Progress in **P**arton **D**istribution **F**unctions

Graeme Watt

CERN PH-TH

CERN Theory Group Retreat
Les Houches Centre de Physique
4th November 2009

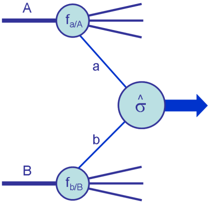
Introduction

- Protons are not elementary particles: made of **partons**.
  $\Rightarrow$ **P**arton **D**istribution **F**unctions (**PDF**s) essential to relate theory to experiment at the LHC (and Tevatron, HERA, ...).
- $f_{a/A}(x, Q^2)$ gives *number density* of partons a in hadron A with momentum fraction x at a hard scale $Q^2 \gg \Lambda^2_{\rm QCD}$.

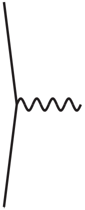
$$\sigma_{AB} = \sum_{a,b=q,g} \int_0^1 \mathrm{d}x_a \int_0^1 \mathrm{d}x_b \; f_{a/A}(x_a, Q^2) \; f_{b/B}(x_b, Q^2) \; \hat{\sigma}_{ab}$$

Outline of talk:

- Sketch of standard pQCD framework
- Determination of PDFs via global fits
- Recent developments from fitting groups
- Implications for Tevatron and LHC



Diagrammatic interpretation of collinear factorisation

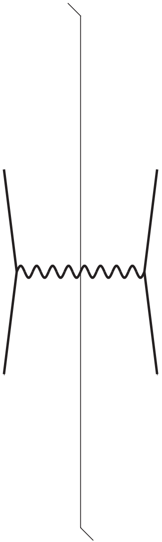


- Drell–Yan production at LO: $q\bar{q} \to V = W/Z/\gamma^*$
- Cut diagram: $|\mathcal{M}|^2 = \mathcal{M}\mathcal{M}^*$
- Large logarithm from collinear gluon emission: $\int_{k_0^2}^{Q^2}(\mathrm{d}k_T^2/k_T^2)\,\frac{\alpha_S}{2\pi}\,P_{q\leftarrow q}(z)$
- Similar collinear logs from other parton splittings.
- DGLAP evolution equation:

$$\frac{\partial f_{a/p}}{\partial \ln Q^2} = \frac{\alpha_S}{2\pi}\sum_{a'=q,g} P_{a\leftarrow a'} \otimes f_{a'/p}$$

- $f_{a/p}(x, Q_0^2) \Rightarrow f_{a/p}(x, Q^2)$

Diagrammatic interpretation of collinear factorisation

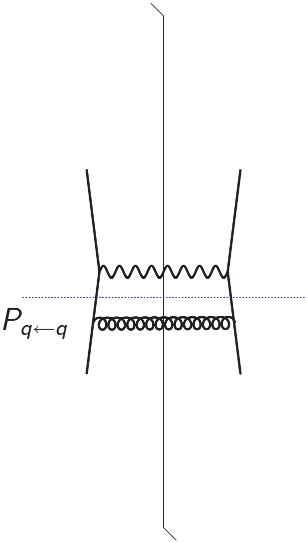


- Drell–Yan production at LO: $q\bar{q} \to V = W/Z/\gamma^*$
- Cut diagram: $|\mathcal{M}|^2 = \mathcal{M}\mathcal{M}^*$
- Large logarithm from collinear gluon emission: $\int_{k_0^2}^{Q^2} (\mathrm{d}k_T^2/k_T^2)\, \frac{\alpha_S}{2\pi}\, P_{q\leftarrow q}(z)$
- Similar collinear logs from other parton splittings.
- DGLAP evolution equation:

$$\frac{\partial f_{a/p}}{\partial \ln Q^2} = \frac{\alpha_S}{2\pi} \sum_{a'=q,g} P_{a\leftarrow a'} \otimes f_{a'/p}$$

- $f_{a/p}(x, Q_0^2) \Rightarrow f_{a/p}(x, Q^2)$

Diagrammatic interpretation of collinear factorisation

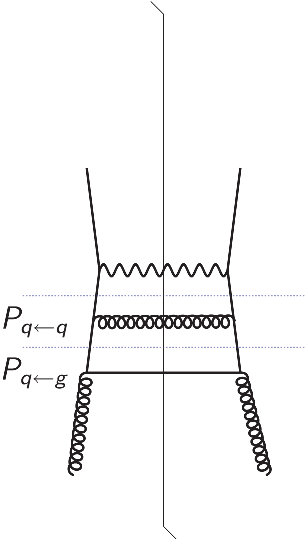


- Drell–Yan production at LO: $q\bar{q} \to V = W/Z/\gamma^*$
- Cut diagram: $|\mathcal{M}|^2 = \mathcal{M}\mathcal{M}^*$
- Large logarithm from collinear gluon emission: $\int_{k_0^2}^{Q^2}(\mathrm{d}k_T^2/k_T^2)\,\frac{\alpha_S}{2\pi}\,P_{q\leftarrow q}(z)$
- Similar collinear logs from other parton splittings.
- DGLAP evolution equation:

$$\frac{\partial f_{a/p}}{\partial \ln Q^2} = \frac{\alpha_S}{2\pi}\sum_{a'=q,g} P_{a\leftarrow a'} \otimes f_{a'/p}$$

- $f_{a/p}(x, Q_0^2) \Rightarrow f_{a/p}(x, Q^2)$

Diagrammatic interpretation of collinear factorisation

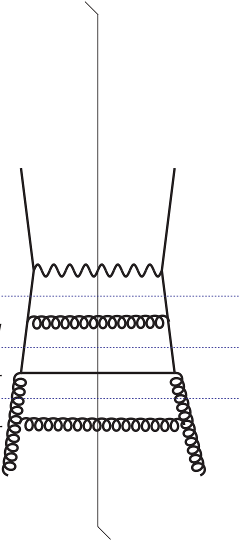


- Drell–Yan production at LO: $q\bar{q} \to V = W/Z/\gamma^*$
- Cut diagram: $|\mathcal{M}|^2 = \mathcal{M}\mathcal{M}^*$
- Large logarithm from collinear gluon emission: $\int_{k_0^2}^{Q^2} (\mathrm{d}k_T^2/k_T^2)\, \frac{\alpha_S}{2\pi}\, P_{q\leftarrow q}(z)$
- Similar collinear logs from other parton splittings.
- DGLAP evolution equation:

$$\frac{\partial f_{a/p}}{\partial \ln Q^2} = \frac{\alpha_S}{2\pi} \sum_{a'=q,g} P_{a\leftarrow a'} \otimes f_{a'/p}$$

- $f_{a/p}(x, Q_0^2) \Rightarrow f_{a/p}(x, Q^2)$

Diagrammatic interpretation of collinear factorisation

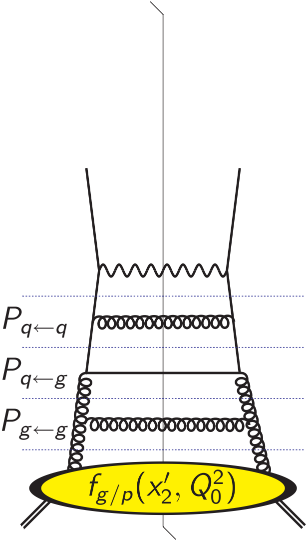


- Drell–Yan production at LO: $q\bar{q} \to V = W/Z/\gamma^*$
- Cut diagram: $|\mathcal{M}|^2 = \mathcal{M}\mathcal{M}^*$
- Large logarithm from collinear gluon emission: $\int_{k_0^2}^{Q^2} (\mathrm{d}k_T^2/k_T^2)\,\frac{\alpha_S}{2\pi}\,P_{q\leftarrow q}(z)$
- Similar collinear logs from other parton splittings.
- DGLAP evolution equation:

$$\frac{\partial f_{a/p}}{\partial \ln Q^2} = \frac{\alpha_S}{2\pi}\sum_{a'=q,g} P_{a\leftarrow a'} \otimes f_{a'/p}$$

- $f_{a/p}(x, Q_0^2) \Rightarrow f_{a/p}(x, Q^2)$

Diagrammatic interpretation of collinear factorisation

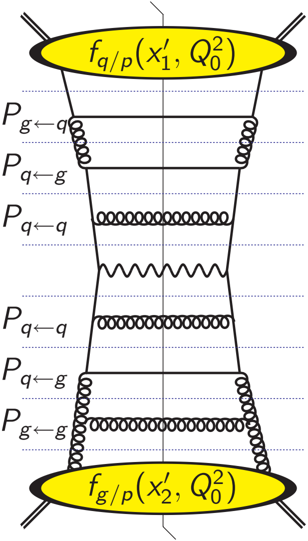


- Drell–Yan production at LO: $q\bar{q} \to V = W/Z/\gamma^*$
- Cut diagram: $|\mathcal{M}|^2 = \mathcal{M}\mathcal{M}^*$
- Large logarithm from collinear gluon emission: $\int_{k_0^2}^{Q^2} (\mathrm{d}k_T^2/k_T^2)\, \frac{\alpha_S}{2\pi}\, P_{q\leftarrow q}(z)$
- Similar collinear logs from other parton splittings.
- DGLAP evolution equation:

$$\frac{\partial f_{a/p}}{\partial \ln Q^2} = \frac{\alpha_S}{2\pi} \sum_{a'=q,g} P_{a\leftarrow a'} \otimes f_{a'/p}$$

- $f_{a/p}(x, Q_0^2) \Rightarrow f_{a/p}(x, Q^2)$

Diagrammatic interpretation of collinear factorisation

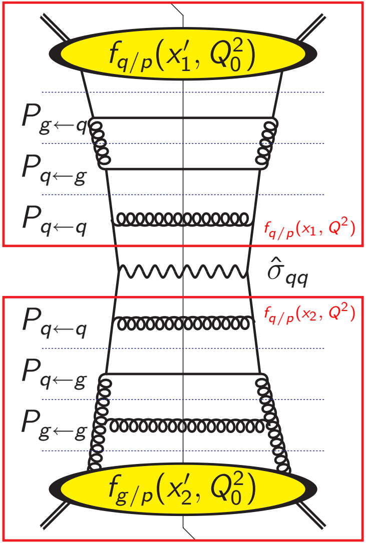


- Drell–Yan production at LO: $q\bar{q} \to V = W/Z/\gamma^*$
- Cut diagram: $|\mathcal{M}|^2 = \mathcal{M}\mathcal{M}^*$
- Large logarithm from collinear gluon emission: $\int_{k_0^2}^{Q^2} (\mathrm{d}k_T^2/k_T^2)\,\frac{\alpha_S}{2\pi}\,P_{q\leftarrow q}(z)$
- Similar collinear logs from other parton splittings.
- DGLAP evolution equation:

$$\frac{\partial f_{a/p}}{\partial \ln Q^2} = \frac{\alpha_S}{2\pi} \sum_{a'=q,g} P_{a\leftarrow a'} \otimes f_{a'/p}$$

- $f_{a/p}(x, Q_0^2) \Rightarrow f_{a/p}(x, Q^2)$

Diagrammatic interpretation of collinear factorisation



- Drell–Yan production at LO: $q\bar{q} \to V = W/Z/\gamma^*$
- Cut diagram: $|\mathcal{M}|^2 = \mathcal{M}\mathcal{M}^*$
- Large logarithm from collinear gluon emission: $\int_{k_0^2}^{Q^2} (\mathrm{d}k_T^2/k_T^2)\, \frac{\alpha_S}{2\pi}\, P_{q\leftarrow q}(z)$
- Similar collinear logs from other parton splittings.
- DGLAP evolution equation:

$$\frac{\partial f_{a/p}}{\partial \ln Q^2} = \frac{\alpha_S}{2\pi} \sum_{a'=q,g} P_{a\leftarrow a'} \otimes f_{a'/p}$$

- $f_{a/p}(x, Q_0^2) \Rightarrow f_{a/p}(x, Q^2)$

Fixed-order collinear factorisation at hadron colliders

- The "standard" pQCD framework: holds up to formally power-suppressed ("higher-twist") terms $\mathcal{O}(\Lambda^2_{\rm QCD}/Q^2)$.
- Expand $\hat{\sigma}_{ab}$, $P_{aa'}$ and β as perturbative series in α_S ($\mu_R = \mu_F = Q$).

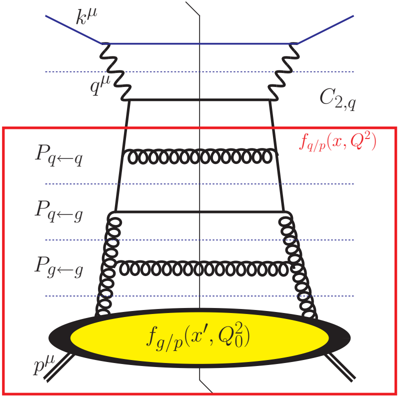
$$\sigma_{AB} = \sum_{a,b=q,g} \left[\hat{\sigma}^{\rm LO}_{ab} + \alpha_S(Q^2)\hat{\sigma}^{\rm NLO}_{ab} + \ldots\right] \otimes f_{a/A}(x_a, Q^2) \otimes f_{b/B}(x_b, Q^2)$$

PDF evolution: $$\frac{\partial f_{a/A}}{\partial \ln Q^2} = \frac{\alpha_S}{2\pi} \sum_{a'=q,g} \left[P^{\rm LO}_{aa'} + \alpha_S P^{\rm NLO}_{aa'} + \ldots\right] \otimes f_{a'/A}$$

α_S evolution: $$\frac{\partial \alpha_S}{\partial \ln Q^2} = -\beta^{\rm LO}\alpha_S^2 - \beta^{\rm NLO}\alpha_S^3 - \ldots$$

Need to extract input values $f_{a/A}(x, Q_0^2)$ and $\alpha_S(M_Z^2)$ from data.
(N.B. Theory not perfect, e.g. resummation of $\ln(1/x)$ terms needed?)

Structure functions in deep-inelastic scattering (DIS)

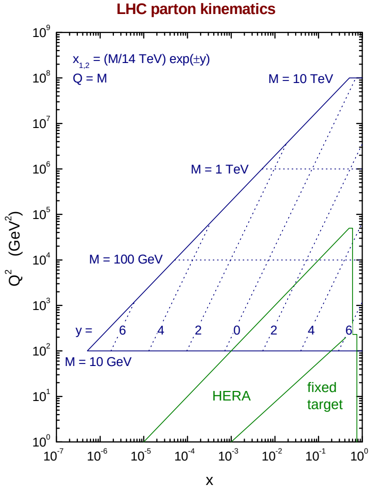


Kinematic variables:

$$Q^2 = -q^2 > 0$$
$$W^2 = (q+p)^2$$
$$s = (k+p)^2$$
$$x_{\rm Bj} = \frac{Q^2}{2p\cdot q} \simeq \frac{Q^2}{Q^2+W^2}$$
$$y = \frac{q\cdot p}{k\cdot p} \simeq \frac{Q^2}{x_{\rm Bj}\, s}$$

$$F_i\left(x_{\rm Bj},Q^2\right) = \sum_{a=q,g}\int_{x_{\rm Bj}}^1 {\rm d}z\; C_{i,a}(z)\,\frac{x_{\rm Bj}}{z} f_{a/p}\left(\frac{x_{\rm Bj}}{z},Q^2\right)$$
$$\equiv \sum_{a=q,g} C_{i,a}\otimes f_{a/p}, \qquad C_{i,a} = C_{i,a}^{\rm LO} + \alpha_S\, C_{i,a}^{\rm NLO} + \ldots$$

From HERA *et al.* to the LHC



- PDFs are **universal**.
- Fit existing data from HERA and fixed-target experiments, together with Tevatron data.
- HERA *ep* (H1, ZEUS).
- Fixed-target experiments:
  ℓp, ℓd
  (BCDMS, NMC, E665, SLAC),
  νN
  (CCFR, NuTeV, CHORUS),
  pp, *pd* (E866/NuSea).
- Tevatron $p\bar{p}$ (CDF, DØ).
- DGLAP evolution gives PDFs at higher Q^2 for LHC.

Paradigm for PDF determination by "global analysis"

1. **Parameterise** the x dependence for each flavour $a = q, g$ at the input scale $Q_0^2 \sim 1\ \mathrm{GeV}^2$ in some flexible form, e.g.

$$xf_{a/p}(x, Q_0^2) = A_a x^{\Delta_a} (1-x)^{\eta_a} (1 + \epsilon_a \sqrt{x} + \gamma_a x),$$

subject to number- and momentum-sum rule constraints.

2. **Evolve** the PDFs to higher scales $Q^2 > Q_0^2$ using the DGLAP (Dokshitzer–Gribov–Lipatov–Altarelli–Parisi) evolution equations.
3. **Convolute** the evolved PDFs with $C_{i,a}$ and $\hat{\sigma}_{ab}$ to calculate theory predictions corresponding to a wide variety of data.
4. **Vary** the input parameters $\{A_a, \Delta_a, \eta_a, \epsilon_a, \gamma_a, \ldots\}$ to minimise

$$\chi^2 = \sum_{i=1}^{N_{\rm pts.}} \left(\frac{\mathrm{Data}_i - \mathrm{Theory}_i}{\mathrm{Error}_i} \right)^2$$

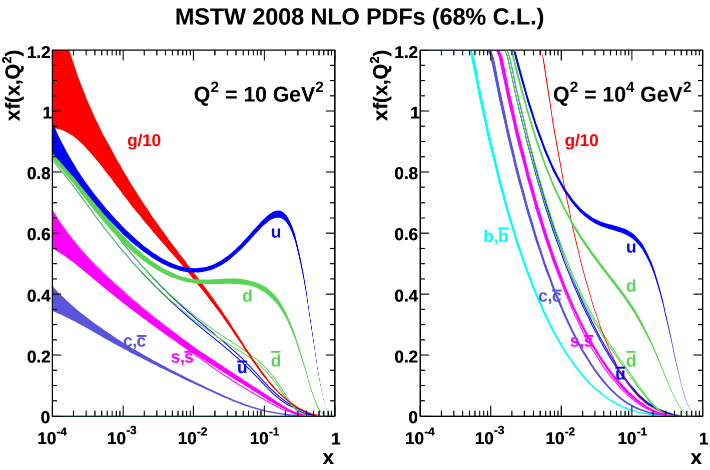
Determination of parton distributions by global analysis

An "industry" for more than 20 years.

Regular updates as new data and theory become available.

1. First NLO fit: **M**artin+**R**oberts+**S**tirling ('88) + **T**horne ('98).
   Recently, "**MSTW**" = **M**R**ST** − Roberts + G.**W**.
   {MRST 2001 LO, MRST 2004 NLO, MRST 2006 NNLO}
   → MSTW 2008 LO, NLO, NNLO fits [arXiv:0901.0002]
2. Other major group: "**CTEQ**" = **C**oordinated **T**heoretical–**E**xperimental Project on **Q**CD.
   - CTEQ6L1 LO [hep-ph/0201195]
   - CTEQ6.6 NLO [arXiv:0802.0007]
   - CTEQ NNLO?
3. Other groups fitting a restricted range of data with fewer free parameters: S. Alekhin *et al.*, HERA experiments (H1, ZEUS).
4. NNPDF Collaboration: see later.

Example of PDFs obtained from global analysis



- Error bands shown are obtained from propagation of experimental uncertainties on the fitted data points.

Criteria for choice of tolerance $T = \sqrt{\Delta\chi^2_{\rm global}}$

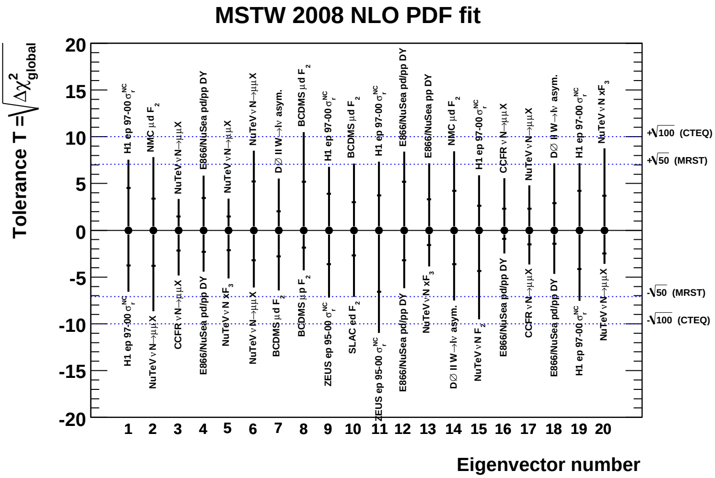
Parameter-fitting criterion

- $T^2 = 1$ for 68% (1-σ) C.L., $T^2 = 2.71$ for 90% C.L.
- **In practice:** minor inconsistencies between fitted data sets, and unknown experimental and theoretical uncertainties, so not appropriate for global PDF analysis.

Hypothesis-testing criterion (proposed by CTEQ)

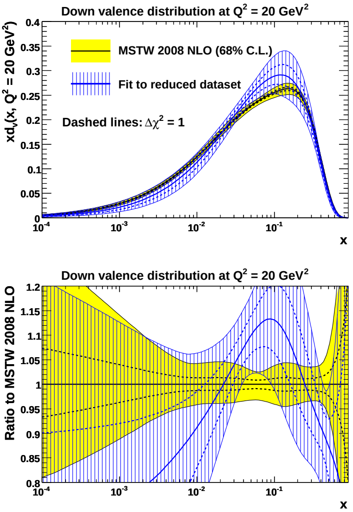
- Much weaker: treat PDF sets obtained from eigenvectors of covariance matrix as alternative hypotheses.
- Determine T^2 from the criterion that each data set should be described within its 90% C.L. limit. Very roughly, a "good" fit has $\chi^2 \simeq N_{\rm pts.} \pm \sqrt{2N_{\rm pts.}}$ for each data set.
- **CTEQ:** $T^2 = 100$ for 90% C.L. limit, **MRST:** $T^2 = 50$.

Dynamic tolerance: different for each eigenvector



- Outer (inner) error bars give tolerance for 90% (68%) C.L.

Test of dynamic tolerance: fit to reduced dataset



- Fit to reduced dataset comprising 589 DIS data points, cf. 2699 data points in global fit.
- Errors given by $T^2 = 1$ don't overlap $\Rightarrow$ inconsistent data sets included in global fit.
- **Dynamic tolerance** $T^2 > 1$ accommodates mildly inconsistent data sets.
- **Issues:** $T^2 > 1$ not statistically rigorous, parameterisation dependence?

Alternative approach: NNPDF Collaboration

NNPDF Collaboration: R. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. Latorre, A. Piccione, J. Rojo, M. Ubiali

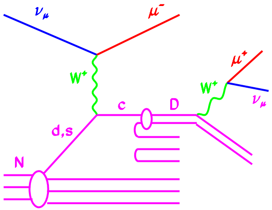
MSTW approach [arXiv:0901.0002] (CTEQ similar)

Parameterisation	$xf_{a/p} \sim A_a x^{\Delta_a}(1-x)^{\eta_a}(1+\epsilon_a\sqrt{x}+\gamma_a x)$
Minimisation	Non-linear least-squares (Marquardt method)
Error propagation	Hessian method with dynamical tolerance
Application	Use best-fit and 40 eigenvector PDF sets

NNPDF approach [arXiv:0808.1231]

Parameterisation	Neural network (37 free parameters per PDF)
Minimisation	Genetic algorithm (stop before overlearning)
Error propagation	Generate $N_{\rm rep} \sim \mathcal{O}(1000)$ MC data replicas
Application	Calculate average and s.d. over $N_{\rm rep}$ PDF sets

NuTeV/CCFR dimuon cross sections and strangeness



$$\frac{\mathrm{d}\sigma}{\mathrm{d}x\mathrm{d}y}(\nu_\mu N \to \mu^+\mu^- X) \propto \frac{\mathrm{d}\sigma}{\mathrm{d}x\mathrm{d}y}(\nu_\mu N \to \mu^- c\, X)$$

- ν_μ and $\bar{\nu}_\mu$ cross sections constrain s and $\bar{s}$, respectively, for $0.01 \lesssim x \lesssim 0.2$.

- Can **relax assumption** made in previous fits that

$$s(x,Q_0^2) = \bar{s}(x,Q_0^2) = \frac{\kappa}{2}\left[\bar{u}(x,Q_0^2) + \bar{d}(x,Q_0^2)\right], \text{ with } \kappa \approx 0.5.$$

- MSTW **parameterise** at input scale of $Q_0^2 = 1$ GeV^2 in the form:

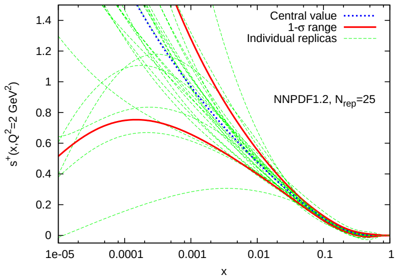
$$xs^+(x,Q_0^2) \equiv xs(x,Q_0^2) + x\bar{s}(x,Q_0^2) = A_+\,(1-x)^{\eta_+}\,xS(x,Q_0^2),$$
$$xs^-(x,Q_0^2) \equiv xs(x,Q_0^2) - x\bar{s}(x,Q_0^2) = A_-\,x^{0.2}(1-x)^{\eta_-}(1-x/x_0).$$

- x_0 fixed by zero strangeness: $\int_0^1 \mathrm{d}x \left[s(x,Q_0^2) - \bar{s}(x,Q_0^2)\right] = 0.$

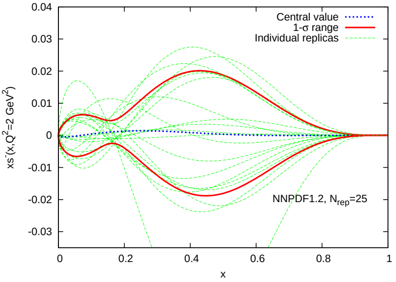
First fits from NNPDF Collaboration

- NNPDF1.0: fit only DIS structure function data. Fix $s = \bar{s} = (\bar{u} + \bar{d})/4$ at $Q_0^2 = 2\ \text{GeV}^2$.
- NNPDF1.1: free strangeness but no νN dimuon data.
- NNPDF1.2: free strangeness and add νN dimuon data.

$s^+ \equiv s + \bar{s}$ at $Q^2 = 2\ \text{GeV}^2$:



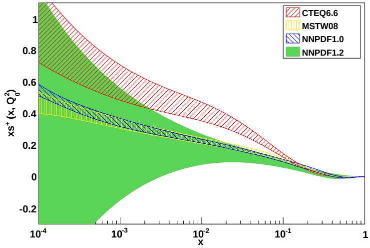
$xs^- \equiv xs - x\bar{s}$ at $Q^2 = 2\ \text{GeV}^2$:



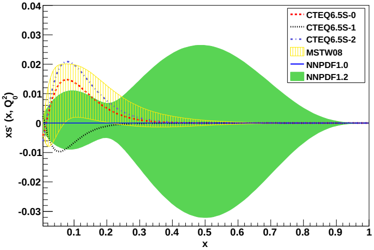
- Data only constrain $0.01 \lesssim x \lesssim 0.2$.

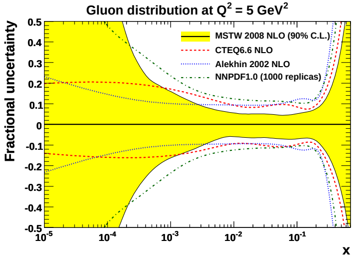
NNPDFs compared to “standard” PDFs

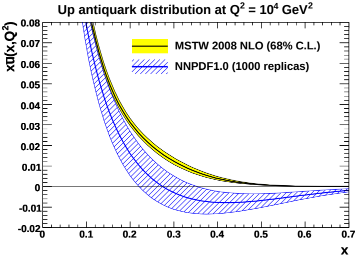
$xs^+ \equiv xs + x\bar{s}$ at $Q^2 = 2$ GeV^2:



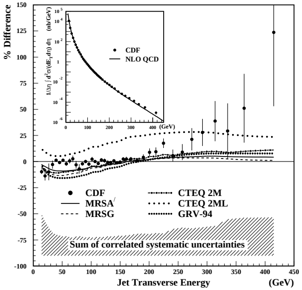
$xs^- \equiv xs - x\bar{s}$ at $Q^2 = 2$ GeV^2:

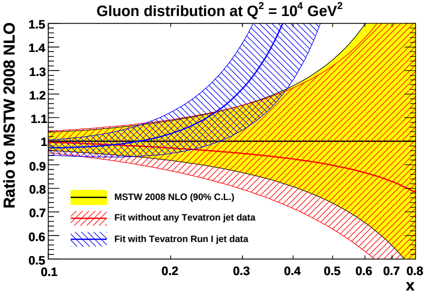






Impact of Tevatron Run II inclusive jet production data

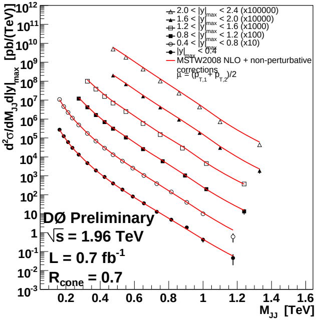




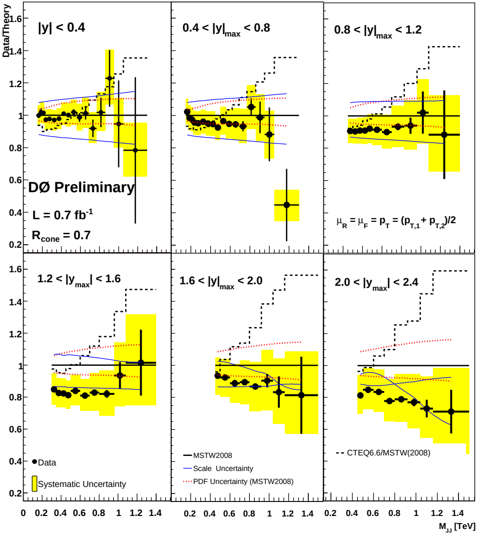
- Initial Tevatron Run I jet data showed an **excess** at high E_T, later **accommodated** by refitting gluon distribution.
- Run I data included in recent PDF fits up to MRST 2006 (and current CTEQ6.6).
- MSTW 2008 is first PDF fit to include Run **II** data: preference for **smaller** gluon distribution at high x.

Description of DØ dijet mass spectrum

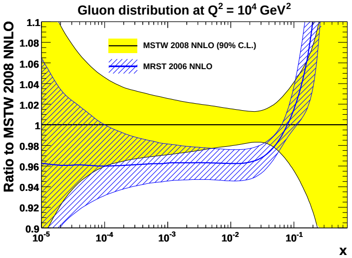
[DØ Note 5919-CONF]



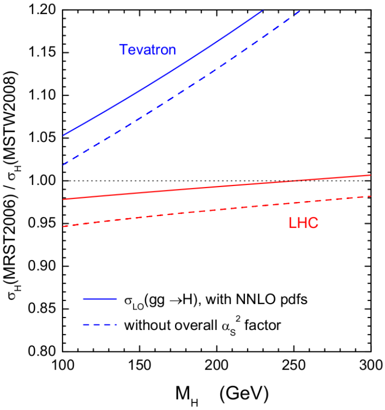
- Data favour **less** gluon at high x (MSTW 2008 over CTEQ6.6).



Implications of new PDFs for Higgs cross sections

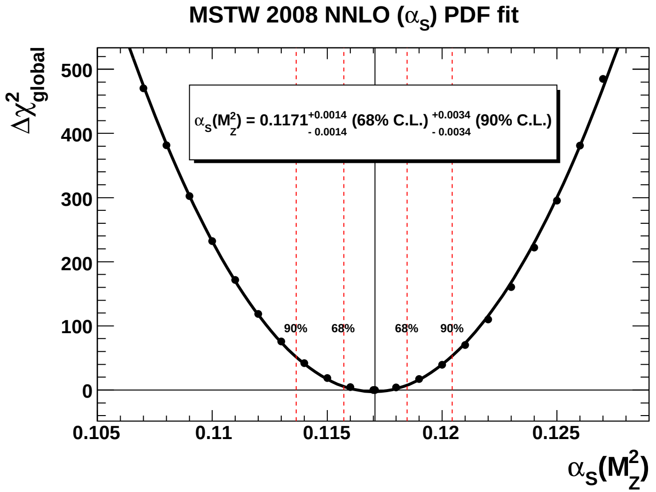


- NNLO trend similar to NLO: **smaller** 2008 gluon at high x, larger 2008 gluon at low x (momentum sum rule).
- $\alpha_S(M_Z^2) = 0.1191$ (2006) $\rightarrow 0.1171$ (MSTW 2008)



- Higgs cross sections **smaller** at Tevatron with 2008 PDFs.
- Used in Tevatron exclusion results (March 2009).

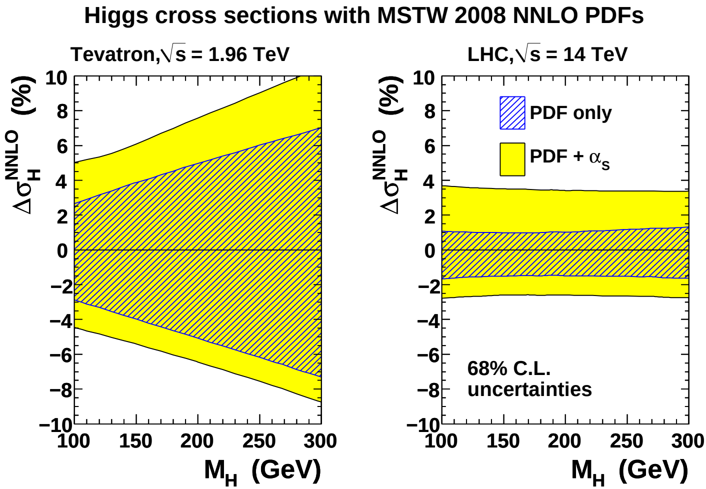
$\Delta\chi^2_{\rm global}$ as a function of $\alpha_S(M_Z^2)$ for the NNLO global fit



- Additional theory uncertainty ($\lesssim |\mathrm{NNLO}-\mathrm{NLO}| = 0.003$).
- cf. PDG world average value of $\alpha_S(M_Z^2) = 0.1176 \pm 0.002$.

Impact of α_S on SM Higgs uncertainty versus M_H

- **Correlation** between PDF and α_S uncertainties in cross section calculations [MSTW, arXiv:0905.3531].

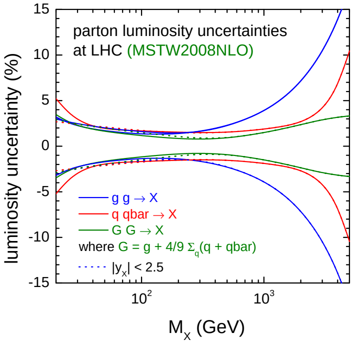


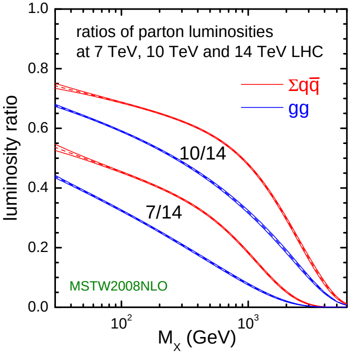
Parton luminosity functions

If $\hat{\sigma}_{ab} = C_X\,\delta\left(\hat{s} - M_X^2\right)$, with $\hat{s} = x_a\,x_b\,s$, then

$$\sigma_{AB} = \sum_{a,b} \int_0^1 \mathrm{d}x_a \int_0^1 \mathrm{d}x_b\; f_{a/A}(x_a, M_X^2)\; f_{b/B}(x_b, M_X^2)\; \hat{\sigma}_{ab} = C_X\,\frac{\partial \mathcal{L}_{ab}}{\partial M_X^2}$$

$$\boxed{\frac{\partial \mathcal{L}_{ab}}{\partial M_X^2} = \int_\tau^1 \frac{\mathrm{d}x}{x}\; f_{a/A}(x, M_X^2)\; f_{b/B}(\tau/x, M_X^2), \quad \tau = \frac{M_X^2}{s}}$$





Summary

- **P**arton **D**istribution **F**unctions (**PDF**s) are a non-negotiable input to all theory predictions at hadron colliders.
- **NNPDF** approach is a promising alternative to the **MSTW**/**CTEQ** approach: fully global fit expected soon.
- **Tevatron Run II jets** prefer smaller high-x gluon than Run I: impact on Higgs cross sections at Tevatron.
- Now possible to consistently calculate combined "PDF+α_S" uncertainty on hadronic cross sections.
- **Parton luminosities** are a simple way to understand basic properties of hadronic cross sections.