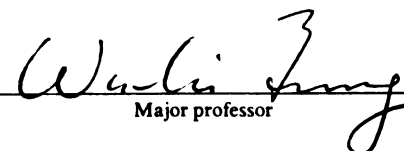This is to certify that the

dissertation entitled

GLOBAL QCD ANALYSIS AND PARTON STRUCTURE OF THE NUCLEON

presented by

Hung-Liang Lai

has been accepted towards fulfillment
of the requirements for

_____Ph.D._____ degree in _Physics_____

_____
Major professor

Date___Aug. 11, 1997___

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.

DATE DUE	DATE DUE	DATE DUE
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

c:\circ\datedue.pm3-p.1

GLOBAL QCD ANALYSIS AND PARTON STRUCTURE OF THE NUCLEON

By

Hung-Liang Lai

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Physics and Astronomy

1997

ABSTRACT

GLOBAL QCD ANALYSIS AND PARTON STRUCTURE OF THE NUCLEON

By

Hung-Liang Lai

Perturbative Quantum Chromodynamics (QCD) provides a unified framework to describe all high energy hard processes, both within and beyond the Standard Model. Global QCD analyses of all available data on high energy processes provide valuable consistency tests of perturbative QCD and furnish the means to determine the parton structure of hadrons. We describe the global analysis program performed by CTEQ collaboration in detail; and discuss the progress on the determination of the parton distributions of the nucleon, especially with the advent of collider data from the $e - p$ collider at HERA, and the $p - \bar{p}$ collider at the Fermilab Tevatron. A series of CTEQ parton distributions is presented, with special attention to the latest CTEQ4 analysis which emphasizes the determination of the more elusive but important gluon distribution. Recent developments pertaining to the elucidation of the highly publicized CDF high E_t inclusive jet events and to the improvement of the global analysis to include heavy quark effects are also described.

For My Parents.

ACKNOWLEDGEMENTS

To my family, for all their support;

To God, for everything.

Contents

List of Tables

List of Figures

xii

xiii

Chapter 1

Introduction to the Standard Model

The Standard Model [1, 2] (SM), incorporating electroweak theory and quantum chromodynamics (QCD), has been very successful in describing and predicting experimental data. It respects local gauge invariance with symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$, contains three generations of leptons and quarks, and incorporates spontanenous electroweak breaking. The $SU(3)_C$ sector, known as QCD, describes the strong interaction which has 8 gauge bosons called gluons with different color quantum numbers. The $SU(2)_L \times U(1)_Y$ sector, known as electroweak theory, describes the electromagnetic and weak interactions, or collectively the electroweak interaction which has $3 + 1$ gauge bosons ($W^{\pm}, Z^0$ and γ). The electromagnetic and weak interactions are unified via the mixing of the neutral $SU(2)_L$ gauge boson and the hypercharge gauge boson of $U(1)_Y$. The Higgs boson (H) is introduced through the process of spontaneous symmetry breaking (the Higgs mechanism) which gives masses to the particles.

The masses of three generations of leptons and quarks in the SM are listed in Table 1.1. The associated gauge bosons are listed in Table 1.2. All the standard model particles have been discovered except the Higgs boson. The detailed properties of each particle can be found in the Particle Data Book [3].

1

Table 1.1: Lepton and Quark Masses

Particle	Symbol	Mass (GeV)	
Electron neutrino	ν_e	0	
Electron	e	0.00051	First
Up quark	u	0.002 to 0.008	Generation
Down quark	d	0.005 to 0.015	
Muon neutrino	ν_μ	0	
Muon	μ	0.106	Second
Charm quark	c	1.0 to 1.6	Generation
Strange quark	s	0.1 to 0.3	
Tau neutrino	ν_τ	0	
Tau	τ	1.78	Third
Top quark	t	180	Generation
Bottom quark	b	4.1 to 4.5	

Table 1.2: Boson Masses

Particle	Symbol	Mass (GeV)	
Photon	γ	0	Electromagnetic Force
W Boson	$W^\pm$	80.33	Charged Weak Force
Z Boson	Z^0	91.187	Neutral Weak Force
Gluon	G	0	Strong Force
Higgs	H	$60 < M_H < 800$	Spontaneous Symmetry Breaking

Table 1.3: Quantum numbers of the fermion spectrum

Chirality	Q	T_W^3	Y	C
ν_{eL}	0	1/2	-1	0
e_L	-1	-1/2	-1	0
u_L	2/3	1/2	1/3	r,g,b
d_L	-1/3	-1/2	1/3	r,g,b
e_R	-1	0	-2	0
u_R	2/3	0	4/3	r,g,b
d_R	-1/3	0	-2/3	r,g,b

1.1 Spin-$\frac{1}{2}$ Fermions and Spin-1 Gauge Bosons

Elementary particles consist of three kinds: leptons, quarks and gauge bosons. Leptons and quarks are spin-$\frac{1}{2}$ fermions. There are six leptons in three generations. Similarly, there are six flavors of quarks in three generations. Under $SU(3)_C$ symmetry, each quark flavor is in a triplet representation, which is labeled as red(r), green(g), or blue(b) color charge:

$$\Psi_q = \begin{pmatrix} q_r \\ q_g \\ q_b \end{pmatrix}. \tag{1.1}$$

Because of the similarity between three generations, only the quantum numbers of the first generation are listed in Table 1.3. The following descriptions, although mentioning only the first generation, can be applied to the second and third generations as well.

Under the $SU(2)_L$ transformation, the left-handed fermions transform as weak isospin (T_W) doublets:

$$\ell_L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \quad q_L = \begin{pmatrix} u \\ d \end{pmatrix}_L, \tag{1.2}$$

whereas right-handed fermions transform as singlets,

$$(e)_R, \quad (u)_R, \quad (d)_R. \tag{1.3}$$

There is no right-handed neutrino. This is allowed because of the massless nature of the Dirac fermion. All leptons and quarks have anti-particles with the quantum number signs resersed.

There are generators associated with each group of the SM and each generator is associated with one gauge boson. Therefore, the $SU(3)_C$ sector of the SM has eight $(3 \times 3 - 1 = 8)$ gauge bosons $G_\mu^a, a = 1, 2, \ldots, 8$, called gluons. $SU(2)_L$ has three $(2 \times 2 - 1 = 3)$ gauge bosons, two of which are charged $SU(2)_L$ gauge bosons $W_\mu^\pm$, and one which is the neutral $SU(2)_L$ gauge boson W_μ^3. $U(1)_Y$ has one gauge boson, the neutral hypercharge gauge boson B_μ.

The W_μ^3 and B_μ gauge bosons are mixed in nature via the following rotation,

$$\begin{pmatrix} Z_\mu^0 \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix}, \tag{1.4}$$

where θ_W, the weak mixing angle, is defined in such a way that A_μ would be the photon field in quantum electrodynamics (QED) which only interacts with electric charged particles; Z_μ^0 is a neutral weak field.

Therefore, there are in total 60 (12 leptons, 36 quarks, and 12 gauge bosons) elementary particles in the SM, and in addition, at least one Higgs boson which will be introduced in a later section.

1.2   Interactions of Fermions and Gauge Bosons

For a massless free fermion (a lepton or a quark) field Ψ, the Lagrangian can be taken as the normal Dirac kinetic energy Lagrangian with ∂_μ replaced by the covariant derivative D_μ,

$$\mathcal{L}_{FK} = \overline{\Psi} i \gamma^\mu \partial_\mu \Psi \rightarrow \overline{\Psi} i \gamma^\mu D_\mu \Psi, \tag{1.5}$$

to preserve the gauge invariance under $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$, where the covariant derivative is

$$D_\mu = \partial_\mu - i g_1 \frac{Y}{2} B_\mu - i g_2 \frac{\tau^j}{2} W_\mu^j - i g_3 \frac{\lambda^a}{2} G_\mu^a. \tag{1.6}$$

The B_μ term represents the $\mathrm{U}(1)_Y$ symmetry. Y, the hypercharge, is the generator of this symmetry. The coupling strength g_1 has to be measured experimentally. The W_μ^j ($j = 1, 2, 3$) term represents the $\mathrm{SU}(2)_L$ symmetry. It acts only on the $\mathrm{SU}(2)_L$ doublets. The τ^j, Pauli matrices, are the generators of the symmetry and $T_W^j = \frac{\tau^j}{2}$ is the Weak isospin. g_2 is the coupling strength. g_1 and g_2 are related through the weak mixing angle θ_W in a simple form,

$$\frac{g_1}{g_2} = \tan \theta_W. \tag{1.7}$$

And $g_2 \sin \theta_W = e$ is related to the fine structure constant α by $\alpha = e^2/4\pi$. Finally, the G_μ^a ($a = 1, 2, \ldots, 8$) term represents the $\mathrm{SU}(3)_C$ symmetry. It acts on quarks, the colored fermions, instead of leptons. the λ^a's are the generators of the symmetry which are (3×3) matrices. The coupling strength $g_3 = g_s$ is universal for all colored fields and has to be measured experimentally. Similar to the α in QED, the coupling constant α_s in QCD is defined as $\alpha_s = g_s^2/4\pi$.

In addition to the interactions between fermions and gauge bosons introduced above, gauge bosons can have self interactions. The kinetic term of the Lagrangian

for gauge bosons can be written as

$$\mathcal{L}_{GK} = -\frac{1}{4}\mathcal{B}_{\mu\nu}\mathcal{B}^{\mu\nu} - \frac{1}{4}\mathcal{W}^i_{\mu\nu}\mathcal{W}^{i\mu\nu} - \frac{1}{4}\mathcal{G}^a_{\mu\nu}\mathcal{G}^{a\mu\nu}, \qquad (1.8)$$

where

$$\mathcal{B}_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \qquad (1.9)$$

$$\mathcal{W}^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g_2 \epsilon^{ijk} W^j_\mu W^k_\nu, \qquad (1.10)$$

$$\mathcal{G}^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_3 f^{abc} G^b_\mu G^c_\nu. \qquad (1.11)$$

ϵ^{ijk} and f^{abc} are defined through the following anti-commutation relations,

$$\left[\frac{\tau^i}{2}, \frac{\tau^j}{2}\right] = i\,\epsilon^{ijk}\frac{\tau^k}{2}, \quad i,j,k = 1,2,3 \qquad (1.12)$$

and

$$\left[\frac{\lambda^a}{2}, \frac{\lambda^b}{2}\right] = i\,f_{abc}\frac{\lambda^c}{2}, \quad a,b,c = 1,2,\ldots,8 \qquad (1.13)$$

for SU(2) and SU(3), respectively.

1.3 Masses and the Higgs Mechanism

So far, the constructed Lagrangians do not contain the masses of the particles in the SM. The Higgs mechanism is developed in order to include the possibility that particles have mass in a gauge invariant way. However, there is a price to pay. Although the $SU(2)_L \times U(1)_Y$ symmetry is preserved for the Lagrangian in the Higgs mechanism, the symmetry for the ground state, *i.e.* the vacuum, is broken. That means the $SU(2)_L$ and $U(1)_Y$ quantum numbers of the vacuum are non-zero. This mechanism is called spontanenous symmetry breaking.

The Lagrangian for the Higgs sector is

$$\mathcal{L}_\Phi = (D_\mu\Phi)^\dagger(D^\mu\Phi) - \frac{\lambda}{2}(\Phi^\dagger\Phi)^2 - \mu^2(\Phi^\dagger\Phi), \qquad (1.14)$$

with

$$D_\mu \Phi = \left(\partial_\mu - i g_1 \frac{Y}{2} B_\mu - i g_2 \frac{\tau^j}{2} W_\mu^j \right) \Phi \qquad (1.15)$$

and

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} v + H + i\phi^0 \\ i\phi^1 - \phi^2 \end{pmatrix} = \begin{pmatrix} \frac{v + H + i\phi^0}{\sqrt{2}} \\ i\phi^- \end{pmatrix}, \qquad (1.16)$$

where H, ϕ^0, ϕ^1 and ϕ^2 are four real scalar fields; the field H is the physical Higgs field; the fields ϕ^0 and $\phi^\pm = (\phi^1 \mp i\phi^2)/\sqrt{2}$ are the unphysical would-be Goldstone bosons; the constant v is the vacuum expectation value of Φ.

It is easy to see from the Lagrangian that the minimum of the potential energy occurs at

$$\Phi^\dagger \Phi = -\frac{\mu^2}{\lambda} = \frac{v^2}{2} \qquad (1.17)$$

with λ and μ^2 taken to be positive and negative, respectively. The Higgs mass, identified by the term $M_H^2 H^2/2$ in the potential energy, is given by

$$M_H = v\sqrt{\lambda}. \qquad (1.18)$$

Let us carry out the algebra further. When Φ gets a vacuum expectation value, the Lagrangian contains extra terms

$$\Phi^\dagger \left(i g_1 \frac{Y}{2} B_\mu + i g_2 \frac{\tau^j}{2} W_\mu^j \right)^\dagger \left(i g_1 \frac{Y}{2} B_\mu + i g_2 \frac{\tau^j}{2} W_\mu^j \right) \Phi. \qquad (1.19)$$

With $Y = -1$ and Φ at the vacuum, it can be simplified as the terms

$$\left(\frac{v g_2}{2} \right)^2 W_\mu^+ W^{-\mu} + \frac{1}{2} \left(\frac{v g_2}{2 \cos \theta_W} \right)^2 Z_\mu^0 Z^{0\mu}. \qquad (1.20)$$

For the charged W, the neutral Z^0 and the photon γ, the expected mass terms in the Lagrangian would be

$$(M_W^2 W_\mu^+ W^{-\mu}) + (M_Z^2 Z_\mu^0 Z^{0\mu})/2 + (M_\gamma^2 A_\mu A^\mu)/2. \qquad (1.21)$$

Thus, we conclude that

$$M_W = \frac{1}{2}g_2 v, M_Z = \frac{M_W}{\cos\theta_W} \text{ and } M_\gamma = 0. \tag{1.22}$$

Consider a quantity

$$\rho = \frac{M_W}{M_Z \cos\theta_W}. \tag{1.23}$$

Equation 1.22 and Equation 1.23 show that the SM predicts $\rho = 1$. It would be a signal of new physics if the measurement of ρ deviates from 1.

With the Higgs field contructed as an SU(2) doublet, we can write an SU(2) invariant interaction of fermions with the Higgs field. For the first generation

$$\mathcal{L}_{fH} = g_u \, \bar{q}_L \, \Phi \, u_R + g_d \, \bar{q}_L \, \Phi_c \, d_R + g_e \, \bar{l}_L \, \Phi_c \, e_R + \text{hermitian conjugate} \tag{1.24}$$

where q_L and l_L are defined in Equation 1.2 and

$$\Phi_c = -i\tau_2 \Phi^* = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \Phi^* = \begin{pmatrix} i\phi^+ \\ \frac{v+H+i\phi^0}{\sqrt{2}} \end{pmatrix}. \tag{1.25}$$

Through similar arguments as for obtaining the gauge boson masses, we can accommodate non-zero fermion masses:

$$m_u = \frac{g_u v}{\sqrt{2}}, m_d = \frac{g_d v}{\sqrt{2}} \text{ and } m_e = \frac{g_e v}{\sqrt{2}}. \tag{1.26}$$

Since g_u, g_d, and g_e are arbitrary parameters in Equation 1.24, the masses have to be measured experimentally. The second and third generation fermion masses can be generated in the same way.

1.4   The Cabibbo-Kobayashi-Maskawa Mixing Matrix

In the electroweak section of the SM, both the quarks and leptons are assigned to left-handed doublets and right-handed singlets. In general, the quark mass eigenstates

are not the same as the quark weak eigenstates. The quark mixing matrix (CKM) is introduced by Kobayashi and Maskawa who generalized the Cabibbo four-quark case [4, 5, 6]. By convention, the three charge 2/3 quarks (u, c, and t) are unmixed, *i.e.* their weak eigenstates are the same as their mass eigenstates. All the mixing is expressed in terms of a 3×3 unitary matrix V operating on the charge $-1/3$ quarks (d, s, and b):

$$\begin{pmatrix} d \\ s \\ b \end{pmatrix}_{\text{Weak}} \equiv \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_{\text{Mass}} \tag{1.27}$$

or in a "standard" parametrization advocated by the Particle Data Group [3]

$$V = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{13}} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{-i\delta_{13}} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{-i\delta_{13}} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{-i\delta_{13}} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{-i\delta_{13}} & c_{23}c_{13} \end{pmatrix}. \tag{1.28}$$

Here $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$, with i and j being "generation" labels, $\{i, j = 1, 2, 3\}$. It reduces to the usual Cabibbo mixing of the first two generations with θ_{12} identified with the Cabibbo angle in the limit $\theta_{13} = \theta_{23} = 0$ where the third generation decouples.

The symmetries of charge conjugation C and parity P hold for strong and electromagnetic interactions, but are violated in the weak interactions. The combined CP symmetry is better preserved, however, it is still violated in the SM as a result of a single phase entering the CKM matrix (*e.g.* the CP violating phase, δ_{13}).

1.5 Comments

In spite of the remarkable descriptive power of the SM, there are some unsettled problems in the theory. Experimentally we have never observed single quarks. What

we measure are mesons and baryons, or many mesons and baryons in jets. The confinement of quarks is not well understood. Though the CKM weak mixing matrix provides a good phenomenological description of data, it gives no satisfactory theoretical explanation of how or why quarks are mixing in that way, nor of the origin of the CP violation. There are 18 parameters in the SM and at least 61 elementary particles as mentioned earlier. Although the SM may be a "true" theory, it is far from being an "ultimate" and "beautiful" theory that physicists are dreaming of.

Chapter 2

The QCD Parton Model

The parton model [7, 8] was first introduced to interpret the observation of scaling
[9] as a consequence of pointlike constituents in the proton (and hadrons in general).
These pointlike contituents are called partons, which are the quarks and gluons of
QCD. The parton model states that any observed hadron is made up of pointlike
constituent particles, its partons. At high energy, every relevant parton i participating
in the hard scattering from an initial-state hadron has momentum $x_i p^\mu$, where x_i is
the momentum fraction carried by parton i and p^μ is the momentum of the parent
hadron (therefore, $0 \leq x_i \leq 1$). The masses of hadrons and partons are neglected and
the cross sections are calculated from the tree level diagrams for partonic scattering
in the naive parton model. In the improved QCD parton model, one can carry out
higher order calculations in α_S. QCD is a non-Abelian gauge theory and its running
coupling α_S is asymptotically free and can be written as

$$\alpha_S(\mu^2) = \alpha_{S0}(\mu^2)\left[1 - \alpha_{S0}(\mu^2)\frac{\beta_2 ln ln(\mu^2/\Lambda_{QCD}^2)}{\beta_1^2 ln(\mu^2/\Lambda_{QCD}^2)} + O(\alpha_{S0}^2(\mu^2))\right], \qquad (2.1)$$

where μ is the renormalization scale; Λ_{QCD} sets the scale for the running coupling;

$$\alpha_{S0}(\mu^2) = \frac{4\pi}{\beta_1 ln(\mu^2/\Lambda_{QCD}^2)} \quad \text{(leading order)}. \qquad (2.2)$$

The beta functions β_1 and β_2 are calculated to be

$$\beta_1 = (11N_c - 2n_f)/3 \quad \text{and} \quad \beta_2 = 102 - 38n_f/3; \qquad (2.3)$$

11

where N_c is the number of colors and n_f the number of flavors.

The QCD parton model has become an essential tool to realize all high energy processes in the Standard Model and in the search for "New Physics". In high energy scattering processes involving at least one hadron in the initial state, the factorization theorem [10] provides the foundation to separate the hard cross section which is calculable, and parton distributions which are not calculable, as shown in Figure 2.1. A master (factorization) formula can be written as
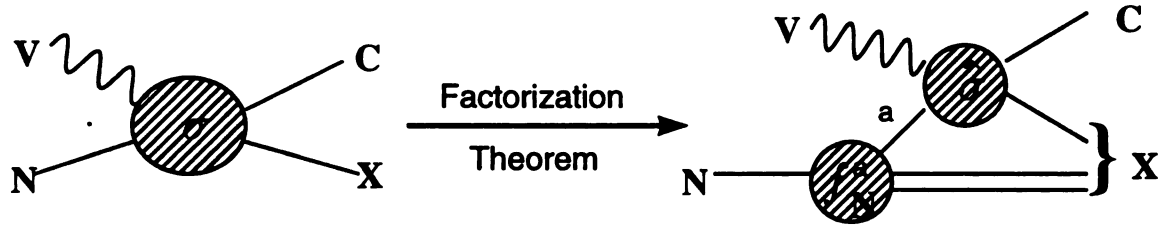
Figure 2.1: Factorization theorem separates the perturbatively calculable hard cross section $\hat{\sigma}$ and the universal parton distributions f_N^a.

$$\sigma_{VN \to CX}(s, Q) = \sum_a f_N^a(x, \mu) \otimes \hat{\sigma}_{Va \to CX}(\hat{s}, Q, \mu), \qquad (2.4)$$

where N denotes a hadron in the initial state; V, the probing particle (which could be γ^*, $W^\pm$, ...); C, a final-state particle if in a semi-inclusive case; X, a set of inclusive or semi-inclusive final states consisting of ordinary or new particles; "a", the parton label; $f_N^a(x, \mu)$, the parton distribution at the factorization scale μ; $\hat{\sigma}_{Va \to CX}$, the perturbatively calculable hard cross-section; and the parton label "a" is to be summed over all possible *active parton species*. "Active" partons, according to this widely accepted credo, include all quanta which can participate effectively in the dynamics at the relevant energy scale [11, 12], here denoted generically by Q (*e.g.*, Q in deep inelastic scattering, p_t in direct photon or jet production, ...)[1].

---

[1]These processes will be discussed in later sections.

Although only the hard cross section $\hat{\sigma}$ is calculable perturbatively, the parton distribution $f_N^a(x, \mu)$ is universal[2], *i.e.*, process independent. Moreover, the parton distributions are governed by the renormalization group equations:

$$\frac{d}{d\mu^2} q_i(x, \mu^2) = \frac{\alpha_S(\mu^2)}{2\pi} \int_x^1 \frac{d\xi}{\xi} \left[q_i(\xi, \mu^2) P_{qq}(\frac{x}{\xi}) + G(\xi, \mu) P_{qG}(\frac{x}{\xi}) \right] + O(\alpha_S^2(\mu^2)) \quad (2.5)$$

for quark (or antiquark) distributions q_i, and

$$\frac{d}{d\mu^2} G(x, \mu^2) = \frac{\alpha_S(\mu^2)}{2\pi} \int_x^1 \frac{d\xi}{\xi} \left[\sum_i q_i(\xi, \mu^2) P_{Gq}(\frac{x}{\xi}) + G(\xi, \mu) P_{GG}(\frac{x}{\xi}) \right] + O(\alpha_S^2(\mu^2))$$

$$(2.6)$$

for gluon distributions G, where $\alpha(\mu^2)$ is defined in Equation 2.1; $P_{qG}(x)$, $P_{Gq}(x)$, and $P_{GG}(x)$ are the calculable splitting functions [14, 15, 16]. That is to say that by knowing Λ_{QCD} and the parton distributions at the scale μ_0, we are able to calculate the parton distributions at any scale μ ($\mu_0 \leq \mu$) perturbatively by solving the above equations. And in turn we are able to obtain the theoretical predictions through Equation 2.4, the factorization theorem. To demostrate the application of the factorization theorem, we will briefly introduce the leading order (LO) results of some processes in the following sections. The full next to leading order (NLO) results can be found elsewhere.

2.1   Deeply Inelastic Scattering

Deeply inelastic scattering (DIS) is the simplest process to study hadron structures through the QCD parton model. It is generically of the form

$$l(k) + h(p) \rightarrow l'(k') + X , \qquad (2.7)$$

---

[2]However, the factorization scheme dependence will come in beyond leading order. It relates to exactly how the colinear singularity is absorbed into the parton distributions. Modified minimal substraction ($\overline{MS}$) scheme is the conventional choice and the default in our discussion, whereas deeply inelastic scattering (DIS) scheme is also used due to its simplicity in dealing with DIS process (see Ref. [13] for a detailed discussion).

where $l(k)$ represents an initial-state lepton of momentum k^μ, $h(p)$ an initial-state hadron of momentum p^μ, $l'(k')$ a final-state lepton of momentum k'^μ, and X an arbitrary hadronic final state. The lepton l and the hadron h interact through the exchange of vector boson V which can be a γ, $W^\pm$, or Z, as illustrated in Figure 2.2. The reaction is deeply inelastic if the magnitude of momentum transfer is large, $i.e.$
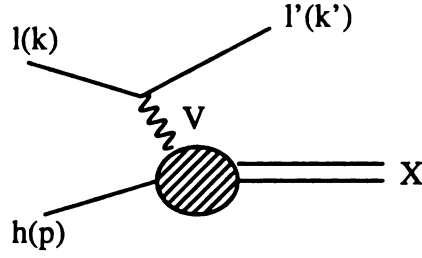
Figure 2.2: Deeply inelastic scattering.

$$Q^2 = -q^2 >> p^2 \ , \tag{2.8}$$

where

$$q^\mu = k^\mu - k'^\mu \ . \tag{2.9}$$

There is a set of kinematic variables used by convention. The momentum transfer has been defined in Equation 2.8 and Equation 2.9. The hard scattering as well as the parton distributions are usually parametrized in terms of the Bjorken scaling variable x,

$$x = \frac{-q^2}{2p \cdot q} = \frac{Q^2}{2m_h \nu} \ , \tag{2.10}$$

where ν is the energy transferred from the lepton to the hadron (with mass m_h) in the hadron (usually the target) rest frame,

$$\nu = \frac{p \cdot q}{m_h} = E_k - E_{k'} \ . \tag{2.11}$$

Another useful dimensionless variable y is defined as

$$y = \frac{p \cdot q}{p \cdot k} = \frac{E_k - E_{k'}}{E_k} \ , \tag{2.12}$$

which measures the ratio of the energy transferred to the hadronic system to the total leptonic energy available in the hadron rest frame.

The cross section for this process, at lowest order in the electroweak interaction but to all orders in the strong interaction, is

$$d\sigma = \frac{d^3k'}{2s|k'|}\frac{c_V^4}{4\pi^2(q^2 - m_V^2)^2}L_{lV}^{\mu\nu}(k,q)W_{\mu\nu}^{Vh}(p,q) \ , \tag{2.13}$$

where c_V is the coupling constant; $L_{lV}^{\mu\nu}$, the leptonic tensor; $W_{\mu\nu}^{Vh}$, the hadronic tensor. The leptonic part and the hadronic part are separated; the only connection is momentum transfer q. The leptonic tensor can be evaluated explicitly from

$$L_{lV}^{\mu\nu}(k,q) = nTr[\not{k}\Gamma_{Vl}^{\mu}(\not{k} - \not{q})\Gamma_{Vl}^{\nu}] \ , \tag{2.14}$$

where Γ_{Vl} is the electroweak vertex connecting lepton l to vector V and the outgoing lepton l', but with the factor c_V^2 removed. To average over lepton spin, we set n equal to $1/2$ for unpolarized $e^{\pm}$ or $\mu^{\pm}$, and 1 for $\nu(\bar{\nu})$.

The hadronic tensor $W_{\mu\nu}^{Vh}$ can be expanded in terms of a set of six independent basis tensors

$$W_{\mu\nu}^{Vh} = -g_{\mu\nu}W_1 + \frac{p_\mu p_\nu}{m_h^2}W_2 - i\frac{\epsilon_{\alpha\beta\mu\nu}p^\alpha q^\beta}{2m_h^2}W_3$$

$$+ \frac{q_\mu q_\nu}{m_h^2}W_4 + \frac{p_\mu q_\nu + q_\mu p_\nu}{2m_h^2}W_5 + \frac{p_\mu q_\nu - q_\mu p_\nu}{2m_h^2}W_6 \ , \tag{2.15}$$

where the scalar coefficient functions $\{W_i\}$ are the invariant hadron structure functions. However, when the hadronic tensor contracts with the leptonic tensor, the hadron structure functions $\{W_4, W_5, W_6\}$ will not appear because those terms are proportional to $O(m_l^2/Q^2)$, as compared to $\{W_1, W_2, W_3\}$, which is neglected. Therefore, only $\{W_1, W_2, W_3\}$ are relevant. Alternatively the functions $\{W_i\}$ are usually replaced by the structure functions $\{F_i\}$ because they satisfy scaling properties, where

$$F_1(x, Q^2) = W_1(x, Q^2) \ , \tag{2.16}$$

$$F_2(x, Q^2) = \frac{\nu}{m_h} W_2(x, Q^2) \ , \tag{2.17}$$

$$F_3(x, Q^2) = \frac{\nu}{m_h} W_3(x, Q^2) \ . \tag{2.18}$$

The differential cross section can then be expressed in terms of the dimensionless variables x and y and the structure functions $\{F_i\}$,

$$\frac{d\sigma}{dxdy} = N^V \left[\frac{y^2}{2} 2x F_1 + (1 - y - \frac{m_h xy}{2E}) F_2 + \delta^V (y - \frac{y^2}{2}) x F_3 \right] \ , \tag{2.19}$$

where δ^V is ± 1 for $W^\pm$ exchange and zero for the photon exchange, and where

$$N^\gamma = 8\pi\alpha^2 \frac{m_h E}{Q^4}, \tag{2.20}$$

$$N^{W^\pm} = \pi\alpha^2 \frac{m_h E}{2 \sin^4 \theta_W (Q^2 + M_W^2)^2} \ . \tag{2.21}$$

By applying the QCD parton model, the structure functions $\{F_i\}$ can be written in terms of parton densities and hard scattering cross section of partons,

$$F_{1,3}^{Vh}(x, Q^2) = \sum_a \int_0^1 \frac{d\xi}{\xi} \hat{F}_{1,3}^{Va}(\frac{x}{\xi}, Q^2, \mu^2) f_{a/h}(\xi, \mu^2) \ , \tag{2.22}$$

$$F_2^{Vh}(x, Q^2) = \sum_a \int_0^1 d\xi \hat{F}_2^{Va}(\frac{x}{\xi}, Q^2, \mu^2) f_{a/h}(\xi, \mu^2) \ . \tag{2.23}$$

Here the $\{\hat{F}_i^{Va}\}$ are the structure functions at the parton level; a, the active parton labelled a; $f_{a/h}$, the distribution of parton a in hadron h; μ, the factorization scale. For the electromagnetic structure functions at the leading order approximation we have

$$2\hat{F}_1^{\gamma a} = \hat{F}_2^{\gamma a} = Q_a^2 \delta(1 - x) \ , \tag{2.24}$$

and

$$\hat{F}_3^{\gamma a} = 0 \ (\text{to all orders}) \ , \tag{2.25}$$

where Q_a is the electric charge of the parton a. Substituting the above results into Equation 2.22 and Equation 2.23, we obtain simple results in terms of parton densities,

$$2xF_1^{\gamma h}(x, Q^2) = F_2^{\gamma h}(x, Q^2) = \sum_a Q_a^2 x f_{a/h}(x, \mu^2) \ . \tag{2.26}$$

Explicitly for the proton struction function in e-p or μ-p collisions, we simply have

$$F_2^{ep}(F_2^{\mu p}) = \frac{x}{9}[4(U + \bar{U}) + (D + \bar{D})] \ , \tag{2.27}$$

where U is the sum of all the up-type quark distributions ($i.e.$, u, c, and t distributions if active) inside the proton and D is the sum of all the down-type quark distributions ($i.e.$, d, s, and b distributions)[3]. For (anti-)neutrino scattering at an isoscalar target N ($e.g.$, a deuteron), we have

$$F_2^{\nu N} = F_2^{\bar{\nu} N} = x[(U + \bar{U}) + (D + \bar{D})] \ , \tag{2.28}$$

and

$$[F_3^{\nu N} + F_3^{\bar{\nu} N}]/2 = [U + D - \bar{U} - \bar{D}] \ . \tag{2.29}$$

The choice of the factorization scale μ^2 is theoretically to separate the long distance, "soft" part, from the short distance, "hard" part. In principle, if we are able to calculate all orders, the choice of scale is irrelevant. Since we only calculate up to finite orders, there is a slight dependence on the factorization scale. However, it makes physical sense to choose the scale of the order of Q^2 as the separation of soft and hard parts. Conventionally the scale μ^2 is chosen to be Q^2.

2.2   Vector Boson Production

Vector boson production (VBP) is a complementary process to probe the hadron structure. The production of virtual or real vector bosons, γ, $W^\pm$, Z, in hadron-hadron collisions can be generated by quark-antiquark annihilation at the order of

---

[3]In terms of equations, $U = u + d(+t)$ and $D = d + s(+b)$, where u is a shorthand for $f_{u/p}$, etc..

α_s^0, which is purely electroweak interactions. The gluons only come in at higher order. The generated vector boson then decays into a pair of leptons (e^+e^-, $\mu^+\mu^-$, $\mu^+\nu_\mu$, $e+\nu_e$, etc.) with large invariance mass. Since the production mechanism was orginally proposed and theoretically analyzed by Drell and Yan [17], it is also called Drell-Yan (DY) production. The process is expressed in the form

$$A(p) + B(p') \to V(q) + X \to l(k) + l'(k') + X \tag{2.30}$$

where initial-state hadron A and B interact and produce an intermediate vector boson V (which can be real or virtual) and then the vector boson decays to the lepton pair l and l', as shown in Figure 2.3. X labels all the undetected hadronic final states.
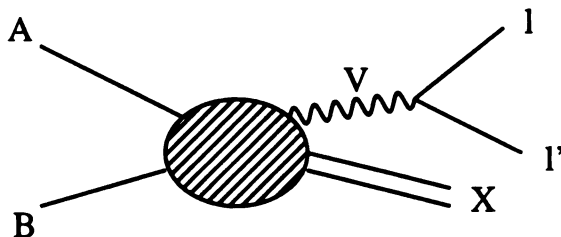
Figure 2.3: Vector boson production.

The inclusive DY cross section is of the form

$$\frac{d\sigma_{AB}^V}{dq^2}(p,p',q) = \sum_{ab} \int_0^1 d\xi \int_0^1 d\xi' f_{a/A}(\xi,\mu^2) f_{b/B}(\xi',\mu^2) \frac{d\hat{\sigma}_{ab}^V(\xi p, \xi' p', q)}{dq^2} . \tag{2.31}$$

We can reorganize in the form

$$\frac{d\sigma_{AB}^V}{dq^2} = \sigma_0^V W_{AB}^V , \tag{2.32}$$

where the factor σ_0^V contains the overall dimensions, while W_{AB}^V is the dimensionless function defined as the integral over the appropriate product of distribution functions times coupling

$$W_{AB}^V = \int_0^1 d\xi \int_0^1 d\xi' \delta(\tau - \xi\xi') D_{AB}^V(\xi,\xi',\mu^2) , \tag{2.33}$$

with the DY scaling variable

$$\tau = \frac{q^2}{s} \, , \tag{2.34}$$

and where s is the total center of mass energy

$$s = (p + p')^2 \, . \tag{2.35}$$

Let us first consider the production of a virtual photon followed by its decay into a lepton pair. Since $q = k + k'$ satisfies $q^2 = Q^2 > 0$, the virtual photon is timelike. q^2, the invariant mass of the lepton pair, is experimentally easily measurable. In the parton model interpretation, the DY process in the electromagnetic case is the annihilation of a quark-antiquark pair to produce the virtual photon. We have

$$\sigma_0^\gamma = \frac{4\pi\alpha^2}{3N_c q^2 s} \, , \tag{2.36}$$

and

$$D_{AB}^\gamma(\xi, \xi', \mu^2) = \sum_q Q_q^2 \left[f_{q/A}(\xi, \mu^2) f_{\bar{q}/B}(\xi', \mu^2) + f_{\bar{q}/A}(\xi, \mu^2) f_{q/B}(\xi', \mu^2) \right] \, , \tag{2.37}$$

which will give W_{AB}^γ. Here a natural factorization scale to choose is of the order of the invariant mass of the lepton pair, e.g., $\mu^2 \sim q^2$. Taking proton-antiproton collisions as an example, we can express $D_{p\bar{p}}^\gamma$ explicitly in terms of parton densities in a simple relation

$$D_{p\bar{p}}^\gamma(\xi, \xi', \mu^2) = \frac{1}{9} \left\{ \sum_{f_U = u,c,(t)} 4 \left[f_U(\xi, \mu^2) f_U(\xi', \mu^2) + \bar{f}_U(\xi, \mu^2) \bar{f}_U(\xi', \mu^2) \right] \right.$$

$$\left. + \sum_{f_D = d,s,b} \left[f_D(\xi, \mu^2) f_D(\xi', \mu^2) + \bar{f}_D(\xi, \mu^2) \bar{f}_D(\xi', \mu^2) \right] \right\} \, . \tag{2.38}$$

In the Z and $W^\pm$ boson productions, the results are slightly more complicated. In the case of Z, we have

$$\sigma_0^Z = \tau \frac{\pi\alpha^2}{192 N_c \sin^4\theta_W \cos^4\theta_W} \frac{1 + (1 - 4\sin^2\theta_W)^2}{(q^2 - M_Z^2)^2 + M_Z^2\Gamma_Z^2} \, , \tag{2.39}$$

where Γ_Z is the total width of the Z boson,

$$\Gamma_Z = \frac{\alpha M_Z}{24 \sin^2 \theta_W \cos^2 \theta_W}(1 - 4\sin^2 \theta_W - 8\cos^2 \theta_W) \, , \tag{2.40}$$

and

$$D_{AB}^Z(\xi, \xi', \mu^2) = \sum_q \left[1 + (1 - 4|Q_q|\sin^2 \theta_W)^2 \right]$$

$$\times \left[f_{q/A}(\xi, \mu^2)f_{\bar{q}/B}(\xi', \mu^2) + f_{\bar{q}/A}(\xi, \mu^2)f_{q/B}(\xi', \mu^2) \right] \, . \tag{2.41}$$

The corresponding results for $W^\pm$ are

$$\sigma_0^W = \tau \frac{\pi \alpha^2}{12 N_c \sin^4 \theta_W} \frac{1}{(q^2 - M_W^2)^2 + M_W^2 \Gamma_W^2} \, , \tag{2.42}$$

where the W width Γ_W is

$$\Gamma_W = \frac{\alpha M_W}{12 \sin^2 \theta_W} \, , \tag{2.43}$$

and

$$D_{AB}^{W^-}(\xi, \xi', \mu^2) = \cos^2 \theta_C \left[f_{\bar{u}/A}(\xi, \mu^2)f_{d/B}(\xi', \mu^2) + f_{\bar{c}/A}(\xi, \mu^2)f_{s/B}(\xi', \mu^2) \right]$$

$$+ \sin^2 \theta_C \left[f_{\bar{u}/A}(\xi, \mu^2)f_{s/B}(\xi', \mu^2) + f_{\bar{c}/A}(\xi, \mu^2)f_{d/B}(\xi', \mu^2) \right] + (A \leftrightarrow B) \, . \tag{2.44}$$

Here θ_W is the weak mixing angle (cf. Sec.1.1) and θ_C is the Cabibbo mixing angle (cf. Sec.1.4). $D_{AB}^{W^+}$ is the same as $D_{AB}^{W^-}$ only with the reversal of quarks and antiquarks in Equation 2.44.

As we can see in Equations (2.32 - 2.38), the parton distributions go into the formula in the form of scalar product, as compared to simple summation in Equations (2.26 - 2.29) of DIS. These different combinations of parton contributions provide complementary information to determine a unique and universal set of parton distributions.

2.3 Direct Photon Production

Up to now, the gluon distribution has not appeared in the leading order formula of DIS and DY processes. Direct photon production, however, gives a "direct" probing of the gluon content inside the hadron in the leading order diagram, as shown in Figure 2.4. Direct photon production is called direct because the photon is produced directly in

Figure 2.4: Leading order diagrams for direct photon production.

the hard scattering process in hadron-hadron collisions, not through the decay of the final states. The inclusive cross section of the process is given by

$$E_\gamma \frac{d^3\sigma}{dp_\gamma^3}(AB \to \gamma + X) = \sum_{a,b} \int dx_a dx_b f_{a/A}(x_a, \mu^2) f_{b/B}(x_b, \mu^2)$$

$$\times \frac{\hat{s}}{\pi} \frac{d\sigma}{d\hat{t}}(ab \to \gamma X)\delta(\hat{s} + \hat{t} + \hat{u}) , \qquad (2.45)$$

where hatted variables, $(\hat{s}, \hat{t}, \hat{u})$, refer to invariants of the partonic subprocess. In the leading order case, the direct photon can be produced by the QCD Compton subprocess $gq \to \gamma q$ and the quark-antiquark annihilation subprocess $q\bar{q} \to \gamma g$. The partonic cross sections are

$$\frac{d\sigma}{d\hat{t}}(gq \to \gamma q) = -\frac{\pi \alpha \alpha_s}{\hat{s}^2} \frac{e_q^2}{3}(\frac{\hat{u}}{\hat{s}} + \frac{\hat{s}}{\hat{u}}) \qquad (2.46)$$

and

$$\frac{d\sigma}{d\hat{t}}(q\bar{q} \to \gamma g) = \frac{\pi \alpha \alpha_s}{\hat{s}^2} \frac{8e_q^2}{9}(\frac{\hat{u}}{\hat{t}} + \frac{\hat{t}}{\hat{u}}) . \qquad (2.47)$$

The full next-to-leading-order $(\alpha\alpha_s^2)$ QCD calculations can be found in reference [18]. To visualize the relative importance of each subprocess contributions in different kinematic region, we plot the production rate at LO for both subprocesses in Figure 2.5 as an example. In $2 \to 2$ $(ab \to cd)$ process, the kinematics (x_a, x_b) of the incoming partons a and b can be expressed in terms of outgoing particles' transverse-x (x_t) and rapidity (y)

$$x_{a,b} = x_t e^{\pm y} , \tag{2.48}$$

where

$$x_t = 2p_t/\sqrt{s} \tag{2.49}$$

and

$$y = \frac{1}{2} ln(\frac{E + p_l}{E - p_l}) . \tag{2.50}$$

Here p_t is the transverse momentum of outgoing particle (i.e. photon in this case); s, the center of mass (CM) energy; E, the energy of outgoing particle in CM frame; p_l, the longitudinal momentum of outgoing particle in CM frame. Therefore, the horizontal axis x_t plotted in the figure is the central probing of x with kinematics (p_t, s). We see that the QCD Compton subprocess gives a significant contribution. It indicates large sensitivity to gluon distribution in this process.

Although direct photon production in hadron collisions is an important process to study in perturbative QCD due to the clean measurement of photons and its sensitivity to the initial state gluon, there are a number of theoretical uncertainties which affect the predictions of the normalization and slope of the measured direct photon p_t spectrum. These effects include: (1) the sensitivity of the theoretical calculations to the choice of factorization and renormalization scales [19]; (2) k_t broadening of the initial state partons due to soft gluon radiation [19]; and (3) photon fragmentation uncertainties [19] and the related issue of photon isolation cuts [20]. More theoretical
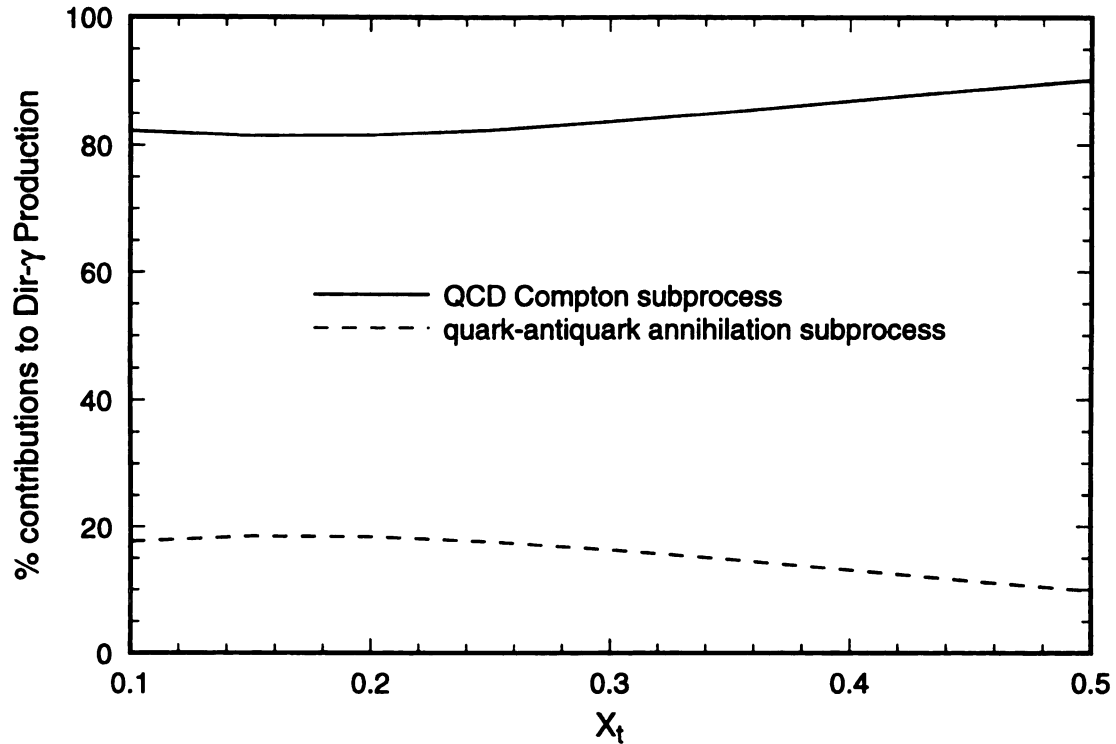
Figure 2.5: Percentage contributions to direct photon production at LO for the QCD Compton subprocess $gq \rightarrow \gamma q$ (solid line) and the quark-antiquark annihilation subprocess $q\bar{q} \rightarrow \gamma g$ (dashed line): ($x_t = 2p_t/\sqrt{s}$).

work needs to be done for a better understanding of the underlying physics.

2.4 Inclusive Jet Production in Hadron Collisions

Because of the property of QCD confinement, quarks and gluons produced in the hard scattering process have to hadronize to observable color neutral particles, which form jets. Since the hadronization process generates a spread of particles by soft gluon interaction with the scattered quarks or gluons, there are various algorithms for defining jets (e.g. cone algorithm, k_t-algorithm, ...) [21, 22] in order to make comparison between theoretically well-defined quantities and experimental measurables. In hadron collisions, jet production is the dominant hard scattering activity,

because it is purely strong interactions, whereas DY and direct photon productions are electroweak interactions. In the lowest order, jets are produced through $2 \to 2$ subprocesses. The relevant Feynman diagrams are shown in Figure 2.6. The cross
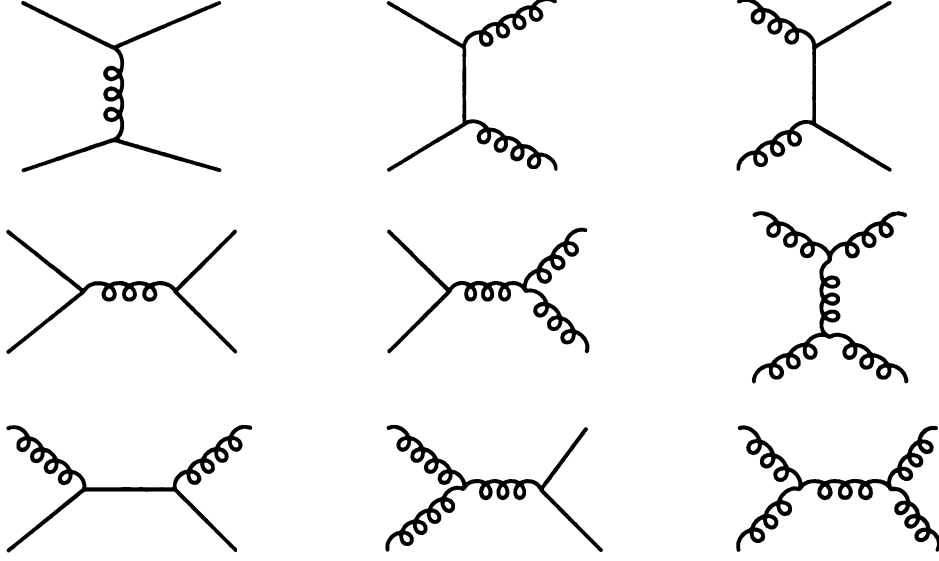
Figure 2.6: Leading order diagrams for jet production in hadron collisions.

section is of the form

$$\frac{d\sigma}{dp_t^2}(AB \to jets) = \sum_{abcd} \int dx_a dx_b f_{a/A}(x_a, \mu^2) f_{b/B}(x_b, \mu^2) \frac{d\hat{\sigma}}{dp_t^2}(ab \to cd) . \qquad (2.51)$$

where the transverse momentum p_t of either scattered parton relative to the beam axis is given by

$$p_t^2 = \frac{\hat{u}\hat{t}}{\hat{s}}, \qquad (2.52)$$

and where $\hat{s} = (a + b)^2$ is the subprocess center of mass energy squared, $\hat{t} = (a - c)^2$ and $\hat{u} = (a - d)^2$. The subprocess cross sections have the form

$$\frac{d\hat{\sigma}}{d\hat{t}}(ab \to cd) = \frac{|M|^2}{16\pi\hat{s}^2} , \qquad (2.53)$$

and $d\hat{\sigma}/d\hat{t}$ is related to $d\hat{\sigma}/dp_t^2$ by

$$\frac{d\hat{\sigma}}{dp_t^2} = \frac{d\hat{\sigma}}{d\hat{t}} \frac{\hat{s}}{\hat{t} - \hat{u}} . \qquad (2.54)$$

A list of the lowest order partonic scattering processes that contribute to two jet production are summarized in Table 2.1, together with their squared matrix elements averaged over spin and color. The full NLO QCD calculations have been carried out in the literature [23, 24]. To illustrate the relative importance of quark-quark, quark-gluon, and gluon-gluon scattering, we plot the percentage contributions at LO for jet production at the Tevatron ($\sqrt{s}$ = 1.8 TeV) in Figure 2.7. We see that gluon-gluon and gluon-quark scattering dominate the small-to-moderate transverse energy (E_t) [4] region, whereas quark-quark scattering contributes more in the large E_t region.
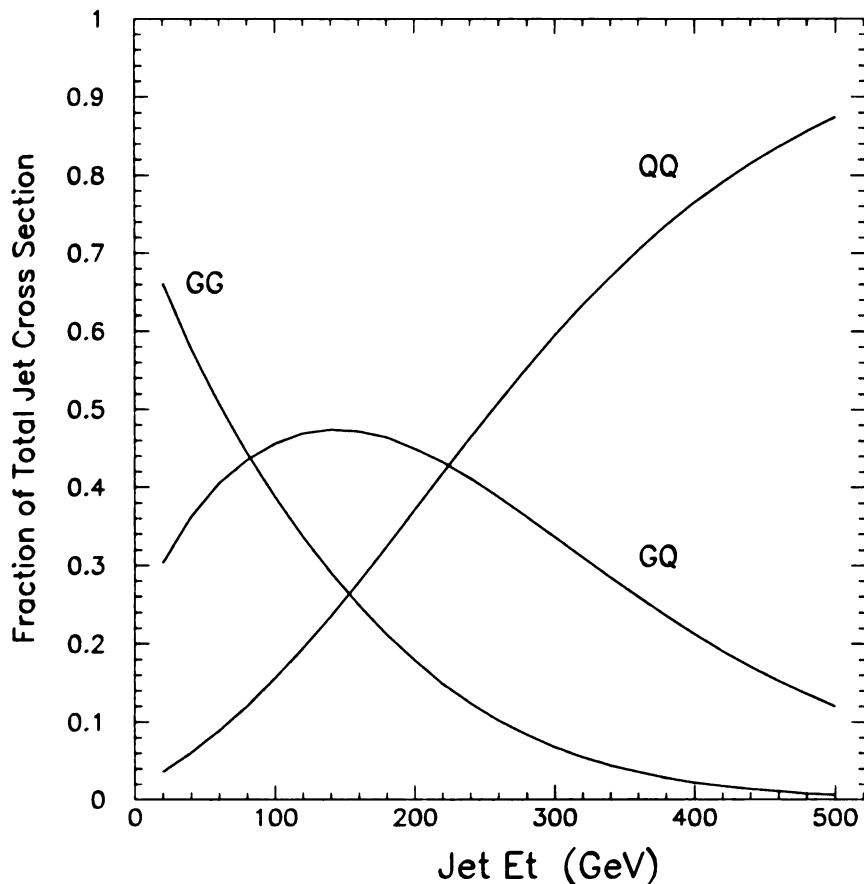
Figure 2.7: Relative contribution to the inclusive jet cross-section due to the various partonic subprocesses.

---

[4] $E_t = p_t$ in massless parton approximation.

Table 2.1: $2 \to 2$ subprocesses and their squared matrix elements: q and q' denote distinct flavors of quark.

Subprocesses	$\dfrac{|M|^2}{16\pi^2\alpha_s^2}$
$qq' \to qq'$	$\dfrac{4}{9}\dfrac{\hat{s}^2+\hat{u}^2}{\hat{t}^2}$
$q\bar{q}' \to q\bar{q}'$	$\dfrac{4}{9}\dfrac{\hat{s}^2+\hat{u}^2}{\hat{t}^2}$
$q\bar{q} \to q'\bar{q}'$	$\dfrac{4}{9}\dfrac{\hat{t}^2+\hat{u}^2}{\hat{s}^2}$
$qq \to qq$	$\dfrac{4}{9}\left(\dfrac{\hat{s}^2+\hat{u}^2}{\hat{t}^2}+\dfrac{\hat{s}^2+\hat{t}^2}{\hat{u}^2}\right)-\dfrac{8}{27}\dfrac{\hat{s}^2}{\hat{u}\hat{t}}$
$q\bar{q} \to q\bar{q}$	$\dfrac{4}{9}\left(\dfrac{\hat{s}^2+\hat{u}^2}{\hat{t}^2}+\dfrac{\hat{u}^2+\hat{t}^2}{\hat{s}^2}\right)-\dfrac{8}{27}\dfrac{\hat{u}^2}{\hat{s}\hat{t}}$
$q\bar{q} \to gg$	$\dfrac{32}{27}\dfrac{\hat{t}^2+\hat{u}^2}{\hat{t}\hat{u}}-\dfrac{8}{3}\dfrac{\hat{t}^2+\hat{u}^2}{\hat{s}^2}$
$gg \to q\bar{q}$	$\dfrac{1}{6}\dfrac{\hat{t}^2+\hat{u}^2}{\hat{t}\hat{u}}-\dfrac{3}{8}\dfrac{\hat{t}^2+\hat{u}^2}{\hat{s}^2}$
$qg \to qg$	$\dfrac{\hat{s}^2+\hat{u}^2}{\hat{t}^2}-\dfrac{4}{9}\dfrac{\hat{s}^2+\hat{u}^2}{\hat{s}\hat{u}}$
$gg \to gg$	$\dfrac{9}{4}\left(\dfrac{\hat{s}^2+\hat{u}^2}{\hat{t}^2}+\dfrac{\hat{s}^2+\hat{t}^2}{\hat{u}^2}+\dfrac{\hat{t}^2+\hat{u}^2}{\hat{s}^2}+3\right)$

Chapter 3

Global QCD Analysis and The CTEQ Program

The factorization theorem, based on the QCD parton model described in Chapter 2, provides the foundation for analyzing high energy hard scattering processes. There are two basic ingredients of calculations used for comparing theoretical predictions with experiments: (1) the perturbatively calculated scattering cross-sections involving the fundamental partons, leptons, and gauge bosons; and (2) the parton distributions inside the incoming hadrons [1]. The universal, i.e. process independent, parton distributions functions (PDF's) are derived from the analysis of data in a variety of hard scattering processes, but governed by the renormalization group equations. Early analyses were limited to deep inelastic lepton nucleon scattering and lepton pair production, as these were the processes for which extensive data sets were available and for which next-to-leading order (NLO) calculations of the hard scattering had been performed. Now the number of available NLO calculations has increased and, simultaneously, data for additional hard scattering processes have become available for a variety of beams and targets. This progress makes it possible to determine the parton

---

[1]If one or more particles are observed in the final state, appropriate fragmentation functions (partons fragmenting into hadrons) are needed. Due to the difficulty of parton identification and the complexity of QCD processes in the final state, the fragmentation functions are currently much less well known as compared to parton distributions.

distributions with a greater precision than was previously possible.

With the wealth of data and corresponding theoretical calculations, true "global analyses" have become possible. In such a program there are two main goals. The first is to determine the parton distributions as precisely as possible, using all available experimental input, and to suggest which new types of data are necessary in order to further improve the precision. Secondly, with an over-constrained set of PDF's it becomes possible to explore whether or not the parton level calculations in perturbative QCD constitute a consistent theoretical framework to account for all the available experimental data relevant for perturbative QCD studies. This may point to areas where improved theoretical treatments are required and, perhaps, uncover areas where various data sets used in the analysis might be mutually inconsistent. Either way, one can expect important progress to be made as the result of careful and critical confrontation of data with theory. Of no less significance is the need for reliable parton distributions for practically all calculations on high energy processes that involve hadrons for studying both SM and new physics.

3.1 What Is Involved in a "Global QCD Analysis"

Here we briefly lay out the necessary elements of performing a global QCD analysis, based on the theoretical framework discussed in Chapter 2.

- A well-defined physical measureable can be written in terms of the convolution of parton distributions and the hard cross sections by the factorization theorem; cf. Equation 2.4:

$$\sigma_{phy} = f \otimes \hat{\sigma} \ . \tag{3.1}$$

- The hard cross sections can be calculated order by order in α_s:

$$\hat{\sigma} = \sum_n \alpha_s^n \hat{\sigma}_n \ . \tag{3.2}$$

- Parton distributions evolve in μ according to the renormalization group equations; cf. Equations (2.5 - 2.6) :

$$\frac{d}{d\mu^2} f = \frac{\alpha_s}{2\pi} P \otimes f \ , \tag{3.3}$$

where the splitting function P is calculable order by order in α_s.

- Since the initial parton distributions are in a non-perturbative physics regime and not calculable, their initial conditions are parametrized at the scale $\mu = Q_0$ with certain functional forms: $f(\mu = Q_0, x) = f_0(x)$.

- Λ_{QCD} is needed for the calculation of α_s; cf. Equations (2.1 - 2.3).

With experiments on the one hand and parameter space (e.g. initial parton distribution parameters and Λ_{QCD}) on the other, based on QCD theory, one can perform a least χ^2 fit by adjusting parameters to obtain parton distributions and the correponding α_s in consistence with data. Most of the modern global analyses [25, 26, 27, 28, 29] use both the hard cross section $\hat{\sigma}$ and the splitting function P in NLO, unless otherwise mentioned.

In the following sections, we will describe the global analysis program performed by CTEQ collaboration in detail.

3.2 Experimental Input

In order to make the comparison of theory with experiment well-defined, we have limited the kinematic range to that where the "leading twist" QCD contributions are

dominant. Thus, we restrict the selection of experimental data to kinematic regions that contain at least one large momentum scale "Q" $> Q_{min}$. In the absence of a reliable theoretical guide in the perturbative formalism, the value Q_{min} is varied within the range $2 - 10$ GeV to test for sensitivity of the results to this choice. We found stable results generally with the following minimum kinematic constraints: for deep inelastic scattering, Q (virtuality of the vector boson) > 2 GeV and W (CM energy) > 3.5 GeV; for lepton-pair production, $Q > 4$ GeV; for direct photon production, $p_t > 4$ GeV; for jet production at Tevatron, $p_t > 50$ GeV.

The high statistics DIS data from NMC [30] on $F_2^n/F_2^p, F_2^p - F_2^n$, and $F_2^{p,d}$ using a muon beam and from CCFR [31] on $F_{2,3}^{F_e}$ using (anti-) neutrinos, combined with the existing BCDMS [32] results provide excellent coverage of the kinematic region $x > 0.01$. New measurements of F_2^p from HERA [33] have extended the kinematic range down to very small x values, approaching 10^{-4}. Although the errors were comparatively large when the results were first published, the extended range provides useful constraints on the behavior of the parton distributions in the small-x region. ("small-x" means $10^{-4} < x < 10^{-2}$.) This is particularly important in light of questions raised concerning the consistency of the structure functions measured in the other experiments in the intermediate region $0.01 < x < 0.1$ [25]. Moreover, with continuously accumulating luminosity, HERA data play a more and more decisive role in the overall determination of parton distributions. As samples to display the accuracy and range of the DIS data, data at $Q^2 = 12$ GeV^2 from BCDMS, NMC, and HERA are shown in Figs. (3.1 - 3.2).

Precision data from the SLAC-MIT series of experiments [34] largely lie outside our kinematic cuts (especially when the cuts are raised above the minimum quoted above); and those data points within the cuts agree well with the BCDMS and NMC data already included. They are thus not used in the analyses reported here. Data
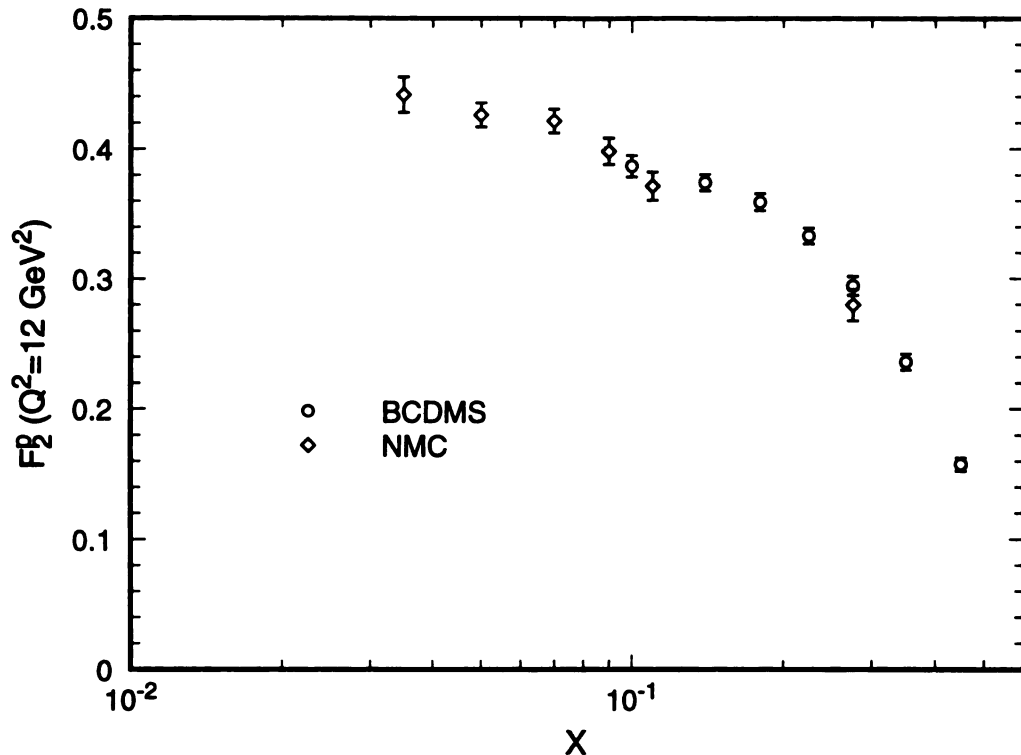
Figure 3.1: F_2^p data at $Q^2 = 12$ GeV^2 from BCDMS and NMC.

from the earlier EMC experiment [35] are excluded since the disagreement between these data with other data sets appears now to be understood as the result of the new NMC analysis. Data from the CDHSW neutrino experiment [36] are also not used since in the (wide) region where they agree with the CCFR results, the latter completely dominate due to much smaller errors; and in the (narrow) region where they disagree, it would be inconsistent to include both sets.

To apply the selected experimental results to the study of the parton structure of the nucleon, the heavy target neutrino data must be converted to their nucleon equivalent. This conversion is done using measured light to heavy target ratios obtained in electron and muon scattering processes by the SLAC-MIT [37], EMC [38], and NMC [39] experiments. There is an uncertainty associated with this procedure, which will be commented on later.
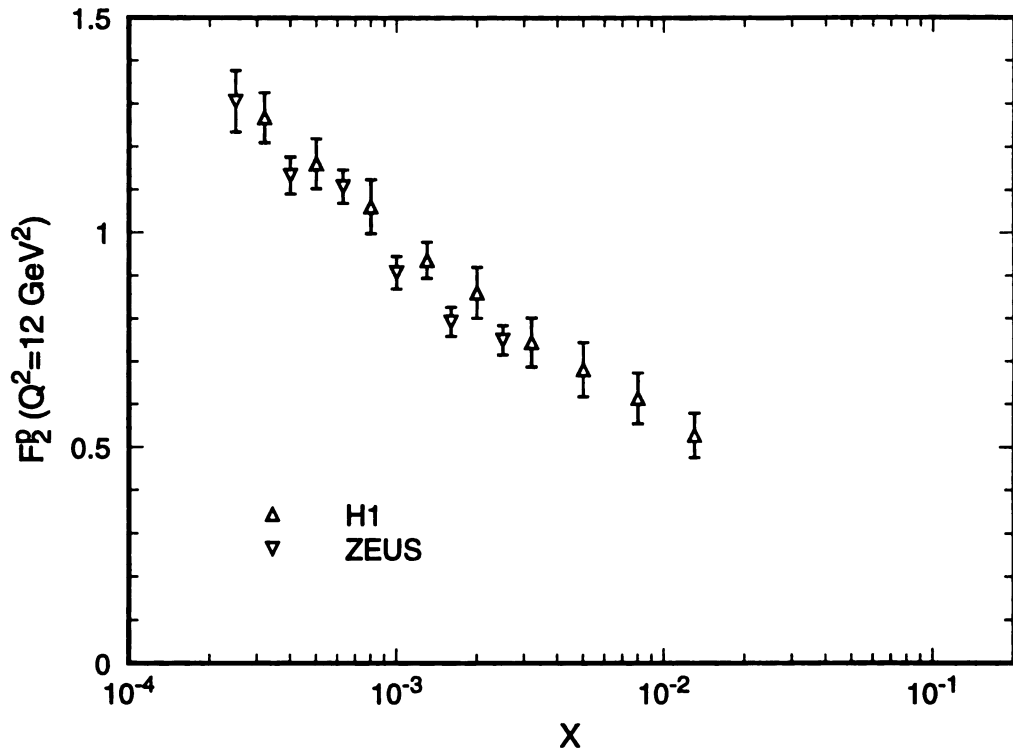
Figure 3.2: F_2^p data at $Q^2 = 12$ GeV^2 from HERA (H1 and ZEUS).

DIS data by themselves are not sufficient to provide a complete set of constraints on the parton content of the nucleon, since the measured nucleon structure functions represent only a few independent combinations of the parton flavors. As discussed in Sec.2.2, Lepton-pair production experiments provide complementary constraints on parton distributions, especially quark and antiquark distributions (through the $q - \bar{q}$ annihilation mechanism). From fixed-target experiments we use the full data set on the double-differential cross-section $d^2\sigma/d\tau dy$ measured by the high statistics E605 experiment at Fermilab [40]. We also include the new collider data on lepton-pairs measured by the CDF Collaboration [41]. Although the errors on these data are comparatively large, they do provide some constraints in the $x \sim 10^{-2}$ region which is beyond the reach of fixed-target experiments.

Two new types of data have become available and have provided valuable informa-

tion on PDF's, notably flavor differentiation of partons, which were not fully covered by earlier data sets. In particular, NA51 [42] measured the difference of cross sections for producing lepton pairs at $y = 0$ from proton and neutron targets. As discussed in [43], this is particularly sensitive to the difference of the $\bar{u}$ and $\bar{d}$ distributions. And the CDF Collaboration has presented new data on the charge asymmetry of the decay leptons in W production [44]. This measurement contributes to the differentiation of the valence u and d quarks as well as the sea-quarks. The effects of including these two data sets will be discussed further in Sec.3.4.

With copious data from DIS, which quarks are "directly" measured, and complementary data from the Drell-Yan process, the gluon distribution remains the major uncertainty. Direct photon production, which is particularly sensitive to the gluon distribution as indicated in Figure 2.5, provides another independent source of information. In addition to the commonly used WA70 data [45], we also include results from the UA6 [46] and E706 [47] experiments. Together these provide coverage of the region $0.27 < x < 0.54$ and, hence, help to constrain the gluon distribution in the middle range of x. The deep inelastic data provide some constraint on the gluon for smaller values of x through the slope of F_2 with respect to Q^2. Additional information at small values of x is provided by direct photon data from various collider experiments. Indeed, the coverage in x extends now down to about 0.02, making a simultaneous analysis of all of the available direct photon data a potentially powerful tool for extracting the gluon distribution. However, there are various theoretical uncertainties as discussed in Sec.2.3. When all those uncertainties are taken into account, existing direct photon data do not place as tight constraints on the gluon distribution as is commonly believed [48]. Full exploitation of the potential of this process in a QCD global analysis will require significant progress in understanding those uncertainties.

An important process that is sensitive to gluons is jet production in hadron-hadron collisions. At leading order, the cross-section is proportional to $\alpha_s^2(Q)\,G(x,Q)\,G(x',Q)$, $\alpha_s^2(Q)\,G(x,Q)\,q(x',Q)$, and $\alpha_s^2(Q)\,q(x,Q)\,q(x',Q)$ for the gluon-gluon, gluon-quark, and quark-quark scattering subprocesses respectively (for details see Sec.2.4). In Figure 2.7, we see that gluons are sensitive to the small-to-moderate E_t region and less sensitive but still sizable in the large E_t region. Experimental measurement of various inclusive jet cross-sections has progressed to an increasingly quantitative level in recent years. For instance, at the Tevatron, good data on single jet production are now available over a wide range of transverse energy, 15 GeV $< E_t < 450$ GeV [49, 50]. NLO QCD calculations of jet cross-sections have also reached a mature stage [23, 51, 52]. Many issues relating to jet definition (which is important for comparing theory with experiment) encountered in earlier stages of jet analysis have been extensively studied and are better understood. For the moderate to large E_t range, the scale dependence of the NLO inclusive jet cross section turns out to be relatively small [53]. Thus, it is natural that inclusive jet data should now be incorporated into a global QCD analysis, and that these data should play a role in constraining the gluon distribution $G(x,Q)$.

The full data sets we use are summarized in Table 3.1. The kinematic map spanning (x,Q) for all data used is shown in Figure 3.3.

3.3 CTEQ Global Analysis Procedures

Our goals in the global analysis program are two-fold. On the one hand, we are seeking a universal set of parton distributions which provide an accurate description of all of the data sets and are therefore suitable for use in the calculation of other high energy processes. On the other, we wish to determine to what degree the theoretical

Table 3.1: The data sets used in the CTEQ global analyses. Data sets marked with 1 in the final column were used for the CTEQ1 and later fits. Those with a 2, 3, or 4 were added for the CTEQ2, CTEQ3, and CTEQ4 analyses respectively. Data points correspond to the most recent data used in the CTEQ4 global analyses. The actual data points used in previous CTEQ1, CTEQ2, CTEQ3 analyses may be less for the same set of experiment due to continuous accumulation over the years. The column labelled $\Delta\sigma$ gives the overall normalization systematic error used in defining the χ^2, as discussed in the text.

Process	Experiment	Measurable	Data Points	$\Delta\sigma$	Set
DIS	BCDMS	$F_{2\ H}^{\mu}, F_{2\ D}^{\mu}$	324	.02	1
	NMC	$F_{2\ H}^{\mu}, F_{2\ D}^{\mu}, F_{2\ n/p}^{\mu}$	297	.02	1
	E665	$F_{2\ H}^{\mu}, F_{2\ D}^{\mu}$	70	.02	4
	H1	$F_{2\ H}^{e}$	172	.02	2
	ZEUS	$F_{2\ H}^{e}$	179	.02	2
	CCFR	$F_{2\ Fe}^{\nu}, x\ F_{3\ Fe}^{\nu}$	126	.02	1
Drell-Yan	E605	$sd\sigma/d\sqrt{\tau}dy$	119	.10	1
	CDF	$sd\sigma/d\sqrt{\tau}dy$	8	.07	2
	NA-51	A_{DY}	1	-	3
W-prod.	CDF	Lepton asym.	9	-	3
Direct γ	WA70	$Ed^3\sigma/d^3p$	8	.10	1
	E706	$Ed^3\sigma/d^3p$	8	.15	1
	UA6	$Ed^3\sigma/d^3p$	16	.10	1
Incl. Jet	CDF	$d\sigma/dE_t$	36	.05	4
	D0	$d\sigma/dE_t$	26	.06	4
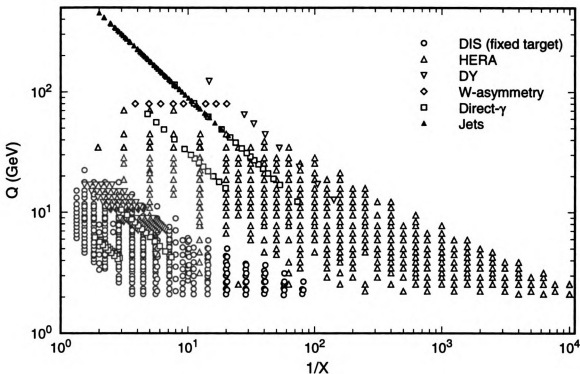
Figure 3.3: Kinematic map in the (x, Q) plane of data points used in the current global analysis.

treatment of the hard scattering processes in the pQCD framework is consistent with all the available experimental results.

To this end, except where otherwise noted, all data sets included in the analysis are treated on the same footing. This is to be contrasted with an often adopted procedure of emphasizing DIS data as the primary source of information (hence, performing a least-χ^2 fit to these data alone), using the other processes only as supplementary constraints. The simultaneous fitting of many different types of data necessitates the inclusion of both systematic as well as statistical errors. The systematic errors include both overall and point-to-point errors. The treatment of the latter poses a particularly difficult problem. The proper treatment of such errors typically differs from one experiment to another and doing this for all experiments requires a prohibitive amount of computer resources. We studied the impact on the global fit of

a full-scale treatment of the (correlated) systematic errors from the high statistics CCFR and BCDMS experiments compared to the common practice of combining the point-to-point systematic and statistical errors in quadrature. The difference is not significant. Thus, we use the latter procedure as an adequate compromise out of practical necessity. (Clearly, a fine-tuning of the final results, including a full treatment of the errors for selected data sets, is possible if necessary.)

The treatment of the overall normalization errors utilized in CTEQ analyses differs from that employed by other groups (including most early PDF's, see [54], and [55]) which usually allow all experimental data sets to be varied freely. In our analysis, with the exception of data which pertain to measured ratios, the normalization (fitting) parameter N_i for each data set i is associated with a fully correlated error ϵ_i given by the experiment: a term of the form $(1 - N_i)^2/\epsilon_i^2$ is then added to the overall χ^2 in the fitting process. This procedure properly takes into account the normalization uncertainties of the experiments, whereas the usual practice mentioned above technically corresponds to assuming infinite normalization errors for all experiments.

The hard cross-sections of all processes included in the analysis are calculated in pQCD to NLO in α_s. We use the $\overline{MS}$ scheme with 5 flavors as the standard; cf. Sec. 3.5 for more details. While such calculations are generally less sensitive than leading-order (LO) results to the choice of the renormalization and factorization scales (denoted jointly by the symbol μ), the residual dependence on these choices provides a potentially important source of theoretical uncertainty. In principle, this uncertainty is one order higher than the approximation used, i.e., next-to-next-to-leading order in our case. In practice, it has been learned that the size of the uncertainty is process-dependent. It is relatively small for DIS and for lepton-pair production and one usually chooses $\mu = Q$, the virtuality of the exchanged virtual vector boson, since this is the natural large scale in the problem. It is also small for inclusive jet production

at large E_t region and we choose $\mu = E_t/2$ in our analyses. On the other hand, the NLO predictions for direct photon production are still sensitive to the choice of μ. It is important to address this issue if quantitative results on the gluon distribution are to be extracted. The common practice of making a specific choice (say $\mu = p_T$) without discussion implicitly introduces a bias into the analysis because of the non-negligible μ-dependence. In this analysis, we have made the first attempt among PDF studies to address this issue by assigning a "theoretical error" to the predictions associated with the choice of μ. The size of this error is estimated by computing the range of predictions spanned by $\mu = p_T/2$ to $\mu = 2p_T$. During the process of fitting, we let the scale parameter μ for direct photon calculation float and add a contribution to the overall χ^2 due to scale uncertainty given by the deviation of μ from p_T divided by the "error" defined above. Although the details of this procedure (such as the central value for μ and the exact range used to estimate the error) may be the subject of debate, it nevertheless represents a reasonable treatment of the theoretical uncertainty which otherwise is simply ignored.

3.4 Relation between PDF's and Observables

The relationship between PDF's and the experimental input is in general quite involved since all parton flavors contribute to the NLO formulas for the hard cross-section; and, in addition, the parton distribution functions always mix as the result of QCD evolution. Nonetheless, simple leading order parton model formulas neglecting small sea-quark contributions are often useful in providing a qualitative guide to analysis strategies. We will review the most relevant relations, with the understanding that they are modified by NLO corrections in practice (to varying degrees for different processes).

Let us first consider deep inelastic scattering. The available high statistics data come in four different types. From Equations (2.26 - 2.29) of Sec.2.1, we have the expressions for those four types, in lowest order approximation with only u, d, and s quarks, given as follows.

$$
\begin{aligned}
F_2^{\mu p}(or F_2^{ep}) &= x[4(u+\bar{u})+(d+\bar{d})+2s]/9 \\
F_2^{\mu n} &= x[4(d+\bar{d})+(u+\bar{u})+2s]/9 \\
F_2^{\nu N} = F_2^{\bar{\nu}N} &= x[(u+\bar{u})+(d+\bar{d})+2s] \\
x[F_3^{\nu N}+F_3^{\bar{\nu}N}]/2 &= x[u+d-\bar{u}-\bar{d}]
\end{aligned}
\tag{3.4}
$$

As noted in [28], these four quantities can be used to extract four combinations of parton distributions, e.g., $u+\bar{u}$, $d+\bar{d}$, s, and $\bar{u}+\bar{d}$, or, equivalently, $u+d$. In particular, these four combinations are sufficient for examining the question of the breaking of SU(3) flavor symmetry (u, d, s) of the quark-antiquark sea. Utilizing the equations given above, the strange sea may be expressed as

$$
xs = \frac{5}{6}F_2^{\nu N} - 3F_2^{\mu N}.
\tag{3.5}
$$

Since the right-hand-side appears as a small difference between two much larger numbers, the relative uncertainty becomes large and, furthermore, is sensitive to the overall systematic errors of the experiments—still, in recent high precision experiments, the latter have been reduced to a level sufficient for the application of relation 3.5. A more direct measure of the strange quark sea is provided by the ν production of charm. Unfortunately, data on this process have not yet been made available in a form independent of experimental corrections. This issue will be discussed in the next chapter.

The question of SU(2) flavor symmetry (u, d) breaking in the sea is not directly addressed by the types of data listed above. Some information is provided by the

Gottfried integral which takes the form

$$I(a, b) = \int_a^b [F_2^{\mu p} - F_2^{\mu n}] \frac{dx}{x}. \tag{3.6}$$

The Gottfried sum rule with naive parton model and SU(2) flavor symmetry predicts $I(0,1) = 1/3$. The NMC Collaboration has measured [56] $I(.004, .8) = 0.236 \pm 0.008$. At lowest order one has

$$I(0,1) = \frac{1}{3} - \frac{2}{3} \int_0^1 (\bar{d} - \bar{u}) dx. \tag{3.7}$$

The experimental result cited above indicates that $\bar{d} > \bar{u}$ when integrated over x. However, information on the x dependence of this SU(2) breaking must be found from another source.

Lepton-pair production (LPP, or the Drell-Yan process) provides direct information on the anti-quark distributions as well as the sea difference between u and d quarks. For simplicity, consider the cross-section

$$\frac{d\sigma}{dQ^2 dy}|_{y=0} \tag{3.8}$$

for LPP in proton collisions on an isoscalar target, such as a deuteron. At lowest order, retaining only the light quark and antiquark contributions, this cross section is proportional to the following product of parton distributions; cf. Equation 2.37 of Sec.2.2:

$$\Sigma_{\mu\mu} = (4u + d)(\bar{u} + \bar{d}) + (4\bar{u} + \bar{d})(u + d) \tag{3.9}$$

where each of the distributions is evaluated at $x = Q/\sqrt{s}$. Note that all terms on the right-hand side are directly proportional to anti-quark distributions (in contrast to DIS where $\bar{q}(x)$ usually is submerged under $q(x)$ for a large part of the x-range). Equation 3.9 can be rewritten as

$$\Sigma_{\mu\mu} = 5(\bar{u} + \bar{d})(u + d) + \frac{3}{2}(\bar{u} + \bar{d})[(u + \bar{u}) - (d + \bar{d})] + \frac{3}{2}(\bar{u} - \bar{d})[(u + d) - (\bar{u} + \bar{d})]. \tag{3.10}$$

In principle, all of the terms except $(\overline{u} - \overline{d})$ are constrained by the deep inelastic data. Therefore, the lepton pair data provide a direct measure of the SU(2) breaking of the sea, i.e. $(\overline{u} - \overline{d})$, when used in conjunction with the deep inelastic data. In fact, the E-605 data on $d^2\sigma/dQ^2 dy$ used in this analysis cover a range in y (centered about zero). This provides even more information than the $y = 0$ case shown above – since the y-dependence extends the x range through the relation $x_{1,2} = \sqrt{Q^2/s}e^{\pm y}$ – but the principle is the same.

All the CTEQ analyses find substantial SU(2) breaking due to the use of the full range of DIS and LPP data. Since the E-605 data constrain the PDF's over a range in x covering approximately 0.10 - 0.6 (when the y-range is taken into account), the SU(2) breaking effects observed are reliable only over this range. To extend these results to lower values of x, additional experimental measurements will be needed.

The NA51 experiment[42] measured the asymmetry between the cross section for producing lepton pairs from proton and neutron targets, designed to probe directly the quantity $(\overline{u} - \overline{d})$. As shown in [43], this quantity can be written as

$$A_{DY} = \frac{(4u_v - d_v)(\overline{u} - \overline{d}) + (u_v - d_v)(4\overline{u} - \overline{d})}{(4u_v + d_v)(\overline{u} + \overline{d}) + (u_v + d_v)(4\overline{u} + \overline{d})} \qquad (3.11)$$

where the subscript v denotes a valence distribution. The NA51 result is $A_{DY} = -0.09 \pm 0.028$ at $y = 0$ and $Q/\sqrt{s} = 0.18$, where the statistical and systematic errors have been added in quadrature. Comparison with Equation 3.11 shows that since $u_v/d_v \approx 2$, one must have $\overline{u} < \overline{d}$. This is consistent with the sign of the breaking indicated by the Gottfried sum result.

Also of interest is the lepton charge asymmetry observed in W production by the CDF Collaboration. Consider the charge asymmetry of W production (before decaying into leptons), defined as

$$A_W(y) = \frac{d\sigma^+/dy - d\sigma^-/dy}{d\sigma^+/dy + d\sigma^-/dy} \qquad (3.12)$$

where the superscript denotes the charge of the W. For $\bar{p}p$ collisions in leading order parton model, $A_W(y)$ is given approximately by

$$A_W(y) \approx \frac{u(x_1)d(x_2) - d(x_1)u(x_2)}{u(x_1)d(x_2) + d(x_1)u(x_2)} \tag{3.13}$$

where $x_{1,2} = x_0 e^{\pm y}$ and $x_0 = M_w/\sqrt{s}$. Letting $R_{du} = d/u$, one can write

$$A_W(y) = \frac{R_{du}(x_2) - R_{du}(x_1)}{R_{du}(x_2) + R_{du}(x_1)}. \tag{3.14}$$

As noted in Ref.[43], in the region of small y (where $R_{du}(x_1) \approx R_{du}(x_2) \approx R_{du}(x_0)$) this asymmetry is directly proportional to the logarithmic slope of the ratio R_{du} in x:

$$A_W(y) \approx -x_0 y \frac{dR_{du}}{dx}(x_0)/R_{du}(x_0). \tag{3.15}$$

For the CDF experiment, $x_0 = 0.044$ and $|y| < 2$, thereby providing information on the ratio of the d and u distributions in the region of x of $(0.01, 0.2)$. Actual data on this process are for the corresponding decaying lepton asymmetry, so the above discussion is relevant only on the qualitative level since Equations (3.13-3.15) are considerably smeared when applied to the measured leptons.

As mentioned earlier, the cross-section of jet production at the Tevatron is proportional to $\alpha_s^2(Q) G(x, Q) G(x', Q)$ and $\alpha_s^2(Q) G(x, Q) q(x', Q)$ for the gluon-gluon and gluon-quark scattering subprocesses respectively. This process gives a direct contraint on the gluon distribution in the region $.05 < x < .5$. In addition to these simple parton model relations, some observables can be sensitive to parton distributions through NLO effects. Two examples come readily to mind: the precise data on DIS, especially HERA data, place important constraints on the gluon distribution $G(x, Q)$ in the small x region through the Q-dependence of the structure functions; and LPP data provide additional constraints on $G(x, Q)$ through the "Compton-scattering" mechanism. These examples caution us against taking simple parton relations too literally under all circumstances.

3.5 Choice of Parametrization

We now address the issue of the parametrization of the initial PDF's at some Q_0 which serves as the non-perturbative input to the global analysis, as described in Sec.3.1. The forms chosen must be flexible enough to account for all experimental input, if possible, yet they should not be under-constrained. Considering the current status of the experimental evidence as discussed above, the parametrization must allow for breaking of both SU(3) and SU(2) flavor symmetry. Our input parton distributions are parametrized at $Q_0 = 1.6$ GeV (which coincides which the charm threshold we use; see below). The Q-dependence of the parton distributions is generated by QCD-evolution using two-loop expressions for the splitting functions and running coupling; cf. Equations (2.1, 2.5, 2.6). In general, the $\overline{\text{MS}}$ factorization scheme is used although, in response to the need for DIS-scheme and leading-order (LO) calculations, we also generate equivalent parton distributions in these schemes. The heavy quark thresholds are taken as 1.6 and 5.0 GeV for the c and b quarks, respectively, and the heavy quark distributions are generated using massless evolution starting from a boundary condition of a vanishing PDF at the appropriate threshold equal to the corresponding quark mass. The renormalization scheme on which this definition of heavy quark parton distribution functions is based has been formulated precisely in references [57, 58] and a global analysis based on this scheme has been carried out recently [59]. In principle, it is possible to have non-zero heavy-quark distributions at threshold – $e.g.$ to have some "intrinsic charm", as has been suggested occasionally in the literature. We do not include this possibility for lack of positive experimental evidence at this time.

The functional forms used for the initial parton distributions in the four rounds of CTEQ analyses vary slightly. We give here the general expressions used in most

of the CTEQ analyses:

$$
\begin{aligned}
xu_v &= a_0^u x^{a_1^u}(1-x)^{a_2^u}(1+a_3^u x^{a_4^u}) \\
xd_v &= a_0^d x^{a_1^d}(1-x)^{a_2^d}(1+a_3^d x^{a_4^d}) \\
xg &= a_0^g x^{a_1^g}(1-x)^{a_2^g}(1+a_3^g x^{a_4^g}) \\
x(\overline{d}+\overline{u})/2 &= a_0^+ x^{a_1^+}(1-x)^{a_2^+}(1+a_3^+ x^{a_4^+}) \\
x(\overline{d}-\overline{u}) &= a_0^- x^{a_1^-}(1-x)^{a_2^-}(1+a_3^- \sqrt{x}+a_4^- x) \\
xs &= \kappa \cdot x(\overline{d}+\overline{u})/2
\end{aligned}
\tag{3.16}
$$

The normalization coefficients a_0^u and a_0^d are fixed by the number sum rules for the valence quarks

$$
\int_0^1 u_v(x)dx = 2 \quad \text{and} \quad \int_0^1 d_v(x)dx = 1 ,
\tag{3.17}
$$

and one of the other normalization coefficients, for instance, the gluon normalization coefficient a_0^g, is fixed by momentum conservation

$$
\sum_{f=all\ partons} \int_0^1 xf(x)dx = 1 .
\tag{3.18}
$$

Furthermore, with the data currently available it is not possible to separately determine the low-x behavior for the sea distributions, so we set the strange quark distribution to be proportional to the average non-strange sea. We have also fixed $\kappa = 1/2$ in most of our fits since the resulting $s(x, Q_0)$ agrees well with the recently published NLO strange quark distribution measured in the most accurate dimuon neutrino experiment [60]. Further reduction of independent parameters could be achieved by assumptions such as $a_1^u = a_1^d = a_1^-$ (motivated by Regge exchange considerations). The viability of such assumptions needs to be tested during the process of the global analysis.

In practice, the series of CTEQ analyses adopted the procedure of starting with a sufficient number of parameters to establish a good fit, then systematically reducing that number to eliminate extraneous degrees of freedom while maintaining good agreement with data. In the most recent CTEQ analyses we found it possible to

obtain excellent overall fit using only 15 independently adjustable *shape parameters* to describe the input distributions. In addition, there are individual normalization parameters for each experiment (constrained by appropriate experimental errors, as described earlier) and the value of Λ_{QCD}.

Applying the PDF's obtained here to generate predictions for processes at new facilities in regions of x and Q^2 beyond those covered in the current global analyses necessarily entails extrapolations in these variables. If one is interested in a region of x below that which was fitted, but at a higher value of Q^2, the "feed down" property of the evolution equations (due to the parton splitting process) provides reasonably reliable extrapolations (cf. [54]) – provided the input distribution functions in this x region are relatively smooth (hence the result is dominated by the nature of the splitting kernel). On the other hand, if one is interested in small x and moderate Q^2, where the PDF's are still dominated by the input functions, the results are in fact only extrapolations, not constrained either by theory or experiment. It is thus important to choose functional forms that smoothly extrapolate into such regions while simultaneously acknowledging the inherent risk of such extrapolations. Sometimes, a given functional form can lead to unintended behavior of the parton distributions beyond the region where data exist. These considerations must be kept in mind, as the parametrization of the non-perturbative initial parton distributions, although guided by certain qualitative "theoretical considerations" (many of which have had to be abandoned in recent years in the face of new experimental results), is ultimately dictated by data and by experience gained in previous global analyses.

The choices shown above are certainly not unique and do, in fact, differ slightly from those used in other work, both by us and by other groups [28]. It is possible to generate fits of comparable quality (in the sense of least-χ^2) using somewhat different functional forms as long as both forms can parametrize the requisite parton distribu-

tion shapes to account for current data. In that case, any remaining difference in the parton distributions can only be resolved by future experiments.

Chapter 4

Results on Previous Generations of Parton Distributions – CTEQ1, CTEQ2, and CTEQ3 Distributions

Four rounds of global analysis based on the general procedures described in chapter 3 have been completed by the CTEQ collaboration [25, 26, 27]. In this chapter, we will describe the progress on previous generations of parton distributions – CTEQ1, CTEQ2, and CTEQ3. The most recent CTEQ4 analyses will be discussed in the next chapter. The CTEQ1 analysis was motivated by the precision DIS data from NMC and CCFR available in 1992, along with existing data at that time. Aside from obtaining several up-to-date sets of parton distributions (the "CTEQ1 distributions"), this analysis uncovered an unexpected inconsistency among existing experiments concerning the flavor dependence of the sea quark distributions. We briefly discuss the relevant points and subsequent developments on this issue in the next section. Next, the advent of data from HERA along with an alternative treatment of the strange sea led to the development of the CTEQ2 distributions which were made available in the fall of 1993. These distributions are described in Sec. 4.2. More recently, in 1994 lepton pair asymmetry data from NA51 and W-decay lepton asymmetry data

48

from CDF prompted refinements of the analysis, resulting in a new set of CTEQ3 distributions which we discuss in detail in Sec. 4.3.

4.1 CTEQ1 Parton Distributions

The CTEQ1 analysis [25] was based on data on cross-sections and structure functions available at the end of 1992. The list of data sets used is given in Table 3.1 with "1" marked in the final column. Very good fits to this wide range of data were obtained—both the overall χ^2 and the χ^2 distribution among the experimental data sets indicate a remarkable degree of consistency and are in much better quantitative agreement with the available data than previous global fits. Five sets of parton distributions were obtained: two best fits in the $\overline{MS}$ and DIS scheme (CTEQ1M and CTEQ1D), one suitable for leading order calculations (CTEQ1L), one fit with a "singular" ($a_1^g = -0.5$ in Equation 3.16) gluon distribution (CTEQ1MS), and one with Λ_{QCD} fixed at a higher ("LEP") value (CTEQ1ML)[1]. The singular (CTEQ1MS) and "LEP"-Λ (CTEQ1ML) distributions were designed to explore the "possible" parton distributions in nature, although they may not be the best fits to available data . Three (CTEQ1M, CTEQ1D and CTEQ1L) of the five sets were obtained from independent fits to the same data sets under the same assumptions except the scheme for calculating the evolution kernel and the hard cross-sections. Thus, they are functionally equivalent in the sense that (when applied in the appropriate scheme) they yield the same physical cross-sections, within errors, for the data included in the analysis. See Ref. [25] for details.

One disturbing feature of the CTEQ1 parton distributions was that the strange quark distribution $s(x,Q)$ obtained was considerably larger in the $x < 0.1$ region than those obtained from leading-order parton model analysis of the dimuon production

---

[1]LEP is an e^+e^- collider operating at the energy of Z-pole.

data from neutrino scattering[2] [61, 62, 63]. It was pointed out that CTEQ1 $s(x, Q)$ behavior follows necessarily from the high precision input data sets on total inclusive structure functions measured by the CCFR and NMC collaborations through the parton model identity $\frac{5}{6}F_2^{\nu N} - 3F_2^{\mu N} = s(x, Q) +$ *small corrections*, cf. Equation 3.5. As remarked earlier, although this combination of structure functions entails using the (small) difference between two larger numbers, the quoted experimental statistical and systematic errors of the relevant high precision DIS experiments are even smaller, hence enabling this relation (which is implicitly embedded in the global analysis calculations) to play a decisive role in the determination of $s(x, Q)$. However, there is considerable sentiment that information obtained on $s(x, Q)$ from neutrino dimuon data should be more reliable than that from the difference of $F_2^{\nu N}$ and $F_2^{\mu N}$ obtained in total inclusive measurements —in spite of the small quoted errors on the latter, because the dimuon production from neutrino scattering provides a "direct" probing on the strange quark content inside the nucleon, as indicated in Footnote 2.

The apparent disagreement with the dimuon results on $s(x, Q)$ imply that either the theoretical input (to the global analysis or to the dimuon analysis) has deficiencies, or some experimental data sets are inconsistent with each other within the quoted errors. Although our global analysis, by itself, cannot resolve this dilemma, it was the insistence on taking available data and their quoted errors seriously which resulted in uncovering this controversial issue.[3] Ref. [25] suggested careful examination of all possible theoretical and experimental sources of this disagreement. Subsequently, CCFR has reanalyzed their dimuon data [60] using the NLO formalism of [64, 58] (which is more consistent with our theoretical framework), resulting in a modified strange quark distribution. Nonetheless, the above disagreement persists.

---

[2] The partonic process is $\nu_\mu + s \rightarrow c + \mu^- \rightarrow s + \mu^+ + \bar{\nu}_\mu + \mu^-$ and its charge conjugation process.

[3] To avoid this inconsistency, one has to either arbitrarily enlarge the quoted experimental errors or overlook (and accept) statistically significant inconsistent fits.

The neutrino dimuon results were not included in the CTEQ analyses because experimental data in this process are not, so far, available in the form of detector-independent physical quantities (i.e. structure functions) which can be included in a global analysis treating all data on the same footing. In view of the resulting discrepancy, the CTEQ2 analysis takes the complementary approach of making direct use (as discussed below) of the strange quark distribution function obtained by the CCFR collaboration from their parton model analysis of the dimuon data, thereby treating this process differently from all the other experimental input. Obviously, neither approach is completely satisfactory. Eventually, we need to understand the source of the discrepancy, and perform a consistent global analysis including measured dimuon structure functions, thereby avoiding a separate treatment of the strange quark.

4.2   CTEQ2 Parton Distributions

The CTEQ2 analysis was initiated after the first measurement of $F_2^{ep}(x, Q)$ from HERA became available [33]. These new data extended the measured range of x by two orders of magnitude. The HERA data provide very useful constraints on the small-x behavior of the parton distributions even with their relatively large initial errors because of the extended reach down to $x \sim 10^{-4}$. We therefore modified the input used in the CTEQ1 analysis by adding the new HERA data in conjunction with: (i) using a parametrized function $s(x, Q_0)$ obtained by the CCFR collaboration in NLO QCD analysis which was allowed to vary within an error band provided by the experiment [60]; (ii) removing the conflicting CCFR and NMC F_2 data between $x = .01$ and $x = .09$ which forced the large strange sea through the charge ratio relation, Equation 3.5, in the previous analysis; (iii) including the same fixed-target lepton-pair and direct photon production data sets; and (iv) adding the new collider

data on lepton-pair production obtained by CDF [41]. The full list of experiments included in CTEQ2 analysis are those in Table 3.1 with the last column marked either 1 or 2.

We obtained global fits to the experimental data mentioned above, again, with remarkable consistency over all data sets. (See Table 4.3 for detailed information on χ^2 distributions.) Six representative sets of parton distributions were selected for use in applications. Following the general CTEQ convention, they are designated as CTEQ2M, CTEQ2MS, CTEQ2MF, CTEQ2ML (for $\overline{MS}$ best fit, Singular, Flat, and "LEP"-Λ respectively)[4], CTEQ2L (Leading order best fit), and CTEQ2D (DIS scheme best fit). The parameters for the initial distribution functions are given in Table 4.1.

We may compare CTEQ2 prediction to new data not included in the global fit. The charge asymmetry in lepton-pair production A_{DY} from NA51 [42], cf. Equation 3.11, is in qualitative agreement. This is shown in Figure 4.1.[5] As discussed in Sec. 3.4, our use of the full set of double differential cross-section $d^2\sigma/dQ^2dy$ data measured by the E605 experiment already constrained the $\bar{d} - \bar{u}$ distribution in the $0.1 < x < 0.5$ region. Thus, the (within about 1 σ) agreement of the CTEQ2 result with the new NA51 data point can be regarded as a reasonable consistency check. (Other work on parton distributions (e.g. MRS D-') tends to use the less comprehensive single differential LPP cross-section $d\sigma/dQ^2$ as a constraint on fits, hence do not take advantage of the full power of the complete E605 data set.)

On the other hand, the lepton asymmetry in W-production $A_W(y)$, cf. Equation 3.12, measured by CDF, conveyed a different message. It was observed that the

---

[4]To be specific: CTEQ2MS (CTEQ2MF) assumes a *singular (flat)* small-x behavior of the form $xf(x,Q_0) \sim x^{-0.5}$ (x^0) for the sea quarks and gluons; and CTEQ2ML fixes $\Lambda^{(5)}$ at 220 MeV. For comparison, the *standard* CTEQ2M has $xf(x,Q_0) \sim x^{-0.26}$ and $\Lambda^{(5)} = 139$ MeV.

[5]The other curves in this figure are obtained from the new CTEQ3M distributions (to be described in the next section) and from the two recent generations of MRS distributions.
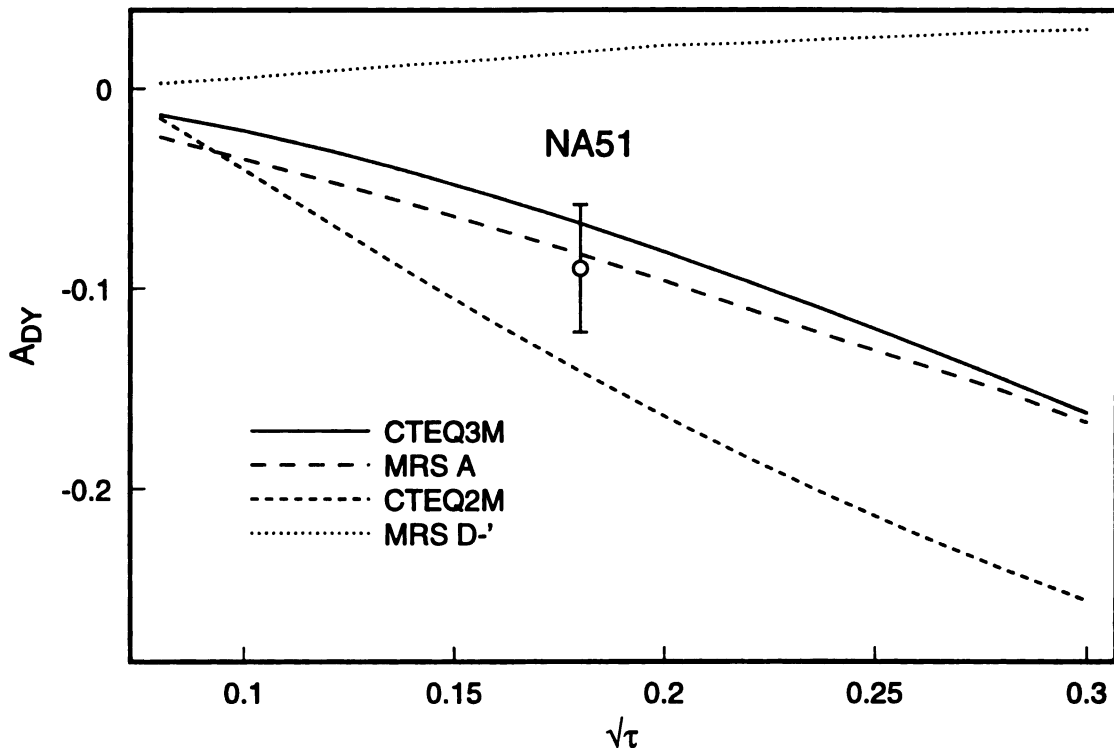
Figure 4.1: A comparison of the data for A_{DY} from NA51 with NLO QCD results obtained from various MRS and CTEQ parton distributions.

predictions of the CTEQ2 distributions were consistently higher than the data, as shown in Figure 4.2. (cf. footnote 5.) Since $A_W(y)$ depends on the x-variation of the ratio d/u, as discussed in Sec.3.4, one naturally turns to data on the ratio of F_2^p/F_2^n in DIS (which also depends on d/u) for a consistency check. It turns out that the CTEQ2 distributions provide an excellent description of the full set of high precision NMC data on F_2^p/F_2^n. In fact, a careful study of the quality of fits to all experimental data sets (cf. Table 4.3) of the CTEQ2 distributions compared to that of other contemporary distributions reveals that CTEQ2 gives a much better overall fit (at least in terms of a substantially lower χ^2)[6] even if others may agree with the specific $A_W(y)$ measurement better. This underlines the fact that $A_W(y)$ is particu-

---

[6]To be specific: using our treatment of experimental errors (close to those specified by the experiments in all cases), the difference in χ^2 is of the order 80-90 (for 920 points) which is evenly distributed in one of the high precision DIS experiments, either BCDMS or CCFR .
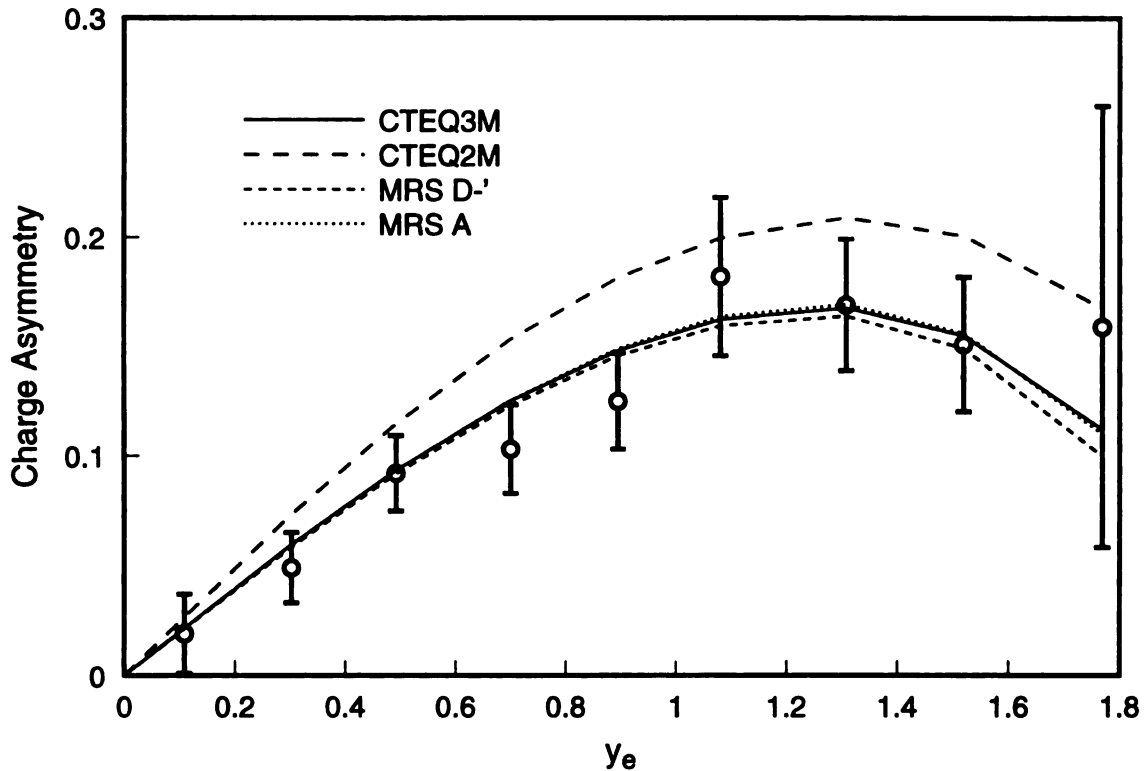
Figure 4.2: The CDF W lepton charge asymmetry data compared to NLO QCD results obtained from various MRS and CTEQ parton distributions.

larly sensitive to *one aspect* of the PDF's – the *slope* of d/u (cf. Sec.3.4) – which is not probed by the other experiments. To study the implication of this fact, we should ask then: Is it possible to vary the CTEQ2 distributions to fit the $A_W(y)$ data and, at the same time, maintain the same quality of agreement with all the other experiments? Or, can we reconcile and understand the interplay of all experiments which play a role in flavor differentiation of the u and d quarks $-F_2^p/F_2^n$, E605, $A_W(y)$ and NA51? This question will be addressed in the next section on CTEQ3 analysis.

One may note that the results in Table 4.3 reveal that the overall χ^2 value in the global fit (including the new data sets mentioned above) for CTEQ2M remains the *lowest* even compared to the two more recent fits which are designed to give better description of the new data. This fact serves as a reminder that total χ^2 is not necessarily the best or only measure of a "good fit" in a global analysis. The balanced

distribution of χ^2's among data sets, particularly those which are sensitive to specific features – such as the $A_W(y)$ measurement to the d/u ratio (relevant for SU(2) flavor differentiation) – must also be taken into account. The new CTEQ3 distributions give a more balanced fit in this sense at the expense of marginally higher total χ^2; hence, they represent an improved general purpose parton distribution set.

Since the CTEQ2 distributions do give such a good global fit to the full data set, the fine-tuning which leads to CTEQ3 only entails very small shifts in the u and d quark distributions, as will be shown in the next two sections. Consequently, for the vast majority of applications which are not sensitive to the precise distinction between u and d quarks, there will hardly be any observable differences in practice.

4.3   CTEQ3 Analysis and Distributions

Previous global analyses have been dominated by experimental data collected at fixed-target energies. The observed sensitivity of the new CDF data on $A_W(y)$ to details of the parton distributions, particularly u and d quarks, ushers in a new stage of global analysis marked by an increasing role for quantitative measurements at hadron colliders.[7] In addition, with the increased number of physical processes included in the analysis, we are approaching the point where all parton flavors will be sufficiently constrained to lead to either an (almost) unique set of PDF's (in the x range covered by the experiments) or evidence for potential inconsistencies. The detailed CTEQ3 analysis is undertaken to respond to this new development and to address the related issues discussed at the end of the last section. All data sets used in CTEQ2 analysis plus the recent NA51, CDF $A_W(y)$ measurements, and the final 1994 ZEUS data on $F_2(x,Q)$ [65], are included in the global fit.

---

[7]Other measurements which will soon play an important role, especially for probing the gluons, are precise data on direct photon production (including photon plus jet) and jet cross-sections (including di-jets).

The specific parametrizations for the initial parton distributions (at $Q_0 = m_c = 1.6\,\text{GeV}$) used in this analysis are discussed in Sec.3.5. The effect of various choices of functional forms and the number of independent parton shape parameters on the predicted behavior of the various processes and on the global analysis have been extensively studied. We found that: the new data do help constrain the flavor dependence of the quark distributions, in particular the u and d, much better than before. From these studies, we have chosen a representative set of new parton distributions – the CTEQ3 distributions, which give a best balanced fit to all available data. Following the general CTEQ convention, the new parton distribution sets in the commonly used factorization schemes will be referred to as CTEQ3M ($\overline{\text{MS}}$), CTEQ3D (DIS), and CTEQ3L (Leading-order) respectively. Since these three sets are equivalent in the sense described in Sec.4.1, in the ensuing discussion, we shall only mention the CTEQ3M distributions explicitly.

The parton distribution shape parameters at $Q_0 = 1.6$ GeV for the CTEQ3 distributions obtained from the global fit are listed in Table 4.2. During the process of this analysis, we started from the full set of (18) parameters introduced in Sec.3.5, then tried to systematically reduce the number of independent parameters while maintaining the quality of the fit as established by benchmarks from the starting fits. The final fit involves 15 parton shape parameters, which is considerably lower than the previous CTEQ analyses.

The total χ^2 is 839 for 850 degrees of freedom, using the available data points of the time in data sets listed in Table 3.1 with the last column marked 1, 2 or 3. The distribution of the χ^2 values among the various processes and data sets is balanced, as summarized in Table 4.3 which also shows the corresponding χ^2 values obtained for some other parton distribution sets in order to indicate where the differences between the various sets lie, as already mentioned in the last section. The normalization

factors for the various experiments emerging from the CTEQ3 global fit are given in Table 4.4.

An overview of the various flavors of CTEQ3M parton distributions at the scale $Q = 5$ GeV is displayed in Figure 4.3. Included (near the bottom of the figure) is the difference between $\bar{d}$ and $\bar{u}$ distributions, which has been a subject of much attention in the last few years. The fact that now we can investigate quantitatively the behavior of such a small difference illustrates the significant progress made by recent high precision experiments and the corresponding theoretical calculations. We will discuss the uncertainty on this difference later.
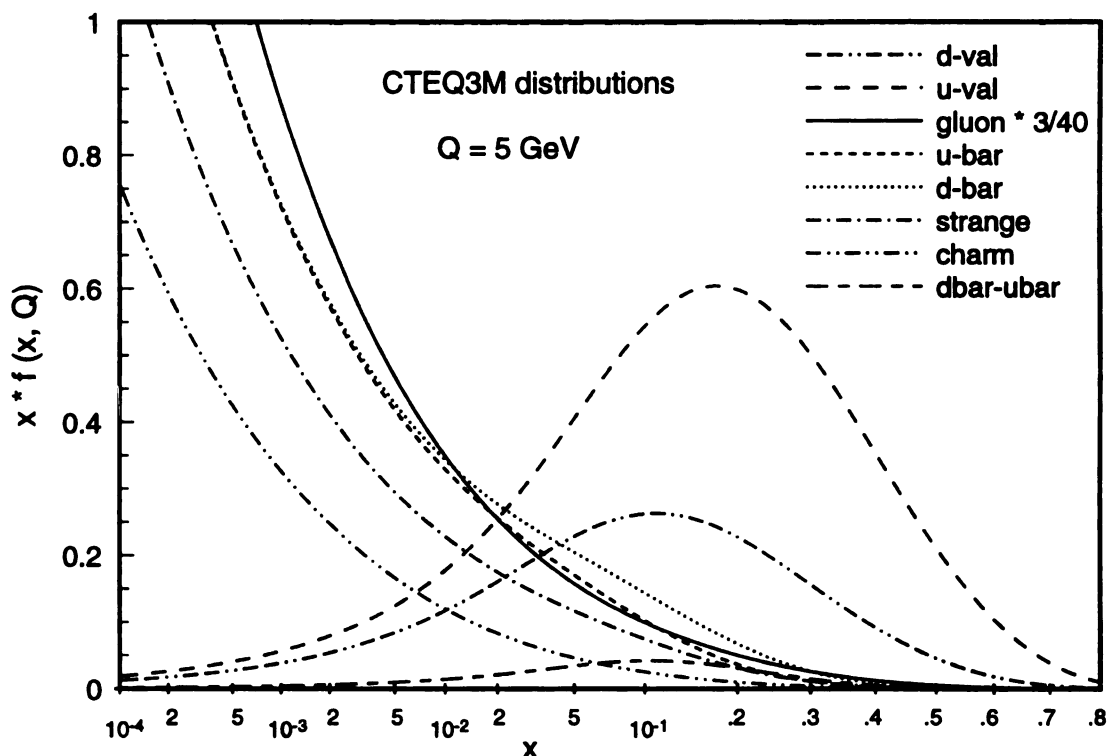
Figure 4.3: An overview of all parton distribution functions at $Q = 5$ GeV from the new CTEQ3M analysis.

Concerning the global analysis which led to the CTEQ3 parton distributions, we notice that:

o The value of $\Lambda_{QCD}^{(5\ flavors)}$ obtained in this round of analysis, 158 MeV (cf. Table

4.2), is similar to the values obtained in previous global fits. It corresponds to a value of $\alpha_s(M_Z^2)$=0.112, in agreement with the value determined from Q-dependence of fixed-target DIS data, but lower than that from global analysis of LEP data [66], reflecting the dominance of DIS in the CTEQ global analysis. However, since the value of Λ is correlated with other parameters in the global fit, particularly those associated with gluon shape which are not well-determined yet, there is still a range of uncertainty on Λ. We found that alternative ways of treating certain processes, e.g. using a particular fixed scale in direct photon calculations, can cause Λ to shift (usually to higher values) by 30-40 MeV.

o The reach into the small-x $(10^{-2} - 10^{-4})$ range and the reduction in errors provided by the 1994 HERA experiments put rather stringent constraints on the effective power a_1 $(-0.35 < a_1 < -0.15)$ for the sea quarks, cf. Equation 3.16. The need to vary this parameter over a certain assumed range, as done in the past, has diminished. To show the development, Figure 4.4 plots the 1994 ZEUS data on F_2 as a function of x at $Q^2 = 15$ GeV^2 compared to CTEQ3M and some previous distribution sets which assume $a_1 = 0$ (MRSD0', CTEQ2MF) or $a_1 = -0.5$ (MRSD-', CTEQ2MS). These previous MRSD sets came before the advent of any HERA data, whereas the CTEQ2 sets were constrained by the early HERA data.

It is important to bear in mind that the values quoted for a_1 from our fit is applicable at the specified scale Q_0 only (1.6 GeV for CTEQ3). The evolution of the parton distributions with increasing scale to an ever softer (i.e. more singular) shape will cause this effective power to increase in absolute value. Thus, comparison with "theoretical expectations" of small-x behavior for fixed *but unspecified* Q, such as those from the BFKL hard pomeron [67], is inherently of limited validity.

o A feature of the CTEQ3 (and CTEQ2) analysis is the inclusion of a theoretical
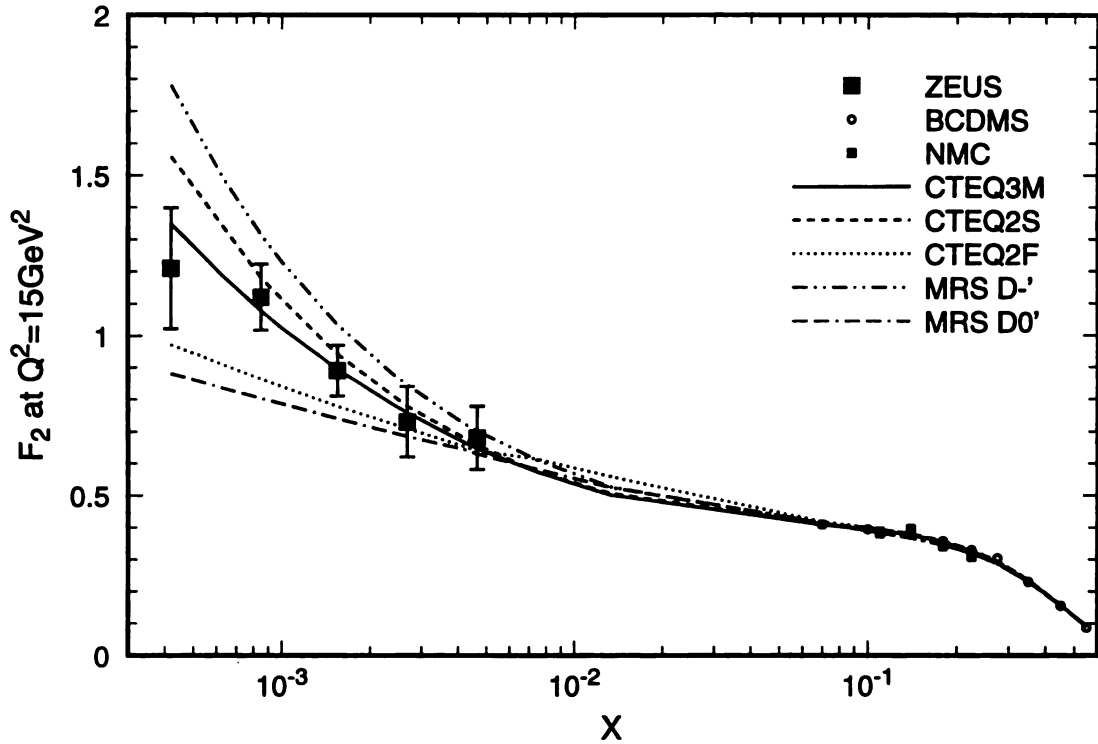
Figure 4.4: Comparison of the '94 ZEUS small-x data at $Q^2 = 15$ GeV^2 with various parton distribution sets with different x exponent values.

parameter representing the uncertainty associated with the choice of scale in direct photon calculations (cf. Sec.3.3). The range of scale μ which gives the optimal overall fit is $\mu = (0.4 \sim 0.5)p_t$. This choice of scale is quite reasonable, since it is of the order of the transverse momentum p_t of the produced photon, the only hard scale in direct photon production.

Table 4.1: CTEQ2 input parton distribution function parameters (at $Q_0 = 1.6$ GeV). The functional forms used in CTEQ2 are described in Equation 3.16, except $xs = a_0^s x^{a_1^s}(1-x)^{a_2^s}$.

Distribution	Parameter	2M	2MS	2MF	2ML	2D	2L
xu_v	a_0^u	.269	.268	.261	.266	.307	.164
	a_1^u	.278	.276	.276	.289	.254	.175
	a_2^u	3.67	3.66	3.66	3.58	3.44	3.32
	a_3^u	29.6	29.1	29.8	30.2	25.5	44.1
	a_4^u	.807	.801	.795	.799	.917	.961
xd_v	a_0^d	1.24	1.32	1.18	1.46	1.17	1.08
	a_1^d	.521	.538	.508	.565	.511	.493
	a_2^d	3.18	3.26	3.24	3.46	3.16	3.00
	a_3^d	-0.85	-0.84	-0.83	-0.59	-0.60	-1.00
	a_4^d	1.82	1.85	2.19	2.32	2.31	2.99
xg	a_0^g	.900	.197	3.05	.825	.711	.521
	a_1^g	-.258	-.500	.000	-.212	-.240	-.259
	a_2^g	5.19	3.82	6.53	4.55	4.84	4.61
	a_3^g	5.13	5.81	2.64	12.0	7.43	16.3
	a_4^g	1.12	.450	2.22	1.62	.960	1.24
$x(\bar{d}+\bar{u})/2$	a_0^+	.0825	.0130	.2540	.1139	.0947	.1127
	a_1^+	-.258	-.500	.000	-.212	-.240	-.259
	a_2^+	8.45	7.62	9.40	9.14	8.76	8.94
	a_3^+	12.7	38.4	13.5	15.2	14.6	17.5
	a_4^+	1.10	0.82	1.60	1.36	1.39	1.58
$x(\bar{d}-\bar{u})$	a_0^-	.111	.105	.114	.117	.121	.103
	a_1^-	.012	.043	.085	.031	.106	.043
	a_2^-	9.53	10.00	9.71	9.95	9.00	9.87
	a_3^-	-14.8	-15.5	-15.7	-15.4	-15.7	-17.7
	a_4^-	49.4	53.8	55.7	51.7	48.2	52.3
xs	a_0^s	.156	.152	.110	.155	.140	.165
	a_1^s	-0.004	0.004	-0.128	0.001	-0.004	-0.001
	a_2^s	6.87	6.85	6.88	6.90	6.90	6.90
	$\Lambda^{(5)}(MeV)$	139	135	135	220	155	143

Table 4.2: CTEQ3 input parton distribution function parameters (at $Q_0 = 1.6$ GeV). The functional forms are described in Equation 3.16. The number of independent parton shape parameters is 15.

Distribution	Parameter	CTEQ3M	CTEQ3D	CTEQ3L
xu_v	a_0^u	1.37	1.36	1.29
	a_1^u	.497	.470	.452
	a_2^u	3.74	3.51	3.51
	a_3^u	6.25	6.19	6.85
	a_4^u	.880	1.04	1.11
xd_v	a_0^d	.801	.837	.858
	a_1^d	.497	.470	.452
	a_2^d	4.19	4.22	4.20
	a_3^d	1.69	2.58	2.54
	a_4^d	.375	.748	.947
xg	a_0^g	.738	.595	.404
	a_1^g	-.286	-.332	-.349
	a_2^g	5.31	5.45	5.59
	a_3^g	7.30	11.0	18.1
$x(\bar{d} + \bar{u})/2$	a_0^+	.0547	.0330	.0451
	a_1^+	-.286	-.332	-.349
	a_2^+	8.34	8.16	7.36
	a_3^+	17.5	23.2	14.5
	a_4^+	1.0	1.0	1.0
$x(\bar{d} - \bar{u})$	a_0^-	.0795	.0702	.0566
	a_1^-	.497	.470	.452
	a_2^-	8.34	8.16	7.36
	a_3^-	0.0	0.0	0.0
	a_4^-	30.0	27.1	29.9
xs	κ	.5	.5	.5
	$\Lambda^{(5)}(MeV)$	158	164	132

Table 4.3: χ^2 and χ^2 per point (in parentheses) in each experiment and overall for previous generations of CTEQ and MRS distributions (at the end of 1994). In the case of the MRS distributions, we have *minimized* the χ^2 by adjusting all the experimental normalizations *freely* while keeping the parton distributions as given by the authors. (See Table 4.4.) These χ^2 values are obtained using the data sets of Table 3.1, employing the same error definitions (except for experimental normalization for which the CTEQ numbers include extra χ^2's for any deviation away from unity as explained in the text); hence they are not necessarily the same as those quoted in the original work which may use a different selection of data points (e.g. for Drell-Yan, and direct photon experiments), apply different error definitions, and adopt different analysis procedures. Large differences in the total χ^2 are mainly associated with the precise BCDMS and CCFR experiments. They may be partially attributed to the influence on the fits due to the $x < 0.09$ data points of CCFR and NMC which are excluded in the CTEQ analyses for consistency considerations, but included in the MRS ones.

expt.	# of pts	CTEQ3M	MRS A	CTEQ2M	MRS D-'
BCDMS^H	168	130.0(0.77)	168.0(1.00)	110.2(0.66)	133.2(0.79)
BCDMS^D	156	187.2(1.20)	215.3(1.38)	174.7(1.12)	162.2(1.04)
NMC^H	52	59.9(1.15)	60.2(1.15)	61.5(1.18)	59.2(1.14)
NMC^D	52	47.2(0.91)	56.0(1.08)	49.1(0.94)	49.7(0.96)
NMC$_R$	89	133.5(1.50)	140.6(1.58)	139.7(1.57)	144.2(1.62)
CCFR F_2	63	69.3(1.10)	68.7(1.09)	58.8(0.93)	95.8(1.52)
CCFR F_3	63	41.0(0.65)	61.7(0.98)	37.2(0.59)	67.4(1.07)
ZEUS	56	27.9(0.50)	40.3(0.72)	27.9(0.50)	74.5(1.33)
H1	21	7.7(0.37)	7.0(0.33)	6.4(0.30)	11.7(0.56)
E605	119	92.3(0.78)	95.9(0.81)	88.1(0.74)	102.6(0.86)
CDF DY	8	3.0(0.38)	1.4(0.18)	2.6(0.32)	2.8(0.34)
CDF A$_W$	9	3.5(0.39)	3.4(0.38)	12.2(1.36)	3.8(0.42)
NA51 A$_{DY}$	1	0.4(0.35)	0.0(0.03)	3.0(3.02)	10.3(10.3)
WA70	39	23.3(0.60)	21.3(0.55)	22.6(0.58)	21.4(0.55)
E706	8	11.8(1.47)	11.2(1.40)	12.2(1.52)	11.3(1.41)
UA6$^{\bar{p}p}$	8	1.8(0.23)	1.6(0.20)	2.2(0.27)	1.5(0.19)
UA6^{pp}	8	6.8(0.85)	6.8(0.85)	7.5(0.94)	6.8(0.85)
Total	920	844	959	816	958

Table 4.4: Normalization factors for experiments obtain in the CTEQ fits according to error treatment procedure described in the text. For comparison, also listed are normalization factors obtained by fitting the same data sets using the (fixed) MRS distributions and allowing all the experimental normalizations to adjust freely.

expt.	CTEQ3M	MRS A	CTEQ2M	MRS D-'
BCDMS	0.988	0.977	0.988	0.969
NMC_{90}	1.008	1.008	1.009	0.996
NMC_{280}	1.021	1.014	1.021	1.000
CCFR	0.975	0.968	0.979	0.958
ZEUS	0.978	1.029	1.003	1.061
H1	0.966	0.978	0.956	1.043
E605	1.098	1.008	1.063	1.052
CDF DY	0.965	0.805	0.971	0.970
WA70	1.010	1.055	1.023	0.977
E706	0.923	0.980	0.960	0.912
$UA6^{\bar{p}p}$	0.858	0.853	0.878	0.813
$UA6^{pp}$	0.853	0.900	0.892	0.834

Chapter 5

Results on CTEQ4 Analyses and the Gluon Distribution

With continuously updated data taken from HERA and new DIS data from NMC and E665, and, moreover, a new type of data – inclusive jet production – from the Tevatron, we have carried out a first systematic study of the uncertainty of parton distributions, especially the gluon distribution, using the CTEQ global analysis framework described in Chapter 3.[1] We discuss phenomenological issues pertinent to extracting $G(x,Q)$ in the global analyses. These factors are systematically taken into account in a series of analyses to gain insight on the current range of uncertainties in $G(x,Q)$. We found that recent, more precise, DIS data have a significant influence in narrowing down the parton distribution functions (PDF's), including $G(x,Q)$; and the inclusion of inclusive jet data from hadron colliders further solidifies knowledge on $G(x,Q)$ over a wide range of x. As the result of this study, we present new sets of CTEQ parton distributions (in $\overline{MS}$, DIS and LO schemes) as well as a series of distributions which give a range of variation of PDF's consistent with current data. We give quantitative information on how these distributions compare to the data sets used in the analysis. In addition, we provide a new set of PDF's with a low initial

---

[1]A similar analysis was carried out earlier [48], focusing on the interpretation of the "high p_t excess" seen in the CDF inclusive jet measurement [49]. See Sec. 5.5 for a discussion of the relation of [48] to the current analysis.

$Q_0 = 0.7$ GeV; and we discuss the previously obtained gluon distributions designed to accommodate the Tevatron high E_t jets [48] in the context of the CTEQ4 analysis.

5.1 Issues in the determination of the gluon distribution

In pQCD, the gluon distribution function is always accompanied by a factor of the strong coupling (i.e. it appears as $\alpha_s G(x, Q)$), both in the hard cross-sections and in the evolution equation for the parton distributions. Thus, the determination of α_s and $G(x, Q)$ is in general a strongly coupled problem. In principle, α_s can be independently extracted from e^+e^- collisions, or from sum rule measurements in deep inelastic scattering. $G(x, Q)$ can then be determined in a global analysis, along with the quark distributions $f^i(x, Q)$, by treating α_s as known. Alternatively, one can try to determine α_s, $G(x, Q)$ and the quark distributions at once in a global analysis. This relies on the full (x, Q) dependence of the wide range of data to differentiate α_s (which controls the overall Q-dependence of all quantities) from the parton distributions (which depend on both x and Q). This method is not as "clean" as the first approach, and comparable precision will require improvements in the theoretical and experimental ingredients of the global analysis. Eventually, however, it is important to demonstrate that the same value of α_s consistently describes all the processes included in the global analysis. Hence, the two approaches are indeed complementary.

It is well known that, at present, the value of α_s determined at high energy colliders, especially LEP, is generally higher than that obtained from analyses of fixed-target DIS data [68]. Since global QCD analyses are up to now dominated by the copious high statistics DIS data, they favor values of α_s close to the lower "DIS value". This situation may change when more and more quantitative results

from hadron collider processes, such as inclusive jet and direct photon production, are included in the global analysis. In the following, we shall explore the range of variation of $G(x, Q)$ when the value of α_s is varied within the currently accepted region, which we shall take to be $0.105 < \alpha_s(M_Z) < 0.122$ [69]. The problem of the determination of α_s in global analysis and the question about consistency of α_s among different processes need to be further studied.

For a quantitative study of $G(x, Q)$, another relevant consideration is: How does the choice of parametrization of the initial gluon distribution at some $Q = Q_0$ affect the results? All global analyses use a generic form:

$$G(x, Q_0) = a_0 \, x^{a_1} \, (1 - x)^{a_2} \, P(x; a_3, ..) \tag{5.1}$$

with $a_{1,2}$ being physically associated with small-x Regge behavior [70] and large-x valence counting rules [71] respectively; and $P(x; a_3, ...)$ being a suitably chosen smooth function depending on one or more parameters. In general, both the number of free parameters and the functional form can have an influence on the global fit. In the CTEQ3 analysis described in Sec. 4.3, an effort was made to minimize the number of free parameters, resulting in an economical set whereby $a_1^G = a_1^{sea}$, and $P_{CTEQ3}(x; a_3) = 1 + a_3 \, x$. We shall refer to this choice as the *minimal set* in the following discussions. In the literature, more degrees of freedom have been assigned to $G(x, Q_0)$. For instance, in CTEQ2 and in recent MRS fits [28], a_1^G is allowed to vary independently of a_1^{sea}; and the function P contains one more free parameter than P_{CTEQ3}: $P_{CTEQ2}(x; a_3, a_4) = 1 + a_3 x^{a_4}$; $P_{MRS}(x; a_3, a_4) = 1 + a_3 \sqrt{x} + a_4 x$. Since two extra degrees of freedom are added, we shall refer to this class of parametrization as *(m+2)* – i.e. *minimal plus two*. The more general parametrization clearly allows a wider range of variation of $G(x, Q_0)$. Some pertinent questions are: whether these general parametrizations are required by current data; and do these parametrizations

give a good indication of the range of variation of $G(x, Q_0)$? We shall investigate these questions in some detail in the next two sections.

Finally, although the PDF's determined from global analysis should, in principle, be universal, they could, in practice, depend on the choice of data sets – in particular, on the choice of "Q_{cut}" values that specify the minimum hard physical scale (Q, P_t, ...) required for data points in the various physical processes to be included in the fit. If the NLO QCD theory is truly applicable in the kinematic range of the data, the parton distributions should be insensitive to the value of Q_{cut}. Since current theory does not predict what value Q_{cut} should take for each process, this point has to be investigated phenomenologically. We found that the obtained PDF's were rather stable when Q_{cut} was varied from 2 to 5 GeV, while keeping all other conditions unchanged.

5.2   Impact of recent DIS data on the global analysis of parton distributions

Since the publication of the CTEQ3 analysis, more accurate and extensive DIS data from NMC [72] and HERA [73, 74] as well as new data from E665 [75] have become available. In comparing the new data with the NLO QCD F_2 computed from CTEQ3M distributions, we find general agreement, except for the small-x region where the more precise recent data show deviations from the theory curves. This is shown in Figure 5.1 for the NMC and H1 data sets respectively.[2] Thus, before obtaining final CTEQ4 sets we first update the CTEQ3 analysis under several different conditions, in order to study the impact of these new DIS data on the global analysis of parton distributions, especially the extraction of the gluon distribution.

The magnitude of the uncertainty in $G(x, Q_0)$ due to the current uncertainty on

---

[2]Results are similar for E665 and ZEUS. Comparison with the full data sets will be presented later, cf. Figure 5.11
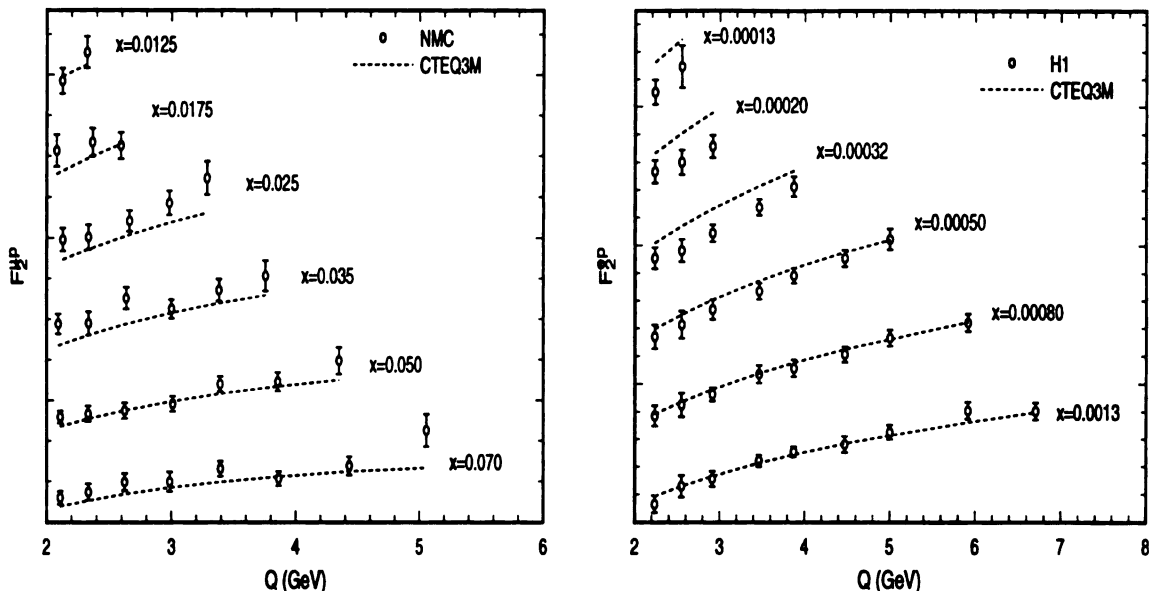
Figure 5.1: Comparison of NLO calculations based on the previous generation CTEQ3M parton distributions with the latest NMC (a) and H1 (b) data in the small-x region where discrepancies appear.

α_s will be investigated by systematically varying the value of α_s over the interval $0.105 < \alpha_s < 0.122$, as mentioned in the previous section. We shall use the shorthand α_s for $\alpha_s(M_Z)$ throughout. In terms of QCD Lambda values, this range of α_s corresponds to $100 < \Lambda_{\overline{MS}}^{(5)} < 280$ (MeV) and $155 < \Lambda_{\overline{MS}}^{(4)} < 395$ (MeV). We shall in general use the $\overline{MS}$ scheme in NLO QCD.

To provide a base-line for comparison, we first obtain a series of such fits under identical conditions and using the same data sets (i.e. pre-1995) as in the CTEQ3 analysis [26]. We shall refer to this as the A-series.[3] By definition, the best fit in this series is the published CTEQ3M fit with $\alpha_s = 0.112$ ($\Lambda_{\overline{MS}}^{(5)} = 158$ MeV). A comparison of the gluon distributions that correspond to these values of α_s are presented in Figure 5.2. In order to render the differences in the various regions of x visible over the range $10^{-4} < x < 1$, part (a) highlights the small-x region by plotting $x\, G(x, Q)$

---

[3]This series of fits were originally obtained in 1994. They have been used in various phenomenological studies related to gluon distributions and α_s determination conducted by CTEQ, CDF, and D0 Collaborations. They have not been formally published.
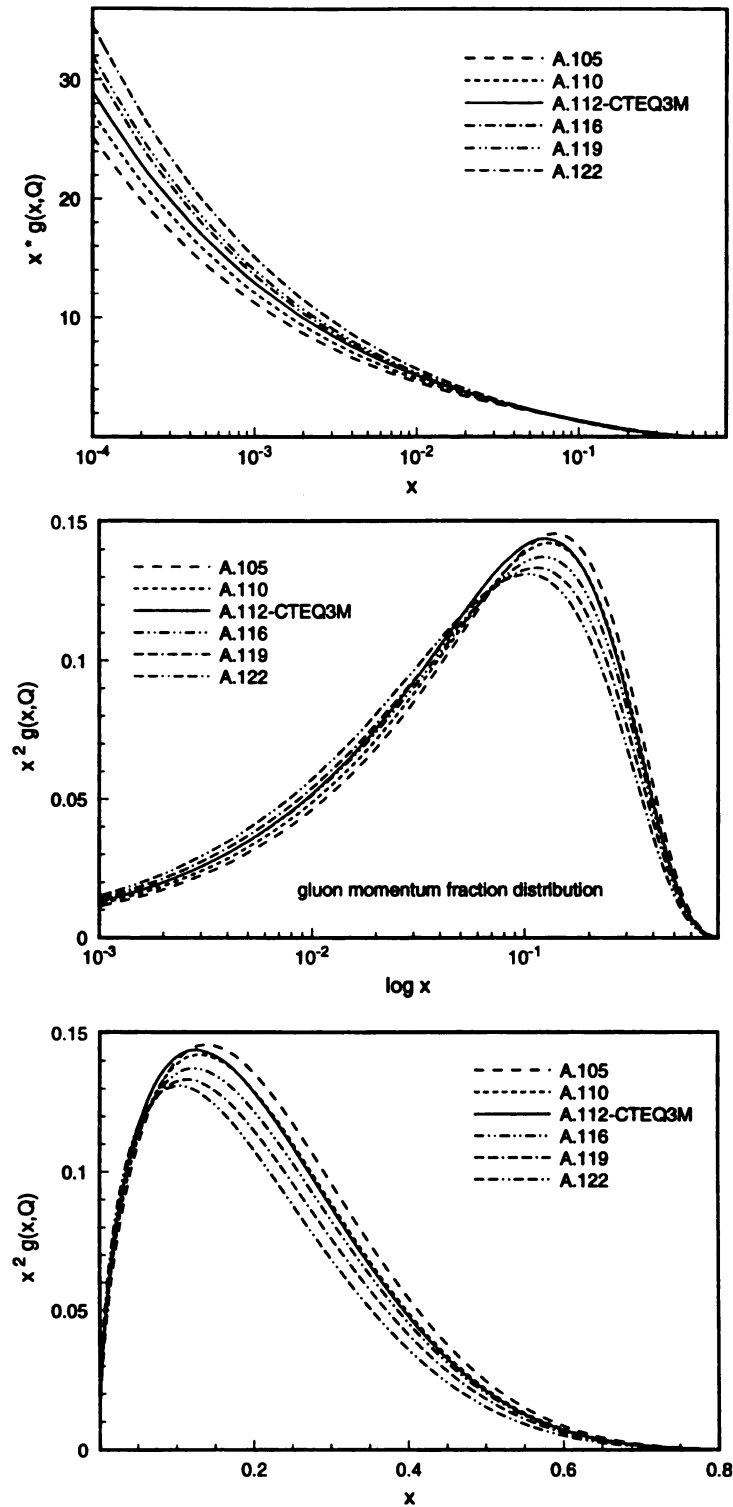
Figure 5.2: Series-A gluon distributions in the small-, medium-, and large-x regions. A.105 refers to the gluon associated with $\alpha_s(M_Z) = 0.105$, and likewise for the other ones.

against log x, part (b) accentuates the medium-x range by plotting $x^2\,G(x,Q)$ vs. log x,[4] and part (c) emphasizes large-x by plotting $x^2\,G(x,Q)$ vs. x. For the many detailed comparisons to follow, these separate plots, though conventional, will prove to be rather cumbersome. We consolidate them into one single less-conventional plot in Figure 5.3 in which all curves are normalized by the function $x^{-1.5}(1-x)^3$, which takes out most of the singular (rapidly vanishing) factors at small (large) x. The scale for the abscissa is chosen to be a function of x (explicitly, $ax + b\ln(1/x)$)
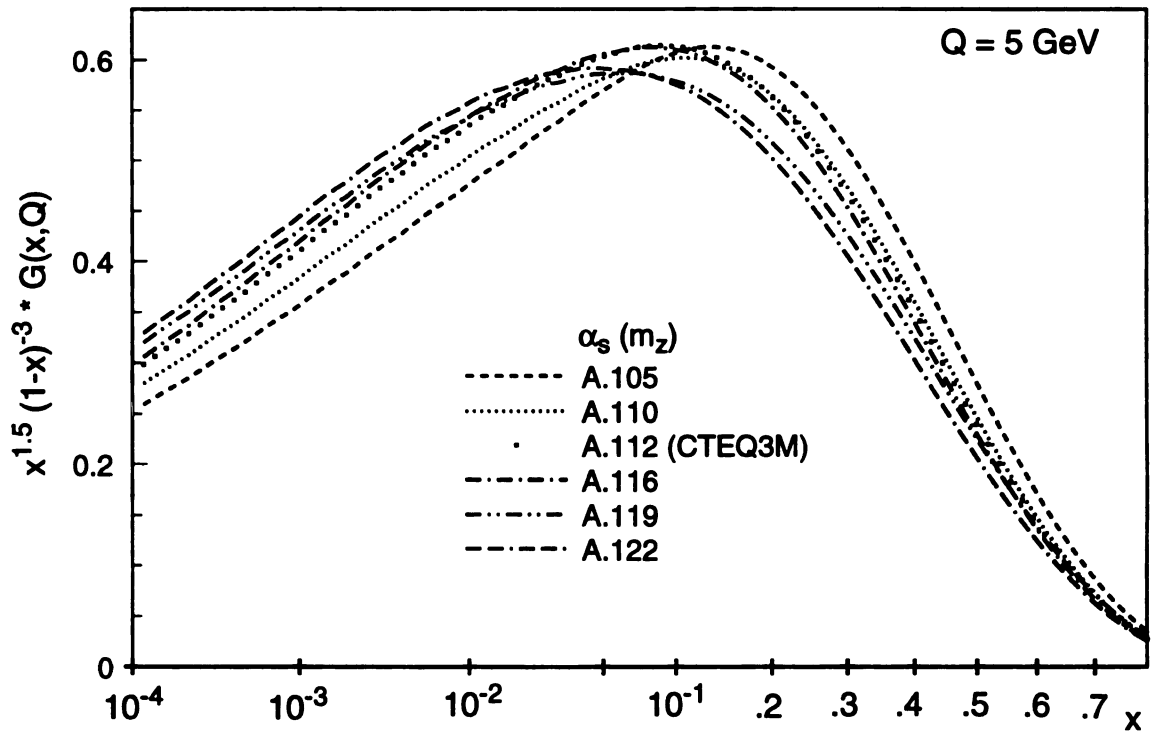
Figure 5.3: Series-A gluon distributions normalized by the function $x^{-1.5}(1-x)^3$ in order to display clearly the behavior of $G(x,Q)$ over the entire x-range. For the same purpose, the horizontal x-axis is drawn with a scale which smoothly changes from log- to linear behavior.

which smoothly interpolates between $\log x$ (at small x) and x (at large x) so that the behavior of $G(x,Q)$ over the full x range is more evenly displayed. We see that all

---

[4]Note that, since $x^2G(x) \cdot d\log x = xG(x) \cdot dx$ = momentum fraction carried within dx, each curve in this plot directly depicts the distribution of the momentum fraction carried by the gluon for that set.

of the features seen in the three plots of Figure 5.2 are evident in this single figure. This will be the format of choice in most subsequent comparisons.

We see in Figure 5.3 that, in the region $x > 0.05$ where the largest concentration of data used for the fit lie, increasing values of α_s lead to decreasing values of $G(x, Q)$ – as expected (particularly for the direct photon data) since the product of the two enters into most cross-section and evolution kernel formulas.[5] As noted before, in the CTEQ3 analysis, and therefore in this series of fits, the initial gluon distribution function is parametrized *minimally* as

$$G(x, Q_0) = a_0\, x^{a_1}\, (1 - x)^{a_2}\, (1 + a_3 x) \tag{5.2}$$

with a_1 set to be the same as that of the sea quarks. Hence there are 3 free gluon parameters – $a_{0,2,3}$ – in the fit. For each α_s, we found the best solution to be quite stable against perturbations in the fitting procedure and starting parameters, indicating the parametrization and the experimental constraints are well-matched. This also results in an orderly variation of $G(x, Q)$ as α_s is varied, as seen in the figure. If one takes the range of α_s used here as representing the current uncertainty on α_s, then the spread of the gluon distribution shown in Figure 5.3 gives the corresponding uncertainty on $G(x, Q)$ (based on the data available prior to 1995, and on the variation of α_s alone). We should mention that, although quark distributions are allowed to vary freely, the valence quark distributions remain practically the same for all of the fits in this series, as they are very much pinned down by the precision DIS data in the region where they dominate the structure functions. On the other hand, the sea quark distributions couple to $G(x, Q)$; thus they do show a systematic variation with α_s, although the variation is somewhat reduced compared to that of the gluon.

Next, we investigate the impact of the new DIS data from NMC [72], E665 [75] and

---

[5]The order is reversed for small x, because of the momentum sum rule.

HERA [73, 74] on F_2 by repeating the same study, with the new data sets replacing the original ones. The resulting series of fits is called the B-series. The quality of these fits (measured in χ^2 values) are similar to those of the A-series. Six representative gluon distributions in this series are shown in Figure 5.4 along with that of CTEQ3M for reference. It is rather striking to note that the spread in $G(x, Q)$ observed above
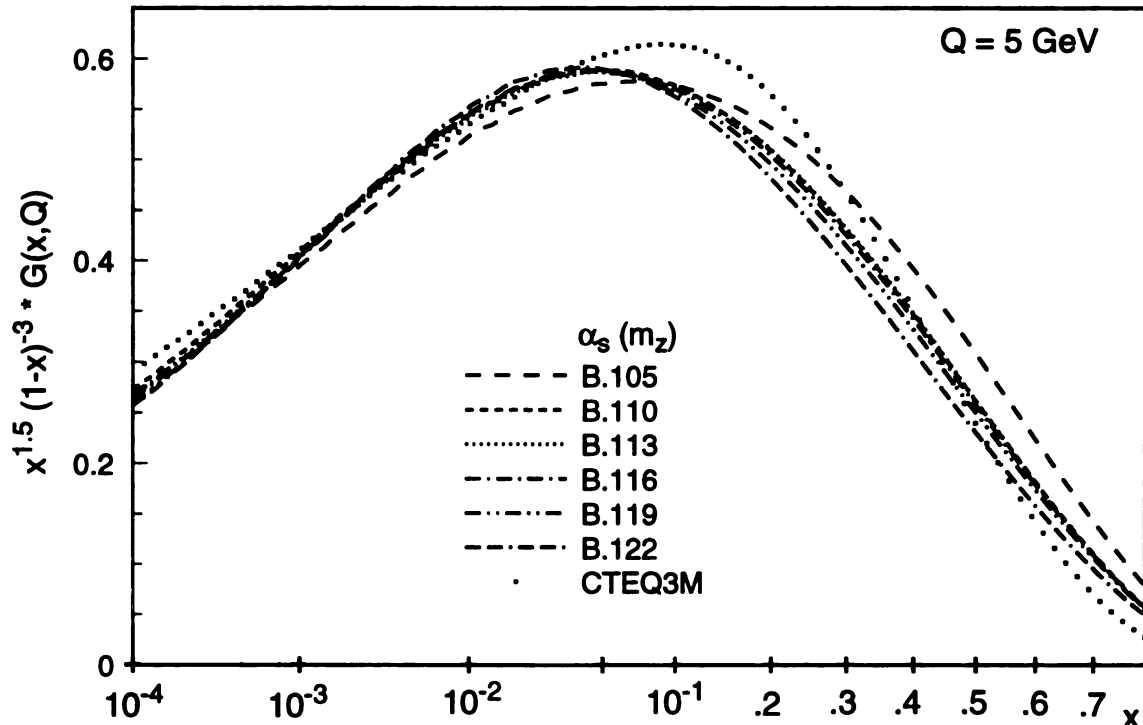
Figure 5.4: Series-B gluon distributions normalized by the function $x^{-1.5}(1-x)^3$ (cf. caption of previous figure.)

in the small-x (< 0.01) region has been practically eliminated. This is precisely the region covered by the HERA experiments. In addition, the new gluons are shifted down from those of the A-series in the region $0.05 < x < 0.3$ where all three DIS experiments contribute. At first glance, this may appear surprising in view of the conventional wisdom that F_2 data are only sensitive to quarks, not gluons. However, we must realize that, first, in the small-x region $G(x, Q)$ is quite large—typically about 20 times bigger than the quark distributions—thus it has a strong influence, directly

and indirectly, on all physical quantities through the hard cross-section and the evolution equation. Moreover, these fits use the minimal parametrization, including the constraint $a_1^G = a_1^{sea}$ which strongly couples the behavior of $G(x, Q)$ at small-x to that of sea quarks. Thus, the much better determined $G(x, Q)$ just reflects the improved accuracy of new data in this region. We note also, the large-x behavior of the new series is somewhat different from the A-series, even if there are no new data in that region. This must be due to the indirect effect of the required changes below $x = 0.1$, induced by the restrictive functional form Equation 5.2, and the constraint imposed by the momentum sum rule. We should point out that the absolute value of the gluon distribution in the region above $x = 0.5$ is very small (about 10^{-3} compared to its value at $x = 0.1$); thus the significance of the observed differences should not be over-emphasized.
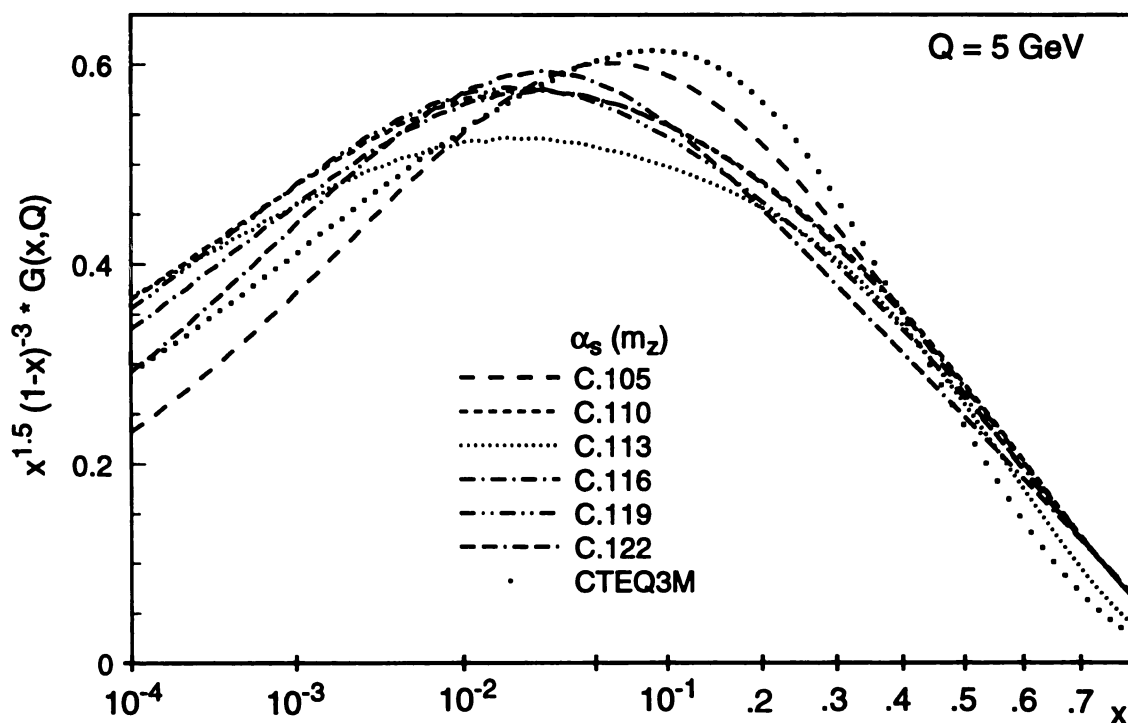
Figure 5.5: Series-C gluon distributions normalized by the function $x^{-1.5}(1 - x)^3$

The minimal parametrization for $G(x, Q_0)$ used above was originally chosen in the CTEQ3 analysis for its economy – all data sets included in these global analyses can be reasonably well fitted with this form. This does not prove that the true gluon distribution must fall within the range shown above; in particular, the true $G(x, Q_0)$ may be more complicated than can be represented by this parametrization. (For instance, all global analyses find it necessary to use one more parameter to describe the valence quarks.) Only experiments probing $G(x, Q)$ in a different way can tell whether our results so far are adequate. Before turning to such additional input, we can obtain a different estimate of the uncertainty on the gluon distribution that is complementary to the width of the "band" shown in Figures 5.3-5.4. We adopt the more general "(m+2)" parametrization of $G(x, Q_0)$ already used in CTEQ2:

$$G(x, Q_0) = a_0 \, x^{a_1} \, (1 - x)^{a_2} \, (1 + a_3 x^{a_4})$$ (5.3)

In addition to introducing the new parameter a_4 compared to Equation 5.2, the parameter a_1 is untied from a_1^{sea} and treated as free. This results in a new series of fits, called the C-series.

With two more free parameters than in the B-series, one would expect (i) to fit the collective data "better" than before and (ii) to find an increased range of variation of the gluon distribution. Indeed, the χ^2 for the fits decreased slightly (by about 10 (/1000 degrees of freedom)) compared to the corresponding ones in the B-series. The gluon distributions at $Q = 5$ GeV in this series for 6 values of α_s are shown in Figure 5.5. First, we see that the range of variation of $G(x, Q)$ in this series is much wider as compared to that of series B, although both include the same improved DIS data. In particular, in the small-x region the very narrow range in series B is very much opened up by the freeing of the a_1 parameter for the gluon – since now $q(x, Q)$ and $\alpha_s G(x, Q)$ can vary independently, the measured F_2 (which depends on

both) no longer constrains each piece tightly as in the B-series. Secondly, we note that the gluon distribution does not vary in a systematic manner as the α_s value is varied – in contrast to the well-constrained case in series A and B. Further study has indicated that, unlike in the other cases, small changes in the fitting process can lead to different solutions for some values of α_s. This suggests that the fits are not entirely stable; or, in other words, the system becomes somewhat under-constrained with the two extra parameters introduced.

These observations point to the need for more experimental input in order to better measure the gluon distribution. We need new data to determine whether the additional degrees of freedom associated with a_1^G and a_4^G are required for the true gluon or whether the restricted form used in series B is already sufficient. If a_1^G and a_4^G are required, these new data could help to stabilize the fits found in the C-series and hence shed light on the possible range of $G(x, Q)$ allowed. From the discussion given in the introduction, it is clear that inclusive jet production data could be used to help resolve these issues, as we will show in the next section. To conclude this section, Table 5.1 summarizes the above described three series of global fits, as well as those including jet data to be discussed next.

Table 5.1: Several series of global fits on which the physics discussions are based. "New DIS data" refers to those becoming available since 1995. Minimal parametrization "m" refers to Equation5.2; and "m+2" refers to Equation5.3. The last column refers to the section number where the specific series is discussed.

Series	New DIS data	Inclusive Jet Data	parame-trization	Section discussed
A			m	5.2
B	x		m	5.2,5.3
C	x		m+2	5.2
CTEQ4A	x	x	m+2	5.4

5.3 Comparison with New Inclusive Jets Cross-section

For studying the impact of inclusive jet production cross-section, we use the recent measurement of $d\sigma/dE_t$ from the CDF [49] and D0 [50] Collaborations. The preliminary data obtained in run 1B of the Tevatron by the two experiments are shown in Figure 1.6. Although data are available for 15 GeV$< E_t < 450$ GeV, we will include in our NLO QCD analysis only data *above 50* GeV because there are a number of potential theoretical and experimental problems that may affect the proper comparison between NLO QCD theory and data for lower E_t. These include (1) scale uncertainty of NLO QCD calculations, which becomes non-negligible at low E_t (cf. Figure 1.7a); (2) ambiguities in the definition of the "underlying event" coming from the proton-antiproton remnants (cf. Figure 1.7b); (3) possible problems in the match between theoretical and experimental jet definitions, such as fragmentation products outside the jet cone; (4) k_t broadening of the initial state partons [19]; and (5) non-perturbative corrections to the theory, which could be of order $1/E_t$ rather than $1/E_t^2$ [76]. All of these affect low E_t jets much more than high E_t jets, as will be illustrated by two examples, one theoretical and one experimental. Figure 1.7a shows the scale-dependence of the NLO QCD calculation as a function of E_t: the theoretical inclusive jet cross-section is shown for several choices of the renormalization and factorization scale ($\mu = \mu_R = \mu_F$) normalized to our standard choice $\mu = E_t/2$.[6] For low E_t, the ratio becomes large and unstable; above $50 - 75$ GeV, the different choices are within 10% and stay relatively constant—they amount to shifts in the overall normalization of the cross-section. Figure 1.7b shows the percentage effect on the inclusive jet cross-section in the CDF experiment due to a ± 30 % change in the

---

[6]The theoretical calculations of jet cross-section in our study are carried out using the EKS program [23].

underlying event correction (in Run IA). Again, the uncertainty becomes large below 50 − 75 GeV.

To emphasize the quantitative aspects of the subsequent analysis, the measured steeply falling $d\sigma/dE_t$ is normalized to the NLO QCD theoretical expectation using the CTEQ3M parton distributions (solid horizontal line) and displayed in Figure 1.8 on a linear plot (with statistical errors only on the data points). In Figure 1.8, we have taken into account the slightly different pseudo-rapidity coverage of the two experiments ($0.1 < |\eta| < 0.7$ for CDF vs. $|\eta| < 0.5$ for D0) by normalizing each data set with respect to the theory values computed within the corresponding η range. In addition, we have allowed a small overall normalization freedom of the two data sets, well-within the quoted uncertainties, for this comparison. This figure shows that the two data sets agree quite well over the entire E_t range, especially when considering the quoted systematic uncertainties (not shown). See Ref. [77] for more discussions. We will discuss the experimental systematic uncertainties in the proper context of the "range" of gluon distributions later. The rise of the data points at high E_t values over the CTEQ3M expectation, more noticeable for the CDF points, has been the subject of much recent discussion and speculation [49, 48, 78]. We will comment on this issue in the context of the global analysis in a later section.

Since most inclusive jet data are collected in the central rapidity region, the x-value of the PDF's probed is around $x_t = 2E_t/\sqrt{s}$. For 50 GeV $< E_t <$ 450 GeV, the x range is approximately $0.06 - 0.5$. Over this range, the relative importance of the three parton subprocesses − quark-quark, quark-gluon, and gluon-gluon − shifts continuously from being gluon-dominated to quark-dominated, as illustrated in Figure 2.7. We should also keep in mind that these jet data probe hadron structure at much higher momentum scales than fixed-target experiments. Due to the nature of the QCD evolution equation, parton distributions at these high momentum scales are

determined by those at lower scales and higher x values. Thus the effective x-range in $G(x, Q_0)$ for some Q_0, say 1.6 GeV as chosen in our analysis, probed by these jet data extends to much higher values than the nominal values mentioned above. Since the quark distributions throughout this range are very well pinned down by DIS experiments, one expects the jet data to be particularly useful in constraining the gluon distribution. The value of α_s has considerable influence on the gluon determination for several reasons. First, the cross-section for medium x_t is proportional to $\alpha_s^2 G^n(x, Q)$ (n = 2,1,0), so that an increase in α_s leads to decrease in $G(x, Q)$. Second, α_s controls the rate of evolution of $G(x, Q)$ and hence affects the slope of the gluon distribution for given measured jet cross-sections. Third, $\alpha_s(\mu)$ itself depends on x through $\mu = E_t/2 = x\sqrt{s}/4$ (at $\eta = 0$), so that the rate of variation of α_s is coupled to the x-dependence of $G(x, Q)$ in the cross-section formula.

We now apply the results obtained in Sec.1.2 to these jet data to see how the latter agree with the predictions of perturbative QCD using these new parton distributions determined by the other processes. Figure 1.9 compares the predictions of the PDF's from the B-series (which incorporate the most recent DIS data and use the minimal parameters for the gluon) with the jet data, using the same "(Data - Theory) / Theory" format as Figure 1.8. We use the set with $\alpha_s(M_Z) = 0.116$ as the "Theory" (horizontal solid line) against which the data points as well as the predictions of the other fits with different α_s values in the series are displayed in this plot. To make these comparisons, we allow an overall relative normalization between theory and data.[7] The normalization factor for the CDF (D0) data set ranges from 0.94 (0.92) to 1.08 (1.06) for $\alpha_s = 0.110$ to 0.122 [79]. The normalization uncertainties quoted by the CDF and D0 experiments are around 5%. Considering the 7 orders of magnitude of variation of the cross-section (Figure 1.6), this is quite remarkable. Within the

---

[7]Such a renormalization, within errors, is usually allowed in global fitting.

minimal parametrization of the gluon used by the B-series, the parton distributions narrowed down by recent precise DIS data (see previous section) are *remarkably consistent* with the new high statistics inclusive hadron-hadron jet data. We also found that the more generally parametrized C-series PDF's give qualitatively similar predictions for jet cross-sections compared to the B-series displayed in Figure 1.9; hence they will not be separately shown.

The important questions at this point are the following: (i) At a more quantitative level, how can these parton distributions be improved by including the jet data in the analysis from the beginning; and (ii) will the addition of the jet data reduce the variation of $G(x, Q)$ when we use the more general (m+2) parametrization?

5.4   New CTEQ parton distribution sets – CTEQ4

To answer these questions, we have performed an extensive study of the interplay of the inclusive jet data with the high-precision DIS and other data within the CTEQ QCD global analysis program. The complete set of processes and experiments used is given in Table 3.1 and a kinematic map of the (x, Q) plane with the data range of the various experiments has been shown in Figure 3.3.

We see the greatly expanded kinematic coverage compared to that of earlier analyses, e.g. CTEQ1, CTEQ2 and CTEQ3: in the direction of small-x due to the HERA experiments, and in the high Q direction due to the Tevatron inclusive jet experiments. As before, all processes are treated consistently to NLO accuracy in pQCD. This new round of global analysis will be referred to as the CTEQ4 analysis.

Building upon studies described in the previous sections, we explored all the issues described in Sec. 1.1, now with jet data also playing a role. Although the quark distributions are coupled to $G(x, Q)$ and α_s, they remain tightly constrained by the

DIS experiments, hence stay very close to those determined before. Thus, our studies concern again mainly the range of variation of $G(x, Q)$ due to uncertainties in α_s and the parametrization of the non-perturbative initial distribution. Since the results from Sec. 1.3 indicate that it is possible to obtain good fits to all the data using the minimal parametrization of the gluon distribution even without taking into account the experimental systematic errors on the inclusive jet data, we anticipate the most important role of the jet data in the new analysis is to constrain the possible range of $G(x, Q)$. Hence, we shall use the more general (m+2) parametrization which allows a wider range of variation of $G(x, Q)$. We shall not include the correlated systematic uncertainties on the jet data since they are not crucial for the present purposes. This point will come up again later.

The new generation of CTEQ4 parton distributions are summarized in Table 1.2. Each set will be described in turn in the following.

Table 5.2: List of new CTEQ4 parton distributions and their characteristics.

PDF set	Description	$\alpha_s(m_z)$	Q_0^2 (GeV^2)
	Standard Sets		
CTEQ4M	$\overline{MS}$ scheme	0.116	2.56
CTEQ4D	DIS scheme	0.116	2.56
CTEQ4L	Leading Order	0.132	2.56
	α_s series		
CTEQ4A1	1	0.110	2.56
CTEQ4A2	2	0.113	2.56
CTEQ4A3	Same as CTEQ4M	0.116	2.56
CTEQ4A4	4	0.119	2.56
CTEQ4A5	5	0.122	2.56
	Specials		
CTEQ4HJ	"Hi-Jet"	0.116	2.56
CTEQ4LQ	"Low Q_0"	0.114	0.49

Standard CTEQ4M parton distributions

We first present the standard fit in the $\overline{MS}$ scheme which we will designate as the CTEQ4M set of parton distributions. The $\alpha_s(m_Z)$ value for this set is 0.116, corresponding to second order $\Lambda^{(5)} = 0.202$ or $\Lambda^{(4)} = 0.296$ GeV. This set gives an excellent fit to all data sets. The total χ^2 for 1297 DIS and DY data points is 1320. Detailed information on the χ^2's for the various experiments, in comparison to those obtained using other current and previous generations of parton distributions are presented in Tables 1.3 and 1.4 respectively. The direct photon and jet data sets are not

Table 5.3: Total χ^2 values and their distribution among the DIS and DY experiments for current generation of parton distributions which take into account the most recent HERA (1996) and NMC (1995) data. In parentheses are the χ^2/point values.

Expt.	#pts	CTEQ4M	CTEQ4HJ	CTEQ4LQ	MRSJ
$BCDMS^H$	168	144.8(0.86)	173.0(1.03)	139.4(0.83)	183.1(1.09)
$BCDMS^D$	156	185.6(1.19)	205.9(1.32)	182.5(1.17)	229.3(1.47)
NMC^H	104	97.3(0.94)	91.7(0.88)	96.0(0.92)	113.4(1.09)
NMC^D	104	93.3(0.90)	90.2(0.87)	97.9(0.94)	122.7(1.18)
NMC_R	89	130.8(1.47)	133.5(1.50)	132.6(1.49)	142.4(1.60)
$E665^H$	35	41.3(1.18)	38.5(1.10)	44.5(1.27)	37.8(1.08)
$E665^D$	35	32.3(0.92)	33.5(0.96)	34.3(0.98)	29.8(0.85)
$CCFRF_2$	63	83.2(1.32)	72.4(1.15)	74.3(1.18)	107.7(1.71)
$CCFRF_3$	63	46.5(0.74)	45.5(0.72)	49.9(0.79)	57.8(0.92)
$ZEUS$	179	243.4(1.36)	232.7(1.30)	268.5(1.50)	252.4(1.41)
$H1$	172	118.9(0.69)	120.2(0.70)	131.9(0.77)	109.6(0.64)
$CDFA_W$	9	4.3(0.48)	3.4(0.38)	3.8(0.42)	3.3(0.37)
$NA51$	1	0.6(0.63)	0.5(0.49)	0.4(0.41)	2.5(2.47)
$E605$	119	97.7(0.82)	101.6(0.85)	100.4(0.84)	97.8(0.82)
Total	1297	1320	1343	1356	1490

included in the χ^2 table since, without including the sizable theoretical uncertainties for the former[8] and experimental systematic errors for the latter, the significance of such χ^2 values would be difficult to evaluate. The comparison of the CDF and D0

---

[8]See Sec. 2.3 and Refs. [19, 48] for discussions on these uncertainties.

Table 5.4: Total χ^2 values and their distribution among the DIS and DY experiments for the previous generation of parton distributions which includes experimental data available in 1995 (MRSA') or before 1995 (CTEQ3M, MRSA). GRV does not perform a full global fit. Since it is used widely, it is included here for reference. In parentheses are the χ^2/point values.

Expt.	#pts	MRSA'	CTEQ3M	MRSA	GRV
BCDMS^H	168	156.9(0.93)	128.7(0.77)	168.0(1.00)	250.3(1.49)
BCDMS^D	156	213.7(1.37)	190.3(1.22)	215.3(1.38)	187.2(1.20)
NMC^H	104	129.0(1.24)	146.6(1.41)	114.4(1.10)	123.8(1.19)
NMC^D	104	151.8(1.46)	137.3(1.32)	135.2(1.30)	115.4(1.11)
NMC$_R$	89	143.3(1.61)	134.4(1.51)	140.6(1.58)	129.0(1.45)
E665^H	35	38.2(1.09)	47.6(1.36)	37.8(1.08)	39.9(1.14)
E665^D	35	29.1(0.83)	44.5(1.27)	29.5(0.84)	29.8(0.85)
CCFR F_2	63	68.0(1.08)	66.2(1.05)	68.7(1.09)	164.4(2.61)
CCFR F_3	63	54.1(0.86)	41.9(0.67)	61.7(0.98)	114.7(1.82)
ZEUS	179	368.7(2.06)	549.5(3.07)	1222.6(6.83)	843.1(4.71)
H1	172	149.5(0.87)	220.2(1.28)	407.6(2.37)	404.2(2.35)
CDF A_W	9	4.2(0.47)	3.0(0.33)	3.7(0.41)	9.6(1.07)
NA51	1	0.1(0.06)	0.4(0.42)	0.01(0.01)	0.01(0.01)
E605	119	93.5(0.79)	92.6(0.78)	95.9(0.81)	90.3(0.76)
Total	1297	1600	1803	2701	2502

jet data to the NLO QCD inclusive jet cross-section calculated with the CTEQ4M distributions is shown in Figure 1.10. And the comparison of the recent NMC, H1, and ZEUS data sets to the fit is shown in Figures (1.11 - 1.13).

From Table 1.3, we see that the CTEQ4M PDF set has the best overall quantitative agreement between NLO QCD theory and global data on high energy scattering. It also represents a significant improvement over the previous generation of parton distributions, as a comparison to Table 1.4 makes clear. Most of the difference is caused by the new precision data from the HERA experiments. Figure 1.10 shows good general agreement of CTEQ4M with the jet data, while the much discussed "high E_t excess" is still noticeable. We will return to this issue in Sec. 1.5 where an alternative "high E_t jet-fit" CTEQ4HJ (included in Table 1.3) will be discussed.

Figures (1.11 - 1.13) explicitly show the improvement of CTEQ4M over CTEQ3M in describing the recent high-precision DIS experiments. Detailed information on the parameters which characterize the initial parton distributions at $Q_0 = 1.6$ GeV (which coincides with our choice of the charm threshold) is given in Sec.3.5. The parameters for CTEQ4M are given in Table 1.5.

Table 5.5: Parameters for the CTEQ4M initial parton distributions at $Q_0 = 1.6$ GeV. The functional forms are described in Equation 3.16 with $\kappa = 1/2$. Also, $\alpha_s(m_z) = 0.116$, corresponding to $\Lambda^{(5)} = 202$ MeV.

Parton	a_0	a_1	a_2	a_3	a_4	% Momentum
xd_v	0.640	0.501	4.247	2.690	0.333	11.2
xu_v	1.344	0.501	3.689	6.402	0.873	30.6
xg	1.123	-0.206	4.673	4.269	1.508	41.7
$x(\bar{d} - \bar{u})$	0.071	0.501	8.041	0.000	30.000	–
$x(\bar{d} + \bar{u})$	0.255	-0.143	8.041	6.112	1.000	13.2
xs	0.064	-0.143	8.041	6.112	1.000	3.3

CTEQ4A-series of parton distributions with varying α_s and $G(x,Q)$

In exploring the range of variation of allowed $G(x,Q)$ by varying the values of α_s, changing the number of parameters for the gluon, and altering the Q_{cut} of data selection, we have found the largest effect is due to the varying of α_s. Hence, in presenting a series of PDF's which give a reasonable representation of the range of possibilities, we use those generated with an α_s range centered around the CTEQ4M value of 0.116, which is close to the current world average [68]. This series will be designated as CTEQ4A-series (shorthand for CTEQ4Alpha)—CTEQ4A1, ...,CTEQ4A5, with CTEQ4A3 being the same as CTEQ4M. The χ^2 per point for the 1297 non-jet data points are (1.07,1.02,1.02,1.07,1.19) for those 5 sets respectively. The higher χ^2 values at low values of α_s are mainly due to the HERA DIS experiments; the higher

χ^2 values at high values of α_s are mainly due to the fixed-target DIS experiments [79]. The difference in χ^2 above minimum, especially for the highest value of α_s, is larger than in previous CTEQ analyses (e.g. CTEQ2ML vs. CTEQ2M) due to the sharply reduced errors on recent DIS data. However, the difference is comparable to that between the MRSJ and CTEQ4M χ^2s, cf, Table 1.3. Because correlations in the experimental errors are not available for all experiments, and hence have not been included in current global analyses; and since theoretical uncertainties are even harder to quantify, pragmatically, we take these χ^2 differences as being acceptable for present purposes.

Figure 1.14 shows the comparison of the CTEQ4A parton distribution sets with the two jet data sets, using CTEQ4M as the common calibration. The overall normalization factor on the jet data sets applied to the various fits range from 0.96 (for CTEQ4A1 on D0 points) to 1.02 (for CTEQ4A5 on CDF points), well within the experimental uncertainty of $\sim$ 5%. Comparing to Figure 1.9 and the range of normalization factors needed there ($0.92 - 1.08$, which is wider than the experimental error), we see the expected improvement of the agreement with the jet data.

The gluon distributions associated with the various values of α_s in this series are shown in Figure 1.15. Comparing the CTEQ4A-series to the C-series (same parametrization form for $G(x, Q_0)$), we see that the constraining influence of the jet data has a rather dramatic effect. The unstable behavior of the various curves observed in the C-series has been replaced by an orderly variation as one steps through the values of α_s within the range explored. We found, indeed, that for each value of α_s, the solution of $G(x, Q)$ is rather unique against perturbations in the fitting procedure.

One concern is that the variation in the CTEQ4A series is too small due to the

lack of treatment of systematic uncertainties in the jet data. To address this issue, we compare the change in the calculated jet cross-sections between the extremes of the CTEQ4A series to the largest E_t-dependent uncertainty in the CDF data. See Figure 1.16. It shows that the range of variation in the CTEQ4A series is about 10% in the moderate E_T range, while the experimental systematic uncertainty is about the same.[9] This observation lends some confidence that this series gives a reasonable estimate of the range of variation of $G(x, Q)$. To the extent that there are sources of uncertainty other than α_s, the variation in $G(x, Q)$ given here may be considered a minimum range. However, our study does indicate that the variation due to the uncertainty of α_s is the dominant one.

Figure 1.17 shows a comparison of some of the new gluon and singlet quark (sum over all flavors) distributions with those of CTEQ3M and MRSJ in the usual form $x f(x, Q)$ without the normalization factor, as in previous figures. On this conventional plot, differences in $G(x, Q)$ can be seen only in the small-x region, and the CTEQ3M and CTEQ4M gluons appear to be indistinguishable. Differences in the singlet quark distribution are more evident near $x = 0.01$. The fact that only small changes in the parton distributions result from adding so much new data in the global analysis is testament to the impressive progress in pinning down these parton distributions that has been made in recent years. These changes, though small, are nonetheless physically significant, as demonstrated by the substantial differences in χ^2 values between the new and old parton distribution sets on the precision experiments given in Tables 1.3 and 1.4.

---

[9]Of course, given the good agreement between the two Tevatron experiments [77], if the CDF jet data requires a significant change due to a systematic error, the D0 data would require the same change. There is little correlation in the apparatus of two experiments, however, there are common features of their analyses, e.g. both using cone algorithm and correction for background events in similar ways, which could contribute to a correlated systematic error between the two experiments.

Other CTEQ4 parton distributions

Along with the standard CTEQ4M $\overline{MS}$ parton distributions, we have also obtained corresponding parton distributions in the "DIS scheme"— CTEQ4D. CTEQ4D uses the same value of α_s (= 0.116) as CTEQ4M; it is obtained by fitting under identical conditions as CTEQ4M except that the hard cross-sections are evaluated in the DIS scheme. The χ^2 values of this fit are comparable to those of CTEQ4M. In addition to these two standard sets, for applications requiring leading order (LO) calculations and low values of the scale Q (LQ), we also provide appropriate parton distribution sets labelled CTEQ4L and CTEQ4LQ respectively. The CTEQ4LQ set can be used for $Q^2 > Q_i^2 = 0.5$ GeV^2. It was obtained by fitting the same data sets as the other PDF sets. Since the proper treatment of low Q data must involve more physics input (such as higher twist effects) than included here, CTEQ4LQ represents only an extrapolation of twist-two QCD physics into the low Q region—it is not intended to be a best fit. However, as demonstrated by the GRV parton distribution sets [80], this kind of extrapolation often turns out to compare rather well with data in the low Q region. Comparison of CTEQ4LQ structure functions to the NMC, E665 and H1 data in the range $1.0 < Q < 3.0$ GeV is shown in Figure 1.18. The parameters for CTEQ4D, CTEQ4L and CTEQ4LQ are given in Tables (1.6 - 1.8).

The rather remarkably consistent picture resulting from this round of CTEQ4 global analysis incorporating jet data from hadron collisions provides a new generation of improved parton distributions for making calculations and predictions on high energy processes both within and beyond the standard model. The more tightly constrained parton distributions can also lay the foundation for more stringent tests of the pQCD framework and provide the basis for discerning signals of new physics.

At present, a remaining area of some uncertainty is the gluon distribution in the

Table 5.6: Parameters for the CTEQ4D initial parton distributions at $Q_0 = 1.6$ GeV. The functional forms are described in Equation 3.16 with $\kappa = 1/2$. Also, $\alpha_s(m_z) = 0.116$, corresponding to $\Lambda^{(5)} = 202$ MeV.

Parton	a_0	a_1	a_2	a_3	a_4	% Momentum
xd_v	0.724	0.490	3.839	1.688	0.338	11.3
xu_v	1.528	0.490	3.554	6.448	1.162	30.4
xg	2.141	-0.058	7.554	36.405	2.223	43.7
$x(\bar{d} - \bar{u})$	0.054	0.490	7.200	0.000	30.000	–
$x(\bar{d} + \bar{u})$	0.154	-0.227	7.200	6.949	1.000	11.7
xs	0.038	-0.227	7.200	6.949	1.000	2.9

Table 5.7: Parameters for the CTEQ4L initial parton distributions at $Q_0 = 1.6$ GeV. The functional forms are described in Equation 3.16 with $\kappa = 1/2$. Also, LO $\Lambda^{(5)} = 181$ MeV

Parton	a_0	a_1	a_2	a_3	a_4	% Momentum
xd_v	0.702	0.443	4.003	2.433	0.622	10.9
xu_v	1.226	0.443	3.465	7.589	1.146	30.1
xg	0.854	-0.305	3.666	1.846	1.968	41.8
$x(\bar{d} - \bar{u})$	0.050	0.443	6.877	0.000	30.000	–
$x(\bar{d} + \bar{u})$	0.201	-0.200	6.877	5.644	1.000	13.8
xs	0.050	-0.200	6.877	5.644	1.000	3.5

"large x" region, beyond say 0.25, where neither the DIS nor the direct photon data give tight constraints. For the DIS process, the sensitivity to the gluon begins below $x = 0.1$. For the direct photon process there are a number of theoretical uncertainties which are not yet under control, as already discussed in Sec.2.3. The noticeable rise of the inclusive jet data points [49] above all "theory" curves shown so far may be related to the conventional choices of parametrization of the non-perturbative function $G(x, Q_i)$, which restricts its behavior in the large x region. This possibility, first raised in Ref. [48], will be discussed next in the context of the CTEQ4 analysis presented above.

Table 5.8: Parameters for the CTEQ4LQ initial parton distributions at $Q_0 = 0.7$ GeV. The functional forms are described in Equation 3.16 with $\kappa = 1/6$. Also, $\Lambda^{(5)} = 174$ MeV

Parton	a_0	a_1	a_2	a_3	a_4	% Momentum
xd_v	0.852	0.573	4.060	4.852	0.693	14.7
xu_v	1.315	0.573	3.281	10.614	1.034	40.4
xg	39.873	1.889	5.389	0.618	0.474	31.2
$x(\bar{d} - \bar{u})$	0.093	0.573	7.293	0.000	30.000	-
$x(\bar{d} + \bar{u})$	0.578	0.143	7.293	1.858	1.000	11.7
xs	0.096	0.143	7.293	1.858	1.000	1.9

5.5 High E_t Jets and Parton Distributions

The higher-than-expected inclusive jet cross-sections, first measured by the CDF collaboration [49] for $E_t > 200$ GeV, were observed in comparison to the predictions of the existing parton distribution sets, including CTEQ3M, as shown in Figure 1.8. It triggered wide excitement as a possible signal of "New Physics" [78], because that was the highest energy scale ever reached, in other way of saying, the shortest distance ever probed where the sub-structure of quarks (if it exsits) could be tested. Before jumped into the conclusion, we investigated the feasibility of accommodating these higher cross-sections in the conventional QCD framework by exploiting the flexibility of $G(x, Q)$ at higher values of x where there are few independent constraints, while maintaining the agreement with other data sets in the global analysis [48]. To do this, it is necessary to (i) provide enough flexibility in the parametrization of $G(x, Q_0)$ to allow for behaviors different from the usual (but arbitrary) choice[10] ; and (ii) focus on the high E_t data points and assign them more statistical weight than their nominal values in order to force a better agreement between theory and experiment. Thus, the

---

[10]For examples, our choice of general functional form in Equation 3.16 does provide the flexibility, whereas the choice of other group ($xg = a_0 x^{a_1}(1 - x)^{a_2}(1 + a_3\sqrt{x} + a_4x)$) failed [81].

spirit of the investigation is not to obtain a "best fit" in the usual sense. Rather, it is (i) to find out whether such solutions exist; and (ii) if they do exist, to quantify how well these solutions agree with other data sets as compared to conventional parton distribution sets. The global analysis work described in Sec. 1.4 without special attention to the high E_t points provides the natural setting to put the results of Ref. [48] in context.

Ref. [48] was performed using the CDF Run-IA data—the only high statistics inclusive jet measurement available at the time. Two illustrative "solutions" of the type described above were presented—one with the normalization fixed at 1.0 with respect to the CDF data, the other with a normalization factor of 0.93. Figure 1.19 compares predictions of the normalization=1.0 PDF set, which we shall refer to as the CTEQ4HJ set, with the more recent Run-1B results of both CDF and D0. For this comparison, an overall normalization factor of 1.01(0.98) for the CDF(D0) data set is found to be optimal in bringing agreement between theory and experiment.[11] The consistency between the two data sets, as well as between theory and experiment, displayed by this comparison appears to be rather remarkable (again, bearing in mind the neglect of systematic errors other than overall normalization). Results shown in Table 1.3 quantify the χ^2 values obtained while accommodating the high E_t jets in the global fit in this particular case. Compared to the best fit CTEQ4M, the overall χ^2 for CTEQ4HJ is indeed slightly higher. But the difference is much smaller than the differences discussed earlier in the CTEQ4A series, and much smaller than the difference between MRSJ and CTEQ4M. Thus the price for accommodating the high E_t jets is negligible. In addition, the difference between CTEQ4HJ and CTEQ4M is almost entirely due to the BCDMS data, even though the BCDMS χ^2 for CTEQ4HJ by itself is quite good. This difference is due to the fact that, in

_____

[11]The change of CDF normalization factor from 1.0 to 1.01 is attributable to the switch from the Run-1A to the Run-1B data set.
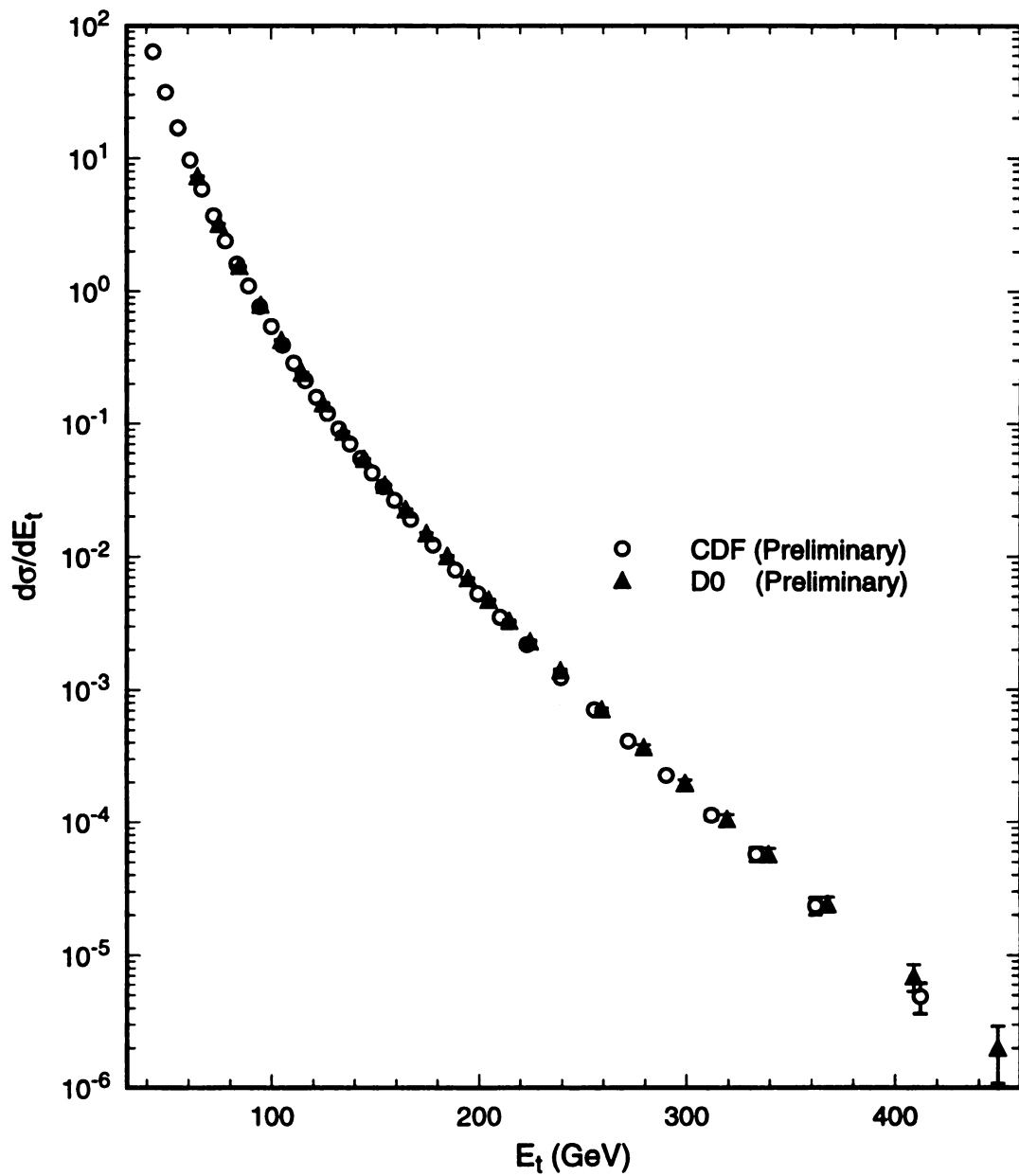
Figure 5.6: Inclusive jet cross-section measured by the CDF and D0 collaborations in Run-1B at the Tevatron. (Averaged over $0.1 < |\eta| < 0.7$ in the case of CDF and $|\eta| < 0.5$ in the case of D0.)
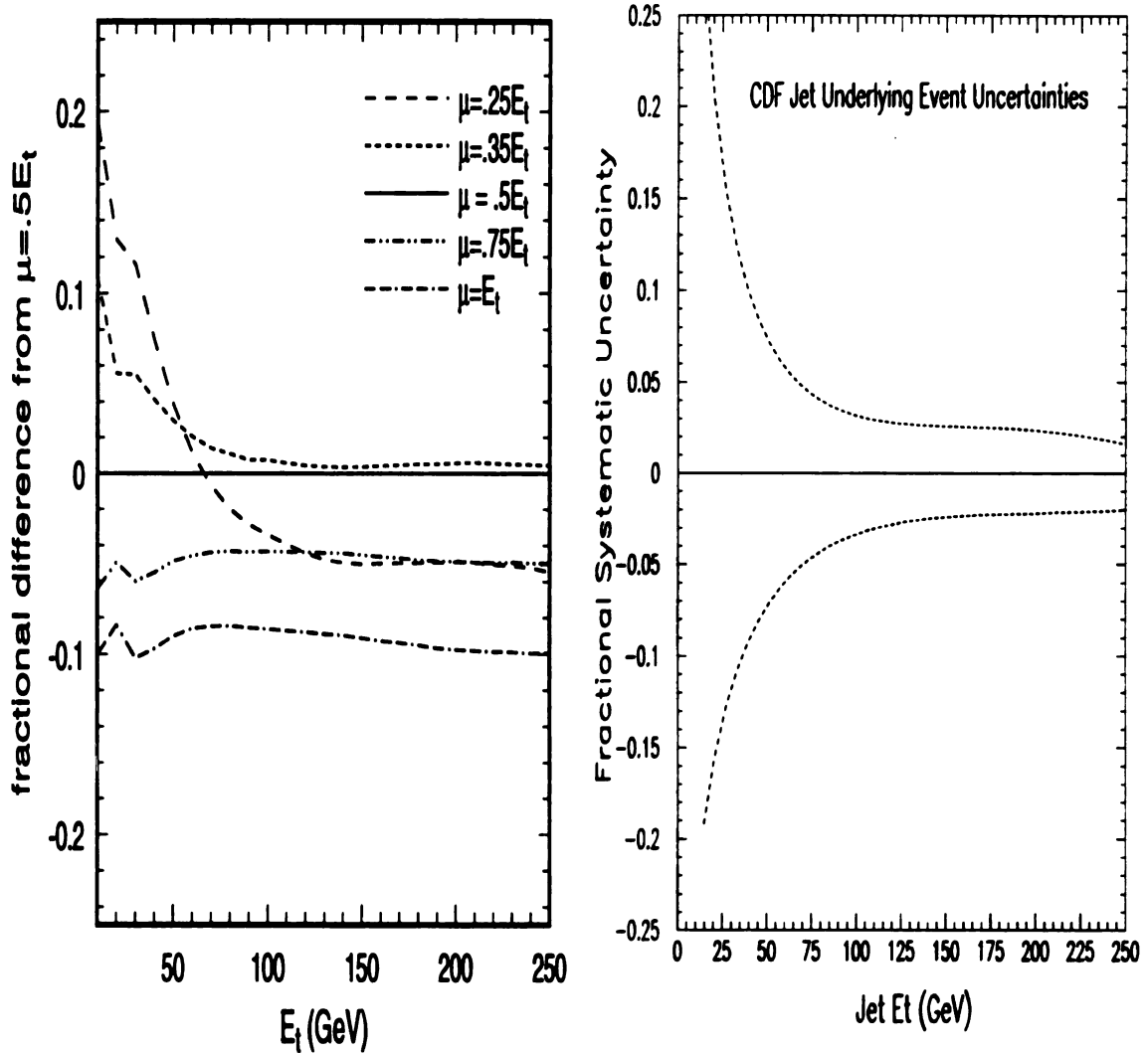
Figure 5.7: Two examples of sources of uncertainties in comparing inclusive jet data with NLO QCD theory: (a) Fractional difference between $d\sigma(E_t, \mu)/dE_t$ and $d\sigma(E_t, \mu = E_t/2)/dE_t$ (for the CDF rapidity coverage $0.1 < |\eta| < 0.7$) as a function of E_t for a variety values of μ; (b) Fractional change in the cross-section due to ± 30 % change in underlying event correction in the CDF experiment.
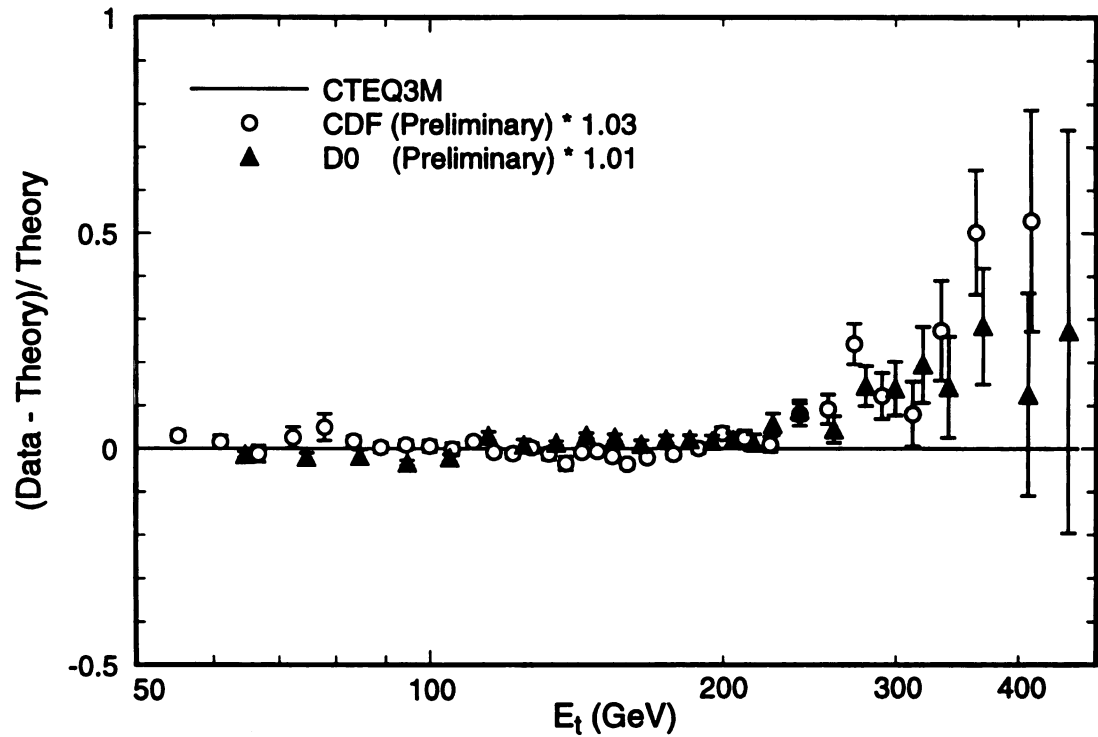
Figure 5.8: Inclusive jet cross-section measured by the CDF and D0 collaborations in Run-1B at the Tevatron normalized to NLO QCD calculations based on CTEQ3M PDF's. The difference in rapidity coverage of the two experiments is taken into account.
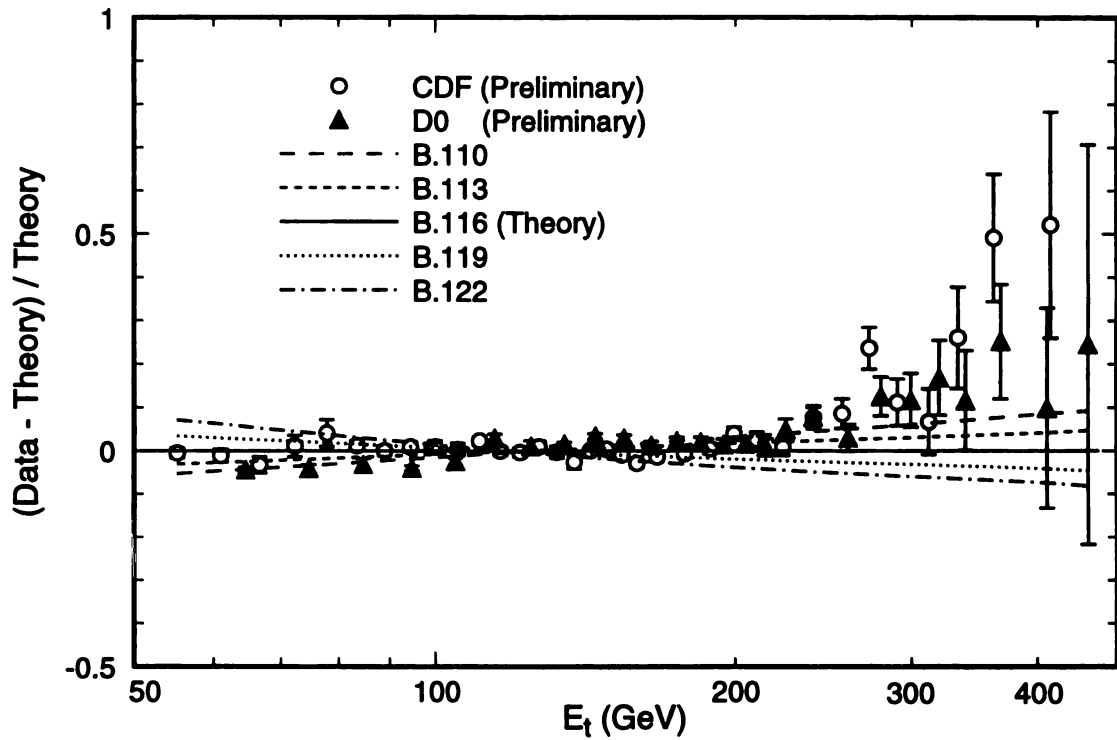
Figure 5.9: Inclusive jet cross-section of CDF and D0 compared to NLO QCD calculations based on the new B-series parton distributions.
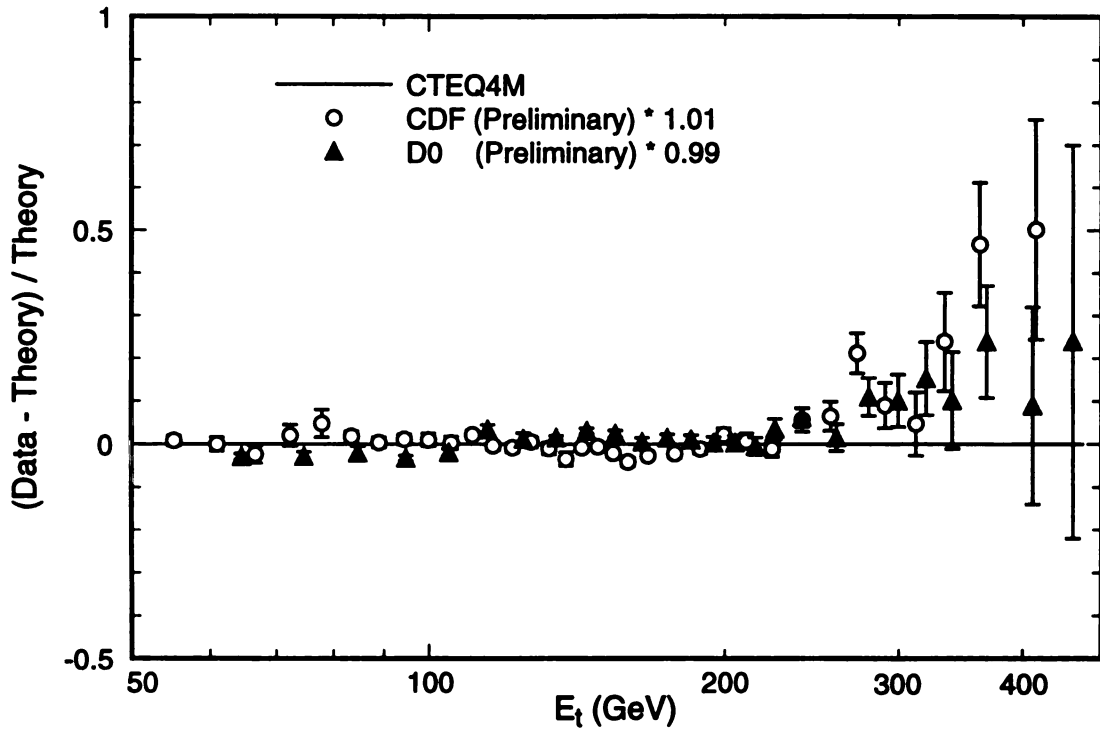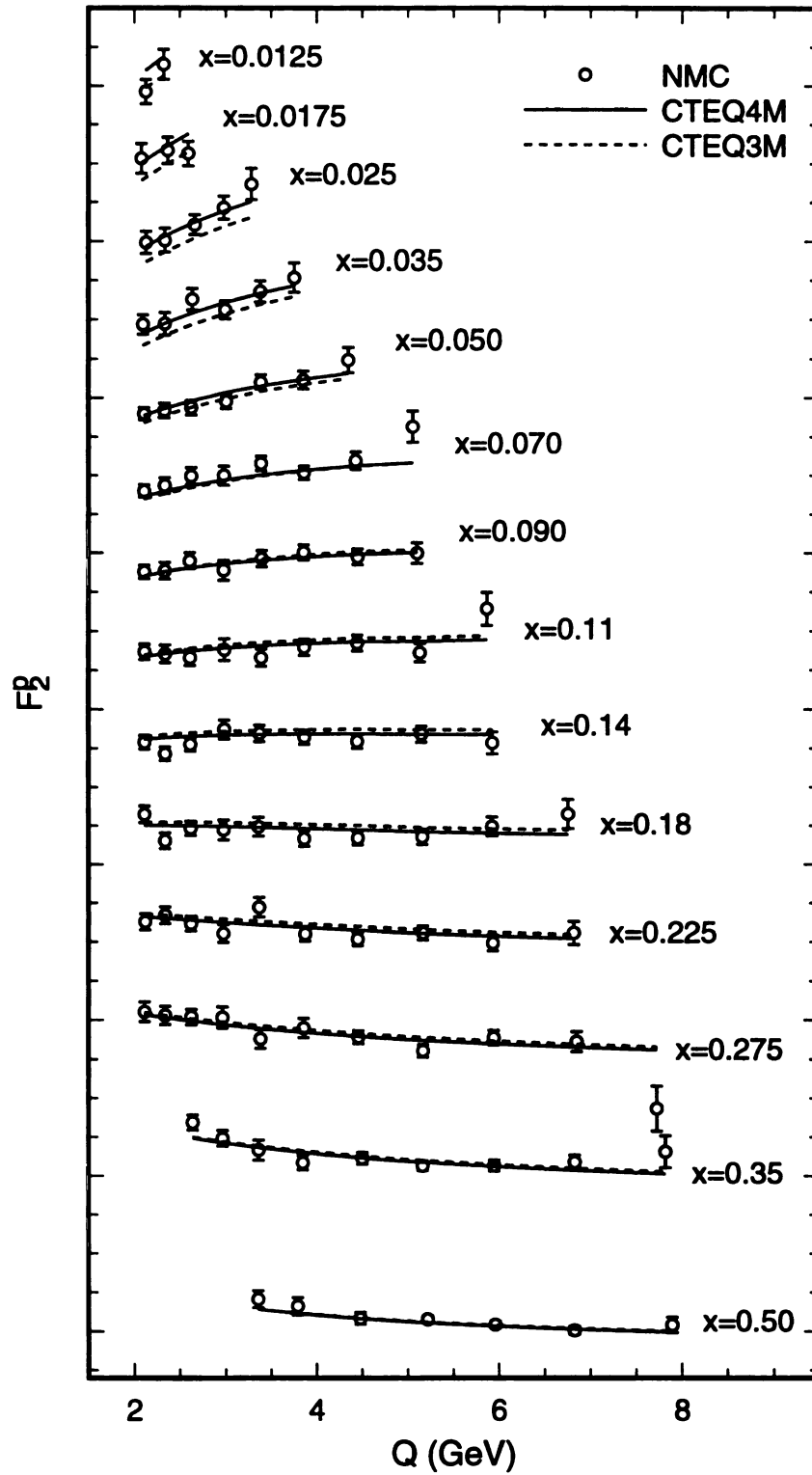
Figure 5.10: Inclusive jet cross-section of CDF and D0 compared to NLO QCD calculations based on the new CTEQ4M parton distributions.

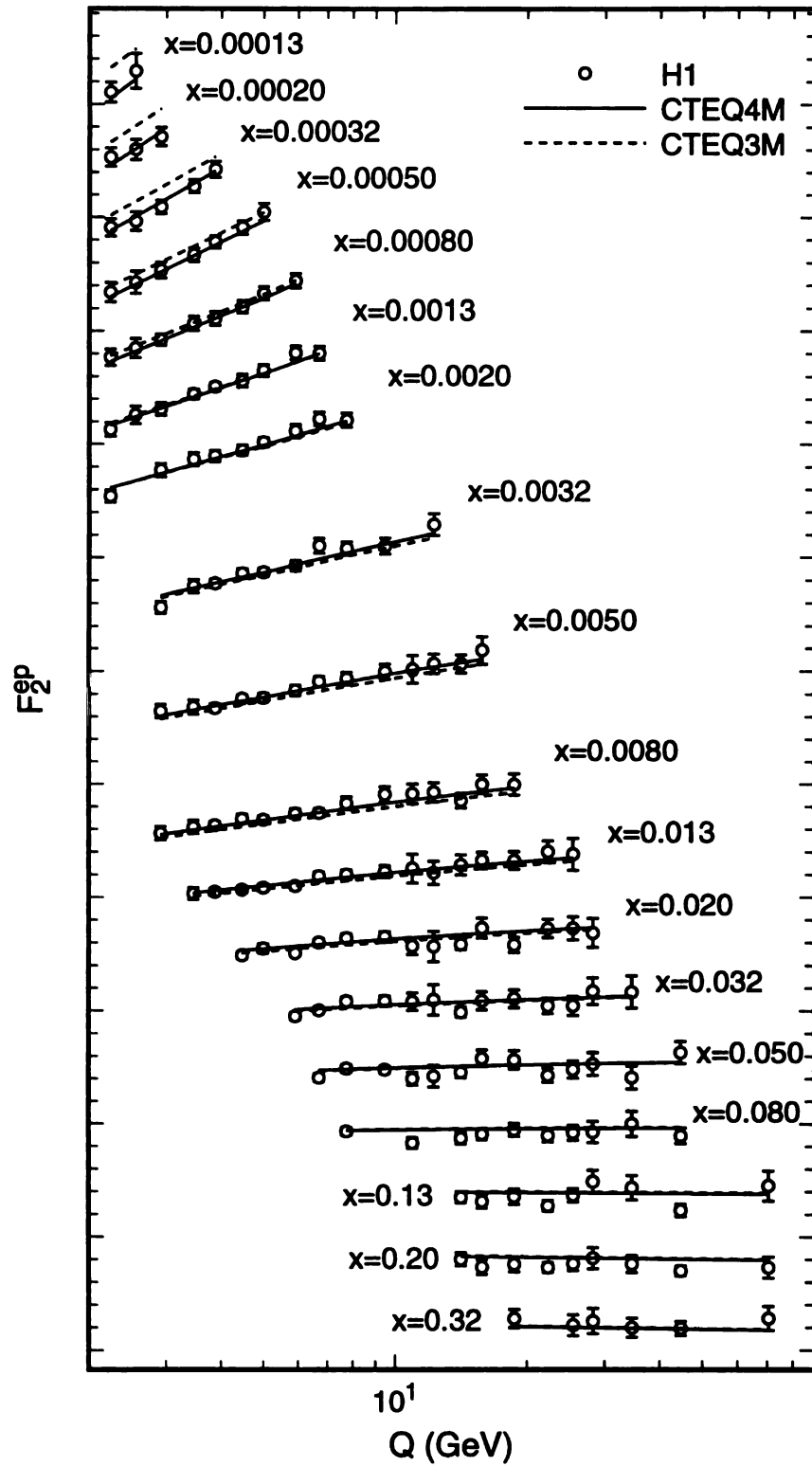Figure 5.11: Comparison of F_2^p data from NMC to NLO QCD calculations based on CTEQ3M and CTEQ4M.

Figure 5.12: Comparison of F_2^p data from H1 to NLO QCD calculations based on CTEQ3M and CTEQ4M. The improvement in the small-x region is evident.
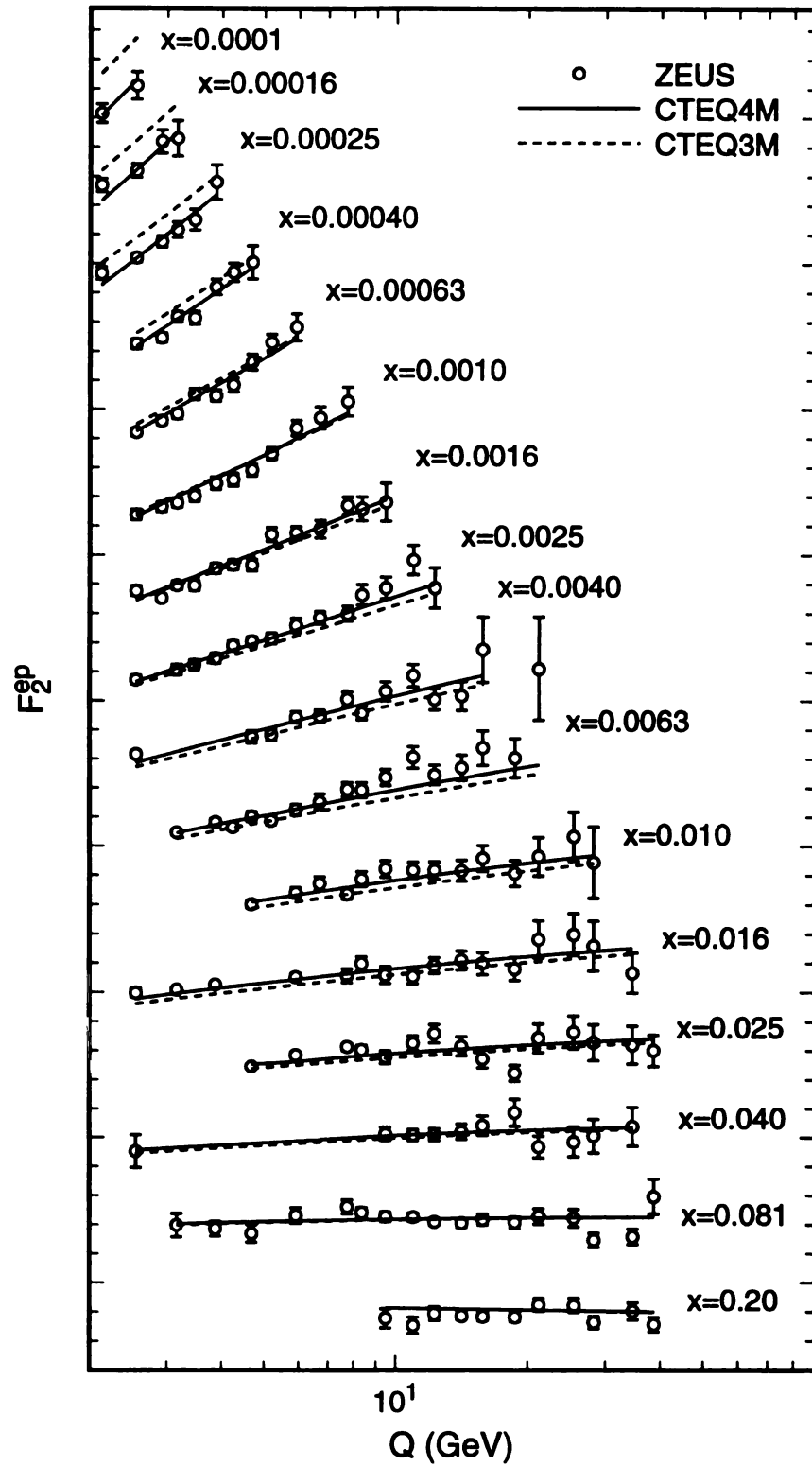
Figure 5.13: Comparison of F_2^p data from ZEUS to NLO QCD calculations based on CTEQ3M and CTEQ4M. The improvement in the small-x region is evident.
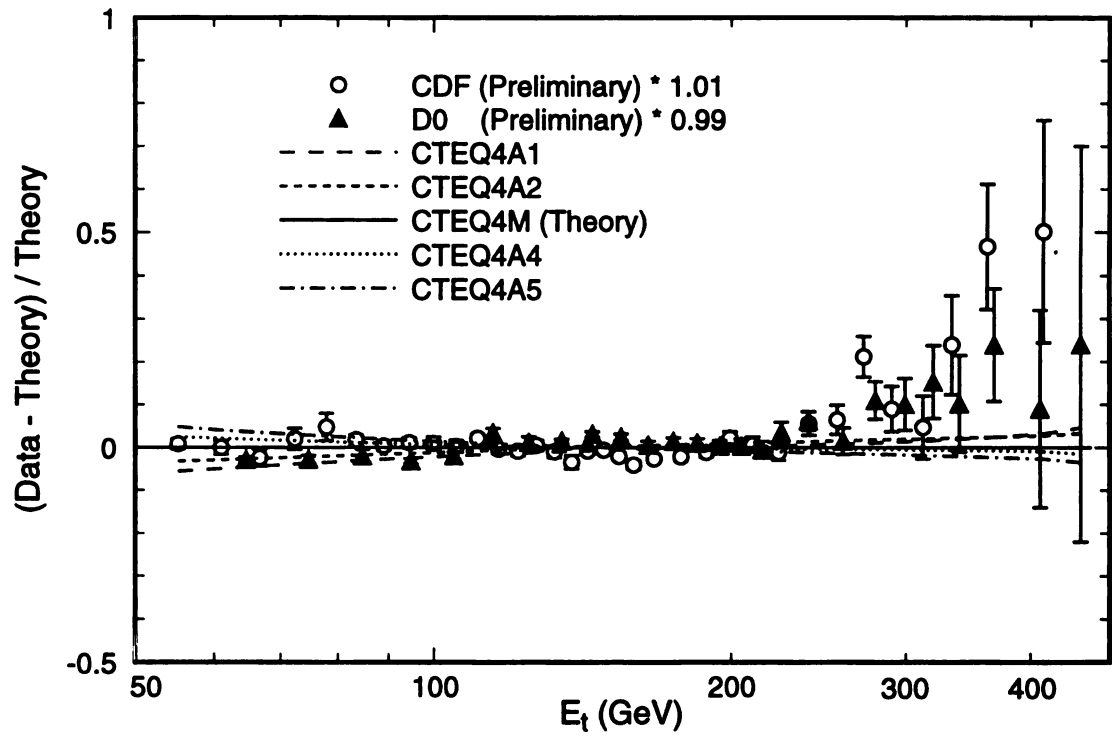
Figure 5.14: Inclusive jet cross-section of CDF and D0 compared to NLO QCD calculations based on the new CTEQ4A series of parton distributions.
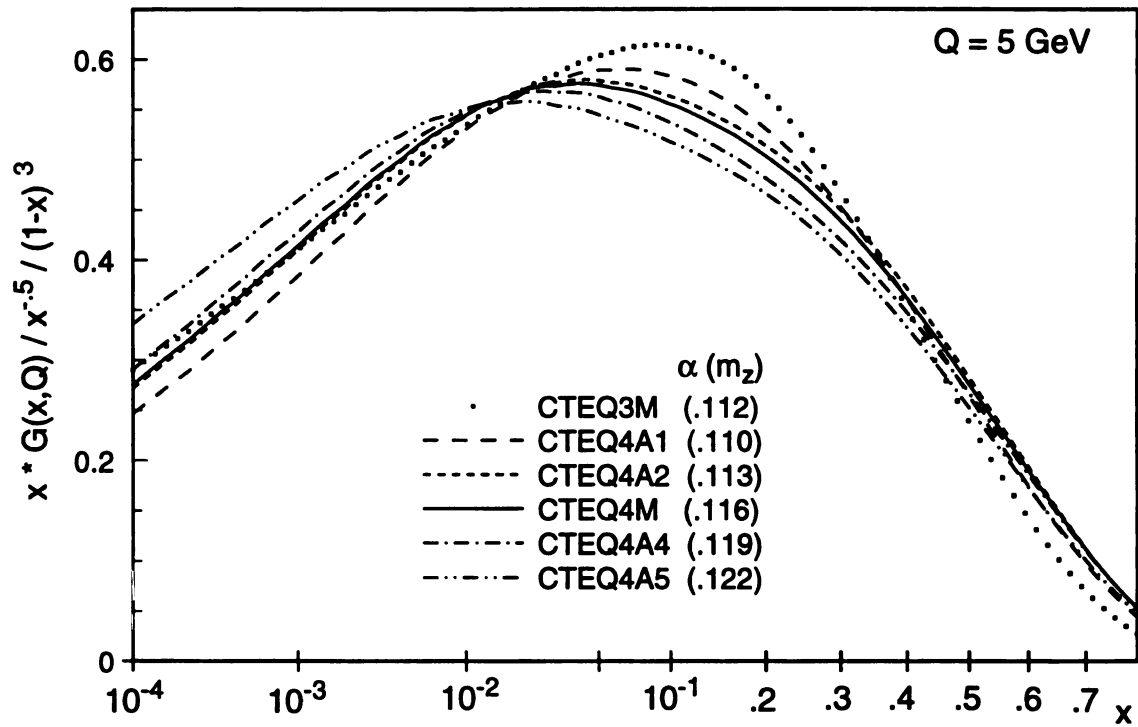
Figure 5.15: Series-CTEQ4A gluon distributions normalized by the function $x^{-1.5}(1-x)^3$.
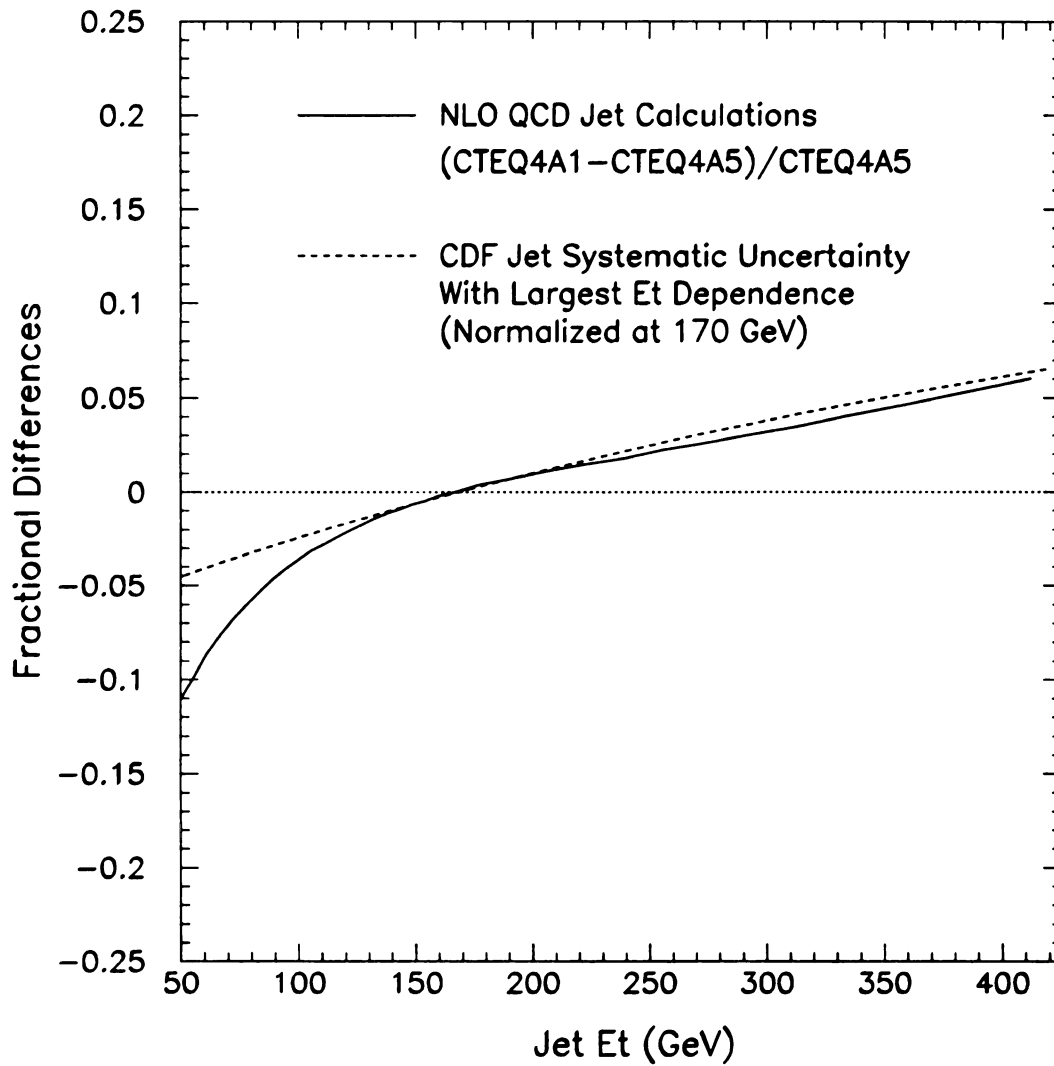
Figure 5.16: Percentage range of variation of the inclusive jet cross-section from the two extreme CTEQ4A PDF sets (CTEQ4A1 and CTEQ4A5) compared to the largest of the E_t dependent systematic uncertainties.
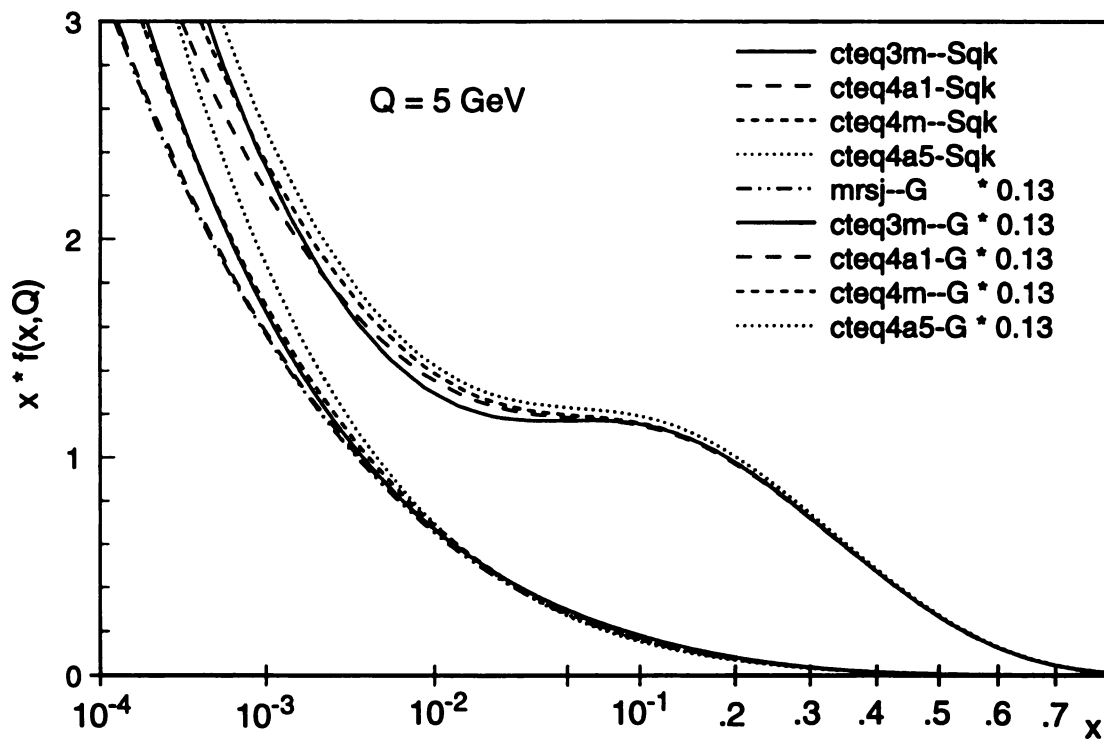
Figure 5.17: Comparison of $x\,G(x,Q)$ and $x\,S(x,Q)$ between some new parton distribution sets and those from CTEQ3M. $S(x,Q)$ is the sum of all quark (singlet quark) distributions. The CTEQ3M and CTEQ4M gluons appear to lie on top of each other. The same is true for the CTEQ4A1 and MRSJ gluons. Differences in $G(x,Q)$ for $x > 0.01$ are not evident in this plot.
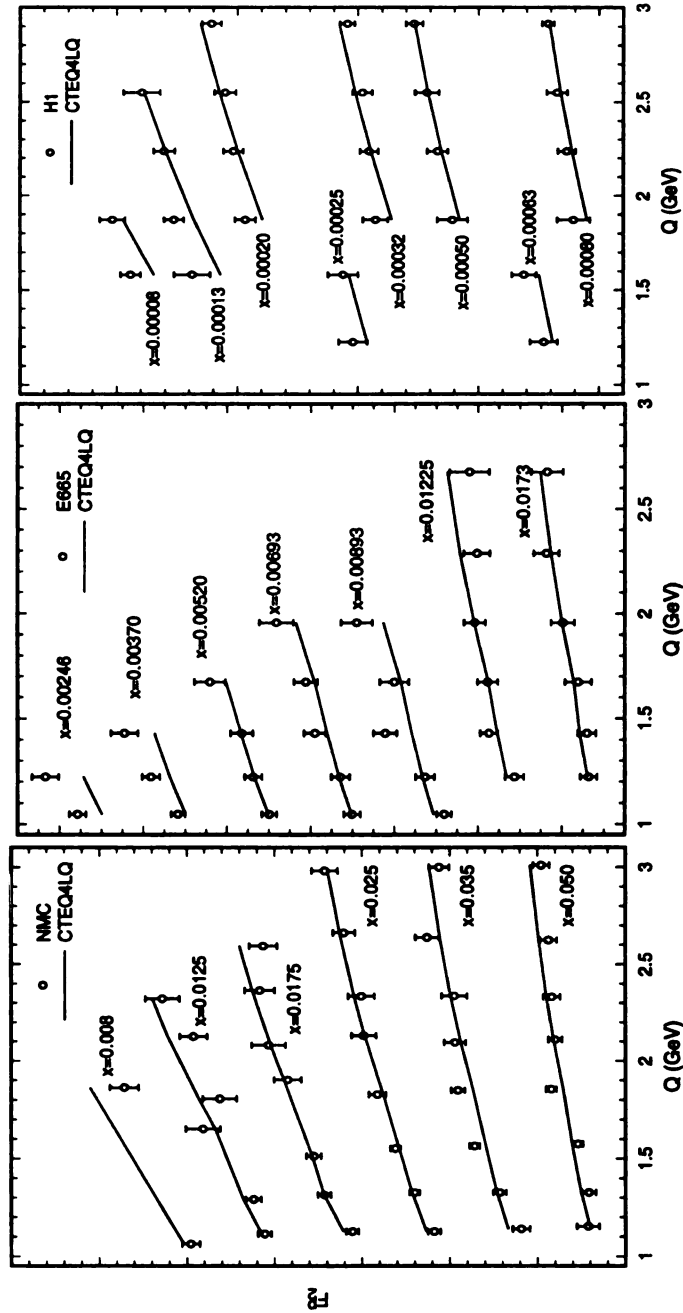
Figure 5.18: Comparison of F_2^p data in the low-Q region from H1, E665 and NMC to NLO QCD calculations based on CTEQ4LQ PDF's. CTEQ4LQ is obtained by fitting to data with $Q > 2$ GeV only. The extrapolation to below $Q = 2$ GeV appears to work remarkably well except for the two lowest x bins of the E665 data shown.
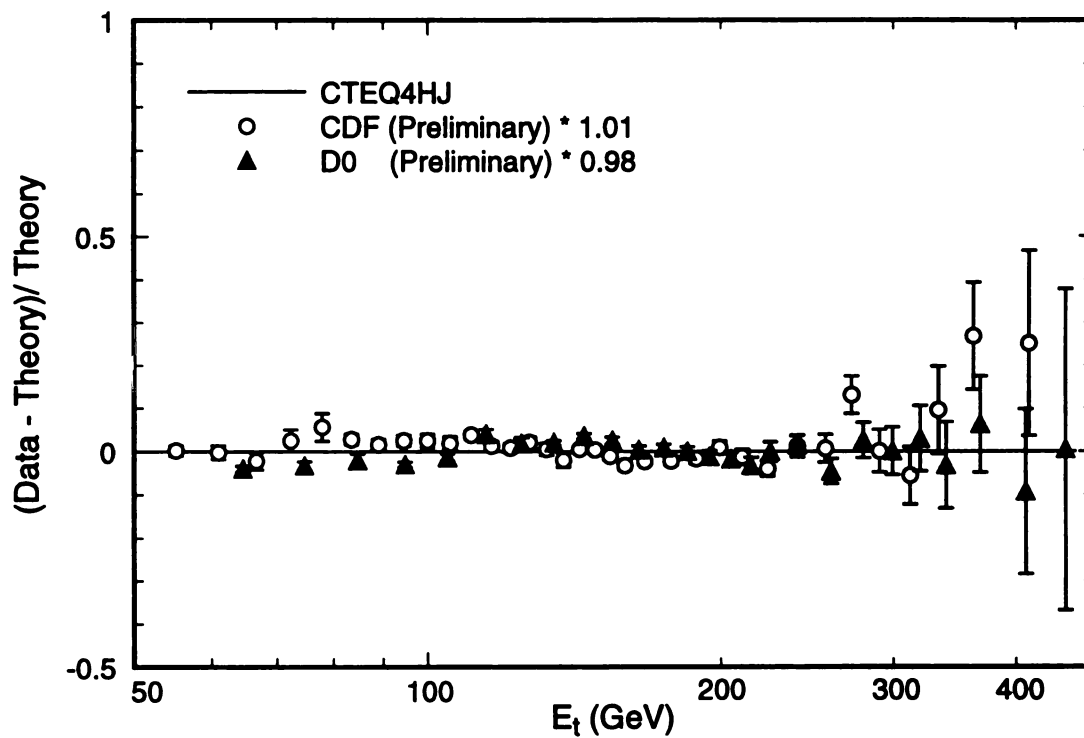
Figure 5.19: Inclusive jet cross-section of CDF and D0 compared to NLO QCD calculations based on the CTEQ4HJ parton distributions.
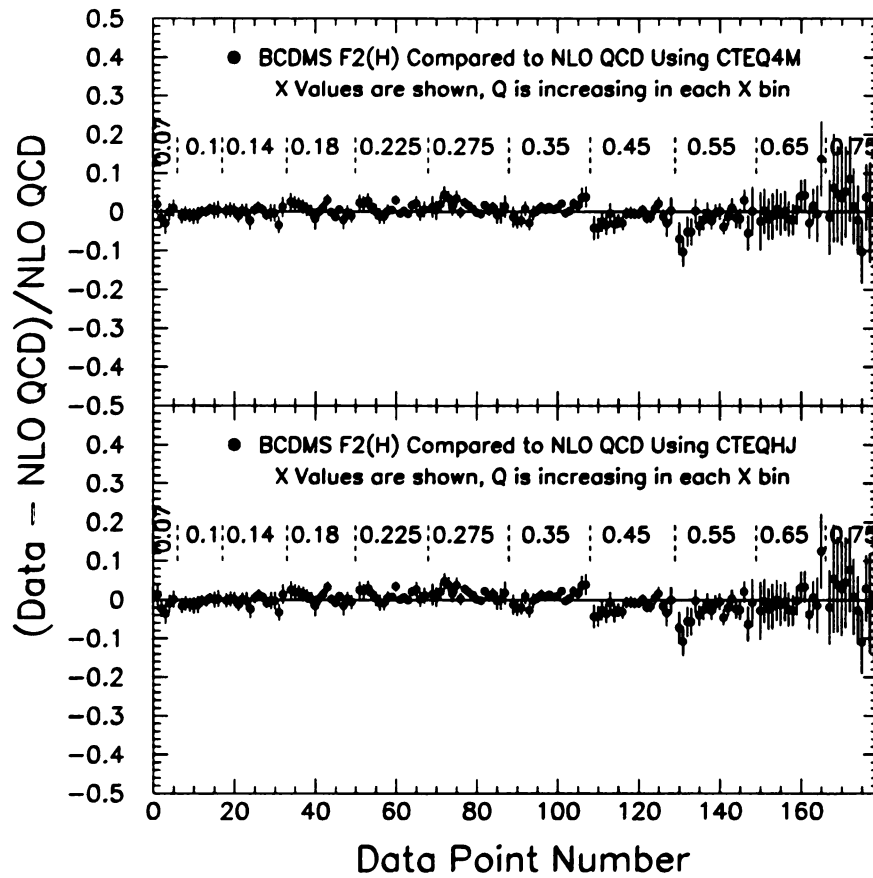
Figure 5.20: Percentage deviation of BCDMS proton data from NLO QCD values based on CTEQ4M and CTEQ4HJ. Both PDF sets give good fits.

Chapter 6

Further Development and Summary

The sequence of CTEQ analyses gives a realistic view of the manner in which progress in theory and experiment interact as the characteristics of the various parton distributions are investigated. The precision of the data and the diversity of physical processes together allow detailed investigations of *fine structures* such as the breaking of flavor symmetry in the sea. The remaining uncertainties on quark distributions concern detailed flavor differentiation, particularly among the sea quarks. New measurements on vector boson production (W-, Z- and continuum lepton-pair) will be valuable, as illustrated by the first CDF results on A_W and NA51 on A_{DY}.

In Figures (6.1 - 6.4), we show the migration of parton distributions at $Q = 5$GeV and $Q = 100$GeV in the various standard sets obtained in the sequence of CTEQ analyses with continuous improvement of measurables and the corresponding calculations. The latest version of CTEQ analysis, CTEQ4, provides an excellent description of a wealth of data covering an extended range in both Q^2 and x compared to what was available just a few years ago. We see significant progress in demonstrating the consistency of the NLO QCD framework, and in narrowing the uncertainties on the elusive but important gluon distribution. Specifically,
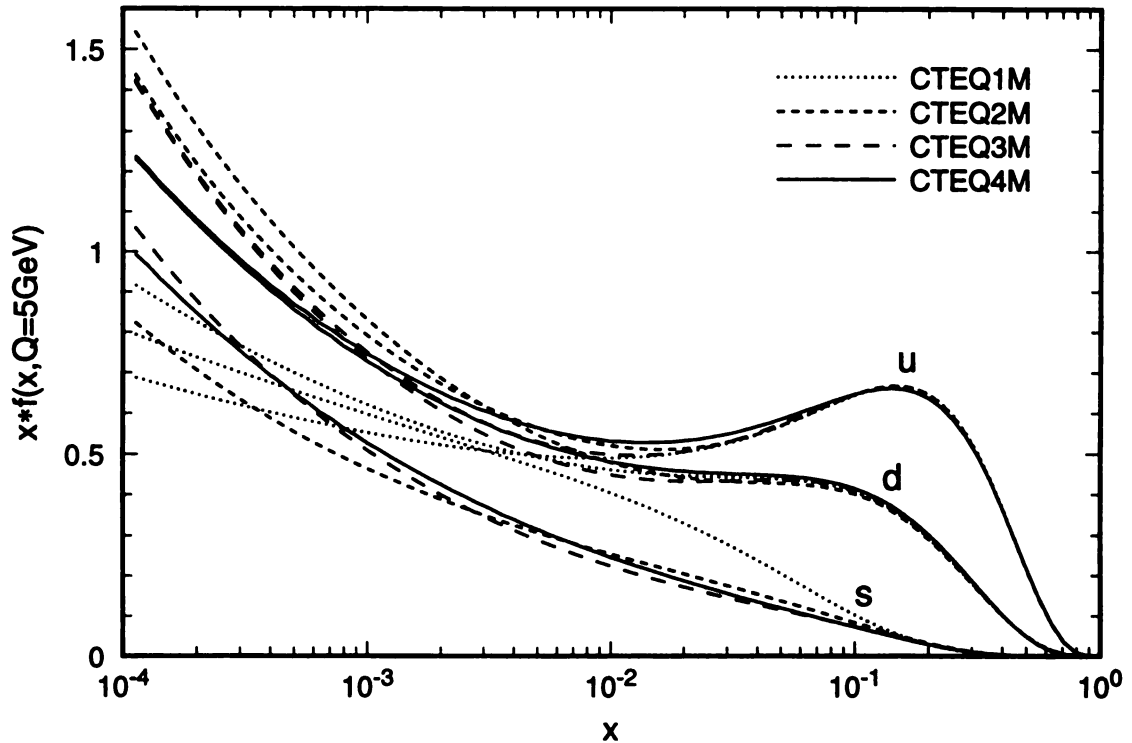
105

Figure 6.1: Migration of CTEQ1-CTEQ4 parton distributions at $Q = 5\text{GeV}$: u, d, and s.

- The recent NMC, E665, H1 and ZEUS data considerably narrow down parton distributions and limit the behavior of the gluon, especially if one uses the minimal form of the gluon parameterization used by CTEQ3;

- The new inclusive jet data agree well with theory predictions based on PDF's determined by the other processes, with the possible exception of the high E_t data points.

- By adding jet data to the global analysis, it is possible to further explore the range of variation of the gluon distribution using a more general parametrization. Although the jet data set covers a limited x-region, its effect is felt over the entire x-range – because it complements the other data sets well.

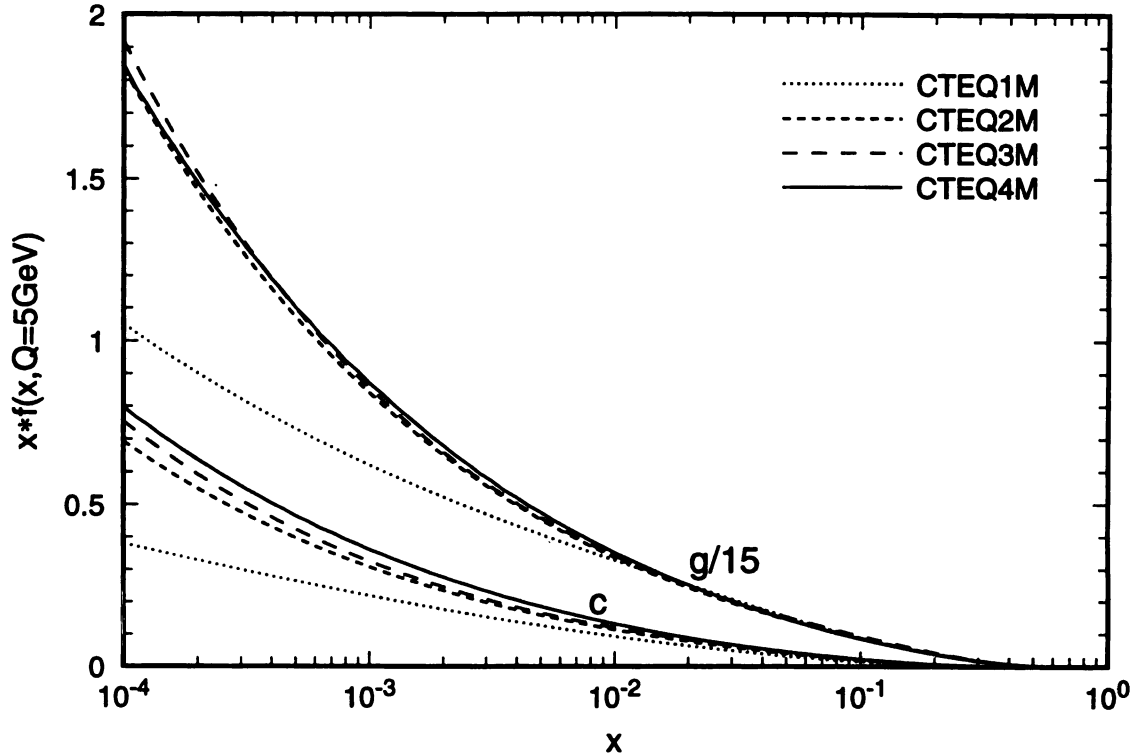- Three sources contributing to the uncertainty of the gluon distribution have

Figure 6.2: Migration of CTEQ1-CTEQ4 parton distributions at $Q = 5\text{GeV}$: g and c.

been investigated: (i) by letting α_s vary over its current range of uncertainty; (ii) by increasing the degree of freedom for parametrizing the non-perturbative initial gluon distribution, and (iii) by varying the Q_{cut} in selecting data for the global fits. The largest effect is due to α_s.

• These studies help to delineate the range of variation of $G(x, Q)$ over the range $10^{-4} < x < 0.25$. Further work is needed in exploring the range of uncertainty of the gluon and other parton distributions by systematically varying the relevant parameters of the global analysis.

• For larger values of x, more definitive experimental results on inclusive jet and direct photon production as well as improved theory are needed for further progress. The observed high E_t "excess" jet cross-section can be accommodated
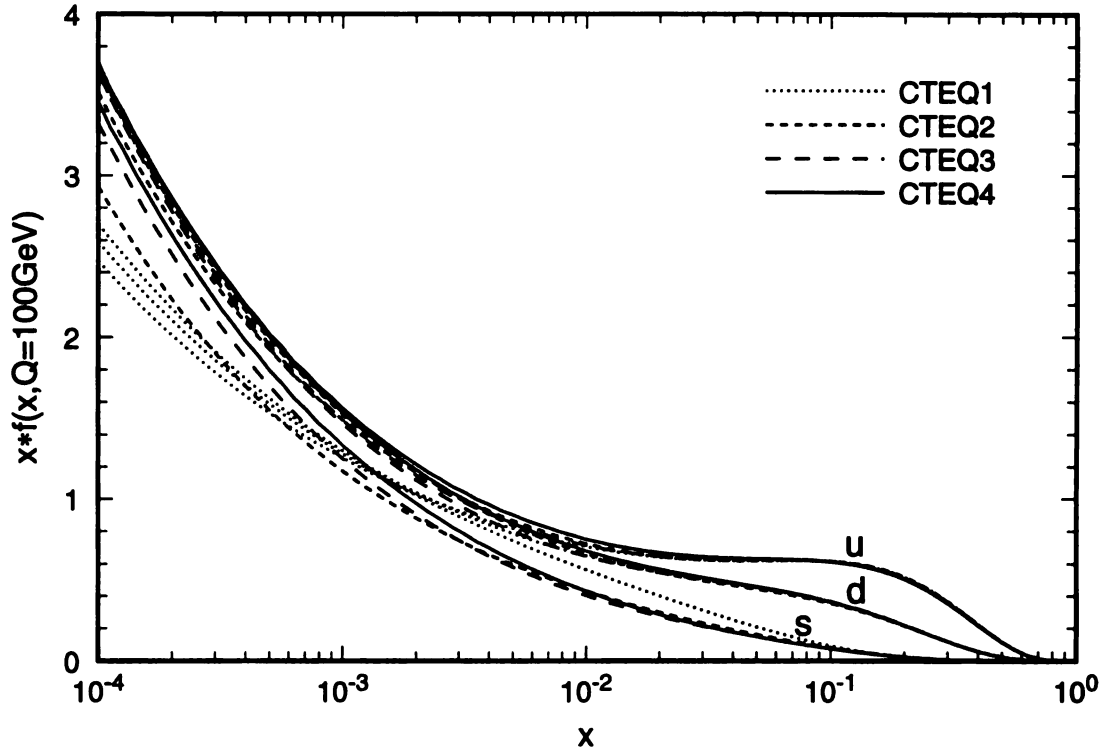
Figure 6.3: Migration of CTEQ1-CTEQ4 parton distributions at $Q = 100\text{GeV}$: u, d, and s.

by a modified gluon distribution, represented by the CTEQ4HJ set, since no other independent measurement constrains it in this range.

Existing global analyses have mostly been performed within the traditional zero-mass parton picture since, up to now, "heavy quark" partons (charm, bottom and top) have played a relatively minor role in measured observables used in the analyses. With the advent of precise data on inclusive F_2 [82] and the direct measurement of the charm component F_2^c [83, 84] from HERA, this is no longer the case. The latter comprises about 25% of the inclusive structure function at small-x, as illustrated in Figure 6.5. It is now necessary to sharpen the formulation of the theory for heavy flavor production used in these global analyses. Thus, it opens a new area which was neglected before but becomes more and more important as measurements probe new kinematics and reach high precision.
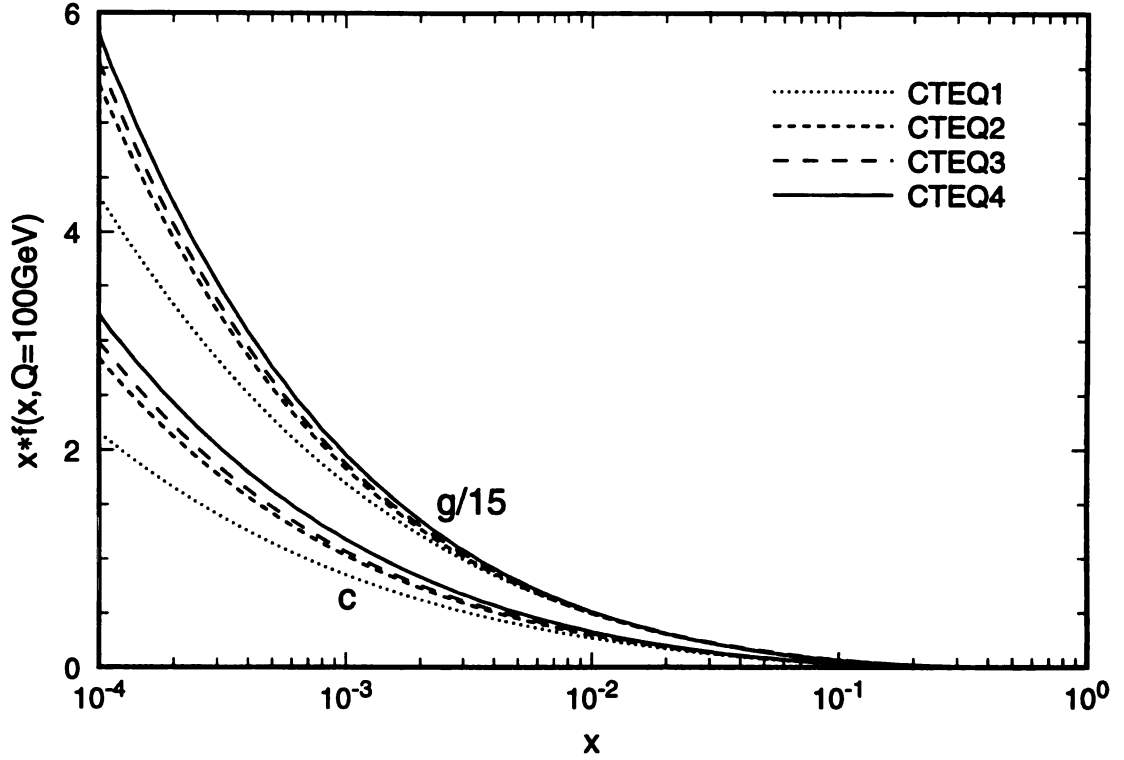
Figure 6.4: Migration of CTEQ1-CTEQ4 parton distributions at $Q = 100\text{GeV}$: g and c.

The leptoproduction formalism of Aivazis, Collins, Olness and Tung (ACOT) [58] represents a natural generalization of the conventional zero-mass QCD parton framework to include heavy quark mass effects. In comparison to recent analyses using the massless approach, significant differences only arise in deep inelastic structure functions at small-x where the charm contribution to the measured cross-section becomes a significant fraction, as shown in Figure 6.6. For the region away from very small-x, data cannot distinguish between the zero-mass scheme and the ACOT scheme in our study [59]. In this sense, the continued use of existing parton distribution sets (such as CTEQ [27] and MRS [29]) obtained from the zero-mass scheme for processes away from the sensitive region is acceptable. However, for processes sensitive to initial or final state heavy quarks, it will be imperative to use the more complete theory, with matching parton distributions, if meaningful physical quantities are to be extracted.
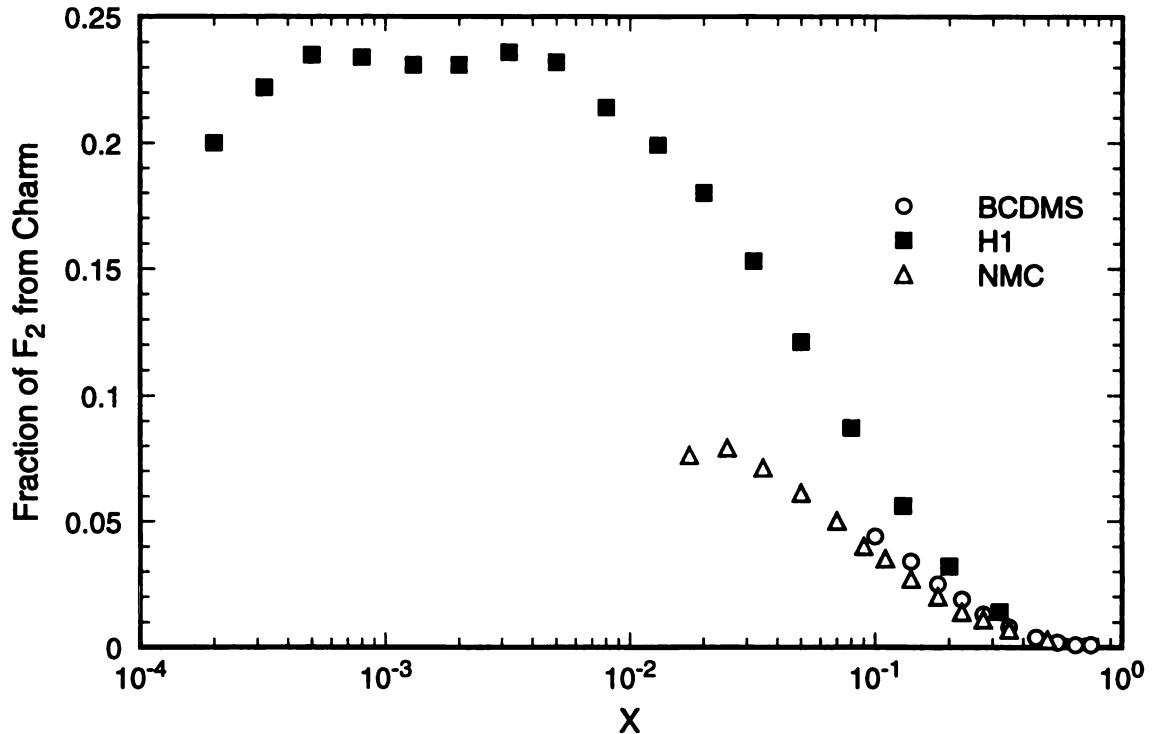
Figure 6.5: Fractional contribution to F_2 from the charm quark for different experimental (x,Q) range. The data shown are calculated using CTEQ4M with average on the Q bins of that particular experiment at fixed x.

Thus, we have extended the global analysis using ACOT scheme for the first time, with the additional semi-inclusive data set F_2^c from H1. We obtain CTEQ4HQ1 parton distributions representing the best fit in this scheme. The change in parton distributions is shown in Figure 6.7. Although we now only include heavy quark mass effects on DIS, we expect little effect on measurables of other processes that are not sensitive to heavy quark contributions and measurables with large energy scale, such as W production and inclusive jets with large E_t, where "heavy quark" becomes relatively light compared to the prevalent energy scale. Our current data in the global analysis generally satisfy the above criteria. For a more complete global analysis considering heavy quark mass effects, current theory can still be improved by including other processes with heavy quark mass effects taken into account, as
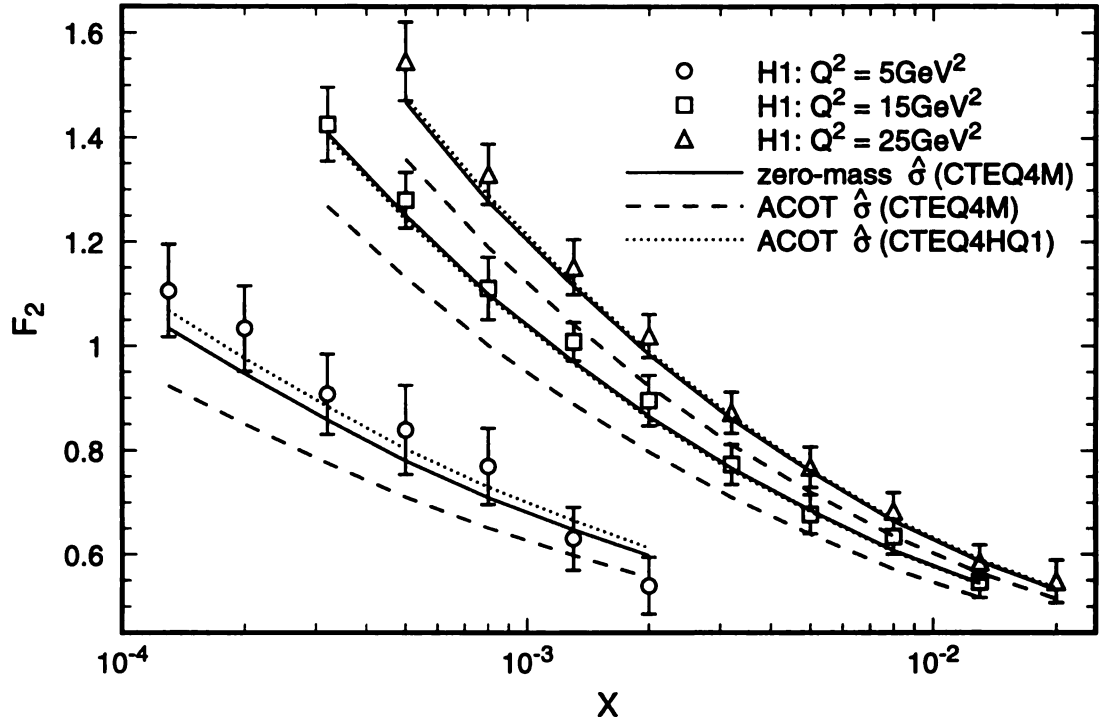
Figure 6.6: Comparison of H1 data in the small-x region with calculations using CTEQ4M parton distributions in the (original) zero-mass scheme (solid line) and with ACOT hard cross-section (dashed line). Also shown is the result of the new fit CTEQ4HQ1 with consistent ACOT scheme (dotted line).

well as to be extended to higher order terms; and detailed phenomenology is yet to be done when both experiment and theory mature.

Figure 6.7: Change in parton distributions due to different treatment of M_c.

Bibliography

[1] D.J. Griffiths, *Introduction to Elementary Particles*, John Wiley & Sons, Inc., 1987.

[2] G. Kane, *Modern Elementary Particle Physics*, Addison-Wesley, Inc., 1987.

[3] Review of Particle Properties, *Phys. Rev.* **D54**, 1 (1996).

[4] N. Cabibbo, *Phys. Rev. Lett.* **10**, 531 (1963).

[5] S.L. Glashow, J.Iliopoulos & L. Maiani, *Phys. Rev.* **D2**, 1585 (1970).
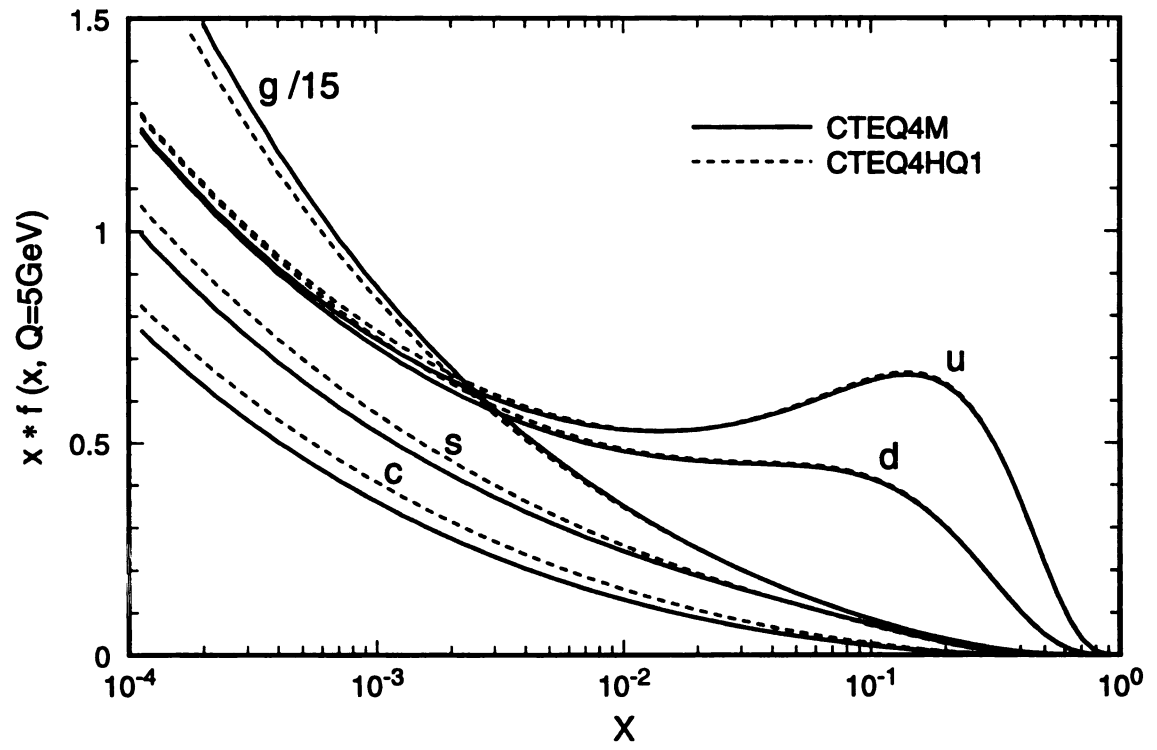
[6] M. Kobayashi and K. Maskawa, *Prog. Theor. Phys.* **49**, 652 (1973).

[7] J.D. Bjorken and E.A. Paschos, *Phys. Rev.* **185**, 1975 (1969).

[8] R.P. Feynman, *Phys. Rev. Lett.* **23**, 1415 (1969).

[9] E.D. Bloom, D.H. Coward, H. DeStaebler, J. Drees, G. Miller, L.W. Mo, R.E. Taylor, M. Breidenbach, J.I. Friedman, G.C. Hartmann, and H.W. Kendall *Phys. Rev. Lett.* **23**, 930 (1969); M. Breidenbach, J.I. Friedman, E.D. Bloom, D.H. Coward, H. DeStaebler, J. Drees, L.W. Mo, and R.E. Taylor, *Phys. Rev. Lett.* **23**, 935 (1969); J.I. Friedman and H.W. Kendall, *Annu. Rev. Nucl. Part. Sci.* **22**, 203 (1972).

[10] D. Amati, R. Petronzio & G. Veneziano, *Nucl. Phys.* **B140**, 54; **B146**, 29 (1978); R.K. Ellis, H. Georgi, M. Machacek, H.D. Politzer & G.G. Ross, *Nucl. Phys.* **B152**, 285 (1979); S.B. Libby & G. Sterman, *Phys. Rev.* **D18**, 3252 (1978); J.C. Collins, D.E. Soper & G. Sterman, in *Perturbative Quantum Chromodynamics*, ed. A.H. Mueller, World Scientific (1989).

[11] E. Witten, *Nucl. Phys.* **B104**, 445 (1976); W.J. Marciano, *Phys. Rev.* **D29** 580, (1984).

[12] J. Collins, F. Wilczek, and A. Zee, *Phys. Rev.* **D18**, 242 (1978); J. Collins, W.-K. Tung *Nucl. Phys.* **B278**, 934 (1986).

[13] R.G. Roberts, *The Structure of the Proton*, Cambridge University Press, 1990; W.K. Tung, in *Proceedings of the workshop on Hadron Structure Functions and Parton Distributions*, World Scientific, 1990.

[14] V.N. Gribov and L.N. Lipotov, *Yad. Fiz.* **15**, 781 (*Sov. J. Nucl. Phys.* **15**, 438) (1972).

[15] L.N. Lipotov, *Yad. Fiz.* **20**, 181 (*Sov. J. Nucl. Phys.* **20**, 94 (1975)) (1974).

[16] G. Altarelli and G. Parisi *Nucl. Phys.* **B126**, 298 (1977).

[17] S.D. Drell, D.J. Levy, and T.M. Yan, *Phys. Rev. Lett.* **22**, 744 (1969); *Phys. Rev.* **187**, 2159 (1969); *Phys. Rev.* **D1**, 1035 (1970); S.D. Drell and T.M. Yan, *Phys. Rev. Lett.* **25** 316, (1970); T.M. Yan and S.D. Drell, *Phys. Rev.* **D1** 2402, (1970).

[18] P. Aurenche *et al.*, *Nucl. Phys.* **B286** 553, (1987).

[19] CTEQ Collaboration: J. Huston *et al.*, *Phys. Rev.* **D51** 6139, (1995).

[20] E.L. Berger, X. Guo, J. Qiu, Phys. Rev. **D53**, 1124 (1996); *Phys. Rev. Lett.* **76**, 2234 (1996)

[21] G. Sterman & S. Weinberg, *Phys. Rev. Lett.* **39**, 1436 (1977).

[22] S. Catani, Yu.L. Dokshitser, M. Olsson, G. Turnock & B.R. Webber, *Phys.Lett.* **B269**, 432 (1991).

[23] S. Ellis, Z. Kunszt, and D. Soper, *Phys. Rev. Lett.* **64** 2121, (1990); S. Ellis, Z. Kunszt, and D. Soper, *Phys. Rev. Lett.* **69**, 3615 (1992).

[24] F. Aversa *et al.*, *Phys. Rev. Lett.*, **65** 401, (1990).

[25] J. Botts, J.G. Morfin, J.F. Owens, Jianwei Qiu, Wu-Ki Tung, and H. Weerts, *Phys. Lett.* **B304**, 159 (1993).

[26] CTEQ Collaboration (H.L. Lai *et al.*), *Phys. Rev.* **D51**, 4763 (1995).

[27] CTEQ Collaboration (H.L. Lai *et al.*), *Phys. Rev.* **D55**, 1280 (1997).

[28] A.D. Martin, W.J. Stirling, and R.G. Roberts, *Phys. Rev.* **D50**, 6734 (1994); *Phys. Lett.* **B354**, 155 (1995).

[29] A.D. Martin, R.G. Roberts, and W.J. Stirling, *Phys. Lett.* **B387**, 419 (1996).

[30] NMC Collaboration (P. Amaudruz *et al.*), *Phys. Lett.* **B295**, 159 (1992).

[31] CCFR Collaboration (W.C. Leung *et al.*), *Phys. Lett.* **B317**, 655 (1993); and (P.Z. Quintas *et al.*), *Phys. Rev. Lett.* **71**, 1307 (1993).

[32] BCDMS Collaboration (A.C. Benvenuti, *et.al.*.), *Phys. Lett.* **B223**, 485, (1989); and *Phys. Lett.* **B237**, 592 (1990).

[33] ZEUS Collaboratio, *Phys. Lett.* **B316**, 412 (1993); H1 Collaboration, *Nucl. Phys.* **B407**, 515 (1993).

[34] L.W. Whitlow *et al.*, *Phys. Lett.* **B282**, 475 (1992); and (A. Bodek *et al.*), *Phys. Rev.* **D20**, 1471 (1979).

[35] J.J. Aubert *et.al.*, *Nucl. Phys.* **B293**, 740 (1987); K. Bazizi, S.J. Wimpenny, and T. Sloan, *Proceedings of the 25th International Conference on High Energy Physics*, Singapore, (1990).

[36] H. Abramowicz *et al.*, *Phys. Rev. Lett.* **57**, 298 (1986); *Z. Phys.* **C28**, 51 (1985); J.P. Berge *et al.* , *Z. Phys.* **C49**, 187 (1991).

[37] J. Gomez *et al.*, *Phys. Rev.* **D49**, 4348 (1994); R.G. Arnold *et al.*, *Phys. Rev. Lett.* **52**, 727 (1984).

[38] EMC Collaboration (M. Arneodo *et al.*), *Nucl. Phys.* **B331** 1 (1990); J. Ashman *et al.*, *Z. Phys.* **C57**, 211 (1993).

[39] NMC Collaboration (R. Seitz and A. Witzmann), in *Proceedings of the XXVIIIth Rencontre de Moriond*, Ed. Tran Thanh Van, Editions Frontieres (1993)

[40] E605 Collaboration (G. Moreno *et al.*), *Phys. Rev.* **D43**, 2815 (1991).

[41] CDF Collaboration (F. Abe *et al.*), *Phys. Rev.* **D49**, 1 (1994).

[42] NA51 Collaboration (A. Baldit *et al.*), *Phys. Lett.* **B332**, 244 (1994).

[43] S.D. Ellis and W.J. Stirling, *Phys. Lett.* B256, 258 (1991).

[44] CDF Collaboration (F. Abe *et al.*), *Phys. Rev. Lett.* **74**, 850 (1995).

[45] M. Bonesini *et al.*, *Z. Phys.* **C38**, 371 (1988).

[46] UA6 Collaboration (A. Bernasconi *et al.*), *Phys. Lett.* **B206**, 163 (1988).

[47] E706 Collaboration (G. Alverson *et al.*), *Phys. Rev.* **D48**, 5 (1993).

[48] CTEQ Collaboration (J. Huston *et al.*), *Phys. Rev. Lett.* **77**, 444 (1996).

[49] CDF Collaboration Run-IA: (Abe *et al.*), *Phys. Rev. Lett.* **77**, 438 (1996); and Run-IB: B. Flaugher, Talk given at APS meeting, Indianapolis, May, 1996.

[50] D0 Collaboration: G. Blazey, Talk given at Rencontre de Moriond, March, 1996; D. Elvira, Talk given at Rome conference on DIS and Related Phenomena, April, 1996.

[51] F. Aversa *et al.*, *Phys. Rev. Lett.* **65**, 401 (1990).

[52] W. Giele *et al.*, *Nucl. Phys.* **B403**, 2121 (1993).

[53] See Refs. [23], [51] and [52].

[54] For a review and detailed references, see J.F. Owens and W.K. Tung, "Parton Distributions of Hadrons" in *Ann. Rev. Nucl. Sci.* **42**, 291 (1992).

[55] A.D. Martin, W.J. Stirling, and R.G. Roberts, *Phys. Rev.* **D47**, 867 (1993); *Phys. Lett.* **B306**, 145 (1993).

[56] NMC Collaboration (M. Arneodo *et al.*), *Phys. Rev.* **D50**, 1 (1994).

[57] J.C. Collins and Wu-Ki Tung, *Nucl. Phys.* **B278**, 934 (1986).

[58] M.G. Aivazis, J.C. Collins, F. Olness & Wu-Ki Tung, *Phys. Rev.* **D50**, 3102 (1994).

[59] H.L. Lai and Wu-Ki Tung, *Z. Phys.* **C74**, 463 (1997).

[60] CCFR Collaboration A. Bazarko in Proceedings of 18th Rencontres de Moriond, Ed. Tran Thanh Van, Editions Frontieres, (1993).

[61] CDHSW Collaboration (H. Abramowicz *et al.*), *Phys. Rev. Lett.* **57**, 298 (1986); *Z. Phys.*, **C28**, 51 (1985).

[62] Charm Collaboration (J.V. Allaby, *et al.*), *Z. Phys.* **C36**, 611 (1987); *Phys. Lett.* **197B**, 281 (1987); *Phys. Lett.* **B213**, 554 (1988).

[63] CCFR Collaboration (K. Lang *et.al.*), *Z. Phys.* **C33**, 483 (1987); Rabinowitz *et al., Phys. Rev. Lett.*, **70** 134 (1993).

[64] M.G. Aivazis, F. Olness & Wu-Ki Tung, *Phys. Rev. Lett.* **65**, 2339 (1990).

[65] ZEUS Collaboration: M. Lancaster, Invited talk given at the 27th International Conference on High Energy Physics, July 1994, Glasgow Scotland.

[66] S. Bethke, *42nd Scottish Universities Summer School in Physics (SUSSP 93): High Energy Phenomenology (NATO Advanced Study Institute)*, St. Andrews, Scotland, Aug 1993 Heidelberg preprint HD-PY-93-07.

[67] T. Jaroscewicz, *Phys. Lett.* **116B**, 291 (1982).

[68] See, for example, Particle Data Group (L. Montanet *et al.*), *Review of Particle Properties*, Phys. Rev. **D50**, 1173 (1994); S. Bethke, p. 213, *30th Rencontres de Moriond: QCD and High Energy Hadronic Interactions*, Ed. J. Trân Thanh Vân. Editions Frontieres (1995).

[69] Some related papers use a wider range of α_s, extending up to 0.125 or 0.130. Such large values are in strong disagreement with the highly precise BCDMS and CCFR experiments. Specifically, in a typical global fit involving around 1000 data points, the χ^2 value increases by about 200 above the best fit, mostly from these two experiments.

[70] D. Gross and F. Wilczek, *Phys. Rev.* **D9**, 980 (1974); V.S. Fadin, E.A. Kuraev, L.N. Lipatov, *Phys. Lett.* **60B**, 50 (1975).

[71] S.J. Brodsky and G.R. Farrar, *Phys. Rev. Lett.* **31**, 1153 (1973); *Phys. Rev.* **D11**, 1309 (1975); V. Matveev, R. Muradyan, A. Tavkhelidze, *Lett. Nuovo Cimento* **5**, 907 (1972); **7** 719 (1973).

[72] NMC Collaboration (M. Arneodo *et al.*), *Phys. Lett.* **B364**, 107 (1995).

[73] H1 Collaboration (S. Aid *et al.*): "1993 data", *Nucl. Phys.* **B439**, 471 (1995); "1994 data", *Nucl. Phys.* **B472**, 32 (1996)

[74] ZEUS Collaboration (M. Derrick *et al.*): "1993 data", *Z. Phys.* **C65**, 379 (1995) ; "1994 data", *Z. Phys.* **C72**, 399 (1996).

[75] E665 Collaboration (M.R. Adams *et al.*), *Phys. Rev.* D54, 3006 (1996).

[76] G. Sterman and G. Korchemsky, *Nucl. Phys.* **B437**, 415 (1995); R. Akhoury and V.I. Zakharov, *Nucl.Phys.* **B465**, 295 (1996); Yu.L. Dokshitser, G. Marchesini, B.R. Webber, *Nucl.Phys.* **B469**, 93 (1996).

[77] H.L. Lai and W.K. Tung, "Comparison of CDF and D0 Inclusive Jet Cross-sections", MSU-HEP-60508, CTEQ-605, e-Print Archive: hep-ph/9605269.

[78] See, for example; J. Ellis and D. A. Ross, *Phys. Lett.* **383**, 187 (1996); V. Barger, K. Cheung, P. Langacker, *Phys. Lett.* **381**, 226 (1996); J. Lykken, *Phys. Rev.* **D54**, 3693 (1996); R.S. Chivukula, A.G. Cohen, E.H. Simmons, *Phys. Lett.* **380**, 92 (1996).

[79] The set with $\alpha_s = 0.105$ is not included here and in subsequent comparisons because it is strongly disfavored by the new H1 and ZEUS data included in the fit to the B- and C-series, by the inclusive jet data used in the CTEQ4 fits, and by the LEP measurements of α_s – even if the total χ^2 of the $\alpha_s = 0.105$ fit on DIS and DY data sets by themselves is comparable to that of the $\alpha_s = 0.122$ set.

[80] M. Gluck, E. Reya, A. Vogt, *Z. Phys.* **C67**, 433 (1995).

[81] E.W.N. Glover, A.D. Martin, R.G. Roberts & W.J. Stirling, *Phys. Lett.* **B381**, 353 (1996).

[82] H1 collaboration (S. Aid *et al.*), *Nucl. Phys.* **B472**, 32 (1996). ZEUS collaboration (M. Derrick *et al.*), *Z. Phys.* **C72**, 399 (1996).

[83] H1 collaboration (C. Adolff *et al.*), *Z. Phys.* **C72**, 593 (1996).

[84] ZEUS collaboration (J. Breitweg *et al.*), DESY-97-089, hep-ex/9706009 (1997).