

# PARTON DISTRIBUTION FUNCTIONS

- Global analysis
- Practical applications

Pavel Nadolsky

Department of Physics  
Southern Methodist University (Dallas, TX)

Lecture 1

# Objectives of these lectures

- introduce the basic methods for determination of PDFs from hadronic scattering data
- convey the richness of ideas encountered in the PDF analysis – contributed by diverse branches of theory, experiment, and mathematics
- discuss how our knowledge of PDFs affects practical applications

Selection of the topics is far from complete – complementary material can be found in excellent lectures on PDFs by A. Cooper-Sarkar, W. Giele, J. Owens, A. Martin, W. Melnitchouk, D. Stump at recent CTEQ schools

# Theoretical essentials

# Parton distribution functions $f_{a/p}(x, Q) \dots$

... are indispensable in computations of inclusive hadronic reactions

- key functions describing the nucleon structure in QCD

- needed for a variety of new physics searches at the LHC and in other experiments

... describe probabilities for finding partons inside parent hadrons

... arise as nonperturbative functions in QCD factorization

... are universal – independent of the hard-scattering process

... cannot be computed systematically

... obey perturbative evolution (DGLAP) equations

... are determined from a global fit to hadron scattering data

# Basic definitions

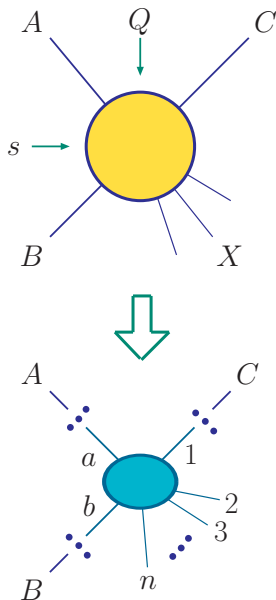
- **Partons** are weakly bound constituents of hadrons with small typical size ( $r \ll r_{nucleon} \approx 1 \text{ fm}$ )

*(Feynman; Bjorken, Paschos - 1969)*

- ▶ assumed to be pointlike at present

## Basic definitions

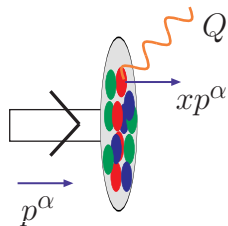
- Partons are most easily detected in **inclusive** hadronic scattering  
 $A + B \rightarrow C + X$  at large collision energy  $\sqrt{s} \gg 1 \text{ GeV}$ , with typical energy transfer  $Q$  of order  $\sqrt{s}$
- Such scattering is dominated by **rare independent collisions**  
 $a + b \rightarrow 1 + 2 + \dots + n$  of a parton  $a$  from  $A$  on a parton  $b$  from  $B$ , proceeding through **perturbative** QCD and electroweak interactions



## Basic definitions

In the simplest (leading-order) interpretation, the PDF  $f_{a/p}(x, Q)$  is a probability for finding a parton  $a$  with 4-momentum  $xp^\alpha$  in a proton with 4-momentum  $p^\alpha$

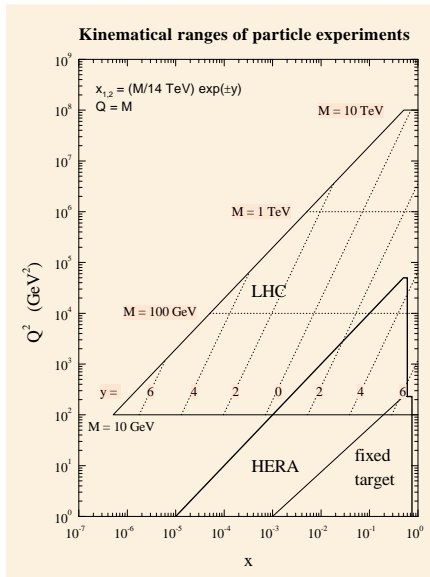
$f_{a/p}(x, Q)$  depends on **nonperturbative** QCD interactions



# Basic definitions

Modern measurements need to know the PDFs with relative accuracy of 1-5% in the whole range  $x \gtrsim 10^{-5}$  and  $1 \lesssim Q \lesssim 10^3$  GeV

Tremendous progress has been made in studying PDFs in this region; yet small  $x$  ( $x \lesssim 10^{-3}$ ) and large  $x$  ( $x > 0.3$ ) are still insufficiently understood





## PDFs and QCD factorization

According to QCD factorization theorems, typical cross sections (e.g., for vector boson production  $p(k_1)p(k_2) \rightarrow [V(q) \rightarrow \ell(k_3)\bar{\ell}(k_4)] X$ ) take the form

$$\sigma_{pp \rightarrow \ell\bar{\ell}X} = \sum_{a,b=q,\bar{q},g} \int_0^1 d\xi_1 \int_0^1 d\xi_2 \hat{\sigma}_{ab \rightarrow V \rightarrow \ell\bar{\ell}} \left( \frac{x_1}{\xi_1}, \frac{x_2}{\xi_2}; \frac{Q}{\mu} \right) f_{a/p}(\xi_1, \mu) f_{b/p}(\xi_2, \mu) + \mathcal{O}(\Lambda_{QCD}^2/Q^2)$$

■  $\hat{\sigma}_{ab \rightarrow V \rightarrow \ell\bar{\ell}}$  is the **hard-scattering cross section**

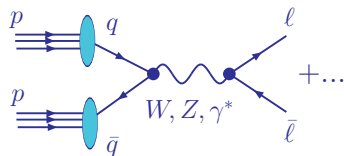
■  $f_{a/p}(\xi, \mu)$  are the **PDFs**

■  $Q^2 = (k_3 + k_4)^2$ ,  $x_{1,2} = (Q/\sqrt{s}) e^{\pm y_V}$  — measurable quantities

■  $\xi_1, \xi_2$  are partonic momentum fractions (integrated over)

■  $\mu$  is a factorization scale (=renormalization scale from now on)

■ Factorization holds up to terms of order  $\Lambda_{QCD}^2/Q^2$



## PDFs and QCD factorization

According to QCD factorization theorems, typical cross sections (e.g., for vector boson production  $p(k_1)p(k_2) \rightarrow [V(q) \rightarrow \ell(k_3)\bar{\ell}(k_4)] X$ ) take the form

$$\sigma_{pp \rightarrow \ell\bar{\ell}X} = \sum_{a,b=q,\bar{q},g} \int_0^1 d\xi_1 \int_0^1 d\xi_2 \hat{\sigma}_{ab \rightarrow V \rightarrow \ell\bar{\ell}} \left( \frac{x_1}{\xi_1}, \frac{x_2}{\xi_2}; \frac{Q}{\mu} \right) f_{a/p}(\xi_1, \mu) f_{b/p}(\xi_2, \mu) + \mathcal{O}(\Lambda_{QCD}^2/Q^2)$$

### Purpose of this arrangement:

- Subtract large collinear logarithms  $\alpha_s^n \ln^k(Q^2/m_q^2)$  from  $\hat{\sigma}$
- Resum them in  $f_{a/p}(\xi, \mu)$  to all orders of  $\alpha_s$

## Operator definitions for PDFs

To all orders in  $\alpha_s$ , PDFs are **defined** as matrix elements of certain correlator functions:

$$f_{q/p}(x, \mu) = \frac{1}{4\pi} \int_{-\infty}^{\infty} dy^- e^{iy^- p^+} \langle p | \bar{\psi}_q(0, y^-, \vec{0}_T) \gamma^+ \psi_q(0, 0, \vec{0}_T) | p \rangle, \text{ etc.}$$

Several types of definitions, or **factorization schemes** ( $\overline{MS}$ , DIS, etc.), exist

They all correspond to the probability density for finding  $a$  in  $p$  at LO; they differ at NLO and beyond

To prove factorization, one must show that  $f_{a/p}(x, \mu)$  correctly captures higher-order contributions for the considered observable

This condition can be violated for multi-scale observables (e.g., DIS or Drell-Yan process at  $x \sim Q/\sqrt{s} \ll 1$ )

## Operator definitions for PDFs

To all orders in  $\alpha_s$ , PDFs are **defined** as matrix elements of certain correlator functions:

$$f_{q/p}(x, \mu) = \frac{1}{4\pi} \int_{-\infty}^{\infty} dy^- e^{iy^- p^+} \langle p | \bar{\psi}_q(0, y^-, \vec{0}_T) \gamma^+ \psi_q(0, 0, \vec{0}_T) | p \rangle, \text{ etc.}$$

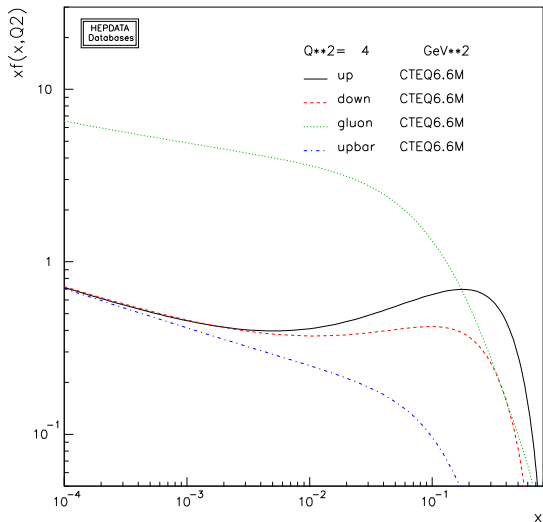
The exact form of  $f_{a/p}$  is not known; but its  $\mu$  dependence is described by **Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP)** equations:

$$\mu \frac{df_{i/p}(x, \mu)}{d\mu} = \sum_{j=g,u,\bar{u},d,\bar{d},\dots} \int_x^1 \frac{dy}{y} P_{i/j} \left( \frac{x}{y}, \alpha_s(\mu) \right) f_{j/p}(y, \mu)$$

$P_{i/j}$  are probabilities for  $j \rightarrow ik$  collinear splittings;  
are known to order  $\alpha_s^3$  (NNLO):

$$P_{i/j}(x, \alpha_s) = \alpha_s P_{i/j}^{(1)}(x) + \alpha_s^2 P_{i/j}^{(2)}(x) + \alpha_s^3 P_{i/j}^{(3)}(x) + \dots$$

# Example of DGLAP evolution



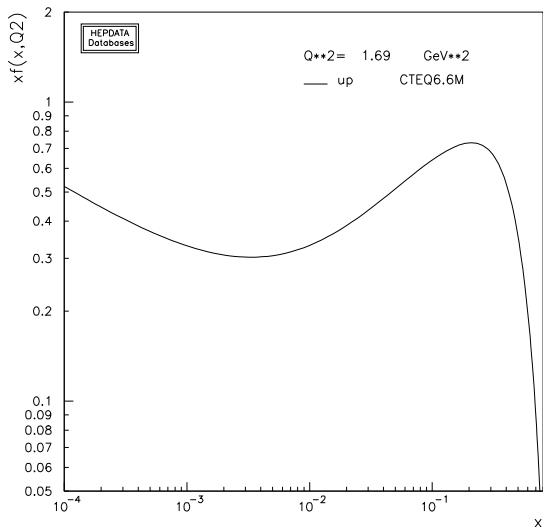
Durham PDF plotter, <http://durpdg.dur.ac.uk/hepdata/pdf3.html>

Compare  $\mu$  dependence of  $u$  quark PDF and the gluon PDF

The  $u$ ,  $d$  PDFs have a characteristic bump at  $x \sim 1/3$  – reminiscent of early valence quark models of the proton structure

The PDFs rise rapidly at  $x < 0.1$  as a consequence of perturbative evolution

# Example of DGLAP evolution

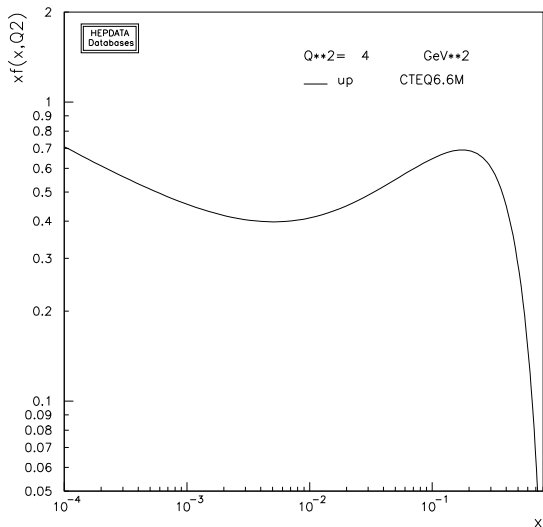


As  $Q$  increases, it becomes more likely that a high- $x$  parton loses some momentum through QCD radiation

$\Rightarrow u(x, Q)$  reduces at  $x \gtrsim 0.1$ , increases at  $x \lesssim 0.1$

Durham PDF plotter, <http://durpdg.dur.ac.uk/hepdata/pdf3.html>

# Example of DGLAP evolution

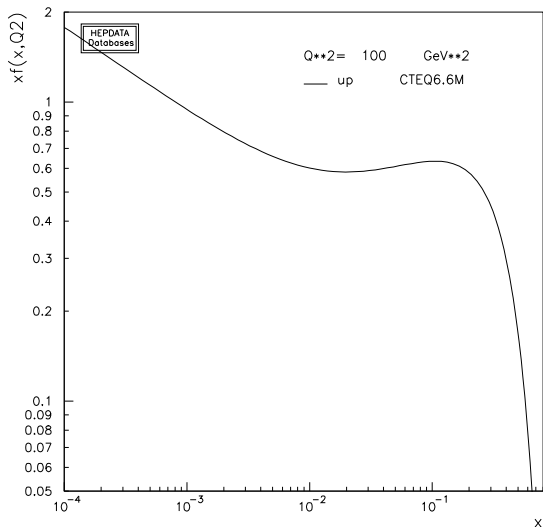


As  $Q$  increases, it becomes more likely that a high- $x$  parton loses some momentum through QCD radiation

$\Rightarrow u(x, Q)$  reduces at  $x \gtrsim 0.1$ , increases at  $x \lesssim 0.1$

Durham PDF plotter, <http://durpdg.dur.ac.uk/hepdata/pdf3.html>

# Example of DGLAP evolution



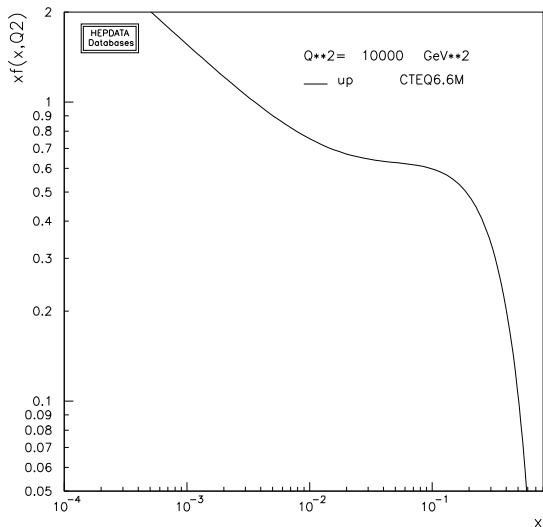
As  $Q$  increases, it becomes more likely that a high- $x$  parton loses some momentum through QCD radiation

$\Rightarrow u(x, Q)$  reduces at  $x \gtrsim 0.1$ , increases at  $x \lesssim 0.1$

Durham PDF plotter, <http://durpdg.dur.ac.uk/hepdata/pdf3.html>



# Example of DGLAP evolution

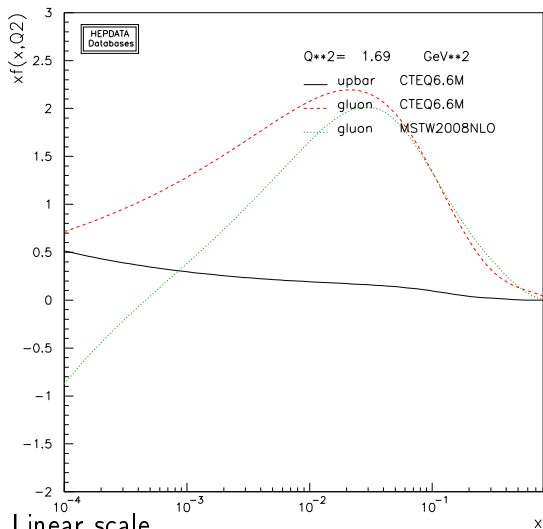


As  $Q$  increases, it becomes more likely that a high- $x$  parton loses some momentum through QCD radiation

$\Rightarrow u(x, Q)$  reduces at  $x \gtrsim 0.1$ , increases at  $x \lesssim 0.1$

Durham PDF plotter, <http://durpdg.dur.ac.uk/hepdata/pdf3.html>

# Example of DGLAP evolution: $\bar{u}$ and gluon PDF



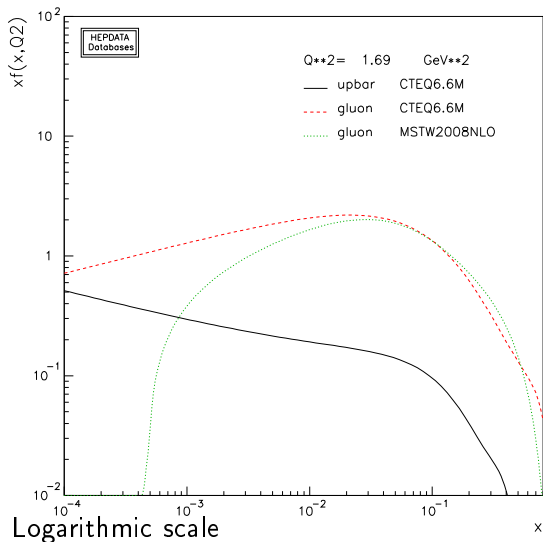
$g(x, Q)$  can become negative at  $x < 10^{-2}$ ,  $Q < 2 \text{ GeV}$

may lead to unphysical predictions

This is an indication that DGLAP factorization experiences difficulties at such small  $x$  and  $Q$

Large  $\ln^k(1/x)$  in  $P_{i/j}(x)$  break PQCD expansion at  $x \sim Q/\sqrt{s} \ll 1$

# Example of DGLAP evolution: $\bar{u}$ and gluon PDF



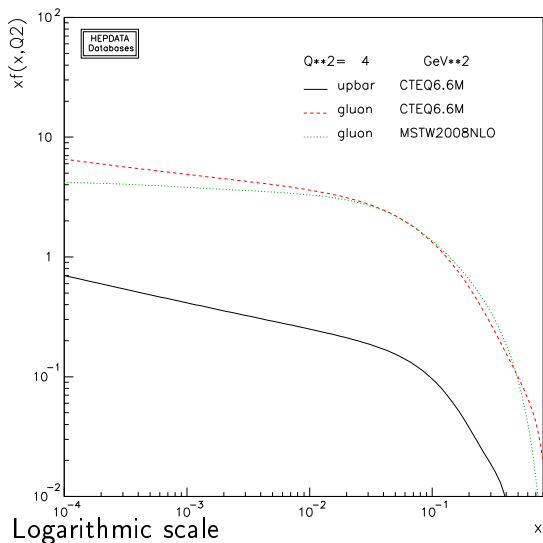
$g(x, Q)$  can become negative at  $x < 10^{-2}$ ,  $Q < 2 \text{ GeV}$

may lead to unphysical predictions

This is an indication that DGLAP factorization experiences difficulties at such small  $x$  and  $Q$

Large  $\ln^k(1/x)$  in  $P_{i/j}(x)$  break PQCD expansion at  $x \sim Q/\sqrt{s} \ll 1$

# Example of DGLAP evolution: $\bar{u}$ and gluon PDF

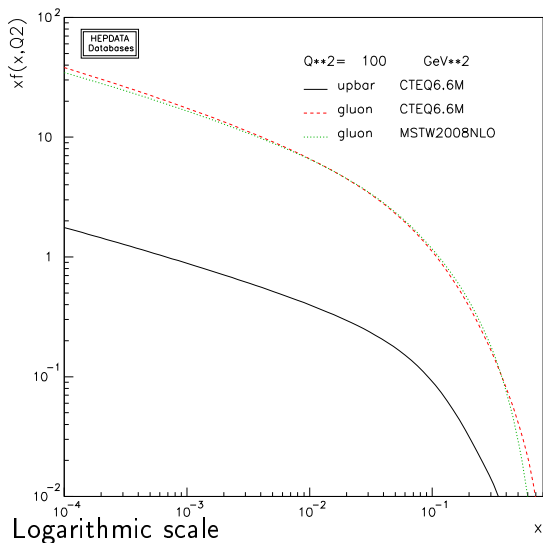


As  $Q$  increases,  $g(x, Q)$  grows rapidly at small  $x$

$\alpha_s(Q)$  becomes small enough to suppress  $\ln^k(1/x)$  terms

small- $x$  behavior stabilizes

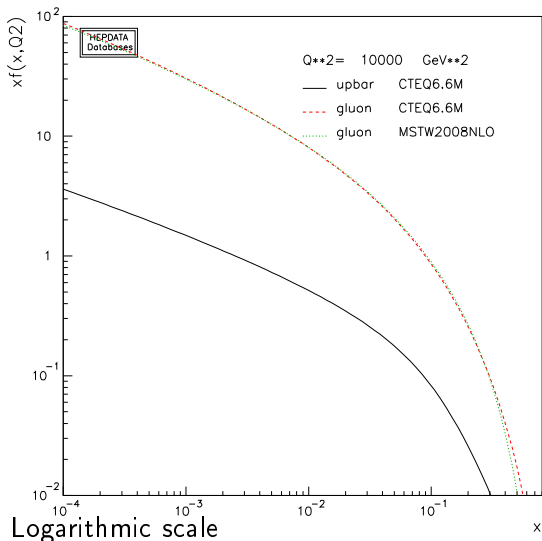
# Example of DGLAP evolution: $\bar{u}$ and gluon PDF



As  $Q$  increases,  $g(x, Q)$  grows rapidly at small  $x$

$\alpha_s(Q)$  becomes small enough to suppress  $\ln^k(1/x)$  terms  
 small- $x$  behavior stabilizes

# Example of DGLAP evolution: $\bar{u}$ and gluon PDF



As  $Q$  increases,  $g(x, Q)$  grows rapidly at small  $x$

$\alpha_s(Q)$  becomes small enough to suppress  $\ln^k(1/x)$  terms

small- $x$  behavior stabilizes

# Universality of PDFs

To all orders in  $\alpha_s$ , PDFs are **defined** as matrix elements of certain correlator functions:

$$f_{q/p}(x, \mu) = \frac{1}{4\pi} \int_{-\infty}^{\infty} dy^- e^{iy^- p^+} \langle p | \bar{\psi}_q(0, y^-, \vec{0}_T) \gamma^+ \psi_q(0, 0, \vec{0}_T) | p \rangle, \text{ etc.}$$

PDFs are **universal** – depend only on the type of the hadron ( $p$ ) and parton ( $q, \bar{q}, g$ )

... can be **parametrized** as

$$f_{i/p}(x, Q_0) = a_0 x^{a_1} (1-x)^{a_2} F(a_3, a_4, \dots) \text{ at } Q_0 \sim 1 \text{ GeV}$$

... predicted by solving DGLAP equations at  $Q > Q_0$

... found from a global fit to the benchmark hadronic data

# Typical features of the PDF analysis

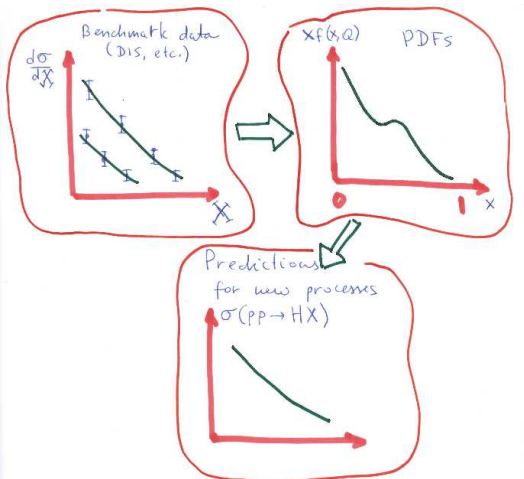
Data set	LO	NLO	NNLO
BCDMS $\mu p F_2$ [32]	165 / 153	182 / 163	170 / 163
BCDMS $\mu d F_2$ [102]	162 / 142	190 / 151	188 / 151
NMC $\mu p F_2$ [33]	137 / 115	121 / 123	115 / 123
NMC $\mu d F_2$ [33]	120 / 115	102 / 123	93 / 123
NMC $\mu n / \mu p$ [103]	131 / 137	130 / 148	135 / 148
E665 $\mu p F_2$ [104]	59 / 53	57 / 53	63 / 53
E665 $\mu d F_2$ [104]	49 / 53	53 / 53	63 / 53
SLAC $ep F_2$ [105, 106]	24 / 18	30 / 37	31 / 37
SLAC $ed F_2$ [105, 106]	12 / 18	30 / 38	26 / 38
NMC/BCDMS/SLAC $F_L$ [32-34]	28 / 24	35 / 31	32 / 31
E866/NuSea $pp$ DY [107]	239 / 184	228 / 184	237 / 184
E866/NuSea $pd/pp$ DY [108]	14 / 15	14 / 15	14 / 15
NuTeV $\nu N F_2$ [37]	49 / 49	49 / 53	46 / 53
CHORUS $\nu N F_2$ [38]	21 / 37	26 / 42	29 / 42
NuTeV $\nu N xF_3$ [37]	62 / 45	40 / 45	34 / 45
CHORUS $\nu N xF_3$ [38]	44 / 33	31 / 33	26 / 33
CCFR $\nu N \rightarrow \mu\mu X$ [39]	63 / 86	66 / 86	69 / 86
NuTeV $\nu N \rightarrow \mu\mu X$ [39]	44 / 40	39 / 40	45 / 40
H1 MB 99 $e^+p$ NC [31]	9 / 8	9 / 8	7 / 8
H1 MB 97 $e^+p$ NC [109]	46 / 64	42 / 64	51 / 64
H1 low $Q^2$ 96-97 $e^+p$ NC [109]	54 / 80	44 / 80	45 / 80
H1 high $Q^2$ 98-99 $e^+p$ NC [110]	134 / 126	122 / 126	124 / 126
H1 high $Q^2$ 99-00 $e^+p$ NC [35]	153 / 147	131 / 147	133 / 147
ZEUS SVX 95 $e^+p$ NC [111]	35 / 30	35 / 30	35 / 30
ZEUS 96-97 $e^+p$ NC [112]	118 / 144	86 / 144	86 / 144
ZEUS 98-99 $e^+p$ NC [113]	61 / 92	54 / 92	54 / 92
ZEUS 99-00 $e^+p$ NC [114]	75 / 90	63 / 90	65 / 90
H1 99-00 $e^+p$ CC [35]	28 / 28	29 / 28	29 / 28
ZEUS 99-00 $e^+p$ CC [36]	36 / 30	38 / 30	37 / 30
H1/ZEUS $ep F_2^{\text{charm}}$ [41-47]	110 / 83	107 / 83	95 / 83
H1 99-00 $e^+p$ incl. jets [59]	109 / 24	19 / 24	—
ZEUS 96-97 $e^+p$ incl. jets [57]	88 / 30	30 / 30	—
ZEUS 98-00 $e^+p$ incl. jets [58]	102 / 30	17 / 30	—
D0 II $pp$ incl. jets [56]	103 / 110	114 / 110	123 / 110
CDF II $pp$ incl. jets [54]	143 / 76	56 / 76	54 / 76
CDF II $W \rightarrow \ell\nu$ asym. [48]	50 / 22	29 / 22	30 / 22
D0 II $W \rightarrow \ell\nu$ asym. [49]	23 / 10	25 / 10	25 / 10
D0 II $Z$ rap. [53]	25 / 28	19 / 28	17 / 28
CDF II $Z$ rap. [52]	52 / 29	49 / 29	50 / 29
All data sets	3066 / 2598	2543 / 2699	2480 / 2615

■ PDFs are not measured directly, but some data sets are more sensitive to specific combinations of PDFs. By constraining these combinations, the PDFs can be disentangled in a combined (global) fit to many diverse processes. Some of these fits involve up to 40 experiments, 5000+ data points, and 100+ free parameters

Data sets and  $\chi^2/d.o.f.$   
from MSTW'2008 analysis

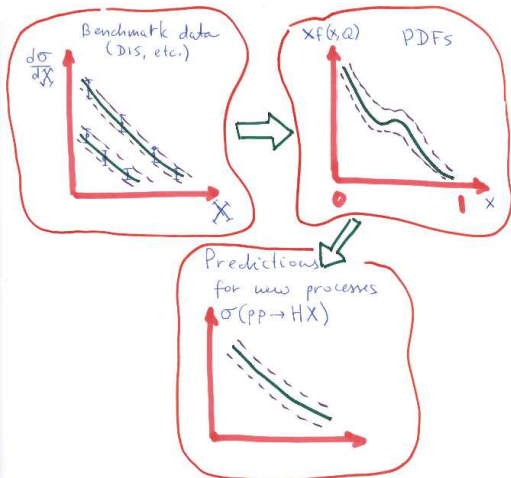


# Typical features of the PDF analysis



We are interested not just in one best fit, but also in the uncertainty of the resulting PDF parametrizations and theoretical predictions based on them. This will be covered in Lecture 2

# Typical features of the PDF analysis



We are interested not just in one best fit, but also in the uncertainty of the resulting PDF parametrizations and theoretical predictions based on them. This will be covered in Lecture 2

## Question to the audience

Which of the SM particles can have a non-zero PDF  $f_{a/p}(x, Q)$  in the proton?

1. light partons  $u, d, s, g$
2. heavy quarks  $c, b, t$
3. photon  $\gamma$ ; leptons  $e, \mu, \tau, \nu$
4. massive electroweak bosons  $W$  and  $Z$

## Question to the audience

Which of the SM particles can have a non-zero PDF  $f_{a/p}(x, Q)$  in the proton?

1. light partons  $u, d, s, g$
2. heavy quarks  $c, b, t$
3. photon  $\gamma$ ; leptons  $e, \mu, \tau, \nu$
4. massive electroweak bosons  $W$  and  $Z$

### Answer

All of them – the PDF can be defined for any particle

However, only partons with mass  $\lesssim 1$  GeV are expected to have a non-negligible  $f_{a/p}(x, Q_0)$  at the initial scale  $Q_0 \approx 1$  GeV

## Boundary conditions at $Q_0$

In practice, independent parametrizations  $f_{a/p}(x, Q_0)$  are introduced for

- $g, u, d, s, \bar{u}, \bar{d}, \bar{s}$  (always)  
contribute  $> 97\%$  of the proton's energy  $E_p$  at  $Q_0$ 
  - ▶ even in this case, the data are usually insufficient for constraining all PDF parameters; some of them can be fixed by hand
  - ▶ e.g.,  $\bar{u} = \bar{d} = \bar{s}$  in outdated fits
- $c$  and or  $b$  (occasionally; in a model allowing nonperturbative “intrinsic heavy-quark production”)
- photons  $\gamma$  (tentatively, in MRST'03 QCD+QED PDFs)
  - ▶ a QCD+QED fit is more complicated than one might think: it must account for violation of charge symmetry by EM effects,

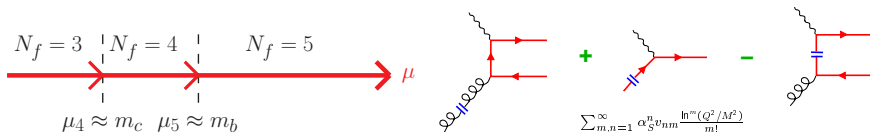
$$u_p(x, Q) \neq d_n(x, Q); \quad d_p(x, Q) \neq u_n(x, Q)$$

# PDFs for heavy flavors

- PDFs for heavy partons  $h$  can be generated via DGLAP evolution at  $Q \geq m$ , using a boundary condition  $f_{h/p}(x, Q) = 0$  at  $Q \leq m$
- In practice:
  - ▶ perturbative PDFs are only introduced for  $c$  and  $b$  quarks
  - ▶ QCD coupling  $\alpha_s(Q)$  and PDFs are evaluated with 5 active flavors at all  $Q \geq m_b$
  - ▶ Moderate logarithmic enhancements may exist in collinear  $t, W, Z$  production at  $Q > 1$  TeV; but, for all practical purposes, it suffices to evaluate these collinear terms as a part of hard cross sections

# General-mass variable-flavor number scheme

- A series of factorization schemes with  $N_f$  active quark flavors in  $\alpha_s(Q)$  and  $f_{a/p}(x, Q)$ 
  - ▶  $N_f$  is incremented sequentially at momentum scales  $\mu_{N_f} \approx m_{N_f}$
- incorporates essential  $m_{c,b}$  dependence near, and away from, heavy-flavor thresholds
- implemented in the latest CTEQ, HERA, MSTW PDF sets

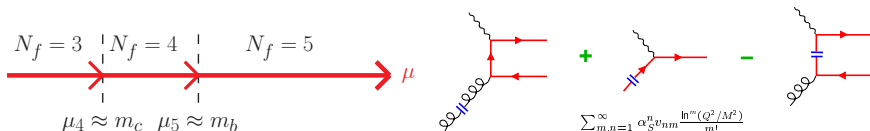


# General-mass variable-flavor number scheme

Proved for *inclusive DIS* by J. Collins (1998)

$$F_2(x, Q, m_c) = \sum_a \int_x^1 \frac{d\xi}{\xi} H_a\left(\frac{\chi}{\xi}, \frac{Q}{\mu}, \frac{m_c}{Q}\right) f_a\left(\xi, \frac{\mu}{m_c}\right) + \mathcal{O}\left(\frac{\Lambda_{QCD}}{Q}\right)$$

- $\lim_{Q \rightarrow \infty} H$  exists and is infrared safe
- collinear logarithms  $\sum_{k,n=1}^{\infty} \alpha_s^k v_{kn} \ln^n(\mu/m_c)$  are resummed in  $f_c(x, \mu/m_c)$
- no terms  $\mathcal{O}(m_c/Q)$  in the remainder



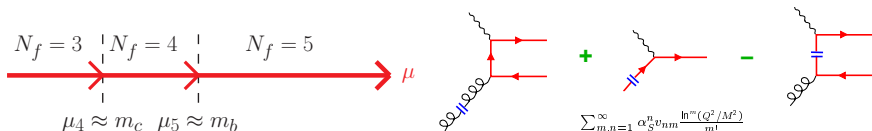


# General-mass variable-flavor number scheme

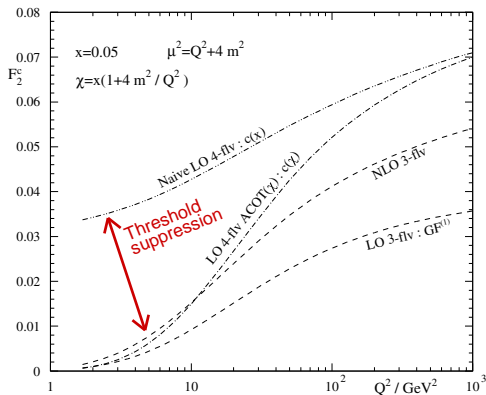
Proved for *inclusive DIS* by J. Collins (1998)

$$F_2(x, Q, m_c) = \sum_a \int_\chi^1 \frac{d\xi}{\xi} H_a\left(\frac{\chi}{\xi}, \frac{Q}{\mu}, \frac{m_c}{Q}\right) f_a\left(\xi, \frac{\mu}{m_c}\right) + \mathcal{O}\left(\frac{\Lambda_{QCD}}{Q}\right)$$

- Works most effectively in DIS and Drell-Yan-like processes; practical implementation requires
  1. efficient treatment of mass dependence, rescaling of momentum fractions  $\chi$  in processes with incoming  $c, b$
  2. physically motivated factorization scale to ensure fast PQCD convergence (e.g.,  $\mu = Q$  in DIS)



# An example of GM-VFN factorization scheme



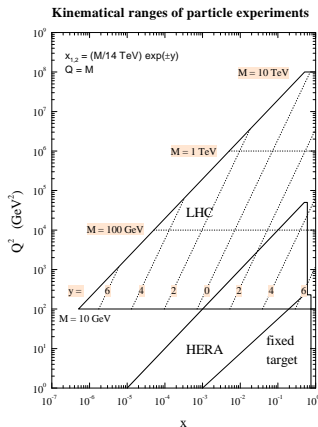
- Charm Wilson coefficient function is suppressed at  $Q \rightarrow m_c$
- To keep agreement with  $F_2$  data,  $u$ ,  $d$ ,  $\bar{u}$ ,  $\bar{d}$  PDF's are enhanced at small  $x$ , as compared to the zero-mass (ZM-VFN) scheme

# Experimental observables



# Selection of experimental data

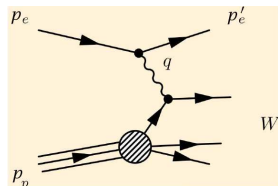
- Lepton pair production ( $pN \xrightarrow{\gamma^*, W, Z} \ell\bar{\ell}' X$ ):  
Tevatron, fixed-target experiments
- Inclusive jet production:  
 $p\bar{p} \rightarrow j X$  (Tevatron),  $ep \rightarrow j(j) X$  (HERA)



Hard cross sections are known at NNLO (two loops) for lepton pair production, NLO (one loop) for jet production

## Neutral-current $ep$ DIS: kinematics

- $s = (p_e + p_p)^2$  – total energy
- $Q^2 = -q^2 = -(p_e - p'_e)^2$  – momentum transfer
- $x = Q^2/(2p_p \cdot q)$  – Bjorken scaling variable
- $y = Q^2/(xs)$  – inelasticity
- $W^2 = Q^2(1 - x)/x$  – energy of the hadronic final state



$$\frac{d^2\sigma(e^\pm p)}{dQ^2 dx} = \frac{2\pi\alpha^2}{Q^4 x} Y_\pm \left( F_2 - \frac{y^2}{Y_\pm} F_L \pm \frac{Y_\mp}{Y_\pm} x F_3 \right),$$

with  $Y_\pm \equiv 1 \pm (1 - y)^2$

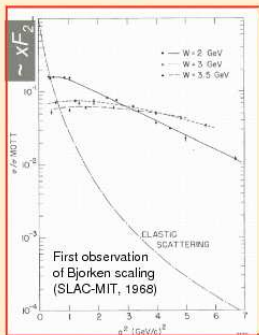
The data is fitted either in the form of  $F_2(x, Q^2)$  or  $d^2\sigma/(dQ^2 dx)$

# Bjorken scaling of $F_2(x, Q^2)$

40 years after its discovery

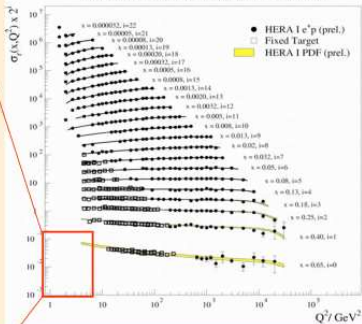


## DIS at HERA



2008

H1 and ZEUS Combined PDF Fit



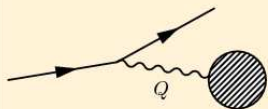
April 2008

HERA Structure Functions Working Group

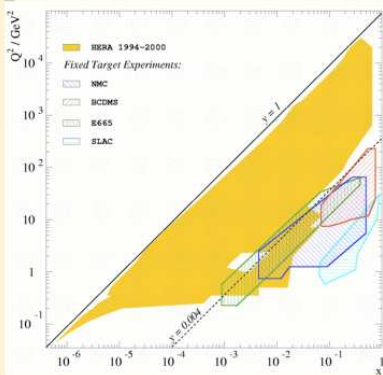
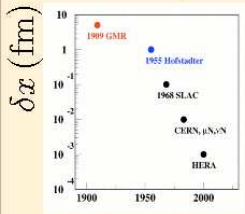
Slide by T. Haas, DESY

# DIS experiments: kinematical reach

## HERA kinematic reach



$$\delta x \approx \frac{200 \text{ MeV}}{Q}$$



Slide by T. Haas, DESY



## PDF combinations in DIS at the lowest order

- Neutral current  $\ell^\pm p$ :

$$F_2^{\ell^\pm p}(x, Q^2) = \frac{4}{9} (u + \bar{u} + c + \bar{c}) + \frac{1}{9} (d + \bar{d} + s + \bar{s} + b + \bar{b})$$

- ▶ PDFs are weighted by the fractional EM quark coupling  $e_i^2 = 4/9$  or  $1/9$
- ▶ 4 times more sensitivity to  $u$  and  $c$  than to  $d$ ,  $s$ , and  $b$
- ▶ No sensitivity to the gluon at this order

- Neutral current ( $\ell^\pm N$ ) DIS on isoscalar nuclei ( $N = (p + n)/2$ ):

$$F_2^{\ell^\pm N}(x, Q^2) = \frac{5}{9} (u + \bar{u} + d + \bar{d} + \text{smaller } s, c, b \text{ contributions})$$

- Charged current ( $\nu N$ ) DIS :

$$F_2^{\nu N}(x, Q^2) = x \sum_{i=u,d,s,\dots} (q_i + \bar{q}_i)$$

$$xF_3^{\nu N}(x, Q^2) = x \sum_{i=u,d,s} (q_i - \bar{q}_i)$$

## DIS at next-to-leading order (NLO) and beyond

Logarithmic corrections to Bjorken scaling ( $Q$  dependence of  $F_2(x, Q^2)$ ) are sensitive to the gluon PDF through DGLAP equations,

$$\mu \frac{df_{i/p}(x, \mu)}{d\mu} = \sum_{j=g,u,\bar{u},d,\bar{d},\dots} \int_x^1 \frac{dy}{y} P_{i/j} \left( \frac{x}{y}, \alpha_s(\mu) \right) f_{j/p}(y, \mu)$$

Thus, when examined at NLO, the DIS data constrains

- $\sum_i e_i^2 (q_i + \bar{q}_i)$  in an amazingly large range  $10^{-5} < x < 0.5$
- $u$  and  $d$  at  $10^{-2} < x < 0.3$
- $g(x, Q)$  at  $x < 0.1$

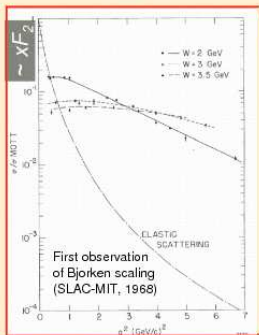
Ability of DIS to separate quarks from antiquarks, or small  $s$ ,  $c$ ,  $b$  contributions from large  $u$  and  $d$  contributions, is limited; more so because of systematic effects in fixed-target DIS experiments (higher-order terms, nuclear corrections, etc.)

# Bjorken scaling of $F_2(x, Q^2)$

40 years after its discovery

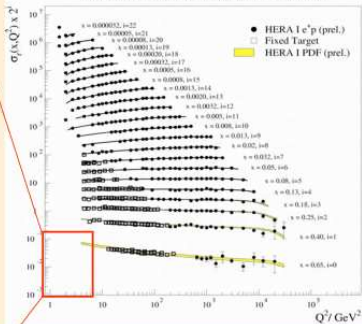


## DIS at HERA



2008

H1 and ZEUS Combined PDF Fit



April 2008

HERA Structure Functions Working Group

Slide by T. Haas, DESY

## DIS at next-to-leading order (NLO) and beyond

Logarithmic corrections to Bjorken scaling ( $Q$  dependence of  $F_2(x, Q^2)$ ) are sensitive to the gluon PDF through DGLAP equations,

$$\mu \frac{df_{i/p}(x, \mu)}{d\mu} = \sum_{j=g,u,\bar{u},d,\bar{d},\dots} \int_x^1 \frac{dy}{y} P_{i/j} \left( \frac{x}{y}, \alpha_s(\mu) \right) f_{j/p}(y, \mu)$$

Thus, when examined at NLO, the DIS data constrains

- $\sum_i e_i^2 (q_i + \bar{q}_i)$  in an amazingly large range  $10^{-5} < x < 0.5$
- $u$  and  $d$  at  $10^{-2} < x < 0.3$
- $g(x, Q)$  at  $x < 0.1$

Ability of DIS to separate quarks from antiquarks, or small  $s$ ,  $c$ ,  $b$  contributions from large  $u$  and  $d$  contributions, is limited; more so because of systematic effects in fixed-target DIS experiments (higher-order terms, nuclear corrections, etc.)

## Constraints on quark sea from $pN \rightarrow \ell^+ \ell^- X$

( $N = p, d, Fe, Cu, \dots$ )

$$\frac{d\sigma_{pp}}{dQ^2 dy} \sim \left(\frac{2}{3}\right)^2 [u_A \bar{u}_B + \bar{u}_A u_B] + \left(-\frac{1}{3}\right)^2 [d_A \bar{d}_B + \bar{d}_A d_B] + \text{smaller terms}$$

$\Rightarrow$  sensitivity to  $\bar{q}(x, Q)$

Assuming charge symmetry between protons and neutrons ( $u_p = d_n$ ,  $u_n = d_p$ ):

$$\frac{d\sigma_{pn}}{dQ^2 dy} \sim \left(\frac{2}{3}\right)^2 [u_A \bar{d}_B + \bar{u}_A d_B] + \left(-\frac{1}{3}\right)^2 [d_A \bar{u}_B + \bar{d}_A u_B] + \text{smaller terms}$$

If deuterium binding corrections are neglected:  $q_d(x) \approx q_p(x) + q_n(x)$

At  $x_A \gg x_B$  (large  $y$ ):  $\bar{q}(x_A) \sim 0$  and  $4u(x_A) \gg d(x_A)$

$$\frac{\sigma_{pd}}{2\sigma_{pp}} \approx \frac{1}{2} \frac{(1 + \frac{d_A}{4u_A})[1 + r]}{(1 + \frac{d_A}{4u_A} r)} \approx \frac{1}{2}(1 + r), \text{ where } r \equiv \bar{d}(x_B)/\bar{u}(x_B)$$

$\therefore \sigma_{pd}/(2\sigma_{pp})$  constrains  $\bar{d}(x, Q)/\bar{u}(x, Q)$  at moderate  $x$

## Constraints on quark sea from $pN \rightarrow \ell^+ \ell^- X$

( $N = p, d, Fe, Cu, \dots$ )

$$\frac{d\sigma_{pp}}{dQ^2 dy} \sim \left(\frac{2}{3}\right)^2 [u_A \bar{u}_B + \bar{u}_A u_B] + \left(-\frac{1}{3}\right)^2 [d_A \bar{d}_B + \bar{d}_A d_B] + \text{smaller terms}$$

$\Rightarrow$  sensitivity to  $\bar{q}(x, Q)$

Assuming charge symmetry between protons and neutrons ( $u_p = d_n$ ,  $u_n = d_p$ ):

$$\frac{d\sigma_{pn}}{dQ^2 dy} \sim \left(\frac{2}{3}\right)^2 [u_A \bar{d}_B + \bar{u}_A d_B] + \left(-\frac{1}{3}\right)^2 [d_A \bar{u}_B + \bar{d}_A u_B] + \text{smaller terms}$$

If deuterium binding corrections are neglected:  $q_d(x) \approx q_p(x) + q_n(x)$

At  $x_A \gg x_B$  (large  $y$ ):  $\bar{q}(x_A) \sim 0$  and  $4u(x_A) \gg d(x_A)$

$$\frac{\sigma_{pd}}{2\sigma_{pp}} \approx \frac{1}{2} \frac{(1 + \frac{d_A}{4u_A})[1 + r]}{(1 + \frac{d_A}{4u_A} r)} \approx \frac{1}{2}(1 + r), \text{ where } r \equiv \bar{d}(x_B)/\bar{u}(x_B)$$

$\therefore \sigma_{pd}/(2\sigma_{pp})$  constrains  $\bar{d}(x, Q)/\bar{u}(x, Q)$  at moderate  $x$

## Constraints on quark sea from $pN \rightarrow \ell^+ \ell^- X$

( $N = p, d, Fe, Cu, \dots$ )

$$\frac{d\sigma_{pp}}{dQ^2 dy} \sim \left(\frac{2}{3}\right)^2 [u_A \bar{u}_B + \bar{u}_A u_B] + \left(-\frac{1}{3}\right)^2 [d_A \bar{d}_B + \bar{d}_A d_B] + \text{smaller terms}$$

$\Rightarrow$  sensitivity to  $\bar{q}(x, Q)$

Assuming charge symmetry between protons and neutrons ( $u_p = d_n$ ,  $u_n = d_p$ ):

$$\frac{d\sigma_{pn}}{dQ^2 dy} \sim \left(\frac{2}{3}\right)^2 [u_A \bar{d}_B + \bar{u}_A d_B] + \left(-\frac{1}{3}\right)^2 [d_A \bar{u}_B + \bar{d}_A u_B] + \text{smaller terms}$$

If deuterium binding corrections are neglected:  $q_d(x) \approx q_p(x) + q_n(x)$

At  $x_A \gg x_B$  (large  $y$ ):  $\bar{q}(x_A) \sim 0$  and  $4u(x_A) \gg d(x_A)$

$$\frac{\sigma_{pd}}{2\sigma_{pp}} \approx \frac{1}{2} \frac{(1 + \frac{d_A}{4u_A})[1 + r]}{(1 + \frac{d_A}{4u_A} r)} \approx \frac{1}{2}(1 + r), \text{ where } r \equiv \bar{d}(x_B)/\bar{u}(x_B)$$

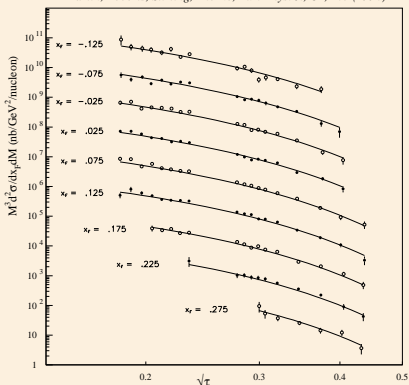
$\therefore \sigma_{pd}/(2\sigma_{pp})$  constrains  $\bar{d}(x, Q)/\bar{u}(x, Q)$  at moderate  $x$

# Theory vs. experiment

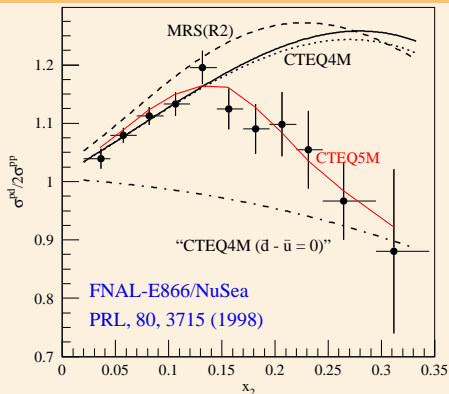
## Cross section at $Q = 4 - 17$ GeV

E605 (p Cu  $\rightarrow \mu^+ \mu^- X$ )  $P_{LAB} = 800$  GeV

Martin, Roberts, Stirling, Thorne, *Eur. Phys. J., C4, 463 (1998)*



## $\sigma_{pd}/(2\sigma_{pp})$ at large $x_F = x_A - x_B$



Theory curves reflect different assumptions about  $\bar{d}/\bar{u}$

The recent PDF fits (starting with CTEQ5M) quantitatively account for the violation of  $SU(2)$  symmetry in the quark sea



## What about $s(x, Q) \neq \bar{s}(x, Q)$ ?

This can be tested in subprocesses

$$W^+ s \rightarrow c \text{ and } W^- \bar{s} \rightarrow \bar{c}$$

In the experiment, charm quarks can be identified by their semileptonic decays,

$$c \rightarrow s \mu^+ \nu \text{ and } \bar{c} \rightarrow \bar{s} \mu^- \bar{\nu}$$

So one sees

$$\nu N \rightarrow \mu^- c X \rightarrow \mu^- \mu^+ X$$

$$\bar{\nu} N \rightarrow \mu^+ c X \rightarrow \mu^+ \mu^- X$$

— SIDIS muon pair  
production (NuTeV)

# Total strangeness and strangeness asymmetry

Denote

$$[q_i](Q) \equiv \int_0^1 x q_i(x, Q) dx \text{ —net moment fraction carried by } q_i$$

and introduce  $s^\pm(x) = s(x) \pm \bar{s}(x)$  (total strangeness and its asymmetry)

It is possible that

$$\int_0^1 s^-(x) dx = 0$$

(a proton has no net strangeness), but

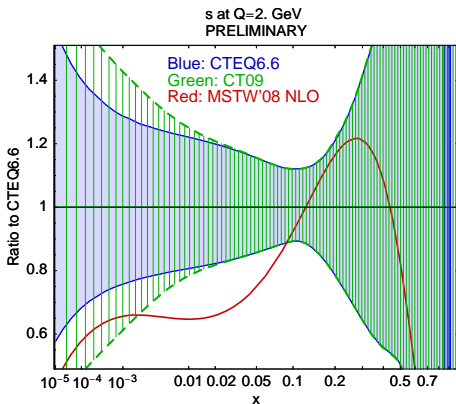
$$[S^-] \equiv \int_0^{-1} x s^-(x) dx \neq 0$$

( $s$  and  $\bar{s}$  have different  $x$  distributions)

A large non-vanishing  $[S^-]$  might explain “the NuTeV weak angle anomaly”

When deriving  $[S^-]$  from the data, one must be careful to avoid being biased by the choice of the parametrization for  $s^-(x, Q)$

# CCFR (inclusive $\nu N$ DIS) and NuTeV (SIDIS dimuon production): constraints on strangeness



$s^+(x, Q)$  is reasonably well constrained at  $x > 0.01$ ; practically unknown at  $x < 0.01$

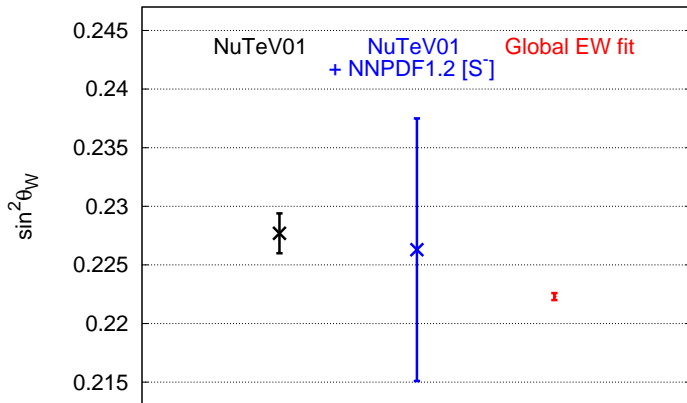
2009 NNPDF estimate (least biased by the parametrization of  $s^-(x, Q)$ ):

$$[S^-](Q^2 = 20 \text{ GeV}^2) = 0 \pm 0.009$$

No statistically significant  $[S^-]$ ; but the PDF error is large enough to eliminate the NuTeV anomaly (!)

# NuTeV, NNPDF1.2, and global EW fit uncertainties on $\sin^2 \theta_w$

Determinations of the weak mixing angle  $\sin^2 \theta_w$



## How is the quark sea flavor symmetry broken?

Several competing mechanisms may contribute

- NNLO DGLAP evolution generally predicts that  $\bar{u}(x, Q_2) \neq \bar{d}(x, Q_2)$ ,  $\bar{s}(x, Q_2) \neq s(x, Q_2)$  at  $Q_2 > Q_1$ , if  $u(x, Q_1) \neq d(x, Q_1)$
- Fermi motion
  - ▶ Since  $d(x, Q)/u(x, Q) < 1$ , expect enhancement in  $\bar{d}(x, Q)/\bar{u}(x, Q)$  at  $x \gtrsim 0.1$
- Nonperturbative meson fluctuations
  - ▶  $p \rightarrow p\pi^0 \rightarrow p$  (or  $uud \rightarrow (udd)(u\bar{d})$ ) suggests suppression of  $\bar{d}(x, Q)/\bar{u}(x, Q)$  ( $n$  carries more momentum than  $\pi^0$  at  $x \rightarrow 1$ )
  - ▶  $p \rightarrow \Lambda K^+ \rightarrow p$  (or  $uud \rightarrow (uds)(u\bar{s})$ ) suggests  $[S^-] > 0$  ( $\Lambda$  carries more momentum than  $K^+$  at  $x \rightarrow 1$ )

# Charged lepton asymmetry at the Tevatron

$p\bar{p} \rightarrow (W \rightarrow e\nu)X$  at  $\sqrt{s} = 1.8 - 1.96$  TeV

$$A_{ch}(y_e) \equiv \frac{\frac{d\sigma^{W^+}}{dy_e} - \frac{d\sigma^{W^-}}{dy_e}}{\frac{d\sigma^{W^+}}{dy_e} + \frac{d\sigma^{W^-}}{dy_e}}$$

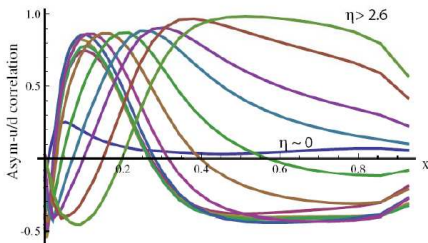
- related to the boson Born-level asymmetry when  $y_e$  is large

$$A_{ch}(y) \xrightarrow{y \rightarrow y_{max}} \frac{r(x_B) - r(x_A)}{r(x_B) + r(x_A)}, \quad r(x) \equiv \frac{d(x, M_W)}{u(x, M_W)}$$

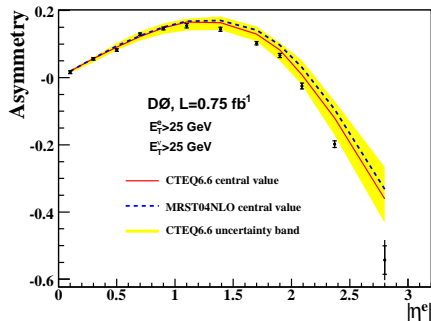
- constrains the PDF ratio  $d(x, M_W)/u(x, M_W)$  at  $x \rightarrow 1$
- In experimental analyses, a selection cut  $p_{Te} > p_{Te}^{min}$  is imposed

# Impact of $W$ charge asymmetry on the PDF's

- Correlation of  $u(x)/d(x)$  and  $A_e$  for each  $\eta_e$  bin



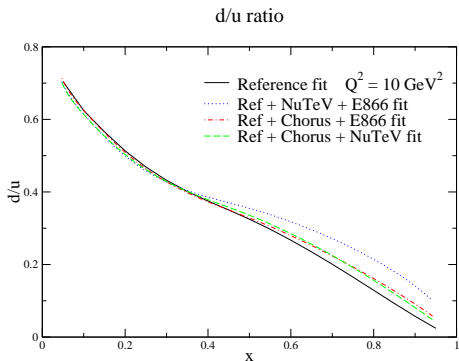
H. Schellman, talk at BNL, 2009



Asymmetry is correlated with the PDF uncertainty of  $d(x, M_W)/u(x, M_W)$  at large  $x$ . Tevatron Run-2 measurements are expected to reduce this uncertainty.

# What do we know about $d(x, Q)/u(x, Q)$ at large $x$ ?

Existing experiments disagree about the behavior of  $d(x, Q)/u(x, Q)$  at  $x > 0.1$



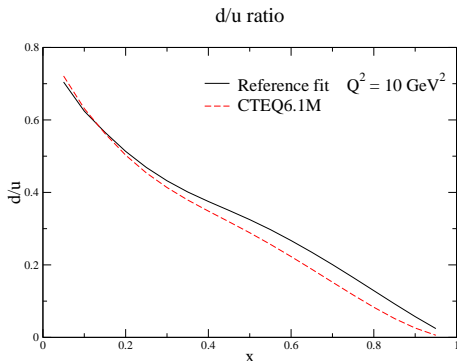
Fits to different combinations of CCFR, NuTeV (inclusive  $\nu N$  DIS) and E866 ( $pp/pd$  Drell-Yan) experiments result in widely ranging predictions for the  $d/u$  ratio

(*J. F. Owens et al., PRD 054030 (2007)*)



# What do we know about $d(x, Q)/u(x, Q)$ at large $x$ ?

Existing experiments disagree about the behavior of  $d(x, Q)/u(x, Q)$  at  $x > 0.1$



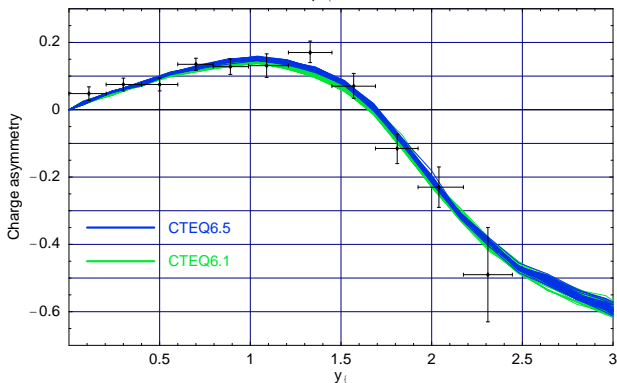
These **fixed-target** experiments are sensitive to the choice of the nuclear corrections for isoscalar nuclei or deuterium – see J. Owens' lecture

Black and wide curves assume different deuterium corrections

## What do we know about $d(x, Q)/u(x, Q)$ at large $x$ ?

Existing experiments disagree about  
the behavior of  $d(x, Q)/u(x, Q)$  at  $x > 0.1$

$25 < p_{Tl} < 35$  GeV

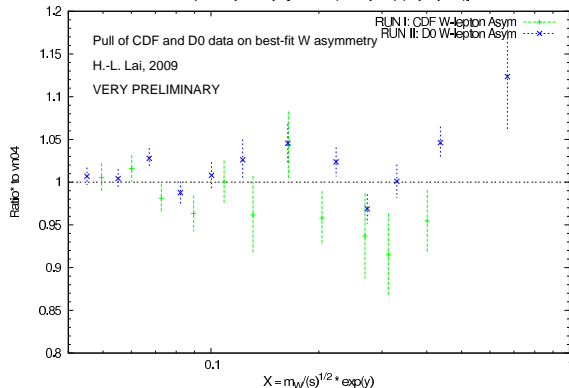


CDF Run-1 and Run-2 ( $0.2 \text{ fb}^{-1}$ )  $W$  asymmetry measurements are consistent with NC  $ep$  DIS, CCFR, E866, CHORUS (not NuTeV)...

# What do we know about $d(x, Q)/u(x, Q)$ at large $x$ ?

Existing experiments disagree about  
the behavior of  $d(x, Q)/u(x, Q)$  at  $x > 0.1$

W-Lepton Asymmetry [Ratio\* = (DatAsy-0.5)/(ThyAsy-0.5)]



... but they don't seem to agree with the more precise  
D0 Run-2 ( $0.75 \text{ fb}^{-1}$ )  $W$  asymmetry measurement  
– this is being investigated

## Remarks about $d(x, Q)/u(x, Q)$ at $x \rightarrow 1$

The large- $x$  asymptotic behavior of  $d/u$  is an old outstanding QCD puzzle

It reflects the spin and isospin structure of the proton wave function

It is tied to the  $x \rightarrow 1$  limit of  $F_2^n(x, Q)/F_2^p(x, Q)$ , known to take specific values (1/4, 3/7, ...) in semi-quantitative models of proton structure

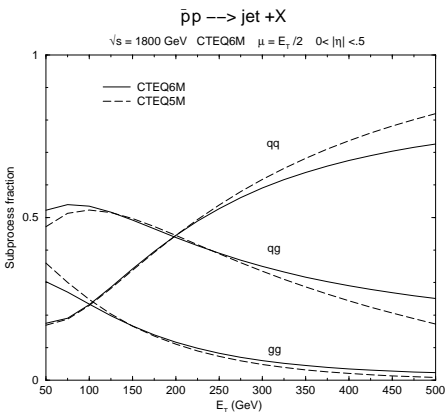
## Remarks about $d(x, Q)/u(x, Q)$ at $x \rightarrow 1$

- The limit  $x \rightarrow 1$  for  $Q^2$  fixed corresponds to the nearly elastic regime of  $ep$  scattering

$$W^2 = m_p^2 + Q^2(1/x - 1) \rightarrow m_p^2, \text{ where } m_p \text{ is the target proton's mass}$$

- Usual factorization (for  $m_p = 0$ ) must be modified to account for
  - ▶ target mass corrections of order  $\mathcal{O}(m_p^2/Q^2)$
  - ▶ higher-twist terms of order  $a(x)/Q^2, b(x)/Q^4$
- PDF fits that account for both kinds of power corrections are being studied at Jefferson Lab (*Accardi, Keppel, Melnitchouk, and collaborators*)
- At  $x > 0.5$ , all -order resummation of threshold  $\ln^p(1-x)$  terms (*Sterman, Laenen, ...*) and target mass corrections (*Accardi, Qiu, 2008*) must be included

# Inclusive jet production at the Tevatron



High- $E_T$  jets are mostly produced in  $qq$  scattering; yet most of the PDF uncertainty arises from  $qg$  and  $gg$  contributions

Here typical  $x$  is of order

$$2E_T/\sqrt{s} \gtrsim 0.1;$$

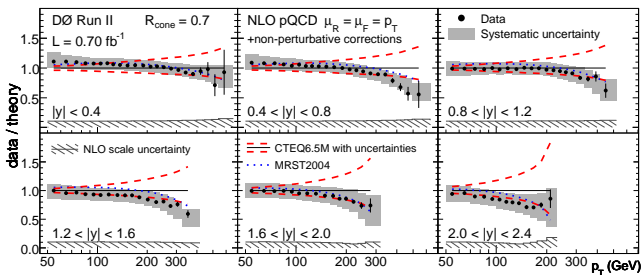
e.g.,  $x \approx 0.2$  for  $E_T = 200 \text{ GeV}$ ,  
 $\sqrt{s} = 1.8 \text{ TeV}$

At such  $x$ ,  $u(x, Q)$  and  $d(x, Q)$  are known very well; uncertainty arises mostly from  $g(x, Q)$

∴ Tevatron jet data

- constrain  $g(x, Q)$  at  $x > 0.1$
- complements HERA constraints on  $g(x, Q)$  at  $x < 0.1$

# Correlated systematic errors in $p\bar{p} \rightarrow jX$



D0 Coll., arXiv:0802.2400  
 ( $700 \text{ pb}^{-1}$ ); CDF results  
 ( $1.13 \text{ fb}^{-1}$ )

■ (Almost)  
 negligible  
 statistical error

- Experimental uncertainty is dominated by correlated shifts of data points due to systematic effects
- There are about 20 systematic errors published by each Run II jet experiment
- The PDF uncertainty would be strongly underestimated if these systematic errors are not included

## Homework assignment

When experimental errors  $\sigma_i$  are uncorrelated, the log-likelihood function  $\chi^2$  of the fit is constructed as

$$\chi^2 = \sum_{i=1}^{N_{points}} \left( \frac{T_i - D_i}{\sigma_i} \right)^2 \quad (1)$$

in terms of the theoretical prediction  $T_i$  and experimental value  $D_i$  at each point

In general, one needs to modify Eq. (1) to include a variety of systematic effects

- Theoretical: dependence on factorization and renormalization scale, higher-order/higher-twist terms, resummation effects, etc.
- Experimental: overall normalization, calibration of detectors, energy resolution, etc.

These are often examined by independent authors and provided to the fitters in some simplified form



# Homework assignment

## Task for tomorrow

Propose a procedure to supplement Eq. (1) with external information about systematic uncertainties. For example,  $N$  systematic errors result in  $N$  unknown correlated shifts of data with respect to theory. Find a optimal combination of the random shifts that minimizes  $\chi^2$ . Such combination can be found **analytically**; when this is done in the global fit, one obtains the most conservative (largest) PDF uncertainty consistent with the correlated shifts

If you have difficulties, read Appendix B in J. Pumplin et al. (CTEQ6), JHEP 0207, 012 (2002)

# Choice of PDF parametrization

# Statistical aspects

# Practical applications