

DETERMINATION OF PARTON DISTRIBUTION FUNCTIONS, IMPACT OF HERA DATA AND CONSEQUENCES FOR THE LHC*

EMMANUELLE PEREZ

CERN, PH Division
Geneva 23, CH-1211 Switzerland

(Received May 12, 2008)

Parton distribution functions are non-perturbative inputs necessary to calculate cross-sections for scattering processes involving hadrons in the initial state. They are obtained by fitting theoretical predictions to various sets of experimental measurements, many of which come from Deep Inelastic Scattering experiments. Among those, the experiments at the HERA electron-proton collider have played a crucial role, from the beginning of the data-taking, with the surprising observation of the strong rise of the structure function F_2 at low Bjorken x in 1993, to its end, with the direct measurement of the structure function F_L in 2008. This paper presents an overview of HERA measurements which constrain most the proton structure, and of our current knowledge of parton distribution functions (pdf) in the proton. The impact of the current pdf uncertainties on predictions for cross-sections at the LHC collider is also discussed.

PACS numbers: 12.38.-t

1. Deep Inelastic Scattering

The scattering of a point-like probe (as a lepton) on a composite target (as a proton) has been used for long to underpin the target contents. In the elastic regime $lp \rightarrow lp$, the scattering, which proceeds mainly via photon exchange as depicted in Fig. 1 (left), depends on the proton structure only via the electric and magnetic form factors of the proton. Elastic scattering dominates the lp cross-section when the virtuality Q^2 of the exchanged photon is small. At larger Q^2 , the proton usually breaks up (inelastic regime) and the cross-section is expressed in terms of structure functions, see *e.g.* [1] for a general introduction.

* Presented at the School on QCD, Low- x Physics, Saturation and Diffraction, Copanello, Calabria, Italy, July 1–14, 2007.

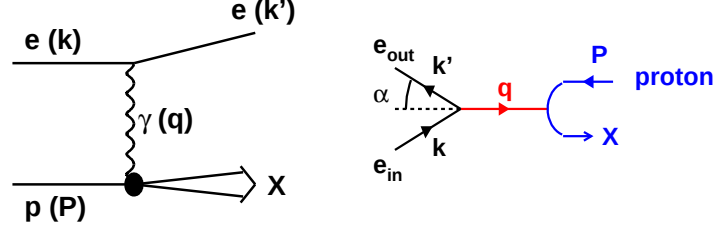


Fig. 1. (left) Generic diagram for neutral current electron–proton scattering, and definition of four-momenta used in the text. The exchanged boson can also be a Z boson. (right) The DIS process viewed in the Breit frame.

1.1. General formalism

The kinematics of the Deep Inelastic Scattering (DIS) process is fully characterized by two variables, usually taken amongst the Lorentz invariant x , Q^2 and y , defined as:

$$Q^2 = -q^2, \quad x = \frac{Q^2}{2(P \cdot q)}, \quad y = \frac{(P \cdot q)}{(P \cdot k)}, \quad (1)$$

where P , k and q denote, respectively, the four-momenta of the incoming proton, of the incoming lepton, and of the exchanged boson (see Fig. 1). These variables are related via $Q^2 = Sxy$, where S is the square of the lepton–nucleon centre-of-mass energy. It is also useful to introduce the squared invariant mass W^2 of the final state X , *i.e.* the centre-of-mass energy in the γp system. The Neutral Current (NC) DIS differential cross-section can be written from the product of the leptonic current $L_{\mu\nu}$, fully calculable, with the hadronic current $W_{\mu\nu}$. The latter can be decomposed along the three tensors which satisfy the requirements of current conservation. This leads to the introduction of three structure functions F_2 , F_L and F_3 , and to the expression

$$\frac{d^2\sigma_{\text{NC}}^{\pm}}{dx dQ^2} = \frac{2\pi\alpha^2}{xQ^4} [Y_+ F_2 \mp Y_- x F_3 - y^2 F_L] \quad (2)$$

for the $e^{\pm}p$ cross-section, with $Y_{\pm} = 1 \pm (1 - y)^2$ (see *e.g.* [2] in these proceedings). Alternatively, this expression can be obtained by considering the contributions to $W_{\mu\nu}$ of the exchange of a boson with a definite helicity $\lambda = +1, -1, 0$. This is most easily seen by considering the DIS process in the “Breit frame”.

The Breit frame is defined by $2x\vec{P} + \vec{q} = \vec{0}$, and the z -axis is taken to be the direction of the photon, see Fig. 1 (right). In the partonic interpretation

of DIS, the hit quark “bounces back” on the photon in this frame, the outgoing quark having the same energy as the incoming quark. Writing the four-vectors of the proton and of the photon, respectively, $P = (E_p, 0, 0, -E_p)$ and $q = (E_\gamma, 0, 0, q_z)$, one gets $E_p = q_z/2x$. From the relations:

$$\begin{aligned} (2xP + q)^2 &= -Q^2 + 4x(P \cdot q) = Q^2 \\ &= (2xE_p + E_\gamma)^2 = (q_z + E_\gamma)^2 \end{aligned}$$

and from $Q^2 = -q^2 = q_z^2 - E_\gamma^2 = (q_z - E_\gamma)(q_z + E_\gamma)$, one deduces that $E_\gamma = 0$: in the Breit frame, the photon carries zero energy. Looking now from the lepton side: the incoming and outgoing electrons have the same energy E_e since $E_\gamma = 0$. They have an opposite p_z and, from $p_z(e_{\text{in}}) = p_z(e_{\text{out}}) + Q$, one deduces that $p_z(e_{\text{in}}) = Q/2$. The product $(P \cdot k) = S/2 = Q^2/(2xy)$ can be written as $(1/2x)(E_e Q + Q^2/2)$, from which one obtains $E_e = Q(2-y)/(2y)$. Combined with $p_z(e_{\text{in}}) = Q/2$, this leads to $\cos \alpha = y/(2-y)$, where the angle α is defined in Fig. 1.

The probability amplitudes for ep scattering can then be written using $d_{1/2,1/2}^{1/2}(\alpha) = \cos \frac{\alpha}{2}$ and $d_{1/2,-1/2}^{1/2}(\alpha) = \sin \frac{\alpha}{2}$. For example, for a transverse boson with $J_z = +1$, the scattering is possible only if the projection of the spin of the incoming (outgoing) electron on the z -axis is $1/2$ ($-1/2$). For a right-handed (left-handed) electron, these two conditions both occur with a probability amplitude $\cos \frac{\alpha}{2}$ ($\sin \frac{\alpha}{2}$), such that the DIS cross-section involving a $J_z = +1$ boson is proportional to $(\cos^2 \frac{\alpha}{2})^2 + (\sin^2 \frac{\alpha}{2})^2$. One then obtains the following y -dependences for the various terms entering the DIS cross-section:

- For transverse bosons: $(\cos^2 \frac{\alpha}{2})^2 + (\sin^2 \frac{\alpha}{2})^2 \propto (1-y)^2 + 1$.
- For longitudinal bosons: $2(\cos \frac{\alpha}{2} \sin \frac{\alpha}{2})^2 \propto 2(1-y)$. This gives us the important result that the contribution of longitudinal photons vanishes at $y = 1$.
- At high Q^2 , the exchange of a Z boson contributes significantly, leading to a parity-violating term: $\propto y(2-y)$.

Putting all the pieces together:

$$d\sigma_{\text{NC}}^\pm \sim [(1-y)^2 + 1] F_T + 2(1-y)F_L \mp [1 - (1-y)^2] xF_3, \quad (3)$$

where the structure functions F_T , F_L and xF_3 contain the dependence on the nucleon structure, for, respectively, the (parity-conserving) contribution due to transverse bosons, to longitudinal bosons, and the parity-violating term. Defining $F_2 = F_T + F_L$, the sum of the first two terms can be rewritten as

$Y_+ F_2 - y^2 F_L$, and one recovers Eq. (2). The reduced cross-section $\tilde{\sigma}^\pm$ for $e^\pm p$ scattering is then defined as:

$$\tilde{\sigma}^\pm = \frac{d\sigma_{\text{NC}}^\pm}{dx dQ^2} \frac{x Q^4}{2\pi\alpha^2} \frac{1}{Y_+} = F_2 \mp \frac{Y_-}{Y_+} x F_3 - \frac{y^2}{Y_+} F_L. \quad (4)$$

A similar expression, with similar structure functions, can be written for Charged-Current (CC) DIS, where a W boson is exchanged instead of a γ or Z , leading to a neutrino in the final state.

1.2. Partonic interpretation of NC DIS

In the quark-parton model (QPM), the ep cross-section is written as an incoherent sum of probabilities of scattering from single free quarks. Writing down the cross-section for eq scattering¹, $d\sigma^{eq}/dy = 2\pi\alpha^2/Q^4 e_q^2 [1 + (1-y)^2]$, denoting by $q_i(x)$ the probability to find a quark q_i in the proton, with momentum fraction x , and identifying with Eq. (2), one obtains:

$$F_2 = \sum_i e_{q_i}^2 x [q_i(x) + \bar{q}_i(x)].$$

Under the simplifying assumption that only the u and d flavors contribute, NC DIS in lepton-proton collisions gives information on $4(u + \bar{u}) + (d + \bar{d})$. In lepton-deuteron collisions², another combination is measured, $\sim u + d + \bar{u} + \bar{d}$, such that having both lp and ld measurements allows to “separate” the flavors. The structure function $x F_3$ is driven by the difference $\sigma(e^+p) - \sigma(e^-p)$ and can be written as $x F_3 \sim \sum_i c_i x [q_i(x) - \bar{q}_i(x)]$, where the c_i depend on the gauge quantum numbers of the partons q_i . It contains a term due to γZ interference, proportional to the product of axial couplings $a_e a_q$, and a term from pure Z exchange, proportional to $a_e v_e a_q v_q$. In the “QCD improved parton model”, $q_i(x)$ is to be replaced by the parton density function (pdf) $q_i(x, Q^2)$ and these two simple expressions for F_2 and $x F_3$ hold at leading order (LO) only. Beyond LO, they can be written as convolutions of pdfs with calculable coefficient functions.

Considering $\gamma q \rightarrow q$ in the Breit frame, one sees that helicity conservation together with the fact that quarks have spin 1/2 requires that the photon be transverse (see Fig. 2). Hence F_L is zero at leading order. A non-zero

¹ The variable y is related to the scattering angle θ^* of the lepton in the eq centre-of-mass frame via $y = (1 - \cos \theta^*)/2$, with $y = 0$ ($y = 1$) corresponding to forward (backward) scattering. The two factors $(1$ and $(1-y)^2$) in $d\sigma^{eq}/dy$ correspond to the two configurations, where the lepton and the quark have the same or the opposite helicity.

² Use is made here of the isospin symmetry relating a density q_n in the neutron to that, q_p , in the proton, *i.e.* $u_n = d_p \equiv d$, $d_n = u_p \equiv u$.

F_L appears at next-to-leading order (NLO), when the final state consists of qg , as may happen when the scattered quark radiates a gluon, or when the incoming quark has a non-zero transverse momentum, as can happen when the interacting quark comes from a gluon splitting $g \rightarrow q\bar{q}$. Hence: at NLO, F_L is given by a convolution involving the “quark singlet” $\sum(q + \bar{q})$ (contribution from $q \rightarrow qg$) and the gluon density (contribution from $g \rightarrow q\bar{q}$); away from the “valence” region, F_L is roughly proportional to the gluon density; and F_L is sensitive to the k_T of quarks produced in $g \rightarrow q\bar{q}$.

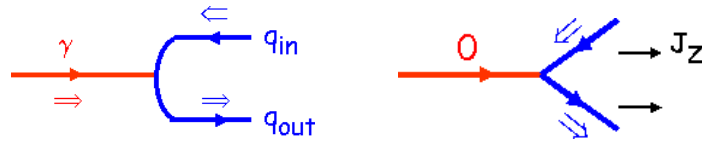


Fig. 2. Illustration of the fact that F_L is zero at LO due to the spin 1/2 of the quarks. Non-zero contributions appear beyond LO.

As illustrated in Fig. 3, the Charged Current DIS e^+p cross-section goes as $(1-y)^2 xD + x\bar{U}$, while the CC e^-p cross-section goes as $(1-y)^2 x\bar{D} + xU$, with $U = u + c$ and $D = d + s + b$. This allows to write a “partonic” expression for the CC structure functions and this shows that:

- the total CC cross-section is much larger in e^-p collisions than in e^+p , because U is larger than D at not too small x , and because the e^+p cross-section is reduced by the angular factor $(1-y)^2$.
- CC DIS brings important information to separate the up and down quark pdfs.

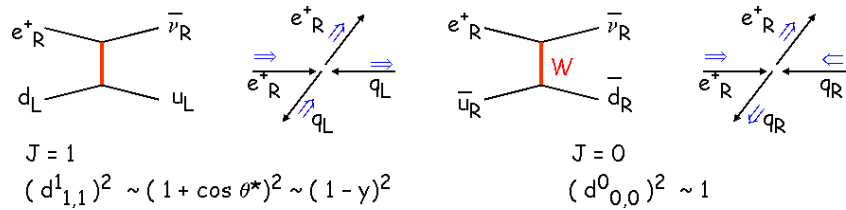


Fig. 3. The different “angular dependences” of CC DIS, depending on the angular momentum J_z of the system.

Neutrino DIS is very similar to CC ep DIS. But because of the very weak cross-sections, experiments used a heavy target, *e.g.* lead (CCFR) or iron (NuTeV), which are nearly isoscalar target (*i.e.* the number of protons is nearly equal to the number of neutrons). When both νN and $\bar{\nu}N$ measurements are available, the sum of νN and $\bar{\nu}N$ cross-sections gives access to the

“CC F_2 ” structure function, $\sim x(U+D+\bar{U}+\bar{D})$, and their difference gives information on the CC parity-violating structure function, $\sim x(U-\bar{U}+D-\bar{D})$. However, the νN data need to be corrected to an isoscalar target, and for nuclear effects (the fact that a density $xf(x)$ in a nuclei differs from $xf(x)$ in the proton). Both corrections appear to be not trivial and model dependent.

2. The HERA collider and experiments

The HERA machine is situated at DESY in Hamburg, Germany. From 1992 until the end of June 2007, it collided 27.6 GeV electrons (or positrons) with 920 GeV protons, leading to a centre-of-mass energy $\sqrt{S} = 318$ GeV. The kinematic domain explored by the two colliding experiments, H1 and ZEUS, is shown in Fig. 4. Compared to fixed target experiments, HERA provides access to much larger values of Q^2 (up to $\sim 10^5$ GeV²), and to much smaller values of x (down to $\sim 10^{-6}$, with most of the low x domain, down to $\sim 10^{-5}$, still in the perturbative region with Q^2 above 1 GeV²). During a first phase of the data-taking (1992 – 2000), called “HERA-I”, both experiments collected ~ 130 pb⁻¹ of unpolarised data, consisting mainly of e^+p data. The second phase, “HERA-II”, started in 2003 after a significant luminosity upgrade, and ended on June 30th, 2007. The experiments collected about 500 pb⁻¹ of e^+p and e^-p polarized data, with a typical lepton beam polarization of 30–40%. The last three months of data taking were dedicated to a “low energy” run, with a much lower proton beam energy, in view of providing a direct measurement of the longitudinal structure function F_L .

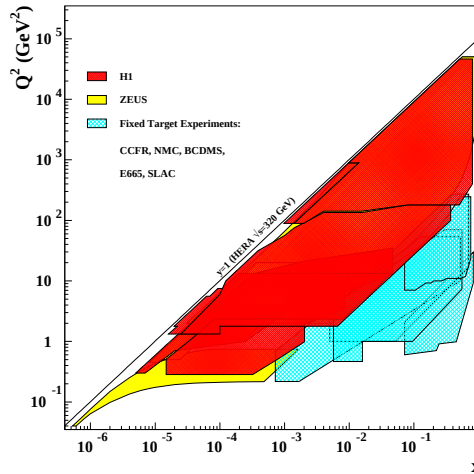


Fig. 4. The kinematic plane in (x, Q^2) accessed by the HERA experiments, and by previous fixed target experiments.

Both the H1 and the ZEUS detectors are asymmetric, reflecting the beam energies asymmetry. By convention, the incoming proton beam defines the positive z -direction, such that a large polar angle for the scattered electron, $\theta_e \sim \pi$, corresponds to a very small deflection. In most of the kinematic plane, both the scattered lepton and the hadronic final state can be detected in the main detector, such that the kinematics of the reaction is overconstrained. Several methods can be used to reconstruct x , Q^2 and y , based on the measurement of the scattered lepton only, on that of the hadronic final state only, or on both. The angular acceptance of the main detector limits however the measurements down to $Q^2 \sim 1.5\text{--}2 \text{ GeV}^2$, and going lower in Q^2 requires special techniques or a dedicated apparatus, as will be seen later.

Fig. 5 (left) shows the spread which existed in parameterizations of F_2 , prior to the first HERA data. The first HERA results, presented in 1993, were based on 30 nb^{-1} of data taken in 1992 and showed a surprising [3], strong rise of F_2 towards low x . Extrapolation from pre-HERA data indicated a “flattish” F_2 at low x — and that was also what came out from Regge-like arguments. However, it was known [4] that in QCD, the gluon distribution should rise at low x , for Q^2 high enough. Approximate solutions of the DGLAP equation indicated indeed that a “flat” gluon distribution at a “starting” scale evolves into a rising distribution at higher scales. What was unknown was “where”, in x and in Q^2 , the rise should start. The “starting scale” Q_0^2 , *i.e.* the scale down to which perturbative QCD was taken to

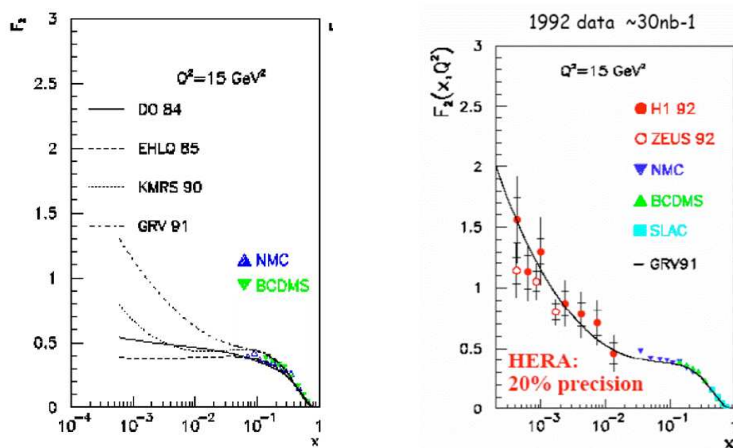


Fig. 5. The first F_2 measurements at HERA are shown on the right. The left plot shows the spread of the theoretical predictions which were consistent with pre-HERA data.

hold, was believed to be a few GeV^2 . And the evolution between this Q_0^2 and $Q^2 = 15\text{--}20 \text{ GeV}^2$ is not long enough to generate a rising gluon from a flattish distribution. Such a rise could only be obtained:

- from a steep input gluon, which could be expected due *e.g.* to large $\ln(1/x)$ terms (resummed in the BFKL evolution equation);
- from pure DGLAP and a flattish gluon, but at a much lower starting scale Q_0^2 .

With the full HERA-I statistics, the statistical precision of these low x and low Q^2 measurements could be reduced below 1%, with a systematic error of about 2%.

With increasing luminosity, an important statistics was accumulated over the whole kinematic domain. Fig. 6 shows an overview of HERA F_2 mea-

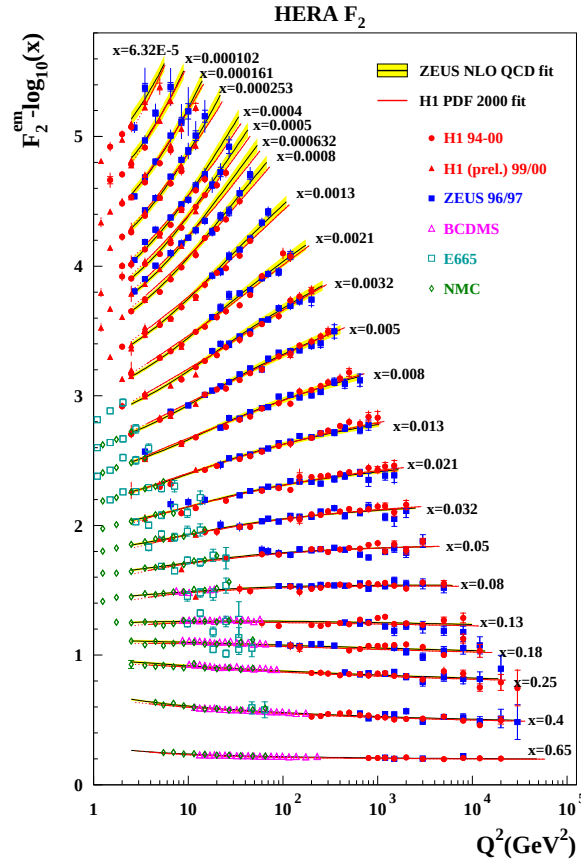


Fig. 6. The logarithmic scaling violations of F_2 , measured in a wide x range [5, 6].

measurements, together with data points from fixed target experiments. One notes the very strong scaling violations observed at low x , which indicate a large gluon density (at leading order within the DGLAP formalism, $\partial F_2 / \partial \ln Q^2 \propto \alpha_s g$). At high x the scaling violations are negative: high x quarks split into a gluon and a lower x quark. The curves overlaid are the result of QCD fits (see later) based on the DGLAP equation. The data show an excellent agreement with DGLAP, over five orders of magnitude in Q^2 and four orders of magnitude in x .

At very high Q^2 , the NC cross-sections are sensitive to the Z -exchange, resulting in $\sigma_{\text{NC}}(e^-p) \neq \sigma_{\text{NC}}(e^+p)$ as was seen in Sec. 1. The NC cross-sections have been measured at high Q^2 both in e^+p and in e^-p collisions, as shown in Fig. 7. The contribution of Z exchange is clearly visible for Q^2 above about 10^3 GeV^2 , with the γ - Z interference being constructive (destructive) in e^-p (e^+p) collisions. The difference between both measurements gives access to the parity violating structure function xF_3 , see Eq. (3) and Fig. 7 (right), which is a direct measure of the valence quark distributions.

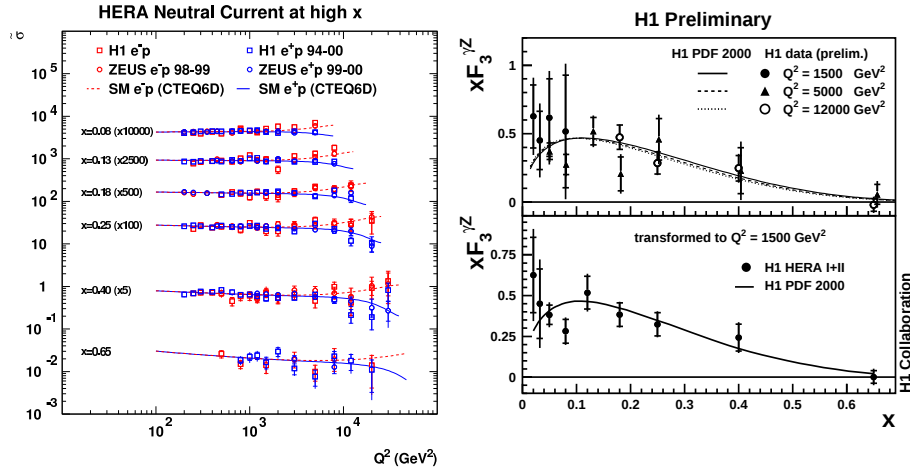


Fig. 7. Left — the NC DIS cross-section as a function of Q^2 for several values of x , measured in e^+p (blue symbols) and in e^-p (red symbols) collisions. The SM predictions are overlaid, as the full and dashed curves, respectively. Right — the extracted structure function xF_3 .

Measurements of charged current DIS provide important constraints on the flavor separation, which are missing from the measurement of F_2 ($\sim 4U + D$) alone. Indeed, σ_{CC}^+ goes as $(1-y)^2 xD + x\bar{U}$ and probes mainly the d density, while σ_{CC}^- goes as $(1-y)^2 x\bar{D} + xU$ and probes mainly the u density. Examples of HERA CC measurements are shown in Fig. 8.

Although the statistics of HERA-I measurements is limited, the precision reaches $\sim 15\%$ for $x \sim 0.1$. With the polarized lepton beams that were available at HERA-II, the measurements, shown in Fig. 8 (right), checked the linear dependence of the cross-section on the lepton polarization P_e , $\sigma_{CC}(e^\pm) \propto (1 \pm P_e)$, as predicted in the Standard Model in which right-handed charged currents do not exist.

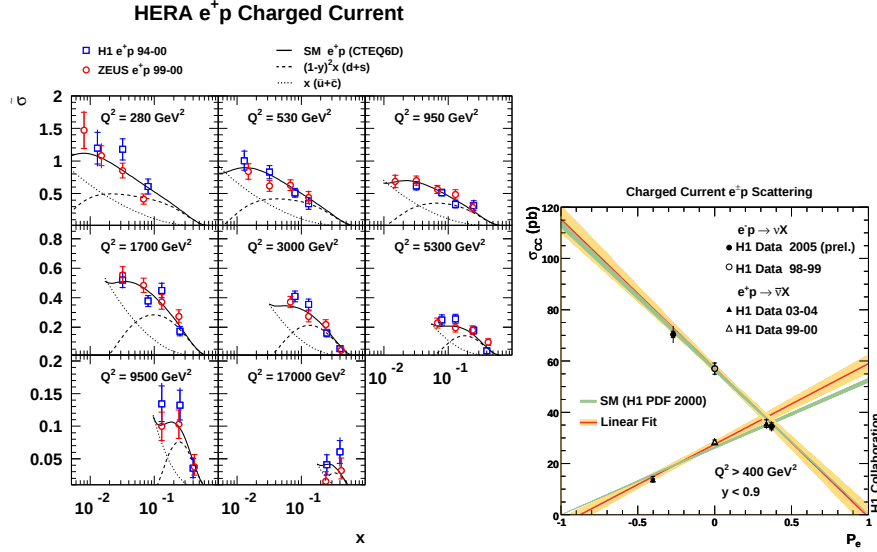


Fig. 8. Left — the CC DIS cross-section measured in e^+p collisions. The overlaid curves show how these measurements disentangle the contributions from up and down quarks [5, 7]. Right — dependence of the CC cross-section on the lepton beam polarization [8].

The charm and beauty contents of the proton have also been measured by the HERA experiments. This can be done via exclusive measurements (exploiting *e.g.* the $D^* \rightarrow D^0 \pi_{\text{slow}} \rightarrow K \pi \pi$ decay chain, or the $b \rightarrow \mu X$ decays), or via semi-inclusive measurements, using silicon vertex devices around the interaction points. Fig. 9 (left) shows the F_2^{bb} measured by the H1 experiment. As the inclusive F_2 , it shows large scaling violations at low x . In Fig. 9 (right), the charm fraction in the proton is shown to be about 25% independently of Q^2 , while the beauty fraction increases rapidly with Q^2 , reaching $\sim 3\%$ at high Q^2 .

As mentioned already, extending the F_2 measurements down to very low Q^2 requires dedicated techniques or detectors. From $Q^2 = 2E_e^0 E_e (1 + \cos \theta_e)$ one sees that, to go down to low Q^2 , one needs to access large angles θ_e , or to lower the incoming electron energy E_e^0 . Large polar angles can be accessed via:

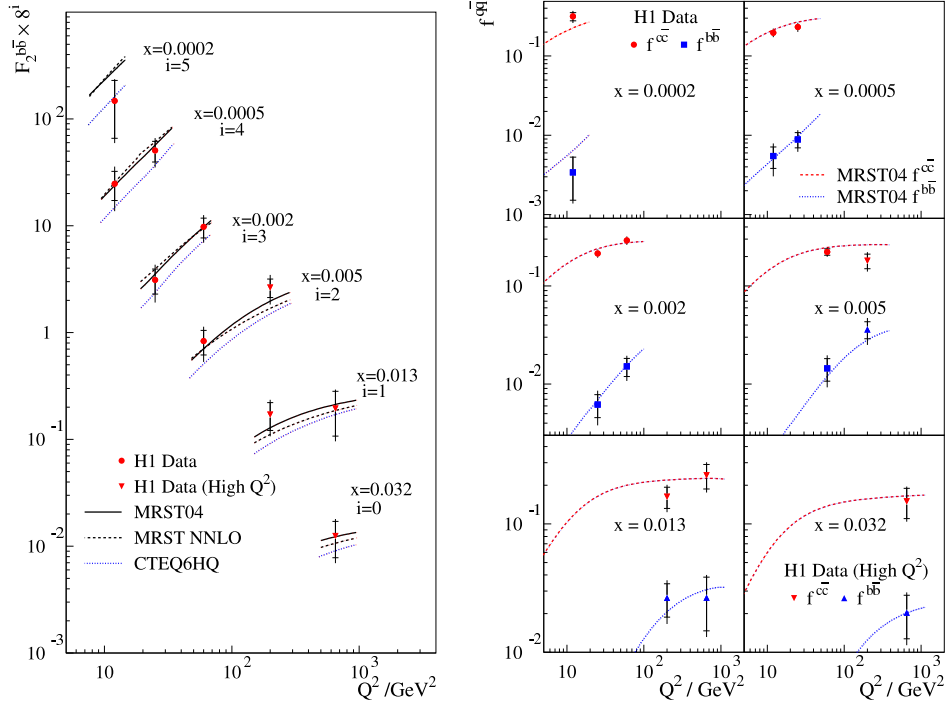


Fig. 9. The measured F_2^{bb} (left) and the fractions of charm and beauty in the proton (right) [9].

- a dedicated apparatus, as the ZEUS Beam Pipe Tracker (BPT), which consisted of a silicon strip tracking detector and an electromagnetic calorimeter very close to the beam pipe;
- shifting the interaction vertex in the forward direction. Two short runs were taken with such a setting, where the nominal interaction point was shifted by 70 cm;
- exploiting QED Compton events: when the lepton is scattered at a large angle θ_e , it may still lead to an observable electron (*i.e.* within the detector acceptance) if it radiates a photon.

Alternatively, initial state radiation events can be used as a way to lower the incoming electron energy, $E_e^0 \rightarrow E_e^0 - E_\gamma$.

All these methods have been exploited at HERA. Fig. 10 shows examples of measurements at lowest Q^2 . In Fig. 10 (left), one notes the good agreement with fixed target data, and that F_2 continues to rise at low x , even at the lowest Q^2 . Figure 10 (right) shows low Q^2 measurements obtained

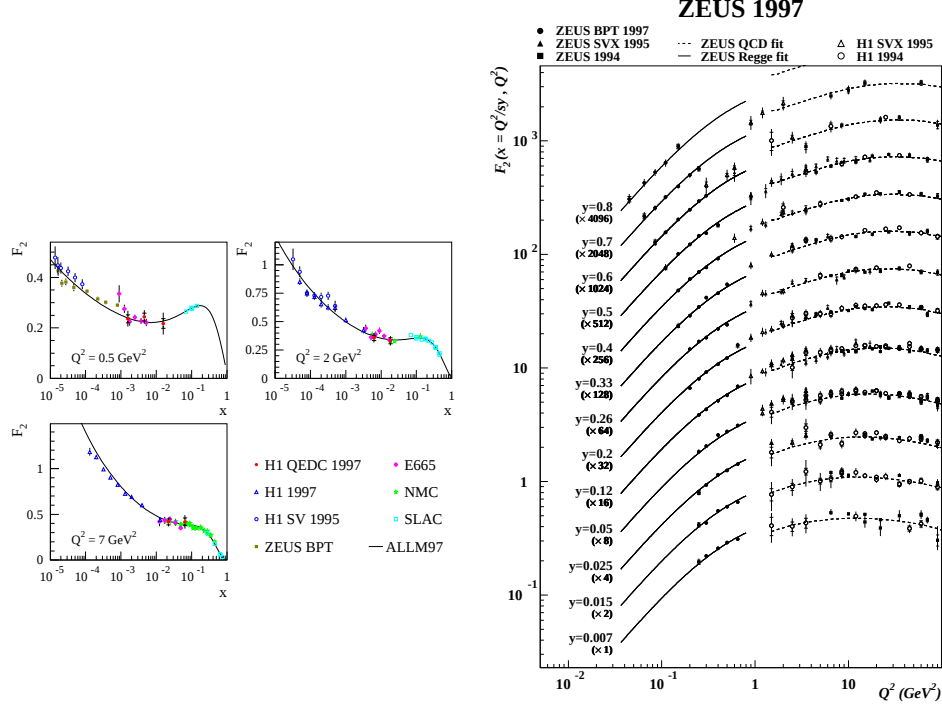


Fig. 10. Examples of low Q^2 measurements at HERA: using QED Compton events, using a dedicated detector and from special runs where the interaction vertex was shifted [10].

using the ZEUS BPT, together with results of the shifted vertex runs. As required by the conservation of the electromagnetic current, $F_2 \propto Q^2$ as Q^2 goes to zero.

3. Determination of parton density functions: QCD fits

3.1. Introduction and generalities

The determination of parton density functions (pdfs) from fits of the DGLAP theory to data from all relevant experiments (so-called “QCD fits”) is carried out by two groups mainly, resulting in the “MRST/MSTW” and the “CTEQ” series. A fit is defined in particular by:

- The order of the fit. Most of the fits are performed at NLO. Leading order fits are still of interest, *e.g.* for Monte Carlo simulations. With the recent, full calculation of the DIS cross-section at NNLO, first NNLO fits are becoming available. For fits performed beyond LO, one needs to choose the renormalization scheme, usually taken to be the

$\overline{MS}$ scheme. One also needs to choose the factorization and renormalization scales, μ_F and μ_R , which will be used in the theoretical calculation. For DIS, a common choice is to take $\mu_F^2 = \mu_R^2 = Q^2$.

- The treatment of heavy flavors (HF) — here the case of the charm quark is taken as an example. In the “zero-mass variable flavor number scheme” (ZM-VFNS), the charm density is set to zero below $Q^2 \sim 4m_c^2$. Above this threshold, the charm is generated by gluon splitting and is treated as massless. The drawback of this approach is that it ignores charm mass effects in the threshold region. In contrast, in the “fixed-flavor number scheme” (FFNS), there is no pdf for the charm and bottom, *i.e.* there are only three active flavors. For W^2 above the production threshold, the DIS production of charm proceeds via photon–gluon fusion, $\gamma g \rightarrow c\bar{c}$. The drawback of this treatment is that the calculations involve terms in $\ln(Q^2/m_c^2)$ which become large at high Q^2 and would need to be resummed. The “state-of-the-art” approach, called “general-mass variable flavor number scheme”, somehow interpolates between the ZM-VFNS and the FFNS. It is not easy to implement, especially at NNLO. To illustrate the importance of HF treatment in QCD fits, note that the main difference between the CTEQ 6.1 and CTEQ 6.5 fits, for which the predictions for W production at the LHC differ by $\sim 8\%$, was precisely an improved treatment of heavy flavors in CTEQ 6.5 [11] (see Sec. 3.3).
- Which datasets and which data points are included in the fit. DIS measurements at very low x , where DGLAP may break down, may be cut away from the fits. The same holds for data points at very low W^2 , which are likely to be affected by large higher twists effects, and for points at very high x , for which one may need a resummation of $\ln(1-x)$ terms.
- The starting scale Q_0^2 of the fit. It should not be too high, in order to keep as much data as possible, but should be high enough to remain in the perturbative domain. Typical values for Q_0^2 range between 1 GeV^2 and a few GeV^2 .
- Which is the set of pdfs that we want to parameterize at the starting scale of the fit. One does not fit the eleven quark and gluon distributions since the data do not contain enough information to disentangle them all. Instead, well-defined combinations of parton pdfs are fitted. In addition to the gluon density, at least two quark distributions are needed (because F_2 is the sum of a singlet and a non-singlet terms, which evolve differently). For example, one can choose to parameterize g , $u_{\text{val}} = u - \bar{u}$ and $d_{\text{val}} = d - \bar{d}$, the total sea distribution $\Sigma\bar{q}$,

and $\bar{d} - \bar{u}$. An alternative choice, used *e.g.* in [5], uses g , $U = u + c$, $D = d + s$, $\bar{U}$ and $\bar{D}$. Once this “flavor decomposition” is chosen, one needs to choose the parameterization, *i.e.* which function will be used to parameterize $xf(x)$ at the starting scale Q_0^2 . For example, $xf(x) = Ax^\alpha(1-x)^\beta(1+p_0x+p_1\sqrt{x}+\dots)$ can be used. The parameterization should be flexible enough to allow for a good fit, but should still avoid unstable fits and secondary minima. One should also keep in mind that in some cases, the choice of the parameterization “hides” some assumptions. For example, the common “valence-like” choice $x\bar{d} - x\bar{u} = Ax^\alpha(1-x)^\beta$ assumes that $x(\bar{d} - \bar{u})$ goes to zero as $x \rightarrow 0$. So far, we have no experimental data to confirm or infirm this assumption.

- Some assumptions are often needed, to supplement the lack of sensitivity of the fitted data. For example, if the fit does not include measurements which are sensitive to $s - \bar{s}$, one will assume that $s = \bar{s}$.
- Usually, “sum rules” are imposed in the fits, which allow the number of free parameters to be reduced. The “number sum rules” read as:

$$\int_0^1 u(x) - \bar{u}(x) dx = 2, \quad \int_0^1 d(x) - \bar{d}(x) dx = 1,$$

$$\int_0^1 q(x) - \bar{q}(x) dx = 0 \quad \text{for} \quad q = s, c, b.$$

And the “momentum sum rule”,

$$\int_0^1 x [g(x) + \Sigma(q(x) + \bar{q}(x))] dx = 1$$

helps fix the gluon normalization, and connects the low x and high x behaviors of the gluon density.

Once this framework is defined, use is made of the DGLAP equations to obtain $xf(x, Q^2)$ at any Q^2 from the parameterized $xf(x, Q_0^2)$. This allows to calculate the theoretical cross-sections of the processes of interest (DIS, Drell–Yan lepton production, jet production, ...) and the parameters can be fitted to the data.

3.2. QCD fits to DIS data

The combination $\sim 4xU + xD$, where xU (xD) denotes the pdf of up-like (down-like) quarks, is well constrained at low and medium Bjorken x via Neutral Current DIS; the scaling violations of F_2 , *i.e.* its logarithmic dependence on the four-momentum transfer squared Q^2 , give access to the gluon density at low and medium x ; additional constraints from Charged Current DIS allow some separation between U and D to be made, although with a limited precision. As a result, QCD fits to HERA data alone give access to all parton distribution functions in the proton. Examples [5, 12] are illustrated in figure 11. The three fits displayed in this figure are in reasonable agreement, within their respective uncertainties. It is interesting however to note that there are differences which are not embedded in the quoted uncertainties. This holds in particular for the valence distributions. The differences between the “H1 fit” and the “ZEUS fit” are mainly due to: a different treatment of heavy quarks; different parameterizations at the starting scale of the fit; and different data included in the fit (the H1 fit uses CC DIS data to perform the flavor separation, while in the ZEUS fit, this separation is brought mainly by fixed-target (BCDMS, NMC and E665) μd data complementing the μp measurements). These three ingredients are part of the definition of any fit, and the uncertainties of the fits obviously do not account for variations around these choices.

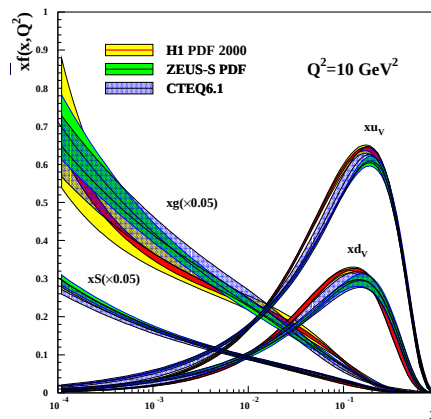


Fig. 11. Distributions of valence quarks densities xu_v and xd_v , of the gluon density xg and of the density of sea quarks xS obtained from various QCD fits.

In [12], the main fit (which includes, in addition to ZEUS data, μp and μd data, the NMC data for F_2^d/F_2^p and CCFR data on xF_3) was compared to a similar fit based on ZEUS data alone. This comparison showed that

the gluon and sea densities are mainly determined by the HERA data for x below ~ 0.1 . However, adding fixed target data reduces the uncertainties on the valence distributions by a factor of about two, compared to what is obtained from HERA data alone. In particular, a reasonable determination of d_{val} requires deuterium data, which are more constraining than HERA high Q^2 charged-current data.

The precision on the gluon density obtained from the ZEUS data alone was shown to improve [13] when jet data from DIS and photoproduction are added in the fit. HERA jet data do not bring strong constraints on the gluon density at high x , due to the limited statistics (constraints at high x are brought by Tevatron jet data). But they are useful for medium x , since HERA jet cross-sections have small systematic uncertainties (typically 5%). Fig. 12 (left) shows cross-sections of inclusive jet production in DIS, as measured by ZEUS. Fig. 12 (right) shows that the inclusion of these data in the fit reduces the uncertainty of the gluon density by a factor of ~ 2 in the mid- x region, $x = 0.01$ – 0.4 . It is also interesting to note that both fits

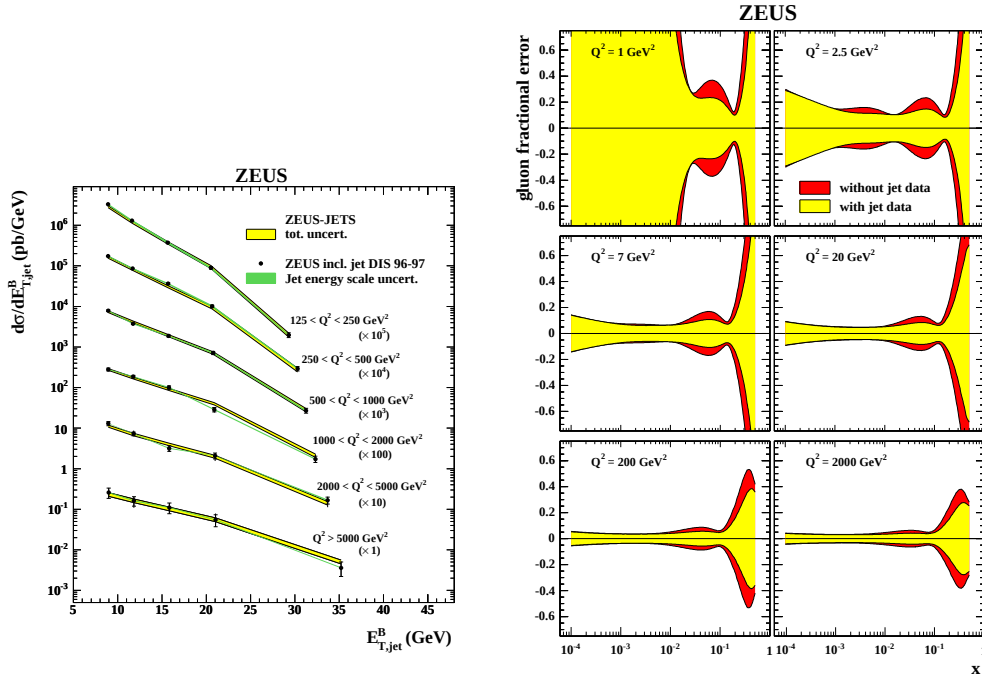


Fig. 12. Left — differential cross-sections of inclusive jet production in DIS. Right — the relative uncertainty of the gluon density as a function of x for several values of Q^2 , when these jet data are included (light, yellow band) or are not included (dark, red band) in a QCD fit to ZEUS inclusive DIS data [13].

with and without the jet data lead to the same shape for the gluon density, indicating that there is no tension between the jet and the inclusive DIS data.

3.3. Global QCD fits

Global QCD fits usually include, in addition to inclusive DIS data: the Tevatron jet data (which constrain the gluon density at high x), the Drell–Yan measurements $pN \rightarrow \mu\mu$ (which constrain the quark densities at high x and the difference $x\bar{d} - x\bar{u}$), the production of muon pairs in neutrino–nucleon scattering ($\nu_\mu s \rightarrow \mu c \rightarrow \mu\mu X$ and c.c., which constrain the s and $\bar{s}$ densities), and the rapidity-binned charge asymmetry of W production at Tevatron (which brings constraints on the d/u ratio at medium x). Recent fits also include HERA jet data and the HERA measurements of $F_2^{c\bar{c}}$ and $F_2^{b\bar{b}}$. This leads to typically ~ 3000 data points in the fits. The latest MSTW07 fit [15] also includes Tevatron Run II data and the CDF Run II W lepton asymmetry measurements. The latest Tevatron jet data tend to slightly lower the gluon density at high x , but, within uncertainties, it is consistent with previous fits. Typically in global fits: the sea densities come from HERA F_2 measurements, the sensitivity on the gluon density at low x is given by F_2 scaling violations measured at HERA, the quark and gluon densities at high x are constrained by Tevatron jet data, and the flavor separation is obtained from fixed target data.

Let us illustrate how non-DIS data can bring specific constraints in QCD fits with the example of Drell–Yan measurements. In global fits, $\bar{d} = \bar{u}$ was a “natural” assumption until the NA51 experiment at CERN reported that $\bar{d} > \bar{u}$ at $x = 0.18$ (some hints had been previously reported by the NMC experiment). This was followed-up by the E866 experiment at Fermilab, which measures di-muon production in pp and pd collisions in a fixed target experiment with $E_{\text{beam}} = 800$ GeV. Let x_1 and x_2 be the Bjorken x variables of the partons involved in the hard subprocess, from the beam and the target nucleon, respectively. Kinematics imposes $x_1 > x_2$, and, in the phase space where $x_1 \gg x_2$, one has:

$$\sigma_{pp} \propto \frac{4}{9}u(x_1)\bar{u}(x_2) + \frac{1}{9}d(x_1)\bar{d}(x_2), \quad \sigma_{pn} \propto \frac{4}{9}u(x_1)\bar{d}(x_2) + \frac{1}{9}d(x_1)\bar{u}(x_2)$$

such that:

$$\frac{\sigma_{pd}}{2\sigma_{pp}} \sim \frac{1}{2} \frac{1 + \frac{1}{4} \frac{d(x_1)}{u(x_1)}}{1 + \frac{1}{4} \frac{d(x_1)}{u(x_1)} \frac{\bar{d}(x_2)}{\bar{u}(x_2)}} \left[1 + \frac{\bar{d}(x_2)}{\bar{u}(x_2)} \right].$$

In the relevant domain of x_1 , the ratio $d(x_1)/u(x_1)$ is quite well known, such that the ratio $\sigma_{pd}/2\sigma_{pp}$ gives access to the ratio $\bar{d}/\bar{u}$ at low x . Fig. 13

shows the measurement of the E866 experiment [14], which extends down to $x_2 \sim 0.03$. Note the spread of the theoretical predictions before these data were included in the fits. The ratio $\bar{d}/\bar{u}$ as extracted by E866 is shown in Fig. 13 (right), and clearly demonstrates that $\bar{d} > \bar{u}$ at low x .

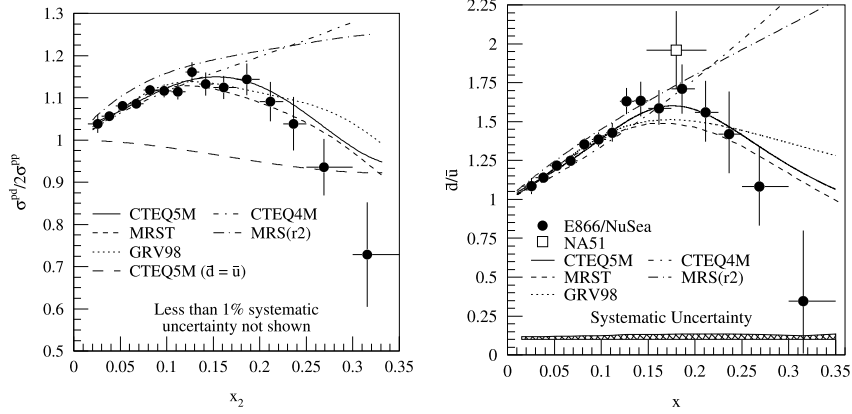


Fig. 13. The ratio $\sigma_{pd}/2\sigma_{pp}$ measured by the E866 experiment (left), and the ratio $\bar{d}/\bar{u}$ extracted from this measurement (right).

The remainder of this paragraph describes some recent updates of global fits.

An important update of the CTEQ global fits [11] comes from the treatment of heavy quarks in a general mass variable flavor number scheme, where mass effects are accounted for both in the kinematics and in the Wilson coefficients. The resulting CTEQ 6.5 pdfs mainly differ from the previous CTEQ 6.1 fit by larger u and d distributions in the region $x \sim 10^{-3}$, for a wide range in Q^2 . This is illustrated in Fig. 14. This results in a $\sim 7\%$ increase of the predicted W and Z cross-sections at the LHC, compared to previous estimates.

The recent inclusion of the final NuTeV dimuon data in global fits, in addition to the previously used CCFR data, allows the strange content of the nucleon to be studied in more details [15, 16]. While previous fits assumed that $s + \bar{s}$ was a constant fraction of the “non-strange sea” $\bar{u} + \bar{d}$ at the starting scale, recent fits give more freedom to $s + \bar{s}$ at high x , and yield indeed an improved χ^2 : the shape of the symmetric strange sea seems to be independent of that of the non-strange sea. Figure 15 shows that $s + \bar{s}$ is smaller than $(\bar{u} + \bar{d})/2$, especially at large x (*i.e.* low W^2), which could be an effect of the strange quark mass. Note that relaxing the assumption on $s + \bar{s}$ in the fits results in a larger uncertainty on these pdfs, which feeds into that of $\bar{u}$ and $\bar{d}$. As a result, the uncertainty on low x antiquarks roughly doubles in the MSTW07 fit, compared to MRST2004.

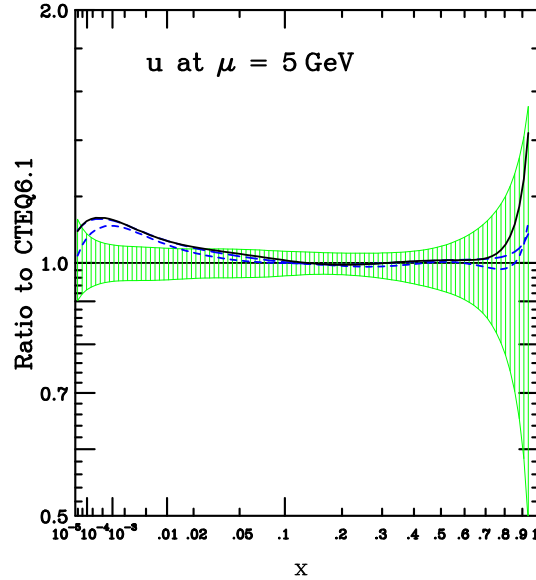


Fig. 14. Ratio of the CTEQ 6.5 u density to that of CTEQ 6.1, at a scale $Q = 5$ GeV (from [11]).

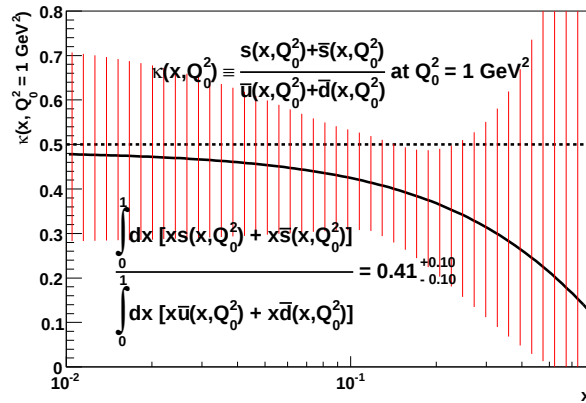


Fig. 15. The ratio of the input $s + \bar{s}$ distribution to that of the non-strange sea $\bar{u} + \bar{d}$ at the starting scale of the MSTW07 fit [15].

Besides these exclusive dimuon data, inclusive cross-sections from neutrino DIS experiments also bring significant constraints on pdfs. Discrepancies between the NuTeV and the older CCFR structure function measurements at high x are now understood by both groups, and the NuTeV data set

is believed to be more reliable. However, CHORUS measurements recently published (obtained with a lead rather than an iron target) also disagree with the NuTeV data at high x . As a result, the latest MSTW07 fit includes the NuTeV and CHORUS data (which replace the CCFR measurements) only for $x < 0.5$. These NuTeV and CHORUS data were analyzed together with the latest Drell–Yan measurements from E866 in [17], in a global fit similar to those performed by the CTEQ Collaboration. This fit yields a d/u ratio which flattens out significantly at high x . A tension is observed at high x : the NuTeV data pull the valence distributions upward (which pulls against the BCDMS and NMC data), while the E866 measurements prefer lower valence distributions at high x . This tension is actually amplified by the nuclear corrections applied to the NuTeV data.

Finally, a first global fit allowing for an “intrinsic charm” component has been carried out in [18]. The existing data have little sensitivity to such a component because it would manifest itself at too large x . As a result, an intrinsic charm component carrying a few percent of the proton momentum is not excluded by the current data. This would enhance drastically *e.g.* the production of a charged Higgs boson ($c\bar{s} \rightarrow H^+$) at the LHC.

3.4. QCD fits and the description of low x and low Q^2 data

An intriguing feature of several QCD fits is that, while at low Q^2 the density of sea quarks continues to rise with decreasing x , the gluon density becomes suppressed. In the fit performed by the ZEUS Collaboration [12], the gluon density even becomes negative³ for $Q^2 \sim 1 \text{ GeV}^2$, see Fig. 16 (left). Problematically, this leads to negative values for the longitudinal structure function F_L at low x . The same feature is observed in the MRST/MSTW fits, where the parameterization of the gluon density at the starting scale can accommodate a negative density at low x : the fits yield a negative F_L at low x , for Q^2 below 2–3 GeV^2 . This may indicate that the approximations done in the NLO QCD calculations are not valid anymore in this low x domain.

If we are not sure of the theoretical framework that should be used at low x , a drastic approach could be to cut out the low x data in the fits. This was tried by the MRST group: a cut $x > x_{\min}$ was made more and more severe, until fits were stable — which was obtained for $x_{\min} \sim 5 \cdot 10^{-3}$! These fits do not describe the HERA data at low x , and give very different predictions from “standard fits”, for many observables at the LHC. Hence, one needs to better understand and overcome the limitations of our calculations at low x .

³ Note that, for Q^2 down to 1 GeV^2 , the fit still describes the rise of F_2 at low x as shown in Fig. 16 (right), even though the gluon density is negative: the approximation $\partial F_2 / \partial Q^2 \propto \alpha_s x g(x, Q^2)$ is not valid at NLO!

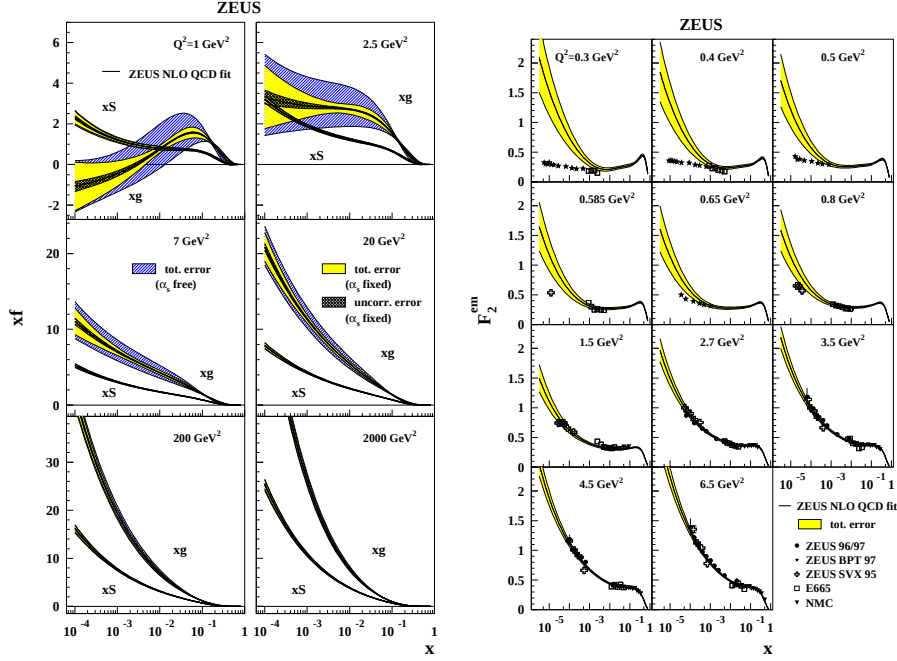


Fig. 16. Left — the gluon and sea densities extracted from the ZEUS NLO fit, shown for different values of Q^2 . Right — compilation of F_2 measurements at low Q^2 .

We know that NLO DGLAP predictions at low x and Q^2 could be wrong due to:

- Large terms in $\alpha_s \ln(1/x)$, which can spoil the convergence of the perturbative expansion. Going to NNLO may cure the problem; or it could be that a full resummation of these $\ln(1/x)$ is necessary for a reliable theoretical prediction. A complete set of NNLO parton distribution functions has recently been obtained by the MSTW group [19], together with their uncertainties. At NNLO, the gluon density at low x and low Q^2 is still negative (even more negative than the NLO gluon). However, this is compensated by positive terms in the $\mathcal{O}(\alpha_s^3)$ coefficient function for F_L . The result is that F_L at NNLO is positive. Recently, a global fit based on a calculation where the $\ln(1/x)$ terms are resummed (NLL BFKL resummation) has been carried out [20]. The resulting predictions for F_L are also positive, and quite different from the NNLO predictions.
- Unitarization (saturation) effects which may tame the low x rise of F_2 , due *e.g.* to gluon recombinations which could make the evolution equations non-linear.

Experimentally, other observables than F_2 are needed to get hints of what is the right approach: the longitudinal structure function F_L , the slopes of F_2 , and, possibly, the study of exclusive final states [21].

The longitudinal structure function F_L : F_L is more directly related to the gluon density than is F_2 , hence it is a good experimental observable to study the importance of the $\ln(1/x)$ terms. Since the DIS cross-sections $d^2\sigma/dxdQ^2$ are proportional to $F_2 - y^2 F_L/Y_+$ (see Eq. (4)), a direct measurement of F_L requires $d^2\sigma/dxdQ^2$ to be measured at two different values of y , such that the contributions of F_2 and F_L can be disentangled. Since $Q^2 = xyS$, this means having measurements at two (or more) different centre-of-mass energies $\sqrt{S}$. The last three months of data taking at HERA have been dedicated to a “low energy” run, where the proton energy was set to 460 and 575 GeV. The low-energy cross-section measurements, combined with the high energy data, yield, in a given (x, Q^2) bin, a measurement of $\sigma \sim F_2 - y^2/(1+(1-y)^2)F_L$ at two different values of $y = Q^2/xs$. Hence, this allows the two structure functions F_2 and F_L to be disentangled directly, and provides a first, direct measurement of F_L . Brand new, preliminary measurements [22] from the H1 experiment are shown in Fig. 17. Confronting these measurements with the theoretical predictions will hopefully tell us whether the NNLO expansion is enough for a reliable prediction of the longitudinal structure function.

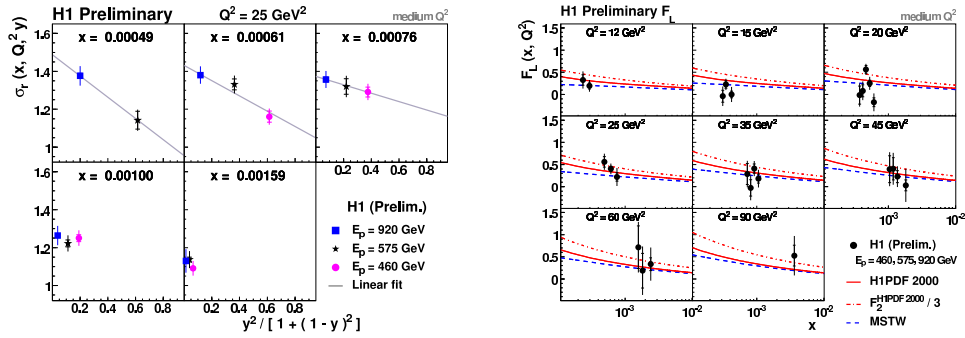


Fig.17. The reduced cross-section measured at different y values and its linear dependence on y^2/Y_+ in the so-called Rosenbluth plot (left), and the extracted values for the longitudinal structure function F_L (right).

The “slopes of F_2 ” and the low Q^2 — high Q^2 transition region: In the double-asymptotic limit, DGLAP predicts that $F_2(x, Q^2)$ is close to $x^{-\lambda(Q^2)}$. A power-behavior is also predicted by the BFKL evolution equation, with $\lambda \sim 0.3 - 0.5$. Experimentally, one can thus extract

$$\lambda(x, Q^2) = \left(\frac{\partial F_2}{\partial \ln x} \right)_{Q^2}$$

and look for a decrease of λ with decreasing x , which may sign a breakdown of the theory due to saturation effects. As shown in Fig. 18 (left), no evidence for such a decrease is observed in the data [23]. This figure shows that the

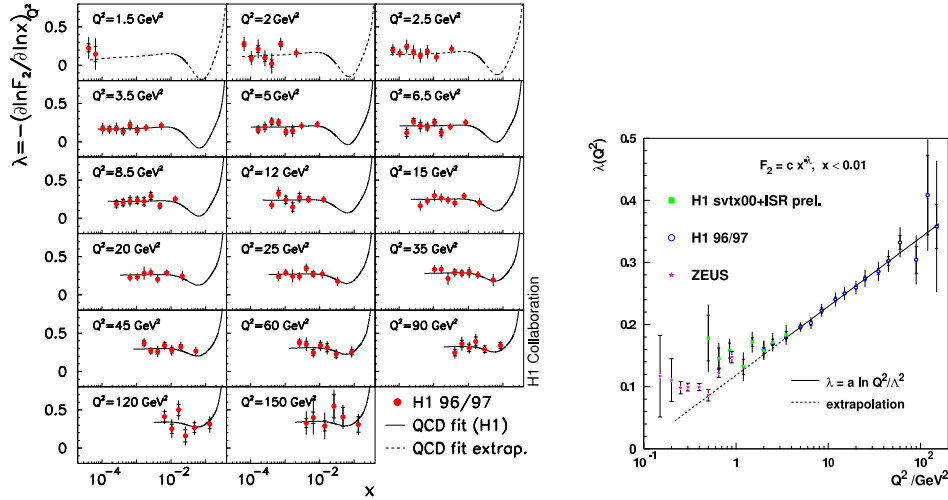


Fig. 18. Left — the derivative of F_2 w.r.t. $\ln x$ as a function of x , in various Q^2 bins, is seen to be independent of x within the errors. The resulting $\lambda(Q^2)$ is shown as a function of Q^2 on the right plot [23].

data can indeed be parameterized by $F_2(x, Q^2) = c(Q^2)x^{-\lambda(Q^2)}$. Figure 18 (right) shows the Q^2 dependence of this $\lambda(Q^2)$. For $Q^2 > 2-3 \text{ GeV}^2$, $\lambda(Q^2)$ depends logarithmically on Q^2 and c is constant to a good approximation, as expected from the DGLAP equations. For Q^2 below $\sim 1 \text{ GeV}^2$, $\lambda(Q^2)$ deviates from a $\ln(Q^2)$ behavior, and tends to a value consistent with Regge theory ($\alpha_{\text{Pom}}(0) - 1 \sim 0.08$). This can be interpreted as a “confinement transition” from partonic degrees of freedom to hadronic degrees of freedom, at a scale of about 0.3 fm. It is interesting to note that dipole models provide a good description of the low Q^2 –high Q^2 transition region, and of Fig. 18 (right) in particular. In the model of Golec-Biernat and Wüsthoff, when the size of the $q\bar{q}$ dipole $\sim 1/Q$ is large compared to the separation of gluons in the target, $R_0 \sim 1/\sqrt{xg(x)}$, the dipole–proton cross-section saturates to a value σ_0 , in such a way that the cross-section exhibits a “geometric scaling” property:

$$\sigma = \sigma_0(1 - \exp(-1/\tau))$$

which involves one variable only, $\tau = Q^2 R_0^2(x)$. This property is indeed borne out by the low x data. While this is not a proof of saturation, this shows that the low x HERA data have many of the attributes of a saturated system. More details can be found in [24].

3.5. Uncertainties on parton densities

A lot of work has been done over the past ~ 5 years to assess rigorously the uncertainties on parton densities extracted from QCD fits [25]. Most of the groups now use the same definition of the χ^2 which is minimized in the fit:

$$\chi^2 = \sum_i \frac{(d_i - \sum_k \beta_{k,i} s_k - t_i)^2}{\alpha_i^2} + \sum_k s_k^2,$$

where d_i denote the measurements, t_i the corresponding theoretical predictions, k labels the sources of systematic uncertainties, $\beta_{k,i}$ is the amount of change of d_i when the source k (for example an energy scale) is shifted by 1σ ($s_k = 1$). Moving coherently the data points by $\beta_{k,i} s_k$ causes a “penalty term” s_k^2 in the χ^2 function. The total uncorrelated uncertainty affecting the measurement i is given by $\alpha_i^2 = \sigma_{i,\text{stat}}^2 + \sigma_{i,\text{uncorr.}}^2$. However, this common χ^2 definition can be used in different ways:

- in the so-called “offset method”: the parameters s_k are set to zero in the central fit — *i.e.*, the central fit is performed without taking into account the correlated systematic errors. Then, for each source of systematic error, s_k is set to ± 1 and the fit is redone. The uncertainty of a given quantity (*e.g.* a parton density) is calculated by adding in quadrature all differences to the quantity obtained in the central fit;
- in the “Hessian method”: the s_k are not fixed, but are parameters of the fit. Technically, they are obtained analytically (the χ^2 is quadratic in s_k , hence $\partial\chi^2/\partial s_k = 0$ leads to a simple matrix equation for the s_k). This means that the central fit is not performed to the “raw data”, but instead, to the data shifted by the optimal setting for the systematic shifts. The error bands on a given quantity are obtained from $\Delta\chi^2 = T^2$ with $T = 1$ or larger.

The offset method gives fitted theoretical predictions which are as close as possible to the “raw” data points. It does not use the full statistical power of the fit to correct the data for the best estimate of the systematic shifts, since it distrusts that systematic uncertainties are Gaussian distributed. The offset method thus appears to be more conservative than the Hessian method. It usually results in larger uncertainties than what is obtained from the Hessian method, when the criterium $\Delta\chi^2 = 1$ is used to get the error bands.

With the Hessian method, model uncertainties (*e.g.* varying α_s , Q_0^2 , ...) are often larger than the fit uncertainties. This is because each “model choice” can result in different values of the systematic shifts — *i.e.*, when changing the “model”, one does not fit the same data points.

The general trend of pdf uncertainties is that: the u quark is much better known than the d quark; the sea and the gluon are well known at medium x ; at high x , the sea and the gluon are poorly known and the valence quarks are reasonably constrained. One should keep in mind that the uncertainties given by the pdf sets do not include additional uncertainties due to: the choice of datasets and cuts applied to the data; the choice of the parameterization; choices done in the “theory”, *e.g.* the treatment of heavy flavors. To illustrate this, Fig. 19 (from [26]) shows the relative uncertainty

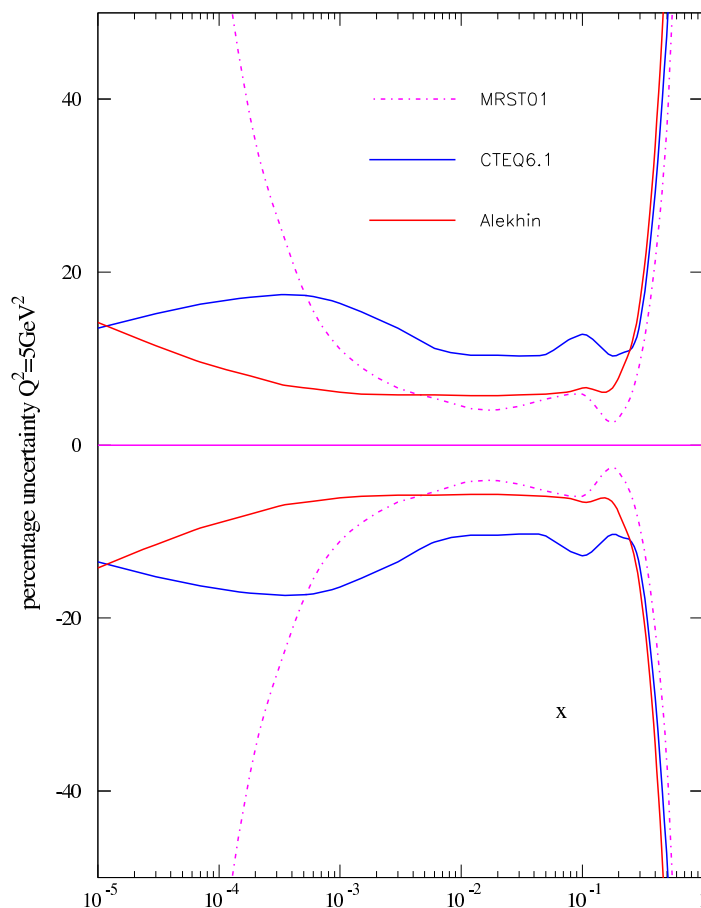


Fig. 19. The uncertainties on the gluon density at $Q^2 = 5 \text{ GeV}^2$ from the fits of MRST, CTEQ and Alekhin (from [26]).

on the gluon density obtained from several fits. The big differences in the shape of these error bands were traced to be due to the parameterization choice: the MRST group chose to allow the gluon density to become negative at low x , while in the CTEQ parameterization, the gluon has a valence-like shape at the starting scale, hence has a very small absolute error, all the uncertainty at higher Q^2 being due to the evolution driven by the higher x gluon, which is well determined.

4. The future: LHC — and beyond

In the near future, the final analyses of the full HERA datasets will reduce further the current uncertainties on the proton structure. The “combined H1 and ZEUS dataset”, resulting from a model-independent averaging of the cross-section measurements, is also very promising [27]. And of course, the LHC proton–proton collider should enter in operation in 2008. With a nominal centre-of-mass energy $\sqrt{S} = 14$ TeV and a luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$, it presents a huge potential for discovering physics beyond the Standard Model.

The current uncertainty on LHC W and Z production due to parton density functions is of $\sim 5\text{--}8\%$, *i.e.* the production of electroweak bosons is not precise enough yet, to be used as a luminosity monitor. Our current knowledge of the pdfs would also limit the precision with which the LHC experiments could measure the W mass. Indeed, at central rapidity, W production at LHC involves quarks with $x = M_W/\sqrt{S} \sim 7 \times 10^{-3}$, which is far away from the “valence region” where quarks are rather well known. However, data of Z production at LHC should allow the pdf uncertainties of W production to be reduced significantly. Measurements of asymmetries in W production will also bring, for the first time, constraints on the valence quarks at low x .

As shown in Fig. 20 for the example of $gg \rightarrow H$, the cross-sections suffer at most a $\sim 10\%$ uncertainty because of pdfs. Hence, the Higgs discovery potential is not much affected by pdf uncertainties. Furthermore, as studied recently in [29], these pdf uncertainties should be further reduced via the correlation of the Higgs production cross-section with Z or $t\bar{t}$ cross-section, depending on the Higgs mass. Similarly, it is unlikely that pdf uncertainties would jeopardize the discovery of a new heavy Z' or W' boson. However, if such a new boson is observed with a mass of 5–6 TeV, the uncertainty on the “partonic luminosity” (currently $\sim 40\%$) would limit the precision with which one would measure the couplings of this new boson. Examples of cases where the uncertainties of the gluon density at high x would reduce considerably the discovery potential for new physics were given in [30], where an extra-dimensional model was used, which would predict deviations in the

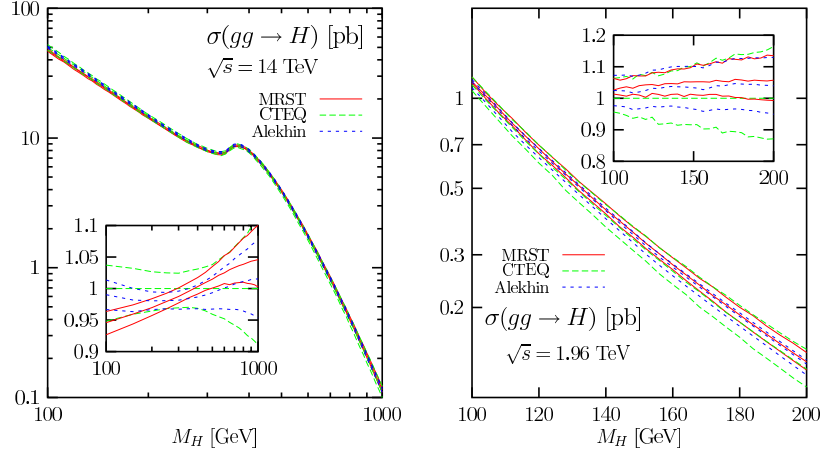


Fig. 20. The CTEQ, MRST and Alekhin pdf uncertainty bands for the NLO cross-section for $gg \rightarrow H$ at the LHC (left) and at the Tevatron (right). The insets show the spread of the predictions, when the cross-sections are normalized to CTEQ6M taken as reference [28].

dijet mass spectrum at high masses. However, the impact of experimental systematic errors, especially that of the jet energy scale, was not taken into account in this analysis. At high masses, the uncertainty due to the energy scale may be even larger than that due to the gluon density, such that the reduction of the discovery potential may not be as dramatic as given in [30]. Conversely, jet data at the LHC may not reduce significantly the uncertainty of the gluon density at high x . For jets with 1 TeV (2 TeV) of transverse momentum, in the rapidity range $1 < \eta < 2$, the pdf uncertainty on the inclusive jet cross-section is of about 15% (25%). In the same kinematic domain, the experimental uncertainty is of 30% (50%) if the jet energy scale is understood within 5%. Fits performed using LHC jet pseudo-data indicate that, to decrease significantly the uncertainty of the gluon density at high x , one should control the jet energy scale at the percent level, which will be quite challenging [31].

Further progress in our understanding of proton structure may also come from the theory side. For example, a better understanding of prompt photon production may allow the corresponding data to be (re-)included in global fits, bringing constraints on the gluon density ($qg \rightarrow q\gamma$). Existing data on the diffractive production of vector mesons, like the J/ψ , could also constrain the low x gluon density in principle: naively, since a two-gluon system is exchanged in the diffractive regime, the cross-section should present a quadratic dependence on the gluon density. This simple picture is however not valid beyond the leading order, and the precise calculation of diffractive

J/ψ production requires “skewed” pdfs to be introduced. Predictions do exist for “next-to-leading order”, but they are not yet as “solid” as DGLAP predictions. Once this is better understood, the inclusion of HERA diffractive J/ψ photoproduction data should bring significant constraints on the gluon density in the range of $x \sim 10^{-4}$ – 10^{-3} .

Finally, on the experimental side, there are several projects, not yet approved, which would bring new information on the nucleon structure. Let us mention:

- The upgrade of the accelerator complex at JLab [32]: fixed-target DIS with a 12 GeV beam, at a large luminosity of $10^{38} \text{ cm}^{-2}\text{s}^{-1}$ would allow x values as large as 0.8 to be accessed, and the ratio d/u to be much better constrained at high x .
- The future Electron Ion Collider [33], where polarized electrons would collide with heavy ions or polarized protons, has a very rich programme on low x physics, QCD at high densities, hard diffraction, polarized pdfs, *etc.*
- The LHeC project [34]: the LHC proton (or ion) beam would be brought in collision with a lepton beam. In a “ring–ring” design, the lepton beam energy would be limited to 70 GeV leading to $\sqrt{S} = 1.4 \text{ TeV}$, but the ep luminosity could reach $10^{33} \text{ cm}^{-2}\text{s}^{-1}$. A larger $\sqrt{S}$ of $\sim 2 \text{ TeV}$ could be achieved in a “linac–ring” design, at the price of a reduced luminosity.

I wish to thank Roberto Fiore and Christophe Royon for organizing this very interesting school and for giving me the opportunity to take part in it. I am very grateful to W.K. Tung and R. Thorne for their help with the preparation of this talk.

REFERENCES

- [1] F. Halzen, A. Martin, *Quarks and Leptons*, Wiley, New York 1984; W.K. Tung, Perturbative QCD and the Parton Structure of the Nucleon, in: *Handbook of QCD*, vol. 2, 887, ed. M. Shifman, World Scientific, 2001; A.M. Cooper, R. Devenish, A. DeRoeck, *Int. J. Mod. Phys. A* **13**, 3385 (1998) [[hep-ph/9712301](#)].
- [2] A. Martin, *Acta Phys. Pol. B* **39**, 2025 (2008), these proceedings [[arXiv:0802.0161](#) [[hep-ph](#)]].
- [3] A.M. Cooper-Sarkar, R. Devenish, *Acta Phys. Pol. B* **34**, 2911 (2003) [[hep-ph/0305165](#)].

- [4] A. De Rujula *et al.*, *Phys. Rev.* **D10**, 1649 (1974).
- [5] H1 Collab., *Eur. Phys. J.* **C30**, 1 (2003).
- [6] ZEUS Collab., *Phys. Rev.* **D70**, 052001 (2004) [[hep-ex/0401003](#)].
- [7] ZEUS Collab., *Eur. Phys. J.* **C32**, 1 (2003) [[hep-ex/0307043](#)].
- [8] H1 Collab., *Phys. Lett.* **B634**, 173 (2006) [[hep-ex/0512060](#)]; ZEUS Collab., *Phys. Lett.* **B637**, 210 (2006) [[hep-ex/0602026](#)].
- [9] H1 Collab., *Eur. Phys. J.* **C45**, 23 (2006) [[hep-ex/0507081](#)].
- [10] H1 Collab., *Phys. Lett.* **B598**, 159 (2004) [[hep-ex/0406029](#)]; ZEUS Collab., *Phys. Lett.* **B487**, 53 (2000) [[hep-ex/0005018](#)].
- [11] W.K. Tung *et al.*, *J. High Energy Phys.* **0702**, 053 (2007).
- [12] ZEUS Collab., *Phys. Rev.* **D67**, 012007 (2002).
- [13] ZEUS Collab., *Eur. Phys. J.* **C42**, 1 (2005).
- [14] R.S. Towell *et al.* [E866 Collab.], *Phys. Rev.* **D64**, 052002 (2001).
- [15] R.S. Thorne *et al.*, [arXiv:0706.0456](#) [[hep-ph](#)].
- [16] H.L. Lai *et al.*, *J. High Energy Phys.* **0704**, 089 (2007).
- [17] J.F. Owens *et al.*, *Phys. Rev.* **D75**, 054030 (2007).
- [18] J. Pumplin *et al.*, *Phys. Rev.* **D75**, 054029 (2007).
- [19] A.D. Martin *et al.*, *Phys. Lett.* **B652**, 292 (2007).
- [20] R. Thorne, C. White, *Phys. Rev.* **D75**, 034005 (2007).
- [21] C. Kiesling, *Acta Phys. Pol. B* **39**, 2125 (2008), these proceedings.
- [22] A. Nikiforov [H1 Collab.], talk presented at the “43rd Rencontres de Moriond on QCD and High Energy Interactions”, March 2008.
- [23] H1 Collab., *Phys. Lett.* **B520**, 183 (2001) [[hep-ex/0108035](#)].
- [24] F. Gelis, *Acta Phys. Pol. B* **39**, 2419 (2008), these proceedings.
- [25] J. Pumplin *et al.*, *J. High Energy Phys.* **07**, 012 (2007); *Phys. Rev.* **D65**, 014013, 014011 (2001).
- [26] R. Thorne, invited talk at PHYSTAT-LHC Workshop on Statistical Issues for LHC Physics, June 2007, [[arXiv:0711.2986](#) [[hep-ph](#)]].
- [27] A. Glazov *et al.*, Proceedings of the HERA-LHC Workshop, [hep-ph/0601012](#), [hep-ph/0601013](#).
- [28] A. Djouadi, S. Ferrag, *Phys. Lett.* **B586**, 345 (2004).
- [29] P.M. Nadolsky *et al.*, [arXiv:0802.0007](#) [[hep-ph](#)].
- [30] S. Ferrag *et al.* [ATLAS Coll.], [hep-ph/0407303](#).
- [31] A.M. Cooper-Sarkar, talk presented at the “15th International Workshop on Deep-Inelastic Scattering and Related Subjects” (DIS2007), [arXiv:0707.1593](#) [[hep-ph](#)].
- [32] “The 12 GeV upgrade of CEBAF”, January 2005, <http://www.jlab.org/12GeV/>
- [33] C. Aidala *et al.*, EIC White Paper prepared for NSAC 2007, <http://www.bnl.gov/eic>
- [34] J.B. Dainton *et al.*, *JINST* **1**, P10001 (2006), [[hep-ex/0603016](#)].