Event Shapes and Power Corrections at HERA

Gavin John McCance

Department of Physics & Astronomy
University of Glasgow

Thesis submitted for the degree of
Doctor of Philosophy

May 2001

Abstract

A measurement of event shape variables in neutral current deep inelastic ep scattering has been made at HERA with the ZEUS detector, using an integrated luminosity of 45.2 pb^{-1}. The variables thrust and broadening, with respect to the photon axis and the thrust axis, as well as the jet-mass and C-parameter, have been measured in the current region of the Breit frame in the kinematic range $10 < Q^2 < 20480$ GeV^2 and $0.0006 < x < 0.6$. The Q dependence of the event shapes have been compared to QCD predictions using Next-to-Leading Order calculations in conjunction with a power correction model to account for hadronisation. The model is tested by extracting the strong coupling constant $\alpha_s(M_Z)$ and a new non-perturbative parameter, $\overline{\alpha_0}(\mu_I)$.

To My Parents

Author's Declaration

This thesis presents measurements of event shape variables using data taken in 1995 - 1997 by the ZEUS detector. The work is all my own, although many of the analysis techniques used reflect the work and efforts of the Glasgow ZEUS group and the ZEUS Collaboration as a whole.

My responsibilities on ZEUS include data quality expert for the Hadronic Final States and QCD working group. This involved me implementing and updating the procedure to monitor the quality of data used by the group; the work is presented in chapter 3. I was jointly responsible for the Monte Carlo production of our group. I also had regular shifts checking the data quality of the ZEUS Central Tracking Detector, which is one of the UK's detector responsibilities, and had regular shifts taking data with the ZEUS experiment.

My first year concluded with a study of Next-to-Leading Order programs for event shapes, results subsequently used in this analysis. My second year was spent implementing the data quality checks, contributing to the Zeus Event Store tag-database, and developing analysis techniques. My third year was spent making the final analysis of event shapes presented in this thesis. An interim result was presented at the Rencontres de Moriond 2000 conference and the UK Institute of Physics 2000 meeting.

Acknowledgements

I would like to thank Prof. David Saxon for giving me the opportunity to work with the ZEUS group at Glasgow University and PPARC for funding me over the three years, particularly generously during my attachment at DESY.

Many thanks to members of the Glasgow ZEUS group who have helped me over the three years. In particular I would like to thank my supervisor Tony Doyle for being constantly enthusiastic about life in general and the analysis in particular, and for patiently answering my innumerable questions. Thanks also to Ian whose help with the analysis was invaluable, and whose sense of humour kept me (moderately) sane during the last few months. Thanks to both of them for correcting this thesis.

I would like to thank members of the Hadronic Final States and QCD working group at DESY, in particular, Juan, Costas, and Stefan whose help with the conference result was much appreciated. My thanks go to Catherine MacIntyre and Susan Ketels for all their help and for sorting out all the various problems that cropped up over the three years.

I would like to mention the DESY crowd, in particular those who I shared an office with - Neil, Rob, Nick, Laurel, Sung-Won, Anna, Claudia and Peter - thanks for putting up with my poor musical taste and occasional tempers! Also, thanks go to the usual suspects - Alex, Eileen, Dave, Rod, John and others - for being (ever) available for a few quick halfs down the local, and making my time in Hamburg as fun as it was.

Thanks to all the Glasgow crowd too, Barbara, Stephen, Keith, the three Daves, and my two suffering flat-mates, Craig and Steve, for reminding me that Scottish beer is best and for being good friends.

I would like to thank my parents for all their love, support and encouragement over the years, and for some superb holidays in the south of France. And finally, thanks to my best friend Kerry (and wee Bruno) for believing in me, putting up with me, and encouraging me all the way; I couldn't ask for more!

February 2001

Glasgow

Outline

In this thesis, a brief introduction to the theoretical background of Quantum Chromodynamics and Deep Inelastic Scattering is made. The event shape variables are defined and the theory of power corrections is introduced in chapter 1. The ZEUS detector is described in chapter 2 with emphasis on the components used in the analysis. Chapter 3 describes the offline reconstruction of the data and the data quality checks made. The various kinematic reconstruction methods are evaluated in chapter 4. The cuts used to reject background events and ensure that the event is well measured are presented. Chapter 5 examines the method used to reconstruct the hadronic final state and gives the hadron level definition for the event shapes. The Monte Carlo models for the simulation of physics events are described in chapter 6. The measurement and correction techniques employed in the analysis, and the various systematic errors which affect the data are presented in chapter 7. The fits to the power correction model of Dokshitzer, Webber et al. are made in chapter 8, and the uncertainties examined. Conclusions are presented in chapter 9.

Contents

List of Figures

List of Tables

Chapter 1

Theoretical Background

In this chapter, a brief introduction to particle physics is presented, followed by a discussion of the Quark Parton Model and the extended version of this model from the theory of Quantum ChromoDynamics. Deep Inelastic Scattering is discussed and the Breit frame of reference is introduced. The event shapes that will be measured in this analysis are defined and discussed. The power correction theory of hadronisation that will be compared to the the measured data is briefly described, together with Next-to-Leading Order predictions for the event shapes.

1.1 Particles and Interactions

The study of high energy particle physics is motivated by the desire to understand the fundamental building blocks of the universe and laws governing their behaviour [1]. The experimental study of matter did not begin until the early 19$^{\text{th}}$ century when John Dalton categorised the existence of individual chemical elements [2]. Subsequent discoveries by Thomson and Rutherford of the electron [3] and the proton [4], and theoretical work by Dirac [5] formed the basis of the Standard Model of particle physics. Its first great success came in 1933 with the discovery of the positron in cosmic rays [6], the essential ingredient of antimatter in the Standard Model. Subsequently, many new particles were discovered and various groups and models were proposed.

The leptons, such as the electron, were grouped into three distinct 'generations', while the theory of quantum electrodynamics (QED) describing their charged interactions, was developed. This theory describes the electromagnetic interactions as the exchange of a virtual photon, γ, between the interacting charged particles. It proposes a coupling constant, α, which governs the strength of the interactions. The theory is not exactly soluble, but relies on perturbative expansions in the coupling constant α. Since α has a relatively small value, ($\alpha \sim 1/137$), only the first few terms need be calculated for accurate solutions to be obtained. Initially, the formalism used to describe QED had problems, in that one part of the model (internal loop diagrams) caused the calculations to diverge and made determination of observables impossible. This was solved by a technique known as *renormalisation* [7], whereby the problematic parts of the theory were factored into the definition of α. This has two practical consequences. Firstly, regardless of how the renormalisation is performed, the coupling 'constant' becomes energy dependent, thereafter referred to as a *running coupling constant*. Secondly, the value of the coupling cannot be determined by theory and must be fixed by experimental observation.

The most successful early model of the observed hadronic states was proposed by Gell-Mann in 1964 [8]. This grouped baryons and mesons according to the experimentally observed quantities of charge, isospin and strangeness. The spin-1/2 baryons, including the proton and neutron, were grouped into an octet, whilst the spin-3/2 baryons were grouped into a decuplet. The model proved successful with the prediction and subsequent discovery of the Ω^- [9].

The model was later reformulated as an $SU(3)$ grouping of *quarks*, with three flavours of quark initially predicted (up, down and strange). The existence of the Ω^- was a problem for this model, since it comprises three spin-1/2 strange quarks with exactly the same quantum numbers, forbidden by the Pauli exclusion principle. An extra degree of freedom, termed colour, was introduced to answer this problem. As colour has never been experimentally observed, it was theorised that hadrons can only form from quarks in such a way that the resulting combination is colourless. The gauge theory used to describe strong nuclear interactions between these quarks is referred to as Quantum ChromoDynamics (QCD), with the coupling constant α_s. In a similar manner to QED, theoretical calculations can be made using perturbative techniques. The theory is also renormalisable [10], so that α_s can only be determined by experiment and is also a function of energy. This latter point proves more problematic for QCD than for QED, since at low energies the coupling constant, $\alpha_s > 1$, meaning that increasing order terms in the perturbation series get larger rather than smaller, preventing the use of perturbative techniques in this energy regime.

1.2 Deep Inelastic Scattering

The first indication of the existence of quarks (or partons) within the proton was made at SLAC in the 1960s [11]. The structure of the proton was examined in Deep Inelastic Scattering (DIS) experiments, as shown in Figure 1.1, where a virtual photon from an electron[1] interacts with one of the putative

[1] In fact, the analysis presented here uses positrons. For the theory relevant to the measurement presented in this thesis, there is no practical difference between using electrons

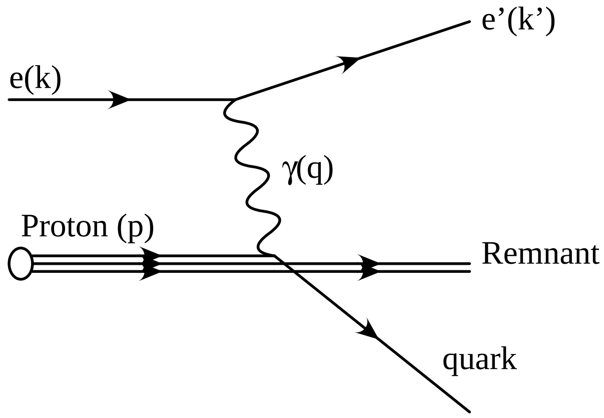


Figure 1.1: Lowest order (QPM) description of a DIS event.

partons within the proton. The process is characterised by two kinematic quantities, Q^2 and x

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

$$x = \frac{-q^2}{2p.q} \tag{1.2}$$

where the four momenta of the incoming electron, incoming proton, exchanged photon, and outgoing electron are denoted k, p, q and k' respectively. Q^2 is the momentum transfer from the electron to the struck parton and, at lowest order, x can be interpreted as the fraction of the proton's momentum carried by the struck parton. The interaction occurs in a very short time, inversely proportional to the virtuality of the exchanged photon, $t \sim 1/Q$, which means that the struck parton behaves as if it was a free particle within the proton; the remainder of the proton, termed the proton remnant, does not take much part in the interaction. An additional kinematic variable y is conventionally defined as

$$y = \frac{q.p}{k.p} \tag{1.3}$$

and positrons, so the two terms are used interchangeably.

and be interpreted as the relative momentum transfer from the electron to the proton in the proton's rest frame. It is related to the other variables via the HERA centre-of-mass energy s by

$$Q^2 = x.y.s. \tag{1.4}$$

The structure of the proton is parameterised by three functions $F_1(x)$, $F_2(x)$ and $xF_3(x)$. $F_2(x)$ was observed at SLAC to be essentially independent of Q^2 at a given x [11]. This was called *scaling* and was taken to indicate point-like substructure in the proton.

The measured differential cross section for neutral current DIS mediated by photon exchange (i.e. $Q^2 << M_{Z^0}^2$) can be expressed in terms of the structure functions as

$$\frac{d^2\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{xQ^4}\left[y^2 xF_1(x) + (1-y)F_2(x)\right] \tag{1.5}$$

Here, only the propagator for photon exchange $\sim 1/Q^4$ is considered, and QED radiative corrections are ignored. $xF_3(x)$ is a pure parity violating term, only present for an electroweak propagator with $Q^2 \sim M_{Z^0}^2$, and is not discussed here.

For spin-1/2 partons, the structure functions F_1 and F_2 are related by the *Callan-Gross relation* [12], as

$$F_2(x) = 2xF_1(x) \tag{1.6}$$

allowing Equation 1.5 to be simplified as

$$\frac{d^2\sigma}{dxdQ^2} = \frac{2\pi\alpha^2}{xQ^4}\left[(1 + (1-y)^2)F_2(x)\right] \tag{1.7}$$

The structure function, $F_2(x)$, can be interpreted as a sum over the momentum probability density functions, $q_i(x)$ of all the quark flavours, and $\bar{q}_i(x)$

of all the antiquark flavours within the proton, expressed as

$$F_2(x) = \sum_i e_i^2 (xq_i(x) + x\bar{q}_i(x)) \tag{1.8}$$

where e_i is the charge on quark flavour i, and the sum i runs over all flavours of quark in the proton. This model forms the Quark Parton Model (QPM) of the proton.

1.3 QCD and Deep Inelastic Scattering

If the QPM were an adequate representation of the structure of the proton, then the momentum sum rule for quarks would hold with

$$\int_0^1 \left(\frac{F_2(x)}{x} \right) dx = 1. \tag{1.9}$$

However, the early measurements of F_2 showed that only $\sim$ 50% of the proton's momentum is carried by the quarks. The remaining momentum is carried by spin-1 particles called *gluons*, which are identified as the mediating particle of the QCD interaction. The QCD-modified picture of the proton then comprises the three *valence* quarks of the QPM, interacting via gluons. These gluons can split to form $q\bar{q}$ pairs of all flavours, known as *sea* quarks. The contributions from the different flavours are suppressed according to their mass, so that the majority of sea quarks are the lightest ones, the up, down and strange quarks. The consequence of these gluons on the structure functions is a logarithmic dependence on Q^2; this *scaling violation* was first observed at SLAC after the measurements were extended to low x ($x < 0.1$) [13] and is due to the dominance of gluons at low x.

The leading order corrections to the QPM as a result of gluons are shown in Figure 1.2, where the quark can radiate a gluon before or after being struck by the exchanged photon (QCD Compton, QCDC), or where a gluon in the

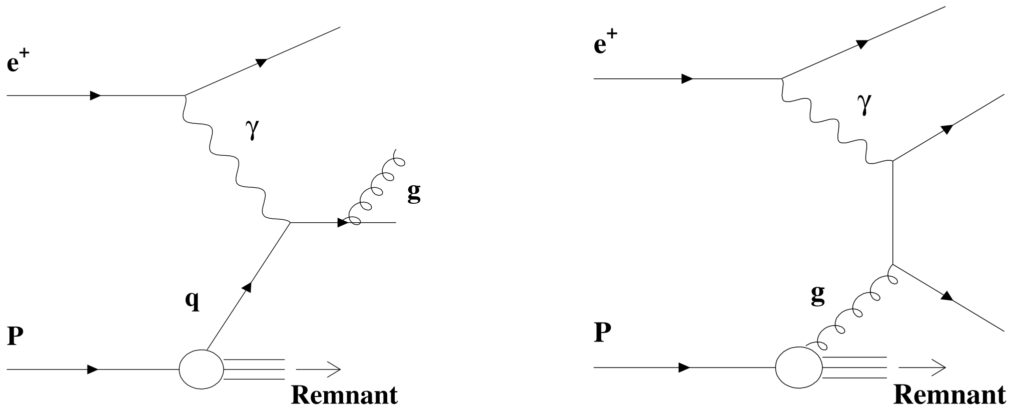


Figure 1.2: The $\mathcal{O}(\alpha_s)$ processes in DIS. a) Final state QCD Compton, b) Boson-Gluon Fusion.

proton may fluctuate into a $q\bar{q}$ pair which then interacts with the exchanged photon (Boson-Gluon Fusion, BGF). These processes are known as $(2+1)$ processes (two outgoing partons and one proton remnant.)

The strength of the interaction in QED is limited by electromagnetic screening as the exchanged photons fluctuate into a e^+e^- pair; this leads to a reduction of the coupling constant at large distances and the observed $1/r^2$ dependence with distance. The significant difference between QCD and QED is that QCD is a non-Abelian gauge theory. This means that the propagators themselves, the gluons, carry the colour charge, so that self-interactions are possible. Since there are 8 gluons which carry approximately double the colour-charge of a quark, the gluon-gluon self-interactions dominate and this self-interaction leads to an 'anti-screening' effect which gives the opposite behaviour with distance for QCD, i.e. the interaction strength increases with increasing distance. To leading order, the coupling constant that measures the strength of the gluon-gluon and quark-gluon interaction, α_s, can be expressed as

$$\frac{1}{\alpha_s(\mu_R^2)} = \frac{\beta_0}{4\pi}\ln(\mu_R^2/\Lambda_{QCD}^2) \tag{1.10}$$

where μ_R is the renormalisation scale; that is, the scale at which divergences in the theory are factored into the definition of the strong coupling. The conventional choice for DIS is to set $\mu_R^2 = Q^2$. β_0 is the first component of the β function which describes the running of α_s with Q^2 as

$$\mu_R^2\frac{\partial\alpha_s}{\partial\mu_R^2} = Q^2\frac{\partial\alpha_s}{\partial Q^2} = \beta(\alpha_s) \tag{1.11}$$

The β function expansion in α_s is given by

$$\beta(\alpha_s) = -\alpha_s\sum_{n=0}^{\infty}\beta_n\left(\frac{\alpha_s}{4\pi}\right)^{n+1} \tag{1.12}$$

where the first two β_n coefficients are

$$\beta_0 = 11 - \frac{2N_f}{3} \tag{1.13}$$

$$\beta_1 = 102 - \frac{38N_f}{3} \tag{1.14}$$

Here N_f is the number of active quark flavours, dependent on the energy of the interaction. In Equation 1.10, Λ_{QCD} is the value of Q at which quarks can be considered to be approximately free particles in the interactions. For $Q >> \Lambda_{QCD}$, $\alpha_s(Q^2)$ is small and the quarks and gluons can be considered as quasi-free particles; this property is referred to as *asymptotic freedom*. Since $\alpha_s(Q^2)$ is relatively small, perturbative methods are applicable, and calculations can be expanded as powers in $\alpha_s(Q^2)$ i.e. logarithmic terms in (Q^2/Λ^2).

As the struck parton moves away from the proton remnant, the interaction is dominated by clouds of virtual gluons. In the relativistic limit, the probability for a parton to radiate another parton is given by the Altarelli-Parisi splitting functions [14], an example of which is

$$P_{q\rightarrow qg}(z) = \frac{4}{3}\left(\frac{1+z^2}{1-z}\right) \tag{1.15}$$

which gives the probability for a quark of momentum p, to radiate a gluon of momentum $(1-z)p$. This leading order correction to the QPM corresponds to the inclusion of the QCD Compton process illustrated in Figure 1.2 a). The cross section, in terms of the transverse momentum of the emitted gluon, p_T^2, now includes a term of the form

$$\sigma \sim P_{q\to qg}(z)\int_0^{p_T^2(max)} \frac{dp_T^2}{p_T^2} \tag{1.16}$$

This has a divergence as $p_T^2 \to 0$, which corresponds to the emission of a collinear gluon. The divergence can be regularised by defining a factorisation scale, μ_F^2 at which the divergent part of the integral is factorised into the definition of the structure function, with the penalty that the structure functions now explicitly depend on the choice of factorisation scale. In DIS, this scale in conventionally chosen to be $\mu_F^2 = Q^2$. The singularity of Equation 1.15 as $z \to 1$, corresponding to the emission of a soft gluon in the real diagram, is cancelled by a term from the corresponding virtual diagram. Similar terms in the cross section appear as a result of the inclusion of BGF in the calculation. There are two consequences of including the $\mathcal{O}(\alpha_s)$ diagrams. One is the dependence of the structure function on Q^2, due to the extra gluon emissions, so that $F_2(x) \to F_2(x,Q^2)$. The other is that the QPM Callan-Gross relation, given in Equation 1.6, no longer holds, leading to a longitudinal structure function, F_L, which is non-zero, defined as

$$F_2 - 2xF_1 \equiv F_L \tag{1.17}$$

Excluding parity violating terms and considering only the photon propagator, the cross section is then given by

$$\frac{d^2\sigma}{dxdQ^2} = \frac{2\pi\alpha^2}{xQ^4}\left[(1+(1-y)^2)F_2(x,Q^2) - y^2F_L(x,Q^2)\right] \tag{1.18}$$

As the outgoing quark showers to produce gluons, its virtuality decreases, leading to an increase in the effective value of α_s. For $Q \sim \Lambda_{QCD} \sim 0.3$ GeV, the coupling is very large, and the quarks and gluons are bound together to form colourless hadrons in a process termed hadronisation. The increase in the stored energy of the field as the distance increases and the interaction energy decreases is called *confinement*. In this low energy and large distance region, since α_s is large, perturbative methods are not applicable in the calculations, and non-perturbative techniques must be employed instead.

1.4 Breit Frame

The interactions in DIS occur between the exchanged photon and one of the partons within the proton, as shown in figure 1.3a) for the QPM situation. The photon direction in the laboratory frame is not convenient for analysis, therefore this frame of reference is not the most appropriate frame in which to study the hadronic final state. Additionally, DIS has the complication of the proton remnant, so it is important to find a frame of reference in which the hadronic system of interest suffers minimal contamination from the proton remnant.

A suitable frame to use is the Breit frame [15], which is defined as the frame where the four momentum of the exchanged virtual photon is purely spacelike, with $q = (0, 0, 0, -Q)$. Figure 1.3b) shows the same DIS interaction in the Breit frame. The Breit frame separates into two parts, the current region, into which the struck quark goes with $p_z < 0$, and the target region, into which the proton remnant goes. In the QPM model, the struck quark comes in with momentum $p_z = Q/2$, interacts with the exchanged boson, and leaves with equal and opposite momentum $p_z = -Q/2$. The proton remnant has momentum $p_z = (1 - x)Q/2x$. The use of the Breit frame simplifies the kinematics considerably, allowing the QCD dynamics to stand out (i.e. any

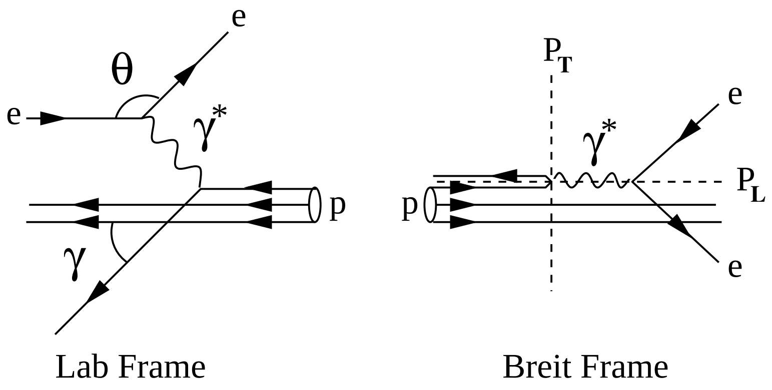


Figure 1.3: a) QPM model for the laboratory frame. b) QPM model boosted to the Breit frame.

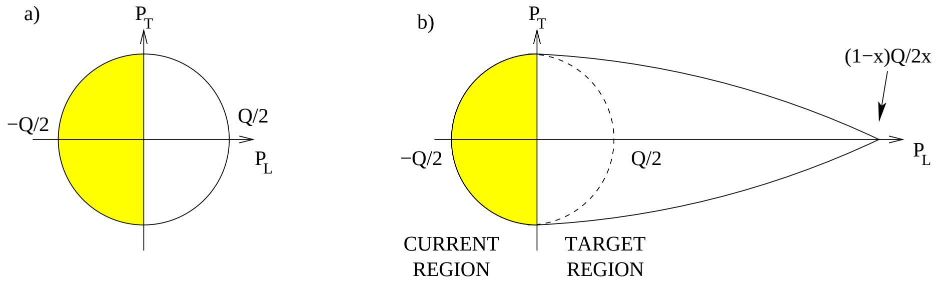


Figure 1.4: a) The shaded part indicates one hemisphere of an e^+e^- event. b) The shaded region indicates the current region of the Breit frame in DIS. The unshaded part with the extended phase space is the target region.

deviation from $p_z = -Q/2$, and $p_x = p_y = 0$ is due to QCD effects.) The frame also achieves the maximal separation between the struck quark and the proton remnant. Theoretically, the target region is not well understood and experimentally the majority of it is lost down the beam pipe, so the measurements in this analysis are confined to the current region.

Figure 1.4 shows the phase space for the current region of the Breit frame compared to one hemisphere of an e^+e^- event; in the QPM, they are identical, with momentum $\sqrt{s}/2 = Q/2$ available to the quark, where $\sqrt{s}$ is the e^+e^- centre-of-mass energy. The basic similarity allows comparison between the time-like process in e^+e^- and the space-like process in DIS. At higher orders, the analogy does not hold due to initial state QCD Compton and BGF, processes which are not present in e^+e^- interactions.

The $\mathcal{O}(\alpha_s)$ processes, QCDC and BGF, can produce two partons in the final state which have large transverse momentum. The QCDC and BGF processes require two more scaling variables to define the event. Integrating over azimuthal angle ϕ gives

$$x_p = \frac{Q^2}{2p'.q} = \frac{Q^2}{Q^2 + m^2} \qquad (x \leq x_p \leq 1) \tag{1.19}$$

$$z_p = \frac{P.p_{jet}}{P.q} = \frac{1}{2}(1 - \cos\theta^*_{jet}) \qquad (0 \leq z_p \leq 1) \tag{1.20}$$

where m is the invariant mass between the two jets, assuming the jets themselves are massless, and $p' = \xi P$ is the momentum of the incoming parton. The momentum of the outgoing jet under consideration is denoted p_{jet}, where θ^*_{jet} is the polar angle of the jet. There are three possible event topologies (topologies II and III are approximately equivalent) as illustrated by Figure 1.5; it is possible for both or neither of the jets to be found in the current region. The cross sections for the two $\mathcal{O}(\alpha_s)$ processes go as

$$d\sigma^{BGF} \propto \frac{1}{z_p(1 - z_p)} \tag{1.21}$$

$$d\sigma^{QCDC} \propto \frac{1}{(1 - x_p)(1 - z_p)} \tag{1.22}$$

As previously stated, the collinear singularity from the real $\mathcal{O}(\alpha_s)$ diagrams as $z_p \to 0$ or 1 is cancelled by an equal singularity in the corresponding virtual diagram. The $x_p \to 1$ singularity from soft initial state QCDC results in an enhancement for low invariant masses m, where both jets are found in the

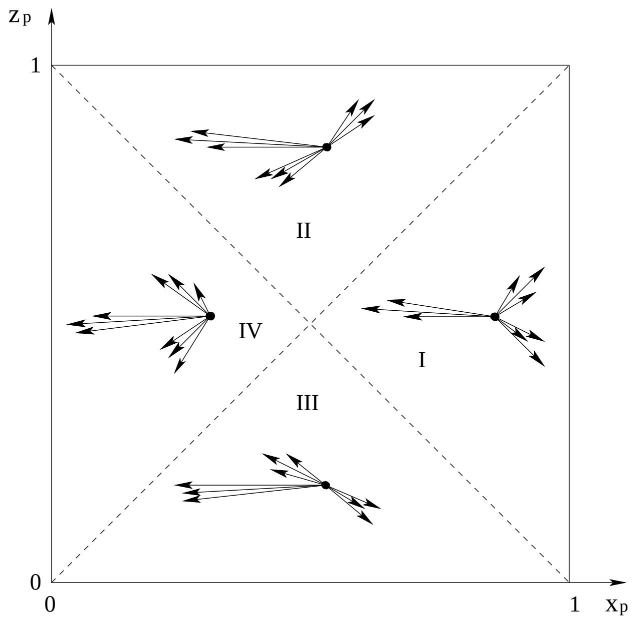


Figure 1.5: The three possible event topologies in the (x_p, z_p) plane. Region I has both jets in the current region. Region II and II have one of the jets in the target region. Region IV has both jets in the target region.

current region (topology I). Hard initial and final state QCD radiations lead to the population of regions II and III. The suppression of large invariant masses by $(1 - x_p)$ is not present in BGF, and jets with large invariant masses can dominate, leading to population of topology regions II, III and IV. At low x, the contribution from BGF, and hence topology IV, becomes more significant.

The transformation from the laboratory frame to the Breit frame is performed by a Lorentz boost followed by a rotation to align the virtual photon axis with the negative z-axis of the Breit frame. The frame travels with velocity $\vec{\beta}$ with respect to the laboratory frame, with $\vec{\beta}$ and γ calculated from the exchanged photon and the incoming proton four-vectors, $q = (q_0, \vec{q})$ and

$p = (p_0, \vec{p})$ respectively.

$$\vec{\beta} = \frac{\vec{q} + 2x\vec{p}}{q_0 + 2xp_0} \tag{1.23}$$

$$\gamma = \sqrt{1 - |\vec{\beta}|^2} \tag{1.24}$$

The particles' four momenta $A = (A_0, \vec{a})$ then transform as

$$A_0' = \gamma(A_0 - \vec{\beta}.\vec{a}) \tag{1.25}$$

$$\vec{a'} = \gamma(\vec{a} - A_0\vec{\beta}) \tag{1.26}$$

where a prime denotes the momenta in the Breit frame.

1.5 Event Shapes

Event shapes are observables which describe the topology of an event. Various properties of the topology can be measured, for example, the extent of collimation of the jet of particles or the broadness of the jet. Event shapes have been investigated in e^+e^- experiments [16], and can be used to extract the strong coupling constant, α_s. Due to the similarity between the current region of the Breit frame in DIS and one hemisphere of an e^+e^- event, most of the event shapes defined in e^+e^- are also applicable to DIS. The *thrust*[2] is defined as

$$\tau_T = 1 - T_T = 1 - \max_{\hat{t}} \frac{\sum |p.\hat{t}|}{\sum |\vec{p}|} \tag{1.27}$$

where $\vec{p}$ denotes the three-vector with components $\vec{p} = (p_x, p_y, p_z)$, and the sum is over all objects in the current region. The calculation must be iterated until the axis which maximises the total longitudinal momentum, T_T,

[2]The *thrust* is denoted by T in Equation 1.27 with $T \to 1$ for collimated events. τ, where $\tau = 1 - T$, is often plotted, however, since this has the same behaviour as other event shapes, i.e. it tends to zero as the events become more collimated.

is found. This axis is referred to as the *thrust axis*; $\hat{t}$ is a unit vector along the thrust axis. For completely collimated events, $\tau_T \rightarrow 0$, increasing to $\tau_T \rightarrow 1/2$ for isotropic events. The *broadening* is defined as

$$B_T = \frac{\sum |\vec{p} \times \hat{t}|}{2 \sum |\vec{p}|} \tag{1.28}$$

and is measured with respect to the already defined thrust axis. It is therefore the total transverse momentum with respect to this axis, and takes a maximum value of 1/2 for isotropic events, falling to zero for completely collimated events.

Since the Breit frame has a defined physical axis, that of the spacelike photon, two extra event shapes can be defined in DIS, namely the thrust and broadening with respect to the photon axis, defined respectively as

$$\tau_\gamma = 1 - T_\gamma = 1 - \frac{\sum |\vec{p}.\hat{n}|}{\sum |\vec{p}|} = 1 - \frac{\sum |p_z|}{\sum |\vec{p}|} \tag{1.29}$$

$$B_\gamma = \frac{\sum |\vec{p} \times \hat{n}|}{2 \sum |\vec{p}|} = \frac{\sum |p_T|}{2 \sum |\vec{p}|} \tag{1.30}$$

where $\hat{n}$ is the unit vector along the photon axis.

The particles used in the definitions are defined as massless particles, i.e. $E^2 = |p|^2 = p_x^2 + p_y^2 + p_z^2$. In the case of real particles, the energy is rescaled to the momentum after the boost to the Breit frame. This is done since comparisons are made with perturbative predictions + power correction theory, both of which deal with massless partons [17].

There are two variables that characterise the correlation between combinations of particles, and are not defined with respect to any axis, the *jet-mass*

$$\rho_0 = \frac{M^2}{4(\sum |\vec{p}|)^2} = \frac{(\sum |\vec{p}|)^2 - (\sum |p_x|)^2 - (\sum |p_y|)^2 - (\sum |p_z|)^2}{4(\sum |\vec{p}|)^2} \tag{1.31}$$

and the *C-parameter*

$$C = 3(\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_1\lambda_3) \tag{1.32}$$

where λ_i $(i = 1, 2, 3)$ are the eigenvalues of the linearised momentum tensor, $\Theta^{\mu\nu}$, defined as

$$\Theta^{\mu\nu} = \frac{\sum_i (\vec{p}_i^{\mu} \vec{p}_i^{\nu} / |\vec{p}_i|)}{\sum_i |\vec{p}_i|} \tag{1.33}$$

where $\mu, \nu = 1, 2, 3$ denote the tensor components. Alternative definitions of the jet mass and C-parameter, more convenient in terms of numerical calculation and understanding, are

$$\rho_0 = \frac{1}{4(\sum |\vec{p}|)^2} \sum_{i,j} |\vec{p}_i||\vec{p}_j|(1 - \cos\theta_{ij}) \tag{1.34}$$

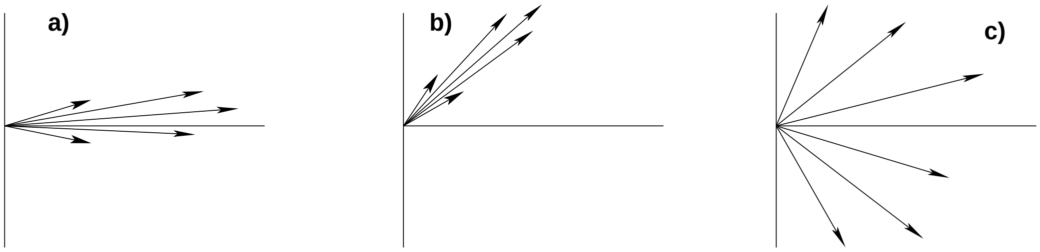
and

$$C = \frac{2}{3(\sum |\vec{p}|)^2} \sum_{i,j} |\vec{p}_i||\vec{p}_j|(1 - \cos^2\theta_{ij}) \tag{1.35}$$

where the sum is the square sum (i.e. each combination counts twice) over all combinations of $\vec{p}_i$ and $\vec{p}_j$, and θ_{ij} is the opening angle between the pair of particles i and j.

The shape definitions all have the characteristic property that they are *collinear safe* to varying degrees. This means that emitting an infinitely soft and collinear particle from one of the other particles does not change the value of the event shape (ie. splitting a particle $p_a \to p_b + p_c$ where $|p_c| \to 0$ will not significantly change the result.) It is the fact the longitudinal momenta enter linearly that makes the definitions generally collinear safe (the broadening suffers slightly here, since this property does not hold exactly for p_T.)

Figure 1.6 show the different possible values that the event shapes can take, from three different event topologies. In a) the event is collimated with respect to the photon axis, which minimises all variables. In c) the event is isotropic and all the event shapes take on maximum values. In b) the event is again collimated, but with the difference that the axis of collimation is not the Breit frame axis, but the 'natural' one as determined by the distribution of the particles. In a) the collimated particles, and hence the direction of the



$\tau_\gamma \to 0$ (min)	$\tau_\gamma \to 1$ (max)	$\tau_\gamma \to 1$ (max)
$B_\gamma \to 0$ (min)	$B_\gamma \to 1/2$ (max)	$B_\gamma \to 1/2$ (max)
$\tau_T \to 0$ (min)	$\tau_T \to 0$ (min)	$\tau_T \to 1/2$ (max)
$B_T \to 0$ (min)	$B_T \to 0$ (min)	$B_T \to 1/2$ (max)
$\rho_0 \to 0$ (min)	$\rho_0 \to 0$ (min)	$\rho_0 \to 1/4$ (max)
$C \to 0$ (min)	$C \to 0$ (min)	$C \to 1$ (max)

Figure 1.6: The values of event shapes for example event topologies in the current region. a) shows collimated particles along the photon axis, b) shows collimated particles but not associated to the photon axis, and c) shows an isotropic distribution of particles.

thrust axis, are all along the photon axis, therefore $\tau_\gamma = \tau_T$ and $B_\gamma = B_T$. In case b) the particles are still very collimated, with $\tau_T \to 0$ and $B_T \to 0$, but since the collimation is not with respect to the photon axis, $\tau_\gamma \to 1$ and $B_\gamma \to 0.5$.

Despite the apparently different original definitions, as given in Equations 1.31 and 1.32, the jet-mass and C-parameter are very similar, as shown in Equations 1.34 and 1.35. Both depend on the angle between all possible pairs of particles, and neither are defined with respect to any axis.

The jet-mass is a measure of the mean scaled invariant mass between pairs of particles. As such, it increases monotonically with increasing angle as an event becomes more isotropic. The minimum value is zero, corresponding to a completely collimated event, rising to 0.25 for two particles back to back in the current region (and hence contained in the z-plane.)

The C-parameter rises monotonically with angular separation, with the $(1-\cos^2)$ term reaching a maximum at an angle of $90°$, falling away symmetrically around $90°$. This property means that it is sensitive to the coplanarity of an event. A completely coplanar event has $C = 0$ and a completely acoplanar event has $C = 1$. It is interesting to note that values of $C > 3/4$ can only come from a configuration when there are three or more particles, and that there is a considerable phase space drop-off in the values of C-parameter for $C > 3/4$.

Two other points form part of the definition of the event shapes

- To be completely infrared safe (i.e. so that the Next-to-Leading Order calculation discussed in the next section does not diverge), a cut on the total energy of the particles in the current region is applied [18], given by

$$\mathcal{E}_{lim} > \zeta Q \tag{1.36}$$

  where $\zeta = 0.1$ for the nominal analysis. Any events failing this cut are not considered for calculation of the event shapes. Recent theoretical discussion [17] has suggested raising the cut to $\zeta = 0.25$ for a best comparison to the power correction theory; the effect of this will be examined in more detail in chapter 8.

- Since most of the event shapes are ambiguous or undefined for less than two particles, only events with two or more particles in the current region are used for the calculation of the event shapes and the definition of the hadronic final state.

1.6 Next-to-Leading Order Calculation

In order to study and test the validity of the power correction theory discussed in the next section, calculations up to $\mathcal{O}(\alpha_s^2)$ are required for the QCD matrix elements. Although the Monte Carlo event generators used to correct the data are only available up to leading order in α_s, there exist parton level calculations up to Next-to-Leading Order (NLO) [19].

The calculations include a large number of new process diagrams, for example, Figure 1.2 but with an extra gluon emission, or with the final state gluon splitting into a quark-antiquark pair. Additionally, virtual loop corrections to the diagrams in Figure 1.2 are considered, for example, a gluon connecting the incoming and outgoing quark lines. These virtual diagrams contain the divergences which are used to cancel the divergences in the real diagrams. The major difficulty in the calculation of observables to $\mathcal{O}(\alpha_s^2)$ is the presence of singularities in the integral, combined with the fact that the integral is too complex to perform analytically. The individual integrals are divergent, and only after they have been combined is the result finite. This is a problem for numerical integration methods.

To reconcile the problem requires a method that allows the divergent parts of the integral to be treated analytically, while the full complexity of the integral is treated numerically. There are two methods for this, the *phase space slicing method* [20] and the *subtraction method* [21]. Both permit arbitrary $(2+1)$ observables to be evaluated. Two NLO programs are used in this analysis, DISENT [22] and DISASTER++ [23]. Both are based on the subtraction method, so this method is discussed. The general form of a NLO cross-section is

$$\sigma = \sum_{a=1}^{m} f_a \otimes (\sigma_a^{LO} + \sigma_a^{NLO}) \tag{1.37}$$

where the sum a runs over all m partons in the final state, and $\otimes$ denotes the convolution of the parton density function f_a with the LO and NLO cross sections. The LO cross section, σ^{LO}, is comparatively straightforward to calculate; the integrals given in section 1.3 are generally finite or can be analytically regularised. To evaluate the NLO cross section, σ^{NLO}, requires the calculation of two cross sections which are individually divergent. The solution is to invent a 'fake' cross section which can be added to one of the cross sections and subtracted from the other, such that the fake cross section cancels point by point in phase space the singular behaviour of the two divergent real cross sections. Clearly this 'fake' cross section must be chosen very carefully, so as to be able to match the singular behaviour of the two different functions. The actual constraints to which this fake cross section must conform are given in [21].

Since the NLO calculation is $\mathcal{O}(\alpha_s^2)$, the maximum number of partons in the current region from which to build an event shape is three, which may be small compared to the average number of hadrons expected in an observed DIS event. However, the majority of the extra hadrons come from collinear or soft branching, and since event shapes are collinear and infrared safe, they are insensitive to these branchings and emissions (to a first approximation), and depend mostly on the hard large-angle emission of the partons.

A comparison between DISENT and DISASTER++, running under the Common NLO Library scheme [24], was presented at the HERA Monte Carlo Workshop [25]. This study, and subsequent studies at the workshop showed that all collinear safe event shapes agreed at the $< 1\%$ level, but found a large discrepancy between the two programs for the jet-broadening with respect to the photon axis, especially at low Q^2. A more detailed study examined smaller discrepancies for the thrust [26]. DISASTER++ agrees better with resummed analytical calculations for thrust [26], so is assumed to be more accurate than DISENT[3]. However, since DISASTER++ takes up to a factor of ten times longer to run a given simulation, the decision to use DISENT as the main NLO program was made, with a set of runs from DISASTER++ as a cross-check.

1.7 Power Corrections

There has been much recent work in the understanding of the non-perturbative process of hadronisation using aspects of perturbation theory as a starting point [27]. The motivation for the analysis presented here is to test the validity of one of these approaches, namely the model of Dokshitzer, Webber et al. [28]. This model exploits the fact that the perturbative series expansion is not expected to converge, even if calculated to all orders of α_s. This is as a result of chains of fermion loops inserted into gluon lines as shown in Figure 1.7; these *renormalons* result in factorial divergences which cause the entire series to diverge. These divergences of the perturbative series are used to describe the leading terms of the non-perturbative hadronisation stage, as described in [28].

The model can be investigated by comparing a NLO parton level prediction with the data. The difference in the distributions is due to the hadronisation correction and can be used in a fit to test the power correction model for hadronisation. In DIS, the power correction for the differential distributions have not generally been calculated yet[4]. Therefore, the test of the theory in this thesis investigates the mean values of the event shapes, for

[3]The resummed calculations for the broadening are not yet available.

[4]Although, recent resummed calculations in [26] present differential distributions for thrust.

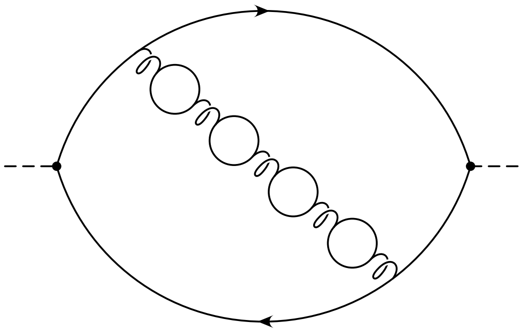


Figure 1.7: Bubble diagram with fermion loops (renormalons) inserted into the gluon lines.

which theoretical predictions are available. The hadronisation corrections to the mean values of the event shapes considered here are observed to go as $\sim 1/Q^p$ where $p = 1$, and are therefore termed *power corrections.*

In general, denoting the mean value of the corrected data as $\langle F \rangle$ and the $\mathcal{O}(\alpha_s^2)$ parton level prediction for the mean as $\langle F \rangle_{NLO}$, the power correction term, $\langle F \rangle_{pow}$, comes in additively as

$$\langle F \rangle = \langle F \rangle_{NLO} + \langle F \rangle_{pow} \tag{1.38}$$

The power correction term from the renormalon calculation, $\langle F \rangle_{pow}$, introduces a new non-perturbative phenomenological constant, $\overline{\alpha_0}$, defined as the mean value of the strong coupling below some low energy infrared matching scale, conventionally chosen as $\mu_I = 2$ GeV. $\overline{\alpha_0}$ is defined by

$$\overline{\alpha_0}(\mu_I) = \frac{1}{\mu_I} \int_0^{\mu_I} \alpha_s(\mu_R^2) d\mu_R \tag{1.39}$$

where μ_R is the renormalisation scale, generally taken to be $\mu_R^2 = Q^2$. The value of μ_I is interpreted as the scale below which the perturbative approach is not valid. The general power correction from this approach is

$$\langle F \rangle_{pow} = a_F \frac{8\mathcal{M}\mathcal{A}_1}{\pi \mu_R} \tag{1.40}$$

Variable	a_F
τ_γ	1
τ_T	1
B_γ	See text
B_T	See text
ρ_0	1/2
C	$3\pi/2$

Table 1.1: a_F factors for the event shape variables.

which has the required $1/Q$ dependence[5]. The variable-dependent coefficient a_F is discussed below. The initial calculation of the theory was carried out to 1-loop accuracy only; this calculation had some ambiguity problems which were solved by extending the calculation to 2-loops. The difference between the calculations turns out to be a universal factor of $\mathcal{M} \simeq 1.49$, called the *Milan factor* [29]. The universal term from the renormalon calculation, $\mathcal{A}_1$, is given by

$$\mathcal{A}_1 = \frac{C_F}{\pi}\left[\overline{\alpha_0}(\mu_I) - \alpha_s(\mu_R) - \frac{\beta_0}{2\pi}\left(\ln\left(\frac{\mu_R}{\mu_I}\right) + \frac{K}{\beta_0} + 1\right)\alpha_s^2(\mu_R)\right] \quad (1.41)$$

where $C_F = 4/3$ and $K = 67/6 - \pi^2/2 - 5N_f/9$. N_f is the number of active quark flavours which is taken to be 5, appropriate for the Q^2 scales under consideration at HERA. β_0 is as defined in Equation 1.13. The constant, $\overline{\alpha_0}$, is used to replace the low-energy part of the mean shape variable while the remainder is the subtraction of the integral, up to μ_I, of the perturbative expression for the average shape. Above this limit, the perturbative expression is taken to be applicable.

[5]Under the assumption that the renormalisation scale $\mu_R = Q$. For convenience, the term 'Q-dependence' will be used throughout, since for all fits except the renormalisation scale determination, $\mu_R = Q$. For the renormalisation scale determination, $\mu_R \neq Q$ is used in the fits.

The constants a_F are given for each of the event shape variables in Table 1.1. A simple constant factor for the broadening with respect to the photon axis has been shown to be insufficient [30]. After a theoretical re-evaluation, a new, more complicated factor for the broadening has been proposed [31]

$$a_F = \frac{\pi}{2\sqrt{2C_F \alpha_{CMW}(\bar{\mu_R})}} + \frac{3}{4} - \frac{\beta_0}{12C_F} + \eta_0 + \mathcal{O}(1) \tag{1.42}$$

where $\eta_0 = -0.614$ is a constant of integration, $\bar{\mu_R} = \mu_R e^{-3/4}$, and β_0 has been evaluated with $N_f = 3$ appropriate for low energy hadronisation. In addition, there are x-dependent terms of $\mathcal{O}(1)$ which are not included in Equation 1.42 since they have not yet been calculated theoretically. These missing terms may be significant at low x where the incoming quark has some transverse momentum. The standard $\overline{\mathrm{MS}}$ renormalisation scheme relates $\alpha_s(\mu_R) \equiv \alpha_{\overline{\mathrm{MS}}}(\mu_R)$ to the more physical $\alpha_{CMW}(\mu_R)$ by the conversion factor

$$\alpha_{CMW} = \alpha_{\overline{MS}}(1 + K\frac{\alpha_{\overline{MS}}}{2\pi}) \tag{1.43}$$

The expression for broadening with respect to the thrust axis is similar to Equation 1.42 [31]

$$a_F = \frac{\pi}{2\sqrt{C_F \alpha_{CMW}(\bar{\mu_R})}} + \frac{3}{4} - \frac{\beta_0}{6C_F} + \eta_0 \tag{1.44}$$

noting that there are no x-dependent terms of $\mathcal{O}(1)$ for B_T.

In the fitting process the value of $\alpha_s(M_Z)$ is used as the input and is evolved to $\alpha_s(Q)$ using the expression

$$\alpha_s(Q) = \frac{\alpha_s(M_Z)}{1 + \alpha_s(M_Z) L^{(n)}\left(\frac{Q}{M_Z}\right)} \tag{1.45}$$

where the 2-loop form for $L^{(n)}$ is

$$L^{(2)}\left(\frac{Q}{M_Z}\right) = \left(\frac{\beta_0}{2\pi} + \frac{\beta_1}{8\pi^2}\right) \ln\left(\frac{Q}{M_Z}\right) \tag{1.46}$$

where β_0 and β_1, defined in Equations 1.13 and 1.14, are evaluated for 5 flavours, appropriate for scales at the mass of the Z^0.

In chapter 8, the corrected data will be used along with a prediction of $\langle F \rangle_{NLO}$ from DISENT and DISASTER++, to fit to Equations 1.38 and 1.40. The values of $\alpha_s(M_z)$ and $\overline{\alpha_0}(\mu_I = 2\text{ GeV})$ will be extracted and used to determine the validity of this theoretical approach.

1.8 Summary

The QPM and the theory of QCD were introduced with reference to DIS. The two $\mathcal{O}(\alpha_s)$ processes, QCD Compton and Boson-Gluon Fusion, were discussed with respect to the hadronic final state. The Breit frame and current region were introduced and the event shapes to be measured in the analysis were defined and discussed. NLO calculations of the event shapes were considered and the theory of power corrections was outlined. This theory will be fitted to the data in order to determine its validity.

Chapter 2

HERA and the ZEUS Detector

A brief introduction to HERA is given, outlining the relevant components. This is followed by a general overview of the ZEUS detector and a detailed description of the detector components used in this analysis.

2.1 HERA

The Hadron Elektron Ring Anlage (HERA) is the world's first electron-proton collider and was constructed at Deutsches Elektronen-Synchrotron (DESY). Its construction was completed in October 1991 and it has been taking data since June 1992. The 6336 m long accelerator is situated under the Volkspark in Hamburg, Germany, as shown in Figure 2.1. The HERA ring consists of two independent storage rings, one for 820 GeV protons, the other for 27.5 GeV electrons, yielding a centre-of-mass energy of $\sqrt{s} \simeq 300$ GeV. The two counter-rotating beams are brought into a single beam pipe and collided at two interaction points. Two general purpose experiments, H1 [32], in the DESY North Hall, and ZEUS [33], in the DESY South Hall, are positioned so as to observe the collisions at these interaction points. The remaining two experimental halls house two fixed target experiments. The HERMES experiment [34] uses the longitudinal polarisation of the electron beam to investigate the spin structure of nucleons, by inelastically scattering a fraction of the electron beam off polarised gas targets (e.g. hydrogen and deuterium). The HERA-B collaboration [35] is studying $\mathcal{CP}$-violation in the $B^0\overline{B^0}$ system. The B^0 mesons are produced by introducing wire targets (e.g. tungsten) into the proton beam halo.

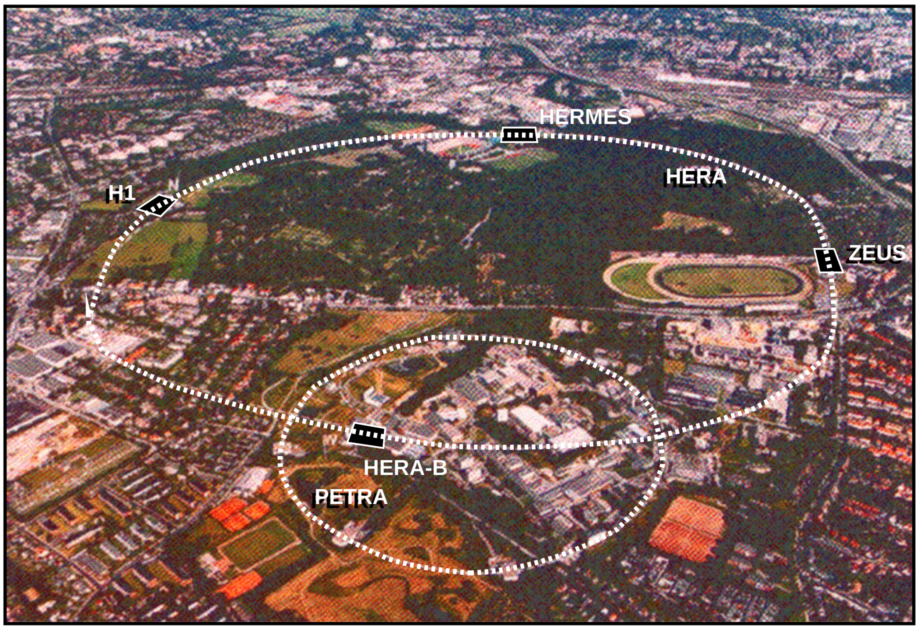


Figure 2.1: Aerial view of the Hamburg Volkspark showing the DESY site, the HERA and PETRA rings, and the four experiments.

2.1.1 Proton acceleration

H^- ions from a 50 MeV linear accelerator are stripped of their electrons and then injected into DESY III. 11 bunches with the final HERA beam spacing of 96 ns are accelerated to an energy of 7.5 GeV. The bunches from DESY

III are then injected into PETRA II where 70 bunches are collected and accelerated to 40 GeV. These are then ready for injection into HERA where 210 bunches are collected and accelerated to the final proton beam energy of 820 GeV. The beam is held in alignment and steered by liquid-helium cooled superconducting quadrupole magnets operating at a temperature of 4 K with a field of 4.65 T. A schematic of the HERA injection system is given in Figure 2.2.

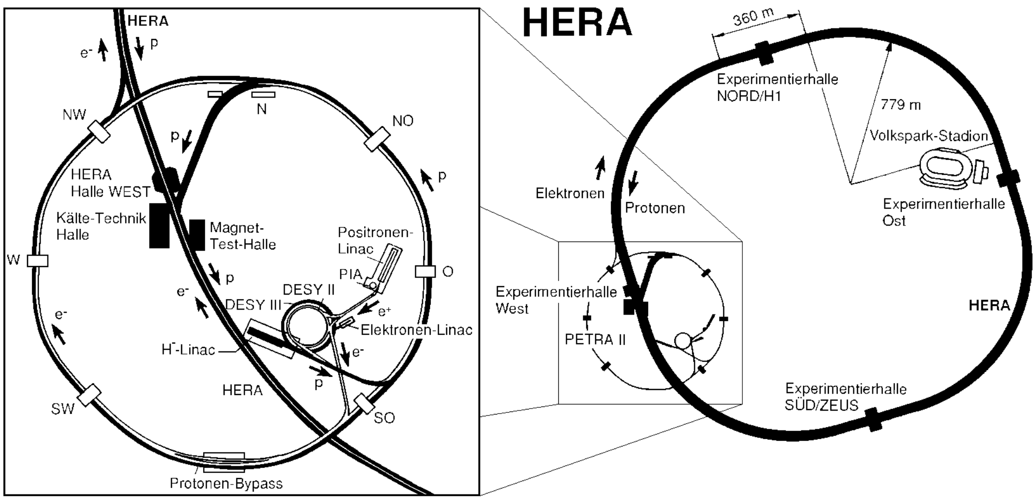


Figure 2.2: Schematic view of the HERA and PETRA accelerator rings. The enlarged section shows the injection system.

2.1.2 Positron acceleration

Positrons with an energy of 500 MeV are collected from a linear accelerator and stored in the PIA (Positron Intensity Accumulator) storage ring until a bunch of approximately 60 mA is accumulated. These are then injected into LINAC II and accelerated to 220 MeV whereupon they are transferred to another linear accelerator, LINAC III. After acceleration to 450 MeV, they are injected into the DESY II ring and accelerated to 7.5 GeV, for injection into PETRA. 70 bunches are accumulated in this way, with the final 96 ns HERA beam spacing, and accelerated to 14 GeV. Three batches of these 70 bunches are injected into HERA for acceleration to the final positron beam energy of 27.5 GeV. In contrast to the superconducting magnets for the proton beam, the positron beam is steered using magnets operating at normal temperatures with a field of 0.17 T. This is a relatively low field and minimises the synchrotron radiation losses from the electron beam. The maximum energy attainable by the positron beam is limited by the synchrotron radiation losses of the beam as it is steered around the ring. Figure 2.3 shows the delivered HERA luminosity from 1994-1997, a total of 800 days of running. The luminosity delivered per year increased as experience was built up in operating the machine.

2.2 ZEUS

2.2.1 Overview

The ZEUS detector was designed as a general purpose detector to study *ep*-scattering. The differing types of events studied at HERA imposed various constraints on the design of the detector:

- The asymmetry between the proton and electron momenta boosts the centre-of-mass (CMS) frame in the forward direction with respect to the laboratory frame and requires an asymmetric detector design.
- The key signature of a neutral current DIS event is the scattered DIS electron. The detector must be able to identify both the angle and

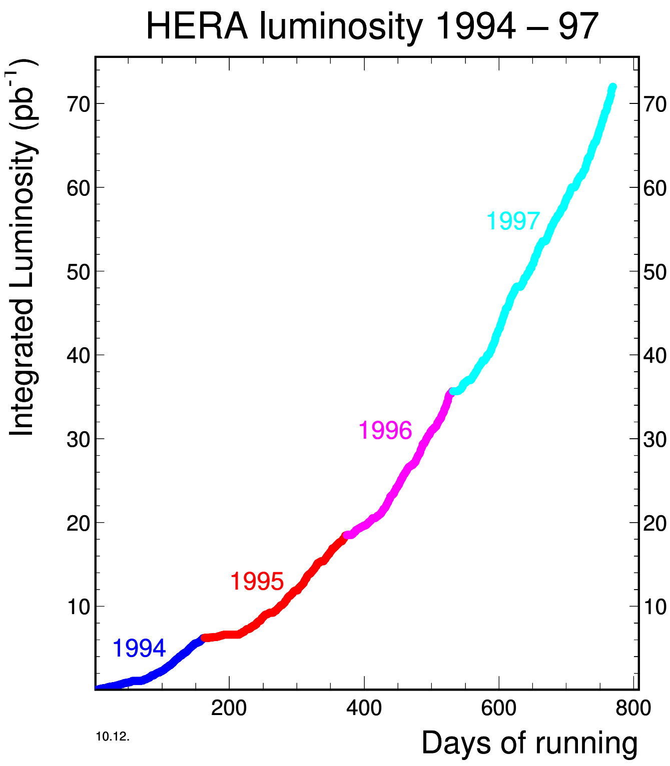


Figure 2.3: The total integrated luminosity delivered by the HERA machine for each year.

energy of the electron to a high degree of precision. This requires the calorimeter to have good electromagnetic energy resolution and the presence of a good tracking detector.

- Identification of the hadronic final state is very important and requires good hadronic energy resolution and sufficient segmentation to allow the jets to be accurately reconstructed. In this analysis, the hadronic final state is used, along with the electron, to precisely determine the kinematics of the events. This requires that the detector be hermetic, covering as much of the 4π solid angle as possible. For charged current analyses, where the outgoing neutrino is undetected, this is the only practical method of reconstructing the kinematics.

- A good tracking detector is required to help identify the scattered electron and to reconstruct precisely the event vertex. It will also improve identification of the hadronic final state. Tracks can be matched to calorimeter deposits and where a match exists, the information from the track can be used to augment the calorimeter information.

- A large magnetic field is required to measure accurately the momentum of high-momentum charged tracks. The field must be sufficiently contained so as not to interfere with the photomultipliers of the calorimeter which do not function correctly under large magnetic fields.

- The detector must be able to work in the high background conditions present in HERA running, being able to efficiently discard background events produced by the proton beam, and cope with the relatively short beam crossing of 96 ns. The design must also incorporate protection of the detector from the relatively high radiation proton background and to shield the tracking detectors from the synchrotron radiation produced by the electron beam.

- For the measurement of luminosity by the Bethe-Heitler process [36] and for the measurement of very low-Q^2 photoproduction events, a detector to tag very low angle electrons and photons is required.

The ZEUS coordinate system is a right-handed orthogonal system with its origin at the nominal HERA interaction point (IP). The positive z-axis points down the proton direction and is termed the *forward direction.* The positive X-axis points towards the centre of HERA, and therefore the positive Y-axis points upwards. The polar angle θ is measured with respect to the positive proton direction; the proton beam has a polar angle of 0° and the electron beam a polar angle of 180°. The *pseudorapidity* is a limiting expression for the rapidity for high energies and is given by

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \tag{2.1}$$

Figures 2.4 and 2.5 shows two cross-sections through the ZEUS detector, one in the yz-plane, parallel to the beam axis, the other in the xy-plane, perpendicular to the beam axis. The design displays a forward-backward

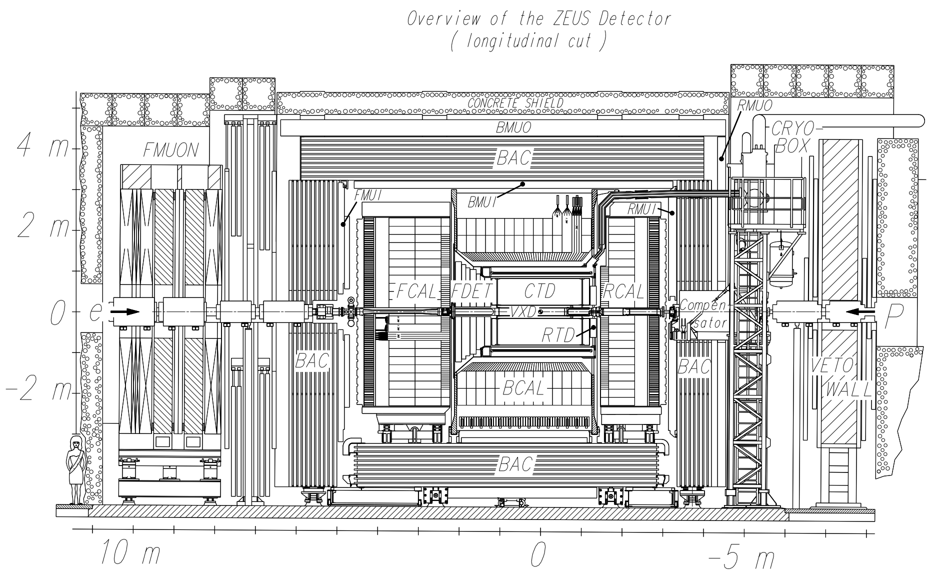


Figure 2.4: Schematic view of the ZEUS detector in the yz plane.

asymmetry and the 'onion-skin' structure of large collider experiments. A brief description of the overall detector is given here, with more detail on the major detectors actually used in the analysis; a fuller description is available in [33]. Describing ZEUS from the inside out:

- The innermost layer is the interaction region and the beam pipe. The

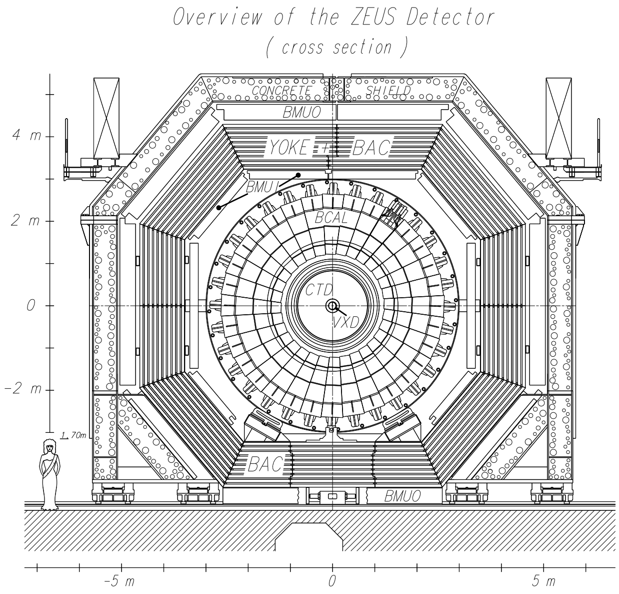


Figure 2.5: Schematic view of the ZEUS detector in the xy plane.

beam pipe is made of aluminium, approximately 170 mm in diameter, and contains a very high vacuum of the order of 10^{-8} torr.

- The vertex detector (VXD) was designed to provide a higher resolution vertex, and to find secondary decay vertices useful for the tagging of charm induced events. Unfortunately, the very high radiation levels near the IP damaged the detector and it was removed at the end of the 1995 run. Consequently, it was not used in this analysis.

- The tracking detectors are used to reconstruct charged tracks with high precision to determine their sign and momentum, and to identify the event vertex. The central tracking detector (CTD) is described in more detail below. There is also a forward tracking detector (FDET) and a rear tracking detector (RTD). The FDET comprises of 3 planar drift chambers (FTD) and a transition radiation detector behind them

(TRD). The TRD uses stacks of polypropylene fibre radiator followed by a drift chamber. The RTD comprises a single drift chamber of the same design as the FTD. In this analysis only the CTD is used to reconstruct tracks and the event vertex, so the other tracking detectors are not discussed further.

- The superconducting solenoid provides a magnetic field of 1.8 T for the inner detectors using an operating current of 5000 A. It is supplied with liquid helium at 4 K from the same cryogenic plant as the HERA superconducting magnets. There is a compensating superconducting solenoid behind the rear calorimeter, to correct for the influence of the magnetic field on the beams.

- The high resolution uranium calorimeter (CAL), which measures the energy of both hadronic and electromagnetic particles by absorption, is discussed in more detail below.

- The Hadron-Electron Separator (HES) is installed inside the FCAL and RCAL and consists of 3 cm $\times$ 3 cm silicon diodes to improve the position resolution of the calorimeter and help in distinguishing hadronic and electromagnetic showers.

- The Small Rear Tracking Detector (SRTD) is designed to improve the position resolution of low-angle scattered electrons. It is positioned on the front face of the RCAL around the beampipe and measures 68 cm $\times$ 68 cm. It consists of two planes of scintillator strips, which have a superior position resolution to that of the calorimeter, of the order of 5 mm in x and y.

- The yoke and Backing Calorimeter (BAC) is made of iron slabs and provides a return path for the solenoid magnetic flux. It is instrumented with proportional chambers and thus acts as a backing calorimeter allowing the measurement of high energy or late showering particles that have escaped the main calorimeter and would otherwise be lost.

- The muon detectors (FMUON, BMUON, RMUON) consist of interleaved sections of proportional counters, streamer tubes and time-of-flight (TOF) counters. Other than as a veto on cosmic induced events, these detectors are not used in this analysis.

- The Veto wall (VETO) is an iron wall behind the RCAL with dimensions 800 cm $\times$ 760 cm $\times$ 87 cm, with a small hole 95 cm^2 to allow the beam pipe through. It is instrumented with hodoscopes on both sides. Its purpose is to shield the detector against particles from the proton beam halo and to veto beam-gas induced events.

- The luminosity monitor does not form part of the main detector, but is a separate detector approximately 100 m downstream in the electron direction. The luminosity is monitored online and measured using Bremsstrahlung photons from the Bethe-Heitler process [36], $ep \rightarrow ep\gamma$. This cross section has been shown experimentally to agree extremely well with the QED calculation. In addition the radiative corrections to the process are known to be less than -0.3%, so that the cross section has very small uncertainties. Therefore, a measurement of the photon rate by the detector gives a very precise determination of the luminosity.

The majority of the electronics are not inside the detector. Instead, most of the readout and all of the trigger system is housed in a three-storey building (Rucksack) next to the detector. The output from the trigger is passed to the computer room in the South Hall and from there to the main DESY computer centre for storage on tape.

2.2.2 Central Tracking Detector

The central tracking detector [37] is a cylindrical drift chamber with an active length of 2 m covering the angular range $15^\circ < \theta < 164^\circ$. It surrounds the beam pipe (and vertex detector) with an internal radius of 18.2 cm and an external radius of 79.4 cm. It measures precisely the momentum and sign of charged tracks and allows the event vertex to be determined with a resolution of 1.4 mm in the z-direction. The design consists of 72 layers arranged into 9 superlayers as shown in Figure 2.6. The five odd numbered superlayers have sense wires strung parallel to the beam axis, while the four even numbered superlayers ('stereo layers') have the wires tilted at a small stereo angle of approximately $\pm 5^\circ$ with respect to the beam axis, allowing the determination of the z-coordinate of the hit with a resolution of $\sigma_z \simeq 1.4$ mm. The position resolution in the $r - \phi$ plane is approximately 180 μm for tracks that pass through all 9 superlayers. The chamber uses a gas mix in the ratio, Argon:CO_2:Ethane = 0.85:0.05:0.1 with trace amounts of ethanol, and has a drift velocity of approximately 50 μm/ns. The wires within each superlayer are arranged azimuthally into cells, each with 8 sense wires, and are designed so that the maximum drift length within a cell is 2.5 cm. The cells are inclined at an angle of 45° with respect to the radial direction, so that the electron drift lines are always perpendicular to high momentum tracks from the vertex, thus optimizing the resolution. This, along with the high drift velocity, ensures that tracks will always cross several sense wire planes so that at least one of the drift time measurements is less than the bunch-crossing time of 96 ns. In this way, the track can be unambiguously

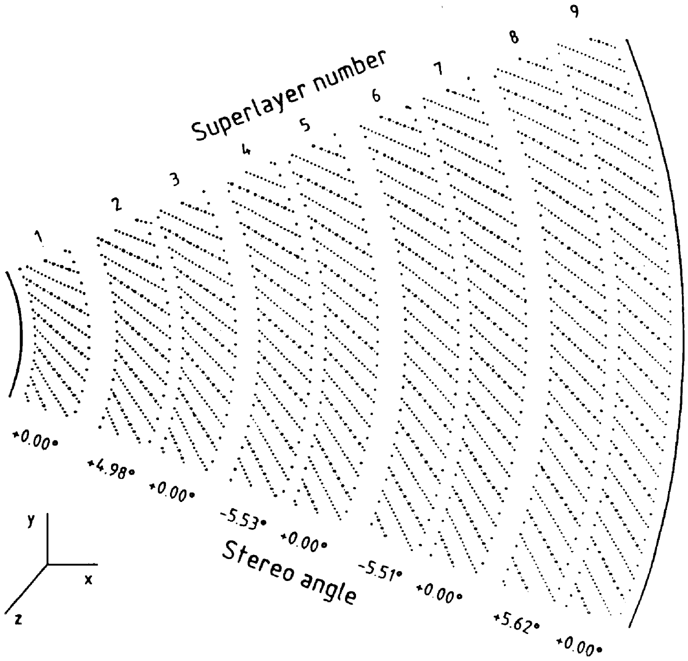


Figure 2.6: Wire layout of one octant of the Central Tracking Detector showing the 9 superlayers. The dots indicate sense and field wires going into the page; the larger dots are sense wires, the smaller dots are field wires. Alternate superlayers are inclined at a stereo angle to aid z-position reconstruction.

assigned to one particular bunch crossing. Sense wires in superlayers 1 and 3 and half of the wires on superlayer 5 are instrumented with a z-by-timing system [38], which compares the arrival time of the signal from both ends of the wires and allows fast determination of the z-coordinate of the track, with a precision of $\sigma_{zbyt} \simeq 3$ cm. The system is used in both the first and second level triggers. In 1996/1997, the momentum resolution of the CTD for a long (>3 superlayers) track was

$$\sigma(p_T)/p_T = 0.005 p_T \oplus 0.016 \tag{2.2}$$

with p_T measured in GeV. The first term is due to the intrinsic position resolution of the hits on a track and the second term is due to multiple Coulomb scattering which dominates at low p_T.

2.2.3 Uranium Calorimeter

The main calorimeter [39, 40, 41] encloses the inner components of the detector and is almost hermetic, covering 99.8 % of the solid angle in the forward direction and 99.5 % of the solid angle in the rear direction. It is a compensating sampling calorimeter with alternating layers of uranium as the absorber and plastic scintillator as the active material. The uranium is actually an alloy of 98.4% ^{238}U with 1.4% Nb to make the alloy harder; there is less than 0.2% ^{235}U. Hadronic showers interact (and often break up) the atoms of the absorbing material; this is particularly true for the neutrons in the hadronic shower, which suffer no Coulomb scattering. In calorimeters where the absorbing material has a medium atomic mass (eg. iron) this energy is generally lost and therefore the signal reaching the photomultipliers is generally less for a hadronic jet composed more of neutral particles than for an equal energy jet composed more of charged hadrons.

However, in the case of uranium or other unstable nuclei, these nuclear interactions with the neutral component of the hadronic shower cause the nuclei to break up producing charged particles, which can undergo elastic reactions with the hydrogen atoms in the scintillator material, returning some of the lost energy. By choosing the uranium layers to be 3.3 mm thick (1 λ_0 radiation length) and the plastic scintillator layers to be 2.6 mm thick, this scattering results in complete compensation, i.e. a hadronic jet of fixed energy produces the same signal in the calorimeter regardless of its charge composition.

The calorimeter is divided into 3 parts as shown in Figure 2.7:

- The forward calorimeter (FCAL), $2.2^\circ < \theta < 39.9^\circ$.
- The barrel calorimeter (BCAL), $36.7^\circ < \theta < 129.1^\circ$.
- The rear calorimeter (RCAL), $128.1^\circ < \theta < 176.5^\circ$.

where θ is the polar angle with respect to the proton beam direction. Each

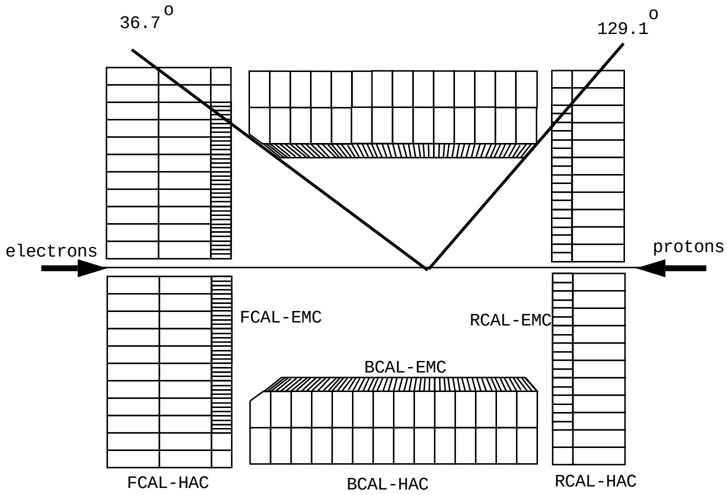


Figure 2.7: The relative position of the three parts of the calorimeter.

part is divided into 3 (2 for RCAL) sections. The inner section is called the electromagnetic calorimeter (EMC) and the outer one or two sections are called the hadronic calorimeter (HAC, or HACI and HACII). Electromagnetic showers generally develop and are absorbed faster than hadronic showers as shown in Figure 2.8. The EMC section is $25X_0$ in thickness for electromagnetic showers to ensure that the entire shower is fully contained in the EMC. This depth is equivalent to one interaction length, λ_0, for hadrons, so that hadrons deposit relatively little of their energy in the EMC. The thickness of the HAC section, in interaction lengths, varies from $6\lambda_0$ in FCAL to

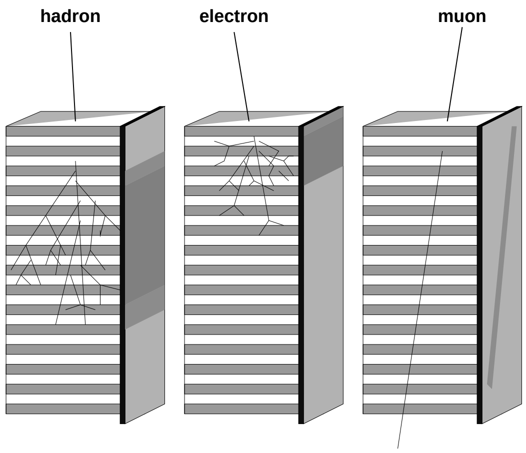


Figure 2.8: Typical shower developments for hadronic, electromagnetic and muonic particles in the calorimeter.

$3\lambda_0$ in RCAL. Each of the calorimeters is further divided into a number of modules, vertical in the FCAL and RCAL and radial in the BCAL. These measure 20 cm $\times$ 20 cm in the HAC sections, 10 cm $\times$ 20 cm in the EMC section of the RCAL and 5 cm $\times$ 20 cm in the EMC sections of the FCAL. Figure 2.9 shows a cut away view of an FCAL module and illustrates the way in which the module is read out. The wavelength shifter transports the light produced by the scintillator material to the two photomultiplier tubes. The difference in the signal heights from these two photomultipliers is used to reconstruct the position of the incident particle relative to the centre of the cell. In test beam conditions the resolution of the calorimeter was found to be

$$\sigma(E)/E = 0.35/\sqrt{E} \oplus 0.02 \quad \text{for hadronic showers} \tag{2.3}$$

and

$$\sigma(E)/E = 0.18/\sqrt{E} \oplus 0.01 \quad \text{for electromagnetic showers.} \tag{2.4}$$

with E measured in GeV. The calorimeter's performance is monitored daily

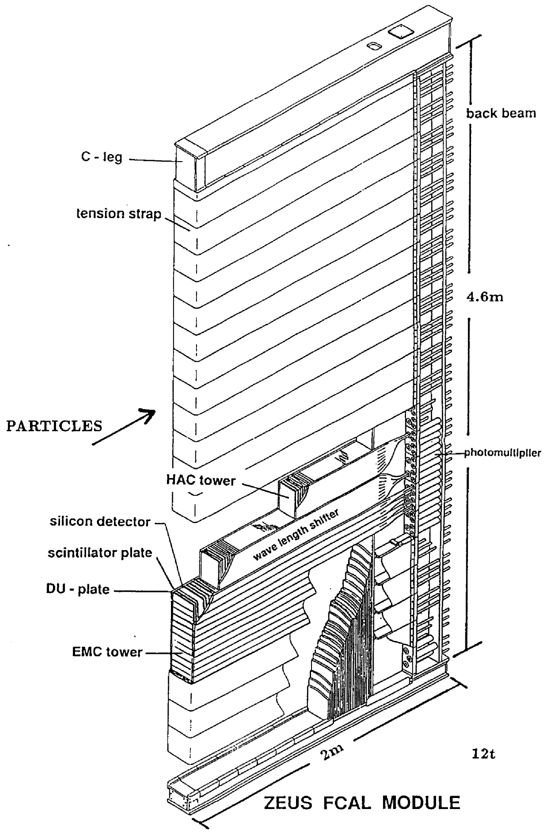


Figure 2.9: Cut-away view of an FCAL module.

using the natural radioactivity of the uranium which provides a very stable background signal. Further details of the tests and design of the calorimeter are available in [42].

The calorimeter also gives very precise timing information, with a resolution better than 1 ns, allowing rejection of beam-gas induced events by the trigger.

2.2.4 ZEUS Trigger

The HERA bunch crossing time of 96 ns corresponds to a crossing rate of 10 MHz which poses considerable challenges for the Data Acquisition (DAQ) system and trigger, given that each event has around 250,000 readout channels to process. The total interaction rate of HERA is dominated by interactions of the proton beam on residual gas in the beampipe; these dominate DIS events by an approximate factor of 100,000, with a rate of 10-100 kHz compared to a few Hz for interesting DIS events. Photoproduction events ($Q^2 \ll 1$ GeV^2) also have a much larger cross section than DIS events and form the other major background to this analysis. The other backgrounds to DIS are interactions of the electron beam on residual beam gas, cosmic induced events, and interactions produced in the proton beam halo. The task of the trigger is to pick out interesting physics events and reject as much of the background as possible. The final rate at which events are written to tape must not exceed a few Hertz. Since it is not feasible for a complicated decision to be made in 96 ns, the data from the components are *pipelined* into 58 bunch crossings, giving a total trigger time of 5.6 μs; special pipelining chips were designed for this [43]. The trigger system used in ZEUS consists of three levels as shown in Figure 2.10. The first level trigger (FLT) operates on each detector component independently; each component having its own pipeline, with a total decision time of 5.6 μs. Any component which displays some interesting feature, possibly from a physics event, will signal that event should be passed onto the second level trigger (SLT). The total rate being passed to the SLT is reduced to approximately 1 kHz. There is an SLT for each component, and since the rate is much lower, they have time to perform more complicated algorithms on the data set and also to do some limited combination of the data set. Any promising events are then

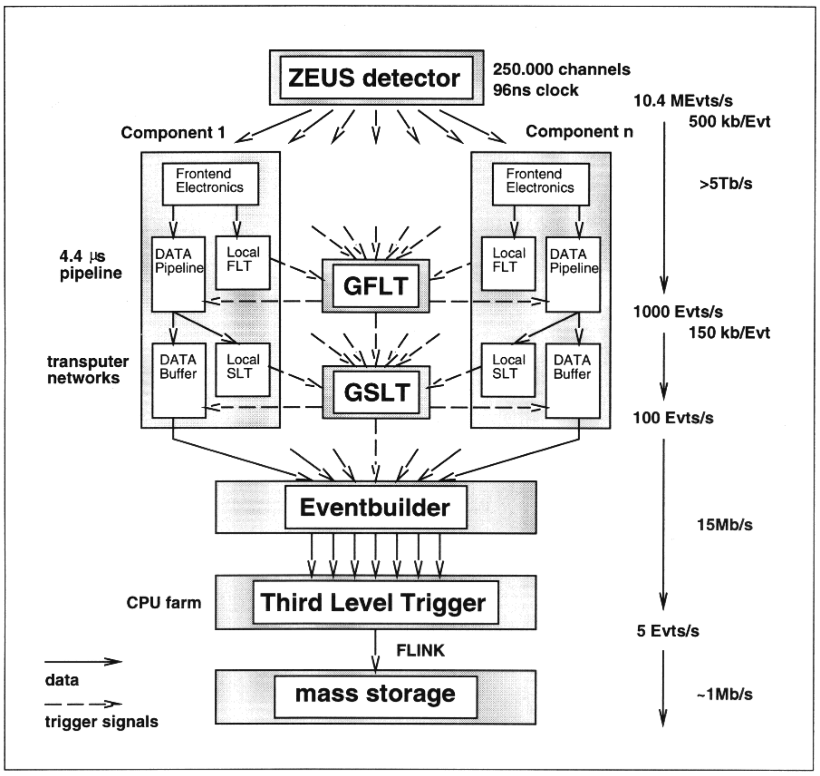


Figure 2.10: The structure of the ZEUS three-level trigger system.

passed to the event builder, where the data from differing components are combined together and put into ADAMO database format. The total rate leaving the event builder is around 100 Hz. These events are passed to the third level trigger (TLT) which can analyse the event as a whole and run complicated algorithms, such as jet-finders and proper kinematic reconstruction, on the data set. Following a positive decision from the TLT, the event is then passed to the DESY computer centre for storage on tape. The final event rate written to tape is 3-5 Hz.

2.3 Summary

The ZEUS detector is one of four experiments located on the HERA *ep* synchrotron at DESY, designed as a general purpose *ep* physics detector. The principal detectors used in this analysis are the CTD, a cylindrical drift chamber used to identify the vertex and hadronic final state, and the compensating uranium CAL, used in conjuction with the CTD to determine the hadronic final state. ZEUS uses a three-level trigger system to select DIS events and reject background from beam-gas interactions and photoproduction.

Chapter 3

Data Reconstruction and Data Quality Monitoring

In this chapter, the methods used to reconstruct the data are examined. The trigger preselection method for the data used in this physics analysis is presented. Work performed for the collaboration to ensure that only high quality data are used in analyses is also presented.

3.1 Offline Reconstruction

The data output from the TLT contains unprocessed information from all the detector components. The TLT has selected events that have some particular physical property, but the raw data as output by the TLT are not suitable for further detailed analysis. For example, only the individual hits in the CTD wires (several thousand in each event) are stored in the raw data; the full algorithm which fits all the tracks and finds the best event vertex must be run offline (there is a simplified, faster version which runs online for the TLT, but this is less accurate.) The purpose of the offline reconstruction program [44] (ZEPHYR) is to process and combine all the raw data and produce physical quantities and objects which are more suitable to physics analysis, applying all the relevant calibration constants to the detectors. It must also correct these quantities for any known deficiencies in the detector at the time the run was taken. For example if a particular CAL cell was malfunctioning, this can be masked out at reconstruction time, so that its faulty signal does not influence the physics results.

The TLT accepts an event using a logical OR of all the TLT bits, i.e. if any of the TLT bits fire, then the event is stored. Checking each of the TLT bits relevant to a particular analysis is quite cumbersome, because there are so many. Instead, the TLT bits are combined together with other information from the reconstruction, to form Data Summary Tape (DST) bits. These are more closely related to the physics under study and a typical analysis will select on one or more of these bits. The analysis presented here requires a logical AND of two DST bits:

- **DST09** *Electron*

  This bit requires that one of four different electron finding algorithms has found an electron candidate in the calorimeter, with $E_{electron} > 4$ GeV. These algorithms are run at reconstruction time, so this bit does not require any information from the TLT.

- **DST11** *Nominal Neutral Current*

  This requires that the total reconstructed $E - p_z > 30$ GeV ($E - p_z$ is defined and discussed in more detail in chapter 4.) This is a very loose cut to reject photoproduction events. The DST bit is also the most inclusive of the DIS DST bits, requiring a logical OR of 13 separate TLT bits which correspond to various properties of neutral current DIS events; if the event exhibits any of these properties then the bit is set so that the event can be analysed further.

All 13 TLT bits that are used by DST11 have vetoes so that they will not fire if there are clear sparks in the calorimeter, or if there is a clear halo muon signal in the detector. One of the 13 TLT bits takes the majority of the events; it requires only a low energy electron ($E_{electron} > 4$ GeV) and a box-cut around the beampipe of 12 cm $\times$ 6 cm to ensure the electron is well measured. In addition, all the TLT bits require certain timing constraints on the event

- The 'up-down' time, which is the difference in time between signals arriving from the top of the BCAL and signals arriving from the bottom of the BCAL, must be less that 8 ns. This is to reject cosmic events where the top part of the calorimeter will be hit by the cosmic particles first.

- The time difference between the signals from the RCAL and the signals from the FCAL must be less than 8 ns, to reject events which have come from proton beam background outside the detector.

- The timing is set so that particles resulting from interactions at the nominal ep beam crossing arrive at the calorimeter at time $t = 0$. The averaged time over all cells with energy greater than 4 GeV in the FCAL (RCAL) must not exceed 8 (6) ns. Again, this is to reduce contamination from background processes not occurring near the nominal interaction point, but with a reconstructed vertex from secondary interactions.

During an analysis run over all the data, the analysis routine can check each event for the relevant DST bits before any further processing is performed.

This permits a basic preselection, and ensures that the analysis program does not waste time analysing events which are clearly not of the desired class. To ensure that this preselection is not too strict and does not reject events that should be kept, a trigger check is performed using Monte Carlo. The efficiency is defined as

$$\text{DST Efficiency} = \frac{\text{Events which passed DST9 AND DST11}}{\left(\begin{array}{c}\text{Events accepted at detector level} \\ \text{without requiring DST9 AND DST11}\end{array}\right)} \tag{3.1}$$

The Monte Carlo study gives 100% indicating that the trigger fires for all the events that should be kept. A more rigorous data-based study and a study made by injecting signals directly into the front-end electronics has shown that the hardware trigger efficiency is higher than 99% for all regions of x and Q^2 [45].

3.2 Zeus Event Store

Although the use of DST bits help reduce the number of events processed, they suffer from two large drawbacks. The major one is that the data format in which the events are stored (ADAMO) requires that the entire event be loaded into memory before anything, even the DST bits, can be checked. There is a large time penalty for this, and since the vast majority of the events are rejected at the DST level, this process can be very inefficient. The second problem is that DST bits are rather crude. There are comparatively few of them, so they must be quite general, and, although better than the raw triggers, they are still somewhat abstracted from the actual physical quantities that are being studied and cut upon. A better solution is to generate, at reconstruction time, a large and varied set of observables that can be used for preselection to augment the existing DST bits. Such a system has been implemented in the ZEUS Event Store [46] (ZES) which is an Objectivity tag-database containing 236 variables which are computed using standard code, approved by the relevant analysis groups. The Objectivity database is independent from the ADAMO database in which the whole event is stored; it is considerably smaller (236 numbers rather than several thousand for a typical event) and so can be loaded into memory and processed much faster. The analysis presented in this thesis uses ZES to check the DST bits (since this is faster than loading the event from ADAMO) and also to make very rough preselection cuts on the event (the actual cuts are described in the next chapter.) Any promising events are then passed to the full analysis routine where tighter selection cuts are imposed. The use of ZES increases the analysis speed by over 80%. The ZES system is to be incorporated into a standard analysis package (ORANGE) for ZEUS which is currently under development.

3.3 Data Quality Monitoring

In ZEUS, physics analyses are divided into five separate working groups. Each group has responsibility for maintaining and monitoring the quality of their own set of triggers and their own set of DST bits, namely those corresponding to the physics interests of the group. The analysis presented here was performed in the QCD and Hadronic Final States (HFS) group. This group focuses primarily on the QCD evolution of the hadronic final state, jet physics and photon structure. During the 1999-2000 running period, the group had 25 DST bits, most of them having a one-to-one correspondence with the group's 27 TLT bits. Although the analysis presented here does not explicitly use the group's DST bits (it relies on basic DST bits from the Structure Function group and upon ZES), the analysis is very sensitive to any problems in the hadronic final state, so the HFS group's Data Quality Monitoring (DQM) is vital.

The DQM serves to check whether all the detector components and triggers are functioning correctly; any problems can then be identified and rectified quickly, so as not to waste time on taking poor quality data. The DQM also monitors whether there is any problematic non-*ep* background (such as proton beam gas) which could affect the group's physics.

For each HERA run, the DQM routines, which run in the main reconstruction program, histogram various quantities to be checked, and these can be compared to a set of reference histograms. Any runs found to have problems can either be excluded entirely from the analysis, or flagged for more careful analysis.

DQM checks are made on both the DST bit selections and on ZES. The QCD/HFS group is mostly concerned with the CAL, so the DQM focuses primarily on that part of the detector. The emphasis is on physical observables; there is also an independent detector DQM for the CAL concerned more with the direct output from the hardware components. This redundancy ensures that any subtle problems are more likely to be observed. The DQM for the CTD, which is also important for the analysis, is checked separately.

A number of general checks are made on the data. Firstly, it is important to minimise proton beam gas events, which generally have a large track activity parallel to the beam direction with not much transverse activity. Figure 3.1 shows the $E - p_z$ distributions, where E is the summed energy in the CAL and p_z is the summed longitudinal momentum. Beam gas tends to have a low $E - p_z$ of a few GeV whereas the DIS events of interest to this

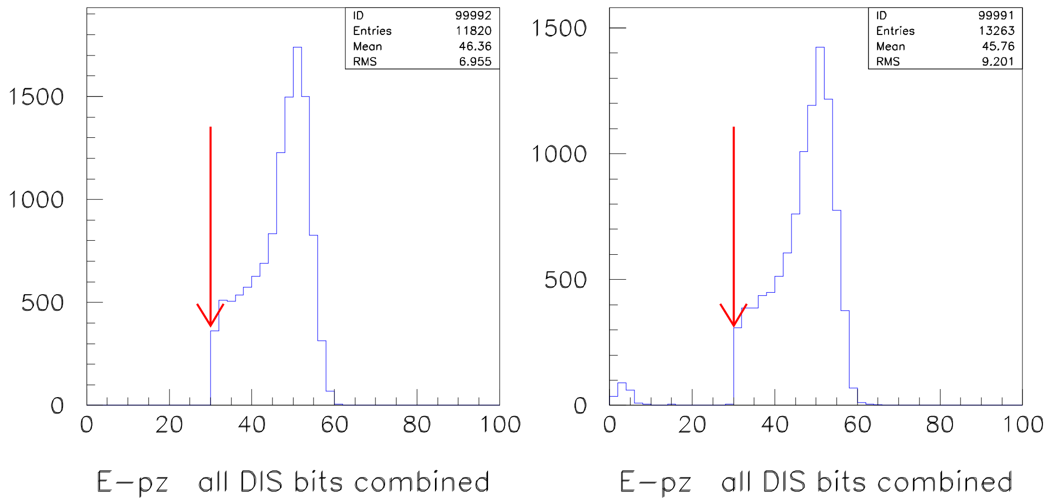


Figure 3.1: The $E - p_z$ distribution for all DIS triggers combined. a) on the left is normal, b) on the right indicates a problem with the trigger, where some events have been incorrectly accepted below the cut at low values of $E - p_z$.

analysis have a larger $E - p_z$. The trigger has a cut on $E - p_z > 30$ GeV for DIS triggers, and the distributions are examined to make sure that the trigger is functioning correctly and that the unwanted beam gas events are being rejected. Figure 3.1a) shows a normal $E - p_z$ distribution, whereas Figure 3.1b) indicates some malfunction of the DIS triggers, where a few low $E - p_z$ beam-gas events have been let through. A check is also made to see that the scattered DIS electron is detected where is should be in $E - p_z$, at about twice the electron energy (~ 55 GeV.) A peak anywhere else will indicate either a problem in the CAL or a serious problem in the reconstruction of $E - p_z$ which would make the run unusable.

The proton bunches are typically 12 cm long in the z-direction and the electron bunches less than 1 mm. Therefore, the actual event vertex where an interaction takes place may be a few cm away from the nominal interaction

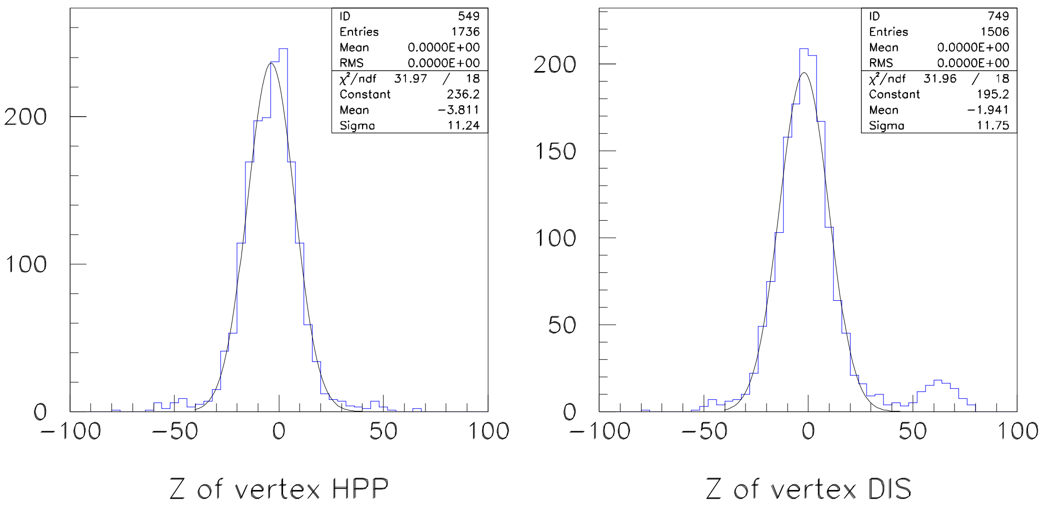


Figure 3.2: The z-position of the vertex for the HPP and DIS branch triggers. Note the small satellite bunch at $z = 60$ cm.

point of $x = y = z = 0$. For a DIS analysis, it is important to make sure that the event vertex has been found correctly, and that the triggers, some of which have a vertex requirement, have actually identified the correct vertex. We want to make sure that there are relatively few so-called 'satellite' bunches that can trail the main bunch by up to 80cm. Runs which have a high proportion of satellite bunches, although not unusable, should be analysed more carefully. Figure 3.2 shows the vertex distributions for photoproduction (HPP) and DIS triggers. The DIS triggers show a large satellite bunch, which is not observed in the photoproduction triggers because these triggers have an explicit vertex cut of $|z_{vtx}| < 50$ cm.[1]

The majority of analyses performed in the QCD/HFS group make use of triggers which identify 'jets' of hadrons in the CAL with particular properties,

[1]Figure 3.2 for the photoproduction trigger shows a few events outside the cut of $|z_{vtx}| < 50$ cm. This is due to the slight mismatch between online and offline vertex selection, similar to the problems discussed later in the chapter.

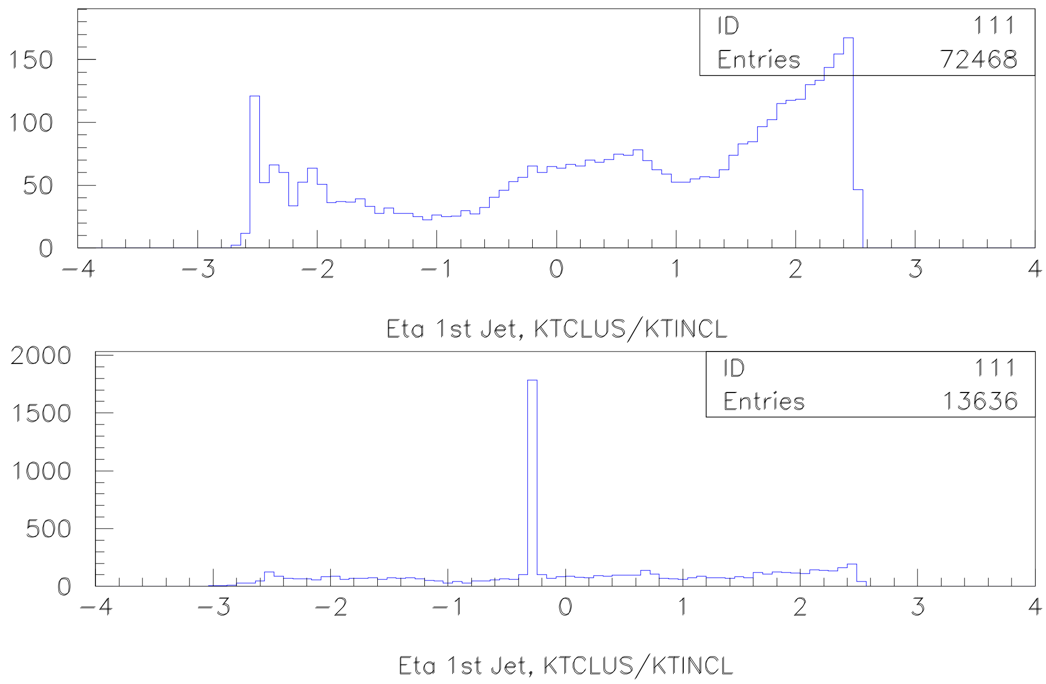


Figure 3.3: The upper plot shows a run with a normal η distribution, and the lower plot a run with a spike in the η distribution, indicating a problem in one of the CAL towers.

in order to study the hadronic final state. The jet triggers must run a jet-finder algorithm over the CAL and make a relatively fast decision (< 0.01 s), so it vital to check that they are working correctly and the CAL cells on which they base the decision are all in full working order. Each of the jet triggers are checked for a variety of properties and compared to a 'normal' run. This is done for both the DST bits and for the ZES database. Figure 3.3 is from the ZES DQM and shows the trigger cross section as a function of pseudorapidity η. The trigger uses one of the ZEUS-standard jet-clustering algorithms, and fires if any jets are found in the CAL. Figure 3.3(upper plot) shows a normal run with a lot of forward jets and some secondary jets in the BCAL at around $\eta = 0$. Figure 3.3(lower plot) shows a problem in the BCAL when one of the photomultiplier tubes was faulty, producing a constant signal ('sparking'). This was consistently misidentified by the jet algorithm as a hadronic jet, making the jet trigger unusable for this run, and any analysis of the hadronic final state impossible. If the data are subsequently reprocessed (i.e. passed through the reconstruction program again), then this faulty CAL cell can be masked out for the run and the data recovered, although the trigger bit will still be unreliable for the run.

During offline reconstruction, the jet and electron quantities that are used online by the TLT to make its decisions are reconstructed fully. Since the TLT has to make a very fast decision, it uses a cruder event reconstruction to save time. The DQM can be used to check decisions made by the TLT based on this crude method against the decision based on the more accurate fully reconstructed event record. Any mismatch between the two is flagged in special 'bug bins' in the DQM. For example, Figure 3.4 shows the pseudorapidity of the electron from the trigger that identifies prompt photon events. The mismatch is due to a less sophisticated electron finder being run at the TLT from that being run in the full reconstruction. The TLT imposes a cut of $-2 < \eta < 2$ on the pseudorapidity of the electron. There are some events measured by the TLT where the electron just falls within this range and will thus fire the trigger. However, when a full and accurate reconstruction is done, the measured pseudorapidity of the electron falls outside the desired range, and this mismatch is flagged by the DQM. Another example of this is shown in Figure 3.5 where the $x - y$ impact position of the DIS electron on the face of the RCAL is plotted. The box-cut at the TLT is shown. Several electrons can be seen to have shifted position inside the cut when the full reconstruction is made. This is not a serious problem for the analysis, but it is important that all analyses use the DQM to identify and understand the

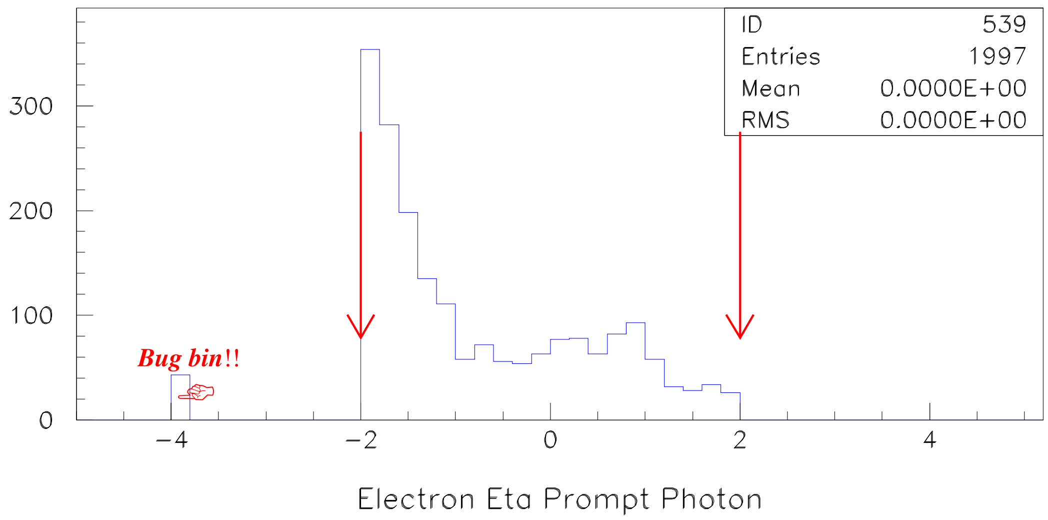


Figure 3.4: Mismatch in η between the TLT electron finders and the full reconstruction electron finders.

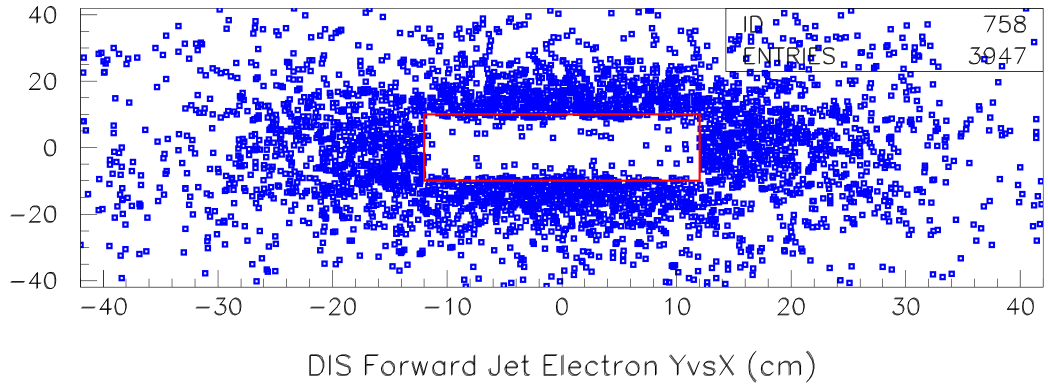


Figure 3.5: Limitations of the fast trigger reconstruction mean that the TLT box-cut is only approximate.

limitations of their trigger.

The final use of the DQM is to monitor the output rates of the various DST bits. These should remain stable with time as long as the trigger or the DST definition is not changed. Figure 3.6 shows the DST output cross-

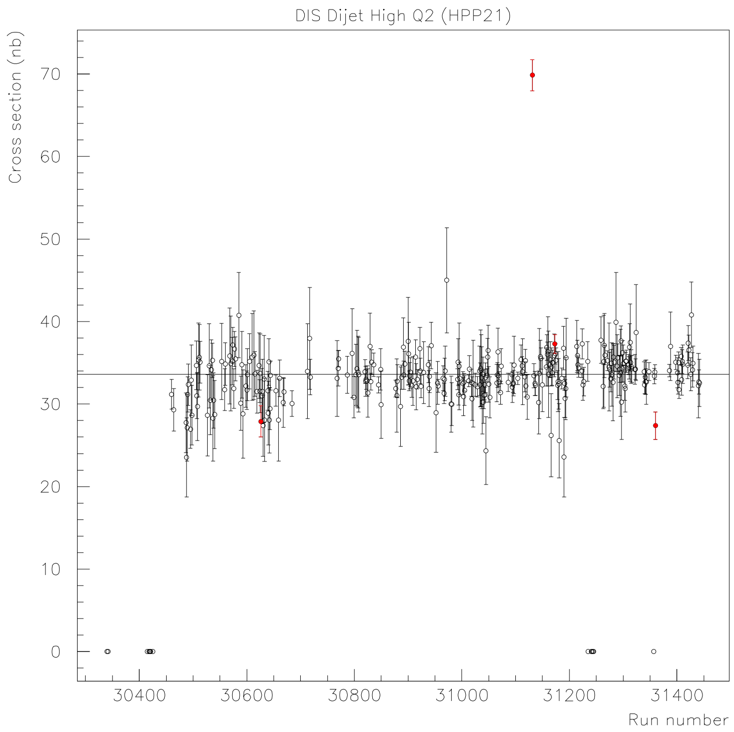


Figure 3.6: Output DST cross-section for the DIS dijet trigger. The solid points are flagged since they are more than 3 standard deviations from the mean.

section for the trigger which identifies DIS dijets. The DQM automatically collects the data and calculates a mean cross section based on the most up-to-date data. Any runs which are more than 3 standard deviations from the mean are flagged and can be investigated in more detail or excluded. The points at zero indicate a detector test when the trigger was switched off, and are not flagged by this process. Typically, a larger than average trigger cross-section is caused by a sparking CAL cell being consistently misidentified as a hadronic jet or a DIS electron. Any computer or hardware problems in the data acquisition system can cause large dead-times and result in a lower than average measured cross-section.

The runs are checked as soon as they come through reconstruction, so that any problems in the data taking can be rectified as soon as possible. This procedure, in conjunction with the individual detector component DQM information, ensures that only high quality data are used in the group's analyses.

3.4 Summary

The data used in the analysis is preselected using loose trigger requirements primarily based on the presence of an isolated DIS electron with some vetoes to limit background events. The trigger efficiency is high. Data quality monitoring procedures have been developed to ensure that the hadronic final state of the events can be correctly reconstructed. This work has led to an understanding of the limitations of the triggers used in the analysis.

Chapter 4

Kinematic Reconstruction and Event Selection

Different methods for reconstructing the kinematic variables are evaluated in terms of their resolutions and by comparison with simulated Monte Carlo. The methods used to find and measure the scattered DIS electron and hence make the boost to the Breit Frame are discussed. The cuts imposed to remove contamination from background and make sure the event is well measured are presented. Finally, the choice of kinematic analysis bins is given, along with the relevant analysis efficiencies and purities.

4.1 Kinematic Reconstruction

Several methods are available at ZEUS for the reconstruction of the kinematic variables x, y, and Q^2 [47]. The methods discussed make use of the measured angles and energies of the scattered DIS electron, (θ'_e and E'_e), those of the current jet, (γ_h and E_h), or a combination of both. The angles θ'_e and γ_h are defined in Figure 4.1.

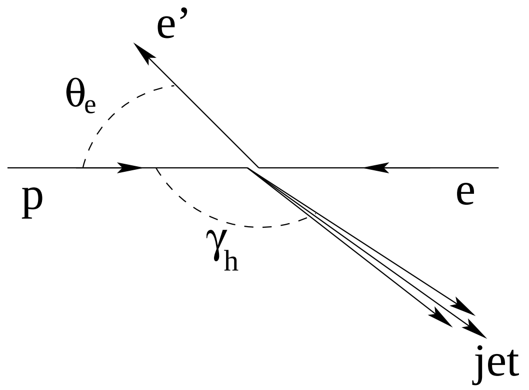


Figure 4.1: The angles θ_e' and γ_h defined with respect to the incoming proton direction.

Isolines of these quantities in the $x - Q^2$ plane are shown in Figure 4.2. These isolines, combined with the intrinsic resolution on each of the quantities provide an understanding of the various methods in different parts of the $x - Q^2$ plane. They help determine how measurement errors in the four variables (θ_e', E_e', γ_h and E_h) will affect the reconstructed values of x and Q^2. Areas on the plane where there are dense isolines indicate that any measurement error on that quantity will lead to a small error on the reconstructed values of x and Q^2. Conversely, the intrinsic resolution is worse in regions where the isolines are sparse, since a small measurement error encompasses a large area in x and Q^2.

Since only two of the three kinematic variables are independent, in the following, only the equations for Q^2 and y are given, with x given by

$$x = \frac{Q^2}{ys} \tag{4.1}$$

where s is the HERA centre-of-mass-energy ($\sqrt{s} \simeq \sqrt{4E_eE_p} \simeq 300$ GeV.) Here, it is assumed that the rest mass of the proton and electron are much smaller than the energies involved and can therefore be neglected.

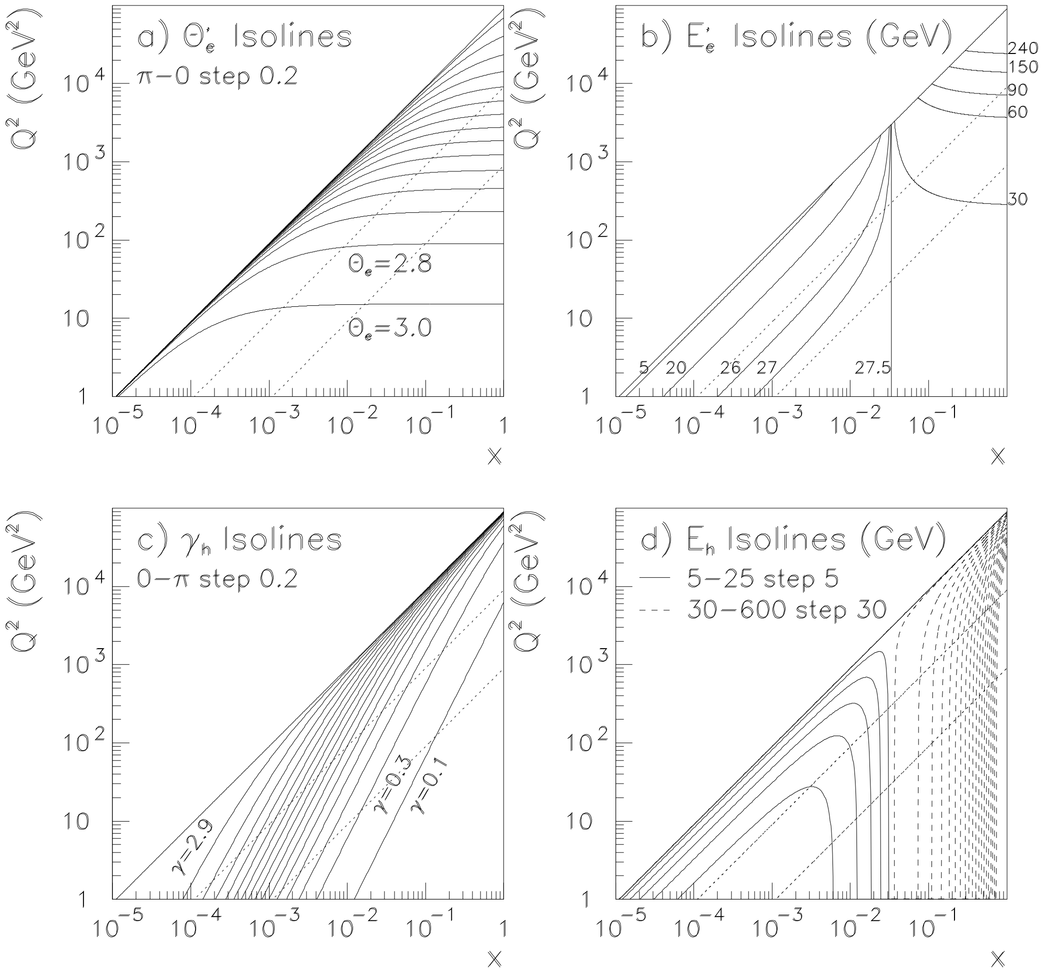


Figure 4.2: Isolines in the $x-Q^2$ plane, for a) the scattered electron angle θ_e', b) the scattered electron energy E_e', c) the hadronic system angle γ_h, and d) the hadronic system energy E_h. The step sizes between the lines are given. The dotted diagonal lines are the values of y for 0.1 and 0.01.

4.1.1 Electron Method

The most straightforward of the reconstruction methods uses the information from the scattered electron only. It uses the energy of the scattered electron, E'_e, and its angle, θ'_e, to determine the kinematics.

$$y_e = 1 - \frac{E'_e}{2E_e}(1 - \cos\theta'_e) \tag{4.2}$$

$$Q^2_e = 2E'_e E_e(1 + \cos\theta'_e) \tag{4.3}$$

where E_e is the incident electron beam energy. Figures 4.2a) and 4.2b) show that this has good resolution in the high y region where the isolines for the electron quantities are dense. However, at low y ($y < 0.1$), where the scattered electron energy is very close to the incident beam energy, the isolines are very sparse indicating poor resolution. Figure 4.3 shows the Q^2 resolution in the chosen analysis bins[1], determined from ARIADNE Monte Carlo. The distribution used to determine the resolution is given[2] by

$$\frac{Q^2_{true} - Q^2_e}{Q^2_{true}} \tag{4.4}$$

The RMS of a Gaussian fit to this distribution is termed the fractional resolution (or just 'resolution'), while the offset of the peak from zero is termed the fractional bias. For the electron method, the Q^2 resolution is between 7-10% depending on the $x - Q^2$ bin.

4.1.2 Jacquet-Blondel Method

The Jacquet-Blondel method [48] uses only the hadronic information in the detector (ie. in practical terms, it uses all the cells in the calorimeter not

[1] The choice of these 16 analysis bins is discussed in section 4.2.

[2] The same definition is used for all resolutions discussed in this chapter.

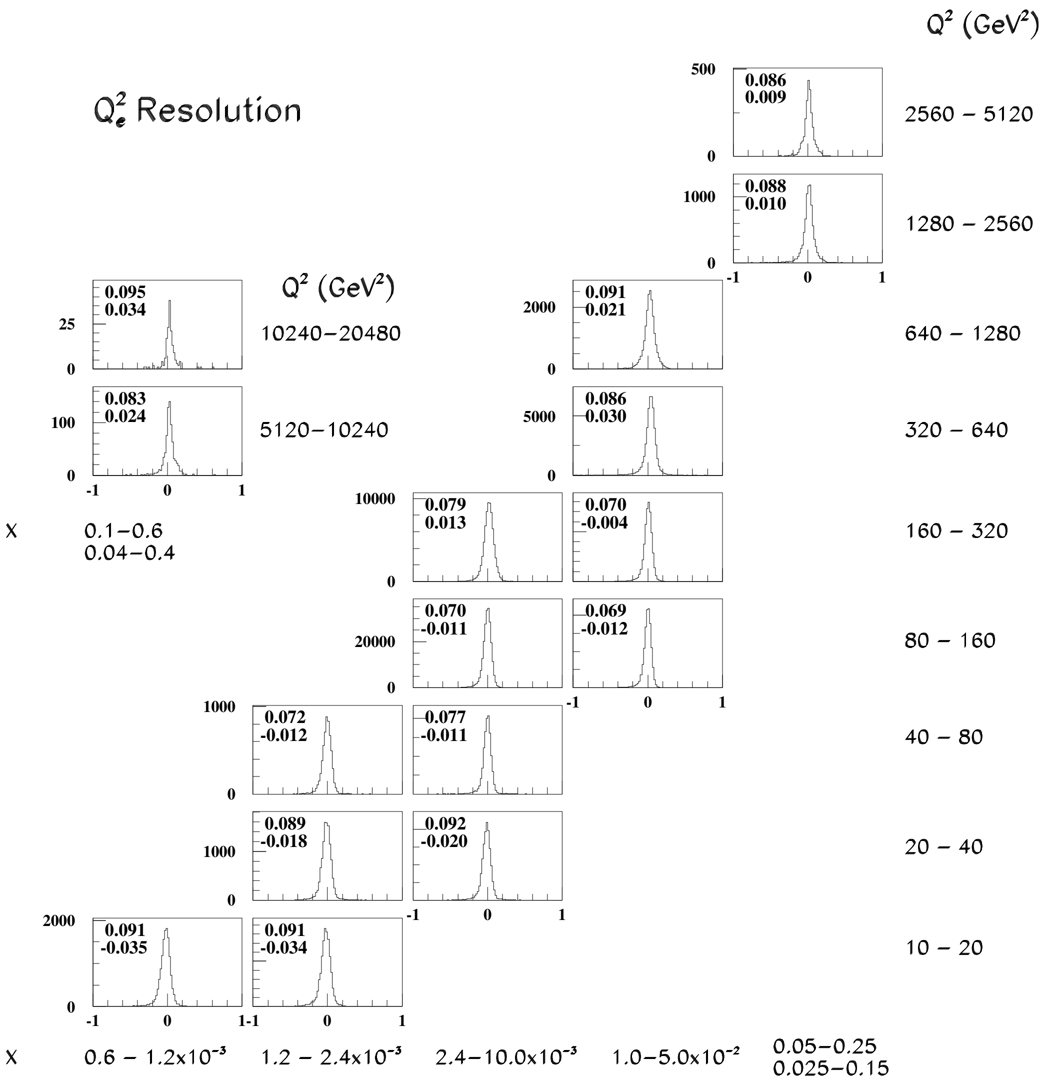


Figure 4.3: Resolution of Q_e^2 as a function of x and Q^2. The top number in each plot is the fractional RMS resolution, the lower number is the fractional bias.

associated with the DIS electron.) It is the only suitable method for charged-current analyses, where the outgoing neutrino is not detected and only the hadronic system is available. The hadronic system is characterised by two energies and one angle

$$\delta_h = \sum_h (E_h - p_{z,h}) \tag{4.5}$$

$$p_{T,h} = \sqrt{\left(\sum_h p_{x,h}^2\right) + \left(\sum_h p_{y,h}^2\right)} \tag{4.6}$$

$$\cos\gamma_h = \frac{p_{T,h}^2 - \delta_h^2}{p_{T,h}^2 + \delta_h^2} \tag{4.7}$$

This method relies on the fact that the detector is almost hermetic (therefore p_T is conserved) with the assumption that the p_T of undetected hadrons lost down the beampipe is negligible. Then

$$y_{JB} = \frac{\delta_h}{2E_e} \tag{4.8}$$

$$Q_{JB}^2 = \frac{p_{T,h}^2}{1 - y_{JB}} \tag{4.9}$$

The resolution of y_{JB} at very low y ($y < 0.05$) is generally better than y_e, however, the resolution of Q_{JB}^2 is poorer than Q_e^2 over the whole kinematic range, as shown in Figure 4.4. It ranges from 40% in the lowest $x - Q^2$ bins to 18% in the highest $x - Q^2$ bins. Additionally, since the method depends on an accurate measurement of the energy of the hadronic system, energy lost in inactive material in front of the calorimeter results in a very biased measurement, with the reconstructed values being on average 30% less than the true value for Q^2.

4.1.3 Double Angle Method

The Double Angle (DA) method [47] determines the kinematics from the angle of the scattered DIS electron, θ_e', and the angle of the hadronic jet γ_h,

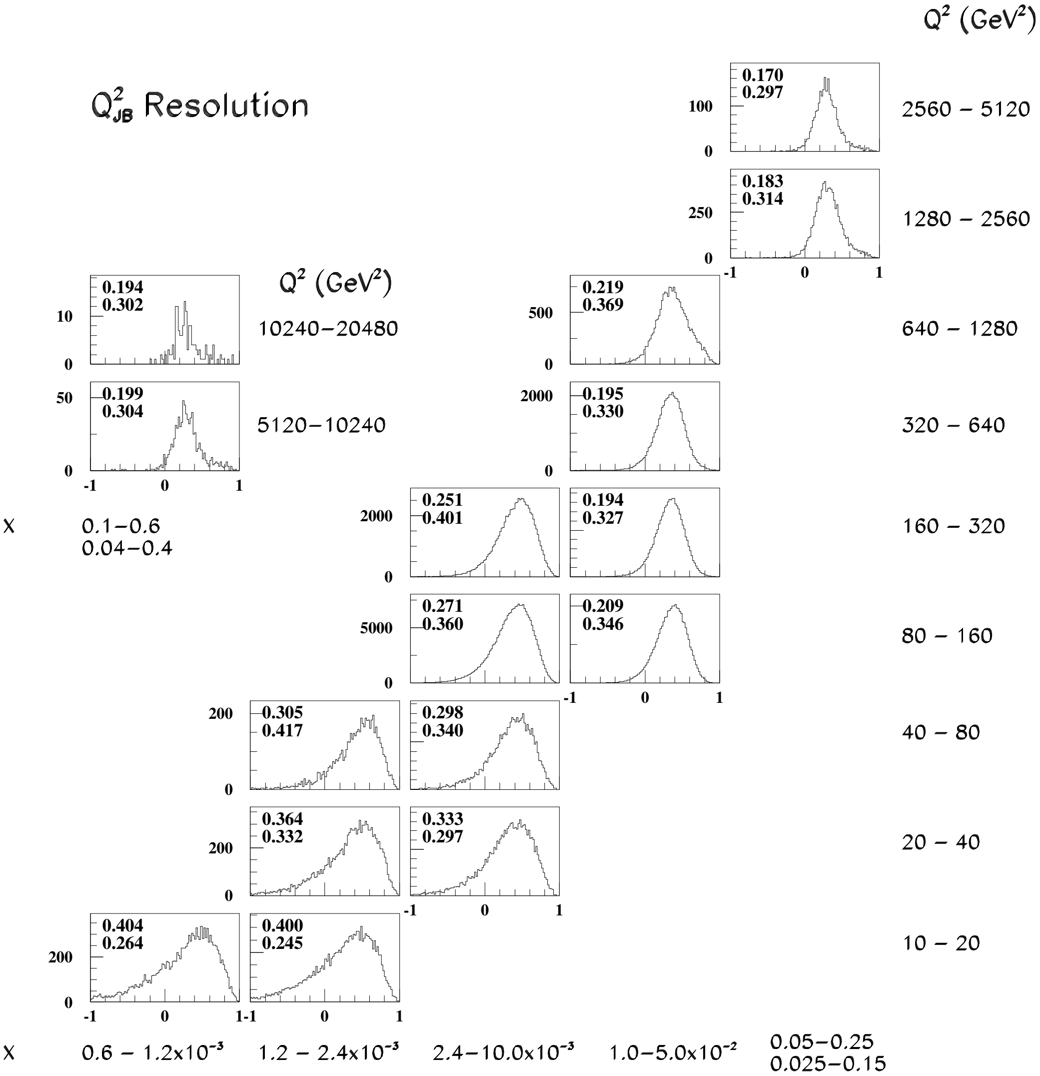


Figure 4.4: Resolution of Q^2_{JB} as a function of x and Q^2. The top number in each plot is the fractional RMS resolution, the lower number is the fractional bias.

leading to

$$y_{DA} = \frac{\sin\theta'_e(1-\cos\gamma_h)}{\sin\gamma_h + \sin\theta'_e - sin(\gamma_h+\theta'_e)} \tag{4.10}$$

$$Q^2_{DA} = 4E_e^2 \frac{\sin\gamma_h(1+\cos\theta'_e)}{\sin\gamma_h + \sin\theta'_e - sin(\gamma_h+\theta'_e)} \tag{4.11}$$

This method has the considerable advantage that it depends only on the angles, so that uncertainties in the hadronic energy scale do not adversely affect the reconstructed variables. The combination of Figure 4.2a) for θ'_e and Figure 4.2c) for γ_h would have dense isolines over a wide range in the $x - Q^2$ plane, indicating that this method has wide applicability. Figure 4.5 shows the resolution for Q^2_{DA} which is approximately a constant 10 - 11% over the entire kinematic range. The methods used to obtain the optimal θ'_e and γ_h are discussed in sections 4.3 and 4.7.5 respectively.

4.1.4 Summary of Reconstruction Techniques

Examining the Q^2 resolution alone would indicate that the electron method is superior to the DA method. However, it is also important to reconstruct x accurately. Figures 4.6 and 4.7 show the resolutions for x_e and x_{DA} respectively. The resolution for x_e varies from 40% at low Q^2 to 12% at highest Q^2 with approximately a 5-10% bias due to energy losses of the electron in dead material in front of the calorimeter In contrast, the x_{DA} resolution varies from 30% at low Q^2 to 11% at highest Q^2. In the intermediate range the resolution is twice as good as the electron method and the bias is less.

Figure 4.8 shows the comparison of data with Monte Carlo for $\log_{10}(Q^2)$ and $\log_{10}(x)$ for $Q^2 > 100$ GeV^2. The representation of Q^2 by the Monte Carlo is good for both methods. The DA method is marginally better at low x and at high x. Because of this, and because of the superior resolution on x, the DA Method is used to reconstruct the (x, Q^2) kinematics of an event.

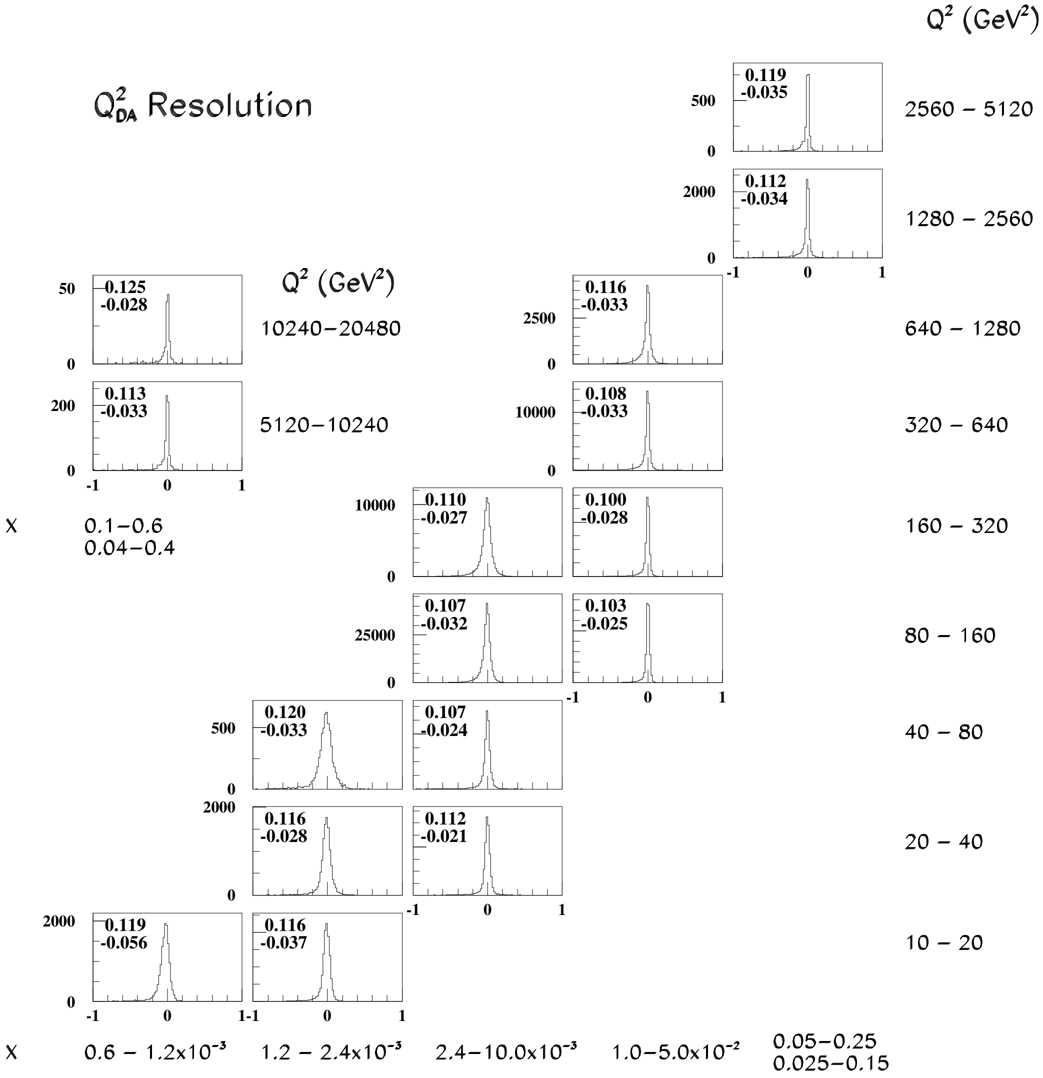


Figure 4.5: Resolution of Q^2_{DA} as a function of x and Q^2. The top number in each plot is the fractional RMS resolution, the lower number is the fractional bias.

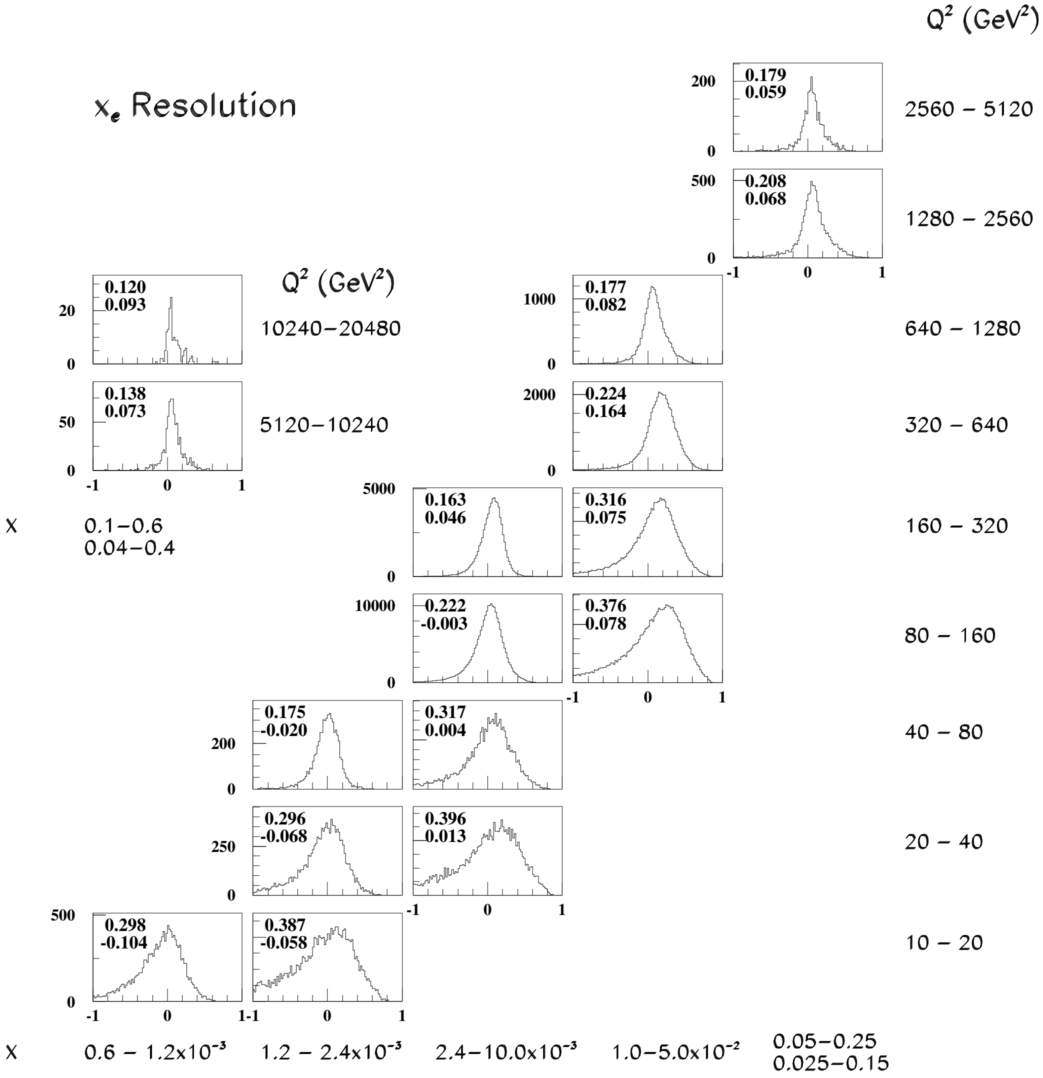


Figure 4.6: Resolution of x_e as a function of x and Q^2. The top number in each plot is the fractional RMS resolution, the lower number is the fractional bias

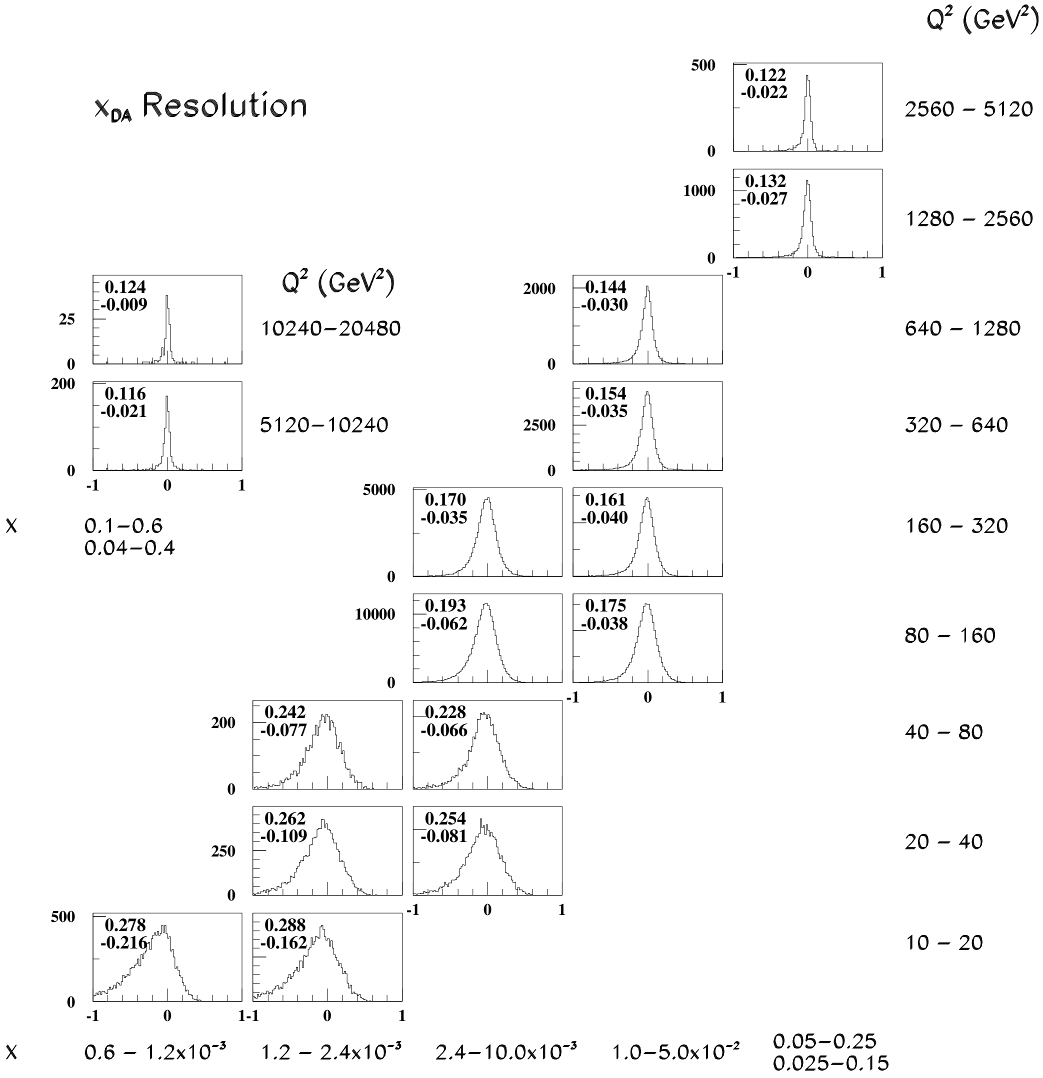


Figure 4.7: Resolution of x_{DA} as a function of x and Q^2. The top number in each plot is the fractional RMS resolution, the lower number is the fractional bias

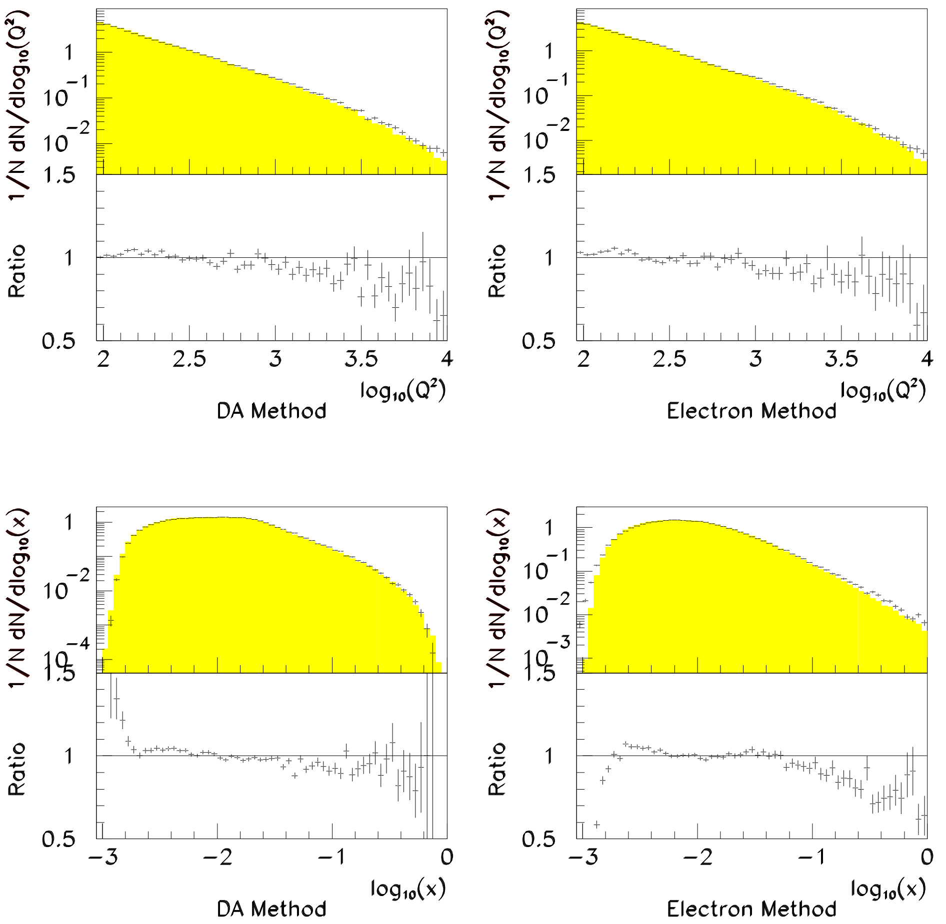


Figure 4.8: Comparison between data and Monte Carlo (ARIADNE) for $\log_{10}(Q^2)$ and $\log_{10}(x)$ for the DA and electron methods. The lower half of each plot shows the ratio of data to Monte Carlo.

The electron method is used as a cross-check, as discussed in chapter 7.

4.2 Selected DIS sample

The kinematic plane for the selected DIS sample is shown in Figure 4.9, and is divided into 16 analysis bins as defined in Table 4.1. The analysis

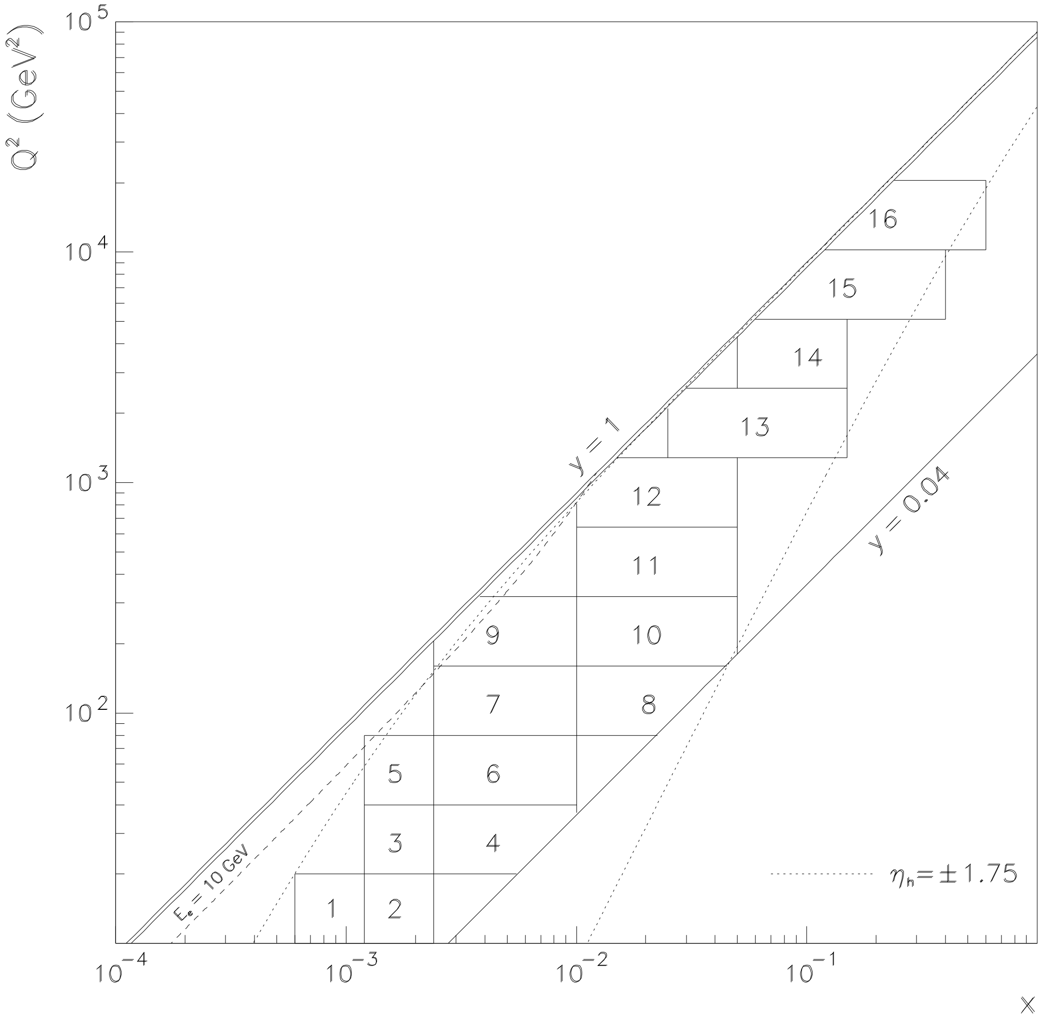


Figure 4.9: The 16 bins in x and Q^2 chosen to measure as much of the kinematic plane as possible.

bins were chosen so as to measure as much of the kinematic plane as possible, while keeping the measured hadronic system within the central range of pseudorapidity ($|\eta| < 1.75$) where it is more accurately measured. The bin sizes were chosen so as to maximise the purity and efficiency in the kinematic bins. The purity and efficiency were determined from ARIADNE Monte Carlo and

Bin	Q^2 range (GeV)	x range	Efficiency(%)	Purity(%)	# events
1	10 - 20	0.0006 - 0.0012	35.8	47.0	25978
2	10 - 20	0.0012 - 0.0024	35.2	47.9	24153
3	20 - 40	0.0012 - 0.0024	51.4	55.8	19261
4	20 - 40	0.0024 - 0.01	58.4	67.5	26862
5	40 - 80	0.0012 - 0.0024	52.0	61.5	9312
6	40 - 80	0.0024 - 0.01	68.9	76.3	20248
7	80 - 160	0.0024 - 0.01	69.2	81.4	60376
8	80 - 160	0.01 - 0.05	67.7	79.2	38405
9	160 - 320	0.0024 - 0.01	56.3	82.4	18542
10	160 - 320	0.01 - 0.05	74.6	84.3	27500
11	320 - 640	0.01 - 0.05	76.7	85.6	13102
12	640 - 1280	0.01 - 0.05	77.8	85.4	5323
13	1280 - 2560	0.025 - 0.15	80.1	86.0	2437
14	2560 - 5120	0.05 - 0.25	77.3	86.5	771
15	5120 - 10240	0.06 - 0.4	79.4	84.9	284
16	10240 - 20480	0.1 - 0.6	80.1	79.2	54

Table 4.1: The (x, Q^2) analysis bins. The event efficiencies, purities and number of events passing all the selection cuts are given. Bins 1-6 are measured using 1995 data only. Bins 7-16 are measured using 1996-1997 data.

defined as

$$\text{Purity} = \frac{\#\text{ events generated } \textit{and} \text{ correctly reconstructed in the } (x,Q^2)\text{ bin}}{\#\text{ events reconstructed in the } (x,Q^2)\text{ bin}} \tag{4.12}$$

$$\text{Efficiency} = \frac{\#\text{ events generated } \textit{and} \text{ correctly reconstructed in the } (x,Q^2)\text{ bin}}{\#\text{ events generated in the } (x,Q^2)\text{ bin}} \tag{4.13}$$

The events are assigned to a kinematic bin based on the reconstructed values of Q^2_{DA} and x_{DA}. At the hadron level in the Monte Carlo, the events are binned according to Q^2_{true} and x_{true}. Both the efficiency and purity are low (around 30% - 40%) in the low Q^2 region. However, for $Q^2 > 80$ GeV^2, the efficiency levels off at around 70% - 80% and the purity at 80% - 90%.

For the low Q^2 bins (bins 1-6) 1995 ZEUS data were used, with a total integrated luminosity of 6.6 pb^{-1}. For the higher Q^2 (bins 7-16), data from the 1996 - 1997 run period were used, with a total integrated luminosity of 38.6 pb^{-1}. Table 4.1 also shows the total number of events accepted after all the selection cuts have been imposed. The selection cuts are defined later in the chapter.

4.3 Electron Finding

The electron finding algorithm [49] used is based on a feed-forward neural network which has been trained using Monte Carlo samples [50]. The smallest geometrical unit available in the calorimeter is a CAL tower (a longitudinal combination of HAC and EMC modules). There are too many of these to practically work with, so towers which have some deposit above the uranium background are first clustered into more complex objects called *tower islands*.

The energy and positional information from all the photomultipliers making up each tower island are transformed and rescaled into a series expansion of Zernike moments as described in [50]. This series is then truncated, keeping only the important terms, which leads to 17 variables describing the energy distribution in the HAC and EMC tower islands; this is to be compared with a standard basic electron finder which will use fewer variables (e.g. only 2, the shower radius and fraction of energy deposited in the EMC section of the calorimeter.)

The neural network algorithm operates on all 17 of these transformed variables, and to each island assigns a probability between 0 and 1 that the island was produced by a single electromagnetically showering particle. The closer this probability is to 1, the more likely the island is to have come from an electromagnetically showering object.

The cuts below ensure the selection of a clean electromagnetic object.

1. **Probability > 0.9.** The probability assigned to the electron candidate must be greater than 0.9. This ensures a purity of greater than 98% in the electron sample [50]. If there is more than one island in the calorimeter with a probability greater than 0.9, then the one with the highest probability is taken.

2. **$\mathbf{E_{tower}}$ > 10 GeV.** The algorithm is optimally efficient (almost 100%) above electron energies of 10 GeV [50]. Around the trigger cut of 4 GeV the algorithm is only ~50% efficient and the Monte Carlo fails to describe the behaviour. In order to ensure optimal selection, a cut of $E_e > 10$ GeV is required for the final analysis.

3. **Matched track** If the candidate is in the BCAL, or in the RCAL (FCAL) with a positional radius of greater than 50 (70) cm, then it is required that a vertex-originating track in the CTD points to the electromagnetic deposit; the extrapolation of the track onto the calorimeter face must be within 10 cm of the centre of the deposit. The radius in the F/RCAL cuts correspond to the good acceptance region for tracks in the CTD (polar angles $0.3 < \theta'_e < 2.8$). Outwith this region, tracks cannot be reliably reconstructed, so no match is required. This requirement limits contamination from high energy photons and π^0s.

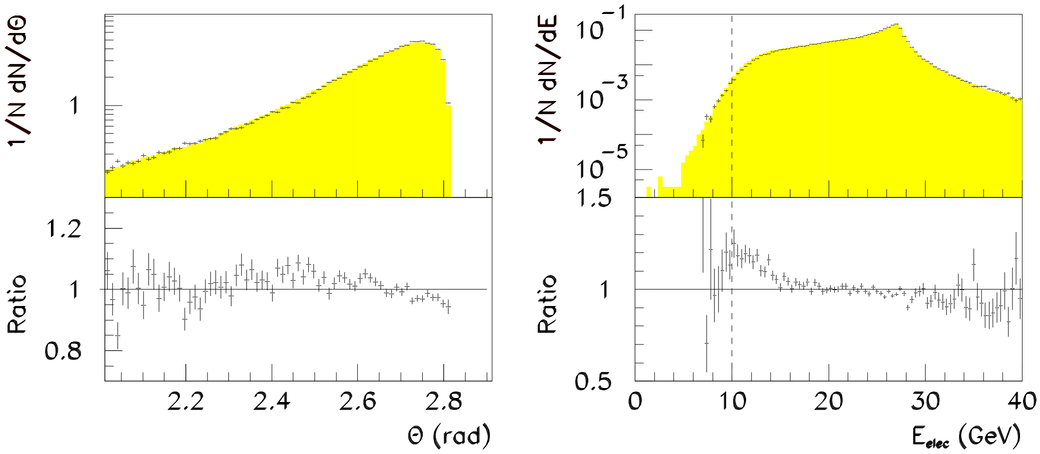


Figure 4.10: Comparison between data and ARIADNE Monte Carlo for the scattered electron polar angle and energy. The lower half of each plot is the ratio of data to Monte Carlo. The 10 GeV cut on the electron energy is indicated.

The four-vector of the electron is reconstructed using the polar angle θ'_e and the energy E'_e. Figure 4.10 shows these two quantities compared to ARIADNE Monte Carlo. Both are well described, except for the electron energy below about 15 GeV, where Monte Carlo underestimates the data by 10-20%. This is due to a poor Monte Carlo simulation of the dead material in front of the calorimeter, and will only affect the bins at high y and low x.

The value of θ'_e is measured, in order of preference, by

1. The angle of the measured track, if there is a track match to the calorimeter deposit.

2. The positional information from the Small Rear Tracking Detector (SRTD) if an electron has been detected in it, and the signal in the detector is not saturated, since the spatial resolution of the SRTD is superior to the RCAL. The angle is then calculated from both this

positional information and the reconstructed event vertex.

3. The positional information from the Forward HES (FHES) or the Rear HES (RHES), if an electron has been detected in these, and the signal is not saturated, since they too have superior spatial resolution to the F/RCAL. The angle is calculated as in 2.

4. The position is taken from the calorimeter cell if nothing else is available or sufficiently well measured. The position is calculated from the nominal geometrical centre, shifted proportional to the energy imbalance between the two photomultipliers on either side of the cell. The angle is calculated as in 2.

Figure 4.11 shows the resulting angular resolution for all the analysis bins in $x - Q^2$. The polar angle is measured very precisely, with resolution of between 0.1 - 0.6%, and almost no bias.

The electron energy, E'_e, can be reconstructed in two ways. The first way is to use the measured value of the energy directly from the calorimeter. Energy losses in inactive material ('dead material') in front of the calorimeter (e.g. cooling pipes and readout wires) cause the measured value to be less than the true value, and make this method unreliable. This can be compensated somewhat by utilising the energy reading in the RCAL presampler and SRTD to correct the CAL measured energy. The second method is to use the reconstructed value of Q^2_{DA}, inverting Equation 4.3 and replacing Q^2_e by Q^2_{DA}, to give

$$E'_{e,DA} = \frac{Q^2_{DA}}{2E_e(1 + \cos\theta'_e)} \tag{4.14}$$

Both methods have similar energy resolutions; the resolution for $E'_{e,DA}$ is shown in Figure 4.12 and is a flat 6-7% over the whole kinematic range. The

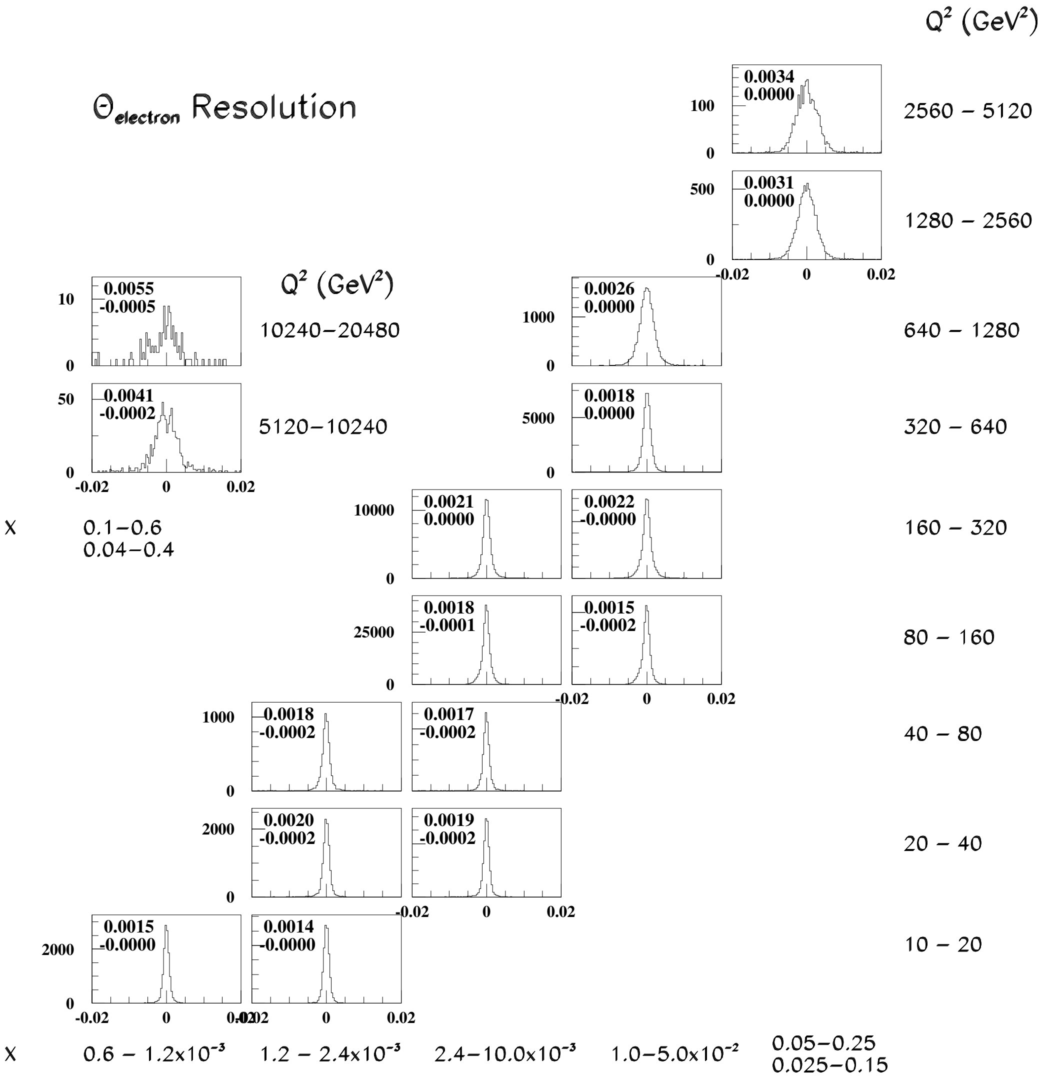


Figure 4.11: Resolution of θ_e' as a function of x and Q^2 using a combination of all available detectors. The top number in each plot is the fractional RMS resolution, the lower number is the fractional bias.

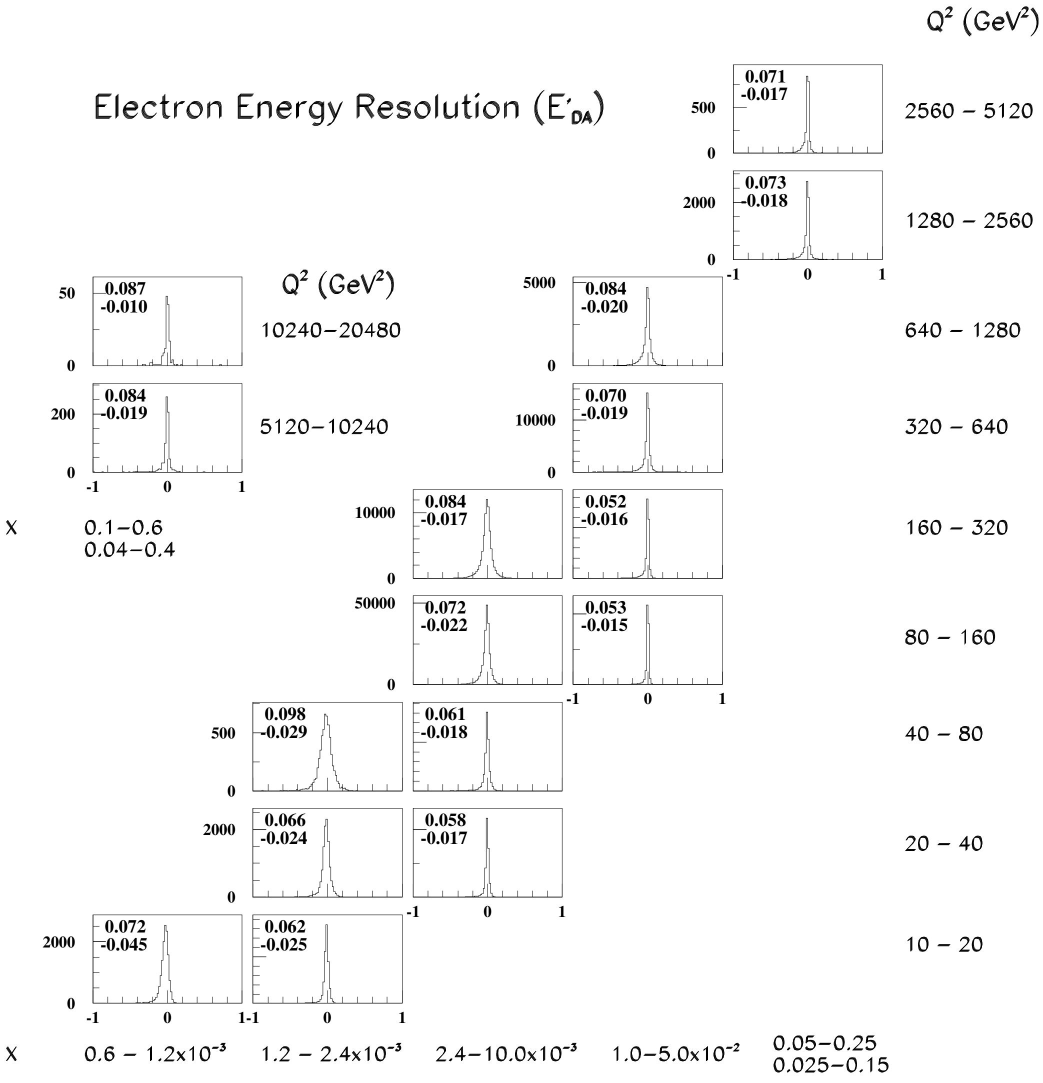


Figure 4.12: Resolution of $E'_{e,DA}$ as a function of x and Q^2 using a combination of all available detectors. The top number in each plot is the fractional RMS resolution, the lower number is the fractional bias.

bias is generally no more than 2.5%. It is vital to measure this electron energy as accurately as possible, since this is the quantity which determines the Lorentz boost to the Breit frame. The Monte Carlo describes the DA reconstructed electron energy better than the directly measured electron energy, so for this reason, we use the DA method to obtain the electron kinematics.

4.4 Breit Frame Boost

The boost vector $\vec{\beta}$, given in Equation 1.23, used to determine the Lorentz boost into the Breit frame, requires the four vector of the virtual photon, q. This is calculated from the four-vector of the scattered electron, e', as determined in the previous section, and the nominal incident electron four-vector, e, as

$$q = e' - e \tag{4.15}$$

The resolution of the magnitude of the boost vector is plotted in Figure 4.13. The boost is determined to within 4 - 6% in most of the kinematic bins.

4.5 Phase Space Definition

Not all of the $x - Q^2$ plane is experimentally accessible. This can be due to limitations of the detector, or excessive background in certain regions of phase space that make the measurement impossible. The measured phase space is limited by the following cuts

1. $\mathbf{Q^2_{DA} > 10\ GeV^2}$

   This is close to the lowest practical Q^2 that can be measured in DIS. It corresponds to a very low angle scattered electron in the RCAL; the

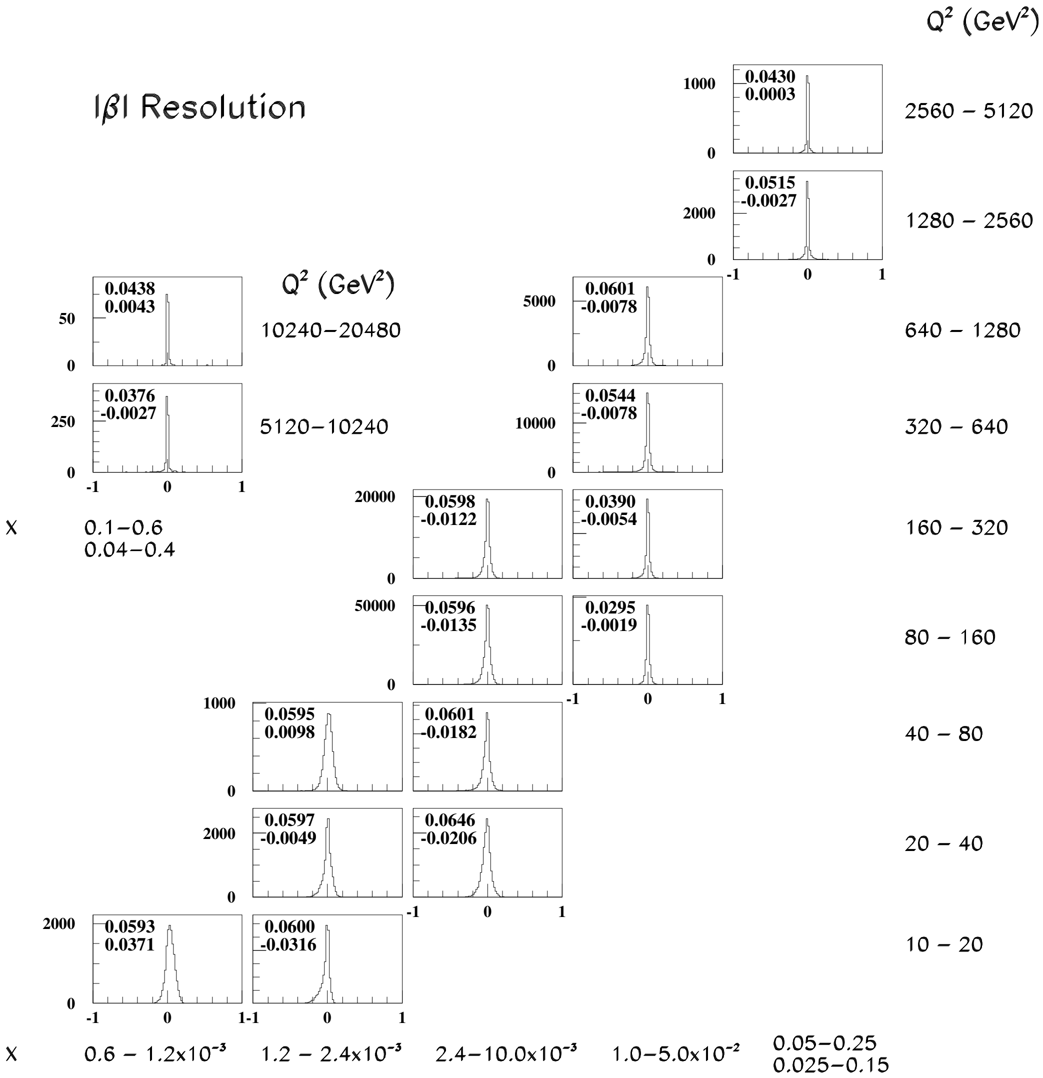


Figure 4.13: Resolution of the boost vector $|\beta|$ as a function of x and Q^2 using a combination of all available detectors. The top number in each plot is the fractional RMS resolution, the lower number is the fractional bias.

beampipe prevents the measurement of much lower Q^2 by any kinematic method that demands the presence of a well measured DIS electron.

2. $\mathbf{y_{JB} > 0.04}$

   At very low values of y, the hadronic jet is very close to the FCAL beampipe, sometimes even lost down the beampipe. In this case, the residual signal from the ^{238}U makes a large relative contribution to y_{JB} (which is essentially a measure of hadronic activity) and distorts the measurement of the hadronic angle γ_h, which in turn, will distort the DA measurements. In order not to be sensitive to this, a cut is made on y_{JB}.

3. $\mathbf{y_e < 0.95}$

   In photoproduction events, where the electron goes undetected down the beampipe, low energy pions from the fragmentation of the current jet can fake an electron signal in the FCAL. These events tend to have very high values of y_e at high Q^2 and are rejected by this cut.

The phase space of the truth level is defined to match the cuts at the detector level, namely,

$$Q^2_{true} > 10 \text{ GeV}^2 \tag{4.16}$$

$$0.04 < y_{true} < 0.95 \tag{4.17}$$

4.6 Background Rejection Cuts

The following cuts are imposed on the data in order to remove the majority of background events from the sample.

4.6.1 Momentum conservation

Although, the ZEUS calorimeter is almost completely hermetic, a large fraction of the energy measured in the detector is lost as the proton remnant goes down the beampipe; equally, a large amount of longitudinal momentum is lost, so neither of these quantities can be used to constrain the event. However, the quantity δ can defined as

$$\delta = \sum_i (E_i - p_{z,i}) = \sum_i E_i (1 - \cos\theta_i) \tag{4.18}$$

where θ_i is the angle of the calorimeter cell with respect to the incoming proton direction and the sum runs over all calorimeter deposits including the DIS electron[3]. This is conserved since the energy lost to the detector from the proton remnant is all in the $+p_z$ direction. Anything in the proton direction has a very low angle θ_i and makes a very small contribution to δ, consequently only the incoming electron makes a contribution to δ, leading to an expected peak at twice the incident electron energy of $2E_e$. Figure 4.14 shows the distribution of δ for data and Monte Carlo, showing that it peaks at approximately $2E_e$. A skew between data and Monte Carlo is noticeable, due to problems in the simulation of the hadron calorimeter energy scale. A cut is made on the reconstructed value of δ such that

$$35 < \delta < 60 \;\; \text{GeV} \tag{4.19}$$

and removes approximately 1% of the events. This cut is imposed for three reasons:

1. The kinematics of the DA method are sensitive to Initial State Radiation (ISR), where the incoming electron radiates a hard photon down

[3]Note the difference to the δ_h used in the evaluation of y_{JB} which only includes hadronic deposits.

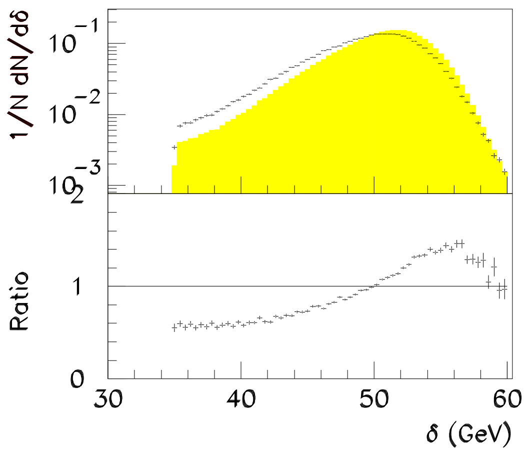


Figure 4.14: The δ distribution.

the beampipe in the electron direction. In this case, δ is substantially reduced, and these events are removed by the lower cut of 35 GeV. For Final State Radiation (FSR), where the scattered DIS electron radiates a photon, this is not a problem since the radiated photon is usually emitted with sufficiently small angle with respect to the electron that it is counted as part of the electron cluster.

2. In photoproduction, the scattered electron goes down the beampipe, again substantially reducing the value of δ. This cut, in combination with the y_e cut, removes practically all of the contamination from photoproduction events.

3. Proton beam gas events which occur inside the detector are not removed by the trigger timing cuts on the calorimeter since the relative timing between signals arriving in different parts of the calorimeter is similar to a real DIS event. However, they do deposit all their energy in the FCAL, leading to low values of δ, and are removed by this cut.

The value of δ should never be above $2E_e \simeq 55$ GeV. The upper cut of 60 GeV is to cut out unphysical tails.

4.6.2 Calorimeter Timing and Vetowall

Timing cuts to remove contamination from proton beam gas background and cosmic events are made at the trigger level as discussed in section 3.1. Events which have a deposit in the veto wall are caused by proton beam gas interactions upstream of the detector. These events are removed from the sample.

4.6.3 Cosmic Muon Rejection

Muons from cosmic showers are rejected first at the trigger level using a timing cut between the upper and lower half of the BCAL, and further by a more sophisticated offline algorithm that combines timing imbalance information with the characteristic shower shape produced by muons in the calorimeter.

4.6.4 Beam Halo Muon Rejection

Halo muons traveling through the BCAL parallel to the beam axis leave characteristic deposits in consecutive towers of the BCAL and no activity in the CTD. This is a clear signal for halo muons and these events are rejected.

4.6.5 QED Compton Rejection

Both elastic and inelastic QED Compton events are searched for. Elastic Compton events require two good electromagnetic candidates for the photon and electron, and further require that these two account for $> 90\%$ of the total energy in the calorimeter, and that the remaining energy not associated with the electromagnetic clusters is less than 3 GeV. Inelastic Compton identification uses similar cuts, but is complicated by the presence of a hadronic system. The final identification is made by topological cuts [51]. Any Compton events identified are rejected from the sample.

4.7 Cleaning Cuts

The following cuts are imposed on the sample to ensure that the event is well measured and contained in the detector, and that all the detector components are operating optimally.

4.7.1 Vertex cut

The reconstructed z-vertex must be close to the nominal interaction point $x = y = z = 0$ for the event to be well contained in the detector. Figure 4.15 shows the reconstructed primary z-vertex distribution compared to ARIADNE Monte Carlo. The distribution is described well in the central region, with about a 15% underestimation by the Monte Carlo in the tails. A cut is made at

$$-50 < z_{\mathrm{vtx}} < 40 \quad \mathrm{cm} \tag{4.20}$$

which successfully removes any satellites bunches from the data (an example of which can be seen in the DQM Z_{vtx} plot in section 3.3.) The cut is asymmetric since during the 1996-1997 running period, HERA ran with the vertex offset by a few cm from the nominal IP.

4.7.2 Box-cut

Electrons at very small angles (low Q^2) are measured in the RCAL very close to the beampipe hole. There is a risk that some part of the electron shower

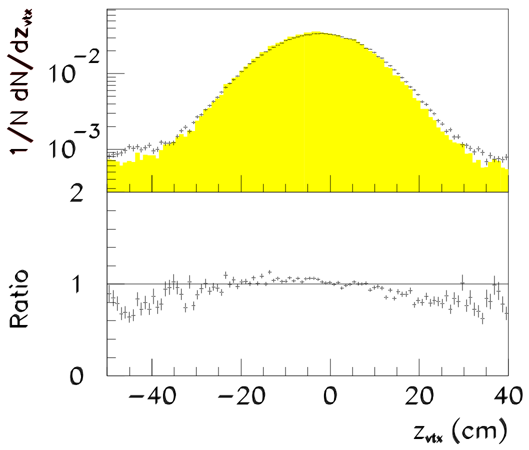


Figure 4.15: The z-vertex distribution of the data compared to Monte Carlo.

will be lost down the beampipe resulting in an incorrect energy measurement and problems in reconstructing the angle correctly. To prevent this, a cut is required on the measured x_e and y_e position of the electron on the RCAL face, given by

$$|x_e| > 16 \quad \text{cm} \quad \text{or} \quad |y_e| > 16 \quad \text{cm} \tag{4.21}$$

4.7.3 Data Quality Monitoring

As discussed in section 3.3, it is important to monitor both the individual detector components and the higher level physics data coming from the detector. The physics checks are discussed in section 3.3; in addition, runs and events were rejected where there were

- problems with the high voltage (HV) or gas system, or where the CTD was off. Runs with very high radiation backgrounds causing many HV trips were also rejected;
- too many sparking cells in the calorimeter, or there were too many cells where neither of the photomultiplier tubes were functioning ('holes');

- any problems with the online trigger or data acquisition system.

These problems are usually caught and corrected by the DQM shift online, with serious problems such as HV trips causing data-taking to stop. Any remaining problems are identified by the DQM procedures, and those runs rejected from the sample.

4.7.4 Calorimeter Spark Rejection

Static sparks between the photomultiplier tube housing and the photomultiplier tube itself can cause large signals in certain calorimeter cells. These are characterised by a large energy imbalance between the two PMTs and can be identified and removed this way. Most of these sparks are rejected by the online trigger. The remaining spark events are rejected offline by a more efficient algorithm. Additionally, any cells which are noisier than the others, averaged over the whole run period, are suppressed.

4.7.5 Backsplash Correction

At low y ($y < 0.3$) and higher Q^2, an overestimation of δ, as defined in Equation 4.5, is observed [52]. Since at low y, the hadronic system is very forward and comparatively low energy, any energy deposited in the rear of the detector will bias the measurement of δ. This energy can be deposited in two ways:

- Backsplash from the calorimeter, where some particles produced in the hadronic shower scatter into the opposite side of the detector.

- Scattering or showering from some material in front of the calorimeter, such as the beampipe or CTD wall.

These effects are poorly simulated in the Monte Carlo, so it cannot be relied upon to correct for this problem. From Equation 4.7 it is clear that an overestimation for δ will overestimate the hadronic angle γ_h and subsequently affect the reconstruction of the DA kinematics.

Both types of deposits have the characteristic feature that they are generally low energy (< 3 GeV) and have angles very far away from the main hadronic system. They are removed by defining

$$\gamma_{max} = \gamma_h + \text{offset} \tag{4.22}$$

Any deposits with energy $E_e < 3$ GeV and polar angle $\theta > \gamma_{max}$ are removed from the event. Since some cells have been removed, a new γ_h must be calculated, and the procedure is then iterated until the final value of γ_h stabilises. The value of the offset is tuned from Monte Carlo. This has been shown to successfully remove the majority of scattered deposits [53]. The final corrected value for γ_h is then used for the determination of DA kinematics.

4.7.6 Energy and Multiplicity Cut

In order for the NLO calculations (discussed in chapter 1) to be stable and reliable, it is required to run them with an energy cut in the current region of the Breit frame, as given in Equation 1.36. Since comparisons are made to this theory, the same cut must be applied in the data, namely

$$\mathcal{E}_{lim} > 0.1Q \tag{4.23}$$

where Q is the momentum transfer and the energy sum runs over all boosted hadrons in the current region of the Breit frame. Figure 4.16 shows the energy in the current region scaled to Q. Indicated is the value of the energy

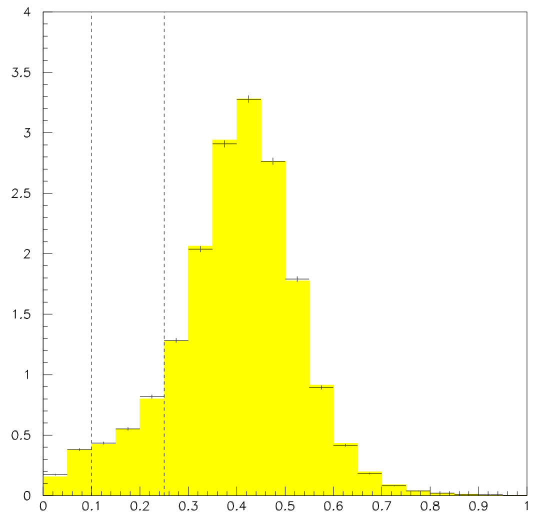


Figure 4.16: Comparison between data and ARIADNE Monte Carlo for $E_{current}/Q$. The current region energy cuts are indicated.

cut at 10% and an additional energy cut at 25%. All the data points in the analysis use a 10% cut unless otherwise stated. The effect of the 25% cut on the fitted results is examined in chapter 8. The QPM peak of $E = Q/2$ is visible, although slightly shifted towards zero due to migrations into the target region. The Monte Carlo gives a good representation of the total energy and the peak position. The tail at low energies is due to soft particles migrating over from the target region. The cut at 10% will remove around half of these particles, with the rest being removed by the 25% cut.

As discussed in chapter 1, the definitions of some of the event shapes are ambiguous if there are less than two partons in the current region of the Breit Frame. Therefore, only events with two or more particles in the current region are used in the calculation of the shapes.

4.8 Summary

An examination of the resolutions and comparison with Monte Carlo indicates that the Double Angle method is the best for reconstructing the kinematics of the events, using the best combination of all available detectors. The electron finding algorithm was presented and shown to agree reasonably with the Monte Carlo simulation. The resolution on the boost is approximately 4-6% over the majority of the measured kinematic plane.

The cuts used to remove contamination from background events were defined, and successfully remove the majority of background. Data quality cuts were introduced and justified. The selected DIS event sample was presented and the efficiencies and purities given. In the region that will be studied in chapter 8, the efficiencies and purities are high.

Chapter 5

Hadronic Final State Reconstruction

Two detectors are used to reconstruct the hadronic final state. Firstly, tracks from charged particles are identified in the CTD, which has excellent spatial and momentum resolution. This detector is limited by an angular acceptance of $|\eta| < 2$ and the fact that it can only detect charged particles.

The calorimeter can be used to detect both charged and neutral particles. It has a good energy resolution, as given in Equation 2.3, although the spatial resolution is inferior to that of the CTD. The optimal resolution on the hadronic final state can be achieved using a combination of the information from these two detectors.

5.1 Track Reconstruction

The standard ZEUS package used to reconstruct tracks in the CTD is called VCTRACK [54]. This program can use tracking information from all of the tracking detectors (CTD, FDET, RTD.) In principle, it is advantageous to use all available information, including that from the FDET and RTD, however, these detectors are not sufficiently well simulated in the Monte Carlo, so the analysis relies on the CTD for track reconstruction.

Each candidate track begins as a track 'seed', consisting of three hits in the outer superlayers (SL6-9). The track seed is then extrapolated inwards picking up hits along the way. The track parameters, which are fitted to a 5 parameter helix model, are continually updated, increasing in precision with every hit. Periodically, any tracks which have too many hits shared with other tracks are rejected; it is normally required that 85% of a track's hits are unique to it.

The tracks are then used for vertex finding [54]. The primary vertex is identified using a full χ^2 fit to all tracks that cross at least one of the inner superlayers. Subsequently, any remaining tracks are refitted in an attempt to match them to secondary decay vertices. Tracks considered for use in the analysis must fulfill the following requirements:

- be fitted to either the primary or a secondary vertex;
- have hits in ≥ 4 superlayers, to ensure that they are long enough to be well measured;
- have a minimum p_T of 0.1 GeV, again to ensure they are well measured.

5.2 Calorimeter Island Reconstruction

The treatment of calorimeter deposits is divided into two stages. Firstly, quality cuts and corrections are applied to the calorimeter cells to ensure optimal measurement. Secondly, the individual CAL cells are clustered into objects more amenable to matching with tracks.

5.2.1 Calorimeter Noise Suppression

The uranium background signal deposits, on average, a total of 1 GeV per event randomly around the calorimeter [55]. Any isolated cell with $E < 100(150)$ MeV in the EMC(HAC) is removed from the analysis of that event.

5.2.2 Calorimeter Energy Correction

Initial studies of the Monte Carlo showed a discrepancy between the data and Monte Carlo in the calorimeter energy scale. Basic correction factors are therefore applied to the measured calorimeter energy in the data to account for these effects.

- In the RCAL, separate correction factors are applied for each cell. For the inner ring of the RCAL, these have been determined from kinematic peak events using the shift in the fitted peak between data and Monte Carlo. For the outer ring, they are obtained by comparing the measured electron energy to the electron energy reconstructed from the DA method. The corrections are typically 2-3%.
- In the BCAL, a global +5% is applied to all cells.
- In the FCAL, no correction is applied.

5.2.3 Backsplash Correction

The treatment of 'backsplash' deposits is described in section 4.7.5, in the discussion of the best measurement of the hadronic angle γ_h. Cells which are identified by this algorithm as backsplash are removed from the determination of the hadronic final state. A systematic check with backsplash cells included is presented in chapter 7.

5.2.4 Calorimeter Clustering

When a particle makes a deposit in the calorimeter, the shower is characterised by a depth and a width. In energetic hadronic showers, there will generally be deposits in the EMC and both HAC sections. Within each CAL section, the shower may extend over several cells. Instead of using individual cells to reconstruct the hadronic final state, the cells are clustered together, so as to have one single object that more closely represents the original particle [56].

However, showers from two or more particles may overlap in the calorimeter making unique identification of each individual particle difficult. This is not so much of a problem for single particle event shapes, so long as the overall momentum distribution from the clustered objects approximates as closely as possible the overall momentum distribution of the original particles. It is more of a problem for the two particle event shapes since these have contributions based on the angle between pairs, and if a pair is misidentified as a single particle, this contribution will be lost. However, in this case, the angle between the two particles is small, so the lost contribution is small as well.

The clustering is performed in two stages. Firstly *cell-islands* are built in each of the separate 8 sections of the calorimeter (FCAL HACI, FCAL HACII, BCAL EMC, . . .). The cells are connected using a 'nearest-neighbour' algorithm as shown in Figure 5.1. A deposit in a cell is connected to the highest energy deposit in the four cells surrounding it. No connection is made on diagonal cells. The centre of each of the cells making up the cell-island is determined from the geometrical centre of the cell, corrected by the energy imbalance between the two photomultiplier tubes. The centre of the whole cell-island is then calculated using a logarithmically weighted centre of gravity as discussed in [57].

The cell-islands in each of the individual sections of the calorimeter are now connected together globally. A typical single hadronic deposit will produce three individual cell-islands, one in the EMC, one in the HACI and one in the HACII section. Also, if the deposit is in the crack region, then it will be spread across two calorimeter sections (eg. the RCAL EMC and the BCAL EMC). Therefore, all the cell-islands belonging to one hadronic deposit need to be reunited, so that the resulting *cone-islands* more closely represent the original particle distribution.

Matching can occur between HACII $\rightarrow$ HACI, HAC $\rightarrow$ EMC and EMC $\rightarrow$ EMC. A probability distribution as a function of the opening angle between two cell-islands exists for each of these matching possibilities. The probability distributions have been determined and tuned from Monte Carlo [57]. If the probability given by the opening angle between two cell-islands is sufficiently high, then the two cell-islands are merged into one cone-island. The process

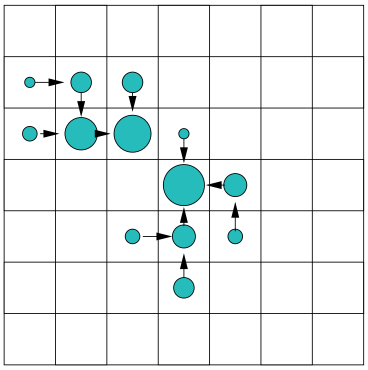


Figure 5.1: Nearest-neighbour cells are connected in the cell-island algorithm.

continues over all cell-islands until they have all been merged into cone-islands. The position of the cone-islands is calculated in the same way as that for the cell-islands.

5.3 Calorimeter to Track Matching

The optimal resolution on the hadronic final state is achieved by using a combination of the measured tracks and the cone-islands, schematically illustrated in Figure 5.2. Tracks passing the quality cuts in section 5.1 are extrapolated onto the calorimeter face using a 'swimming' algorithm from VCTRACK [54]. For each track-island combination, a Distance of Closest Approach (DCA) is calculated, from the track impact point on the face of

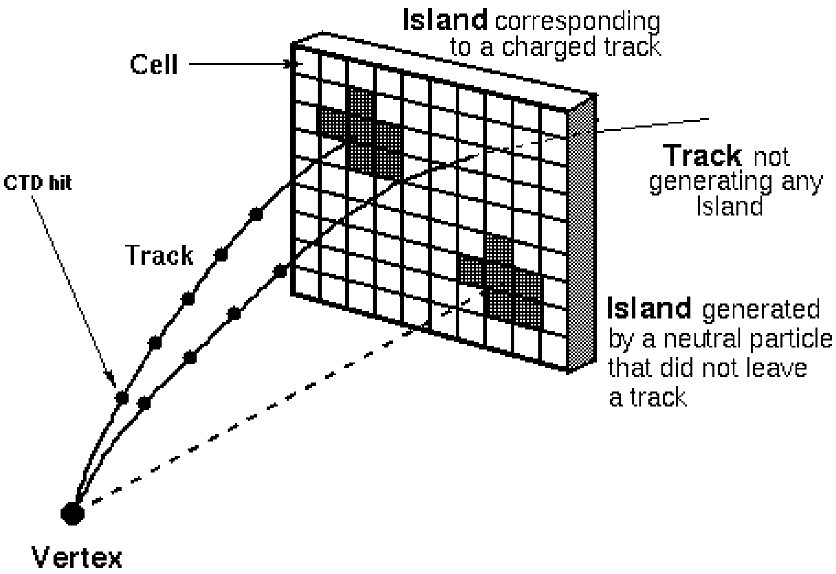


Figure 5.2: Matching of calorimeter deposits to tracks in the CTD provides optimal resolution.

the calorimeter to the centre of the cone-island. A match is declared if

$$\mathrm{DCA} < 20\ \mathrm{cm} \quad .\mathrm{OR.} \quad \mathrm{DCA} < R \tag{5.1}$$

where R is the radius of the cone-island on the face of the calorimeter.

All the calorimeter deposits and tracks are processed in this way, and if any matches exist, the optimal information is taken for the single combined object. If no match is obtained for an object, then the unmatched track or cone-island is used directly. The objects thus matched are referred to as (Zeus Unidentified Flow Objects) ZUFOs and used in the determination of the final state [57] [58]. Figure 5.3 shows the relative fraction of ZUFOs for each type, considering ZUFOs in the current region only.

The Monte Carlo represents the behaviour of the ZUFOs algorithm adequately, as shown in Figure 5.3, although it slightly underestimates the fraction of events where the cone-islands were replaced by the tracks and

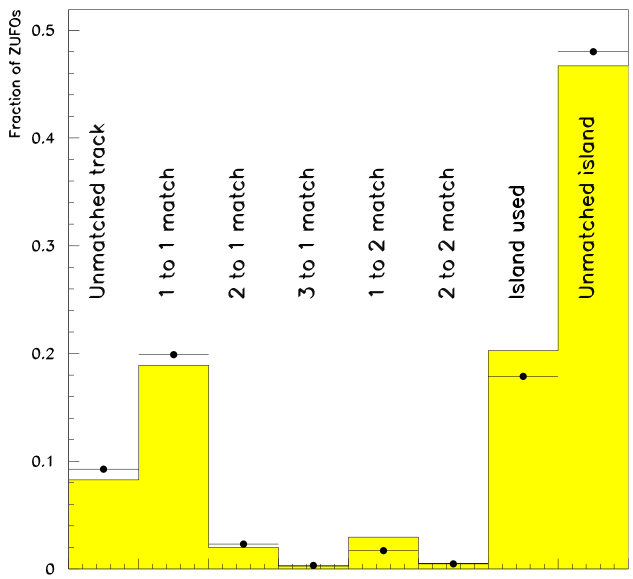


Figure 5.3: Fraction of ZUFOs of particular types. The shaded histogram is Monte Carlo, the points are data.

underestimates the fraction of cone-islands that could not be matched to any tracks. There are three main classes of ZUFO:

1. 43% of the ZUFOs have some match between a cone-island and a track. The decision about whether to substitute the track information for the cone-island information is based on two requirements. Firstly, it is necessary to check that the energy deposit in the calorimeter comes only from the track, and not from any overlapping neutral deposits. In this case, taking just the track energy would lose the overlapping neutral energy. To substitute tracking information for cone-island information, it is required that

$$\frac{E_{\text{island}}}{p_{\text{track}}} < \beta + n\left(\frac{\Delta E_{\text{island}}}{\Delta p_{\text{track}}}\right) \qquad \beta = 1.0,\ n = 1.2 \tag{5.2}$$

where the n is to allow for the finite resolutions on the cone-island energy and track momenta; the resolutions are given in Equations 2.2 and 2.3. The second requirement is that the momentum resolution of the track is better than the energy resolution of the cone-island

$$\frac{\Delta p_{\text{track}}}{p_{\text{track}}} < \alpha\left(\frac{\Delta E_{\text{island}}}{E_{island}}\right) \qquad \alpha = \begin{cases} 1.0 & \text{in general CAL.} \\ 1.2 & \text{in CAL crack region.} \end{cases} \tag{5.3}$$

Figure 5.4 shows that the electromagnetic (hadronic) calorimeter energy resolution is superior to the tracking resolution above 9 (15) GeV for central tracks ($\eta = 0$). The factor $\alpha = 1.2$ in the crack regions increases the likelihood of taking the track information in these regions, since the calorimeter resolution there is worse than the rest of the calorimeter. If the four-vector is taken from the track, then the pion mass is assigned.

In a small fraction of the ZUFOs, the matching is more complicated. Two or three tracks can point to one cone-island (2 to 1 and 3 to

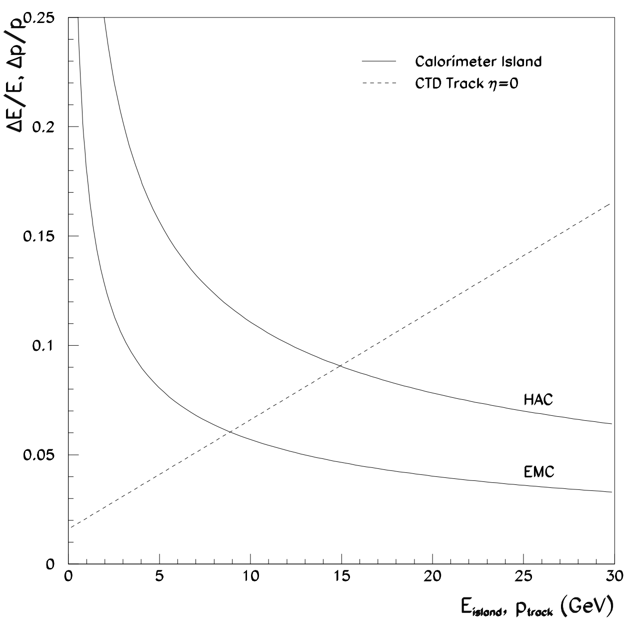


Figure 5.4: Comparison of track momentum resolution and calorimeter island energy resolution for central tracks ($\eta = 0$).

1 match). Alternatively, the deposit from one or two tracks can be split into two cone-islands (1 to 2 and 2 to 2 matches.) In all cases, the energies and momenta of all cone-islands and tracks involved are summed, and Equations 5.2 and 5.3 applied to the sums.

Of the 43% of ZUFOs with some track match, 20% comes from ZUFOs with an unambiguous 1-to-1 track match for which the tracking information was taken. A further 5% comes from more complicated configurations for which the tracking was taken. The remaining 18% are ZUFOs which failed the track replacement criteria and retained the information from the cone-island ('Island used' in Figure 5.3.)

2. 48% of the ZUFOs come from cone-islands to which no track could be matched. These deposits come predominantly from neutral particles

and angular regions where the CTD acceptance is poor. For this type, the angle which the centre of the cone-island makes with the vertex is used, along with the cone-island energy, to determine the four-vector. The majority of these are caused by photons, so zero mass is assigned to the object.

3. 9% of the ZUFOs come from unmatched tracks. These are generally low energy tracks that could not be unambiguously matched to a cone-island. The four-vector is measured from the tracking angle and momentum. The majority of charged tracks are from pions, so a pion mass is assigned to these objects.

The energy distribution of the ZUFOs is shown in Figure 5.5a), compared to ARIADNE Monte Carlo. There is a clear deficit in the Monte Carlo in the high energy region. This is due to a poor simulation of the energy deposits in the inner F/RCAL where no tracking information is available to augment

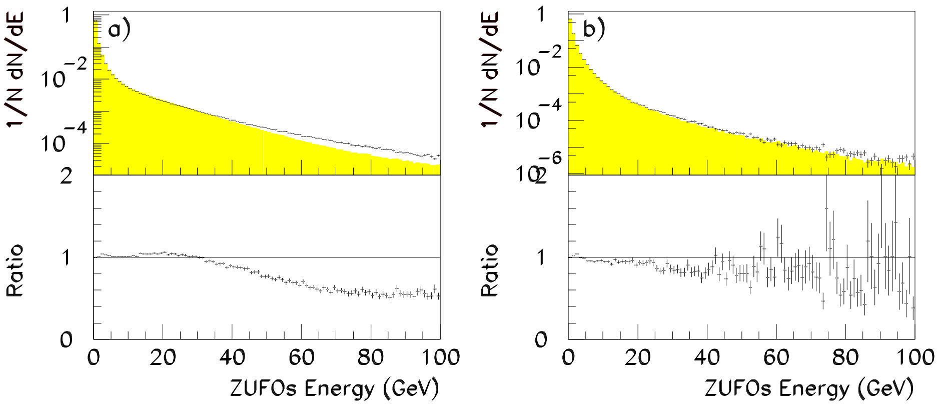


Figure 5.5: a) shows the comparison between data (points) and ARIADNE Monte Carlo (shaded) for the energy of the ZUFOs before the η and p_T cuts. b) shows the same distribution after the cuts.

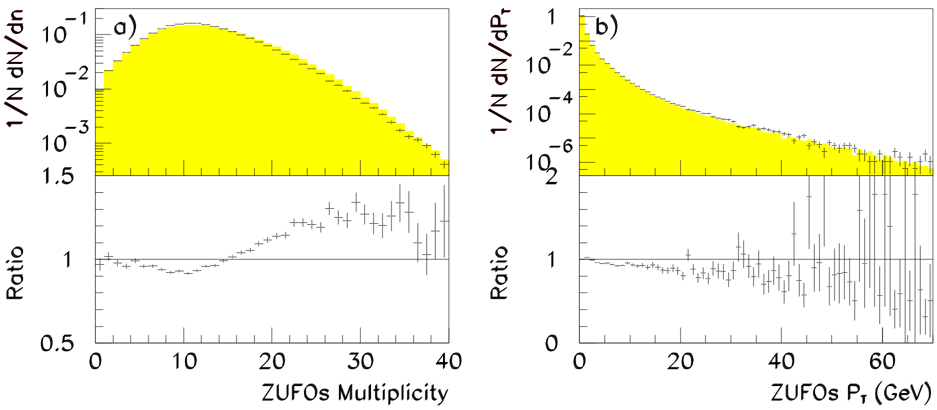


Figure 5.6: a) shows the global multiplicity of ZUFOs compared to ARIADNE Monte Carlo. b) shows the ZUFOs p_T spectrum.

the calorimeter deposits. It is therefore a requirement that all ZUFOs are in the good acceptance region of the CTD, given by

$$\left.\begin{array}{r}|\eta_h| < 1.75\\ p_{T,h} > 0.15 \text{ GeV}\end{array}\right\} \text{over all ZUFOs.} \tag{5.4}$$

The resulting agreement of the energy distribution with Monte Carlo is shown in Figure 5.5b). Both the energy and relative fraction of the ZUFOs are adequately simulated in the Monte Carlo. A further check is made on the global multiplicity and the p_T spectrum of the final ZUFOs as shown in Figure 5.6. The multiplicity plot shows a small tendency for the Monte Carlo to overestimate the number of ZUFOs in an event (by around 3-4%). This is not serious since the combined four-vector for any hadronic deposit that is split in the Monte Carlo will very closely represent the single four-vector that would be found in the data. The p_T spectrum of the ZUFOs is very well simulated. The ZUFOs thus selected in the lab frame are boosted into the Breit Frame as discussed in section 4.4. After boosting, the energy of the ZUFOs is rescaled assuming the objects are massless i.e. by setting

$E' = \sqrt{p_x'^2 + p_y'^2 + p_z'^2}$, as discussed in chapter 1. The sums for event shapes in Equations 1.27 - 1.35 then run over all the ZUFOs in the current region of the Breit Frame ($p_z' < 0$). The event shapes are calculated when there are two or more ZUFOs in the current region.

5.4 Hadron Level Definition

When making a comparison of hadronic final state variables to Monte Carlo predictions, it is necessary to define what is meant by the final state of the Monte Carlo. A number of different definitions are possible based on whether particular particles are assigned stable or allowed to decay. The standard e^+e^- definition that is chosen [16], is to take as stable all particles with mean lifetimes $\tau > 3.0 \times 10^{-10}$ s, subject to the following criteria

1. 'Unstable' particles (i.e. lifetimes $\tau < 3.0 \times 10^{-10}$) are rejected.
2. If a particle marked as 'stable' has any ancestor in the decay chain which is also marked as 'stable' then the ancestor is taken instead.

This definition is designed to keep the leptons, photons, charged pions, nucleons, and K_L^0 particles in the final state. Everything else, such as K_S^0 and π^0 particles should be allowed to decay. There are two rare exceptions to point 1. Very occasionally, in less than 0.1% of the events, an entire decay chain is marked as 'unstable', in which case the first particle in the decay chain is taken. The other exception is caused by a rare bug in the simulation ($<$ 0.1% of events), when an 'unstable' particle, often a strongly decaying particle, comes directly from the JETSET fragmentation but is not subsequently decayed by GEANT. In this case, the 'unstable' particle is taken as part of the final state, since this is the only possible option.

The hadron level thus selected consists of approximately 47.5% photons, 38.5% charged pions, 7.5% charged kaons and K_L^0, 5% baryons, and about 1.5% electrons, muons, taus and neutrinos. The residual contamination from 'unstable' π^0, K_S^0, Σ and η particles, as a result of the two exceptions, is less than 0.4%.

The selected hadrons are boosted to the Breit Frame using the true photon four-vector. All those in the current region ($p_z < 0$) are used to determine the event shapes. After boosting, the energy component of the four-vectors of the hadrons are rescaled by setting them as massless, and event shapes are calculated when there are two or more stable hadrons in the current region.

5.5 Event Shape Resolutions

The fractional resolutions of the event shapes, determined from ARIADNE Monte Carlo, are given in Figures 5.7 to 5.12. The resolution of event shape F is defined as

$$\frac{F_{true} - F_{reconstructed}}{F_{true}} \tag{5.5}$$

The resolution is variable dependent and Q^2 dependent. Generally, the region of $Q^2 > 80$ GeV^2 has a superior resolution to the lower Q^2 region. The thrusts generally have the best resolution, from 20% in the $Q^2 > 80$ GeV^2 region to 11% in the highest Q^2 region for thrust with respect to the photon axis. Thrust with respect to the thrust axis maintains a better resolution throughout, around 10%, increasing to 5% in the higher Q^2 region. It is reasonable that T_T has a better resolution, since it is not sensitive to the misreconstruction of the photon axis. It has some sensitivity due to the current region cut, but at high values of T_T, the jet is collimated, so there is less chance of particles migrating across into the target region. Broadening with respect to the photon axis has a resolution of around 20%, degrading to around 25% as Q^2 increases. The broadening with respect to the thrust axis has a resolution of around 25-30%. The broadenings have more comparable resolutions since they both suffer equally from migrations into the target region. C-parameter has a resolution of 30-35% with up to a 10% bias to lower values of C at low Q^2. The jet-mass has the worst resolution of around 40% and has a strong bias (16%) towards lower values of jet-mass at low Q^2. These biases are consistent with the merging of multiple hadrons into one single ZUFO, with the resultant loss of that pair's contribution to the variable.

5.6 Summary

The optimal resolution on the hadronic final state has been achieved by using a combination of tracking and calorimeter information. Calorimeter deposits were clustered using a cone-island clustering technique and then matched to tracks. Criteria applied to the resultant combined object determine whether the track or calorimeter information is most precise. A momentum and angular cut is applied to the combined objects to ensure that they are sufficiently well simulated in the Monte Carlo. A standard definition was adapted for the hadron level and its composition presented. The resulting resolutions achieved on the event shapes were presented and vary from 10% - 40%.

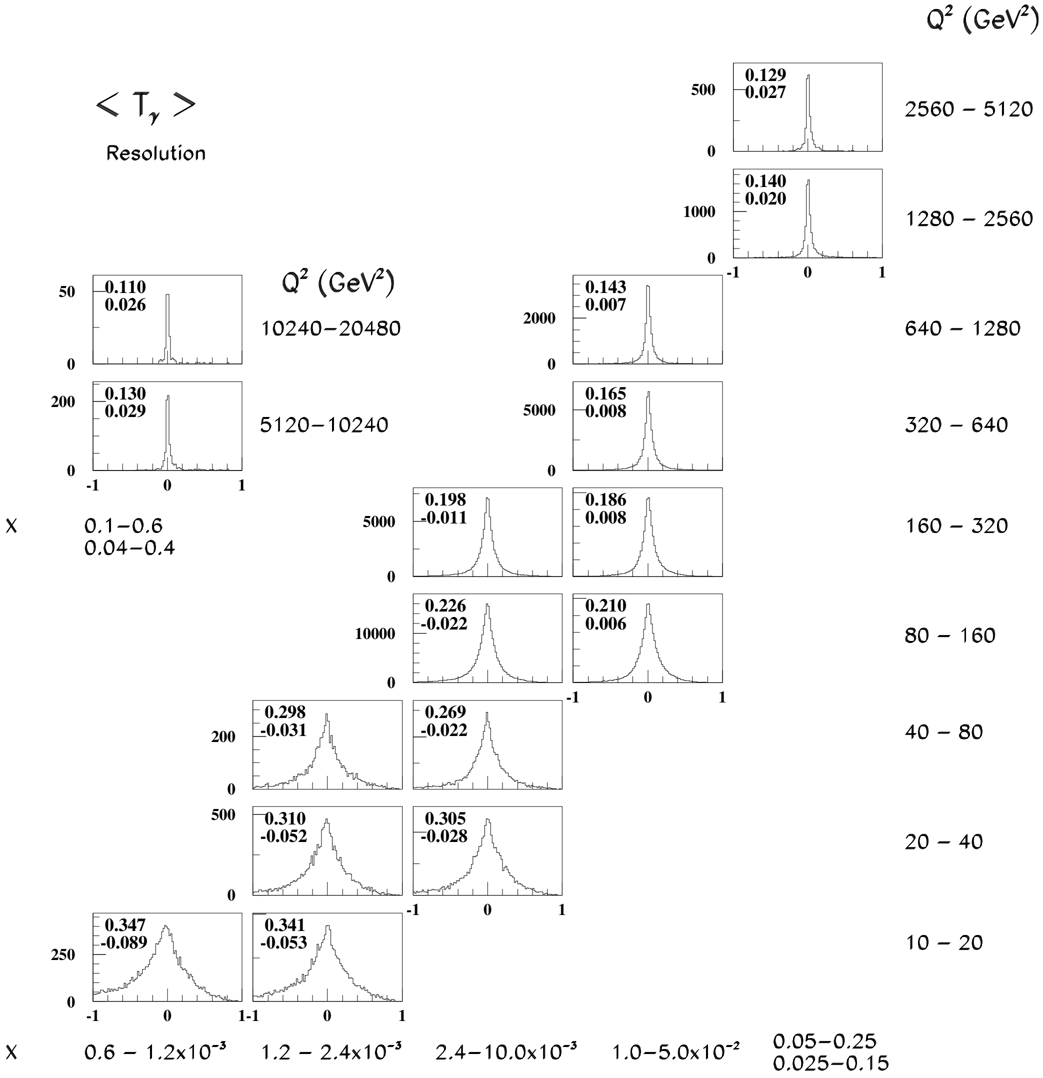


Figure 5.7: Resolution of τ_γ as a function of x and Q^2. The top number in each plot is the fractional RMS resolution, the lower number is the fractional bias.

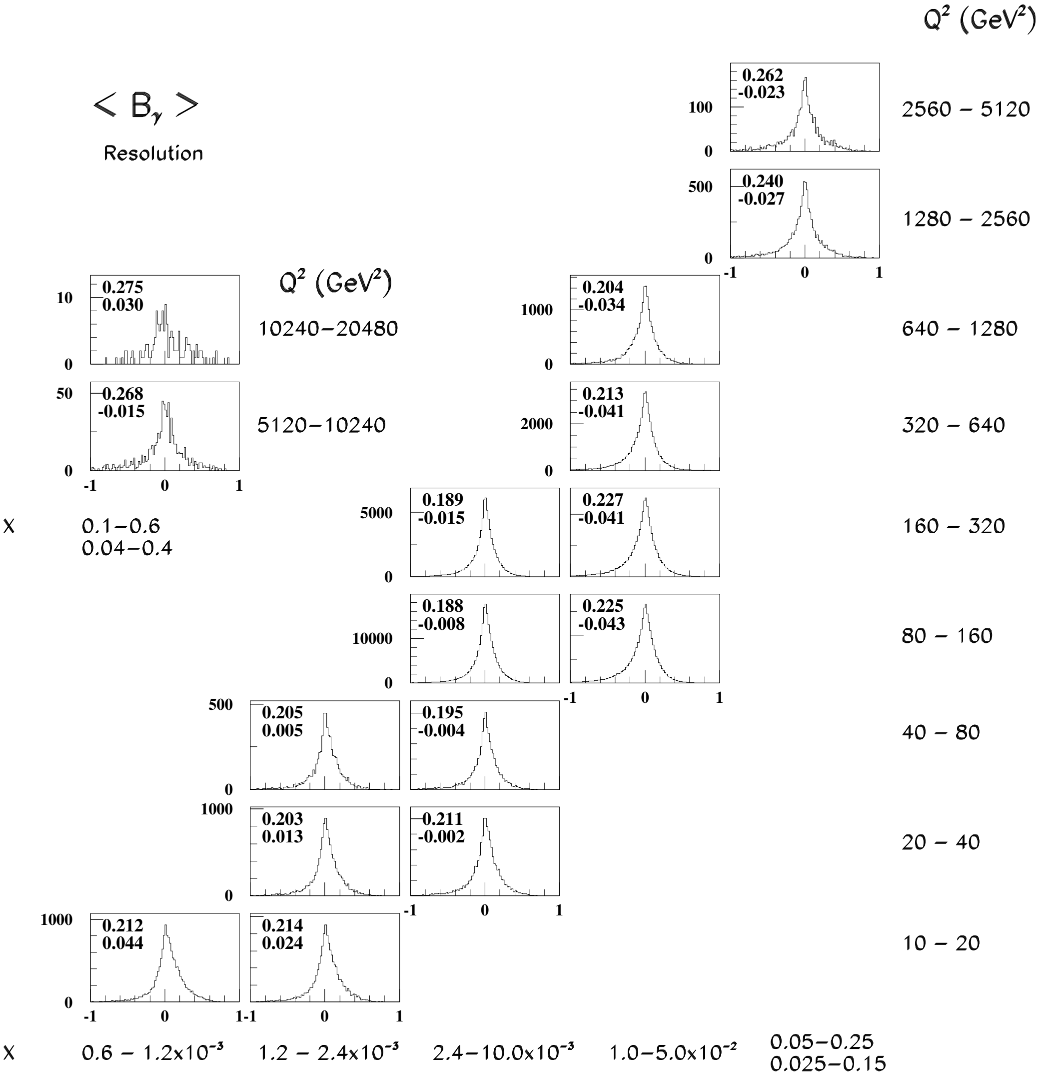


Figure 5.8: Resolution of B_γ as a function of x and Q^2. The top number in each plot is the fractional RMS resolution, the lower number is the fractional bias.

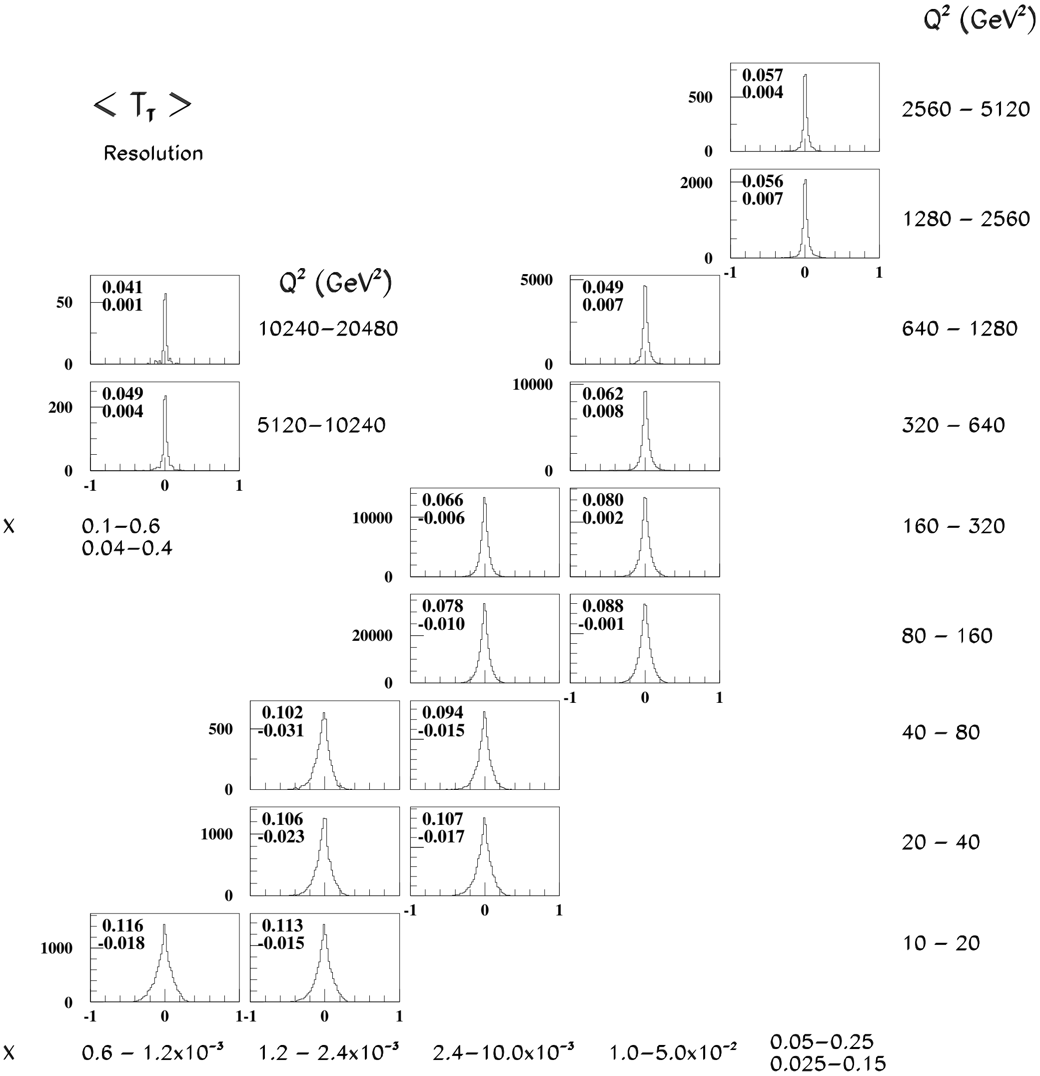


Figure 5.9: Resolution of τ_T as a function of x and Q^2. The top number in each plot is the fractional RMS resolution, the lower number is the fractional bias.

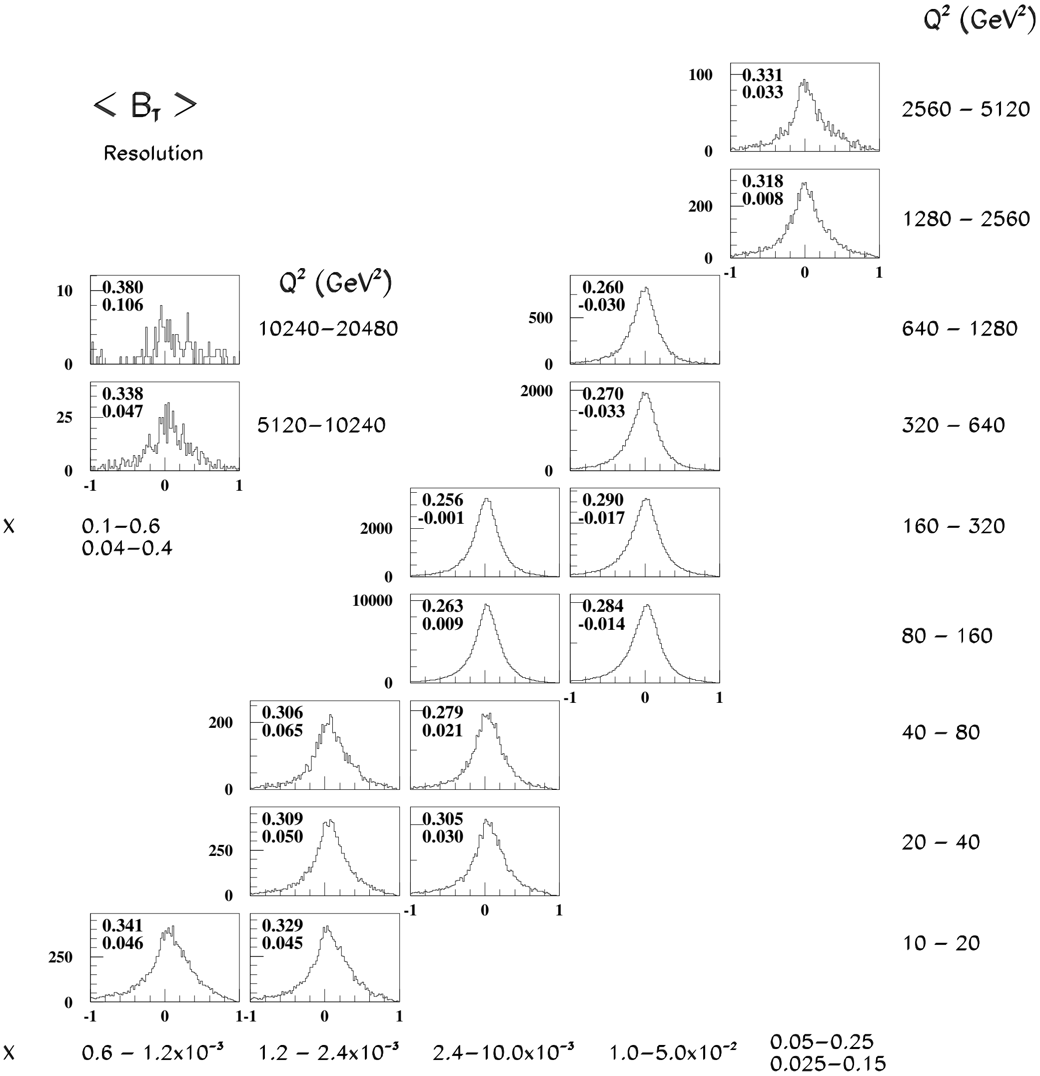


Figure 5.10: Resolution of B_T as a function of x and Q^2. The top number in each plot is the fractional RMS resolution, the lower number is the fractional bias.

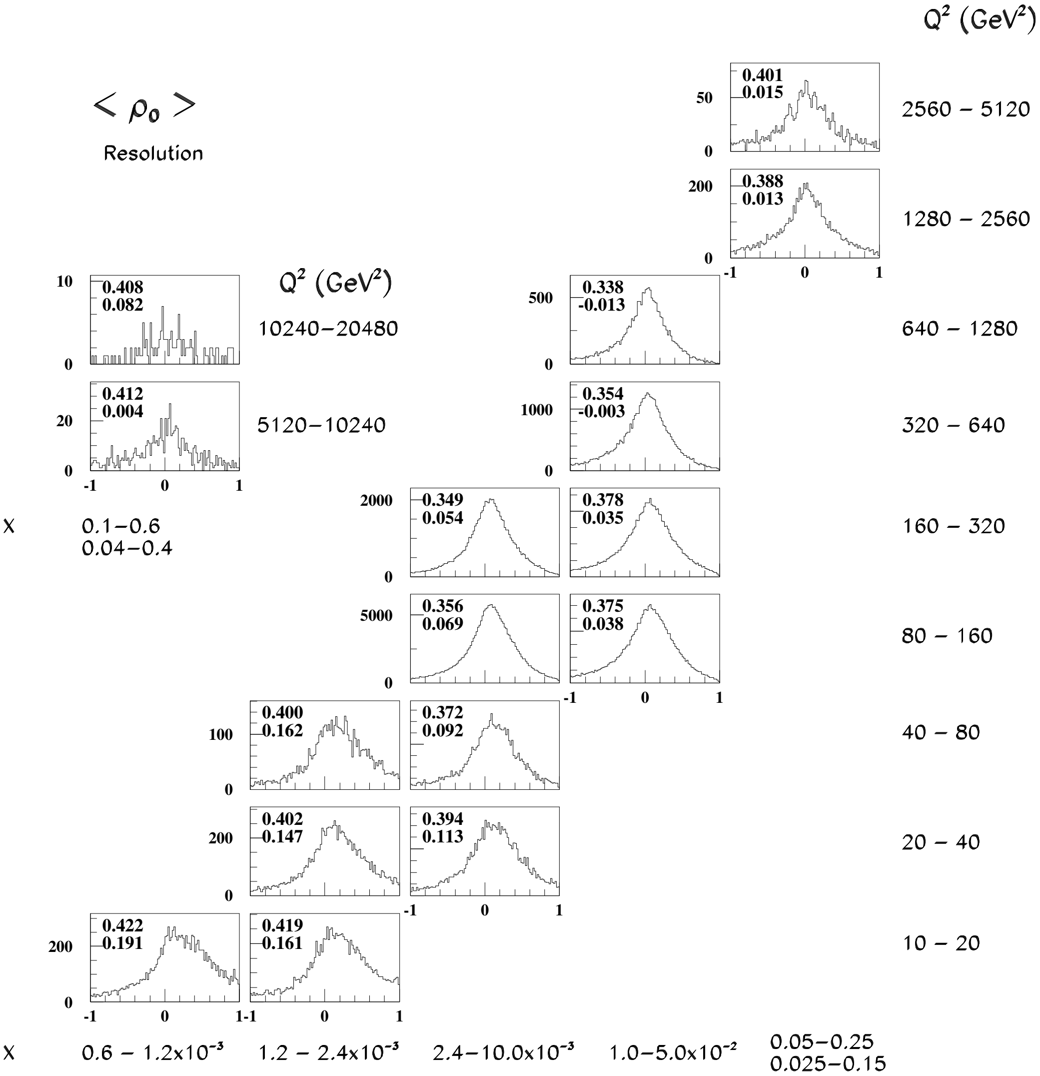


Figure 5.11: Resolution of ρ_0 as a function of x and Q^2. The top number in each plot is the fractional RMS resolution, the lower number is the fractional bias.

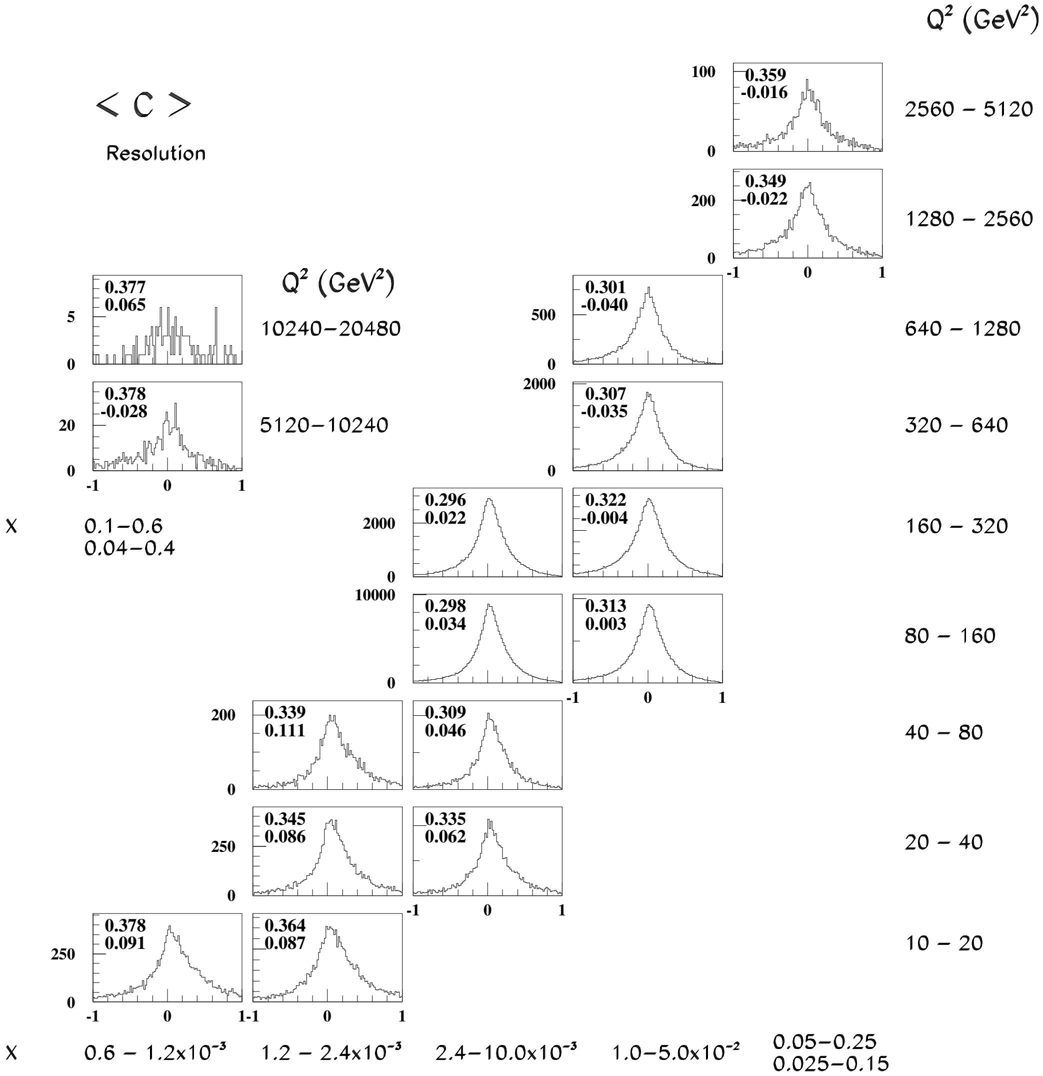


Figure 5.12: Resolution of C-parameter as a function of x and Q^2. The top number in each plot is the fractional RMS resolution, the lower number is the fractional bias.

Chapter 6

Event and Detector Simulation

This chapter describes the phenomenological models used to describe and correct the measured data distributions. The simulation of the detector is also discussed. Both of these models can be implemented using Monte Carlo methods.

6.1 Introduction

The basic matrix elements for the lepton-quark hard scattering process are calculable only up to leading-order (LO) in α_s (i.e. $\mathcal{O}(\alpha_s)$.) The situation is complicated by the necessity to evolve from the matrix element to the full partonic final state, and then to the hadrons which are actually detected in the experimental apparatus. This last process takes place at large distances and low energies, where α_s is of $\mathcal{O}(1)$, and therefore standard perturbative methods are not applicable.

Due to these theoretical difficulties, the only practical way to simulate interactions is to use phenomenological predictions that are generated using some model or combination of models. The models are probabilistic and produce individual events, therefore a large number of events is generally required to ensure that all possible physics processes have been included in the sample. This probabilistic approach is implemented by Monte Carlo methods, where a chain of random numbers is used as a basis for the simulation of the production of particles.

It is also necessary to correct for the finite resolutions and imperfections in the experimental apparatus. These effects will tend to obscure the real hadronic system, and in order to compare the data to theoretical models, it is essential to unfold the data, correcting for detector effects, so that the true underlying distribution can be recovered. The generated events are taken and passed into a full Monte Carlo simulation of the detector. The simulation includes a full accurate geometrical and compositional description of the detector and simulates all the scattering and other physics processes that will occur inside it. It also simulates the read out and triggering system. The output from this simulation is then passed to the same reconstruction program as the data, and from then on can be analysed as if it were data. Tagged onto each fully simulated event, is the real underlying event that was originally generated by the phenomenological model program. Comparison of these underlying 'true' events to the 'reconstructed' events which have been passed through the detector simulation provides information on the detector response.

6.2 Monte Carlo Models

The simulation of the underlying physics event takes advantage of the factorisation property of QCD. This allows different processes in the interaction to be calculated separately according to the model of choice, with the results being combined to give the overall result. The prediction of a DIS event can be factorised into four main parts.

- The stage which deals with the partonic structure of the proton, f.
- The analytical calculation of the matrix elements for the hard scattering process $\hat{\sigma}$, at LO in α_s, where the underlying physics process takes place to produce partons.
- The parton shower $\mathcal{P}$ which simulates higher order QCD effects which are missing from the hard scatter calculation.
- A phenomenological model $\mathcal{H}$ to describe the evolution of the partons into observable hadrons, a process termed hadronisation.

The final result for any observable is then given by the convolution

$$\sigma = f \otimes \hat{\sigma} \otimes \mathcal{P} \otimes \mathcal{H}. \tag{6.1}$$

6.2.1 Parton Distribution Function (PDF)

The partonic distributions inside the proton are not known a priori from theory. However, if the distribution is measured at a particular scale, $f(x, Q_0^2)$, the DGLAP approach enables them to be evolved up to different scales, $f(x, Q_i^2)$ [14][59]. Several PDFs exist, based on fits to measured data for the proton structure function, F_2 [60], and other data. The PDFs used in the generator programs employed in this analysis are the CTEQ-4D distributions [61], which describe well the fitted data.

6.2.2 Hard Scatter

The hard subprocess, $\hat{\sigma}$, can, in principle, be exactly calculated to any order in perturbative QCD (pQCD); in DIS, calculations currently exist up to

$\mathcal{O}(\alpha_s^2)$. However, these results need to be matched to the parton showers and this has currently been implemented only up to $\mathcal{O}(\alpha_s)$, corresponding to the QCD-Compton and Boson-Gluon-Fusion (BGF) final state diagrams, given in Figure 1.2.

The generator program LEPTO [62] is based on the Standard Model and includes the full electroweak cross sections calculated to $\mathcal{O}(\alpha_s)$. Higher orders are implemented via parton showers as described in the next section. QED radiation effects are simulated by the program HERACLES [63], which simulates both single photon emission from the lepton and quark line, and the complete virtual one-loop electroweak corrections.

The combination of LEPTO and HERACLES is called DJANGOH [63], and is the primary Monte Carlo used in the analysis. The secondary Monte Carlo used in the analysis, HERWIG [64], also has a full database of electroweak hard subprocesses, although it currently lacks any treatment of QED radiative effects.

6.2.3 Colour Dipole Model and Parton Showers

There are two approaches used to simulate higher order effects. The first approach, employed in DJANGOH, uses the Colour Dipole Model (CDM) [65] as implemented in ARIADNE [66]. In this model, the struck quark is treated as point-like and the proton remnant as an extended object. A colour dipole is formed between them which can radiate a gluon, thereby splitting into two dipoles, one between the struck quark and the radiated gluon, and the other between the radiated gluon and the proton remnant. Further emissions can occur from these two dipoles, forming an extra dipole with every emission. The CDM is a good approximation when the emissions are ordered in p_T^2, therefore ARIADNE imposes ordered emissions, with $p_{T1}^2 > p_{T2}^2 > p_{T3}^2 > \ldots$.

When the diminishing p_T of the colour dipoles reaches a specified cut-off (around $Q_0 = 0.6$ GeV), the emissions stop, and the partons thus created form the partonic final state.

The second approach, employed by HERWIG (and the MEPS option in DJANGOH) is to use parton showers. These use the $\mathcal{O}(\alpha_s)$ diagrams from the hard subprocess as a starting point for initial state (proton) and final state (quark) QCD radiation, termed respectively Initial State Parton Showers (ISPS) and Final State Parton Showers (FSPS). The probabilities for the initial and final state parton showers can be calculated from splitting functions if the hard scale Q, and the initial parton energies are known. For the initial state parton shower, these probabilities cannot be explicitly calculated at the hadronisation scale, but are evolved up from the hard scale to the hadronisation scale using DGLAP evolution equations [14][59]. The final partonic system is then passed to the hadronisation model.

6.2.4 Hadronisation

There are also two possible hadronisation models. The first, as used in DJANGOH, uses the program JETSET [67] to implement the Lund String Model. This model assumes the property of linear confinement. It models the colour field between two partons moving apart as the stretching of a linear flux tube. The tube stretches until it becomes thermodynamically favourable for it to break into two or more hadrons. The breaking is performed in such a way that the resulting hadrons are colourless and conserve charge.

The second method, as implemented in HERWIG, assumes that the cluster hadronisation model used is local in colour and independent of the hard process and the energy. After the parton showering, all the outgoing gluons are split non-perturbatively into quark-antiquark pairs or diquarkantidiquark pairs. Each quark can then be colour connected to an anti-quark or di-quark pair to form colourless objects ('clusters') which then decay isotropically into hadrons.

6.2.5 Diffractive Contribution

There is a class of DIS events observed at HERA called diffractive events [68][69] which are characterised by a large rapidity gap between the main hadronic system and a second hadronic system deposited in the forward region. The model used to describe these events from Ingelmann and Schlein [70] involves a proton coupling to a spacelike object called a *pomeron* with its own partonic structure. The virtual photon from the lepton then probes the partonic structure of the pomeron rather than the proton directly.

Diffractive events can be characterised by the pseudorapidity of the furthest forward hadronic deposit, η_{max}. Figure 6.1a) shows η_{max} compared to ARIADNE Monte Carlo. A clear deficit can be seen in the forward region, $\eta_{max} < 2.5$. It is possible to run ARIADNE with the diffractive processes included; however, this has been shown not to describe well the diffractive events observed at HERA; instead it is preferable to use RAPGAP [71] to describe the diffractive DIS process, while using ARIADNE to describe all other DIS processes.

The distribution in Figure 6.1a) is used to fit the fraction of RAPGAP events to be included in the overall Monte Carlo sample. Figure 6.2 shows the results of a Kolmogorov fit between the measured data and the Monte Carlo sample, as a function of the percentage of RAPGAP in the Monte Carlo sample. The best fit is at approximately 14%, so this is the fraction used in the analysis. The final distribution of η_{max} from this combined sample is shown in Figure 6.1b). The region $\eta_{max} < 2.5$ is better fitted by the Monte

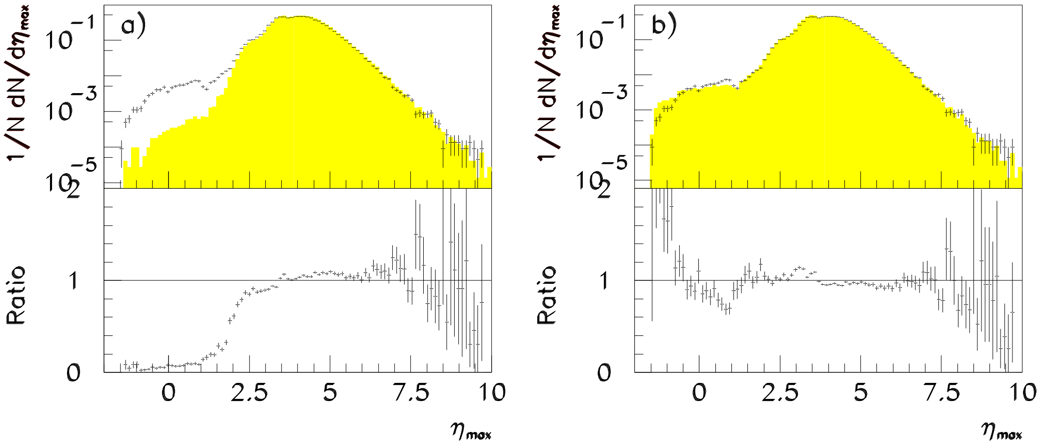


Figure 6.1: a) The η_{max} distribution from ARIADNE alone. b) The η_{max} distribution with 14% of RAPGAP added to the ARIADNE sample. The lower half of the plots show the ratio of data to Monte Carlo.

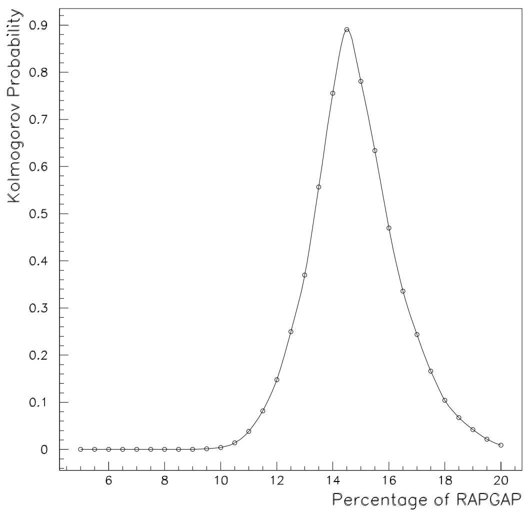


Figure 6.2: The Kolmogorov probability of fitting the η_{max} distribution, as a function of percentage of RAPGAP in the overall Monte Carlo sample.

Carlo, although is still not perfect, with a 10-20% excess in the very forward region, $\eta_{max} < 0$, [1].

The data that is fitted to the power corrections is an inclusive DIS sample, so it will contain approximately 14% diffractive events. Neither the NLO calculation or the power corrections being tested contain any diffractive component. The event shapes are defined in the current region and so they should have minimal sensitivity to the diffractive process which takes place in the target region. The residual sensitivity of the event shapes to the diffractive component, when it is close in rapidity to the current region, is minimal and well within the quoted systematic errors [72].

6.3 Detector Simulation

The hadrons which are produced from the final state fragmentation process are passed directly to the Monte Carlo simulation of the ZEUS detector. In ZEUS, this is performed by a program called MOZART, which uses version 3.13 of the GEANT simulation tool [73]. This contains a detailed simulation of all detector components and a physics database of all possible physics processes that each final state particle can undergo. Each particle in the final state is tracked, from the interaction point, through all the detectors in ZEUS, possibly decaying into two or more lighter particles. The decays are statistical in nature, therefore a large number of events is required to model adequately the detector response and all possible decay modes.

Particles with a very short lifetime (most strongly decaying particles) are decayed inside the fragmentation program JETSET. However, the detector simulation GEANT is left to decay e.g. π, Λ and D particles, since these

[1] Corresponding to low mass diffractive states. e.g. ρ, ω, and ϕ.

Year	Generator Program	Q^2_{true} cut	No. of events (K)	Proton PDF
1995	ARIADNE Radiative	$Q^2 > 7$	660	CTEQ-4D
	HERWIG	$Q^2 > 7$	590	CTEQ-4D
	RAPGAP Diffractive	$Q^2 > 6$	200	CTEQ-4D
1996	ARIADNE Radiative	$Q^2 > 70$	1600	CTEQ-4D
&	HERWIG	$Q^2 > 70$	1600	CTEQ-4D
1997	LEPTO	$Q^2 > 70$	800	CTEQ-4D
	RAPGAP Diffractive	$Q^2 > 70$	600	CTEQ-4D

Table 6.1: The number of Monte Carlo events generated in each sample.

decays have a longer lifetime and often occur somewhere within the volume of the detector.

The detector simulation evolves from year to year, since each year modifications are made to the detector. During the 1996 - 1997 run period, no major modifications were made to the detector. However, between 1995 and 1996, the vertex detector was removed, subsequently reducing the amount of dead material the particles must transverse before entering the CTD. Additionally, new components were added into the BCAL, resulting in changes in the dead material map of the detector, and the RCAL was moved a few centimetres to accommodate new detectors. These modifications had to be updated in the Monte Carlo simulation. Because of this, the Monte Carlo samples are split; one for the 1995 running period and one for the 1996 - 1997 running period. Table 6.1 shows the different Monte Carlo samples used in the analysis.

The primary Monte Carlo used was DJANGOH v.1.1. This used LEPTO v.6.5.1 for the hard scatter, HERACLES v.4.6.1 for the QED radiation, ARIADNE v.4.08 to simulate the Colour Dipole Model, and JETSET 7.410 for the fragmentation. The ARIADNE included the high-Q^2 phase-space modification, as proposed in [74]. 14% of the final sample was RAPGAP which used HERACLES, ARIADNE and JETSET internally. The parameterisation of the pomeron used was that obtained by the H1 Collaboration using a fit to their data [69]. In the remainder of the thesis, this combined Monte Carlo model is referred to as 'ARIADNE'.

The secondary Monte Carlo, used as a systematic check, was HERWIG v.5.9 which simulates all the processes itself, except for the diffractive process and QED radiation. A 14% sample of RAPGAP was added to HERWIG to take account of diffractive effects. HERACLES cannot yet be interfaced to HERWIG, so QED radiation effects are absent from this generator. The effect of these missing corrections is examined in chapter 7.

A cross-check is made with a third Monte Carlo, DJANGOH using MEPS parton showering. This differs from the DJANGOH ARIADNE only in that it uses a parton showering approach rather than the Colour Dipole Model. This model will be referred to as 'LEPTO'. It was only generated for the 1996-1997 run period.

6.4 Summary

The theoretical models used in the simulation of DIS events have been described. Three models are used in the analysis. The primary Monte Carlo is DJANGOH with the ARIADNE (CDM) model. The model dependent systematic check is performed with HERWIG, and a subsequent cross-check performed with DJANGOH using MEPS parton showering. For each, a 14% fraction of RAPGAP is used to describe diffractive processes. The predictions from the models were passed into a full description of the ZEUS detector in order to determine the resolutions and biases of the apparatus.

Chapter 7

Event Shape Measurement

In this chapter, the thrust, broadening, jet-mass and C-parameter event shape distributions are presented. The purities and efficiencies are given and the bin-by-bin correction technique discussed. Systematic checks on the measurements are discussed and their effects presented.

7.1 Differential Distributions

The event shapes are measured differentially as a function of (x, Q^2) and each of the event shapes. The differential bin sizes represent a balance between having as many bins as possible, whilst maintaining reasonable efficiencies and purities, as well as statistics. The same bins are chosen for each of the (x, Q^2) bins, although the binning was optimised for the $Q^2 > 80$ GeV^2 bins where the default fits were performed.

Figures 7.1 and 7.2 show typical efficiency plots for an event shape variable, in this case the C-parameter. The purity, $\mathcal{P}$, and efficiency, $\mathcal{E}$, are defined as

$$\mathcal{P} = \frac{\#\text{ events generated } \textit{and} \text{ correctly reconstructed in the } (x,Q^2,F) \text{ bin}}{\#\text{ events reconstructed in the } (x,Q^2,F) \text{ bin}} \tag{7.1}$$

$$\mathcal{E} = \frac{\#\text{ events generated } \textit{and} \text{ correctly reconstructed in the } (x,Q^2,F) \text{ bin}}{\#\text{ events generated in the } (x,Q^2,F) \text{ bin}} \tag{7.2}$$

For $Q^2 > 80$ GeV^2, the efficiency and purity are reasonable, at approximately 30 - 40% and are also reasonably flat. In the $Q^2 < 80$ GeV^2 region, the efficiencies and purities are less than 30%, due to poor resolution on the boost vector at low (x, Q^2). They are limited by migrations between the (x, Q^2) bins and migrations into and out of the measured kinematic region (typically a 60-70% effect), and migrations across the differential bins themselves (typically a 30% effect.)

7.1.1 Uncorrected Differential Distributions

The uncorrected differential distributions for each event shape, and the comparison with reconstructed level ARIADNE and HERWIG, are given in Figures 7.3 to 7.8. For an arbitrary event shape (denoted F), the measured distribution is defined as

$$\frac{1}{N_{\rm rec}} \left(\frac{dn}{dF}\right)_{\rm rec} \tag{7.3}$$

where $N_{\rm rec}$ is the total number of events in the reconstructed kinematic (x, Q^2) bin *before* the energy cut in the current region, and $n_{\rm rec}$ is the number of events in the bin *after* all selection cuts. The thrust and broadening with respect to the photon axis are well simulated by ARIADNE over the whole Q^2 region. HERWIG describes the distributions reasonably but overestimates the fraction of collimated events ($T \to 1$) in the lower Q^2 bins. Thrust

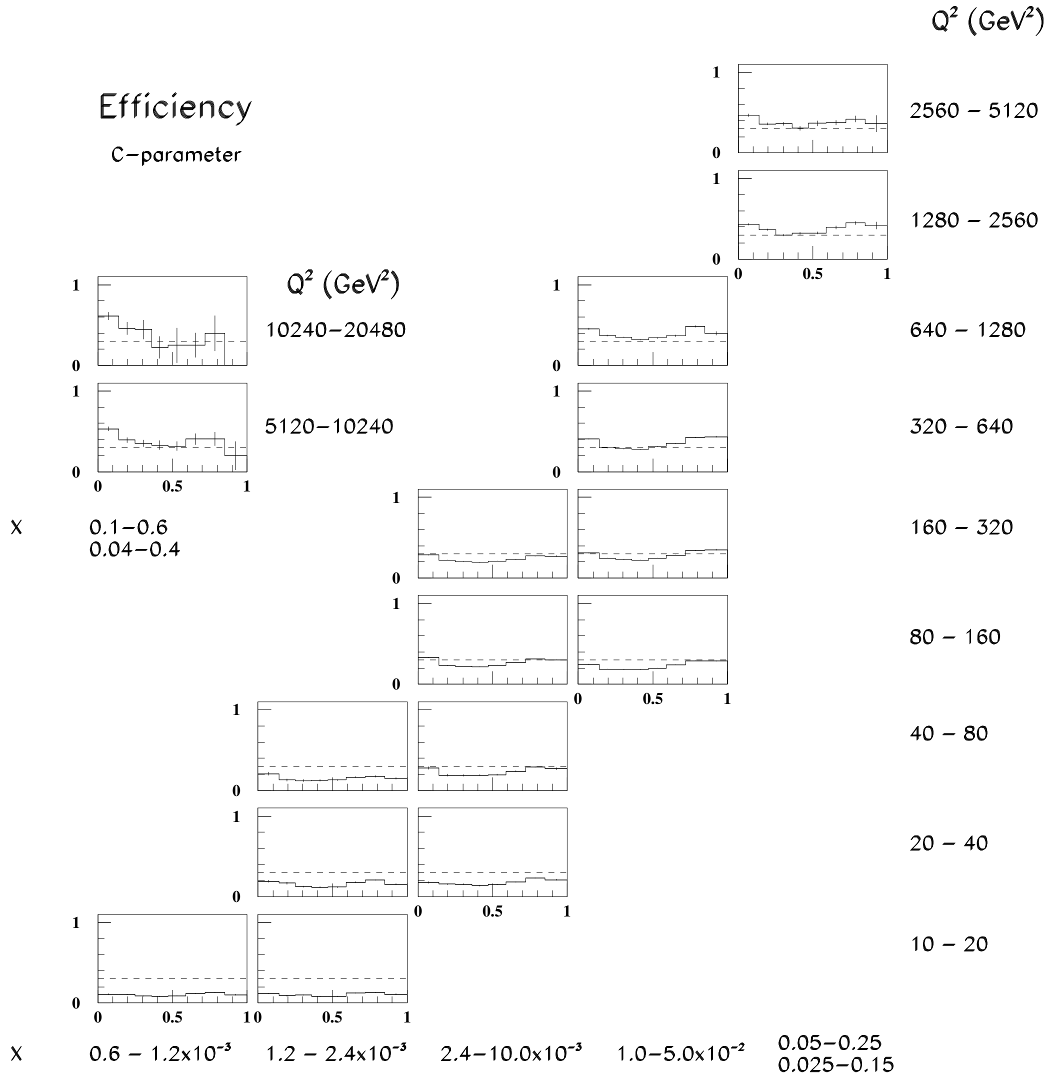


Figure 7.1: Efficiency, $\mathcal{E}$, in chosen differential bins for the C-parameter. The dotted line indicates 30% efficiency.

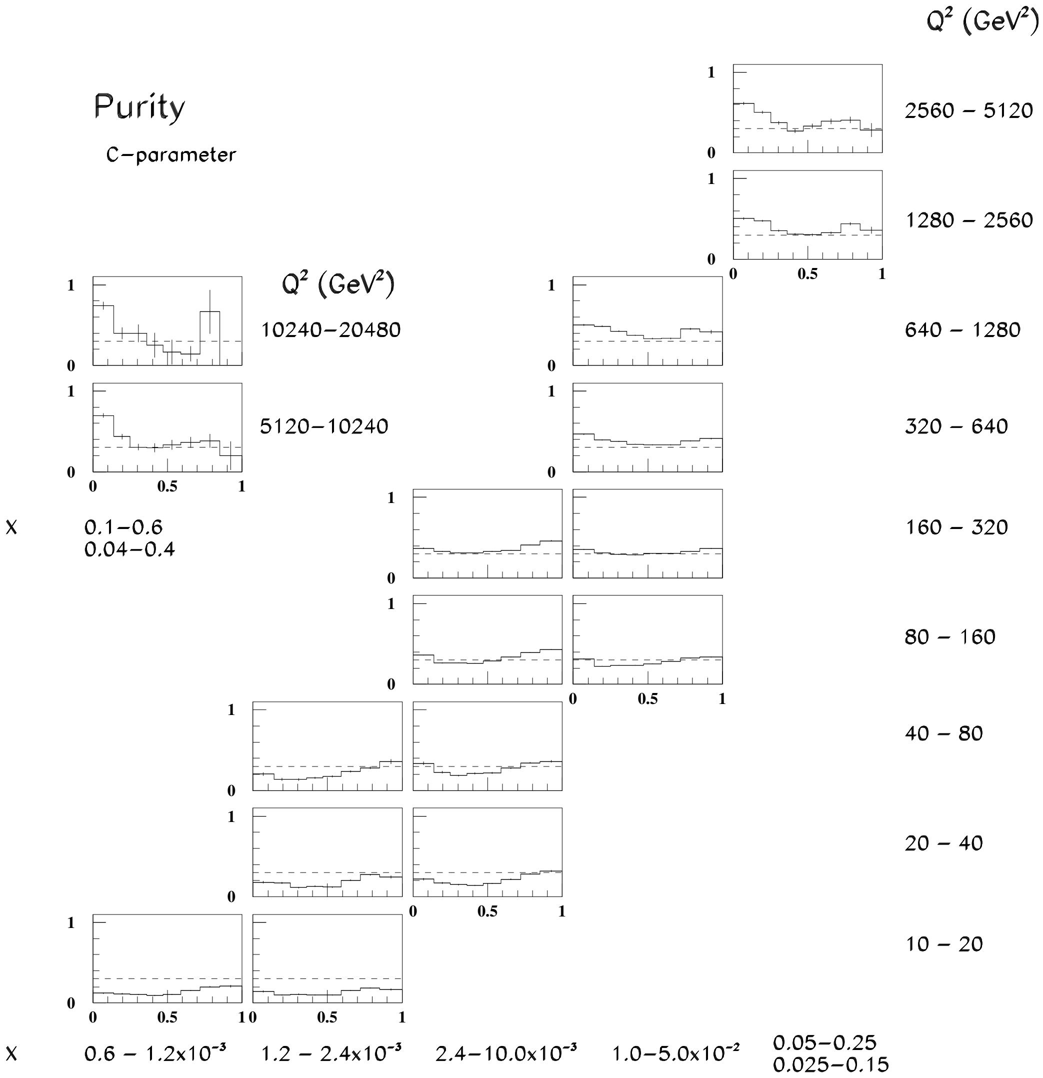


Figure 7.2: Purity, $\mathcal{P}$, in chosen differential bins for the C-parameter. The dotted line indicates 30% purity.

and broadening with respect to the thrust axis exhibit similar characteristics. Both the jet-mass and C-parameter are well simulated by ARIADNE, in particular the phase-space drop off at 3/4 for C. HERWIG poorly represents the shape of the C and ρ_0 distributions, overestimating at low C and underestimating at high C. This indicates that HERWIG has problems modelling the multiplicity of the events.

In conclusion, ARIADNE agrees well with the data and therefore was chosen as the central Monte Carlo for data correction in this analysis. HERWIG agrees reasonably, but has problems reproducing the multiplicity of the events and problems with the shape; data correction with HERWIG is used as a systematic check.

7.1.2 Differential Data Correction

In order to account for the detector and physics effects described below, the differential data are corrected using a bin-by-bin linear correction technique. Correction factors were obtained using the ARIADNE Monte Carlo ratio of the true underlying event, to the event which had been passed through the detector simulation, subject to the same trigger, reconstruction and analysis techniques as the data. The individual bin correction factors are given by

$$\mathcal{K}(F) = \frac{1}{N_{\text{gen}}} \left(\frac{dn}{dF}\right)_{\text{gen}} \bigg/ \frac{1}{N_{\text{rec}}} \left(\frac{dn}{dF}\right)_{\text{rec}} \tag{7.4}$$

with N_{gen} and n_{gen} defined in a similar way as Equation 7.3. Using this method corrects for inefficiencies in the reconstruction of the hadronic final state, migrations between differential bins and between kinematic bins, QED radiative effects, the effect of the selection cuts, and the effect of the difference between what can practicably be measured in the detector and the defined hadronic final state (acceptance effects).

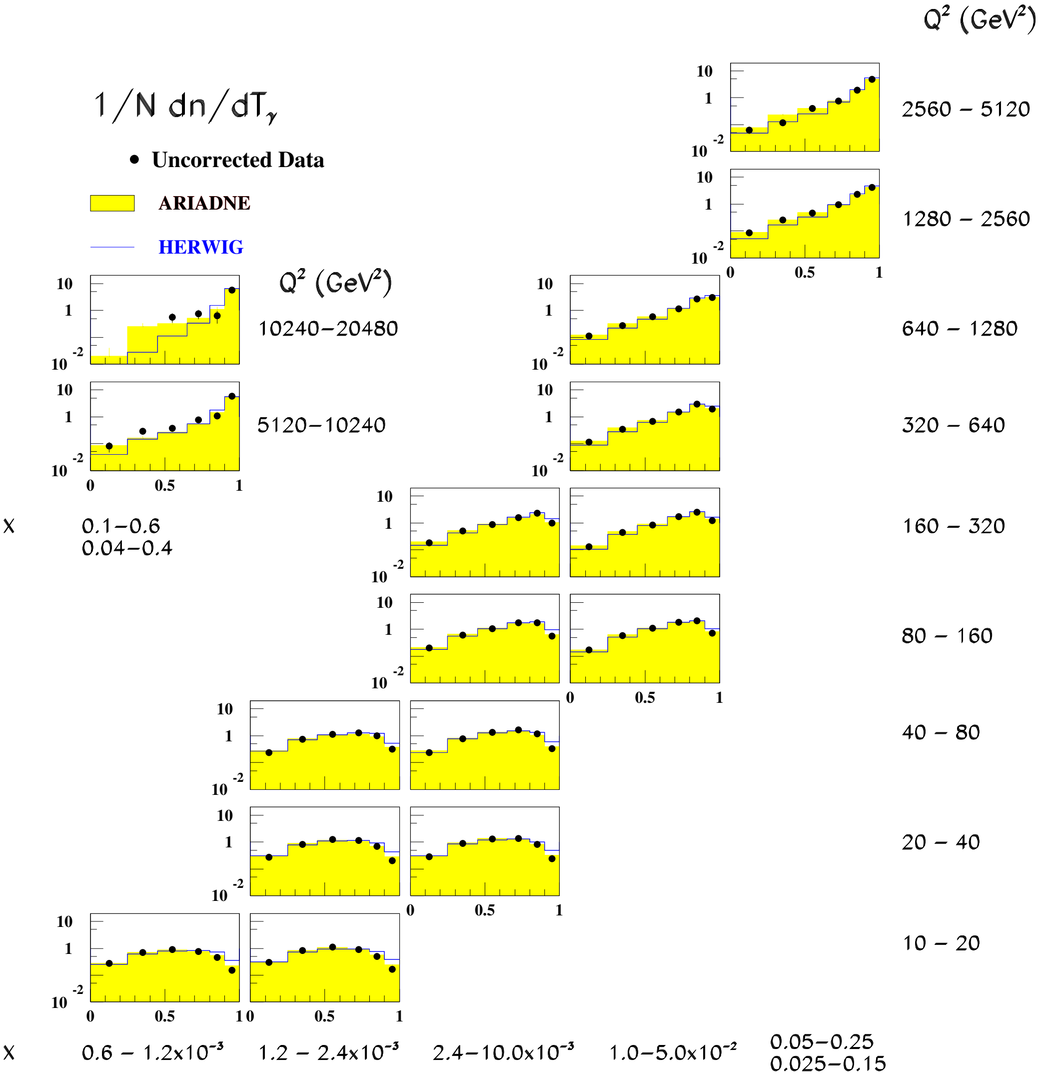


Figure 7.3: Uncorrected differential distributions for thrust with respect to the photon axis, T_γ. The points are data, the shaded histogram is ARIADNE, and the solid line is HERWIG.

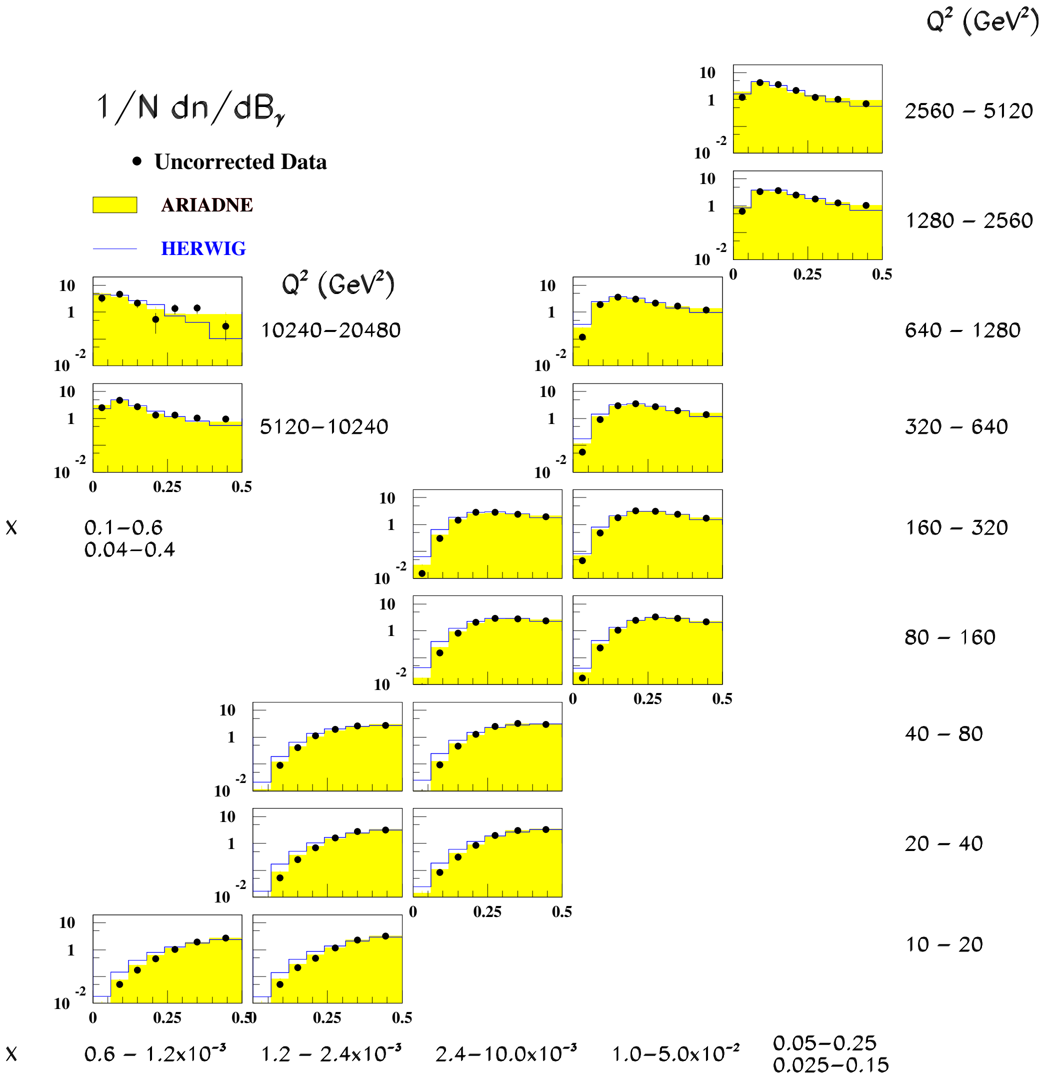


Figure 7.4: Uncorrected differential distributions for jet broadening with respect to the photon axis, B_γ. The points are data, the shaded histogram is ARIADNE, and the solid line is HERWIG.

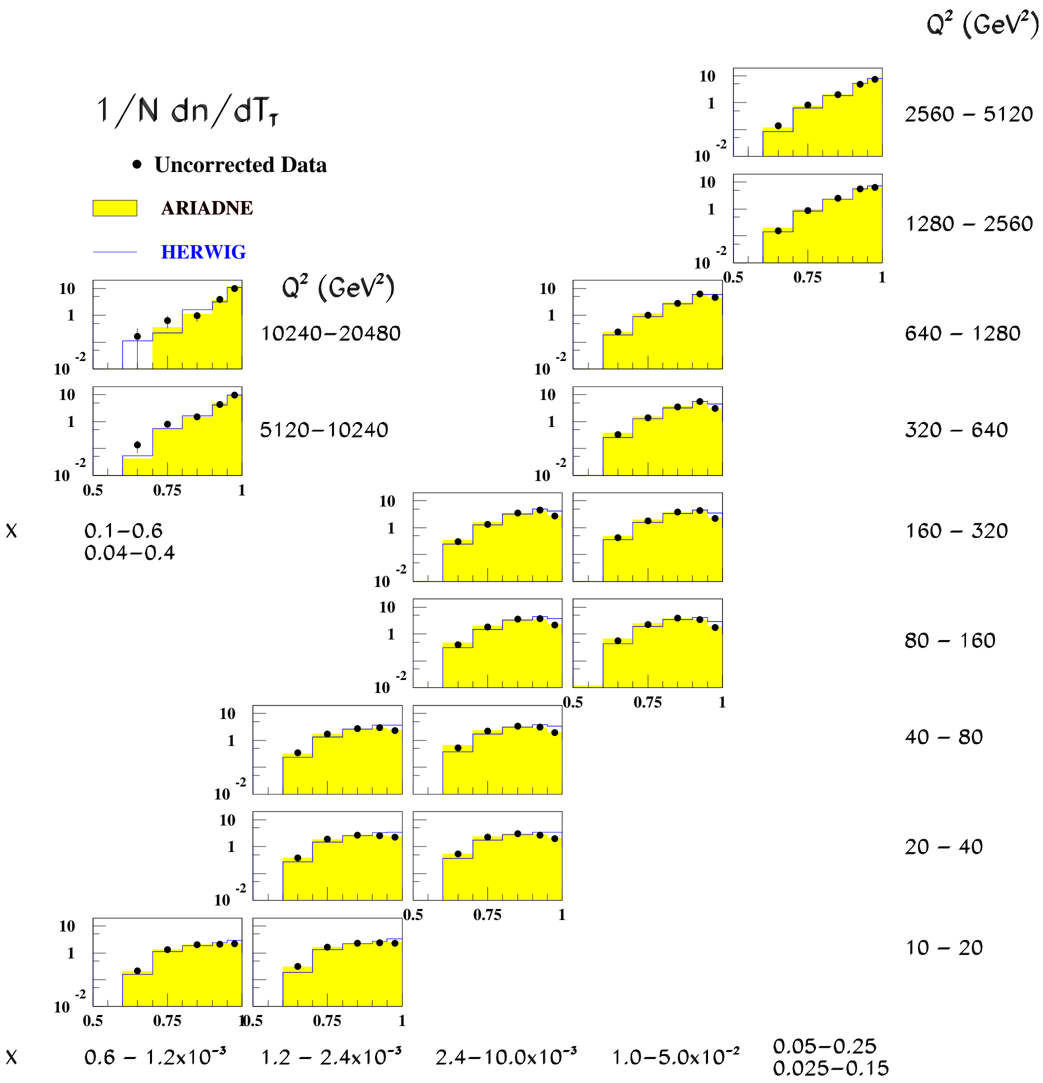


Figure 7.5: Uncorrected differential distributions for thrust with respect to the thrust axis, T_T. The points are data, the shaded histogram is ARIADNE, and the solid line is HERWIG.

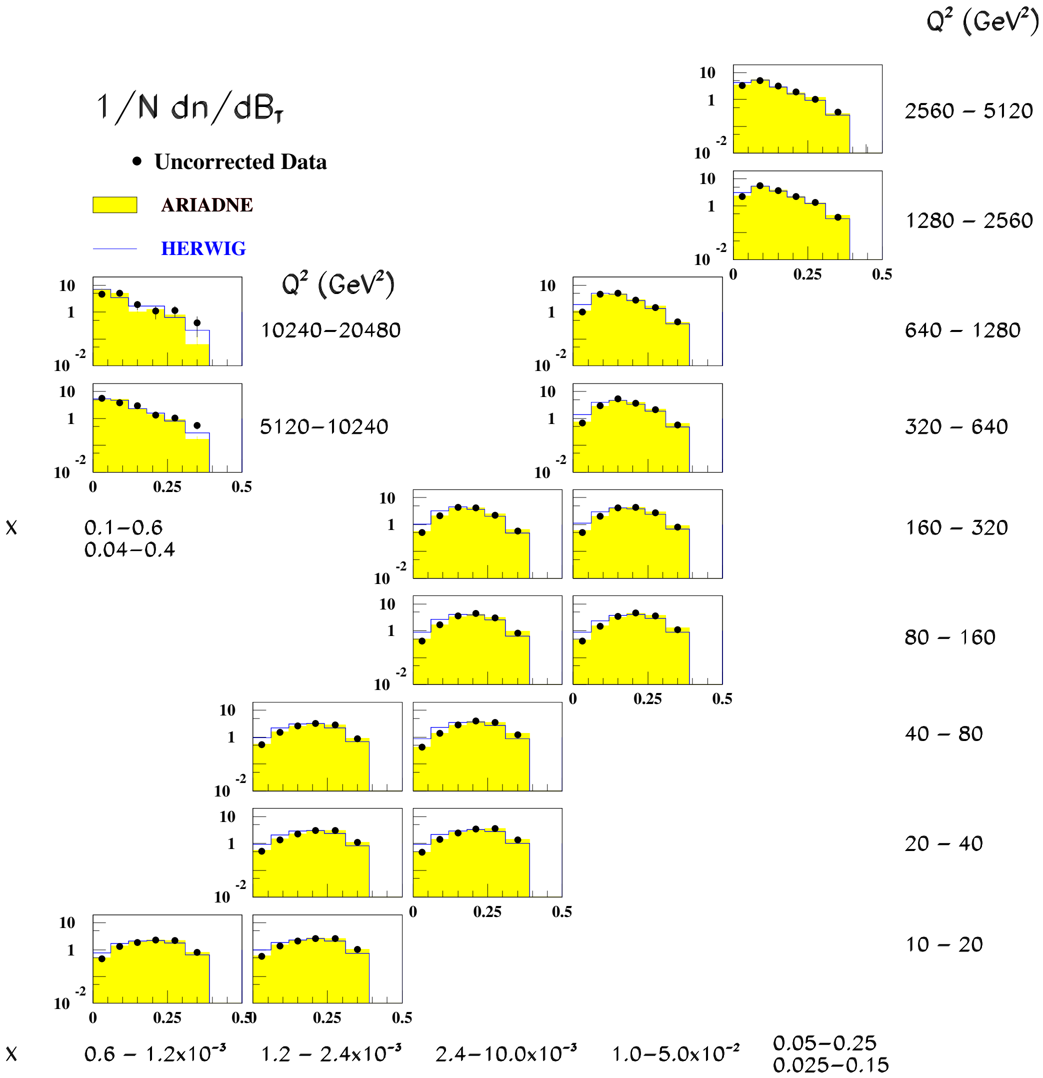


Figure 7.6: Uncorrected differential distributions for jet-broadening with respect to the thrust axis, B_T. The points are data, the shaded histogram is ARIADNE, and the solid line is HERWIG.

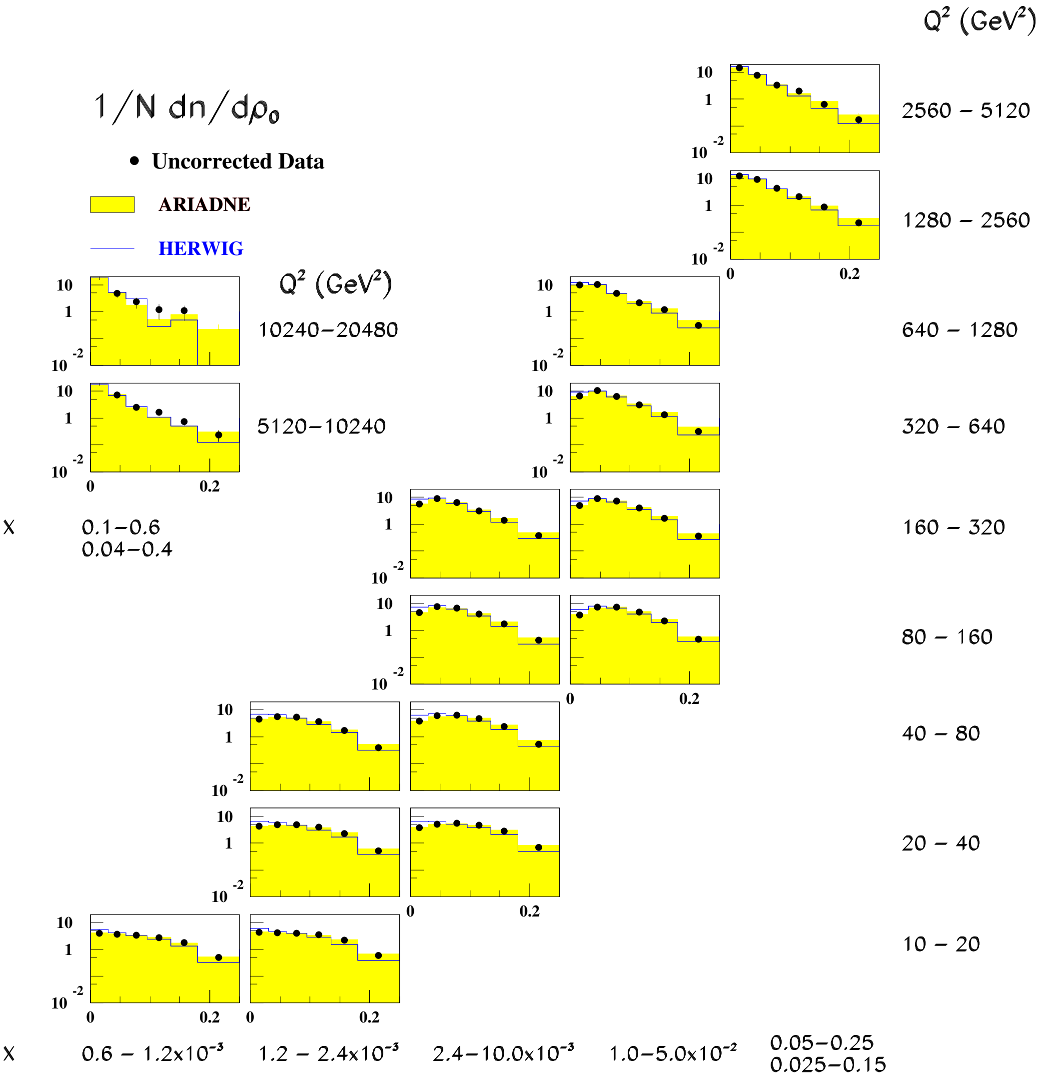


Figure 7.7: Uncorrected differential distributions for jet mass, ρ_0. The points are data, the shaded histogram is ARIADNE, and the solid line is HERWIG.

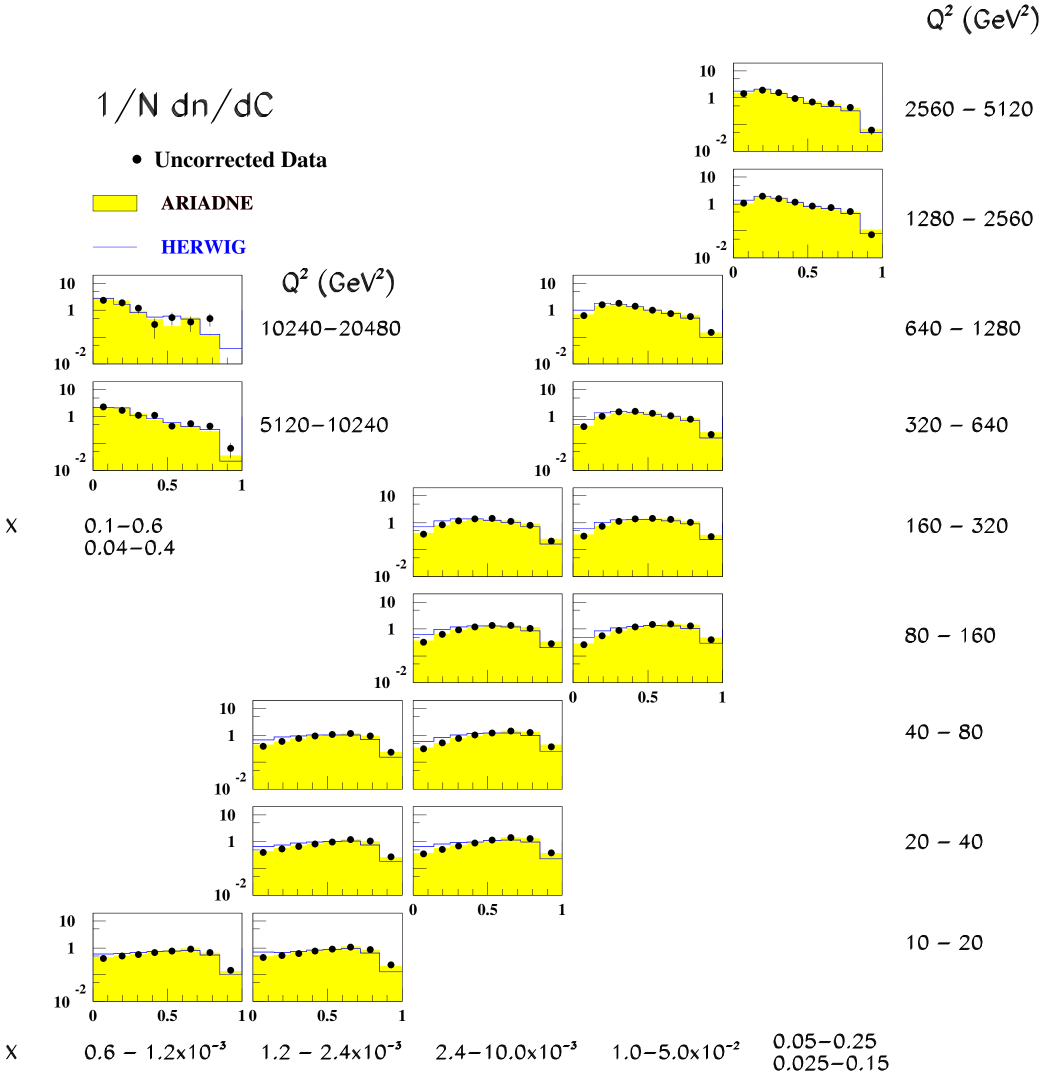


Figure 7.8: Uncorrected differential distributions for C-parameter. The points are data, the shaded histogram is ARIADNE, and the solid line is HERWIG.

For the bin-by-bin correction technique to be applicable, the correction factors should be close to unity. Figure 7.9 shows the bin-by-bin correction factors for the C-parameter. In the intermediate range of Q^2 ($80 < Q^2 < 5120$) the correction factors are all close to unity and reasonably flat. At very high Q^2, the factors are much more variable due to poor statistics in the Monte Carlo and the fact that the distribution is falling more sharply. At low Q^2 and high y, the factors exhibit some functional dependence, becoming larger as $C \rightarrow 1$, the largest factor being $\sim$2. This is due to the fact that the events are broader, so more ZUFOs are lost in the η and p_T cuts, requiring the Monte Carlo to correct for these losses.

The correction factors are then applied to the reconstructed data histograms. The errors on the correction factor are calculated, taking into account migrations into and out of the bins, as discussed in Appendix A. The fractional error on the corrected data is calculated by taking the fractional data error in quadrature with the fractional error on the correction factor.

Figures 7.10 to 7.15 show the data corrected with ARIADNE, compared to hadron level ARIADNE and HERWIG. The systematics errors combined in quadrature, as discussed in section 7.3, are plotted with the data points. At low Q^2, the particle distributions are quite isotropic, exemplified by the low values of thrust and high values of broadening; the distribution for thrust is also quite flat, indicating that that no particular event topology is favoured. As Q^2 increases, the events become more collimated, with a peak developing as T_γ and $T_T \rightarrow 1$. The broadening distributions, B_γ and B_T, exhibit a significant change in shape; at low Q^2, they peak towards $B \rightarrow 1$, indicating broad events. However, as Q^2 increases, the peak moves to a value of $B \sim 0.1$. The jet-mass, ρ_0, decreases as the events become more collimated and the angle between the particles decreases. The C-parameter shows similar

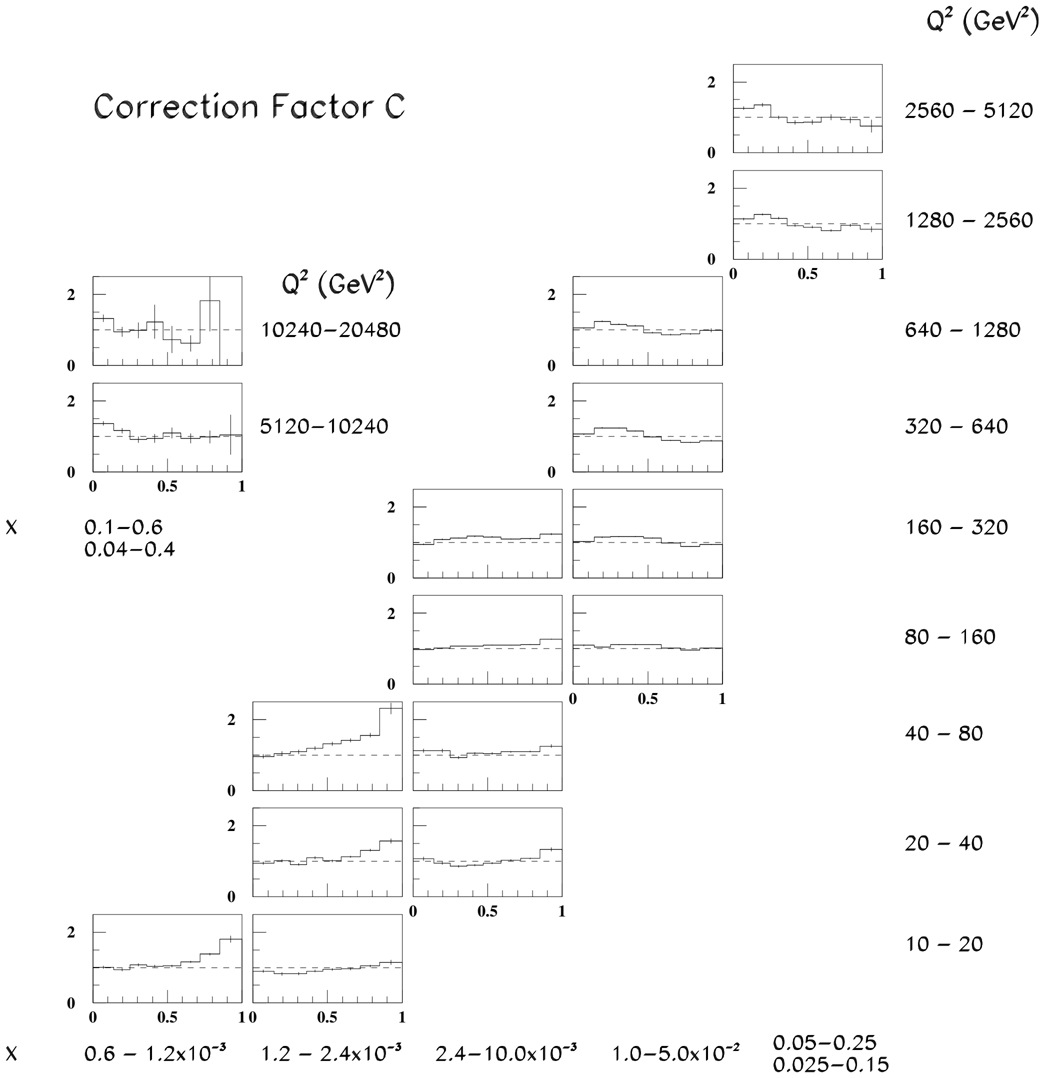


Figure 7.9: Differential bin-by-bin correction factors for the C-parameter. The dotted line indicates a factor of unity.

behaviour as the events becomes more collimated and hence more coplanar; it also exhibits the expected phase-space drop off at $C \sim 3/4$, since values of $C > 3/4$ are only possible when there are four or more particles in the final state.

The agreement with hadron level ARIADNE is good over the full Q^2 range. HERWIG overestimates the fraction of collimated events in the thrust and broadening and poorly represents the multiplicity, resulting in an incorrect shape for the C-parameter and jet-mass distributions.

7.2 Mean Event Shapes

The mean event shapes, denoted as $\langle F \rangle$, are defined as

$$\langle F \rangle = \frac{1}{\sigma} \int_0^{F_{max}} F \frac{d\sigma}{dF} dF. \tag{7.5}$$

and plotted against the mean momentum transfer, $\langle Q \rangle$. Figure 7.16 shows the uncorrected mean distributions for the event shapes compared to reconstructed ARIADNE and HERWIG Monte Carlo.

The data for the means show more clearly that the events become more collimated as Q increases, with $F \to 0$ for all variables (it is for this reason that the thrust T is replaced by $\tau = 1 - T$; so that it displays the same behaviour with Q as the other event shapes). An x-dependence is apparent in the data where there are two points measured over the same Q^2 range but different x ranges. The ARIADNE Monte Carlo represents the data well, including the x-dependence. For all variables, HERWIG Monte Carlo represents the data well in the high Q region. At low Q, it represents the thrust and broadening with respect to the photon axis well, but for the other variables, the shape is not reproduced, with HERWIG being less than the data by up to 20%.

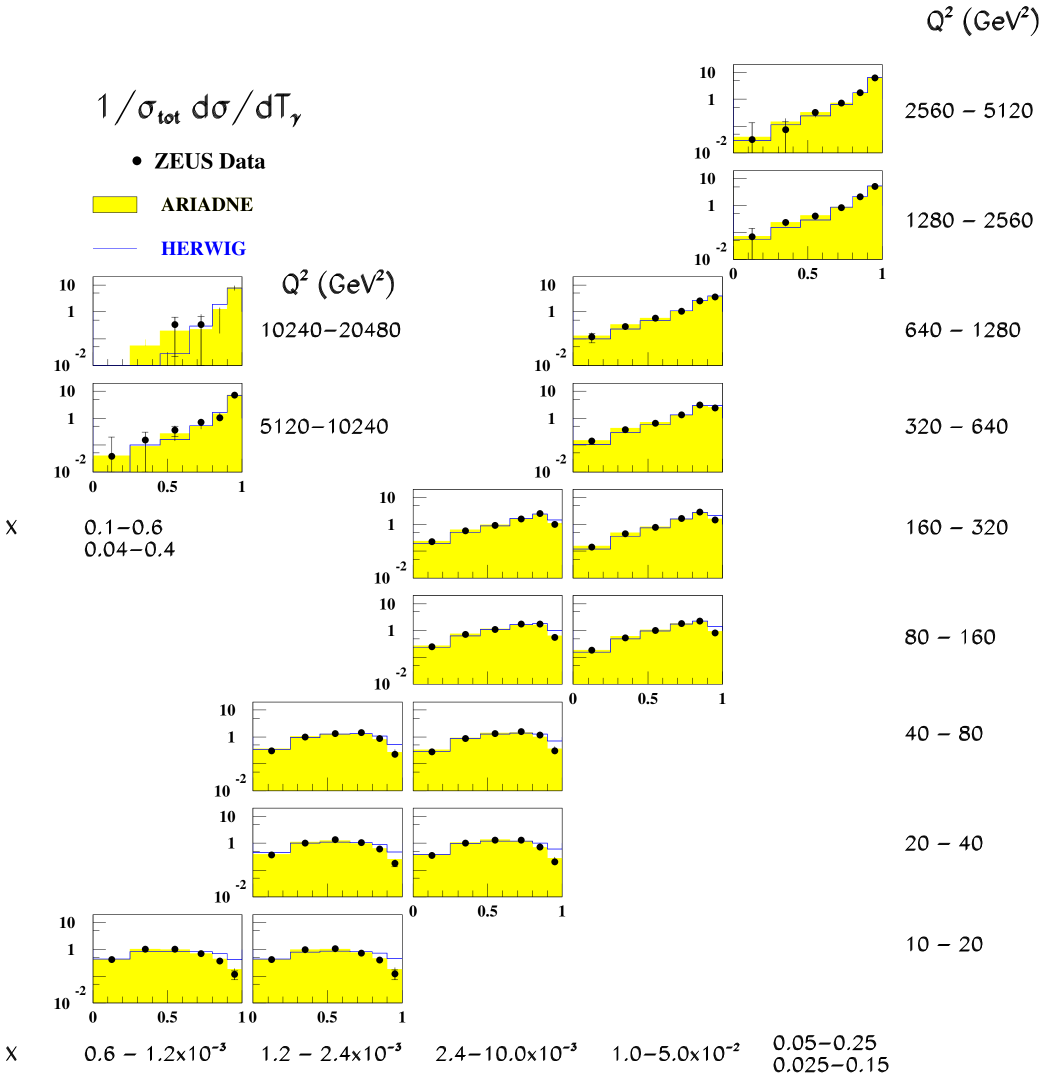


Figure 7.10: Corrected differential distributions for thrust with respect to the photon axis, T_γ. The points are data, corrected with ARIADNE; the outer error bars represent statistical$\oplus$systematic errors, the inner bars represent statistical errors only. The shaded histogram is hadron level ARIADNE and the solid line is hadron level HERWIG.

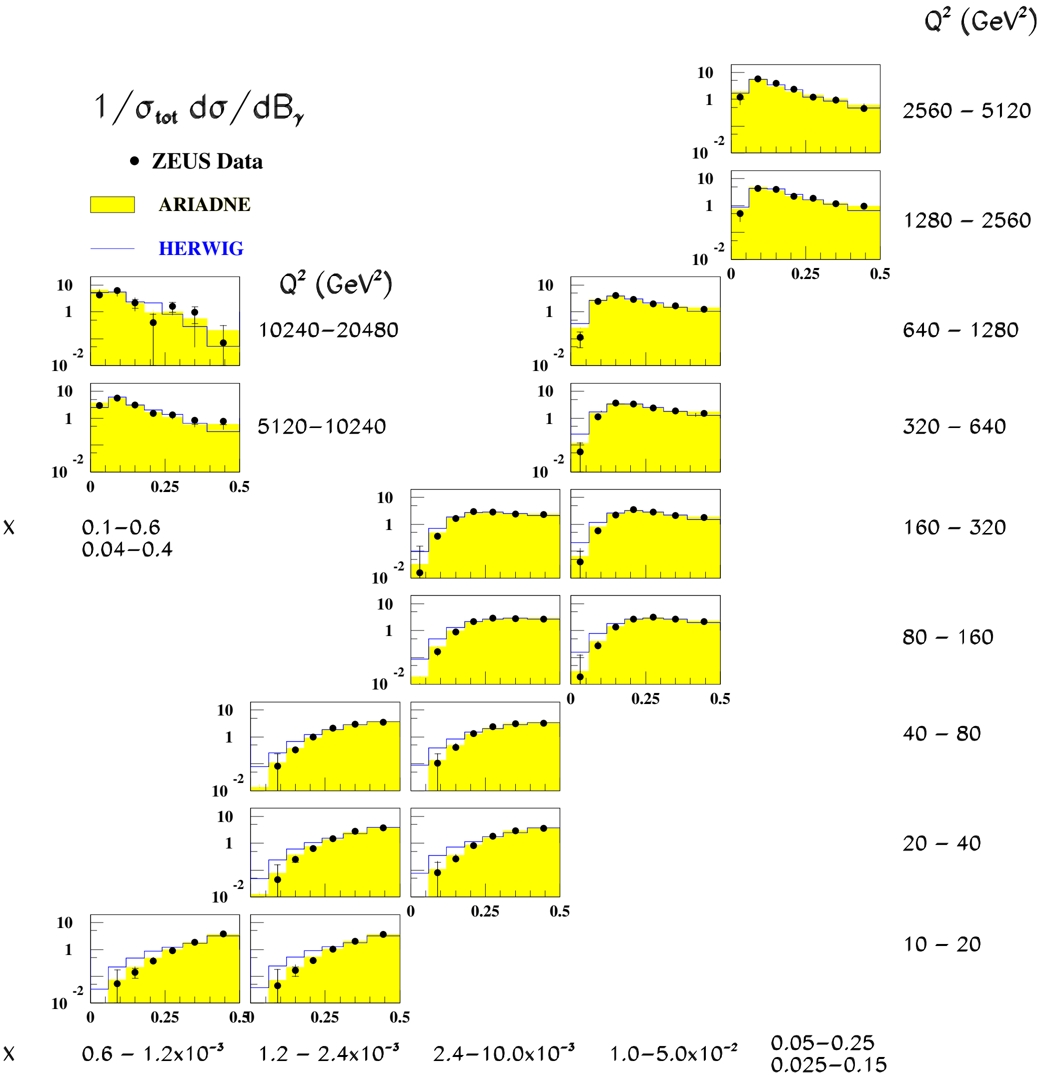


Figure 7.11: Corrected differential distributions for jet broadening with respect to the photon axis, B_γ. The points are data, corrected with ARIADNE; the outer error bars represent statistical$\oplus$systematic errors, the inner bars represent statistical errors only. The shaded histogram is hadron level ARIADNE and the solid line is hadron level HERWIG.

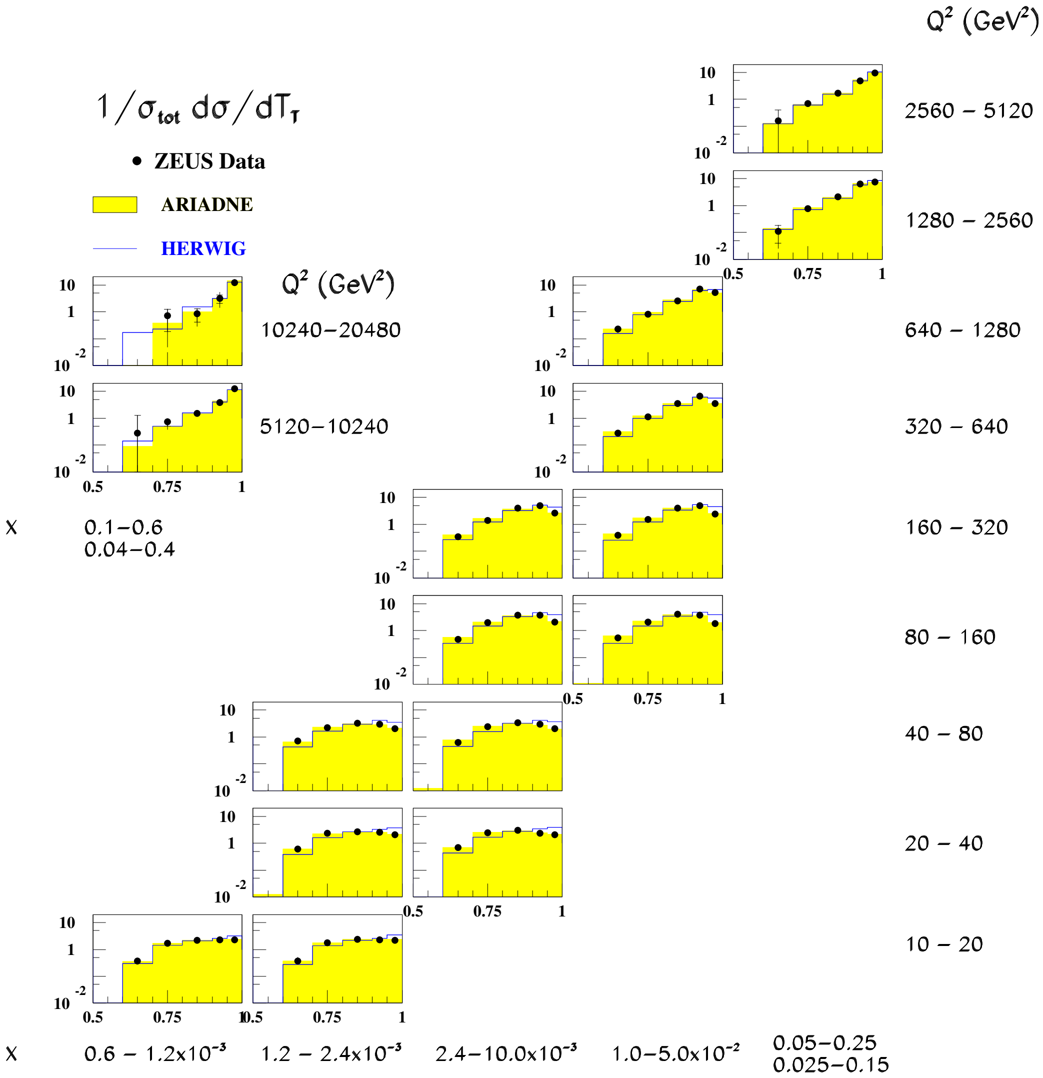


Figure 7.12: Corrected differential distributions for thrust with respect to the thrust axis, T_T. The points are data, corrected with ARIADNE; the outer error bars represent statistical$\oplus$systematic errors, the inner bars represent statistical errors only. The shaded histogram is hadron level ARIADNE and the solid line is hadron level HERWIG.

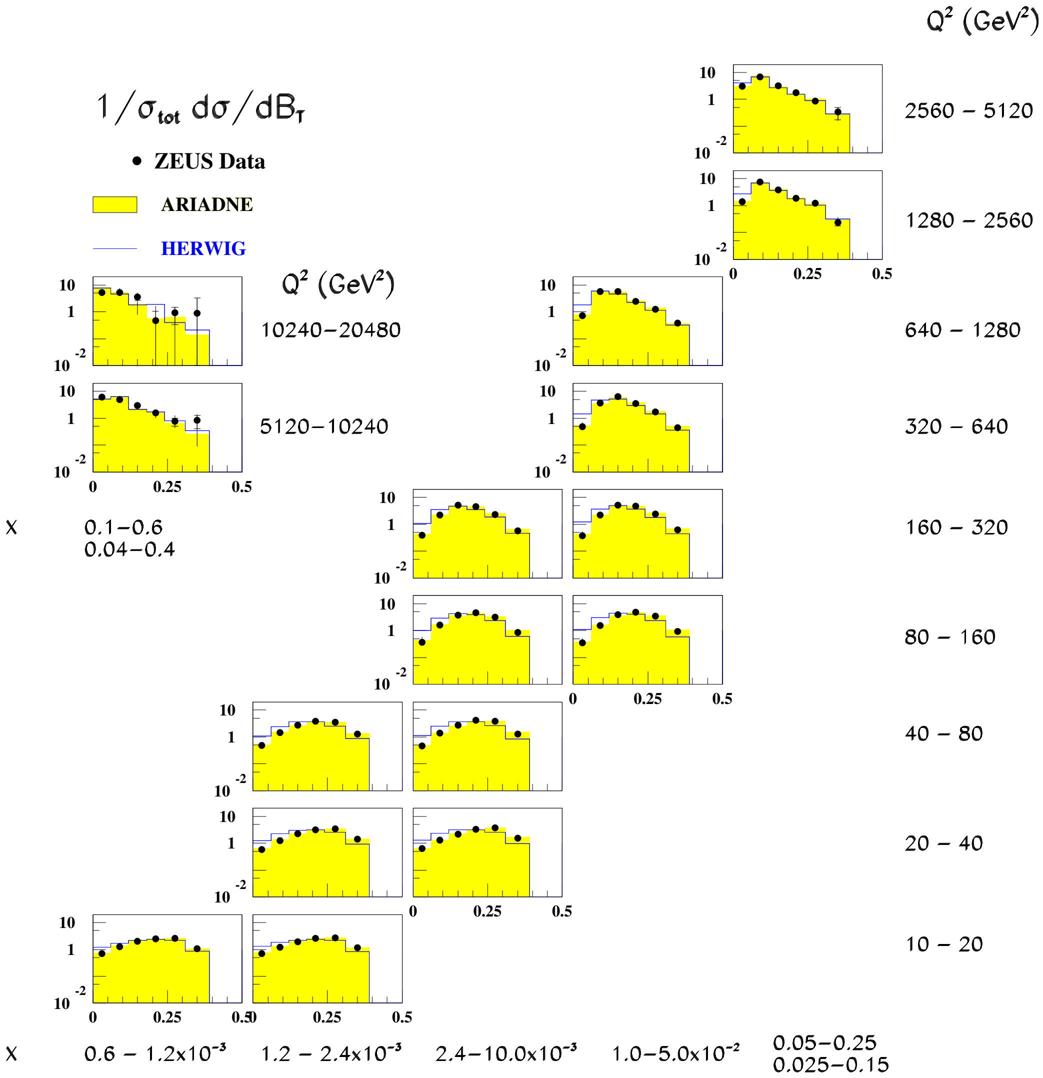


Figure 7.13: Corrected differential distributions for jet-broadening with respect to the thrust axis, B_T. The points are data, corrected with ARIADNE; the outer error bars represent statistical$\oplus$systematic errors, the inner bars represent statistical errors only. The shaded histogram is hadron level ARIADNE and the solid line is hadron level HERWIG.

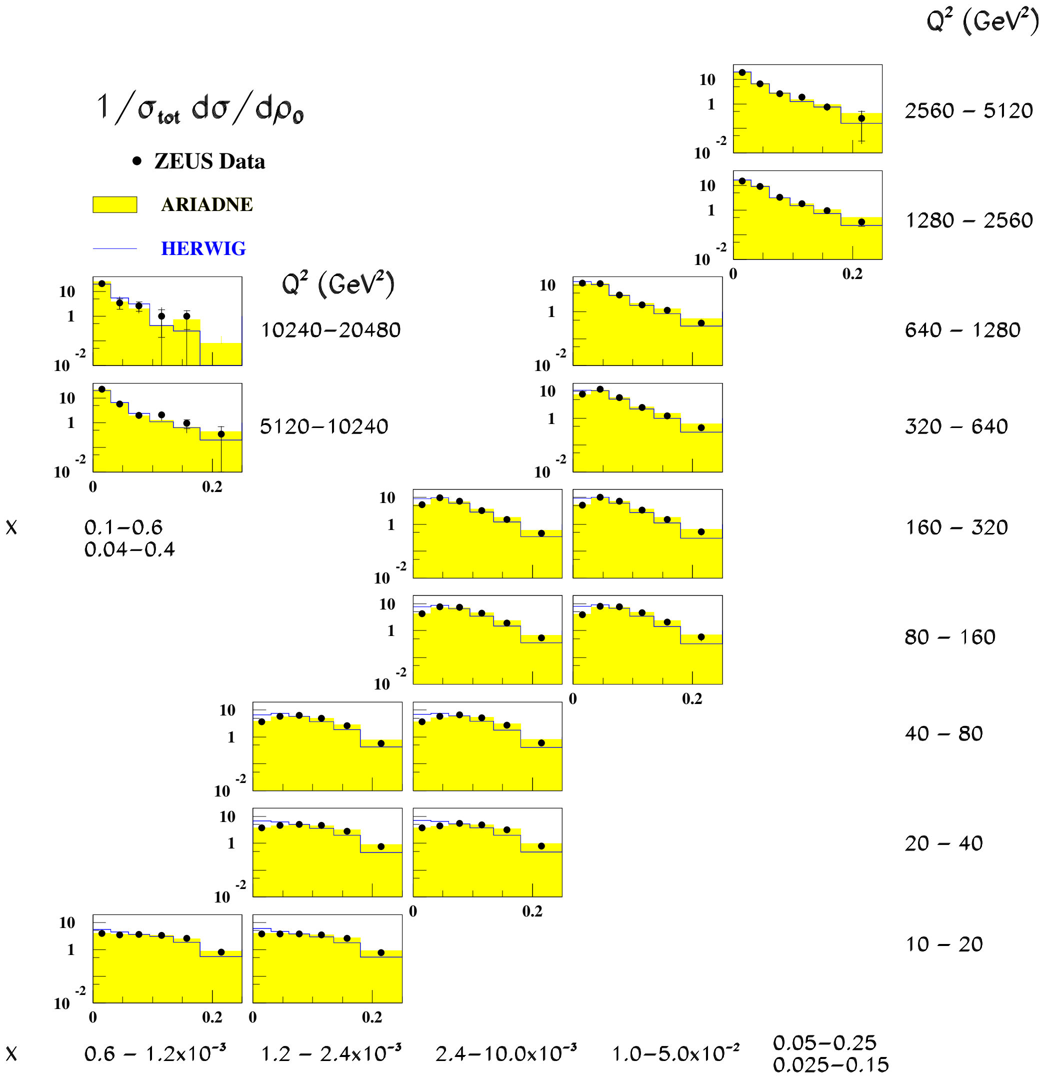


Figure 7.14: Corrected differential distributions for jet mass, ρ_0. The points are data, corrected with ARIADNE; the outer error bars represent statistical$\oplus$systematic errors, the inner bars represent statistical errors only. The shaded histogram is hadron level ARIADNE and the solid line is hadron level HERWIG.

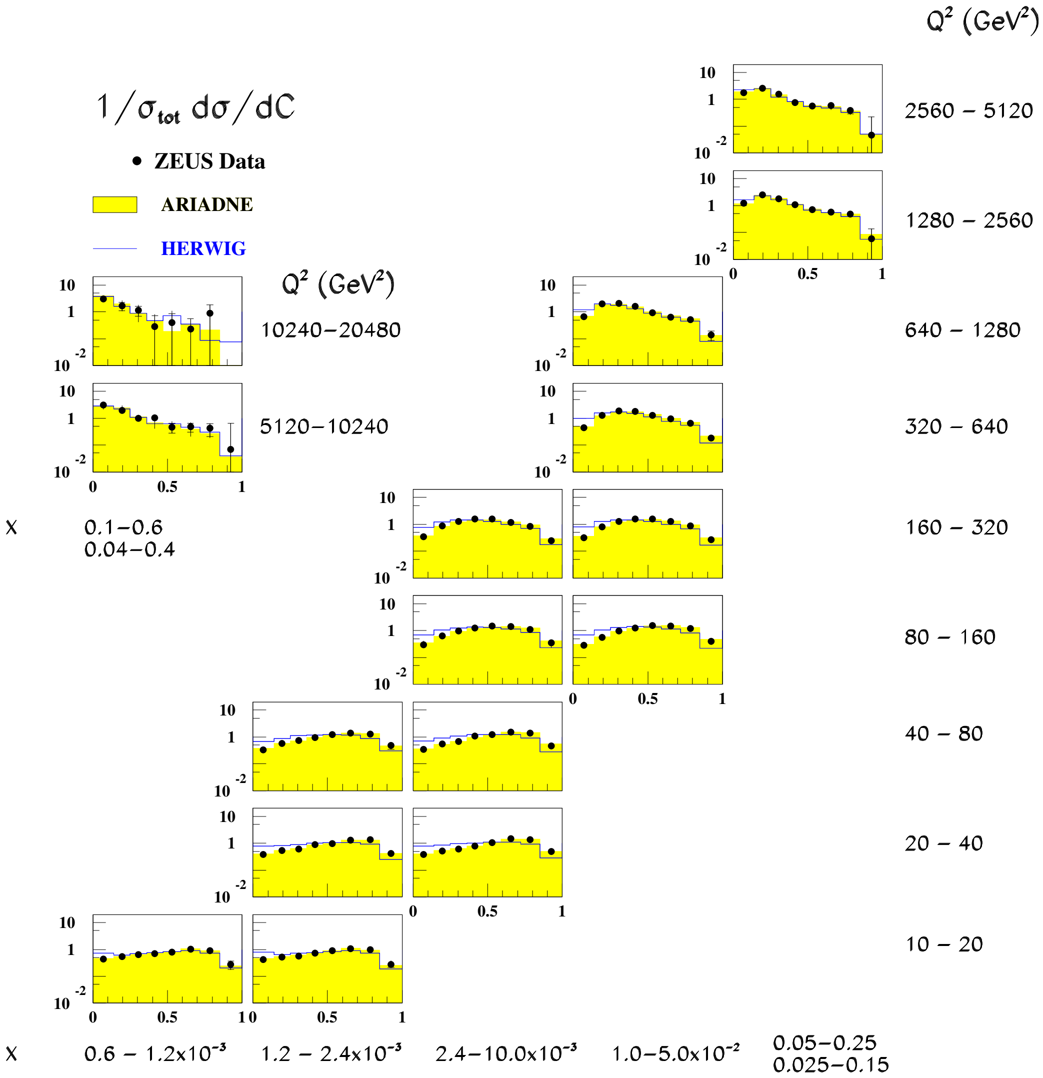


Figure 7.15: Corrected differential distributions for C-parameter. The points are data, corrected with ARIADNE; the outer error bars represent statistical$\oplus$systematic errors, the inner bars represent statistical errors only. The shaded histogram is hadron level ARIADNE and the solid line is hadron level HERWIG.

Correction factors are obtained from the ARIADNE Monte Carlo in a manner similar to the bin-by-bin corrections used in the differential distributions, as a simple ratio between the reconstructed mean and the generated hadron level mean.

$$\mathcal{K}\langle F\rangle = \frac{\langle F\rangle_{\mathrm{gen}}}{\langle F\rangle_{\mathrm{rec}}} \tag{7.6}$$

The corrected mean distributions are given in Figure 7.17, with the correction factors and tables of results given in Appendix B. For the variables that are independent of the photon axis, there is little x-dependence at a given Q, whereas there is a large x-dependence for thrust and broadening with respect to the photon axis. The ARIADNE Monte Carlo generally agrees well over the whole Q range. The HERWIG Monte Carlo agrees well for the thrust and broadening with respect to the photon axis, but fails to represent the data for the other variables by as much as 30% in the lower Q region.

The difference in behaviour between the photon axis variables and the other variables with respect to x is understandable in terms of the initial state parton showers. The QPM assumes zero transverse momentum for the incoming quark; in reality, the parton showers give the incoming quark some transverse momentum with respect to the Breit frame axis, quite independent from the fragmentation shape of the outgoing quark. The phase-space for ISPS and hence the amount of transverse momentum increases as a function of $\ln(1/x)$. This will affect the variables with respect to the Breit frame axis more than the other variables, which are either axis independent (C and ρ_0) or only depend on the fragmentation axis (τ_T and B_T) [1]. Therefore, any changes in x will affect the τ_γ and B_γ most, since these are the variables

[1] The Breit frame axis only affects these variables indirectly through the current region cut-off.

most sensitive to ISPS.

In general, the correction factors on the means, obtained from ARIADNE, are all close to unity (within 15%). This, combined with the high efficiency and purity for each of the (x, Q^2) bins (as given in Table 4.1) indicates that the correction technique used is appropriate. Appendix B gives tables of the mean data and the correction factors.

A comparison with H1 data [30] is shown in Figure 7.18. The H1 data start at $\langle Q \rangle > 7$ GeV and are available for all the variables except B_T. They were measured using an energy cut in the current region of the Breit frame of $\mathcal{E}_{lim} > 0.10Q$. In general, there is very good agreement between the two data sets, although there appears to be a small (few percent) difference in some of the high Q bins, most noticeable in the variables with respect to the photon axis. H1 choose to integrate over x (with $0.05 < y < 0.8$), and generally, the H1 data lie between the two ZEUS points in the region where the x-dependence has been measured.

7.3 Systematic Checks

When making a measurement, it is necessary to consider what the effect of particular choices of cuts or methods made during an analysis have on the final result. Cuts are made to ensure that the analysed data can be well measured, either because of detector limitations, or to reduce contamination from background sources. For the areas of phase space that have been removed by cuts, the Monte Carlo is then trusted to correct back to the defined hadron level, providing it adequately represents the quantity being cut upon. In order to investigate the effect that the value of a cut has on the data, the cut value is conventionally moved by approximately one standard deviation, whilst maintaining the same truth level. Additionally, during an analysis, certain algorithms are applied, in specific orders, with predefined calibration constants. The data are also corrected using a specific Monte Carlo model, and the sensitivity of the corrected data to the model chosen must be investigated. It is impractical to vary every single cut and technique used, so only the major ones are investigated. The following systematic effects were examined.

- $40 < \delta < 60$ GeV. The lower cut on the value of δ is changed from 35 GeV to 40 GeV, since the cut is partly motivated by the necessity to remove photoproduction events and proton beam gas induced events, neither of which is exactly simulated by the Monte Carlo.

- $y_e < 0.8$. This is moved from $y_e < 0.95$. This shift is somewhat larger than the resolution of y_e and is introduced to examine the effect of background photoproduction in the data compared to the Monte Carlo, in the region $y_e > 0.8$.

- $y_{JB} > 0.05$ increased from $y_{JB} > 0.04$ to account for the imperfect simulation of noise in the Monte Carlo.

- $|\eta| < 1.5$ and $p_T > 0.2$ GeV. The p_T and η cuts were tightened from the nominal 1.75 and 0.15 GeV to examine the effect of varying the fiducial cuts for the tracking.

- The p_T and η cuts were removed completely. This examines the effect of using more unmatched islands in the analysis.

- The energy of all the unmatched deposits in the calorimeter was scaled by a factor $\pm$ 3%, 1%, and 2% in the FCAL, BCAL and RCAL respectively. This is the degree of accuracy to which the electromagnetic energy scale is known and corresponds primarily to the scaling of π^0's. The ZUFOs algorithm limits the calorimeter energy scale dependence since the charged component (mostly π^+/π^-'s) will often use the tracking information.

- Backsplash cells included. The default analysis removed all cells considered as backsplash from the hadronic final state; this systematic check examines the effect on the final result of including them.

- Monte Carlo hadronisation model dependence. The analysis of the final state is likely to be sensitive to the hadronisation model used to correct the data. To investigate this, the data were corrected with HERWIG instead of ARIADNE.

For each of the systematic checks discussed above, the entire data and Monte Carlo set is re-analysed, keeping everything the same except for the quantity or method being varied. The positive systematic shifts and negative systematic shifts are combined in quadrature separately.

An additional two cross-checks were performed on the data. These cross-checks are not considered as systematic checks and the final quoted values of systematic errors do not include them.

- An investigation of the electron method for reconstructing the kinematic variables. Both the (x, Q^2) and electron four-vector were determined using the corrected electron energy and angle alone. This method was not used as a direct systematic check since it is known to suffer from poorer intrinsic resolution[2] using the ZEUS detector.

[2]Therefore the choice of the DA method is clearly motivated and not arbitrary. In general, all of the systematics considered are those for which the choice for their use contained some degree of arbitrariness. A method which has a clear motivation and clear

Number	Systematic	Cross-Check
1		Correction using LEPTO
2	Correction using HERWIG	
3	Backsplash included	
4	CAL energy scale up	
5	CAL energy scale down	
6	$40 < \delta < 65$ GeV	
7	$\|\eta\| < 1.5$, $p_T > 0.2$ GeV	
8	No η and p_T cuts	
9	$y_e < 0.8$	
10	$y_{JB} > 0.05$	
11		Electron method kinematics

Table 7.1: List of systematics checks and cross-checks, numbered as presented in Figures 7.19 - 7.24.

- A check of the parton showering versus Colour Dipole Model approach. The data were corrected using DJANGOH with the MEPS option ('LEPTO'), instead of the nominal DJANGOH with the ARIADNE option. Due to limited Monte Carlo availability, this was only checked for the 1996-1997 data. Both the LEPTO and HERWIG Monte Carlos are based on the parton showering model, therefore only one, HERWIG, was used as a systematic check; LEPTO was used solely as a cross-check.

The percentage systematic deviations on the mean values of the event shapes are shown in Figures 7.19 to 7.24. The two cross-checks are also presented. The systematic numbering scheme on the plots is given in Table 7.1. The dotted lines on the plots show the fractional statistical uncertainty on the central mean value. The systematics dominate the errors on the analysis in all but the highest and second highest Q^2 bins.

superiority to alternative methods need not be systematically varied.

The dominant systematic shift on all the variables, over the entire kinematic range, is from the model dependence due to HERWIG (systematic number 2). For the variables with respect to the photon axis (τ_γ and B_γ), this is an approximately 1-5% effect, with the shift increasing to over 10% for the highest Q^2 bins, where the Monte Carlo statistics are more limited. The shift is 5-6% in the thrust and broadening with respect to the thrust axis, and 5-10% in the C-parameter and jet-mass. Some difference between ARIADNE and HERWIG is expected, since it is already known from a measurement of the scaled momentum distributions ($\ln(1/x_p)$) that HERWIG has problems in modelling the momentum distribution of the data in the current region of the Breit frame [75]. The other systematic shifts are from the backsplash cut (3) and the η and p_T cuts (7,8). These cuts change the multiplicity of the event, and since this is not exactly simulated by the Monte Carlo (see Figure 5.6), it would be expected to have some effect on the final result. The three systematics each give up to a 1% effect in most of the variables, increasing to a 1-2% effect in the jet-mass. This is understandable, since the jet-mass is sensitive to the changes in the multiplicity of the events. The systematic from the CAL energy scale is generally small. This is partly due to the definition of the event shapes; since they are defined with a scale in the denominator, the energy dependence tends to cancel. Also, the use of the ZUFOs algorithm, which takes tracking information for around a third of the hadrons, limits the dependence on the CAL energy scale.

The cross-check using DJANGOH with LEPTO (cross-check 1) instead of ARIADNE, in general produces effects of a comparable size to the HERWIG systematic, up to 10% depending on the variable and the kinematic bin. Correcting with the electron energy and angle directly (cross-check 11), instead of the DA method, gives an approximately 1-3% effect in the kinematic bins below $Q^2 \sim 320$ GeV. In the higher Q^2 bins, this rises to 5% for most of the variables.

7.4 Radiative and Electroweak Corrections

7.4.1 Radiative Corrections

The HERWIG Monte Carlo does not include ISR and FSR QED radiative corrections, and hence an investigation was made to determine the size of this effect. This was done by correcting the measured data with non-radiative ARIADNE (i.e. ARIADNE in which the HERACLES program has been turned off), and comparing this to the nominal analysis with the radiative corrections included. Figure 7.25 shows the typical size of this effect for the mean event shapes. The effect is generally negligible for all the variables except those with respect to the photon axis, and is about 10% for τ_γ and B_γ at the highest Q^2. In the case of ISR, the true Q^2 is always lower than the naïvely measured Q^2. In the nominal analysis with ARIADNE this is corrected for; however, in HERWIG (and ARIADNE with the corrections switched off) no correction to Q^2 is made, so the whole event shape curve will be pushed to higher values of Q^2. However, the deficiencies in HERWIG are mainly at the low Q^2, and it is clear from Figure 7.25 that this deficit is not caused by the absence of QED radiative effects.

7.4.2 Electroweak Corrections

Only the photon propagator term is implemented in the two NLO calculation programs; effects due to the electroweak Z^0 propagator are absent. In order to examine the size of this effect on the event shapes, a run of standard

DJANGOH Monte Carlo was generated and compared to a run with the Z^0 and $\gamma - Z^0$-interference terms switched off. There was no significant difference ($<< 1\sigma$) for all the event shapes [72]. It can therefore be concluded that the absence of electroweak terms in the NLO programs is not significant for the analysis.

7.5 Summary

The differential event shape data and the method used to correct them were presented. ARIADNE gives a good description of the data. HERWIG is adequate but has problems describing the observed shape. The mean event shapes were presented, and a strong x-dependence noted in the variables with respect to the photon axis, due to Initial State Parton Showering. ARIADNE gives a good description of the mean data, including the observed x-dependence. The description of the data by HERWIG was generally poor at low Q^2 and reasonable at high Q^2.

Systematic studies on the mean event shapes were presented. The systematic errors dominate over the statistical errors, except in the bins with $Q^2 > 5120$ GeV^2. The dominant systematic effect was the model dependence from HERWIG at around 5% on average. The absence of radiative corrections in HERWIG do not account for this difference. The cuts which affect the multiplicity shifted the mean results by around 1% on average.

Two cross-checks were performed. One with DJANGOH using the LEPTO option, resulting in shifts of a similar order but typically smaller than HERWIG, and the other from using the electron method to obtain the event kinematics rather than the DA method, which resulted in an average 3-5% effect. In general, the quoted systematic errors encompass the variations due to these cross-checks.

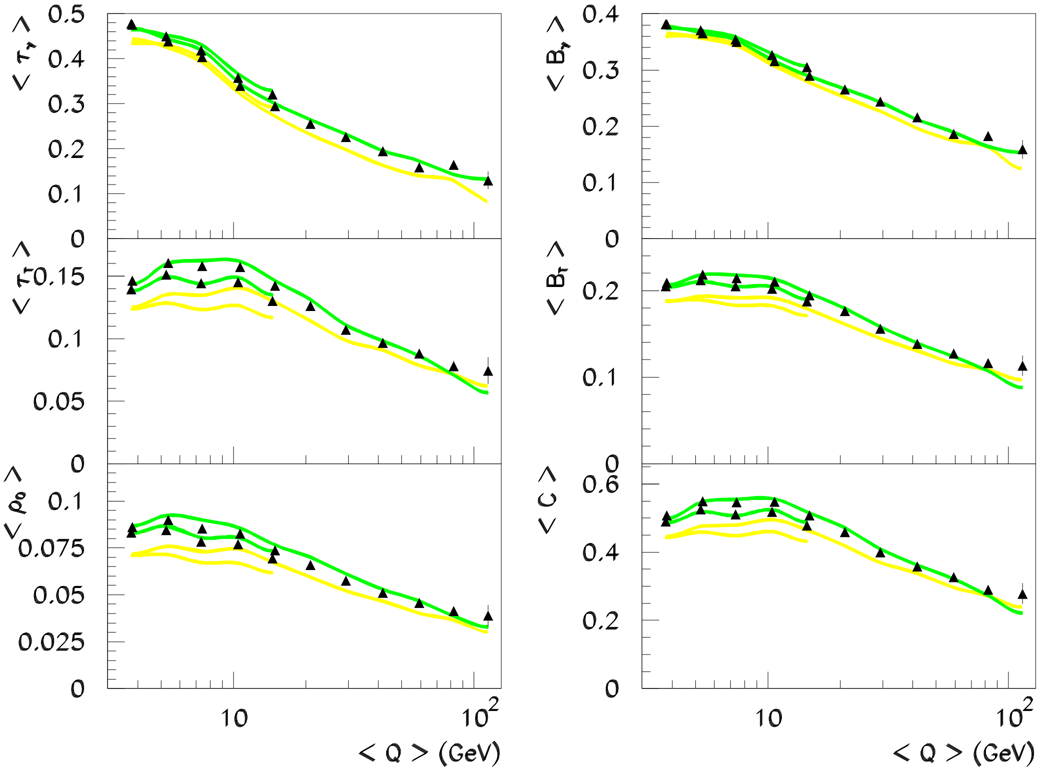


Figure 7.16: Uncorrected mean event shape distributions, $\langle F \rangle$ vs. $\langle Q \rangle$. The filled triangles are data, the dark (green) shaded band is reconstructed ARIADNE Monte Carlo, and the light (yellow) shaded band is reconstructed HERWIG Monte Carlo.

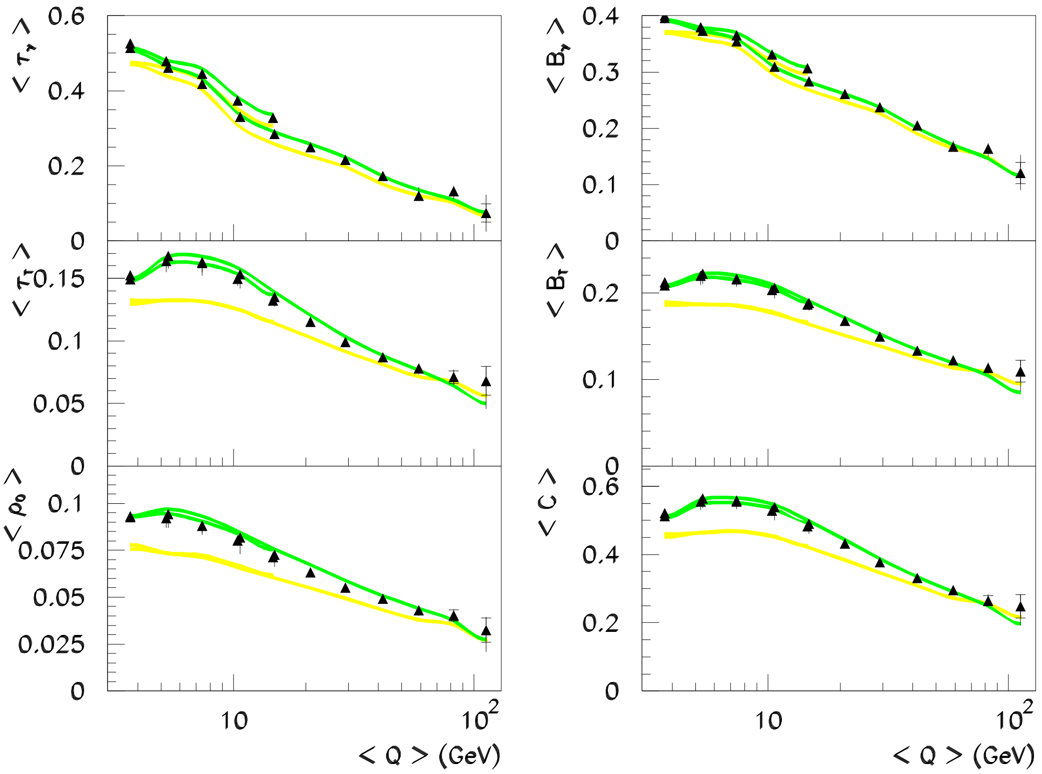


Figure 7.17: Corrected mean event shape distributions, $\langle F \rangle$ vs. $\langle Q \rangle$. The filled triangles are data, the dark (green) shaded band is hadron level ARIADNE Monte Carlo, and the light (yellow) shaded band is hadron level HERWIG Monte Carlo. The outer error bars on the data points are statistical$\oplus$systematic errors, the inner error bars are statistical only.

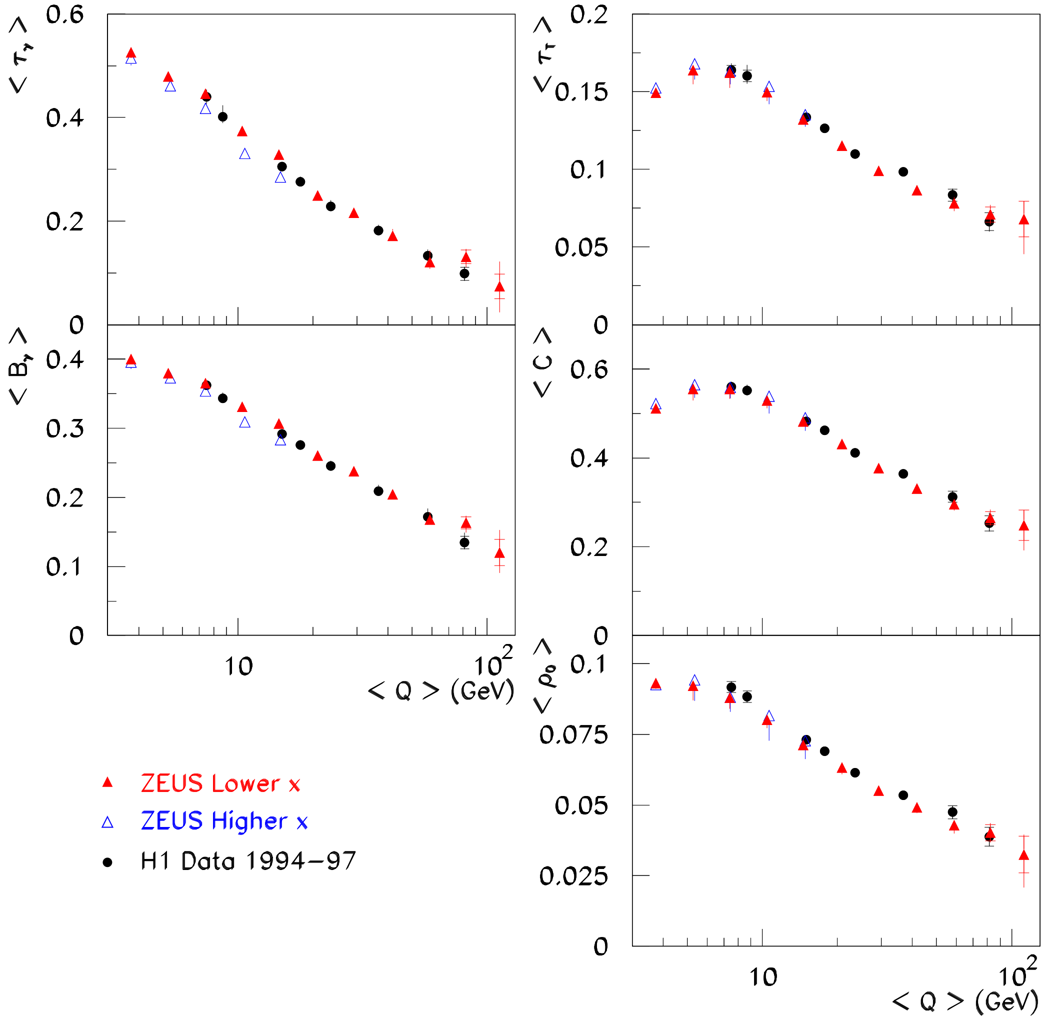


Figure 7.18: The red and blue triangles are ZEUS mean event shape data for low-x and high-x data points respectively. The black circles are H1 mean event shape data.

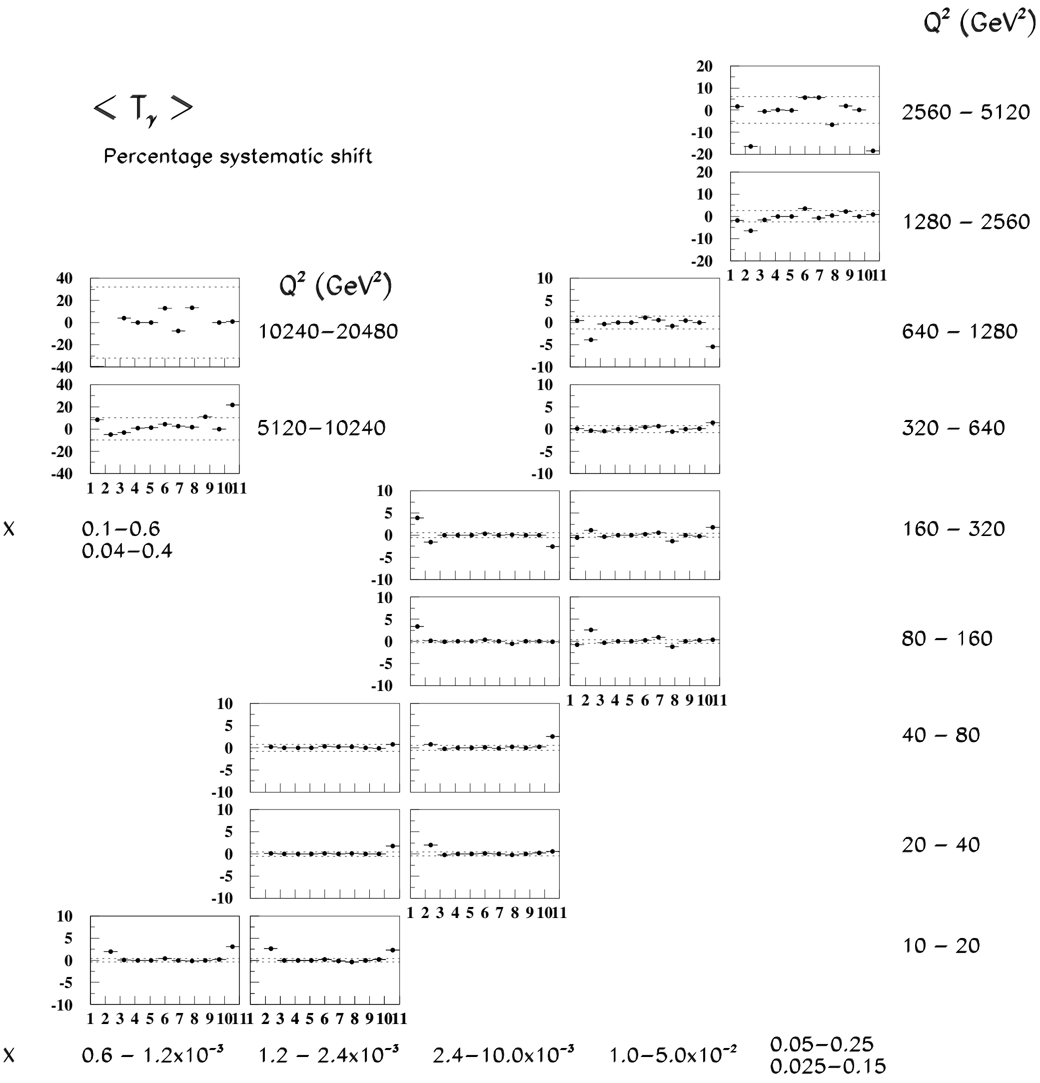


Figure 7.19: Percentage systematic deviations for $\langle T_\gamma \rangle$. Refer to Table 7.1 for the numbering of the systematics. The dotted line is the statistical error on the central data point. Cross-check number 1 (for LEPTO) is missing in the low Q^2 data, since no LEPTO Monte Carlo was available at low Q^2.

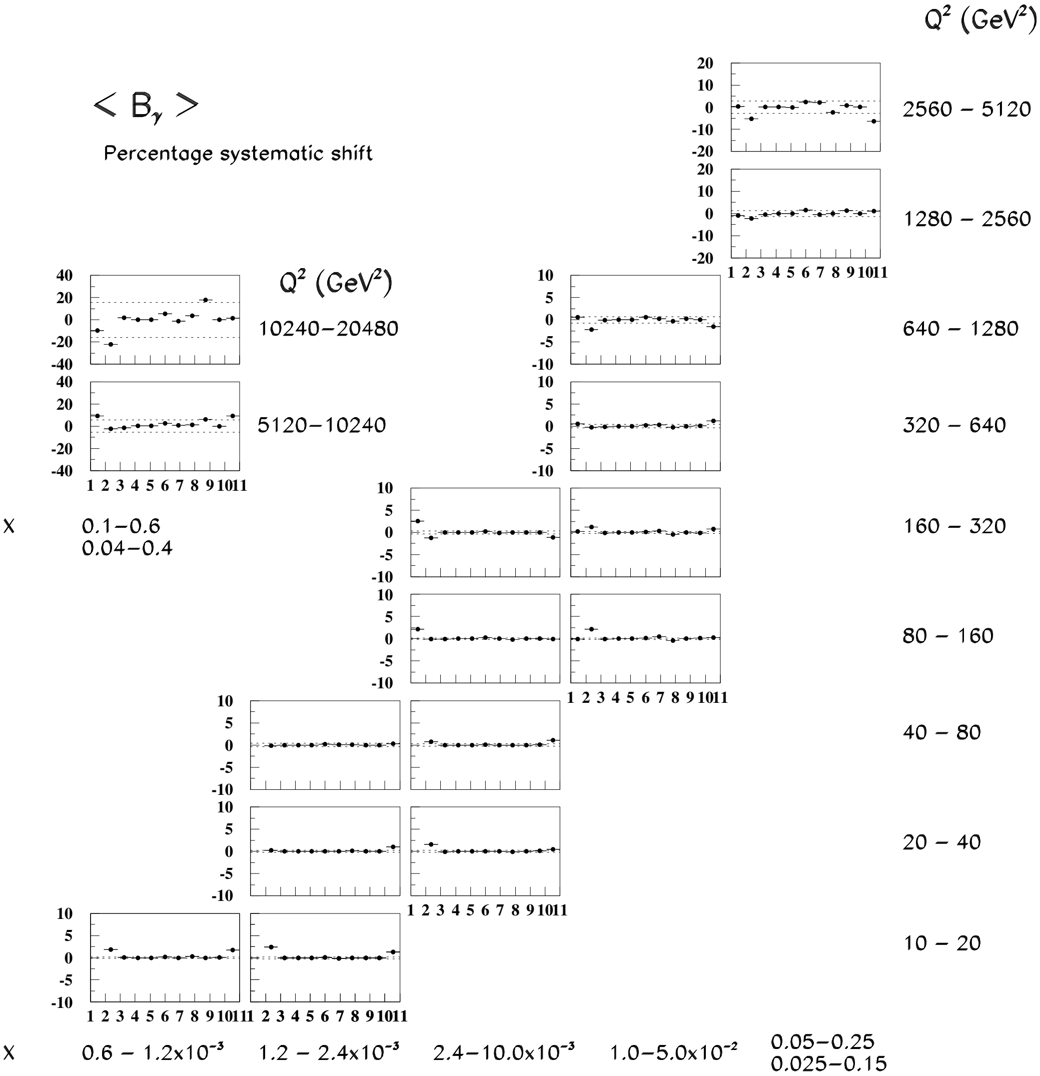


Figure 7.20: Percentage systematic deviations for $\langle B_\gamma \rangle$. Refer to Table 7.1 for the numbering of the systematics. The dotted line is the statistical error on the central data point. Cross-check number 1 (for LEPTO) is missing in the low Q^2 data, since no LEPTO Monte Carlo was available at low Q^2.

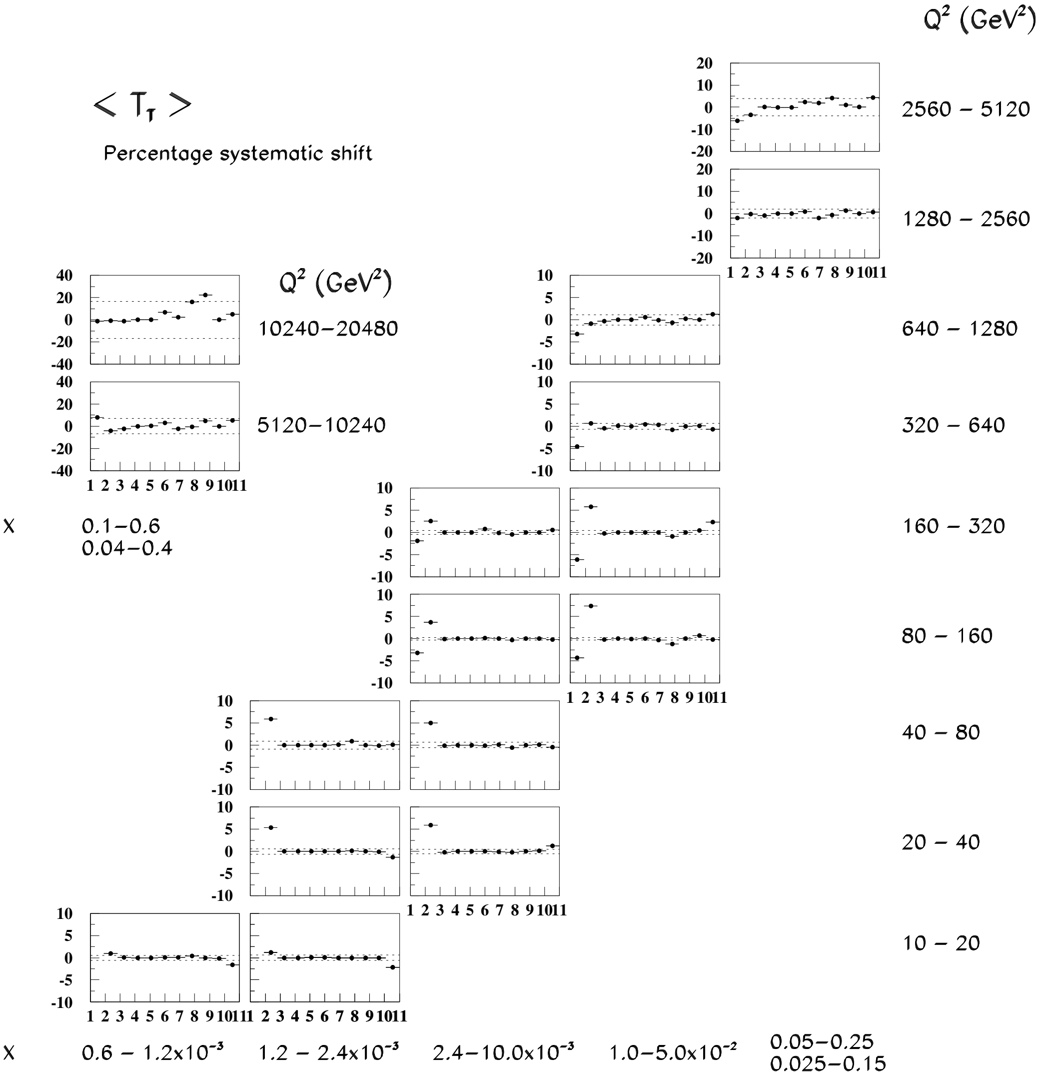


Figure 7.21: Percentage systematic deviations for $\langle T_T \rangle$. Refer to Table 7.1 for the numbering of the systematics. The dotted line is the statistical error on the central data point. Cross-check number 1 (for LEPTO) is missing in the low Q^2 data, since no LEPTO Monte Carlo was available at low Q^2.

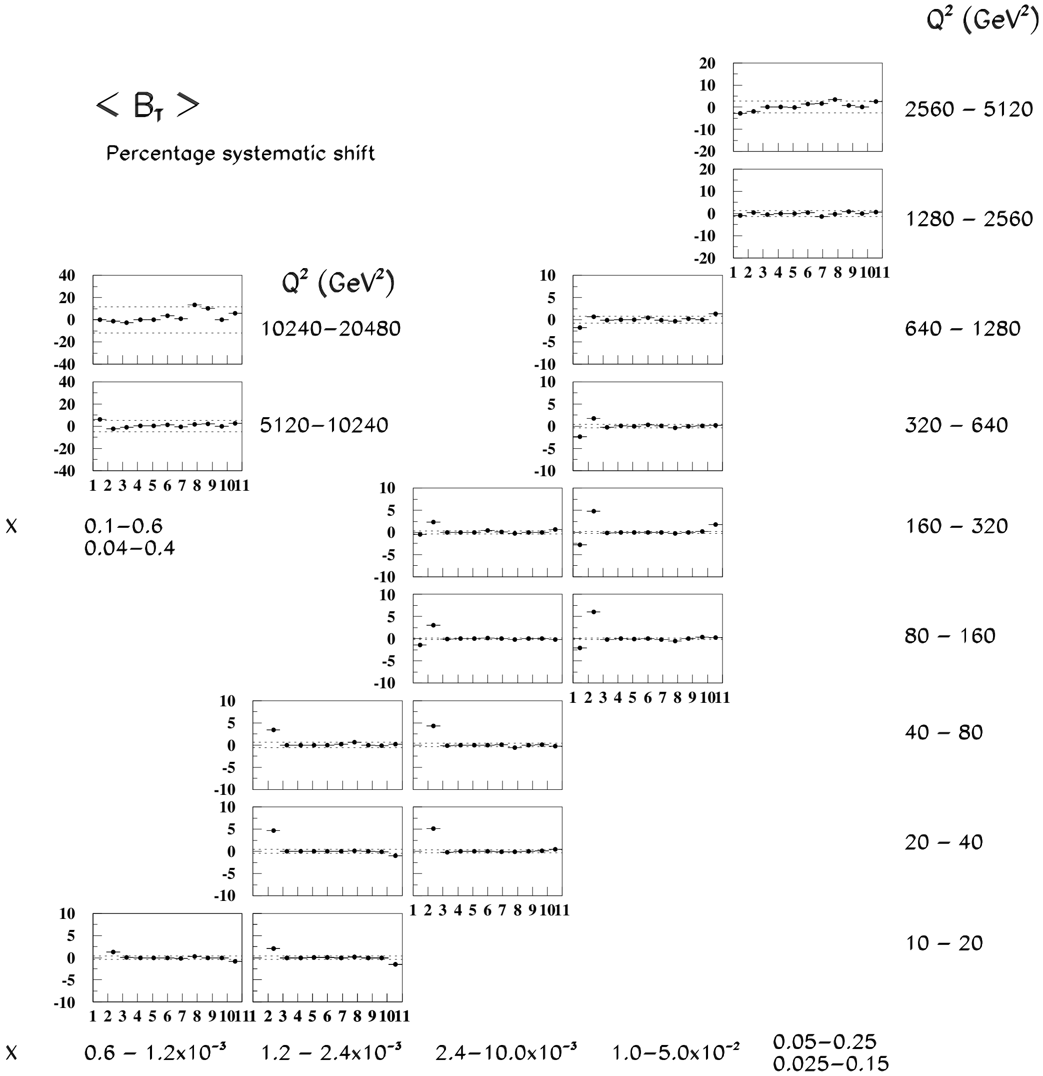


Figure 7.22: Percentage systematic deviations for $\langle B_T \rangle$. Refer to Table 7.1 for the numbering of the systematics. The dotted line is the statistical error on the central data point. Cross-check number 1 (for LEPTO) is missing in the low Q^2 data, since no LEPTO Monte Carlo was available at low Q^2.

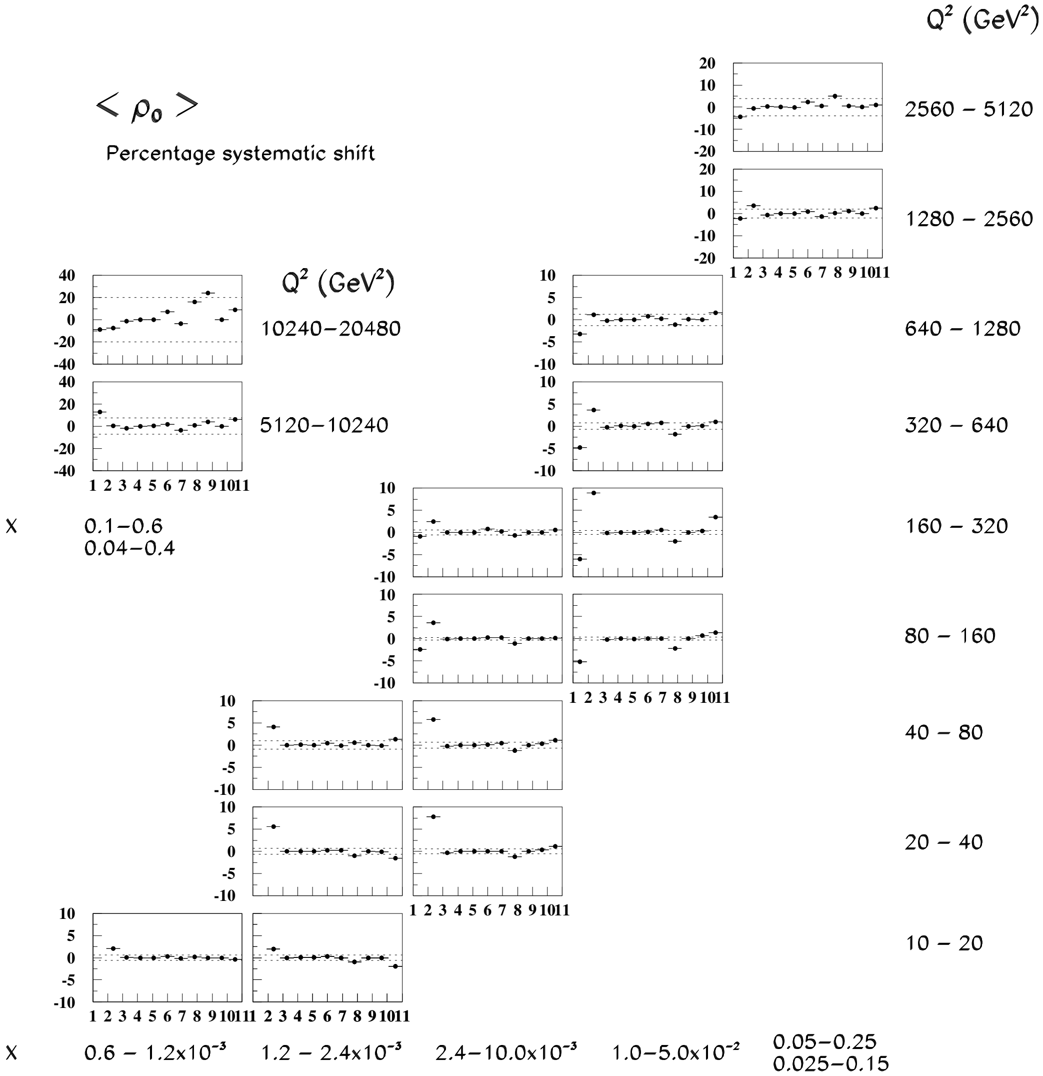


Figure 7.23: Percentage systematic deviations for $\langle \rho_0 \rangle$. Refer to Table 7.1 for the numbering of the systematics. The dotted line is the statistical error on the central data point. Cross-check number 1 (for LEPTO) is missing in the low Q^2 data, since no LEPTO Monte Carlo was available at low Q^2.

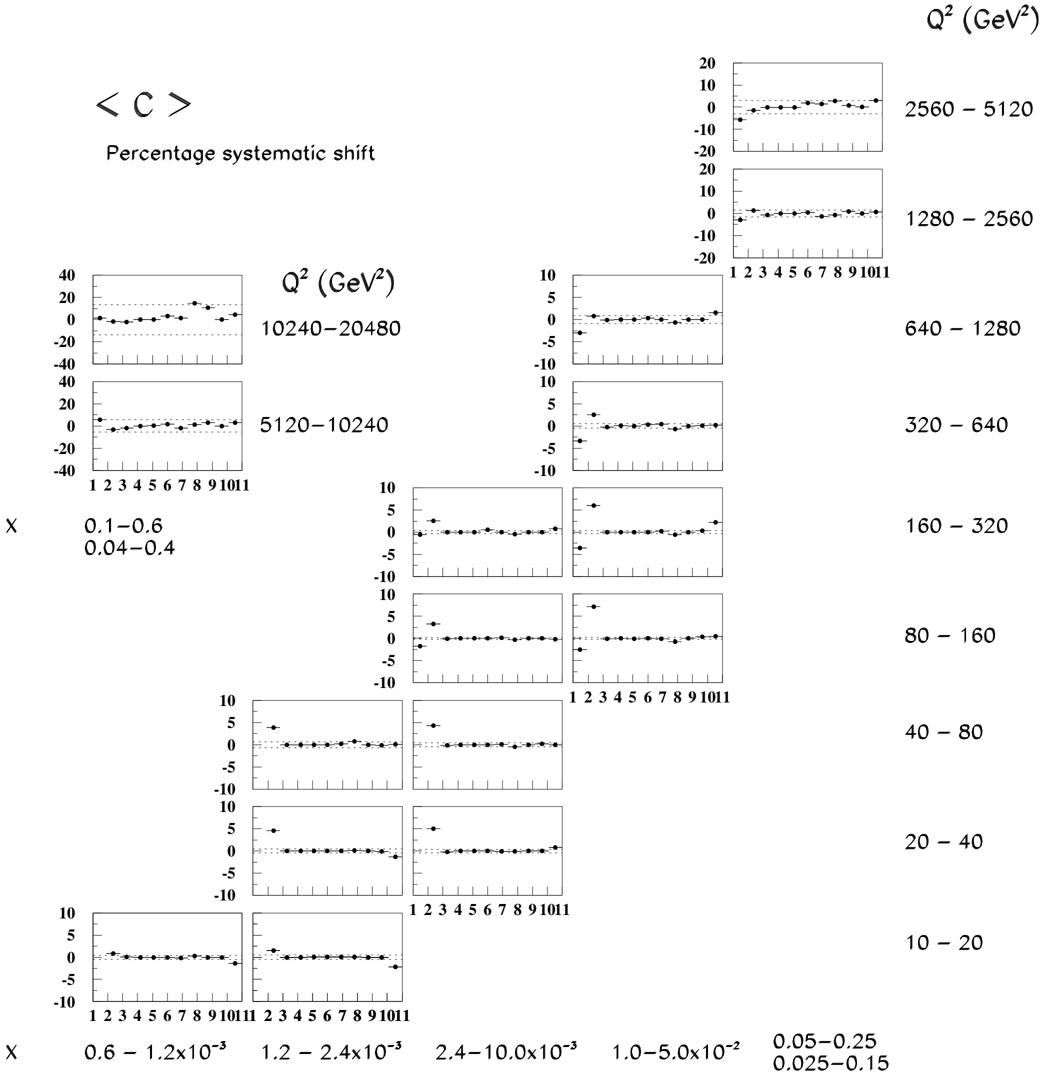


Figure 7.24: Percentage systematic deviations for $\langle C \rangle$. Refer to Table 7.1 for the numbering of the systematics. The dotted line is the statistical error on the central data point. Cross-check number 1 (for LEPTO) is missing in the low Q^2 data, since no LEPTO Monte Carlo was available at low Q^2.

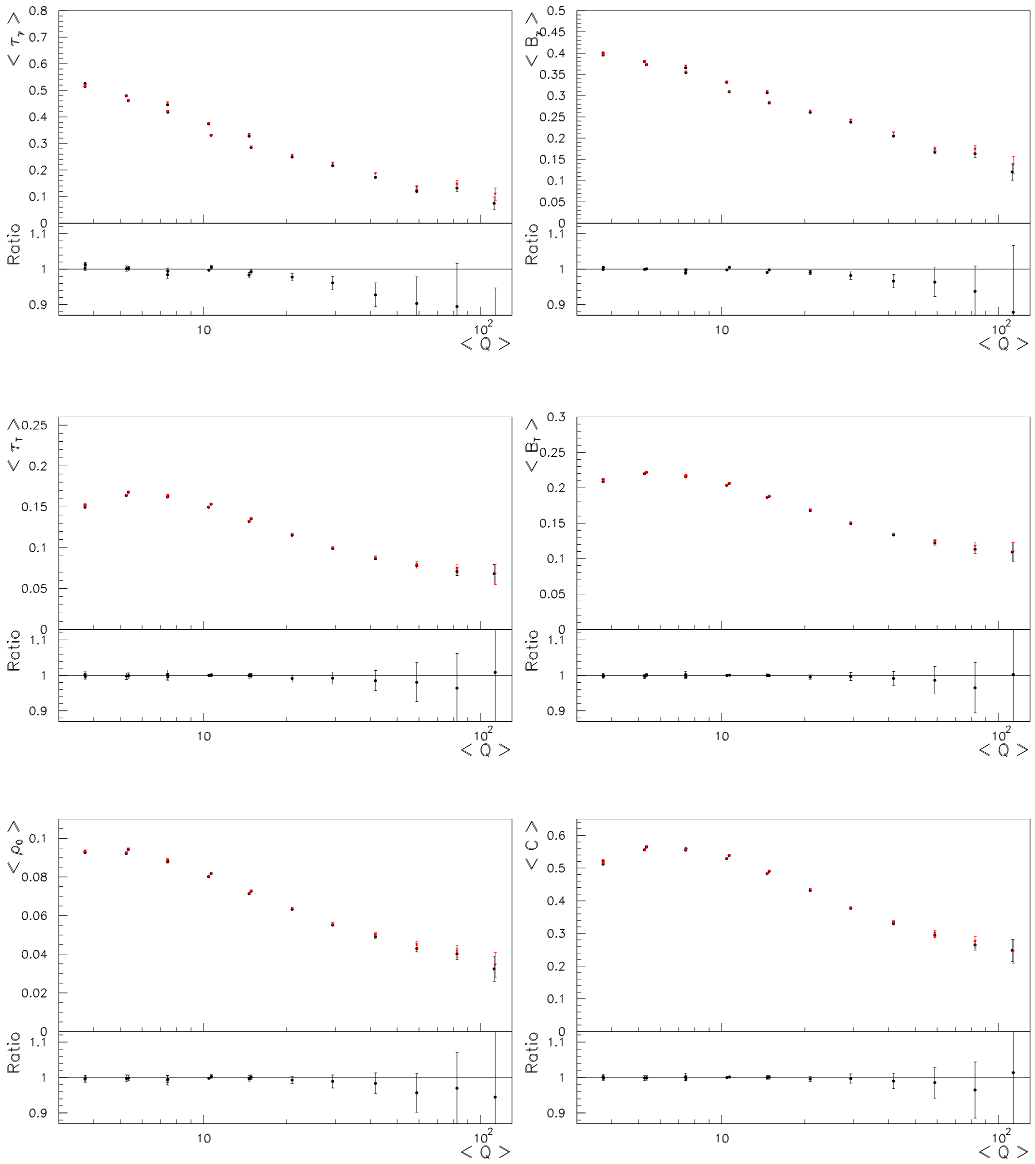


Figure 7.25: Comparison between mean event shapes corrected with the radiative ARIADNE (black circles) and corrected with the non-radiative ARIADNE (red triangles). The lower plots show the ratio of corrected to uncorrected data.

Chapter 8

QCD Fits

In this chapter, data are fitted to the QCD theory, incorporating the parton level NLO perturbative prediction and the power correction model of Dokshitzer and Webber et al., as described in chapter 1, and the values for $\alpha_s(M_Z)$ and $\overline{\alpha_0}(\mu_I)$ are extracted. The statistical errors associated with the NLO prediction, and the experimental systematic and theoretical uncertainties are propagated through to the fit parameters $\alpha_s(M_Z)$ and $\overline{\alpha_0}(\mu_I)$.

8.1 Fit Procedure

The theoretical prediction to which the data are fitted, as given in Equation 1.38, is composed of two parts which are added together. Firstly the NLO parton level prediction from programs such as DISENT, and secondly, the power correction term to account for the hadronisation.

8.1.1 Power Correction

Figure 8.1 gives an indication of the size of the typical power corrections to event shapes. The corrected differential data for the C-parameter from

Chapter 7 are plotted, compared to the $\mathcal{O}(\alpha_s^2)$ parton level prediction from DISENT. The difference between the two plots is largest at low Q^2 where the shape of the NLO distribution is quite different from that of the data. The NLO prediction has a preference for much lower values of C than the data, consequently the mean value of C is much lower. This difference must be accounted for by the power correction. The correction to the mean value reduces as Q^2 increases, becoming small in the highest Q^2 bins. Since the differential power corrections are generally unavailable, the QCD fits have been performed on the mean values of the event shapes.

8.1.2 NLO error calculation

The two NLO programs employed in this analysis, DISENT and DISASTER++, are both based on the subtraction method, as discussed in Chapter 1. The NLO programs are not event generators. Instead they are large integration programs, and the 'events' produced are points on an integration grid. Each 'event' is composed of a number of contributions (typically 10-13), real parts of the cross section and virtual counter-terms. The programs give numerically unstable results if insufficient events are generated, therefore the distributions used in this analysis were all generated for 10 million integration 'events' to ensure that the results were numerically stable.

Additional care was taken with the NLO programs to obtain correctly the error on the mean values. The error on the mean based on summed events is conventionally calculated as

$$\sigma_{\langle F\rangle} = \sqrt{\langle F\rangle^2 - \langle F^2\rangle} \tag{8.1}$$

As mentioned, each 'event' is composed of a number of different contribution terms, real and virtual parts. Since both types of terms can be divergent, they

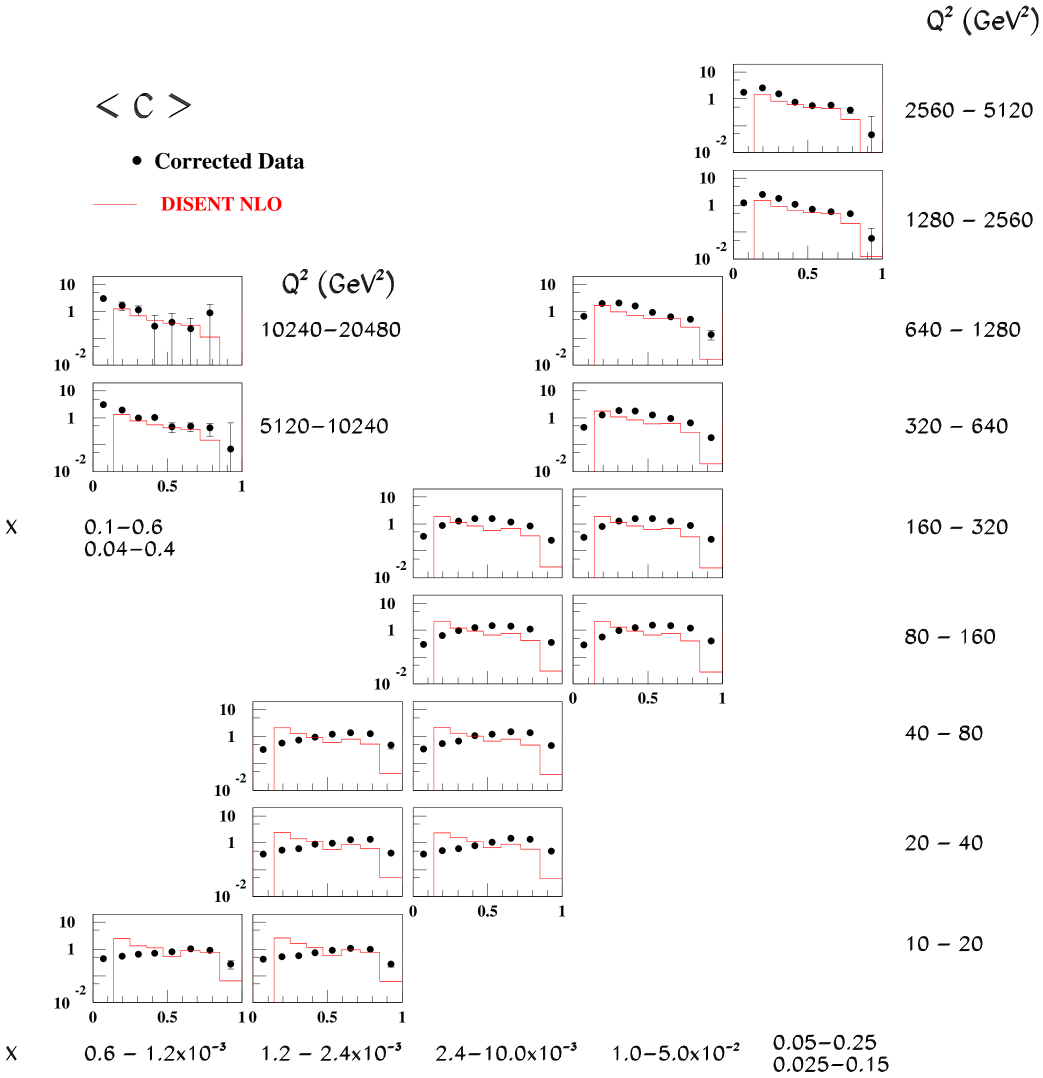


Figure 8.1: Comparison of corrected data to NLO parton level prediction. The solid points are corrected differential data for the C-parameter, in the 16 kinematic bins of (x, Q^2). The histogram is the $\mathcal{O}(\alpha_s^2)$ parton level prediction from DISENT.

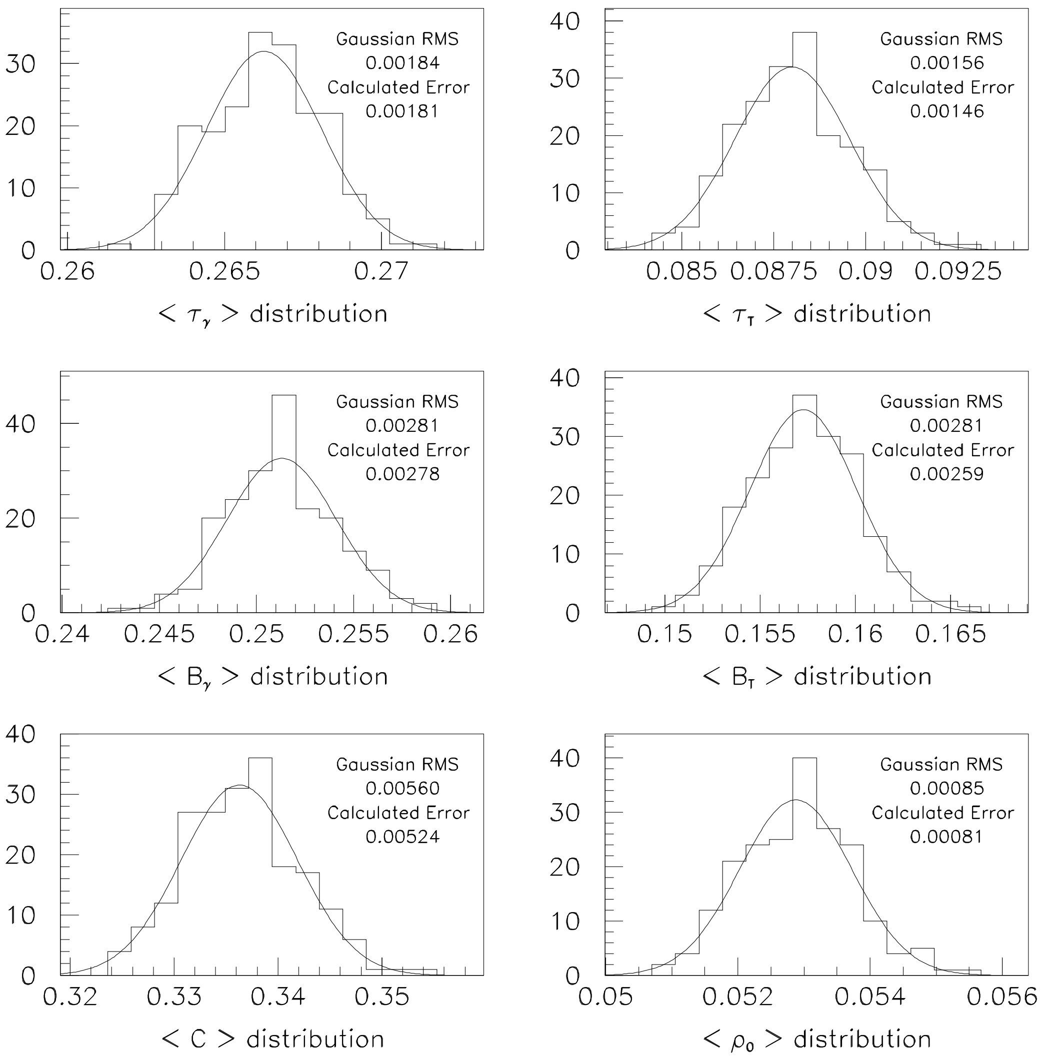


Figure 8.2: Results of generating the event shapes 200 times using different random number seeds. The results are fitted to a Gaussian. The RMS of the Gaussian agrees well with the average returned by the error calculation in Equation 8.1.

can be very large (with only their difference guaranteed to be convergent). If the sum of square of the contributions, $\langle F^2 \rangle$, is taken, these 'unreal' divergent terms will unreasonably dominate the error.

Instead, the contributions which comprise a single event must be added together, and any cancellations performed, before the sum of the squares is histogrammed. To check that this approach produces a reasonable error calculation, 200 test runs of the DISENT NLO program were generated using the same parameters, but with different random number seeds and the results for the generated mean values of the event shapes were histogrammed. Figure 8.2 shows the resulting distributions fitted to a Gaussian. The RMS of the Gaussian distributions agree to within 10% with the value of the errors determined from Equation 8.1. In addition, Figure 8.3 shows that the error returned exhibits the expected $1/\sqrt{N}$ behaviour as the number of generated events increases.

8.1.3 Parameterisation of the NLO in α_s.

The parton densities used in the NLO prediction are evaluated at a fixed α_s. This presents a problem for the extraction of α_s, since the NLO prediction cannot be evaluated consistently for arbitrary values of α_s.

Therefore, five separate NLO predictions are generated using the CTEQ-4A(1-5) series of parton density functions [61]. Each of these have been evaluated at a fixed value of α_s. For each variable and for each of the kinematic bins, the five generated points were parameterised as a linear function of α_s

$$\langle F \rangle_{NLO} = c_1 + c_2 \cdot \alpha_s(M_Z) \tag{8.2}$$

Figure 8.4 shows an example of this parameterisation in kinematic bins 7 to 16 for the jet-mass. For the lower Q^2 kinematic bins, there is a strong variation

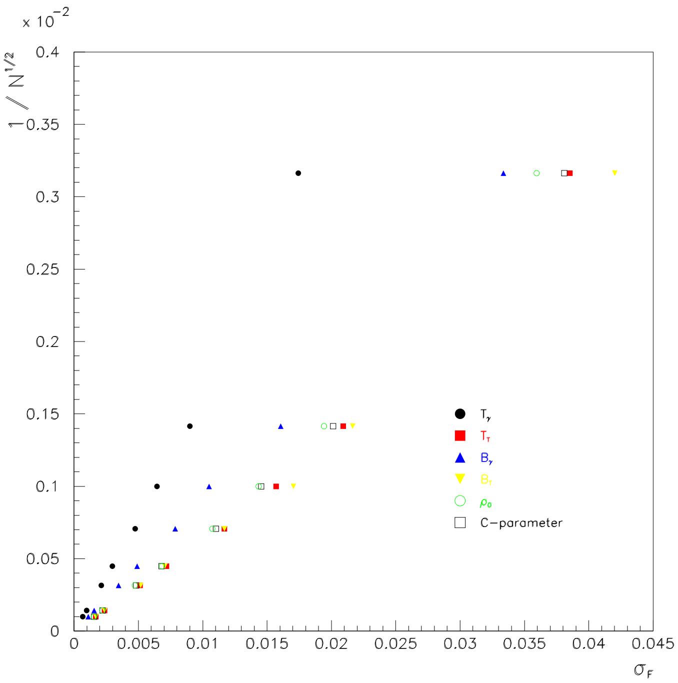


Figure 8.3: Value of the calculated error versus $1/\sqrt{N}$, where N is the number of events generated.

of the NLO prediction with α_s (up to 30%). The low values obtained for the χ^2 show that a linear fit is appropriate, and that a higher order polynomial in α_s does not need to be considered. The x-dependence of the NLO prediction varies as a function of α_s. For example, in bins 7 and 8, at around $\alpha_s \simeq 0.123$ there is no x-dependence, whereas, at $\alpha_s \simeq 0.110$ an x-dependence of a few percent has developed. In all the fits that follow, a linear parameterisation of the NLO using Equation 8.2 provided a good description of the generated points and was used to obtain the perturbative NLO component of the fit.

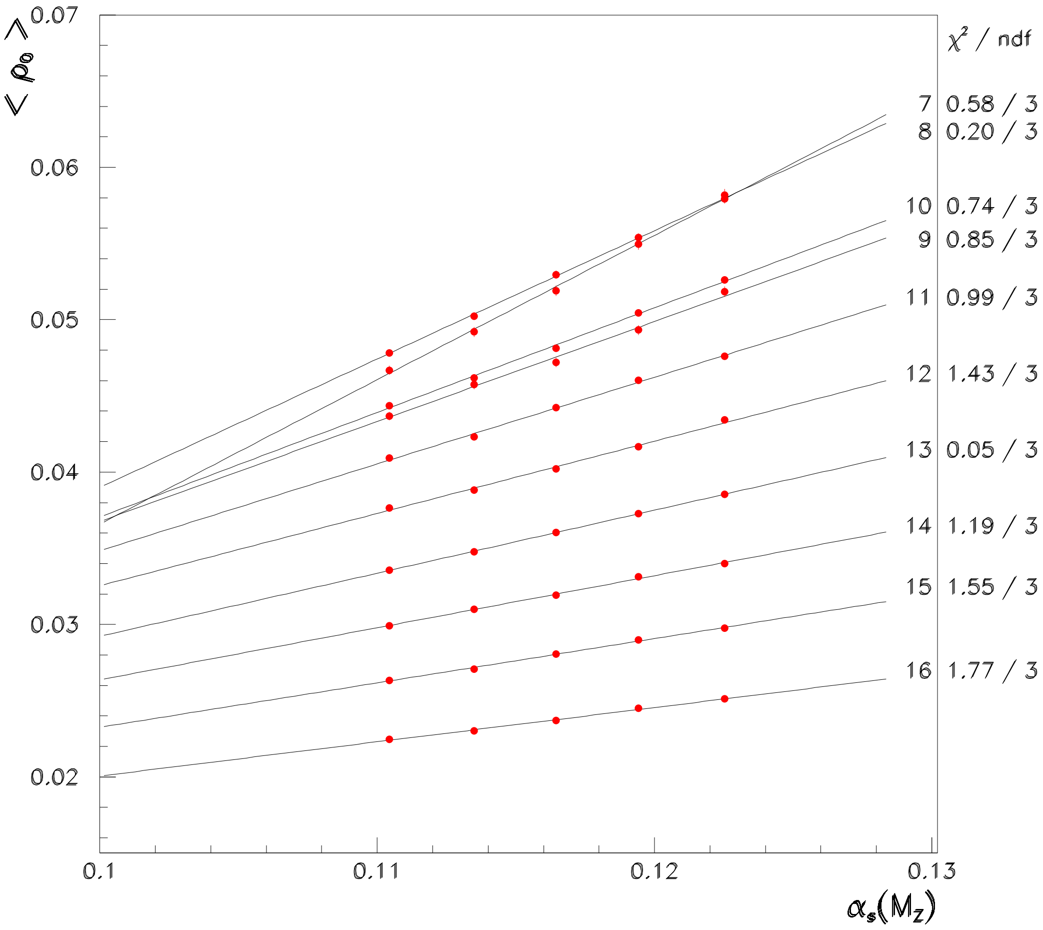


Figure 8.4: The five runs of DISENT at different values of fixed α_s for kinematic bins 7 to 16. Each kinematic bin is parameterised as a linear function of α_s given in Equation 8.2. The χ^2 values for the parameterisation are given.

8.2 Fit Results

The fits are performed using the MINUIT [76] package of the CERN library [77], with the NLO prediction given by Equation 8.2. The fit program varies both parameters, $\alpha_s(M_Z)$ and $\overline{\alpha_0}(\mu_I)$, simultaneously, until the best χ^2 is found, as defined by

$$\chi^2(\alpha_s, \overline{\alpha_0}) = \sum_{i=7}^{16} \frac{\left[\langle F \rangle_{NLO,i} + \langle F \rangle_{pow,i} - \langle F \rangle_{data,i}\right]^2}{\sigma_i^2} \tag{8.3}$$

where σ_i is the statistical error of the data points combined in quadrature with the error of the parameterisation of the NLO; the power correction under consideration is assumed to have no error. The errors and correlations returned by the NLO calculation are fully propagated through the parameterisation into the calculation for σ_i.

The minimisation is performed in two dimensions using a variable metric method. This depends on the first-derivatives of the fit function, which are determined numerically. In the rare case that this method fails, for example, in a region of phase space where the first-derivatives are poorly determined from numerical techniques, then the fit program switches to the Simplex method for continued minimisation. In the case of local or unphysical minima, two improvement algorithms which search for extra minima are available. Should this fail, the allowed values of α_s and $\overline{\alpha_0}$ can be fixed to within physical limits, e.g. $\alpha_s > 0$. The parameter errors and correlations are calculated fully by the program, mapping out the minimum and tracing a contour $\chi^2 = \chi^2_{min} + 1$ around the minimum point. The quoted errors on the two parameters are then projections of this contour onto the parameter axes. The contours themselves can also be plotted on the $(\alpha_s, \overline{\alpha_0})$ plane.

Unless otherwise stated, only the corrected data from bin 7 and above are used in the fits (corresponding to $Q > 9$ GeV.) This is because the

Variable	$\alpha_s(M_Z)$	$\overline{\alpha_0}(\mu_I)$	χ^2/ndf
τ_γ	0.1303 ± 0.0018	0.4171 ± 0.0105	29.5 / 8
B_γ	0.1006 ± 0.0012	0.5790 ± 0.0034	217.8 / 8
τ_T	0.1289 ± 0.0012	0.5003 ± 0.0017	30.7 / 8
B_T	0.1174 ± 0.0011	0.4569 ± 0.0026	29.9 / 8
ρ_0	0.1287 ± 0.0014	0.4681 ± 0.0018	17.7 / 8
C	0.1296 ± 0.0008	0.4438 ± 0.0016	18.5 / 8

Table 8.1: Fitted results for $\alpha_s(M_Z)$ and $\overline{\alpha_0}(\mu_I)$ for the fit using the NLO prediction from DISENT and the $\mathcal{E}_{lim} > 0.10Q$ energy cut. Only bins with $Q > 9$ GeV were fitted.

power correction theory is not expected to be applicable at low energies. The initial analysis uses the DISENT prediction with a $\mathcal{E}_{lim} > 0.10Q$ energy cut in the current region of the Breit frame. This cut was motivated so that a comparison with H1 event shape data data could be made (see Figure 7.18). The effect of using a $\mathcal{E}_{lim} > 0.25Q$ cut and changing to DISASTER++ is examined later.

8.2.1 Initial Fit Results

The $\mathcal{E}_{lim} > 0.10Q$ fit to data with $Q > 9$ GeV, using DISENT, is shown in Figure 8.5 with the fitted values of α_s and $\overline{\alpha_0}$ given in Table 8.1. For the thrust and broadening with respect to the photon axis, Figure 8.5 shows that the data have a significant x-dependence which is poorly modelled by the fit. In B_γ, the x-dependence is not followed at all, leading to a very poor χ^2/ndf for this fit of 217.8/8. In T_γ, the x-dependence is modelled, but at the expense of having almost no power correction term; the fitted value

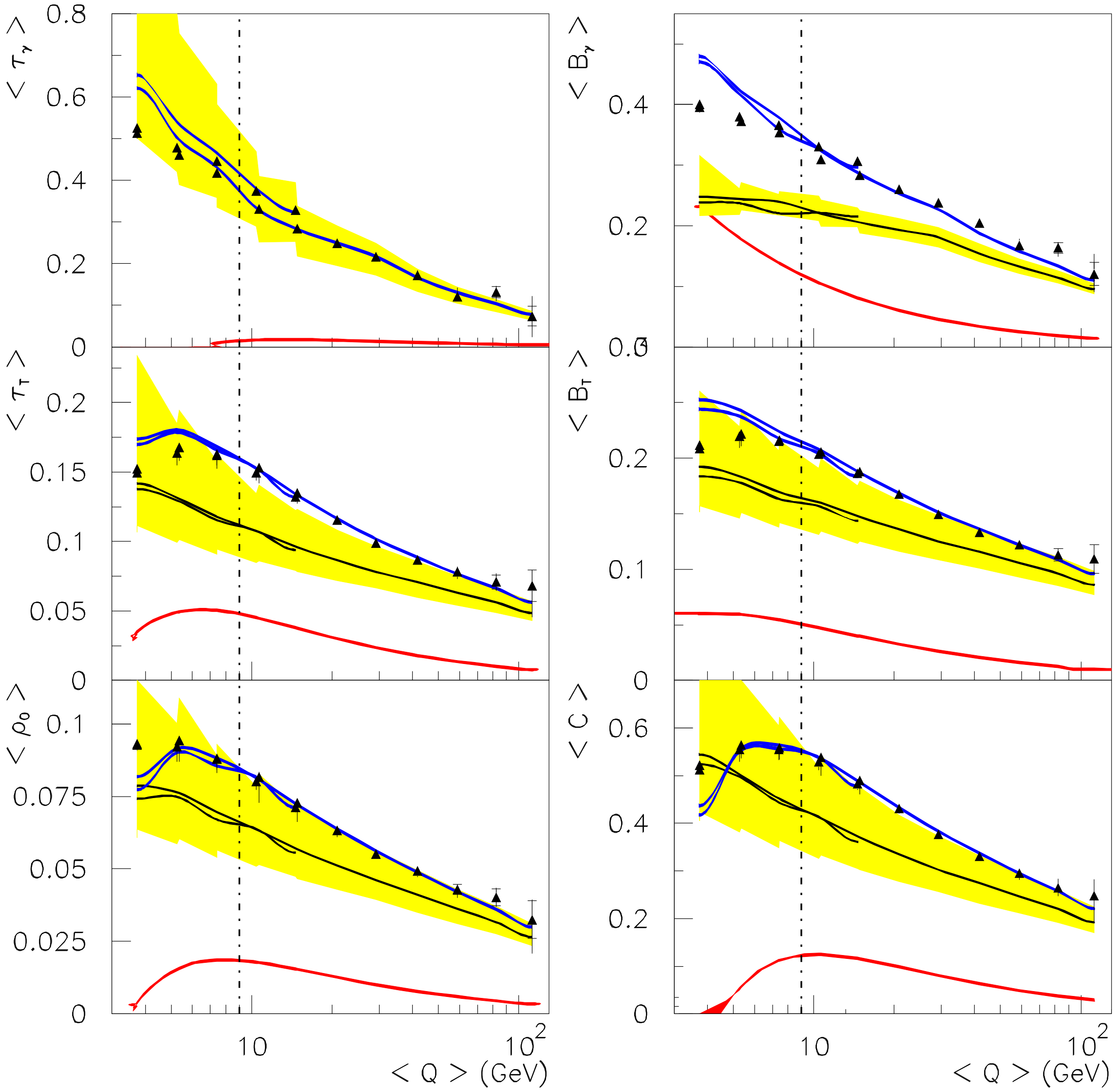


Figure 8.5: Fitted event shape data for $\mathcal{E}_{lim} > 0.10Q$ energy cut. The data are shown as triangles with statistical and systematic errors. The lower (red) line is the power correction. The line in the middle of the light (yellow) shaded band is the NLO prediction from DISENT; the shaded band indicates the renormalisation scale uncertainty. The upper (blue) line is the sum of the NLO and the power correction, fitted to the data. In T_γ the power correction is very small, so the NLO curve has been omitted for clarity. The fit is made for data with $Q > 9$ GeV, and the curves are extrapolated back into the unfitted region.

of $\overline{\alpha_0}$ is around 15% lower than that for the other variables. The thrust is theoretically expected to have the same hadronisation correction regardless of axis; this is clearly not the case, with the fitted values of $\overline{\alpha_0}$ differing by 16% between T_γ and T_T. Excluding B_γ, the variables fit similar values of α_s (to within a few standard deviations), 5-10% higher than the world average, with an approximately 15% spread in the fitted values of $\overline{\alpha_0}$, around 0.45.

Figure 8.5 also shows the typical renormalisation scale uncertainty, represent by the shaded band on the plot. This uncertainty is obtained by varying the renormalisation scale in the NLO generator program up ($\mu_R = 2Q$) and down ($\mu_R = Q/2$), holding $\alpha_s(M_Z)$ fixed to the fitted value for each variable to illustrate the size of the effect. The scale uncertainty increases as $\langle Q \rangle$ decreases, this being another reason to exclude the data below $Q \simeq 9$ GeV. B_γ apparently displays a much smaller renormalisation scale dependence than the other variables. However, this is thought to be a phase space limitation of the integration in DISENT, associated with the DISENT problem for B_γ, as discussed in [25]. Figure 8.5 illustrates that the theoretical scale uncertainty is the dominant uncertainty on the fit, larger by a factor 5 or 6 than the experimental systematic errors on the data points. The effect of the analysis systematics and the theoretical scale uncertainties on the fitted values of α_s and $\overline{\alpha_0}$ will be examined in section 8.3.

8.2.2 x-dependence

In order to examine the stability of the fits with respect to the observed x-dependence, the fits have been repeated, once including only the low-x bins (bins 7 and 9), and once with only the high-x bins included (bins 8 and 10.) The results are summarised in a contour plot in the $(\alpha_s, \overline{\alpha_0})$ plane shown in Figure 8.6. The axis-independent variables, the C-parameter and jet-mass,

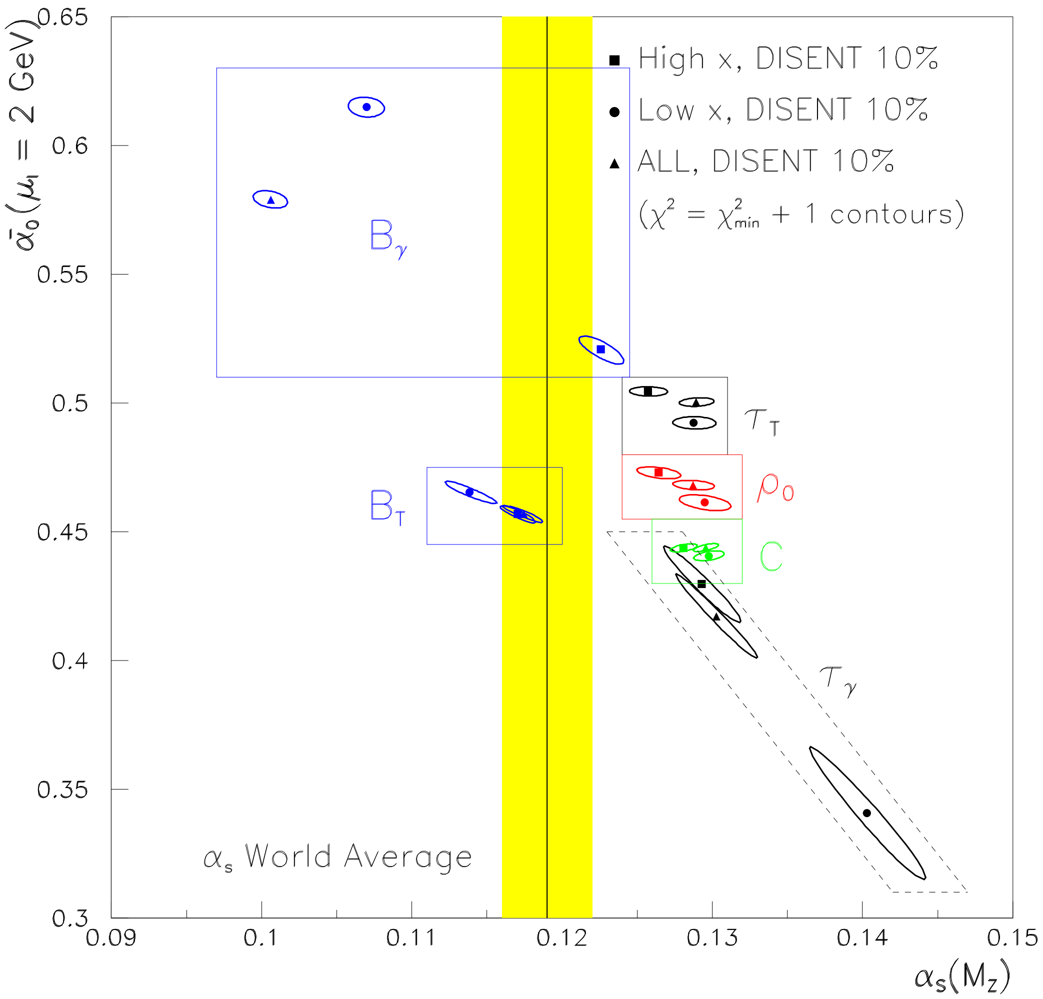


Figure 8.6: The fitted values of α_s and $\overline{\alpha_0}$ for DISENT with the $\mathcal{E}_{lim} > 0.10Q$ energy cut. The contours are the 1σ statistical contours only. Each fit has been repeated three times (for clarity, a box has been placed around each set), one for all data above 9 GeV, one fitting the low x points above 9 GeV only and the other fitting the high x points above 9 GeV only. The shaded vertical band is the world average range for $\alpha_s(M_Z)$.

ρ_0, show little dispersion with x, and are consistent within two standard deviations. The thrust and broadening with respect to the thrust axis (τ_T and B_T) show some dispersion, but are also consistent within two or three standard deviations. A significant anti-correlation between α_s and $\overline{\alpha_0}$ can also be seen, particularly in the variables with respect to the photon axis.

The thrust with respect to the photon axis shows a two standard deviation dispersion when the low-x data is fitted, otherwise it is consistent within statistics. The instability of the broadening, B_γ, with respect to x, is most clearly shown by the wide dispersion of the three fitted points, ranging from an $\alpha_s(M_Z)$ of 0.09 to 0.12, and an $\overline{\alpha_0}(\mu_I)$ of 0.52 to 0.62. Additionally, for the fit using all the data, there exists a second minimum with a similarly poor χ^2/ndf of 238.8/8, at $(\alpha_s, \overline{\alpha_0}) = (0.1381, 0.4945)$. This solution is a result of the x-dependence being fitted in a different way. Neither solution can adequately describe x-dependence. The instability with x and problems with secondary minima suggest that the data are not being modelled correctly by the combination of DISENT with the power correction theory.

8.2.3 Effect of Energy Cut

The energy cut in the current region of the Breit frame $\mathcal{E}_{lim} > 0.10Q$ can be varied. As mentioned in Chapter 1, recent work [17] has suggested that an energy cut of $\mathcal{E}_{lim} > 0.25Q$ is more appropriate for comparison with power correction theory. Figure 8.7 shows the effect of changing the energy cut from $\mathcal{E}_{lim} > 0.10Q$ to $\mathcal{E}_{lim} > 0.25Q$. The fitted values of $\overline{\alpha_0}(\mu_I)$, given in Table 8.2, are reasonably unaffected (within 2 standard deviations) by the change in the energy cut (although T_γ is lower by 6%). However, $\alpha_s(M_Z)$ appears to be more sensitive to the cut, in particular, the thrust with respect to the thrust axis, the jet-mass and the C-parameter where $\alpha_s(M_Z)$ decreases by

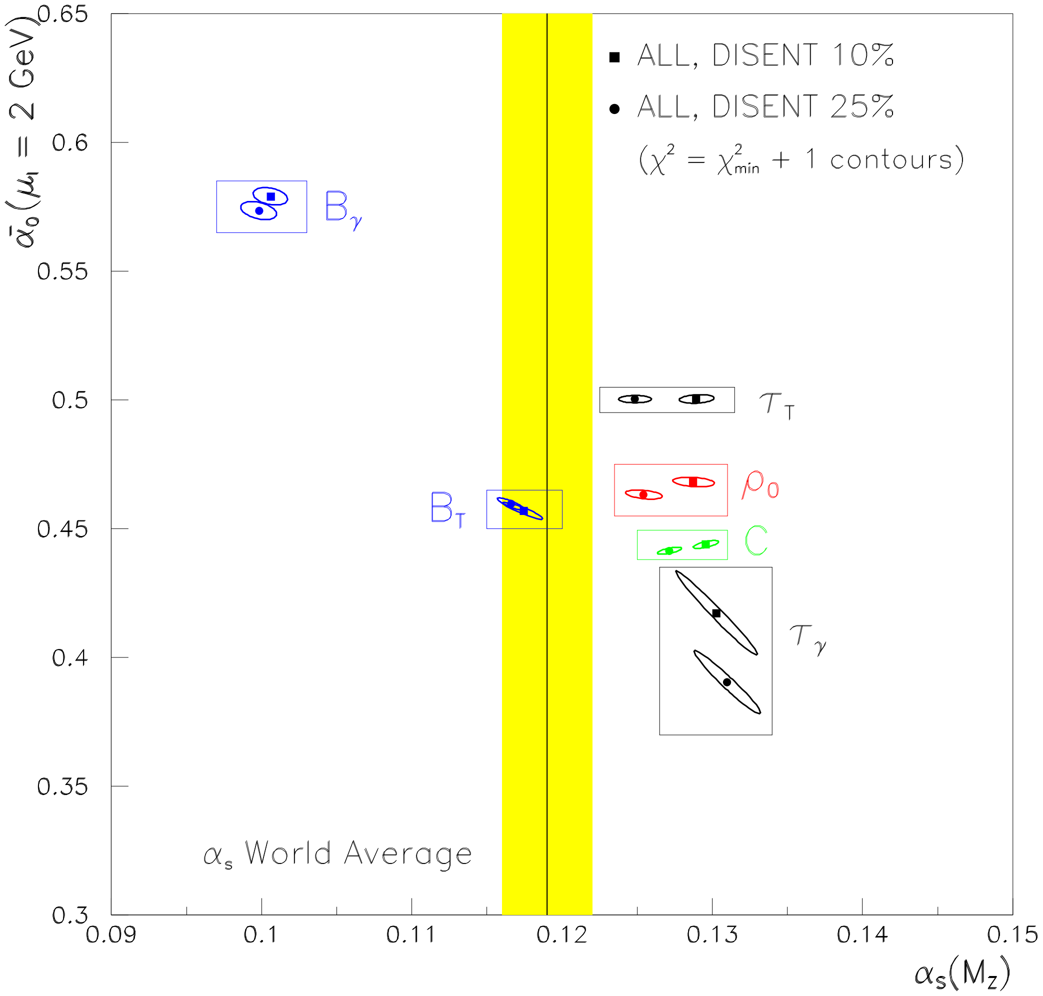


Figure 8.7: The fitted values of α_s and $\overline{\alpha_0}$ using the NLO prediction from DISENT with the $\mathcal{E}_{lim} > 0.10Q$ and $\mathcal{E}_{lim} > 0.25Q$ energy cuts. The shaded vertical band is the world average range for $\alpha_s(M_Z)$. The boxes are for clarity to group the variables.

Variable	$\alpha_s(M_Z)$	$\overline{\alpha_0}(\mu_I)$	χ^2/ndf
τ_γ	0.1310 ± 0.0017	0.3905 ± 0.0093	14.2 / 8
B_γ	0.0998 ± 0.0012	0.5734 ± 0.0035	247.3 / 8
τ_T	0.1248 ± 0.0011	0.5003 ± 0.0014	25.1 / 8
B_T	0.1166 ± 0.0009	0.4596 ± 0.0023	23.2 / 8
ρ_0	0.1254 ± 0.0012	0.4633 ± 0.0017	16.2 / 8
C	0.1271 ± 0.0008	0.4414 ± 0.0014	11.9 / 8

Table 8.2: Fitted results for $\alpha_s(M_Z)$ and $\overline{\alpha_0}(\mu_I)$ for the central fit using the NLO prediction from DISENT and the $\mathcal{E}_{lim} > 0.25Q$ energy cut. Only bins with $Q > 9$ GeV were fitted.

several standard deviations.

To examine whether the $\mathcal{E}_{lim} > 0.25Q$ cut has any effect on the problems of modeling the x-dependence, the three fits with different x ranges are repeated. The resulting fitted contours are shown in Figure 8.8. In general, the larger energy cut decreases the dispersion for all of the variables except B_γ. The poor modelling of B_γ with respect to x is still evident, suggesting that this is an intrinsic problem of the DISENT with power correction model and not just an artifact of a lower energy cut.

The fitted values for α_s and $\overline{\alpha_0}$ for the $\mathcal{E}_{lim} > 0.25Q$ cut are presented in Table 8.2. They display a similar range to the $\mathcal{E}_{lim} > 0.10Q$ energy cut, clustering around the world average for α_s with a 15% spread in $\overline{\alpha_0}$. The values of the χ^2 are slightly better (except for the jet broadening with respect to the photon axis), indicating a preferential fit for this energy cut. For this reason, combined with the reduced dispersion with respect to x, and because of the general theoretical preference for the larger energy cut, the remainder

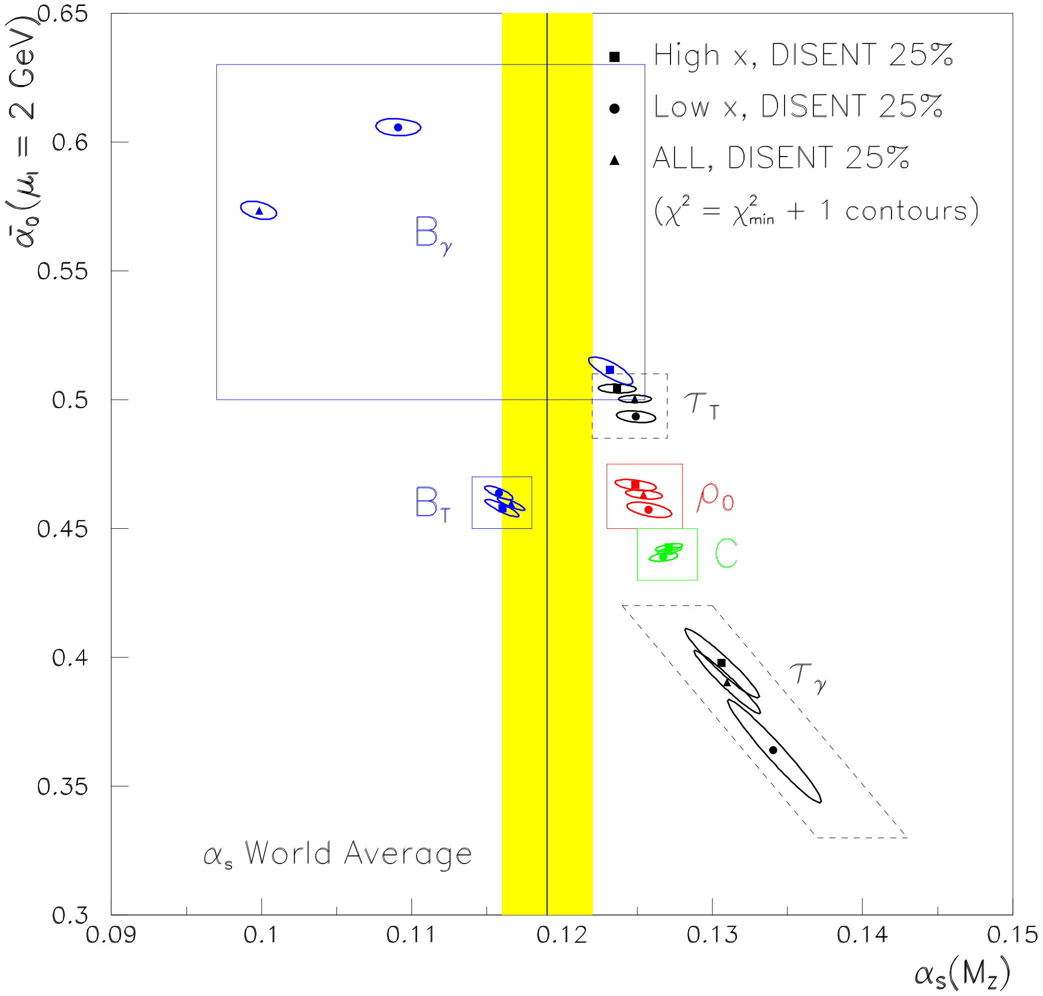


Figure 8.8: The fitted values of α_s and $\overline{\alpha_0}$ using the NLO prediction from DISENT and a $\mathcal{E}_{lim} > 0.25Q$ energy cut. The contours are the 1σ statistical contours only. Each fit has been repeated three times (for clarity, a box has been placed around each set), one for all data above 9 GeV, one fitting the low x points above 9 GeV only and the other fitting the high x points above 9 GeV only. The shaded vertical band is the world average range for $\alpha_s(M_Z)$.

Variable	$\alpha_s(M_Z)$	$\overline{\alpha_0}(\mu_I)$	χ^2/ndf
τ_γ	0.1354 ± 0.0029	0.3287 ± 0.0188	19.7 / 8
B_γ	0.1271 ± 0.0027	0.4593 ± 0.0172	14.3 / 8
τ_T	0.1259 ± 0.0014	0.4844 ± 0.0020	22.4 / 8
B_T	0.1160 ± 0.0014	0.4566 ± 0.0042	10.7 / 8
ρ_0	0.1271 ± 0.0016	0.4440 ± 0.0031	16.7 / 8
C	0.1275 ± 0.0011	0.4274 ± 0.0017	12.4 / 8

Table 8.3: Fitted results for $\alpha_s(M_Z)$ and $\overline{\alpha_0}(\mu_I)$ using the NLO prediction from DISASTER++ and a $\mathcal{E}_{lim} > 0.25Q$ energy cut.

of the fit analysis was performed using the $\mathcal{E}_{lim} > 0.25Q$ energy cut.

8.2.4 DISASTER++

As first discovered by a study comparing DISENT and DISASTER++ for event shapes [25], the DISENT NLO program has a problem in the jet-broadening with respect to the photon axis, B_γ, as discussed in Chapter 1 and [26]. To examine this effect, and also the general effect on other variables, the fits have been repeated using DISASTER++ (with a $\mathcal{E}_{lim} > 0.25Q$ energy cut).

The comparison with the $\mathcal{E}_{lim} > 0.25Q$ DISENT-based fit is shown in Figure 8.9 with the fitted values given in Table 8.3. With the exception of B_γ, the variables are fitted with the same value of $\alpha_s(M_Z)$ (within statistics) as the fit performed using the DISENT NLO prediction. For all variables, the fitted value of $\overline{\alpha_0}(\mu_I)$ is systematically smaller by approximately 4%. It is also interesting to note that the errors for DISASTER++ are larger than those of DISENT, despite both simulations being run for the same number of integration points. This is due to the distribution of weights in the integration phase space of the DISASTER++ calculation, which results in the occasional very large weight. The χ^2/ndf value of 14.3/8 for the B_γ fit is significantly better than that from the DISENT fit (more than would be expected from simply an increased statistical error on the NLO calculation) and comparable with the other variables. This indicates that the x-dependence is better described by the combination of DISASTER++ and the power correction model. To examine this explicitly, the three fits over differing x-ranges were repeated and the resultant contours plotted in Figure 8.10. The fitted results for α_s and $\overline{\alpha_0}$ for B_γ are compatible with the other variables, and there is no significant instability with x. Although the phase space integration problem makes the errors very large, it is clear that the combination of DISASTER++ and the power correction model provides a much better description of the jet-broadening than the model incorporating DISENT. The other variables display the same degree of stability with respect to x as in the DISENT-based model.

8.3 Systematic and Theoretical Uncertainties

In this section, the effect of the analysis systematics, as described in section 7.3, on the fitted values of α_s and $\overline{\alpha_0}$ is examined. Since some of the systematic effects on the fitted values are likely to be correlated, the appropriate method is to fit each of the systematics separately and then combine the result at the end. The renormalisation scale and factorisation scale of the NLO calculation are varied independently to obtain a theoretical scale uncertainty.

Although the DISASTER++ model is preferable for the broadening with respect to the photon axis, the calculation takes a factor 10 longer in com-

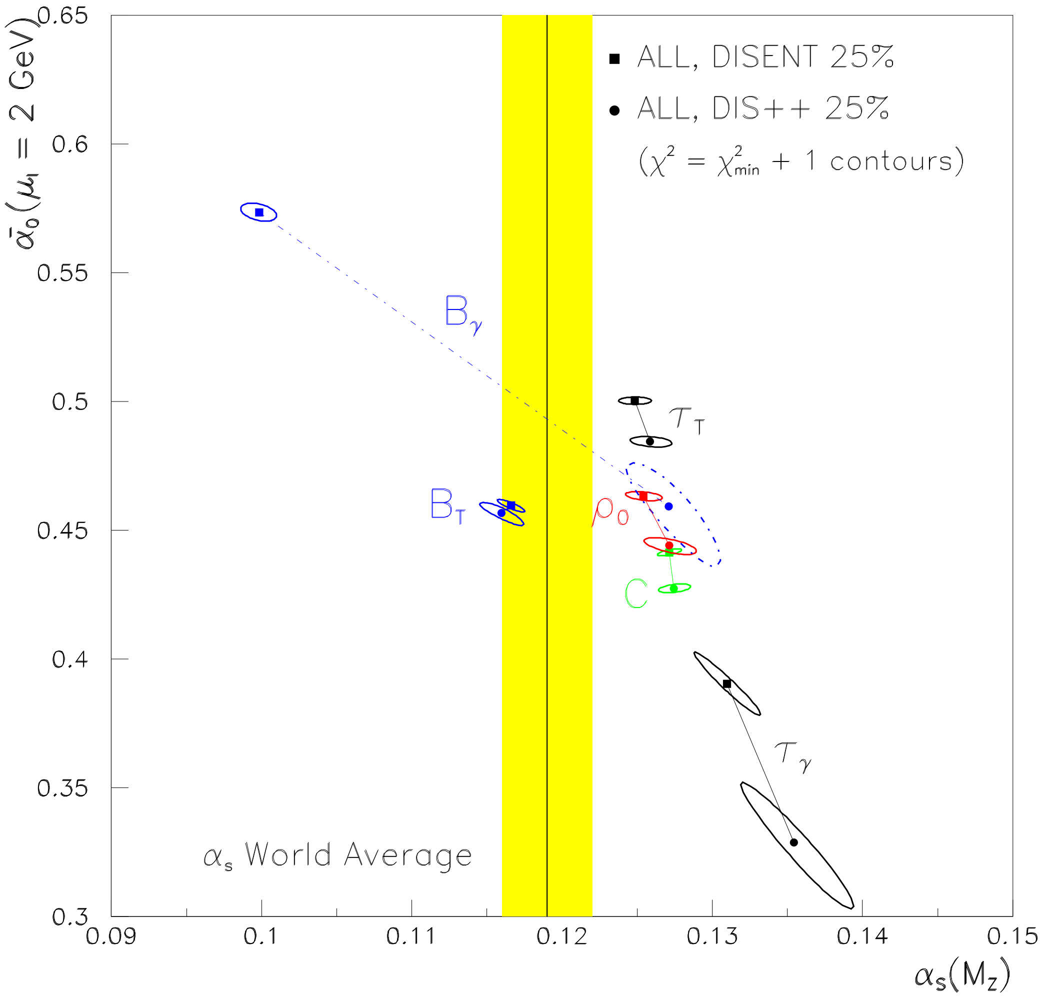


Figure 8.9: The fitted values of α_s and $\overline{\alpha_0}$ for the DISENT with the $\mathcal{E}_{lim} > 0.25Q$ energy cuts compared to DISASTER++ with the $\mathcal{E}_{lim} > 0.25Q$ energy cut. For clarity, the results for each variable are connected by a line. The DISASTER++ contour for B_γ is dot-dashed to help distinguish it from the other contours.

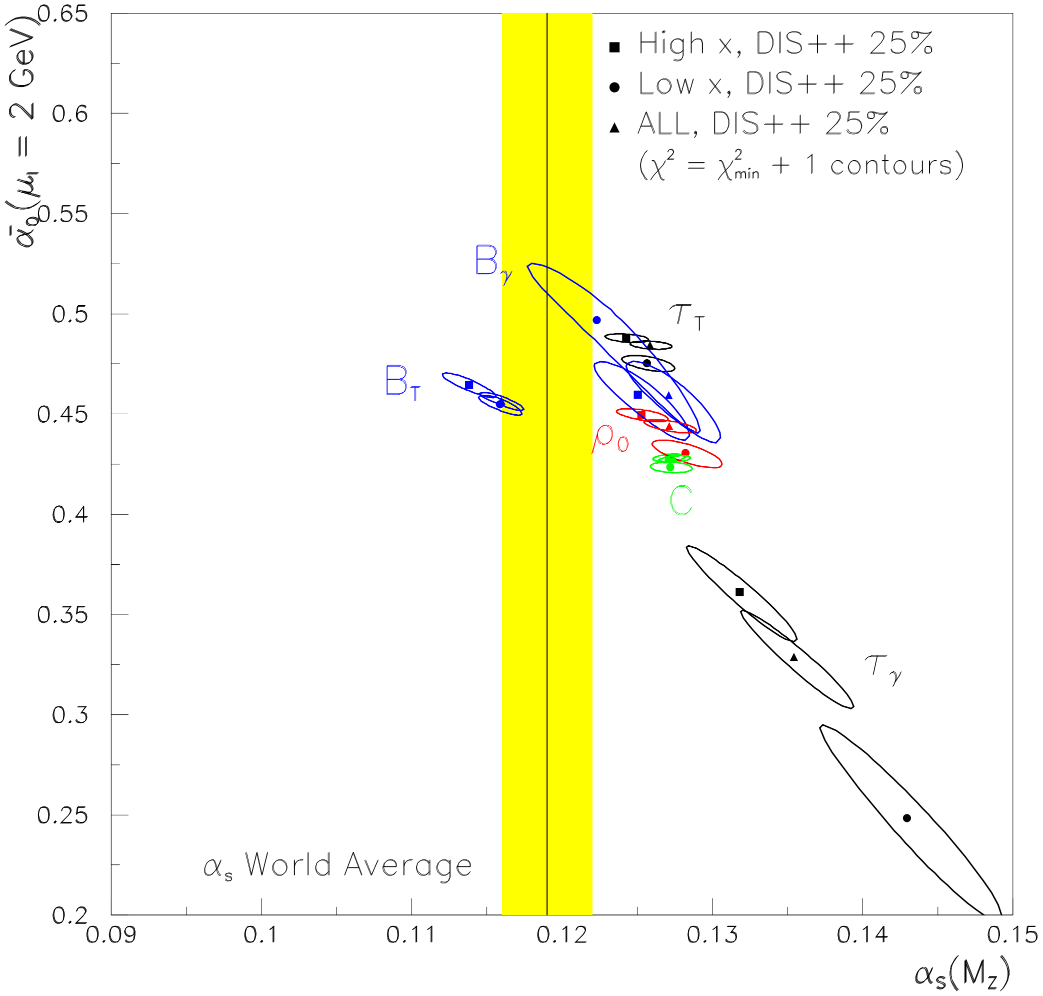


Figure 8.10: The fitted values of α_s and $\overline{\alpha_0}$ using the NLO prediction from DISASTER++ and a $\mathcal{E}_{lim} > 0.25Q$ energy cut. The contours are the 1σ statistical contours only. Each fit has been repeated three times, one for all data above 9 GeV, one fitting the low x points above 9 GeV only and the other fitting the high x points above 9 GeV only. The shaded vertical band is the world average range for $\alpha_s(M_Z)$.

puter time, preventing the renormalisation and factorisation scale calculations from being run. Therefore, the full systematic analysis was performed using the DISENT model with the $\mathcal{E}_{lim} > 0.25Q$ energy cut. Figures 8.11 to 8.16 show the effect of each of the systematic checks and cross-checks on the fitted values. The two cross-checks using the electron method to reconstruct the kinematics, and correcting the data with LEPTO, are plotted above the line. The experimental systematic uncertainties are plotted below the line and are included in the final quoted values of α_s and $\overline{\alpha_0}$. The combined uncertainties from varying the renormalisation and factorisation scales are quoted separately.

The dominant experimental uncertainty on the data points, and hence on the fitted values, is the model dependence from HERWIG (the cross-check from LEPTO has a similar large effect). The effect can be more clearly seen in Figure 8.17 which shows the data corrected using ARIADNE, HERWIG and LEPTO. Both HERWIG and LEPTO systematically increase the fitted value of α_s, but give shifts in opposite directions for the value of $\overline{\alpha_0}$. HERWIG and LEPTO share a similar parton fragmentation model (parton showering) differing from the Colour Dipole Model employed by ARIADNE, so it is the parton fragmentation model that appears to determine the systematic uncertainty on the fitted value of α_s. HERWIG and LEPTO differ in the hadronisation model, and it is this difference that leads to the systematic uncertainty on the fitted value of $\overline{\alpha_0}$. The model-dependent systematic shifts of the variables with respect to the photon axis are dominated by the reconstruction of the Breit frame axis, and are less sensitive to this effect (although the fitted value of α_s generally increases). The only other significant systematic effect is from varying the η and p_T cuts in the thrust with respect to the photon axis, but this effect is completely dominated by the model dependence systematic when combined in quadrature.

The electron method cross-check results in a systematic shift of a similar order of magnitude to the model dependence for the variables with respect to the photon axis, since these variables depend explicitly on the kinematic reconstruction of the Breit frame axis. The effect is comparatively small for the other variables.

The effect of varying the factorisation scale in the NLO model is generally smaller than the model dependence systematic, shifting the fitted values by 1-4%. This is small as expected, since the current region is somewhat insensitive to the factorisation scale. As illustrated by the shaded band in Figure 8.5, the dominant uncertainty on the fit comes from varying the renormalisation scale of the NLO prediction. This results in an approximately 15-20% shift in the fitted values of $\alpha_s(M_Z)$ and up to a 40% shift in the fitted values of $\overline{\alpha_0}(\mu_I)$. Table 8.4 summarises the final fitted values of $\alpha_s(M_Z)$ and $\overline{\alpha_0}(\mu_I)$, with the systematic checks combined in quadrature to form the experimental systematic error and the factorisation and renormalisation scale uncertainties combined in quadrature to form the theory error.

8.4 Statistical ⊕ Systematic Fit

The method discussed above allows the correlations between the systematic errors to be taken into account, but has the disadvantage that the final contour plots give the $\chi^2 = \chi^2_{min} + 1$ contours for the statistical errors only. An alternative method can be used as a cross-check to produce plots with systematic contours that are approximately correct, and to examine whether the dominant systematic errors result in significantly different central fitted values.

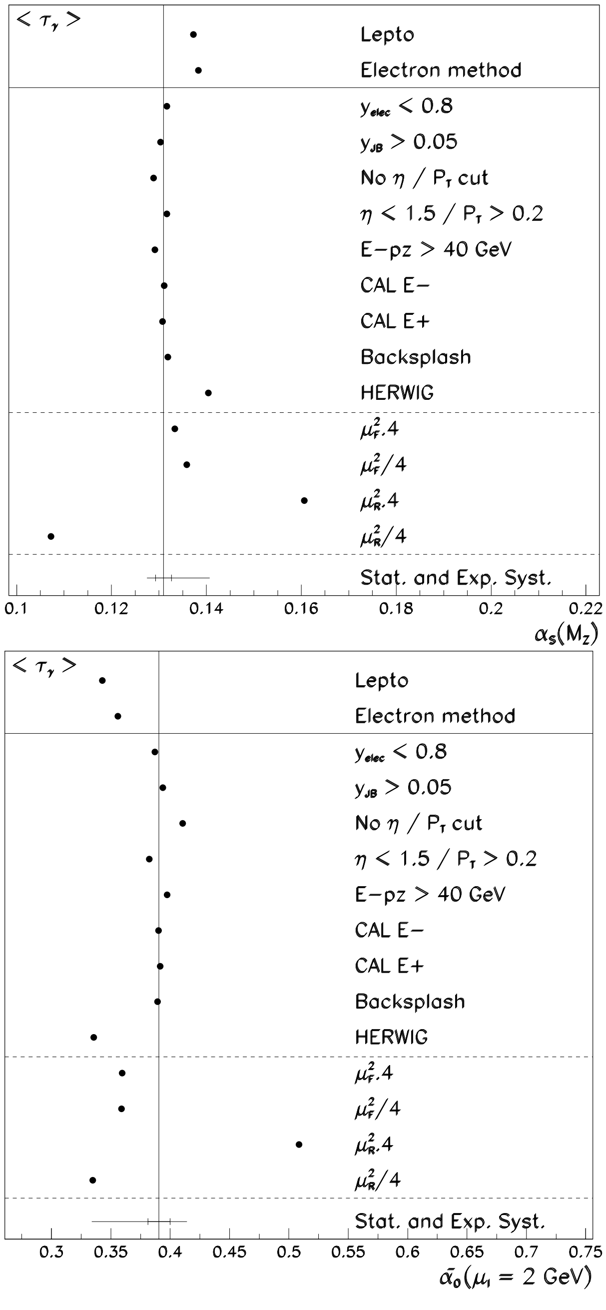


Figure 8.11: Systematic shifts in the fitted values of α_s and $\overline{\alpha_0}$ for τ_γ. The inner error bar at the bottom represents the statistical error, the outer represents the systematic error combined in quadrature with the statistical error.

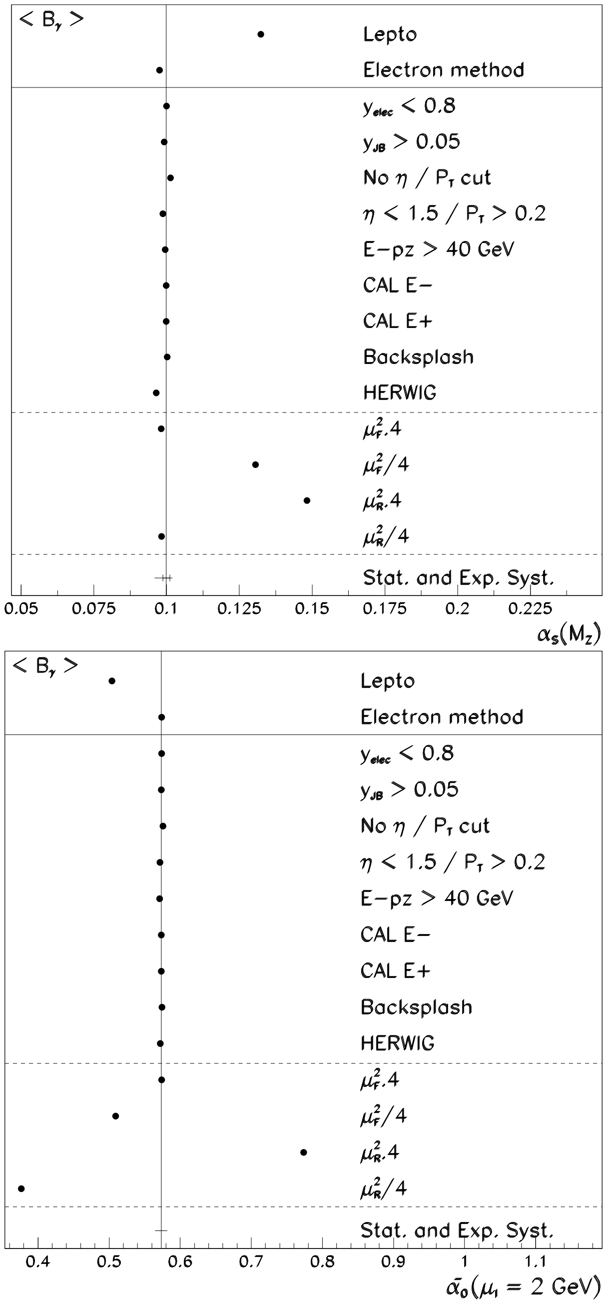


Figure 8.12: Systematic shifts in the fitted values of α_s and $\overline{\alpha_0}$ for B_γ. The inner error bar at the bottom represents the statistical error, the outer represents the systematic error combined in quadrature with the statistical error. The statistical error on $\overline{\alpha_0}$ is too small to show up on the plot.

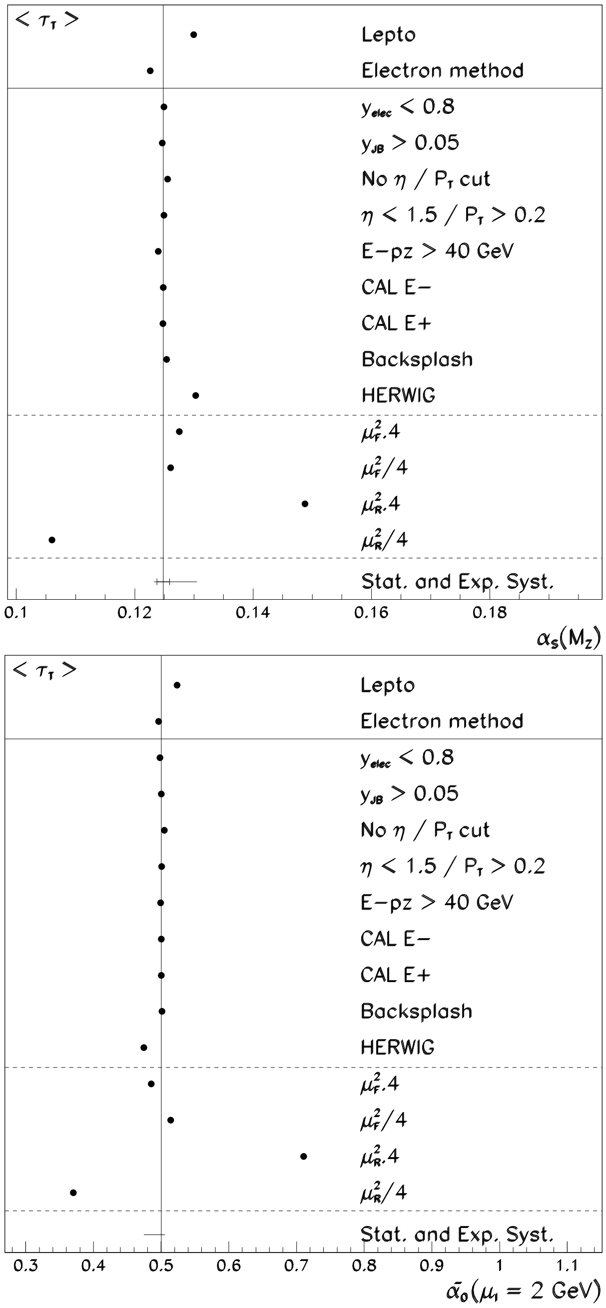


Figure 8.13: Systematic shifts in the fitted values of α_s and $\overline{\alpha_0}$ for τ_T. The inner error bar at the bottom represents the statistical error, the outer represents the systematic error combined in quadrature with the statistical error. The statistical error on $\overline{\alpha_0}$ is too small to show up on the plot.

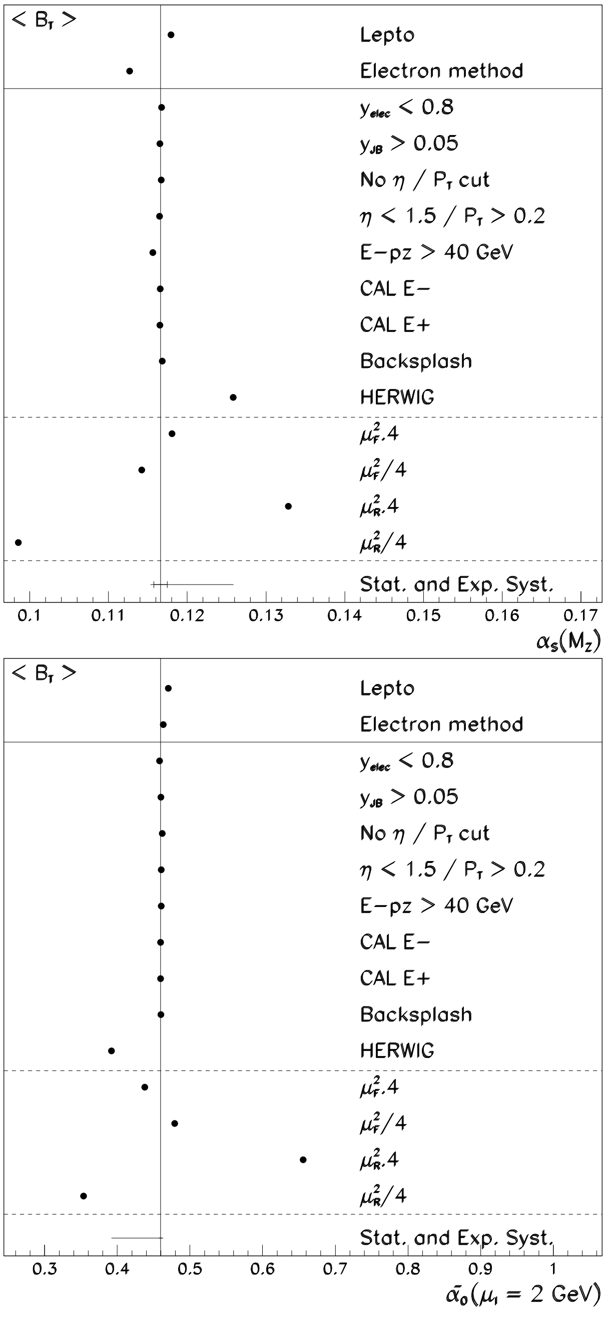


Figure 8.14: Systematic shifts in the fitted values of α_s and $\overline{\alpha_0}$ for B_T. The inner error bar at the bottom represents the statistical error, the outer represents the systematic error combined in quadrature with the statistical error. The statistical error on $\overline{\alpha_0}$ is too small to show up on the plot.

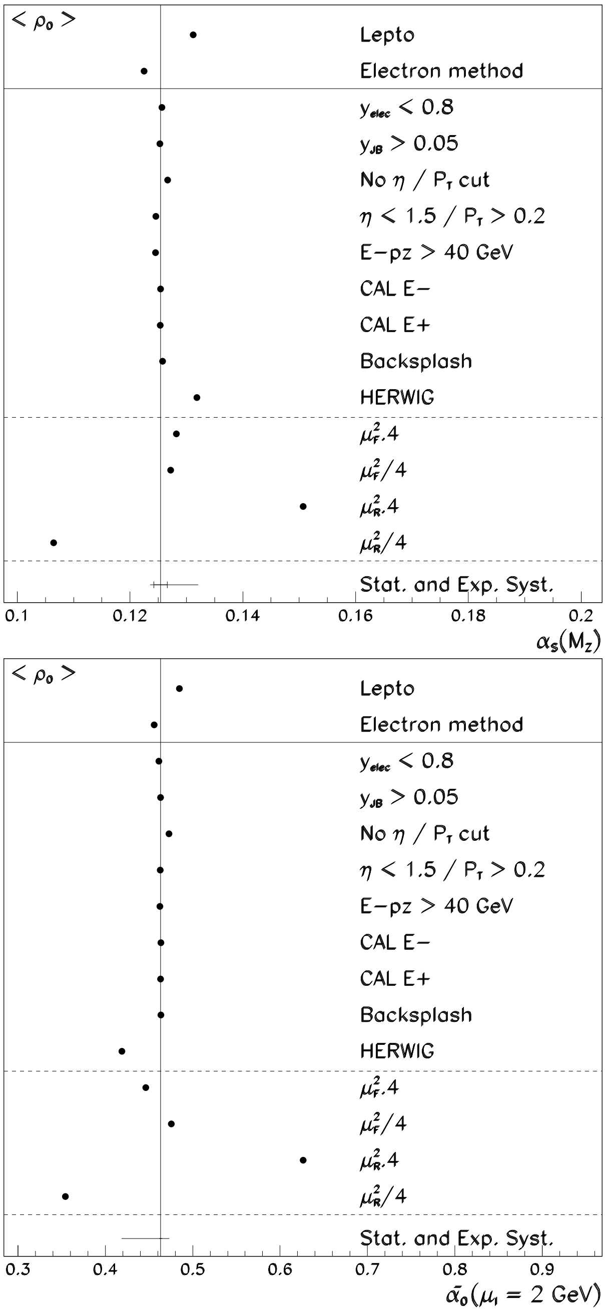


Figure 8.15: Systematic shifts in the fitted values of α_s and $\overline{\alpha_0}$ for ρ_0. The inner error bar at the bottom represents the statistical error, the outer represents the systematic error combined in quadrature with the statistical error. The statistical error on $\overline{\alpha_0}$ is too small to show up on the plot.

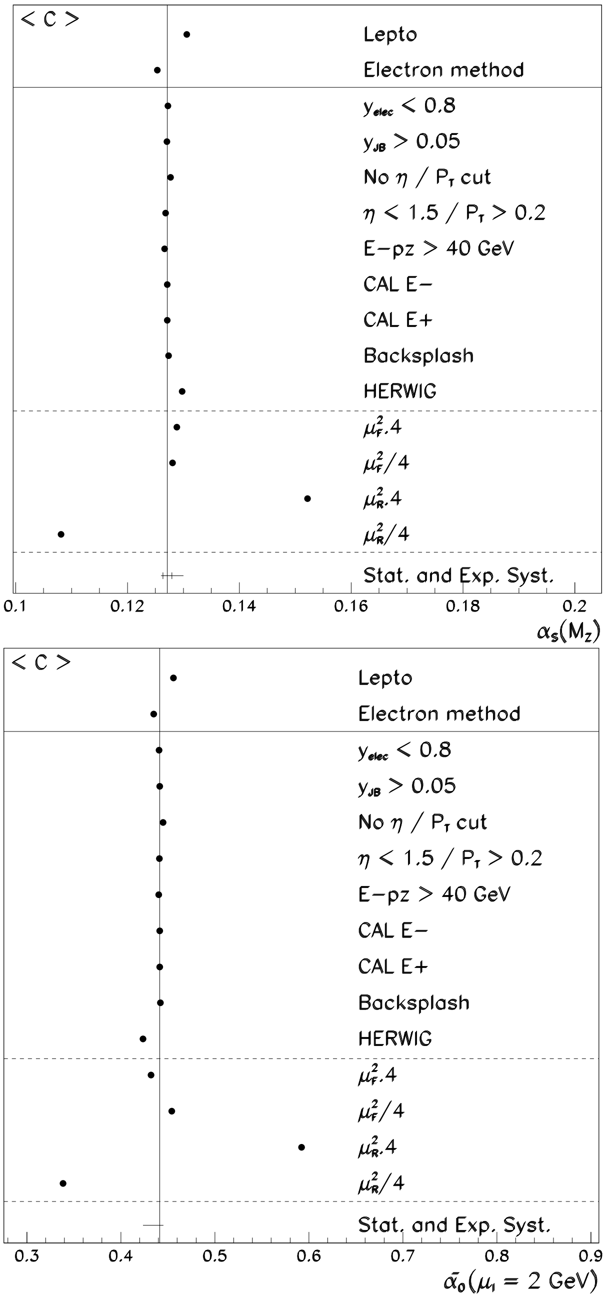


Figure 8.16: Systematic shifts in the fitted values of α_s and $\overline{\alpha_0}$ for C-parameter. The inner error bar at the bottom represents the statistical error, the outer represents the systematic error combined in quadrature with the statistical error. The statistical error on $\overline{\alpha_0}$ is too small to show up on the plot.

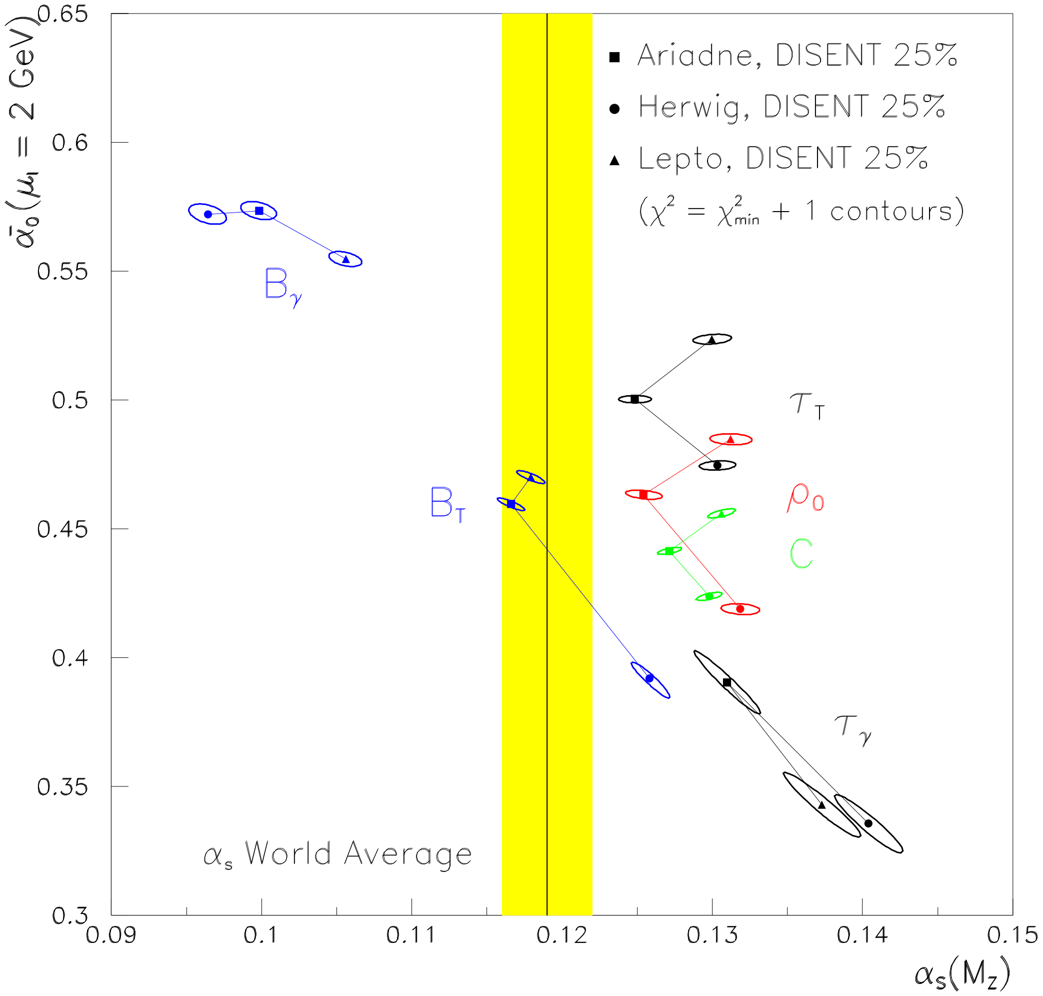


Figure 8.17: The fitted values of α_s and $\overline{\alpha_0}$ using the NLO prediction from DISENT and a $\mathcal{E}_{lim} > 0.25Q$ energy cut. The contours are the 1σ statistical contours only for data corrected with ARIADNE, HERWIG and LEPTO. The shaded vertical band is the world average range for $\alpha_s(M_Z)$. For clarity, the results for each of the variables are connected by a line.

Variable	$\alpha_s(M_Z)$	$\overline{\alpha_0}(\mu_I)$	χ^2/ndf
τ_γ	$0.1310 \pm 0.0017^{+0.0095+0.0301}_{-0.0029-0.0237}$	$0.3905 \pm 0.0093^{+0.0214+0.1179}_{-0.0556-0.0709}$	14.2/8
B_γ	$0.0998 \pm 0.0012^{+0.0015+0.0573}_{-0.0037-0.0023}$	$0.5734 \pm 0.0035^{+0.0025+0.2000}_{-0.0032-0.2070}$	247.3/8
τ_T	$0.1248 \pm 0.0011^{+0.0056+0.0241}_{-0.0009-0.0188}$	$0.5003 \pm 0.0014^{+0.0048+0.2105}_{-0.0258-0.1310}$	25.1/8
B_T	$0.1166 \pm 0.0009^{+0.0092+0.0163}_{-0.0010-0.0182}$	$0.4596 \pm 0.0023^{+0.0028+0.1969}_{-0.0677-0.1081}$	23.2/8
ρ_0	$0.1254 \pm 0.0012^{+0.0066+0.0255}_{-0.0013-0.0190}$	$0.4633 \pm 0.0017^{+0.0097+0.1634}_{-0.0445-0.1104}$	16.2/8
C	$0.1271 \pm 0.0008^{+0.0027+0.0251}_{-0.0006-0.0190}$	$0.4414 \pm 0.0014^{+0.0036+0.1512}_{-0.0177-0.1034}$	11.9/8

Table 8.4: Final fitted results for $\alpha_s(M_Z)$ and $\overline{\alpha_0}(\mu_I)$ using the NLO prediction from DISENT and a $\mathcal{E}_{lim} > 0.25Q$ energy cut. The statistical, systematic and theoretical uncertainties are given as $x \pm \text{statistical}^{+\text{systematic}+\text{theory}}_{-\text{systematic}-\text{theory}}$. The quoted χ^2 is that from the $\mathcal{E}_{lim} > 0.25Q$ central statistical fit.

The central data points were taken and assigned an error based on the statistical and systematic errors combined in quadrature. Since the systematic error is generally asymmetric (due to the dominance of the single model dependence systematic), the error was made symmetric and the fit performed with these errors. Figure 8.18 shows the resulting fits for a $\mathcal{E}_{lim} > 0.25Q$ energy cut using DISENT, compared to the fit for statistical errors only. In all cases except B_γ, the fitted value of $\overline{\alpha_0}$ decreases slightly. Here, all variables are consistent with the statistical fit except B_T, which differs in the fitted value of α_s by slightly more than 1σ, and B_γ which yields significantly different values of α_s. This instability is not unexpected due to the problems of the DISENT with power correction model in representing B_γ. The fits are repeated with DISASTER++ and the results given in figure 8.19. All variables are consistent with the statistical fit.

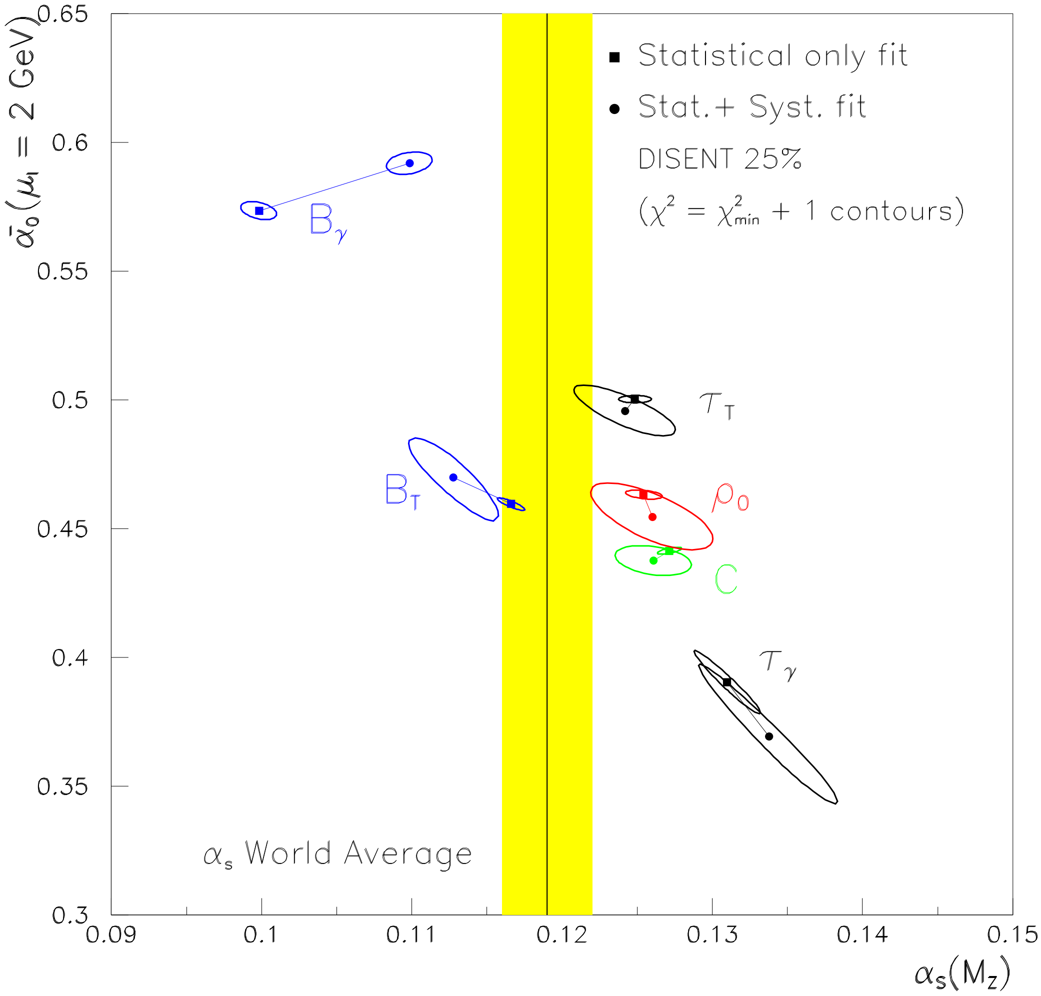


Figure 8.18: Cross-check comparing central fit using DISENT and a $\mathcal{E}_{lim} > 0.25Q$ energy cut (smaller contours) with the fit performed using the systematic and statistical errors combined in quadrature (larger contours). For clarity, the results for each of the variables are connected by a line.

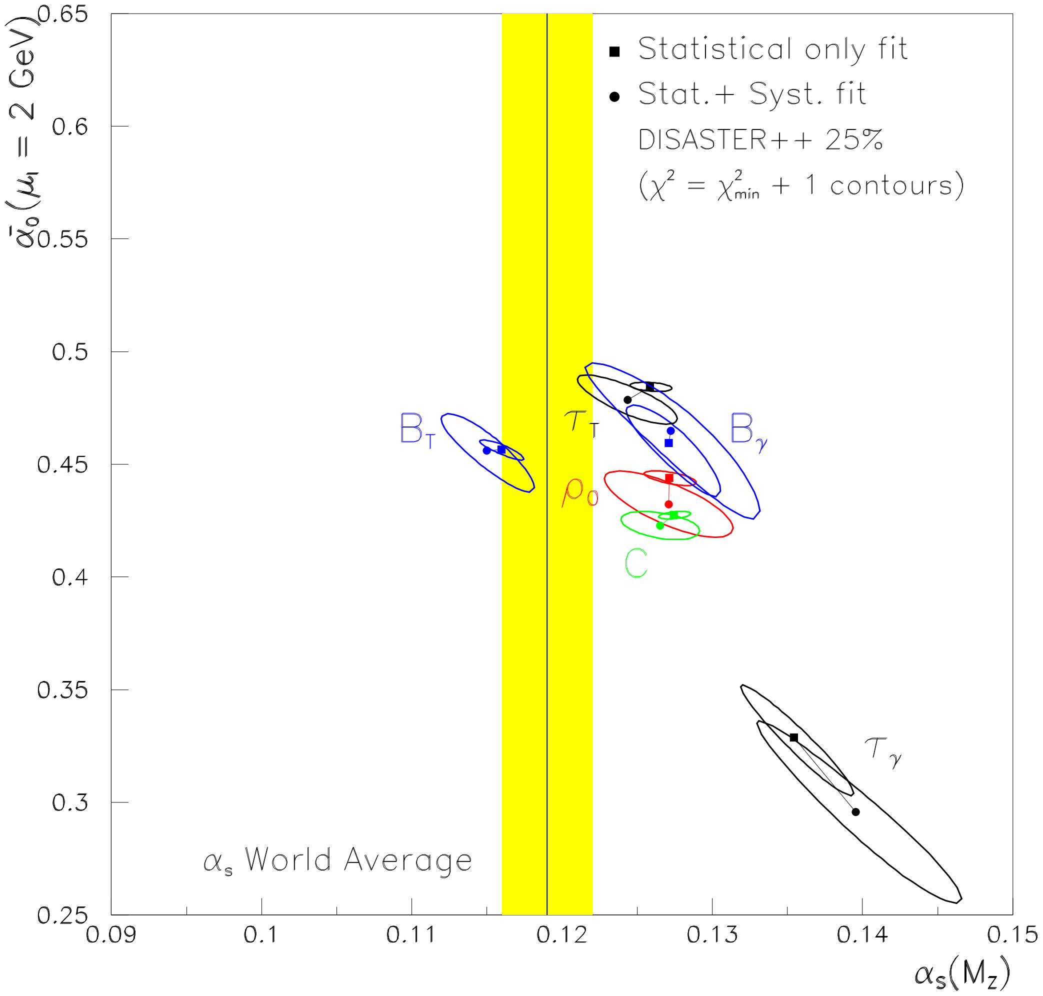


Figure 8.19: Cross-check comparing central fit using DISASTER++ and a $\mathcal{E}_{lim} > 0.25Q$ energy cut (smaller contours) with the fit performed using the systematic and statistical errors combined in quadrature (larger contours). For clarity, the results for each of the variables are connected by a line.

8.5 Effect of varying the Q Limit

The stability of the fit was investigated as a function of the minimum Q in the fit, Q_0. This was performed by lowering and raising the cut, by including extra bins, or by removing bins from the fit. Figures 8.20 and 8.21 show the variation, as a function of the lower Q limit, of the fitted values of $\alpha_s(M_Z)$ and $\overline{\alpha_0}(\mu_I)$ respectively. In general, the α_s and $\overline{\alpha_0}$ are relatively insensitive to the lower Q limit and all the variations are within the quoted systematic errors.

8.6 Comparison with other experiments

Figure 8.22 shows the ZEUS data compared to H1 data [30] and combined e^+e^- data [16]. The H1 result presents all variables except broadening with respect to the thrust axis. For τ_T, C and ρ_0, the two experiments are consistent within systematic errors. The fitted values of thrust and broadening with respect to the photon axis differ between the experiments; this is likely due to the strong anti-correlation between α_s and $\overline{\alpha_0}$ in the fits. The e^+e^- data were only available for the comparison of τ_T and C, (τ_z and B_z are DIS specific, B_T uses an incompatible definition, and ρ_0 was unavailable.) The perturbative part of the calculation used in the e^+e^- result included resummed higher order contributions which are absent from the perturbative part in the DIS fit. These theoretical resummed calculations have only just become available for the thrust in DIS [26]. The fitted values of $\overline{\alpha_0}$ are consistent between the experiments, however, the fitted values of α_s are considerably lower than in DIS, more consistent with the world average.

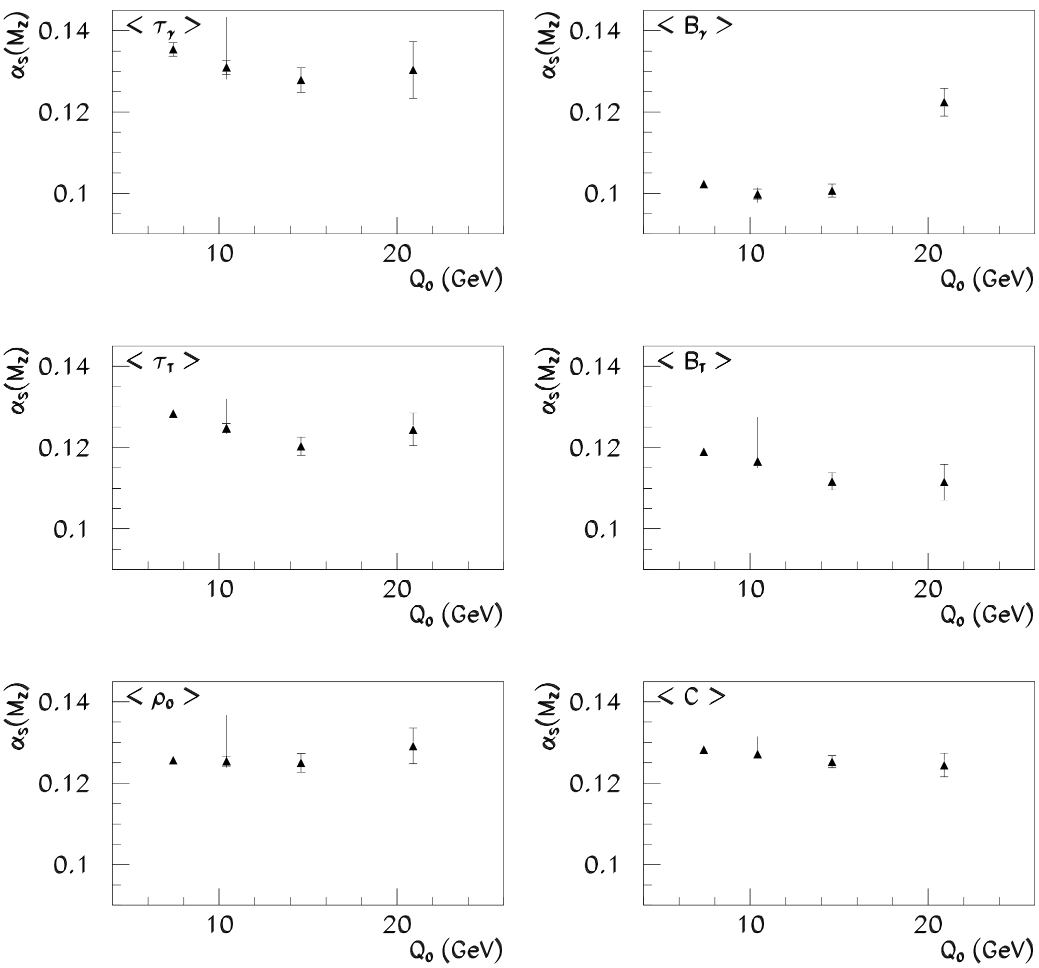


Figure 8.20: The extracted values of $\alpha_s(M_Z)$ as a function of the lower Q cut. All points show the statistical errors. The second point is the central analysis with the outer error bars being the systematic errors combined in quadrature with the statistical errors.

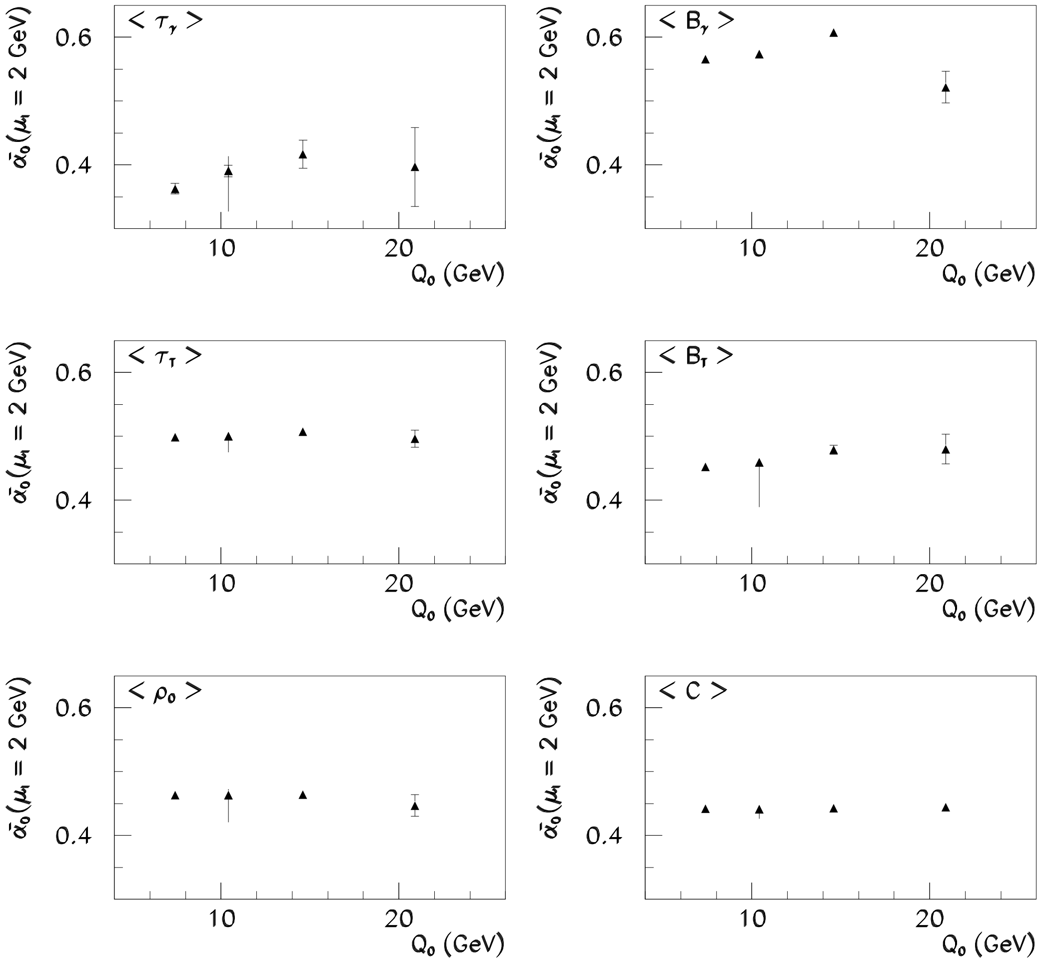


Figure 8.21: The extracted values of $\overline{\alpha_0}(\mu_I)$ as a function of the lower Q cut. All points show the statistical errors. The second point is the central analysis with the outer error bars being the systematic errors combined in quadrature with the statistical errors.

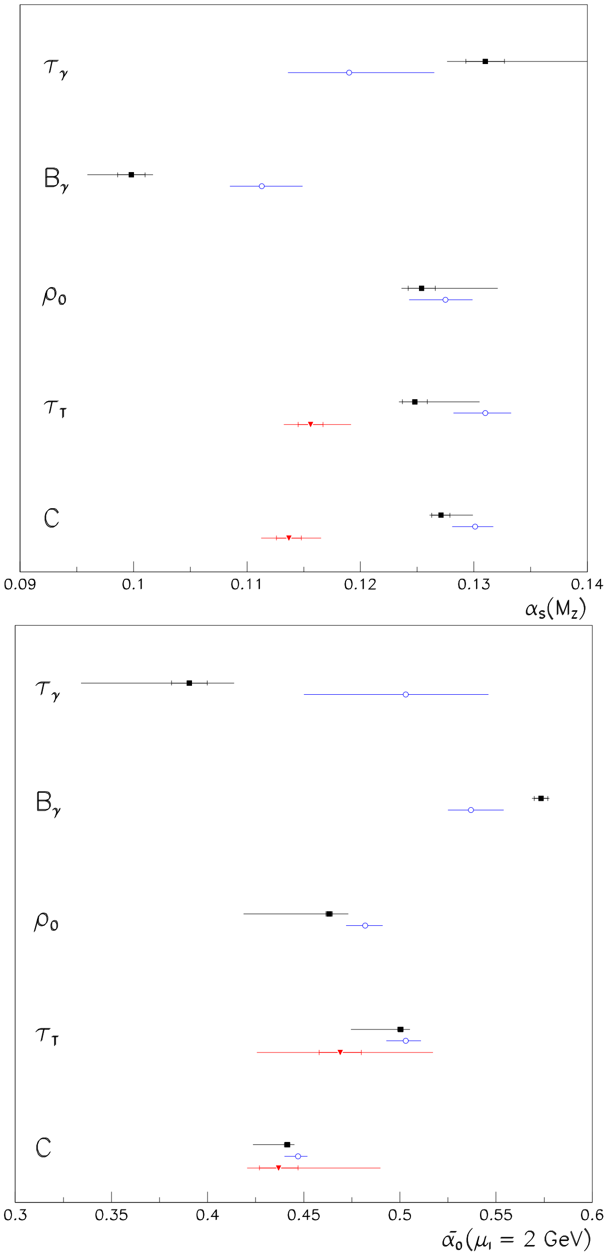


Figure 8.22: Comparison for the fitted values of α_s and $\overline{\alpha_0}$ of ZEUS data (black points) to H1 data (blue open points) and combined e^+e^- data (red triangles). The e^+e^- data were only available for τ_T and C. The inner bars are statistical errors only, and the outer bars are statistical in quadrature with systematic errors. The H1 points from [30] are quoted with statistical $\oplus$ systematic errors only. Note that the dominant theory error from the renormalisation and factorisation scale dependence is omitted from all experimental points.

8.7 Summary

The corrected mean data event shape data are fitted in order to test the hadronisation model of Dokshitzer, Webber et al. [28]. The data are fitted for $\mathcal{E}_{lim} > 0.10Q$ and $\mathcal{E}_{lim} > 0.25Q$, the energy cut in the current region. A significant instability with x is observed for B_γ using the combination of DISENT with the power correction theory. The DISASTER++ based model provides a better description of B_γ, consistent with the other variables.

The extracted values of $\alpha_s(M_Z)$ are reasonably consistent with each other (within a few standard deviations), and on average around 5% higher than the world average of 0.119. A 15% spread around 0.45 in the fitted values of $\overline{\alpha_0}(\mu_I)$ is observed. The experimental systematic effects are dominated by the dependence from the Monte Carlo model used to correct the data. Uncertainties in the parton fragmentation model (CDM or parton showering) systematically limit the determination of α_s, while the hadronisation model limits $\overline{\alpha_0}$. However, the dominant uncertainty on the fit is from varying the renormalisation scale of the NLO prediction, which is a factor 2-5 larger than the experimental systematic effects. The fitted values are insensitive to the lower Q limit. Other than the variables with respect to the photon axis which suffer from large anti-correlations in the fit, the ZEUS and H1 data agree within systematic errors. The data from e^+e^- experiments agree for the fitted values of $\overline{\alpha_0}$ but fit a lower value of α_s.

Chapter 9

Conclusions

Event shape distributions have been studied in the current region of the Breit frame for deep inelastic scattering. The thrust and broadening with respect to both the virtual photon and the thrust axis, and two axis-independent variables, the jet-mass and C-parameter, have been studied in the kinematic range $10 < Q^2 < 20480$ GeV^2 and $0.0006 < x < 0.6$. Both differential and mean distributions have been measured. The achieved experimental precision is typically 5%. The dominant systematic effect on the data points is from the model dependence, at approximately 5%.

The corrected data are well described by the Colour Dipole approach of ARIADNE, with the high-Q^2 phase space modification. HERWIG is good at high Q^2 but gives a poor description at low Q^2. A significant x-dependence is observed in the variables measured with respect to the photon axis, due to the transverse momentum of the incoming quark. This effect is well described by ARIADNE, and reasonably described by HERWIG.

The mean values of the event shapes have been fitted to a model combining NLO parton level perturbative predictions with the hadronisation power-correction model of Dokshitzer, Webber, et al. The fitted values of $\alpha_s(M_Z)$

and a new non-perturbative parameter, $\overline{\alpha_0}(\mu_I)$, are extracted in order to determine the validity of the model. The experimental precision on the fitted value for α_s is limited by the parton fragmentation model used to correct the data and the experimental precision on $\overline{\alpha_0}$ by the hadronisation model used. These effects are dominated (by a factor 2-5) by the renormalisation scale uncertainty of the NLO prediction.

Both the DISENT and DISASTER++ NLO programs coupled with the power correction theory provide a good description of all the variables except the jet broadening with respect to the photon axis. For B_γ, a significant instability with x is observed with the DISENT model and the fitted value of α_s is around 25% lower than the other variables. The B_γ fit with DISASTER++ is more stable with x and a value of α_s more consistent with the other variables is fitted.

Overall, the data support the concept of approximately universal power corrections, fitting a consistent value of $\overline{\alpha_0}$ to within 15%, (in agreement with e^+e^-). However, the fitted values of α_s are higher than the accepted world average value by around 5%, suggesting that the current model is inadequate. This is possibly due to missing resummed higher order corrections in the perturbative calculation; these are present in the e^+e^- result where the fit is consistent with the world average for α_s.

Further possible work in this area include a more detailed examination of the DISASTER++ calculation, including the renormalisation scale dependence. In order to resolve the question of whether the missing higher order calculations in DIS are the cause of the large fitted values of α_s, further theoretical work is required to provide these calculations. Finally, the measurement of other event shapes such as jet-based variables would also provide further valuable input.

References

[1] Demokritos of Abdera, public discussions and references therein, Thrace (420 BC).

[2] J. Dalton, *A new system of Chemical Philosophy*, Vol I (1808).

[3] J. J. Thomson, Phil. Mag. **44**, 293 (1897).

[4] E. Rutherford, Phil. Mag. VI, **37**, 581 (1919).

[5] P. A. M. Dirac, Proceedings of the Royal Society of London, **A117**, 610 (1928).

[6] C. D. Anderson, Science **76**, 238 (1932).

[7] R. P. Feynman, Phys. Rev. **76**, 749 (1949).

[8] M. Gell-Mann, Phys. Lett. **8**, 214 (1964).

[9] V.E. Barnes et al., PRL **12** 204 (1964).

[10] G. 't Hooft, Nucl. Phys. **B35**, 167 (1971).

[11] J. D. Bjorken, E. A. Paschos, Phys. Rev. **185**, 1975 (1969).

[12] C. G. Callan, D. Gross, Phys. Rev. Lett. **22**, 156 (1969).

[13] R.E. Taylor, Proceedings of the Int. Symposium on Lepton and Photon Interactions at High Energies, 679 (1975).

[14] G. Altarelli, G. Parisi, Nucl. Phys. **B126**, 298 (1997).

[15] R. P. Feynman, *Photon-Hadron Interactions*, (1972).

[16] JADE Collaboration, P.A. Movilla Fernandez et al., Eur. Phys. J. **C1**, 461 (1998).

[17] G. P. Salam, Private communication (1999).

[18] M. H. Seymour, Private communication (1998).

[19] S. Catani, M. H. Seymour, Phys. Lett. **B378**, 287 (1996).

[20] K. Fabricius, G. Kramer, G. Schierholz, I. Schmitt, Zeit. Phys. **C11**, 315 (1981).

[21] S. Catani, M. H. Seymour, Nucl. Phys. **B485** 291 (1997).

[22] M. H. Seymour, DISENT Program Manual, Version 0.1, (1997).

[23] Dirk Graudenz, DISASTER++ Program Manual, Version 1.0.1 (1997).

[24] T. Hadig, G. J. McCance, GLAS-PPE/1999-24, Proceedings of the Monte Carlo Generators for HERA Physics Workshop, DESY-PROC-1999-02, 125 (1999).

[25] G. J. McCance, GLAS-PPE/1999-25, Proceedings of the Monte Carlo Generators for HERA Physics Workshop, DESY-PROC-1999-02, 151 (1999).

[26] V. Antonelli, M. Dasgupta, G. P. Salam, JHEP **0007**, 54 (2000).

[27] Yu .L. Dokshitzer, B. R. Webber, Phys. Lett. **B352**, 451 (1995).

[28] Yu. L. Dokshitzer, B. R. Webber, Phys. Lett. **B404**, 321 (1997)
Yu. L. Dokshitzer, A. Lucenti, G. Marchesini, G. P. Salam,
Nucl. Phys. **B511**, 396 (1998).

[29] Yu. L. Dokshitzer, JHEP **9805** 3 (1998); M. Dasgupta et al.,
JHEP **9907** 12 (1999).

[30] H1 Collab., C. Adloff et al., Eur. Phys. J. **C14**, 255 (2000);
H1 Collab., C. Adloff et al., Eur. Phys. J. **C18** 417 (2000);
H-U. Martyn, talk at ICHEP 2000, hep-ex/0010046 (2000).

[31] Yu. L. Dokshitzer, G. Marchesini, G. P. Salam,
Eur.Phys.J. **C3**, 1 (1999).

[32] Nucl. Instr. and Meth. **A386**, 310 (1997).

[33] ZEUS Collaboration, *The ZEUS Detector Status Report 1993*,
DESY (1993).

[34] K. Ackerstaff et al., Nucl. Instr. and Meth. **A417** (1998) 230.

[35] HERA-B Collaboration, Technical Design Report,
DESY-PRC 95-001 (1995).

[36] H. Bethe and W. Heitler, Proc. Roy. Soc., **A146** 83 (1934).

[37] B. Foster et al., Nucl. Instr. and Meth. **A338** 254 (1994).

[38] D.S. Bailey et al., Nucl. Instr. and Meth. **A396** 320 (1997).

[39] A. Anderson et al., Nucl. Instr. and Meth. **A309** 101 (1991).

[40] B. Bernstein et. al, Nucl. Instr. and Meth. **A336** 23 (1993).

[41] M. Derrick et al., Nucl. Instr. and Meth. **A309** 77 (1991).

[42] J. Kruger, *The Uranium Scintillator Calorimeter for the ZEUS-detector at the Electron Proton Collider HERA*, DESY (1992).

[43] L. Hervas, DESY-F35D 91-001 (1991).

[44] E. Tscheslog, ZEUS Note 91-037 (1991).

[45] P. de Jong, Thesis, NIKHEF (1994).

[46] A. Fox-Murphy, ZEUS weekly meeting, DESY (3 May 1999).

[47] S. Bentvelsen, J. Engelen, P. Kooijman,
Proceedings of Physics at HERA Workshop, Vol. 1, 23 (1991).

[48] F. Jacquet, A. Blondel, Proceedings of an *ep* facility for Europe,
DESY 79-048 391, (1979).

[49] H. Abramowicz, A. Caldwell, R. Sinkus, DESY 95-054 (1995).

[50] R. Sinkus, T. Voss, Nucl. Inst. Methods **A389**, 160 (1997).

[51] D. Acosta, ZEUS 94-098, (1994).

[52] J. Grosse-Knetter, ZEUS 97-039 (1997).

[53] J. Grosse-Knetter, ZEUS 98-031 (1998).

[54] G.F. Hartner, ZEUS 98-058 (1998).

[55] S. Kananov, ZEUS 97-018 (1997).

[56] H. Beier, A. Caldwell, ZEUS 95-040 and references therein (1995).

[57] G. Briskin, Thesis, Tel Aviv University,
DESY-THESIS-1998-036 (1998).

[58] N. Tuning, ZEUS Note under preparation (2000).

[59] V. Gribov, L. Lipatov, Sov. Jour. Nucl. Phys. **15**, 438 (1972);
L. Lipatov, Sov. Jour. Nucl. Phys. **20** 94 (1975).

[60] ZEUS Collab. J. Breitweg et al., F_2 paper presented at ICHEP-00 (2000).

[61] CTEQ Collab., Phys. Rev. **D55**, 1280 (1997).

[62] G.Ingelman, A. Edin and J. Rathsman,
Comput. Physics Communications **101** 108 (1997).

[63] H. Spiesberger, DJANGOH v1.1 Manual
`http://www.desy.de/~hspiesb/djangoh.html`

[64] G. Marchesini, B.R. Webber, G. Abbiendi, I.G. Knowles, M.H. Seymour and L. Stanco, Comput. Phys. Commun. **67** 465 (1992).

[65] B. Andersson, G. Gustapson, L. Lönnblad,
Nucl. Phys. **B339** 393 (1990).

[66] ARIADNE program manual, first published by L. Lönnblad,
Comput. Phys. Commun. **71**, 15 (1992).

[67] T. Sjöstrand, Comput. Phys. Commun. **82**, 74 (1994).

[68] ZEUS Collab., J. Breitweg et al., Eur. Phys. J. **C6**, 43 (1993).

[69] H1 Collab., C. Adloff et al., Zeit. Phys. **C76**, 613 (1997).

[70] G. Ingelman and P.E. Schlein, Phys. Lett. **B152**, 256 (1985).

[71] H. Jung, RAPGAP program manual
`http://www.quark.lu.se/~hannes/rapgap.html` (2000).

[72] A. Doyle, C. Glasman, S. Hanlon, G. McCance, I. Skillicorn, R. Waugh, U. Wollmer, ZEUS Note under preparation (2001).

[73] R. Brun et al, GEANT3, CERN DD/EE/84-001 (1987).

[74] L. Lönnblad, Proceedings of the Monte Carlo Generators for HERA Physics Workshop, DESY-PROC-1999-02, 47 (1999).

[75] ZEUS Collab. J. Breitweg et al., Eur. Phys. J. **C11**, 251 (1999).

[76] MINUIT program manual, CERN Program Library Long Writeup D506, (1994).

[77] CERNLIB program manual, CERN Program Library Short Writeups, (1996).

Appendix A

Evaluation of Statistical Errors

The evaluation of statistical uncertainties on the correction factors, purities and efficiencies are complicated by migrations between bins when going from the generated truth level to the reconstructed level. As an example, consider the correction factor, κ, for one of the differential bins for an event shape F

$$\kappa(F) = \frac{1}{N_{\rm gen}} \left(\frac{dn}{dF}\right)_{\rm gen} \Bigg/ \frac{1}{N_{\rm rec}} \left(\frac{dn}{dF}\right)_{\rm rec} \tag{A.1}$$

where $\left(\frac{dn}{dF}\right)_{\rm gen}$ is the number of Monte Carlo events generated at truth level in a particular bin in F, x and Q^2, and $\left(\frac{dn}{dF}\right)_{\rm rec}$ is the number of Monte Carlo events reconstructed in the same bin (i.e. after the detector simulation) in F, x and Q^2. $N_{\rm gen}$ ($N_{\rm rec}$) are the total number of events generated (reconstructed) in a given x and Q^2 bin.

For the evaluation of the error on $\kappa(F)$, it is convenient to define three statistically independent quantities.

1. The total number of events generated *and* correctly reconstructed in the bin is denoted C.

2. The total number of events generated in the bin *but not* reconstructed

in that bin is denoted D. These measured events have been smeared out of the correct bin by the detector response.

3. The total number of events reconstructed in the bin *but not* generated in the bin is denoted E. These are events from other bins, or events from outside the true kinematic range, which due to measurement errors have been measured incorrectly in this bin.

Then, the total number of events generated in the bin is $a = C + D$, and the total number of events reconstructed in the bin is $b = C + E$. Clearly, it is desirable to choose the bins in such as way as to minimise D and E, since these come only from mismeasurements.

The correction factor can then be re-expressed as

$$\kappa(F) = \frac{a(F)}{b(F)} = \frac{C+D}{C+E} \tag{A.2}$$

and the purity ($\mathcal{P}$) and efficiency ($\mathcal{E}$) as

$$\mathcal{P} = \frac{C}{C+E} \tag{A.3}$$

$$\mathcal{E} = \frac{C}{C+D} \tag{A.4}$$

The error on the correction factor is then given by standard error propagation methods as

$$\delta\kappa = \sqrt{\left(\frac{d\kappa}{dC}\right)^2 \delta_C^2 + \left(\frac{d\kappa}{dD}\right)^2 \delta_D^2 + \left(\frac{d\kappa}{dE}\right)^2 \delta_E^2} \tag{A.5}$$

where the quantities C, D and E are independent, so there is no correlation term. Assuming the errors on the total number of events in the bin are given as $\delta_C = \sqrt{C}$, etc. then substituting and differentiating gives the final result as

$$\delta\kappa = \sqrt{\frac{C^2(D+E) + D^2(C+E) + E^2(C+D) + 2CDE}{(C+E)^4}} \tag{A.6}$$

The error in purity can be found by setting $D = 0$, and then the error in efficiency by symmetry using $E \to D$.

$$\delta\mathcal{P} = \sqrt{\frac{CE}{(C+E)^3}} \tag{A.7}$$

$$\delta\mathcal{E} = \sqrt{\frac{CD}{(C+D)^3}} \tag{A.8}$$

Appendix B

Tables of Data

Tables of mean data are presented in this appendix. References to the relevant section or chapter in the main text are given.

B.1 Mean Event Shape Data

The corrected and uncorrected mean values of the event shapes, and the correction factors are given in Tables B.1 to B.6. The mean data are plotted, and the correction technique discussed in chapter 7. The tables also give the positive and negative systematic deviations on the event shape means, combined in quadrature, as discussed in section 7.3.

Bin	Uncorrected Mean	stat. error	Correction Factor	Corrected Mean	stat. error	Positive Systematic	Negative Systematic
1	0.4780	0.0013	1.1000	0.5258	0.0021	0.0010	-0.0105
2	0.4752	0.0013	1.0823	0.5143	0.0021	0.0019	-0.0135
3	0.4499	0.0015	1.0649	0.4790	0.0024	0.0000	-0.0009
4	0.4387	0.0013	1.0509	0.4611	0.0020	0.0012	-0.0092
5	0.4182	0.0022	1.0661	0.4458	0.0035	0.0004	-0.0021
6	0.4039	0.0014	1.0349	0.4180	0.0023	0.0011	-0.0032
7	0.3582	0.0008	1.0442	0.3740	0.0010	0.0019	-0.0016
8	0.3398	0.0010	0.9737	0.3309	0.0012	0.0040	-0.0091
9	0.3200	0.0016	1.0254	0.3281	0.0018	0.0052	-0.0012
10	0.2950	0.0012	0.9646	0.2845	0.0014	0.0039	-0.0035
11	0.2558	0.0017	0.9739	0.2492	0.0019	0.0021	-0.0018
12	0.2256	0.0027	0.9582	0.2162	0.0031	0.0086	-0.0030
13	0.1949	0.0039	0.8828	0.1721	0.0044	0.0117	-0.0073
14	0.1590	0.0062	0.7584	0.1206	0.0072	0.0213	-0.0100
15	0.1642	0.0116	0.8016	0.1316	0.0132	0.0076	-0.0161
16	0.1298	0.0199	0.5727	0.0743	0.0238	0.0415	-0.0446

Table B.1: Mean event shape results for τ_γ.

Bin	Uncorrected Mean	stat. error	Correction Factor	Corrected Mean	stat. error	Positive Systematic	Negative Systematic
1	0.3815	0.0005	1.0497	0.4004	0.0009	0.0001	-0.0073
2	0.3818	0.0005	1.0362	0.3956	0.0009	0.0007	-0.0094
3	0.3709	0.0006	1.0245	0.3800	0.0010	0.0002	-0.0011
4	0.3651	0.0005	1.0216	0.3730	0.0008	0.0004	-0.0058
5	0.3548	0.0010	1.0307	0.3657	0.0015	0.0005	-0.0009
6	0.3496	0.0006	1.0132	0.3543	0.0010	0.0003	-0.0028
7	0.3265	0.0004	1.0145	0.3313	0.0004	0.0008	-0.0007
8	0.3155	0.0005	0.9810	0.3095	0.0006	0.0013	-0.0067
9	0.3049	0.0008	1.0060	0.3067	0.0009	0.0036	-0.0007
10	0.2897	0.0006	0.9773	0.2831	0.0007	0.0014	-0.0036
11	0.2658	0.0009	0.9799	0.2605	0.0011	0.0010	-0.0010
12	0.2437	0.0016	0.9763	0.2380	0.0018	0.0053	-0.0018
13	0.2159	0.0024	0.9489	0.2049	0.0027	0.0050	-0.0042
14	0.1865	0.0041	0.8993	0.1677	0.0048	0.0095	-0.0055
15	0.1827	0.0077	0.8944	0.1634	0.0088	0.0045	-0.0109
16	0.1588	0.0164	0.7581	0.1204	0.0191	0.0266	-0.0232

Table B.2: Mean event shape results for B_γ.

Bin	Uncorrected Mean	stat. error	Correction Factor	Corrected Mean	stat. error	Positive Systematic	Negative Systematic
1	0.1395	0.0005	1.0710	0.1494	0.0008	0.0002	-0.0015
2	0.1466	0.0006	1.0398	0.1524	0.0009	0.0001	-0.0018
3	0.1513	0.0006	1.0821	0.1638	0.0010	0.0001	-0.0089
4	0.1606	0.0005	1.0452	0.1679	0.0008	0.0005	-0.0100
5	0.1442	0.0009	1.1229	0.1620	0.0014	0.0002	-0.0096
6	0.1579	0.0006	1.0322	0.1630	0.0010	0.0010	-0.0081
7	0.1451	0.0003	1.0307	0.1496	0.0004	0.0004	-0.0055
8	0.1575	0.0004	0.9729	0.1533	0.0005	0.0019	-0.0114
9	0.1304	0.0006	1.0140	0.1322	0.0007	0.0006	-0.0035
10	0.1424	0.0005	0.9491	0.1352	0.0006	0.0012	-0.0078
11	0.1260	0.0007	0.9153	0.1153	0.0008	0.0011	-0.0009
12	0.1070	0.0010	0.9257	0.0990	0.0012	0.0010	-0.0006
13	0.0965	0.0015	0.8982	0.0867	0.0017	0.0021	-0.0014
14	0.0879	0.0027	0.8885	0.0781	0.0030	0.0027	-0.0040
15	0.0779	0.0045	0.9105	0.0709	0.0050	0.0037	-0.0039
16	0.0745	0.0106	0.9130	0.0680	0.0114	0.0011	-0.0194

Table B.3: Mean event shape results for τ_T.

Bin	Uncorrected Mean	stat. error	Correction Factor	Corrected Mean	stat. error	Positive Systematic	Negative Systematic
1	0.2050	0.0005	1.0175	0.2086	0.0008	0.0003	-0.0028
2	0.2094	0.0005	1.0119	0.2119	0.0008	0.0001	-0.0043
3	0.2122	0.0006	1.0350	0.2197	0.0009	0.0001	-0.0104
4	0.2184	0.0005	1.0164	0.2219	0.0008	0.0005	-0.0114
5	0.2051	0.0008	1.0504	0.2154	0.0013	0.0002	-0.0076
6	0.2144	0.0005	1.0094	0.2164	0.0009	0.0013	-0.0093
7	0.2022	0.0003	1.0064	0.2035	0.0003	0.0003	-0.0063
8	0.2104	0.0004	0.9787	0.2059	0.0004	0.0011	-0.0125
9	0.1873	0.0005	0.9957	0.1865	0.0006	0.0004	-0.0045
10	0.1947	0.0004	0.9659	0.1881	0.0005	0.0004	-0.0090
11	0.1763	0.0006	0.9522	0.1678	0.0007	0.0008	-0.0030
12	0.1559	0.0010	0.9603	0.1497	0.0011	0.0005	-0.0012
13	0.1383	0.0016	0.9635	0.1332	0.0018	0.0020	-0.0015
14	0.1271	0.0029	0.9611	0.1221	0.0033	0.0024	-0.0051
15	0.1159	0.0051	0.9755	0.1130	0.0057	0.0031	-0.0033
16	0.1132	0.0117	0.9662	0.1094	0.0128	0.0029	-0.0193

Table B.4: Mean event shape results for B_T.

Bin	Uncorrected Mean	stat. error	Correction Factor	Corrected Mean	stat. error	Positive Systematic	Negative Systematic
1	0.0831	0.0003	1.1205	0.0932	0.0005	0.0001	-0.0019
2	0.0861	0.0004	1.0762	0.0927	0.0006	0.0009	-0.0019
3	0.0846	0.0004	1.0905	0.0922	0.0006	0.0009	-0.0051
4	0.0899	0.0003	1.0484	0.0943	0.0005	0.0012	-0.0074
5	0.0784	0.0005	1.1226	0.0880	0.0008	0.0002	-0.0037
6	0.0852	0.0003	1.0341	0.0881	0.0006	0.0011	-0.0050
7	0.0771	0.0002	1.0405	0.0802	0.0002	0.0009	-0.0029
8	0.0826	0.0002	0.9897	0.0817	0.0003	0.0018	-0.0089
9	0.0695	0.0003	1.0259	0.0713	0.0004	0.0005	-0.0019
10	0.0737	0.0003	0.9870	0.0727	0.0003	0.0015	-0.0065
11	0.0660	0.0004	0.9589	0.0633	0.0004	0.0011	-0.0024
12	0.0575	0.0006	0.9575	0.0551	0.0007	0.0006	-0.0008
13	0.0512	0.0009	0.9623	0.0492	0.0010	0.0008	-0.0019
14	0.0456	0.0015	0.9412	0.0429	0.0017	0.0002	-0.0024
15	0.0415	0.0026	0.9705	0.0402	0.0029	0.0017	-0.0017
16	0.0388	0.0057	0.8361	0.0325	0.0065	0.0027	-0.0098

Table B.5: Mean event shape results for ρ_0.

Bin	Uncorrected Mean	stat. error	Correction Factor	Corrected Mean	stat. error	Positive Systematic	Negative Systematic
1	0.4903	0.0015	1.0447	0.5122	0.0023	0.0011	-0.0043
2	0.5085	0.0015	1.0267	0.5221	0.0024	0.0003	-0.0082
3	0.5252	0.0017	1.0579	0.5556	0.0026	0.0004	-0.0257
4	0.5500	0.0014	1.0264	0.5645	0.0022	0.0011	-0.0286
5	0.5107	0.0024	1.0868	0.5551	0.0037	0.0007	-0.0217
6	0.5473	0.0016	1.0206	0.5586	0.0025	0.0030	-0.0244
7	0.5182	0.0009	1.0208	0.5289	0.0010	0.0013	-0.0172
8	0.5482	0.0011	0.9822	0.5385	0.0012	0.0042	-0.0385
9	0.4782	0.0016	1.0095	0.4827	0.0018	0.0019	-0.0126
10	0.5075	0.0013	0.9664	0.4904	0.0015	0.0027	-0.0297
11	0.4589	0.0019	0.9407	0.4317	0.0021	0.0034	-0.0111
12	0.3990	0.0029	0.9448	0.3770	0.0033	0.0023	-0.0035
13	0.3580	0.0044	0.9254	0.3312	0.0050	0.0058	-0.0055
14	0.3265	0.0079	0.9048	0.2954	0.0089	0.0046	-0.0111
15	0.2899	0.0134	0.9128	0.2646	0.0151	0.0114	-0.0097
16	0.2786	0.0309	0.8917	0.2484	0.0339	0.0066	-0.0458

Table B.6: Mean event shape results for C.