Parton distribution functions and inclusive vector boson cross sections

Graeme Watt

University College London

YETI'09, Durham, 14th January 2009

[A.D. Martin, W.J. Stirling, R.S. Thorne and G.W., arXiv:0901.0002]

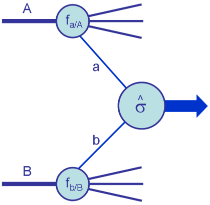
Introduction

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

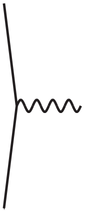
$$\sigma_{AB} = \sum_{a,b=q,g} \int_0^1 \mathrm{d}x_a \int_0^1 \mathrm{d}x_b \; f_{a/A}(x_a, Q^2) \; f_{b/B}(x_b, Q^2) \; \hat{\sigma}_{ab}$$

Outline of talk:

- Sketch of standard pQCD framework
- Determination of PDFs via global fits
- Recent developments [arXiv:0901.0002]
- Inclusive W and Z cross sections at LHC



Diagrammatic interpretation of collinear factorisation



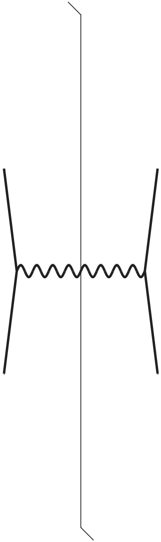
- Drell–Yan production at LO: $q\bar{q} \to V = W/Z/\gamma^*$
- Cut diagram: $|\mathcal{M}|^2 = \mathcal{M}\mathcal{M}^*$
- Large logarithm from collinear gluon emission: $\int_{k_0^2}^{Q^2} (\mathrm{d}k_T^2/k_T^2)\, \frac{\alpha_S}{2\pi}\, P_{q\leftarrow q}(z)$
- Similar collinear logs from other parton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

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

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

Factorisation $\Rightarrow$ Evolution $\Rightarrow$ Resummation

Diagrammatic interpretation of collinear factorisation



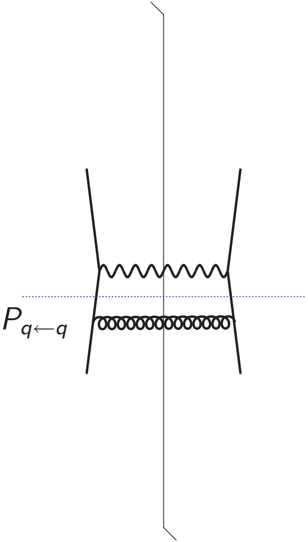
- Drell–Yan production at LO: $q\bar{q} \to V = W/Z/\gamma^*$
- Cut diagram: $|\mathcal{M}|^2 = \mathcal{M}\mathcal{M}^*$
- Large logarithm from collinear gluon emission: $\int_{k_0^2}^{Q^2} (\mathrm{d}k_T^2/k_T^2)\, \frac{\alpha_S}{2\pi}\, P_{q\leftarrow q}(z)$
- Similar collinear logs from other parton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

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

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

Factorisation $\Rightarrow$ Evolution $\Rightarrow$ Resummation

Diagrammatic interpretation of collinear factorisation



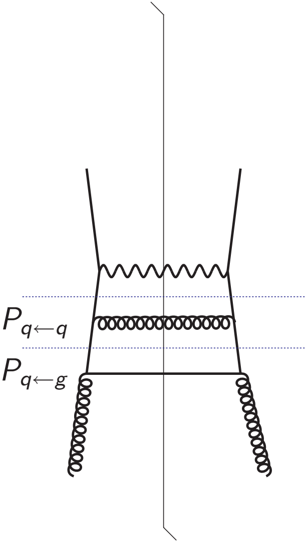
- Drell–Yan production at LO: $q\bar{q} \to V = W/Z/\gamma^*$
- Cut diagram: $|\mathcal{M}|^2 = \mathcal{M}\mathcal{M}^*$
- Large logarithm from collinear gluon emission: $\int_{k_0^2}^{Q^2} (\mathrm{d}k_T^2/k_T^2)\, \frac{\alpha_S}{2\pi}\, P_{q\leftarrow q}(z)$
- Similar collinear logs from other parton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

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

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

Factorisation ⇒ Evolution ⇒ Resummation

Diagrammatic interpretation of collinear factorisation



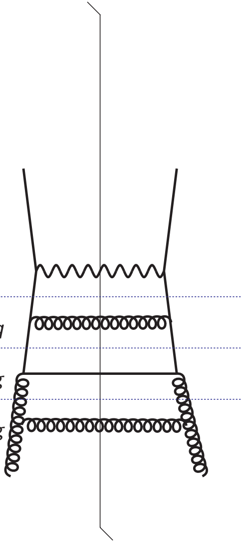
- Drell–Yan production at LO: $q\bar{q} \to V = W/Z/\gamma^*$
- Cut diagram: $|\mathcal{M}|^2 = \mathcal{M}\mathcal{M}^*$
- Large logarithm from collinear gluon emission: $\int_{k_0^2}^{Q^2}(\mathrm{d}k_T^2/k_T^2)\,\frac{\alpha_S}{2\pi}\,P_{q\leftarrow q}(z)$
- Similar collinear logs from other parton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

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

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

Factorisation $\Rightarrow$ Evolution $\Rightarrow$ Resummation

Diagrammatic interpretation of collinear factorisation



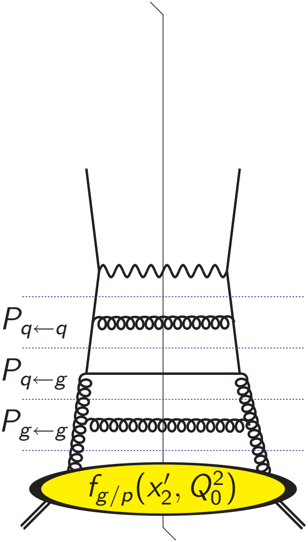
- Drell–Yan production at LO: $q\bar{q} \to V = W/Z/\gamma^*$
- Cut diagram: $|\mathcal{M}|^2 = \mathcal{M}\mathcal{M}^*$
- Large logarithm from collinear gluon emission: $\int_{k_0^2}^{Q^2}(\mathrm{d}k_T^2/k_T^2)\,\frac{\alpha_S}{2\pi}\,P_{q\leftarrow q}(z)$
- Similar collinear logs from other parton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

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

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

Factorisation $\Rightarrow$ Evolution $\Rightarrow$ Resummation

Diagrammatic interpretation of collinear factorisation



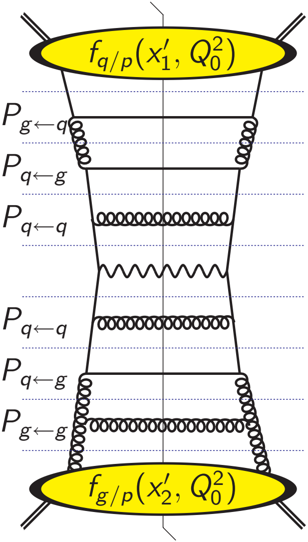
- Drell–Yan production at LO: $q\bar{q} \to V = W/Z/\gamma^*$
- Cut diagram: $|\mathcal{M}|^2 = \mathcal{M}\mathcal{M}^*$
- Large logarithm from collinear gluon emission: $\int_{k_0^2}^{Q^2} (\mathrm{d}k_T^2/k_T^2)\, \frac{\alpha_S}{2\pi}\, P_{q\leftarrow q}(z)$
- Similar collinear logs from other parton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

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

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

Factorisation $\Rightarrow$ Evolution $\Rightarrow$ Resummation

Diagrammatic interpretation of collinear factorisation



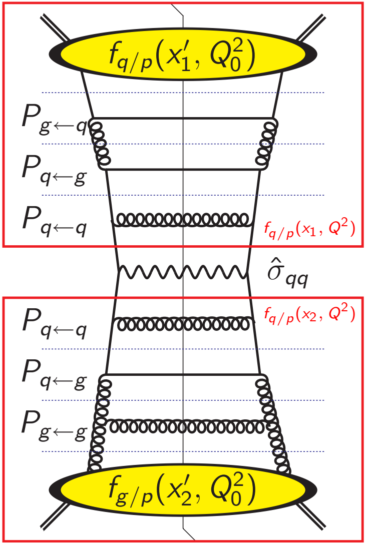
- Drell–Yan production at LO: $q\bar{q} \rightarrow V = W/Z/\gamma^*$
- Cut diagram: $|\mathcal{M}|^2 = \mathcal{M}\mathcal{M}^*$
- Large logarithm from collinear gluon emission: $\int_{k_0^2}^{Q^2} (\mathrm{d}k_T^2/k_T^2)\, \frac{\alpha_S}{2\pi}\, P_{q\leftarrow q}(z)$
- Similar collinear logs from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

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

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

Factorisation ⇒ Evolution ⇒ Resummation

Diagrammatic interpretation of collinear factorisation

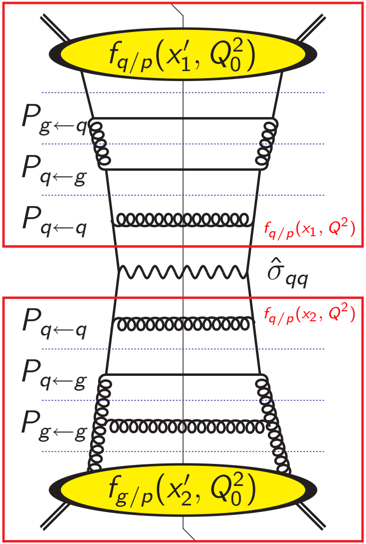


- Drell–Yan production at LO: $q\bar{q} \to V = W/Z/\gamma^*$
- Cut diagram: $|\mathcal{M}|^2 = \mathcal{M}\mathcal{M}^*$
- Large logarithm from collinear gluon emission: $\int_{k_0^2}^{Q^2} (\mathrm{d}k_T^2/k_T^2)\,\frac{\alpha_S}{2\pi}\,P_{q\leftarrow q}(z)$
- Similar collinear logs from other parton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

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

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

Diagrammatic interpretation of collinear factorisation



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

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

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

Factorisation ⇒ Evolution ⇒ Resummation

LO DGLAP splitting functions and sum rules

Splitting functions $P_{a\leftarrow a'}(z)$ at leading-order

$$P^{\rm LO}_{q\leftarrow q}(z) = C_F\left[\frac{1+z^2}{(1-z)_+}+\frac{3}{2}\,\delta(1-z)\right],\quad \int_0^1 {\rm d}z\,\frac{f(x)}{(1-z)_+} = \int_0^1 {\rm d}z\,\frac{f(z)-f(1)}{1-z}$$

$$P^{\rm LO}_{g\leftarrow q}(z) = C_F\left[\frac{1+(1-z)^2}{z}\right],\quad P^{\rm LO}_{q\leftarrow g}(z) = T_R\left[z^2+(1-z)^2\right]$$

$$P^{\rm LO}_{g\leftarrow g}(z) = 2\,C_A\left[\frac{z}{(1-z)_+}+\frac{1-z}{z}+z(1-z)\right]+\frac{1}{6}\,(11\,C_A-4\,n_f\,T_R)\,\delta(1-z)$$

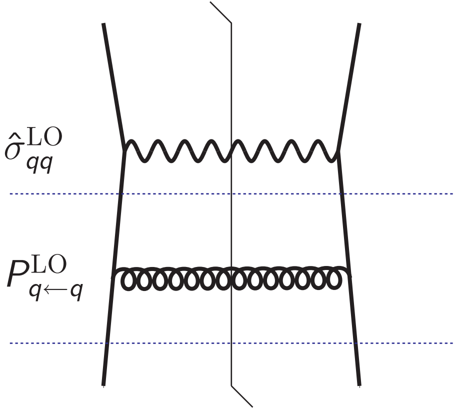
Number- and momentum-sum rules (preserved by evolution)

$$\int_0^1 {\rm d}x\,\left[f_{q/p}(x,Q_0^2)-f_{\bar{q}/p}(x,Q_0^2)\right] = n_q \qquad (n_u = 2,\, n_d = 1,\, n_s = 0)$$

$$\sum_{a=q,\bar{q},g}\int_0^1 {\rm d}x\; x f_{a/p}(x,Q_0^2) = 1$$

Higher-order corrections in collinear factorisation

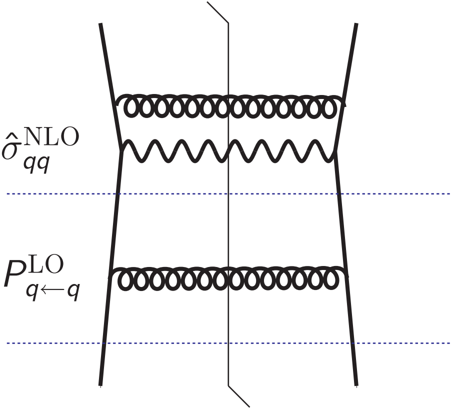
- Consider hard-scattering and last evolution step at LO.



- NLO correction to $\hat{\sigma}_{qq}$.
- Collinear gluon emission already included in $P_{q\leftarrow q}^{\rm LO}$.
- Need to subtract collinear divergence (plus certain finite terms).
- Defines **factorisation scheme**, usually $\overline{\rm MS}$: work in $4-2\epsilon$ dimensions, then divergence appears as pole $P_{q\leftarrow q}^{\rm LO}(z)/\epsilon$.
- Also need NLO corrections to $P_{a\leftarrow a'}$ in same scheme.
- Similarly at NNLO. Include all $\mathcal{O}(\alpha_S^2)$ corrections to both $\hat{\sigma}_{ab}$ and $P_{a\leftarrow a'}$.

Higher-order corrections in collinear factorisation

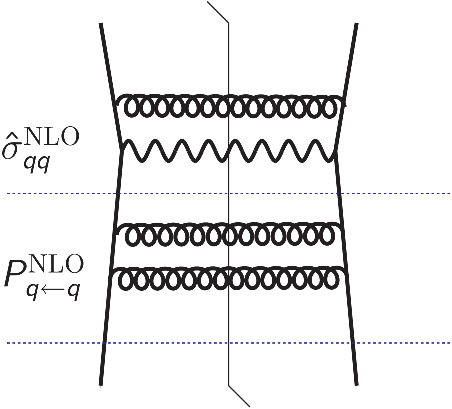
- Consider hard-scattering and last evolution step at LO.



- NLO correction to $\hat{\sigma}_{qq}$.
- Collinear gluon emission already included in $P_{q\leftarrow q}^{\rm LO}$.
- Need to subtract collinear divergence (plus certain finite terms).
- Defines **factorisation scheme**, usually $\overline{\rm MS}$: work in $4-2\epsilon$ dimensions, then divergence appears as pole $P_{q\leftarrow q}^{\rm LO}(z)/\epsilon$.
- Also need NLO corrections to $P_{a\leftarrow a'}$ in same scheme.
- Similarly at NNLO. Include all $\mathcal{O}(\alpha_S^2)$ corrections to both $\hat{\sigma}_{ab}$ and $P_{a\leftarrow a'}$.

Higher-order corrections in collinear factorisation

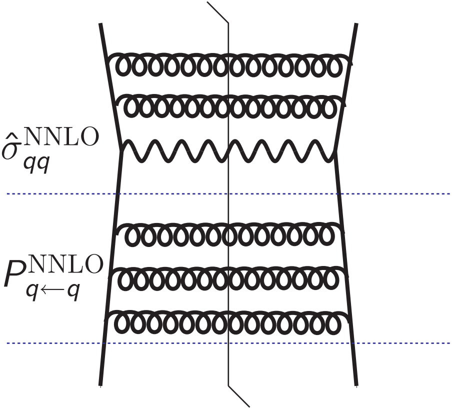
- Consider hard-scattering and last evolution step at LO.



- NLO correction to $\hat{\sigma}_{qq}$.
- Collinear gluon emission already included in $P^{\mathrm{LO}}_{q\leftarrow q}$.
- Need to subtract collinear divergence (plus certain finite terms).
- Defines **factorisation scheme**, usually $\overline{\mathrm{MS}}$: work in $4 - 2\epsilon$ dimensions, then divergence appears as pole $P^{\mathrm{LO}}_{q\leftarrow q}(z)/\epsilon$.
- Also need NLO corrections to $P_{a\leftarrow a'}$ in same scheme.
- Similarly at NNLO. Include all $\mathcal{O}(\alpha_S^2)$ corrections to both $\hat{\sigma}_{ab}$ and $P_{a\leftarrow a'}$.

Higher-order corrections in collinear factorisation

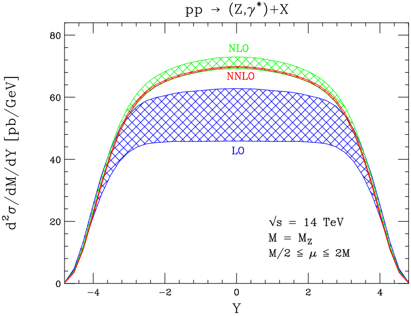
- Consider hard-scattering and last evolution step at LO.



- NLO correction to $\hat{\sigma}_{qq}$.
- Collinear gluon emission already included in $P_{q\leftarrow q}^{\rm LO}$.
- Need to subtract collinear divergence (plus certain finite terms).
- Defines **factorisation scheme**, usually $\overline{\rm MS}$: work in $4-2\epsilon$ dimensions, then divergence appears as pole $P_{q\leftarrow q}^{\rm LO}(z)/\epsilon$.
- Also need NLO corrections to $P_{a\leftarrow a'}$ in same scheme.
- Similarly at NNLO. Include all $\mathcal{O}(\alpha_S^2)$ corrections to both $\hat{\sigma}_{ab}$ and $P_{a\leftarrow a'}$.

NNLO calculations for inclusive vector boson production

- NNLO splitting functions calculated by Moch, Vermaseren and Vogt [hep-ph/0403192, hep-ph/0404111]
- NNLO vector boson (W, Z, γ^*) total cross sections known for a while [Hamberg, van Neerven, Matsuura, '91 (Erratum, '02)]



[hep-ph/0312266]

- NNLO rapidity distributions calculated by Anastasiou, Dixon, Melnikov and Petriello [hep-ph/0312266] available in VRAP code.
- Spin correlations included (allowing cuts on leptonic decay products) by Melnikov and Petriello [hep-ph/0609070] available in FEWZ code.

Aside: relation of PDF evolution to parton shower

Collinear factorisation forms the **basis** for parton showers implemented in Monte Carlo event generators such as HERWIG/PYTHIA/SHERPA.

$$\mathrm{d}\hat{\sigma}_n = \frac{\mathrm{d}k_{T,n}^2}{k_{T,n}^2}\,\mathrm{d}z_n\,\frac{\alpha_S}{2\pi}\,P_{a\leftarrow a'}(z_n)\,\mathrm{d}\hat{\sigma}_{n-1}$$

Fixed-order collinear factorisation

- Emissions are **strongly** ordered: $Q^2 \gg k_{T,n}^2 \gg k_{T,n-1}^2 \gg \dots$
- Can be improved to higher orders for both $P_{a\leftarrow a'}$ and $\hat{\sigma}_{ab}$.

Parton shower in event generators

- Hard-scattering as in fixed-order approach, but then add additional backwards evolution steps: "undo" PDF evolution.
- Relax **strong** ordering: $Q^2 > k_{T,n}^2 > k_{T,n-1}^2 > \dots$
- Only $P_{a\leftarrow a'}^{\mathrm{LO}}$ in parton shower, but some higher-order effects.
- Higher-orders in $\hat{\sigma}_{ab}$ now possible (MC@NLO, POWHEG).

Aside: relation of PDF evolution to parton shower

Collinear factorisation forms the **basis** for parton showers implemented in Monte Carlo event generators such as HERWIG/PYTHIA/SHERPA.

$$\mathrm{d}\hat{\sigma}_n = \frac{\mathrm{d}k_{T,n}^2}{k_{T,n}^2}\,\mathrm{d}z_n\,\frac{\alpha_S}{2\pi}\,P_{a\leftarrow a'}(z_n)\,\mathrm{d}\hat{\sigma}_{n-1}$$

Fixed-order collinear factorisation

- Emissions are **strongly** ordered: $Q^2 \gg k_{T,n}^2 \gg k_{T,n-1}^2 \gg \ldots$
- Can be improved to higher orders for both $P_{a\leftarrow a'}$ and $\hat{\sigma}_{ab}$.

Parton shower in event generators

- Hard-scattering as in fixed-order approach, but then add additional backwards evolution steps: "undo" PDF evolution.
- Relax **strong** ordering: $Q^2 > k_{T,n}^2 > k_{T,n-1}^2 > \ldots$
- Only $P_{a\leftarrow a'}^{\mathrm{LO}}$ in parton shower, but some higher-order effects.
- Higher-orders in $\hat{\sigma}_{ab}$ now possible (MC@NLO, POWHEG).

Aside: PDFs for use in LO Monte Carlo event generators

- Which PDFs ($f^X, X = \mathrm{LO}, \mathrm{NLO}, \ldots$) to use if only a LO $\hat{\sigma}$ is available? Define the "truth" to be $f^{\mathrm{NLO}} \otimes \hat{\sigma}^{\mathrm{NLO}}$, and

$$\mathrm{K}(X) = \frac{f^{\mathrm{NLO}} \otimes \hat{\sigma}^{\mathrm{NLO}}}{f^X \otimes \hat{\sigma}^{\mathrm{LO}}}$$

- A. Sherstnev and R. Thorne have studied modified LO PDFs (LO* and LO**) which give K(LO*) or K(LO**) much closer to 1 than either K(LO) or K(NLO) for a variety of processes.

  LO* [arXiv:0711.2473] = LO PDF fit with violation of momentum-sum rule and NLO α_S.

  LO** [arXiv:0807.2132] = same but also modified α_S scale in PDF evolution, similar to in parton shower.

- LO* and LO** now available in LHAPDF library.
- Will be updated soon to include PDF uncertainty sets.

Fixed-order collinear factorisation at hadron colliders

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

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

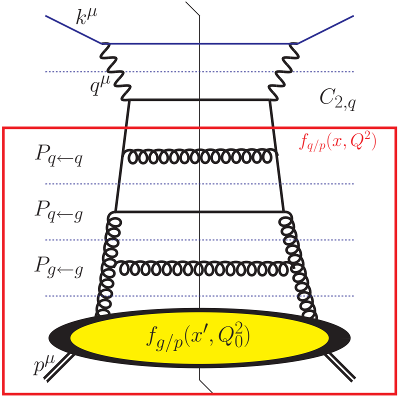
PDF evolution: $$\frac{\partial f_{a/A}}{\partial \ln Q^2} = \frac{\alpha_S}{2\pi} \sum_{a'=q,g} \left[P^{\rm LO}_{aa'} + \alpha_S P^{\rm NLO}_{aa'} + \ldots\right] \otimes f_{a'/A}$$

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

Need to extract input values $f_{a/A}(x, Q_0^2)$ and $\alpha_S(M_Z^2)$ from data.

- Works well for **inclusive** processes with **one** hard scale. Fails for processes with **two** disparate hard scales, e.g.
  1. $p_T^2 \ll M^2 \Rightarrow \ln(M^2/p_T^2)$ terms $\Rightarrow$ RESBOS, parton shower
  2. $M^2 \ll s \Rightarrow \ln(s/M^2)$ terms $\Rightarrow$ high-energy/small-x/BFKL

Structure functions in deep-inelastic scattering (DIS)



Kinematic variables:

$$Q^2 = -q^2 > 0$$

$$W^2 = (q+p)^2$$

$$s = (k+p)^2$$

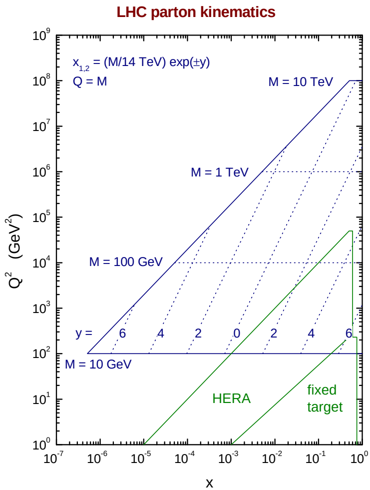
$$x_{\rm Bj} = \frac{Q^2}{2p \cdot q} \simeq \frac{Q^2}{Q^2 + W^2}$$

$$y = \frac{q \cdot p}{k \cdot p} \simeq \frac{Q^2}{x_{\rm Bj}\, s}$$

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

$$\equiv \sum_{a=q,g} C_{i,a} \otimes f_{a/p}, \qquad C_{i,a} = C_{i,a}^{\rm LO} + \alpha_S\, C_{i,a}^{\rm NLO} + \ldots$$

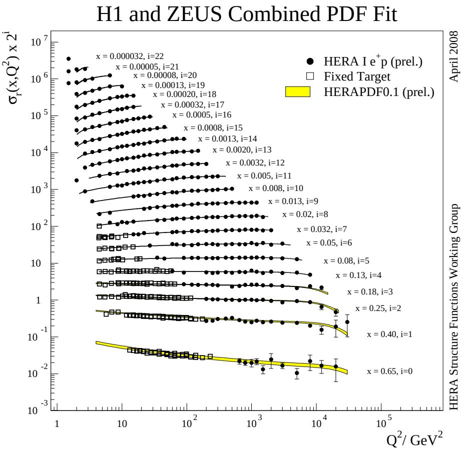
From HERA *et al.* to the LHC



[Plot by J. Stirling]

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

HERA reduced cross section $\sigma_r(x, Q^2) \approx F_2(x, Q^2)$



- H1 and ZEUS measurements combined to reduce errors.
- At leading-order:

$$C_{2,q}^{\rm LO}(z) = e_q^2\,\delta\,(1-z)$$
$$C_{2,g}^{\rm LO} = 0$$
$$\Rightarrow\; F_2^{\rm LO}(x, Q^2) = \sum_q e_q^2\, x f_{q/p}$$

$$\frac{\partial F_2^{\rm LO}(x, Q^2)}{\partial \ln Q^2} = \frac{\alpha_S}{2\pi}\left[P_{q\leftarrow q}^{\rm LO} \otimes F_2^{\rm LO} + \sum_q e_q^2\, P_{q\leftarrow g}^{\rm LO} \otimes f_{g/p}\right]$$

Which processes constrain different PDFs?

Process	Subprocess	Partons	x range
$\ell^\pm\,\{p,n\} \to \ell^\pm\,X$	$\gamma^* q \to q$	$q,\bar{q},g$	$x \gtrsim 0.01$
$\ell^\pm\,n/p \to \ell^\pm\,X$	$\gamma^*\,d/u \to d/u$	d/u	$x \gtrsim 0.01$
$pp \to \mu^+\mu^-\,X$	$u\bar{u}, d\bar{d} \to \gamma^*$	$\bar{q}$	$0.015 \lesssim x \lesssim 0.35$
$pn/pp \to \mu^+\mu^-\,X$	$(u\bar{d})/(u\bar{u}) \to \gamma^*$	$\bar{d}/\bar{u}$	$0.015 \lesssim x \lesssim 0.35$
$\nu(\bar{\nu})\,N \to \mu^-(\mu^+)\,X$	$W^* q \to q'$	$q,\bar{q}$	$0.01 \lesssim x \lesssim 0.5$
$\nu\,N \to \mu^-\mu^+\,X$	$W^* s \to c$	s	$0.01 \lesssim x \lesssim 0.2$
$\bar{\nu}\,N \to \mu^+\mu^-\,X$	$W^* \bar{s} \to \bar{c}$	$\bar{s}$	$0.01 \lesssim x \lesssim 0.2$
$e^\pm\,p \to e^\pm\,X$	$\gamma^* q \to q$	$g,q,\bar{q}$	$0.0001 \lesssim x \lesssim 0.1$
$e^+\,p \to \bar{\nu}\,X$	$W^+\,\{d,s\} \to \{u,c\}$	d,s	$x \gtrsim 0.01$
$e^\pm p \to e^\pm\,c\bar{c}\,X$	$\gamma^* c \to c,\ \gamma^* g \to c\bar{c}$	$c,\ g$	$0.0001 \lesssim x \lesssim 0.01$
$e^\pm p \to \text{jet} + X$	$\gamma^* g \to q\bar{q}$	g	$0.01 \lesssim x \lesssim 0.1$
$p\bar{p} \to \text{jet} + X$	$gg, qg, qq \to 2j$	g,q	$0.01 \lesssim x \lesssim 0.5$
$p\bar{p} \to (W^\pm \to \ell^\pm\nu)\,X$	$ud \to W, \bar{u}\bar{d} \to W$	$u,d,\bar{u},\bar{d}$	$x \gtrsim 0.05$
$p\bar{p} \to (Z \to \ell^+\ell^-)\,X$	$uu, dd \to Z$	d	$x \gtrsim 0.05$

Paradigm for PDF determination by "global analysis"

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

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

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

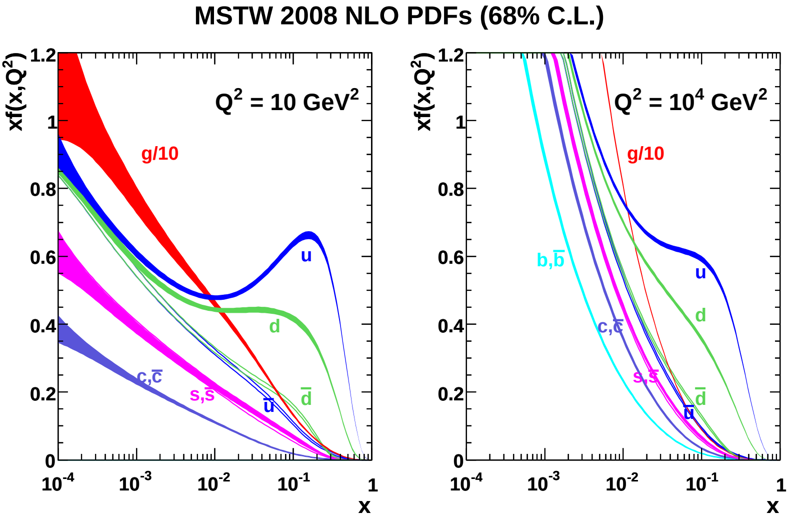
$$\chi^2 = \sum_{i=1}^{N_{\text{pts.}}}\left(\frac{\text{Data}_i - \text{Theory}_i}{\text{Error}_i}\right)^2.$$

Determination of parton distributions by global analysis

An "industry" for more than 20 years. Regular updates as new data and theory become available.

1. Major group: "**MRST**" = **M**artin+**R**oberts+**S**tirling+**T**horne. Recently, "**MSTW**" = **M**R**ST**−Roberts+G.**W**.
   {MRST 2001 LO, MRST 2004 NLO, MRST 2006 NNLO}
   → MSTW 2008 LO, NLO, NNLO [arXiv:0901.0002]
2. Other major group: "**CTEQ**" = **C**oordinated **T**heoretical–**E**xperimental Project on **Q**CD. Last CTEQ6.6 analysis by groups at Michigan State/Taiwan/Washington.
   - CTEQ6L1 LO [hep-ph/0201195]
   - CTEQ6.6 NLO [arXiv:0802.0007]
3. Other groups fitting a restricted range of data with fewer free parameters: S. Alekhin *et al.*, HERA experiments (H1, ZEUS).
4. NNPDF Collaboration: see later.

Example of PDFs obtained from global analysis



Public access to PDF sets

- In principle, just need input parameterisation at scale Q_0^2, then can DGLAP-evolve with public codes to higher scales Q^2.
  1. PEGASUS [A. Vogt, hep-ph/0408244] : Mellin-space.
  2. HOPPET [G. Salam, arXiv:0804.3755] : x-space.
- In practice, simpler and more efficient for fitting groups to provide "pre-evolved" grids in (x, Q^2) with interpolation code.

CTEQ	http://hep.pa.msu.edu/cteq/public/cteq6.html (Fortran)
MRST	http://durpdg.dur.ac.uk/hepdata/mrs.html (Fortran)
MSTW	http://projects.hepforge.org/mstwpdf/ (Fortran, C++, Mathematica)
LHAPDF	http://projects.hepforge.org/lhapdf/ (Fortran, C++ wrapper)

- Les Houches Accord PDF (LHAPDF): common interface to all PDF sets (originally by W. Giele, maintained by M. Whalley).
- On-line PDF plotting (also by M. Whalley): http://durpdg.dur.ac.uk/hepdata/pdf3.html

Data sets fitted in MSTW 2008 NLO analysis [arXiv:0901.0002]

Data set	χ^2 / $N_{\rm pts.}$
H1 MB 99 e^+p NC	9 / 8
H1 MB 97 e^+p NC	42 / 64
H1 low Q^2 96–97 e^+p NC	44 / 80
H1 high Q^2 98–99 e^-p NC	122 / 126
H1 high Q^2 99–00 e^+p NC	131 / 147
ZEUS SVX 95 e^+p NC	35 / 30
ZEUS 96–97 e^+p NC	86 / 144
ZEUS 98–99 e^-p NC	54 / 92
ZEUS 99–00 e^+p NC	63 / 90
H1 99–00 e^+p CC	29 / 28
ZEUS 99–00 e^+p CC	38 / 30
H1/ZEUS $e^\pm p$ $F_2^{\rm charm}$	107 / 83
H1 99–00 e^+p incl. jets	19 / 24
ZEUS 96–97 e^+p incl. jets	30 / 30
ZEUS 98–00 $e^\pm p$ incl. jets	17 / 30
DØ II $p\bar{p}$ incl. jets	114 / 110
CDF II $p\bar{p}$ incl. jets	56 / 76
CDF II $W \to l\nu$ asym.	29 / 22
DØ II $W \to l\nu$ asym.	25 / 10
DØ II Z rap.	19 / 28
CDF II Z rap.	49 / 29

Data set	χ^2 / $N_{\rm pts.}$
BCDMS μp F_2	182 / 163
BCDMS μd F_2	190 / 151
NMC μp F_2	121 / 123
NMC μd F_2	102 / 123
NMC $\mu n/\mu p$	130 / 148
E665 μp F_2	57 / 53
E665 μd F_2	53 / 53
SLAC ep F_2	30 / 37
SLAC ed F_2	30 / 38
NMC/BCDMS/SLAC F_L	38 / 31
E866/NuSea pp DY	228 / 184
E866/NuSea pd/pp DY	14 / 15
NuTeV νN F_2	49 / 53
CHORUS νN F_2	26 / 42
NuTeV νN xF_3	40 / 45
CHORUS νN xF_3	31 / 33
CCFR $\nu N \to \mu\mu X$	66 / 86
NuTeV $\nu N \to \mu\mu X$	39 / 40
All data sets	**2543 / 2699**

- Red = New w.r.t. MRST 2006 fit.

Input parameterisation in MSTW 2008 NLO fit

At input scale $Q_0^2 = 1\ \mathrm{GeV}^2$ (notation: $f_{a/p} \equiv a$):

$$
\begin{aligned}
xu_v &= A_u\, x^{\eta_1}(1-x)^{\eta_2}(1+\epsilon_u\sqrt{x}+\gamma_u\, x)\\
xd_v &= A_d\, x^{\eta_3}(1-x)^{\eta_4}(1+\epsilon_d\sqrt{x}+\gamma_d\, x)\\
xS &= A_S\, x^{\delta_S}(1-x)^{\eta_S}(1+\epsilon_S\sqrt{x}+\gamma_S\, x)\\
x(\bar{d}-\bar{u}) &= A_\Delta\, x^{\eta_\Delta}(1-x)^{\eta_S+2}(1+\gamma_\Delta\, x+\delta_\Delta\, x^2)\\
xg &= A_g\, x^{\delta_g}(1-x)^{\eta_g}(1+\epsilon_g\sqrt{x}+\gamma_g\, x)+A_{g'}\, x^{\delta_{g'}}(1-x)^{\eta_{g'}}\\
x(s+\bar{s}) &= A_+\, x^{\delta_S}(1-x)^{\eta_+}(1+\epsilon_S\sqrt{x}+\gamma_S\, x)\\
x(s-\bar{s}) &= A_-\, x^{\delta_-}(1-x)^{\eta_-}(1-x/x_0)
\end{aligned}
$$

- A_u, A_d, A_g and x_0 are determined from sum rules.
- 20 parameters allowed to go free for error propagation, cf. 15 for MRST error PDF sets.

Strangeness in the proton

- Protons have **no** valence strange quarks, i.e.

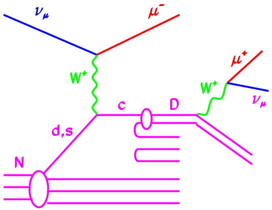
$$\int_0^1 \mathrm{d}x \ \left[s(x,Q^2) - \bar{s}(x,Q^2)\right] = 0.$$

- But this does **not** necessarily mean that $s(x,Q^2) = \bar{s}(x,Q^2)\ \forall x$.
- **Assumption** in most recent PDF fits is that

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

  with $\kappa \approx 0.5$, justified by a simplified analysis of dimuon data by the CCFR experiment [hep-ex/9406007].
- Updated CCFR/NuTeV dimuon cross sections now available $\Rightarrow$ include in global fit to constrain s and $\bar{s}$ directly.

NuTeV/CCFR dimuon cross sections and strangeness



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

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

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

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

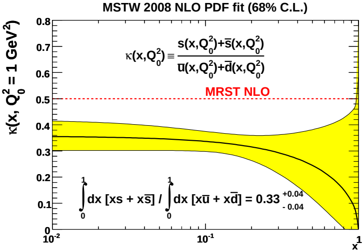
- **Parameterise** at input scale of $Q_0^2 = 1$ GeV^2 in the form:

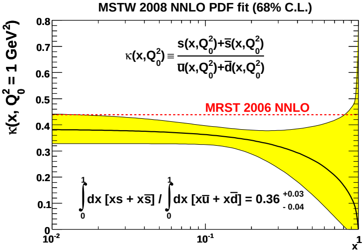
$$xs(x,Q_0^2) + x\bar{s}(x,Q_0^2) = \tilde{A}_+ \,(1-x)^{\tilde{\eta}_+}\, xS(x,Q_0^2),$$
$$xs(x,Q_0^2) - x\bar{s}(x,Q_0^2) = A_-\, x^{0.2}(1-x)^{\eta_-}(1-x/x_0).$$

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

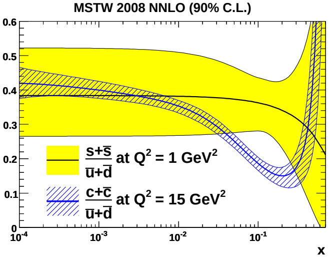
Ratio of strange sea to non-strange sea: $(s+\bar{s})/(\bar{u}+\bar{d})$

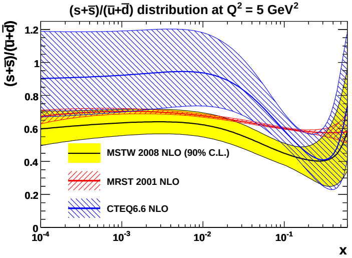
- Mass suppression of strange sea similar at NLO and NNLO:





- CTEQ6.6 larger at small x (more flexible where no data):





Aside: the NuTeV $sin^2\theta_W$ "anomaly"

- NuTeV extraction [hep-ex/0110059] of $\sin^2\theta_W$ from

$$R^- \equiv \frac{\sigma(\nu_\mu N \to \nu_\mu X) - \sigma(\bar{\nu}_\mu N \to \bar{\nu}_\mu X)}{\sigma(\nu_\mu N \to \mu^- X) - \sigma(\bar{\nu}_\mu N \to \mu^+ X)} \approx \frac{1}{2} - \sin^2\theta_W$$

was $\sim 3\sigma$ above the Standard Model prediction.

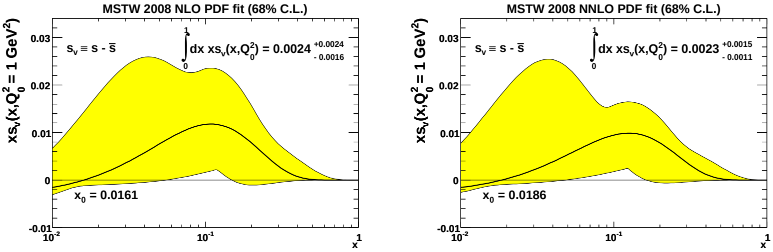
- New Physics? Or neglected PDF-related uncertainties, e.g.
  1. **Isospin violation?** ($u^p \neq d^n$, $d^p \neq u^n$)

$$R^- \approx \frac{1}{2} - \sin^2\theta_W + (1 - \frac{7}{3}\sin^2\theta_W)\frac{\int_0^1 \mathrm{d}x\, x\,[(u_v^p - d_v^n) - (d_v^p - u_v^n)]}{2\int_0^1 \mathrm{d}x\, x(u_v + d_v)}$$

    - Generated by QED corrections to parton evolution.
    - Found by MRST [hep-ph/0411040] to remove a little more than 1σ of the total discrepancy in the NuTeV $\sin^2\theta_W$.
  2. **Strange sea asymmetry?** ($s \neq \bar{s}$)

$$R^- \approx \frac{1}{2} - \sin^2\theta_W - (1 - \frac{7}{3}\sin^2\theta_W)\frac{\int_0^1 \mathrm{d}x\, x(s - \bar{s})}{\int_0^1 \mathrm{d}x\, x(u_v + d_v)}$$

Strange sea asymmetry: $xs_v = xs - x\bar{s}$



	$\int_0^1 \mathrm{d}x\, x(s-\bar{s})$		
Q^2/GeV^2	LO	NLO	NNLO
1	$0.0026^{+0.0022}_{-0.0017}$	$0.0024^{+0.0024}_{-0.0016}$	$0.0023^{+0.0015}_{-0.0011}$
10	$0.0019^{+0.0016}_{-0.0012}$	$0.0018^{+0.0018}_{-0.0012}$	$0.0016^{+0.0011}_{-0.0009}$
100	$0.0016^{+0.0014}_{-0.0010}$	$0.0015^{+0.0015}_{-0.0010}$	$0.0013^{+0.0010}_{-0.0007}$

- Slight preference for a positive asymmetry, but consistent with zero within 90% C.L. limit. (68% C.L. shown here.)
- Positive value of $\int_0^1 \mathrm{d}x\, x(s-\bar{s}) \sim 0.007$ would bring the NuTeV $\sin^2\theta_W$ to the world average. Values here reduce anomaly.

Tevatron data on inclusive jet production in PDF fits

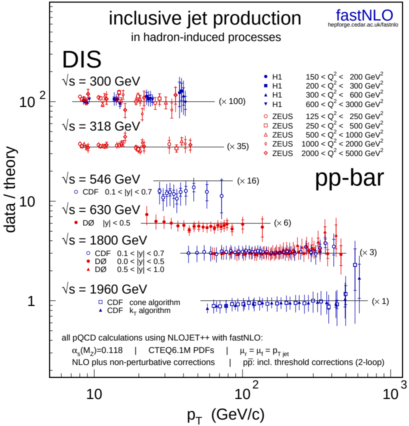
- Early CDF Run I data at central rapidity indicated an **excess** at high-E_T using NLO QCD with available PDFs of the time, taken initially as a possible sign of quark compositeness.
- Accommodated within global fit by modified high-x gluon distribution. Subsequent DØ Run I data measured E_T distribution at range of rapidities: confirmed new global fits.
- **Tevatron jet data** now routinely used in global fits: constrain the high-x gluon important for new physics searches at LHC.

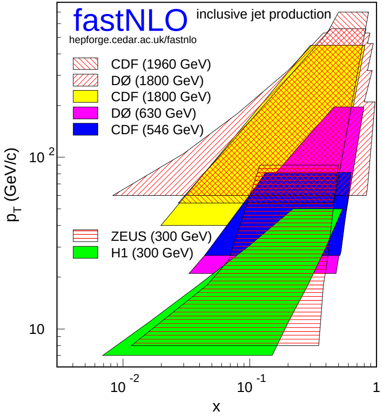
$$\sigma_{p\bar{p}} = \alpha_S^2(Q^2) \sum_{a,b=q,g} \left[\hat{\sigma}_{ab}^{\rm LO} + \alpha_S(Q^2)\hat{\sigma}_{ab}^{\rm NLO}\right] \otimes f_{a/p}(x_a, Q^2) \otimes f_{b/\bar{p}}(x_b, Q^2)$$

- **Problem:** Need multidimensional phase space integration with cuts on e.g. E_T and η. $\Rightarrow$ takes hours/days of CPU time.
- **Solution:** Interpolate PDFs and α_S so that they can be factorised from $\hat{\sigma}_{ab}$. Replaces convolution with multiplication. Phase space integration only needs to be done once, with result stored in a grid for later use during a PDF fit.

FASTNLO: fast pQCD calculations for PDF fits

Grids corresponding to Tevatron and HERA kinematic cuts provided by FASTNLO[1] project.





[1] [Kluge, Rabbertz, Wobisch, hep-ph/0609285]

Treatment of jet data in MRST/MSTW analyses

MRST 2001–2006

- Fit six "pseudogluon" points at $Q^2 = 2000$ GeV^2 inferred from Tevatron Run I inclusive jet data.
- Comparison to actual jet data, calculated at LO with a K-factor, only made **after** the fit.

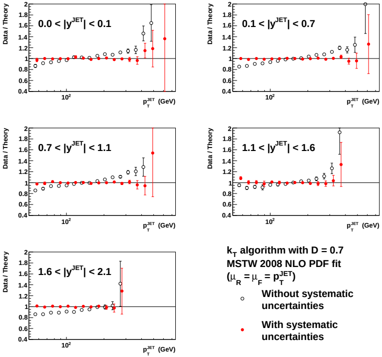
MSTW 2008

- Fit to Tevatron Run II and HERA DIS inclusive jet data.
- Complete treatment of correlated systematic errors.
- Use FASTNLO code to calculate NLO cross sections.
- Full NNLO $\hat{\sigma}_{ab}$ not yet known: include 2-loop threshold ($x_T = 2E_T/\sqrt{s}$) corrections[a] for Tevatron jet data at NNLO and exclude HERA DIS jet data.

[a] [Kidonakis, Owens, hep-ph/0007268]

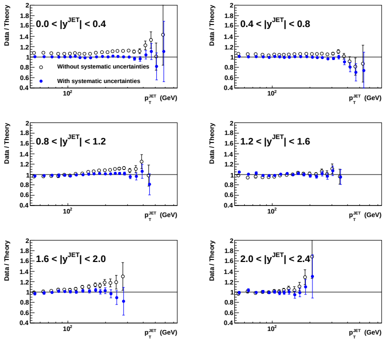
Description of Tevatron Run II inclusive jet data

CDF Run II inclusive jet data, χ^2 = 56 for 76 pts.



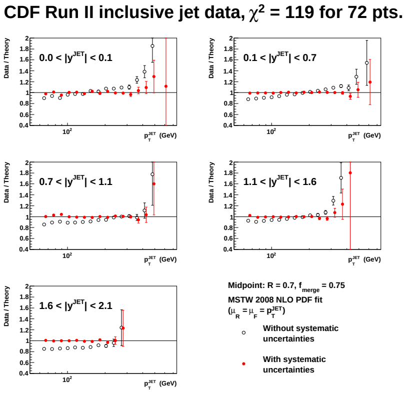
[hep-ex/0701051]

D∅ Run II inclusive jet data (cone, R = 0.7)
MSTW 2008 NLO PDF fit ($\mu_R = \mu_F = p_T^{JET}$), χ^2 = 114 for 110 pts.



[arXiv:0802.2400]

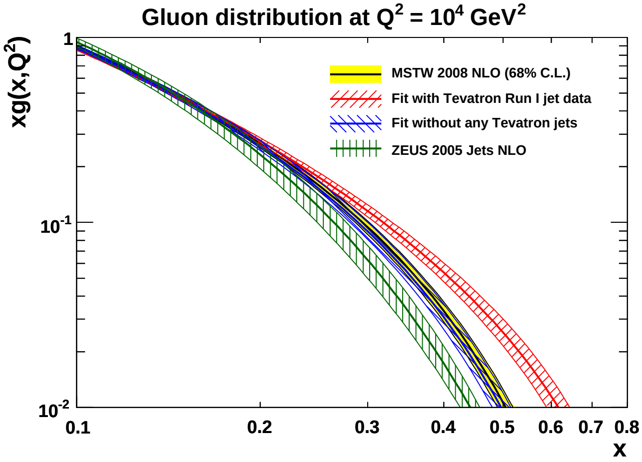
Dependence on jet clustering algorithm



[arXiv:0807.2204]

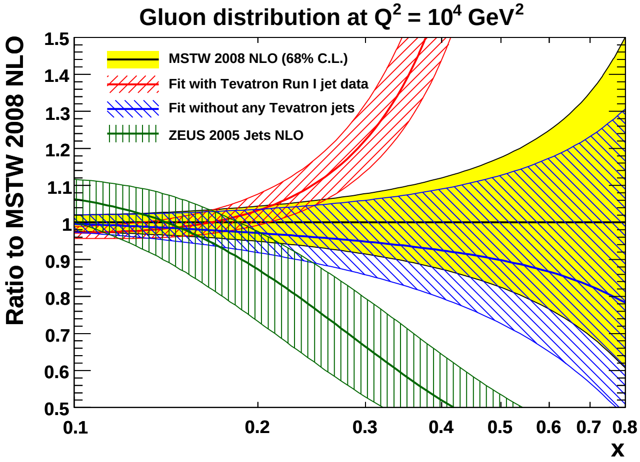
- Fit data from DØ (cone) and CDF (k_T).
- But also good description of CDF (cone): see plot.
- Replacing CDF (k_T) by CDF (cone) and refitting does not significantly change χ^2 or gluon distribution.
- $\Rightarrow$ All three data sets on inclusive jet production from Tevatron Run II are consistent, but cone algorithm disfavoured for theoretical reasons (lack of infrared safety).

Impact of Run II jet data on high-x gluon distribution



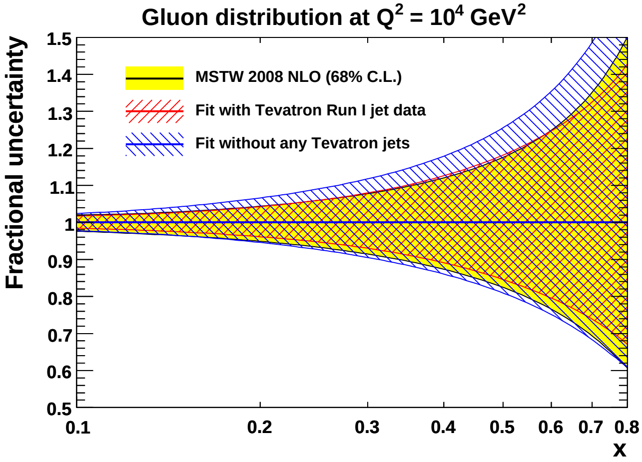
- Run II jet data prefer smaller gluon distribution at high x.

Impact of Run II jet data on high-x gluon distribution



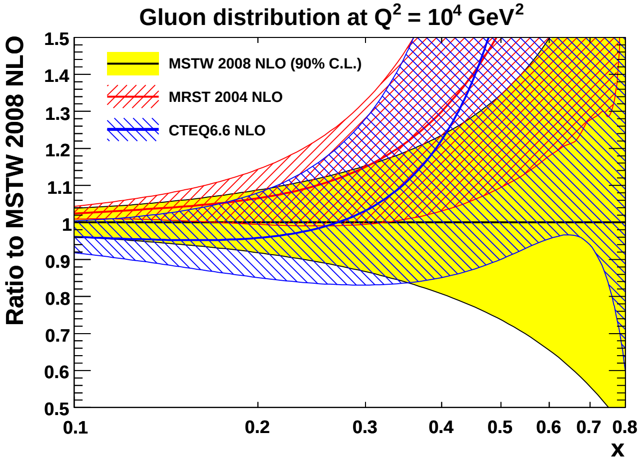
- ZEUS 2005 fit omits fixed-target data which constrain high-x gluon.

Impact of Run II jet data on high-x gluon distribution



- Errors not significantly reduced with inclusion of Run II data.

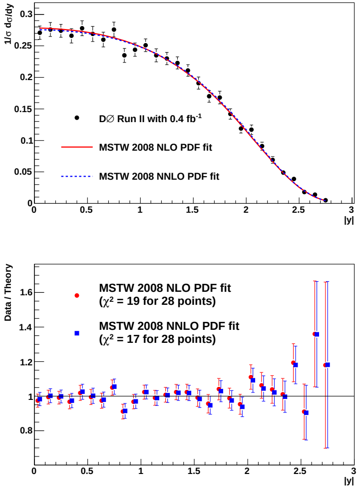
Impact of Run II jet data on high-x gluon distribution



- Smaller high-x gluon than previous MRST and CTEQ fits.

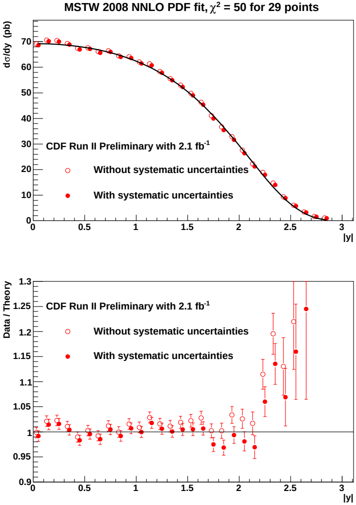
Z/γ^* rapidity distributions from Tevatron Run II

Z/γ* rapidity shape distribution from D∅



[hep-ex/0702025]

Z/γ* rapidity distribution from CDF



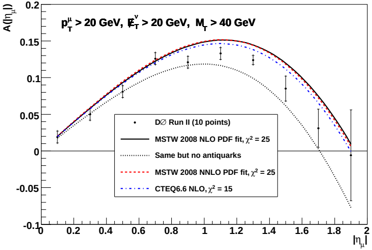
[CDF Preliminary, May 2008]

$W \to l\nu$ charge asymmetry from Tevatron Run II

$$A(\eta_l) = \frac{\mathrm{d}\sigma(l^+)/\mathrm{d}\eta_l - \mathrm{d}\sigma(l^-)/\mathrm{d}\eta_l}{\mathrm{d}\sigma(l^+)/\mathrm{d}\eta_l + \mathrm{d}\sigma(l^-)/\mathrm{d}\eta_l}$$

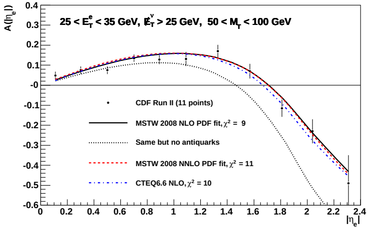
- Mainly constrains **down** quark.
- Antiquarks important at low E_T^l.

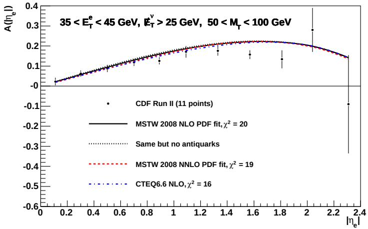
DØ data on lepton charge asymmetry from W→μν decays



[arXiv:0709.4254]

CDF data on lepton charge asymmetry from W→eν decays

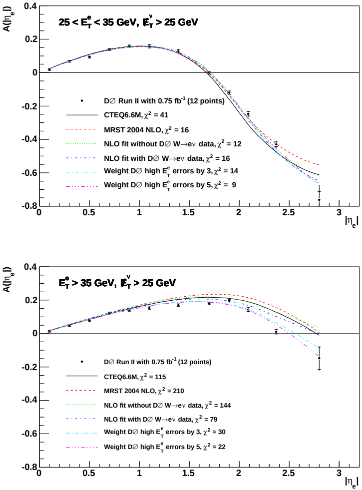




[hep-ex/0501023]

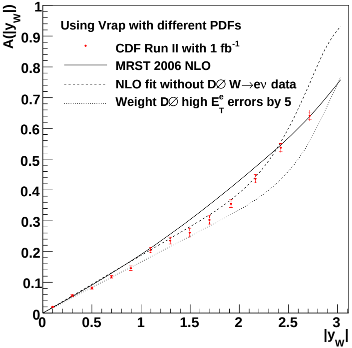
$W \to l\nu$ charge asymmetry from Tevatron Run II

DØ data on lepton charge asymmetry from W→eν decays



[arXiv:0807.3367]

- Problems fitting recent DØ data $\Rightarrow$ not included in final 2008 fit.
- Inconsistencies between DØ and CDF: under investigation.



[CDF Preliminary, August 2007]

Uncertainties in global PDF analysis

"Theoretical" errors

- *Examples:* input parameterisation form, neglected higher-order and higher-twist QCD corrections, electroweak corrections, choice of cuts, nuclear corrections, heavy flavour treatment.
- Difficult to quantify *a priori*. Often correction only known after an improved treatment/calculation is available.

"Experimental" errors

- If all the above sources of "theoretical" errors are fixed, how do we propagate the experimental uncertainties on the fitted data points through to the PDF uncertainties, or to derived quantities such as cross sections, ratios of cross sections, experimental acceptances (etc.)?
- Generally use the **Hessian** method.

Uncertainties in global PDF analysis

“Theoretical” errors

- *Examples:* input parameterisation form, neglected higher-order and higher-twist QCD corrections, electroweak corrections, choice of cuts, nuclear corrections, heavy flavour treatment.
- Difficult to quantify *a priori*. Often correction only known after an improved treatment/calculation is available.

“Experimental” errors

- If all the above sources of “theoretical” errors are fixed, how do we propagate the experimental uncertainties on the fitted data points through to the PDF uncertainties, or to derived quantities such as cross sections, ratios of cross sections, experimental acceptances (etc.)?
- Generally use the **Hessian** method.

Traditional propagation of experimental uncertainties

- Assume $\chi^2_{\rm global}$ is quadratic about the global minimum $\{a_i^0\}$:

$$\Delta\chi^2_{\rm global} \equiv \chi^2_{\rm global} - \chi^2_{\rm min} = \sum_{i,j} H_{ij}(a_i - a_i^0)(a_j - a_j^0),$$

where the Hessian matrix has components

$$H_{ij} = \frac{1}{2}\frac{\partial^2\,\chi^2_{\rm global}}{\partial a_i \partial a_j}\bigg|_{\rm min}$$

- Uncertainty on quantity $F(\{a_i\})$ from linear error propagation:

$$\Delta F = T\sqrt{\sum_{i,j}\frac{\partial F}{\partial a_i}C_{ij}\frac{\partial F}{\partial a_j}},$$

where $C \equiv H^{-1}$ is the covariance matrix, and $T = \sqrt{\Delta\chi^2_{\rm global}}$ is the tolerance for the required confidence interval.

Eigenvector PDF sets (pioneered by CTEQ)

- Convenient to diagonalise covariance (or Hessian) matrix:

$$\sum_j C_{ij} v_{jk} = \lambda_k v_{ik},$$

where λ_k is the kth eigenvalue and v_{ik} is the ith component of the kth orthonormal eigenvector ($k = 1, \ldots, N_{\rm parameters}$).

- Expand parameter displacements from minimum in basis of rescaled eigenvectors $e_{ik} \equiv \sqrt{\lambda_k} v_{ik}$:

$$a_i - a_i^0 = \sum_k e_{ik} z_k.$$

- Then can show, using the orthonormality of v_{ik}, that

$$\Delta\chi^2_{\rm global} = \sum_{i,j} H_{ij}(a_i - a_i^0)(a_j - a_j^0) = \sum_k z_k^2,$$

i.e. $\sum_k z_k^2 \le T^2$ is the interior of a hypersphere of radius T.

Eigenvector PDF sets (pioneered by CTEQ)

- Convenient to diagonalise covariance (or Hessian) matrix:

$$\sum_j C_{ij} v_{jk} = \lambda_k v_{ik},$$

where λ_k is the kth eigenvalue and v_{ik} is the ith component of the kth orthonormal eigenvector ($k = 1, \ldots, N_{\rm parameters}$).

- Expand parameter displacements from minimum in basis of rescaled eigenvectors $e_{ik} \equiv \sqrt{\lambda_k} v_{ik}$:

$$a_i - a_i^0 = \sum_k e_{ik} z_k.$$

- Then can show, using the orthonormality of v_{ik}, that

$$\Delta\chi^2_{\rm global} = \sum_{i,j} H_{ij}(a_i - a_i^0)(a_j - a_j^0) = \sum_k z_k^2,$$

i.e. $\sum_k z_k^2 \leq T^2$ is the interior of a hypersphere of radius T.

How to calculate uncertainties from eigenvector PDF sets?

- Fitting groups produce eigenvector PDF sets $S_k^{\pm}$ with parameters a_i shifted from the global minimum:

$$a_i(S_k^{\pm}) = a_i^0 \pm t\, e_{ik},$$

with t adjusted to give the desired $T = \sqrt{\Delta\chi^2_{\rm global}}$.

- Then users can calculate uncertainties on a quantity F with

$$\Delta F = \frac{1}{2}\sqrt{\sum_k \left[F(S_k^+) - F(S_k^-)\right]^2},$$

or to account for asymmetric errors (S_0 = central PDF set):

$$(\Delta F)_+ = \sqrt{\sum_k \left[\max(F(S_k^+) - F(S_0), F(S_k^-) - F(S_0), 0)\right]^2}$$

$$(\Delta F)_- = \sqrt{\sum_k \left[\max(F(S_0) - F(S_k^+), F(S_0) - F(S_k^-), 0)\right]^2}$$

How to calculate uncertainties from eigenvector PDF sets?

- Fitting groups produce eigenvector PDF sets $S_k^\pm$ with parameters a_i shifted from the global minimum:

$$a_i(S_k^\pm) = a_i^0 \pm t\, e_{ik},$$

with t adjusted to give the desired $T = \sqrt{\Delta\chi^2_{\rm global}}$.

- Then users can calculate uncertainties on a quantity F with

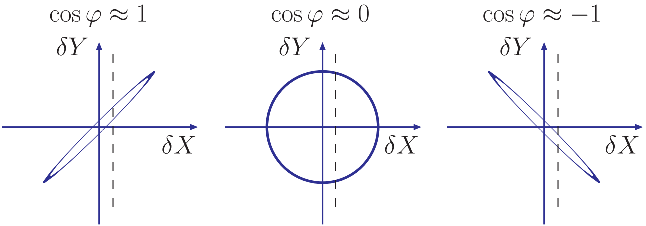
$$\Delta F = \frac{1}{2}\sqrt{\sum_k \left[F(S_k^+) - F(S_k^-)\right]^2},$$

or to account for asymmetric errors (S_0 = central PDF set):

$$(\Delta F)_+ = \sqrt{\sum_k \left[\max(F(S_k^+) - F(S_0), F(S_k^-) - F(S_0), 0)\right]^2}$$

$$(\Delta F)_- = \sqrt{\sum_k \left[\max(F(S_0) - F(S_k^+), F(S_0) - F(S_k^-), 0)\right]^2}$$

PDF uncertainties in 2-D: correlation cosines



- See e.g. CTEQ6.6 paper [Nadolsky *et al.*, arXiv:0802.0007].
- Define a correlation cosine between two quantities X and Y:

$$\cos\phi_{XY} = \frac{1}{4\,\Delta X\,\Delta Y}\sum_k \left[X(S_k^+) - X(S_k^-)\right]\left[Y(S_k^+) - Y(S_k^-)\right].$$

- Then tolerance ellipse in the X–Y plane is given by:

$$\delta X \equiv X - X(S_0) = \Delta X\,\cos\theta,$$
$$\delta Y \equiv Y - Y(S_0) = \Delta Y\,\cos(\theta + \phi_{XY}).$$

- Simple way to investigate PDF-induced (anti)correlations between different cross sections and PDF flavours.

PDF reweighting for Monte Carlo event generators

- To calculate PDF uncertainties on an observable, need to evaluate separately with ∼40 eigenvector PDF sets.
- In many cases not practical to ensure that statistical fluctuations (in Monte Carlo integration) are negligible compared to small differences between PDF sets.
- Instead, can reweight the hard-scatter in already-generated Monte Carlo events by multiplying by a factor:

$$\frac{f_{a/p}^{(i)}(x_1, Q^2)\, f_{b/p}^{(i)}(x_2, Q^2)}{f_{a/p}^{(0)}(x_1, Q^2)\, f_{b/p}^{(0)}(x_2, Q^2)},$$

  to go from the central set (0) to an alternative set (i).
- Neglects PDF dependence of Sudakov factor in parton shower, but only ratio of PDFs in Sudakov factor, so PDF uncertainty generally small [S. Gieseke, hep-ph/0412342].
- MCFM can store separate weights and calculate PDF uncertainties on distributions automatically.

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

Parameter-fitting criterion

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

Hypothesis-testing criterion (proposed by CTEQ)

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

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

Parameter-fitting criterion

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

Hypothesis-testing criterion (proposed by CTEQ)

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

Choice of tolerance by MRST [hep-ph/0211080]

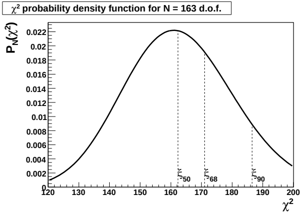
"We estimate $\Delta\chi^2 = 50$ *to be a conservative uncertainty (perhaps of the order of a* 90% *confidence level or a little less than* 2σ*) due to the observation that* **an increase of 50 in the global** χ^2, *which has a value* $\chi^2 = 2328$ *for* 2097 *data points,* **usually signifies that the fit to one or more data sets is becoming unacceptably poor***. We find that* **an increase** $\Delta\chi^2$ **of 100 normally means that some data sets are very badly described** *by the theory."*

- Fairly qualitative statements.
- $\Rightarrow$ Study more quantitatively in new MSTW analysis using same procedure applied in original CTEQ6 analysis.

Determination of 90% C.L. region following CTEQ6

- Define 90% C.L. region for each data set n (with N data points) as

$$\chi_n^2 < \left(\frac{\chi_{n,0}^2}{\xi_{50}}\right)\xi_{90}$$



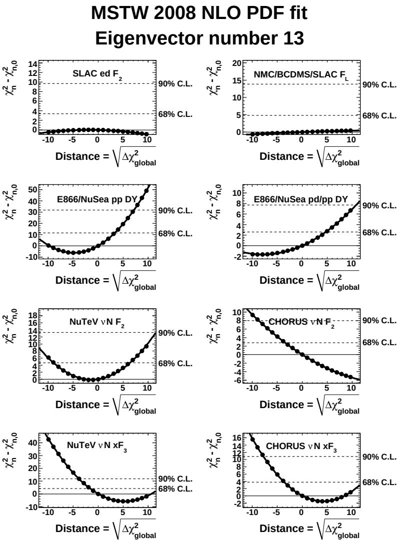
- ξ_{90} is the 90th percentile of the χ^2-distribution with N d.o.f., i.e.

$$\int_0^{\xi_{90}} \mathrm{d}\chi^2\ P_N(\chi^2) = 0.90,$$

where the probability density function is $\quad P_N(\chi^2) = \frac{(\chi^2)^{N/2-1}\,\mathrm{e}^{-\chi^2/2}}{2^{N/2}\,\Gamma(N/2)}.$

- $\xi_{50} \simeq N$ is the most probable value of the χ^2-distribution.
- $\chi_{n,0}^2$ for data set n is evaluated at the **global** minimum.
- **Rescale** by a factor $\chi_{n,0}^2/\xi_{50}$ since this often deviates from 1.

Change in χ_n^2 when moving along eigenvector number 13



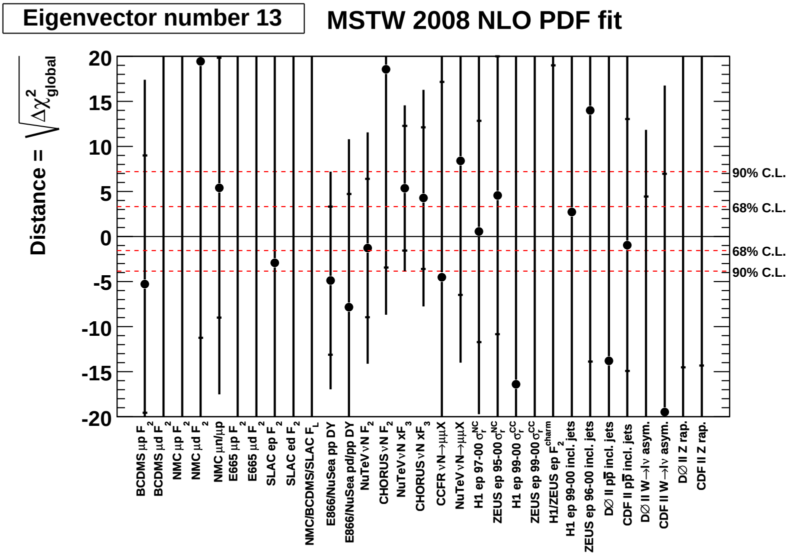
1. Plot difference in χ_n^2 compared to value at global minimum ($\chi_{n,0}^2$) when moving along each eigenvector direction.
2. Define 90% C.L. region by:

$$\chi_n^2 < \left(\frac{\chi_{n,0}^2}{\xi_{50}}\right) \xi_{90}$$

and similarly for 68% C.L.

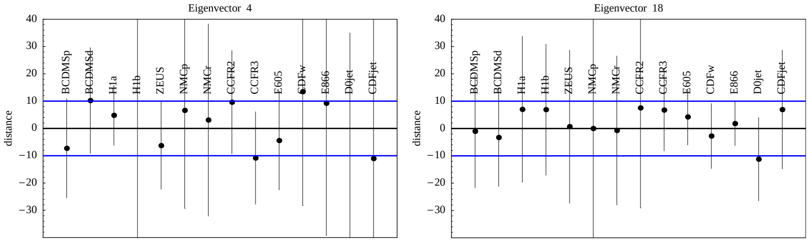
3. Record values of $\sqrt{\Delta\chi^2_{\rm global}}$ for which χ_n^2 is minimised, together with the 90% and 68% C.L. limits.

Determination of tolerance for eigenvector number 13



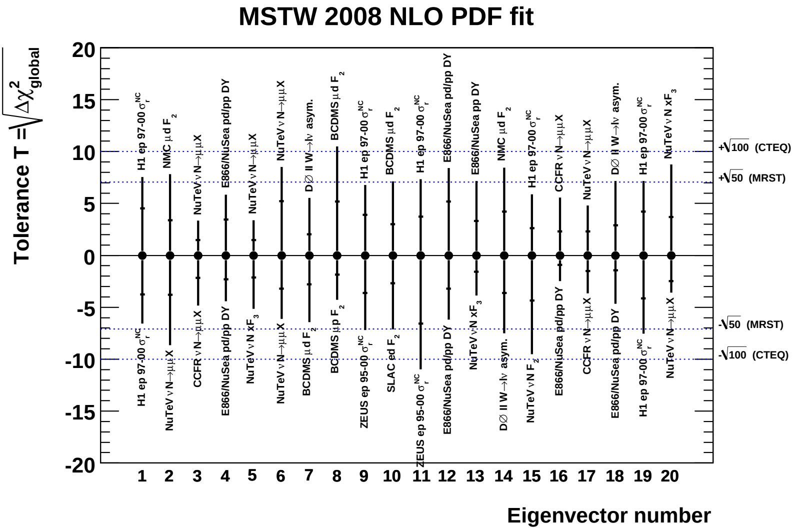
Choice of tolerance by CTEQ [hep-ph/0201195]

- Corresponding plots from the original CTEQ analysis.

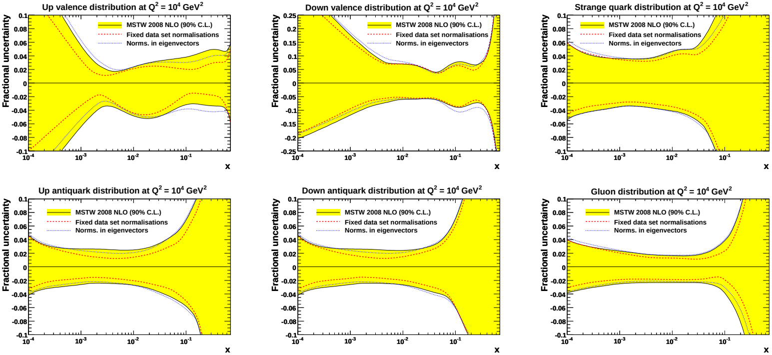


- A rough "average" over all eigenvectors gives $T = 10$ ...
- ... But $T = 10$ exceeds the 90% C.L. limits of some data sets.

Dynamic tolerance: different for each eigenvector

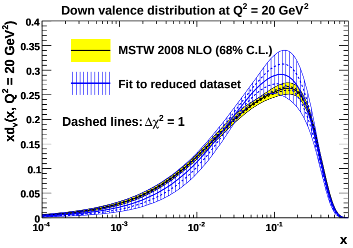


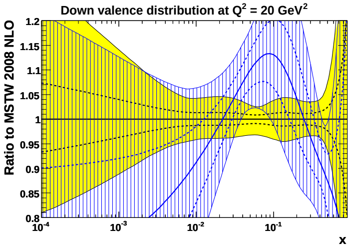
Effect of free data set normalisations on PDF uncertainties



- Float relative normalisations of individual data sets for best fit.
- Should **refit** these normalisations for each of the eigenvector PDF sets: wasn't done in previous MRST and CTEQ fits.
- Increase in PDF uncertainty from free data set normalisations.

Test of dynamic tolerance: fit to reduced dataset





- Fit to reduced dataset comprising 589 DIS data points, cf. 2699 data points in global fit.
- Errors given by $T^2 = 1$ don't overlap $\Rightarrow$ inconsistent data sets included in global fit.
- **Dynamic tolerance** $T^2 > 1$ (partially) accommodates inconsistent data sets.

Alternative approach: NNPDF Collaboration

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

MSTW approach [arXiv:0901.0002]

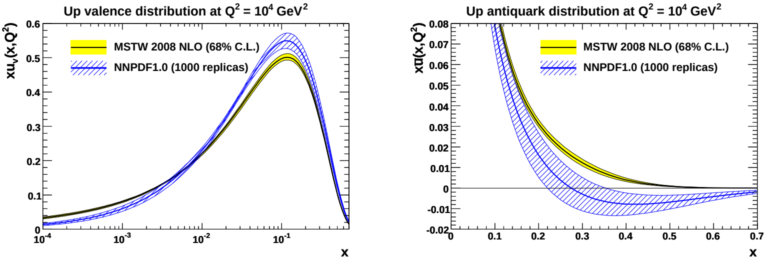
Parameterisation	$xf_{a/p} \sim A_a x^{\Delta_a}(1-x)^{\eta_a}(1+\epsilon_a\sqrt{x}+\gamma_a x)$
Minimisation	Non-linear least-squares (Marquardt method)
Error propagation	Hessian method with dynamical tolerance
Application	Use best-fit and 40 eigenvector PDF sets

NNPDF approach [arXiv:0808.1231]

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

First results from NNPDF1.0 [arXiv:0808.1231]

- Fit restricted set of only DIS structure function data (SLAC, BCDMS, NMC, H1, ZEUS, CHORUS).
- Inadequate treatment of heavy quarks (zero-mass variable flavour number scheme, see later).



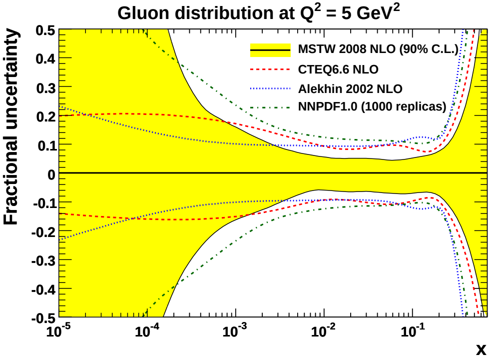
- Up valence: relative data set normalisations fitted by MSTW.
- Up antiquark: NNPDF1.0 negative by ~ 2-σ, no Drell–Yan.

Parameterisation dependence of low-x gluon

- PDFs lose probabilistic interpretation beyond LO.
- Negative small-x gluon distribution preferred at low scales.
- MRST/MSTW parameterise as:

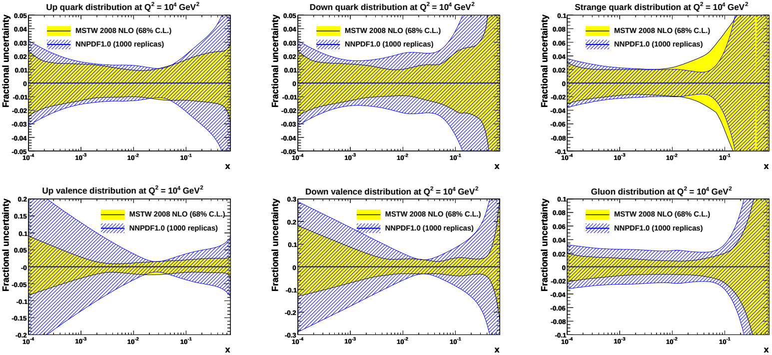
$$xg(x, Q_0^2) = xg_1(x, Q_0^2) + xg_2(x, Q_0^2) \sim A_g\, x^{\delta_g} + A_{g'}\, x^{\delta_{g'}}$$

$$\Rightarrow \Delta g(x, Q_0^2) \sim \pm g_1(x, Q_0^2)\, \Delta\delta_g \ln(1/x) \pm g_2(x, Q_0^2)\, \Delta\delta_{g'} \ln(1/x)$$



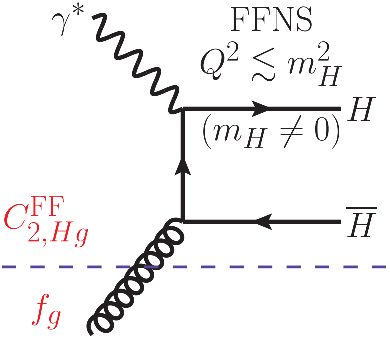
- Other groups (CTEQ, Alekhin) parameterise with a valence-like $xg(x, Q_0^2) \sim x^{\delta_g}$: less freedom at small x.

Uncertainties for MSTW 2008 and NNPDF1.0



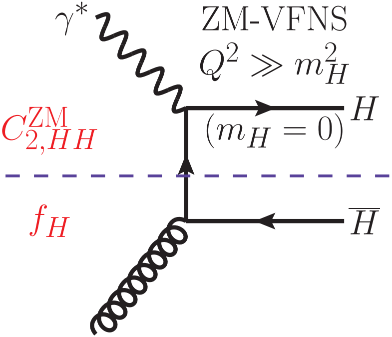
- NNPDF1.0 has fixed $s = \bar{s} = (\bar{u} + \bar{d})/4$ at $Q_0^2 = 2$ GeV^2.
- NNPDF1.1 [arXiv:0811.2288]: free strangeness but no νN dimuon data to constrain, so huge PDF uncertainties.
- **Conclusion:** NNPDF approach looks promising, but PDFs not yet directly comparable to those from standard approach.

Heavy quark contribution to DIS structure function F_2



Fixed flavour number scheme

- No heavy quark PDF.
- Includes $\mathcal{O}(m_H^2/Q^2)$ terms.
- No resummation of $\alpha_S \ln(Q^2/m_H^2)$ terms.



Zero-mass variable flavour number scheme

- Use heavy quark PDF.
- Mass dependence neglected.
- Resums $\alpha_S \ln(Q^2/m_H^2)$ terms similar to light quarks.

General-mass variable flavour number scheme (GM-VFNS)

Recent review by R. Thorne and W.-K. Tung [arXiv:0809.0714]

- Interpolate between two well-defined regions.
- FFNS for $Q^2 \leq m_H^2$, ZM-VFNS for $Q^2 \gg m_H^2$.
- Some freedom/ambiguity in swapping $\mathcal{O}(m_H^2/Q^2)$ terms between different orders for $Q^2 > m_H^2$.
- Different choices also possible in ordering of expansion:

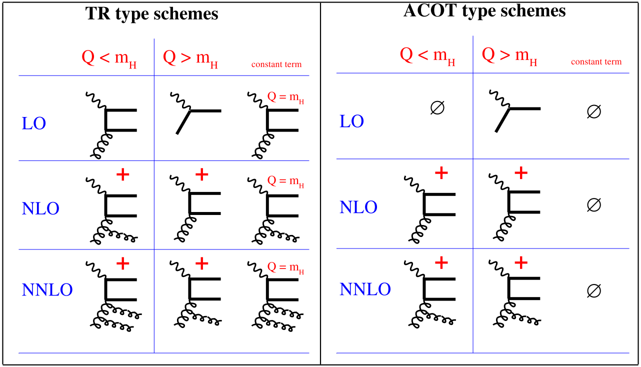


Figure from F. Olness and I. Schienbein [arXiv:0812.3371]

Impact of GM-VFNS on W and Z cross sections at LHC

- W and Z cross sections at the LHC are sensitive to light sea quark PDFs at $x \sim 0.006$ (for $y = 0$), determined mainly from HERA F_2 data.

CTEQ6.1 NLO (ZM-VFNS) $\rightarrow$ CTEQ6.5 NLO (GM-VFNS)

- 8% increase in W and Z cross sections at LHC.

MRST 2004 $\rightarrow$ MRST 2006 [arXiv:0706.0459]

- The MRST group have used a GM-VFNS since 1998.
- At NNLO, PDFs are discontinuous at $Q^2 = m_H^2$, i.e. $f_{j/p}^{n_f+1}(x, m_H^2) = f_{j/p}^{n_f}(x, m_H^2) + \alpha_S^2 \sum_k A_{jk}(x) \otimes f_{k/p}^{n_f}(x, m_H^2)$.
- Neglected in MRST NNLO fits prior to 2006.
- GM-VFNS at NNLO worked out by R. Thorne [hep-ph/0601245]. Changes needed also at LO and NLO for consistency.
- 2004 NNLO $\rightarrow$ 2006 NNLO: 6% increase in W and Z cross sections.
- 2004 NLO $\rightarrow$ 2006 NLO: 3% increase in W and Z cross sections.
- Increase at NNLO is a **correction**, but at NLO is a **scheme change**.

$W \equiv W^+ + W^-$ and Z total cross sections at LHC

LHC, $\sqrt{s} = 10$ TeV	$B_{l\nu} \cdot \sigma_W$ (nb)	$B_{l^+l^-} \cdot \sigma_Z$ (nb)	R_{WZ}
MSTW 2008 LO	$12.57^{+0.13}_{-0.19}$ $\left(^{+1.1\%}_{-1.5\%}\right)$	$1.163^{+0.011}_{-0.017}$ $\left(^{+1.0\%}_{-1.5\%}\right)$	$10.81^{+0.02}_{-0.02}$ $\left(^{+0.2\%}_{-0.2\%}\right)$
MSTW 2008 NLO	$14.92^{+0.31}_{-0.24}$ $\left(^{+2.1\%}_{-1.6\%}\right)$	$1.390^{+0.029}_{-0.022}$ $\left(^{+2.1\%}_{-1.5\%}\right)$	$10.73^{+0.02}_{-0.02}$ $\left(^{+0.2\%}_{-0.2\%}\right)$
MSTW 2008 NNLO	$15.35^{+0.26}_{-0.25}$ $\left(^{+1.7\%}_{-1.6\%}\right)$	$1.429^{+0.024}_{-0.022}$ $\left(^{+1.7\%}_{-1.6\%}\right)$	$10.74^{+0.02}_{-0.02}$ $\left(^{+0.2\%}_{-0.2\%}\right)$

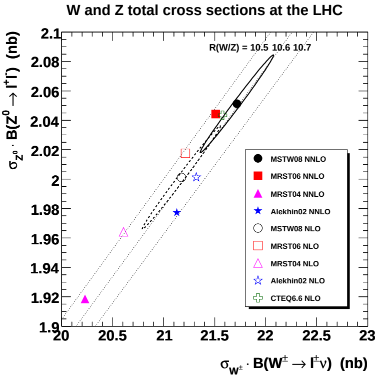
LHC, $\sqrt{s} = 14$ TeV	$B_{l\nu} \cdot \sigma_W$ (nb)	$B_{l^+l^-} \cdot \sigma_Z$ (nb)	R_{WZ}
MSTW 2008 LO	$18.51^{+0.22}_{-0.32}$ $\left(^{+1.2\%}_{-1.7\%}\right)$	$1.736^{+0.019}_{-0.028}$ $\left(^{+1.1\%}_{-1.6\%}\right)$	$10.66^{+0.02}_{-0.02}$ $\left(^{+0.2\%}_{-0.2\%}\right)$
MSTW 2008 NLO	$21.17^{+0.42}_{-0.36}$ $\left(^{+2.0\%}_{-1.7\%}\right)$	$2.001^{+0.040}_{-0.032}$ $\left(^{+2.0\%}_{-1.6\%}\right)$	$10.58^{+0.02}_{-0.02}$ $\left(^{+0.2\%}_{-0.2\%}\right)$
MSTW 2008 NNLO	$21.72^{+0.36}_{-0.36}$ $\left(^{+1.7\%}_{-1.7\%}\right)$	$2.051^{+0.035}_{-0.033}$ $\left(^{+1.7\%}_{-1.6\%}\right)$	$10.59^{+0.02}_{-0.03}$ $\left(^{+0.2\%}_{-0.3\%}\right)$

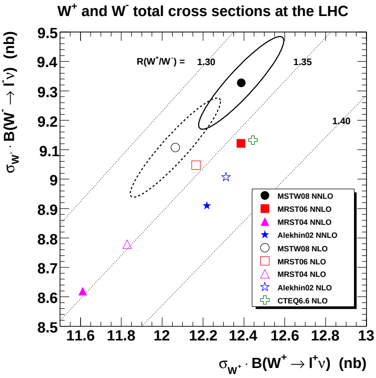
LHC, $\sqrt{s} = 10$ TeV	$B_{l\nu} \cdot \sigma_{W^+}$ (nb)	$B_{l\nu} \cdot \sigma_{W^-}$ (nb)	$R_\pm$
MSTW 2008 LO	$7.35^{+0.08}_{-0.12}$ $\left(^{+1.1\%}_{-1.6\%}\right)$	$5.22^{+0.06}_{-0.09}$ $\left(^{+1.1\%}_{-1.7\%}\right)$	$1.408^{+0.015}_{-0.012}$ $\left(^{+1.0\%}_{-0.8\%}\right)$
MSTW 2008 NLO	$8.62^{+0.18}_{-0.14}$ $\left(^{+2.1\%}_{-1.7\%}\right)$	$6.30^{+0.14}_{-0.11}$ $\left(^{+2.2\%}_{-1.7\%}\right)$	$1.367^{+0.012}_{-0.010}$ $\left(^{+0.9\%}_{-0.7\%}\right)$
MSTW 2008 NNLO	$8.88^{+0.15}_{-0.15}$ $\left(^{+1.7\%}_{-1.6\%}\right)$	$6.47^{+0.11}_{-0.11}$ $\left(^{+1.7\%}_{-1.6\%}\right)$	$1.373^{+0.012}_{-0.010}$ $\left(^{+0.8\%}_{-0.7\%}\right)$

LHC, $\sqrt{s} = 14$ TeV	$B_{l\nu} \cdot \sigma_{W^+}$ (nb)	$B_{l\nu} \cdot \sigma_{W^-}$ (nb)	$R_\pm$
MSTW 2008 LO	$10.69^{+0.14}_{-0.19}$ $\left(^{+1.3\%}_{-1.8\%}\right)$	$7.83^{+0.10}_{-0.14}$ $\left(^{+1.2\%}_{-1.8\%}\right)$	$1.366^{+0.013}_{-0.010}$ $\left(^{+0.9\%}_{-0.8\%}\right)$
MSTW 2008 NLO	$12.06^{+0.24}_{-0.21}$ $\left(^{+2.0\%}_{-1.8\%}\right)$	$9.11^{+0.19}_{-0.16}$ $\left(^{+1.2\%}_{-1.6\%}\right)$	$1.325^{+0.011}_{-0.009}$ $\left(^{+0.8\%}_{-0.7\%}\right)$
MSTW 2008 NNLO	$12.39^{+0.22}_{-0.21}$ $\left(^{+1.8\%}_{-1.7\%}\right)$	$9.33^{+0.16}_{-0.16}$ $\left(^{+1.7\%}_{-1.7\%}\right)$	$1.328^{+0.011}_{-0.009}$ $\left(^{+0.8\%}_{-0.7\%}\right)$

$W \equiv W^+ + W^-$ and Z total cross sections at LHC

(Error ellipses use 68% C.L. PDFs $\Rightarrow$ less than 68% C.L. in 2-D.)





- W and Z highly correlated: $\cos\phi_{WZ} \approx 0.99$.
- Little change in going from MRST 2006 $\to$ MSTW 2008.

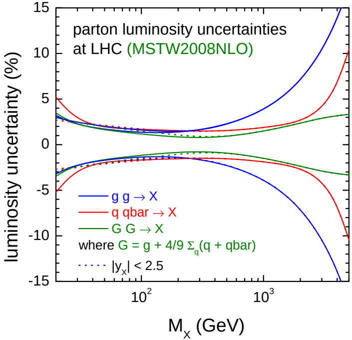
- W^+ and W^- less correlated: $\cos\phi_{W^+W^-} \approx 0.91$.
- Decrease in $R(W^+/W^-) \sim u/d$ from MRST 2006 $\to$ MSTW 2008.

PDF uncertainties in parton luminosity functions

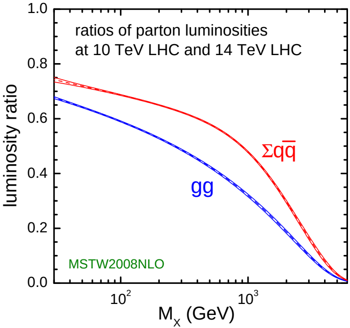
If $\hat{\sigma}_{ab} = C_X\,\delta\left(\hat{s} - M_X^2\right)$, with $\hat{s} = x_a\,x_b\,s$, then

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

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



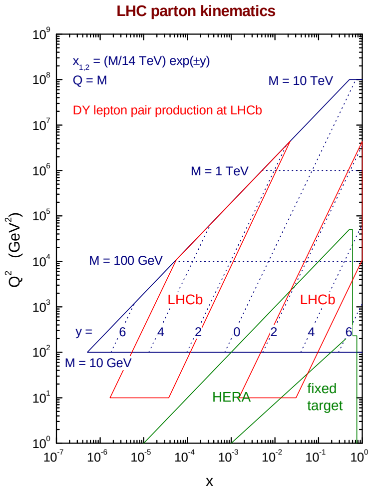
[Plots by J. Stirling]

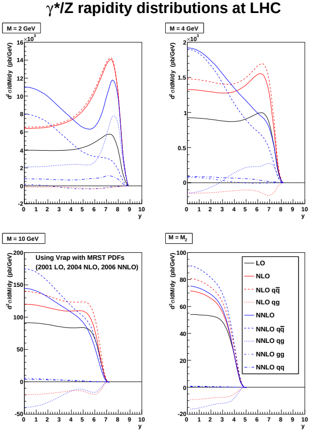


- $\sim 0.3\%$ PDF uncertainty for ratio $\sigma_{W,Z}(10\,\mathrm{TeV})/\sigma_{W,Z}(14\,\mathrm{TeV})$

Precision measurements at high rapidity from LHCb

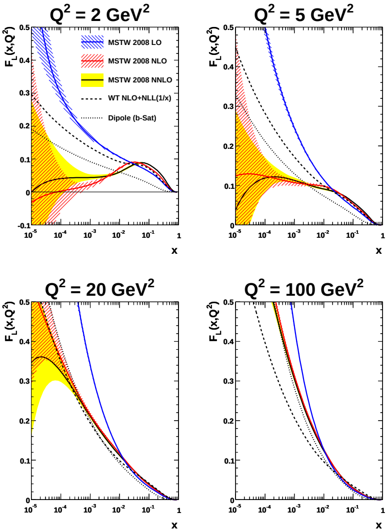
[MSTW, arXiv:0808.1847; R. McNulty, arXiv:0810.2550]

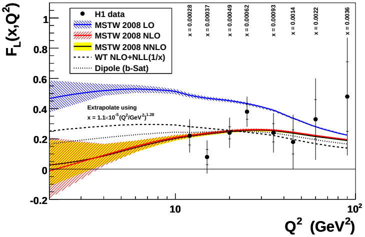




- Low-mass Drell–Yan perturbatively unstable at high rapidities.

Longitudinal structure function at HERA



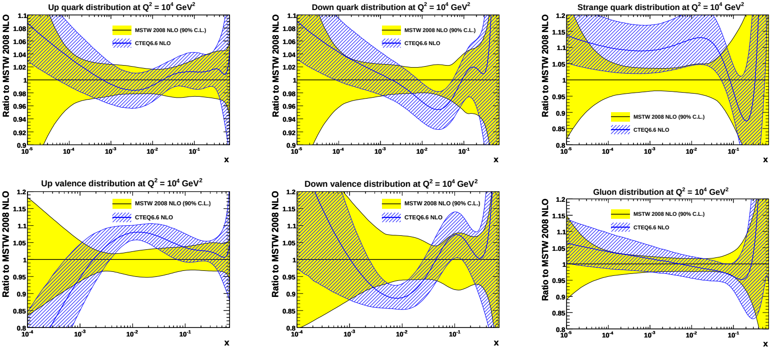


- Small-x resummation stabilises result for F_L.
- First analytic results for resummed Drell–Yan coefficient functions by S. Marzani and R. Ball [arXiv:0812.3602].

Summary

- **P**arton **D**istribution **F**unctions (**PDF**s) are a non-negotiable input to all theory predictions at hadron colliders.
- **MSTW 2008** (LO, NLO, NNLO) PDF fits [arXiv:0901.0002] are the most comprehensive to date: supersede MRST sets.
- **First** PDF fits available to include Tevatron Run II data: changes to high-x gluon and down quark distributions.
- **Improved** "dynamic tolerance" controlling propagation of experimental errors through to PDF uncertainties ...
- ... **But** there are sometimes uncertainties in PDFs which are not quantifiable *a priori*, e.g. choice of data to fit (Tevatron Run I vs. Run II jets), details of heavy quark treatment, ..., and only realised later with improved data or calculations.
- ***W* and *Z* total cross sections at LHC** have NNLO PDF uncertainties from data fitted of around 2%, but uncertainties largely cancel in **ratios** of total cross sections: $\sigma(W)/\sigma(Z)$, $\sigma(W^+)/\sigma(W^-)$, $\sigma(10\,\mathrm{TeV})/\sigma(14\,\mathrm{TeV})$.

Comparison to CTEQ6.6 NLO



Comparison to MRST 2006 NNLO