Z. Phys. C – Particles and Fields 44, 291–306 (1989)

Zeitschrift für Physik C Particles and Fields

Quark momentum distributions from *ep* collisions: strategies and prospects

G. Ingelman and R. Rückl*
Deutsches Electronen-Synchrotron DESY, Notkestrasse 85, D-2000 Hamburg 52, Federal Republic of Germany

Received 6 March 1989

Abstract. We present the results of a comprehensive investigation of possibilities to extract structure functions and momentum distributions of individual quark flavours from cross-section measurements at *ep* colliders. Approximate determinations based on single cross-sections are examined and compared with an exact unfolding procedure which, however, requires the combination of neutral and charged current data for both electron and positron beams. Using Monte Carlo methods we evaluate the precision that can be expected from the forthcoming HERA experiments in measuring the shape in Bjorken-x of structure functions and quark distributions.

1 Introduction

Deep-inelastic lepton–nucleon scattering is the most direct way of exploring the substructure of hadronic matter. With the new electron–proton collider HERA one will possess a powerful tool to penetrate to distances as small as 10^{-3} fm. Moreover, it will become possible to examine structure functions at very small values of the Bjorken-variable x in the range 10^{-4} to 10^{-2}. This promises exciting opportunities to improve and extend the present knowledge of deep-inelastic physics [1], to put the standard model to further tests, and to trace a possible substructure of quarks and leptons. The outcome of such studies will also have implications on the phenomenology of future hadron colliders by checking the necessary extrapolations from present-day energies into the multi-TeV range.

From the experience with fixed target experiments, extraction of nucleon structure functions and determination of quark distributions from differential cross-sections measurable at *ep* colliders are not expected to be easy experimental tasks. Previous phenomenological studies [2–6] have elucidated the main difficulties and emphasized the specific problems of collider experiments. Motivated by the importance of the subject we continue and extend these investigations in the present paper. It is systematically and comprehensively examined which distribution functions can be obtained from *ep* collisions and the achievable precision is estimated for each case. We consider unfolding methods, which require the combination of different cross-sections, as well as approximate methods that hold in limited kinematical regions but, on the other hand, only use a single cross-section. The advantages and disadvantages of these methods are discussed. Which procedure is more suitable depends on the data available and on the aim of the study, such as measuring the momentum distribution of a particular constituent quark or testing scaling violations predicted by QCD. Here, we concentrate on the determination of the shape in Bjorken-x, and present quantitative results which illustrate the potential of neutral current (NC) and charged current (CC) measurements at HERA.

As usual, the kinematics of inclusive *ep* scattering is characterized by the *ep* centre-of-mass energy $\sqrt{s}$, $s=(p_e+P)^2$, the squared momentum transfer $Q^2 \equiv -q^2 = -(p_e-p_l)^2$, the Bjorken-variable $x=Q^2/2P\cdot q$ and the variable $y \equiv P\cdot q/P\cdot p_e = Q^2/xs$. In these definitions p_e, p_l and P denote the four-momenta of the incoming and scattered lepton and the incoming proton, respectively. The statistics of realistic event samples is simulated by Monte Carlo techniques, while detector effects are mainly taken into account by cuts in x, y and Q^2 within which systematic effects on the cross-sections are expected to be less than 10% [7]. In order to exploit the available statistics, we average the resulting x-distributions of structure functions and quark densities over y. Because of the Q^2 evolution in QCD, x-distribution averaged over the y (or Q^2) range of HERA experiments are expected to differ in shape

* Present address: Sektion Physik, University of Munich, Theresienstr. 37, D-8000 München 2, FRG

from the corresponding distributions at lower values of Q^2. We indicate the size of these effects and point out the cases in which a comparison of HERA results with data from present-day fixed target experiments should reveal these changes. Prospects of more stringent quantitative QCD tests including a measurement of $\Lambda_{\overline{\rm MS}}$ are discussed in [4, 8].

The paper is organized as follows. In Sect. 2 we set up the theoretical framework and provide the necessary formalism. Section 3 examines different strategies to extract structure functions and quark distributions from *ep* cross-sections. Important details of the Monte Carlo event simulation and numerical illustrations of structure function measurements at HERA are presented in Sect. 4. Finally, in Sect. 5, we make some concluding remarks.

2 Theoretical cross-sections

For our analysis we adopt the parton model expressions for the inclusive NC and CC cross-sections obtained in lowest order of the electroweak couplings and leading order QCD. Taking $Q^2 \geqq 100\,\mathrm{GeV}^2$, one can safely neglect the proton mass and primordial transverse momenta of the partons as well as higher twist operators. Contributions of this kind are suppressed by inverse powers of Q^2. We also do not have to worry about heavy quark thresholds. It has been shown that, at $Q^2 \geqq 100\,\mathrm{GeV}^2$ and $x \geqq 0.01$, contributions from $c\bar{c}$ production have a leading-logarithmic behaviour similar to that of light quarks, while contaminations from $b\bar{b}$, $t\bar{t}$ and presumably also from $t\bar{b}$ production are negligible [9]. Therefore, it is perfectly appropriate for the present study to assume four massless quark flavours (u, d, s, c).

Writing the differential cross-sections for unpolarized NC scattering, $ep \to eX$, in the general form

$$\frac{d\sigma_{\rm NC}(e^{\mp})}{dx\,dy} = \frac{4\pi\alpha^2 s}{Q^4}\left[y^2 x F_1(x,Q^2) + (1-y)F_2(x,Q^2) \pm \left(y - \frac{y^2}{2}\right) x F_3(x,Q^2)\right] \tag{1}$$

where α is the electromagnetic finestructure constant, and using the above approximations one finds the following relations of the NC structure functions $F_i(x,Q^2)$ to the quark and antiquark density distributions $q_f(x,Q^2)$ and $\bar{q}_f(x,Q^2)$:

$$\begin{aligned} F_2(x,Q^2) &= 2xF_1(x,Q^2) \\ &= \sum_{f=u,d,s,c} A_f(Q^2)[xq_f(x,Q^2) + x\bar{q}_f(x,Q^2)] \\ xF_3(x,Q^2) &= \sum_{f=u,d,s,c} B_f(Q^2)[xq_f(x,Q^2) - x\bar{q}_f(x,Q^2)]. \end{aligned} \tag{2}$$

The flavour-dependent coefficients A_f and B_f are given by

$$\begin{aligned} A_f(Q^2) &= e_f^2 - 2e_f v_e v_f P_Z + (v_e^2 + a_e^2)(v_f^2 + a_f^2)P_Z^2 \\ B_f(Q^2) &= -2e_f a_e a_f P_Z + 4 v_e v_f a_e a_f P_Z^2 \end{aligned} \tag{3}$$

where e_f is the electric charge of a fermion f, while $v_f = (T_{3f} - 2e_f \sin^2\theta_W)/\sin 2\theta_W$ and $a_f = T_{3f}/\sin 2\theta_W$ are the vector and axial vector couplings to the Z^0 boson expressed in terms of the third component of the weak isospin T_{3f} and the Weinberg angle θ_W. The convention is such that $e_u = \frac{2}{3}$, $T_{3e} = -\frac{1}{2}$, $T_{3u} = \frac{1}{2}$, etc. Moreover, by the appearance in (3) of the factors $P_Z = Q^2/(Q^2 + m_Z^2)$ denoting the ratio of the Z and γ propagators, one can easily recognize the individual contributions from γ exchange, Z exchange and $\gamma - Z$ interference. In a similar notation, the differential cross-sections for unpolarized CC scattering, $ep \to \nu X$, take the form

$$\begin{aligned} \frac{d\sigma_{\rm CC}(e^-)}{dx\,dy} &= \frac{\pi\alpha^2 s}{4\sin^4\theta_W (Q^2 + m_W^2)^2} \cdot \left[\sum_{f=u,c} xq_f(x,Q^2) + (1-y)^2 \sum_{f=d,s} x\bar{q}_f(x,Q^2)\right] \\ \frac{d\sigma_{\rm CC}(e^+)}{dx\,dy} &= \frac{\pi\alpha^2 s}{4\sin^4\theta_W (Q^2 + m_W^2)^2} \cdot \left[\sum_{f=u,c} x\bar{q}_f(x,Q^2) + (1-y)^2 \sum_{f=d,s} xq_f(x,Q^2)\right]. \end{aligned} \tag{4}$$

To the order considered here, one can replace the W and Z couplings in the above formulas by the Fermi constant G_F, using the relations $G_F = \pi\alpha/[\sqrt{2}\sin^2\theta_W m_W^2]$ and $m_Z = m_W/\cos\theta_W$. Furthermore, the quark distribution functions introduced above obey the Altarelli–Parisi evolution equations of QCD in leading-logarithmic approximation [10].

Concerning higher order effects, we want to add the following remarks. As $Q^2 \geqq 100\,\mathrm{GeV}^2$, the neglect of non-leading QCD corrections [10] can be justified to some extent by asymptotic freedom. Moreover, for studies of x-distributions averaged over y or Q^2, next-to-leading corrections to the evolution equations should certainly be less relevant than for a detailed analysis of scaling violations. This anticipation is confirmed by the numerical results shown later. Regarding $\alpha_s(Q^2)$ corrections to the parton model formulas, we have studied the influence of the longitudinal structure function

$$F_L \equiv F_2 - 2xF_1 \tag{5}$$

on the NC cross-sections in the absence of contribu-

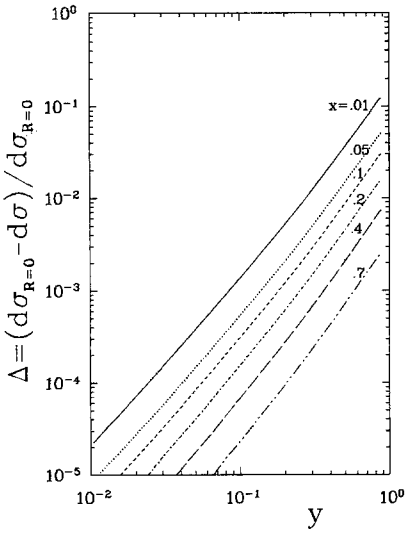

Fig. 1. Relative shift, (8), of the differential ep cross-section $d\sigma_\gamma/dx\,dy$ due to the $\mathcal{O}(\alpha_s)$ longitudinal structure function F_L predicted in QCD

tions from the Z boson. In that case, (1) reduces to

$$\frac{d\sigma_\gamma}{dxdy} = \frac{4\pi\alpha^2 s}{Q^4} F_2^{\mathrm{em}}(x,Q^2)\left[1 - y + \frac{y^2}{2}\frac{1}{1+R}\right] \tag{6}$$

with

$$F_2^{\mathrm{em}}(x,Q^2) = \sum_{f=u,d,s,c} e_f^2\,[xq_f(x,Q^2) + x\bar{q}_f(x,Q^2)] \tag{7}$$

and $R = F_L/(F_2 - F_L)$. Evaluating R from F_L to order α_s as described in [11] we have calculated the relative shift

$$\Delta = \frac{\left.\frac{d\sigma_\gamma}{dxdy}\right|_{R=0} - \frac{d\sigma_\gamma}{dxdy}}{\left.\frac{d\sigma_\gamma}{dxdy}\right|_{R=0}} = \frac{y^2}{1+(1-y)^2}\frac{R}{1+R} \tag{8}$$

to the above cross-section and display in Fig. 1 the y-dependence of Δ at fixed values of x. Although the effect of $R \neq 0$ reaches a few percent at large y, and even exceeds 10% at $y \geqq 0.8$ and $x \leqq 0.01$, the resulting structure functions and quark densities are shifted insignificantly when averaged over suitable ranges in y as done in the later applications. Thus, as far as QCD is concerned the leading-logarithmic approximation is certainly sufficient for our purposes.

In contrast, higher order electroweak corrections, particularly the electromagnetic radiative effects, are known to be important and must carefully be taken into account. Complete one-loop calculations have been performed by different groups and, although some differences are still being discussed, there is a general consensus on the results [12]. Notwithstanding the difficulties in implementing these corrections in the experimental analysis, we take the attitude that the problem can be solved, and assume that our lowest order analysis is applied to the data after correction for the electroweak radiative effects.

Returning to the master formulas (1–4), we note that due to the rising strength of the weak neutral current interactions with increasing Q^2 a new structure function, xF_3, appears in the cross-sections for $ep \to eX$, and the F_2 structure function deviates from the familiar expression, (7), valid at present-day energies. Moreover, the Z boson propagator induces scaling violations by powers of Q^2/m_Z^2 which cannot simply be factored out, in contrast to the W propagator which enters the CC cross-section as an overall factor. In the upper Q^2 region accessible at HERA energies, these scale dependences become much stronger than the logarithmic scaling violations predicted by QCD, and therefore complicate QCD tests as discussed in [4, 8]. Also the shape in x of structure functions averaged over y is affected. As a consequence, one must be quite careful when comparing measurements from future ep collider experiments with existing (low-Q^2) data. Finally, it is rather clear from (1–4) that the separation of different quark flavours and of valence and sea components is generally not straightforward, but requires more involved approaches. The new features appearing in high energy ep collisions have been one of the main motivations for the work reported here.

3 Strategies for extracting quark distributions

Having discussed the NC and CC cross-sections provided by theory, we shall now investigate the problem of extracting individual quark distributions or simple combinations thereof from these basic observables. The aim is to systematically clarify which distributions can in principle be obtained from measurements with electron and positron beams at fixed ep c.m. energy. There are essentially two different procedures [4, 6]. One can start from a single cross-section and consider restricted kinematical regions where certain structure functions or quark distributions dominate. This method is relatively straightforward, but rather selective in the distributions which can be obtained. Moreover, it only yields approximate results. As a first sophistication, one can apply this procedure to sums and differences of e^-p and e^+p cross-sections and, in this way, obtain additional combinations of quark distributions. More ambitiously, one can try to strictly unfold a desired quark distribution or structure function by using the NC *and* CC cross-sections for e^-p *and* e^+p scattering. Although the gain in flexibility is considerable, the necessity to combine four different cross-sections makes this procedure subject to larger statistical uncertainties and relative normalization errors. It is therefore not a priori clear whether such a complete unfolding procedure really gives better results than simpler approximate extraction methods.

In the following, we derive and examine approxi-

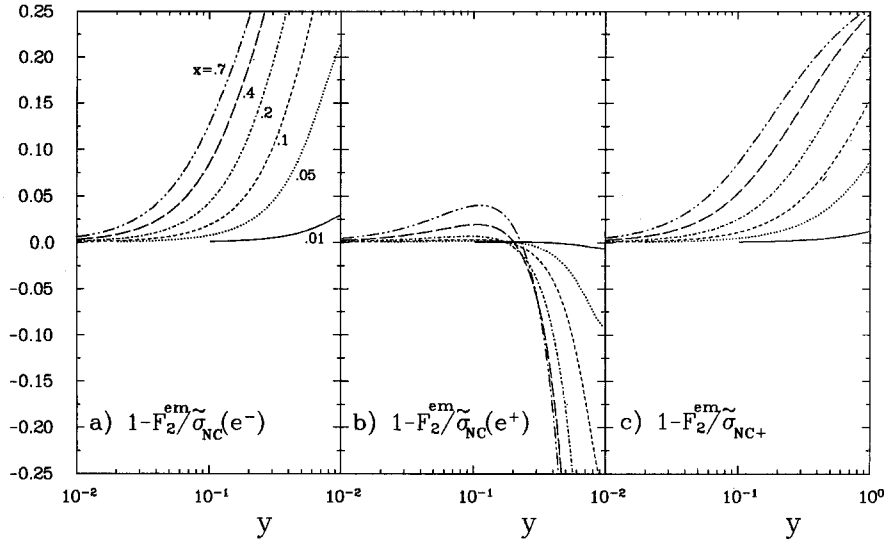

Fig. 2a–c. Deviation of the rescaled differential NC cross-sections, (9) and (11), from the electromagnetic structure function $F_2^{\rm em}$. The various curves in **b and c** correspond to the same values of x as indicated in **a**. Because of the overall cut $Q^2 \geq 100\,\mathrm{GeV}^2$, $y \geq 0.1$ (0.02) for $x = 0.01$ (0.05)

mate as well exact relations for distribution functions of single quark flavours and various combinations of flavours in terms of rescaled cross-sections. The results of this formal study will be used in the Monte Carlo simulation of a real data analysis discussed in the next section. For numerical purposes, we take the parametrization EHLQ1 [13] for the quark densities, and the values $\alpha = 1/137$, $\sin^2\theta_W = 0.226$, $m_W = 81.4\,\mathrm{GeV}$ and $m_Z = 92.5\,\mathrm{GeV}$ for the electroweak parameters. The ep c.m. energy is assumed to be $\sqrt{s} = 314\,\mathrm{GeV}$.

3.1 *Approximate methods*

3.1.1 Use of NC cross-sections. It is convenient to factor out the dominant Q^2 dependence from the differential NC cross-sections given in (1) and to define

$$\tilde{\sigma}_{\rm NC}(e^{\mp}) \equiv \frac{Q^4}{2\pi\alpha^2 s Y_+}\frac{d\sigma_{\rm NC}(e^{\mp})}{dx\,dy} = F_2(x,Q^2) \pm \frac{Y_-}{Y_+} xF_3(x,Q^2) \tag{9}$$

where $Y_\pm \equiv 1 \pm (1-y)^2$. For sufficiently small values of Q^2, Z-exchange can be neglected relative to γ-exchange and consequently xF_3 vanishes, while F_2 approaches the electromagnetic structure function $F_2^{\rm em}$ as can be seen from (2, 3, 7). Figure 2a, b show the departure of $\tilde{\sigma}_{\rm NC}(e^{\mp})$ from $F_2^{\rm em}$ at fixed x as a function of y. We see that $\tilde{\sigma}_{\rm NC}(e^+)$ is better suited for a determination of $F_2^{\rm em}$ than $\tilde{\sigma}_{\rm NC}(e^-)$ because of a partial cancellation between the Z contributions to F_2 and the xF_3 term in (1) as already noted in [4]. Quantitatively, for $x \leq 0.7$ and $y \leq 0.3$, the approximation

$$F_2^{\rm em} = \sum_f e_f^2 (xq_f + x\bar{q}_f) \approx \tilde{\sigma}_{\rm NC}(e^+) \tag{10}$$

is accurate to better than 5%.

By combining e^-p and e^+p cross-sections one can separate the structure functions F_2 and xF_3 given in (2):

$$\tilde{\sigma}_{\rm NC+} \equiv \tfrac{1}{2}(\tilde{\sigma}_{\rm NC}(e^-) + \tilde{\sigma}_{\rm NC}(e^+)) = F_2 \tag{11}$$

$$\tilde{\sigma}_{\rm NC-} = \frac{Y_+}{2Y_-}(\tilde{\sigma}_{\rm NC}(e^-) - \tilde{\sigma}_{\rm NC}(e^+)) = xF_3. \tag{12}$$

Such a separation is useful for studies of electroweak issues, in particular, since xF_3 is a direct probe of $\gamma - Z$ interference. Here, the interesting question is whether or not one can gain further information on quark distributions from the above combinations. This is not really the case for the sum $\tilde{\sigma}_{\rm NC+}$ which still involves a superposition of quark densities with coefficients containing electroweak couplings and propagators. Although $\tilde{\sigma}_{\rm NC+}$ is a better approximation of $F_2^{\rm em}$ than $\tilde{\sigma}_{\rm NC}(e^-)$, the best opportunity to determine $F_2^{\rm em}$ is provided by positron–proton collisions. This fact can be clearly seen by comparing Fig. 2c with Fig. 2a, b. On the other hand, the difference of the e^-p and e^+p cross-sections or, equivalently, xF_3 can serve as an interesting and independent probe of the valence quark structure of the proton. Using (3) and dropping the quadratic Z contributions proportional to P_Z^2 which are negligible in the kinematical region under consideration, one finds

$$x(e_u u_v - e_d d_v) \approx \frac{2\sin^2 2\theta_W}{P_Z}\tilde{\sigma}_{\rm NC-} \tag{13}$$

where $u_v = u - \bar{u}$ and $d_v = d - \bar{d}$ denote the valence up and down quark distributions, respectively. The quality of this approximation is better than 1% up to $Q^2 = 10^3\,\mathrm{GeV}^2$ and then worsens to about 4% at the HERA kinematic limit in Q^2. Apart from $e_u u_v - e_d d_v$ and $F_2^{\rm em}$ there are no other, comparatively simple structure functions which can be obtained from NC cross-sections alone.

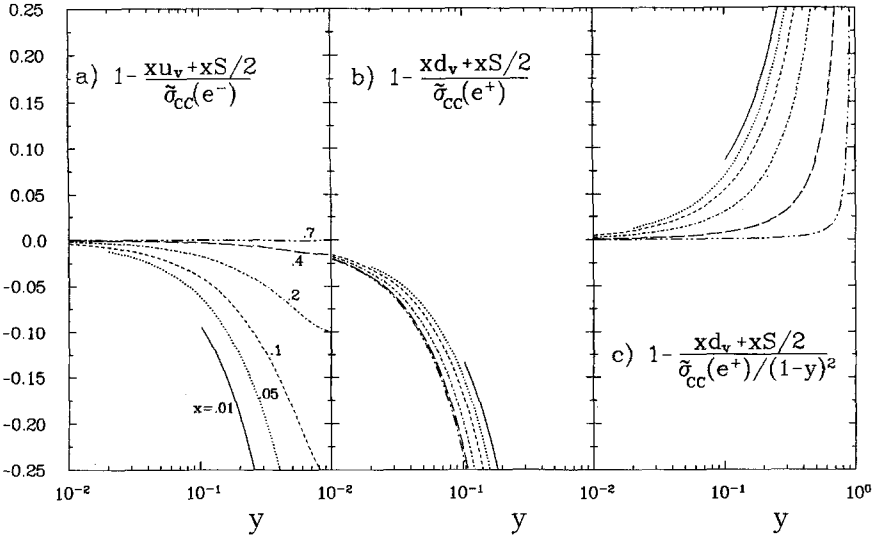

Fig. 3a–c. Deviation of rescaled CC cross-sections, (14), from the quark momentum distributions $x(u_v + S/2)$ **a** and $x(d_v + S/2)$ **b** and **c**. The explanations on the curves given in Fig. 2 apply here correspondingly

3.1.2 Use of CC cross-sections. Similarly as in the NC case we divide the CC cross-sections given in (4) by an overall factor and define the quantities

$$\tilde{\sigma}_{CC}(e^-) \equiv \frac{4\sin^4\theta_W(Q^2+m_W^2)^2}{\pi\alpha^2 s}\frac{d\sigma_{CC}(e^-)}{dxdy}$$
$$= x(u + c + (1-y)^2(\bar{d}+\bar{s}))$$
$$\tilde{\sigma}_{CC}(e^+) \equiv \frac{4\sin^4\theta_W(Q^2+m_W^2)^2}{\pi\alpha^2 s}\frac{d\sigma_{CC}(e^+)}{dxdy}$$
$$= x(\bar{u} + \bar{c} + (1-y)^2(d+s)). \quad (14)$$

Here, an obvious and more convenient notation has been used for the quark densities $q_f(x, Q^2)$. Evidently, at sufficiently large x one can neglect the sea quarks and make use of the valence quark approximation to (14). Then,

$$xu_v \approx \tilde{\sigma}_{CC}(e^-) \quad (15)$$

$$xd_v \approx \frac{\tilde{\sigma}_{CC}(e^+)}{(1-y)^2}. \quad (16)$$

Another interesting and feasible limit is $y \to 0$. In this case, (14) approximately yields

$$x(u_v + \tfrac{1}{2}S) \approx \tilde{\sigma}_{CC}(e^-) \quad (17)$$

$$x(d_v + \tfrac{1}{2}S) \approx \tilde{\sigma}_{CC}(e^+) \quad (18)$$

where S denotes the total quark sea defined as

$$S \equiv \sum_f (q_f + \bar{q}_f) - u_v - d_v. \quad (19)$$

We assume S to be symmetric in quarks and antiquarks of a given flavour. The deviations of the above approximations from the exact relations (14) are displayed in Fig. 3a, b. While (17) is applicable in a sufficiently large region of y such that a reasonable determination of $u_v + S/2$ can be expected, this procedure seems to be of little use for $d_v + S/2$. The reason is easy to understand. In the former case, only sea contributions are neglected, whereas in the latter case the deviations also come from a valence quark. More precisely, in (17, 18) we have disregarded the terms $-Y_-(\bar{d}+\bar{s})$ and $-Y_-(d+s)$, respectively. The problem can be partly cured by taking instead of (18) the approximation

$$x(d_v + \tfrac{1}{2}S) \approx \frac{\tilde{\sigma}_{CC}(e^+)}{(1-y)^2} \quad (20)$$

such that only the sea quark contribution $Y_-/(1-y)^2$ $(\bar{u}+\bar{c})$ is neglected when going from (14) to (20). The validity of this choice is illustrated in Fig. 3c. We find that for $x \geqq 0.05$ and $y \leqq 0.15$ the left- and right-hand sides of (17, 20) deviate from each other by less than about 10%.

Furthermore, from the sum and difference of the CC cross-sections for e^-p and e^+p scattering one gets

$$\tilde{\sigma}_{CC+} \equiv \tilde{\sigma}_{CC}(e^-) + \tilde{\sigma}_{CC}(e^+) = x(U + (1-y)^2 D) \quad (21)$$

$$\tilde{\sigma}_{CC-} \equiv \tilde{\sigma}_{CC}(e^-) - \tilde{\sigma}_{CC}(e^+) = x(u_v - (1-y)^2 d_v) \quad (22)$$

with

$$U = u + \bar{u} + c + \bar{c}$$
$$D \equiv d + \bar{d} + s + \bar{s}. \quad (23)$$

Equation (21) suggests to study the possibility of extracting the structure function

$$F_s \equiv xU + xD = \sum_f x(q_f + \bar{q}_f) \quad (24)$$

which, being a total singlet under flavour transformations, plays an important role in QCD studies [10]. Similarly as discussed above, one can again consider different approximations of F_s in terms of $\tilde{\sigma}_{CC+}$ or, more generally, $\tilde{\sigma}_{CC}(e^{\mp})$ which all coincide with F_s as

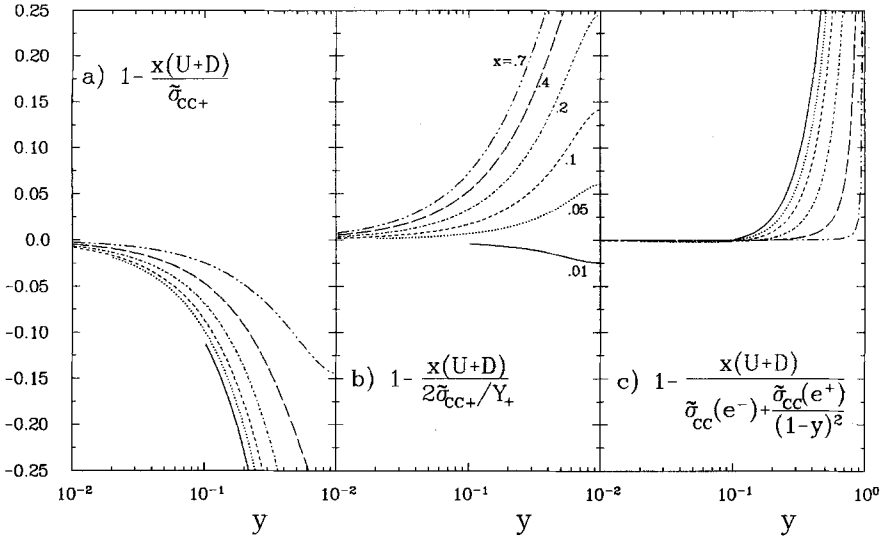

Fig. 4a–c. Deviation of the combinations, (25), of CC cross-sections from the singlet structure function $F_s = x(U+D)$. The explanations on the curves given in Fig. 2 apply here correspondingly

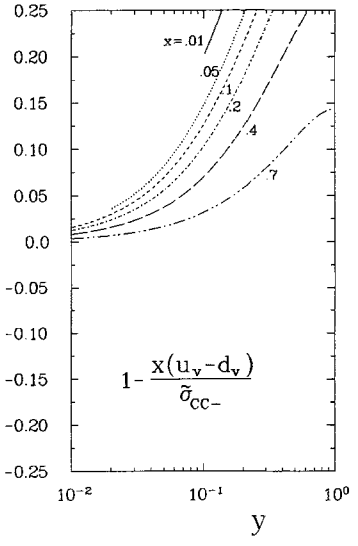

Fig. 5. Deviation of the difference, (22), of the CC electron and positron cross-sections from the quark momentum distribution $x(u_v - d_v)$

$y \to 0$. The three most obvious choices are:

$$F_s \approx \begin{cases} \tilde{\sigma}_{CC+} \\ \dfrac{2}{Y_+}\tilde{\sigma}_{CC+} \\ \tilde{\sigma}_{CC}(e^-) + \tilde{\sigma}_{CC}(e^+)/(1-y)^2. \end{cases} \quad (25)$$

Here, the approximations involve the neglect of the terms $-Y_-D$, $+Y_-/Y_+(U-D)$ and $-Y_-(\bar{d}+\bar{s}) + Y_-/(1-y)^2(\bar{u}+\bar{c})$, respectively. The important point is that these approximations break down more or less rapidly as y increases and one should, of course, use the one with the largest region of validity. Figure 4 illustrates the accuracy of the approximations, (25). As expected from the previous examination of the relations (17) and (18) the most promising way to determine F_s from CC measurements is via the last relation in (25). In this case, F_s is reproduced perfectly for $y \leqq 0.1$, and still with an accuracy of better than 7% for $y \leqq 0.3$.

Finally, also the difference of the CC cross-sections, (22), is rather interesting since it involves only valence quarks. Unfortunately, it is difficult to separate u_v from d_v in (22), at least at fixed ep c.m. energy. What one can probe is the difference of u_v and d_v. For that purpose, one has to restrict y to small enough values such that it becomes reasonable to take

$$x(u_v - d_v) \approx \tilde{\sigma}_{CC-}. \quad (26)$$

On the other hand, one cannot be too restrictive without paying in statistics. A feasible compromise is a cut at $y = 0.15$. For this choice, the approximation is better than 20% at $x \geqq 0.05$ as shown in Fig. 5, an accuracy which turns out to be sufficient compared with the expected statistical errors.

To conclude, the theoretical considerations presented in this section suggest that it should be possible to extract from the CC cross-sections alone acceptable approximations of the following structure functions: $u_v + S/2$, $d_v + S/2$, $u_v - d_v$ abd the singlet F_s. At large x, this correspond to measurements of the valence quark distributions u_v, d_v and $x(u_v \pm d_v)$.

3.2 Exact unfolding procedure

A more complete separation and determination of structure functions requires the use of all four NC and CC cross-sections measurable in $e^{\mp}p$ scattering [4, 5]. At least formally, one can solve the four equations (1) and (4) for four particular quark distributions. A convenient basis is provided by the valence quark densities u_v and d_v, and the structure functions U and D defined in (23). For this choice, the solutions read

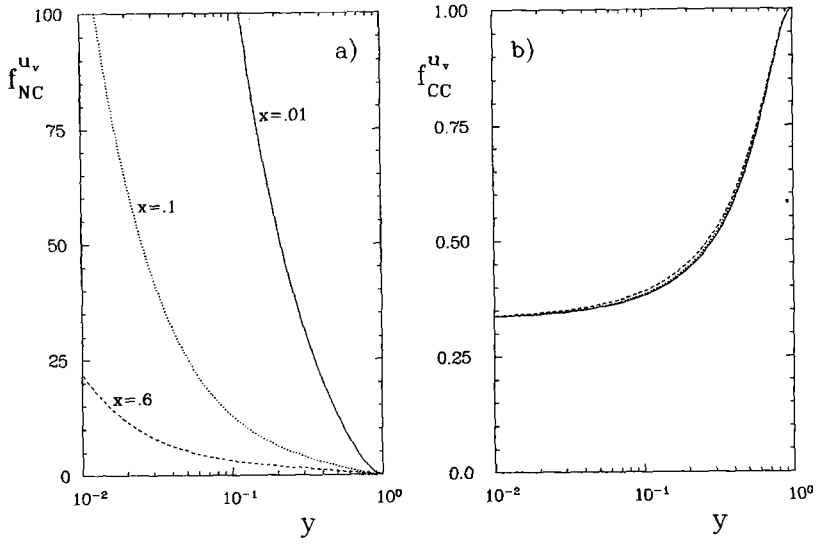

Fig. 6 a, b. Coefficient functions entering the unfolding relation (27) for the valence up-quark distribution, xu_v. The curves in **b** correspond to the same values of x as indicated in **a**

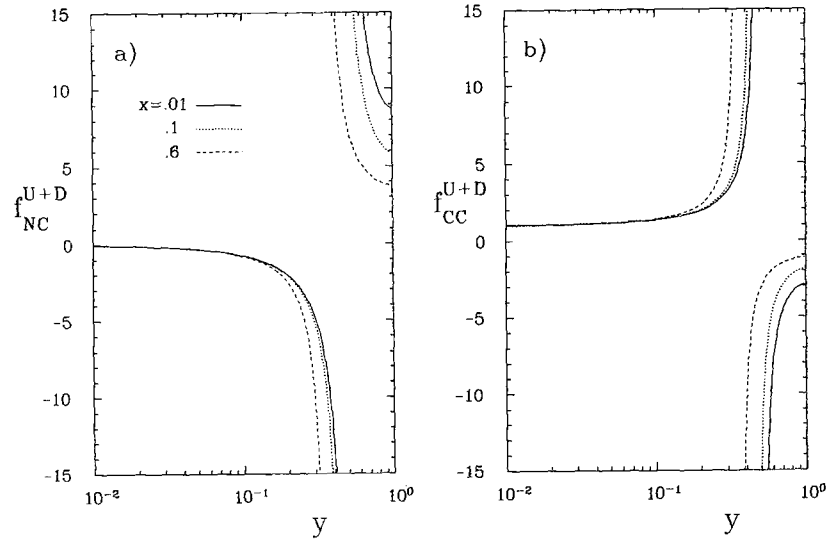

Fig. 7 a, b. Coefficient functions entering the unfolding relation (32) for the total singlet structure function, $F_s = x(U+D)$. The various curves correspond to fixed values of x as indicated in **a**

as follows:

$$u_v = f_{\mathrm{NC}}^{u_v}\tilde{\sigma}_{\mathrm{NC}-} + f_{\mathrm{CC}}^{u_v}\tilde{\sigma}_{\mathrm{CC}-} \tag{27}$$

$$f_{\mathrm{NC}}^{u_v} = \frac{(1-y)^2}{(1-y)^2 B_u + B_d};\quad f_{\mathrm{CC}}^{u_v} = \frac{B_d}{(1-y)^2 B_u + B_d}$$

$$d_v = f_{\mathrm{NC}}^{d_v}\tilde{\sigma}_{\mathrm{NC}-} + f_{\mathrm{CC}}^{d_v}\tilde{\sigma}_{\mathrm{CC}-} \tag{28}$$

$$f_{\mathrm{NC}}^{d_v} = \frac{1}{(1-y)^2 B_u + B_d};\quad f_{\mathrm{CC}}^{d_v} = \frac{-B_u}{(1-y)^2 B_u + B_d}$$

$$xU = f_{\mathrm{NC}}^{U}\tilde{\sigma}_{\mathrm{NC}+} + f_{\mathrm{CC}}^{U}\tilde{\sigma}_{\mathrm{CC}+} \tag{29}$$

$$f_{\mathrm{NC}}^{U} = \frac{(1-y)^2}{(1-y)^2 A_u - A_d};\quad f_{\mathrm{CC}}^{U} = \frac{-A_d}{(1-y)^2 A_u - A_d}$$

$$xD = f_{\mathrm{NC}}^{D}\tilde{\sigma}_{\mathrm{NC}+} + f_{\mathrm{CC}}^{D}\tilde{\sigma}_{\mathrm{CC}+} \tag{30}$$

$$f_{\mathrm{NC}}^{D} = \frac{-1}{(1-y)^2 A_u - A_d};\quad f_{\mathrm{CC}}^{D} = \frac{A_u}{(1-y)^2 A_u - A_d}.$$

Here, $\tilde{\sigma}$ are the quantities derived from the differential NC and CC cross-sections as specified in (9, 14), and $A_{u,d}$ and $B_{u,d}$ are the coefficients given in (3). The arguments of these functions are obvious and have, therefore, been omitted in (27–30).

From the above solutions one can construct other interesting distributions, for example, the non-singlet structure function

$$x(u_v + d_v) = f_{\mathrm{NC}}^{u_v+d_v}\tilde{\sigma}_{\mathrm{NC}-} + f_{\mathrm{CC}}^{u_v+d_v}\tilde{\sigma}_{\mathrm{CC}-} \tag{31}$$

$$f_{\mathrm{NC}}^{u_v+d_v} = \frac{Y_+}{(1-y)^2 B_u + B_d};$$

$$f_{\mathrm{CC}}^{u_v+d_v} = \frac{B_d - B_u}{(1-y)^2 B_u + B_d},$$

the singlet structure function

$$F_s \equiv x(U+D) = f_{\mathrm{NC}}^{U+D}\tilde{\sigma}_{\mathrm{NC}+} + f_{\mathrm{CC}}^{U+D}\tilde{\sigma}_{\mathrm{CC}+} \tag{32}$$

$$f_{\mathrm{NC}}^{U+D} = \frac{-Y_-}{(1-y)^2 A_u - A_d};\quad f_{\mathrm{CC}}^{U+D} = \frac{A_u - A_d}{(1-y)^2 A_u - A_d}$$

and quark sea distributions such as

$$xU_{\text{sea}} = x(U - u_v), \quad xD_{\text{sea}} = x(D - d_v),$$
$$xS = xU_{\text{sea}} + xD_{\text{sea}}. \tag{33}$$

The relations (27–32) provide explicit examples for the manipulations which must be applied to the measured cross-sections in order to unfold the weak propagators and flavour-dependent electroweak couplings and to extract a desired quark distribution.

The practicability of the method proposed here depends among other things, on the properties of the coefficient functions f_{NC} and f_{CC} appearing in the relations (27–32). This point has been discussed briefly in our previous paper [5]. For completeness of the present study, and since here we use x and y as independent variables rather than x and Q^2 as in [5], we display these functions in Figs. 6 and 7 for two typical cases and recall the main features. The coefficients f^V_{NC} and f^V_{CC} of the unfolding relations for the valence quark distributions $V = u_v$, d_v and $u_v + d_v$ are well behaved functions of x and y. For different V the coefficients f^V_{NC} differ only by simple kinematical factors such as $(1-y)^2$ or Y_+, and also the functions f^V_{CC} are comparable with each other in shape and absolute magnitude. However, except at large y and/or x, $|f^V_{\text{NC}}| \gg |f^V_{\text{CC}}|$ as can be seen for $V = u_v$ in Fig. 6. This fact is explained by $B_{u,d} \sim P_Z = Q^2/(Q^2 + m_Z^2)$ being rather small unless $Q^2 \gtrsim \mathcal{O}(m_Z^2)$. As a consequence, the NC cross-sections play a very significant role in determining valence quark densities despite the smallness of the difference $\tilde{\sigma}_{\text{NC}-} \sim \tilde{\sigma}_{\text{NC}}(e^-) - \tilde{\sigma}_{\text{NC}}(e^+)$. This difference is mainly due to $\gamma - Z$ interference and, hence, becomes sizeable only at large values of $Q^2 = xys$. In summary, the reconstruction of valence quark distributions from data with the help of (27, 28, 31) should improve with increasing y until the event rates, which decrease with y, become too low. A feasible region is $y \geqq 0.15$ as demonstrated later.

In contrast, the coefficients f^F_{NC} and f^F_{CC} ($F = U, D, U + D$) are quite similar to each other, apart from different overall signs, as evident from the expressions (29, 30, 32). Other characteristic features can be inferred from Fig. 7 where we have plotted f^{U+D}_{NC} and f^{U+D}_{CC}. Most noteworthy are the divergencies caused by the vanishing of the denominator $(1-y)^2 A_u(Q^2) - A_d(Q^2)$ of $f^F_{\text{NC,CC}}$. For sufficiently low Q^2, $A_{u,d}(Q^2) \simeq e^2_{u,d}$ so that the problem occurs at $y = 1 - \sqrt{A_d/A_u} \simeq 1 - |e_d/e_u| = 1/2$. This case is exemplified by the low-x curves in Fig. 7. As x increases at fixed y, Q^2 increases and the divergencies are shifted to lower values of y due to the Z-boson contribution to $A_{u,d}(Q^2)$ given in (3). Thus, the unfolding procedure for U, D and F_s is not applicable in the region at and just below $Y \simeq 1/2$. However, this subtlety is not a real problem since one can simply exclude this problematic region when averaging over y, and thereby avoid the divergencies without significant loss in statistical precision. Only if one wants to determine distributions in x and Q^2 one has to be more careful [8]. Away from the divergencies, the coefficients $f^F_{\text{NC,CC}}$ are comparatively small in size and, moreover, multiply sums of NC and CC cross-sections. The latter observation together with Fig. 7 suggests to use data in the region $y \leqq 0.3$ for extracting distributions such as U, D and F_s.

Finally, the unfolding relations for sea quark distributions involve coefficients of the type $f^V_{\text{NC,CC}}$ ($V = u_v, d_v, u_v + d_v$) *and* $f^F_{\text{NC,CC}}$($F = U, D, U + D$). In more practical terms, one subtracts valence quark distributions from the structure functions U, D or $U + D$ as indicated in (33). Hence, firstly one has to avoid the region in y where the coefficients $f^F_{\text{NC,CC}}$ diverge, and secondly one has to go to large y in order to reconstruct valence quark densities. The only remaining region in y where sea quark densities can be unfolded is therefore $y \geqq 1/2$. This means that one cannot expect high statistical precision. Yet, for $y \geqq 0.6$ useful results can be obtained as shown in the next section.

4 Prospects of structure function measurements at HERA

In order to evaluate the prospects for the determination of structure functions and quark distributions from ep collisions using the strategies described in the last section, we apply these procedures to Monte Carlo (MC) data samples that may simulate real data obtainable in a few years of experimentation at HERA. As stated in the beginning, the present study is concerned with the problem of measuring the shape in x of quark distribution functions.

4.1 Monte Carlo procedure

The Monte Carlo program LEPTO [14] simulates ep collision events based on the theoretical cross-sections given in Sect. 2. For the input quark distributions we use the EHLQ1 parametrization [13] and for the electroweak parameters we take the values specified in Sect. 3. We only consider the deep inelastic region $Q^2 \geqq 100\,\text{GeV}^2$ where mass and higher twist effects should be negligible, and also restrict the x-region to $x \geqq 0.01$ where the usual $\log Q^2$ evolution equations in QCD should be valid [10, 15].

Sets of NC and CC events are generated corresponding to collisions of 30 GeV electrons and positrons with 820 GeV protons for an integrated luminosity of $200\,\text{pb}^{-1}$ per lepton beam. For the kinematic region just defined, and with the additional cut $y \geqq 0.01$, the total statistics for an $e^-(e^+)$ beam is 632k (626k) NC and 10.6k (4.9k) CC events. The number of events as well as the corresponding cross-sections are recorded in suitable bins of x and y. We note that for a given bin of y, the full range in x is kinematically allowed. Since we aim at the determination of x-distributions (and are not interested here in a detailed study of Q^2 dependences) it is thus advantageous to choose y, and not Q^2, as the

second independent variable. In order to assure samples with reasonable statistics, the bin-size is increased as x and y increases. For y we have chosen six bins per decade which are equally large on a logarithmic scale, i.e. $\Delta \log y = 0.167$. In x we have taken bins with $\Delta x = 0.05$ for $x \leqq 0.5$, $\Delta x = 0.1$ for $0.5 \leqq x \leqq 0.8$, and the bin $0.8 \leqq x \leqq 1$. For NC events we have in some cases exploited the high statistics available at low x and made a finer binning in the region $0.01 \leqq x \leqq 0.1$ (four equal bins on a log scale).

Systematic errors arising from the energy calibration and resolution, granularity and beam hole of the detectors as well as other possible sources have not been included in our Monte Carlo simulation. However, these effects have been investigated in some detail in [7]. There are, in principle, two ways to determine the values of x, y and Q^2 for a given event: either from an energy and angle measurement of the scattered electron or from the total hadron flow. The ranges of the above variables for which the systematic shifts of the differential cross-sections can be kept below 10% have been identified in [7]. The 'safe' regions are:

for electron measurements: $5 \times 10^{-5} \leqq x \leqq 0.6$,

$$Q^2 \geqq 5\,\mathrm{GeV}^2, \quad y \geqq 0.1 \tag{34}$$

for hadron measurements: $0.01 \leqq x \leqq 0.5$,

$$Q^2 \geqq 100\,\mathrm{GeV}^2, \quad y \geqq 0.03. \tag{35}$$

CC events only allow measurements through the hadron flow. For NC events, on the other hand, one can use either measurement such that the accessible phase space is larger than in the CC case and the reconstructed values of x, y and Q^2 can be cross-checked. In order to make sure that our results based on statistical considerations are not spoiled by too large systematic uncertainties, we apply the cuts detailed in (34, 35) in our analysis. In addition, we indicate MC results at $x \geqq 0.5$ or 0.6 in order to illustrate the statistical power which would be available also in the high-x region.

To gain statistics in the x-shape measurements we average the values at fixed x of a structure function $F(x, y_i)$, obtained through one of the relations given in Sect. 3, over suitable y-bins using the errors $\pm \varepsilon_i$ on $F(x, y_i)$ as weights:

$$F(x) = \frac{\sum_i F(x, y_i)/\varepsilon_i^2}{\sum_i 1/\varepsilon_i^2}; \quad \varepsilon(x) = \frac{1}{\sqrt{\sum_i 1/\varepsilon_i^2}}. \tag{36}$$

The averaged result $F(x) \pm \varepsilon(x)$ refers to an average value of y,

$$\bar{y} = \frac{\sum_i y_i/\varepsilon_i^2}{\sum_i 1/\varepsilon_i^2}, \tag{37}$$

which in general varies with x. The errors ε_i are obtained from the statistical errors on the number of NC and CC events in a given bin in x and y, after propagation through the relevant unfolding relation. Schematically,

$$\varepsilon_i = \sqrt{\sum_k f_k^2 \frac{\tilde{\sigma}_k^2}{N_k}} \tag{38}$$

where the sum runs over the included data samples $(e^{\mp}; \mathrm{NC}, \mathrm{CC})$ and $\tilde{\sigma}_k, N_k$ and f_k denote the correspond-

Table 1. Summary of structure function extraction

Structure function	Figure no.	approximation[a] single[b]	comb.[c]	exact[a] unfolding	important y-region	measurability
xu_v	8a, b	15	—	27	low/high	**
xd_v	—	16	—	28	—	—
$x(u_v + d_v)$	9a, b	—	21	31	low/high	**
$x(u_v - d_v)$	10	—	26	—	very low	*
$x(e_u u_v - e_d d_v)$	11	—	13	—	high	*
$x(u_s + S/2)$	12a	17	—	—	very low	***
$x(d_s + S/2)$	12b	20	—	—	very low	***
F_s	13a, b	—	25c	32	low	***
xU	14a	—	—	29	low	***
xD	14b	—	—	30	low	*
xU_{sea}	15a	—	—	33	very high	*
xD_{sea}	—	—	—	33	—	—
xS	15b	—	—	33	very high	*
F_2^{em}	16	10	—	—	low	****

[a] Numbers refer to equation numbers in the text
[b] Extraction from a single cross-section (NC or CC, $e^- p$ or $e^+ p$)
[c] Extraction from a combination of $e^- p$ and $e^+ p$ cross-sections (NC or CC)

ing scaled cross-sections, numbers of events and coefficient functions in the reconstruction relations. For bins with a few events only, we use Poisson statistics instead of Gaussian to obtain proper error estimates.

Finally, we remark on a technical problem arising from the finite bin size. The differential cross-sections $\tilde{\sigma}_k$ in a certain bin are to be multiplied by known functions f_k to be evaluated at an (x, y) point representative for that bin. The choice of this point is actually quite essential since some of the functions f_k vary significantly with x and y even within the rather small bins we employ. We have observed that using a bin edge can produce catastropic shifts, but also the bin center or the mean value of (x, y) determined from the data is not an appropriate choice. In order to completely avoid this problem we formally consider the cross-section in a bin as a sum over the events in that bin. The relevant coefficient functions f_k can then be applied to each event, using the generated (x, y) value of the event. Experimentally, this may not be very practical and one would instead fit the measured cross-sections to the centers of the (x, y) bins which are then used for evaluating the functions f_k. The final result should be the same and, hence, we have employed the former method in our simulation for simplicity.

4.2 Results

We are now ready to present and discuss the main results of our studies. Table 1 summarizes the structure functions which, following the formal considerations of Sect. 3, can be reconstructed from cross-section measurements for NC and CC ep scattering at fixed c.m. energy. The list is rather complete as we think. Indicated in Table 1 are the relations used to reconstruct a particular structure function from the Monte Carlo generated data samples described in the last subsection. It can furthermore be inferred from Table 1 whether a given extraction method yields the desired structure function as a strict result or in a certain approximation, and whether the extraction requires only a single cross-section or the combination of different cross-sections. The resulting distributions are plotted in all cases where the statistical uncertainties allow a meaningful determination of the shape in x. The figure numbers are also given in Table 1, for convenience.

Column six of Table 1 characterizes the regions in y which have turned out most suitable for reconstruction. Of course, the boundaries of these regions are not uniquely defined. Taking into account the results of Sect. 3, the statistical errors and properties of the weighted averaging specified in (36), we find the following choice optimal:

(i) $0.03 \leqq y \leqq 0.15$ 'very low' y
(ii) $0.03 \leqq y \leqq 0.3$ 'low' y
(iii) $0.15 \leqq y \leqq 1$ 'high' y
(iv) $0.65 \leqq y \leqq 1$ 'very high' y.

The labels in quotation marks refer to the classification used in Table 1. The lower boundary $y = 0.03$ in (i) and (ii) is set by the experimental reconstruction of the kinematic variables from the hadron flow measurement as discussed in [7] and stated in (35). The upper boundaries in (iii) and (iv) should lie as near to the phase space limit $y = 1$ as possible. However, a cut $y \leqq 0.9$ which partly avoids the very large radiative corrections predicted for $y \to 1$ [4, 12] would also be acceptable. Less trivial is the choice of the remaining boundaries. The upper boundary in (i) is a compromise: lowering it would decrease the statistical accuracy of the results shown in Figs. 10 and 12; raising it would restrict the applicability of the approximations, (17, 20, 26), to larger values of x. This can be seen, for example, in Fig. 2 of [6] where the result of an extraction of $x(u_v + S/2)$ is displayed similarly as in Fig. 12a of the present paper but using the range $0.03 \leqq y \leqq 1$. In contrast, the upper boundary

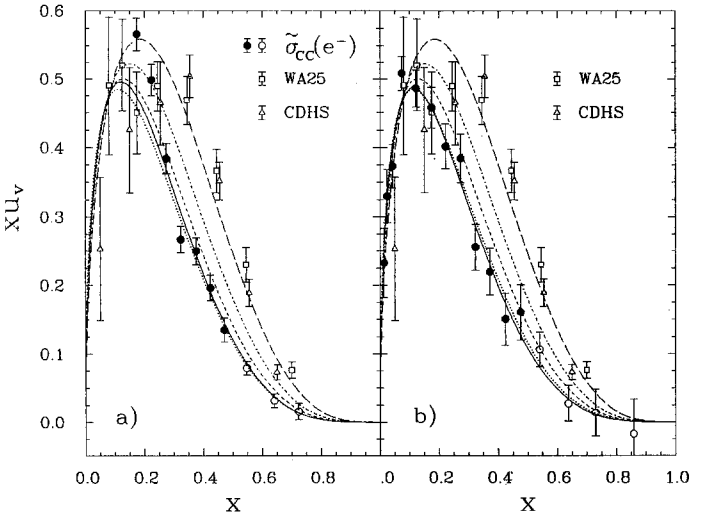

Fig. 8 a, b. The valence up-quark distribution $xu_v(x)$: **a** Approximate determination from the CC e^-p cross-section, (15), averaged over $0.03 \leqq y \leqq 0.3$; **b** Extraction from the NC and CC $e^{\mp}p$ cross-sections, (27), averaged over $0.15 \leqq y \leqq 1$. The Monte Carlo results corresponds to an integrated luminosity of $200\,\text{pb}^{-1}$ for e^-p and equally for e^+p collisions, and are drawn as full circles in case the phase space restrictions (34, 35) are satisfied. The curves represent $xu_v(x, Q^2)$ evaluated from the input distributions [13] for $Q^2 = 10$ (long-dashed), 10^2 (dash-dotted), 10^3 (dashed), 10^4 (dotted) GeV^2 and $\overline{Q^2} = x\bar{y}s$ (full), $\bar{y}$ being the average value of y associated with the MC data. Also plotted are existing measurements of xu_v in neutrino scattering at $Q^2 = 11\,\text{GeV}^2$ [17] (open squares) and $Q^2 = 15\,\text{GeV}^2$ [18] (open triangles) with statistical and systematic errors added in quadrature

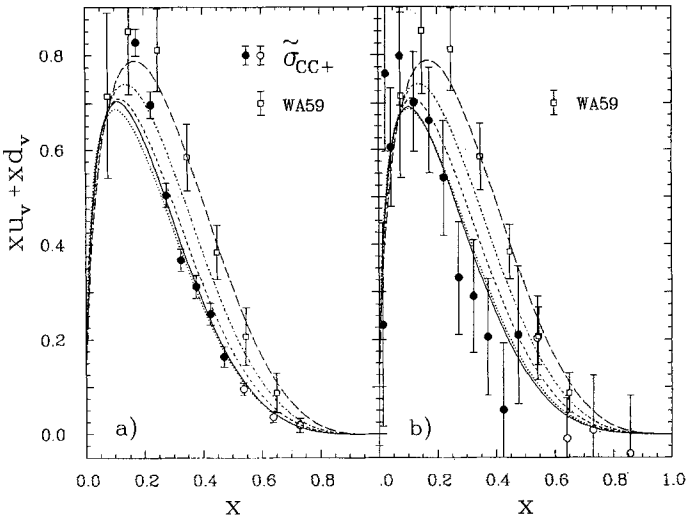

Fig. 9 a, b. The total valence quark distribution $x(u_v + d_v)$: **a** Approximate determination from the sum of CC $e^- p$ and $e^+ p$ cross-sections, (21), averaged over $0.03 \leqq y \leqq 0.3$; **b** Extraction from the NC and CC $e^{\mp} p$ cross-sections, (31), averaged over $0.15 \leqq y \leqq 1$. Further explanations on the MC result and the curves are given in Fig. 8. Also plotted is an existing measurement of $x(u_v + d_v)$ in neutrino scattering at $Q^2 = 11.3\,\mathrm{GeV}$ [19] (open squares) with statistical and systematic errors added in quadrature

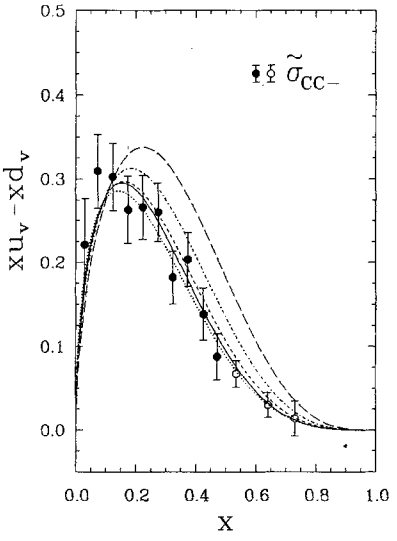

Fig. 10. Approximate determination of the difference of the valence up- and down-quark distributions, $x(u_v - d_v)$, from the difference of the CC $e^- p$ and $e^+ p$ cross-sections, (26), averaged over $0.03 \leqq y \leqq 0.15$. Further explanations on the MC result and the curves are given in Fig. 8

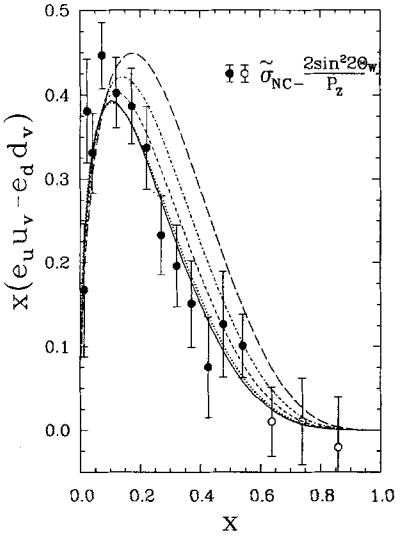

Fig. 11. Approximate determination of the valence quark distribution $x(e_u u_v - e_d d_v)$ from the difference of the NC $e^- p$ and $e^+ p$ cross-sections, (13), averaged over $0.15 \leqq y \leqq 1$. Further explanations on the MC result and the curves are given in Fig. 8

in (ii) is not so significant. If $y \leqq 1$ were chosen instead of $y \leqq 0.3$ the changes to Figs. 8a, 9a, 13a and 16 would hardly be visible. The reason is simply that in these cases the statistical uncertainties of the resulting distributions increase with y and hence the high-y region contributes little to the weighted averages, (36). This is also true for the remaining cases of this class displayed in Figs. 13b, 14a and 14b, except that one should obviously omit the region around $y \simeq 1/2$ where the coefficients appearing in the relevant unfolding relations (29, 30, 32) are divergent. For a similar reason one may include data at $y \leqq 0.15$ in the analysis of class (iii), although the results shown in Figs. 8b, 9b and 11 cannot be significantly improved by extending the analysis to low values of y because of increasing statistical errors. This holds correspondingly for class (iv) and the region in y below $y \simeq 1/2$, the bin where some of the coefficients in the relations (33) for the sea quark distributions are divergent.

At this point we should stress that the departure of an approximately reconstructed distribution from the actually desired structure function is not determined by how much a relation such as (10) is maximally violated in the range of y used for the reconstruction. It is rather the deviation at the average value $\bar{y}$ given by (37) which counts. A good example is provided by F_2^{em}. As can be seen from Figs. 2b and 16, $\tilde{\sigma}_{NC}(e^+)$ deviates from F_2^{em} by about 5% at $x \simeq 0.5$ and $y \simeq 0.3$, whereas after averaging over y in the range $0.03 \leqq y \leqq 0.3$ the deviation is typically below the 1%

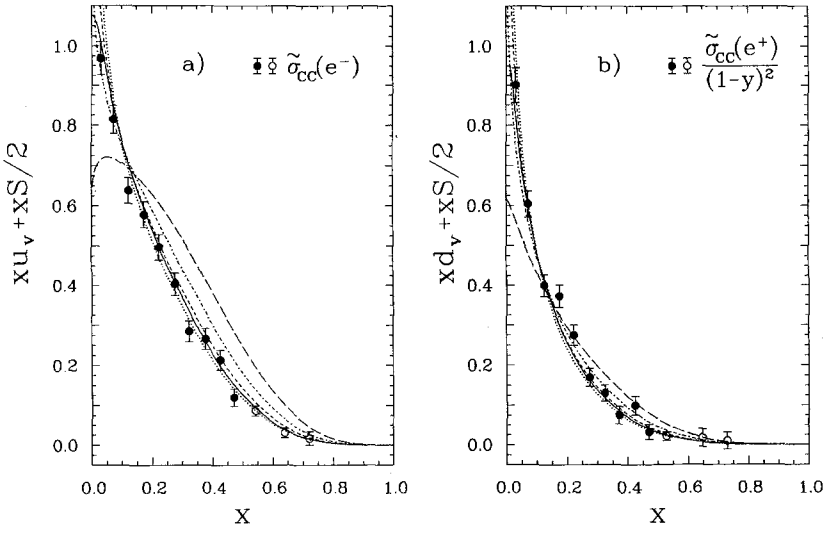

Fig. 12 a, b. Approximate determination of the quark distributions **a** $x(u_v + S/2)$ from the CC $e^- p$ cross-section, (17), and **b** $x(d_v + S/2)$ from the CC $e^+ p$ cross-section, (20), averaged over $0.03 \leqq y \leqq 0.15$. Further explanations on the MC result and the curves are given in Fig. 8

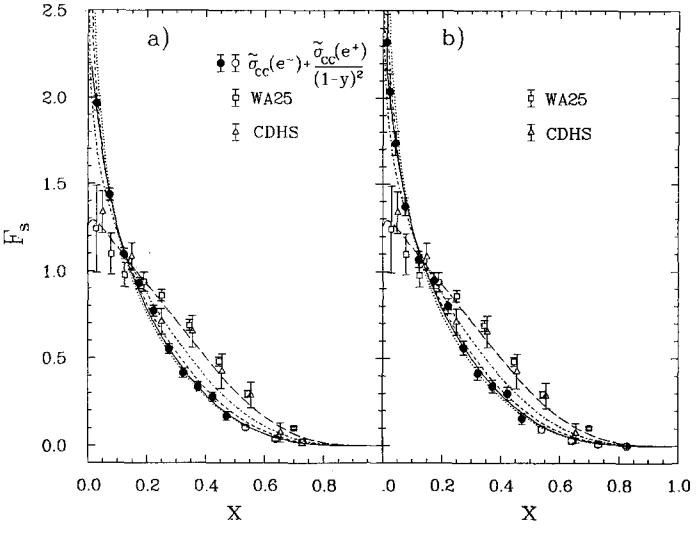

Fig. 13 a, b. The total flavour singlet structure function $F_s = \sum_f x(q_f + \bar{q}_f)$: **a** Approximate determination from the combined CC $e^- p$ and $e^+ p$ cross-sections, (25c), averaged over $0.03 \leqq y \leqq 0.3$; **b** Extraction from the NC and CC $e^\mp p$ cross-sections, (32), averaged over $0.03 \leqq y \leqq 0.3$. Further explanations on the MC result and the curves are given in Fig. 8. Also plotted are existing (approximate) measurements of F_s in neutrino scattering at $Q^2 = 11\,\mathrm{GeV}^2$ [17] (open squares) and $Q^2 = 15\,\mathrm{GeV}^2$ [18] (open triangles) with statistical and systematic errors added in quadrature

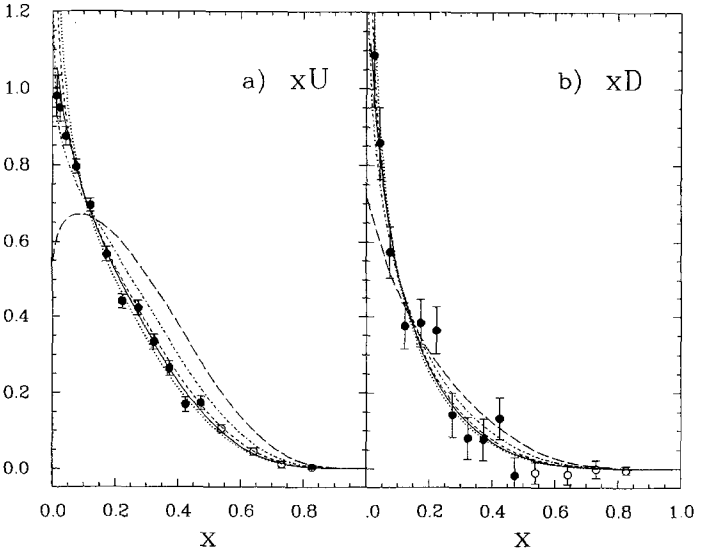

Fig. 14 a, b. Separate extraction of **a** the total up-type quark distribution $xU = x(u + \bar{u} + c + \bar{c})$, (29), and **b** the total down-type quark distribution $xD = x(d + \bar{d} + s + \bar{s})$ (30), from the NC and CC $e^\mp p$ cross-sections, averaged over $0.03 \leqq y \leqq 0.3$. Further explanations on the MC result and the curves are given in Fig. 8

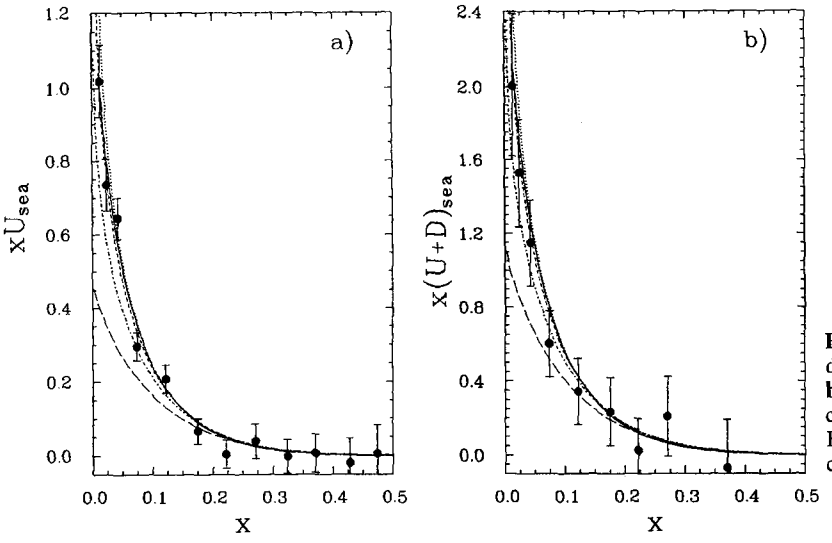

Fig. 15 a, b. Separate extraction of the sea quark distributions for **a** the up-type quark flavours and **b** all quark flavours from the NC and CC $e^{\mp}p$ cross-sections, (33), averaged over $0.65 \leqq y \leqq 1$. Further explanations on the MC result and the curves are given in Fig. 8

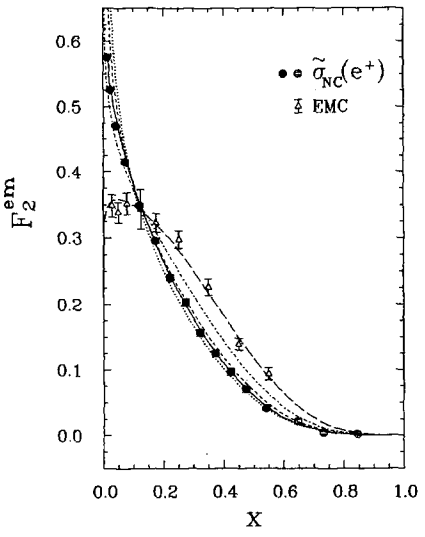

Fig. 16. Determination of the electromagnetic structure function F_2^{em} from the NC e^+p cross-section, (10), averaged over $0.03 \leqq y \leqq 0.3$. Further explanations on the MC result and the curves are given in Fig. 8. Also plotted is an existing measurement of F_2^{em} in muon scattering at $Q^2 = 11.5\,\mathrm{GeV}^2$ [20] (open triangles) with statistical and systematic errors added in quadrature

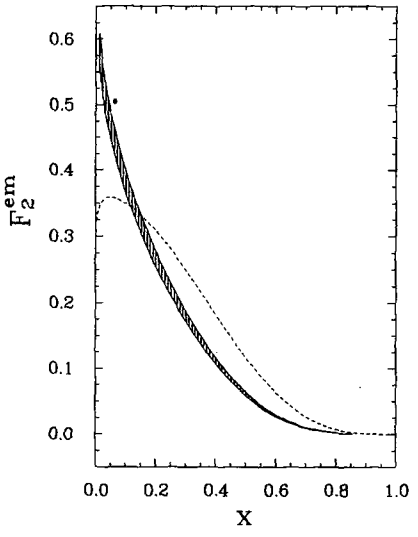

Fig. 17. Systematic uncertainty on F_2^{em} (shaded band) due to errors of $\pm 1\%$ and $\pm 2\%$ in the calibration of electromagnetic and hadronic calorimeters, respectively. The dashed curve represents F_2^{em} at $Q^2 = 10\,\mathrm{GeV}^2$ and indicates the distribution of existing data (see also Fig. 16)

level as expected for $\bar{y} \simeq 0.07$. It should also be pointed out that the selection (i) to (iv) of regions in y can be improved and optimized in several ways. One obvious possibility are x-dependent cuts in y as suggested by the studies illustrated in Figs. 2 to 7. Since such sophistications do not change our main conclusions, we have kept the procedure as simple as possible.

Finally, in the last column of Table 1, it is roughly indicated how well the listed functions can be extracted from the data. We repeat that in this evaluation systematic uncertainties are only considered through cuts on the kinematical variables. In this sense, our numerical results may be somewhat on the optimistic side.

Before commenting in turn on Figs. 8 to 16, it may be helpful to clarify a few general points. Each figure is labeled by the symbol of the structure function (see Table 1) which one wants to extract. If an approximate method is used the actual quantity represented by the Monte Carlo data is specified in the figure, e.g. $\tilde{\sigma}_{CC}(e^-)$ in Fig. 8a dealing with the structure function xu_v. If a distribution is unfolded from the cross-sections according to the procedure described in Sect. 3.2, the Monte Carlo data are supposed to be identical to the desired structure function. The Monte Carlo results are shown with full symbols when they lie within the 'safe' kinematical regions, (34, 35), whereas results for $x \geqq 0.5$ or 0.6 are drawn as open symbols to indicate

that measurements are still possible from the statistical point of view, but are presumably affected by larger systematic errors [7]. Furthermore, the broken curves in each figure show theoretical distributions in x for fixed $Q^2 = 10, 10^2, 10^3, 10^4\,\mathrm{GeV}^2$ calculated from the input quark distributions [13] in the Monte Carlo generation. The full curve, drawn in each figure, represents the same structure function as the broken curves in that figure, but is calculated for the average value $\overline{Q^2} = \bar{y}xs$ given by (37) which is associated with the MC result and which varies bin-by-bin in x. In other words, a difference between the full curve and the MC data points (except for statistical fluctuations) indicate the breakdown of a particular approximation. In contrast, differences between the full and the broken curves arise from different Q^2-scales used in the evaluation. These auxiliary curves are useful for answering the following questions: What is the region of validity of a certain approximation? What is the typical range of Q^2 associated with a reconstructed quark distribution? Is it possible to observe a significant difference in shape due to the evolution from low Q^2 to the Q^2 range of HERA experiments? Finally, wherever possible we have also plotted real data from fixed-target experiments, preferentially for $Q^2 \approx 10\,\mathrm{GeV}^2$ and with statistical and systematic errors added in quadrature. Comparison of these data with the MC results illustrates the potential of HERA and allows for a judgement of the observability of global evolution effects.

The most important observations can be summarized as follows:

(I) The valence up-quark density can be obtained at $x \gtrsim 0.25$, i.e. in the valence quark region, directly from the CC e^-p cross-section (Fig. 8a), while for smaller values of x the more demanding unfolding procedure based on (27) has to be used (Fig. 8b). As expected, the statistical accuracy of the latter is somewhat worse than the accuracy of the approximate, but direct extraction (in spite of the twice as long running time needed for the exact unfolding). Note that both methods yield u_v at high values of Q^2, typically around $10^4\,\mathrm{GeV}^2$, i.e. two to three orders of magnitude above presently accessible scales. As illustrated in Fig. 8, it should be possible to check the evolution of xu_v predicted by QCD at least in a rough way.

(II) A separate determination of the valence down quark density is more difficult. The unfolding method defined by relation (28) does not seem to yield a statistically significant measurement of xd_v. However, useful constraints can be derived from the CC e^+p cross-section in the valence quark region $x \geqq 0.25$. In this region, the measurement illustrated in Fig. 12b can be considered a determination of xd_v

(III) The sum and difference of the valence up and down quark distributions, $x(u_v \pm d_v)$, can be extracted in various ways as demonstrated in Figs. 9 and 10. Because of the problem with the determination of the valence down-quark density from (28), the achievable statistical accuracy on $x(u_v + d_v)$ in Fig. 9b is not as good as for xu_v (Fig. 8a) but still satisfactory. Nevertheless, a comparison of HERA results with existing data should allow the observation of the evolution predicted by QCD, in particular in the valence quark region (Fig. 9a).

(IV) Additional information on the valence quark distributions can be obtained from the NC structure function xF_3. Fig. 11 shows the expected measurement of the distribution $x(e_u u_v - e_d d_v)$ using the relation (13).

(V) The e^-p and e^+p CC cross-sections yield nice measurements of the distributions $xu_v + xS/2$ and $xd_v + xS/2$, respectively, where S is the total sea quark density in the proton. With the restriction $y \leqq 0.15$ good results can be obtained for all $x \geqq 0.01$ as shown in Fig. 12. The average scale associated with such a measurement varies from $Q^2 \approx 10^2 - 10^3\,\mathrm{GeV}^2$ at small x to $Q^2 \approx 10^3 - 10^4\,\mathrm{GeV}^2$ at large x.

(VI) In (I) to (IV) we have considered pure non-singlet distributions. It is also interesting to see how well one can expect to determine the total flavour singlet structure function $F_s = \sum_f x(q_f + \bar{q}_f)$. The two best possibilities which we have found are illustrated in Fig. 13. An approximate measurement of F_s is provided by the CC cross-section for e^+p and e^-p using relation (25c). The MC result is shown in Fig. 13a. If one also uses measurements of the NC cross-sections one can unfold F_s without any approximation as specified in (32), and obtain the result shown in Fig. 13b. By comparing Figs. 13a and b one clearly sees that the approximate method using CC cross-sections alone is favourable from the statistical point of view. In fact, it seems to provide a very good measurement of F_s at average values of Q^2 similar to the ones mentioned in (V). Also shown in Fig. 13 are existing data on F_s at lower Q^2. More accurately, these data are also approximations of F_s to the extent that the difference of the strange and charm quark sea is neglected. Figure 13 shows that one can hope to observe a clear change in the shape of F_s due to the evolution in QCD.

(VII) Not only is it possible to extract F_s, one can also separate the charge 2/3 from the charge 1/3 quark distributions, xU and xD, respectively, using the unfolding relations (29) and (30). As can be seen from Fig. 14, the momentum distribution for the up-type quarks can be determined with a better statistical precision than the down-type quark distribution in reminiscence of what we have observed in the case of the valence quarks u_v and d_v. The average Q^2 scale associated with these measurements again varies with x as described in (V).

(VIII) On the other hand, a separation of sea quarks from valence quarks is more difficult. As explained in Sect. 3, such a separation appears possible only at large y. Nevertheless, the charge 2/3 quark sea distribution xU_{sea} and the total quark sea distribution xS can be obtained with satisfactory statistical accuracy as shown in Fig. 15a, b. However, we have

found no way to a meaningful determination of the charge 1/3 quark sea distribution xD_{sea}. Furthermore, it is interesting to note the high scale $Q^2 \simeq 10^4\,\mathrm{GeV}^2$ associated with measurements such as the ones exemplified in Fig. 15.

(IX) Finally, F_2^{em} is the structure function that can be determined most accurately. The statistical errors for a run of $200\,\mathrm{pb}^{-1}$ are, in fact, so small that the corresponding error bars lie inside the MC data symbols drawn in Fig. 16. Moreover, the departure of F_2^{em} from the actually measured distribution $\tilde{\sigma}_{NC}(e^+)$ is smaller than the statistical errors as long as $Q^2 \lesssim 3000\,\mathrm{GeV}^2$ or $y \lesssim 0.3$. This can be seen from the precise agreement between the theoretical expectation on $F_2^{em}(x, \overline{Q^2})$ represented by the full curve in Fig. 16 and the MC data on $\tilde{\sigma}_{NC}(e^+)$. The clear change in shape of the MC result as compared to the existing EMC data on F_2^{em} at $Q^2 \simeq 10\,\mathrm{GeV}^2$ shows rather impressively the observability of the QCD evolution. We should emphasize that F_2^{em} plays the most important role for more quantitative QCD tests [4, 8]. In addition, F_2^{em} can also be measured very well in the region $10^{-4} \leqq x \leqq 10^{-2}$ as discussed in [4, 6], and thus may serve as a sensitive probe of low-x physics.

Finally, we note that in cases where our study overlaps with the investigation in [4] the results generally agree.

5 Summary and conclusions

In this paper we have studied strategies and prospects for extracting stucture functions and individual quark distributions from NC and CC cross-sections measurable at *ep* colliders such as HERA. We have concentrated on possibilities which exist at fixed *ep* centre-of-mass energy. The available methods can be characterized as follows:

(i) approximate determination from a single cross-section (NC or CC, e^-p or e^+p),
(ii) approximate determination from combinations of cross-sections (NC or CC, e^-p and e^+p),
(iii) exact unfolding using the four independent NC and CC cross-sections for $e^{\mp}p$ scattering.

The main objective of our study has been the determination of the shape of structure functions in the Bjorken-variable x, and not so much a detailed analysis of the Q^2 dependence expected in QCD. It is thus useful to average the resulting x-distributions over suitable ranges in y taking the inverse statistical errors as weights. This improves the statistical accuracy, but washes out details of scaling violations. Nevertheless, it is still possible to check the gross effect of the Q^2 evolution. Fistly, the average value $\bar{y}$ defined in (37), varies relatively little with x so that the average scale parameter $\overline{Q^2} = x\bar{y}s$ is roughly proportional to x, a correlation that can easily be taken into account when confronting data with theory. Secondly, the values of $\overline{Q^2}$ typical for HERA experiments lie one to two orders of magnitude above the presently accessible range in Q^2. Hence, one expects sizeable differences in the corresponding quark momentum distributions due to QCD evolution. For a number of structure functions these differences should be clearly observable as demonstrated by Monte Carlo simulations in Sect. 4. Most promising in this respect are the valence quark distributions u_v and $u_v + d_v$, the singlet structure function F_s and the electromagnetic structure function F_2^{em}.

Furthermore, we have (a) systematically investigated which structure functions and quark distributions can be extracted from *ep* collisions, (b) clarified the advantages and disadvantages of different approaches to extract the same structure function, and (c) compared results which can be expected from HERA measurements with existing fixed-target measurements. Realistic estimates of the statistical accuracy of structure function measurements at HERA are obtained by Monte Carlo event simulation assuming an integrated luminosity of $200\,\mathrm{pb}^{-1}$ for e^-p and the same for e^+p scattering. This should correspond to one or two years running time per beam configuration. The kinematical variables x, y and Q^2 are restricted to a region that has been found safe in the sense that the systematic errors on the cross-section measurements are expected to be less than 10%. Table 1 represents a summary of our results. It should be pointed out that the possibility to measure differential cross-sections down to low values of y is essential for the extraction of several structure functions as emphasized in Table 1. The experimentally favoured range given in [7] is $y \geqq 0.1$ or 0.03 depending on whether the kinematical variables are reconstructed from the scattered lepton or the total hadron flow. Thus, the latter method is of great importance, not only in charged current events. Apart from phase space cuts, systematic errors have been neglected since their inclusion would require a detailed detector simulation. Moreover, we have assumed the correction of the data for radiative effects to be a solvable problem that will not drastically change the conclusions presented here.

In order to get a rough idea about systematic uncertainties on structure funtions that remain even after restricting the kinematical regions used for extraction we have estimated the influence of errors in the energy calibration of the calorimeters which is expected to be the most severe problem. Since F_2^{em} is the structure function with the smallest statistical errors, it is most vulnerable to systematic problems. Therefore, we have choosen F_2^{em} as a test case. Following [16] we introduce shifts in F_2^{em} arising from systematic calibration errors of $\pm 1\%$ in the electromagnetic and $\pm 2\%$ in the hadronic energy measurements as considered realistic for the H1 detector. The procedure and results are worked out in more detail in [8], since this question is particularly important for quantitative QCD tests studied there. Here, we simply indicate the resulting uncertainty in F_2^{em} by the shaded band in Fig. 17. Although the

systematic uncertainty is clearly larger than the very small statistical errors in F_2^{em}, barely visible in Fig. 16, it would not upset the x-shape measurement.

Thus, we can conclude in summary, that the prospects for determining the shape in x of structure functions and quark distributions at *ep* colliders such as HERA look rather promising. Even with x-distributions that are averaged over y (or Q^2), a global check of the Q^2 evolution expected in QCD appears possible by comparison with existing data from fixed target experiments which have a substantially lower mean Q^2 scale than the typical scales associated with HERA experiments. For a detailed analysis of QCD scaling violations the data must, of course, be binned in Q^2. In agreement with [4] we find that the most promising structure function for this purpose is F_2^{em}. A quantitative analysis involving a proper fit to determine Λ_{QCD} is in progress and will be presented in a forthcoming paper [8].

Acknowledgements. We want to thank J. Blümlein, F. Eisele, M. Glück, M. Klein and E. Reya for interesting discussions and helpful suggestions.

References

1. For recent reviews on deep inelastic scattering experiments, see e.g. T. Sloan: Proc. Int. Europhysics Conference on High Energy Physics, Uppsala, 1987, O. Botner (ed.): European Physical Society Vol. II, p. 857 (1987); R. Voss: Proc. Int. Symposium on Lepton and Photon Interactions at High Energies, Hamburg, 1987, W. Bartel, R. Rückl (eds.) p 581 Amsterdam: North-Holland 1988, T. Sloan, G. Smadja, R. Voss: Phys. Rep. 162 (1988) 45
2. F. Brasse: Proc. of the Workshop on Experimentation at HERA, Amsterdam, 1983, DESY HERA 83/20, p. 499
3. M. Klein, T. Riemann: Z. Phys. C—Particles and Fields 24 (1984) 151
4. J. Blümlein, M. Klein, T. Naumann, T. Riemann: Proc. of the HERA Workshop, Hamburg, 1987, (ed.) R.D. Peccei, DESY Hamburg Vol. 1, p. 67 (1988); J. Blümlein, M. Klein, T. Naumann: Berlin preprint PHE 88-12 and to appear in: New Theories in Physics, Proc. Int. Conf. High Energy Physics, Kazimierz, Poland, 1988, Z. Adjuk, (ed.) Singapore World Scientific
5. G. Ingelman, R. Rückl: Phys. Lett. 201B (1988) 369
6. G. Ingelman, R. Rückl: Proc. of the HERA Workshop, Hamburg, 1987, R.D. Peccei (ed.) DESY Hamburg Vol. 1, p. 107 1988
7. J. Feltesse: Proc. of the HERA Workshop, Hamburg, 1987, R.D. Peccei (ed.) DESY Hamburg, Vol. 1, p. 33 1988
8. J. Blümlein, G. Ingelman, M. Klein, R. Rückl: in preparation
9. M. Glück, R.M. Godbole, E. Reya: Z. Phys. C—Particles and Fields 38 (1988) 441; M. Glück: Proc. of the HERA Workshop, Hamburg, 1987 R.D. Peccei (ed.) DESY Hamburg, Vol. 1, p. 119 1988
10. G. Altarelli, G. Parisi: Nucl. Phys. B126 (1977) 298; G. Altarelli: Phys. Rep. 81 (1982) 1
11. A.M. Cooper-Sarker et al.: Z. Physik C—Particles and Fields 39 (1988) 281
12. For discussions of radiative corrections in deep inelastic scattering and references to original papers, see e.g. W. Hollik: Proc. of the HERA Workshop, Hamburg, 1987 R.D. Peccei (ed.) DESY Hamburg, Vol. 2, p. 579 (1988); H. Spiesberger, ibid p. 605; J. Kripfganz, H. Perlt, ibid p. 645; C. Kiesling, ibid p. 653
13. E. Eichten, I. Hinchliffe, K. Lane, C. Quigg: Rev, Mod. Phys. 56 (1984) 579; 58 (1986) 1047
14. G. Ingelman: LEPTO version 5.2, DESY preprint in preparation
15. L.V. Gribov, E.M. Levin, M.G. Ryskin: Phys. Rep. 100 (1983) 1; J. Kwiecinski: Z. Phys. C—Particles and Fields 29 (1985) 561
16. J. Feltesse: H1 Internal Report H1-4/85-04
17. D. Allasia et al. WA25 Coll.: Z. Phys. C—Particles and Fields 28 (1985) 321
18. H. Abramowicz et al. CDHS Coll.: Z. Phys. C—Particles and Fields 25 (1984) 29
19. K. Varvell et al. WA59 Coll.: CERN/EP 87-46
20. J.J. Aubert et al. EMC Coll.: Nucl. Phys. B259 (1985) 189