Parton distributions: HERA–Tevatron–LHC

Graeme Watt[*]

Theory Group, Physics Department, CERN, CH-1211 Geneva 23, Switzerland
E-mail: Graeme.Watt@cern.ch

The parton distribution functions (PDFs) are a non-negotiable input to almost all theory predictions at hadron colliders. In this talk, I introduce PDF determination by global analysis and discuss selected topics concerning recent relevant data from HERA and the Tevatron, before giving some prospects for the LHC. The combination of H1 and ZEUS cross sections reduces uncertainties and will be an important input to future global PDF analyses. The theoretical description of the heavy-quark contribution to structure functions at HERA has a significant influence on predictions at the LHC. New W and Z data from the Tevatron Run II provide important PDF constraints, but there are currently problems describing the latest data on the lepton charge asymmetry from $W \to \ell\nu$ decays. The Tevatron Run II jet production data prefer a smaller high-x gluon than the previous Run I data, which impacts on predictions for Higgs cross sections at the Tevatron. It is now possible to consistently calculate a combined "PDF+α_S" uncertainty on hadronic cross sections, which is around 2–3% for the W and Z total cross sections at the LHC, reflecting their potential as a "standard candle" to measure machine luminosity. Parton luminosity functions are useful quantities for studying properties of hadronic cross sections. Precision measurements at the LHC will provide further constraints on PDFs as data accumulates in the early running period.

---

[*]Speaker.

1. Introduction

Protons are not elementary particles: they are made of *partons* (quarks and gluons). Parton distribution functions (PDFs) are therefore essential to relate theory to experiment at HERA, the Tevatron and the LHC [1]. Each PDF $f_{a/A}(x, Q^2)$ intuitively gives the *number density* of partons a in a hadron A with momentum fraction x at a hard scale $Q^2 \gg \Lambda^2_{\text{QCD}}$. The "standard" perturbative QCD framework at hadron colliders is fixed-order collinear factorisation, which holds up to formally power-suppressed ("higher-twist") terms $\mathcal{O}(\Lambda^2_{\text{QCD}}/Q^2)$. A hadronic cross section σ_{AB} can be written as a sum of partonic cross sections $\hat{\sigma}_{ab}$, each expanded as a perturbative series in the running strong coupling $\alpha_S(Q^2)$, convoluted with a PDF for each hadron, i.e.

$$\sigma_{AB} = \sum_{a,b=q,g} \left[\hat{\sigma}^{\text{LO}}_{ab} + \alpha_S \, \hat{\sigma}^{\text{NLO}}_{ab} + \alpha_S^2 \, \hat{\sigma}^{\text{NNLO}}_{ab} + \ldots \right] \otimes f_{a/A}(x_a, Q^2) \otimes f_{b/B}(x_b, Q^2), \quad (1.1)$$

where $\otimes$ indicates a convolution over the momentum fraction $x_{a,b}$. The scale dependence of the PDFs is given by the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) evolution equations:

$$\frac{\partial f_{a/A}}{\partial \ln Q^2} = \frac{\alpha_S}{2\pi} \sum_{a'=q,g} \left[P^{\text{LO}}_{aa'} + \alpha_S P^{\text{NLO}}_{aa'} + \alpha_S^2 P^{\text{NNLO}}_{aa'} + \ldots \right] \otimes f_{a'/A}, \quad (1.2)$$

while the running of $\alpha_S(Q^2)$ satisfies the renormalisation group equation. However, the input values $f_{a/A}(x, Q_0^2)$ and $\alpha_S(Q_0^2)$ to the evolution equations are incalculable by perturbative QCD and so need to be extracted from data. Structure functions in deep-inelastic scattering (DIS) can similarly be written in terms of perturbative coefficient functions, $C_{i,a}$, convoluted with PDFs, i.e.

$$F_i\left(x, Q^2\right) = \sum_{a=q,g} \left[C^{\text{LO}}_{i,a} + \alpha_S C^{\text{NLO}}_{i,a} + \alpha_S^2 C^{\text{NNLO}}_{i,a} + \ldots \right] \otimes f_{a/A}. \quad (1.3)$$

Since the PDFs are *universal*, they can be determined from a wide range of existing data, for example, from the HERA *ep* collider (H1 and ZEUS experiments), from fixed-target experiments in ℓp and ℓd scattering (BCDMS, NMC, E665, SLAC), νN scattering (CCFR, NuTeV, CHORUS), pp and pd scattering (E866/NuSea), together with $p\bar{p}$ collider data from the Tevatron (CDF, DØ).

The paradigm for PDF determination by "global analysis" is to parameterise the x dependence of $f_{a/A}(x, Q_0^2)$ for each flavour $a = q, g$ at the input scale $Q_0^2 \sim 1 \text{ GeV}^2$ in some flexible form, subject to number- and momentum-sum rule constraints. The PDFs are then evolved to higher scales $Q^2 > Q_0^2$ using the DGLAP evolution equations. The evolved PDFs are convoluted with $C_{i,a}$ or $\hat{\sigma}_{ab}$ to calculate theory predictions corresponding to a wide variety of data. The input parameters are then varied to minimise a global goodness-of-fit measure (χ^2).

The determination of parton distributions by global analysis has been an "industry" for more than 20 years, with regular updates as new data and new theory become available. The first NLO fit was done by the **M**artin–**R**oberts–**S**tirling group (1987), later joined by **T**horne (1998), until the retirement of **R**oberts (2005) and the addition of G.**W**. (2006). The previous "MRST" fits have recently been superseded by the "MSTW 2008" (LO, NLO and NNLO) fits [2]. The other major group is "CTEQ" (**C**oordinated **T**heoretical–**E**xperimental **P**roject on **Q**CD), and the latest public fits are CTEQ6L1 at LO [3], CTEQ6.6 at NLO [4], while a NNLO fit is still forthcoming. Other groups generally fit a more restricted range of data with fewer free parameters [5, 6, 7]. The NNPDF Collaboration [8, 9] use an interesting alternative approach to determine PDFs from a neural network parameterisation to avoid bias due to a particular functional form of the input.

2. HERA

Existing HERA data already provide one of the most important inputs to global PDF analyses, especially in providing a strong constraint on the small-x gluon and sea-quark distributions. The separate H1 and ZEUS inclusive cross-section measurements in DIS have recently been *combined* to improve accuracy [7]. The averaging procedure gives a large reduction in the correlated systematic uncertainties due to the different properties of the two detectors, leading effectively to "cross-calibration" of the systematic uncertainties between the two experiments. These new HERA combined data will prove invaluable in the next generation of global fits. A PDF fit *only* to these HERA data has already been performed [7], but with only 10 free PDF parameters used to determine the central fit compared to, for example, 28 free PDF parameters for the "MSTW 2008" fit, reflecting the incomplete flavour separation provided by fitting *only* to the HERA inclusive data.

Heavy quarks, particularly charm, can contribute a sizeable amount to the total DIS structure function F_2. There are two well-defined regions in which to calculate the heavy-quark contribution. In the *fixed flavour number scheme* (FFNS), valid for $Q^2 \lesssim m_H^2$, the heavy-quark mass (m_H) is retained in the calculation of the hard-scattering coefficient function, and there is no heavy-quark PDF which would resum $\alpha_S \ln(Q^2/m_H^2)$ terms in a similar way as for light quarks. In the *zero-mass variable flavour number scheme* (ZM-VFNS), valid for $Q^2 \gg m_H^2$, this heavy-quark PDF is introduced, and the mass dependence is neglected in the coefficient function. A *general-mass VFNS* (GM-VFNS) interpolates between these two well-defined limits, using a FFNS for $Q^2 \leq m_H^2$ and a ZM-VFNS for $Q^2 \gg m_H^2$, although there are ambiguous $\mathscr{O}(m_H^2/Q^2)$ terms in the intermediate region of $Q^2 > m_H^2$. The calculation of W and Z cross sections at the LHC is directly influenced by the treatment of heavy quarks in DIS, since the relevant sea-quark PDFs are determined from HERA data. The change from CTEQ6.1 (ZM-VFNS) to CTEQ6.5 (GM-VFNS) gave a 8% increase in $\sigma_{W,Z}$ at the LHC. The MRST group have used a GM-VFNS since 1998, but the change from MRST 2004 to MRST 2006 introduced the first precise definition of a GM-VFNS at NNLO, including in particular the (correct) discontinuities in the NNLO PDF evolution at $Q^2 = m_H^2$, leading to a 6% increase in $\sigma_{W,Z}$ at the LHC. Pre-2006 MRST *NNLO* (but not NLO) PDF sets should therefore now be considered obsolete due to the incomplete heavy-flavour treatment.

Heavy-flavour structure function data from HERA are reaching impressive precision, particularly for the charm structure function $F_2^{c\bar{c}}$, where the separate H1 and ZEUS measurements have also been combined using the same procedure as for the inclusive cross sections. For both charm and beauty structure functions, there is good agreement between the data and theoretical predictions using different varieties of GM-VFNS, with the $F_2^{c\bar{c}}$ data having some discriminating power at the lowest Q^2 values (but still $Q^2 > m_c^2$), where the GM-VFNS predictions exhibit the largest variation. The systematic uncertainty in the particular choice of GM-VFNS, and its effect on hadronic cross sections, remains to be fully quantified, but work is in progress towards achieving this goal.

The longitudinal proton structure function, F_L, was measured at HERA using data taken in the last few months of running in 2007 when the proton beam energy was lowered. The NLO and NNLO calculations tend to undershoot the HERA data at the lowest Q^2 and x values where the theory predictions are perturbatively unstable, while small-x resummation [10] aids the description. Small-x resummation will be important at the LHC, for example, in low-mass Drell–Yan production, where the fixed-order theory predictions are also seen to be perturbatively unstable.

3. Tevatron

Data from the Tevatron Run II are playing an increasing rôle in global PDF fits. The primary types of data used are the Z rapidity distributions [11, 12], included for the first time in the "MSTW 2008" fit [2], the $W \to \ell \nu$ charge asymmetry, and cross sections for inclusive jet production.

3.1 $W \to \ell \nu$ charge asymmetry

The W charge asymmetry at the Tevatron, as a function of the W rapidity (y_W), is

$$A_W(y_W) = \frac{\mathrm{d}\sigma(W^+)/\mathrm{d}y_W - \mathrm{d}\sigma(W^-)/\mathrm{d}y_W}{\mathrm{d}\sigma(W^+)/\mathrm{d}y_W + \mathrm{d}\sigma(W^-)/\mathrm{d}y_W} \approx \frac{u(x_1)d(x_2) - d(x_1)u(x_2)}{u(x_1)d(x_2) + d(x_1)u(x_2)}, \tag{3.1}$$

where $x_{1,2} = (M_W/\sqrt{s}) \exp(\pm y_W)$, constraining mainly the d/u ratio of the proton PDFs. However, experimentally, the W rapidity cannot be directly reconstructed since the longitudinal momentum of the decay neutrino is, in general, unknown. Therefore, the quantity which is traditionally measured instead is the *lepton* charge asymmetry, as a function of the lepton pseudorapidity η_ℓ, i.e.

$$A_\ell(\eta_\ell) = \frac{\mathrm{d}\sigma(\ell^+)/\mathrm{d}\eta_\ell - \mathrm{d}\sigma(\ell^-)/\mathrm{d}\eta_\ell}{\mathrm{d}\sigma(\ell^+)/\mathrm{d}\eta_\ell + \mathrm{d}\sigma(\ell^-)/\mathrm{d}\eta_\ell}. \tag{3.2}$$

Global PDF fits have previously used Tevatron Run I data on A_ℓ [13]. The MSTW 2008 fit [2] was the first to instead use Run II data [14, 15], provided in two E_T^e bins for the case of CDF data on A_e [14]. The latest DØ data on A_e [16] and A_μ [17] are badly described by current NLO PDFs, especially for $p_T^\ell > 35$ GeV, while refitting the PDFs causes tension with a number of other data sets, although this tension is reduced with modified deuteron corrections. It is not possible to describe both the DØ A_e [16] and A_μ [17] data simultaneously. The effect of NNLO corrections [18, 19] (or p_T^W-resummation, as implemented in RESBOS) is small, but acts in the right direction. CDF have recently determined A_W [20] using a new technique to obtain the neutrino's longitudinal momentum by constraining the $\ell \nu$ mass to M_W. The MSTW 2008 PDFs using VRAP [21] give a good description (better than the previous MRST 2006 PDFs) of the CDF A_W data, while modified fits to the new DØ A_ℓ data [16, 17] tend to undershoot the CDF A_W data. Before the new precise A_ℓ data can be usefully included in global PDF fits, more work is needed to qualify and resolve the apparent discrepancies between (i) CDF and DØ data, (ii) A_e and A_μ data, and (iii) data and theory.

3.2 Inclusive jet production

The Tevatron Run I data on inclusive jet production were included in global PDF fits up to MRST 2006 (and the current CTEQ6.6) as an important constraint on the high-x gluon distribution. The MSTW 2008 analysis [2] was the first to include Run II jet data [22, 23], finding a preference for a *smaller* gluon distribution at high x than that obtained with the previous Run I data. Fitting only to Run I jet data gives a bad description of Run II jet data, and vice versa, while fitting neither gives a similar description as only fitting Run II jet data. Some similar findings have been made by the CTEQ group [24], although with a little less discrepancy and change in gluon. There is therefore some apparent inconsistency between the Run I and Run II jet data, while the Run II jet data are slightly more consistent with the rest of the data included in the global fit. The final MSTW 2008 analysis therefore dropped the Run I jet data from the fit. There is only a slight change in the

gluon if the CDF Run II data obtained using the k_T jet algorithm [22] are replaced by the CDF Run II data obtained using the cone-based Midpoint jet algorithm [25]. The new smaller high-x gluon is also preferred by the DØ Run II dijet mass spectrum [26], especially at high rapidities, where the data prefer MSTW 2008 over CTEQ6.6.

The NNLO trend is similar to NLO, with the caveat that the exact NNLO jet cross sections are unavailable, so 2-loop threshold corrections are used instead. The smaller high-x gluon (and smaller α_S) in MSTW 2008 compared to MRST 2006 means that the predicted Higgs cross sections at the Tevatron are also smaller. The MSTW 2008 NNLO PDFs were used for the Tevatron exclusion results from last March [27] and last November [28], whereas previous results used MRST 2002 NNLO PDFs, which fit Tevatron Run I jet data and also had an incomplete heavy-flavour treatment.

4. LHC

It is common to determine the strong coupling α_S at the same time as the PDFs. For example, the MSTW 2008 NNLO analysis obtained $\alpha_S(M_Z^2) = 0.1171 \pm 0.0014$ from only experimental uncertainties [29], with an additional theory uncertainty ($\lesssim 0.003$), cf. the Particle Data Group world average value of $\alpha_S(M_Z^2) = 0.1176 \pm 0.002$. The same value of α_S should be used in subsequent cross-section calculations. However, since the PDFs and α_S are correlated, the uncertainty on a hadronic cross section due to both PDFs and α_S cannot simply be obtained by adding the two separate uncertainties in quadrature. A prescription has recently been developed [29] to allow consistent calculation of the combined "PDF+α_S" uncertainty on a hadronic cross section. The additional uncertainty due to α_S is particularly important for processes where multiple powers of α_S appear at lowest-order, such as Higgs production via gluon–gluon fusion or inclusive jet production, both of which enter at $\mathcal{O}(\alpha_S^2)$ at the LHC.

The W and Z total cross sections at the LHC are a potential "standard candle" for determination of the machine luminosity. The NNLO total cross sections using MSTW 2008 NNLO PDFs have a "PDF+α_S" uncertainty of around 2–3%, while the additional uncertainty from varying the renormalisation and factorisation scales is less than 1%. Dependence on other theoretical uncertainties, such as heavy-quark masses and the specific choice of GM-VFNS used in the PDF fit, is currently under study. Most uncertainties largely cancel in the W/Z and W^+/W^- ratios.

The parton luminosity function, $\partial \mathcal{L}_{ab}/\partial M_X^2$, can be interpreted as the appropriate convolution of PDFs for production of a final state with invariant mass M_X from initial-state partons a and b. It proves very useful when studying properties of hadronic cross sections, for example, the PDF uncertainty or the dependence on different LHC beam energies [30, 31].

Of course, as data begins accumulating at the LHC, precision measurements will provide further constraints on PDFs. In particular, measurement of low-mass Drell–Yan production at high rapidity by LHCb [32, 33] may extend the small-x reach of HERA, although as already noted, useful inclusion in PDF fits may require small-x resummation.

The importance of PDFs can only increase now that we have firmly entered the LHC era.

Acknowledgments

References

[1] Slides:
http://indico.cern.ch/contributionDisplay.py?contribId=8&confId=34002

[2] A. D. Martin *et al.*, *Eur. Phys. J. C* **63** (2009) 189 [arXiv:0901.0002].

[3] J. Pumplin *et al.*, *JHEP* **0207** (2002) 012 [hep-ph/0201195].

[4] P. M. Nadolsky *et al.*, *Phys. Rev. D* **78** (2008) 013004 [arXiv:0802.0007].

[5] P. Jimenez-Delgado and E. Reya, *Phys. Rev. D* **79** (2009) 074023 [arXiv:0810.4274].

[6] S. Alekhin *et al.*, *Phys. Rev. D* (to be published) [arXiv:0908.2766].

[7] H1 and ZEUS Collaborations, *JHEP* (to be published) [arXiv:0911.0884].

[8] R. D. Ball *et al.* [NNPDF Collaboration], *Nucl. Phys. B* **809** (2009) 1 [arXiv:0808.1231].

[9] R. D. Ball *et al.* [NNPDF Collaboration], *Nucl. Phys. B* **823** (2009) 195 [arXiv:0906.1958].

[10] C. D. White and R. S. Thorne, *Phys. Rev. D* **75** (2007) 034005 [hep-ph/0611204].

[11] V. M. Abazov *et al.* [DØ Collaboration], *Phys. Rev. D* **76** (2007) 012003 [hep-ex/0702025].

[12] T. Aaltonen *et al.* [CDF Collaboration], Submitted to *Phys. Rev. Lett.* [arXiv:0908.3914].

[13] F. Abe *et al.* [CDF Collaboration], *Phys. Rev. Lett.* **81** (1998) 5754 [hep-ex/9809001].

[14] D. E. Acosta *et al.* [CDF Collaboration], *Phys. Rev. D* **71** (2005) 051104 [hep-ex/0501023].

[15] V. M. Abazov *et al.* [DØ Collaboration], *Phys. Rev. D* **77** (2008) 011106 [arXiv:0709.4254].

[16] V. M. Abazov *et al.* [DØ Collaboration], *Phys. Rev. Lett.* **101** (2008) 211801 [arXiv:0807.3367].

[17] DØ Collaboration, DØ Note 5976-CONF, August 26, 2009,
http://www-d0.fnal.gov/Run2Physics/WWW/results/prelim/EW/E31/.

[18] K. Melnikov and F. Petriello, *Phys. Rev. D* **74** (2006) 114017 [hep-ph/0609070].

[19] S. Catani *et al.*, *Phys. Rev. Lett.* **103** (2009) 082001 [arXiv:0903.2120].

[20] T. Aaltonen *et al.* [CDF Collaboration], *Phys. Rev. Lett.* **102**, 181801 (2009) [arXiv:0901.2169].

[21] C. Anastasiou *et al.*, *Phys. Rev. D* **69** (2004) 094008 [hep-ph/0312266].

[22] A. Abulencia *et al.* [CDF Collaboration], *Phys. Rev. D* **75** (2007) 092006 [hep-ex/0701051].

[23] V. M. Abazov *et al.* [DØ Collaboration], *Phys. Rev. Lett.* **101** (2008) 062001 [arXiv:0802.2400].

[24] J. Pumplin *et al.*, *Phys. Rev. D* **80** (2009) 014019 [arXiv:0904.2424].

[25] T. Aaltonen *et al.* [CDF Collaboration], *Phys. Rev. D* **78** (2008) 052006 [arXiv:0807.2204].

[26] DØ Collaboration, DØ Note 5919-CONF, April 29, 2009,
http://www-d0.fnal.gov/Run2Physics/WWW/results/prelim/QCD/Q14/.

[27] CDF and DØ Collaborations, arXiv:0903.4001.

[28] CDF and DØ Collaborations, arXiv:0911.3930.

[29] A. D. Martin *et al.*, *Eur. Phys. J. C* **64** (2009) 653 [arXiv:0905.3531].

[30] http://projects.hepforge.org/mstwpdf/plots/plots.html

[31] C. Quigg, *LHC Physics Potential vs. Energy*, arXiv:0908.3660.

[32] R. S. Thorne *et al.*, *Parton Distributions and QCD at LHCb*, arXiv:0808.1847.

[33] R. McNulty [LHCb Collaboration], *Potential PDF sensitivity at LHCb*, arXiv:0810.2550.