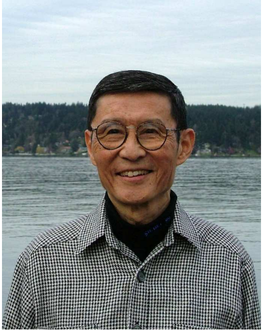
Parton distribution functions and their collider applications

Pavel Nadolsky (SMU)
J. Huston, H.-L. Lai, J.Pumplin, D. Stump, W.-K. Tung, C.-P. Yuan
progress on **CT09 PDF set**

Rik Yoshida and Steve McGill (ANL)
HERAPDF0.2 set

Wu-Ki Tung

1939-2009



- Professor of Physics at IIT, Michigan State University and University of Washington; well known for his work on hadronic physics
- A founder and long-term leader of the Coordinated Theoretical-Experimental Project on QCD (CTEQ)

Fermilab-Pub-90/74
IIT-PHY-90/11

PARTON DISTRIBUTIONS FROM A GLOBAL QCD ANALYSIS OF DEEP INELASTIC SCATTERING AND LEPTON-PAIR PRODUCTION

Jorge G. Morfín[a*,c] and Wu-Ki Tung[a,b,†]

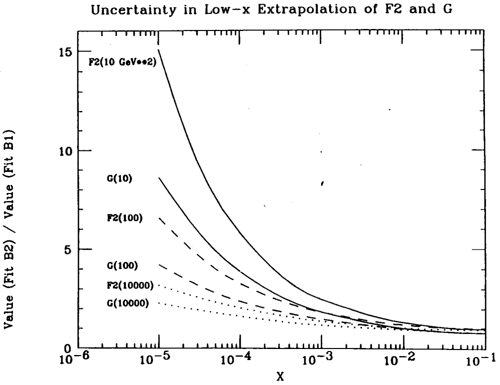
[a]Fermi National Accelerator Laboratory, Batavia, IL 60510
[b]Illinois Institute of Technology, Chicago, IL 60616
[c]Lab. de Fisica d'Altes Energies, Uni. Autonoma de Barcelona, Spain

ABSTRACT

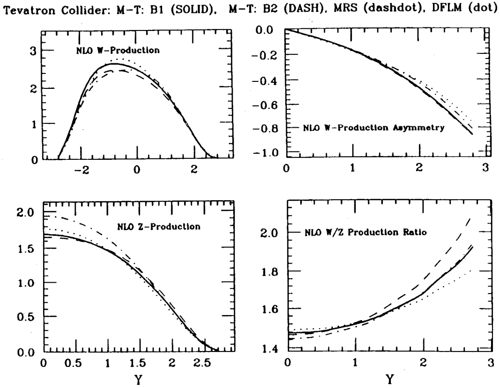
Parton Distribution Functions consistent with neutrino and muon deep inelastic scattering as well as Drell-Yan pair production results have been extracted. This analysis incorporates experimental systematic errors which are the dominant errors in recent deep inelastic scattering experiments. The dependence of the results on factors such as kinematic cuts in the data, heavy target corrections, and choice of initial functional form are also explored. The form adopted is motivated by perturbative QCD and particularly useful in exploring the small-x extrapolation of the distributions. This is crucial for studying the range of predictions for Collider, HERA, and SSC/LHC cross sections. Representative distribution function sets are presented in a very compact parametrized form both in the DIS and MS-bar renormalization schemes.

†This work is partially supported by the National Science Foundation under Grant No.PHY89-05161. This work was also supported by Argonne Natioanl Laboratory during the 1988 - 89 academic year when the author was on Sabbatical leave at the Laboratory from IIT.
* Permanent address.

The 1990 paper by Morfin and Tung pioneered the global QCD analysis of **p**arton **d**istribution **f**unctions (in parallel with the effort by Martin, Roberts, and Stirling in Europe)



Many common themes of ongoing PDF studies (interplay of constraints from different experiments, PDF uncertainties, predictions for “precision observables”, ...) are already present in the M-T paper



Many common themes of ongoing PDF studies (interplay of constraints from different experiments, PDF uncertainties, predictions for "precision observables", ...) are already present in the M-T paper

Parton distribution functions in 2009

Parton distribution functions $f_{a/p}(x, Q)$...

- ...are universal nonperturbative functions needed for many perturbative QCD calculations
- ... are **parametrized** as

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

- ... are found from **Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations** at $Q > Q_0$:

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

with $P_{i/j}$ known to order α_s^3 (NNLO):

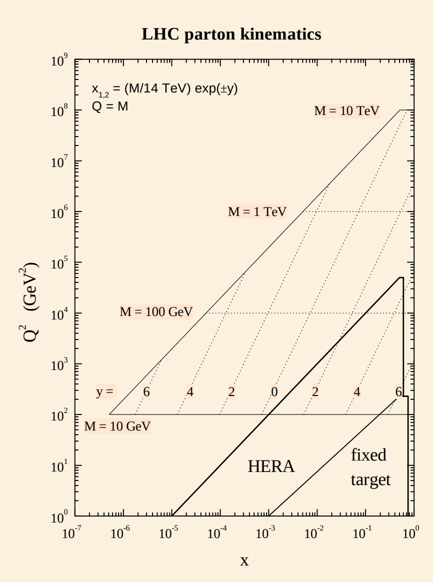
$$P_{i/j}(x, \alpha_s) = \alpha_s P_{i/j}^{(1)}(x) + \alpha_s^2 P_{i/j}^{(2)}(x) + \alpha_s^3 P_{i/j}^{(3)}(x) + ...$$

- Free parameters a_i **and their uncertainties** are determined from a global fit to hadron scattering data

Parton distributions for the Large Hadron Collider

PDF's must be determined in a wide (x, Q) range with accuracy $\sim 1\%$ for purposes of...

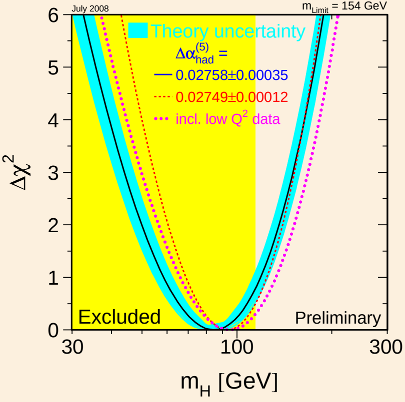
- monitoring of the LHC luminosity, calibration of detectors
- tests of electroweak symmetry breaking (EWSB)
- searches for Higgs bosons, supersymmetry, etc
- discrimination between new physics models
- precision tests of hadronic structure



Key Tevatron/LHC measurements require trustworthy PDFs

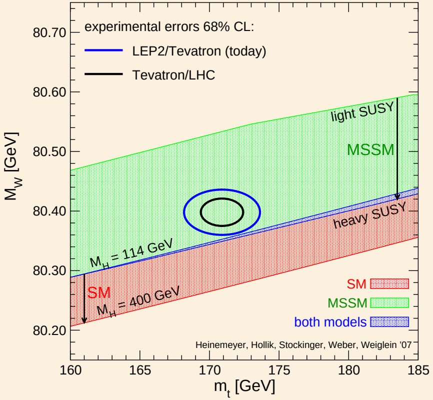
For example, leading syst. uncertainties in tests of electroweak symmetry breaking are due to insufficiently known PDFs

EW precision fits



A large part of δM_H arises from $\delta_{PDF} M_W$

EW fits + direct Higgs searches



SM band: $114 \leq M_H \leq 400$ GeV
SUSY band: random scan

Origin of differences between PDF sets

1. Corrections of wrong or outdated assumptions

lead to significant differences between new (≈post-2007) and old (≈pre-2007) PDF sets

- inclusion of (N)NLO QCD, heavy-quark hard scattering contributions
  - CTEQ6.6 and MSTW'2008 PDFs implement complete heavy-quark treatment; previous PDFs are obsolete without it
  - "NNLO" contributions are not automatically equivalent to better theory; to claim that, instabilities at small x or near heavy-quark thresholds must be also "tamed"
- relaxation of ad hoc constraints on PDF parametrizations
- improved numerical approximations

Origin of differences between PDF sets

2. PDF uncertainty

a range of allowed PDF shapes for plausible input assumptions, **partly** reflected by the PDF error band

is associated with

- the choice of fitted experiments
- experimental errors propagated into PDF's
- handling of inconsistencies between experiments
- choice of factorization scales, parametrizations for PDF's, higher-twist terms, nuclear effects,...

leads to non-negligible differences between the newest PDF sets

Nucleon PDFs: selection of experimental data

DIS-based analyses $\Rightarrow$ focus on the most precise (HERA DIS) data

- NC DIS, CC DIS, NC DIS jet, c and b production *(H1, ZEUS,* ***HERAPDF****)*
- some fixed-target DIS and Drell-Yan data, compatible with HERA DIS at $\Delta\chi^2 = 1$ level *(S. Alekhin)*

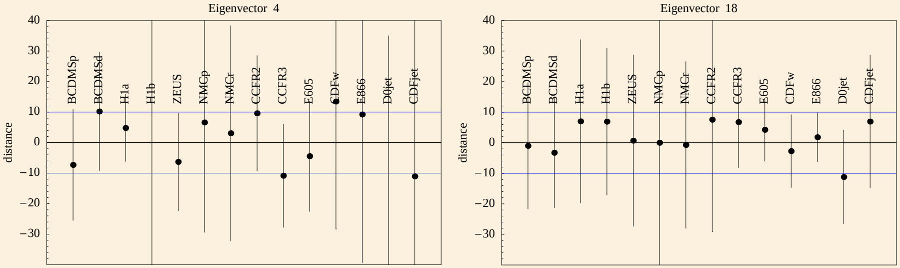
Global analyses *(****CT09****, MSTW'2008, NNPDF1.1)*
$\Rightarrow$ focus on completeness, reliable flavor decomposition

- all HERA data + fixed-target DIS data
  - notably, CCFR and NuTeV νN DIS constraining $s(x, Q)$
- low-Q Drell-Yan (E605, E866), Run-1 W lepton asymmetry, Run-2 Z rapidity *(CT09, MSTW'08, upcoming NNPDF2.0)*
- Tevatron Run-2 jet production, W asymmetry *(CT09, MSTW'08)*

Confidence intervals in global PDF analyses

CTEQ6 tolerance criterion (2001)

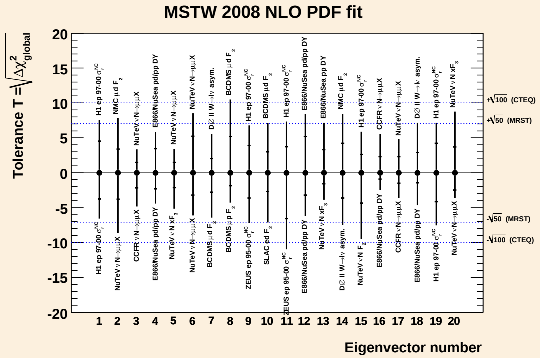
- acceptable values of PDF parameters must agree at ≈90% c.l. with all experiments included in the fit, *for a plausible range of theoretical assumptions*
- is realized by accepting all PDF fits with $\Delta\chi^2 < T^2 \approx 100$
- this criterion is modified in the new CT09 fit *(Pumplin et al., arXiv:0904.2424)*



Confidence intervals in global PDF analyses

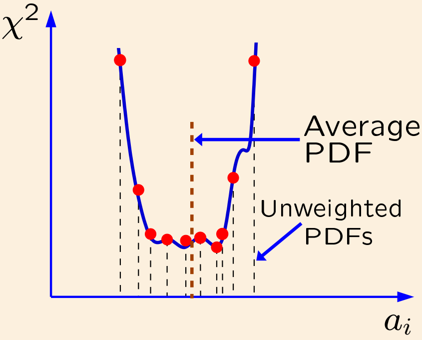
MSTW tolerance criterion (2008)

- an evolved version of the original tolerance criterion
- T^2 is calculated **independently** for each PDF eigenvector
- is close on average to $T^2 \approx 50$ *(but for which assumptions?)*



Confidence intervals in global PDF analyses

Neural Network PDF



A very general approach that

- realizes stochastic sampling of the probability distribution in PDF parameter space

(Alekhin; Giele, Keller, Kosower)

- parametrizes PDF's by flexible neural networks
- does not rely on smoothness of χ^2 or Gaussian approximations

High Precision PDFs from Combined HERA I Data and the LHC

S. Magill and R. Yoshida, ANL
ANL-IIT Theory Institute
21 May, 2009

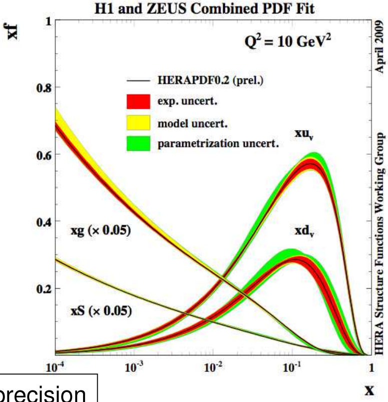
New HERA I combined data and PDF

❖ Combine the measured H1 and ZEUS cross sections. Double statistics and take advantage of complementary measurement techniques which result in reduced systematic uncertainties. No physics or model assumptions are made in the combination.

❖ The averaging method includes uncertainties related to the averaging procedure itself as well as full systematic error correlations.

❖ The combined HERA-I cross sections are used as the sole input in a QCD analysis to extract new proton PDFs:

$\Rightarrow$ HERAPDF0.2



Final HERA I PDF: low x part has final HERA precision

$0.05\ GeV^2 < Q^2 < 10^5\ GeV^2$ $10^{-6} < x < 0.65$

716 data points each from ZEUS and H1

Fit for data points (716 of them)

And j systematic uncertainties

$$\chi_e^2(\{\mu\},\{r\}) = \sum_{i=1}^{N} \left(\frac{m_i^e - \mu_i - \sum_{j=1}^{Ke} \beta_{ji}^e r_j^e}{\sigma_i^e} \right)^2 + \sum_{j=1}^{Ke} \left(r_j^e\right)^2$$

m_i^e = measured cross section in bin i by exp e

μ_i^e = true cross section in bin i

σ_i^e = statistical uncertainty in bin i by exp e

β_{ji}^e = correlated syst. unc. in bin i by exp e

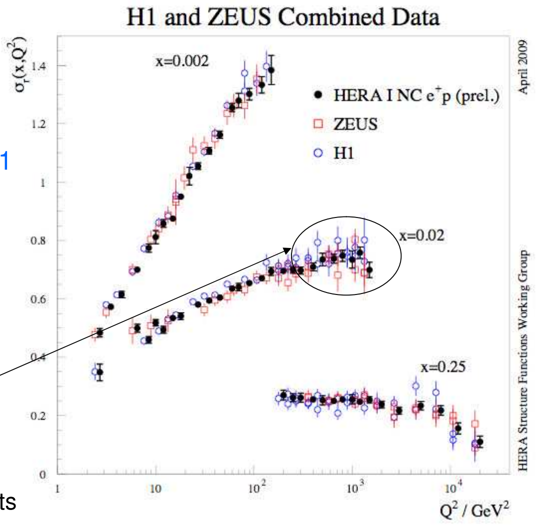
$s_i, r_j \sim N(0,1)$

(Multiplicative uncertainties are handled differently—see backup)

Averaged Cross Sections

$\chi^2/ndf = 699/716$

Sample of NC $e^{+}p$ data showing the ZEUS and H1 data and the combined data as a result of the averaging procedure



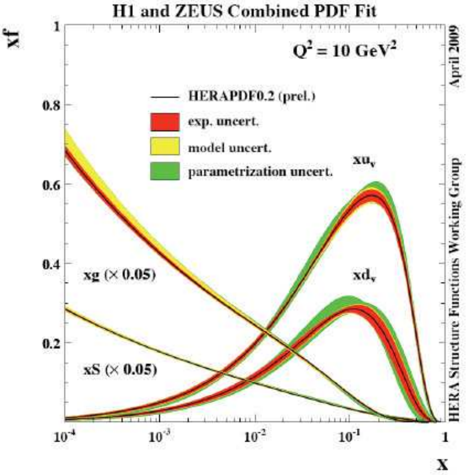
Dramatic improvement. Systematic uncertainties cancel between experiments

HERAPDF0.2

- Red: experimental uncertainties
- Yellow: model uncertainties
- Green: pdf parametrization uncertainties

Observations:

- High-x and valence are mostly affected by the PDF parametrisation
  - The procedure to estimate PDF parametrisation uncertainty addresses the high-x region
  - Low-x region interesting to investigate



QCD Analysis Model Framework

- Calculations in Heavy Flavour scheme (Thorne-Roberts Variable Flavour Number Scheme)
  - An improved theoretical treatment of heavy quarks that takes the quark masses into account
- NLO predictions using DGLAP evolution equations
  - QCDNUM17.02 (M. Botje): quicker, more accurate at high-x and can do NNLO fits
- Starting scale $Q_0^2 < M_c^2 \rightarrow Q_0^2 = 1.9$ GeV^2
  - Implies new starting sea fractions

- Differences between HERAPDF0.1 (DIS 2008) and HERAPDF0.2:

	HERAPDF0.1	HERAPDF0.2
Scheme	ZM-VFNS	TR-VFNS
Evolution	QCDNUM16.12	QCDNUM17.02
Order	NLO	NLO
Q_0^2	4 GeV^2	1.9 GeV^2
$f_s = s/D$	0.33	0.31
$f_c = c/U$	0.15	0.00
Renorm. and Fact. scales	Q^2	Q^2
Q^2_{min}	3.5 GeV^2	3.5 GeV^2
$\alpha_S(M_Z)$	0.1176	0.1176
M_c	1.4 GeV	1.4 GeV
M_b	4.75 GeV	4.75 GeV

- Fit for PDFs: $\mathrm{gluon}, u_{val}, d_{val}, \bar{U} = \bar{u} + \bar{c}, \bar{D} = \bar{d} + \bar{s} + \bar{b}$

PDF Parametrisation

- A generic functional form has been considered:

  $$xf(x,Q_0^2)=Ax^B(1-x)^C(1+Dx+Ex^2)$$

  - 6 parameters are fixed by the model assumptions
  - The optimum number of parameters are chosen by saturation of the χ^2 (i.e. only parameters that significantly contribute to χ^2 are let to vary)
- This results in 10 free parameters for the central fit (χ^2/dof=576/592)
  - All PDFs ≥ 0
  - Valence not too low compared to sea distribution at high x
  - Fit is stable with respect to the error treatment (correlations)

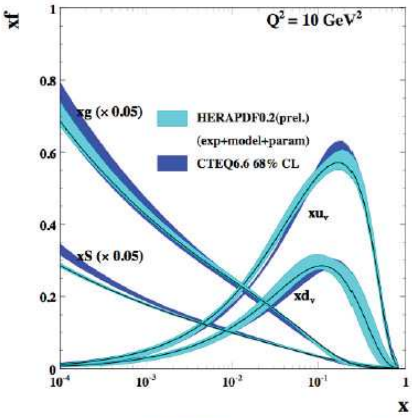
PDF	A	B	C	D	E
xg	sum rule	FIT	FIT	-	-
xu_{val}	sum rule	FIT	FIT	FIT for HERAPDF0.1	FIT
xd_{val}	sum rule	$=B_{u_{val}}$	FIT	-	-
$x\overline{U}$	$\lim_{x\to 0} \overline{U}/\overline{D} \to 1$	FIT	FIT	-	-
$x\overline{D}$	FIT	$=B_{\overline{U}}$	FIT	-	-

- Remark:
  - For HERAPDF0.1 the optimal parametrisation consisted of 11 free parameters
    - Include Du_{val}

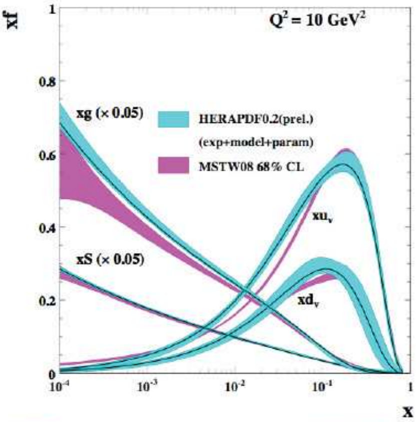
HERAPDF0.2 vs CTEQ/MSTW

- We compare HERAPDF0.2 to the global fits (at 68% CL)
  - ⇒ The new combined HERA-I data provides a strong constraint on PDFs

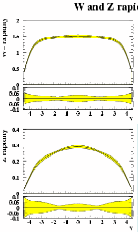
CTEQ6.6



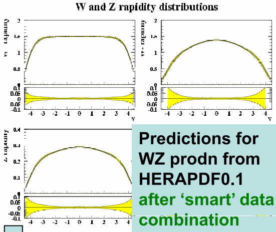
MSTW08



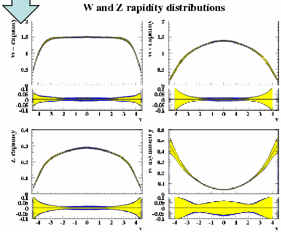
Effect on W/Z ratio at the LHC



Predictions for WZ prodn from including HERA data in PDF fit **before 'smart' data combination**



Predictions for WZ prodn from HERAPDF0.1 **after 'smart' data combination**

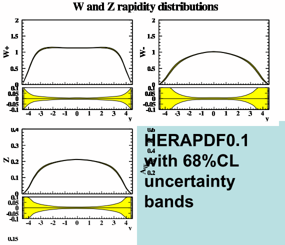


Experimental uncertainties are becoming so small that model dependence assumes greater importance- here add model errors to HERAPDF0.1 from input assumptions to PDF fitting

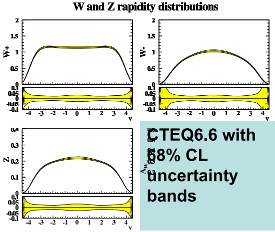
A. Cooper-Sarkar, Workshop on early LHC data, London, March 2009

Effect on W/Z ratio at the LHC

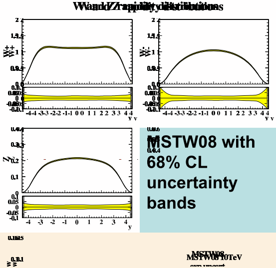
Compare HERAPDF to CTEQ6.6 and MSTW08 for W/Z predictions for 10TeV



HERAPDF0.1 with 68%CL uncertainty bands



CTEQ6.6 with 68% CL uncertainty bands



MSTW08 with 68% CL uncertainty bands

The new HERA combined data reduce the uncertainty in the central region- should be fed into CTEQ/MSTW fits

A. Cooper-Sarkar, Workshop on early LHC data, London, March 2009

HERAPDF0.2: summary

- All H1 and ZEUS NC and CC measurements from HERA I have been combined.
- A remarkable cancellation of systematic uncertainties take place between H1 and ZEUS data.
- The precision at low-x is ~1% and is very likely the ultimate precision from HERA in this kinematic region.
- A new QCD fit (HERAPDF0.2) has been made with the new combined data. It has higher precision than the previously released combined fit.
- The W cross-section prediction for LHC using the HERA PDFs will be better than 3%.
- HERA plans further combinations of their cross-section results. These include
  - HERA II polarized CC and NC cross-sections: these will impact the high x region.
  - Heavy flavor cross-sections, in particular charm.

End of HERAPDF part

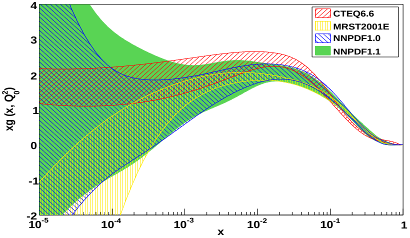
HERAPDF vs. the other PDF sets

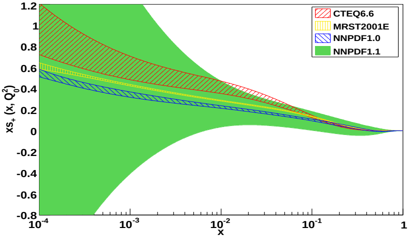
- The H1+ZEUS sample has a much smaller systematical uncertainty than the H1 and ZEUS samples individually
- Nominally, very small uncertainty compared to CTEQ-MSTW-NNPDF!
- However:
  - insufficient PDF flavor separation (neutral-current DIS probes only $4/9\,(u+\bar{u}+c+\bar{c})+1/9\,(d+\bar{d}+s+\bar{s})$)
  - very rigid PDF parametrizations $\Rightarrow$ less flexibility to probe all allowed PDF behavior, notably at small x
  - typical gluon forms, e.g., $g(x,Q_0)=Ax^B(1-x)^C\,(1+Dx)$, are ruled out by the Tevatron jet data *(Pumplin et al., arXiv:0904.2424)*

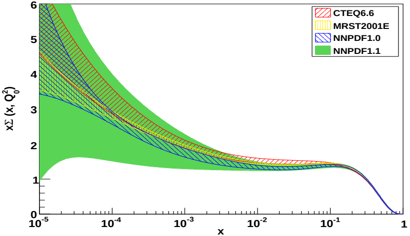
HERAPDF vs. the other PDF sets

- The H1+ZEUS sample has a much smaller systematical uncertainty than the H1 and ZEUS samples individually
- Nominally, very small uncertainty compared to CTEQ-MSTW-NNPDF!
- However:
  - insufficient PDF flavor separation (neutral-current DIS probes only $4/9\,(u+\bar{u}+c+\bar{c})+1/9\,\left(d+\bar{d}+s+\bar{s}\right)$)
  - very rigid PDF parametrizations $\Rightarrow$ less flexibility to probe all allowed PDF behavior, notably at small x
  - typical gluon forms, e.g., $g(x,Q_0)=Ax^B(1-x)^C\,(1+Dx)$, are ruled out by the Tevatron jet data *(Pumplin et al., arXiv:0904.2424)*

NNPDF1.1 vs. other PDFs at $Q = \sqrt{2}$ GeV (arXiv:0811.2288)



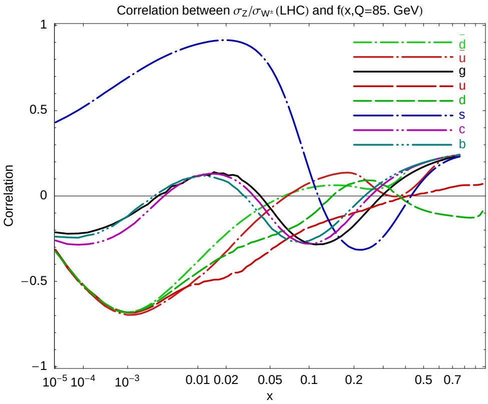




At $x \lesssim 10^{-3}$, gluon g, strangeness $s_+ = (s + \bar{s})/2$, **and** singlet $\Sigma = \sum_i (q_i + \bar{q}_i)$ PDFs are **poorly** constrained;

determined by a "theoretically motivated" functional form in CTEQ/MSTW, flexible neural net in NNPDF; g, s_+ can be < 0!

Strangeness and σ_Z/σ_W at the LHC



The PDF uncertainty in σ_Z/σ_W is mostly driven by $s(x,Q)$; increases by a factor of 3 compared to CTEQ6.1 as a result of free strangeness in CTEQ6.6

a stumbling block in the precision measurement of W boson mass M_W at the LHC

Correlation analysis for collider observables

(J. Pumplin et al., PRD 65, 014013 (2002); P.N. and Z. Sullivan, hep-ph/0110378)

A technique based on the Hessian method

For $2N$ PDF eigensets and two cross sections X and Y:

$$\Delta X = \frac{1}{2}\sqrt{\sum_{i=1}^{N}\left(X_i^{(+)} - X_i^{(-)}\right)^2}$$

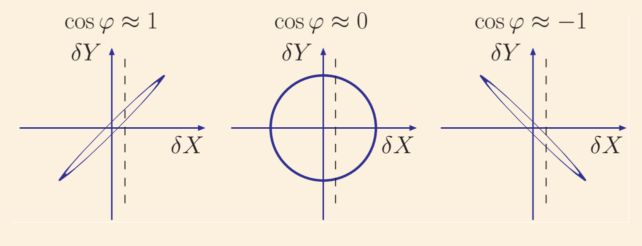
$$\cos\varphi = \frac{1}{4\Delta X\,\Delta Y}\sum_{i=1}^{N}\left(X_i^{(+)} - X_i^{(-)}\right)\left(Y_i^{(+)} - Y_i^{(-)}\right)$$

$X_i^{(\pm)}$ are maximal (minimal) values of X_i tolerated along the i-th PDF eigenvector direction; $N = 22$ for the CTEQ6.6 set

Correlation angle φ

Determines the parametric form of the $X - Y$ correlation ellipse

$$X = X_0 + \Delta X \cos\theta$$
$$Y = Y_0 + \Delta Y \cos(\theta + \varphi)$$



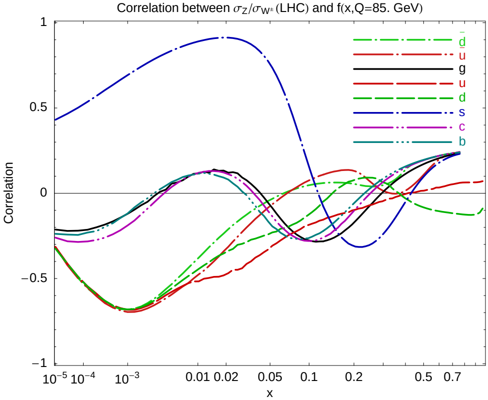
X_0, Y_0: best-fit values

$\Delta X, \Delta Y$: PDF errors

$\cos\varphi \approx \pm 1$: Measurement of X imposes tight constraints on Y

$\cos\varphi \approx 0$: Measurement of X imposes loose constraints on Y

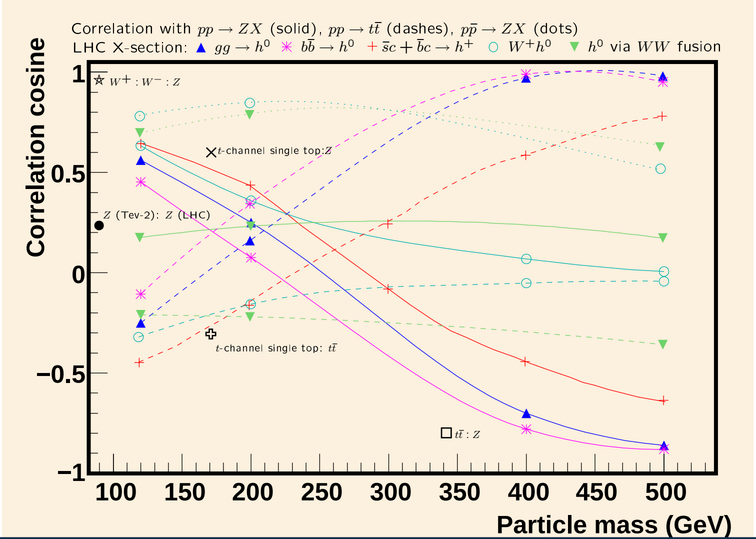
Strangeness and σ_Z/σ_W at the LHC



The PDF uncertainty in σ_Z/σ_W is mostly driven by $s(x,Q)$; increases by a factor of 3 compared to CTEQ6.1 as a result of free strangeness in CTEQ6.6

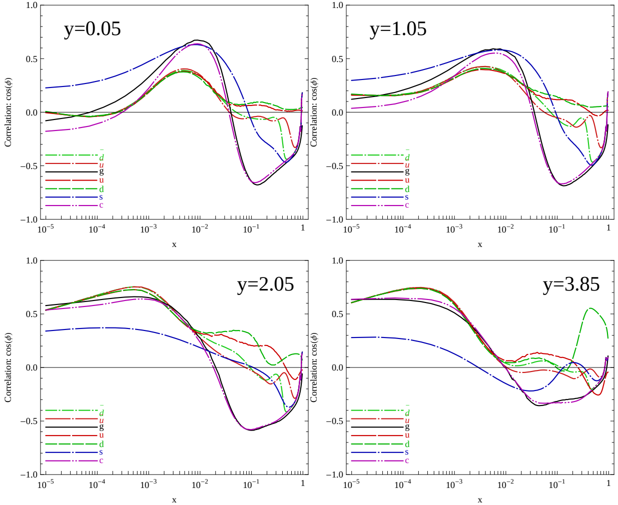
a stumbling block in the precision measurement of W boson mass M_W at the LHC

$\cos\varphi$ for various NLO Higgs production cross sections in SM and MSSM



Correlations between $d\sigma(pp \to Z^0 X)/dy$ and PDF's

$\cos\varphi$ between $d\sigma(pp \to Z^0 X)/dy$ at the LHC ($\sqrt{s} = 10$ TeV) and PDFs $f(x, Q = 85$ GeV)

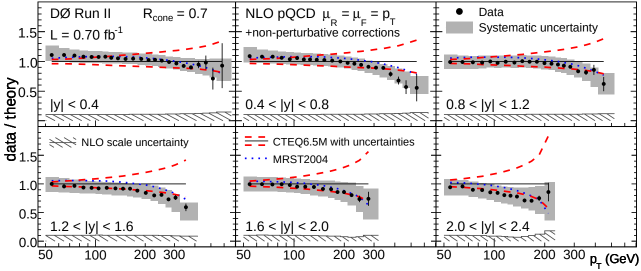


Notice the change in sensitivity to parton flavors and the shift in the most relevant x range

Toward CT09 PDF analysis

- An update of CTEQ6.6 study *(PRD 78, 013004 (2008))*
- New experimental data in the fit
  - CDF Run-2 and D0 Run-2 inclusive jet production
    - preliminarily explored in *J. Pumplin et al., arXiv:0904.0424; P.N., in preparation*
  - CDF Run-2 lepton asymmetry
  - CDF Z rapidity distribution
  - low-Q Drell-Yan p_T (E288, E605, R209) and Tevatron Run-1, Run-2 Z p_T distributions
- updated procedure for PDF error estimates

Inclusive jet production in Tevatron Run-2

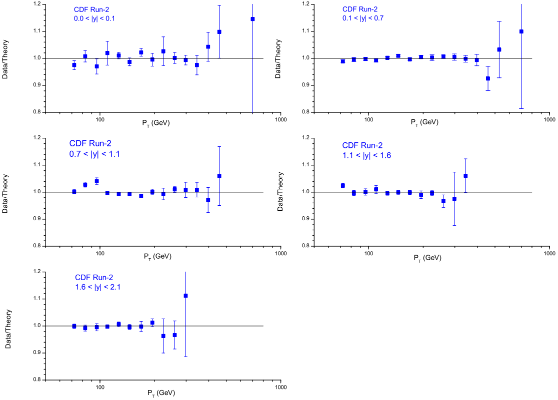


D0 Coll., arXiv:0802.2400 (700 pb^{-1}); CDF results (1.13 fb^{-1})

- (Almost) negligible statistical error
- MidCone/k_T algorithm samples, corrected to parton level
- D0 paper:
  - "There is a tendency for the data to be lower than the central CTEQ prediction..."
  - "...but they lie mostly within the CTEQ uncertainty band"
- non-negligible effect on the CTEQ gluon PDF?

Impact of Run-2 jet data on CT09 fit

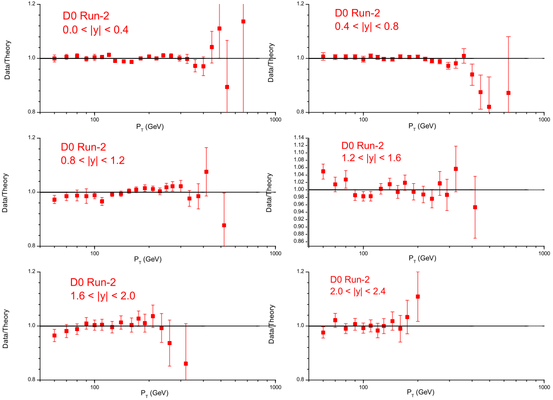
- CT09 fit includes **all four** Run-1 and Run-2 jet data samples
- Excellent quality of the fit: $\chi^2 = 2756$ for 2898 data points



(Shifted CDF-Run 2 data)/CT09

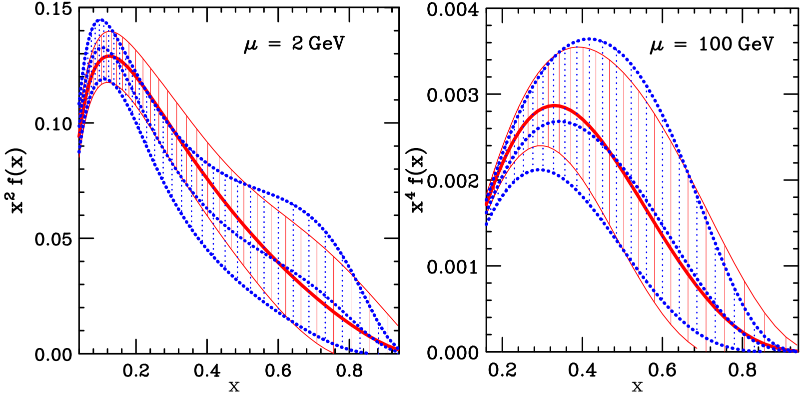
Impact of Run-2 jet data on CT09 fit

- CT09 fit includes **all four** Run-1 and Run-2 jet data samples
- Excellent quality of the fit: $\chi^2 = 2756$ for 2898 data points



(Shifted D0-Run 2 data)/CT09

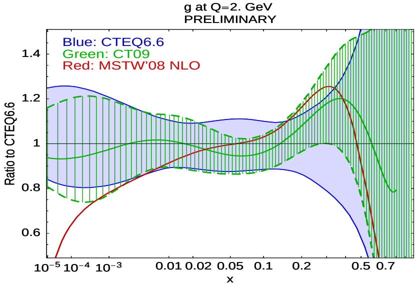
CT09 and CTEQ6.6 are generally compatible

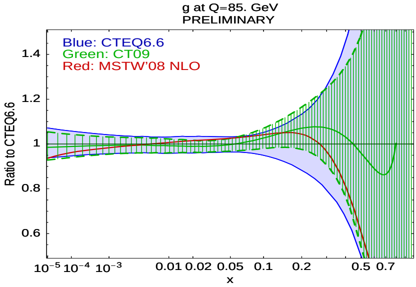


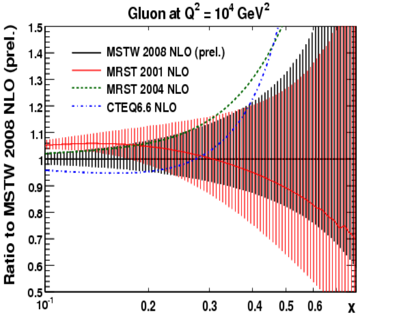
Gluon PDF: CT09 (red), CT66 (blue)

CT09 PDF uncertainty is about the same as CT66 (compensation between the Run-2 jet constraints and more flexible $g(x, \mu)$)

CT09 gluon vs. CT66 and MSTW'08 NLO







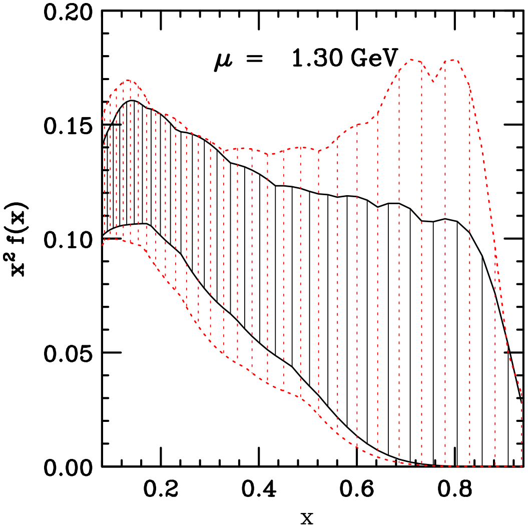
- Run II jet data prefer smaller gluon distribution at high x.

G. Watt, 2008

⇑

Not really supported by CT09 analysis

Constraints of Run-2 data on CT09 PDFs



Black band: $g(x, \mu)$ from CT09 fit

Red band: The same fit **without** Run-2 jet data

This band is wider than CT66 because of additional free parameters

Run-2 data impose tangible constraints on the allowed range of $g(x, \mu)$

Self-consistency of CT09 fit

Are there tensions in the fit?

1. Are the Run-2 jet data consistent with theory?
   1.1 Are the PDF parametrizations too flexible/too rigid?
2. Are the new data consistent with other experiments?
3. Are the new data consistent with one another?

J. Pumplin et al. (arXiv:0904.2424) find that

- individual data sets, and data and theory are generally consistent with one another
- abnormalities exist in the agreement of D0 Run-1 set with other data sets

Correlated systematic errors (CSE) in jet production
P. Nadolsky, in preparation

CSE for inclusive jets are important. PDF errors are underestimated without them. CTEQ takes them into account since 2000. CSE are provided in two forms:

1. $N_{pt} \times N_\lambda$ correlation matrix $\beta_{k\alpha}$ for N_λ random systematic parameters λ_α

$$\chi^2 = \sum_{e=\{\text{expt.}\}} \left[\sum_{k=1}^{N_{pt}} \frac{1}{s_k^2} \left(D_k - T_k - \sum_{\alpha=1}^{N_\lambda} \lambda_\alpha \beta_{k\alpha} \right)^2 + \sum_{\alpha=1}^{N_\lambda} \lambda_\alpha^2 \right]$$

D_k are T_k are data and theory

s_k is the stat.+syst. uncorrelated error

2. $N_{pt} \times N_{pt}$ covariance matrix $C = I + \beta\beta^T$

$$\chi^2 = \sum_{e=\{\text{expt.}\}} (D-T)^T C^{-1} (D-T)$$

Comparison of CSE's for four jet experiments

- β (used by CDF Run-1 and 2, D0 Run-2) has several practical advantages compared to C (used by D0 Run-1)

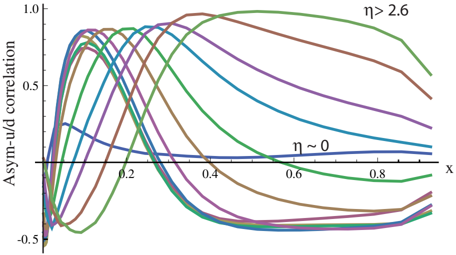
- Plausibility of β can be checked by the principal component analysis (PCA) of β
  - Typically only $\approx N_\lambda/2$ combinations of λ_α (found by PCA) are relevant for χ^2; $\text{rank}\left[\beta\beta^T\right] \approx N_\lambda/2 \ll N_{pt}$

- C is a large ($N_{pt} \times N_{pt}$) matrix provided as a "black box"; plausibility of C is harder to verify. C provided by D0 Run-1 has irregularities revealed by PCA
  - $\text{rank}\,[C - I] = \text{rank}\left[\beta\beta^T\right] \approx N_{pt} = 90$ – too large

- This suggests that D0 Run-1 CSE's are overestimated; may explain consistently small $\chi^2_{\text{D0 Run-1}}/N_{pt} \sim 0.3$ in fits, other peculiarities of D0 Run-1 data observed in the weight scan

CDF and D0 Run-2 W asymmetry $A_\ell(y)$

New CDF and D0 Run-2 W lepton asymmetry (in bins of electron p_{Te} and η_e) ; probes u/d in a range of large x values



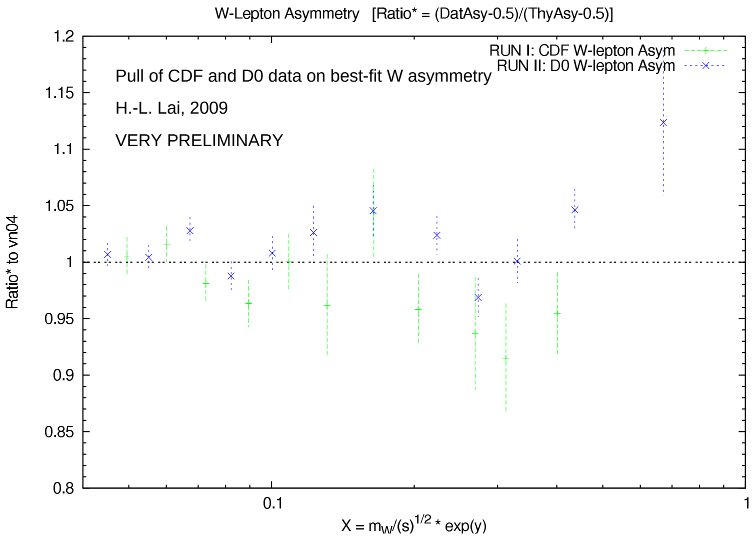
Correlation of $A_\ell(y)$ in different η_e bins ($p_{Te} > 35$ GeV) with $u(x)/d(x)$ *(H. Schellman)*

We find that CDF and D0 $A_\ell(y)$ data disagree in a similar kinematical range (confirming a similar MSTW finding)

CDF Run-2 $A_\ell(y)$ agrees ok with the other data

CT09 includes only CDF Run-2 $A_\ell(y)$

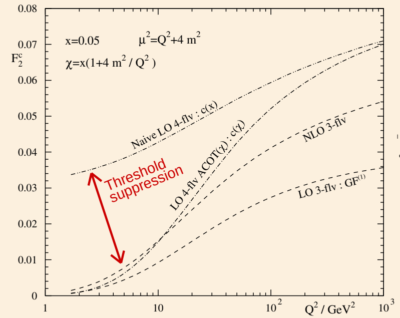
A preliminary fit to CDF and D0 $A_\ell(y)$

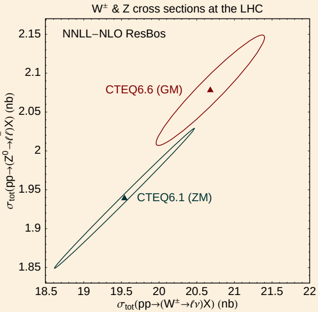


Role of heavy flavors in PDF analyses

General-mass (GM) scheme(s), currently adopted by CTEQ, MSTW and HERA analyses, strive to provide consistent description of c, b scattering both near heavy-quark thresholds and away from them

In 2006 *(CTEQ6.5, hep-ph/0611254)*, it was realized that the GM (and not zero-mass) treatment of c, b mass terms in DIS is essential for predicting precision W, Z, cross sections at the LHC

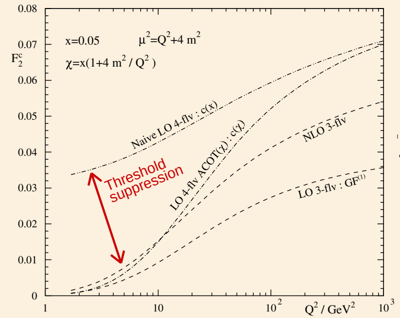


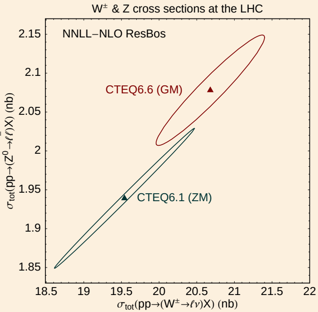


Role of heavy flavors in PDF analyses

At NLO, the dominant mass effect is mostly kinematical; propagates from DIS into W, Z cross sections through changes in $u(x)$, $d(x)$

Thorne and Tung (arXiv:0809.0714): can this kinematical effect be **approximately** introduced in the **widely used** ZM scheme, while preserving ZM hard cross sections?





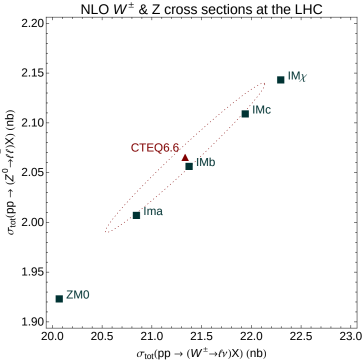
Kinematically improved ZM schemes

At NLO, such schemes were indeed developed (P. N., Tung, arXiv:0903:2667)

They depend on a free parameter λ, tuned to approximate either a ZM DIS cross section or a GM DIS cross section

They can be viewed either

- as improved ZM formulations with realistic c, b kinematics; or
- as simplified GM formulations with approximate ZM hard cross sections



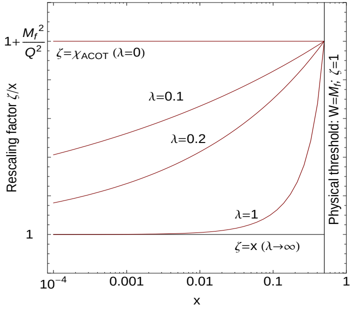
We propose to call them "**intermediate-mass** (IM) schemes"

Generalized rescaling variable ζ

Realistic GM behavior of PDF's is reproduced by a **generalized rescaling variable** $\zeta(\lambda)$:

$$x = \zeta \Big/ \left(1 + \zeta^{\lambda} M_f^2/Q^2\right), \text{ with } 0 \leq \lambda \lesssim 1$$

- $\zeta \to 1$ as $W \to M_f^2$
- $\lambda = 0 : \zeta \equiv \chi = x\left(1 + M_f^2/Q^2\right)$ – the ACOT-χ variable
- $\lambda \gtrsim 1$: $\zeta \approx x$ (no rescaling)
- $\zeta \approx x$ for $Q^2 \gg M_f^2$



ZM fits with $\lambda \approx 0.15$ closely reproduce GM results

PDF reweighting in Monte-Carlo integration

If $X_i^{(\pm)}$ and $\Delta X^2 = \sum_{i=1}^{N} \left(X_i^{(+)} - X_i^{(-)} \right)^2 / 4$ are computed in $2N = 44$ independent Monte-Carlo runs with $\bar{N}$ events each, their resulting estimates are given by

$$\overline{X}_i^{(\pm)} = X_i^{(\pm)} + \overline{\delta}_i^{(\pm)} \sim X_i^{(\pm)} + \frac{c}{\overline{N}^{1/2}} \text{ and}$$

$$\overline{\Delta X}^2 = \frac{1}{4} \sum_{i=1}^{N} \left(\overline{X}_i^{(+)} - \overline{X}_i^{(-)} \right)^2 \sim \Delta X^2 + \frac{c' N}{\overline{N}^{1/2}}$$

$\overline{\delta}_i^{(\pm)}$ is a **random** MC error dependent on the input PDF, arising, e.g., from importance sampling

As a result of the PDF dependence of $\overline{\delta}_i^{(\pm)}$, the error $\overline{\Delta X}^2 - \Delta X^2$ is increased by a factor $N \sim 22$

PDF reweighting in Monte-Carlo integration

- PDF reweighting generates the same sequence of events to compute each of $2N$ cross sections
  - all $\overline{\delta}_i^{(\pm)}$ are the same
  - $\overline{\Delta X}^2 = \Delta X^2$
- In multi-loop calculations, PDF reweighting saves CPU time drastically by reducing slow computations of hard-scattering matrix elements

FROOT: a simple interface for Monte-Carlo PDF reweighting

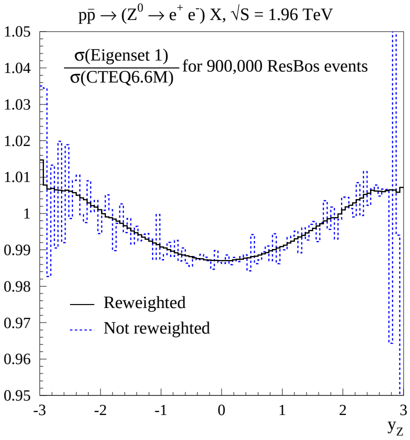
- Written in C, can be linked to standalone FORTRAN/C/C++ programs
- Simple – 170 lines of the code
- Writes the output directly into a ROOT ntuple; no need in intermediate PAW ntuples
- Flexible; new columns (branches) with PDF weights or events can be added into an existing ntuple
- Kinematical cuts, selection conditions can be imposed a posteriori in interactive or batch ROOT sessions
- implemented in MCFM, ResBos; additional libraries for ROOT analysis of reweighted ntuples are on the way

FROOT: a simple interface for Monte-Carlo PDF reweighting

```
// These are the C functions accessible from Fortran.


extern "C" {
  //Initialization of the ROOT file
  void initrootnt_(const char *title, const char *access, int ltitle, int laccess);
  void reinitrootnt_(const char *access, int laccess);
  void addntbranch_(float *element, const char *chtag, int ltag);
  void fillntbranch_(const char *chtag, int ltag);
  int getnumbranches_();
  void rootntoutp_();
  void printnt_();
  void teststr_(const char *str, int lstr);

}//extern "C"
```

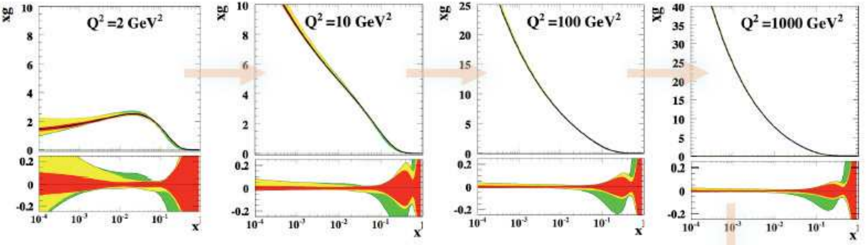


Other ongoing work

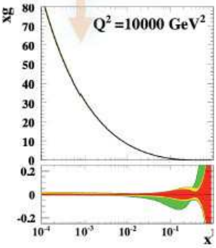
- Combined fit of PDF's and Drell-Yan p_T distributions
- PDF's for leading-order Monte-Carlo programs
- consistency and implementation of heavy-flavor contributions in the global fit at NNLO
- constraints on new physics (strongly interacting superpartners, etc.)
- public C++/ROOT libraries for PDF reweighting for (N)/NLO calculations and efficient error analysis

Backup: Rik's slides

Gluon Evolution

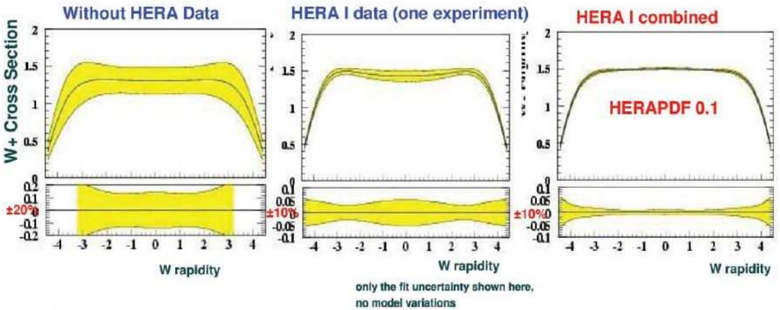


- Near the starting scale gluon is valence like
  - The model uncertainties are large in low x region
    - Mostly due to Q_0^2 variations
  - The PDF param. uncertainty dominates high x



- Impressive precision at higher Q^2

Parton densities from combined data



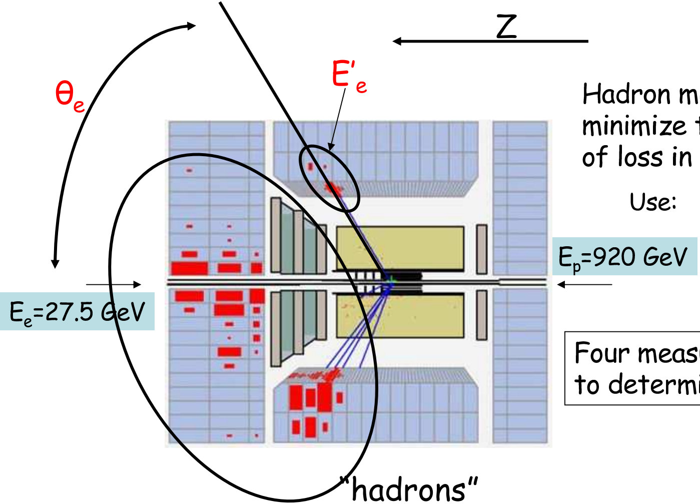
A test on a standard candle process at LHC

HERA PDFs 0.1 available in LHAPDF

HERAPDF0.2 has factor of 2 smaller uncertainty than 0.1 (more data) at low x
Available soon in LHAPDF

How do the systematics cancel??

Systematics correlated across the kinematic plane, but uncorrelated between experiments cancel.



Hadron measurement: minimize the influence of loss in the beampipe

Use: $\Sigma(E-P_z)_h$

ΣP_{Th}

Four measurements to determine x and Q^2

Additional kinematic constraints:

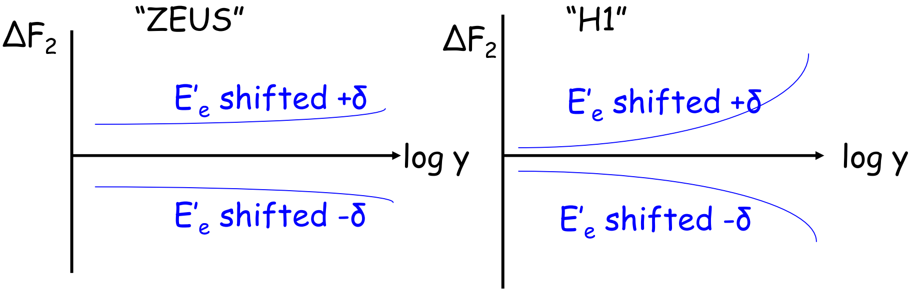
$$\sum(E-P_z)_h + (E-P_z)_e = 2 \cdot Ee = 55\ GeV$$

$$\sum P_{Th} = P_{Te}$$

HERA detectors over-constrain the kinematics.
How ZEUS and H1 actually reconstruct x and Q^2 depend on the particular characteristics of each detector.
The procedures are, in practice, quite different, and relatively complicated. Roughly speaking:

H1: θ_e and E'_e at high y, and use $\sum(E-P_z)_h$ at low y.
ZEUS: θ_e and θ_h { $\tan(\theta_h/2) = \sum(E-P_z)_h / \sum P_{Th}$ }

This is what happens schematically



- ZEUS and H1 have similarly sized uncertainties.
- ZEUS and H1 have differently "shaped" uncertainty correlations.
- ZEUS and H1 have different best measured regions.

➔ You win big from the fit

Averaging Procedure

- Swim all points to a common x-Q^2 grid
- Moved 820 GeV data to 920 GeV p-beam energy
  (except for data points with y>0.35)
- Calculate average values and uncertainties
- Evaluate "procedural uncertainties

Additive error sources:

$$\chi^2_{\exp}(\boldsymbol{m}, \boldsymbol{b}) = \sum_i \frac{\left[m^i - \sum_j \Gamma^i_j b_j - \mu^i\right]^2}{\Delta_i^2} + \sum_j b_j^2.$$

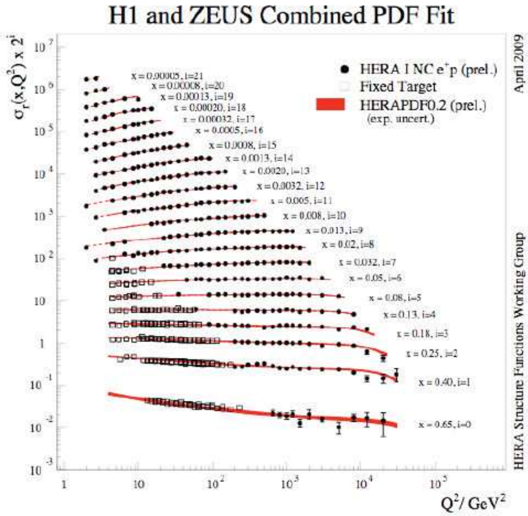
For multiplicative error sources small biases to lower cross section values may occur - avoided by modifying the χ^2 definition as follows:

$$\chi^2_{\exp}(\boldsymbol{m}, \boldsymbol{b}) = \sum_i \frac{\left[m^i - \sum_j \gamma^i_j m^i b_j - \mu^i\right]^2}{\delta^2_{i,\mathrm{stat}}\left(m^i - \sum_j \gamma^i_j m^i b_j\right) + \left(\delta_{i,\mathrm{uncor}}\, m^i\right)^2} + \sum_j b_j^2.$$

$$\gamma^i_j = \Gamma^i_j/\mu^i \qquad \delta_{i,\mathrm{stat}} = \Delta_{i,\mathrm{stat}}/\mu^i \qquad \delta_{i,\mathrm{uncor}} = \Delta_{i,\mathrm{uncor}}/\mu^i$$

Fit Results on Cross Sections

- (χ^2 /dof = 576/592)
- Plots show the extended kinematic range of the HERA data as compared to the fixed target measurements
- Plots include experimental uncertainties on both data and fit



Procedural Uncertainties

Three procedural uncertainties are introduced:

1. Additive vs Multiplicative nature of the error sources
   (Typically below 0.5%)
2. Correlated systematic unc. for the photoproduction background
   (Few % only at high-y)
3. Correlated systematic unc. for the hadronic energy scale
   (At the % level)

In fact, a more general study of the possible correlated systematic uncertainties between H1 and ZEUS has been performed:

- Identified 12 possible uncertainties of common origin
- Compare 2^{12} averages, taking all pairs as corr./uncorr. in turn

Mostly negligible except for photoproduction and hadronic energy scale

Data Sets

HERAPDF0.2 includes complete HERA-I inclusive NC and CC DIS data:

⇒ Ep=820 ($\sqrt{s}$=300) and Ep=920 ($\sqrt{s}$=320) GeV, L=240 pb^{-1}

⇒ 1% precision for combined data in Q^2 = 10-100 GeV^2 region

HERAPDF0.1 includes :

- CC e^-p data: H1 98, ZEUS 98
- CC e^+p data: H1 94-97, H1 99-00, ZEUS 94-97, ZEUS 99-00
- NC e^-p data: H1 98, ZEUS 98
- NC e^+p data: ZEUS 96-97, ZEUS 99-00, H1 99-00 "high Q^2"

New data sets:

- H1 95-00 "low Q^2" $0.2 \leq Q^2 \leq 12$ GeV^2
- H1 96-00 "bulk" $12 \leq Q^2 \leq 150$ GeV^2
- ZEUS BPC/BPT, SVX95 ($0.045 \leq Q^2 \leq 17$ GeV^2)

110 correlated systematic error sources

3 "procedural uncertainties" related to the averaging procedure

Model Uncertainties

- Variation of the heavy quark thresholds:
  - Mc = 1.4 GeV → 1.35 - 1.50 GeV
    - varied with Q^2_0 (1.77 - 2.19) GeV^2
  - Mb = 4.75 GeV → 4.30 - 5.00 GeV

- Variation of the sea fractions:
  - f_s = s/D = 0.31 → 0.23 - 0.38
  - f_c = c/U = 0.00 → specified by TR-VFNS

- Variation of the starting scale of evolution of PDFs:
  - Q^2_0= 1.9 GeV^2 → 1.5 - 2.5 GeV^2:
    - for Q^2_0= 2.5 GeV^2 vary f_s=0.32 and Mc=1.6 GeV because $Q^2_0 < Mc^2$
    - for Q^2_0= 1.5 GeV^2 vary f_s=0.29

- Variation of the minimum Q^2 cut on data:
  - Q^2_{min}= 3.5 GeV^2 → 2.5 - 5.0 GeV^2

PDF Parametrisation Uncertainties

PDFs are parametrised using the following general functional form:

$$xf(x,Q_0^2) = Ax^B(1-x)^C(1+Dx+Ex^2)$$

- The optimization procedure leads to the choice of PDF parametrisation with 10 parameter with χ^2 /dof = 576/592
  - Similar optimization procedure used for H12009 PDF

PDF	A	B	C	D	E
xg	sum rule	FIT	FIT	-	-
xu_{val}	sum rule	FIT	FIT	-	FIT
xd_{val}	sum rule	$=B_{u_{val}}$	FIT	-	-
$x\overline{U}$	$\lim_{x\to 0} \overline{U}/\overline{D} \to 1$	FIT	FIT	-	-
$x\overline{D}$	FIT	$=B_{\overline{U}}$	FIT	-	-

Current strategy to determine PDF parametrisation uncertainty is to test alternative parametrisations with similar or better χ^2 which have been discarded due to additional optimisation requirements:

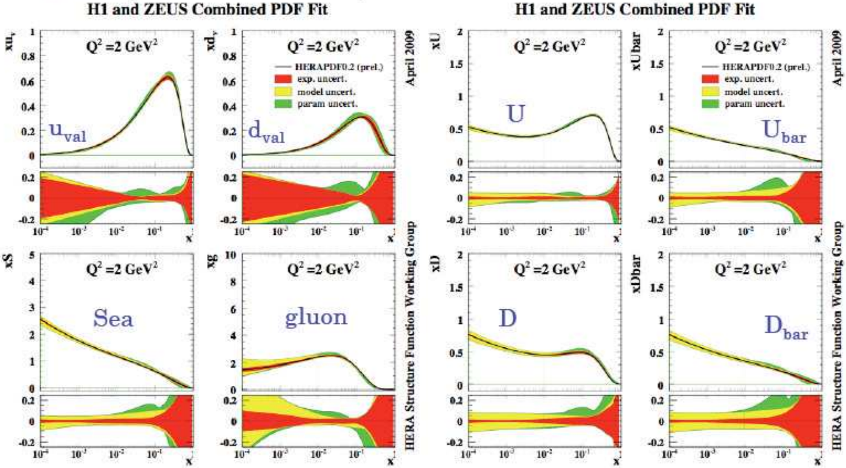
- Reasonable shape for valence and sea distributions at high-x
- All PDFs >0
- Stability in error treatment (correlated vs uncorrelated)

- Envelope of all these fits is formed and used as PDF parametrisation error
  - 7 fits out of all possible 11 parameter fits obtained by adding one additional parameter to the central fit parametrisation choice were used for the envelope

Note: the procedure addresses the high-x region

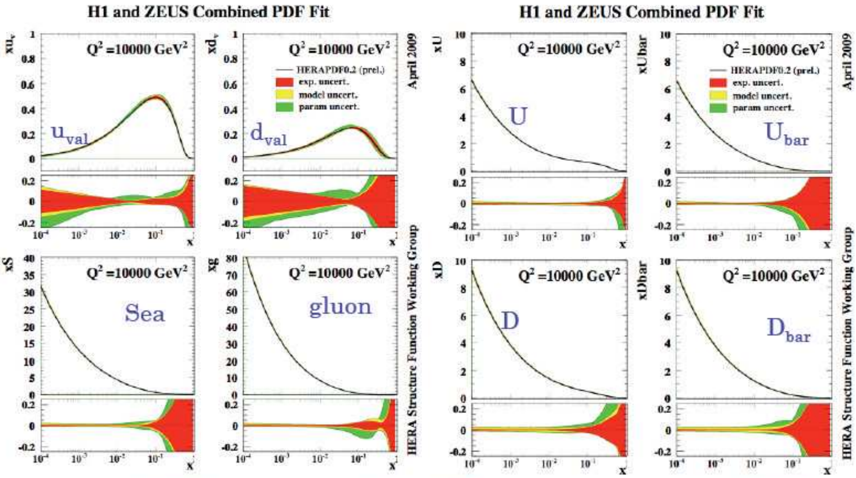
HERAPDF0.2 at Q^2=2 GeV^2

- At the starting scale gluon is valence like
- Q_0^2, Q^2_{min} dominate the model uncertainty of gluon and valence PDFs
- PDF parametrisation uncertainty dominates valence PDFs and high x region



HERAPDF0.2 at Q^2=10000 GeV^2

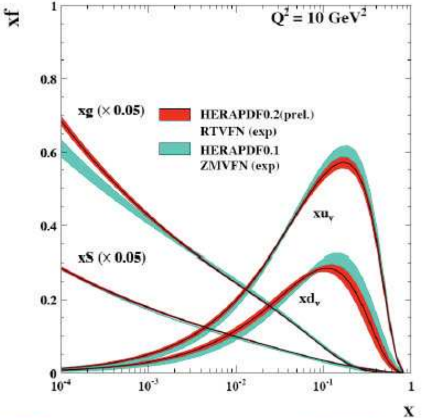
- PDF parametrisation uncertainty dominates valence PDFs and high x region
- Impressive precision at the scale relevant to LHC



HERAPDF0.1 vs HERAPDF0.2

- For consistency, when comparing the HERA PDF sets only the experimental errors are used:
  - The model uncertainties of the two PDF sets are not identical
  - HERAPDF0.1 did not consider the uncertainty due to PDF parametrisation

- Observations:
  - Errors are smaller for HERAPDF0.2
  - d_{val} is softer
  - Gluon is steeper:
    - This is expected due to the heavy flavour treatment
      HERAPDF0.1 - massless quarks (ZM-VFNS)
      HERAPDF0.2 - massive quarks (TR-VFNS)



Backup: Jets

Comparison of NLO theoretical calculations

- NLO theoretical uncertainties are at the level 10-20% *(D. Soper)*
- **NLO** inclusive jet **cross sections** are currently available from (at least) **two groups**:
  - Ellis-Kunszt-Soper - in CTQ6.6 and our earlier fits
  - NLOJet++ *(Nagy)* + FastNLO *(Kluge, Rabbertz, Wobisch)* - in CT09 and MSTW'08
- Jon P. explored
  - agreement between EKS and FastNLO
  - dependence on the choice of scale, jet algorithms, and partial threshold resummation corrections
- **The overall agreement/stability at NLO is satisfactory, although not perfect**

CT09 uses FastNLO for $\mu = p_T/2$ without the threshold resummation correction

Comparison of K=NLO/LO from EKS and FastNLO

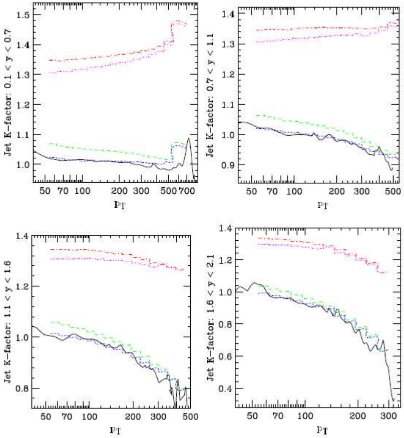


Figure 1: Theory calculations for the ratio K = NLO/LO from FastNLO and EKS. FastNLO with $\mu = p_T$: $R_{sep} = 2.0$ (long dash dot), $R_{sep} = 1.3$ (short dash dot); FastNLO with $\mu = p_T/2$: $R_{sep} = 2.0$ (long dash), $R_{sep} = 1.3$ (short dash); EKS with $\mu = p_T/2$, $R_{sep} = 1.3$ (solid).

Scale dependence of NLO cross section

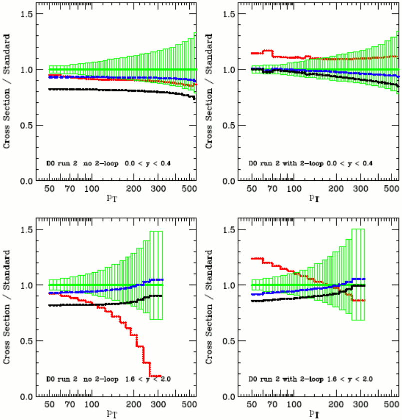


Figure 2: Effect of scale choice on predicted cross section with $R_{\rm sep} = 1.3$: $\mu = 2\,p_T$ (short dash), p_T (long dash), $p_T/2$ (solid), $p_T/4$ (dotted), relative to our Standard Choice ($\mu = p_T/2$, $R_{\rm sep} = 1.3$, no "two-loop" correction). Right panels include the "two-loop" resummation correction. Uncertainty bands from PDFs are shown for comparison.

χ^2 weighting scans

All questions are explored using the χ^2 reweighting technique
(Collins, Pumplin, hep-ph/0105207)

$$\chi^2 = \sum_{\text{jet expts.}} w_i \chi_i^2 + \chi^2_{\text{non-jet}} = w\chi^2_{\text{jet}} + \chi^2_{\text{non-jet}}$$

$w_i = 0$: experiment i is not included

$w_i = 1$: common choice

$w_i \gg w_{j \neq i}$: only experiment i matters in the fit

Self-consistency of CT09 fit

CDF_I (33 pts)		$D0_I$ (90 pts)		CDF_{II} (72 pts)		$D0_{II}$ (110 pts)		$\Delta\chi^2$
Wt	χ^2	Wt	χ^2	Wt	χ^2	Wt	χ^2	non-jet
0	55.4	0	115.3	0	99.5	0	134.0	0.0
1	52.6	1	47.0	0	105.6	0	138.3	11.8
0	56.6	0	82.2	1	85.6	1	124.1	6.2
1	52.1	1	59.4	1	88.5	1	121.5	9.6
0	58.4	0	60.9	10	79.6	10	120.4	39.9
1	54.8	1	58.8	10	80.3	10	120.0	39.4
10	54.1	10	35.6	0	112.9	0	156.7	24.1
10	53.1	10	38.6	1	102.6	1	142.3	21.9
10	51.6	10	49.7	10	82.8	10	120.9	39.6
10	49.5	0	73.5	0	110.4	0	125.3	12.5
50	47.3	0	74.0	0	123.9	0	139.3	80.5
0	58.6	10	32.1	0	122.7	0	172.2	25.2
0	66.8	50	30.6	0	140.0	0	189.1	58.6
1	59.6	1	67.5	10	75.2	1	130.9	32.0
1	63.4	1	70.4	50	71.6	1	140.0	92.9
1	50.6	1	60.0	1	93.0	10	116.5	20.6
1	50.5	1	61.6	1	96.6	50	112.6	113.8

Table 1: χ^2 for jet experiments with various weights

- Individual data sets, and data and theory are generally consistent with one another
- abnormalities in the agreement of D0 Run-1 set with other data sets

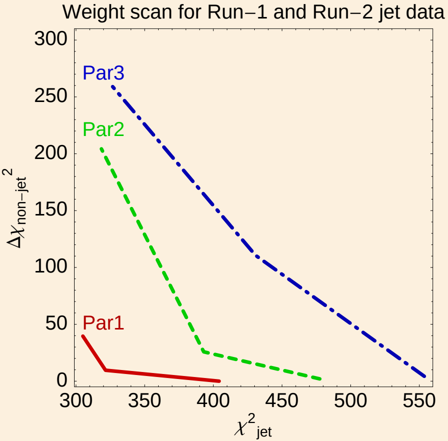
Dependence on the gluon PDF parametrization

CT09 uses a more flexible $g(x, \mu_0)$ ("par 1") than CT66

- Par 1: $g(x, \mu_0) = A_0 x^{A_1} (1-x)^{A_2} \times e^{A_3 x + A_4 x^2 + A_5 x^{1/2}}$
- CT66: par 1 with $A_2 = 4$, $A_5 = 0$
- Par 2: $A_4 = A_5 = 0$
- Par 3 (H1-like): $g(x, \mu_0) = A_0 x^{A_1} (1-x)^{A_2} (1 + A_3 x)$

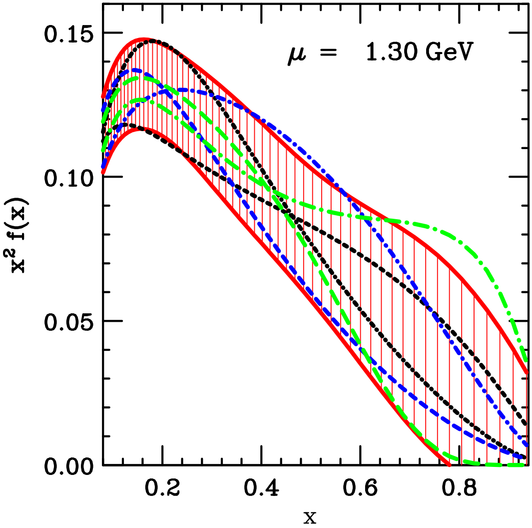
CT09 (par1) form provides the best χ^2 and vanishing tension with the non-jet data

Par 2 and 3 are disfavored



$w_{jet} = 0$ (right ends); 1 (bend); 10 (left end)

Lagrange multiplier method vs. Hessian method



The Hessian method (48 error PDFs — red band) underestimates the true $\delta_{PDF} g(x, Q)$ suggested by χ^2 (revealed by the LM method — individual lines)

Backup: IM scheme

Intermediate-mass scheme: a basic recipe

- Start with PQCD factorization for DIS, in a form applicable both in ZM or GM schemes

$$F_\lambda(x,Q^2) = \sum_{a,b}\int_\zeta^1 \frac{d\xi}{\xi}\, f_a(\xi,\mu)\, C^a_{b,\lambda}\left(\frac{\zeta}{\xi},\frac{Q}{\mu},\frac{m_i}{\mu},\alpha_s(\mu)\right)$$

- sum over **initial-state** active flavors a **prescribed by the factorization scheme for the PDFs**
- sum over **final-state** quark flavors b **physically produced at the given scattering energy**
- evaluate convolutions over the kinematical range $\zeta \le \xi \le 1$ determined by a **rescaling variable** ζ
- use **zero-mass** Wilson coefficients $C^a_{b,\lambda}$=$C^a_{b,\lambda}\left(\frac{\zeta}{\xi},\frac{Q}{\mu},0,\alpha_s(\mu)\right)$, evaluated at ζ/ξ
- keep $\mu > m_Q$ in heavy-quark channels for all Q, to guarantee applicability of the (subtracted) $\overline{MS}$ expression for $C^a_{b,\lambda}$; e.g., use $\mu = \sqrt{Q^2+m_i^2}$ in $q_i\gamma^* \to q_i$ and $g\gamma^* \to q_i\bar{q}_i$ channels

Rescaling variables in heavy-flavor DIS

Rescaled light-cone variable ζ is a simple way to approximate exact scattering kinematics in processes where the exact momentum conservation is absent

- **General-mass scheme:** dominant heavy-quark mass effects are approximated in LO $c\gamma^* \to c$ (or $sW \to c$) by a **rescaling variable** χ :

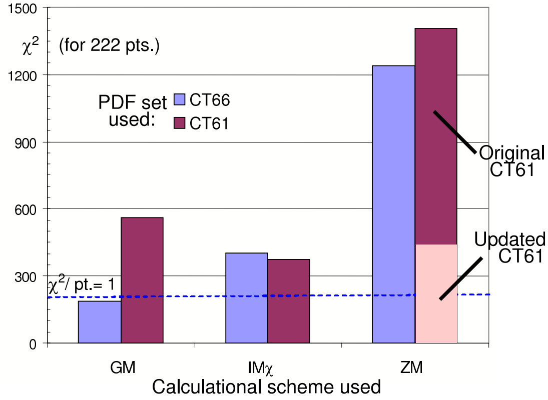
$$x = \chi / \left(1 + M_f^2/Q^2\right), \text{ with } M_f^2 = 4m_c^2\,(m_c^2) \text{ in NC (CC) DIS}$$

Barnett, Haber, Soper; Tung, Kretzer, Schmidt

- It is natural to try $\zeta = \chi$ both in $c\gamma^* \to c$ and $g\gamma^* \to c\bar{c}$ in the IM scheme; however, it leads to **excessive suppression** of charm scattering in NC DIS at small x (for given Q), where threshold effects should be less pronounced, while the PDF variation is rapid

Comparison to c, b SIDIS data from CT6.6 data sample

ZM/GM/IM Wilson coefficients with ZM (CT6.1M) and GM (CT6.6M) PDF's

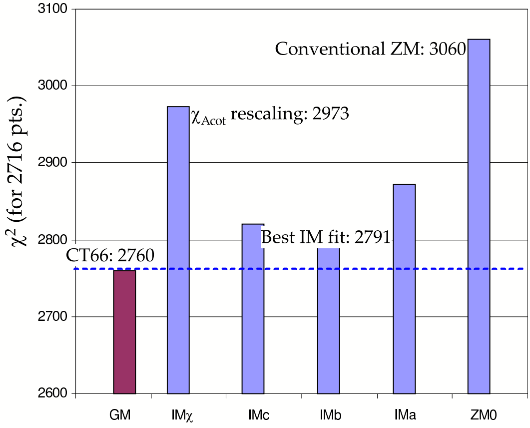


- CT6.1 are refitted including the latest HERA c, b data (χ^2 improved)
- GM+CT6.6 gives the best fit; ZM+CT6.6 the worst
- χ^2 for IM+CT6.6 (IM+CT6.1) is better than ZM+CT6.6 (ZM+6.1)

All IMλ+CT6.x fits produce close χ^2 – only IMχ is shown

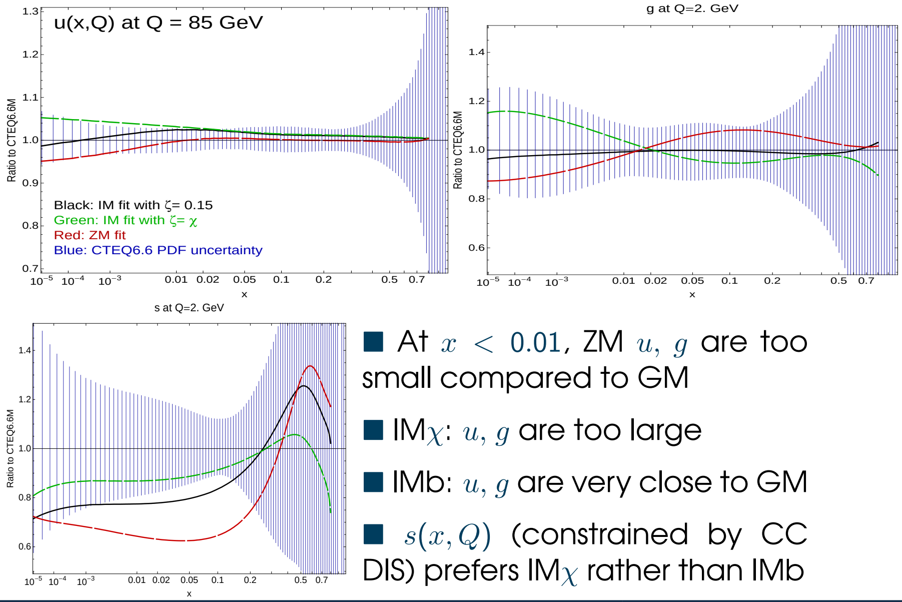
IM formulation works!

Comparison to the full CT6.6 data sample



- GM+CT6.6 still gives the best fit; ZM+(new CT6.1) the worst
- Quality of the best IM fit (IMb) approaches that of GM+CT6.6
- Some IM fits ($\lambda = 0.1 - 0.2$) are clearly very good

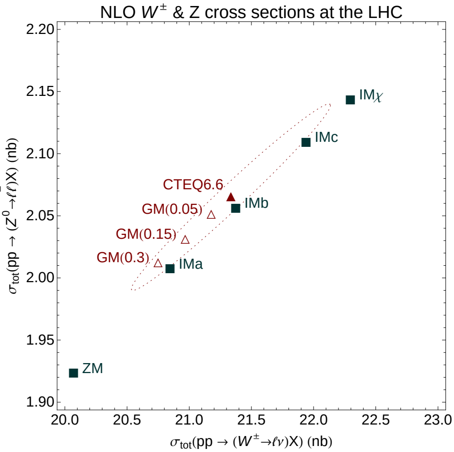
PDFs in GM/IMb/ZM/IMχ formulations



- At $x < 0.01$, ZM u, g are too small compared to GM
- IMχ: u, g are too large
- IMb: u, g are very close to GM
- $s(x, Q)$ (constrained by CC DIS) prefers IMχ rather than IMb

Dependence on the rescaling ζ in W, Z production at the LHC

- IMλ predictions span a wide range between ZM and IMχ
- IM predictions with $\lambda = 0.05 - 0.3$ are compatible with CT6.6M at $\approx 90\%$ c.l.
- ζ variable can be also introduced in the GM scheme; in this case, the standard choice $\zeta = \chi$ (CT6.6, or $\lambda = 0$) gives the best χ^2
- GM PDF's with $\lambda \lesssim 0.3$ agree with CT6.6M within the CT6.6 uncertainty



Conclusions

- The full GM heavy-quark kinematical dependence is approximated well **at NLO** by an **effective** IM scheme with ZM Wilson coefficients and generalized rescaling variable ζ
    - It remains to be seen if this scheme is viable **beyond NLO**
- The IM formulation can be applied to easily implement the leading heavy-quark mass effects in ZM calculations
- Variations in the form of ζ lead to an additional theoretical uncertainty that must be provided with GM predictions