Experimental determination of Parton Distributions

T. Carli, A. Cooper-Sarkar, J. Feltesse, A. Glazov, C. Gwenlan, M. Klein, T. Laštovička
G. Laštovička-Medin, S. Moch, B. Reisert G. Salam, F. Siegert

1 Introduction [1]

With HERA currently in its second stage of operation, it is possible to assess the potential precision limits of HERA data and to estimate the potential impact of the measurements which are expected at HERA-II, in particular with respect to the PDF uncertainties.

Precision limits of the structure function analyses at HERA are examined in [1]. Since large amounts of luminosity are already collected, the systematic uncertainty becomes most important. A detailed study of error sources with particular emphasis on correlated errors for the upcoming precision analysis of the inclusive DIS cross section at low Q^2 using 2000 data taken by the H1 experiment is presented. A new tool, based on the ratio of cross sections measured by different reconstruction methods, is developed and its ability to qualify and unfold various correlated error sources is demonstrated.

An important issue is the consistency of the HERA data. In section 3, the H1 and ZEUS published PDF analyses are compared, including a discussion of the different treatments of correlated systematic uncertainties. Differences in the data sets and the analyses are investigated by putting the H1 data set through both PDF analyses and by putting the ZEUS and H1 data sets through the same (ZEUS) analysis, separately. Also, the HERA averaged data set (section 4) is put through the ZEUS PDF analysis and the result is compared to that obtained when putting the ZEUS and H1 data sets through this analysis together, using both the Offset and Hessian methods of treating correlated systematic uncertainties.

The HERA experimental data can not only be cross checked with respect to each other but also combined into one common dataset, as discussed in section 4. In this respect, a method to combine measurements of the structure functions performed by several experiments in a common kinematic domain is presented. This method generalises the standard averaging procedure by taking into account point-to-point correlations which are introduced by the systematic uncertainties of the measurements. The method is applied to the neutral and charged current DIS cross section data published by the H1 and ZEUS collaborations. The averaging improves in particular the accuracy due to the cross calibration of the H1 and ZEUS measurements.

The flavour decomposition of the light quark sea is discussed in [2]. For low x and thus low Q^2 domain at HERA only measurement of the photon exchange induced structure functions F_2 and F_L is possible, which is insufficient to disentangle individual quark flavours. A general strategy in this case is to assume flavour symmetry of the sea. [2] considers PDF uncertainties if this assumption is released. These uncertainties can be significantly reduced if HERA would run in deuteron-electron collision mode.

The impact of projected HERA-II data on PDFs is estimated in section 7. In particular, next-to-leading order (NLO) QCD predictions for inclusive jet cross sections at the LHC centre-of-mass energy are presented using the estimated PDFs. A further important measurement which could improve understanding of the gluon density at low x and, at the same time, provide consistency checks of the low Q^2 QCD evolution is the measurement of the longitudinal structure function F_L. Perspectives of this measurement are examined in section 5, while the impact of this measurement is also estimated in section 7.

Further improvements for consistently including final-state observables in global QCD analyses are discussed in section 8. There, a method for "a posteriori" inclusion of PDFs, whereby the Monte Carlo run calculates a grid (in x and Q) of cross section weights that can subsequently be combined with an arbitrary PDF. The procedure is numerically equivalent to using an interpolated form of the PDF. The

[1] Subsection coordinators: A. Glazov, S. Moch

main novelty relative to prior work is the use of higher-order interpolation, which substantially improves the tradeoff between accuracy and memory use. An accuracy of about $0.01\,\%$ has been reached for the single inclusive cross-section in the central rapidity region $|y| < 0.5$ for jet transverse momenta from $1\ \ 00$ to $5\ \ 000\mathrm{Ge\,V}$. This method will make it possible to consistently include measurements done at HERA, Tevatron and LHC in global QCD analyses.

2 Precision Limits for HERA DIS Cross Section Measurement [2]

The published precision low Q^2 cross section data [3] of the H1 experiment became an important data set in various QCD fit analyses [3–6]. Following success of these data the H1 experiment plans to analyse a large data sample, taken during 2000 running period[3], in order to reach precision limits of low Q^2 inclusive cross sections measurements at HERA. The precision is expected to approach 1% level.

The aim of this contribution is to calculate realistic error tables for 2000 H1 data and pursue paths how to reach such a high precision. Correlated error sources are studied in particular and a new tool, based on the ratio of cross sections measured by different reconstruction methods, is developed. All errors, including correlated errors, are treated in the same manner as in [3]. Error tables are provided and used in QCD fit analysis, see Sec 7, in order to study the impact of the new data on PDFs. The new data are expected to reach higher precision level than [3] due to the following reasons

- Larger data statistics - Statistical errors will decrease by factor of $1.5 - 2$, compared to [3], depending on the kinematic region.
- Very large Monte Carlo simulations (MC) - Due to a progress in computing a number of simulated events can be significantly increased in order to minimise statistical error of MC, to understand uncorrelated errors and to estimate correlated errors more precisely.
- During past years increasing knowledge, arriving from various H1 analyses, enabled better understanding of the detector and its components as well as improving quality of MC.
- Data taking in 2000 was particularly smooth. Both HERA and H1 were running at peak performance for HERA-I running period.

This contribution uses existing 2000 data and MC ntuples along with the full analysis chain. It applies all preliminary technical work done on these data, including calibration, alignment, trigger studies etc. Quoted errors are assumed to be achieved in the final version of analysis yet the analysis has not been finalised, all the numbers in the paper are preliminary and may change in the publication.

The uncertainties of the cross section measurement are divided into a number of different types. Namely, these are *statistical uncertainties* of the data, *uncorrelated systematics* and *correlated systematics*. The term 'correlated' refers to the fact that cross section measurements in kinematic bins are affected in a correlated way while different correlated systematic error sources are considered uncorrelated among each other. The classification of the systematic errors into types is sometimes straightforward (MC statistics is uncorrelated error source) but sometimes is rather arbitrary (radiative corrections are assumed to be uncorrelated error source). The main goal of this classification is to preserve correlation between data points while keeping the treatement as simple as possible.

The cross section uncertainties depend on the method used to reconstruct event kinematics. There are various methods existing, involving a measurement of the scattered electron as well as of the hadronic finale state. In the following two of them, so called *electron method* and *sigma method*, are employed [7]. The electron method uses only the measurement of the scattered electron, namely its energy and polar angle, while the sigma method uses both the scattered electron and the hadronic final state. An advantage of the sigma method is a proper treatment of QED radiation from the incoming beam electron (ISR).

[2] Contributing authors: G. Laštovička-Medin, A. Glazov, T. Laštovička

[3] Data statistics will be increased further by adding data taken in year 1999.

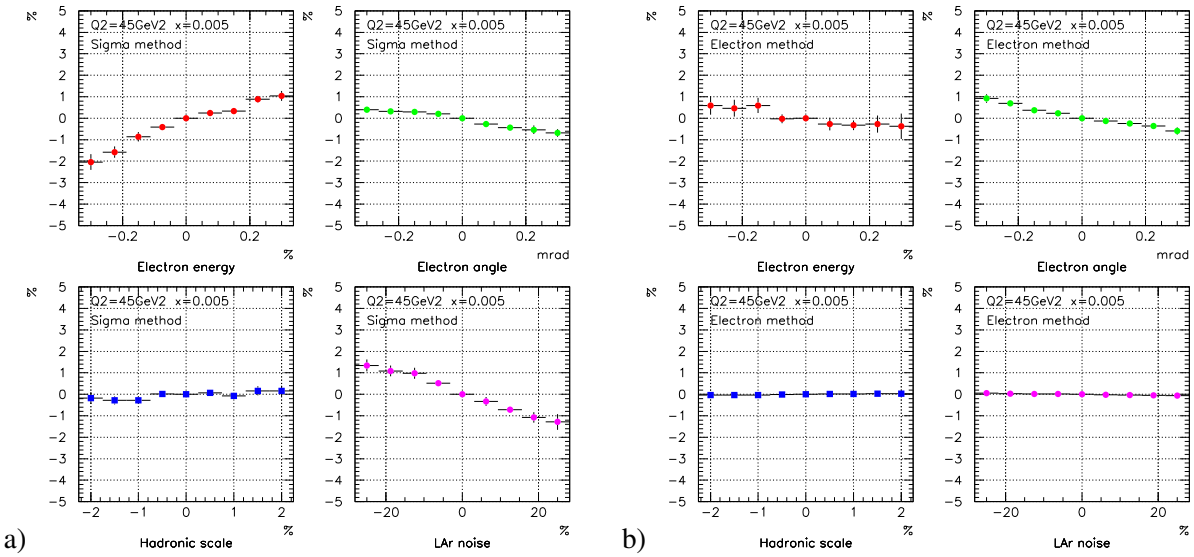

Fig. 1: A scan of the cross section measurement change in % depending on a variation of (from top-left) electron energy, electron polar angle, hadronic final state calibration scale and noise level in LAr calorimeter (bottom-right). The sigma method (a) and the electron method (b) were used to reconstruct kinematics of events.

The *statistical uncertainty* of the data is typically 0.5-1%, depending on the kinematic region analysed and the definition of the kinematic bins. In the following we adapt the bin definition used in [3], apart from merging bins at low y which was done in the published data in order to increase statistics.

The *uncorrelated systematics* consists from various contributions. A cross section uncertainty due to the Monte Carlo statistics is the one with very good potential to be minimised. In the following we assume 100 million simulated events to be used in analysis of 2000 data. Estimates were calculated with available 12 million simulated events and corresponding statistical errors scaled by a factor of $\sqrt{1\ \ 0\emptyset 1\ \ .2}$ As a result the uncertainty is very small and typically on the level of few permile.

Additional contributions to the uncorrelated systematics are efficiencies. We assume for trigger efficiency 0.3% and backward tracker tracker efficiency 0.3% uncertainty. Radiative corrections are expected to affect the final cross section by 0.4%.

Effect of *correlated uncertainties* on the cross section measurement is studied in the following manner. Particular source of correlated uncertainty, for instance the scattered electron energy measurement, is varied by assumed error and the change of the measured cross section is quoted as the corresponding cross section measurement error. An example of cross section change on various correlated error source is shown in Fig. 1 for bin of $Q^2 = 4\ \ 5\text{GeV}^2$ and $x = 0.00\emptyset 5$. The kinematics of events was reconstructed with the sigma method (a) and the electron method (b). Errors are calculated as so called standard errors of the mean in calculation of which the available Monte Carlo sample was split into nine sub-samples. It is clearly seen that the cross section measurement with the sigma method in this kinematic bin is particularly sensitive to the electron energy measurement (top-left) and to noise description in LAr calorimeter (bottom-right). On the contrary, the electron polar angle measurement and the calibration of the hadronic final state play a little role. The electron method is mainly sensitive to the electron energy measurement. The importance of the systematic sources vary from bin to bin.

There are five individual sources contributing to the correlated cross section uncertainties:

- Uncertainties of 0.15% at $E_e = 2\ \ 7$GeV and 1% at 7 GeV are assigned to the electron energy scale for the backward calorimeter. The uncertainty is treated as a linear function of E_e interpolating between the results at 27 GeV and 7 GeV.
- The uncertainty on the scattered electron polar angle measurement is 0.3 mrad . The corresponding

Table 1: An example of the error table for $Q^2 = 25\,\mathrm{GeV}^2$ for 2000 data, large Monte Carlo sample and suppressed systematic errors compared to [1], see text for details. Absolute errors are shown. The table format is identical to the one published in [1].

Q^2	x	y	σ_r	R	F_2	Tot.(%)	Sta.	Uncorr.	Corr.	E_e	θ	Ehad	Noise	γp
25	0.0005	0.493	1.391	0.261	1.449	0.88	0.47	0.63	0.41	0.19	0.21	0.22	0.15	0.13
25	0.0008	0.308	1.251	0.261	1.268	0.91	0.43	0.62	0.51	0.34	0.37	0.02	0.04	0
25	0.0013	0.19	1.138	0.248	1.143	0.94	0.44	0.62	0.56	0.45	0.33	0.03	0.02	0
25	0.002	0.123	1.041	0.236	1.042	0.9	0.45	0.62	0.47	0.13	0.45	0.03	0.05	0
25	0.0032	0.077	0.842	0.254	0.843	1.42	0.5	0.63	1.17	0.74	0.36	0.17	0.8	0
25	0.005	0.049	0.745	0.243	0.745	1.17	0.52	0.63	0.83	0.59	0.42	0.25	0.33	0
25	0.008	0.031	0.667	0.225	0.667	1.22	0.56	0.64	0.87	0.43	0.35	0.66	0.09	0
25	0.013	0.019	0.586	0.214	0.586	2.02	0.65	0.66	1.8	0.67	0.57	1.43	0.65	0
25	0.02	0.012	0.569	0.159	0.569	5.77	0.86	0.71	5.66	0.83	0.52	3.51	4.33	0
25	0.032	0.008	0.553	0.065	0.553	10.64	1.34	0.88	10.52	0.93	0.64	3.86	9.72	0

Table 2: An example of the full error table for $Q^2 = 25\,\mathrm{GeV}^2$, published H1 data. The definition of kinematic bins is not identical to that in Table 1, some bins were merged to enlarge statistics.

Q^2	x	y	σ_r	R	F_2	Tot.(%)	Sta.	Uncorr.	Corr.	E_e	θ	Ehad	Noise	γp
25	0.0005	0.553	1.345	0.248	1.417	2.41	1.04	1.81	1.21	-1.04	-0.37	0.25	0.04	-0.41
25	0.0008	0.346	1.242	0.243	1.263	1.94	0.67	1.62	0.85	-0.6	-0.6	0.04	0.02	-0.07
25	0.0013	0.213	1.091	0.238	1.097	1.78	0.66	1.36	0.93	-0.64	-0.69	0	0	0
25	0.002	0.138	0.985	0.236	0.987	2.89	0.76	1.43	2.4	1.78	-0.7	0.17	1.34	0
25	0.0032	0.086	0.879	0.234	0.88	2.78	0.79	1.46	2.23	1.8	-0.77	-0.23	0.92	0
25	0.005	0.055	0.754	0.234	0.754	2.38	0.85	1.49	1.64	1.01	-0.58	0.16	1.03	0
25	0.008	0.034	0.663	0.234	0.663	2.52	0.92	1.54	1.78	1.11	-0.68	-0.72	0.84	0
25	0.0158	0.018	0.547	0.226	0.547	3.71	0.85	1.49	3.29	1.36	-0.88	-2.44	-1.42	0
25	0.05	0.005	0.447	0.148	0.447	7.54	1.28	3.35	6.64	0.99	-0.68	-3.28	-5.62	0

error on the cross section measurement is typically well below 1% but may be larger at lowest values of Q^2.

- The uncertainty on the hadronic energy scale comprises a number of systematic error sources corresponding to the $E-p_z$ decomposition: an uncertainty of the hadronic energy scale calibration of 2% for the central and forward calorimeter, an uncertainty of 3% for the fraction carried by tracks and a 5% uncertainty of the hadronic energy scale measured in backward calorimeter.
- The uncertainty on the hadronic energy scale is further affected by the subtracted noise in the calorimetery. The noise is described to the level of 10% and the corresponding error is propagated to the cross section uncertainty. The largest influence is in the low y region, which is measured with the sigma method.
- The uncertainty due to the photoproduction background at large y is estimated from the normalisation error of the PHOJET simulations to about 10%. At low and medium values of $y \lesssim 0.5$ it is negligible.

The total systematic error is calculated from the quadratic summation over all sources of the uncorrelated and correlated systematic uncertainties. The total error of the DIS cross section measurement is obtained from the statistical and systematical errors added in quadrature.

An example of the full error table for kinematic bin of $Q^2 = 2\ 5\mathrm{GeV}^2$ is shown in Table 1. For a comparison the corresponding part of the published data from [3] is presented in Table 2. One can see that precision about 1% can be reached especially in four lowest x bins, where the electron method was used to reconstruct the event kinematics. The key contributions to the seen improvement in the cross section measurement precision are the electron energy measurement, very large Monte Carlo statistics, well understood noise in LAr calorimeter and precisely controlled efficiencies entering the analysis.

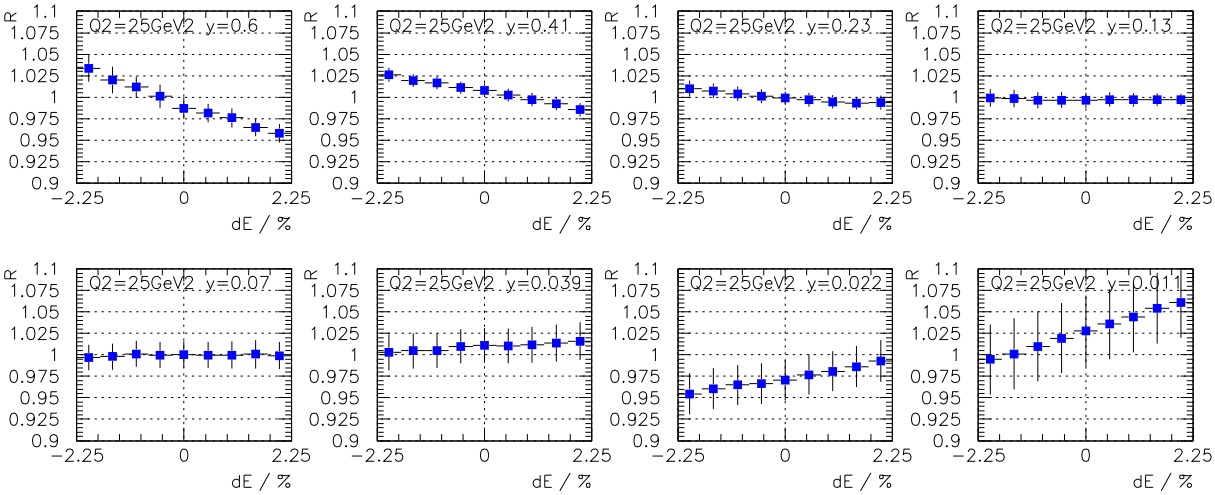

Fig. 2: A scan of the cross section ratio R in bins of Q^2 and y as a function of the hadronic final state calibration variation.

Full error table, covering the kinematic region of $5 \leq Q^2 \leq 1\ \ 5\,\mathrm{GeV}^2$ and $0.01 \leq y \leq 0.6$ was produced. The electron method was applied for kinematic bins at $y > 0.1$ while the sigma method otherwise. The measurement of the proton structure function F_2 was simulated using fractal parametrisation [8] for central values, accounting for all sources of correlated and uncorrelated errors. This table was used to estimate effect of precise low Q^2 data on the determination of proton PDFs from QCD fits.

The fact that different kinematics reconstruction methods are affected differently by the correlated systematic uncertainties may be employed as a tool to estimate these uncertainties. We define

$$R_i = \frac{\sigma_r^{el,\,i}}{\sigma_r^{\Sigma,\,i}} \tag{1}$$

to be the cross section measurement ratio, where the reduced cross section $\sigma_r^{el,\,i}$ and $\sigma_r^{\Sigma,\,i}$ is measured using the electron method and the sigma method, respectively. Kinematic bins, indexed by i, cover a region of the analysis phase space where both reconstruction methods are applicable for the measurement. The statistical error of R_i measurement is again evaluated by splitting the sample to a number of sub-samples and calculating the standard error of the mean. An example of a scan of the cross section ratio R_i dependence on the hadronic final state calibration variation in a bin of $Q^2 = 2\ 5\,\mathrm{GeV}^2$ and various inelasticity y is shown in Fig. 2.

An error of a particular correlated uncertainty source j can be estimated by searching for lowest $\chi^2 = \sum_i (R_i(\alpha_j) - 1\ ^2)/\ q_i^2$, where summation runs over kinematic bins, σ_i is the error of R_i measurement and α_j is the variation of the source j. However, since there is a number of correlated error sources the correct way to find correlated uncertainties is account for all of them.

Unfolding of the correlated error sources can be linearised and directly solved by minimising the following function:

$$\mathcal{L} = \sum_i \frac{1}{\sigma_i^2}(R_i + \sum_j \alpha_j \frac{\partial R_i}{\partial \alpha_j} - 1\ ^2) \tag{2}$$

The partial derivatives $\frac{\partial R_i}{\partial \alpha_j}$ for systematic source α_j are obtained from linear fits to distributions as shown in Fig. 2. Parameters α_j and their respective errors are obtained by matrix inversion technique.

The procedure was tested on available Monte Carlo sample for 2000 H1 data. Half of the sample, six million events, was used to simulate data. Full analysis chain was applied to measure the cross section and thus R_i. Kinematic bins were selected according to $1\ \ 5 \leq Q^2 \leq 60\,\mathrm{GeV}^2$ and $0.01\ \ 1 \leq y \leq 0.6$, i.e.

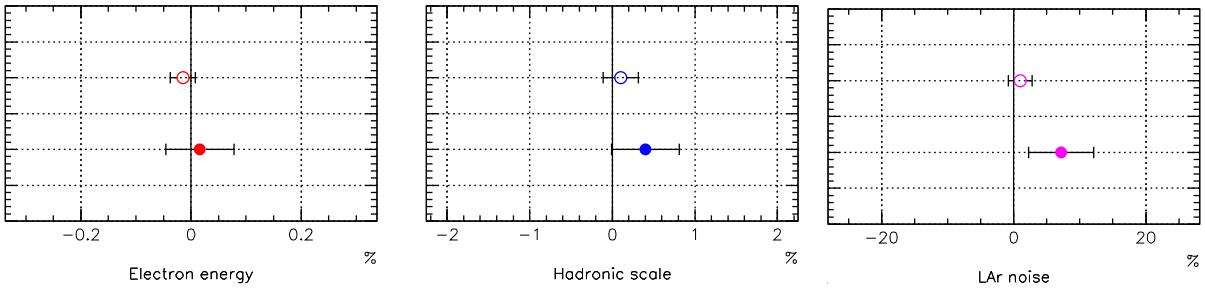

Fig. 3: Errors on the electron energy measurement (top-left), hadronic scale calibration (top-right) and noise in LAr calorimeter (bottom-left). Open points correspond to χ^2 scan in one correlated error source. Closed points show the result of complete unfolding, taking into account correlations.

in the main region of the data. The results are shown in Fig. 3. Closed points correspond to unfolded errors of the electron energy measurement (top-left), hadronic final state calibration and noise in the LAr calorimeter (bottom-left). There is no sensitivity observed to the electron polar angle measurement. All values are within statistical errors compatible with zero, as expected. For the final analysis the statistical errors are expected to be approximately three times smaller due to the significantly larger statistics than used for the presented study. This will enable the method to gain sufficient control over systematic correlated errors. Apart from being able to evaluate calibration of the scattered electron and of the hadronic final state, it gives a very good handle on the LAr calorimeter noise.

For a comparison, open points in Fig. 3 correspond to a χ^2 scan in one correlated error source. The statistical errors are smaller, as expected, and compatible with zero. However, the unfolding method is preferred since it takes into account all correlated error sources correctly.

In summary, a study of the DIS cross section uncertainties realistically achievable at HERA has been performed. For $x \in 0.001 - 0.01$ a precision of 1 % can be reached across for a wide range of $Q^2 \in 5 - 1\ 5000\ \text{GeV}^2$, allowing improved estimate of W, Z production cross section in the central rapidity region of LHC. The accuracy of the DIS cross section measurement can be verified using different kinematic reconstruction methods available at the HERA collider.

3 Comparison and combination of ZEUS and H1 PDF analyses [4]

Parton Density Function (PDF) determinations are usually global fits [4,5,9], which use fixed target DIS data as well as HERA data. In such analyses the high statistics HERA NC e^+p data, which span the range $6.3 \times 10^{-5} < x < 0.65, 2.7 < Q^2 < 30{,}000\ \text{GeV}^2$, have determined the low-x sea and gluon distributions, whereas the fixed target data have determined the valence distributions and the higher-x sea distributions. The ν-Fe fixed target data have been the most important input for determining the valence distributions, but these data suffer from uncertainties due to heavy target corrections. Such uncertainties are also present for deuterium fixed target data, which have been used to determine the shape of the high-x d-valence quark.

HERA data on neutral and charged current (NC and CC) e^+p and e^-p inclusive double differential cross-sections are now available, and have been used by both the H1 and ZEUS collaborations [10, 11] in order to determine the parton distributions functions (PDFs) using data from within a single experiment. The HERA high Q^2 cross-section data can be used to determine the valence distributions, thus eliminating uncertainties from heavy target corrections. The PDFs are presented with full accounting for uncertainties from correlated systematic errors (as well as from statistical and uncorrelated sources). Peforming an analysis within a single experiment has considerable advantages in this respect, since the global fits have found significant tensions between different data sets, which make a rigorous statistical

[4] Contributing authors: A. Cooper-Sarkar, C. Gwenlan

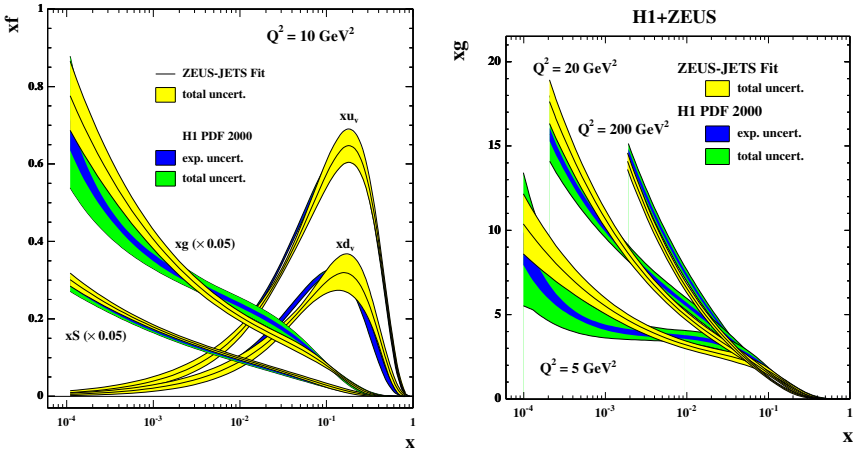

Fig. 4: Left plot: Comparison of PDFs from ZEUS and H1 analyses at $Q^2 = 10\text{GeV}^2$. Right plot: Comparison of gluon from ZEUS and H1 analyses, at various Q^2. Note that the ZEUS analysis total uncertainty includes both experimental and model uncertainties.

treatment of uncertainties difficult.

Fig. 4 compares the results of the H1 and ZEUS analyses. Whereas the extracted PDFs are broadly compatible within errors, there is a noticeable difference in the shape of the gluon PDFs. Full details of the analyses are given in the relevant publications, in this contribution we examine the differences in the two analyses, recapping only salient details.

3.1 Comparing ZEUS and H1 published PDF analyses

The kinematics of lepton hadron scattering is described in terms of the variables Q^2, the invariant mass of the exchanged vector boson, Bjorken x, the fraction of the momentum of the incoming nucleon taken by the struck quark (in the quark-parton model), and y which measures the energy transfer between the lepton and hadron systems. The differential cross-section for the NC process is given in terms of the structure functions by

$$\frac{d^2\sigma(e^{\pm}p)}{dxdQ^2} = \frac{2\pi\alpha^2}{Q^4x}\left[Y_+\,F_2(x,Q^2) - y^2\,F_L(x,Q^2) \mp Y_-\,xF_3(x,Q^2)\right], \tag{3}$$

where $Y_\pm = 1 \pm (\quad + y)^2$. The structure functions F_2 and xF_3 are directly related to quark distributions, and their Q^2 dependence, or scaling violation, is predicted by pQCD. At $Q^2 \leq 1\quad 000\text{GeV}^2$ F_2 dominates the charged lepton-hadron cross-section and for $x \leq 1\quad 0^2$, F_2 itself is sea quark dominated but its Q^2 evolution is controlled by the gluon contribution, such that HERA data provide crucial information on low-x sea-quark and gluon distributions. At high Q^2, the structure function xF_3 becomes increasingly important, and gives information on valence quark distributions. The CC interactions enable us to separate the flavour of the valence distributions at high-x, since their (LO) cross-sections are given by,

$$\frac{d^2\sigma(e^+p)}{dxdQ^2} = \frac{G_F^2M_W^4}{(Q^2+M_W^2)^2 2\pi x}x\left[(\bar{u}+\bar{c}) + (\quad + y)^2(d+s)\right],$$

$$\frac{d^2\sigma(e^-p)}{dxdQ^2} = \frac{G_F^2M_W^4}{(Q^2+M_W^2)^2 2\pi x}x\left[(u+c) + (\quad + y)^2(\bar{d}+\bar{s})\right].$$

For both HERA analyses the QCD predictions for the structure functions are obtained by solving the DGLAP evolution equations [12–15] at NLO in the $\overline{\text{M}}$ Sscheme with the renormalisation and factorization scales chosen to be Q^2. These equations yield the PDFs at all values of Q^2 provided they are

input as functions of x at some input scale Q_0^2. The resulting PDFs are then convoluted with coefficient functions, to give the structure functions which enter into the expressions for the cross-sections. For a full explanation of the relationships between DIS cross-sections, structure functions, PDFs and the QCD improved parton model see ref. [16].

The HERA data are all in a kinematic region where there is no sensitivity to target mass and higher twist contributions but a minimum Q^2 cut must be imposed to remain in the kinematic region where perturbative QCD should be applicable. For ZEUS this is $Q^2 > 2.5$ GeV^2, and for H1 it is $Q^2 > 3.5$ GeV^2. Both collaborations have included the sensitivity to this cut as part of their model errors.

In the ZEUS analysis, the PDFs for u valence, $xu_v(x)$, d valence, $xd_v(x)$, total sea, $xS(x)$, the gluon, $xg(x)$, and the difference between the d and u contributions to the sea, $x(\bar{d} - \bar{u})$, are each parametrized by the form

$$p_1 x^{p_2}(1 + x)^{p_3} P(x), \tag{4}$$

where $P(x) = 1 + p_4 x$, at $Q_0^2 = 7\text{GeV}^2$. The total sea $xS = 2x(\bar{u} + \bar{d} + \bar{s} + \bar{c} + \bar{b})$, where $\bar{q} = q_{sea}$ for each flavour, $u = u_v + u_{sea}$, $d = d_v + d_{sea}$ and $q = q_{sea}$ for all other flavours. The flavour structure of the light quark sea allows for the violation of the Gottfried sum rule. However, there is no information on the shape of the $\bar{d} - \bar{u}$ distribution in a fit to HERA data alone and so this distribution has its shape fixed consistent with the Drell-Yan data and its normalisation consistent with the size of the Gottfried sum-rule violation. A suppression of the strange sea with respect to the non-strange sea of a factor of 2 at Q_0^2, is also imposed consistent with neutrino induced dimuon data from CCFR. Parameters are further restricted as follows. The normalisation parameters, p_1, for the d and u valence and for the gluon are constrained to impose the number sum-rules and momentum sum-rule. The p_2 parameter which constrains the low-x behaviour of the u and d valence distributions is set equal, since there is no information to constrain any difference. When fitting to HERA data alone it is also necessary to constrain the high-x sea and gluon shapes, because HERA-I data do not have high statistics at large-x, in the region where these distributions are small. The sea shape has been restricted by setting $p_4 = 0$ for the sea, but the gluon shape is constrained by including data on jet production in the PDF fit. Finally the ZEUS analysis has 11 free PDF parameters. ZEUS have included reasonable variations of these assumptions about the input parametrization in their analysis of model uncertainties. The strong coupling constant was fixed to $\alpha_s(M_Z^2) = 0.1176$ [17]. Full account has been taken of correlated experimental systematic errors by the Offset Method, as described in ref [9, 18].

For the H1 analysis, the value of $Q_0^2 = 4\text{GeV}^2$, and the choice of quark distributions which are parametrized is different. The quarks are considered as u-type and d-type with different parametrizations for, $xU = x(u_v + u_{sea} + c)$, $xD = x(d_v + d_{sea} + s)$, $x\bar{U} = x(\bar{u} + \bar{c})$ and $x\bar{D} = x(\bar{d} + \bar{s})$, with $q_{sea} = \bar{q}$, as usual, and the the form of the quark and gluon parametrizations given by Eq. 4. For $x\bar{D}$ and $x\bar{U}$ the polynomial, $P(x) = 1.0$, for the gluon and xD, $P(x) = (1 + p_4 x)$, and for xU, $P(x) = (1 + p_4 x + p_5 x^3)$. The parametrization is then further restricted as follows. Since the valence distributions must vanish as $x \to 0$, the low-x parameters, p_1 and p_2 are set equal for xU and $x\bar{U}$, and for xD and $x\bar{D}$. Since there is no information on the flavour structure of the sea it is also necessary to set p_2 equal for $x\bar{U}$ and $x\bar{D}$. The normalisation, p_1, of the gluon is determined from the momentum sum-rule and the p_4 parameters for xU and xD are determined from the valence number sum-rules. Assuming that the strange and charm quark distributions can be expressed as x independent fractions, f_s and f_c, of the d and u type sea, gives the further constraint $p_1(\bar{U}) = p_1(\bar{D})(1 - f_s)/(1 - f_c)$. Finally there are 10 free parameters. H1 has also included reasonable variations of these assumptions in their analysis of model uncertainties. The strong coupling constant was fixed to $\alpha_s(M_Z^2) = 0.1180$ and this is sufficiently similar to the ZEUS choice that we can rule it out as a cause of any significant difference. Full account has been taken of correlated experimental systematic errors by the Hessian Method, see ref. [18].

For the ZEUS analysis, the heavy quark production scheme used is the general mass variable flavour number scheme of Roberts and Thorne [19]. For the H1 analysis, the zero mass variable flavour

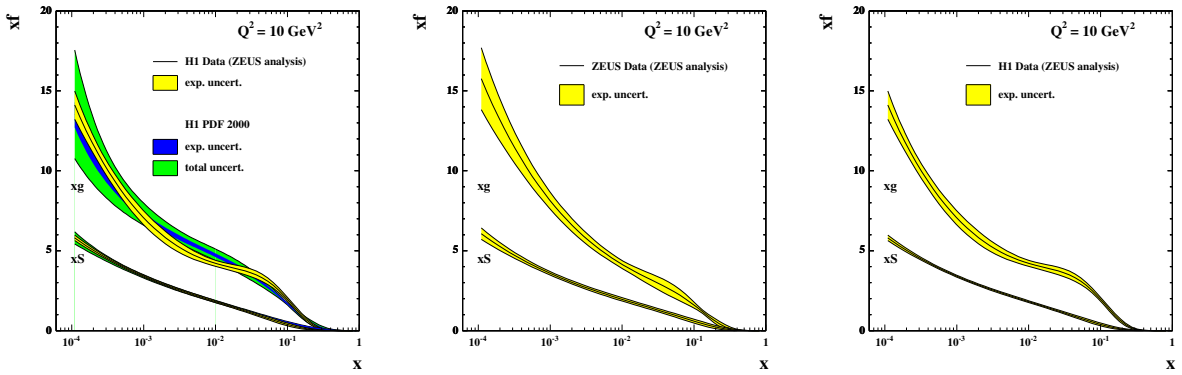

Fig. 5: Sea and gluon distributions at $Q^2 = 10\text{GeV}^2$ extracted from different data sets and different analyses. Left plot: H1 data put through both ZEUS and H1 analyses. Middle plot: ZEUS data put through ZEUS analysis. Right plot: H1 data put through ZEUS analysis.

number scheme is used. It is well known that these choices have a small effect on the steepness of the gluon at very small-x, such that the zero-mass choice produces a slightly less steep gluon. However, there is no effect on the more striking differences in the gluon shapes at larger x.

There are two differences in the analyses which are worth further investigation. The different choices for the form of the PDF parametrization at Q_0^2 and the different treatment of the correlated experimental uncertainties.

3.2 Comparing different PDF analyses of the same data set and comparing different data sets using the same PDF analysis.

So far we have compared the results of putting two different data sets into two different analyses. Because there are many differences in the assumptions going into these analyses it is instructive to consider:(i) putting both data sets through the same analysis and (ii) putting one of the data sets through both analyses. For these comparisons, the ZEUS analysis does NOT include the jet data, so that the data sets are more directly comparable, involving just the inclusive double differential cross-section data. Fig. 5 compares the sea and gluon PDFs, at $Q^2 = 10\text{GeV}^2$, extracted from H1 data using the H1 PDF analysis with those extracted from H1 data using the ZEUS PDF analysis. These alternative analyses of the same data set give results which are compatible within the model dependence error bands. Fig. 5 also compares the sea and gluon PDFs extracted from ZEUS data using the ZEUS analysis with those extracted from H1 data using the ZEUS analysis. From this comparison we can see that the different data sets lead to somewhat different gluon shapes even when put through exactly the same analysis. Hence the most of the difference in shape of the ZEUS and H1 PDF analyses can be traced back to a difference at the level of the data sets.

3.3 Comparing the Offset and Hessian method of assessing correlated experimental uncertainties

Before going further it is useful to discuss the treatment of correlated systematic errors in the ZEUS and H1 analyses. A full discussion of the treatment of correlated systematic errors in PDF analyses is given in ref [16], only salient details are recapped here. Traditionally, experimental collaborations have evaluated an overall systematic uncertainty on each data point and these have been treated as uncorrelated, such that they are simply added to the statistical uncertainties in quadrature when evaluating χ^2. However, modern deep inelastic scattering experiments have very small statistical uncertainties, so that the contribution of systematic uncertainties becomes dominant and consideration of point to point correlations between systematic uncertainties is essential.

For both ZEUS and H1 analyses the formulation of the χ^2 including correlated systematic uncertainties is constructed as follows. The correlated uncertainties are included in the theoretical prediction, $F_i(p\ s)$, such that

$$F_i(p\ s) = F_i^{\mathrm{NLOQCD}}(p) + \sum_\lambda s_\lambda \Delta_{i\lambda}^{\mathrm{sys}}$$

where, $F_i^{\mathrm{NLOQCD}}(p)$, represents the prediction from NLO QCD in terms of the theoretical parameters p, and the parameters s_λ represent independent variables for each source of systematic uncertainty. They have zero mean and unit variance by construction. The symbol $\Delta_{i\lambda}^{\mathrm{sys}}$ represents the one standard deviation correlated systematic error on data point i due to correlated error source λ. The χ^2 is then formulated as

$$\chi^2 = \sum_i \frac{[F_i(p\ s) - F_i(\ \mathrm{mea}\ s)]^2}{\sigma_i^2} + \sum_\lambda s_\lambda^2 \tag{5}$$

where, $F_i(\ \mathrm{mea}\ s)$ represents a measured data point and the symbol σ_i represents the one standard deviation uncorrelated error on data point i, from both statistical and systematic sources. The experiments use this χ^2 in different ways. ZEUS uses the Offset method and H1 uses the Hessian method.

Traditionally, experimentalists have used 'Offset' methods to account for correlated systematic errors. The χ^2 is formluated without any terms due to correlated systematic errors ($s_\lambda = 0$ in Eq. 5) for evaluation of the central values of the fit parameters. However, the data points are then offset to account for each source of systematic error in turn (i.e. set $s_\lambda = +1$ and then $s_\lambda = -1$ for each source λ) and a new fit is performed for each of these variations. The resulting deviations of the theoretical parameters from their central values are added in quadrature. (Positive and negative deviations are added in quadrature separately.) This method does not assume that the systematic uncertainties are Gaussian distributed. An equivalent (and much more efficient) procedure to perform the Offset method has been given by Pascaud and Zomer [20], and this is what is actually used. The Offset method is a conservative method of error estimation as compared to the Hessian method. It gives fitted theoretical predictions which are as close as possible to the central values of the published data. It does not use the full statistical power of the fit to improve the estimates of s_λ, since it choses to mistrust the systematic error estimates, but it is correspondingly more robust.

The Hessian method is an alternative procedure in which the systematic uncertainty parameters s_λ are allowed to vary in the main fit when determining the values of the theoretical parameters. Effectively, the theoretical prediction is not fitted to the central values of the published experimental data, but these data points are allowed to move collectively, according to their correlated systematic uncertainties. The theoretical prediction determines the optimal settings for correlated systematic shifts of experimental data points such that the most consistent fit to all data sets is obtained. Thus, in a global fit, systematic shifts in one experiment are correlated to those in another experiment by the fit. In essence one is allowing the theory to calibrate the detectors. This requires great confidence in the theory, but more significantly, it requires confidence in the many model choices which go into setting the boundary conditions for the theory (such as the parametrization at Q_0^2).

The ZEUS analysis can be performed using the Hessian method as well as the Offset method and Fig. 6 compares the PDFs, and their uncertainties, extracted from ZEUS data using these two methods. The central values of the different methods are in good agreement but the use of the Hessian method results in smaller uncertainties, for a the standard set of model assumptions, since the input data can be shifted within their correlated systematic uncertainties to suit the theory better. However, model uncertainties are more significant for the Hessian method than for the Offset method. The experimental uncertainty band for any one set of model choices is set by the usual χ^2 tolerance, $\Delta\chi^2 = 1$, but the acceptability of a different set of choices is judged by the hypothesis testing criterion, such that the χ^2 should be approximately in the range $N \pm \sqrt{(\ N)}$, where N is the number of degrees of freedom. The PDF parameters obtained for the different model choices can differ by much more than their experimental uncertainties, because each model choice can result in somewhat different values of the systematic

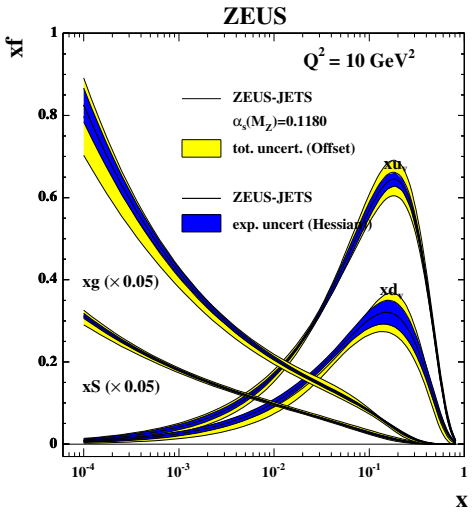

Fig. 6: PDFs at $Q^2 = 10\text{GeV}^2$, for the ZEUS analysis of ZEUS data performed by the Offset and the Hessian methods.

uncertainty parameters, s_λ, and thus a different estimate of the shifted positions of the data points. This results in a larger spread of model uncertainty than in the Offset method, for which the data points cannot move. Fig 4 illustrates the comparability of the ZEUS (Offset) total uncertainty estimate to the H1 (Hessian) experimental plus model uncertainty estimate.

Another issue which arises in relation to the Hessian method is that the data points should not be shifted far outside their one standard deviation systematic uncertainties. This can indicate inconsistencies between data sets, or parts of data sets, with respect to the rest of the data. The CTEQ collaboration have considered data inconsistencies in their most recent global fit [4]. They use the Hessian method but they increase the resulting uncertainty estimates, by increasing the χ^2 tolerance to $\Delta\chi^2 = 100$ to allow for both model uncertainties and data inconsistencies. In setting this tolerance they have considered the distances from the χ^2-minima of individual data sets to the global minimum for all data sets. These distances by far exceed the range allowed by the $\Delta\chi^2 = 1$ criterion. Strictly speaking such variations can indicate that data sets are inconsistent but the CTEQ collaboration take the view that all of the current world data sets must be considered acceptable and compatible at some level, even if strict statistical criteria are not met, since the conditions for the application of strict criteria, namely Gaussian error distributions, are also not met. It is not possible to simply drop "inconsistent" data sets, as then the partons in some regions would lose important constraints. On the other hand the level of "inconsistency" should be reflected in the uncertainties of the PDFs. This is achieved by raising the χ^2 tolerance. This results in uncertainty estimates which are comparable to those achieved by using the Offset method [18].

3.4 Using both H1 and ZEUS data in the same PDF analysis

Using data from a single experiment avoids questions of data consistency, but to get the most information from HERA it is necessary to put ZEUS and H1 data sets into the same analysis together, and then questions of consistency arise. Fig 7 compares the sea and gluon PDFs and the u and d valence PDFs extracted from the ZEUS PDF analysis of ZEUS data alone, to those extracted from the ZEUS PDF analysis of both H1 and ZEUS data. It is noticeable that, for the low-x sea and gluon PDFs, combining the data sets does not bring a reduction in uncertainty equivalent to doubling the statistics. This is because the data which determine these PDFs are systematics limited. In fact there is some degree of tension between the ZEUS and the H1 data sets, such that the χ^2 per degree of freedom rises for both data sets when they are fitted together. The Offset method of treating the systematic errors reflects this

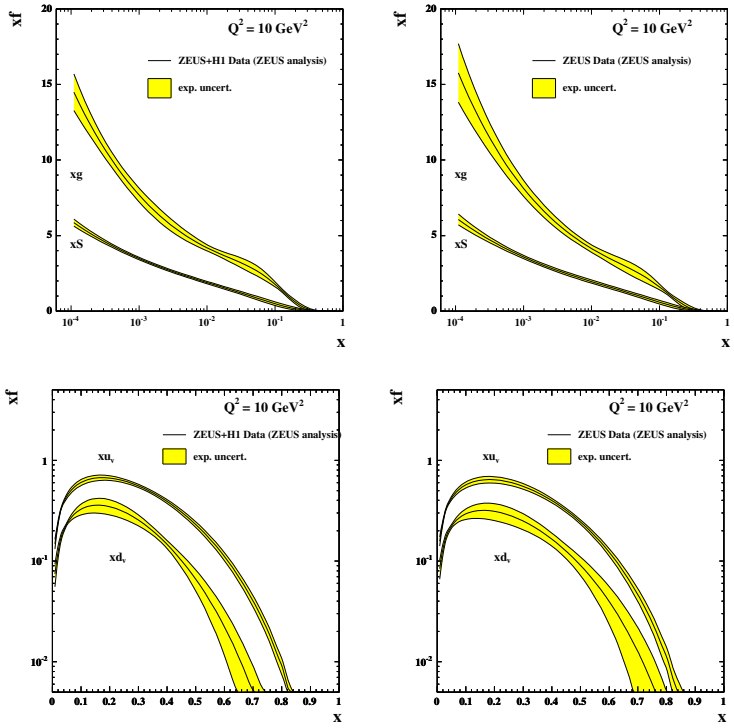

Fig. 7: Top plots: Sea and gluon distributions at $Q^2 = 10\text{GeV}^2$ extracted from H1 and ZEUS data using the ZEUS analysis (left) compared to those extracted from ZEUS data alone using the ZEUS analysis (right). Bottom Plots: Valence distributions at $Q^2 = 10\text{GeV}^2$, extracted from H1 and ZEUS data using the ZEUS analysis (left) compared to those extracted from ZEUS data alone using the ZEUS analysis (right).

tension such that the overall uncertainty is not much improved when H1 data are added to ZEUS data. However, the uncertainty on the high-x valence distributions is reduced by the input of H1 data, since the data are still statistics limited at high x.

3.5 Combining the H1 and ZEUS data sets before PDF analysis

Thus there could be an advantage in combining ZEUS and H1 data in a PDF fit if the tension between the data sets could be resolved. It is in this context the question of combining these data into a single data set arises. The procedure for combination is detailed in the contribution of S. Glazov to these proceedings (section 4). Essentially, since ZEUS and H1 are measuring the same physics in the same kinematic region, one can try to combine them using a 'theory-free' Hessian fit in which the only assumption is that there is a true value of the cross-section, for each process, at each x, Q^2 point. The systematic uncertainty parameters, s_λ, of each experiment are fitted to determine the best fit to this assumption. Thus each experiment is calibrated to the other. This works well because the sources of systematic uncertainty in each experiment are rather different. Once the procedure has been performed the resulting systematic uncertainties on each of the combined data points are significantly smaller than the statistical errors. Thus one can legitimately make a fit to the combined data set in which these statistical and systematic uncertainties are simply combined in quadrature. The result of making such a fit, using the ZEUS analysis, is shown in Fig. 8. The central values of the ZEUS and H1 published analyses are also shown for comparison. Looking back to Fig. 7 one can see that there has been a dramatic reduction in the level of uncertainty compared to the ZEUS Offset method fit to the separate ZEUS and H1 data sets. This result is very promising. A preliminary study of model dependence, varying the form of the polynomial, $P(x)$, used in the PDF paremtrizations at Q_0^2, also indicates that model dependence is relatively small.

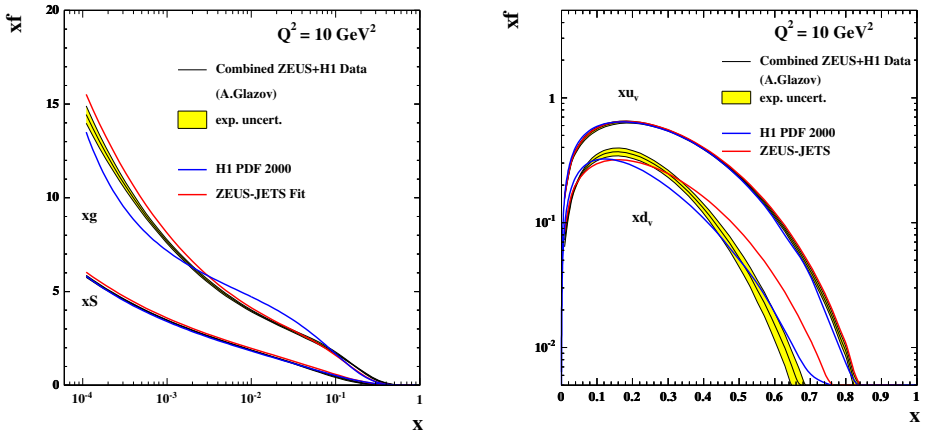

Fig. 8: Left plot: Sea and gluon distributions at $Q^2 = 10\text{GeV}^2$, extracted from the combined H1 and ZEUS data set using the ZEUS analysis. Right plot: Valence distributions at $Q^2 = 10\text{GeV}^2$, extracted from the combined H1 and ZEUS data set using the ZEUS analysis.

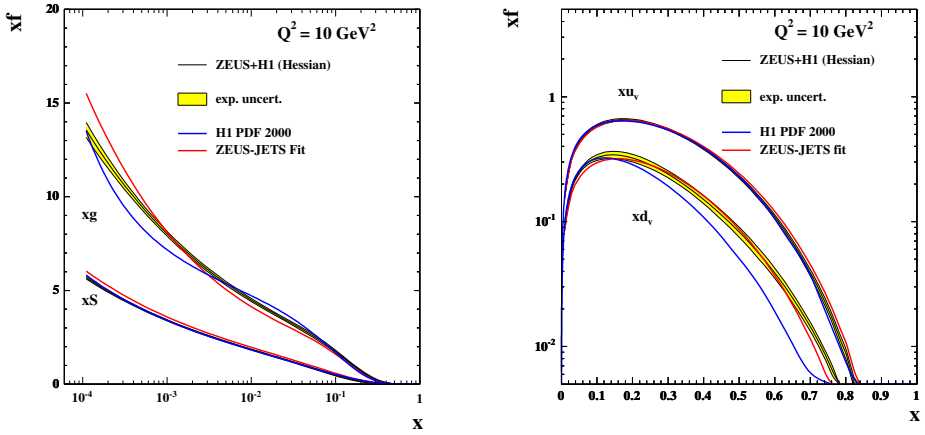

Fig. 9: Left plot: Sea and gluon distributions at $Q^2 = 10\text{GeV}^2$, extracted from the H1 and ZEUS data sets using the ZEUS analysis done by Hessian method. Right plot: Valence distributions at $Q^2 = 10\text{GeV}^2$, extracted from the H1 and ZEUS data sets using the ZEUS analysis done by Hessian method.

The tension between ZEUS and H1 data could have been resolved by putting them both into a PDF fit using the Hessian method to shift the data points. That is, rather than calibrating the two experiments to each other in the 'theory-free' fit, we could have used the theory of pQCD to calibrate each experiment. Fig. 9 shows the PDFs extracted when the ZEUS and H1 data sets are put through the ZEUS PDF analysis procedure using the Hessian method. The uncertainties on the resulting PDFs are comparable to those found for the fit to the combined data set, see Fig. 8. However, the central values of the resulting PDFs are rather different- particularly for the less well known gluon and d valence PDFs. For both of the fits shown in Figs. 8 and 9 the values of the systematic error parameters, s_λ, for each experiment have been allowed to float so that the data points are shifted to give a better fit to our assumptions, but the values of the systematic error parameters chosen by the 'theory-free' fit and by the PDF fit are rather different. A representaive sample of these values is given in Table 3. These discrepancies might be somewhat alleviated by a full consideration of model errors in the PDF fit, or of appropriate χ^2 tolerance when combining the ZEUS and H1 experiments in a PDF fit, but these differences should make us wary about the uncritical use of the Hessian method.

Table 3: Systematic shifts for ZEUS and H1 data as determine by a joint pQCD PDF fit, and as determined by the theory-free data combination fit

Syatematic uncertainty s_λ	in PDF fit	in Theory-free fit
ZEUS electron efficiency	1.68	0.31
ZEUS electron angle	-1.26	-0.11
ZEUS electron energy scale	-1.04	0.97
ZEUS hadron calorimeter energy scale	1.05	-0.58
H1 electron energy scale	-0.51	0.61
H1 hadron energy scale	-0.26	-0.98
H1 calorimeter noise	1.00	-0.63
H1 photoproduction background	-0.36	0.97

4 Averaging of DIS Cross Section Data [5]

The QCD fit procedures (Alekhin [6], CTEQ [4], MRST [5], H1 [11], ZEUS [9]) use data from a number of individual experiments directly to extract the parton distribution functions (PDF). All programs use both the central values of measured cross section data as well as information about the correlations among the experimental data points.

The direct extraction procedure has several shortcomings. The number of input datasets is large containing several individual publications. The data points are correlated because of common systematic uncertainties, within and also across the publications. Handling of the experimental data without additional expert knowledge becomes difficult. Additionally, as it is discussed in Sec. 3, the treatment of the correlations produced by the systematic errors is not unique. In the Lagrange Multiplier method [20] each systematic error is treated as a parameter and thus fitted to QCD. Error propogation is then used to estimate resulting uncertainties on PDFs. In the so-called "offset" method (see e.g. [9]) the datasets are shifted in turn by each systematic error before fitting. The resulting fits are used to form an envelope function to estimate the PDF uncertainty. Each method has its own advantages and shortcomings, and it is difficult to select the standard one. Finally, some global QCD analyses use non-statistical criteria to estimate the PDF uncertainties ($\Delta\chi^2 \gg 1$). This is driven by the apparent discrepancy between different experiments which is often difficult to quantify. Without a model independent consistency check of the data it might be the only safe procedure.

These drawbacks can be significantly reduced by averaging of the input structure function data in a model independent way before performing a QCD analysis of that data. One combined dataset of deep inelastic scattering (DIS) cross section measurements is much easier to handle compared to a scattered set of individual experimental measurements, while retaining the full correlations between data points. The averaging method proposed here is unique and removes the drawback of the offset method, which fixes the size of the systematic uncertainties. In the averaging procedure the correlated systematic uncertainties are floated coherently allowing in some cases reduction of the uncertainty. In addition, study of a global $\chi^2/\ dof$of the average and distribution of the pulls allows a model independent consistency check between the experiments. In case of discrepancy between the input datasets, localised enlargement of the uncertainties for the average can be performed.

A standard way to represent a cross section measurement of a single experiment is given in the case of the F_2 structure function by:

$$\chi^2_{exp}(\left\{F_2^{i,\,ture}\right\},\{\alpha_j\}) \;=\; \sum_i \frac{\left[F_2^{i,\,ture} - \left(F_2^i + \sum_j \frac{\partial F_2^i}{\partial \alpha_j}\alpha_j\right)\right]^2}{\sigma_i^2} + \sum_j \frac{\alpha_j^2}{\sigma^2_{\alpha_j}}. \qquad (6)$$

Here F_2^i (σ_i^2) are the measured central values (statistical and uncorrelated systematic uncertainties) of the

[5] Contributing author: A. Glazov

F_2 structure function[6], α_j are the correlated systematic uncertainty sources and $\partial F_2^i / \partial \alpha_j$ are the sensitivities of the measurements to these systematic sources. Eq. 6 corresponds to the correlated probability distribution functions for the structure function $F_2^{i,\,true}$ and for the systematic uncertainties α_j. Eq. 6 resembles Eq. 5 where the theoretical predictions for F_2 are substituted by $F_2^{i,\,true}$.

The χ^2 function Eq. 6 by construction has a minimum $\chi^2 = 0$ for $F_2^{i,\,true} = F_2^i$ and $\alpha_j = 0$. One can show that the total uncertainty for $F_2^{i,\,true}$ determined from the formal minimisation of Eq. 6 is equal to the sum in quadrature of the statistical and systematic uncertainties. The reduced covariance matrix $cov(F_2^{i,\,true}, F_2^{j,\,true})$ quantifies the correlation between experimental points.

In the analysis of data from more than one experiment, the χ^2_{total} function is taken as a sum of the χ^2 functions Eq. 6 for each experiment. The QCD fit is then performed in terms of parton density functions which are used to calculate predictions for $F_2^{i,\,true}$.

Before performing the QCD fit, the χ^2_{total} function can be minimised with respect to $F_2^{i,\,true}$ and α_j. If none of correlated sources is present, this minimisation is equivalent to taking an average of the structure function measurements. If the systematic sources are included, the minimisation corresponds to a generalisation of the averaging procedure which contains correlations among the measurements.

Being a sum of positive definite quadratic functions, χ^2_{total} is also a positive definite quadratic and thus has a unique minimum which can be found as a solution of a system of linear equations. Although this system of the equations has a large dimension it has a simple structure allowing fast and precise solution.

A dedicated program has been developed to perform this averaging of the DIS cross section data (http://www.desy.de/~h1zeus/herapdf/averaging.html.gz). This program can calculate the simultaneous averages for neutral current (NC) and charged current (CC) electron- and positron-proton scattering cross section data including correlated systematic sources. The output of the program includes the central values and uncorrelated uncertainties of the average cross section data. The correlated systematic uncertainties can be represented in terms of (i) covariance matrix, (ii) dependence of the average cross section on the original systematic sources together with the correlation matrix for the systematic sources, (iii) and finally the correlation matrix of the systematic sources can be diagonalised, in this case the form of χ^2 for the average data is identical to Eq. 6 but the original systematic sources are not preserved.

The first application of the averaging program has been a determination of the average of the published H1 and ZEUS data [3, 11, 21–28]. Nine individual NC and CC cross section measurements are included from H1 and seven are included from ZEUS. Several sources of systematic uncertainties are correlated between datasets, the correlations among H1 and ZEUS datasets are taken from [11] and [10], respectively. No correlations are assumed between H1 and ZEUS systematic uncertainties apart from a common $0.5\ \%$ luminosity measurement uncertainty. The total number of data points is 1153 (552 unique points) and the number of correlated systematic sources, including normalisation uncertainties, is 43.

The averaging can take place only if most of the data from the experiments are quoted at the same Q^2 and x values. Therefore, before the averaging the data points are interpolated to a common Q^2, x grid. This interpolation is based on the H1 PDF 2000 QCD fit [11]. The interpolation of data points in principle introduces a model dependency. For H1 and ZEUS structure function data both experiments employ rather similar Q^2, x grids. About $2\ 0\%$ of the input points are interpolated, for most of the cases the correction factors are small (few percent) and stable if different QCD fit parametrizations [4, 5] are used.

The cross section data have also been corrected to a fixed center of mass energy squared $S = 1\ 01\ 5\,\mathrm{GeV}^2$. This has introduced a small correction for the data taken at $S = 9\,05\ 3\,\mathrm{GeV}^2$. The correction is based on H1-2000 PDFs, it is only significant for high inelasticity $y > 0.6$ and does not exceed 6%.

[6]The structure function is measured for different Q^2 (four momentum transfer squared) and Bjorken-x values which are omitted here for simplicity.

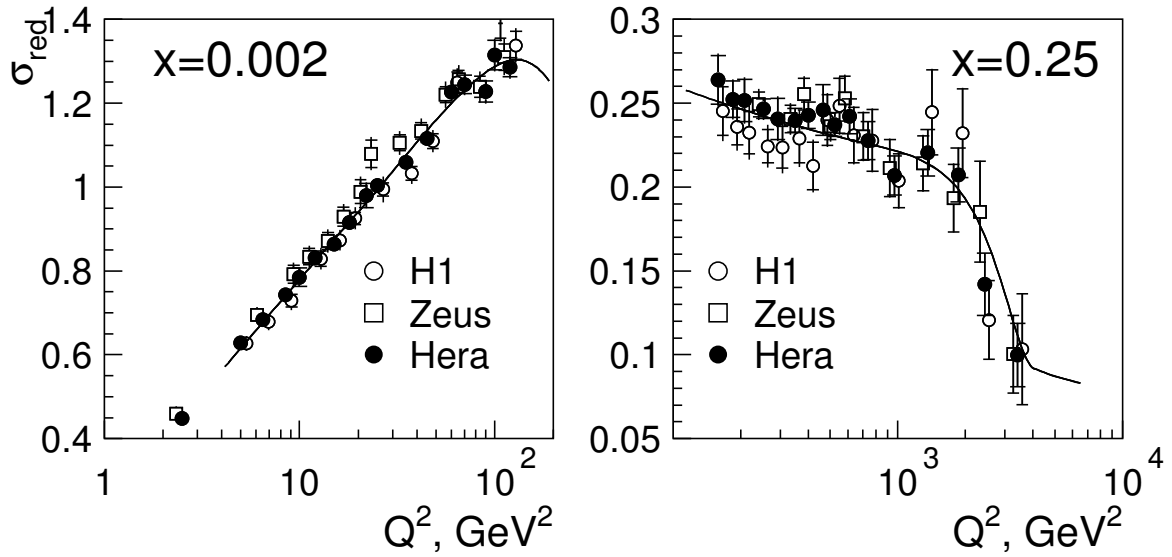

Fig. 10: Q^2 dependence of the NC reduced cross section for $x = 0.002$ and $x = 0.25$ bins. H1 data is shown as open circles, ZEUS data is shown as open squares and the average of H1 and ZEUS data is shown as filled circles. The line represents the expectation from the H1 PDF 2000 QCD fit.

The HERA data sets agree very well: $\chi^2/\ dof$ for the average is $5\ \ 2/601$. The distribution of pulls does not show any significant tensions across the kinematic plane. Some systematic trends can be observed at low $Q^2 < 5\ \ 0\text{GeV}^2$, where ZEUS NC data lie systematically higher than the H1 data, although this difference is within the normalisation uncertainty. An example of the resulting average DIS cross section is shown in Fig. 10, where the data points are displaced in Q^2 for clarity.

A remarkable side feature of the averaging is a significant reduction of the correlated systematic uncertainties. For example the uncertainty on the scattered electron energy measurement in the H1 backward calorimeter is reduced by a factor of three. The reduction of the correlated systematic uncertainties thus leads to a significant reduction of the total errors, especially for low $Q^2 < 1\ \ 0\text{GeV}^2$, where systematic uncertainties limit the measurement accuracy. For this domain the total errors are often reduced by a factor two compared to the total errors of the individual H1 and ZEUS measurements.

The reduction of the correlated systematic uncertainties is achieved since the dependence of the measured cross section on the systematic sources is significantly different between H1 and ZEUS experiments. This difference is due mostly to the difference in the kinematic reconstruction methods used by the two collaborations, and to a lesser extent to the individual features of the H1 and ZEUS detectors. For example, the cross section dependence on the scattered electron energy scale has a very particular behaviour for H1 data which relies on kinematic reconstruction using only the scattered electron in one region of phase space. ZEUS uses the double angle reconstruction method where the pattern of this dependence is completely different leading to a measurement constraint.

In summary, a generalised averaging procedure to include point-to-point correlations caused by the systematic uncertainties has been developed. This averaging procedure has been applied to H1 and ZEUS DIS cross section data. The data show good consistency. The averaging of H1 and ZEUS data leads to a significant reduction of the correlated systematic uncertainties and thus a large improvement in precision for low Q^2 measurements. The goal of the averaging procedure is to obtain HERA DIS cross section set which takes into account all correlations among the experiments.

5 The longitudinal structure function F_L [7]

5.1 Introduction

At low x the sea quarks are determined by the accurate data on $F_2(x, Q^2)$. The charm contribution to F_2 is directly measured while there is no separation of up and down quarks at low x which are assumed to have the same momentum distribution, see [2]. Within this assumption, and setting the strange sea to be a fraction of the up/down sea, the proton quark content at low x is determined. The gluon distribution $xg(x, Q^2)$, however, is determined only by the derivative $\partial F_2 / \partial \ln Q^2$ which is not well measured [3]. It is thus not surprising that rather different gluon distributions are obtained in global NLO analyses, as is illustrated in Figure 11. The figure displays the result of recent fits by MRST and CTEQ on the gluon distribution at low and high Q^2. It can be seen that there are striking differences at the initial scale, $Q^2 = 5\,\mathrm{GeV}^2$, which at high Q^2 get much reduced due to the evolution mechanism. The ratio of these distributions, however, exhibits differences at lower x at the level of 10% even in the LHC Higgs and W production kinematic range, see Figure 12. One also observes a striking problem at large x which is beyond the scope of this note, however. In a recent QCD analysis it was observed [3] that the dependence of the gluon distribution at low x, $xg \propto x^{b_G}$, is correlated to the value of $\alpha_s(M_Z^2)$, see Figure 13.

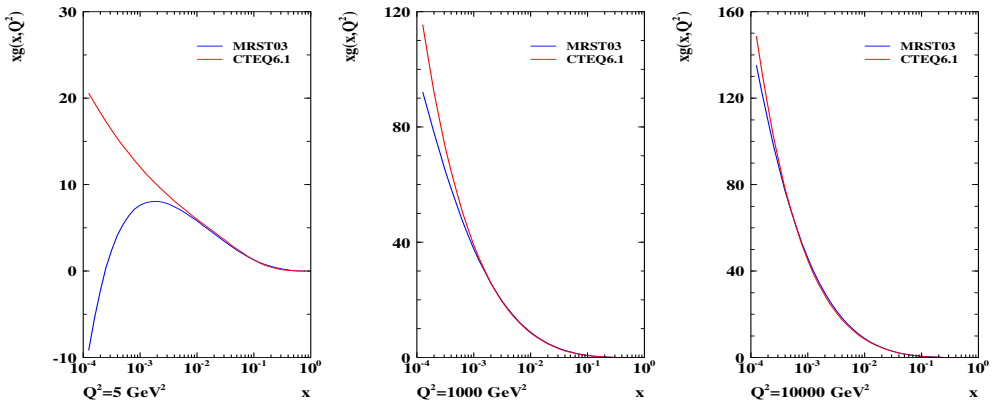

Fig. 11: Gluon momentum distributions determined by MRST and CTEQ in NLO QCD, as a function of x for $Q^2 = 5\,\mathrm{GeV}^2$, close to the initial scale of the fits, and at higher Q^2 as the result of the DGLAP evolution.

In the Quark-Parton Model the longitudinal structure function $F_L(x, Q^2)$ is zero [29]. In DGLAP QCD, to lowest order, F_L is given by [30]

$$F_L(x, Q^2) = \frac{\alpha_s}{4\pi} x^2 \int_x^1 \frac{dz}{z^3} \cdot \left[\frac{16}{3} F_2(z, Q^2) + 8 \sum e_q^2 \left(1 - \frac{x}{z}\right) z g(z, Q^2) \right] \tag{7}$$

with contributions from quarks and from gluons. Approximately this equation can be solved [31] and the gluon distribution appears as a measurable quantity,

$$xg(x) = 1.8 \left[\frac{3\pi}{2\alpha_s} F_L(\ldots x) - F_2(\ldots x) \right] \simeq \frac{8.3}{\alpha_s} F_L, \tag{8}$$

determined by measurements of F_2 and F_L . Since F_L , at low x, is not much smaller than F_2 , to a good approximation F_L is a direct measure for the gluon distribution at low x.

Apart from providing a very useful constraint to the determination of the gluon distribution, see also Sect. 7, a measurement of $F_L(x, Q^2)$ is of principal theoretical interest. It provides a crucial test of QCD to high orders. A significant departure of an F_L measurement from the prediction which is based on the measurement of $F_2(x, Q^2)$ and $\partial F_2 / \partial \ln Q^2$ only, would require theory to be modified. There are known reasons as to why the theoretical description of gluon radiation at low x may differ

[7] Contributing authors: J. Feltesse, M. Klein

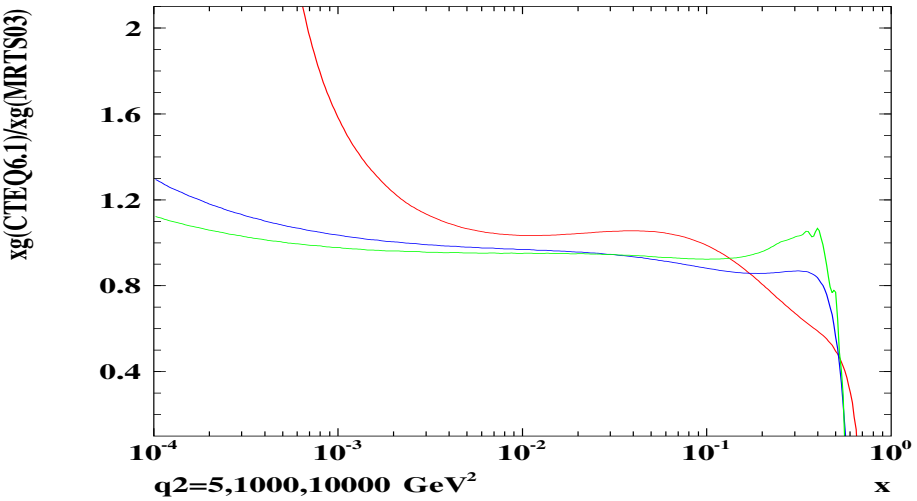

Fig. 12: Ratio of the gluon distributions of CTEQ to MRST as a function of x for low and large Q^2.

from conventional DGLAP evolution: the neglect of $\ln(1/x)$, in contrast to BFKL evolution, or the importance of NLL resummation effects on the gluon splitting function (see [32]). Furthermore recent calculations of deep inelastic scattering to NNLO predict very large effects from the highest order on F_L contrary to F_2 [33].

Within the framework of the colour dipole model there exists a testable prediction for $F_L(x, Q^2)$, and the longitudinal structure function, unlike F_2, may be subject to large higher twist effects [34].

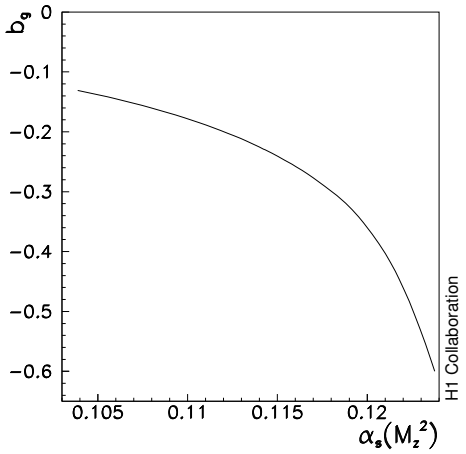

Fig. 13: Correlation of the low x behaviour of the gluon distribution, characterised by the power x^{-b_g}, with the strong coupling constant α_s as obtained in the H1 NLO QCD fit to H1 and BCDMS data.

5.2 Indirect Determinations of F_L at Low x

So far first estimates on $F_L(x, Q^2)$ at low x have been obtained by the H1 Collaboration. These result from data on the inclusive $ep \to eX$ scattering cross section

$$\frac{Q^4 x}{2\pi\alpha^2 Y_+} \cdot \frac{d^2\sigma}{dx dQ^2} = [F_2(x, Q^2) - f(y) \cdot F_L(x, Q^2) \quad \mp \sigma_r \tag{9}$$

obtained at fixed, large energy, $s = 4E_e E_p$. The cross section is defined by the two proton structure functions, F_2 and F_L, with $Y_+ = 1 + (\quad \mp y)^2$ and $f(y) = y^2/Y_+$. At fixed s the inelasticity y is

fixed by x and Q^2 as $y = Q^2/\ sx$ Thus one can only measure a combination $F_2 - f(y)F_L$. Since HERA accesses a large range of y, and $f(y)$ is large only at large $y > 0.4$, assumptions have been made on F_L to extract F_2 at larger y. Since the cross section measurement accuracy has reached the few per cent level [3], the effect of the F_L assumption on F_2 at lowest x has been non-negligible. The determination of $F_2(x, Q^2)$ has thus been restricted to a region in which $y < 0.6$. The proton structure function $F_2(x, Q^2)$ is known over a few orders of magnitude in x rather well, from HERA and at largest x from fixed target data. Thus H1 did interpret the cross section at higher y as a determination of $F_L(x, Q^2)$ imposing assumptions about the behaviour of $F_2(x, Q^2)$ at lowest x. These were derived from QCD fits to the H1 data [35] or at lower Q^2, where QCD could not be trusted, from the derivative of F_2 [36]. Recently, with the established x behaviour [37] of $F_2(x, Q^2) = c(Q^2)x^{-\lambda(Q^2)}$, a new method [36] has been used to determine F_L . This "shape method" is based on the observation that the shape of σ_r, Eq. 9, at high y is driven by $f \propto y^2$ and sensitivity to F_L is restricted to a very narrow range of x corresponding to $y = 0.3 - 0.9$. Assuming that $F_L(x, Q^2)$ in this range, for each bin in Q^2, does not depend on x, one obtains a simple relation, $\sigma_r = c\,x^{-\lambda} - fF_L$. which has been used to determine $F_L(x, Q^2)$. Figure 14 shows the existing, preliminary data on $F_L(x, Q^2)$ at low Q^2 from the

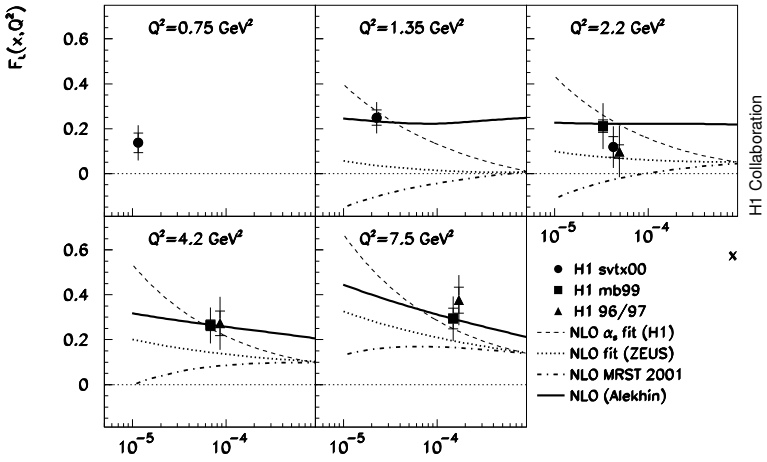

Fig. 14: Data on the longitudinal structure function obtained using assumptions on the behaviour of the other structure function F_2 in comparison with NLO QCD fit predictions. The data labeled svtx00 and mb99 data are preliminary.

H1 Collaboration in comparison with predictions from NLO DGLAP QCD fits to HERA and further cross section data. One can see that the accuracy and the x range of these $F_L(x, Q^2)$ determinations are rather limited although the data have some discriminative power already.

5.3 Backgrounds and Accuracy

The longitudinal structure function contribution to σ_r represents a small correction of the cross section in a small part of the kinematic range only. The demands for the F_L measurement are extremely high: the cross section needs to be measured at the per cent level and the scattered electron be uniquely identified up to high y. The method of unfolding F_2 and F_L consists in a measurement of σ_r at fixed x and Q^2 with varying s. This allows both structure functions to be determined from a straight line variation of σ_r as a function of $f(y)$, see [38].

At large y, corrresponding to low x, and low Q^2 the scattering kinematics at HERA resembles that of a fixed target scattering experiment: the electron scattered off quarks at very low x ("at rest") is going in the backward detector region, i.e. in the direction of the electron beam. The scattered electron is accompanied by part of the hadronic final state which is related to the struck quark. High inelasticities $y \simeq 1 - E'_e/\ E_e$ demand to identify scattered electrons down to a few GeV of energy E'_e. Thus a

considerable background is to be isolated and removed which stems from hadrons or photons, from the $\pi_0 \to \gamma\gamma$ decay. These particles may originate both from a genuine DIS event but to a larger extent stem from photoproduction processes, in which the scattered electron escapes mostly non recognised in electron beam direction. Removal of this background in H1 is possible by requiring a track associated to the Spacal cluster, which rejects photons, and by measuring its charge which on a statistical basis removes the remaining part of the background as was demonstrated before [3, 36].

The scattered electron kinematics, E'_e and θ_e, can be accurately reconstructed using the high resolution Spacal calorimeter energy determination and the track measurements in the Backward Silicon Tracker (BST) and the Central Jet Drift Chamber (CJC). Reconstruction of the hadronic final state allows the energy momentum constraint to be imposed, using the "$E - p_z$" cut, which removes radiative corrections, and the Spacal energy scale to be calibrated at large E'_e using the double angle method. At low energies E'_e the Spacal energy scale can be calibrated to a few % using the π_0 mass constraint and be cross checked with the BST momentum measurement and with QED Compton events. The luminosity is measured to 1-2%. Any common normalisation uncertainty may be removed, or further constrained, by comparing cross section data at very low y where the contribution of F_L is negligible.

Subsequently two case studies are presented which illustrate the potential of measuring F_L directly in unfolding it from the large F_2 contribution to the cross section, a study using a set of 3 low proton beam energies and a simulation for just one low Ep data set combined with standard 920 GeV data. Both studies use essentially the same correlated systematic errors and differ slightly in the assumptions on the background and efficiency uncertainties which regard the errors on cross section ratios. The following assumptions on the correlated systematics are used: $\delta E'_e / E'_e = 0.003$ at large E_e linearly rising to 0.03 at 3 GeV; $\delta\theta_e = 0.2$ mrad in the BST acceptance region and 1 mrad at larger angles; $\delta E_h / E_h = 0.02$. These and further assumed systematic uncertainties represent about the state of analysis reached so far in inclusive low Q^2 cross section measurements of H1.

5.4 Simulation Results

A simulation has been performed for $E_e = 27.6$ GeV and for four different proton beam energies, $E_p = 920, 575, 465$ and 400 GeV assuming luminosities of 10, 5, 3 and 2 pb^{-1}, respectively. The beam energies are chosen such that the cross section data are equidistant in $f(y)$. If the luminosity scales as expected as E_p^2, the low E_p luminosities are equivalent to 35 pb^{-1} at standard HERA settings. Further systematic errors regard the residual radiative corrections, assumed to be 0.5%, and the photoproduction background, 1-2% depending on y. This assumption on the background demands an improvement by a factor of about two at high y which can be expected from a high statistics subtraction of background using the charge assignment of the electron scattering candidate. An extra uncorrelated efficiency correction is assumed of 0.5%. The resulting cross section measurements are accurate to 1-2%. For each Q^2 and x point this choice provides up to four cross section measurements. The two structure functions are then obtained from a fit to $\sigma_r = F_2 + f(y)F_L$ taking into account the correlated systematics. This separation provides also accurate data of F_2, independently of F_L . The simulated data on F_L span nearly one order of magnitude in x and are shown in Figure 15. For the chosen luminosity the statistical and systematic errors on F_L are of similar size. The overall accuracy on $F_L(x, Q^2)$, which may be obtained according to the assumed experimental uncertainties, is thus estimated to be of the order of 10-20%.

Based on recent information about aspects of the machine conditions in a low proton beam energy mode, a further case study was performed [39] for only one reduced proton beam energy. In this simulation, for the standard electron beam energy of $E_e = 27.6$ GeV, proton beam energies of $E_p = 920$ and 460 GeV were chosen with luminosities of 30 and 10 pb^{-1}, respectively. According to [40] it would take about three weeks to change the configuration of the machine and to tune the luminosity plus 10 weeks to record 10 pb^{-1} of good data with High Voltage of trackers on. Uncertainties besides the correlated errors specified above are assumed for photo-production background subtraction varying from 0% at y=0.65 to 4% at y = 0.9, and of 0.5% for the residual radiative corrections. An overall uncertainty of 1% is assumed

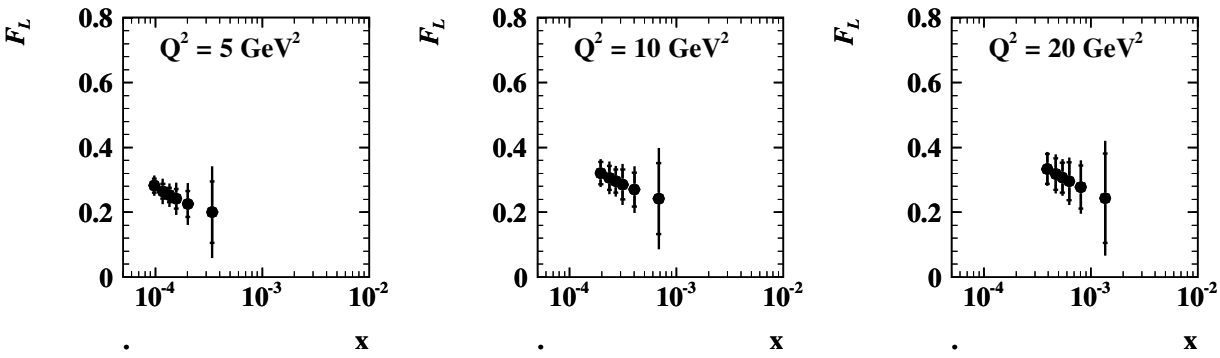

Fig. 15: Simulated measurement of the longitudinal structure function $F_L(x, Q^2)$ using the H1 backward apparatus to reconstruct the scattered electron up to maximum inelasticities of $y = 0.9$ corresponding to a mimimum electron energy of E'_e of about 3 GeV. The inner error bar is the statistical error. The full error bar denotes the statistical and systematic uncertainty added in quadrature.

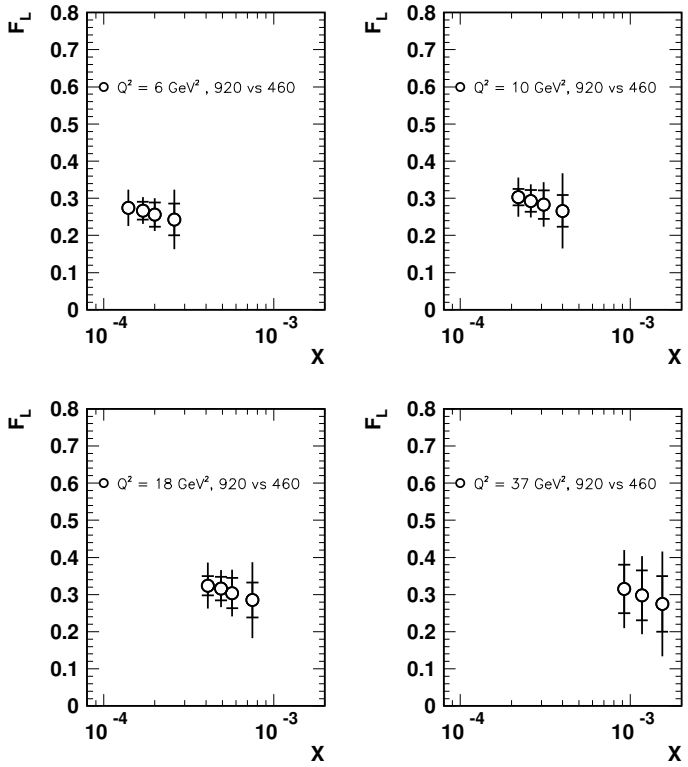

Fig. 16: Simulated measurement of the longitudinal structure function $F_L(x, Q^2)$ for data at 920 GeV (30 pb^{-1}) and 460 GeV (10 pb^{-1}). The inner error bar is the statistical error. The full error bar denotes the statistical and systematic uncertainty added in quadrature.

on the measurement of the cross section at low beam energy settings, which covers relative uncertainties on electron identification, trigger efficiency, vertex efficiency, and relative luminosity.

To evaluate the errors two independent methods have been considered an analytic calculation and a fast Monte-Carlo simulation technique. The two methods provide statistical and systematic errors which are in excellent agreement. The overall result of this simulation of F_L is displayed in Figure 16. In many bins the overall precision on $F_L(x, Q^2)$ is around or below 20%. It is remarkable that the overall precision would stay below 25% even if the statistical error or the larger source of systematic uncertainty would turn out to be twice larger than assumed to be in this study.

5.5 Summary

It has been demonstrated with two detailed studies that a direct measurement of the longitudinal structure function $F_L(x, Q^2)$ may be performed at HERA at the five sigma level of accuracy, in the x range from $1\ 0^4$ to $1\ 0^3$ in four bins of Q^2. This measurement requires about three months of running and tuning time at reduced proton beam energy. In addition it would provide the first measurement of the diffractive longitudinal structure function at the three sigma level (see the contribution of P. Newman in the summary of Working Group 4). The exact choice of the parameters of such a measurement are subject to further studies. In conclusion an accurate measurement of $F_L(x, Q^2)$ is feasible, it requires efficient detectors, dedicated beam time and analysis skills. It would be the right sign of completion to have measured F_2 first, in 1992 and onwards, and to conclude the HERA data taking with a dedicated measurement of the second important structure function $F_L(x, Q^2)$, which is related to the gluon density in the low x range of the LHC.

6 Determination of the Light Quark Momentum Distributions at Low x at HERA [8]

Based on the data taken in the first phase of HERA's operation (1993-2000), the HERA collider experiments have measured a complete set of neutral (NC) and charged (CC) current double differential $e^{\pm}p$ inclusive scattering cross sections, based on about $120\,\mathrm{pb}^{-1}$ of positron-proton and $15\,\mathrm{pb}^{-1}$ of electron-proton data. The NC and CC deep inelastic scattering (DIS) cross sections for unpolarised $e^{\pm}p$ scattering are determined by structure functions and quark momentum distributions in the proton as follows:

$$\sigma^{\pm}_{NC} \sim Y_+ F_2 \mp Y_- x F_3, \tag{10}$$

$$F_2 \simeq e_u^2 x(U + \overline{U}) + e_d^2 x(D + \overline{D}), \tag{11}$$

$$xF_3 \simeq 2x[a_u e_u (U - \overline{U}) + a_d e_d (D - \overline{D})\ ,] \tag{12}$$

$$\sigma^{+}_{CC} \sim x\overline{U} + (\ \ + y)^2 x D, \tag{13}$$

$$\sigma^{-}_{CC} \sim xU + (\ \ + y)^2 x\overline{D}. \tag{14}$$

Here $y = Q^2/\ sx$is the inelasticity, $s = 4E_eE_p$ and $Y_{\pm} = 1 \pm (\ \ + y)^2$. The parton distribution $U = u + c + b$ is the sum of the momentum distributions of the up-type quarks with charge $e_u = 2/3$ and axial vector coupling $a_u = 1/2$, while $D = d + s$ is the sum of the momentum distributions of the down type quarks with charge $e_d = -1/3$, $a_d = -1/2$. Similar relationships hold for the anti-quark distributions $\overline{U}$ and $\overline{D}$.

As is illustrated in Fig. 17 the H1 experiment [11] has determined all four quark distributions and the gluon distribution xg. The accuracy achieved so far by H1, for $x = 0.01, 0.4$ and 0.65, is $1\ \ \%3\ \%7\%$ for the sum of up quark distributions and $2\ \%\ 1\ \ 0\%3\ 0\%$for the sum of down quark distributions, respectively. The extracted parton distributions are in reasonable agreement with the results obtained in global fits by the MRST [5] and CTEQ [4] collaborations. The H1 result is also consistent with the pdfs determined by the ZEUS Collaboration [10] which uses jet data to improve the accuracy for the gluon distribution and imports a $\overline{d} - \overline{u}$ asymmetry fit result from MRST. New data which are being taken (HERA II) will improve the accuracy of these determinations further. At the time this is written, the available data per experiment have been grown to roughly $150\,\mathrm{pb}^{-1}$ for both e^+p and e^-p scattering, and more is still to come. These data will be particularly important to improve the accuracy at large x, which at HERA is related to high Q^2.

As is clear from the above equations, the NC and CC cross section data are sensitive directly to only these four quark distribution combinations. Disentangling the individual quark flavours (up, down, strange, charm and beauty) requires additional assumptions. While informations on the c and b densities are being obtained from measurements of $F_2^{c\overline{c}}$ and $F_2^{b\overline{b}}$ of improving accuracy, the determination of the strange quark density at HERA is less straightforward and may rest on $sW^+ \to c$ and strange (Φ) particle

[8] Contributing authors: M. Klein, B. Reisert

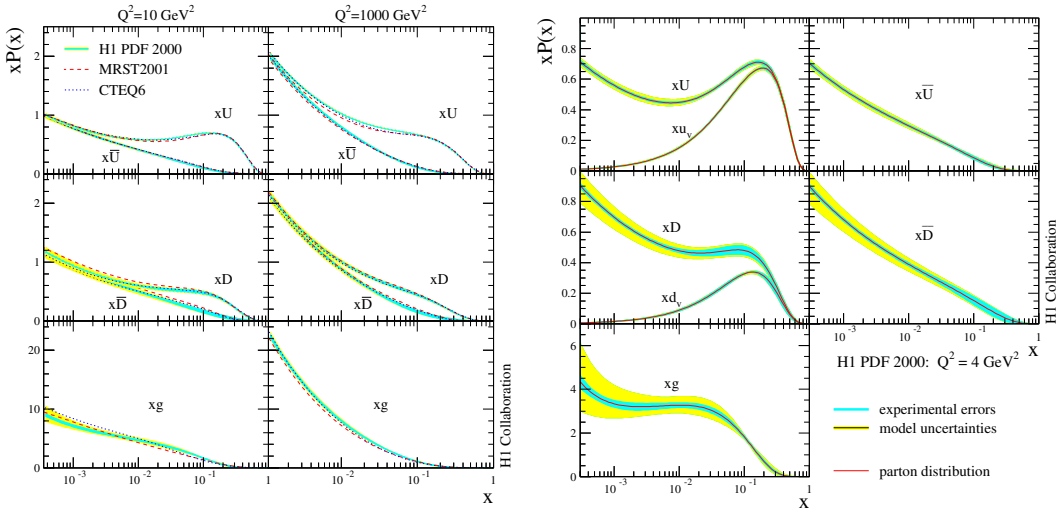

Fig. 17: Determination of the sum of up, anti-up, down and anti-downquark distributions and of the gluon distribution in the proton based on the H1 neutral and charged current cross section data. Left: for Q^2 of 10 and 1000 GeV^2 compared with results from MRST and CTEQ; Right: the parton distributions with their experimental and model uncertainties as determined by H1 at the starting scale $Q_0^2 = 4\,\mathrm{GeV}^2$.

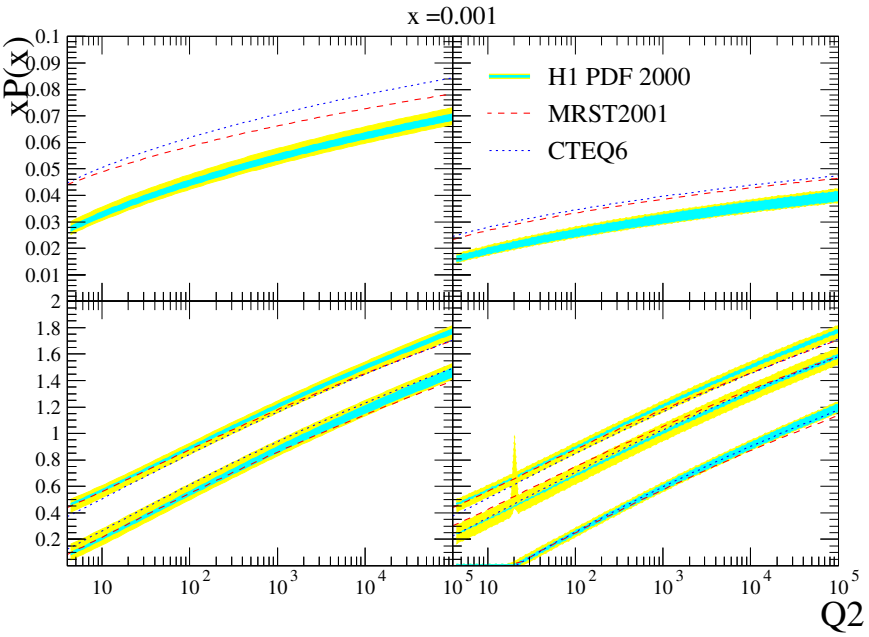

Fig. 18: Parton distributions and their uncertainties as determined by H1 extrapolated to the region of the LHC, for $x = 0.001$ near to the rapidity plateau. Top left: u valence; top right: d valence; bottom left: $\overline{u}$ and below c; bottom right, in decreasing order: $\overline{d}$, s, b. The results are compared with recent fits to global data by MRST and CTEQ. Note that at such small x the valence quark distributions are very small. With increasing Q^2 the relative importance of the heavy quarks compared to the light quarks increases while the absolute difference of the quark distributions is observed to be rather independent of Q^2. The beauty contribution to the cross section thus becomes sizeable, amounting to about 5% for $pp \to HW$.

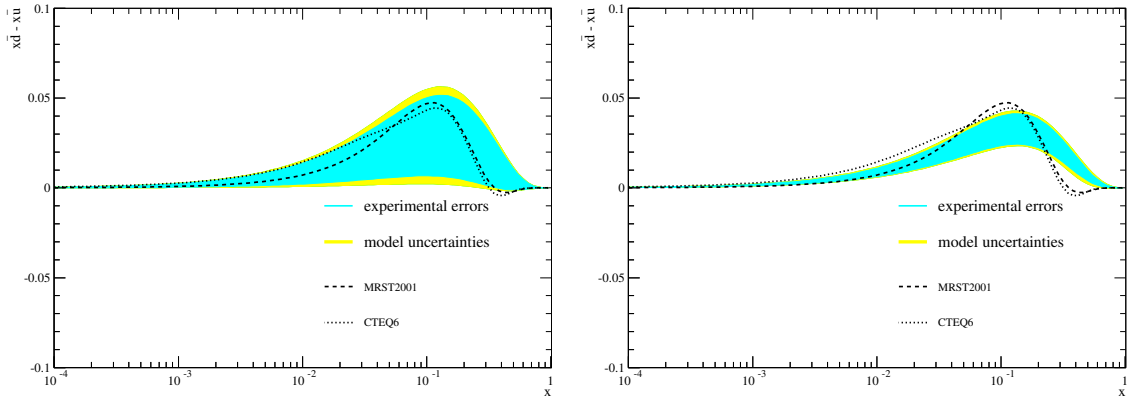

Fig. 19: Determination of the difference $x(\overline{d}-\overline{u})$ in the H1 PDF 2000 fit to the H1 data (left) and the H1 and the BCDMS μp and μD data (right). The sea quark difference is enforced to tend to zero at low x. The global fit results of MRST and CTEQ include Drell Yan data which suggest a sea quark asymmetry at $x \sim 0.1$.

production [41]. The relative contributions from the heavy quarks become increasingly important with Q^2, as is illustrated in Fig. 18.

The larger x domain is dominated by the valence quarks. At HERA the valence quark distributions are not directly determined but extracted from the differences $u_v = U - \overline{U}$ and $d_v = D - \overline{D}$. Note that this implies the assumption that sea and anti-quarks are equal which in non-perturbative QCD models may not hold. A perhaps more striking assumption is inherent in these fits and regards the sea quark asymmetries at low x which is the main subject of the subsequent discussion.

Fig. 19 shows the difference $x\overline{d} - x\overline{u}$ as determined in the H1 PDF 2000 fit based on the H1 data alone (left) and using in addition the BCDMS proton and deuteron data (right). One observes a trend of these fits to reproduce the asymmetry near $x \sim 0.1$ which in the MRST and CTEQ fits, shown in Fig. 19, is due to fitting the Drell Yan data from the E866/NuSea experiment [42]. While this enhancement is not very stable in the H1 fit [43] and not significant either, with the BCDMS data an asymmetry is observed which reflects the violation of the Gottfried sum rule.

In the H1 fit [11] the parton distributions at the initial scale $Q^2 = 4\,\mathrm{GeV}^2$ are parameterised as $xP = A_P x^{B_P}(1+x)^{C_P} \cdot f_P(x)$. The function f_P is a polynomial in x which is determined by requiring "χ^2 saturation" of the fits, i.e. starting from $f_P = 1$ additional terms $D_P x$, $E_P x^2$ etc. are added and only considered if they cause a significant improvement in χ^2, half integer powers were considered in [43]. The result for fitting the H1 data has been as follows: $f_g = (1 + D_g x)$, $f_U = (1 + D_U x + F_U x^3)$, $f_D = (1 + D_D x)$ and $f_{\overline{U}} = f_{\overline{D}} = 1$. The parton distributions at low x are thus parameterised as $xP \to A_P x^{B_P}$. The strange (charm) anti-quark distribution is coupled to the total amount of down (up) anti-quarks as $\overline{s} = f_c \overline{D}$ ($\overline{c} = f_c \overline{U}$). Two assumptions have been made on the behaviour of the quark and anti-quark distributions at low x. It has been assumed that quark and anti-quark distributions are equal and, moreover, that the sea is flavour symmetric. This implies that the slopes B of all four quark distributions are set equal $B_U = B_D = B_{\overline{U}} = B_{\overline{D}}$. Moreover, the nomalisations of up and down quarks are the same, i.e. $A_{\overline{U}}(1 + f_c) = A_{\overline{D}}(1 + f_s)$, which ensures that $\overline{d}/\overline{u} \to 1$ as x tends to zero. The consequence of this assumption is illustrated in Fig. 19. While the DIS data suggest some asymmetry at larger x, the up-down quark asymmetry is enforced to vanish at lower x. This results in a rather fake high accuracy in the determination of the four quark distributions at low x, despite the fact that at low x there is only one combination of them measured, which is $F_2 = x[4U(+\overline{U}) + (D + \overline{D})/9]$. If one relaxes both the conditions on the slopes and normalisations, the fit to the H1 data decides to completely remove the down quark contributions as is seen in Fig. 20 (left plot).

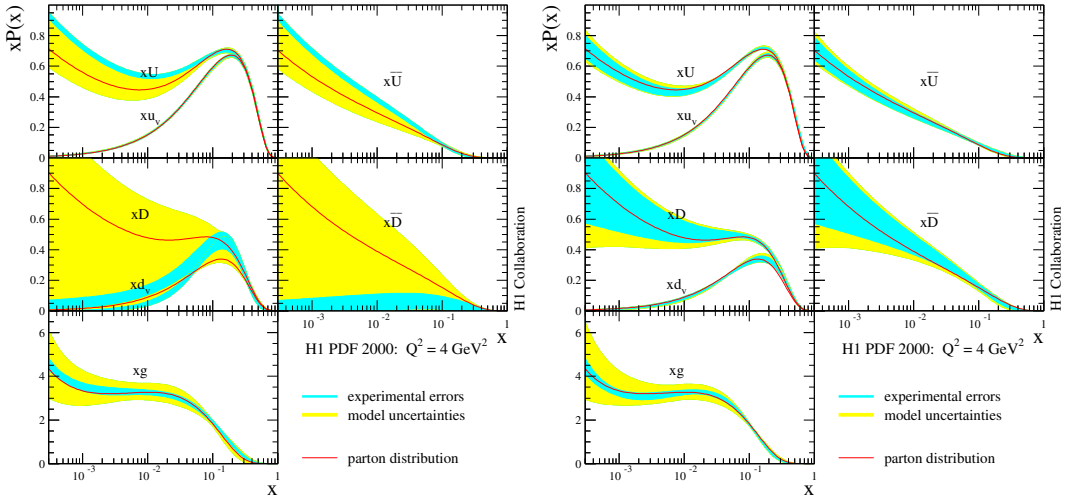

Fig. 20: Determinations of the quark and gluon momentum distributions releasing the constraint $x\overline{d} = x\overline{u}$ at low x, from the H1 NC and CC data alone (left) and from the H1 ep and the BCDMS μp and μD data (right). Since at low $x < 0.01$ there is no further constraint than that given from F_2 the uncertainties of $\overline{U}$ and in particular of $\overline{D}$ become sizeable.

In DIS the up and down quark asymmetry can be constrained using deuteron data because the nucleon structure function determines a different linear combination according to $F_2^N = 5x(U + \overline{U} + D + \overline{D})/1 \;\; \$ x(c + \overline{c} - s - \overline{s})/6$ with $N = (p+n)/2$. Unfortunately, there are only data at rather large x available. The effect of including the BCDMS data on the low x behaviour of the parton distributions is illustrated in Fig. 20 (right plot). It restores some amount of down quarks at low x , the errors, however, in particular of the down quarks, are still very large. The result is a large sea quark asymmetry uncertainty, which is shown in Fig. 21. At HERA a proposal had been made [44] to operate the machine in electron-deuteron mode. Measuring the behaviour at low x would not require high luminosity. Such data would constrain [9] a possible sea quark asymmetry with very high accuracy, as is also shown in Fig. 21.

Deuterons at HERA would require a new source and modest modifications to the preaccelerators. The H1 apparatus could be used in its standard mode with a forward proton detector added to take data at half the beam energy. Tagging the spectator protons with high accuracy at HERA, for the first time in DIS, one could reconstruct the electron-neutron scattering kinematics essentially free of nuclear corrections [44]. Since the forward scattering amplitude is related to diffraction one would also be able to constrain shadowing to the per cent level [47]. The low x measurements would require small luminosity amounts, of less than $50\,\mathrm{pb}^{-1}$. Long awaited constraints of the d/u ratio at large x and Q^2 would require extended running, as would CC data. Besides determining the parton distributions better, the measurement of the singlet F_2^N structure function would give important constraints on the evolution and theory at low x [48]. It would also result in an estimated factor of two improvement on the measurement of α_s at HERA [49]. For the development of QCD, of low x physics in particular, but as well for understanding physics at the LHC and also for superhigh energy neutrino astrophysics, HERA $e\,D$ data remain to be important.

[9] Constraints on the sea quark distributions may also be obtained from $W^+/\ W^-$ production at the TeVatron. However, the sensitivity is limited to larger $x \geq 0.1$ [45] since $W's$ produced in collisions involving sea quarks of smaller x will be boosted so strongly, that their decay products are not within the acceptance of the collider detectors. W^+ and W^- production at the LHC has been discussed in [46].

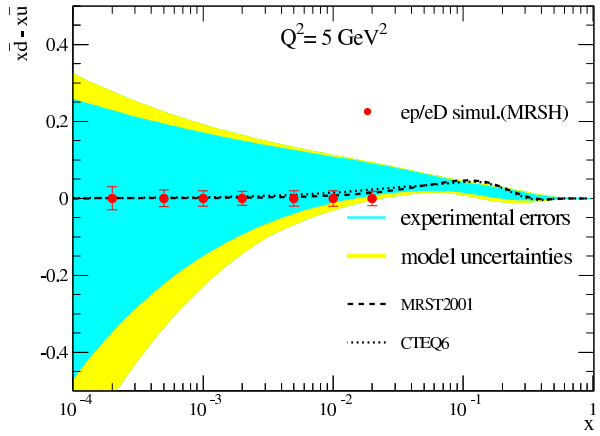

Fig. 21: Simulation of the difference of sea quark distributions, here assumed to be zero, at low x based on additional 20 pb^{-1} of electron-deuteron data at HERA. The error band represents the uncertainty of the H1 NLO QCD fit to the H1 ep and the BCDMS μp and μd data without the constraint $\overline{d} = \overline{u}$ at low x. The dashed curves represent calculations using recent global fits by MRST and by CTEQ.

7 Impact of future HERA data on the determination of proton PDFs using the ZEUS NLO QCD fit [10]

7.1 PDF fits to HERA data

Recently, the ZEUS Collaboration have performed a combined NLO QCD fit to inclusive neutral and charged current DIS data [23–28] as well as high precision jet data in DIS [50] and γp scattering [51]. This is called the ZEUS-JETS PDF fit [10]. The use of only HERA data eliminates the uncertainties from heavy-target corrections and removes the need for isospin symmetry assumptions. It also avoids the difficulties that can sometimes arise from combining data-sets from several different experiments, thereby allowing a rigorous statistical treatment of the PDF uncertainties. Furthermore, PDF uncertainties from current global fits are, in general, limited by (irreducible) experimental systematics. In contrast, those from fits to HERA data alone, are largely limited by the statistical precision of existing measurements. Therefore, the impact of future data from HERA is likely to be most significant in fits to only HERA data.

7.2 The ZEUS NLO QCD fit

The ZEUS-JETS PDF fit has been used as the basis for all results shown in this contribution. The most important details of the fit are summarised here. A full description may be found elsewhere [10]. The fit includes the full set of ZEUS inclusive neutral and charged current $e^{\pm}p$ data from HERA-I (1994-2000), as well as two sets of high precision jet data in e^+p DIS ($Q^2 >> 1$ GeV^2) and γp ($Q^2 \sim 0$) scattering. The inclusive data used in the fit, span the kinematic range $6.3 \times 10^{-5} < x < 0.65$ and $2.7 < Q^2 < 30000$ GeV^2.

The PDFs are obtained by solving the NLO DGLAP equations within the $\overline{\text{MS}}$ scheme. These equations yield the PDFs at all values of Q^2 provided they are input as functions of x at some starting scale Q_0^2. The resulting PDFs are convoluted with coefficient functions to give predictions for structure functions and, hence, cross sections. In the ZEUS fit, the $xu_v(x)$ (u-valence), $xd_v(x)$ (d-valence), $xS(x)$ (total sea-quark), $xg(x)$ (gluon) and $x(\bar{d}(x) - \bar{u}(x)$ PDFs are parameterised at a starting scale of $Q_0^2 = 7$ GeV^2 by the form,

$$xf(x) = p_1 x^{p_2} (1 + x)^{p_3} P(x), \tag{15}$$

[10] Contributing authors: C. Gwenlan, A. Cooper-Sarkar, C. Targett-Adams.

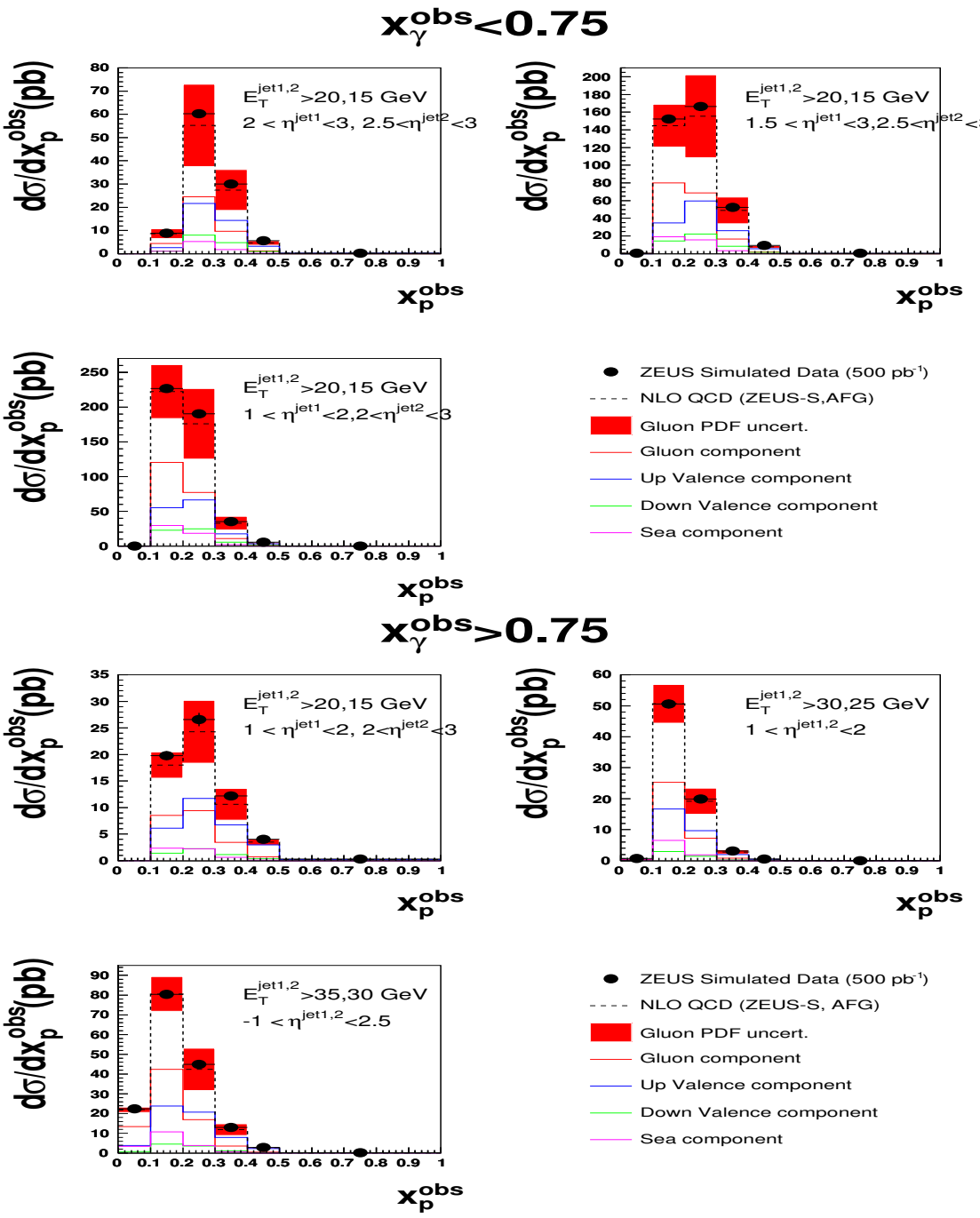

Fig. 22: The optimised jet cross sections included in the HERA-II projected fit. The solid points show the simulated data generated using the NLO QCD programme of Frixione-Ridolfi, using the CTEQ5M1 proton and the AFG photon PDFs. The error bars show the statistical uncertainties, which correspond to $500\ \mathrm{pb}^{-1}$ of HERA data. Systematic uncertainties have been neglected. The dashed line shows the NLO QCD prediction using the ZEUS-S proton and AFG photon PDFs. The shaded band shows the contribution to the cross section uncertainty arising from the uncertainty in the gluon distribution in the proton.

Table 4: The data-sets included in the ZEUS-JETS and HERA-II projected PDF fits. The first column lists the type of data and the second gives the kinematic coverage. The third column gives the integrated luminosities of the HERA-I measurements included in the ZEUS-JETS fit. The fourth column gives the luminosities assumed in the HERA-II projection. Note that the 96-97 NC and the 94-97 CC measurements have not had their luminosity scaled for the HERA-II projection.

data sample	**kinematic coverage**	**HERA-I** $\mathcal{L}$ (p b^{-1})	**HERA-II** $\mathcal{L}$ (p b^{-1}) **(assumed)**
96-97 NC e^+p [23]	$2.7 < Q^2 < 30000$ GeV^2; $6.3 \cdot 10^{-5} < x < 0.65$	30	30
94-97 CC e^+p [24]	$280 < Q^2 < 17000$ GeV^2; $6.3 \cdot 10^{-5} < x < 0.65$	48	48
98-99 NC e^-p [25]	$200 < Q^2 < 30000$ GeV^2; $0.005 < x < 0.65$	16	350
98-99 CC e^-p [26]	$280 < Q^2 < 17000$ GeV^2; $0.015 < x < 0.42$	16	350
99-00 NC e^+p [27]	$200 < Q^2 < 30000$ GeV^2; $0.005 < x < 0.65$	63	350
99-00 CC e^+p [28]	$280 < Q^2 < 17000$ GeV^2; $0.008 < x < 0.42$	61	350
96-97 inc. DIS jets [50]	$125 < Q^2 < 30000$ GeV^2; $E_{\rm T}^{Breit} > 8$ GeV	37	500
96-97 dijets in γp [51]	$Q^2 \lesssim 1$ GeV^2; $E_{\rm T}^{jet1,2} > 14, 11$ GeV	37	500
optimised jets [52]	$Q^2 \lesssim 1$ GeV^2; $E_{\rm T}^{jet1,2} > 20, 15$ GeV	-	500

where $P(x) = (1 + p_4 x)$. No advantage in the χ^2 results from using more complex polynomial forms. The normalisation parameters, $p_1(u_v)$ and $p_1(d_v)$, are constrained by quark number sum rules while $p_1(g)$ is constrained by the momentum sum rule. Since there is no information to constrain any difference in the low-x behaviour of the u- and d-valence quarks, $p_2(u_v)$ has been set equal to $p_2(d_v)$. The data from HERA are currently less precise than the fixed target data in the high-x regime. Therefore, the high-x sea and gluon distributions are not well constrained in current fits to HERA data alone. To account for this, the sea shape has been restricted by setting $p_4(S) = 0$. The high-x gluon shape is constrained by the inclusion of HERA jet data. In fits to only HERA data, there is no information on the shape of $\bar{d} - \bar{u}$. Therefore, this distribution has its shape fixed consistent with Drell-Yan data and its normalisation set consistent with the size of the Gottfried sum rule violation. A suppression of the strange sea with respect to the non-strange sea of a factor of 2 at Q_0^2 is also imposed, consistent with neutrino induced dimuon data from CCFR. The value of the strong coupling has been fixed to $\alpha_s(M_{\rm Z}) = 0.118$. After all constraints, the ZEUS-JETS fit has 11 free parameters. Heavy quarks were treated in the variable flavour number scheme of Thorne & Roberts [19]. Full account was taken of correlated experimental systematic uncertainties, using the Offset Method [9, 18].

The results of two separate studies are presented. The first study provides an estimate of how well the PDF uncertainties may be known by the end of HERA-II, within the currently planned running scenario, while the second study investigates the impact of a future HERA measurement of F_L on the gluon distribution. All results presented, are based on the recent ZEUS-JETS PDF analysis [10].

7.3 PDF uncertainty estimates for the end of HERA running

The data from HERA-I are already very precise and cover a wide kinematic region. However, HERA-II is now running efficiently and is expected to provide a substantial increase in luminosity. Current estimates suggest that, by the end of HERA running (in mid-2007), an integrated luminosity of 700 p b^{-1} should be achievable. This will allow more precise measurements of cross sections that are curently statistically limited: in particular, the high-Q^2 NC and CC data, as well as high-Q^2 and/or high-$E_{\rm T}$ jet data. In

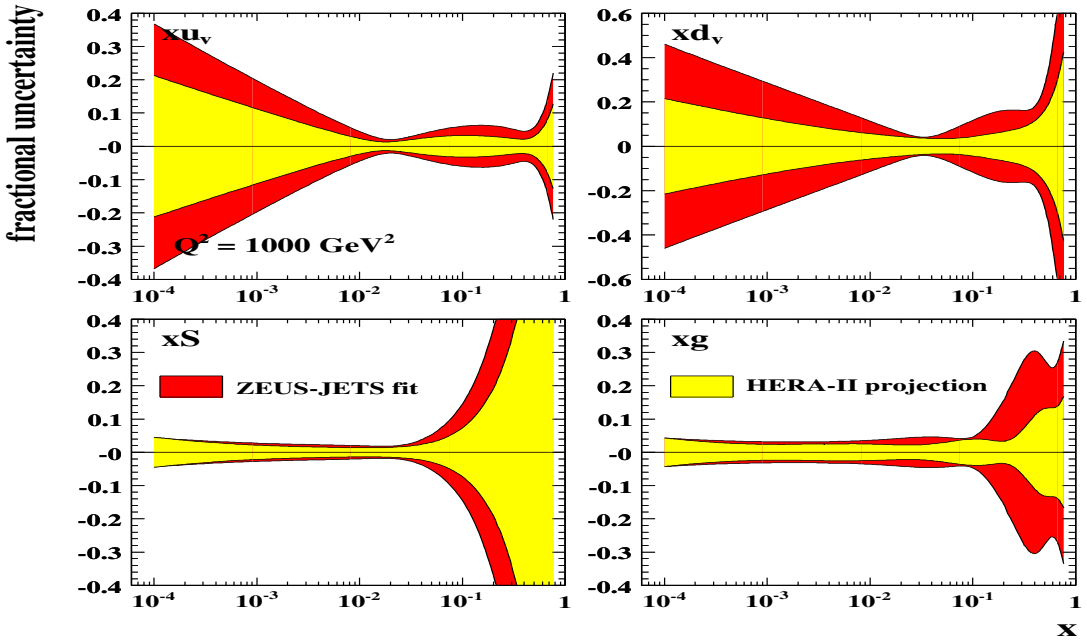

Fig. 23: The fractional PDF uncertainties, as a function of x, for the u-valence, d-valence, sea-quark and gluon distributions at $Q^2 = 1000\ \mathrm{GeV}^2$. The red shaded bands show the results of the ZEUS-JETS fit and the yellow shaded bands show the results of the HERA-II projected fit.

addition to the simple increase in luminosity, recent studies [52] have shown that future jet cross section measurements, in kinematic regions optimised for sensitivity to PDFs, should have a significant impact on the gluon uncertainties. In this contribution, the effect on the PDF uncertainties, of both the higher precision expected from HERA-II and the possibility of optimised jet cross section measurements, has been estimated in a new QCD fit. This fit will be referred to as the "HERA-II projection".

In the HERA-II projected fit, the statistical uncertainties on the currently available HERA-I data have been reduced. For the high-Q^2 inclusive data, a total integrated luminosity of 700 p $\bar{b}^{-1}$ was assumed, equally divided between e^+ and e^-. For the jet data, an integrated luminosity of 5 00p $\bar{b}^{-1}$ was assumed. The central values and systematic uncertainties were taken from the published data in each case. In addition to the assumed increase in precision of the measurements, a set of optimised jet cross sections were also included, for forward dijets in γp collisions, as defined in a recent study [52]. Since no real data are yet available, simulated points were generated using the NLO QCD program of Frixione-Ridolfi [53], using the CTEQ5M1 [4] proton and AFG [54] photon PDFs. The statistical uncertainties were taken to correspond to 5 00p $\bar{b}^{-1}$. For this study, systematic uncertainties on the optimised jet cross sections were ignored. The simulated optimised jet cross section points, compared to the predictions of NLO QCD using the ZEUS-S proton PDF [9], are shown in Fig. 22.

Table 4 lists the data-sets included in the ZEUS-JETS and HERA-II projected fits. The luminosities of the (real) HERA-I measurements and those assumed for the HERA-II projection are also given.

The results are summarised in Fig. 23, which shows the fractional PDF uncertainties, for the u- and d-valence, sea-quark and gluon distributions, at $Q^2 = 1\ 000\mathrm{GeV}^2$. The yellow bands show the results of the ZEUS-JETS fit while the red bands show those for the HERA-II projection. Note that the same general features are observed for all values of Q^2. In fits to only HERA data, the information on the valence quarks comes from the high-Q^2 NC and CC cross sections. The increased statistical precision of the high-Q^2 data, as assumed in the HERA-II projected fit, gives a significant improvement in the valence uncertainties over the whole range of x. For the sea quarks, a significant improvement in the

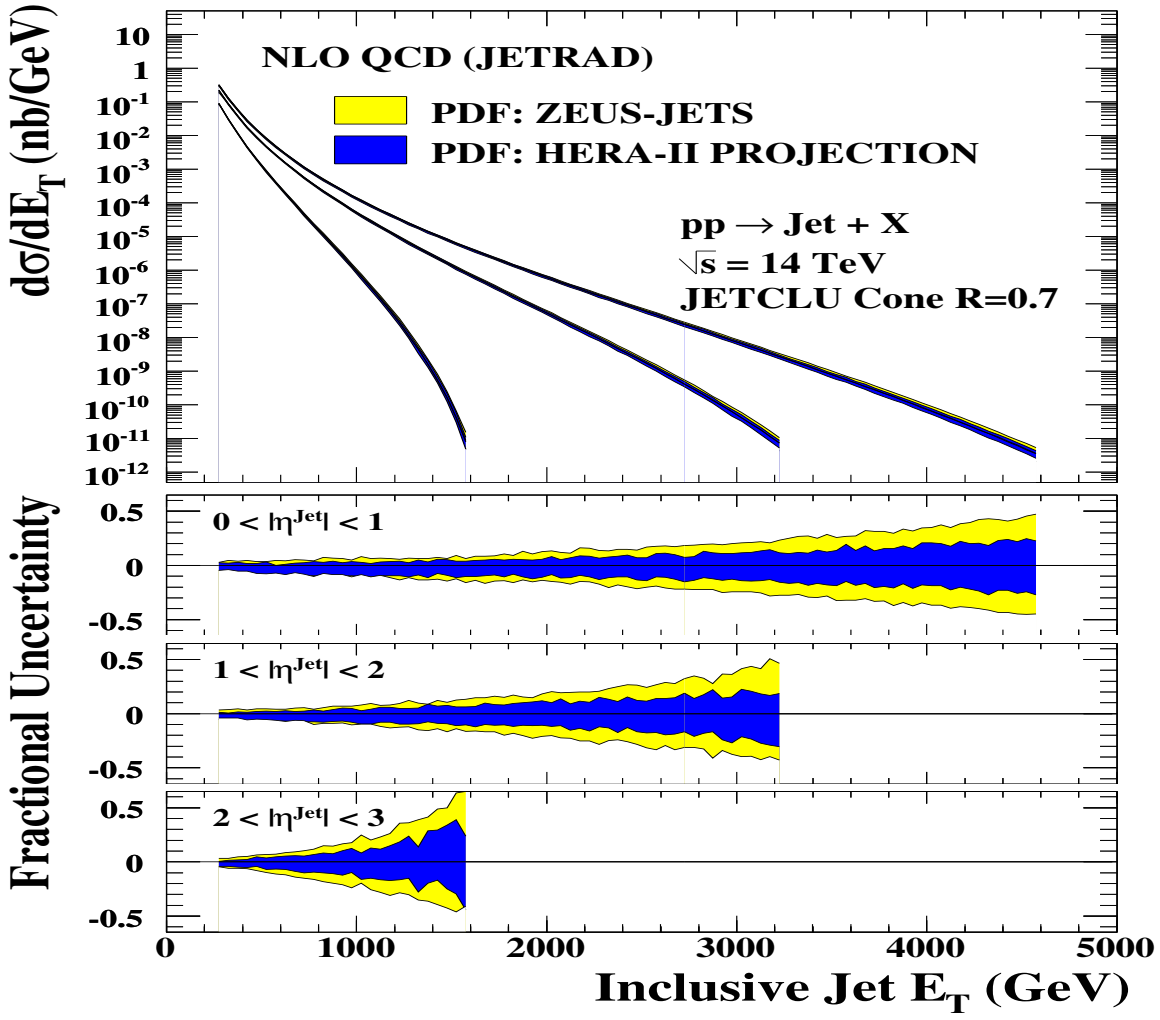

Fig. 24: NLO QCD inclusive jet cross section predictions at $\sqrt{s}$=14 TeV in three regions of pseudo-rapidity. The yellow and blue bands show the PDF uncertainties from the ZEUS-JETS and HERA-II projected fits, respectively.

uncertainties at high-x is also observed. In contrast, the low-x uncertainties are not visibly reduced. This is due to the fact that the data constraining the low-x region tends to be at lower-Q^2, which are already systematically limited. This is also the reason why the low-x gluon uncertainties are not significantly reduced. However, the mid-to-high-x gluon, which is constrained by the jet data, is much improved in the HERA-II projected fit. Note that about half of the observed reduction in the gluon uncertainties is due to the inclusion of the simulated optimised jet cross sections.

Inclusive jet cross sections at the LHC

The improvement to the high-x partons, observed in the HERA-II projection compared to the ZEUS-JETS fit, will be particularly relevant for high-scale physics at the LHC. This is illustrated in Fig. 24, which shows NLO QCD predictions from the JETRAD [55] programme for inclusive jet production at $\sqrt{s} = 14$ TeV. The results are shown for both the ZEUS-JETS and the HERA-II projected PDFs. The uncertainties on the cross sections, resulting from the PDFs, have been calculated using the LHAPDF interface [56]. For the ZEUS-JETS PDF, the uncertainty reaches ~ 5 % at central pseudo-rapidities, for the highest jet transverse energies shown. The prediction using the HERA-II projected PDF shows a marked improvement at high jet tranverse energy.

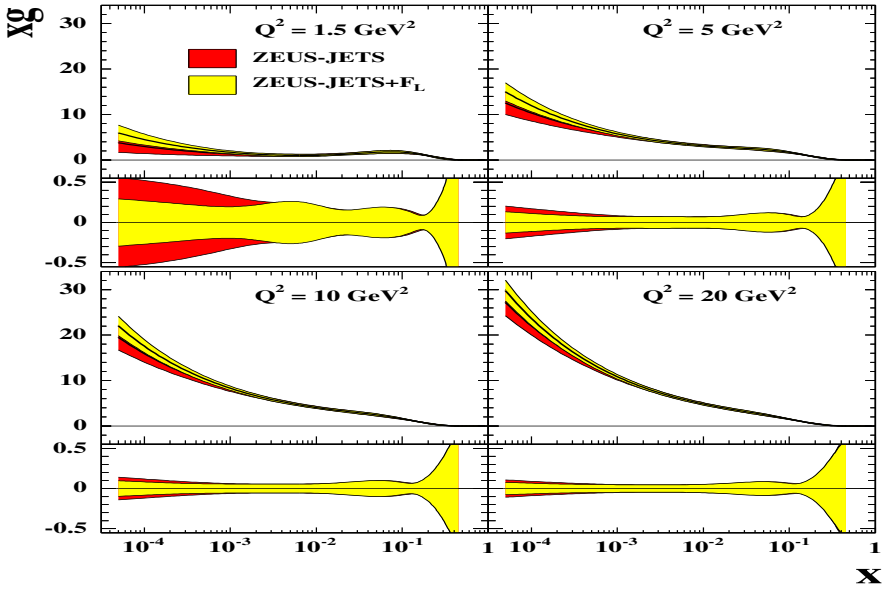

Fig. 25: The gluon PDFs, showing also the fractional uncertainty, for fits with and without inclusion of the simulated F_L data, for $Q^2 = 1.5$, 5, 10 and 20 GeV^2. The red shaded bands show the results of the ZEUS-JETS fit and the yellow shaded band show the results of the ZEUS-JETS+F_L fit.

7.4 Impact of a future HERA measurement of F_L on the gluon PDF

The longitudinal structure function, F_L, is directly related to the gluon density in the proton. In principle, F_L can be extracted by measuring the NC DIS cross section at fixed x and Q^2, for different values of y (see Eqn. 3). A precision measurement could be achieved by varying the centre-of-mass energy, since $s = Q^2/\ xy \approx\ 4E_eE_p$, where E_e and E_p are the electron and proton beam energies, respectively. Studies [38] (Sec. 5) have shown that this would be most efficiently achieved by changing the proton beam energy. However, such a measurement has not yet been performed at HERA.

There are several reasons why a measurement of F_L at low-x could be important. The gluon density is not well known at low-x and so different PDF parameterisations can give quite different predictions for F_L at low-x. Therefore, a precise measurement of the longitudinal sturcture function could both pin down the gluon PDF and reduce its uncertainties. Furthermore, predictions of F_L also depend upon the nature of the underlying theory (e.g. order in QCD, resummed calculation etc). Therefore, a measurement of F_L could also help to discriminate between different theoretical models.

Impact on the gluon PDF uncertainties

The impact of a possible future HERA measurement of F_L on the gluon PDF uncertainties has been investigated, using a set of simulated F_L data-points [38]. (see Sec. 5). The simulation was performed using the GRV94 [57] proton PDF for the central values, and assuming $E_e = 2\ 76$ GeV and $E_p = 9\ 2\ 0\ 5\ \ 7\ 5\ 4\ \ 6\ 5$and $4\ \ 0\ 0$GeV, with luminosities of 10, 5, 3 and 2 $\mathrm{p\ b}^{-1}$, respectively. Assuming that the luminosity scales simply as E_p^2, this scenario would nominally cost $3\ 5\mathrm{p\ b}^{-1}$ of luminosity under standard HERA conditions. However, this estimate takes no account of time taken for optimisation of the machine with each change in E_p, which could be considerable. The systematic uncertainties on the simulated data-points were calculated assuming a $\sim\ 2\ \%$precision on the inclusive NC cross section measurement. A more comprehensive description of the simulated data is given in contribution for this proceedings, see Sec. 5.

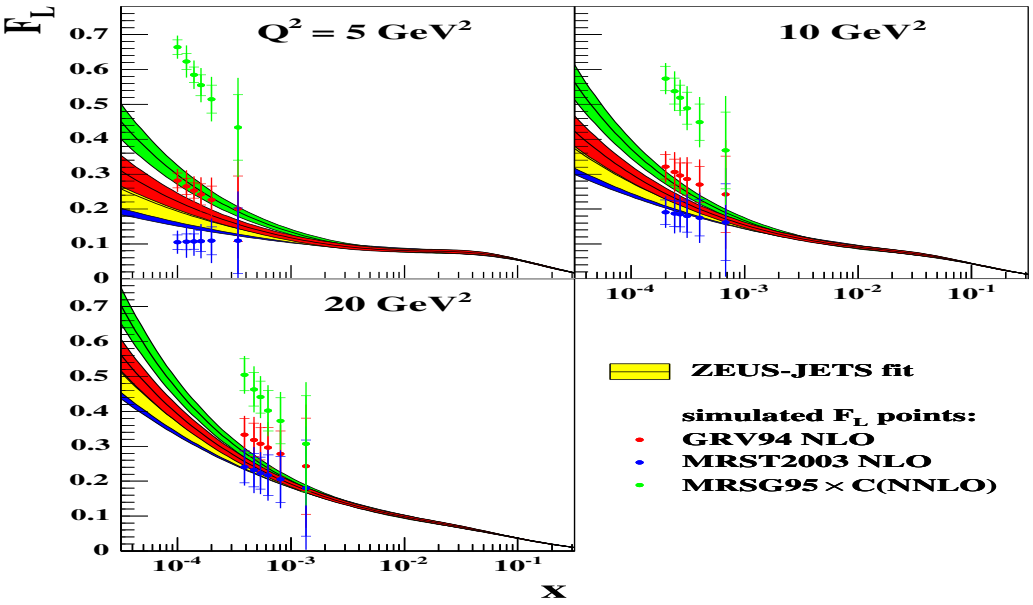

Fig. 26: The distribution of the longitudinal structure function F_L at Q^2 =5, 10 and 20 GeV^2. The blue, red and green points show the simulated F_L data-points, respectively labelled maximum, middle and minimum in Table 5. The blue, red and green shaded bands show the NLO QCD predictions, in the case where the data-points of the corresponding colour have been included in the fit. For comparison, the yellow shaded band shows the prediction of the ZEUS-JETS fit.

The simulated data were included in the ZEUS-JETS fit. Figure 25 shows the gluon distribution and fractional uncertainties for fits with and without inclusion of the simulated F_L data. The results indicate that the gluon uncertainties are reduced at low-x, but the improvement is only significant at relatively low $Q^2 \lesssim 2$ 0GeV^2.

Discrimination between theoretical models

In order to assess whether a HERA measurement of F_L could discriminate between theoretical models, two more sets of F_L data-points have been simulated [58], using different theoretical assumptions. The first of the two sets was generated using the MRSG95 [59] proton PDF, which has a large gluon density. The PDFs were then convoluted with the NNLO order coefficient functions, which are large and positive. This gives the "maximum" set of F_L data-points. In contrast, the second set has been generated using the MRST2003 [60] proton PDF, which has a negative gluon at low-x and low-Q^2, thus providing a "minimum" set of F_L data. The original set of F_L points described in the previous subsection lies between these two extremes. The details of all three sets are summarised in Table 5.

Figure 26 shows the results of including, individually, each set of simulated F_L data into the ZEUS NLO QCD fit. The results show that the NLO fit is relatively stable to the inclusion of the extreme sets of data. This indicates that a measurement of F_L could discriminate between certain theoretical models. However, it should be noted that the maximum and minimum models studied here were chosen specifically to give the widest possible variation in F_L. There are many other alternatives that would lie between these extremes and the ability of an F_L measurement to discriminate between them would depend both on the experimental precision of the measurement itself, as well as the theoretical uncertainties on the models being tested.

Table 5: Summary of the PDFs used to generate the simulated F_L data-points. The extreme maximum F_L points were generated using the MRSG95 PDF, and convoluted with NNLO coefficient functions. The middle points were generated using the GRV94 PDF, and the extreme minimum points were generated using the MRST2003 PDF, which has a negative gluon at low-x.

	PDF	QCD order of coefficient functions
Maximum F_L	MRSG95	NNLO
Middle F_L	GRV94	NLO
Minimum F_L	MRST2003	NLO

8 A Method to Include Final State Cross-sections Measured in Proton-Proton Collisions to Global NLO QCD Analysis [11]

The Large Hadron Collider (LHC), currently under construction at CERN, will collide protons on protons with an energy of 7 TeV. Together with its high collision rate the high available centre-of-mass energy will make it possible to test new interactions at very short distances that might be revealed in the production cross-sections of Standard Model (SM) particles at very high transverse momentum (P_T) as deviation from the SM theory.

The sensitivity to new physics crucially depends on experimental uncertainties in the measurements and on theoretical uncertainties in the SM predictions. It is therefore important to work out a strategy to minimize both the experimental and theoretical uncertainties from LHC data. For instance, one could use single inclusive jet or Drell-Yan cross-sections at low P_T to constrain the PDF uncertainties at high P_T. Typical residual renormalisation and factorisation scale uncertainties in next-to-leading order (NLO) calculations for single inclusive jet-cross-section are about $5-10\%$ and should hopefully be reduced as NNLO calculations become available. The impact of PDF uncertainties on the other hand can be substantially larger in some regions, especially at large P_T, and for example at $P_T = 2\,000$ GeV dominate the overall uncertainty of 20%. If a suitable combination of data measured at the Tevatron and LHC can be included in global NLO QCD analyses, the PDF uncertainties can be constrained.

The aim of this contribution is to propose a method for consistently including final-state observables in global QCD analyses.

For inclusive data like the proton structure function F_2 in deep-inelastic scattering (DIS) the perturbative coefficients are known analytically. During the fit the cross-section can therefore be quickly calculated from the strong coupling (α_s) and the PDFs and can be compared to the measurements. However, final state observables, where detector acceptances or jet algorithms are involved in the definition of the perturbative coefficients (called "weights" in the following), have to be calculated using NLO Monte Carlo programs. Typically such programs need about one day of CPU time to calculate accurately the cross-section. It is therefore necessary to find a way to calculate the perturbative coefficients with high precision in a long run and to include α_s and the PDFs "a posteriori".

To solve this problem many methods have been proposed in the past [3,10,61–64]. In principle the highest efficiencies can be obtained by taking moments with respect to Bjorken-x [61,62], because this converts convolutions into multiplications. This can have notable advantages with respect to memory consumption, especially in cases with two incoming hadrons. On the other hand, there are complications such as the need for PDFs in moment space and the associated inverse Mellin transforms.

Methods in x-space have traditionally been somewhat less efficient, both in terms of speed (in the 'a posteriori' steps — not a major issue here) and in terms of memory consumption. They are, however, somewhat more transparent since they provide direct information on the x values of relevance. Furthermore they can be used with any PDF. The use of x-space methods can be further improved by using methods developed originally for PDF evolution [65,66].

[11] Contributing authors: T. Carli, G. Salam, F. Siegert.

8.1 PDF-independent representation of cross-sections

Representing the PDF on a grid

We make the assumption that PDFs can be accurately represented by storing their values on a two-dimensional grid of points and using n^{th}-order interpolations between those points. Instead of using the parton momentum fraction x and the factorisation scale Q^2, we use a variable transformation that provides good coverage of the full x and Q^2 range with uniformly spaced grid points:[12]

$$y(x) = \ln\frac{1}{x} \quad \text{and} \quad \tau(Q^2) = \ln\ln\frac{Q^2}{\Lambda^2}. \tag{16}$$

The parameter Λ is to be chosen of the order of Λ_{QCD}, but not necessarily identical. The PDF $q(x, Q^2)$ is then represented by its values $q_{i_y,\, i_\tau}$ at the 2-dimensional grid point $(i_y\,\delta y, i_\tau\,\delta\tau)$, where δy and $\delta\tau$ denote the grid spacings, and obtained elsewhere by interpolation:

$$q(x,Q^2) = \sum_{i=0}^{n}\sum_{\iota=0}^{n'} q_{k+i,\,\kappa+\iota}\; I_i^{(n)}\left(\frac{y(x)}{\delta y} - k\right)\; I_\iota^{(n')}\left(\frac{\tau(Q^2)}{\delta\tau} - \kappa\right), \tag{17}$$

where n, n' are the interpolation orders. The interpolation function $I_i^{(n)}(u)$ is 1 for $u = i$ and otherwise is given by:

$$I_i^{(n)}(u) = \frac{(-1)^{n-i}}{i!\,(n-i)!}\,\frac{u(u-1)\ldots(u-n)}{u-i}\,. \tag{18}$$

Defining $\text{int}(u)$ to be the largest integer such that $\text{int}(u) \le u$, k and κ are defined as:

$$k(x) = \text{int}\left(\tfrac{y(x)}{\delta y} - \tfrac{n-1}{2}\right), \quad \kappa(x) = \text{int}\left(\frac{\tau(Q^2)}{\delta\tau} - \frac{n'-1}{2}\right). \tag{19}$$

Given finite grids whose vertex indices range from $0 \ldots N_y - 1$ for the y grid and $0 \ldots N_\tau - 1$ for the τ grid, one should additionally require that eq. (17) only uses available grid points. This can be achieved by remapping $k \to \max(0, \min(N_y - 1 - n, k))$ and $\kappa \to \max(0, \min(N_\tau - 1 - n', \kappa))$.

Representing the final state cross-section weights on a grid (DIS case)

Suppose that we have an NLO Monte Carlo program that produces events $m = 1 \ldots N$. Each event m has an x value, x_m, a Q^2 value, Q_m^2, as well as a weight, w_m, and a corresponding order in α_s, p_m. Normally one would obtain the final result W of the Monte Carlo integration from:[13]

$$W = \sum_{m=1}^{N} w_m \left(\frac{\alpha_s(Q_m^2)}{2\pi}\right)^{p_m} q(x_m, Q_m^2). \tag{20}$$

Instead one introduces a weight grid $W_{i_y,\, i_\tau}^{(p)}$ and then for each event updates a portion of the grid with:
$i = 0 \ldots n,\ \iota = 0 \ldots n'$:

$$W_{k+i,\,\kappa+\iota}^{(p_m)} \to W_{k+i,\,\kappa+\iota}^{(p_m)} + w_m\, I_i^{(n)}\left(\frac{y(x_m)}{\delta y} - k\right) I_\iota^{(n')}\left(\frac{\tau(Q_m^2)}{\delta\tau} - \kappa\right), \tag{21}$$
$$\text{where } k \equiv k(x_m),\ \kappa \equiv \kappa(Q_m^2).$$

[12] An alternative for the x grid is to use $y = \ln(x/(1-x))$ with a a parameter that serves to increase the density of points in the large x region.

[13] Here, and in the following, renormalisation and factorisation scales have been set equal for simplicity.

The final result for W, for an arbitrary PDF, can then be obtained *subsequent* to the Monte Carlo run:

$$W = \sum_p \sum_{i_y} \sum_{i_\tau} W^{(p)}_{i_y,\, i_\tau} \left(\frac{\alpha_s\left(Q^{2(i_\tau)}\right)}{2\pi} \right)^p q\left(x^{(i_y)}, Q^{2(i_\tau)}\right) , \tag{22}$$

where the sums index with i_y and i_τ run over the number of grid points and we have have explicitly introduced $x^{(i_y)}$ and $Q^{2(i_\tau)}$ such that:

$$y(x^{(i_y)}) = i_y\, \delta y \quad \text{and} \quad \tau\left(Q^{2(i_\tau)}\right) = i_\tau\, \delta\tau. \tag{23}$$

Including renormalisation and factorisation scale dependence

If one has the weight matrix $W^{(p)}_{i_y,\, i_\tau}$ determined separately order by order in $\alpha_{\rm s}$, it is straightforward to vary the renormalisation μ_R and factorisation μ_F scales a posteriori (we assume that they were kept equal in the original calculation).

It is helpful to introduce some notation relating to the DGLAP evolution equation:

$$\frac{dq(x,Q^2)}{d\ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi}(P_0 \otimes q)\ x(Q^2) + \left(\frac{\alpha_s(Q^2)}{2\pi}\right)^2 (P_1 \otimes q)\ x(Q^2) + \ldots, \tag{24}$$

where the P_0 and P_1 are the LO and NLO matrices of DGLAP splitting functions that operate on vectors (in flavour space) q of PDFs. Let us now restrict our attention to the NLO case where we have just two values of p, $p_{\rm LO}$ and $p_{\rm NLO}$. Introducing ξ_R and ξ_F corresponding to the factors by which one varies μ_R and μ_F respectively, for arbitrary ξ_R and ξ_F we may then write:

$$\begin{aligned} W(\xi_R,\xi_F) = \sum_{i_y}\sum_{i_\tau} & \left(\frac{\alpha_s\left(\xi_R^2 Q^{2(i_\tau)}\right)}{2\pi} \right)^{p_{\rm LO}} W^{(p_{\rm LO})}_{i_y,\, i_\tau} q\left(x^{(i_y)}, \xi_F^2 Q^{2(i_\tau)}\right) + \\ & \left(\frac{\alpha_s\left(\xi_R^2 Q^{2(i_\tau)}\right)}{2\pi} \right)^{p_{\rm NLO}} \Big[\left(W^{(p_{\rm NLO})}_{i_y,\, i_\tau} + 2\pi\beta_0 p_{\rm LO} \ln\xi_R^2\, W^{(p_{\rm LO})}_{i_y,\, i_\tau} \right) q\left(x^{(i_y)}, \xi_F^2 Q^{2(i_\tau)}\right) \\ & - \ln\xi_F^2\, W^{(p_{\rm LO})}_{i_y,\, i_\tau} (P_0 \otimes q)\left(x^{(i_y)}, \xi_F^2 Q^{2(i_\tau)}\right) \Big], \end{aligned} \tag{25}$$

where $\beta_0 = (11 - 2n_f)/(12\pi)$ and N_c (n_f) is the number of colours (flavours). Though this formula is given for x-space based approach, a similar formula applies for moment-space approaches. Furthermore it is straightforward to extend it to higher perturbative orders.

Representing the weights in the case of two incoming hadrons

In hadron-hadron scattering one can use analogous procedures with one more dimension. Besides Q^2, the weight grid depends on the momentum fraction of the first (x_1) and second (x_2) hadron.

In the case of jet production in proton-proton collisions the weights generated by the Monte Carlo program as well as the PDFs can be organised in seven possible initial state combinations of partons:

$$\text{g g:} \quad F^{(0)}(x_1,x_2;Q^2) = G_1(x_1)G_2(x_2) \tag{26}$$

$$\text{q g:} \quad F^{(1)}(x_1,x_2;Q^2) = \left(Q_1(x_1) + \overline{Q}_1(x_1)\right) G_2(x_2) \tag{27}$$

$$\text{g q:} \quad F^{(2)}(x_1,x_2;Q^2) = G_1(x_1)\left(Q_2(x_2) + \overline{Q}_2(x_2)\right) \tag{28}$$

$$\text{q r:} \quad F^{(3)}(x_1,x_2;Q^2) = Q_1(x_1)Q_2(x_2) + \overline{Q}_1(x_1)\overline{Q}_2(x_2) - D(x_1,x_2) \tag{29}$$

$$\text{q q:} \quad F^{(4)}(x_1,x_2;Q^2) = D(x_1,x_2) \tag{30}$$

$$\mathrm{q\bar{q}}: \quad F^{(5)}(x_1,x_2;Q^2) = \overline{D}(x_1,x_2) \tag{31}$$
$$\mathrm{q\bar{r}}: \quad F^{(6)}(x_1,x_2;Q^2) = Q_1(x_1)\overline{Q}_2(x_2) + \overline{Q}_1(x_1)Q_2(x_2) - \overline{D}(x_1,x_2), \tag{32}$$

where g denotes gluons, q quarks and r quarks of different flavour $q \neq r$ and we have used the generalized PDFs defined as:

$$G_H(x) = f_{0/H}(x,Q^2), \qquad Q_H(x) = \sum_{i=1}^{6} f_{i/H}(x,Q^2), \quad \overline{Q}_H(x) = \sum_{i=-6}^{-1} f_{i/H}(x,Q^2),$$

$$D(x_1,x_2) = \sum_{\substack{i=-6 \\ i\neq 0}}^{6} f_{i/H_1}(x_1,Q^2) f_{i/H_2}(x_2,Q^2), \tag{33}$$

$$\overline{D}(x_1,x_2,\mu_F^2) = \sum_{\substack{i=-6 \\ i\neq 0}}^{6} f_{i/H_1}(x_1,Q^2) f_{-i/H_2}(x_2,Q^2),$$

where $f_{i/H}$is the PDF of flavour $i = -6 \ldots 6$ for hadron H and H_1 (H_2) denotes the first or second hadron[14].

The analogue of eq. 22 is then given by:

$$W = \sum_p \sum_{l=0}^{6} \sum_{i_{y_1}} \sum_{i_{y_2}} \sum_{i_\tau} W^{(p)\ l}_{i_{y_1},\, i_{y_2},\, i_\tau} \left(\frac{\alpha_s\left(Q^{2(i_\tau)}\right)}{2\pi} \right)^p F^{(l)}\left(x_1^{(i_{y_1})}, x_2^{(i_{y_1})}, Q^{2(i_\tau)}\right). \tag{34}$$

Including scale depedence in the case of two incoming hadrons

It is again possible to choose arbitrary renormalisation and factorisation scales, specifically for NLO accuracy:

$$W(\xi_R,\xi_F) = \sum_{l=0}^{6} \sum_{i_{y_1}} \sum_{i_{y_2}} \sum_{i_\tau} \left(\frac{\alpha_s\left(\xi_R^2 Q^{2(i_\tau)}\right)}{2\pi} \right)^{p_{\mathrm{LO}}} W^{(p_{\mathrm{LO}})\ l}_{i_{y_1},\, i_{y_2},\, i_\tau} F^{(l)}\left(x_1^{(i_{y_1})}, x_2^{(i_{y_1})}, \xi_F^2 Q^{2(i_\tau)}\right) +$$
$$\left(\frac{\alpha_s\left(\xi_R^2 Q^{2(i_\tau)}\right)}{2\pi} \right)^{p_{\mathrm{NLO}}} \Big[\left(W^{(p_{\mathrm{NLO}})\ l}_{i_{y_1},\, i_{y_2},\, i_\tau} + 2\pi\beta_0 p_{\mathrm{LO}} \ln\xi_R^2\, W^{(p_{\mathrm{LO}})\ l}_{i_{y_1},\, i_{y_2},\, i_\tau} \right) F^{(l)}\left(x_1^{(i_{y_1})}, x_2^{(i_{y_1})}, \xi_F^2 Q^{2(i_\tau)}\right) \tag{35}$$
$$- \ln\xi_F^2\, W^{(p_{\mathrm{LO}})\ l}_{i_{y_1},\, i_{y_2},\, i_\tau} \left(F^{(l)}_{q_1\to P_0\otimes q_1}\left(x_1^{(i_{y_1})}, x_2^{(i_{y_1})}, \xi_F^2 Q^{2(i_\tau)}\right) + F^{(l)}_{q_2\to P_0\otimes q_2}\left(x_1^{(i_{y_1})}, x_2^{(i_{y_1})}, \xi_F^2 Q^{2(i_\tau)}\right) \right) \Big],$$

where $F^{(l)}_{q_1\to P_0\otimes q_1}$ is calculated as $F^{(l)}$, but with q_1 replaced wtih $P_0\otimes q_1$, and analogously for $F^{(l)}_{q_2\to P_0\otimes q_2}$.

8.2 Technical implementation

To test the scheme discussed above we use the NLO Monte Carlo program NLOJET++ [67] and the CTEQ6 PDFs [4]. The grid $W^{(p)\ l}_{i_{y_1},\, i_{y_2},\, i_\tau}$ of eq. 34 is filled in a NLOJET++ user module. This module has access to the event weight and parton momenta and it is here that one specifies and calculates the physical observables that are being studied (e.g. jet algorithm).

Having filled the grid we construct the cross-section in a small standalone program which reads the weights from the grid and multiplies them with an arbitrary α_{s} and PDF according to eq. 34. This program runs very fast (in the order of seconds) and can be called in a PDF fit.

[14] In the above equation we follow the standard PDG Monte Carlo numbering scheme [17] where gluons are denoted as 0, quarks have values from 1-6 and anti-quarks have the corresponding negative values.

The connection between these two programs is accomplished via a C++ class, which provides methods e.g. for creating and optimising the grid, filling weight events and saving it to disk. The classes are general enough to be extendable for the use with other NLO calculations.

The complete code for the NLOJET++ module, the C++ class and the standalone job is available from the authors. It is still in a development, testing and tuning stage, but help and more ideas are welcome.

The C++ class

The main data members of this class are the grids implemented as arrays of three-dimensional ROOT histograms, with each grid point at the bin centers[15]:

$$\mathrm{TH3D}[\mathrm{p}][\mathrm{l}]_{[x_1,x_2,Q^2]}(\mathrm{x}) \tag{36}$$

where the l and p are explained in eq. 34 and i_{obs} denotes the observable bin, e.g. a given P_T range[16].

The C++ class initialises, stores and fills the grid using the following main methods:

- *Default constructor:* Given the pre-defined kinematic regions of interest, it initializes the grid.
- *Optimizing method:* Since in some bins the weights will be zero over a large kinematic region in x_1, x_2, Q^2, the optimising method implements an automated procedure to adapt the grid boundaries for each observable bin. These boundaries are calculated in a first (short) run. In the present implementation, the optimised grid has a fixed number of grid points. Other choices, like a fixed grid spacing, might be implemented in the future.
- *Loading method:* Reads the saved weight grid from a ROOT file
- *Saving method:* Saves the complete grid to a ROOT file, which will be automatically compressed.

The user module for NLOJET++

The user module has to be adapted specifically to the exact definition of the cross-section calculation. If a grid file already exists in the directory where NLOJET++ is started, the grid is not started with the default constructor, but with the optimizing method (see 8.2). In this way the grid boundaries are optimised for each observable bin. This is necessary to get very fine grid spacings without exceeding the computer memory. The grid is filled at the same place where the standard NLOJET++ histograms are filled. After a certain number of events, the grid is saved in a root-file and the calculation is continued.

The standalone program for constructing the cross-section

The standalone program calculates the cross-section in the following way:

1. Load the weight grid from the ROOT file
2. Initialize the PDF interface[17], load $q(x, Q^2)$ on a helper PDF-grid (to increase the performance)
3. For each observable bin, loop over $i_{y_1}, i_{y_2}, i_\tau, l, p$ and calculate $F^l(x_1, x_2, Q^2)$ from the appropriate PDFs $q(x, Q^2)$, multiply α_s and the weights from the grid and sum over the initial state parton configuration l, according to eq. 34.

[15] ROOT histograms are easy to implement, to represent and to manipulate. They are therefore ideal in an early development phase. An additional advantage is the automatic file compression to save space. The overhead of storing some empty bins is largely reduced by optimizing the x_1, x_2 and Q^2 grid boundaries using the NLOJET++ program before final filling. To avoid this residual overhead and to exploit certain symmetries in the grid, a special data class (e.g. a sparse matrix) might be constructed in the future.

[16] For the moment we construct a grid for each initial state parton configuration. It will be easy to merge the qg and the gq initial state parton configurations in one grid. In addition, the weights for some of the initial state parton configurations are symmetric in x_1 and x_2. This could be exploited in future applications to further reduce the grid size.

[17] We use the C++ wrapper of the LHAPDF interface [56].

8.3 Results

We calculate the single inclusive jet cross-section as a function of the jet transverse momentum (P_T) for jets within a rapidity of $|y| < 0.5$. To define the jets we use the seedless cone jet algorithm as implemented in NLOJET++ using the four-vector recombination scheme and the midpoint algorithm. The cone radius has been put to $R = 0.7$, the overlap fraction was set to $f = 0.5$. We set the renormalisation and factorization scale to $Q^2 = P_{T,\,max}^2$ where $P_{T,\,max}$ is the P_T of the highest P_T jet in the required rapidity region[18].

In our test runs, to be independent from statistical fluctuations (which can be large in particular in the NLO case), we fill in addition to the grid a reference histogram in the standard way according to eq. 20.

The choice of the grid architecture depends on the required accuracy, on the exact cross-section definition and on the available computer resources. Here, we will just sketch the influence of the grid architecture and the interpolation method on the final result. We will investigate an example where we calculate the inclusive jet cross-section in $N_{\rm obs} = 1\ 00$ bins in the kinematic range $1\ 00 \le P_T \le 5\ 000$ GeV. In future applications this can serve as guideline for a user to adapt the grid method to his/her specific problem. We believe that the code is transparent and flexible enough to adapt to many applications.

As reference for comparisons of different grid architectures and interpolation methods we use the following:

- *Grid spacing in $y(x)$:* $1\ 0^{-5} \le x_1, x_2 \le 1.0$ with $N_y = 3\ 0$
- *Grid spacing in $\tau(Q^2)$:* $1\ 00$ GeV $\le Q \le 5\ 000$ GeV with $N_\tau = 3\ 0$
- *Order of interpolation:* $n_y = 3$, $n_\tau = 3$

The grid boundaries correspond to the user setting for the first run which determines the grid boundaries for each observable bin. In the following we call this grid architecture $3\ 0^3$x$3\ 0$x$1\ 00(,33)$ Such a grid takes about $3\ 00$ Mbyte of computer memory. The root-file where the grid is stored has about $5\ 0$ Mbyte.

The result is shown in Fig. 27a). The reference cross-section is reproduced everywhere to within 0.05 % The typical precision is about 0.01 % At low and high P_T there is a positive bias of about 0.04 % Also shown in Fig. 27a) are the results obtained with different grid architectures. For a finer x grid ($5\ 0^2$x$3\ 0$x$1\ 00(,33)$ the accuracy is further improved (within 0.005 %) and there is no bias. A finer ($3\ 0^3$x60x$1\ 00(,33)$ as well as a coarser ($3\ 0^3$x$1\ 0$x$1\ 00(,33)$ binning in Q^2 does not improve the precision.

Fig. 27b) and Fig. 27c) show for the grid ($3\ 0^3$x$3\ 0$x$1\ 00$) different interpolation methods. With an interpolation of order $n = 5$ the precision is 0.01 % and the bias at low and high P_T observed for the $n = 3$ interpolation disappears. The result is similar to the one obtained with finer x-points. Thus by increasing the interpolation order the grid can be kept smaller. An order $n = 1$ interpolation gives a systematic negative bias of about 1 % becoming even larger towards high P_T.

Depending on the available computer resources and the specific problem, the user will have to choose a proper grid architecture. In this context, it is interesting that a very small grid $1\ 0^2$x$1\ 0$x$1\ 00(,55)$ that takes only about 1 Mbyte computer memory reaches still a precision of 0.5 % if an interpolation of order $n = 5$ is used (see Fig. 27d)).

We have developed a technique to store the perturbative coefficients calculated by an NLO Monte Carlo program on a grid allowing for a-posteriori inclusion of an arbitrary parton density function (PDF)

[18] Note that beyond LO the $P_{T,\,max}$ will in general differ from the P_T of the other jets, so when binning an inclusive jet cross section, the P_T of a given jet may not correspond to the renormalisation scale chosen for the event as a whole. For this reason we shall need separate grid dimensions for the jet P_T and for the renormalisation scale. Only in certain moment-space approaches [62] has this requirement so far been efficiently circumvented.

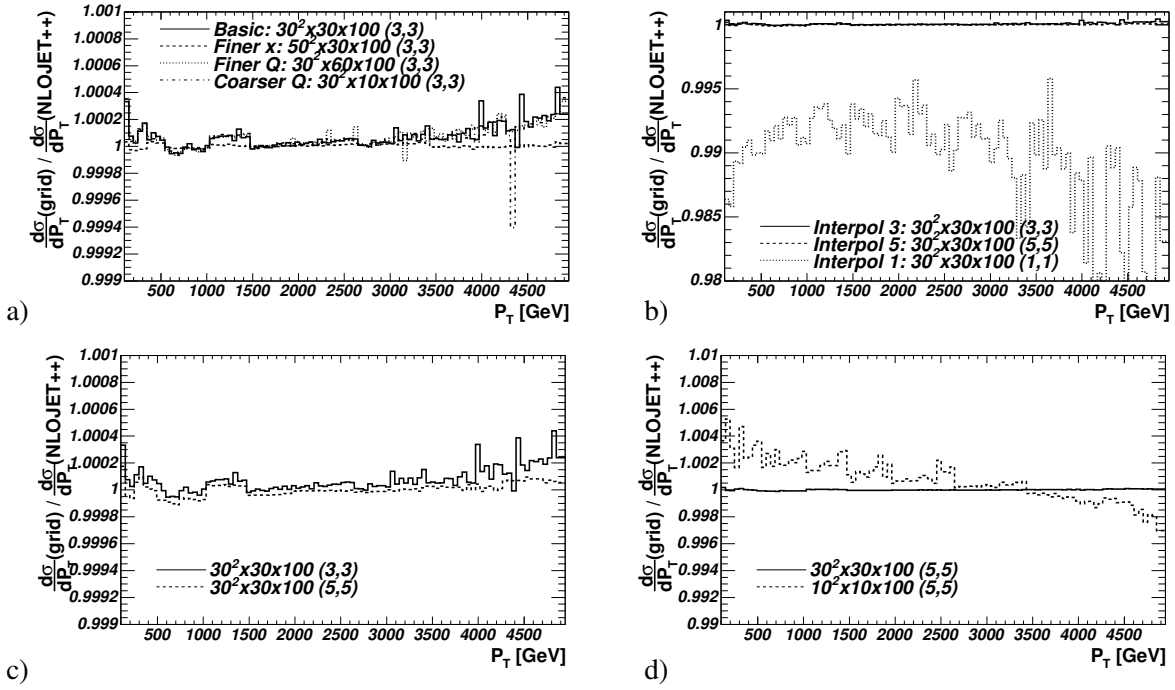

Fig. 27: Ratio between the single inclusive jet cross-section with 100 P_T bins calculated with the grid technique and the reference cross-section calculated in the standard way. Shown are the standard grid, grids with finer x and Q^2 sampling (a) with interpolation of order 1, 3 and 5 (b) (and on a finer scale in c)) and a small grid (d).

set. We extended a technique that was already successfully used to analyse HERA data to the more demanding case of proton-proton collisions at LHC energies.

The technique can be used to constrain PDF uncertainties, e.g. at high momentum transfers, from data that will be measured at LHC and allows the consistent inclusion of final state observables in global QCD analyses. This will help increase the sensitivity of LHC to find new physics as deviations from the Standard Model predictions.

Even for the large kinematic range for the parton momentum fractions x_1 and x_2 and of the squared momentum transfer Q^2 accessible at LHC, grids of moderate size seem to be sufficient. The single inclusive jet cross-section in the central region $|y| < 0.5$ can be calculated with a precision of 0.01 % in a realistic example with $1\ 00$bins in the transverse jet energy range $1\ 00 \leq P_T \leq 5\ 000$GeV. In this example, the grid occupies about $3\ 00$Mbyte computer memory. With smaller grids of order $1\ 0$Mbyte the reachable accuracy is still 0.5 %This is probably sufficient for all practical applications.

References

[1] G. Laštovička-Medin et al., *Precision limits for hera dis cross section measurement*. These proceedings.

[2] M. Klein, B. Reisert, *Determination of the light quark momentum distributions at low x at hera*. These proceedings.

[3] Adloff, C. et al., Eur. Phys. J. **C21**, 33 (2001).

[4] Pumplin, J. et al., JHEP **07**, 012 (2002).

[5] Martin, A. D. et al., Eur. Phys. J **C23**, 73 (2002).

[6] Alekhin, S., Phys. Rev. **D68**, 014002 (2003).

[7] Bassler, U. Thesis, Université Pierre et Marie Curie, Paris, 2003.

[8] Laštovička, T., Eur. Phys. J. **C24**, 529 (2002);
Laštovička, T., Acta Phys. Polon. **B33**, 2867 (2002).

[9] ZEUS Coll., Chekanov, S. et al., Phys. Rev. **D67**, 012007 (2003).

[10] ZEUS Coll., Chekanov, S. et al., Eur.Phys.J **C42**, 1 (2005).

[11] Adloff, C. et al., Eur. Phys. J. **C30**, 1 (2003).

[12] Gribov, V.N. and Lipatov, L.N., Sov.J.Nucl.Phys **15**, 438 (1972).

[13] Altarelli, G. and Parisi, G., Nucl.Phys. **B126**, 298 (1977).

[14] Lipatov, L. N., Sov.J.Nucl.Phys **20**, 94 (1975).

[15] Dokshitzer, Yu. L., JETP **46**, 641 (1977).

[16] Devenish, R. C. E. and Cooper-Sarkar, A. M., *Deep Inelastic Scattering*. Oxford Unviersity Press, Oxford, 2004.

[17] Eidelman, S. et al., Phys. Lett. **B592**, 1 (2004).

[18] Cooper-Sarkar, A. M., J. Phys. **G28**, 2669 (2002).

[19] Thorne, R.S. and Roberts, R.G., Phys. Rev **D57**, 6871 (1998).

[20] Pascaud, C. and Zomer, F., *Qcd analysis from the proton structure function f2 measurement: Issues on fitting, statistical and systematic errors*. LAL-95-05.

[21] Adloff, C. et al., Eur. Phys. J. **C13**, 609 (2000).

[22] Adloff, C. et al., Eur. Phys. J. **C19**, 269 (2001).

[23] Chekanov, S. et al., Eur. Phys. J. **C21**, 443 (2001).

[24] Breitweg, J. et al., Eur. Phys. J. **C12**, 411 (2000).

[25] Chekanov, S. et al., Eur. Phys. J. **C28**, 175 (2003).

[26] ZEUS Coll., Chekanov, S. et al., Phys. Lett. **B539**, 197 (2002).

[27] Chekanov, S. and others, Phys. Rev. **D70**, 052001 (2004).

[28] ZEUS Coll., Chekanov, S. et al., Eur. Phys. J. **C32**, 16 (2003).

[29] Callan, Curtis G. , Jr. and Gross, David J., Phys. Rev. Lett. **22**, 156 (1969).

[30] Altarelli, Guido and Martinelli, G., Phys. Lett. **B76**, 89 (1978).

[31] Cooper-Sarkar, Amanda M. and Ingelman, G. and Long, K. R. and Roberts, R. G. and Saxon, D. H., Z. Phys. **C39**, 281 (1988).

[32] G. Altarelli et al., *Resummation*. These proceedings.

[33] Moch, S. and Vermaseren, J. A. M. and Vogt, A., Phys. Lett. **B606**, 123 (2005).

[34] Bartels, Jochen and Golec-Biernat, K. and Peters, K., Eur. Phys. J. **C17**, 121 (2000).

[35] Adloff, C. and others, Phys. Lett. **B393**, 452 (1997).

[36] Adloff, C. and others, *Determination of the longitudinal proton structure function f(l)(x,q**2) at low q**2*. Preprint CH1prelim-03-043. To appear in the proceedings of 32nd International Conference on High-Energy Physics (ICHEP 04), Beijing, China, 16-22 Aug 2004.

[37] Adloff, C. and others, Phys. Lett. **B520**, 183 (2001).

[38] Klein, M., *On the future measurement of the longitudinal structure function at low x at hera*. Prepared for 12th International Workshop on Deep Inelastic Scattering (DIS 2004), Strbske Pleso, Slovakia, 14-18 Apr 2004.

[39] Feltesse, J., *On a measurement of the longitudinal structure function f_l at hera*. Preprint in preparation. Talk at the Ringberg Workshop (October 2005), to be published in the proceedings.

[40] Willeke, J, *Prospects for operating hera with lower proton energy at hera*. Preprint unpublished memo (October 2005).

[41] Chekanov, S. et al., Phys. Lett. **B553**, 141 (2003).

[42] Towell, R. S. et al., Phys. Rev. **D64**, 052002 (2001).

[43] Portheault, B. Thesis, Univ. Paris XI Orsay, March 2005.

[44] Alexopoulos, T. and others, *Electron deuteron scattering with hera, a letter of intent for an experimental programme with the h1 detector*. DESY-03-194;
Abramovicz, H. et al., *A new experiment for hera*. 2003. MPP-2003-62;
Willeke, F. and Hoffstaetter, G. Talks at the Workshop on the Future of DIS, Durham 2001.

[45] Acosta, D. et al., Phys. Rev. **D71**, 051104 (2005);
Heinemann, B. Talk at this workshop.

[46] Stirling, J. Talk at the Binn workshop on LHC Physics, 2003.

[47] Strikman, M., private communication.

[48] Forte, S., private communication.

[49] Botje, M. and Klein, M. and Pascaud, C. Preprint hep-ph/9609489. HERA Physics Workshop 1996/97.

[50] Chekanov, S. et al., Phys. Lett. **B547**, 164 (2002).

[51] Chekanov S. et al., Eur. Phys. J. **C23**, 615 (2002).

[52] Targett-Adams, C., private communication.

[53] Frixione, S. and Ridolfi, G., Nucl. Phys., **B507**, 315 (1997).

[54] Aurenche, P. and Guillet, J. and Fontannaz, M., Z. Phys. **C64**, 621 (1994).

[55] Giele, W. T. and Glover, E. W. N. and Kosower, D. A., Nucl. Phys. **B403**, 633 (1993).

[56] Whalley, M. R. and Bourilkov, D. and Group, R. C., *The les houches accord pdfs (lhapdf) and lhaglue*. Preprint hep-ph/0508110, 2005.

[57] Glück, M. and Reya, E. and Vogt, A., Z. Phys. **C67**, 433 (1995).

[58] Thorne, R., private communication.

[59] Martin, A. D. and Stirling, W. J. and Roberts, R. G., Phys. Lett. **B354**, 155 (1995).

[60] Martin, A. D. and Roberts, R. G. and Stirling, W. J. and Thorne, R. S., Eur. Phys. J. **C35**, 2004 (2004).

[61] Graudenz, D. and Hampel, M. and Vogt, A. and Berger, C., Z. Phys. **C70**, 77 (1996).

[62] Kosower, D. A., Nucl. Phys. **B520**, 263 (1998).

[63] Stratmann, M. and Vogelsang, W., Phys. Rev. **D64**, 114007 (2001).

[64] Wobisch, M. Thesis, RWTH Aachen, PITHA 00/12 and DESY-THESIS-2000-049, December 2000.

[65] Ratcliffe, P. G., Phys. Rev. **D63**, 116004 (2001).

[66] Dasgupta, M. and Salam, G. P., Eur. Phys. J. **C24**, 213 (2002).

[67] Nagy, Z., Phys. Rev. **D68**, 094002 (2003);
Nagy, Z., Phys. Rev. Lett. **88**, 122003 (2002);
Nagy, Z. and Trocsanyi, Z., Phys. Rev. Lett. **87**, 082001 (2001).