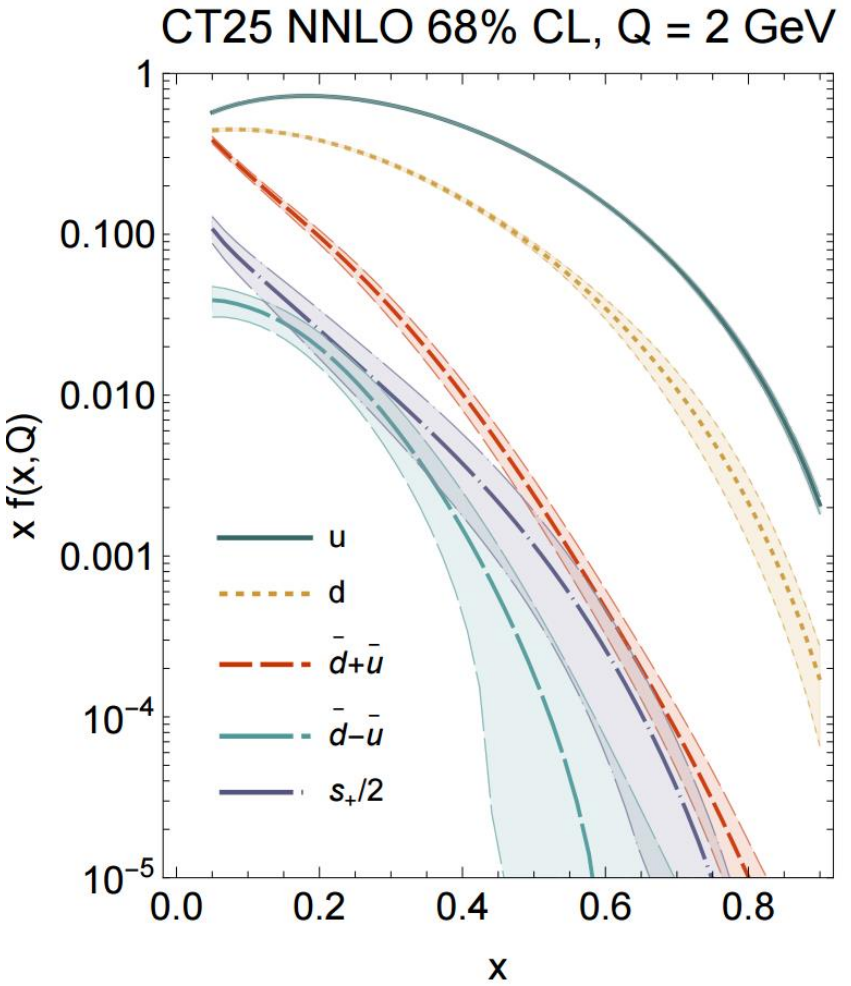
# Parton Distribution Functions in the EIC era

## Lecture 1

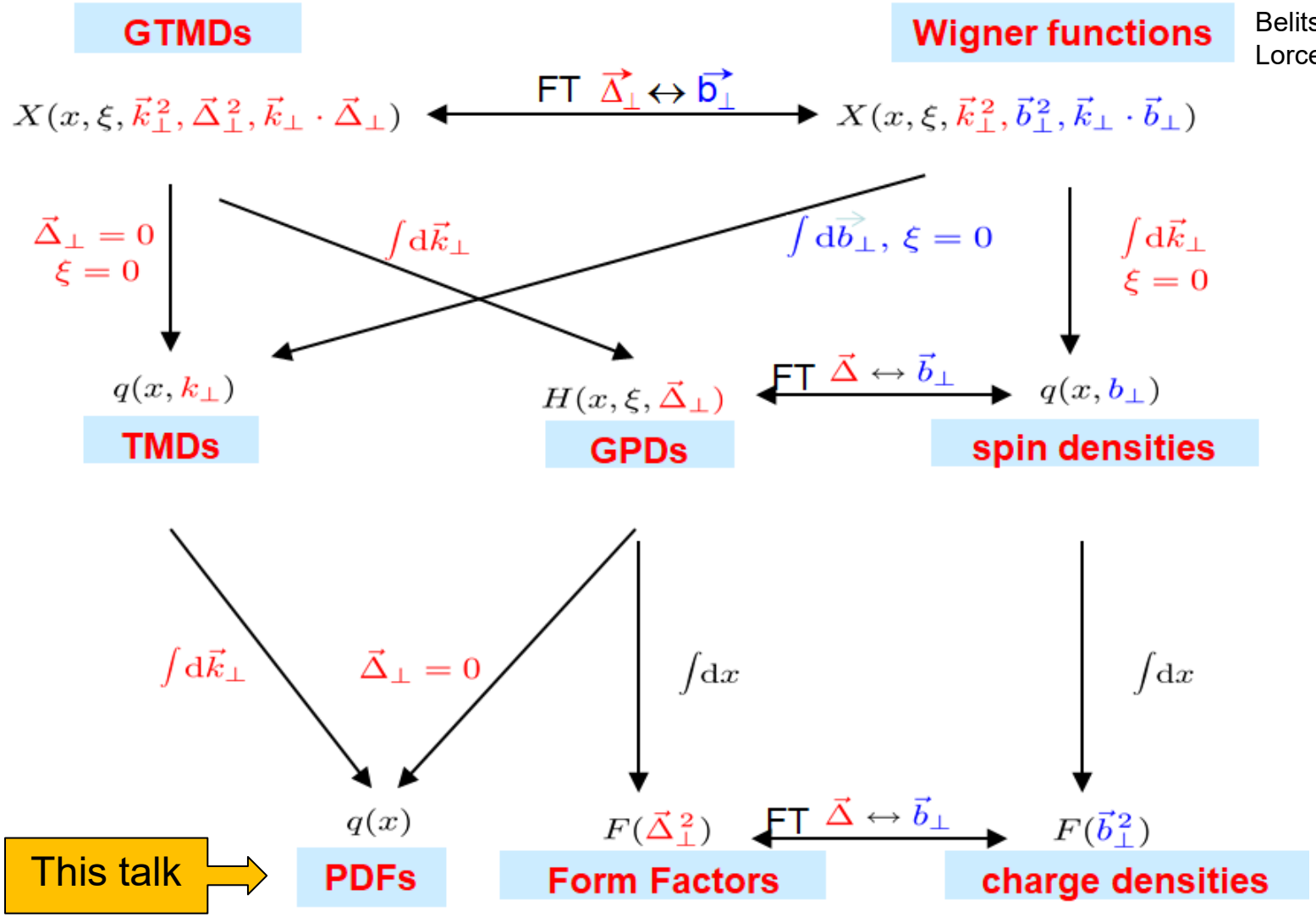
**Pavel Nadolsky**
**Michigan State University**
**CTEQ-TEA (Tung Et Al.) group**



# The inner world of a hadron


A short-distance probe (virtual photon, heavy boson, gluon) resolves increasingly small structures inside the nucleon.

# Nonperturbative functions for hadron structure



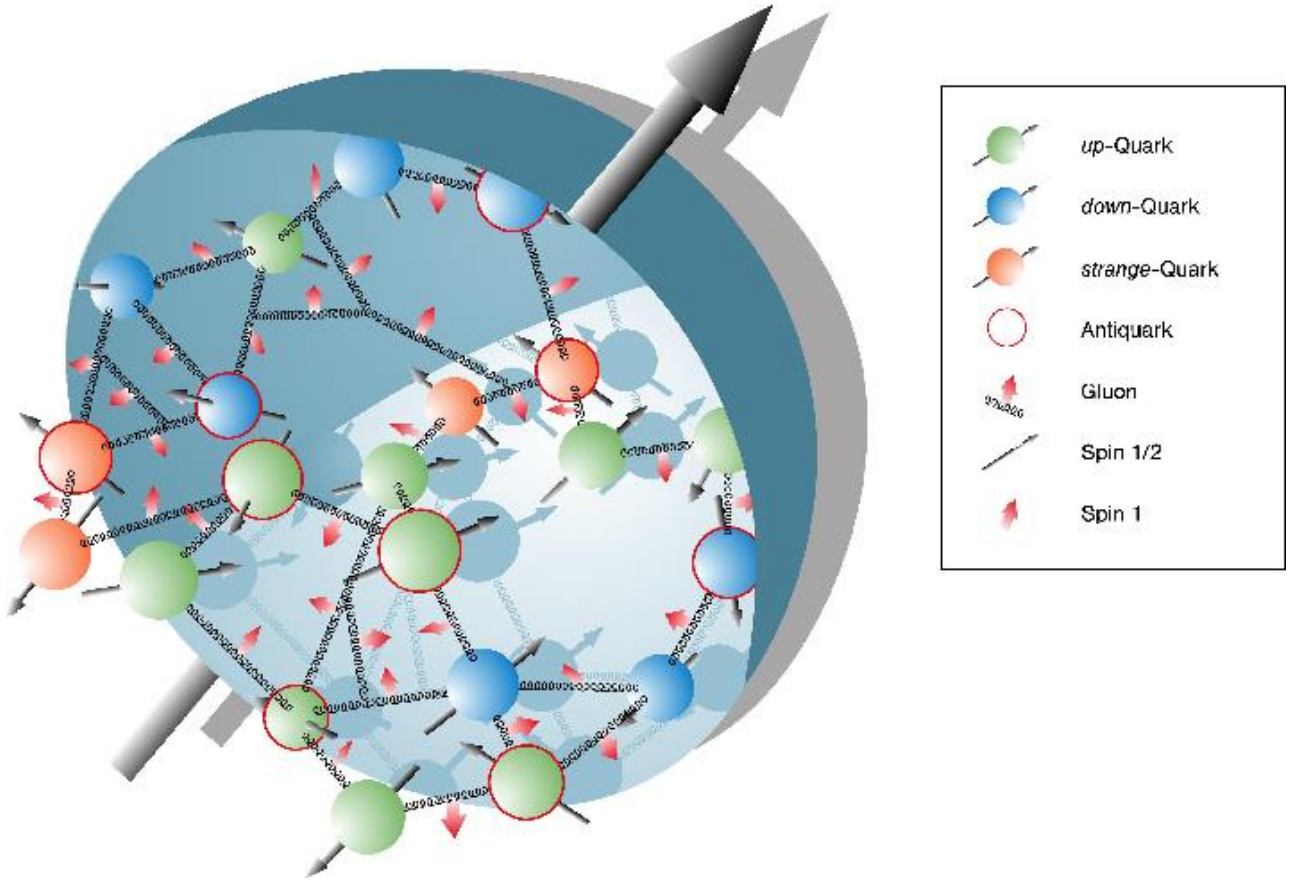

Belitsky, Ji,Yuan (2003)
Lorce, Pasquini (2011)

Figure: B. Pasquini, C. Lorcé

# Collinear PDFs: “the simplest” in QCD theory, the richest in information theory

Collinear PDFs $f_{a/p}(x, Q)$ are precisely measured nonperturbative functions describing the hadron structure at energy (resolution) scales $Q > 1$ GeV

Many considerations discussed here apply to spin-dependent and nuclear PDFs.

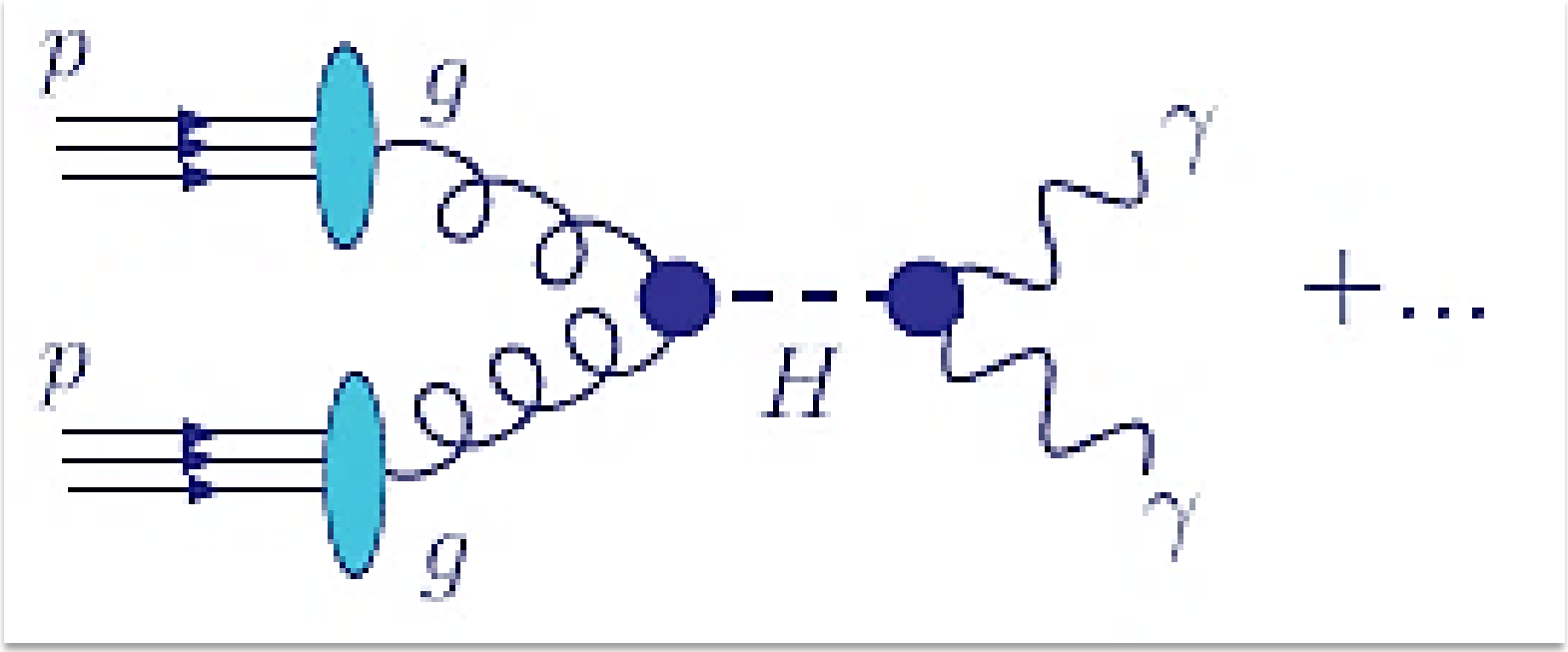


# 1. Collinear QCD factorization

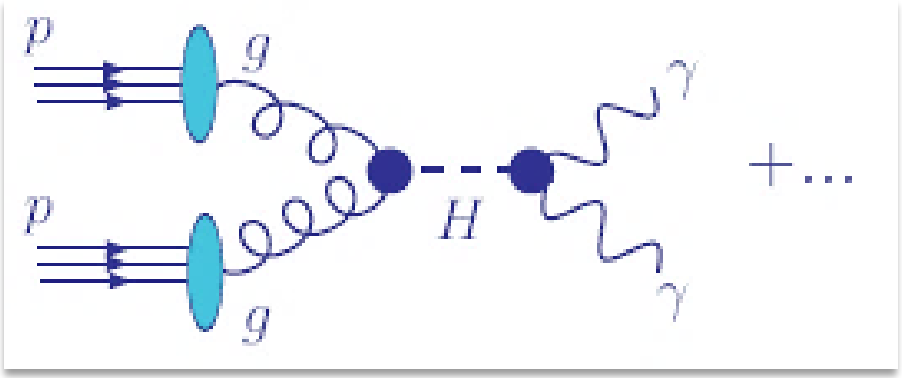
# PDFs and QCD factorization

$pp \to (\mathrm{H}^0 \to \gamma\gamma)X$: Higgs boson production and diphoton decay at the LHC

Feynman diagram at the leading $\alpha_s$ order, can be extended to higher orders



# Collinear factorization of LHC cross sections

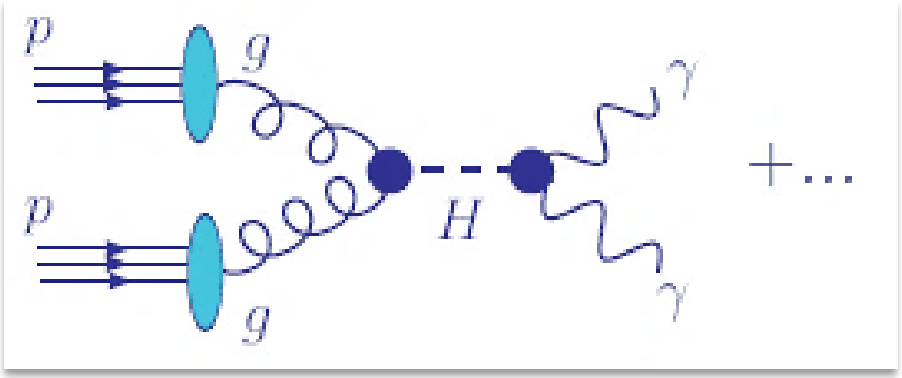



$$\sigma_{pp\to H\to\gamma\gamma X}(Q) = \sum_{a,b=g,q,\bar{q}} \int_0^1 d\xi_a \int_0^1 d\xi_b \hat{\sigma}_{ab\to H\to\gamma\gamma}\left(\frac{x_a}{\xi_a}, \frac{x_b}{\xi_b}, \frac{Q}{\mu_R}, \frac{Q}{\mu_F}; \alpha_s(\mu_R)\right)$$
$$\times f_a(\xi_a, \mu_F) f_b(\xi_b, \mu_F) + O\left(\frac{\Lambda_{QCD}^2}{Q^2}\right)$$

$\hat{\sigma}_{ab\to H\to\gamma\gamma}$ is the perturbative hard cross section; computed order-by-order in $\alpha_s(\mu_R)$

**$f_a(x, \mu_F)$ is the nonperturbative distribution for parton $a$ with momentum fraction $x$, at scale $\mu_F$**

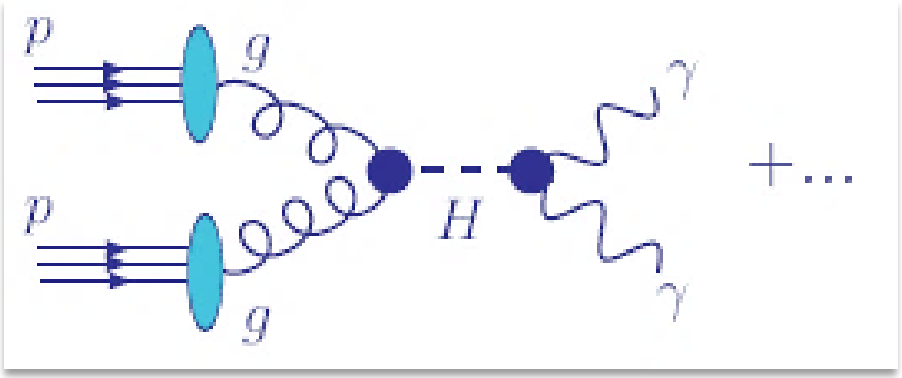
# Collinear factorization of LHC cross sections




$$\sigma_{pp\to H\to\gamma\gamma X}(Q) = \sum_{a,b=g,q,\bar{q}} \int_0^1 d\xi_a \int_0^1 d\xi_b \hat{\sigma}_{ab\to H\to\gamma\gamma}\left(\frac{x_a}{\xi_a}, \frac{x_b}{\xi_b}, \frac{Q}{\mu_R}, \frac{Q}{\mu_F}; \alpha_s(\mu_R)\right)$$
$$\times f_a(\xi_a, \mu_F) f_b(\xi_b, \mu_F) + O\left(\frac{\Lambda_{QCD}^2}{Q^2}\right)$$

- $Q^2 = (k_3 + k_4)^2$, $x_{a,b} = (Q/\sqrt{s})e^{\pm y}$ —measurable quantities (not too small)
- $\xi_a$, $\xi_b$ are partonic momentum fractions (integrated over)
- $\mu_F$ is a factorization scale (=renormalization scale $\mu_R$ from now on)

# Collinear factorization of LHC cross sections




$$\sigma_{pp\to H\to\gamma\gamma X}(Q) = \sum_{a,b=g,q,\bar{q}} \int_0^1 d\xi_a \int_0^1 d\xi_b \hat{\sigma}_{ab\to H\to\gamma\gamma}\left(\frac{x_a}{\xi_a}, \frac{x_b}{\xi_b}, \frac{Q}{\mu_R}, \frac{Q}{\mu_F}; \alpha_s(\mu_R)\right)$$
$$\times f_a(\xi_a, \mu_F) f_b(\xi_b, \mu_F) + O\left(\frac{\Lambda_{QCD}^2}{Q^2}\right)$$

- $\mu_{\mathrm{R}} = \mu_F \equiv \mu$ is naturally set to be of order $Q$
- Factorization holds up to terms of order $\Lambda_{QCD}^2/Q^2$

**Purpose of this arrangement:**

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

# Operator definitions for PDFs

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

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

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

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

To prove factorization, one must show that  correctly captures higher-order contributions for the considered observable

This condition can be violated for multi-scale observables

(e.g., DIS or Drell-Yan process at $x \sim Q/\sqrt{s} \ll 1$)

# Operator definitions for PDFs

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

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

The exact form of $f_{a/p}$ is not known; but its $\mu$ dependence is described by **D**okshitzer-**G**ribov-**L**ipatov-**A**ltarelli-**P**arisi (**DGLAP**) equations

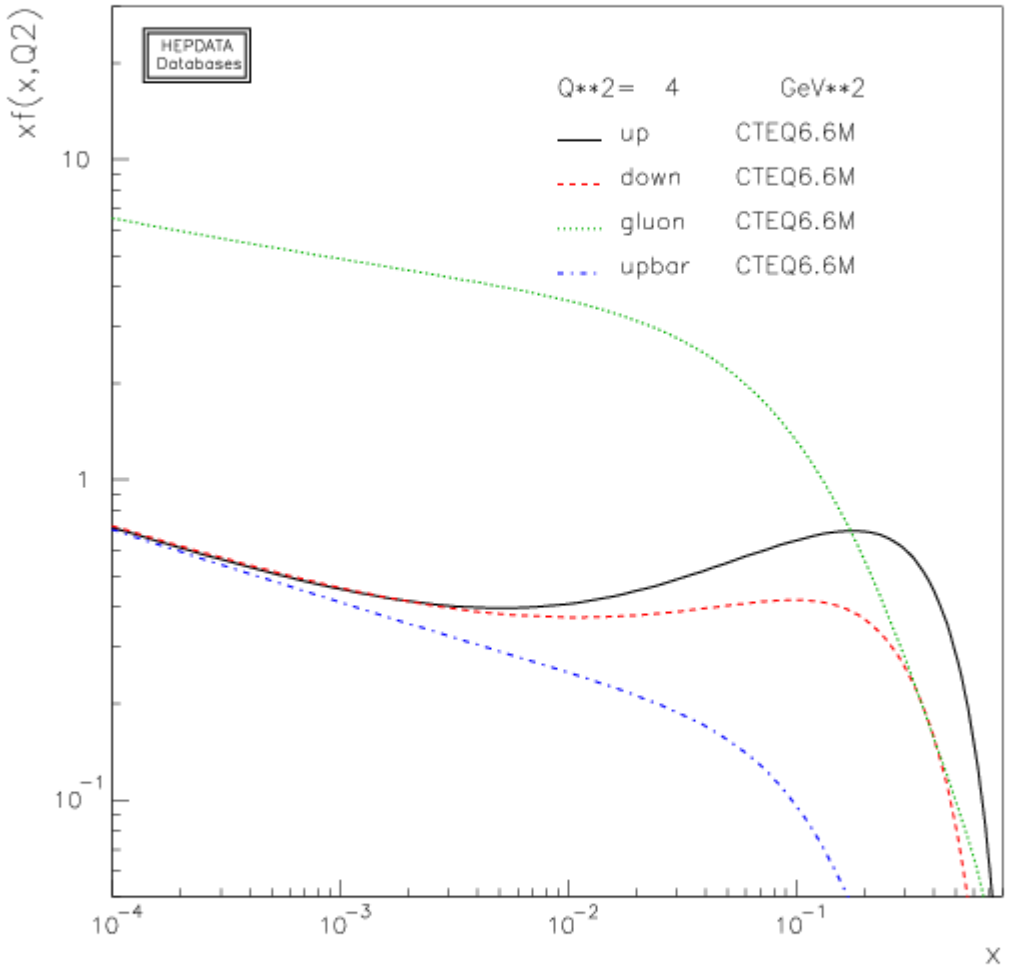
$$\mu\frac{df_{i/p}(x,\mu)}{d\mu} = \sum_{j=g,u,\bar{u},d,\bar{d},....}\int_x^1 \frac{dy}{y}P_{i/j}\left(\frac{x}{y},\alpha_s(\mu)\right)f_{j/p}(y,\mu)$$

$P_{i/j}$ are probabilities for $j \to ik$ collinear splittings;

are known to order NNLO exactly and N3LO approximately:

$$P_{i/j}(x,\alpha_s) = \alpha_s P_{i/j}^{(1)}(x) + \alpha_s^2 P_{i/j}^{(2)}(x) + \alpha_s^3 P_{i/j}^{(3)}(x) + \alpha_s^4 P_{i/j}^{(4,\mathrm{approx})}(x)\ldots$$

# Example of DGLAP evolution

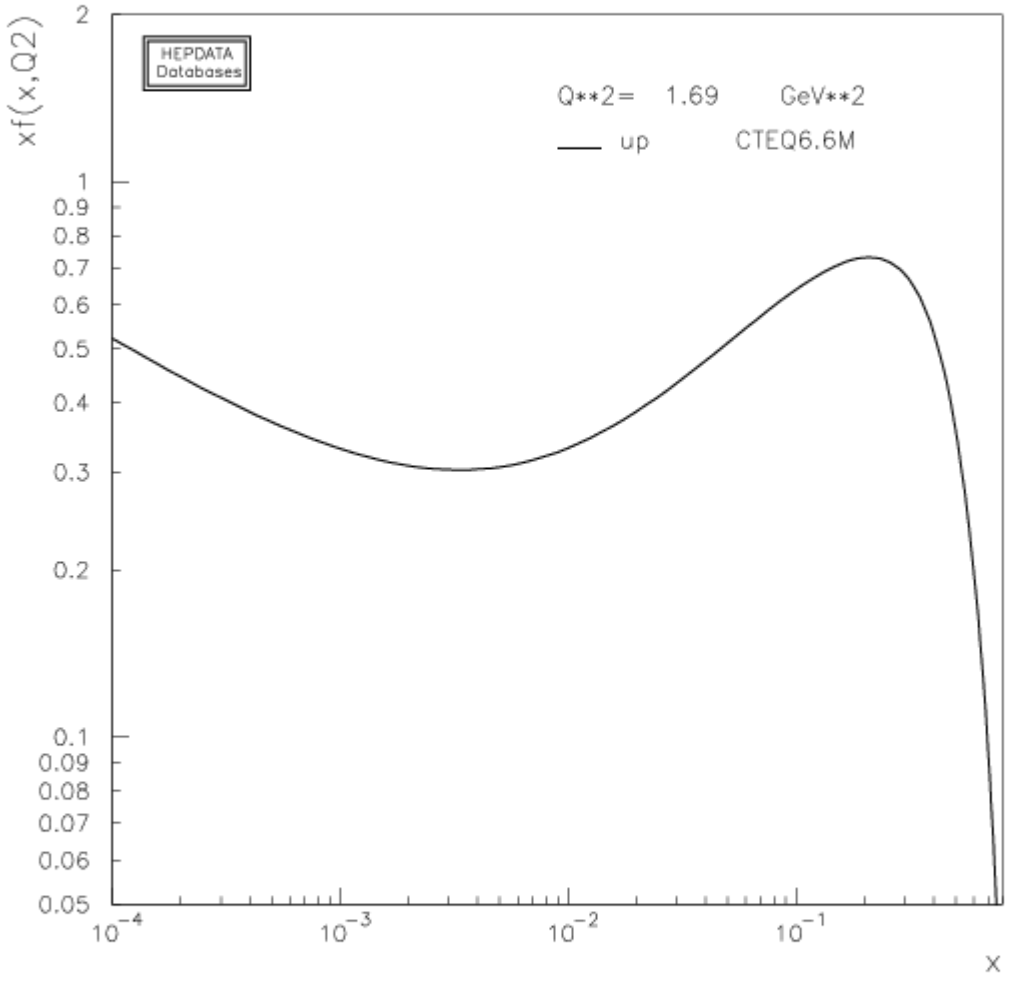



Compare $\mu$ dependence of $u$ quark PDF and the gluon PDF

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

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

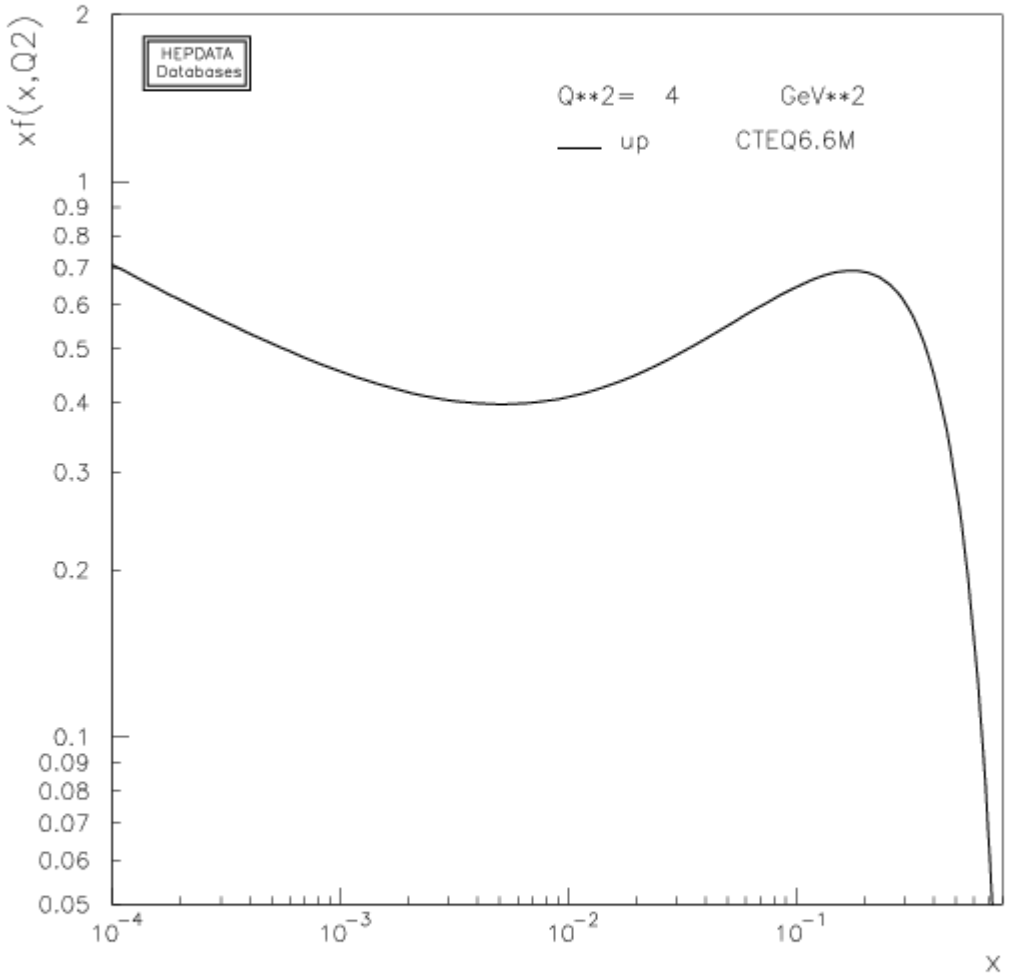
# Example of DGLAP evolution




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

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

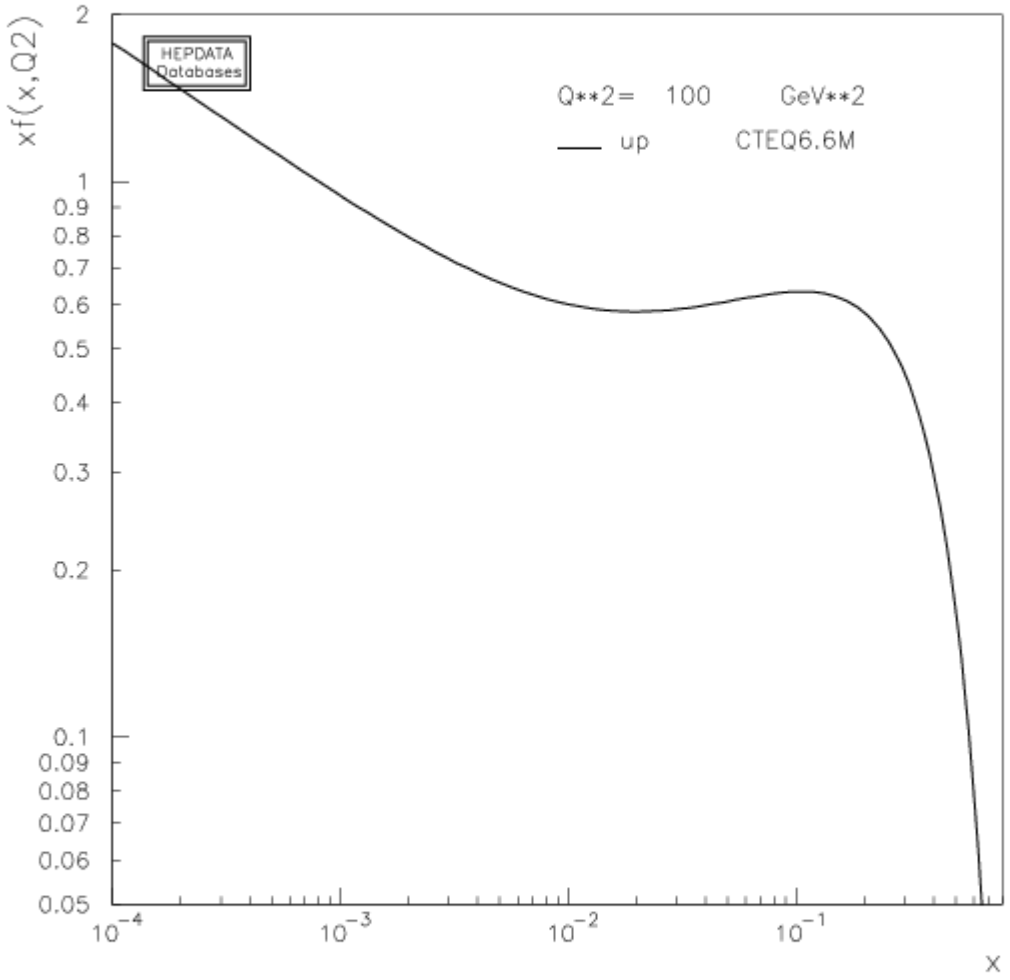
# Example of DGLAP evolution




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

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

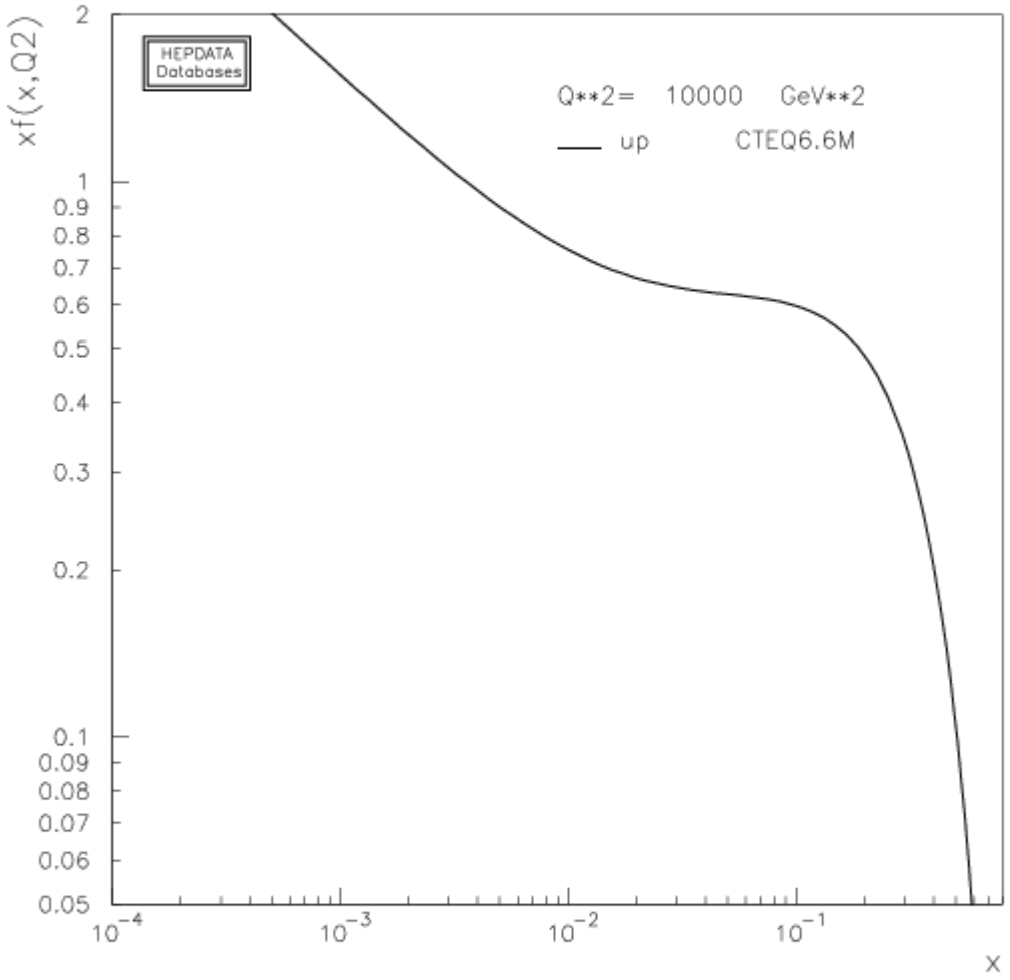
# Example of DGLAP evolution




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

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

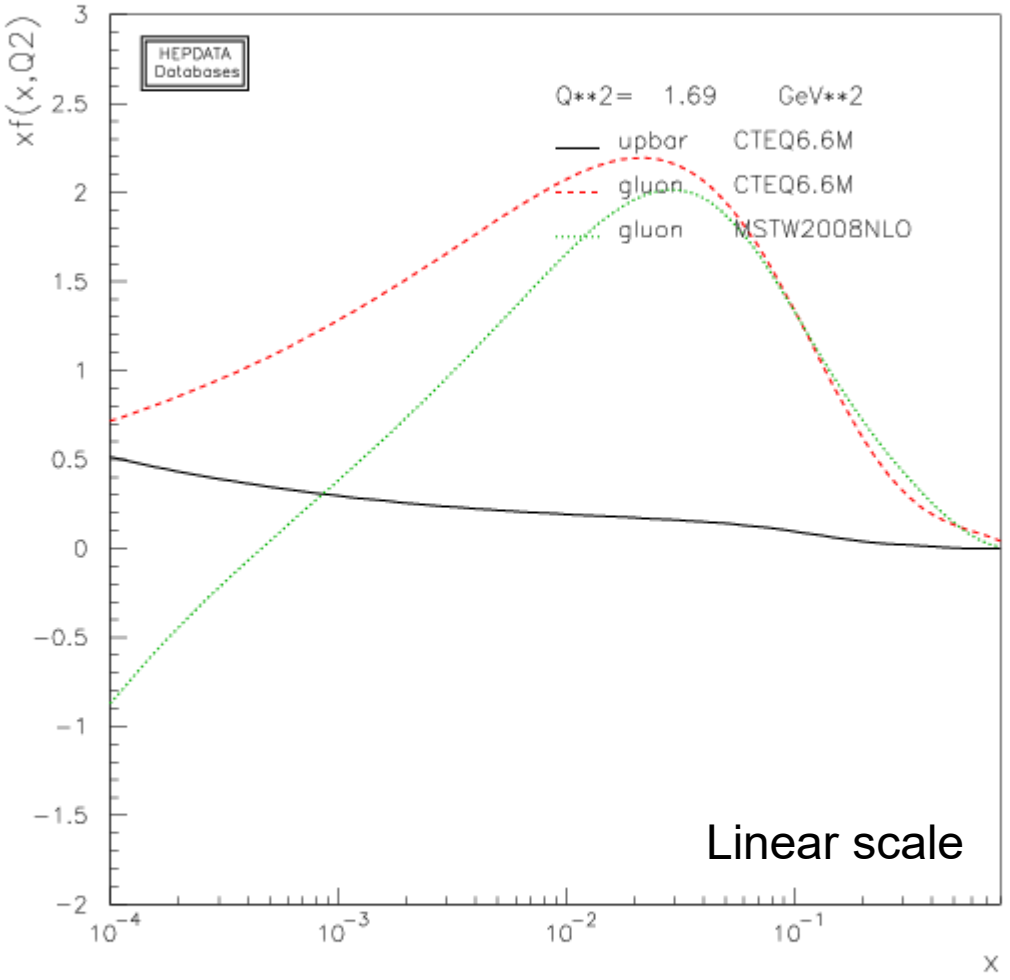
# Example of DGLAP evolution




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

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

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



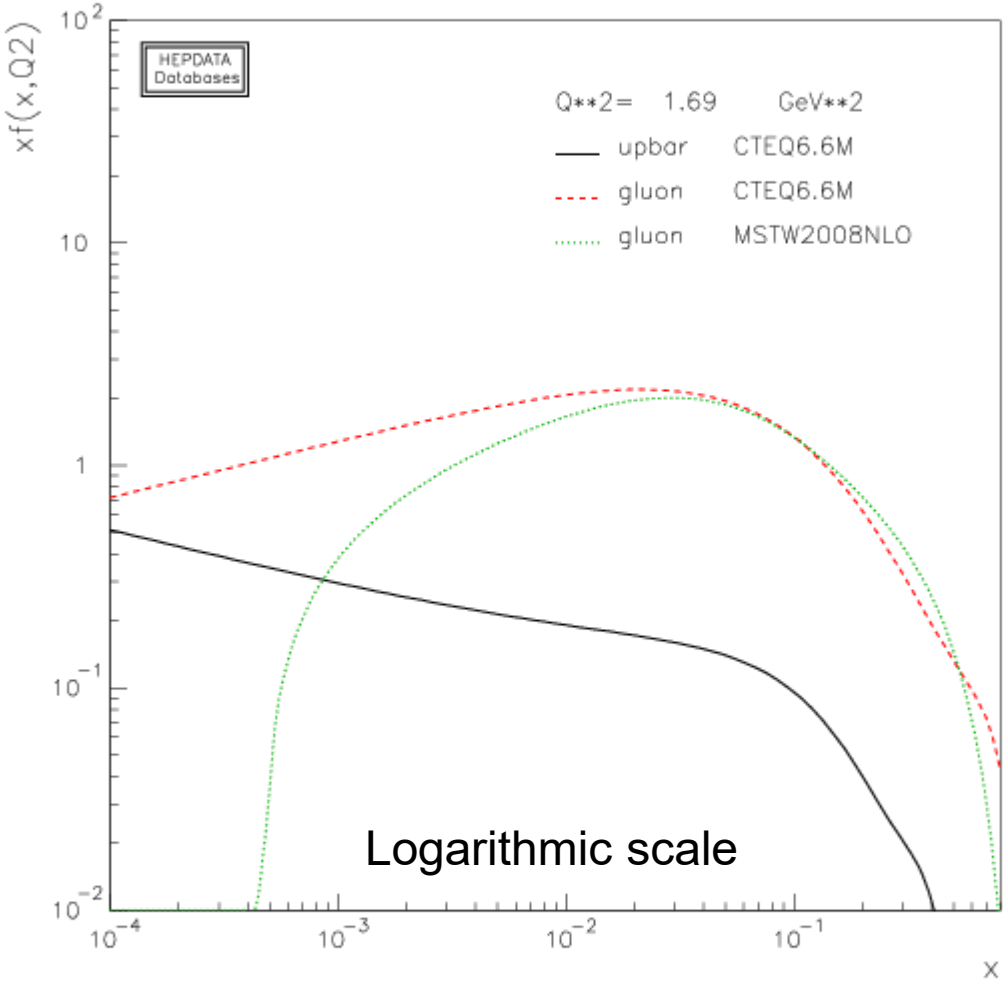

$g(x, Q)$ can become negative at $x < 10^{-2}$, $Q < 2$ GeV;

may lead to unphysical predictions

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

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

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



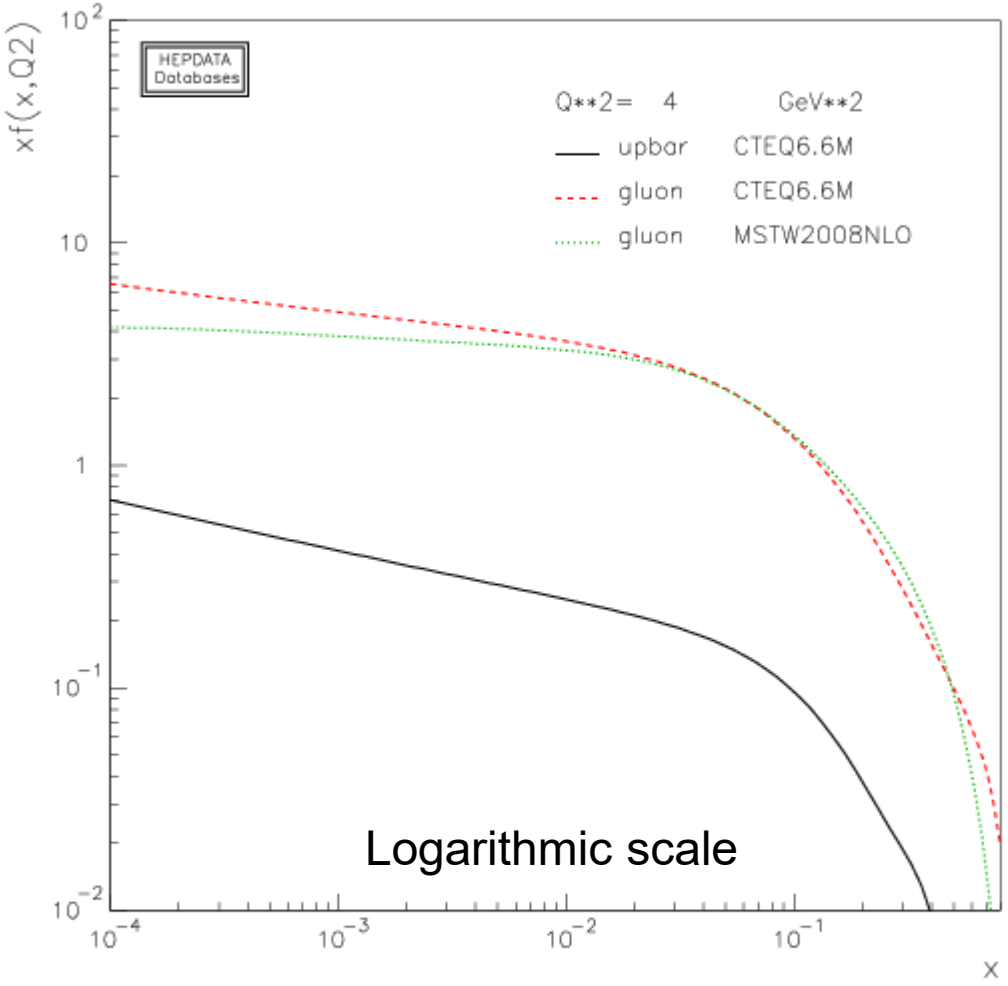

$g(x, Q)$ can become negative at $x < 10^{-2}$, $Q < 2$ GeV;

may lead to unphysical predictions

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

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

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

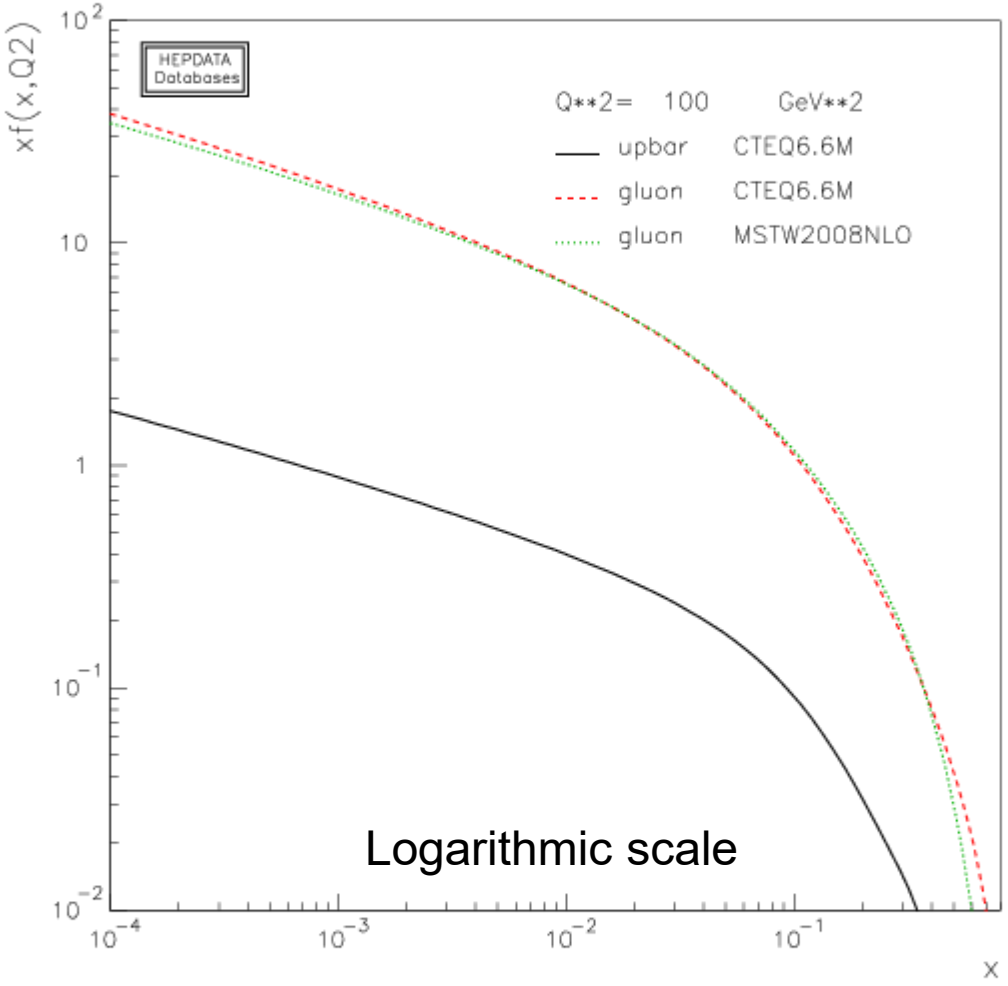



As $Q$ increases, $g(x, Q)$ grows rapidly at small $x$

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

small-$x$ behavior stabilizes

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

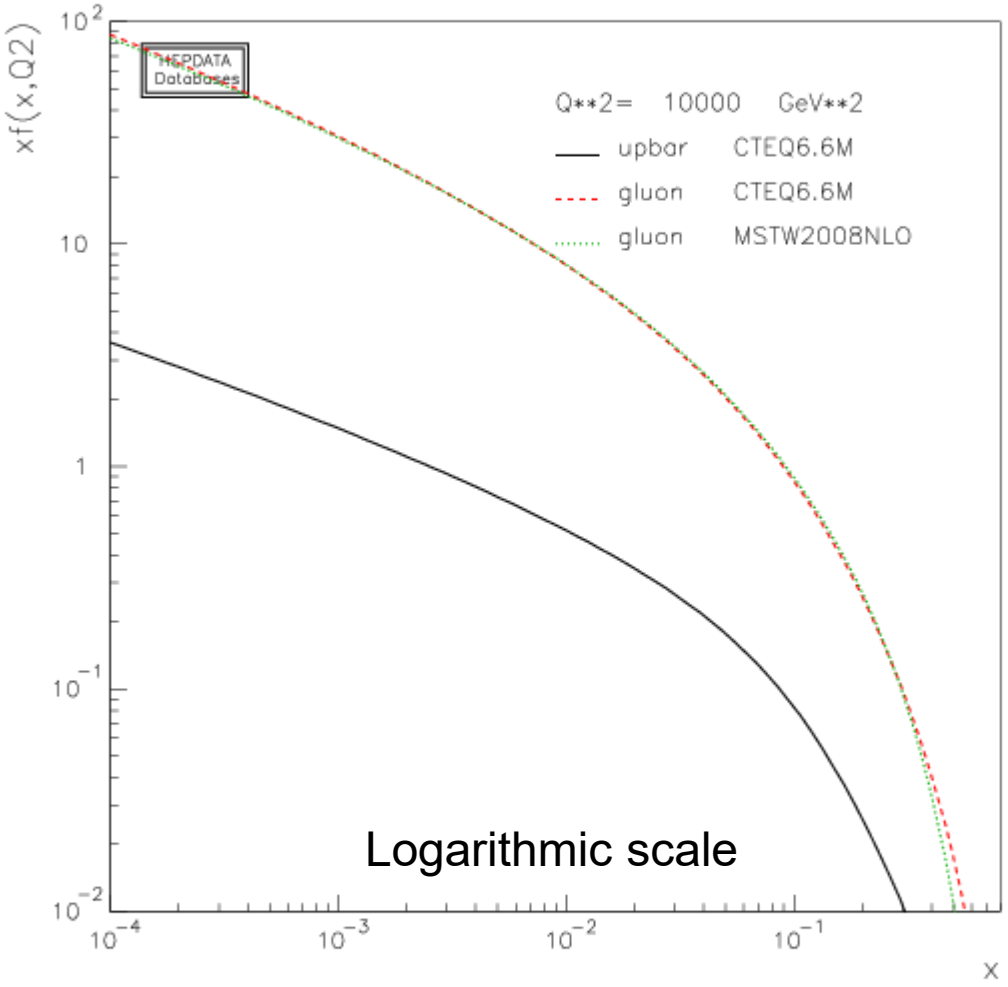



As $Q$ increases, $g(x, Q)$ grows rapidly at small $x$

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

small-$x$ behavior stabilizes

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




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

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

small-$x$ behavior stabilizes

# Universality of PDFs

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

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

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

…can be **parametrized** as

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

…predicted by solving DGLAP equations at $\mu > Q_0$

# Factorized QCD predictions

Lepton-hadron scattering

$$\sigma = \sum_{a} \hat{\sigma}_a \otimes f_{a/p}$$

Hadron-hadron scattering

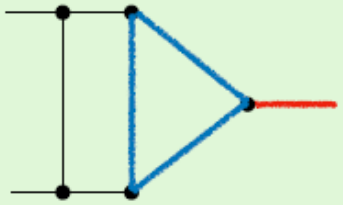
$$\sigma = \sum_{a_1,a_2} \hat{\sigma}_{a_1 a_2} \otimes f_{a_1/p_1} \otimes f_{a_2/p_2}$$

The accuracy in determination of PDFs $f_{a/p}$ must match the accuracy of hard cross sections $\hat{\sigma}$

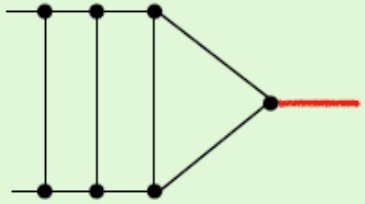
# The multi-loop frontier

F. Maltoni, Snowmass'21 meeting

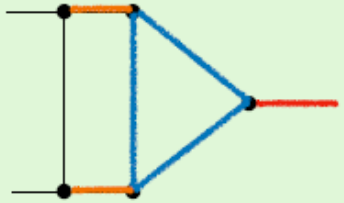
DONE



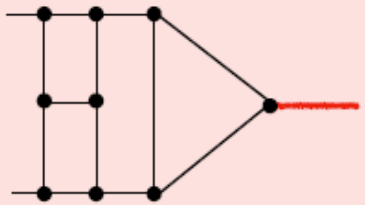
H,Z,W at N2LO$_{QCD}$



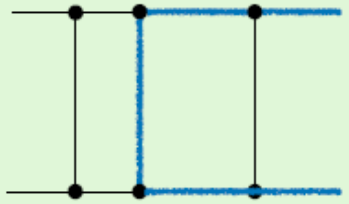
H,Z,W at N3LO



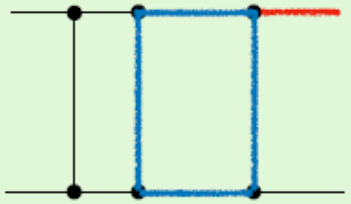
H,Z,W at NLO2$_{EWxQCD}$



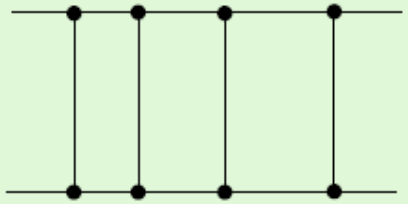
H,Z,W at N4LO$_{QCD}$



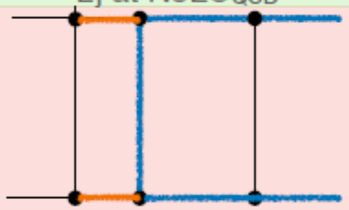
tt at N2LO$_{QCD}$



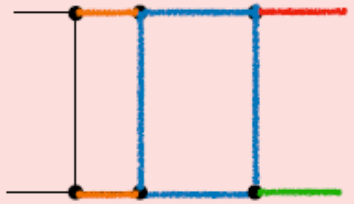
H+j at NLO$_{QCD}$



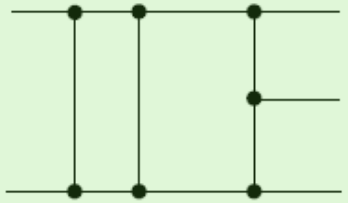
2j at N3LO$_{QCD}$



tt at NLO2$_{EWxQCD}$

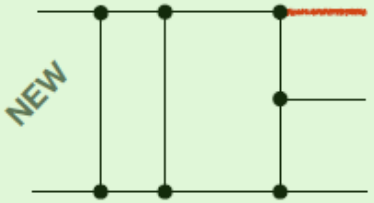


ZH at N2LO$_{EW}$

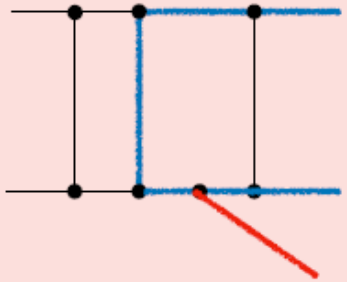


3j at N2LO$_{QCD}$

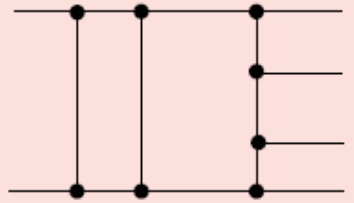
NEW



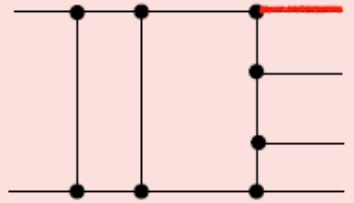
Vbb at N2LO$_{QCD}$



ttH at N2LO$_{QCD}$



4j at N2LO$_{QCD}$



V+3j at N2LO$_{QCD}$

As of 22 July 2022.
FAST MOVING FRONTIER

Growing number of Z, W, H $2 \to 1$ hard cross sections are computed at N3LO QCD

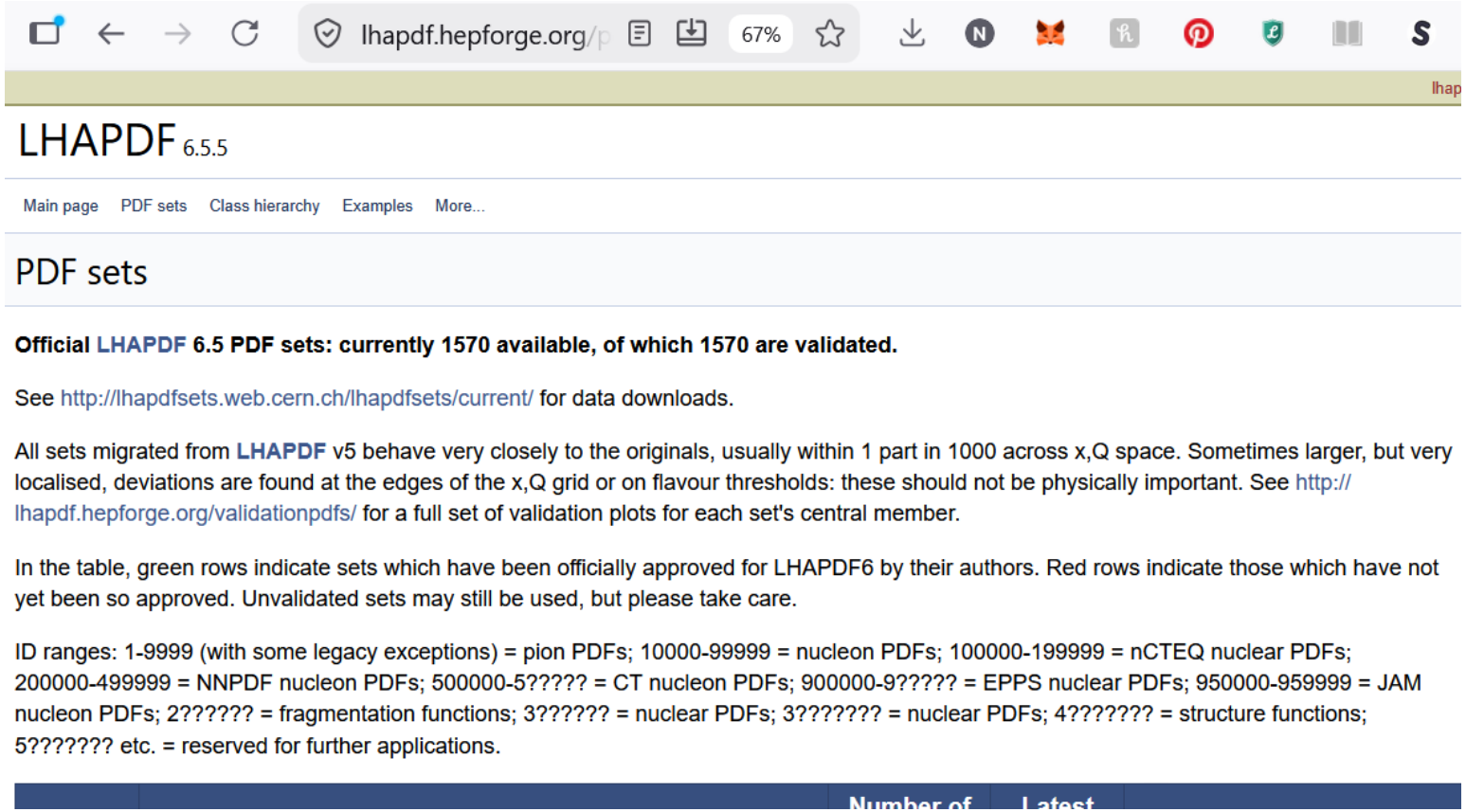
**PDFs must have comparable accuracy (NNLO or N3LO)**

TO DO

# Where do the PDFs come from?

# Where do the PDFs come from?

Practical answer: from the Les Houches Accord PDF library (lhapdf.hepforge.org)




Thousands of PDF sets are provided by the LHAPDF C++ library and can be linked to your computer code. Which one should you use?

# Where do the PDFs come from?

- From a combination of BIG, medium, and **small** experiments
- Complementarity in
  - kinematical ranges
  - systematics

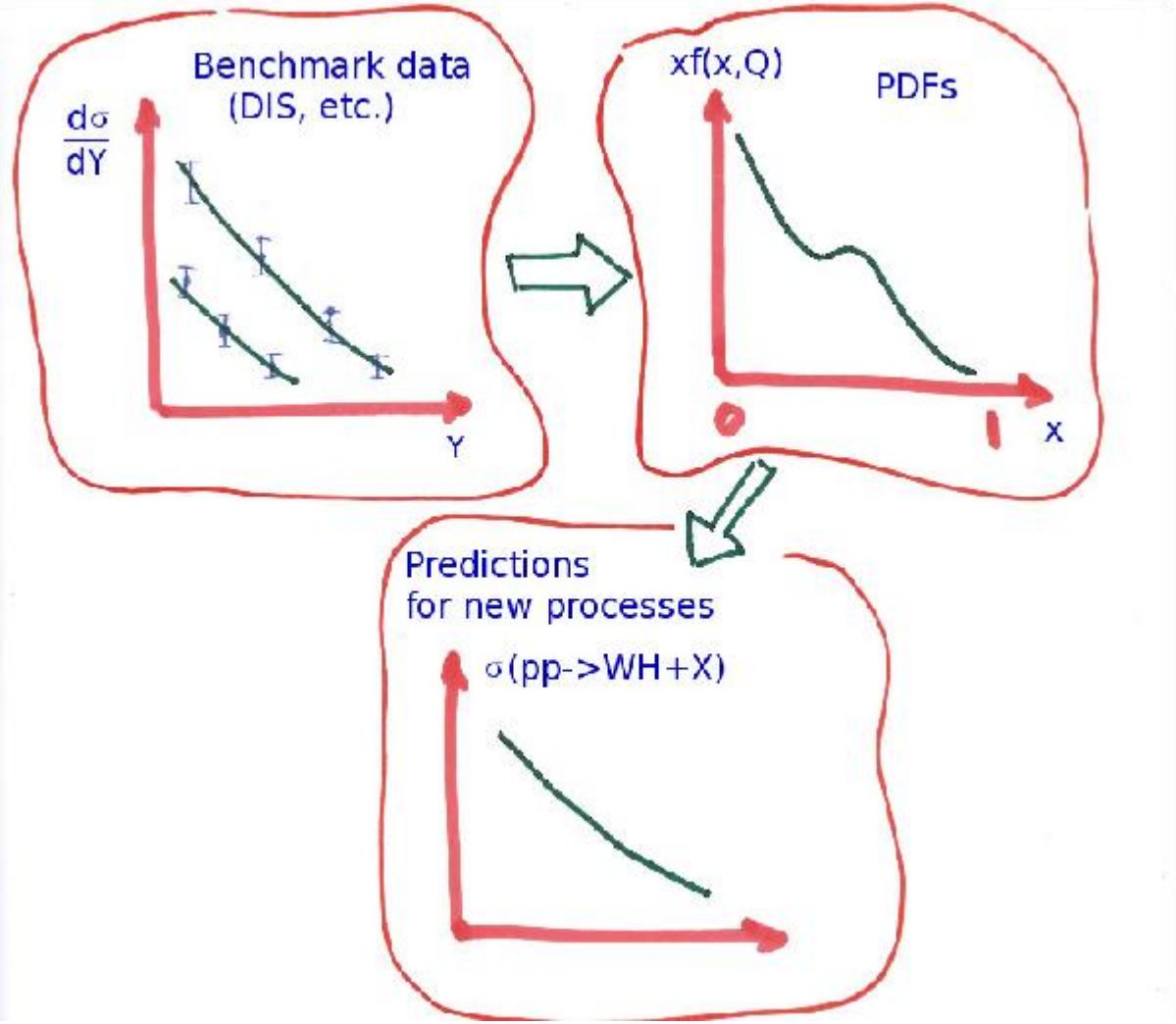
LHC
Tevatron

HERA
RHIC
EIC

Fixed-target experiments

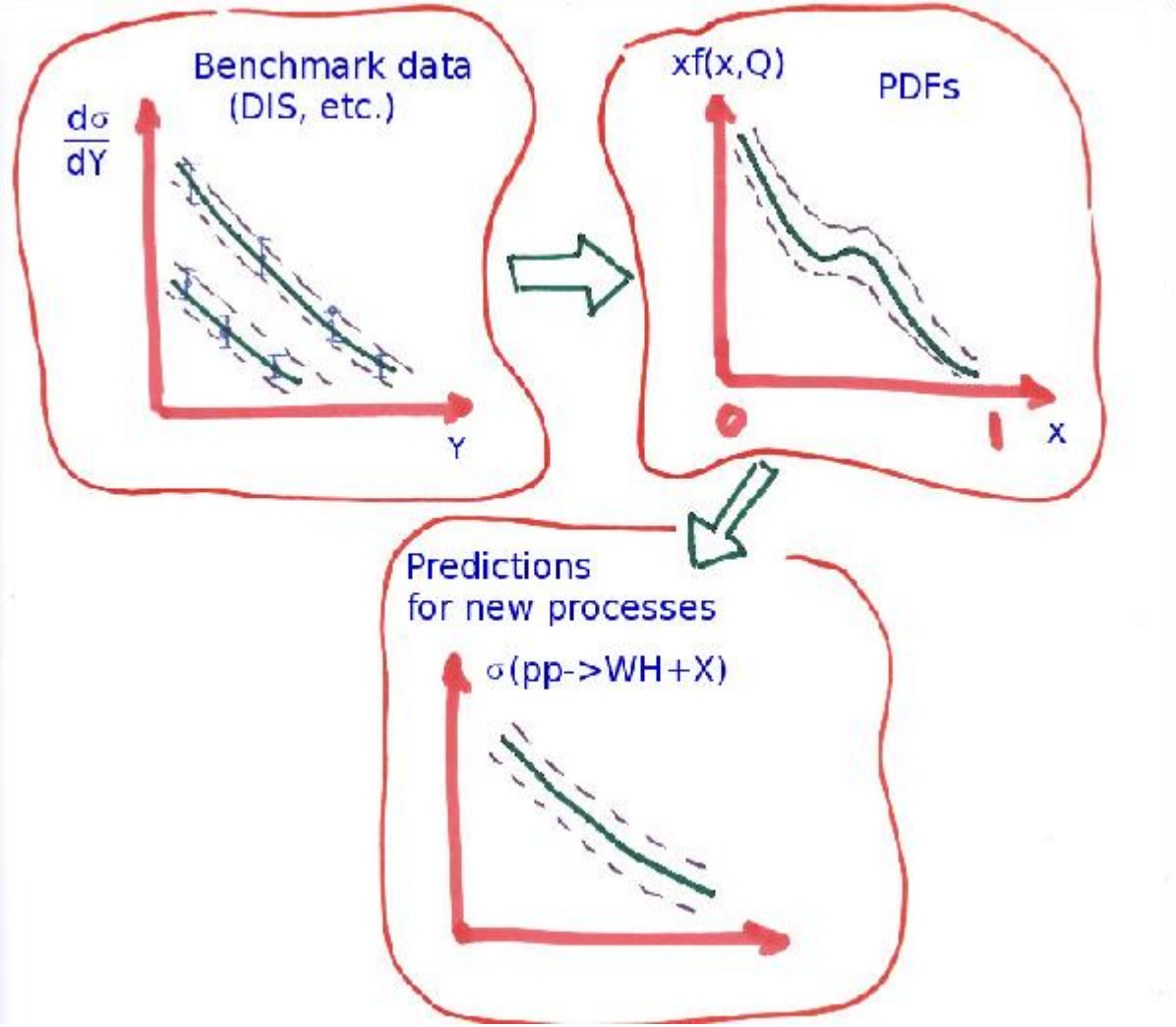
+ lattice QCD

# The flow of the global analysis




PDFs are not measured directly, but some data sets are sensitive to specific combinations of PDFs. By constraining these combinations, the PDFs can be disentangled in a combined (global) fit.

# The flow of the global analysis

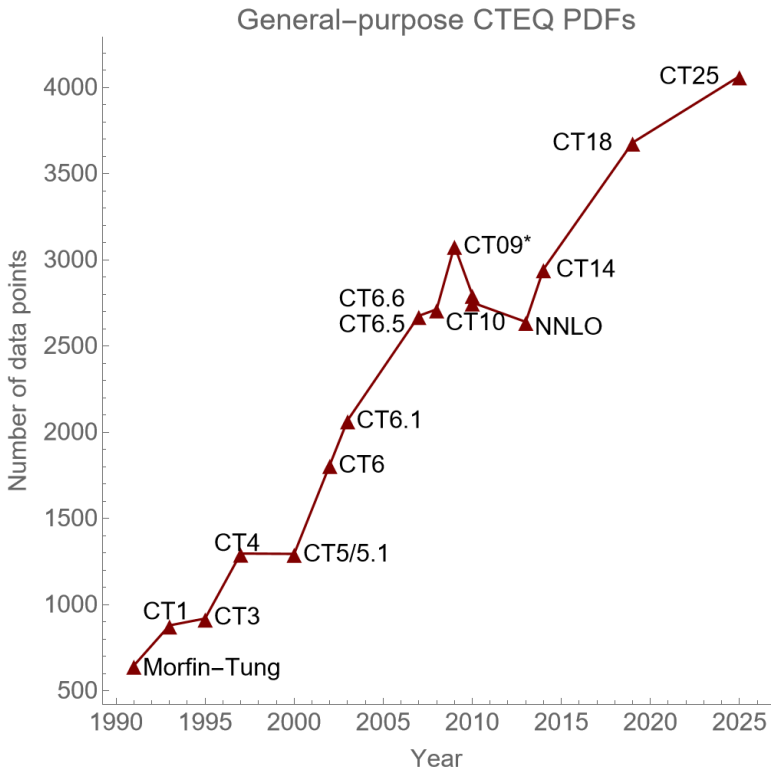


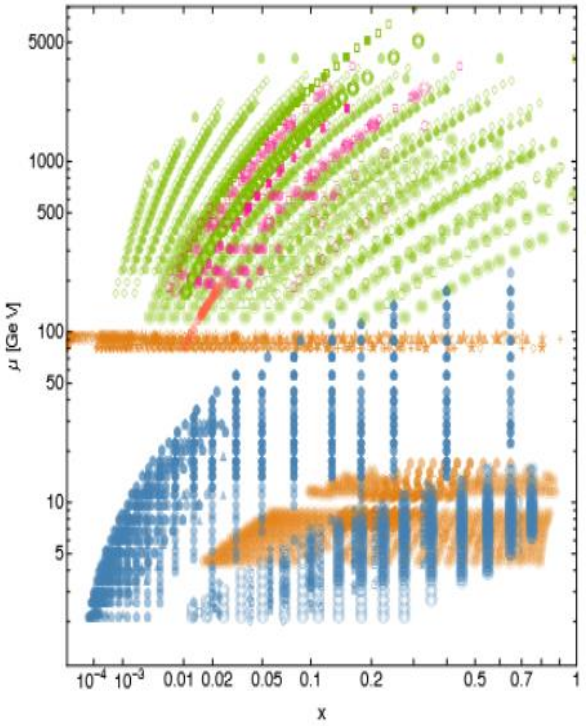

PDFs are not measured directly, but some data sets are sensitive to specific combinations of PDFs. By constraining these combinations, the PDFs can be disentangled in a combined (global) fit.

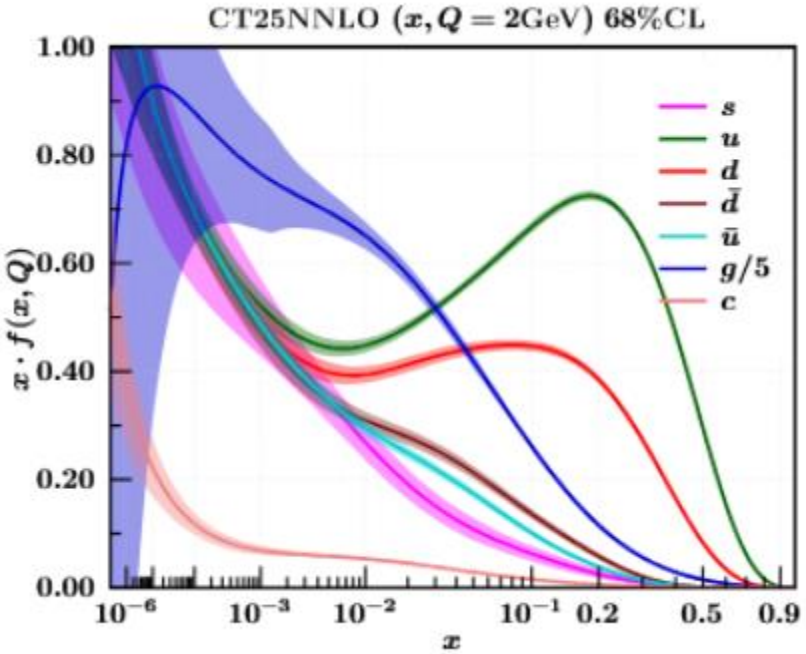
# The CT25 NNLO global fit of nucleon data

**Announced in** 2512.19779 (proceedings);
**journal publication in preparation**

A culmination of several years of studies since the release of CT18 NNLO
Four ensembles of CT25 NNLO PDFs are published on cteq-tea.gitlab.io




4066 data points from CT18 baseline datasets
+13 precision LHC data sets at 5.02, 8, 13 TeV based on arXiv:2305.10733, 2307.11153, 2412.00350
+SeaQuest DY data

PDFs are provided with uncertainties

# A question to you (think for 1 minute)

Among Standard Model particles, which particles can have a non-zero PDF?

# Boundary conditions at $Q_0 \approx 1$ GeV

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

- $g, u, d, s, \bar{u}, \bar{d}, \bar{s}$ (always)

contribute $> 97\%$ of the proton's energy $E_p$ at $Q_0$ - even in this case, the data are usually insufficient for constraining all PDF parameters; some of them can be fixed by hand - e.g., $\bar{u} = \bar{d} = \bar{s}$ in outdated fits

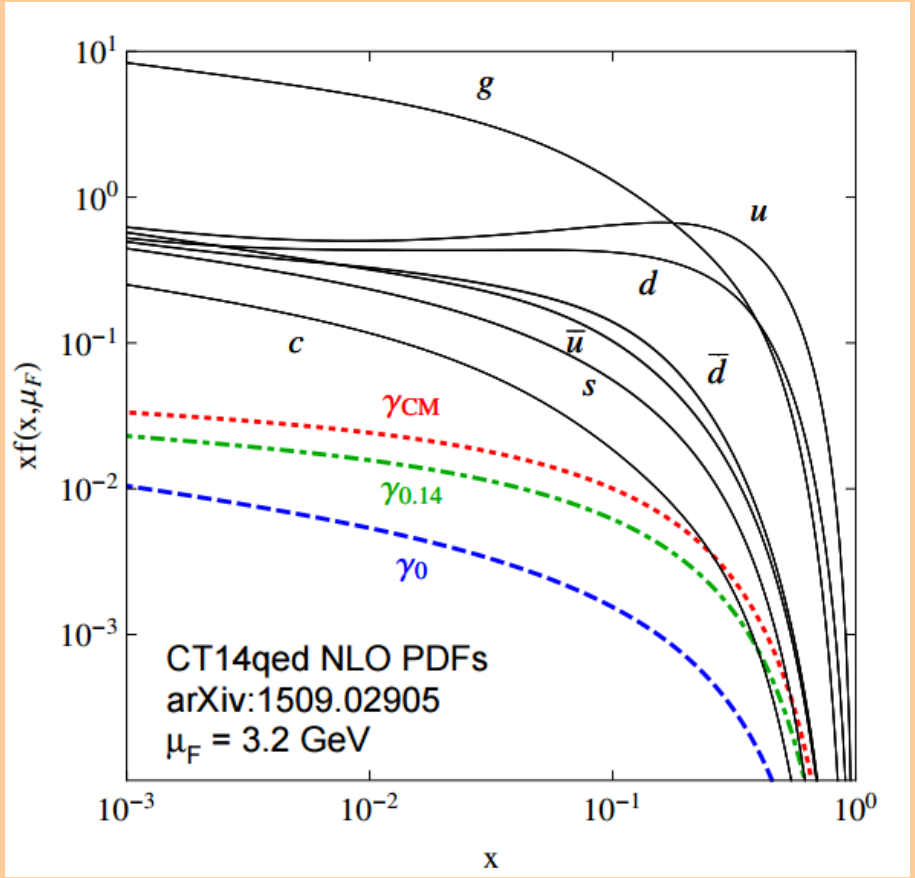
- $c$ and or $b$ (occasionally; in a model allowing nonperturbative ``intrinsic heavy-quark production'')

- photons $\gamma$ (in QCD+QED PDFs by CT, xFitter, LUX, MRST, NNPDF… groups)

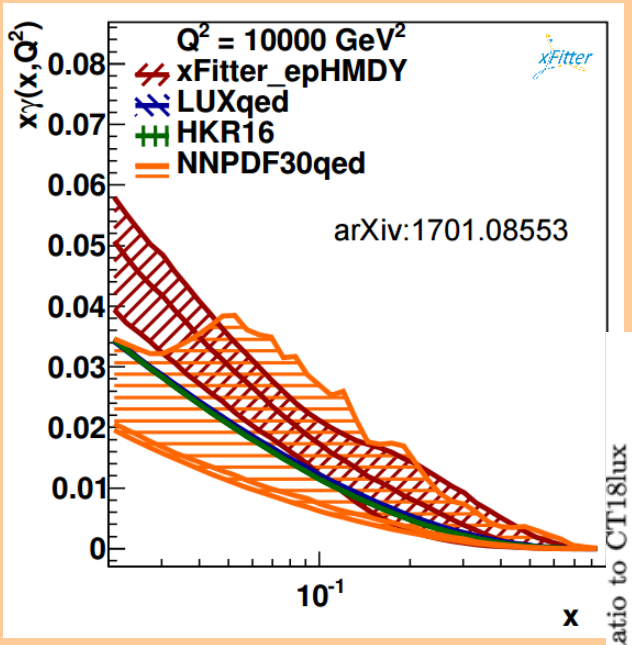
# Example: QCD+QED PDFs

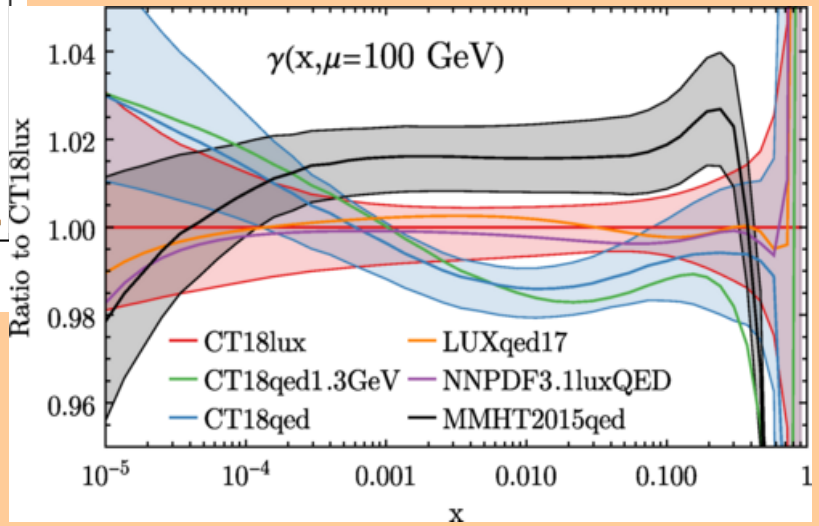
The LUXqed group (Manohar et al., 1607.04266, 1708.01256) derives $f_{\gamma/p}(x, Q)$ from DIS inclusive cross section ⇒ very small uncertainty on $f_{\gamma/p}(x, Q)$

**Before LUXqed**




**After LUXqed**



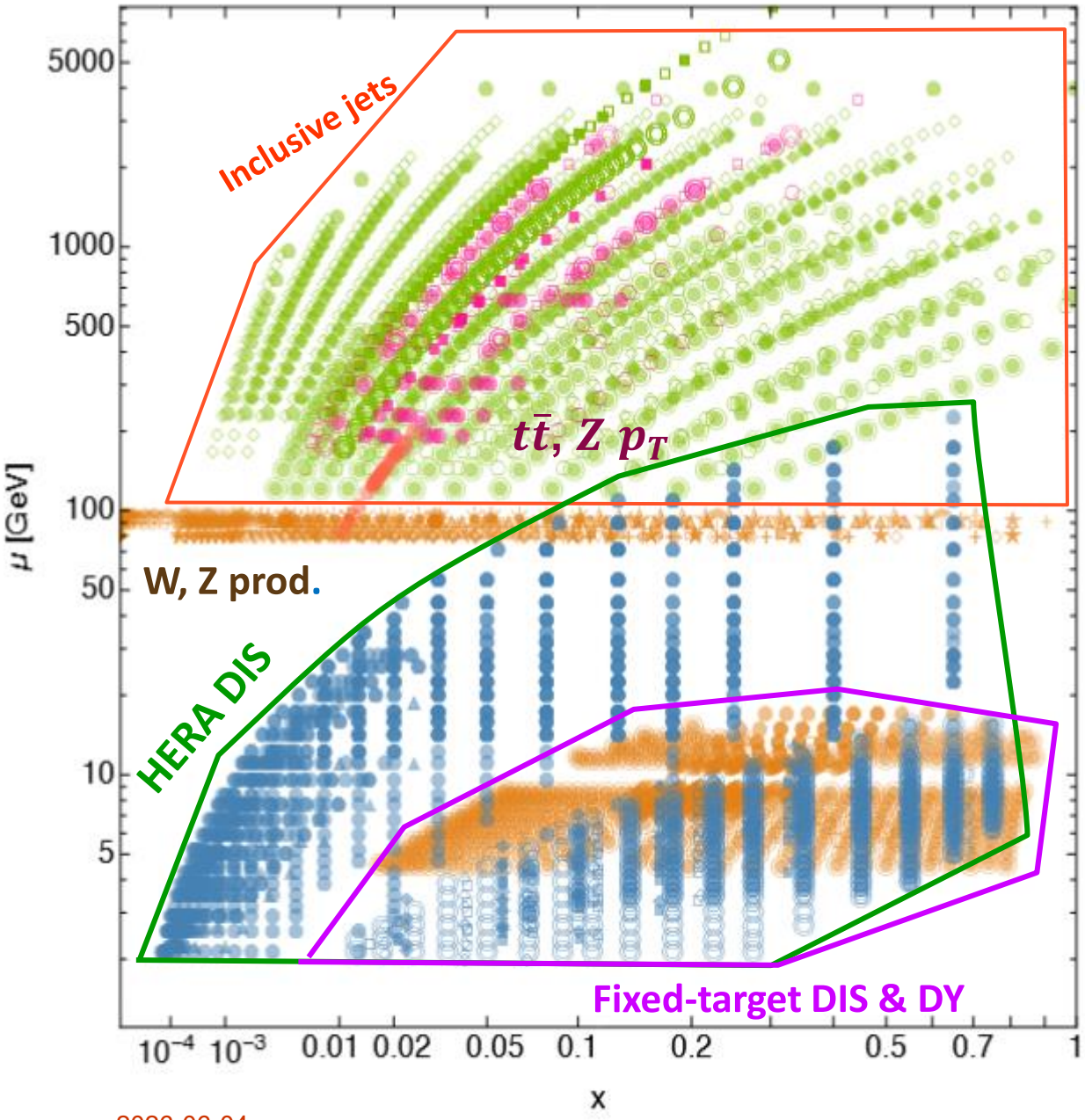



# Another question [1 minute]

Given a QCD observable $\mathcal{O}$, can you tell which parton flavors drive the PDF uncertainty on $\mathcal{O}$? How?

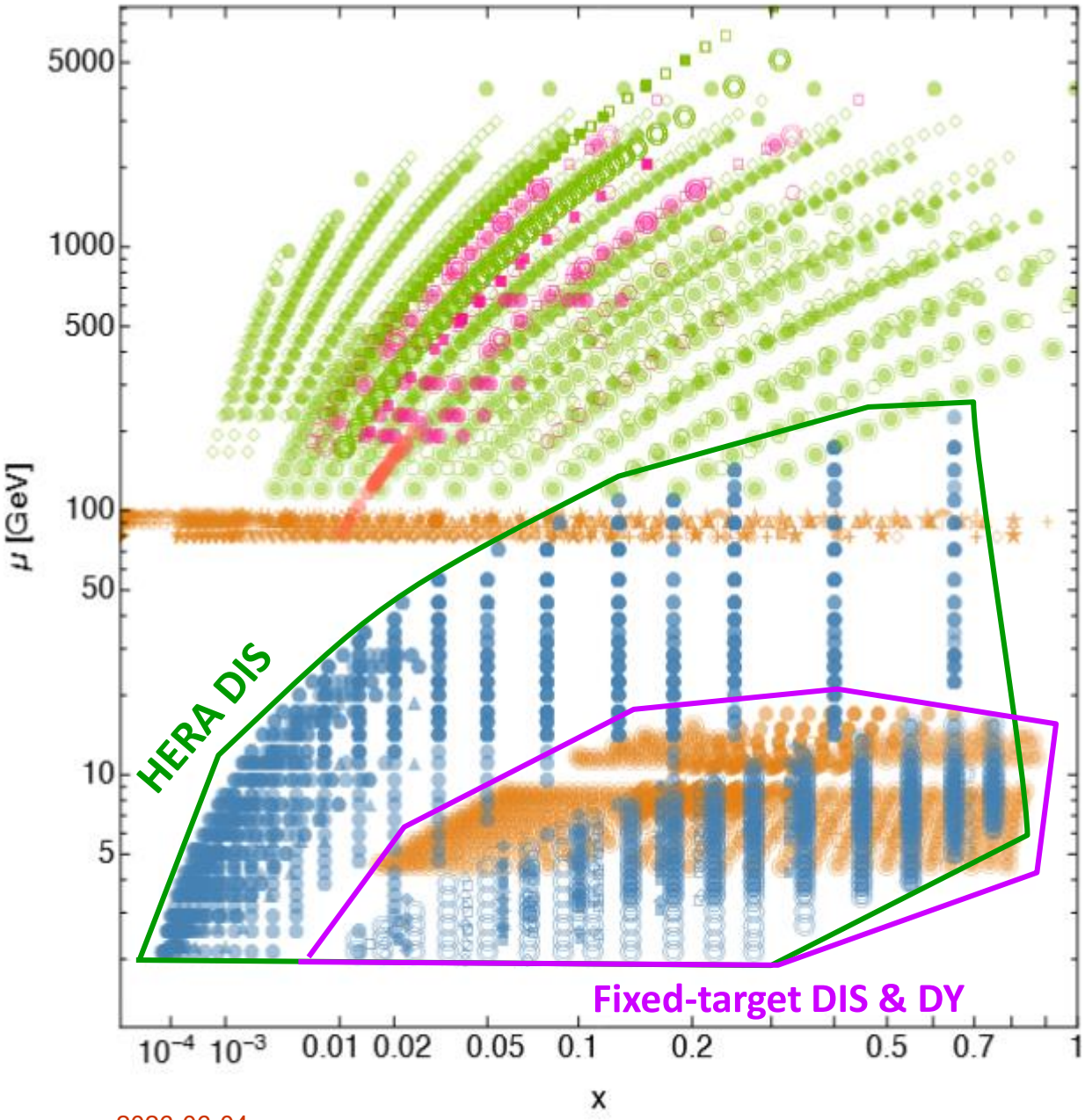
# 2.Experimental observables constraining the PDFs

# Data sets in CT25 NNLO analysis




- HERA DIS combined
- BCDMS $F_2^p$
- BCDMS $F_2^d$
- NMC d/p ratio
- NuTeV $\sigma_{\mu\mu}$ $\nu$
- NuTeV $\sigma_{\mu\mu}$ $\bar{\nu}$
- CCFR $\sigma_{\mu\mu}$ $\nu$
- CCFR $\sigma_{\mu\mu}$ $\bar{\nu}$
- H1/ZEUS charm
- H1/ZEUS bottom
- E605 Drell–Yan
- E866 Drell–Yan ratio
- E866 pp Drell–Yan
- E906 Drell–Yan ratio
- CDF $A_l^{charge}$
- CDF $A_e^{charge}$
- DØ $A_\mu^{charge}$
- DØ $A_e^{charge}$
- DØ Z
- CDF Z
- CMS 7 TeV $A_\mu^{charge}$
- CMS 7 TeV $A_e^{charge}$
- LHCb 7 TeV $W/Z$
- LHCb 8 TeV $Z \to e^+e^-$
- ATLAS 7 TeV $W/Z$ [2016]
- CMS 8 TeV $A_l^{charge}$
- LHCb 8 TeV $W/Z$
- ATLAS 8 TeV $W$ $A_\mu^{charge}$
- CMS 13 TeV Z
- LHCb 13 TeV Z
- ATLAS 8 TeV Z 3D
- ATLAS 5.02 TeV $W/Z$
- LHCb 8 TeV $W$
- ATLAS 7 TeV incl. jet
- CDF incl.jet
- DØ incl.jet
- ATLAS 8 TeV incl.jet
- ATLAS 13 TeV incl.jet
- CMS 13 TeV incl.jet
- CMS 7 TeV incl.jet
- CMS 8 TeV incl.jet
- ATLAS 8 TeV $Z$ $pT$
- CMS 8 TeV $t\bar{t}$
- ATLAS 8 TeV $t\bar{t}$
- ATLAS 13 TeV all–hadronic $y_{t\bar{t}}$
- CMS 13 TeV $ll$ $y_{t\bar{t}}$
- ATLAS 13 TeV $lj$ $m_{t\bar{t}}$+$y_{t\bar{t}}$+$y_{t\bar{t}}^B$+$H_{T,t\bar{t}}$
- CMS 13 TeV $lj$ $m_{t\bar{t}}$

# Data sets in CT25 NNLO analysis




- **Inclusive deep-inelastic scattering**
- At HERA:

neutral-current $e^{\pm}p \to e^{\pm}X$;

charged-current $ep \to \nu X$

- the largest data set in the fit
- Fixed-target experiments
- $eN$, $\mu N$, $\nu N$ scattering

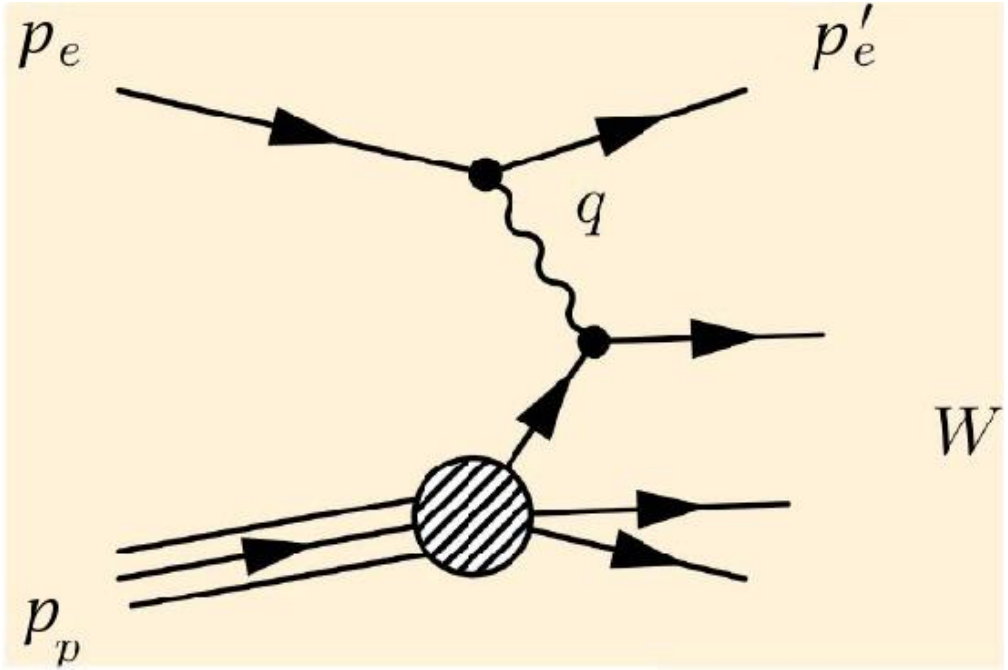
# Neutral-current $ep$ DIS: kinematics

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

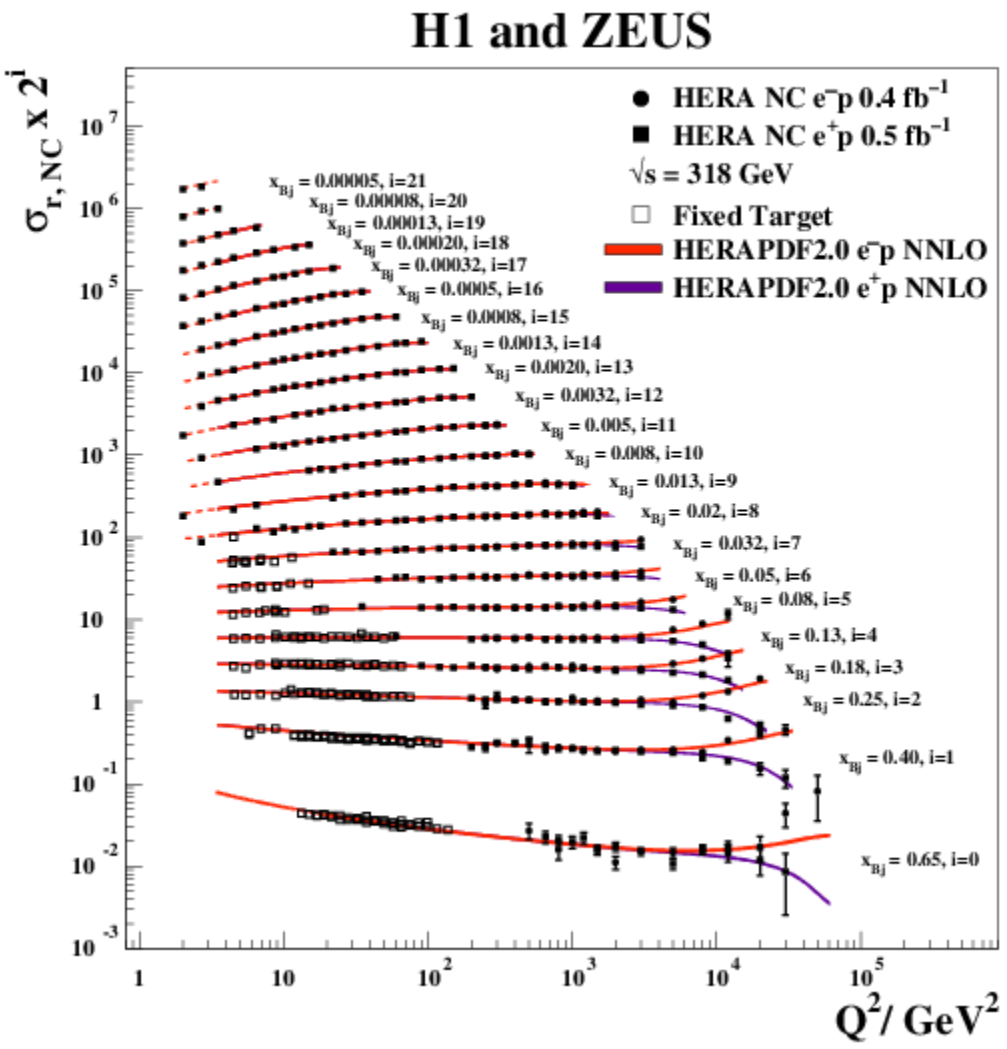
$$\frac{d^2\sigma(e^{\pm}p)}{dQ^2dx} = \frac{2\pi\alpha^2}{Q^4x} Y_+ \left( F_2 - \frac{y^2}{Y_+} F_L \pm \frac{Y_-}{Y_+} xF_3 \right),$$

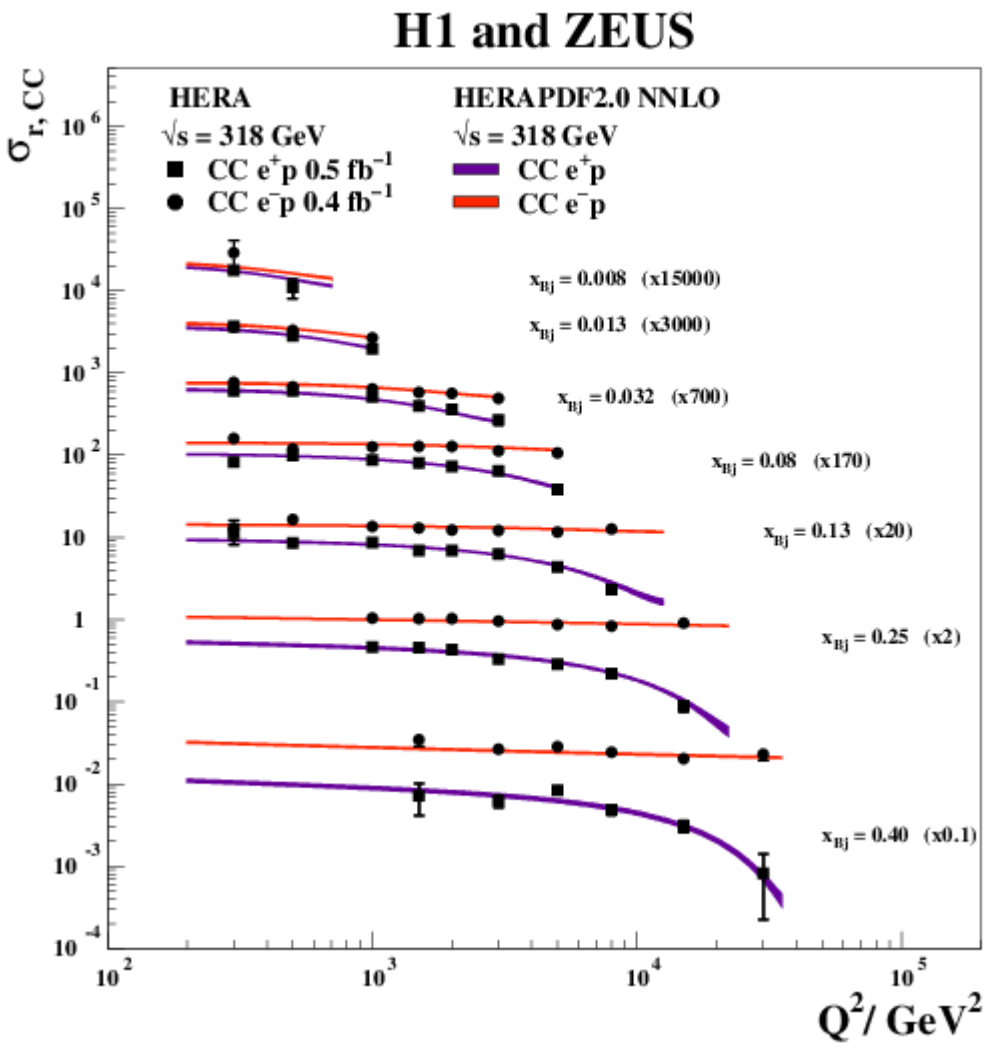
with $Y_\pm \equiv 1 \pm (1-y)^2$

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



# Final combined inclusive cross sections from HERA (2015)



# PDF combinations in DIS at the lowest order

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

$$F_2^{\ell^{\pm}p}(x,Q^2) = \left(\frac{2}{3}\right)^2 (u + \bar{u} + c + \bar{c}) + \left(-\frac{1}{3}\right)^2 \left(d + \bar{d} + s + \bar{s} + b + \bar{b}\right)$$

- PDFs are weighted by the fractional EM quark coupling $e_i^2 = 4/9$ or $1/9$
  - 4 times more sensitivity to $u$ and $c$ than to $d$, $s$, and $b$
- No sensitivity to the gluon at this order
- Neutral current ($\ell^{\pm}N$ (DIS on isoscalar nuclei ($N = (p+n)/2$)

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

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

$$F_2^{\nu N}(x,Q^2) = x \sum_{i=u,d,s\ldots} (q_i + \bar{q}_i) \qquad xF_3^{\nu N}(x,Q^2) = x \sum_{i=u,d,s} (q_i - \bar{q}_i)$$

DIS on nucleus is sensitive to nuclear corrections!

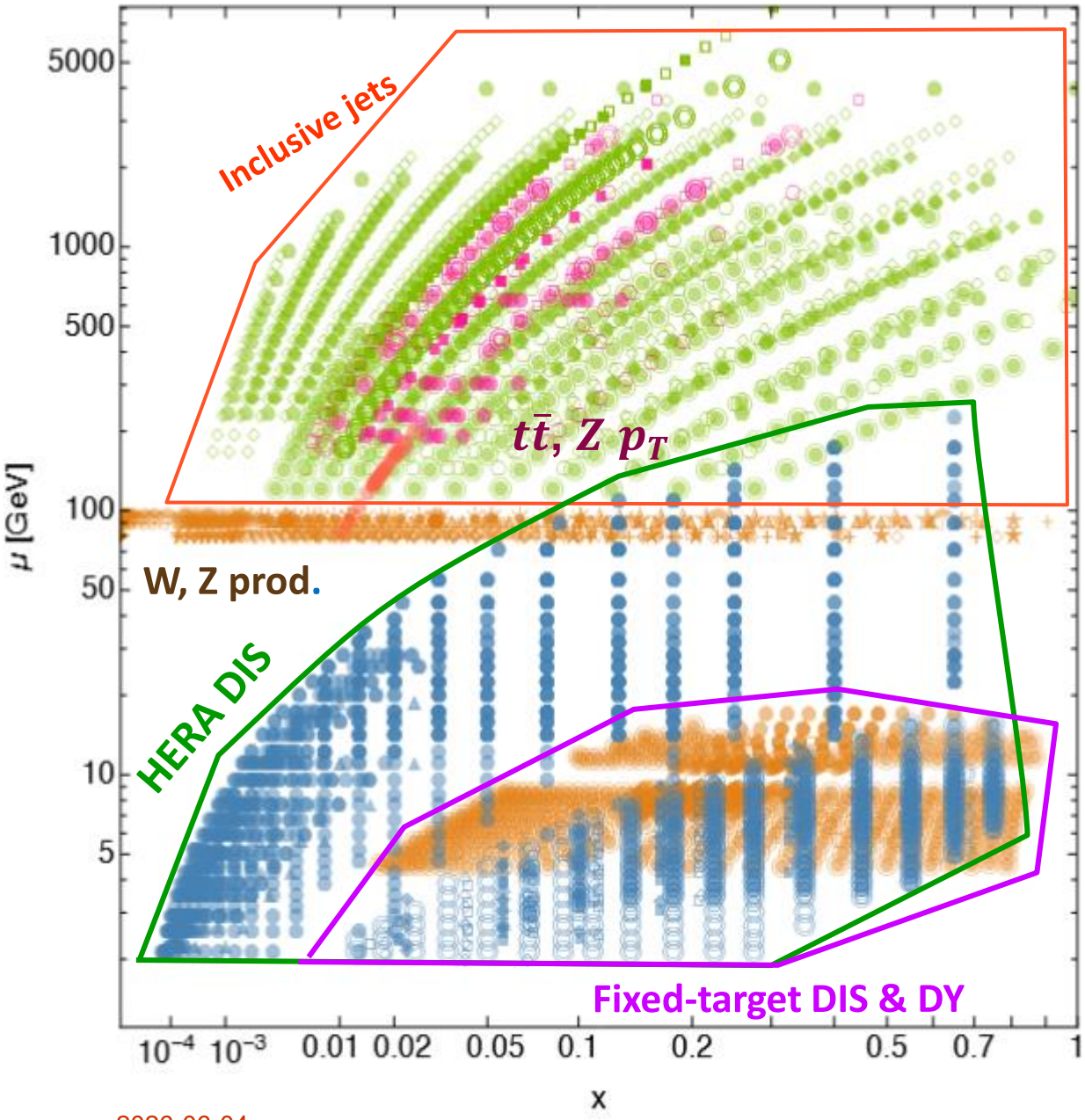
# DIS at next-to-leading order (NLO) and beyond

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

Thus, when examined at NLO, the DIS data constrains

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

DIS cannot fully separate quarks from antiquarks, or $s$, $c$, $b$ contributions from $u$ and $d$ contributions; fixed-target DIS experiments affected by higher-order terms, nuclear corrections,…




The modern PDF fits include

**Inclusive deep-inelastic scattering…**

**+ Semi-inclusive DIS:** - charm production $ep \to ecX$ (HERA)

- $\mu\mu$ production $\nu N \to \mu(c \to \mu)X$ )NuTeV, NOMAD, …)

**+ Lepton pair production**}

(LHC/Tevatron, fixed-target experiments)

**+ Inclusive jet production:** $p\bar{p} \to jX$ )LHC/Tevatron), $ep \to j(j)X$) HERA)

**+ $t\bar{t}$, $Wc$, SIDIS, etc. production**

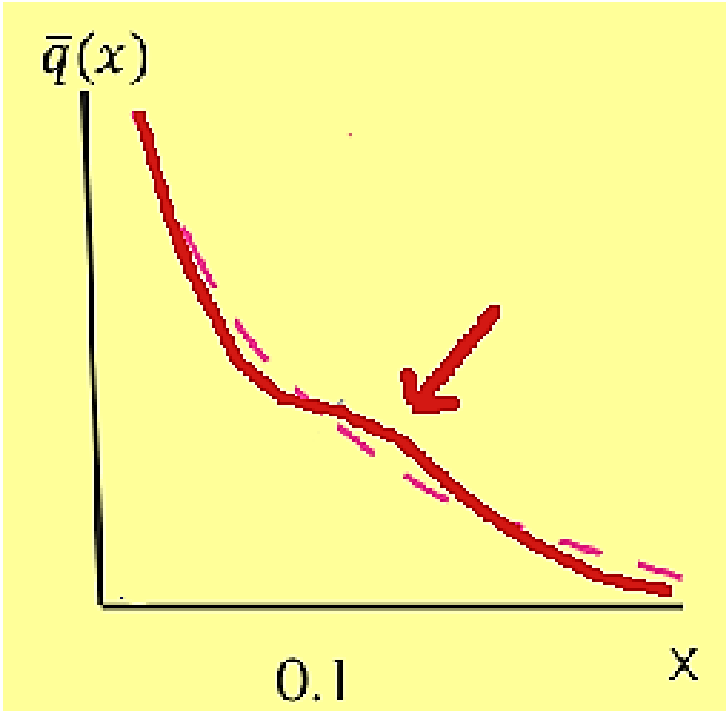
# $SU(2)_{flavor}$ and charge symmetry breaking

$$\bar{d}(x,Q_0) \neq \bar{u}(x,Q_0) \quad or \quad q_{sea}(x,Q_0) \neq \bar{q}(x,Q_0)$$

May be caused by

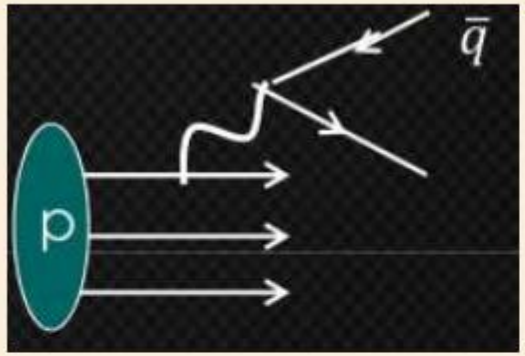
- DGLAP evolution
- Fermi motion
- Electromagnetic effects
- Nonperturbative meson fluctuations
- Chiral symmetry breaking
- Instantons

1% accuracy can distinguish between these effects.

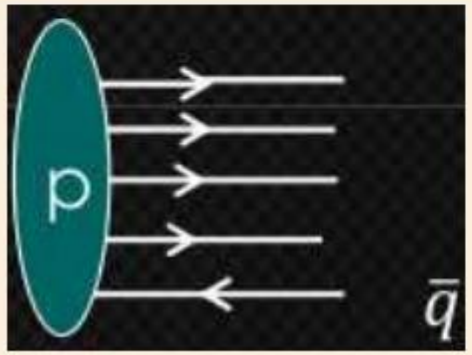


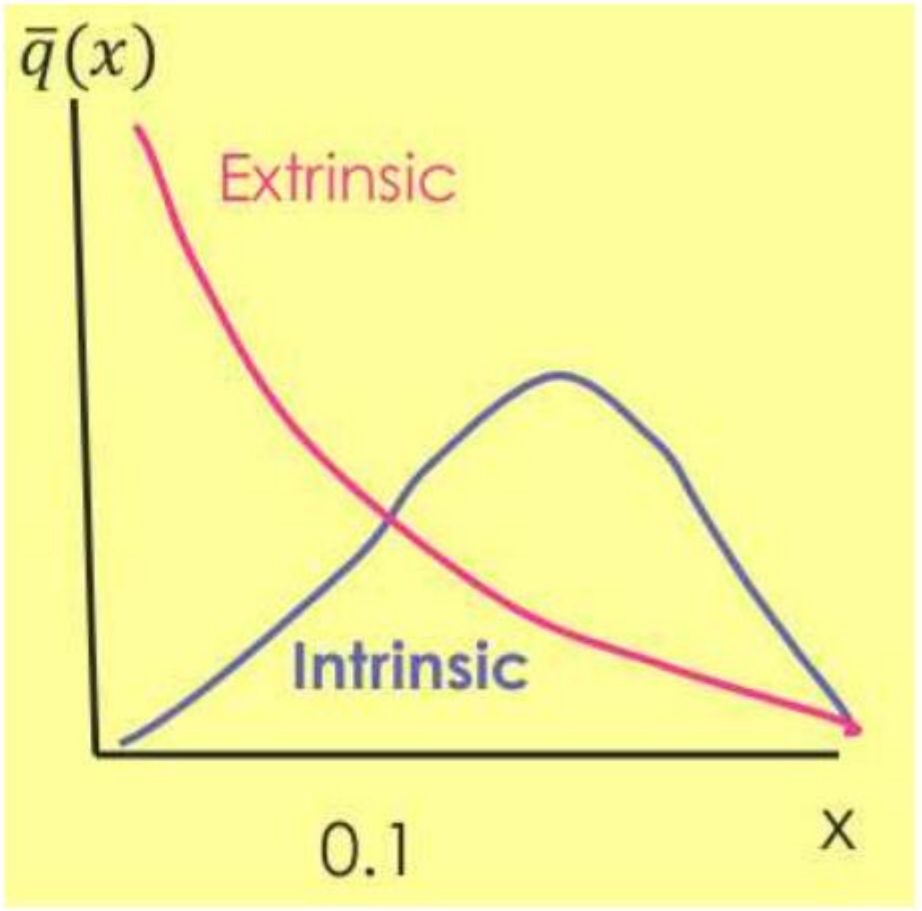
# Extrinsic and intrinsic sea PDFs

## "Extrinsic" sea




## "Intrinsic" sea



# Extrinsic and intrinsic sea PDFs

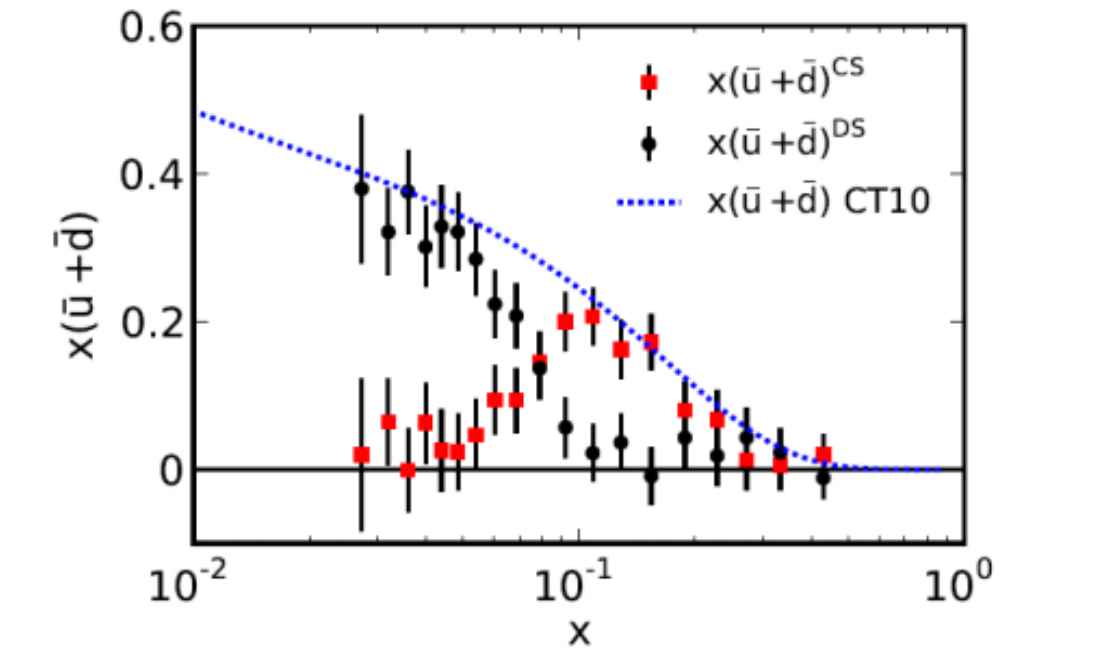



FIG. 5: $x(\bar{u}^{cs}(x)+\bar{d}^{cs}(x))$ obtained from Eq. (1) is plotted together with $x(\bar{u}(x)+\bar{d}(x))$ from CT10 and $\frac{1}{R}x(s(x)+\bar{s}(x))$ which is taken to be $x(\bar{u}^{ds}(x)+\bar{d}^{ds}(x))$.

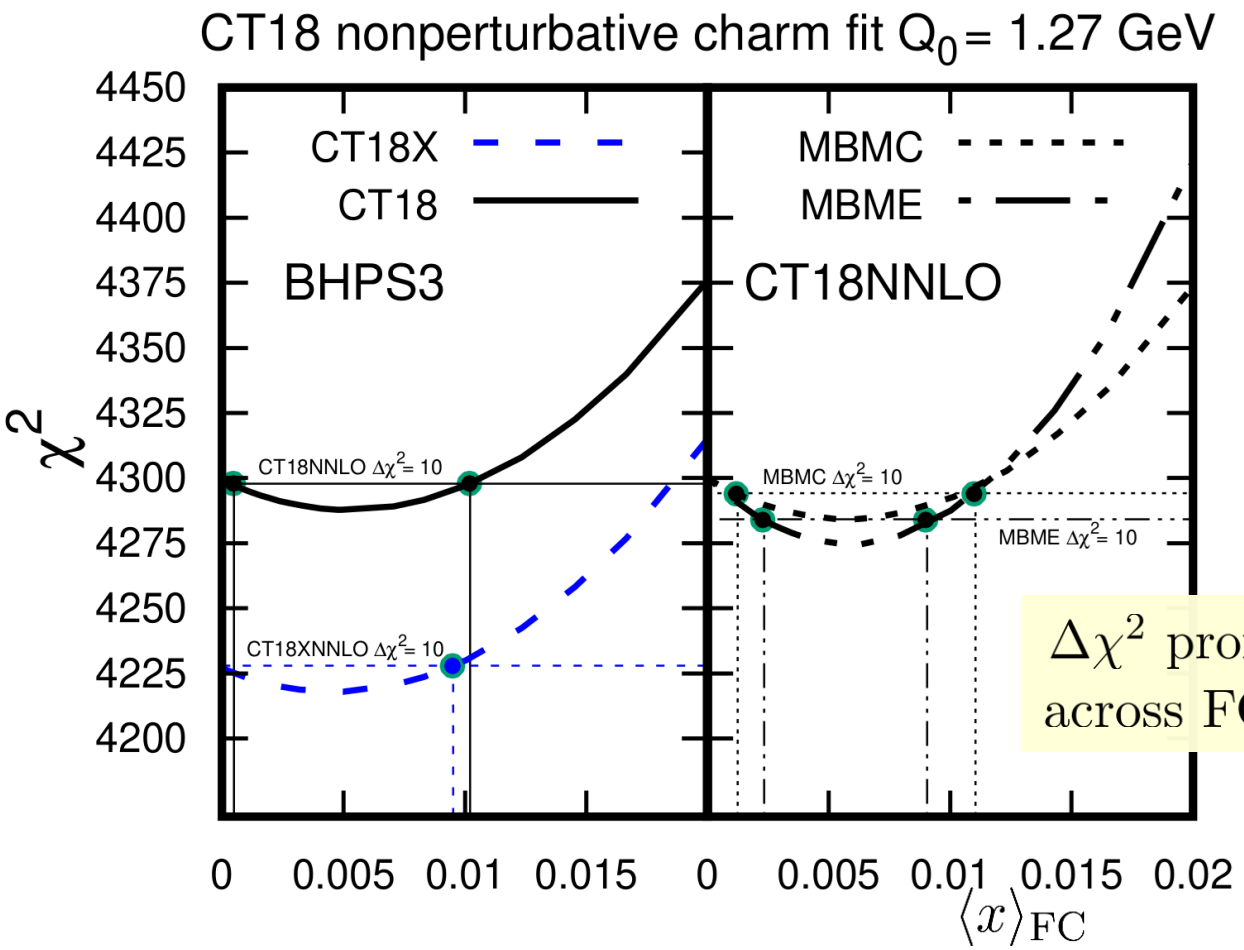
Intrinsic/fitted charm (IC/FC) can carry up to 1% of the proton momentum

Preference for IC in NNPDF4.0 fits, no preference in CT18 FC analysis

Smooth $\bar{u}+\bar{d}$ parametrizations can hide existence of two components
*Liu, Chang, Cheng, Peng, 1206.4339*

# signal for FC in CT18 study, but with shallower $\Delta\chi^2$ than CT14 IC

## CT18 FC NNLO is compatible with zero




$\Delta\chi^2$ profiles similar across FC scenarios

FC uncertainty quantified by normalization via $\langle x_{FC} \rangle$ for each input IC model

- $\langle x \rangle_{FC} \approx 0.5\%$ $(\Delta\chi^2 \gtrsim -25)$ in CT18 FC

  vs.

  $\langle x \rangle_{FC} \approx 0.8 - 1\%$ $(\Delta\chi^2 \gtrsim -40)$ in CT14 IC

$\langle x \rangle_{\mathrm{FC}}$

# Constraints on quark sea from $pN \to \ell^+\ell^- X$ ($N = p,\ d,\ \mathrm{Fe},\ \mathrm{Cu}, \ldots$)

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

⇒ sensitivity to $\bar{q}(x, Q)$

# Constraints on quark sea from $pN \to \ell^+\ell^- X$ $(N = p,\ d,\ \mathrm{Fe},\ \mathrm{Cu}, ..)$

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

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

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

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

# Constraints on quark sea from $pN \to \ell^+\ell^- X$ $(N = p,\ d,\ \mathrm{Fe},\ \mathrm{Cu}, ..)$

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

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

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

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

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

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

# Constraints on quark sea from $pN \to \ell^+\ell^- X$ $(N = p,\ d,\ \mathrm{Fe},\ \mathrm{Cu}, ..)$

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

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

# Experimental evidence for SU(2) symmetry breaking

E866 Drell-Yan pair production:

$\bar{d}(x) - \bar{u}(x) \neq 0$ at $x > 0.1$

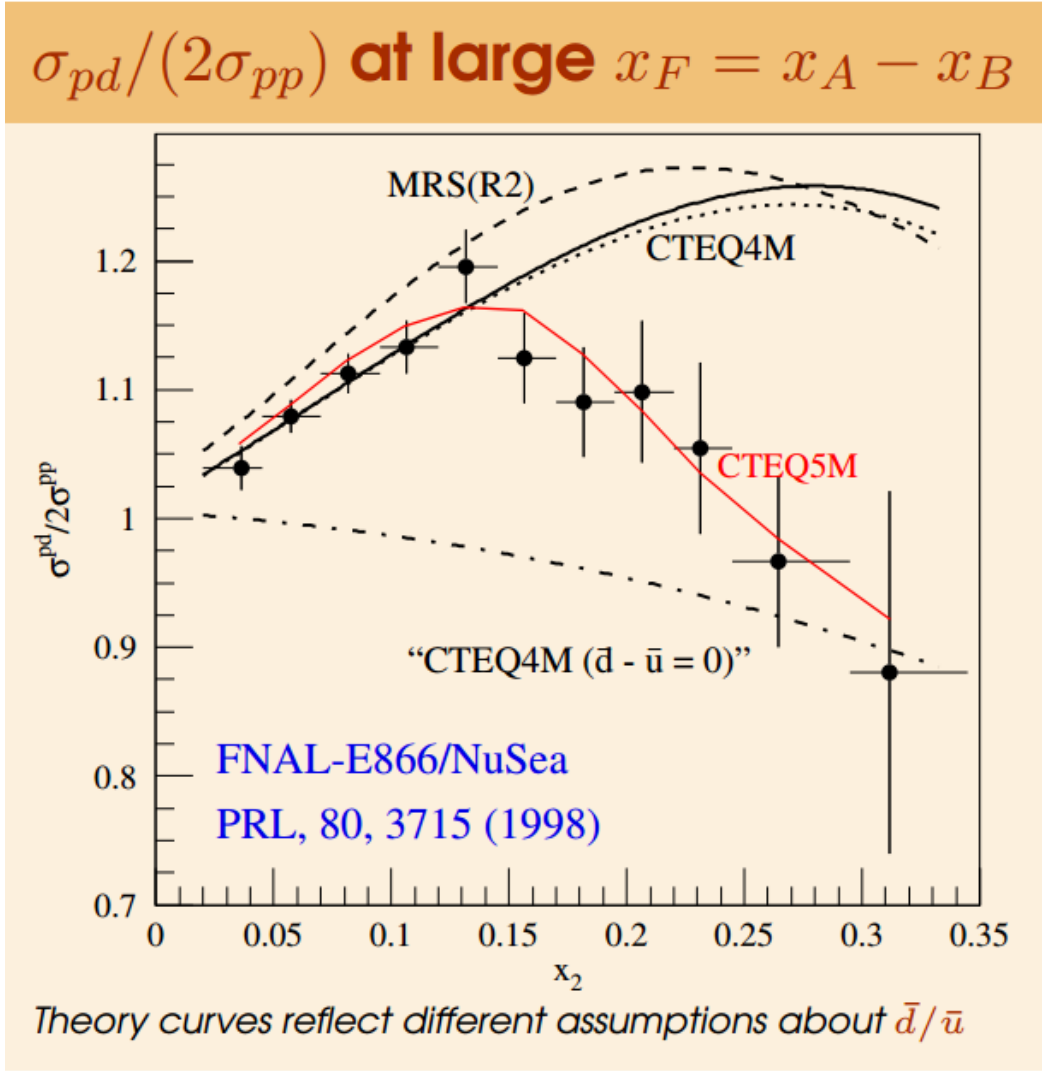
(large difference)

LHC $W/Z$ production:

$\bar{d}(x) - \bar{u}(x) \neq 0$ at $x < 0.1$

(a few percent)

PDF fits (e.g., CTEQ5M) quantitatively account for the violation of $SU(2)$ symmetry in the quark sea



# Charged lepton asymmetry in $AB \to (W \to e\nu_e)X$ $(A, B = p \text{ or } \bar{p})$

$y_e$ and $\eta \approx y_e$ are rapidity and pseudorapidity of an electron from $W$ decay

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

$A_{ch}(y_e)$ relates to the boson asymmetry $A_{ch}(y) = \frac{\left(d\sigma^{W^+}/dy\right) - \left(d\sigma^{W^-}/dy\right)}{\left(d\sigma^{W^+}/dy\right) + \left(d\sigma^{W^-}/dy\right)}$, where

$$\left(d\sigma^{W^+}/dy\right) \propto u_A(x_A, M_W)\bar{d}_B(x_B, M_W) + \bar{d}_A(x_A, M_W)u_B(x_B, M_W) + \ldots$$
$$\left(d\sigma^{W^-}/dy\right) \propto \bar{u}_A(x_A, M_W)d_B(x_B, M_W) + d_A(x_A, M_W)\bar{u}_B(x_B, M_W) + \ldots$$

# Charged lepton asymmetry in $AB \to (W \to e\nu_e)X$ $(A, B = p \text{ or } \bar{p})$

$y_e$ and $\eta \approx y_e$ are rapidity and pseudorapidity of an electron from $W$ decay

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

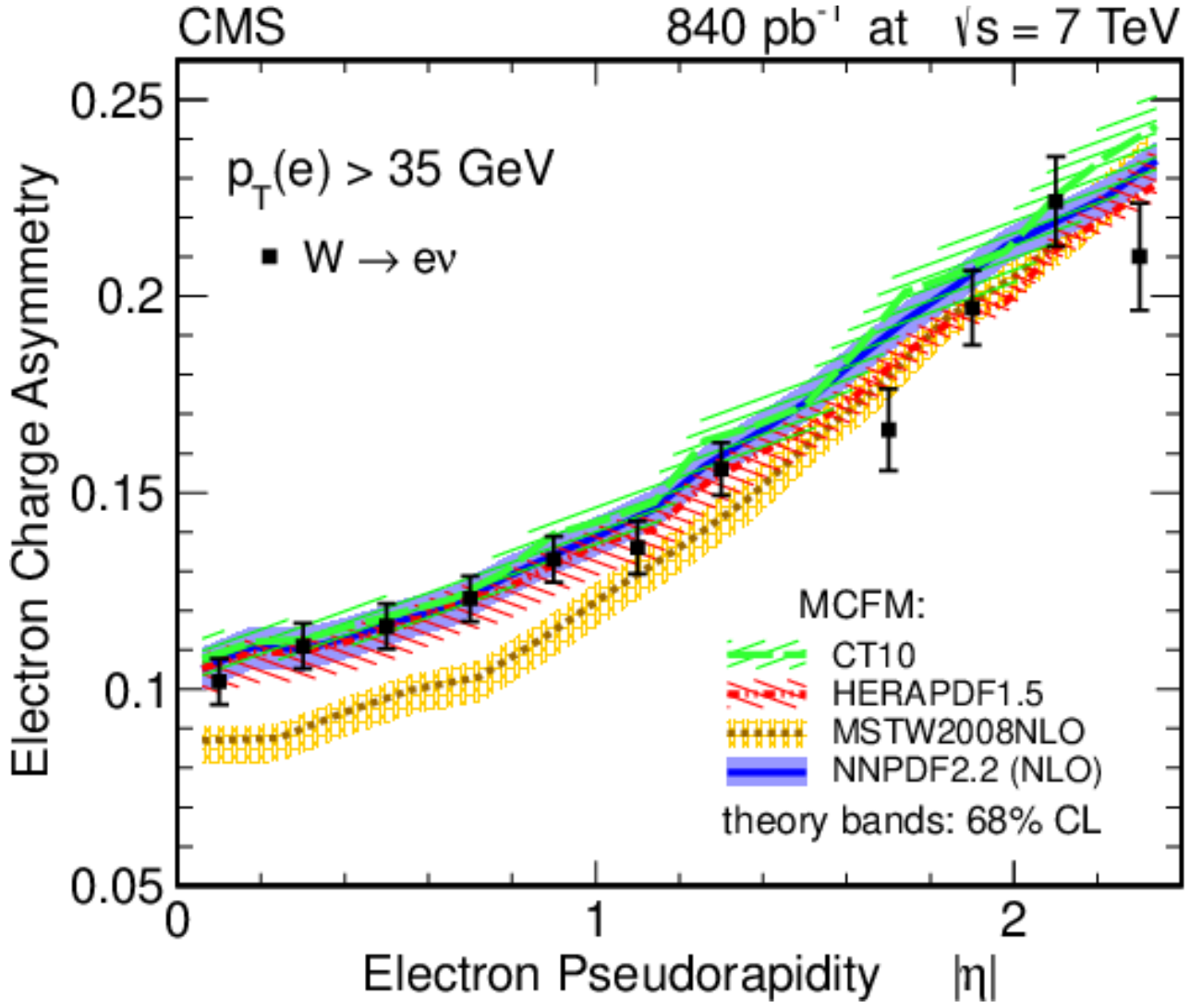
$\therefore A_{ch}(y_e)$ constrains PDF ratios at $Q \approx M_W$:

- $d/u$ at $x \to 1$ at the Tevatron 1.96 TeV $(p\bar{p})$;
- $d/u$ at $x > 0.1$ **and** $\bar{u}/\bar{d}$ at $x \sim 0.01$ at the LHC $(pp)$

# Charge asymmetry at the Tevatron and LHC

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

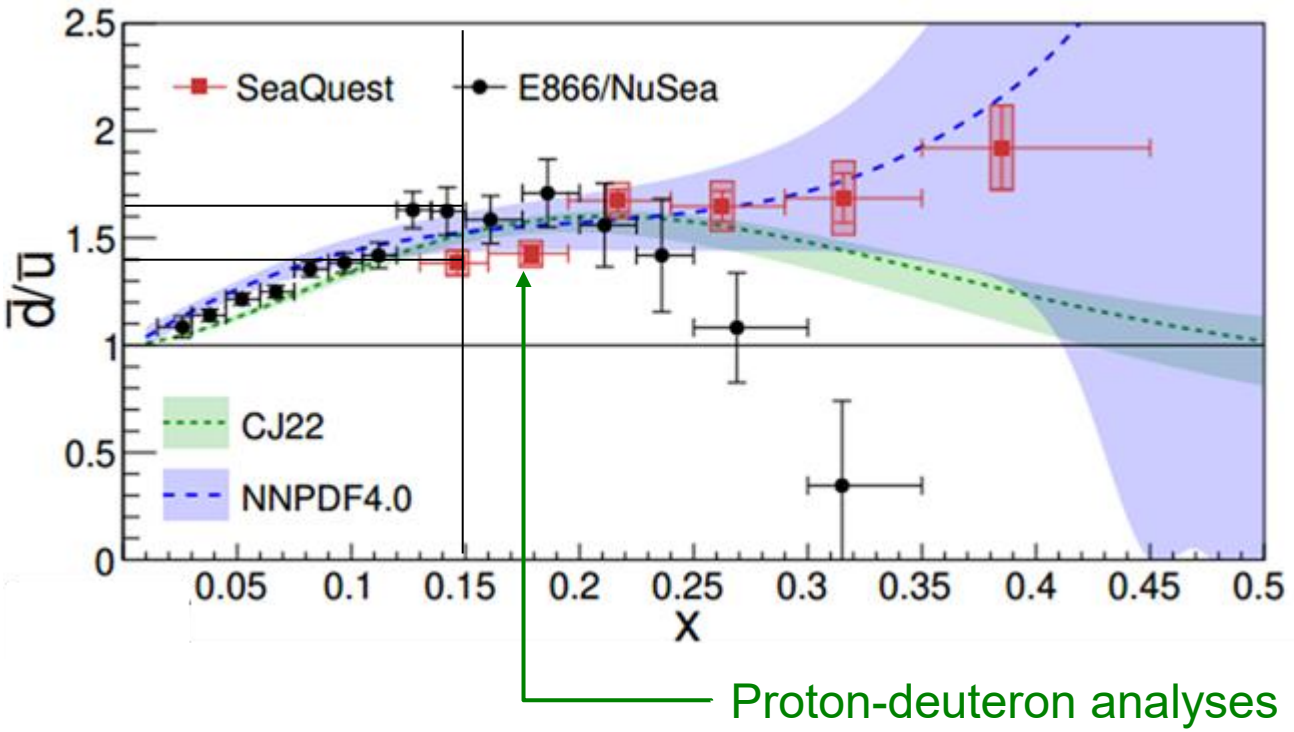
The asymmetry is sensitive to down-strange separation



# 2025: How big is the $\bar{d} - \bar{u}$ asymmetry in a proton?

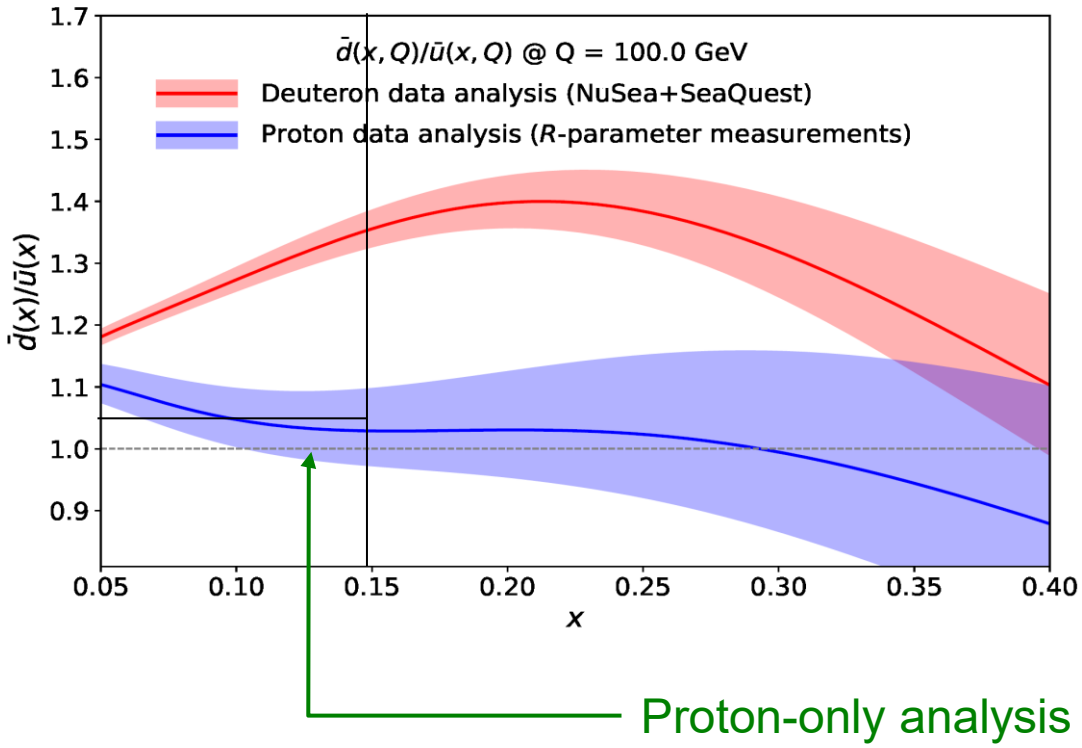
Final E906/SeaQuest $\sigma_{pd}/(2\sigma_{pp})$ DY ratio

arXiv:2512.17564




CMS, forward-backward asymmetry of Drell-Yan pairs

W. Ma et al., arXiv: 2510.08941



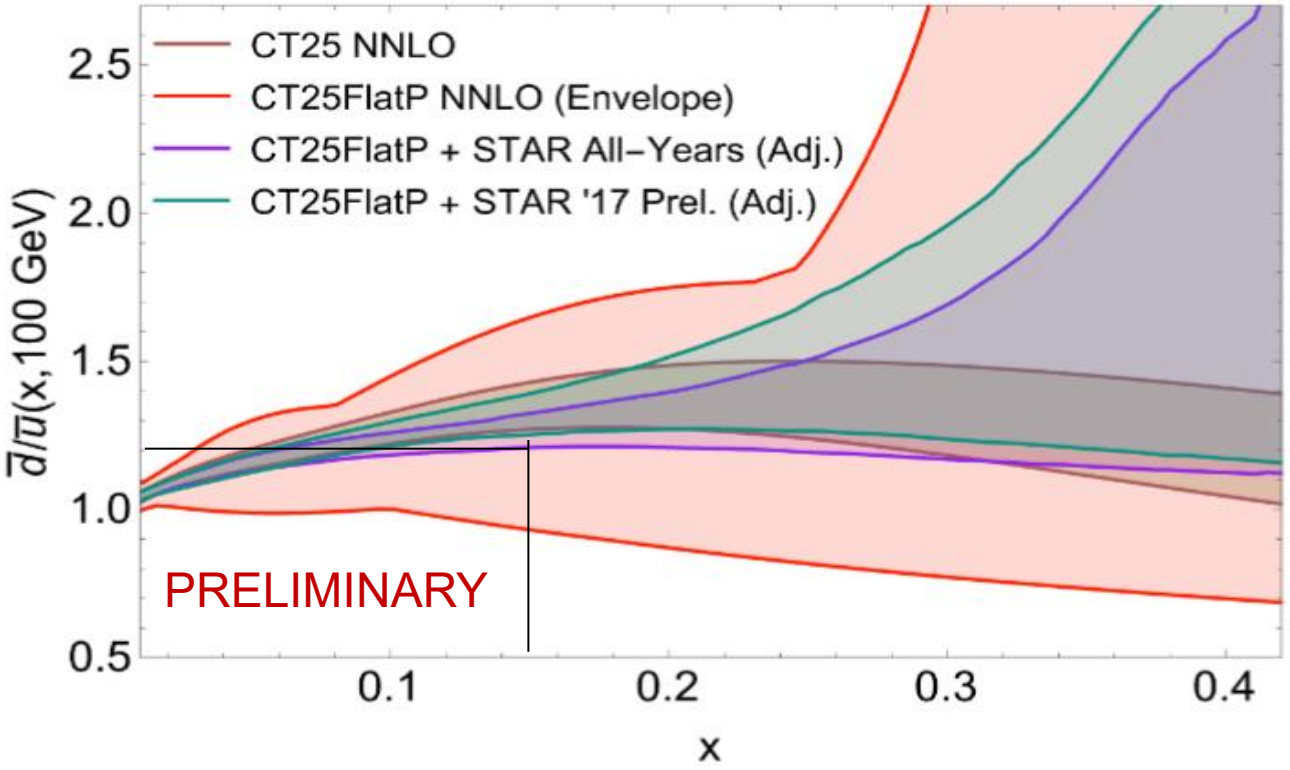

Note: the E906 $\bar{d}/\bar{u}$ extraction depends on assumed $\bar{d} + \bar{u}$ at $x > 0.1$, which differs considerably among PDF groups

$\bar{d}/\bar{u} \approx 1$ at $x \approx 0.15$

Uncertainty bands are displayed for $\Delta\chi^2 = 1$ with one PDF parametrization choice

# Reweighting by $W^+/W^-$ cross section ratio from STAR

M. Ponce Chavez, arXiv:2606.xxxxx



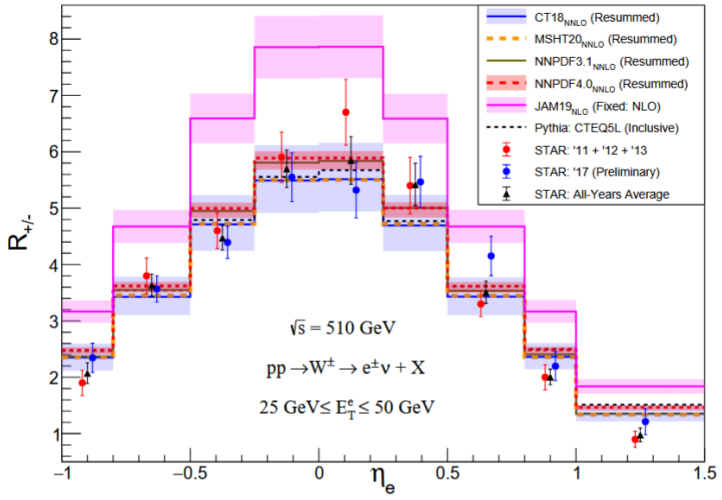




Figure 6: Experimental and resummed $\sigma_{W^+}/\sigma_{W^-}$ cross section ratios for various PDF sets. Uncertainty bands are given at 68% confidence level. Systematic and statistical uncertainties are combined in quadrature for all data points, and the veto to inclusive correction is not applied.

CT25FlatP proton-only PDF set is not biased by the deuteron data ⇒ only weak constraint on $\bar{d}/\bar{u}$

Reweighting with tentative 2017 data ($\Delta\chi^2 = 1$), with many PDF parametrizations, prefers $\bar{d}/\bar{u} \approx 1.3$ at $x = 0.15$

**How big is the $\bar{d} - \bar{u}$ asymmetry in a proton?**

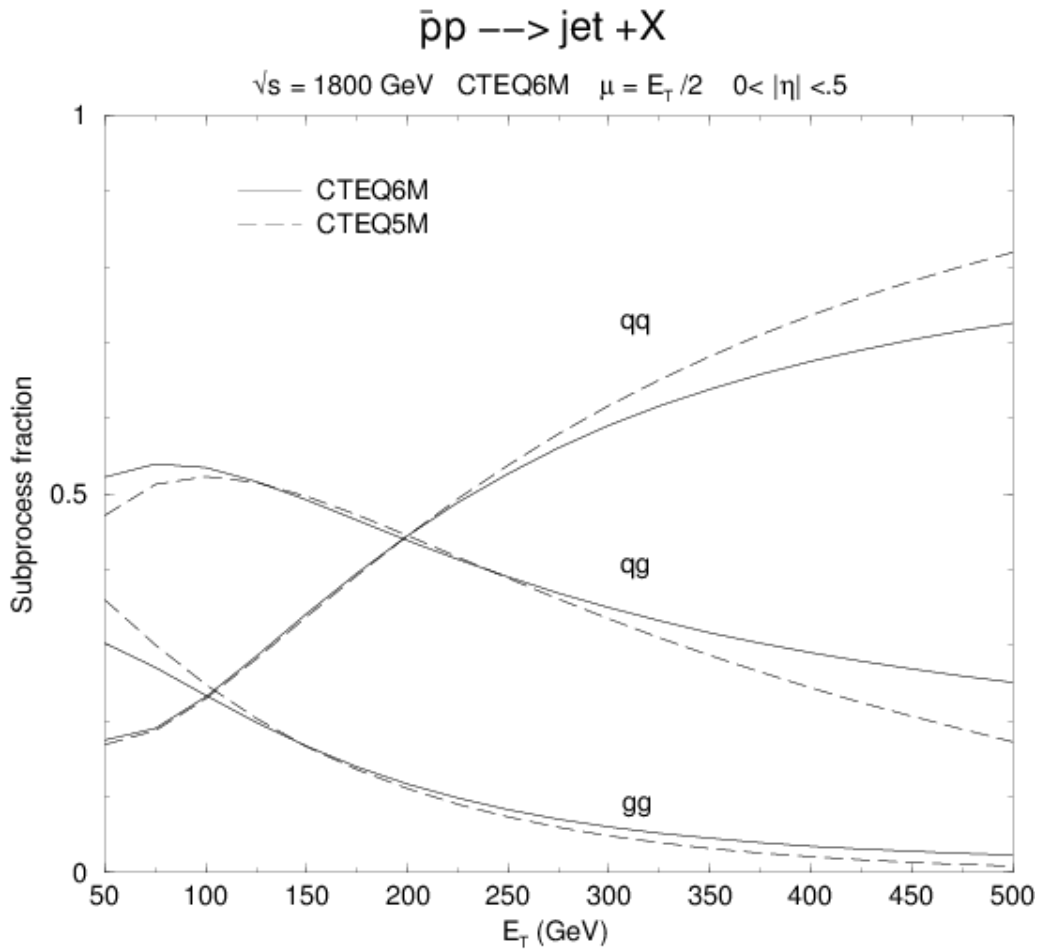
We will know after including E906 (2025), LHC $A_{FB}$, RHIC $W^+/W^-$ ratio in the same global fit

# Inclusive jet production, $pp \to \text{jet} + X$

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

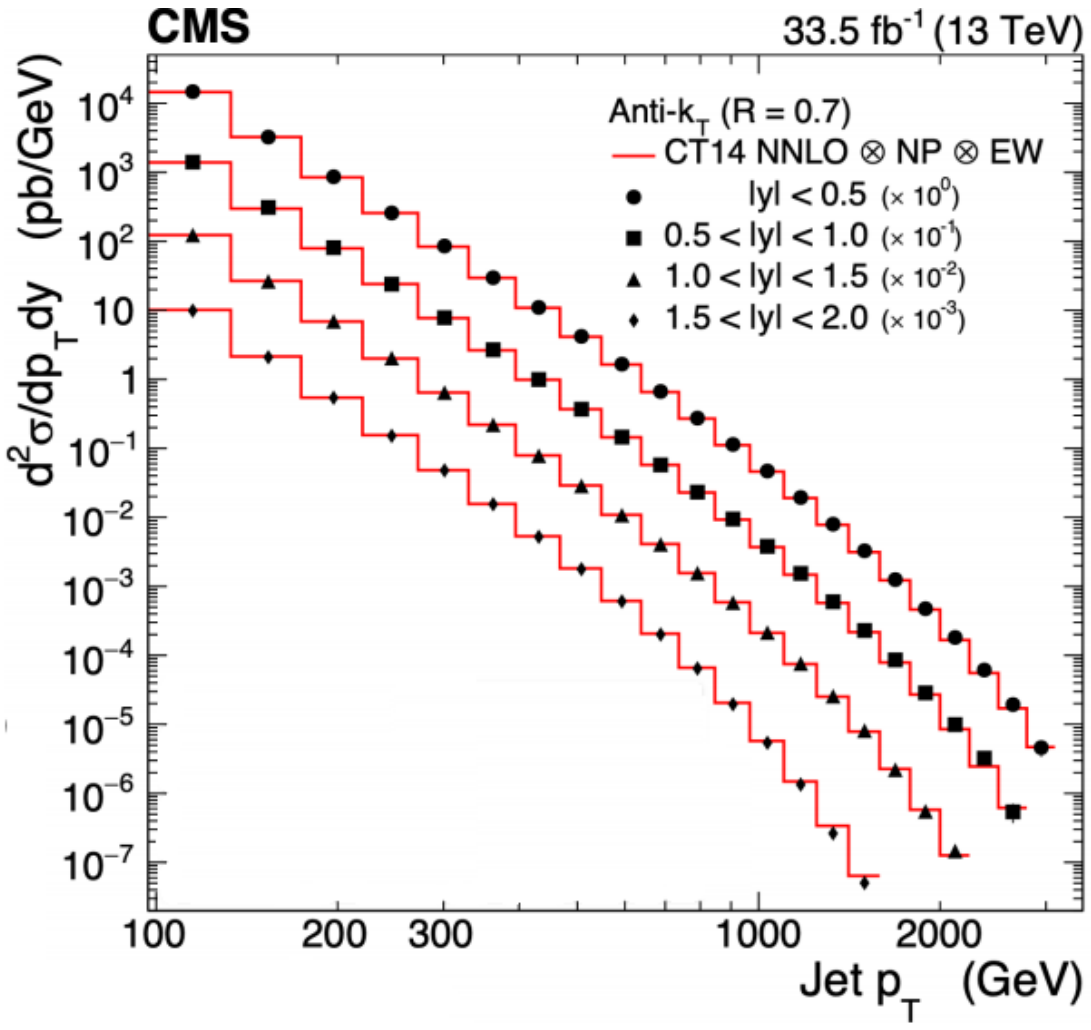
Here typical $x$ is of order $2E_T/\sqrt{s} \gtrsim 0.1$; e.g., $x \approx 0.2$ for $E_T = 200$ GeV, $\sqrt{s} = 1.8$ TeV

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



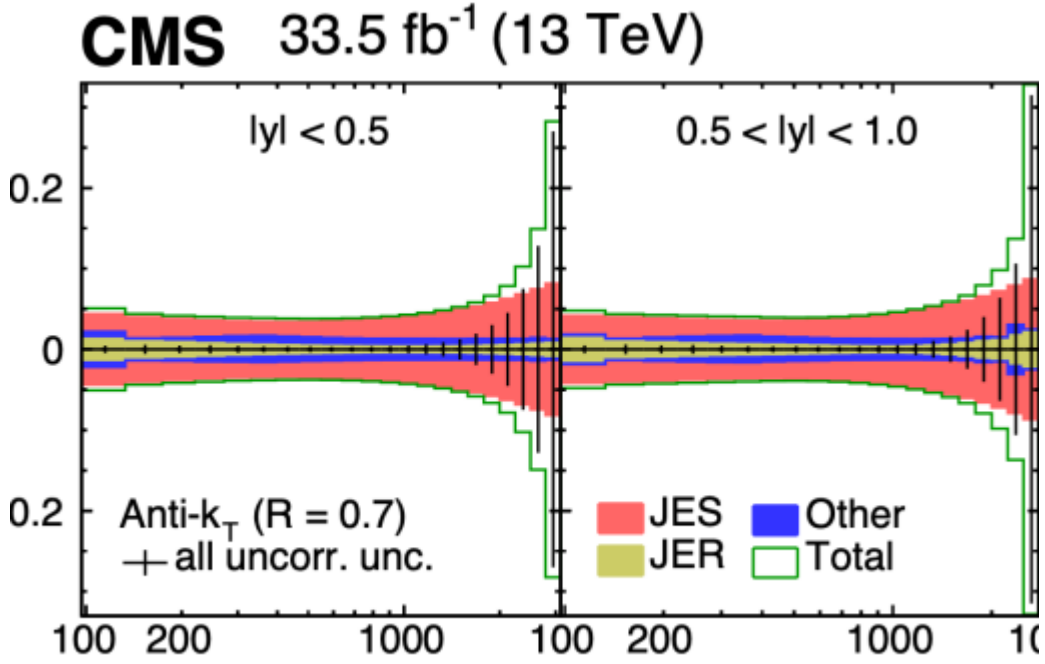
# Inclusive jet and dijet production at LHC 13 TeV

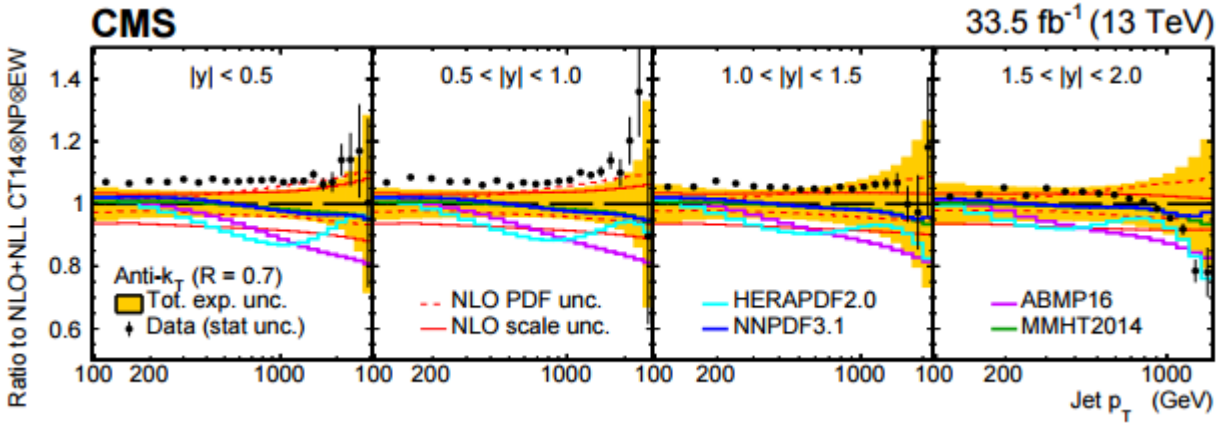
- The cross sections span 13 orders of magnitude
- (Almost) negligible statistical error



# Inclusive jet and dijet production at LHC 13 TeV

- The cross sections span 13 orders of magnitude
- (Almost) negligible statistical error
- Systematic uncertainties dominate, both from the experiment (up to 90 correlated sources of uncertainty) and NLO theoretical cross section (QCD scale dependence)
- The PDF uncertainty would be strongly underestimated if these systematic errors are not included

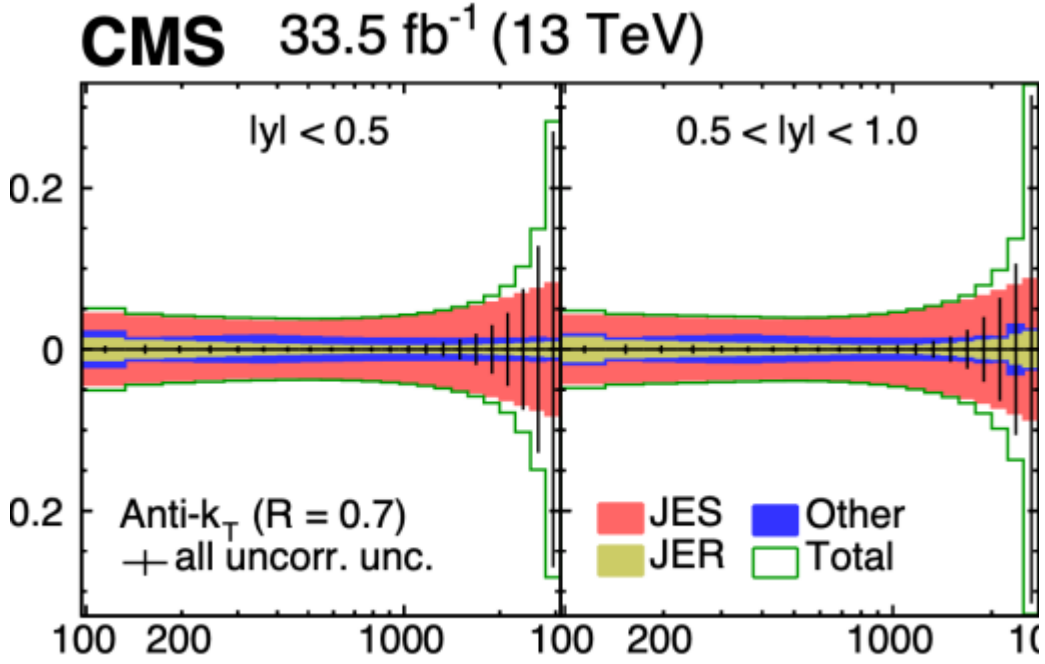


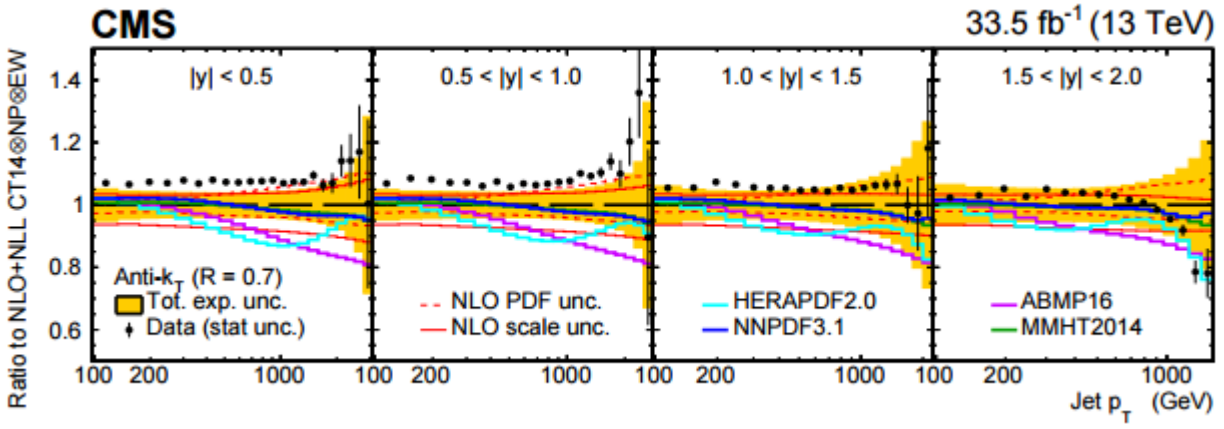



# Inclusive jet and dijet production at LHC 13 TeV

- The cross sections span 13 orders of magnitude
- (Almost) negligible statistical error

- Lecture 2 will discuss how to include the correlated systematic errors into the PDF analysis



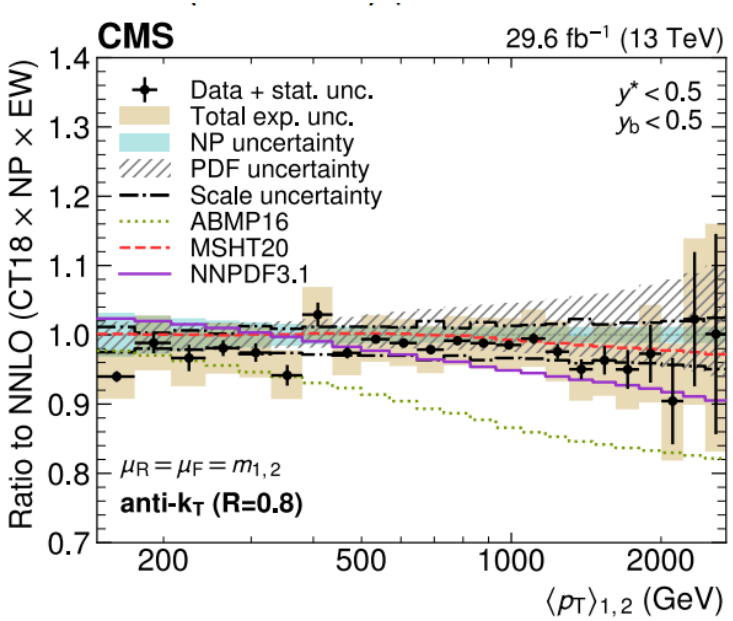






FIGURES: ../eps/jets/CMS7/InclusiveSpectrum

# inclusive jet vs. dijet data sets

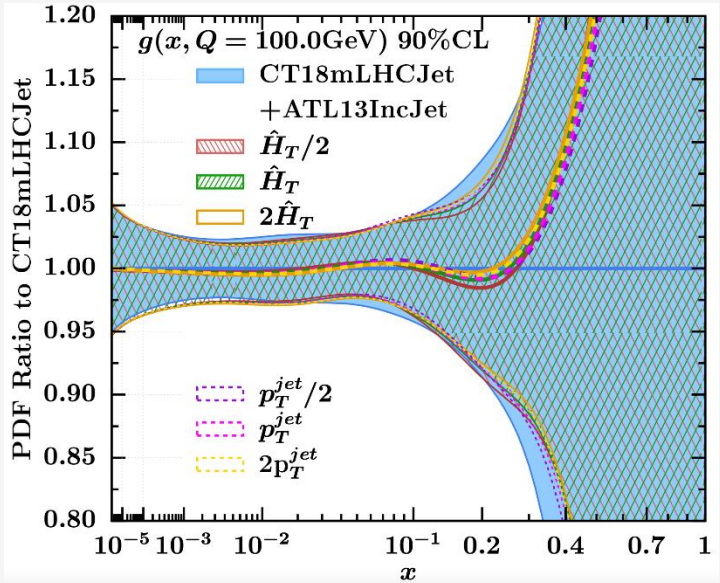
Several PDF fits use dijet datasets



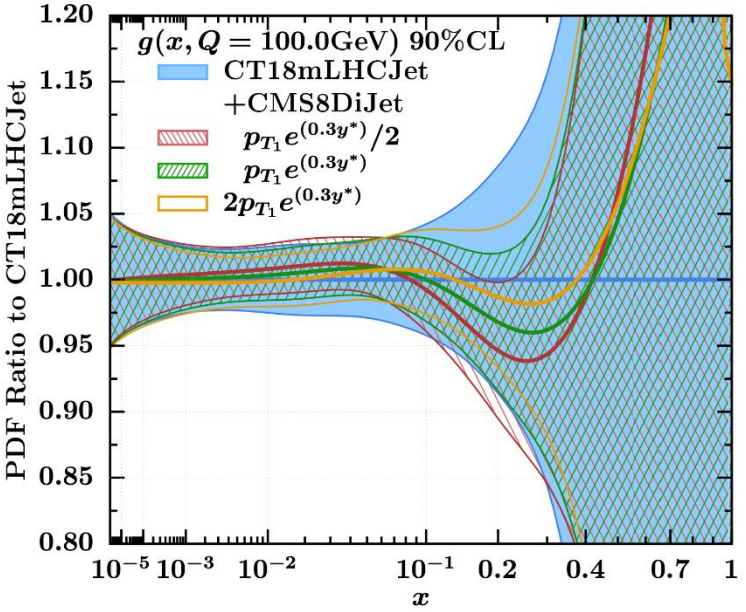

Ablat et al., PRD111 (2025) 3, 036033

- Impact of LHC inclusive jets on $g(x,Q)$ is relatively independent of the scale choice. **Impact of dijets substantially depends on scale choices**

**+ inclusive jets: small scale dependence, a harder $g(x,Q)$**




**+ dijets: significant scale, dependence, varied pulls on $g(x,Q)$**



# Recap

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

… are nonperturbative QCD functions describing the structure of hadrons in high-energy scattering according to the method of QCD factorization

… are related to probabilities for finding partons inside parent hadrons

… cannot be computed systematically

… are universal – independent of the hard-scattering process

… obey perturbative evolution (DGLAP) equations

… are determined from select hadronic experiments, used to make predictions for other experiments

# Homework assignment
# Electron-neutron deep inelastic scattering

Electron—neutron DIS experiments are harder to do than electron-proton experiments because one cannot make a target of free neutrons. Nevertheless, the essential data can be inferred from electron—deuteron scattering, and it is found that

$$\int_0^1 dx\ F_2^{ep}(x,Q) \approx 0.18, \int_0^1 dx\ F_2^{en}(x,Q) \approx 0.12 \quad at\ Q \sim 1\ \mathrm{GeV}.$$

Let $p_q \equiv \int_0^1 dx\, x\big(f_{q/h}(x) + f_{\bar{q}/h}(x)\big)$ denote the momentum fraction of a proton carried by quarks and antiquarks of flavor $q$. The momentum sum rule for a proton states that $p_u + p_d + p_s + \cdots + p_g = 1$.

1. Working at the leading order in $\alpha_s$ ,use these results to show that, if we keep only $u$ and $d$ (anti)quarks, we get $p_u = 2p_d$ .Find $p_u$ and $p_d$ .
   [Hint: how are $f_{u/n}(x)$ and $f_{d/n}(x)$ related to $f_{u/p}(x) \equiv u(x)$ and $f_{d/p}(x) \equiv d(x)$? ]
2. Suppose you also know that $p_s \approx 3\%$ .What fraction of the proton's momentum is carried by the gluon?

# Backup slides

# PDFs for heavy flavors

PDFs for heavy partons $h$ can be generated via DGLAP evolution at $Q \geq m$, using a boundary condition $f_{h/p}(x, Q) = 0$ at $Q \leq m$

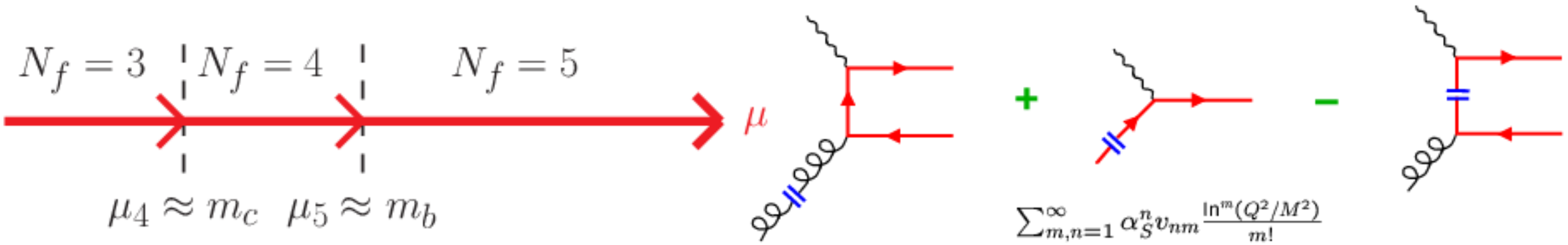
In practice:

- PDFs are usually introduced for $c$ and $b$ quarks
- starting from $O(\alpha_s^2)$, an initial condition $f_{c/p}(x, Q_0) \neq 0$ can be generated at $Q_0 = m_c$ from twist-4 intrinsic charm DIS terms *(arXiv:1707.00657)*
- QCD coupling $\alpha_s(Q)$ and PDFs are evaluated with 5 active flavors at all $Q \geq m_b$
- Logarithmic enhancements may exist in collinear $t$, $W$, $Z$ production at $Q \gtrsim 1$ TeV; PDFs for $t$, $W$, $Z$ ``partons'' may be introduced at such $Q$

# General-mass variable-flavor number scheme

A series of factorization schemes with $N_f$ active quark flavors in $\alpha_s(Q)$ and $f_{a/p}(x,Q)$ - $N_f$ is incremented sequentially at momentum scales $\mu_{N_f} \approx m_{N_f}$

- incorporates essential $m_{c,b}$ dependence near, and away from, heavy-flavor thresholds
- implemented in all latest PDF fits except ABM

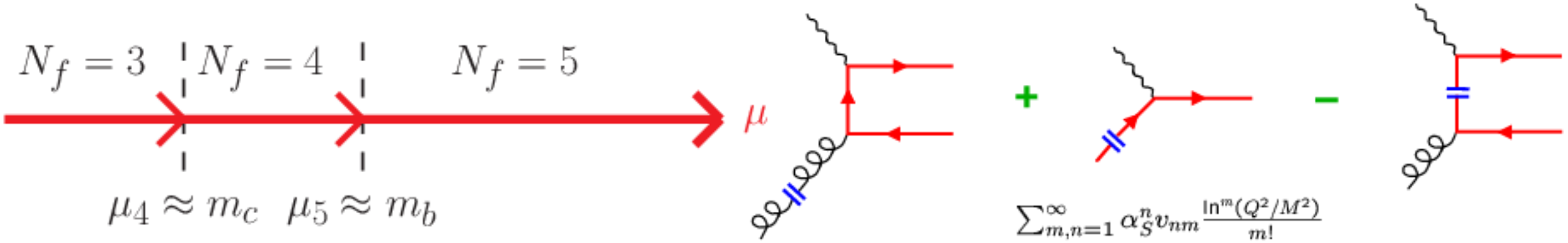


# General-mass variable-flavor number scheme

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

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

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

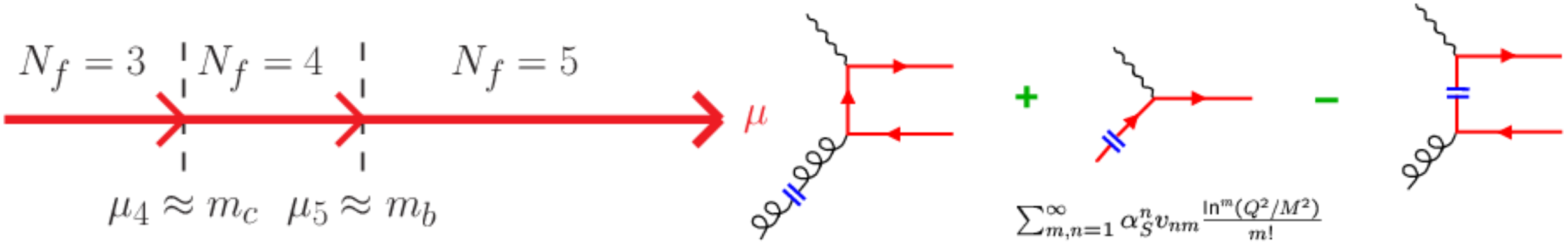


# General-mass variable-flavor number scheme

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

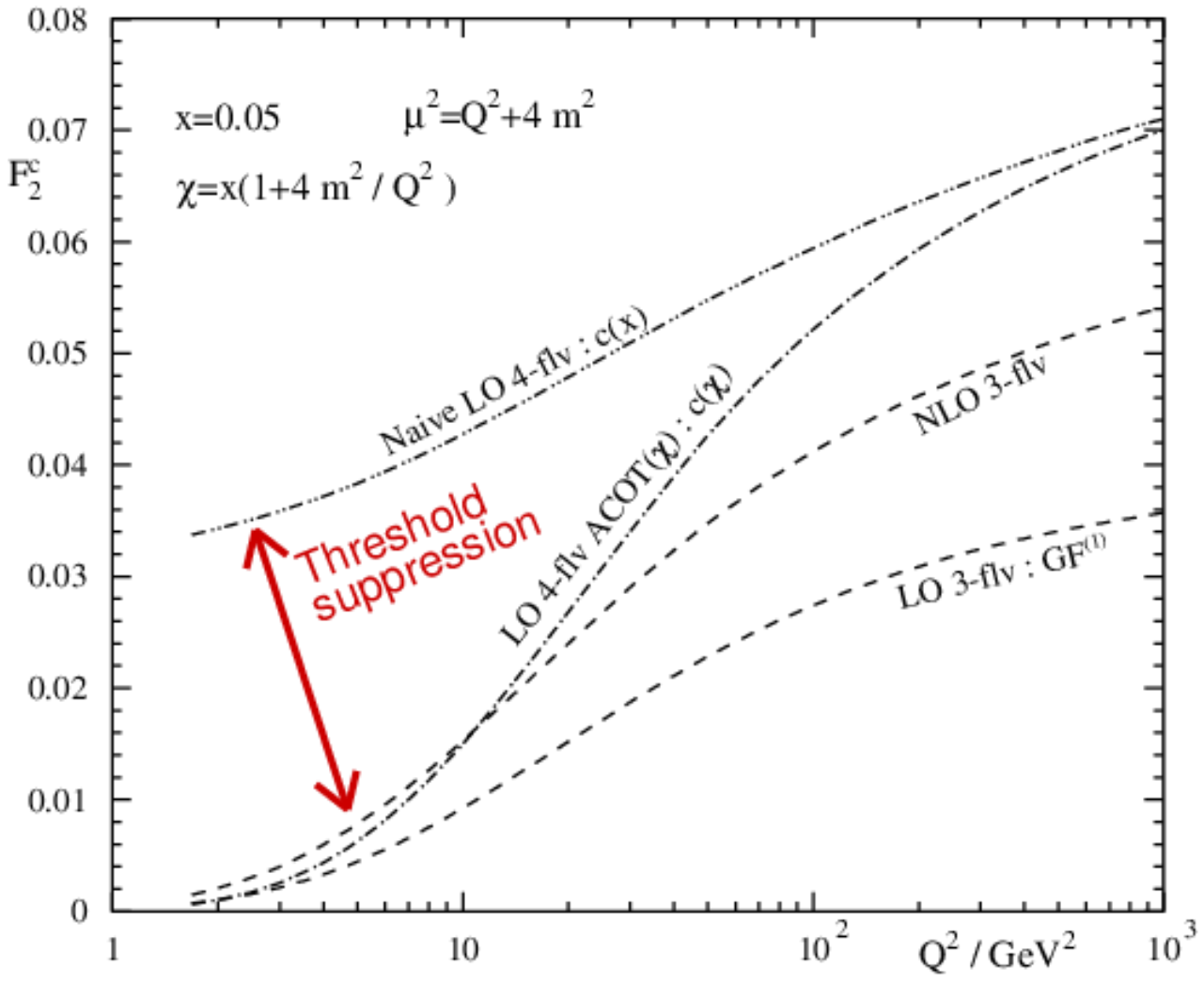
$$F_2(x,Q,m_c) = \sum_a \int_\chi^1 \frac{d\xi}{\xi} C_a\left(\frac{\chi}{\xi}, \frac{Q}{\mu}, \frac{m_c}{Q}\right) f_a\left(\xi, \frac{\mu}{m_c}\right) + \mathcal{O}\left(\frac{\Lambda_{QCD}}{Q}\right)$$

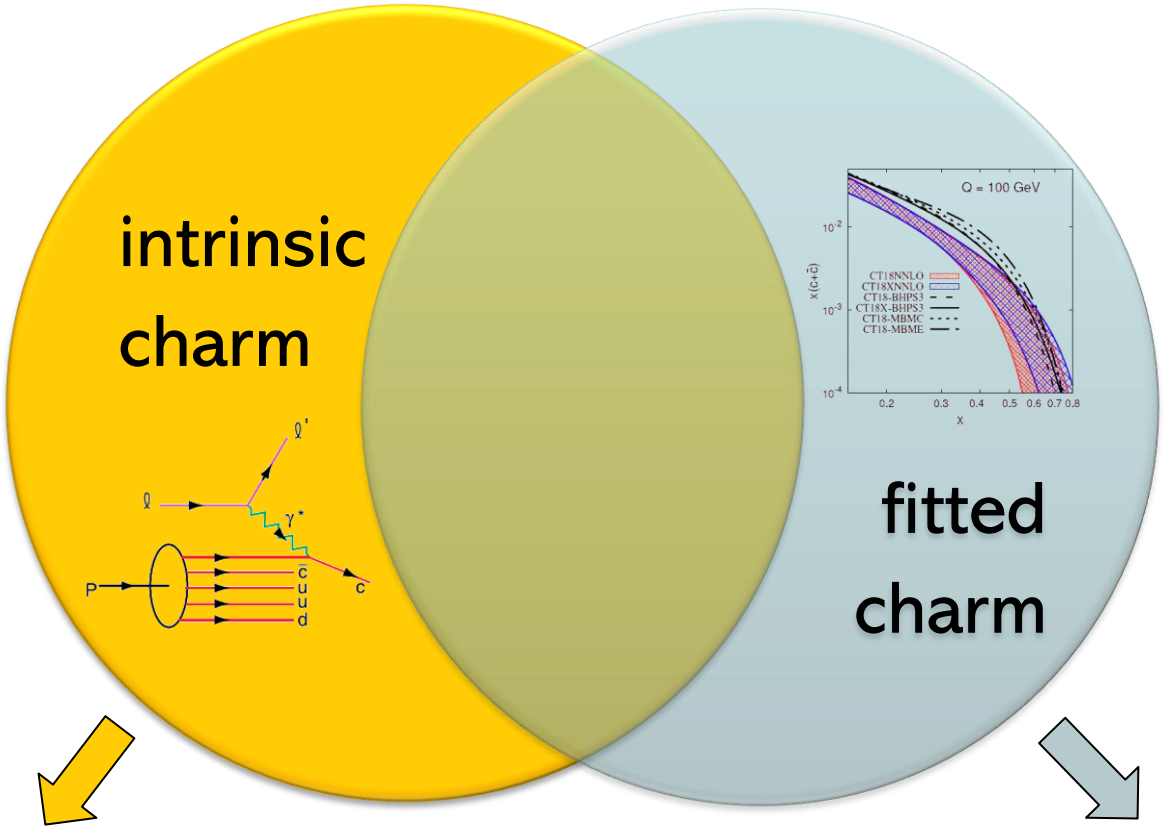
Works most effectively in DIS and Drell-Yan-like processes; practical implementation requires - efficient treatment of mass dependence, rescaling of momentum fractions $\chi$ in processes with incoming $c$, $b$ - physically motivated factorization scale to ensure fast PQCD convergence (e.g., $\mu = Q$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



## An example of a GM-VFN factorization scheme

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






- The concept of nonperturbative methods
- Can refer to a component of the hadronic Fock state or the type of the hard process
- Predicts a typical enhancement of the charm PDF at $x \gtrsim 0.2$

Connection?

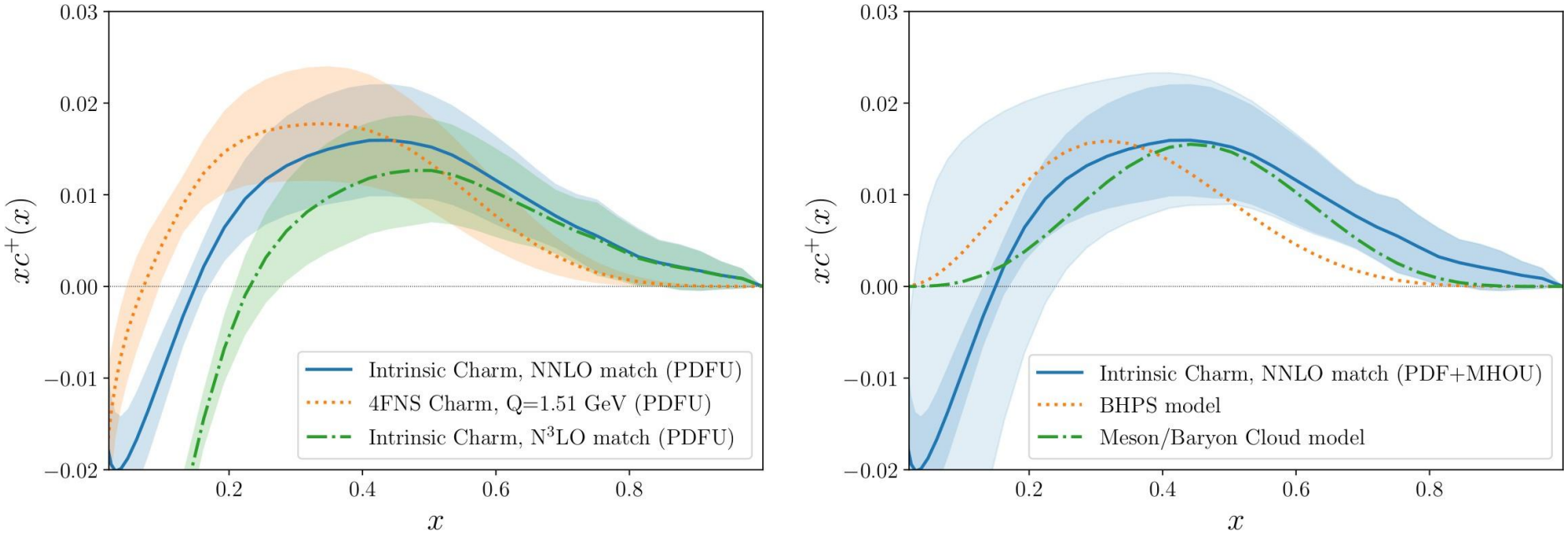
- A charm PDF parametrization at scale $Q_0 \approx 1$ GeV found by global fits [CT, NNPDF, …]
- Arises in perturbative QCD expansions over $\alpha_s$ and operator products
- May absorb process-dependent or unrelated radiative contributions

# NNPDF IC PDFs and moments

*NNPDF, Nature 608 (2022) 7923, 483.*

**Preference for non-zero NNPDF IC PDF at $0.2 \lesssim x \lesssim 0.7$**

Large perturbative instability from MHOU in DGLAP affects $x < 0.2$




→ MHOU excluded to obtain a nominal charm fraction, $\langle x \rangle_{\rm FC} = 0.62 \pm 0.28\%$

→ if MHOU is included, consistency with zero: $\langle x \rangle_{\rm FC} = 0.62 \pm 0.61\%$

# CT18 FC charm PDFs *M. Guzzi, T. J. Hobbs, K. Xie, et al., PLB 843 (2023) 137975; update on T.-J. Hou et al. JHEP* 02 (2018) 059

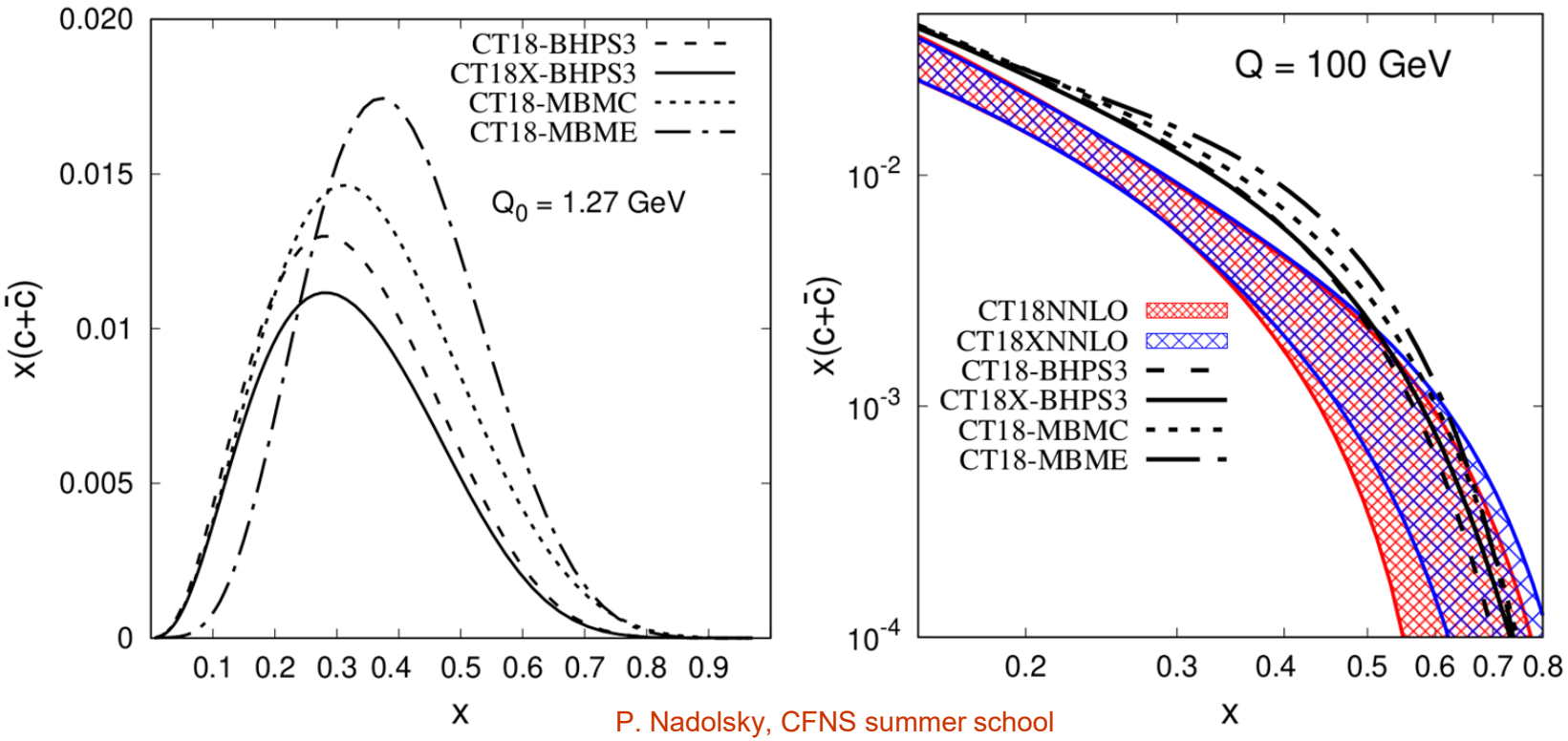
FC scenarios traverse range of high-*x* behaviors from IC models

→ fit implementation of BHPS from CT14IC (BHPS3) on CT18 or CT18X (NNLO)

→ fit two MBMs: MBMC (confining), MBME (effective mass) on CT18

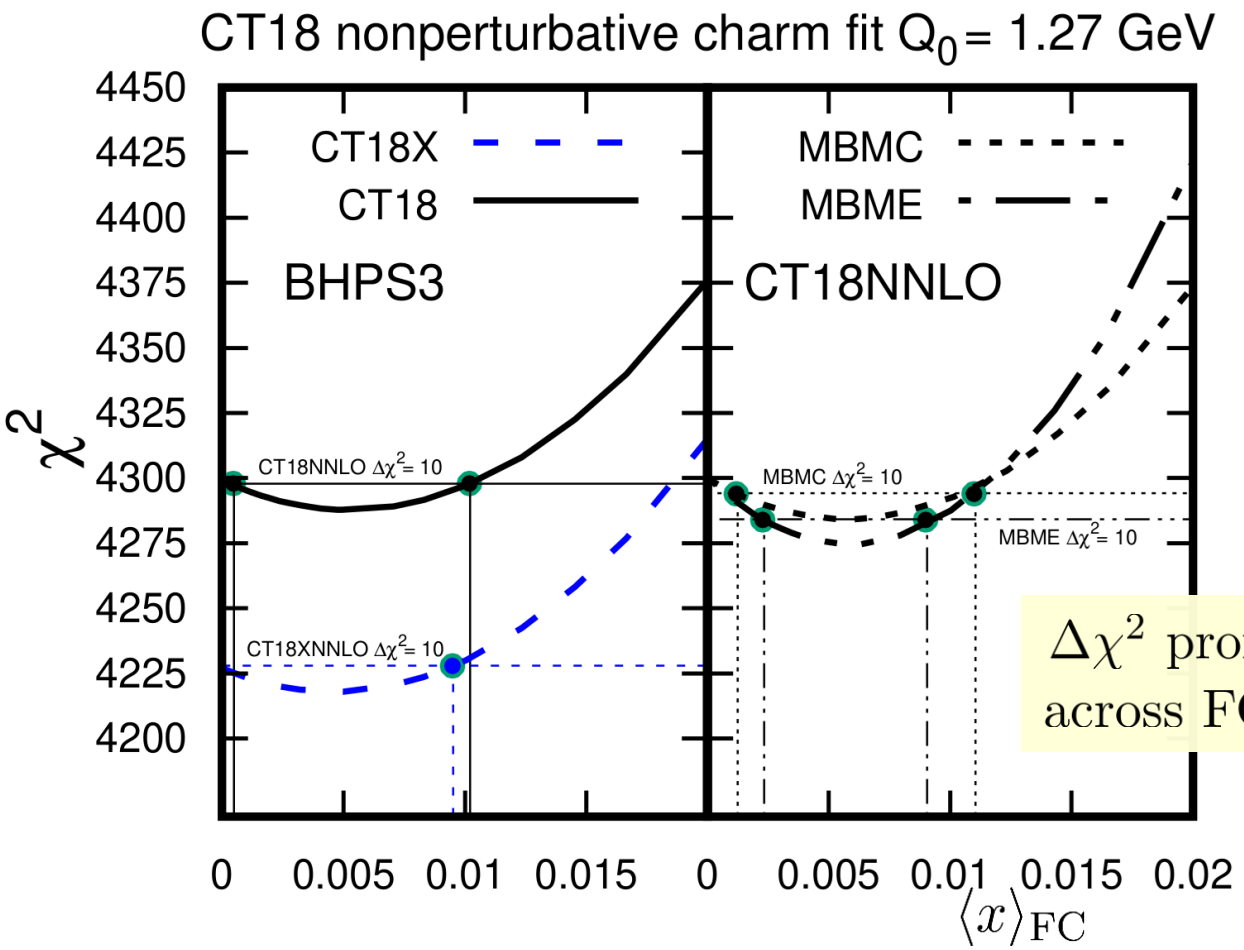
investigate constraints from newer LHC data in CT18

central fits



# signal for FC in CT18 study, but with shallower $\Delta\chi^2$ than CT14 IC

## CT18 FC NNLO is compatible with zero




FC uncertainty quantified by normalization via $\langle x_{FC} \rangle$ for each input IC model

- $\langle x \rangle_{FC} \approx 0.5\%$ $(\Delta\chi^2 \gtrsim -25)$ in CT18 FC

  vs.

  $\langle x \rangle_{FC} \approx 0.8 - 1\%$ $(\Delta\chi^2 \gtrsim -40)$ in CT14 IC

$\Delta\chi^2$ profiles similar across FC scenarios

$\langle x \rangle_{\rm FC}$

# FC PDF moments

**even restrictive uncertainties give moments consistent with zero**

→ broaden further for default CT tol.

→ lattice may give $\langle x \rangle_{c+}$, $\langle x^2 \rangle_{c-}$

$$\langle x \rangle_{\rm FC} \equiv \langle x \rangle_{c+}[Q_0 = 1.27\,{\rm GeV}]$$

$$= 0.0048\,^{+0.0063}_{-0.0043}\,(^{+0.0090}_{-0.0048}),\ \text{CT18 (BHPS3)}$$

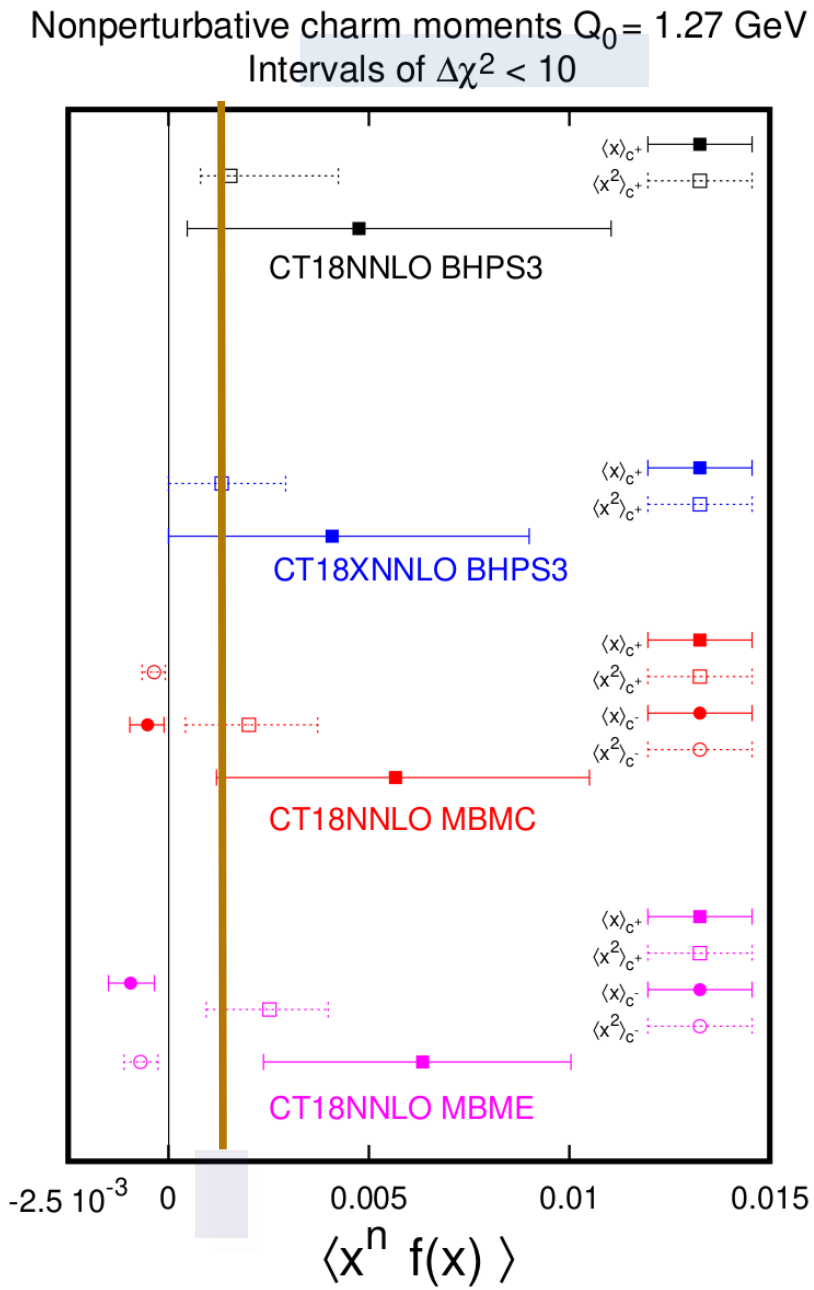
$$= 0.0041\,^{+0.0049}_{-0.0041}\,(^{+0.0091}_{-0.0041}),\ \text{CT18X (BHPS3)}$$

$$= 0.0057\,^{+0.0048}_{-0.0045}\,(^{+0.0084}_{-0.0057}),\ \text{CT18 (MBMC)}$$

$$= 0.0061\,^{+0.0030}_{-0.0038}\,(^{+0.0064}_{-0.0061}),\ \text{CT18 (MBME)}$$

$\Delta\chi^2 \leq 10$ (restrictive tolerance)

$\Delta\chi^2 \leq 30$ (~CT standard tolerance)

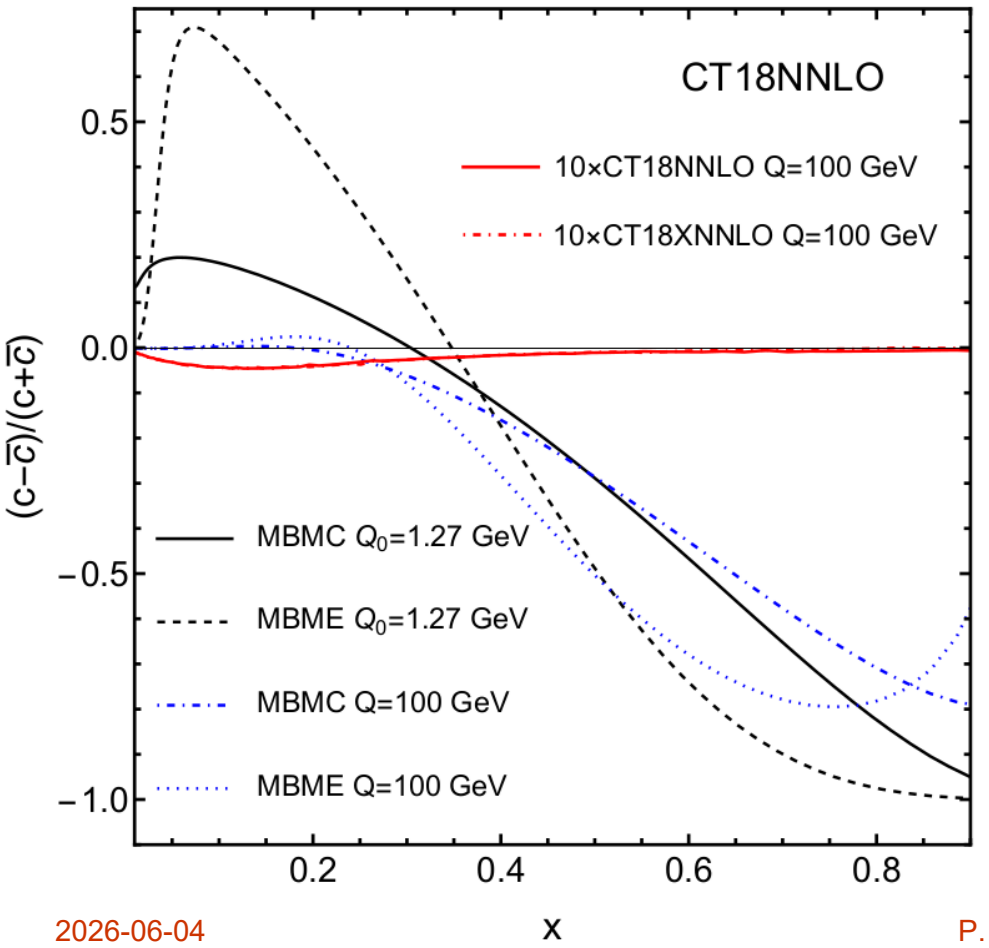


# Charm-anticharm asymmetries can be a robust signal of nonperturbative charm

pQCD only very weakly breaks $c = \bar{c}$ through HO corrections

→ large(r) charm asymmetry would signal nonpert dynamics, IC

→ MBM breaks $c = \bar{c}$ through hadronic interactions

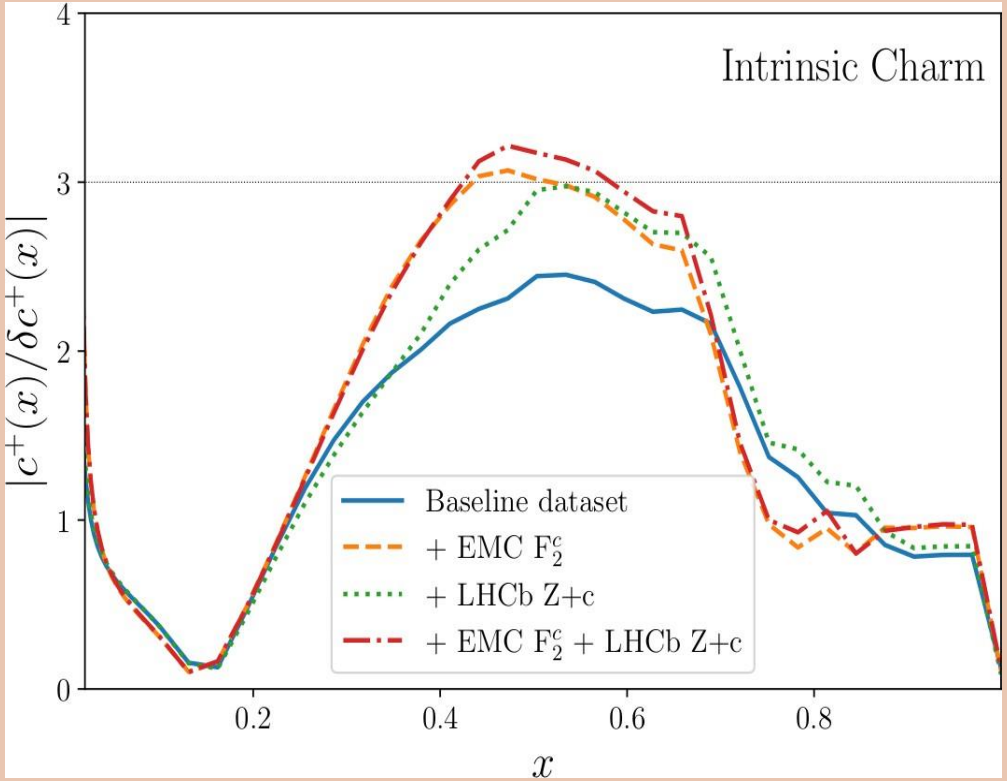



consider two MBM models as ***examples*** (<u>not predictions</u>)

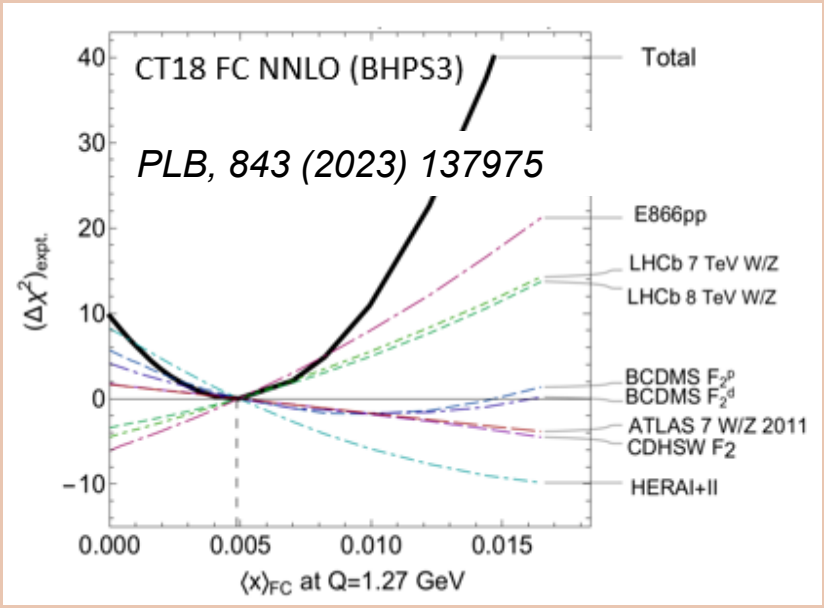
→ asymptotically small, but ratio can be bigger; will be hard to extract from data

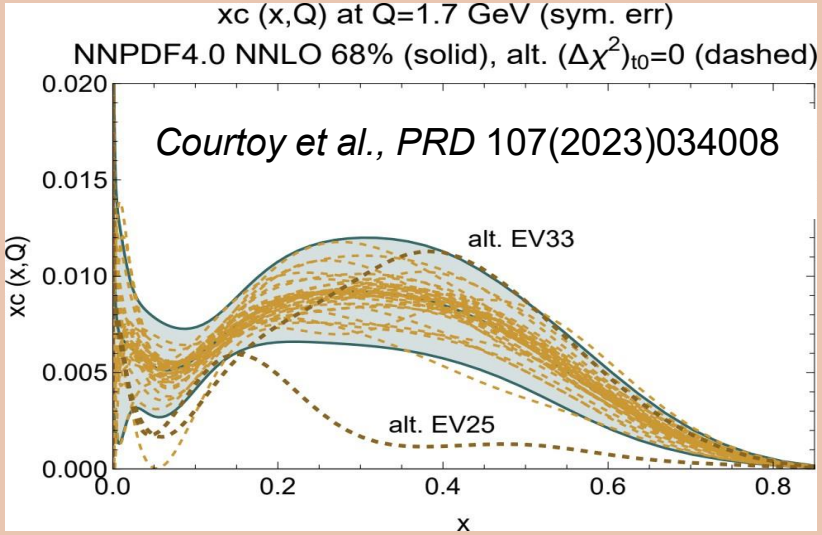
**T. J. Hobbs**, J. T. Londergan, W. Melnitchouk, *Phys. Rev. D* 89 (2014) 074008

# How significant is the non-zero FC?




**NNPDF states a $3\sigma$ evidence for $f_{IC}(x, Q_0) \neq 0$** based on the combined constraints from the baseline fit, LHCb $Z + c$ analysis, and EMC $F_2^c$ data




CTEQ-TEA authors find larger uncertainties in each of these sources.

This conclusion is also supported by

- **Lagrange multiplier scans** in the CT18 FC fit [upper figure]
- **hopscotch sampling** of MC replicas in the NNPDF4.0 fitting code [lower figure].

**Consequently, $f_{FC}(x, Q_0) \approx 0$ is allowed with high confidence.**