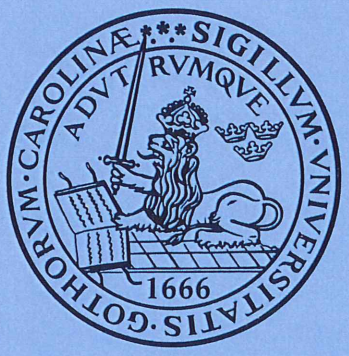
Studies of Various Aspects of the Proton Structure in Deep Inelastic Scattering at HERA and Identification of Quark and Gluon Jets

Thesis for the degree of Doctor of Philosophy in Physics
by

Morten Nyberg-Werther



DEPARTEMENT OF PHYSICS, UNIVERSITY OF LUND
LUND 1994

Studies of Various Aspects of the Proton Structure in Deep Inelastic Scattering at HERA and Identification of Quark and Gluon Jets

Thesis for the degree of Doctor of Philosophy in Physics
by

Morten Nyberg-Werther

DEPARTEMENT OF PHYSICS, UNIVERSITY OF LUND
LUND 1994

KRONTRYCK - ESLÖV

ISBN 91-628-1451-6
LUNFD6/(NFFL-7090)1994

Studies of Various Aspects of the Proton Structure in Deep Inelastic Scattering at HERA and Identification of Quark and Gluon Jets

By due permission of the faculty of mathematics and natural science at the University of Lund to be publicly discussed at the lecture hall B at the department of Physics, December 9 1994, at 10.15 a.m. for the degree of Doctor of Philosophy.

by

Morten Nyberg-Werther
Department of Physics, University of Lund, Sölvegatan 14, S-223 62 Lund, Sweden

This thesis is based on the following articles, included as Appendices A-D:

A The H1 Forward Muon Spectrometer,
Nucl. Instr. and Meth. A 340 (1994) 304-308

B Deep-Inelastic Scattering on Intrinsic Charm Quarks in the Proton,
Phys. Rev. D, vol. 47 nr. 11 1993, 4872-4882

C A measurement of the Gluon Density in the Proton at Low x,
H1 Note, 94-

D Jet Identification based on Probability Calculations using Bayes' Theorem
to be submitted to Zeitschrift für Physik

Contents

Abstract

This thesis is based on work done as a member of the H1 detector collaboration at DESY during 1989-1994 and consists of three main topics.
The possibility to probe the predicted intrinsic charm component in the proton by deep inelastic scattering is investigated in a simulation study, including HERA and a fixed target experiment. For existing experiments the predicted statistics are relatively small, but may still contribute to a solution of the issue.
A determination of the gluon density has been obtained in the region $5 \cdot 10^{-3} < x < 8 \cdot 10^{-2}$ by measuring the cross section of photon gluon fusion events in deep inelastic scattering with the H1 detector at the ep-collider HERA. This direct measurement of the gluon density was based on an integrated luminosity of 242 nb^{-1} and was performed in a kinematic region previously not accessible.
The problem of identifying jets at LEP and HERA has been studied. The identification was based on jet energies and framentation properties, but instead of working with the separation variables directly, these have been used to calculate probabilities for having a gluon (or quark) according to Bayes' theorem.

Preface

In my teenage years I eagerly read anything that had the least to do with science, space and the future. A lot of it was of course just cheap space operas and similar stuff, but still it managed to arouse a deeply felt interest in technology and science. Well, you can imagine what I felt some years later when I actually got a lot closer to what used to be fiction to me, and I started to look around for the phasers and the hyperspace-ships. Eventually I got over the initial disappointment and actually started to feel at home, and at least there were some slightly mad people around, just as my books had predicted. Now, when I'm a lot wiser, I realize how little I actually understand - modern physics is such a complex framework and a single person has little impact on it, with some exceptions of course. Will we ever understand the underlying picture? In "the hitch-hiker's guide to the galaxy" there is mention of a theory which states that if someone ever finds the meaning of life, the universe will immediately self-destruct and be replaced with something even more absurd. According to another theory, this has already happened.

This thesis summarizes my life in the years from 1989 to the present day (1994) and reports only the good results. I am extremely pleased that I actually got a chance to fulfil my boyhood dreams, not everyone get such a possibilty.

Morten Werther
Lund, October 1994

Chapter 1

HERA, the Machine and the Physics

The **Hadron-Elektron-Ring-Anlage**, HERA, at DESY in Hamburg is the first electron-proton storage ring ever built. In 1991 it was completed and in the spring of 1992 the first physics collisions were observed in the two HERA experiments H1 and ZEUS. The design energies of the colliding electron and proton beams are 30 GeV and 820 GeV respectively, but in the first years HERA was operated with an electron beam energy of 26.7 GeV thus giving a center-of-mass energy of 296 GeV. The time between bunch crossings is 96 ns allowing for a total of 210 bunches per ring, but in 1993 HERA operated with 84 paired bunches (10 in 1992), which gave beam currents typically of the order 10 mA. Some additional unpaired (pilot) bunches in each beam were used for background studies. Although conditions have improved even more in 1994, HERA has yet to reach its design luminosity of $1.5 \cdot 10^{31} cm^{-2} s^{-1}$ yielding an integrated luminosity of 100 pb^{-1} per year.

Experiments in deep inelasticc scattering (DIS) have played an important role in the understanding of the fundamental forces and the substructure of matter. At HERA, the proton structure functions and thus the valence quark, sea quark and gluon densities in the proton can be measured over a much larger kinematic range than ever before, especially in the interesting region of small momentum fractions. Detailed comparisons of various aspects of the hadronic final state, e.g. particle distributions and jet production, can be made with predictions of QCD models.

At very low momentum transfers, $Q^2 \rightarrow 0$, the exchanged photon is quasi-real. In this so-called photoproduction domain the partonic content of the photon can be investigated. Photoproduction reactions producing heavy quarks pose another challenging possibility for interesting physics, since they, for example, can be used to extract the gluon density in the proton.

The large accessible energies available at HERA also makes it a hunting ground for new physics. For instance searches for leptoquarks, leptogluons and excited leptons and quarks can be made.

1.1 Kinematics

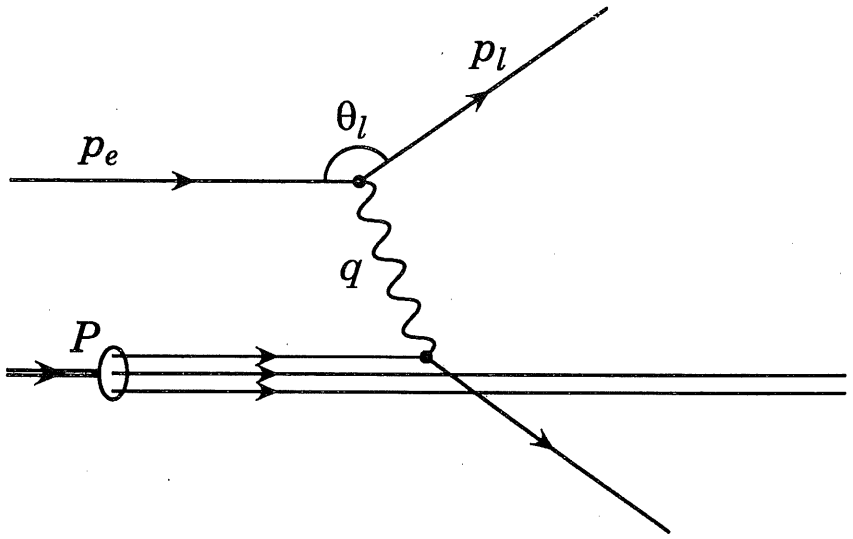


Figure 1.1: *The lowest order deep inelastic scattering process.*

In fig. 1.1 the inclusive deep inelastic scattering process $e + p \rightarrow e(\nu) + X$ is shown. Charged current events, where the exchanged boson is a $W^{\pm}$, and neutral current events in which the exchanged boson is a Z^0, are due to the larger masses of these bosons suppressed as compared to neutral currents with virtual photon exchange. Only at very large momentum transfers do these contributions become important. The kinematics of the basic process above can be determined from two independent variables. The three standard deep inelastic variables are:

- Q^2, the four-momentum transfer squared between the incoming and outgoing lepton carried by the exchanged photon.
- x, the Bjorken scaling variable which in the quark parton model (with massless partons) is interpreted as the fraction of the proton momentum carried by the struck quark.
- Bjorken y, which in the proton rest frame represents the fractional energy loss of the electron.

In Lorentz-invariant notation they are defined as:

$$\begin{aligned} Q^2 &\equiv -q^2 = -(p_e - p_\ell)^2 \\ x &\equiv \frac{Q^2}{2P \cdot q} \\ y &\equiv \frac{P \cdot q}{P \cdot p_e} \end{aligned} \tag{1.1}$$

where p_e, p_ℓ and P are the four-vectors of the incoming electron, the scattered lepton and the incoming proton respectively. Using the center-of-mass energy squared, $s = (p_e + P)^2$, and neglecting particle masses, the kinematic variables are related through the expression:

$$Q^2 = xys \tag{1.2}$$

The kinematic variables Q^2, x and y can be experimentally measured in terms of the incoming electron energy, E_e, and the scattered lepton energy, E_ℓ, and angle θ_ℓ:

$$\begin{aligned} Q^2 &\approx 4E_e E_\ell \cos^2(\theta_\ell/2) \\ y &\approx \frac{E_e - E_\ell \sin^2(\theta_\ell/2)}{E_e} \end{aligned} \tag{1.3}$$

and x is then given by eq. 1.2. In H1 notation the forward $(+z)$ direction is the direction in which the proton beam is moving, and polar angles (θ), such as the scattered electron angle, are measured with respect to this direction. The reconstruction of Q^2 based on the scattered electron is superior to all other available methods, but the reconstruction of y, for which the precision of the measurement basically deteriorates as $1/y$, cannot be accurately determined below $y \leq 0.05$. An alternative measurement of the kinematic variables is obtained from the hadronic final state according to the Jaquet-Blondel method [1]. For instance y can be measured by the relation:

$$y_h = \frac{1}{2E_e} \sum_{hadrons} (E_h - p_{z,h}) \tag{1.4}$$

Using mixed methods or other schemes like the double angle method [2], the accessible kinematic region is widely increased due to the complementarity of different methods, see fig 1. 2.

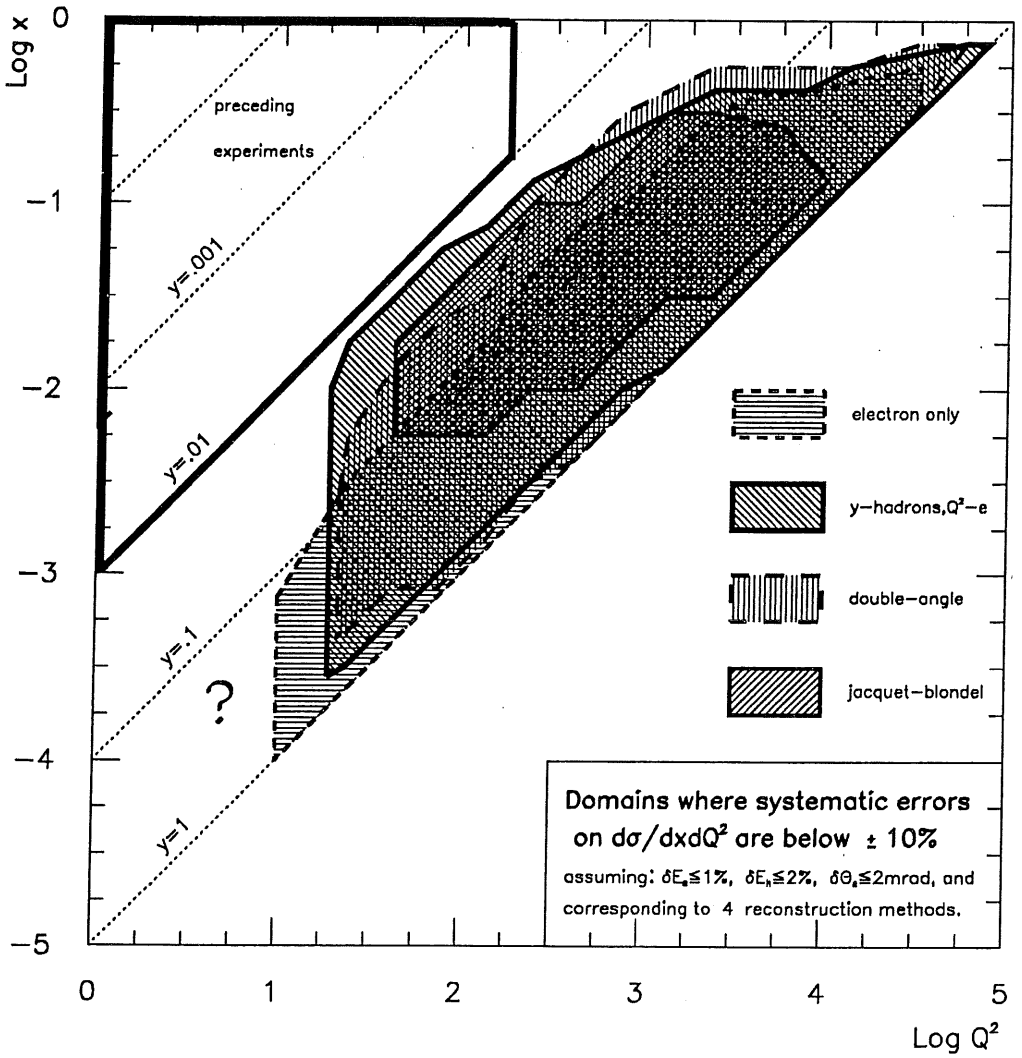


Figure 1.2: *The shaded regions indicate where in the x, Q^2 plane different kinematic reconstruction methods can be used with reasonable errors on the measurement of the differential cross-section [3]. The area in the upper left corner indicates the range accessed by previous fixed target experiments.*

Chapter 2

The H1 detector

To allow for a good measurement of the event kinematics, one of the main aims of the H1 detector [4] is to give a clear identification of the scattered electron and a good measurement of its energy. Furthermore, it is also important to have a good homogeneity for energy flow measurements and measurement of the event kinematics based on the hadronic system. In order to obtain this, the most sensitive detector components were placed inside a large superconducting magnetic coil so that the amount of dead material in front of them was minimized. Inside the coil we therefore find the electromagnetic and hadronic liquid argon calorimeter, and closest to the interaction region, the high resolution tracking system. Outside the coil is the instrumented iron yoke and, in the forward (proton) direction, the forward muon system. The general design concept does not differ much from the $e^+e^-/p\bar{p}$ detectors, but the imbalance of the beam energies, which leads to a boosted centre-of-mass, gives the need for an asymmetric detector. Thus the H1 detector is considerably more massive and has more instrumentation in the proton direction. This is apparent from fig. 2. 1 which shows a longitudinal cut through the detector.

2.1 Detection of the scattered electron

Scattered electrons from DIS are detected in the backward electromagnetic calorimeter (BEMC) for $Q^2 < 100GeV^2$ and in the liquid argon calorimeter for larger Q^2 values, where the scattering angle is larger. The BEMC is a lead-scintillator sandwich calorimeter with photodiode read-out and an energy resolution of $\sigma/E \approx 10\%/\sqrt{E} \oplus 3\%$. The backward proportional chamber (BPC) located in front of the BEMC provides the angular measurement of the electron together with track and vertex position data determined from particle tracks in the central tracker. In the liquid argon calorimeter, the electron is identified based on the transverse and longitudinal shower shapes and a link to a charged track, and here the resolution is approximately $\sigma/E \approx 12\%/\sqrt{E} \oplus 1\%$.

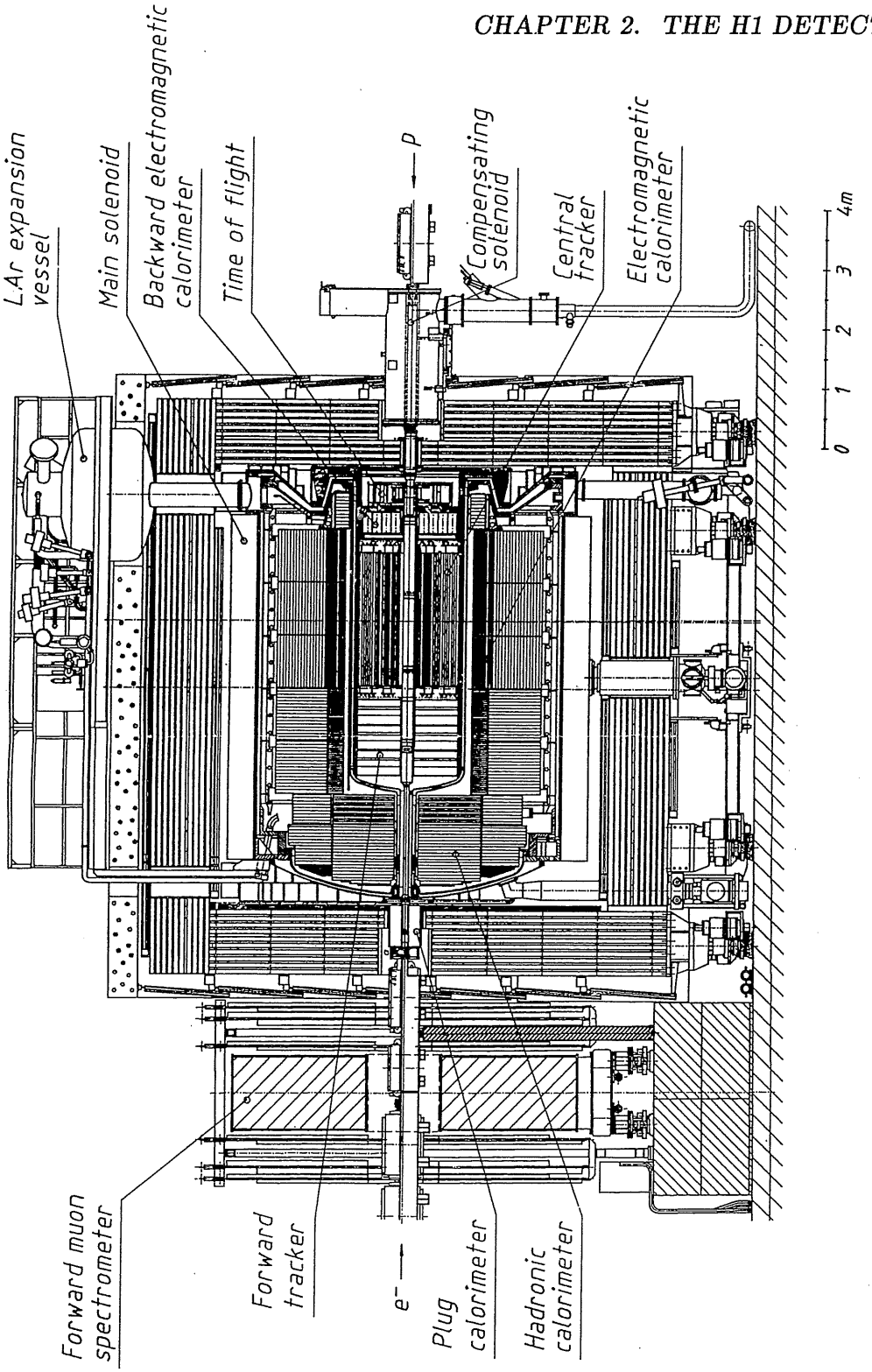


Figure 2.1: *Longitudinal cross section of the H1 detector.*

2.2 Detection of hadrons and jets

Hadrons and jets are mainly detected in the liquid argon calorimeter, which covers the angular range $4° \leq \theta \leq 153°$. Mainly the hadronic calorimeter, but to some extent also

the electromagnetic parts, provide the energy measurement of hadrons with a resolution which has been determined in a test beam to be $\sigma/E \approx 50\%/\sqrt{E} \oplus 2\%$. Its depth in terms of absorption lengths ranges between 4.5 and 8 depending on the polar angle.

2.3 Tracking

The tracking system provides triggering, track reconstruction and also particle identification. Due to the asymmetric electron and proton beam energies there will be many particles produced at small polar angles, and therefore the tracking system has been divided into two mechanically distinct detectors, the central tracking detector (CTD) and the forward tracking detector (FTD), see fig. 2. 2.

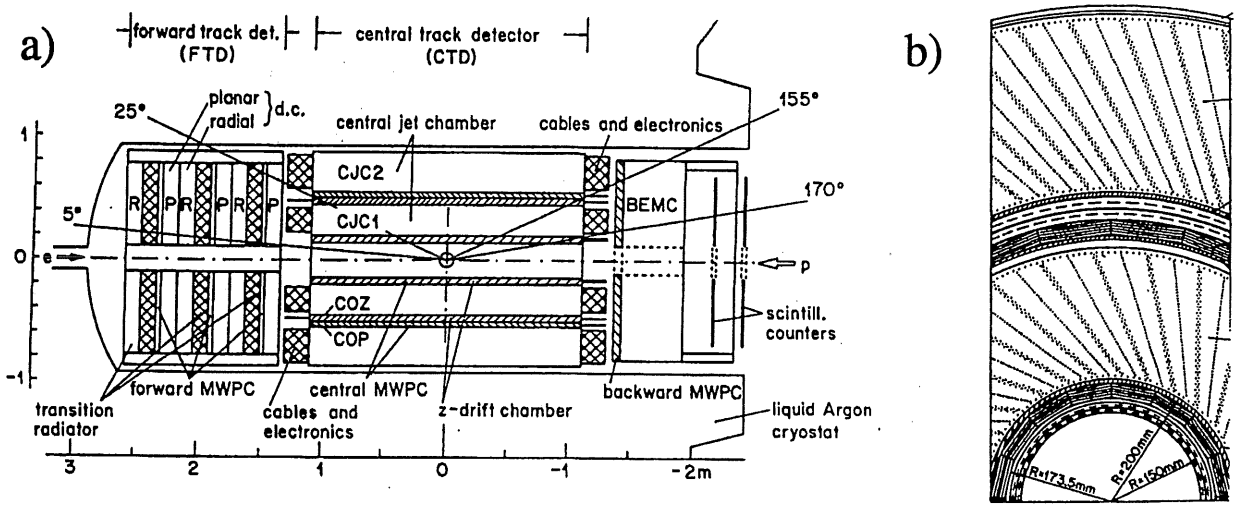


Figure 2.2: *a) The central and forward tracking detectors shown in a cut along the beam and b) the central tracker in a perpendicular view.*

2.3.1 Central Tracking Detector

The CTD roughly covers the angular range $25° \leq \theta \leq 155°$ and consists of two large jet drift chamber modules, CJC1 with a radial dimension of $20.3cm \leq r \leq 45.1cm$ and CJC2 with $53cm \leq r \leq 84.4cm$. Wires are strung parallel to the beam axis and the wire planes are tilted about 30° with respect to the radial direction (see fig. 2. 2b) to resolve the left-right ambiguity. Transverse particle momenta can be determined with a precision of $\sigma_p/p^2 \approx 0.003$. The z-coordinate of a hit is measured using the method of charge division, giving a resolution of $\sigma_z \approx 22$ mm. The measurement of the specific energy loss, dE/dx, is used to provide particle identification. Inside and outside of CJC1 there are two thin drift chambers, the central inner (CIZ) and outer (COZ) z-chambers, which determine the z-coordinate considerably better than the CJC ($\sigma_z \approx 350\mu$m) and thus complement the

measurement of charged track momenta in the jet chambers. There are also two multiwire proportional chambers, the CIP and the COP, which are mainly used for triggering and can therefore have a coarser pad read-out.

2.3.2 Forward Tracking Detector

Charged tracks at polar angles below $\sim 25°$ and above $\sim 155°$ do not traverse the full bending plane radius of the magnetic field. In the forward region the loss of transverse track length is instead compensated for by having a high density of accurate space points. The forward tracking detector, with an acceptance between $6° \leq \theta \leq 25°$, consists of an assembly of three nearly identical supermodules. Each supermodule includes (in increasing z) three different orientations of planar drift chambers for measuring θ, a multi-wire proportional chamber (FWPC) for fast triggering, a passive transistion radiator followed by a radial wire drift chamber which provides $r\phi$ information and measures the transition radiation, thereby enhancing the particle identification obtained from measuring dE/dx.

2.3.3 Scintillators

The time-of-flight (ToF) system, two plastic scintillator hodoscopes placed in the electron direction at $z \approx -2m$ from the interaction point, is used for separation at the trigger level of genuine ep events from beam-wall and beam-gas interactions occuring upstream from the detector. In addition, there are two double scintillator walls further away from the interaction region, one at $z = -6.5m$ and the other covering the near beam area at $z = -8.1m$. These, so-called veto-walls, are used for background rejection, but also to monitor rates after filling and during runs.

2.4 Calorimetry

The trackers are surrounded by the liquid argon calorimeter (LAC) covering the range $4° \leq \theta \leq 153°$ and segmented into an electromagnetic part and a hadronic part. To cover the very forward region between the beam-pipe and the LAC there is a compact calorimeter (the PLUG) and the backward region is covered by an electromagnetic calorimeter (the BEMC). Outside these, the instrumented iron acts as a tail catcher system to provide a rough measurement of "leaking" hadronic energy.

2.4.1 Liquid Argon Calorimeter

The LAC is segmented along the beam axis in eight self-supporting modules, "wheels", housed in one large cryostat. Each of the six barrel wheels are segmented in azimuthal angle in eight identical octants, see fig. 2. 3, called "stacks", and each stack is divided into an electromagnetic (EMC) and hadronic section (HAC). The hadronic stacks are made of welded stainless steel absorber plates with high voltage planes glued to the sides. Between the steel plates are the gaps for the active medium (liquid argon) and a readout board

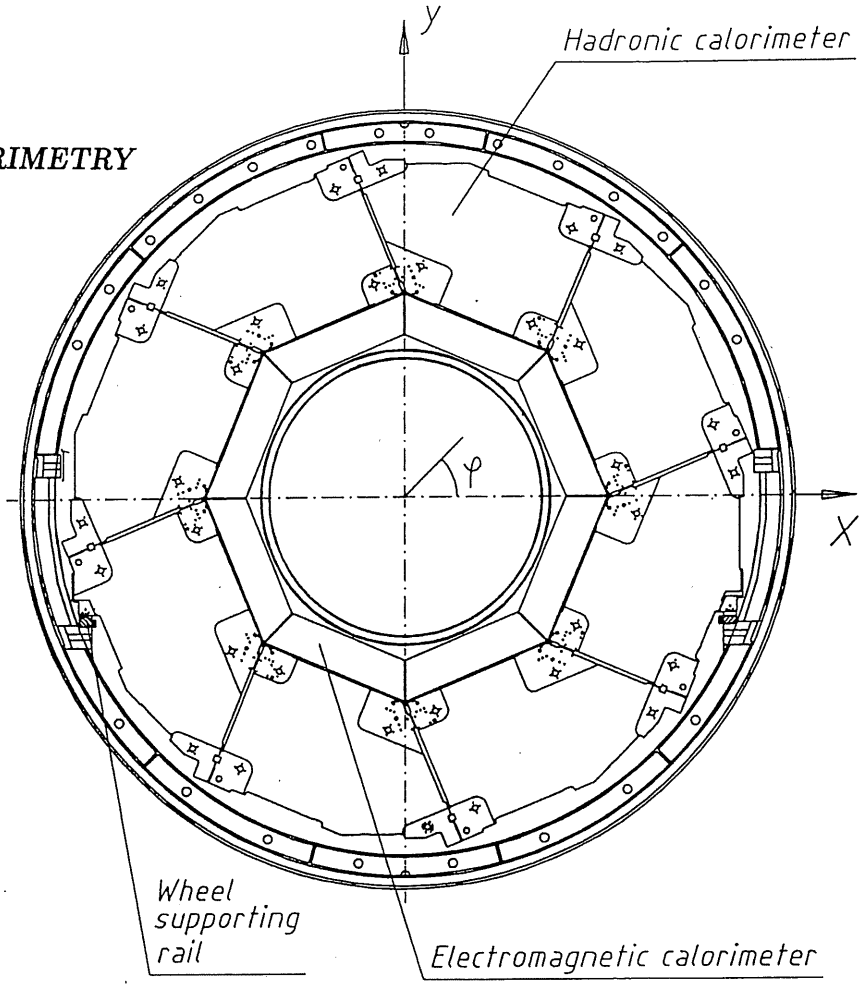


Figure 2.3: *Transversal cross-section of the liquid argon calorimeter.*

with pads on both sides to collect the charges deposited in the gaps. The orientation of the absorbers is such that particles are incident with angles not smaller than 45°. Onto (inside) the structure defined by the HAC the EMC stacks are mounted, consisting of a sandwich arrangement of G10 (epoxy+fiberglass)-lead-G10 separated by spacers to define the liquid argon gaps. The thickness of the EMC in terms of interaction lengths (X_0) varies between 20 and 30, and for the whole LAC the depth is between 4.5 and 8 interaction lengths (λ).
Precalibration with π, e and μ test-beams at CERN was done for all the different types of calorimeter stacks, and great care was taken to ensure that the calibration constants thus obtained could be transfered to the final H1 modules. To check the precalibration one has for instance used the three central barrel wheels to study cosmic muons and in that way the electromagnetic energy scale was verified to ±8%. In DIS where both the scattered electron and the hadronic jet are detected in the LAC, a direct comparison between the hadronic and electromagnetic energy checking the $p_\perp$ balance can be made.

2.4.2 Backward Electromagnetic Calorimeter

The backward electromagnetic calorimeter, BEMC, is a conventional lead-scintillator sandwhich calorimeter mainly designed for measuring the energy and direction of small angle scattered electrons from DIS. It covers scattering angles from $151° \leq \theta \leq 177°$ and has full azimuthal acceptance in that range. Its 88 calorimeter stacks, 56 with a quadratic

cross-section (16 by 16 cm^2) and the remaining 32 with a trapezoidal or triangular shape, are mounted parallell to the beam. Each stack consists of 50 active sampling layers of plastic scintillator interleaved with 49 layers of lead, the total depth being 22 X_0. Initial calibration was performed with $5GeV$ electrons at the DESY synchrotron and a $120GeV$ muon beam at CERN. During *ep* collisions, the elastic peak in the scattered electron energy spectrum given by the BEMC is compared to the result of a simulation program, taking all detector effects into account, to check and improve calibration.

2.4.3 Plug Calorimeter

The plug calorimeter (PLUG) was designed to close the gap of acceptance for energy flows between the beam pipe ($\theta \sim 0.6°$) and the forward part of of the LAC ($\theta \sim 3°$). Due to geometrical limitations, this is a very compact calorimeter consisting of nine copper absorber plates interleaved with eight sensitive layers of silicon detectors and its diameter is $0.7m$. The energy resolution, affected by coarse sampling and leakage, is $\sigma/E \approx 150\%/\sqrt{E}$.

2.4.4 Tail Catcher

To measure the hadronic energy leaking out of the LAC, eleven of the sixteen limited streamer tube layers of the instrumented iron are equipped with readout pads. The pads vary in size from $30 \times 30cm^2$ in the endcaps up to about $50 \times 40cm^2$ in the barrel region. Calibration is done using cosmic muons traversing opposite pairs of endcap or barrel modules, and the general hadronic energy constant was given by test beam measurements at CERN where also the energy resolution was measured to be $\sigma/E \approx 100\%/\sqrt{E}$.

2.5 Muon System

The dominating source of muons will be semileptonic decays of heavy quarks. This means that it will be important that the muon system can measure muons within jets, and in H1 this is accomplished by having calorimeters absorbing the hadronic energy and a surrounding system to measure the single ionizing tracks from muons. In the instrumented iron there are a total of sixteen streamer tube layers, a triple layer both in front and after the iron and the remaining layers in between the ten iron sheets in the yoke. The spatial resolution of a few millimeters roughly matches the multiple scattering in the material in front of the muon system and the momentum resolution is $\sigma/p \approx 35\%$. In the barrel region there is a threshold energy of 1.2 GeV for a muon to reach the iron, 2 GeV to penetrate it, and in the forward direction the threshold is 2.5 GeV. In the very forward region between $3° \leq \theta \leq 17°$ there is an additional forward muon spectrometer (see appendix A). Here, driftchambers are placed on both sides of a toroid magnet (average field of 1.6 T) which leads to a resolution of $24\% \leq \sigma/p \leq 36\%$ for muons in the momentum range $5GeV \leq p \leq 200GeV$.

2.6 Luminosity System

The luminosity system has several functions; it provides a fast relative luminosity measurement, electron beam monitoring for the HERA machine, an absolute luminosity measurement in the interaction region with a precision of 5%, tagging of photoproduction events and energy measurement for very small angle scattered electrons and radiative photons. The luminosity measurement is based on the rate of Bethe-Heitler events, $e+p \rightarrow e+\gamma+p$, a process which has a large and calculable cross-section. The luminosity monitor detects scattered electrons and outgoing photons in coincidence and therefore consists of two arms, the electron tagger and the photon detector. Since the angular distributions for both e and γ are strongly peaked in the direction of the primary e-beam, the detectors are placed close to the beamline and very far from the interaction region.

2.7 Trigger System

The trigger system is designed to select interesting ep collision events and reject the outnumbering background, of which there are three basic types: synchrotron radiation from the e-beam, proton gas interaction and stray protons which produce particle showers by hitting the beamtube or other apertures around the accelerator. Since there are 96 ns between bunch crossings, there must be very low deadtime in the system, and therefore a pipelined system to keep the detector information stored during first level trigger calculations was chosen. Most of the subdetectors produce trigger information which enables a first level trigger decision based on main physics quantities, but to allow decisions of increasing complexity a multi-level concept is used. Thus the deadtime free level 1 trigger is followed by two levels of triggers operating during the primary deadtime of the front end readout, one hardware trigger and one software trigger, and finally an event filter system with access to the full event information operating on a fast processor farm.

2.8 Data Handling and Simulation

Data from the H1 experiment is transfered via a fast link to the IBM ES 9000 mainframe on the DESY site. The mainframe stores and manages raw, simulated and reconstructed data as well as the data summary tapes (DST) for physics analysis. Furthermore, two Silicon Graphics computers are used for number crunching tasks, for instance data reconstruction during data taking with a speed that matches the data logging rate of 5 Hz.
In all H1 software, the dynamic memory management package BOS is used together with the FPACK package for input/output and data transfer. A complete detector simulation program based on the GEANT framework including the full geometry of the detector and the beamline within $\pm 100m$ around the interaction region has been developed. For graphics applications, like the H1 event display program (see fig. 2. 4), a system called LOOK is used. The physics analysis is simplified by the general tool package named

H1PHAN, where the complicated underlying bank structures of data are accessed via simple buffers called Q-vectors.

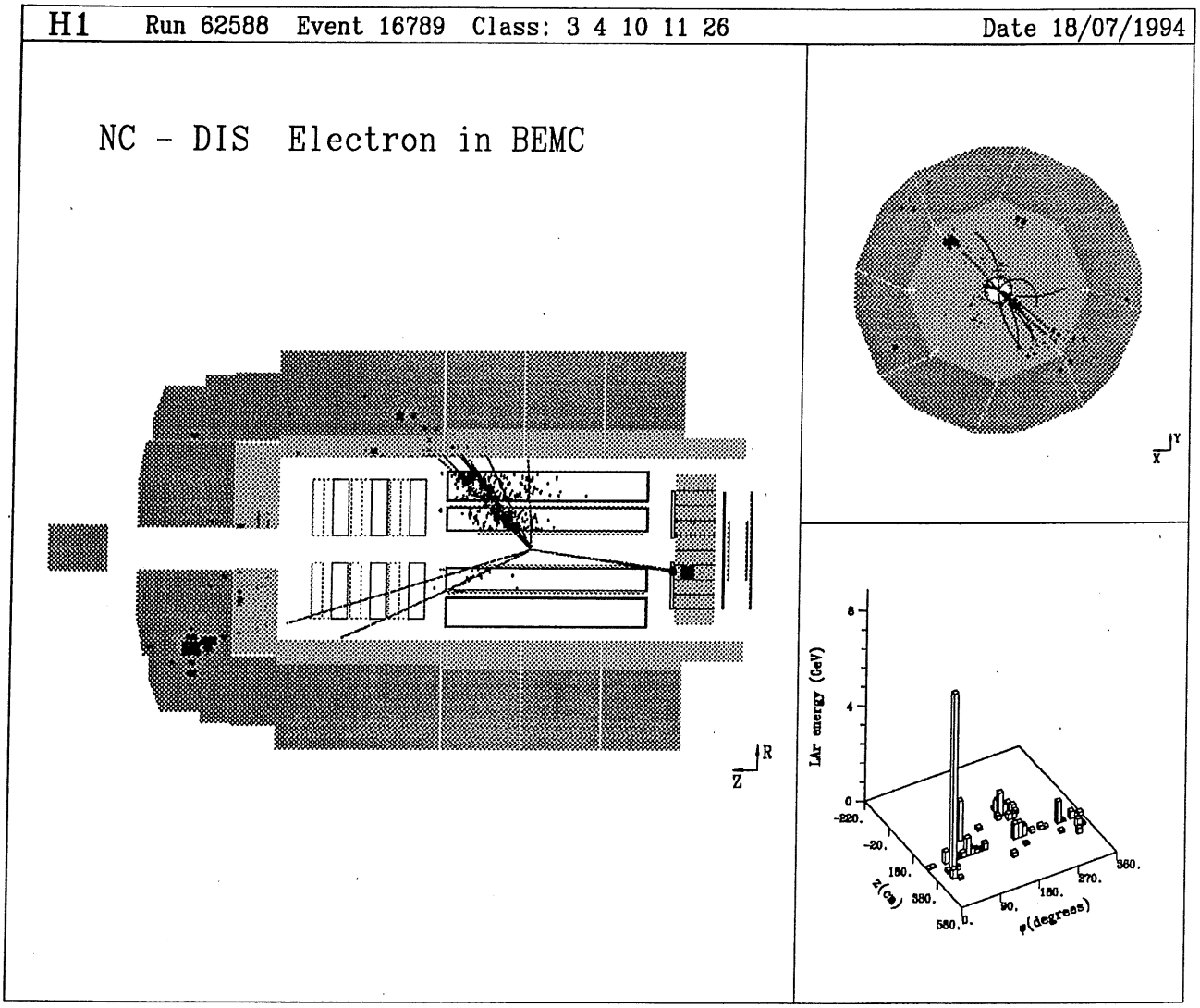


Figure 2.4: *Example of a neutral current event (candidate 2+1 jet event) displayed with the H1 event display program*

Chapter 3

Intrinsic Charm

The simulation study presented in this chapter is based on the paper in appendix B. Since no new efforts have been put into the subject, it is reviewed in the form of a summary. In the paper, the possibility to study the intrinsic charm (IC) hypothesis at HERA and at a fixed target experiment is investigated. Backgrounds to this process are large and therefore ways to suppress them are suggested.

3.1 Introduction

Inspired by the large observed charm cross-section in hadron collisions it was suggested that there could exist a non-perturbative $c\bar{c}$ component in the proton [5]. The probabilty to find a so-called intrinsic charm quark in the proton with a given momentum fraction x was proposed to be of the form (with a 1% nomalisation):

$$c(x) = 18x^2\{\frac{1}{3}(1-x)(1+10x+x^2) + 2x(1+x)lnx)\} \tag{3.1}$$

From this expression it can be computed that the intrinsic charm quarks generally have large momentum fractions, with a mean of $\bar{x} \approx 2/7$. This characteristic property of IC can be intuitively understood by viewing the proton in an infinite momentum frame where the five quarks in the $|uudc\bar{c}\rangle$ state are moving together with the same velocity, and the charm quarks with their larger mass consequently must take larger momentum fractions. It is important to distinguish between intrinsic quarks and the normal sea quark pairs generated by large momentum transfer processes in perturbative QCD. The latter is a short- lived fluctuation carrying only small momentum fractions of the proton, whereas intrinsic quarks exist over a time-scale independent of any probe momentum and are associated with the bound state dynamics of the proton.
Deep inelastic scattering (DIS) can be used to probe both of these proton components. The dominant source of charm production is the photon gluon fusion (PGF) process $\gamma g \rightarrow c\bar{c}$, and this will be the most severe background to scattering on intrinsic charm quarks. In the Monte Carlo event simulations, muons from semileptonic decays were used to tag the charm quark. To these muons, we have also considered the background

source of muons from decaying pions and kaons, since their large production cross-section outbalance the small decay probability.

3.2 Event Simulation

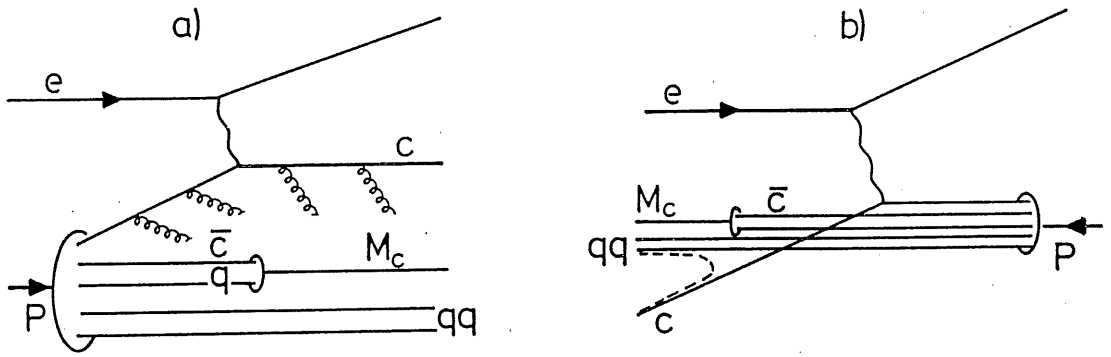


Figure 3.1: *Deep inelastic scattering on an intrinsic charm quark illustrating (a) the simulation model and (b) the typical event topology in the HERA lab frame.*

The deep inelastic scattering event generator LEPTO 5.2 [6] was used to simulate complete intrinsic charm scattering events. As input distribution we used (3.1) normalized as it is to a 1% probability for intrinsic charm, but to include a proper Q^2-evolution we introduced a correction to it by a parametrization of the results in [9]. (As has been shown later in [10] this correction may have been overestimated.) Hadronization was performed in the Lund string model as implemented in JETSET [7]. In IC events, the string is spanned between the scattered charm quark (alt. antiquark) and a diquark (quark) in the remnant. In addition to the string system, the remnant also contains the partner charm antiquark (quark) which forms a charmed meson (baryon) with the remaining quark (diquark), see fig 3. 1. The photon gluon fusion background was simulated with AROMA [8] based on exact matrix elements for heavy quark production in leading order. Finally, to estimate the muon rate from pion and kaon decays in normal DIS events, LEPTO was used together with version 7.3 of JETSET which allows for particles to decay within a volume defined by the user.

3.3 Intrinsic Charm at HERA

The two HERA experiments H1 and ZEUS cover most of the solid angle, except for the regions around the beampipe, and we have in our simulations chosen to make an accep-

tance cut of 4° both in the eletron and proton beam direction. As previously explained, the dominant signal for IC will be in the large x domain ($x \gtrsim 0.1$). This will clearly lead to problems in terms of the reconstruction of the kinematical variables. As can be seen in fig. 3. 2a, the scattered electron energy spectrum is dominated by the kinematic peak, the energy of the scattered electron is $\sim 30\ GeV$ in this x-region at $Q^2 < 100\ GeV^2$, which means that (Bjorken) y will be poorly reconstructed using electron variables. This can be somewhat improved by moving towards larger Q^2, but on the other hand this will also lead to a reduction of the cross-section from the $1/Q^4$ propagator-factor in the neutral current cross-section expression. The muon from the semileptonic charm decay will essentially emerge at the same polar angle as the scattered charm quark, which is also the current jet direction. In fig. 3. 2b we see that in the x-region of interest this will put us in conflict with the detector acceptance. If it was possible to lower the polar angle cut there would also be a substantial rate increase.

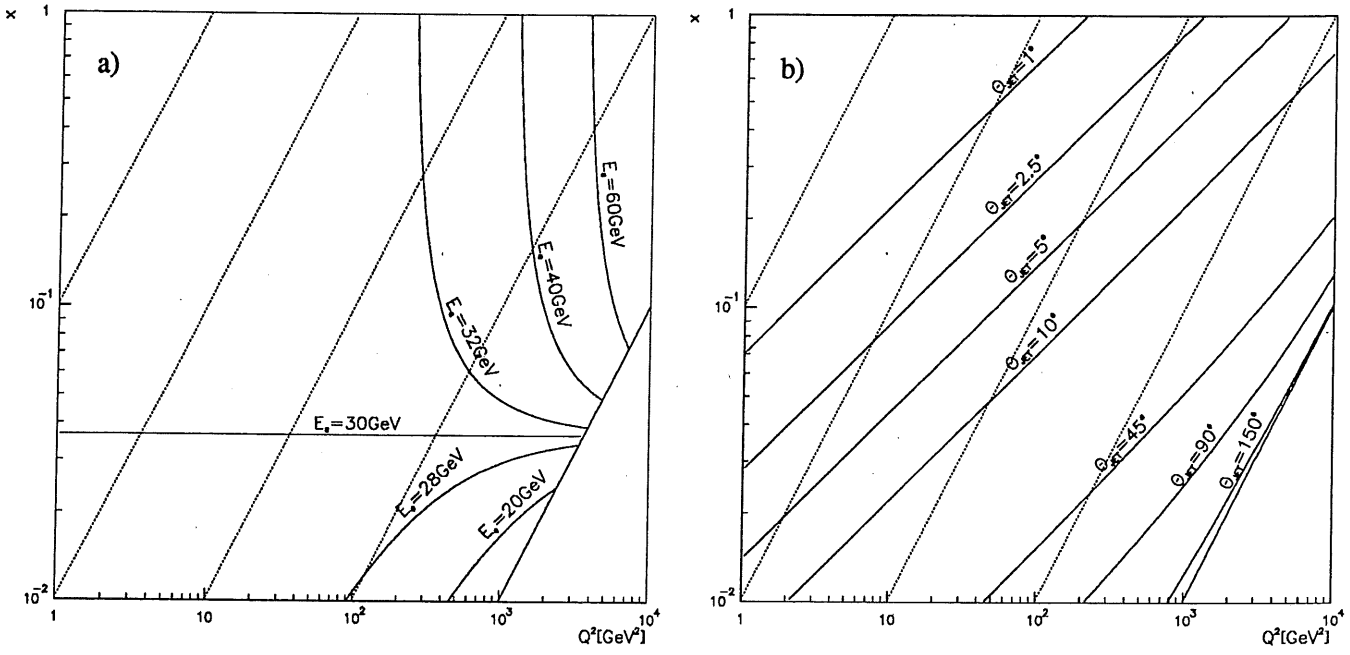


Figure 3.2: *(a) Isolines for constant scattered electron energies and (b) isolines for constant polar angles of the current jet, i.e. the scattered quark in the kinematic x, Q^2 plane for HERA. Also shown: lines of constant y (dotted) from left to right $10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}$, and full line $y = 1$.*

Specifying the requirements $4° \leq \theta \leq 176°$ and $E_\mu > 4GeV$ for a muon to be correctly identified, we end up with a signal cross-section of ≈ 1.1 pb ($\sigma_{tot}^{IC} \approx 200$ pb) for IC and a background of ≈ 1.9 pb from PGF in the region $0.1 \leq x \leq 0.5$, $10^1 \leq Q^2 \leq 10^3$. Further cuts can be imposed to improve the signal/background ratio. Noting that there

in general will be an extra charm quark jet in PGF events and that this will give rise to extra activity in the detector, we were able to define additional cuts and could thus reach a signal/background ratio of ~ 2 with only a 20% reduction of the signal cross-section. The background from decaying pions and kaons is unfortunately not so easily mastered. It could be argued that once the pion and kaon momentum spectra are measured well enough this background can be subtracted, since both the decay and the basic DIS events from which they stem are well-known processes that could be controlled. With HERA operating at the full design luminosity of about $100pb^{-1}$ per year, there would be approximately 100 IC events collected in one year for a 1% component of IC in the proton.

3.4 Intrinsic Charm at Fixed Target Experiments

The problem with the limited acceptance for small angle scattered intrinsic charm quarks at HERA can generally be reduced at a fixed target experiment. Here, the beam hole in the detector can be significantly smaller and there is more freedom with respect to the arrangement of individual detector components. In the simulations we have assumed conditions as at the Fermiliab muon beam with $E_\mu \approx 470GeV$, giving $\sqrt{s} \approx 30GeV$, so that the total cross-section not will be much affected by the charm quark mass threshold. Thus, the total IC cross-section is only 30% lower than at HERA.

The μp center-of-mass system is strongly boosted along the the muon beam direction and as a consequence both the scattered muon and the current jet will come out at small angles. In the kinematic region of interest ($0.1 \leq x \leq 0.5$, $10^1 \leq Q^2 \leq Q^2_{max}$) this means that the polar angle of the scattered muon is varying between 0.4° and 5° and the current jet between $\sim 1°$ and 20° (see fig. 3. 3). Assuming that the detector covers the angular range 0.4° to 10° and requiring that muons have a minimum energy of $10GeV$ in order for them to be detected, we are left with an observable IC cross-section in the above-mentioned x, Q^2-region of ≈ 5 pb compared to ≈ 3 pb for the background from PGF. Although this cross-section is larger than what could be obtained at HERA by a factor ~ 4, one must remember to take the luminosity into account. Using the Fermilab experiment E665 [11] as an example, we estimated the integrated luminosity for a typical target to be of the order $15pb^{-1}$ per year. Since the E665 experimental program included studies of hadronic final states, the hadron filter and muon detector was placed far away (26 m) from the target. Hence, background from pion and kaon decaying to muons is large and the acceptance for high energy muons is limited to $\theta \lesssim 3°$. In a dedicated IC experiment one could place the hadron filter close to the target to minimize such background and put the spectrometer at a suitable distance such that the acceptance for high energy muons would be at least $0.4° \leq \theta_\mu \leq 5°$. Furthermore, with a solid target the luminosity could be increased with roughly an order of magnitude. With such an experimental situation one could get more than 500 IC events per year (assuming a 1% IC content), with a chance to probe the IC component to the 0.1% level.

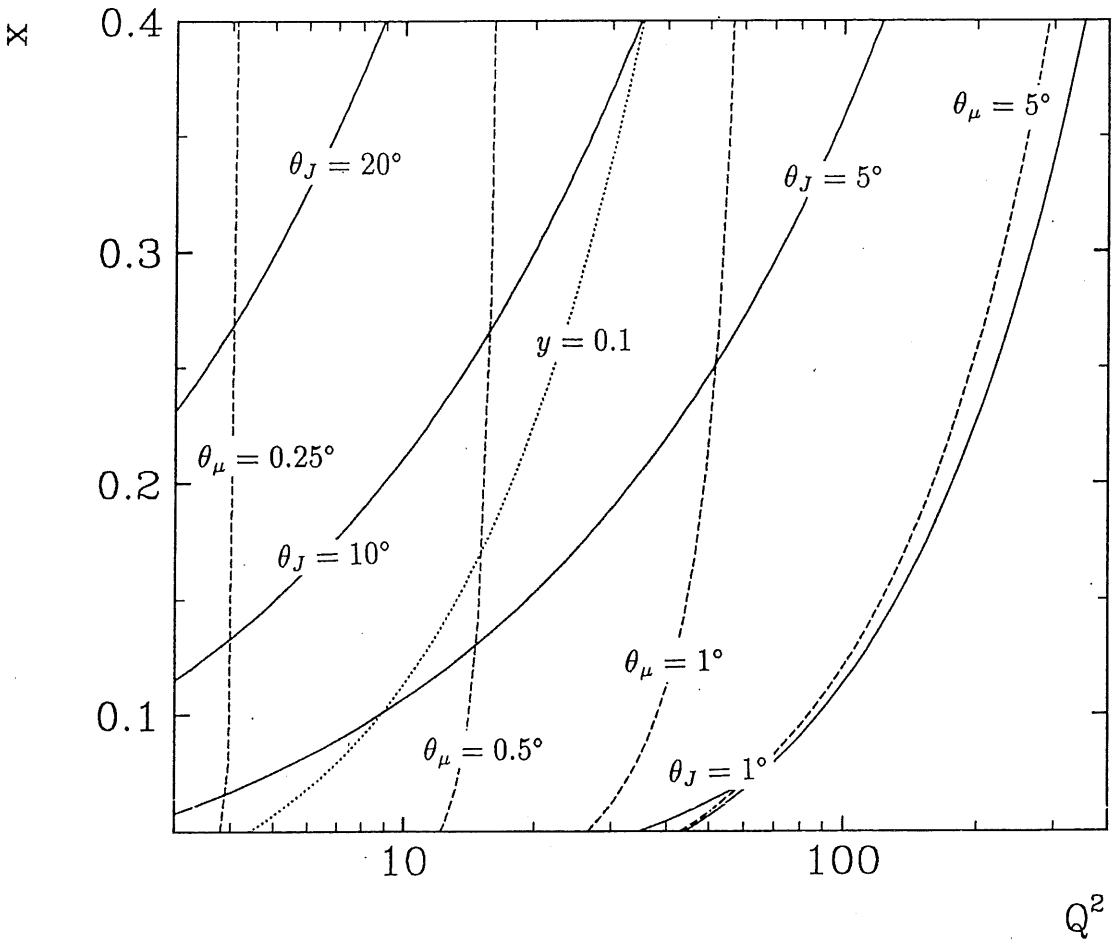


Figure 3.3: *Isolines for constant polar angles θ_J of the current jet, i.e. the scattered quark, (full lines) and θ_μ of the scattered muon (dashed lines) in the kinematic x, Q^2 plane for DIS of 470 GeV muons on a fixed nucleon target. Lines of constant y are also shown.*

Chapter 4

The Gluon Density in the Proton

This chapter is a summary of the paper in appendix C in which the details of the analysis are described. By measuring the cross-section for 2+1 jet events in deep inelastic scattering at HERA, we have calculated the corrected cross-section for photon gluon fusion at small momentum fractions, x, of the proton, or more precisely, in the range $0.005 < x < 0.08$. We have also estimated the gluon density in this range and note that it is strongly increasing with decreasing x.

4.1 Introduction

Why do we study structure functions? One answer lies in the fact that high energy lepton-hadron and hadron-hadron interaction cross-sections due to the QCD factorization theorems can be calculated as a sum of integrals convoluting fundamental parton interaction cross-sections $\hat{\sigma}$ with universal parton distributions. Thus, by measuring structure functions we can estimate the production rates of different hard processes. Another answer lies in the possibility for finding a new interesting QCD dynamics, expected to become apparent at x of $\sim 10^{-3}$. For instance, the evolution of the gluon density, g(x), with the Lipatov equation suggests that the gluon density behaves as $xg(x) \sim x^{-\lambda}$ when $x \to 0$ with λ typically ~ 0.5. Such an increase in $xg(x) = G(x)$ cannot go on indefinitely as x decreases. If the density of gluons becomes too large they can no longer be treated as essentially free, and new effects like screening or shadowing could be manisfested.

The proton structure function F_2 enters in the Born cross-section for the inclusive deep inelastic scattering process according to:

$$\frac{d^2\sigma}{dx\,dQ^2} = \frac{2\pi\alpha^2}{xQ^4}(1+(1-y)^2)F_2(x,Q^2) \tag{4.1}$$

neglecting longitudinal contributions and Z^0 exchange, since they are normally small in the kinematic region studied at HERA so far. The structure function F_2 itself can further be expressed in leading order at small Q^2 as:

$$F_2(x,Q^2) = \sum_q e_q^2 x f_q(x,Q^2) \tag{4.2}$$

From this we see that the distribution of quarks in the proton, $f_q(x,Q^2)$, in principle could be determined by measuring values of F_2 (and the other structure functions which were neglected above). Parton distributions are typically determined from global fits not only to DIS data, but also to for instance neutrino scattering and hadron-hadron collision data. In eq. (4.2) we have already introduced scaling violation; the structure function has an explicit dependence on the probe momentum transfer Q^2, and this has its origin in the evolution of the "bare" parton densities. The sum of all possible chains connecting the initiating parton and the "struck" quark leads to a power series in $\alpha_s(Q^2)\log Q^2$ governed by the Altarelli-Parisi equations [13]. Through this evolution there is also the possibility for a struck quark to have a gluon "ancestor" and one can therefore study the gluon density by measuring the scaling violations of F_2, as in [12]. An alternative approach, which can be applied at HERA, is to measure the gluon density directly from the rate of the $\mathcal{O}(\alpha_s)$ gluon initiated process photon gluon fusion (PGF). We may symbolically express the relation between the PGF cross-section and the gluon density in the following way:

$$\sigma_{PGF} \sim \int dx_g G(x_g,Q^2)\hat{\sigma}_{LO} \tag{4.3}$$

where $\hat{\sigma}_{LO}$ is the $\gamma g \to q\bar{q}$ parton cross-section.

4.2 Kinematics

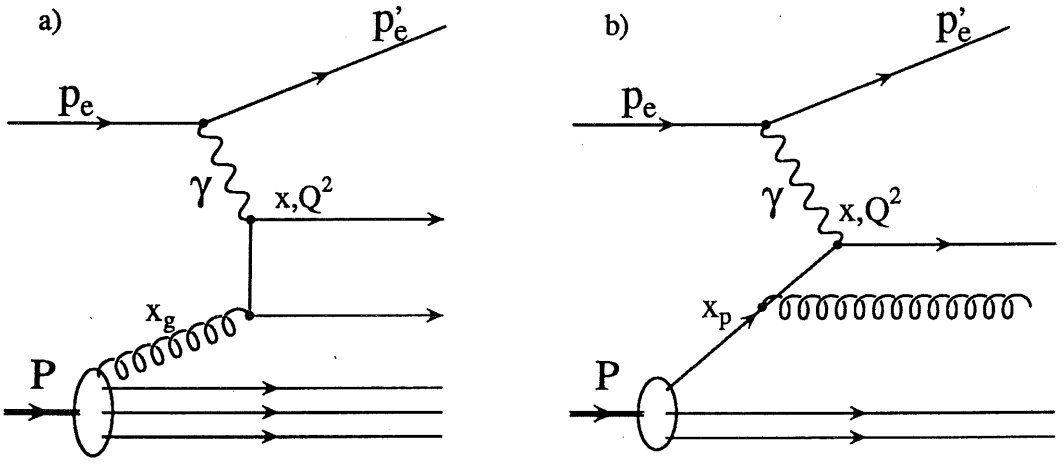


Figure 4.1: *a) The gluon induced photon gluon fusion process and b) the quark induced QCD-Compton process with relevant kinematic variables indicated.*

For the analysis of the gluon density at HERA, there are two basic $\mathcal{O}(\alpha_s)$ contributions to consider, PGF, fig. 4. 1a, and QCD-Compton, fig. 4. 1b. Both of these processes in general lead to 2+1 jets, where the +1 denotes the jet from the proton remnant. To unambigously define what we mean by a 2+1 jet event, especially when we apply the notation for cross-section calculations, we must define a jet clustering algorithm to be applied to the partonic or hadronic final state. In the analysis described below, we have chosen the JADE algorithm, since this is implemented in PROJET [17], a Monte Carlo program for parton level cross-section calculations. In the JADE algorithm, clustering (using one of several schemes) of four-vectors continues until the invariant mass of all remaining (jet) pairs fulfil the condition $(p_i + p_j)^2 = m_{ij}^2 > M_{cut}^2$. Due to the presence of the proton remnant at HERA, the mass cut is normally defined by relating it to W^2, the invariant mass squared of the hadronic system, via a dimensionless parameter y_{cut} giving $M_{cut}^2 = y_{cut}W^2$. As W^2 increases this cut will lead to a decresing resolving power. In our analysis we have considered the possibilty to resolve the hard sub-process with a fixed resolution over the full W^2 range, proven in [14] to be feasible. In order for the analysis to be sensitive to small x, we want to measure as small invariant masses of the hard sub-process ($\sqrt{\hat{s}}$) as possible (see eq. (4.4)), but we also have to stay clear of the collinear limit where the matrix element calculations give divergences and we have to take the limited resolution of the detector into account by assuring that we have clear jet-structures. The fixed cut was thus somewhat arbitrarily chosen to be $M_{cut}^2 = 100GeV^2$ as a sort of relevant compromise.
Experimentally, the challenge lies in reconstructing the necessary kinematical quantities, in this case x, Q^2 and x_g (x_p), the momentum fraction of the proton carried by the gluon (parton). In particular for the latter, we need to reconstruct the two jets from the hard sub-process well. By doing that, we obtain x_g in terms of $\hat{s} = (p_{jet1} + p_{jet2})^2$ as:

$$x_g = x(1 + \frac{\hat{s}}{Q^2}) \tag{4.4}$$

or alternatively measuring the (pseudo-)rapidities of the two jets in the hadronic centre-of-mass, i.e. the centre-of-mass system of the exchanged photon and the proton, which gives:

$$x_g = \frac{W^2}{W^2 + Q^2}(\frac{Q^2}{W^2} + e^{-(\eta_{jet1} + \eta_{jet2})}) \tag{4.5}$$

For a derivation of eqs. (4.4) and (4.5), see appendix C.

4.3 Event Selection and Jet Reconstruction

To obtain the optimum result in our analysis, we want the largest possible rate for PGF, with only small contributions from other processes like QCD-Compton, and at the same time we wish to obtain sensitivity in as large an x_g range as possible. The full trigger and data selection procedure that optimized these conditions is described in appendix C. The data sample analysed corresponds to an integrated luminosity of $242pb^{-1}$ and was

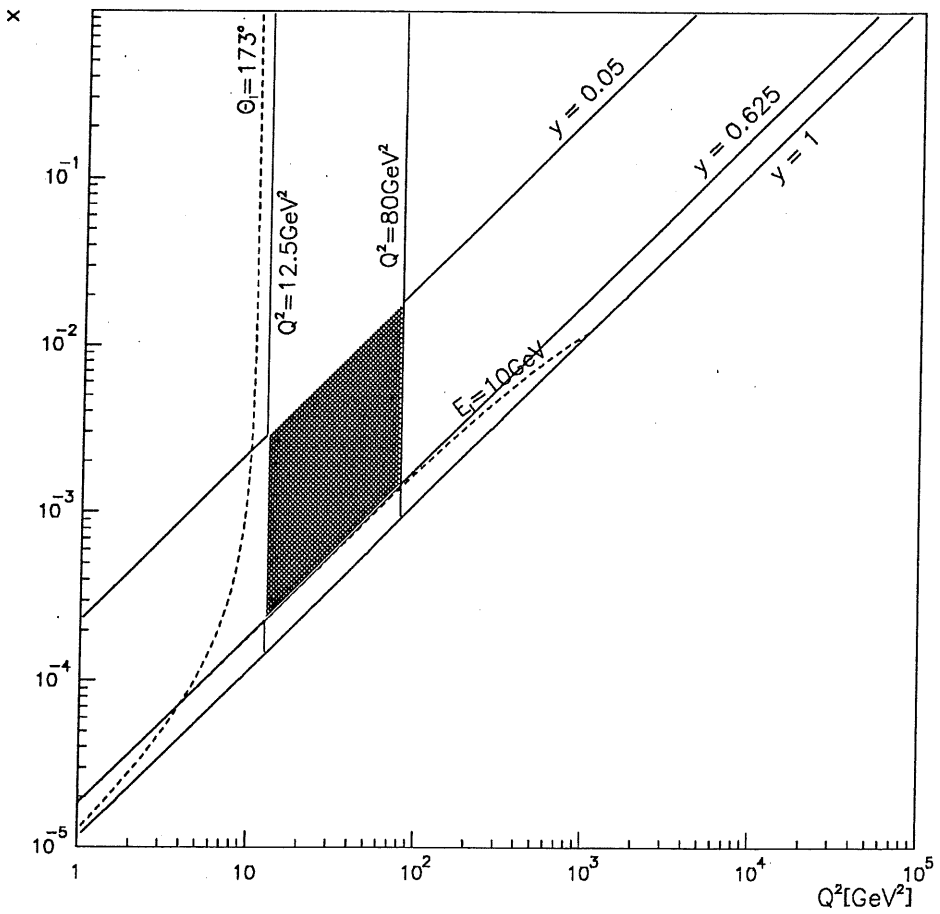


Figure 4.2: *The x, Q^2 region covered by this analysis.*

collected in the H1 detector during the 1993 HERA runs. The main off-line requirements on this sample leading to constraints on the kinematic variables are:

- The scattered electron energy $E_e > 10GeV$ to remove photoproduction background.
- $12.5 < Q^2 < 80$ to ensure that the sacttered electron is well inside the BEMC acceptance.
- $y_{el} \geq 0.05$ in order to properly reconstruct the event kinematics from the scattered electron.
- $y_{had}/y_{el} > 0.5$ to further suppress photoproduction background. (See fig. 4.3)

The kinematic region in which the analysis is performed is consequently defined to be (see fig. 4. 2):

$$\begin{aligned} 12.5 < \quad & Q^2 \quad < 80GeV^2 \\ 0.05 < \quad & y \quad < 0.625 \end{aligned}$$

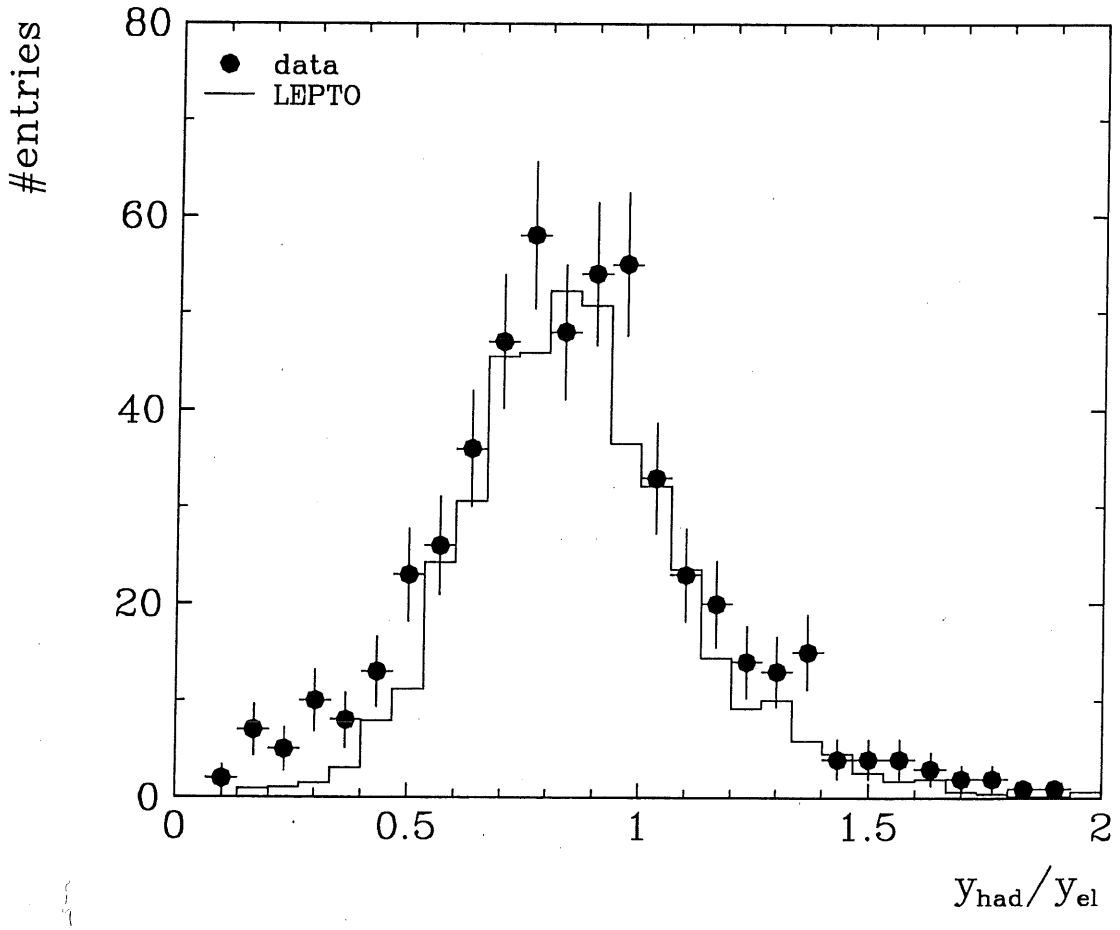


Figure 4.3: *The comparison of data and Monte Carlo for y_{had}/y_{el} shows an excess for data att small y_{had}/y_{el} consistent with being photoproduction background.*

On the remaining sample of almost pure DIS events we apply a jet algorithm and require these additional cuts to be fulfilled by the jet final state:

- 2 (+1) jets
- $m_{ij} > 10GeV$, thereby using the same definition of a 2+1 jet event on the hadronic and partonic final state.
- $10° \leq \theta_{jet} \leq 150°$ so that the jets are measured within the coverage of the liquid argon calorimeter and avoid the forward region dominated by activity from initial parton shawors.
- $\mid \eta_{jet1} - \eta_{jet2} \mid < 2$ to increase the performance of the jet algorithm and thereby the precision of the $\sqrt{\hat{s}}$ determination.

4.3.1 Performance of the Jet Algorithm

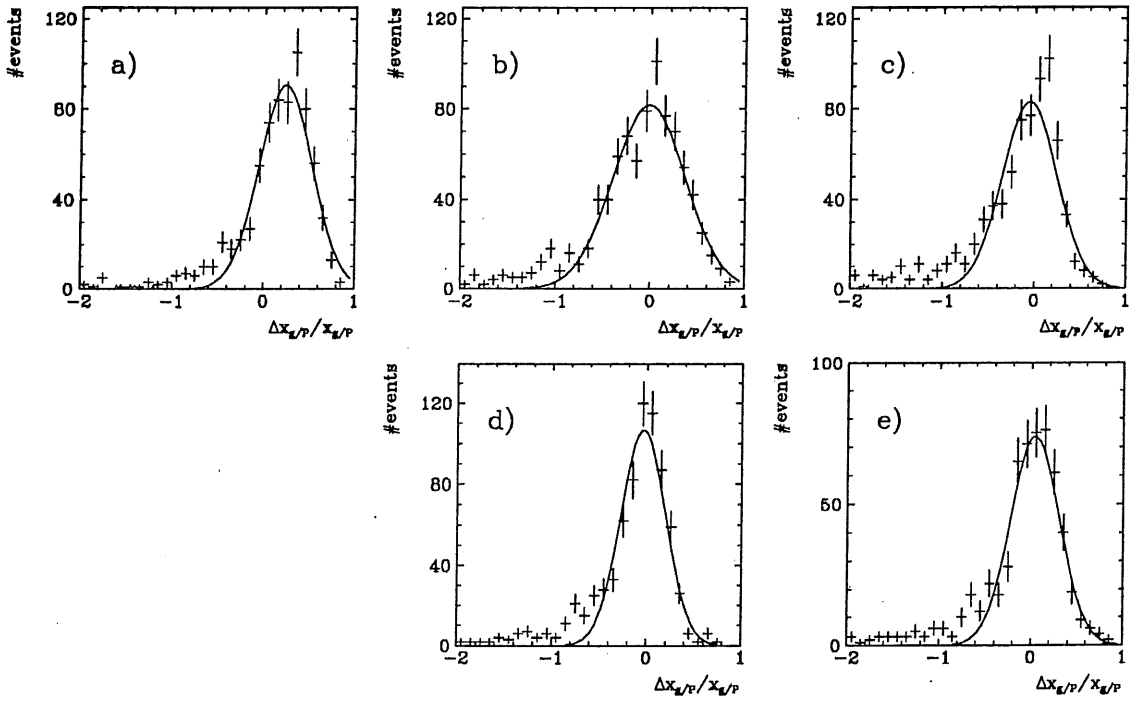


Figure 4.4: *Relative error in the x_g reconstruction for: a) the method based on eq. (4.4) b) as a), but corrected for the systematic shift c) eq. (4.5) corrected for the systematic shift d) the combined result and e) the combined result from JADE.*

We have tested and compared the results of two jet algorithms, the JADE algorithm using a fixed resolution cut of $10GeV$ in the laboratory frame, and a cone algorithm [15] using a cone size of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 1$ and an E_t-cut of $3.5GeV$ applied in the hadronic centre-of-mass. Since we are operating on calorimeter clusters only, and threshold cuts are made to suppress noise in the calorimeter, jet energies will be underestimated. For this and other reasons there will be systematic shifts in the reconstructon of x_g, see fig 4. 4a. We used the results from Monte Carlo event simulations to calculate a systematic correction factor for eqs (4.4) and (4.5) respectively. The resolution given by the equation based on rapidities is better than that of eq. (4.4), but on the other hand the tails are more pronounced. (see figs. 4. 4b and c) For well reconstructed events, both methods should give consistent results and we therefore select these events by requiring that the two results do not differ significantly. We also define a combined result, which is simply the mean of the two reconstructed x_g-values. The latter approach is clearly superior as compared to using either of the previous methods, see fig. 4. 4d. The results presented so far are all obtained with the cone algorithm. JADE also performs fairly well, but the resolution is nevertheless inferior, see fig. 4. 4e. It may seem rather involved to use a cone algorithm to search for jets, and then, after the clustering has been made, apply the

JADE type m_{ij}-cut to the found jets, The reason for this procedure is simply the better performance thus obtained. A lot of effort was spent to get the best possible resolution and this is due to the fact that we are very sensitive to migration. The $\hat{s}$ distribution is a steeply falling function ($\sim 1/\hat{s}$) and poor reconstruction will lead to a large and uncontrollable migration of badly reconstructed low $\hat{s}$ events into our final sample. We are unfortunately not able to completely suppress this background, which we for obvious reasons will call faked 2+1 jet events.

4.4 Monte Carlo Generation

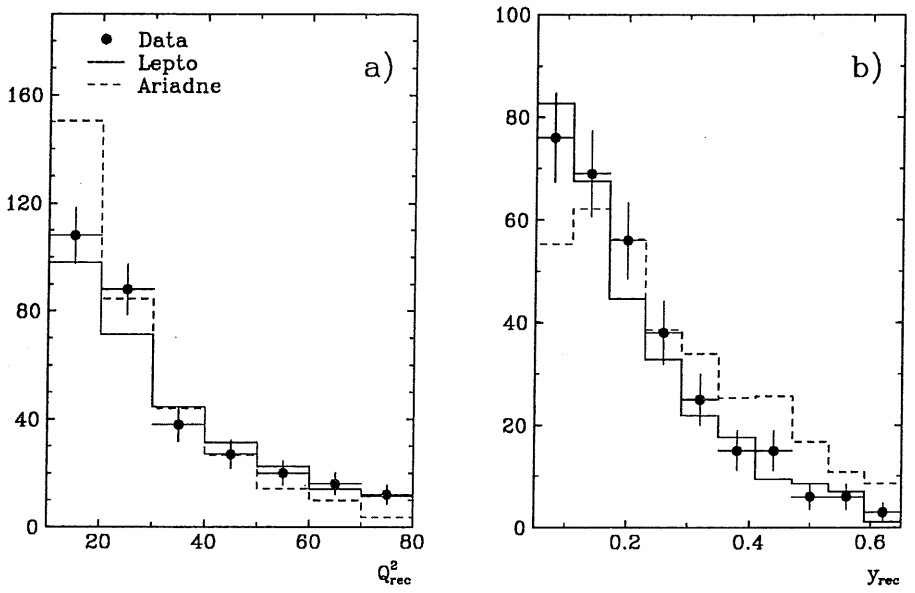


Figure 4.5: *Control plots showing a) the y distribution and b) the Q^2 ditribution of data, LEPTO MEPS and ARIADNE. The curves are normalized to the luminosity of data.*

To relate quantities measured in the detector to quantities calculated on the parton level, we need to compute acceptance correction-factors. In doing this, we not only have to rely on a good simulation of the detector response, but also on a specific Monte Carlo event generator that subsequently should have the same qualitative behaviour as the data we wish to investigate. For this purpose we have used LEPTO 6.1 [6] which includes the necessary first order α_s matrix elements onto which parton showers are added to account for some aspects of higher order corrections (called the MEPS option). It can be seen from fig. 4.4 that for our 2+1 jet sample, LEPTO has a good correspondence with data in terms of the y- and Q^2-distributions. The LEPTO sample used corresponds to an integrated luminosity of $\sim$ 4 times that of the data. Other control plots can be seen in appendix C, for instance it is noteworthy that the energy flow plots are very well described by LEPTO for our jet sample. For comparison, the y and Q^2 distributions as obtained with

ARIADNE [16] are also presented in fig. 4. 5. It is evident from this plot that ARIADNE fails to reproduce our data in a reliable way, and we have therefore disregarded it in the following.

4.5 Cross-sections and Unfolding

There are three contributions to the observed 2+1 jet cross-section; the PGF events we wish to measure (our signal), QCD-Compton events and the faked 2+1 jet events. The kinematic limitations of our data sample leads to a measurable range in x_g of $0.005 < x_g < 0.08$, which we subdivide into three bins to match the resolution and the statistical precision of data. LEPTO was used to calculate the acceptances in each bin for PGF and QCD-Compton by taking the ratio of the PGF (or QCD-Compton) 2+1 jet cross-section in the fully simulated and reconstructed sample surviving all cuts and the parton level cross-section for the same process:

$$A_{PGF} = \frac{\sigma^{PGF}_{LEPTO,full\ sim}}{\sigma^{PGF}_{LEPTO,parton\ level}} \quad ; \qquad A_{QCD-C} = \frac{\sigma^{QCD-C}_{LEPTO,full\ sim}}{\sigma^{QCD-C}_{LEPTO,parton\ level}} \tag{4.6}$$

Now the corrected PGF cross-section can be computed by taking the total observed cross-section for 2+1 jet events and subtracting the QCD-Compton contribution and the cross-section of faked 2+1 jet events as estimated with the LEPTO sample:

$$\sigma^{PGF}_{obs} = \sigma_{obs} - A_{QCD-C} \cdot \sigma_{QCD-C} - \sigma_{fake} \tag{4.7}$$

and

$$\sigma^{PGF}_{corr} = \frac{\sigma^{PGF}_{obs}}{A_{PGF}} \tag{4.8}$$

We calculate parton-level cross-sections with PROJET. Since we are at $x \sim 10^{-2}$ where the quark distributions are fairly well constrained, σ_{QCD-C} is not sensitive to different input parametrizations. The total background subtraction amounts to $\sim 30\%$. The corrected PGF cross-sections are shown in fig. 4. 6a.

For the final step of the analysis we rely more heavily on the LO calculations made with PROJET. We estimate $xg(x) = G(x)$ from the following relation:

$$G(x_g) = \frac{\sigma^{PGF}_{corr}}{\sigma^{PGF}_{MC}} \cdot G(x_g)_{MC} \tag{4.9}$$

where σ^{PGF}_{MC} is the Monte Carlo (PROJET) calculated PGF cross section and $G(x_g)_{MC}$ is the parametrisation of the density function used in the Monte Carlo program. The final result is shown in fig. 4. 6b.

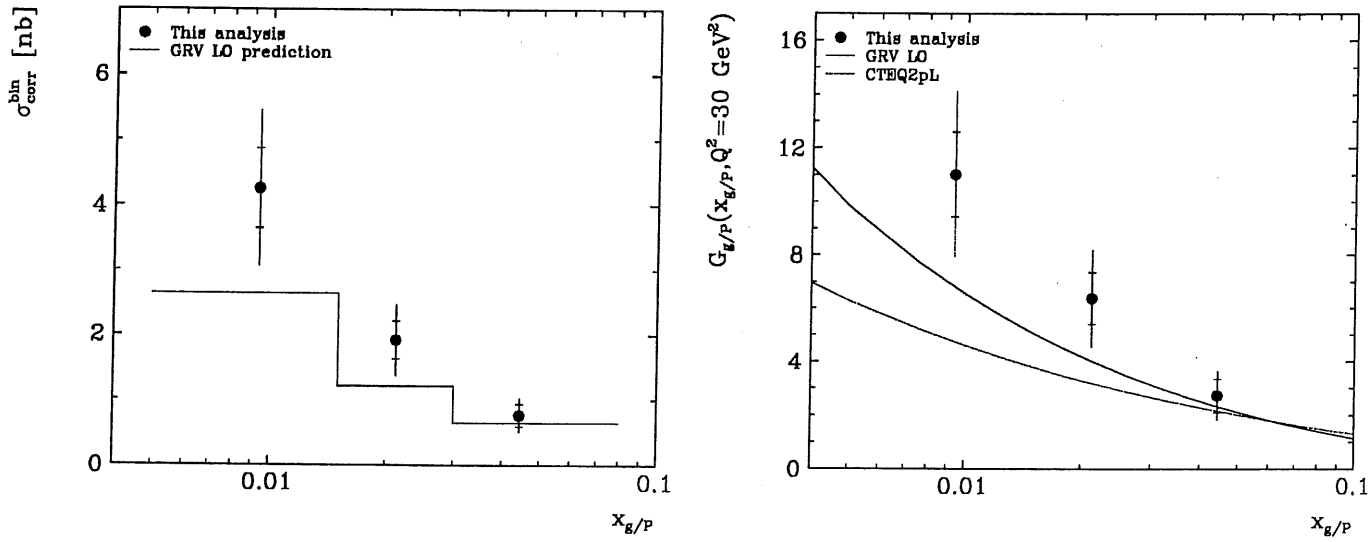


Figure 4.6: *a) The corrected cross-section for photon gluon fusion and b) the unfolded gluon density as a function of the fractional momentum of the gluon.*

4.6 Errors and Uncertainties

With the present level of statistics in data the mean error in the gluon density is below 20%. This should be compared to the systematical error which was estimated to be $\sim 24\%$, not including theoretical uncertainties. We can thus conclude that it is more important to get a better understanding of systematic effects in our analysis, than to increase the statistical precision by re-doing the study with the data collected in 1994 (corresponding to an increase in the integrated luminosity of an order of magnitude). Briefly we have investigated the following main sources for systematic errors:

- The hadronic energy scale is measured to a precision of 5%. By applying a scale variation of that size to the data analysis, the end-result is shifted 10%.

- The E_T-cut used in the cone algorithm was taken to be $3.5GeV$. Below $3GeV$ jet-structures are no longer clear and at $4GeV$ we start to loose too much statistics. A variation of the E_T-cut in this interval affects the result with 7%.

- Using a different jet algorithm on the reconstructed calorimeter clusters should not affect the result, since it is only used for acceptance calculations. If we use the JADE algorithm instead of the cone algorithm we get variations of the order 6% (see fig. 4. 7).

- Our procedure is aimed at only being sensitive to the hard matrix element process. With LEPTO we tested this assumption by calculating acceptances on the hadron level with and without parton showers included in the event simulation. The estimated systematic error comes out to be 15% in this case.

- The effect of different input parametrizations on the acceptances was investigated on the hadron level. Two extreme cases (MRS $D^{0'}$ and MRS $D^{-'}$) where tested, leading to a 5% effect.
- Radiative corrections, estimated by using a Monte Carlo program (DJANGO [18]) including radiative corrections to deep inelastic scattering, are a the level of a few percent. Since we do not make any explicit correction of the measured cross-section due to QED radiation, we include a systematic error of 5% to account for it.
- Global effects, like uncertainties in the luminosity measurement (5%) and detector efficiency calculations (6%) also contribute.

In the calculation of corrected cross-sections, we have not made any explicit assumptions about the order in α_s in which the analysis is carried out (except for the calculation of the small QCD-Compton contribution which is to be subtracted from the observed cross-section). When the last unfolding step is introduced in order to extract the gluon density, we make an explicit leading order determination by using eq. (4.9) and PROJET cross-section calculations. Thus we become sensitive to the truncation of the perturbation series. By varying the renormalization scale (the scale that enters into α_s) from the nominal chosen scale Q^2, we get an *ad hoc* estimate of the theoretical uncertainties. A variation in the interval $0.25 \cdot Q^2$ to $16 \cdot Q^2$ leads to a 30% change of the calculated LO cross-section.

We have also checked our result for a serious dependence on a diffractive contribution. Since the total diffractive contribution was estimated to be approximately 8%, with an equal distribution over the full x_g range, we concluded that this was not the case.

4.7 Summary and Outlook

The measurement of the corrected PGF cross-section at $x_g \sim 10^{-2}$ and the subsequent unfolding of the gluon density in leading order, shows a considerable increase of gluons with decreasing fractional momenta of the proton. Such a steep rise is expected from the Lipatov evolution equations, but could in fact also be obtained with the standard Altarelli-Parisi evolution. There remains also the question of the usefulness of the result in terms of universality. JADE is not factorizable in the formulation used in this analysis, so perhaps the next step would be to apply it in its factorizable formulation [19] [20] in a similar analysis. It will also be interesting to make a comparison with the results of a gluon density determination from J/Ψ and open charm in photoproduction. These studies should soon become feasible with the present fast increase of integrated luminosity.

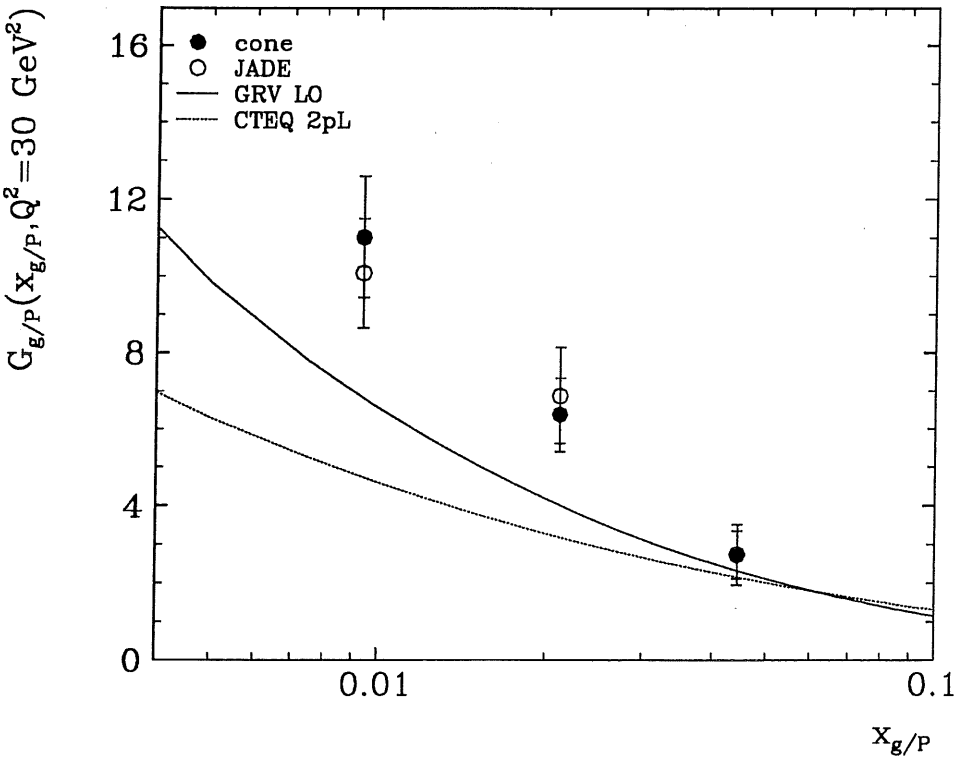


Figure 4.7: *a) The unfolded gluon density as obtained using the cone algorithm (filled circles) and JADE (open circles). Statistical errors indicated.*

Chapter 5

Identification of Jets from Gluons and Quarks

The main strategies for the methods reported in the paper "Jet Identification based on Probability Calculations using Bayes' Theorem" (appendix *IV*) are outlined in this chapter.

5.1 Introduction

We have in this study investigated strategies to separate gluon jets from quark jets. At hand, we have a few properties which can be used to identify a jet:

- Kinematic properties of the production process. For instance gluon radiation from a quark, a bremsstrahlung like process, leading to the gluons generally having less energy than the quarks.
- Instrinsic properties, like colour charge, which will influence the fragmentation of quarks and gluons into hadrons.
- Identification of heavy quarks can be made based on their specific deacay properties using methods like D*-tagging, secondary vertex reconstruction or tagging using semileptonic decays into leptons.

We consider heavy quarks to be a special case, and have thus chosen to base the separation of gluon and quark jets on the differences in their fragmentation properties and energies.

5.2 Event Simulation and Jet Reconstruction

Two event configurations containing quarks and gluons were investigated, 3 jet events at LEP, which are always of the type $q\bar{q}g$, and 2(+1) jet events at HERA, which at Bjorken $x > 0.1$ to a good approximation are of the qg type (QCD-Compton events).

The LEP events were simulated with the parton shower option of the Monte Carlo program JETSET, and the HERA events were simulated with the event generators LEPTO and HERWIG [21]. By comparing results from JETSET and LEPTO the process dependence of our method can be checked. Correspondingly, we check the fragmentation model dependence by comparing results from HERWIG and LEPTO.

We have used LUCLUS [22] with $d_{join} = 4GeV$ for jet clustering in all three cases. When calculating the fragmentation variables, represented by the so-called Fodor moments [23], we only take the "core" of the jet into account. The core is defined by picking particles in a jet in descending order of momentum with respect to the jet axis, until 80% of the total energy is reached. By doing this, sensitivity to unwanted effects, like the Lorentz frame in which the clustering is performed, should decrease.

5.3 Methodology

A distinction is made between separation based on jet energies and separation based on fragmentation variables. We motivate this with the following arguments:

- Separation based on jet energies is a process dependent method, whereas separation based on fragmentation variables in principle is not, and can thus be treated in a more general way.
- A result based on combined energy and fragmentation can not be used to test a model of either of the two.
- The discriminating information in fragmentation variables is stored in many variables and in their correlations. It may therefore be advantageous to treat this more complicated case separately.

There is additional information stored in the event, meaning that since we are only considering events in which we know the exact number of quarks and gluons, we will also "force" our method to give the correct number of quarks and gluons for each event. Consequently we always identify all the jets in an event. The basic tool for introducing event based identification is the calculation of probabilities for a certain jet in an event to be either a gluon or a quark. Using Bayes' theorem we include the condition on the correct number of quarks and gluons in an event, with given values of the jet energies or fragmentation variables, in these probabilities. A further advantage of probabilities is that they are easily combined (easy in the mathematical sense). Thus, the separate identification of the jets based on energy and fragmentation variables can be combined in the end.

A neural network is used for the case of identification based on fragmentation variables. The multi-dimensional charachter of this problem makes neural networks ideal, since it is basically a classification problem. There is furthermore the advantage that the

output of the neural network directly can be interpreted as a conditional probability given the training sample. Special care is taken with respect to the training sample; quark and gluon jets are processed through the neural network individually and with an equal and flat distribution in energy. In this way we take the energy dependence of the fragmentation variables into account, but the energy information in itself does not contain any discriminating information.

Acknowledgements

It is not easy to express the amount of gratitude I feel towards all my colleagues who have helped me get this far, at least not without giving the impression that you are reading one of the novels written by Barbara Cartland. Unfortunately there is not room enough to mention you all.

My supervisor, Leif Jönsson, has given me patient support and advice through all these years. I even think he had some faith in me at times when I didn't.

Gunnar Ingelman has been a constant source of wisdom. His patience with me and all my questions is certainly admirable.

Thanks are definitely due to Hermann Küster and Joachim Stier for a perfect collaboration on the gluon density analysis.

Christian (we must have had a million discussions, and OK, I was wrong sometimes) and Håkan added to the joy of being an H1er in Lund.

I thank all the people at the home deparment in Lund for creating such a special atmosphere. (More wine-tasting sessions, please.) Of course all the colleagues in the H1 collaboration deserve a mentioning, and for the hospitality of DESY in Hamburg I am also grateful. Financial support from the University of Lund is gratefully acknowledged.

My wife, Shanti, who has gone through thick and thin with me. I will never be able to thank you enough.

My family, you have followed me all the way.

My dear friends (those of you who are left), I resolve to spend more time with you all from now on.

Bibliography

[1] F. Jacquet, A. Blondel, Proceedings of the study of an ep facility for Europe 79/48 (1979), p. 391-394, U. Amaldi ed.

[2] S. Bentvelsen, J. Engelen, P. Kooijman, in [24] vol. 1, p. 23

[3] G. Bernardi, W. Hildesheim, in [24] vol. 1, p. 79

[4] H1 collaboration, I. Abt. et al., Preprint, DESY 93-103

[5] S.J. Brodsky, P. Hoyer, C. Peterson, N. Sakai, Phys. Lett. 93B (1980) 451
S.J. Brodsky, C. Peterson, Phys. Rev. D23 (1981) 2745

[6] G. Ingelman, LEPTO version 5.2, unpublished program manual
G. Ingelman, LEPTO version 6.1, in [24] vol. 3, p. 1366

[7] T. Sjöstrand, JETSET 6.3, Comp. Phys. Comm. 39 (1986) 347, *ibid.* 43 (1987) 367
T. Sjöstrand, JETSET 7.3, program manual, CERN-TH.6488/92

[8] G. Ingelman, G.A. Schuler, AROMA 1.2, unpublished program manual, see also
G. Ingelman, G.A. Schuler, AROMA 1.4, in [24] vol. 3, p. 1346
G. Ingelman, G.A. Schuler, Zeitschrift für Physik C40 (1988) 299

[9] E. Hoffman R. Moore, Z. Phys. C20 (1983) 71

[10] Yu. A. Golubkov, Preprint DESY 94-060, 1994

[11] M.R. Adams et al., E665 collaboration, Nucl. Instr. and Meth. A291 (1990) 533

[12] H1 collaboration, I. Abt. et al., Phys. Lett. B321 (1994) 161

[13] G. Altarelli and G. Parisi, Nucl. Phys. 126 (1977) 297.

[14] V. Hedberg, G. Ingelman, C. Jacobsson and L. Jönsson, Z. Phys. C63 (1994) 49

[15] F.Abe et al., Phys. Rev. D45 (1992) 1448.
M.Seymour, LU TP 93-7, May 1993.

[16] L. Lönnblad, ARIADNE version 4.03, Comp. Phys. Comm. 71 (1992) 15

[17] D. Graudenz, PROJET program manual, unpublished and J. Rathsman, private communication.

[18] G.A. Schuler and H. Spiesberger, in [24] vol. 3 and references therein.

[19] B.R. Webber, Factorization and Jet Clustering Algorithms for Deep Inelastic Scattering, CERN-TH. 6871/93

[20] A. Edin, The Gluon Density of the Proton from 3-jet Events in *ep* Collisions at HERA, TSL/ISV-93-0087

[21] G. Marchesini, B.R. Webber et al., Comp. Phys. Comm. 67 (1992) 465

[22] T. Sjöstrand, Comp. Phys. Comm. 28 (1983) 22

[23] Fodor, Phys. Rev. D41 (1990) 1726

[24] Proceedings 'Physiscs at HERA', Eds. W. Buchmüller, G. Ingelman, DESY Hamburg 1992, vol. 1-3

Appendix A

The H1 Forward Muon Spectrometer

Nuclear Instruments and Methods in Physics Research A 340 (1994) 304–308
North-Holland

NUCLEAR
INSTRUMENTS
& METHODS
IN PHYSICS
RESEARCH
Section A

The H1 forward muon spectrometer

H. Cronström [a], V. Hedberg [a], C. Jacobsson [a], L. Jönsson [a,*], H. Lohmander [a], M. Nyberg [a], I. Kenyon [b], H. Phillips [b], P. Biddulph [c], P. Finnegan [c], J. Foster [c], S. Gilbert [c], C. Hilton [c], M. Ibbotson [c], A. Mehta [c], P. Sutton [c], K. Stephens [c], R. Thompson [c]

[a] *Physics Department, University of Lund, Lund, Sweden*
[b] *School of Physics and Space Research, University of Birmingham, Birmingham, UK*
[c] *Physics Department, University of Manchester, Manchester, UK*

Received 21 September 1993

The H1 detector started taking data at the electron–proton collider HERA in the beginning of 1992. In HERA 30 GeV electrons collide with 820 GeV protons giving a strong boost of the centre-of-mass system in the direction of the proton, also called the forward region. For the detection of high momentum muons in this region a muon spectrometer has been constructed, consisting of six drift chamber planes, three either side of a toroidal magnet. A first brief description of the system and its main parameters as well as the principles for track reconstruction and T_0 determination is given.

1. General description

The purpose of the forward muon spectrometer is to measure high energy muons in the range of polar angles $3° \leqslant \theta \leqslant 17°$. The detector consists of drift chamber planes, either side of a toroidal magnet. The design specifications aim at measuring the momenta of muons in the range between 5 and 200 GeV/c, the lower limit being given by the amount of material the muons have to penetrate and the influence on the momentum resolution of the multiple Coulomb scattering in the magnet iron. The upper limit is set by the magnetic field strength of the toroid together with the spatial resolution of the drift chambers. The expected momentum resolution at 5 GeV/c is 24% and deteriorates slowly to 36% at 200 GeV/c above which there is a danger of misidentifying the charge of the muon. Muon momenta below 5 GeV/c will be measured in the forward tracker.

Fig. 1a shows schematically the detector arrangement and the toroid magnet. The drift chamber planes, which increase in size from about 4 m diameter for the first detector plane to 6 m diameter for the last, are all divided into octants which are formed from individual drift cells accurately mounted on Al-frames. The orientation of the drift cells is such that four of the planes essentially measure the polar angle (θ) and thereby provide the momentum of the traversing muon whilst the remaining two measure the azimuthal angle (ϕ). Each plane consists of a double layer of drift cells such that each layer is displaced with respect to the other by half a cell width (Fig. 1b). This arrangement enables the resolution of left–right ambiguities and also the determination of T_0 as will be explained below. The total number of drift cells is 1520.

The toroid is 1.2 m thick and constructed out of eight semicircular steel modules with an inner radius of 0.65 m and an outer radius of 2.9 m. Each of the twelve rectangular coils which provide the field consists of 15 turns of watercooled Cu-tube, 11.5×11.5 mm^2. At a current of 150 A the field strength varies from about 1.75 T at the inner radius to about 1.5 T at the outer radius. Field measurements made in the centre of each coil show a variation of less than 1%. A more detailed description of the toroid magnet can be found in ref. [1].

2. Chamber design

All drift cells have a rectangular cross section with a depth of 2 cm, a width of 12 cm and lengths between 40 and 240 cm. With a central sense wire the maximum drift distance becomes 6 cm. The cells have 50 μm thick nichrome wires except for the inner short cells where the diameter is 40 μm. For cells longer than 1.5 m there is a wire support in the middle. As illustrated in Fig. 2 each cell comprises two PCB planes, copper-coated on both sides, and 0.7 mm thin extruded Al-

* Corresponding author.

SSDI 0168-9002(93)E1039-Z

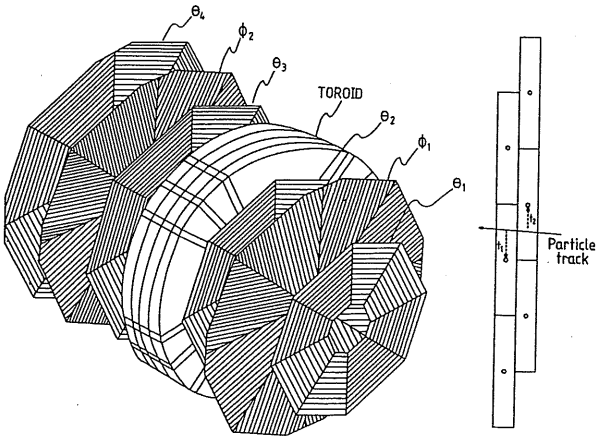


Fig. 1. (a) A schematic view of the forward muon spectrometer and (b) the cell structure of a double layer.

profiles to minimise the dead space between cells. The outer copper surface of the PCB is kept at ground to form a screened box while the inner surface has been machined to give 4 mm wide strips. These are used as drift electrodes connected to a 230 MΩ metal film resistor chain mounted on the end cap to obtain a uniform drift field. The end caps are made of moulded Noryl with high precision holes to locate the crimp pins for fixing the sense wires and provide holes for the gas connections. One end contains the high voltage distribution resistor chain and the sense wire readout connection isolated via a 1 nF ceramic disc capacitor. Sense wires of adjacent cells are linked together via a 330 Ω resistor at the other end forming the equivalent of a U-shaped cell which then is read out at both ends. This allows not only a determination of the track position transverse to the sense wire from the measurement of the drift time but also the coordinate along the wire by charge division measurement and thereby giving information on which cell of a coupled pair was hit.

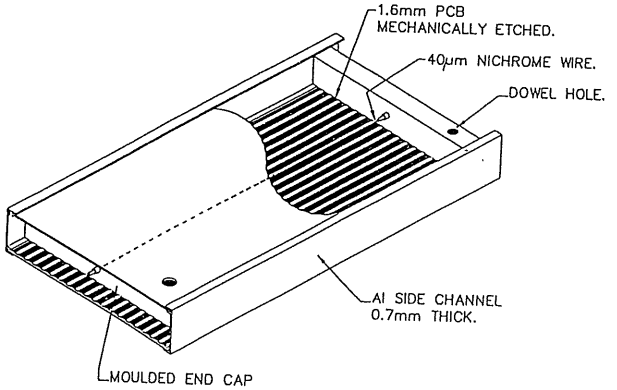


Fig. 2. The construction of a drift cell.

3. The chamber gas and high voltage system

The choice of gas for the drift chambers was determined by several requirements. One is the desire to work in a drift voltage range where the drift velocity is constant. Further the gas has to be fast enough for the pulse to arrive in time for the trigger and finally it should be nonflammable for safety reasons. Currently the so-called FMS gas (forward muon spectrometer gas), which is a mixture of 92.5% argon, 5% CO_2 and 2.5% methane, has been chosen for the chambers. The gas is mixed and purified in a recirculator [2]. The chambers have a total gas volume of 4 m^3, and with a small overpressure of about 0.2 mbar measured at the output, the return gas flow is typically 90% of the input and the oxygen content is measured to be of order 100 ppm. For the FMS gas the drift velocity as a function of the drift field voltage, corrected for atmospheric pressure, is shown in Fig. 3. An average drift field of 480 V/cm gives the desired drift velocity of ~ 5 cm/μs. The drift field is defined by an increasing positive potential from the cathode at ground to +2.88 kV on the centre electrode at the position of the sense wire. The sense wire is kept typically at 4.21 kV for the 40 μm wires and at 4.26 kV for the 50 μm wires. The gas gain is controlled by the difference in voltage between the sense wire and the drift field close to the wire.

A 120-channel CAEN 127 system supplies distribution boxes on the detector with high voltage via 50 m long coaxial cables. One 6 kV 1 mA module supplies drift voltage to an entire octant, feeding 20–40 individual resistor chains. Similarly an 8 kV 200 μA module supplies the sense voltage to all but the 12 innermost cells of a θ-octant, which in case of bad beam conditions might be set to a lower voltage. For the φ-octants the central section which is close to the beam tube can be moved outwards mechanically by remote controls and thus there is no need for any special HV arrangement.

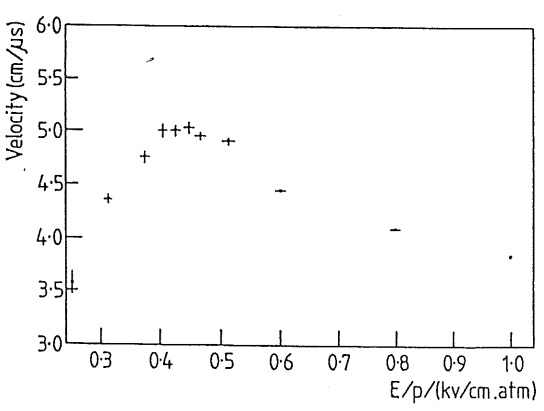


Fig. 3. The drift velocity versus the drift voltage, corrected for atmospheric pressure, for the FMS gas.

There is a continuous monitoring of the gas composition and flow rates as well as of the high voltage, communicated via an Apple Macintosh II ci in the control room. From this work station it is also possible to control the high voltage of the detector and the toroid magnet.

4. The readout system

The signals are read out into 8 channel preamplifiers of standard H1 design [3] mounted close to the cells. The output pulses are driven down 50 m long coaxial cables to F1001 flash analogue to digital converters (FADC) [3] run in common stop mode. Signals are digitised with the equivalent of 10 bit resolution (8 bit nonlinear response) in 9.6 ns time bins, phase locked to the HERA beam crossing frequency. These are stored in a circular buffer with a depth of 256 time bins. Each crate of 256 FADCs is controlled by a scanner [4] which on receiving the first trigger stops the buffer and scans the preceding 256 digitisings for hits. Each scanner then transfers the zero suppressed data to a front end processor where the pulses are analysed to extract start times and charge contents.

5. The charge–time analysis

Only the rising edge and peak region of a pulse is used to get the time and charge information. A pulse is said to start when there are two successively rising digitisings above threshold. The end of a pulse is taken as the second successive digitising after the peak which is below threshold, or eight 9.6 ns time bins from the start of the pulse, whichever occurs first. The arrival time of the pulse is obtained by extrapolating a line fitted to the steepest part of the leading edge back to the intercept with the background level. With a test setup, looking at cosmic muons, this method gave a resolution of < 200 μm as illustrated in Fig. 4. This result was obtained with a gas mixture of 90% argon and 10% propane providing a drift velocity of 4 cm/μs. However, to satisfy the gas requirements specified in section 3, we have, as mentioned earlier, chosen the FMS gas with a drift velocity of ~ 5 cm/μs, resulting in an expected resolution of ~ 250 μm. Pairs of pulses which originate from the same hit are associated by requiring the difference of their arrival times to be less than the full propagation time through the two sense wires of the linking resistor.

The collected charge is found by integrating the digitisings of the pulses from the two wire ends over intervals of the same length, with subtraction of a constant background. A correction for fractional time bins was found to be important since the start times for the two pulses are subject to variable propagation delays. With cosmic muons in the test setup we found a charge-division versus distance characteristics which was linear to about 1%, which is well matched to the resolution.

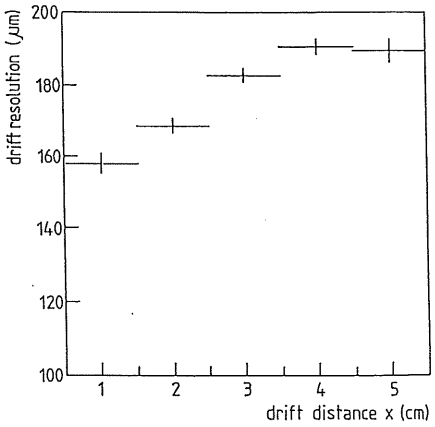


Fig. 4. The space resolution of a drift cell as a function of drift distance for the gas mixture 90% argon and 10% propane (drift velocity 4 cm/μs).

6. Track reconstruction

The space points obtained from the charge–time analysis of the chamber hits are used in a three-step procedure for track reconstruction which starts with the pairing of hits in each double layer followed by association of pairs into straight track segments and finally the linking of track segments through the toroid to form full tracks and thus provide a momentum measurement. Pair finding in the double layers is decisive due to the displacements of cells which results in the sum of drift times being a constant (compare Fig. 1b). A vertex pointing requirement is applied as selection criteria but also unpaired hits are kept to be considered in the track segment finding where we demand three out of four hits in the θ-layers. The measuring errors of the space points for a pair define a cone which is extrapolated to the other θ-layer on the same side of the toroid. In the area defined by the cone, hits are tried for segment fits and are selected by a χ^2-cut. In the future the information from the ϕ-layers will also be used.

For the linking procedure each pretoroid segment is tracked through the magnetic field of the toroid, taking into account energy loss and multiple Coulomb scattering in the magnet iron. By doing this for a minimal reconstructable momentum of 2.5 GeV/c in the spectrometer and for either of the two muon charges possible, regions in the θ-layers after the toroid are defined inside which segment candidates for linking are consid-

ered. From the crossing angle of two linked segments an estimate of the momentum is made. Starting from the pretoroid segment and the estimated momentum the tracking is repeated as the momentum is changed in small steps around the estimated value. Each post-toroid segment obtained from the tracking is compared to the actual segment found and a χ^2 is calculated. The minimum of the χ^2 variation with momentum defines the momentum corresponding to the best fit.

Fig. 5 shows a schematic side view of the H1-detector with the central and forward tracking devices and the calorimeter all surrounded by the instrumented iron. The forward muon spectrometer can be seen to the left of the main H1-detector. A clear muon track originating from the vertex region can be followed through the various subdetectors extending all the way to the end of the forward muon spectrometer. The track coordinates are given by the radial distance from the beam line, R, and the longitudinal Z-coordinate in the direction of the proton beam. However, since the instrumented iron only gives information on the vertical position, X, a radius coordinate cannot be extracted and consequently this track segment is plotted in X, Z-coordinates. This results in the apparent non-alignment of that particular track segment.

7. Drift velocity and T_0 determination

Beam halo muons are used to determine the drift velocity. From the uniform population of the total number of tracks (N) over the full drift distance (ΔY), recorded in a run, a rectangular distribution is expected if the drift velocity is constant. However, due to field variations close to the sense wire, dependence on the angle of the track, the possibility of tracks traversing only the corner of a cell etc., the drift velocity will

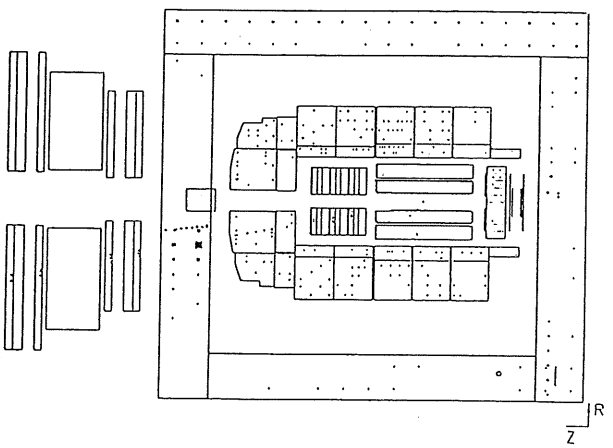


Fig. 5. A side view of a genuine event with a muon penetrating the complete detector.

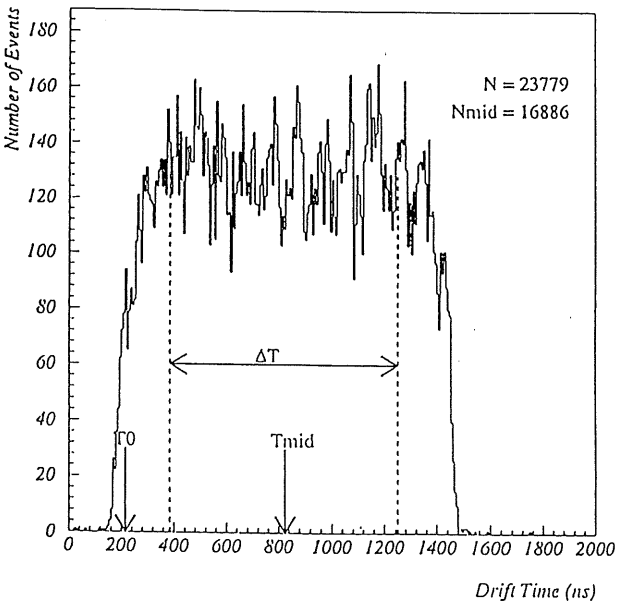


Fig. 6. The drift time distribution for beam halo tracks used to extract T_0 with the FMS-gas.

be altered and cause a smearing of the distribution (Fig. 6). In spite of this smearing, the drift time (T_{mid}), corresponding to half the drift distance (3 cm), can be defined as the time which leaves equal number of tracks above and below. Taking an arbitrary time interval (ΔT) symmetrically around T_{mid}, where the distribution is still flat, we can count the number of tracks (N_{mid}) in this interval and use it for a determination of the drift velocity by the following expression:

$$v = \Delta Y N_{\mathrm{mid}}/(\Delta T N).$$

The result is $v = 4.926 \pm 0.039$ cm/μs.

T_0 is determined from the specific geometry of the detector which makes one of the following check sums true for each track.

$$T_1 + T_2 + T_3 + T_4 = 4T_{\mathrm{mid}},$$

$$T_1 + T_2 - T_3 - T_4 = 0,$$

where T_1, T_2, T_3 and T_4 are drift times in the four θ-layers. The first check sum thus will provide an independent measurement of T_{mid}. T_0 can now be determined from the expression:

$$T_0 = (3\ \mathrm{cm}/v) - T_{\mathrm{mid}}.$$

The widths of the two check sum distributions can be used to find the spatial resolution of the chambers.

8. Chamber alignment

The drift chambers must be aligned with respect to each other and to the rest of the detector. The cells of a θ-layer are positioned on its supporting Al-frame to a

precision of ~ 50 μm along the drift direction and to ~ 1 mm in the two other directions. This is better than the achievable resolution and therefore we only have to consider the alignment of the full octants.

Simulation studies and analysis of a small sample of real data have shown that beam halo tracks are suitable for providing the two translational and one rotational quantities which are needed to specify the position of the octant in the plane transverse to the beam direction. Further studies with angle tracks together with the survey will determine the relative positions of the octants along the direction of the beam [5].

Acknowledgements

We wish to acknowledge the substantial effort from the mechanical and electronic workshops at the Universities of Lund and Manchester, together with the contributions of the engineers from the PAG group at the Rutherford Laboratory. We thank the Swedish and United Kingdom funding agencies for their financial support to the construction of the detector. Further we acknowledge the technical facilities provided at the installation phase by DESY.

References

[1] I. Abt et al., DESY Preprint 93–103 (1993).
[2] G. Kessler and W. von Schröder, private communication.
[3] J. Bürger et al., Nucl. Instr. and Meth. A 279 (1989) 217.
[4] R.J. Ellison and D. Mercer, H1 Internal Note 4/87–63.
[5] P. Sutton, Ph.D. Thesis, University of Manchester (1993).

Appendix B

Deep-Inelastic Scattering on Intrinsic Charm Quarks in the Proton

PHYSICAL REVIEW D VOLUME 47, NUMBER 11 1 JUNE 1993

Deep-inelastic scattering on intrinsic charm quarks in the proton

G. Ingelman
Deutsches Elektronen-Synchrotron DESY, Notkestrasse 85, D-2000 Hamburg 52
and Department of Radiation Sciences, Uppsala University, Box 535, S-751 21 Uppsala, Sweden

L. Jönsson and M. Nyberg
Department of Physics, Lund University, Sölvegatan 14, S-223 62 Lund, Sweden
(Received 21 December 1992)

Possibilities of probing the predicted intrinsic charm quark component in the proton by deep-inelastic lepton scattering are investigated. Total cross sections are found to be sizable, but the observable rate, e.g., through muons from semileptonic charm decays, depends strongly on the experimental situation. The DESY *ep* collider HERA has acceptance losses close to the proton beam, along which the *ep* system is strongly boosted, whereas for fixed target muon scattering, e.g., at Fermilab, the acceptance can be made much better. Backgrounds are calculated and ways to suppress them developed, resulting in acceptable signal/background ratios. For existing experiments the predicted statistics are relatively small, but may still contribute to a solution of the open intrinsic charm problem, whereas a dedicated experiment should settle the issue.

PACS number(s): 13.60.Hb, 12.38.Qk, 14.20.Dh

I. INTRODUCTION

The charm production cross section in hadron collisions has been observed [1] to be larger and have a flatter distribution in Feynman x than anticipated based on leading-order perturbative QCD diagrams and simple hadronization models. As a solution to this apparent problem the hypothesis of intrinsic charm (IC) was introduced [2] by assuming the existence of a $c\bar{c}$ pair as a *nonperturbative* component in the bound-state proton. This means that the Fock-state decomposition of the wave function, $|p\rangle=\alpha|uud\rangle+\beta|uudc\bar{c}\rangle+\cdots$, contains a small, but finite, probability β^2 for such an intrinsic quark-antiquark pair. Viewed in an infinite momentum frame, all nonperturbative (long-lived) components must move with essentially the same velocity in order that the proton can "stay together" for an appreciable time. The larger mass of the charmed quarks then implies that they take a larger fraction of the proton momentum. For definiteness we shall assume the model form for the intrinsic charm quark density distribution proposed in Ref. [2]

$$c(x)=18x^2\{\tfrac{1}{3}(1-x)(1+10x+x^2)+2x(1+x)\ln x\}\ , \tag{1}$$

having a mean value $\bar{x}=2/7$ as compared to $\bar{x}=1/7$ for the light quark distribution in the $|uudc\bar{c}\rangle$ state. This distribution is derived by assuming a five-quark Fock state wave function which varies inversely with the invariant mass of the intermediate state. The normalization factor is chosen to correspond to 1% probability for intrinsic charm as first suggested [2] to explain the original data [1]. The actual normalization of heavy quark Fock components in the proton is the key unknown, although it should decrease as $1/m_Q^2$.

More recent calculations have shown that the next-to-leading-order QCD corrections to the conventional perturbative mechanism for hadroproduction of open charm give a sizable increase of the cross section (although the shapes of some important differential distributions are not much affected), see [3] and references therein. In addition, the x_F spectra become harder than predicted by lowest-order fusion processes when string hadronization effects and coalescence with spectator quarks are taken into account [4,5]. This gives less of a need for an intrinsic charm component to describe the total charm production cross section, but some combination of coalescence and intrinsic charm contributions seems to be needed in order to describe the x_F and nuclear dependence of the charm data [5]. It is also interesting to note that measured cross sections for J/ψ production at large x_F in πN and pN collisions [6] appear to be in excess of that predicted by the conventional fusion subprocesses; an effect which cannot be accounted for by string hadronization or coalescence effects. Further data [7] shows that the J/ψ produced in $\pi^- N$ collisions becomes strongly longitudinally polarized at large x_F, suggesting a change in the production mechanism as the quarkonium state receives a high fraction of the beam momentum. The presence of an intrinsic charm contribution can account for the measured x_F distribution as well as the observed diffractivelike nuclear target dependence of the hidden charm cross sections [8,9].

The "intrinsic" quark sea should be contrasted against the "extrinsic" sea generated by large momentum-transfer processes in the perturbative QCD evolution of the parton distribution functions. Being perturbative, the latter is a short-lived fluctuation which can only be put on-shell by a large momentum-transfer external probe. The extrinsic quarks will only carry small momentum fractions of the proton as is characteristic for normal sea

quarks. This is a normal leading-twist contribution as opposed to the higher-twist nature of the intrinsic heavy quark component. Since the intrinsic contributions are associated with multiparton correlations in the hadron wave function, the intrinsic charm quarks may carry a large fraction of the proton's momentum.

Both of these charm quark components in the proton can be probed by deep-inelastic scattering (DIS). The dominant charm production process is the photon-gluon fusion (PGF) process $\gamma g \to c\bar{c}$. Data from the European Muon Collaboration on $F_2^{\text{charm}}(x)=2e_c^2 xc(x)$ were found to be globally well described by this conventional process, but could not exclude an IC component below 0.6% [10]. A refined theoretical analysis [11], which took proper account of the reduction of $c(x)$ at large x due to the charm threshold at low Q^2 and the QCD evolution, showed no conflict between the data and the IC model. On the contrary, it gave some evidence for an intrinsic charm component. At $x \gtrsim 0.25$, where PGF is expected to drop fast and IC should give its main contribution, the data seem to indicate the presence of IC at the level of approximately 0.3% [11].

In the following we investigate to what extent this rather unclear situation can be improved by observing deep-inelastic scattering on intrinsic charm quarks in the proton using experiments at higher energies. In particular we make detailed calculations, based on Monte Carlo event simulation, for fixed target scattering at Fermilab energies and ep collisions at the DESY HERA. Charm can favorably be tagged through its semileptonic decay into muons. The branching ratio of approximately 10% is large compared to most other charm branching ratios. Furthermore, the efficiency for detecting and identifying muons is high. The reconstruction of specific charm states requires a completely different experimental situation with excellent particle identification and momentum measurement as well as good photon detection. Even so the reconstruction efficiency turns out to be low due to the large number of charm decay channels, all with small branching ratios, and due to the high combinatorial background.

As a potential background to the charm muon signal we consider, in addition to the PGF process, also muons from π and K decays, since their large production cross section in normal DIS can outbalance the small decay probability within the detector volume. Muons can also be produced in pairs from various QED radiation processes [12]. These have, however, either a small cross section or involve small momentum transfers and will then have different topologies compared to the IC signal. We have, therefore, not considered these backgrounds in detail.

II. MODELS FOR SIMULATING SIGNAL AND BACKGROUNDS

The basic scattering on an intrinsic charm quark is identical to normal DIS, when quark masses can be neglected, with the kinematic variables $Q^2 \equiv -q^2 = -(p_e - p_{e'})^2$, $x \equiv Q^2/2P\cdot q$ and $y \equiv P\cdot q / P\cdot p_e$ having their normal definition and meaning. Furthermore, the electroweak cross sections are obtained from the normal formulas [13] by simply keeping only a charm quark distribution which is now taken from Eq. (1). It is only meaningful to consider neutral current interactions, since the charged current exchange has a smaller cross section and turns the charm quark into a lighter flavor which gives a less clear signature. The cross section for ep or μp scattering on charm is then given by

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

where in the case of intrinsic charm

$$F_2^c(x,Q^2) = F_2^{\text{IC}}(x,Q^2) = 2e_c^2 xc(x,Q^2) \tag{3}$$

and only photon exchange is taken into account since one can neglect the region of large Q^2 where Z exchange contributes.

We have adopted the ep scattering event generator LEPTO 5.2 [14] to simulate complete intrinsic charm scattering events. When implementing the intrinsic charm quark density distribution equation (1) we have kept the 1% normalization; rescaling to a different value can be trivially made in our final results. We take the Q^2 evolution of $xc(x)$ into account through the leading logarithm QCD, since it has the important effect to reduce the function in the characteristic large-x region as shown in Fig. 1(a). The proper evolution equation, including the charm quark mass, is rather involved and we have adopted a suitable parametrization of the numerical results in Ref. [11] by a polynomial with an amplitude proportional to $lnlnQ^2$. We normally require $Q^2 \geq 10$ GeV^2 to avoid corrections for mass (threshold) effects in Eq. (1) [11]. This minimum Q^2 also ensures a proper DIS situation where the scattered lepton can be detected and used to

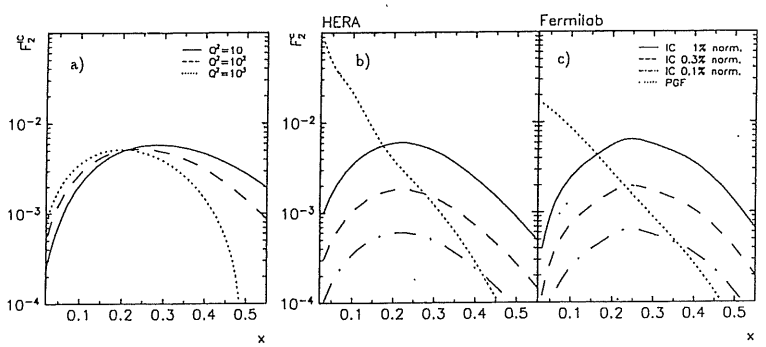


FIG. 1. (a) The intrinsic charm structure function $F_2^{\text{IC}}(x,Q^2)=\frac{8}{9}xc(x,Q^2)$ obtained from Eq. (1) taking QCD Q^2 evolution into account based on Ref. [11]. (b) and (c) The charm structure function F_2^c extracted using Eq. (2) from the differential cross section obtained from the model simulations of intrinsic charm in the proton (with the indicated normalizations) and photon-gluon fusion into $c\bar{c}$ at (b) HERA and (c) Fermilab energies. Q^2 is integrated from 10 GeV^2 to the kinematic limit (its mean value is smaller for IC).

reconstruct the kinematics of the event.

Although there are theoretical arguments for the form of the intrinsic charm distribution in Eq. (1) it is not necessarily fully correct. In order to investigate the sensitivity to variations in the x shape we have switched off the QCD evolution, thus making the distribution somewhat harder, and found that our final results, in the form of cross sections, decrease about 5% (for Fermilab) to 10% (for HERA). These small changes do not alter the conclusions of our study.

In order to have a proper description of the hadronic final state we include QCD corrections by standard initial- and final-state parton showers [14]. A primordial transverse momentum of order m_c is introduced through a Gaussian distribution with a width of 1 GeV. Hadronization is finally performed using the Lund string model Monte Carlo JETSET 6.3 [15]. Here, a complication arises since the scattered charm quark (antiquark) leaves a more complex proton remnant which, in addition to the three valence quarks, contains the "partner" charm antiquark (quark) to conserve the charm quantum number, cf. Fig. 2(a). Following the general strategy in LEPTO we let this partner pick up a random quark (diquark) to form a charm meson (baryon), taking, from the complete spectator system, an energy-momentum fraction chosen according to the characteristically hard charm fragmentation function of Peterson *et al.* [16], which is related theoretically to the intrinsic charm quark distribution. The remaining spectator diquark (quark) will then be connected to the scattered charm quark (antiquark) with the Lund string via any gluons emitted in the parton shower (Fig. 2). The details of this modeling of the target remnant is not important for our purposes, since it only affects the leading particles in the "spectator jet" which are lost in the beam pipe in an *ep* collider and are very soft in a fixed target configuration.

The PGF background is simulated using AROMA 1.2 [17], which is based on the exact matrix elements for $\gamma g \to c\bar{c}$ in leading-order QCD [18]. The next-to-leading-order corrections to this process have been available for some time for the photoproduction case ($Q^2 \approx 0$) [19] and have very recently been calculated also for DIS [20,21], which is of direct interest in our case. These corrections increase the cross section by 50–100% [21]. Although there are some changes in the shapes of differential distributions, one may as a first approximation regard the corrections as an overall normalization change. These order $\alpha^2\alpha_s^2$ cross sections are not implemented in general purpose Monte Carlo generators, giving complete final states which are needed for our study of the experimental measurement of charm production. Our results are therefore based on the leading-order formalism, which should provide a good approximation with respect to the shapes of differential distributions, but the overall normalization may need to be adjusted by a simple rescaling. The influence of higher-order corrections on the event properties are, however, included in AROMA through multiple parton emission in terms of parton shower development from the $c\bar{c}$ state. To get a complete final state of observable hadrons the Lund string model for hadronization is applied in its Monte Carlo form [15]. To check the rate of muons from decays of noncharm particles we simulate normal DIS events using LEPTO 6.1 [14] and JETSET 7.3 [15] and allow particles to decay within the volume of the detector under consideration.

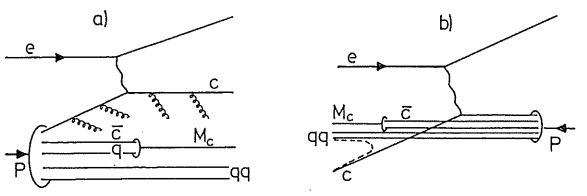


FIG. 2. Deep-inelastic scattering on an intrinsic charm quark illustrating (a) the simulation model and (b) the typical event topology in the HERA laboratory frame.

III. *ep* COLLISIONS AT HERA

In the HERA collider at DESY, 30-GeV electrons collide with 820-GeV protons giving a c.m. system (c.m.s.) energy $\sqrt{s} = 314$ GeV. The invariant mass W of the hadronic final state is normally far above the charm production threshold [$W^2 = Q^2(1-x)/x + m_p^2 \gg \sim 4m_c^2$] and also Q^2 can easily be chosen large enough that charm quark mass effects can be safely neglected (cf. Sec. II and [11]). The two general purpose experiments, H1 and ZEUS, cover essentially the whole solid angle except for $\sim 4°$ around the beam pipes. Data taking has started this year and should later give data samples of 100 pb^{-1}/year with the design luminosity. In this section we present results obtained from the application of our simulation models for the IC signal and the background processes to the case of *ep* scattering at HERA.

An overview can be obtained in terms of the effective charm structure function F_2^c, shown in Fig. 1(b), which is extracted from Eq. (2) where the differential cross section has been generated using our model simulations for IC and PGF, respectively (accounting for mass threshold effects). As can be seen, the IC contribution dominates over PGF at large x, but the crossover point depends on the overall normalization of the IC component. (The mean Q^2 in an x bin is, however, somewhat lower for IC than for PGF and will therefore give a correspondingly larger cross section when the $1/Q^4$ factor from Eq. (2) is applied to F_2^2.) In the following we investigate the properties of IC events and how an enriched sample of them can be obtained.

A. Event topology

The event topology at HERA is strongly influenced by the large boost of the *ep* c.m.s. system in the direction of the proton beam. Large momentum fractions x of the incoming quark imply that it is scattered to a small angle (unless Q^2 is very large, which is suppressed in the neutral current (NC) cross section). This event topology, which is characteristic for scattering on intrinsic charm quarks, is illustrated in Fig. 2(b). Based on our simulation we show in Fig. 3 representative rapidity distributions for those bins in x and Q^2 where the IC contribution is relatively large. At the parton level (top row) we

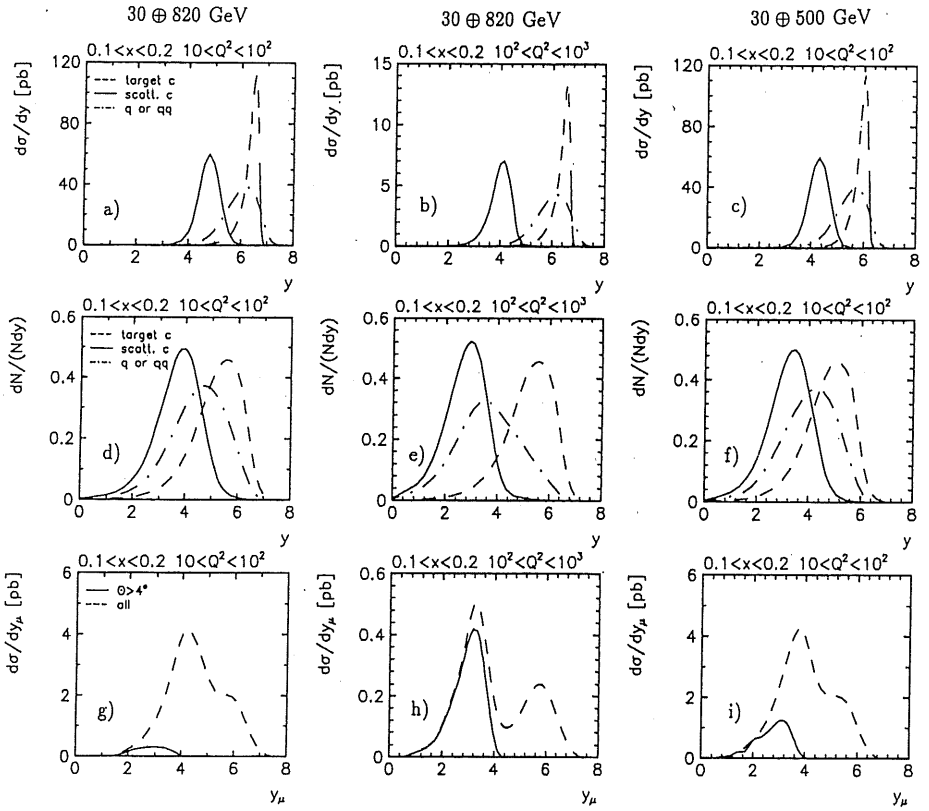


FIG. 3. Rapidity distributions for IC events in the HERA laboratory frame. The first two columns are for the nominal HERA beam energies (30⊕820 GeV) and the rightmost column for a lower proton energy (30⊕500 GeV). The top row (a)–(c) are at the parton level, the middle row (d)–(f) are at the hadron level (normalized to unit area), and the bottom row (g)–(i) are muons from semileptonic charm decays comparing all muons (dashed curve) and the detectable ones outside the beam pipe (full curve). Cf. Sec. III A and Fig. 2.

notice that the charm hadron formed from the proton remnant (target c) is emitted at large forward rapidities (i.e., along the proton beam) as is also the case for the remaining proton fragment (q or qq) although this has a somewhat broader rapidity distribution. The scattered IC quark emerges at relatively smaller, but still very much forward, rapidities. With increasing Q^2 [Fig. 3(b)] or reduced proton energy [Fig. 3(c)] it moves towards lower rapidities. After parton showers and hadronization the same picture holds (middle row in Fig. 3) although the distributions are shifted down in rapidity. With a beam pipe cut of $\theta \geq 4°$, which corresponds to pseudorapidity $\eta \leq 3.4$, it is obvious that there will be large acceptance losses not only of the proton remnant jet (as usual) but also of the current quark jet. This is also the case for the muons from semileptonic charm decays as shown in the bottom row of Fig. 3. It is evident that the basic scattering kinematics for IC at an ep collider is not completely compatible with the possible acceptance coverage of an experiment. This is especially true at low Q^2 and for $x \geq 0.2$ where the cross section is largest.

A way to reduce the inconveniences caused by the boost would simply be to reduce the proton-beam energy. This is evident from the example of $E_p = 500$ GeV (on $E_e = 30$ GeV) shown in the rightmost column in Fig. 3 and to be compared with the same x, Q^2 bin in the leftmost one. The total cross section for IC is decreased by just a few percent, but particle distributions are shifted to lower rapidities such that the observable cross sections are actually increased. The gain in the detectable muon rate is clearly seen by comparing Figs. 3(g) and 3(i).

The observability of the hadronic final state varies with x and Q^2, as found from the reconstruction of jets using the JADE algorithm on all particles outside the beam pipe [22]. At low Q^2 most of the hadrons are escaping down the beam pipe which makes a reliable jet reconstruction difficult. As Q^2 increases or x decreases the current jet will move into the detector volume and two-jet events become predominant where the second jet often originates from the target remnant [cf. Figs. 3(d) and 3(e)].

Although the PGF background events normally have another topology with two charm jets plus the spectator, the forward boost also affects these events such that one charm jet may be lost in the beam pipe making these events more like IC events. Normal DIS events with muons from π and K decays have the same overall topology as IC, but may differ in the details.

We have here only considered muons to tag charm. Electrons and positrons from semileptonic charm decays would have the same distributions and could in principle also be used, but they are harder to identify when they are not well isolated from hadrons. There are also more background $e^{\pm}$ from various other particle decays.

B. Cross sections and signal-to-background ratios

Table I gives the cross sections for both intrinsic charm and the background processes in relevant x, Q^2 bins, with the ones we consider most promising for an IC search based on muons emphasized in boldface. Compared to the PGF background (obtained from leading-order QCD as discussed above), which is expected to give

TABLE I. Cross sections.

	σ_{tot} (pb) Q^2 (GeV^2)			σ_μ ($\theta_\mu > 4°, E_\mu > 4$ GeV) (pb) Q^2 (GeV^2)		
x	10^1–10^2	10^2–10^3	10^3–10^5	10^1–10^2	10^2–10^3	10^3–10^5
			IC cross sections			
0.05–0.1	24.0	3.40	0.33	0.72	0.24	0.02
0.1–0.2	59.0	6.60	0.67	**0.40**	**0.49**	0.05
0.2–0.5	75.0	8.10	0.57	**0.03**	**0.24**	0.05
0.5–1.0	1.3	0.29	0.00	0.00	0.00	0.00
			PGF $c\bar{c}$ cross sections			
0.05–0.1	170.0	50.0	4.50	8.02	4.90	0.38
0.1–0.2	49.0	19.0	2.00	**0.40**	**1.40**	0.21
0.2–0.5	5.9	3.3	0.43	**0.00**	**0.12**	0.04
0.5–1.0	0.0	0.0	0.00	0.00	0.00	0.00
			DIS cross sections (μ from π, K decays)			
0.05–0.1	6300	660	55.0	26.00	12.00	0.90
0.1–0.2	5700	560	53.0	5.60	8.30	1.20
0.2–0.5	3800	430	41.0	0.15	2.10	1.00
0.5–1.0	80	40	3.4	0.00	0.01	0.03
			IC cross sections (30⊕500 GeV)			
0.05–0.1	24.0	3.3	0.26	1.10	0.18	0.01
0.1–0.2	59.0	6.6	0.59	**1.30**	**0.49**	0.04
0.2–0.5	75.0	8.8	0.53	**0.47**	**0.57**	0.04
0.5–1.0	1.3	0.3	0.00	0.00	0.00	0.00
			PGF $c\bar{c}$ cross sections (30⊕500 GeV)			
0.05–0.1	180.0	46.0	3.50	8.70	3.30	0.22
0.1–0.2	49.0	17.0	1.70	**1.40**	**1.60**	0.16
0.2–0.5	5.7	3.2	0.38	**0.03**	**0.27**	0.03
0.5–1.0	0.0	0.0	0.00	0.00	0.00	0.00

a larger overall cross section but decrease faster with x due to the softness of the gluon distribution, we find that the IC cross section is smaller for $x \leq 0.1$, but of similar magnitude as PGF for $0.1 \leq x \leq 0.2$ and then substantially larger at higher x. For $x \geq 0.5$ or $Q^2 \geq 10^3$ GeV^2 the absolute IC cross section is too small to be measurable with the expected HERA luminosity, and we are therefore left with the region $0.1 \leq x \leq 0.5$, $10^1 \leq Q^2 \leq 10^3$ for further investigation. Some differential cross sections for IC in this region are shown in Fig. 4. The characteristic hard x distribution and decrease with Q^2 (photon propagator) are strongly shifted to lower x and higher Q^2 when including the requirement of having both the scattered electron and charm quark within the detector ($4° \leq \theta \leq 176°$), in which case they have sizable energies and transverse momenta, which favors their detection [Figs. 4(c) and 4(d)]. The latter distribution reflects the Q^2 dependence since with naive parton level kinematics the quark and electron have balancing transverse momenta given by $p_\perp = \sqrt{Q^2(1-y)}$. Since $x \geq 0.1$ and $Q^2 \leq 10^3$ in our kinematic region, $y = Q^2/xs$ becomes ≤ 0.1, i.e., the electron energy is lowered by at most 10% leading to the narrow peak in the electron energy distribution Fig. 4(c). As y approaches smaller values it becomes increasingly difficult to reconstruct the kinematic variables from the electron alone. However, the range in which the kinematics can be reconstructed can be extended by also using the information from the hadronic system via, e.g., the double-angle method [23].

In the favored region for IC, the cross section for charm production and observable muons in the PGF background process is of similar magnitude, compare, e.g., the boldface bins in Tables I(a) and I(b) (where muons are required to be within the detector acceptance

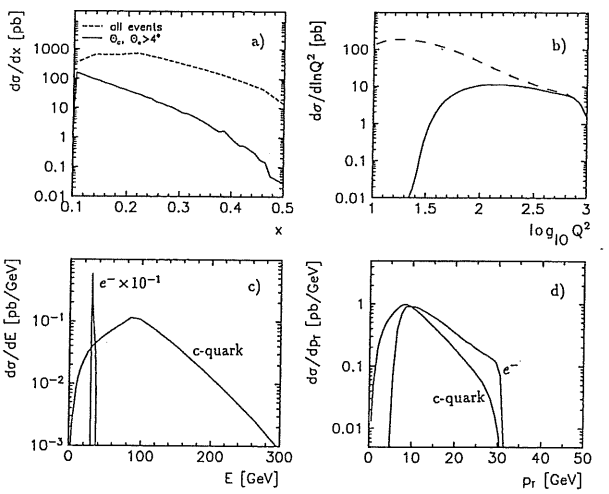


FIG. 4. Intrinsic charm cross-sections differential in (a) x, (b) $\log_{10} Q^2$ and (c) energy, (d) $p_\perp$ of scattered electron and quark for all events (dashed curves) and for events where both the scattered electron and the charm quark are within the detector ($4° \leq \theta \leq 176°$) (full curves).

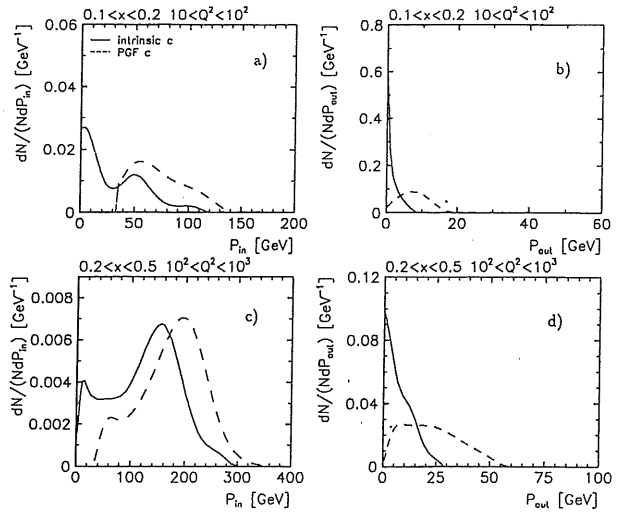


FIG. 5. Distributions of P_{in} and P_{out} for IC and PGF$\rightarrow c\bar{c}$ events for two of the bins in x and Q^2 marked in Tables I(a) and I(b).

and have energy above 4 GeV to ensure their identification). Although this may be sufficient to prove an excess in the charm production over the conventional process, one would like to obtain a cleaner IC sample. This can be based on the expected difference in event topology as discussed above. Letting the current jet direction (taken as the reconstructed jet at the largest polar angle) and the beam axis define a plane, we consider the sum of all particle momenta in this plane (P_{in}) and out of the plane (P_{out}). The differences observed in these quantities, see Fig. 5, reflect the extra c quark jet in PGF which gives additional transverse momenta and makes the PGF events less planar. (The peak at small P_{in} in Figs. 5(a) and 5(c) is due to the partial loss of a jet in the beam pipe.) A cut, in particular in P_{out}, can then be applied to suppress the background without affecting the signal much, as exemplified in Table II based on the muon sample.

Muons from ordinary meson decays constitutes a problem since the low decay probability in the detector is compensated by the large cross section for π and K production in normal DIS events as indicated in Table I(c). Given the softer energy and $p_\perp$ distribution of these muons, see Fig. 6, an increased cut in muon energy (or a cut in $p_\perp$) would improve the signal-to-background ratio, but a cut which gives an acceptable ratio would lead to a significant loss of the signal cross section. We have checked that these remaining high-energy background muons cannot be excluded based on their tracks having

TABLE II. Separation of IC and PGF.

x	Q^2 (GeV^2)	P_{out} cut (GeV)	σ_{IC} (pb)	Signal / background
0.1–0.2	10^1–10^2	7.5	0.38	3.5
0.1–0.2	10^2–10^3	10.0	0.33	1.0
0.2–0.5	10^2–10^3	20.0	0.24	4.4

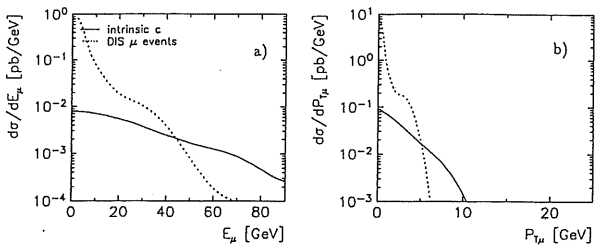


FIG. 6. Energy and transverse momentum distributions of muons in IC events (full curves) and from π, K decays in normal DIS events (dotted curve). $0.2<x<0.5, 10^2<Q^2<10^3$.

observable kinks or being incompatible with originating from the primary vertex. However, since the muons in IC come from the decay of a leading particle they should be more strongly correlated with the quark jet direction as compared to the majority of decay muons from π, K which are not necessarily leading particles. We have therefore compared the distance $\Delta R=\sqrt{(\eta_j-\eta_\mu)^2+(\phi_j-\phi_\mu)^2}$ between the muons and the reconstructed current jet and indeed found such a difference, but the background suppression due to this difference alone is still not sufficient. Thus, there is no simple cut to avoid this background, but one may investigate the feasibility of neural networks, which have proven useful for classification problems based on multidimensional inputs. The alternative way is to subtract such decay muons since they come from well-known decay processes of particles whose production it should be possible to control sufficiently well based on direct measurements and, perhaps, complemented with tuned Monte Carlo event generators. Therefore, we do not consider this background to be a major problem.

IV. MUON SCATTERING ON FIXED TARGET

From the last section it is evident that the main problem in observing DIS on charm quarks in *ep* collider experiments is due to the limited acceptance coverage close to the downstream proton beam pipe. This does not only apply when tagging charm through semileptonic decays into muons, but also in attempts to reconstruct charm particles through invariant masses (e.g., using $D^*\rightarrow D\pi$) since tracking and momentum measurements of high-energy particles at small forward angles close to the beam pipe are very difficult. These experimental problems can in general be reduced in fixed target experiments since a much smaller hole in the detector is needed to let through the primary lepton beam. A further advantage in a fixed target configuration is that it allows much more freedom to optimize the detector arrangement so as to obtain the best possible signal-to-background ratio. In particular, the hadron absorber can be placed close to the target in order to minimize the amount of muons from π and K decays.

In order for the cross section for IC not to be suppressed by the charm particle masses, the lepton-beam energy cannot be too low. For a beam energy of $E_e=30$ GeV (as for the HERMES experiment at HERA) the to-

tal IC cross section is only ~8 pb, so electron accelerators are practically excluded and we are left with muon beams from proton accelerators. However, once the lepton-beam energy is above a few hundred GeV, i.e., $\sqrt{s} \gtrsim 20$ GeV, the total IC cross section does not vary much with increasing energy. At Fermilab, with $E_\mu \simeq 470$ GeV ($\sqrt{s} = 30$ GeV), the IC cross section is still 20% higher than at CERN with $E_\mu = 280$ GeV ($\sqrt{s} = 23$ GeV), but only 30% lower than at HERA ($\sqrt{s} = 314$ GeV). Thus, the gain in cross section at collider energies is not essential.

These overall relations of cross sections at Fermilab and HERA energies are demonstrated in Figs. 1(b) and 1(c), which show the effective charm structure functions F_2^c (see the definition and comments in the second paragraph of Sec. III). Threshold effects are included in the simulations and one can therefore make a proper comparison between the different energies. Whereas the IC cross section does not increase much in this energy range, the PGF background does increase significantly. In the latter process, the $c\bar{c}$ pair is produced in the photon-gluon subsystem which has a lower invariant mass as a result of the, typically small, fractional energy of the proton carried by the gluon. Hence, the charm mass threshold gives a stronger effect in this case, resulting in a stronger energy dependence. Comparing F_2^{IC} and F_2^{PGF} in Figs. 1(b) and 1(c), it is evident that the signal-to-background ratio is more favorable at Fermilab energies than at HERA and that the crossover point in x, over which IC dominates, is favorably located at smaller x values in the Fermilab case.

As far as luminosity is concerned, the many-orders-of-magnitude larger density of protons in a liquid or pressurized gas target compared to a circulating proton beam are compensated by the lower number of muons per second impinging on the target as compared to the number of electrons in an *ep* collider. To evaluate the attainable luminosities with the Fermilab muon beam we use parameters typical for the E665 experiment [24]: muon flux of 2×10^7 muons in a 20 s spill length with 57 s cycle time. The typical target thickness used was 10 g/cm^2 resulting in luminosities of the order 10^{30} cm^{-2}s^{-1}. Assuming a 50% efficient running time, this would in turn give an integrated luminosity of about 15 pb^{-1} per year. Comparing this with the design luminosity 10^{31} cm^{-2}s^{-1} at HERA, giving 100 pb^{-1} per year, and considering the similar magnitude of the IC cross section, it is clear that the statistics attainable in a fixed target experiment can only be comparable to or larger than at HERA if either a denser target is used and/or the acceptance losses are correspondingly smaller. In order to clarify the acceptance situation for the fixed target case one needs to investigate the detailed properties of the IC and the background events.

A. Event topology

In a fixed target experiment the μp c.m.s. is strongly boosted along the muon beam and hence the scattered muon and the current jet will appear at rather small forward angles. As an example, in order to give definite numerical results, we will in what follows take the case of a 470-GeV muon beam on a fixed proton target corresponding to a feasible experiment at Fermilab. The polar angles of the scattered muon and the current jet, i.e., the scattered quark using quark-parton model kinematics, is shown in Fig. 7 for the region in x, Q^2 of interest for IC. In order to determine the event kinematics, i.e., x and Q^2, one needs to measure the scattered muon and therefore, according to Fig. 7, the region $0.25° \lesssim \theta_\mu \lesssim 5°$ has to be covered. Similarly, Fig. 7 shows that the angle of the scattered quark varies between ~1° and 25°. A muon from the decay of a scattered charm quark would essentially emerge at the same angle since the charmed particle momentum is normally much larger than the transverse momenta generated in its decay.

The distribution of muons from IC decays in terms of their energy and angle is shown in Fig. 8. As expected from basic kinematics of the scattered quark, there is a correlation such that higher muon energies occur at smaller angles. Furthermore, the muon angle tends to increase and its energy decrease for larger x [compare Figs. 8(a) and 8(b)]. The scattered intrinsic charm quark (antiquark) will leave its partner charm antiquark (quark), from the intrinsic $c\bar{c}$ pair, in the target remnant. As described in Sec. II, the latter will therefore produce a charmed meson (baryon) in the target fragmentation region, i.e., at small laboratory energies but possibly much larger angles than hadrons in the current jet. Semileptonic decays of such charmed hadrons gives rise to the component of low-energy muons extending to larger angles as shown in Figs. 8(c) and 8(d). These muons provide another signature for IC and one should therefore attempt to detect them, too.

The identification of muons is normally based on their ability to penetrate large amounts of material. With the exception of neutrinos, which are not seen in the detector in any case, this is not true for other particles. The experimental technique is thus to use an absorbing material

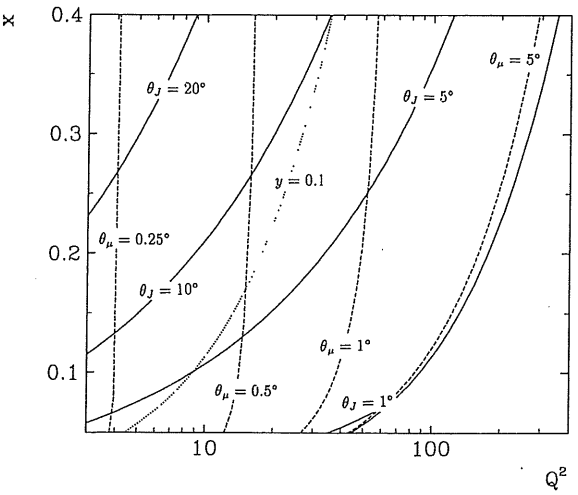


FIG. 7. Isolines for constant polar angles θ_J of the current jet, i.e., the scattered quark (full lines), and θ_μ of the scattered muon (dashed lines) in the kinematic x, Q^2 plane for DIS of 470 GeV muons on a fixed nucleon target. Lines of constant y are also shown.

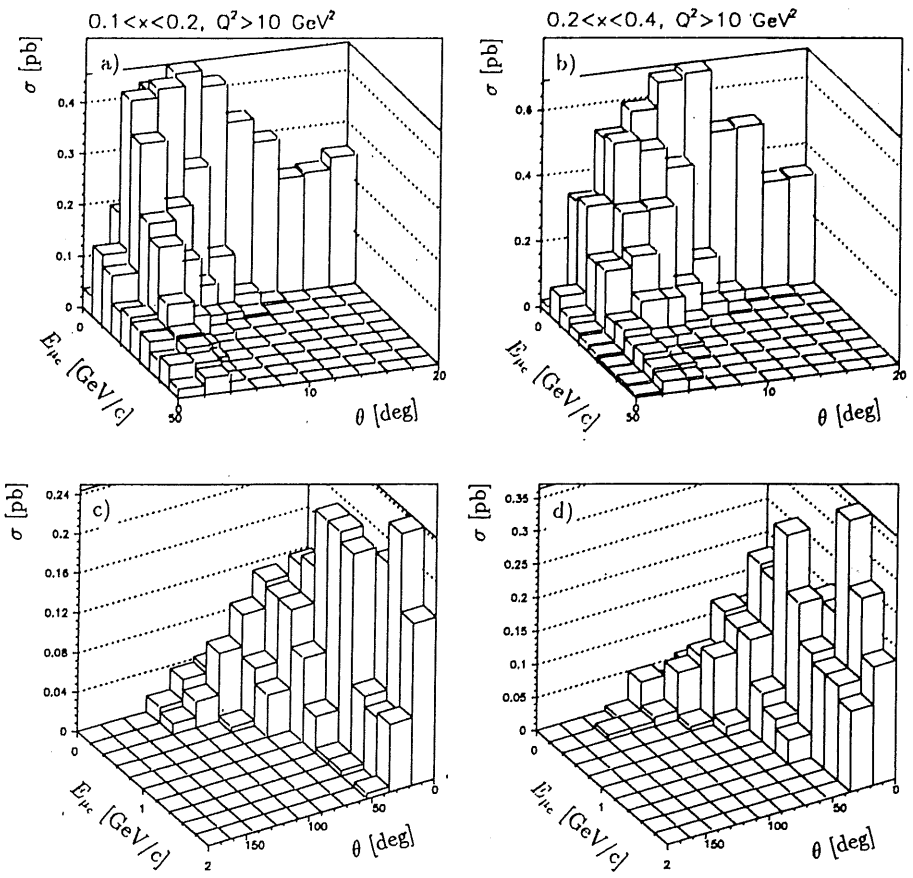


FIG. 8. Differential cross section (pb/bin) of muons in IC events vs muon energy and polar angle for DIS of 470 GeV muons on a fixed nucleon target. The large energy muons at smaller angles originate from the decay of the scattered intrinsic charm quark (antiquark), whereas those at lower energies and larger angles [better seen in (c) and (d) with adjusted scales] are dominantly from the decay of the partner charm antiquark (quark) left over in the target remnant.

of appropriate thickness to filter away all other particles except the muons.

B. Cross sections and signal-to-background ratios

The cross section for the IC signal and the leading-order PGF background is given in Table III for the most relevant bins in x and Q^2. For the total cross sections no experimental restrictions have been included, but for the muon cross sections we have assumed a fully efficient detector covering the polar angle region from 0.25° to 10° where the lower limit corresponds to $Q^2 \sim 5$ GeV^2 (see Fig. 7). The muons were required to have an energy of at least 10 GeV, based on our estimations of the conditions at the E665 experiment, which should allow them to pass a thick filter giving a safe absorption of the hadrons. From Fig. 8 it is clear that if this energy requirement can be somewhat relaxed a sizable gain in the muon rate will result. For example, taking a minimum muon energy of 5 GeV will increase the cross section by 60%, which has to be balanced against the increased probability for hadrons passing through a correspondingly less thick absorber.

Although it is clear from Table III that a lower limit for Q^2 of 5 GeV^2 gives a substantially higher total cross section compared to $Q^2 \geq 10$ GeV^2, there is only a small gain in observable muons. This is due to the resulting lower y values giving lower current jet energies, which in turn results in lower decay muon energies such that the 10-GeV cut is more noticeable. Because of this minor difference in the observable cross section we can allow the lower angular limit to increase to 0.4°, which corresponds to $Q^2 \sim 10$ GeV^2 as is realized from Fig. 7. This of course will simplify the experimental situation. From Fig. 8 we notice that a cut on the muon energy at 10 GeV will remove essentially all muons above 10°, which we therefore have chosen as our upper limit in the angle.

When comparing with the corresponding results at an *ep* collider (Table I) one should note that although the larger c.m.s. energy gives a larger *total* IC cross section, the *observable* muon cross section is smaller due to the re-

TABLE III. Cross sections.

	σ_{tot} (pb) Q^2 (GeV^2)			σ_μ (pb) Q^2 (GeV^2)		
x	5–10	10–45	45–	5–10	10–45	45–
		IC cross sections				
0.05–0.1	21.0	15.0	1.0	0.60	0.99	0.08
0.1–0.2	49.0	43.0	6.0	0.47	**1.70**	0.50
0.2–0.4	6.1	54.0	11.0	0.02	**1.60**	**0.81**
0.4–0.6		3.9	2.7		0.08	**0.16**
		PGF $c\bar{c}$ cross sections				
0.05–0.1	36.0	89.0	8.2	1.30	8.50	1.10
0.1–0.2	4.9	26.0	9.5	0.05	**1.70**	1.20
0.2–0.4	0.1	2.9	2.7	0.00	**0.10**	**0.29**
0.4–0.6		0.0	0.1		0.00	**0.01**

duced acceptance along the downstream proton beam pipe. In the favored x, Q^2 regions the cross section at the fixed target can be a factor ~ 4 larger, with a muon energy cut at 10 GeV, which does not quite outbalance the lower luminosity when using a low-density target.

In order to illustrate the influence of the muon background from decaying pions and kaons, we have in Fig. 9 plotted the resulting signal/background ratio as a function of the decay distance available before the hadron absorber. To suppress this background the hadron filter should obviously be placed as close as possible to the target.

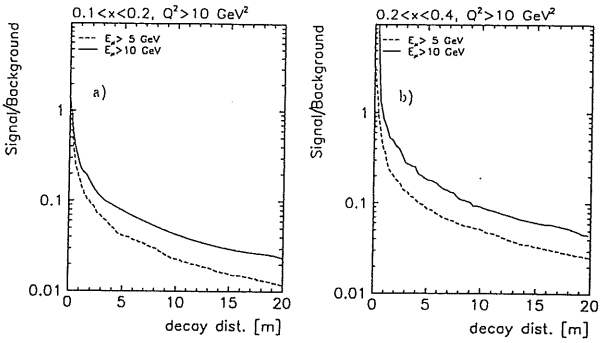


FIG. 9. Ratio of muons from the decay of scattered intrinsic charm quarks and from π and K decays in normal DIS events vs the available decay distance before the hadron absorber in the experiment. The muons are required to have a minimum energy of 5 GeV (dashed curve) or 10 GeV (full curve) as example criteria for identification in a muon detector. In the kinematic region used in (a) and (b), the signal cross section is 4.6 pb requiring $E_\mu > 10$ GeV and 7.5 pb requiring $E_\mu > 5$ GeV.

C. A dedicated intrinsic charm experiment

With the luminosities discussed above the rate of observable IC events, when tagged through decay muons, are about the same at the fixed target and the collider configuration. However, the prospects for improvements seem better in the fixed target case. The muon-scattering experiment E665 [24] at Fermilab, with 470-GeV muons on a liquid or pressurized gas target, is unfortunately not very well designed for IC measurements. Since the experimental program involved a study of the produced hadronic system in a spectrometer, the hadron filter and muon detector was forced to be placed far downstream (26 m) of the target. This not only gives a large muon background from π and K decays, but also limits the detection of high-energy muons to within $\sim 3°$. This is no problem for the high-energy scattered muon, but for muons from charm decays it results in a substantial acceptance loss as can be seen from Fig. 8. For intrinsic charm we are only interested in detecting and measuring muons and do not care about the hadrons produced. Thus, the muon detector can be placed close to the target to increase the acceptance. Furthermore, a more dense target could preferably be chosen to increase the luminosity. By using iron the luminosity could easily be increased by one to two orders of magnitude without running the risk of absorbing the muons of interest, but one has to consider the increased radiative corrections. With such an arrangement the target would also serve as a first hadron filter.

The most energetic hadrons determine the thickness of the filter, but since in the IC process the muons are produced over a broad momentum range, one has to split up the filter into a sandwich structure where absorbing plates are interleaved with detector planes in order not to absorb the lower momentum muons before they can be identified. The muons from decays of the scattered charm quark will have a similar or higher energy compared to the hadrons in the current jet and should therefore be possible to identify. Also the low-energy muons at larger angles, arising from charm in the target remnant, should in principle be detectable since hadrons at those angles should also come from the target remnant hadronization and therefore have similar or even lower energies.

The requirements for a dedicated intrinsic charm experiment, at beam conditions such as at Fermilab, are set by the following criteria.

(i) In the kinematic region of interest the scattered muon is emitted in the angular range $\sim 0.4°-5°$ and with momenta between 50 GeV/c and 400 GeV/c. Since the scattered muon has to be momentum determined to extract the kinematics of the event a magnetic spectrometer with high bending power and good spatial resolution is needed.

(ii) The incoming muon beam has a certain width of typically ± 3 cm which implies that the spectrometer has to be placed at a distance from the target of at least 5 m to allow muons scattered by 0.4° to exit the primary beam.

(iii) The momentum spread of the Fermilab primary muon beam has a sigma of about 60 GeV/c at a nominal beam energy of 500 GeV. Therefore, the magnetic spectrometer must be able to distinguish the scattered muon from primary muons of comparable momenta. This can be achieved as in the E665 experiment with two dipole magnets with opposite polarities or with a toroid magnet with a central hole for the primary beam.

(iv) The decay muon from the scattered intrinsic charm quark emerges in the angular range 1°–20° and with momenta between a few and 50 GeV/c, while the muons from charm in the target remnant would be softer and have emission angles extending above 50°. It is, strictly speaking, not necessary to measure the momenta of these muons since they are only used to tag the presence of charm. To identify these muons over such a momentum range needs a sandwich arrangement of absorbers and detectors as described above. The detector could have one forward part covering angles up to $\sim 45°$ and a barrel part for the angular range $\sim 45°-90°$ and it should be placed close to the target to minimize the background from decaying pions and kaons.

With a solid target giving one order of magnitude higher luminosity one could get more than 500 IC events per year. Since this number applies to a 1% normalization of the IC component in the proton, one could then reach the level 0.1%, or lower, and thus be more sensitive

than the present evidence based on European Muon Collaboration (EMC) data. Including the possibility to detect muons from the target remnant would add more data. It would also give the possibility to observe two muons in the same event, one from the scattered charm quark and one from the remaining charm in the target remnant, giving an even more characteristic signature for intrinsic charm.

V. CONCLUSIONS

We have found that the total cross section for scattering on intrinsic charm quarks are quite reasonable both at fixed target and HERA energies. However, since the cross section is dominant at large x and rather small Q^2, this implies a typical event configuration at an ep collider where the scattered charm quark, and hence its decay muon, emerge at small forward angles giving large acceptance losses due to the beam pipe. Any possibility to cover smaller angles will give a significant improvement, e.g., lowering the cut $\theta \geq 4°$ (applicable in the H1 detector) to 3° would increase the observable muon cross section by 70%. Alternatively, one can increase the observable cross section by lowering the proton-beam energy, since this would give a less forward boosted ep system. For example, $E_p = 500$ GeV would more than double the observable muon rate.

These acceptance problems are not present in a fixed target experiment which is designed to measure muons over a large angular region. Thus, the slightly smaller total cross section, due to the smaller c.m.s. energy, is more than compensated for the observable muon cross section, which is in fact a factor ~ 4 larger. The absolute level of the cross section for the muon signal from intrinsic charm in the preferred range $0.1 \leq x \leq 0.5$, $Q^2 \gtrsim 10$ GeV^2 is around 1 pb at HERA and 4 pb at Fermilab when a 1% normalization of the intrinsic charm quark density distribution is assumed. At HERA this should give a useful event sample with an integrated luminosity of 100 pb^{-1} attainable in a year. Thus with statistics of, say, 500 pb^{-1} obtainable in some years running it should be possible to probe the intrinsic charm content of the proton down to a level of 0.1%, i.e., somewhat below the level indicated by the analysis [11] of the EMC data.

A possible increase of the luminosity in a future upgrade of HERA would, of course, be very useful whereas an increased energy has no effect. By combining the HERA proton beam with a possible linear electron accelerator, with beam energy $\gtrsim 250$ GeV, we will face the problem of not being able to detect the scattered electron since it will emerge at very small angles. Only at quite large Q^2 this would be possible but then the IC cross section becomes too small.

In a dedicated fixed target experiment with a low-density target the observable event rate is similar to HERA, but it can be increased with a more dense target. With integrated luminosities of ~ 150 pb^{-1}/year one would get a factor of 6 more statistics compared to HERA and should be able to probe the 0.1% level of an intrinsic charm component in the proton in only one year of running.

There are two kinds of backgrounds to be considered. The first, and most serious one, is charm production through the photon-gluon fusion process. In the HERA case it is of the same magnitude as the signal in the preferred x, Q^2 range, while it is almost a factor 2 smaller in the fixed target situation. Given the different topology of these events we have demonstrated how they can be suppressed to an acceptable level. The second background is due to muons from decaying pions and kaons in normal DIS events. A fixed target experiment can be designed to minimize such a contribution by making the flight distance to the muon detector short. At HERA these muons are much more frequent than muons from charm, but they can be strongly suppressed by an energy cut and the remainder can be subtracted once the π and K momentum spectra are known sufficiently well.

We note that intrinsic beauty may also be present in the proton, but should be suppressed by the factor $m_b^2/m_c^2 \approx 10$ relative to intrinsic charm. Together with an increased heavy quark production threshold, this gives a total cross section which is about 5% of the IC cross section at HERA and still lower at fixed target energies. Although a larger fraction of the muons from bottom decays may enter a HERA detector, due to an increased transverse momentum, the observable muon cross sections will be uncomfortably small.

Finally, it is interesting to note that a dedicated fixed target experiment for intrinsic charm could also be used with the muon beam replaced by a proton beam. This would open the possibility to investigate the predicted J/ψ production [8,25] from the intrinsic charm quark pair. The basic idea is here that the $c\bar{c}$ pair has a smaller transverse extension ($\sim 1/m_c$) than the light quarks and a target nucleus may therefore act as a "filter" that absorbs the valence quarks, leaving a forward $c\bar{c}$ pair hadronizing into a J/ψ which can be detected through its decay to a muon pair at small forward angles.

ACKNOWLEDGMENTS

We are grateful to Stan Brodsky for several inspiring discussions and helpful comments and to H. Schellman for useful information on the E665 experiment. This work was supported by the Swedish Natural Science Research Council.

[1] D. Drijard *et al.*, Phys. Lett. **81B**, 250 (1979); **85B**, 452 (1979); K. L. Giboni *et al.*, *ibid.* **85B**, 437 (1979); W. Lockman *et al.*, *ibid.* **85B**, 443 (1979); A. Chilingarov *et al.*, *ibid.* **83B**, 136 (1979).

[2] S. J. Brodsky, P. Hoyer, C. Peterson, and N. Sakai, Phys. Lett. **93B**, 451 (1980); S. J. Brodsky and C. Peterson, Phys. Rev. D **23**, 2745 (1981).

[3] M. L. Mangano, P. Nason, and G. Ridolfi, Nucl. Phys. **B373**, 295 (1992); G. Altarelli *et al.*, *ibid.* **B308**, 724 (1988).

[4] B. Andersson, H. U. Bengtsson, and G. Gustafson, Lund University Report No. LU-TP 83-4, 1983 (unpublished).

[5] R. Vogt, S. J. Brodsky, and P. Hoyer, Nucl. Phys. **B383**,

643 (1992).
[6] NA3 Collaboration, J. Badier *et al.*, Z. Phys. C **20**, 101 (1983).
[7] C. Biino *et al.*, Phys. Rev. Lett. **58**, 2523 (1987).
[8] S. J. Brodsky, P. Hoyer, A. H. Mueller, and W.-K. Tang, Nucl. Phys. **B369**, 519 (1992).
[9] R. Vogt, S. J. Brodsky, and P. Hoyer, Nucl. Phys. **B360**, 67 (1991).
[10] J. J. Aubert *et al.*, Phys. Lett. **110B**, 73 (1982); Nucl. Phys. **B213**, 31 (1983).
[11] E. Hoffman and R. Moore, Z. Phys. C **20**, 71 (1983).
[12] V. Ganapathi and J. Smith, Phys. Rev. D **19**, 801 (1979).
[13] G. Ingelman *et al.*, in *Proceedings of the HERA Workshop,* Hamburg, 1987, edited by R. D. Peccei (DESY, Hamburg, 1988), Vol. 1, p. 3.
[14] G. Ingelman, LEPTO version 5.2 (unpublished); LEPTO version 6.1, in *Proceedings of "Physics at HERA"* [26], Vol. 3, p. 1366.
[15] T. Sjöstrand, Comput. Phys. Commun. **39**, 347 (1986); **43**, 367 (1987); JETSET 7.3, program manual, Report No. CERN-TH.6488/92 (unpublished).
[16] C. Peterson *et al.*, Phys. Rev. D **27**, 105 (1983).
[17] G. Ingelman and G. A. Schuler, AROMA 1.2 (unpublished); see also AROMA 1.4, in *Proceedings of "Physics at HERA"* [26], Vol. 3, p. 1346; Z. Phys. C **40**, 299 (1988).
[18] G. A. Schuler, Nucl. Phys. **B299**, 21 (1988).
[19] R. K. Ellis and P. Nason, Nucl. Phys. **B312**, 551 (1989); J. Smith and W. L. van Neerven, *ibid.* **B374**, 36 (1992).
[20] E. Laenen *et al.*, Phys. Lett. B **291**, 325 (1992); Nucl. Phys. **B392**, 162 (1993).
[21] E. Laenen *et al.*, Nucl. Phys. **B392**, 229 (1993).
[22] For jet finding in *ep* collider events, see, e.g., P. N. Burrows, G. Ingelman, and E. Ros, Z. Phys. C **39**, 257 (1988); M. Fleischer *et al.*, in *Proceedings of "Physics at HERA"* [26], Vol. 1, p. 303.
[23] S. Bentvelsen *et al.*, in *Proceedings of "Physics at HERA"* [26], Vol. 1, p. 23.
[24] E665 Collaboration, M. R. Adams *et al.*, Nucl. Instrum. Methods A **291**, 533 (1990).
[25] S. J. Brodsky and P. Hoyer, Phys. Rev. Lett. **63**, 1566 (1989).
[26] *Proceedings of "Physics at HERA,"* edited by W. Buchmüller and G. Ingelman (DESY, Hamburg, 1992), Vols. 1–3.

Appendix C

A Measurement of the Gluon Density in the Proton at Low x

H1-10/94-

A Measurement of the Gluon Density in the Proton at Low x

L. Jönsson[a], H. Küster[b], M. Nyberg-Werther[a], J. Stier[c]

[a] Physics Department, Lund University, Sölvegatan 14, S-223 62 Lund, Sweden
[b] III Physikalisches Institut, RWTH Aachen, Sommerfeldstrasse, D-5100 Aachen, Germany
[c] Deutsches Electronen-Synchrotron, Notkestrasse 85, D-22603 Hamburg, Germany

Abstract

A determination of the gluon density in the proton has been made in the region $5 \cdot 10^{-3} < x < 8 \cdot 10^{-2}$ by measuring the cross section of photon gluon fusion events in deep inelastic scattering with the H1 detector at the ep-collider HERA. This direct measurement of the gluon density was based on an integrated luminosity of 242 nb^{-1} and was performed in a kinematic region previously not accessible. The data points show a considerable increase in the gluon density with decreasing fractional momenta of the gluons.

1. Introduction

Deep inelastic lepton-nucleon scattering experiments have played a fundamental role for the understanding of the structure of matter. Ever since the discovery of the proton's parton content in the late sixties, extensive studies have been made over the years at accelerators providing increasingly higher energies, to obtain more detailed knowledge about the parton properties within the nucleons. Although it was soon realized that about 50% of the nucleon momentum was carried by gluons, a direct measurement of their momentum distribution has so far been constrained to large fractional momenta. Instead one has tried to extract information about the gluons from measurements made on the sea-quark distribution, relying upon the assumption that sea-quark pairs are produced by gluon decays in an evolution process starting with gluon emission from the valence quarks.

The electron-proton collider HERA has considerably extended the kinematic region available to investigations of the nucleon constituents. Especially it permits measurements of processes directly initiated by gluons in the proton. In order to extract the gluon density in the proton, we have in this analysis used photon-gluon fusion events where a gluon, emitted from one of the quarks in the proton, interacts with the virtual photon from the scattered electron to produce a quark-antiquark pair (Fig. 3.a). The typical final state thus contains two jets in addition to the proton fragment which to a large extent disappears undetected down the beam pipe. Such events are denoted (2+1) jet events to account for both the jets of the hard subsystem and the spectator jet originating from the proton fragment. The same final state is expected from the QCD Compton process, where a gluon is emitted by the scattered quark (Fig. 3.b). This process gives a background contribution which, however, can be accurately determined in the kinematic region which has been used for this analysis. Here the QCD-Compton process is initiated by quarks carrying a momentum fraction of the proton of the order of 10^{-2}, a region in which the quark density is well measured by previous experiments. The contribution from resolved photon processes can be neglected because of the Q^2 requirement in the selection of deep inelastic scattering (DIS) events.

2. The H1 Detector

A detailed description of the H1 detector is to be found elsewhere [1]. The construction concept is similar to that of other general purpose detectors at collider machines. Closest to the interaction point there are tracking devices surrounded by a calorimeter consisting of an electromagnetic and a hadronic section. Outside these detectors a superconducting coil provides a magnetic field parallel to the beam line and finally the instrumented magnet iron gives a rough measurement of the energy leaking out of the calorimeter and signals the presence of a muon track. At HERA one specifically has to pay attention to the forward direction i.e. the direction of the proton beam since the imbalance of the beam energies results in a strong boost of the events in this direction. Here we will only describe the detector parts which are of relevance to the measurement of the gluon density.

The backward electromagnetic calorimeter (BEMC) covers the angular range $155^o < \theta < 177^o$ where θ is defined with respect to the proton direction. Together with the track information from the backward proportional wire chamber (BPC), sitting right in front of the BEMC, and the reconstructed vertex position, the energy and scattering angle of the scattered electron can be measured in the kinematic region considered here.

The BEMC consists of 22.5 radiation lengths deep lead-scintillator sandwich stacks, each read out by two oppositely positioned wavelength shifter bars. This system gives an energy resolution of $\sigma(E)/E \approx 0.1/\sqrt{E}[GeV] \oplus 0.42/E[GeV] \oplus 0.03$. By adjusting the measured electron energy spectrum to the kinematic peak, the BEMC energy scale is known to an accuracy of 1.7%. The BPC has four wire planes giving a spatial resolution which together with the precision in the vertex reconstruction results in an angular resolution better than 5 mrad.

The hadronic final state is measured by the liquid argon calorimeter (LAr). The LAr calorimeter extends over the angular range $4^o < \theta < 153^o$ with a complete azimuthal coverage. The total depth varies between 4.5 and 8 interaction lengths. Test beam measurements have given a hadronic energy resolution of $\sigma(E)/E \approx 0.5/\sqrt{E}[GeV] \oplus 0.02$ [2]. The absolute hadronic energy scale has at present been measured to 5%.

3. Kinematics

In HERA 26.7 GeV electrons collide with 820 GeV protons which results in a centre-of-mass energy of 296 GeV. Fig. 1. shows the Feynman diagrams of the boson-gluon fusion process (PGF) and the QCD-Compton process (QCD-C) with the relevant kinematic variables indicated.

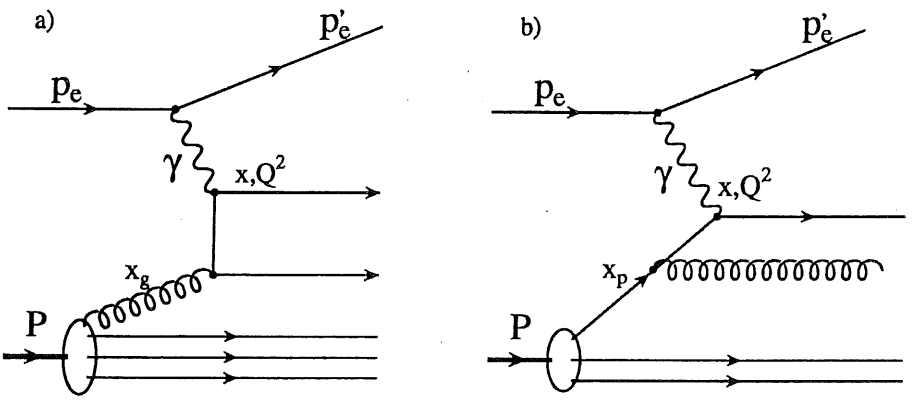


Figure 1: *The Feynman diagrams for a) the Photon-Gluon fusion process and b) the QCD-Compton process*

The kinematics of an event can be determined from two independent Lorentz invariant variables which are either of the two Bjorken scaling variables x or y and the

photon momentum transfer squared, Q^2. These variables are defined with the help of the four-momenta of the proton, P, the incoming and outgoing electron, p_e and p'_e, and the exchanged photon, q. Experimentally they are deduced from measurements of the energy, E'_e, and polar angle, θ_e, of the scattered electron and from the energy, E_h, and longitudinal momentum component, $p_{z,h}$, of the hadronic system according to the following relations:

$$Q^2 \equiv -q^2 = -(p_e - p'_e)^2 = -4E_e E'_e \cos^2(\theta_e/2)$$

$$y \equiv \frac{Pq}{2Pp_e} \quad \Rightarrow \quad y_{el} = 1 - (E'_e/E_e)\sin^2(\theta_e/2)$$

$$y_{hadr} = \sum_{hadrons} \frac{E_h - p_{z,h}}{2E_e}$$

$$x \equiv \frac{-q^2}{2Pq} = \frac{Q^2}{ys}$$

The centre-of-mass energy squared is given by $s = 4E_eE_p$, where E_e and E_p are the energies of the incoming electron and proton respectively.

If the momentum fraction carried by the parton entering the hard subprocess is denoted x_p and the momenta of the two partons produced are p_1 and p_2, the following relation is valid:

$$(x_pP + q)^2 = (p_1 + p_2)^2 = \hat{s}$$

Since $q^2 \equiv -Q^2$ and the proton mass can be neglected:

$$x_p = \frac{\hat{s} + Q^2}{2Pq}$$

Expressed in the Bjorken scaling variables one gets:

$$x_p = \frac{\hat{s} + Q^2}{ys} = x(1 + \frac{\hat{s}}{Q^2}) \tag{1}$$

Consequently, by measuring the hadronic energy deposited in the calorimeter, the invariant mass of the hard subsystem can be determined and thereby x_p. In the following we will call this method **the energy method.**

A complementary way of extracting x_p is based on the measurement of the jet directions in the hadronic centre-of-mass system, which is the rest system of the exchanged

photon and the proton. In this system the four-momenta of the proton and the photon are:

$$P = \left[\frac{E_e E_p y}{1-x}\right]^{1/2}(1,0,0,-1)$$

$$q = \left[\frac{E_e E_p y}{1-x}\right]^{1/2}(1-2x,0,0,1)$$

if the z-axis is pointing along the photon direction. We now use the relations $m_p\nu = 2E_eE_py$ and $2m_p\nu = Q^2/x$, where $\nu = E_e - E'_e$ is the energy transfer to the hadronic final state in the proton rest frame, to express the four-vectors in terms of the invariant mass squared of the hadronic system, $W^2 = 2m_p\nu - Q^2 + m_p^2$, and the momentum transfer squared, Q^2.

$$P = \frac{1}{2W}(W^2+Q^2,0,0,-(W^2+Q^2))$$

$$q = \frac{1}{2W}(W^2-Q^2,0,0,W^2+Q^2)$$

The energy of the interacting parton is x_pE_p and the rapidity of the photon-parton system is:

$$\eta = \frac{1}{2}ln\frac{E_\gamma + E_{parton} + p_{z,\gamma} + p_{z,parton}}{E_\gamma + E_{parton} - p_{z,\gamma} - p_{z,parton}}$$

$$\Rightarrow \quad e^{2\eta} = \frac{E_\gamma + E_{parton} + p_{z,\gamma} + p_{z,parton}}{E_\gamma + E_{parton} - p_{z,\gamma} - p_{z,parton}} = \frac{W}{-(Q^2/W) + 2x_pE_p}$$

$$\Rightarrow \quad x_p = \frac{W^2}{W^2+Q^2}\left(\frac{Q^2}{W^2} + e^{-2\eta}\right) = \frac{W^2}{W^2+Q^2}\left(\frac{Q^2}{W^2} + e^{-(\eta_{jet1}+\eta_{jet2})}\right)$$

This method will be called **the rapidity method.**

4. Trigger and Data Selection

The analysis is performed on DIS events which are selected in the detector by applying the trigger requirement of a local energy deposition, or equivalently a cluster, of more than 4 GeV in the BEMC detector. The proton beam induced background entering the BEMC from the wrong direction is removed already at the trigger level by using a TOF system behind the BEMC. The total event sample collected corresponds to an integrated luminosity of $242nb^{-1}$. After the event reconstruction has been performed, a clean DIS sample is obtained by applying the following off-line requirements:

- the scattered electron energy $E'_e > 10GeV$ which corresponds to $y \lesssim 0.625$ and has the effect of removing most of the photoproduction background, the events of which have the electron faked by a photon in the BEMC overlapped by a charged hadron.

- the momentum transfer squared $12.5 < Q^2 < 80GeV^2$ to ensure that the scattered electron is well inside the BEMC.

- $y \gtrsim 0.05$, in order to be able to reconstruct the event kinematics from the scattered electron.

- at least one charged track from the hadronic final state for the determination of the vertex position along the beam which had to be within 30 cm from the nominal position to suppress beam induced background.

- the hit in the BPC matching the cluster centre-of-gravity in the BEMC to within 5 cm and the lateral spread of the cluster to be less than 5 cm in radius.

- $y_{hadr}/y_{el} > 0.5$ to further suppress the photoproduction backgroud.

The data has to be corrected for the total relative efficiency of the cuts which are applied in the event selection procedure, as obtained from a comparison between data and Monte Carlo. Further details can be found in [7]. The 6% error in this efficiency will contribute to the overall normalisation error.

Out of this sample we select (2+1) jet events by applying a cone algorithm [3] with $\Delta R = 1$ to the events in the hadronic centre-of-mass system and by requiring the transverse jet energy $E_T > 3.5GeV$. Both reconstructed jets have to fall inside the angular range $10^o < \theta_{jet} < 150^o$ in the lab. system so as to be well within the coverage of the liquid argon calorimeter. A further purpose of the lower cut in angle is to remove the very forward region which is dominated by the proton fragments and initial state parton radiation which otherwise may give rise to a separate jet. The corrected invariant mass (see section 6.) of the two jets of the hard scattering system is required to be $\sqrt{\hat{s}_{rec}} > 10GeV$ to ensure well defined jet structures. Finally the difference in pseudorapidity between the two jets must be $\Delta\eta < 2$ in order to improve the determination of $\sqrt{\hat{s}_{rec}}$.

The resulting sample of (2+1) jet events contains 305 events consisting of PGF events, QCD Compton events and faked (2+1) jet events. What we denote faked events will be described in the following section. Since we have found no efficient method to distinguish between PGF and QCD Compton events we have to restrict ourself to a kinematic region

which is dominated by the PGF process. The region which was chosen for this analysis is constrained by defining ranges in the following two kinematic variables:

$$\begin{aligned} 12.5 < \quad & Q^2 \quad < 80 GeV^2 \\ 0.05 < \quad & y \quad < 0.625 \end{aligned}$$

which gives $2 \cdot 10^{-4} \lesssim x \lesssim 2 \cdot 10^{-2}$ and since the $\hat{s}$ region in which we have reasonable statistics is $10 GeV \lesssim \sqrt{\hat{s}} \lesssim 50 GeV$, we will be able to cover the range $0.005 < x_{g/p} < 0.08$.

5. Monte Carlo Generation

Within the kinematic region defined above we have used the Monte Carlo event generator LEPTO 6.1 [4] to calculate the detector acceptance for PGF and QCD Compton events and the cross section for faked (2+1) jet events. We also extract the resolution in the $x_{g/p}$ reconstruction. The LEPTO generator is based on QCD matrix element calculations up to the first order in α_s and the inclusion of parton showers accounts for higher order processes (MEPS). The QCD matrix element gives divergencies for soft and collinear emission which technically is avoided by defining a smallest invariant mass, m_{ij} of the created parton pairs. The normal implementation of this mass cut is through a cut parameter, y_c, which is defined by $m_{ij}^2 \geq y_c W^2$ and thus results in a correlation between m_{ij} and W. Depending on the choice of y_c, m_{ij} might vary between masses close to the divergence limit at low W and very high values at high W. This will thus result in high m_{ij} cuts at high W values although jet structures can still be resolved below this cut. It has also been demonstrated in [5] that high values of W do not necessarily correspond to high values of the hard subsystem mass $\sqrt{\hat{s}}$ for (2+1) jet events. We have therefore used a parametrisation for the cut in m_{ij} which follows the divergence limit with a 2 GeV margin. On the other hand we recall that in the event selection we have used a constant mass cut of 10 GeV and thus the two cuts are not equivalent. In the W range covered by this analysis the Monte Carlo cut is always below 10 GeV so that the Monte Carlo generated (2+1) jet events occasionally have hard subsystems with invariant masses below 10 GeV. If such a system, due to the limited resolution in the $\hat{s}$ determination, is given a mass greater than 10 GeV, we call it a faked event. Also events which are generated as (1+1) jet events but which are reconstructed as (2+1) jet events with an invariant mass of more than 10 GeV are denoted faked events.

A generated sample, corresponding to about four times the statistics of the data, was subject to full detector simulation followed by event reconstruction.

In order to get confidence in the results provided by the Monte Carlo generator we have made a number of control plots where we have compared the predictions of the Monte Carlo program with our experimental data. In Fig. 2 are shown distributions of pseudorapidity in the laboratory system and transverse energy for the most forward going and the most backward going jets respectively. Fig. 3a-d shows the energy flow with

respect to the jet axis as a function of the azimuthal angle and pseudorapidity for the two jets. The well described energy flow in the backward region (Figs. 3a and c) indicates that there is no significant contribution from resolved processes in our DIS sample. We have in particular looked into the Monte Carlo description of the energy flow in the forward region, $2 < \eta < 3$, which might contain a large contribution from initial state radiation. Although we have previously observed [16] that the MEPS model is not able to reproduce the energy flow in this rapidity region for an inclusive DIS sample, it is demonstrated from Fig. 3e that the energy flow of our jet sample is well described by MEPS. It can be noticed that not only the jet profile exhibits good agreement between data and Monte Carlo but also the level of the underlying energy flow. Other control plots, like the Q^2, W^2, y and x distributions, give further evidence for the ability of MEPS to reproduce the data.

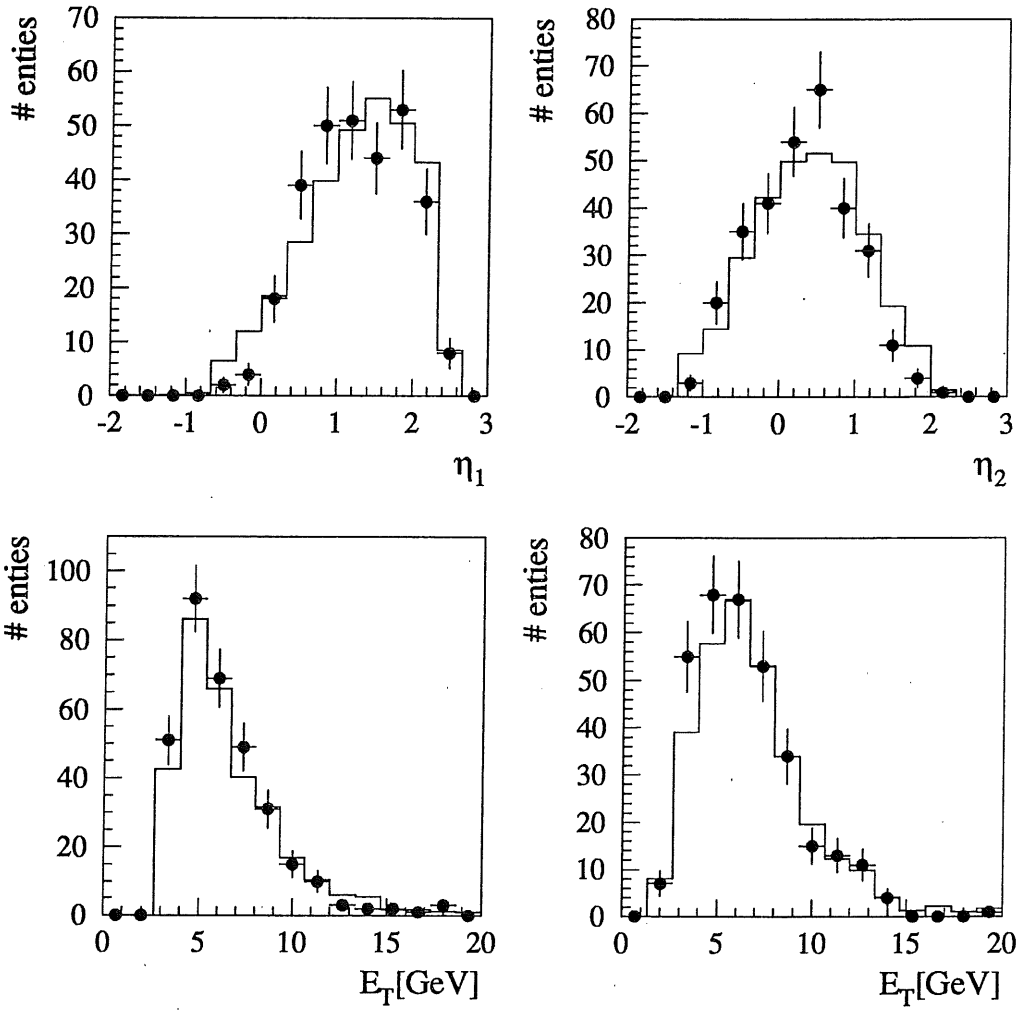


Figure 2: *Comparisons of data with Monte Carlo generated distributions on a) η and b) E_T for the most forward going jet and c) η and b) E_T for the most backward going jet in (2+1) jet events.*

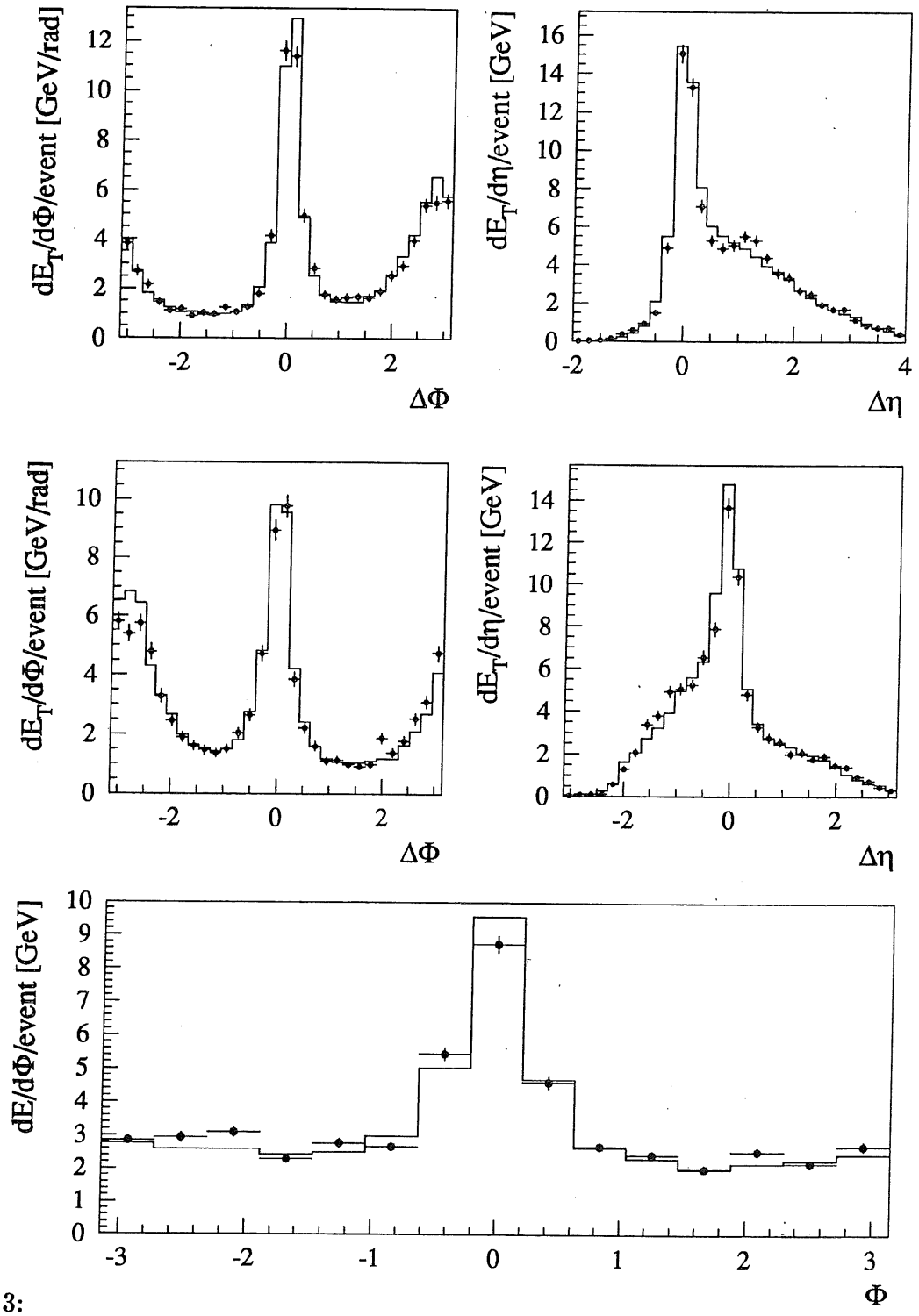


Figure 3:
Energy flow with respect to the jet axis in the laboratory frame as a function of a) η and b) ϕ for the most backward going jet, and as a function of c) η and d) ϕ for the most forward going jet. The energy flow in the region $2 < \eta < 3$ is shown in e).

6. Reconstruction of $x_{g/p}$

The Monte Carlo sample has been used to investigate the precision which can be obtained in the $x_{g/p}$ reconstruction, where $x_{g/p}$ gives the momentum fraction of the of the proton carried by the interacting gluon in the photon-gluon fusion process. One essential condition is to use a jet reconstruction algorithm which provides a good separation between the spectator jet and the jets of the hard subsystem. Since our event sample populates the region of small x and Q^2 values, the spectator jet will carry away most of the available energy and at least one of the other jets will proceed close to the spectator jet. For example a mistake by the jet algorithm in which one of the relatively energetic particles of the spectator jet is included into the hard subsystem might cause a strong impact on the reconstructed $\hat{s}$ value and thereby completely distort the $x_{g/p}$ reconstruction. It is therefore important to optimize the resolution parameters of the jet algorithms to give the best possible performance in this respect.

The threshold which has to be applied in order to eliminate the noise in the calorimeter leads to an underestimation of the jet energy and thereby of the reconstructed $x_{g/p}$ value for the energy method. The Monte Carlo results have been used to determine the systematic shift in $x_{g/p}$ and the distribution of the relative error in $x_{g/p}$ is shown in Fig. 4a for the energy method after correction for the shift.

The influence of colour strings spanned between the partons, initial state parton showers and some experimental effects lead to a tendency by the jet algorithms to systematically shift the reconstructed jet directions towards the proton direction. The resulting shift in $x_{g/p}$ has been corrected for in Fig. 4b which shows the relative error in the $x_{g/p}$ reconstruction for the rapidity method.

Comparing the results of the two methods we notice that the fitted Gaussian distributions result in a worse resolution for the energy method.

However, for properly reconstructed events both methods are expected to give consistent results whereas a misassignment of particles by the jet algorithm might have different impacts on the $x_{g/p}$ reconstruction from the two methods. Therefore an improved result should be obtained by only selecting events where the result of the two methods agree within a predefined accuracy. As a measure of this accuracy we have used $\Delta\sqrt{\hat{s}}$ which is given by the difference between the directly extracted $\hat{s}$ value from the energy method and the one obtained by introducing the $x_{g/p}$ value from the rapidity method into the expression (1). We have required a reconstruction agreement given by $\Delta\sqrt{\hat{s}_{rec}} \leq 10GeV$ and then we have simply taken the mean value of the two reconstructed values of $x_{g/p}$ according to:

$$x_{g/p}^{comb} = \frac{x_{g/p}^{E} + x_{g/p}^{\eta}}{2}$$

As shown in Fig. 4c the result of the combined method proves to be considerably better than those of the energy and rapidity methods separately in the respect that the tail has been suppressed and that the resolution has improved.

In Fig. 4d the correlation between the reconstructed and true $x_{g/p}$ is given for the combined method.

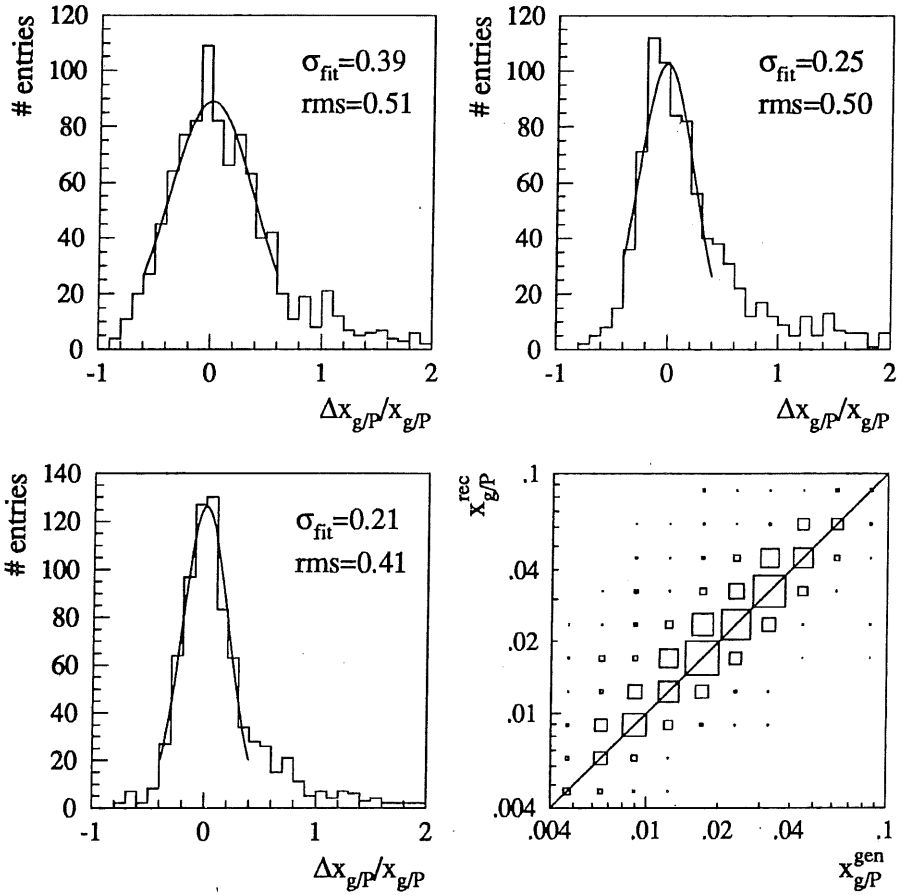


Figure 4: *The reconstruction of $x_{g/p}$ for the a) energy method, b) the rapidity method and c) the combined method. The correlation between the reconstructed and true $x_{g/p}$ for the combined method is shown in d).*

7. Cross Section Calculations and Unfolding of $G(x_{g/p}, Q^2)$

Within the acceptance region previously defined ($12.5 < Q^2 < 80GeV^2, 0.05 < y < 0.625, m_{ij} > 10GeV$) we can use Monte Carlo programs to calculate predictions for LO ($\mathcal{O}(\alpha_s)$) (2+1) jet cross-sections, given a certain structure function parametrization. Since LEPTO is not optimized for this task, we have used the PROJET [8] and DISJET [9] programs which give results that are in perfect agreement. Although LEPTO has only been used to extract the acceptance for (2+1) jet events and to estimate the background from faked (2+1) jet events in this analysis, we have checked that LEPTO, with a careful parameter setting, delivers cross sections which are consistent with those from the other programs. Fig. 7. shows the Monte Carlo event composition of PGF, QCD Compton and faked (2+1) jet events in each $x_{g/p}$ bin in the LEPTO sample. The overall signal to background ratio in the whole $x_{g/p}$ range comes out to be about 2:1.

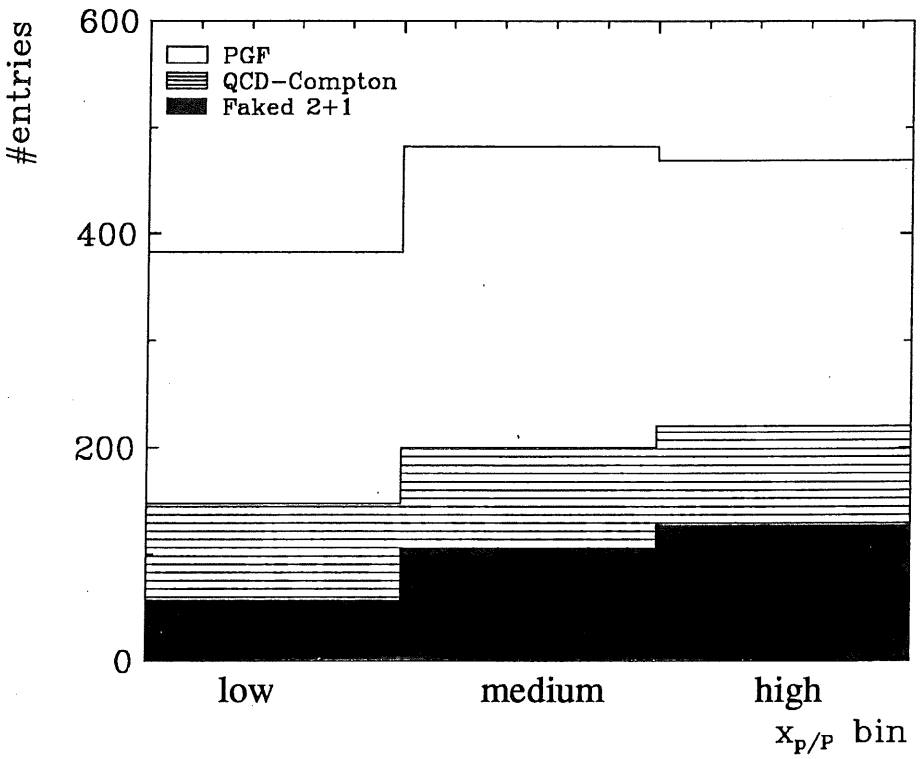


Figure 5: *The number of PGF, QCD Compton and faked (2+1) jet events in the LEPTO sample.*

If we define the PGF acceptance as the cross-section for PGF events in our fully simulated sample, surviving all the cuts, divided by the PGF cross-section on the parton level provided by LEPTO (and similarly for QCD Compton) we have:

$$A_{PGF} = \frac{\sigma^{PGF}_{LEPTO,full\ sim}}{\sigma^{PGF}_{LEPTO,parton\ level}} \quad ; \qquad A_{QCD-C} = \frac{\sigma^{QCD-C}_{LEPTO,full\ sim}}{\sigma^{QCD-C}_{LEPTO,parton\ level}}$$

We can now calculate a corrected PGF cross section, σ^{PGF}_{corr}, from our observed cross-section for (2+1) jets according to:

$$\sigma^{PGF}_{obs} = \sigma_{obs} - A_{QCD-C} \cdot \sigma_{QCD-C} - \sigma_{fake}$$

and

$$\sigma^{PGF}_{corr} = \frac{\sigma^{PGF}_{obs}}{A_{PGF}}$$

where σ_{obs}^{PGF} is the measured cross section after subtraction of the background cross section for QCD Compton events calculated with PROJET (σ_{QCD-C}) corrected for acceptance (A_{QCD-C}) and of the estimated cross section for faked (2+1) jet events (σ_{fake}). The subtraction constitutes about 30% of the total observed cross-section. By correcting the observed PGF cross section for acceptance (A_{PGF}) the corrected PGF cross section (σ_{corr}^{PGF}) is obtained.

The $x_{g/p}$ range covered by our kinematic region has been subdivided into three bins matching our reconstruction accuracy. For each of the three $x_{g/p}$ bins the corrected cross section has been calculated. The result on the corrected cross section is shown in Fig. 6.

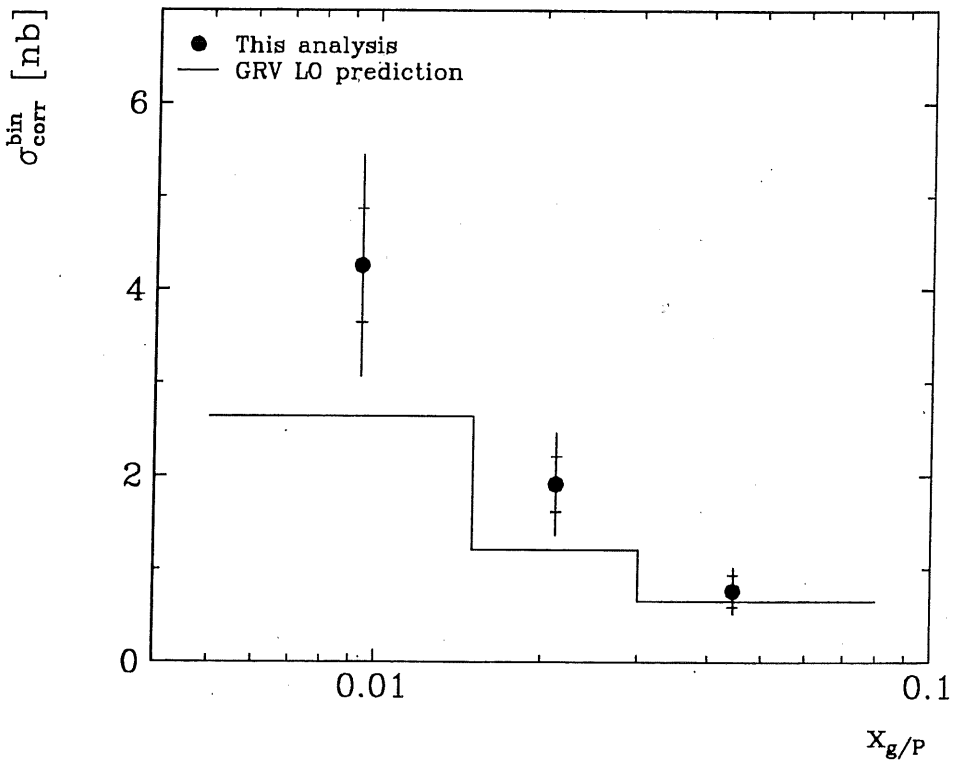


Figure 6: *The corrected PGF cross section as a function of the fractional gluon momentum.*

The final step of the analysis is the unfolding of the gluon density, $G(x_{g/p}) = x \cdot g(x)$, which is related to the corrected cross section in the following way:

$$G(x_{g/p}) = \frac{\sigma_{corr}^{PGF}}{\sigma_{MC}^{PGF}} \cdot G(x_{g/p})_{MC}$$

where σ_{MC}^{PGF} is the Monte Carlo (PROJET/DISJET) calculated PGF cross section and $G(x_{g/p})_{MC}$ is the parametrisation of the density function used in the Monte Carlo program.

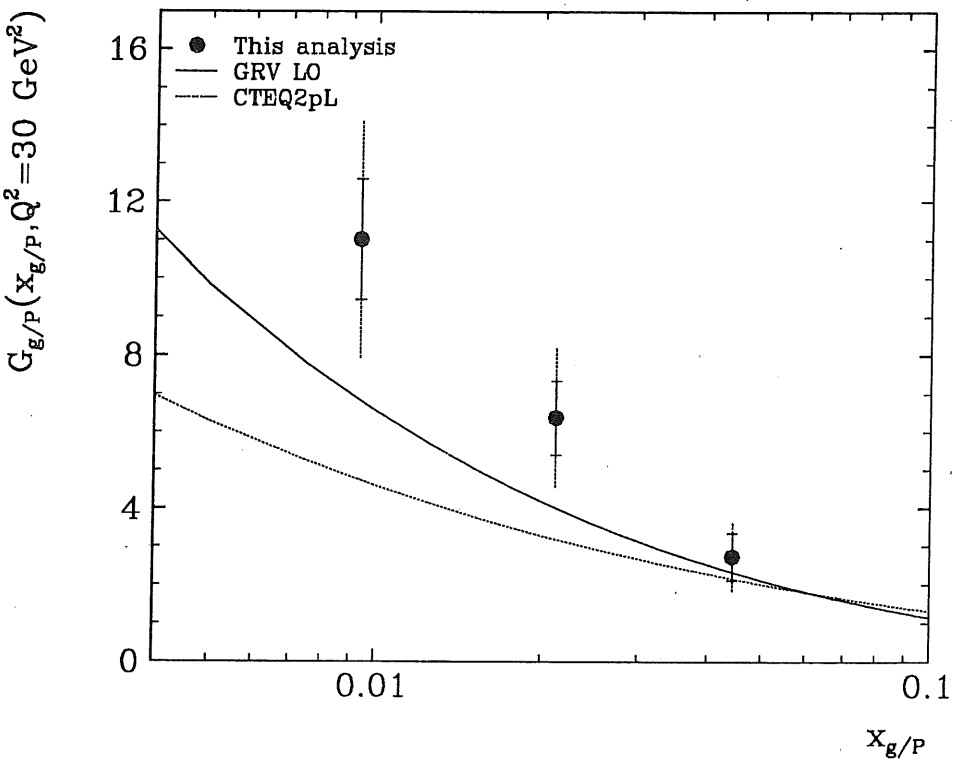


Figure 7: *The gluon density as a function of the fractional gluon momentum.*

8. Discussion of the Results

Our results on the gluon density as a function of the fractional gluon momentum is presented in Fig. 7. A considerable rise with decreasing fractional momentum of the gluon is observed. Such a behaviour is expected from a leading log(1/x) summation of soft gluons [10] (the Lipatov effect) but can also obtained from the more conventional evolution in Q^2 [11]. The various parametrizations of the gluon distribution are all phenomenological extrapolations based on some known distribution at a fixed Q^2 value, giving considerable uncertainties in the low $x_{g/p}$ region.

The data points are compared to two different parametrizations of the gluon density in LO. The GRV model [12] assumes the gluons and sea quarks to be valence like at $Q^2 = 0.3$ GeV^2 and the growth of the gluon density with decreasing $x_{g/p}$ values is due to the radiation of low x partons generated according to the Altarelli-Parisi equation [11]. The CTEQ parametrizations [13] are based on an input distribution function at $Q^2 = 4$ GeV^2 and assumes an $x^{-\lambda}$ dependence of the gluon density. In CTEQ the $\bar{u}, \bar{d}$ and s sea quark distributions are freely and independently parametrized whereas in other parametrizations it is assumed that the sea quark distribution is driven by the gluons. This difference should, however, not influence the gluon density distributions.

The error bars in the data points reflect the statistical and the total errors, the latter obtained by adding the statistical and systematic errors in quadrature. At present we are

source	contribution [%]
hadronic energy scale	10
variation of E_T cut	7
alternative MC model:	
LEPTO ME	15
variation of pdf's:	
MRS D0, D-, H	5
alt. jet algorithm:	
JADE	6
luminosity	5
detector efficiencies	6
rad. corrections	5
α_s	5
total systematic error:	24

Table 1: Sources of systematic errors and their contribution to the uncertainty of the gluon density measurement

dominated by systematic errors where for the highest $x_{g/p}$ bin the largest contribution comes from the uncertainty in the determination of the hadronic energy scale. The dominating error for the lowest $x_{g/p}$ bin is given by the variation in the unfolding procedure using the pure matrix element option of LEPTO to calculate acceptances.

The data points lie systematically above both parametrization curves. We have investigated whether the 8% contribution of diffractive events to our data sample changes the behaviour of the density function but this is not the case. In spite of the large errors in the data it seems that the CTEQ2'L leading order parametrization does not describe the data which would imply a larger gluon content in the proton at small $x_{g/p}$ than expected from an extrapolation of a simple parametrisation of high x data assuming a Lipatov behaviour. A decisive test of the GRV parametrization require a better understanding of the systematics both on the experimental level and in theory.

In the unfolding procedure we introduced cross-sections calculated in LO. At any fixed order in perturbation theory the dependence on the residual renormalisation scale, μ_r, acts as an estimateor of the theoretical uncertainty associated with the truncated perturbation series [14]. We implemented an *ad hoc* variation by multiplying, μ_r, with a factor ρ^2, where ρ^2 was set to $\frac{1}{4}$ and 16. This simple procedure gives cross-section variations of the order of 30%.

9. Systematic Errors

Table 1 gives an overview over the estimated averaged systematic uncertainties of the gluon density measurement from various sources.

The absolute hadronic energy scale is determined from studies of the transverse momentum balance between the scattered electron and the hadronic jet giving presently a precision of 5% [16], which translates into a 10% variation of the gluon density measurement.

A variation of the cut in the transverse jet energy between 3 GeV, below which no clear jet structures are observed, and 4 GeV, above which our statistics is poor, gives a contribution to the systematic error of the gluon density which amounts to 7%.

We estimated the uncertainty related to the choice of a specific Monte Carlo model (LEPTO MEPS) by repeating the analysis using the LEPTO ME model (first order matrix element followed by JETSET fragmentation, i.e. no parton showers). This model predicts considerably more narrow jet structures and significantly less energy in the detector around the proton direction and is not able to give a good descrition of data. This sort of extreme model choice yielded a 15 % change in the unfolded gluon density.

Running the LEPTO model with different parton density parametrizations MRS $D^{0'}$, MRS $D^{-'}$, and MRS H [15] led to 5 % variations consistent with the statistical accuracy of the generated Monte Carlo samples.

In order to be convinced that the final result is not critically affected by the choice of jet algorithm we repeated the analysis using the JADE algorithm [17] in the laboratory frame with a fixed mass cut of 10 GeV, the result of which gave 6% discrepancy in the measured gluon density. We attribute the discrepancy to the fact that the Monte Carlo is not able to give a perfect description of all aspects of data and take the observed variations as a measure of the systematic uncertainty related to this fact The reason why we used the cone algorithm to extract the final results was the better $x_{g/p}$ resolution and the more suppressed tails.

The radiative correction has been estimated on the hadron level for each $x_{g/p}$ bin by calculating the cross section with and without including QED radiation into the DJANGO Monte Carlo program [18]. The resulting corrections for all three bins are on the percent level with an estimated uncertainty of 5%.

A variation of the lower θ_{jet} cut of $\pm 2^{o}$ had no influence on the result.

The 5% uncertainty in the luminosity measurement [19] directly propagates into the gluon density measurement as does the 6 % unceratinty on various detector efficiencies mentioned in section 4. Finally, the uncertainty in the strong coupling constant α_s [20] contributes a 5 % uncertainty.

10. Conclusions

We have made a direct measurement of the gluon density in the proton from a determination of the cross section for the gluon induced photon gluon fusion process. Using the H1 detector at the *ep* collider HERA, data in a new kinematic domain has been obtained

in which a new QCD dynamics might appear. Our data is consistent with a steep rise of the gluon density as the fractional gluon momentum, $x_{g/p}$, decreases.

A comparison has been done between our data and two existing parametrizations of the gluon density in leading order. The CTEQ2'L model seems to be ruled out by our data points whereas a more precise conclusion about the GRV parametrization will require more statistics and above all a better understanding of the systematic errors.

Acknowledgements: We have profited from many constructive discussions with G. Ingelman and members of the Lund theory group. Our colleagues J. Gayler and G. Grindhammer in the H1 group have contributed valuable help and advice.

References

[1] H1 Collaboration, I. Abt et al., DESY preprint 93-103.

[2] H1 Calorimeter Group, B. Andrieu et al., Nucl.Instrum.Meth. A336 (1993) 499.

[3] F.Abe et al., Phys. Rev. D45 (1992) 1448.
M.Seymour, LU TP 93-7, May 1993.

[4] G. Ingelman, LEPTO version 6.1, in Proceedings 'Physics at HERA', Eds W. Buchmüller, G. Ingelman, DESY Hamburg 1992, vol. 3 and references therein.

[5] V. Hedberg, G. Ingelman, C. Jacobsson and L. Jönsson, Z. Phys. C63 (1994) 49

[6] H1 Collaboration, I. Abt et al., DESY preprint 94-033.

[7] H1 Collaboration, I. Abt et al., Nucl. Phys. B407 (1993) 515.

[8] D. Graudenz, PROJET program manual, unpublished and J. Rathsman, private communication.

[9] T. Brodkorb and E. Mirkes, DISJET program manual, MAD/PH/821, 1994, and J. Rathsman, private communication.

[10] E.A. Kuraev, L.N. Lipatov and V.S. Fadin, Sov. Phys. JETP 45 (1972) 199. Y.Y. Balitsky and L.N. Lipatov, Sov. J. Nucl. Phys. 28 (1978) 282.

[11] G. Altarelli and G. Parisi, Nucl. Phys. 126 (1977) 297.

[12] M. Glück, E. Reya, A. Vogt, Phys. Rev. D46 (1992) 1973.

[13] J. Botts et al., Phys. Lett. B304 (1993) 159.

[14] S.D. Ellis, Z. Kunszt, D.E. Phys. Rev. Lett. 69 (1992) 3615.

[15] A.D. Martin, W.J. Stirling, R.G. Roberts, Phys. Lett. B306 (1993) 145, *ibid.* B309 (1993) 492.

[16] H1 Collaboration, I. Abt et al., DESY preprint 94-033 (1944), submitted to Z. Phys. C

[17] JADE Collaboration, W. Bartel et al., Z. Phys. C33 (1986) 23.

[18] G.A. Schuler and H. Spiesberger, 'Physics at HERA', Eds W. Buchmüller, G. Ingelman, DESY Hamburg 1992, vol. 3 and references therein.

[19] H1 Collaboration, T. Ahmed et al., Phys. Lett. B299 (1993) 374.

[20] Particle Data Group, K. Hikasa et al., Phys. Rev. D45 (1992).

Appendix D

Jet Identification based on Probability Calculations using Bayes' Theorem

Jet Identification based on Probability Calculations using Bayes' Theorem

C. Jacobsson[a], L. Jönsson[a], G. Lindgren[b], M. Nyberg-Werther[a]

[a] Physics Department, Lund University, Sölvegatan 14, S-223 62 Lund, Sweden
[b] Department of Mathematical Statistics, Lund University, Box 118, 221 00 Lund, Sweden

Abstract

The problem of identifying jets at LEP and HERA has been studied. Identification using jet energies and fragmentation properties was treated separately in order to investigate what degree of quark-gluon separation could be achieved from either of these approaches. In the case of the fragmentation based identification, a test of the dependence on the jet production process and the fragmentation model was done. Instead of working with the separation variables directly, these have been used to calculate probabilities for having a specific type of jet according to Bayes' theorem. This offers a direct interpretation of the performance of the jet identification and provides simple means of combining the results from the energy and fragmentation based identifications.

1 Introduction

In many tests of QCD based on processes producing jets it is of great importance to be able to identify whether a jet originates from a quark or a gluon. Different criteria like specific decay properties, prior knowledge of the short range dynamics of the process or differences in the topology of jets due to the hadronisation can be used in such an identification.

Especially heavy quarks can be identified from their decay properties by using adequate particles in the decay chain to tag the flavour of the heavy quark. Fast leptons from semileptonic decays have been used as well as charged kaons and D-mesons. Further the long decay time of weak decays offers the possibility of reconstructing the secondary decay vertex by using high resolution vertex detectors.

The short range dynamics defines the kinematic properties of the process. For example, the fact that gluons are produced from primary quarks in a bremsstrahlung like process implies that the gluon jets usually are less energetic than the quark jets in an event.

The topology of jets are due to features of the partons which are related to their intrinsic properties, like mass and colour charge. Such differences influence the way the partons fragment into final state hadrons forming jets.

In this analysis we have studied jet separation from a general aspect and therefore concentrated on differences in the jet energies and in those properties of jets that are related to the fragmentation process. The optimal cut in jet energy for a separation between quarks and gluons obviously depends on how much energy is available for a certain process and how many jets are produced in that process. In e^+e^- collisions the energy available for jet production is well defined while for ep- and pp-collisions the energy involved in the hard scattering subprocess is varying from event to event. In order to describe the shape of jets a large number of fragmentation variables are available. In principle the fragmentation of a parton should not depend on the way it has been produced if we restrict ourselves to consider the jet core which makes a possible influence of the colour strings less important. We have thus made an attempt to find a process independent method to identify quarks and gluons by using suitable fragmentation variables alone.

The neural network method has previously been used with various input variables for the purpose of separating gluon jets from quark jets [1]. In this analysis we have been inspired by the strategy of [2] to distinguish between identification based on jet energies and identification based on fragmentation properties, using a neural network in the latter case. In a final step the two methods have been combined in order to improve the result by using all the available information. This can easily be done if one instead of working with the various separation variables directly, converts these into probabilities for a jet to be a quark or a gluon and applies cuts in the combined probabilities.

Most previous attempts to perform jet identification have been based on studies of individual jets. Another approach is to isolate the specific event type of interest and use the additional information contained in the knowledge of the exact number of quarks (q), antiquarks ($\bar{q}$) and gluons (g) for that event. Experimentally this method is possible for 3-jet events in e^+e^- collisions which must be of the type $q\bar{q}g$ whereas it is a good approximation to assume 4-jet

events to consist of a $q\bar{q}gg$ configuration since the $q\bar{q}q\bar{q}$ contribution is strongly suppressed. In the case of *ep*-collisions (2+1) jet events denote events with two jets in the hard scattering system in addition to the jet from the spectator quarks. The spectator jet is of no direct interest in the study of the hard subsystem motivating why we in the following will treat such events as 2-jet events. The final state of these events is either of qg type (the QCD-Compton process) or $q\bar{q}$ type (the Boson-Gluon fusion process), although in certain kinematic regions the $q\bar{q}$-events can be neglected and we are left with a clean sample of qg-events. For higher jet multiplicities the situation becomes less clear and an event based identification can not be easily applied. In this study we have thus concentrated on 3-jet events generated at the LEP centre of mass energy, 91.2 GeV, and (2+1) jet events of the qg-type from simulated collisions between 820 GeV protons and 26.7 GeV electrons at HERA. The Feynman diagrams for the processes investigated here are shown in Fig. 1.

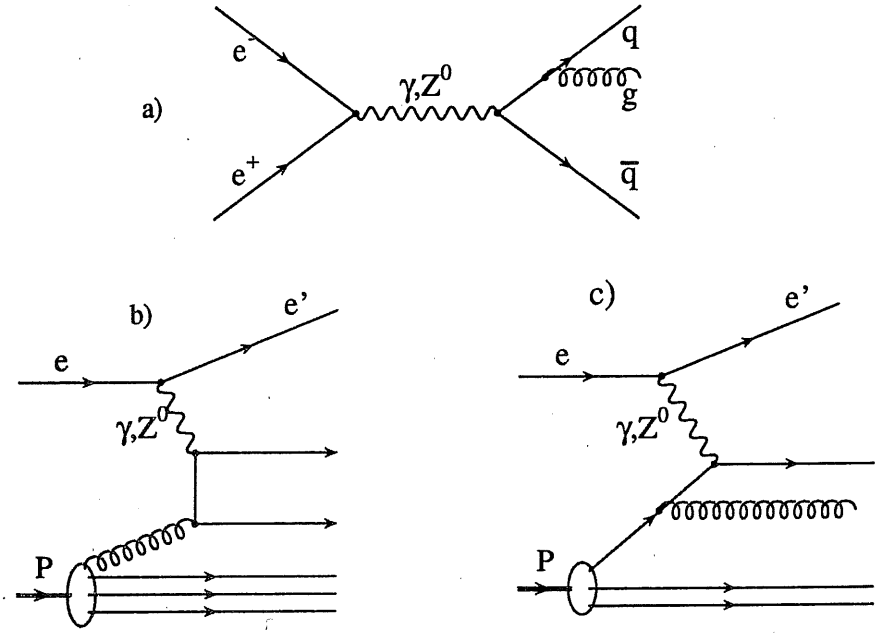


Figure 1: *The Feynman diagrams for a) a 3-jet event from an e^+e^--collision b) a BGF event from an ep-collision and c) a QCD-Compton event from an ep-collision.*

2 Event generation

In order to investigate whether our jet identification based on fragmentation properties is process independent we have applied our method to both Monte Carlo generated e^+e^--events and *ep*-events. The generation of *ep*-events has been done with the Monte Carlo (MC) programs LEPTO and HERWIG which are known to reproduce Deep Inelastic Scattering (DIS) data from previous fixed target lepton-nucleon scattering experiments and also to give a fair description of the limited data on jet physics presently available at HERA. The Monte Carlo program JETSET has proven to give a good description of various e^+e^- data and was thus used to produce such events. Since the Monte Carlo generators we have used cover jet production in e^+e^- as well as *ep* collisions and in addition use two different fragmentation schemes

(see below), we are able to test both the process and model dependence. This is done by training the network with samples from either of the generators and compare the results when it is applied to a test sample from the same generator and from one of the other generators respectively, according to the following procedure.

A comparison of the results from the network when trained on event samples from HERWIG and LEPTO respectively and subsequently tested on an event samples generated by LEPTO, will provide the model dependence. On the other hand if the respective event samples from JETSET and LEPTO are used to train the network which is then tested on a sample from JETSET the process dependence will come out. Finally if the network trained on the samples from HERWIG and JETSET respectively is applied to the test sample from JETSET we obtain both the model and the process dependence.

The basic concept for these generators is that hard scattering processes can be factorized into an elementary hard process, initial and final state radiation, and a hadronization process. This general scheme can be used to describe a large variaty of QCD and electroweak processes by applying different elementary subprocess matrix elements.

The JETSET program [3] describes e^+e^- annihilation into hadronic final states using two alternative approaches. One is the calculation of explicit matrix elements (ME) up to the second order in α_s and the other is based on parton shower (PS) emission which allows the production of an arbitrary number of jets. At present there is no scheme to connect the matrix element description with subsequent parton showers. In this analysis we have used the parton shower option based on the coherent evolution scheme by Marchesini and Webber since it seems to better reproduce the multi jet rates from experimental data. A parton shower is based on the branchings $q \to qg, g \to gg$ and $g \to q\bar{q}$ as given by Altarelli-Parisi evolution equations in the leading logarithm approximation of perturbative QCD. The evolution is performed in an iterative manner with decreasing virtuality of the partons until all parton masses have evolved below some minimum mass. This leads to an ordering in angle in the sense that angles between two emitted partons decrease with consequtive branching. The hadronisation is performed according to the Lund string model [4] [3].

LEPTO [5] simulates the basic DIS neutral and charged current processes and we have chosen to only consider the dominant neutral current process where a reconstruction of the event kinematics from the scattered electron can be made. In addition to the leading order Quark Parton Model (QPM) process, $\gamma^* q \to q$, where no jet identification is needed and therefore is of no interest for this analysis, also first order (α_s) processes, i.e. the QCD-Compton process, $\gamma^* q \to qg$, and the boson-gluon fusion process, $\gamma^* g \to q\bar{q}$, are calculated from QCD matrix elements. In order to avoid divergencies from soft and collinear parton emission a cut-off in the invariant mass of any two partons, m_{ij}, is implemented. Higher order corrections are then included by adding parton showers according to the same scheme, based on the Altarelli-Parisi evolution equation, as in JETSET. The amount of initial and final state radiation is determined by the virtual mass of the initiating parton just before and after the boson vertex. The initial state radiation is performed by a backward evolution scheme from the hard vertex which is controlled by the parton density function specified to the program. We have used the MRS H parametrisation of the density function as it describes recent results on the proton structure function F_2 at low Bjorken-x from HERA. The Lund string fragmentation is used to produce the hadronic final state.

Similar to the LEPTO program, the simulation of *ep*-collisions by the HERWIG generator [6] is done using matrix elements to describe the processes up to the first order in the strong coupling constant and higher order emissions are introduced by parton showers which are generated with essentially Q^2 as mass scale. The upper limit for the shower evolution variables are related to energies and angles rather than to parton virtualities. The backward evolution process produces coherent initial state parton showers with full QCD cascading of all emitted partons. The same parton density function was used as for LEPTO. The final state coherent showers include soft gluon interference and azimuthal correlations due to the gluon spin. The emitted gluons are split into quark antiquark pairs (or eventually into diquark antidiquark pairs) between which there are colour lines forming colour-singlet clusters. The clusters created in this way are fragmented into hadrons through a longitudinal splitting of the high mass clusters.

As already mentioned this study has been limited to simulated 3-jet events from e^+e^--interactions at LEP and (2+1) jet events from generated *ep*-collisions at HERA. No detector simulation has been made but for the analysis of HERA events the usual beam pipe cut was introduced, excluding the regions in polar angle below 4° and above 176° not covered by the detector. Due to the event topologies of deep inelastic scattering processes most of the spectator jet will disappear undetected down the forward cone while for a majority of events the scattered electron will proceed inside the backward cone. No such cut is necessary at LEP since no specific activity is expected in these regions and since the jet analysis in any case is limited to the barrel region of the detector.

3 Jet reconstruction

In order to reconstruct particle jets the LUCLUS algorithm [7], based on the combination of energy clusters, was used. A careful study of the reconstruction quality as a function of the resolution parameter in the algorithm showed that a value $d_{join} = 4$ GeV was relevant (see [8]). In the HERA analysis the clustering was done in the so-called hadronic center-of-mass system i.e. in the center-of-mass system of the incoming proton and the exchanged virtual photon. In order to reconstruct the spectator jet in the best possible way, a pseudoparticle is added to each event to represent the fraction of the proton fragment lost in the beampipe. The momentum of this pseudoparticle is given by the difference of the longitudinal momentum of the initial state and the measured longitudinal momentum of the final state, as described in [9].

In a Monte Carlo generator including parton showers, many partons can contribute to a jet and one needs a method to establish if the reconstructed jet should be regarded as originating from a quark or a gluon. For *ep*-events, where the ME forms the basis of the processes and PS are added to simulate higher order corrections, we simply checked which reconstructed jet was closest to the original parton from the ME, according to: $\min(|\bar{P}_{jet1} - \bar{P}_q| + |\bar{P}_{jet2} - \bar{P}_g|, |\bar{P}_{jet2} - \bar{P}_q| + |\bar{P}_{jet1} - \bar{P}_g|)$. Since the e^+e^--events are generated with PS alone we have to extract the momentum vectors of the jets for both the parton level and the hadron level by applying the LUCLUS jet algorithm. A comparison of the momentum vectors in pairs on the parton and hadron level, identified which jet on the hadron level corresponds to a certain jet on the parton level. If the jet on the parton level contains an odd number of quarks it is

considered a quark initiated jet while if the number of quarks in the jet is even it is defined as a gluon jet.

4 Event selection

4.1 Selection of HERA events

Since we want to concentrate on the (2+1) jet events we have only considered events in which the jet-algorithm found exactly two jets + the spectator jet. In the hadronic centre-of-mass system we required the minimum energy of each jet to be 5 GeV and the invariant mass of the two jets to be larger than 15 GeV to ensure that the selected events had a reasonably clear jet-structure. We also required a minimum number of four particles to be assigned to each jet, since the jet-variables which were used to study the fragmentation are not meaningful for jets with too few particles. In order for the two hard jets to be well inside the acceptance region of the HERA experiments and to have a separation in space from the proton remnant, both jets had to be reconstructed within the region of polar angles $10° < \theta < 160°$ as measured in the laboratory system. The two jets also had to be separated by less than two units of pseudorapidity. It has been previously shown [8] that this is necessary in order to cut down the background of q-type events which otherwise enters the (2+1) jet sample.

Finally, we only use events produced within certain limits of the kinematic variables generally used to describe DIS events. These are Q^2, the momentum transfer squared, the Bjorken-x and -y scaling variables and W^2, the invariant mass squared of the hadronic system:

$$Q^2 \equiv -q^2 = -(p_e - p_l)^2, \quad x \equiv \frac{Q^2}{2P \cdot q}, \quad y \equiv \frac{P \cdot q}{P \cdot p_e}, \quad W^2 \equiv (q+P)^2 = Q^2 \frac{1-x}{x} + m_p^2$$

(For a description of these variables and how they can be measured at HERA see for example [10]).

The cross-section falls rapidly with increasing x and Q^2, for both the qg- and the $q\bar{q}$-type of events, but qg-events dominate in the region of high-x-values. At $x \geq 0.1$ an almost pure sample of qg-events is produced, $\frac{qg}{qq} > 8$, and one can therefore concentrate on separating quark jets from gluon jets in this region. This sample was used in our attempts to identify gluon jets at HERA. At lower values of x ($x < 0.1$) a mixture of qg- and $q\bar{q}$- events is produced and one has to deal with the problem of separating the two event types.

4.2 Selection of LEP events

In the selection of Monte Carlo generated 3-jet events from e^+e^--collisions at LEP, we required each jet to have an energy of more than 5 GeV in order to have reasonably collimated flows of particles. Exactly as for the jets from ep-collisions the invariant mass of any jet pair, m_{ij}, should exceed 15 GeV to give an observable 3-jet topology. Also in analogy with the treatment of ep-collision events we required each jet in an e^+e^--event to contain at least 4 particles since the same fragmentation variables are going to be used in both cases. The energy sum of all

three jets in an event was required to be greater than or equal to 90 GeV in order to assure that some fraction of a jet was not escaping detection. We have assumed a LEP detector with full azimuthal coverage but restricted the jets to fall inside the range $40° < \theta < 140°$ of polar angle. This is the barrel region which is normally well covered by both the tracking system and the calorimetry of a detector.

5 Identification of gluon jets using jet-energy

Based on the assumption that gluon jets carry less energy than quark jets and using the knowledge about the number of quarks and gluons in the event type under investigation, we wish to calculate the probability for a jet to originate from a gluon. In doing this we recall that there is a difference between e^+e^--collisions and ep-collisions in the sense that the energy sum of quarks and gluons is constant and equal to $\sqrt{s}$ for e^+e^--processes while the energy entering the hard scattering subprocess in ep-collisions is varying from event to event. An identification which includes jet energies will therefore always be process dependence.

The probability of emitting a gluon with a certain energy in an e^+e^--collision is directly obtained from first order ME calculations. The conditional probability of jet1 in a 3-jet event to originate from a gluon, provided the jets have the energies E_1, E_2 and E_3, is given by:

$$P_{gqq}^{E_{123}}(jet1, jet2, jet3|E_1, E_2, E_3) = \frac{E_2^2 + E_3^2}{(E_{cm} - 2E_2)(E_{cm} - 2E_3)} \cdot \tag{1}$$
$$\left(\frac{E_1^2 + E_2^2}{(E_{cm} - 2E_1)(E_{cm} - 2E_2)} + \frac{E_2^2 + E_3^2}{(E_{cm} - 2E_2)(E_{cm} - 2E_3)} + \frac{E_3^2 + E_1^2}{(E_{cm} - 2E_3)(E_{cm} - 2E_1)}\right)^{-1}$$

where E_{cm} is the centre-of-mass energy of the e^+e^--collision.

The probability of the scattered quark to radiate a gluon of a certain energy in an ep-interaction is not so easily accessible from the matrix element and we have therefore used Monte Carlo generated energy distributions to extract the density functions for the gluon and the quark, f_g and f_q, in a qg-event.

$$f_g^E = \frac{f_{gq}^E(E_g, E_q)}{f_{g+q}^E(E_g + E_q)}$$

$$f_q^E = \frac{f_{gq}^E(E_q, E_g)}{f_{g+q}^E(E_g + E_q)}$$

where f_{gq}^E is the joint density function given the gluon and the quark have the energies E_g and E_q respectively. The total energy of the quark gluon system, $E = E_g + E_q$, is a random variable with the density $f_{g+q}^E(E_g + E_q)$. Since we know that the *a priori* probability for having a quark or a gluon is equal, then the probability for jet1 to be a gluon jet is given by Bayes' theorem according to:

$$P_{gq}^{E_{12}}(jet1, jet2|E_1, E_2) = \frac{0.5 f_g^E}{0.5 f_g^E + 0.5 f_q^E} = \frac{f_{gq}^E(E_1, E_2)}{f_{gq}^E(E_1, E_2) + f_{gq}^E(E_2, E_1)} \tag{2}$$

with E_1 and E_2 giving the energies of jet1 and 2 respectively.

In each event the jet with the highest probability, $P_{g,max}$, is selected to originate from a gluon. The allowed range of gluon probabilities is for 2-jet events $\frac{1}{2} < P_{g,max} < 1$ and for 3-jet events $\frac{1}{3} < P_{g,max} < 1$, where the lower bounds correspond to equal probabilities for all jets in the event to be a gluon. A general definition of the allowed range would thus be $\frac{1}{\#jets} < P_{g,max} < 1$, giving the limits within which a cut (P_{cut}) can be specified in order to enhance the purity of gluon jets in our selected sample. For processes containing more than one gluon the procedure can in principle be repeated to find a second gluon in the remaining sample and so on. The probability bounds for the second gluon will then be $\frac{1}{\#jets-1} < P_{g,max} < 1$ and for the i:th gluon candidate $\frac{1}{\#jets+1-i} < P_{g,max} < 1$. From the generated MC data we can now check whether the jet with the value $P_{g,max}$ and thereby identified as a gluon jet, actually was initiated by a gluon or a quark. If we, as an example, for 3-jet events from e^+e^- collisions plot the frequency of the jet to originate from a quark and a gluon respectively as a function of $P_{g,max}$, we get the distributions shown in Fig. 2. The figure should be interpreted in the following way. If the jet with the highest probability

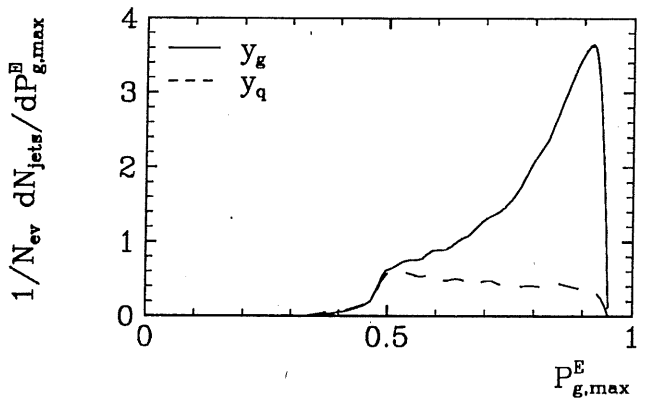


Figure 2: *The $P^E_{g,max}$ distributions for quarks and gluons in a 3 jet event.*

of being a gluon jet, in an event, has the probability value $P_{g,max}$, then the probability of it to really originate from a gluon is given by the value of the gluon distribution (y_g) at $P_{g,max}$ divided by the summed values of the gluon and quark distributions ($y_g + y_q$) also at $P_{g,max}$. It is then clear that the efficiency and purity for identifying the gluon jet, and thereby also the quark jet(s), in events containing only one gluon, can be obtained by using the information from all jets in the event in the following way:

$$Efficiency = \frac{\int_{P_{cut}}^{1} (y_g + y_q)\, dP_{g,max}}{\int_{\frac{1}{\#jets}}^{1} (y_g + y_q)\, dP_{g,max}} \tag{3}$$

$$Purity = \frac{\int_{P_{cut}}^{1} y_g\, dP_{g,max}}{\int_{P_{cut}}^{1} (y_g + y_q)\, dP_{g,max}} \tag{4}$$

where $P_{cut} \geq \frac{1}{\#jets}$.

6 Identification of gluon jets using fragmentation properties

Due to the fact that gluons according to QCD carry a stronger colour charge than quarks it is expected that there will be differences in their fragmentation. A large number of variables sensitive to these differences have been suggested for the purpose of performing quark-gluon separation. Above we have derived the probability formalism for a separation using the jet energies alone and here we will go through the same procedure for an identification from fragmentation variables, using a neural network.

One of the reason for using a neural network is that we want to simultaneously take into accout the effect of several variables and their correlations. We have chosen a network, as implemented in the program package JETNET 2.0 [11], using the method of backpropagation, well suited for this kind of pattern recognition. We have varied the number of hidden layers, nodes and values of the learning rate but this did not cause the final results to change significantly. We therefore decided to use one hidden layer and one output node ($0 = quark; 1 = gluon$)

To enable a process independent identification it is essential that the result does not depend on the jet energy, and one could therefore either try to select variables which are completely uncorrelated with the jet energy or train the neural network in such a way that the jet energy by itself does not give any discrimination. In the latter case it is from a neural network point of view important to be careful in the use of energy dependent fragmentation variables. Although we want the network to be sensitive to the energy dependence of the fragmentation variables it should not be affected by the jet energies themselves. Since the difference in the jet energy distributions is the most dominant effect it might be picked up even implicitly by the neural network. Therefore a good training strategy will help to emphasize the learning on the more subtle fragmentation properties.

Among the variables we have investigated it turns out that those most sensitive to differences in the fragmentation properties all have a considerable energy dependence. We thus trained the neural network with equal and flat energy distributions for quarks and gluons in order to prevent the network to be influenced by the jet energies themselves. Such an artificial training sample is obtained by using individual jets taken from the Monte Carlo generated events. In the following we denote this method **balanced energy training.**

A jet algorithm has to be applied before the fragmentation sensitive variables can be calculated. To avoid a dependence of the fragmentation variables on the detailed reconstruction of a jet by different jet algorithms we only consider particles in the jet core. The jet core is defined by taking the particles of a jet in descending order of P_l, the longitudinal momentum with respect to the jet axis, until we reach 80% of the total jet energy. Since the jet algorithm is not Lorentz invariant, using only the jet core also leads to an insensitivity of the frame in which the clustering takes place, which is very important for the HERA events. We also want the fragmentation variables to be experimentally useful, i.e. they should not be heavily affected by detector smearing and poor event reconstruction.

Variables describing fragmentation properties are normally based on the relation between single particles in a jet and the jet axis. One set of variables we have tested and found to provide the best separation between quarks and gluons are the so called **Fodor moments**

[12]:

$$F_{nm}(E_{jet}) = \sum (\frac{P_T}{E_{jet}})^n \eta^m$$

where P_T and η are the transverse momentum and the pseudorapidity of a particle in a jet with respect to the jet-axis and E_{jet} is the jet energy. The sum is taken over all particles in the jet core. The three lowest moments have an obvious interpretation. F_{00} is the multiplicity of the jet, F_{01} is the pseudorapidity sum of all particles in the jet, and F_{10} is the transverse momentum sum of all particles, scaled by the jet energy. A careful study of the Fodor moments reveals that the results based on the different generators gives general agreement only for some of the moments. This is, however, a necessary condition in order to obtain an independence of both the jet production process and the fragmentation model used, and consequently we have concentrated on these moments. The mean values of the moments F_{11}, F_{15}, F_{20} and F_{31} exhibit similar behaviour as a function of energy for all the generators except in the low energy range of the moment F_{15} where the JETSET curves fall below the others. This is illustrated in Fig. 6a-d. A separation cut between quarks and gluons common for all the generators can thus in principle be found for the moments F_{11}, F_{20} and F_{31} over the full energy range whereas this is not true for the moment F_{15}. However, as can be observed from Fig. 6f the Fodor moment distributions for quarks and gluons are significantly overlapping which in any case prevents a completely clean separation. As an example of a moment where the curves from the various generators are widely spread and therefore makes the choice of a common separation cut difficult we show the moment F_{00} in Fig. 6e. In our selection we have avoided the higher Fodor moments since they will, from an experimental point of view, be very sensitive to the reconstruction quality of the energy and direction of the particles. Instead of using the notation $F_{nm,i}$ to specify the value of fragmentation variable F_{nm} for jet i, we will simplify the notation by letting F_i represent the values of all used fragmentation varibles for jet i.

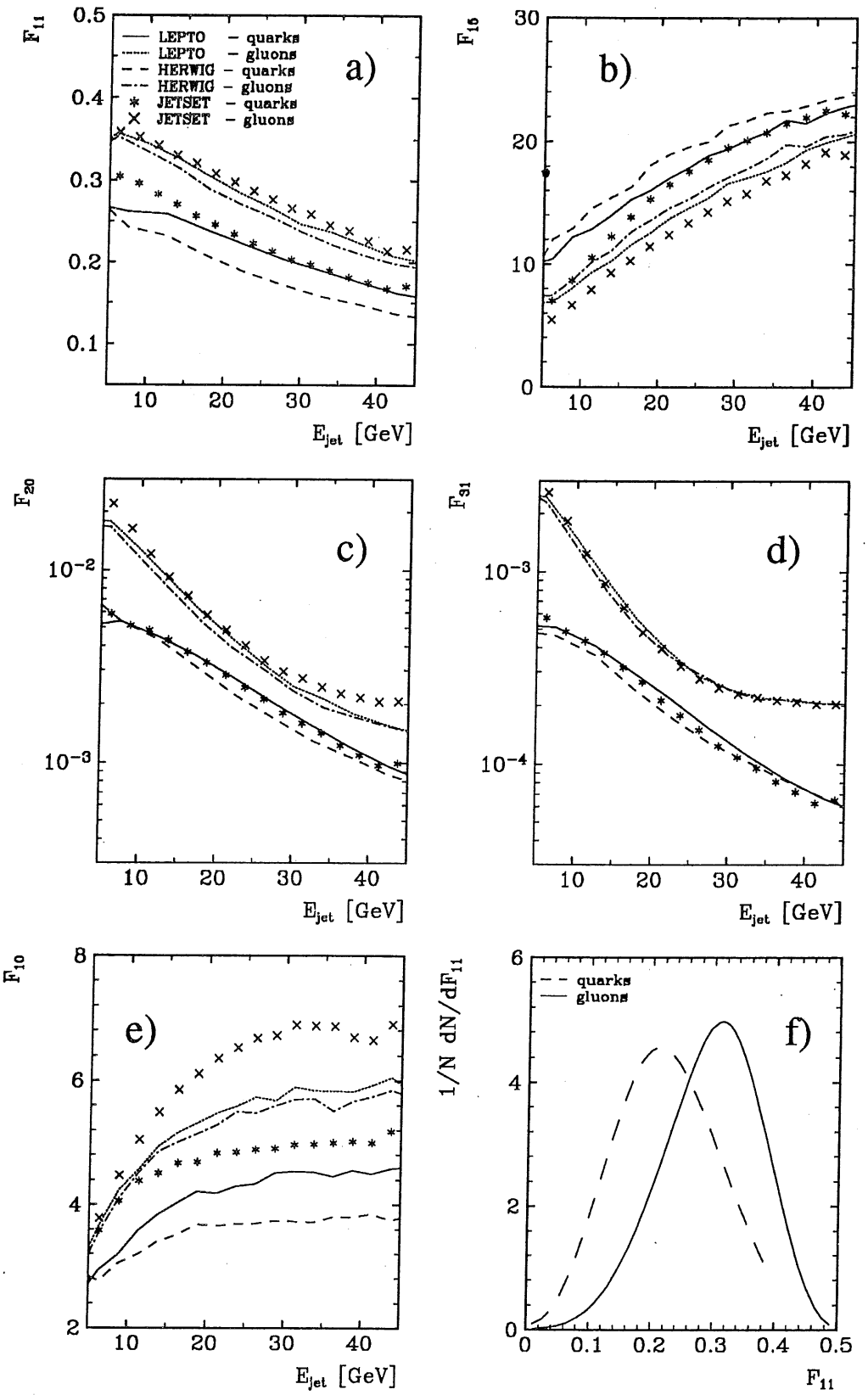


Figure 3: *The Fodor moments a) F_{11}, b) F_{15}, c) F_{20}, d) F_{31} and e) F_{00} as a function of the jet energy. In f) is shown the F_{11}-distributions for quarks and gluons in LEPTO at $E_{jet} \approx 25$ GeV*

The output of the neural network will simply be the conditional probability for a jet to be a gluon (or a quark depending on how the output is defined), given the input variables and the composition of the training sample [13]). Using Bayes' theorem the probability for jet1 to be a gluon jet can be expressed through the density functions, given the values of the fragmentation variables are F_1 at an energy $\tilde{E}_1$, in the following way:

$$P_g^F(jet1|F_1) = \frac{\tilde{f}_g^E(\tilde{E}_1)f_g^F(F_1)}{\tilde{f}_g^E(\tilde{E}_1)f_g^F(F_1) + \tilde{f}_q^E(\tilde{E}_1)f_q^F(F_1)} = \frac{f_g^F(F_1)}{f_g^F(F_1) + f_q^F(F_1)} \tag{5}$$

where $P_g^F(jet1|F_1)$ thus is identical to the network output. $\tilde{f}^E$ denotes the density function of the artificial energy distribution used in the balanced energy training which implies $\tilde{f}_g^E(E_1) = \tilde{f}_q^E(E_1)$. Since the jet energy ($E_1$) is given as input to the network together with the fragmentation variables (F_1), the correct notation of the density function should be $f^F(F_1|E_1)$. However, in the balanced energy sample the energy will only provide information on the energy dependence of the fragmentation variables, as already explained, which means that we are in reality only considering the fragmentation variables. In order not to confuse the reader we have therefore decided to use the simplified notation $f^F(F_1)$.

The probability (5) based on the fragmentation properties is also valid in the case where we instead of treating individual jets make use of the fact that we know the number of quark and gluon jets in the event. Thus it is not necessary to train the network specifically for this situation. For 2-jet events the output of a network trained on the quark and the gluon in a pair according to the balanced energy method, is just a simple function of the output from a network trained with individual jets, assuming the jets to fragment independently. Again, due to the balanced energy training, the combined density functions $\tilde{f}_{gq}^E(E_1, E_2) = \tilde{f}_{qg}^E(E_1, E_2)$. Since the fragmentation of the quark and the gluon is independent, given the energies E_1 and E_2, we consequently have $f_{gq}^F(F_1, F_2|E_1, E_2) = f_g^F(F_1|E_1)f_q^F(F_2|E_2) = f_g^F(F_1)f_q^F(F_2)$ using our simplified notation.

$$P_{gq}^{F_{12}}(jet1, jet2|F_1, F_2) = \frac{f_g^F(F_1)f_q^F(F_2)}{f_g^F(F_1)f_q^F(F_2) + f_g^F(F_2)f_q^F(F_1)} \tag{6}$$

Dividing the nominator and the denominator by $[f_g^F(F_1)+f_q^F(F_1)][f_g^F(F_2)+f_q^F(F_2)]$ and using the fact that for individual jets the quark and gluon probabilities are related as $P_q^F(jet1|F_1) = 1 - P_g^F(jet1|F_1)$ we get:

$$P_{gq}^{F_{12}}(jet1, jet2|F_1, F_2) = \frac{P_g^F(jet1|F_1)[1 - P_g^F(jet2|F_2)]}{P_g^F(jet1|F_1)[1 - P_g^F(jet2|F_2)] + [1 - P_g^F(jet1|F_1)]P_g^F(jet2|F_2)}$$

where $P_{gq}^{F_{12}}(jet1, jet2|F_1, F_2)$ is now expressed in single jet probabilities which are identical to the network output values. The 3-jet events from e^+e^--collisions contain a quark and an anti-quark which are identical from a fragmentation point of view and the probability for a gluon jet is obtained by a simple extension of the expression for the 2-jet events:

$$P_{gqq}^{F_{123}}(jet1, jet2, jet3|F_1, F_2, F_3) = \frac{f_g^{F_1} f_q^{F_2} f_q^{F_3}}{f_g^{F_1} f_q^{F_2} f_q^{F_3} + f_g^{F_2} f_q^{F_1} f_q^{F_3} + f_g^{F_3} f_q^{F_1} f_q^{F_2}}$$

with $f_g^{F_i} = f_g^F(F_i)$, i=1,2,3. The probability for a gluon jet in a 3-jet event can be expressed in terms of individual jet probabilites using a similar procedure as for the 2-jet case.

$$P_{gqq}^{F_{123}}(jet1, jet2, jet3|F_1, F_2, F_3) =$$

$$\frac{P_g^{F_1}(1-P_g^{F_2})(1-P_g^{F_3})}{P_g^{F_1}(1-P_g^{F_2})(1-P_g^{F_3})+P_g^{F_2}(1-P_g^{F_1})(1-P_g^{F_3})+P_g^{F_3}(1-P_g^{F_1})(1-P_g^{F_2})} \quad (7)$$

where $P_g^{F_i} = P_g^F(jeti|F_i)$, i=1,2,3.

7 Jet identification using jet-energy and fragmentation properties

We now want to extract the conditional probability for a jet being a gluon jet given both the jet energies as well as the values of the fragmentation variables which were used as input to our neural network. In the (2+1) jet case we obtain this by using the equations (2) and (6):

$$P_{gq}^{C_{12}}(jet1, jet2|E_1, E_2, F_1, F_2) = \frac{f_{gq}^{E_{12}} f_g^F(F_1) f_q^F(F_2)}{f_{gq}^{E_{12}} f_g^F(F_1) f_q^F(F_2) + f_{gq}^{E_{21}} f_g^F(F_2) f_q^F(F_1)}$$

where

$$f_{gq}^{E_{ij}} = \frac{f_{gq}^E(E_i, E_j)}{f_{gq}^E(E_i, E_j) + f_{gq}^E(E_j, E_i)}$$

If we now divide all the terms with $(f_{gq}^{E_{12}} + f_{gq}^{E_{21}})(f_g^{F_1} + f_g^{F_1})(f_q^{F_2} + f_q^{F_2})$ we obtain

$$P_{gq}^{C_{12}}(jet1, jet2|E_1, E_2, F_1, F_2) = \frac{P_{gq}^{E_{12}} P_g^{F_1} P_q^{F_2}}{P_{gq}^{E_{12}} P_g^{F_1} P_{qg}^{F_2} + P_q^{E_{12}} P_q^{F_1} P_g^{F_2}} = \frac{P_{gq}^{E_{12}} P_g^{F_1}(1-P_g^{F_2})}{P_{gq}^{E_{12}} P_g^{F_1}(1-P_g^{F_2}) + (1-P_{gq}^{E_{12}})(1-P_g^{F_1})P_g^{F_2}}$$

After division of all the terms with $P_g^{F_1}(1-P_g^{F_2}) + (1-P_g^{F_1})P_g^{F_2}$ we finally get

$$P_{gq}^{C_{12}}(jet1, jet2|E_1, E_2, F_1, F_2) = \frac{P_{gq}^{E_{12}} P_{gq}^{F_{12}}}{P_{gq}^{E_{12}} P_{gq}^{F_{12}} + P_{gq}^{E_{21}} P_{gq}^{F_{21}}} \quad (8)$$

where we have used $P_{gq}^{E_{12}} = 1 - P_{gq}^{E_{21}}$ and $P_{gq}^{F_{12}} = 1 - P_{gq}^{F_{21}}$.

In a completely analogue way the corresponding conditional probability can be obtained for a 3-jet event.

$$P_{gqq}^{C_{123}}(jet1, jet2, jet3|E_1, E_2, E_3, F_1, F_2, F_3) = \frac{P_{gqq}^{E_{123}} P_{gqq}^{F_{123}}}{P_{gqq}^{E_{123}} P_{gqq}^{F_{123}} + P_{gqq}^{E_{321}} P_{gqq}^{F_{321}} + P_{gqq}^{E_{231}} P_{gqq}^{F_{231}}} \quad (9)$$

8 Results

The final results presented in this section are based on the analysis of events i.e. the quark and gluon content is defined by the event type generated. This is in contrast to the artificial training samples for the neural network which were created out of a selection of individual jets from the generated events.

Fig. 4a shows the neural network output for individual jets in 3-jet events from e^+e^--collisions, based on the fragmentation variables discussed in section 6 and their energy dependence. The x-axis gives directly the probability for a jet from a quark and a gluon respectively to be identified as a gluon jet. Using the expression (7) we can calculate the event based conditional probability for a jet to originate from a gluon given the values F_1, F_2, F_3 for the fragmentation variables and the energies $\tilde{E}_1, \tilde{E}_2, \tilde{E}_3$, specified in order to account for the energy dependence of the fragmentation variables. From these calculations we select for each event the jet with the highest probability of being a gluon jet, giving a distribution as shown in Fig. 4b. We note that the distributions populate just the allowed region $1/3 < P_{g,max} < 1$.

Fig. 4c presents the probability for a jet being a gluon jet as obtained from the ME calculation according to 2 and Fig. 4d gives the distributions of jets with the highest conditional probability in the event of coming from a gluon. In agreement with what has been indicated previously in the text Figs. 4b and d confirm that the jet energies are much more efficient in identifying jets than are the fragmentation properties.

The event based combined conditional probability for gluon jet identification given the jet energies E_1, E_2, E_3 and the values F_1, F_2, F_3 for the fragmentation variables is given in Fig. 4e and finally the jet per event having the highest combined probability of being produced by a gluon is plotted in Fig. 4f. The corresponding plots for (2+1) jet events from *ep*-collisions are shown in Fig 5a-f.

From the $P_{g,max}$ distributions we can now calculate the purity and efficiency as a function of P_{cut} according to the expressions (3) and (4). In Fig. 6 the purity is plotted versus the efficiency separately for an identification based on the jet energies and on the fragmentation variables as well as for a combination of the two. Fig. 6a shows the results for 3-jet events generated with JETSET and for which the neural network has also been trained with a sample generated with JETSET. The corresponding results for (2+1) jet events are shown in Fig. 6b and c for samples generated with LEPTO and HERWIG respectively using networks trained with jets from the same generators. **It has to be stressed that the results presented here correspond to an identification of all jets in the events and they can therefore not be directly compared to results on individual jet identification.** From a comparison of Figs. 6a-c it is seen that the energy based identification of 3-jet events in LEP, with an *a priori* probability of 33%, and (2+1) jet events at HERA, with an *a priori* probability of 50%, are both 78-79% at 100% efficiency. As an increasingly harder P_{cut} is made in the $P_{g,max}$ distributions the identification improves faster for the *ep* events than for the e^+e^- events. The identification using fragmentation variables gives in general much worse separation between quarks and gluons but it is normally a complementary method in the sense that events which are well separated by the jet energies are not necessarily those which are well separated by the fragmentation properties. Consequently an improved result is obtained if the identifications from energy and fragmentation are combined. Only for the HERWIG sample (Fig. 6c) the fragmentation based identification is essentially equally good as the one based on jet energies and the combined result is also somewhat better than for the LEPTO sample (Fig. 6c). The better result in the HERWIG case compared to the LEPTO case must be due to the different fragmentation models used. Although the fragmentation seems to give significantly better results for the *ep* events than for the e^+e^- events over the full efficiency range this is not reflected in the combined results, the relative improvement with respect to the jet energy results being approximately the same for both types of events. This might indicate a higher

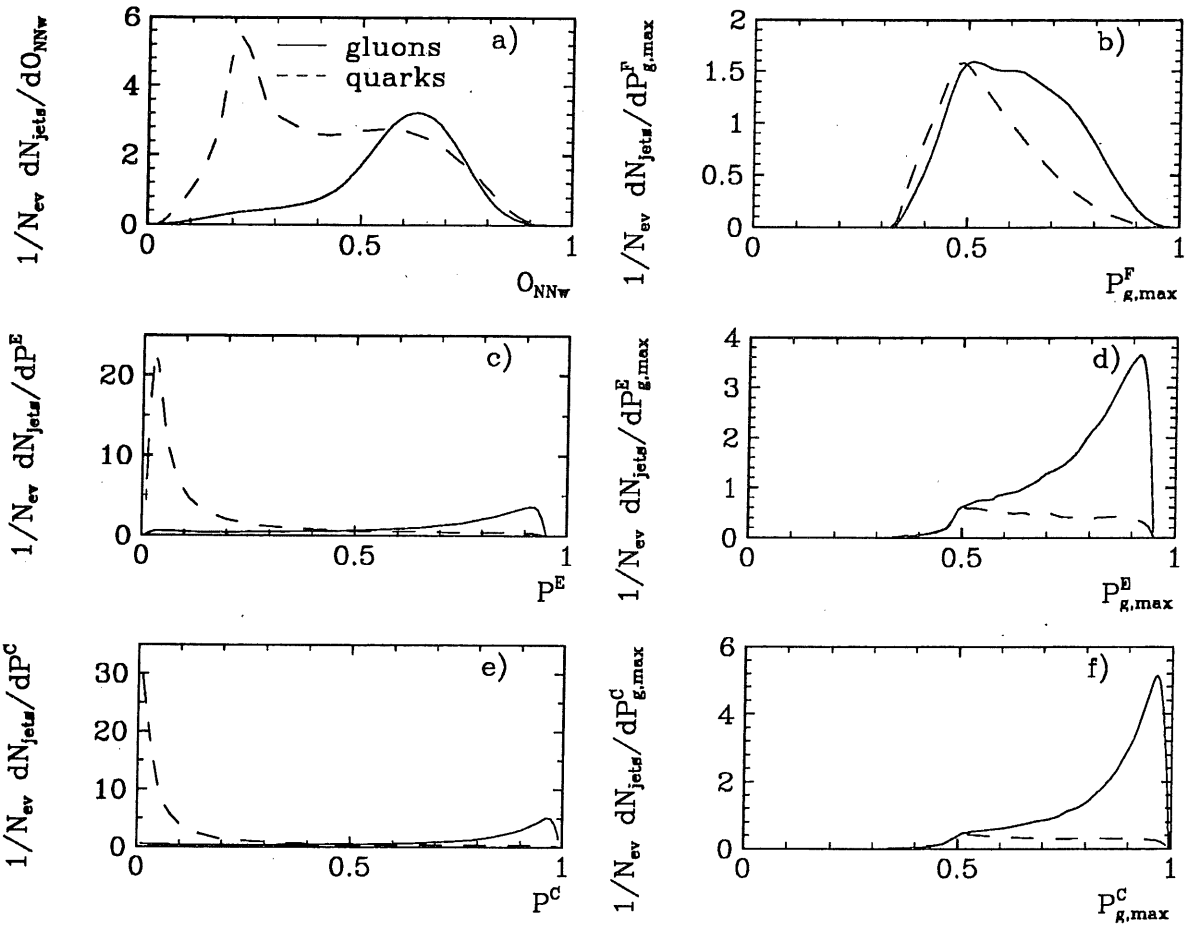


Figure 4: *Results on 3-jet events from e^+e^- shown as a) the neural network output for individual jets b) the $P^F_{g,max}$ distribution from the fragmentation c) the gluon probability obtained from the ME d) the $P^E_{g,max}$ distribution from the jet energies e) the combined conditional probability distribution according to equation (7) f) the $P^C_{g,max}$ distribution from the combined conditional probability.*

degree of complementarity for the e^+e^- events than for ep events.

In order to investigate the process and model dependence we also present results from samples using a generator different from the one used in the training of the neural network. These results are given in Figs. 6d-f. A comparison between Figs. 6a and d illustrates the process dependence while a comparison of the Figs. 6b and f provides the model dependence. From the Figs. 6a and e the effects of both the model and process dependence can be extracted.

According to Fig. 6a and d the fragmentation result is better when a test sample from JETSET is presented to a network trained on LEPTO compared to a network trained on JETSET. The explanation of this behaviour is that the energy dependence of some Fodor moments are different for LEPTO and JETSET, causing the network to implicitly pick up this energy dependence in spite of the balanced energy training. It seems thus difficult to obtain a process independent identification of the jets. We have previously observed that a

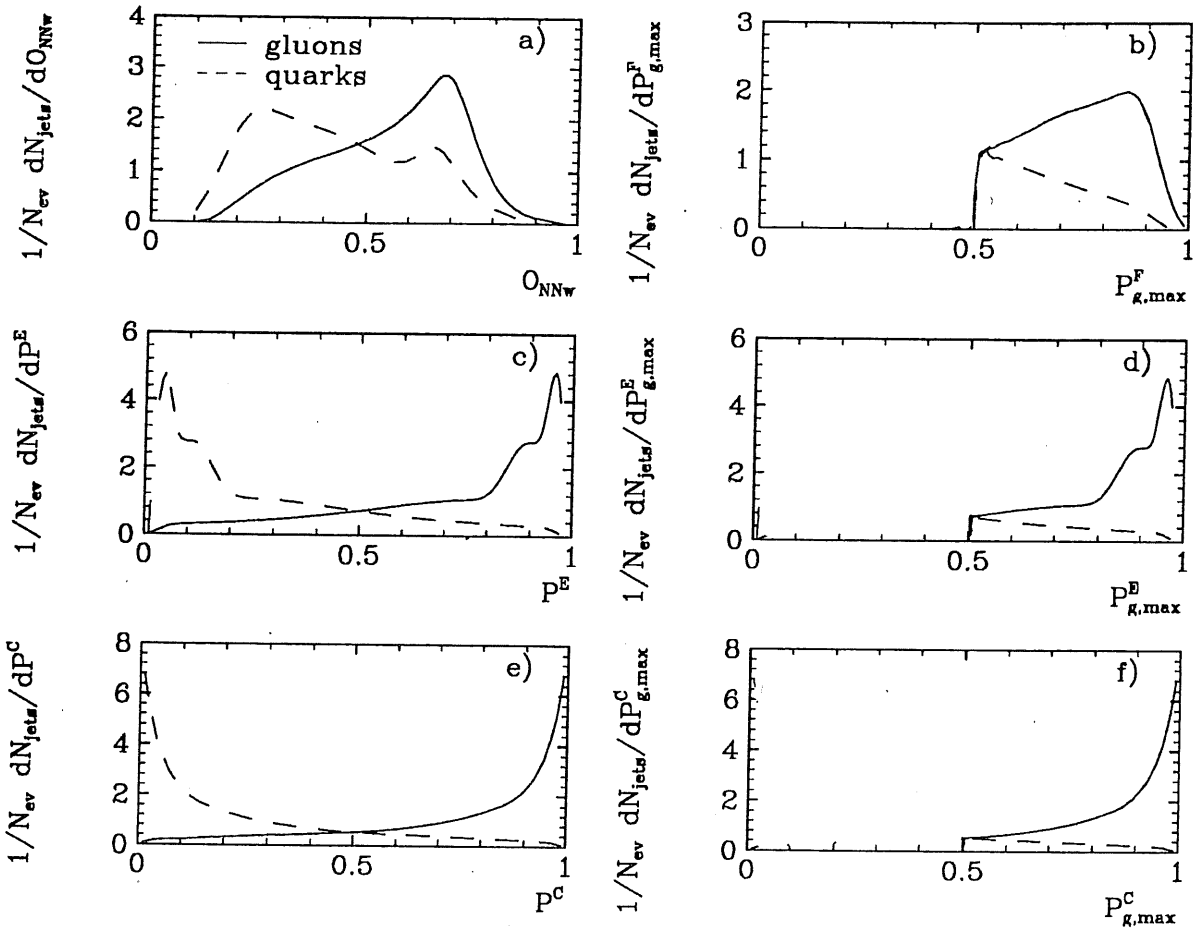


Figure 5: *Results on (2+1)-jet events from ep-collisions shown as a) the neural network output for individual jets b) the $P^F_{g,max}$ distribution from the fragmentation c) the gluon probability distribution according to equation (2) d) the $P^E_{g,max}$ distribution from the jet energies e) the combined conditional probability distribution according to equation (8) f) the P_g^{max} distribution from the combined conditional probability.*

network trained on a HERWIG sample and applied on a HERWIG test sample gives a much higher degree of identification than a network trained on a LEPTO sample and tested on a LEPTO sample. If we now compare the Figs. 6b and f we notice that the curves are almost identical, indicating that the separation between quarks and gluons are optimized in the same way by the two networks trained on LEPTO and HERWIG.

9 Conclusions

We have studied the problem of identifying jets using the jet energies and fragmentation variables separately. The conditional probability for a jet to originate from a gluon (or a quark) can be calculated from Bayes' theorem provided the density functions for gluons and

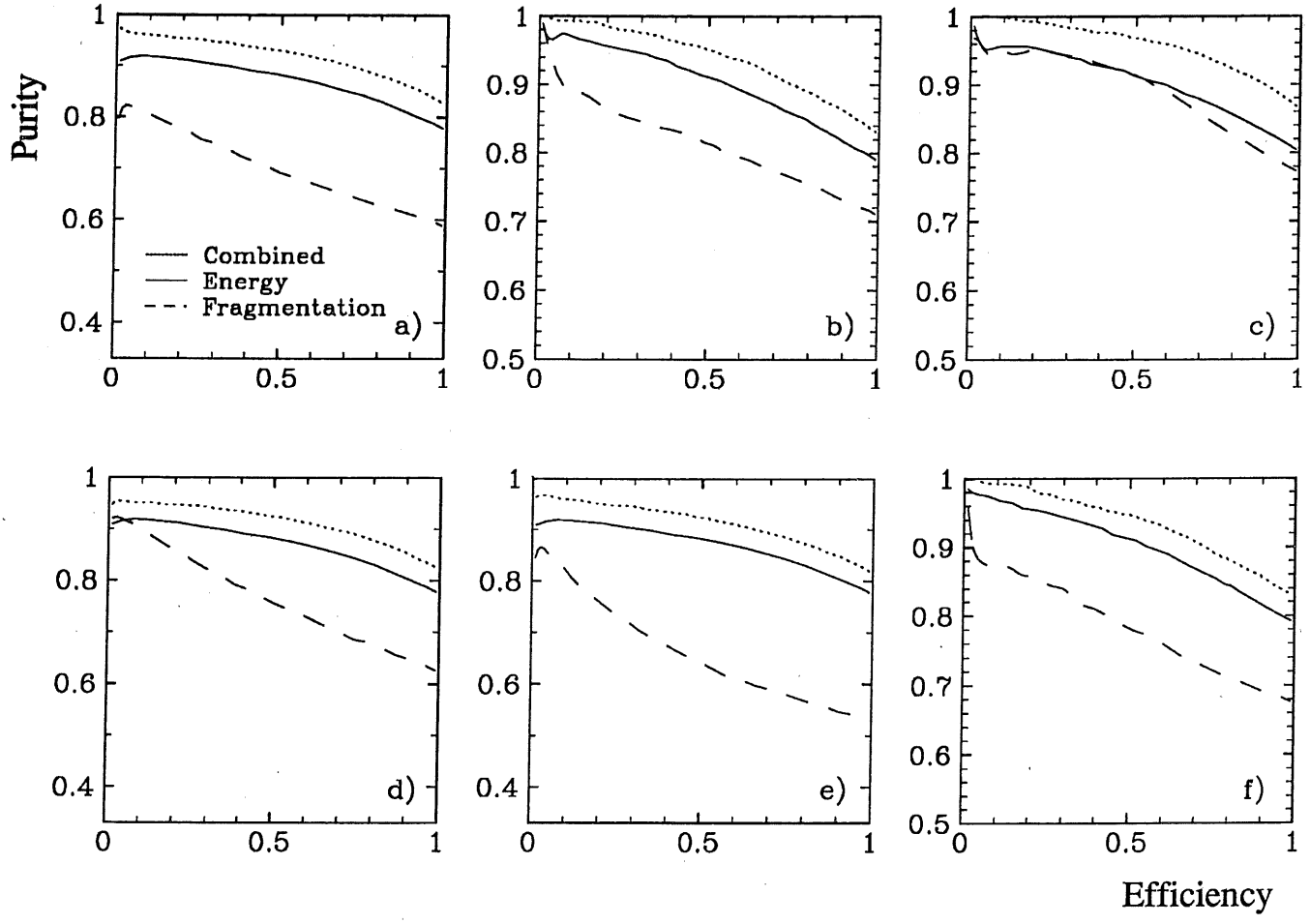


Figure 6: *The purity as a function of the efficiency for a network a) trained on a JETSET sample and tested on a JETSET sample b) trained on a LEPTO sample and tested on a LEPTO sample c) trained on a HERWIG sample and tested on a HERWIG sample d) trained on a LEPTO sample and tested on a JETSET sample e) trained on a HERWIG sample and tested on a JETSET sample f) trained on a HERWIG sample and tested on a LEPTO sample.*

quarks with respect to jet energies and fragmentation variables are known. The formalism to extract these probabilities, for the event types investigated here, has been presented. The advantage in working with probabilities is the simple interpretation of the results and the procedure to combine the results from the energy and fragmentation based identification methods.

A large number of fragmentation variables have been studied of which we found the Fodor moments to give the best performance. Three different Monte Carlo generators were used to test the model and process dependences. The results show a significant process dependence but only a minor model dependence.

The event based level of identification using the jet energies is about 80% at 100% efficiency whereas it is 60% and 70% for e^+e^- 3-jet events and *ep* QCD-Compton events respectively,

also at 100% efficiency, using fragmentation variables. The identification of the jets in a complete event is a much stronger requirement than the identification of individual jets and therefore a direct comparison is not possible.

Acknowledgements: The analysis described in this paper as well as the way the results should be presented have gone through several metamorphoses as the work has proceeded. In the initial phase we had many animated discussions with O. Bärring and V. Hedberg which forced us to find convincing arguments for our analysis strategy.

A Appendix

Bayes' Theorem

If the entire event space is composed of the subsets $B_i, (i = 1...n)$, with no elements in common, then the subsets are said to be *mutually exclusive* and *exhaustive*, which means that

$$\sum_{i=1}^{n} P(B_i) = 1 \tag{10}$$

Provided A is also a set that belongs to the event space, Bayes' theorem states

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)} \tag{11}$$

This theorem can be proven by starting from the definition of the conditional probability, $P(A \cap B) = P(B|A)P(A) = P(A|B)P(A)$, where $P(B|A)$ should be interpreted as the probability that the event B occurs under the condition that A has already occured. For our subset B_i we thus get

$$P(A \cap B_i) = P(B_i|A)P(A) = P(A|B_i)P(B_i)$$

$$\Rightarrow P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} \tag{12}$$

The elements of a set might be classified according to more than one criterion so that we for example have $\sum_{i=1}^{m} P(A_i) = \sum_{j=1}^{n} P(B_j) = 1$. If some of the criteria are being neglected in the classification we can define the *marginal probability* for A_i according to

$$P(A_i) = \sum_{j=1}^{n} P(A_i \cap B_j) = \sum_{j=1}^{n} P(A_i|B_j)P(B_j) \tag{13}$$

where we again have used the definition of the conditional probability for the second step. Using this expression we can now rewrite the equation (12) to obtain equation (11).

If we now consider the (2+1) jet case we could take the subset B_i to represent the four configurations of quark- and gluon jets possible if the jets are identified individually in the event. These are $B_i = \{qq, qg, gq, gg\}$, where for example qg means that jet1 is a gluon jet and jet2 a quark jet. Since we have selected events which have one quark and one gluon in the final state (QCD-Compton events) we introduce this information into our probabilities by considering only the allowed configurations, and define the two subsets $B_1 = \{qg\}$, $B_2 = \{gq\}$. Assuming that the two jets are in a state C, which might refer to energy and/or fragmentation variable values, we can compute the probability of having a qg configuration in the state C, by using Bayes' theorem.

$$P(gq|C) = \frac{P(gq)P(C|gq)}{P(gq)P(C|gq) + P(qg)P(C|qg)} \tag{14}$$

This formula is valid when state C has a positive probability $P(C) > 0$. When C is defined in terms of a continously varying quantity, the discrete probabilities in Bayes formula should

be replaced by probability density functions. For example, if C represents a continous energy variable E, then (14) reads

$$P(gq|E) = \frac{P(gq)f_{gq}(E)}{P(gq)f_{gq}(E) + P(qg)f_{qg}(E)} \tag{15}$$

where f_{gq} and f_{qg} are the joint energy density functions for pairs gq and qg, respectively.

For the 3-jet case, we have 8 possible configurations of quarks and gluons. After deleting the impossible combinations the remaining subsets are $B_i = \{gqq, qgq, qqg\}$. The final expression becomes

$$P(gqq|C) = \frac{P(gqq)P(C|gqq)}{P(gqq)P(C|gqq) + P(qgq)P(C|qgq) + P(qqg)P(C|qqg)} \tag{16}$$

References

[1] L. Lönnblad, Peterson and Rögnvaldsson, Phys. Rev. Lett. 65(1990)1321.
L. Lönnblad, Peterson and Rögnvaldsson, Nucl. Phys. B349(1991)675.

[2] Csabai, Czako and Fodor, Nucl. Phys. B374(1992)288.

[3] T. Sjöstrand, Computor Phys. Comm. 39 (1986) 347, it ibid 43 (1987) 367.

[4] B. Andersson, G. Gustafson, G. Ingelman, T. Sjöstrand, Phys. Rep. 97(1983)31.

[5] G. Ingelman, LEPTO version 6.1, in Proceedings 'Physics at HERA', Eds W. Buchmüller, G. Ingelman, DESY Hamburg 1992, vol. 1-3 and references therein.

[6] G. Marchesini, B.R. Webber, G. Abbiendi, I.G. Knowles, M.H. Seymour and L. Stanco, Computer Phys. Comm. 67(1992)465.

[7] T. Sjöstrand, Computer Phys. Comm. 28 (1983) 227

[8] V. Hedberg, G. Ingelman, C. Jacobsson, L. Jönsson, Z. Phys. C63 (1994) 49.

[9] D. Graudenz, N. Magnussen in Proceedings 'Physics at HERA', Eds W. Buchmüller, G. Ingelman, DESY Hamburg 1992, vol. 1, p. 261.

[10] K. C. Hoeger, Measurement of x, y, Q^2 in Neutral Current Events, in Proceedings 'Physics at HERA', Eds W. Buchmüller, G. Ingelman, DESY Hamburg 1992, vol. 1, p. 43

[11] L. Lönnblad, C.Peterson, H. Pi and T. Rögnvaldsson, Comp. Phys. Comm. 67(1991)193.

[12] Fodor, Phys. Rev. D41(1990)1726.

[13] M.D. Richard, R.P. Lippman, Neural Comp 3 (1991) 461.