H1prelim-14-041
ZEUS-prel-14-005
April 24, 2014

Combined Measurement of Inclusive $e^{\pm}p$ Scattering Cross Sections at HERA

H1 and ZEUS Collaborations

Abstract

A combination is presented of all inclusive deep inelastic cross sections measured by the H1 and ZEUS collaborations in neutral and charged current unpolarised $e^{\pm}p$ scattering at HERA. The data correspond to a luminosity of about $1\,\mathrm{fb}^{-1}$ and span six orders of magnitude in negative four-momentum-transfer squared, Q^2, and Bjorken x. They include data taken at proton beam energies of 920, 820, 575 and 460 GeV. The combination method used takes the correlations of systematic uncertainties into account, resulting in improved accuracy.

.

1 Introduction

Deep inelastic scattering (DIS) of electrons[1] on protons at HERA has been central to the exploration of proton structure and quark–gluon interaction dynamics as prescribed by perturbative Quantum Chromodynamics (QCD). HERA operated at a centre-of-mass energy of up to $\sqrt{s} \simeq 320\,\text{GeV}$. This enabled the two collaborations, H1 and ZEUS, to explore a large phase space in x and Q^2. The kinematic range for neutral current (NC) interactions was $0.045 \leq Q^2 \leq 50000\,\text{GeV}^2$ and $6 \cdot 10^{-7} \leq x$ for values of the inelasticity, $y = Q^2/(sx)$, between 0.005 and 0.95. The kinematic range for charged current (CC) interactions was $200 \leq Q^2 \leq 50000\,\text{GeV}^2$ and $1.3 \cdot 10^{-2} \leq x \leq 0.40$ for values of y between 0.037 and 0.76.

HERA was operated in two phases: HERA I, from 1992–2000, and HERA II, from 2002–2007. It was always operated with an electron beam energy of $E_e \simeq 27.5\,\text{GeV}$. For most of HERA I and II, the proton beam energy was $E_p = 920\,\text{GeV}$, resulting in the highest centre-of-mass energy of $\sqrt{s} \simeq 320\,\text{GeV}$. The total luminosity collected by both H1 and ZEUS was aproximately $500\,\text{pb}^{-1}$, divided about equally between e^+p and ep scattering. In HERA I, each experiment collected about $100\,\text{pb}^{-1}$ of e^+p and $15\,\text{pb}^{-1}$ of e^-p data. The HERA I data was the basis of a combination published previously [**?**]. The paper presented now is based on the combination of all published H1 [**? ? ? ? ? ? ? ?**] and ZEUS [**? ? ? ? ? ? ? ? ? ? ? ? ? ?**] measurements from both HERA I and II on inclusive DIS in NC and CC reactions. This includes data taken at $\sqrt{s} = 319$, 301, 252 and 255 GeV, corresponding to proton beam energies of $E_p = 920$, 920, 575 and 460 GeV. The HERA II measurements were made with polarised beams, but individually averaged to obtain cross sections for unpolarised beams used as inputs to the combination.

The combination was performed using the packages HERAverager [**? ?**] and HERAfitter [**? ?**]. It is based on a method introduced in [**?**] and extended in [**?**]. HERAverager not only combines data, but also provides a model-independent check of the consistency of the data. The correlated systematic uncertainties and global normalisations are averaged such that one coherent data set is obtained. Since H1 and ZEUS have employed different experimental techniques, using different detectors and methods of kinematic reconstruction, the combination leads to a significantly reduced uncertainty.

Analyses of the x and Q^2 dependences of the NC and CC DIS cross sections measured at HERA have determined sets of quark and gluon momentum distributions in the proton, both from H1 [**?**] and ZEUS [**?**] and from the combined HERA I inclusive data [**?**]. In such analyses, the lower-Q^2 NC data constrain the low-x sea quark and gluon distributions. The high-Q^2 CC data, together with the difference between NC e^+p and e^-p cross sections at high Q^2, constrain the valence quark distributions. The use of the HERA CC data allows the down quark distribution in the proton to be determined without assuming isospin symmetry. In addition, the use of HERA data alone for the determination of parton distribution functions (PDFs) eliminates the need for heavy target corrections, which must be applied to DIS data from nuclear targets. The new combined HERA data were used to determine a new set of parton distributions termed HERAPDF2.0. Consistency of the input data allowed the experimental uncertainty of the HERAPDF2.0 set to be determined using rigorous statistical methods. Uncertainties resulting from model assumptions and from the choice of PDF parametrisation were also considered.

[1] In this paper, the word electron refers to both electrons and positrons, unless otherwise stated.

2 Cross Sections and Parton Distributions

The NC deep inelastic $e^{\pm}p$ scattering cross sections are at tree level given by a linear combination of generalised structure functions. For unpolarised beams, they can be expressed as

$$\sigma^{\pm}_{r,\mathrm{NC}} = \frac{\mathrm{d}^2\sigma^{e^{\pm}p}_{\mathrm{NC}}}{\mathrm{d}x\mathrm{d}Q^2} \cdot \frac{Q^4 x}{2\pi\alpha^2 Y_+} = \tilde{F}_2 \mp \frac{Y_-}{Y_+} x\tilde{F}_3 - \frac{y^2}{Y_+}\tilde{F}_L \,, \tag{1}$$

where the electromagnetic coupling constant, α, the photon propagator and a helicity factor are absorbed in the definitions of $\sigma^{\pm}_{r,\mathrm{NC}}$ and $Y_{\pm} = 1 \pm (1-y)^2$. The structure functions, $\tilde{F}_2$, $\tilde{F}_L$ and $x\tilde{F}_3$, depend on the electroweak parameters as [**?**]

$$\begin{aligned} \tilde{F}_2 &= F_2 - \kappa_Z v_e \cdot F_2^{\gamma Z} + \kappa_Z^2 (v_e^2 + a_e^2) \cdot F_2^Z \,, \\ \tilde{F}_L &= F_L - \kappa_Z v_e \cdot F_L^{\gamma Z} + \kappa_Z^2 (v_e^2 + a_e^2) \cdot F_L^Z \,, \\ x\tilde{F}_3 &= \kappa_Z a_e \cdot xF_3^{\gamma Z} - \kappa_Z^2 \cdot 2 v_e a_e \cdot xF_3^Z \,, \end{aligned} \tag{2}$$

where v_e and a_e are the vector and axial-vector weak couplings of the electron to the Z boson, and $\kappa_Z(Q^2) = Q^2/[(Q^2 + M_Z^2)(4\sin^2\theta_W \cos^2\theta_W)]$. In HERAfitter, the values of $\sin^2\theta_W = 0.2315$ and $M_Z = 91.187\,\mathrm{GeV}$ were used for the electroweak mixing angle and the Z boson mass.

At low Q^2, the contribution of Z exchange is negligible and

$$\sigma^{\pm}_{r,\mathrm{NC}} = \tilde{F}_2 - y^2 \tilde{F}_L / Y_+ \,. \tag{3}$$

The contribution of the term containing the longitudinal structure function $\tilde{F}_L$ is only significant for large values of y.

In the Quark Parton Model (QPM), gluons are not present and $\tilde{F}_L = 0$ [**?**]. The other functions in equation 2 become

$$\begin{aligned} (F_2, F_2^{\gamma Z}, F_2^Z) &= [(e_u^2, 2e_u v_u, v_u^2 + a_u^2)(xU + x\overline{U}) + (e_d^2, 2e_d v_d, v_d^2 + a_d^2)(xD + x\overline{D})] \,, \\ (xF_3^{\gamma Z}, xF_3^Z) &= 2[(e_u a_u, v_u a_u)(xU - x\overline{U}) + (e_d a_d, v_d a_d)(xD - x\overline{D})] \,, \end{aligned} \tag{4}$$

where e_u and e_d denote the electric charge of up- or down-type quarks while $v_{u,d}$ and $a_{u,d}$ are the vector and axial-vector weak couplings of the up- or down-type quarks to the Z boson. The terms xU, xD, $x\overline{U}$ and $x\overline{D}$ denote the sums of parton distributions for up-type and down-type quarks and anti-quark, respectively. Below the b quark mass threshold, these sums are related to the quark distributions as follows

$$xU = xu + xc \,, \qquad x\overline{U} = x\bar{u} + x\bar{c} \,, \qquad xD = xd + xs \,, \qquad x\overline{D} = x\bar{d} + x\bar{s} \,, \tag{5}$$

where xs and xc are the strange and charm quark distributions. Assuming symmetry between the quarks and anti-quarks in the sea, the valence quark distributions can be expressed as

$$xu_v = xU - x\overline{U} \,, \qquad xd_v = xD - x\overline{D} \,. \tag{6}$$

The reduced cross sections for inclusive unpolarised CC $e^{\pm}p$ scattering are defined as

$$\sigma^{\pm}_{r,\mathrm{CC}} = \frac{2\pi x}{G_F^2} \left[\frac{M_W^2 + Q^2}{M_W^2}\right]^2 \frac{\mathrm{d}^2\sigma^{e^{\pm}p}_{\mathrm{CC}}}{\mathrm{d}x\mathrm{d}Q^2} \,. \tag{7}$$

In HERAfitter the values of $G_F = 1.16637 \times 10^{-5}$ GeV^{-2} and $M_W = 80.41$ GeV were used for the Fermi constant and W-boson mass. In analogy to equation 1, CC structure functions are defined such that

$$\sigma^{\pm}_{r,CC} = \frac{Y_+}{2} W_2^{\pm} \mp \frac{Y_-}{2} x W_3^{\pm} - \frac{y^2}{2} W_L^{\pm} . \tag{8}$$

In the QPM, $W_L^{\pm} = 0$ and $W_2^{\pm}$, $xW_3^{\pm}$ represent sums and differences of quark and anti-quark distributions, depending on the charge of the lepton beam:

$$W_2^+ = x\overline{U} + xD , \qquad xW_3^+ = xD - x\overline{U} , \qquad W_2^- = xU + x\overline{D} , \qquad xW_3^- = xU - x\overline{D} . \tag{9}$$

From these equations, it follows that

$$\sigma^+_{r,\mathrm{CC}} = x\overline{U} + (1-y)^2 xD , \qquad \sigma^-_{r,\mathrm{CC}} = xU + (1-y)^2 x\overline{D} . \tag{10}$$

The combination of NC and CC measurements makes it possible to determine both the combined sea quark distribution functions, $x\overline{U}$ and $x\overline{D}$, and the valence quark distributions, xu_v and xd_v.

3 Measurements of Inclusive DIS Cross Sections

3.1 Detectors

The H1 [**?** **?** **?**] and ZEUS [**?**] detectors both had nearly 4π hermetic coverage[2]. They were built following similar physics considerations but the collaborations opted for different technical solutions, both for the calorimeters and the tracking detectors. The luminosity upgrade for HERA II made significant changes in both detectors necessary. The HERA machine had to be extended into the experimental area with final-focus magnets inside the detectors. This required some detector elements to be retracted. As a result, the acceptance for low-Q^2 events was reduced.

The most relevant components of the H1 detector for these measurements were the liquid argon calorimeter (LAr), in the polar angular range $4^\circ < \theta < 154^\circ$, the backward lead-scintillator calorimeter (SpaCal) with a coverage of $153^\circ < \theta < 177^\circ$ and the inner tracking system enclosed by the two calorimeters.

Each of the calorimeters had an inner electromagnetic and an outer hadronic part. Depending on the polar angle the thickness of the LAr's electromagnetic section varied between 20 and 30 radiation lengths and that of the hadronic sections ranged from 4.5 to 8 nuclear interaction lengths. For the SpaCal the corresponding figures were 27.5 radiation lengths and 2 nuclear interaction lengths, respectively. The relative energy resolutions σ_E, as measured with test beams,

[2]Both experiments used a right-handed Cartesian coordinate system, with the Z axis pointing in the proton beam direction, referred to as the forward direction, and the X axis pointing towards the centre of HERA. The coordinate origins were at the nominal interaction points. The pseudorapidity was defined as $\eta = -ln(tan(\theta/2))$, where the polar angle, θ, was measured with respect to the proton beam direction.

are $\sigma_E \simeq 0.11/\sqrt{E/\mathrm{GeV}} \oplus 0.01$ (LAr) and $\sigma_E \simeq 0.07/\sqrt{E/\mathrm{GeV}} \oplus 0.01$ (SpaCal) for electromagnetic particles and $\sigma_E \simeq 0.50/\sqrt{E/\mathrm{GeV}} \oplus 0.02$ (LAr) and $\sigma_E \simeq 0.70/\sqrt{E/\mathrm{GeV}} \oplus 0.01$ (SpaCal) for hadronic energy deposites. The LAr was surrounded by a superconducting coil providing a solenoidal magnetic field of 1.16 T to enable the momentum measurement of charged particles passing the inner tracking system. The instrumented iron return yoke of the solenoid was used for measuring the energy leakage of high energetic hadronic showers in the LAr and for muon detection.

The inner tracking system consisted of the central tracking detector (CTD), the forward tracking detector (FTD) and the backward drift chamber (BDC), which was replaced by the backward proportional chamber (BPC) for the HERA-II running period. The CTD measured the trajectories of charged particles in two cylindric drift chambers (CJC). A further drift chamber (COZ) between the two drift chambers of the CJC improves the z coordinate reconstruction. During the HERA-I running period an additional drift chamber (CIZ) attached to the inner wall of the inner CJC was used for the same purpose. Sets of multiwire proportional chambers between the inner CJC and the beam line (CIP) and between the two CJCs (COP) served mainly for trigger purposes. The components of the inner tracking system closest to the ep interaction point were a set of silicon detectors: the central silicon tracker (CST) and the backward silicon tracker (BST) which were supplemented by the forward silicon tracker (FST) during the HERA-II running period. The CTD measures charged particle trajectories in the polar angular range $15^\circ < \theta < 165^\circ$ with a transverse momentum resolution of $\sigma(p_T)/p_T \simeq 0.002\, p_T/\mathrm{GeV} \oplus 0.015\%$ for particles passing both CJCs. The FTD consisted of a set of drift chamber modules of different orientation. It mainly served for improving the measurement of the hadronic final state. The BDC/BPC in front of the SpaCal improved the x and y reconstruction of the position of electromagnetic showers in the SpaCal.

The main component of the ZEUS detector [**?**] was a compensating uranium–scintillator calorimeter (CAL) [**?**] consisting of three parts: forward (FCAL), barrel (BCAL) and rear (RCAL). Each part was segmented into one electromagnetic section (EMC) and either one (in RCAL) or two (in BCAL and FCAL) hadronic sections (HAC). Under test-beam conditions, the energy resolutions were $0.18\sqrt{E/\mathrm{GeV}}$ and $0.35\sqrt{E/\mathrm{GeV}}$ for the EMC and HAC sections, respectively. The timing resolution of the CAL was ~ 1 ns for energy deposits greater than 4.5 GeV. A Scintillator-tile presampler was mounted in front of the CAL. The RCAL was instrumented at a depth of 3.3 radiation lengths with a silicon-pad hadron-electron separator (HES).

Charged particles were tracked in the central tracking detector (CTD) [**?**] which operated in a magnetic field of 1.43 T provided by a thin superconducting solenoid, positioned between the BCAL and the presampler. The CTD consisted of 72 cylindrical drift chamber layers, organised in nine superlayers covering the polar-angle region $15^\circ < \theta < 164^\circ$. Planar drift chambers provided additional tracking in the forward and rear directions. The small angle rear tracking detector (SRTD), consisting of two orthogonal planes of scintillator strips, was used to measure electrons at large θ_e. In HERA II, the drift chamber based tracking detectors were complemented by a silicon microvertex detector (MVD) [**?**], consisting of three active layers in the barrel and four disks in the forward region. For CTD–MVD tracks that passed through all nine CTD superlayers, the transverse momnentum resolution was $\sigma(p_T)/p_T = 0.0029 p_T \oplus 0.0081 \oplus 0.0012/p_T$, with p_T in GeV. In HERA I, the angular coverage in the electron beam direction was extended with a tungsten–scintillator calorimeter (BPC)[**?**], located behind the RCAL at Z = -294 cm

close to the beam axis, and a silicon microstrip tracking device (BPT) [**?**] installed in front of the BPC.

Both experiments measured the luminosity using the Bethe–Heitler reaction $ep \to e\gamma p$. In HERA I, H1 and ZEUS both had photon taggers positioned about 100 m down the electron beam line and achieved accuracies on the luminosity measurement of about 1–2% for this period.

For the HERA II period, both H1 [**?**] and ZEUS upgraded their luminosity detectors. The ZEUS luminosity detector consisted of independent leadscintillator calorimeter [**?** **?**] and magnetic spectrometer [**?**] systems. The fractional systematic uncertainty on the measured luminosity for ZEUS was 1.8% for most of the HERA II period. H1 determined the overall normalisation for the HERA II measurements [**?**] using a precision measurement of the QED Compton process [**?**].

3.2 Reconstruction of Kinematics

The deep inelastic ep scattering cross sections of the inclusive neutral and charged current reactions depend on the centre-of-mass energy, $\sqrt{s}$, and on the two kinematic variables, Q^2 and x. Usually, x is obtained from the measurement of the inelasticity, y, and from Q^2 and s through the relationship $x = Q^2/(sy)$. The specialty of the HERA collider experiments is the ability to determine the NC event kinematics from the scattered electron, e, or from the hadronic final state, h, or from a combination of the two. The choice of the most appropriate kinematic reconstruction method for a given phase space region is based on resolution, possible biases of the measurements and effects due to initial or final state radiation. The optimisation led to different choices for the two experiments. The usage of different reconstruction techniques contributes to an improved accuracy when combining data sets.

The "electron method" is applied on NC scattering events. The quantities y and Q^2 are calculated using only the variables measuered for the scattered electron:

$$y_e = 1 - \frac{\Sigma_e}{2E_e} \ , \qquad Q^2_e = \frac{P^2_{T,e}}{1 - y_e} \ , \qquad x_e = \frac{Q^2_e}{sy_e} \ , \tag{11}$$

where $\Sigma_e = E'_e(1 - \cos\theta_e)$, E'_e is the energy of the scattered electron, θ_e is its angle with respect to the proton beam, and $P_{T,e}$ is its transverse momentum.

For CC scattering, the reconstruction of the hadronic final state, h yields similar relations [**?**]:

$$y_h = \frac{\Sigma_h}{2E_e} \ , \qquad Q^2_h = \frac{P^2_{T,h}}{1 - y_h} \ , \qquad x_h = \frac{Q^2_h}{sy_h} \ , \tag{12}$$

where $\Sigma_h = (E - P_Z)_h = \sum_i (E_i - p_{Z,i})$ is the hadronic $E - P_Z$ variable with the sum extending over the reconstructed hadronic final state particles, i, and $P_{T,h} = \left|\sum_i \boldsymbol{p}_{\perp,i}\right|$ is the total transverse momentum of the hadronic final state with $\boldsymbol{p}_{\perp,i}$ being the transverse momentum vector of the particle i. The hadronic scattering angle, θ_h, is computed as

$$\tan\frac{\theta_h}{2} = \frac{\Sigma_h}{P_{T,h}} \ , \tag{13}$$

which, within the QPM, corresponds to the direction of the struck quark.

In the "sigma method" [**?**], the total $E - P_Z$ variable,

$$E - P_Z = E'_e(1 - \cos\theta_e) + \sum_i (E_i - p_{Z,i}) = \Sigma_e + \Sigma_h \ , \tag{14}$$

is introduced. For events without initial or final state radiation, $E - P_Z = 2E_e$. Thus, equations 11 and 12 become

$$y_\Sigma = \frac{\Sigma_h}{E - P_Z} \ , \qquad Q^2_\Sigma = \frac{P^2_{T,e}}{1 - y_\Sigma} \ , \qquad x_\Sigma = \frac{Q^2_\Sigma}{s y_\Sigma} \ . \tag{15}$$

An extension of the sigma method [**?** **?**] provides

$$y_{\Sigma'} = y_\Sigma \ , \qquad Q^2_{\Sigma'} = Q^2_\Sigma \ , \qquad x_{\Sigma'} = \frac{Q^2_\Sigma}{2E_p(E - P_z)y_\Sigma} = \frac{Q^2_\Sigma}{2E_p\Sigma_h} \ . \tag{16}$$

This modification takes radiation at the lepton vertex into account by replacing the electron beam energy in the calculation of $x_{\Sigma'}$ in a way similar to its replacement in the calculation of y_Σ.

In the hybrid "e-sigma method" [**?** **?** **?**], Q^2_e and x_Σ are used to reconstruct the event kinematics as

$$y_{e\Sigma} = \frac{Q^2_e}{s x_\Sigma} = \frac{2E_e}{E - P_Z} y_\Sigma \ , \qquad Q^2_{e\Sigma} = Q^2_e \ , \qquad x_{e\Sigma} = x_\Sigma \ . \tag{17}$$

The "double angle method" [**?** **?**] is used to reconstruct Q^2 and x from the electron and hadronic scattering angles as

$$y_{DA} = \frac{\tan(\theta_h/2)}{\tan(\theta_e/2) + \tan(\theta_h/2)} \ , \qquad Q^2_{DA} = 4E^2_e \cdot \frac{\cot(\theta_e/2)}{\tan(\theta_e/2) + \tan(\theta_h/2)} \ , \qquad x_{DA} = \frac{Q^2_{DA}}{s y_{DA}} \ . \tag{18}$$

This method is largely insensitive to hadronisation effects. To first order, it is also independent of the detector energy scales. However, the hadronic angle is not as well determined as electron angle due to particle loss in the beampipe.

In the "PT method" of reconstruction [**?**], the well-measured electron variables are used to obtain a good event-by-event estimate of the loss of hadronic energy by employing $\delta_{PT} = P_{T,h}/P_{T,e}$. This improves both the resolution and uncertainties on the reconstructed y and Q^2. The PT method uses all measured variables to optimise the resolution over the entire kinematic range measured. A variable θ_{PT} is introduced as

$$\tan\frac{\theta_{PT}}{2} = \frac{\Sigma_{PT}}{P_{T,e}} \ , \quad \text{where} \quad \Sigma_{PT} = 2E_e \frac{C(\theta_h, P_{T,h}, \delta_{PT}) \cdot \Sigma_h}{\Sigma_e + C(\theta_h, P_{T,h}, \delta_{PT}) \cdot \Sigma_h} \ . \tag{19}$$

The variable θ_{PT} is then substituted for θ_h in the formulae for the double angle method to determine x, y and Q^2. The detector-specific function, C, is calculated using Monte Carlo simulations as $\Sigma_{\text{true,h}}/\Sigma_h$, depending on θ_h, $P_{T,h}$ and δ_{PT}.

The methods of the kinematic reconstruction used by H1 and ZEUS for the individual data sets is given in Table 1 as part of their specification.

3.3 Data Samples

A summary of the 41 data sets used in the combination is given in Table 1. HERA was always operated with an electron beam energy of $E_e \simeq 27.5\,\text{GeV}$. In the first years, until 1997, the proton beam energy, E_p, was set to 820 GeV. In 1998 it was increased to 920 GeV. In 2007, it was lowered to 575 GeV and 460 GeV.

The very low-Q^2 region is covered by data from HERA I. The lowest, $Q^2 \geq 0.045\ \text{GeV}^2$, data come from the measurements of ZEUS using the BPC and BPT. The Q^2 range from 0.2 GeV^2 to 1.5 GeV^2 is covered using special HERA I runs, in which the interaction vertex position was shifted forward, bringing backward scattered electrons with larger angles into the acceptance of the detectors [**?** **?** **?**]. The lowest Q^2 for the shifted-vertex data was reached using events, in which the electron energy was reduced by initial state radiation [**?**].

The $Q^2 \geq 1.5\ \text{GeV}^2$ range was covered by HERA I and HERA II data in various configurations. The high-Q^2 data from HERA I were kept as in the previously published combination[**?**]. However, for high Q^2, the high statistics data from HERA II were essential, especially for $e^- p$ scattering, where the integrated luminosity for HERA I was very limited.

The 2007 data with lowered proton energies[**?** **?** **?**] were included in the combination and provide data with reduced $\sqrt{s}$ and Q^2 up to $800\,\text{GeV}^2$.

4 Combination of the Measurements

The combination of the data was performed with the HERAverager [**?** **?**] and HERAfitter [**?** **?**] tools.

4.1 Averaging Data Points

The averaging of the data points was performed using the HERAverager [**?** **?**] tool which is based on a χ^2 minimisation method [**?**]. This method assumes that there is one and only one correct value for the cross section of each process at each point of the phase space. These values are estimated by optimising a vector, $\boldsymbol{m}$. The χ^2 function used takes into account the correlated and uncorrelated systematic uncertainties[3] of the H1 and ZEUS cross-section measurements and allows for shifts of the data to accomodate the correlated uncertainties. For a single data set, ds, the χ^2 is defined as

$$\chi^2_{\text{exp},ds}(\boldsymbol{m},\boldsymbol{b}) = \sum_{i,ds} + \sum_{j,b} = \sum_i \frac{\left[m^i - \sum_j \gamma^i_j m^i b_j - \mu^i\right]^2}{\delta^2_{i,\text{stat}}\,\mu^i\left(m^i - \sum_j \gamma^i_j m^i b_j\right) + \left(\delta_{i,\text{uncor}}\, m^i\right)^2} + \sum_j b_j^2\,, \tag{20}$$

where μ^i is the measured value at a point, i, and γ^i_j, $\delta_{i,\text{stat}}$ and $\delta_{i,\text{uncor}}$ are the relative correlated systematic, relative statistical and relative uncorrelated systematic uncertainties, respectively.

[3] The original double-differential cross-section measurements were published with their statistical and systematic uncertainties. The systematic uncertainties were classified as either point-to-point correlated or point-to-point uncorrelated.

For the reduced cross-section measurements, $\mu^i = \sigma_r^i$, i runs over all point on the (x, Q^2) plane for which a measurement exists in ds. The vector $\boldsymbol{b}$ represents the shifts with respect to the correlated systematic uncertainties; the summations over j extends over all correlated systematic uncertainties.

Equation 20 takes into account that the quoted uncertainties are based on measured cross sections, which are subject to statistical fluctuations. Under the assumptions, that the statistical uncertainties are proportional to the square root of the number of events and that the systematic uncertainties are proportional to $\boldsymbol{m}$, the minimisation of $\chi^2_{\text{exp},ds}$ from equation 20 with respect to $\boldsymbol{m}$ provides an unbiased estimator of the true values.

The leading systematic uncertainties on the cross-section measurements used for the combination arose from the uncertainties on the acceptance corrections and luminosity determinations. This indicates that both the correlated and uncorrelated systematic uncertainties are of multiplicative nature, i.e. they increase proportionally to the central values. In equation 20, the multiplicative nature of these uncertainties is taken into account by multiplying the relative errors γ_j^i and $\delta_{i,\text{uncor}}$ by the estimate m^i. For the inclusive DIS cross-section measurements, the background contributions were small and thus, it is justified to take the square root of the number of events used to determine σ_r^i as the statitical uncertainty. The expected number of events is calculated from the estimator $\boldsymbol{m}$. Corrections due to the shifts allowed to accomodate the correlated systematic uncertainties are introduced through the term $\delta^2_{i,\text{stat}}\,\mu^i\left(m^i - \sum_j \gamma_j^i m^i b_j\right)$.

For the combination of several datasets, a total χ^2 function is defined as:

$$\chi^2_{\text{tot}} = \sum_{ds}\sum_{i,ds} + \sum_{j,b}\ , \tag{21}$$

with $\sum_{i,ds}$ and $\sum_{j,b}$ as introduced in equation 20. The averaging of the data is performed such that equation 21 takes a form similar to equation 20 and $\boldsymbol{m}$ again is an estimator of the true cross sections:

$$\chi^2_{\text{tot}}\left(\boldsymbol{m}, \boldsymbol{b'}\right) = \chi^2_{\text{min}} + \sum_{i=1}^{N_M} \frac{\left[m^i - \sum_j \gamma_j^{i,\text{ave}} m^i b'_j - \mu^{i,\text{ave}}\right]^2}{\delta^2_{i,\text{ave,stat}}\,\mu^{i,\text{ave}}\left(m^i - \sum_j \gamma_j^{i,\text{ave}} m^i b'_j\right) + \left(\delta_{i,\text{ave,uncor}}\, m^i\right)^2} + \sum_j (b'_j)^2\ , \tag{22}$$

where $\mu^{i,\text{ave}}$ is the average value at point, i, and $\gamma_j^{i,\text{ave}}$, $\delta_{i,\text{ave,stat}}$ and $\delta_{i,\text{ave,uncor}}$ are its relative correlated systematic, relative statistical and relative uncorrelated systematic uncertainties, respectively. The value of χ^2_{min} corresponds to the minimum of equation 21. The ratio $\chi^2_{\text{min}}/n_{\text{dof}}$ is a measure of the consistency of the data sets. The number of degrees of freedom, n_{dof}, is calculated as the difference between the total number of measurements and the number of averaged points, N_M. The systematic uncertainties b'_j are obtained from the original shifts, b_j, by an orthogonal transformation [**?**].

Some of the measurements were originally reported with asymmetric systematic uncertainties. They were symmetrised by the collaborations before entering the combination procedure. The comination was found to be insensitive to the details of the symmetrisation procedures[**?**]. An overall normalisation uncertainty of 0.5% due to uncertainties on higher order corrections to the Bethe-Heitler process was assumed for all data sets which were normalised with data from the luminosity monitors.

The experimental uncertainties which are treated as point-to-point correlated uncertainties γ_i^j may be common for CC and NC data as well as for several data sets of the same experiment. A full table of the correlations of the systematic uncertainties across the data sets can be found elsewhere [**?**]. The systematic uncertainties are treated as independent between H1 and ZEUS. All the NC and CC cross-section data from H1 and ZEUS are combined in one simultaneous minimisation. Therefore resulting shifts of the correlated systematic uncertainties propagate coherently to both CC and NC data.

4.2 Common $\sqrt{s}$-Values and (x, Q^2)-Grids

The data were taken at several $\sqrt{s}$ and the cross sections were published for different (x, Q^2) grids. In order to average a set of data points, the points have to be translated to a common $\sqrt{s}_{\rm com}$ and a common $(x_{\rm grid}, Q^2_{\rm grid})$. The translation requires the ratio of the double differential cross sections at $(x_{\rm grid}, Q^2_{\rm grid})$ and (x, Q^2). The determination of these ratios is described in the next section. Here, the choice of $\sqrt{s}_{\rm com}$ and grid points is decribed.

Three common center-of-mass values, $\sqrt{s}_{{\rm com},i}$, with $\sqrt{s}_{\rm com,1} = 318\,{\rm GeV}$ ($E_p = 820\,{\rm GeV}$ and $E_p = 920\,{\rm GeV}$), $\sqrt{s}_{\rm com,2} = 252\,{\rm GeV}$ ($E_p = 575\,{\rm GeV}$) $\sqrt{s}_{\rm com,3} = 225\,{\rm GeV}$ ($E_p = 460\,{\rm GeV}$) were chosen to combine data. An exception was made for data with $E_p = 820\,{\rm GeV}$ or $E_p = 920\,{\rm GeV}$ which were not translated to $\sqrt{s}_{\rm com,1}$ if $y \geq 0.35$. Such data were kept separately at $\sqrt{s}$ =301 and 319 GeV, respectively.

Two common $(x_{\rm grid}, Q^2_{\rm grid})$ grids were chosen, one for data at $\sqrt{s}_{\rm com,1}$ and one for data at $\sqrt{s}_{\rm com,2}$ and $\sqrt{s}_{\rm com,3}$. The two grids have a different structure in y such that the translation corrections are minimised. Figure 1 depicts the grids. For a given data point with $\sqrt{s}_{\rm com,2}$ or $\sqrt{s}_{\rm com,3}$, the grid point was chosen such that it is closest in Q^2 and y. For a given data point with $\sqrt{s}_{\rm com,1}$, the grid point was chosen such that it is closest in Q^2 and x.[4]

Over most of the phase space, it was ensured that separate measurements from the same data set were not translated to the same grid point. Only 9 (8) grid points accumulated two (three) points from the same dataset. Up to 10 datasets were available for a given process. The vast majority of grid points accumulated data from both H1 and ZEUS measurements, in many cases six measurements from six different datasets. However, there are grid points where only one measurement was available. It should be noted that in these cases, the combination procedure nevertheless introduced shifts with respect to the original measurements due to the correlation of systematic uncertainties.

4.3 Combination Procedure

The combination procedure is iterative. Each iteration has a first step, in which the data are translated to the common $\sqrt{s}$ values and (x, Q^2) grids and a second step, in which they are averaged.

[4]The grid points closest in y were chosen for data points from $\sqrt{s}_{\rm com,1}$ datasets marked with ^{*y} or $^{*y.5}$ in Table 1 for all y or $y > 0.5$, respectively.

For the translation, predictions for the ratios of the double differential cross section at the (x, Q^2) and $\sqrt{s}$ where the measurements took place and the $(x_{\text{grid}}, Q^2_{\text{grid}})$ to which they are translated are needed. These predictions, T_{grid}, were obtained from the data themselves by performing fits to the data using the HERAfitter tool. For $Q^2 \geq 3\,\text{GeV}^2$, a QCD fit within the DGLAP formalism was performed. In addition, a fit using the fractal model[5] [**?** **?**] was performed for $Q^2 \leq 4.9\,\text{GeV}^2$. For $Q^2 < 3\,\text{GeV}^2$, the fit to the fractal model was used[6] to obtain factors $T_{\text{grid-FM}}$. For $Q^2 > 4.9\,\text{GeV}^2$, the QCD fit was used to provide $T_{\text{grid-QCD}}$. For $3\,\text{GeV}^2 \leq Q^2 \leq 4.9\,\text{GeV}^2$, the factors were averaged as $T_{\text{grid}} = (1-0.53(Q^2-3\,\text{GeV}^2))T_{\text{grid-FM}} + 0.53(Q^2 - 3\,\text{GeV}^2)T_{\text{grid-QCD}}$.

The averaging of the data was done as decribed in section 4.1.

In the first interation the fits to provide the T_{grid} values are performed on the uncombined data. Starting with the second iteration, the fits are performed on combined data. The process was stopped after the third iteration. It was tested that further iterations do not induce significant changes in the cross sections.

4.4 Procedural Uncertainties

The χ^2 definition from equation 20 treats all systematic uncertainties as multiplicative, i.e. their size is expected to be proportional to the "true" values $\boldsymbol{m}$. While this generally is a good assumption for normalisation uncertainties, it might not be for other uncertainties. Therefore an alternative averaging was performed, in which only the normalisation uncertainties were taken as multiplicative while all other uncertainties were treated as additive. The differences between this alternative average and the nominal averages were taken as correlated procedural uncertainties $\delta_{\text{ave,rel}}$. The typical values of $\delta_{\text{ave,rel}}$ for the $\sqrt{s} = 320\,\text{GeV}$ (lower-$\sqrt{s}$) combination were below 0.5% (1%) for medium-Q^2 data, increasing to a few percent for low- and high-Q^2 data.

The H1 and ZEUS collaborations at some stage used similar methods to calibrate the detectors. They also employed similar Monte Carlo simulation models. These similar approaches led to correlations between the H1 and ZEUS measurements, especially for the HERA I period. This was investigated in depth for the combination of HERA I data [**?**]. The important correlations for this period were found to be related to the background from photoproduction and hadronic energy scales. The correlations between the experiments for the HERA II period were considered much less important, because both experiments developed diverging methods to deal with these issues. In addition, the correlations between HERA I and HERA II were weak, especially for ZEUS, because of new methods and changes in the detector. The correlations for the HERA I period were taken into account as before [**?**]. The differences between the nominal average and the averages in which systematic sources for the photoproduction background and hadronic energy scale are considered to be correlated are taken as additional procedural uncertainties $\delta_{\text{ave},\gamma\text{p}}$ and $\delta_{\text{ave,had}}$. Typical values of $\delta_{\text{ave},\gamma\text{p}}$ and $\delta_{\text{ave,had}}$ are below 1% (0.5%) for NC (CC) scattering. For low-Q^2 data, they can reach a few percent.

[5] The *ansatz* of the fractal model is based on the self-similar properties in x and Q^2 of the proton structure function at low x. They are represented by two continuous, variable and correlated fractal dimensions.

[6] A cross check was performed using the colour dipole model [**?**] as implemented in HERAfitter. The results did not change significantly.

5 Cross Sections

The averaged NC and CC reduced cross sections together with statistical, uncorrelated systematic and procedural uncertainties will be provided in the final paper. The full information about correlation between cross-section measurements will be available elsewhere [**?**]. The total integrated luminosity of the combined data set corresponds to about 500 pb^{-1} for both e^+p and e^-p. In total, 2927 data points were combined to 1307 cross-section measurements. The data showed good consistency, with $\chi^2/n_{\mathrm{dof}} = 1685/1620$.

For data points, k, contributing to point i on the (x, Q^2)-grid, pulls $\mathrm{p}^{i,k}$ were defined as

$$\mathrm{p}^{i,k} = \frac{\mu^{i,k} - \mu^{i,\mathrm{ave}}\left(1 - \sum_j \gamma_j^{i,k} b_{j,\mathrm{ave}}\right)}{\sqrt{\Delta_{i,k}^2 - \Delta_{i,\mathrm{ave}}^2}}\,, \tag{23}$$

where $\Delta_{i,k}$ and $\Delta_{i,\mathrm{ave}}$ are the statistical and uncorrelated systematic uncertainties added in quadrature for the point k and the average, respectively. The distribution of pulls shows no exceptional tensions for all datasets, i.e. processes across the kinematic plane, as demonstrated in Figure 2.

There are in total 162 sources of correlated systematic uncertainty including global normalisations characterising the separate data sets. None of these systematic sources shifts by more than $2.4\,\sigma$ of the nominal value in the averaging procedure.

The influence of several correlated systematic uncertainties was reduced significantly for the averaged result. For example, the uncertainty due to the H1 LAr calorimeter energy scale was reduced by 55% while the uncertainty due to the ZEUS photoproduction background is reduced by 70%. There are two main reasons for this significant reduction. Since H1 and ZEUS use different reconstruction methods, described in section 3.2, similar systematic sources influence the measured cross section differently as a function of x and Q^2. Therefore, requiring the cross sections to agree at all x and Q^2 constrains the systematics efficiently. In addition, for certain regions of the phase space, one of the two experiments has superior precision compared to the other. For these regions, the less precise measurement is fitted to the more precise one, with a simultaneous reduction of the correlated systematic uncertainty. This reduction propagates to the other average points, including those which are based solely on the measurement from the less precise experiment.

Over most of the phase space, the precision of the H1 and ZEUS measurements are about equal and the systematic uncertainties are reduced uniformly. The total uncertainty is typically around 1% for $20 < Q^2 < 100$ GeV^2, less than X% for $100 < Q^2 < 500$ GeV^2 and less than X% for $500 < Q^2 < 3000$ GeV^2.

Figures 3 and 4 show the averaged NC e^+p reduced cross sections together with the input data from H1 and ZEUS for e^+p scattering and together with the equivalent result from the HERA I combination [**?**], respectively. Figures 5 and 6 depict the results for NC e^-p scattering. The benefit of averaging is enormous and the improvement with respect to HERA I due to the high-Q^2 data impressive, especially for e^-p scattering. Figure 7 shows combined NC data for e^+p and e^-p. The physics potential is obivous.

Figures 8, 9 and Figures 10, 11 show the averaged CC cross sections together with the input data from H1 and ZEUS and the comparison to the HERA I combination results for e^+p and e^-p

scattering, respectively. Again, both the power of averaging and the improved precision due to the high statistics data from HERA II is demonstrated.

Figures 12 and 13 demonstrate the power of combination for the data with lowered proton beam energy. This part of the phase space is sensitive to the gluon density in the nucleon.

6 Conclusions

The result of a combination of all inclusive deep inelastic cross sections measured by the H1 and ZEUS collaborations in neutral and charged current unpolarised $e^{\pm}p$ scattering at HERA was presented. The combination based on a total luminosity of about $1\,\mathrm{fb}^{-1}$ of data produced cross section measurements of very high precisions which are one of the legacies of the HERA experiments.

7 Acknowledgements

We are grateful to the HERA machine group whose outstanding efforts made these experiments possible. We appreciate the contributions to the construction and maintenance of the H1 and ZEUS detectors of many people who are not listed as authors. We thank our funding agencies for financial support, the DESY technical staff for continuous assistance and the DESY directorate for their support and for the hospitality they extended to the non-DESY members of the collaborations.

Data Set		x Grid from	x Grid to	Q^2/GeV^2 Grid from	Q^2/GeV^2 Grid to	$\mathcal{L}$ pb^{-1}	e^+/e^-	$\sqrt{s}$ GeV	x,Q^2 from equations	Ref.
HERA I E_p = 820 GeV and E_p = 920 GeV data sets										
H1 svx-mb	95-00	0.000005	0.02	0.2	12	2.1	e^+p	301, 319	11,15,16	[**?**]
H1 low Q^2	96-00	0.0002	0.1	12	150	22	e^+p	301, 319	11,15,16	[**?**]
H1 NC	94-97	0.0032	0.65	150	30000	35.6	e^+p	301	17	[**?**]
H1 CC	94-97	0.013	0.40	300	15000	35.6	e^+p	301	12	[**?**]
H1 NC	98-99	0.0032	0.65	150	30000	16.4	e^-p	319	17	[**?**]
H1 CC	98-99	0.013	0.40	300	15000	16.4	e^-p	319	12	[**?**]
H1 NC HY	98-99	0.0013	0.01	100	800	16.4	e^-p	319	11	[**?**]
H1 NC	99-00	0.0013	0.65	100	30000	65.2	e^+p	319	17	[**?**]
H1 CC	99-00	0.013	0.40	300	15000	65.2	e^+p	319	12	[**?**]
ZEUS BPC	95	0.000002	0.00006	0.11	0.65	1.65	e^+p	300	11	[**?**]
ZEUS BPT	97	0.0000006	0.001	0.045	0.65	3.9	e^+p	300	11, 17	[**?**]
ZEUS SVX	95	0.000012	0.0019	0.6	17	0.2	e^+p	300	11	[**?**]
ZEUS NC	96-97	0.00006	0.65	2.7	30000	30.0	e^+p	300	19	[**?**]
ZEUS CC	94-97	0.015	0.42	280	17000	47.7	e^+p	300	12	[**?**]
ZEUS NC	98-99	0.005	0.65	200	30000	15.9	e^-p	318	18	[**?**]
ZEUS CC	98-99	0.015	0.42	280	30000	16.4	e^-p	318	12	[**?**]
ZEUS NC	99-00	0.005	0.65	200	30000	63.2	e^+p	318	18	[**?**]
ZEUS CC	99-00	0.008	0.42	280	17000	60.9	e^+p	318	12	[**?**]
HERA II E_p = 920 GeV data sets										
H1 NC	03-07	0.0008	0.65	60	30000	182	e^+p	319	11, 17	[**?**][1]
H1 CC	03-07	0.008	0.40	300	15000	182	e^+p	319	12	[**?**][1]
H1 NC	03-07	0.0008	0.65	60	50000	151.7	e^-p	319	11, 17	[**?**][1]
H1 CC	03-07	0.008	0.40	300	30000	151.7	e^-p	319	12	[**?**][1]
H1 NC med Q^2 *y.5	03-07	0.0000986	0.005	8.5	90	97.6	e^+p	319	11	[**?**]
H1 NC low Q^2 *y.5	03-07	0.000029	0.00032	2.5	12	5.9	e^+p	319	11	[**?**]
ZEUS NC	06-07	0.005	0.65	200	30000	135.5	e^+p	318	11,12,18	[**?**]
ZEUS CC	06-07	0.0078	0.42	280	30000	132	e^+p	318	12	[**?**]
ZEUS NC	05-06	0.005	0.65	200	30000	169.9	e^-p	318	18	[**?**]
ZEUS CC	04-06	0.015	0.65	280	30000	175	e^-p	318	12	[**?**]
ZEUS NC nominal *y	06-07	0.000092	0.008343	7	110	44.5	e^+p	318	11	[**?**]
ZEUS NC satellite *y	06-07	0.000071	0.008343	5	110	44.5	e^+p	318	11	[**?**]
HERA II E_p = 575 GeV data sets										
H1 NC high Q^2	07	0.00065	0.65	35	800	5.4	e^+p	252	11, 17	[**?**]
H1 NC low Q^2	07	0.0000279	0.0148	1.5	90	5.9	e^+p	252	11	[**?**]
ZEUS NC nominal	07	0.000147	0.013349	7	110	7.1	e^+p	251	11	[**?**]
ZEUS NC satellite	07	0.000125	0.013349	5	110	7.1	e^+p	251	11	[**?**]
HERA II E_p = 460 GeV data sets										
H1 NC high Q^2	07	0.00081	0.65	35	800	11.8	e^+p	225	11, 17	[**?**]
H1 NC low Q^2	07	0.0000348	0.0148	1.5	90	12.2	e^+p	225	11	[**?**]
ZEUS NC nominal	07	0.000184	0.016686	7	110	13.9	e^+p	225	11	[**?**]
ZEUS NC satellite	07	0.000143	0.016686	5	110	13.9	e^+p	225	11	[**?**]

Table 1: The 41 data sets from H1 and ZEUS used for the combination. The markers *y.5 and *y in the collumn "Data Set"are explained in a footnote in section 4.2. The marker [1] for [**?**] indicates that published cross section were scaled by a factor of 1.018 [erratum–[**?**]].
Luminosities are quoted as given by the collaborations; H1 luminosities are given for the data within the Z-vertex acceptance; ZEUS luminosities are given without any acceptance cut. The equations used for the reconstruction of x and Q^2 are given in section 3.2.

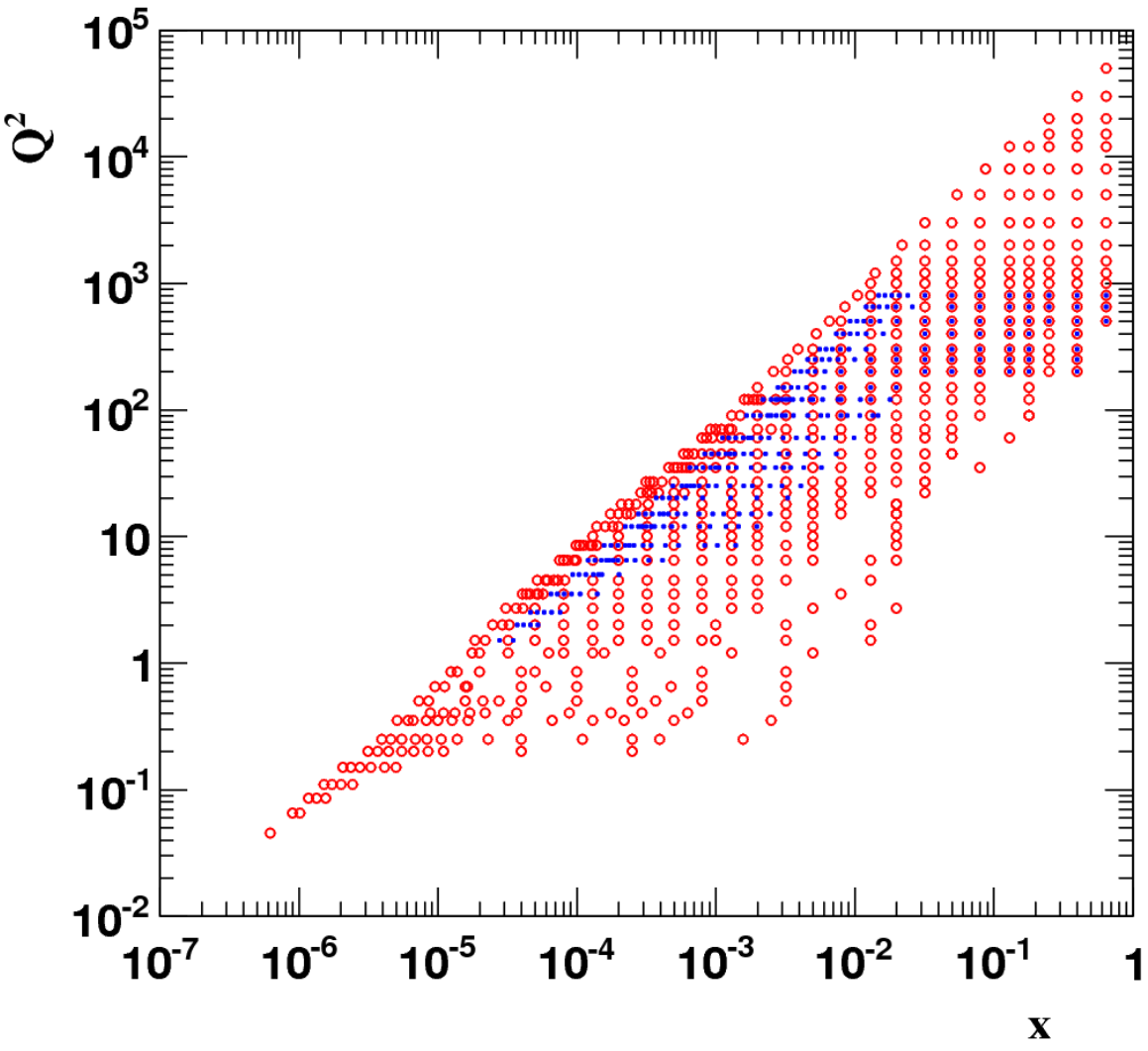

Figure 1: The points of the two grids for $\sqrt{s}_{\mathrm{com},1} = 318\,\mathrm{GeV}$ (big open circles) and $\sqrt{s}_{\mathrm{com},2} = 252\,\mathrm{GeV}$ as well as $\sqrt{s}_{\mathrm{com},3} = 225\,\mathrm{GeV}$ (small filled squares) are shown. The latter grid has a finer binning in x in accordance with its special structure in y.

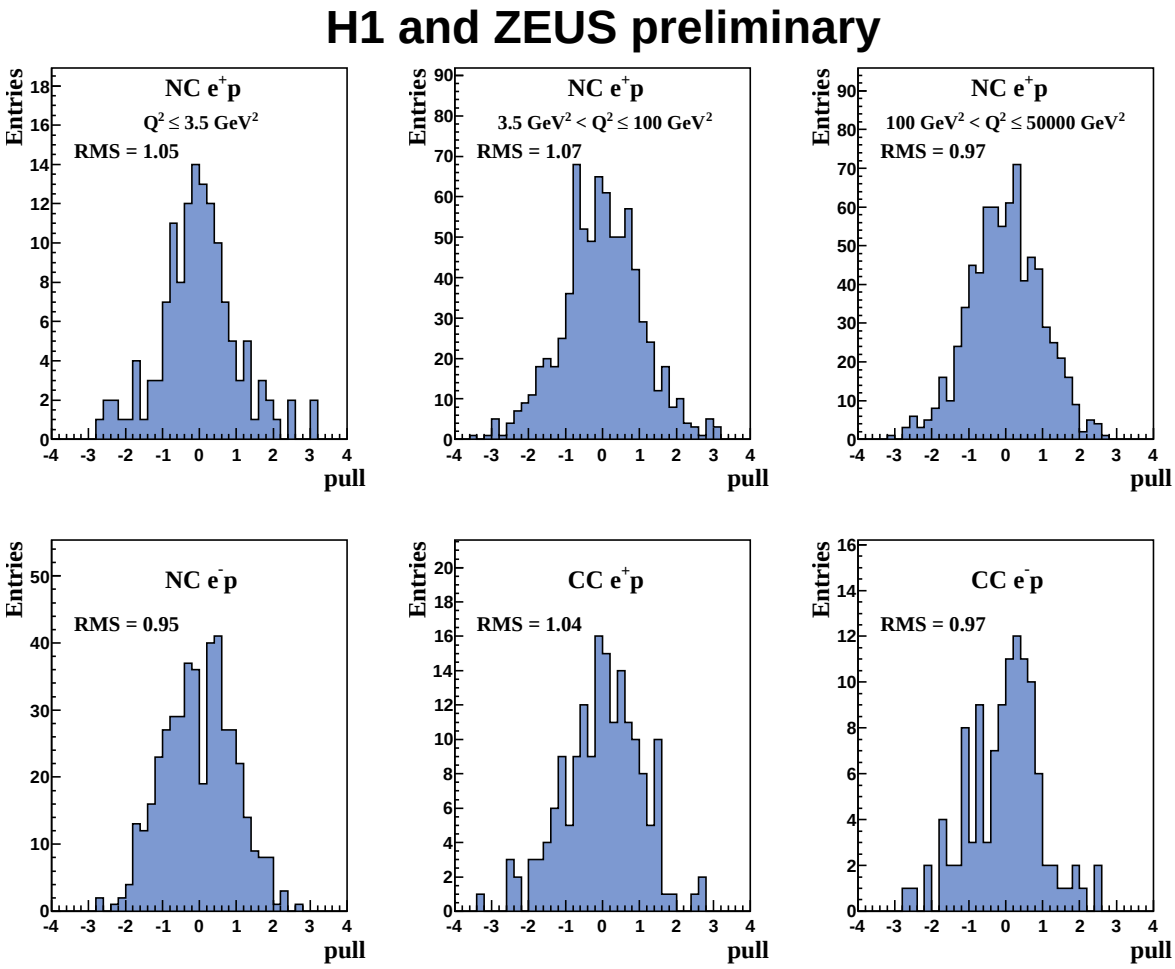

Figure 2: Distribution of pulls p for the following samples: a) NC e^+p for $Q^2 < 3.5\,\mathrm{GeV}^2$; b) NC e^+p for $3.5 \leq Q^2 < 100\,\mathrm{GeV}^2$; c) NC e^+p for $Q^2 \geq 100\,\mathrm{GeV}^2$; d) NC e^-p; e) CC e^+p and f) CC e^-p. There are no entries outside the histogram ranges. RMS gives the root mean square of each distribution calculated as $\overline{p^2}$. The curves show the results of binned log-likelihood Gaussian fits to the distributions.

Figure 3: HERA combined NC e^+p reduced cross section as a function of Q^2 for six selected x-bins compared to the separate H1 and ZEUS data which were the input to the averaging procedure. The individual measurements are displaced horizontally for better visibility. Errors bars represent the total uncertainties.

Figure 4: HERA combined NC e^+p reduced cross section as a function of Q^2 for six selected x-bins compared to the results from HERA I alone [**?**]. The two measurements are displaced horizontally for better visibility. Errors bars represent the total uncertainties.

Figure 5: HERA combined NC $e^- p$ reduced cross section as a function of Q^2 for four selected x-bins compared to the separate H1 and ZEUS data which were the input to the averaging procedure. The individual measurements are displaced horizontally for better visibility. Errors bars represent the total uncertainties.

Figure 6: HERA combined NC $e^- p$ reduced cross section as a function of Q^2 for four selected x-bins compared to the results from HERA I alone [**?**]. The two measurements are displaced horizontally for better visibility. Errors bars represent the total uncertainties.

Figure 7: HERA combined NC e^+p and e^-p reduced cross sections as a function of Q^2 for four selected x-bins. Errors bars represent the total uncertainties.

Figure 8: HERA combined CC e^+p reduced cross section as a function of of x for 10 Q^2 bins compared to the separate H1 and ZEUS data which were the input to the averaging procedure. The individual measurements are displaced horizontally for better visibility. Errors bars represent the total uncertainties.

Figure 9: HERA combined CC e^+p reduced cross section as a function of of x for 10 Q^2 bins to the results from HERA I alone [**?**]. The individual measurements are displaced horizontally for better visibility. Errors bars represent the total uncertainties.

Figure 10: HERA combined CC $e^- p$ reduced cross section as a function of of x for 10 Q^2 bins compared to the separate H1 and ZEUS data which were the input to the averaging procedure. The individual measurements are displaced horizontally for better visibility. Errors bars represent the total uncertainties.

Figure 11: HERA combined CC $e^- p$ reduced cross section as a function of of x for 10 Q^2 bins to the results from HERA I alone [**?**]. The individual measurements are displaced horizontally for better visibility. Errors bars represent the total uncertainties.

Figure 12: HERA combined NC e^+p reduced cross section at $E_p = 460$ GeV running as a function of x for five selected Q^2 bins compared to the separate H1 and ZEUS data which were the input to the averaging procedure. The individual measurements are displaced horizontally for better visibility. Errors bars represent the total uncertainties.

Figure 13: HERA combined NC e^+p reduced cross section at $E_p = 575$ GeV running as a function of x for five selected Q^2 bins compared to the separate H1 and ZEUS data which were the input to the averaging procedure. The individual measurements are displaced horizontally for better visibility. Errors bars represent the total uncertainties.