The first measure of relative partonic contributions in DIS at HERA using an average hadronic charge

Robert Kenny Griffiths

Department of Physics,

Queen Mary and Westfield College,

University of London

November 1998

A thesis submitted in accordance with the regulations

for the Degree of Doctor of Philosophy

in the University of London

Abstract

High Energy Deep Inelastic Scattering events at the H1 detector at the HERA collider are used for analysis in the current region of the Breit frame. The analysis is concerned with extracting a correlation between the hadrons produced by the fragmentation of a struck parton in DIS events and with that struck parton's charge.

We define a fragmentation variable, $x_p = 2p_h/Q$, where Q is the four-momentum transfer between the lepton and the struck quark, and p_h is the momentum that a hadron, h, possesses. The hadronic tracks are weighted according to their momenta, so that more energetic tracks have a bigger weight. The entire data-sample is normalised by an average weight. We further define a variable, C_j, to be a normalised momentum-weighted charge sum of the whole hadronic system for a given event.

The C_j distributions show nodal structure at points corresponding to net charges of -1, 0, $+1$. To emphasise this behaviour, we take the ratio of the distribution's mean to its standard deviation, and maximise it with respect to an initiating quark.

The method outlined in this analysis may be used to provide a statistical method of decomposing proton structure functions into partonic structure functions. Consistency checks on this method are also presented in this thesis. The method's reliability was high, and was robust under all performed tests.

Further analysis using the Mandelstam variable $\hat{s}$, i.e. the invariant centre-of-mass energy of the hard sub-process, was performed in the hope that it would provide information about the different kinematics used in various Monte Carlo simulations to model some classes of physics events, e.g., Boson-Gluon Fusion, QCD-Compton. Splitting the sample into two sub-groups (one with high $\hat{s}$, where we are clearly in the perturbative QCD region; the other low $\hat{s}$) provided no further information than already obtained in an inclusive treatment. An investigation of events having different classes of $C_j = 0.0$ also provided no additional information.

The method described in this thesis can be used to extract the relative partonic contributions in DIS at HERA and the results obtained are presented.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other institute of learning.

Contents

List of Figures

List of Tables

I Mam a Dad

Preface

The thesis presents an analysis performed using data collected at the H1 detector on the HERA machine at DESY, Hamburg during 1994.

The Standard Model of particle physics pertaining to Deep Inelastic Scattering (DIS) and its background contamination are described in chapter 1. Proton structure functions and Quantum Chromodynamic (QCD) models are also discussed, and a description is given of the Breit frame of reference.

The components of the H1 detector are discussed in chapter 2. Particular attention is paid to those parts of the detector used for this analysis. Software used for data collection and processing is also described.

QCD models are further detailed in chapter 3 whilst describing the different uses of computer simulations, known as Monte Carlo. Various Monte Carlo models are used and their different fragmentation schemes are discussed and contrasted. The treatment is concluded by a discussion of some shortcomings of the simulations discovered during the completion of this thesis.

Data selection criteria are presented in chapter 4, with a brief discussion of the technical and physics reasons that necessitate such selections, and a comparison of data and Monte Carlo spectra.

As stated in the abstract, the analysis is concerned with extracting a correlation between the hadrons produced by the fragmentation of a struck parton in DIS events and with that struck parton's charge. A full treatment of the analysis is presented, including the motivation and aims, a suitable definition and subsequent testing of some candidate observables, up to the final implementation and use of the chosen variable. The variable, C_j, is a normalised momentum-weighted charge sum over the hadrons of a given event, where the momentum information, p_h, is embodied within the fragmentation variable, $x_p = 2p_{\mathrm{hadron}}/Q$, where Q is the magnitude of the four-momentum transfer between the lepton and the struck quark. Various Monte Carlo programs are used to produce "ideal" C_j distributions for each parton, and these are then used to fit to an experimentally-obtained data C_j distribution, with coefficients reflecting the contribution from each parton species.

Further investigations on events having $C_j \equiv 0.0$, and variations of C_j with the hard sub-process centre of mass energy are also described. Results are presented and discussed, concluding with a treatment of the ramifications and possible further uses of the methods outlined.

Chapter 1

Theory

1.1 Introduction

The HERA machine at DESY collides 820 GeV protons and 27.6 GeV positrons resulting in a total centre-of-mass energy of $\sim$ 301 GeV . This large energy not only increases the kinematic phase-space at high 4-momentum transfer, Q^2, but also at low Q^2 and low Bjorken x. At high Q^2, conventionally taken to mean Q^2 >100 GeV^2 for DIS at H1, the physics region is expanded by at least an order of magnitude over fixed target experiments [1], and at low Q^2, defined to be 12.0 $\leq Q^2 \leq$ 80.0 GeV^2 in DIS at H1, it is possible to probe to an order of magnitude smaller than fixed-target experiments. This enlarged phase-space is rich in physics, the principles of which will be described in this chapter. Figure 1.1 illustrates the range of Q^2 and x achieved during 1994 running at H1, with various kinematic selections (described in section 1.2) also shown. These contours include the areas as defined by parts of the detector; i.e., 157.0° $\leq \theta_{\mathrm{electron}} \leq$172.5° for the Backward Electromagnetic Calorimeter (see section 2.5.2 for a fuller description) and also $\theta_{\mathrm{electron}} \leq$150.0° for the liquid argon calorimeter (see section 2.5.1) ensures good containment of the electron, whereas 10.0$\leq \theta_{\mathrm{quark}} \leq$150° ensures that the quark jet remains within the Liquid Argon calorimeter. Other contours are kinematic constraints which ensure event quality and track purity, e.g. requiring that the scattered electron has energy of at least 14 GeV , or that the invariant mass-squared of the hadronic final state must be greater than 4400 GeV^2 . It is these data, collected during 1994 running, that are used in the analysis presented in this thesis.

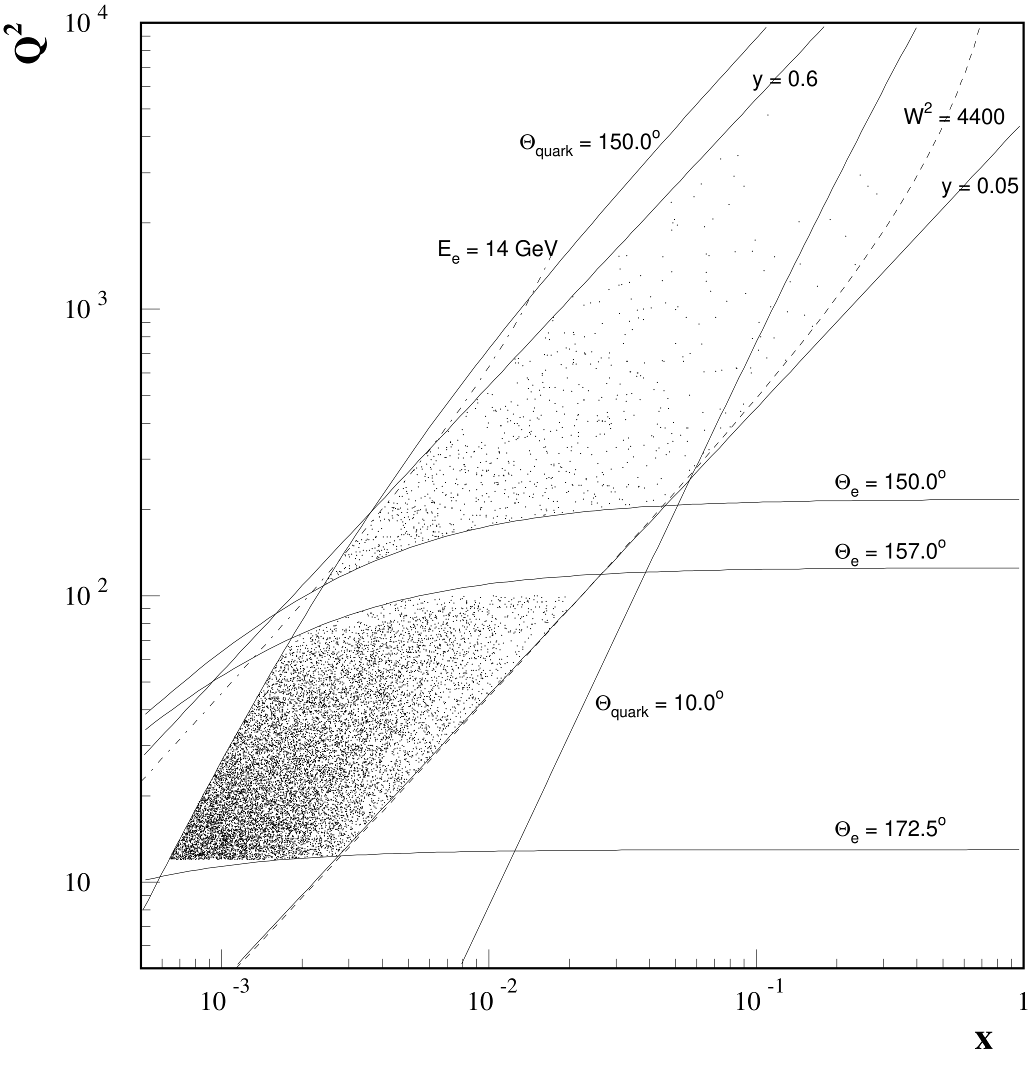

Figure 1.1: The $Q^2 - x$ plane at H1

1.2 Deep Inelastic Scattering

The deep inelastic scattering (DIS) process may be represented by $ep \to l' \ X$, where l is either an electron (positron) or a neutrino (ν) depending on whether the interaction is by the exchange of, respectively, a neutral current (γ,Z) or a charged current ($W^{\pm}$). A typical, lowest order (or Born term) NCDIS event is shown in figure 1.2.

A cartoon diagram of a typical DIS event, for both low and high Q^2 cases is shown in figure 1.3. A Boson-Gluon fusion event and possible QCD Compton events are shown in figures 1.4 and 1.5 respectively. Both of these processes can affect the determination of Q^2 which is vital in performing the boost to the Breit frame if, for example, one of the QCD Compton gluons is not detected, or similarly if one of the quarks in the BGF event travels down the beam-pipe.

The kinematics of such processes at fixed centre of mass energy, $\sqrt{s}$, can be described completely by the use of two independent variables. The variables used at H1 (and those conventionally used) are Q^2, the negative four-momentum transfer squared, and the scaling quantity x. The third variable, another scaling quantity, y, is constrained by the other two quantities according to the formula

$$\mathrm{Q}^2 = sxy \tag{1.1}$$

We define the kinematic variables x, y, and Q^2 (as enumerated above the detector on figure 1.2) and these are expressed, in 4-vector notation, in equations 1.2, 1.3 and 1.4 below

$$\mathrm{Q}^2 = -q^2 = -(k - k')^2 \tag{1.2}$$

$$y = \frac{P{\cdot}q}{P{\cdot}k} \tag{1.3}$$

$$x = \frac{\mathrm{Q}^2}{2P{\cdot}q} \tag{1.4}$$

We can further define a variable for the invariant mass-squared of the hadronic final state, W^2, as follows

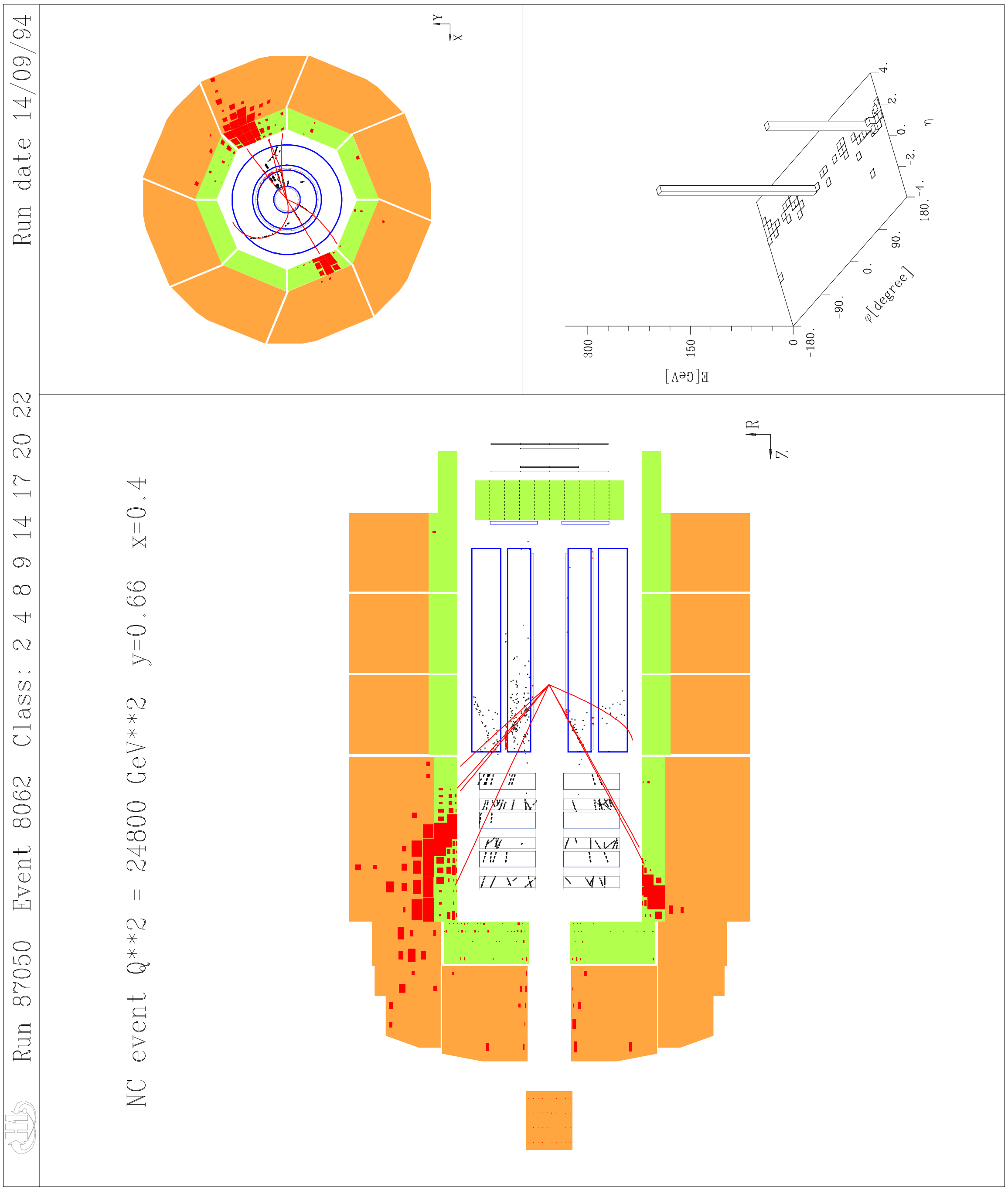

Figure 1.2: A typical neutral current deep inelastic scattering event

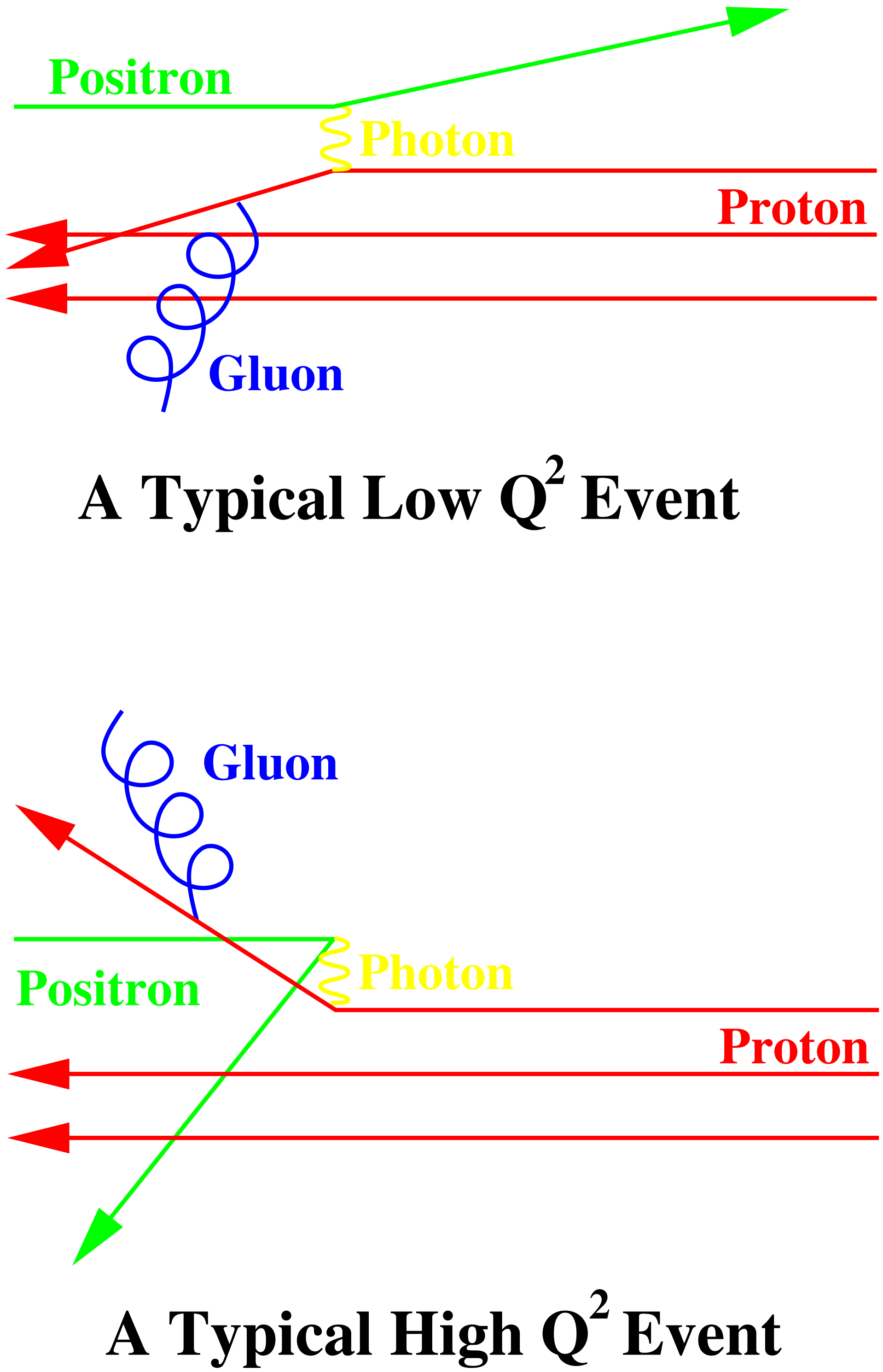

Figure 1.3: A cartoon illustrating a DIS event at both low and high Q^2

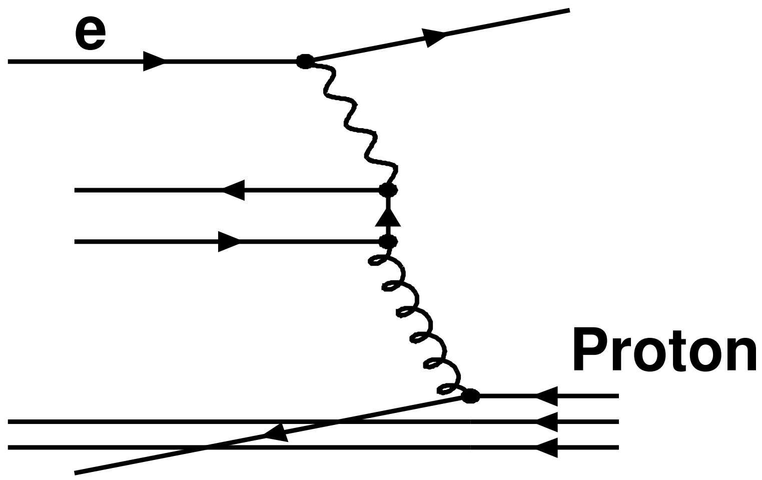

Figure 1.4: A diagram illustrating boson-gluon fusion

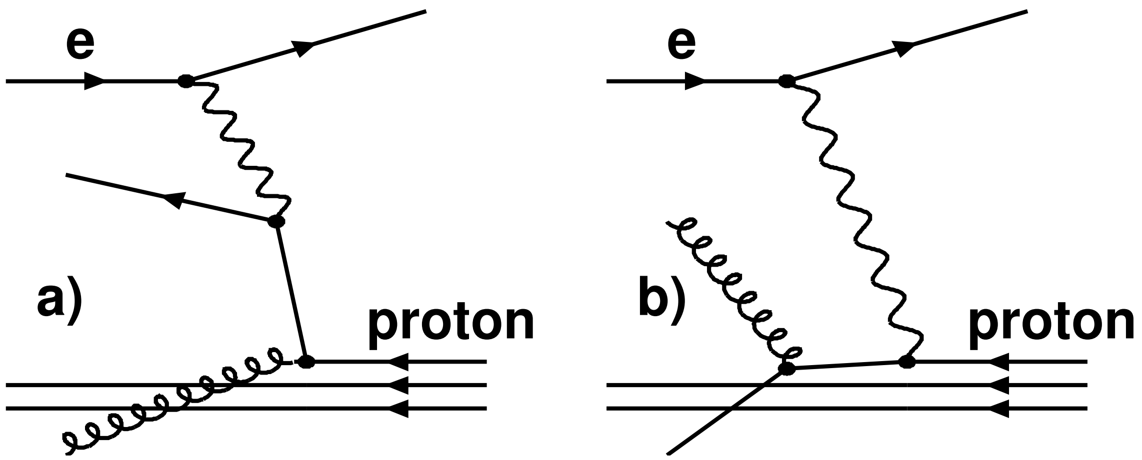

Figure 1.5: A diagram illustrating both initial- and final-state QCD radiation, labelled a) and b) respectively

$$W^2 = (\mathbf{P} + \mathbf{q})^2 \tag{1.5}$$

which, neglecting the mass of the proton, becomes

$$W^2 = \mathrm{Q}^2 \left(\frac{1-x}{x}\right) \tag{1.6}$$

or

$$x \sim \frac{\mathrm{Q}^2}{\mathrm{Q}^2 + W^2} \tag{1.7}$$

providing another method to measure the value of x.

The Lorentz invariant scaling quantities x and y are, respectively, the momentum fraction of the proton in the infinite momentum frame carried by the struck (naïve Quark Parton Model) quark, and the fraction of the energy transferred from the initial electron to the hadronic system, in the rest frame of the proton. This quantity, y, can be thought of as the inelasticity of the system.

The H1 apparatus is a "4π" detector and measures not only the scattered electron, but also the hadronic final state produced by the struck parton and the proton remnant. By measuring both of these quantities, the kinematics are over-constrained and can therefore be determined or calculated in a variety of ways. These methods include electron-only, hadronic-only, double-angle and mixed methods. The electron-only method determines Q^2 and y from the scattered electron's energy and polar angle; the hadron-only method, also known as the Jacquet-Blondel method, determines Q^2 and y from all the DIS final state hadrons. The double angle method requires the polar angles of both the scattered electron and current jet (struck quark) to determine Q^2 and y. The mixed method determines Q^2 from the scattered electron energy but the y it obtains from the Jacquet-Blondel, or hadronic method. The electron method is used at H1 because it has the best resolution for x in the kinematic range studied. A full description of these methods may be found elsewhere [2].

Current beam energies achieve a centre-of-mass energy of $s = 87,600$ GeV^2 , where

$$s = (k + p)^2 \sim 4E_e E_p \tag{1.8}$$

It can be seen that, using $\mathrm{Q}^2 = sxy$, values of x less than $\sim 10^{-3}$ can be reached in the region where DIS is easily the dominant process, *i.e.*, $\mathrm{Q}^2 > 0.4$ GeV^2 , for the massless (real) photon Q^2 is also a direct measure of "virtuality".

Not all processes at HERA depend on extremely "virtual" photons however. Most of the cross-section is at $Q^2 \sim 0$ (the transition point usually being $Q^2 > 0.4$ GeV^2), termed *quasi-real* photons, give rise to another branch of study at HERA, namely that of Photoproduction.

As this analysis is concerned with the analysis of non-diffractive DIS events, both photoproduction and large rapidity gap (diffractive) events are considered to be backgrounds that we wish to eliminate, and a brief description of both of these categories may be found below in sections 1.3 and 1.4.

1.3 Photoproduction

Photoproduction is the term used to describe a just off-mass-shell photon interacting with the proton. The cross-section for such a process is far higher than that for DIS and is a major source of background to DIS measurements. This is because the hadron production is so large that misidentification of hadrons as electrons is possible.

Photoproduction can be categorised into 3 types of interaction :-

Direct processes where the photon couples directly to a quark in the proton. Examples are Boson-Gluon Fusion (BGF) or QCD Compton as in DIS.

Vector Dominance Model is where the photon fluctuates into a low mass vector meson, e.g., ρ, ω, ϕ which then interacts with the proton. These processes can be both soft and hard.

Resolved Processes where structure is observed inside the photon, and one of its virtual quarks or gluons interacts with a valence parton. As this is a process which resolves quarks and gluons, a "photon remnant" jet is formed at very low angle, *i.e.*, along the photon direction almost parallel to the electron beam direction. This jet is produced in addition to those produced by the partonic interaction. The highest momentum exchange is not at the photon vertex anymore, but at the vertex where the virtual parton interacts with the proton's parton.

As has been mentioned above, the photon can exhibit partonic structure and, because of this, a structure function for the photon, F_γ can be modelled or used. A major analysis at H1 is the measurement of the photon structure function.

Many photoproduction processes have their equivalents in DIS, e.g., QCDC or BGF, which contaminate the pure DIS signal. These contaminants must be eliminated in any study of DIS physics. The usual separation between the classes of physics are "low-Q^2" for photoproduction and "high-Q^2" for DIS, though the individual interpretations of "low" and "high" vary.

1.4 Diffractive or Large Rapidity Gap Events

This is a class of DIS physics where the proton either does not fragment or fragments into a low mass system. Diffractive events are recognised by their characteristic signature, *i.e.*, no energy in the forward region of the detector, and are known as rapidity gap events. A "rap-gap" cartoon is shown in figure 1.6. A scattered electron is still detected in the backward region (in the BEMC, see section 2.5.2) implying some kind of interaction, and jets are often seen, implying a hard process at some point. As this is not the proton, another theory was proposed involving the exchange of a colourless object, often called the Pomeron, emitted from the proton. To date, it is still unclear what exactly constitutes a Pomeron, but one way often proposed is a gluon ladder, which is colourless. Diffractive events are not considered in this analysis due to the difficulty of modelling these in any way other than with an *ansatz*.

1.5 Proton Structure Functions

Neutral Current DIS events may be described by a differential cross-section [3] for inclusive ep scattering (ignoring electroweak effects (F_3) and radiative QED) as follows

$$\frac{d^2\sigma}{dx dQ^2} = \frac{4\pi\alpha^2}{Q^4}\left[(1-y)\frac{F_2(x,Q^2)}{x} + \frac{y^2}{2}\cdot\frac{2xF_1(x,Q^2)}{x}\right] \tag{1.9}$$

We shall use $F_2(x, Q^2) = 2xF_1(x, Q^2)$, the equality known as the Callan-Gross relation [4], which implies that the point-like constituents of the proton have spin-1/2 (this follows from a consideration of helicity conservation.)

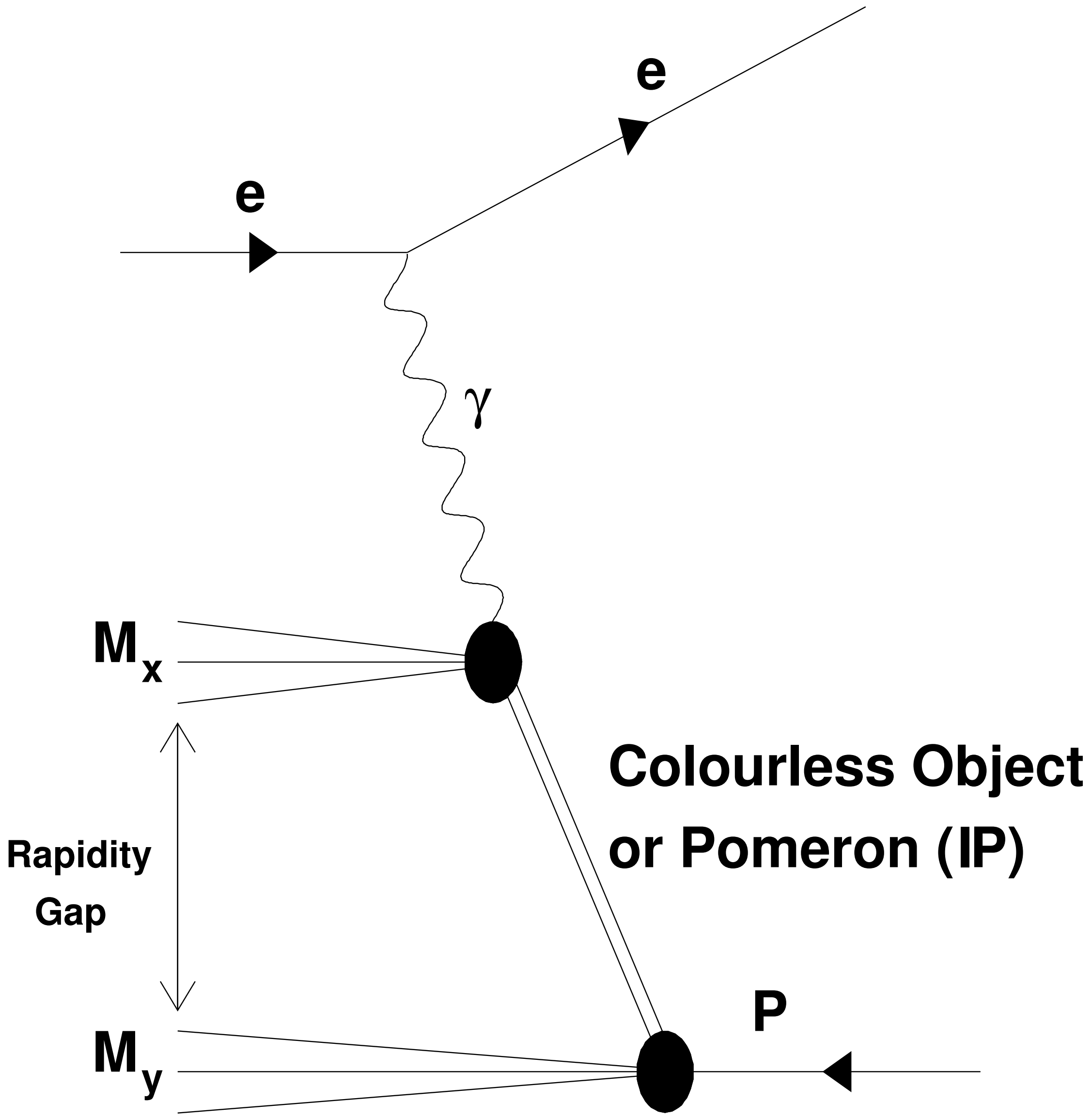

Figure 1.6: Illustration of a typical rapidity gap event

Equation 1.9 and the Callan-Gross relation are valid in the limit of the Quark-Parton Model. This is a model in which the proton is thought of, and treated as, a collection of free point-like charged partons that do not interact with one-another. QCD interactions violate the Callan-Gross relation. Ignoring these interactions, DIS can be treated as elastic scattering between electrons and partons, the cross-section for which can be calculated fully using QED.

The proton structure function, F_2 [5] is described as a sum of the individual quark and antiquark distribution or density functions, $q_i(x)$, $\bar{q}_i(x)$ for quarks and anti-quarks, of charge e_i times that of the electron

$$F_2^{\mathrm{ep}}(x, \mathrm{Q}^2) = \sum_i {e_i}^2 x[q_i(x) + \bar{q}_i(x)] \tag{1.10}$$

These partonic distribution functions, although modelled, are still experimentally unknown for the most part (especially light quarks) leading only to *inclusive* measurements of the structure function. This thesis will outline a possible method of obtaining these parton distribution functions.

1.6 QCD Models

QCD processes may be calculated, for massless particles, according to exact first order ($\mathcal{O}(\alpha_s)$) matrix elements (ME). Turning off $\mathcal{O}(\alpha_s)$ matrix elements corresponds to just modelling the QPM. The higher orders are simulated by means of, e.g., a leading logarithm approximation. The struck quark can emit partons either before or after the vertex, at least in an incoherent approach, but the way in which it becomes off-mass-shell is different. Before the vertex, it becomes negatively virtual, whereas after the vertex it is positively off-mass-shell, and then returns to its mass-shell by radiating particles which form the final state shower. The amount of gluon radiation can vary with either Q^2 or W^2. These scales are totally different in magnitude, and a fuller description of how the two approaches are combined is dealt with elsewhere [6].

1.7 The Breit Frame

The Breit frame [7] is a frame of reference co-linear with the hadronic centre of mass frame (see figure 1.7), but further boosted along the z axis such that the incoming virtual photon is space-like, *i.e.*, it has no energy or transverse momentum. The four-momentum of the photon is entirely contained in the z component, $P_z = -Q$, such that

$$q_{Bf} = (0, 0, 0, -Q) \tag{1.11}$$

The negative z-direction is defined to be that of the incoming photon, called the current hemisphere [8], and the massless (QPM-like) parton's trajectory will be along the positive z-axis. This direction is called the target hemisphere. Previous analyses [6], [9] have shown that the current hemisphere of the Breit frame can be compared to one half of an e^+e^- hemisphere in which the centre of mass energy E^* is equivalent to Q (see figure 1.8). The current hemisphere is defined by $cos\theta_{Bf} \leq 0.0$. In the Breit frame, and according to the Quark Parton Model (QPM), the massless quark will have momentum $p_z = +Q/2$ before, and $p_z = -Q/2$ after the interaction with the virtual photon, as shown in both figures 1.7 and 1.8. Figure 1.9 is a cartoon which illustrates the extent of $\cos\theta_{BF}$ and how it relates to the current and target hemispheres.

Free quarks cannot be seen in nature, consequently, the struck quark must somehow create other quarks which can then combine with each other to form hadrons in processes known as fragmentation and hadronisation. A dimensionless fragmentation variable may be defined in the Breit frame, in accordance with e^+e^- practice, such that

$$x_p = \frac{P_h}{Q/2} \tag{1.12}$$

where P_h is the momentum of the fragmented hadron.

The Breit frame is *not* the centre of mass system at H1 and has different properties to other CMS frames. The Breit frame is boosted along the z-axis and can be compared most readily to one half of an e^+e^- event. Comparisons with, say, the hadronic CMS will show differences such as the current hemisphere of the Breit frame being populated mainly by the fragments of the struck quark (unlike the hadronic

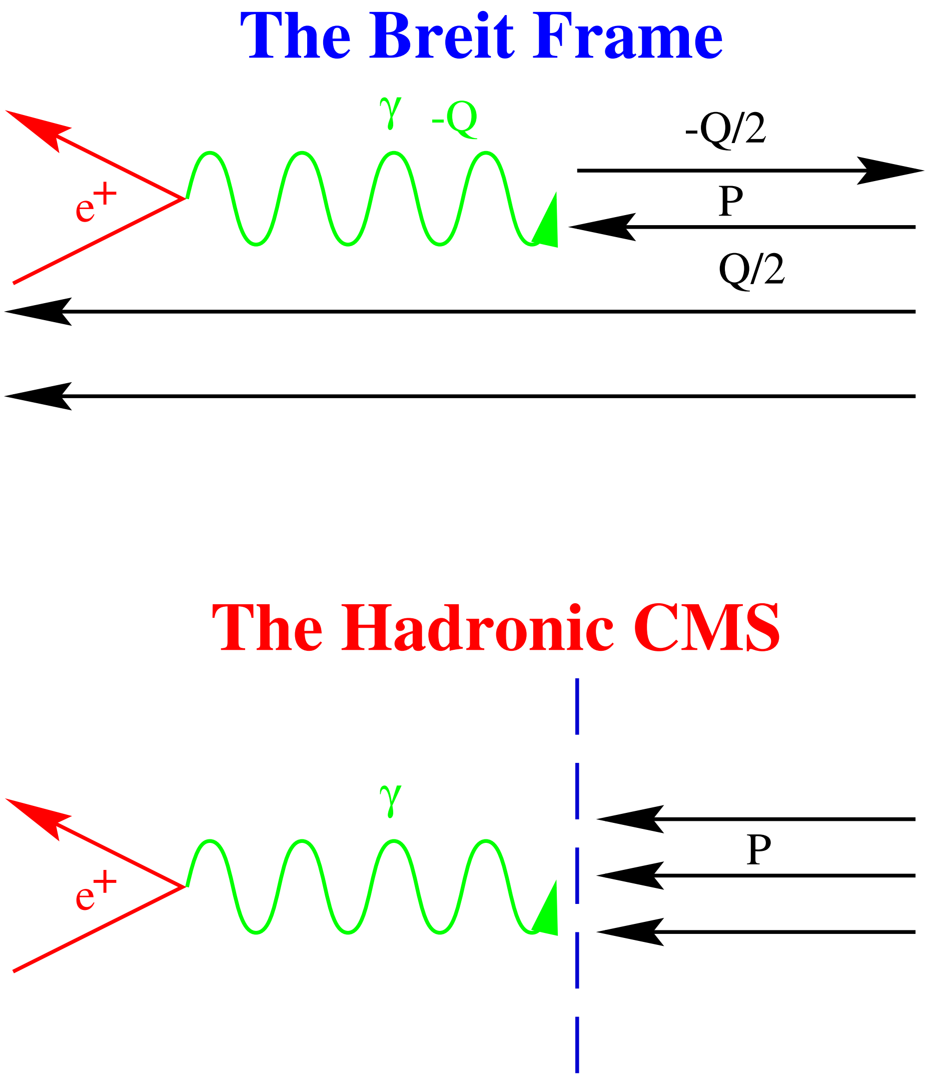

Figure 1.7: Difference between the Breit frame and the hadronic centre of mass

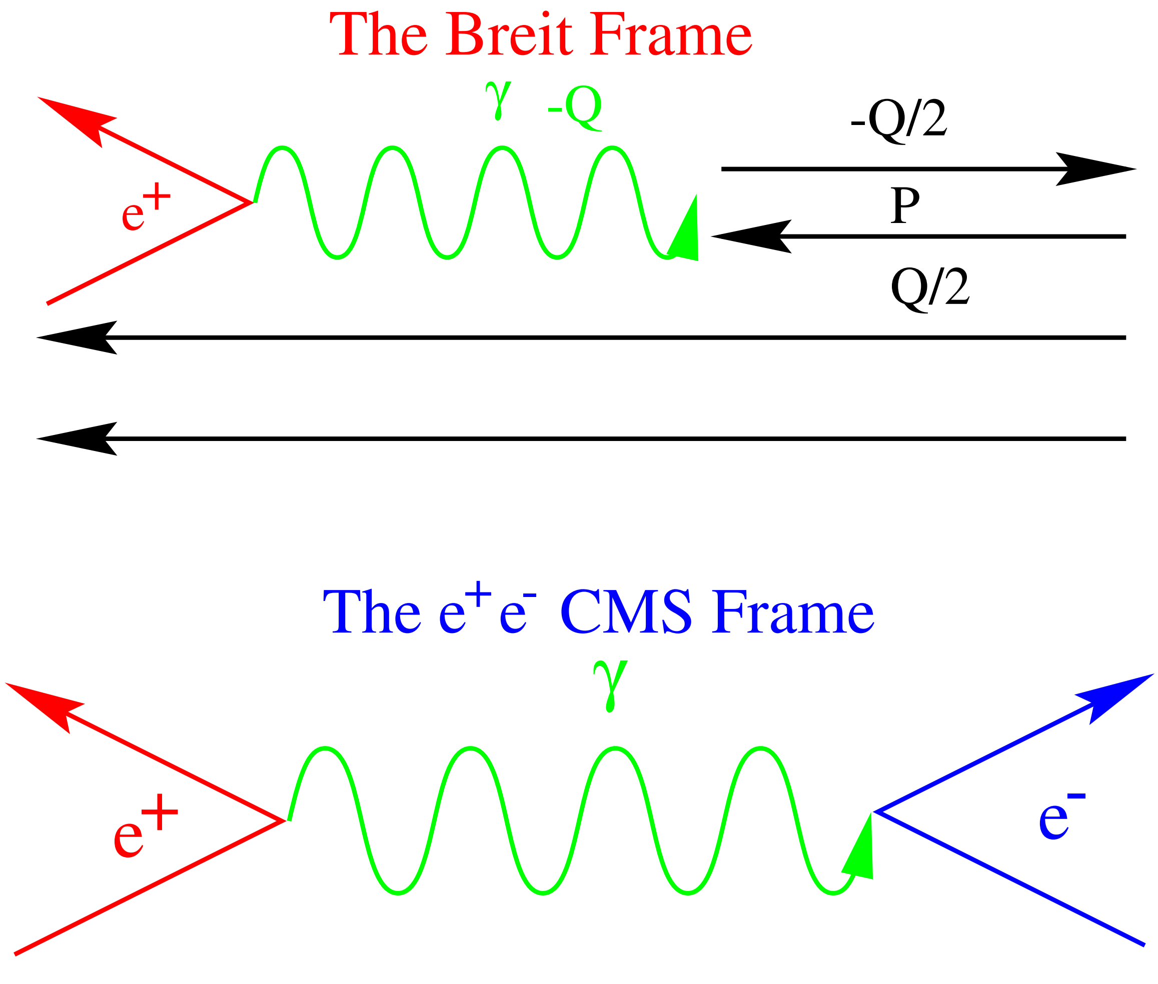

Figure 1.8: Comparison of the Breit frame and the e^+e^- centre of mass system

CMS). The remaining portion of the proton is called the "spectator" system (as it plays no part in the interaction) and it moves with much higher momentum into the target hemisphere.

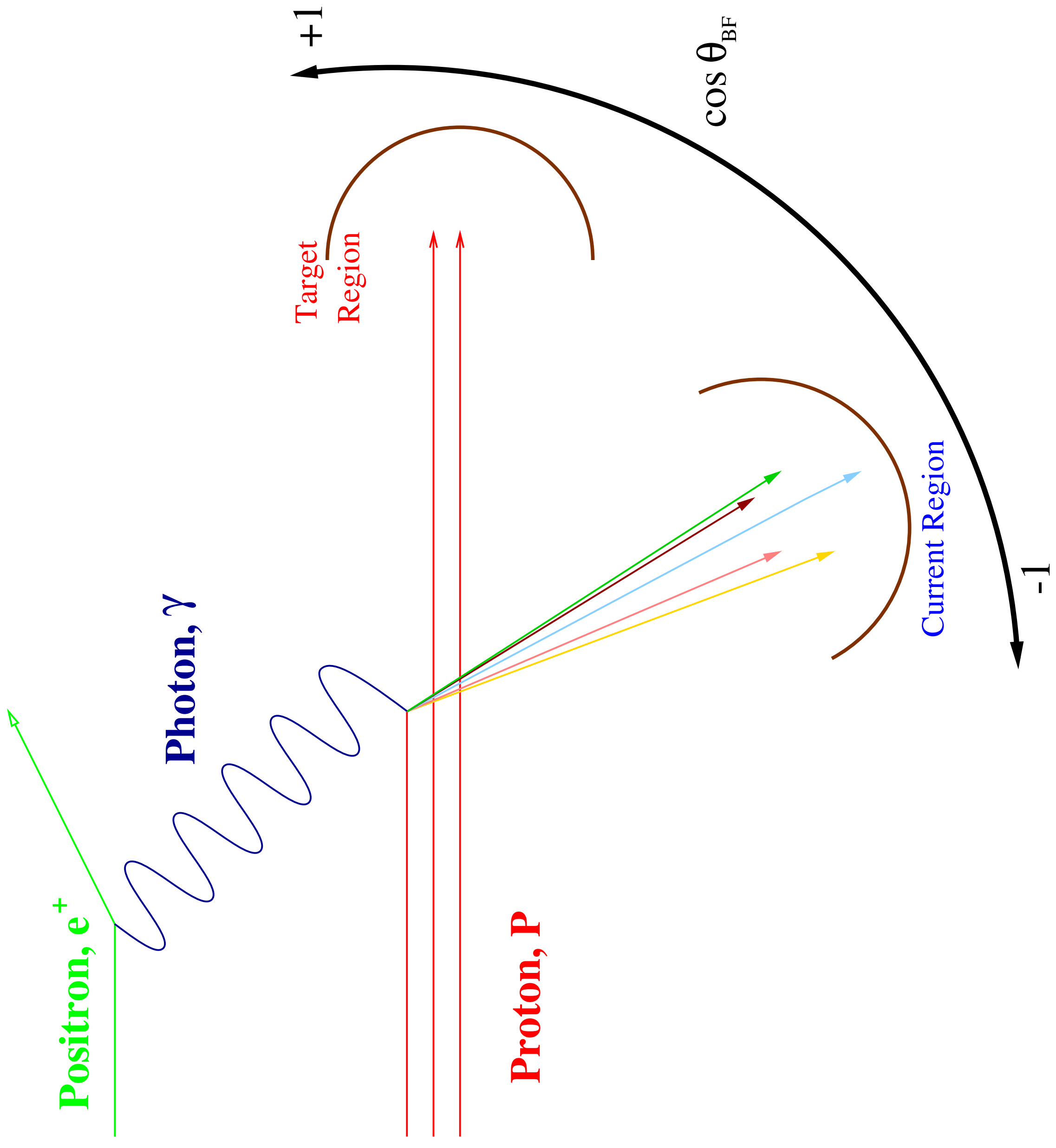

Figure 1.9: The Breit frame

Chapter 2

Experimental Setup

2.1 Detector Introduction

The H1 detector is one of four independent experiments at the HERA facility, DESY, Hamburg. Two of the experiments, H1 and ZEUS, were designed to be colliding-beam detectors. The other two experiments, HERMES and HERA-B, are effectively fixed-target experiments making use of the electron[1] and proton beams respectively. Design energies were 820 GeV and 30 GeV for the proton and electron beams respectively, though due to technical problems with the beam polarisation, the maximum electron energy was capped at 26.7 GeV in 1994 and 27.6 GeV in 1995 and 1996.

Both the proton and electron beams are stored in independent beam-pipes in HERA (Hadron Elektron Ring Anlage), being brought into co-incidence (at zero crossing angle) at the four detector sites named above. The tunnel housing the beam-pipes (the magnets in the proton one being super-conducting) is 6.3 km in circumference, and is situated approximately 30 m underneath the ground of the north-Hamburg suburbs of Lurup, Bahrenfeld and Osdorf.

[1] electrons and positrons will be referred to generically as electrons, though they will be distinguished in specific cases

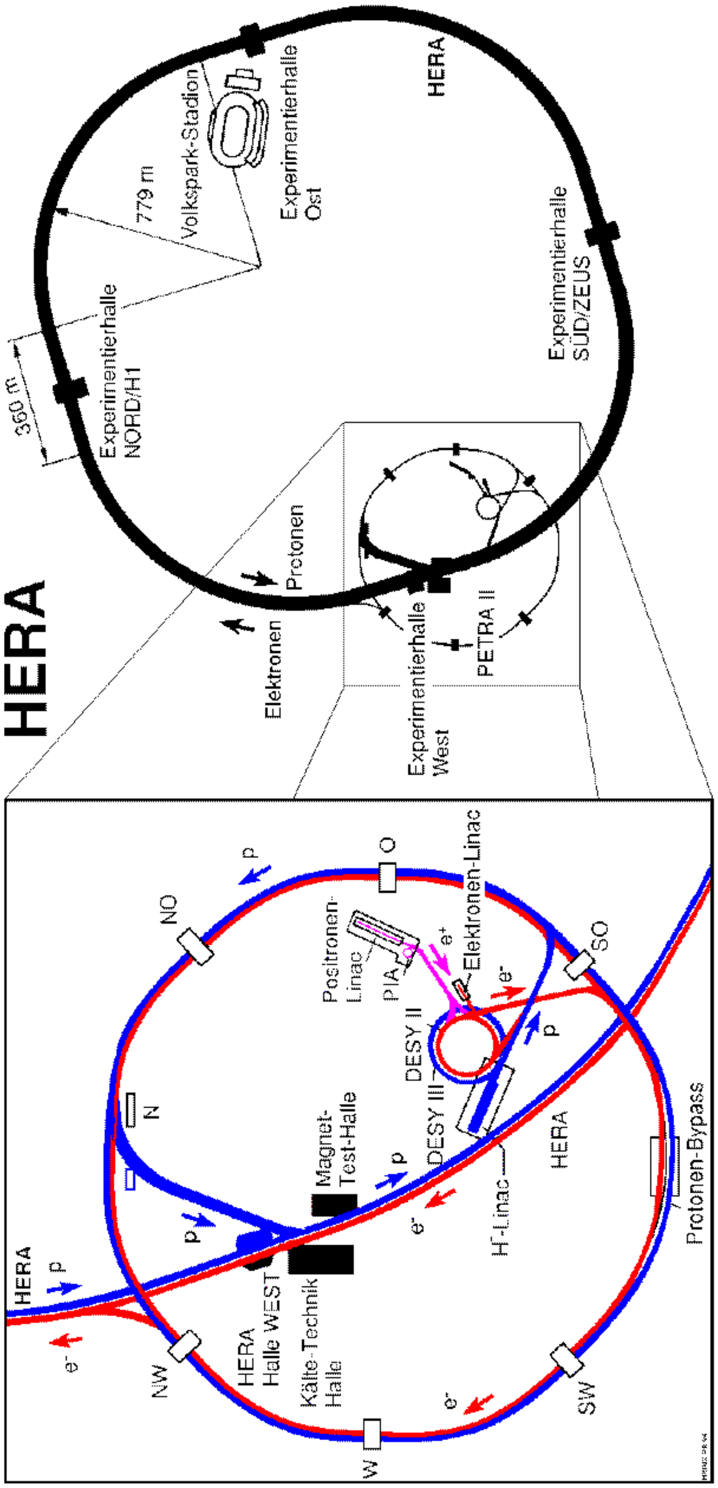

Figure 2.1: A map of the HERA facility at DESY

2.2 Filling The Machine

2.2.1 Electrons

A 500 MeV linear accelerator (linac) injects electrons into a small storage ring where, after 60 mA has been accumulated into a single bunch, they are injected into DESY II - another accelerating ring. In DESY II, they are accelerated to 7 GeV and subsequently injected into the (modified) PETRA II ring. This process is then repeated until approximately 70 bunches, $\sim$ 28.8 m apart, have been collected. The entire contents of the PETRA II ring are then injected into the main HERA ring. HERA allows room for up to 210 bunches each of protons and electrons to be stored. This is never achieved in practice however, as many "RF buckets" are required for proton/electron pilot bunches (no matching/corresponding electron/proton bucket); and also completely empty buckets. The average number of filled buckets is $\sim$ 178 giving a typical electron current of 35 mA . The pilot bunches are used for background monitoring and other calibrations, and the empty buckets are used for other calibrations, e.g., (post 1995) SpaCal LED (Energy).

2.2.2 Protons

The protons are obtained from Hydrogen by ionisation and are then fed into a 50 MeV linac which injects into DESY II where they are accelerated to 7.5 GeV , transferred to PETRA II; accelerated further to $\sim$ 40 GeV before finally being injected into HERA for final ramping to 820 GeV . The average number of proton bunches is $\sim$ 185, giving a few pilot bunches, and leaving $\sim$ 10 - 20 empty bunches for SpaCal LED calibration. The average proton current is 65 mA .

Once both beams have reached their collision energy, the HERA machine group steer the beams so that luminosity running may begin. A "Luminosity Run" is declared when HERA detects ep collisions and the background rate and synchrotron radiation is low.

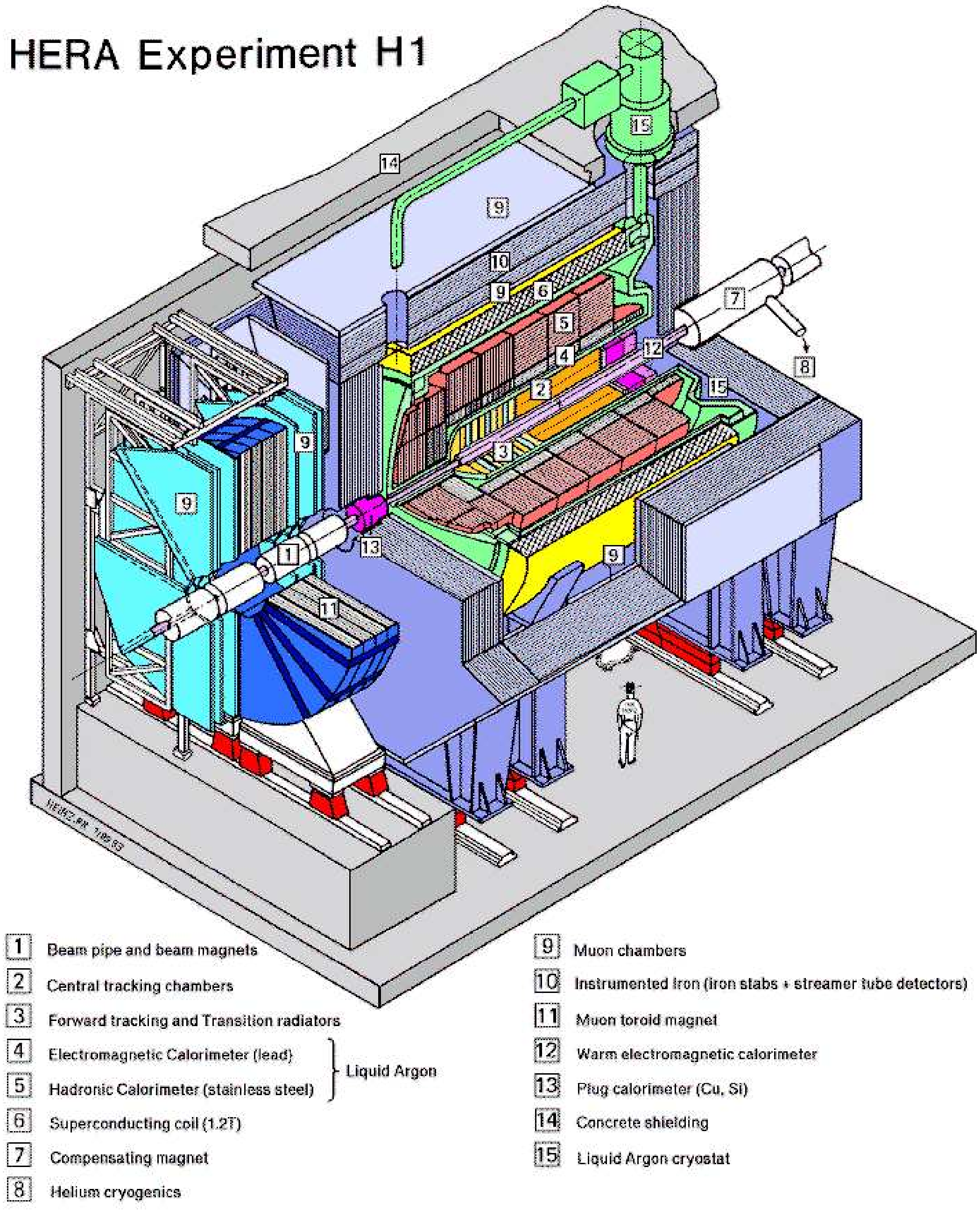

Figure 2.2: The H1 Detector

2.3 The H1 Detector

The layout of the detector is shown in figure 2.2. The co-ordinate system defined for use in H1 has its origin at the nominal interaction point (IP) and is defined such that the $+z$ direction lies along the direction of the proton beam, with the $+x$ direction pointing towards the middle of the ring and $+y$ pointing vertically upwards. The polar angle, θ, goes from $\theta = 0$ at the forward part of the detector to $\theta = \pi$ in the backward region, and the azimuthal angle, ϕ, goes from -π to $+\pi$. At the centre of the detector are the tracking chambers consisting of two cylindrical jet and z-drift chambers, with forward tracking chambers (three planar and three radial modules) in the $+z$ direction. Surrounding these tracking chambers are the superconducting solenoid, giving an uniform field of 1.15 T in which the charged particles can be tracked, and the calorimeters. There is also a 'plug' detector in the forward direction and a backward calorimeter. The outermost detectors are the muon chambers which enclose the entire inner detector. There are more detectors in the $+z$ direction due to the huge asymmetry in beam energies, and therefore all collisions having a boost along the proton ($+z$) direction.

Downstream of the interaction vertex, behind the BEMC (Incept - 1994) / SpaCal (1995 - present) lies the ToF, the time of flight device which was designed and built by QMW engineers and physicists, the veto wall, and the lumi systems. The veto wall, as its name suggests, vetoes background events out of time with the interaction, *i.e.*, beam-gas, and beam-wall events. The ToF measures the time from the interaction point to the ToF and passes or vetoes an event.

The Luminosity measuring system lies further downstream, and consists of an electron tagger at -33 m and a photon tagger at -103 m. These are necessary devices as the luminosity (used in cross-section calculations) is determined by measuring the rate of the (calculable) Bethe-Heitler *bremsstrahlung* process [10] ep $\rightarrow$ ep γ.

In 1994, the backward calorimeter and timing was done by the BEMC and the ToF. The BEMC was a lead-glass electromagnetic calorimeter, with the ToF consisting of scintillator arrays. Since then, in data taken in 1995 and onwards, the H1 detector has been upgraded and the BEMC and ToF have been replaced by the SpaCal (Spaghetti Calorimeter) which comprises of an electromagnetic *and* a hadronic calorimeter and also gives precise timing information. In 1996, the SpaCal was further improved by the addition of an insert allowing particle detection and calorimetry down to 178°. The SpaCal detector consists of lead-scintillating fibres which look like strips of spaghetti, hence the name.

The areas of the H1 detector that were vital to this analysis were the Central and Forward Trackers, described in sections 2.4.1 and 2.4.2 respectively; the Backward Electromagnetic and Liquid Argon Calorimeters, described in sections 2.5.2 and 2.5.1 respectively; the Time of Flight device, as described in section 2.10); and the H1 software chain, described in section 2.12.

2.4 Tracking

The tracking is handled by three major components: The Central Tracking Detector (sub-section 2.4.1); The Forward Tracking Detector (sub-section 2.4.2); and the Backward Proportional Drift-Chamber (sub-section 2.4.3).

The tracking system of H1 is responsible for providing simultaneous track triggering, reconstruction, and particle identification for all events. The tracking detectors were designed to reconstruct jets which had high particle densities and also to measure the momentum of the scattered isolated charged particles to a precision of $\frac{\sigma_p}{p^2} = 3 \times 10^{-3}$ GeV $^{-1}$ and their angles to a precision of $\sigma_\theta \approx 1$ mrad.

2.4.1 Central Tracking Detector

Two large concentric drift chambers form the CTD at H1. These are known as CJC1 and CJC2 which are, respectively, the inner and outer Central Jet Chambers. The chambers' sense wires are strung parallel to the beam axes (the z-direction) with each group of sense wires, called a drift cell, inclined at 30° with respect to the radial direction. CJC1 has 30 drift cells each of 24 wires, whereas CJC2 contains 60 drift cells with a complement of 32, 2.2 m long sense wires. These drift cells and wires are illustrated in Figure 2.3.

The drift cells are inclined with respect to the beam to ensure that (in the presence of the magnetic field) the electrons, ionised by a charged particle, drift in a roughly perpendicular direction to the track direction thus achieving optimum resolution. Each "stiff" track, i.e., $p_t > 170$ MeV [11], crosses the plane of the sense wire at least once in both CJC1 and 2. This, combined with the 30° inclination, resolves drift-chamber ambiguities since wrong mirror track segments in neighbouring cells just do not match up, and these track segments do not point back to the event vertex. Another advantage of such a layout for the detector is that the drift-field is uniform to

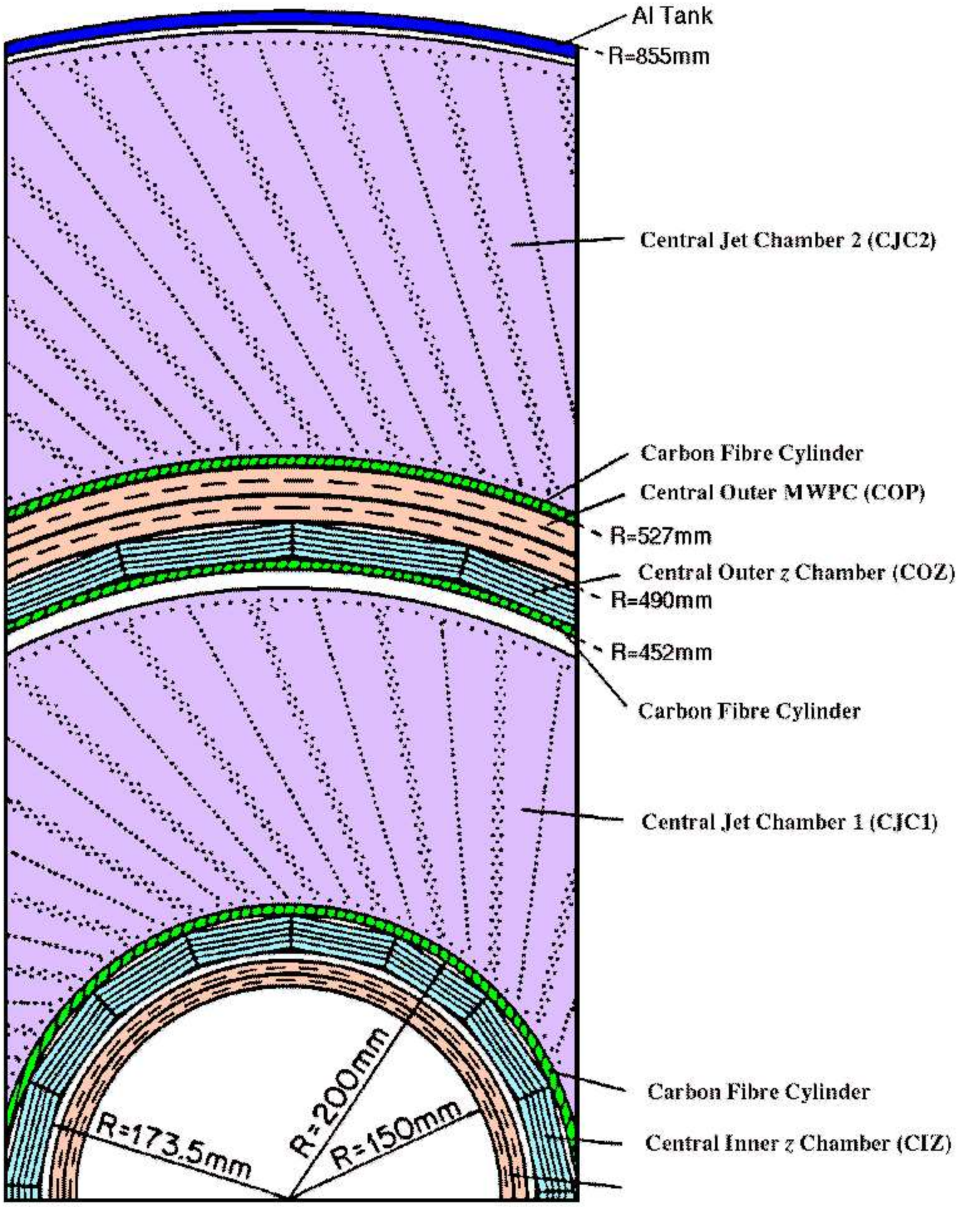

Figure 2.3: Transverse section of the CJC

better than 1 % apart from at the first and last sense wire, where the wire geometry produces local distortions of ≤ 2 % [12]. This introduces only negligible variations of the drift velocity, but there are systematic shifts in the drift time measurement. However, as all tracks cross the sense wire planes, these systematic errors cancel (approximately) on either side. The distance along the wire is measured by charge-division, and a resolution of 1 % of the wire length in z is achieved, whereas the space point resolution in the $r - \phi$ plane (the drift co-ordinates) is 170 μm.

A better determination of z can be achieved by making use of the dedicated z-chambers, CIZ and COZ, which are the Inner and Outer z-chambers, respectively. These chambers lie at the inner radii of the jet chambers and deliver track elements with typical resolutions of 300 μm in z and $8°$ in ϕ.

2.4.2 Forward Tracking Detector

The Forward Tracker is, as the name implies, in the forward region of the H1 detector. It is comprised of three identical supermodules, each of which contain four different kinds of detector. These detectors are two kinds of drift chamber - "planar" and "radial" - a multiwire proportional chamber (MWPC) and a transition radiation detector (TRD). A cross-section of the detector is shown in Figure 2.4.

The first detectors, as viewed by the incoming proton beam, are the planar chambers, of which there are four sets, each with three sub-layers of cells arranged so that the wires extend in the xy plane. Each of these three subsets are rotated with respect to the next layer by $60°$ in ϕ. No information about the hit position along the wire is available due to the fact that the wire is only read-out at one end.

The MWPCs are 20 (16+4) annular sections (around the beam pipe) which are divided into azimuthal sections of $22.5°$ for the inner 16 cathode pads, and $11.25°$ for the outer 4 cathode pads. An excellent timing resolution of 20 ns (FWHM) has been measured when all three (or at least two) modules of the MWPC registers a hit. This puts the track in the polar-angle range of $6.6° \leq \theta \leq 18.0°$ ($5.1° \leq \theta \leq 21.6°$) and allows discrimination of hits from successive bunch crossings, and as a consequence, the MWPCs (in conjunction with the central proportional chambers) form the basis of many first level (L1) triggers.

The TRDs were constructed to discriminate between electrons and pions using the method of dE/dx (in conjunction with the radial chambers). The TRDs are made

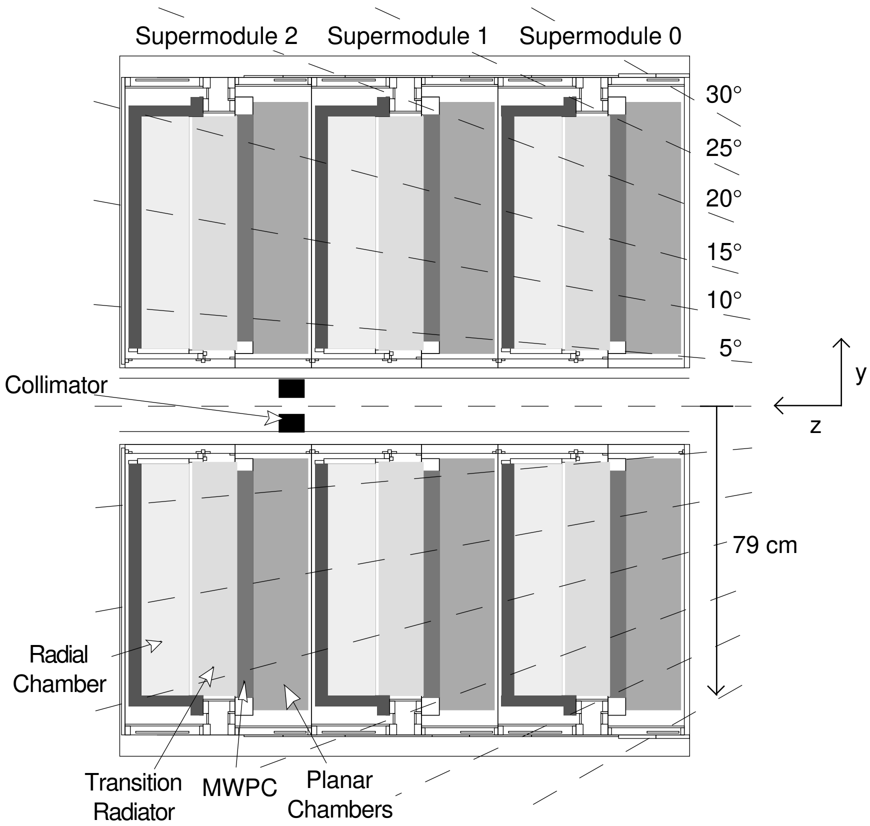

Figure 2.4: Longitudinal cross-section through the Forward Tracker

from 400 polypropylene sheets, are self-supporting and enclosed in their own gas envelope as their gases are different from those used in the rest of the FTD. The polypropylene layers provide the necessary dielectric interfaces for X-ray (or TR) emission. The energy and yield of the X-rays are influenced by many factors, such as dielectric constant and layer thickness, and the values chosen for the FTD were such that the X-ray spectrum peaks at around 6 keV for 20 GeV electrons. Using this design for the TR and also the radial chambers, the interface between them having been optimised for TR transmission, a discrimination between electrons and pions at the level of 90 % electron acceptance with less than 10 % pion contamination is possible, up to track energies of 80 GeV passing through all three modules of the FTD.

2.4.3 Backward Proportional Drift Chamber

The BPC is situated between the CTD and the BEMC (see subsection 2.5.2) and was designed with the express purpose of providing an accurate spatial resolution of the scattered electron in low Q^2 DIS (see section 1.2) events. The BPC has an angular resolution in θ of 0.5 mrad due mainly to its construction of four wire layers strung in the $x - y$ plane, each of which is rotated by 45° with respect to its neighbour. The detection efficiency is 98 %.

2.5 Calorimetry

Tracking detectors may provide information about the momentum of the particle, but provide no information when a neutral particle traverses their volume; this information is provided by the many calorimeters within H1. Together, the trackers and calorimeters provide complementary and complete knowledge about the trajectory, momentum, and energy of all particles - even the neutral particles as they too deposit energy in the calorimeter! Most of the calorimeters at H1 sit within the magnetic field produced by the solenoid. This helps to minimise the dead material between the interaction vertex and the sampling calorimeters. The calorimeters at H1 are the Liquid Argon Calorimeter (LAr), the Plug Calorimeter, the Tail-catcher (or Instrumented Iron) and the Backward electromagnetic calorimeter (BEMC), which was superseded in 1995 by the SpaCal, (Spaghetti Calorimeter) which is an electromagnetic and hadronic calorimeter. Figure 2.5 shows a schematic of the H1 detector with all of the calorimeters (and other detectors) labelled and it provides an alternative diagram to figure 2.2

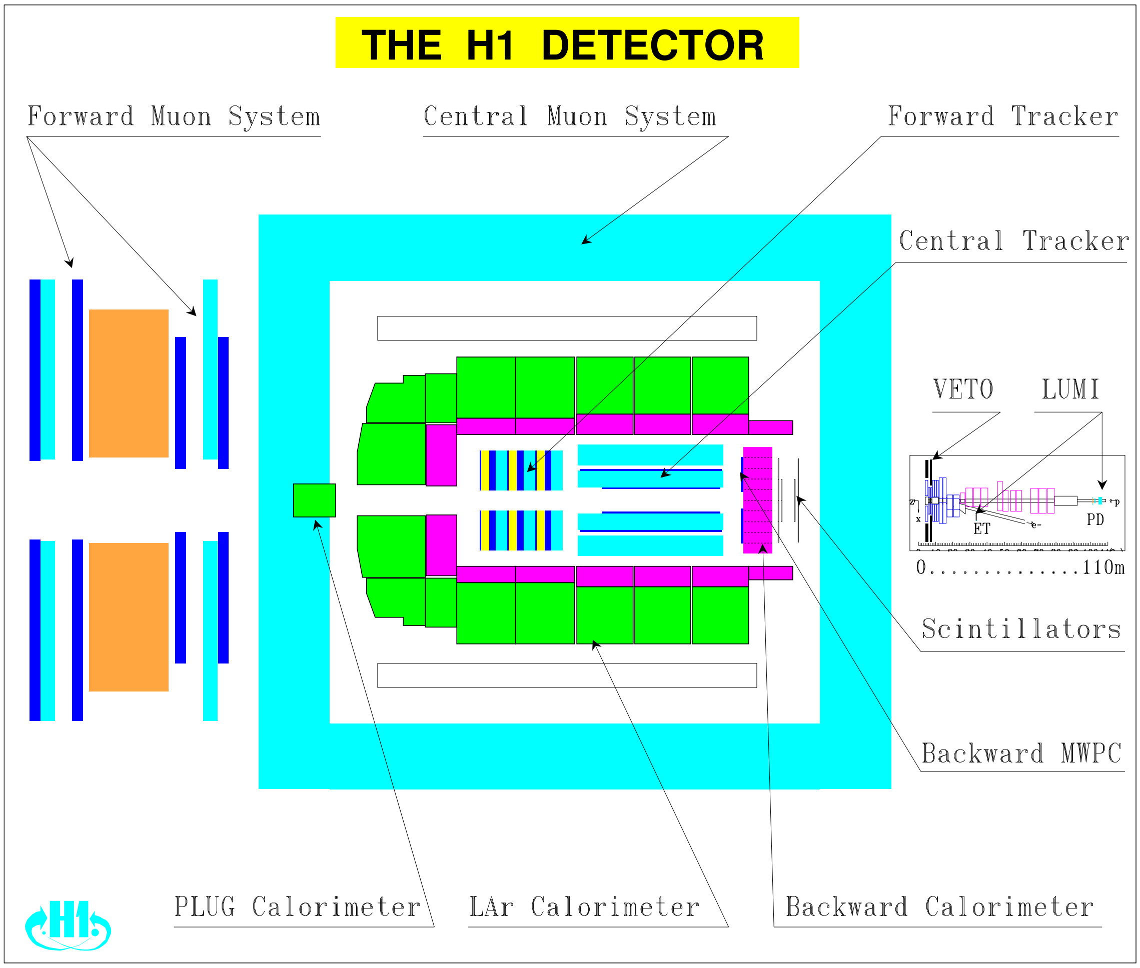

Figure 2.5: A schematic of the H1 detector illustrating the major detector components, providing a complementary picture to figure 2.2.

The LAr calorimeter covers the polar angular region $4^\circ \leq \theta \leq\sim 153^\circ$, and is complemented in the forward region by the plug extending from the beam-pipe hole to the LAr calorimeter, or $\sim 0.7^\circ \leq \theta \leq 4^\circ$. Complementing the LAr calorimeter in the backward region is BEMC, extending from $151^\circ \leq \theta \leq 177^\circ$. The tail-catcher is used to provide a rough calorimetric measurement of the particles leaking out of the LAr, and is based on the analog readout of the pads of the limited streamer tubes that constitute the "instrumented" iron yoke.

2.5.1 The Liquid Argon Calorimeter

The Liquid Argon Calorimeter is the largest of the H1 calorimeters. The primary motivations and reasons for choosing a LAr system were the ease with which it could be calibrated, a fine granularity (useful for off-line e/π separation and energy flow measurement), and also its homogeneity of response.

The LAr calorimeter measures the scattered lepton in "High" Q^2 NC DIS events, as well as the hadronic final state. The detector is effectively double-layered, the closest to the beam pipe being the electromagnetic section, outside of, or around which is the hadronic section. The calorimeter is divided further along the z-axis into eight "wheels", each of which is subdivided into eight components (or octants) in ϕ.

The electromagnetic section has lead absorber plates whilst the hadronic section's are stainless steel. The absorption length varies from 20-30 radiation lengths in the former to 4.5-8 interaction lengths in the latter. The sampling medium in both cases is LAr. "Sampling Calorimeters" are those which have alternate layers of passive absorber followed by active instrumented regions. They work by virtue of the fact that incident particles lose energy by inelastic collisions in the absorbing material. These inelastic collisions produce a shower of secondary particles, the number produced being proportional to the energy of the initiating particle. The secondary particles are sampled in the active region as the shower develops. There are differences in the showering mechanism between electromagnetic and hadronic interactions; Electromagnetic showers (usually from electrons) consist of photons (γ) and electrons from *bremsstrahlung* and pair-production. Hadronic showers are due to inelastic nuclear collisions and exhibit a larger extent both in the transverse and longitudinal directions than electromagnetic showers.

The LAr calorimeter is "non-compensating", which means that its response to hadrons is about 30 % smaller than that for electrons of the same energy. This is a known feature, and is routinely corrected for in the off-line reconstruction (see 2.12) of the events. Test beam studies (at the CERN SPS) were used for determination of the energy resolution, and they were found to lie in the range $0.10 \oplus 0.01 \leq \frac{\sigma(E)}{E} \leq \frac{0.13}{\sqrt{E(GeV\ )}} \oplus\ 0.01$ for the electromagnetic portion, and $\sim \frac{0.5}{\sqrt{E(GeV\ )}} \oplus\ 0.01$ for the hadronic section.

The energy scale is known to within 3 % for the electromagnetic calorimeter and to within 6 % for the hadronic part. The energy scale is determined differently for the two sections. The electromagnetic section compares the scattered lepton energy in the LAr and the CJC and measurements of transverse momentum balance in NC DIS are used for the hadronic section.

2.5.2 BEMC

As the acronym suggests, the BEMC is a purely electromagnetic calorimeter in the backward region of H1. The requirements of the BEMC are to trigger on, and measure the energy of the scattered lepton in typical low Q^2 DIS events, where "low" is taken to mean $2 \leq Q^2 \leq 100$ GeV^2 . This Q^2 range is determined by the BEMC's angular coverage, *i.e.*, $151\,^\circ < \theta < 177\,^\circ$ (see chapter 1). Contrary to its name, the BEMC also measures hadronic activity from low- (to medium-) x, high-y final states, e.g., hadronic jets from photoproduction. Such events dominate the total cross-section.

Another feature contrary to its name is the ability to provide not only calorimetric information but also directional information. The BEMC can do this due to its fine granularity, which comes from its construction from 88 segments (in stacks) mounted in an aluminium barrel, each of which is composed of 50 Pb-scintillating layers 22.5 radiation lengths in depth (or 1 hadronic interaction length).

Test beam studies at the CERN SPS facility were used for calibration, and the average noise per stack was measured to be 150 MeV with an energy resolution due to sampling fluctuations of 10 %$/\sqrt{E}$ GeV . There is also another factor degrading the resolution, and that is the intercalibration between each stack which is estimated to be 4 % [6]. Thus, the energy resolution can be expressed as

$$\frac{\sigma_E}{E} = \frac{\sigma_{noise}}{E} \oplus \frac{\sigma_{sampling}}{\sqrt{(E)}} \oplus \sigma_{constant} \tag{2.1}$$

where $\sigma_{noise} = 0.45$, $\sigma_{sampling} = 0.1$, and $\sigma_{constant} = 0.04$.

There is yet another factor to the resolution, and that is an uncertainty in the calibration, which is estimated to be 2 % [6]. This figure is derived by comparing the energy seen to that calculated from the angles of the electron and the hadronic final state's momentum vector.

The hadronic response is not so well-measured due, in no small part, to the fact that

~ 30 % of the total hadronic flow into the BEMC produce signals below threshold, *i.e.*, just escapes detection. This is because of the low interaction length of the device. Of those hadrons that do interact, typically 30 % of their energy is deposited in the BEMC. A resolution[2] of $\sim 80\ \%/\sqrt{E}$ can be achieved only by complementing the measurement with information from the instrumented iron.

2.5.3 The Plug Calorimeter

The angular range in the very forward part of the detector $0.6^\circ \leq \theta \leq\sim 3^\circ$ is instrumented by the Plug calorimeter, which "plugs" the gap between the end of the LAr acceptance and the beam pipe. The Plug is used mainly to contain the hadronic final state particles in charged current events that would normally disappear down the beam-pipe at low angle. It is built from Copper and Silicon, *i.e.*, 9 layers of Cu absorber and between each pair is sandwiched a layer of Si (making 8 in total), the total depth of the whole thing is approximately 4.3 interaction lengths (λ_I).

The Plug sits in an area of the detector with very high energy-density, and consequently suffers from radiation damage. Its energy resolution is only a poor $150\ \%/\sqrt{E}$, however, this is adequate for the physics needs of H1. Plug data are not used in this analysis.

2.6 The Tailcatcher or Instrumented Iron

Energy that leaks from the LAr and BEMC is "caught" and measured by the hadronic tailcatcher. The tailcatcher's elements, streamer tubes, are sandwiched between layers of iron in the return yoke of the magnet. Readout pads are also placed in the gaps between the layers of iron. The orientation of the tubes and pads are along the beam-axis in the barrel region, and perpendicular to the beam-axis in the forward and backward regions. The instrumented iron covers $5^\circ \leq \theta \leq 175^\circ$ and has a depth, λ_I of ~ 4.5 . The hadronic energy resolution, from experimentally-verified Monte Carlo simulation, is, as with the Plug, $150\ \%/\sqrt{E}$. The Central Muon Detectors (see figure 2.5) are also contained within the iron.

[2]obtained from simulation studies

2.7 Forward Muon Spectrometer

Energetic muons that survive the passage through the iron (in the forward region), *i.e.*, $5 \leq P_\mu \leq 200$ GeV/c , can be detected using the forward muon spectrometer. The angular range covered by the Forward Muon Spectrometer is $3° \leq \theta \leq 17°$. The detector consists of three double-layered drift chambers arranged in octants (one ϕ octant sandwiched between two θ octants) which are mounted on an aluminium frame on each side of a toroidal magnet. The Forward Muon Spectrometer lies outside the field of the H1 main solenoid and is consequently unaffected by its axial field.

The lower momentum value is that which guarantees that a muon will penetrate through the iron to the Forward Muon Spectrometer and depends also on the momentum resolution, whereas the upper limit is governed by the strength of the forward muon toroid's field and the spatial resolution of the drift chambers. The Forward Muon Spectrometer is illustrated in figure 2.6.

The Forward Muon Spectrometer is also used to detect particles emanating from secondary interactions with the collimators inside the beam-pipe. The collimators help to protect the experiment from synchrotron radiation.

2.8 The Magnet

The magnet surrounds the H1 calorimeter and provides an uniform, axial magnetic field of 1.15 T. This it achieves by using a superconducting solenoid (made of Nb-Ti and Cu) which operates at a nominal current of 5517 A, as well as by use of an iron return yoke. The area covered by the solenoid is 3.6 m long and has a radius of 1.6 m, and it is centred about the nominal vertex.

To prevent the H1 magnet from adversely affecting the conditions of the HERA beams, and to preserve $e^{\pm}$ polarisation, H1 also has a compensating solenoid (870 A) at $z = -4.4$ m. This also has the effect of minimising defects in the beam quality due to misalignment of the main solenoid.

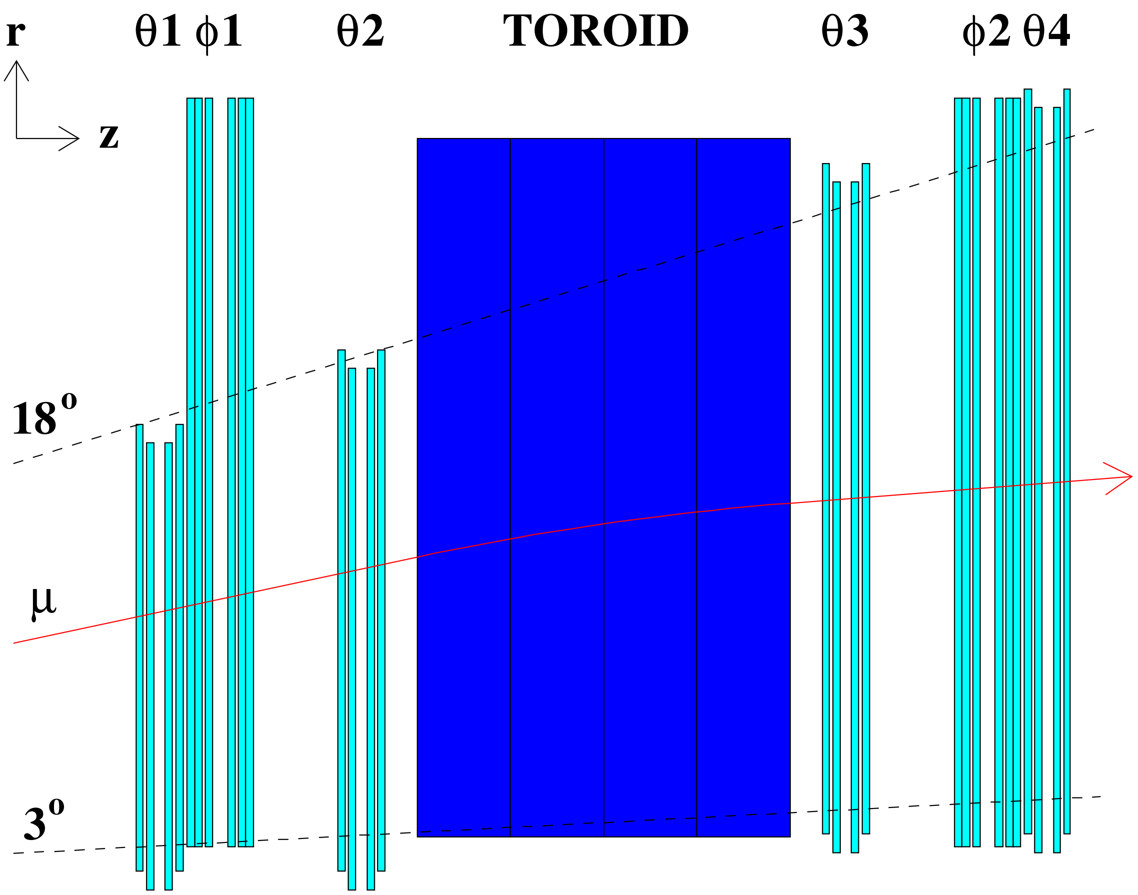

Figure 2.6: A cross-section through the forward muon spectrometer. The solid line is the trajectory of a muon which has traversed the toroid. The acceptance of the detector is shown by the dotted lines.

2.9 The Luminosity System

The luminosity at H1 is calculated from the rate of Bethe-Heitler events, where $ep \rightarrow ep\,\gamma$. These events have a large and precisely calculable cross-section [10]. Unfortunately, beam-gas and beam-pipe interactions, $eA \rightarrow eA\,\gamma$, have an almost identical signature and a larger cross-section. The background can be measured experimentally by use of the pilot bunches in both the electron and proton beams. The luminosity is calculated as follows:-

$$\mathcal{L} = \frac{R_{tot} - (\frac{I_{tot}}{I_0})R_0}{\sigma_{vis}} \tag{2.2}$$

where R_{tot} is the total rate of the bremsstrahlung events, R_0 is the rate in the electron pilot bunches, I_{tot} and I_0 are the corresponding electron beam currents and σ_{vis} is the visible part of the $ep \rightarrow ep\,\gamma$ cross-section (once the trigger efficiency and detector acceptance have been accounted for).

The luminosity (lumi) monitor consists of two main parts; namely the electron tagger, ET, at $z = -33.4$ m, and the photon detector, PD, at $z = -102.9$ m. The lumi monitor detects both the $e^{\pm}$ and the photon in coincidence using these two arms. This is shown in figure 2.7. The reason that the detectors are so far away from the interaction vertex is that the distribution for both the electrons and photons are strongly peaked in the direction of the $e^{\pm}$ beam. At ~ 30 GeV the polar angles are of the order $\theta \sim \mathcal{O}(m/E) \sim 17\ \mu$ radians.

Scattered electrons are deflected by a series of magnets, between $z = -5.8$ m and $z = -23.8$ m, and exit the beam-pipe (at their exit window) at $z = -27.3$ m so that they may strike the ET. The bremsstrahlung photons, of course, are not deflected by the magnets and exit through their window at $z = -92.3$ m (where the beam-pipe bends upwards), pass through a lead filter and Čerenkov counter, then finally into the PD.

2.10 ToF & The Veto Wall

The Time of Flight device, ToF, is composed of two walls of scintillating plastic which is sandwiched between lead plates. The ToF scintillators (labelled Scintillators in figure 2.5) provide a fast veto on background events (induced by the proton beam)

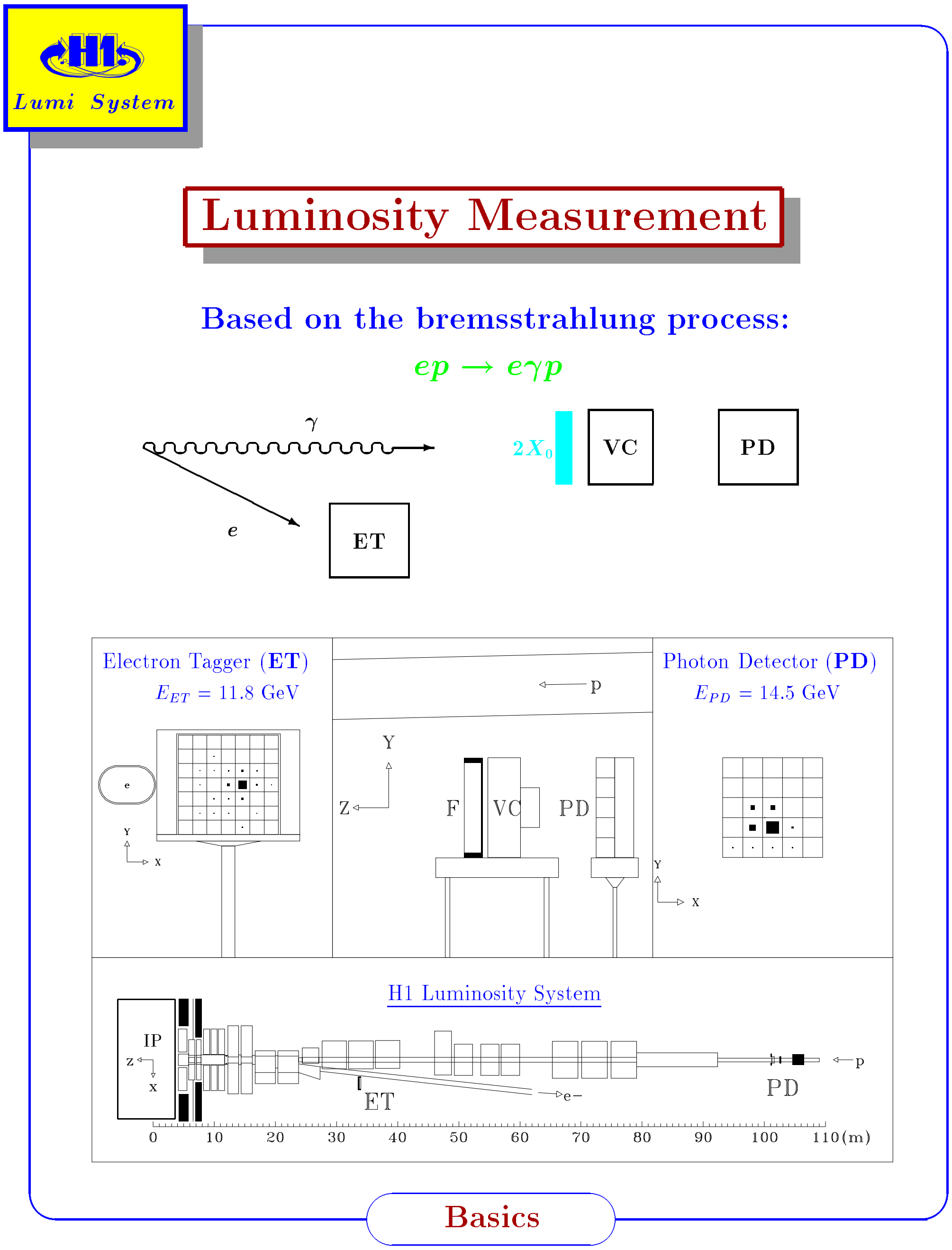

Figure 2.7: The luminosity measuring system

coming from interactions before the nominal interaction vertex. The lead is there to shield the scintillator from high synchrotron radiation flux. The ToF is situated at approximately 2 m behind the BEMC, which allows a separation between physics and background of $\sim$ 13 ns compared to a resolution in the ToF of $\sim$ 4 ns. The level 1 trigger makes use of the ToF to remove background events, and reduces the trigger rates by 99 % giving a massive improvement in both the efficiency and the minimisation of dead-time.

Two further scintillators, known as the Veto Walls (labelled VETO in figure 2.5), are located at $z = -6.5$ m and $z = -8.1$ m respectively and help to further suppress background events (mainly from halo muons).

2.11 Triggering and Data Acquisition (DAQ)

The task of collecting data from the H1 detector is not a trivial one, due mainly to the exceedingly short time (96 ns) between bunch crossings but also compounded by the complexity of the very different parts of the detector. The bunch-crossing time is considerably shorter than the time required to read out all of the individual branches of the detector components ($\sim$ 1 ms), and therefore, to capitalise on all of the available luminosity, some triggering system is required which will only initialise a complete read-out when there is a probability that the last collision seen was *ep* and not some other interaction. It is also necessary to store data coming from the detector until it is needed, and this is achieved by use of “pipe-lines”. These pipe-lines are merely front-end digitising units which store the output data passed to them where, at some later time when the data are required, the signals may be read-off.

A complete dump of all the raw data in H1 involves obtaining the contents of over a quarter of a million analogue readout channels. Each of these triggered events has an approximate size of 3 Mb. The event size and event rate that can be handled safely by “The Farm” (see section 2.11.5) are 50-100 Kb and 50 Hz respectively, consequently a great deal of data reduction and compression is necessary in order to decrease both the event size (from $\sim$ 3 Mb to $\sim$ 100 kb) and rate (from several MegaHertz to $\sim$ 50 Hz). The method employed by H1 is a combination of trigger selections operating in series and will be described in the next sections, 2.11.1, 2.11.2, 2.11.3, 2.11.4, 2.11.5 and 2.12.

2.11.1 The H1 Trigger

There are five nominal levels of H1 trigger, L1 through to L5. The higher the number, the more sophisticated (and usually software-based) are the selection criteria. Hand in hand with this sophistication goes the time it takes for a decision to be made, ranging from 1 μs at L1 to $\mathcal{O}$(ms) for L4 (and even longer for the off-line L5).

2.11.2 Level 1

A custom-built hardware system forms the L1 trigger. It obtains data from nine different trigger subsystems (obtained from the subdetectors of H1) and the output of each is called a trigger element, TE. These TEs are then fed to the main Central Trigger Logic, CTL, where some logical-OR decisions are made to combine them into subtriggers. Any and each of these subtriggers is sufficient to initiate an event readout, which is achieved by sending an L1KEEP signal in order to stop the pipelines and thus begin accruing deadtime. The nine branches that comprise the L1 trigger are Calorimeter, CTD, FTD, Forward Muon, MWPCs, Muon (Iron), Luminosity, Forward Muon Trigger, Subsystem Triggers, e.g., DC R-ϕ. An L1KEEP decision takes 2 μs to make, equivalent to 22 bunch crossings or HERA clock cycles.

DIS *ep* reaction selections are based primarily on calorimetric triggers. Such events are characterised by large transverse energy, $\mathrm{E_t}$, deposits stemming from jets and the primary scattered electron. Similar event signatures are also expected from "Beyond the Standard Model" physics. The trigger observables seen in the calorimeter cover a huge spectrum, from the narrow, localised energy deposits of the electron to the large global energy sums of $\mathrm{E_t}$ or missing energy, and the trigger deals with all of these conditions.

The MWPCs provide information such as wire hit-maps and photo-tube pulse heights extremely rapidly, thus are an ideal choice for the L1 trigger. The sampling of the data is performed by FADCs operating at 104Mhz, or 10 times that of the HERA clock frequency. This excellent time resolution affords a very precise determination of the event time, $\mathrm{t_0}$, from the correct bunch crossing identification, thereby allowing a good determination of the position of the vertex. The vertex position measurement is complemented by information from the central z-chamber, used to provide a vertex z-position. Vertices which lie outside the nominal interaction region are clearly background events of one kind or another, and thus can be rejected safely.

Different kinds of background can be vetoed in many ways, but those applied at L1 need to be fast decisions. Some decisions are derived from vertex position measurement as mentioned above, others from the ToF system, and yet more from the central tracker. Tracks can usually be traced back to their origin with very little difficulty, and this geometric origin of the event is the main constraint which is used to suppress background. The central tracker can be used to eliminate beam-wall and synchrotron radiation backgrounds due to its ability to calculate the distance of closest approach (DCA) from the nominal beam axis. If the DCA is less than 2 cm, then the event is kept, otherwise the ambiguities in the spatial resolution become non-trivial and events which do not originate at the IP will contaminate the sample.

The ToF and Veto Wall systems remove background originating from up-stream (backwards proton direction) interactions. Two timing windows are used, based on the HERA clock, called the ToF_IA and ToF_BG for Interaction and Background respectively. Hits in these windows decide whether the event contains physics information (ToF_IA) or not (ToF_BG), and this fact decides whether the event is subsequently kept or rejected, respectively.

Once all this information has been provided to the CTL by the various sub-detectors in the binary form of trigger decisions of YES or NO, they are combined and synchronised to the correct HERA clock pulse. This is a non-trivial task, as different signals take a different amount of time to get to the CTL due to different cable lengths and signal delays, but the HERA clock cycle is encoded in the information sent to the CTL, which it then uses to re-synchronise all the decisions.

If, after 2 μs, the L1 trigger is satisfied, the readout is enabled. This involves freezing the pipelines of data, and stopping H1 from taking any more until the pipelines have been re-enabled again. It is at this point that H1 accrues deadtime.

2.11.3 Level 2

A more advanced decision is made by the intermediate level triggers, L2 and L3. Both of these operate within the dead-time of the experiment, and are called *synchronous* in H1 terminology. The L2 is similar to the L1 in that it is hardware based, but makes more complicated decisions involving more factors and combinations of signals before deciding whether to keep or reject the event. For example, the CJC trigger elements are dealt with in a more sophisticated manner dealing with complex pattern matching and the determination of the momentum of the tracks. The event topology is also considered at L2. All of this takes about 20 μs after the interaction has taken place. If, and only if, L2 issues an L2KEEP does the readout of the more time-consuming parts of the detector take place. These include the zero-suppression of the drift chamber signals and the analogue to digital conversion of the calorimeter signals and the calorimeter DSP processing.

2.11.4 Level 3

L3, however, is software based, and performs yet more complicated tasks in order to arrive at a decision. The typical time for an L3 decision is about 800 μs. If, however, L3 decides to reject an event, then the readout is aborted and the experiment becomes alive again within a few μs. Events which survive both L2 and L3 are then passed to the Central DAQ to be managed. The maximum rate which the CDAQ can safely handle, within the bounds of dead-time, is 50 Hz. Therefore, provided the L3 accept is less than this, no further dead-time will be accumulated by the CDAQ.

Neither L2 nor L3 were operated during 1994.

2.11.5 Level 4 Filter Farm

The CDAQ system incorporates the L4 filter farm as an asynchronous software trigger. The L4 has access to the raw data of the entire event upon which to base a decision. In principle, this allows the possibility of an online trigger selection including full knowledge of the intrinsic detector resolution. The L4 farm comprises of 32 MIPS boards (later supplanted by PowerPC boards), each of which processes an event to completion before a decision is reached. The L4 modules use one or the other of fast algorithms written especially for the farm or parts of the standard offline reconstruction program. In order to accelerate the decision-making speed, software modules are only run if their output is required to arrive at a decision. Rejection techniques used are based mainly on getting a more precise interaction vertex position, using the 3D reconstructed vertex finder which, by this stage, is available. Use of this further aids the suppression of proton-induced background, whereas another algorithm is used for synchrotron radiation.

The most important thing to remember about H1 is that, at all stages, it is an asynchronous experiment, and that different branches, read out at different times, take varying amounts of time to be collected. This is also true at the L4 stage, where events are processed as and when the MIPS boards collect them in true parallel fashion. To overcome this feature, the event number is always kept as part of the event record to ensure that this asynchronous behaviour does not impair any future analysis. The final work that the CDAQ must do is to transmit the information to the data storage facilities at the DESY main site. There is a maximum logging rate, below which no increase in primary dead time will occur, but in order to achieve a logging rate lower than this critical rate, the L4 must reject somewhere between 70 to 80 % of input events.

2.12 H1 Software

Once events have successfully passed all of the trigger criteria, they are fully reconstructed by a programme called H1REC [13] and the output written to a "Production Output Tape", POT. These POT data (still) contain a huge fraction of background events which passed the trigger selections, and thus, further refinement to the cuts must now be imposed. These selection refinements make use of algorithms that look for characteristics of associated physics events which background events do not have. Roughly 1 % of non-physics events are "monitor triggers" which are used to check trigger consistencies among other things. Events which pass the characteristic tests of a physics class are assigned a physics *class number* and compressed onto a Data Summary Tape, or DST. This stage is referred to as Level 5, or L5.

H1REC is a programme which takes, e.g., raw track hits (as digitisations on wires) and energy deposits in calorimeters etc. and "reconstructs" the track that caused such a pattern of hits. Once this reconstruction has taken place, we now have access to information about the energy of the track, as well as the information about its curvature provided by the trackers.

For simplicity, convenience and also a uniform, consistent treatment of ***all*** data, *i.e.*, real data and Monte Carlo simulated data, H1REC can be run in exactly the same fashion on both.

The final link in the chain of H1 software is the package H1PHAN [14]. It uses the data obtained from H1REC (optionally with H1SIM [15] beforehand) on either

POT or DST data and returns meaningful physics quantities, e.g., four-vectors. H1PHAN can determine the event kinematics, perform basic particle identification, and reconstruct jets. The end user must provide the necessary code to do the actual physics analysis, but this is written within the framework of H1PHAN which provides an interface between the user and the cryptic commands necessary to access "BOS" banks [16] where the information is actually stored. H1PHAN's calls can also be used to access information about generated Monte Carlo (chapter 3) in a manner very similar to dealing with data.

Chapter 3

Monte Carlo Simulations

3.1 What is a Monte Carlo?

A Monte Carlo is a software tool which tries to simulate real physics by solving the equations governing these processes. Analytical solutions are, in practice, impossible and the multidimensional integration is best done by using random numbers coupled with the known laws of physics to simulate genuine data. The output, called the generated Monte Carlo, can then be run through a chain of software which makes it look like data gathered from the detector, and then reconstructs it like data whereupon it is called reconstructed Monte Carlo.

The data naturally come in digitised form, but the Monte Carlo data are generated in terms of 4-vectors and cannot be treated in the same way as real data. H1SIM [15] achieves this. This program is based on GEANT [17], a CERN detector simulation program, which provides an accurate picture of what the detector actually looks like to a particle which is traversing its volume. The whole detector, including its geometry, different materials, layers and supporting structures can all be simulated by H1SIM, and the effective output is the digitised hits the simulated track would have left on the detector had it actually gone through it. The program tracks the "particle" through the detector volume and scatters it randomly but according to known probabilistic physical laws when it encounters material. It can also shower it in the calorimeter, or make it lose energy in the instrumented iron. Once H1SIM has finished its calculations, the generated Monte Carlo now looks like "raw data" and, now that the generated Monte Carlo has these simulated track banks associated with it, it can be processed in exactly the same way as the real data, and reconstructed by H1REC (see chapter 2.12).

Monte Carlo, or simulated data is used in two ways. It provides a valuable tool for studies of detector efficiencies and acceptances for which the data can be corrected, and also to test results against the predictions of a particular theory.

Reconstructed data and generated Monte Carlo are not compared to one another because of the fact that reconstructed data incorporates detector inefficiencies and limited acceptances within it. For a true comparison to be made, the detector must be simulated in software so that its response to the passage of a particle through it may be modelled. The output of this process mimics real data and can then be reconstructed in a manner analogous to data. This reconstructed Monte Carlo can now be compared to reconstructed data. If a good comparison is found, the ratio of the generated Monte Carlo to the reconstructed Monte Carlo can be formed for a given quantity which is a parameterisation of the detector's response for that quantity. This "function" can be used to correct the data for all the detector's inefficiencies and limited acceptances (for that given quantity), in effect correcting it back to a detector-less form which can be compared to the generated Monte Carlo. This allows different detectors, e.g., H1 and Zeus, to compare their findings on a given accelerator, and also for physics results to be published in a detector-independent way.

The second use of a Monte Carlo is as a test, e.g., for the different theoretical descriptions of QCD, fragmentation and hadronisation. All Monte Carlo programs have a multitude of parameters which can be tuned to provide a better description of the data, and these parameters can only be tuned once the Monte Carlo simulations have been compared to real physics. It must be stressed, however, that great care must be taken when using Monte Carlo in this way because of the many different *ansätze*, or assumptions, of particular models, especially in non-perturbative regimes.

H1 makes use of many different kinds of Monte Carlo programs, with a wide range of generating conditions, but those most suitable for Deep Inelastic Scattering (DIS) studies are called MEAR [18], MEPS [19], DJANGO [20] [18] [21] and HERWIG [22]. MEAR and MEPS rely on a program called LEPTO [19] to provide the Leading Order Matrix Elements for the hard sub-process, but differ in their higher order approximations. The exact matrix elements for the hard subprocess and QCD processes up to order α_s are included in the matrix element. Higher order QCD emission is then simulated using approximate methods until the partons have been evolved to such a degree that their virtuality is low, usually $\sim$ 1-2 GeV . A hadronisation model is then required to form hadrons (and also to decay the short-lived ones) from these low-virtuality partons. MEAR uses the ARIADNE [18] program (a member of the "Lund family of Monte Carlo generators"), which utilizes the theory of the Colour Dipole Model coupled with other programs which handle the hard interactions, hadronisation and particle decays, e.g., JETSET [23]. MEPS uses the LEPTO [19] program (another member of the "Lund family") which uses a parton showering approach to calculate higher order QCD radiation. Hadronisation is performed using the LUND String Model implemented in JETSET [23]. DJANGO is very similar to MEAR in that it uses LEPTO and ARIADNE, but it also has the QED-radiation model of HERACLES [21] interfaced to the program. HERACLES provides both initial- and final-state QED radiation, and thus DJANGO may be compared with real data directly without the need for factoring-out, or correcting for, QED-radiative effects. HERWIG was written to provide an "all-in-one" QCD Monte Carlo which has no need of any external programs and stands for Hadron Emission Reactions With Interacting Gluons. The hard sub-process is calculated from a Matrix Element scheme (much like LEPTO does) but the fragmentation and hadronisation is done by a cluster algorithm. It does not, however, include QED radiation.

3.2 Matrix Elements and LEPTO

The matrix-element method, in which Feynman diagrams are calculated order by order, is, in principle, the correct approach because it takes into account the exact kinematics as well as the full interference and other terms. However, this procedure becomes particularly difficult in higher orders (especially the loop corrections) and increasingly difficult the higher the order desired. This means that most QCD calculations are only performed to the one-loop correction level. Evidence from PETRA/PEP [23] indicates that multiple soft gluon emission plays a significant role in building up the event structure.

LEPTO is used to calculate $eq \longrightarrow eq$ elastic scattering, *i.e.*, the rather simplistic Quark-Parton Model. LEPTO can also include QCD corrections using exact first order matrix elements and higher orders in the leading logarithm parton cascade (see 3.3). It is a general and flexible Monte Carlo which was designed to simulate the scattering of leptons and nucleons and to integrate the cross-sections. As has been mentioned above, LEPTO is based on the leading order electroweak cross-sections for the underlying parton level scattering process. The fragmentation of the produced partons into hadrons is performed using the Lund string hadronisation model [24].

3.3 MEPS, Parton Showers and JETSET

In order to take higher than first order QCD effects into account, the method used is the Parton Shower (PS) approach. This is not an exact treatment, as in fixed order Matrix Element (ME) calculations, but rather can be extended to arbitrarily high orders in α_s using the leading logarithm approach. What happens in this approach is that an arbitrary number of branchings of one parton into two (or sometimes more) can be joined which yields a description of multijet events. Parton showering is therefore expected to give a good description of the substructure of jets. By using this method, there is no explicit upper limit to the number of partons involved.

JETSET is used with LEPTO to construct the MEPS Monte Carlo, *i.e.*, Matrix Elements with Parton Showering.

3.4 MEAR, ARIADNE and Colour Dipole Model

The Colour Dipole Model is based on the fact that gluon emission can be thought of as radiation from the colour dipole between two quarks, or a quark antiquark pair. Softer gluons may be emitted between the newly-formed colour dipoles between the original particles and the first gluon.

In DIS, the CDM is different to conventional parton cascade models in that it does not divide the QCD cascade into an initial and a final state. The CDM assumes, instead, that all the radiation can be described as coming from the colour dipole formed between the struck quark and the hadron remnant. DIS differs from e^+e^- in that, in e^+e^- both q and $\bar{q}$ can be treated as point-like, whereas in DIS, only the struck quark can be treated as point-like, as the hadron remnant is an extended object.

In e^+e^- annihilation, the exact $O(\alpha_s)$ matrix element governs the first emission in the dipole cascade. This situation is arguably better than in other QCD shower programs, where different matching procedures of differing complexity are required to reproduce the form of the matrix element.

In DIS, the situation is not so straight-forward. There is the added complexity of boson-gluon fusion (sometimes called photon-gluon fusion at low Q^2), or BGF events, which is not described at all by the colour dipole model. The colour dipole model only considers the radiation between the struck quark and the proton remnant; it does not account for the incoming gluon splitting into a quark-antiquark pair. To overcome the problem, a first order matrix element is used to generate the boson-gluon fusion process and further emission is performed by dipole radiation. The gluons emitted by the colour dipoles do not obey strong angular ordering and this is the essential difference between MEAR and MEPS, in effect the difference between BFKL and DGLAP evolution respectively.

3.5 Initial- and Final-State QED Radiation and HERACLES

Electroweak corrections of $O(\alpha_s)$ consist of one-loop virtual corrections as well as of effects from real bremsstrahlung, *i.e.*, radiative events $ep \rightarrow e\gamma X$.

The separation of radiation into initial- and final-state showers is (largely) arbitrary, but there are situations where it is appropriate, e.g. when modelling e^+e^- physics in the reaction $e^+e^- \longrightarrow q\bar{q}$. In such an interaction, the QED radiation in the initial state is far greater than the final state. The QCD radiation is far more prevalent than the QED radiation for the $q\bar{q}$ and the latter can therefore be ignored. Recent studies at LEP, however, have made measurements of QED effects in the final state. In ep physics, however, the distinction is not so clear-cut due to the presence of a charged lepton in both the initial- and final-states (charged current interactions excepted).

HERACLES is used to simulate the process $lp \rightarrow l'X$ including electroweak corrections of $O(\alpha_s)$, and the bremsstrahlung process $lp \rightarrow l'\gamma X$. Electroweak corrections of $O(\alpha_s)$ are either one-loop virtual corrections, or real bremsstrahlung events where $ep \rightarrow e\gamma X$. These photons are, in the main, radiated from the incoming or scattered lepton. When HERACLES is combined with ARIADNE, the resulting program is called DJANGO.

3.6 DJANGO

DJANGO is a simplified name for the Monte Carlo resulting from the interfacing of MEAR (see section 3.4) and HERACLES (see section 3.5). In DJANGO, events are generated randomly from all the individual contributions according to their partial cross-sections. In HERACLES, only the scattered lepton and sometimes a radiative photon is added to the event record. Only when this portion is completed are the DJANGO6 routines called which both simulate QCD effects and construct the hadronic part of the event.

3.7 HERWIG and Cluster Algorithms

HERWIG was written with the express aim of constructing as complete a treatment as possible (which needed no or little recourse to external programs) of perturbative effects and a simple non-perturbative treatment. HERWIG starts from the hardest scattering and works "outwards" to link this back to a proton and a lepton. This method is valid for *strong* ordering in ${P_t}^2$ up to some hard scale, which, in Deep Inelastic Scattering, is Q^2. Contrary to some other Monte Carlo programs dealing with co-linear partons, where an x value is picked and evolved to some random hard scale, HERWIG picks the hard scale and evolves backwards to the starting point. In addition to this you also get soft gluons produced. The parton shower is an event-by-event implementation of the evolution equations of Altarelli-Parisi [25]. These soft gluons may well affect the strongly-ordered behaviour of the splitting, however, as they are soft (long wavelength) gluons, they are unable to resolve emitters, therefore can be treated as coherent. Quantum mechanical interference gives the same angular ordering in the gluon, and thus all the soft gluons can be replaced by one on-shell gluon, with strong angular ordering going with the P_t ordering, *i.e.*, with the largest angle nearest the hard sub-process.

In HERWIG, the final state hadrons are created using cluster fragmentation from the quarks, diquarks and gluons produced in the parton shower process. This model (like string fragmentation) uses the QCD property of confinement between coloured objects. Confinement causes a parton to be close in phase-space (and real space) to a parton of opposite colour; this is true both for the production of $q\bar{q}$ pairs and also the more complex situation in a parton shower. It is possible, therefore, to associate the colour-confined partons into colourless objects called clusters which may be comprised of $q\bar{q}$ pairs or diquark-antidiquark systems. Any gluons present at the end of the parton shower are split into $q\bar{q}$ pairs which are subsequently clustered in the same fashion.

All of the theoretical considerations in the above paragraphs assume that the partons have zero mass, but what about heavy quarks? Heavy quarks can also be modelled by HERWIG due to the fact that they look and behave like (light) quarks at large angles.

The naïve cluster model used by HERWIG decays colour singlet clusters according to pure phase space. This has the following advantages :-

- The decay is presumed to be isotropic;

- Higher masses are automatically suppressed;

- It is *fully* specified by phase space, *i.e.*, there are no adjustable parameters.

However, some clusters are too heavy to decay isotropically and, as a result, longitudinal splitting is obtained. This splitting of the heavy clusters continues until the clusters are no longer too heavy to decay isotropically via the cluster model. An illustration of such a process is shown schematically in figure 3.1, and is referred to as a "trouser" diagram in HERWIG parlance. It can be seen that many clusters (five in figure 3.1) are created from what would have been one very heavy cluster. The initiating gluon splits into a q$\bar{\text{q}}$ pair, each of these radiating gluon bremsstrahlung which, in turn, may radiate further gluon bremsstrahlung and/or splitting into more q$\bar{\text{q}}$ pairs. Finally, clusters are formed which can decay isotropically.

Only about 15% of the clusters do actually get split, but they account for more than 50% of the hadrons! This is because the heavy clusters really are very heavy indeed, leading to multiple trousers.

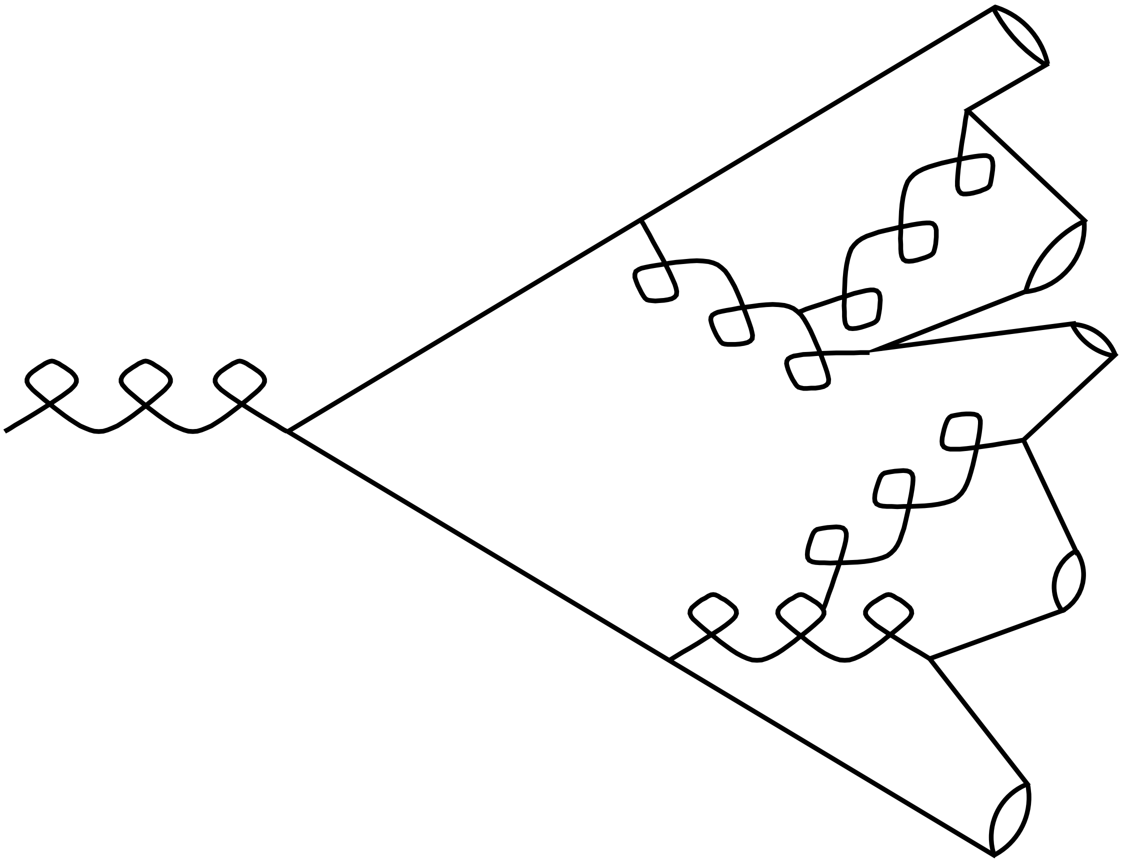

Figure 3.1: Schematic example of a “trouser” diagram in HERWIG

3.8 Failures Of The Monte Carlo

Serious faults were found with two of the above Monte Carlo, which rendered them useless to this analysis. These two Monte Carlo were DJANGO6 and HERWIG5.8d.

3.8.1 DJANGO6

It was noticed quite early that no distributions existed for $\bar{d}$ or $\bar{u}$ quarks, and the distributions for u and d quarks did not look anything like those of MEAR or MEPS. The fact that there was no antimatter in DJANGO6 was quite puzzling, to say the least, and by various means of logical deduction, we came to the conclusion that the most probable cause was that somewhere in the code, an absolute value of the Particle Data Group identification [26] code was being taken rather than the true value. This postulate was in fact verified, and the fault has been corrected and will be available for use in future versions of the program.

It was also found that p_t was not conserved in an alarmingly large fraction of events, but once again, an explanation was found [27] and the problem resolved.

There were also many problems with the event record, specifically, the records could not be matched with any physical process as laid-down in the generator-level descriptions. For example, about 1 % of events were something like $u \longrightarrow \bar{u}$! This problem was also communicated to the librarian (as well as numerous other problems from H1 users) and future versions will have had this problem also corrected.

3.8.2 HERWIG 5.8d

It is believed that, from the expectations at the start of this thesis, if a graph were to be drawn of C_j *versus* the actual charge of the parton (see section 5.2, specifically equation 5.2), some kind of relationship should become obvious (see chapter 5), *i.e.*, we are searching for a correlation of a quark charge with a hadron charge. The relationship we expect is a higher value of mean C_j with higher value of parton charge, the value decreasing in line with the parton charge so that the most negative value of mean C_j is associated with the most negative parton charge. If we look at a graph of the mean C_j divided by the RMS width, σ of the distributions for

MEPS and MEAR in figure 3.2 then we do indeed see this relationship, but *not* for HERWIG 5.8d.

3.8.3 Charge Symmetry

In order to investigate the charge carried by each hadron, we plot the product of both the momentum in the Breit frame and the charge of each track. By doing this at the generator level, and obtaining 4 distributions (one for each of the charged parton types) it is possible to test charge conjugation symmetry conservation. We expect, with infinite statistics, that the distribution for +2/3 quarks should be identical to -1 times that of the -2/3 quarks. By plotting the positive quark distributions superimposed on the reversed negative quark distributions, any anomalies or differences should be immediately apparent. The next six plots, figures 3.3 to 3.8 inclusive, show the positively charged distributions as full histograms with the negatively-charged, reversed distributions as black dots. Figures 3.3 and 3.4 have which charges are plotted written on them; the same convention is used for the other diagrams even though it is not explicitly written on them.

MEAR and MEPS behaved exactly as we expected, as shown in figures 3.3 and 3.4, and no anomalies were found (only within the error bars), but this is *not* what we saw in HERWIG 5.8d shown in figure 3.5. Large differences were found between the two distributions, implying a breaking of the law of charge-conjugation symmetry. This anomaly led to many discussions with the author of the program [28] who provided many suggestions for further tests. A summary of the discussion as well as the investigations suggested are given below.

Initially, it was postulated that the actual discrepancy was not so much between positive and negative charges but between valence and non-valence flavours, *i.e.*, it is expected that the difference between d and $\bar{d}$ quarks is considerably greater than between s and $\bar{s}$. This difference arises because in events in which there is no initial-state radiation, a struck sea quark is colour-connected to the fast-moving proton diquark remnant, while a struck sea anti-quark is connected to the slower-moving valence quark remnant. Thus, the string-drag effect (present in HERWIG even though string language is not explicitly used) accelerates them differently. Unfortunately, this was not found to be the case, as can be seen in figures 3.6 and 3.7, because the version used was HERWIG 5.8d. The change from 5.8d to 5.9 incorporated the most important changes from an internal-only ZEUS version (5.8c+),

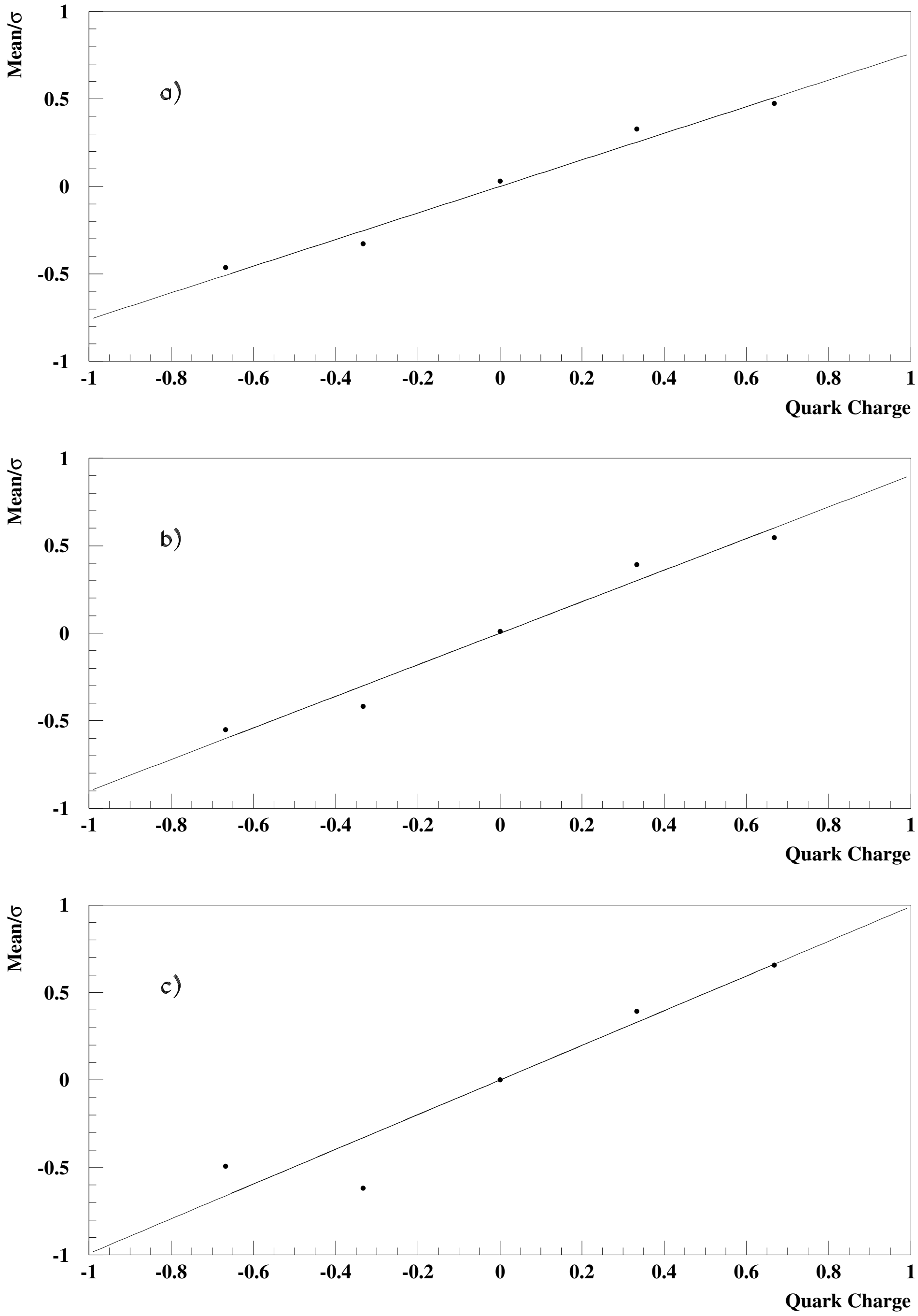

Figure 3.2: Graph showing how the mean/σ varies with initiating parton charge for MEPS, MEAR and HERWIG labelled a), b) and c) respectively. The -1/3 charge in HERWIG is clearly more negative than the -2/3 charge

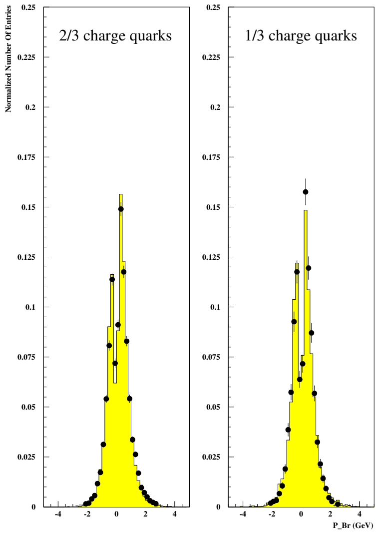

Figure 3.3: Product of track momentum in the Breit frame and track charge for the MEAR64 Monte Carlo simulation. On the left are the $\pm$ 2/3 charge plots, with the $\pm$ 1/3 on the right. Note the asymmetry in the distributions, with a greater number of positive tracks reflecting the charge of the struck quark (N.B. Negative quark distributions, being the solid black dots, are reversed!). Another feature worth noting is that the $\pm$ 2/3 distributions have a more densely populated positive tail, *i.e.*, compare the tails at $\mathrm{P_{Br}} = 3$

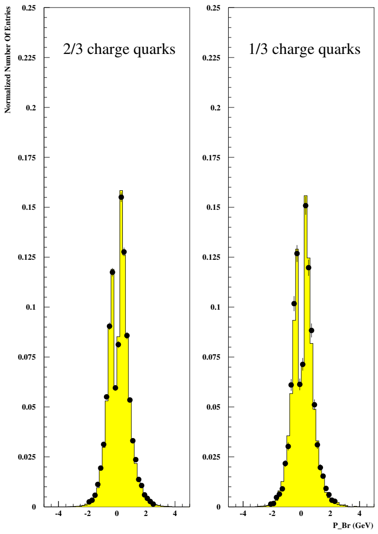

Figure 3.4: Product of track momentum in the Breit frame and track charge for the MEPS64 Monte Carlo simulation. Note the similarity between these distributions and those for the MEAR64 Monte Carlo above, *i.e.*, the asymmetry about zero and the higher extent in the positive $\mathrm{P_{Br}}$ tail.

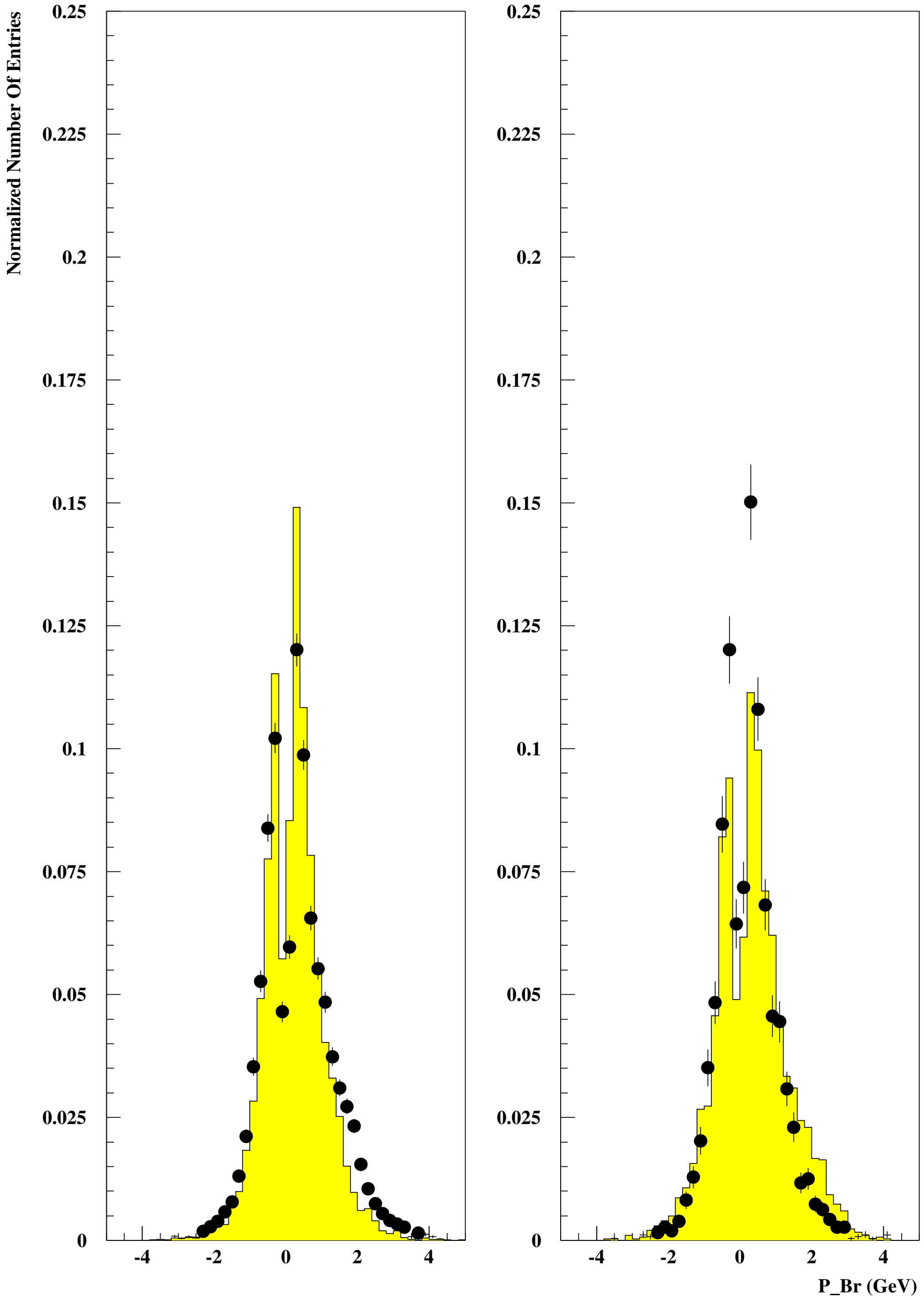

Figure 3.5: Product of track momentum in the Breit frame and track charge for the HERWIG 5.8d Monte Carlo simulation. Note the difference between these distributions and the ones for MEAR64 (Figure 3.3) and MEPS64 (Figure 3.4). The asymmetry about zero is still present, though not so well modelled, however, there is considerable disparity in the tails for both the $\pm\, 1/3$ *and* the $\pm\, 2/3$ distributions.

where a similar effect was observed in photoproduction, and affects the way unusually heavy clusters (e.g. remnant clusters) fragment. The original model was set up assuming that only a very small fraction of clusters is heavy, therefore not much care was taken about how they fragment. Unfortunately, in small-x DIS most events contain at least one heavy cluster (the remnant) and thus the results depend much more critically on the treatment. In the new version, the heavy clusters split in a chain-like fashion into lighter clusters before decaying into hadrons. This being the case, the amount of drag exerted on the struck quark is much less, and therefore the quarks and antiquarks should have much more similar distributions. This is exactly what is seen in figures 3.6 and 3.7.

This meant that there was still a problem with HERWIG5.8d which was not resolved by discussions with the author. Therefore, it was decided that usage of this Monte Carlo should be suspended until this problem was resolved.

The resolution (with the code changes from 5.8c+ as mentioned above) arrived in HERWIG 5.9, and this was tested for the same anomaly, and the distributions were now found to be nearly identical. This is shown in figure 3.8.

This charge-anomaly is now corrected, but there remains another difficulty with HERWIG (any version) which impedes our use of it as a Monte Carlo model.

3.8.4 Gluons and HERWIG

When the hard matrix element correction stage of HERWIG was under development, the authors were very concerned [28] that users would attempt to use the (unphysical) information about the nature of the generated event. To circumvent this possibility, the information is deliberately destroyed even though the information *is* available at *very* early stages of event generation (before it becomes useful for this analysis). The main intention of doing this [28] was to help people to realise that the information is potentially unphysical, and of such a different nature to that in LEPTO or ARIADNE, that comparisons should not be made.

The information that this analysis (for example) requires is fundamentally model dependent, and it is in this sense that we must be clear about what we are attempting to do. There are two types of model-dependence which we must distinguish at this point.

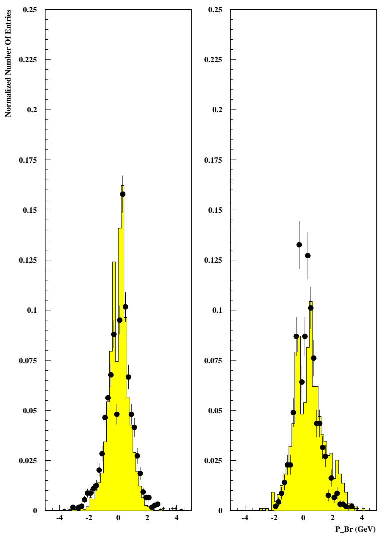

Figure 3.6: Product of track momentum in the Breit frame and track charge for sea quarks in the HERWIG 5.8d Monte Carlo simulation. The expected asymmetry about zero is very poorly modelled, and the shape of the distributions in the tails is significantly different.

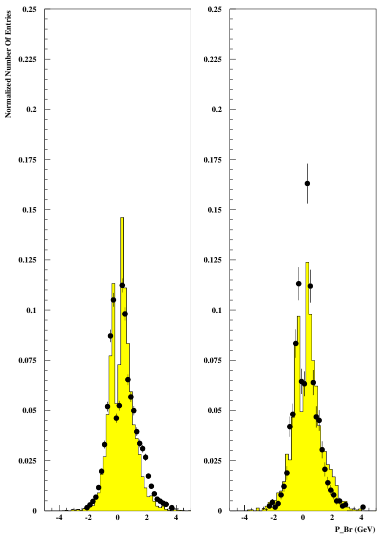

Figure 3.7: Product of track momentum in the Breit frame and track charge for valence quarks in the HERWIG 5.8d Monte Carlo simulation. Comparing these distributions with those for the sea quarks above, we see that not only are the shape of the tails significantly different from the charge-reversed partner distribution (black dots), but that the asymmetry about zero is not as expected either.

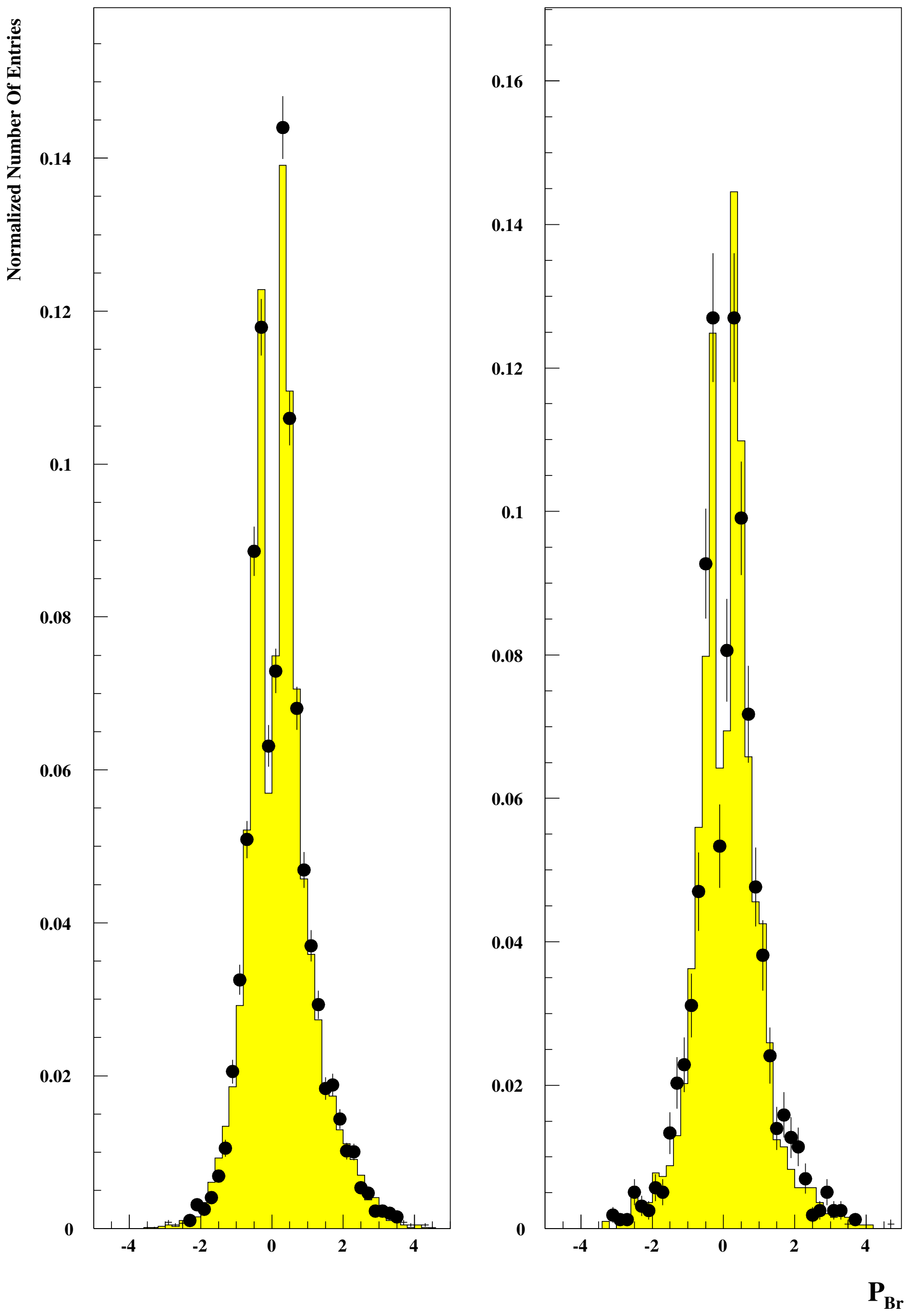

Figure 3.8: Product of track momentum in the Breit frame and track charge for the HERWIG 5.9 Monte Carlo simulation. Comparing these distributions with the earlier ones shown for HERWIG 5.8d (e.g. figure 3.5) we see that charge-conjugation symmetry is restored to HERWIG.

1. No models are perfect representations of QCD; different models approximate it in different ways.

2. Within any particular model, one can make different choices for how one labels particular physical effects.

Number 1 is a valuable tool as it allows physicists to get an estimate of the errors that any particular model makes when approximating a theory. Number 2, however, is completely misleading because if one were to take notice of the labelling, then it is entirely possible that two models which give identical physical predictions are actually different because an identical event could be labelled differently in different models. One could therefore draw the conclusion that a given phenomenon is poorly understood even if all the available models give predictions in perfect agreement with data!

Matrix element matching is a particular example of number 2. What normally happens is that one picks a completely arbitrary surface in phase space and defines one side to be the "parton shower emission" and the other the "hard sub-process". Consider two events, either side of the border but arbitrarily close to one another so that any physical definition will be unable to distinguish the two. The two events should be completely indistinguishable with both parton cascades yielding identical parton-level results which subsequently hadronise to give identical hadron-level results.

HERWIG, LEPTO and ARIADNE all use different borders between these two regions, and thus any analysis which uses this unphysical information about how particular events were generated would give different results in the different models. The main reason that the information was included was to help physicists to realise that this is the case, but unfortunately such model-dependence is often interpreted as the first kind as opposed to what it really is, the second kind. In an endeavour to circumvent this problem and to obtain a systematic error for the model dependency *á la* method number 1, an attempt to define our own phase space was tried, and the approach used is documented in section 6.6.

Chapter 4

Data Selection

4.1 Introduction

Data selection is an important part of any analysis, and serious thought must be given to how exactly this should be done. We must consider whether or not our selections will bias the sample in any way, and if so, how and by how much, and also whether or not this can be accounted for.

Event selection begins by considering what general features of the chosen channel (or channels) can be distinguished from other classes of physics and/or background, and how best to translate event characteristics into quantities measured by the detector upon which we can cut. Typical low Q^2 characteristics are energy deposits from the scattered lepton in the backward electromagnetic calorimeter (section 2.5.2) with hadronic energy deposited in the Liquid Argon Calorimeter (section 2.5.1) as well as tracks through the Central and Forward Tracking Detectors (sections 2.4.1 and 2.4.1, respectively). A typical signature for a high Q^2 event is electromagnetic energy from the scattered lepton deposited in the liquid argon calorimeter.

Therefore, before any data may be analysed, a consistent set of selection criteria and cuts must be satisfied in order to maintain a high purity, high quality sample of events. Such criteria are necessary for eliminating background interactions such as beam-wall events or cosmic muons. These criteria must be laid out for data and both reconstructed and generated Monte Carlo simulations separately. Details of how to obtain neutral current DIS events from all recorded data will be set forth here.

4.2 Event Classification

Event classification at H1 is performed at both L4 (see section 2.11.5) and L5 (see section 2.12) by the module ECLASS. Events which satisfy the classification scheme (for all physics classes) are written to a production output tape (POT) or, if the module ECLDST has also been used, a data summary tape (DST). NC DIS data are then selected from these DSTs by choosing the appropriate event class or classes. A candidate may have more than one event class, depending on how many event-class definitions it may satisfy. Candidates for NC DIS events are further divided into two sub-classes, those of low Q^2 and high Q^2. The definitions that ECLASS uses to differentiate between the two sub-classes are:

- Low Q^2 events are those with a scattered electron candidate in the BEMC. They are assigned the class NCLQSQ.
- High Q^2 events are those with a scattered electron candidate in the LAr. They are assigned the class NCHQSQ.

4.2.1 The NCLQSQ class

The criteria for low Q^2 neutral current DIS candidates are as follows: The most energetic cluster in the BEMC has $E_{clus} > 8$ GeV and has an associated BPC hit. It must be in a region of good acceptance within the BEMC and it must be well-contained within that region of good acceptance, *i.e.*, $157° < \theta < 173°$.

4.2.2 The NCHQSQ class

High Q^2 neutral current DIS candidates are initially chosen from events where the reconstructed cluster energy in the LAr is greater than 5 GeV . The selection is further refined by making additional requirements on these clusters. These include only accepting those clusters whose $E_0 > 5$ GeV , where E_0 is the energy of the cluster on the electromagnetic energy scale; and also rejecting those clusters which either lie within 5° of a phi crack or have an "electromagnetic" to total energy fraction of less than 50 %.

Before classification can be finalised, a few further requirements are made. The detector is "cut" into pieces, with the following criteria on the cluster with the largest $E^e_{T_0}$, where $E^e_{T_0}$ is the *transverse* energy of the cluster on the electromagnetic scale:

- $10° \leq \theta < 45°$ and $E^e_{T_0} \geq 8$ GeV
- $45° \leq \theta < 160°$ and $E^e_{T_0} \geq 5$ GeV

The event is then considered for an electron candidate, provided that there is, in fact, at least one candidate without a muon track behind the cluster with the largest $E^e_{T_0}$, and which has a $P_{t_{miss}}$ for the event of < 40 GeV . The halo muon flag must *not* be set.

It might appear at a cursory glance that these cuts are very loose, allowing large missing P_t and a low electromagnetic energy fraction, but this is to ensure that good high Q^2 events are not rejected due to detector and reconstruction effects.

Now that we have a sample of data selected, we can impose further cuts and select good events on which physics analyses can be performed.

4.3 Event Selection

Initial rejection is of events where there is a known malfunction or problem with the apparatus. These problems are flagged in the data and are easily removed. This produces a sample of data which is of good quality, where all the high voltages of the detectors are on, and the "noise", *i.e.*, random signals in the detector, can be accounted for. Further selection, based more on physics considerations, can now be applied.

4.3.1 Electron Selection

Photoproduction events are a large contribution to the contamination of DIS events, mainly due to the fact that low energy pions can be misidentified as scattered electrons in the BEMC. To avoid such misidentifications, the scattered electron's energy is required to be greater than 14 GeV . This removes (nearly) all photoproduction events yielding a sample which is better than 99.7% pure DIS [8].

Definitions of electron topologies for both low and high Q^2 samples have been given above, and these are now applied at this level.

4.3.2 "Struck Quark" Selection

To assure good detector acceptance for charged tracks in DIS events, the "struck quark" must lie within the polar region $10^\circ < \theta_{quark} < 150^\circ$. The "struck quark" is, of course, an artificial construct and cannot be detected (due to QCD confinement) but its 4-vector can be calculated from the knowledge of the electron's kinematics, and it is this which is referred to when the term "struck quark" is used.

4.3.3 Vertex Selection

A requirement on the vertex position is applied to ensure that the kinematic variables of the event, as determined from the scattered lepton, are well determined and also to discriminate against beam-gas interactions. The position is given by the H1 tracking system, and the cut is -30 cm$< z_{vertex} < +30$ cm.

4.3.4 Diffractive Events

The definition of diffractive events has been discussed in section 1.4. They, unlike most DIS events, leave very little trace in the forward detectors (even though they have enormous forward energy!) and are not included in the standard QCD models used in this analysis. As a consequence, they are not considered in this analysis, but as only $\sim$ 6 % of all DIS events are diffractive, the loss is acceptable. The safe criterion needed to remove these events are as follows: No summed calorimetric cluster in the forward region of the LAr should have less than 0.5 GeV of energy, where the forward region is defined (see section 2.5.1) as $4.4^\circ < \theta_{\mathrm{LAr\ cluster}} < 15^\circ$.

4.3.5 Hadronic Activity

A DIS event is characterised by a high level of hadronic activity in the central part of the detector. To ensure that the hadronic activity seen is true DIS and not some background contribution, a minimum value of the invariant mass of the hadronic system is required. This selection is applied to kinematics calculated using both electron and hadronic (Jacquet-Blondel) methods. The low y region is also susceptible to QED-radiative corrections and background from photoproduction, and in order to suppress both of the above, and to maintain good kinematic reconstruction, a cut on y is also applied. W^2 and y are also related by the approximate relation $W^2 \approx sy$. These criteria are laid-out on the Energy Flow and Particle Spectra Working Group Web Pages [29], and are summarized below.

- $W^2 > 4400$ GeV^2
- $0.05 < y < 0.6$

These selections yield an useful data sample of $\sim$ 50k events whereas the Monte Carlo used yielded useful samples of $\sim$ 30k events each.

4.4 Track Selection

At this point, any and all events that have passed the criteria laid out above are accepted for analysis. Not all tracks, however, should be accepted for analysis and the following selection criteria are imposed in order to obtain a set of good tracks in each event.

Most of the H1 tracking is performed by the central and forward tracking detectors, as described in sections 2.4.1 and 2.4.2. H1 refers to tracks coming from the primary vertex as primary tracks, whereas all other vertices, e.g., decays of hadrons or from interactions, give rise to secondary tracks. Different selection criteria are applied to a track depending on which part or parts of the detector it passes through, *i.e.*, tracks passing through the central tracker only are subject to central track selection, likewise forward and combined (*mutatis mutandis*) where a combined track is a track passing through both central and forward trackers.

4.4.1 Central Tracks

To avoid reconstruction problems due to poor track finding efficiency at low momentum where the tracks curl back on themselves, it is important that all tracks pass through both chambers of the CJC. This can be assured with a transverse momentum cut.

Mis-interpretation of the scattered lepton as a hadron may occur if the BEMC is included in the acceptance region, and thus a cut on θ should be applied to constrain the acceptance to the CJC region.

To ensure that only high quality tracks are considered for the analysis, the vertex must be constrained not only in the z-direction, as mentioned in section 4.3.3, but also in the $x - y$ plane, *i.e.*, the distance of closest approach to the z-axis must be constrained. Another consideration for high quality tracks is the actual track length itself. This must be long enough to eliminate tracks reconstructed from randomly associated points.

High energy tracks do not curve much in the magnetic field, consequently their momentum measurement, p, will have a large associated error, δp. We require that the largest error that the momentum can have to be no greater than the value of the momentum itself.

The selections are collated and presented below:

1. $p_t^{\mathrm{min}} > 0.15$ GeV
2. $\theta < 158^\circ$
3. $\mathrm{DCA}_{\mathrm{max}} < 2.0$ cm
4. Length > 10.0 cm
5. $\delta p/p < 1.0$

4.4.2 Forward Tracks

Different criteria are needed for considering forward tracks, but the one common selection is that on the error in momentum, *i.e.*, , $\delta p/p < 1.0$.

The forward tracker's acceptance region must also be selected, *i.e.*, $8° < \theta < 30°$.

A minimum momentum is also required to suppress low momentum scattering, e.g., between different sub-detectors, *i.e.*, CJC end-wall, or in non-sensitive materials.

Tracks are associated with the primary vertex by extrapolating them back to the vertex. To maintain the quality of the extrapolation, the distance between the track and the vertex in the $x-y$ plane, R_0, must be less than 2.0 cm. This is an analogous selection to the DCA cut for central tracks. Another selection for the vertex is the $\chi^2_{vertexfit}$. Tracks with a low probability of originating from the vertex are rejected.

Another χ^2 cut is on the track fit. This can be used to suppress those tracks associated with wrong hits in the forward tracker.

The last requirement is a technical one, and it stipulates that there must be at least one primary segment from the planar chambers. This is necessary as these are much less contaminated by random signals than the radial segments.

All of these criteria may be found (with a fuller description) elsewhere [30] and are succinctly summarised below:

1. $p^{min} > 0.5$ GeV
2. $\delta p/p < 1.0$
3. $8° < \theta < 30°$
4. $R_0 < 2.0$ cm
5. $\chi^2_{vertexfit} < 10$
6. $\chi^2_{trackfit} < 25$
7. Number of hits in the planar chambers > 1

4.4.3 Combined Tracks

Combined tracks are those which pass through both central and forward tracking detectors. The selection criteria are taken from those defined above and are listed below:

1. $p_t^{min} > 0.15\ GeV$
2. $5.7^\circ \leq \theta \leq 28.6^\circ$
3. $\chi^2_{vertexfit} < 10$
4. $\chi^2_{linkfit} < 25$
5. $\delta p/p < 1.0$

The selection criteria described in the above sections are applied to both reconstructed data and simulated Monte Carlo data, and it can be seen in figures 4.1, 4.2, 4.3 and 4.4 that Monte Carlo and data are consistent within errors at both low Q^2 and high Q^2. It is not always a certainty that data and Monte Carlo simulation agree - especially if the model is wrong! - but such close agreement is a good indicator that the models used in both the MEAR64 and MEPS64 Monte Carlo simulations are a good description of the data.

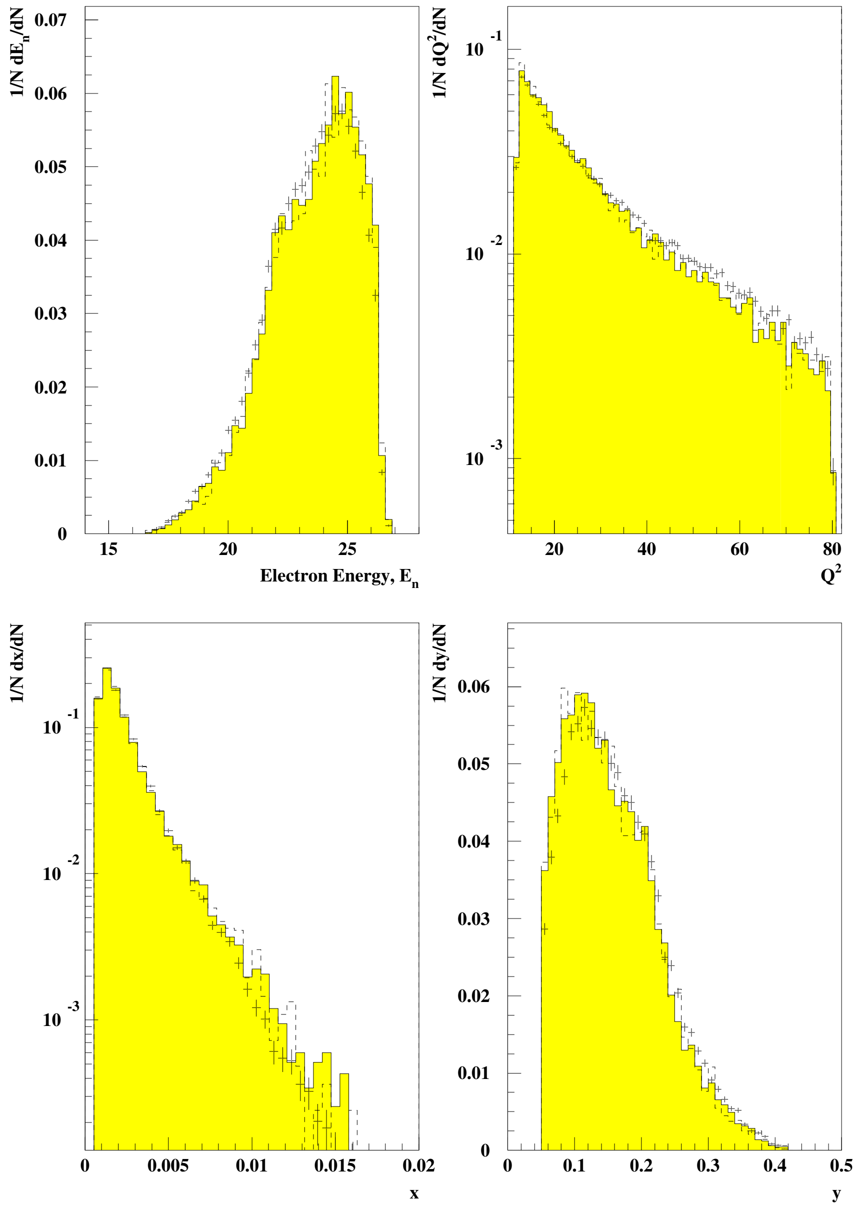

Figure 4.1: Comparison of selected data and Monte Carlo event quantities at low Q^2

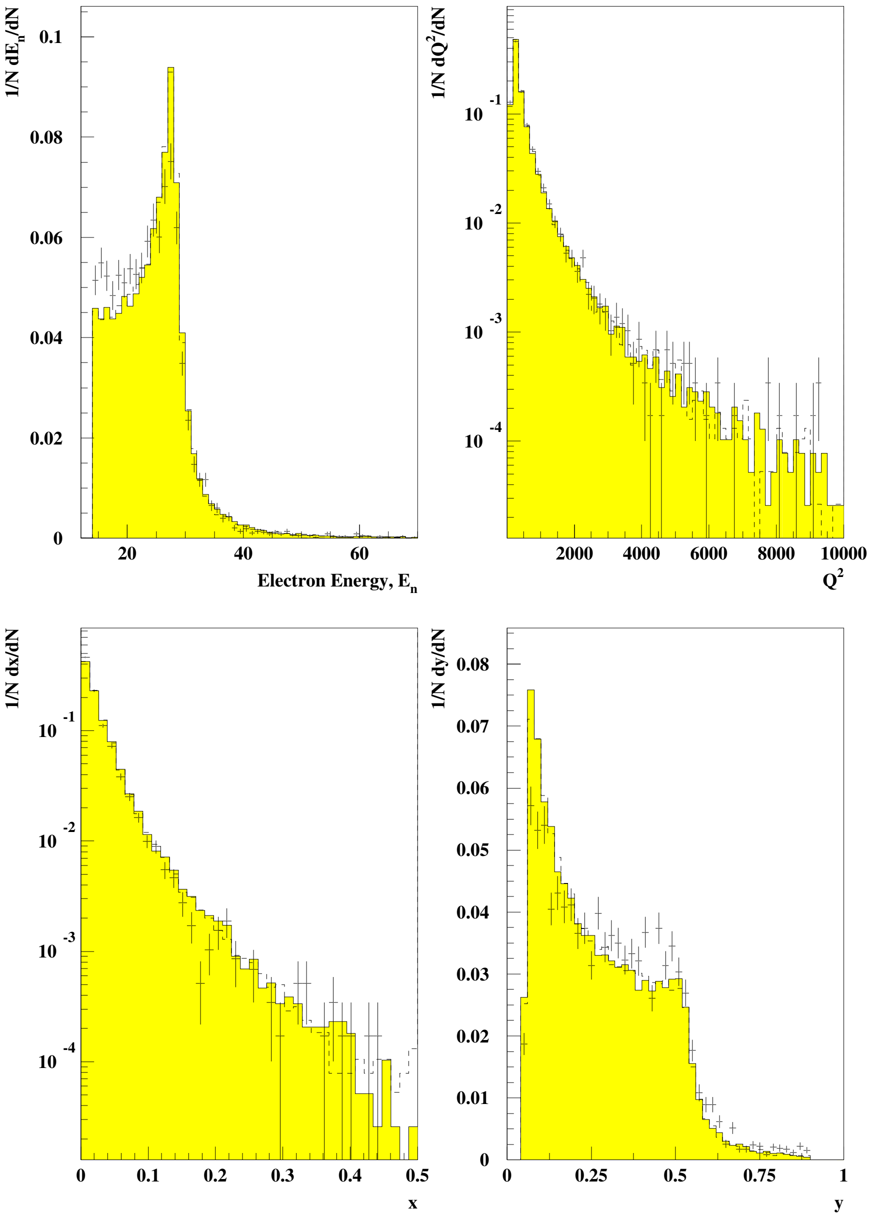

Figure 4.2: Comparison of selected data and Monte Carlo event quantities at high Q^2

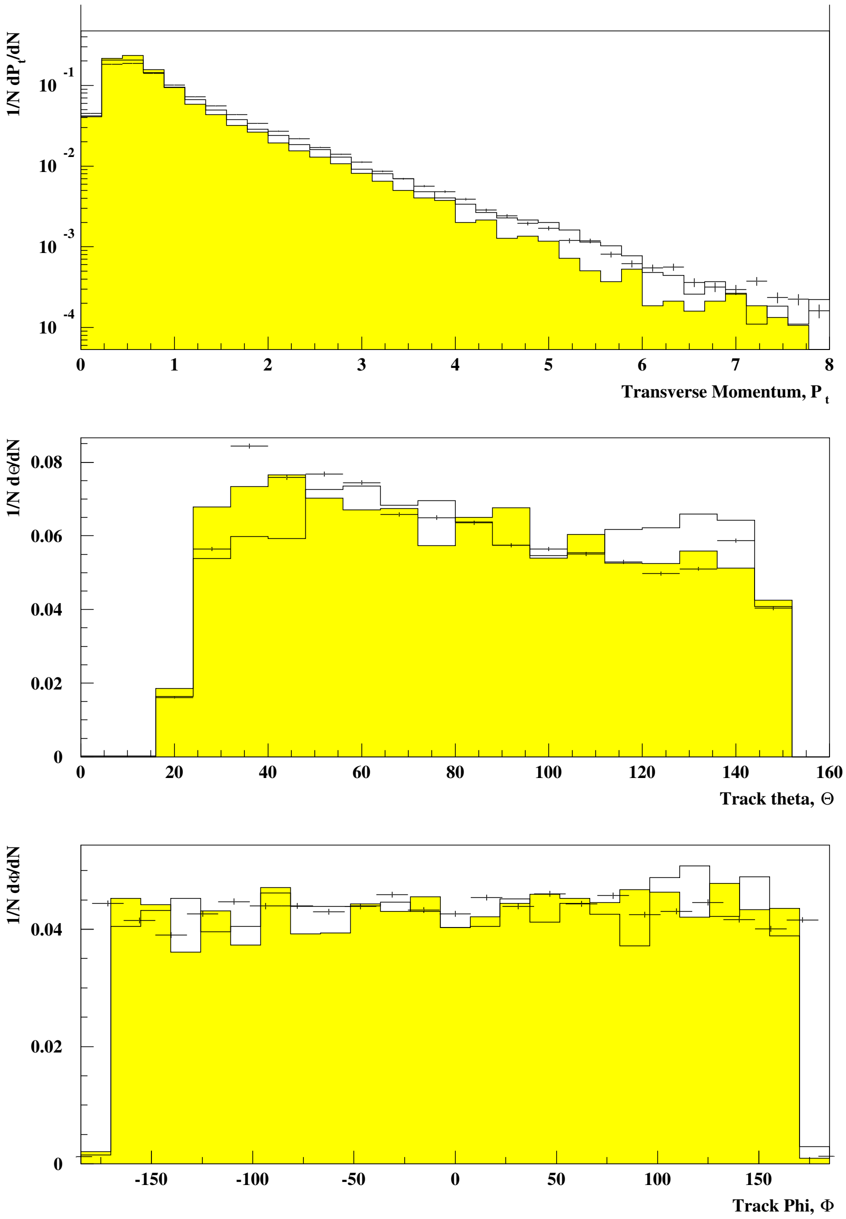

Figure 4.3: Comparison of selected data and Monte Carlo track quantities at low Q^2

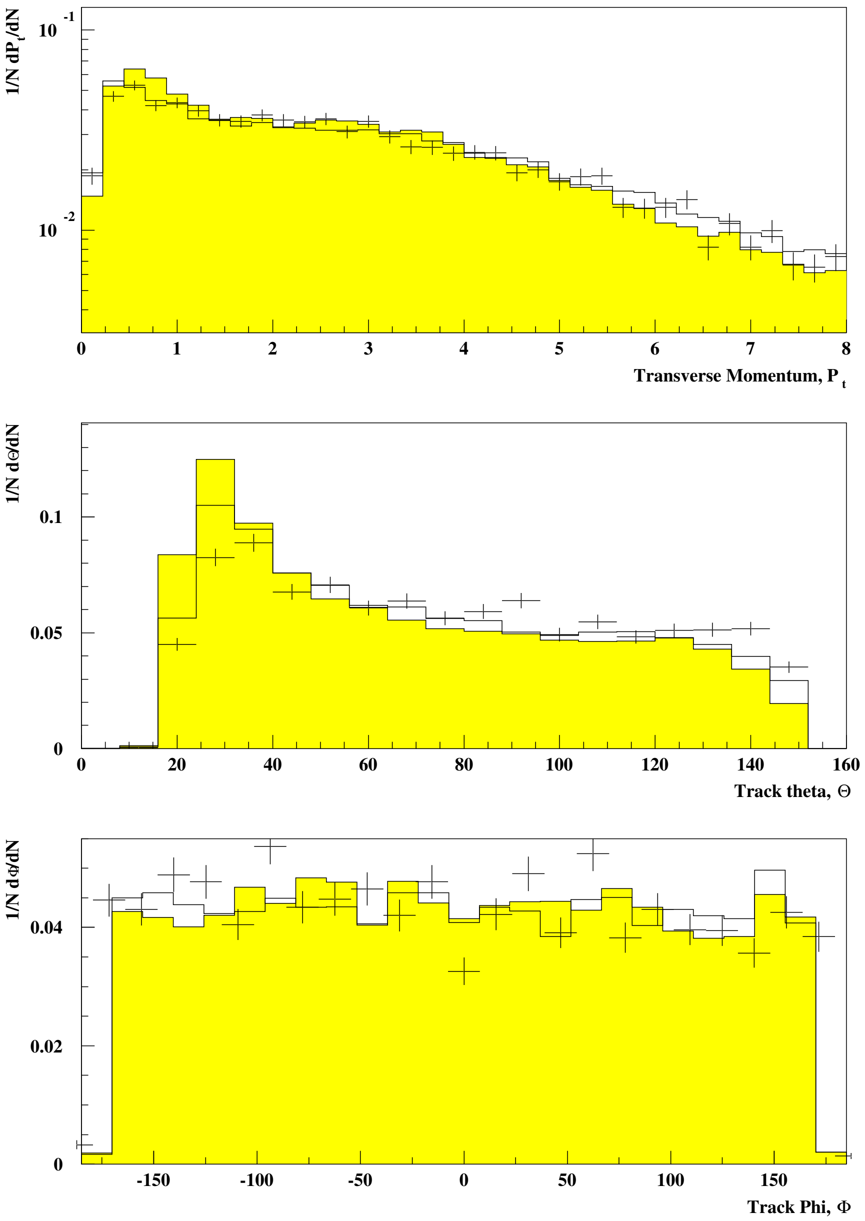

Figure 4.4: Comparison of selected data and Monte Carlo track quantities at high Q^2

Chapter 5

Analysis I

5.1 Introduction

The aim of the thesis is to provide an insight into, and an understanding of, the nature of the constituents of the proton when probed by a photon in e^+p Deep Inelastic Scattering. The assumption is that the charge of the fragmented hadrons bears some relation to the charge of the struck quark (mentioned in section 3.8.2). Clearly, complications arise due to the fact that hadrons have integer charges whereas quarks have fractional charges. Fractional charges have never been observed in nature due to QCD confinement of quarks within hadrons, and thus the fragmentation of the struck quark and the subsequent hadronisation will tend to dilute the "memory" of the struck quark's charge. At high energies, the amount of colour charge "seen" in the colour cloud of the quark is different to that "seen" at low energies. This "anti-screening" of the colour charge is referred to as *asymptotic freedom* and means that, at short range, the quark behaves as if it were essentially free. Thus, DIS of a proton is essentially elastic scattering from a quasi-free quark.

5.2 Charge Alternation Of Hadrons

Preliminary studies [31] indicated that the leading hadron carries much of the information about the charge of the struck quark, where the leading hadron is defined as the hadron that has the greatest momentum fraction of the hadrons resulting from the hard sub-process. The diagram illustrating this point (from [31]) is reproduced here as figure 5.1)[1].

Consider the simple case of ordering the hadrons in momentum so that the highest momentum hadron is called the first rank, the next highest momentum the second rank *etc.* An investigation into the behaviour of the charge in successive orders of hadrons as a function of the preceding order may then be performed on data, and the results can be compared to a simple model. The model is illustrated in figure 5.2, where the quarks used for the illustration are the valence quarks, u and d. More generally, any $+\frac{2}{3}$ quark could be used for an u quark, similarly any $-\frac{1}{3}$ quark for a d quark. Let us assume for the moment that the struck quark is the left-most u quark in the upper half of figure 5.2. The diagram shows two possible paths to creating a hadron: a) an $u\bar{u}$ pair is created; b) a $d\bar{d}$ pair is created. If path a) is chosen, a neutral particle such as a π^0 is created leaving behind another u quark and thus the upper half of the diagram can be used once more. However, if path b) is chosen, a hadron such as a π^+ is created from a $u\bar{d}$ combination, leaving behind a d quark. The options for d quarks are illustrated in the lower half of the diagram where similar paths to the upper half are shown. The model makes use of the fact that there is a colour force between the struck quark and the partons left behind in the proton remnant. This colour force is sometimes referred to as a string or colour flux tube. The creation of $q\bar{q}$ pairs is possible when the stored energy in the string becomes large enough for this process to be kinematically allowed. Further strings are created between these subsequent quarks, and the process repeats itself until no more pair production is viable.

According to this model, the leading hadron should have the same charge sign as the struck quark, *i.e.*, an u quark will give rise to a positively charged hadron *no matter* how many neutral particles are created first. This leaves a d quark to continue the fragmentation in the next rank yielding this time a negatively charged hadron. This process, again, is independent of how many neutral particles are created in the mean time. A struck d quark will yield a negatively charged hadron in the first rank and will result in a positive hadron being produced in the second rank due to the u quark resulting from the hadronisation process. The pattern then repeats in a charge-alternating way down through successive ranks of hadrons. What is observed in data at the second rank when we have stipulated the charge of the first is shown in figure 5.3.

Although the outcome is not entirely as predicted by the model, there **is** a significant

[1] Note that this is based on 1993 data

difference between the amount of positive hadrons seen in the second rank if the first rank is negative, and *vice versa.* A variable describing the alternation, C_a, can be defined as follows

$$C_a = \frac{E - U}{E + U} \tag{5.1}$$

where E is the number of events *expected* by our model ($+-$ or $-+$), and U is the number of *unexpected* events in a given rank ($++$ or $--$). For example, if the struck quark is positive, then we expect the first rank hadron to be positively charged. Given that this is the case, we then expect the second rank hadron to be negatively charged. C_a values are tabulated in tables 5.1, 5.2, and 5.3 for Data and two types of Reconstructed Monte Carlo, as well as the percentage contribution from each of the expected and unexpected classes. Note that the errors are larger in the Monte Carlo because there are fewer statistics than in the data. The middle two columns show an excess of events at the expense of the outer two columns, given that random behaviour is expected to yield 25% in each column.

It can be seen that the C_a values are *significantly* different from zero *i.e.*, about seven standard deviations. This confirms our supposition that the charge of hadrons in one generation, n, is influenced by the charge of the hadron in the preceding generation, $n - 1$. It is *possible* that the cause of (or even amplification of) these effects are resonances, e.g. ρ° particles, though it is unlikely to be the the main contribution to any explanation as the effects are also observed well-away from any resonant mass regions such as the ρ° mass [31].

The method presented thus far only uses information from one generation of hadrons at a time, whereas a far more general method is one in which the information from **all** the generations of hadrons can be used at the same time. At the most general level, we can define an average charge, C_j, for a collection of hadrons in an event, j, such that

$$C_j = \sum_i w_{ij} q_i \tag{5.2}$$

where q_i is the charge of an hadronic track, and w_{ij} is a dimensionless weight of some kind. It remains to experiment on simulated Monte Carlo data to decide what form this weight should take, and also over which particles the sum should be performed, in order to best get a correlation with the charge of the scattered quark.

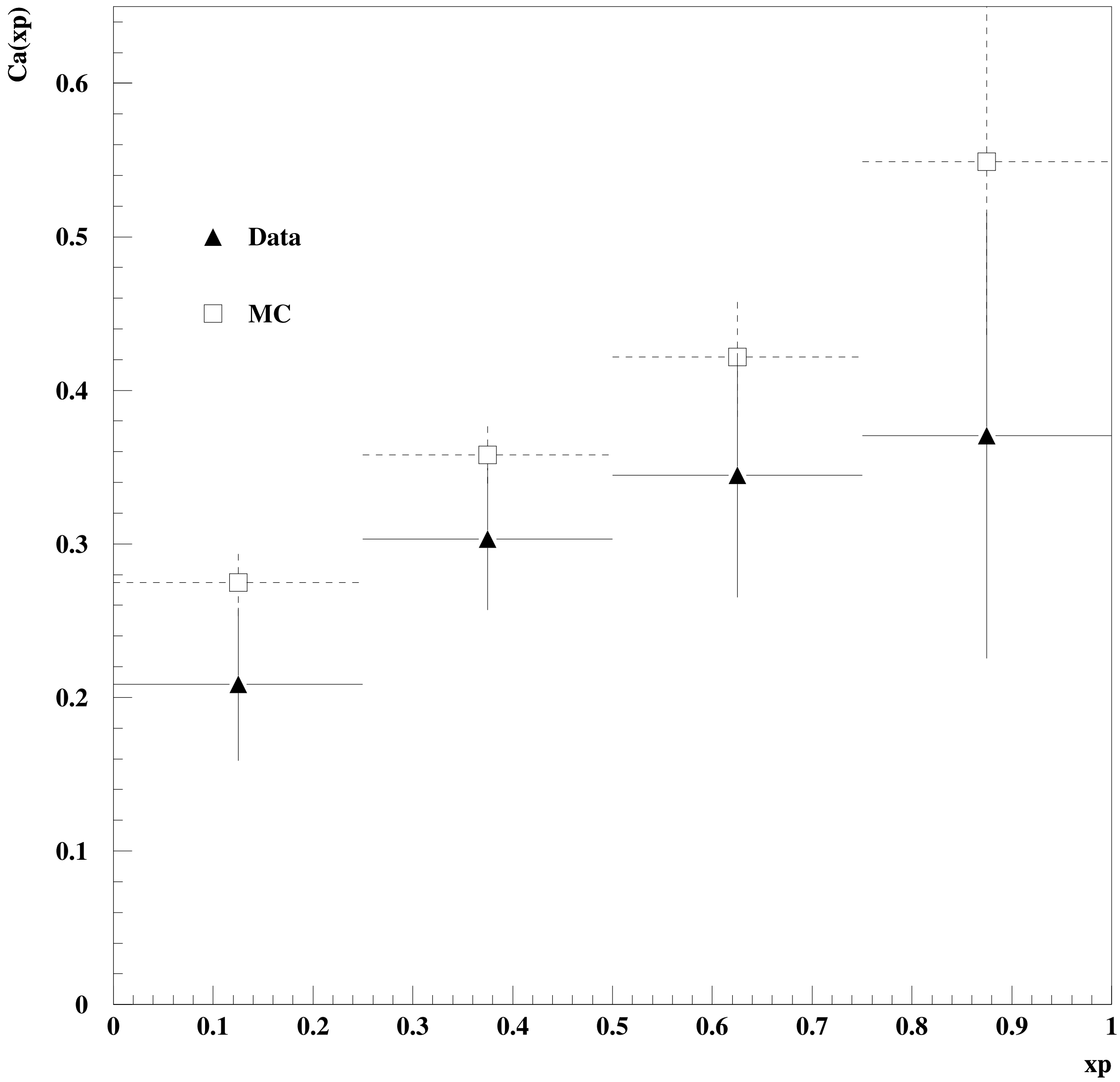

Figure 5.1: Charge alternation as a function of the particle's momentum fraction, x_p

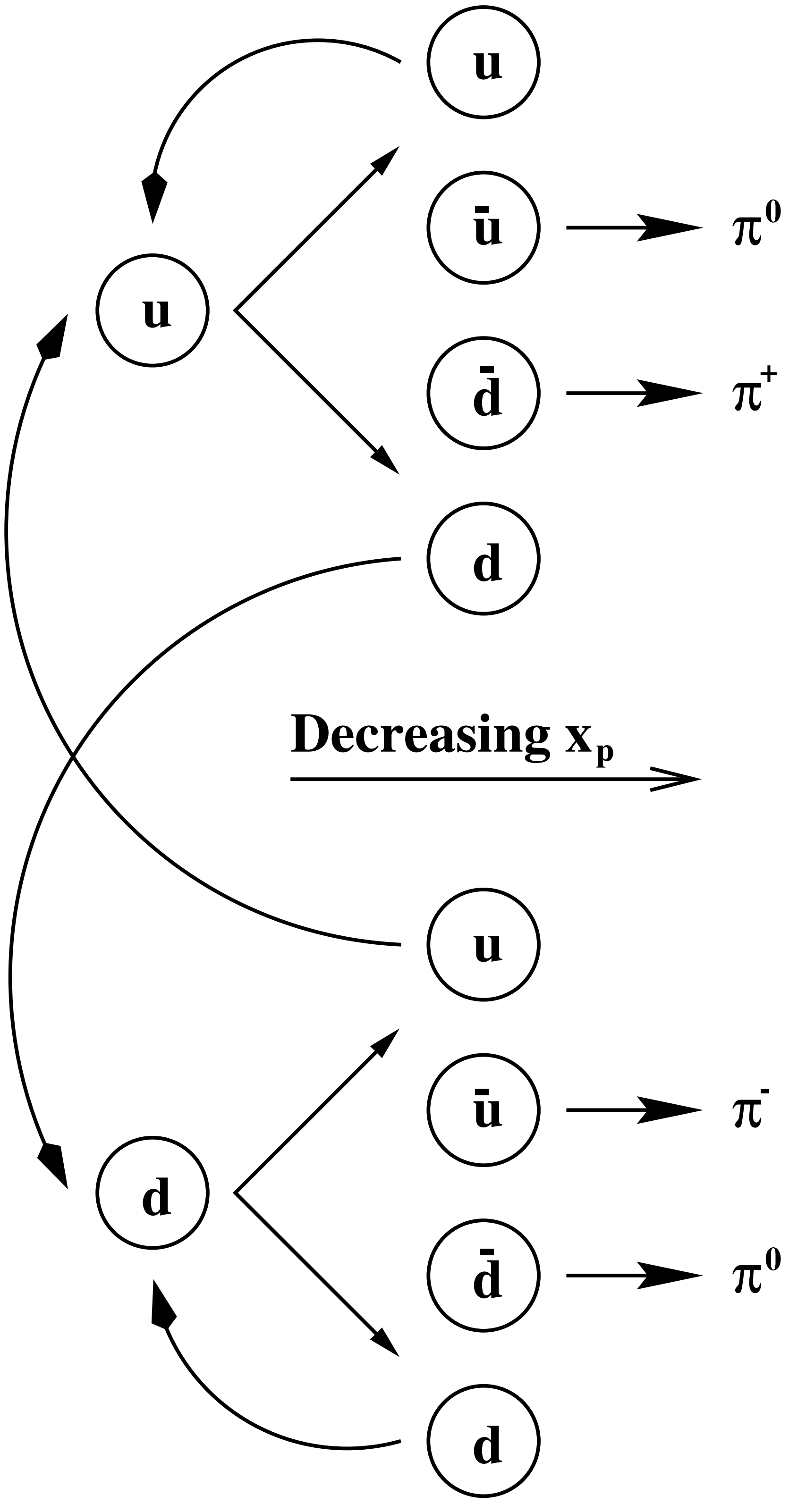

Figure 5.2: A simple model to illustrate charge alternation

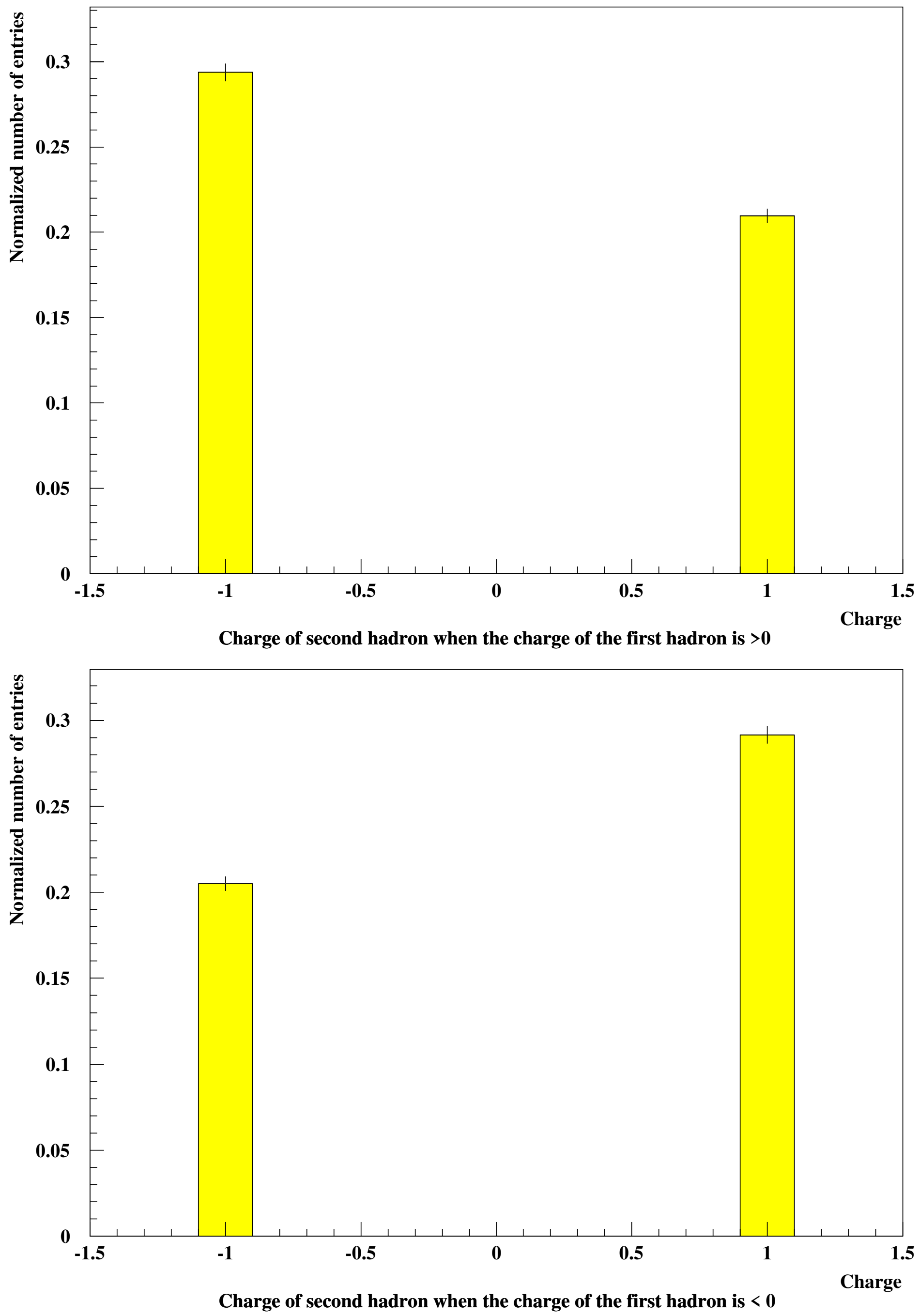

Figure 5.3: Charge alternation in data for n=2 when constraints have been imposed on n=1. Charge of n=1 hadron > 0 in the top diagram; Charge of n=1 hadron < 0 in the bottom diagram.

Generation	Q(n) = 1	Q(n) = -1	Q(n) = 1	Q(n) = -1	C_a	σ_{C_a}
	Q(n-1) = 1	Q(n-1) = 1	Q(n-1) = -1	Q(n-1) = -1		
2	21.7%	29.5%	28.6%	20.2%	0.161	0.022
3	21.6%	28.7%	28.3%	21.4%	0.141	0.019
4	22.1%	27.9%	28.2%	21.8%	0.123	0.017

Table 5.1: Table of C_a values for DATA

Generation	Q(n) = 1	Q(n) = -1	Q(n) = 1	Q(n) = -1	C_a	σ_{C_a}
	Q(n-1) = 1	Q(n-1) = 1	Q(n-1) = -1	Q(n-1) = -1		
2	20.9%	29.4%	29.2%	20.5%	0.168	0.036
3	22.3%	27.9%	28.9%	20.9%	0.152	0.033
4	22.3%	28.8%	27.7%	21.2%	0.134	0.030

Table 5.2: Table of C_a values for the MEPS64 Monte Carlo

Generation	Q(n) = 1	Q(n) = -1	Q(n) = 1	Q(n) = -1	C_a	σ_{C_a}
	Q(n-1) = 1	Q(n-1) = 1	Q(n-1) = -1	Q(n-1) = -1		
2	21.5%	28.6%	29.8%	20.1%	0.171	0.033
3	22.3%	29.1%	28.5%	20.2%	0.136	0.027
4	21.9%	28.6%	28.1%	21.4%	0.131	0.027

Table 5.3: Table of C_a values for the MEAR64 Monte Carlo

5.2.1 Selecting The Charged Tracks

This thesis concerns itself with analysing the interaction between electrons (positrons) and quarks through the exchange of a virtual boson, and trying to associate the pattern of charged hadrons in the final state with the charge of the struck quark. Some method of charged track selection must be used in order to determine which ones to use. It is possible to use a jet-finding algorithm, e.g., the cone algorithm [32], but use of any such jet-finder will introduce biases into the sample according to the definition of the searching algorithm [33]. A far easier method of charged-track selection is to just accept all the particles in a given region, e.g. the current region of a "standard" frame such as the $\gamma^{*}p$ centre-of-mass frame. Other such "standard" frames are the lab frame and the Breit (or $\gamma^{*}q$ CMS) frame. As has already been stated above, we are dealing with the interactions between electrons and quarks and a frame which is also eq oriented would be an ideal choice. The Breit frame is just such a frame and has already been described in section 1.7. Many of the Breit frame's properties have been thoroughly researched in numerous other publications [6], [9], [34], [35] and a more in-depth treatment of the Breit frame may be found elsewhere [36].

After choosing the Breit frame as our reference frame, a natural choice of variable to use to begin to form the weight became clear; this was the dimensionless fragmentation variable, x_p. Its properties have also been well-researched [6], [9] and these have been assumed for the work presented in this thesis. Some of these properties include: the current hemisphere of the Breit frame is approximately half the average charged hadronic multiplicity of an $e^{+}e^{-}$ interaction at an energy of $E^{*} = Q$; x_p is independent of the difficulties involved with any jet or thrust axis determination.

It has been shown (see section 5.2, figure 5.1) that most information is carried by the leading hadron, therefore it would be prudent to emphasise its contribution to the final value of C_j. To achieve this, the fragmentation variable, x_p, can be raised to some power, κ. This power is presumably positive, but this and its magnitude are left for empirical determination by means of a simulated data analysis. The fragmentation variable raised to some power, κ, thus constitutes our initial choice of weight, w_{ij}.

The figures shown in these early sections reflect values of κ and a summation range after the optimising procedures laid out in sections 5.5 through 5.7 have been performed. This does not alter the conclusions drawn in the early days of the analysis, but does yield clearer diagrams.

Now that we have defined a variable which can be used, in part at least, to form the weight, we may construct a C_j histogram. Figure 5.4 shows just such a histogram - the C_j distribution obtained from data (solid histogram) which also shows the Monte Carlo simulation as an overlay (dots). The graph exhibits nodal structure at points corresponding to net charges of -1, 0, and $+1$. The zero bin (the large central spike) is potentially information-rich and is assigned its own special discussion (see the next section, 5.3). Details of the investigations performed using the contents of the zero bin can be found in sections 6.7 and 6.8.

5.3 The zero bin of the C_j histogram

Boson-gluon fusion events tend to have a large angle between the two produced quarks, due mainly to the fact that they have a larger hard sub-process invariant energy than Quark-Parton Model events, for example. This is illustrated in section 6.6, especially figure 6.5, where further details may be found. This process is one of those responsible for producing empty current regions, and thus the zero bin can be used as a handle for determining the gluon density within the proton. The contents of the bin having $C_j \equiv 0.0$ could therefore be quite an important tool for a refined determination of the gluon proportions. Some investigations were performed using this method, and they are described in sections 6.7 and 6.8.

The $C_j \equiv 0.0$ bin can be split into three categories, each of which will be discussed below. The three categories are: those events having no charged tracks in the current region of the Breit frame; those events which have no energy in the current region of the Breit frame; those events which are "accidentally" zero because of equal or cancelling signed momenta.

5.3.1 No Energy In The Current Region Of The Breit Frame

Such events occur when there is no energy in our acceptance (see section 5.6.2) which is indicative of a large split between the two outgoing partons from a BGF event - wide enough in fact that the tracks are pushed completely out of the region of interest. This is typical of an event with an high $\hat{s}$ (see section 6.6) between the two tracks, though it does not *mean* it! It only implies it's a possibility.

5.3.2 No Charged Tracks In The Current Region Of The Breit Frame

This case is where the products of the interaction are wholly neutral in nature, thus providing no possibility of charged-track information from the trackers. Consequently, no momentum measurement can be made using the usual techniques. The neutrals do, however, make their presence known in the electromagnetic calorimeter through showering interactions thus providing an energy measurement. Without the momentum measurement, however, no C_j value can be constructed for that event and hence it is assigned an identically zero value.

5.3.3 Accidental Zeroes

Consider when two hadrons are the result of a hard sub-process; one has charge +q and the other -q. Consider further that their share of the available momentum is also very nearly equal such that their momentum-weighted charge sum is ~ 0.0, or at least, small enough to be counted within the zero bin. In reality, though these fall into the zero-bin, the values are not *identically* zero, and in our investigations, no "accidental" zeroes were found in any event sample. This fact means that we can limit our further investigation to those classes described above in sections 5.3.1 and 5.3.2.

5.4 Charge Tagging Of Quarks

The C_j distribution for uncorrected data has already been shown in figure 5.4. The equivalent distribution for reconstructed MEPS64 Monte Carlo (shown as an overlay to figure 5.4) as well as the distributions from a known struck parton are shown in figure 5.5. It can be seen that, although nodes and spikes are seen with pronounced humps reflecting the parton charge, the distributions are both similar and widespread (note the symmetry in the distributions about $C_j = 0.0$). It can also be seen, in figure 5.5, that the mean values follow a pattern (as already shown in section 3.8.2) in that the distributions tend to be more positive the more positive the initiating parton, but that the spread of the distribution is far larger than the difference between the means. Figure 5.6 shows all five C_j distributions in one three-

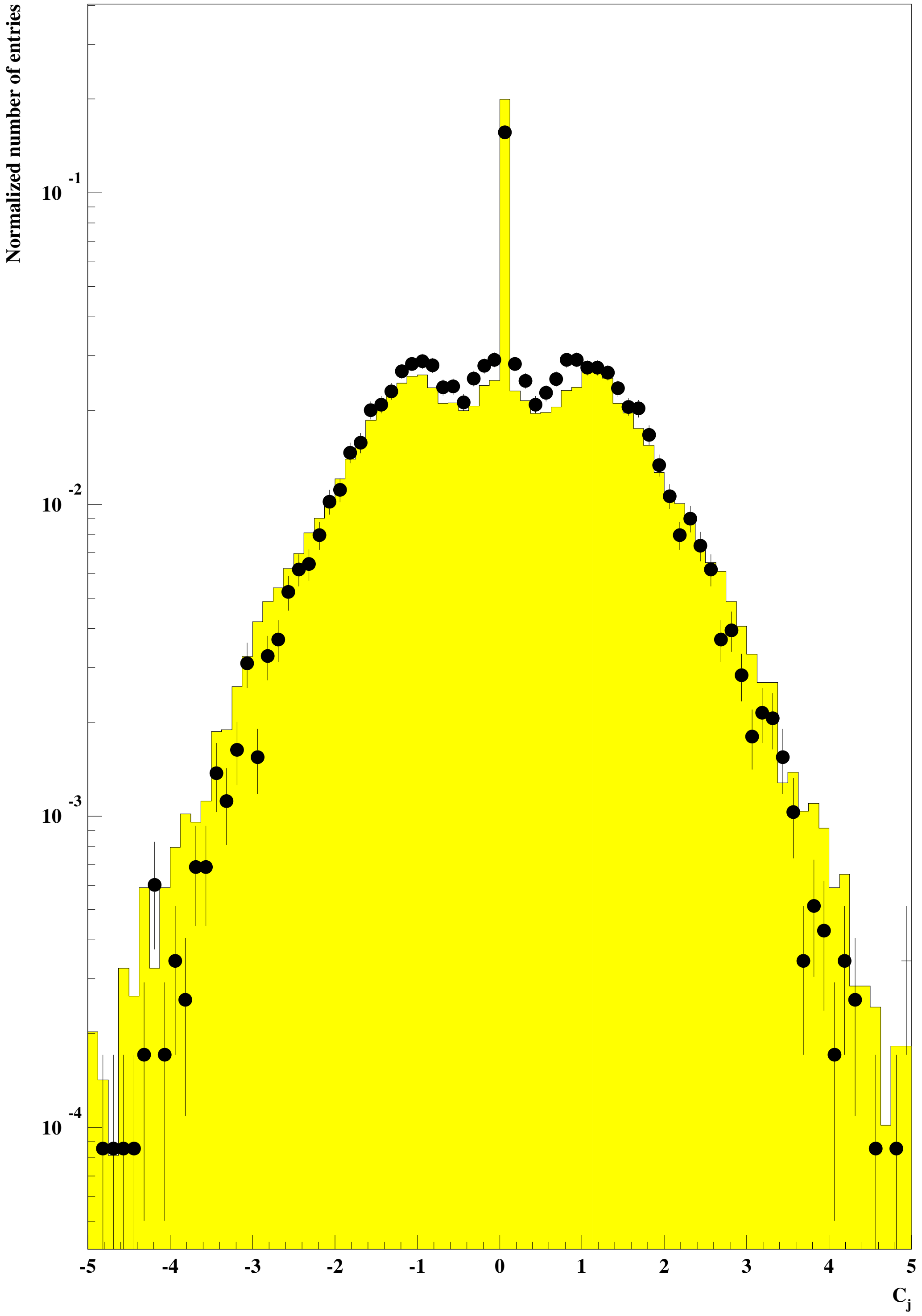

Figure 5.4: A C_j distribution for data overlaid with the C_j distribution for reconstructed MEPS64 Monte Carlo

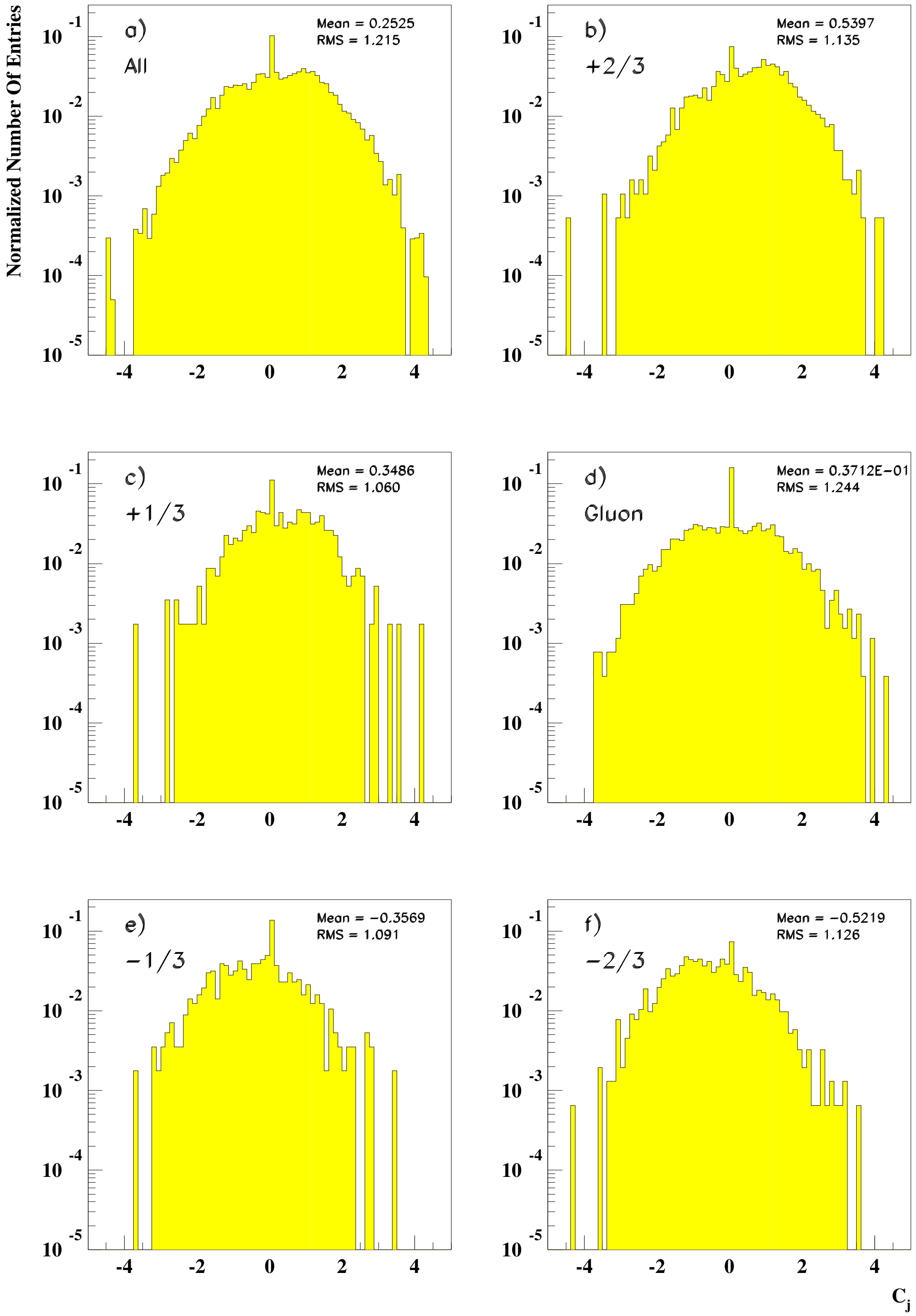

Figure 5.5: A C_j distribution for reconstructed MEPS64 Monte Carlo showing the individual contributions from each parton type

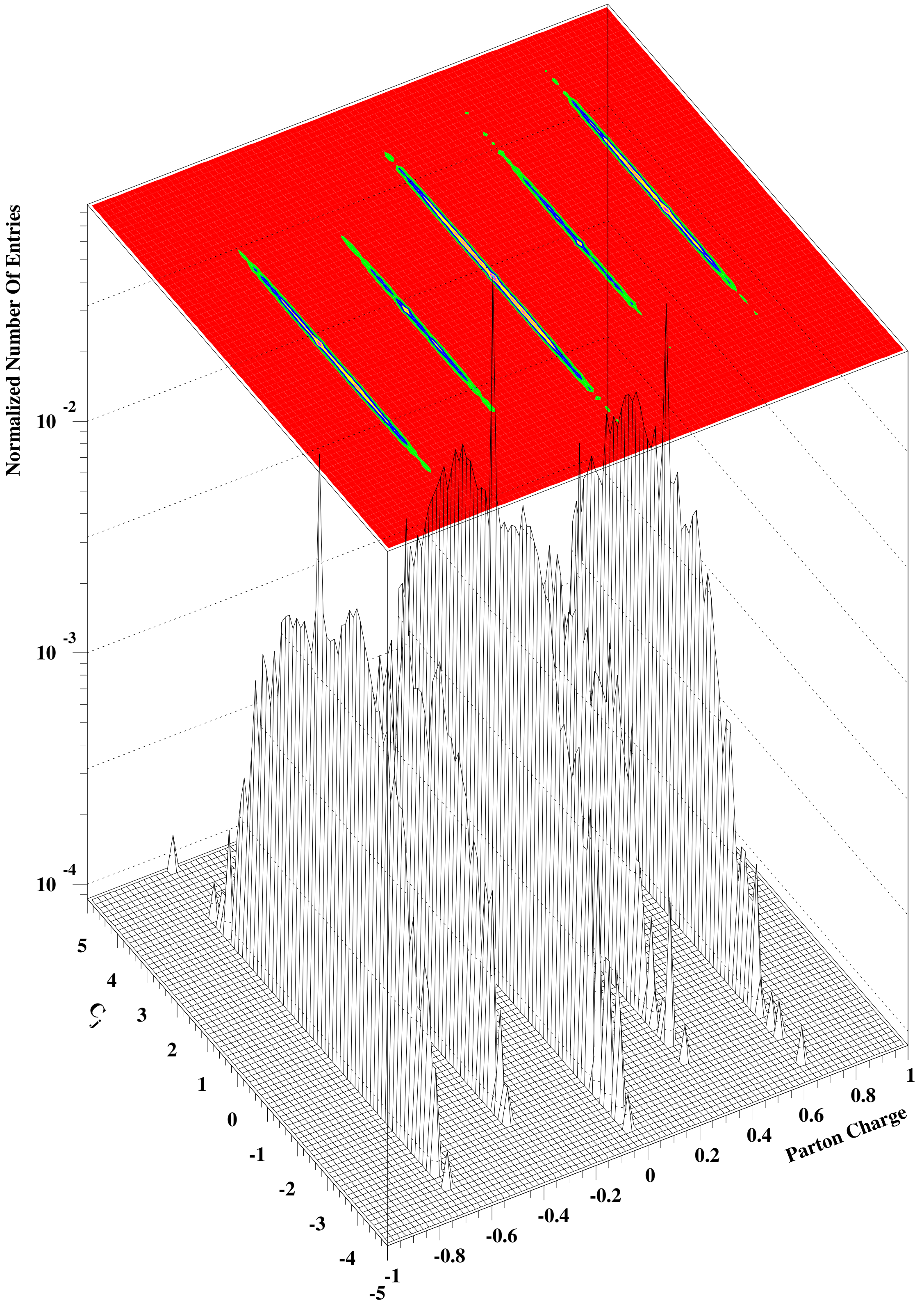

Figure 5.6: A 3-dimensional view of the individual components of the C_j distributions

dimensional plot, overlaid on which is the contour plot for each distribution. This shows more clearly the pattern alluded to when describing figure 5.5. The increased number of events in a given region (reflecting the charge of the initiating parton) can be seen again and clearer in figure 5.7 where the extent of the, e.g., negative portion of the distribution is larger for the -2/3 quarks and that of the positive distribution is largest for the +2/3 quarks. Figures 5.7 and 5.8 also show clearly that the spread in C_j is *far* larger than the difference between the mean values of C_j. These diagrams (figures 5.5, 5.6, 5.7 and 5.8) together effectively rule out charge tagging on an event-by-event basis. However, as has been mentioned above (and mentioned in section 3.8.2), the *mean* values do conform to a pattern. The *mean* values can be used in a statistical approach to determine the *proportion* of a given sample initiated by a certain parton.

The mean alone is only a handle on this information, whereas a quantity similar to the number of standard deviations from zero would be useful also (and also sample-size-independent). Such a quantity is the mean divided by the width, mean/σ, which can be thought of as like the mean divided by the error on that mean, and which is also independent of the width of (and hence any normalisation to) the sample.

It can be seen in figure 5.8 (and which has already been used in section 3.8.2) that the error on the mean, $\sigma/\sqrt{N}$, is much less than the difference between the means and thus differentiation between the parton types is possible. The error bars are so small, in fact, that they cannot be seen on this plot.

5.5 Defining The Tuning Parameters

C_j is to be a quantity calculated by an algorithm *for each event* related to the scattered quark in that event. Experimentally, on data and also in Monte Carlo studies, we find that in fact, and not unexpectedly, the value of C_j has considerable hadronisation scatter (as shown already in figures 5.4, 5.5, 5.6 and 5.7). In order to obtain a better quantity with which to fine-tune the definition, we must therefore fall back on the distribution of the mean value of C_j for all of the events in a given sample, $< C_j >$, characterised by a mean and a spread, σ. The fluctuation in the value of C_j prevents the tagging of individual events, but we can use a *distribution* of C_j to perform a statistical tagging (see sections 5.9 *et seq.* for a description).

To obtain a "signal", or a measure of significance for our value of C_j, we take the

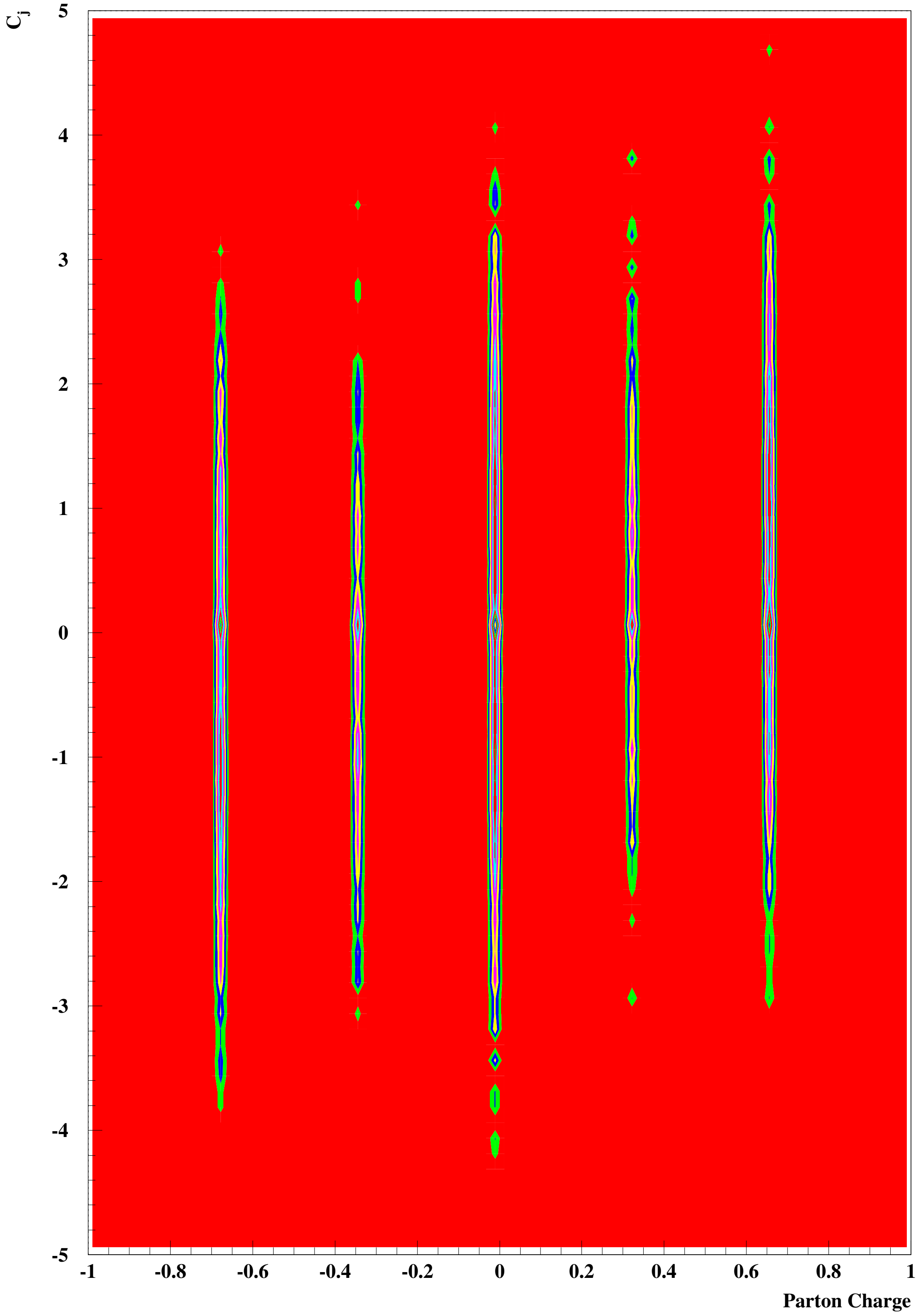

Figure 5.7: A 2-dimensional projection illustrating the spread of the C_j

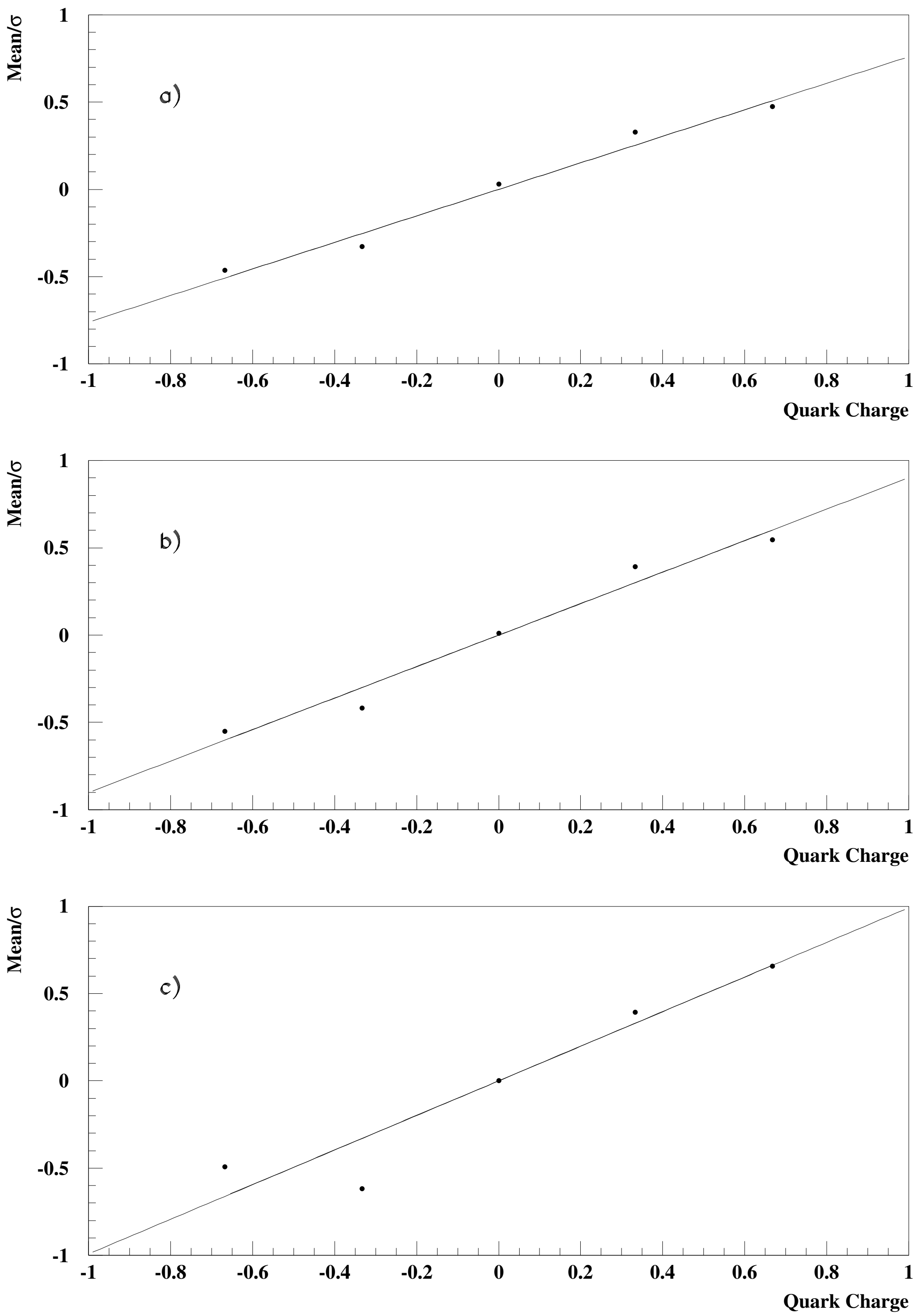

Figure 5.8: The mean of the partonic C_j^m distributions divided by the spread, mean/σ for MEPS (a), MEAR (b) and HERWIG (c) Monte Carlo models. The error bars are shown on this plot but are too small to be seen.

ratio of the mean to the standard deviation, Mean/σ. This provides us with a quantity that we can optimize or "tune" by varying parameters. In simulated data, it is possible to construct a C_j for struck partons of a particular charge type, termed C_j^m, where m can be +2/3, +1/3, 0 (for a gluon), -1/3 or -2/3. In Monte Carlo simulations, the incoming target particle entering into the hard sub-process is well known, and it is the identification of this particle which we are trying to deduce in this analysis. We have assigned a charge type of zero to gluons (as they are electrically neutral) to signify boson-gluon fusion events. It is expected that the C_j^m signal for a +2/3 charge scattered quark will be more positive than that for one of charge +1/3, a gluon should give rise to a distribution symmetric about $C_j = 0$, and that the -2/3 charge scattered quark will have the most negative value of C_j^m. This is indeed observed as shown in figure 5.8. A "saturation effect" is clearly visible in figure 5.8, *i.e.*, the variation of $< C_j >$ with parton charge is *not* linear. Rather, $< C_j >$ plateaus at both the most positive and most negative quark charges. This behaviour is seen in all of the Monte Carlo simulations apart from HERWIG 5.8d and it was this method therefore that was the basis for our investigations into the HERWIG Monte Carlo, detailed in section 3.8.2.

The intention is to be able to say, at least on a statistical level, what the scattered parton's charge was purely from looking at the hadronic final state and the associated C_j value. To do this, it is necessary to be able to differentiate as much as possible the C_j^m values for positive and negative values of m. This can be achieved by maximising the difference between the most positive and most negative "signals", *i.e.*, the +2/3 and -2/3 charge quarks. This maximisation is achieved by varying the statistical weight, w_{ij} in some way. There are two obvious ways in which this weight can be altered; by varying the fragmentation variable's exponent, κ, and by changing our acceptance in the Breit Frame, *i.e.*, the extent of the summation, i.

5.6 Tuning The Parameters

The values of $< C_j^{+\frac{2}{3}} >$ and $< C_j^{-\frac{2}{3}} >$ (for +2/3 and -2/3 charged quarks respectively) are obtained for a given Monte Carlo sample of DIS events passing the cuts and selection criteria described in chapter 4, and the difference between them, $(< C_j^{+\frac{2}{3}} > - < C_j^{-\frac{2}{3}} >)$ is then varied to find the maximum difference as a function of the track weighting algorithm, w_{ij}. Experimental details of how this signal is maximised are presented below in sections 5.6.1 and 5.6.2.

5.6.1 Changing The Exponent, κ

By varying the exponent, κ, of the fragmentation variable, x_p, it is possible to alter the relative importance of the leading hadron compared to the the next-to-leading hadron (and all subsequent hadrons) and thus affect the value of C_j. A range of κ was investigated and a maximum was observed. This range was investigated further in smaller steps, and $< C_j^m >$ for each parton type obtained. Section 5.6 described the intended method of investigation, using a difference of mean C_j values. This difference, $(< C_j^{+\frac{2}{3}} > - < C_j^{-\frac{2}{3}} >)$, is plotted as a function of the exponent κ and the results for various Monte Carlo models are shown in figure 5.9.

The optimal value of κ was deemed to be 0.4 and this value was selected for all subsequent analyses.

5.6.2 Changing the acceptance

The variable C_j can also be tuned in another way, and that is to change the acceptance of hadrons in the Breit frame. The initial choice was to accept only hadrons in the current hemisphere of the Breit frame, but it has been shown [8] that there is some degree of contamination from the target hemisphere into the current hemisphere, and also a similar amount of migration from the current into the target hemisphere. This migration and contamination is at the $<5\%$ level at high Q^2 and at the $\simeq$ 20 - 25% level at low Q^2 [8]. The method of Monte Carlo correction used in this thesis compensates for migration between hemispheres caused by resolution effects. It is instructive to investigate the sensitivity of C_j^m, and in particular the variation of the difference, $(< C_j^{+\frac{2}{3}} > - < C_j^{-\frac{2}{3}} >)$, with the acceptance region in the Breit frame. A method similar to that in section 5.6.1 was used for this investigation, and "Cut-off" track values from $-1.0 \leq \cos\theta_B \leq 1.0$ in steps of 0.1 were used. The results are shown in figure 5.10. Note that the plot shown here in figure 5.10 is a *cumulative* plot, unlike that of figure 5.9.

The optimal value was not so easy to choose as in section 5.6.1 due to the lack of a distinct peak, but a plateau was observed in all the Monte Carlo simulated data up to $\cos\theta \approx 0.4$, and consequently this value was chosen as our new acceptance region. It is clear that the value is insensitive to the acceptance region in the Breit frame.

To ensure that there was no change to the optimal value of κ, given this new ac-

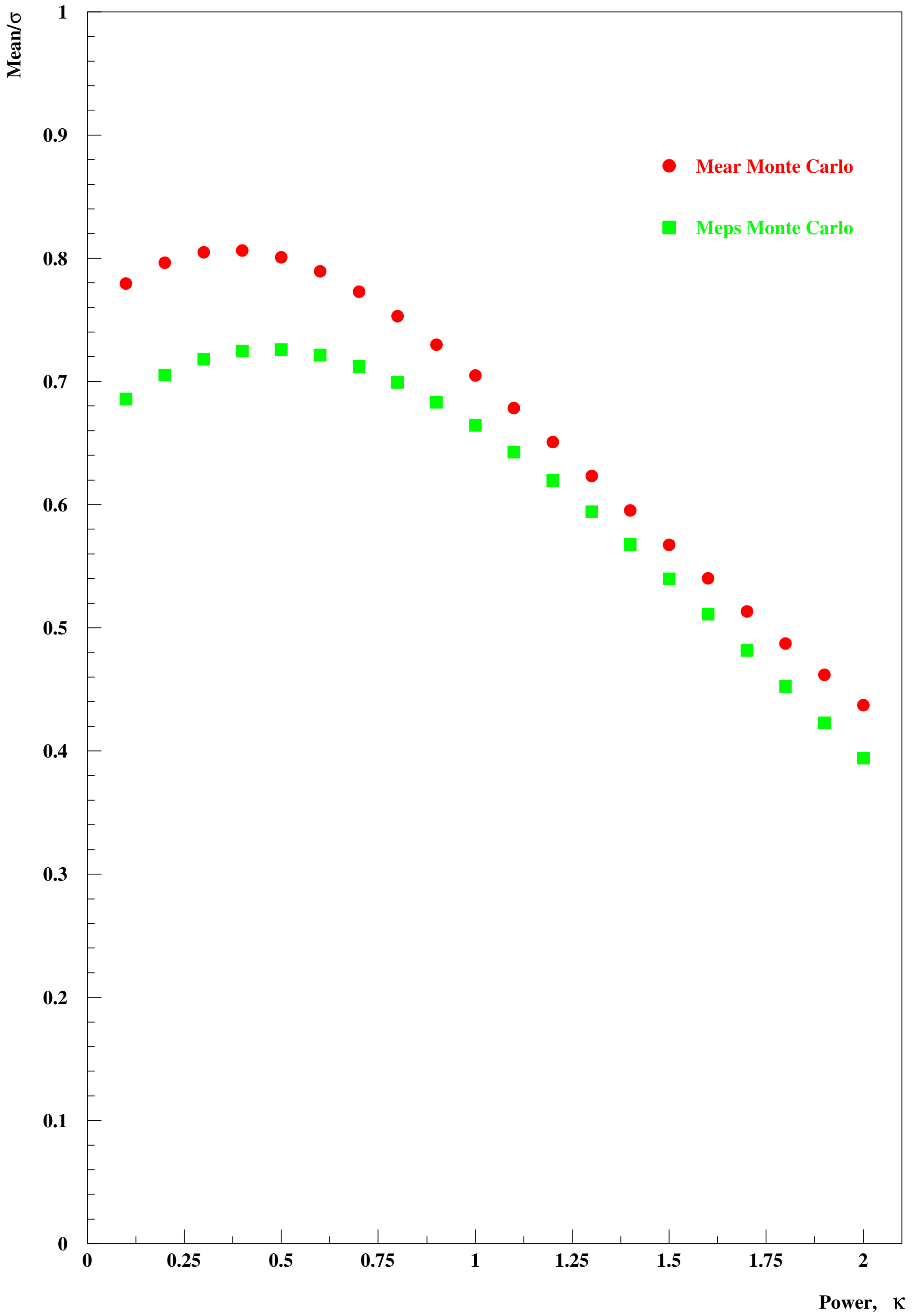

Figure 5.9: $(C_j^{+\frac{2}{3}} - C_j^{-\frac{2}{3}})$ as a function of κ

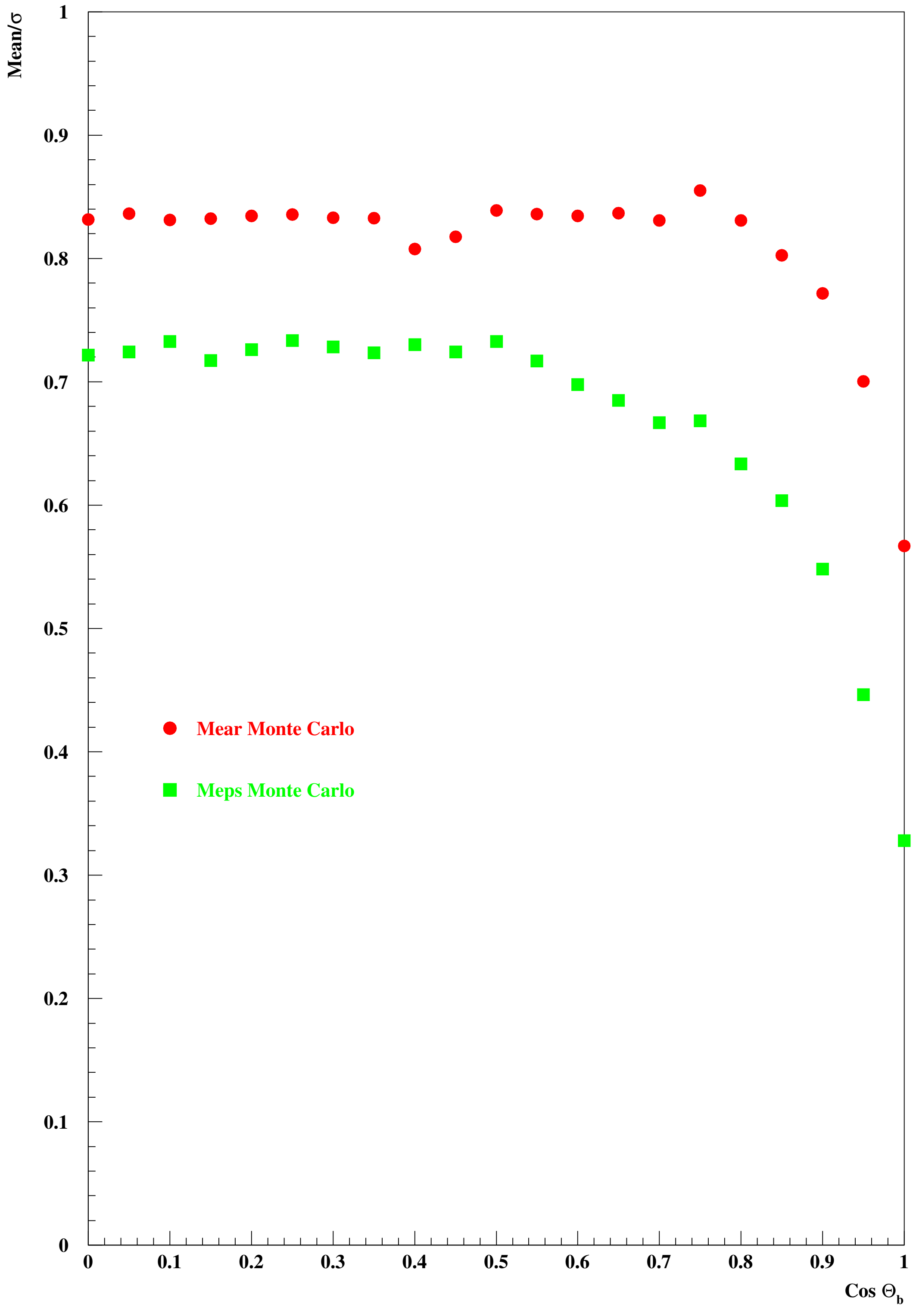

Figure 5.10: $(C_j^{+\frac{2}{3}} - C_j^{-\frac{2}{3}})$ as a function of $cos\theta_B$

ceptance region, the tests in section 5.6.1 were repeated. No significant change was found, and the optimal values for both κ and $\cos\theta_B$ were fixed at 0.4 for the subsequent analysis.

5.6.3 Changing The Fragmentation Variable

It is possible that our choice of variable for the weight is not the best one, and that another might give rise to a clearer or larger signal. Another candidate for the variable is known as z in e^+e^- analyses. We would then have $w_{ij} = z^\kappa$ and

$$z = \left(\frac{(E+p_z)_h}{(E+p_z)_q}\right) \tag{5.3}$$

which contains directional as well as kinematic information. Another advantage is that this ratio is Lorentz-invariant to the $\pm z$ boosts. Similar investigations to sections 5.6.1 and 5.6.2 were performed using this weight, but it was found that the signal obtained was not as large as when the fragmentation variable was used. Figure 5.11 illustrates this point for the variation with the exponent, κ, in the MEAR64 Monte Carlo only as this Monte Carlo gives the larger "signal" of MEAR or MEPS. It was decided, therefore, to remain with our initial choice of fragmentation variable, x_p.

5.7 Normalisation Of C_j

The C_j thus far discussed and obtained from the above definitions is sensitive to the different multiplicities and kinematics of each event. It is highly desirable to have a distribution of C_j which would be insensitive to such phase space differences between low and high Q^2 data; between data and Monte Carlo simulations. This calls for some kind of normalisation.

An event with one very high momentum track, for example, should not be treated with the same significance as an event in which there is one low momentum track. The assertion in the discussion of section 5.2 is that high momenta tracks carry more information about the struck quark than tracks of low momentum. This has been borne out experimentally by the tests of section 5.6 where a positive value of

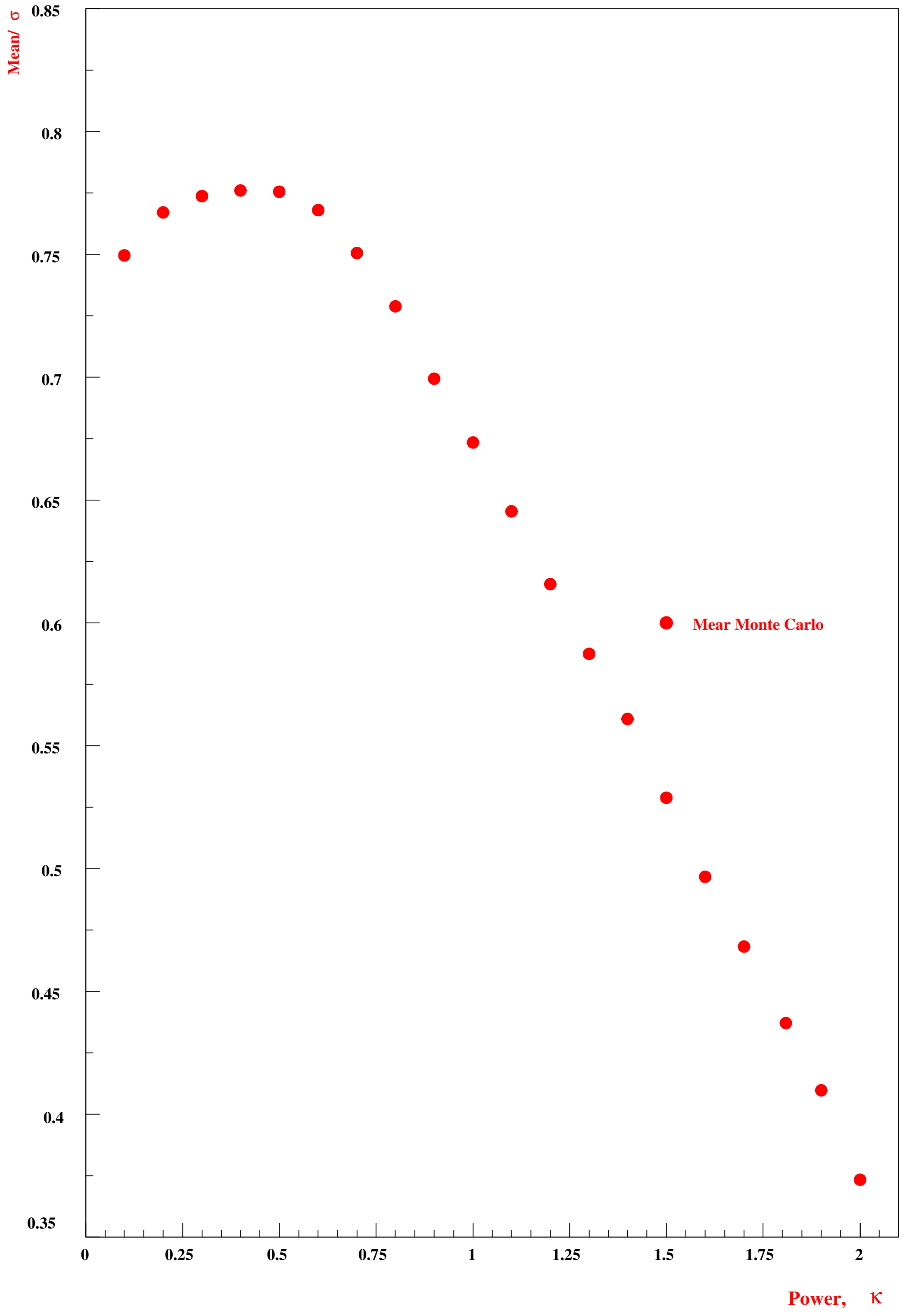

Figure 5.11: The "signal" when using a different weight as modelled by the MEAR64 Monte Carlo

κ was found to give the best "signal". A $\kappa > 0$ means that the particle with the largest momentum contributes the most to the signal. Therefore C_j, as defined in equation 5.2, actually does emphasise the harder hadrons, but does not account for the differences in track multiplicity or kinematics. Consequently, without some kind of normalisation, this method cannot be used.

The contribution that the weighted momentum gives when charge is not considered can be used to calculate, on an event by event basis, the mean weight of each event. This track normalisation may be expressed as shown in equation 5.4.

$$< w_{event} >_j = \sum_i \left(\frac{1}{N_{tracks}}\right) w_{i,j} \tag{5.4}$$

This can then be used to normalise the contribution to C_j and will redress the (unintentional) bias caused by the different charge multiplicities in each event. This quantity however, treats all events equally and just scales the quantity C_j by the same amount. Consider two events, both with one positive track, but one with $x_i = 0.9$ and the other with $x_i = 0.1$. Both of these events would yield a C_j value of 1. However, it is known that the contribution from the harder of the two tracks is the most significant, and hence this initial approach cannot be used as this method does not fit our requirements.

The next candidate to normalise with is merely to take the average of all the event weights in the sample, as calculated in equation 5.4. This can be written as shown in equation 5.5, or expressed in full in equation 5.6.

$$< w_{sample} > = \sum_j \left(\frac{1}{N_{events}}\right) < w_{event} >_j \tag{5.5}$$

$$< w_{sample} > = \sum_j \left(\frac{1}{N_{events}}\right) \left(\sum_i \left(\frac{1}{N_{tracks}}\right) w_{i,j}\right) \tag{5.6}$$

Though this mean weight for the sample, $< w_{sample} >$, is a dimensionless number, it incorporates all the pertinent kinematic information about any given sample within it. The behaviour of $< w_{sample} >$ with Q^2 is tabulated below in table 5.4.

It can be seen that, as we move towards higher and higher values of Q^2 and multiplicities increase, the mean value of $< w_{sample} >$ decreases as expected. This effect is observed in both data and Monte Carlo with disagreements at the less than 10%

Data Type	Range in Q^2	Mean sample weight, $< w_{sample} >$		
		Data	MEPS64	MEAR64
Low Q^2	$12.0 \leq Q^2 < 25.0$	0.452	0.475	0.475
Low Q^2	$25.0 \leq Q^2 < 48.0$	0.450	0.469	0.479
Low Q^2	$48.0 \leq Q^2 < 65.0$	0.440	0.461	0.474
Low Q^2	$65.0 \leq Q^2 < 80.0$	0.433	0.458	0.471
High Q^2	$100.0 \leq Q^2 < 350.0$	0.421	0.397	0.418
High Q^2	$350.0 \leq Q^2 < 800.0$	0.369	0.353	0.375
High Q^2	$800.0 \leq Q^2 < 2000.0$	0.337	0.311	0.333
High Q^2	$Q^2 > 2000$	0.283	0.265	0.282

Table 5.4: A table showing how the mean value of w_{sample} changes as a function of Q^2 for both low and high Q^2 data and several Monte Carlo simulations

level. Errors on the numbers range from 4% in the more populated, low Q^2 bins to 16% in the higher Q^2 bins.

This, then, is the method chosen as the normalisation factor, and all results presented in this thesis have incorporated this “sample” normalisation.

5.8 Valence Quark Effects

At high enough x_{Bj}, valence quarks should dominate the cross-section [4]. Furthermore, the interaction cross-section varies as the square of the electromagnetic charge and thus an up quark is four times more likely to interact with a photon than a down quark. Coupled with the fact that there are also two up quarks and only one down quark, the cross-section for eu is (naïvely) eight times that of ed. Such a dominance of up quarks should manifest itself as a rise in the mean value of C_j as $x_{Bj} \to 1$. Similar behaviour should also be seen in Q^2 through the x-dependence of Q^2 at fixed W^2.

This behaviour is indeed seen in data and is also modelled very well in the Monte Carlo simulation as shown in figure 5.12 for x_{Bj} and figure 5.13 for Q^2. The dependence on Q^2 is not so strong as that on x_{Bjorken}, as is expected, because any variation in C_j with Q^2 is only there through the dependence of Q^2 on x_{Bjorken}. This bolsters confidence in the chosen method.

5.9 How To Use C_j

The intention is to use a Monte Carlo model (several, if possible) at the generator level to find the shape of the C_j^m distributions for *each parton-charge species*. These ideal, parton-specific distributions will be referred to as "theoretical" C_j^m distributions, and summed proportions of each are fitted to a sample of data in order to determine the corresponding proportions of each parton type in the proton.

5.10 Obtaining Theoretical Parton Distributions

In order to obtain useful theoretical distributions, a large Monte Carlo sample is required, *i.e.*, one in which statistical error bars are small compared to those of the data samples, in order that we do not fit data to statistical fluctuations. Unfortunately, the volume of 1994 Monte Carlo is insufficient to provide us naturally with the smooth distributions that we require. Consequently, some means of alleviating this shortcoming and making full use of the available Monte Carlo must be found.

We expect the fragmentation of quarks to be symmetric under charge conjugation, and evidence confirms (section 3.8.3) that this symmetry is present (in the limit of infinite statistics) in the Monte Carlo models used in this analysis. That is, we expect that the distribution of positive and negative fragmented hadrons from, e.g., $+\frac{2}{3}$ quarks to be identical to the distribution of negative and positive hadrons from $-\frac{2}{3}$ quarks, respectively. This rationale led to the method used to illustrate the problem with HERWIG 5.8d discussed in section 3.8.2. It is thus possible to reverse the $C_j^{-\frac{2}{3}}$ distribution and add it to the $C_j^{+\frac{2}{3}}$ distribution, similarly the $C_j^{\pm\frac{1}{3}}$ quark distributions and gluon distribution (which is mirror-reversed and added to itself). The three resulting distributions are then smoothed using the HBOOK spline fitting routine [37] resulting in three smoothed distributions for $C_j^{\frac{2}{3}}$, $C_j^{\frac{1}{3}}$ and C_j^0.

In order to obtain a smooth distribution, a spline fit can be used so that statistical fluctuations can be averaged over. The discontinuity of the $C_j^m \equiv 0.0$ bin would affect the spline fit adversely, and thus all instances of $C_j^m \equiv 0.0$ are removed.

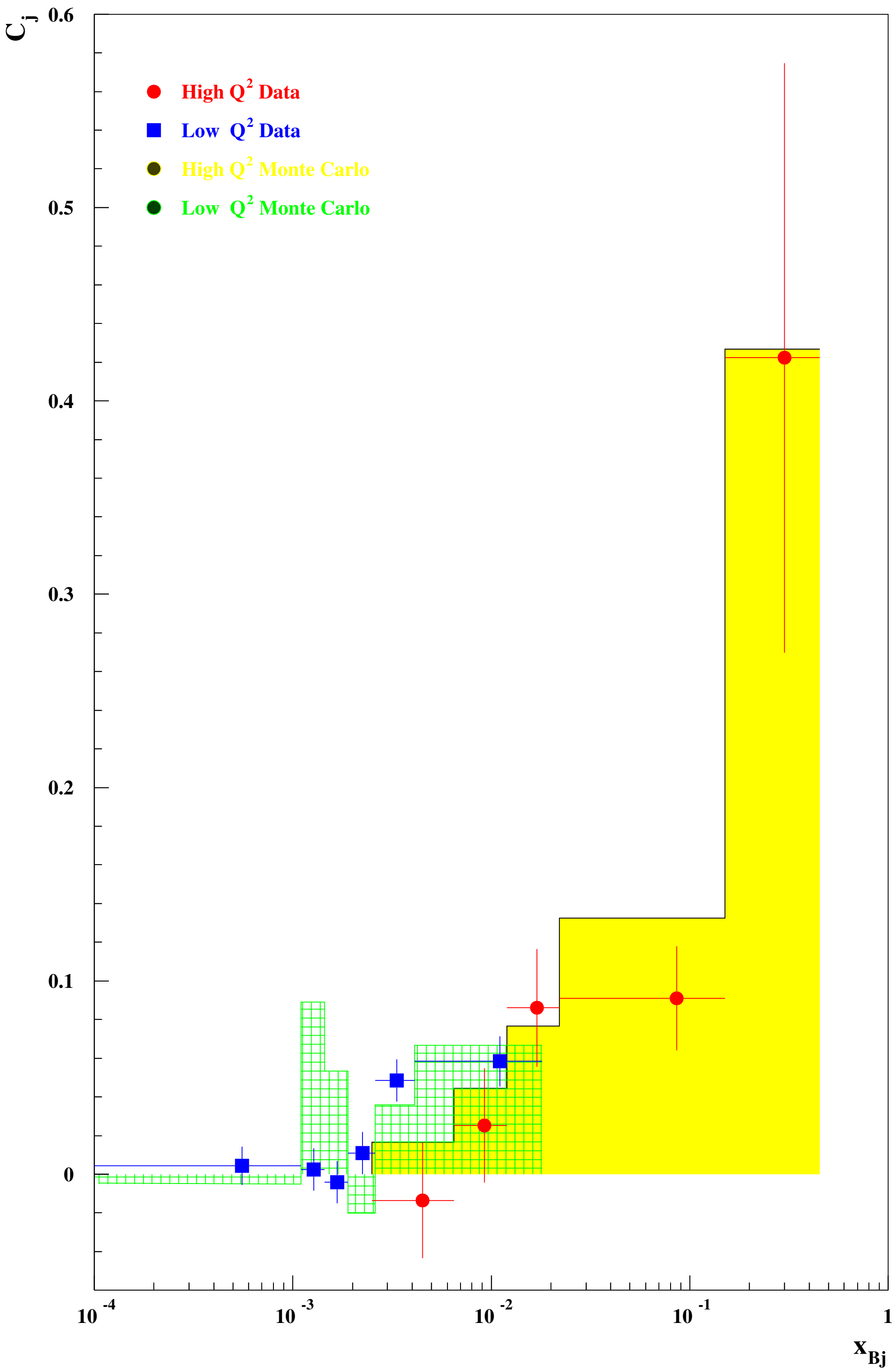

Figure 5.12: Variation in $< C_j >$ as a function of $x_{Bjorken}$

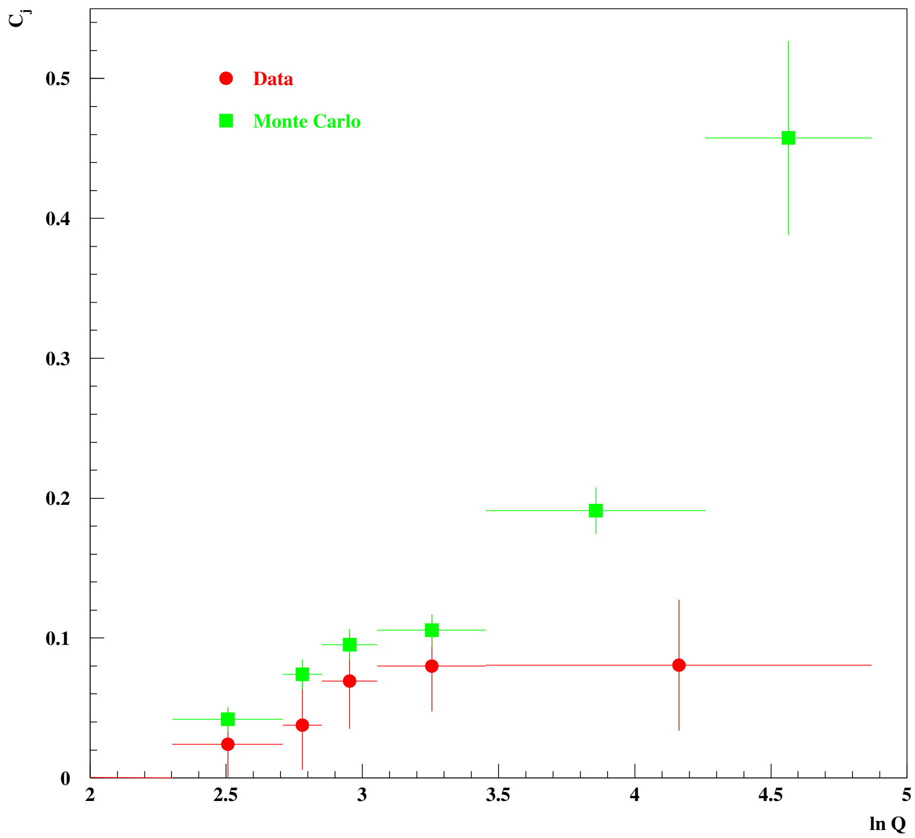

Figure 5.13: Variation in $< C_j >$ as a function of Q^2 at high Q^2

These $C_j^m \equiv 0.0$ events are a special case and are either events in which there is no energy in the current region, or in which there is no charge multiplicity (see section 5.3), and they are put back in to the distribution at a later stage, *i.e.*, after the non-zero contributions to C_j have been smoothed. A method which utilises the breakdown of the $C_j^m \equiv 0.0$ bin in the analysis is described in sections 6.7 and 6.8.

Once again, we invoke symmetry under reversal of charge conjugation to reclaim the negatively charged quark distributions, $C_j^{\frac{2}{3}}$ and $C_j^{\frac{1}{3}}$, and after re-inserting the $C_j \equiv 0.0$ bin into each histogram we are left with five smoothed theoretical parton distributions. These five distributions can then be used, in varying proportions, to fit to one data distribution. Figures 5.14 and 5.15 show the distributions for the MEAR64 Monte Carlo both before and after smoothing, as well as the χ^2/bin associated with the smoothing process. It can be seen that the χ^2/bin is small in nearly all bins.

5.10.1 The Make-up Of A Proton

A proton is known to consist of two up and one down valence quarks. It also has a contribution from sea quarks, *i.e.*, those that owe their existence to quark pair production. Completing the line-up of "constituents" are the gluons, similarly radiated off colour-charged valence quarks. All of these axioms can be stated as the following set of integrals (equation 5.7) over constituent distribution functions, called sum rules, where the subscript v refers to valence quark distribution functions and the subscript s refers to sea quark and sea antiquark distribution functions. The sum rules simply state that the integrals of the parton distribution functions over all x yield the constituents of the proton, *i.e.*, two valence up quarks and one valence down quark.

$$\begin{aligned} \int_0^1 (u_v(x) + (u_s(x) - \bar{u}_s(x)))dx &= 2 \\ \int_0^1 (d_v(x) + (d_s(x) - \bar{d}_s(x)))dx &= 1 \\ \int_0^1 (s_s(x) - \bar{s}_s(x))dx &= 0 \end{aligned} \tag{5.7}$$

We limit our consideration to light quarks only, the contribution from heavier quarks being assumed to be negligible. We also make the assumption that the strange

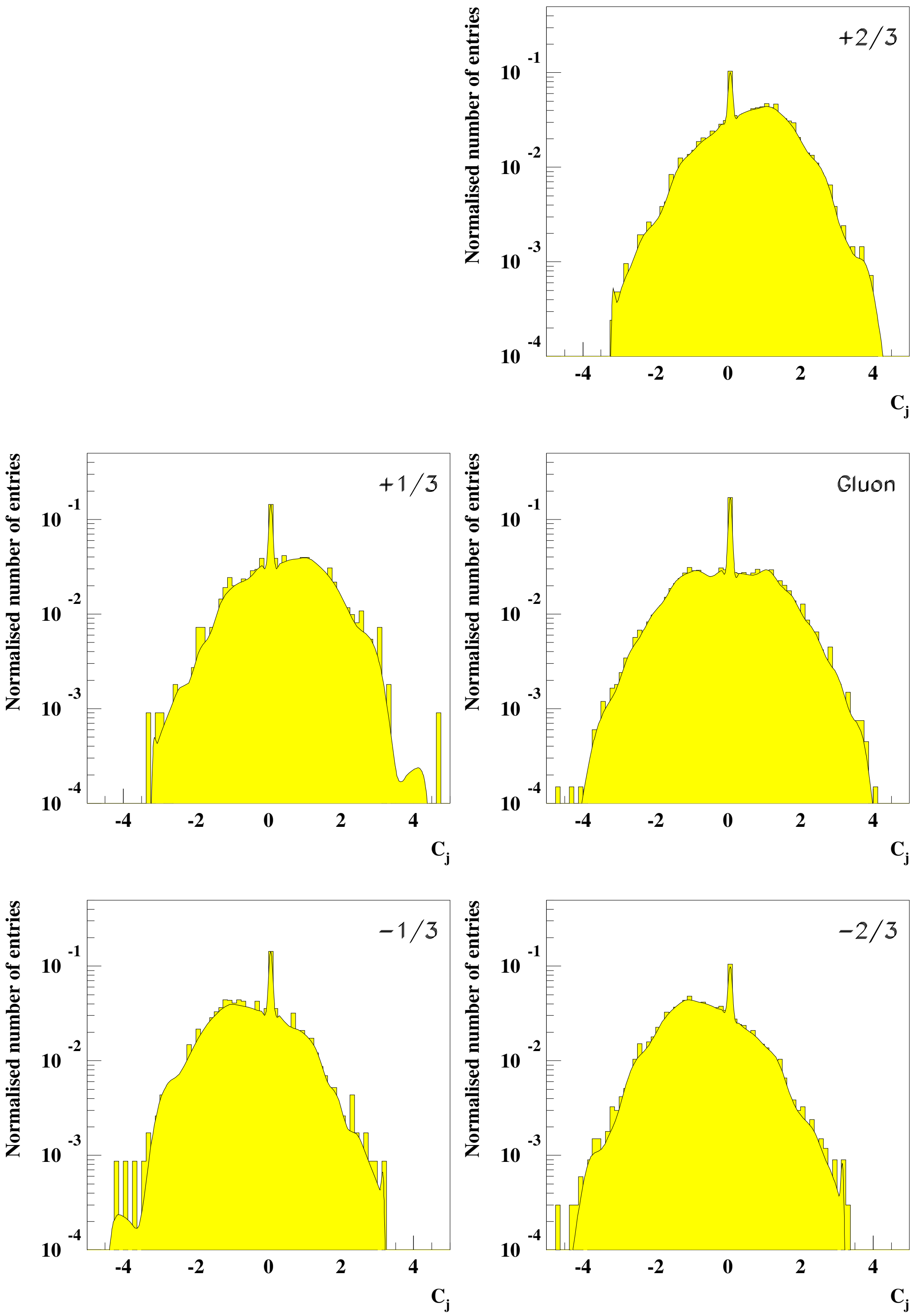

Figure 5.14: Comparison of MEAR64 Monte Carlo (solid histogram) with the smoothed output (line overlay) after charge symmetry reversal

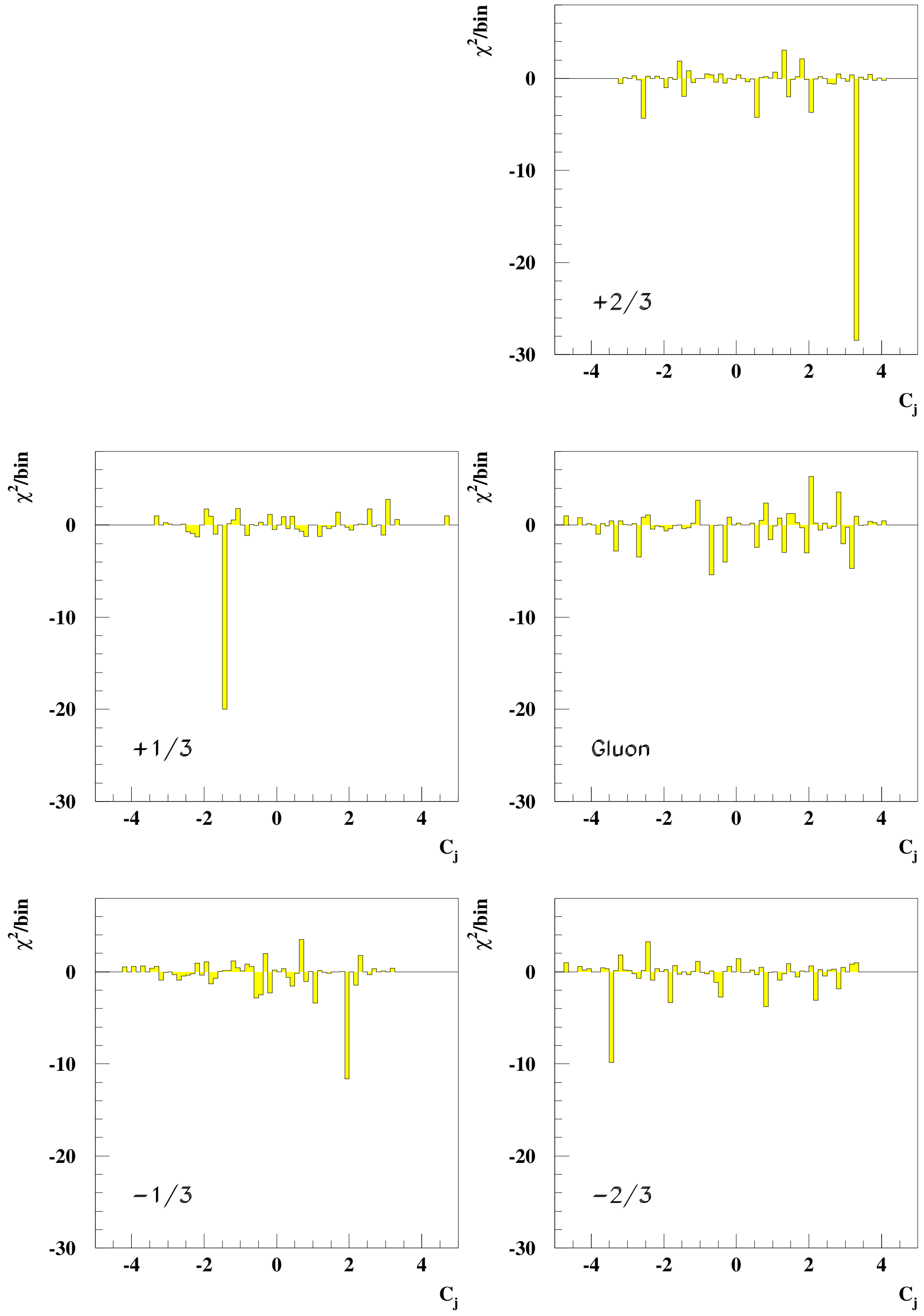

Figure 5.15: The χ^2/bin for the smoothing process for each parton type

quarks appear with the same frequency in the sea as the down quarks *were they to be massless*, but that they are penalised for being considerably more massive than the down quarks. One can introduce a *strangeness suppression factor*, ϵ_s, for the strange quarks so that both they and the down quarks may be treated in the same fashion. The parton distribution functions for the strange quarks may be expressed as written in equations 5.8 where the strangeness suppression factor, ϵ_s, is obtained from the input parameters to the Monte Carlo and has value $\epsilon_s = 0.24$.

$$\begin{aligned} s_s(x) &= \epsilon_s d_s(x) \\ \bar{s}_s(x) &= \epsilon_s \bar{d}_s(x) \end{aligned} \tag{5.8}$$

What this factor effectively does is to *enhance* the $q = \pm\frac{1}{3}$ sea quark contribution by subsuming the strange quark contribution.

We are now in a position to be able to assess the contribution to deep inelastic scattering from each parton charge-type in the classes of Sea Quarks (Q_s), Valence Quarks (Q_v) and Gluons (G). The proportional contribution to each class may be parameterised by the two unknowns, a and b, for the Sea quarks and Valence quarks respectively[2]. The contribution to the cross-section is parameterised to the square of the parton's charge and to the amount of each parton in the proton.

Consider the contribution of the up quark to the photon-interaction cross-section. The photon couples as the square of the charge and thus the interaction of a photon with an u quark is four times more probable, or alternately, happens four times as often as that of the down quark. Couple this with the fact that there are twice as many valence up quarks as there are valence down quarks in the proton, and one finds that the interaction cross-section for γu_v is *eight* times that of γd_v. This is where the $8b$ and $1b$ in table 5.5 come from, respectively. For the sea quarks, u and d appear in equal proportion (assuming they have equal masses) and thus the γu_s interaction cross-section is only four times that of the valence γd_v. The γd_s interaction is effectively enhanced by subsuming the γs_s contribution by means of the strangeness suppression factor, ϵ_s, and hence the sea interaction cross-sections are $4a$ and $(1 + \epsilon_s)a$ for the $\gamma q^{\frac{2}{3}}$ and $\gamma q^{\frac{1}{3}}$ respectively. It is from such considerations that table 5.5 is created.

[2] a and b are in fact parton distribution functions, but it is not necessary to explicitly use them at this point

If we divide the amount of each type by the total number, N, then we obtain the *proportional contribution* to the proton from each of the three classes, Q_s, Q_v and Gluons, as shown in table 5.5.

Class Of Proton	Parton Charge Types					Sum
Contribution	$P(+\frac{2}{3})$	$P(+\frac{1}{3})$	$P(0)$	$P(-\frac{1}{3})$	$P(-\frac{2}{3})$	Total
Q_s	4a	(1+ϵ_s)a	0	(1+ϵ_s)a	4a	(10 + 2ϵ_s)a
Q_v	8b	0	0	1b	0	9b
G	0	0	$1 - Q_v - Q_s$	0	0	$1 - Q_v - Q_s$

Table 5.5: The *proportional contribution* from each parton type to the different classes of gluons and sea and valence quarks

The total contribution to the proton's make-up by each class can then be written as set forth in equations 5.9.

$$\begin{aligned} Q_s &= (10 + 2\epsilon_s)a \\ Q_v &= 9b \\ G &= 1 - (10 + 2\epsilon_s)a - 9b = 1 - Q_s - Q_v \end{aligned} \tag{5.9}$$

We now have expressions for the contributions to the proton make-up from its valence quarks, from the sea quarks and from the gluons in terms of the (unknown) variables a and b by recasting the expressions in terms of Q_s and Q_v. The last stage is to cast C_j^m for each parton type in terms of Q_s, Q_v and G. This simply involves taking the *fractional* contribution for each parton charge type, e.g. $+\frac{2}{3}$ quarks, from each class of contributor to the proton's make-up, i.e., Q_s, Q_v or Gluons, and adding them together. As an example, consider the construction of $C_j^{-\frac{1}{3}}$ from table 5.5. The contribution from Q_s is $(1+\epsilon_s)a$ divided by the total amount of sea quarks, $(10+2\epsilon_s)a$; the contribution from Q_v is just $1b$ from a total of $9b$; there is no gluonic contribution. This method also (rather nicely) eliminates the dependency on the variables a and b. The expressions for each C_j^m are written below in equations 5.10.

$$\begin{aligned} C_j^{+\frac{2}{3}} &= \frac{4}{(10 + 2\epsilon_s)} \cdot Q_s + \frac{8}{9} \cdot Q_v \\ C_j^{+\frac{1}{3}} &= \frac{(1 + \epsilon_s)}{(10 + 2\epsilon_s)} \cdot Q_s \end{aligned}$$

$$
\begin{aligned}
C_j^0 &= 1 - Q_s - Q_v \\
C_j^{-\frac{1}{3}} &= \frac{(1+\epsilon_s)}{(10+2\epsilon_s)} \cdot Q_s + \frac{1}{9} \cdot Q_v \\
C_j^{-\frac{2}{3}} &= \frac{4}{(10+2\epsilon_s)} \cdot Q_s
\end{aligned} \tag{5.10}
$$

Chapter 6

Analysis II

6.1 Fitting a C_j distribution

Utilising the equations set out in section 5.10.1, specifically equations 5.10, (and a constant strangeness suppression factor of 0.24) it is possible to find the best fit from the MINUIT [38] minimisation package. This is achieved by varying two of the three fractions Q_s, Q_v and G, coupled with the fact that $Q_s + Q_v + G = 1$ to reduce the number of parameters by one, and subsequently minimising the χ^2 per bin as defined in equation 6.1. The error on the data, (E_{Dat}), dominates because the error on the fit (E_{Fit}) is small as fluctuations have been smoothed out. It must be remembered that the variables, Q_s, Q_v, G, are *fractions*, and before comparisons between theory and data can be made the theoretical fractions must be multiplied by the number of data entries. It is found that the fits are insensitive to the precise value of the strangeness suppression factor and consequently, this method cannot be used to derive a value for the strangeness suppression factor, ϵ_s.

$$\chi^2 = \sum_{bins} \left(\frac{Data - Fit}{E_{Dat}} \right)^2 \tag{6.1}$$

6.2 Initial Results

Presented below in table 6.2 are the results obtained from fits performed to the *data* using the procedures and theory laid out in the above sections. All the steps will now be discussed.

C_j distributions are obtained in both Generated and Reconstructed Monte Carlo and their ratio, G/R (which parameterises the response of the detector), and is known as the acceptance-correction factor, is obtained. This is subsequently used to "correct" the data back to the Generated level.

Ideal C_j distributions are obtained from the Generator-level description using the idea of charge-conjugation symmetry coupled with a spline-smoothing technique which reduces statistical fluctuations due to non-infinite statistics of Monte Carlo simulated data.

These are then used, in varying proportions, to fit to the acceptance-corrected data, using the prescription and formulae laid out in section 5.10.1.

In the results presented below in table 6.2, the same Monte Carlo model is used for both the G/R (acceptance) correction and also for obtaining the "theoretical" parton distributions, *i.e.*, if MEPS64 is used to model the detector response, then MEPS64 is also used to form the "theoretical" parton distributions. "Cross-corrections" and other tests for consistency of results which were performed are presented later in this chapter.

The mean x_{Bj} and mean Q^2 for the results in table 6.2 are presented in table 6.1.

6.3 Further Analysis

Set forth in this chapter are the details of the analyses performed on the consistency of the results presented in section 6.2, as well as subsequent investigations into the behaviour of C_j at high $\hat{s}$ (section 6.6) and the contents of the zero bin (sections 6.7 and 6.8).

Data or Monte Carlo?	$< x_{Bj} >$	$< Q^2 >$
Low Q^2 Data	0.234×10^{-2}	$29.68\ GeV^2$
Low Q^2 MEAR64 Monte Carlo	0.238×10^{-2}	$28.30\ GeV^2$
Low Q^2 MEPS64 Monte Carlo	0.238×10^{-2}	$28.27\ GeV^2$
High Q^2 Data	1.184×10^{-2}	$497.4\ GeV^2$
High Q^2 MEAR64 Monte Carlo	1.209×10^{-2}	$492.4\ GeV^2$
High Q^2 MEPS64 Monte Carlo	1.208×10^{-2}	$492.1\ GeV^2$

Table 6.1: Table of mean values of x_{Bj} and Q^2 both for the data sample that is fitted and also for the Monte Carlo samples which are used to fit to that data

6.4 Consistency Checks

To ensure that our fitting procedure is both consistent and trustworthy, many checks should be performed. The easiest way to double-check the results of the fit would be to input a known distribution, say, a reconstructed Monte Carlo which has been acceptance-corrected back to the generated level, which we call "false data", and compare the "true" values, *i.e.*, the proportions known to be present in the Monte Carlo, with those obtained from the fitting program. As the input quantity is known, any systematic differences between true values and fitted values will show up when inspecting the results. These differences can then be used as systematic errors in the final results to show sensitivity to this essential Monte Carlo dependence. Good agreement is obtained using both MEPS and MEAR (descriptions of each may be found in sections 3.3 and 3.4 respectively) Monte Carlo simulations, as tabulated in 6.3. *Note that the numbers in table 6.3 give the results of fitting Monte Carlo to Monte Carlo and are meant to be different to those in table 6.2 which fits Monte Carlo to data.*

Note the difference in the value of the χ^2 between the results for data and Monte Carlo in tables 6.2 and 6.3, respectively. If the fit were perfect, we would expect a χ^2 of one per degree of freedom, and the χ^2 values for the data are all between one and two per degree of freedom indicating that the fits are reasonable. The χ^2 values for the Monte Carlo are less than one per degree of freedom. This usually indicates that the errors are too large, and this is the case when we attempt to fit

Monte Carlo	Q_s	Q_v	Gluon	χ^2/dof
MEPS 64 Low Q^2	67 ±10%	5 ±6%	28 ±8%	1.28
MEAR 64 Low Q^2	59 ±5%	6 ±5%	37 ±5%	1.89
MEPS 64 High Q^2	78 ±10%	22 ±13%	0 ±9%	1.98
MEAR 64 High Q^2	51 ±10%	20 ±9%	30 ± 9%	1.58

Table 6.2: Parton fractions obtained when acceptance-corrected ***data*** were fitted using theoretical distributions obtained from both MEAR64 and MEPS64 Monte Carlo models at both low and high Q^2

Monte Carlo	Q_s	Q_v	Gluon	χ^2/dof
MEAR 64 Low Q^2 Fit	47 ±10%	6 ±4%	47 ±9%	0.16
MEAR 64 Low Q^2 Actual	54%	5%	41%	
MEPS 64 Low Q^2 Fit	58 ±9%	9 ±5%	33 ±7%	0.41
MEPS 64 Low Q^2 Actual	59%	5%	36%	

Table 6.3: Comparison of the proportions in fitted Monte Carlo with the actual proportions in that Monte Carlo at low Q^2

a distribution to itself. The fact that we obtain small χ^2 values when doing such a fit is another good indication that the procedure is correct.

These two tests, however, are not enough to provide conclusive proof that the method works reliably at all times. A procedure to obtain more combinations of theoretical models and acceptance corrections is to use various Monte Carlo simulations in different combinations together, e.g. using MEPS theoretical parton distributions to model MEAR acceptance-corrected Reconstructed MEAR Monte Carlo. This provides more permutations using the available Monte Carlo data, and all of the possible permutations are tabulated in table 6.4. The nomenclature at this point can be potentially a source of great confusion, therefore a method number has been ascribed to each of the different tests for easy reference.

"False Data" is the Monte Carlo equivalent of the reconstructed (real) data. It

Method Number	"False data" to be fitted	$\frac{G}{R}$-correction by which Monte Carlo?	Using Theoretical distribution of
1a	MEPS	MEPS	MEPS
2a	MEPS	MEAR	MEPS
3a	MEPS	MEPS	MEAR
4a	MEPS	MEAR	MEAR
1b	MEAR	MEPS	MEPS
2b	MEAR	MEAR	MEPS
3b	MEAR	MEPS	MEAR
4b	MEAR	MEAR	MEAR

Table 6.4: Possible combinations of Monte Carlo models

will be acceptance-corrected back to the generator level (just as is done with real data) by the Monte Carlo in the 3$^{\text{rd}}$ column, labelled "G/R-correction by which Monte Carlo?". The "theoretical" distributions are those obtained from the spline-smoothing technique described in section 5.10.

Such tests can only been performed using MEAR and MEPS Monte Carlo (both using Lepto 6.4) as we cannot use either HERWIG 5.8d or DJANGO for reasons discussed in section 3.8.

However, this method still does not provide us with a large *range* of parton proportions with which to test our program. What is required is some means of entering as many different combinations of parton distributions as possible and to confirm that the fitting program models all of these variations well. Rather than generating vast quantities of Monte Carlo using different parton distributions, a technique of re-weighting already existing simulation can be used.

6.4.1 Re-weighting Monte Carlo

Though the Monte Carlo models we use have a certain proportion of partons derived from parton distribution functions, it is possible to put in different proportions as we please for testing purposes. In this manner, we can test each method in table 6.4 several times. This allows a good spread of results to be obtained. It is possible (in the HBOOK package) to fill a histogram with variable weights, therefore, if we know the proportions of partons in the generated histogram and the proportions we would like to test, then a simple calculation will yield the necessary weights to be applied to each parton-type histogram. This gives a good degree of flexibility when testing the fitting program, and allows a large range of all the quantities to be checked.

6.4.2 Which Range Of Variables?

As there are three variables that can be changed, it was decided that one of the variables should be held constant at some value, e.g., the gluon at some fixed value, leaving two "free" variables. It was decided that the proportion of valence quarks, Q_v, would be chosen freely thus leaving Q_s to be determined by unitarity. Imposing such a choice in no way biases our results because of the unitarity constraint; had we chosen the proportion of sea quarks, Q_s, as the fraction to be varied, then Q_v would have been the complementary value instead. This process can then be repeated at several values of Q_v per fixed gluon value. It is then possible to repeat further the whole procedure at a different fixed gluon value. Four different values of Q_v at each gluon value were chosen, and the test repeated at three fixed gluon values resulting in twelve tests per value of Q_v *per method.*

From tables 6.5, 6.6, 6.7, 6.8, 6.9, 6.10, 6.11 and 6.12 and figures 6.1, 6.2, 6.3 and 6.4 it can be seen that the agreement between actual values and fitted values is excellent at all of the chosen gluon values. It can be seen in figure 6.3 that there exists occasionally some confusion between gluons and sea quarks and that a higher value of Q_s will (necessarily) yield a lower value of G. This is only to be expected, and is in fact quite "natural"; the distinction between gluons and sea quarks being defined arbitrarily by kinematics. There does not appear to be any confusion between sea and valence quarks, however. Consistency between true values and fitted values can be achieved by taking systematic errors of 5% to both MEPS-fitted data and MEAR-fitted data.

Method Number	Input Proportions Q_s	Input Proportions Q_v	Input Proportions G	Output Proportions Q_s	Output Proportions Q_v	Output Proportions G
1a	0.65±0.01	0.10±0.01	0.25±0.01	0.65±0.082	0.112±0.048	0.237±0.064
1a	0.50±0.01	0.25±0.01	0.25±0.01	0.529±0.081	0.232±0.048	0.239±0.064
1a	0.35±0.01	0.40±0.01	0.25±0.01	0.407±0.081	0.353±0.048	0.240±0.064
1a	0.20±0.01	0.55±0.01	0.25±0.01	0.284±0.081	0.479±0.048	0.238±0.064
1a	0.60±0.01	0.05±0.01	0.35±0.01	0.596±0.083	0.080±0.048	0.323±0.066
1a	0.45±0.01	0.20±0.01	0.35±0.01	0.475±0.083	0.201±0.048	0.325±0.066
1a	0.30±0.01	0.35±0.01	0.35±0.01	0.353±0.083	0.323±0.048	0.324±0.066
1a	0.15±0.01	0.50±0.01	0.35±0.01	0.231±0.083	0.448±0.048	0.321±0.067
1a	0.50±0.01	0.05±0.01	0.45±0.01	0.509±0.085	0.088±0.048	0.403±0.069
1a	0.35±0.01	0.20±0.01	0.45±0.01	0.387±0.085	0.210±0.048	0.403±0.069
1a	0.20±0.01	0.35±0.01	0.45±0.01	0.265±0.085	0.333±0.048	0.401±0.069
1a	0.05±0.01	0.50±0.01	0.45±0.01	0.143±0.085	0.460±0.048	0.397±0.069

Table 6.5: Comparison of input and output proportions using method number 1a

Method Number	Input Proportions Q_s	Input Proportions Q_v	Input Proportions G	Output Proportions Q_s	Output Proportions Q_v	Output Proportions G
1b	0.65±0.01	0.10±0.01	0.25±0.01	0.679±0.081	0.112±0.048	0.209±0.063
1b	0.50±0.01	0.25±0.01	0.25±0.01	0.555±0.081	0.233±0.049	0.212±0.063
1b	0.35±0.01	0.40±0.01	0.25±0.01	0.430±0.081	0.357±0.049	0.213±0.063
1b	0.20±0.01	0.55±0.01	0.25±0.01	0.303±0.081	0.485±0.048	0.212±0.064
1b	0.60±0.01	0.05±0.01	0.35±0.01	0.630±0.082	0.080±0.048	0.290±0.065
1b	0.45±0.01	0.20±0.01	0.35±0.01	0.504±0.082	0.204±0.048	0.292±0.065
1b	0.30±0.01	0.35±0.01	0.35±0.01	0.377±0.082	0.330±0.048	0.293±0.065
1b	0.15±0.01	0.50±0.01	0.35±0.01	0.249±0.082	0.460±0.048	0.291±0.066
1b	0.50±0.01	0.05±0.01	0.45±0.01	0.544±0.084	0.088±0.048	0.369±0.068
1b	0.35±0.01	0.20±0.01	0.45±0.01	0.415±0.084	0.215±0.048	0.370±0.068
1b	0.20±0.01	0.35±0.01	0.45±0.01	0.285±0.084	0.345±0.048	0.370±0.068
1b	0.05±0.01	0.50±0.01	0.45±0.01	0.154±0.084	0.480±0.048	0.366±0.068

Table 6.6: Comparison of input and output proportions using method number 1b

Method Number	Input Proportions			Output Proportions		
	Q_s	Q_v	G	Q_s	Q_v	G
2a	0.65±0.01	0.10±0.01	0.25±0.01	0.607±0.069	0.069±0.041	0.324±0.056
2a	0.50±0.01	0.25±0.01	0.25±0.01	0.474±0.068	0.197±0.041	0.329±0.056
2a	0.35±0.01	0.40±0.01	0.25±0.01	0.337±0.068	0.331±0.041	0.332±0.056
2a	0.20±0.01	0.55±0.01	0.25±0.01	0.198±0.067	0.471±0.041	0.331±0.056
2a	0.60±0.01	0.05±0.01	0.35±0.01	0.536±0.072	0.052±0.041	0.412±0.059
2a	0.45±0.01	0.20±0.01	0.35±0.01	0.407±0.070	0.178±0.041	0.415±0.057
2a	0.30±0.01	0.35±0.01	0.35±0.01	0.279±0.069	0.311±0.041	0.410±0.057
2a	0.15±0.01	0.50±0.01	0.35±0.01	0.155±0.068	0.450±0.041	0.396±0.057
2a	0.50±0.01	0.05±0.01	0.45±0.01	0.444±0.072	0.070±0.041	0.486±0.060
2a	0.35±0.01	0.20±0.01	0.45±0.01	0.322±0.071	0.199±0.041	0.478±0.059
2a	0.20±0.01	0.35±0.01	0.45±0.01	0.211±0.070	0.334±0.041	0.455±0.058
2a	0.05±0.01	0.50±0.01	0.45±0.01	0.116±0.068	0.476±0.041	0.408±0.056

Table 6.7: Comparison of input and output proportions using method number 2a

Method Number	Input Proportions			Output Proportions		
	Q_s	Q_v	G	Q_s	Q_v	G
2b	0.65±0.01	0.10±0.01	0.25±0.01	0.643±0.068	0.064±0.042	0.294±0.056
2b	0.50±0.01	0.25±0.01	0.25±0.01	0.506±0.068	0.193±0.042	0.301±0.056
2b	0.35±0.01	0.40±0.01	0.25±0.01	0.364±0.068	0.329±0.041	0.307±0.056
2b	0.20±0.01	0.55±0.01	0.25±0.01	0.218±0.067	0.474±0.041	0.309±0.056
2b	0.60±0.01	0.05±0.01	0.35±0.01	0.574±0.070	0.046±0.041	0.380±0.057
2b	0.45±0.01	0.20±0.01	0.35±0.01	0.438±0.069	0.176±0.041	0.386±0.057
2b	0.30±0.01	0.35±0.01	0.35±0.01	0.300±0.068	0.313±0.041	0.387±0.057
2b	0.15±0.01	0.50±0.01	0.35±0.01	0.164±0.068	0.457±0.041	0.379±0.057
2b	0.50±0.01	0.05±0.01	0.45±0.01	0.475±0.071	0.065±0.041	0.460±0.059
2b	0.35±0.01	0.20±0.01	0.45±0.01	0.341±0.070	0.200±0.041	0.459±0.059
2b	0.20±0.01	0.35±0.01	0.45±0.01	0.212±0.069	0.341±0.041	0.447±0.058
2b	0.05±0.01	0.50±0.01	0.45±0.01	0.097±0.068	0.490±0.041	0.413±0.057

Table 6.8: Comparison of input and output proportions using method number 2b

Method Number	Input Proportions Q_s	Input Proportions Q_v	Input Proportions G	Output Proportions Q_s	Output Proportions Q_v	Output Proportions G
3a	0.65±0.01	0.10±0.01	0.25±0.01	0.900±0.062	0.100±0.079	0.000±0.049
3a	0.50±0.01	0.25±0.01	0.25±0.01	0.791±0.064	0.209±0.082	0.000±0.051
3a	0.35±0.01	0.40±0.01	0.25±0.01	0.681±0.067	0.319±0.086	0.000±0.055
3a	0.20±0.01	0.55±0.01	0.25±0.01	0.569±0.072	0.431±0.093	0.000±0.060
3a	0.60±0.01	0.05±0.01	0.35±0.01	0.857±0.096	0.069±0.043	0.074±0.084
3a	0.45±0.01	0.20±0.01	0.35±0.01	0.742±0.096	0.180±0.043	0.078±0.084
3a	0.30±0.01	0.35±0.01	0.35±0.01	0.626±0.097	0.290±0.043	0.083±0.084
3a	0.15±0.01	0.50±0.01	0.35±0.01	0.507±0.097	0.403±0.043	0.090±0.085
3a	0.50±0.01	0.05±0.01	0.45±0.01	0.738±0.100	0.076±0.043	0.186±0.089
3a	0.35±0.01	0.20±0.01	0.45±0.01	0.622±0.100	0.187±0.043	0.191±0.089
3a	0.20±0.01	0.35±0.01	0.45±0.01	0.504±0.100	0.299±0.043	0.197±0.089
3a	0.05±0.01	0.50±0.01	0.45±0.01	0.366±0.100	0.409±0.043	0.226±0.090

Table 6.9: Comparison of input and output proportions using method number 3a

Method Number	Input Proportions Q_s	Input Proportions Q_v	Input Proportions G	Output Proportions Q_s	Output Proportions Q_v	Output Proportions G
3b	0.65±0.01	0.10±0.01	0.25±0.01	0.900±0.049	0.100±0.059	0.000±0.033
3b	0.50±0.01	0.25±0.01	0.25±0.01	0.789±0.050	0.211±0.061	0.000±0.034
3b	0.35±0.01	0.40±0.01	0.25±0.01	0.677±0.052	0.323±0.064	0.000±0.037
3b	0.20±0.01	0.55±0.01	0.25±0.01	0.562±0.055	0.438±0.068	0.000±0.041
3b	0.60±0.01	0.05±0.01	0.35±0.01	0.902±0.095	0.069±0.043	0.029±0.082
3b	0.45±0.01	0.20±0.01	0.35±0.01	0.783±0.094	0.182±0.043	0.034±0.082
3b	0.30±0.01	0.35±0.01	0.35±0.01	0.663±0.095	0.297±0.043	0.040±0.083
3b	0.15±0.01	0.50±0.01	0.35±0.01	0.538±0.095	0.414±0.043	0.048±0.084
3b	0.50±0.01	0.05±0.01	0.45±0.01	0.786±0.098	0.076±0.043	0.138±0.087
3b	0.35±0.01	0.20±0.01	0.45±0.01	0.664±0.098	0.192±0.043	0.144±0.087
3b	0.20±0.01	0.35±0.01	0.45±0.01	0.539±0.098	0.310±0.043	0.151±0.087
3b	0.05±0.01	0.50±0.01	0.45±0.01	0.397±0.099	0.428±0.043	0.175±0.088

Table 6.10: Comparison of input and output proportions using method number 3b

Method Number	Input Proportions			Output Proportions		
	Q_s	Q_v	G	Q_s	Q_v	G
4a	0.65±0.01	0.10±0.01	0.25±0.01	0.802±0.082	0.057±0.038	0.141±0.070
4a	0.50±0.01	0.25±0.01	0.25±0.01	0.680±0.081	0.176±0.038	0.145±0.071
4a	0.35±0.01	0.40±0.01	0.25±0.01	0.548±0.082	0.300±0.037	0.152±0.072
4a	0.20±0.01	0.55±0.01	0.25±0.01	0.407±0.082	0.429±0.037	0.164±0.074
4a	0.60±0.01	0.05±0.01	0.35±0.01	0.691±0.084	0.049±0.037	0.260±0.074
4a	0.45±0.01	0.20±0.01	0.35±0.01	0.565±0.084	0.166±0.037	0.269±0.074
4a	0.30±0.01	0.35±0.01	0.35±0.01	0.432±0.084	0.289±0.037	0.279±0.075
4a	0.15±0.01	0.50±0.01	0.35±0.01	0.290±0.085	0.418±0.037	0.293±0.076
4a	0.50±0.01	0.05±0.01	0.45±0.01	0.547±0.088	0.074±0.037	0.379±0.078
4a	0.35±0.01	0.20±0.01	0.45±0.01	0.413±0.087	0.194±0.037	0.393±0.078
4a	0.20±0.01	0.35±0.01	0.45±0.01	0.269±0.087	0.321±0.037	0.410±0.079
4a	0.05±0.01	0.50±0.01	0.45±0.01	0.115±0.087	0.454±0.036	0.431±0.080

Table 6.11: Comparison of input and output proportions using method number 4a

Method Number	Input Proportions			Output Proportions		
	Q_s	Q_v	G	Q_s	Q_v	G
4b	0.65±0.01	0.10±0.01	0.25±0.01	0.848±0.081	0.047±0.038	0.105±0.069
4b	0.50±0.01	0.25±0.01	0.25±0.01	0.726±0.081	0.168±0.038	0.106±0.070
4b	0.35±0.01	0.40±0.01	0.25±0.01	0.593±0.081	0.294±0.038	0.113±0.071
4b	0.20±0.01	0.55±0.01	0.25±0.01	0.447±0.082	0.428±0.037	0.125±0.073
4b	0.60±0.01	0.05±0.01	0.35±0.01	0.747±0.083	0.038±0.037	0.215±0.072
4b	0.45±0.01	0.20±0.01	0.35±0.01	0.620±0.083	0.159±0.037	0.221±0.073
4b	0.30±0.01	0.35±0.01	0.35±0.01	0.484±0.083	0.286±0.037	0.230±0.074
4b	0.15±0.01	0.50±0.01	0.35±0.01	0.337±0.084	0.420±0.037	0.243±0.075
4b	0.50±0.01	0.05±0.01	0.45±0.01	0.610±0.086	0.063±0.037	0.326±0.076
4b	0.35±0.01	0.20±0.01	0.45±0.01	0.473±0.086	0.188±0.037	0.339±0.077
4b	0.20±0.01	0.35±0.01	0.45±0.01	0.325±0.086	0.320±0.037	0.354±0.077
4b	0.05±0.01	0.50±0.01	0.45±0.01	0.167±0.087	0.460±0.037	0.373±0.079

Table 6.12: Comparison of input and output proportions using method number 4b

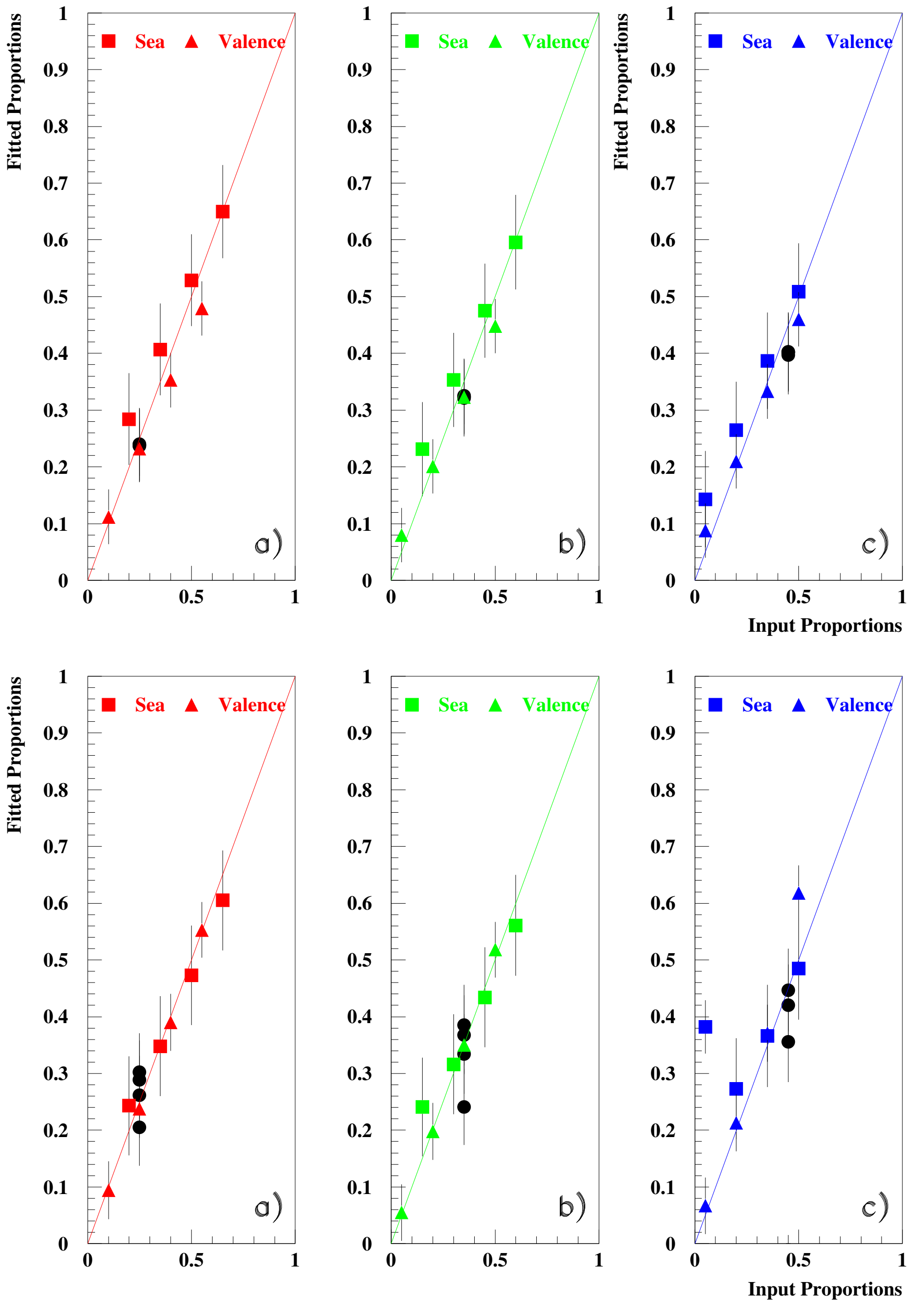

Figure 6.1: Comparison of fitted values with input values for Monte Carlo methods 1a and 1b (see table 6.4 for the definitions of the methods)

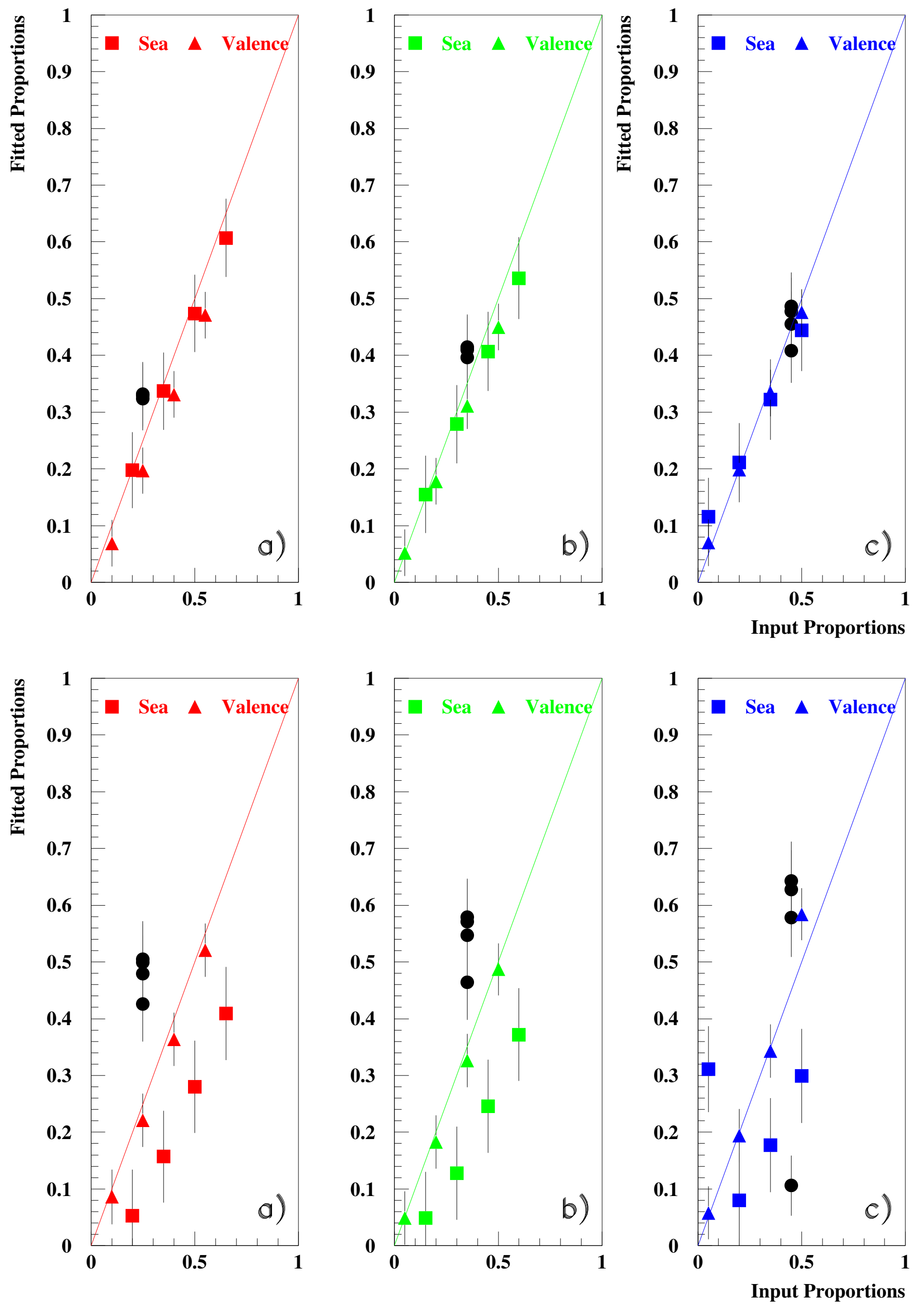

Figure 6.2: Comparison of fitted values with input values for Monte Carlo methods 2a and 2b (see table 6.4 for the definitions of the methods)

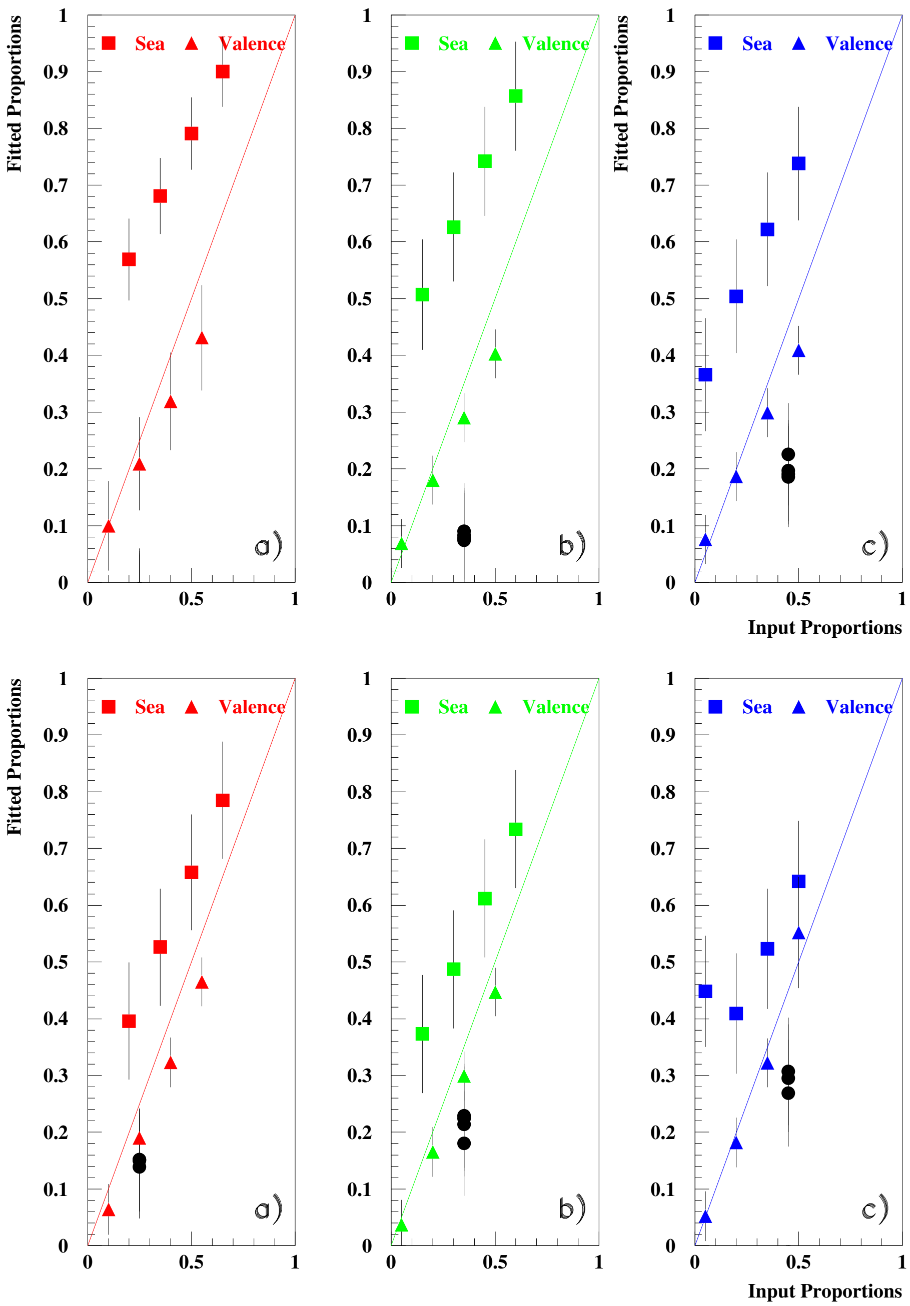

Figure 6.3: Comparison of fitted values with input values for Monte Carlo methods 3a and 3b (see table 6.4 for the definitions of the methods)

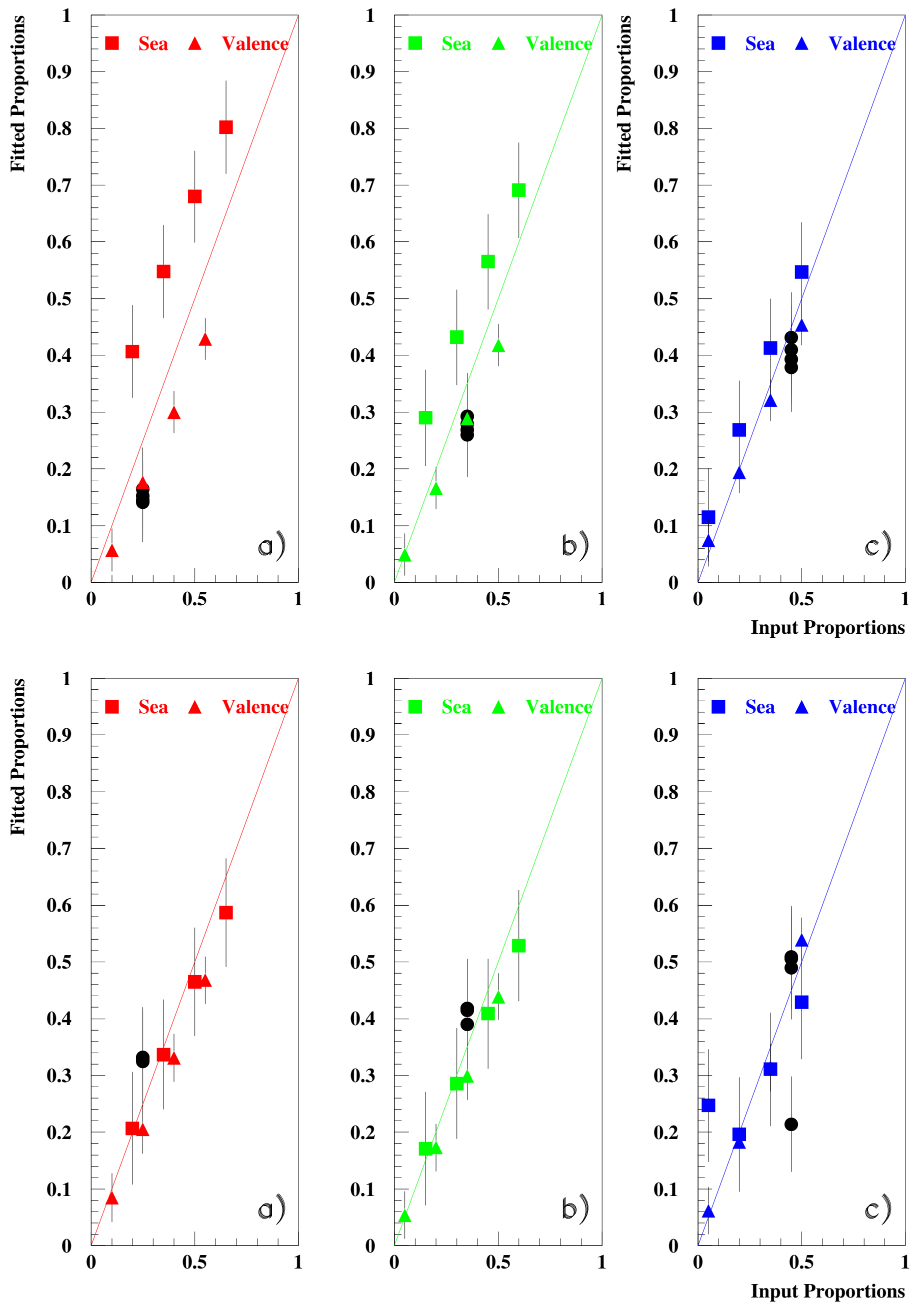

Figure 6.4: Comparison of fitted values with input values for Monte Carlo methods 4a and 4b (see table 6.4 for the definitions of the methods)

6.5 An Attempt At Improving The Signal

We now have fitted results for the proton using both MEPS and MEAR Monte Carlo simulations. It is possible that an improvement, or a better fit with smaller errors could be found by utilising, for example, the information contained in the zero bin, or by attempting to define our own region of phase space by making use of the variation of the parton distributions with $\hat{s}$. The attempts at improving the signal are documented in the following sections.

6.6 C_j At High $\hat{s}$

It is known that the C_j distributions of Monte Carlo generated gluons and partons vary with the hard sub-process centre of mass energy, $\hat{s}$. For example, generation of Boson-Gluon Fusion (BGF) and QCD Compton (QCDC) begin at different values of $\hat{s}$, *i.e.*, 1.8 GeV for BGF and 0.8 GeV for QCDC, whereas QPM events have an $\hat{s}$ of zero. This means that there is a gap of ≈ 0.8 GeV in Monte Carlo where nothing is simulated due mainly to the difficulties of modelling non-perturbative QCD. Some data, however, *do* lie in this range which means that they do not get modelled by any simulation.

In an attempt to get around this difficulty, and also to define exactly the phase space in which the fit is to be performed, it was decided to impose an $\hat{s}$ cut at some value well away from the non-perturbative region. The choice of $\hat{s}$ is largely arbitrary, but a sensible choice would be one which separates light and heavy flavours. Heavy flavours are produced predominantly by boson-gluon fusion, and thus a direct handle on the gluon density in the proton can be obtained through charm studies, e.g., J/ψ analyses. The charm threshold, $m^2_{c\bar{c}}$, was thus used as the $\hat{s}$ cut-off. This also had the inadvertent bonus of approximately equalising the statistics in both the low and high $\hat{s}$ samples. In order to take advantage of this new cut that has been imposed, it might be prudent to have two distributions from which to fit the data, *i.e.*, one set of distributions for high $\hat{s}$ events, and one set for low.

It is a straightforward procedure to split the sample into two sets of theoretical parton distributions, and relatively easy to add one more parameter; the proportion of high $\hat{s}$ events; to the MINUIT fitting procedure. Unfortunately, splitting the sample enlarges the error bars by at least a factor of $\sqrt{2}$, and the paucity of the

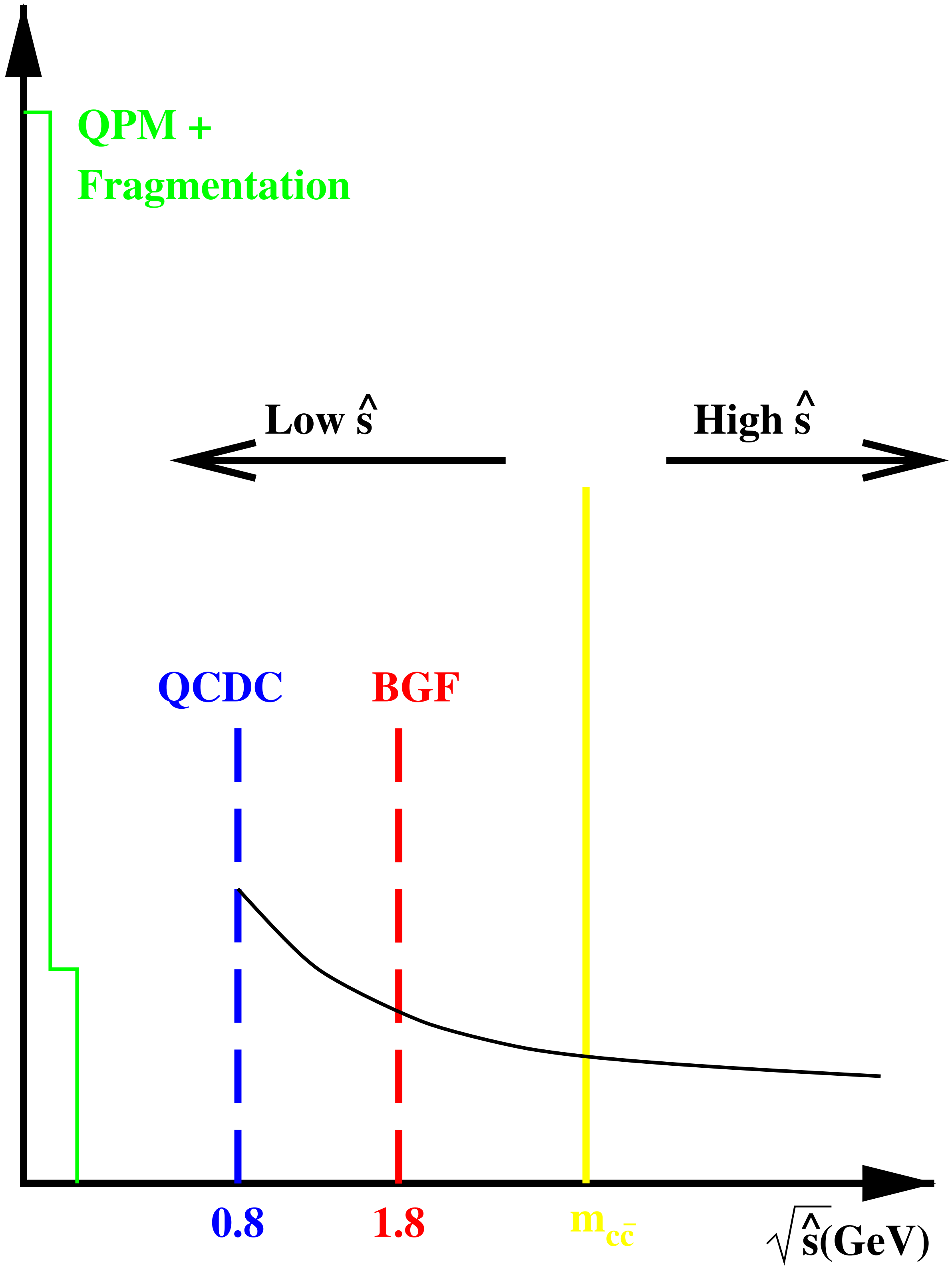

Figure 6.5: Cartoon illustrating the different values of $\hat{s}$ in a Monte Carlo

statistics means that the error bars are not small. Another difficulty encountered in trying to do this $\hat{s}$ analysis is that only the MEPS Monte Carlo has an easily-accessible value of $\hat{s}$. This means that we are wholly-dependent on the theoretical description of this Monte Carlo, but, having said that, we can now stipulate exactly to which partonic kinematic range we are fitting.

Results for MEPS (Lepto 6.4) for both un-split and split $\hat{s}$ cases are tabulated for comparison in 6.13. This approach must assume that the fraction of sea quarks, Q_s, is the same at both low and high $\hat{s}$.

Monte Carlo	Q_s	Q_v	Gluon	High $\hat{s}$ fraction	χ^2/dof
Meps 6.4	$60 \pm 12\%$	$10 \pm 5\%$	$30 \pm 11\%$	$44 \pm 10\%$	0.58
Non-split	$58 \pm 9\%$	$9 \pm 5\%$	$33 \pm 7\%$	N/A	0.41

Table 6.13: Results for $\hat{s}$-split distributions

The final results are in very good agreement with those obtained by an all-inclusive (or non $\hat{s}$-split) fit, and a ratio of 44% indicates that the partons react at both low and high $\hat{s}$ approximately equally. Given that the error on the fitted value is ~ 10 %, it is unreliable to state that there might be a genuine bias towards low $\hat{s}$ values. The actual proportion of high ŝ events in the MEPS64 Monte Carlo is 42% $\pm$ 1% which means that the fit recovers this value very well. Unfortunately, therefore, we must conclude that splitting the sample into low and high $\hat{s}$ portions adds no extra information to that which we already know.

6.7 Re-Performing The Analysis Utilising The Information Contained Within The Zero Bin

Having identified two distinct classes of $C_j \equiv 0.0$, as discussed in section 5.3, it is possible to separate them from the zero bin leaving distributions as shown in figure 6.6. This requires the addition of two further bins to the histogram, one containing $C_j \equiv 0.0$ events with no energy and the other with the rest, *i.e.*, energy but no charged tracks.

Two methods were thought of to take advantage of any potentially new information which may be gleaned from this zero bin division. The first, and most obvious, is to repeat the analysis described in chapter 5 above to investigate whether splitting the contents of the zero bin affects the results in any way. The second method is to perform an investigation only on the contents of the zero bin. These methods are now described here and in section 6.8.

Once all the necessary histograms have been re-created, the analysis can proceed in exactly the same manner as before. However, this time the fitting program will account for the fact that there are now two different categories of $C_j \equiv 0.0$ and fit to them accordingly. It is hoped that this somewhat more refined fit will lead to a result which will be more accurate, *i.e.*, one with smaller error bars.

The results obtained when the MEPS64 Monte Carlo is fitted to itself (labelled *new*) are shown in table 6.14 together with those for the identical test when the zero bin was not split (labelled *old*), as already shown in table 6.2.

Method	Q_s	Q_v	Gluon	χ^2/dof
Old	58±9%	9±5%	33±7%	0.41
New	56±7%	9±5%	35±4%	0.41

Table 6.14: Table comparing two methods of obtaining partonic proportions

It can be seen that the results are virtually identical, and that the only difference is that 2% of the sea quarks in the unsplit fit have now been identified as gluons in the split fit. This is consistent with the fact that at low $\hat{s}$ it is extremely difficult to distinguish between a sample of gluonic and sea quark events. A further 2% error should therefore be added to the systematic error.

6.8 Probability Of $C_j \equiv 0.0$ By Gluonic Initiation

Details of how to decompose events having $C_j \equiv 0.0$ were alluded to in section 5.3. It has been stated (also in section 5.3) that empty current regions, *i.e.*, $C_j \equiv 0.0$, are an indication of boson-gluon fusion events. However, not all empty current hemispheres are initiated by gluons; some are initiated by quarks. If it were possible to determine the probability that a gluon were responsible for initiating a $C_j \equiv 0.0$ event, this would be a good handle for determining the gluon density within the proton. Answering this question, however, is a non-trivial task!

Probabilities for each sub-class of $C_j \equiv 0.0$ must be calculated for both gluons and

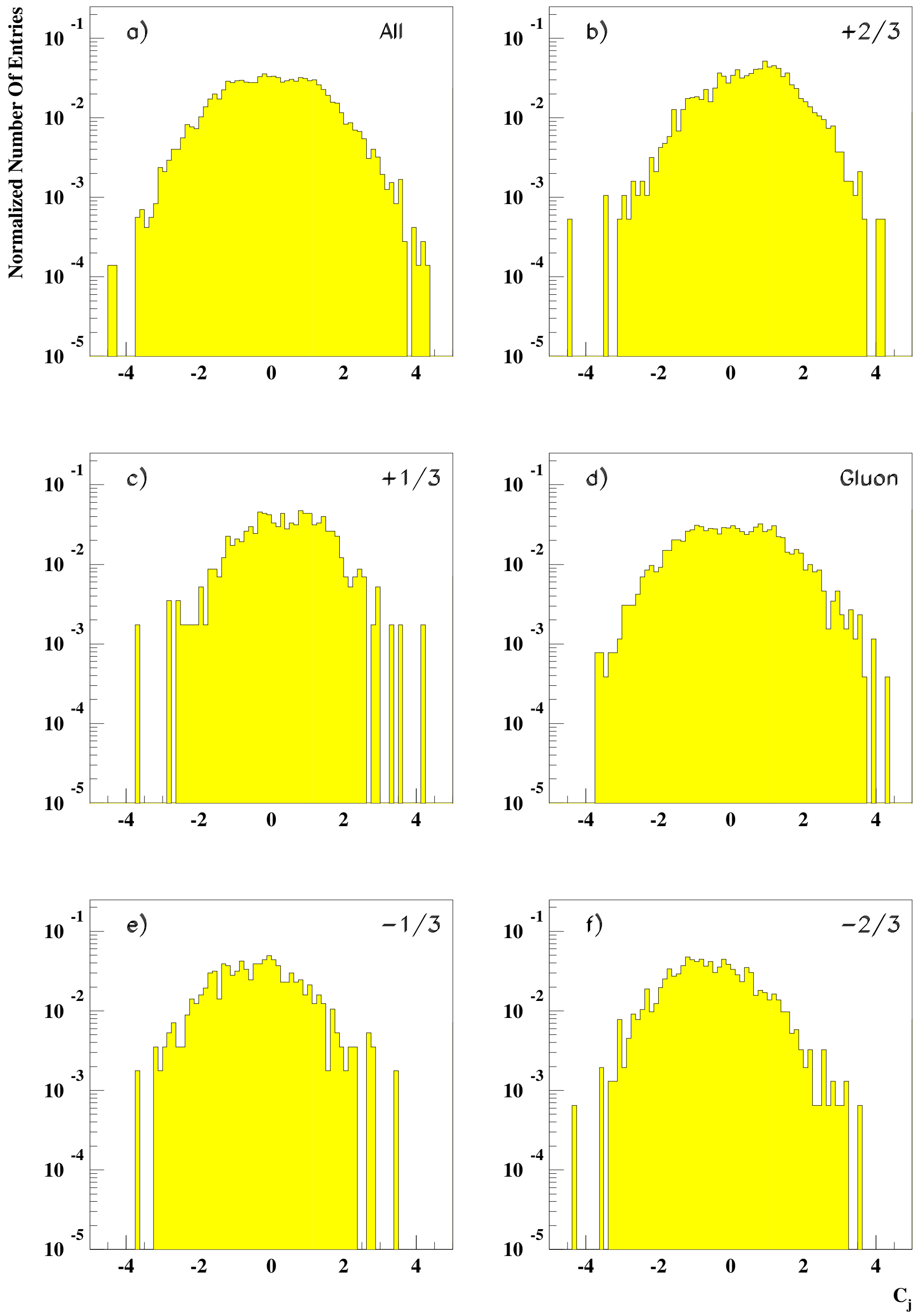

Figure 6.6: C_j distributions for MEPS64 Monte Carlo where the zero bin has been separated and shifted to the extreme right of the plot. Those events with zero energy in the current region of the Breit frame are in the left-hand bin of the right-hand pair (penultimate bin), whereas those events with no charge-multiplicity in the current region of the Breit frame are in the extreme right hand bin

non-gluons, *i.e.*, quarks. These must then be combined, using Bayesian statistics [39] [40], to give the probability that; given that $C_j \equiv 0.0$; a gluon initiated such an event.

The probabilities that will be required are the following:

- Probability that a gluon creates an event with no charged tracks in the current region of the Breit frame, $P(g \cap C)$;
- Probability that a gluon creates an event with no energy in the current region of the Breit frame, $P(g \cap E)$;
- Probability that a quark, *i.e.*, a non-gluon creates an event with no charged tracks in the current region of the Breit frame, $P(\bar{g} \cap C)$;
- Probability that a quark creates an event with no energy in the current region of the Breit frame, $P(\bar{g} \cap E)$.

These quantities are unknown. What we *do* have, however, are the probabilities that, for example, a gluon was responsible for initiating a certain event *given that* there was no charged-multiplicity, and also the probability of obtaining such an event in the first place, *i.e.*, $P(g \mid C)$ and $P(C)$ respectively. Using the method of Bayes [40], we can combine these to give us the required probability, as shown in the table of equations labelled 6.2.

$$\begin{aligned}
P(g \cap C) &= P(g \mid C) \cdot P(C) \\
P(g \cap E) &= P(g \mid E) \cdot P(E) \\
P(\bar{g} \cap C) &= P(\bar{g} \mid C) \cdot P(C) \\
P(\bar{g} \cap E) &= P(\bar{g} \mid E) \cdot P(E)
\end{aligned} \tag{6.2}$$

The probability that a gluon initiated an empty current region event may then be deduced using equation 6.3. Similarly, the probability that a non-gluon initiated an empty current hemisphere event may be obtained from 6.4.

$$P(g) = P(g \cap C) + P(g \cap E) \tag{6.3}$$

$$P(\bar{g}) = P(\bar{g} \cap C) + P(\bar{g} \cap E) \tag{6.4}$$

The values obtained for each of these categories can be placed in a table for each Monte Carlo simulation available, the template for which is shown in table 6.15.

Parton Species, x	$P(E)$	$P(C)$	$P(x \cap C) + P(x \cap E)$
Gluons	$P(g \cap E)$	$P(g \cap C)$	$P(g)$
Quarks	$P(\bar{g} \cap E)$	$P(\bar{g} \cap C)$	$P(\bar{g})$

Table 6.15: A template table illustrating the component probabilities used to calculate gluonic initiation probabilities

The probabilities for the charged and non-charged parton types, i.e., quarks and gluons respectively, for both classes of zero can then be summed to yield, finally, the probability that a gluon was the initiator of an empty current hemisphere event. It can be seen in both tables 6.16 and 6.17 for MEPS64 and MEAR64 respectively, that there is no strong preference for any class or type. The gluons are responsible for just over half of the empty current hemisphere events, but with an error of $\sim 5\%$ one cannot draw any firm conclusions from this. It can also be seen that the empty current hemisphere events are divided into the two classes approximately equally. This, again, means that there is no difference, as far as our modelling is concerned, between events with no charged multiplicity in the current region of the Breit frame and those with no energy in the current region of the Breit frame.

The only conclusion that we can draw, therefore, is that this method provides us with no additional information than we already have from the C_j analysis.

Parton Species, x	P(E)	P(C)	$P(x \cap C) + P(x \cap E)$
Gluons	0.34	0.21	0.55
Quarks	0.17	0.28	0.45

Table 6.16: A table showing the probability that a given quark or gluon was responsible for initiating either an event with no charged multiplicity in the current region of the Breit frame or no energy in the current region of the Breit frame. The Monte Carlo model used was reconstructed MEPS64

Parton Species, x	P(E)	P(C)	$P(x \cap C) + P(x \cap E)$
Gluons	0.34	0.19	0.53
Quarks	0.24	0.23	0.47

Table 6.17: A table showing the probability that a given quark or gluon was responsible for initiating either an event with no charged multiplicity in the current region of the Breit frame or no energy in the current region of the Breit frame. The Monte Carlo model used was reconstructed MEAR64

Chapter 7

Conclusions

This thesis has presented the first measure of relative partonic contributions in DIS at HERA using an average hadronic charge. The method of constructing an average hadronic charge has been described in detail, as well as the necessary theoretical information required for a successful completion of the analysis.

Details have been given of the method of data selection and the choice and implementation of Monte Carlo simulation models. Where these Monte Carlo models were found to be deficient, the attempts at correcting the faults have also been described.

The initial "signal" was found to be only a few standard deviations above the noise, and the methods employed to emphasise and enhance the signal to a more distinct level have been described in detail. The optimal signal was achieved using the fragmentation variable, x_p, raised to a power, $\kappa = 0.4$ to sum over all charged particles in a region of the Breit frame described by $\cos\theta_{Bf} \leq 0.4$. The normalisation chosen was a *sample* normalisation which not only maintains the emphasised contribution from the leading hadron, but also makes the average charge variable, C_j, independent of the sheer size, different multiplicities and kinematics of each sample.

The five different partonic charge-types have been combined in various proportions into three classes of contributor to the proton, namely sea quarks, valence quarks and gluons, Q_s, Q_v and G respectively. The relative contributions from each have been found using a fitting routine at both low and high Q^2, and the values, averaged over the different Monte Carlo models, are shown in table 7.1.

We see that the nominal value for the proportion of valence quarks increases with

Sample	Q_s	Q_v	Gluon
Low Q^2 Data	63 ±11%	6 ±9%	33 ±10%
High Q^2 Data	64 ±15%	21±16%	15 ±9%

Table 7.1: Relative partonic contributions at both low and high Q^2 in DIS at HERA

increasing Q^2, as expected by theory; the proportion of sea quarks remains constant and the gluonic contribution decreases as Q^2 increases. The errors, however, are large enough so that the nominal values are compatible with each other.

Consistency checks were performed on these results by using a technique involving the re-weighting of Monte Carlo distributions. This was the method used both to test the robustness of the method, and also to provide an estimate of the systematic error on the fitted values. The systematic error due to the ambiguities of the fit itself are 3% for MEPS64 and 5% for MEAR64.

An attempt at a specific definition of phase-space using a cut on $\hat{s}$ was investigated. Only the MEPS64 Monte Carlo contained the required information about the hard subprocess, *i.e.*, the Mandelstam variable, $\hat{s}$. Consequently, the results were wholly dependent on the *ansatz* in this Monte Carlo. However, having said this, the results obtained when using the proportion of high $\hat{s}$ events as an extra variable in a fit to Monte Carlo data were in excellent agreement with the same fit when performed on the non $\hat{s}$-split Monte Carlo sample. This change is smaller than the errors on the results, and therefore we conclude that this method yielded no further useful information.

Two distinct classes of events which had the property of $C_j \equiv 0.0$ were identified. Two methods have been described which utilised the information potentially contained within these classes. The first method, in essence a repeat of the first analysis but with the $C_j \equiv 0.0$ classes separated and shifted so that they contribute separately to the fit, gave identical results (within errors) to the fitted results obtained when the $C_j \equiv 0.0$ events were treated inclusively. This method was, consequently, not pursued any further. The second investigation was performed only on the contents of the $C_j \equiv 0.0$ bins as empty current hemispheres are indications of gluonic events. A method was devised to calculate the probability that a given empty current hemisphere was initiated by a gluon. Converting the probabilities into the percentages yields the result that $54\% \pm 1\%$ of all empty current hemispheres are initiated by gluons. This analysis was unsuccessful in providing a convincing method of determining the proportion of either gluons or quarks within the proton.

This analysis has demonstrated that a method of determining relative partonic contributions using an average hadronic charge is possible, yielding credible values for the proportions of gluons and both sea and valence quarks which behave in a manner consistent with expectations.

Such an analysis could be used to take analyses currently limited to inclusive descriptions of the proton a stage further, e.g., rather than having $F_2^p(x, \mathrm{Q}^2)$, it is possible to define $F_2^a(x, \mathrm{Q}^2)$ where a is either gluons, sea or valence quarks such that

$$F_2^p(x, \mathrm{Q}^2) = F_2^{\mathrm{Gluon}}(x, \mathrm{Q}^2) + F_2^{\mathrm{Q_s}}(x, \mathrm{Q}^2) + F_2^{\mathrm{Q_v}}(x, \mathrm{Q}^2)$$

With increased statistics such as those in the integrated 1996 and 1997 (*et seq.* years at HERA, further investigations into C_j could be refined. These include parameterising the behaviour of C_j with x and Q^2 to obtain a functional form describing how C_j varies with both x and Q^2; determining C_j^q rather than C_j^{valence} or C_j^{sea}, all with much smaller errors; marrying the analysis with other Breit frame investigations, e.g., multiplicity analyses. Unfortunately, this requires not only a much larger volume of data, but a considerably larger volume of Monte Carlo simulation than was available for this analysis. When the volume of Monte Carlo simulation far exceeds the data, theoretical distributions with far smaller errors can be achieved. Another natural investigation would be to see how the relative partonic contributions vary with different parton distribution functions.

In conclusion, this analysis has defined a method of determining relative partonic contributions and has made their first determination in DIS at HERA. With increased integrated statistics and higher luminosity, a more precise determination will be possible as well as a host of other interesting analyses not currently possible, some of which have been mentioned above.

Bibliography

[1] A. K. Mavroidis, University of London Thesis, 1995, p31

[2] U. Bassler et al., Progress on Kinematical Variable Reconstruction, H1 note H1-03/93-274 (1993)

[3] A measurement of the proton structure function $F_2(x, q^2)$, H1 Collaboration, Nucl. Phys. 439B (1995) 471

[4] Quarks & Leptons, F. Halzen & A. Martin, John Wiley & Sons (Wiley International Edition)

[5] Measurement of the proton structure function $F_2(x, q^2)$ in the low x region at HERA, H1 Collaboration, Nucl. Phys. 407B (1993) 515

[6] Dave Kant, University of London Thesis, 1996

[7] K. H. Streng, T. F. Walsh and P. M. Zerwas, Zeit. Phys. C2 (1979) 237

[8] H1 Collaboration, A Study of the Fragmentation of Quarks in ep Collisions at HERA Nucl. Phys. 445B (1995) 3

[9] Paul Dixon, Lancaster University Thesis, 1997

[10] H. Bethe and W. Heitler, Proc. Roy. Soc. A 146 (1934) 83

[11] Eram Rizvi, University of London Thesis, 1997, p10

[12] The tracking, calorimeter and muon detectors of the H1 experiment at HERA, H1 Collaboration, Nucl. Instr. Meth. A386 (1997) 310 (Vol. 1) *ibid.* 348 (Vol. 2)

[13] H1 Internal Software

[14] The H1PHAN On-line Manual, http://www-h1.desy.de/icas/imanuals/

[15] H1SIM On-line Manual, http://www-h1.desy.de/icas/imanuals/

[16] BOS Bank Descriptions, http://www-h1.desy.de/h1/iww/iswnotes/sw012.e

[17] R. Brun et al., GEANT Long Write-up, CERN Program Library, W5103 (1989)

[18] Ariadne version 4, Leif Lönnblad,
http://www-h1.desy.de/icas/imanuals/ariadne4.ps.
Originally published as L. Lönnblad, Computer Phys. Comm. 71 (1992) 15

[19] Lepto version 6.1 (Version 6.4 is internal release only)
http://www-h1.desy.de/icas/imanuals/lepto61.ps

[20] Combined QED and QCD radiative effect in deep inelastic lepton proton scattering: the Monte Carlo generator DJANGO6
http://www-h1.desy.de/icas/imanuals/django6.ps

[21] HERACLES. An event generator for ep interactions at HERA including radiative processes, http://www-h1.desy.de/icas/imanuals/heracles44.ps

[22] HERWIG : a Monte Carlo event generator for simulating hadron emission reactions with interfering gluons (version 5.1 - April 1991, G. Marchesini, B. R. Webber, G. Abbiendi, I. G. Knowles, M. H. Seymour, L. Stanco, Computer Phys. Comm. 67 (1992) 465

[23] Torbjörn Sjöstrand, Computer Phys. Comm. 39 (1986) 347
Torbjörn Sjöstrand and M. Bengtsson, Computer Phys. Comm. 43 (1987) 367
http://www-h1.desy.de/icas/imanuals/pythia57.ps

[24] The Lund String Model, B. Andersson and W. Hofmann, Phys. Lett. 169B (1986) 364

[25] G. Altarelli and G. Parisis, Nucl. Phys. 126B (1977) 298

[26] Physical Review D, Particles and Fields, The American Physical Society through the American Institute of Physics, 1 July 1996, PART I, Review of Particle Physics, p.170

[27] D. Kant and G. Thompson, *Private Communication*

[28] M. H. Seymour, *Private Communication*

[29] http://www-h1.desy.de/h1/iww/iwork/iflow/analyses.html

[30] H1 Collaboration, Measurement of Charged Particle Transverse Momentum Spectra in Deep Inelastic Scattering, Nucl. Phys. 485B (1997) 3

[31] R. K. Griffiths, First Year Report, University of London (Unpublished), 1995

[32] S. D. Ellis, J. Phys. G17 (1991) 1552

[33] K. T. Donovan & G. Thompson, Algorithmic bias on observed hadronic jet properties, Zeit. Phys. C73 (1997) 641

[34] G. Thompson & D. Kant, Comparing Multiplicities in the Breit Frame and the Hadronic CMS, H1-08/95-452

[35] D. Kant & G. Thompson, Charged Track Multiplicity Distributions in the Breit Frame, H1-04/96-475

[36] K. H. Streng, T. F. Walsh and P. M. Zerwas, Quark and gluon jets in the Breit frame of lepton-nucleon scattering, Zeit. Phys. C2 (1979) 237

[37] Spline fitting routine HSPLI1, CERN Program Library Long Writeups Y250, HBOOK Reference Manual, Version 4.24

[38] Minuit Reference Manual, CERN Program Library Long Writeup D506, Version 92.1 (March 1992)

[39] Rev. Thomas Bayes, "An essay towards solving a problem in the doctrine of chances", Philosophical Transactions **53**, 370 (1763), Reprinted in Biometrika **45**, 296 (1958)

[40] Harrison Prosper, "Techniques and concepts of high-energy physics IX, 1996", ed. by Thomas Ferbel, NATO ASI Series (1997)

Acknowledgements

First and foremost, the biggest thanks of all must go to my parents, without whom ***none*** of this would ever have been accomplished. I am grateful for the support, both moral and financial, and the encouragement throughout all the years. I can't say any more than *Diolch yn fawr iawn.*

I would like to thank my supervisor, *Professor* Graham Thompson for his help, guidance and support over the course of my Ph.D. For providing the money to pay for the Ph. D. fees, and also for giving me a grant to survive and for paying the COLA in Hamburg, I must thank P. P. A. R.C. from the bottom of my heart! Others who were of invaluable (though not financial!) help to me over the years were "Burping" Dave Kant, Eram Rizvi, Thanos "Ddanki" Mavroidis, Gerry "Guitar" Lopez, Paul Dixon, Michael Hapke, Keith "What? Eh??" Donovan, Murrough Landon, Eric Eisenhandler and Tim Nicholls. Eric, Murrough and Tim deserve an extra mention for two reasons: firstly because of their superb and ever-ready help and tutillage when I was the UK expert for the SpaCal - especially the TDC System - and secondly because they, like me, are dedicated Apple Macintosh fans! I would like to thank Derek Newman-Coburn for his helpful tips on proper posture, hair care and women, and also for his expert wine knowledge - especially when it came to which ones to avoid in the *Stuttgarter Weinfest* in Hamburg! - Your friendship is sorely missed, Derek. Other members of the department who deserve a mention for making the working environment a friendly and easy-going one are Brian Morgan, Peter Crew & Ted Lee, Paul Harrison, Paul Kyberd, Alex Martin, Steve Lloyd, Simon Robins, Andy Yeaman, James King, Martin Evans, Dan Odoom, Caroline Hodd, Kostas Markopoulos, Dai Hughes, John Back, Dan Azzopardi, Dan Traynor and John Turney. Thanks to you all for never ever letting a dull moment get past our door!! Well, not without a fight anyway!

The members of *"The group with no name"*, sometimes referred to as BHAG, also deserve a special mention. Not only did they provide guidance when needed and offered helpful tips and hints, but they also sat through a stultifying number of my talks with unfailing diligence and a superhuman show of patience! You'll get your reward some day!! Extra-special thanks for answering the weirdest of my questions and for the software assistance (as well as for being Dani Behr's neighbour - can I come round for tea sometime?) goes to the shining star of the Rutherford Lab, Dave Sankey!

I was fortunate enough to spend a (considerable!) period of my Ph. D. in Hamburg, which I thoroughly enjoyed! Life in Hamburg was just great, both in and out of the lab. People at DESY who deserve a mention, are: Eram Rizvi, a veritable mine of useful information with whom I was glad to share an office; the somewhat mad-cap Paul "How did that happen?" Sutton who is the most entertaining and fun person ever to grace *Gebaüde 2g* and a top bloke to have around when on shift!!; Lee West, for lengthy discussions on F1 team tactics and the best video games for the office PowerMac, as well as the occasional work-related question!!; Kirstee Hewitt for her utterly incredible stories, her love of beards(!) and for being Welsh!; The SpaCal group deserve a medal for putting up with me! Thanks for the help and guidance to Gerhard Müller, Vincent Boudry and Ekaterini Tzamariudaki.

Thanks also to: Andy Mehta, Paul Newman and Julian Phillips - "The Leopard Story" will become legendary!; Stuart Cocks, Theo Holtom, "Little Dave" Milstead (who deserves special praise for actually showing some interest in my analysis!)l; Dave Cussans for sorting out most of my Monte Carlo queries; Chris Hilton, Ben Waugh and Molly Anderson who must surely deserve the Peace Prize for putting up with Pete Bispham and John Lomas! John's *unique* viewpoints and perspective on life were refreshing, and I have to thank Pete especially, as he was indirectly responsible for fixing my choice of experiment for a Ph.D. - thanks for the advice!; Jane Bromley, Dave Waters, James Tickner and Hugh Tallini.

I went on a fortnight's summer school to the NATO Advanced Study Institute on Techniques and Concepts of High Energy Physics *IX*, St. Croix, US Virgin Islands in the summer of '96 - thanks to NATO, PPARC and the department for getting me there! Not only did I have a fantastic time in a tropical paradise, met some fun and interesting people, but some of the concepts that I learned are actually used in this thesis!! Who could ask for more? I can't let talk of the ASI go by without a special mention of one of the participants, George Michail, an exceptionally gifted physicist destined for great things, whose life was tragically cut short by a drunk driver less than 6 months after the school.

During my stay in Hamburg, I attended German lessons at "der, die, das" (http://der-die-das.com or mail <language@der-die-das.com>) which I thoroughly enjoyed! If ever you're in Hamburg and looking for a German school, Der Die Das comes highly recommended!! Not only were my teachers, Friederike Timm, Nicola Thiermann and Daniela Kurtz, absolutely fantastic at their job, but they were and still are good friends. *Ich freue mich darüber, daß wir uns getroffen haben.*

Back in London, there was, of course, a considerable amount of socialising outside of laboratory, and it was these very people that were mostly responsible for keeping me sane over the last four (and some of you many many more...) years! There are just *sooooo* many people to thank for various combinations of keeping my spirits up, listening to me ranting, feeding me, and just being there when I needed them, that a complete catalogue would probably fill another volume of this thesis! Nevertheless, I must try! In no particular order of importance - purely a geographical one!!, and with most sincere apologies if you think you should be in here and you're not, the roll of honour looks something like this...

My oldest and truest friends from home, those terrible twins Damian and Jason Walford Davies, Meinir Evans, Huw Smith and Rhodri Davies for always being ready for a "swift one" before closing time and who are a great laugh on evenings out. Evenings out with these people didn't always occur in Aberystwyth, however, and have unfortunately occurred at home with ever-decreasing frequency over the last 4 years. That's not to say that we don't go out in other locations though...Rhodri's harp improvisation gigs in London (and the subsequent drinks!) are an unique experience and should be attended at least once per lifetime! Jason & Meinir have been excellent hosts in Bangor on many an occasion. The weekends have invariably led to much revelry and laughter (and a source of many a good story - most with plausible deniability! Well, that's my story and I'm sticking to it!) and provided a much-needed break from the "Big City Life" - *Diolch yn dalpe*! Damian deserves big thanks not only for inviting me to dine on high table at Balliol (and then inviting me back again!!), but also for introducing me to Sam "Kiss me" Kate Phillips who is a great laugh and has become a true friend and trusted confidante.

Two years of my stay in London were spent as a Senior Member at Commonwealth Hall - an excellent place to be a student and a top place for a great social life!! I thoroughly enjoyed my stay there, and I'd like to thank the warden, Terry Bishop, and his wife Georgina for giving me the job and also for their hospitality. I would like to thank the Bursar, Liz Parkin, and the assistant Bursar, Cheryl Trickett, for their friendship and ready help. Liz's generosity (especially when I was a visitor from Hamburg) was overwhelming and I am truly grateful. During my two excellent years at Commie, I met a staggering number of people, though only a handful of these are worthy of mention! Firstly, I would like to thank Kathy, Gareth, Selena and Dave - (some of) my fellow senior members at Commie - for not taking the job too seriously but for being seriously good laughs! Other Commie "inmates" that I would like to thank for their continuing friendship are Andy Taylor and Jennifer Pickurel (now Taylor!), Phil Dack, Lew Gold(water), Sally Edwards, Sam Weston and John Ingham, Steve Terrill, James Black and Andre Benz; all of whom are jointly responsible for my current well-being and sanity (though others would claim that they did a bad job on the latter!). Though primarily a student residence, the undergraduates moved out of Commie during the summer months. This led to an American summer school invasion! The Americans were (mostly) truly lovely, and I would like to send my special thanks to Kalie Bannon, Marcie Cowley, Maggie Frump, Sara Lidgey and Julie Slusher for turning good summers into great summers!

My social life in London was not based entirely in and around Commie however, and I have enjoyed singing with many choirs over the years. In the ULU Chorus, I would like to thank Tamsin and Kirstie (another set of twins!), Angela, Noelia, Casey and Whitney, Alexis, Amanda, Nick, Bill, Dave & Robin for being a good bunch of fun people to sing with! In the Gwalia Male Voice, thanks must go to both Bill Pritchard and John Evans who surely must be the only two people on earth to know and tell worse jokes than I do! From the 1998 eisteddfod choir, as well as from SWS (Social, Welsh & Sexy), I would like to thank Margaret Edwards, Betty Lynn and Nesta Wynne for "looking after me" so well!

If I forgot anyone, I do apologise, but the acknowledgments were getting long enough anyway!

That's all, folks!
Cue the music and let the fat lady sing!