Submitted to the
International Europhysics Conference on High Energy Physics
July 17 - 23, 2003, Aachen, Germany

Abstract: 511 **Session: HP, NPFT**

A search for QCD-instanton induced events in deep inelastic ep scattering at HERA

ZEUS Collaboration

12th June 2003

Abstract

A search for QCD-instanton induced events in deep-inelastic ep scattering (DIS) has been performed with the ZEUS detector at the HERA accelerator, using data corresponding to an integrated luminosity of $38.3\,\mathrm{pb}^{-1}$. A kinematic range defined by cuts on the photon virtuality, $Q^2 > 120\,\mathrm{GeV}^2$, and on the Bjorken scaling variable, $x > 10^{-3}$, has been investigated. QCD-instanton induced events were modelled by the Monte Carlo generator QCDINS. In an instanton-enhanced subsample, sphericity S has been fitted by a combination of signal and background Monte Carlo. The fits are statistically compatible with an instanton fraction of zero. A background dependent 2σ upper limit at the theoretically expected value of the instanton fraction is obtained. Attributing all data events in an instanton enhanced subsample to a signal, a background independent most conservative 2σ limit of 30 pb is obtained. This cross section is a factor 3.5 above the theoretical expectation.

1 Introduction

In the Standard Model, both the strong and the electroweak interaction are described by non-abelian gauge theories. In such theories, the ground state has a rich topological structure. Its simplest building blocks are non-perturbative fluctuations of the gauge-fields [1], called instantons [2]. They can be interpreted as tunneling processes of the gauge fields between topologically distinct types of vacuum states. The name instanton refers to the fact that these fluctuations are localised in space and Euclidean time. In quantum chromodynamics (QCD) a number of theoretical problems can be addressed by taking instanton effects into account, for a review see e.g. [3].

Although instantons are theoretically required by the Standard Model, they have not been observed experimentally. While in electroweak interactions instantons might play a role only at center of mass energies $\gg 10\,\mathrm{TeV}$ [4, 5], in QCD instanton-effects are expected to become sizable at much lower energies, because the strong coupling constant α_s is much larger than the equivalent parameter in electroweak theory.

QCD instantons are predicted to have short distance implications [5–8], see also [9]. In particular, they can induce characteristic events in neutral current (NC) deep-inelastic ep scattering (DIS) at HERA.

HERA therefore offers a unique chance to discover instantons. This would constitute a confirmation of an essential non-perturbative Standard Model prediction, connected with the QCD vacuum. Results of such a search have recently been reported by the H1 collaboration [10]. The strategy of this paper is as follows: Since instanton-induced events are predicted to contribute only $\approx 0.7\%$ to the cross section of the inclusive sample used in this search, it is essential to find variables that efficiently discriminate between instanton and normal DIS events (see section 6). Statistical discrimination methods (section 7) have then been used to obtain event samples with a large fraction of instanton events. These samples have been used to obtain new conservative upper limits on the instanton production cross section, close to the theoretical prediction.

2 Characteristics of instanton-induced events

Ringwald and Schrempp [5, 7, 8, 11–13] have identified kinematic regions in DIS which allow a perturbative calculation of instanton-induced processes. These processes lead to a characteristic final state, that allows instanton-induced events to be distinguished from events stemming from normal DIS processes. They predict [12] a non-negligible rate of these events. Fig. 1 shows a diagram of instanton-induced events at HERA. Such diagrams

form the basis of the perturbative calculation of Ringwald and Schrempp. The incoming lepton emits a photon, which in turn transforms into a quark-antiquark pair. One of these quarks hadronises to form the current jet with four-vector q''. The other quark, with four-vector q', fuses with a gluon (four-vector g) from the proton. The phenomenological characteristics of instanton-induced events, as worked out in [5], can be summarised as follows: In the hard subprocess exactly one $q\bar{q}$ pair of each of the n_f quark flavours, which are kinematically accessible, participates in the hard subprocess of each event. This gives rise to a high-multiplicity final state with large transverse energy. Particles stemming from the instanton are expected to be isotropically distributed in their centre of mass frame.

Instanton-induced events in DIS are simulated using the Monte Carlo (MC) generator QCDINS [12], version 2.0. QCDINS simulates the hard subprocess in the presence of an instanton. For the Description of parton showers and hadronisation, HERWIG [14] is used.
The simulation of the hard subprocess is accomplished by applying instanton-perturbation theory around the one-instanton solution. By comparing the results obtained from instanton perturbation theory to those from high quality lattice simulations of QCD, the fiducial region of the QCDINS MC program has been derived [8] in terms of the kinematic variables Q', x', describing the hard subprocess of instanton-induced events, see Fig. 1.

$$Q' \geq Q'_{\min} = 30.8\ \Lambda_{\overline{\rm MS}}^{(n_f)} = \sqrt{113\,\mathrm{GeV}^2} \qquad x' \geq x'_{\min} = 0.35$$

$$\text{with} \quad Q'^2 = -q'^2 = -(q - q'')^2 \qquad x' = Q'^2/(2g \cdot q')$$

In order to restrict the QCDINS sample to the region, where the calculation is reliable, these cuts on the generated x' and Q'^2 values were made when generating the QCDINS sample. Non-planar contributions [7], which are not taken account of in the calculation, are suppressed by applying a further cut on the photon virtuality,

$$Q^2 = -q^2 = -(e - e')^2 \geq Q^2_{\min} = Q'^2_{\min} \quad ,$$

where e and e' denote the four-momenta of the incoming and outgoing positron, respectively. The uncertainty of the value of the strong coupling constant gives rise to an uncertainty of the QCDINS predicted cross section of $\approx 30\%$ within the cuts on x', Q'^2, Q^2 [15].

3 Experimental setup

The data sample used in this search corresponds to an integrated luminosity of $38.3 \pm 0.6\,\mathrm{pb}^{-1}$ collected with the ZEUS detector at the HERA collider. During the years

1996/1997, positrons of energy $E_e = 27.5\,\text{GeV}$ were collided with protons of energy $E_p = 820\,\text{GeV}$. A detailed description of the ZEUS detector can be found elsewhere [16]. The main detector components used in the search presented are the central track detector (CTD) [17], operating in a magnetic field of 1.43 T provided by a thin superconducting solenoid, and the uranium-scintillator calorimeter (CAL) [18].

Tracking information is provided by the CTD, in which the momenta of tracks in the polar-angle[1] region $15° < \theta < 164°$ are reconstructed. The CTD consists of 72 cylindrical drift chamber layers, organised in nine superlayers. The relative transverse momentum, p_T, resolution for full-length tracks can be parametrised as $\sigma(p_T)/p_T = 0.0058p_T \oplus 0.0065 \oplus 0.0014/p_T$, with p_T in GeV.

The CAL covers 99.7% of the total solid angle. It is divided into three parts with a corresponding division in θ, as viewed from the nominal interaction point: forward (FCAL, $2.6° < \theta < 36.7°$), barrel (BCAL, $36.7 < \theta < 129.1°$), and rear (RCAL, $129.1 < \theta < 176.2°$). Each of the CAL parts is subdivided into towers, which in turn are segmented longitudinally into one electromagnetic (EMC) and one (RCAL) or two (FCAL,BCAL) hadronic (HAC) sections. The smallest subdivision of the CAL is called a cell. Under test-beam conditions, the CAL single-particle relative energy resolution was $\sigma(E)/E = 18\%/\sqrt{E}$ for electrons and $\sigma(E)/E = 35\%/\sqrt{E}$ for hadrons, with E in GeV.

The luminosity was measured using the Bethe-Heitler reaction $ep \to e\gamma p$. The resulting small-angle energetic photons were measured by the luminosity monitor [19], a lead-scintillator calorimeter placed in the HERA tunnel at $Z = -107\,\text{m}$.

A three-level trigger [20] was used to select events online. A high Q^2 NC DIS sample was selected requiring a positron with an energy greater than 4 GeV at a position outside a radius of 25 cm on the calorimeter at the third level of the trigger system.

4 Simulation of standard DIS events

Standard DIS events were simulated using the HERACLES [21] program with DJANGOH version 6 [22], which includes EW initial- and final state radiative corrections. To describe the hard subprocess, the version based on the Colour Dipole model (CDM) [23] as implemented in the program ARIADNE [24] was chosen, because this model is known to yield overall the best description of most final state data [25].

[1] The ZEUS coordinate system is a right-handed Cartesian system, with the Z axis pointing in the proton beam direction, referred to as the "forward direction", and the X axis pointing left towards the center of HERA. The coordinate origin is at the nominal interaction point. The pseudorapidity is defined as $\eta = -\ln\left(\tan\frac{\theta}{2}\right)$, where the polar angle, θ, is measured with respect to the proton beam direction.

In the Colour Dipole model, QCD radiation is described in the form of parton showers evolving from independently radiating colour dipoles, with no difference being made between initial and final state radiation. To simulate the hadronisation phase, the Lund string fragmentation [26] was applied as implemented in JETSET [27]. In order to improve the description of the hadronic final state, the width of the transverse momentum distribution of primary hadrons (i.e. JETSET parameter PARJ(21)) was lowered to 0.28 GeV. In order to describe the diffractive contribution to the inclusive neutral current sample, 12% of diffractive RAPGAP events were added. The percentage was determined from a fit of the distribution of the variable $\eta_{\max}$. Here, $\eta_{\max}$ is the pseudorapidity of the calorimeter energy deposit with the smallest polar angle and an energy above 400 MeV. The CTEQ4 [28] parton distribution functions were used.
For a cross-check of the results, the generator HERWIG [14] has been used as an alternative model, because HERWIG is the program providing the description of parton showers and hadronisation for the QCDINS MC sample. In HERWIG, coherence effects in the final-state cascade are included by angular ordering of successive parton emissions, and a clustering model is used for hadronisation [29].

The generated events were passed through the GEANT-based [30] ZEUS detector- and trigger-simulation programs [31] [32]. They were reconstructed and analysed by the same program chain as the data.

5 Event Selection and Reconstruction

5.1 Reconstruction of kinematic variables

Both track and calorimeter information were used for event reconstruction. Calorimeter cells were first grouped to form clusters and these clusters were then associated with tracks, where possible, to form energy-flow objects, EFO's, associated with the hadrons formed in the interaction [33]. The hadronic final state comprises all EFOs, that do not stem from the scattered positron.
Positron identification was performed from the pattern of energy deposits in the CAL [34]. The positron energy was calculated according to the double-angle method [35]:

$$E_{\mathrm{e,DA}} = 2E_e \, \frac{\sin\gamma_h}{\sin\theta + \sin\gamma_h - \sin(\theta+\gamma_h)} \ ,$$

where E_e is the positron beam energy, θ the positron polar angle and γ_{h} the hadronic scattering angle, measured in the laboratory frame. The hadronic scattering angle is

defined by

$$\cos\gamma_{\rm h} = \frac{p_{\rm T,h}^2 - \delta_{\rm h}^2}{p_{\rm T,h}^2 + \delta_{\rm h}^2} \quad \text{with} \quad p_{T,h}^2 = (\sum_{\rm EFOs} p_x)^2 + (\sum_{\rm EFOs} p_y)^2 \quad \text{and} \quad \delta_{\rm h} = \sum_{\rm EFOs} (E - p_z) \quad .$$

Both the transverse momentum, $p_{T,h}$, of the hadronic final state and $\delta_{\rm h}$ are calculated by summing over components of all EFO four-momenta (p_x, p_y, p_z, E), except the one of the positron. To calculate the positron polar angle, tracking information was used, if the polar angle in the CAL was above 0.3 and if the track traversed more than three CTD superlayers. Otherwise these angles were found from calorimeter information alone.

The kinematic region investigated is defined by cuts on the virtuality of the photon, Q^2, and on the Bjorken scaling variable $x = Q^2/(2P \cdot q)$, where q denotes the four-momentum of the photon emitted by the positron, and P the four-momentum of the proton, cf. Fig. 1. The variables Q^2 and x were reconstructed [35] using the double-angle method. To ensure a good reconstruction of $Q^2_{\rm DA}$ and $x_{\rm DA}$, a cleaning cut on $y = (q \cdot P)/(e \cdot P)$, the fraction of the positron energy transferred to the proton in its rest frame, is applied, where y is reconstructed according to the Jacquet-Blondel method [36]:

$$y_{\rm JB} = \delta_{\rm h}/(2E_e) \quad .$$

In order to reconstruct the kinematic variable Q'^2 (see Fig. 1), EFOs were assigned to the current jet and to the instanton part of the hadronic final state. The current jet was identified by applying the k_T-cluster algorithm [37] in the longitudinally invariant form [38] on all EFOs in the hadronic centre of mass frame (hcms), defined by $\vec{P} + \vec{q} = 0$. The direction, in which the photon moves, is chosen as negative z-direction. Monte Carlo studies at parton level [39] were used to select the best current jet candidate. These studies have shown, that in the QCDINS enriched region at high transverse jet momenta, the current quark on average has a larger pseudorapidity than the partons assigned to the instanton. Therefore, the current jet was chosen from the list of jets provided by the k_T cluster algorithm as follows: The list was sought for jets with a pseudorapidity $\eta_{\rm jet}^{\rm hcm}$ less than the transverse energy E_T-weighted mean pseudorapidity of all EFOs in the event (except those attributed to the positron):

$$\eta_{\rm jet}^{\rm hcm} < \frac{\sum_{\rm EFOs} E_T \eta}{\sum_{\rm EFOs} E_T} \quad ,$$

where the jet pseudorapidity is calculated from the jet four-momentum, obtained by summing the four-momenta of all the EFOs assigned to the jet.

Of the jets fulfilling that requirement, the one with the highest transverse energy was chosen as current jet.
The QCDINS sample has been generated with default cuts on the generated Q'^2 and

x' applied. Additionally, a cut on the reconstructed variable Q'^2_{DA} was made. It was calculated using the expression

$$Q'^2_{\mathrm{DA}} = Q^2_{\mathrm{DA}} + \frac{W^2_{\mathrm{DA}} - Q^2_{\mathrm{DA}}}{W_{\mathrm{DA}}} \cdot \sum_{i \in \mathrm{Jet}} E_i - \frac{W^2_{\mathrm{DA}} + Q^2_{\mathrm{DA}}}{W_{\mathrm{DA}}} \cdot \sum_{i \in \mathrm{Jet}} p_{z,i} - M^2_{\mathrm{Jet}} \ ,$$

with

$$W_{\mathrm{DA}} = \sqrt{Q^2_{\mathrm{DA}}(1 - x_{\mathrm{DA}})/x_{\mathrm{DA}} + m_p^2} \ ,$$

the current jet mass M_{Jet} and the proton mass m_p.
This particular form was found, based on MC investigations, to give a reasonably accurate reconstruction of Q'^2. Even so, the distribution of the reconstructed Q'^2 has a tail in the direction of overestimating the true value.

5.2 Selection of neutral current DIS events

A radius cut of 36 cm for the impact point of the positron track on the rear calorimeter surface was required. Further cuts were applied as listed below, defining what will be referred to as fiducial sample in what follows:

- kinematic cuts
  - $Q^2_{\mathrm{DA}} > 120\,\mathrm{GeV}^2$;
  - $x_{\mathrm{DA}} > 10^{-3}$;
  - $y_{\mathrm{JB}} > 0.05$;
- vertex cut:
  - z position of the event vertex, $|z_{\mathrm{Vtx}}| < 50\,\mathrm{cm}$;
- cuts to ensure the quality of the positron reconstruction:
  - $E_{\mathrm{e,DA}} > 10\,\mathrm{GeV}$;
  - $E_{\mathrm{cne}} < 5\,\mathrm{GeV}$;
    E_{cne} is the energy not stemming from the positron, which is found inside a cone around the track of the positron candidate. This cone is defined by having the (η, ϕ)-radius $R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.8$ at the impact point of the positron track on the calorimeter surface;
- suppression of photoproduction events:
  - $y_{\mathrm{el}} < 0.90$;

- $35\,\mathrm{GeV} < \sum_{\mathrm{efos}}(E - p_z) < 65\,\mathrm{GeV}$,
  where the sum runs over components of all efo four-momenta (p_x, p_y, p_z, E).
- DCA $< 10\,\mathrm{cm}$:
  The distance of closest approach, DCA, is calculated between the positron track and the centre of the cluster of CAL cells assigned to it.

• restriction to the fiducial region of the QCDINS MC:

- $Q'^2_{\mathrm{DA}} > 140\,\mathrm{GeV}^2$.

The cut on Q'^2_{DA} is made to restrict the data sample to a region where the calculation implemented in QCDINS is reliable.

No cut was made on the variable x' for the normal DIS MC sample and the data, whereas the QCDINS Monte Carlo includes a cut $x' > 0.35$. The reason is, that x' cannot be reconstructed well from the data. Therefore, these samples might contain a slightly larger contribution from instanton-induced events than predicted from the QCDINS sample used in the search. However, lattice calculations show a steep decrease of the instanton contribution towards small values of $x' < 0.35$, corresponding to a small instanton anti-instanton separation, suggesting that this region can be neglected [8,13,15].

These cuts select 91846 events in the data, 596 events of the QCDINS signal MC sample and 88320 and 76366 events for DJANGOH with RAPGAP and HERWIG, respectively, if numbers of events in the MC samples are normalised to data luminosity.
Within these cuts the predicted instanton cross section is 8.9 pb.
There is a discrepancy in the number of events measured by HERWIG and DJANGOH with RAPGAP. It can be traced to the different hadronization routines in the two Monte Carlo programs. In the subsequent analysis, numbers of standard DIS Monte Carlo events are normalised to the number of data events in the fiducial sample, where the predicted QCDINS contribution is negligible ($\approx 0.7\%$).

6 Definition of discriminating variables

Variables which are effective for distinguishing instanton-induced events from standard DIS events should have widely different distributions and should not depend strongly on the particular DIS Monte Carlo used. Two kinds of such variables can be distinguished:

(i) Variables connected with the kinematics of an instanton-induced event. The variables chosen are:

Q'^2_{DA} , (see sect. 2) and p_T^{Jet}, the transverse energy of the current jet in the hadron CMS.

(ii) Variables connected with the final state particles of the instanton system, the so called shape variables.
These variables are calculated from a subset of hadronic final state EFOs, which is assigned to the instanton, referred to as "instanton part of the hadronic final state" or "instanton region" in what follows. The instanton region comprises all EFOs which have not been assigned to the current jet, and which lie in the hemisphere opposite to the outgoing proton, with polar angle $\theta_{\rm efo}^{\rm hcm} > 90°$. Once the instanton part of the hadronic final state has been identified, two multiplicities are determined as shape variables:

- $N_{\rm efo}$, the multiplicity of EFOs assigned to the instanton region, and

- $N_{\rm eft}$, the multiplicity, of tracks, which have been used in constructing these EFOs.

The QCDINS distribution peaks at higher multiplicities than the normal DIS MC and the data. At high multiplicities, the normal DIS description overshoots the data, at low multiplicities, it is below the data points.

- C, the circularity is a measure of the isotropy of EFOs assigned to the instanton in the hadronic centre of mass frame with respect to the photon-proton axis. To determine the circularity, the normalised two-dimensional momentum tensor,

$$M_{\alpha\beta}^{(2D)} = \frac{\sum_j p_{j,\alpha} p_{j,\beta}}{\sum_j \left(p_{j,x}^2 + p_{j,y}^2\right)} \quad \text{with } \alpha, \beta = x,\, y \ ,$$

is computed from the hcms-three-momenta of all EFOs assigned to the instanton. From its eigenvalues λ_1, λ_2, with $\lambda_1 > \lambda_2$, the circularity is obtained: C $=2(1-\lambda_2)$.

- S, the sphericity of the instanton part of the event in its centre of mass frame. The sphericity S is a measure of how isotropically a collection of three-momenta is distributed, large values corresponding to a more isotropic distribution.
The normalised momentum tensor, calculated from the EFOs assigned to the instanton in the instanton region cms frame,

$$M_{\alpha\beta}^{(3D)} = \frac{\sum_j p_{j,\alpha} p_{j,\beta}}{\sum_j \left(p_{j,x}^2 + p_{j,y}^2 + p_{j,z}^2\right)} \quad \text{with } \alpha, \beta = x,\, y,\, z \quad ,$$

has eigenvalues $Q_1, \ldots, Q_3$ with $0 \le Q_1 \le Q_2 \le Q_3$. The sphericity is found according to the definition S $= 3/2 \cdot (Q_1 + Q_2)$. Within these cuts, in which the instanton-induced contribution is expected to be negligible, the sphericity data distribution is well described by the standard DIS Monte Carlo.

• ϵ' is a measure of the η density of EFOs assigned to the instanton part of the event in the hcms. All EFOs assigned to the instanton part of the event are sorted with respect to η in the hcms, yielding $N_{\rm efo}$ values $\eta_1, \eta_2, \ldots, \eta_{N_{\rm efo}}$.
The closed interval $[\eta_i, \eta_{i+k}]$, $i + k \leq N_{\rm EFO}$, thus contains $k + 1$ EFOs. The variable ϵ' is then defined by

$$\epsilon' = \epsilon_c - b(N_{\rm efo} - 30) \quad \text{with} \quad b = 0.339$$

$$\text{and} \quad \epsilon_c = \frac{k}{N_{\rm efo} - k} \sum_{i=1}^{N_{\rm efo}-k} \frac{1}{\eta_{i+k} - \eta_i} \quad \text{where} \quad k = \begin{cases} N_{\rm EFO}/2 & \text{for } N_{\rm EFO} \text{ even}, \\ (N_{\rm EFO}+1)/2 & \text{for } N_{\rm EFO} \text{ odd} \end{cases} .$$

7 Instanton enhancement

In order to enhance the instanton contribution and thus increase the sensitivity to a possible signal, cuts were applied on the following variables: $Q'^2_{\rm DA}$, the transverse momentum $p_T^{\rm Jet}$ of the current jet in the hadronic centre of mass frame, the multiplicities $N_{\rm efo}$ and $N_{\rm eft}$, the sphericity S and the circularity C of the instanton part of the hadronic final state and the EFO correlation variable ϵ', as defined in section 6. The following figures show the distributions of these variables: Distributions of $N_{\rm efo}$ and $N_{\rm eft}$ are shown in Fig. 2. The distribution of C is shown in Fig. 2 (d). Fig. 2 (c) shows the sphericity distribution. The distribution of ϵ' is shown in Fig. 3. There is a reasonable agreement between data and standard DIS Monte Carlo, however there are also discrepancies on a scale, which makes discrimination between standard DIS and instanton induced events difficult. This indicates the need to enhance the fraction of instanton induced events in the data.

Several enhancement methods were used, one of them, the Fisher algorithm, is described in detail below. From the numbers

$N_O(I_O)$: number of standard DIS MC (QCDINS) events in the fiducial sample

$N_E(I_E)$: number of standard DIS MC (QCDINS) events in the enhanced sample

the efficiencies r_I of QCDINS and r_N of normal DIS MC were obtained according to

$$r_I = \frac{I_E}{I_O} \quad , \qquad r_N = \frac{N_E}{N_O} \quad .$$

In order to compare the performance of the different enhancement methods, values of the separation power,

$$P_s = \frac{r_I}{r_N} = \frac{I_E/I_O}{N_E/N_O} \quad ,$$

were compared for different samples of similar QCDINS efficiency.

The number of instanton-induced events has been enhanced compared to standard DIS events by one-dimensional cuts, labelled 1D in the tables, by requiring $p_T^{\rm Jet} > 4\,{\rm GeV}$, $Q'^2_{\rm DA} < 250\,{\rm GeV}^2$, $N_{\rm eft} > 7$, $C > 0.5$, $S > 0.5$.

A second method (Fisher algorithm) is cutting on a linear combination of input discriminating variables x_i. The coefficients were obtained from the algorithm [40,41], additionally requiring $Q'^2_{\rm DA} < 250\,{\rm GeV}^2$.
For n input variables x_i, $i = 1, 2, \ldots, n$, the mean values $\overline{x_i}^s$ for the signal and $\overline{x_i}^b$ for the background were determined. The correlation matrices between the variables were calculated according to

$$V_{ik,s} = \frac{1}{N} \sum_{\rm events} (x_i^s - \overline{x_i}^s)(x_k^s - \overline{x_k}^s)$$

for the signal sample, and equivalently for the background ($V_{ik,b}$). These were averaged, $\overline{V}_{ik} = \frac{1}{2}(V_{ik,s} + V_{ik,b})$. The resulting matrix was inverted. The Fisher discriminant is defined as a sum over the input variables,

$$t = \sum_{i=1}^{n} w_i x_i \quad \text{with weights} \quad w_i = \sum_k (\overline{V}^{-1})_{ik} (\overline{x_k}^s - \overline{x_k}^b) \quad .$$

The method was then applied to the discriminating variables S, C, $p_T^{\rm Jet}$,$N_{\rm efo}$, $N_{\rm eft}$ and ϵ'. Requiring to keep about 10% of the QCDINS sample yields a cut value of $t = 9.5$. A distribution of this discriminant is shown in Fig. 4. The resulting numbers are shown in Table 1.
Measured by the separation power P_s at fixed instanton efficiency r_I, the Fisher algorithm performs better than a procedure using one-dimensional cuts on the discriminating variables. It also performs slightly better than a combination of optimally chosen two-dimensional cuts. Looking at the table, it has to be noted, that the number of events is slightly smaller than the predicted DIS background. Therefore no meaningful DIS background subtraction, based on absolute event numbers, can be made. A key to understanding this difficulty could lie in the fact, that the discriminating cuts leave only about 0.3% of the number of events in the original fiducial sample, and in such an extreme region of phase space absolute event numbers cannot be expected to be very accurate.

8 Fits to the sphericity distribution

8.1 Procedure

The discussion following Table 1 has shown, that no reliable background subtraction, based on event numbers, is possible. However, the shape of discriminant variables could still contain additional information, independent of absolute numbers. In particular, the distribution of the sphericity S is rather different for instanton-induced and standard DIS events, and contains therefore additional information on the instanton contribution to the data. A fit of this distribution can then be used to extract the fraction of instanton events in the data.

The predicted fraction $f_{\mathrm{I,th}}$ of instanton-induced events within a given data sample is defined as the ratio of the cross sections for instanton-induced events predicted by the QCDINS MC program, σ_I, and the cross section of the data, σ_D:

$$f_{\mathrm{I,th}} = \frac{\sigma_I}{\sigma_D} \quad \text{where} \quad \sigma_I = \frac{N_c}{N_g}\sigma_g \quad \text{and} \quad \sigma_D = \frac{N_D}{\mathcal{L}} \quad .$$

Here, N_c is the number of QCDINS events within the cuts that constrain the sample, N_g is the number of events generated, and σ_g denotes the QCDINS generated cross section. The data cross section is obtained from the number N_D of data events in the sample and the luminosity $\mathcal{L}$. In the inclusive NC DIS sample, the predicted fraction of instanton-induced events is 0.7%. The predictions for the different samples used for the fit approach are given in table 2.

For the fit, the function χ^2 is considered as negative log likelihood function. It is given by:

$$\chi^2(f_{\mathrm{I}}) = \sum_{i=1}^{n_{\mathrm{bins}}} \frac{\{n^*_{iD} - [f_{\mathrm{I}} \cdot n^*_{iI} + (1 - f_{\mathrm{I}}) \cdot n^*_{iN}]\}^2}{{\sigma^*_{iD}}^2 + f_{\mathrm{I}}^2 {\sigma^*_{iI}}^2 + (1 - f_{\mathrm{I}})^2 {\sigma^*_{iN}}^2} \quad ,$$

with the fraction of instanton-induced events in the enhanced sample, f_I, as free parameter of the fit. In this expression, n^*_{iD}, n^*_{iN} and n^*_{iI} are the number of data, standard DIS and QCDINS events within bin i, respectively, σ^*_{iD}, σ^*_{iN} and σ^*_{iI} the respective statistical errors and the asterisk indicates, that each of the distributions has been normalised to 1 within the cuts considered. The statistical error σ^*_i for each of the samples is given by the expression $\sigma^*_i = \sqrt{n_i}/n$, where n_i is the number of events in bin i and n the total number of events in the sample considered.

8.2 Results

Fits to the sphericity S of the instanton region were carried out with an instanton-enhancement sample prepared by the Fisher algorithm. In this application the variables

used were C, $p_T^{\rm Jet}$, $N_{\rm efo}$, and $N_{\rm eft}$. The Fisher discriminant t in this application is required to exceed the value 7.5.

Fig. 5 (a) shows the sphericity distribution in the instanton-enhanced sample. The χ^2 function is shown in Fig. 5 (b) and (c).
The fits give an acceptable value of χ^2 for an instanton fraction of zero. Therefore, only 2σ upper limits on the instanton contribution can be set. According to the maximum likelihood approach, 2σ limits, $f_{\rm I,lim}$, are determined from

$$\chi^2\left(f_{\rm I,lim}\right) = \chi^2_{\rm min} + 4$$

or

$$\chi^2\left(f_{\rm I,lim}\right) = \chi^2(f_I = 0) + 4$$

in case the minimum χ^2 is found to be at a negative, hence unphysical, f_I-value. The resulting 2σ-range of allowed values of the instanton fraction, is indicated by the shaded regions in the χ^2 plots. The results are summarized in Table 2.

Various sources of systematic uncertainty have been considered and checks on the method performed. In each case, the fits have been repeated under changed conditions for both the standard DIS Monte Carlo simulations. The changes are generally smaller than those derived from a comparison of the two DIS Monte Carlos.

For both DIS Monte Carlos, the most likely fit values of the instanton fraction lie below the theoretical expectation, and the fits are compatible with a zero instanton fraction. Therefore only 2σ upper limits on the instanton fraction are given. They are, for both Monte Carlos, below the predicted fraction, with DJANGO giving a more restrictive limit than HERWIG. The study of systematics leads to the same result.

In giving limits on the instanton cross section, one is faced with the difficulty, that the result depends on the MC model used, even though one compares only event shapes and not absolute number of events. Therefore the most conservative approach possible is required, considering the MC, which gives the least stringent limits (HERWIG) and considering all systematic effects. On this basis, one obtains a 2σ upper limit at the value of the predicted instanton cross section. One has to note, however, that the predicted cross-section also has uncertainties, e.g. from the uncertainty of α_s.

9 Background independent limits

As shown in section 7, there are indications that the standard DIS Monte Carlo predictions are not reliable enough in the extreme region of phase space, in which the instanton-induced events are located. In order to be independent of the standard DIS prediction,

an extremely conservative limit is set by assuming, that all data events belong to the instanton signal, in other words, that the standard DIS background is zero.

Cuts have been chosen such that about 10% of the fiducial QCDINS sample are kept, in order not to be sensitive to the precise modelling of the QCDINS Monte Carlo in the tails of distributions, and for comparison with the result of the H1 Collaboration. The results obtained from the different enhancement methods were described in section 7 and are listed in table 1.

When assuming that all observed events belong to the signal, the cut on the Fisher discriminant at $\approx 10\%$ QCDINS efficiency yields an extremely conservative 2σ limit of $179 + 2 \cdot \sqrt{179} = 206$ events. Scaled to the cuts of the fiducial sample, $Q^2 > 120\,\mathrm{GeV}^2$, $x > 10^{-3}$, $0.05 < y < 0.9$, $Q'^2 > 140\,\mathrm{GeV}^2$, this corresponds to a cross section of 30 pb. The limit, using statistical errors only, is a factor 3.4 above the QCDINS prediction of 61 events. Systematic uncertainties were taken into accout by considering a $\pm 3\%$ change in the CAL energy scale, a change in the definition of the instanton region (see Sect. 6), and using Q^2_{el} instead of Q^2_{DA}. This changes the factor to 3.5, if the systematic errors are added quadratically to twice the statistical error.

This factor of 3.5 is conservative, not only because the background has been assumed to be zero, but also because the QCDINS prediction underestimates the actual instanton-induced contribution that might be contained in the data: The QCDINS prediction of 61 events was obtained with cuts $x' > 0.35$ and $Q'^2 > 113\,\mathrm{GeV}^2$ at event generation. Since these variables cannot be reconstructed well, the data might include some instanton-induced events with real x', Q'^2 values below these cut values. Removing these events from the data sample would lower the factor of 3.5. Also, by lowering the fraction of instanton events kept, one would get a smaller limit, still closer to the theoretical value.

Our results can be compared with those published by the H1 collaboration [10]. In a region of phase space differing from ours by a lower range of Q^2, H1 found in an instanton enhanced sample an excess in the data, which is however of about the same size as the difference between the standard DIS Monte Carlo samples. Setting the background to zero, H1 set a most conservative 95% confidence level upper limit of 221 pb for the cross section , within their kinematical region. This value is about a factor 5 above the theoretical prediction.

It appears, that our limits are somewhat more restrictive than the ones of H1, but one also has to note, that our measurement was made at larger values of Q^2, where a comparison with theory may be more reliable.

10 Conclusion

A search for QCD-instanton-induced events has been performed in neutral current deep-inelastic ep scattering based on $38\,\mathrm{pb}^{-1}$ in the kinematic range $Q^2 > 120\,\mathrm{GeV}^2$, $x > 10^{-3}$. Cuts on the Fisher discriminant have been used to obtain instanton-enhanced subsamples. Using these enhanced subsamples, a fit to the sphericity for the fraction of instanton events in the sample is compatible with zero. A background dependent 2σ upper limit at the theoretically expected value for this fraction was obtained.

Assuming all the data events to belong to an instanton signal, a most conservative 2σ limit to the instanton cross-section has been set, independent of the standard DIS background. Scaled to $Q^2 > 120\,\mathrm{GeV}^2$, $x > 10^{-3}$, $0.05 < y < 0.9$, $Q'^2 > 140\,\mathrm{GeV}^2$, this cross section is 30 pb. Including systematic uncertainties, it is a factor of 3.5 above the predicted value.

11 Acknowledgements

We thank the DESY Directorate for their strong support and encouragement. The HERA machine group and the DESY computing staff are gratefully acknowledged for their outstanding operation of the collider and of the data-analysis environment. We appreciate the contributions to the construction and maintenance of the ZEUS detector by many people who are not listed as authors. It is a pleasure to thank F. Schrempp for helpful discussions.

References

[1] A. Belavin et al., Phys. Lett. **B 59**, 85 (1975).

[2] G. 't Hooft, Phys. Rev. Lett. **37**, 8 (1976);
G. 't Hooft, Phys. Rev. **D 14**, 3432 (1976). Erratum in Phys. Rev. D18, 2199 (1978);
G. 't Hooft, Phys. Rep. **142**, 357 (1986).

[3] T. Schäfer and E. Shuryak, Rev. Mod. Phys. **70**, 323 (1998).

[4] A. Ringwald et al., Nucl. Phys. **B 365**, 3 (1991);
M. Gibbs et al., Z. Phys. **C 66**, 285 (1995);
A. Ringwald, Preprint DESY-02-158 (hep-ph/0210209), 2002.

[5] A. Ringwald and F. Schrempp, *Proc. 8th Int. Seminar, Vladimir,Russia, 1994*, D. Grigoriev et al. (ed.), p. 170. World Scientific, Singapore (1995).

[6] I. Balitsky and V. Braun, Phys. Lett. **B 314**, 237 (1993).

[7] S. Moch et al., Nucl. Phys. **B 507**, 134 (1997).

[8] A. Ringwald and F. Schrempp, Phys. Lett. **B 459**, 249 (1999).

[9] E. Witten, Nucl. Phys. **B 149**, 285 (1979).

[10] H1 Coll., C. Adloff et al., Preprint DESY-02-062 (hep-ex/0205078), 2002. Acc. by Eur. Phys. J. **C**.

[11] A. Ringwald and F. Schrempp, Phys. Lett. **B 438**, 217 (1998).

[12] A. Ringwald and F. Schrempp, Comp. Phys. Comm. **132**, 267 (2000).

[13] A. Ringwald and F. Schrempp, Phys. Rev. Lett. **B 503**, 331 (2001);
M. Gibbs, A. Ringwald and F. Schrempp, *Proc. DIS 1995 (Paris)*, J.-F. Laporte and Y. Sirois (eds.), p. 341.).

[14] G. Marchesini, B.R. Webber, Nucl. Phys. **B 310**, 461 (1988);
G. Marchesini et al., Comp. Phys. Comm. **67**, 465 (1992).

[15] A. Ringwald and F. Schrempp, *Proc. DIS 2000, Liverpool*, J. A. Gracey and T. Greenshaw (eds.), p. 318. (2000).

[16] ZEUS Coll., U. Holm (ed.), *The ZEUS Detector*. Status Report (unpublished), DESY (1993), available on `http://www-zeus.desy.de/bluebook/bluebook.html`.

[17] N. Harnew et al., Nucl. Inst. Meth. **A 279**, 290 (1989);
B. Foster et al., Nucl. Phys. Proc. Suppl. **B 32**, 181 (1993);
B. Foster et al., Nucl. Inst. Meth. **A 338**, 254 (1994).

[18] M. Derrick et al., Nucl. Inst. Meth. **A 309**, 77 (1991);
A. Andresen et al., Nucl. Inst. Meth. **A 309**, 101 (1991);
A. Caldwell et al., Nucl. Inst. Meth. **A 321**, 356 (1992);
A. Bernstein et al., Nucl. Inst. Meth. **A 336**, 23 (1993).

[19] J. Andruszków et al., Preprint DESY-92-066, DESY, 1992.

[20] C. Youngman, Preprint DESY-92-105, 1992;
W. H. Smith et al., *The ZEUS Trigger System* (unpublished). ZEUS-89-084, 1989.

[21] A. Kwiatkowski, H. Spiesberger and H.-J. Möhring, Comp. Phys. Comm. **69**, 155 (1992). Also in *Proc. Workshop Physics at HERA*, 1991, DESY, Hamburg.

[22] K. Charchula, G.A. Schuler and H. Spiesberger, Comp. Phys. Comm. **81**, 381 (1994).

[23] G. Gustafson and U. Petterson, Nucl. Phys. **B 306**, 746 (1988);
G. Gustafson, Phys. Lett. **B 175**, 453 (1986);
B. Andersson et al., Z. Phys. **C 43**, 625 (1989).

[24] L. Lönnblad, Comp. Phys. Comm. **71**, 15 (1992).

[25] M. Kuhlen, *QCD and the Hadronic Final State in Deep Inelastic Scattering at HERA*. Habilitation Thesis, Universität Hamburg, Report MPI-PHE-97-33, 1997.

[26] X. Artru and G. Mennessier, Nucl. Phys. **B 70**, 93 (1974);
M. G. Bowler, Z. Phys. **C 11**, 169 (1981);
B. Andersson et al., Z. Phys. **C 20**, 317 (1983);
B. Andersson et al., Nucl. Phys. **B 264**, 29 (1986);
B. Andersson et al., Phys. Rep. **97**, 31 (1983).

[27] T. Sjöstrand, Comp. Phys. Comm. **82**, 74 (1994);
T. Sjöstrand, Preprint CERN-TH.7112/93,LU TP 95-20, 1993.

[28] H.L. Lai et al., Phys. Rev. **D 55**, 1280 (1997).

[29] B. R. Webber, Nucl. Phys. **B 238**, 492 (1984).

[30] R. Brun, Preprint CERN-DD/EE 84-1, 1987.

[31] T. Haas, *Generating Monte Carlo events with MOZART* (unpublished). ZEUS-92-021, 1992.

[32] L. Wai and R. Yoshida, *Running MOZART-ZGANA-ZEPHYR chain using FUNNEL* (unpublished). ZEUS-93-065, 1993.

[33] ZEUS Coll., J. Breitweg et al., Eur. Phys. J. **C 6**, 43 (1999).

[34] H. Abramowicz, A. Caldwell and R. Sinkus, Nucl. Inst. Meth. **A 365**, 508 (1995);
R. Sinkus and T. Voss, Nucl. Inst. Meth. **A 391**, 360 (1997).

[35] S. Bentvelsen, J. Engelen and P. Kooijman, *Proc. Workshop on Physics at HERA*, W. Buchmüller and G. Ingelman (eds.), Vol. 1, p. 23. Hamburg, Germany, DESY (1992);
K.C. Höger, *Proc. Workshop on Physics at HERA*, W. Buchmüller and G. Ingelman (eds.), Vol. 1, p. 43. Hamburg, Germany, DESY (1992).

[36] F. Jacquet and A. Blondel, *Proceedings of the Study for an ep Facility for Europe*, U. Amaldi (ed.), p. 391. Hamburg, Germany (1979). Also in preprint DESY 79/48.

[37] S.Catani et al., Nucl. Phys. **B406**, 187 (1993).

[38] S.D. Ellis and D.E. Soper, Phys. Rev. **D 48**, 3160 (1993).

[39] T. Carli et al., *Proc. Workshop on Monte Carlo Generators for HERA Physics*, p. 329. DESY, Hamburg, Germany (1999). Also in preprint DESY-PROC-1999-02, available on `http://www.desy.de/~heramc/`.

[40] R. A. Fisher, Ann. Eug. **7**, 179 (1936).

[41] V. Blobel and E. Lohrmann, *Statistische und numerische Methoden der Datenanalyse*. Teubner, Stuttgart, Leipzig, 1998.

	r_I [%]	DATA	QCDINS	DJA, RGP	P_S	HERWIG	P_S
1D	9.1	252 ± 16	54.6 ± 0.9	340.4 ± 9.4	24	184.6 ± 7.9	45
$t > 9.5$	10.3	179 ± 13	61.3 ± 0.9	261.4 ± 8.3	36	227.4 ± 8.7	41
$t > 10.0$	5.6	76 ± 8.7	33.5 ± 0.7	104.7 ± 5.2	49	89.0 ± 5.5	58
$t > 10.5$	2.8	33 ± 5.7	16.7 ± 0.5	34.8 ± 3.0	73	35.1 ± 3.4	73

Table 1: *Numbers of events within instanton enhancing cuts chosen such that a fraction $r_I \approx 10\%$ of the QCDINS sample within offline cuts is kept. For the Fisher discriminant t, harder cuts are also tried, yielding a higher separation power P_S. The normal DIS MC is normalised such that the number of events in the inclusive sample agrees with the data, cf. section 5.2. Statistical errors are given. "DJA, RGP" stands for the combination of DJANGOH and RAPGAP normal DIS Monte Carlos. "1D" corresponds to the combination of one-dimensional cuts and "Fisher" to the cut on the Fisher discriminant.*

	$f_{\mathrm{I,th}}$	fit minimum		2σ upper limits on f_I	
		Hwg	DjaRgp	Hwg	DjaRgp
1D	0.088	-0.003	-0.051	0.082	0.045
F.	0.23	0.043	-0.045	0.189	0.116

Table 2: *Results of fits of the sphericity S of the instanton part of the event within different sets of instanton enhancing cuts before systematic studies. Fit minima and limits on the instanton fraction f_I are given and compared to the predicted fraction $f_{\mathrm{I,th}}$. "1D" corresponds to the combination of one-dimensional cuts, and "F." to cuts on the Fisher discriminant, and $Q'^2_{\mathrm{DA}} < 250\,GeV^2$. "Hwg" stands for HERWIG and "DjaRgp" for DJANGOH with RAPGAP.*

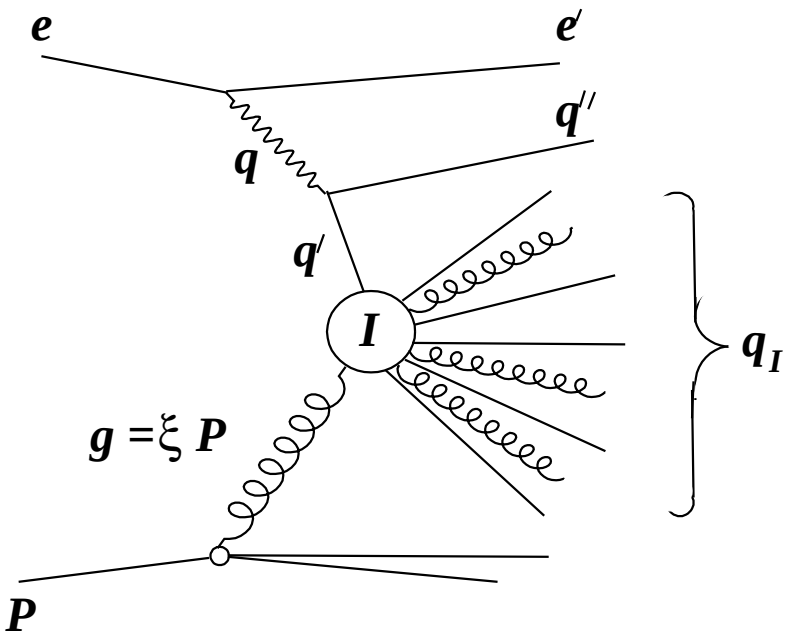

Figure 1: *Kinematics of instanton-induced events at HERA.*

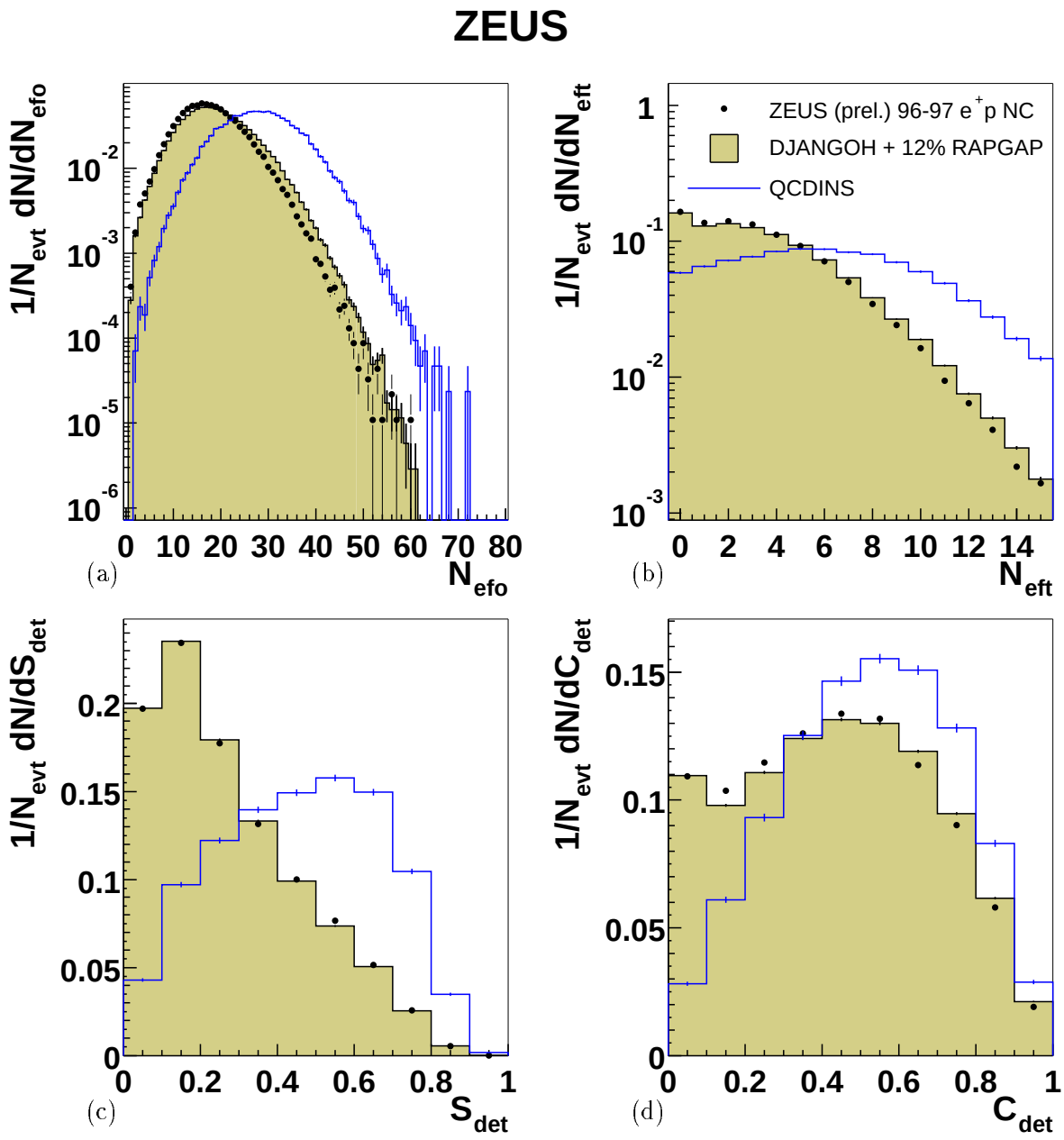

Figure 2: *Distributions of variables calculated from the instanton region, that are used as input for the Fisher discriminant, shown for the inclusive sample: (a): efo multiplicity $N_{\rm efo}$, (b): efo track multiplicity $N_{\rm eft}$, (c): sphericity S, reconstructed in the instanton cms, (d): circularity C, reconstructed in the hcms.*

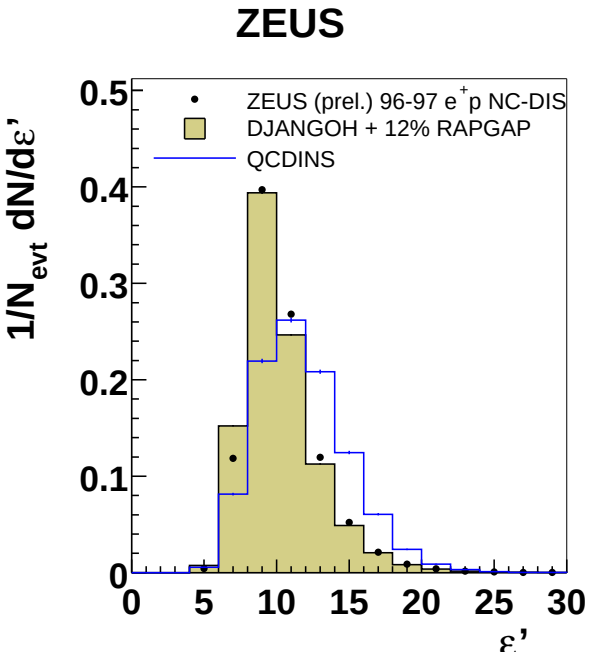

Figure 3: *Distribution of the shape variable ϵ'.*

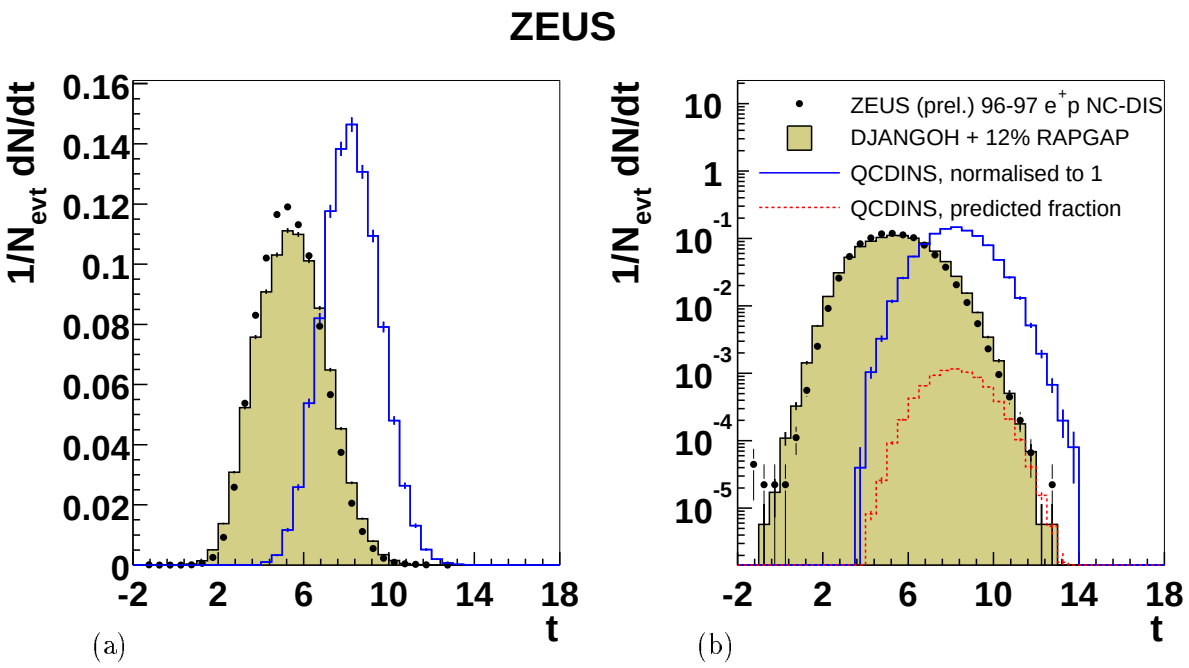

Figure 4: *Fisher discriminant t calculated from the variables S, C, $\log_{10} p_T^{\mathrm{Jet}}$, N_{efo}, N_{eft} and ϵ'. Shown are (a) linear and (b) logarithmic plots for the inclusive sample with an additional cut $Q'^2_{\mathrm{DA}} < 250\ GeV^2$. In (b), the QCDINS distribution, normalised to the predicted fraction of 0.79%, is also shown.*

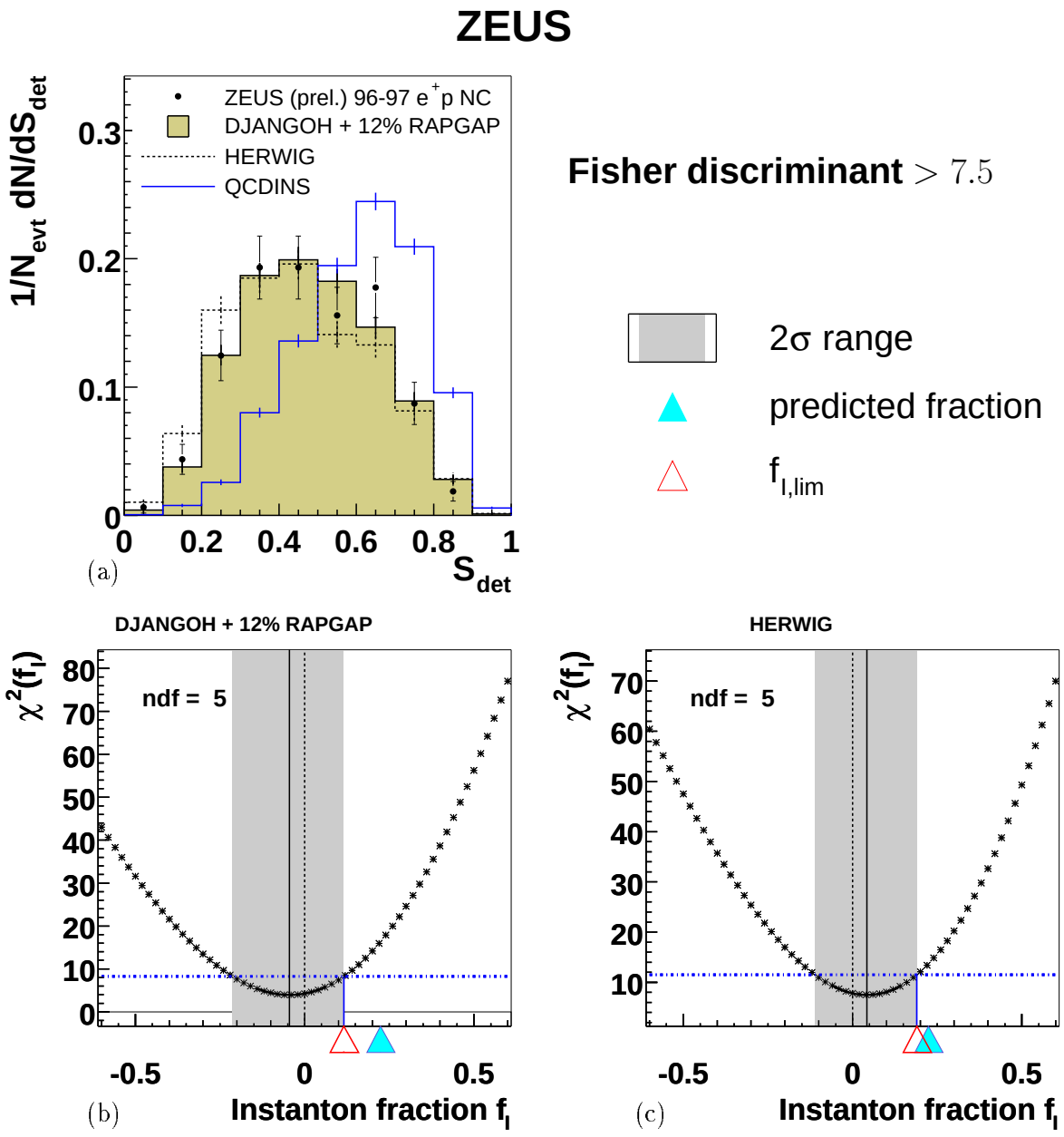

Figure 5: *Sphericity S of the instanton part of the hadronic final state in its CM system for the instanton-enhanced sample, obtained from a cut on a Fisher discriminant based on 5 input variables, a cut $Q'^2_{\mathrm{DA}} < 250\ GeV^2$ and a cut $S > 0.2$. (The cut on S itself is not applied in (a), but the first two bins are not used in the fit). The lower plots show the χ^2 value obtained from a fit of sphericity as function of the instanton fraction in the sample for (b) DJANGOH with RAPGAP and (c) HERWIG.*