CTEQ activities at MSU/SMU/Taiwan/Washington

Pavel Nadolsky
in collaboration with
J. Huston, H.-L. Lai, J. Pumplin, D. Stump, W.-K. Tung, C.-P. Yuan

Southern Methodist University
Dallas, TX, U.S.A.

April 28, 2009

DIS2009

Our group's other presentations at DIS'2009

- Combined fit of PDF's and Drell-Yan p_T distributions

  (H.-L. Lai, next talk)

- PDF's for leading-order Monte-Carlo programs

  (H.-L. Lai, Structure functions WG, 16:28)

- Heavy flavors *(P.N., Heavy-quark WG, 17:15)*

- This talk: everything else

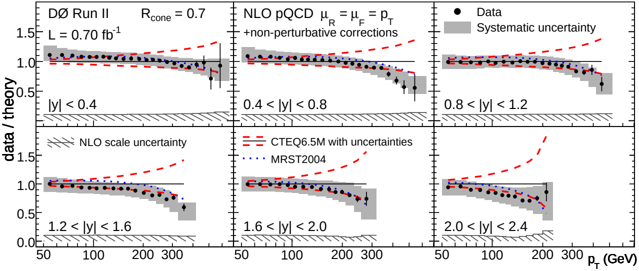
  - I will only have time to quickly flip some slides through – feel free to look them up on the computer and ask questions in the end

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

DIS2009

Inclusive jet production in Tevatron Run-2



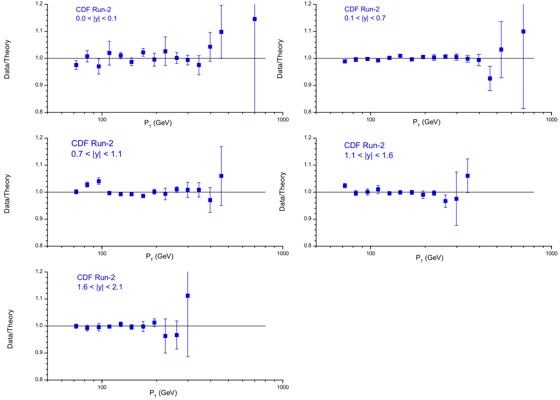
D0 Coll., arXiv:0802.2400 (700 pb^{-1}); CDF results (1.13 fb^{-1})

- (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

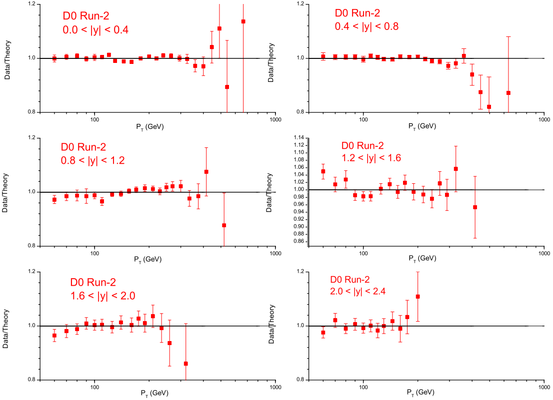


(Shifted CDF-Run 2 data)/CT09

DIS2009

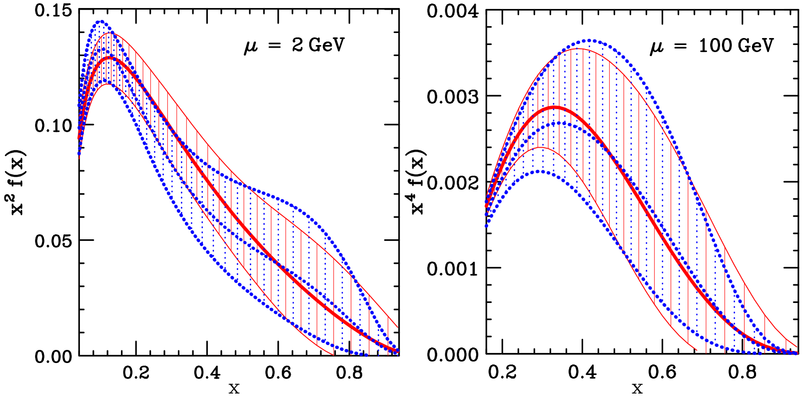
Impact of Run-2 jet data on CT09 fit

- 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

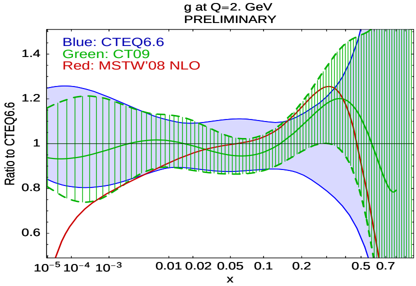
CT09 and CTEQ6.6 are generally compatible

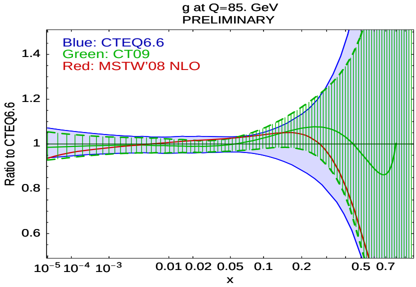


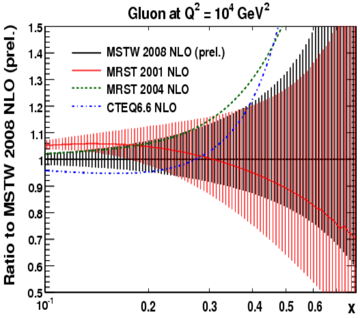
Gluon PDF: CT09 (red), CT66 (blue)

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

Impact of Run-2 jet data on global fits

Several issues affect the ability of the Run-2 data to constrain the PDF's

- reliability of theoretical predictions
- compatibility of the Run-1 and Run-2 measurements
- role of correlated systematic errors
- role of PDF parametrizations
- method for the computation of the PDF uncertainty (Hessian method; Lagrange multiplier; random sampling...)

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

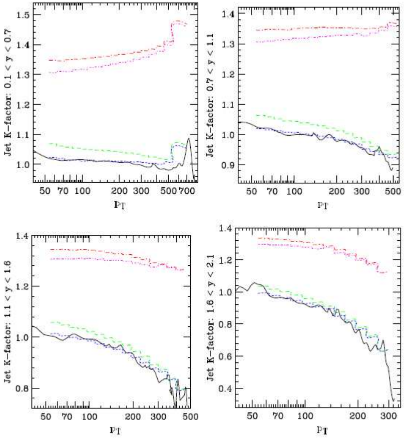


Figure 1: Theory calculations for the ratio K = NLO/LO from FastNLO and EKS. FastNLO with $\mu = p_T$: $R_{\rm sep} = 2.0$ (long dash dot), $R_{\rm sep} = 1.3$ (short dash dot); FastNLO with $\mu = p_T/2$: $R_{\rm sep} = 2.0$ (long dash), $R_{\rm sep} = 1.3$ (short dash); EKS with $\mu = p_T/2$, $R_{\rm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

Scale dependence of NLO cross section

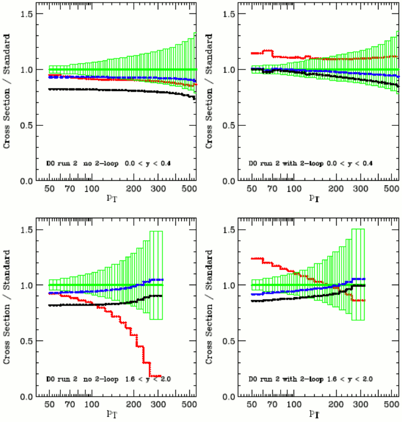


Figure 2: Effect of scale choice on predicted cross section with $R_{sep} = 1.3$: $\mu = 2\,p_T$ (short dash), p_T (long dash), $p_T/2$ (solid), $p_T/4$ (dotted), relative to our Standard Choice ($\mu = p_T/2$, $R_{sep} = 1.3$, no "two-loop" correction). Right panels include the "two-loop" resummation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

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?

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

DIS2009

Self-consistency of CT09 fit

$\mathrm{CDF_I}$ (33 pts)		$\mathrm{D0_I}$ (90 pts)		$\mathrm{CDF_{II}}$ (72 pts)		$\mathrm{D0_{II}}$ (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

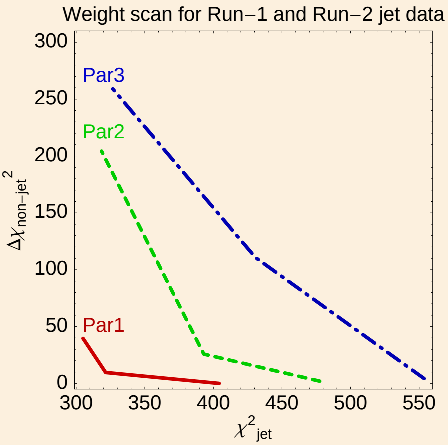
Dependence on the gluon PDF parametrization

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)

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

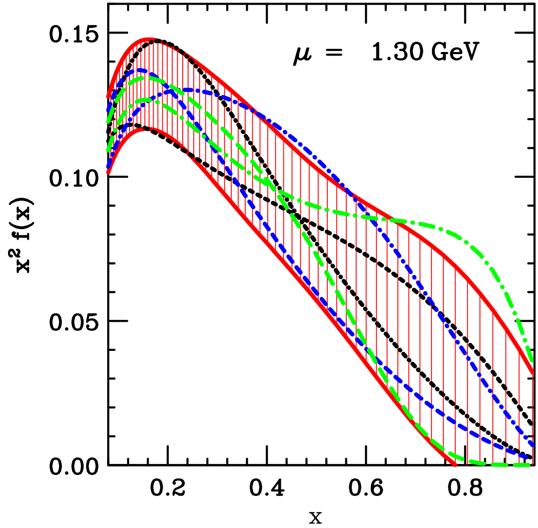
- β (used by CDF Run-1 and 2, D0 Run-2) has several practical advantages compared to C (used by D0 Run-1)

- 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; $\mathrm{rank}\left[\beta\beta^T\right] \approx N_\lambda/2 \ll N_{pt}$

- 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
  - $\mathrm{rank}\,[C - I] = \mathrm{rank}\left[\beta\beta^T\right] \approx N_{pt} = 90$ – too large

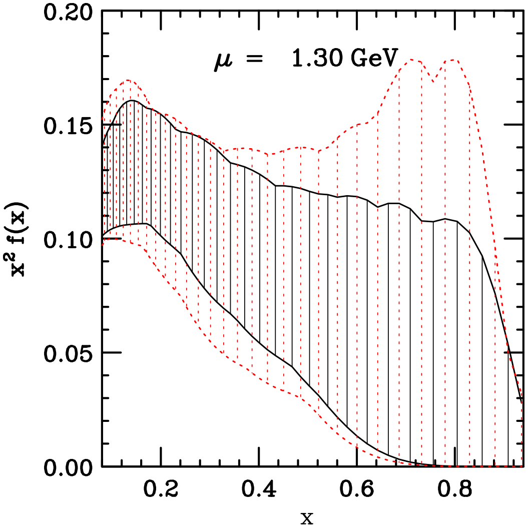
- This suggests that D0 Run-1 CSE's are overestimated; may explain consistently small $\chi^2_{\text{D0 Run-1}}/N_{pt} \sim 0.3$ in fits, other peculiarities of D0 Run-1 data observed in the weight scan

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)

Constraints of Run-2 data on CT09 PDFs



Black band: $g(x,\mu)$ from CT09 fit ("par 1")

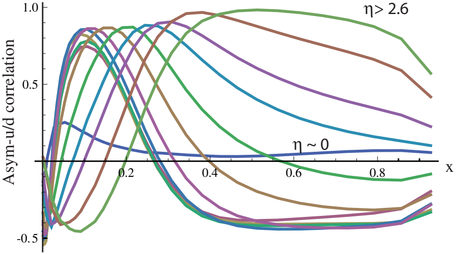
Red band: A "par 1" fit **without** Run-2 jet data

This band is wider than CT66 because of additional free parameters

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

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)*

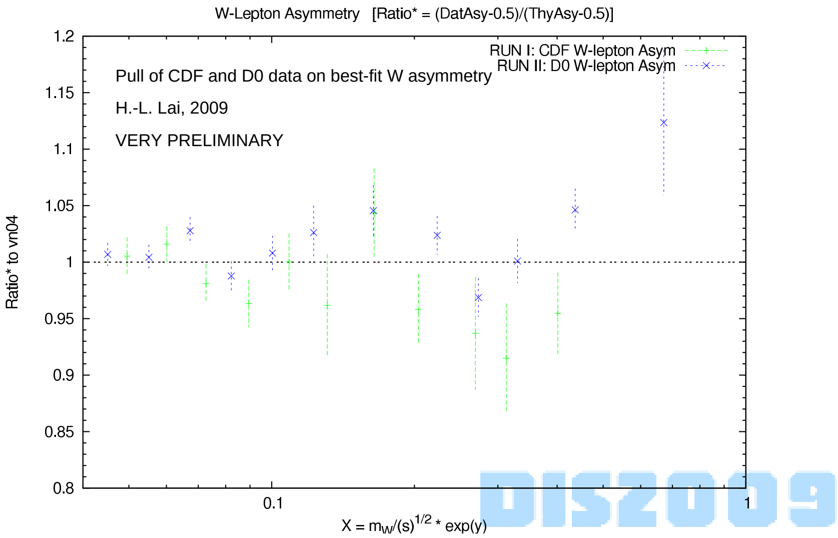
We find that CDF and D0 $A_\ell(y)$ data disagree in a similar kinematical range (confirming a similar MSTW finding)

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

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

DIS2009

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



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:

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

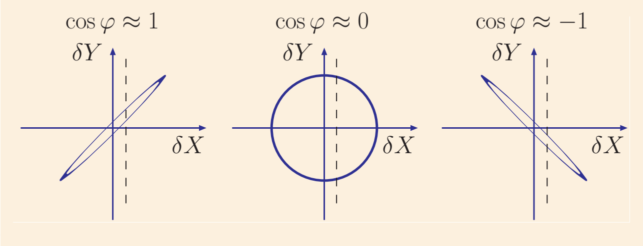
$$\cos\varphi = \frac{1}{4\Delta X\,\Delta Y}\sum_{i=1}^{N}\left(X_i^{(+)} - X_i^{(-)}\right)\left(Y_i^{(+)} - Y_i^{(-)}\right)$$

$X_i^{(\pm)}$ 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

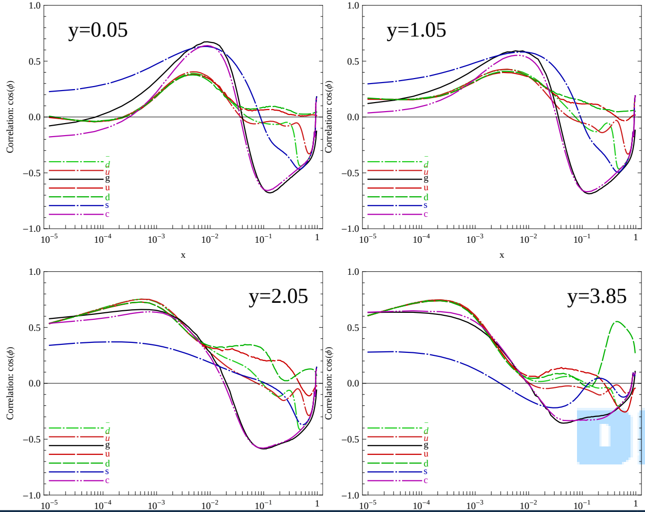
$\Delta X, \Delta Y$: PDF errors

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

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

Ongoing work on correlations

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



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

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

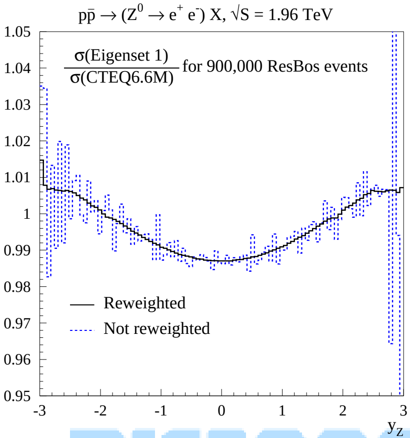
- Written in C, can be linked to standalone FORTRAN/C/C++ programs
- Simple – 170 lines of the code
- Writes the output directly into a ROOT ntuple; no need in intermediate PAW ntuples
- Flexible; new columns (branches) with PDF weights or events can be added into an existing ntuple
- Kinematical cuts, 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"
```



An exploratory technique for error analysis and propagation

Unconstrained (combinations of) PDF parameters a_i (N_{flat} flat directions in PDF space, corresponding to $\partial\chi/\partial a_i = 0$) may result in a large PDF uncertainty; CTEQ/MSTW fix such a_i by hand

Example: $s(x, \mu_0)$ at $x \to 0$ ($N_{flat} = 1$); $R_s \equiv \lim_{x\to 0} \left[(s+\bar{s}) / (\bar{u}+\bar{d})\right]$ is set to a fixed value, different in CTEQ and MSTW PDF's

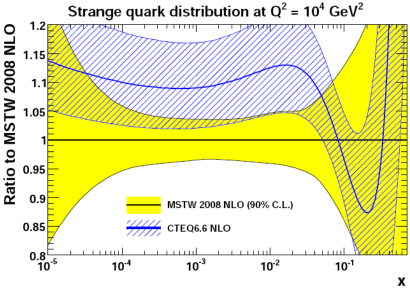
A better way is to provide a theoretically plausible uncertainty range $(\Delta X)_{flat}$ due to the flat directions (simple to do if $N_{flat} = 1$)

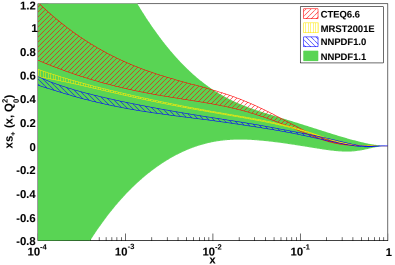
$(\Delta X)_{flat}$ can be added in **quadrature** to the usual Hessian error from $2N_{Hessian}$ error PDF's, **if the end user wishes to**

$$\Delta X^2 = \sum_{i=1}^{N_{Hessian}} \Delta X_i^2 + \Delta X_{flat}^2 \qquad (1)$$

$(\Delta X)_{flat}$ should be large enough to minimize subjective bias; constrained to exclude unphysical solutions

Unknown strangeness at $x \lesssim 10^{-2}$





Current CTEQ/MSTW/NNPDF errors on $s_+(x)$ at $x \to 0$ are unsatisfactory

- CTEQ, MSTW: too narrow and subjective
- NNPDF: too broad and potentially unphysical

CT09 strangeness vs CT66 and MSTW08NLO

with $0.2 \leq R_s \leq 1.5$ (imposed by Eq. (1))