Parton Distribution Functions

2024 CTEQ Summer School on QCD and Electroweak Phenomenology

Emanuele R. Nocera

Università degli Studi di Torino and INFN, Torino

27-28 August 2024

Table of contents

DISCLAIMER

These lectures contain a personal selection of topics and are certainly not exhaustive

Bibliography

1. Textbooks on perturbative QCD
   - J. Campbell, J. Huston, F. Krauss, *The Black Book of QCD*, Oxford (2018)
   - J.C. Collins, *Foundations of perturbative QCD*, Cambridge (2011)
   - R.K. Ellis, W.J. Stirling, B.R. Webber, *QCD and Collider Physics*, Cambridge (1996)
2. Reviews on Parton Distribution Functions
   - K. Kovarik and P.M. Nadolsky, Rev.Mod.Phys. **92** (2020) 045003
   - J.J. Ethier and E.R. Nocera, Ann.Rev.Nucl.Part.Sci. **70** (2020) 43
   - J. Gao, L. Harland-Lang and J. Rojo, Phys.Rept. **742** (2018) 1
   - S. Forte and G. Watt, Ann.Rev.Nucl.Part.Sci. **63** (2013) 291
   - P. Jimenez-Delgado, W. Melnitchouk and J. F. Owens, J. Phys. G40 (2013) 093102
   - E.C. Aschenauer, R.S. Thorne, R. Yoshida (rev.), *Structure Functions*, PDG, ch. 8
3. Specific topics not addressed above
   - more journal references as we proceed through these lectures

DISCLAIMER

These lectures will focus on collinear leading-twist Parton Distribution Functions

Transverse-momentum-dependent distributions will not be covered here

Parton Distribution Functions

Lecture 1: What are Parton Distribution Functions and how we can determine them from experimental data

Outline

1.1 Why Parton Distribution Functions?
the LHC as a laboratory for precision QCD and discovery

1.2 How are PDFs related to Physical Observables?
factorisation, evolution
properties of splitting functions, theoretical constraints

1.3 How can we determine PDFs?
how to formulate the problem and how to solve it

1.4 Which data constrain which PDFs and how?
overview of experimental data: from HERA to the LHC
which constraints different scattering processes put on PDFs

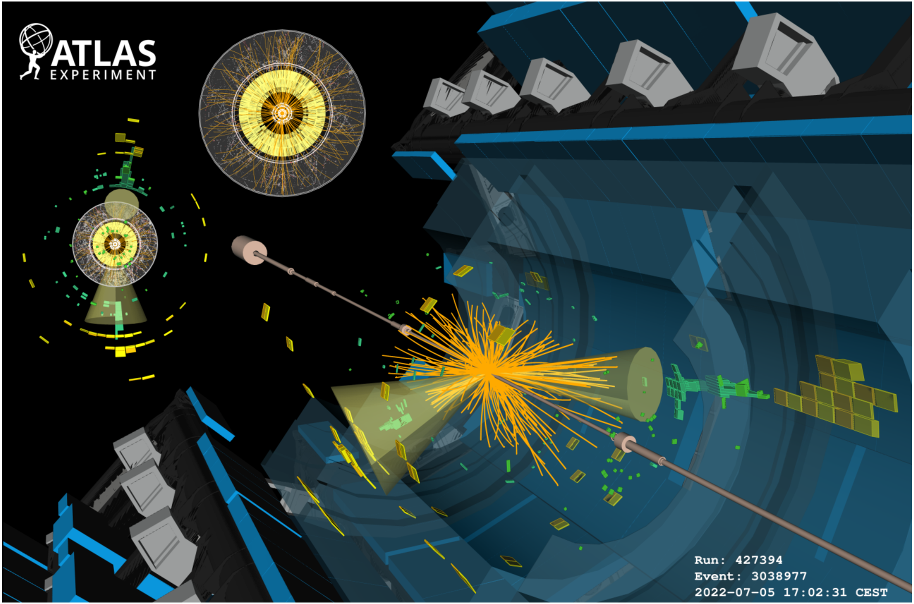
I will focus on the phenomenological determination of PDFs

I will not talk about Lattice QCD nor of models of nucleon structure

See also lectures by D. Soper and A. Cooper-Sarkar

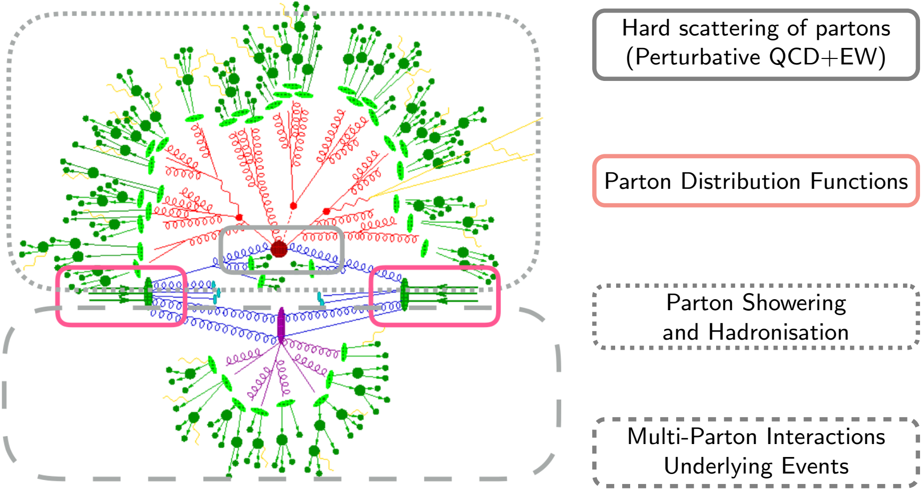
1.1 Why Parton Distribution Functions?

First Collisions of LHC Run III



[Image credit: ATLAS collaboration]

A Laboratory for Quantum Chromodynamics



[Plot by courtesy of SHERPA]

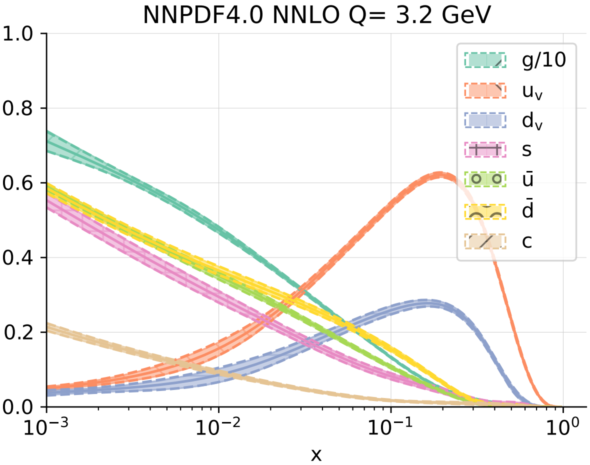
$$\sigma(\tau, Q^2, \mathbf{k}) = \sum_{ij} \int_\tau^1 \frac{dz}{z} \hat{\sigma}_{ij} \left(\frac{\tau}{z}, \alpha_s(Q^2), \mathbf{k} \right) \mathcal{L}_{ij}(z, Q^2) \quad \mathcal{L}_{ij}(z, Q^2) = (f_i^{h_1} \otimes f_j^{h_2})(z, Q^2)$$

Parton Distribution Functions

PDFs express the likelihood of a quark or gluon (partons) to enter a collision

That is, $x\times$PDFs are momentum fraction distributions for each parton

Dependence on x is non-perturbative (fit); dependence on Q^2 is perturbative (DGLAP)



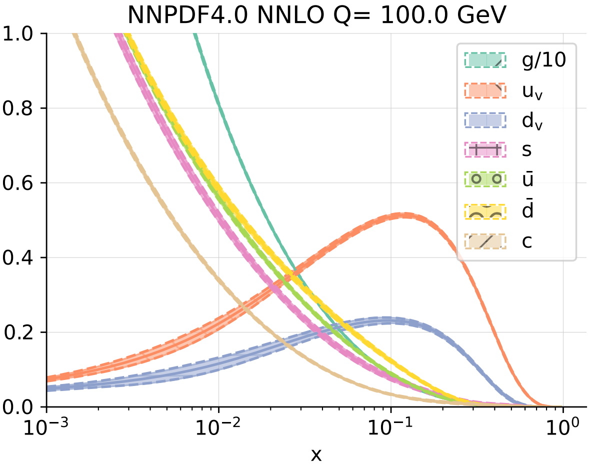
[Plot from the PDG Review of Particle Physics]

Parton Distribution Functions

PDFs express the likelihood of a quark or gluon (partons) to enter a collision

That is, $x\times$PDFs are momentum fraction distributions for each parton

Dependence on x is non-perturbative (fit); dependence on Q^2 is perturbative (DGLAP)



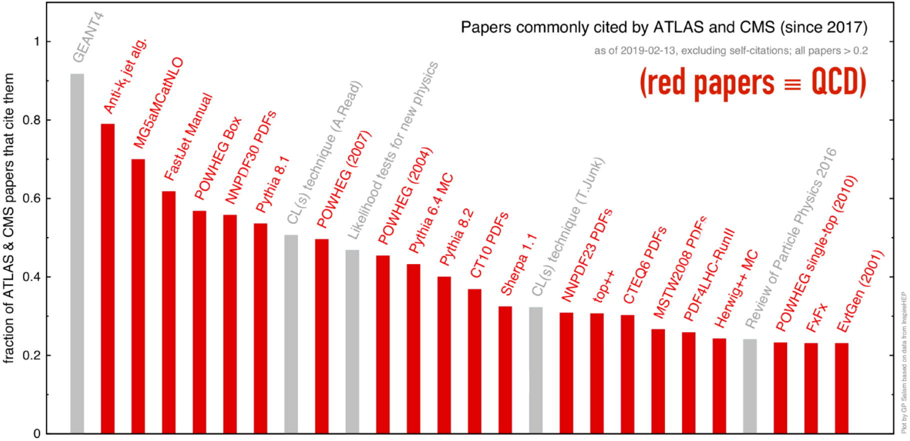
[Plot from the PDG Review of Particle Physics]

LHC, QCD and PDFs

The LHC is a Proton Collider – Any interaction contains a strong interaction

Quantum Chromodynamics (QCD) is the main actor

Within QCD, Parton Distribution Functions (PDFs) play a leading role



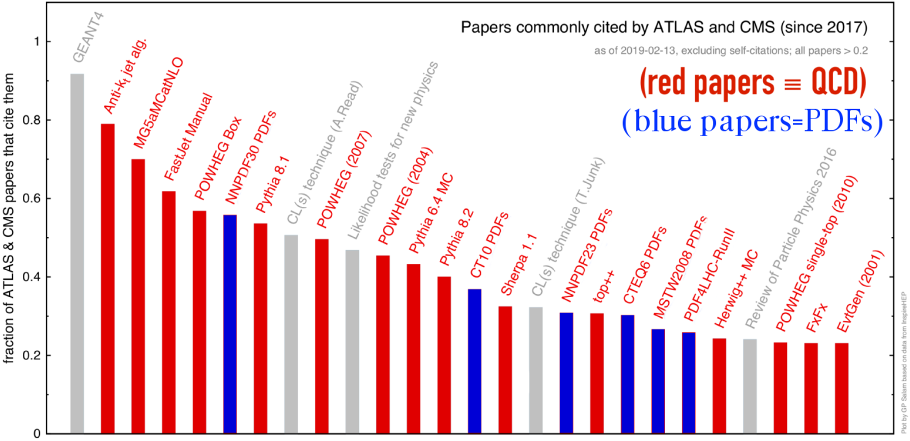
[Plot by courtesy of G. Salam]

LHC, QCD and PDFs

The LHC is a Proton Collider – Any interaction contains a strong interaction

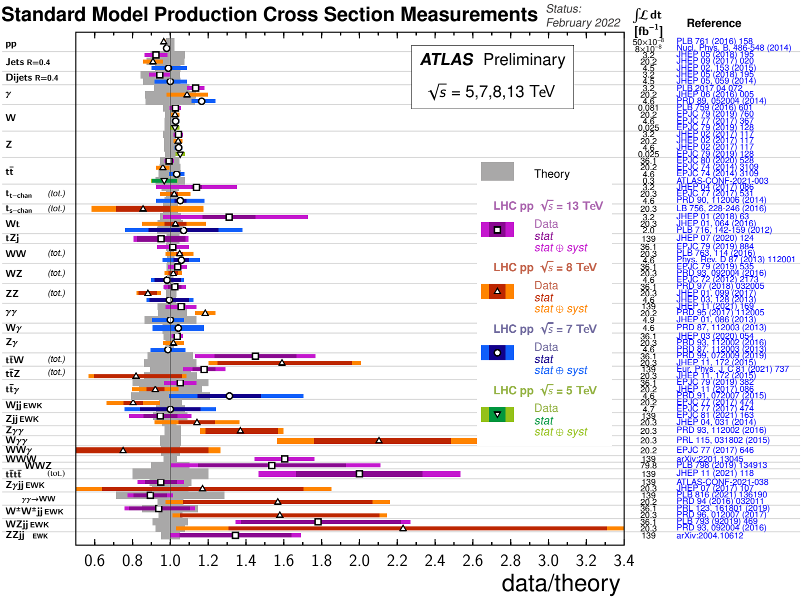
Quantum Chromodynamics (QCD) is the main actor

Within QCD, Parton Distribution Functions (PDFs) play a leading role



[Plot by courtesy of G. Salam]

Physics at the LHC as Precision Physics

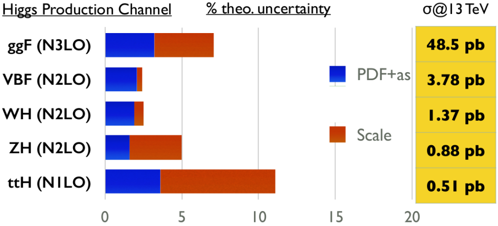


[Plot from ATLAS Collaboration web page]

PDFs as a Tool: Making Predictions with PDFs

PDF uncertainty is often the dominant source of uncertainty in LHC cross sections

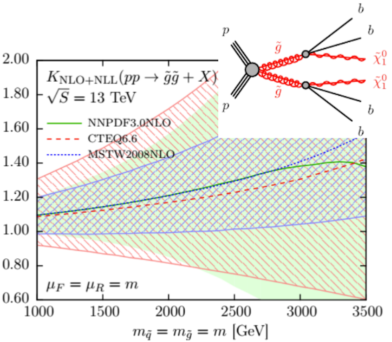
Precision



Channel	$m_{W^+} - m_{W^-}$ [MeV]	Stat. Unc.	Muon Unc.	Elec. Unc.	Recoil Unc.	Bckg. Unc.	QCD Unc.	EW Unc.	PDF Unc.	Total Unc.
$W \to e\nu$	−29.7	17.5	0.0	4.9	0.9	5.4	0.5	0.0	24.1	30.7
$W \to \mu\nu$	−28.6	16.3	11.7	0.0	1.1	5.0	0.4	0.0	26.0	33.2
Combined	−29.2	12.8	3.3	4.1	1.0	4.5	0.4	0.0	23.9	28.0

[Plot from the CERN Yellow Report 2016]

Discovery



[EPJC 76 (2016) 53]

Higgs boson characterisation

Determination of SM parameters, such as the mass of the W boson

Searches for beyond SM physics at large invariant mass of the final state

1.2 How are PDFs related to Physical Observables?

Factorisation of Physical Observables

1. Factorisation theorems apply to sufficiently inclusive scattering processes

short-distance interaction partonic hard interaction process-specific kernel	$\xleftrightarrow[\text{scale } \mu]{\text{factorisation scheme}}$	long-distance interaction nucleon structure universal PDFs

2. Physical observables can be written as convolutions of matrix elements and PDFs

$$F_I(x,\mu^2) = \sum_i \int_x^1 \frac{dz}{z} C_{Ii}(z,\alpha_s(\mu^2)) f_i\left(\frac{x}{z},\mu^2\right) \qquad \text{ONE HADRON}$$

$$\sigma(\tau,\mu^2,\mathbf{k}) = \sum_{ij} \int_\tau^1 \frac{dz}{z} \hat{\sigma}_{ij}\left(\frac{\tau}{z},\alpha_s(\mu^2),\mathbf{k}\right) \mathcal{L}_{ij}(z,\mu^2) \qquad \text{TWO HADRONS}$$

$$\mathcal{L}_{ij}(z,\mu^2) = (f_i^{h_1} \otimes f_j^{h_2})(z,\mu^2)$$

$$f \otimes g = \int_x^1 \frac{dz}{z} f\left(\frac{x}{z}\right) g(z)$$

3. The matrix elements C_{If} and $\hat{\sigma}_{ij}$ can be computed perturbatively

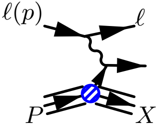
$$C_{Ii}(y,\alpha_s) = \sum_{k=0} a_s^k C_{Ii}^{(k)}(y) \qquad \hat{\sigma}_{ij}(y,\alpha_s) = \sum_{k=0} a_s^k \hat{\sigma}_{ij}^{(k)}(y) \qquad a_s = \alpha_s/(4\pi)$$

4. Because of factorisation, all of these quantities depend on μ^2; usually $Q^2 = \mu^2$

Factorisation Kinematics

ONE HADRON

$$F_I(x,\mu^2) = \sum_{i=q,\bar{q},g} \int_0^1 dz \int_x^1 dy \delta(x-yz) C_{Ii}(z,\mu^2) f_i(y,\mu^2)$$



scale: $Q^2 = -q^2$

scaling variable (hadronic): $x = x_B = \frac{Q^2}{2P\cdot q}$

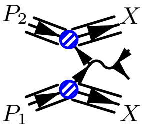
scaling variable (partonic): $z = \frac{Q^2}{2p\cdot q}$

incoming parton momentum fraction: $y \to p = yP$

P: proton momentum q: gauge boson momentum p parton momentum

TWO HADRONS

$$\sigma(\tau,\mu^2,\mathbf{k}) = \sum_{i,j=q,\bar{q},g} \int_0^1 dz \int_\tau^1 dx_1 dx_2 \delta(\tau - x_1 x_2 z) \hat{\sigma}_{ij}(\tau/z,\mu^2,\mathbf{k}) f_i^{h_1}(x_1,\mu^2) f_j^{h_2}(x_2,\mu^2)$$



scale: M^2

scaling variable (hadronic): $\tau = \frac{M^2}{s}$ $s = (P_1+P_2)^2$

scaling variable (partonic): $z = \frac{M^2}{\hat{s}}$ $\hat{s} = (p_1+p_2)^2$

incoming parton momentum fractions: $x_{1,2} \to p_{1,2} = x_{1,2} P_{1,2}$

$P_{1,2}$: proton momenta M^2: final state mass $p_{1,2}$ parton momenta

Theoretical constraints

1. Momentum sum rule (momentum conservation)

$$\int_0^1 dx x \left[\sum_{q=1}^{n_f} \left(f_q(x,Q^2) + f_{\bar{q}}(x,Q^2) \right) + f_g(x,Q^2) \right] = 1$$

2. Valence sum rules (baryon number conservation)

$$\int_0^1 dx \left[f_u(x,Q^2) - f_{\bar{u}}(x,Q^2) \right] = 2$$

$$\int_0^1 dx \left[f_d(x,Q^2) - f_{\bar{d}}(x,Q^2) \right] = 1$$

$$\int_0^1 dx \left[f_q(x,Q^2) - f_{\bar{q}}(x,Q^2) \right] = 0 \qquad q = s, c, b, t$$

3. Isospin symmetry of the strong interaction

$$f_u^p = f_d^n \qquad f_{\bar{u}}^p = f_{\bar{d}}^n$$

4. Positivity of cross sections [PRD 105 (2022) 076010; EPJ C84 (2024) 335]
   → PDFs should be positive-definite at LO
   → beyond LO, PDFs ought not be positive, however they are positive above for Q^2 large

5. Integrability of non-singlet PDFs
   → follows from operator product expansion

PDF evolution: DGLAP equations

1. A set of $(2n_f+1)$ integro-differential equations, n_f is the number of active partons

$$\frac{\partial}{\partial \ln \mu^2} f_i(x,\mu^2) = \sum_j^{n_f} \int_x^1 \frac{dz}{z} P_{ji}\left(z, \alpha_s(\mu^2)\right) f_j\left(\frac{x}{z}, \mu^2\right)$$

2. They are often written in a convenient PDF basis

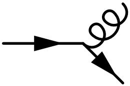
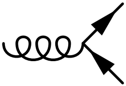
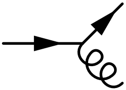
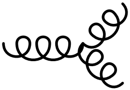
$$f_\pm = (f_q \pm f_{\bar{q}}) - (f_{q'} \pm f_{\bar{q}'}) \qquad f_v = \sum_i^{n_f} (f_q - f_{\bar{q}}) \qquad f_\Sigma = \sum_i^{n_f} (f_q + f_{\bar{q}})$$

$$\frac{\partial}{\partial \ln \mu^2} f_{\pm,v}(x,\mu^2) = P^{\pm,v}(x,\mu_F^2) \otimes f_{\pm,v}(x,\mu^2)$$

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

3. The splitting functions P can be computed perturbatively

$$P_{ji}(z,\alpha_s) = \sum_{k=0} a_s^{k+1} P_{ji}^{(k)}(z), \qquad a_s = \alpha_s/(4\pi)$$

$P_{qq}^{(0)}$  $P_{gq}^{(0)}$  $P_{qg}^{(0)}$  $P_{gg}^{(0)}$ 

Splitting Functions and Anomalous Dimensions

Perform the Mellin transform $\gamma_{ji}(N,\alpha_s(\mu^2)) \equiv \int_0^1 dx x^{N-1} P_{ji}(x,\alpha_s(\mu^2))$

$$\frac{\partial}{\partial \ln \mu^2} f_{\pm,v}(N,\mu^2) = \gamma_{\pm,v}(N,\mu_F^2) \cdot f_{\pm,v}(N,\mu^2)$$

$$\frac{\partial}{\partial \ln \mu^2} \begin{pmatrix} f_\Sigma(N,\mu^2) \\ f_g(N,\mu^2) \end{pmatrix} = \begin{pmatrix} \gamma_{qq} & 2n_f\gamma_{qg} \\ \gamma_{gq} & \gamma_{gg} \end{pmatrix} \cdot \begin{pmatrix} \Sigma(N,\mu^2) \\ g(N,\mu^2) \end{pmatrix}$$

How many different anomalous dimensions are there?

LO: $\gamma_{qq} = \gamma_{\pm,v} \to$ 4 independent splitting functions
NLO: $\gamma_{qq} \neq \gamma_+ \neq \gamma_- \to$ 6 independent splitting functions
NNLO: $\gamma_- \neq \gamma_v \to$ 7 independent splitting functions

Which PDF combinations evolve independently?

LO: f_g, f_Σ, and any $2n_f - 1$ linear combinations of f_q and $f_{\bar{q}}$
NLO: f_g, f_Σ, any $n_f - 1$ linear combinations of $f_q - f_{\bar{q}}$, and of $f_q + f_{\bar{q}}$
NNLO: as NLO, and $f_V = \sum_q^{n_f}(f_q - f_{\bar{q}})$

A common choice

$$f_g,\ f_\Sigma = \sum_q^{n_f}(f_q + f_{\bar{q}}),\ f_V = \sum_q^{n_f}(f_q - f_{\bar{q}})$$

iterative NS combinations of $f_{q^+} = f_q + f_{\bar{q}}$ and of $f_{q^-} = f_q - f_{\bar{q}}$

$$T_3 = f_{u^+} - f_{d^+} \qquad T_8 = f_{u^+} + f_{d^+} - 2f_{s^+} \qquad T_{15} = f_{u^+} + f_{d^+} + f_{s^+} - 3f_{c^+} \ \ldots$$

$$V_3 = f_{u^-} - f_{d^-} \qquad V_8 = f_{u^-} + f_{d^-} - 2f_{s^-} \qquad V_{15} = f_{u^-} + f_{d^-} + f_{s^-} - 3f_{c^-} \ \ldots$$

Anomalous dimensions: perturbative accuracy

LO (1977)

[NPB 126 (1977) 298]

$$
\begin{aligned}
\gamma_{\rm ps}^{(0)}(N) &= 0 \\
\gamma_{\rm qg}^{(0)}(N) &= 2n_f\left(\mathbf{N}_- + 4\mathbf{N}_+ - 2\mathbf{N}_{+2} - 3\right)S_1 \\
\gamma_{\rm gq}^{(0)}(N) &= 2C_F\left(2\mathbf{N}_{-2} - 4\mathbf{N}_- - \mathbf{N}_+ + 3\right)S_1 \\
\gamma_{\rm gg}^{(0)}(N) &= C_A\left(4(\mathbf{N}_{-2} - 2\mathbf{N}_- - 2\mathbf{N}_+ + \mathbf{N}_{+2} + 3)S_1 - \frac{11}{3}\right) + \frac{2}{3}n_f
\end{aligned}
$$

NLO (1980)

[NPB 175 (1980) 27; PLB 97 (1980) 437]

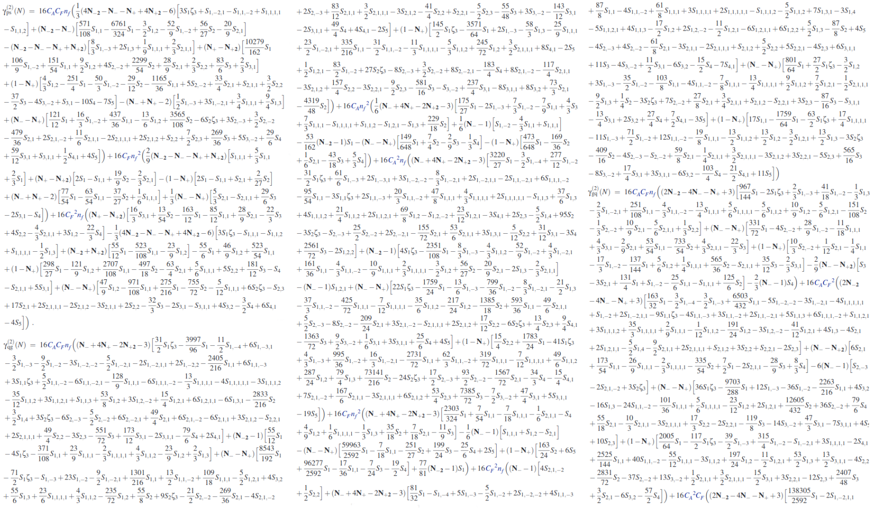
$$
\begin{aligned}
\gamma_{\rm ps}^{(1)}(N) = {} & 4C_F n_f\Big(\frac{20}{9}(\mathbf{N}_{-2} - \mathbf{N}_-)S_1 - (\mathbf{N}_+ - \mathbf{N}_{+2})\Big[\frac{56}{9}S_1 + \frac{8}{3}S_2\Big] + (1-\mathbf{N}_+)\Big[8S_1 - 4S_2\Big] \\
& - (\mathbf{N}_- - \mathbf{N}_+)\Big[2S_1 + S_2 + 2S_3\Big]\Big) \\
\gamma_{\rm qg}^{(1)}(N) = {} & 4C_A n_f\Big(\frac{20}{9}(\mathbf{N}_{-2} - \mathbf{N}_-)S_1 - (\mathbf{N}_- - \mathbf{N}_+)\Big[2S_1 + S_2 + 2S_3\Big] - (\mathbf{N}_+ - \mathbf{N}_{+2})\Big[\frac{218}{9}S_1 \\
& + 4S_{1,1} + \frac{44}{3}S_2\Big] + (1-\mathbf{N}_+)\Big[27S_1 + 4S_{1,1} - 7S_2 - 2S_3\Big] - 2(\mathbf{N}_- + 4\mathbf{N}_+ - 2\mathbf{N}_{+2} - 3)\Big[S_{1,-2} \\
& + S_{1,1,1}\Big]\Big) + 4C_F n_f\Big(2(\mathbf{N}_+ - \mathbf{N}_{+2})\Big[5S_1 + 2S_{1,1} - 2S_2 + S_3\Big] - (1-\mathbf{N}_+)\Big[\frac{43}{2}S_1 + 4S_{1,1} - \frac{7}{2}S_2\Big] \\
& + (\mathbf{N}_- - \mathbf{N}_+)\Big[7S_1 - \frac{3}{2}S_2\Big] + 2(\mathbf{N}_- + 4\mathbf{N}_+ - 2\mathbf{N}_{+2} - 3)\Big[S_{1,1,1} - S_{1,2} - S_{2,1} + \frac{1}{2}S_3\Big]\Big) \qquad (3.7) \\
\gamma_{\rm gq}^{(1)}(N) = {} & 4C_A C_F\Big(2(2\mathbf{N}_{-2} - 4\mathbf{N}_- - \mathbf{N}_+ + 3)\Big[S_{1,1,1} - S_{1,-2} - S_{1,2} - S_{2,1}\Big] + (1-\mathbf{N}_+)\Big[2S_1 \\
& - 13S_{1,1} - 7S_2 - 2S_3\Big] + (\mathbf{N}_{-2} - 2\mathbf{N}_- + \mathbf{N}_+)\Big[S_1 - \frac{22}{3}S_{1,1}\Big] + 4(\mathbf{N}_- - \mathbf{N}_+)\Big[\frac{7}{9}S_1 + 3S_2 + S_3\Big] \\
& + (\mathbf{N}_+ - \mathbf{N}_{+2})\Big[\frac{44}{9}S_1 + \frac{8}{3}S_2\Big]\Big) + 4C_F n_f\Big((\mathbf{N}_{-2} - 2\mathbf{N}_- + \mathbf{N}_+)\Big[\frac{4}{3}S_{1,1} - \frac{20}{9}S_1\Big] - (1-\mathbf{N}_+)\Big[4S_1 \\
& - 2S_{1,1}\Big]\Big) + 4C_F^2\Big((2\mathbf{N}_{-2} - 4\mathbf{N}_- - \mathbf{N}_+ + 3)\Big[3S_{1,1} - 2S_{1,1,1}\Big] - (1-\mathbf{N}_+)\Big[S_1 - 2S_{1,1} + \frac{3}{2}S_2 \\
& - 3S_3\Big] - (\mathbf{N}_- - \mathbf{N}_+)\Big[\frac{5}{2}S_1 + 2S_2 + 2S_3\Big]\Big) \\
\gamma_{\rm gg}^{(1)}(N) = {} & 4C_A n_f\Big(\frac{2}{3} - \frac{16}{3}S_1 - \frac{23}{9}(\mathbf{N}_{-2} + \mathbf{N}_{+2})S_1 + \frac{14}{3}(\mathbf{N}_- + \mathbf{N}_+)S_1 + \frac{2}{3}(\mathbf{N}_- - \mathbf{N}_+)S_2\Big) \\
& + 4C_A^2\Big(2S_{-3} - \frac{8}{3} - \frac{14}{3}S_1 + 2S_3 - (\mathbf{N}_{-2} - 2\mathbf{N}_- - 2\mathbf{N}_+ + \mathbf{N}_{+2} + 3)\Big[4S_{1,-2} + 4S_{1,2} + 4S_{2,1}\Big] \\
& + \frac{8}{3}(\mathbf{N}_+ - \mathbf{N}_{+2})S_2 - 4(\mathbf{N}_- - 3\mathbf{N}_+ + \mathbf{N}_{+2} + 1)\Big[3S_2 - S_3\Big] + \frac{109}{18}(\mathbf{N}_- + \mathbf{N}_+)S_1 + \frac{61}{3}(\mathbf{N}_- \\
& - \mathbf{N}_+)S_2\Big) + 4C_F n_f\Big(\frac{1}{2} + \frac{2}{3}(\mathbf{N}_{-2} - 13\mathbf{N}_- - \mathbf{N}_+ - 5\mathbf{N}_{+2} + 18)S_1 + (3\mathbf{N}_- - 5\mathbf{N}_+ + 2)S_2 \\
& - 2(\mathbf{N}_- - \mathbf{N}_+)S_3\Big) .
\end{aligned}
$$

Numerical solution (LO, NLO , NNLO and aN^{3}LO) of DGLAP implemented in open-source software: EKO [EPJ C82 (2022) 976] and APFEL++ [CPC 185 (2014) 1647]

Anomalous dimensions: perturbative accuracy

NNLO (2004)

[NPB 691 (2004) 129]

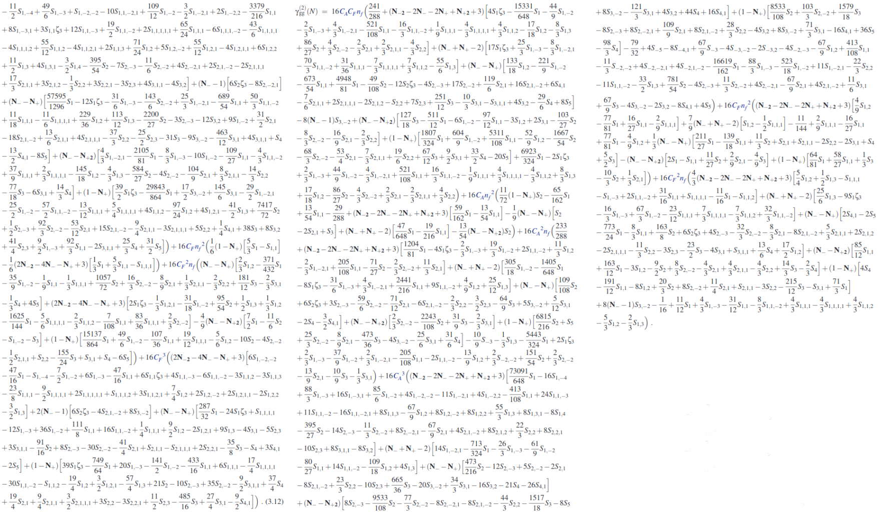


aN^{3}LO (2020 - ongoing)

Anomalous dimensions: perturbative accuracy

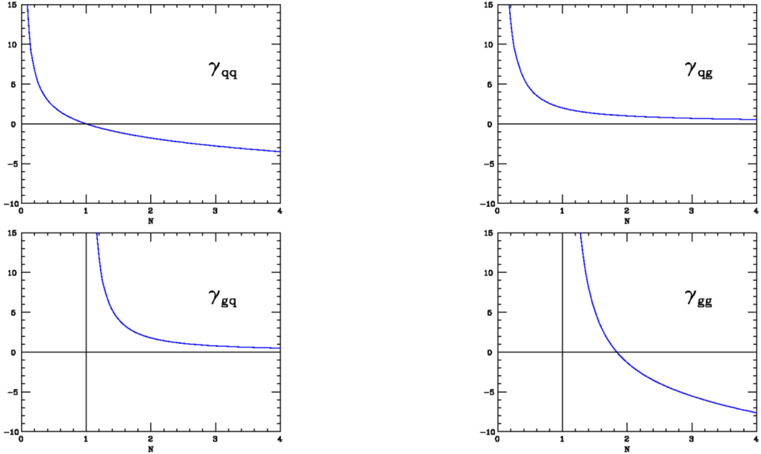
NNLO cont'd (2004)

[NPB 691 (2004) 129]



aN^{3}LO (2020 - ongoing)

Anomalous Dimensions: Scale Dependence at LO



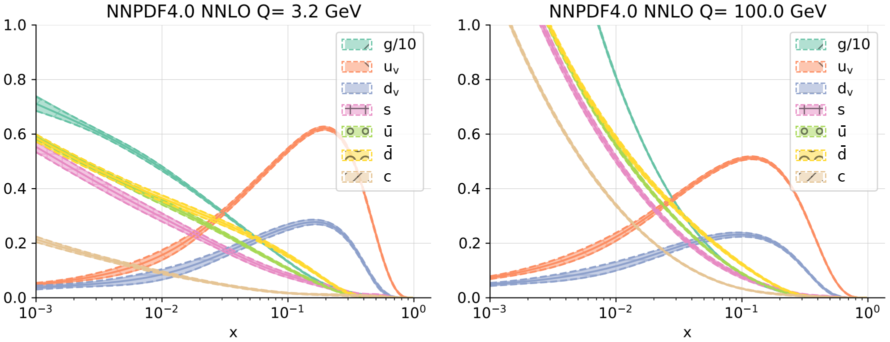
As Q^2 increases, PDFs decrease at large x and increase at small x due to radiation

Gluon sector singular at $N = 1$, therefore the gluon grows fater at small x

$\gamma_{qq}(1) = 0$ follows from baryon number conservation (beyond LO, $\gamma_{qq}(1) = \gamma_{q\bar{q}}(1)$)

$\gamma_{qq}(2) + \gamma_{qg}(2) = \gamma_{gq}(2) + \gamma_{gg}(2) = 0$ follows from momentum conservation

PDFs: Qualitative features



The valence bump follows from sum rules

The small-x growth of the gluon PDF follows from singularity of γ_{gg} at $N = 1$

The similar small-x rise of all PDFs follows from singlet-gluon mixing

PDF depletion at large x and Q^2 follows from sign change of anomalous dimensions

Valence does not evolve multiplicatively because $\gamma_- \neq \gamma_v$

Valence does not vanish at all scales

1.3 How can we determine PDFs?

PDF determination in statistical language

Inverse problem

Given a set of data D, determine $p(f|D)$ in the space of functions $f : [0,1] \to \mathbb{R}$.

The expectation value and uncertainty of each physical observable $\mathcal{O}$ that depends on a PDF set $[f]$ are functional integrals of the PDFs

$$\langle \mathcal{O}[f] \rangle = \int \mathcal{D}f \, p(f|D) \, \mathcal{O}[f] \qquad \text{expectation value}$$

$$\sigma_{\mathcal{O}}[f] = \left[\int \mathcal{D}f \, p(f|D) \, \left(\mathcal{O}[f] - \langle \mathcal{O}[f] \rangle\right)^2 \right]^{1/2} \qquad \text{uncertainty}$$

THE PROBLEM IS ILL-DEFINED

We want to determine infinite-dimensional objects, the PDFs, from a finite set of data

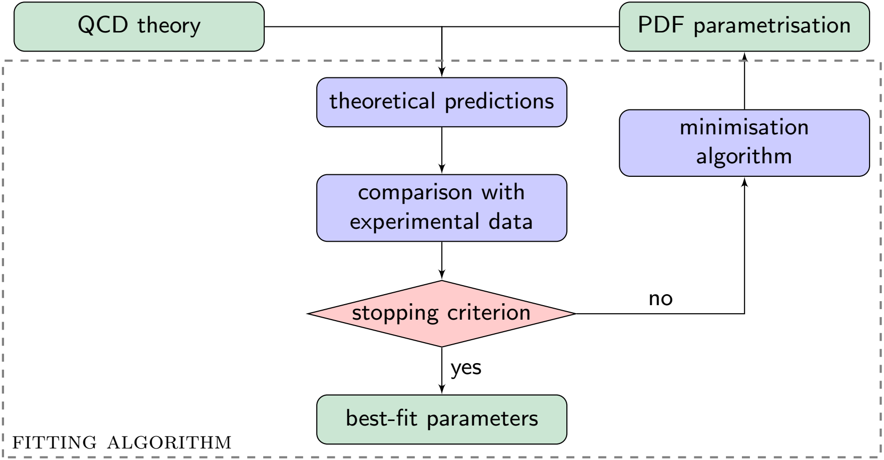
Solution: parametric regression

Approximate $p(f|D)$ with its projection in the space of parameters $p(\boldsymbol{\theta}|D,\mathcal{H})$

Determine $p(\boldsymbol{\theta}|D,\mathcal{H}) \propto p(D|\boldsymbol{\theta},\mathcal{H})p(\boldsymbol{\theta}|\mathcal{H})$ as MAP $\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|D,\mathcal{H})$

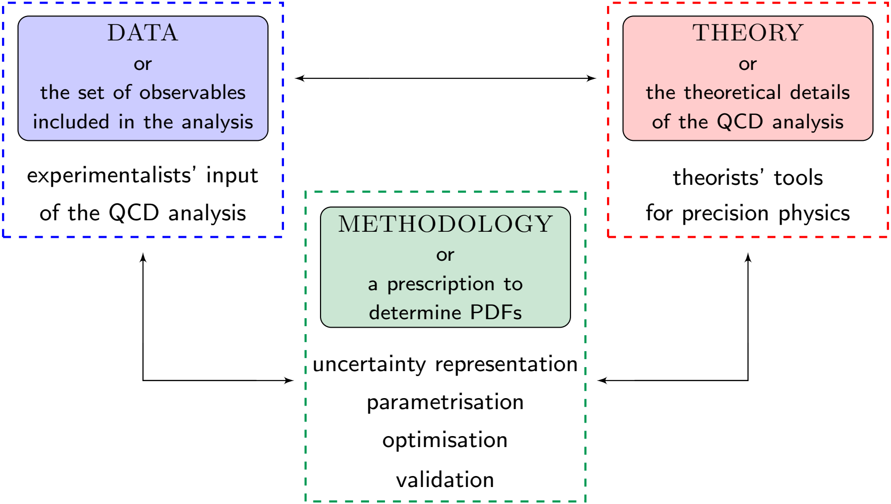
Determining PDFs from (LHC) experimental data

$$\sigma(\tau, Q^2\mathbf{k}) = \sum_{ij}\int_\tau^1 \frac{dz}{z}\hat{\sigma}_{ij}\left(\frac{\tau}{z}, \alpha_s(Q^2), \mathbf{k}\right)\mathcal{L}_{ij}(z, Q^2) \quad \mathcal{L}_{ij}(z, Q^2) = (f_i^{h_1}\otimes f_j^{h_2})(z, Q^2)$$



$$\chi^2 = \sum_{i,j}^{N_{\rm dat}}[T_i[\{\vec{a}\}] - D_i](\mathrm{cov}^{-1})_{ij}[T_j[\{\vec{a}\}] - D_j]$$ with $\{\vec{a}\}$ the set of parameters

The ingredients of PDF determination



Each of these ingredients is a source of uncertainty in the PDF determination
Each of these ingredients require to make choices which lead to different PDF sets

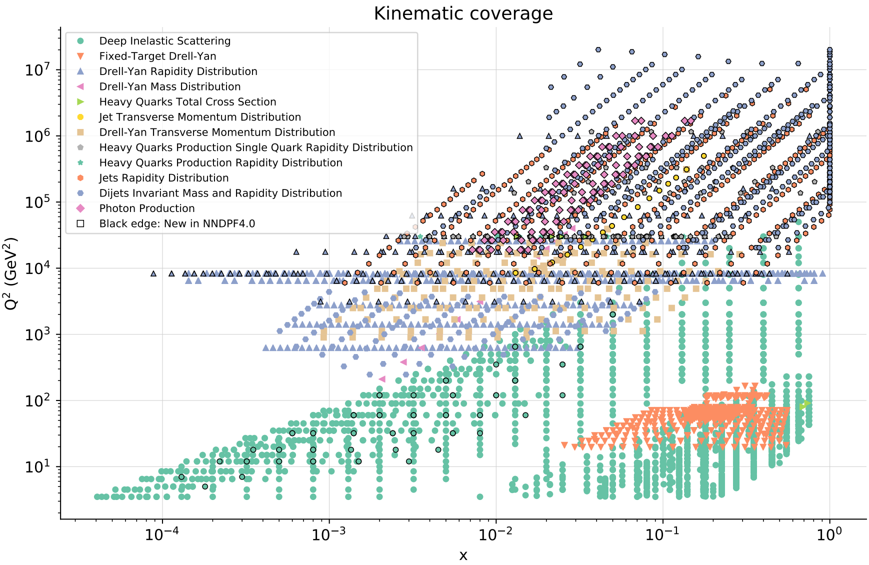
Overview of current PDF determinations

	NNPDF4.0	MSHT20	CT18	HERAPDF2.0	CJ22	ABMP16
Fixed-target DIS	☑	☑	☑	☒	☑	☑
JLAB	☒	☒	☒	☒	☑	☒
HERA I+II	☑	☑	☑	☑	☑	☑
HERA jets	☑	☒	☒	☑	☒	☒
Fixed target DY	☑	☑	☑	☒	☑	☑
Tevatron W, Z	☑	☑	☑	☒	☑	☑
LHC vector boson	☑	☑	☑	☒	☑	☑
LHC $W+c$ $Z+c$	☑	☒	☒	☒	☒	☒
Tevatron jets	☑	☑	☑	☒	☑	☒
LHC jets	☑	☑	☑	☒	☒	☒
LHC top	☑	☑	☒	☒	☒	☑
LHC single t	☑	☒	☒	☒	☒	☒
LHC prompt γ	☑	☒	☒	☒	☒	☒
statistical treatment	Monte Carlo	Hessian $\Delta\chi^2$ dynamical	Hessian $\Delta\chi^2$ dynamical	Hessian $\Delta\chi^2=1$	Hessian $\Delta\chi^2=1.645$	Hessian $\Delta\chi^2=1$
parametrisation	Neural Network	Chebyschev pol.	Bernstein pol.	polynomial	polynomial	polynomial
HQ scheme	FONLL	TR′	ACOT-χ	TR′	ACOT-χ	FFN
accuracy	aN^{3}LO	aN^{3}LO	NNLO	NNLO	NLO	NNLO
latest update	EPJ C82 (2022) 428	EPJ C81 (2021) 341	PRD 103 (2021) 014013	EPJ C82 (2022) 243	PRD 107 (2023) 113005	PRD 96 (2017) 014011

All PDF sets are available as (x, Q^2) interpolation grids through the LHAPDF library

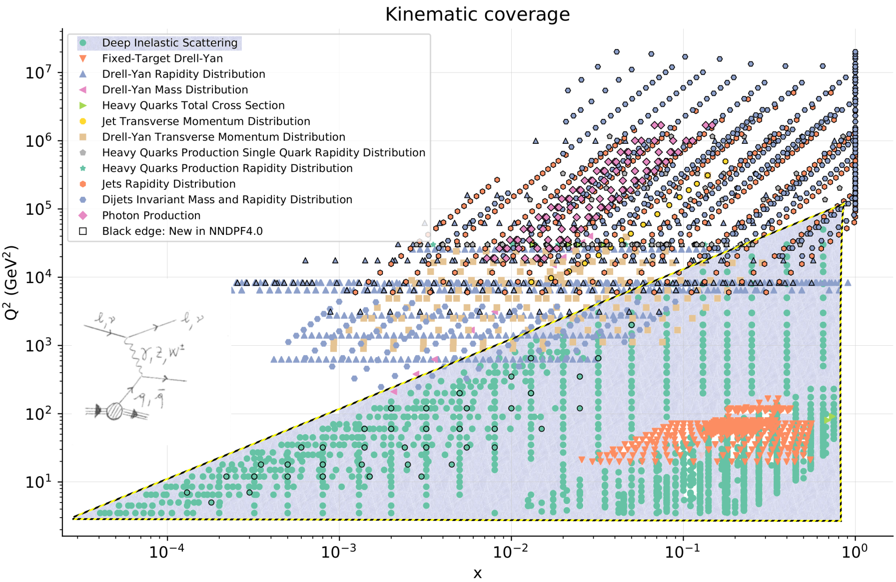
1.4 Data

Overview of experimental data



$$N_{\mathrm{dat}} = 4618$$

Deep Inelastic Scattering



Deep Inelastic Scattering

Re-write the cross section in terms of structure functions (F_2, F_3, F_L)

$$\frac{d^2\sigma^i}{dxdy} \propto Y_+ F_2^i \mp Y_- x F_3^i - y^2 F_L^i \quad Y_\pm = 1 \pm (1-y)^2 \quad F_L^i = F_2^i - 2xF_1^i \quad i = \mathrm{NC}(\gamma, \mathrm{Z}, \gamma\mathrm{Z}), \mathrm{CC}(\mathrm{W}^\pm)$$

NC DIS ($\ell p \to \ell X$) at LO (NMC, SLAC, BCDMS, HERA)

$$\left[F_2^\gamma, F_2^{\gamma Z}, F_2^Z\right] = x \sum_q \left[e_q^2, 2e_q g_V^q, (g_V^q)^2 + (g_A^q)^2\right] (q + \bar{q})$$

$$\left[F_3^\gamma, F_3^{\gamma Z}, F_3^Z\right] = \sum_q \left[0, 2e_q g_A^q, 2g_V g_A^q\right] (q - \bar{q})$$

$$\left[F_L^\gamma, F_L^{\gamma Z}, F_L^Z\right] = [0, 0, 0]$$

CC DIS ($\ell^- p \to \nu X$ or $\bar{\nu} p \to \ell^+ X$) at LO (CHORUS, NuTeV, HERA)

$$\left[F_2^{W^-}, F_2^{W^-}\right] = 2x \left[(u + \bar{d} + \bar{s} + c \ldots), (d + \bar{u} + \bar{c} + s \ldots)\right]$$

$$\left[F_3^{W^-}, F_3^{W^+}\right] = 2 \left[(u - \bar{d} - \bar{s} + c \ldots), (d - \bar{u} - \bar{c} + s \ldots)\right]$$

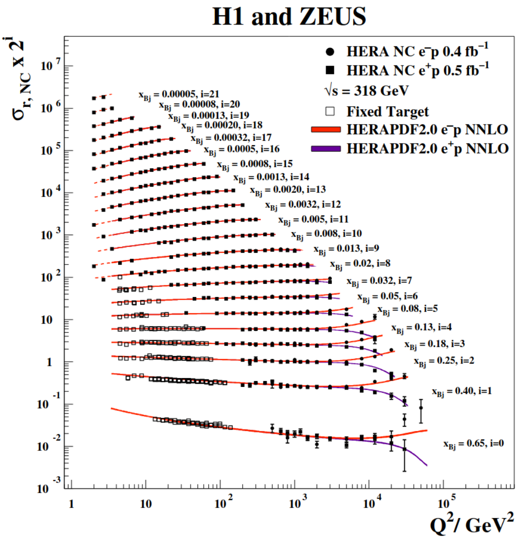
$$\left[F_L^{W^+}, F_L^{W^-}\right] = [0, 0]$$

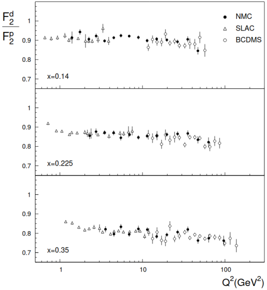
isospin ($p \to n$): $u^p = d^n \qquad d^p = u^n \qquad \bar{u}^p = \bar{d}^n \qquad \bar{d}^p = \bar{u}^n$

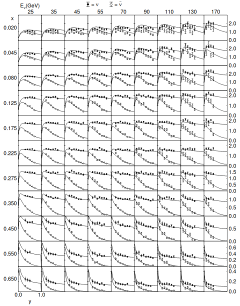
deuteron target approximated as the average of one proton and one neutron

$$Q^2 \geq Q^2_{\mathrm{min}} \sim 1\ \mathrm{GeV}^2 \qquad W^2 = Q^2(1-x)/x \geq W^2_{\mathrm{min}} \sim 10\ \mathrm{GeV}^2$$

Deep Inelastic Scattering



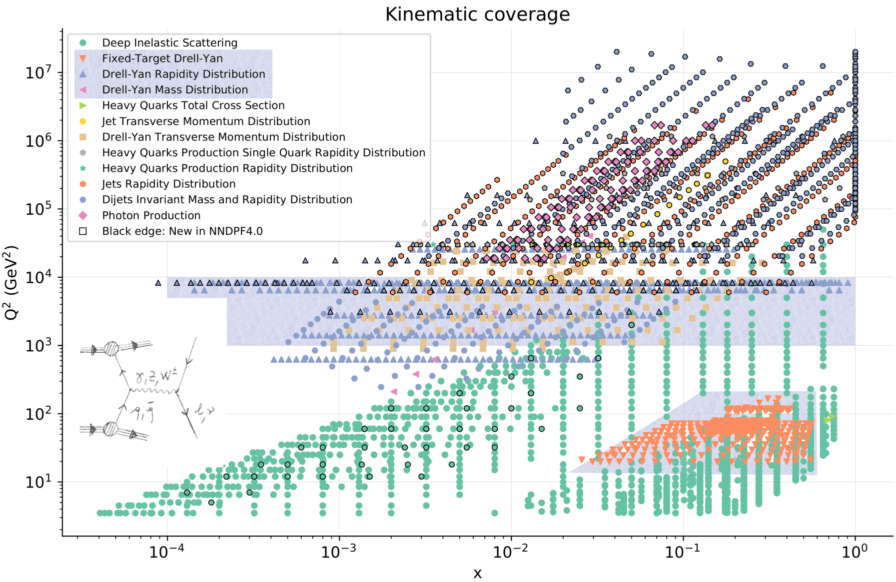




[HERA, EPJ C75 (2015) 580]

[NMC, NPB 487 (1997) 3; CHORUS, PLB 632 (2006) 65]

Drell-Yan


Kinematic coverage
Deep Inelastic Scattering
Fixed-Target Drell-Yan
Drell-Yan Rapidity Distribution
Drell-Yan Mass Distribution
Heavy Quarks Total Cross Section
Jet Transverse Momentum Distribution
Drell-Yan Transverse Momentum Distribution
Heavy Quarks Production Single Quark Rapidity Distribution
Heavy Quarks Production Rapidity Distribution
Jets Rapidity Distribution
Dijets Invariant Mass and Rapidity Distribution
Photon Production
Black edge: New in NNDPF4.0
Q^2 (GeV^2)
x
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
10^{5}
10^{6}
10^{7}

Drell-Yan

Work out the cross section differential in the rapidity of the lepton pair

$$\frac{d\sigma^i}{dy} \propto A^i \mathcal{L}^i \qquad i = \mathrm{NC}(\gamma, \mathrm{Z}, \gamma\mathrm{Z}), \mathrm{CC}(\mathrm{W}^\pm)$$

NC DY (γ, Z) at LO

$$\left[A^\gamma \mathcal{L}^\gamma, A^Z \mathcal{L}^Z\right] = \left[\sum_q e_q^2 q\bar{q}, \sum_{q,q'} |V_{qq'}^{\mathrm{CKM}}| q\bar{q}'\right]$$

CC DY $(W^\pm)$ at LO

$$\left[A^{W^\pm} \mathcal{L}^{W^\pm}\right] = \left[\sum_q \left((g_V^q)^2 + (g_A^q)^2\right) q\bar{q}\right]$$

isospin $(p \to n)$: $u^p = d^n \qquad d^p = u^n \qquad \bar{u}^p = \bar{d}^n \qquad \bar{d}^p = \bar{u}^n$

deteron target approximated as the average of one proton and one neutron

different experiments measure different cross section combinations

$$\frac{\sigma_{pn}^Z}{\sigma_{pp}^Z} \approx \frac{4/9u\bar{d} + 1/9d\bar{u}}{4/9u\bar{u} + 1/9d\bar{d}} \to \frac{\bar{d}}{\bar{u}}$$

DY p/d asymmetry (NuSea, SeaQuest)

$$\frac{\sigma_{p\bar{p}}^{W^+}}{\sigma_{p\bar{p}}^{W^-}} \approx \frac{ud + \bar{d}\bar{u}}{du + \bar{u}\bar{d}} \to \frac{ud}{du}$$

$W^\pm$ asymmetry (CDF, D0)

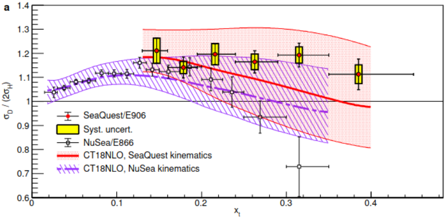
$$\sigma_{pp}^{W^+} \approx u\bar{d} + c\bar{s} \quad \sigma_{pp}^Z \approx u\bar{u} + d\bar{d} + s\bar{s} \to s, \bar{s}$$

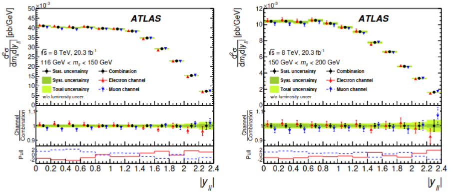
$W^\pm$ and Z production (ATLAS, CMS, LHCb)

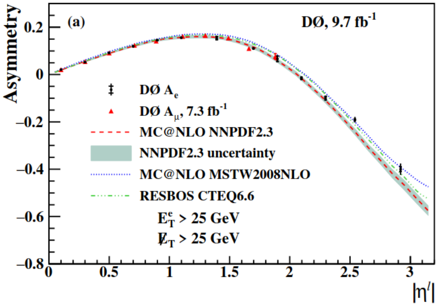
$$\frac{\sigma_{pp}^{W^+}}{\sigma_{pp}^{W^-}} \approx \frac{u\bar{d} + \bar{d}u}{d\bar{u} + \bar{u}d} \to \bar{u} - \bar{d}$$

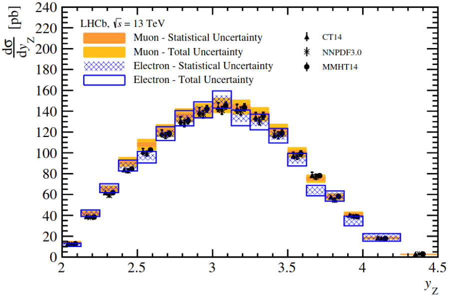
$W^\pm$ muon asymmetry (ATLAS, CMS)

Drell-Yan





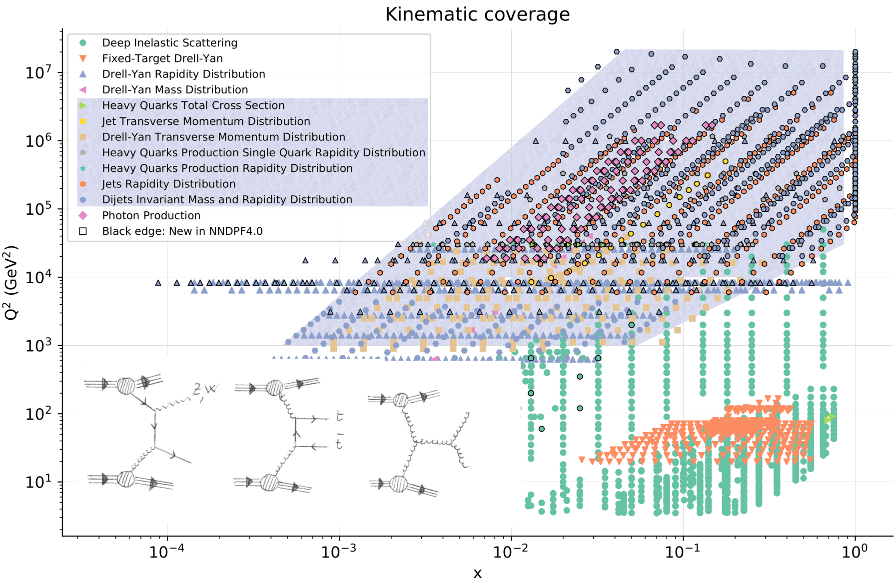




[SeaQuest, Nature 590 (2021) 7847] [ATLAS, JHEP 08 (2016) 009]

[D0, PRD 91 (2015) 032007] [LHCb, JHEP 09 (2016) 136]

Zp_T, $t\bar{t}$, Single-Inclusive Jet, and Dijet Production



Zp_T, $t\bar{t}$, Single-Inclusive Jet, and Dijet Production

Various differential distributions, all proportional to the gluon PDF at LO

Zp_T production: $\frac{d\sigma^2}{dp_T^Z dm_{\ell\ell}}$, $\frac{d\sigma^2}{dp_T^Z dy_Z}$ (ATLAS) $\frac{d\sigma}{dp_T^Z}$ (CMS)

need one final-state parton, then initial-state quark and gluon are on the same footing

wide p_T range, constraints on a wide x (typically intermediate) and Q^2 range

$t\bar{t}$ production: $\frac{d\sigma}{dp_T^t}$, $\frac{d\sigma}{dy^t}$, $\frac{d\sigma}{dy^{t\bar{t}}}$, $\frac{d\sigma}{dm^{t\bar{t}}}$ (ATLAS and CMS)

process initiated by two gluons in the initial state

differential cross sections reconstructed at parton level (additional systematics)

normalise by $\sigma_{\rm tot}^{t\bar{t}}$ (systematics largely cancel, but loose control on PDF shape)

particle-level cross sections are theoretically more complicated

wide rapidity range, constraints on the large-x region

single-inclusive jet and dijet production: $\frac{d\sigma^2}{dy dp_T}$, $\frac{d\sigma^2}{dy_{1,2} dm_{1,2}}$ (HERA, ATLAS and CMS)

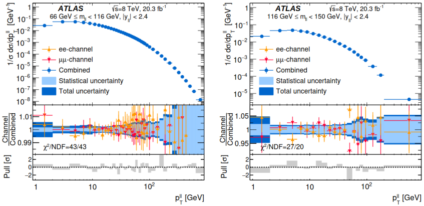
process initiated by two gluons in the initial state

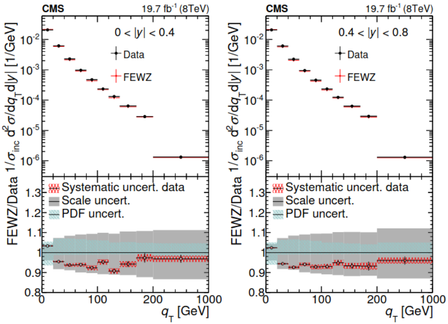
be careful with systematic uncertainties (mostly driven by jet energy reconstruction)

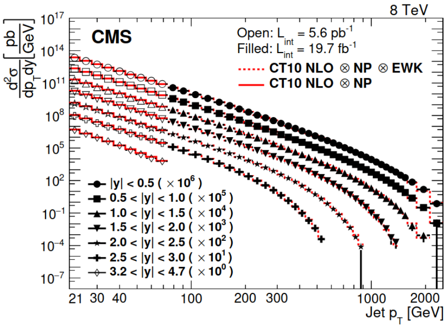
can also be measued in DIS

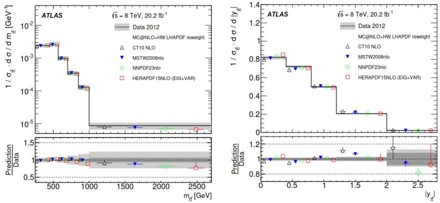
wide rapidity range, constraints on the large-x region

Zp_T, $t\bar{t}$, Single-Inclusive Jet, and Dijet Production









[ATLAS, EPJ C76 (2016) 291]

[CMS PhysLB 749 (2015) 187]

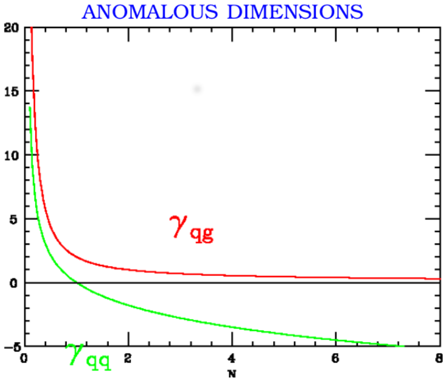
[CMS, JHEP 03 (2017) 156]

[ATLAS, EPJ C76 (2016) 538]

Scaling Violations and Heavy Flavour Production

Scale dependence in DIS
of singlet structure function

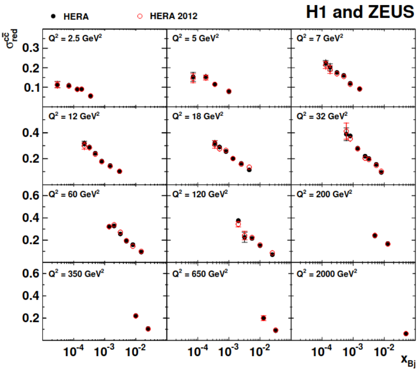
$$\frac{d}{d\ln(Q^2)}F_2^{\Sigma} \approx \frac{\alpha_s}{2\pi}\left[\gamma_{qq}f_{\Sigma} + 2n_f\gamma_{gq}f_g\right]$$



the gluon PDF can be determined at small x
from DIS scaling violations (from HERA)

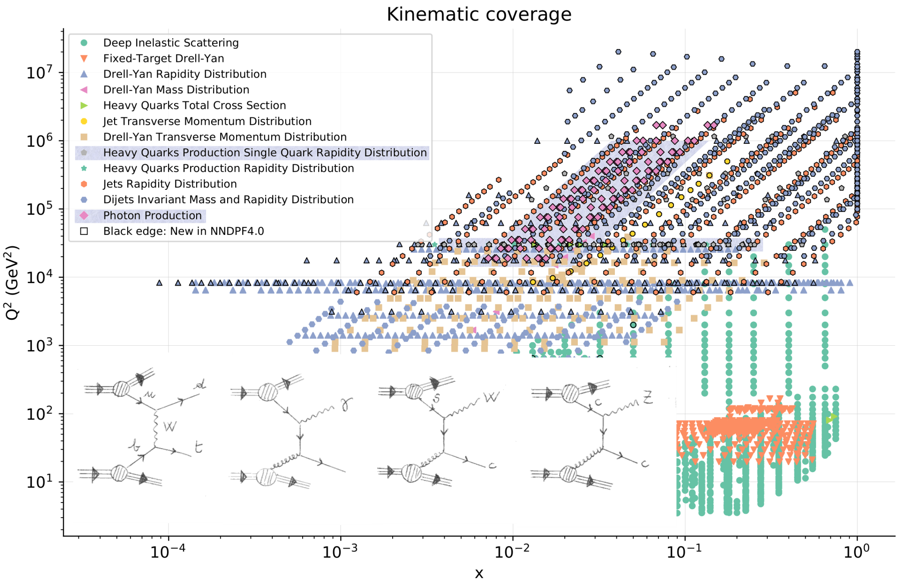
Heavy quark production in DIS
initiated by gluons

$$\sigma_{\rm red}^{c,b} = F_2^{c,b} - \frac{y^2}{1+(1-y)^2}F_L^{c,b}$$



the gluon PDF is determined by tagging
a c or a b quark in the final state (HERA)

Single t, Direct γ, W, Z+c-jets



Single t, Direct γ, W, Z+c-jets

Other processes, currently limited by experimental uncertainties

Single t production (t-channel): $\frac{d\sigma}{dp_T^t}$, $\frac{d\sigma}{dy^t}$ $\frac{d\sigma}{dp_T^{\bar{t}}}$, $\frac{d\sigma}{dy^{\bar{t}}}$ (ATLAS, CMS)

partonic cross sections similar to CC DIS; t reconstructed from Wb decay

potential sensitivity to $\bar{u}$ and $\bar{d}$, also through ratios of t and $\bar{t}$ production

potential currently limited by large experimental uncertainties

Prompt γ production: $\frac{d\sigma^2}{dE_T^\gamma y^\gamma}$ (ATLAS and CMS)

gluon-quark-initiated Compton scattering

potential sensitivity to the gluon PDF

potential currently limited by large experimental uncertainites

W, Z + charm-tagged jets: $\frac{d\sigma}{dy}$ (ATLAS and CMS)

$W + c$: sensitivity to strange PDF (and $s - \bar{s}$ asymmetry)

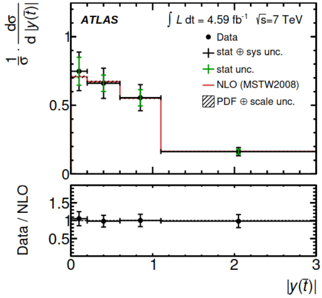
$W + Z$: sensitivity to charm PDF (including intrinsic charm and $c - \bar{c}$ asymmetry)

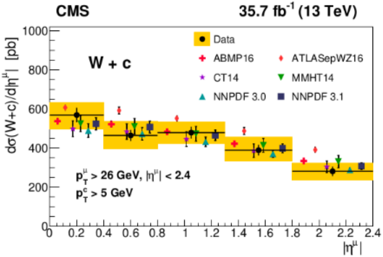
be careful with systematic uncertainties (due to jet tagging algorithm)

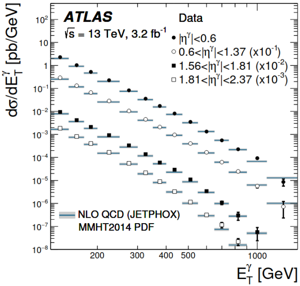
More exclusive processes: double gauge boson production, multijet production, ...

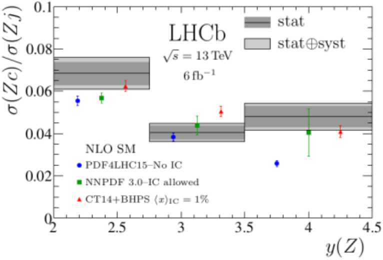
generally less precise and potentially contaminated by BSM physics

Single t, Prompt γ, $W, Z+c$-jets









[ATLAS, PRD 90 (2014) 112006; PLB 770 (2017) 473]

[CMS, EPJ C79 (2019) 269 ; LHCb, PRL 128 (2022) 082001]

All the data together

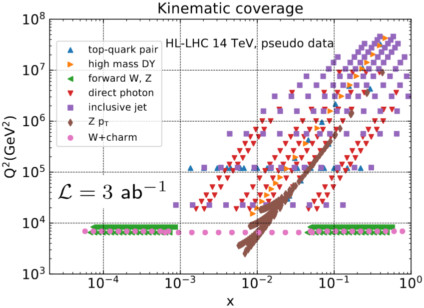
	Hadronic Process	Partonic Process	PDFs probed	x coverage
Lepton-nucleon	$\ell^{\pm}\{p,n\} \to \ell^{\pm} + X$	$\gamma^* q \to q$	$q, \bar{q}, g$	$x \gtrsim 0.01$
	$\ell^{\pm} n/p \to \ell^{\pm} + X$	$\gamma^* d/u \to d/u$	d/u	$x \gtrsim 0.01$
	$\nu(\bar{\nu})N \to \mu^-(\mu^+) + X$	$W^* q \to q'$	$q, \bar{q}$	$0.01 \lesssim x \lesssim 0.5$
	$\nu N \to \mu^- \mu^+ + X$	$W^* s \to c$	s	$0.01 \lesssim x \lesssim 0.2$
	$\bar{\nu} N \to \mu^+ \mu^- + X$	$W^* \bar{s} \to \bar{c}$	$\bar{s}$	$0.01 \lesssim x \lesssim 0.2$
	$e^{\pm} p \to e^{\pm} + X$	$\gamma^* q \to q$	$g, q, \bar{q}$	$0.0001 \lesssim x \lesssim 0.1$
	$e^+ p \to \bar{\nu} + X$	$W^+\{d,s\} \to \{u,c\}$	d, s	$x \gtrsim 0.01$
	$e^{\pm} p \to e^{\pm} c\bar{c} + X$	$\gamma^* c \to c, \gamma^* g \to c\bar{c}$	c, g	$0.0001 \lesssim x \lesssim 0.1$
	$e^{\pm} p \to jet(s) + X$	$\gamma^* g \to q\bar{q}$	g	$0.01 \lesssim x \lesssim 0.1$
Proton-(anti)proton	$pp \to \mu^+ \mu^- + X$	$u\bar{u}, d\bar{d} \to \gamma^*$	$\bar{q}$	$0.015 \lesssim x \lesssim 0.35$
	$pn/pp \to \mu^+ \mu^- + X$	$(u\bar{d})/(u\bar{u}) \to \gamma^*$	$\bar{d}/\bar{u}$	$0.015 \lesssim x \lesssim 0.35$
	$p\bar{p}(pp) \to jet(s) + X$	$gg, qg, qq \to 2jets$	g, q	$0.005 \lesssim x \lesssim 0.5$
	$p\bar{p} \to (W^{\pm} \to \ell^{\pm}\nu) + X$	$ud \to W^+, \bar{u}\bar{d} \to W^-$	$u, d, \bar{u}, \bar{d}$	$x \gtrsim 0.05$
	$pp \to (W^{\pm} \to \ell^{\pm}\nu) + X$	$u\bar{d} \to W^+, d\bar{u} \to W^-$	$u, d, \bar{u}, \bar{d}, (g)$	$x \gtrsim 0.001$
	$p\bar{p}(pp) \to (Z \to \ell^+\ell^-) + X$	$uu, dd(u\bar{u}, d\bar{d}) \to Z$	$u, d(g)$	$x \gtrsim 0.001$
	$pp \to (W + c) + X$	$gs \to W^- c, g\bar{s} \to W^+ \bar{c}$	$s, \bar{s}$	$x \sim 0.01$
	$pp \to (Z + c) + X$	$gc \to Zc, g\bar{c} \to Z\bar{c}$	$c, \bar{c}$	$x \sim 0.01$
	$pp \to t\bar{t} + X$	$gg \to t\bar{t}$	g	$x \sim 0.1$
	$pp \to t, \bar{t} + X$	$gq \to t, \bar{t}q$	u, d	$x \sim 0.01$
	$pp \to \gamma + X$	$gq \to \gamma q$	g	$x \sim 0.01$

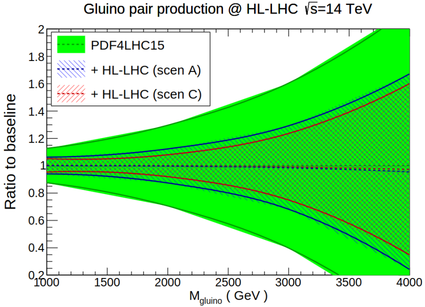
Overview of current PDF determinations

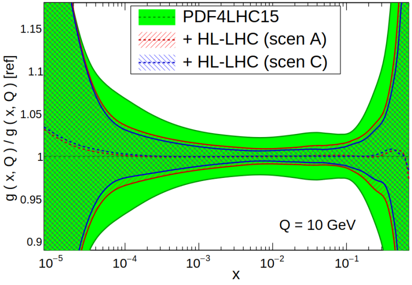
	NNPDF4.0	MSHT20	CT18	HERAPDF2.0	CJ22	ABMP16
Fixed-target DIS	✓	✓	✓	✗	✓	✓
JLAB	✗	✗	✗	✗	✓	✗
HERA I+II	✓	✓	✓	✓	✓	✓
HERA jets	✓	✗	✗	✓	✗	✗
Fixed target DY	✓	✓	✓	✗	✓	✓
Tevatron W, Z	✓	✓	✓	✗	✓	✓
LHC vector boson	✓	✓	✓	✗	✓	✓
LHC $W + c$ $Z + c$	✓	✗	✗	✗	✗	✗
Tevatron jets	✓	✓	✓	✗	✓	✗
LHC jets	✓	✓	✓	✗	✗	✗
LHC top	✓	✓	✗	✗	✗	✓
LHC single t	✓	✗	✗	✗	✗	✗
LHC prompt γ	✓	✗	✗	✗	✗	✗
statistical treatment	Monte Carlo	Hessian $\Delta\chi^2$ dynamical	Hessian $\Delta\chi^2$ dynamical	Hessian $\Delta\chi^2 = 1$	Hessian $\Delta\chi^2 = 1.645$	Hessian $\Delta\chi^2 = 1$
parametrisation	Neural Network	Chebyschev pol.	Bernstein pol.	polynomial	polynomial	polynomial
HQ scheme	FONLL	TR′	ACOT-χ	TR′	ACOT-χ	FFN
accuracy	aN^{3}LO	aN^{3}LO	NNLO	NNLO	NLO	NNLO
latest update	EPJ C82 (2022) 428	EPJ C81 (2021) 341	PRD 103 (2021) 014013	EPJ C82 (2022) 243	PRD 107 (2023) 113005	PRD 96 (2017) 014011

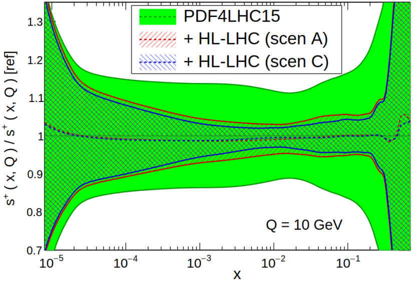
All PDF sets are available as (x, Q^2) interpolation grids through the LHAPDF library

Impact of future data: HL-LHC



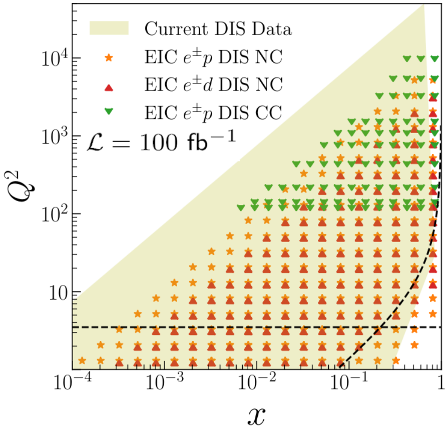






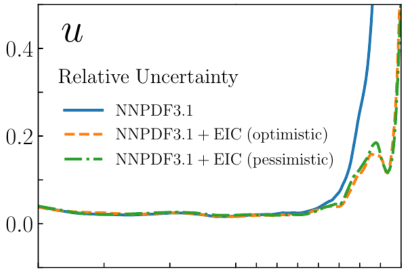
[EPJ C78 (2018) 962]

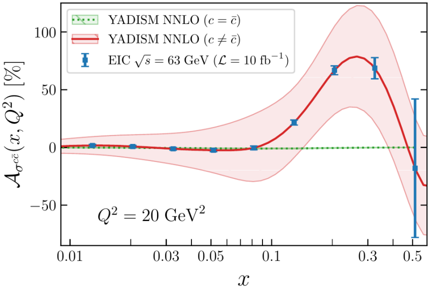
Impact of future data: EIC



$E_\ell \times E_p$ [GeV]: 18× 275; 10× 100; 5× 100

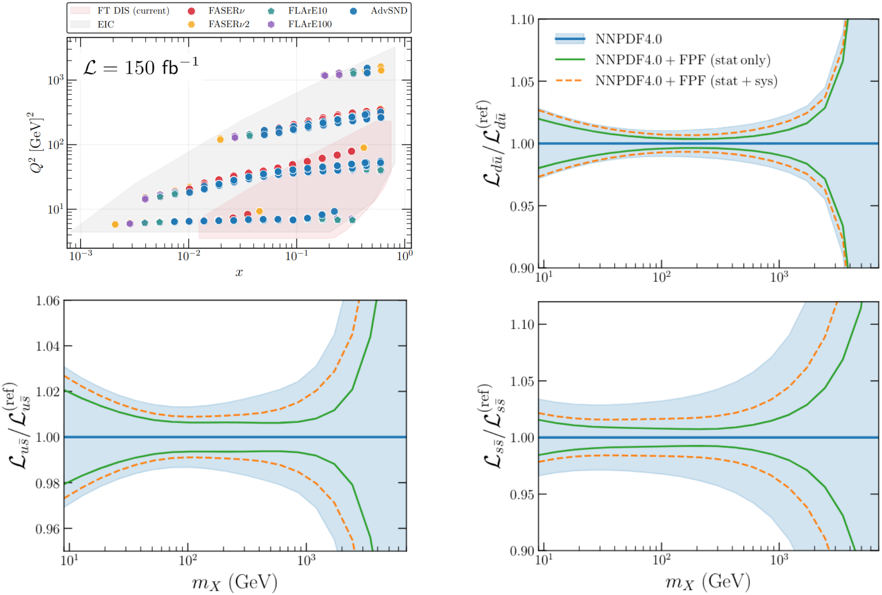
$$\mathcal{A}_{\sigma^{c\bar{c}}} = \frac{\sigma^{c}_{\rm red} - \sigma^{\bar{c}}_{\rm red}}{\sigma^{c\bar{c}}_{\rm red}}$$





[PRD 103 (2021) 096005; PRD 109 (2024) L091501]

Impact of future data: FPF



[EPJ C84 (2024) 369]

1.5 Summary of Lecture 1

Summary of Lecture 1

1. Parton Distribution functions are a key ingredient of the LHC program
   $\longrightarrow$ PDFs are often the dominant source of uncertainty in theoretical predictions
   $\longrightarrow$ limiting factor for precision and discovery
2. PDFs are related to physical observales via factorisation and evolution
   $\longrightarrow$ qualitative PDF features are driven by this theoretical framework
   $\longrightarrow$ valence peak follows from valence sum rules and kinematic vanishing
   $\longrightarrow$ small-x rise follows from rise of anomalous dimensions
   $\longrightarrow$ correlation of small-x rise and large-x depletion follow from momentum conservation
3. PDFs are determined from experimental data by means of parametric regression
   $\longrightarrow$ need to define data, theory, and methodology
4. Different physical observables constrain different PDF combinations
   $\longrightarrow$ fixed-target NC DIS: u and d
   $\longrightarrow$ fixed-target CC DIS: s and $\bar{s}$
   $\longrightarrow$ HERA NC and CC DIS: u, $\bar{u}$, d, $\bar{d}$, g (scaling violations and tagged DIS)
   $\longrightarrow$ fixed-target DY: u and d at large x
   $\longrightarrow$ collider DY: u, $\bar{u}$, d, $\bar{d}$, s
   $\longrightarrow$ collider DY+c: s (W) and c (Z)
   $\longrightarrow$ Zp_T, $t\bar{t}$, jets: g

Lecture 2: Theoretical and methodological accuracy in PDF determination