

PARTON DISTRIBUTION FUNCTIONS (PDFs)

Katerina Lipka

RECOMMENDED LITERATURE

- **Particle Physics Reference Library editor: H. Schopper Volume 1: Theory and Experiments**

Chapter 4 QCD Guido Altarelli and Stefano Forte: The Theory of Strong Interactions

https://link.springer.com/chapter/10.1007/978-3-030-38207-0_4

- **The Structure of the Nucleon, by Anthony W. Thomas, Wolfram Weise ISBN/EAN: 9783527635047**

PDFs: INPUT TO EVENT DESCRIPTION/MC SIMULATIONS

The structure of an event – 1

Warning: schematic only, everything simplified, nothing to scale, ...

From the
Introduction lecture

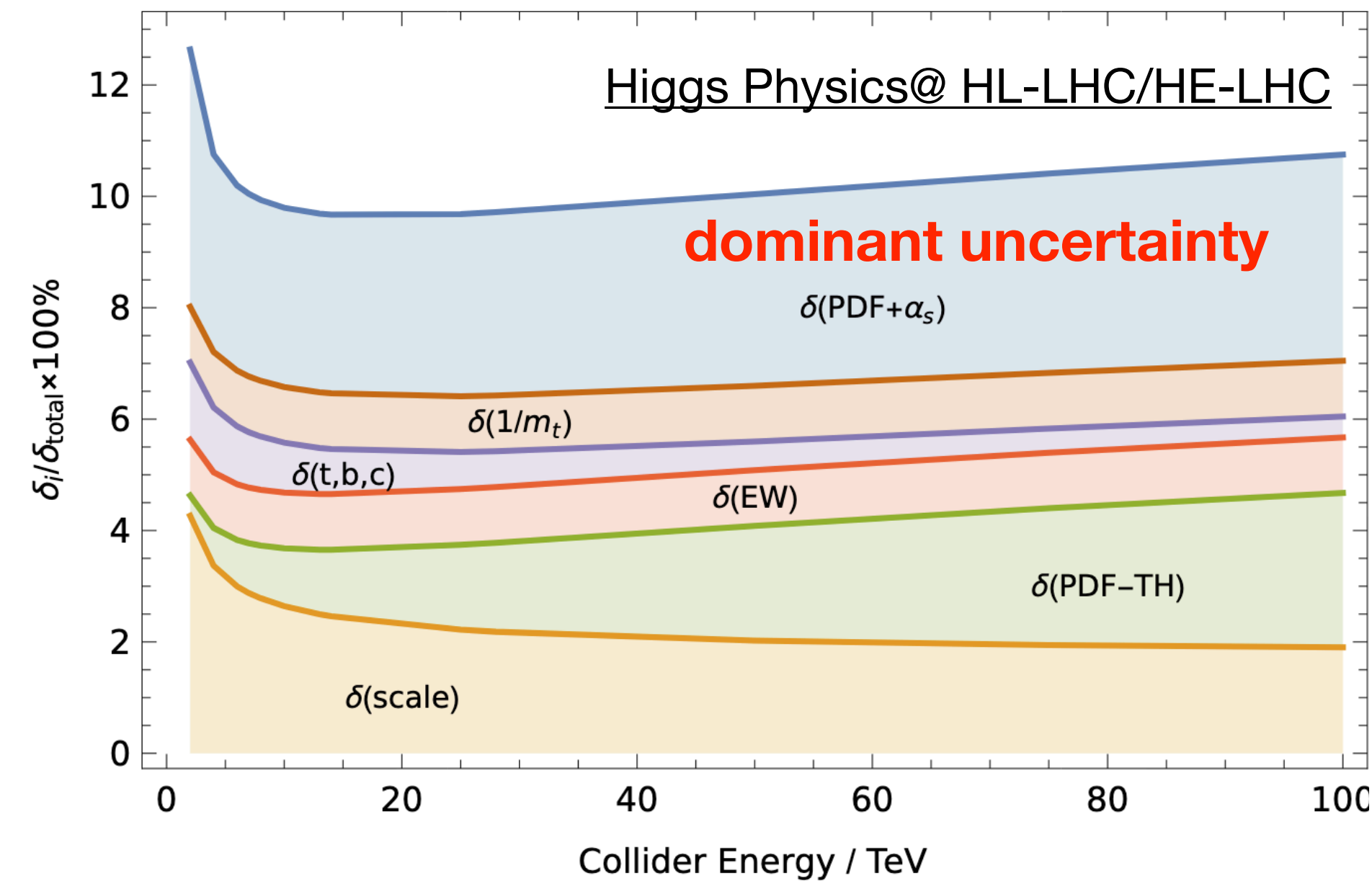


Incoming beams: parton densities

PDFs BECOME A PROBLEM FOR LHC

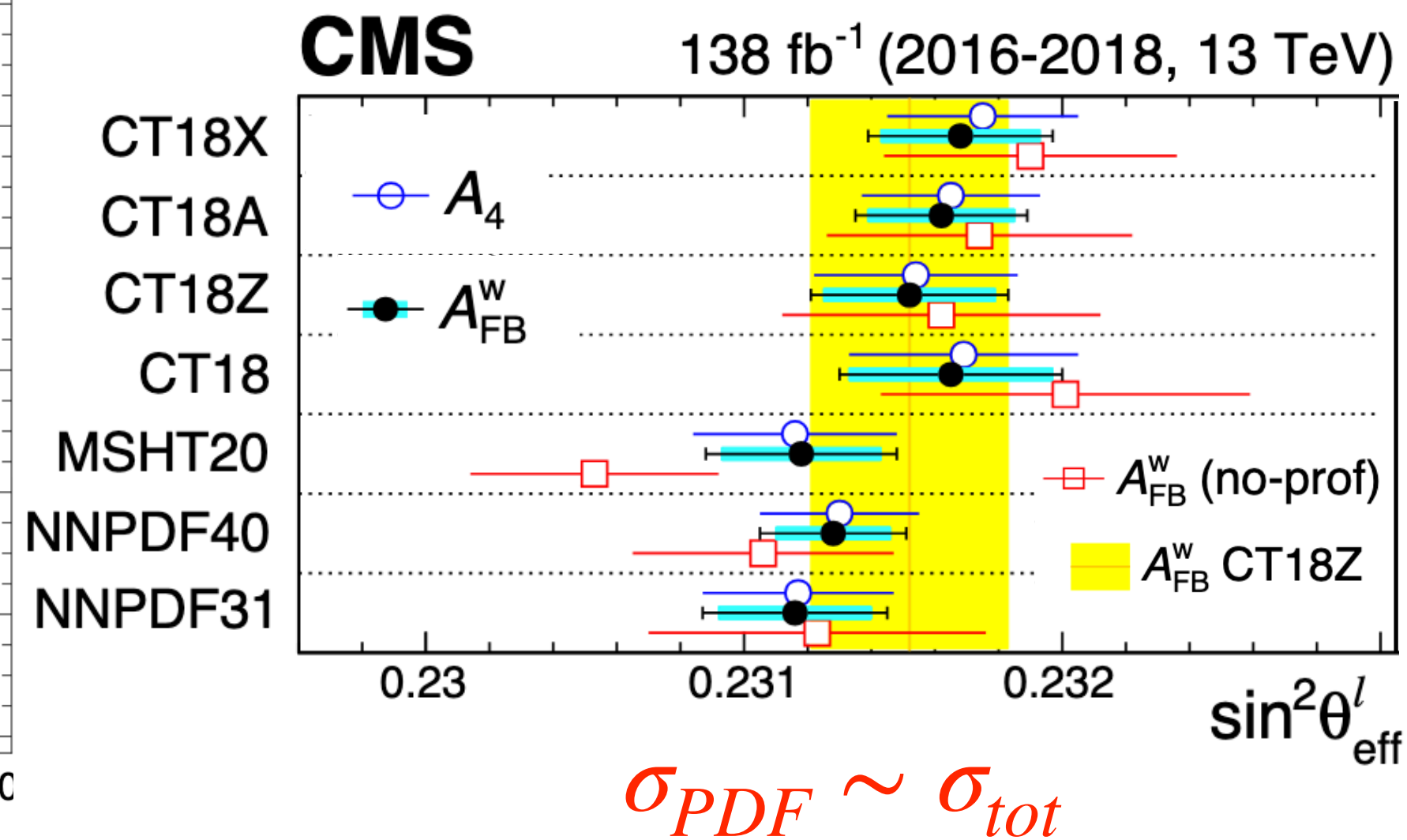
dilute precision of SM parameters and bias the BSM interpretations

error in inclusive Higgs cross section



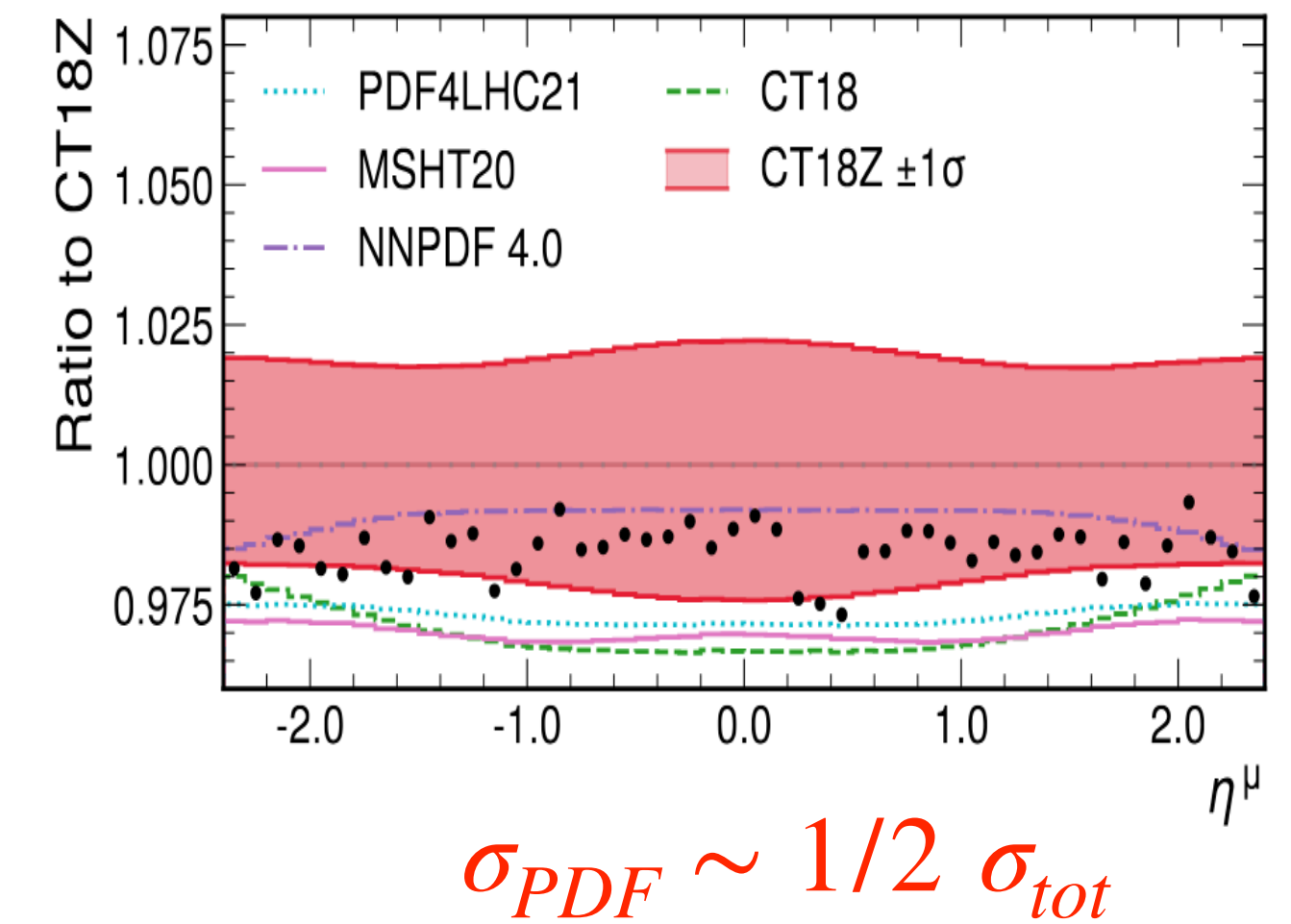
Electroweak mixing angle

PLB 866 (2025) 139526



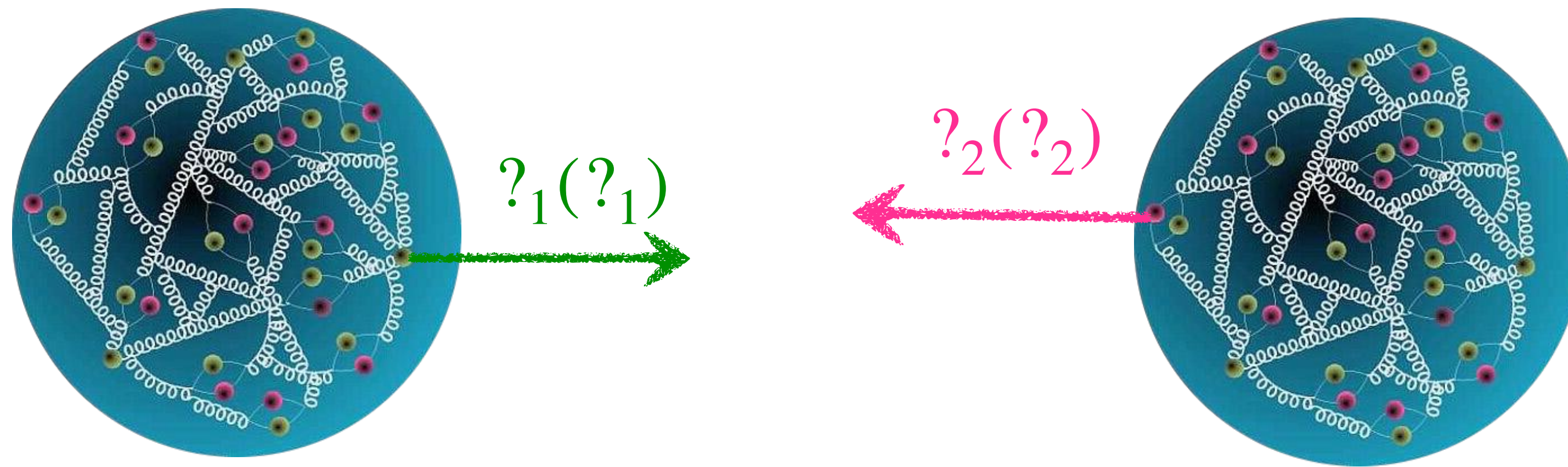
W mass

arXiv:2412.13872



PDFs: INPUT TO EVENT DESCRIPTION/MC SIMULATIONS

to make a reasonable prediction, we need to know:

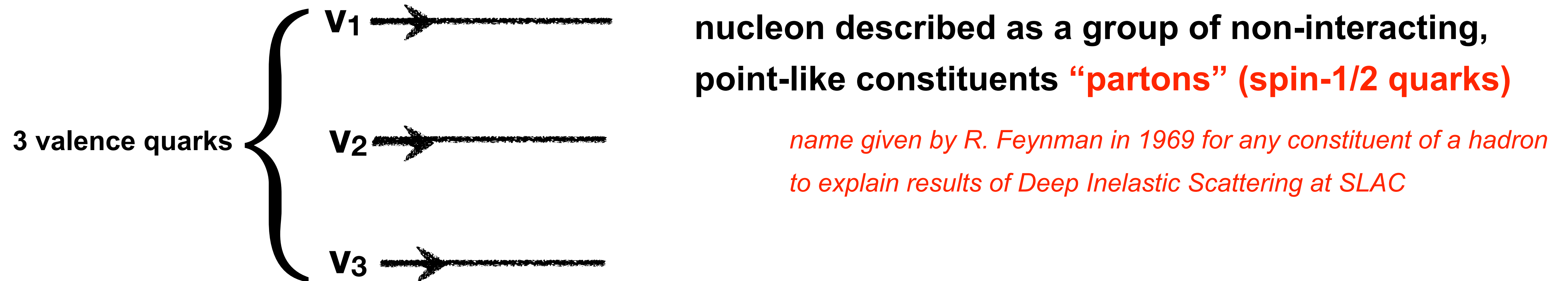


which constituent of one proton interacts with which constituent of another proton?

what are their momenta and (sometimes) mass ?

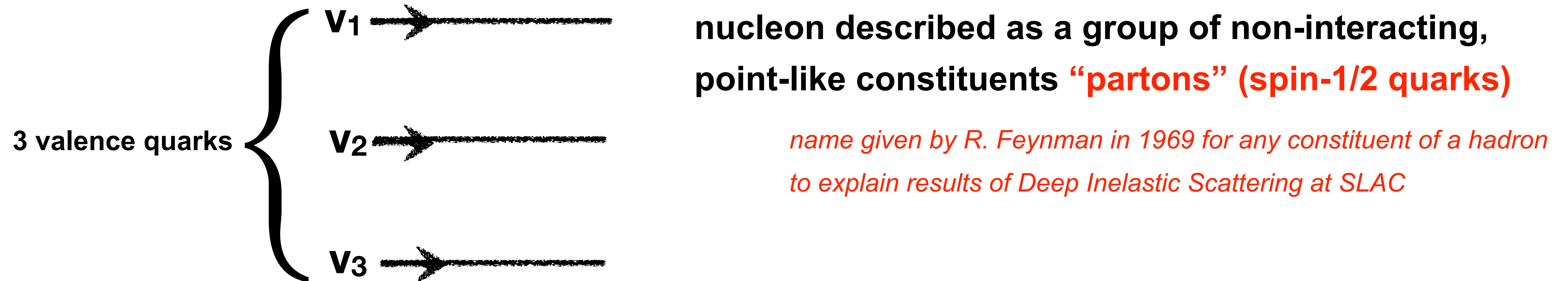
as of now, we know only the probability distributions

HISTORY: QUARK-PARTON MODEL (QPM)



Nucleon structure: 3 valence quarks which define the nucleon's quantum numbers

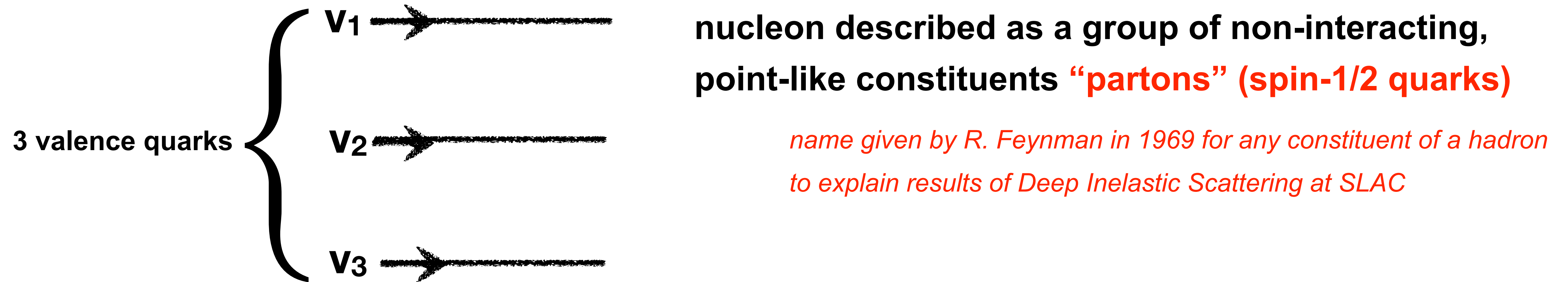
HISTORY: QUARK-PARTON MODEL (QPM)



Nucleon structure studied first in lepton-nucleon **deep inelastic scattering (DIS)** [$e^{\pm}N$ or νN]

QPM: cross section of DIS = sum of scattering cross sections off individual partons, weighted with a **probability to interact with a certain parton, carrying a longitudinal fraction x of proton momentum**

HISTORY: QUARK-PARTON MODEL (QPM)



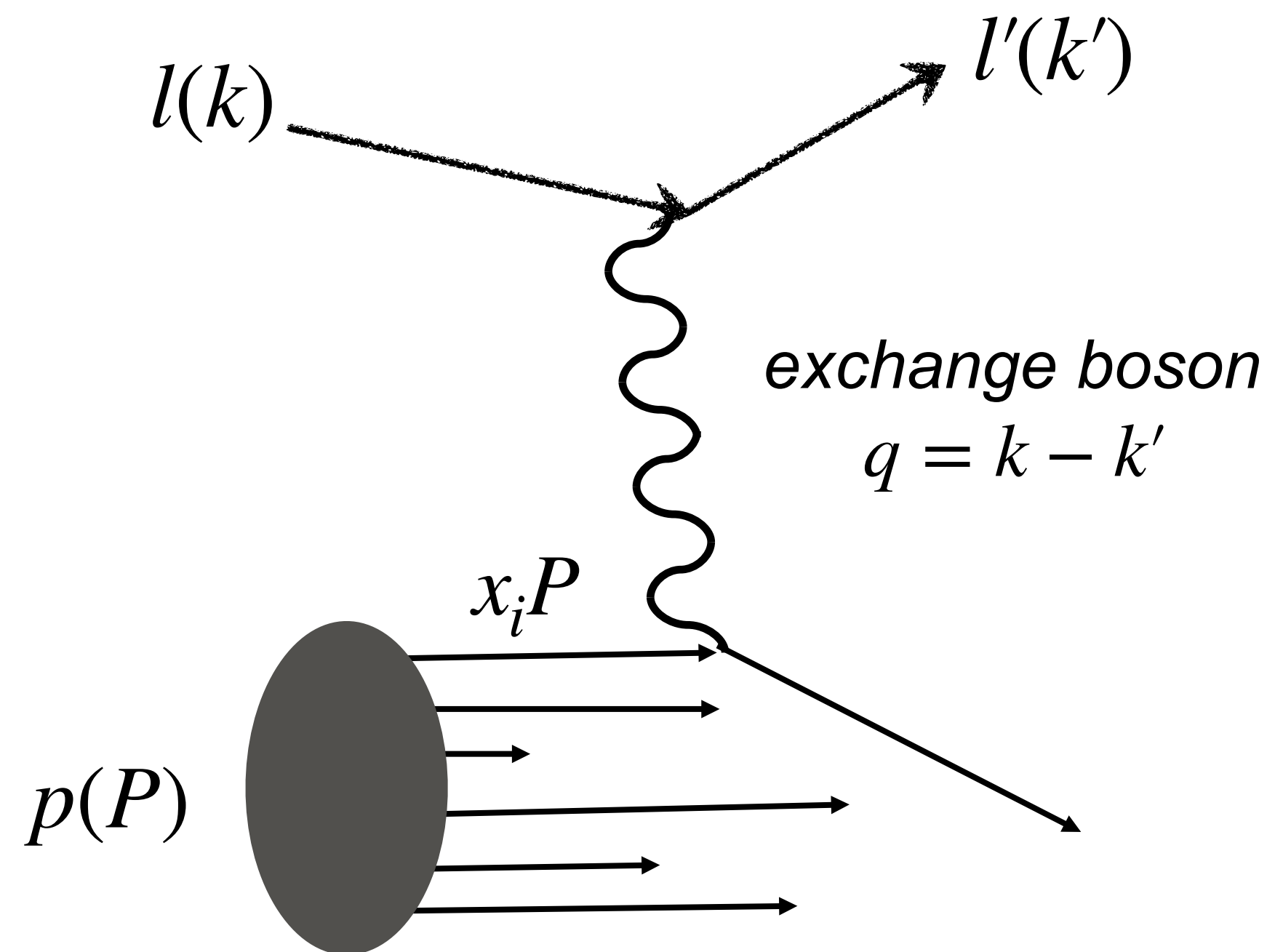
Nucleon structure studied first in lepton-nucleon **deep inelastic scattering (DIS)** [$e^{\pm}N$ or νN]

QPM: cross section of DIS = sum of scattering cross sections off individual partons, weighted with a
probability to interact with a certain parton, carrying a longitudinal fraction x of proton momentum

→ **Parton Distribution Functions in QPM: $f_i(x)$**

PHYSICS MEANING OF x in QPM

infinite-momentum frame (IMF): Partons are massless, non-interacting, have no transverse momenta

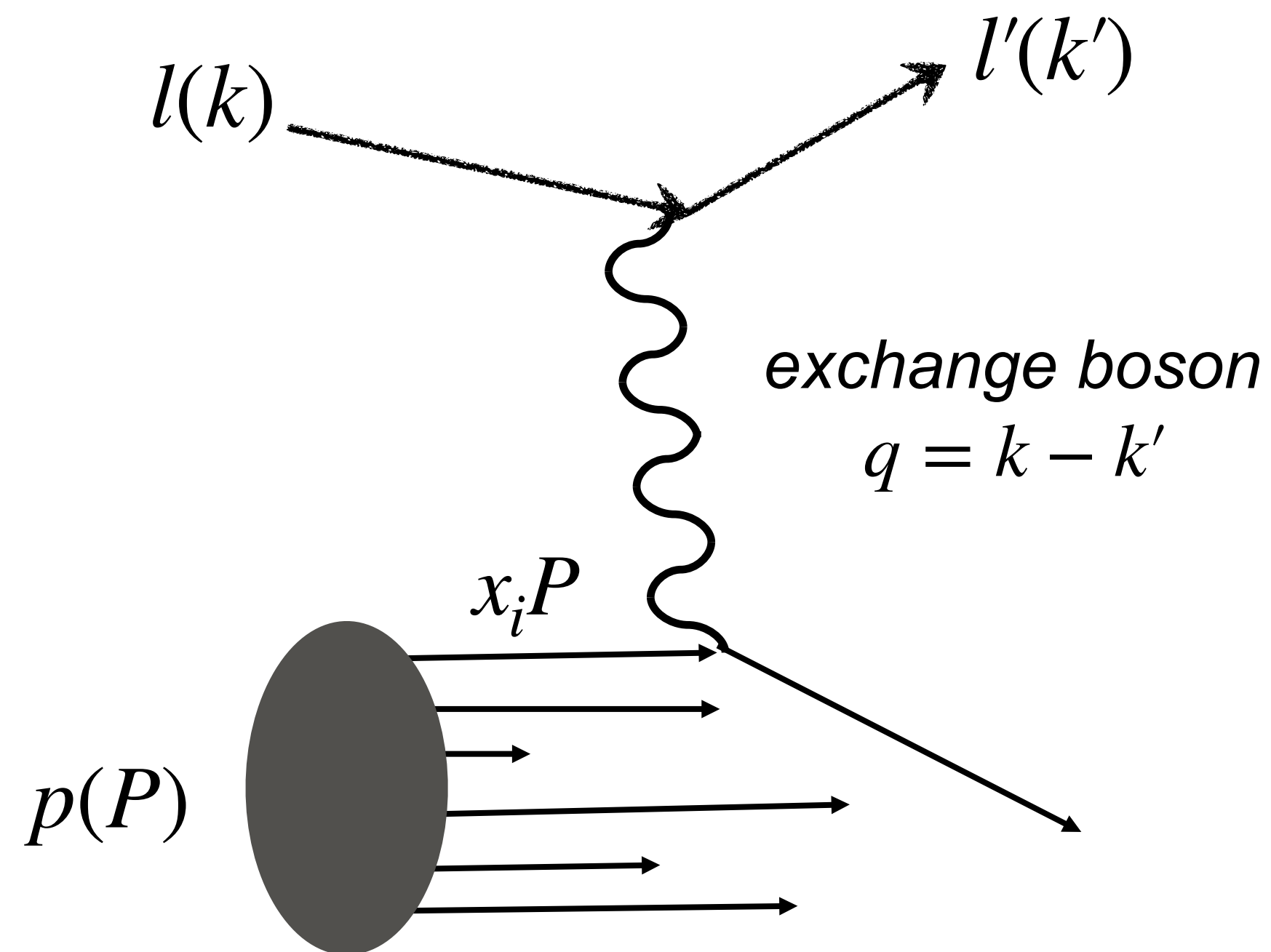


Parton i carries a fraction x_i of the proton momentum:

$$x = \frac{-q^2}{2p \cdot q}, \quad 0 < x_i < 1, \quad \sum_i x_i = 1$$

PHYSICS MEANING OF x in QPM

infinite-momentum frame (IMF): Partons are massless, non-interacting, have no transverse momenta



Parton i carries a fraction x_i of the proton momentum:

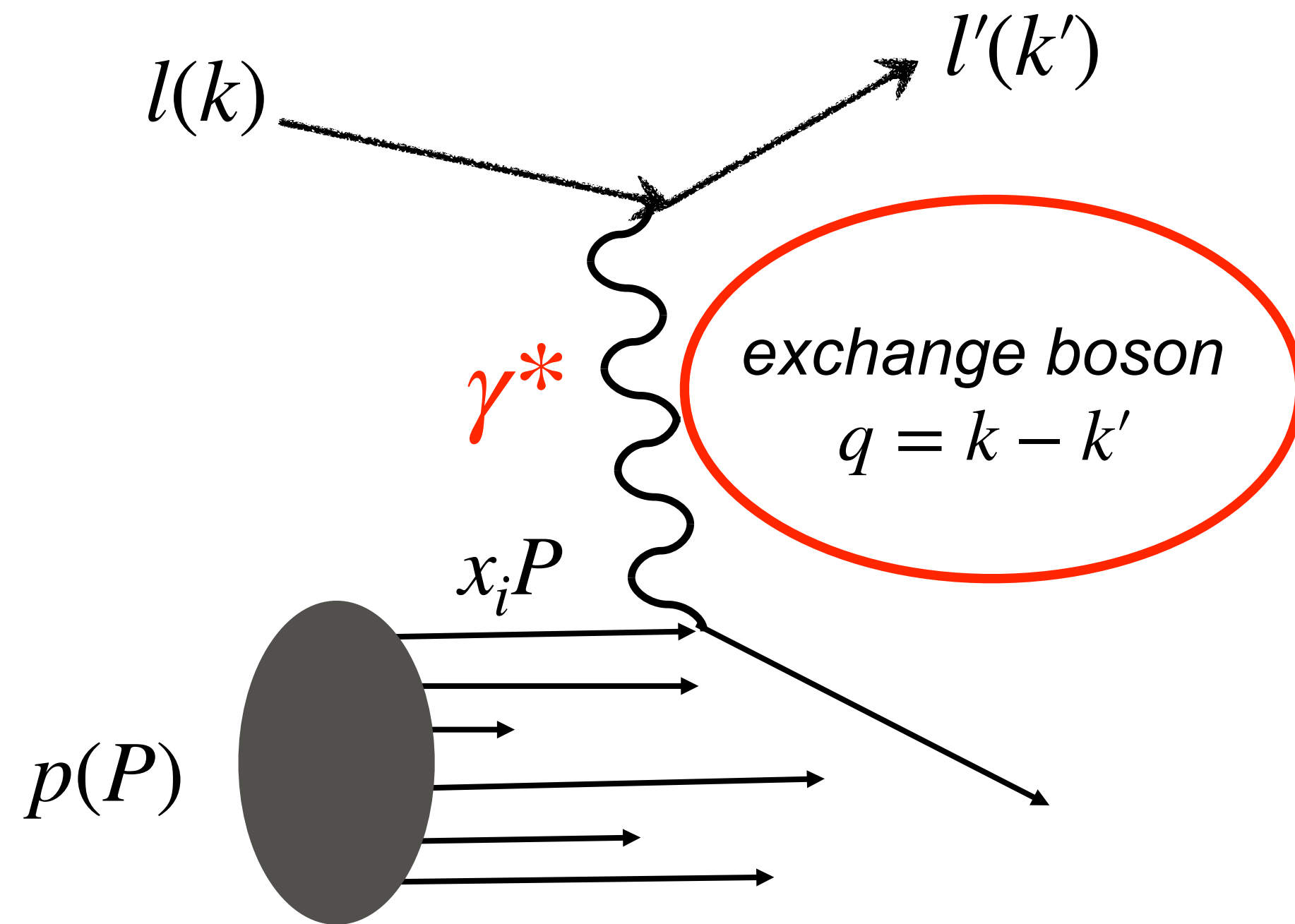
$$x = \frac{-q^2}{2p \cdot q}, \quad 0 < x_i < 1, \quad \sum_i x_i = 1$$

x - Bjorken scaling variable

James Bjorken: scattering off a point-like constituent (in IMF) is independent on resolution, 'scales' with x

RESOLUTION MEASURE: ONE-PHOTON EXCHANGE

infinite-momentum frame (IMF): Partons are massless, non-interacting, have no transverse momenta



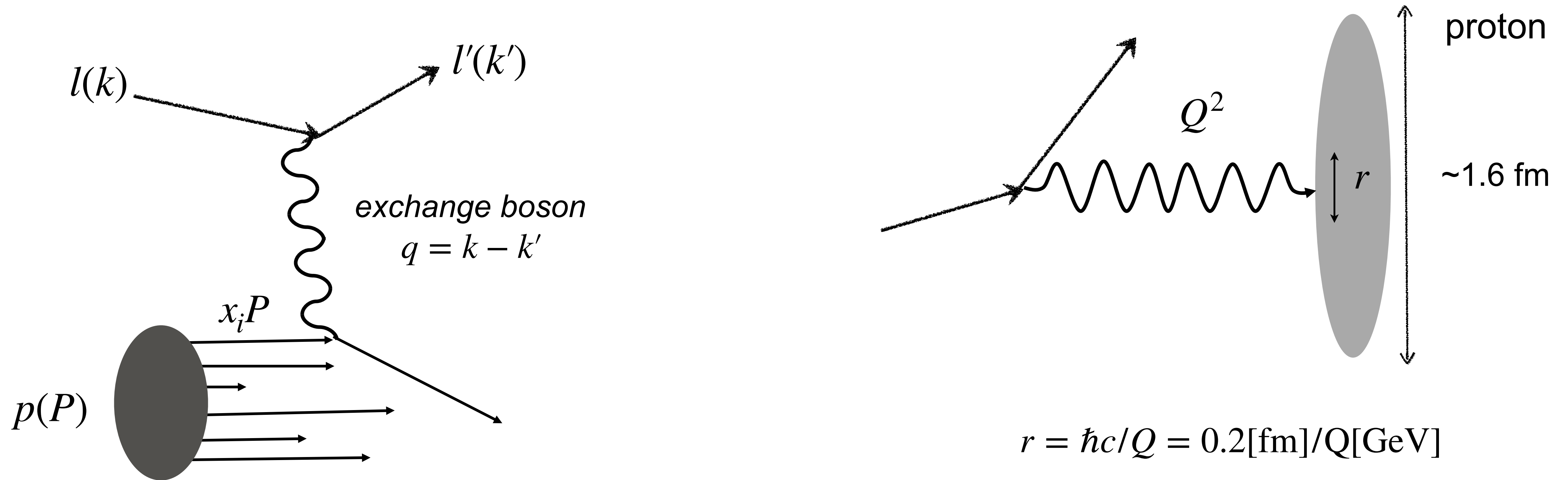
at energies $< m_Z$, purely electromagnetic
exchange of a photon with **virtuality** $Q^2 = -q^2$

Parton i carries a fraction x_i of the proton momentum:

$$x = \frac{-q^2}{2p \cdot q}, \quad 0 < x_i < 1, \quad \sum_i x_i = 1$$

PHYSICS MEANING OF x AND Q^2 (QPM)

infinite-momentum frame (IMF): Partons are massless, non-interacting, have no transverse momenta



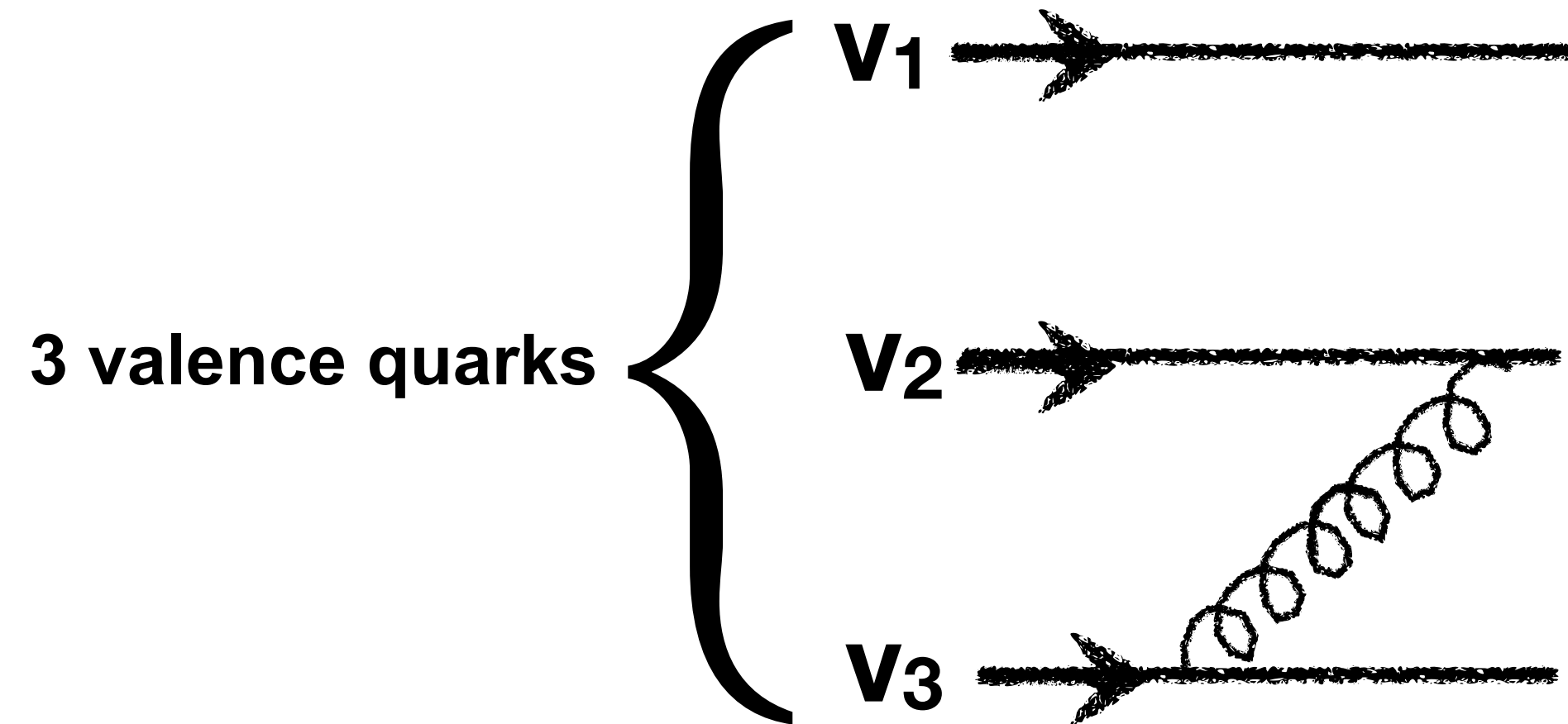
Parton i carries a fraction x_i of the proton momentum:

$$x = \frac{-q^2}{2p \cdot q}, \quad 0 < x_i < 1, \quad \sum_i x_i = 1$$

4-momentum transfer Q^2 defines the distance r which can be resolved

defines resolution of the scattering

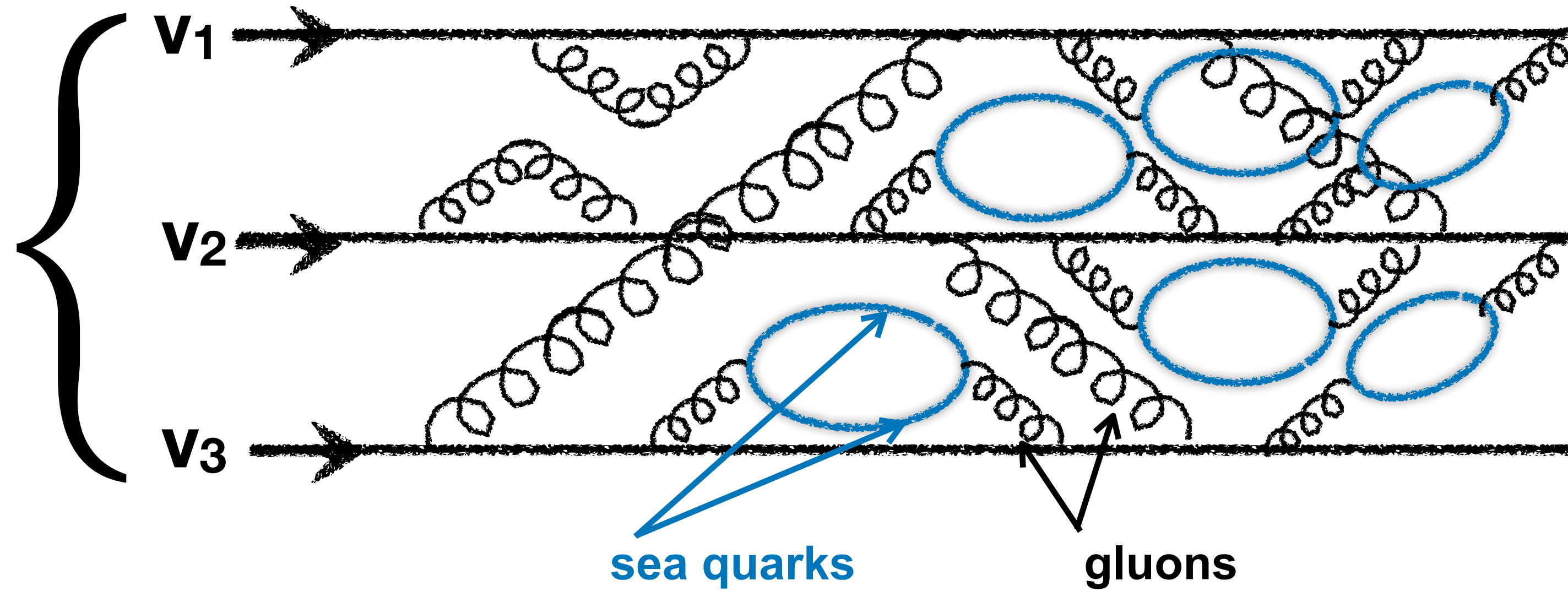
QCD-CORRECTED QPM



quarks do interact via gluon exchange

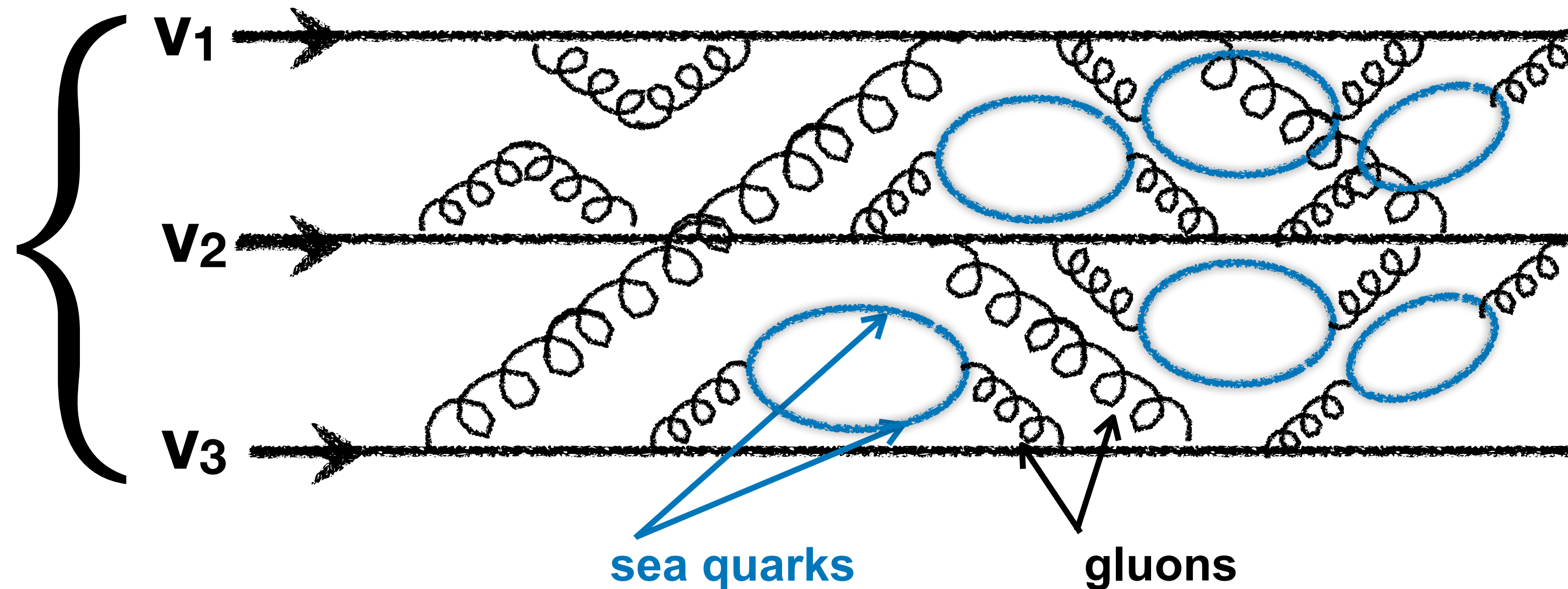
QCD radiation corrections have to be considered

QCD-CORRECTED QPM



nucleon = valence quarks + gluons + sea quarks

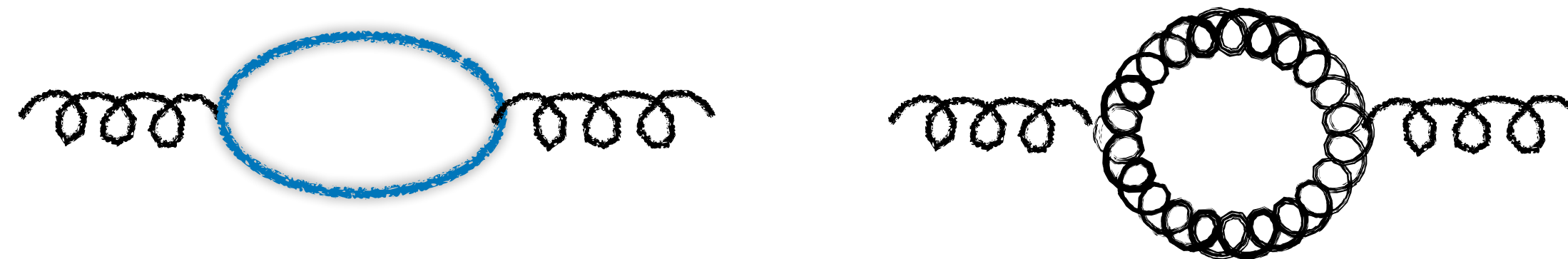
QCD-CORRECTED QPM



QCD
radiation
corrections

Cross-section calculation:

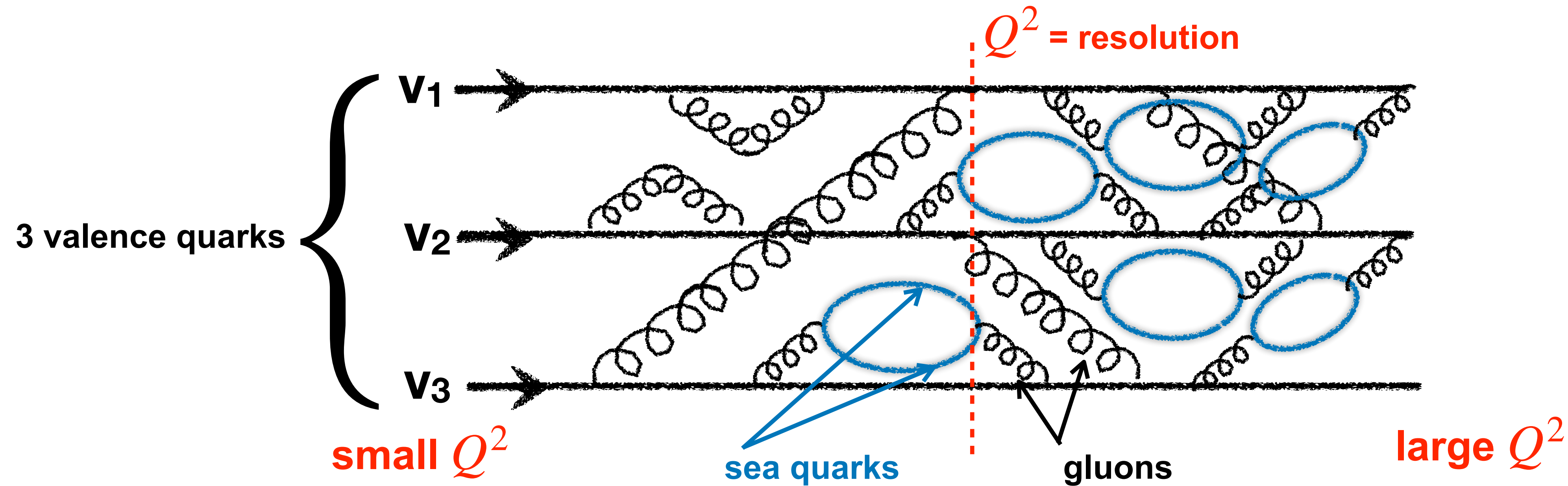
QCD vacuum
polarisation diagrams



lead to divergent terms

cross section becomes infinite $\rightarrow$ renormalisation needed $\rightarrow$ additional dependence on Q^2

QCD-CORRECTED QPM

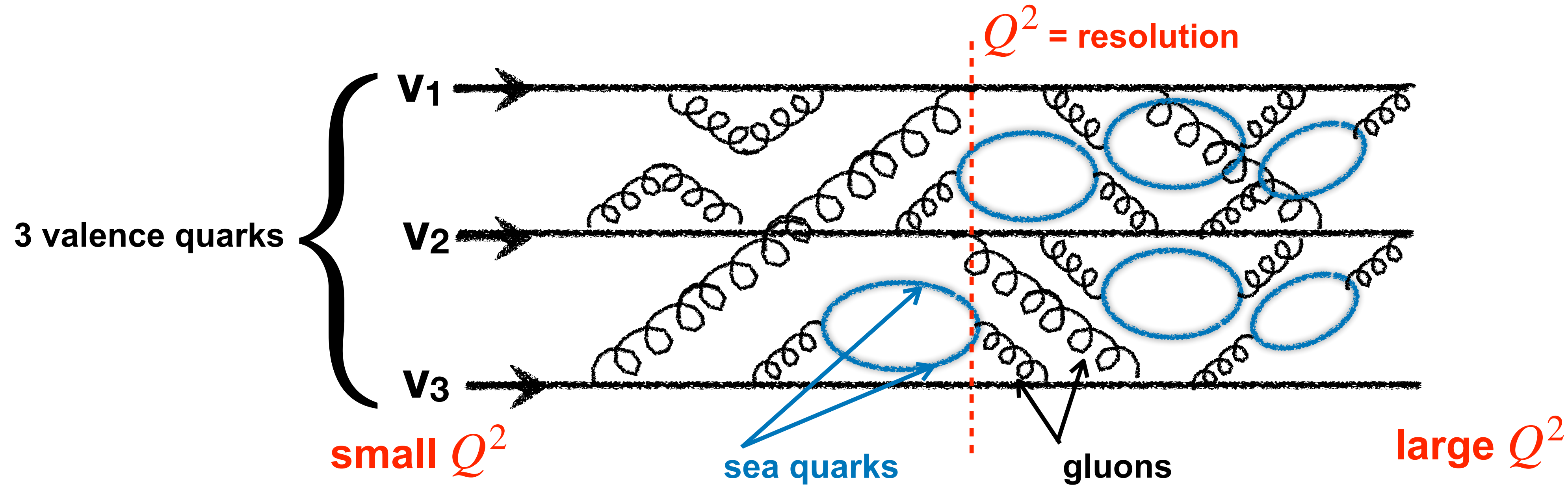


Partons = quarks and gluons

Scaling violation: number (+kind) of probed partons in a proton of momentum fraction x depends on Q^2

- small Q^2 : coarse resolution, only valence quarks
- large Q^2 : high resolution, also gluon and sea quarks

QCD-CORRECTED QPM



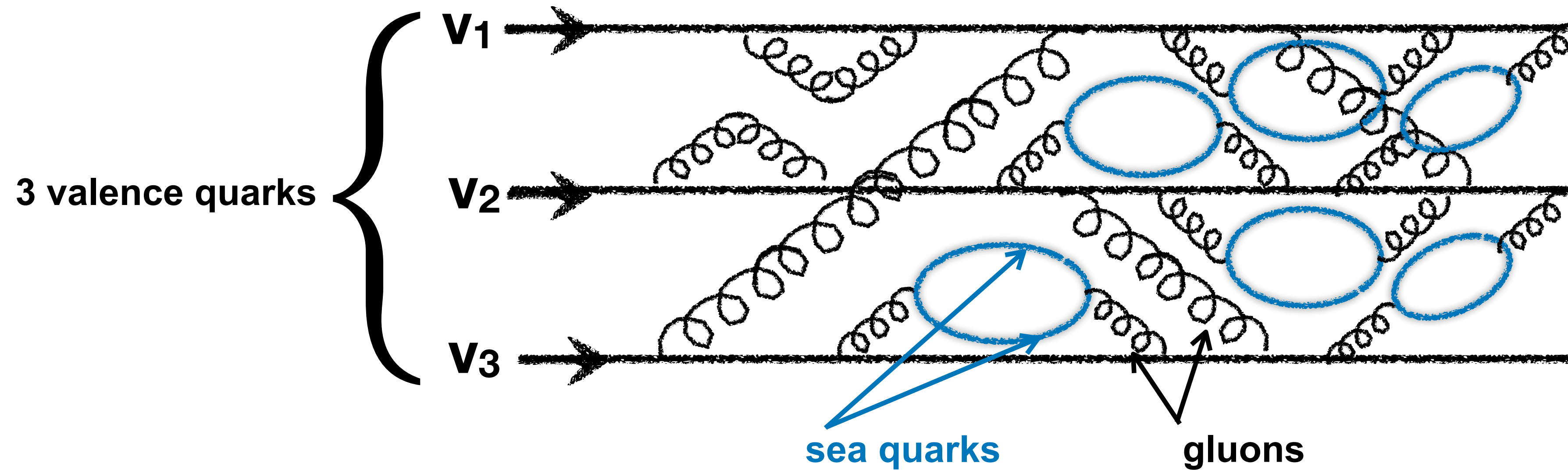
Partons = quarks and gluons

Scaling violation: number (+kind) of probed partons in a proton of momentum fraction x depends on Q^2

- small Q^2 : coarse resolution, only valence quarks
- large Q^2 : high resolution, also gluon and sea quarks

Note: here, a parton's transverse momentum is assumed negligible wrt Q
[collinear factorisation]

QCD-CORRECTED QPM



Renormalisation: *bare* $\frac{y}{x}$ *effective*

quark density: $q_0(y) \longrightarrow q(y, Q) = q_0(y) + \Delta q(y, Q)$

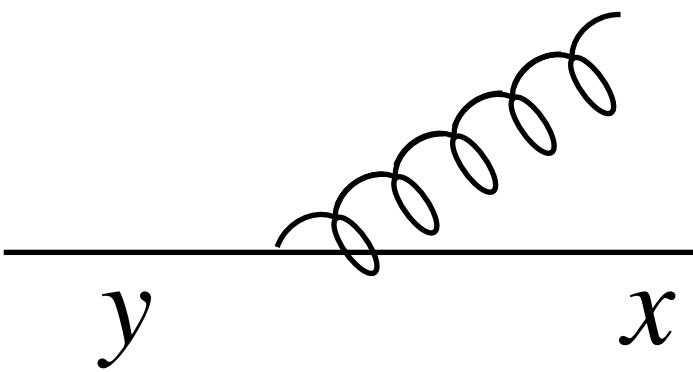
Q^2 evolution

splitting function

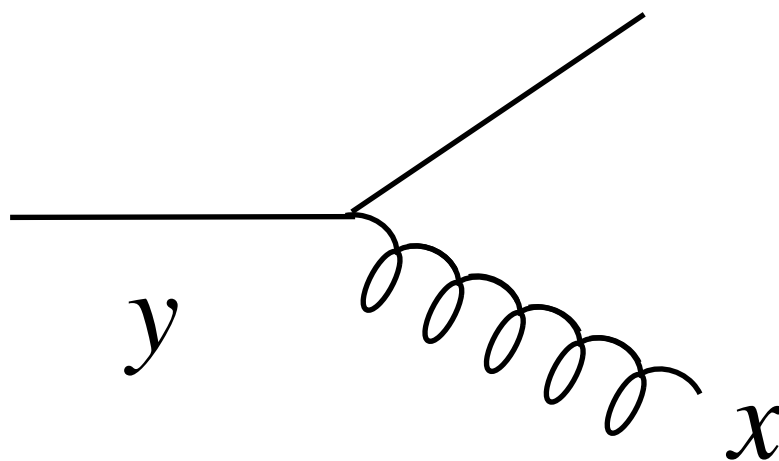
$$\frac{\alpha_S}{2\pi} \log Q^2 \int_x^1 dy \frac{q_0(y)}{y} \cdot P\left[\frac{x}{y}\right]$$

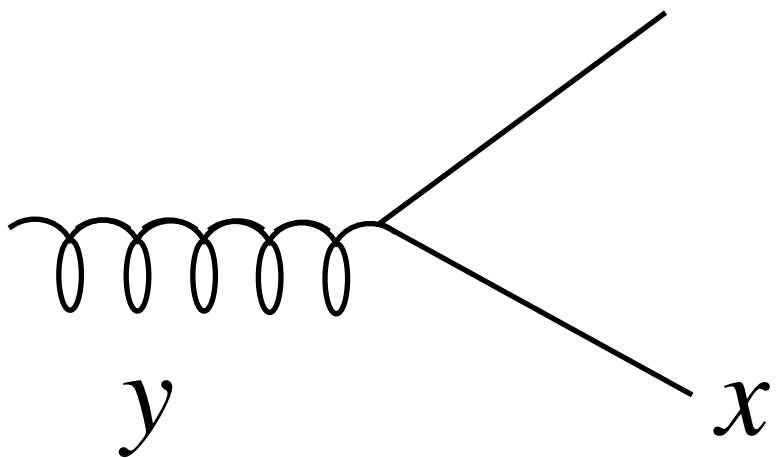
SPLITTING FUNCTIONS (LO QCD)

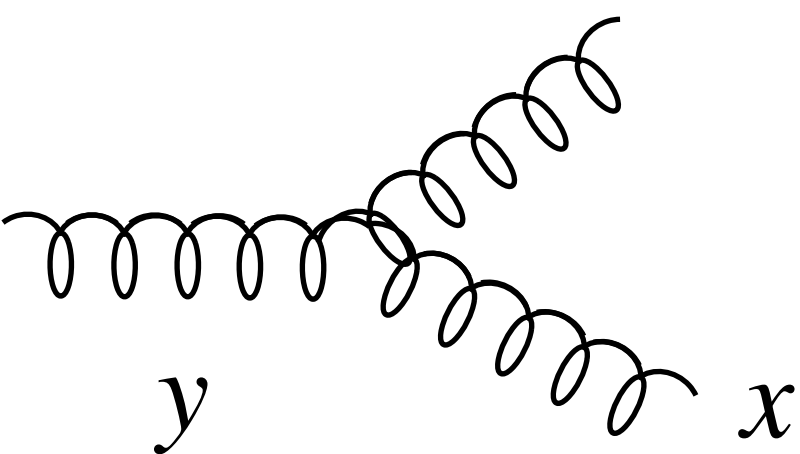
[Altarelli-Parisi notation]

$$P_{qq}\left(z = \frac{x}{y}\right)$$


P_{ji} probabilities of a parton i (with a proton momentum fraction y)
to radiate a parton j (with a proton momentum fraction x)

$$P_{Gq}\left(z = \frac{x}{y}\right)$$


$$P_{qG}\left(z = \frac{x}{y}\right)$$


$$P_{GG}\left(z = \frac{x}{y}\right)$$


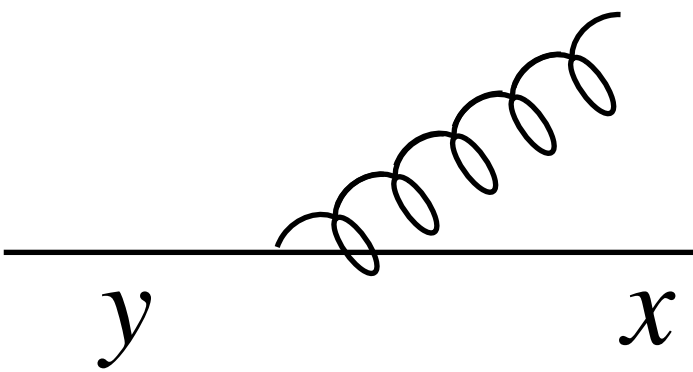
derived from the QCD vertices by using QCD constants:

$$C_F = \frac{N_c^2 - 1}{2N_c} = \frac{4}{3}, \quad C_A = N_c = 3, \quad T_R = \frac{1}{2}$$

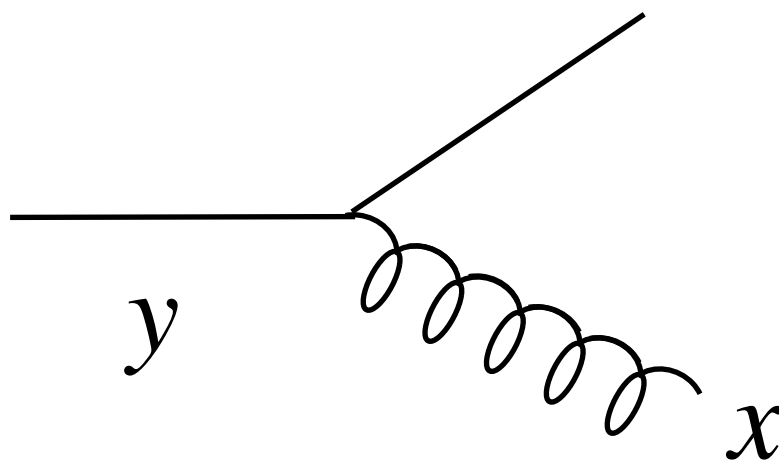
and number of flavours n_f

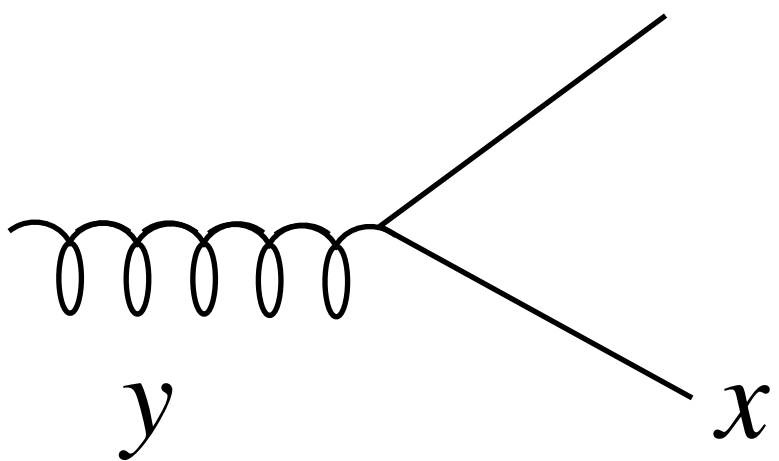
SPLITTING FUNCTIONS (LO QCD)

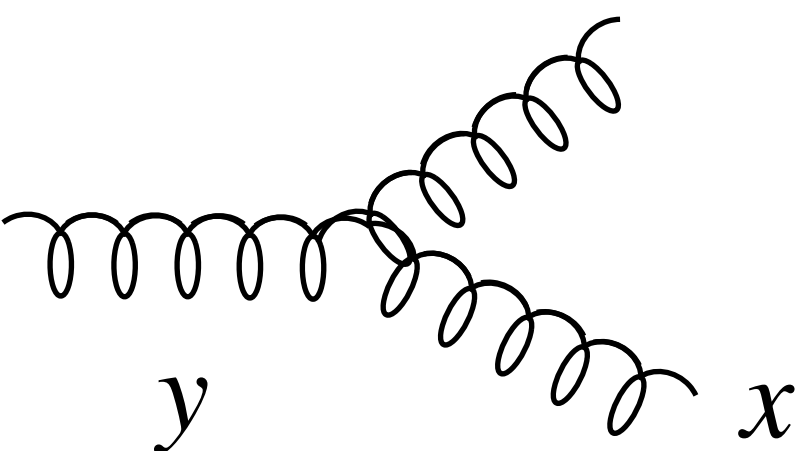
[Altarelli-Parisi notation]

$$P_{qq}\left(z = \frac{x}{y}\right)$$


P_{ji} probabilities of a parton i (with a proton momentum fraction y)
to radiate a parton j (with a proton momentum fraction x)

$$P_{Gq}\left(z = \frac{x}{y}\right)$$


$$P_{qG}\left(z = \frac{x}{y}\right)$$


$$P_{GG}\left(z = \frac{x}{y}\right)$$


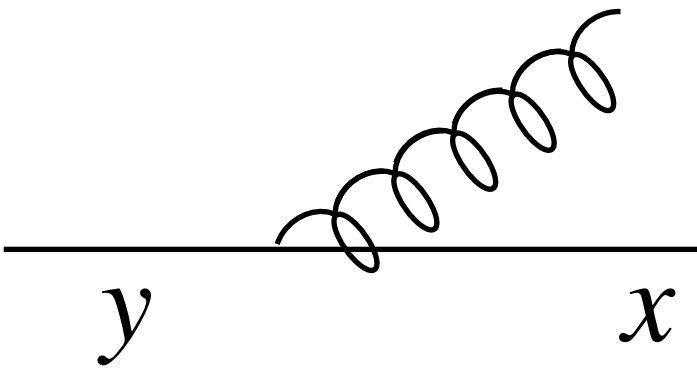
derived from the QCD vertices by using QCD constants:

$$C_F = \frac{N_c^2 - 1}{2N_c} = \frac{4}{3}, \quad C_A = N_c = 3, \quad T_R = \frac{1}{2}$$

and number of flavours n_f

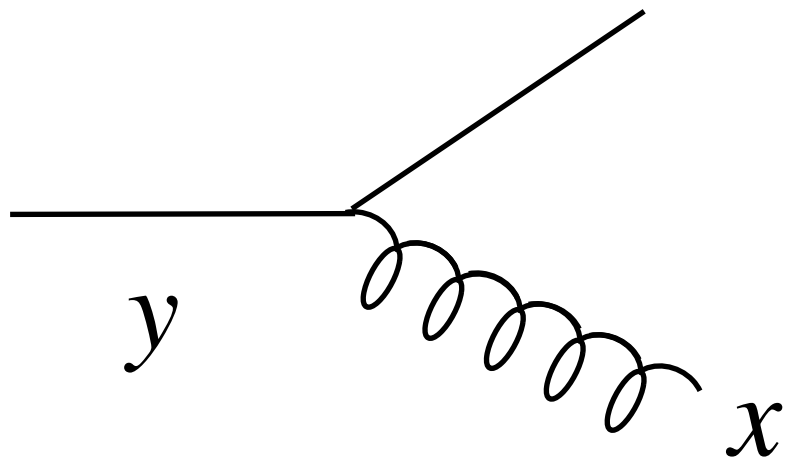
**independent on a process the parton density is taking part in
→ universal !**

SPLITTING FUNCTIONS (LO QCD)

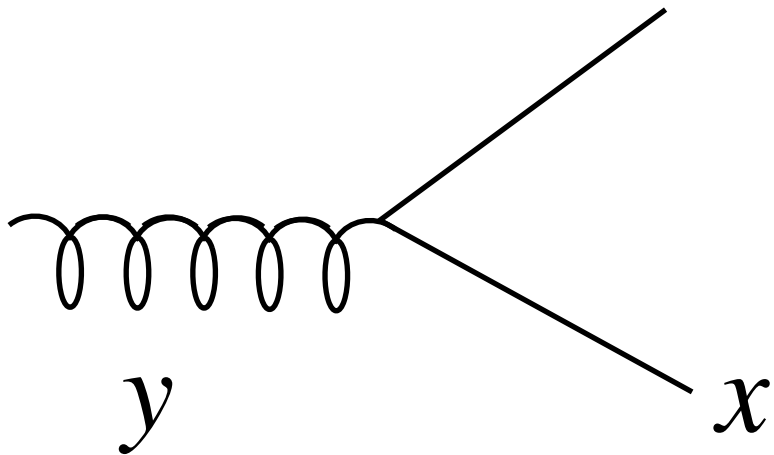
$$P_{qq}\left(z = \frac{x}{y}\right)$$


$$P_{qq}(z) = C_F \left[\frac{1+z^2}{1-z} \right]_+ + \frac{3}{2} C_F \delta(1-z)$$

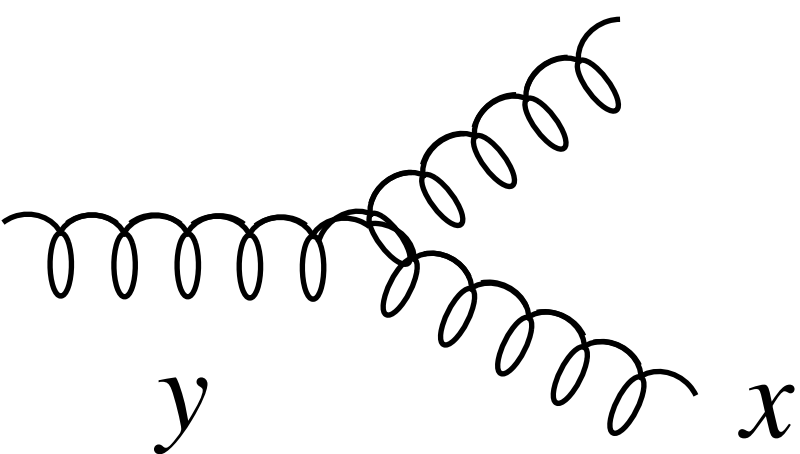
$$\int_0^1 dz [g(z)]_+ f(z) = \int_0^1 dz g(z) (f(z) - f(1))$$

$$P_{Gq}\left(z = \frac{x}{y}\right)$$


$$P_{Gq}(z) = T_R (z^2 + (1-z)^2)$$

$$P_{qG}\left(z = \frac{x}{y}\right)$$


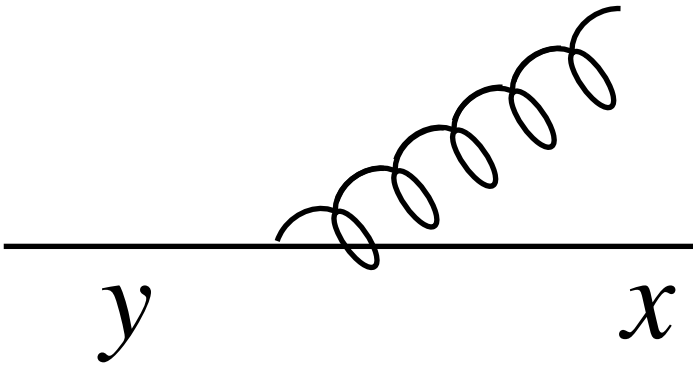
$$P_{qG}(z) = C_F \frac{1 + (1-z)^2}{z}$$

$$P_{GG}\left(z = \frac{x}{y}\right)$$


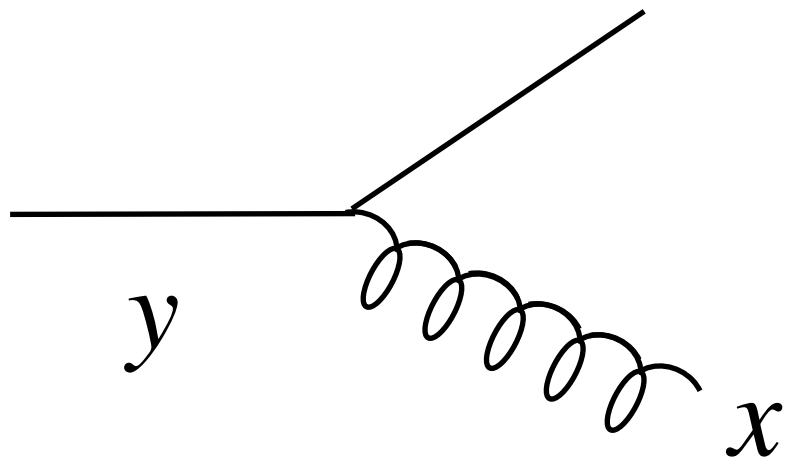
$$P_{GG}(z) = 2C_A \left[\frac{z}{(1-z)_+} + \frac{1-z}{z} + z(1-z) \right] + \frac{\beta_0}{2} \delta(1-z)$$

with $\beta_0 = \frac{11C_A - 4T_R n_f}{3}$

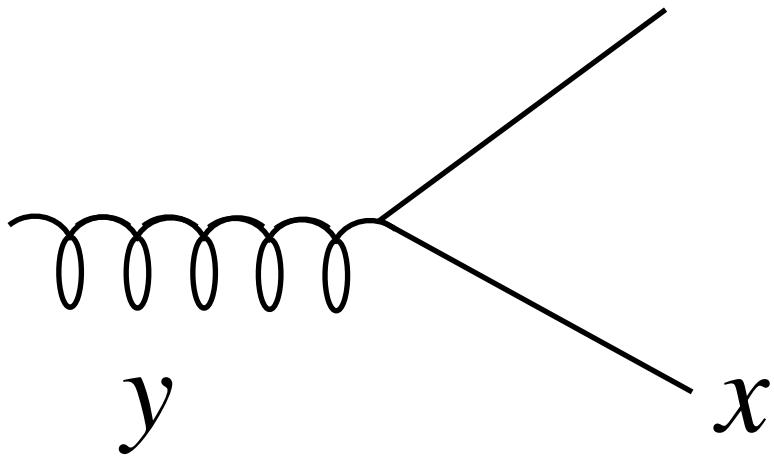
SPLITTING FUNCTIONS

$$P_{qq}\left(z = \frac{x}{y}\right)$$


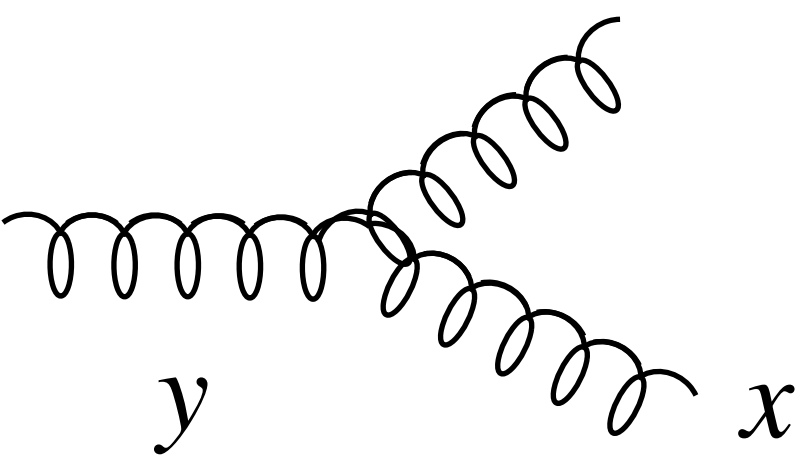
Beyond LO: power series in α_S

$$P_{Gq}\left(z = \frac{x}{y}\right)$$


$$P_{qq, qG, Gq, GG} = P_{qq, qG, Gq, GG}^0(z) + \frac{\alpha_S}{2\pi} P_{qq, qG, Gq, GG}^1(z) + \dots$$

$$P_{qG}\left(z = \frac{x}{y}\right)$$


Since recently, calculated up to N³LO [S.-O. Moch et al]

$$P_{GG}\left(z = \frac{x}{y}\right)$$


PARTON DISTRIBUTION FUNCTIONS IN QCD

Evolution equations formulated by Dokshitzer, Gribov, Lipatov, Altarelli and Parisi: **DGLAP**

Evolution starts at a scale Q_0^2 to any other $Q^2 > Q_0^2$

Which parton distributions?

Singlet $\Sigma(x, Q^2) = \sum_{i=1}^{n_f} [q_i(x, Q^2) + \bar{q}_i(x, Q^2)]$

Gluon $g(x, Q^2)$

Non-singlet $q^{NS}(x, Q^2) = q_i(x, Q^2) - \bar{q}_i(x, Q^2)$

(can not mix with the gluon PDF, e.g. valence - there are only 3 valence quarks)

DGLAP EVOLUTION EQUATIONS

general form:

$$\frac{\partial f_j(x, \mu^2)}{\partial \ln \mu^2} = \frac{\alpha_s(\mu^2)}{2\pi} \sum_i \left(P_{ji} \otimes f_i \right)(x, \mu^2)$$


$$\int_x^1 \frac{dy}{y} P_{ji}\left(\frac{x}{y}\right) f_i(y, \mu^2) = \int_x^1 \frac{dz}{z} P_{ji}(z) f_i\left(\frac{x}{z}, \mu^2\right)$$

DGLAP EVOLUTION EQUATIONS

general form:

$$\frac{\partial f_j(x, \mu^2)}{\partial \ln \mu^2} = \frac{\alpha_s(\mu^2)}{2\pi} \sum_i \left(P_{ji} \otimes f_i \right)(x, \mu^2)$$

singlet
gluon

$$\frac{\partial}{\partial \ln \mu^2} \begin{pmatrix} \Sigma(x, \mu^2) \\ g(x, \mu^2) \end{pmatrix} = \frac{\alpha_s(\mu^2)}{2\pi} \begin{pmatrix} P_{qq}(z) & 2n_f P_{qg}(z) \\ P_{gq}(z) & P_{gg}(z) \end{pmatrix} \otimes \begin{pmatrix} \Sigma(x, \mu^2) \\ g(x, \mu^2) \end{pmatrix}$$

Gluon and singlet equations are coupled

non-singlet

$$\frac{\partial q_{NS}(x, \mu^2)}{\partial \ln \mu^2} = \frac{\alpha_s(\mu^2)}{2\pi} \left(P_{NS} \otimes q_{NS} \right)(x, \mu^2)$$

at LO = P_{qq}

DGLAP EVOLUTION EQUATIONS

Interpretation as **probabilities** of parton branching: splitting functions are **positive** definite in x !

Physics consequence: the momentum and quantum numbers are **conserved** in evolution

DGLAP EVOLUTION EQUATIONS

Interpretation as **probabilities** of parton branching: splitting functions are **positive** definite in x !

Physics consequence: the momentum and quantum numbers are **conserved** in evolution

→ **Sum Rules**

- **baryon number/charge conservation:** number of valence quarks must remain constant

$$\frac{d}{d \ln \mu^2} \int_0^1 dx [q_i(x, \mu^2) - \bar{q}_i(x, \mu^2)] = 0 \quad \longrightarrow \quad \int_0^1 dz P_{qq}^0(z) = 0$$

[valence quark number does not evolve with scale] [real emission does not change the quark number]

DGLAP EVOLUTION EQUATIONS

Interpretation as **probabilities** of parton branching: splitting functions are **positive** definite in x !

Physics consequence: the momentum and quantum numbers are **conserved** in evolution

→ **Sum Rules**

- **baryon number/charge conservation:** number of valence quarks must remain constant

$$\frac{d}{d \ln \mu^2} \int_0^1 dx [q_i(x, \mu^2) - \bar{q}_i(x, \mu^2)] = 0 \quad \Rightarrow \quad \int_0^1 dz P_{qq}^{(NS)}(z) = 0$$
$$P_{qq}(z) = C_F \left[\frac{1+z^2}{1-z} \right]_{+} + \frac{3}{2} C_F \delta(1-z)$$

→ assure convergence as $z \rightarrow 1$

DGLAP EVOLUTION EQUATIONS

Interpretation as **probabilities** of parton branching: splitting functions are **positive** definite in x !

Physics consequence: the momentum and quantum numbers are **conserved** in evolution

→ Sum Rules

- **baryon number/charge conservation:** $\int_0^1 dz P_{qq}^{(NS)}(z) = 0$
- **momentum conservation:** $\sum_i \int_0^1 dx x f_i(x, \mu^2) = 1 \quad \rightarrow \quad \sum_i \int_0^1 dz z P_{ij}(z) = 0 \quad \text{for each } i$

DGLAP EVOLUTION EQUATIONS

Interpretation as **probabilities** of parton branching: splitting functions are **positive** definite in x !

Physics consequence: the momentum and quantum numbers are **conserved** in evolution

→ Sum Rules

- **baryon number/charge conservation:** $\int_0^1 dz P_{qq}^{(NS)}(z) = 0$
- **momentum conservation:** $\sum_i \int_0^1 dx x f_i(x, \mu^2) = 1 \quad \rightarrow \quad \sum_i \int_0^1 dz z P_{ij}(z) = 0 \quad \text{for each } i$

explicitly for
quarks and gluons

$$\int_0^1 dz z (P_{qq}^S(z) + P_{Gq}(z)) = 0$$

$$\int_0^1 dz z (2n_f P_{qG}(z) + P_{GG}(z)) = 0$$

momentum is only redistributed in the evolution (not gained and not lost)

PARTON DISTRIBUTION FUNCTIONS IN QCD

Definition (collinear factorisation):

non-perturbative $f_i(x, \mu^2)$ represents the number density of partons of type i carrying a longitudinal momentum fraction x of the hadron, when probed at a factorisation scale μ^2 . Its evolution with μ^2 is governed by the perturbative DGLAP equations.

PARTON DISTRIBUTION FUNCTIONS IN QCD

Definition (collinear factorisation):

non-perturbative $f_i(x, \mu^2)$ represents the number density of partons of type i carrying a longitudinal momentum fraction x of the hadron, when probed at a factorisation scale μ^2 . Its evolution with μ^2 is governed by the perturbative DGLAP equations.

Problem:

while the evolution in μ^2 is perturbative, $f_i(x, \mu_0^2)$ @ starting scale μ_0 is a non-perturbative object. The x -dependence cannot be calculated from first principles, only certain integral constraints are imposed by QCD sum rules

→ PDFs must be fitted to experimental data [making use of factorisation]

QCD FACTORISATION

$$\sigma_{kh \rightarrow X} = \hat{\sigma}_i^n [\text{pQCD}] \otimes \text{PDF}_h(\text{PDF}_k) [\text{non - perturbative}] \text{ at factorisation scale } \mu_f$$

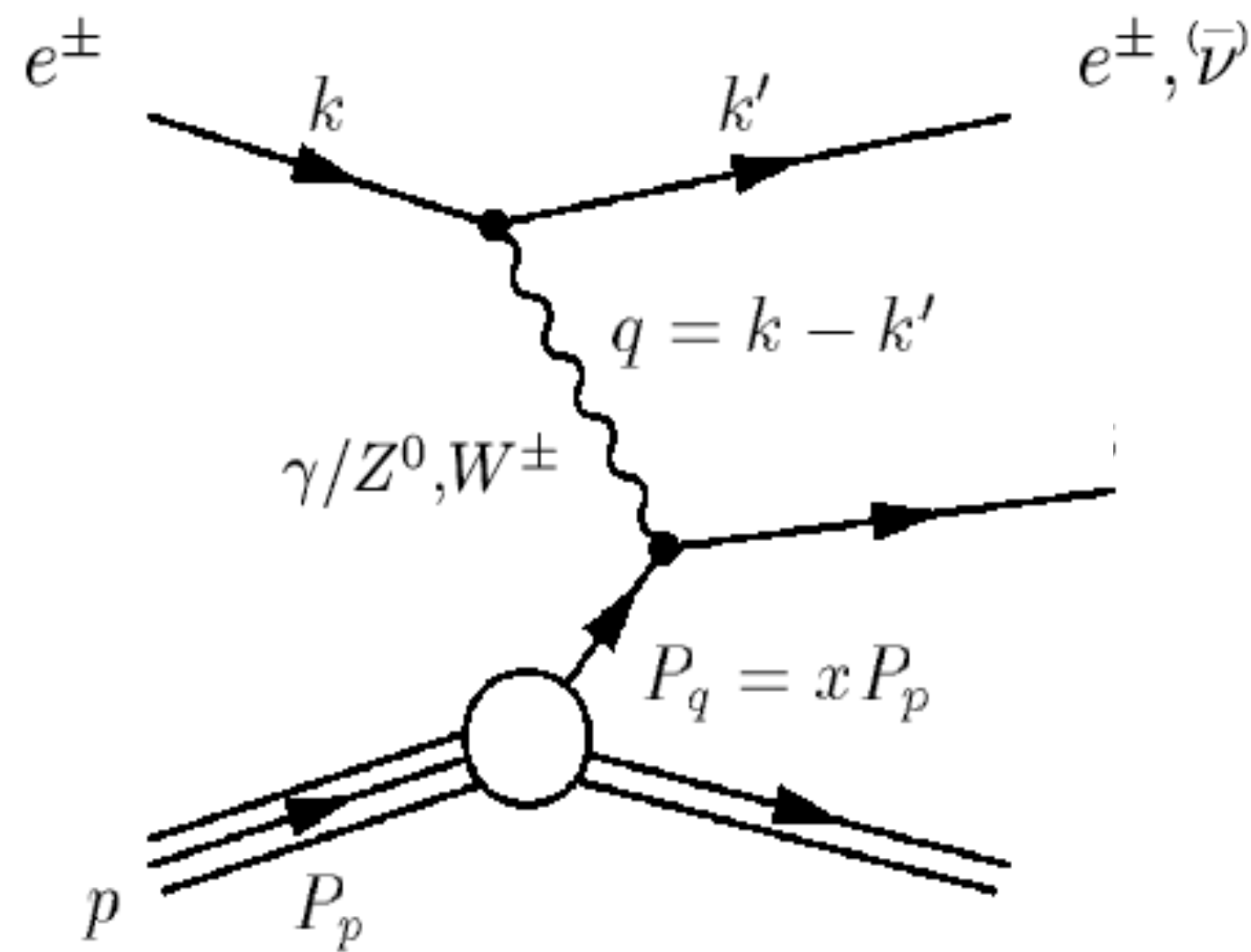
a scattering of a particle k off a hadron h (e.g. nucleon)

QCD FACTORISATION

$$\sigma_{kh \rightarrow X} = \hat{\sigma}_i^n [\text{pQCD}] \otimes \text{PDF}_h [\text{non-perturbative}] \text{ at factorisation scale } \mu_f$$

a scattering of a particle k off a hadron h (e.g. nucleon)

Simplest example: Deep Inelastic Scattering (DIS): $k = l \rightarrow lN \rightarrow l'X$



QCD FACTORISATION [DIS]

$$\sigma_{DIS} = \hat{\sigma}_i^n [\text{pQCD}] \otimes \text{PDFs [non – perturbative] at factorisation scale } \mu_f$$



**contain UV+collinear
singularities**

QCD FACTORISATION [DIS]

$$\sigma_{DIS} = \hat{\sigma}_i^n \text{ [pQCD]} \otimes \text{PDFs [non – perturbative] at factorisation scale } \mu_f$$



UV renormalisation !

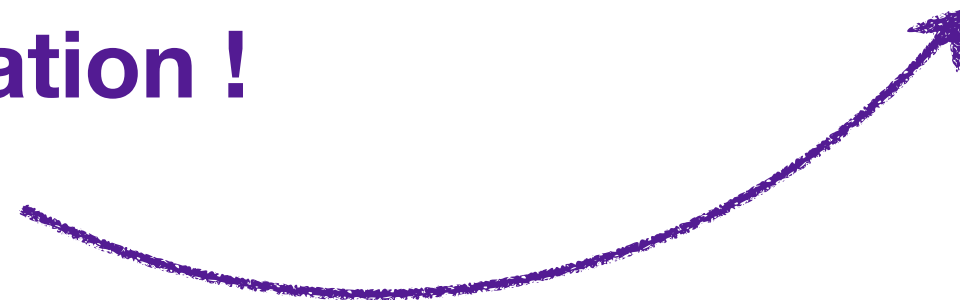
QCD FACTORISATION [DIS]

$$\sigma_{DIS} = \hat{\sigma}_i^n [\text{pQCD}] \otimes \text{PDFs [non - perturbative] at factorisation scale } \mu_f$$



UV renormalisation !

**soft and collinear divergences
subtracted and absorbed into PDFs**



QCD FACTORISATION [DIS]

$$\sigma_{DIS} = C_i^n [\text{pQCD}] \otimes \text{PDFs [non - perturbative]} \text{ at factorisation scale } \mu_f$$

**short-distance
coefficient functions
(finite !)**

**soft and collinear divergences
subtracted and absorbed into PDFs**

QCD FACTORISATION [DIS]

$$\sigma_{DIS} = C_i^n [\text{pQCD}] \otimes \text{PDFs [non – perturbative]} \text{ at factorisation scale } \mu_f$$

short-distance
coefficient functions
(finite !)

soft and collinear divergences
subtracted and absorbed into PDFs

DIS cross sections
or
structure functions

F_a ($a = 2, 3, L, \dots$) for a scattering off a hadron h , via exchange of a boson V

$$F_a^{V,h}(x, Q^2) = \sum_{i=q,\bar{q},G} \int_x^1 dz C_a^{V,i} \left(\frac{x}{z}, \frac{Q^2}{\mu_r^2}, \frac{\mu_f^2}{\mu_r^2}, \alpha_s(\mu_r^2) \right) f_{i,h}(z, \mu_f, \mu_r)$$

sum over all partons
(quarks, antiquarks, gluons)

QCD FACTORISATION [DIS]

$$\sigma_{DIS} = C_i^n [\text{pQCD}] \otimes \text{PDFs [non – perturbative]} \text{ at factorisation scale } \mu_f$$

short-distance
coefficient functions
(finite !)

soft and collinear divergences
subtracted and absorbed into PDFs

DIS cross sections
or
structure functions

F_a ($a = 2, 3, L, \dots$) for a scattering off a hadron h , via exchange of a boson V

$$F_a^{V,h}(x, Q^2) = \sum_{i=q,\bar{q},G} \int_x^1 dz C_a^{V,i} \left(\frac{x}{z}, \frac{Q^2}{\mu_r^2}, \frac{\mu_f^2}{\mu_r^2}, \alpha_s(\mu_r^2) \right) f_{i,h}(z, \mu_f, \mu_r)$$

(Wilson) **coefficients calculated in pQCD**
for a boson V , parton i and structure function a

Parton distribution functions,
specific for parton i and hadron h ,
universal for V and a

QCD FACTORISATION [DIS]

$$\sigma_{DIS} = C_i^n [\text{pQCD}] \otimes \text{PDFs [non – perturbative]} \text{ at factorisation scale } \mu_f$$

short-distance
coefficient functions
(finite !)

soft and collinear divergences
subtracted and absorbed into PDFs

DIS cross sections
or
structure functions

F_a ($a = 2, 3, L, \dots$) for a scattering off a hadron h , via exchange of a boson V

$$F_a^{V,h}(x, Q^2) = \sum_{i=q,\bar{q},G} \int_x^1 dz C_a^{V,i} \left(\frac{x}{z}, \frac{Q^2}{\mu_r^2}, \frac{\mu_f^2}{\mu_r^2}, \alpha_S(\mu_r^2) \right) f_{i,h}(z, \mu_f, \mu_r)$$

Attention:

beyond LO, **only valid in a certain**
factorisation/renormalisation **scheme!**

Terms are “re-shuffled” between C_a^i and f_i
Scheme defines this shuffling!

QCD FACTORISATION [DIS]

$$\sigma_{DIS} = C_i^n [\text{pQCD}] \otimes \text{PDFs [non – perturbative]} \text{ at factorisation scale } \mu_f$$

short-distance
coefficient functions
(finite !)

soft and collinear divergences
subtracted and absorbed into PDFs

DIS cross sections
or
structure functions

F_a ($a = 2, 3, L, \dots$) for a scattering off a hadron h , via exchange of a boson V

$$F_a^{V,h}(x, Q^2) = \sum_{i=q,\bar{q},G} \int_x^1 dz C_a^{V,i} \left(\frac{x}{z}, \frac{Q^2}{\mu_r^2}, \frac{\mu_f^2}{\mu_r^2}, \alpha_S(\mu_r^2) \right) f_{i,h}(z, \mu_f, \mu_r)$$

Attention:

beyond LO, **only valid in a certain**
factorisation/renormalisation **scheme!**

PDFs in different schemes and at different
orders are **different quantities!** Must be
only used consistently with coefficients!

QCD FACTORISATION [DIS]

$$\sigma_{DIS} = C_i^n [\text{pQCD}] \otimes \text{PDFs [non - perturbative]} \text{ at factorisation scale } \mu_f$$

short-distance
coefficient functions
(finite !)

soft and collinear divergences
subtracted and absorbed into PDFs

DIS cross sections
or
structure functions

F_a ($a = 2, 3, L, \dots$) for a scattering off a hadron h , via exchange of a boson V

$$F_a^{V,h}(x, Q^2) = \sum_{i=q,\bar{q},G} \int_x^1 dz C_a^{V,i} \left(\frac{x}{z}, \frac{Q^2}{\mu_r^2}, \frac{\mu_f^2}{\mu_r^2}, \alpha_S(\mu_r^2) \right) f_{i,h}(z, \mu_f, \mu_r)$$

Convolution range: physics at a given x only influenced by larger momentum fractions.

→ definite predictions at large x without any knowledge of the small- x

PDF EXTRACTION: QCD ANALYSIS

Observable
(measurable in experiment)

Calculable in pQCD
to certain order

Parameters
of interest

$$\sigma_{kh \rightarrow X}(x, Q^2) = \sum_{i=q, \bar{q}, G} \int_x^1 dz \, C_i \left(\frac{x}{z}, \frac{Q^2}{\mu_r^2}, \frac{\mu_f^2}{\mu_r^2}, \alpha_S(\mu_r^2) \right) f_{i,h}(z, \mu_f, \mu_r)$$

Data-driven *QCD analyses* are inverse problems that aim to infer PDFs

PDF EXTRACTION CHOICES

Observable
(measurable in experiment)

Calculable in pQCD
to certain order

Parameters
of interest

$$\sigma_{kh \rightarrow X}(x, Q^2) = \sum_{i=q, \bar{q}, G} \int_x^1 dz \, C_i \left(\frac{x}{z}, \frac{Q^2}{\mu_r^2}, \frac{\mu_f^2}{\mu_r^2}, \alpha_S(\mu_r^2) \right) f_{i,h}(z, \mu_f, \mu_r)$$

- **which data to chose?**
- **data from different processes:**
within one experiment, measurements might be statistically or systematically correlated
- **data from different experiments:**
data might be systematically correlated (through e.g. similar assumptions in the MC simulations)
possible inconsistencies of the data sets have to be accounted for

PDF EXTRACTION CHOICES

Observable
(measurable in experiment)

Calculable in pQCD
to certain order

Parameters
of interest

$$\sigma_{kh \rightarrow X}(x, Q^2) = \sum_{i=q, \bar{q}, G} \int_x^1 dz \, C_i \left(\frac{x}{z}, \frac{Q^2}{\mu_r^2}, \frac{\mu_f^2}{\mu_r^2}, \alpha_S(\mu_r^2) \right) f_{i,h}(z, \mu_f, \mu_r)$$

Number of *active* flavours in the nucleon. How to treat “heavy” quarks c and b ?

PDF EXTRACTION CHOICES

Observable
(measurable in experiment)

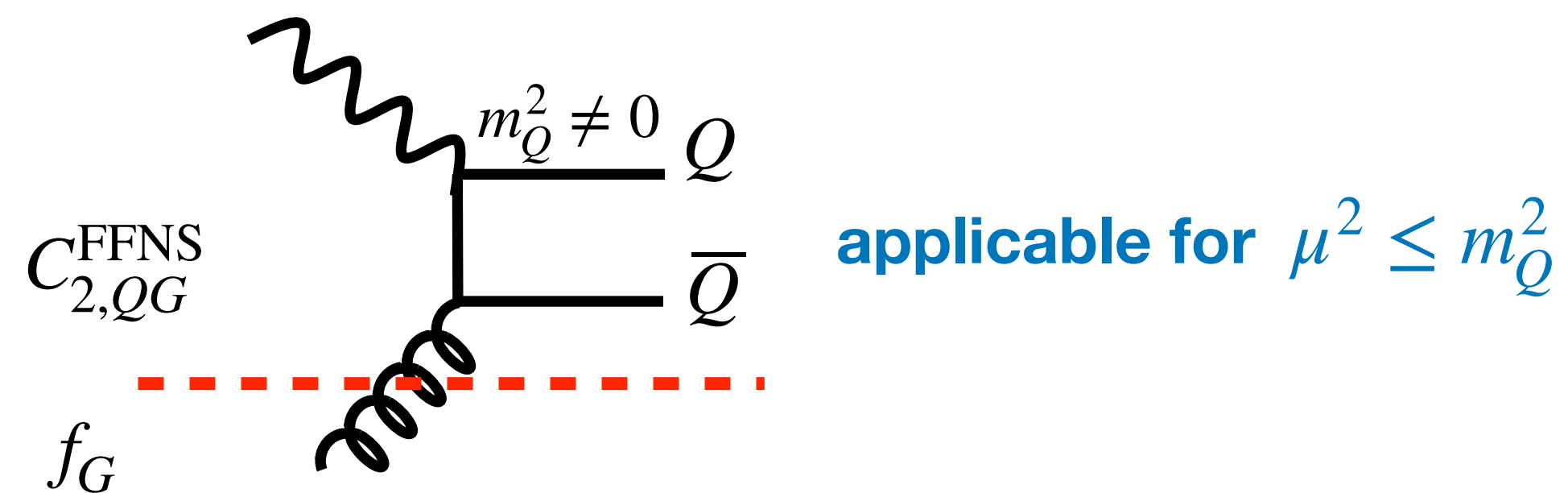
Calculable in pQCD
to certain order

Parameters
of interest

$$\sigma_{kh \rightarrow X}(x, Q^2) = \sum_{i=q, \bar{q}, G} \int_x^1 dz \, C_i \left(\frac{x}{z}, \frac{Q^2}{\mu_r^2}, \frac{\mu_f^2}{\mu_r^2}, \alpha_S(\mu_r^2) \right) f_{i,h}(z, \mu_f, \mu_r)$$

Fixed Flavour Number Scheme (FFNS)

Heavy quark Q is massive, produced in gluon fusion



- no Q PDF (absorbed in f_G)
- C_2 include terms $\mathcal{O}(m_Q^2/Q^2)$

PDF EXTRACTION CHOICES

Observable
(measurable in experiment)

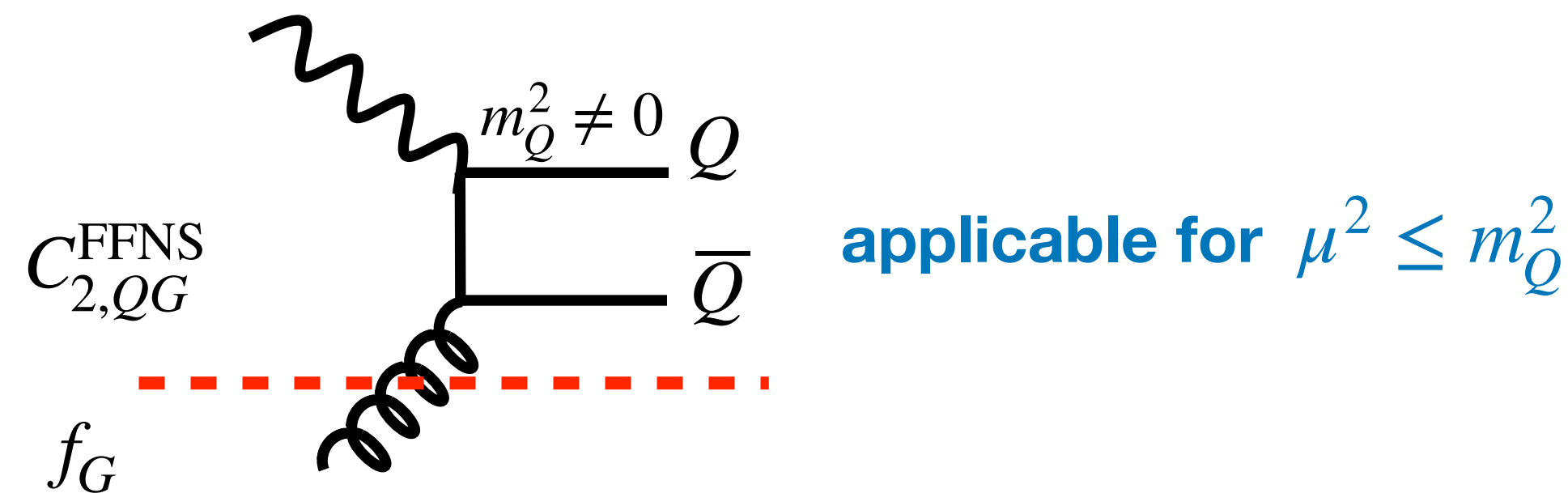
Calculable in pQCD
to certain order

Parameters
of interest

$$\sigma_{kh \rightarrow X}(x, Q^2) = \sum_{i=q, \bar{q}, G} \int_x^1 dz C_i \left(\frac{x}{z}, \frac{Q^2}{\mu_r^2}, \frac{\mu_f^2}{\mu_r^2}, \alpha_s(\mu_r^2) \right) f_{i,h}(z, \mu_f, \mu_r)$$

Fixed Flavour Number Scheme (FFNS)

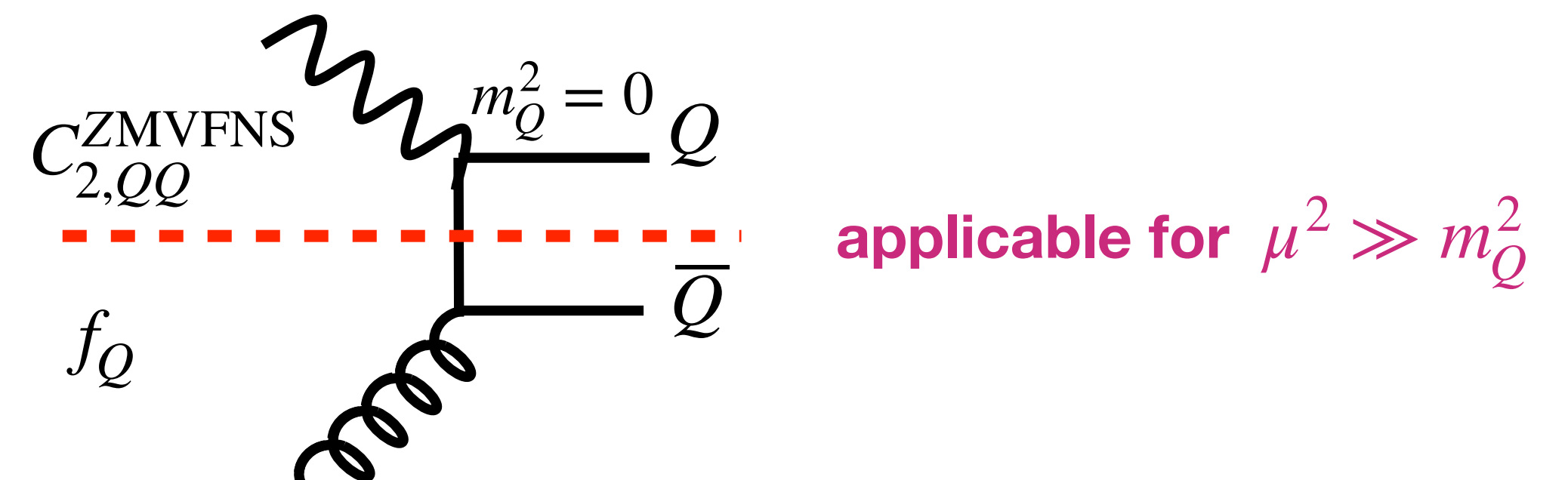
Heavy quark Q is massive, produced in gluon fusion



- no Q PDF (absorbed in f_G)
- C_2 include terms $\mathcal{O}(m_Q^2/Q^2)$

Zero-mass Variable Flavour Number Scheme (VFNS)

Heavy Quark Q is massless



- Q PDF present in the nucleon
- terms $\alpha_s \ln(Q^2/m_Q^2)$ resummed into PDF

PDF EXTRACTION CHOICES

Observable
(measurable in experiment)

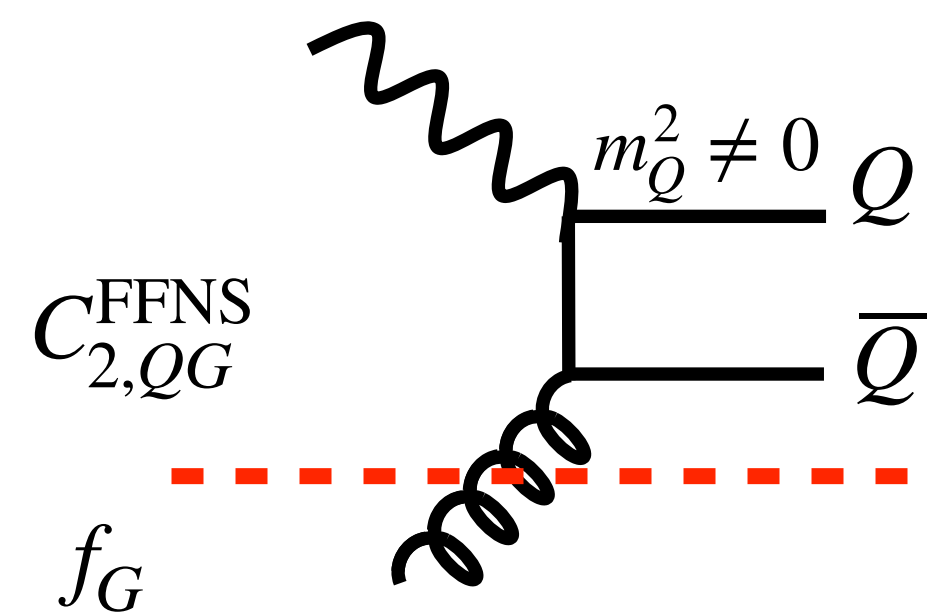
Calculable in pQCD
to certain order

Parameters
of interest

$$\sigma_{kh \rightarrow X}(x, Q^2) = \sum_{i=q, \bar{q}, G} \int_x^1 dz C_i \left(\frac{x}{z}, \frac{Q^2}{\mu_r^2}, \frac{\mu_f^2}{\mu_r^2}, \alpha_S(\mu_r^2) \right) f_{i,h}(z, \mu_f, \mu_r)$$

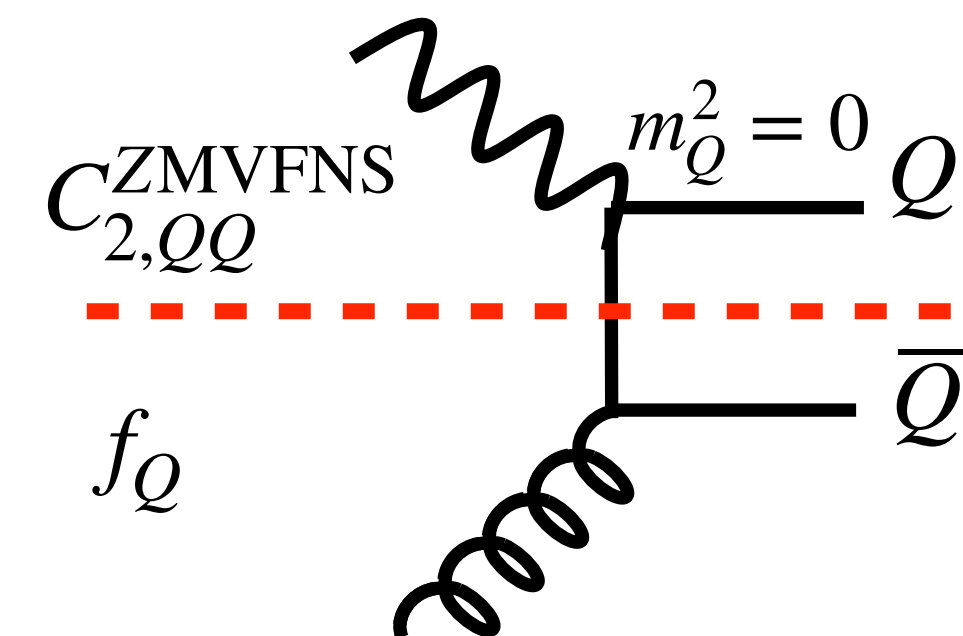
Fixed Flavour Number Scheme (FFNS)

Heavy quark Q is massive, produced in gluon fusion



Zero-mass Variable Flavour Number Scheme (VFNS)

Heavy Quark Q is massless



Generalised Mass VFNS

Match both schemes to get description over whole energy range

(involves assumptions and additional matching scale, diluted m_Q interpretation)

ORIGINS OF PDF **UNCERTAINTIES**

Observable
(measurable in experiment)

Calculable in pQCD
to certain order

Parameters
of interest

$$\sigma_{kh \rightarrow X}(x, Q^2) = \sum_{i=q, \bar{q}, G} \int_x^1 dz \, C_i \left(\frac{x}{z}, \frac{Q^2}{\mu_r^2}, \frac{\mu_f^2}{\mu_r^2}, \alpha_S(\mu_r^2) \right) f_{i,h}(z, \mu_f, \mu_r)$$

data-related:

- **experimental**
- **methodical**

scheme-related:

- model (related to m_Q)
- phenomenological (scheme choice)
- **theoretical (C_i within a scheme)**

+ perturbative order-related

- choice/variation of μ_r, μ_f ambiguous
- is $\mu_r = \mu_f = Q$ ("natural" scale)?
- **theoretical** (missing higher-orders)

ORIGINS OF DIFFERENT PDF SETS

Observable
(measurable in experiment)

Calculable in pQCD
to certain order

Parameters
of interest

$$\sigma_{kh \rightarrow X}(x, Q^2) = \sum_{i=q, \bar{q}, G} \int_x^1 dz \, C_i \left(\frac{x}{z}, \frac{Q^2}{\mu_r^2}, \frac{\mu_f^2}{\mu_r^2}, \alpha_S(\mu_r^2) \right) f_{i,h}(z, \mu_f, \mu_r)$$

data-related:

- **experimental**
- **methodical**

scheme-related:

- model (related to m_Q)
- phenomenological (scheme choice)
- **theoretical (C_i within a scheme)**

perturbative order-related

- choice/variation of μ_r, μ_f ambiguous
- is $\mu_r = \mu_f = Q$ ("natural" scale)?
- **theoretical** (missing higher-orders)

Several dedicated phenomenology groups provide PDFs based on particular choices

MAJOR PDF FITTING GROUPS (*alphabetic order*)

ABKM/ABM/ABMP global PDF fitting group

Heavy flavour scheme: FFNS

Latest sets ABMP16 [arXiv:1701.05838], ABMP16tt [EPJC 85 (2025) 2, 162]

CT/CTEQ/CJ/TEA (<https://cteq.gitlab.io/>) global PDF fitting group

Heavy flavour scheme: GM-VFNS in ACOT implementation

Latest CT18 sets [[arXiv:1912.10053](https://arxiv.org/abs/1912.10053)]

HERAPDF (https://www.desy.de/h1zeus/combined_results/) only DIS collider data from HERA

Heavy flavour scheme: different schemes, VFNS in the Thorne-Roberts implementation

Latest HERAPDF2.0 sets [EPJC 75 (2015) 12, 580]

MMHT/MSTW/MSHT global PDF fitting group

Heavy flavour scheme: VFNS in a Thorne-Roberts implementation

Latest set MSHT20: [SciPost Phys. 17 (2024) 1, 026]

NNPDF (<https://nnpdf.mi.infn.it/>) global PDF fitting group

Heavy flavour scheme: VFNS and FFNS

Latest NNPDF4.0 [EPJC 82 (2022) 5, 428]

OPEN-SOURCE PDF-FIT RELATED TOOLS

ABM/ABMP : OPENQCDRAD [PRD86 (2012) 054009]: numerical computation of all hard scattering cross sections (DIS structure function calculation including heavy quark contributions, W and Z production)

CT/CTEQ/CJ: Fantômas (2025) toolkit for exploring the parametrization dependence of (PDFs) and other correlator functions in QCD

HERAPDF : HERAFitter/xFitter: a framework for the global QCD analyses, SM parameter extraction, benchmarking of different Flavour schemes, SMEFT fits

NNPDF: APFEL (PDF evolution library); **n3fit**: global QCD analysis code; **Colibri (2025)** fast PDF fit code + other tools by NNPDF (<https://nnpdf.mi.infn.it/>)

QCD ANALYSIS: BASIC IDEA

Select data and theory



Parametrize PDFs $f(x, \mu_0^2)$
at a starting scale μ_0^2



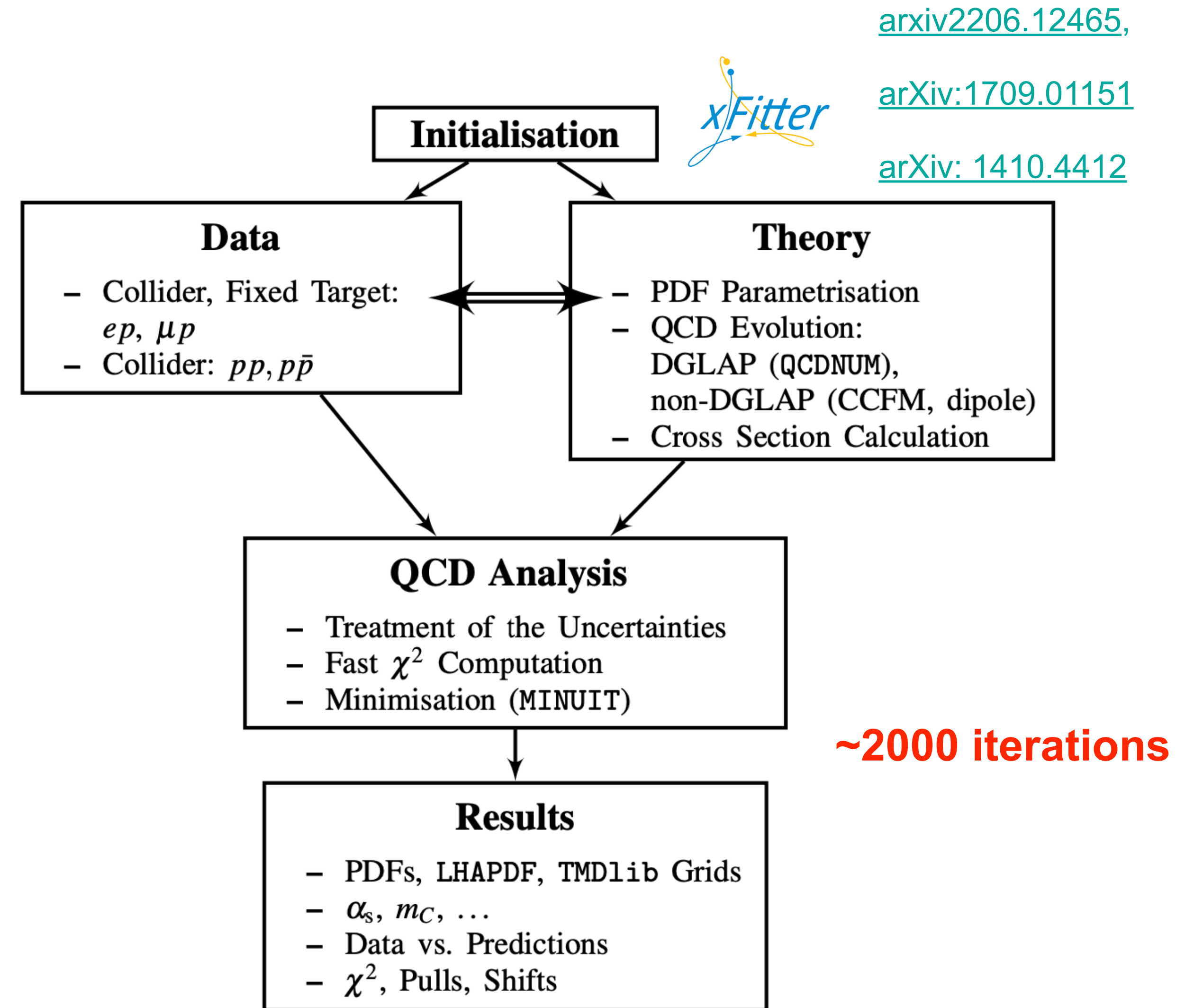
Perform a fit
(minimize χ^2)



Output PDFs $f(x, \mu^2)$

QCD ANALYSIS: BASIC IDEA

Example:



Select data and theory



Parametrize PDFs $f(x, \mu_0^2)$
at a starting scale μ_0^2



Perform a fit
(minimize χ^2)



Output PDFs $f(x, \mu^2)$

STEP1: THEORY [PERTURBATIVE ORDER]

Select data and **theory**



Parametrize PDFs $f(x, \mu_0^2)$
at a starting scale μ_0^2



Perform a fit
(minimize χ^2)



Output PDFs $f(x, \mu^2)$

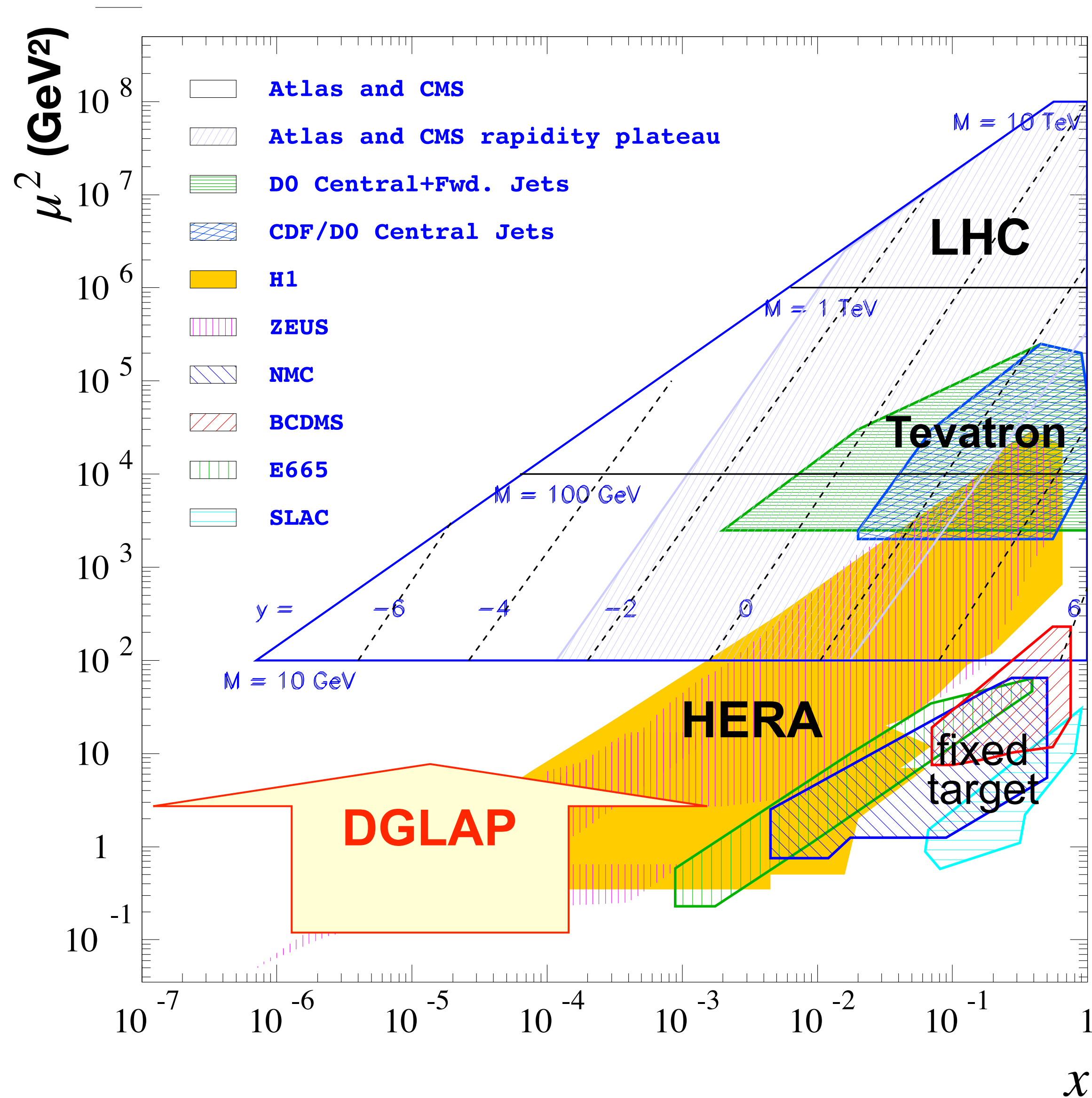
- state-of-the-art NNLO in QCD with NLO QED corrections pushing perturbative precision to N3LO...

Theory Ingredients [not complete !]		Availability
Splitting functions		Partial N3LO
coefficient functions (partonic cross sections)	• DIS, light flavors	Full N3LO
	• NC DIS, heavy flavors	Full N3LO (Blümlein et al.), not yet in fitting codes
	• Vector boson production	Full N3LO for some processes, fixed N3LO/NLO K-factor tables
	• CC DIS, jet, $t\bar{t}$ production	N2LO
	• $pp \rightarrow W + c, pp \rightarrow Z + b, pp \rightarrow b$	NLO (massive); NNLO (ZM)

T. Hobbs, QCD@LHC2025

- issues: quark mass, final states with additional jets, full color, resummation, scale choice and variations (theory nuisance parameters?)
- specific to PDF fits: availability of **fast interpolation grids**
fastNLO, APPLgrid, APPLfast, PineAPPL

STEP1: SELECT DATA



PDFs are intrinsic property of the nucleon

→ **process independent**

Remind: x -dependence is deduced from the data:

coverage in x is important !

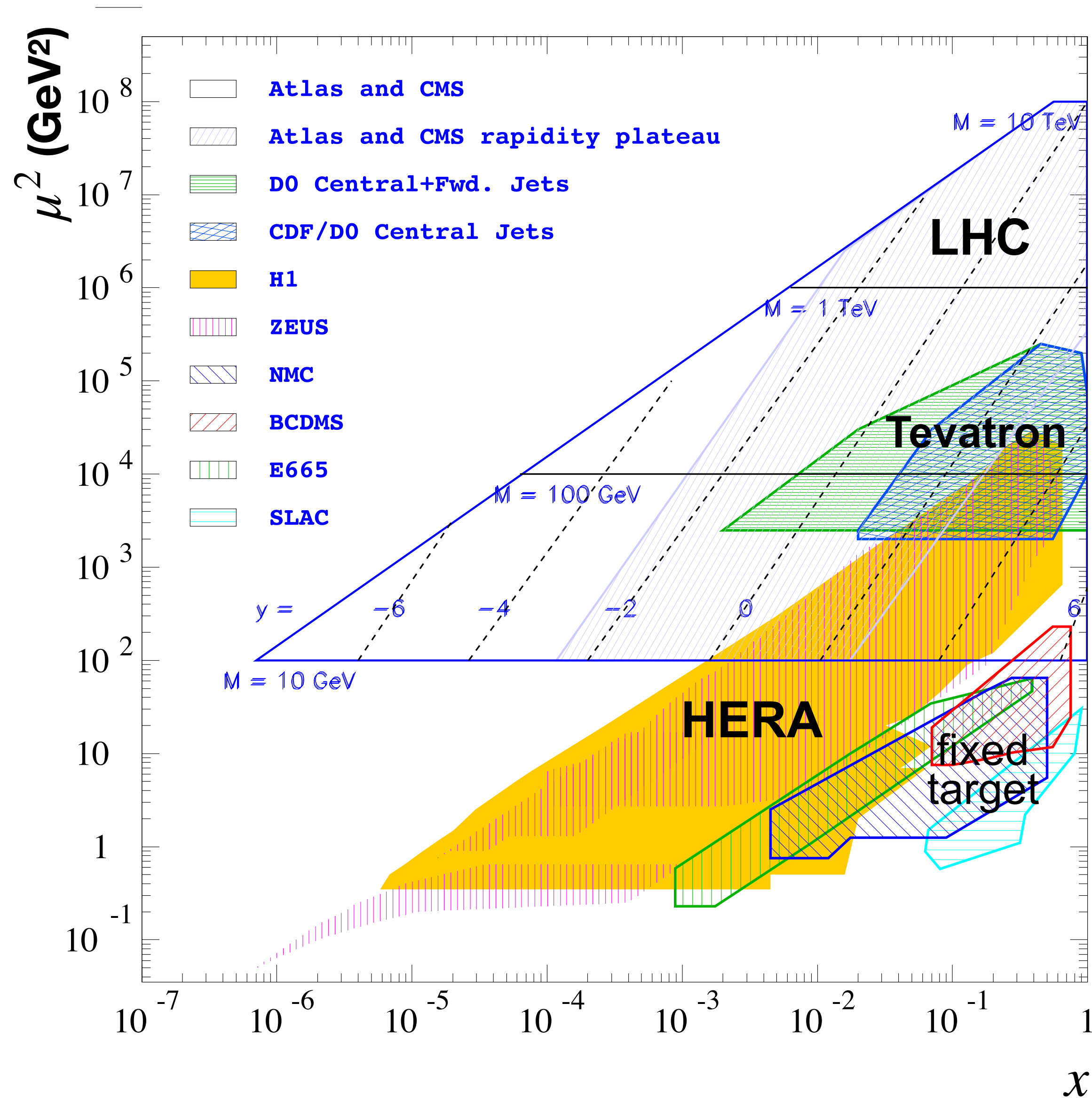
STEP1: DATA FOR PDF EXTRACTION

ℓN Deep Inelastic Scattering (DIS)

- **fixed target experiments**
- SLAC: ep , $\sqrt{s} = 2 - 5$ GeV; NMC, BCDMS@ SPS (CERN) : μN , $\sqrt{s} \sim 12 - 22$ GeV; E665 at Tevatron (Fermilab): μN , $\sqrt{s} \sim 30$ GeV

→ **PDFs at high x**

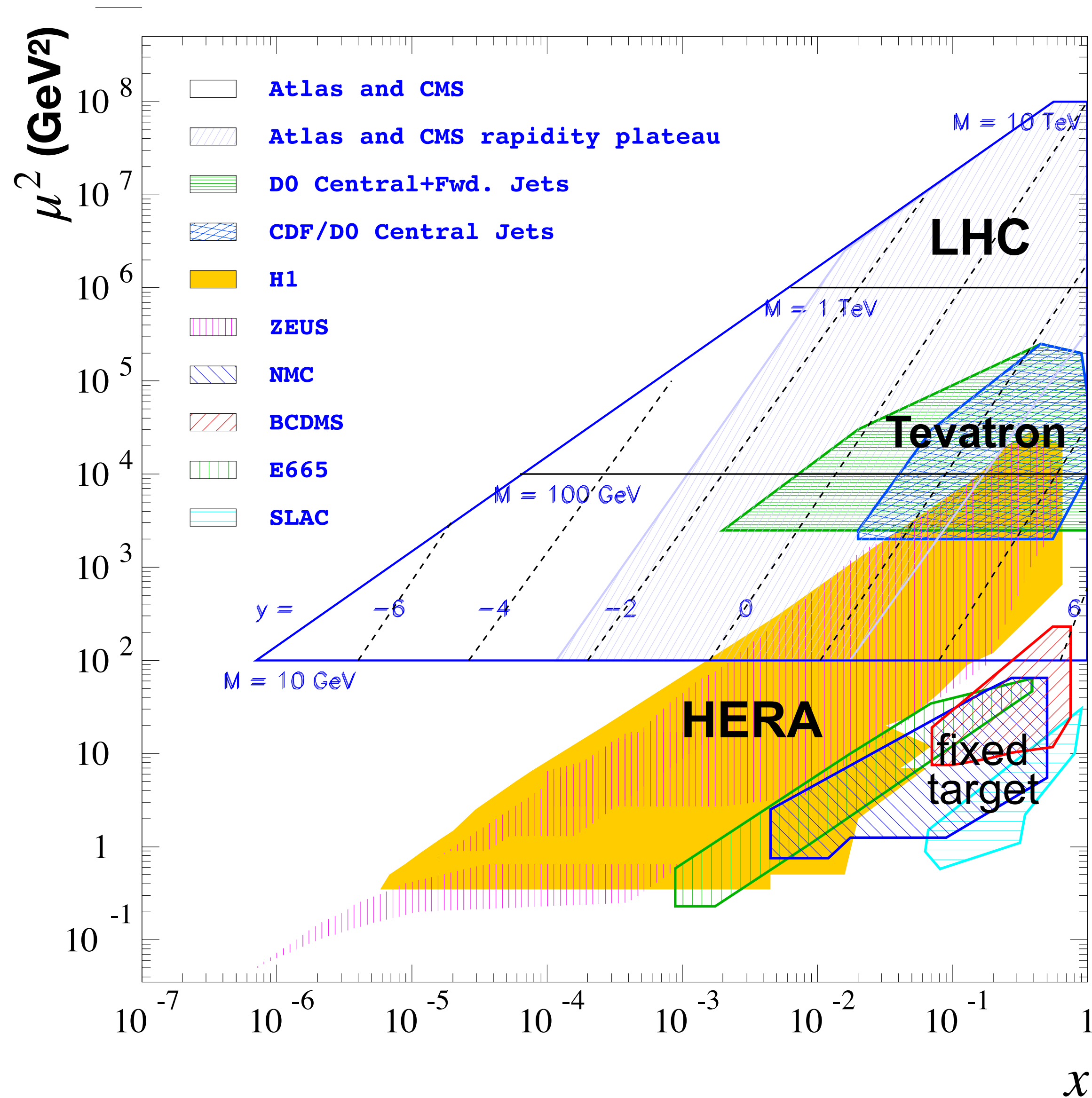
→ **flavour decomposition, $s(x)$ for $x > 0.01$**



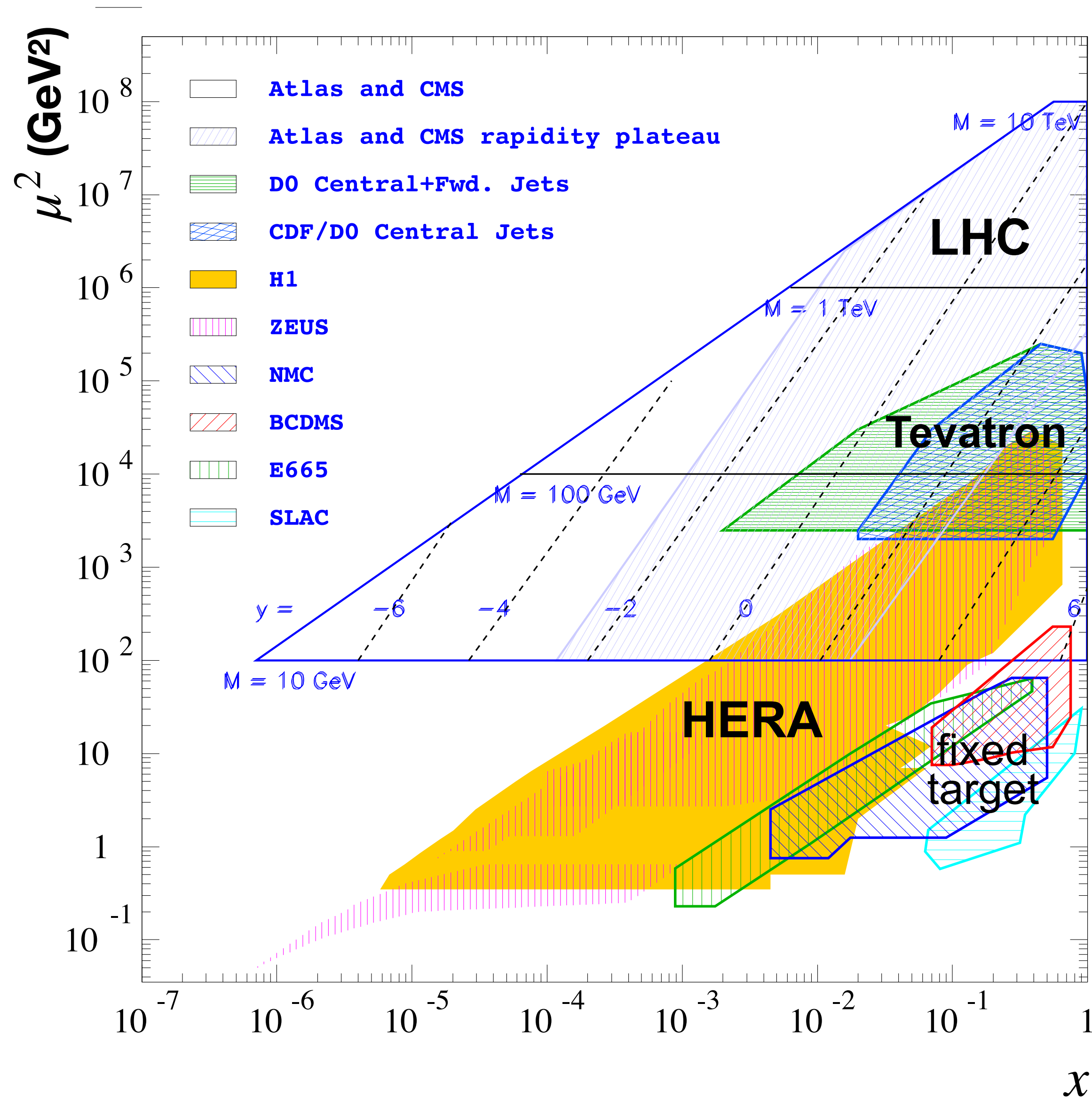
STEP1: DATA FOR PDF EXTRACTION

$\ln$ Deep Inelastic Scattering (DIS)

- **fixed target experiments**
- SLAC: ep , $\sqrt{s} = 2 - 5$ GeV; NMC, BCDMS@ SPS (CERN) : μN , $\sqrt{s} \sim 12 - 22$ GeV; E665 at Tevatron (Fermilab): μN , $\sqrt{s} \sim 30$ GeV
- **DIS in collisions**
H1, ZEUS@HERA (DESY) $\sqrt{s} \sim 225 - 318$ GeV
- **valence and gluon, flavour separation@medium x**
- **jets and heavy quarks: gluon, α_S , m_c , m_b**



STEP1: DATA FOR PDF EXTRACTION



ℓN Deep Inelastic Scattering (DIS)

- fixed target experiments
- SLAC: ep , $\sqrt{s} = 2 - 5 \text{ GeV}$; NMC, BCDMS@ SPS (CERN) : μN , $\sqrt{s} \sim 12 - 22 \text{ GeV}$; E665 at Tevatron (Fermilab): μN , $\sqrt{s} \sim 30 \text{ GeV}$

DIS in collisions

H1, ZEUS@HERA (DESY) $\sqrt{s} \sim 225 - 318 \text{ GeV}$

→ valence and gluon, flavour separation@medium x

→ jets and heavy quarks: gluon, α_S , m_c , m_b

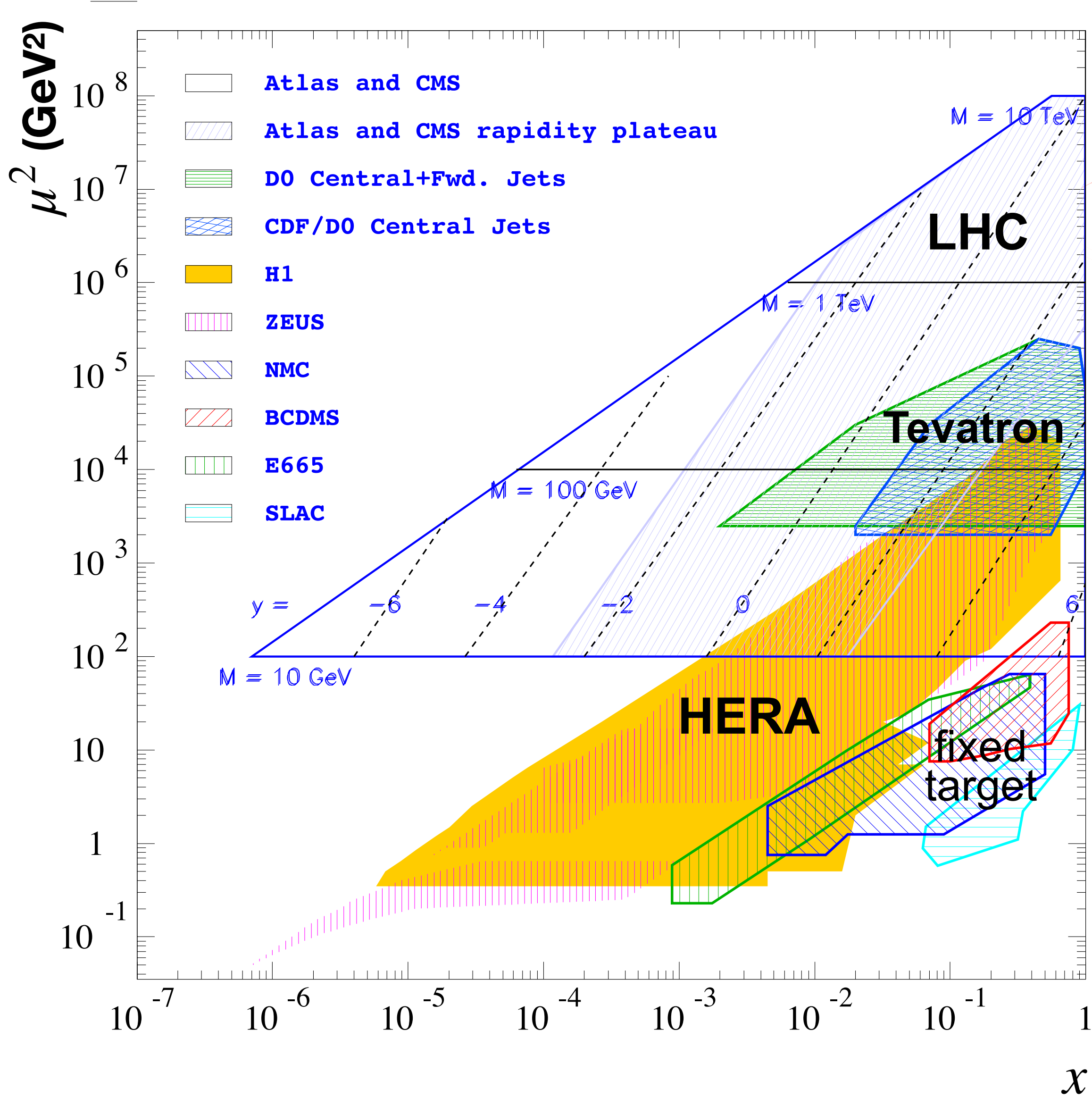
Hadronic collisions

- $p\bar{p}$: CDF/D0@ Tevatron $\sqrt{s} = 1.8, 1.96 \text{ TeV}$
- pp : ATLAS/CMS/LHCb@LHC $\sqrt{s} \sim 2 - 13.6 \text{ TeV}$

→ flavour separation @ medium-high x

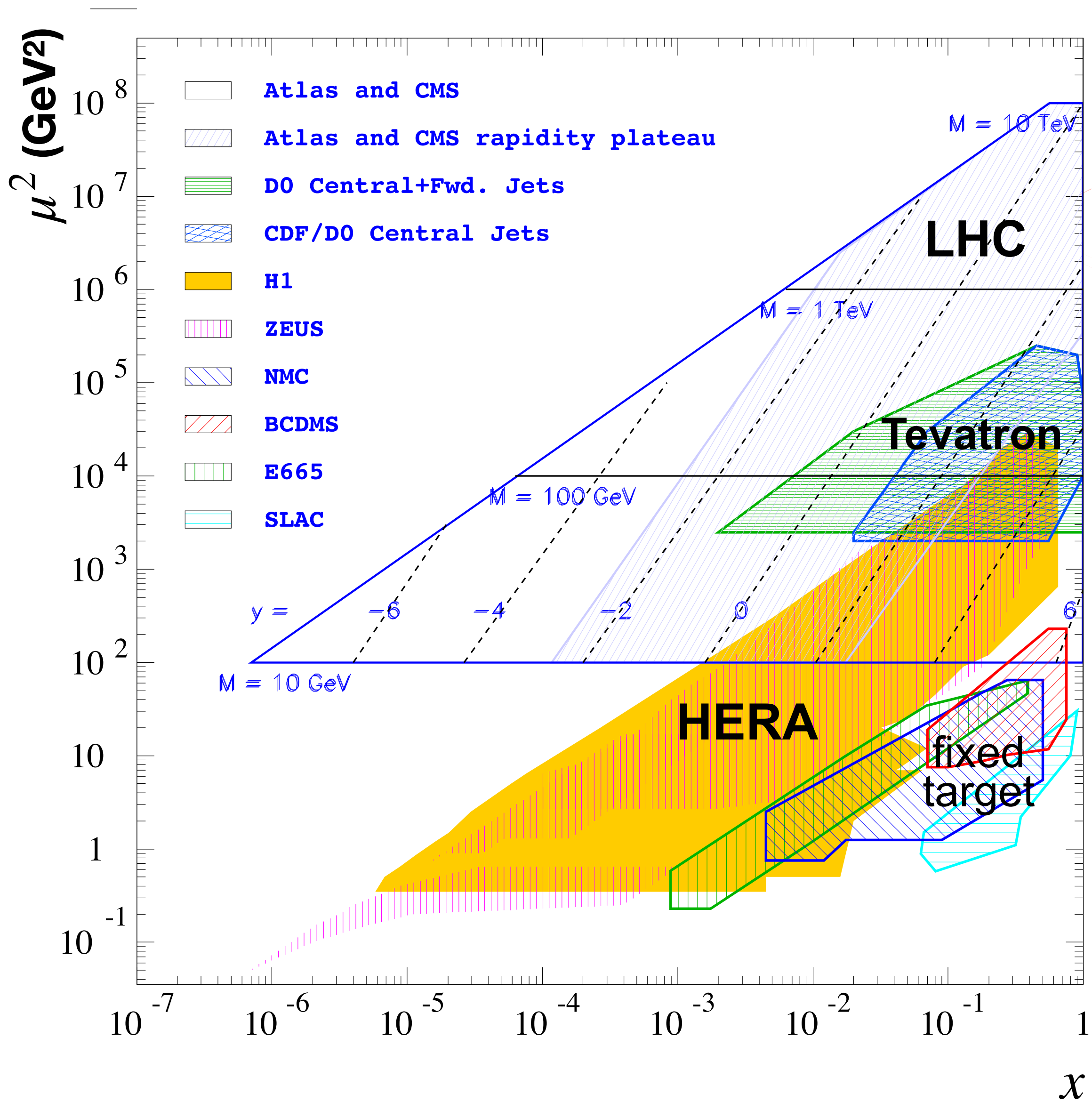
→ gluon @ high and low x , α_S , m_t ,

STEP1: DATA FOR PDF EXTRACTION



	Process	Subprocess	Partons	x range
Fixed Target	$\ell^\pm \{p, n\} \rightarrow \ell^\pm + X$	$\gamma^* q \rightarrow q$	$q, \bar{q}, g$	$x \gtrsim 0.01$
	$\ell^\pm n/p \rightarrow \ell^\pm + X$	$\gamma^* d/u \rightarrow d/u$	d/u	$x \gtrsim 0.01$
	$pp \rightarrow \mu^+ \mu^- + X$	$u\bar{u}, d\bar{d} \rightarrow \gamma^*$	$\bar{q}$	$0.015 \lesssim x \lesssim 0.35$
	$pn/pp \rightarrow \mu^+ \mu^- + X$	$(u\bar{d})/(u\bar{u}) \rightarrow \gamma^*$	$\bar{d}/\bar{u}$	$0.015 \lesssim x \lesssim 0.35$
	$\nu(\bar{\nu}) N \rightarrow \mu^-(\mu^+) + X$	$W^* q \rightarrow q'$	$q, \bar{q}$	$0.01 \lesssim x \lesssim 0.5$
	$\nu N \rightarrow \mu^- \mu^+ + X$	$W^* s \rightarrow c$	s	$0.01 \lesssim x \lesssim 0.2$
	$\bar{\nu} N \rightarrow \mu^+ \mu^- + X$	$W^* \bar{s} \rightarrow \bar{c}$	$\bar{s}$	$0.01 \lesssim x \lesssim 0.2$
Collider DIS	$e^\pm p \rightarrow e^\pm + X$	$\gamma^* q \rightarrow q$	$g, q, \bar{q}$	$0.0001 \lesssim x \lesssim 0.1$
	$e^+ p \rightarrow \bar{\nu} + X$	$W^+ \{d, s\} \rightarrow \{u, c\}$	d, s	$x \gtrsim 0.01$
	$e^\pm p \rightarrow e^\pm c\bar{c} + X$	$\gamma^* c \rightarrow c, \gamma^* g \rightarrow c\bar{c}$	c, g	$10^{-4} \lesssim x \lesssim 0.01$
	$e^\pm p \rightarrow e^\pm b\bar{b} + X$	$\gamma^* b \rightarrow b, \gamma^* g \rightarrow b\bar{b}$	b, g	$10^{-4} \lesssim x \lesssim 0.01$
	$e^\pm p \rightarrow \text{jet} + X$	$\gamma^* g \rightarrow q\bar{q}$	g	$0.01 \lesssim x \lesssim 0.1$
Tevatron	$p\bar{p} \rightarrow \text{jet} + X$	$gg, qg, q\bar{q} \rightarrow 2j$	g, q	$0.01 \lesssim x \lesssim 0.5$
	$p\bar{p} \rightarrow (W^\pm \rightarrow \ell^\pm \nu) + X$	$ud \rightarrow W^+, \bar{u}\bar{d} \rightarrow W^-$	$u, d, \bar{u}, \bar{d}$	$x \gtrsim 0.05$
	$p\bar{p} \rightarrow (Z \rightarrow \ell^+ \ell^-) + X$	$uu, dd \rightarrow Z$	u, d	$x \gtrsim 0.05$
	$p\bar{p} \rightarrow t\bar{t} + X$	$q\bar{q} \rightarrow t\bar{t}$	q	$x \gtrsim 0.1$
LHC	$pp \rightarrow \text{jet} + X$	$gg, qg, q\bar{q} \rightarrow 2j$	g, q	$0.001 \lesssim x \lesssim 0.5$
	$pp \rightarrow (W^\pm \rightarrow \ell^\pm \nu) + X$	$u\bar{d} \rightarrow W^+, d\bar{u} \rightarrow W^-$	$u, d, \bar{u}, \bar{d}, g$	$x \gtrsim 10^{-3}$
	$pp \rightarrow (Z \rightarrow \ell^+ \ell^-) + X$	$q\bar{q} \rightarrow Z$	$q, \bar{q}, g$	$x \gtrsim 10^{-3}$
	$pp \rightarrow (Z \rightarrow \ell^+ \ell^-) + X, p_\perp$	$gq(\bar{q}) \rightarrow Zq(\bar{q})$	$g, q, \bar{q}$	$x \gtrsim 0.01$
	$pp \rightarrow (\gamma^* \rightarrow \ell^+ \ell^-) + X, \text{Low mass}$	$q\bar{q} \rightarrow \gamma^*$	$q, \bar{q}, g$	$x \gtrsim 10^{-4}$
	$pp \rightarrow (\gamma^* \rightarrow \ell^+ \ell^-) + X, \text{High mass}$	$q\bar{q} \rightarrow \gamma^*$	$\bar{q}$	$x \gtrsim 0.1$
	$pp \rightarrow W^+ \bar{c}, W^- c$	$sg \rightarrow W^+ c, \bar{s}g \rightarrow W^- \bar{c}$	$s, \bar{s}$	$x \sim 0.01$
	$pp \rightarrow t\bar{t} + X$	$gg \rightarrow t\bar{t}$	g	$x \gtrsim 0.01$
	$pp \rightarrow D, B + X$	$gg \rightarrow c\bar{c}, b\bar{b}$	g	$x \gtrsim 10^{-6}, 10^{-5}$
	$pp \rightarrow J/\psi, \Upsilon + pp$	$\gamma^*(gg) \rightarrow c\bar{c}, b\bar{b}$	g	$x \gtrsim 10^{-6}, 10^{-5}$
	$pp \rightarrow \gamma + X$	$gq(\bar{q}) \rightarrow \gamma q(\bar{q})$	g	$x \gtrsim 0.005$

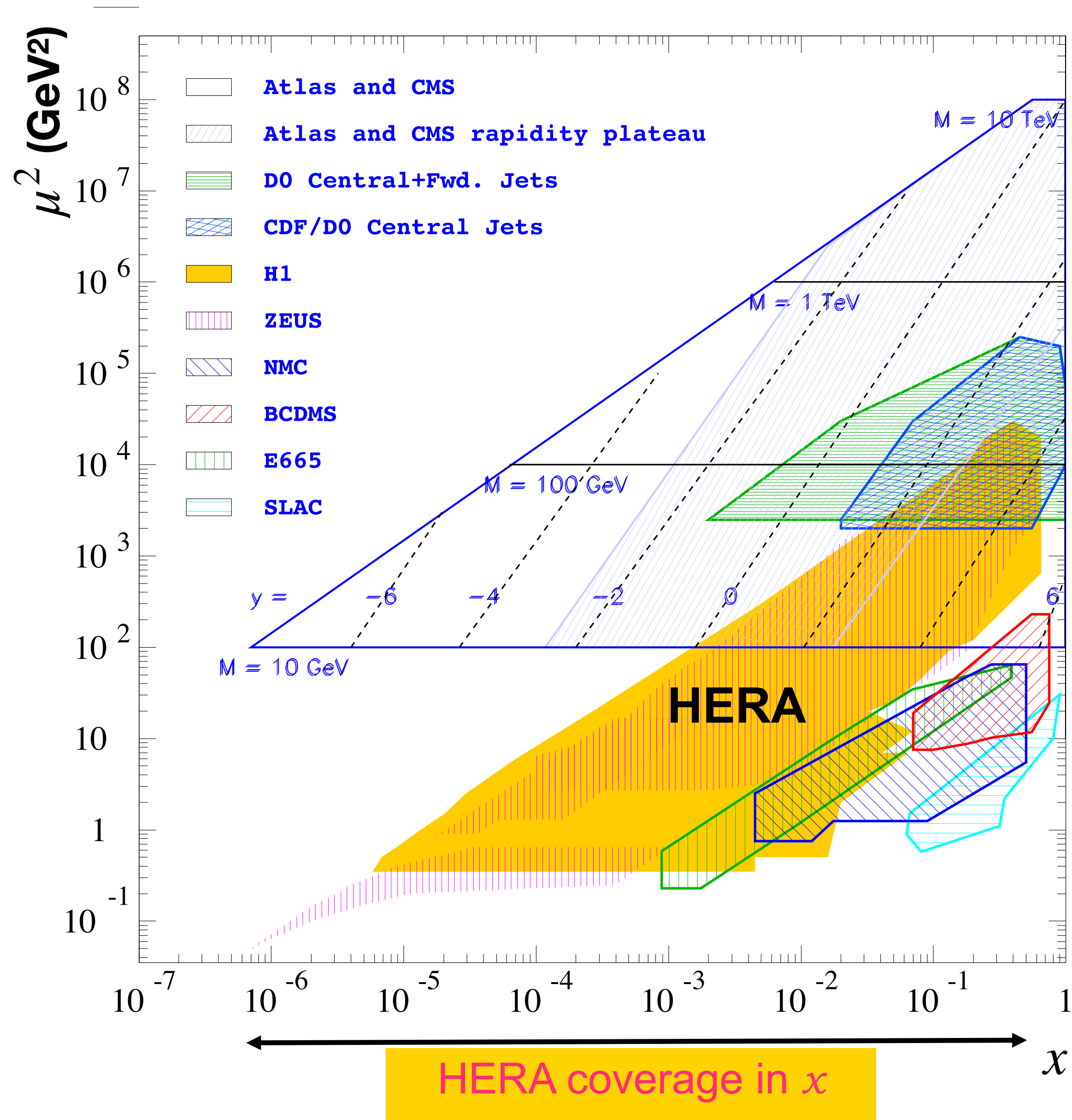
STEP1: DATA FOR PDF EXTRACTION



	Process	Subprocess	Partons	x range
Fixed Target	$\ell^\pm \{p, n\} \rightarrow \ell^\pm + X$	$\gamma^* q \rightarrow q$	$q, \bar{q}, g$	$x \gtrsim 0.01$
	$\ell^\pm n/p \rightarrow \ell^\pm + X$	$\gamma^* d/u \rightarrow d/u$	d/u	$x \gtrsim 0.01$
	$pp \rightarrow \mu^+ \mu^- + X$	$u\bar{u}, d\bar{d} \rightarrow \gamma^*$	$\bar{q}$	$0.015 \lesssim x \lesssim 0.35$
	$pn/pp \rightarrow \mu^+ \mu^- + X$	$(u\bar{d})/(u\bar{u}) \rightarrow \gamma^*$	$\bar{d}/\bar{u}$	$0.015 \lesssim x \lesssim 0.35$
	$\nu(\bar{\nu}) N \rightarrow \mu^-(\mu^+) + X$	$W^* q \rightarrow q'$	$q, \bar{q}$	$0.01 \lesssim x \lesssim 0.5$
	$\nu N \rightarrow \mu^- \mu^+ + X$	$W^* s \rightarrow c$	s	$0.01 \lesssim x \lesssim 0.2$
	$\bar{\nu} N \rightarrow \mu^+ \mu^- + X$	$W^* \bar{s} \rightarrow \bar{c}$	$\bar{s}$	$0.01 \lesssim x \lesssim 0.2$
	$e^\pm p \rightarrow e^\pm + X$	$\gamma^* q \rightarrow q$	$g, q, \bar{q}$	$0.0001 \lesssim x \lesssim 0.1$
Collider	$e^+ n \rightarrow \bar{\nu} + X$	$W^+ \{d, s\} \rightarrow \{u, c\}$	d, s	$x \gtrsim 0.01$
				$10^{-4} \lesssim x \lesssim 0.01$
				$10^{-4} \lesssim x \lesssim 0.01$
				$0.01 \lesssim x \lesssim 0.1$
Tevatron				$0.01 \lesssim x \lesssim 0.5$
				$x \gtrsim 0.05$
				$x \gtrsim 0.05$
				$x \gtrsim 0.1$
LHC	$pp \rightarrow (\gamma^* \rightarrow \ell^+ \ell^-) + X$, Low mass	$q\bar{q} \rightarrow \gamma^*$	$q, \bar{q}, g$	$x \gtrsim 10^{-4}$
	$pp \rightarrow (\gamma^* \rightarrow \ell^+ \ell^-) + X$, High mass	$q\bar{q} \rightarrow \gamma^*$	$\bar{q}$	$x \gtrsim 0.1$
	$pp \rightarrow W^+ \bar{c}, W^- c$	$sg \rightarrow W^+ c, \bar{s}g \rightarrow W^- \bar{c}$	$s, \bar{s}$	$x \sim 0.01$
	$pp \rightarrow t\bar{t} + X$	$gg \rightarrow t\bar{t}$	g	$x \gtrsim 0.01$
	$pp \rightarrow D, B + X$	$gg \rightarrow c\bar{c}, b\bar{b}$	g	$x \gtrsim 10^{-6}, 10^{-5}$
	$pp \rightarrow J/\psi, \Upsilon + pp$	$\gamma^*(gg) \rightarrow c\bar{c}, b\bar{b}$	g	$x \gtrsim 10^{-6}, 10^{-5}$
	$pp \rightarrow \gamma + X$	$gq(\bar{q}) \rightarrow \gamma q(\bar{q})$	g	$x \gtrsim 0.005$

50 – 60 individual data sets
4000 – 5000 measurements (data points)

STEP1: DATA FOR PDF EXTRACTION



- **DIS in collisions**

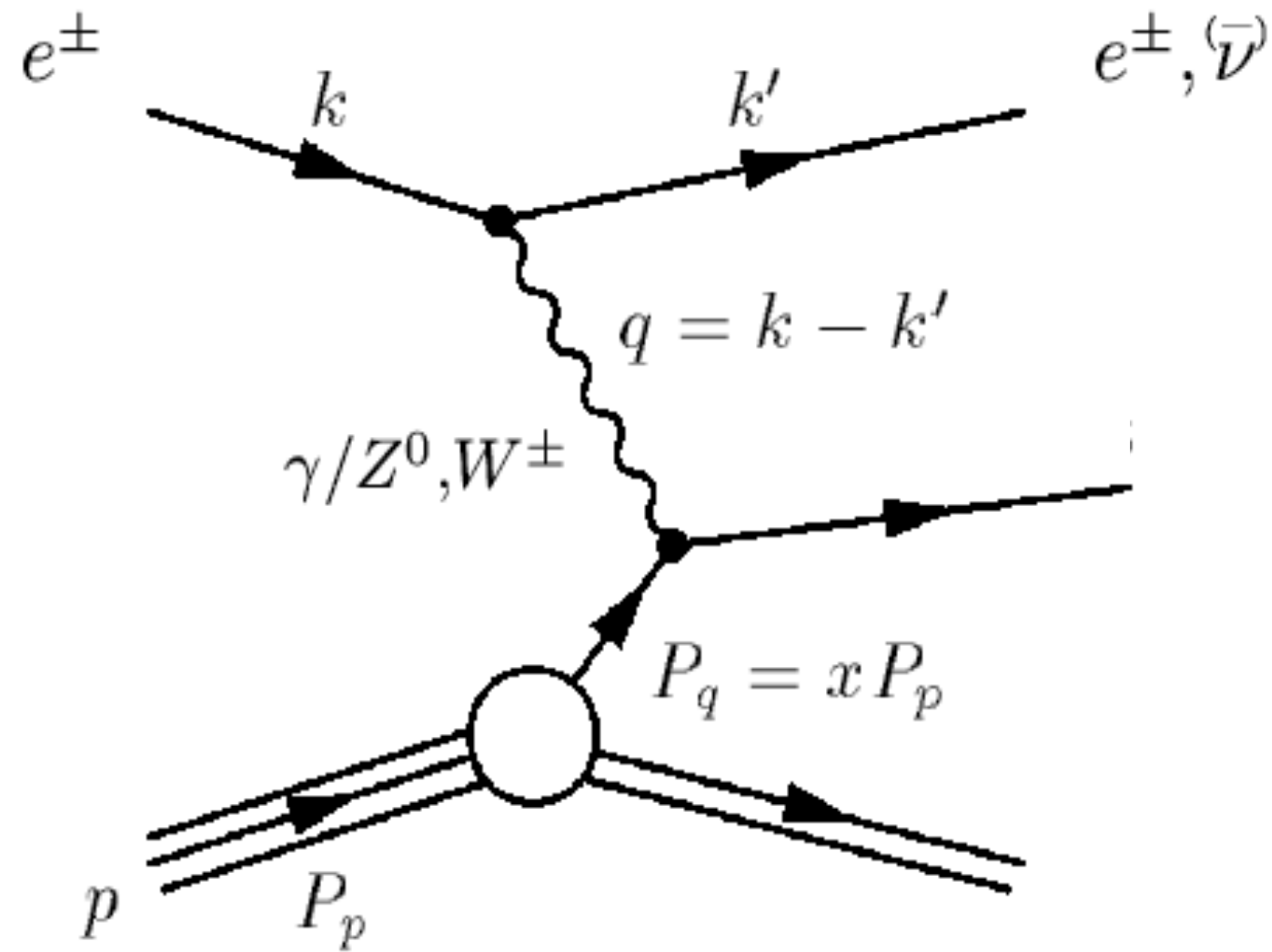
H1, ZEUS@HERA (DESY) $\sqrt{s} \sim 225 - 318$ GeV

point-like probe, large span in x :

excellent constraints at low and medium x

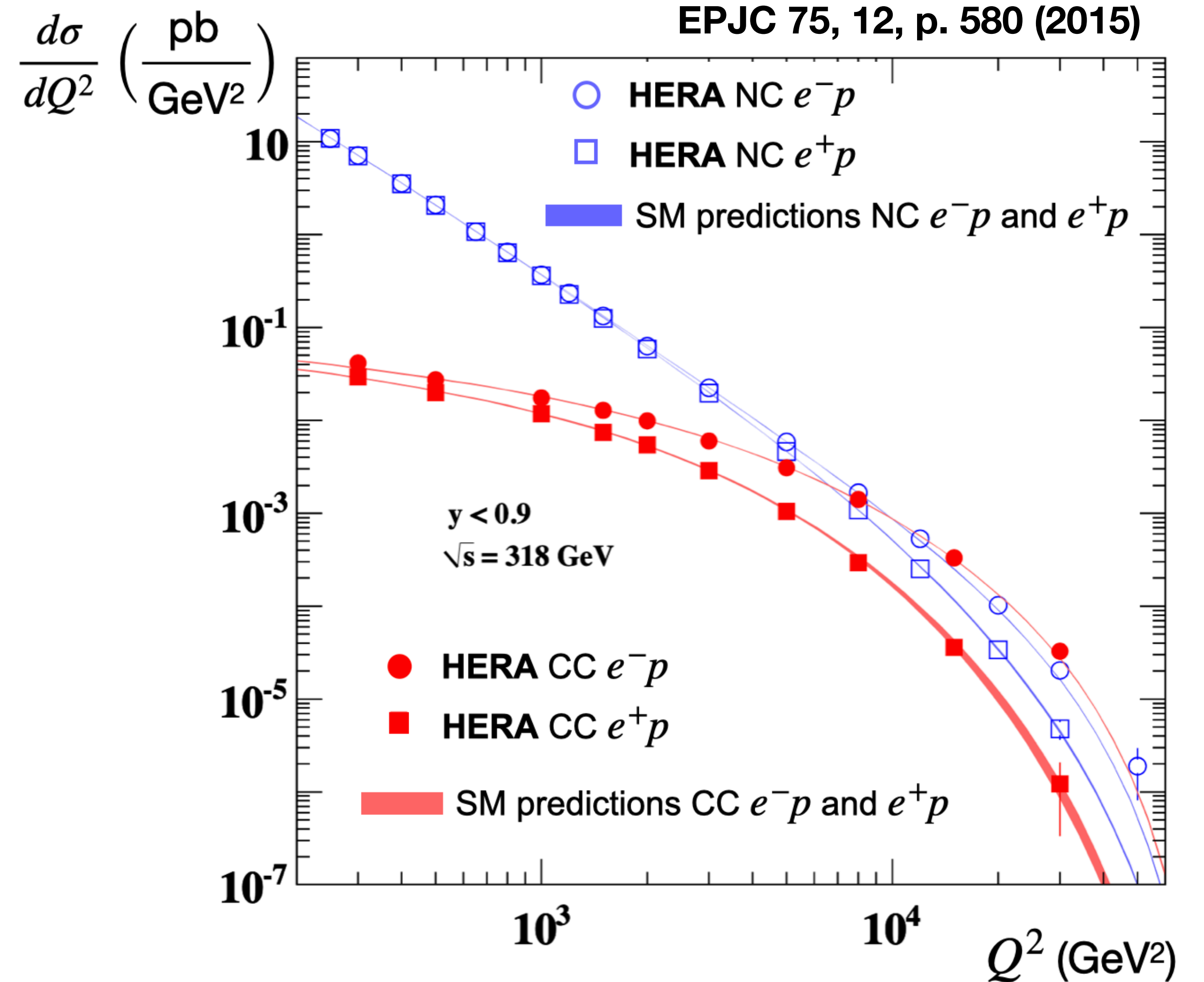
backbone of all available PDFs

DIS CROSS SECTIONS AND PDFs

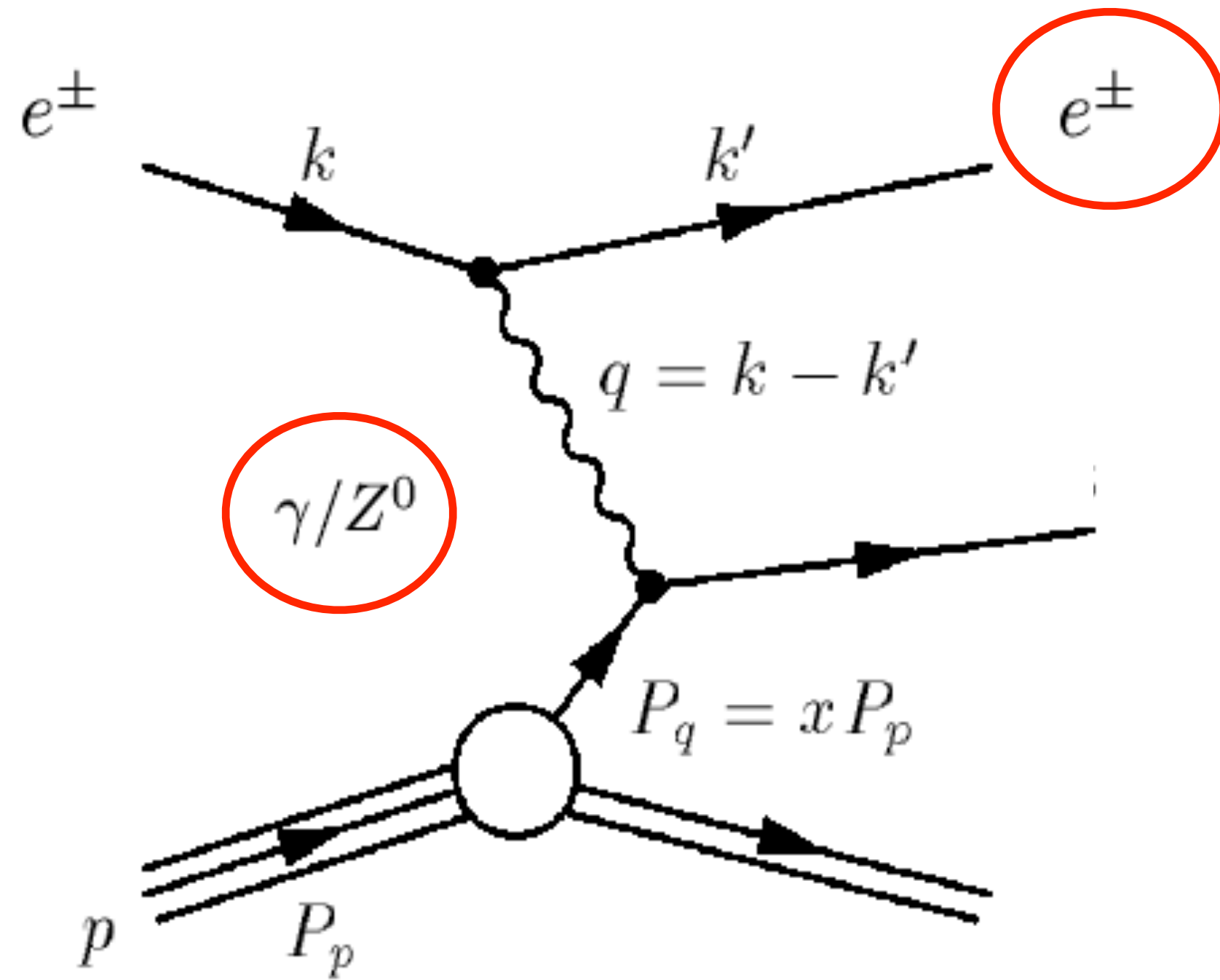


Exchanged boson:

- γ^* , Z boson: Neutral Current (dominant)
- $W^\pm$ boson: Charged Current



DIS CROSS SECTIONS AND PDFs: NEUTRAL CURRENT



$$\frac{d^2\sigma_{NC}(e^\pm p)}{dx dQ^2} = \frac{2\pi\alpha^2}{x} \left[\frac{1}{Q^2} \right]^2 \phi_{NC}^\pm(x, Q^2)$$

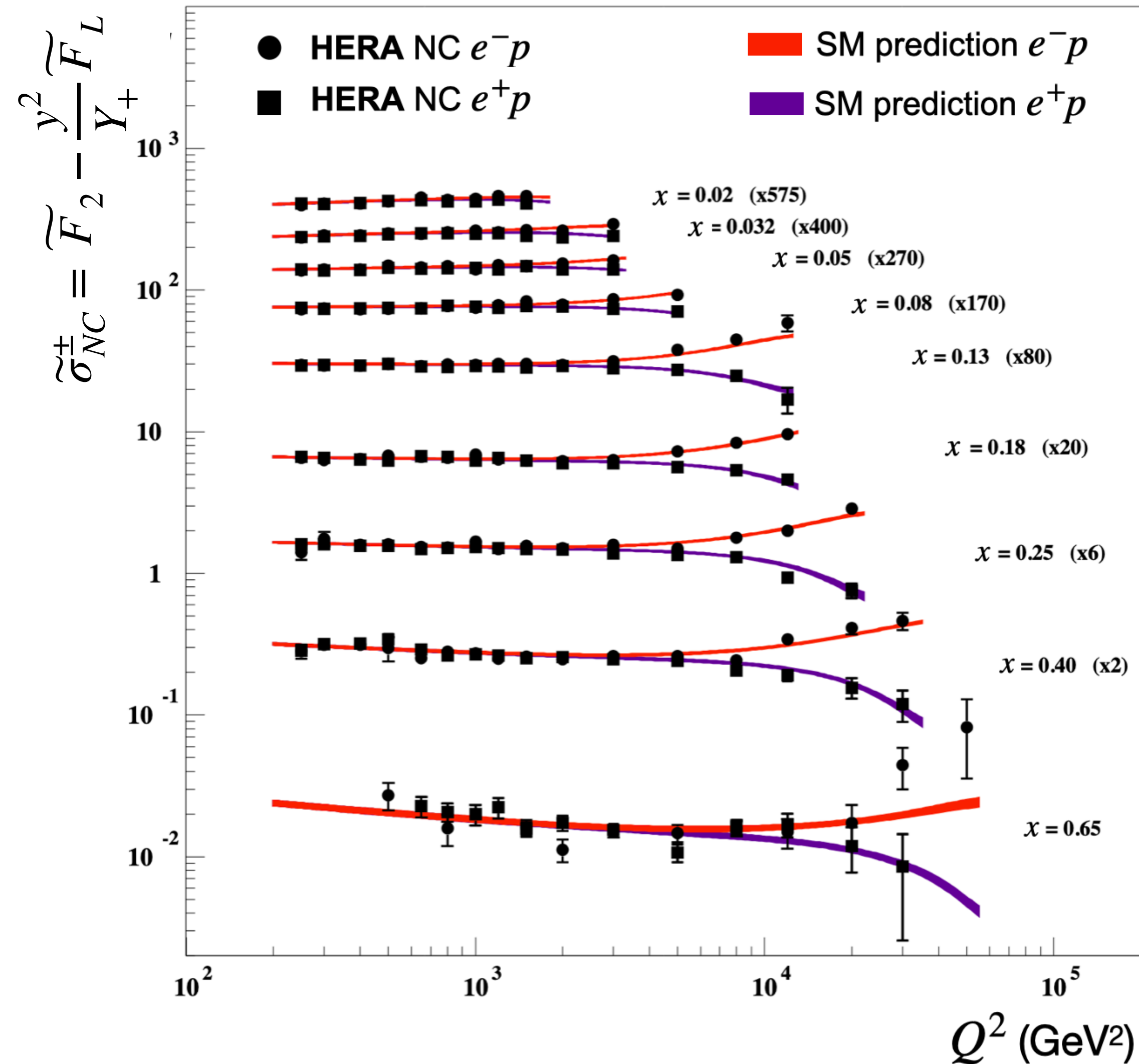
structure function term

kinematics: $x, Q^2, y = \frac{p \cdot q}{p \cdot k}$ “inelasticity”,

$$Y_\pm = 1 \pm (1 - y)^2 \text{ (related to boson polarization)}$$

DIS CROSS SECTIONS AND PDFs: NEUTRAL CURRENT

EPJC 75, 12, p. 580 (2015)



$$y = \frac{p \cdot q}{p \cdot k} \text{ "inelasticity", } Y_{\pm} = 1 \pm (1 - y)^2$$

$$\frac{d^2\sigma_{NC}(e^{\pm}p)}{dx dQ^2} = \frac{2\pi\alpha^2}{x} \left[\frac{1}{Q^2} \right]^2 \phi_{NC}^{\pm}(x, Q^2)$$

$$Y_+ \widetilde{F}_2^{\pm}(x, Q^2) \mp Y_- x \widetilde{F}_3^{\pm}(x, Q^2) - y^2 \widetilde{F}_L^{\pm}(x, Q^2)$$

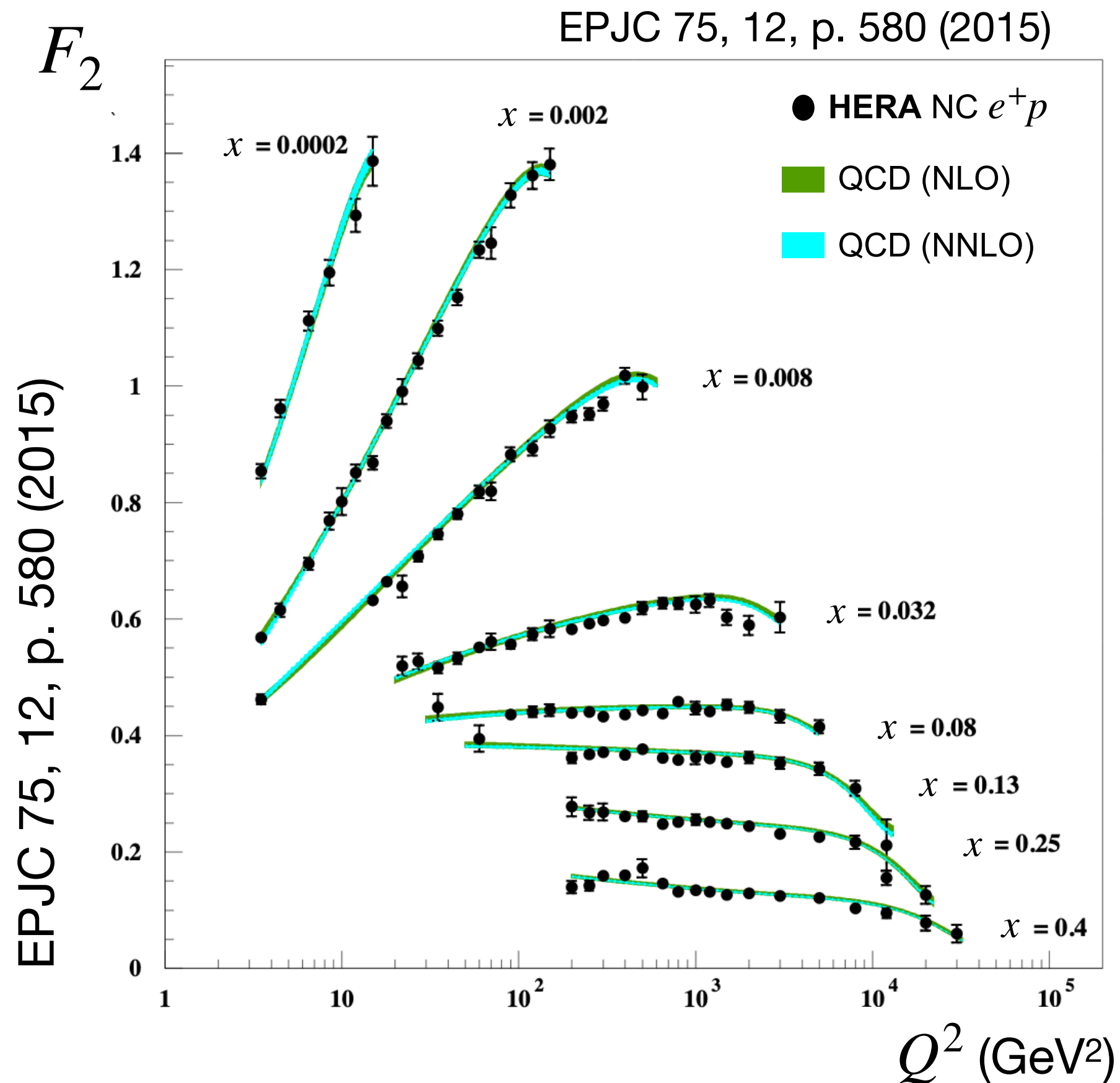
Dominant term at LO,

Sensitive to the sum of quark and antiquark distributions,
distinguishes between γ^* , Z exchange and γZ interference

$$[F_2, F_2^{\gamma Z}, F_2^Z] = x \sum_q [e_q^2, 2e_q v_q, v_q^2 + a_q^2] \cdot (q + \bar{q})$$

quark charge
vector coupling
axial coupling

DIS CROSS SECTIONS AND PDFs: NEUTRAL CURRENT



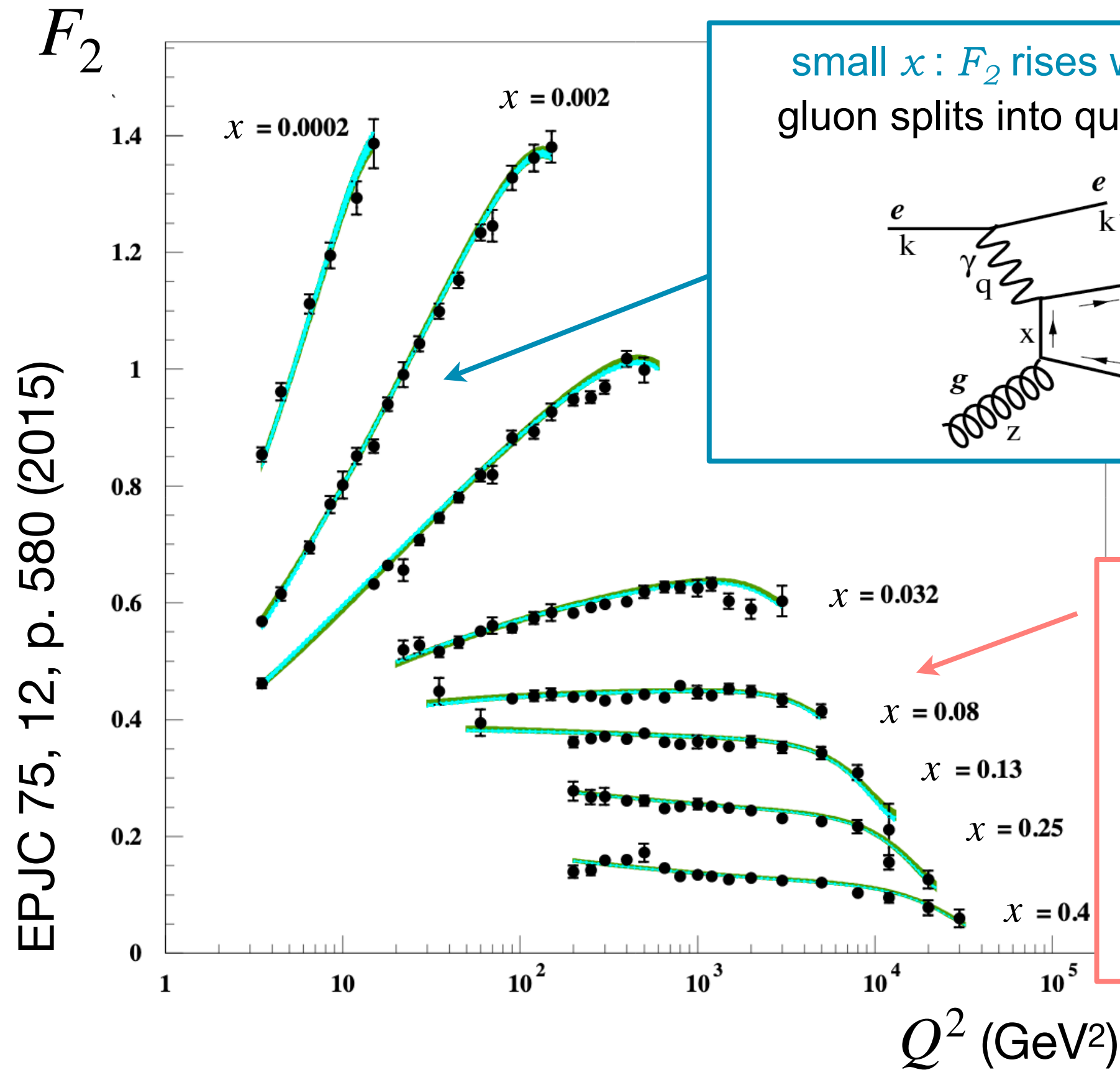
$$\frac{d^2\sigma_{NC}(e^\pm p)}{dx dQ^2} = \frac{2\pi\alpha^2}{x} \left[\frac{1}{Q^2} \right]^2 \phi_{NC}^\pm(x, Q^2)$$

$$Y_+ \widetilde{F}_2^\pm(x, Q^2) \mp Y_- x \widetilde{F}_3^\pm(x, Q^2) - y^2 \widetilde{F}_L^\pm(x, Q^2)$$

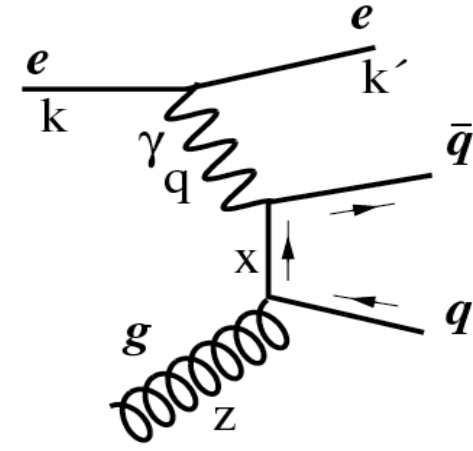
Scaling violations :

indirect probe of the gluon distribution

DIS CROSS SECTIONS AND PDFs: NEUTRAL CURRENT



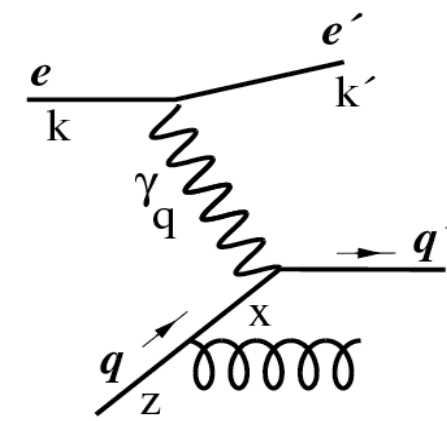
small x : F_2 rises with Q^2
gluon splits into quark pair



$$\frac{d^2\sigma_{NC}(e^\pm p)}{dx dQ^2} = \frac{2\pi\alpha^2}{x} \left[\frac{1}{Q^2} \right]^2 \phi_{NC}^\pm(x, Q^2)$$

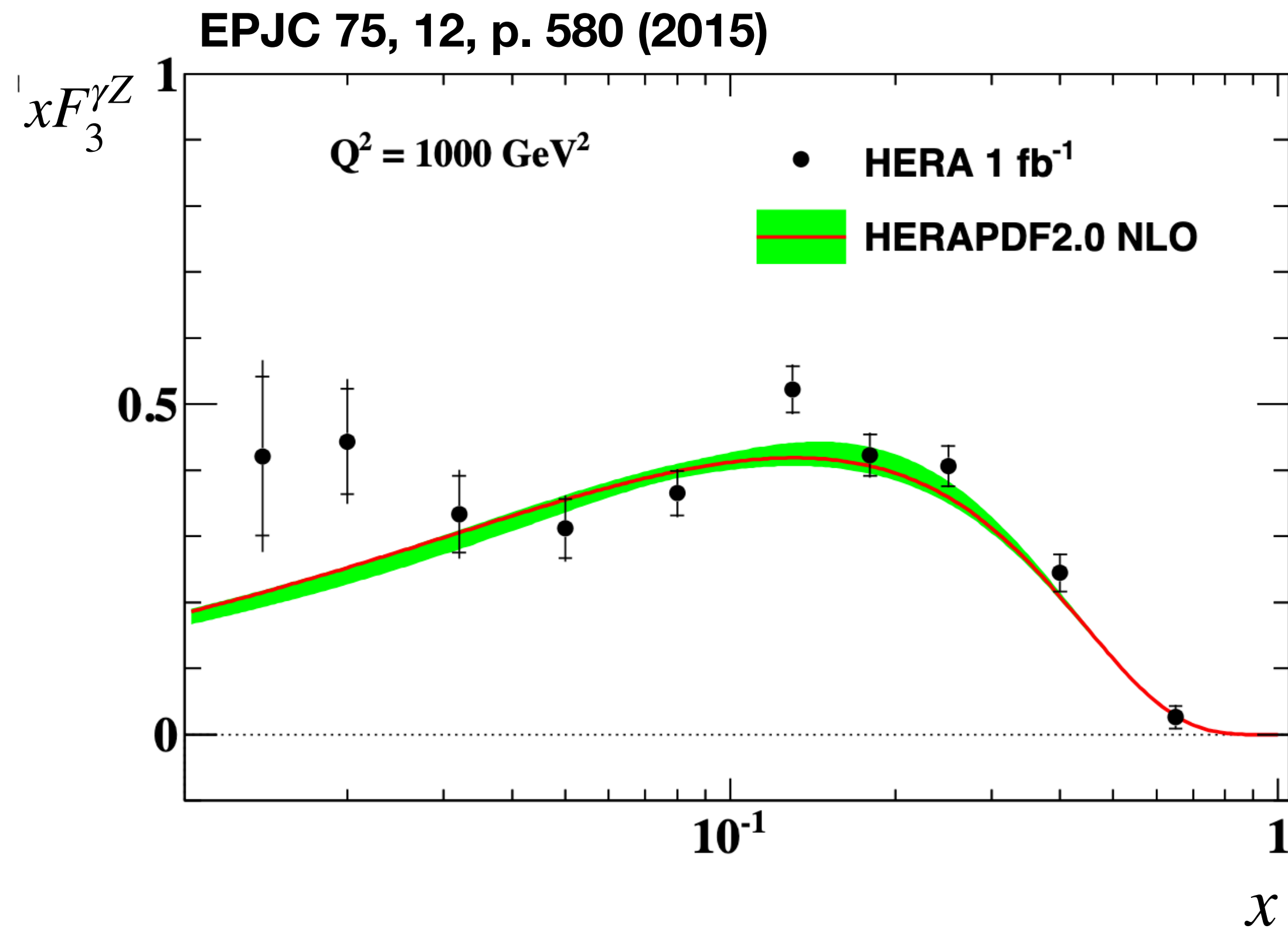
$$Y_+ \widetilde{F}_2^\pm(x, Q^2) \mp Y_- x \widetilde{F}_3^\pm(x, Q^2) - y^2 \widetilde{F}_L^\pm(x, Q^2)$$

large x : F_2 falls with Q^2
quarks radiate gluons



Scaling violations :
indirect probe of the gluon distribution

DIS CROSS SECTIONS AND PDFs: NEUTRAL CURRENT



HERA kinematics : $x F_3^{\gamma Z} \approx x/3(2u_v + d_v)$

$$\frac{d^2 \sigma_{NC}(e^\pm p)}{dx dQ^2} = \frac{2\pi\alpha^2}{x} \left[\frac{1}{Q^2} \right]^2 \phi_{NC}^\pm(x, Q^2)$$

$$Y_+ \widetilde{F}_2^\pm(x, Q^2) \mp Y_- x \widetilde{F}_3^\pm(x, Q^2) - y^2 \widetilde{F}_L^\pm(x, Q^2)$$

Parity-violating term, unique to weak Z-boson exchange

$$[x F_3^{\gamma Z}, x F_3^Z] = x \sum_q [2e_q a_q, 2v_q a_q] \cdot (q - \bar{q})$$

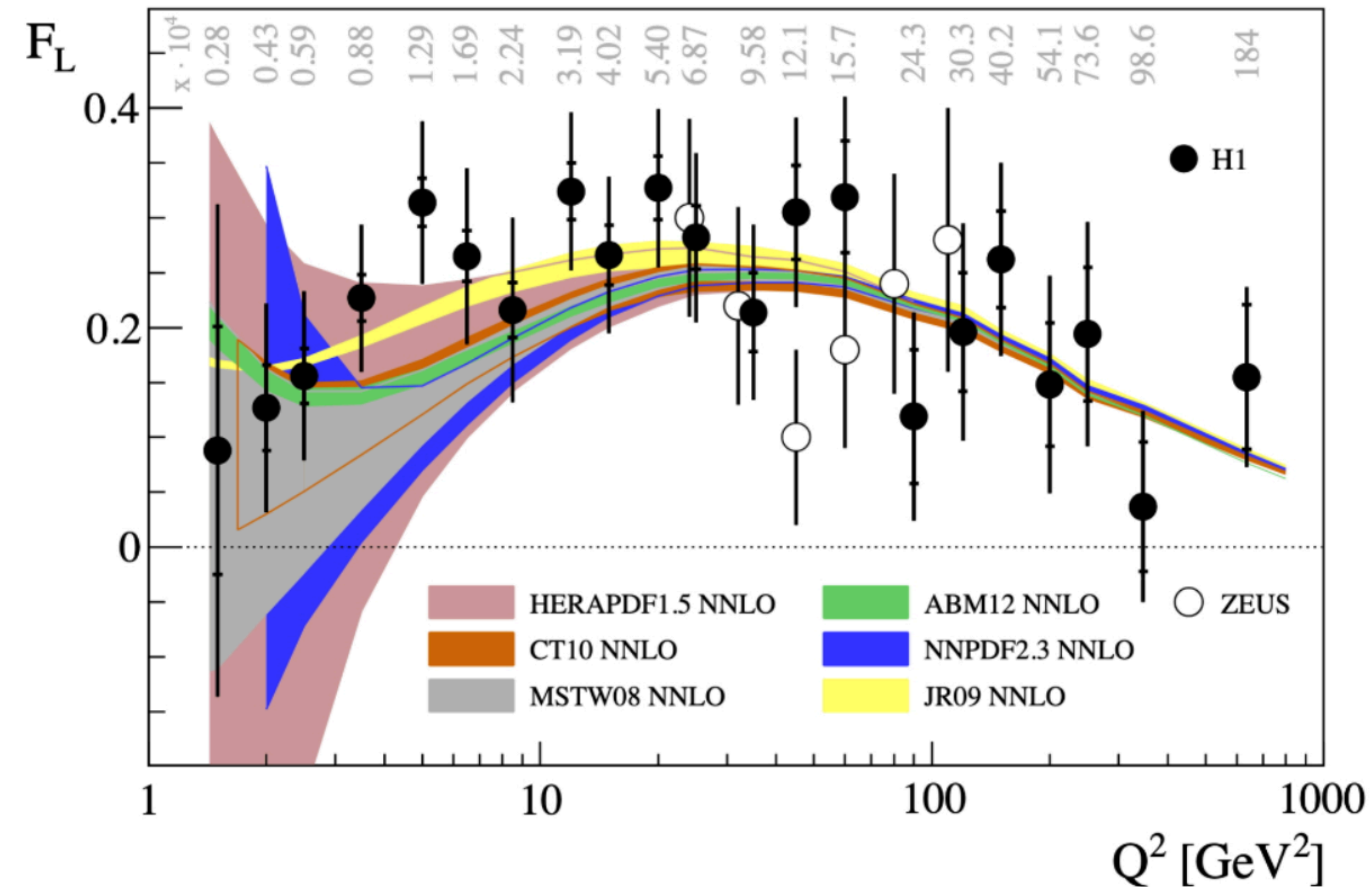
quark charge

vector coupling

axial coupling

DIS CROSS SECTIONS AND PDFs: NEUTRAL CURRENT

EPJC 74, 4, p. 2814 (2014)



$$\frac{d^2\sigma_{NC}(e^\pm p)}{dx dQ^2} = \frac{2\pi\alpha^2}{x} \left[\frac{1}{Q^2} \right]^2 \phi_{NC}^\pm(x, Q^2)$$

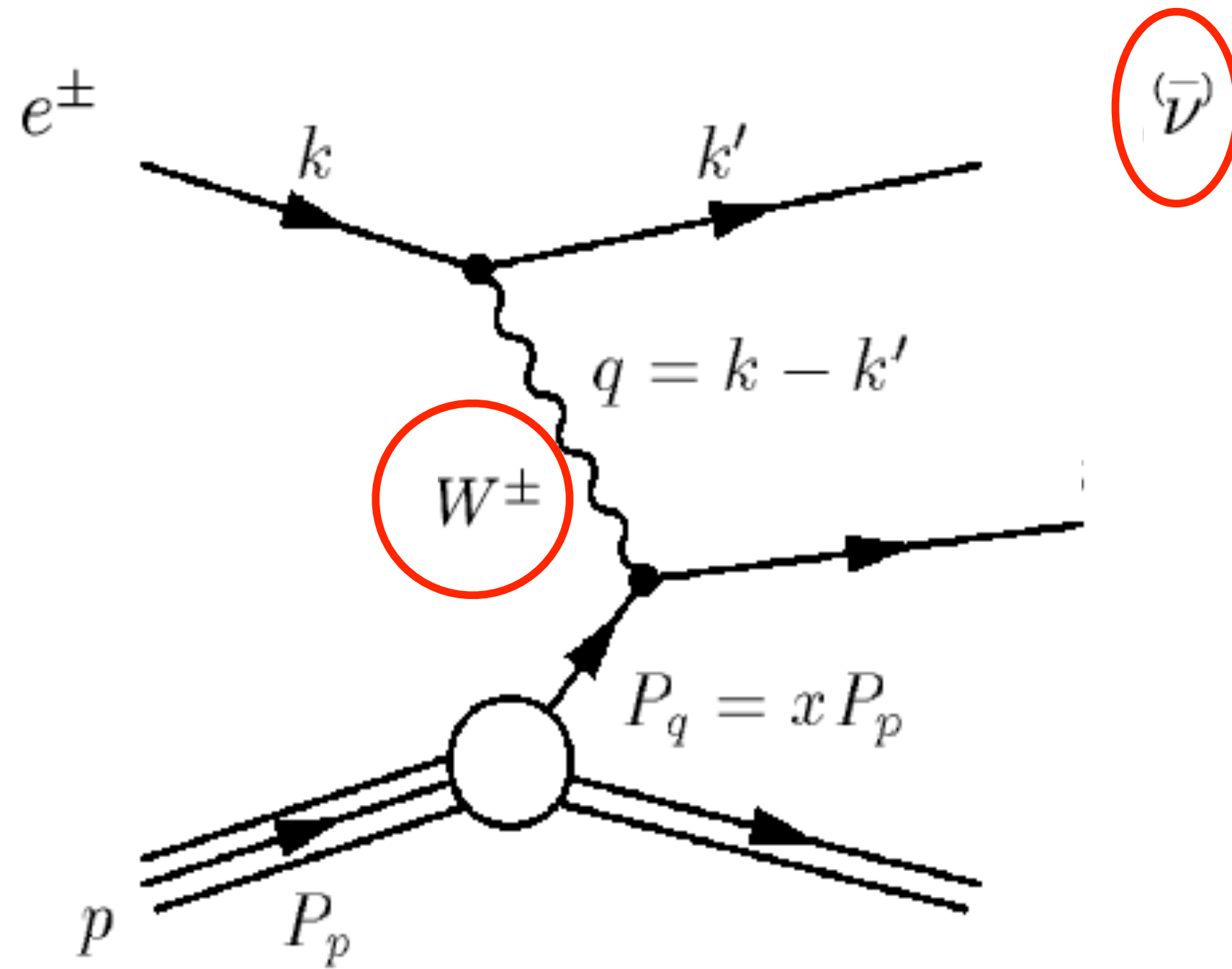
$$Y_+ \widetilde{F}_2^\pm(x, Q^2) \mp Y_- x \widetilde{F}_3^\pm(x, Q^2) - y^2 \widetilde{F}_L^\pm(x, Q^2)$$

Originates from interaction of longitudinally polarized γ^* with the parton content of the nucleon. In QMP $F_L = 0$.

In QCD, directly related to the gluon distribution:

$$F_L(x, Q^2) = \frac{\alpha_S}{4\pi} x^2 \int_x^1 \frac{dz}{z^3} \left[\frac{16}{3} F_2 + 8 \sum_q e_q^2 \left(1 - \frac{x}{z} \right) z g(z) \right]$$

DIS CROSS SECTIONS AND PDFs: CHARGED CURRENT



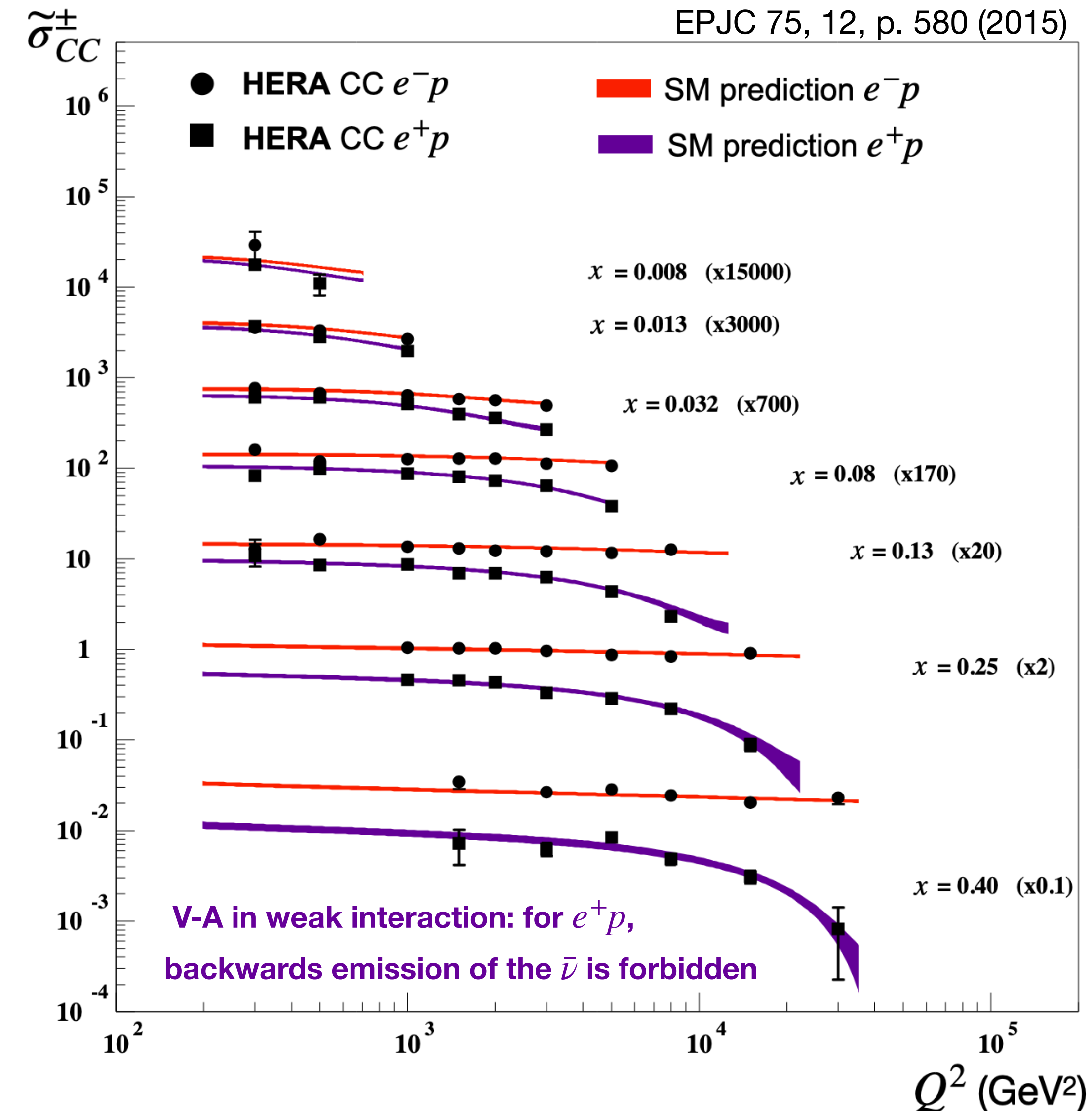
kinematics: $x, Q^2, y = \frac{p \cdot q}{p \cdot k},$

$$Y_{\pm} = 1 \pm (1 - y)^2$$

$$\frac{d^2\sigma_{CC}(e^\pm p)}{dx dQ^2} = \frac{G_F^2}{2\pi x} \left[\frac{m_W^2}{Q^2 + m_W^2} \right]^2 \phi_{CC}^\pm(x, Q^2)$$

$$Y_+ \widetilde{W}_2^\pm(x, Q^2) + y^2 \widetilde{W}_L^\pm(x, Q^2) \mp Y_- x \widetilde{W}_3^\pm(x, Q^2)$$

DIS CROSS SECTIONS AND PDFs: CHARGED CURRENT



$$\frac{d^2\sigma_{CC}(e^\pm p)}{dx dQ^2} = \frac{G_F^2}{2\pi x} \left[\frac{m_W^2}{Q^2 + m_W^2} \right]^2 \phi_{CC}^\pm(x, Q^2)$$

$$Y_+ \widetilde{W}_2^\pm(x, Q^2) + y^2 \widetilde{W}_L^\pm(x, Q^2) \mp Y_- x \widetilde{W}_3^\pm(x, Q^2)$$

Flavour separation in weak processes:

$$e^-p \rightarrow \nu X \begin{cases} W_2^- = 2x[u(x) + \bar{d}(x) + \bar{s}(x) + c(x)] \\ xW_3^- = 2x[u(x) + c(x) - \bar{d}(x) - \bar{s}(x)] \end{cases}$$

W^- probes more abundant u valence quarks

$$e^+p \rightarrow \nu X \begin{cases} W_2^+ = 2x[d(x) + s(x) + \bar{u}(x) + \bar{c}(x)] \\ xW_3^+ = 2x[d(x) + s(x) - \bar{u}(x) - \bar{c}(x)] \end{cases}$$

W^+ probes d valence quarks

(INCLUSIVE) DIS CROSS SECTIONS AND PDFs

Neutral Current $e^\pm p$

$$F_2(x, Q^2) \propto x \sum q_f(x, Q^2) + \bar{q}_f(x, Q^2) \quad \gamma^* \text{ exchange dominant: valence+sea; gluon indirectly via scaling violations}$$

$$F_3(x, Q^2) \propto x \sum^f q_f(x, Q^2) - \bar{q}_f(x, Q^2) \quad Z: \text{exchange at high } Q^2: \text{valence}$$

$$F_L(x, Q^2) \propto x \sum^f g(x, Q^2) \quad \text{direct access to the gluon distribution}$$

Charged Current $e^\pm p$

$$e^+p: \quad \frac{d^2\sigma^{CC}(e^+p)}{dx dQ^2} = \frac{G_F^2}{2\pi x} \left[\frac{m_W^2}{m_W^2 + Q^2} \right]^2 \{ \bar{u} + \bar{c} + (1-y)^2(d + s) \} \quad : \text{d-type quarks}$$

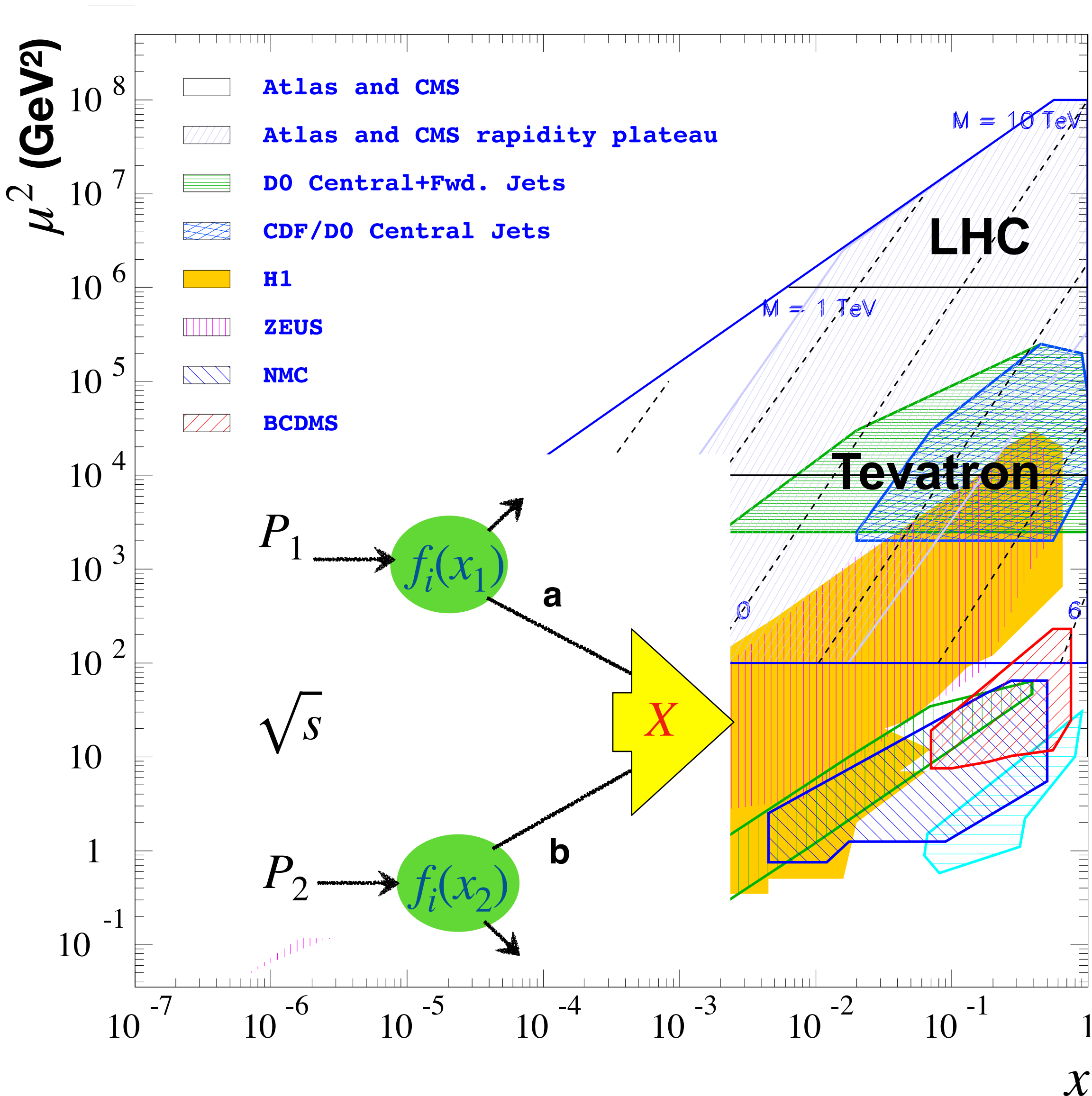
flavour separation

$$e^-p: \quad \frac{d^2\sigma^{CC}(e^-p)}{dx dQ^2} = \frac{G_F^2}{2\pi x} \left[\frac{m_W^2}{m_W^2 + Q^2} \right]^2 \{ u + c + (1-y)^2(\bar{d} + \bar{s}) \} \quad : \text{u-type quarks}$$

Deficits of collider DIS:

- quark flavour decomposition less constrained at medium-high x ,
- d -valence less precise than u -valence, gluon at high x not well constrained,
- direct probes of the gluon statistically limited

TEVATRON AND LHC IN PDF FITS



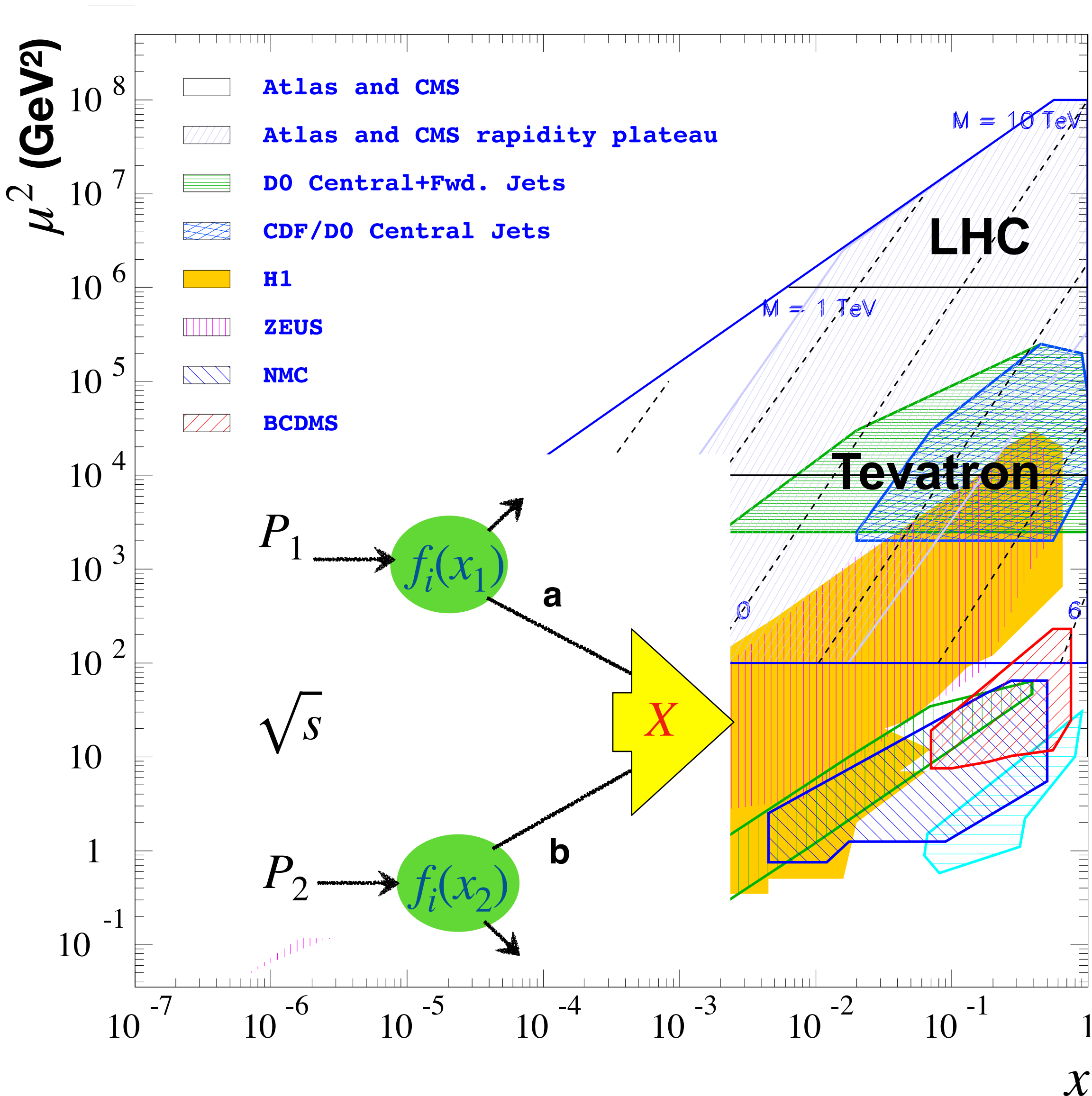
partonic center-of mass energy $\hat{s} = x_1 x_2 s$

energy scale for the process can be e.g.
a mass of produced final state, M_X

$$\tau = x_1 x_2 = M_X^2/s$$

Note: cannot disentangle x_1 and x_2 - PDFs can not be extracted by using $pp/p\bar{p}$ data alone

TEVATRON AND LHC IN PDF FITS



partonic center-of mass energy $\hat{s} = x_1 x_2 s$

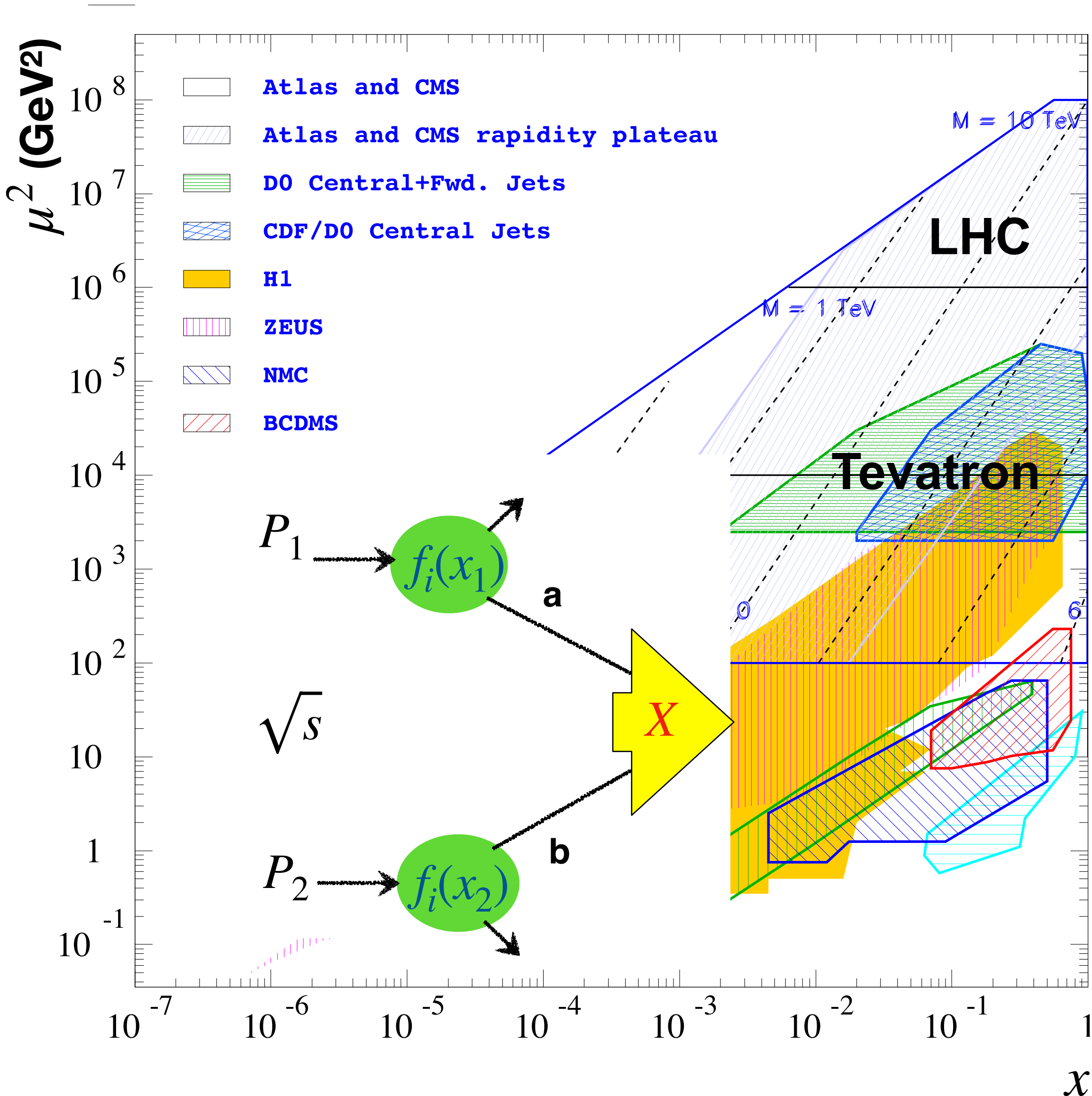
energy scale for the process can be e.g.
a mass of produced final state, M_X

$$\tau = x_1 x_2 = M_X^2/s$$

Partonic cross section: $\hat{\sigma}_{12 \rightarrow X} = C_X \delta(\hat{s} - M_X^2)$

$$\hat{\sigma}^n = c_0 \alpha_S^k \cdot \{1 + c'_1 \alpha_S + c'_2 \alpha_S^2 + \dots\} = c_0 \alpha_S^k + c_1 \alpha_S^{k+1} + c_2 \alpha_S^{k+2} + \dots$$

TEVATRON AND LHC IN PDF FITS



partonic center-of mass energy $\hat{s} = x_a x_b s$

energy scale for the process can be e.g.
a mass of produced final state, M_X

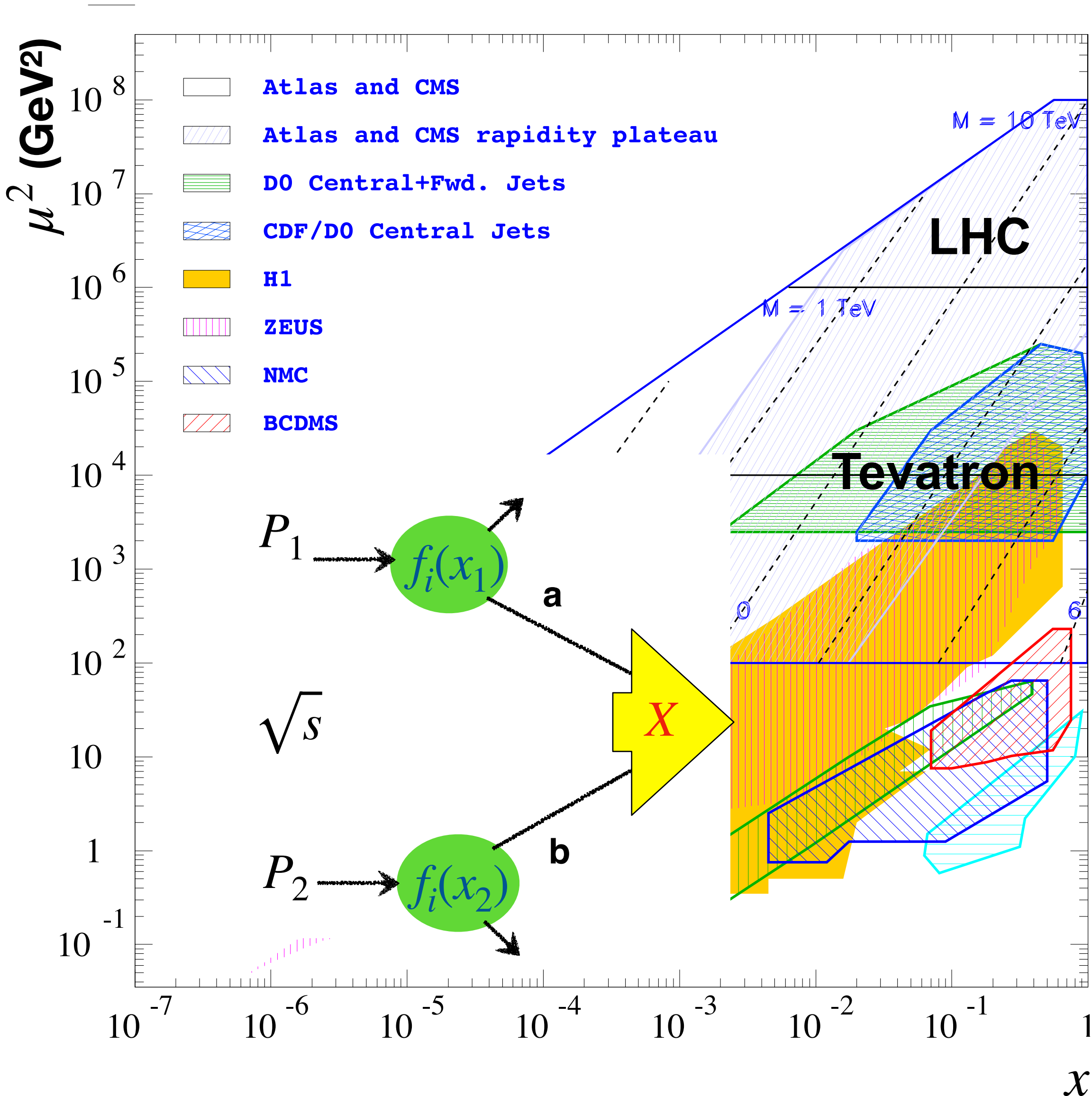
$$\tau = x_a x_b = M_X^2/s$$

Partonic cross section: $\hat{\sigma}_{12 \rightarrow X} = C_X \delta(\hat{s} - M_X^2)$

Process cross section:

$$\sigma_X = \int_0^1 dx_1 \int_0^1 dx_2 f_1(x_1, M_X^2) f_1(x_2, M_X^2) C_X \delta(x_1 x_2 - \tau) \equiv C_X \left[\frac{1}{s} \frac{\partial L_{ab}}{\partial \tau} \right]$$

TEVATRON AND LHC IN PDF FITS



partonic center-of mass energy $\hat{s} = x_a x_b s$

energy scale for the process can be e.g.
a mass of produced final state, M_X

$$\tau = x_a x_b = M_X^2/s$$

Partonic cross section: $\hat{\sigma}_{12 \rightarrow X} = C_X \delta(\hat{s} - M_X^2)$

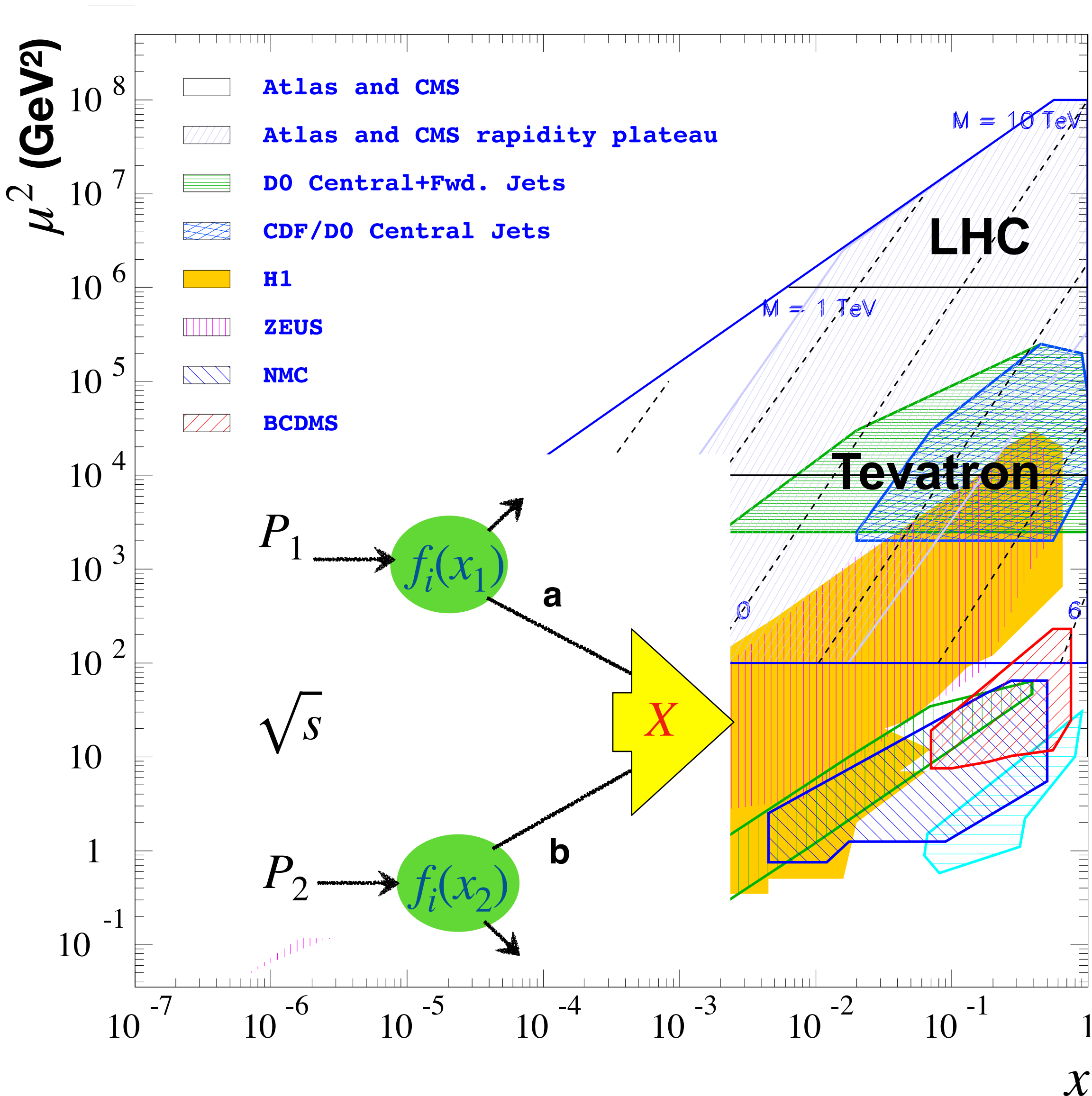
Process cross section:

$$\sigma_X = \int_0^1 dx_1 \int_0^1 dx_2 f_1(x_1, M_X^2) f_2(x_2, M_X^2) C_X \delta(x_1 x_2 - \tau) \equiv C_X \left[\frac{1}{s} \frac{\partial L_{ab}}{\partial \tau} \right]$$

Parton Luminosity function:

$$\frac{\partial L_{12}}{\partial \tau} = \int_0^1 \int_0^1 dx_1 dx_2 f_1(x_1, M_X^2) f_2(x_2, M_X^2) \delta(x_1 x_2 - \tau)$$

TEVATRON AND LHC IN PDF FITS



partonic center-of mass energy $\hat{s} = x_a x_b s$

energy scale for the process can be e.g.
a mass of produced final state, M_X

$$\tau = x_a x_b = M_X^2/s$$

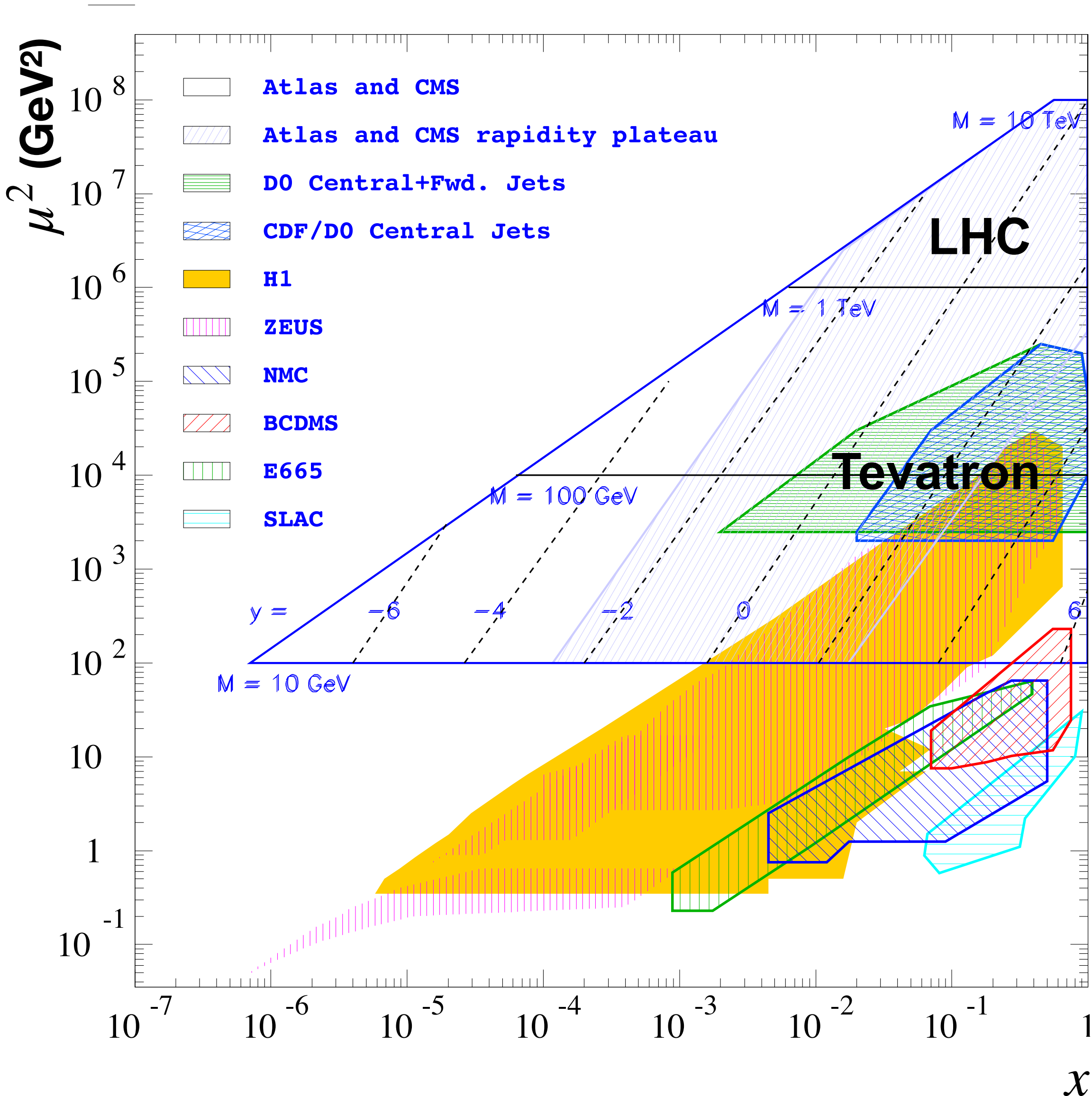
$$\hat{\sigma}_{ab \rightarrow X} = C_X \delta(\hat{s} - M_X^2)$$

Interpretation of the $pp/p\bar{p}$ data depends strongly on the PDFs

$$\underbrace{\sigma_{pp \rightarrow X}}_{\text{observed}} = \sum_{ij} \underbrace{f_i(x_1, \mu_F) \cdot f_j(x_2, \mu_F)}_{\text{proton structure}} \otimes \underbrace{\sigma_{12 \rightarrow X}\left(x_1, x_2, \alpha_S(\mu_R), \frac{M_X}{\mu_R}, \frac{M_X}{\mu_F}\right)}_{\text{partonic cross section}}$$

Uncertainties in both PDF enter multiplicatively !

IMPACT OF pp ($\bar{p}p$) ON PDFs



Data of Tevatron/LHC data have strong impact on the PDFs

[personal selection is based on illustration,
not on the date of the publication]

Most exhaustive investigations
of the LHC data in different global PDF fits:

ABMPtt: [arXiv:2407.00545](https://arxiv.org/abs/2407.00545)

MSHT20: [arXiv:2012.04684](https://arxiv.org/abs/2012.04684), [update](#)

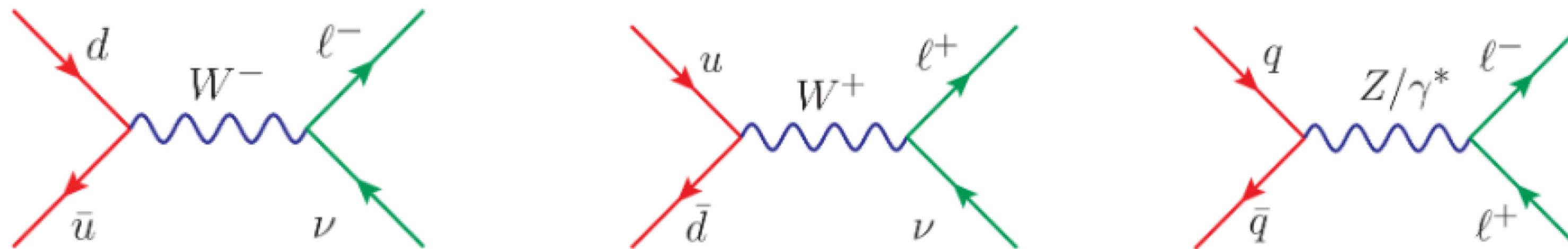
CT18: [arxiv:1912.10053](https://arxiv.org/abs/1912.10053), [update coming soon](#)

NNPDF40: [arXiv:2109.02653](https://arxiv.org/abs/2109.02653)

EWK BOSONS @ TEVATRON AND PDFs

DIS proton data mostly constrain the u-type PDF, while d - type PDFs are only constrained at high Q^2 , where precision is limited

Drell-Yan (W, Z) production in pp and $p\bar{p}$ collisions provides additional information on the d -quark



EWK BOSONS @ TEVATRON AND PDFs

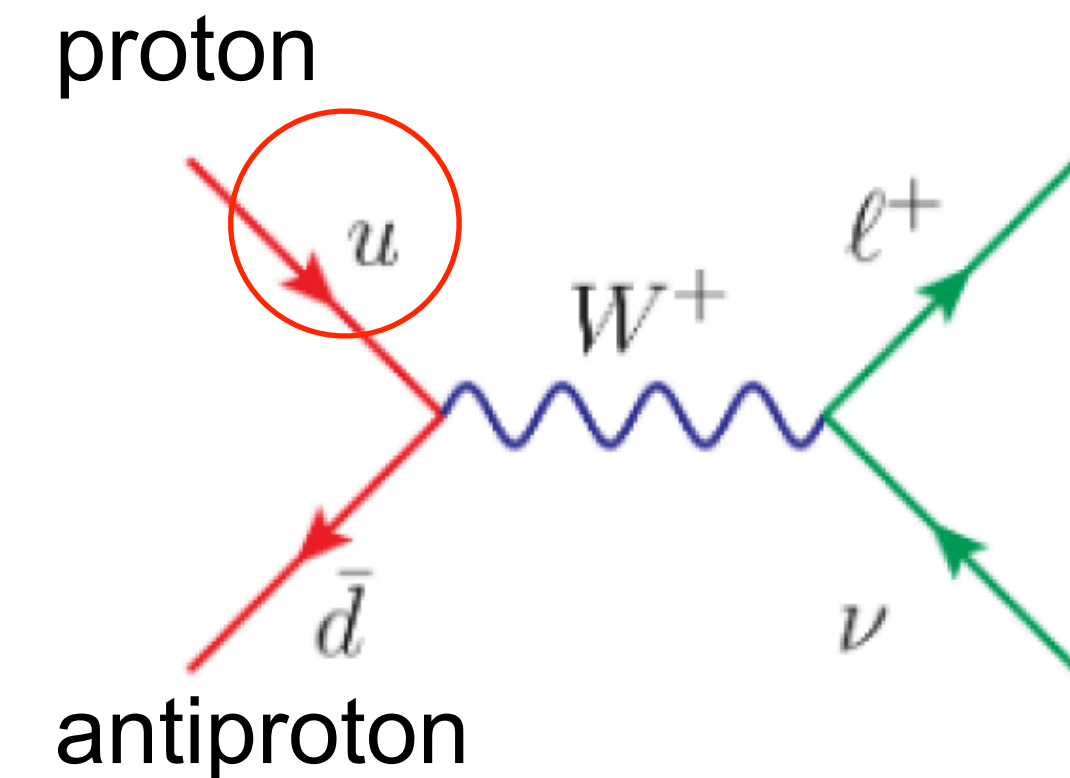
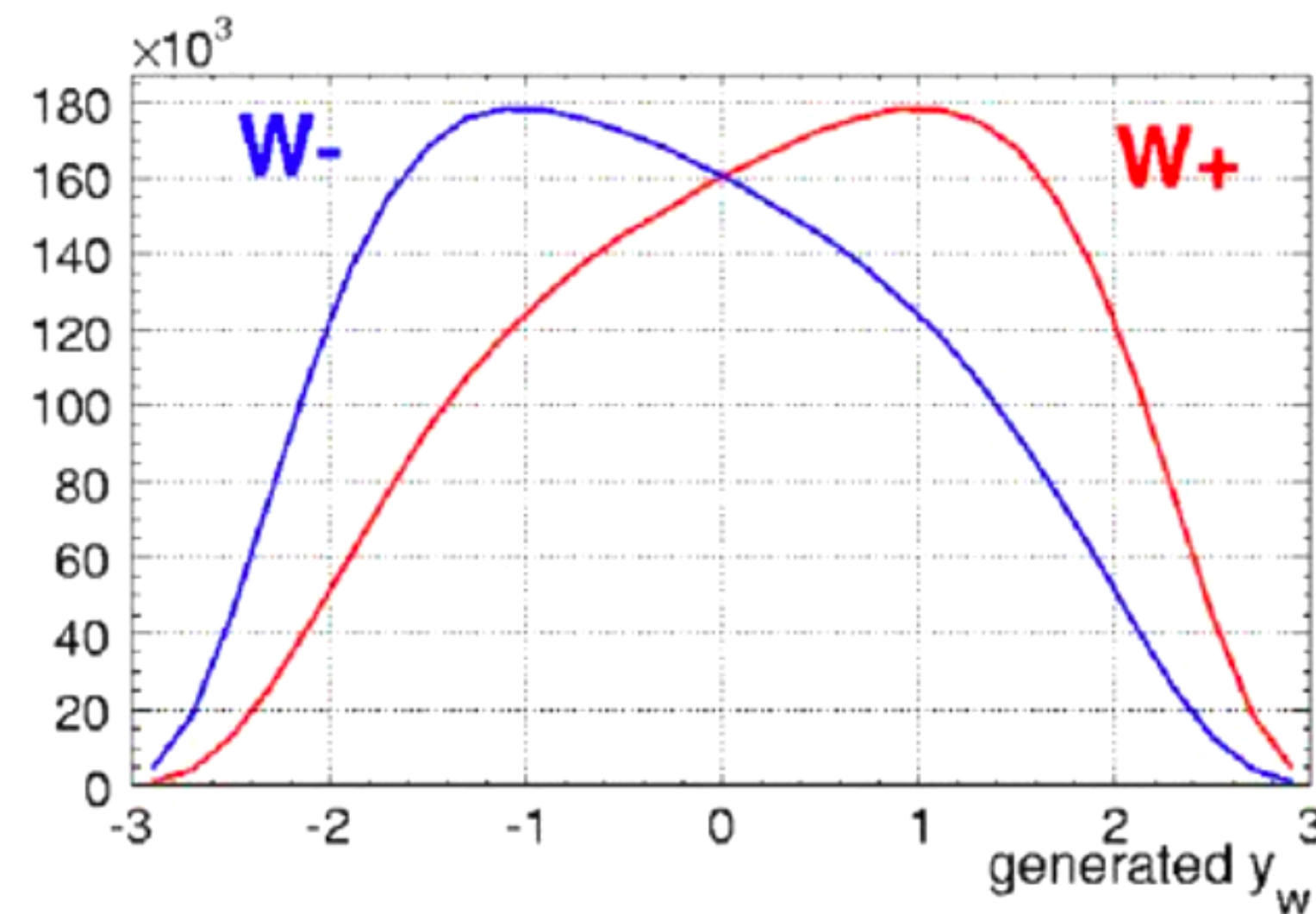
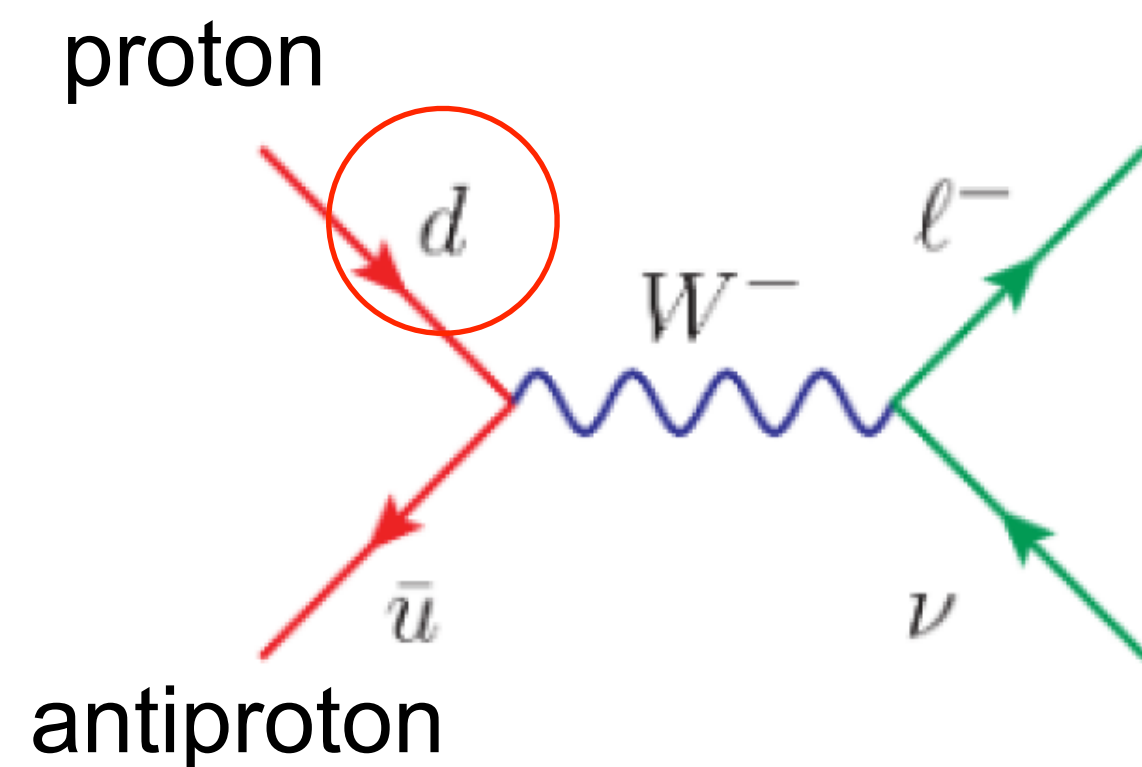
DIS proton data mostly constrain the u-type PDF, while d - type PDFs are only constrained at high Q^2 , where precision is limited

Drell-Yan (W, Z) production in pp and $p\bar{p}$ collisions provides additional information on the d -quark where HERA data is missing/imprecise

EWK BOSONS @ TEVATRON AND PDFs

DIS proton data mostly constrain the u -type PDF, while d - type PDFs are only constrained at high Q^2 , where precision is limited

In the proton, u quarks carry larger fraction of the proton's momentum than d quarks $\rightarrow$ asymmetry



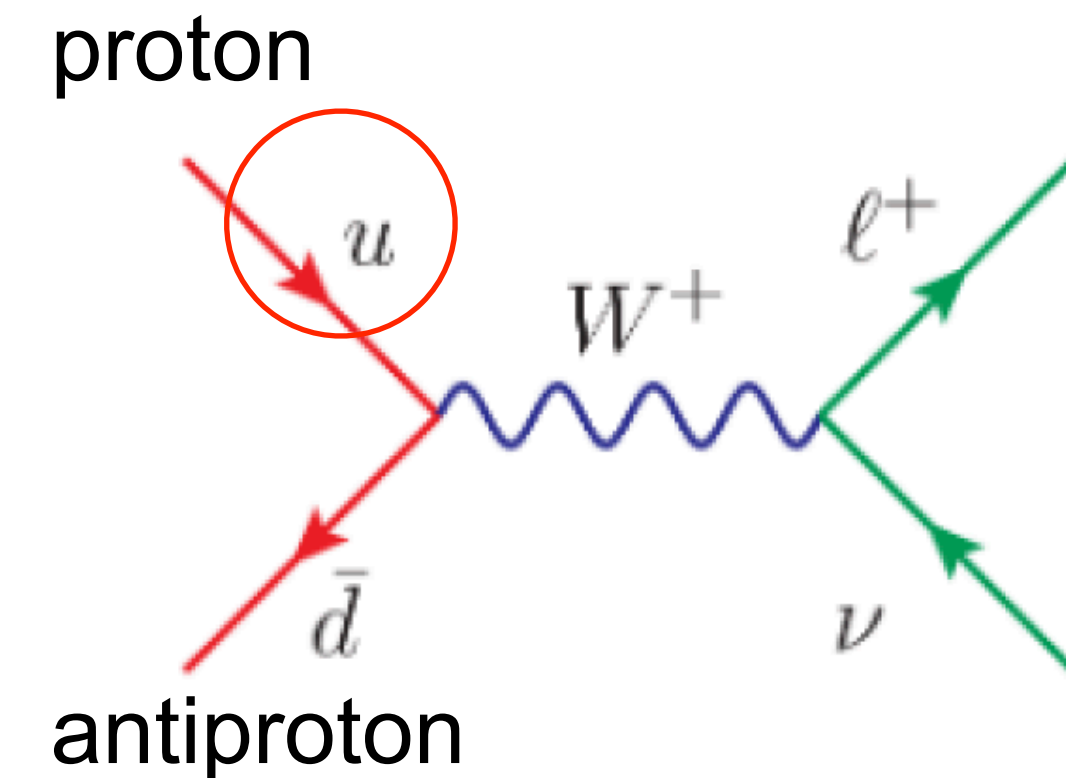
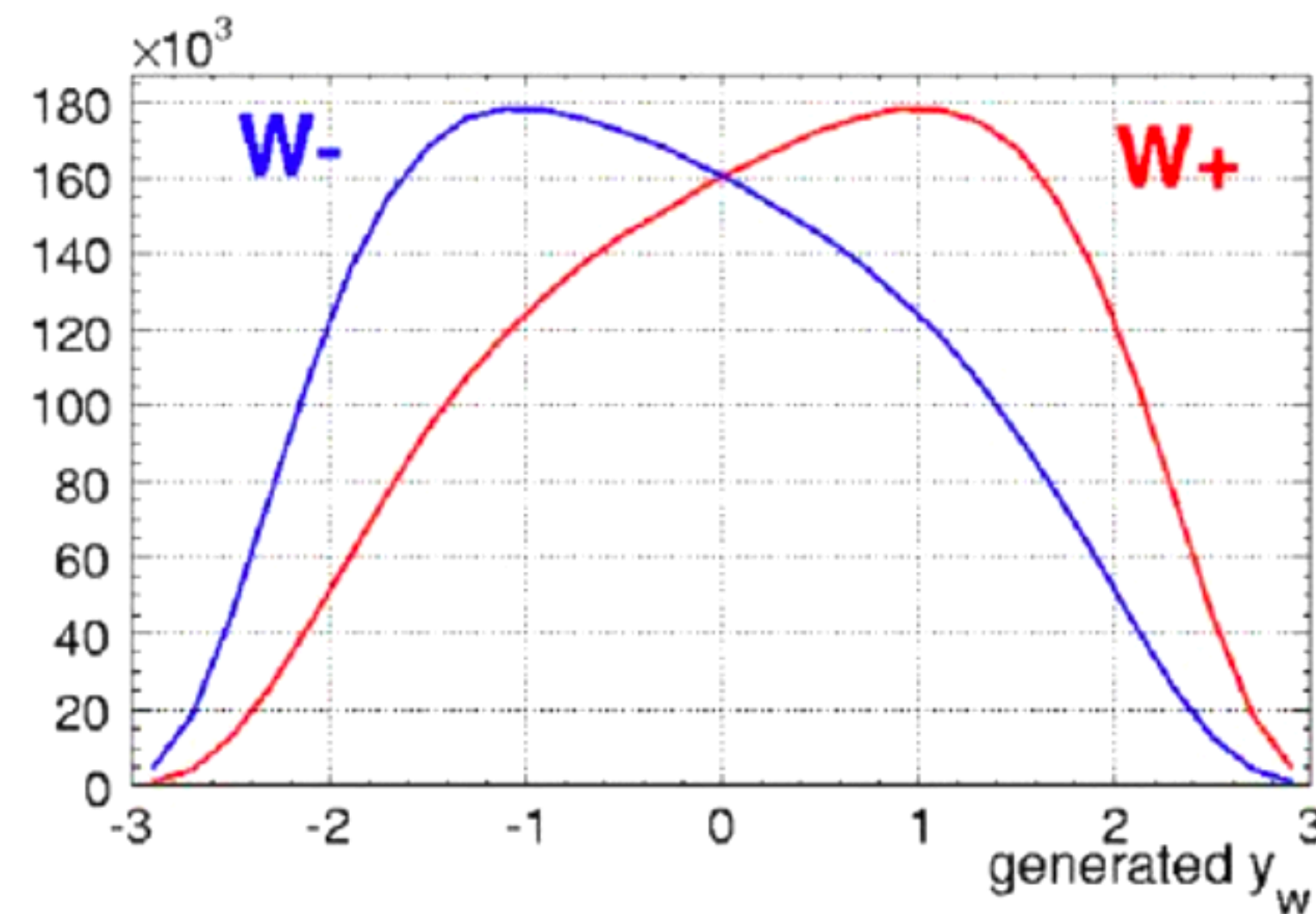
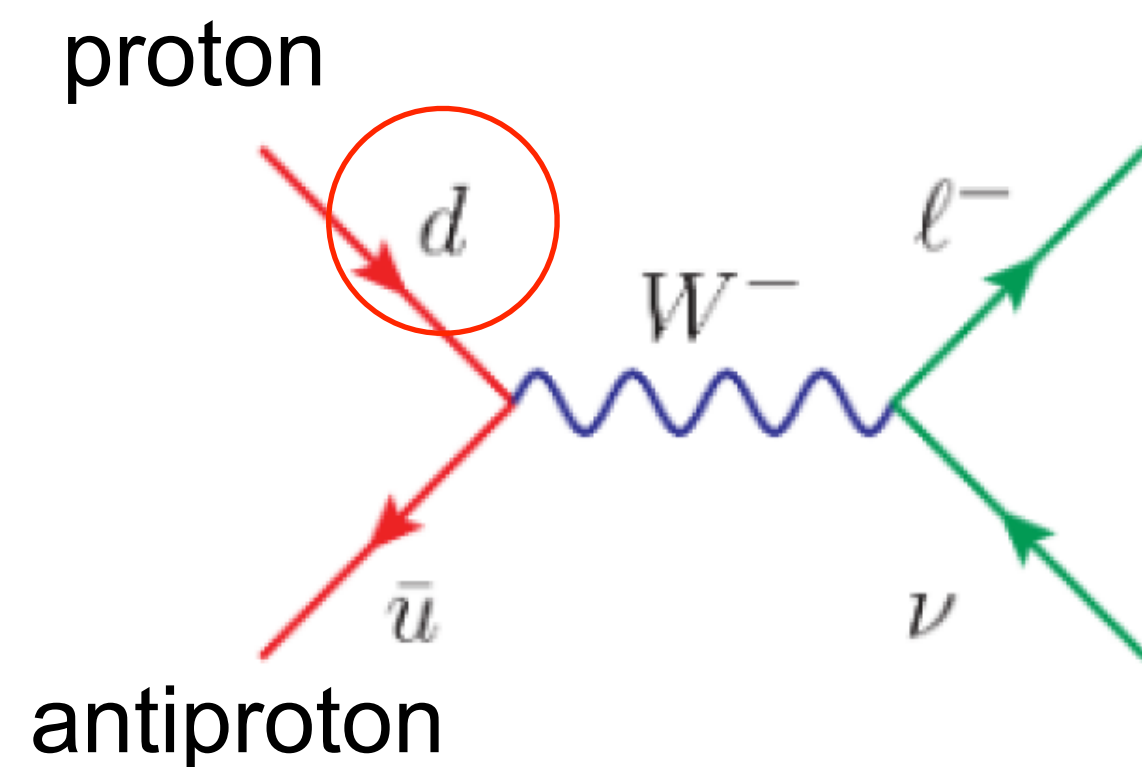
W^- boosted more along the $\bar{p}$ beam direction (more backward in rapidity)

W^+ boosted more along the p beam direction (more forward in rapidity)

EWK BOSONS @ TEVATRON AND PDFs

DIS proton data mostly constrain the u -type PDF, while d - type PDFs are only constrained at high Q^2 , where precision is limited

In the proton, u quarks carry larger fraction of the proton's momentum than d quarks $\rightarrow$ asymmetry

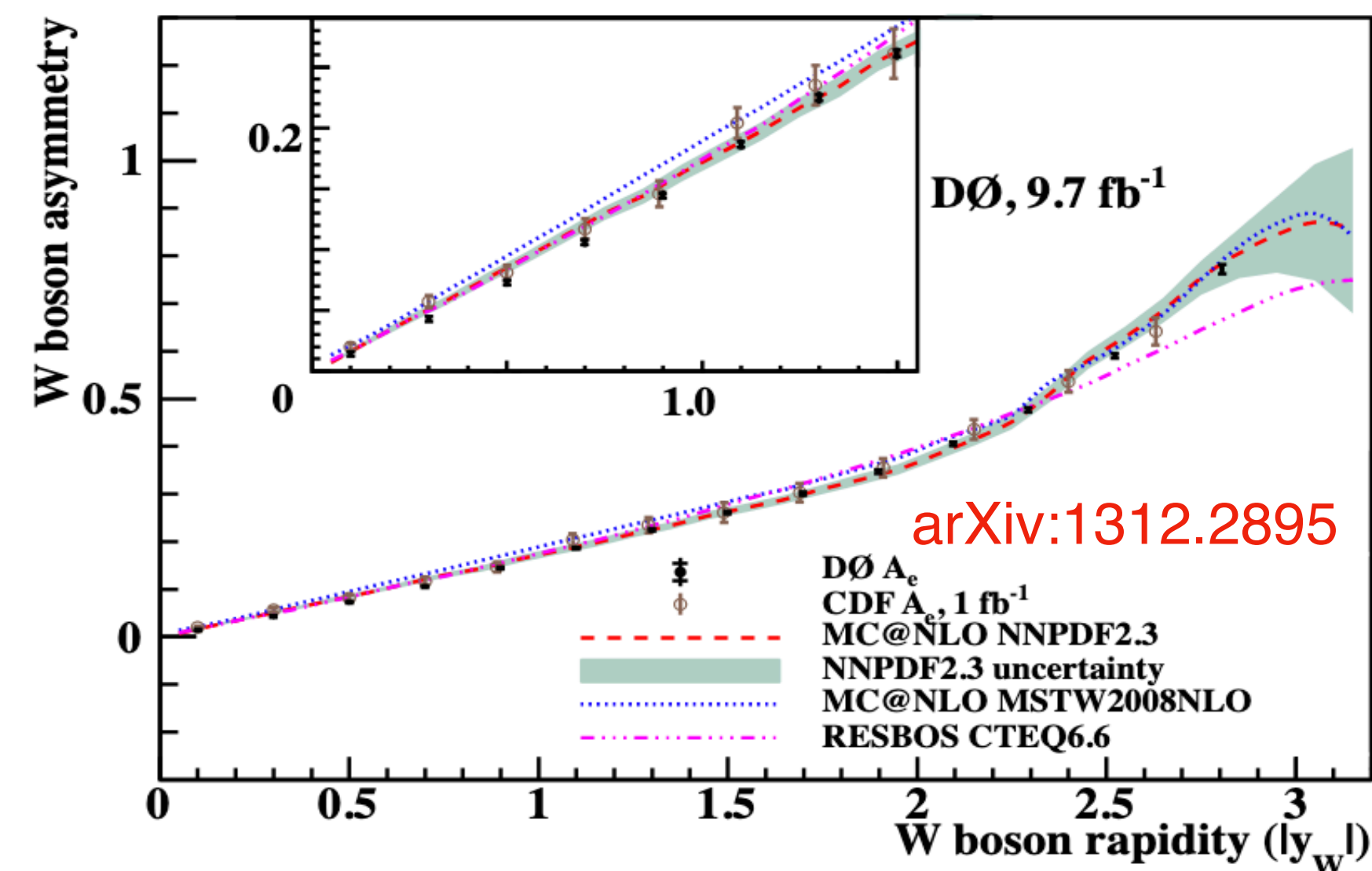
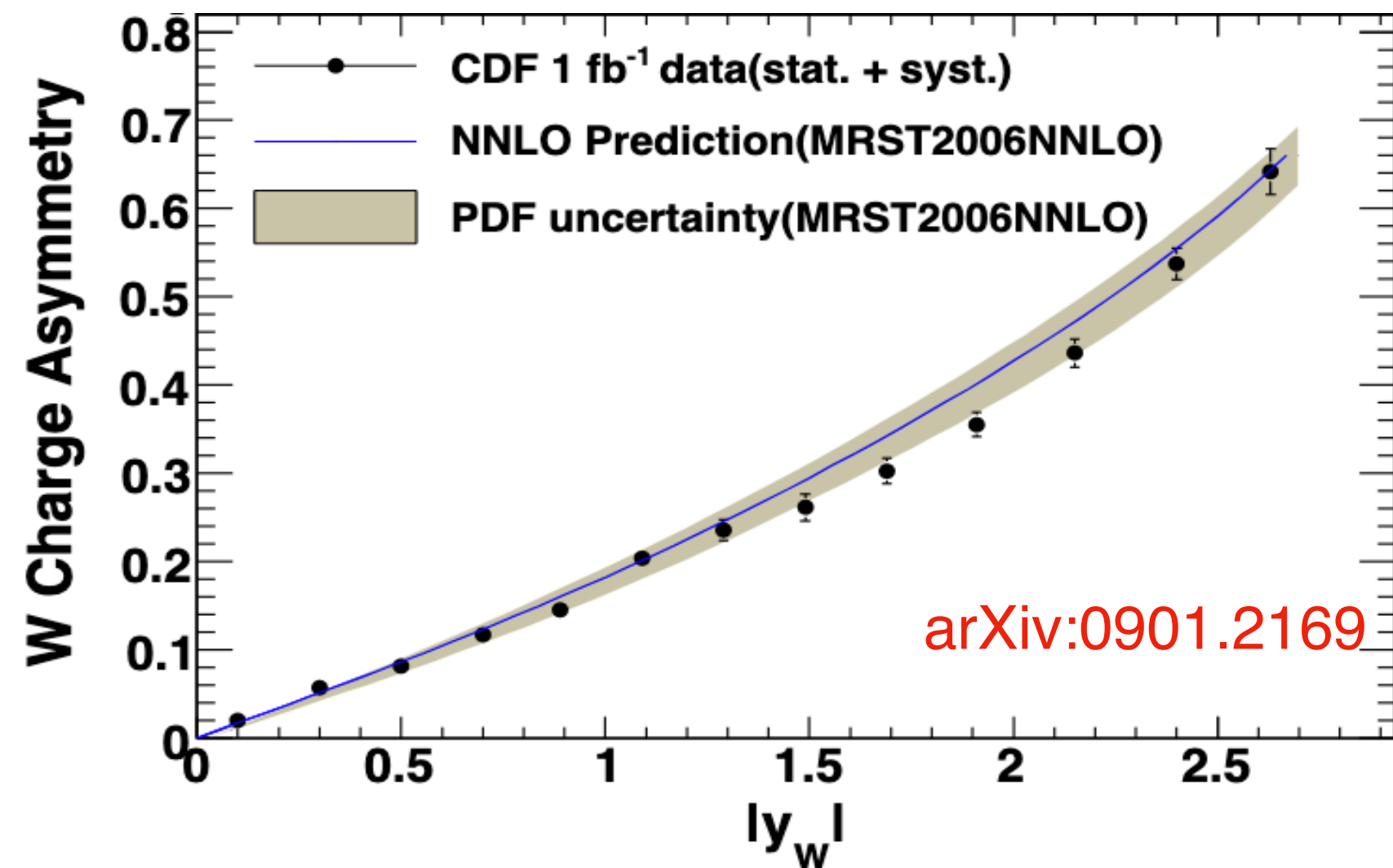


W charge asymmetry:
$$A(y_W) = \frac{d\sigma_+ - d\sigma_-}{d\sigma_+ + d\sigma_-} \approx \frac{u(x_+)d(x_-) - d(x_+)u(x_-)}{u(x_+)d(x_-) + d(x_+)u(x_-)}$$

$$x_{\pm} = \frac{m_{W^{\pm}}}{\sqrt{s}} \cdot \exp(\pm y_{W^{\pm}})$$

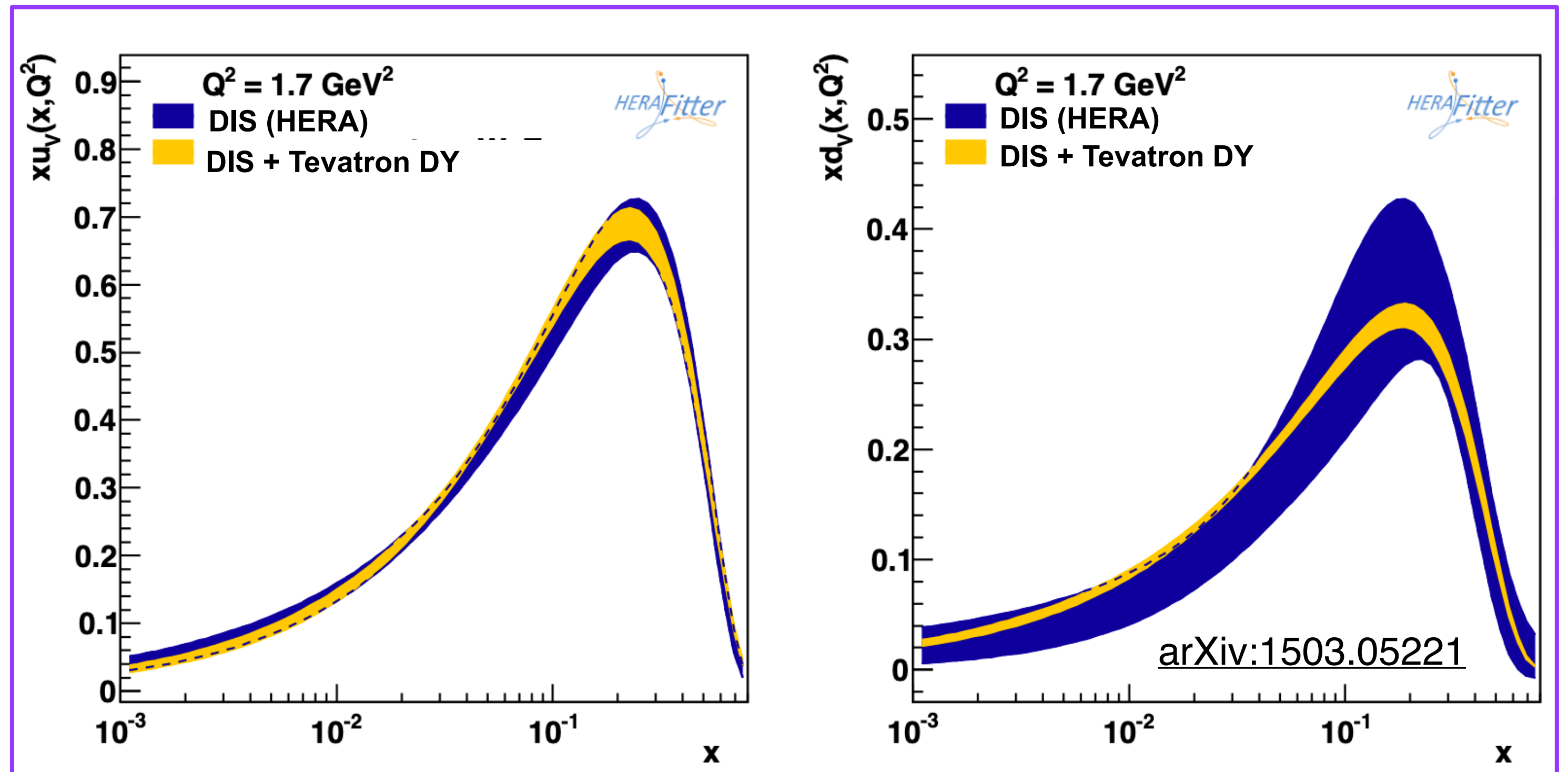
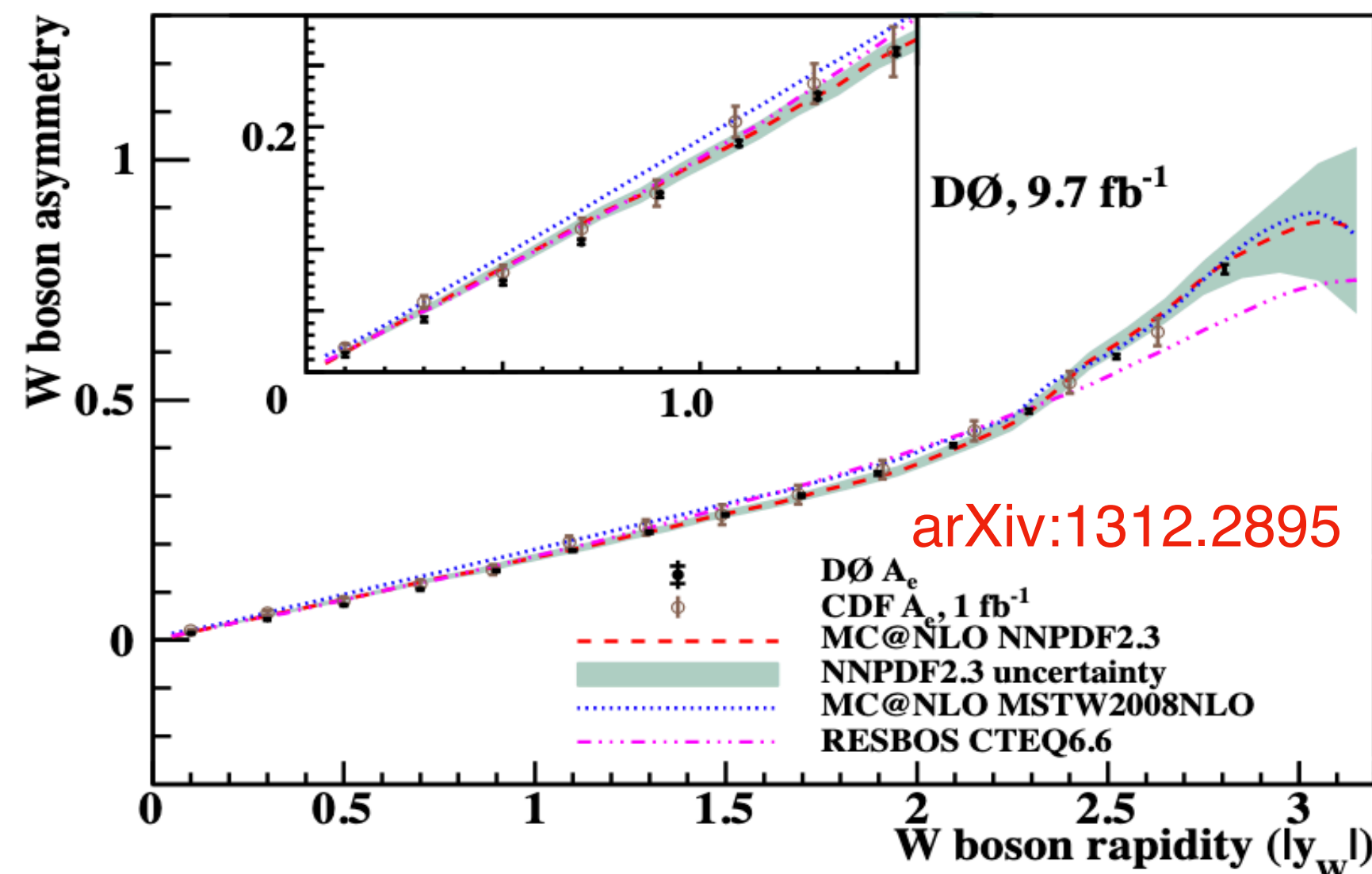
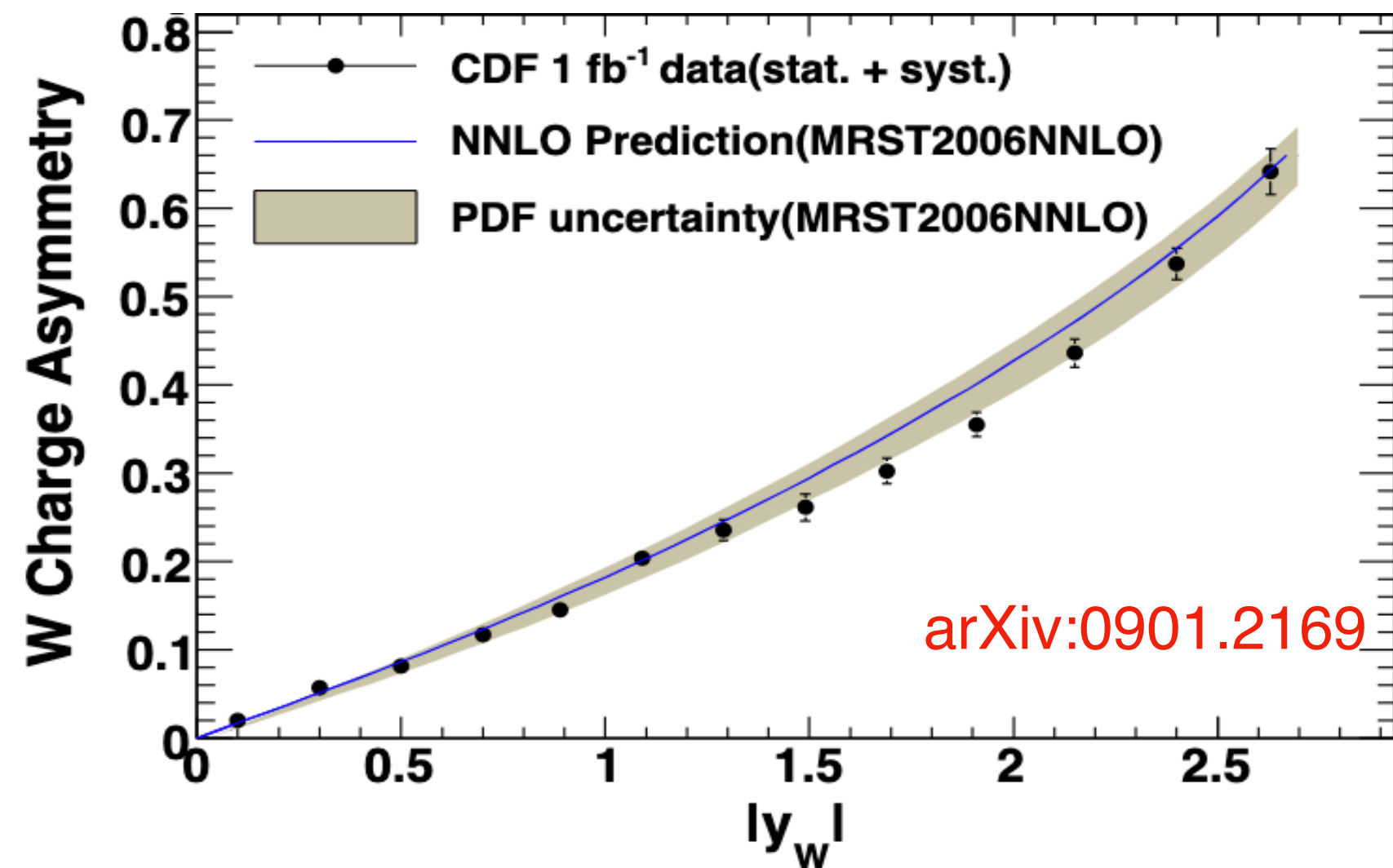
EWK BOSONS @ TEVATRON AND PDFs

Tevatron measurements of W asymmetry $A(y_W) = \frac{d\sigma_+ - d\sigma_-}{d\sigma_+ + d\sigma_-}$



EWK BOSONS @ TEVATRON AND PDFs

Tevatron measurements of W asymmetry $A(y_W) = \frac{d\sigma_+ - d\sigma_-}{d\sigma_+ + d\sigma_-}$



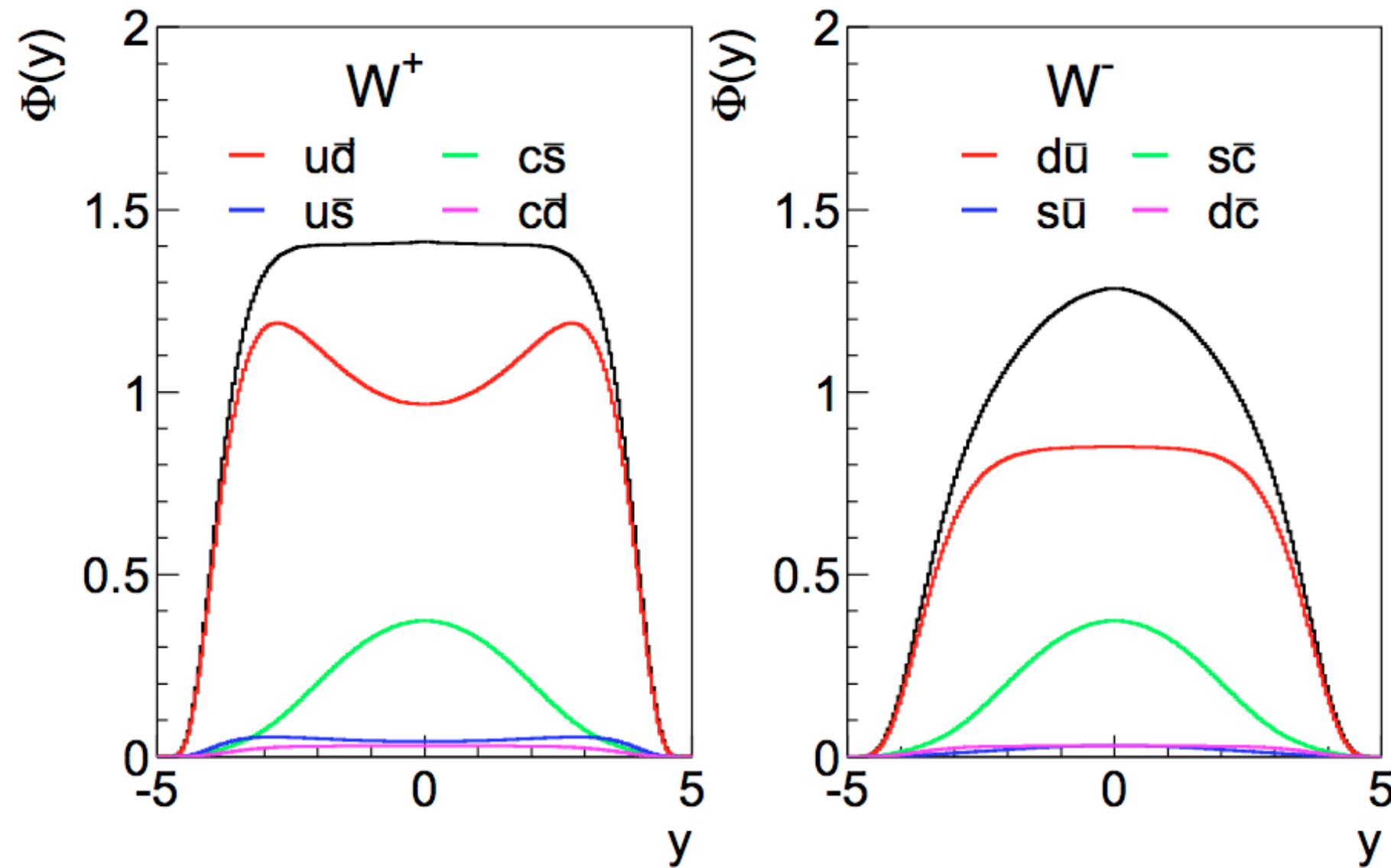
significant improvement in the d-valence

ELECTROWEAK BOSONS @ LHC AND PDFs

LO:

$$W^+ \approx 0.95(u\bar{d} + c\bar{s}) + 0.05(u\bar{s} + c\bar{d})$$

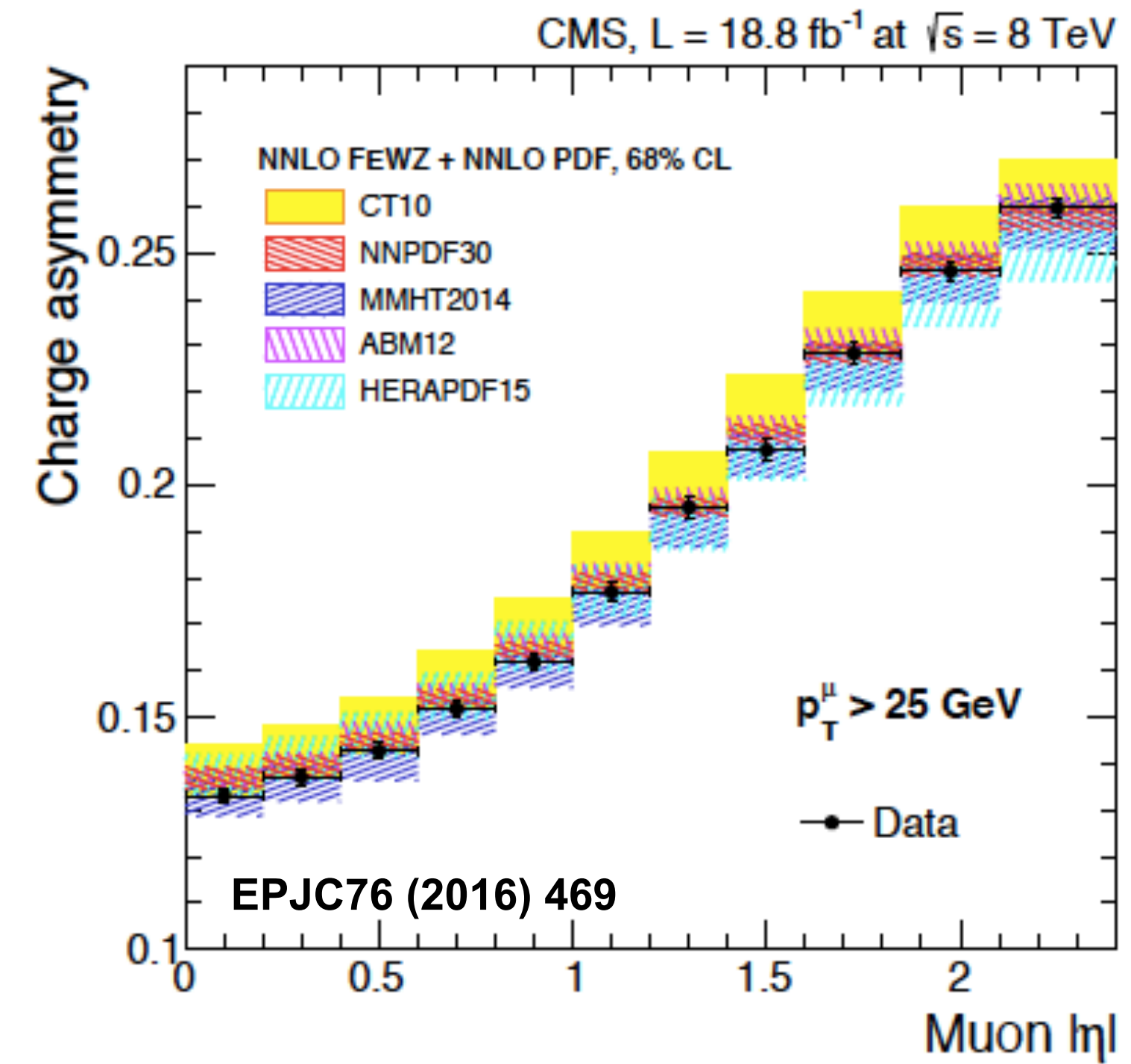
$$W^- \approx 0.95(d\bar{u} + s\bar{c}) + 0.05(d\bar{c} + s\bar{u})$$



W charge asymmetry: constrains valence

$$A_W = \frac{\sigma_{W^+} - \sigma_{W^-}}{\sigma_{W^+} + \sigma_{W^-}} \approx \frac{u_v - d_v}{u_v + d_v + 2u_{sea}}$$

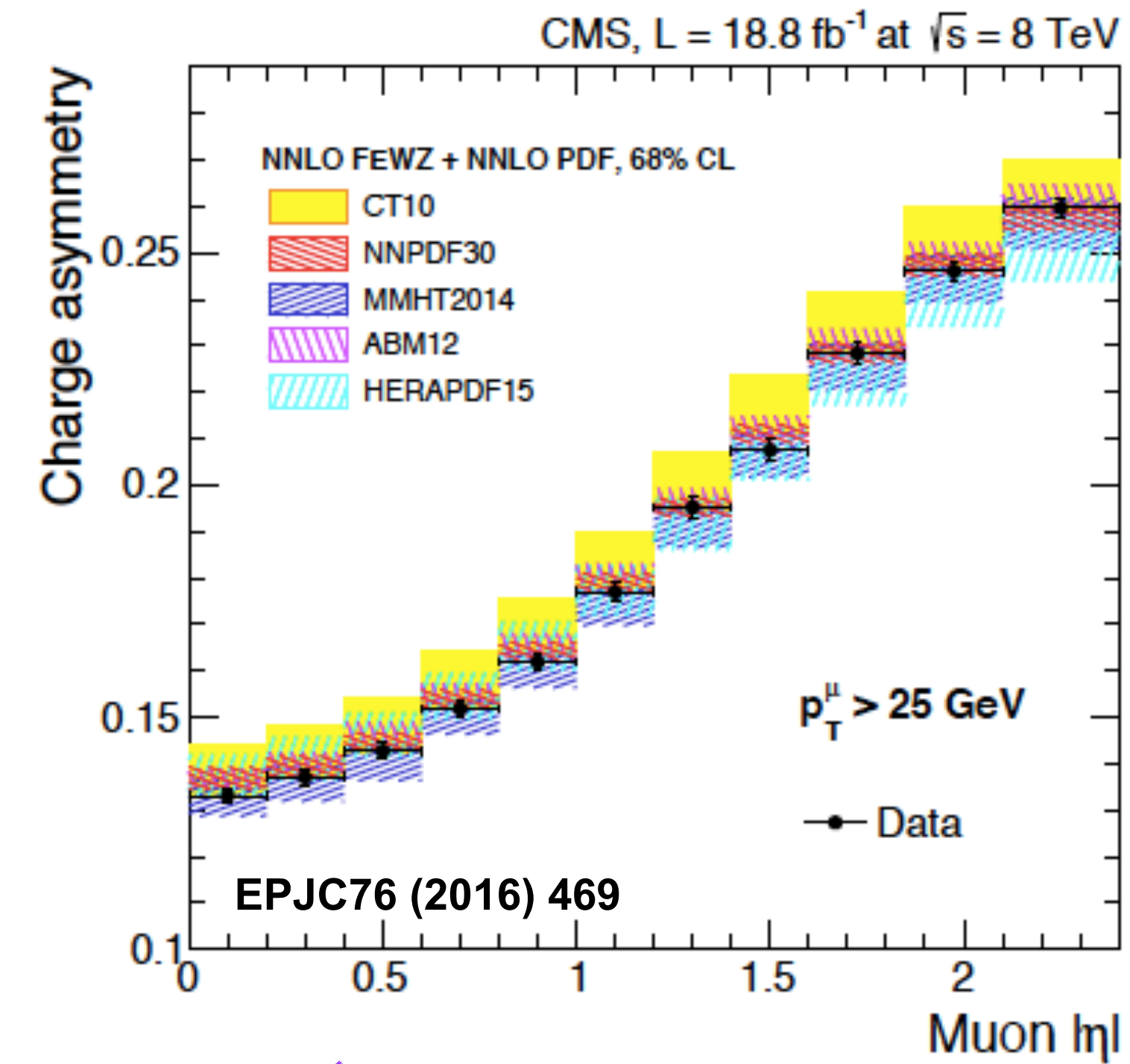
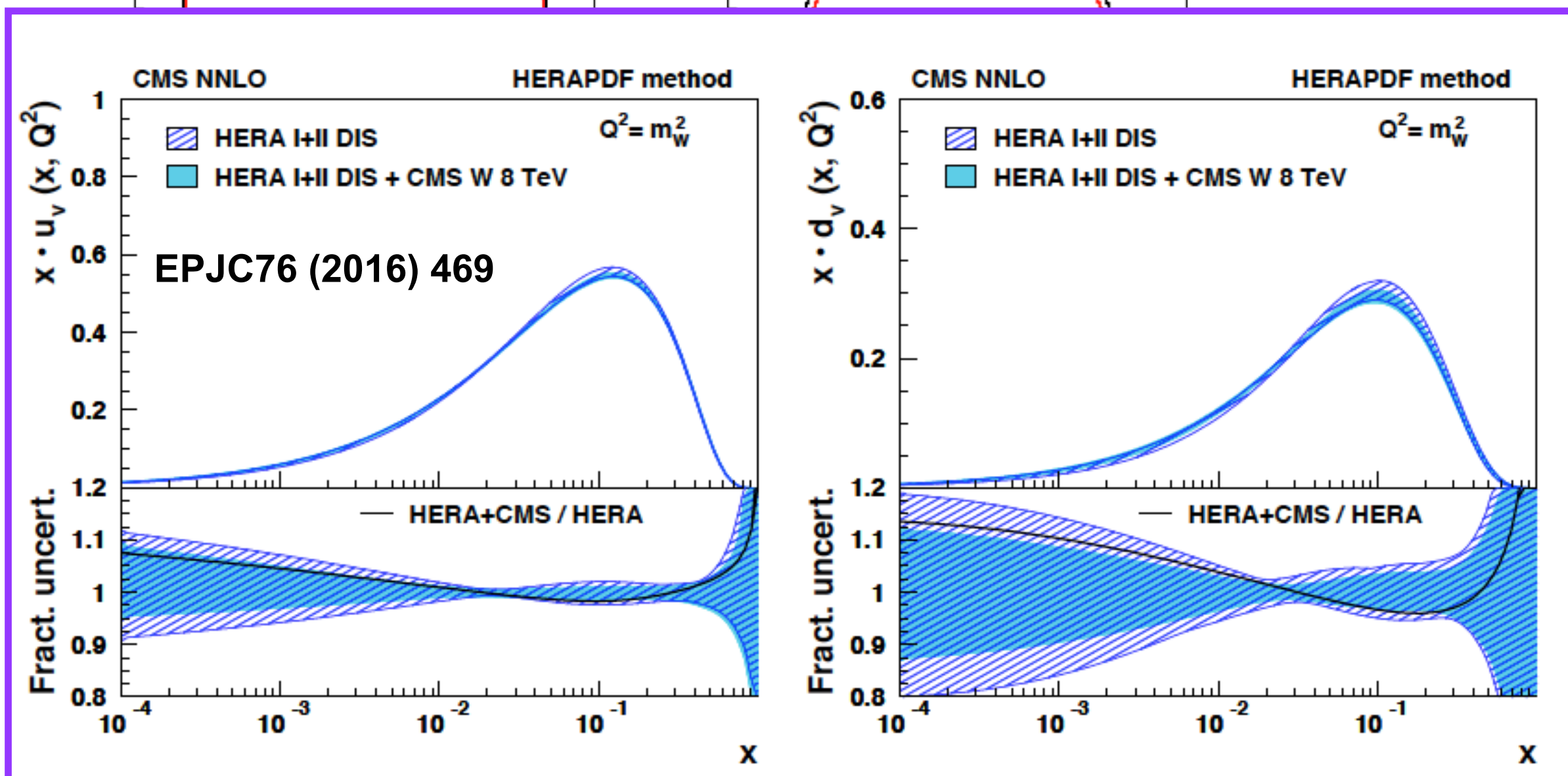
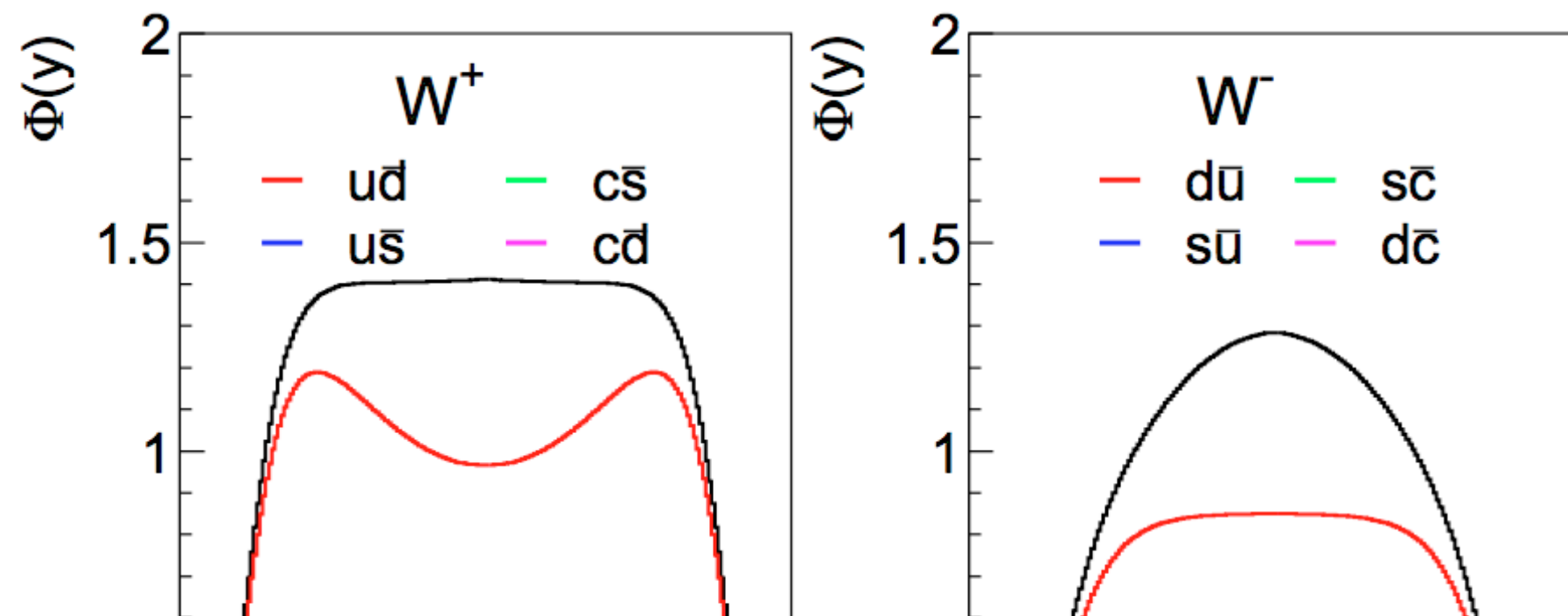
Example: measurement at $\sqrt{s} = 13$ TeV



ELECTROWEAK BOSONS @ LHC AND PDFs

LO: $W^+ \approx 0.95(u\bar{d} + c\bar{s}) + 0.05(u\bar{s} + c\bar{d})$
 $W^- \approx 0.95(d\bar{u} + s\bar{c}) + 0.05(d\bar{c} + s\bar{u})$

Example: measurement at $\sqrt{s} = 13$ TeV



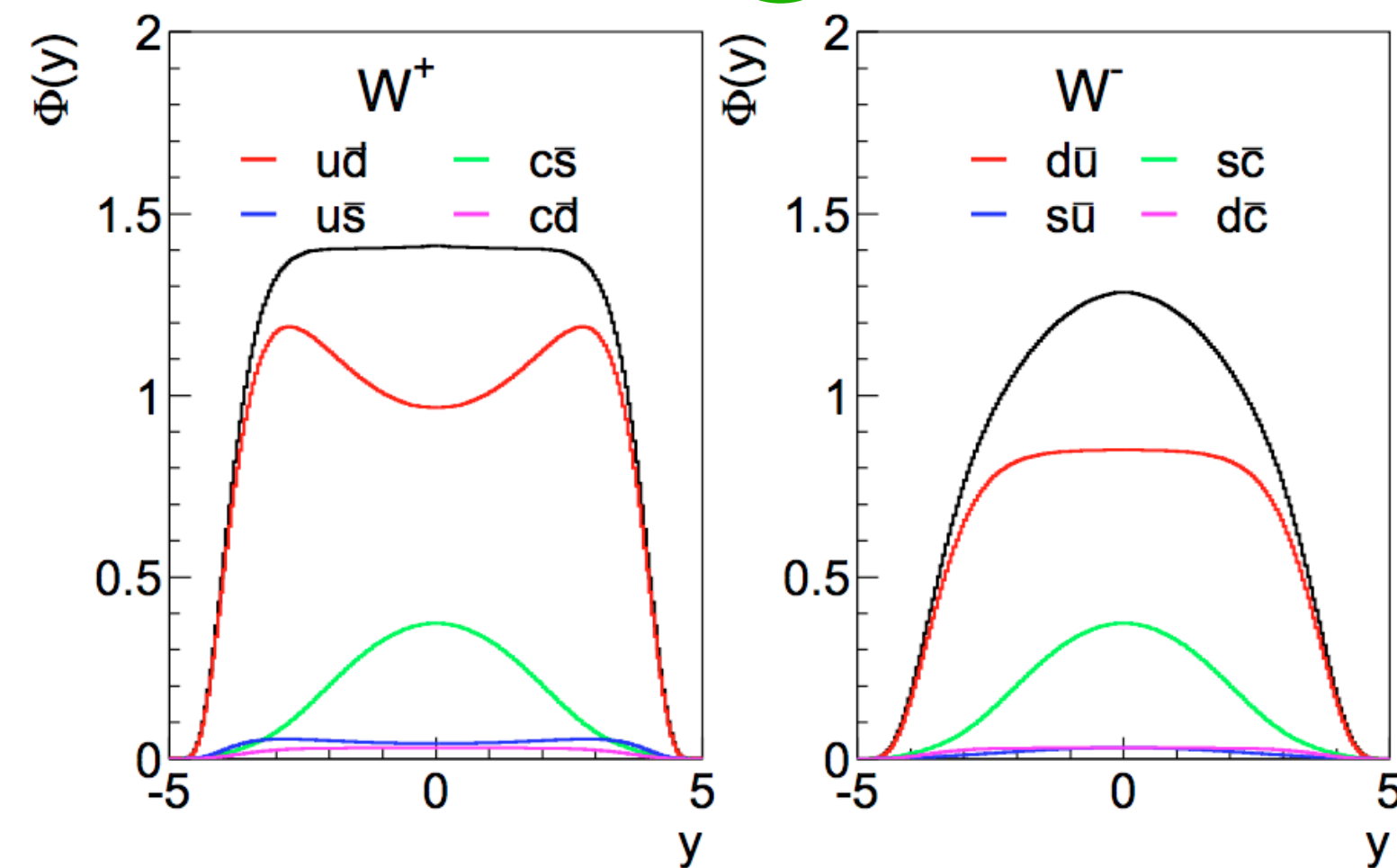
once added to DIS data in a PDF fit:
 significant reduction of uncertainty in the valence

W and Z AT THE LHC: u, d AND s QUARKS

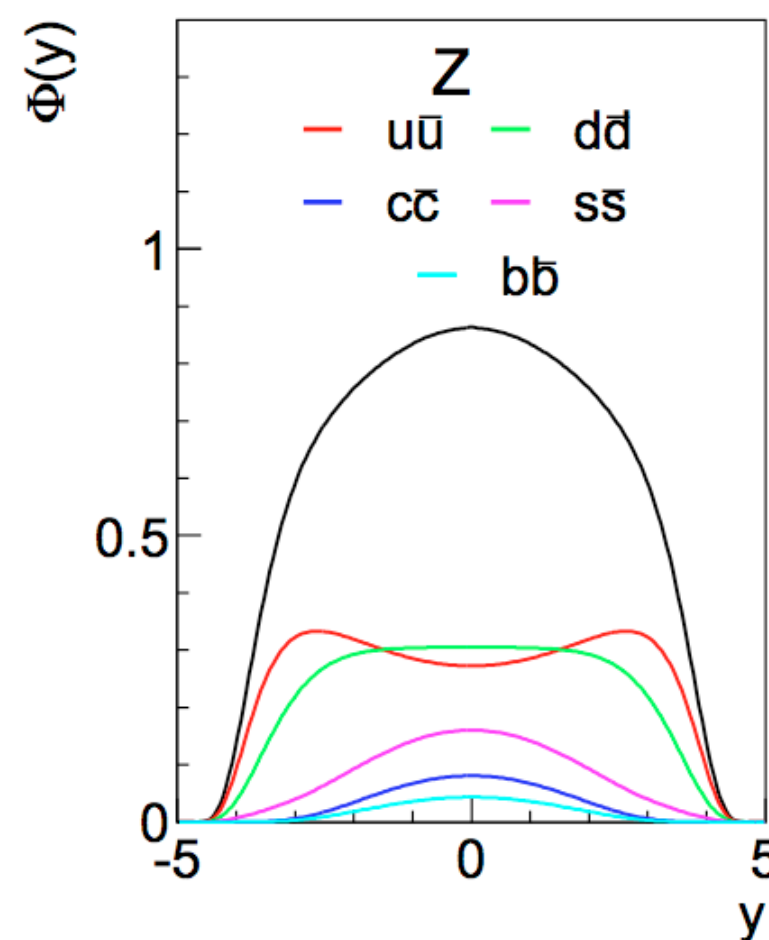
$$W^+ \approx 0.95(u\bar{d} + c\bar{s}) + 0.05(u\bar{s} + c\bar{d})$$

$$W^- \approx 0.95(d\bar{u} + s\bar{c}) + 0.05(d\bar{c} + s\bar{u})$$

DY also probes $s, \bar{s}$ quarks indirectly



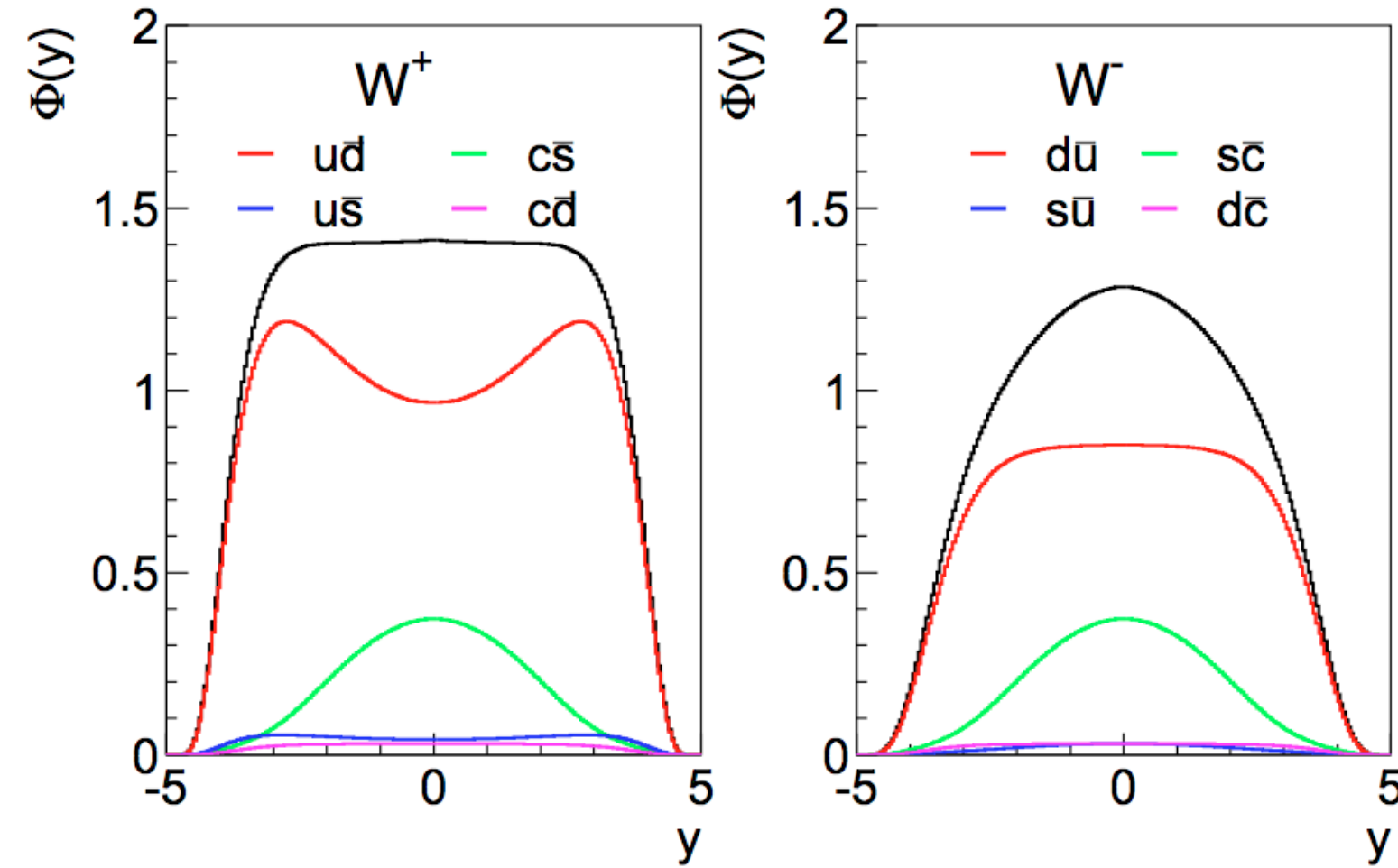
$$Z \approx 0.29(u\bar{u} + c\bar{c}) + 0.37(d\bar{d} + s\bar{s} + b\bar{b})$$



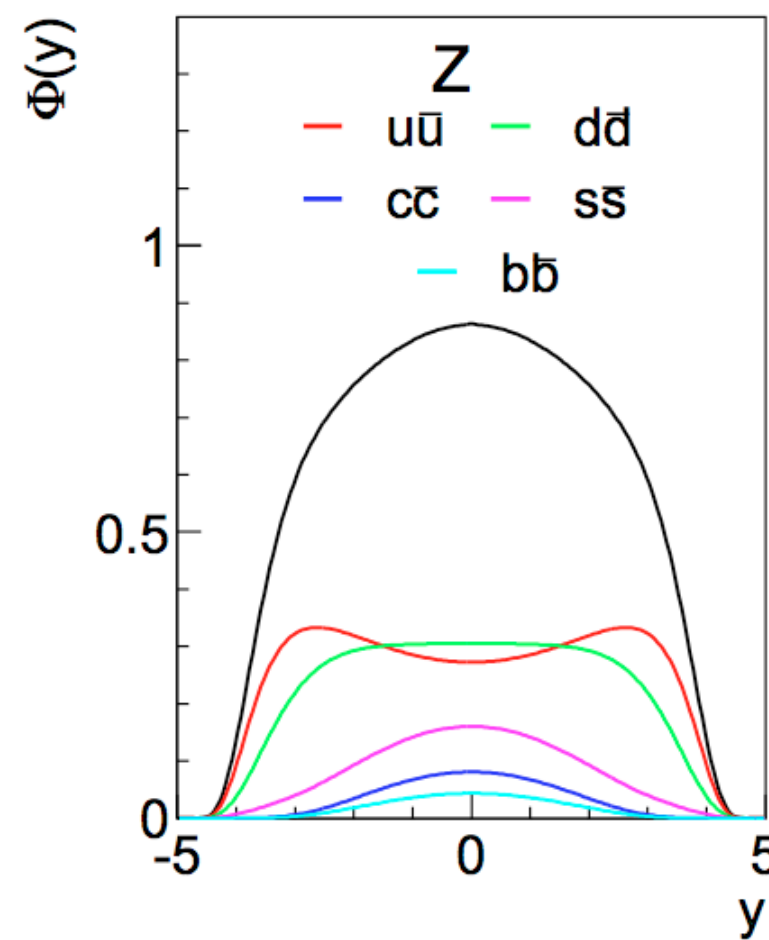
W and Z AT THE LHC: u, d AND s QUARKS

$$W^+ \approx 0.95(u\bar{d} + c\bar{s}) + 0.05(u\bar{s} + c\bar{d})$$

$$W^- \approx 0.95(d\bar{u} + s\bar{c}) + 0.05(d\bar{c} + s\bar{u})$$

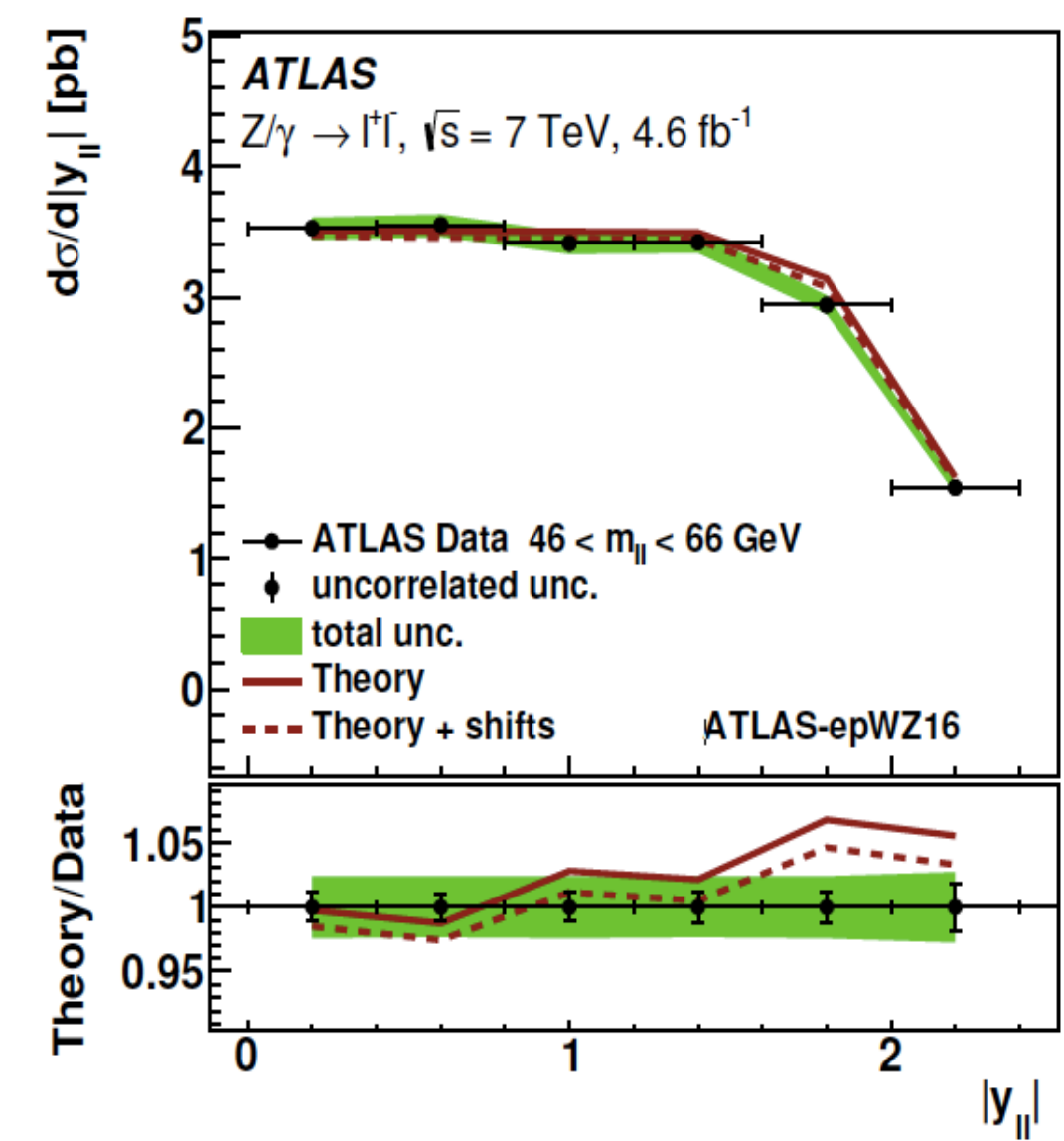
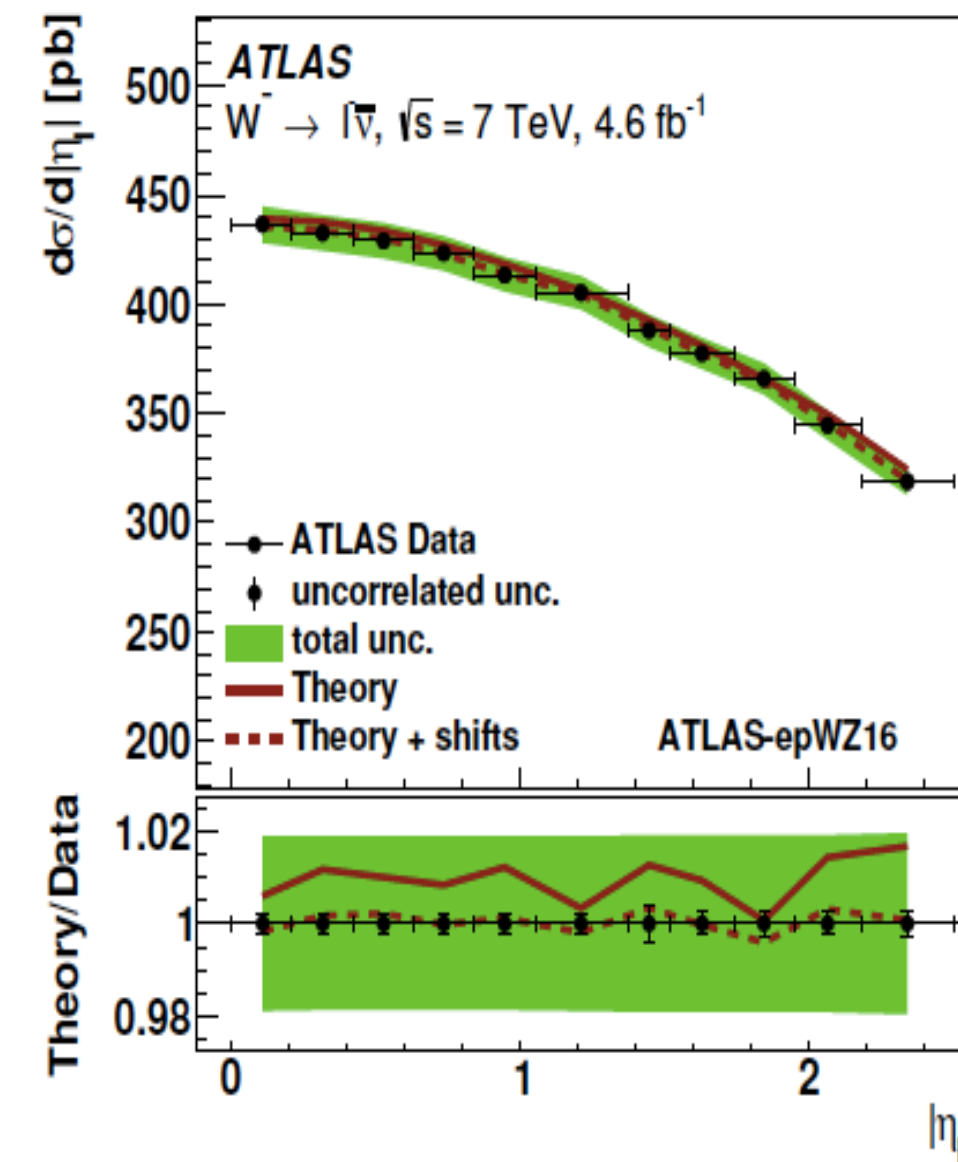
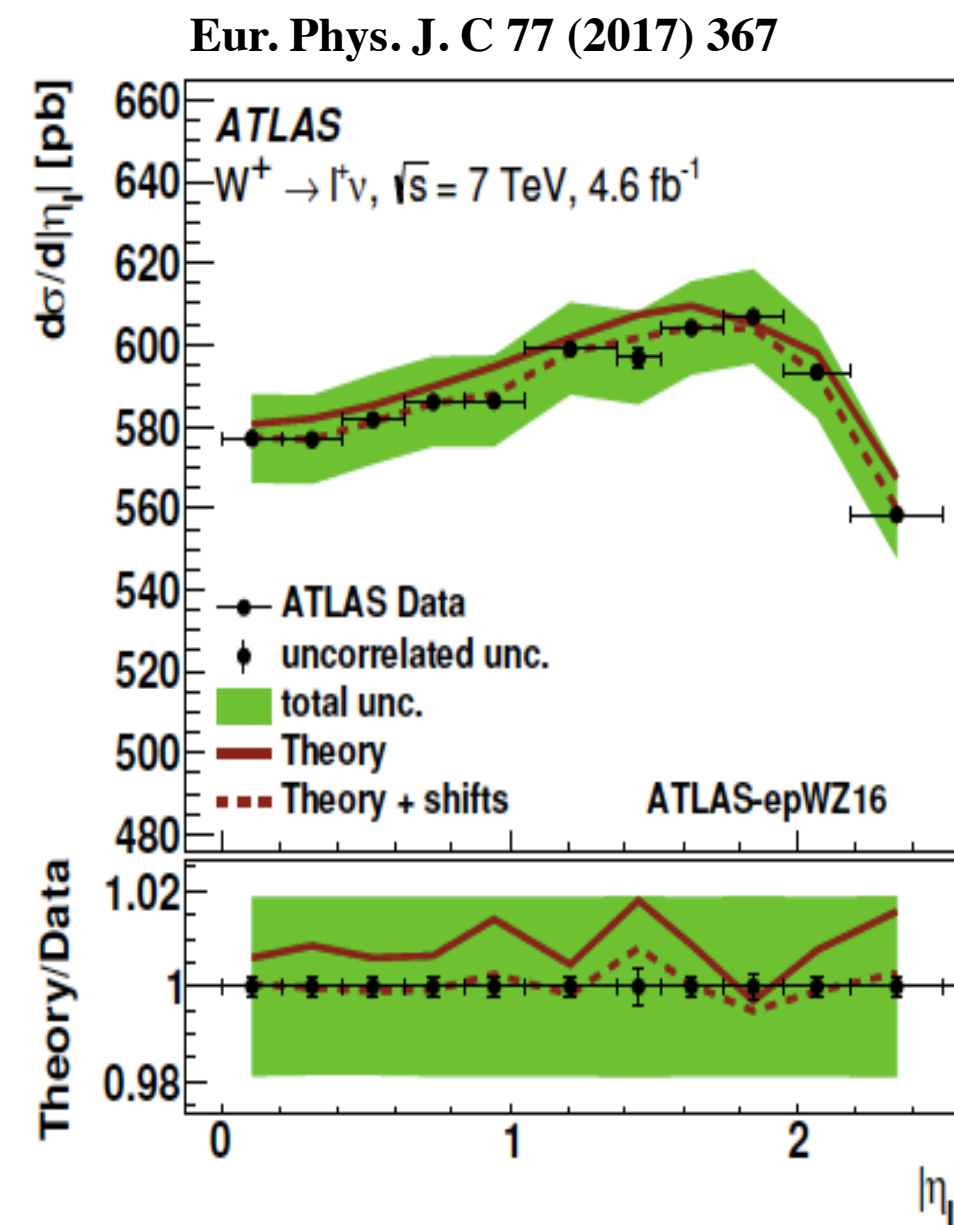


$$Z \approx 0.29(u\bar{u} + c\bar{c}) + 0.37(d\bar{d} + s\bar{s} + b\bar{b})$$



DY also probes $s, \bar{s}$ quarks indirectly

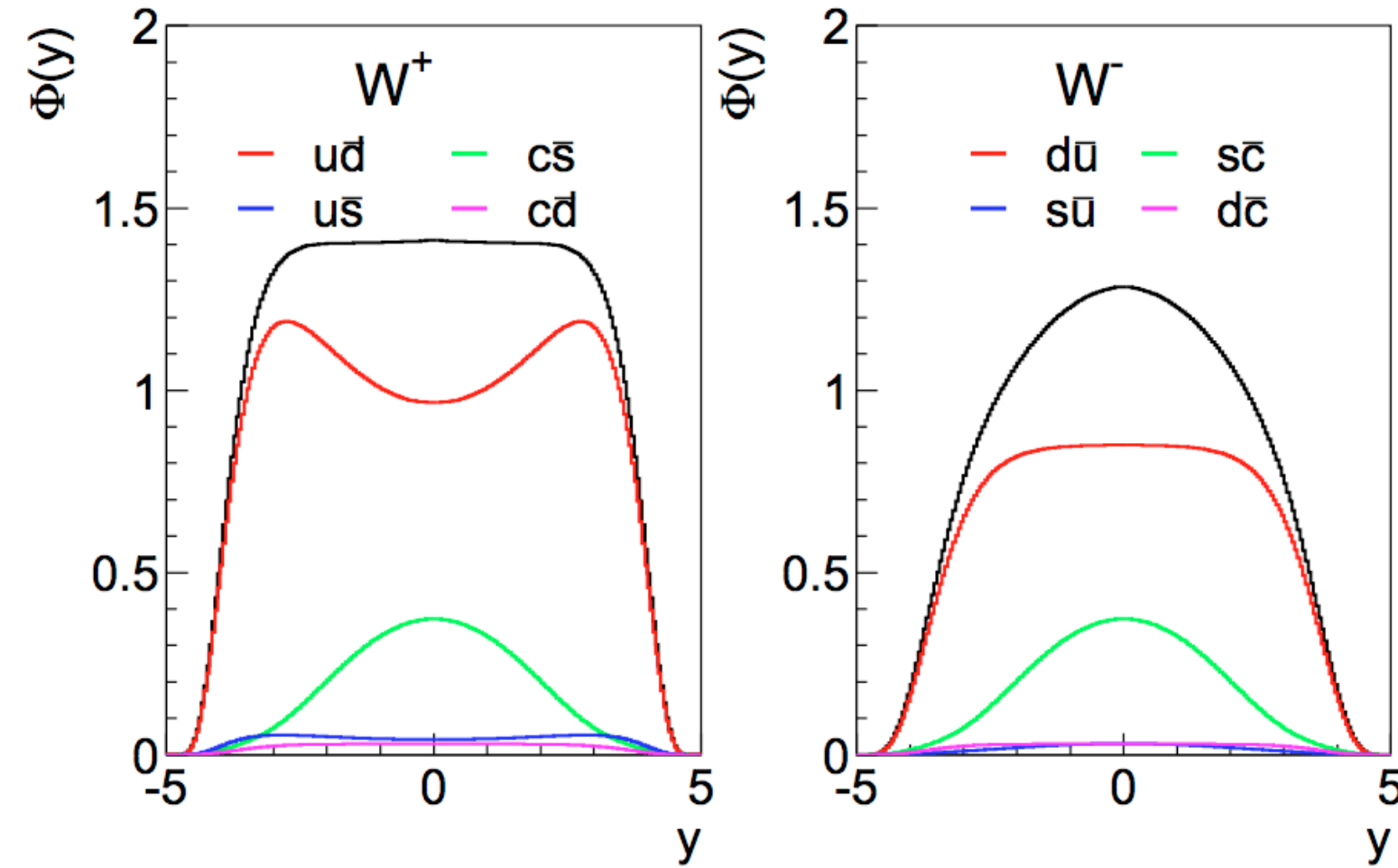
LHC measurements of lepton rapidity in $W^\pm$ and Z/γ^* production



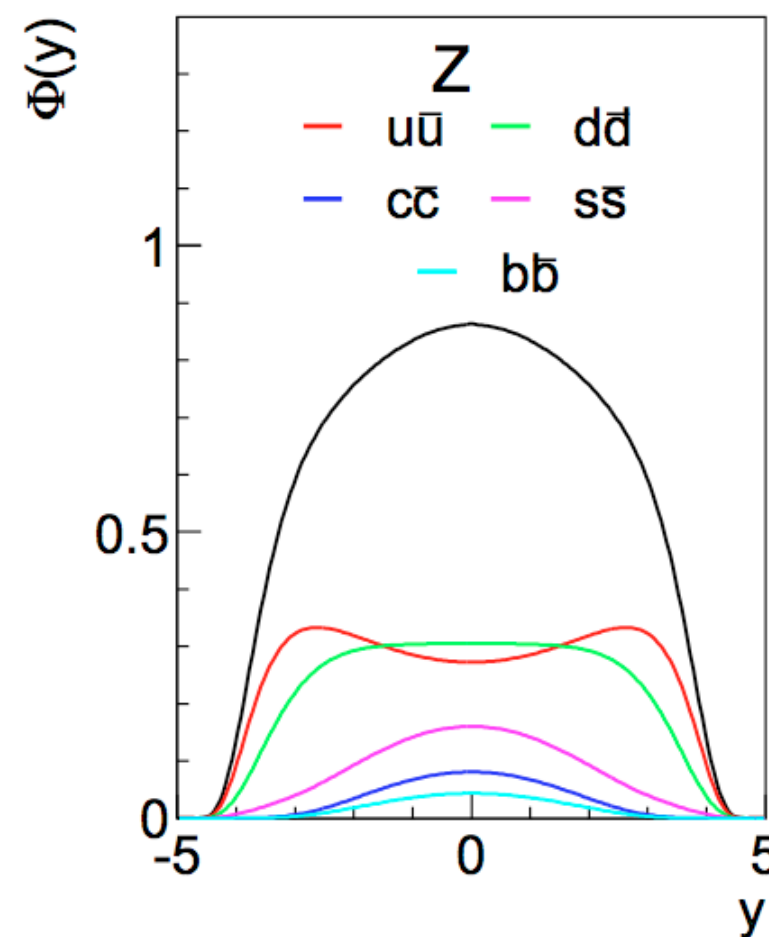
W and Z AT THE LHC: u, d AND s QUARKS

$$W^+ \approx 0.95(u\bar{d} + c\bar{s}) + 0.05(u\bar{s} + c\bar{d})$$

$$W^- \approx 0.95(d\bar{u} + s\bar{c}) + 0.05(d\bar{c} + s\bar{u})$$

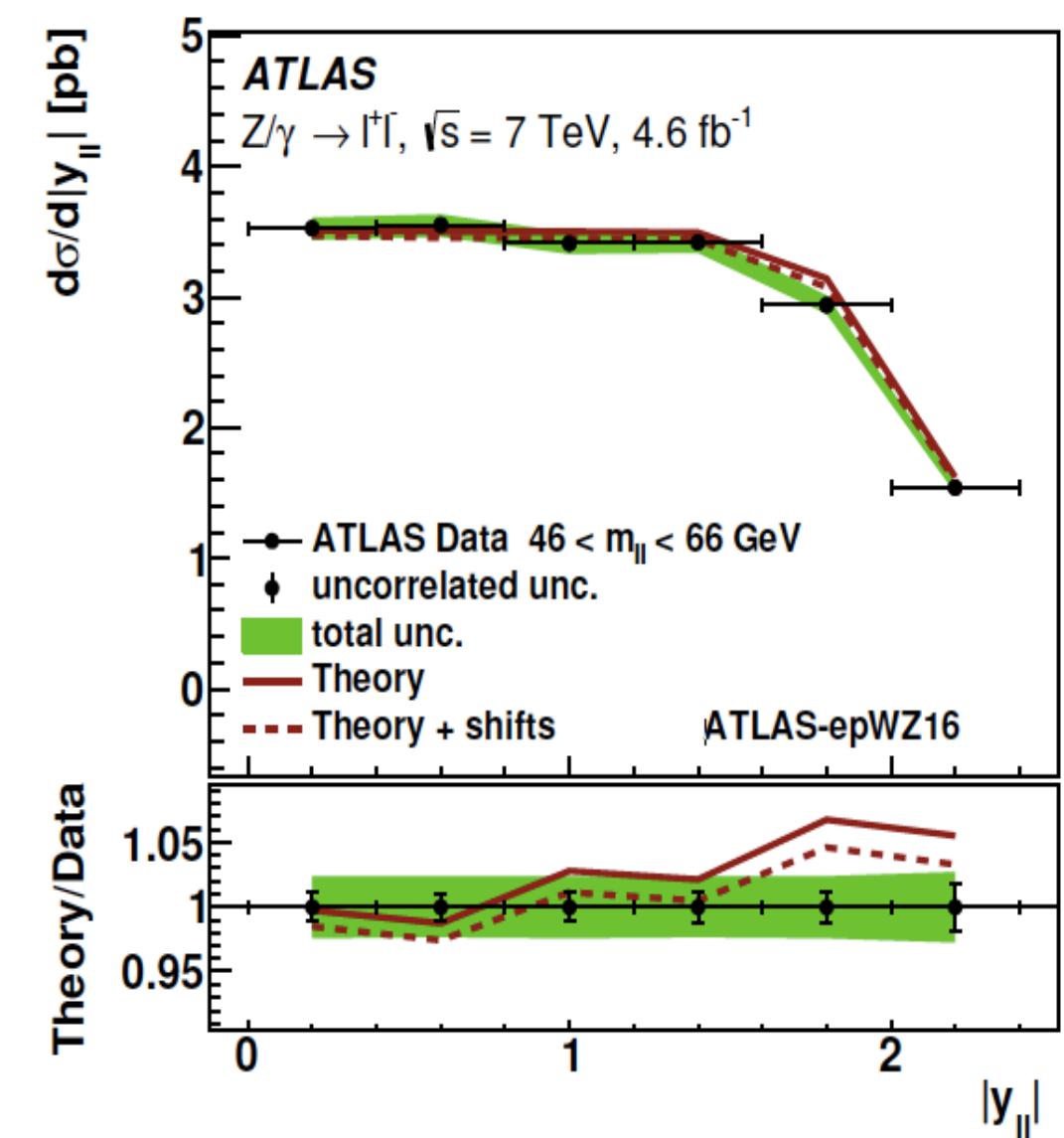
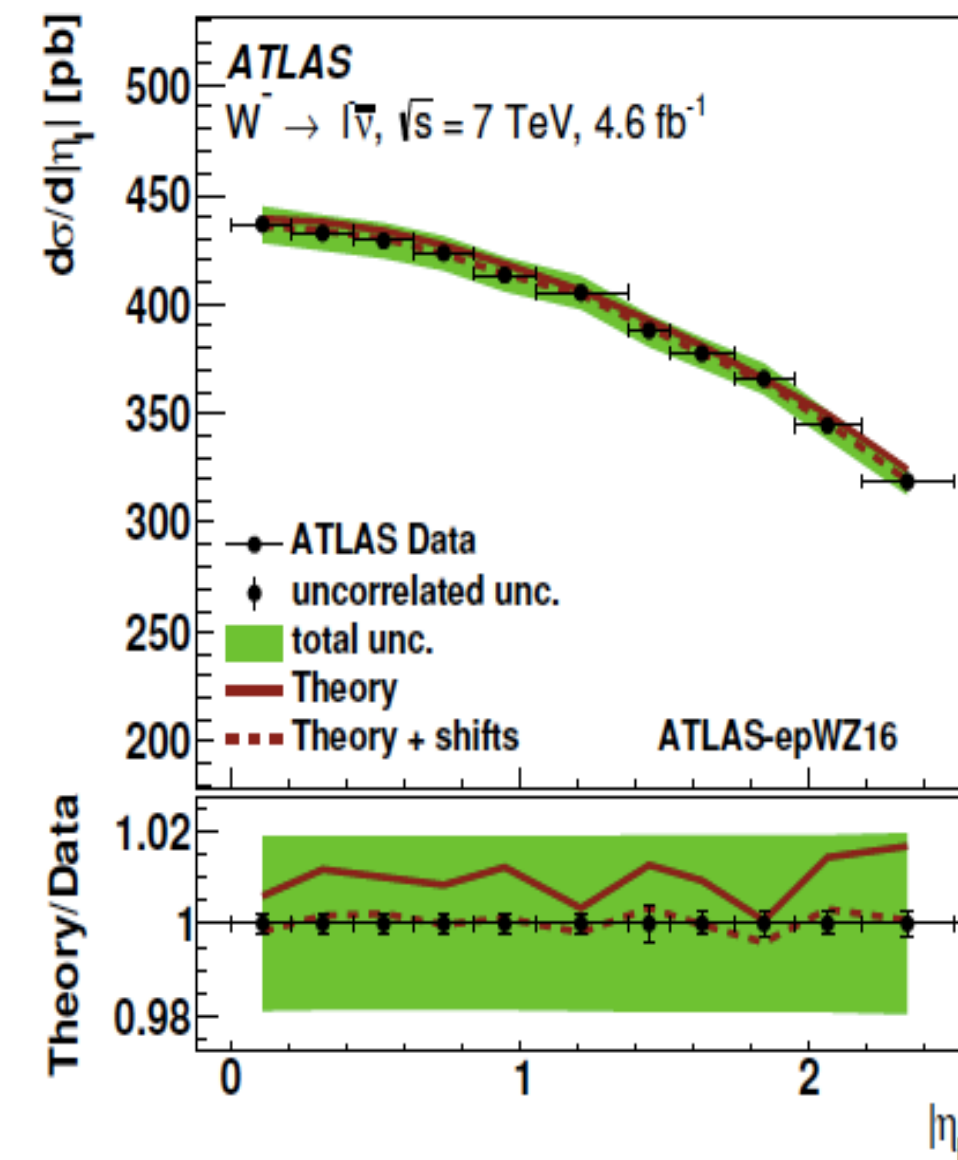
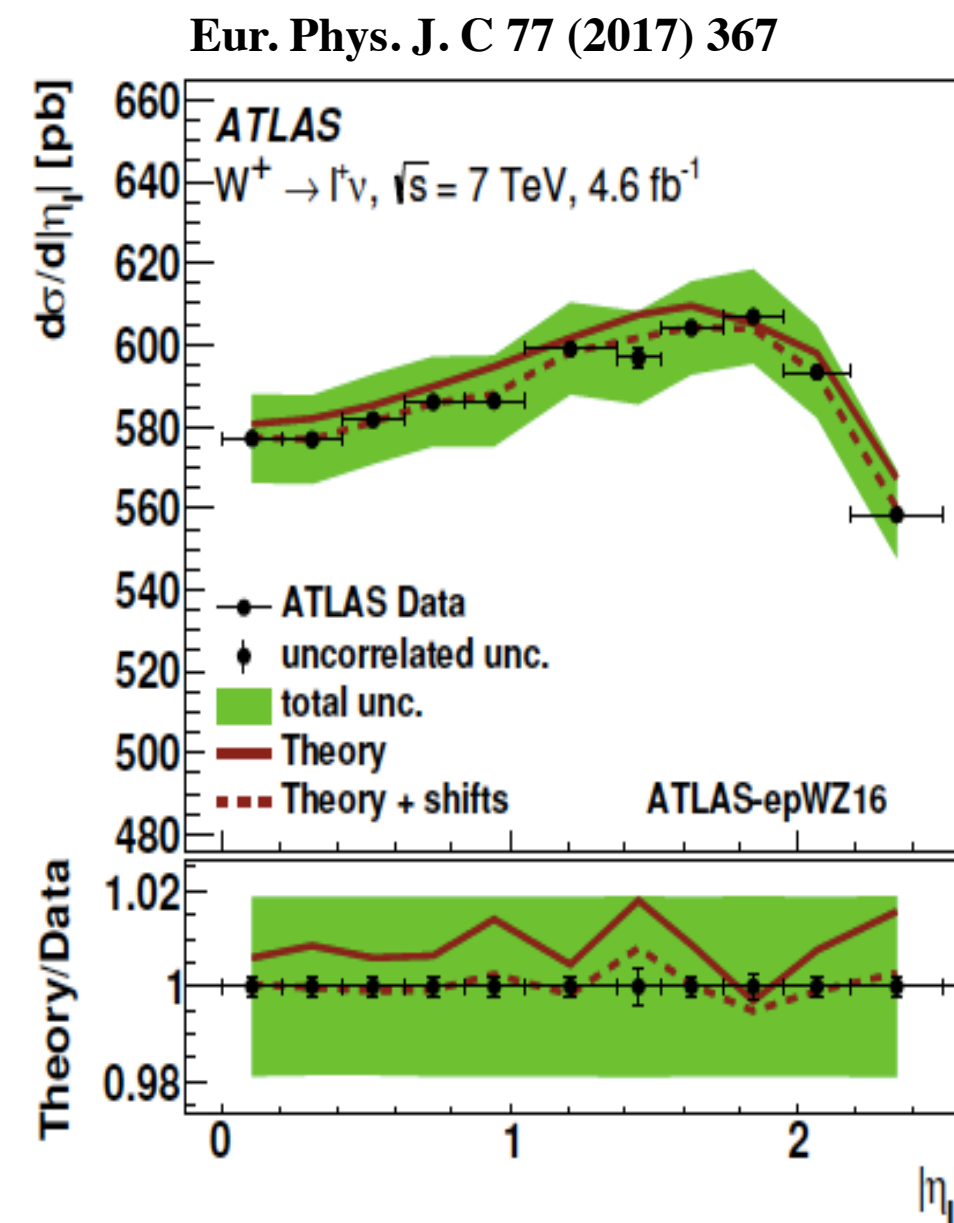


$$Z \approx 0.29(u\bar{u} + c\bar{c}) + 0.37(d\bar{d} + s\bar{s} + b\bar{b})$$



DY also probes $s, \bar{s}$ quarks indirectly

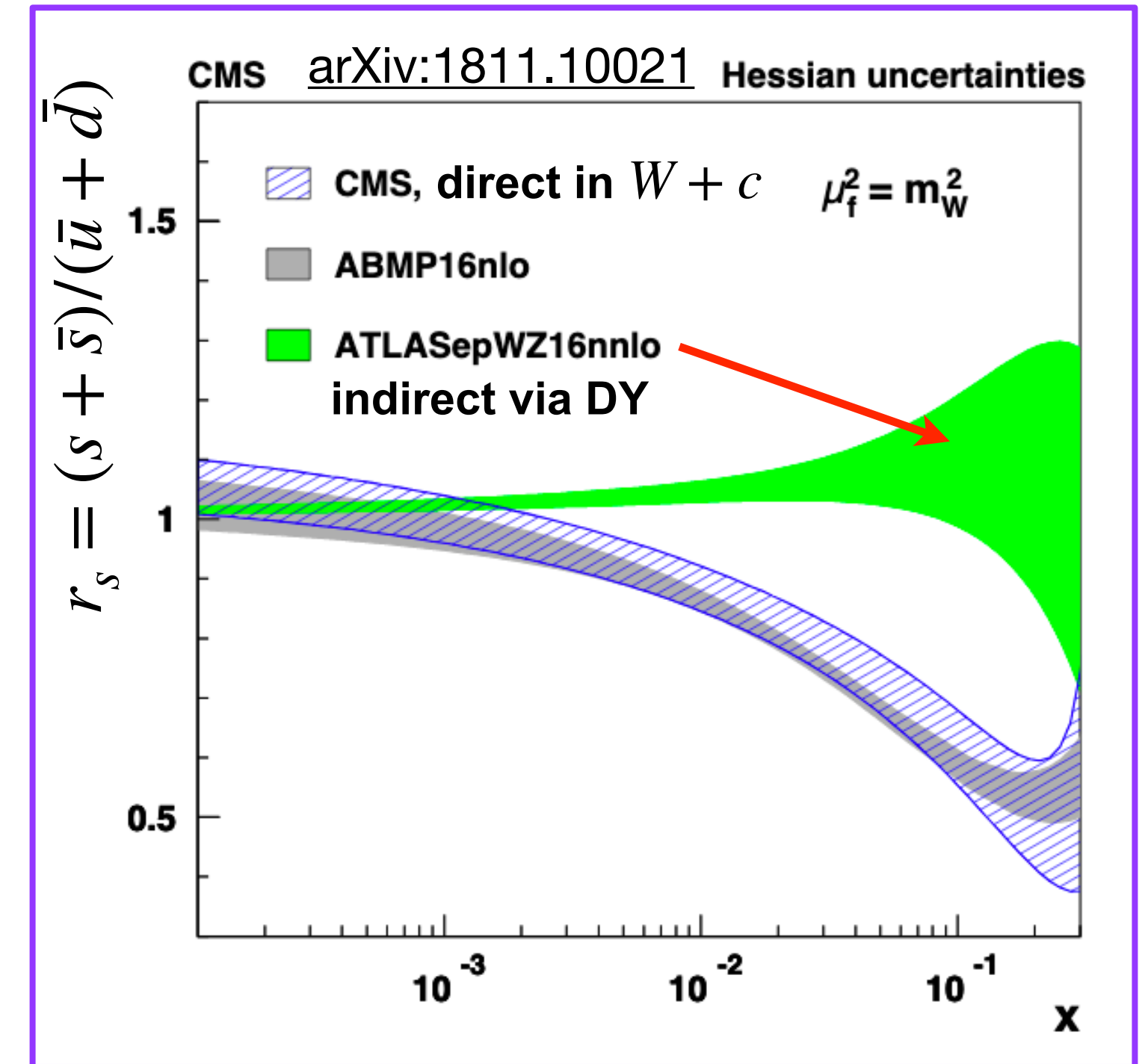
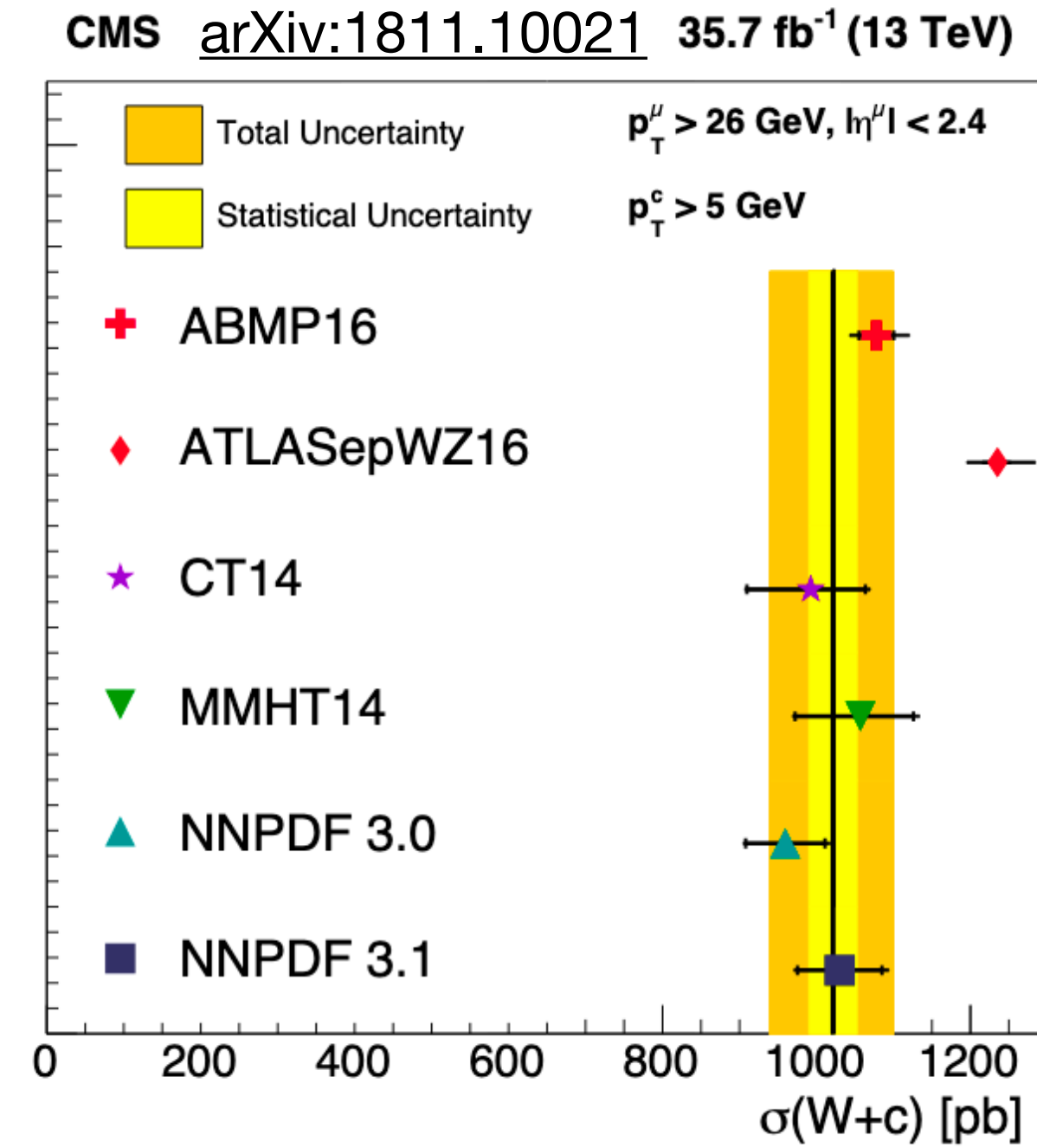
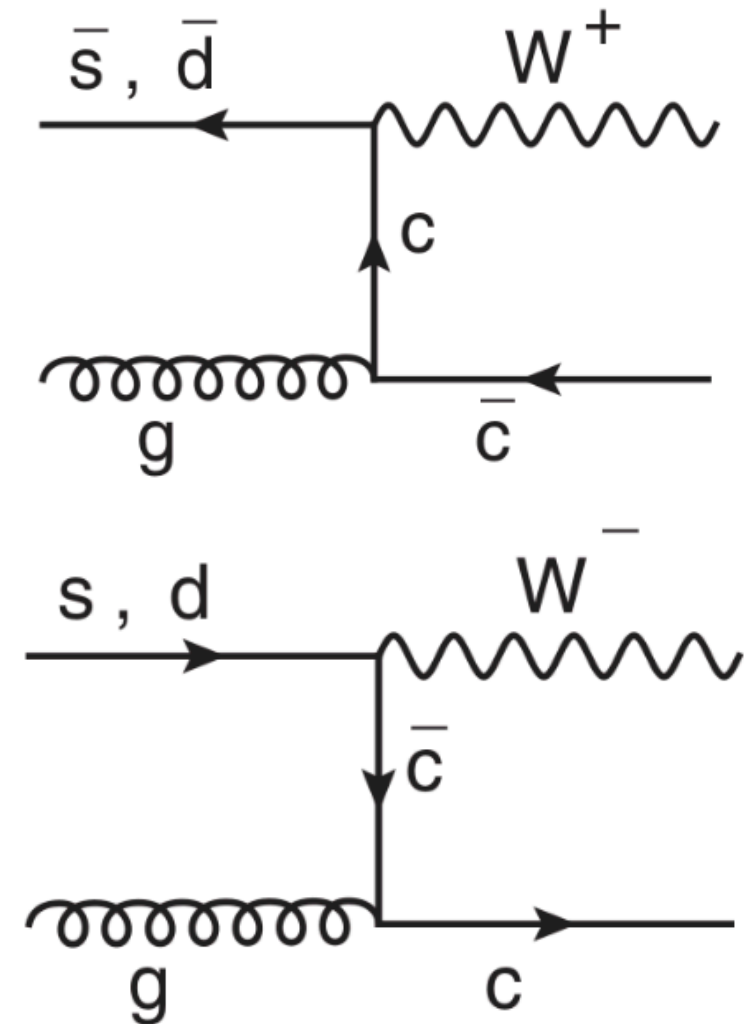
LHC measurements of lepton rapidity in $W^\pm$ and Z/γ^* production



pointed first to unsuppressed strange quark in the sea
(in disagreement with global PDFs)

$W + c$ AT THE LHC: DIRECT PROBE OF s QUARK

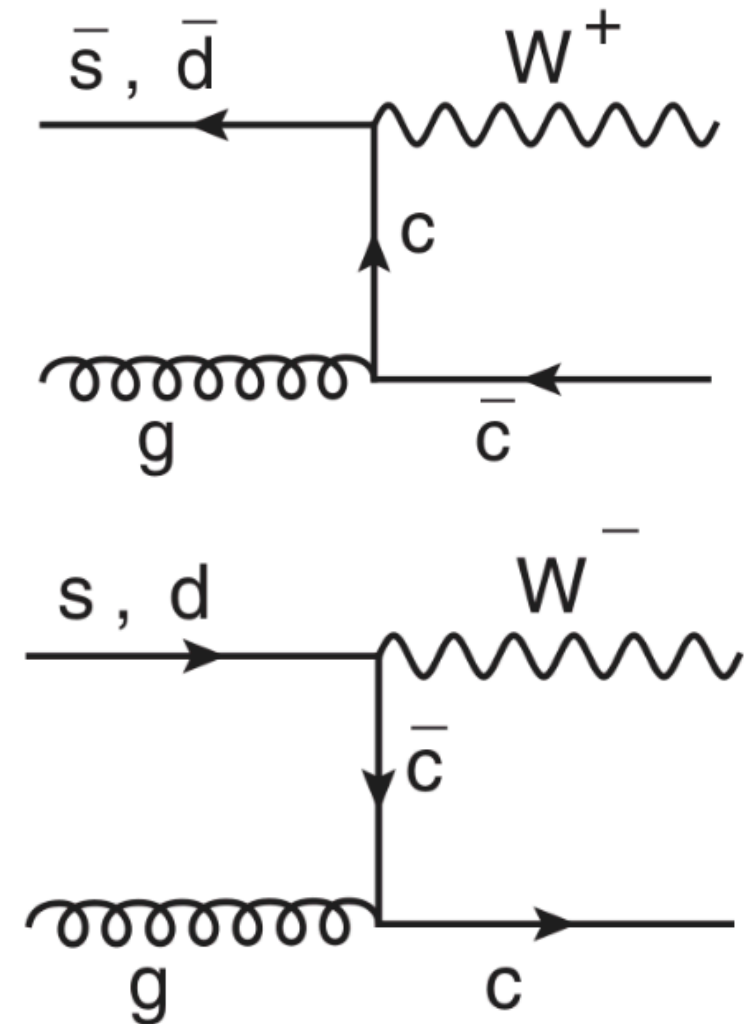
$W + c$ produced in sg fusion:



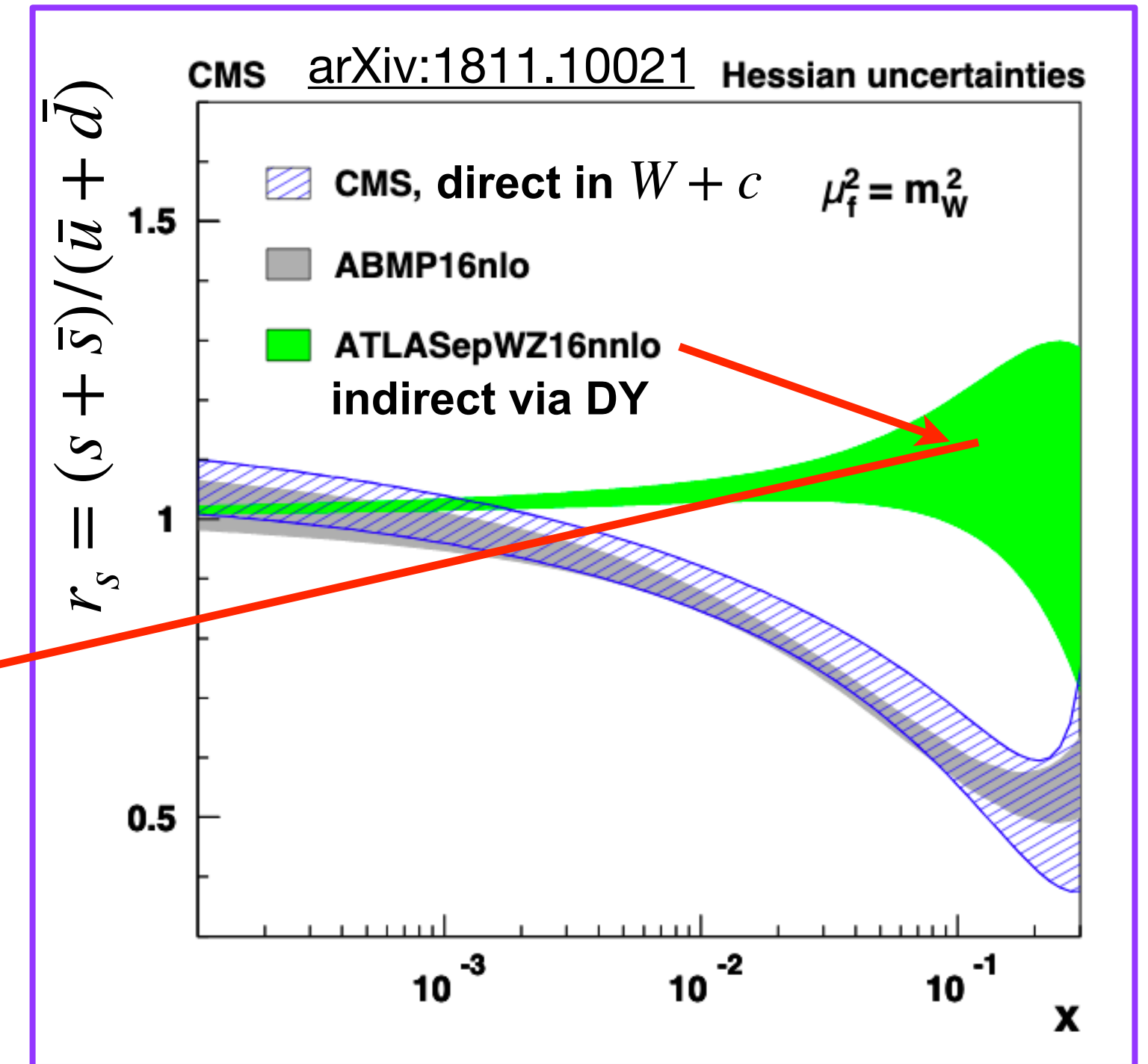
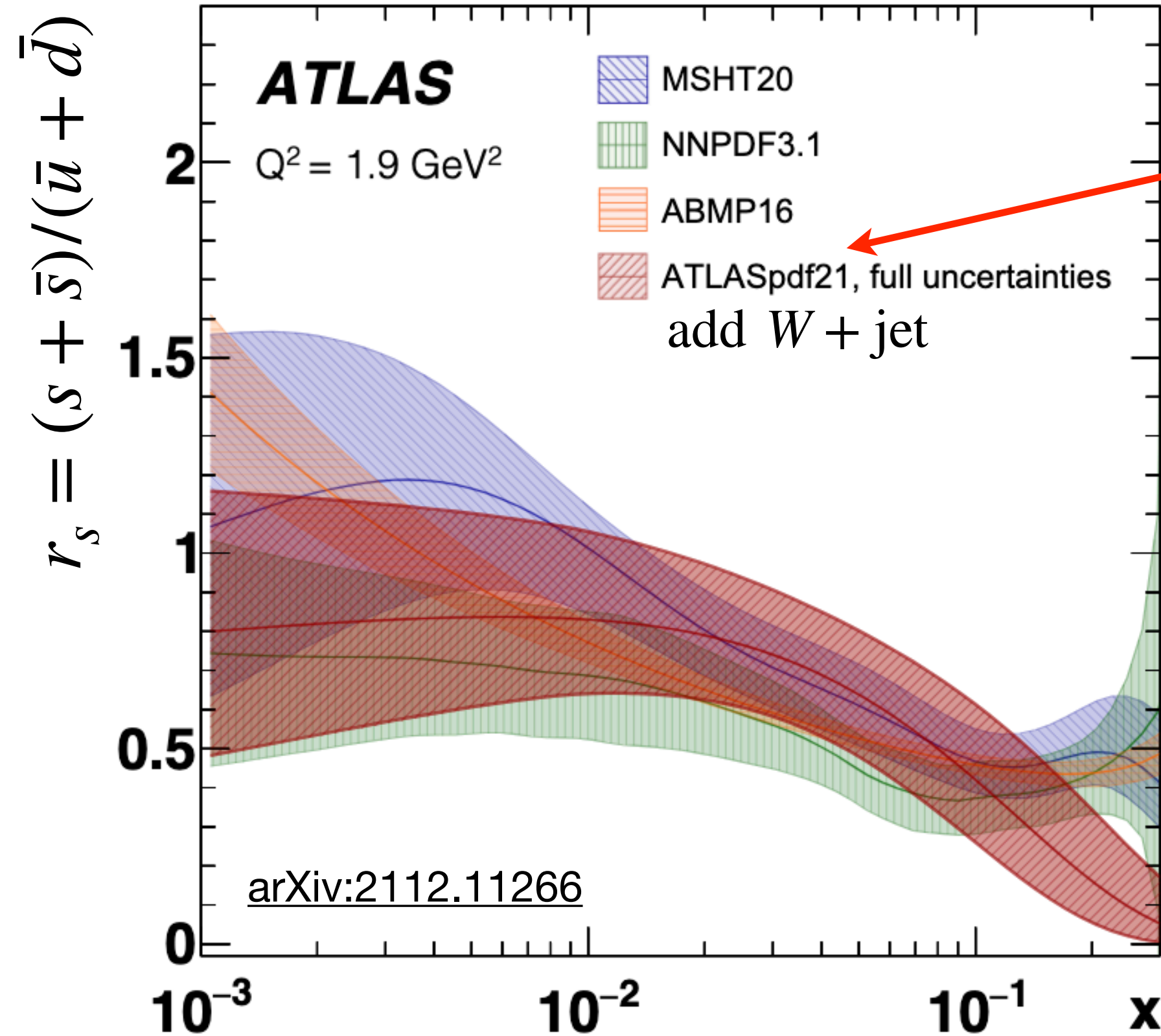
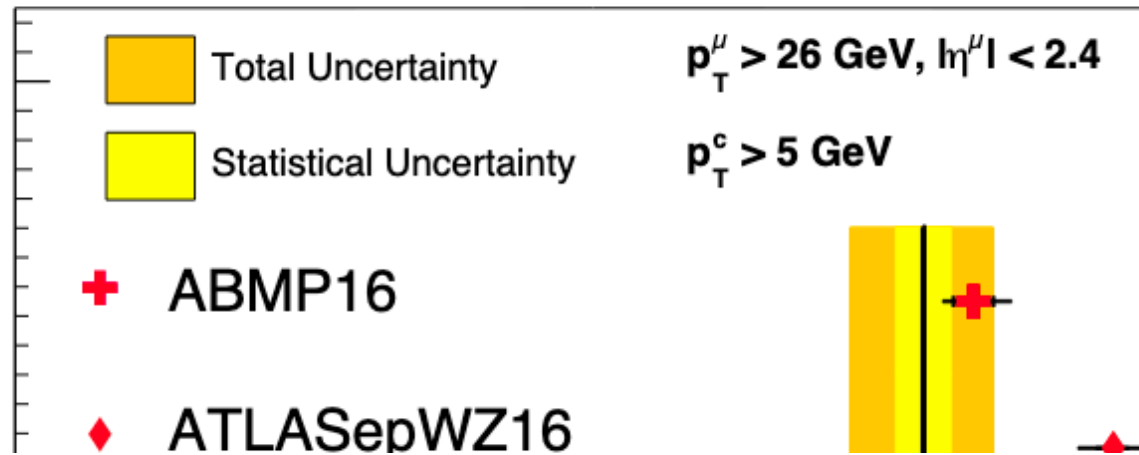
once added to DIS data in PDF fit: strangeness suppression agrees with global PDFs (not with first indirect extraction)

$W + c$ AT THE LHC: DIRECT PROBE OF s QUARK

$W + c$ produced in sg fusion:



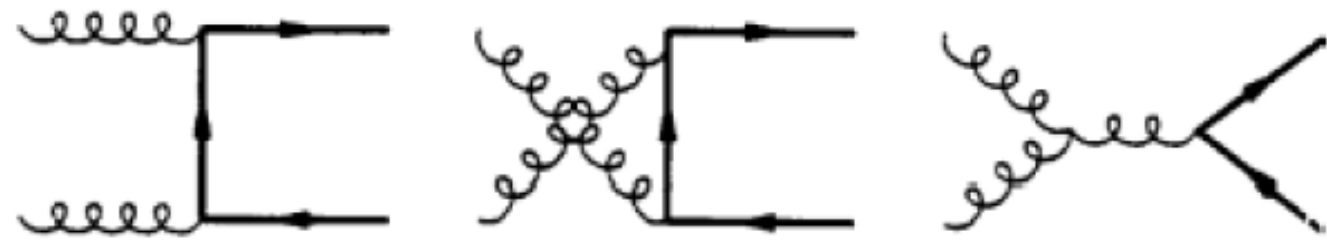
CMS arXiv:1811.10021 35.7 fb⁻¹ (13 TeV)



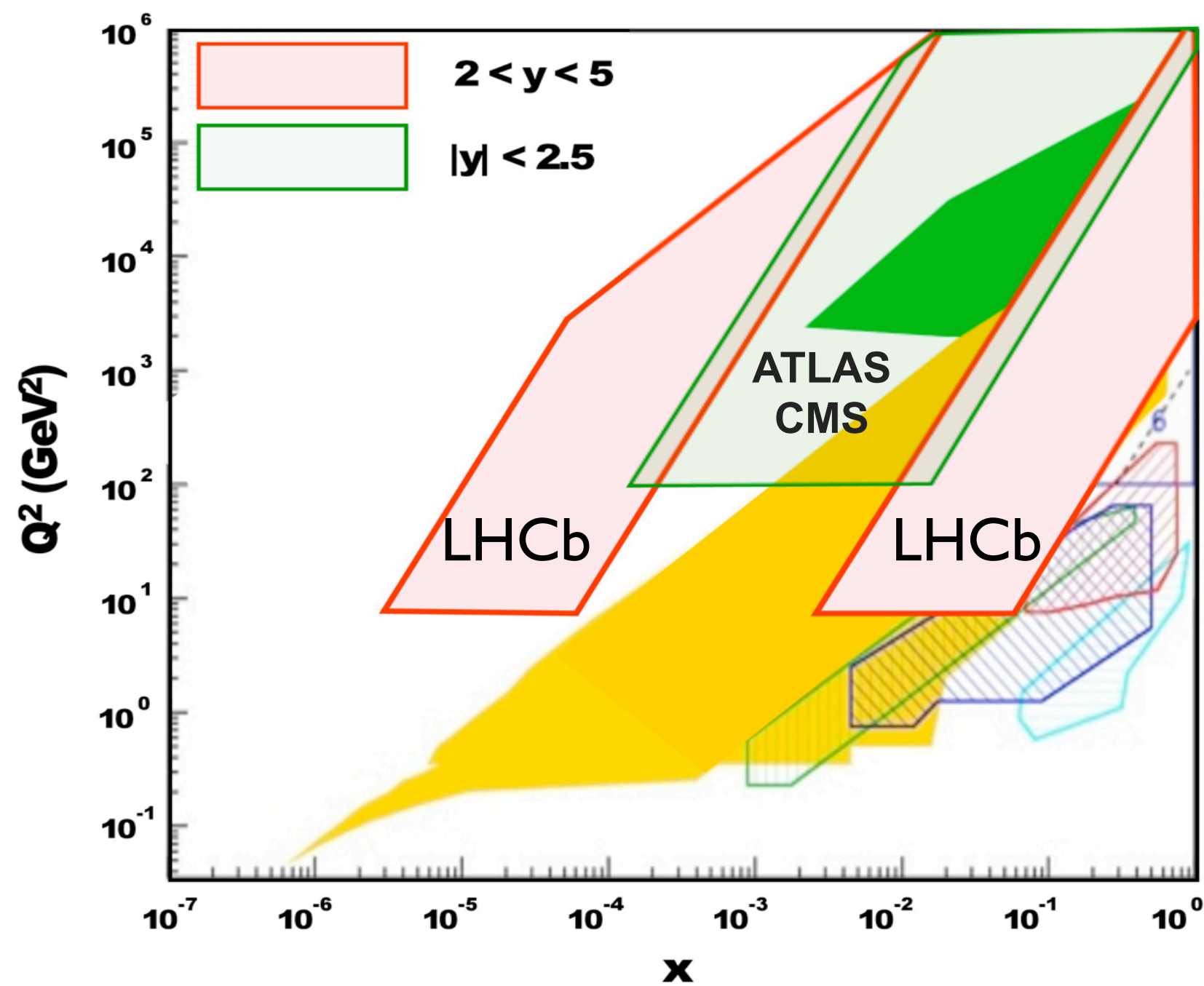
meanwhile, the disagreement resolved by adding $W + \text{jet}$ data to DY measurements

HEAVY QUARKS AT THE LHC: GLUON, m_Q , AND α_S

Heavy-quark production in pp :

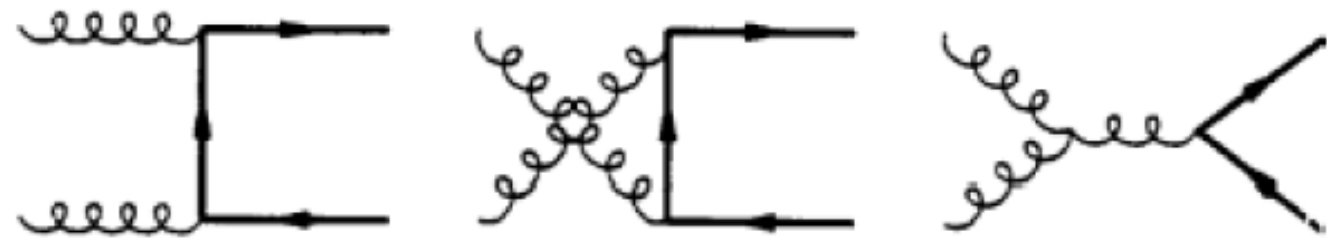


Forward heavy-flavour production at
the LHC probes gluon down to $x \sim 10^{-5}$



HEAVY QUARKS AT THE LHC: GLUON, m_Q , AND α_S

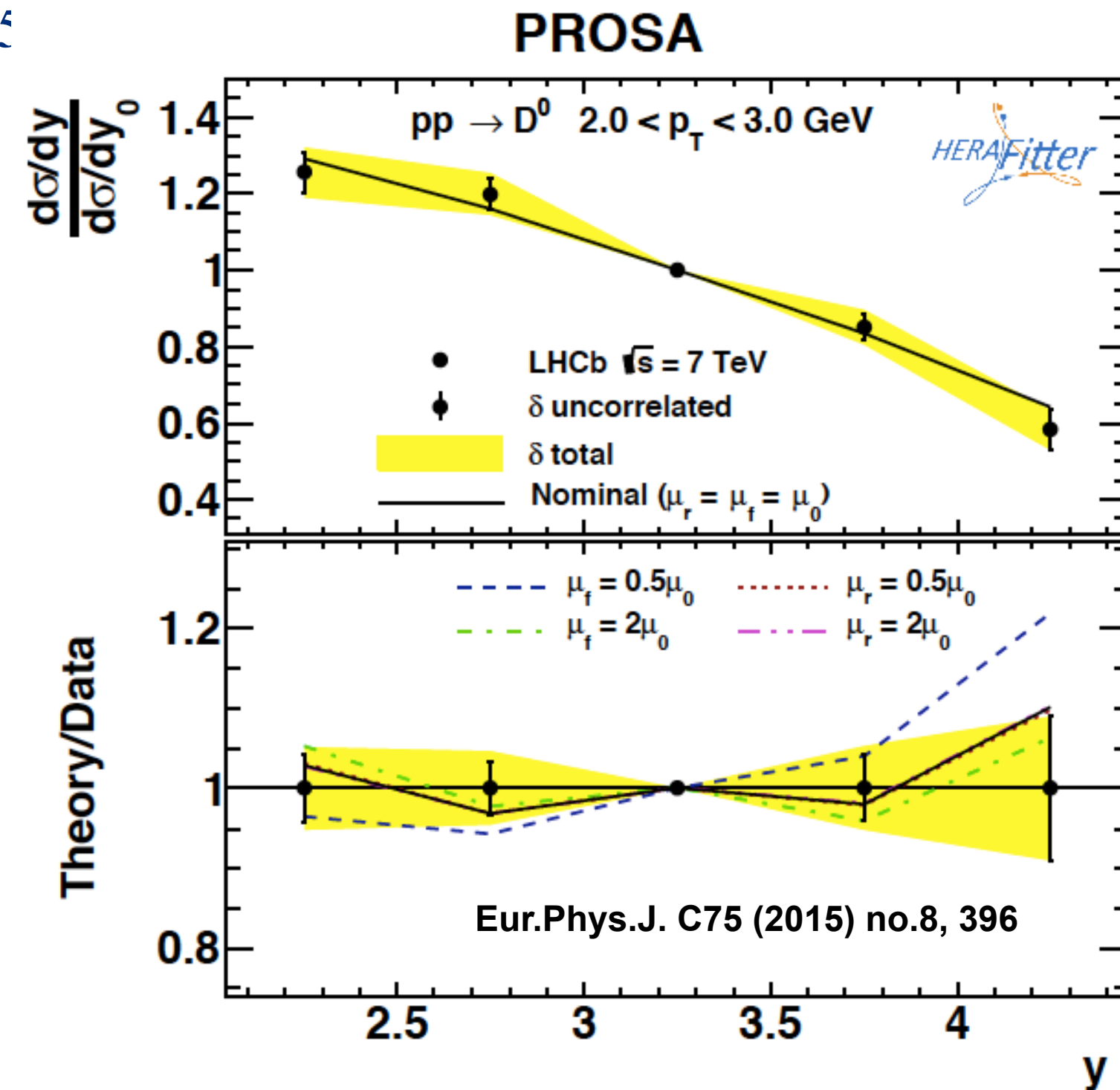
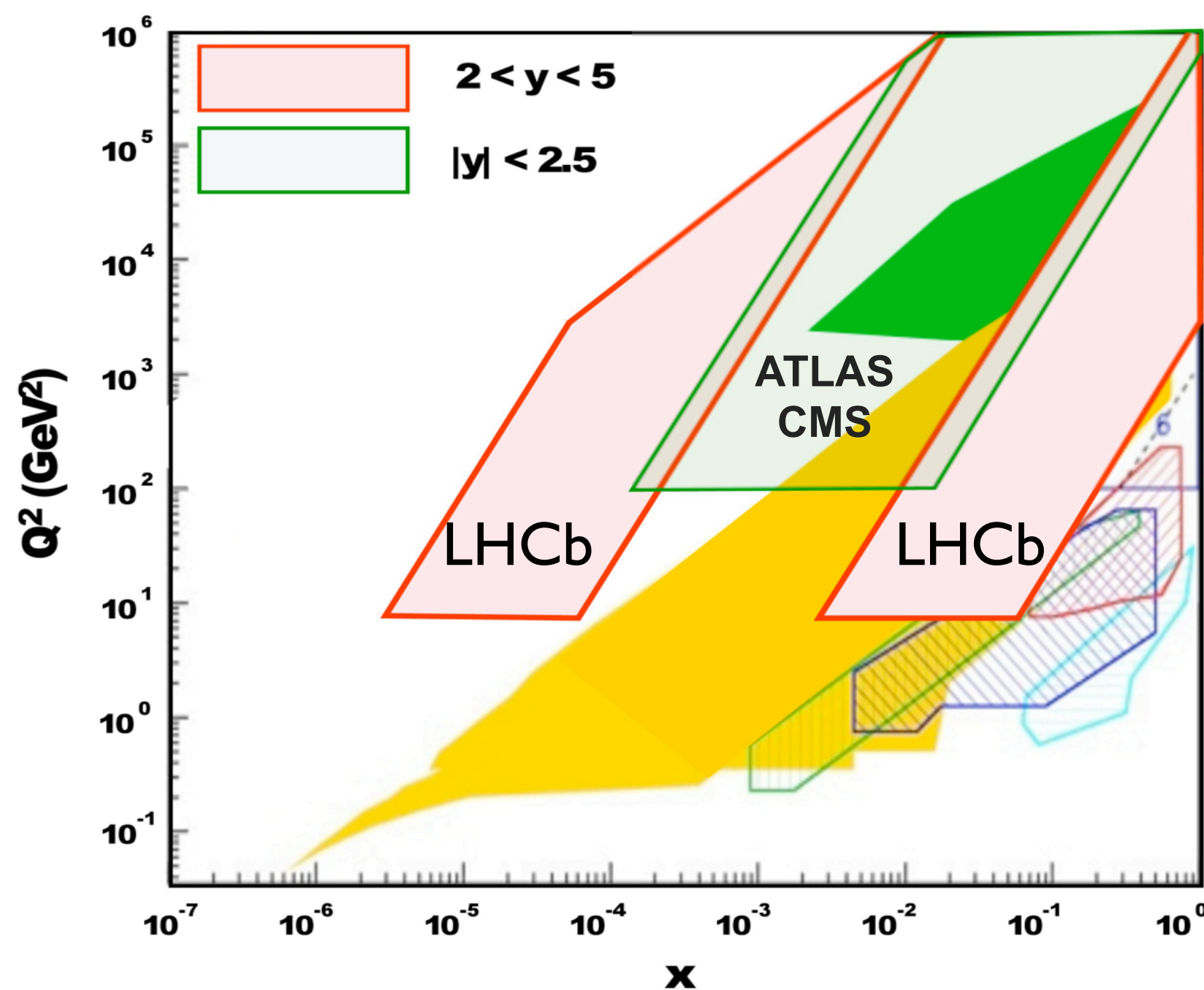
Heavy-quark production in pp :



c [NPB871 (2013) 1] at $\sqrt{s} = 7$ TeV

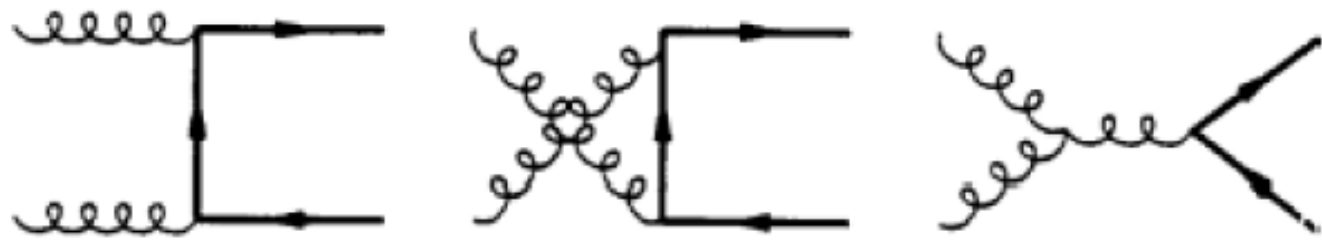
b [JHEP 08 (2013) 117] at $\sqrt{s} = 7$ TeV

Forward heavy-flavour production at the LHC probes gluon down to $x \sim 10^{-5}$



HEAVY QUARKS AT THE LHC: GLUON, m_Q , AND α_S

Heavy-quark production in pp :

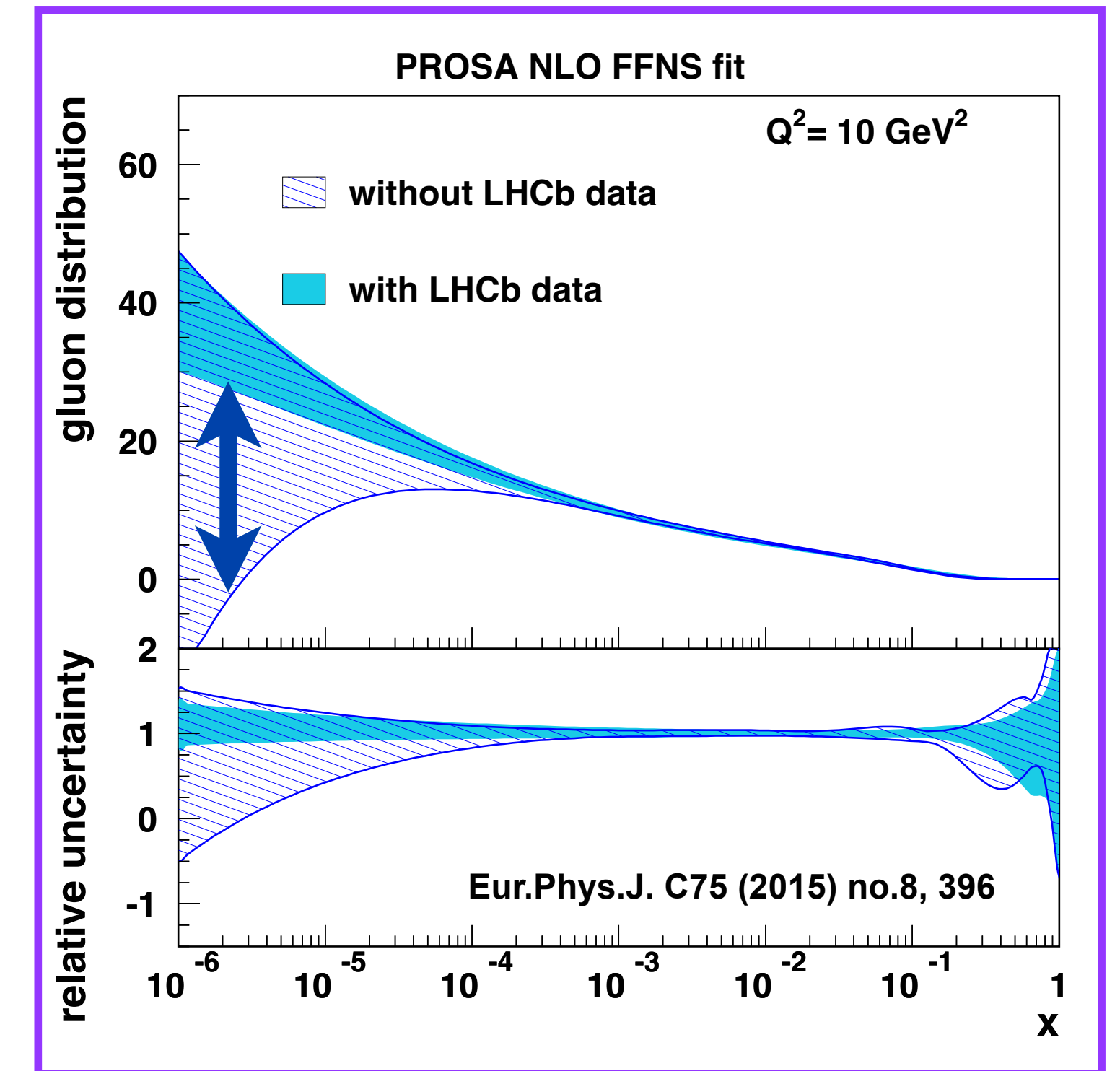
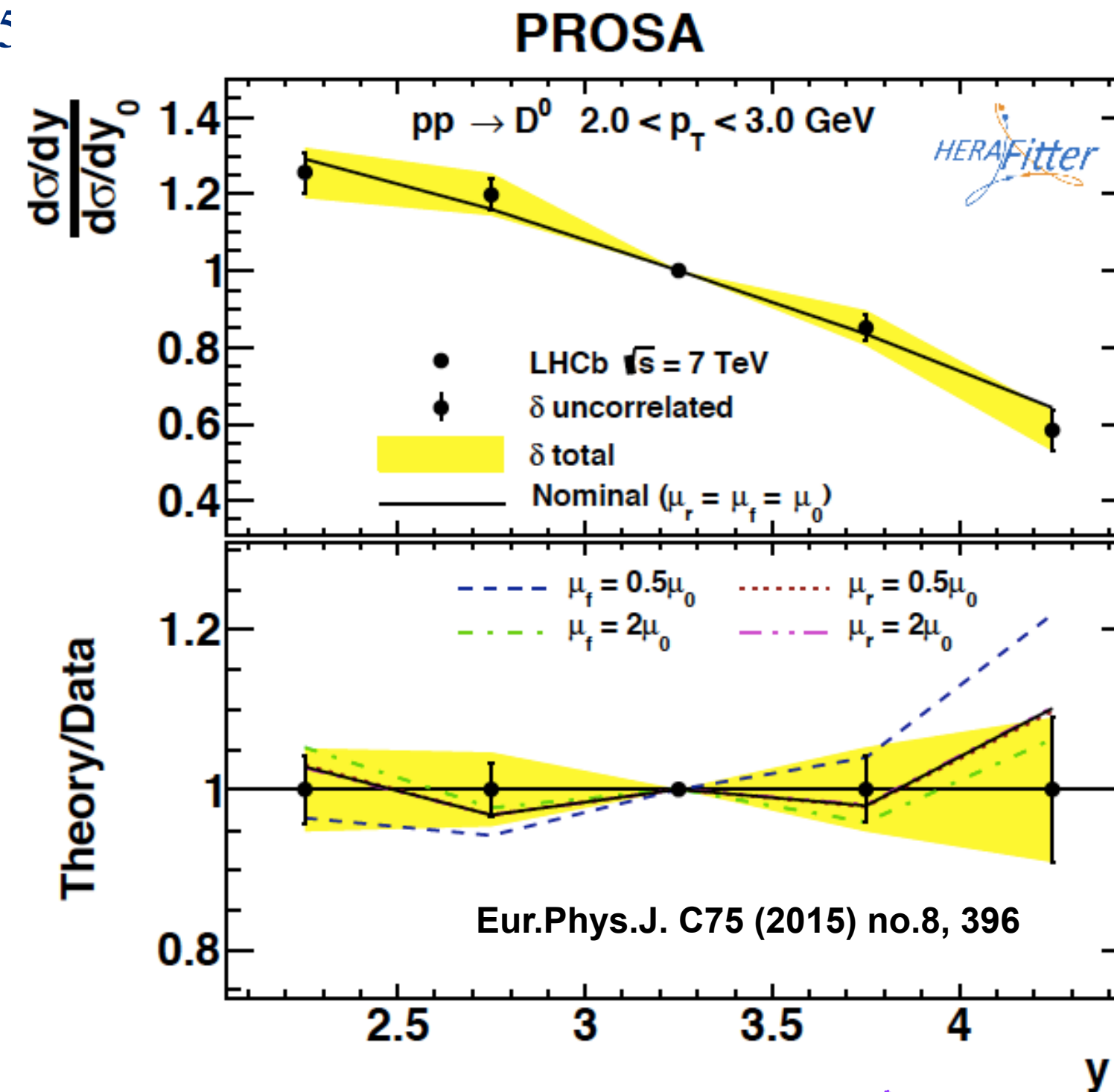
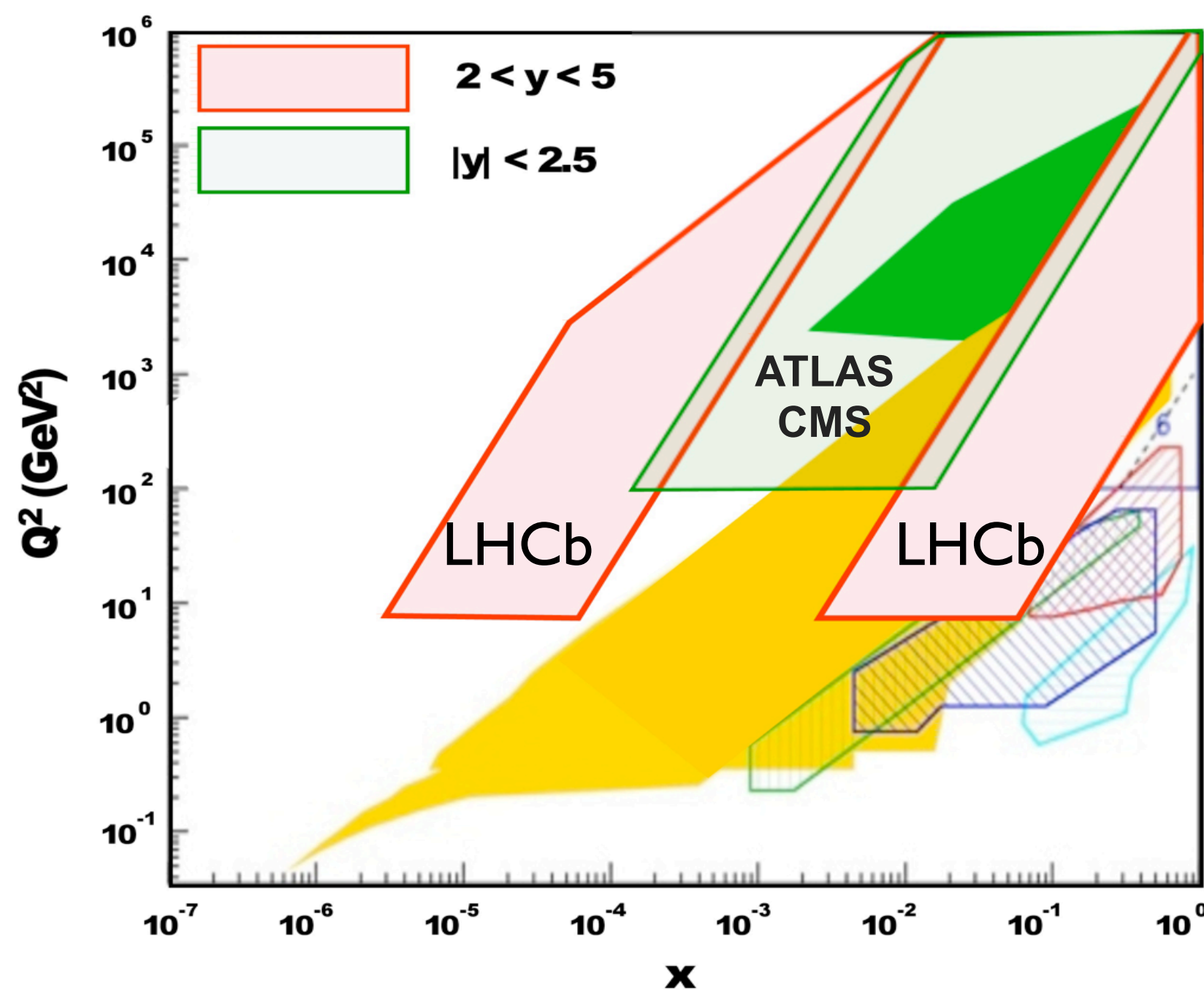


c [NPB871 (2013) 1] at $\sqrt{s} = 7$ TeV

b [JHEP 08 (2013) 117] at $\sqrt{s} = 7$ TeV

Use in PDF fit [EPJC 75 (2015) 8, 396]

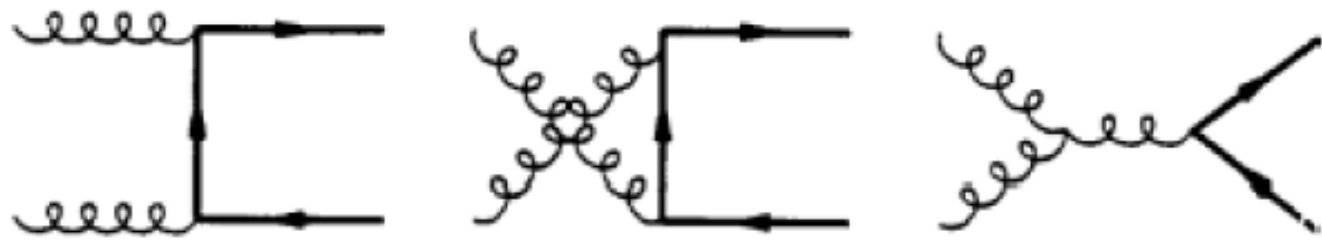
Forward heavy-flavour production at the LHC probes gluon down to $x \sim 10^{-5}$



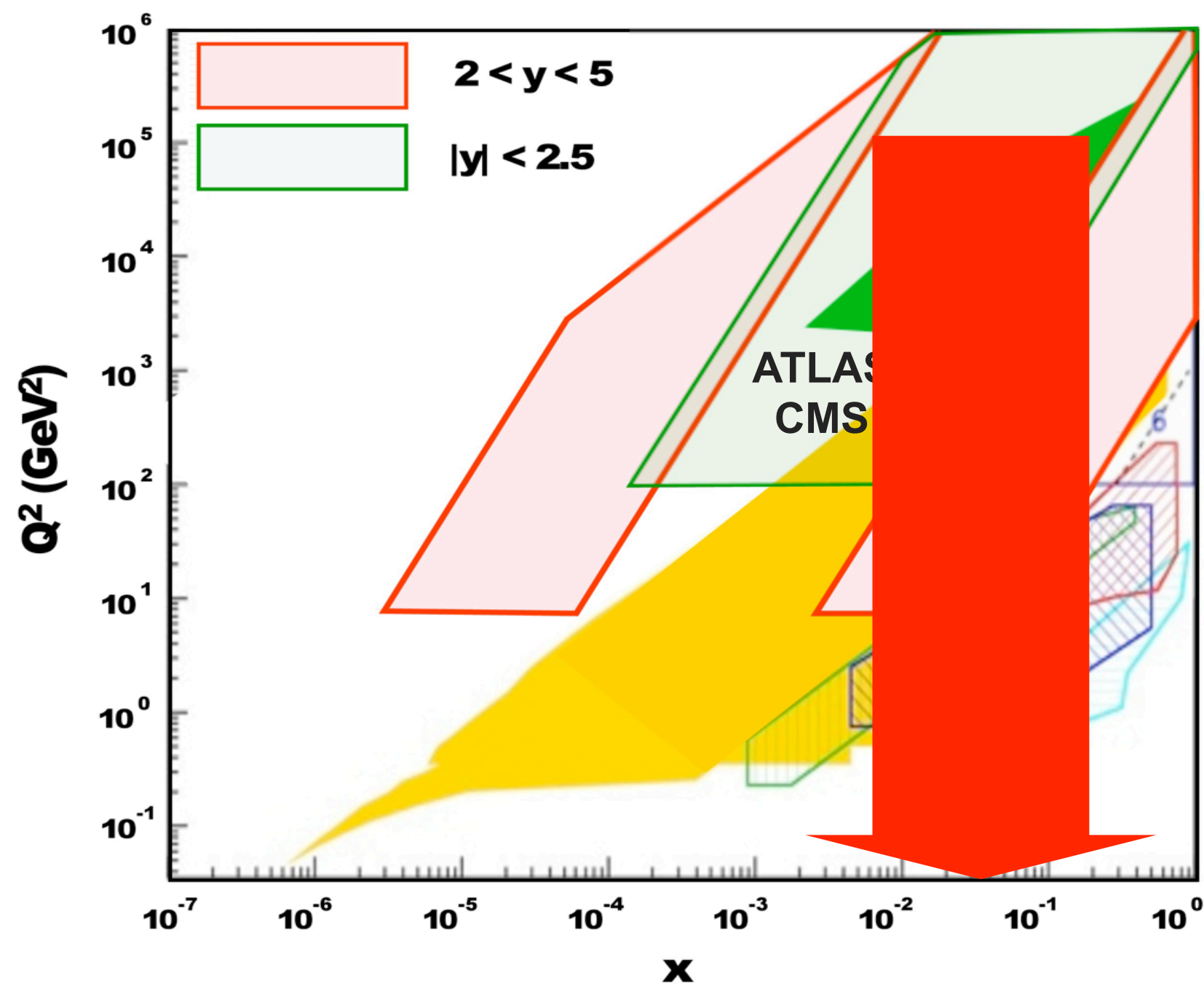
reduced uncertainty in $g(x)$ at very low x !

TOP QUARKS AT THE LHC: GLUON, m_Q , AND α_S

Heavy-quark production in pp :

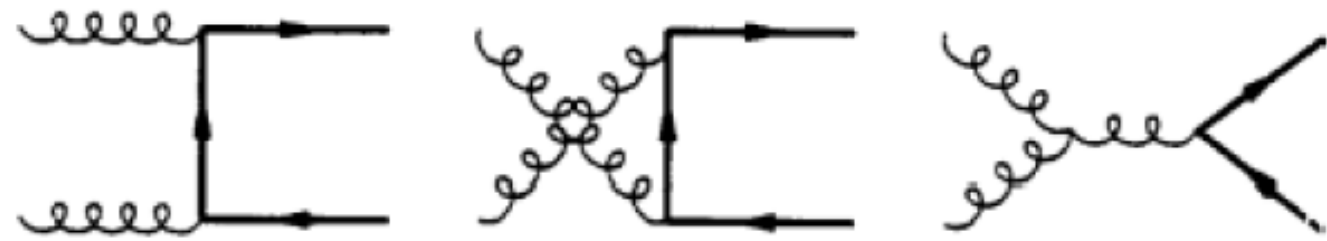


Central $t\bar{t}$ production at
the LHC probes gluon at high x

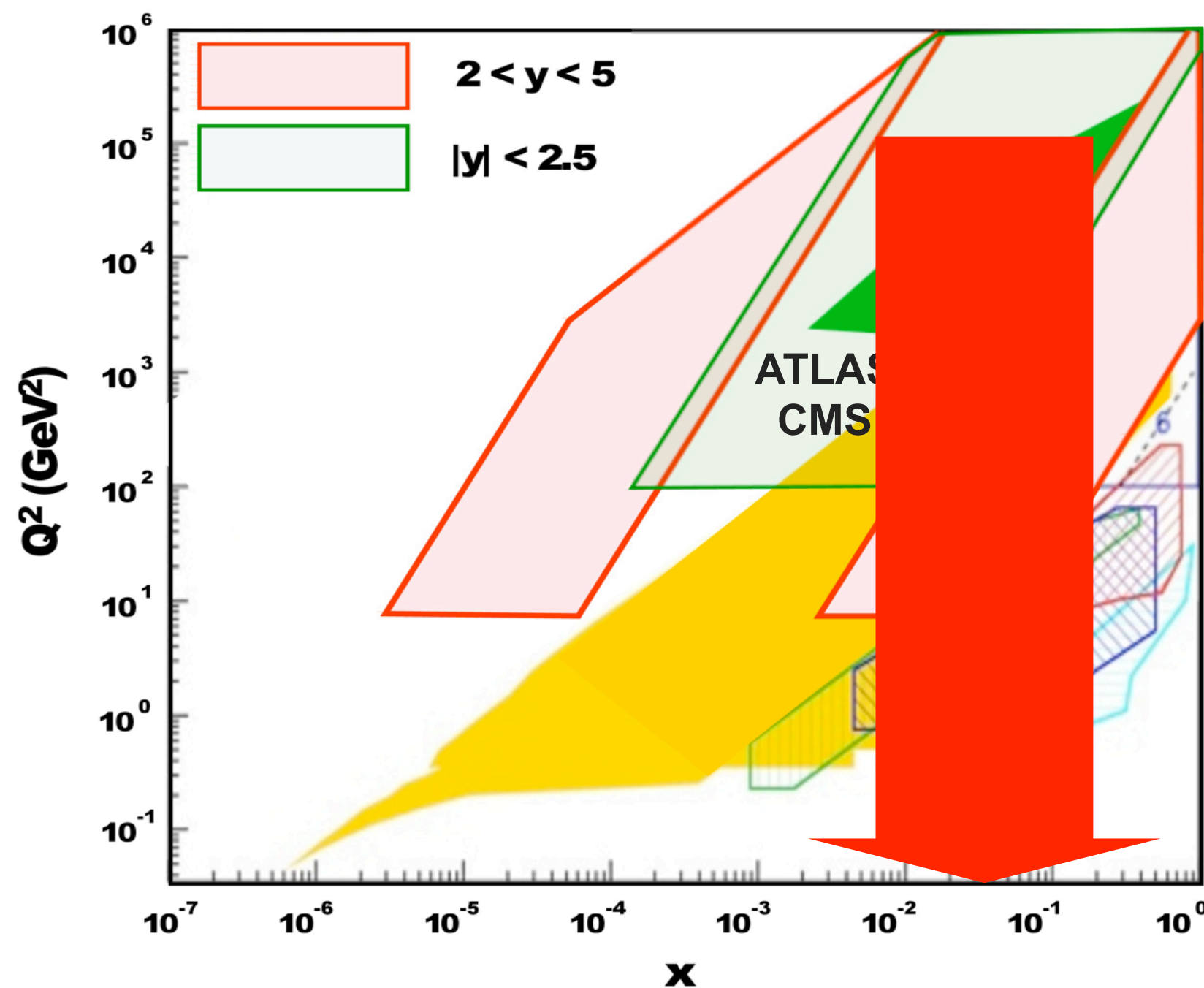


TOP QUARKS AT THE LHC: GLUON, m_Q , AND α_S

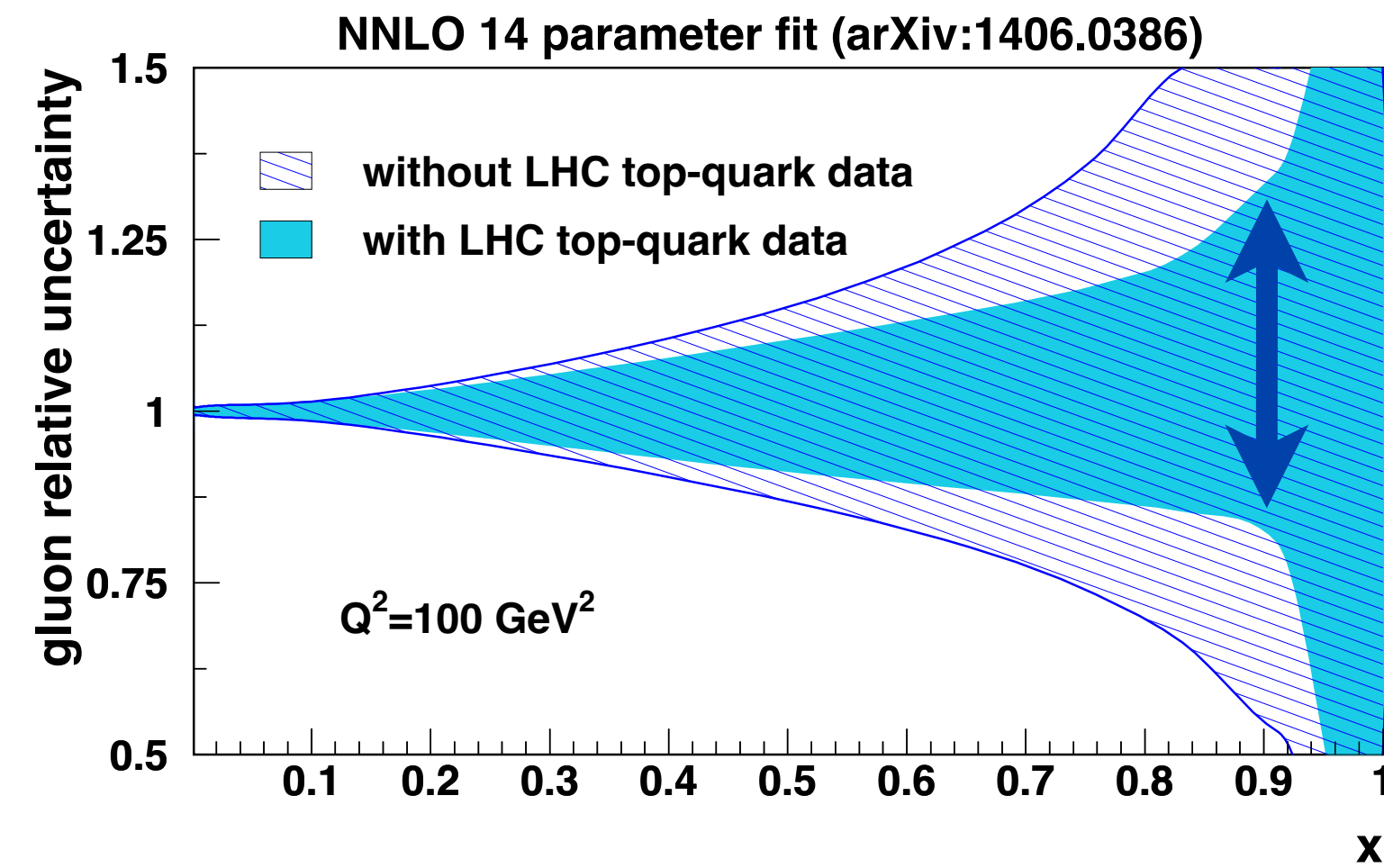
Heavy-quark production in pp :



Central $t\bar{t}$ production at
the LHC probes gluon at high x



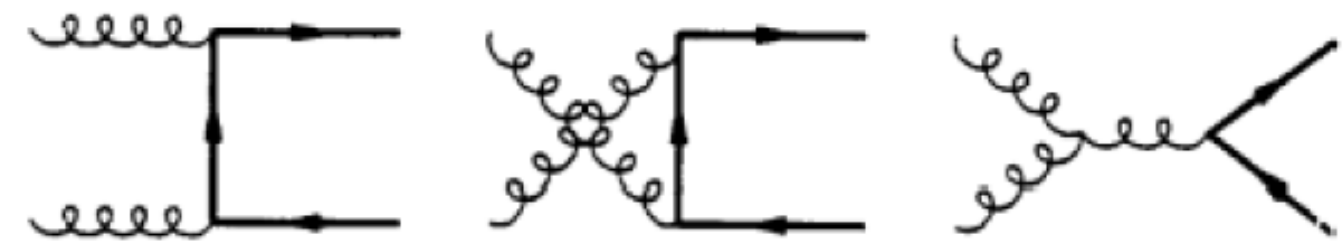
First QCD analysis @ NNLO using differential $t\bar{t}$: more precise $g(x)$



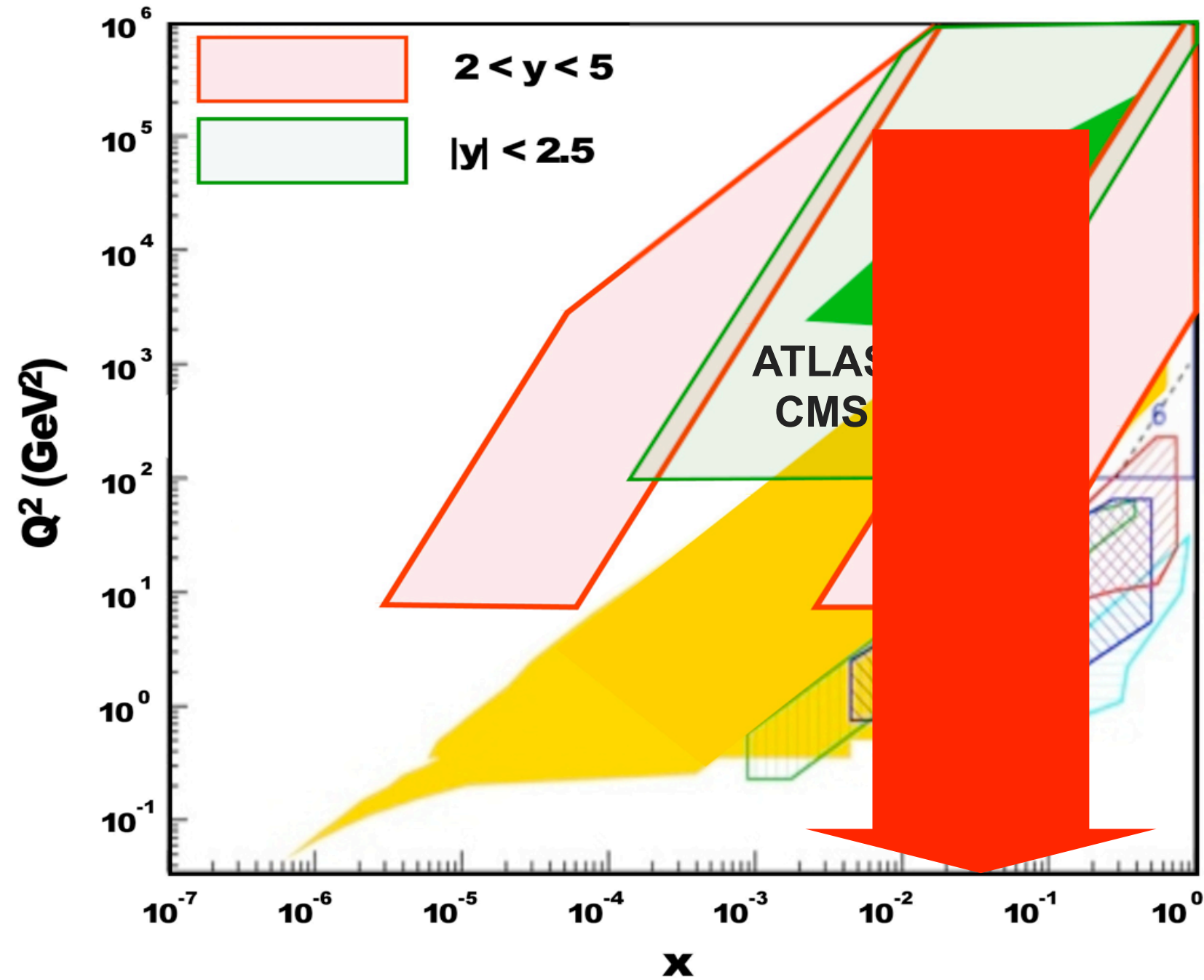
By now, usage of $t\bar{t}$ in
PDF fits well-established

TOP QUARKS AT THE LHC: GLUON, m_Q , AND α_S

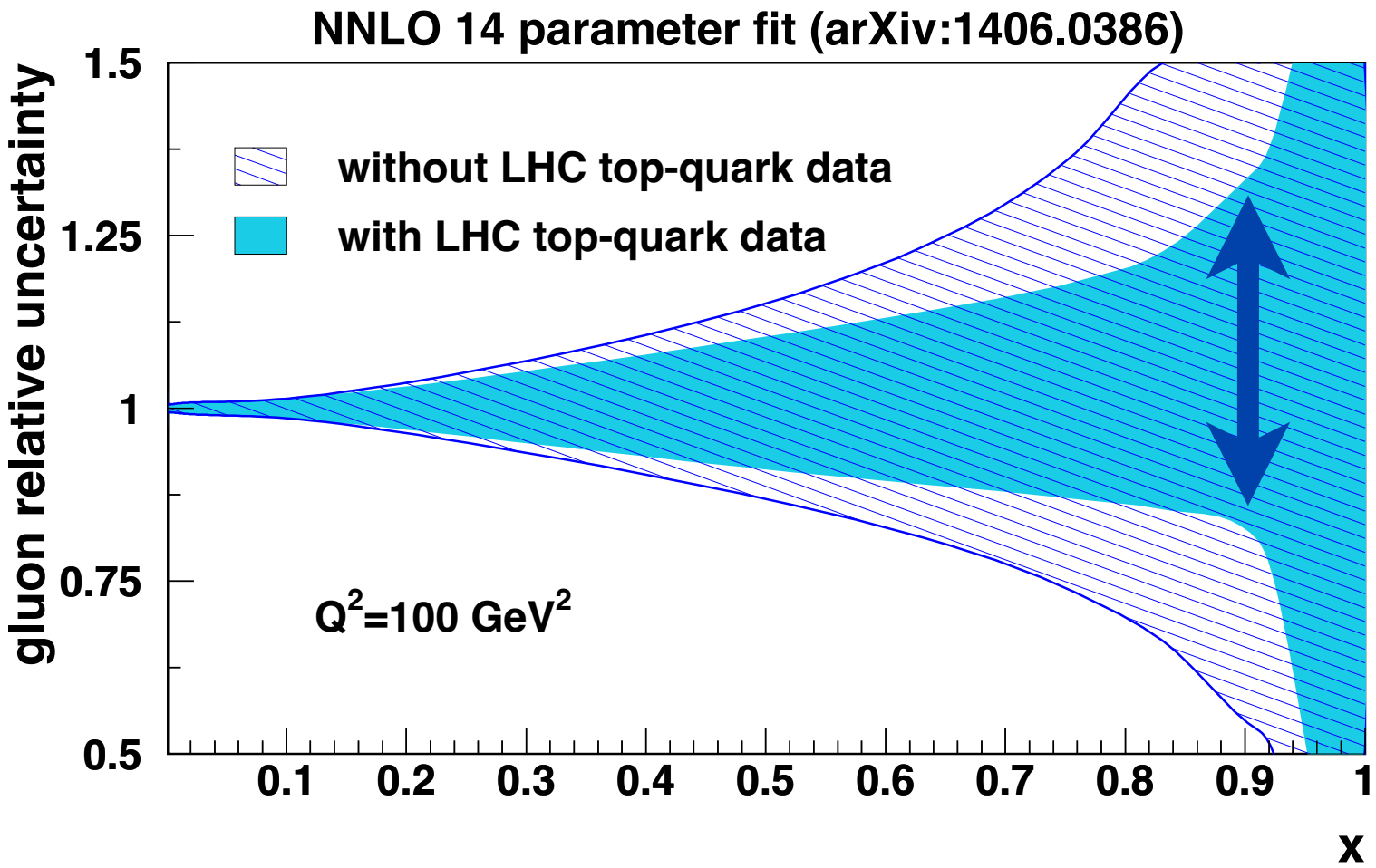
Problem: correlation of α_S , $g(x)$, m_t



Central $t\bar{t}$ production at the LHC probes gluon at high x

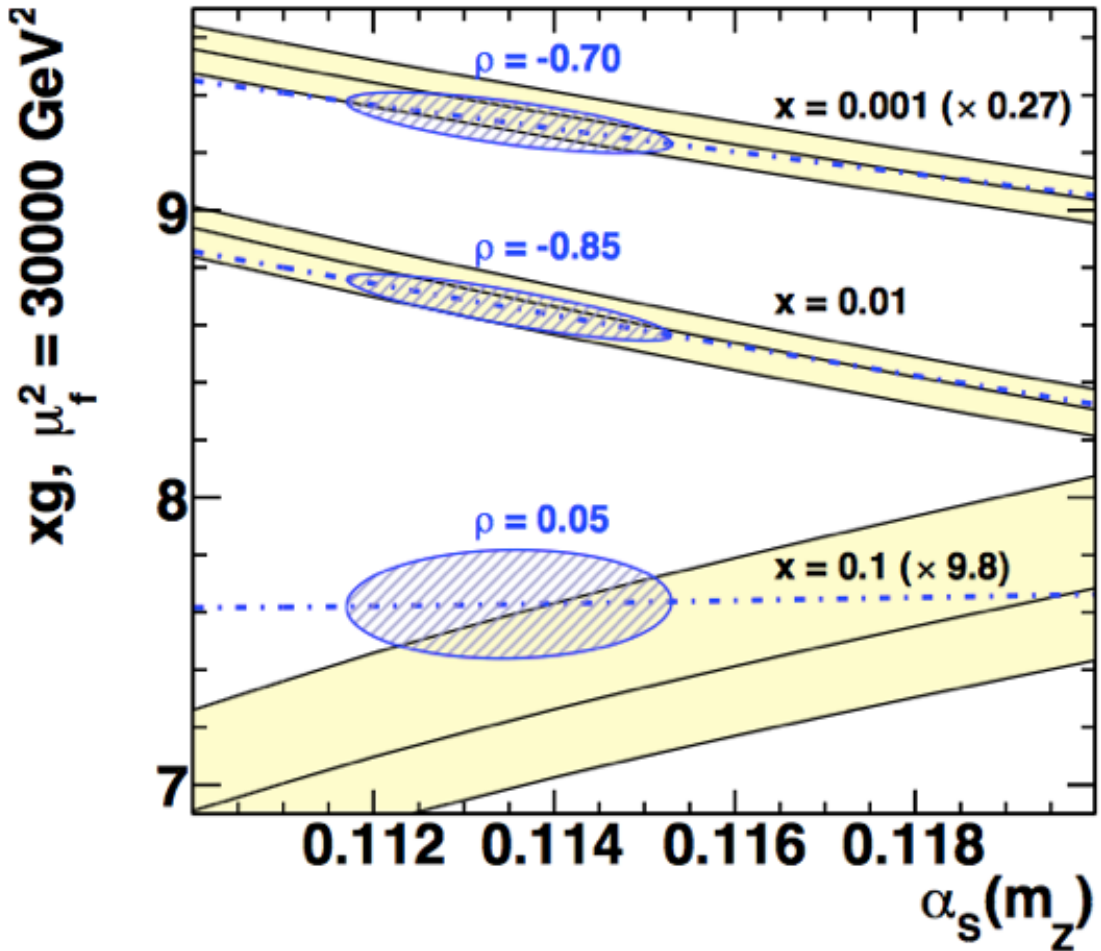
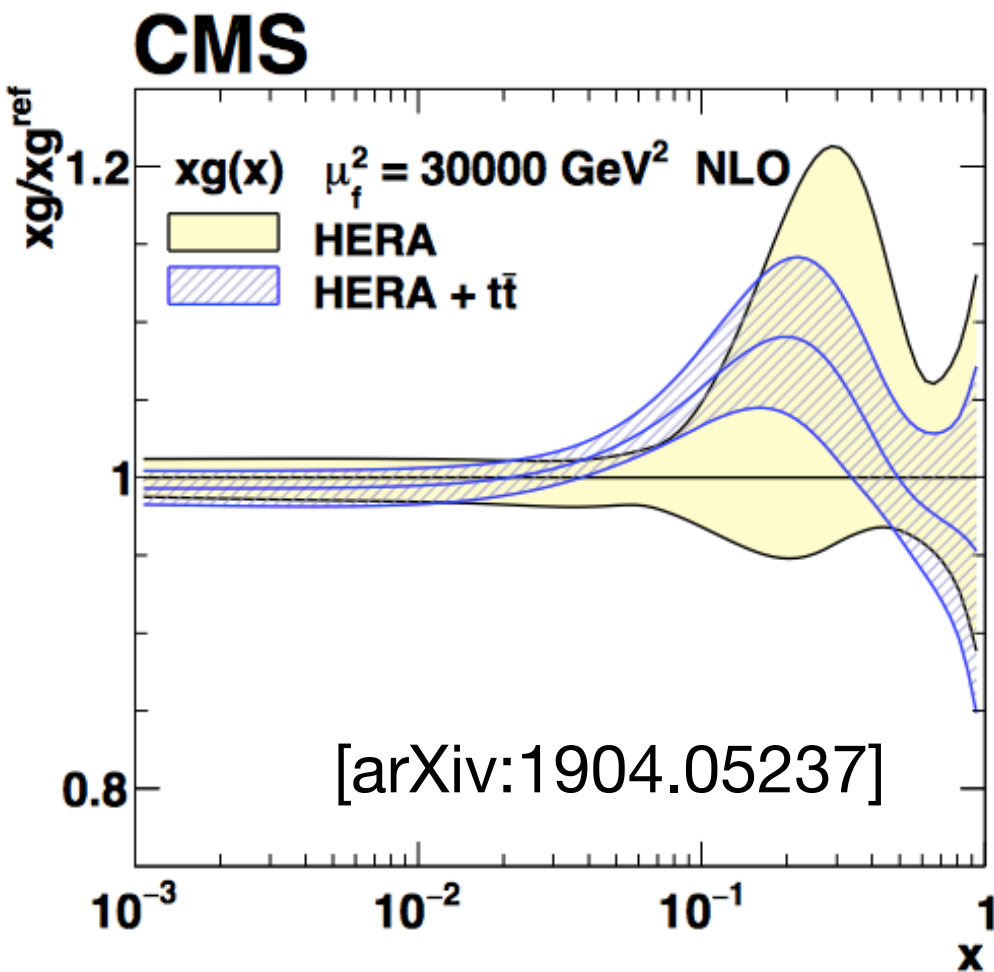


First QCD analysis @ NNLO using differential $t\bar{t}$: more precise $g(x)$



By now, usage of $t\bar{t}$ in PDF fits well-established

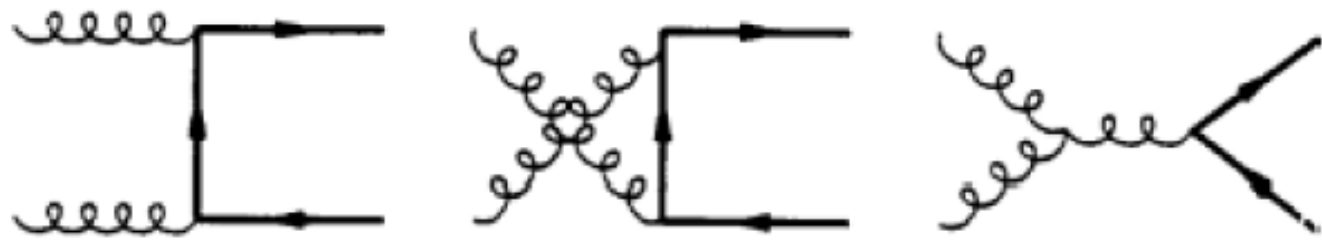
3-D $t\bar{t}$ cross sections: fit α_S , m_t , $g(x)$ together



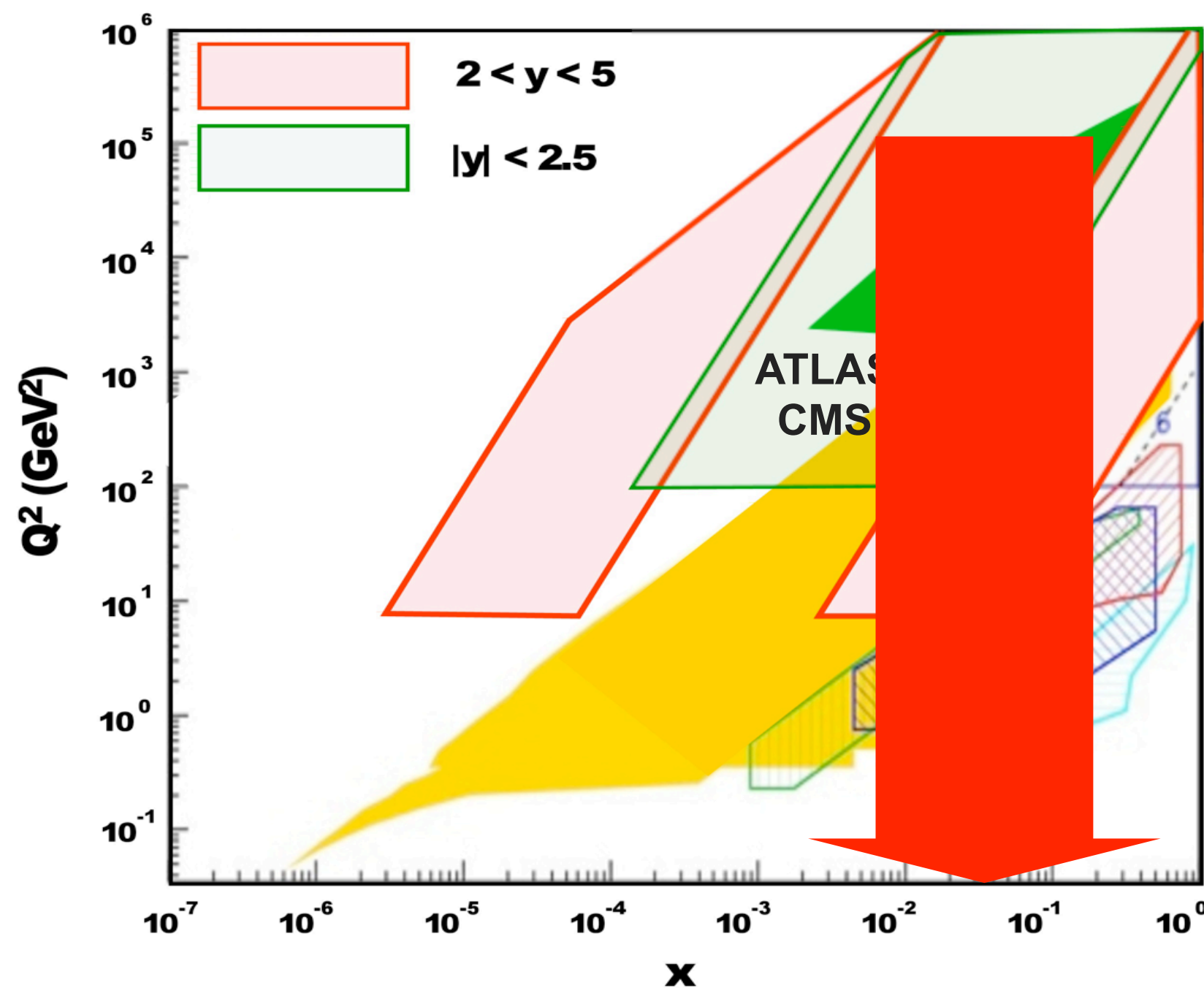
reduce parameter correlations!

TOP QUARKS AT THE LHC: GLUON, m_Q , AND α_S

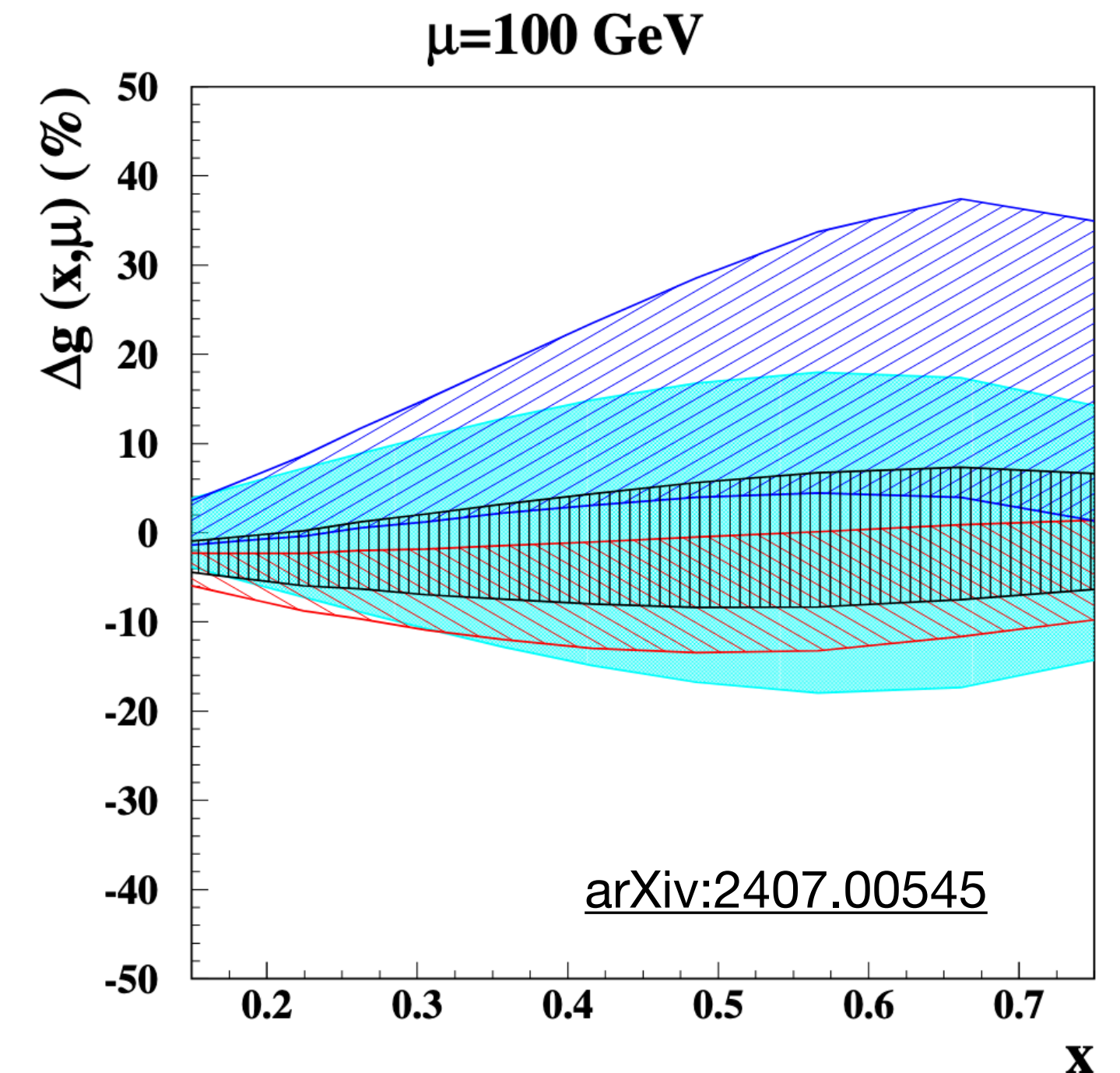
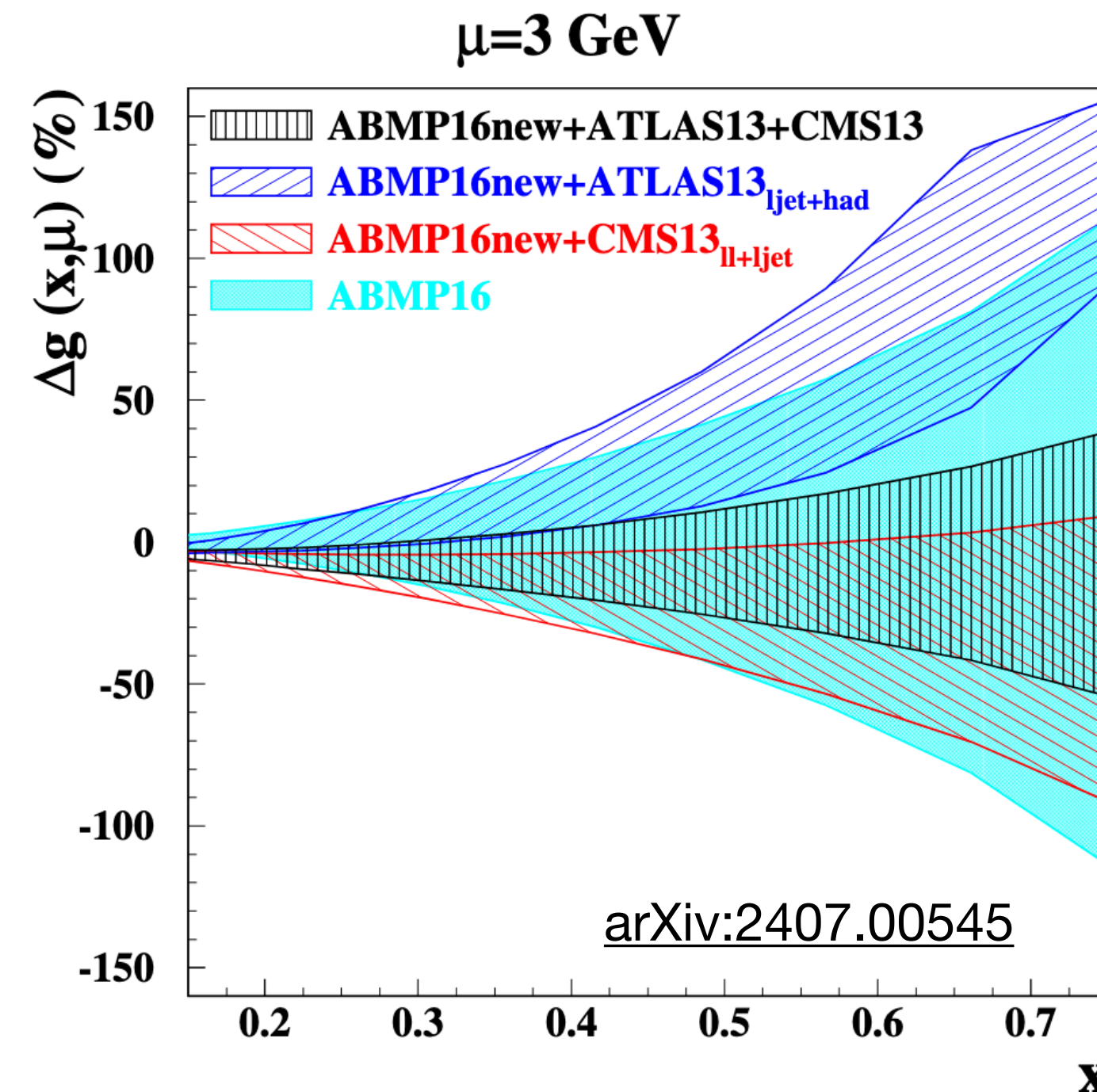
Problem: correlation of $\alpha_S, g(x), m_t$



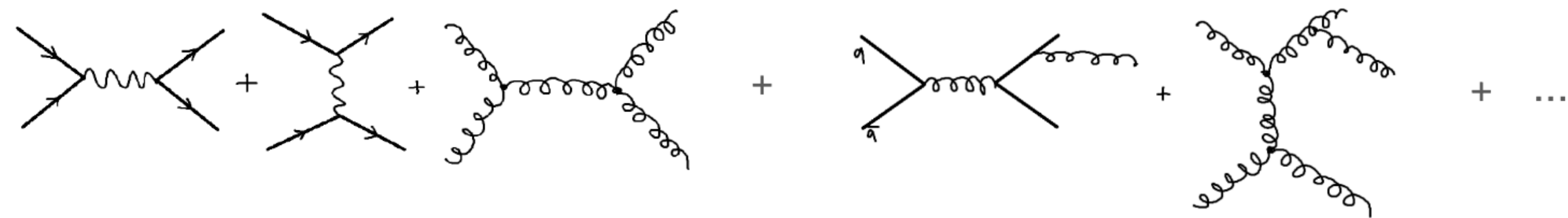
Central $t\bar{t}$ production at
the LHC probes gluon at high x



Most recent: most of LHC $t\bar{t}$ data included into ABMPtt fit

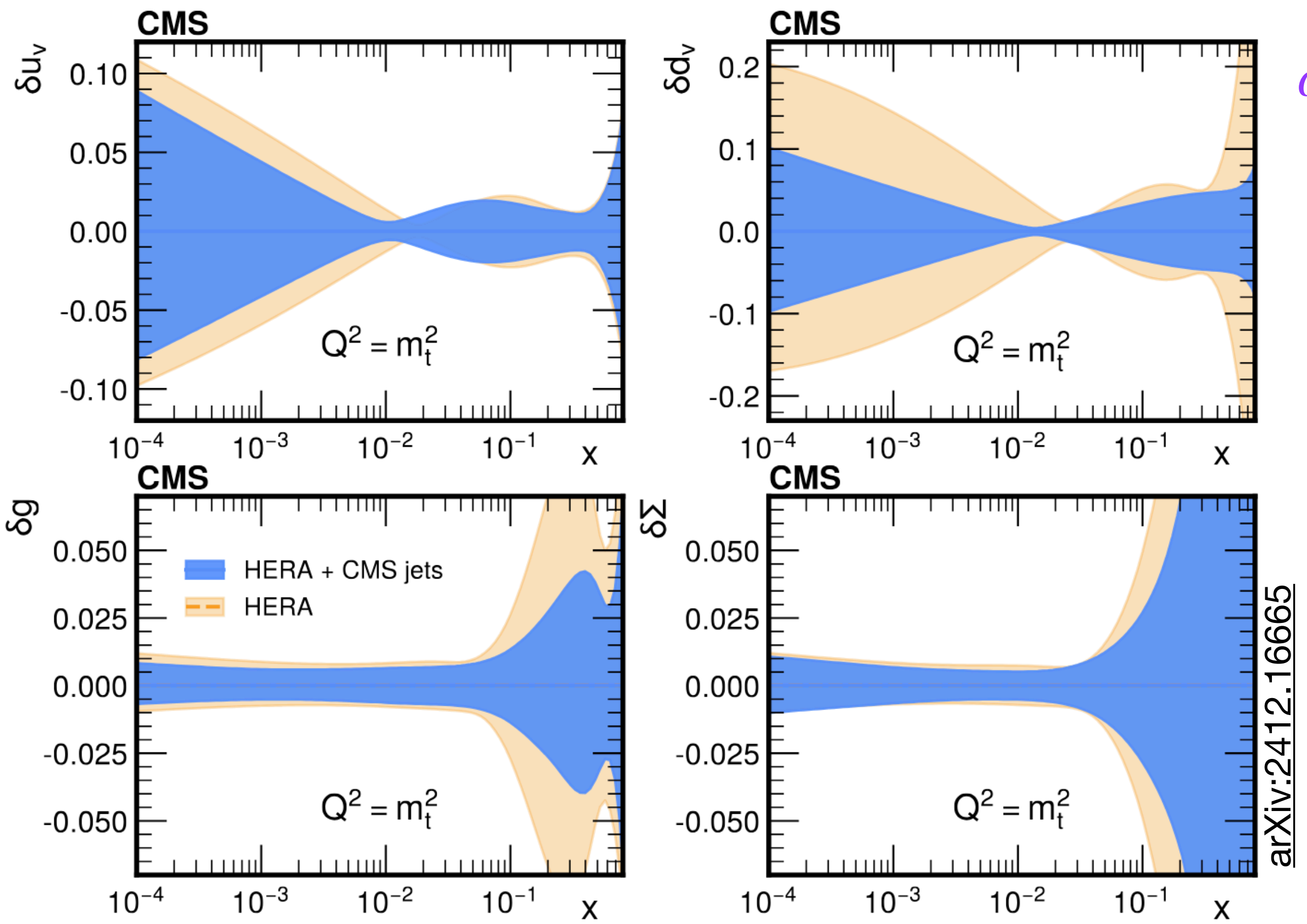


JETS AT THE LHC: DIRECT PROBE OF GLUON AND α_S



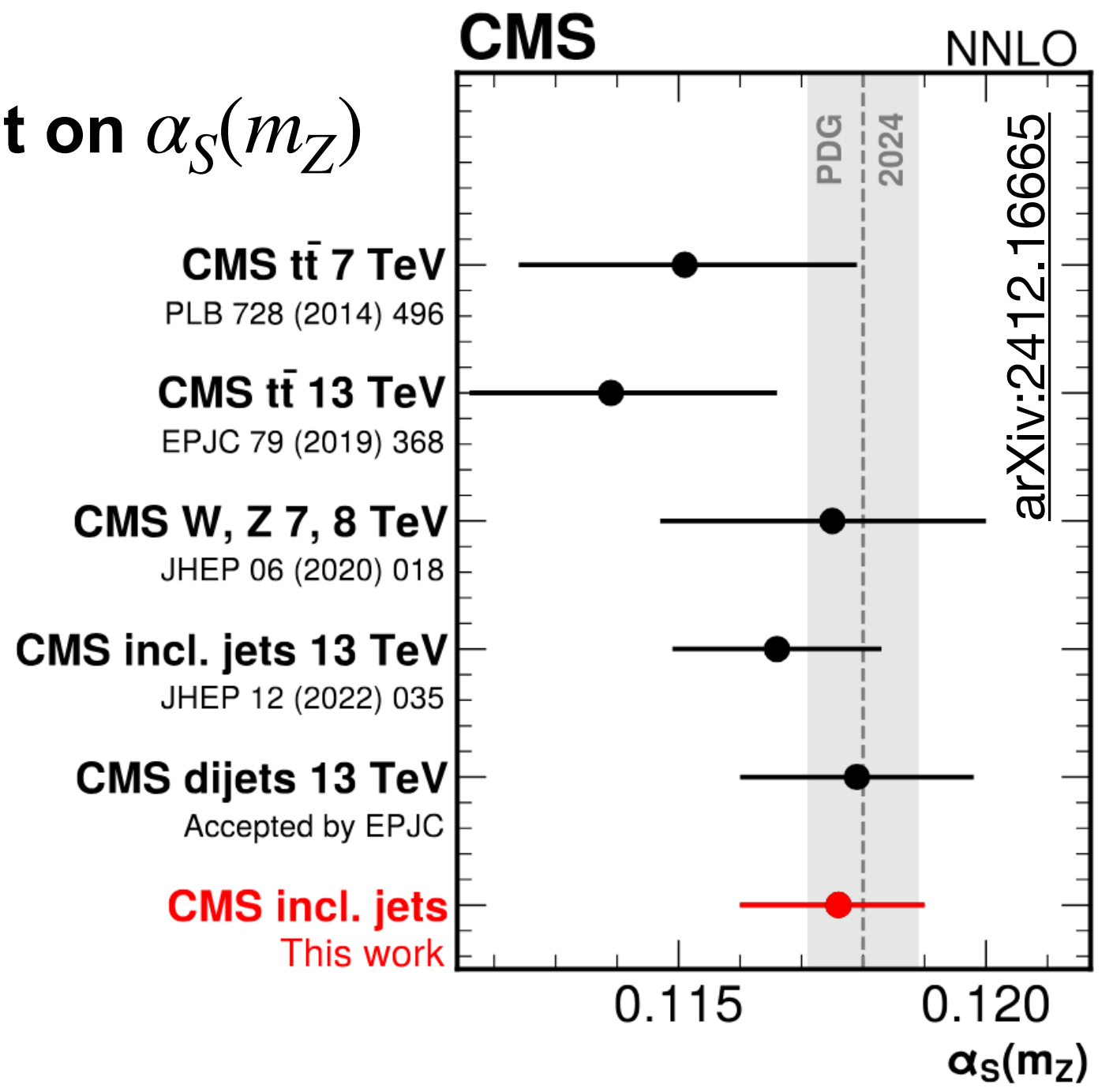
$\sqrt{s}$ [TeV]	$\mathcal{L}$ [fb ⁻¹]	N_p	p_T [GeV]	$ y $
2.76	0.0054	80	74–592	0.0–3.0
7	5.0	130	114–2116	0.0–2.5
8	20	165	74–1784	0.0–3.0
13	33.5	78	97–3103	0.0–2.0

Inclusive jet measurements at different $\sqrt{s}$: sensitivity to light quarks and gluon at NNLO!



α_S is fitted as a free parameter $\rightarrow$ correlation with PDF mitigated

best jet result on $\alpha_S(m_Z)$



arXiv:2412.16665

arXiv:2412.16665

STEP2: PARAMETRISATION

Select data and theory



Parametrize PDFs $f(x, \mu_0^2)$
at a starting scale μ_0^2



Perform a fit
(minimize χ^2)



Output PDFs $f(x, \mu^2)$

“From point of view of fitting, PDFs are one dimensional, x –dependent functions”

L. Harland-Lang, PDF4LHC2025

STEP2: PARAMETRISATION

Select data and theory



Parametrize PDFs $f(x, \mu_0^2)$
at a starting scale μ_0^2



Perform a fit
(minimize χ^2)



Output PDFs $f(x, \mu^2)$

generic form: $f_i(x, \mu_0^2) = A_i x^{B_i} (1 - x)^{C_i} \times [\dots]$

A : overall normalisation

B : small- x behaviour, C : shape for $x \rightarrow 1$

constraints:

- cross section must be positive
- $x \in [0, 1]$
- end-point $f(x = 1) = 0$
- integral constraints from Sum Rules (quark number, momentum)

STEP2: PARAMETRISATION

Select data and theory



Parametrize PDFs $f(x, \mu_0^2)$
at a starting scale μ_0^2



Perform a fit
(minimize χ^2)



Output PDFs $f(x, \mu^2)$

explicit form: $f_i(x, \mu_0^2) = A_i x^{B_i} (1 - x)^{C_i} \times P_i(x)$

HERAPDF: standard polynomials $P(x) = 1 + D \cdot x + E \cdot x^2 + F \cdot x^3$

ABMP: Bernstein polynomials $P(x) = P(y(x)) = \sum_{k=0}^{n_i} c_{i,k} \binom{n_i}{k} [y(x)]^k [1 - y(x)]^{n_i-k}$

$$\text{mapping function } y(x) = \frac{\ln(1/x) - \ln(1/x_{\max})}{\ln(1/x_{\min}) - \ln(1/x_{\max})}$$

CT/TEA: combination of splines and Bernstein polynomials

MSHT: Chebyshev polynomials $P(x) = P(y(x)) = 1 + \sum_{k=1}^{N_i} a_{i,k} T_k(y)$

$$\text{mapping function } y(x) = \frac{2 \ln(1/x) - \ln(1/x_{\max}) - \ln(1/x_{\min})}{\ln(1/x_{\max}) - \ln(1/x_{\min})}$$

STEP2: PARAMETRISATION

Select data and theory



Parametrize PDFs $f(x, \mu_0^2)$
at a starting scale μ_0^2



Perform a fit
(minimize χ^2)



Output PDFs $f(x, \mu^2)$

explicit form: $f_i(x, \mu_0^2) = A_i x^{B_i} (1-x)^{C_i} \times P_i(x)$

HERAPDF: standard polynomials $P(x) = 1 + D \cdot x + E \cdot x^2 + F \cdot x^3$

14 free parameters

ABMP: Bernstein polynomials $P(x) = P(y(x)) = \sum_{k=0}^{n_i} c_{i,k} \binom{n_i}{k} [y(x)]^k [1-y(x)]^{n_i-k}$

25 free parameters

$$\text{mapping function } y(x) = \frac{\ln(1/x) - \ln(1/x_{\max})}{\ln(1/x_{\min}) - \ln(1/x_{\max})}$$

CT/TEA: combination of splines and Bernstein polynomials

29 free parameters

MSHT: Chebyshev polynomials $P(x) = P(y(x)) = 1 + \sum_{k=1}^{N_i} a_{i,k} T_k(y)$

52 free parameters

$$\text{mapping function } y(x) = \frac{2 \ln(1/x) - \ln(1/x_{\max}) - \ln(1/x_{\min})}{\ln(1/x_{\max}) - \ln(1/x_{\min})}$$

STEP2: PARAMETRISATION

Select data and theory



Parametrize PDFs $f(x, \mu_0^2)$
at a starting scale μ_0^2



Perform a fit
(minimize χ^2)



Output PDFs $f(x, \mu^2)$

explicit form: $f_i(x, \mu_0^2) = A_i x^{B_i} (1 - x)^{C_i} \times P_i(x)$

flexibility is a trade-off:

- too few free parameters - bias
- too many free parameters : several local minima, bad convergence

different parameterisations must lead to similar results

STEP2: PARAMETRISATION

Select data and theory



Parametrize PDFs $f(x, \mu_0^2)$
at a starting scale μ_0^2



Perform a fit
(minimize χ^2)



Output PDFs $f(x, \mu^2)$

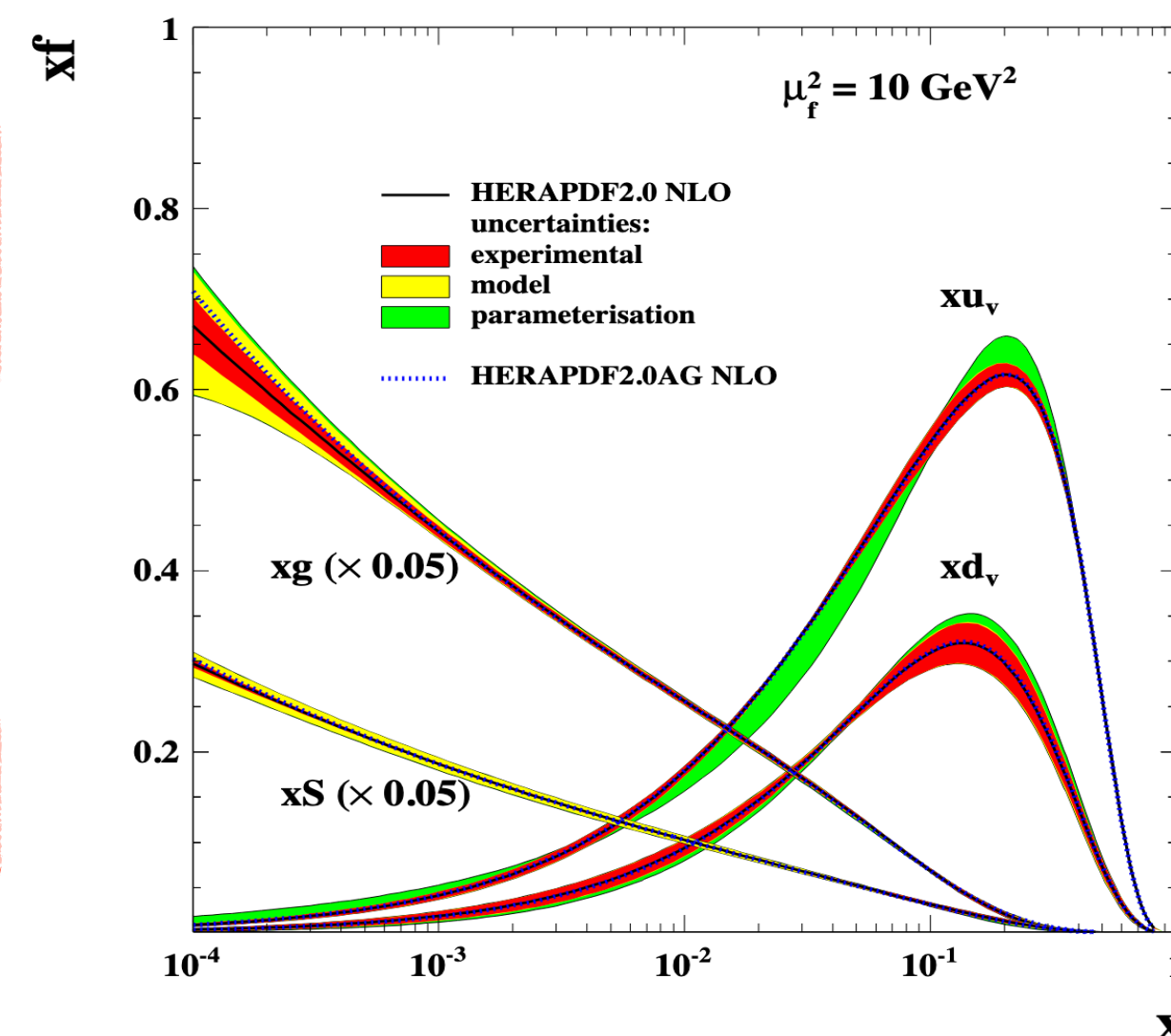
explicit form: $f_i(x, \mu_0^2) = A_i x^{B_i} (1-x)^{C_i} \times P_i(x)$

flexibility is a trade-off:

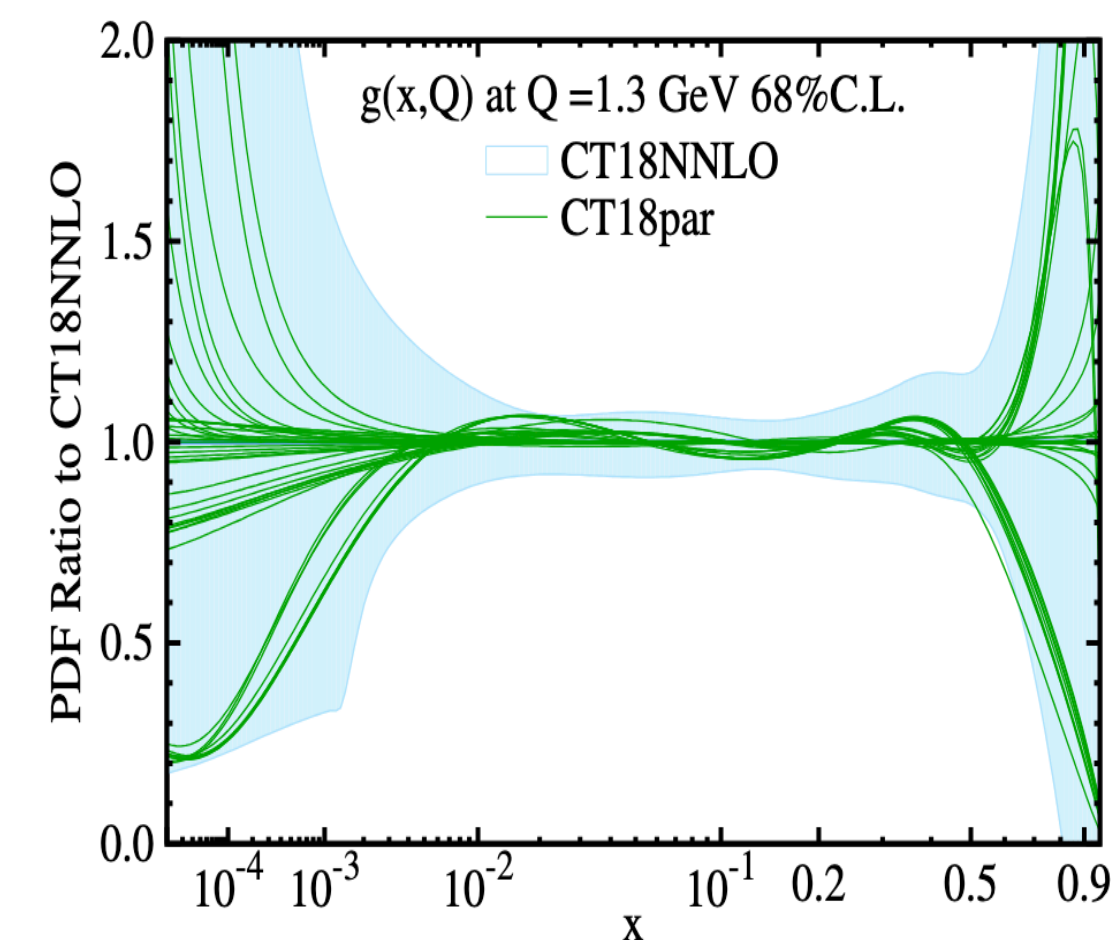
- too few free parameters - bias
- too many free parameters : several local minima, bad convergence

parametrisation uncertainty?

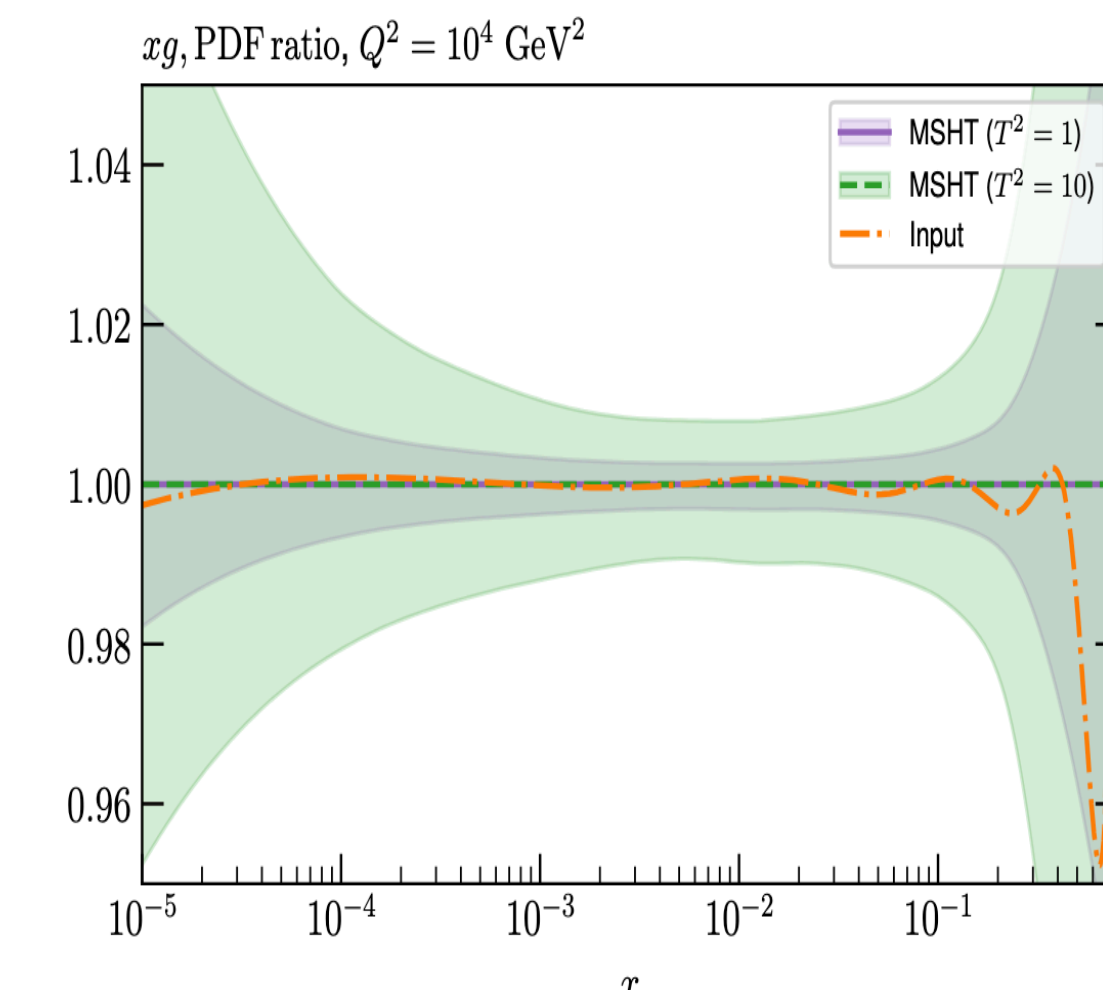
HERAPDF: extra uncertainty



CT18: implicit (envelope of alternative fits)



MSHT: closure tests, no extra uncertainty



L. Harland-Lang, PDF4LHC2025

STEP2: PARAMETRISATION

Select data and theory



Parametrize PDFs $f(x, \mu_0^2)$
at a starting scale μ_0^2



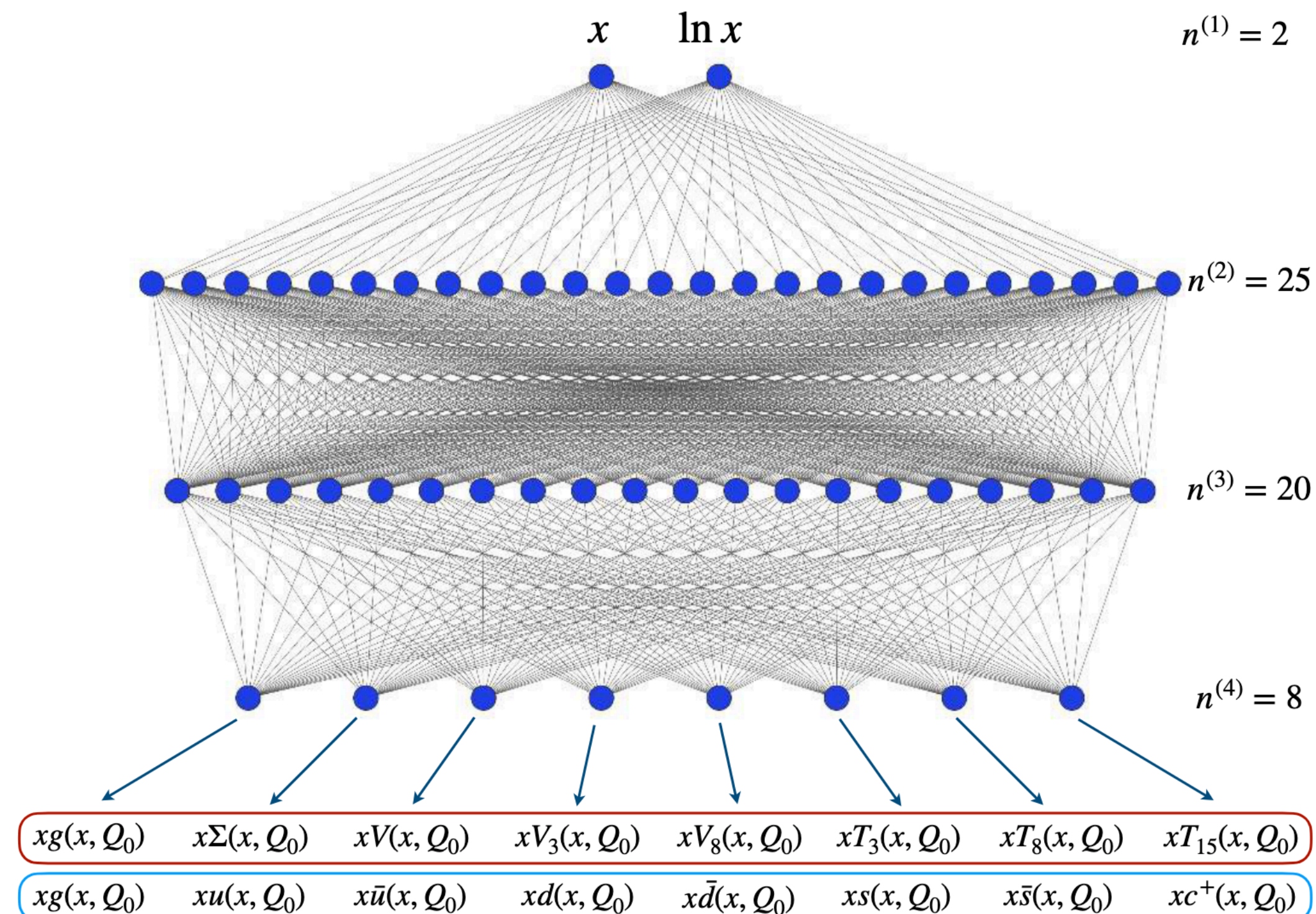
Perform a fit
(minimize χ^2)



Output PDFs $f(x, \mu^2)$

Neural Networks: $f_i(x, \mu_0^2) = A_i x^{B_i} (1-x)^{C_i} \times \text{NN}$ [NNPDF]

- MC sampling of data (generation of replicas of experimental data)
- training: set of PDFs parametrised by neural networks on each of the replicas
- validation: fit stops if fit quality stops improving (based on random selected validation data)



STEP2: PARAMETRISATION

Select data and theory



Parametrize PDFs $f(x, \mu_0^2)$
at a starting scale μ_0^2



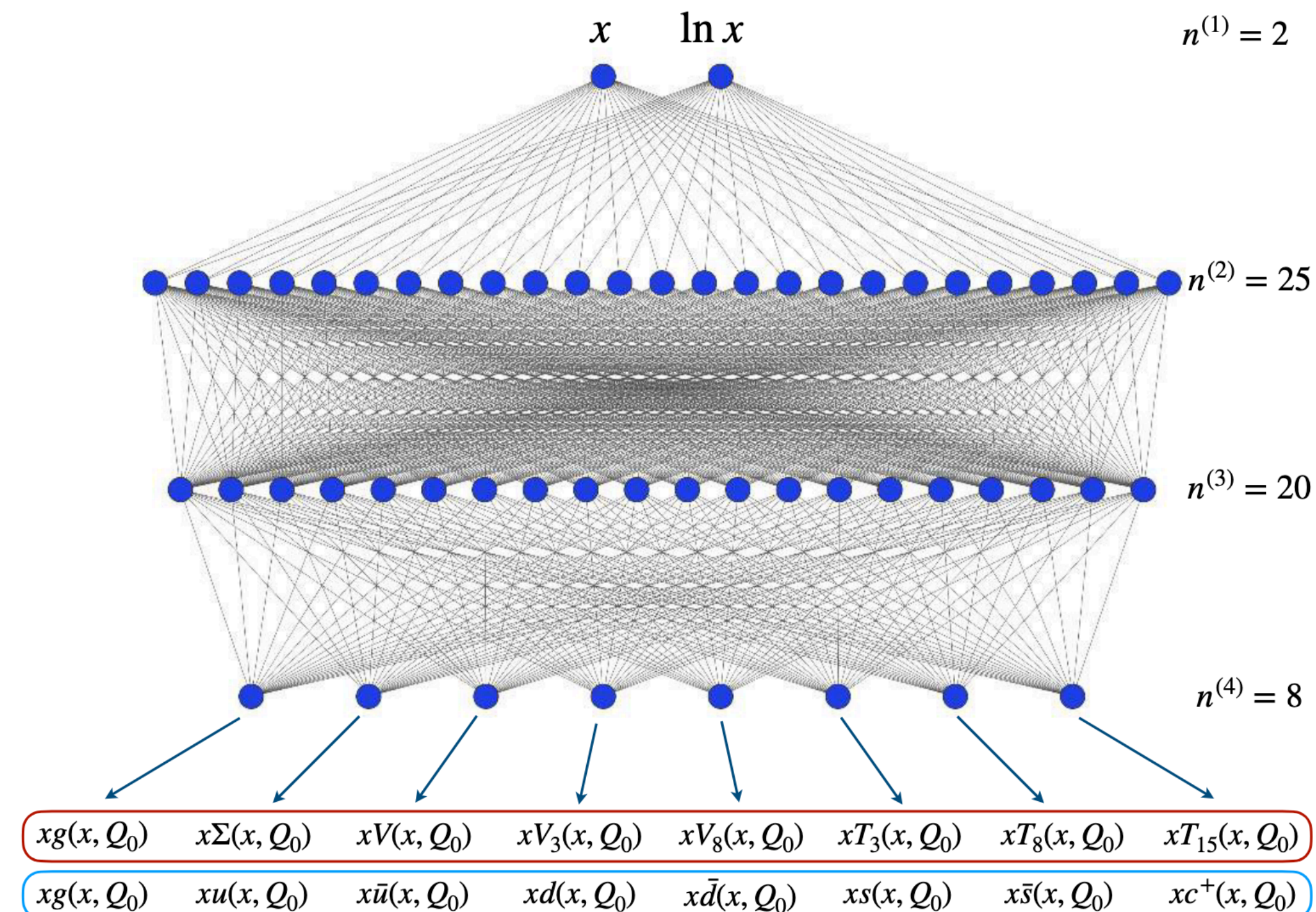
Perform a fit
(minimize χ^2)



Output PDFs $f(x, \mu^2)$

Neural Networks: $f_i(x, \mu_0^2) = A_i x^{B_i} (1 - x)^{C_i} \times \text{NN}$ [NNPDF]

- MC sampling of data (generation of replicas of experimental data)
- training: set of PDFs parametrised by neural networks on each of the replicas
- validation: fit stops if fit quality stops improving (based on random selected validation data)



763 free parameters
[NNPDF4.0]

STEP2: PARAMETRISATION

Select data and theory



Parametrize PDFs $f(x, \mu_0^2)$
at a starting scale μ_0^2



Perform a fit
(minimize χ^2)



Output PDFs $f(x, \mu^2)$

Neural Networks: $f_i(x, \mu_0^2) = A_i x^{B_i} (1 - x)^{C_i} \times \text{NN}$ [NNPDF]

more flexible (so less biased) parametrisation,
needs robust optimisation + stopping (avoid over/under-fitting)

parametrisation bias addressed via NN.

NN parameters chosen via NNPDF hypermoptimisation technique.

STEP 3 : MINIMISATION

Select data and theory



Parametrize PDFs $f(x, \mu_0^2)$
at a starting scale μ_0^2



Perform a fit
(minimize χ^2)



Output PDFs $f(x, \mu^2)$

minimise χ^2 , which measures difference of data and theory:

[example: HERAPDF]

$$\chi^2(T, b_{\text{exp}}) = \sum_{i=1}^{N_{\text{data}}} \frac{\left[\overset{\text{measurements}}{M_i} - \overset{\text{theory predictions}}{T_i \left(1 - \sum_j \gamma_{ij}^{\text{exp}} b_{j,\text{exp}} \right)} \right]^2}{\underset{\substack{\text{uncorrelated} \\ \text{systematic uncertainties}}}{\delta_{i,\text{uncor}}^2 T_i^2} + \underset{\substack{\text{statistical} \\ \text{uncertainties}}}{\delta_{i,\text{stat}}^2 M_i T_i \left(1 - \sum_j \gamma_{ij}^{\text{exp}} b_{j,\text{exp}} \right)}} + \sum_{j=1}^{N_{\text{exp.sys}}} \underset{\substack{\text{nuisance parameters} \\ \text{related to correlated} \\ \text{uncertainty sources}}}{b_{j,\text{exp}}^2}$$

impact of each source $b_{j,\text{exp}}$
on theory point T_i

STEP 3 : MINIMIZATION

Select data and theory



Parametrize PDFs $f(x, \mu_0^2)$
at a starting scale μ_0^2



Perform a fit
(minimize χ^2)



Output PDFs $f(x, \mu^2)$

minimise χ^2 , which measures difference of data and theory:

[example: HERAPDF]

$$\chi^2(T, b_{\text{exp}}) = \sum_{i=1}^{N_{\text{data}}} \frac{\left[\overset{\text{measurements}}{M_i} - \overset{\text{theory predictions}}{T_i \left(1 - \sum_j \gamma_{ij}^{\text{exp}} b_{j,\text{exp}} \right)} \right]^2}{\underset{\substack{\text{uncorrelated} \\ \text{systematic uncertainties}}}{\delta_{i,\text{uncor}}^2 T_i^2} + \underset{\substack{\text{statistical} \\ \text{uncertainties}}}{\delta_{i,\text{stat}}^2 M_i T_i \left(1 - \sum_j \gamma_{ij}^{\text{exp}} b_{j,\text{exp}} \right)}} + \sum_{j=1}^{N_{\text{exp.sys}}} \underset{\substack{\text{nuisance parameters} \\ \text{related to correlated} \\ \text{uncertainty sources}}}{b_{j,\text{exp}}^2}$$

impact of each source $b_{j,\text{exp}}$ on theory point T_i

**Alternatively (equivalent),
covariance matrix is used:**

$$\chi^2 = \frac{1}{2} (T - M)^T C^{-1} (T - M)$$

STEP 3 : MINIMIZATION

Select data and theory



Parametrize PDFs $f(x, \mu_0^2)$
at a starting scale μ_0^2



Perform a fit
(minimize χ^2)



Output PDFs $f(x, \mu^2)$

minimise χ^2 , which measures difference of data and theory:

[example: HERAPDF]

$$\chi^2(T, b_{\text{exp}}) = \sum_{i=1}^{N_{\text{data}}} \frac{\left[\overset{\text{measurements}}{M_i} - \overset{\text{theory predictions}}{T_i \left(1 - \sum_j \gamma_{ij}^{\text{exp}} b_{j,\text{exp}} \right)} \right]^2}{\underset{\substack{\text{uncorrelated} \\ \text{systematic uncertainties}}}{\delta_{i,\text{uncor}}^2 T_i^2} + \underset{\substack{\text{statistical} \\ \text{uncertainties}}}{\delta_{i,\text{stat}}^2 M_i T_i \left(1 - \sum_j \gamma_{ij}^{\text{exp}} b_{j,\text{exp}} \right)}} + \sum_{j=1}^{N_{\text{exp.sys}}} \underset{\substack{\text{nuisance parameters} \\ \text{related to correlated} \\ \text{uncertainty sources}}}{b_{j,\text{exp}}^2}$$

impact of each source $b_{j,\text{exp}}$
on theory point T_i

- **iterative procedure**
- **fast interpolation grids** ← predictions T are called at each step of minimisation !

STEP 3 : MINIMIZATION

Select data and theory



Parametrize PDFs $f(x, \mu_0^2)$
at a starting scale μ_0^2



Perform a fit
(minimize χ^2)



Output PDFs $f(x, \mu^2)$

minimise χ^2 , which measures difference of data and theory:

[example: HERAPDF]

$$\chi^2(T, b_{\text{exp}}) = \sum_{i=1}^{N_{\text{data}}} \frac{\left[\overset{\text{measurements}}{M_i} - \overset{\text{theory predictions}}{T_i \left(1 - \sum_j \gamma_{ij}^{\text{exp}} b_{j,\text{exp}} \right)} \right]^2}{\underset{\substack{\text{uncorrelated} \\ \text{systematic uncertainties}}}{\delta_{i,\text{uncor}}^2 T_i^2} + \underset{\substack{\text{statistical} \\ \text{uncertainties}}}{\delta_{i,\text{stat}}^2 M_i T_i \left(1 - \sum_j \gamma_{ij}^{\text{exp}} b_{j,\text{exp}} \right)}} + \sum_{j=1}^{N_{\text{exp.sys}}} \underset{\substack{\text{nuisance parameters} \\ \text{related to correlated} \\ \text{uncertainty sources}}}{b_{j,\text{exp}}^2}$$

impact of each source $b_{j,\text{exp}}$
on theory point T_i

minimum $\chi^2 \rightarrow$ output PDFs

QCD ANALYSIS: BASIC IDEA

Select data and theory



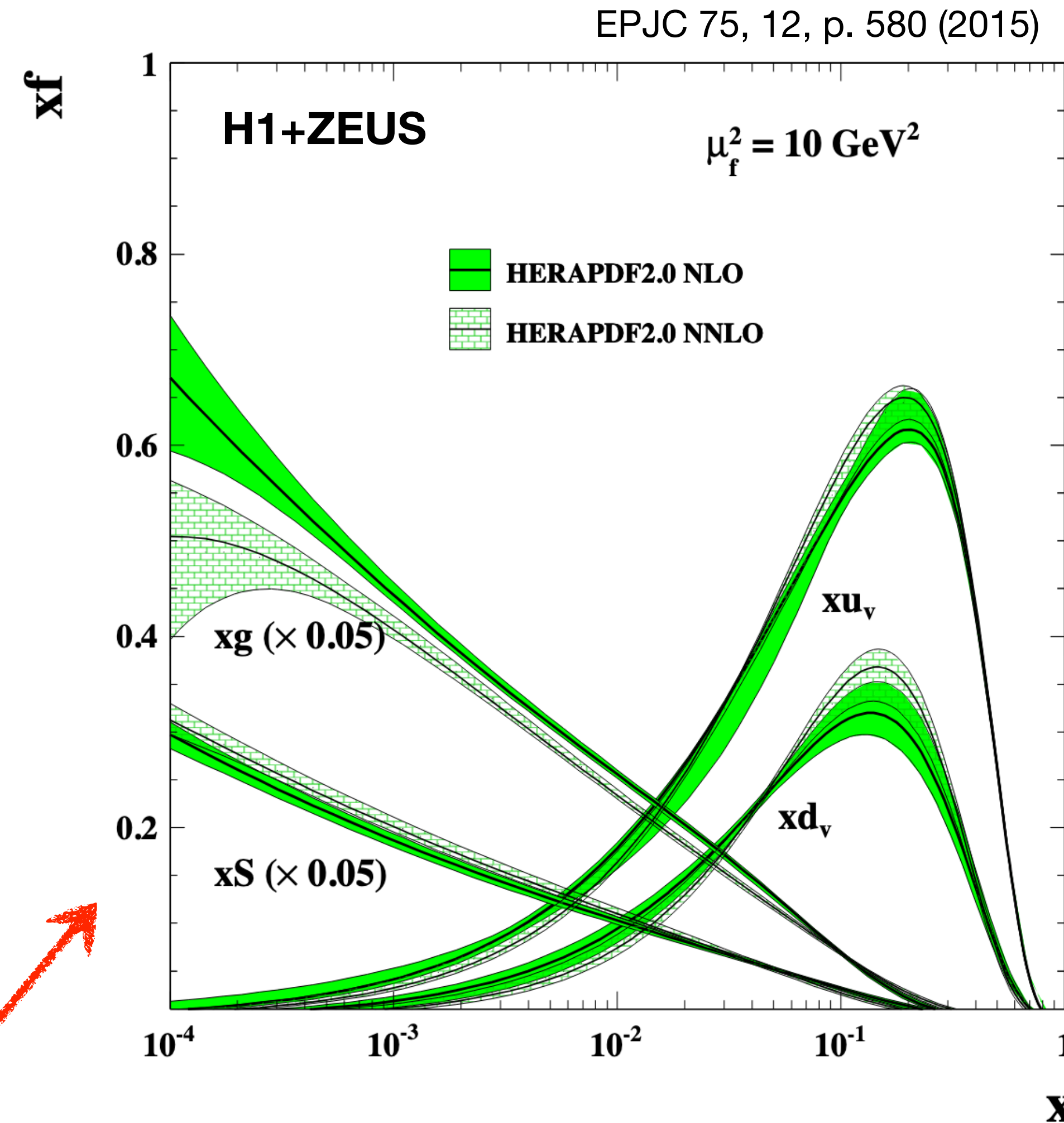
Parametrize PDFs $f(x, \mu_0^2)$
at a starting scale μ_0^2



Perform a fit
(minimize χ^2)



Output PDFs $f(x, \mu^2)$



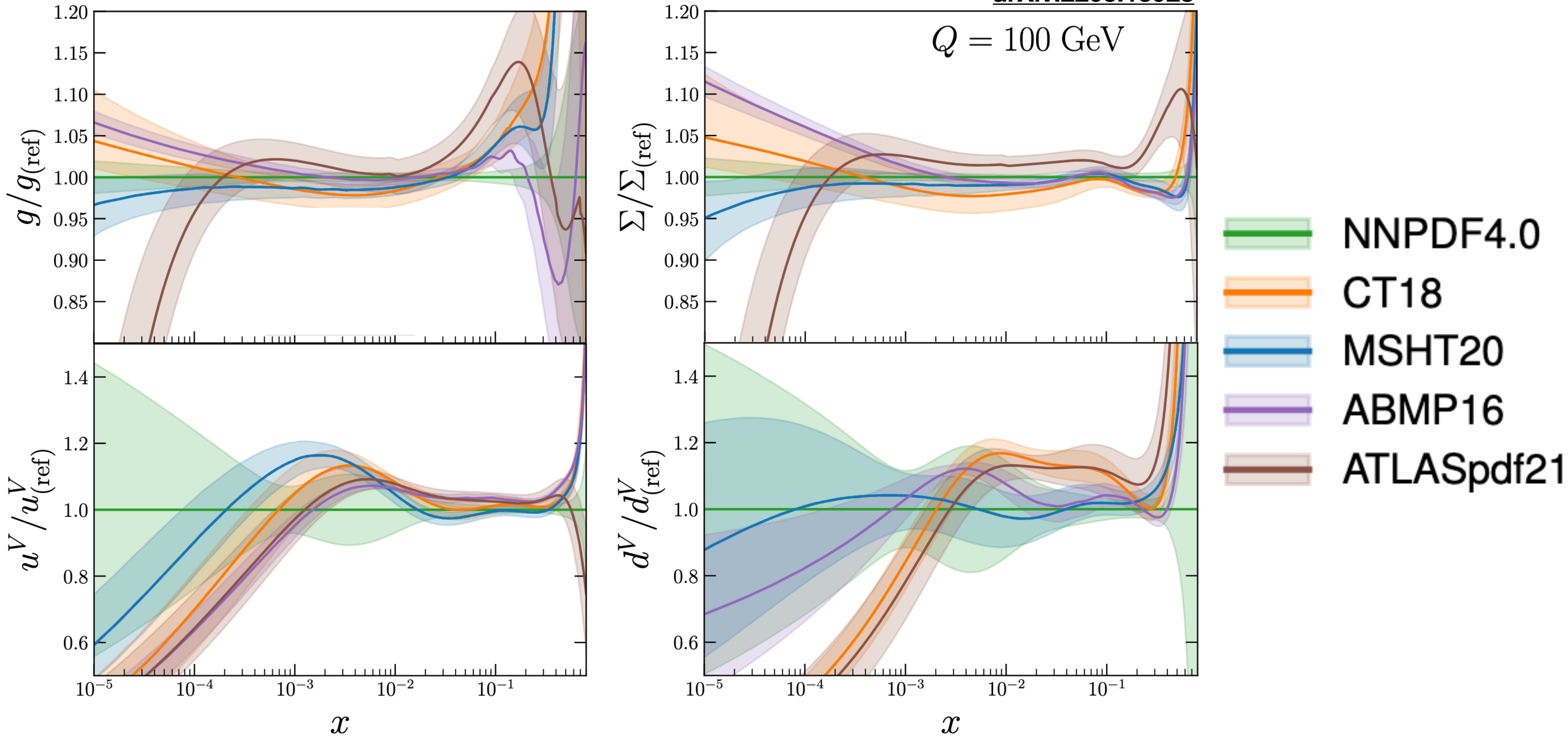
Example: only HERA data used

- gluon and sea dominant @low x
- PDFs at different orders differ

PDFs FROM DIFFERENT GROUPS [as a ratio to NNPDF4.0]

arXiv:2203.13923

$Q = 100 \text{ GeV}$



PDF UNCERTAINTIES

Originate from uncertainties in the experimental measurements and theoretical approximations

Method 1 (J. Pumplin et al) : Hessian uncertainties based on expansion of χ^2 , assuming

(near the best fit point (global minimum χ_0^2)

- χ^2 function is approximately quadratic
- the first derivatives are zero (since at a minimum)
- the behaviour is dominated by the second derivatives (the Hessian matrix)

used by ABMP, CT, HERAPDF, MHST

PDF UNCERTAINTIES

Originate from uncertainties in the experimental measurements and theoretical approximations

Method 1 (J. Pumplin et al) : Hessian uncertainties based on expansion of χ^2

change of χ^2 from the
global minimum

$$\Delta\chi^2 = \chi^2 - \chi_0^2 = \sum_{i=1}^{N_p} \sum_{j=1}^{N_p} H_{ij} (a_i - a_i^{(0)}) (a_j - a_j^{(0)})$$

number of free
parameters in the fit

PDF parameters

(e.g. A_i, B_i, C_i in $xf(x) = Ax^B(1 - x^C)P(x)$)

PDF parameters at the global minimum

elements of Hessian matrix
(the matrix of second derivatives)

$$H_{ij} = \frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_i \partial a_j} \bigg|_{a=a^{(0)}}$$

PDF UNCERTAINTIES

Originate from uncertainties in the experimental measurements and theoretical approximations

Method 1 (J. Pumplin et al) : Hessian uncertainties based on expansion of χ^2

change of χ^2 from the global minimum

number of free parameters in the fit

PDF parameters
(e.g. A_i, B_i, C_i in $xf(x) = Ax^B(1 - x^C)P(x)$)

PDF parameters at the global minimum

elements of Hessian matrix
(the matrix of second derivatives)

$$\Delta\chi^2 = \chi^2 - \chi_0^2 = \sum_{i=1}^{N_p} \sum_{j=1}^{N_p} H_{ij} (a_i - a_i^{(0)}) (a_j - a_j^{(0)})$$
$$H_{ij} = \frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_i \partial a_j} \bigg|_{a=a^{(0)}}$$

Interpretation: if PDF parameters move away from their best-fit values, fit quality gets worse, which can be calculated using the curvature information stored in the Hessian matrix

PDF UNCERTAINTIES

Originate from uncertainties in the experimental measurements and theoretical approximations

Method 1 (J. Pumplin et al) : Hessian uncertainties based on expansion of χ^2

change of χ^2 from the
global minimum

$$\Delta\chi^2 = \chi^2 - \chi_0^2 = \sum_{i=1}^{N_p} \sum_{j=1}^{N_p} H_{ij}(a_i - a_i^{(0)})(a_j - a_j^{(0)}) \quad H_{ij} = \frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_i \partial a_j} \Big|_{a=a^{(0)}}$$

In practice:

- Hessian matrix is diagonalised to find eigenvector directions,
- Eigenvalues determine how "steep" χ^2 rises in each direction,
- PDF uncertainty (eigenvector) sets are generated by moving along the eigenvector directions

until $\Delta\chi^2$ reaches a certain Tolerance $T = \sqrt{\Delta\chi_{global}^2}$

PDF UNCERTAINTIES

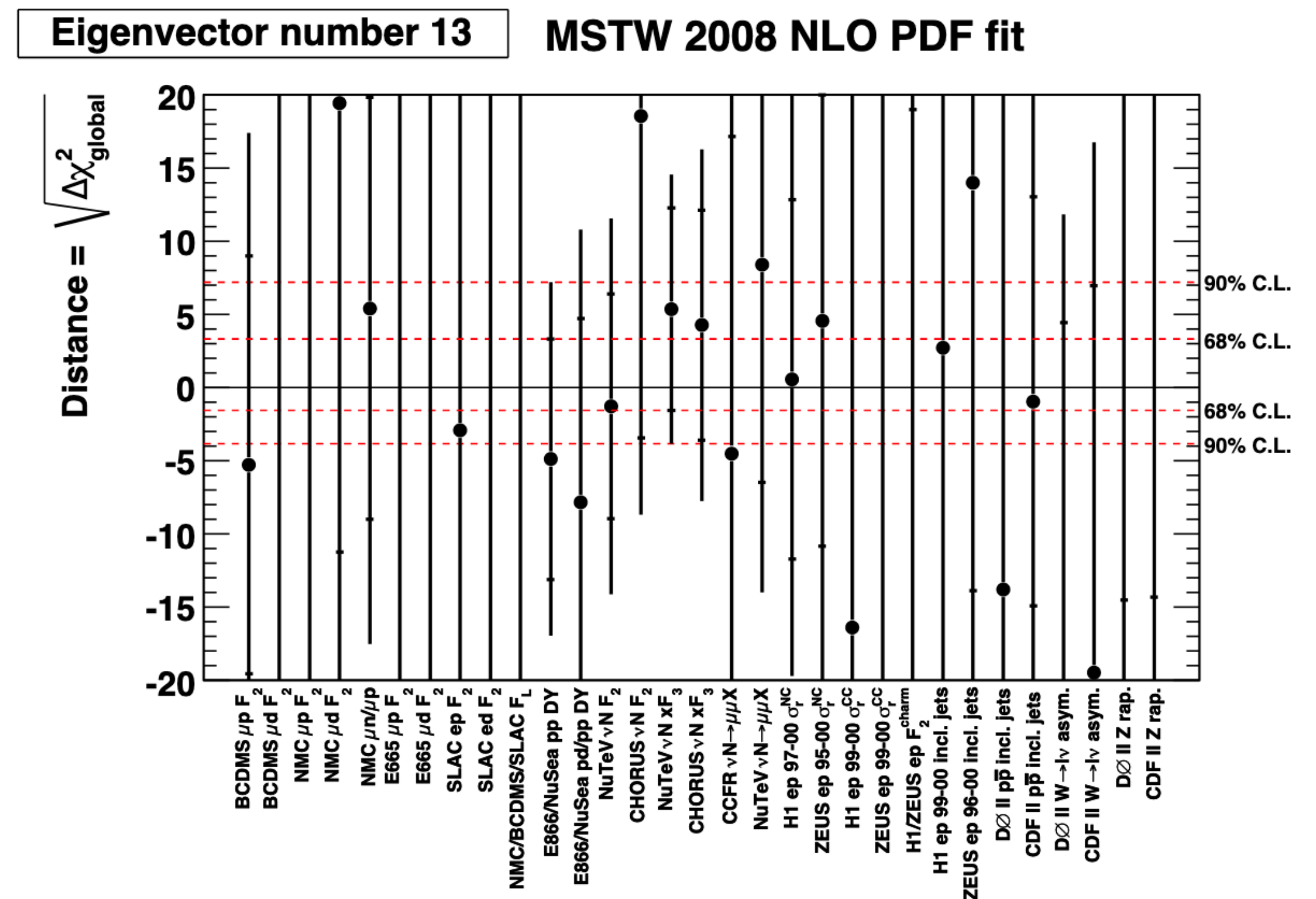
Originate from uncertainties in the experimental measurements and theoretical approximations

Method 1 (J. Pumplin et al) : Hessian uncertainties based on expansion of χ^2

[arXiv:0901.0002](https://arxiv.org/abs/0901.0002)

Ideally, Tolerance is $\Delta\chi^2 = \pm 1$ for 68% CL

- appropriate if fitting consistent data sets with ideal Gaussian errors to a well-defined theory
- not always the case due to not fully compatible data sets
[missing correlations of uncertainties, underestimated uncertainties/correlations, non-Gaussian uncertainties...]



PDF UNCERTAINTIES

Originate from uncertainties in the experimental measurements and theoretical approximations

Method 1 (J. Pumplin et al) : Hessian uncertainties based on expansion of χ^2

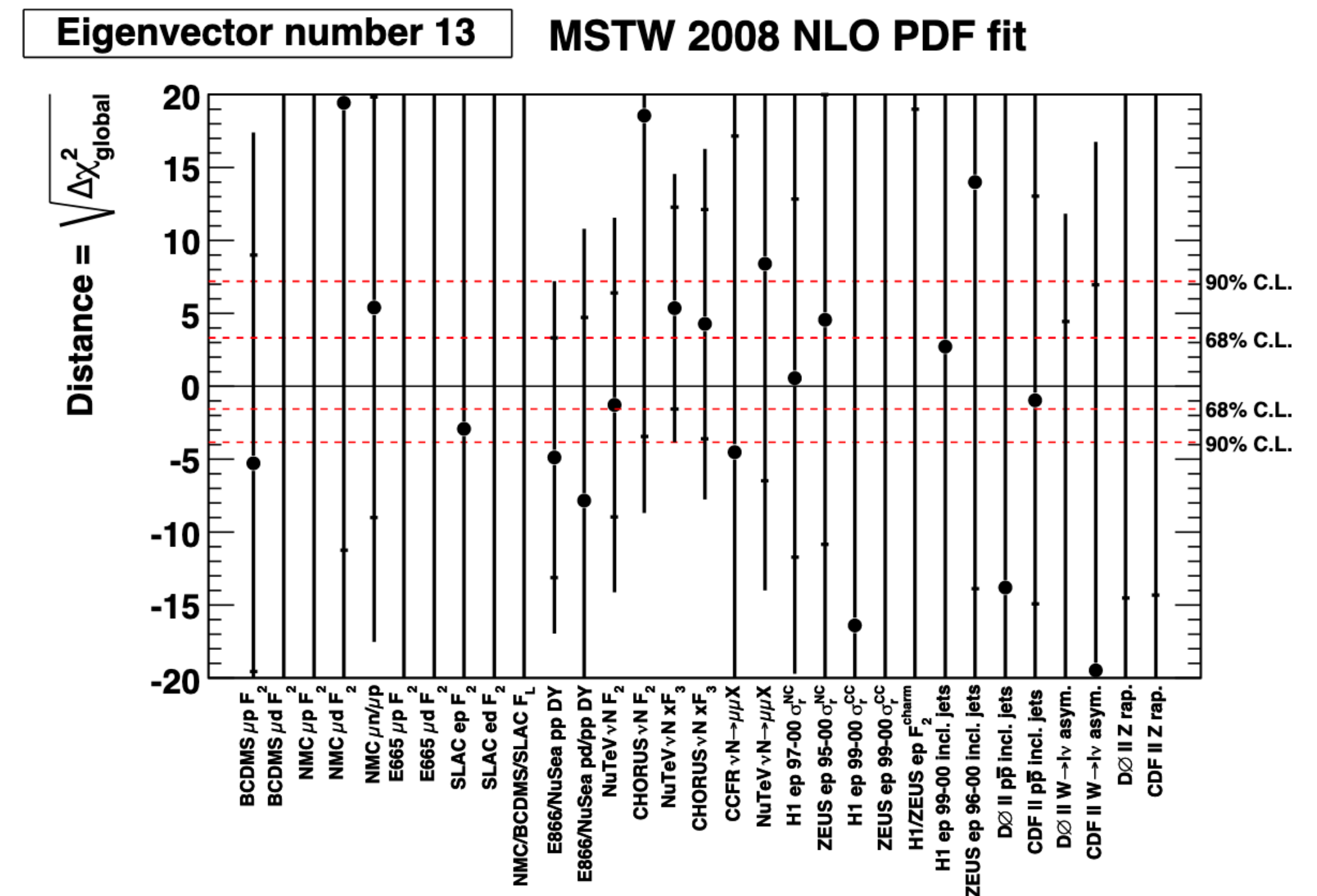
[arXiv:0901.0002](https://arxiv.org/abs/0901.0002)

Ideally, Tolerance is $\Delta\chi^2 = \pm 1$ for 68% CL

- appropriate if fitting consistent data sets with ideal Gaussian errors to a well-defined theory
- not always the case due to not fully compatible data sets

Solutions:

- use only consistent data and $T = 1$ [ABMP, HERAPDF]
- increase tolerance [CT, MSHT]



PDF UNCERTAINTIES

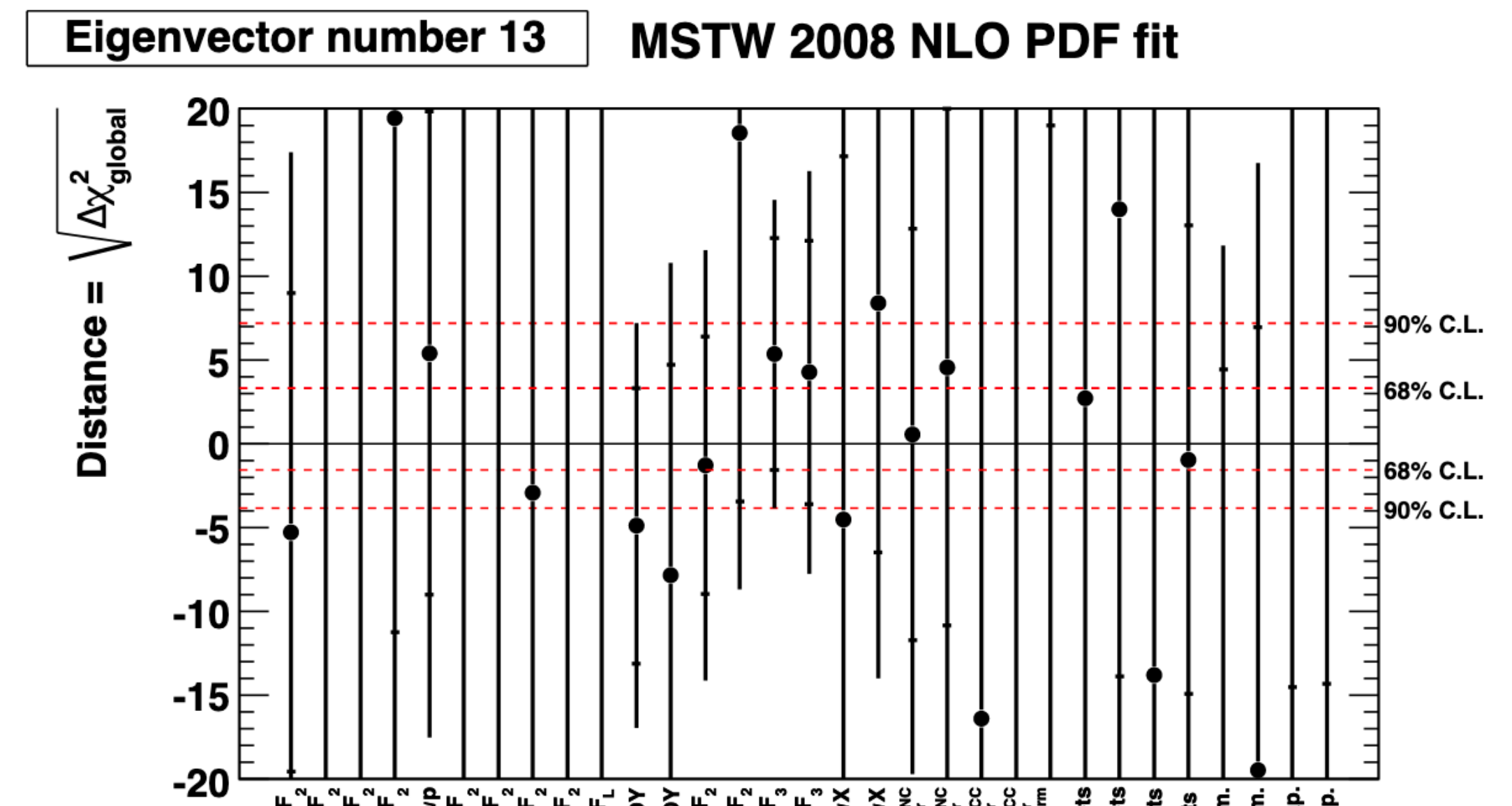
Originate from uncertainties in the experimental measurements and theoretical approximations

Method 1 (J. Pumplin et al) : Hessian uncertainties based on expansion of χ^2

[arXiv:0901.0002](https://arxiv.org/abs/0901.0002)

Ideally, Tolerance is $\Delta\chi^2 = \pm 1$ for 68% CL

- appropriate if fitting consistent data sets with ideal Gaussian errors to a well-defined theory
- not always the case due to not fully compatible data sets



long-standing problem, becomes more pressing with more high-precision LHC data in the PDF fits

→ **most recent discussions at the most recent PDF4LHC meeting**

PDF UNCERTAINTIES

Originate from uncertainties in the experimental measurements and theoretical approximations

Method 2: Monte Carlo (used mostly by NNPDF group)

- generate MC replicas of the experimental data (> 100 -1000)
- perform a separate fit using each replica

Expectation values

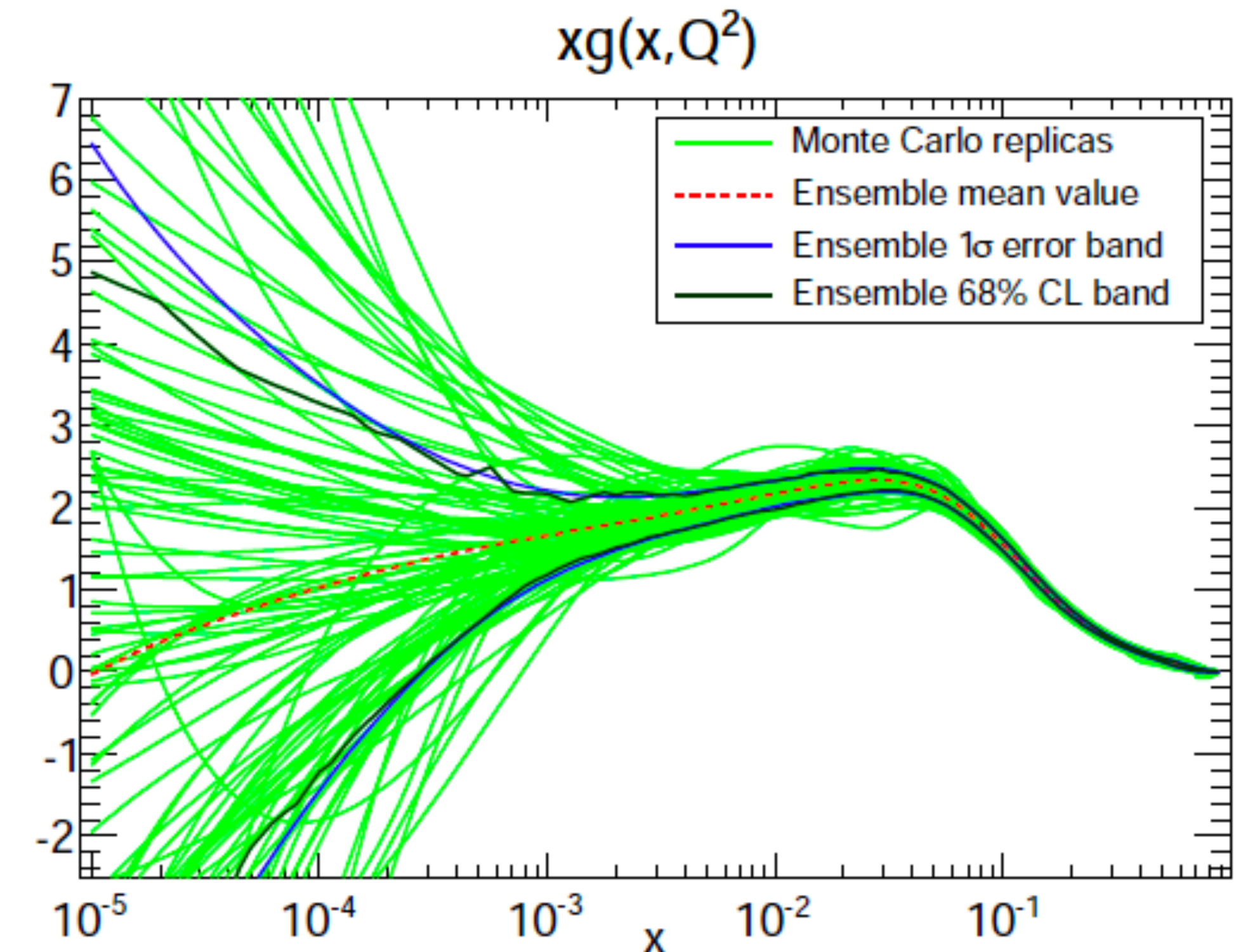
for any variable F_k (of k replica):

$$\langle F \rangle = \frac{1}{N_{rep}} \sum_i^{N_{rep}} F_k$$

Uncertainty

from deviation of
individual replicas
from $\langle F \rangle$

$$\sigma^2[F] = \frac{1}{N_{rep} - 1} \sum_{i=1}^{N_{rep}} (F^{(k)} - \langle F \rangle)^2$$



PDF UNCERTAINTIES

Originate from uncertainties in the experimental measurements and theoretical approximations

Method 2: Monte Carlo (used mostly by NNPDF group)

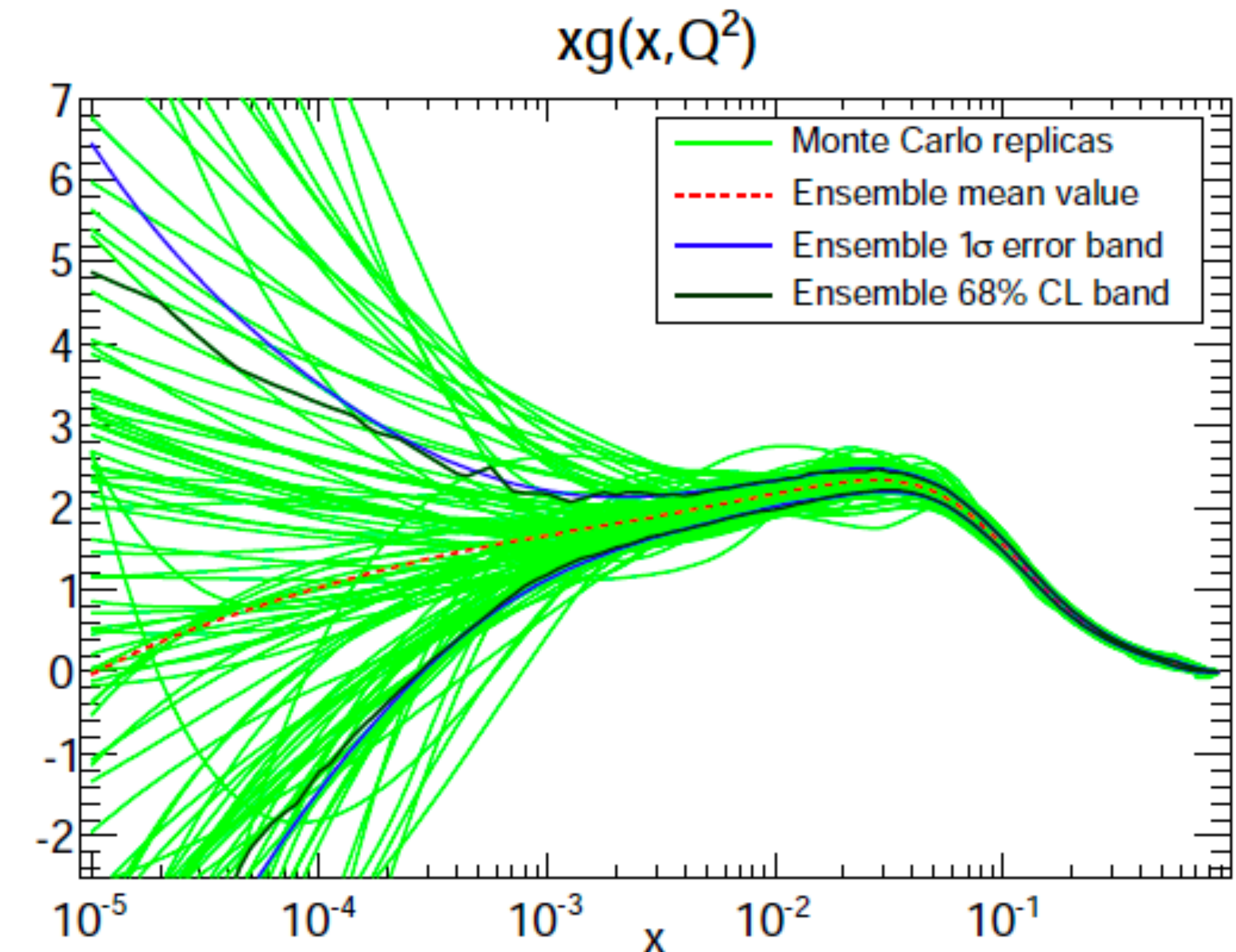
- generate MC replicas of the experimental data ($> 100 - 1000$)
- perform a separate fit using each replica

Expectation values

$$\langle F \rangle = \frac{1}{N_{rep}} \sum_i^{N_{rep}} F_k$$

Uncertainty

$$\sigma^2[F] = \frac{1}{N_{rep} - 1} \sum_{i=1}^{N_{rep}} (F^{(k)} - \langle F \rangle)^2$$



Feature: using this approach for a fixed parametrisation, it is equivalent to $T = 1$

[\(Recent study by MSHT group - fit NNPDF theory + data choices with fixed MSHT parameterisation\)](#)

PDF FITTING METHODOLOGY

Questions addressed currently by the PDF fitters:

- to what extent is tolerance required by tensions in the global fits?
- what is contribution of the parametrisation uncertainty?
- is there a need for tolerance between NN and explicit parametrisation approaches?

PDF FITTING METHODOLOGY

Questions addressed currently by the PDF fitters:

- to what extent is tolerance required by tensions in the global fits?
- what is contribution of the parametrisation uncertainty?
- is there a need for tolerance between NN and explicit parametrisation approaches?

MSHT Study [[arXiv:2407.07944](https://arxiv.org/abs/2407.07944)]:

“makes use of the public NNPDF fitting code to perform a PDF fit to exactly the same data and theory settings that are used in the NNPDF4.0 NNLO public release, but with the MSHT20 fixed parameterisation applied instead of a neural network...”

“the resulting PDFs and various predicted benchmark cross sections due to the public NNPDF4.0 releases **are not compatible** with these MSHT fits within the nominal NNPDF uncertainties.”

“either the NNPDF uncertainties are too aggressive, or the MSHT uncertainties are too conservative, or the truth lies somewhere in between”.

TAKE-HOME MESSAGES

Many interpretations of the LHC measurements are limited by PDFs

PDFs are universal, non-perturbative quantities with perturbative scale evolution, determined from data

Different approaches of PDF fits exist, they differ by

choice of data, flavour scheme, parametrisation, and methodology

Recommendation of PDF fitters is to use individual PDFs for data interpretation

Once using PDFs as an input, respect the perturbative order and scheme

Your measurement might be used in a PDF fit: work out statistical and systematic correlations!