \qquad (4.1)$$

E' is the energy of the positron after the photon radiation, m_p and m_e are the mass of the proton and the positron respectively and α_{em} is the QED coupling strength.

The monitor consists in a lead-scintillator calorimeter situated 107 m down the beam line with respect to the nominal interaction point. Its resolution is $\sigma(E)/E = \frac{18.5\%}{\sqrt{E}}$. The acceptance of the calorimeter $\mathcal{A}$ (determined by Monte Carlo studies), together with the total rate R of radiative photons observed in a run and the integrated cross section (quoted above) of the process yield the total luminosity of the run by the formula:

$$\mathcal{L} = \frac{R}{\mathcal{A}\sigma}. \qquad (4.2)$$

4.3 Data taking

The data used in this study were collected by the ZEUS detector in 1995, with an integrated luminosity of 6.36 pb^{-1}, which corresponds to nearly three times as much luminosity than in 1994. No big detector upgrade has been undertaken between 1993 and 1997, but the high luminosity available from the HERA accelerator has made it difficult to trigger on every physical process, especially after 1996. The trigger channel used in this work (see next section) was prescaled in the following years (1996 and 1997), leading to numerous statistical problems in the analysis. This is why we restricted ourselves to 1995 data set in this work.

Chapter 5

Event Selection

5.1 On-line Triggers

Most of the events seen by the ZEUS detector at HERA do not come from the ep scattering but are the results of interactions between the particles in the beam and the residual gas in the beam pipe, or of halo interactions (interactions between the protons and positrons and particles trapped in the beam lines). Some of the events observed are also the result of cosmic rays passing through the detector.

Because of these high background rates, it is necessary to build a sophisticated trigger system which filters the interesting physics events out of the noise. Furthermore, these high rates require a decision time short enough, so that the readout electronics and the transfer of data to the storage system can handle them. As it is impossible to make such a decision within the HERA bunch crossing time of 96 ns, the trigger system is subdivided into three levels: the GFLT, GSLT and TLT (global first and second level triggers and third level trigger). The purpose of it is to reduce the incoming rate, which is on the order of 10 MHz to a few Hz at the end of the chain. Each trigger has to make a decision on whether to keep or to reject the event. The first level trigger uses simple algorithms which allow it to make crude decisions in a very short time. The higher levels use more elaborate algorithms and criteria so that they can make a finer selection in a slightly longer period. In order to reduce the dead time due to the decision making, the first level trigger is pipelined, which means that the data are stored in 52 buffers at a rate of 10.4 MHz before being send to the GSLT. The various components of the GFLT process in parallel the algorithms in order to select the event and send back their decision to the GFLT which then passes

the events further down the chain or clears the buffers, according to a decision based on a combination of all the components decisions. Once the event has been processed by the GFLT and the GSLT, it is transferred to the event builder which combines all the information of the various components into one event record. The data is then transferred to the third level trigger (TLT) which makes the final decision and reduces the data set to a few events per second. A simplified diagram of the three-level trigger system at ZEUS is shown in figure 5.1. Also shown is the expected reduction rate at each level.

The description of the various cuts made to select the events considered in this analysis is presented below.

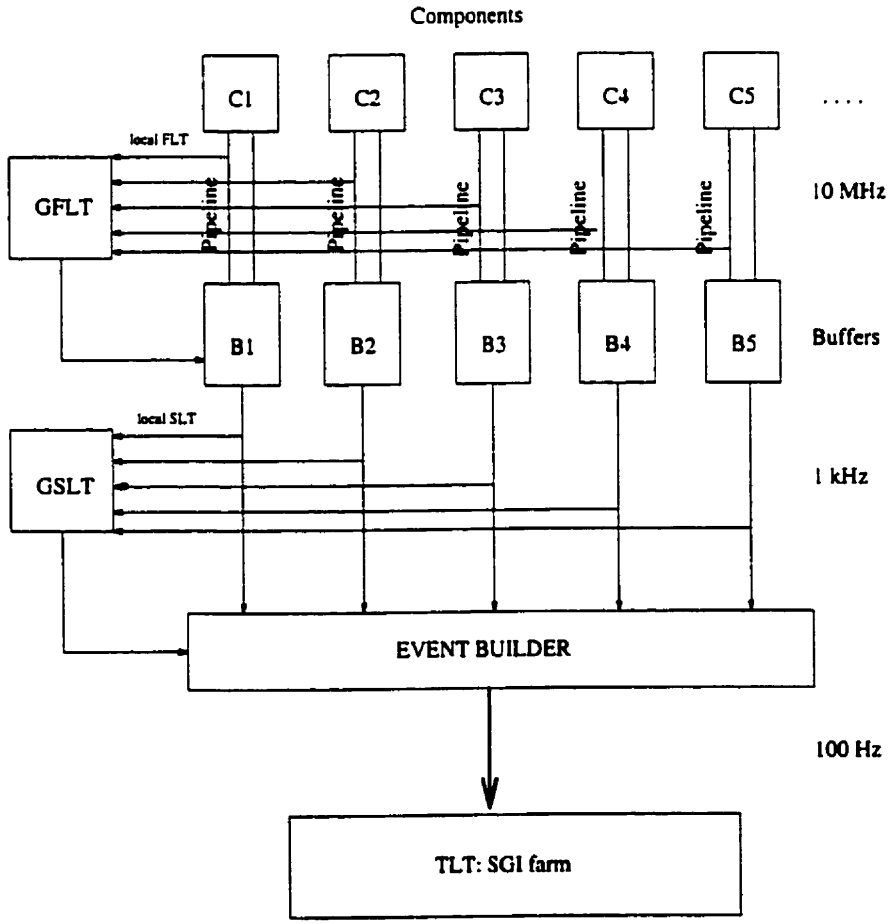

Figure 5.1: *Schematic organization of the three-level trigger system in ZEUS.*

5.1.1 GFLT

At the first level trigger, the information on the energy deposit in the ZEUS calorimeter is responsible for selecting the events which will be further processed. The calorimeter is therefore the most crucial component of the ZEUS detector as it appears at the bottom of the event selection chain. The GFLT consists of a logical "OR" between the various parts of the calorimeter, the selection being based on several energy thresholds: the total energy of the electromagnetic sections has to be larger than 4.8 GeV in the barrel (central) region (BEMC), or 3.4 GeV in the rear calorimeter(REMC). The events are also triggered if an isolated positron (defined as a set of up to four trigger towers surrounded by silent triggers) is found with an energy larger than 2 GeV. An output rate of 1 kHz is expected after the first level trigger. The rates are adjustable by varying the thresholds, depending on the luminosity and the running conditions.

5.1.2 GSLT

The main goal of the second level trigger is to reject the beam gas background to obtain a cleaner physics sample and lower rates. As for the first level trigger, the main component for the event selection is the calorimeter. In Deep Inelastic Scattering, two variables are used to trigger on the events:

- The total transverse energy of the calorimeter, calculated using the variable $\delta \equiv E - P_Z = \sum_i E_i(1 - cos\theta_i)$, where the sum is carried over all the calorimeter cells. In a fully contained DIS event, following energy conservation, the quantity δ must be equal to twice the energy of the incoming positron, that is $2 \cdot 27.5 = 55$ GeV. In photoproduction, where the positron goes undetected down the beam pipe, the δ distribution peaks at lower values, around 30 GeV. In DIS, the events are selected by applying a loose cut on this variable: $E - P_Z + E_\gamma > 24$ GeV, where E_γ is the energy of a photon detected in the lumi monitor, to take into account possible electromagnetic radiation of the positron. The $E - P_Z$ distribution in DIS and in photoproduction is shown in figure 5.2, together with the nominal value of the cut (see section 5.3).

- The calorimeter timing offers a powerful way of discriminating beam gas from physics events, thanks to the good time resolution of the measurement ($\tilde{1}$ ns for

cells with an energy above 1 GeV). In an *ep* scattering process, as the origin of time in an event is taken at the nominal interaction point, the timing difference between the FCAL and the RCAL will be small. On the other hand, if the events are coming from the proton-beam gas and given that the proton beam moves in the forward direction, the rear calorimeter will record an early energy deposition. This is why the cuts $T_{FCAL} - T_{RCAL} > 8$ ns and $|T_{RCAL}| < 8$ ns are used to reject beam-gas events. The events originating from the positron beam gas (interaction of a positron with a beam gas particle) are removed by cutting on the FCAL timing: $|T_{FCAL}| < 8$ ns. This cut provides with a very powerful way of discriminating between the beam gas background and the outcome of the *ep* collision as is shown in figure 5.3.

An additional loose cut on the Z vertex allows further beam-gas related background rejection (the event vertex reconstructed by the CTD is required to be within 100 cm from the nominal interaction point, if there are at least two tracks pointing towards it in the CTD).

The SLT reduces the event rate to about 100 Hz and transfers the various parameters of the event to the event builder.

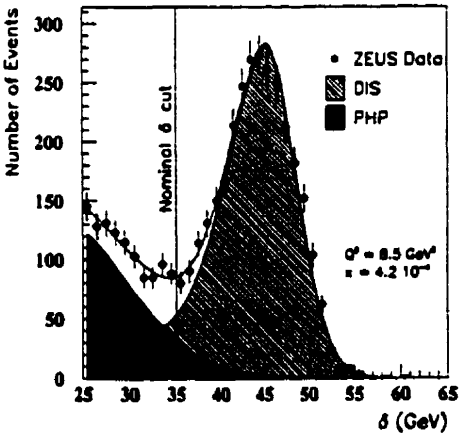

Figure 5.2: *$E-P_Z$ distribution in the ZEUS data and in two Monte Carlo simulations. The DIS Monte Carlo model is shown as a gray histogram, while the photoproduction one is in black.*

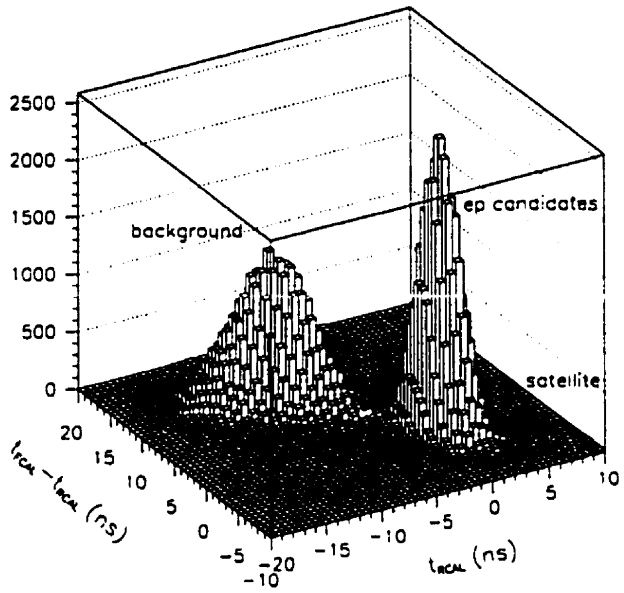

Figure 5.3: *RCAL and FCAL time distributions.*

5.1.3 TLT

The final on-line event selection is performed by the third-level trigger (TLT). At this stage, more elaborate algorithms can be used for the selection (like jet or positron finders) as the processing time constraints are relaxed. In the present study, the events triggered are required to have $E - P_Z + E_\gamma > 30$ GeV (similarly to the SLT requirement). A positron, found with the on-line algorithm such as 'Local' [80] or 'Elec5' [81], is required to be found with an energy larger than 4 GeV. Additional cuts on the position of the detected positron were applied.

In 1995, because of the high rate due to the high luminosity, some of the trigger slots used were pre-scaled. In practice, the selection requirement on the events were tightened (the box cut applied to the positron was increased from (12, 6) cm to (14, 14) cm in the (X, Y) plane - for the reason for this box cut, see section 5.2). This can cause problems when computing cross sections as the observed cross section per run might vary from one period to another and depend on the reduction factors coming from the preselection. In figure 5.4, the first plot shows how this affected the rate of events over luminosity per run: a clear step shows up for the run numbers above 12750. However, just by applying a cut on Q^2 ($Q^2 > 7$ GeV^2), the effect practically disappears, the average cross section per run is approximately flat and does not depend any more on the run number. As the kinematic region considered in this analysis is constrained to be

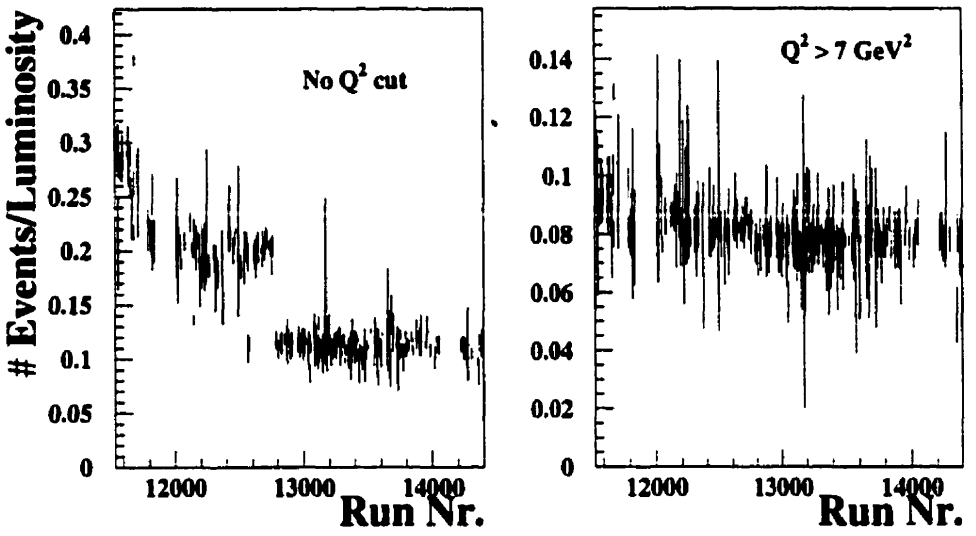

Figure 5.4: *Number of events per run, divided by the luminosity of the run (in nb^{-1}) without and with a Q^2 cut in the 1995 data, using the TLT as explained in the text.*

$Q^2 > 10$ GeV^2, the effect of the prescaling is not important.

After applying the on-line selection, the data sample in 1995 contains 7.6×10^6 events.

5.2 Positron Reconstruction

Before turning towards the off-line cuts which determine the final event sample, we shall consider one of the crucial point of this analysis, that is the reconstruction of the positron energy and position. In DIS, the accuracy of the reconstructed kinematics of an event depends on the reconstruction of the scattered lepton parameters (energy and position) since the method used to determine the kinematic variables x, y and Q^2 is based on the positron variables (see appendix A).

The reconstruction algorithm used for the positron finding is called SINISTRA and is described in detail in [82]. This algorithm is a neural network simulation based on calorimeter cells and more specifically on the concept of islands: these are clusters of

cells, gathered around a single local maximum which approximate a single particle shower in the calorimeter. Only the cells neighboring (in the 3 dimensions) the local maximum are taken into account. To increase efficiency, only well isolated islands are considered, for which all the surrounding cells have no significant energy deposit. The positron finder is based on a probabilistic method: the energies in the PMT's of the tower surrounding the local maximum constitute a "neural" network, from which a probability P_{SI} is extracted. Only electromagnetic cells are considered in the elaboration of the network. If the probability is close to 0, the algorithm assumes that the island has been created by a hadron. If the probability is close to one, it assumes that the island has been created by a positron (it can also be a photon). The energy of the positron is then the sum of the energies of the cells in the island. In this analysis, the probability cut of the algorithm, in order to identify the island as a positron, has been set to $P_{SI} > 0.9$.

The position of the positron is also calculated by an algorithm finder, except in the cases where the positron has been detected in the SRTD, in which case, only the SRTD information is taken into account. In the first case, the position of the positron is determined by fitting the energy weighted distribution of the shower. This fit becomes however inaccurate close to the beam pipe. The positron scattered at low angles (close to the beam line) can also lose part of its energy in the beam pipe. To reject those positrons, a "box" cut has been performed on the positron position in the (X, Y) plane: $|X| > 13$ cm and $|Y| > 8$ cm. The positions of the scattered positrons with hits in the SRTD and with calorimeter information only are shown in figure 5.5. The SRTD hits are shown as black dots. The shape of these hits shows the characteristic acceptance of the SRTD.

5.3 Off-line Selection

We saw up to now how the detector selects on-line the event sample and how the reconstruction of the positron energy and position is described, which allows in order to determine accurately thereafter the kinematic variables in the events and at the reconstructed level of the Monte Carlo simulation.

Another set of cuts is applied off-line to further constraint the data sample. These

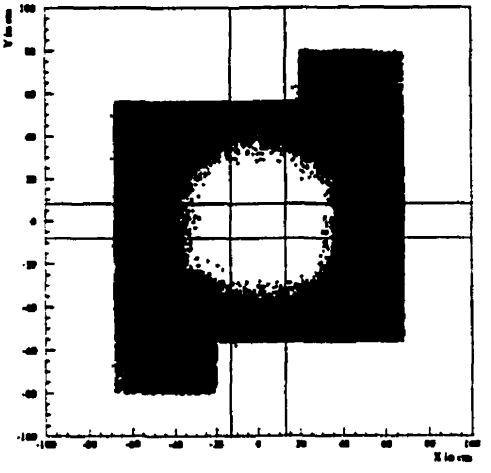

Figure 5.5: *Positron position in the calorimeter and the SRTD in the transverse (X,Y) plane. The positron is found with the SINISTRA algorithm. The gray dots represent hits for the positrons in the calorimeter, while the black dots are hits of the positrons in the SRTD.*

cuts can be roughly divided into two groups: the "cleaning cuts" which are applied at the reconstructed level only and which provide us with a cleaner, background-free sample of the experimental and simulated data, and the "phase-space cuts", which determine the value of the total cross section within the phase space of investigation. Unlike the former set, these latter cuts are applied at both levels, reconstructed and generated level and are used as well in the theoretical calculations. The first set of cuts does not appear at the theoretical level (which does not know about the detector). The loss of events due to these cuts is therefore part of the overall efficiency of the reconstruction (see section 7.3.3).

Cleaning Cuts

After the on-line selection, not all the runs and events are conserved for data processing. Some of the events (or even entire runs) are removed on behalf of criteria like holes (local losses of sensitivity) in the calorimeter during the data taking period, part of the trigger being off, high voltage or timing problems. These runs are removed in the first stage of the processing period to save computer time and space.

Some basic, general checks are then performed to remove characteristic events which can occur in parallel to the normal *ep* scattering:

- Compton events: the elastic QED compton process occurs when an electron or positron scatters off a real photon, and both particles are detected in the calorimeter. The proton is left unscattered and goes undetected in the forward beam pipe. The signature for this type of event is the presence of two electromagnetic condensates and the fact that the energy ratio of each one of the condensates to the total energy in the calorimeter is larger than a certain value (here we take 0.9). These events are removed from the sample. The effect is however quite small in the current sample (on the order of 1%).
- Events with a cosmic muon in the detector are rejected.
- Events with sparking cells, i.e. cells which can produce wrong signals due to sparks recorded in the PMT's are rejected if a high energy isolated cell is found in the calorimeter.

The other cleaning cuts performed on the reconstructed sample (data and simulation) are:

- $35 < E - P_Z < 65$ GeV: the variable $E - P_Z$ has been described earlier (section 5.1.2). The lower cut is however performed in order to remove photoproduction background, where the positron disappears undetected, and the event is not fully contained in the detector (therefore the variable $E - P_Z$ peaks at lower values). The upper cut is set to remove beam-gas events and misreconstructed vertex which can artificially increase the value of $E - P_Z$.
- To further remove beam-gas background, a cut on the vertex position has been performed: $|VTX(Z)| < 50$ cm, with respect to the nominal interaction point.
- The variable y_{el} is proportional to the scattering angle of the positron (see appendix A). In order not to mix the reconstructed positron with a low energy, forward going photon, the following cut is applied: $y_{el} < 0.8$.
- Finally, the cut on the positron position, described in section 5.2, is applied: $|X| > 13$ cm or $|Y| > 8$ cm.

Phase Space Cuts

The second set of cuts, which determine the phase space under investigation in the forward jet study and in the study of the angular correlation between jets, is:

- The positron is required to be found with an energy larger than 10 GeV.
- The y_{JB} variable (see appendix A) gives the relative amount of hadronic energy in the central region of the detector. Any measurement with a large amount of energy in the forward region and no energy in the central region would be too sensitive to the proton remnant, whose contribution lies mainly in the forward region. Besides, the y variable represents (in the QPM) the fraction of energy transferred from the leptonic to the hadronic system. A large y ensures an easier interpretation of the results in terms of perturbative physics. In the di-jet analysis (angular correlation between jets, chapter 7), the measured jets lie in the central region of pseudorapidity, therefore a cut of $y_{JB} > 0.04$ is enough to ensure a sufficient amount of energy in the central region. In the forward jet study, the jet probed lies mostly in the FCAL. To ensure a reasonable amount of energy in the BCAL, the cut has been raised to $y_{JB} > 0.1$.
- A Q^2 cut is applied in the di-jet analysis, $Q^2 > 10$ GeV^2, in order to make sure that the process can be treated in a perturbative manner. In the forward jet study, the phase space under investigation is restricted to $Q^2 > 12.5$ GeV^2 by other cuts (see section 6.2.2 for the detailed explanation).
- The goal of both analyses is to probe a wide range in x, therefore the kinematic cuts for the two analyses are:
  - Angular correlation between jets: $10^{-4} < x < 10^{-2}$.
  - Forward jet study: $4.5 \times 10^{-4} < x < 4.5 \times 10^{-2}$.

  The exact value of these cuts depends on the resolution of the x variable and the choice of the binning. This will be discussed in each of the relevant analysis chapters (7 and 8).

The phase space region under investigation in each one of the analyses is shown in figure 5.6. In figure 5.7, the main kinematic variables x, y and Q^2 are compared to the predictions of some of the Monte Carlo models discussed in the previous chapter: ARIADNE 4.08, LEPTO 6.5 and HERWIG 5.9 (the cuts are the ones used in the

di-jet analysis. The same conclusions stand when using the other set of cuts). The comparison is performed at the detector level, all cleaning cuts being included. The agreement between data and the various models is very good, which does not come as a surprise as we are only evaluating here the total cross section of the phase space under consideration, driven only by the structure function F_2 of the proton in DIS. These plots are therefore an evidence on how well F_2 is reproduced by the various Monte Carlo models.

The final samples considered, including the jet cuts will be discussed in chapter 6.

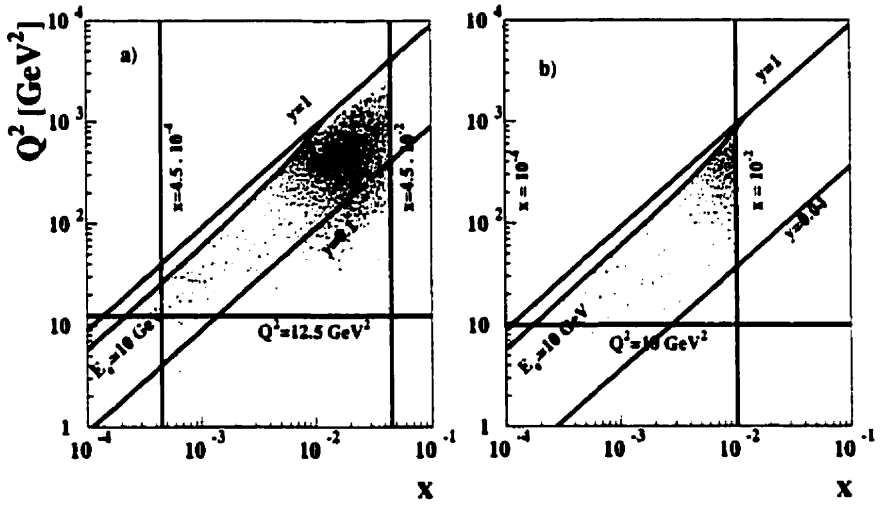

Figure 5.6: *Kinematic boundaries for the phase space region under investigations for a) the forward jet study and b) the study of the correlation between jet. The lines correspond to the kinematic cuts. Some points are outside the range. This is due to the bin size and the resolution of the kinematic variables.*

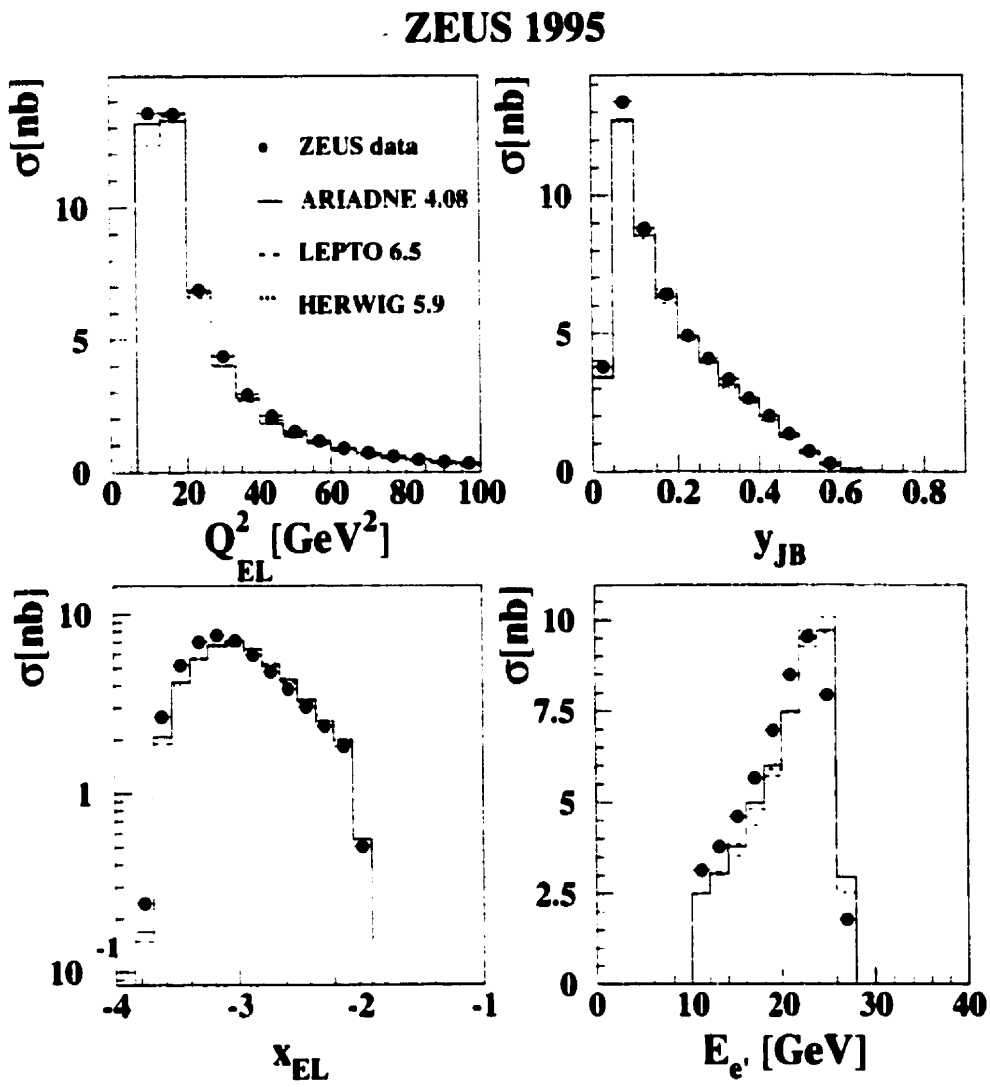

Figure 5.7: *General kinematic variables in data and Monte Carlo simulations after the cleaning cuts and within the phase space defined by the di-jet analysis (angular correlation between jets) that is* $Q^2 > 10\ GeV^2$, $10^{-4} < x < 10^{-2}$, $E_{e'} > 10\ GeV$ *and* $y > 0.04$. *The data are shown as black dots and the Monte Carlo simulations as histograms. Statistical errors only.*

Chapter 6

Jet Physics in DIS at HERA

6.1 Introduction to Jet Studies in QCD

6.1.1 Hadronic Final States Observables and Jets

Unlike lepton, photons, W or Z bosons, partons cannot be either observed directly or reconstructed in a detector. Their very existence arises from theoretical concepts, in particular from group theory. But in order to relate the experimental observations to the theory, some quantities need to be defined, which should carry information about the underlying parton processes. In addition to obvious experimental constraints (the quantities must be detected unambiguously and separated from the background present in the detector), these observables must obey some theoretical requirements without which no stable conclusion can be drawn. The most important of them is to make sure that effects due to soft (low energy) and collinear particles are small. As has been shown in section 2.3.1, these particles lead to divergent partonic cross section in the standard Feynman graphs and special techniques (cut-off, dimensional regularization) have to be applied to remove these divergencies. This is why any observable whose value depends on these soft and collinear particles cannot have an unambiguous theoretical treatment. This leads to the property of "infra-red safety": A quantity S which depends on a certain number of particles momenta $(p_1^\mu, p_2^\mu, \cdots, p_i^\mu, \cdots, p_n^\mu)$ is said to be infra-red safe if its value does not change when a particle is added which is either collinear to one of the p_i^μ or which has a vanishing momentum ($p_i^\mu = 0$).

This can be translated by the mathematical formula:

$$\mathcal{S}_{n+1}(p_1^\mu, p_2^\mu, \cdots, p_i^\mu, \cdots, (1-\lambda)p_n^\mu, \lambda p_n^\mu) = \mathcal{S}_n(p_1^\mu, p_2^\mu, \cdots, p_i^\mu, \cdots, p_n^\mu) \qquad (6.1)$$

where λ is a real number between 0 and 1.

For instance the multiplicity of charged particles in an event is not an infra-red safe quantity but the transverse momentum P_T spectrum of these particles is. The search for "infra-red safe" quantities in deep inelastic scattering is one of the most important efforts to relate theory to experiment. A standard example of "infra-red safe" quantity is the total inclusive cross section of the process, σ_{tot}: as the soft and collinear contributions cancel out when summing over all contributions, adding a soft or collinear particle does not change the value of σ_{tot}. Other quantities can be derived from the so-called "hadronic final state" of a process, which is, in HERA, one of the possible states of the hadronic energy in which the proton can fluctuate after the collision with the probe, the virtual photon. This corresponds to the projection of the proton wave function on a particular state. The sphericity of an event or the angular distribution of the energy in the detector are examples of hadronic final state quantities which implement the requirement of "infra-red safety". These quantities relate the global properties of an event to the theoretical framework and allow predictions to be tested accurately on an event-by-event basis. They can be predicted by perturbative QCD but the calculations are more difficult and less rigorous than for σ_{tot} and the influence of hadronization can be large so that a definitive statement on parton level properties is sometimes hard to make and dependent on long range physics (see chapter 2).

An attempt to improve the correspondence between quarks, gluons and the final states observables leads to the definition of the "jets". These objects are collimated subsystems in an event, often (but not always) clustered according to the energy deposition in the detector. They differ from the previous quantities by the fact that they are closely related to the history of a single parton evolution and therefore less likely to be sensitive to the non-perturbative fragmentation processes. This argument relies on the property of local "hadron-parton" duality [84], that is the observation that long range processes, which are responsible for the hadronization mechanism, create a small amount of transverse energy in the detector and therefore do not affect the topology of the high transverse momentum partons; the jets which are subsequently observed by clustering final state particles (hadrons) will therefore have a transverse momentum close to the formerly created partons.

Like the former observables, jets are known to be infra-red safe quantities and there-

fore reliable theoretical predictions on the jet cross sections or on the angular distribution of the jets can be made. Experimentally, jets are usually defined as clusters of calorimeter cells. They can also be defined in the theoretical frameworks as a clusters of four-momenta partons coming from QCD radiations or of final state particles after hadronization. But the correspondence between the real parton four-momenta and the jet energy and position is not perfect, as the jet is always defined according to an algorithm which merges the various particles in the event. The relation between the partons and the jets might strongly depend on the choice of the algorithm.

6.1.2 Standard Definitions and Jet Algorithms

One of the first attempts to use jets to relate experimental results to theoretical predictions was performed by the JADE collaboration using the "minimal invariant mass" or JADE algorithm [85]. To define a jet, this algorithm loops over any pair of particles p_i in the sample. If the minimum invariant mass of a parton pair is smaller than a fraction y of the center of mass energy of the event (ys, where s is the center of mass energy of the collision), the partons are merged into a new object:

$$(p_i + p_j)^2 \;=\; 2E_iE_j(1 - cos\theta_{ij})) \;<\; ys, \tag{6.2}$$

where i and j run over all the particles in an event. The algorithm then keeps merging particles until all pairs of "particles" have an invariant mass larger than ys. All the thus remaining pseudo-particles are then called "jets" (experimentally, at HERA, one introduces an extra particle carrying infinite momentum in the direction of the incoming beam in order to simulate the proton remnant and merge the particles close to the proton beam line to the remnant). This simple algorithm has however shortcomings when describing soft parton emissions (that is emissions of partons with a small amount of energy, often collinear to the parent parton) and leads to the creation of spurious ("ghost") jets which momenta does not coincide with an approximately collinear set of particles [86]; in figure 6.1, one can see an example of such creation of "ghost" jet: the two soft gluons are merged into a jet whose momentum does not coincide with any of the hard partons momenta. Such a behavior creates large discrepancies between the jet cross sections and the NLO predictions at parton level.

To overcome these problems, a new algorithm has been suggested: the "Durham" or "k_T" algorithm [87] which clusters particles in the same way as the Jade algorithm

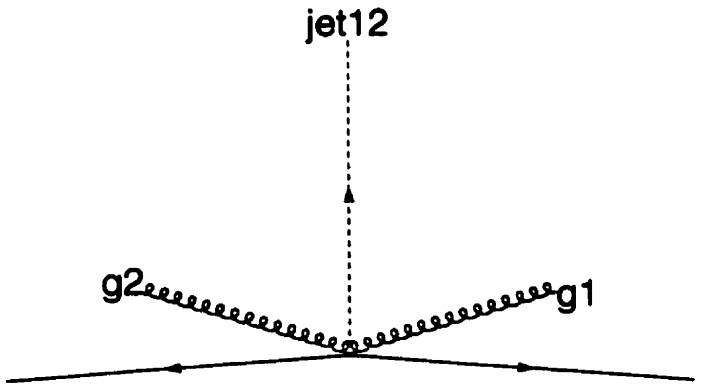

Figure 6.1: *The gluons 1 and 2 are merged into the jet 12. The event is then classified as a three-jet event, although there are only two hard partons.*

but instead of comparing the invariant mass of the two particles, it compares the minimum relative transverse energy of the two-parton system to a cut-off y_{cut}:

$$2min(E_i^2, E_j^2)(1 - cos\theta_{ij}) \; < \; y_{cut}, \tag{6.3}$$

The y_{cut} introduced here depends on some arbitrary scale of the process (which can also be taken as a constant). As in the previous case, a pseudoparticle of infinite momentum in the z direction can be introduced to simulate the proton remnant. The jet fractions which are obtained with this algorithm are in better agreement with the predictions of perturbative QCD. However, this algorithm has some experimental drawbacks which will be briefly discussed in chapter 8 and its performances in the forward region are limited.

The two previous kinds of jet algorithms belong to the so-called "cluster" type. Because of their theoretical and experimental limitations, a second class of algorithm is preferred: the "cone" type [88]. In the previous cases, all the energy available in the collision was assigned to a jet (which could possibly be the proton remnant). Here, a fraction of the total available energy, contained in a fixed solid angle, is called a jet if the total transverse energy in this solid angle is larger than some cut-off (usually a few GeV. The value of this cut-off $E_{T,min}$ and the comparison with the scale of the process will be discussed later). Therefore some of the available energy of the process can be found outside the jets.

One of the first attempts to use such algorithm was performed by Steerman and

Weinberg [89]: they classified a final state as a two-jet system if all but a fraction ϵ of the total energy available was contained in a pair of cones of half angles δ. This allowed comparisons of jet cross sections with leading order predictions. At next-to-leading order, the contribution of the three-jet systems become significant and the comparison breaks down.

A more flexible way to define a jet with the cone algorithm is to introduce a distance R, as $R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$, where $\Delta\phi$ is the difference of azimuthal angle between the object to be merged and the "seed" of the jet (which later becomes the center of the cone) and $\Delta\eta$, the difference in pseudorapidity. All the objects within a cone radius R (usually taken to 1 in HERA) are merged and a jet is found if the sum of the transverse energies of these objects is larger than the cut-off $E_{T,min}$. The kinematic parameters of the jets are defined according to the Snowmass convention [90] or "p_T" scheme:

$$E_T = \sum_{i \in Jet} E_{T,i}. \tag{6.4}$$

$$\eta = \sum_{i \in Jet} E_{T,i}\eta_i / E_T. \tag{6.5}$$

$$\phi = \sum_{i \in Jet} E_{T,i}\phi_i / E_T. \tag{6.6}$$

In this case, the convention defines the transverse momentum as

$$P_T = \sqrt{P_x^2 + P_y^2} = \sqrt{(E_T cos\phi)^2 + (E_T sin\phi)^2} = E_T. \tag{6.7}$$

The jets are therefore massless by construction. This algorithm is easy to use and fairly intuitive (in this study, the algorithm used is PUCELL. Its exact implementation of the algorithm in the current study will be described in the appendix B). It has however shortcomings from a theoretical side. The properties of the cone jets depend on the number of particles in the jet, and differences can occur when comparing partons (where there are usually a few particles only) to hadrons (where the number of particles in a jet is usually quite large) [91]. Depending on its width, a one-jet event at parton level (which occurs for instance when clustering two partons into a jet) can become a two-jet event at hadron level. To deal with this problem, a phenomenological parameter, R_{sep} has been introduced, which sets a limit to the maximal distance between two partons in a jet [92] (instead of $2R$). R_{sep} is selected

at both experimental and theoretical levels and its value is tuned in the central region of the detector so that the jet rates at both levels match.

6.1.3 Examples of Jet Physics at HERA in DIS

For more than a decade, jets have been used extensively in Deep Inelastic Scattering. These objects are ideal to probe perturbative QCD, as the amount of hadronization correction is expected to be small [93]. The determination of QCD parameters, like the strong coupling parameter $\alpha_s(Q^2)$ has been presented as an ideal field to test their performance [94]. But jets are also used to probe non-perturbative quantities like the gluon density of the proton [95] and the internal properties of jets can be used to study the different fragmentation properties of quarks and gluons, in particular through jet shapes [96].

Determination of α_s

As we saw in section 2.2.1, in the QPM (quark parton model), the process $\gamma q \rightarrow q$ gives rise to a one-parton system, that is a one-jet system if the parton is resolved as a jet. In the first order of α_s, the diagrams pictured in figure 2.3 give rise to a two-jet system. In the HERA convention, such a system is called "2+1" jet event, the "2" referring to the two hard jets and "+1", to the associated proton remnant. In the first order of α_s, two diagrams contribute to the total cross section: the Boson Gluon Fusion ($\gamma q \rightarrow q\bar{q}$) and the QCD Compton ($\gamma q \rightarrow qg$). The determination of the strong coupling α_s can be made by measuring the rate of 2+1 jet events, starting at high Q^2, where the QCDC dominates. As will be shown in the next subsection, at low Q^2 and low x, the BGF process is used to determine the gluon density of the proton. The 2+1 rate is defined as: $R_{2+1} = \frac{\sigma_{2+1}}{\sigma_{tot}}$, where σ_{2+1} is the cross section for the 2+1 jet production and σ_{tot} is the total cross section in the same phase space under consideration. At leading order, the 2+1 jet cross section can be expressed as:

$$\frac{d\sigma_{2+1}^{LO}}{dxdy} = c_{31} \cdot \alpha_s, \tag{6.8}$$

where c_{ij} contain the hard matrix elements coefficients and the contribution of the parton density. The first index i refers to the jet multiplicity including the remnant

and the second index j refers to the order of α_s. α_s is therefore the only free parameter in this expression.

However, we saw in section 2.1.3 that at leading order in α_s, the process was very dependent on the renormalization and factorization scales. In order to determine the renormalization scheme unambiguously, the cross section has to be evaluated in the NLO in α_s ($\mathcal{O}(\alpha_s^2)$):

$$\frac{d\sigma_{2+1}^{NLO}}{dxdy} = c_{31} \cdot \alpha_s + c_{32} \cdot \alpha_s^2, \tag{6.9}$$

The 2+1 jet rate in the experimental data must be corrected to the parton level in order to be compared to the NLO calculations, which are provided by adjustable user routines.

In 1994, the ZEUS experiment measured the jet rates using the Jade algorithm and compared the results to two programs: DISJET [97] and PROJET [98]. In order to describe the 2+1 jet kinematics, 5 independent kinematic variables are needed (remember than in the QPM, the 1+1 jet cross section depends only on two independent variables x and Q^2); these variables are x, y (which were already defined in the QPM), x_p, ϕ and z, where, using the same convention than in section 2.2.1:

$$x_p = \frac{Q^2}{p \cdot q} = \frac{x}{\xi}, \tag{6.10}$$

where ξ is the fraction of the proton's four momentum P carried by the incoming quark of momentum p:

$$p = \xi P, \tag{6.11}$$

$$z_j = \frac{p \cdot p_j}{p \cdot q} \approx \frac{E_j - p_{z,j}}{\sum_{i=1,2}(E_i - p_{z,i})}, \tag{6.12}$$

where p_j is the four momentum of parton j. The variable z is then taken as $z = min_{1,2}(z_1, z_2)$.

Finally ϕ is the azimuthal angle between the parton plane and the lepton scattering plane in the hadronic center of mass.

To determine the α_s parameter, the measurement was performed in the kinematic region [94]: $120 < Q^2 < 3600$ GeV^2, $0.01 < x < 0.1$ and $0.1 < y < 0.95$. An additional cut has been set on the z variable in order to remove regions sensitive to infrared

and collinear singularities: $0.1 < z < 0.9$. The values of the jet rates in different Q^2 bins can be found in [94]. The jet rates were compared to the NLO calculations as a function of several jet variables (z, x_p, the transverse momentum p_T and the invariant mass of the two-jet system m_{ij}). In figure 6.2, one can see that the data are in good agreement with the theory calculations. Finally, the determination of α_s has been performed by varying the scale Λ_{QCD} until the best fit to R_{2+1} was obtained (see figure 6.3).

The value of α_s at the mass of the Z^o was found to be (for $\lambda_{QCD} = 200$ GeV):

$$\alpha_s(M_{Z^0}) = 0.117 \pm 0.005(stat)^{+0.004}_{-0.005}(syst_{exp}) \pm 0.007(syst_{theory}),$$

which is close to the world average 0.119 ± 0.002 [99].

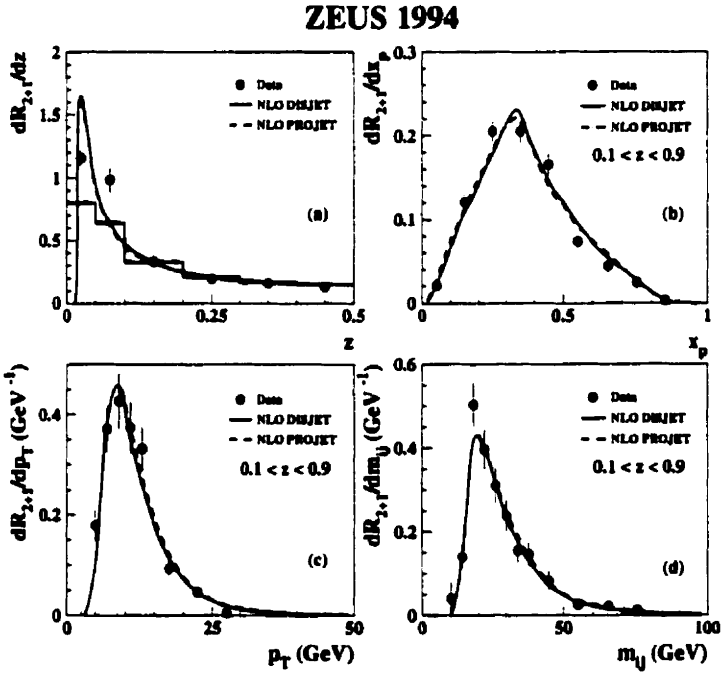

Figure 6.2: *Comparison of the 2+1 jet rate as a function of four kinematic variables with the NLO predictions: a) z, b) x_p, c) p_T and d) m_{ij}. The data points are corrected to the parton level and are plotted with statistical errors only.*

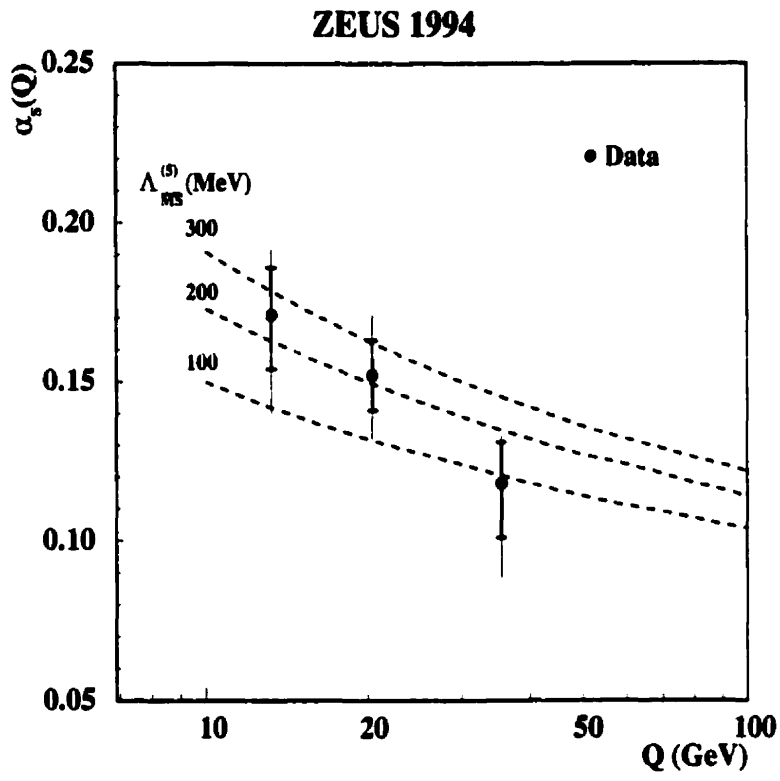

Figure 6.3: *Measured α_s in three different Q^2 regions. The dashed curves represent α_s with $\Lambda_{QCD} = 100, 200$ and 300 GeV. The statistical error bars correspond to the inner bars and the sum in quadrature of statistical and systematical errors are the thin error bars.*

Gluon Density from Jets

In the QPM, half of the momentum of the proton is carried by the gluons. Measuring the gluon content of the proton is therefore an important test of QCD. These measurements can be performed indirectly by fitting the Q^2 slopes of the structure function F_2 of the proton [100], but are limited in the x range. A direct method to measure the momentum distribution of the gluon content of the proton from the 2+1 jet rate has been proposed in [101]. At leading order in α_s, two processes contribute to the two-jet production as we saw in the previous section: the BGF and the QCDC processes. With an appropriate set of cuts, one can obtain a sample enriched with BGF events, where the gluon couples to the exchanged boson to give rise to two quark jets (see figure 2.3). At leading order, the momentum fraction of the gluon participating in the interaction, x_g is given by:

$$x_g = x(1 + m_{ij}^2/Q^2), \tag{6.13}$$

where m_{ij} is the invariant mass of the quark-antiquark pair.

At low x, the 2+1 jet cross section can be written as:

$$\sigma_{2+1} = \alpha_s(G(x,Q^2) \cdot C_{BGF} + A_{QCDC}), \tag{6.14}$$

where C_{BGF} is a coefficient which can be computed in pQCD and A_{QCDC} is the background due to the QCDC process. The strong coupling is assumed to be known and taken from the Particle Data Book (PDG) world average. The function $G(x,Q^2)$ can therefore be determined when compared to the Monte Carlo prediction using the formula:

$$G(x,Q^2)_{meas} = \frac{f(x)_{meas} G(x,Q^2)_{MC}}{f(x)_{MC}} \frac{\mathcal{L}_{MC}}{\mathcal{L}_{DATA}}, \tag{6.15}$$

where the subscript MC refers to the distribution obtained in the Monte Carlo simulation, $f(x)$ is the distribution of x_g as a function of x and $\mathcal{L}$ is the luminosity of the sample. To make this comparison, the data must be corrected to parton level as in the previous section. This involves, as we said earlier, large uncertainties due to hadronization effects.

The ZEUS collaboration has measured the x_g distribution, using the cone algorithm (radius of 1) in the laboratory system and in the center of mass frame to determine the 2+1 jet cross section [95]. The measurement covered the kinematic range: $0.01 < x < 3 \cdot 0.03$, $Q^2 > 10$ GeV^2 and $0.005 < x_g < 0.1$. The jets have been selected if their pseudorapidity η was less than 2 in the laboratory frame and -0.5 in the HCM frame (for experimental purposes: the forward region is quite difficult to handle as we are going to see in the following). Their transverse energy was required to be larger than 3.5 GeV in the laboratory frame and 4 GeV in the HCM. The background due to the QCDC process and to the QPM (where one of the jets is coming from energy fluctuations due to hadronization effects) was determined by the Monte Carlo simulation.

The H1 experiment performed a similar analysis [102] at smaller x: $0.0003 < x < 0.0015$. The x_g range was: $0.002 < x_g < 0.2$. The measurement confirmed the steep rise of the gluon density at small x (which has been seen by indirect measurements).

Recent progress on this analysis focus on comparing the data to the NLO corrections which are now available in programs like MEPJET [53], DISENT [54] or DISASTER++ [55]. These allow a more precise comparison with the 2+1 jet rate in the data. However until now, the data revealed a jet rate about 30% higher than the NLO predictions [103] (this problem will be addressed in the chapter 7).

Other attempts focus on measuring the gluon densities at lower x and x_g where indirect measurements are limited by large systematic errors. However, the various effects leading to the 2+1 jet production are not yet clear in this extreme region and a better understanding of perturbative QCD and hadronization is needed before completing this effort. The current analysis is an example of such an attempt to understand this difficult region.

Jet Shapes in DIS

Not only the jet rates and cross sections can be predicted by pQCD, but also the internal structure of the jets. Using a cluster algorithm, one can for instance increase the resolution power of the jet to observe smaller "sub-jets" within a jet. This is consistent with the theoretical concept that the jet width is not only arising from smearing in angle due to hadronization but also from the clustering of several partons within a jets [104, 105]. PQCD also predicts that the width of a jet depends on the type of parton (quark or gluon) which initiates the process; thus gluon jets are predicted to be broader than quark jets, which is due to a property of the gluon-gluon coupling which stronger than the gluon-quark coupling [106].

In 1995, the ZEUS collaboration has measured the differential and integrated jet shape for neutral and charged current in DIS, using the cone algorithm (radius 1). The differential jet shape is defined as the average fraction of the jet transverse energy that lies inside an annulus in the η-ϕ plane of inner (outer) radius $r - \Delta r/2$ ($r + \Delta r/2$):

$$\rho(r) = \frac{1}{N_{jets}} \sum_{jets} \frac{E_{T,jet}(r - \Delta r/2, r + \Delta r/2)}{E_{T,jet}(r = R)}, \tag{6.16}$$

where the radius of the cone jet is R and $E_{T,jet}(r - \Delta r/2, r + \Delta r/2)$ is the transverse energy in the given annulus. N_{jets} is the total number of jets in the sample.

The integrated jet shape is defined as the average fraction of the jet transverse energy that lies inside an inner cone of radius r concentric with the jet axis:

$$\psi(r) = \frac{1}{N_{jets}} \sum_{jets} \frac{E_T^{jet}(r)}{E_T^{jet}(r = R)}. \tag{6.17}$$

The measurement was performed for $Q^2 > 100$ GeV^2. The jets were required to have a transverse energy $E_T^{jet} > 14$ GeV a pseudorapidity lying within: $-1 < \eta^{jet} < 2$. Figures 6.4 and 6.5 present respectively the differential and integrated jet shapes compared to various Monte Carlo predictions in bins of E_T and η. The jets become narrower with increasing E_T, which is consistent with the QCD predictions. All the models are in good agreement with the data, except PYTHIA which exhibits too narrow jets. These results were compared to the $p\bar{p}$ measurements from D0 and CDF [107, 108] and to the e^+e^- measurements from OPAL [109]. The HERA jets were found to be narrower than the $p\bar{p}$ jets, which is consistent with theoretical expectations, as there is a larger gluon jet contribution at the Tevatron than at HERA. The measured jet shapes of ZEUS were similar to the e^+e^-, where the jets are mostly quark initiated. This shows the universality of the QCD radiation for a quark-initiated process.

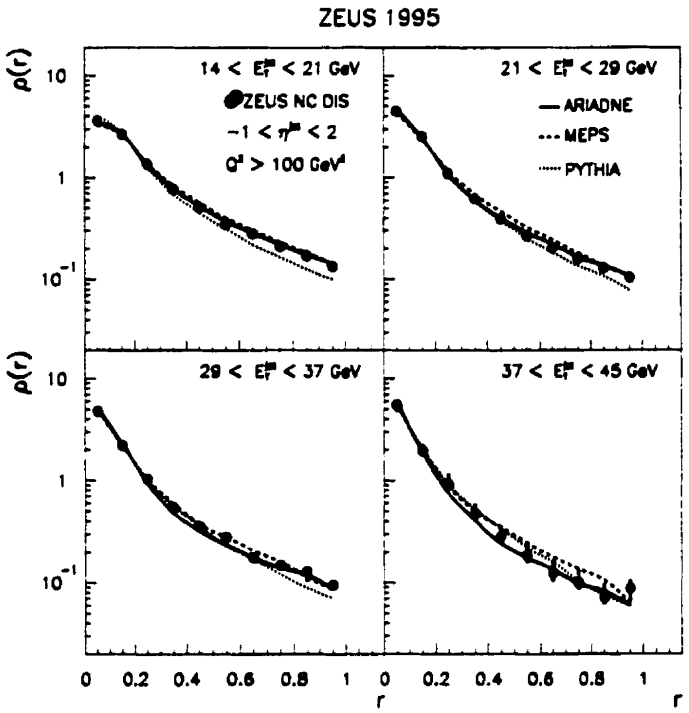

Figure 6.4: *Measured differential jet shapes corrected to hadron level, as a function of the transverse energy of the jet, E_T^{jet}, and the pseudo-rapidity of the jet, η^{jet}, compared to three Monte Carlo predictions: PYTHIA (dotted lines), ARIADNE (solid lines) and MEPS (dashed lines). The data are shown as black dots. The errors which are displayed are the sum in quadrature of the statistical and the systematic errors.*

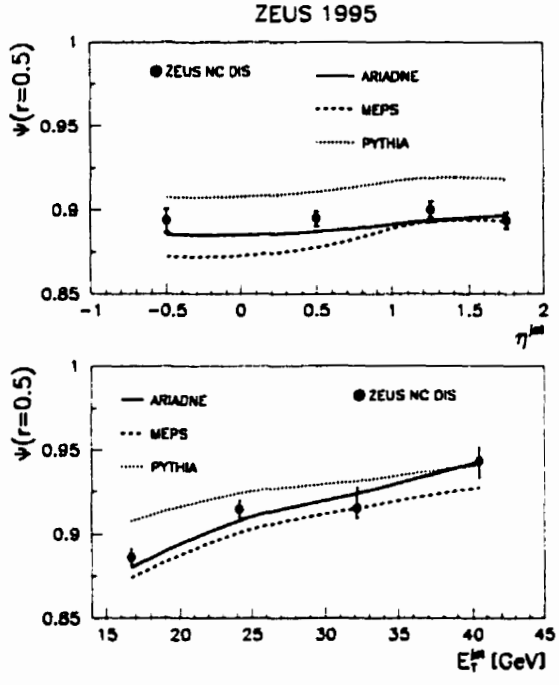

Figure 6.5: *Measured integrated jet shapes corrected to hadron level and compared to three Monte Carlo predictions: PYTHIA (dotted lines), ARIADNE (solid lines) and MEPS (dashed lines), as in figure 6.4.*

6.2 Jets as a Probe of Parton Dynamics at Low x.

The low x regime is one of the most difficult and challenging phase space regions to study for both inclusive and final state analyses. As an example, out of the three analyses presented in the previous sections, two (the jet shape analysis and the α_s determination) limit themselves to the extreme high-Q^2/high-x region, and the third one (determination of the gluon density), which probes lower Q^2 values, has experimental and theoretical uncertainties such that it cannot go down beyond $x \approx 10^{-2}$. As it was shown in section 2.3.1, the reason is that multi-gluon production dominates at these values of x and Q^2, so that the standard perturbative picture of the parton evolution breaks down. Contributions from non-perturbative effects, such as Soft Colour Interaction (see section 3.3.3) or hadronization effects make it difficult to interpret the results as pQCD effects. A better understanding of this kinematic regime is therefore mandatory to extract information on the various parameters and distributions predicted by pQCD.

The use of hard jets in this region is an excellent alternative to inclusive analyses. It should decrease the sensitivity to hadronization effects and improve the interpretation in terms of pQCD predictions. As it was mentioned in chapter 2, two jet analyses (among others) were suggested to study the parton dynamics: the angular correlation between jets [110] and the forward jet production [111]. The strategy in particular aims to discriminate between the two leading order set of equations describing parton evolution in DIS: the DGLAP and the BFKL pictures. A description of these two analyses is given below.

6.2.1 Angular Correlation Between Jets

We saw in section 2.2 that the evolution equation describes the time scale between the long-range processes such as hadronization and the short-range processes such as leading order QCD matrix element calculations. This intermediate phase space can be pictured as in figure 2.4. At low x, the BGF is the main diagram contributing to the jet production, this is why the t-channel propagator, which initiates the process, in figure 2.4 is drawn as a gluon ladder. In figure 6.6, the two possible cases of such processes are pictured. In case a), either there is no parton emission from the gluon ladder or these parton emissions do not yield a large overall transverse momentum. This is typically the case at high x, where the momentum of the gluon propagator is

large, or in the DGLAP picture, where the parton emissions with large k_T are suppressed due to the requirement of strong ordering in transverse momenta. The two leading order jets are carrying all the transverse momentum of the process and must therefore be strongly back-to-back correlated in the hadronic center of mass system in order to conserve energy and momentum. In case b), some transverse energy is carried by the gluon rungs emitted from the gluon ladder. This can happen if the k_T of the partons is randomly distributed along the gluon ladder as in the BFKL picture. There, the two leading order jets are carrying only part of the total transverse momentum of the process. In the HCM frame, this will translate as a weakening of the azimuthal correlation between these two jets as the remaining transverse momentum can be compensated by the k_T of the higher order partons. This effect must increase as x (or x_g) becomes smaller because multi-gluon emissions are becoming predominant. Therefore, at small x, instead of having $\Delta\phi$ distributions peaked at the value of π, broader azimuthal distributions are expected. However, in practice, these distributions can also be accounted for NLO processes where a third hard jet can balance the transverse momentum of the other two leading order jets.

Calculation at leading and next-to-leading order have been performed in [110] using the PROJET package and the $\Delta\phi$ dependence of the di-jet cross section has been evaluated for $Q^2 > 10$ GeV^2. The jets were required to have $E_T^2 > 10$ GeV^2. For the NLO calculations, the third jet was required to be $E_T^2 < 10$ GeV^2, so that it simulates a relatively soft parton. It should be added that in these calculations, both the NLO and the BFKL predictions have been made at parton level without using a jet algorithm. A one-to-one correspondence between the parton and the jet was assumed. Figures 6.7 show the dependence of the differential cross section on the variable ϕ for two values of Q^2 and three values of x. In the region where $\Delta\phi \sim \pi$, the two jets usually have low momenta. This leads to larger hadronization corrections and this is why this region is not shown in the figures. The NLO distributions are falling sharply for $\Delta\phi < 2.8$, whereas the BFKL curve exhibits much broader tails. At the parton level, this distinctive effect can be interpreted as a clear signal of the BFKL dynamics.

In order to reproduce this effect in the data, the cone algorithm as described earlier is used with a radius of 1, in order to determine the two leading order jets. The jets are found in the laboratory frame and then boosted to the HCM frame to find out the values of the kinematic parameters (E_T, E and η) in this frame.

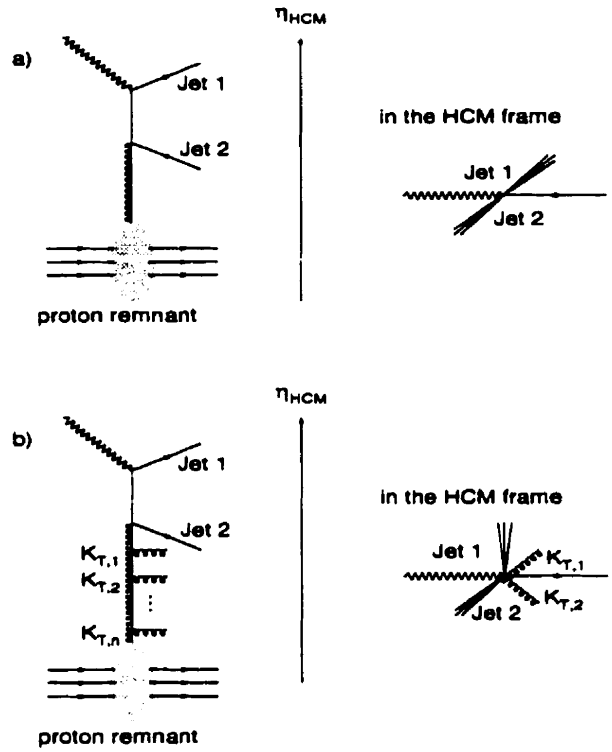

Figure 6.6: *a) Dijet production through the BGF process. No initial state parton emission, or no k_T production in the gluon ladder: in the HCM, the two jets are back-to-back correlated. b) Dijet production with parton emissions. The overall k_T produced by the initial state emissions balances the k_T of the two leading order jets in the hadronic center of mass frame.*

In the DGLAP scheme, the parton emissions are ordered both in x_{jet} and in k_T, while they are only ordered in x_{jet} in the BFKL scheme. In order to select the two leading order partons, the jets are therefore sorted in the HCM according to their pseudorapidity η (which is equivalent to sort the jets by x_{jet}) and the two jets with the highest η are marked as the leading order partons. These jets are compared if they obey the following criteria:

- In order to have reasonably hard jets, the transverse momentum in the laboratory frame is required to be $E_{T,1} > 4$ GeV, and $E_{T,2} > 6$ GeV. The indices 1

and 2 are not specific for the jet selected. This only means that one of the jets is required to have a transverse energy larger than 4 GeV and the other, larger than 6 GeV. The reason of this asymmetric cut will be explained in section 7.3.2.

- As the ϕ distribution is observed in the HCM frame, the transverse energy of the jets in this frame is also required to be within $E_{T,1}^{HCM} > 4$ GeV, and $E_{T,2}^{HCM} > 6$ GeV.

- In order to reconstruct accurately the jets, in a region far enough from the proton remnant and where the position resolution is not degraded by the size of the calorimeter cells, the pseudorapidity η of the jets in the laboratory frame is required to be within $-2 < \eta < 2.2$.

- A cut is applied on the difference between the pseudorapidity of the jets in the HCM frame: $\Delta\eta^{HCM} > 1.2$. The reason of this cut will be explained in the next chapter, as well as its consequences on the theoretical predictions.

- Finally, to make sure that the selected jets are coming from the box diagram and not from radiations from the gluon ladder, a requirement is set on the pseudorapidity of the jets in the HCM frame: $\eta^{HCM} > 0$.

The last cut was suggested by theorists working the subject [112]. Unfortunately, there is not so far a BFKL Monte Carlo model which could optimize these cuts. All the information relies on theoretical calculations.

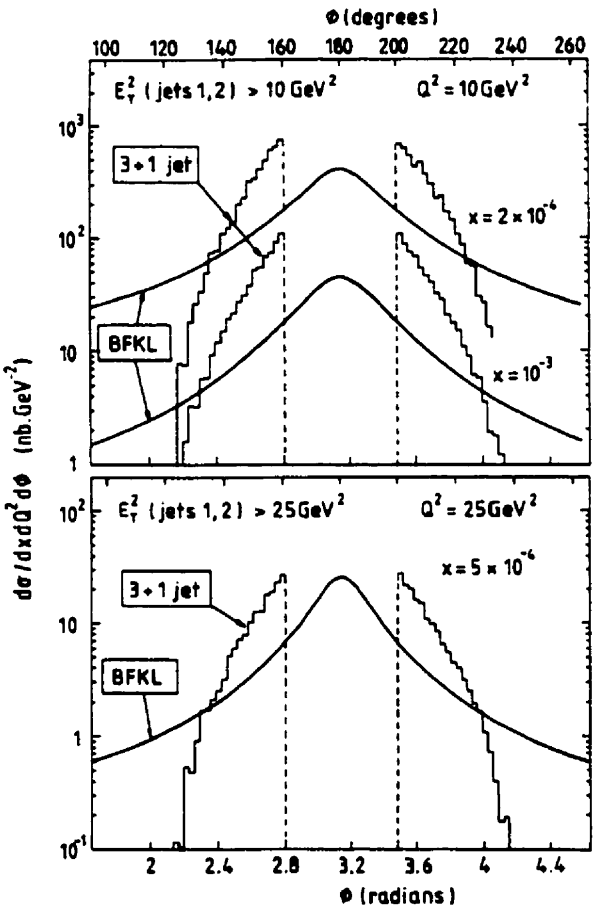

Figure 6.7: *Differential two-jets cross section as a function of $\Delta\phi$, the difference in azimuthal angle between the two jets in the HCM frame. The curves represent the BFKL predictions while the histograms show the NLO calculations (using the program PROJET) for various values of x and Q^2.*

6.2.2 Forward Jets

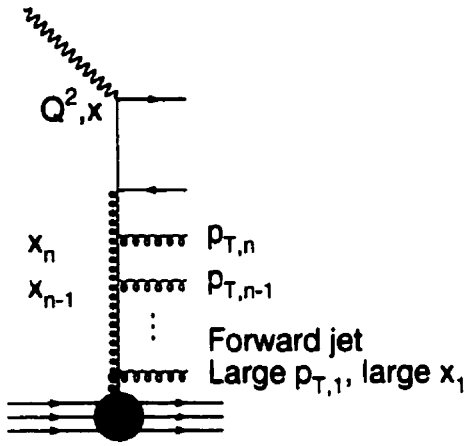

Figure 6.8: *Forward jet production in DIS. p_T is the transverse momentum of the emitted parton and x_i is the fraction of longitudinal momentum. The forward jet is the parton with the largest x_{jet}.*

A second attempt to probe the parton evolution is the study of forward jets at low x. The method was proposed by Mueller in [111] and studied by the Durham group in [38]. In order to probe the parton dynamics at low x, the strategy is to look at the bottom of the gluon ladder pictured in figure 6.8. The last parton emitted is treated inclusively as a probe of the dynamics of the process. The idea comes once again from the difference in the ordering of the parton emissions between the DGLAP and the BFKL schemes. In the DGLAP picture, the last parton emitted has the smallest transverse momentum p_T of all (because of the strong ordering in p_T requirement) and the largest fraction of longitudinal momentum, x_{jet}. This also means that this jet is the most forward one, as $x_{jet} = p_z/p_{proton} = E \cdot cos\theta/p_{proton}$. The larger x_{jet} is, the closer the process is to the so-called high energy limit [113]. This can be shown by looking at the parton-photon center of mass energy (x_1 is the fraction of longitudinal momentum of this first jet, $p_1 \simeq x_1 p$ is the momentum of the parton 1 and q is the momentum of the photon):

$$s_{\gamma 1} \simeq 2p_1 \cdot q \simeq 2x_1 p \cdot q \simeq \frac{x_{jet}}{x} Q^2. \tag{6.18}$$

Therefore, if $x_{jet} \gg x$, the partonic cross section becomes very large. Besides, because of the ordering in the transverse energy of the emitted partons, if this parton has a

momentum p_T comparable to the scale of the process (here Q^2), this emission will be suppressed in the DGLAP scheme. On the other hand, the BFKL scheme predicts a random p_T distribution along the gluon ladder. The evolution is nevertheless strongly ordered in x_{jet}. The last emitted parton, with the highest x_{jet} can therefore have a p_T^2 of the order of Q^2.

Therefore the two following requirements suppress the phase space for a DGLAP-based evolution, while they enhance the contributions of a possible BFKL-type process:

- $p_T^2 \sim Q^2$
- $x_{jet} \gg x$

The forward jet cross section has been formally calculated in [113] and numerical estimations have been performed. It has been found to be:

$$\sigma_{forward\ jet} \sim \left(\frac{x_{jet}}{x}\right)^{4ln2\frac{\alpha_s N_c}{\pi}} \left(\frac{Q^2}{p_t^2}\right)^{\mu}. \tag{6.19}$$

Where μ is an exponent depending on the factorization scale of the process and N_c is the number of active colors. Let's note that the forward jet cross section grows like the gluon density: in section 2.3.3, we showed that $xg(x,Q^2) \approx x^{-\lambda}$, with $\lambda = \frac{12\alpha_s}{\pi} ln2 \approx 0.5$.

Expanding the exponent in eq. 6.19, we obtain:

$$\left\{1 + \frac{\alpha_s N_c}{\pi} 4ln2 \log\left(\frac{x_{jet}}{x}\right) + \frac{1}{2}\left[\frac{\alpha_s N_c}{\pi} 4ln2 \log\left(\frac{x_{jet}}{x}\right)\right]^2 + ...\right\} \left(\frac{Q^2}{p_t^2}\right)^{1+\mu}. \tag{6.20}$$

This is the standard perturbative expansion in terms of leading $\ln(\frac{1}{x})$ predicted by the BFKL equation. The first term of this expansion ('1') corresponds to the emission of a forward parton on top of the fixed matrix elements calculated from the quark box to which the photon couples. It is therefore similar to the NLO calculations in the DGLAP perturbation theory. The second term corresponds to the emission of a second parton between the forward going parton and the quark box. Each higher term of the perturbative expansion corresponds to a new gluon emitted between the

forward jet and the leading order partons coming from the matrix elements. Because it is a leading $\ln(\frac{1}{x})$ expansion, each new gluon is emitted far in x_{Jet} (and therefore in rapidity) from the previously emitted partons, leading to parton emissions strongly ordered in rapidity. Nevertheless, because of the NLO corrections to the BFKL kernel (see chapter 2), we expect that this strong ordering in rapidity breaks down. The results would be a total cross section growing as: $\sigma \sim x^{-(\lambda+\Delta\lambda)}$, where $\Delta\lambda$ is the outcome of the NLO corrections (large and negative) to the BFKL kernel. This would result in a much flatter and smaller cross section. However, so far, the NLO corrections have only be evaluated for the total cross section [32]. Their application to specific hadronic final states is not yet certain and the above estimation can at best be taken as a rough estimate.

Experimentally, to select the forward jets in the data, the cone algorithm with a radius 1 is used, as in the previous analysis, and the following cuts are set:

- $x_{jet} > 0.036$. This cut selects high energy jets at the bottom of the gluon ladder (it also makes sure that we probe higher orders of α_s). At low x, it enhances the contributions from a BFKL-type evolution. The value of 0.036 is set to be roughly consistent with the other kinematic cuts (E_T and η).

- $0.5 < {E_T}^2/Q^2 < 2$. Together with the previous one, this cut suppresses the phase space for DGLAP evolution. Within the cone algorithm, the transverse energy and the transverse momentum are equal.

- The jet selected is required to lie in the target region of the Breit frame[1]: $p_{z,Breit} > 0$. This cuts prevents the leading order jets from the quark box from contributing to the jet cross section. This can happen at high x, where $x_{jet} \sim x$.

- To make sure that we nevertheless observe hard jets, a cut on the transverse energy $E_T > 5$ GeV is necessary.

- The pseudorapidity is required to lie within the range: $0 < \eta < 2.6$. The lower cut is not really needed as the requirement on x_{jet} selects jets with $\eta > 0.5$. The upper cut is needed for experimental reasons. First, the jets have to be fully contained in the calorimeter and the last cell is at $\eta \sim 3.7$. Besides, the correlation between detected and true jets is getting worse at $\eta > 2.6$ (see figure 6.9). This is due to the bad position resolution in the forward region of

[1]For a definition of the Breit frame, see appendix C

the calorimeter, as an effect of the granularity of the cells (see section 4.2.1). Finally, because of the original boost, the proton remnant deposits a lot of energy in the forward region and the separation of the forward jet from the remnant is quite difficult in this region. This issue will be treated in the next section.

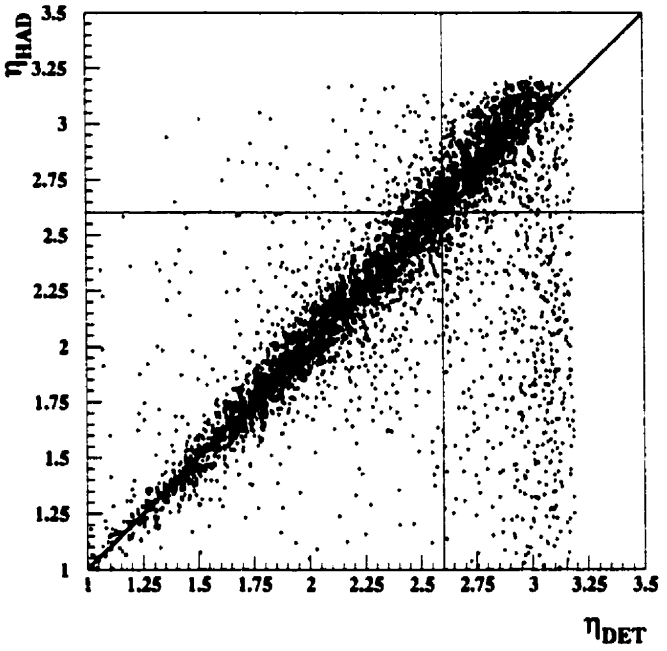

Figure 6.9: *Correlation between the pseudorapidities at the detector level and at the hadron or true level. All the cuts (section 6.2.2) but the one on the jet pseudorapidity have been applied. The vertical and horizontal lines correspond to the value of the η cut. At $\eta_{Jet} > 2.6$, the correlation starts to worsen.*

6.3 Some Properties of the Jets in the Forward Region

The forward region of the ZEUS calorimeter is known to be very challenging to study and it is excluded from most of the physics analyses made at ZEUS. The three main reasons for this are: the hole in the detector due to the beam pipe and the energy loss which follows, the presence of the proton remnant, and the reconstruction accuracy

which is limited by the coarse granularity of the calorimeter cells. As we want to study jets which can be very close to the beam pipe, it is important to check that their behavior can be predicted from a theoretical and an experimental sides. These jets might also receive contributions from the proton remnant. One of the goals of this section is to estimate how large these contributions can be. Unfortunately, there is no model for the behavior of the proton remnant, the dissociation being bound by too many uncertainties (the boost, the effect of the hadronization...). So far, it has not been possible to make a quantitative prediction for the average amount of energy which arises from the remnant. The following arguments will therefore be mainly qualitative; their purpose is to convince ourselves that the measured jets are not merely energy fluctuations of the remnant. The study will be performed with the cuts defined in previous section (forward jets).

6.3.1 Energy Flow around the Jets

The first check which is performed is the study of the energy flow around the forward jets. All the calorimeter cells with a significant energy deposition are considered with respect to the position of the center of the jet and averaged over all events with a forward jet. As the cone algorithm has a radius defined as a function of the pseudorapidity η and the azimuthal angle ϕ, it is natural to express the position of the calorimeter cells as a function of these two variables. In the following, the difference between the pseudorapidity of the jet and the cell pseudorapidity is called $\Delta\eta$: $\Delta\eta = \eta_{jet} - \eta_{cell}$. The forward direction (direction forward with respect to the jet position) lies therefore in the negative values of $\Delta\eta$. The difference between the azimuth of the jet and the azimuth of the cell is called $\Delta\phi$: $\Delta\phi = \phi_{jet} - \phi_{cell}$. In figure 6.10, the transverse energy flow in the 1995 data with respect of the jet position is plotted as a function of $\Delta\eta$ and $\Delta\phi$. This transverse energy flow is plotted in various η bins, where η is the pseudorapidity of the forward jet. The energy belonging to the jet is plotted as the shaded area. Up to η of 2.5, the jet appears to be very well separated from the energy activity in the rest of the calorimeter (represented by the blank histogram). For $\eta = 2.5$, the jet can still be fairly well separated from the rest of the event, but this becomes more difficult for $2.6 < \eta < 2.8$. There, the jet looses its sharp peak structure and becomes much wider. An excess of energy which does not belong to the jet is visible in the region forward to the center of the cone and can barely be distinguished from the energy inside the jet.

To better estimate the effect of the amount of energy forward to the jet's center (which could possibly arise from a contribution of the proton remnant), a cross section of the previous plot is shown in figure 6.11. Here, the transverse energy flow is plotted versus $\Delta\eta$ only, the amount of transverse energy being integrated over the full $\Delta\phi$ range (full line histogram) and the region where $-1 < \Delta\phi < 1$ (dashed histogram). The characteristic peak structure of the transverse energy appears in all the η bins, indicating the jettiness of the object found. However, for $\eta > 2$, a second peak appears on the left of the jet axis (that is in a region which is forward with respect to the center of the jet). This second peak moves closer to the jet center in the high η regions and finally merges with the primary peak for $\eta > 2.6$. This structure is very well reproduced in the Monte Carlo simulation (see figure 6.12).

Interpreting this feature in terms of proton remnant is not so straightforward. One has first to remember the structure and the shape of the calorimeter cells (see chapter 4). In the forward region, these cells becomes wider and the gap between two cells becomes large. The position of a given cell in the calorimeter is given from the data base by the center of this cell. Therefore, the wider in angle the cell is, the larger the uncertainty on the position becomes. Part of the gap between the two peaks can come from the empty space which lies between two cells in the forward region of the calorimeter . When the jet is very forward (typically $\eta > 2.6$), it merges these very wide cells and the position reconstruction becomes unaccurate, which may explain its flattening and widening. However the contribution from the energy of the proton remnant in the forward region is also quite important and must be taken into consideration for a correct interpretation of the results.

6.3.2 Jet Shapes

The contribution of the proton remnant and the accuracy of the jet reconstruction should also reflect themselves in the internal structure of the jet. A usual way to look at the internal properties of the jet is to study the jet shape. In section 6.1.3, the integral and differential jet shapes were defined: they picture the width of the jet and reflect how well it is collimated. At low x, we expect a much more important contribution from gluons than from quarks, and therefore, for reasons quoted above, we expect the jets to be quite large.

In figures 6.13 and 6.14, the differential jet rates in E_T and η bins are plotted for the data and the Monte Carlo simulations. The differential jet rate is defined as above in equation 6.16. All the cuts of the forward jet analysis, defined in section 6.2.2, have been applied (except for the η cut on the last plot on figure 6.14). The first noticeable feature of these plots is that the differential jet shape is more irregular than in the plots presented in section 6.1.3, especially at high η and low E_T. Furthermore, the amount of energy observed in the higher r/R bins of each plot is much larger than in the previous plots, meaning that a substantial amount of energy lies outside the cone radius of 1. However, the core of the jets is still hard which means that the energy is in average collimated around the center. Another characteristic feature of these plots is that the jets tend to becomes broader when they are more forward (cf. figure 6.14) and less energetic (cf.figure 6.13). This is in agreement with the QCD predictions for jets.

These features are quite well reproduced by the ARIADNE model, while LEPTO does not describe the differential shapes as well and tends to find more energy in the tails of the jets.

The integrated jet shapes are shown in figures 6.15, 6.16 and 6.17. The same definition as in section 6.1.3 is taken. In figure 6.17, the integrated jet shape at a radius $r = 0.5R$ is shown (half-way to the full width of the jets). These plots lead to the same conclusion as the previous ones: the width of the jet decreases with increasing pseudorapidity and with increasing transverse energy. The increase of the integrated width with increasing E_T is well in agreement with the QCD prediction [114]. Once again, the ARIADNE Model reproduces the jet shape fairly well, whereas the LEPTO Monte Carlo yields systematically wider jets. If the width of the jet is a signature of its hardness, and therefore its relation to the hard perturbative physics, the data and ARIADNE exhibits more perturbative-like jets than LEPTO.

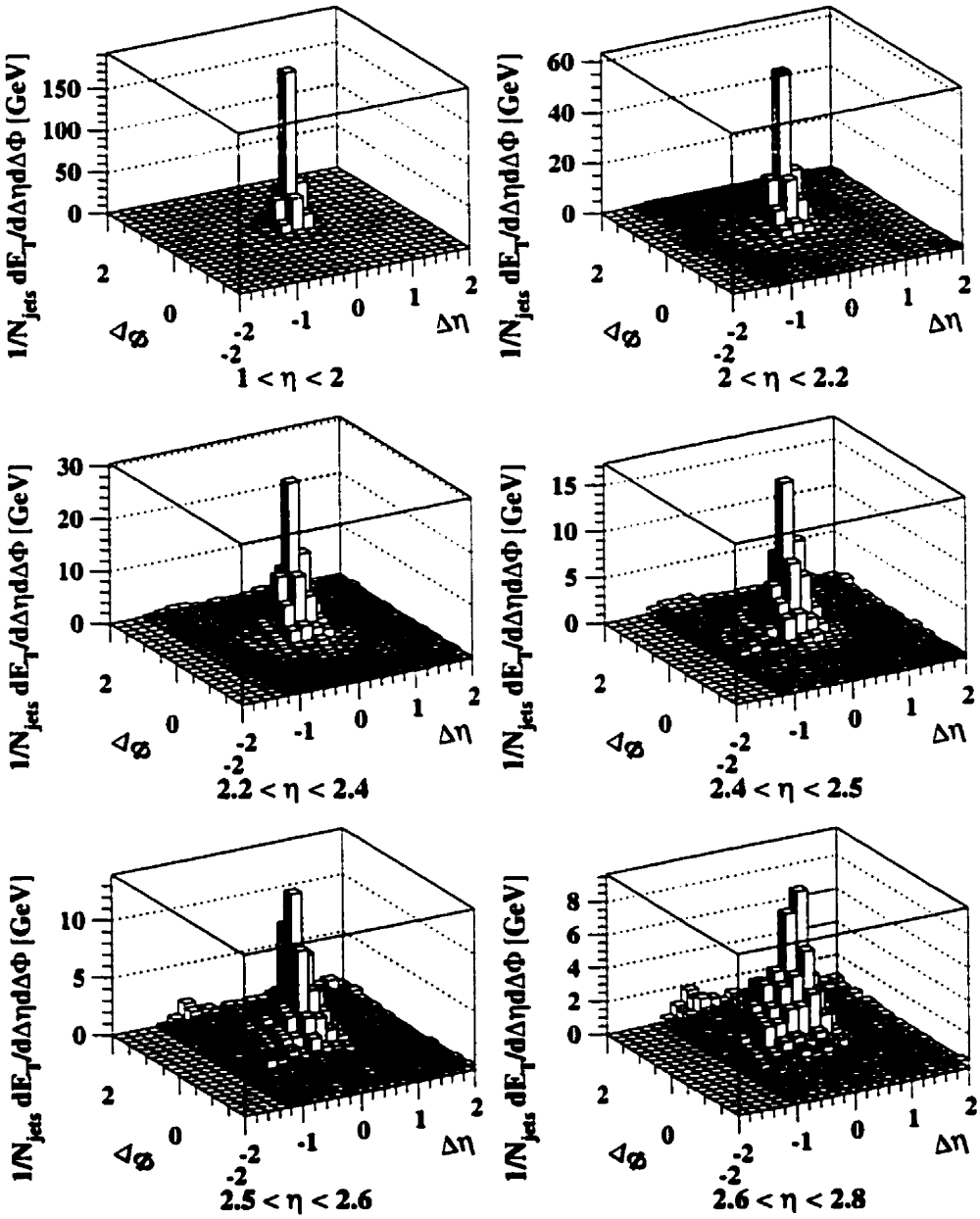

Figure 6.10: *Transverse energy flow around the center of the jet as a function of $\Delta\eta$ and $\Delta\phi$. The shaded region represents the cells belonging to the jet, while the open histograms picture all the cells in the calorimeter.*

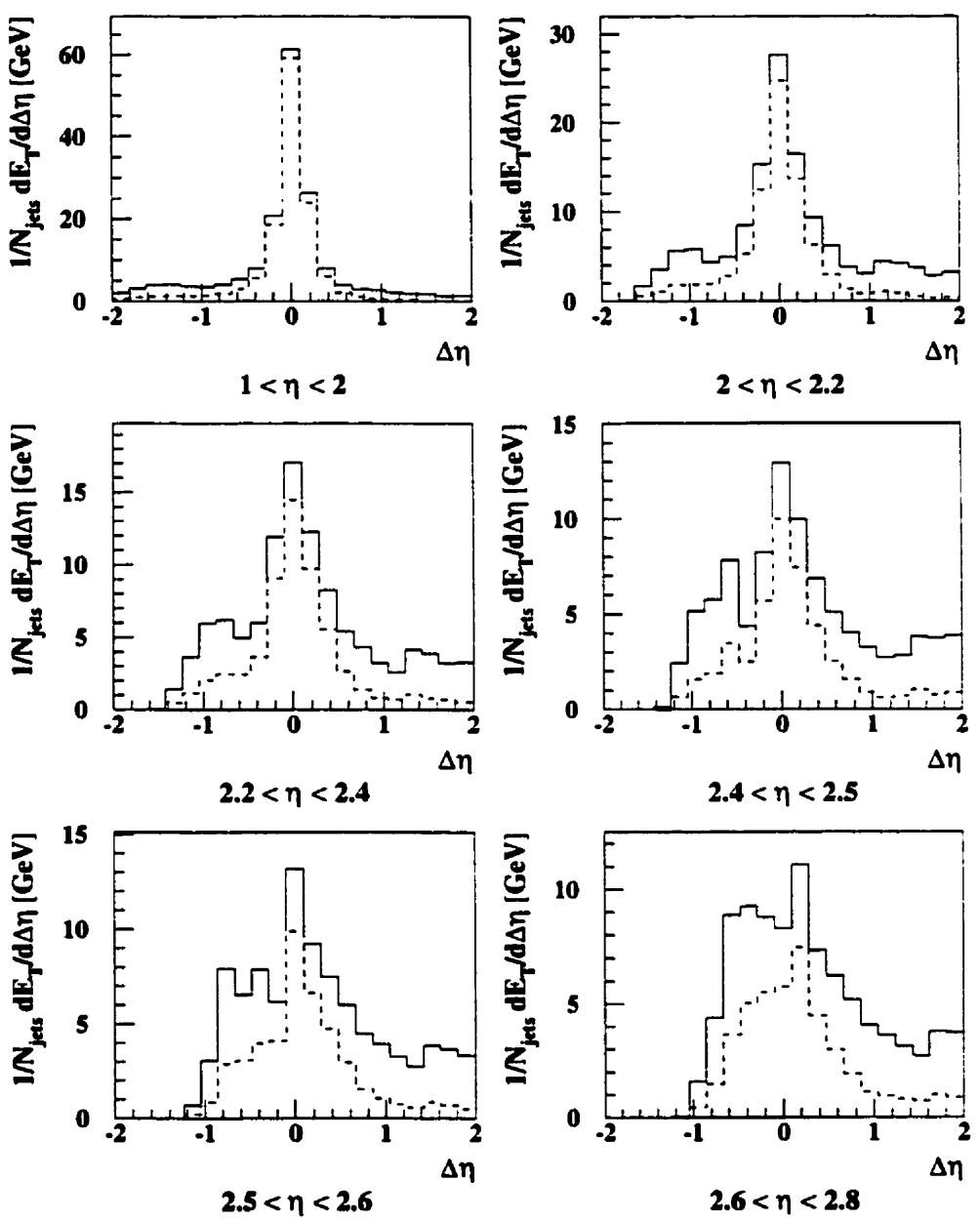

Figure 6.11: *Transverse energy flow in the calorimeter around the jet in the data. The shaded histograms show the cells belonging to the jet. The full line histograms show all the cells in the calorimeter, integrated over all $\Delta\phi$. The dashed histograms picture the cells in the calorimeter integrated over $|\Delta\phi| < 1$.*

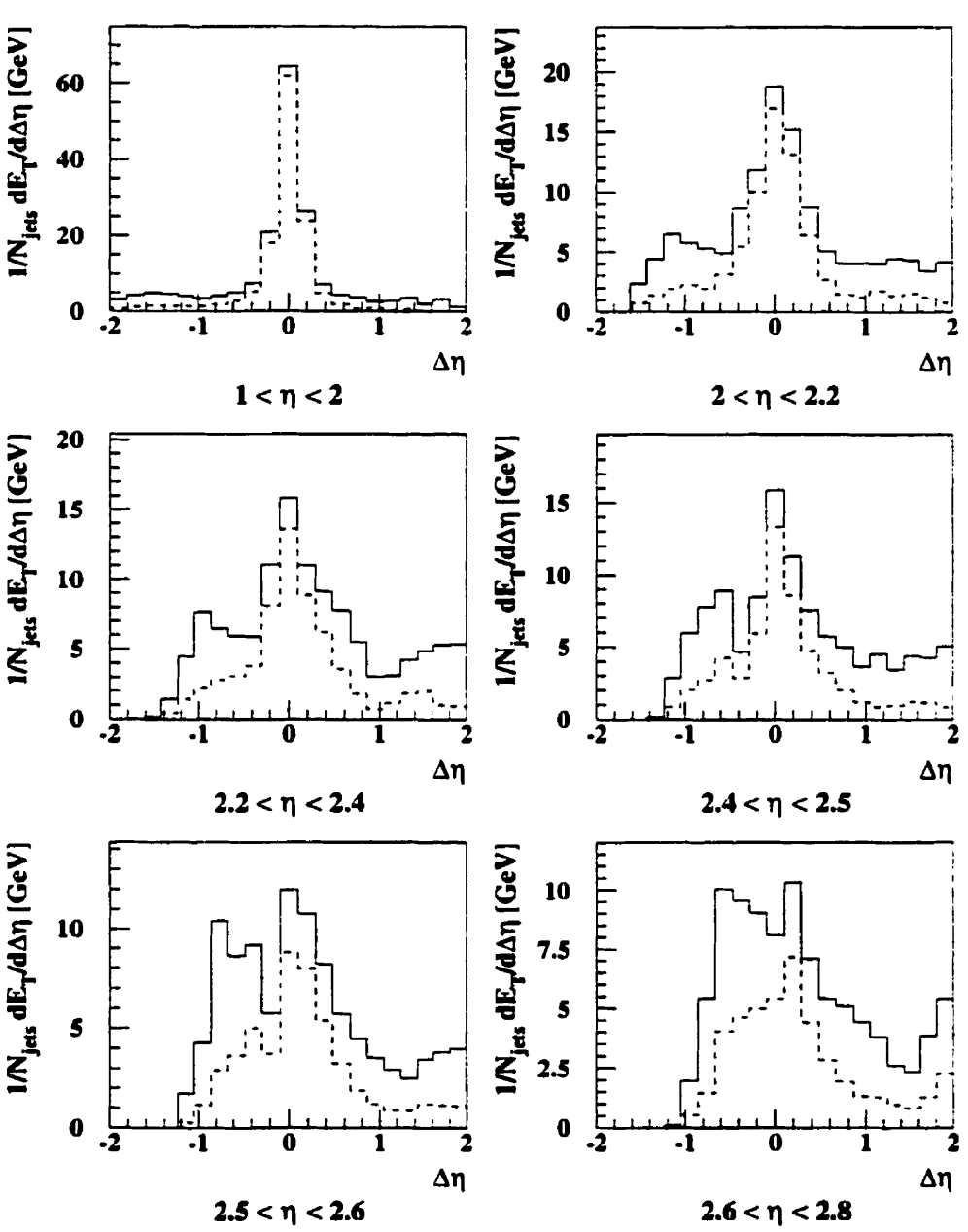

Figure 6.12: *Transverse energy flow in the calorimeter around the jet in the Monte Carlo simulation (detector level). The shaded histograms show the cells belonging to the jet. The full line histograms show all the cells in the calorimeter, integrated over all* $\Delta\phi$. *The dashed histograms picture the cells in the calorimeter integrated over* $|\Delta\phi| < 1$.

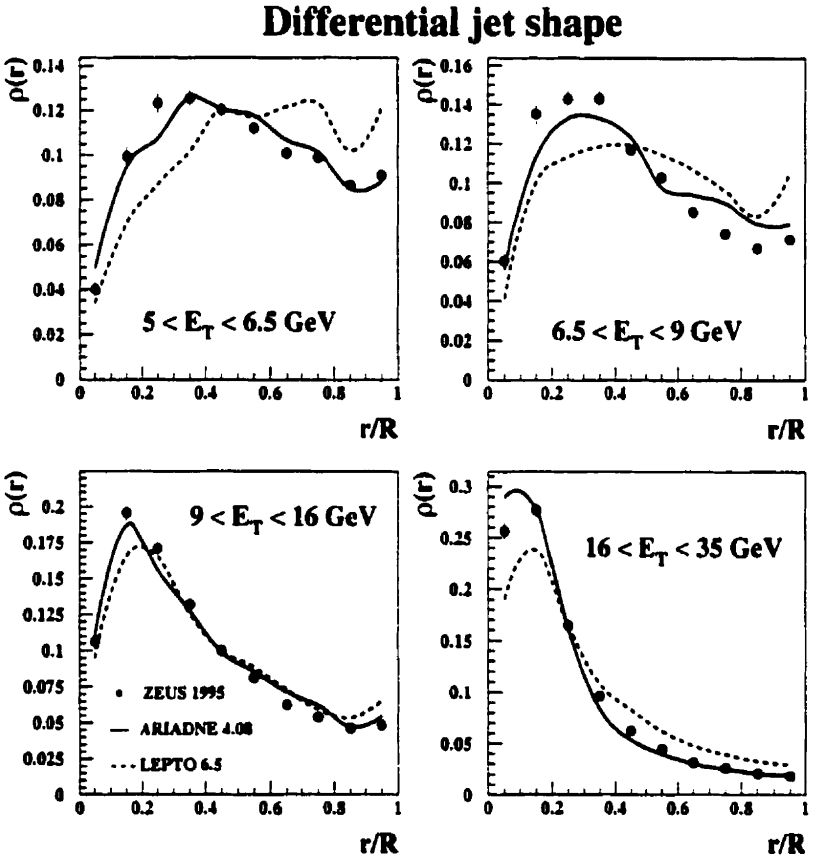

Figure 6.13: *Differential jet shape in various E_T bins. The data are shown as dots, ARIADNE, as a full line and LEPTO as a dashed line. Statistical errors only.*

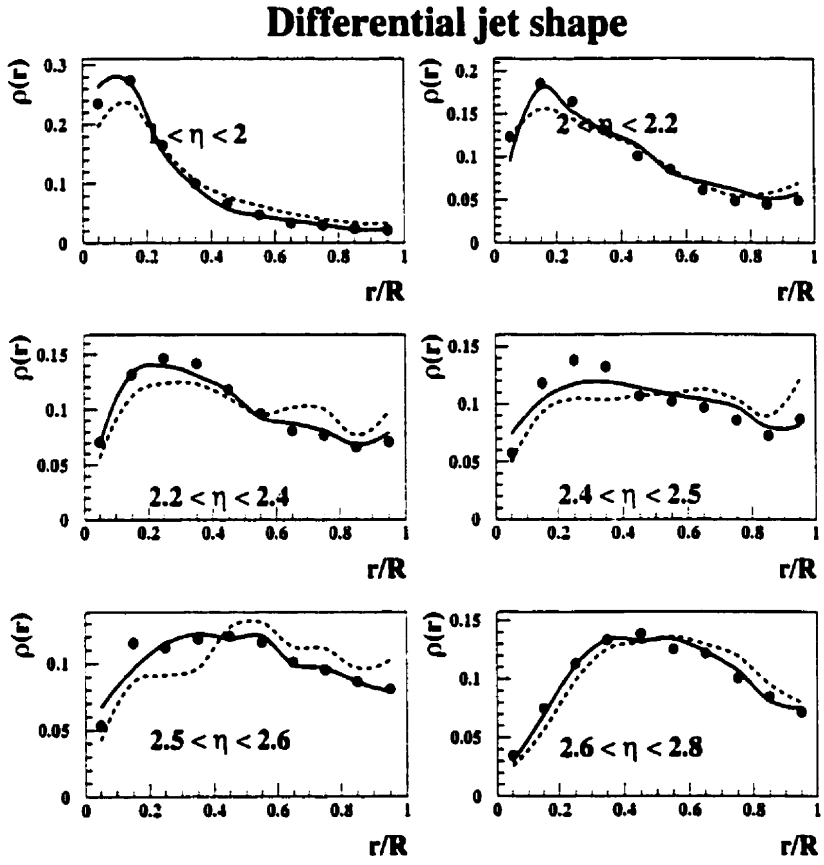

Figure 6.14: *Differential jet shape in various η bins. The data are shown as dots, ARIADNE, as a full line and LEPTO as a dashed line. Statistical errors only.*

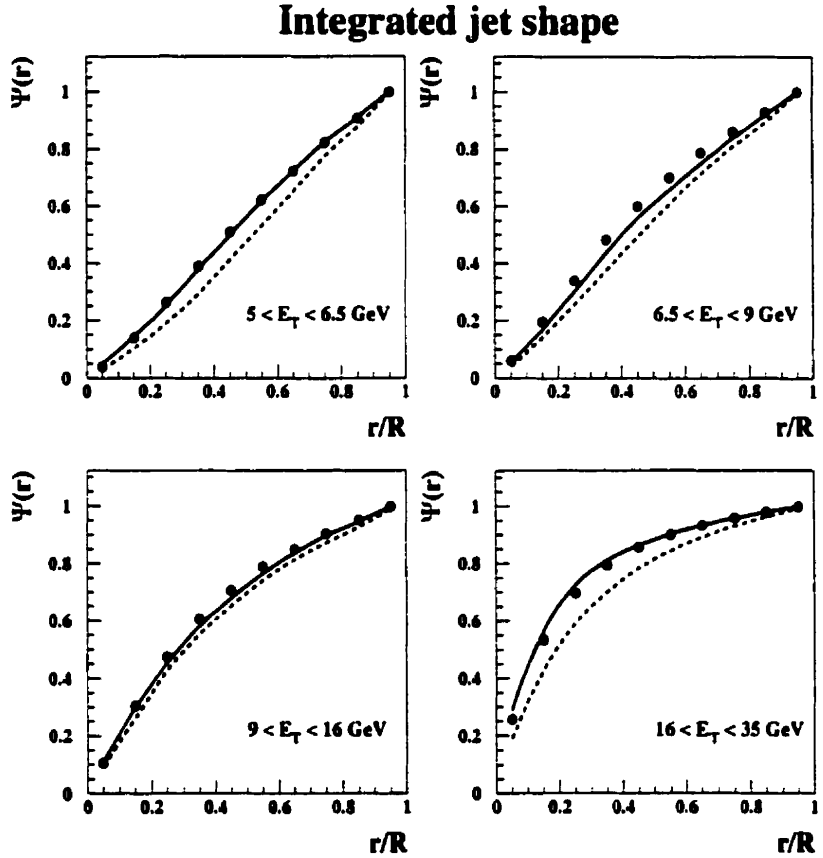

Figure 6.15: *Integrated jet shape in various E_T bins. The data are shown as dots, ARIADNE, as a full line and LEPTO as a dashed line. Statistical errors only.*

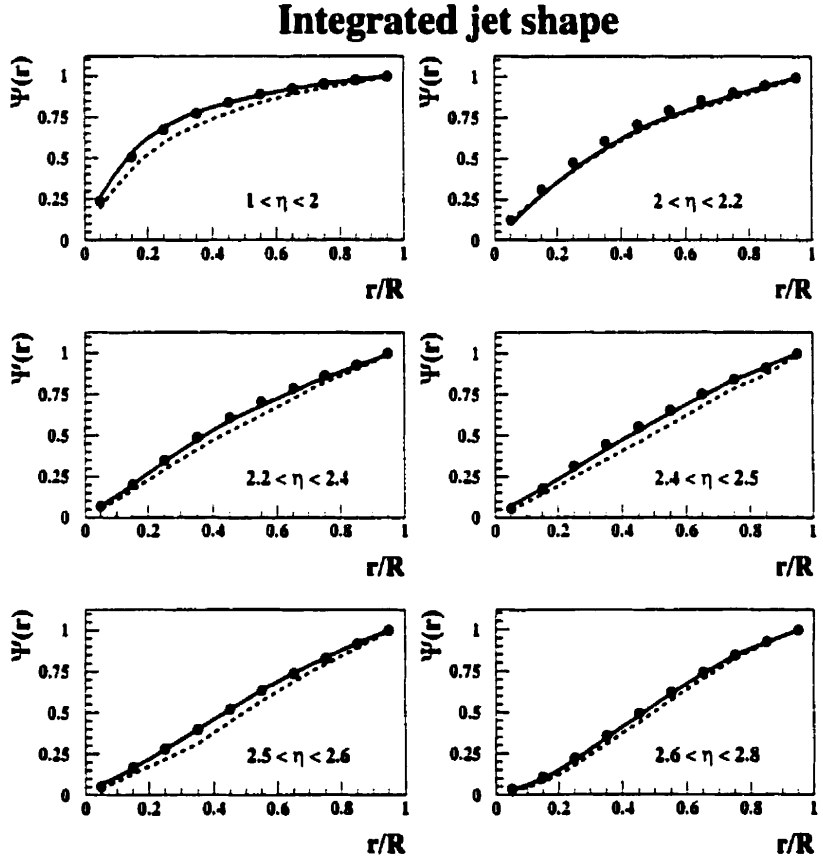

Figure 6.16: *Integrated jet shape in various η bins. The data are shown as dots, ARIADNE, as a full line and LEPTO as a dashed line. Statistical errors only.*

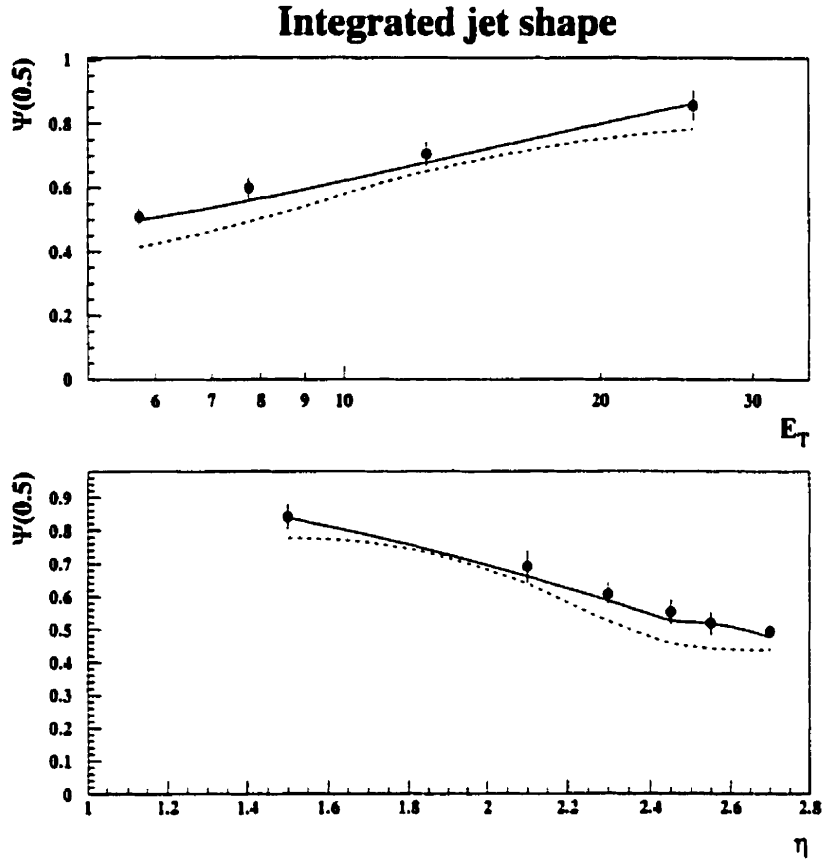

Figure 6.17: *Integrated jet shape for $r = 0.5R$. The data are shown as dots, ARIADNE, as a full line and LEPTO as a dashed line. Statistical errors only.*

Chapter 7

Angular Correlation between Jets

7.1 Introduction

In the previous chapters, we explained why a better understanding of the low-x physics is a crucial step towards a global understanding of pQCD and how the jet observables could help us reaching this goal. In this chapter, the angular correlation between two jets in the hadronic center of mass frame will be studied. The correlation between the azimuthal angles of the two leading order jets, in this frame, is supposed to give a hint on the underlying parton processes [109]. The main purpose is to test the validity range of the DGLAP evolution picture. According to these evolution equations, the overall transverse momentum carried away by the higher order partons (beyond leading order in α_s) is small. Therefore the two leading order jets must be approximately back-to-back in the transverse plane over the entire kinematic range. In another type of evolution scheme, the BFKL evolution picture, the transverse momentum of the high-order partons could be comparable to the one of the leading order jets. Therefore, the correlation between the leading-order jets should become weaker when the multigluon emissions become large. This is the case when the scaling variable x or the momentum fraction carried by the gluon x_g (described in the section 6.2.1) is small. To study these processes, some cuts were applied which were described in the section 6.2.1 and are summarized in the table 7.1. The aim of this analysis is to study the azimuthal angle $\Delta\phi$ between the two leading order jets in bins of x and x_g and see if, going to small x and x_g values, one can really observe a significant weakening of the azimuthal correlation.

The strategy used in this analysis is standard in high-energy physics and will be the

same for the next study on forward jet: the data will be first compared to the Monte Carlo predictions at detector level. The Monte Carlo model which describes the data best will be used for the detector corrections. Once corrected back to hadron level, the data will be compared to the various simulations and some first conclusions on the validity of the models will be drawn. Finally, the data will be corrected for hadronization effects (to "parton level") in order to be compared to the NLO calculations.

$E_{e'} > 10$ GeV
$y > 0.04$
$Q^2 > 8$ GeV^2
$E_{T,Jet1}^{Lab} > 4$ GeV, $E_{T,Jet2}^{Lab} > 6$ GeV
$E_{T,Jet1}^{HCM} > 4$ GeV, $E_{T,Jet2}^{HCM} > 6$ GeV
$-2 < \eta_{Jet}^{Lab} < 2.2$
$\eta_{Jet}^{HCM} > 0$
$\Delta\eta^{HCM} > 1.2$
$10^{-4} < x < 10^{-2}$
$10^{-3} < x_g < 10^{-1}$

Table 7.1: *Selected phase space region for the cross section measurement.*

7.2 Comparison between the Data and the Monte Carlo Simulations

The data which are obtained in ZEUS need to be corrected for detector effects: all the subcomponents of the detector have a finite acceptance, a certain resolution and the background arising from the beam gas can contribute to spoil the accuracy of the measurement. The usual method to correct the data is to use the Monte Carlo model, which includes a full simulation of the detector through the GEANT package (see chapter 3). The geometrical acceptance and the resolution of each one of the subcomponents of the detector are implemented in the simulation, as well as the

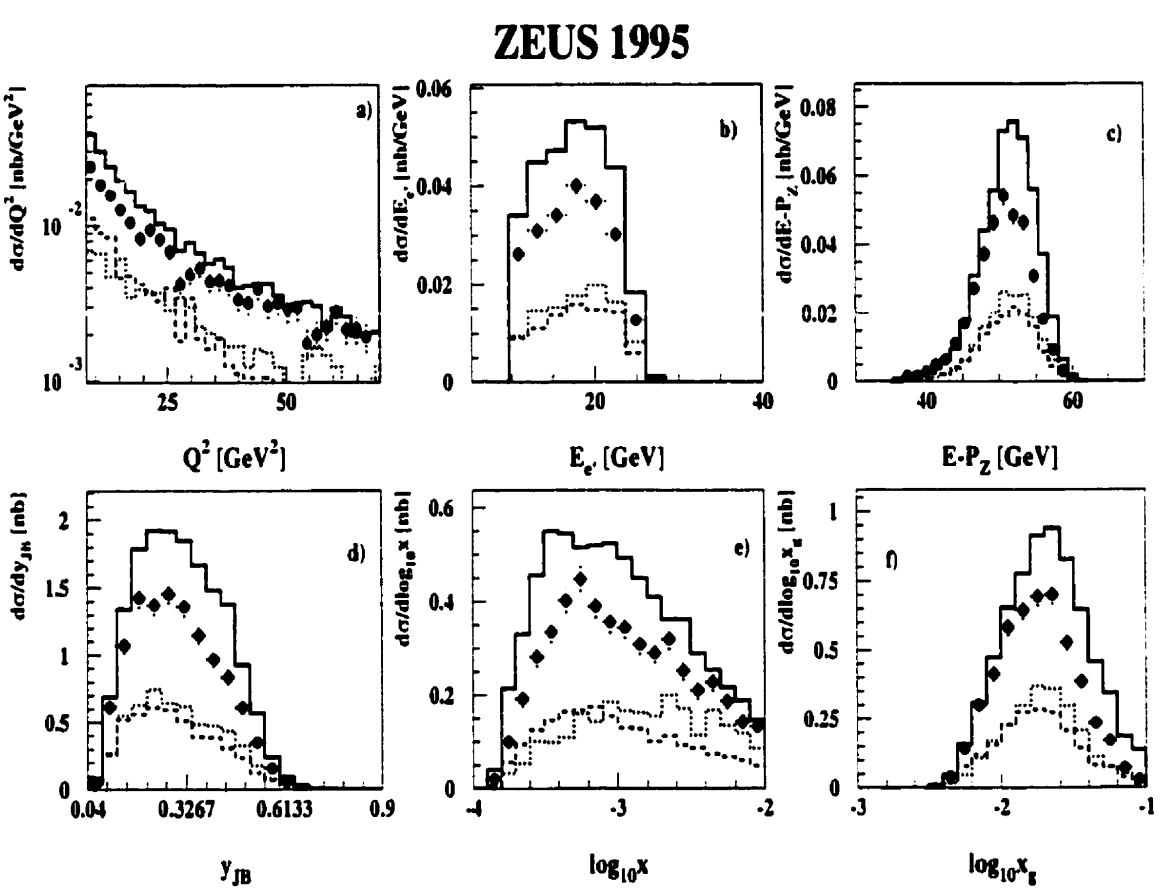

Figure 7.1: *Comparison between the data (shown as full dots) and Monte Carlo simulations for six variables: a) Q^2, b) $E_{e'}$, c) y_{JB}, d) $E - P_Z$, e) $log_{10}x$, f) $log_{10}x_g$. The data are displayed as full dots. ARIADNE is shown as a full line, LEPTO as a dashed line and HERWIG as a dotted line. Errors shown are statistical only.*

beam gas background. The comparison between the "true" level, where no detector is present, and the detector level (after the simulation of all the subcomponents) provides the tool for the corrections. The technics and results of these corrections will be discussed in the next section. But before correcting the data with the Monte Carlo simulation, one needs to make sure that the model describes fairly accurately all the distributions observed in the data. If this is not the case, the results of the correction might be inaccurate or completely wrong.

It is extremely important that the Monte Carlo model describes the data in the particular phase space under investigation. We saw in section 5.3 that all the models

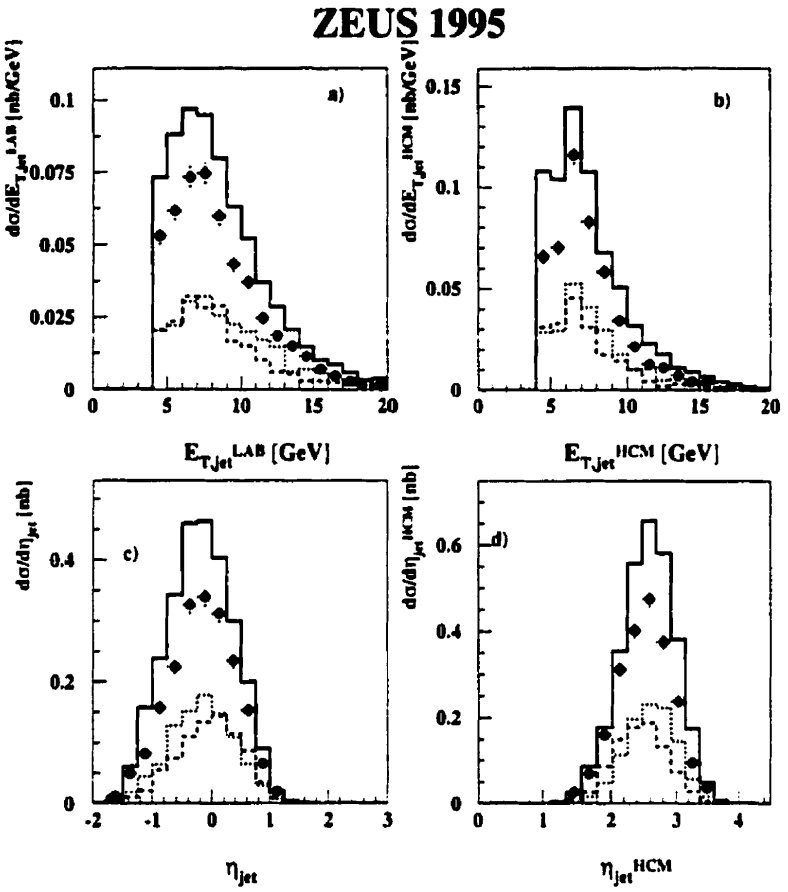

Figure 7.2: *Comparison between the data (shown as full dots) and Monte Carlo simulations for four jets variables: a) $E_{T,jet}^{lab}$, b) $E_{T,jet}^{HCM}$, c) η^{lab}, d) η^{HCM}. The data are displayed as full dots. ARIADNE is shown as a full line, LEPTO as a dashed line and HERWIG as a dotted line. Errors shown are statistical only.*

provide a good description of the overall DIS kinematic variables. This is not enough to safely use any of these models to correct the data. For example, a model like LEPTO which does not have the super-symmetry implemented could not be used in a leptoquark search, but could still describe the general kinematics of the event in this type of analysis. In figure 7.1, the differential cross section of the events at detector level, within the phase space defined by the cuts in table 7.1, is compared to six general kinematic variables: Q^2, the energy of the scattered lepton $E_{e'}$, y_{JB}, $E - P_Z$, x and x_g. The data points are compared to three Monte Carlo models: ARIADNE 4.08, LEPTO 6.5 and HERWIG 5.9. The striking feature of these plots is that none of the Monte Carlo describes the absolute normalization of the data. ARIADNE pre-

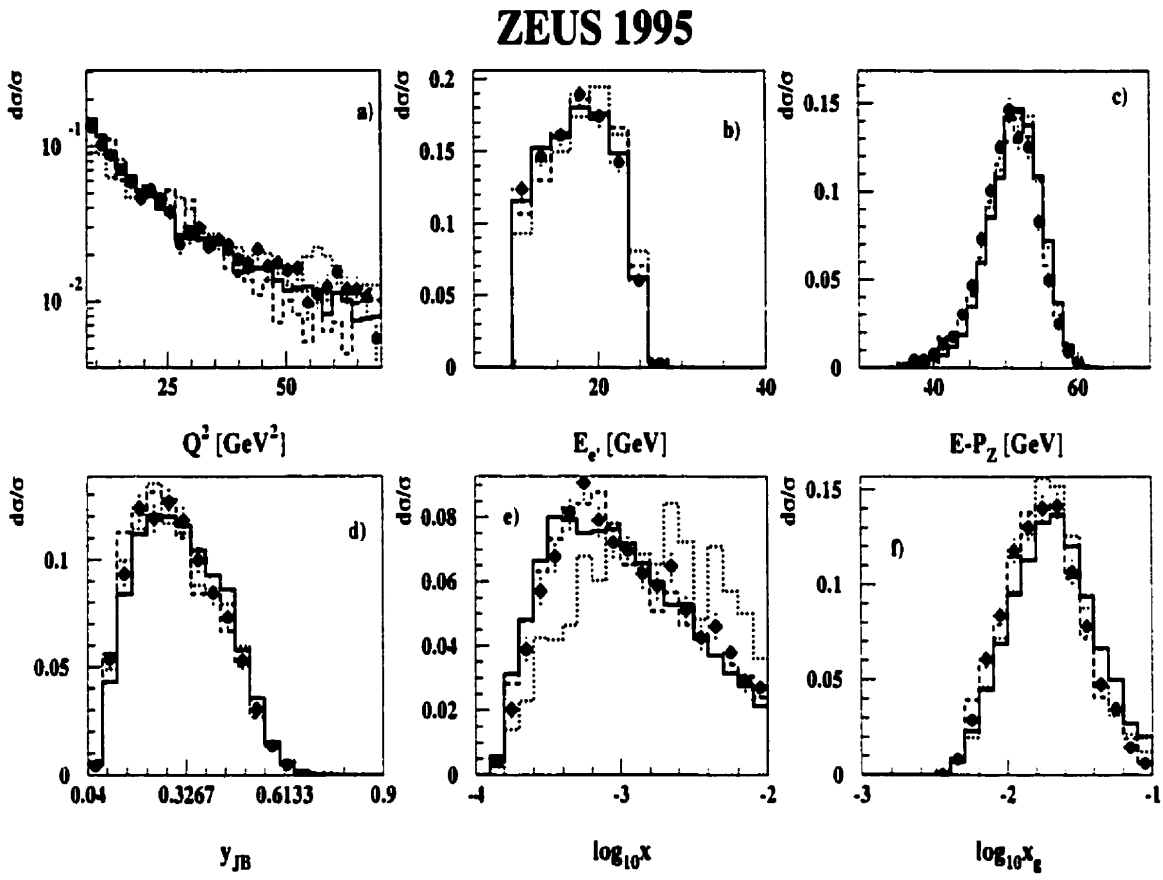

Figure 7.3: *Shape comparison between the data (shown as full dots) and Monte Carlo simulations for six variables: a) Q^2, b) $E_{e'}$, c) y_{JB}, d) $E - P_Z$, e) $log_{10}x$, f) $log_{10}x_g$. The data are displayed as full dots. ARIADNE is shown as a full line, LEPTO as a dashed line and HERWIG as a dotted line. Errors shown are statistical only.*

dicts a cross section which is 25% too large, whereas LEPTO and HERWIG predict a much too small cross sections (by about 40%). In figure 7.2, the comparison is made between some jet variables in the data and in the Monte Carlo models: the transverse energy of the jets in the laboratory frame, the transverse energy in the HCM frame, the pseudorapidity in the laboratory frame and the pseudorapidity in the HCM frame (all these variables are used in the selection of the final sample). There again, none of the models can reproduce the absolute normalization of the data in this phase space region. The reason for this is still unclear. All the Monte Carlo models have been tuned to describe the hadronic final states in several distributions, like the hadronic energy flow versus the pseudo-rapidity, but fail to describe these ones. This is also

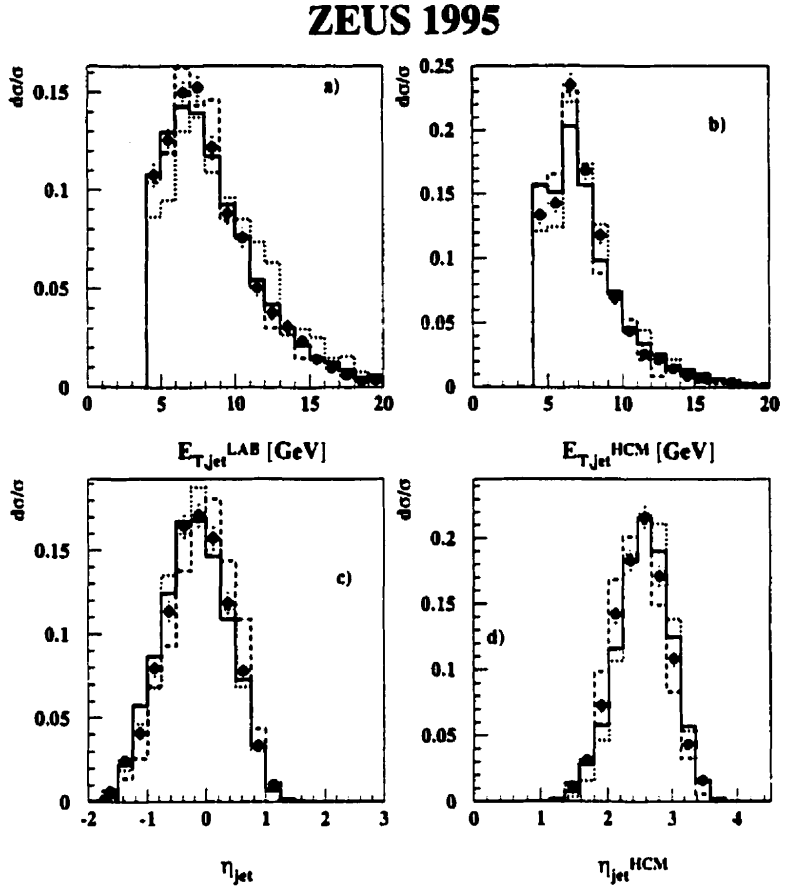

Figure 7.4: *Shape comparison between the data (shown as full dots) and Monte Carlo simulations for four jets variables: a) $E_{T,jet}^{lab}$, b) $E_{T,jet}^{HCM}$, c) η^{lab}, d) η^{HCM}. The data are displayed as full dots. ARIADNE is shown as a full line, LEPTO as a dashed line and HERWIG as a dotted line. Errors shown are statistical only.*

the case in other di-jet analyses (like the determination of the gluon density [114]).

This normalization between data and Monte Carlo is however not critical. The correction factors are ratios between the true and the detector levels and the absolute normalization cancels out to the first order. It is however crucial that the shape of each one of the distributions agrees with the data, otherwise, migrations of events from bin to bin are not properly taken into account and the corrections might be biased. The shape comparison between data and Monte Carlo simulations is shown in figures 7.3 (for the general kinematic variables quoted above) and 7.4 (for the jet variables). Here ARIADNE and LEPTO give an accurate description of the shape

of most of the variables considered (except for the x_g distribution in ARIADNE and both pseudorapidity distributions in LEPTO which seem shifted). These models can therefore be used for correcting the variables for detector effects. HERWIG, however, fails completely in describing the shape of the x distribution, so it can not be used to determine the detector corrections.

7.3 Detector Correction

Once the Monte Carlo models which describe the experimental data, have been determined, the correction procedure is fairly straightforward. The first step is the jet energy correction, in order to minimize the differences between the model predictions and the experimental data, using the Monte Carlo simulations. Then, the same simulations are used to evaluate the acceptance correction, purity and efficiency of the sample, by a simple comparison between what has been generated (the true physics process) and what has been observed after the detector simulation. These numbers will be used to correct the data for both acceptance and migration.

7.3.1 Jet Energy Correction

The energy correction is a quite usual step in the study of hadronic final states, and in particular in jet analyses. This is not strictly speaking mandatory for the final results. As a matter of fact, the full acceptance correction should also take into account the effects of energy loss in the calorimeter due to dead material (see chapter 4). For instance, let's consider a jet generated with a transverse energy E_T, larger than some minimum value $E_T^{cut-off}$ to ensure that the jet comes from hard processes and not merely from energy fluctuation inside the calorimeter. Let's assume that due to dead material, the jet loses part of its energy and is observed with an energy $E'_T < E_T^{cut-off}$. The jet will not participate in the observed jet distribution but will be included in the generated jet distribution. After looping over all the jets (and assuming that the energy loss is the only cause of discrepancy between true and reconstructed level), the number of jets ending up in the observed distribution will be smaller than the number of jets ending up in the true distribution. Therefore, the amount of jets which are lost due to energy loss in the detector is part of the overall efficiency of the detector reconstruction.

If the model would describe the data perfectly, there would not be a need for such a correction. Unfortunately no model can pretend to reproduce exactly the energy loss within the detector so that discrepancies between simulation and reality are always present. Besides, there is at ZEUS an uncertainty on the total energy scale when comparing data and Monte Carlo simulation [115] (the data are off by 5% in the BCAL and 3% in the RCAL, compared to the GEANT simulations. The energy scale discrepancy in the FCAL is known less accurately but is believed to be on the order of 3% as in the RCAL). This systematic shift in the energy distribution could therefore bias the correction.

Therefore such a correction has to be applied for any energy deposition in the calorimeter and in particular for the jets. In figure 7.5, the relative difference between detector level and generator level jets if displayed (after all the cuts listed in table 7.1. The energy of the generated jets is 13%-16% larger than the energy of the reconstructed jets. This discrepancy is due to the energy loss of the jets in dead material in front of the calorimeter (see section 4.2.1). A standard and unified method for correcting this effect has yet to be determined in ZEUS. A number of ZEUS notes exist on the subject(see [116], [117], [118], [119] and [120] for a non-exhaustive list), but it is not the object of the present research to perform a comparative study of the various methods which have been tested so far. The method presented here (the same will be used in the next analysis "forward jets in DIS") has the advantage of providing corrections which are little affected by the statistical fluctuations.

These corrections are known to be dependent on the geometry of the detector, so they are determined in various regions of pseudorapidity. Five regions have been selected, with approximately the same amount of statistics and dead material: $-2 < \eta < -1.2$, $-1.2 < \eta < -0.5$, $-0.5 < \eta < 1.0$, $1.0 < \eta < 1.5$, $1.5 < \eta < 2$ and $2 < \eta < 2.2$.

In each one of these regions, the correction factors $C(E_{(T)}^{Det})$ are averaged over the polar and azimuthal angle of the jets, which is reasonable as the physics is symmetric in ϕ. The corrections are computed independently for E and E_T in different bins of E and E_T. The (transverse) energy of the detector level jets was then corrected using the general formula: $E_{(T)}^{CORR} = C(E_{(T)}^{Det}) \times E_{(T)}^{Det}$, where $C(E_{(T)}^{Det}) = \frac{E_{(T)}^{TRUE}}{E_{(T)}^{Det}}$. At the detector level, the sample used for the correction is obtained by applying all the

selection cuts to the events except the energy and transverse energy cuts which are to be a little bit looser. The latter is important in order not to bias the correction by effects which are due to jets which will not end up in the final sample. The hadron jet is then selected within some loose cuts and matched to the acceptable detector jet by applying a distance criteria in the (η,ϕ) space: the hadron jet which has the smaller distance $R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ with respect to the detector jet is chosen to compute the correction factor. An additional cut is performed to reject events where the closest hadron jet is still too far from the detector jet: $R < 1$. In order not to bias the sample, a gaussian distribution is fitted to the ratio $\frac{E^{TRUE}_{(T)}}{E^{Det}_{(T)}}$ in each $E^{Det}_{(T)}$) and η^{Det} bin, so that the hadron jets which are wrongly matched to the detector jets are sitting on the tail of the distributions and thus do not contribute to the overall correction factor. In fig. 7.6, the values of these correction factors are shown as a function of the transverse energy in several bins of pseudorapidity (as the shape of the correction was flat as a function of the energy, the raw data were simply multiplied by a single number in the various η bins to account for the total energy loss). The correction is applied by multiplying bin by bin the energy (or transverse energy) of the jet observed in the detector by the result of the fit.

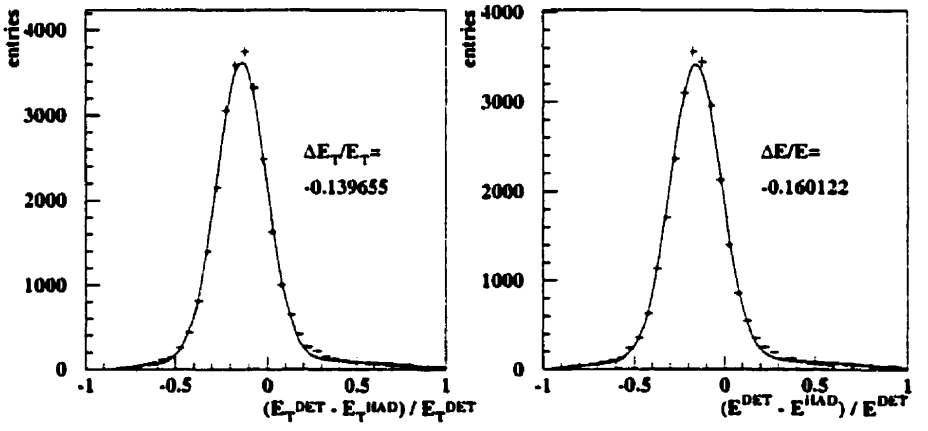

Figure 7.5: *Relative difference between the true and the observed (transverse) energy in the ARIADNE 4.08 model. The systematic shift between true and observed energy distribution is due to energy loss in dead material.*

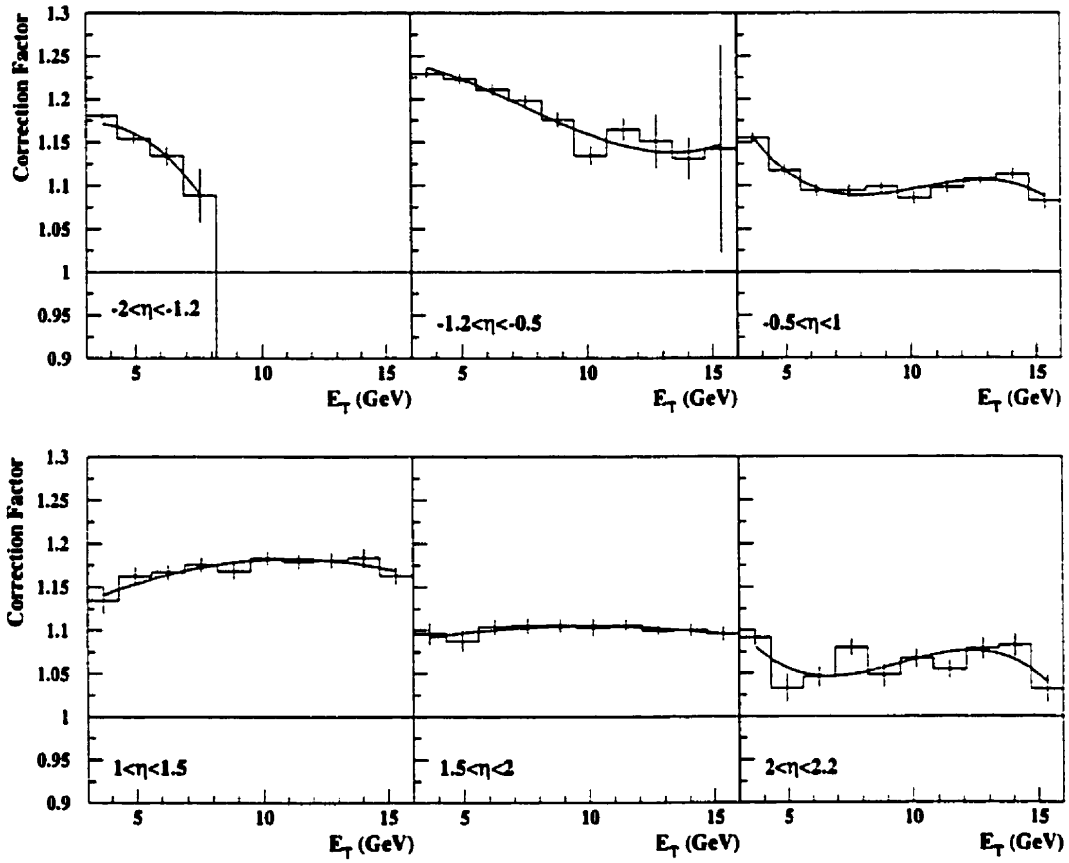

Figure 7.6: *Transverse energy correction in various η bins. The result of the fit is used to correct the detector level distributions (in the data and in the Monte Carlo model).*

7.3.2 Resolution of the Kinematic Variables

Before looking at the final value of the acceptance correction which will be applied to the data, it is important to evaluate the various contributions to this acceptance. The main one is the resolution of the various kinematic variables which are used in the cuts. These resolutions limit the purity and efficiency of the sample and can lead to large acceptance corrections.

The estimation of these resolutions is also important in order to determine the bin width of the final distributions. This bin width should be chosen large enough as to minimize migrations from bin to bin (so that the final acceptance looks flat). In figure 7.7, we show the relative resolutions for the two variables x and x_g. The relative resolution for an observable X is defined as:

$$\frac{\delta X}{X} = \frac{X^{Det} - X^{Had}}{X^{Det}}. \tag{7.1}$$

The x resolution is discontinuous in the two largest x bins. This is a feature of the electron reconstruction method: the resolution on the the reconstruction of the value of x depends on the value of y [121]. At small y, the x resolution diverges with the electron method. At large x, smaller values of y are sampled and therefore the x resolution worsens a lot. The average resolution is around 16%. The x_g resolution is much worse (around 23 %), because x_g is a combination of event and jet variables which all have a large resolution. In figures 7.8 and 7.9, the resolution of the azimuthal angle $\Delta\phi^{HCM}$, between the two highest η^{HCM} jets, is shown. The resolution is defined by:

$$\delta\Delta\phi = \Delta\phi^{Det}_{HCM} - \Delta\phi^{Had}_{HCM}, \tag{7.2}$$

In figure 7.8, all the cuts have been applied except the one on $\Delta\eta^{HCM}$ (the difference between the pseudorapidity of the two jets in the HCM). The striking feature of this plot is that a second peak appears to the left of the main one, in the low $\delta\phi^{HCM}$ bins. This means that in most of the events, the two selected detector jets are not identified as the two hadron level jets. In these bins, the two detector jets are very close to each other, separated only in pseudorapidity in the HCM. When two jets are close in pseudorapidity and in azimuth, their distance in the (η,ϕ) plane is small and, providing that this distance is also small in the laboratory frame (which is the case for the cone algorithm as it is longitudinally invariant), it might happen that they merge to give one jet once reconstructed at detector level. In figure 7.8, the second

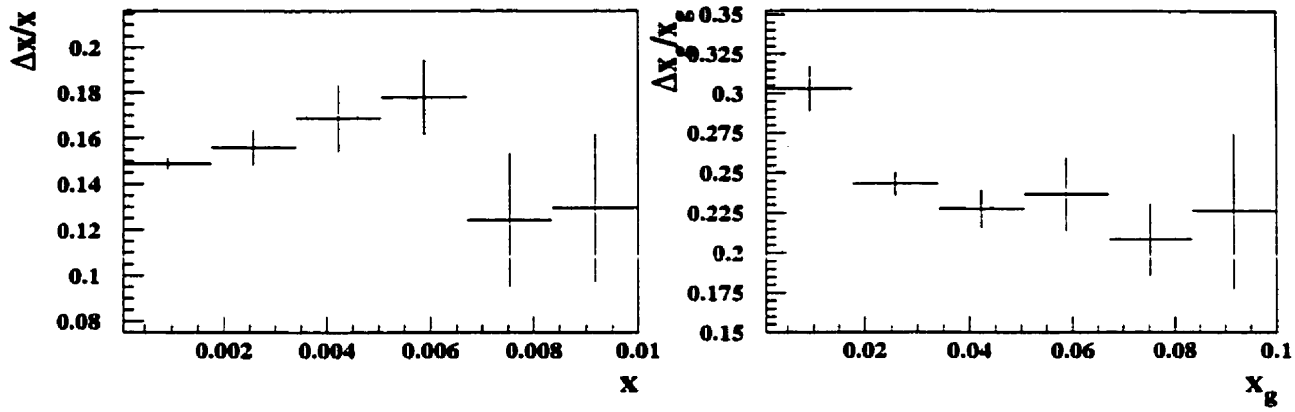

Figure 7.7: *Relative resolution for two variables x and x_g. All the cuts listed in table 7.1 have been applied. These values were obtained by fitting a gaussian to the distribution in each bin. The error quoted is the error on the fit.*

peak appears for negative values of $\delta\Delta\phi$, so the azimuthal angle between the two jets at detector level is much smaller than the one for the two jets at hadron level. What actually happens is pictured in figure 7.10. Most of the leading order jets are emitted back to back. After reconstruction, it might happen that a generated jet is reconstructed as two smaller jets. In the limit where the azimuth between the two reconstructed jets is small, the configuration presented in figure 7.10 might dominate and give rise to a large amount of di-jets at detector level unrelated to the two hadron level jets. The only solution to this problem is to make sure that the two jets are far enough in pseudorapidity, so that these "broken" jets do not contribute. This is achieved by applying a cut on the difference on the pseudorapidity between the two jets at detector and hadron level as in figure 7.9. In this plot, the second peak is much reduced (although not completely suppressed). Most of the jets found at detector level are there effectively related to the hadron level jets.

Finally, it must be added that cutting on a minimum $\Delta\eta$ in the HCM might not be a correct thing to do. As a matter of fact, the wider the gap in pseudorapidity between the jets becomes, the larger is the probability for a BFKL-type evolution between the two jets. The two jets would therefore not be any more at the top of the gluon ladder and the difference of azimuthal angle would not reflect the parton evolution. This is however the only viable solution from an experimental point of view with this type

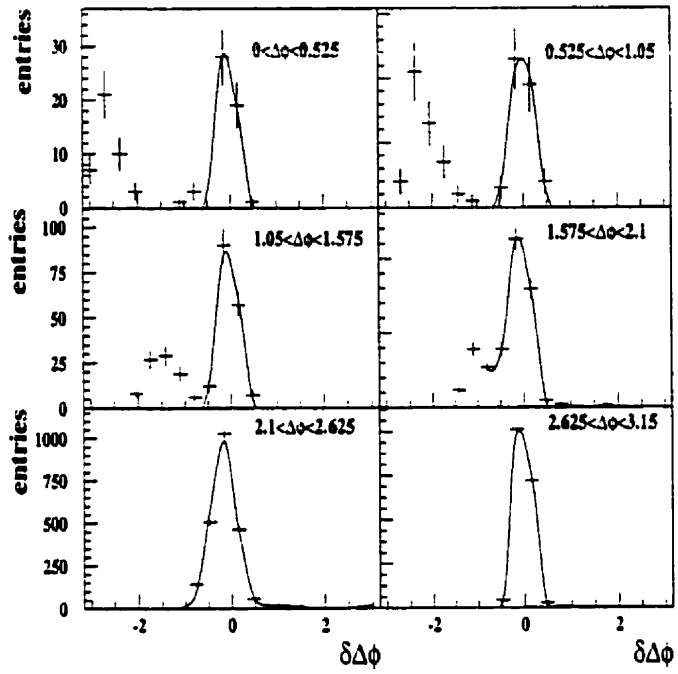

Figure 7.8: *Resolution for the variable $\Delta\phi$, the azimuthal angle between the two leading order jets before the cut in $\Delta\eta$. $\delta\Delta\phi = (\Delta\phi)^{Det} - (\Delta\phi)^{Had}$. All the cuts listed in table 7.1 have been applied except the cut on $\Delta\eta^{HCM}$.*

of algorithm. With another choice of radius (for instance R=0.7, as in D0 [38]), the situation is similar, although the number of uncorrelated jets is a little bit smaller. Nevertheless, in any case, the cut on $\Delta\eta^{HCM}$ is mandatory. We must add that the analysis was also performed with the k_T algorithm but the resolutions of the jet variables were even worse.

All these considerations, together with the requirement that there should be a reasonable amount of statistics in each bin, lead to the selected bin size displayed in table 7.2.

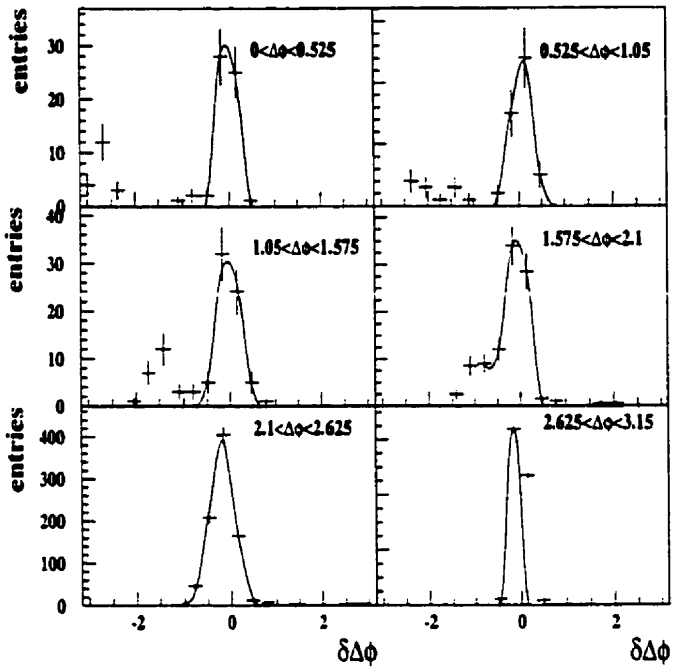

Figure 7.9: *Resolution for the variable $\Delta\phi$, the azimuthal angle between the two leading order jets after the cut in $\Delta\eta$. $\delta\Delta\phi = (\Delta\phi)^{Det} - (\Delta\phi)^{Had}$. All the cuts listed in table 7.1 have been applied.*

As we are mainly interested in studying the properties of the di-jet system in each one of the x and x_g bins, it is important to check that these properties do not vary from bin to bin due to the detector resolution. In tables 7.3 and 7.4, the relative or absolute resolutions for various jet kinematic variables is shown in the three bins of x and x_g (this is the result of a gaussian fit applied in each one of the distributions). As expected from the detector specifications, the transverse energy resolutions are around 15% while the η and ϕ resolutions vary from 5 to 10% over the whole detector range. It is important to stress that the variation of the resolution of a given variable

x range	0.0001-0.0006	0.0006-0.0020	0.0020-0.0100
x_g range	0.0010-0.0075	0.0075-0.0200	0.0200-0.1000

$\Delta\phi$ range	0.0-0.78	0.78-1.18	1.18-1.57	1.57-1.96	1.96-2.35	2.35-2.75	2.75-3.14

Table 7.2: *Final bin size for x, x_g and $\Delta\phi$ after resolution considerations.*

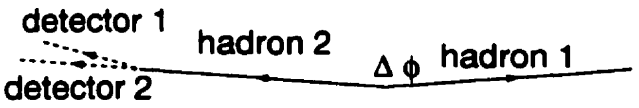

Figure 7.10: *Example of di-jet configuration in which the two hadron level jets 1 and 2 do not match the detector level jets 1 and 2. The latter come from hadron jet 1 which "breaks up" at the reconstruction level.*

from one x (x_g) bin to another is quite small (maximum 2-3%), so that the detector response does not influence too much the behavior of the di-jet system from one bin to another.

x range	$0.0001 < x < 0.0006$	$0.0006 < x < 0.002$	$0.002 < x < 0.01$
$\delta E_T^{lab}/E_T^{lab}$	0.150 ± 0.003	0.150 ± 0.003	0.145 ± 0.004
$\delta E_T^{HCM}/E_T^{HCM}$	0.160 ± 0.003	0.160 ± 0.003	0.170 ± 0.004
$\delta\eta^{lab}$	0.063 ± 0.001	0.058 ± 0.001	0.051 ± 0.002
$\delta\eta^{HCM}$	0.096 ± 0.002	0.120 ± 0.003	0.120 ± 0.004
$\delta\phi^{HCM}$	0.098 ± 0.002	0.110 ± 0.003	0.140 ± 0.005

Table 7.3: *(Relative) resolution for five jet variables E_T^{lab}, E_T^{HCM}, η^{lab}, η^{HCM}, ϕ^{HCM}, as a function of x.*

7.3.3 Acceptance correction, Purity, Efficiency

Finally, an acceptance correction is performed and the events are corrected for efficiencies and purities. The final cross section is expressed in bins of x, x_g and $\Delta\phi$, therefore the correction factors are expressed in these bins.

In order to correct the data, a bin-by-bin method has been applied: the acceptance correction factor in each bin is the ratio of generated over the reconstructed number of events in this bin ($N_{di-jets}$ means number of di-jets):

x_g range	$0.001 < x_g < 0.0075$	$0.0075 < x_g < 0.02$	$0.02 < x_g < 0.1$
$\delta E_T^{lab}/E_T^{lab}$	0.150 ± 0.003	0.150 ± 0.003	0.150 ± 0.002
$\delta E_T^{HCM}/E_T^{HCM}$	0.150 ± 0.002	0.160 ± 0.003	0.160 ± 0.002
$\delta\eta^{lab}$	0.061 ± 0.005	0.065 ± 0.002	0.055 ± 0.001
$\delta\eta^{HCM}$	0.084 ± 0.009	0.110 ± 0.002	0.120 ± 0.002
$\delta\phi^{HCM}$	0.080 ± 0.009	0.120 ± 0.003	0.110 ± 0.003

Table 7.4: *(Relative) resolution for five jet variables E_T^{lab}, E_T^{HCM}, η^{lab}, η^{HCM}, ϕ^{HCM}, as a function of x_g.*

$$acceptance = \frac{N_{di-jets}\ at\ hadron\ level\ in\ true\ variables\ in\ the\ relevant\ phase\ space}{N_{di-jets}\ at\ detector\ level\ in\ reconstructed\ variables\ after\ all\ the\ cuts}. \tag{7.3}$$

In this method, the number of jets in one bin is assumed to be independent from the number of jets in the other bins when comparing hadron and detector levels. However, this is not rigorously true as a jet generated in one bin can be reconstructed in another bin because of the finite resolution of the detector. This migration effect has been taken care of by taking the size of the x, x_g and $\Delta\phi$ bins at least twice as large as the resolution of each one of these variables (see previous section).

The correction factors, purities and efficiencies are shown (in bins of x and x_g) in fig. 7.11 and 7.12 are defined on an event-by-event basis and with the ARIADNE simulation in the following way:

Purity is the ratio of events generated and detected in one bin over the total number of detected events in this bin:

$$purity = \frac{N_{di-jets}\ found\ at\ detector\ level\ AND\ at\ hadron\ level\ after\ all\ cuts}{N_{di-jets}\ found\ at\ detector\ level\ after\ all\ cuts}. \tag{7.4}$$

Efficiency is the ratio of events generated and detected in one bin over the total number of generated events in this bin:

$$efficiency = \frac{N_{di-jets}\ found\ at\ detector\ level\ AND\ at\ hadron\ level\ after\ all\ cuts}{N_{di-jets}\ found\ at\ hadron\ level\ after\ all\ cuts}. \tag{7.5}$$

Therefore, a jet which is found at detector level is called an impurity if there is no corresponding jet at hadron level, within the set of cuts of the analysis and inside the same x and ϕ bin. Similarly, if there is no jet in the detector, although there is one fulfilling all the selection criteria at the generator level, this will reduce the efficiency of the di-jet detection. Both quantities are calculated after energy and transverse energy correction. They reflect the experimental limits of the di-jet detection.

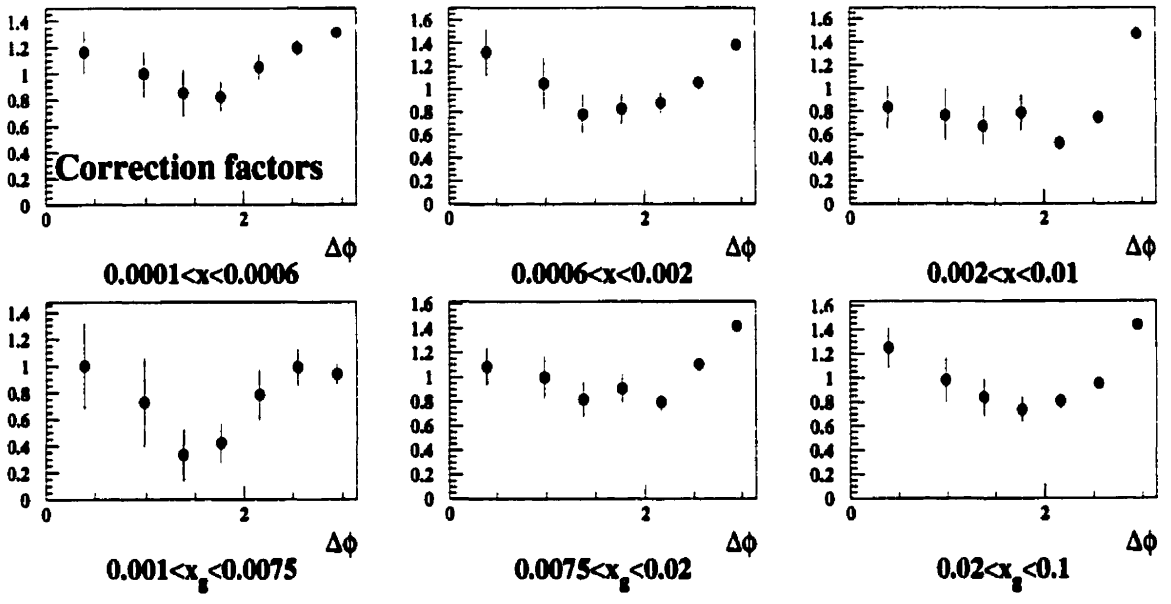

Figure 7.11: *Acceptance correction for the di-jet system as a function of x and x_g.*

In figure 7.12, we also show the effect of the migration on the correction factors, efficiencies and purities: the black dots represent the values of these quantities when no migration between the bins is considered either in $\Delta\phi$ or in x, the open dots show these same values when migration between $\Delta\phi$ bins is considered and the full squares show these values when migrations are allowed both in x and in $\Delta\phi$. Although the values of the bins are consistent with the resolution of the displayed variable, the correction due to bin-by-bin migration is quite large, especially in $\Delta\phi$. This means that the tails of the gaussian distributions are large and ultimately, the resolution of the variable is not gaussian. Figure 7.12 also shows that the effect of migration in the variable x_g is extremely large. This indicates that the x_g resolution is very coarse and once again, the resolution distribution is barely a gaussian. The first bins of the x_g distributions show no purity or efficiency. This means that no event has been generated in this x_g bin and reconstructed in the same bin. Applying a bin-by-bin

correction to these events is meaningless as the value of the corrected cross section would only reflect the way the Monte Carlo simulation is handling the bin-to-bin migration rather than the behaviour of the experimental data. Because of that, we decided not to present the x_g distributions.

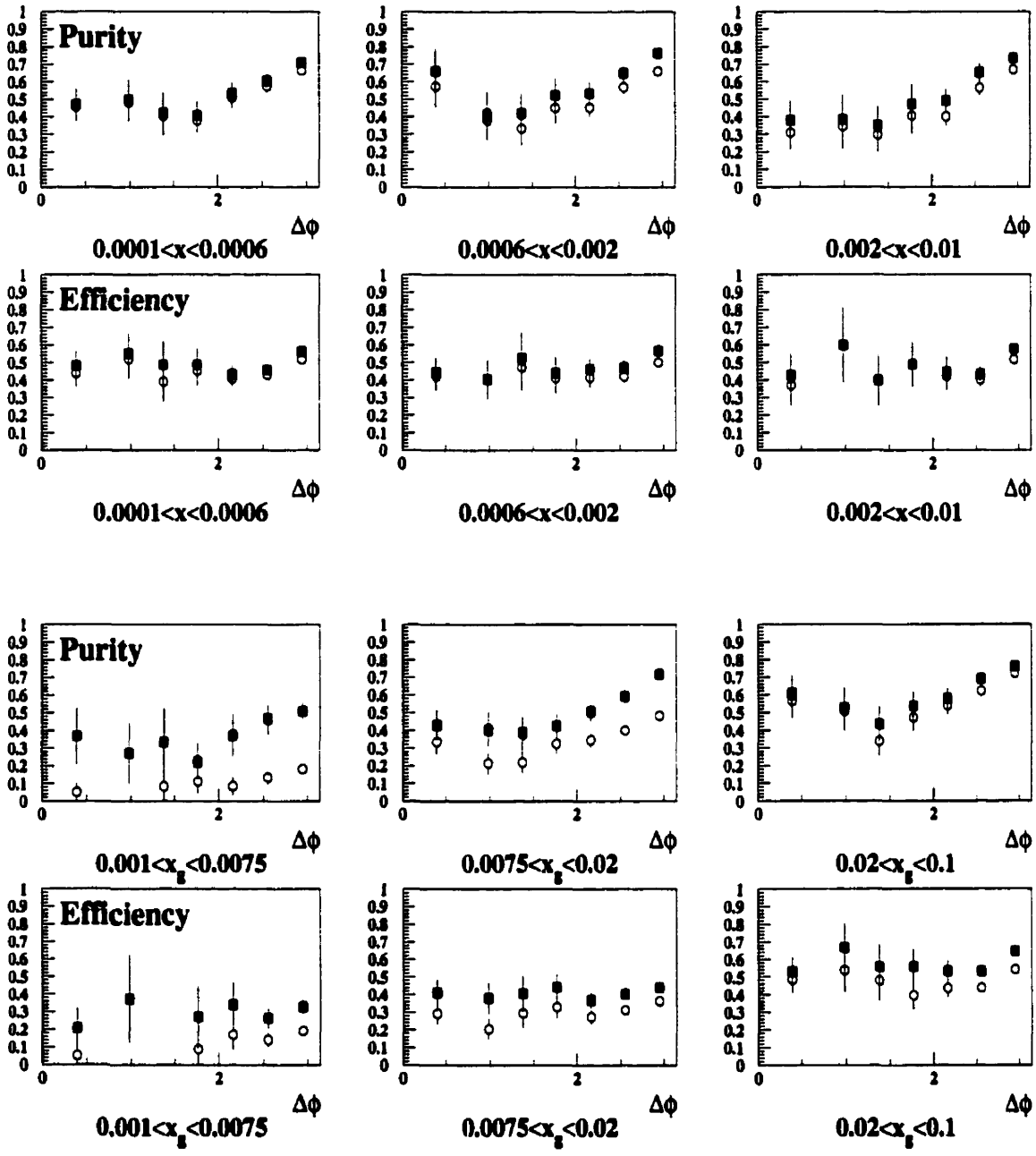

Figure 7.12: *Purity and efficiency for the di-jet system as a function of x and x_g. The various symbols correspond to the constraint on the migration: with full dots, no migration is allowed between the bins, in open dots, migration is allowed between $\Delta\phi$ bins and in full square, migration is allowed between x and $\Delta\phi$ bins.*

7.4 Azimuthal Correlation between the two Leading Order Jets

7.4.1 Detector and Hadron Level Comparison

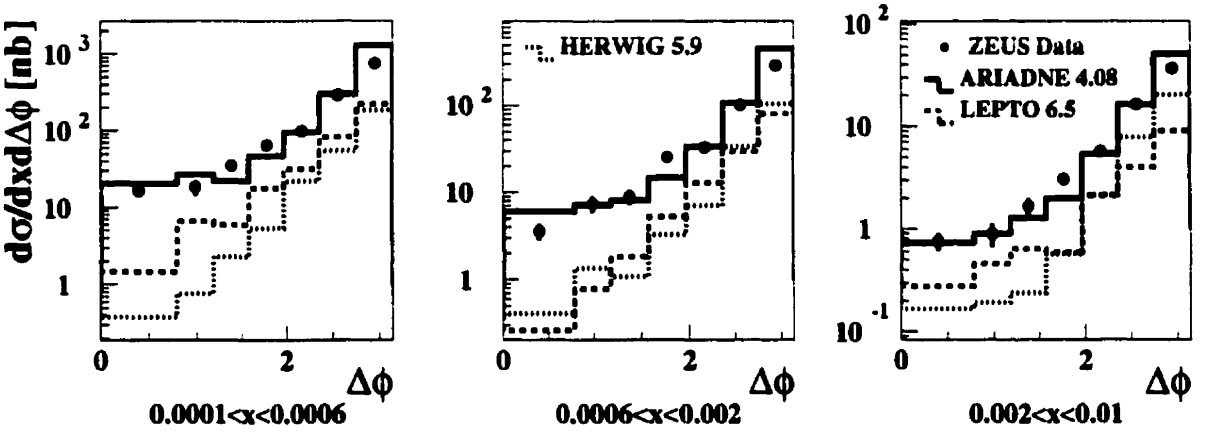

Figure 7.13: *Uncorrected di-jet cross section as a function of $\Delta\phi$ in three x bins in data and Monte Carlo. The data are shown as full dots and the Monte Carlo simulations as histograms: ARIADNE is shown as a full line, LEPTO as a dashed line and HERWIG as a dotted line. For all plots, statistical errors are shown as thick error bars, and statistical and systematic errors added in quadrature as thin error bars.*

The azimuthal correlation between the two leading order jets in the data compared to the Monte Carlo predictions at detector level is shown in figure 7.13. The binning in x and in $\Delta\phi$ was defined in the previous section. The detector level cross section is defined as:

$$\frac{d\sigma^{Det}}{d\Delta\phi dx} = \frac{N_{events}}{\mathcal{L}_{data}\Delta x\Delta(\Delta\phi)}. \tag{7.6}$$

$\mathcal{L}_{data}$ is the luminosity in the sample, Δx the bin size in the variable x and $\Delta(\Delta\phi)$, the bin size in $\Delta\phi$.

As in section 7.2, no Monte Carlo simulation can describe the data. ARIADNE leads to a cross section larger than the data's in the largest $\Delta\phi$, while the MEPS models,

LEPTO and HERWIG predict too small cross sections.

To get rid of the under-determination due to the detector effects, the hadron level cross section of the di-jet system in bins of x and $\Delta\phi$ is defined as:

$$\frac{d\sigma^{had}}{d\Delta\phi dx} = \frac{N_{events}}{\mathcal{L}_{data}\mathcal{A}_{corr}\Delta x\Delta(\Delta\phi)}. \tag{7.7}$$

$\mathcal{L}_{data}$, Δx and $\Delta\phi$ are defined above. $\mathcal{A}_{corr}$ is the acceptance correction which is shown in figure 7.11, using the ARIADNE simulation.

The corrected cross section is shown in figure 7.14 for the three x bins considered.

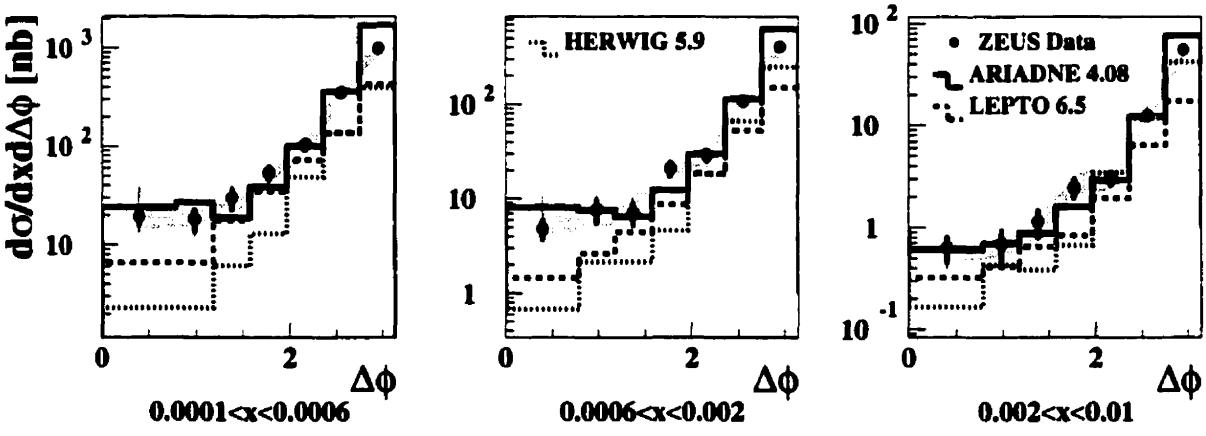

Figure 7.14: *Di-jet cross section corrected for detector effects (hadron level) as a function of $\Delta\phi$ in three x bins in data and Monte Carlo. The data are shown as full dots and the Monte Carlo simulations as histograms: ARIADNE is shown as a full line, LEPTO as a dashed line and HERWIG as a dotted line. The shaded area corresponds to the systematic error due the energy scale uncertainty. Statistical errors are shown as thick error bars, and statistical and systematic errors added in quadrature as thin error bars.*

The conclusions are the same than for the uncorrected results: no Monte Carlo simulation describes the absolute value of the data cross section over the full x range. From this figure, it is also difficult to judge on the amount of correlation between the two jets and compare it in different x bins. For a clearer analysis of the effect, the

distributions are normalized to the highest $\Delta\phi$ bin in the data and in the Monte Carlo simulations. The results are shown in figure 7.15. The $\Delta\phi$ distribution in LEPTO in the highest x bin is broader than in ARIADNE. However, in this bin, no weakening of the azimuthal correlation between jets is expected, as the evolution length, in the DGLAP picture as well as in the BFKL scheme, is small. Therefore, this effect can only be due to the combined effects of hadronization smearing and angular resolution of the jets. In the highest x bin, the data points lie in between ARIADNE and LEPTO predictions, so no conclusion can be drawn here. In the lowest two x bins, the $\Delta\phi$ distributions stay the same within the error bars. It is therefore impossible to make a statement on a possible weakening of the correlation between the two jets, either in the data or in the Monte Carlo simulations.

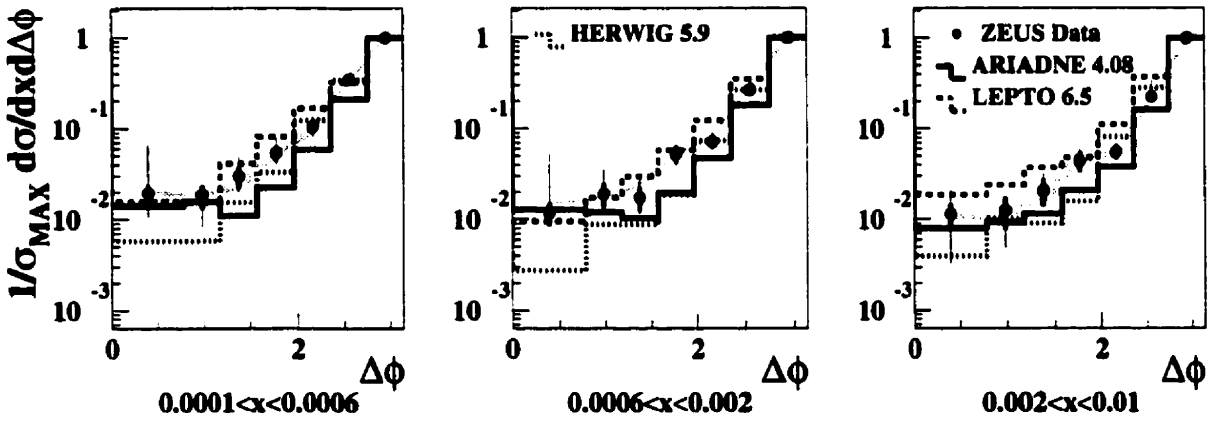

Figure 7.15: *Di-jet cross section corrected for detector effects (hadron level) as a function of $\Delta\phi$ in three x bins in data and Monte Carlo. All the histograms have been normalized to the highest $\Delta\phi$ bin. The data are shown as full dots and the Monte Carlo simulations as histograms: ARIADNE is shown as a full line, LEPTO as a dashed line and HERWIG as a dotted line. The shaded area corresponds to the systematic error due the energy scale uncertainty. Statistical errors are shown as thick error bars, and statistical and systematic errors added in quadrature as thin error bars.*

In figure 7.16, the distribution of the mean $\Delta\phi$ at hadron level is plotted in bins of x. The mean value is defined here as:

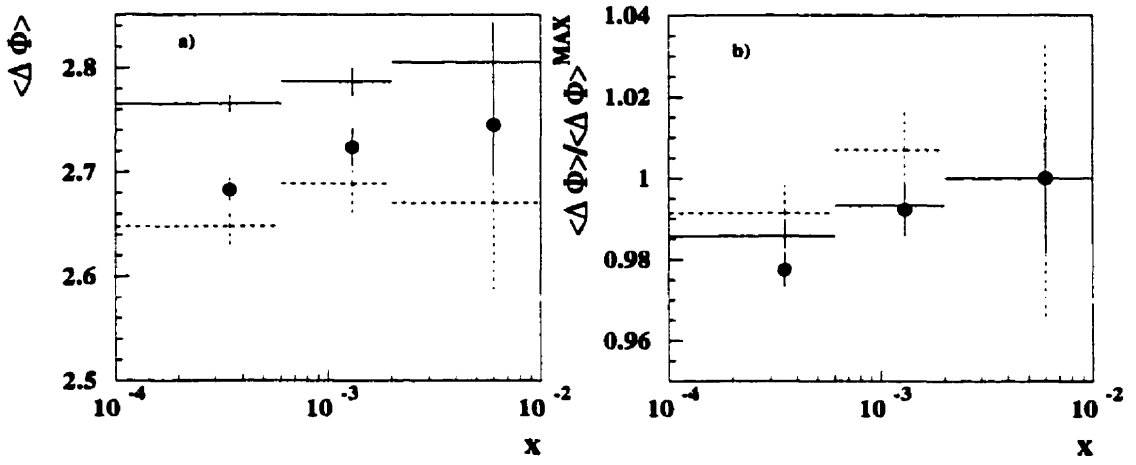

Figure 7.16: *a) Mean value of $\Delta\phi$ plotted versus x. b) Mean value of $\Delta\phi$ normalized to the largest x bin. The data are shown as full dots and the Monte Carlo simulations are shown as histograms: ARIADNE is shown as a full line and LEPTO as a dashed line. Errors shown are statistical only.*

$$< \Delta\phi >= \frac{\sum_i \Delta\phi_i \omega_i}{\sum_i \omega_i}, \tag{7.8}$$

where ω_i is the weight of the distribution in the bin $\Delta\phi$ (that is the number of events in this bin) where i runs over all the $\Delta\phi$ bins. The error quoted is the variance of the mean sample, taken as the variance of the $\Delta\phi$ distribution divided by the sample size, that is the number of entries in the distribution (for a demonstration of this general result, see [123]). In figure a), the mean value is plotted as a function of x, while in figure b) the distributions of the mean has been normalized to the largest x bin. As observed in figure 7.15, the distribution seems to fall down faster in ARIADNE than in LEPTO, however the effect is not significant as it lies within the error bars.

7.4.2 Parton Level Comparison

To be able to relate the results of the analysis to the theoretical predictions, we must correct for the hadronization uncertainties. In practice however, correcting for hadronization effects is not as easy as correcting for detector effects. Every event generator has its own way of treating the parton shower evolution and the fundamental

concept of parton can vary from one generator to another. As an attempt, the data have been corrected to parton level using the ARIADNE model as it describes the data somewhat better than LEPTO (see figure 7.13). The corrected cross section is then defined as:

$$\frac{d\sigma^{parton}}{dxd\Delta\phi} = \frac{N_{events}}{\mathcal{L}_{data}\mathcal{A}_{corr}^{parton}\mathcal{A}_{corr}\Delta x\Delta(\Delta\phi)}, \tag{7.9}$$

where $\mathcal{A}_{corr}^{parton}$ is the amount of hadronization correction. The corrected cross sections can also be compared to NLO programs like MEPJET, DISENT or DISTASTER++. However, it is known that in the NLO calculations, there are divergences which appear mostly when the two leading order jets are correlated back-to-back. When plotting the di-jet cross section as a function of any variable other than $\Delta\phi$, the divergencies are not visible as they are smeared out in each bin. They are however well apparent in the $\Delta\phi$ bins as shown in figure 7.17. A negative cross section is even found around $\Delta\phi \sim \pi$. This is due to infra-red divergences inherent in the theoretical calculations. In order to be able to include the largest $\Delta\phi$ bin, we must then use a binning large enough, so that the divergences cancel out. In practice, one can achieve this by taking the largest $\Delta\phi$ bin such that the number of events entering this bin is at least equal to the sum of all the events in the distribution, excluding this bin [124]. This is the case for our plots, so normalizing all the distributions to the last bin in the NLO calculations is safe. This problem is believed to be the cause of the bad description of the di-jet rate by the various NLO packages (see the previous chapter). Recently, the H1 collaboration showed that taking asymmetric transverse energy cuts for the di-jet system reduces significantly these divergencies [125]. The use of these asymmetric cuts has indeed lead to a good description of the di-jet rate in the data by the NLO calculations. This is why the two jets have been selected with different transverse momentum cuts (see table 7.1).

The final cross sections corrected with ARIADNE 4.08 are shown in figure 7.18. Together with the data cross sections and the Monte Carlo predictions, the NLO calculations as produced by the MEPJET package are shown. The values of the NLO cross sections agree with the data in the last bins of $\Delta\phi$ (for $\Delta\phi \sim \pi$) but disagree at small $\Delta\phi$ values. The small bump in the MEPJET in the smallest x bin is simply due to poor Monte Carlo statistics and should not be taken too seriously.

In figure 7.19, all the cross sections have been normalized to the highest $\Delta\phi$ bin.

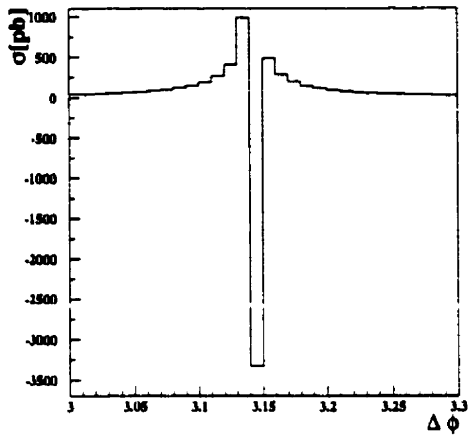

Figure 7.17: *Di-jet cross section as a function of $\Delta\phi$ in the NLO calculations (MEPJET package) in the back-to-back range. All the cuts listed in table 7.1 have been applied.*

No sizable difference is observed between the various Monte Carlo simulations and the data. The $\Delta\phi$ distribution in the data is broader and wider than the MEPJET predictions in the low x bins, which is a sign that the jets are more correlated in MEPJET. However, the dependence on the model and the similarity of the behaviors of LEPTO and ARIADNE at hadron and parton levels prevent us from drawing strong conclusions from this effect.

7.4.3 Systematic Uncertainties

A detailed explanation of the calculation and the size of the systematic uncertainties can be found later in chapter 8. We try here to determine the main contributions to these errors of this analysis. Four main types of systematics are studied and included in the evaluation of the hadron and parton level cross sections:

- the dependence of the corrections on the model: to study this systematic effect, all the corrections have been performed using the LEPTO model (versus ARIADNE).
- the energy scale of the calorimeter: the energy scale has been changed by $\pm 5\%$ to account for the uncertainty of the correction (see section 4.2.1).

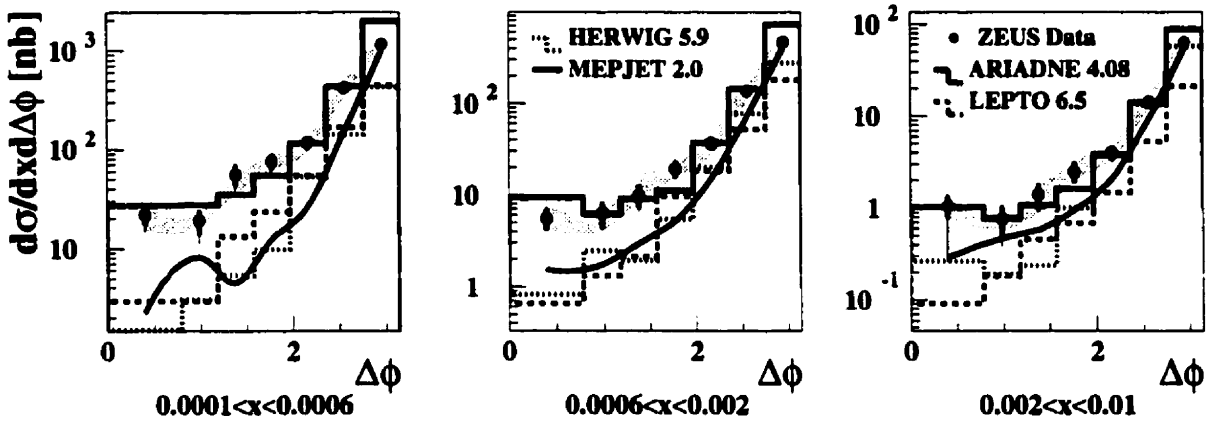

Figure 7.18: *Di-jet cross section corrected for detector and hadronization effects (parton level) as a function of $\Delta\phi$ in three x bins in data and Monte Carlo. The data are shown as full dots and the Monte Carlo simulations as histograms: ARIADNE is shown as a full line, LEPTO as a dashed line and HERWIG as a dotted line. MEPJET is shown as a full line. The shaded area corresponds to the systematic error due the energy scale uncertainty. Statistical errors are shown as thick error bars, and statistical and systematic errors added in quadrature as thin error bars.*

- Systematics on the boost: as the analysis is performed in the HCM frame, it is important to check the influence of the boost. The error on the boost from the laboratory frame to the HCM frame depends mainly on the error of the determination of the energy of the positron. To study this error, we computed the positron energy in a different way, using the double-angle method (see appendix A). The value of the energy of the positron, using the Q^2_{DA} method is: $E_{e'} = \frac{Q^2_{DA}}{2E_e(1+cos\theta_e)}$, where E_e is the value of the positron beam and θ_e is the scattering polar angle of the positron.

- The resolution on the jet variables was tested by adding or subtracting the value of each jet cut by one standard deviation from the resolution (one sigma) of the variable considered. The summary of these changes is in table 7.5.

All the systematic effects but the ones related to the energy scale have been added in quadrature to the statistical errors. The energy scale systematic, being correlated with the other systematic uncertainties, is shown as a shaded band. In all the bins, the dominant systematic effects are the dependences on the model and on the boost.

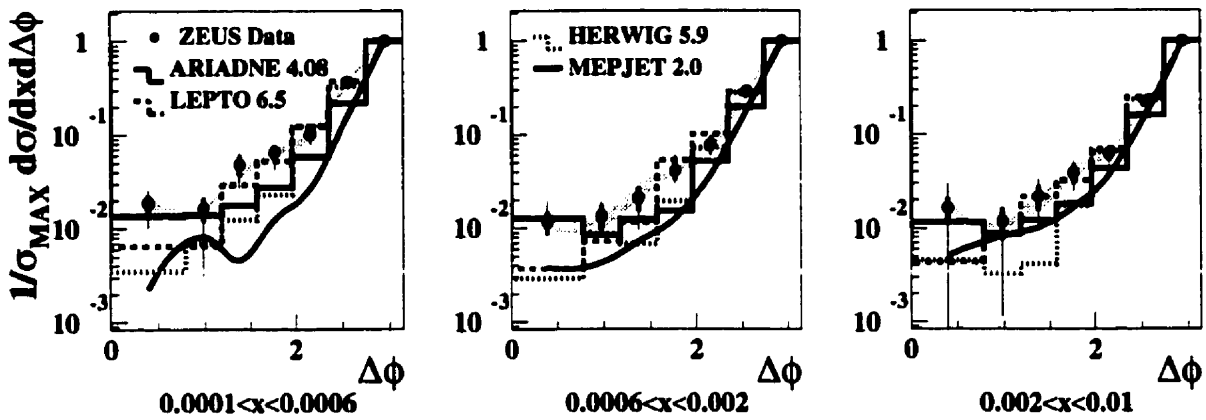

Figure 7.19: *Di-jet cross section corrected for detector and hadronization effects (parton level) as a function of $\Delta\phi$ in three x bins in data and Monte Carlo. All the histograms have been normalized to the highest $\Delta\phi$ bin. The data are shown as full dots and the Monte Carlo simulations as histograms: ARIADNE is shown as a full line, LEPTO as a dashed line and HERWIG as a dotted line. MEPJET is shown as a full line. The shaded area corresponds to the systematic error due the energy scale uncertainty. Statistical errors are shown as thick error bars, and statistical and systematic errors added in quadrature as thin error bars.*

The value of the systematics, together with the value of the data cross section corrected to hadron level is shown in table 7.6.

Jet variable	Nominal cuts	Changed cuts
$E_{T,LAB}^{Det}$	min 4 (6) GeV	min 4.6/3.4 (6.9/5.1) GeV
$E_{T,HCM}^{Det}$	min 4 (6) GeV	min 4.64/3.36 (6.95/5.05) GeV
η_{LAB}^{Det}	max 2.2	max 2.26/2.13
η_{HCM}^{Det}	min 0.	min $0.1/-0.1$

Table 7.5: *Systematic checks on the jet variables. In parenthesis are shown the values of the transverse energy cut for the second jet (asymmetric cuts).*

$\Delta\phi$-range	$\frac{d\sigma}{dxd\Delta\phi} \pm$ stat. $\pm$ syst. [nb]		
	$0.0001 < x < 0.0006$	$0.0006 < x < 0.002$	$0.002 < x < 0.010$
0 – 0.78	$19.4 \pm 4.1^{+19.0}_{-4.6}$	$4.8 \pm 1.2^{+6}_{-0.2}$	$0.63 \pm 0.2^{+0.1}_{-0.1}$
0.78 – 1.18	$18.4 \pm 5.2^{+0.8}_{-3.8}$	$7.8 \pm 2.3^{+2.2}_{-1}$	$0.70 \pm 0.3^{+0.0}_{-0.1}$
1.18 – 1.57	$30.1 \pm 7.9^{+7}_{-3}$	$7.08 \pm 2.0^{+1.8}_{-0.6}$	$1.15 \pm 0.4^{+0.2}_{-0.0}$
1.57 – 1.96	$53.6 \pm 10^{+8}_{-5}$	$21.2 \pm 4.2^{+1.2}_{-1.4}$	$2.45 \pm 0.6^{+0.3}_{-0.1}$
1.96 – 2.36	$103 \pm 14^{+11}_{-8}$	$29.1 \pm 4.2^{+2.5}_{-1.6}$	$3.05 \pm 0.6^{+0.2}_{-0.1}$
2.36 – 2.75	$351 \pm 26^{+7}_{-13}$	$109 \pm 8.5^{+5.1}_{-4.8}$	$12.5 \pm 1.1^{+2.3}_{-0.1}$
2.75 – 3.14	$1000 \pm 41^{+22}_{-55}$	$408 \pm 16^{+17}_{-7.6}$	$55.4 \pm 2.8^{+0.9}_{-1.9}$

Table 7.6: *Cross section values and errors for the corrected data (hadron level).*

7.5 Summary

The azimuthal correlation between the two leading order jets has been evaluated and corrected to the hadron and parton levels in three bins of x. The cross section is however affected by large migrations in the $\Delta\phi$ bins. Moreover, because of the internal resolution of the jet (that is the jet radius), a cut must be applied in $\Delta\eta^{HCM}$ which can hide the weakening of the azimuthal correlation between the leading order jets in the data when reaching small value of x. No major difference between the DGLAP-based models as in LEPTO, and ARIADNE, which implements a BFKL-like, non-ordered parton evolution, has been observed.

At hadron level, the data are consistent with both models within the error bars. HERWIG, however exhibits a much more correlated di-jet system. But this model does not reproduce properly the detector level distributions in the data and must therefore be rejected in our kinematic range. As HERWIG and LEPTO implement the same kind of DGLAP-based mechanism in the parton shower, this is a hint that the non-perturbative effects are dominant in this analysis. The interpretation of the results is therefore quite model-dependent. Moreover, because the DGLAP-based models and

the ARIADNE simulation are consistent with each other and with the experimental data, testing new models like LDC or RAPGAP is not necessary for this analysis.

The data have been corrected to parton level, and compared to NLO calculations. In general, the experimental di-jet system seems to exhibit a smaller amount of azimuthal correlation than the one obtained with the NLO calculations, for the low values of x. This is consistent with the predictions of BFKL. However, the model dependence, included into the correction to the parton level and the agreement of the data with LEPTO at hadron level prevent a strong conclusion to be made out of this observation.

The main obstacle to this analysis is the fact that the azimuthal resolution of the jet is of the same order as the size of the expected effect. The measurement is therefore bound by a fundamental uncertainty, as to whether the effect observed is due to the dynamics of the partonic process or to a feature of the reconstruction of the di-jet system. Moreover, the choice of the cuts does not enable us to isolate a phase space region small enough to distinguish between the contribution from a non-ordered parton evolution, and the contribution from the standard picture, ordered in transverse momentum.

Chapter 8

Forward Jet Cross Section

8.1 Introduction

The previous chapter showed that an indirect measurement of the parton evolution, through the study of the azimuthal correlation between the two leading order jets, did not give a positive BFKL signal. The main problem was that the resolution of the jet (its cone radius) was of the order of the resolution of the effect, that is the azimuthal correlation between the jets. In this chapter, a second attempt is performed to study the parton dynamics by looking directly at the outcome of the parton evolution, i.e. following a parton arising from the evolution process rather than the effects of this parton evolution on the leading order jets.

The strategy of the analysis was explained in section 6.2.2. The aim is to study the forward jet cross section and compare it to various evolution models. If the jet, emitted in the forward direction, has a transverse energy comparable to the four-momentum of the exchange boson, the jet cross section must be suppressed in the low x limit in the usual DGLAP picture. In the BFKL scheme, however, the cross section is expected to rise as a power of $\frac{1}{x}$, and therefore a clear signal might be expected. We saw in chapter 6 that the rise of the forward jet cross section, in the BFKL picture, is weakened by NLO corrections. However, no prediction on the size of these corrections has been made yet and it is interesting to evaluate the cross section in the data.

The cuts for the forward jet analysis have been explained in section 6.2.2. They are summarized in table 8.1. The final purpose is to evaluate the forward jet cross section

in bins of x, and probe in particular the low x region. The plan of this chapter is essentially the same than the previous one: the cross sections at detector level are first compared to the various Monte Carlo predictions. Then the various corrections are applied to the data: energy corrections, acceptance, purity and efficiency. The various systematic errors are then discussed. Finally the cross sections are obtained at hadron and parton level, to be discussed and interpreted in the light of the Monte Carlo predictions and the various theoretical calculations.

$E_{e'} > 10$ GeV
$y > 0.1$
$\eta_{Jet} < 2.6$
$E_{T,Jet} > 5$ GeV
$x_{Jet} > 0.036$
$0.5 < E^2_{T,Jet}/Q^2 < 2$
$p_{Z,Jet}(Breit) > 0$ GeV/c
$4.5 \cdot 10^{-4} < x < 4.5 \cdot 10^{-2}$

Table 8.1: *Selected phase space region for the cross section measurement.*

8.2 Comparison between the Data and the Monte Carlo Simulations

In this analysis, we are mostly interested in the absolute value of the jet cross section in the selected phase space. In figure 8.1, some general kinematic variables (Q^2, y_{JB}, $E_{e'}$ and $E - P_Z$) are compared to three Monte Carlo predictions: ARIADNE 4.08, LEPTO 6.5 and HERWIG 5.9. To obtain these plots, all the cuts presented in table 8.1 have been applied. Contrary to what was shown in the previous chapter, ARIADNE describes fairly well the data. Both the shape and the absolute normalization are well reproduced. On the other hand, the MEPS-based models, LEPTO and HERWIG, exhibits a cross section which is in average 40% too small compared to the data. This is a first hint that the physics implemented in these Monte Carlo models does not yield a good description of the hard process in the phase space under consideration.

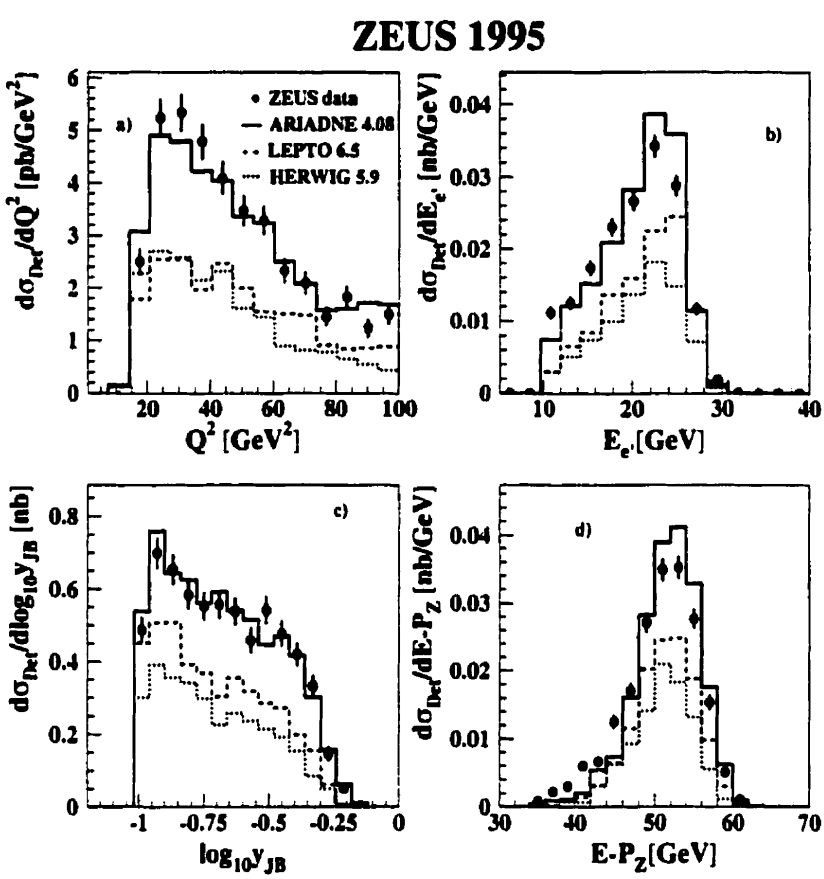

Figure 8.1: *Shape comparison between the data (shown as full dots) and Monte Carlo simulations for four variables: a) Q^2, b) $E_{e'}$, c) y_{JB}, d) $E - P_Z$. The data are displayed as full dots. ARIADNE is shown as a full line, LEPTO as a dashed line and HERWIG as a dotted line. Statistical errors only.*

In figure 8.2, the jet variables, which are used in the cuts (i.e. E_T, η, x_{jet} and E_T^2/Q^2), are compared to the Monte Carlo predictions. In these plots, all the cuts defined in table 8.1 have been applied to the selection procedure, except the one on the plotted variable. This allows a study of the cross section outside the region where the contributions due to a BFKL-type dynamic are expected to dominate. The shaded area corresponds to the region outside the kinematic limits of the phase space defined by table 8.1.

Once again, the variables are fairly well described by ARIADNE, whereas LEPTO and HERWIG yield a cross section too small by around 40%. The last plot of figure 8.2,

d), is particularly revealing. The variable E_T^2/Q^2 is plotted in an extended range and compared to the various Monte Carlo predictions. In section 6.2.2, the importance of this variable was stressed: for the forward jets, it gives an indication on the type of ordering which is considered. One can basically distinguish three regions in this plots, which correspond to three types of evolution mechanism:

- $Q^2 >> E_T^2$: in this region, the scale of the process is set by the four momentum of the exchange boson. This is consistent with the DGLAP picture. The forward jet has there a low E_T with respect to Q^2 and the models and the experimental data converge to the same value. This corresponds to the high Q^2 limit where the DGLAP scheme is known to give correct results.

- $Q^2 \sim E_T^2$: this is the region we are interested in. There, the DGLAP mechanism is suppressed due to the requirement of strong ordering and the two MEPS models exhibit much smaller cross sections than the data, while ARIADNE, which has not implemented the requirement of strong ordering, describes the data correctly.

- $Q^2 << E_T^2$: this is the most intriguing phase space region. The scale of the process is set by the transverse momentum of the forward jet. A backward evolution is possible, from the quark box to the virtual boson, which is similar to the photoproduction case. In this region, none of the Monte Carlo models describes the data.

As ARIADNE describes the data best, this Monte Carlo simulation is used to perform the detector corrections.

8.3 Detector Correction

8.3.1 Energy Correction

As for the previous analysis, an energy correction is applied to the data as well as to the detector level of the Monte Carlo simulation. The method is very similar to the one presented in the section 7.3.1 so we refer the reader to it for more details. The difference is that the energy correction, here, is averaged over all the pseudorapidity range of the jets, rather than dividing the sample on different η bins. This is justified

as the η range of the jets is small and limited to the forward calorimeter (cf. figure 8.2). The value of the energy corrections is shown in figure 8.3. Once again, the (transverse) energy of the jet found in the detector is multiplied by the corresponding value from the fit.

8.3.2 Resolution of the Kinematic Variables

The phase space of this analysis is constrained by a variety of cuts which are determined by the performance of the detector. Each cut can indeed be associated with a certain purity and efficiency which depends on the resolution of the variable on which this cut is applied. Therefore, when all the cuts are applied to get the final sample, the overall purity and efficiency are reduced by a certain amount which depends on the number of cuts applied and the resolution of each variable used in the cuts.

As the final distribution is investigated in bins of x, it is very important to determine its resolution, in order to obtain the bin size of the final distribution. In table 8.2, the resolution of x in each selected bin is shown. The width of the bins is set to be at least twice the resolution in x in order to account for migrations from bin to bin. The last bins are larger in order to have a reasonable amount of statistics.

x range ($\times 10^{-3}$)	0.45-0.8	0.8-1.4	1.4-2.5	2.5-4.5	4.5-8	8-14	14-25	25-45
$\sigma(x)$ ($\times 10^{-3}$)	0.07	0.16	0.41	0.78	1.5	3.6	4.7	12

Table 8.2: *Size of the bins in x and resolution of x in each of the bins.*

The resolution of four jet variables is shown in figure 8.4: E_T, η, x_{jet} and E_T^2/Q^2. As in chapter 7, the absolute and relative resolution on the variable X are defined as:

$$\Delta X = X^{Det} - X^{Had}. \tag{8.1}$$

$$\frac{\Delta X}{X} = \frac{X^{Det} - X^{Had}}{X^{Had}}, \tag{8.2}$$

The typical relative E_T resolution of the jet is 11 % whereas the η_{Jet} resolution is around 0.1. The x_{Jet} resolution is around $6.5 \cdot 10^{-3}$. The worst resolution is in E_T^2/Q^2: it is around 25%. As all these variables are included in the analysis cuts, these resolutions strongly affect the purities and efficiencies of the final sample, especially the E_T^2/Q^2 resolution (see below).

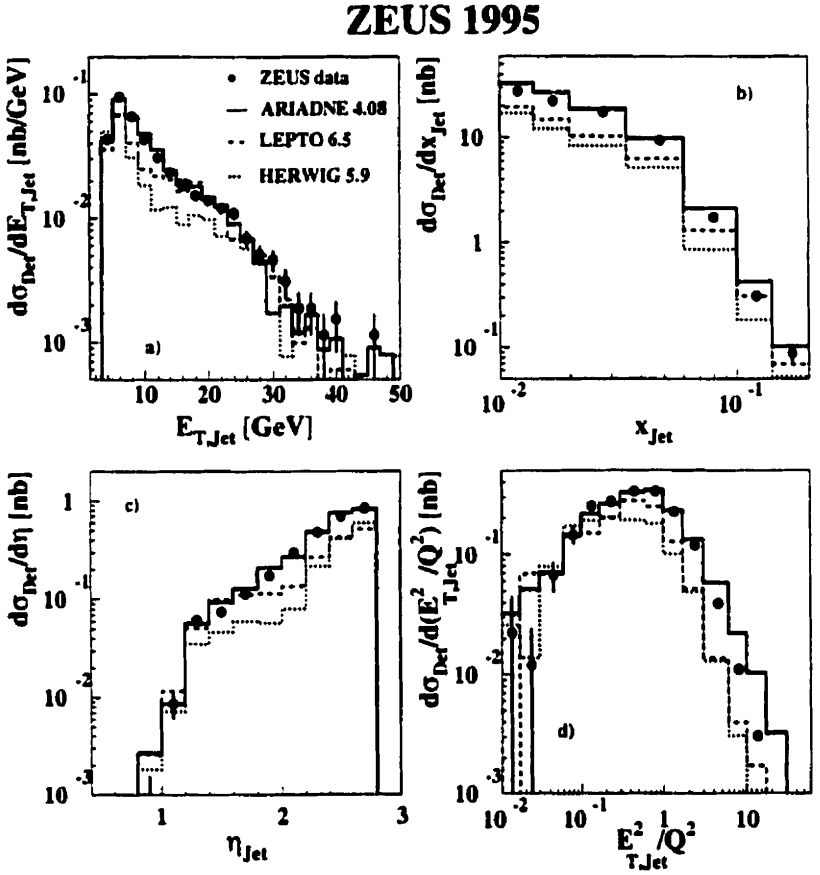

Figure 8.2: *Shape comparison between the data (shown as full dots) and Monte Carlo simulations for four jet variables: a) E_T, b) η, c) x_{jet}, d) E_T^2/Q^2. The data are displayed as full dots. ARIADNE is shown as a full line, LEPTO as a dashed line and HERWIG as a dotted line. Statistical errors only. The shaded region corresponds to the value of the variable outside the kinematic domain defined by the cuts of table 8.1.*

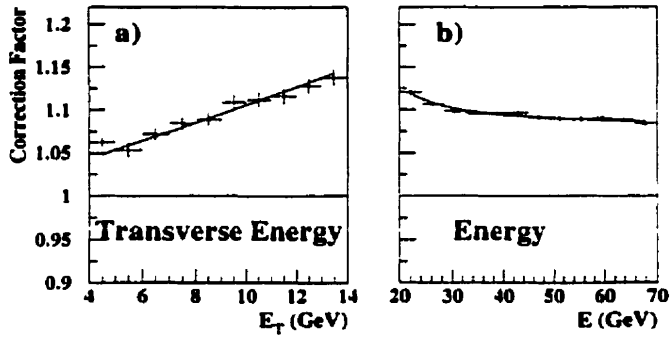

Figure 8.3: *Jet energy correction factors, defined as $E^{Had}_{(T)}/E^{Det}_{(T)}$, as a function of the energy and the transverse energy. The Monte Carlo used for this plot is ARIADNE 4.08.*

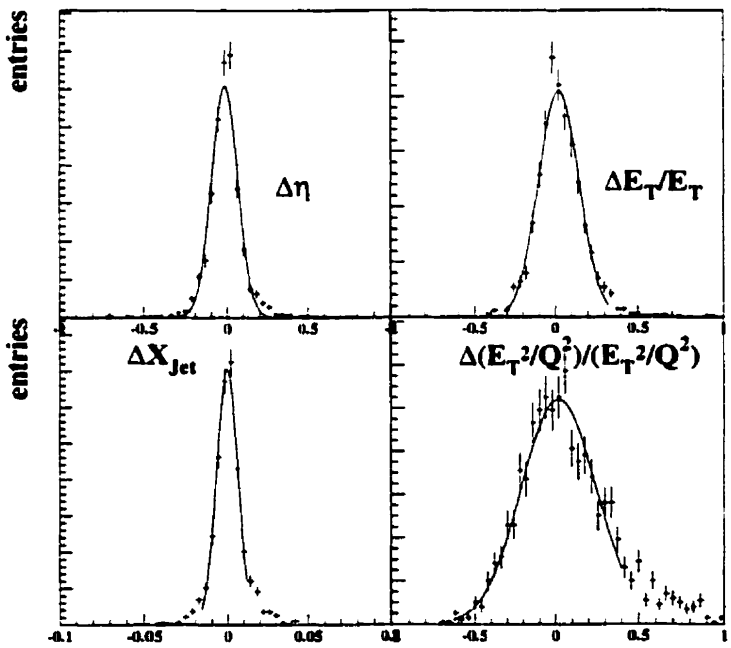

Figure 8.4: *Resolution for 4 jets variables. Δ = ()$_{TRUE}$ - ()$_{DETECTOR}$. ARIADNE was used to produce this plot.*

8.3.3 Acceptance Correction, Purity, Efficiency

The bin-by-bin correction procedure, similar to the one used in the azimuthal correlation analysis, is performed here. We refer the reader to section 7.3.3 for more information. The value of the purities, efficiencies and acceptance correction factors are shown in figure 8.5. The purities and efficiencies are typically on the order of 20-30%. These values are limited by the resolution of the cuts which are applied on the jet variables, in particular the cut on E_T^2/Q^2. When this cut is removed, the values of the purity and efficiency increases by 30-50%,depending on the x bin (see figure 8.6).

As a consistency check, the same values are computed using the LEPTO model. They are shown in figure 8.7. Both efficiency and purity are slightly smaller than the values obtained with ARIADNE but the value of the correction factors remain fairly consistent showing that the choice of the model has only a small influence on the detector corrections.

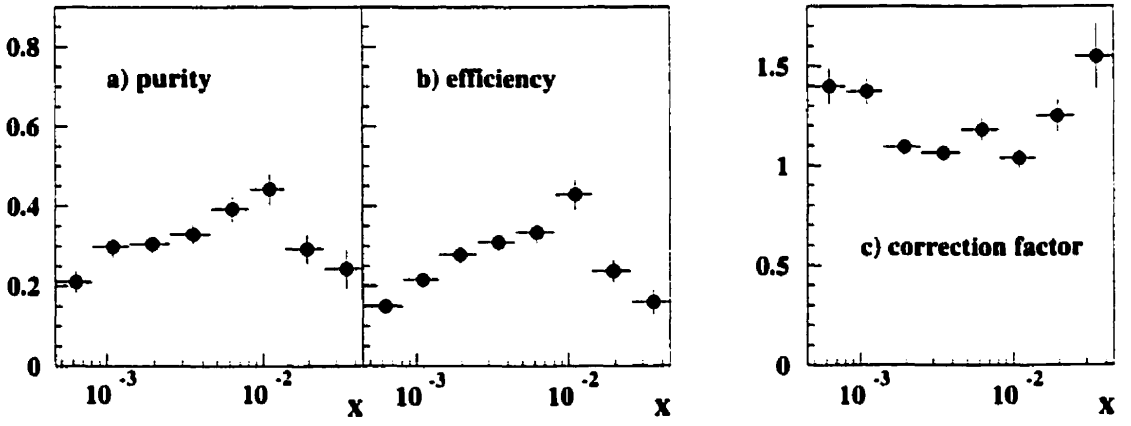

Figure 8.5: *Hadron to detector purities, efficiencies and (bin by bin) correction factors of the event sample for the analysis based on the cone algorithm. These values have been obtained with the ARIADNE model.*

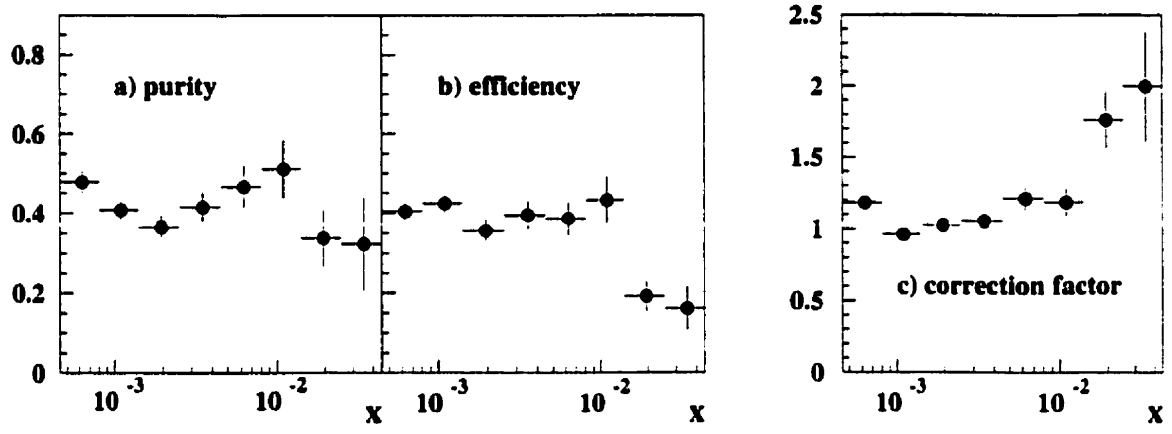

Figure 8.6: *Hadron to detector purities, efficiencies and (bin by bin) correction factors of the event sample for the analysis based on the cone algorithm. The cut on the variable E_T^2/Q^2 is removed. These values have been obtained with the ARIADNE model.*

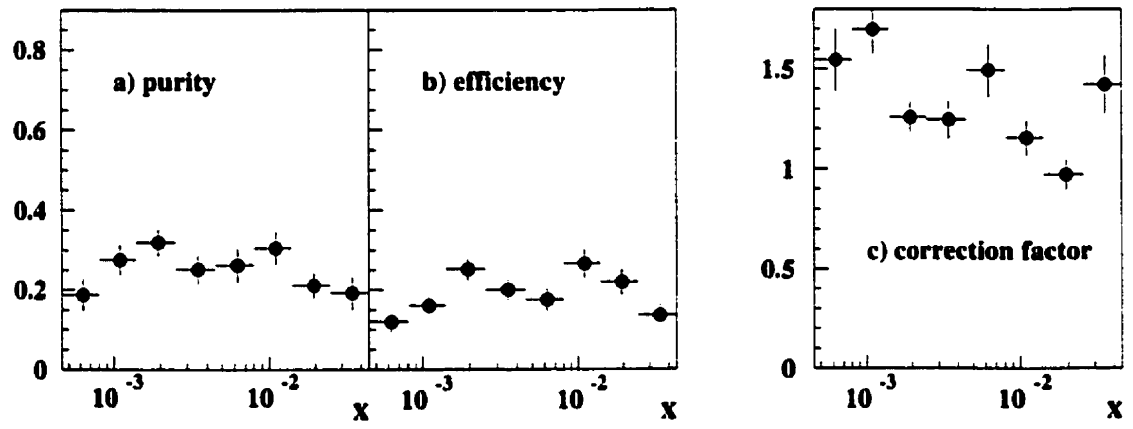

Figure 8.7: *Hadron to detector purities, efficiencies and (bin by bin) correction factors of the event sample for the analysis based on the cone algorithm. These values have been obtained with the LEPTO Monte Carlo.*

8.4 Forward jet Cross Sections

Before turning towards the final aim of the study, the comparison of the corrected results to the theoretical calculations, we shall have a brief look at the uncorrected distributions. In section 8.2, some uncorrected distributions were presented and compared to various Monte Carlo distributions. The behavior of the E_T^2/Q^2 distribution turned out to be quite meaningful on the behavior of the parton evolution at various values of Q^2.

The uncorrected differential jet cross section is calculated as:

$$\frac{d\sigma}{dx} = \frac{N_{events}}{\mathcal{L}_{data}\Delta x}, \tag{8.3}$$

where $\mathcal{L}_{data}$ is the integrated luminosity of the sample, Δx is the bin size in the variable x and N_{events}, the number of events in the Δx bin.

In figure 8.8, the uncorrected differential forward jet cross section in the data is compared to the three Monte Carlo predictions obtained with ARIADNE, LEPTO and HERWIG. The trend which appears in the plots shown in the previous section is confirmed: ARIADNE describes the data best, while LEPTO and HERWIG, the two Monte Carlo models based on the DGLAP equations, exhibit significantly smaller cross sections.

Figure 8.9 pictures a typical event containing a forward jet in the ZEUS detector. A large amount of transverse energy is seen in the forward calorimeter, corresponding to the forward jet, together with a fraction of energy in the barrel calorimeter (which can be attributed to the leading order parton or current jet). A positron is clearly reconstructed in the rear calorimeter.

8.4.1 Hadron Level Comparison

The cross section corrected to hadron level is calculated in the following way:

$$\frac{d\sigma}{dx} = \frac{N_{events}}{\mathcal{L}_{data}\mathcal{A}_{corr}\Delta x}. \tag{8.4}$$

$\mathcal{L}_{data}$ and Δx have been defined above and $\mathcal{A}_{corr}$ is the acceptance correction which is shown in figure 8.5. The values of the cross section corrected to the hadron level

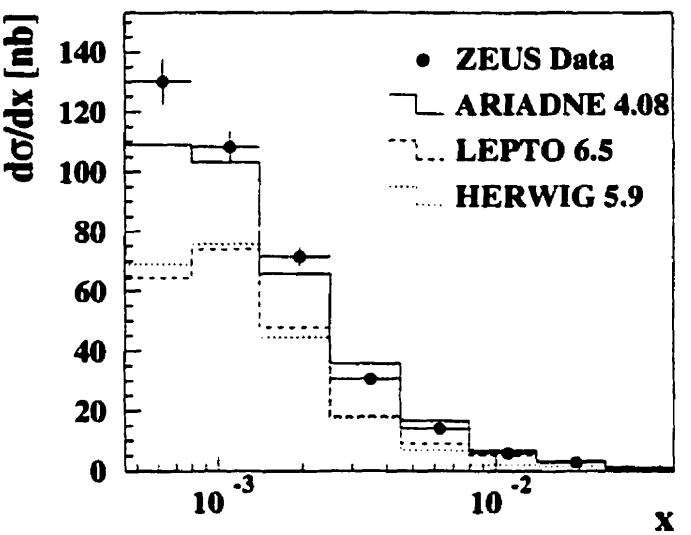

Figure 8.8: *Differential cross section as a function of x for the uncorrected data (CONE algorithm and all cuts applied). Data are shown as full dots and Monte Carlo as histograms. Only statistical errors are shown.*

are displayed in table 8.3, together with the statistical and systematic errors.

The data corrected to hadron level are compared to the Monte Carlo simulations in figure 8.10a). There, four Monte Carlo simulations are plotted together with the data: ARIADNE 4.08, LEPTO 6.5, HERWIG 5.9 and the new model LDC 1.0 described in section 3.2.3.

As in the previous section, ARIADNE is the only model which reproduces the data in absolute cross section, including the low-x limit. The two MEPS-based models exhibit a smaller cross section. This is expected as these Monte Carlo simulations do not implement the type of hard physics which is probed here. The larger value of the cross section measured in the data is a first clear experimental hint that higher order contributions to the leading order DGLAP equations contribute significantly to the partonic processes.

The figure 8.10 b) shows the same differential cross section but with a logarithmic y scale. This is made in order to emphasize on the high-x region. LEPTO, ARI-

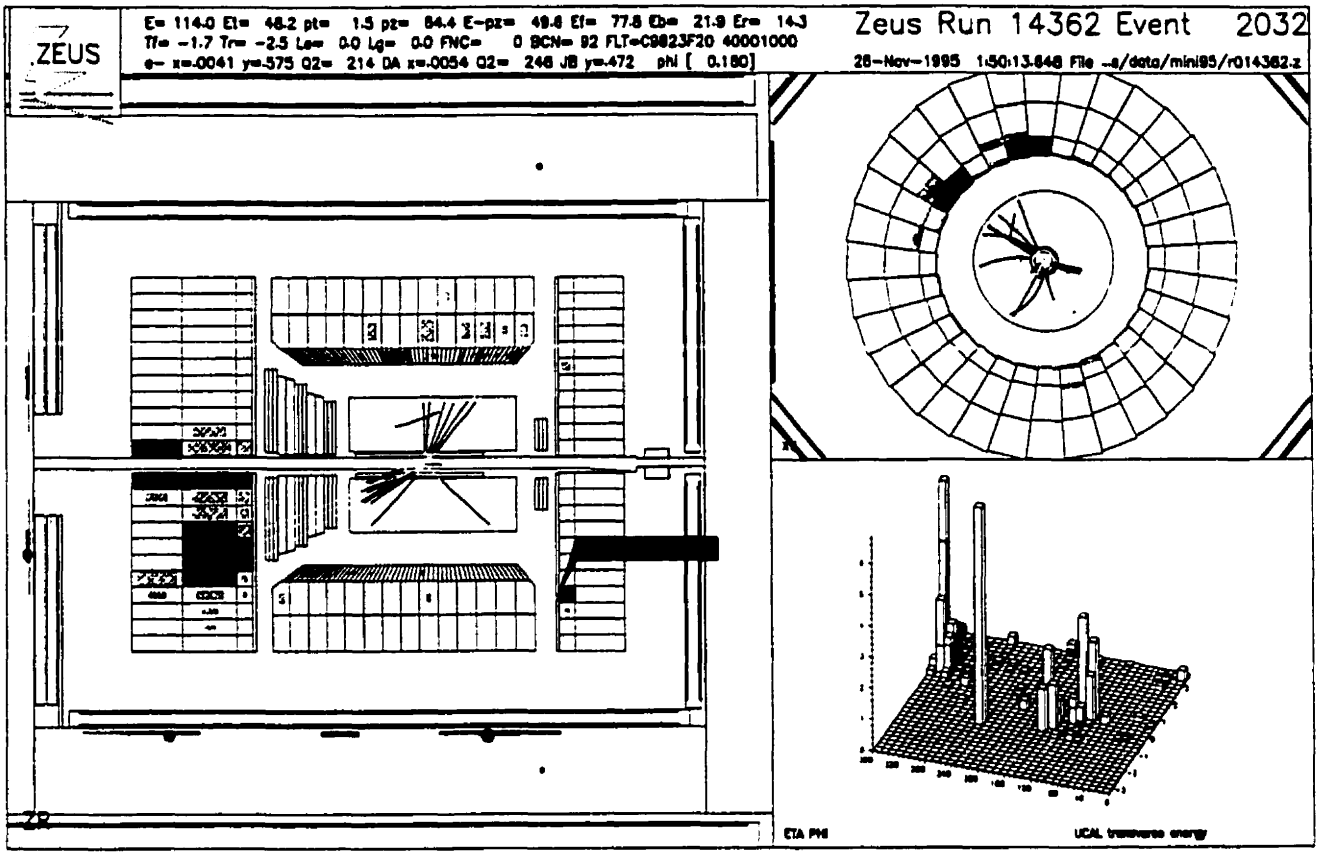

Figure 8.9: *Picture of an event found in the ZEUS detector and containing a forward jet. The display is divided in three parts corresponding to three different views in the ZEUS detector: the XZ plane is the largest one, the top right one is the transverse cross section containing the track pattern in the CTD and energy deposited in the BCAL and on the bottom right side, the amount of transverse energy deposited in the calorimeter is shown in the (η,ϕ) plane.*

ADNE and the corrected data converge, while HERWIG does not. In this region, the convergence is indeed expected between the data and the DGLAP-based models given that the evolution parameter for BFKL, that is the evolution length in which the BFKL-type dynamics can contribute, is suppressed by a factor $ln\frac{x_{jet}}{x}$), where x is on the same order as x_{jet}. HERWIG however still does not agree with the data. A tentative explanation is given below.

The LDC model, which implements the CCFM equation, does not describe the data either at low x. This is a puzzle since this Monte Carlo simulation implements the type of hard physics expected to give rise to the BFKL-like effects. The comparison between LDC and the data in other analyses (jet rates, transverse energy flow) shows

the same kind of results: LDC predicts a cross section (or absolute rate) which generally lies between ARIADNE and LEPTO (cf. [125]). As the structure function used to generate the LDC events is different from the other simulations (cf. chapter 3), an attempt was performed to compare the jet rates only. The forward jet rate is defined by:

$$R_{fj} = \frac{\sigma_{fj}}{\sigma_{tot}},$$

where σ_{tot} is the total cross section determined in the phase space under consideration and σ_{fj}, the cross section for the events containing a forward jet. As both cross sections depend on the structure function F_2 of the proton, the ratio must be independent of F_2 in the first order. The forward jet rate in the data is plotted in figure 8.11 and compared to ARIADNE, LEPTO and LDC. Here, it can be seen that the forward jet rate in the LDC model lies in between ARIADNE and LEPTO. The previous result (for the total cross section) is therefore confirmed and LDC does not describe the data results.

In the high-x region of figure 8.10 a), LDC does not converge to the value of the data cross section but to HERWIG's predictions. This might result from the requirement of strong ordering in the emission angle of the partons along the gluon ladder, requirement which is similar to the implementation of the colour coherence in HERWIG. In this case also, the emitted parton is supposed to lie within a cone around the parent parton. At high-x, both features of HERWIG and LDC must be similar [126].

x-range	$\frac{d\sigma}{dx} \pm$ stat. $\pm$ syst. [nb]	syst. ($E_{T,Jet}$-scale) [nb]
$4.5 \cdot 10^{-4} - 8.0 \cdot 10^{-4}$	$114.0 \pm 9.7^{+29.0}_{-15.0}$	$(-5.9, +18)$
$8.0 \cdot 10^{-4} - 1.4 \cdot 10^{-3}$	$96.2 \pm 6.5^{+8.2}_{-8.2}$	$(-8.1, +7.8)$
$1.4 \cdot 10^{-3} - 2.5 \cdot 10^{-3}$	$77.8 \pm 4.7^{+5.2}_{-6.9}$	$(-4.2, +7.0)$
$2.5 \cdot 10^{-3} - 4.5 \cdot 10^{-3}$	$34.4 \pm 2.2^{+3.8}_{-1.9}$	$(-2.1, +2.6)$
$4.5 \cdot 10^{-3} - 8.0 \cdot 10^{-3}$	$14.1 \pm 1.0^{+2.5}_{-1.2}$	$(-1.2, +1.3)$
$8.0 \cdot 10^{-3} - 1.4 \cdot 10^{-2}$	$6.53 \pm 0.54^{+0.10}_{-0.70}$	$(-0.7, +0.2)$
$1.4 \cdot 10^{-2} - 2.5 \cdot 10^{-2}$	$2.65 \pm 0.25^{+0.30}_{-0.30}$	$(-0.03, +0.20)$
$2.5 \cdot 10^{-2} - 4.5 \cdot 10^{-2}$	$0.65 \pm 0.09^{+0.10}_{-0.40}$	$(-0.00, +0.05)$

Table 8.3: *Cross section values and errors for the corrected data after all the cuts defined in table 8.1. The last column shows the systematic error due to the energy scale uncertainty of the calorimeter, which is not included in the central column. It corresponds to the shaded band in figure 8.10.*

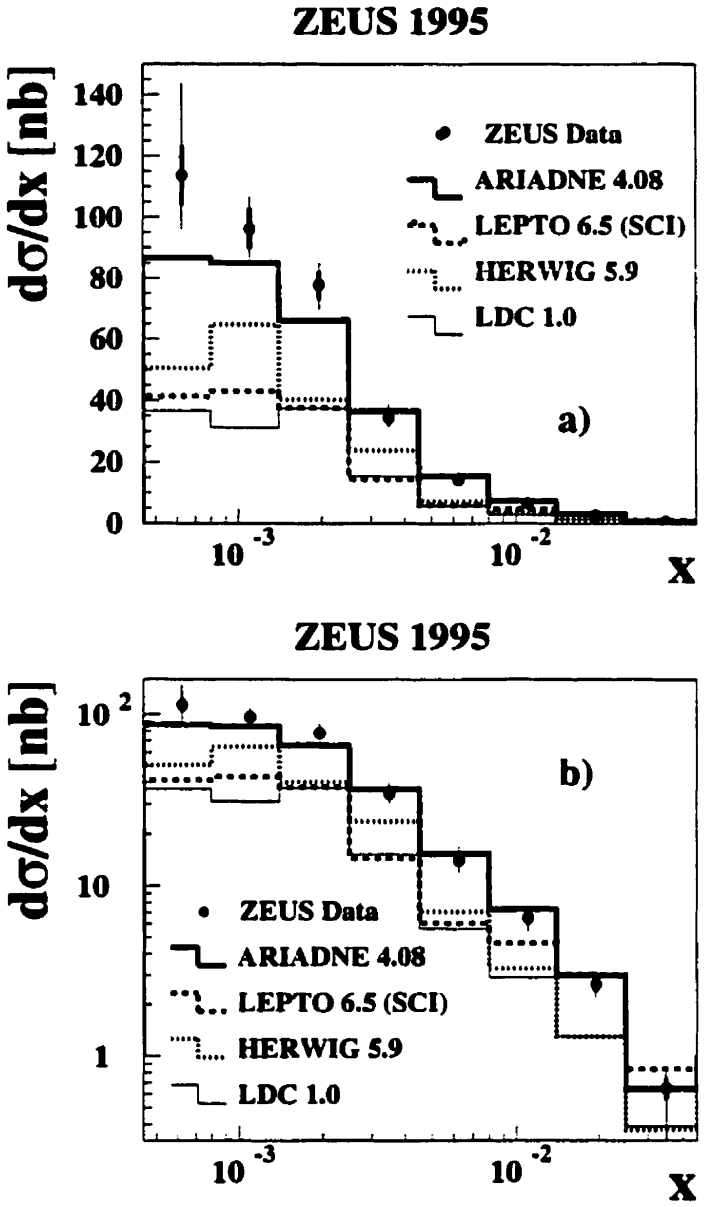

Figure 8.10: *a) Differential hadron level forward cross section as a function of x. b) Same as a) but in double logarithmic scale. The cross section is measured in the region:* $\eta_{Jet} < 2.6$, $x_{Jet} > 0.036$, $0.5 < E^2_{T,Jet}/Q^2 < 2$, $E_{T,Jet} > 5$ *GeV*, $E_{e'} > 10$ *GeV*, $y > 0.1$. *Statistical errors are shown as thick error bars, and statistical and systematic errors added in quadrature as thin error bars. The errors due to the uncertainty of the jet energy scale are correlated for the different x bins and therefore not included in the error bars but given as the shaded band.*

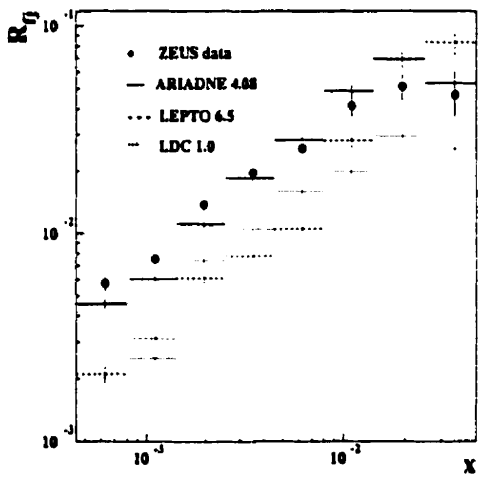

Figure 8.11: *Corrected forward jet rate in the data compared to three Monte Carlo simulations. The data are shown as full dots. Errors shown are statistical only.*

8.4.2 Comparison at Parton Level and Discussion of the Results

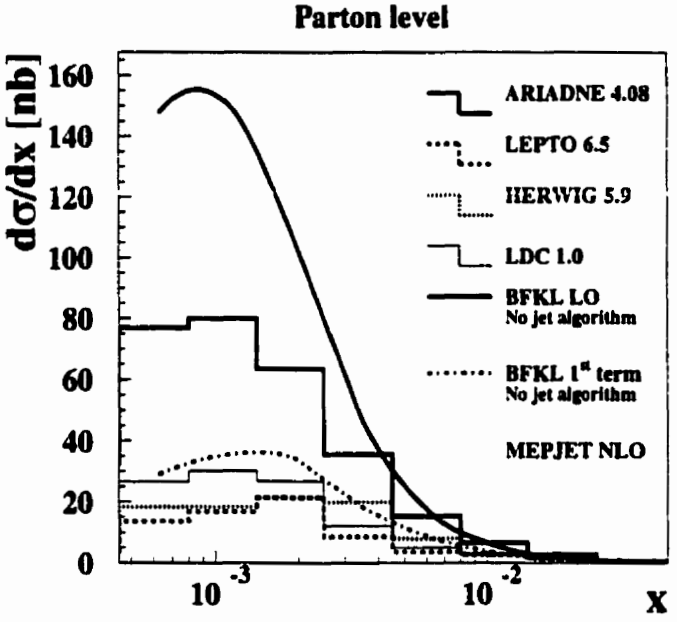

Figure 8.12: *Predictions of the NLO and the BFKL calculations together with the Monte Carlo predictions at parton level. The theory predictions are shown as smooth curves, while the Monte Carlo cross sections are shown as histogram: ARIADNE is shown as a thick line, LEPTO as a dashed line, HERWIG as a dotted line and LDC as a thin line. The band shown for the NLO predictions corresponds to the possible range on the renormalization and factorization scales.*

The ultimate goal of this study is to probe the perturbative expansion of the partonic evolution. The BFKL calculations, as well as the MEPJET NLO predictions, only take into account this perturbative region, ignoring the hadronization phase.

In the Monte Carlo simulation, the parton level, as defined in chapter 3 is not such a well defined quantity. The partons are taken at the end of the parton shower (or dipole radiation in the case of ARIADNE), at a scale where non-perturbative effects can be large. The comparison between the parton level in the various Monte Carlo used and the theoretical calculations is shown in figure 8.12 within the set of cuts defined in table 8.1. The NLO predictions within the factorization (renormalization) scale $0.25K_T^2 < \mu^2_{F(R)} < 2K_T^2$ are in good agreement with the parton level of the

DGLAP-based Monte Carlo simulations[1], meaning that in this kinematical domain, the parton showers discussed in section 3.2.1 reproduce correctly the NLO effects in terms of shapes and total cross sections. These cross sections are however strongly suppressed with respect to the full BFKL calculations. The predictions using only the first term of the BFKL calculations comes closer to the DGLAP and NLO predictions. This corresponds to the term in $ln(x/x_{jet})$ in the BFKL resummation, and is therefore the equivalent to the NLO corrections to the DGLAP equations. The slight discrepancy which is observed between the BFKL first term and the MEPJET predictions can be due to a normalization uncertainty in the BFKL picture and is currently under investigation (one can show that the absence of jet algorithm in the NLO prediction does not change significantly the value of the forward jet cross section [114]). ARIADNE lies in between the DGLAP-based predictions and the BFKL predictions. The LDC Monte Carlo exhibits a larger cross section than the MEPS Monte Carlo simulations, but still much smaller than ARIADNE. It is indeed consistent with the NLO predictions. Efforts are currently performed to understand the reasons for this discrepancy with the BFKL predictions.

Before comparing these predictions to the data, care must be taken of the hadronization corrections. To this end, a bin-by-bin correction is used and the partonic cross section in the data is defined by:

$$\frac{d\sigma}{dx} = \frac{N_{events}}{\mathcal{L}_{data}\mathcal{A}_{corr}^{parton}\mathcal{A}_{corr}\Delta x}. \tag{8.5}$$

$\mathcal{A}_{corr}^{parton}$ is the correction from the hadron to the parton level. These corrections are however model dependent. In figure 8.13, the hadronization corrections are plotted for the four Monte Carlo simulations we compared to the data. The striking feature of the plot is the disagreement between the corrections for the DGLAP-based models (HERWIG and LEPTO) and the BFKL-like models (ARIADNE and LDC). The latter present small and flat corrections over the full x range, whereas the former have very large corrections at small x. This does not really come as a surprise, as the DGLAP-based models do not have the kind of hard physics which can describe the topology of the event within the selected phase space. In order to generate a jet of order α_s^2 or higher with a transverse energy comparable to the Q^2 of the event, the DGLAP-based simulations would need to implement NLO corrections. Without that,

[1] K_T^2 is the sum of the transverse momenta of the partons in the Breit frame. The choice of the scale K_T^2 is performed in accordance with the minimal sensitivity criteria, see [127].

the jets which are generated in the parton shower have a small transverse momentum by construction.

The correct procedure would require to take into account the model dependence of the hadron-to-parton correction factors into the calculation of the systematic errors. Lacking two models which describe correctly the data at hadron level, we are bound to estimate this error with one of the MEPS model. The difference between the data corrected to parton level with ARIADNE and with LEPTO is shown in figure 8.14. In case a), the data corrected with ARIADNE, agree fairly well with this Monte Carlo and are much above the DGLAP-based predictions. In the second case (b), the correction factors are calculated with LEPTO and the data cross sections lie much below the ARIADNE predictions (but are still larger than the LEPTO predictions by a factor 1.5 to 2).

Of course, correcting the data cross sections with a Monte Carlo which does not describe them is not quite a valuable procedure. However, so far, only one Monte Carlo could match the data, so that a large uncertainty remains in the interpretation of the data corrected to parton level.

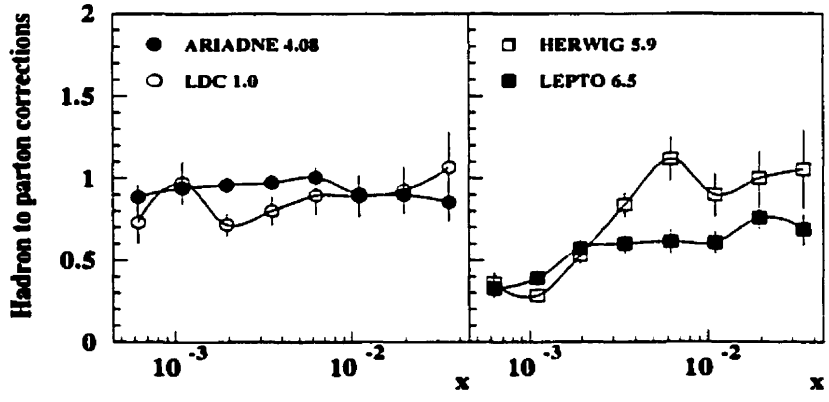

Figure 8.13: $\mathcal{A}_{corr}^{parton}$ *Hadron to parton correction factors for the phase space region defined by the cuts listed in table 8.1, for four different Monte Carlo simulations: ARIADNE 4.08, LEPTO 6.5, HERWIG 5.9 and LDC 1.0.*

The above results yield finally another puzzling fact: from the hadron-to-parton correction factors, it is seen that LEPTO and HERWIG predict a large amount of jets

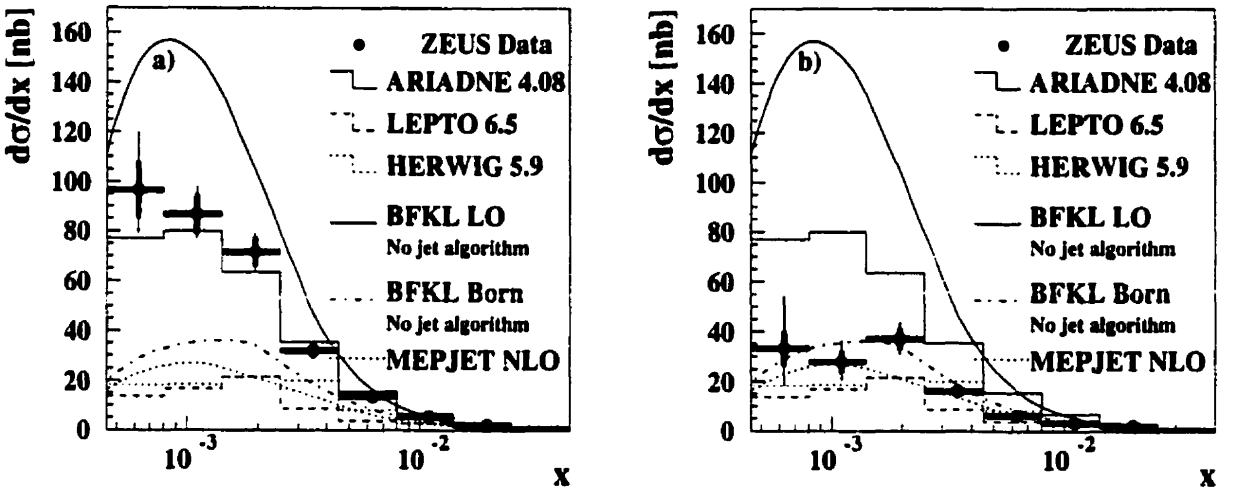

Figure 8.14: *Parton level differential cross section as a function of x for the data corrected for detector and hadronization effects. Data are shown as full dots and Monte Carlos as histograms (the cuts listed in table 8.1 have been applied. The thick error bars on the data points are the statistical errors and the full error bar (in thin line) is the systematic + statistical error added in quadrature. The energy scale error is represented as a hashed area around the data points. a) Data corrected with the ARIADNE Monte Carlo. b) Data corrected with the LEPTO Monte Carlo.*

at the hadron level which do not have a corresponding partner at the parton level. That is, most of the jets predicted by LEPTO at hadron level and at low x arise from some "energy fluctuation" of non-perturbative nature. To estimate the origin more exactly, we try to compare LEPTO generated jets with and without the Soft Colour Interactions described in section 3.3.3. The comparison with the data cross sections corrected with LEPTO (with the flag SCI turned on) and the LEPTO predictions with and without SCI is shown in figure 8.15. At hadron level (a), the absence of SCI makes a big difference (almost a factor two in the lowest x bins), whereas at parton level (b), the discrepancy is hardly noticeable (but this is expected as the SCI is a long range effect, taking place after the parton shower). Thus, a large number of the jets that are predicted by LEPTO at hadron level arise from a purely non-perturbative phenomenon, as a consequence of an excess of energy in the forward region produced by a different string topology (see section 3.3.3). This is confirmed by the comparison of the hadronization corrections shown in figure 8.16. The hadron-to-parton corrections are shown for ARIADNE, LEPTO with SCI and LEPTO without SCI. There again, the hadronization corrections are much larger when the SCI flag is turned on

instead of off.

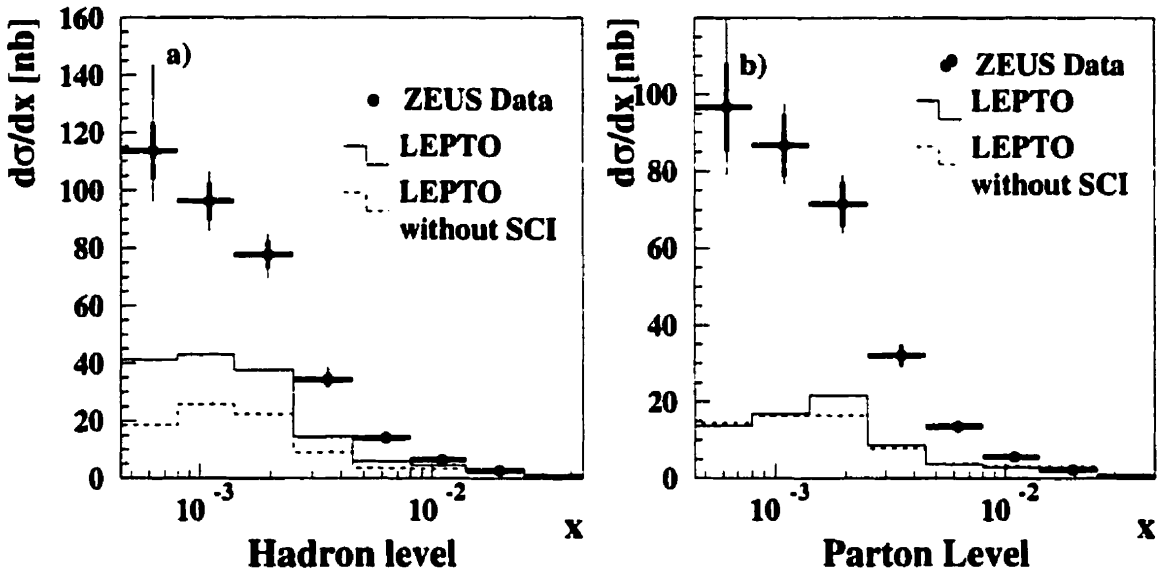

Figure 8.15: *Forward jet cross section (corrected with LEPTO including soft colour interactions) compared to LEPTO with and without SCI at hadron level (a) and parton level(b). All the cuts defined on table 8.1 have been applied. The data are shown as full dots and the Monte Carlo simulations as histograms: LEPTO with SCI is shown as a full line histogram and LEPTO without SCI is shown as a dashed line histogram. The statistical errors are shown as thick error bars. The systematic error added in quadrature with the statistical errors are shown as thin lines.*

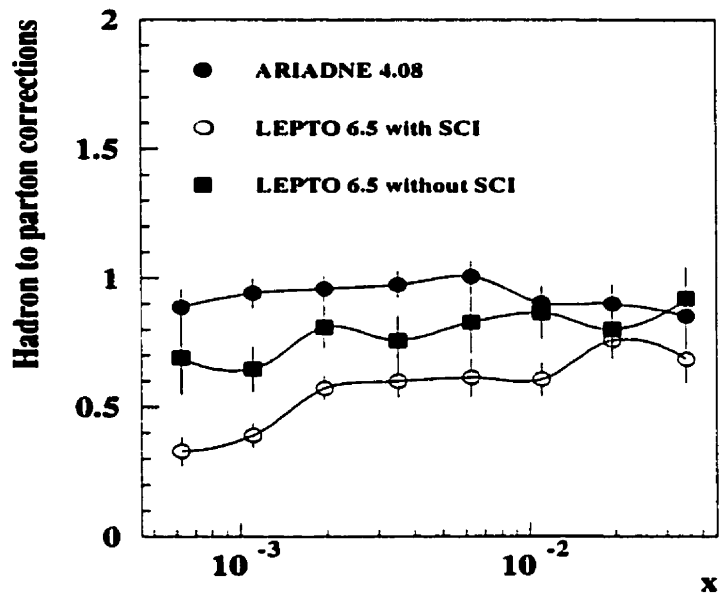

Figure 8.16: *Hadron to parton corrections for the phase space region defined by the cuts listed in table 8.1 for ARIADNE 4.08, LEPTO 6.5 with and without SCI.*

8.4.3 General Systematic Checks

Some systematic checks are performed in order to test the stability of the results against fluctuations arising from an inaccurate description of the data from the model. In order to evaluate the amplitude of these fluctuations, the Monte Carlo simulation with which the detector corrections have been performed, ARIADNE, is used. The systematic checks are then added in quadrature as they are expected to be independent from each other (except for the energy scale determination, see below). The value of the systematic error for each cut (if this error is significant) is shown in figure 8.17 and 8.18.

- A first category of checks deals with the "cleaning cuts" described in section 5.3 (i.e. the event selection cuts which are corrected for at hadron level), in order to test their influence on the hadron level cross sections. As these cuts are corrected for, their value should not matter. But because of the small deviations of the model from the data, the change in the correction factors does not always compensate the change in the data cross section. The list of these checks is given in table 8.4. All these changes prove to have a very insignificant effect and are therefore not displayed in the summary figure 8.17 (except for the $E - P_Z$ cut which also tests the photoproduction background).

- The reconstruction of the vertex by the CTD is checked by moving the Z position of the reconstructed vertex by a value equal to the resolution (in the beam direction) of the CTD ($\pm 0.4\ cm$). The position of the cells is then a little bit shifted, which leads to different value on the position of the positron and the jets. The effect of this change is shown in figure 8.17.

- The uncertainty on the energy scale of the calorimeter was taken into account by adding $\pm 5\%$ of energy of the jets; this corresponds in average to the uncertainty on the total energy scale in the ZEUS calorimeter (see section 7.3.1). This yields a systematic error of about 1 to 15 %. As this error is expected to be correlated with the others, in particular, with the checks on the jet variables, it is not added in quadrature as the other but represented as a band around the jet cross sections.

- The uncertainty on the positron energy scale is checked by increasing and decreasing the positron energy by $\pm 1\%$. This is negligible in the lowest x bins but

yields an error of 2 to 10% in the highest ones.

- To test the jet reconstruction, all the jet variables used in the determination of the final cross section are changed at detector level (in the data and in the Monte Carlo simulation) by one standard deviation of the mean. The summary of these changes is shown in table 8.5 and the result of the changes are displayed in figure 8.18.

- A special treatment on radiative corrections is also performed. It will be detailed in the next subsection.

- Finally, to test the dependence on the model itself, the correction to hadron level is performed with the LEPTO model instead of ARIADNE. But we face a problem here: LEPTO does not describe the behavior of all the jet variables, in particular E_T^2/Q^2. Using this model to correct the data would result on an overestimate of the systematic error. In order to turn around the difficulty, LEPTO has been reweighted according to the E_T^2/Q^2 distribution obtained in ARIADNE. The weight (which corresponds to the ratio of cross sections normalized to the shape, between LEPTO and ARIADNE) has been applied at both hadron and detector levels in LEPTO. If no hadron level jet was found in the event (within the set of cuts presented in table 8.1), the hadron level jet which was the closest to the detector level jet in the (η, ϕ) plane was used to get the weight. This yields a systematic error much smaller than the one which is obtained without reweighting.

Nominal cuts	Changed cuts
$35 < E - P_Z < 65GeV$	$40 < E - P_Z < 65GeV$
$\|VTX_Z\| < 50$ cm	$\|VTX_Z\| < 60(40)$ cm
$y_{el} < 0.8$	$y_{el} < 0.95$
Box cut (13cm,8cm)	Box cut (14cm,14cm)

Table 8.4: *List of systematic checks on the cleaning cuts. On the last row, the box cut corresponds to the cut performed on the positron position as measured by the calorimeter or the SRTD.*

Nominal cuts	Changed cuts
$E_T^{Det} > 5$ GeV	$E_T^{Det} > 5.5/4.5$ GeV
$X_{JET}^{Det} > 0.036$	$X_{JET}^{Det} > 0.042/0.03$
$\eta^{Det} < 2.6$	$\eta^{Det} < 2.7/2.5$
$E_T^2/Q^{2Det} < 2$	$E_T^2/Q^{2Det} < 2.4/1.6$
$E_T^2/Q^{2Det} > 0.5$	$E_T^2/Q^{2Det} > 0.6/0.4$

Table 8.5: *List of systematic checks on the jet variables.*

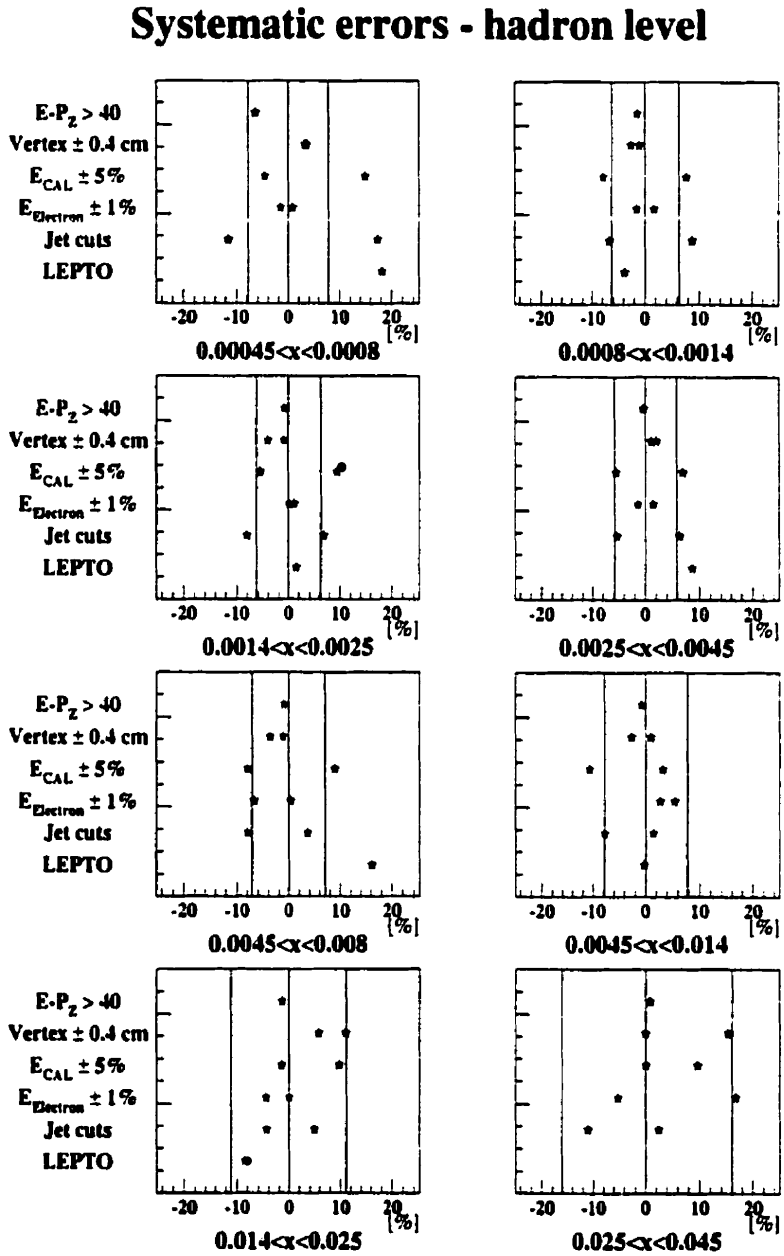

Figure 8.17: *Relative systematic error for each of the general checks done in each of the x bins. In the last x bin, the error on the model dependence is around -30% and does not appear on this plot.*

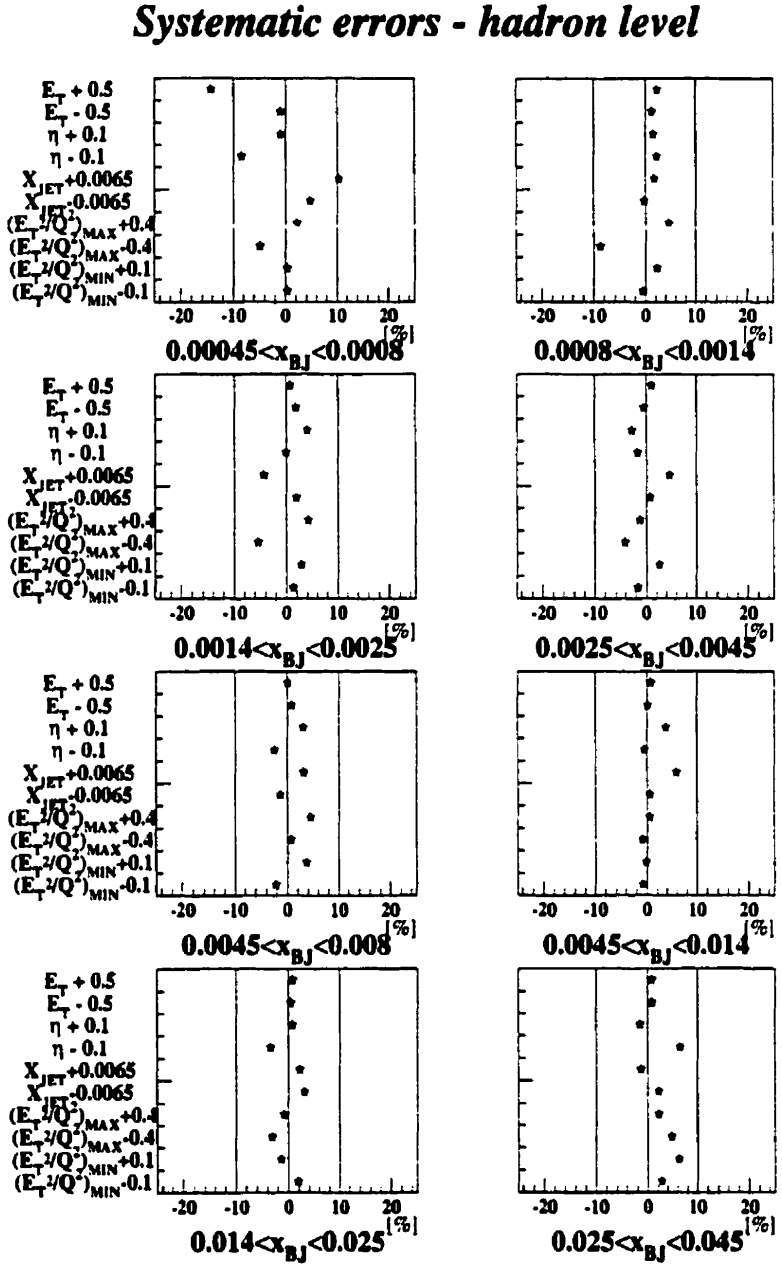

Figure 8.18: *Relative systematic error for each of the checks on the jets selection cuts done in each of the x bins.*

8.4.4 Radiative Corrections

Most theoretical calculations and Monte Carlo predictions in Deep Inelastic Scattering are performed at the Born level of QED. This means that, in these calculations, the scattering lepton does not emit a Bremsstrahlung photon before (initial state) or after (final state) the collision with the proton. This is justified by the fact that QED radiations are suppressed by an order α_{em} and are not expected to contribute much in the total cross section.

In practice however, the Bremsstrahlung radiations can have sizable effects on both the value of the kinematic parameters and the total cross section. Although initial and final state radiations are undistinguishable from a theoretical point of view, their effects are different on the correction factors: the photon emitted after the collision between the lepton and the proton is often emitted collinear to the scattered lepton and therefore the measured leptonic energy in the calorimeter includes both the energy of the scattered lepton and the energy of the radiated photon. On the other hand the initial state radiation have a more dramatic effects on the kinematical variables than the final state Bremsstrahlung.

The implementation of QED effects in a Monte Carlo package is used to correct data and Monte Carlo for the distortions of the kinematic variables Q^2, x and y created by a radiative photon. These radiations also affect the absolute value of the cross section of the events in the generator. To estimate the size of such a variation at generator level, two samples were generated, one with QED radiations and one without. This check has been performed with the HERACLES package [128], on events generated with ARIADNE. HERACLES generates QED radiations to the order α_{em}. No NLO QED radiations are implemented, but these ones are known to have a very small effect. The relative difference between the two sample cross sections is shown in figure 8.19. All the cuts listed in table 8.1 have been applied. The two cross sections, with and without radiative corrections, are very similar. There is no systematic shift between these cross sections and the relative difference is $\pm 10\%$ or less. Therefore, the QED effects on the forward jet cross section can be neglected.

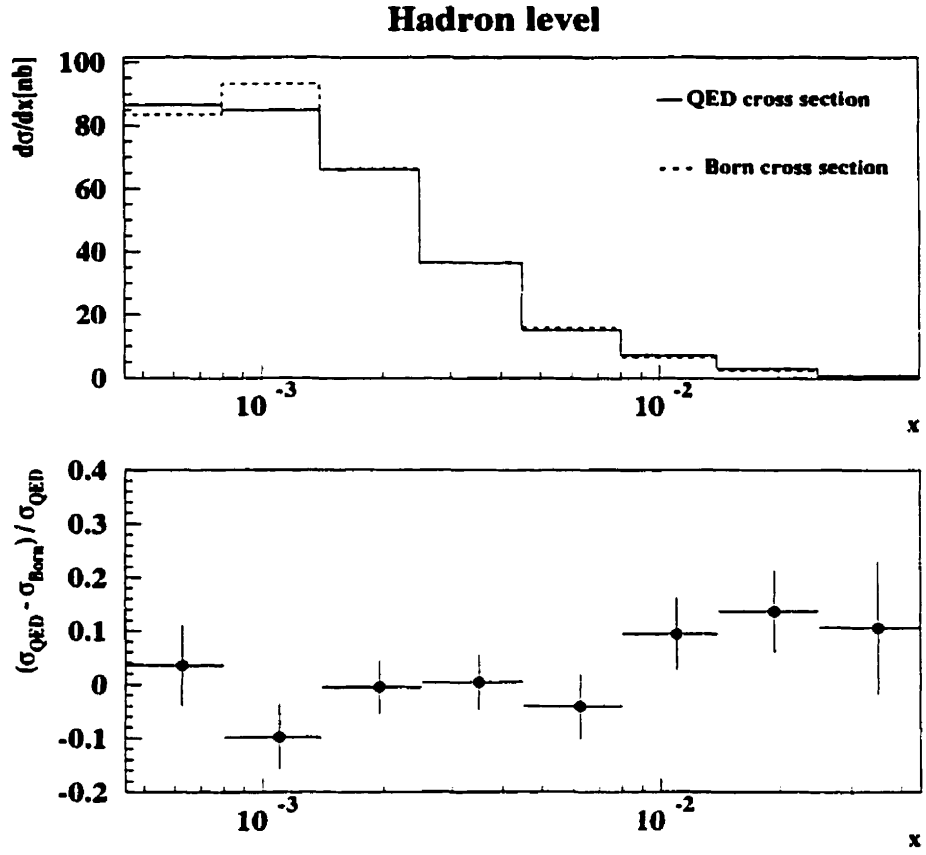

Figure 8.19: *Cross section comparison between the forward jets generated with radiative corrections (σ_{QED}) and without(σ_{Born}). The upper plot pictures the differential cross section, while the bottom plot is the relative difference between the cross sections.*

8.4.5 Dependence of the Results on the Jet Algorithm Used

Although the value of the jet cross sections at detector and hadron levels depend on a large number of factors, an inclusive study of all the parameters involved in the determination of the forward jet cross section would be a long and tedious task beyond the scope of the present study. The aim of this section is just to present two of the factors which may change, if not the physics message of the analysis, at least the absolute meaning that one could give to it.

One of the crucial problems of this analysis is that, in order to evaluate a parton level effect, we make use of an algorithm which is supposed to improve the agreement

between the parton and the hadron levels. As we saw in section 8.4, this was not a trivial, one-to-one correspondence, so that different algorithms may give different answers. In figure 8.20, the data cross sections (obtained with the algorithm PUCELL) are compared to the predictions af ARIADNE performed with three different jet algorithm: PUCELL (which was used all over the analysis) and two algorithms frequently used at ZEUS: EUCELL and PXCONE. The difference between the algorithms lies mainly in the way they treat the energy shared between two overlapping jets. For instance, the minimum fraction of energy shared required for merging two jets is different from one algorithm to another. In figure 8.20 a), all the algorithms are used with the same cone radius 1. There is about a 30% spread on the value of the cross section. When the jets are generated with different radii, all the results can match as in figure 8.20 b), leading to the following correspondence: PUCELL (R=1) = PXCONE (R=1.2) = EUCELL (R=0.9). It must be added here that the radius of the jets were properly tuned to some well known distributions, but as we saw in chapter 6, the internal structure of the jets is different in the forward region. Finally, the parton level calculations (either BFKL or CCFM) are also subject to to theoretical uncertainties which can lead to comparable discrepancies.

Similarly, varying the cone radius of the PUCELL algorithm by $\pm$ 20% can lead to a 30% spread of the cross section, as can be seen in figure 8.21. There is no definitive "best choice" for the determination of the cone radius or the algorithm used. The most natural criteria is the correspondence between hadron and parton level jets.

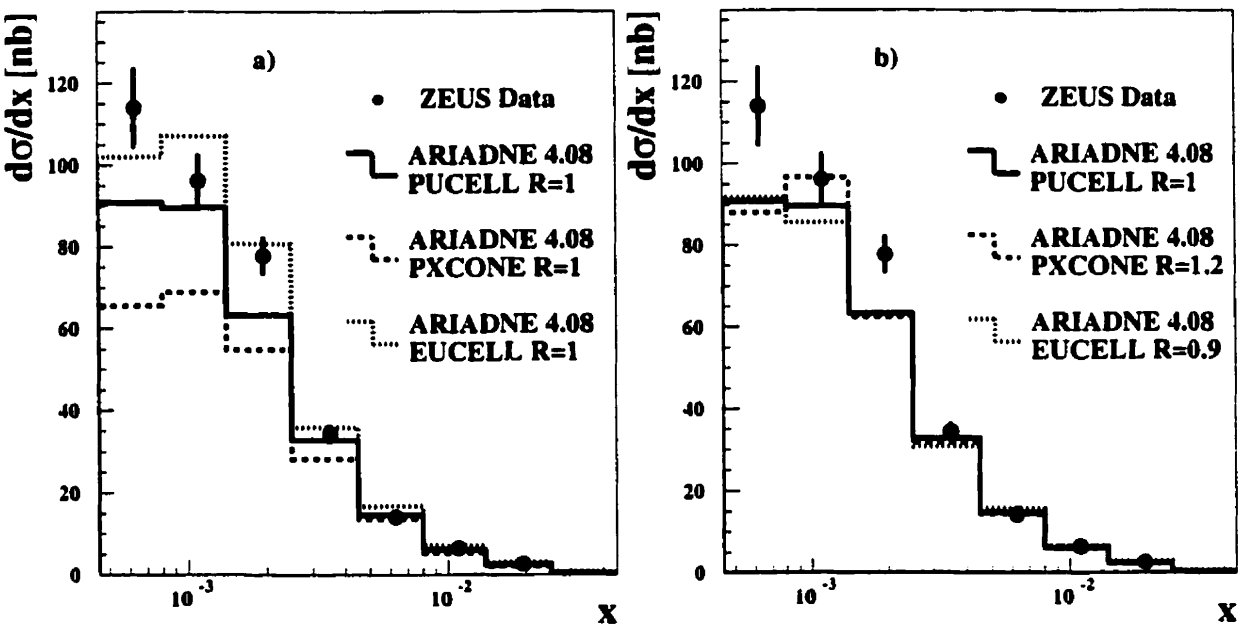

Figure 8.20: *Hadron level jet cross sections in the data (obtained with the PUCELL algorithm) compared to the predictions of ARIADNE 4.08 obtained with three different jet algorithms: PUCELL, PXCONE and EUCELL. a) All the algorithms use the same radius. b) Different radii are used.*

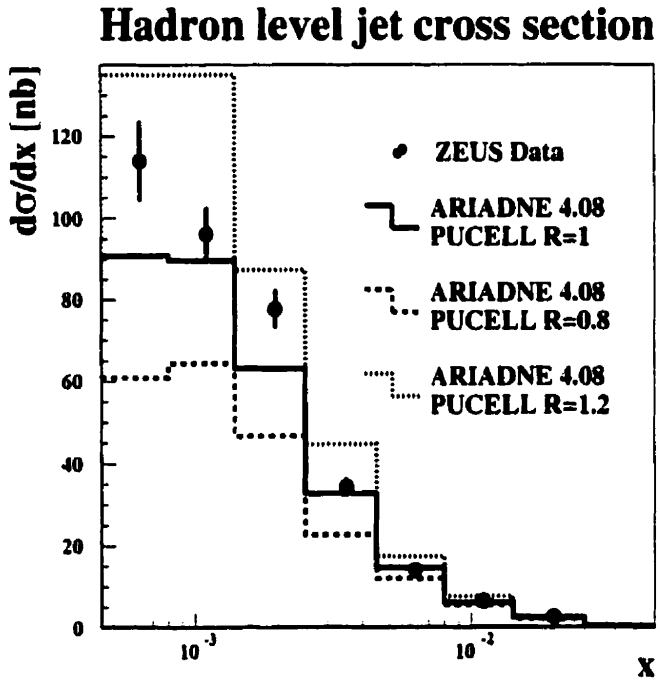

Figure 8.21: *Hadron level jet cross sections in the data (obtained with the PUCELL algorithm) compared with the predictions by ARIADNE 4.08 obtained with PUCELL and three cone radii: 0.8, 1 and 1.2.*

8.4.6 An Alternative Explanation: the Equivalent Photon Approximation (EPA)

Recently, another attempt to explain the excess in the forward jet cross section with respect to the standard DGLAP model predictions (and the NLO calculations) has been presented [129]. The evolution mechanism that it involves is based upon the equivalent photon approximation (EPA): if the scale which is probed by the hadron jet is larger than the four-momentum squared Q^2 of the photon, one might then be able to resolve the structure of the photon, even at moderate Q^2 ($Q^2 > 10$ GeV^2). In this case, Q^2 ceases to be the natural scale of the process and a DGLAP-like process can take place, picturing an evolution from the hard jet (which sets the scale) to the photon side. This backward evolution is pictured in figure 8.22.

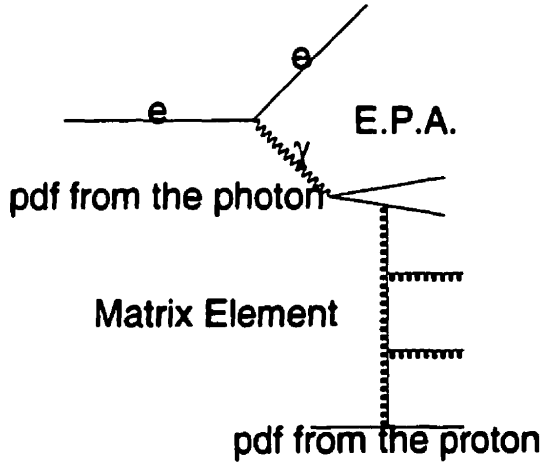

Figure 8.22: *Schematic description of the equivalent photon approximation. The factorization scale on the proton side being larger than Q^2, one can resolve the structure of the photon. Here PDF means Parton Distribution Function.*

In order to generate this kind of events, a new Monte Carlo simulation has been presented, RAPGAP [56], which uses the "Schüler-Sjöstrand", SaSgam, parameterization [130] for the structure function of the photon at high Q^2. Within this framework, two samples must be added (as in photoproduction): a direct part, which pictures the standard DGLAP evolution, as in LEPTO and HERWIG, and the resolved part, which implements the EPA. In figure 8.23, the comparison between the data cross sections corrected at hadron level and the predictions obtained by RAPGAP and

LEPTO is presented. The direct process is consistent with LEPTO's prediction, while the sum of direct and resolved processes describes well the data (at least, as good as ARIADNE). There is however some nuance to bring to this result: in order to have any contribution from the photon structure, the factorization scale for the photon side, μ^2, has to be much larger than the Q^2 of the photon (in figure 8.23, the scale is taken to be $p_t^2+Q^2$, p_t^2 being the sum squared of the transverse momenta of the leading order partons). Similarly, the "amount of resolved photon" which contributes to the forward jet cross section changes when the scale is changed: in figure 8.24, the same process is considered with three different scales: $Q^2/2 + p_t^2$, $Q^2 + p_t^2$ and $4 \cdot Q^2 + p_t^2$. Whereas the direct component of the process does not change much, the

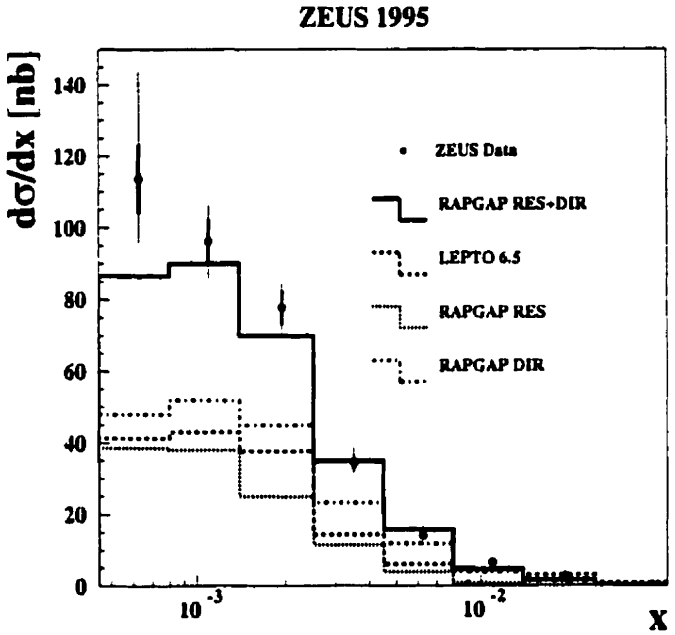

Figure 8.23: *Differential hadron level jet cross sections in the data (obtained with the PUCELL algorithm) compared with the predictions by RAPGAP 2.06 implementing direct and resolved processes.*

resolved cross section depends considerably on the scale. Indeed there is no resolved part when the scale is set to Q^2, while the resolved component dominates over the direct one when the scale is set to $4 \cdot Q^2 + p_t^2$. The phenomenological input to the model is therefore very important and this makes the predictive power uncertain. However, a resolved component in the photon is certainly a possible explanation for the forward jet excess and NLO corrections for the resolved part of the process are expected and could improve the accuracy of the predictions.

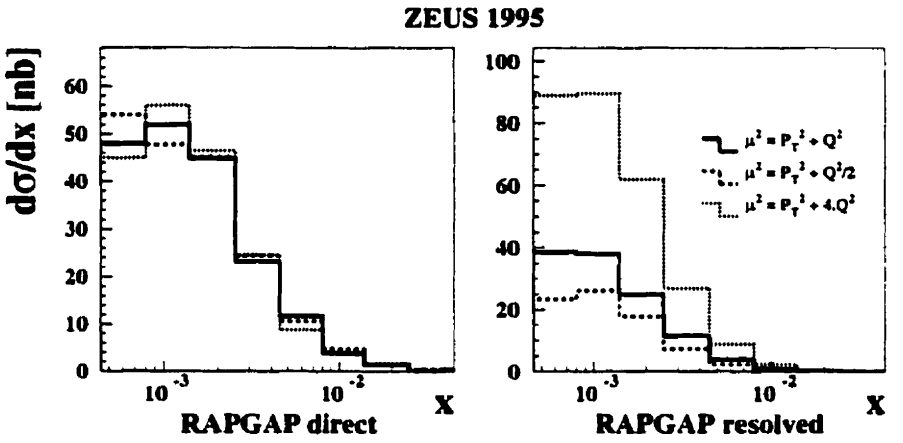

Figure 8.24: *Hadron level jet cross sections obtained with RAPGAP 2.06 (resolved and direct processes) for three different choices of scale.*

8.5 Extending the Kinematic Region of Interest

In section 8.1 of this chapter, the E_T^2/Q^2 distribution of the forward jet was shown to be a good estimator for the various parton evolution mechanisms which can take place in DIS. Three kinematic regimes have been determined, when comparing the various data points to the models at detector level: the region where the DGLAP picture is valid, $Q^2 > E_T^2$, the region where the contributions from higher order processes and BFKL-type processes dominate, $Q^2 \sim E_T^2$ and the region where the hard scale is set by the transverse energy of the jets $E_T^2 > Q^2$.

For a better understanding of these various regimes, the E_T^2/Q^2 distribution of figure 8.2 has been corrected to hadron level. As the resolution of the variable E_T^2/Q^2 is quite large in the forward jet analysis, the bin size has been doubled. The correction factors are shown in figure 8.25. The average efficiencies and purities are 40%. The maximum values of the purities and efficiencies lie in the high E_T limit, which is expected as the high E_T jets are easy to measure and provide a clean sample. The minimum value of the purities and efficiencies is found in the region $E_T^2 \sim Q^2$, which corresponds to the phase space under investigation in the rest of this chapter. This is due to the fact that both E_T and Q^2 have relatively low values. In this plot are also shown the values for efficiencies and purities where the events are not required to be generated and reconstructed in the same bin (see section 7.3.3). This is equivalent to say that the the migration from bin to bin of E_T^2/Q^2 between generated and reconstructed level is not included in the definition of purity and efficiency. From the difference between the two values (with and without the requirement of being generated and reconstructed in the same bin), one can infer the average amount of bin-to-bin migration of the sample, which is typically between 10% and 20%.

Once the data have been corrected, they are compared to some Monte Carlo models in figure 8.26. The same simulations as for the x distribution are shown: ARIADNE 4.08, LEPTO 6.5, HERWIG 5.9 and LDC 1.0 (the systematics checks are similar to those which were performed for the previous distribution). The same trend as at the detector level is observed: the same three kinematic regimes shows up in the corrected E_T^2/Q^2 distribution. The DGLAP-based models, as well as ARIADNE and LDC are valid for $Q^2 > E_T^2$, while only ARIADNE describes the data at $E_T^2 \sim Q^2$. In the third region, the LDC model describes the data best, although it exhibits a smaller cross section over the full range. LDC is the only model which treats forward evolution (from the on-shell parton at the proton side to the virtual photon) and

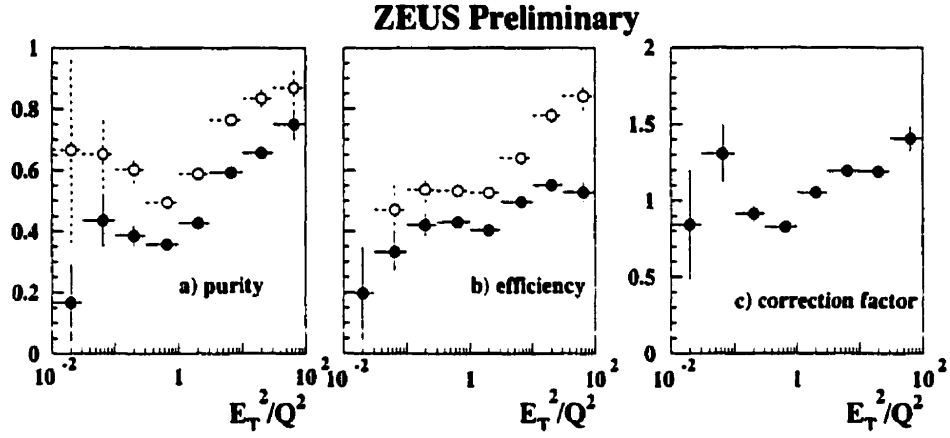

Figure 8.25: *Efficiency, purity and correction factors for the E_T^2/Q^2 distribution of the forward jets. All the cuts listed in table 8.1 , but the one on E_T^2/Q^2 have been applied. The black dots correspond to the real purity and efficiency. The open dots correspond to the values of purities and efficiencies when bin-to-bin migration is allowed.*

backward evolution (from the hard, virtual jet to the "softer" photon) on a similar footing, as the evolution properties are symmetric with respect to the photon-proton axis. However, since LDC does not reproduce the data cross section in a region where it should ($E_T^2/Q^2 \sim 1$), to conclude on an eventual backward evolution towards the photon side is still too premature. Similarly, the comparison with the RAPGAP model (see section 8.4.6) is made in figure 8.27. There also, the additional resolved contribution from the photon improves the description of the data. As in the case of LDC, the data exhibits smaller cross sections than the model. Once again, the predicted cross section depends on the scale selected and no strong conclusion can be drawn.

8.6 The k_T Analysis

This analysis has also been performed with the k_T algorithm (see [131]), but the experimental results have proven to be disappointing. In the analysis, the resolution parameter, y_{cut} is set equal to 0.5 (see section 6.1.2). The absolute scale used during the clustering process is $Q^2{}_{DA}$ at the reconstructed level and $Q^2{}_{True}$ at the generator level. The jet clustering is done in the Breit frame (the four momenta of the particles are boosted to the Breit frame where the clustering is done, the jets are then boosted back to the laboratory frame where the cuts are applied and the cross sections are

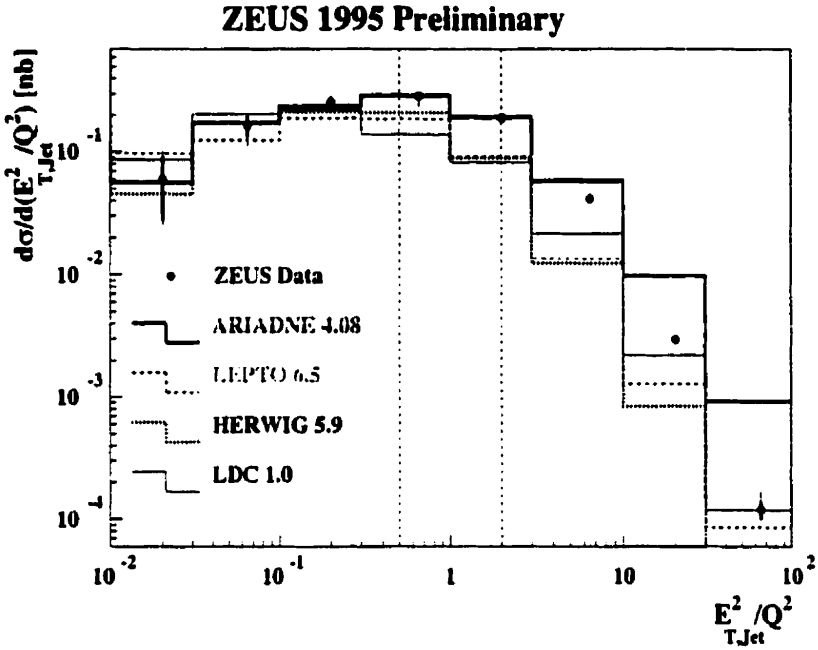

Figure 8.26: *Differential hadron level forward cross section as a function of E_T^2/Q^2. All the cuts listed in table 8.1 , but the one on E_T^2/Q^2 have been applied. The two vertical lines show the limits of each kinematical regime (see text). The statistical errors are shown as thick error bars, and statistical and systematic errors added in quadrature as thin error bars. The correlated errors due to the uncertainty of the jet energy scale are given as the shaded band.*

evaluated). The exact cuts listed on table 8.1 are applied to the jet.

The comparison between the jet cross sections at various level of the reconstruction process (detector level, hadron level and parton level) is shown in figure 8.28. The difference between hadron and parton level cross sections indicates that the objects found at hadron level are mostly not coming from hard processes (as in the case of LEPTO, with the cone algorithm).

The cross section at detector level is much larger than at hadron level, especially in the smallest x bins, indicating that most of the objects found after clustering the cells of the calorimeter not only do not correspond to any hard physics, but are not even related to any generator level quantity. This translates into the values of purities and efficiencies of the jets, shown in figure 8.29. At low x, they are typically of the order of

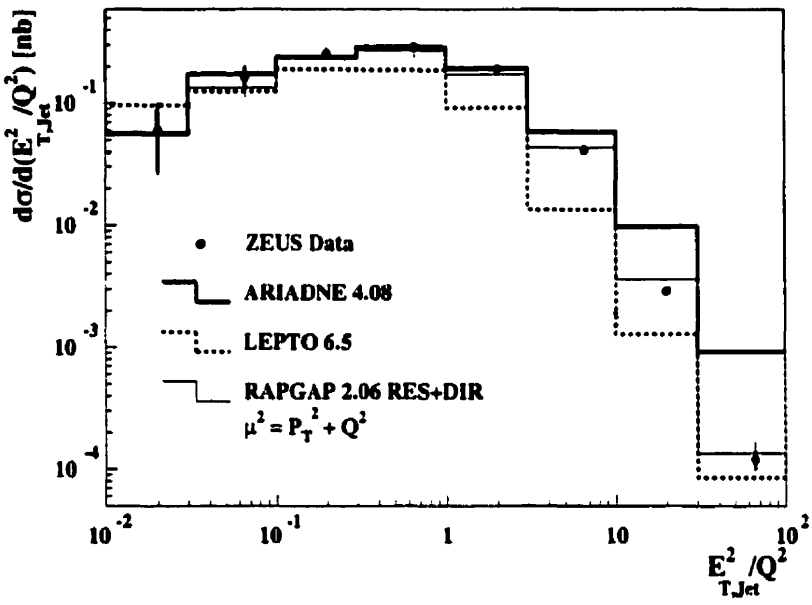

Figure 8.27: *Differential hadron level forward cross section as a function of E_T^2/Q^2, compared to the RAPGAP predictions (scale $\mu^2 = P_T^2 + Q^2$). The statistical errors are shown as thick error bars, and statistical and systematic errors added in quadrature as thin error bars. The errors due to the uncertainty of the jet energy scale are given as the shaded band.*

10% or less and the correction factor is about 2. This time, not only the resolution of each variable used in the cuts enters into account in the determination of the purities and efficiencies, but the nature of the jet found in the calorimeter is not related to the hadron level. This can be shown when comparing quantities found at hadron and detector level. In figure 8.30, the correlation between the pseudorapidity of the "true" jet, found at generator level is plotted versus the pseudorapidity of the detected jet. For $\eta > 2.2$, the k_T algorithm finds after the reconstruction a large amount of jets which are not correlated with the jets found at the generator level. The jets at hadron level are in average more forward than the detector jets and therefore do not survive the maximum η cut of 2.6. These hadron level jets are actually receiving contributions from the forward region, possibly from the proton remnant, which shift the overall pseudorapidity to the forward region. Other evidences exist which make us think that the k_T algorithm, in this implementation, can not make a clear distinction between the forward jet and the remnant in the forward region of rapidity (cf. [131]). One of the explanations which can be advanced is that the Q^2 scale used by the algorithm

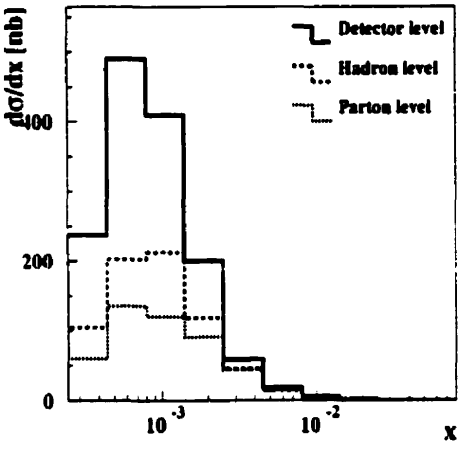

Figure 8.28: *Differential parton, hadron and detector level forward jet cross sections obtained with the ARIADNE 4.08 simulations, using the k_T algorithm.*

to perform the clustering in the Breit frame is not suited in this kinematic region. As a matter of fact, in the phase space region probed, one does not expect that Q^2 sets the scale of the hard process (in fact this should not be the case by construction, as the analysis is performed in the region where the transverse energy squared of the jet is on the order of Q^2). The behavior of the correction factors suggests a similar explanation, as they come closer to one at high x, where Q^2 is larger and might be the correct scale of the process.

The determination of the real scale of the process and its implementation into the k_T algorithm, to find forward jets in DIS, is therefore not straightforward and then no strong conclusion can be drawn from this analysis.

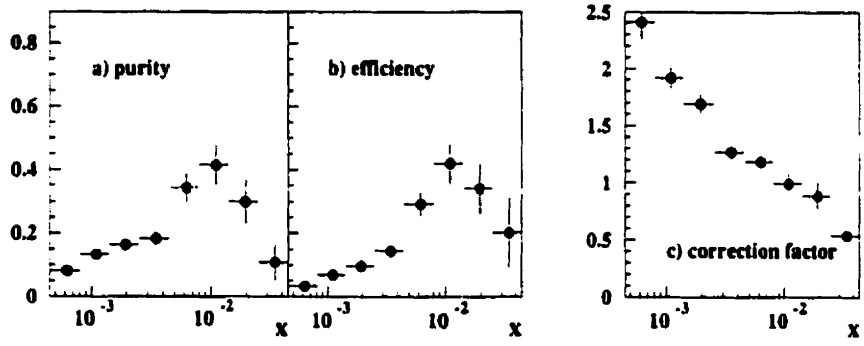

Figure 8.29: *Purity, efficiency and correction factors for the forward jets obtained with the k_T algorithm.*

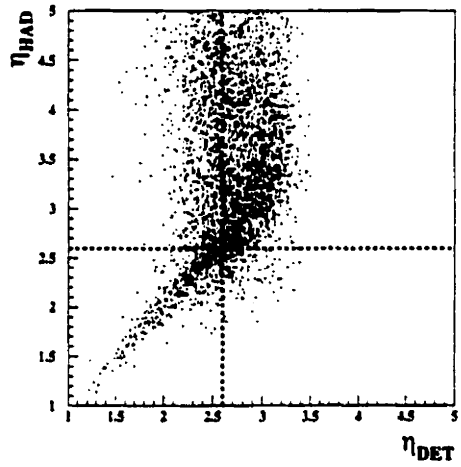

Figure 8.30: *Correlation between the pseudorapidity of the hadron and the detector level jets with the k_T algorithm. All the cuts listed in table 8.1, but the one on η_{jet} are applied. The dashed lines show the value of the pseudorapidity cuts.*

8.7 Summary

The forward jet cross section has been measured in the 1995 data with the cone algorithm, using the ZEUS detector. The internal structure of the jets has been studied and the data cross sections have been corrected for detector effects using standard Monte Carlo simulations. The comparison between the corrected data and the Monte Carlo simulation for $E_T^2/Q^2 \sim 1$ shows that the DGLAP-based MEPS models, LEPTO and HERWIG, fail to describe the absolute jet cross section in the data, while the ARIADNE model, which does not implement the same type of parton evolution, strongly ordered in E_T, describes the data accurately. This strong discrepancy is a hint that in the low x limit, there are hard processes not implemented in the MEPS models. The nature of these hard processes can only be determined when the hadronization effects are subtracted from the measurements. The LDC model, which implements literally the CCFM equation and should therefore match the BFKL predictions at low x, is unable to reproduce the experimental cross sections.

At high x, the LEPTO and ARIADNE model converge and reproduce the experimental cross sections, but LDC and HERWIG still exhibit smaller cross sections. One explanation could be the requirement of strong angular ordering which is implemented in both latter models.

The NLO calculation, as implemented in MEPJET, are in good agreement with the predictions of the DGLAP-based processes, LEPTO and HERWIG, and show much smaller cross sections than ARIADNE. For the comparison to the data, a hadronization correction has been applied, but it has been found very model dependent, as the correction factors are very different between LEPTO and ARIADNE. The existence of large correction factors in LEPTO at low x is again an evidence that the hard processes which are probed are not implemented in this model. The Soft Colour Interaction is an example of a non-perturbative process which contributes significantly to the forward jet rate in LEPTO.

An attempt to use a contribution of the resolved photon into the determination of the cross section successfully describes the value of the forward jet cross section but is bound to a large scale dependence, which make the conclusions uncertain.

The study of the forward jet cross section in an extended range of E_T^2/Q^2 revealed three kinematic regions, depending on the scale of the process: I) $E_T^2/Q^2 \ll 1$: the scale is set by Q^2 and all the predictions converge to the experimental distributions, II) $E_T^2/Q^2 \sim 1$: the DGLAP approximation is not valid any more and ARIADNE is the only model describing the data, III) $E_T^2/Q^2 \gg 1$, the hard scale is set by the transverse energy of the jet, as in photoproduction, and no model reproduces the data, except the RAPGAP model, very dependent on the factorization scale.

The measurement of the forward jet cross section is bound by several uncertainties, such as the dependence of the result on phenomenological parameters (e.g. the radius of the cone) or the uncertainty on the corrections to parton level which currently makes the comparison with theoretical calculations performed at the parton level valid only within to about $\pm$ 30%. The ideal method of comparison would be to use the exact same algorithm in the theoretical calculations as in the data, but the BFKL calculations do not implement this scheme yet. This present level of accuracy is also expected in the calculations, so that the results are still meaningful in terms of hard physics processes.

Chapter 9

Conclusions

Low-x physics in Deep Inelastic Scattering has long been known as one of the best place to study the dynamic of parton evolution. In this region of the phase space, the parton emissions, which can be treated perturbatively thanks to the property of asymptotic freedom, are extremely sensitive to the approximations which are performed in the various parton evolution schemes. The standard DGLAP evolution, which resums log Q^2 terms only and then predicts parton emissions strongly ordered in transverse momentum, is expected to fail in this region, while the new pictures, based either on the summation of logarithms of $\frac{1}{x}$ (BFKL scheme) or on the angular ordering of the parton emissions (CCFM scheme) are both expected to yield good results. The theoretical studies which have been performed on the comparison between the various schemes predict that at low values of x, the difference is dramatic.

However, this phase space region is also extremely sensitive to the long-range physics, dominated by the property of confinement and non-calculable through the standard methods, so that the extraction of the proper QCD effect from the huge non-perturbative background is the main challenge of all analyses whose goal is to study the parton dynamic at low x. The choice of the variable used in order to probe the parton dynamic is then of a crucial importance and the effects of non-perturbative physics on this variable must be checked carefully. As a matter of fact, the influence of long-range physics on inclusive quantities like the proton structure function and the transverse energy flow can not be disentangled from the perturbative evolution. Therefore no conclusion on which type of parton evolution takes place and which set of approximations is valid can be drawn from such inclusive variables.

Jet variables are sensitive to QCD effects and less influenced by non-perturbative physics. However the experimental definition of a jet is dependent on a certain algorithm. Up to now, the jets were used at HERA only in the high Q^2 range and with fairly large transverse momenta, so that the systematics of the jet algorithm were reduced by the fact that the jets probed were very collimated. At low x, these effects are not so well known. Although a lower cut on the transverse momentum is supposed to reduce the effects of soft physics on the final cross sections or angular distributions, this problem might bounce back through the dependence on the algorithm. However, jets are closely related to the details of the parton evolutions as their four momenta are supposed to reproduce those of the hard partons through the property of local parton-hadron duality and are therefore privileged objects to probe them.

In this work, the study of the parton dynamics at low x has been performed with two methods using the jets observables. The first method is an attempt to probe indirectly the underlying parton evolution by measuring their effects on the angular correlation of the two leading order jets. As the analysis focuses on the tail of the azimuthal correlation, in the region where both jets are very close to each other in the transverse plane, the uncertainties due to the cone jet algorithm must however be removed by a cut which reduces the sensitivity to the various modes of parton evolutions. When comparing the measured distributions to the DGLAP-based Monte Carlo simulations, LEPTO or HERWIG, no large discrepancy is found, and the ARIADNE model which implements some of the features of a BFKL-like evolution, is also consistent with the former Monte Carlo models.

The second measurement aims at looking directly at the partonic emissions by focusing on a single parton, at a higher stage of the evolution (and for this reason, named forward jet). The emission rate of this type of parton is largely suppressed in a DGLAP-based picture, so that the value of the forward jet cross section can give an indication of the validity range of this model. Indeed, the measured hadron level jet cross sections exhibit a clear excess with respect to LEPTO and HERWIG in the low-x limit, even with all the non-perturbative effects turned on. This cross section agrees moreover with the predictions of the ARIADNE simulation, which is attributed to the effects of a parton evolution, non-ordered in transverse momenta. The disagreement with a first release of a CCFM-based simulation like the LDC model is still a puzzle, as this picture implements the kind of physics which is expected in this kinematic

region. The answer may come from the high x limit, where both LDC and HERWIG converge in cross sections and lie much below the data cross sections. One hypothesis is that the coherence effects, strictly implemented in each one of these models and handled by the angular ordering of the parton emissions, are restricting the available phase space for emission, leading to too small cross sections.

An alternative explanation of this excess of forward jets with respect to the DGLAP predictions was provided by the RAPGAP model, which predicts that, when probing high E_T jets, one can resolve the structure of the photon. The forward jets would then be the consequence of the existence of a "photon remnant" which has yet to be studied in detail in DIS. The attractive feature of this model is that it reproduces the data cross sections over the full E_T^2/Q^2 range, while all the other models fail for $E_T^2/Q^2 > 2$. This is indeed the region where the effects of a resolved photon are expected to appear as E_T is the hardest scale of the process which makes this case similar to photoproduction. On the other hand, for $E_T^2/Q^2 < 0.5$, Q^2 is the hard scale and all the models converge.

Ideally, the study of the parton dynamic should involve a parton level comparison between the data and the theoretical predictions (for all the models). In practice, this comparison is bound to very large uncertainties due to the model dependence of the hadronization correction. This is once again the consequence of the influence of the large non-perturbative effects at low x, which affect the jets properties. A better understanding of these effects is mandatory for a more conclusive study of this kinematic region.

In summary, as expected, the jets are less sensitive to hadronization effects than the other observables and more sensitive to the perturbative evolution, so that the forward jet analysis provides us with the evidence that HERA reaches the region where the standard DGLAP models are no longer valid. The next generation of event generators, with a more accurate implementation of non-perturbative effects such as soft colour interaction and colour coherence should improve the interpretation of these results in terms of hard physics and reduce the uncertainty due to the hadronization phase.

Appendix A

Reconstruction of the Kinematic Variables

In chapter 2, the main variables which are determining the kinematics of the event were presented: the virtuality of the scattered photon Q^2, the scaling variables x (which corresponds to the momentum fraction of the struck quark in the QPM) and y, the energy transfer from the leptonic to the hadronic system (in the rest frame of the target hadron). These variables are defined by the set of equations:

$$Q^2 = (k - k')^2 = (-p + p')^2 = -q^2, \tag{A.1}$$

$$x = \frac{Q^2}{2p \cdot q}, \tag{A.2}$$

$$y = \frac{q \cdot p}{k \cdot p}. \tag{A.3}$$

Here, the variables k and k' correspond respectively to the positron momenta before and after the collision p and p' are likewise the proton momenta before and after the collision while q is the momentum of the scattered photon in the laboratory frame. These variables are displayed in figure A.1, together with θ_e, the angle of the scattered positron and γ_h, the angle of the 'current' jet, namely the struck parton in the QPM. To determine these variables from a HERA event, the three most common methods are: the electron method, which uses only the information given by the scattered positron, the Jacquet-Blondel method, which uses only the hadronic energy information of the event and the double angle method, which reduces the energy fluctuations by considering only the angles of the two systems.

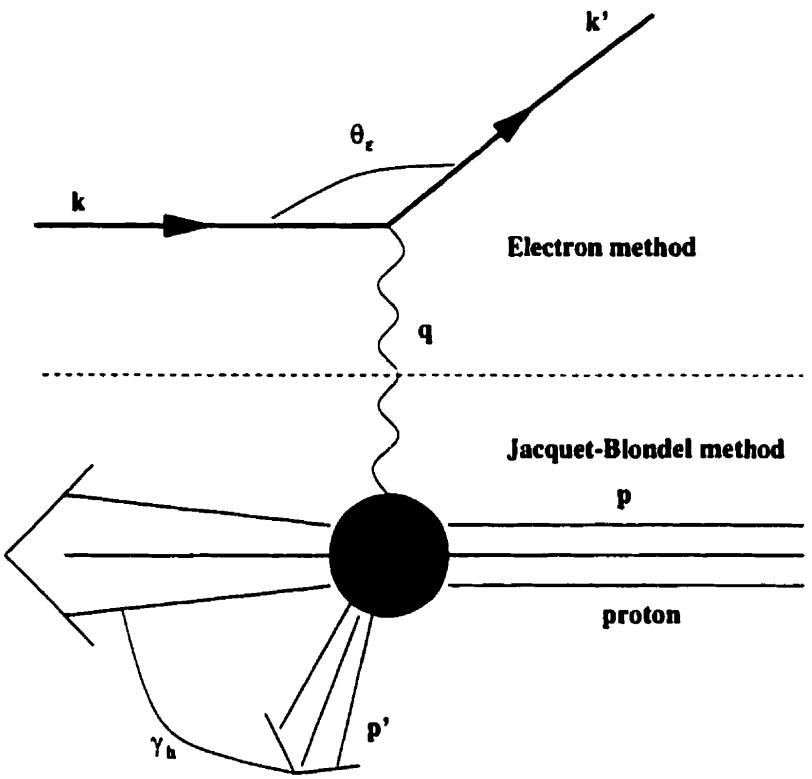

Figure A.1: *Deep inelastic scattering between the virtual photon and the proton in the QPM.*

The conservation of energy and momentum, which enables the determination of the kinematic variables reads:

$$k\begin{pmatrix} E \\ 0 \\ 0 \\ -E \end{pmatrix} = k'\begin{pmatrix} E' \\ E'sin\theta_e cos\phi_e \\ E'sin\theta_e sin\phi_e \\ -E'cos\theta_e \end{pmatrix} + q\begin{pmatrix} E_q \\ q_x \\ q_y \\ q_z \end{pmatrix} \tag{A.4}$$

$$p\begin{pmatrix} E_p \\ 0 \\ 0 \\ E_p \end{pmatrix} = p'\begin{pmatrix} E_h \\ E_h sin\gamma_h cos\phi_h \\ E_h sin\gamma_h sin\phi_h \\ E_h cos\gamma_h \end{pmatrix} - q\begin{pmatrix} E_q \\ q_x \\ q_y \\ q_z \end{pmatrix} \tag{A.5}$$

A.1 The Electron Method

This is the easiest method to obtain x, y and Q^2. By simply identifying the positron variables with the variables presented in equations A.1, A.2 and A.3, we obtain the

following set of equations:

$$Q^2_{el} = 2EE'(1 + cos\theta_e) \tag{A.6}$$

$$y_{el} = 1 - \frac{E'}{2E}(1 - cos\theta_e) \tag{A.7}$$

$$x_{el} = \frac{E}{P}\left(\frac{E'(1 + cos\theta_e)}{2E - E'(1 - cos\theta_e)}\right) \tag{A.8}$$

Here P is the momentum of the incoming proton. The method gives the best results for $y \sim 1$, where the energy fluctuations do not affect too much the values of x and Q^2 (see [121]). Usually, a cut on y ($y > 0.01$) is needed in order to have good resolutions for the kinematic variables. Overall, the electron method gives the best resolution of x and Q^2 over all the kinematic range considered in this analysis and is therefore the one we chose to use in this thesis.

A.2 The Jacquet-Blondel Method

This method uses exclusively the information from the hadronic system to determine the kinematic variables. It is useful in Charged Current events (CC) where the scattered neutrino can not be measured. Otherwise, this method is dependent on the hadronic energy fluctuations and usually gives a poor resolution for x and Q^2, as it relies on the resolution of the hadronic energy which is in all cases poorer than the positron resolution.

The variables are determined by identifying the variables on equations A.1, A.2 and A.3 with the hadronic variables:

$$y_{JB} = \frac{E_h(1 - cos\gamma_h)}{2E} \tag{A.9}$$

$$Q^2_{JB} = \frac{(\sum_i p_{x,i})^2 + (\sum_i p_{y,i})^2}{1 - y_{JB}} \tag{A.10}$$

$$x_{JB} = \frac{Q^2_{JB}}{sy_{JB}} \tag{A.11}$$

Although the Jacquet-Blondel method is rarely used to determine x and Q^2, the variable y_{JB} is often identified with y.

A.3 The Double Angle Method

The third reconstruction method relies only on the angles of the leptonic and hadronic system after the collision. The purpose is to have a reconstruction method very little dependent of the energy fluctuations of each one of the systems, and therefore, of the calibration of the calorimeter.

In the QPM, the conservation equations listed in A.4 and A.5 can be written:

Energy conservation:	$E_p + E = E' + E_h \rightarrow xP + E = E' + E_h.$
Longitudinal momentum conservation:	$xP - E = E'cos\theta_e + E_h cos\gamma_h.$
Transverse momentum conservation:	$E'sin\theta_e = E_h sin\gamma_h.$

Therefore, the energy of the scattered positron can be expressed as a function of γ_h, θ_e and E:

$$E'_{DA} = 2E\frac{sin\gamma_h}{sin\theta_e + sin\gamma_h - sin(\theta_e + \gamma_h)} \tag{A.12}$$

Substituting the positron energy in the equations given by the electron method, one obtains the various kinematic variables as a function of θ_e and γ_h:

$$y_{DA} = \frac{sin\theta_e(1 - cos\gamma_h)}{sin\gamma_h + sin\theta_e - sin(\theta_e + \gamma_h)} \tag{A.13}$$

$$Q^2_{DA} = 4E^2\frac{sin\gamma_h(1 + cos\theta_e)}{sin\gamma_h + sin\theta_e - sin(\theta_e + \gamma_h)} \tag{A.14}$$

$$x_{DA} = \frac{E}{E_p}\frac{sin\gamma_h + sin\theta_e + sin(\theta_e + \gamma_h)}{sin\gamma_h + sin\theta_e - sin(\theta_e + \gamma_h)} \tag{A.15}$$

The main problem with this method is to determine the angle γ_h. The calculation uses the integrated energy and position of the hadronic system:

$$\gamma_h = \frac{(\sum_i P_{x,i})^2 + (\sum_i P_{y,i})^2 - (\sum_i E_i - E_{z,i})^2}{(\sum_i P_{x,i})^2 + (\sum_i P_{y,i})^2 + (\sum_i E_i - E_{z,i})^2} \tag{A.16}$$

Here P_i and E_i are the momentum and energy of the hadronic cell i (the positron cells are removed from this calculation). This can be shown to be in first order independent of the energy fluctuations. However, at low x, the determination of the position of the hadronic system becomes inaccurate and the method gives poor resolution. In figure A.2, the relative resolution of x and Q^2 is shown for the two reconstruction methods 'electron' and 'double angle' using the Monte Carlo "True" value. The resolutions are plotted in the kinematic region of the forward jet analysis, namely $Q^2 > 12.5$ GeV^2, $4.5 \cdot 10^{-4} < x < 4.5 \cdot 10^{-2}$, $E_{e'} > 10$ GeV and $y > 0.1$.

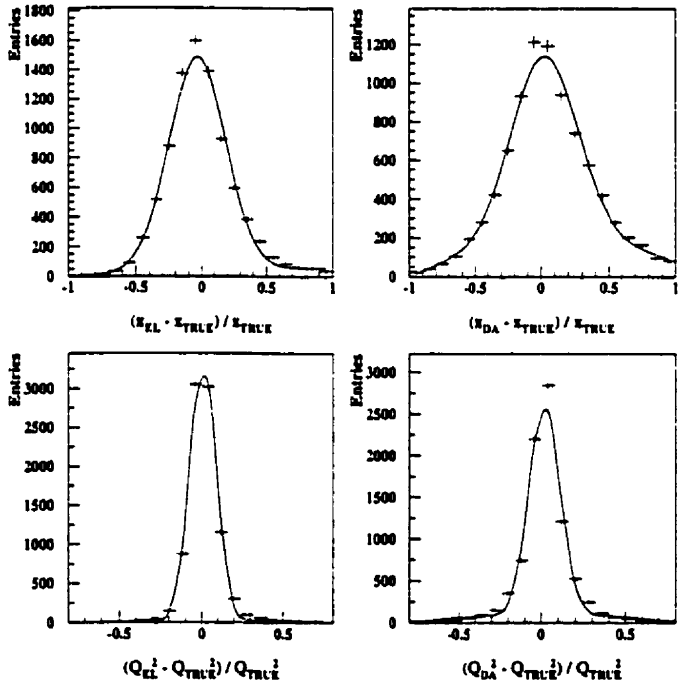

Figure A.2: *Relative resolution for the variables x and Q^2 with the two reconstruction methods 'electron' (EL) and 'double angle' (DA). The cuts applied are $Q^2 > 12.5$ GeV^2, $4.5 \cdot 10^{-4} < x < 4.5 \cdot 10^{-2}$, $E_{e'} > 10$ GeV and $y > 0.1$.*

Appendix B

The PUCELL Jet Algorithm

The jet finding performed in the two analyses presented here is based on the cone algorithm, implementing the standard Snowmass convention (see section 6.1.2). The jets are found either by applying the algorithm on calorimeter cells or on the four momenta of the particle generated in the Monte Carlo simulation.

This algorithm, PUCELL [132] is one of the most commonly used within the ZEUS collaboration. It proceeds via the following steps:

Step 1: The cells (particles) are sorted out according to their transverse energy. A pointer is set to all cells which have a transverse energy larger than a certain seed energy ($E_{T,cell} > E_T^{seed}$ (=0.5 GeV)). These are the seed cells.

Step 2: Around each seed cell, the distance to each other cell, $R = \sqrt{(\eta^{seed} - \eta^{cell})^2 + (\phi^{seed} - \phi^{cell})^2}$, is computed. if $R < 1$, then the cell is merged into an new object, a "pre-cluster", whose transverse energy, pseudorapidity and azimuthal angle are defined by the equations 6.2, 6.3, 6.4, yielding the quantities: η^{jet}, ϕ^{jet}, $E_{T,jet}$. The list of all cells contributing to the pre-cluster is kept in an array.

Step 3: The seed is then reset with $\eta^{seed} = \eta^{jet}$ and $\phi^{seed} = \phi^{jet}$.

Step 4: The cells are merged as in step 2 with the new seed. The list of cells thus obtained is compared to the one obtained in step 2. If the lists are the same, the merging procedure is stopped and the algorithm moves to step 5. If the list differs, the algorithm goes back to step 3. A maximum number of iterations (75) is given above which the jet finding is stopped.

Step 5: Once step 3 is complete, some of the cells might be common to several cluster. The energy of the overlapping region is estimated by summing up the energies of the cells in the overlap region. If the sum of the energies is larger than a fraction (usually set to 75%) of the energy of the smallest cluster, the two jets are merged. If it is smaller, each cell is assigned to its closest cluster.

Step 6: The cluster is now formed and its position and energy are now re-evaluated according to equations 6.2, 6.3 and 6.4. The jet is then formed by taking only cells within a radius of 1 from the center.

Step 7: All the remaining jets are re-ordered according to increasing transverse energy.

Appendix C

Reference Frames

In the various analyses presented in this study, different reference frames were used to describe the various variables involved. The phase space which is probed is mainly defined by Lorentz invariant parameters, Q^2 and x, so that the total inclusive DIS cross section is frame independent (this is true for F_2 as well). However, in jet studies, the phase space is of particular importance as the kinematic variables used to define jets are usually not Lorentz invariant.

The three frames appearing in this work are: the laboratory frame, the photon-proton Center of Mass frame (or Hadronic Center of Mass frame or HCM frame) and the Breit frame. They are described below and the correspondence between them is given.

C.1 The Laboratory Frame

The most natural frame to perform an operation in an analysis (like jet finding) is the laboratory frame, that is the frame of the ZEUS detector. The Z direction is set by the direction of the proton beam, Y points upwards (in the vertical plane) and X lies in the horizontal plane, pointing towards the center of the HERA ring.

In this frame, the momenta of the proton and of the positron are:

$$P_{proton} = \begin{pmatrix} 820.0 \\ 0 \\ 0 \\ 820.0 \end{pmatrix}, \; P_{positron} = \begin{pmatrix} 27.52 \\ 0 \\ 0 \\ -27.52 \end{pmatrix} \tag{C.1}$$

The four momentum of the photon is defined as the difference between the momentum of the scattered positron and $P_{positron}$. In order to conserve energy and momentum, the transverse momentum of the hadronic system must balance the transverse momentum of the scattered positron.

In both studies performed here, the jets are found in this frame. This means that the clustering algorithm is performed on four-momenta defined by the cells of the detector. This is more natural from an experimental point of view as the cell configuration in another frame might be very different from the one in the detector. Besides, the reconstruction of the jets in the laboratory frame is independent from the reconstruction of the scattered positron, which is not the case in the other frames.

C.2 The Center of Mass Frame (HCM)

In general, from the theoretical point of view, it is easier to perform calculations without taking into account the boost due to the difference of momentum between the two particles taking part in the collision. This problem does not exist in the LEP e^+e^- environment, but it is crucial in the ep collisions.

The hadronic center of mass system accounts for this. If $\vec{P}$ is the momentum of the proton and $\vec{q}$, the momentum of the exchanged boson in the laboratory frame, $\vec{P'}$ and $\vec{q'}$ their momenta in the HCM frame, it is defined by the condition:

$$\vec{P'} + \vec{q'} = 0. \tag{C.2}$$

So the boost is defined by the β:

$$\vec{\beta} = \frac{\vec{P} + \vec{q}}{E_{proton} + E_{\gamma}} \tag{C.3}$$

The transformations are defined as usual, with the γ parameter: $\gamma = \frac{1}{\sqrt{1-\beta^2}}$, using the following set of equations [1]:

$$E' = \gamma(E - \vec{\beta} \cdot \vec{p}) \tag{C.4}$$

$$\vec{p'} = \gamma(\vec{p} - \vec{\beta}E) \tag{C.5}$$

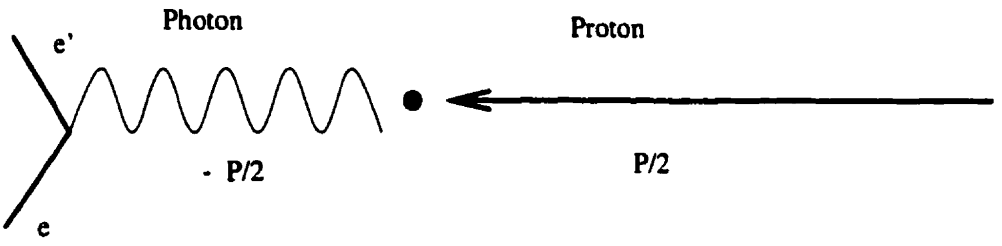

Figure C.1: *The Hadronic Center of Mass (HCM) system.*

The Z axis is usually defined by the position of the exchanged boson.

In the HCM frame, the total hadronic transverse momentum is 0 by construction. The total invariant mass of the system is $W^2 = (P + q)^2$. In the HCM frame, the current hemisphere is defined by the direction of the struck quark and the target (fragmentation) hemisphere is defined by the direction of the proton remnant. Both are not necessary collinear to each other. This would be only true in the QPM, where the struck quark does not carry an intrinsic transverse energy within the proton.

C.3 The Breit Frame

The Breit frame has been defined by analogy with the e^+e^- experiments, where the quark-antiquark pair is created back-to-back, with the same overall momentum (see figure C.2).

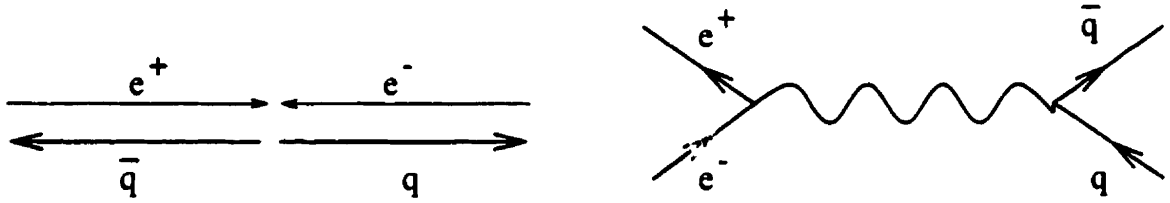

Figure C.2: e^+e^- *scattering.*

In an *ep* scattering process, the Breit frame is defined so that the current and target hemisphere (in the QPM) are collinear and carry the same momentum. It can be shown that this condition is fulfilled if the exchanged boson is completely space-like,

[1] the prime(') indices refer to the coordinate system in the frame of transformation

that is with no energy and a momentum $-Q$ (so that $Q^2 = -q^2$), so that we can write its four-momentum as: $(0,0,0,-Q)$. In this case, if the proton comes with a four momentum $(E_p,0,0,E_p)$. and the struck quark as a four momentum xP (in the QPM) so that $xP = (Q/2,0,0,Q/2)$, then because of momentum and energy conservation, the outgoing quark momentum will be $xP = (Q/2,0,0,-Q/2)$ (see figure C.3), according the equation:

$$\begin{pmatrix} Q/2 \\ 0 \\ 0 \\ Q/2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ -Q \end{pmatrix} = \begin{pmatrix} Q/2 \\ 0 \\ 0 \\ -Q/2 \end{pmatrix} \tag{C.6}$$

In this case, the outgoing quark is reverted, reminding of the back-to-back correlation of the e^+e^-. Assuming that the rest of the proton is not affected by the collision (which is a good approximation in the QPM), the Breit frame defines two hemispheres in the *ep* collision: the "target" (or fragmentation) hemisphere, which is defined for $p_{Z,Breit} > 0$ and is oriented along the proton direction in the Breit Frame, and the "current" hemisphere (for $p_{Z,Breit} < 0$), which is defined along the struck (leading order) quark. This distinction between target and current is only exact in the QPM and is approximative at leading order of α_s. Nevertheless, the frame is useful in practice to identify the jets which are probed (either LO jets or part of the gluon ladder).

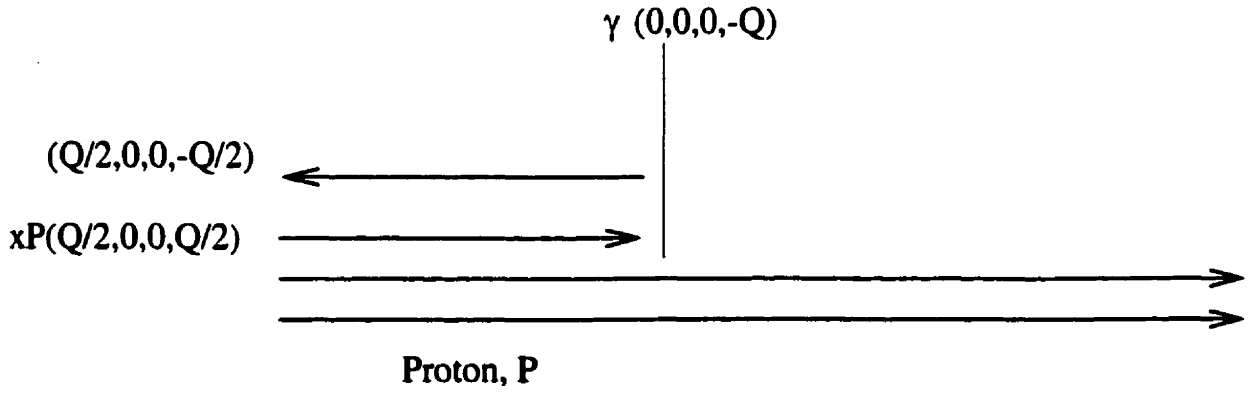

Figure C.3: *The Breit frame (in the QPM).*

In order to perform the transformation to the Breit frame, the photon direction has first to be rotated to lie in the negative Z' direction. The positron scattering plane usually defines the (X',Z'). The Lorentz transformation to the Breit frame is defined by the $\vec{\beta}$ vector:

$$\vec{\beta} = \frac{\vec{q} + 2x\vec{p}}{E_\gamma + 2xE_{proton}}, \tag{C.7}$$

where E_γ is the energy of the photon in the laboratory frame. The boost to the HCM or to the Breit frame have many theoretical advantages, as the calculations are easier in these frames and the interpretation in terms of hard and soft physics are more straight forward. Some jet algorithms (for instance the k_T [86]) cluster the four momenta of the particles found in the Breit frame as the comparison between the transverse momentum of the jet and the scale of the process is easier there (cf. chapter 6). Nevertheless, the experimental uncertainty due to the boost, and in particular the large dependence of the results on the reconstruction of the positron, is often the dominant contribution of the inaccuracy of the measurement (see for instance [133]). This is why this study has been performed in the laboratory frame.

Appendix D

Glossary

Expression	Meaning	Page
QCD	Quantum Chromodynamics; theory of the strong interaction	2
DIS	Deep Inelastic Scattering	2
QED	Quantum Electrodynamics; theory of the electroweak interaction	7
NLO	Next-to-Leading Order corrections	10
QPM	Quark-Parton Model; lowest order in DIS	11
QCDC	QCD Compton; first order in α_s in DIS	16
BGF	Boson Gluon Fusion; first order in α_s in DIS	16
DGLAP	Dokshitzer-Gribov-Lipatov-Altarelli-Parisi; evolution equations	21
LLA	Leading Logarithm Approximation	21
DLLA	Double Leading Logarithm Approximation	23
CTEQ	Coordinated Theoretical Experimental project on QCD; set of structure functions	24
MRS	Martin-Roberts-Stirling; set of structure functions	24
GRV	Glück-Reya-Vogt; set of structure functions	24
BFKL	Balitzki-Fadin-Kuraev-Lipatov; evolution equation	27
CCFM	Ciafaloni-Catani-Fiorani-Marchesini; evolution equation	32
HCM	Hadronic Center of Mass frame	34
LEPTO	Monte Carlo simulation based on the MEPS model	35

Expression	Meaning	Page
ARIADNE	Monte Carlo simulation based on the colour dipole model	35
MEPS	Matrix Element-Parton Shower model (LEPTO and HERWIG)	44
MEPJET	Next-to-Leading order calculation package by Mirkes et al.	45
DISENT	Next-to-Leading order calculation package by Seymour et al.	45
DISASTER	Next-to-Leading order calculation package by D. Graudenz	45
ISR	Initial State Radiations	47
FSR	Final State Radiations	47
HERWIG	Monte Carlo simulation based on the MEPS model and using the CLUSTER model for hadronization	47
CDM	Colour Dipole Model (ARIADNE)	47
LDC	Linked Dipole Chain model	49
SCI	Soft Colour Interactions in the LEPTO simulation	52
FCAL	Forward Calorimeter	60
BCAL	Barrel Calorimeter	60
RCAL	Rear Calorimeter	60
EMC	Electromagnetic cells in the calorimeter	62
HAC	Hadronic cells in the calorimeter	62
SRTD	Small Rear Tracking Detector	64
CTD	Central Tracking Detector	65
GFLT	Global First Level Trigger	69
GSLT	Global Second Level Trigger	69
TLT	Third Level Trigger	71
PUCELL	Type of Cone algorithm used to define a jet	83
PDF	Parton Distribution Function	174

Bibliography

[1] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language,* Nucl. Phys. B126 (1977) 298-331.

[2] M. Gell-Mann, Phys. Rev. 125 (1962) 1067;
Y. Ne'eman, Nucl. Phys. 26 (1961) 222.

[3] M. Gell-Mann, Phys. Lett. 8 (1964) 214;
G. Zweig, CERN-8192/TH 401(1964).

[4] O.W. Greenberg, Phys. Rev. Lett. 13 (1964) 598.

[5] K. Wilson, *Confinement of Quarks,* Phys. Rev. D10 (1974) 2445-2459.

[6] B.R. Martin and G. Shaw, *Particle Physics,* John Wiley and Sons (1992).

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

[8] R.N. Cahn and G. Goldhaber, *The Experimental Foundations of Particle Physics,* Cambridge University Press (1991).

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

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

[11] BCDMS Collab., A.C. Benvenuti et al., Phys. Lett. B223 (1989) 485.

[12] NMC Collab., P Amaudruz et al., Phys. Lett. B295 (1992) 159.

[13] R.G. Roberts, *The Structure of the Proton,* Cambridge University Press (1994).

[14] C.G. Callan and D. Gross, Phys. Rev. Lett. 22 (1969) 156.

[15] ZEUS Collaboration, M. Derrick et al., Z. Phys. C72 (1996) 399.

[16] M. Kuhlen, *QCD and the Hadronic Final State in Deep Inelastic Scattering at HERA,* Habilitationsschrift, Universität Hamburg (1997).

[17] Yu.L. Dokshitzer, V.A. Khoze, A.H. Mueller, S.I. Troyan, *Basics of Perturbative QCD,* Editions FRONTIERES, 1994.

[18] V.N. Gribov and L.N. Lipatov, Sov. J. Nucl. Phys. 15 (1972) 438 and 675;
Yu.L. Dokshitzer, Sov. Phys. JETP 46(1977) 641;
G. Altarelli and G. Parisi, Nucl. Phys. B126 (1977) 297.

[19] A.H. Mueller and J. Qiu, Nucl. Phys. B 268 (1986) 427-452.

[20] A.H. Mueller, J.Phys. G. 19 (1993) 1463.

[21] M.G. Ryskin and E.M. Levin, Phys.Rep. 189 (1990) 267;
A.H. Mueller, Nucl. Phys. B335 (1990) 115;
W.Zhu et al., Phys Rev. D 54 (1996) 847.

[22] A.De Rujula et al., Phys.Rev. D 10 (1974) 1649.

[23] H. Lai et al, Phys. Rev D51 (1995) 4763;
H. Lai et al, Phys. Rev D55 (1997) 1280.

[24] A.D. Martin, R.G. Roberts and W.J. Stirling, Phys. Rev D47 (1993) 867;
A.D. Martin, R.G. Roberts and W.J. Stirling, Phys. Rev D50 (1994) 6734;
A.D. Martin, R.G. Roberts and W.J. Stirling, Phys. Lett. B387 (1996) 419.

[25] A.M. Cooper-Sarkar, R.C.E. Devenish and A. De Roeck, Int.J.Mod.Phys. A13:3385-3586 (1998).

[26] M. Glück, E. Reya and A. Vogt, Z. Phys. C48 (1990) 471;
M. Glück, E. Reya and A. Vogt, Z. Phys. C53 (1992) 127;
M. Glück, E. Reya and A. Vogt, Z. Phys. C67 (1995) 433.

[27] E. Kuraev, L.N. Lipatov and V.S. Fadin, Sov. Phys. JETP 45 (1977) 199;
Y.Y. Balitzki and L.N. Lipatov, Sov. J. Nucl. Phys. 28 (1978) 822.

[28] J.R. Forshaw and D.A. Ross, *Quantum Chromodynamics and the Pomeron* Cambridge University Press (1997).

[29] J. Kwiecinski, J. Phys. G. 19 (1993) 1443.

[30] M. Ciafaloni, hep-ph/9709390, in Proc. Workshop "New Trends in HERA physics", Schloß Ringberg, Tegernsee 1997.

[31] J. Stirling, Proceedings of the Workshop "Deep Inelastic Scattering 1998", Bruxelles, April 1998 (ed. World Scientific).

[32] V.S. Fadin, in *Proceedings of the* 5^{th} *International Workshop on Deep Inelastic Scattering and QCD;* Chicago, IL, USA, eds. J. Repond and D. Krakauer, (American Institute of Physics, 1997), p. 924;
V.S. Fadin and L.N. Lipatov, hep-ph/9802290;//V.S. Fadin Proceedings of the workshop "Deep Inelastic Scattering 1998", Bruxelles, April 1998 (ed. World Scientific).

[33] J. Bartels, *private communication.*

[34] M. Ciafaloni, Nucl. Phys. B296 (1988) 49;
S. Catani, F. Fiorini and G. Marchesini, Phys. Lett. B234 (1990) 339;
Nucl. Phys. B336 (1990) 18.

[35] Yu.L. Dokshitzer, V.A. Khoze, A.H. Mueller and S.I. Troyan, Rev. Mod. Phys 60 (1988) 373.

[36] M. Ciafaloni, Nucl. Phys. B296 (1987) 249;
S. Catani, F. Fiorani and G. Marchesini, Phys. Lett. 234 B (1990) 339;
S. Catani, F. Fiorani and G. Marchesini, Nucl. Phys. B336 (1990) 18.

[37] G.P. Salam, Proceedings of the Workshop "Deep Inelastic Scattering 1998", Bruxelles, April 1998 (ed. World Scientific).

[38] J. Kwiecinski, A.D. Martin and P.J. Sutton, Phys. Rev. D46 (1992) 921;
Phys. Lett. B287 (1992) 254.

[39] The D0 collaboration, *AZIMUTHAL DECORRELATION OF JETS WIDELY SEPARATED IN RAPIDITY in D0* FERMILAB-CONF-96-304-EH, Sep 1996. 3pp. Presented at 1996 Annual Divisional Meeting (DPF 96) of the Division of Particles and Fields of the American Physical Society, Minneapolis, MN, 10-15 Aug 1996.

[40] J. Kwiecinski, A.D. Martin, P.J. Sutton and K. Golec-Biernat, Phys. Rev. D 50 (1994) 217;
J. Kwiecinski et al, Phys. Lett. B335 (1994) 220.

[41] H1 Collab., S. Aid et al., Phys. Lett. B356 (1995) 118.

[42] N. Pavel, Proc. of the Int. Workshop on DIS and Related Phenomena, ed. G. D'Agostini and A. Nigro, Rome 1996;
T. Haas, Proceedings of the Workshop on DIS and QCD, Paris, April 1995.

[43] G. Ingelman, Proc. 1991 Workshop on Physics at HERA, ed. W. Buchmuller and G. Ingelman (DESY, Hamburg, 1992), Vol. 3, 1366;
M. Bengtsson, G. Ingelman and T. Sjöstrand, Nucl. Phys. B301 (1988) 554;
G. Ingelman, A. Edin and J. Rathsman Computer Physics Communications 101 (1997) 108-134.

[44] L. Lönnblad, Lund preprint, LU TP-89-10;
L. Lönnblad, Comp. Phys. Comm. 71 (1992) 15.

[45] T. Carli, *Proceedings of the XVI Conference on Physics in Collision*, Mexico City, June 1996.

[46] M. Kuhlen, Phys. Lett. B382 (1996) 441;
M. Kuhlen, in *Proc. of Workshop on Future Physics at HERA vol1., DESY (1995).*

[47] The H1 collaboration, C. Adloff et al., Nucl. Phys. B485 (1997) 3.

[48] T. Wengler, Proceedings of the Workshop "Deep Inelastic Scattering 1998", Bruxelles, April 1998 (ed. World Scientific).

[49] J. Kwiecinski, S.C. Lang and A.D. Martin, Phys. Rev. D55 (1996) 1273.

[50] A.H. Mueller, Journ. of Phys. G 17, 1443 (1991).

[51] J.G. Kröner, E. Mirkes and G.A. Schüler, Int. J. Mod. Phys. A4 (1989) 1781.

[52] M. Seymour, *Monte Carlo Event Generators;* CTEQ Summer School. Interlaken, Lake Como, Wisconsin. 27 May - 4 June 1997.

[53] Erwin Mirkes, *THEORY OF JETS IN DEEP INELASTIC SCATTERING.* TTP-97-39, Oct 1997. 121pp, hep-ph/9711224;
E. Mirkes and D. Zeppenfeld, Phys. Lett. B380 (1996) 205 [hep-ph/9511448].

[54] S. Catani and M.H. Seymour, Nucl. Phys. B485 (1997) 291-419, Erratum-ibid. B510 (1997) 503-504.

[55] DISASTER++: D. Graudenz, hep-ph/9710244,
http://wwwcn.cern.ch/ graudenz/disaster.html.

[56] A.H. Mueller, Nucl. Phys. B213 (1983) 85. Erratum quoted *ibid.*, B241 (1984) 141;
Yu.L. Dokshitzer and S. I. Troyan. "Proceedings of the XIX Winter School of the LNPI, volume 1, page 144, Leningrad, 1984.

[57] H. Jung, The RAPGAP Monte Carlo for Deep Inelastic Scattering, version 2.06, 1997, http://www-h1.desy.de/~jung/rapgap.html.

[58] G. Marchesini et al., Cavendish-HEP-90/26, DESY 91-048;
G. Marchesini et al., Comp. Phys. Comm. 67 (1992) 465.

[59] R. D. Field and S. Wolfram, Nucl. Phys. B213 (1983) 65;
B. R. Webber, Nucl. Phys. B238 (1984) 492.

[60] T. Sjöstrand, Comput. Phys. Commun. 39 (1986) 347;
T. Sjöstrand and M. Bengtsson, Comput. Phys. Commun. 43 (1987) 367.

[61] B. Andersson et al., Z. Phys. C43 (1989) 625.

[62] H1 Collab., C. Adloff et al., preprint hep-ex/9806029.

[63] J. Rathsman, Phys. Lett. B393 (1997) 181-187.

[64] B. Andersson et al., Nucl. Phys. B463 (1996) 217.

[65] H. Kharraziha and L. Lönnblad, JHEP 03(1998)006, preprint hep-ph/9709424.

[66] B. Andersson et al., Phys. Rep. C97 (1983) 31.

[67] T. Sjöstrand, Nucl. Phys. B248 (1984) 469;
G. Ingelman, Physica Scripta Vol. 33(1986) 39.

[68] A. Edin, G. Ingelman and J. Rathsman, Phys. Lett. B366 (1996) 371-378;
A. Edin, G. Ingelman and J. Rathsman, J. Phys. G22 (1996) 943-946.

[69] M. Derrick et al., ZEUS collaboration, Phys. Lett. B315 (1993) 481;
H1 collaboration, Nucl. Phys. B429 (1994) 477.

[70] A. Edin, G. Ingelman, J. Rathsman, Z.Phys.C75 (1997) 57-70.

[71] T. Ahmed et al., H1 Collaboration, Z. Phys. C63 (1994) 377.

[72] ZEUS Collaboration, The ZEUS Detector, Status Report 1993, February 1993.

[73] R. Brun et al., GEANT program manual, CERN program library (1992).

[74] ZGANA, ZEUS trigger simulation library, Els de Wolf et al.

[75] ZEUS Collaboration, Status Report 1993 - The Blue Book, ZEUS 93, DESY (1993).

[76] N. Harnew et al., Nucl. Inst. and Meth. A 279 (1989), 290;
B. Foster et al., Nucl. Phys. B (Proc. Suppl.) (1993), 181;
B. Foster et al., Nucl. Inst. and Meth. A 338 (1994), 254.

[77] M. Derrick et al., Z. Phys. C69 (1996) 607.

[78] M. Derrick et al., Nucl. Inst. and Meth. A 309 (1991), 77;
A. Andresen et al., Nucl. Inst. and Meth. A 309 (1991), 101;
A. Bernstein et al., Nucl. Inst. and Meth. A 336 (1993), 23.

[79] A. Bamberger et al., Nucl.Inst. and Meth. 382 (1996) 419.

[80] H. Bethe, W. Heitler, Proc.Roy.Soc., A146 (1934) 83.

[81] LOCAL electron finder. ZEUS Phantom user library.

[82] LOCAL electron finder. ZEUS Phantom user library.

[83] H. Abramovicz, A. Caldwell and R. Sinkus, Nucl. Inst. Meth. A365 (1995) 508.

[84] Y.I. Azimov et al., Z. Phys. C27(1985)65.

[85] JADE collaboration, S Bethke et al., Phys. Lett. B213 (188) 235.

[86] N. Brown and W.J. Stirling, Phys. Lett. B252 (1990) 657.

[87] Yu.L. Dokshitzer, J. Phys. G17 (1991) 1537.

[88] M. Seymour, Z. Phys. C62(1994) 127.

[89] G. Sterman and S. Weinberg, Phys. Rev. Lett. 39(1997) 1436.

[90] J. Huth et al., Proc. of the 1990 DPF Summer Study on High Energy Physics, Snowmass, Colorado, edited by E.L. Berger (World Scientific, Singapore, 1992) p. 134.

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

[92] B. Abbott et al., D0 collaboration, FERMILAB-Pub-97/242-E.

[93] M. Seymour, Jets in QCD, CERN Preprint CERN-TH-95-176.

[94] ZEUS Collab., M. Derrick et al., Phys. Lett. B363 (1995) 201.

[95] J. Repond, *Direct Measurement of the Gluon Density of the Proton at Large x*, ZEUS Note 94 137;
G. Cases, *Inclusive two-jet cross sections in Deep Inelastic Scattering Neutral Current events at HERA and measurement of the gluon density of the proton*, ZEUS Note 95 116.

[96] ZEUS Collaboration, DESY 98-038, submitted to the European Physical Journal.

[97] T. Brodkorb and J.G. Körner, Z. Phys. C54 (1992) 519;
T. Brodkorb and E. Mirkes, Z. Phys. C66 (1995) 141.

[98] D. Graudenz, CERN-TH.7420/94 (1994).

[99] Particle Data Group, C. Caso et al., Euro. Phys. Journal C 3 (1998).

[100] ZEUS Collaboration, M. Derrick et al., Phys. Lett. B345 (1995) 576-588, H1 and ZEUS Collaborations (L. Lindemann for the collaboration), ucl.Phys.Proc.Suppl.64 (1998) 179-183.

[101] A. Edin, *The gluon density of the proton from 3-jet events in ep collisions at HERA*, Undergraduate thesis, TSL/ISV-93-0087.

[102] H1 Collaboration, Nucl. Phys. B449 (1995) 3.

[103] D. Mikunas for ZEUS collab., Proc. 1997 DIS97 workshop. Chicago 1997, eds J. Repond and D. Krakauer.

[104] OPAL Collaboration, Z. Phys. C63 (1994) 363.

[105] ALEPH Collaboration, Phys. Lett. B346 (1995) 389.

[106] S.J. Brodsky and J. Gunion, Phys. Rev. Lett. 37 (1976) 402;
K.Shizuya and S.H.H. Tye, Phys. Rev. Lett. 41 (1978) 787.

[107] D0 Collaboration, Phys. Lett. B357 (1995) 500.

[108] CDF Collaboration, Phys. Rev. Lett. 70 (1993) 713.

[109] OPAL Collaboration, Z.Phys. C63 (1994) 197.

[110] A.J. Askew, D. Graudenz, J. Kwiecinski, A.D. Martin, Pys. Lett. B338 (1994) 92-97.

[111] A.H. Mueller, Nucl. Phys. Proc. Suppl 18C (1991) 125-132;
A.H. Mueller, Journ. of Phys. G 17, (1991) 1443.

[112] J. Kwiecinsky, *private communication.*

[113] J. Bartels, V. Del Duca, A.De Roeck, D. Graudenz and M. Wüsthoff, Phys. Lett. B384 (1996) 300-306.

[114] G. Kramer and S.G. Salesch, Phys. Lett. B317 (1993)218;
G. Kramer and S.G. Salesch, Phys. Lett. B333 (1994)219.

[115] Jose Repond, *private communication.*

[116] A. Caldwell, W. Liu, B. Straub, ZEUS-note 98-018;
F. Pelucchi, S. Schlenstedt, J.H. Vossebeld, ZEUS-note 96-104.

[117] J. Repond, *How to Determine Correction Factors,* ZEUS Note 98-019.

[118] J. Repond, *Jet Energy Corrections,* ZEUS Note 96-133.

[119] J. Grosse-Knetter, *Energy Correction for Islands,* ZEUS Note 97-039.

[120] B. Lu, L. Mo, *Calorimeter Energy Corrections for the 1994 ZEUS Data,* ZEUS Note 95-098.

[121] S. Magill, *Energy Corrections to Jets in the ZEUS Calorimeter,* ZEUS Note 94-042.

[122] Stan Bentvelsen, Jos Engelen and Paul Kooijman, Proc. 1991 HERA Workshop, ed. W. Buchmüller and G. Ingelman (October 29-30, 1991).

[123] W.R. Leo, *Techniques for Nuclear and Particle Physics Experiments* Springer-Verlag, 1987.

[124] E. Mirkes and S. Willfahrt, Phys. Lett. B414 (1997) 205-209.

[125] H1 col., Proc. 1998 DIS98 workshop. Bruxelles 1998, eds World Scientific.

[126] H. Kharraziha and L. Lönnblad, Lund preprint, LU TP-97-21, hep-ph/9709424.

[127] L. Lönnblad, *private communication.*

[128] E. Mirkes and D. Zeppenfeld, Phys. Lett. B380 (1996) 205-212.

[129] A. Kwiatkowski, H. Spiesberger and H.J. Möhring, Proc. of Workshop on Physics at HERA vol3, DESY (1992).

[130] H. Jung, Talk given at Madrid Workshop on low x Physics, Miraflores de la Sierra, June 18-21, 1997.

[131] G. Schuler,T. Sjöstrand, Phys. Lett. B380 (1996) 105.

[132] J. Hartmann et al., *Forward Jet Production in Deep Inelastic Scattering Events at HERA,* ZEUS-Note 98-050 (1998).

[133] PUCELL jet finder. ZEUS Phantom user library.

[134] V. Noyes, Ph. D. Thesis, Glasgow University, November 1994, unpublished.