

# 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

## Lecture 1. What are PDFs and how we can determine them from data

- why Parton Distribution Functions?
- how are PDFs related to Physical Observables?
- how can we determine PDFs?
- which data constrain which PDFs and how?

## Lecture 2. Theoretical and methodological accuracy in PDF determination

- higher-order corrections and theory uncertainties
- heavy flavour schemes and intrinsic charm
- the photon PDF and electroweak corrections
- parametrisation, optimisation, uncertainty representation
- validation of uncertainties and benchmarks

## 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

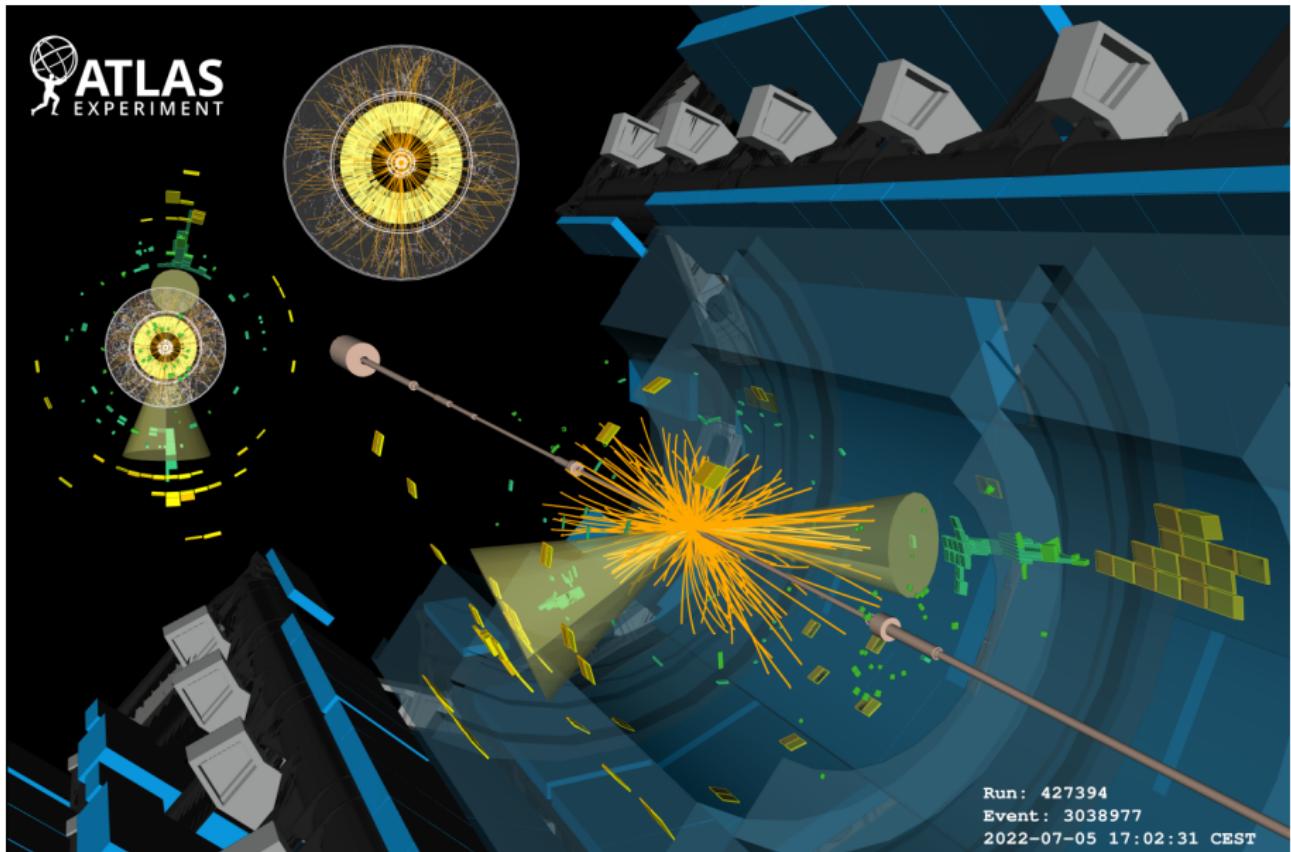
I will focus on the phenomenological determination of PDFs

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

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

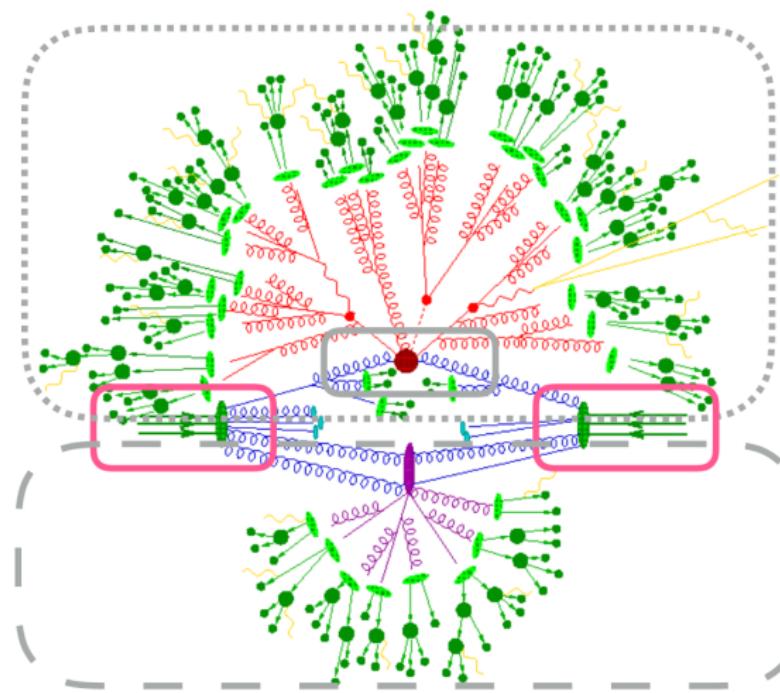
## 1.1 Why Parton Distribution Functions?

# First Collisions of LHC Run III



[Image credit: ATLAS collaboration]

# A Laboratory for Quantum Chromodynamics



Hard scattering of partons  
(Perturbative QCD+EW)

Parton Distribution Functions

Parton Showering  
and Hadronisation

Multi-Parton Interactions  
Underlying Events

[Plot by courtesy of SHERPA]

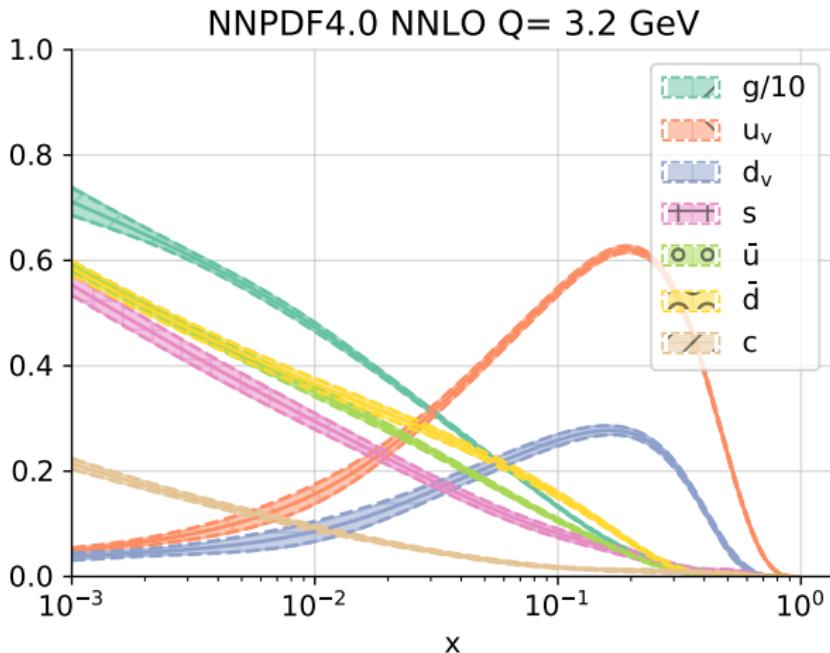
$$\sigma(\tau, Q^2, \mathbf{k}) = \sum_{ij} \int_{\tau}^1 \frac{dz}{z} \hat{\sigma}_{ij} \left( \frac{\tau}{z}, \alpha_s(Q^2), \mathbf{k} \right) \mathcal{L}_{ij}(z, Q^2) \quad \mathcal{L}_{ij}(z, Q^2) = (f_i^{h_1} \otimes f_j^{h_2})(z, Q^2)$$

# Parton Distribution Functions

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

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

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



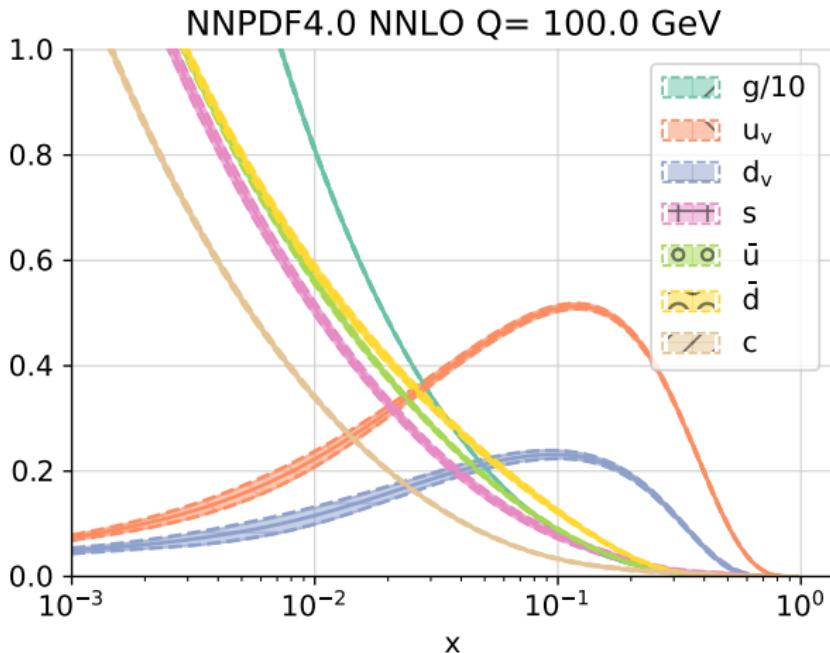
[Plot from the PDG Review of Particle Physics]

# Parton Distribution Functions

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

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

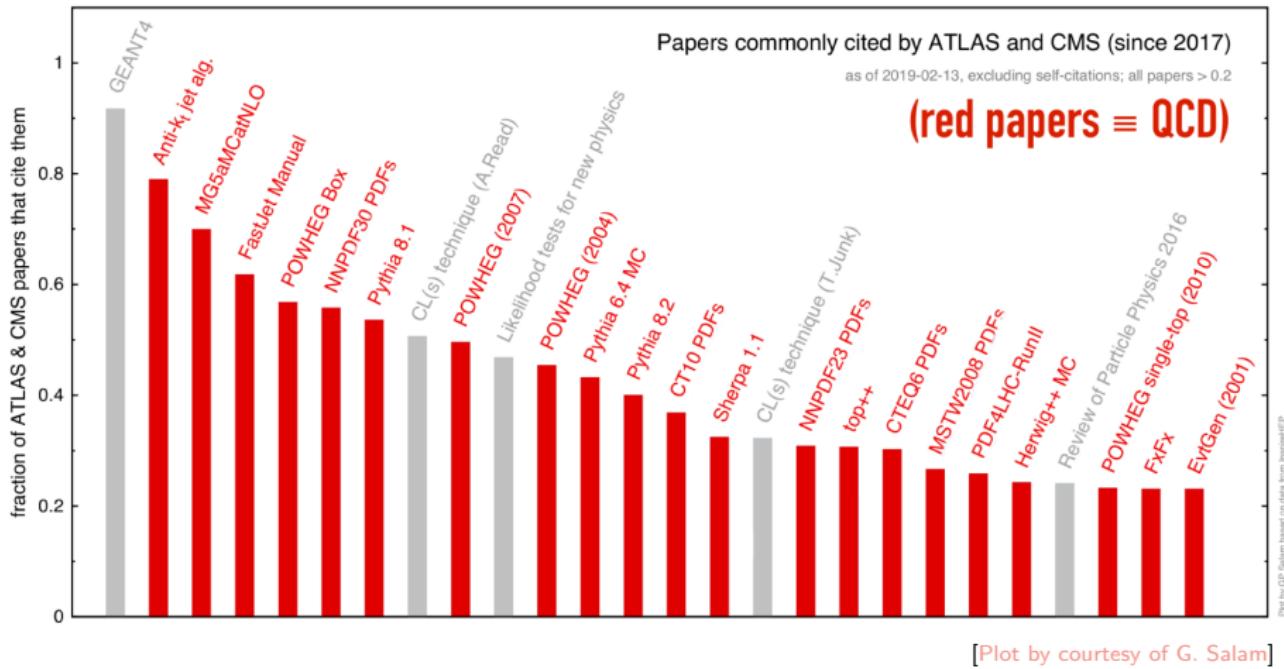
Dependence on  $x$  is non-perturbative (fit); dependence on  $Q^2$  is perturbative (DGLAP)



[Plot from the PDG Review of Particle Physics]

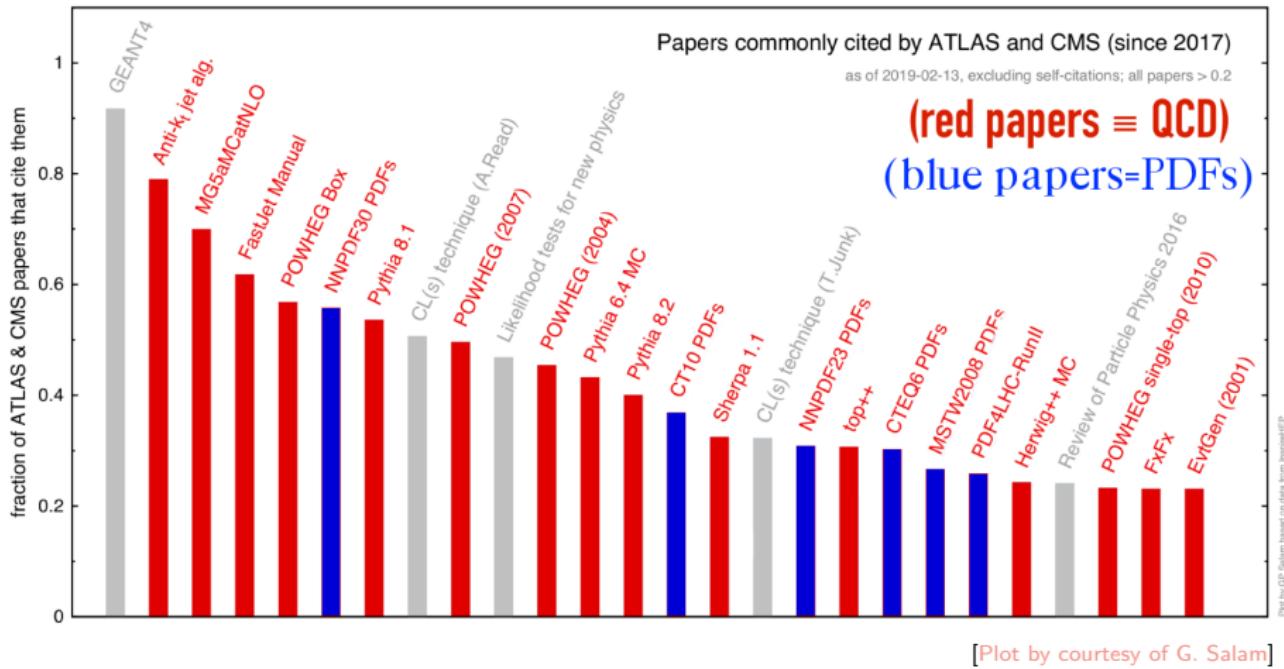
# LHC, QCD and PDFs

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



# LHC, QCD and PDFs

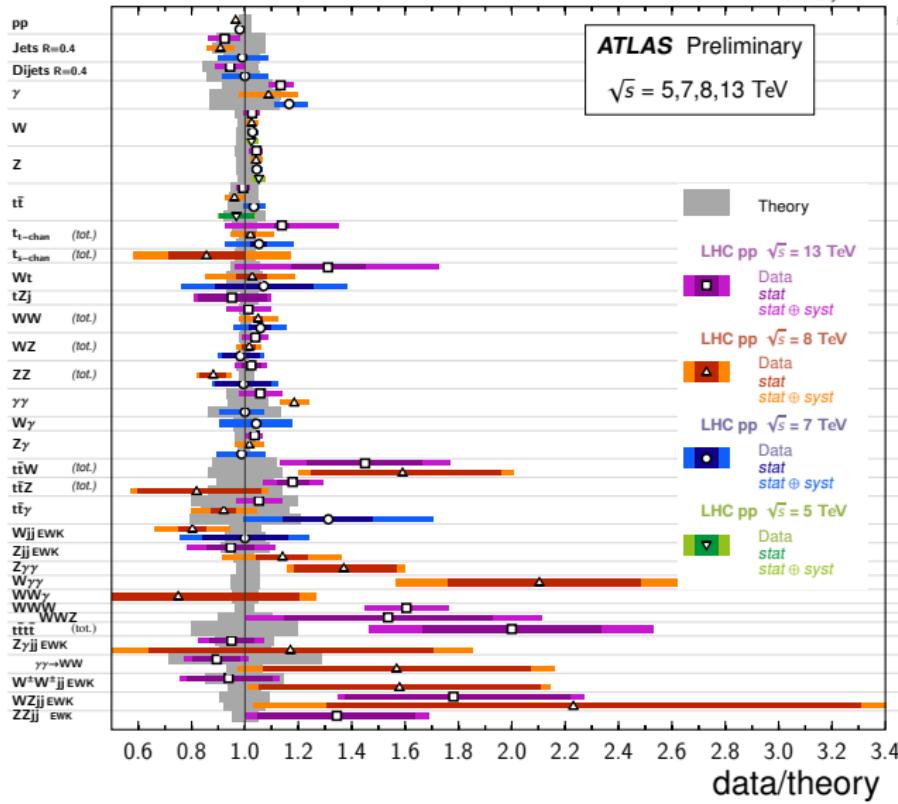
The LHC is a Proton Collider – Any interaction contains a strong interaction  
Quantum Chromodynamics (QCD) is the main actor  
Within QCD, Parton Distribution Functions (PDFs) play a leading role



# Physics at the LHC as Precision Physics

## Standard Model Production Cross Section Measurements

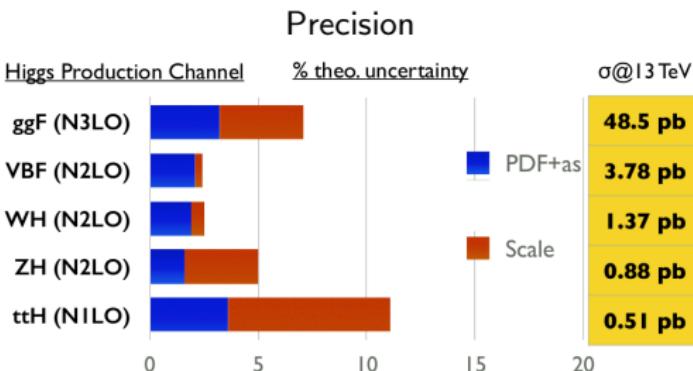
Status:  
February 2022



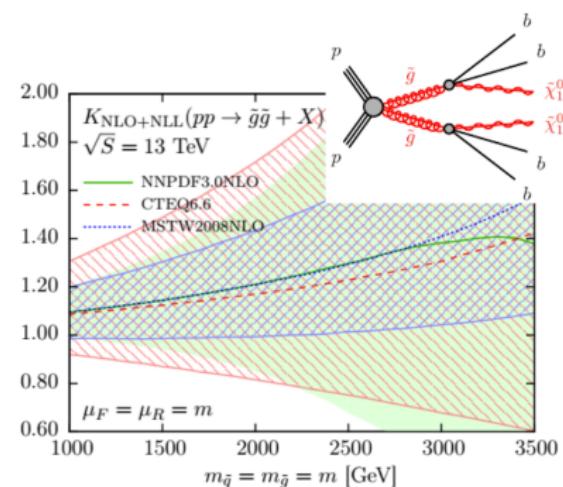
[Plot from ATLAS Collaboration web page]

# PDFs as a Tool: Making Predictions with PDFs

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



### Discovery



Channel	$m_{W^+} - m_{W^-}$ [MeV]	Stat. Unc.	Muon Unc.	Elec. Unc.	Recoil Unc.	Bckg. Unc.	QCD Unc.	EW Unc.	PDF Unc.	Total Unc.
$W \rightarrow e\nu$	-29.7	17.5	0.0	4.9	0.9	5.4	0.5	0.0	24.1	30.7
$W \rightarrow \mu\nu$	-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]

[EPJC 76 (2016) 53]

## Higgs boson characterisation

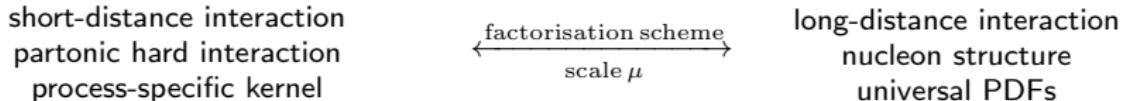
Determination of SM parameters, such as the mass of the  $W$  boson

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



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

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

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

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

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

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

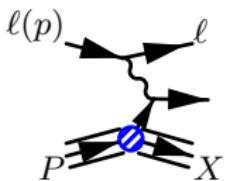
$$C_{Ii}(y, \alpha_s) = \sum_{k=0} a_s^k C_{Ii}^{(k)}(y) \quad \hat{\sigma}_{ij}(y, \alpha_s) = \sum_{k=0} a_s^k \hat{\sigma}_{ij}^{(k)}(y) \quad a_s = \alpha_s/(4\pi)$$

- ④ Because of factorisation, all of these quantities depend on  $\mu^2$ ; 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:  $y \rightarrow p = yP$

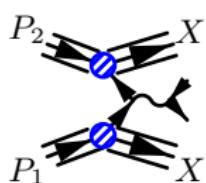
$P$ : proton momentum

$q$ : 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:  $x_{1,2} \rightarrow p_{1,2} = x_{1,2} P_{1,2}$

$P_{1,2}$ : proton momenta

$M^2$ : final state mass

$p_{1,2}$  parton momenta

# Theoretical constraints

- ① Momentum sum rule (momentum conservation)

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

- ② Valence sum rules (baryon number conservation)

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

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

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

- ③ Isospin symmetry of the strong interaction

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

- 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

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

- ② They are often written in a convenient PDF basis

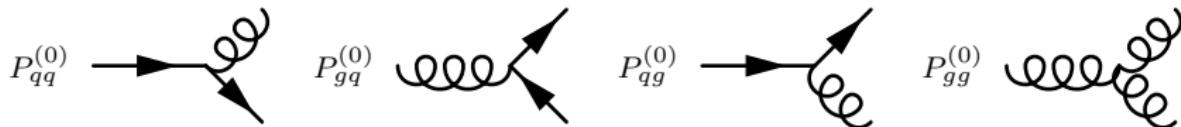
$$f_{\pm} = (f_q \pm f_{\bar{q}}) - (f_{q'} \pm f_{\bar{q}'}) \quad f_v = \sum_i^{n_f} (f_q - f_{\bar{q}}) \quad f_{\Sigma} = \sum_i^{n_f} (f_q + f_{\bar{q}})$$

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

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

- ③ The splitting functions  $P$  can be computed perturbatively

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



# Splitting Functions and Anomalous Dimensions

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

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

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

How many different anomalous dimensions are there?

LO:  $\gamma_{qq} = \gamma_{\pm, v} \rightarrow 4$  independent splitting functions

NLO:  $\gamma_{qq} \neq \gamma_+ \neq \gamma_- \rightarrow 6$  independent splitting functions

NNLO:  $\gamma_- \neq \gamma_v \rightarrow 7$  independent splitting functions

Which PDF combinations evolve independently?

LO:  $f_g$ ,  $f_\Sigma$ , and any  $2n_f - 1$  linear combinations of  $f_q$  and  $f_{\bar{q}}$

NLO:  $f_g$ ,  $f_\Sigma$ , any  $n_f - 1$  linear combinations of  $f_q - f_{\bar{q}}$ , and of  $f_q + f_{\bar{q}}$

NNLO: as NLO, and  $f_V = \sum_q^n (f_q - f_{\bar{q}})$

A common choice

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

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

$$T_3 = f_{u+} - f_{d+} \quad T_8 = f_{u+} + f_{d+} - 2f_{s+} \quad T_{15} = f_{u+} + f_{d+} + f_{s+} - 3f_{c+} \dots$$

$$V_3 = f_{u-} - f_{d-} \quad V_8 = f_{u-} + f_{d-} - 2f_{s-} \quad V_{15} = f_{u-} + f_{d-} + f_{s-} - 3f_{c-} \dots$$

# Anomalous dimensions: perturbative accuracy

LO (1977)      NLO (1980)

[NPB 126 (1977) 298]

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

$$\gamma_{\text{ps}}^{(0)}(N) = 0$$

$$\gamma_{\text{qg}}^{(0)}(N) = 2n_f(\mathbf{N}_- + 4\mathbf{N}_+ - 2\mathbf{N}_{+2} - 3)S_1$$

$$\gamma_{\text{gg}}^{(0)}(N) = 2C_F(2\mathbf{N}_{-2} - 4\mathbf{N}_- - \mathbf{N}_+ + 3)S_1$$

$$\gamma_{\text{gg}}^{(0)}(N) = C_A \left( 4(\mathbf{N}_{-2} - 2\mathbf{N}_- - 2\mathbf{N}_+ + \mathbf{N}_{+2} + 3)S_1 - \frac{11}{3} \right) + \frac{2}{3}n_f$$

$$\begin{aligned} \gamma_{\text{ps}}^{(1)}(N) &= 4C_F n_f \left( \frac{20}{9}(\mathbf{N}_{-2} - \mathbf{N}_-)S_1 - (\mathbf{N}_+ - \mathbf{N}_{+2}) \left[ \frac{56}{9}S_1 + \frac{8}{3}S_2 \right] + (1 - \mathbf{N}_+) \left[ 8S_1 - 4S_2 \right] \right. \\ &\quad \left. - (\mathbf{N}_- - \mathbf{N}_+) \left[ 2S_1 + S_2 + 2S_3 \right] \right) \end{aligned}$$

$$\begin{aligned} \gamma_{\text{qg}}^{(1)}(N) &= 4C_A n_f \left( \frac{20}{9}(\mathbf{N}_{-2} - \mathbf{N}_-)S_1 - (\mathbf{N}_- - \mathbf{N}_+) \left[ 2S_1 + S_2 + 2S_3 \right] - (\mathbf{N}_+ - \mathbf{N}_{+2}) \left[ \frac{218}{9}S_1 \right. \right. \\ &\quad \left. + 4S_{1,1} + \frac{44}{3}S_2 \right] + (1 - \mathbf{N}_+) \left[ 27S_1 + 4S_{1,1} - 7S_2 - 2S_3 \right] - 2(\mathbf{N}_- + 4\mathbf{N}_+ - 2\mathbf{N}_{+2} - 3) \left[ S_{1,-2} \right. \\ &\quad \left. + S_{1,1,1} \right] \Big) + 4C_F n_f \left( 2(\mathbf{N}_- - \mathbf{N}_{+2}) \left[ 5S_1 + 2S_{1,1} - 2S_2 + S_3 \right] - (1 - \mathbf{N}_+) \left[ \frac{43}{2}S_1 + 4S_{1,1} - \frac{7}{2}S_2 \right] \right. \\ &\quad \left. + (\mathbf{N}_- - \mathbf{N}_+) \left[ 7S_1 - \frac{3}{2}S_2 \right] + 2(\mathbf{N}_- + 4\mathbf{N}_+ - 2\mathbf{N}_{+2} - 3) \left[ S_{1,1,1} - S_{1,2} - S_{2,1} + \frac{1}{2}S_3 \right] \right) \end{aligned} \quad (3.7)$$

$$\begin{aligned} \gamma_{\text{gg}}^{(1)}(N) &= 4C_A C_F \left( 2(2\mathbf{N}_{-2} - 4\mathbf{N}_- - \mathbf{N}_+ + 3) \left[ S_{1,1,1} - S_{1,-2} - S_{1,2} - S_{2,1} \right] + (1 - \mathbf{N}_+) \left[ 2S_1 \right. \right. \\ &\quad \left. - 13S_{1,1} - 7S_2 - 2S_3 \right] + (\mathbf{N}_{-2} - 2\mathbf{N}_- + \mathbf{N}_+) \left[ S_1 - \frac{22}{3}S_{1,1} \right] + 4(\mathbf{N}_- - \mathbf{N}_+) \left[ \frac{7}{9}S_1 + 3S_2 + S_3 \right] \\ &\quad + (\mathbf{N}_+ - \mathbf{N}_{+2}) \left[ \frac{44}{9}S_1 + \frac{8}{3}S_2 \right] \Big) + 4C_F n_f \left( (\mathbf{N}_{-2} - 2\mathbf{N}_- + \mathbf{N}_+) \left[ \frac{4}{3}S_{1,1} - \frac{20}{9}S_1 \right] - (1 - \mathbf{N}_+) \left[ 4S_1 \right. \right. \\ &\quad \left. - 2S_{1,1} \right] \Big) + 4C_F^2 \left( (2\mathbf{N}_{-2} - 4\mathbf{N}_- - \mathbf{N}_+ + 3) \left[ 3S_{1,1} - 2S_{1,1,1} \right] - (1 - \mathbf{N}_+) \left[ S_1 - 2S_{1,1} + \frac{3}{2}S_2 \right. \right. \\ &\quad \left. - 3S_3 \right] - (\mathbf{N}_- - \mathbf{N}_+) \left[ \frac{5}{2}S_1 + 2S_2 + 2S_3 \right] \Big) \end{aligned}$$

$$\begin{aligned} \gamma_{\text{gg}}^{(1)}(N) &= 4C_A n_f \left( \frac{2}{3} - \frac{16}{3}S_1 - \frac{23}{9}(\mathbf{N}_{-2} + \mathbf{N}_{+2})S_1 + \frac{14}{3}(\mathbf{N}_- + \mathbf{N}_+)S_1 + \frac{2}{3}(\mathbf{N}_- - \mathbf{N}_+)S_2 \right) \\ &\quad + 4C_A^2 \left( 2S_{-3} - \frac{8}{3}S_1 + 2S_3 - (\mathbf{N}_{-2} - 2\mathbf{N}_- - 2\mathbf{N}_+ + \mathbf{N}_{+2} + 3) \left[ 4S_{1,-2} + 4S_{1,2} + 4S_{2,1} \right] \right. \\ &\quad \left. + \frac{8}{3}(\mathbf{N}_+ - \mathbf{N}_{+2})S_2 - 4(\mathbf{N}_- - 3\mathbf{N}_+ + \mathbf{N}_{+2} + 1) \left[ 3S_2 - S_3 \right] + \frac{109}{18}(\mathbf{N}_- + \mathbf{N}_+)S_1 + \frac{61}{3}(\mathbf{N}_- \right. \\ &\quad \left. - \mathbf{N}_+)S_2 \right) + 4C_F n_f \left( \frac{1}{2} + \frac{2}{3}(\mathbf{N}_{-2} - 13\mathbf{N}_- - \mathbf{N}_+ - 5\mathbf{N}_{+2} + 18)S_1 + (3\mathbf{N}_- - 5\mathbf{N}_+ + 2)S_2 \right. \\ &\quad \left. - 2(\mathbf{N}_- - \mathbf{N}_+)S_3 \right). \end{aligned}$$

Numerical solution (LO, NLO, NNLO and  $\alpha N^3 LO$ ) 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]

$$\begin{aligned}
g_{\mu}^{(2)}(N) = & 16C_F C_F n_f \left[ \frac{1}{3}(4N_+ - N_- - N_{+4} - N_{-2}) - 35(S_0^+ - S_{1,-1} - S_{1,1,-2} + S_{1,1,1,1}) \right. \\
& - S_{1,1,2,1} ] + [N_- - 2N_+] \left[ \frac{371}{102} S_{1,-1} - \frac{6761}{324} S_{1,-2} - \frac{5}{2} S_{1,-2} - \frac{56}{9} S_2 - \frac{20}{9} S_2 \right] \\
& - (N_- - 2N_+ - N_{+4} - N_{-2}) \left[ \frac{8}{11} S_{1,-3} - \frac{1}{3} S_{1,-4} - \frac{2}{3} S_{2,-1} \right] + [N_+ - N_{-2}] \left[ \frac{10279}{162} S_1 \right. \\
& + \frac{106}{9} S_{1,-2} + \frac{151}{54} S_{1,-3} + \frac{9}{2} S_{1,-4} + 25S_{2,-1} + \frac{2299}{54} S_2 + \frac{28}{9} S_{2,-2} + \frac{83}{162} S_3 + \frac{3}{5} S_{3,-1} \\
& + (1 - N_{+4}) \left[ \frac{4}{3} S_{1,-2} - \frac{25}{24} S_1 - \frac{50}{3} S_{1,-2} - \frac{29}{12} S_2 - \frac{1165}{36} S_{1,-1} + 5S_{2,-2} - \frac{3}{4} S_{2,-1} + S_{2,1,1,-2} + \frac{3}{4} S_{2,2} \right. \\
& - \frac{37}{2} S_3 - 4S_{3,-1} - 2S_{3,-2} - 3S_{3,-3} - 10S_4 - 7S_5 ] - (N_+ - N_{-2}) \left[ \frac{1}{2} S_{1,-3} + 3S_{1,-4} - \frac{3}{4} S_{1,1,1,-2} + \frac{9}{4} S_{1,1,2} \right] \\
& + [N_+ - N_{-2}] \left[ \frac{121}{12} S_1 + \frac{16}{3} S_{1,-2} + \frac{437}{36} S_{1,-3} - \frac{13}{6} S_{1,-4} - \frac{3565}{108} S_2 - 3S_{2,-1} + 3S_{2,-2} \right. \\
& - \frac{479}{36} S_{2,1,1} + 2S_{2,1,2} + \frac{11}{6} S_{2,1,3} - \frac{25}{12} S_{2,1,4} + 2S_{2,1,5} + S_{2,1,6} + \frac{269}{36} S_3 + \frac{55}{6} S_4 + \frac{24}{9} S_5 \\
& + \frac{59}{162} S_{5,1} + S_{5,1,1,1} + 2S_{5,1,1,2} + 4S_5 \right] + 16C_F e^{\gamma_E} \left[ \frac{7}{9} (N_+ - N_- - N_{+4} - N_{-2}) [S_{1,1,1,1} + 5S_1 \right. \\
& + \frac{2}{3} S_2] + (N_+ - N_{-2}) [2S_1 - \frac{19}{2} S_2 - \frac{2}{3} S_3] - (1 - N_+) [2S_1 - 2S_2 - 2S_3 - \frac{2}{3} S_4] \\
& + (N_+ - N_{-2}) \left[ \frac{77}{24} S_1 - \frac{63}{34} S_{1,-1} - \frac{37}{27} S_{1,-2} + \frac{1}{3} S_{1,1,1,1} \right] [S_{1,1,1,1} + 5S_{1,1,2,1} + \frac{29}{6} S_1 \\
& - 2S_{3,1,1,1,1} - S_1] + 16C_F e^{\gamma_E} [N_+ - N_{-2}] \left[ \frac{16}{13} S_{1,-1} + \frac{13}{54} S_2 - \frac{163}{12} S_3 - \frac{85}{9} S_{4,1} - \frac{28}{3} S_{5,1} - \frac{22}{3} S_5 \right. \\
& + 4S_{2,1,2} - \frac{4}{3} S_{2,1,3,1} + 3S_{2,1,4} - \frac{22}{3} S_4 - \frac{1}{3} S_{2,1,5} - N_{+4} - N_{-2}] \left[ 35(S_1 - S_{1,1,-1} - S_{1,1,1,1} \right. \\
& + S_{1,1,1,1,1} - \frac{1}{2} S_2] + (N_+ - N_{-2}) \left[ \frac{523}{12} S_1 - \frac{23}{108} S_2 - \frac{5}{6} S_3 + \frac{46}{9} S_4 - \frac{523}{27} S_5 \right. \\
& + (1 + N_{+4}) \left[ \frac{27}{28} S_1 - \frac{121}{9} S_2 - \frac{2707}{108} S_3 - \frac{497}{18} S_4 - \frac{63}{2} S_5 + \frac{55}{6} S_6 + \frac{181}{12} S_7 - S_8 \right] \\
& - S_{2,1,1,1} + 5S_{1,1,1} ] + (N_+ - N_{-2}) \left[ \frac{47}{9} S_{1,-1} - \frac{271}{108} S_{1,-2} - \frac{216}{72} S_{1,-3} - \frac{7}{12} S_{1,-4} - \frac{1}{2} S_{1,-5} \right. \\
& + 17S_{2,1} + 2S_{2,1,1,1} - 2S_{2,1,2,1} + 2S_{2,1,3,1} + 2S_{2,1,4,1} - 2S_{2,1,5,1} - S_{2,1,6,1} + 4S_{2,1,7,1} \\
& - 4S_5 \Big) \Big].
\end{aligned}$$

$$\begin{aligned}
g_{\mu}^{(2)}(N) = & 16C_F C_F n_f \left[ \frac{31}{2} (N_+ - 4N_- - 2N_{+4} - 2N_{-2}) - \frac{3997}{3} S_1 - \frac{11}{2} S_2 - 4S_{3,1,-3} \right. \\
& - \frac{3}{2} S_{1,-3} - \frac{9}{2} S_{1,-2} - 3S_{1,-1,2} - \frac{2}{2} S_{1,-1,3} - 2S_{1,-1,4} - \frac{2405}{96} S_{1,-5} + 26S_1 + 6S_{1,1,-3} \\
& + 3S_{1,1,5} + \frac{5}{2} S_{1,1,2} - 6S_{1,1,3,1} - \frac{9}{2} S_{1,1,4} - 6S_{1,1,5,1,1} - \frac{128}{3} S_{1,1,1,1,1} - 3S_{1,1,1,1,1,1} - 3S_{1,1,1,2} \\
& - \frac{35}{12} S_{1,1,2,1} + 3S_{1,1,2,1,1} + 3S_{1,1,2,1,2} - \frac{15}{4} S_{1,1,2,1,3} + 6S_{1,1,2,1,4} - \frac{3}{2} S_{1,1,2,1,5} \\
& + \frac{3}{2} S_{1,1,2,1,6} + 3S_{1,1,2,1,7} - \frac{5}{2} S_{1,1,2,1,8} + 6S_{2,1,-2,1} - \frac{49}{2} S_{2,1,-2,2} + 6S_{1,1,-2} - 6S_{1,1,1,-2} + 3S_{1,1,2,-2} \\
& + 2S_{2,1,1,1,1} + \frac{49}{2} S_{2,1,1,2,1} - \frac{3}{2} S_{2,1,1,2,2} - \frac{55}{2} S_{2,1,1,2,3} + [N_+ - N_{-2}] \left[ \frac{11}{12} S_1 \right. \\
& - 4S_{5,1} - \frac{108}{108} S_{5,1,1} + 2S_{5,1,1,1} + \frac{4}{3} S_{5,1,1,2} + 2S_{5,1,1,3} + [N_+ - N_{-2}] \left[ \frac{8453}{192} S_1 \right. \\
& - \frac{71}{72} S_1 S_3 - \frac{1}{3} S_1 S_4 - \frac{9}{2} S_1 S_5 - \frac{1}{3} S_1 S_6 - \frac{1}{2} S_1 S_7 - \frac{1}{2} S_1 S_8 - \frac{1}{2} S_1 S_9 - \frac{1}{2} S_1 S_{10} \\
& + \frac{55}{6} S_{1,1,3} + \frac{23}{6} S_{1,1,4} + \frac{4}{3} S_{1,1,5} - \frac{23}{72} S_{1,1,6} + \frac{55}{8} S_2 + 9S_3 - \frac{21}{8} S_4 - \frac{269}{26} S_5 - \frac{45}{36} S_6 \\
& + \frac{1}{6} S_{7,1} + \frac{23}{6} S_{1,1,1,1} + \frac{4}{3} S_{1,1,2,1} - \frac{235}{72} S_{1,1,2,2} + \frac{55}{8} S_2 + 9S_3 - \frac{21}{8} S_4 - \frac{269}{26} S_5 - \frac{45}{36} S_6 \\
& + \frac{1}{2} S_{2,2} ] + (N_+ - 4N_- - 2N_{+4} - 2N_{-2}) \left[ \frac{81}{32} S_1 - S_{1,-4} + 5S_{1,-5} - \frac{5}{2} S_{1,-6} + 2S_{1,-7} + 4S_{1,-8} \right. \\
& + \frac{1}{2} S_{2,2} ] + (N_+ - 4N_- - 2N_{+4} - 2N_{-2}) \left[ \frac{81}{32} S_1 - S_{1,-4} + 5S_{1,-5} - \frac{5}{2} S_{1,-6} + 2S_{1,-7} + 4S_{1,-8} \right. \\
& + \frac{1}{2} S_{2,2} \Big) \Big].
\end{aligned}$$

aN<sup>3</sup>LO (2020 - ongoing)

# Anomalous dimensions: perturbative accuracy

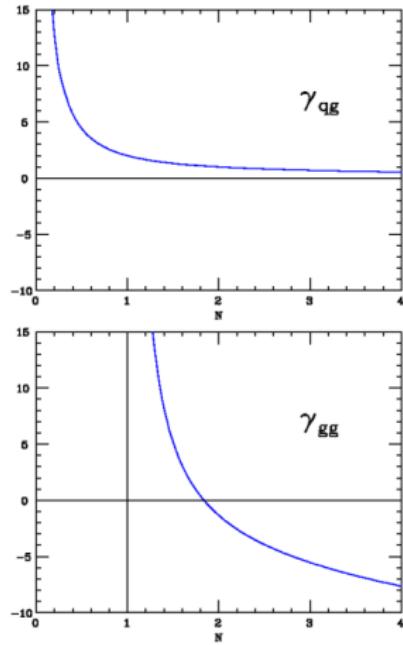
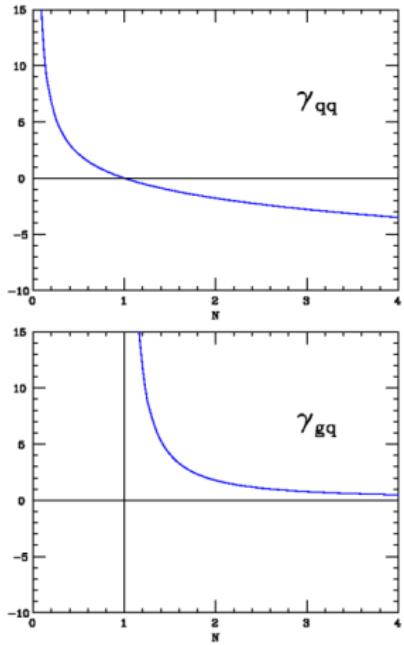
## NNLO cont'd (2004)

[NPB 691 (2004) 129]

$$\begin{aligned}
& -\frac{11}{2}S_{1,-4} + \frac{49}{6}S_{1,-3} + S_{1,-2,-2} - 10S_{1,-1,-2} + \frac{109}{12}S_{1,-2,-3} - \frac{3}{2}S_{1,-2,-1} + 2S_{1,-2,-2} - \frac{3379}{216}S_{1,-1} \\
& + 8S_{1,-3,1} + 13S_{1,\zeta_5} + 12S_{1,1,-3} + \frac{1}{9}S_{1,-1,-2} + 2S_{1,1,1,-1} + \frac{65}{24}S_{1,1,-1,-2} - \frac{43}{6}S_{1,1,1,1} \\
& - 4S_{1,1,2} + \frac{55}{2}S_{1,1,2,-1} - 4S_{1,1,2,1} + \frac{71}{12}S_{1,2,-1} + 5S_{1,2,-2} + \frac{55}{6}S_{1,2,-1,-2} + 12S_{1,2,1,-1} - 4S_{1,2,1,1} + 6S_{1,2,2} \\
& + \frac{11}{2}S_{1,3,-1} + 4S_{1,3,1} - \frac{3}{2}S_{1,4,-1} + \frac{395}{54}S_{2,-2} - 7S_{2,-3,-1} + \frac{11}{6}S_{2,-2,-1} + 2S_{2,-2,-2} - 2S_{2,1,1} \\
& + \frac{17}{3}S_{2,1,2} + 3S_{2,1,2,-1} - \frac{1}{3}S_{2,1,2,1} + 3S_{2,2,1,-1} + 4S_{1,1,1,-1} - 4S_{1,1,2} + [N_{-1}] \left[ 6S_{2,1,-1} - 8S_{2,-1,1} \right] \\
& + [N_{-}, N_{+}] \left[ 37595S_{1,-1} - 125S_{1,\zeta_5} - \frac{31}{6}S_{1,-3} - 14S_{1,-2} + 2S_{1,-1,-1} - 54S_{1,1} + \frac{50}{3}S_{1,1,-2} \right. \\
& \left. + [N_{-}, N_{+}] \right] \left[ \frac{1}{12}S_{1,1} - 2 + \frac{229}{36}S_{1,2,-1} + \frac{16}{12}S_{1,2,-2} - \frac{27}{2}S_{1,2,-3} - \frac{1}{2}S_{1,2,-4} + 31S_{1,2,-1} \right. \\
& + \frac{11}{18}S_{1,3,-1} - \frac{5}{6}S_{1,3,1,1} + \frac{37}{36}S_{1,3,2,-1} - \frac{3}{2}S_{1,3,2,-2} - \frac{2}{3}S_{1,3,2,-3} - 31S_{1,3,1,-1} - 95S_{1,3,-2} - \frac{1}{2}S_{1,3,1,1} + S_{1,3,1,2} \\
& - 18S_{1,3,2,-1} + \frac{6}{2}S_{1,3,2,1} - 4S_{1,3,2,1,-1} - \frac{3}{2}S_{1,3,2,1,-2} - \frac{2}{3}S_{1,3,2,1,-3} - 31S_{1,3,1,1} + 8S_{1,3,1,2} + S_{1,3,1,3} \\
& - \frac{13}{2}S_{1,3,-1} - 8S_{1,3,1} + [N_{-}, N_{+}] \left[ \frac{4}{3}S_{1,4,-1} - 8S_{1,4,-2} - 8S_{1,4,-3} - 10S_{1,4,-4} - \frac{19}{2}S_{1,4,-5} + \frac{4}{3}S_{1,4,-6} \right. \\
& \left. - \frac{37}{3}S_{1,4,-7} + \frac{145}{4}S_{1,4,-8} - \frac{584}{9}S_{1,4,-9} - \frac{104}{3}S_{1,4,-10} - \frac{8}{3}S_{1,4,-11} - \frac{14}{3}S_{1,4,-12} - \frac{29}{3}S_{1,4,-13} \right. \\
& \left. - \frac{77}{18}S_{1,4,-14} + \frac{14}{3}S_{1,4,-15} + [N_{-}, N_{+}] \right] \left[ \frac{3}{2}S_{1,\zeta_5} - \frac{29843}{16}S_{1,-1} + \frac{17}{3}S_{1,-2} + 14S_{1,-3} - \frac{29}{3}S_{1,-4} \right. \\
& - 18S_{1,-5} + 8S_{1,-6} + [N_{-}, N_{+}] \left[ \frac{864}{13}S_{1,-7} + \frac{2}{3}S_{1,-8} + 6S_{1,-9} - \frac{2}{3}S_{1,-10} \right. \\
& - 25S_{1,-11} - \frac{57}{2}S_{1,-12} - \frac{13}{3}S_{1,-13} + \frac{5}{2}S_{1,-14,1,1} + \frac{97}{12}S_{1,-14,2,1} - \frac{41}{2}S_{1,-14,2,2} - 7417S_{1,-14,3,1} \\
& + \frac{1}{2}S_{1,-14,3,2} - \frac{92}{3}S_{1,-14,3,3} + 15S_{1,-14,3,4} - \frac{9}{4}S_{1,-14,3,5} - 3S_{1,1,1,1} - 5S_{1,1,1,2} + \frac{1}{4}S_{1,1,1,3} + 38S_{1,-14,3,6} \\
& + 41S_{1,-14,3,7} + \frac{9}{4}S_{1,-14,3,8} - \frac{2}{3}S_{1,-14,3,9} - 2S_{1,-14,3,10} - \frac{1}{3}S_{1,-14,3,11} - \frac{2}{3}S_{1,-14,3,12} \\
& - \frac{1}{6}(2N_{-} - 4N_{+})S_{1,-14,3,13} + [N_{-}, N_{+}] \left[ \frac{1}{3}S_{1,-14,3,14} + 16C_F \eta^2 \left( \frac{1}{3}(N_{-} - N_{+}) \right) \left[ \frac{5}{3}S_{1,-14,3,15} \right] \right. \\
& - \frac{35}{9}S_{1,-14,3,16} + \frac{1}{3}S_{1,-14,3,17} + \frac{1057}{72}S_{1,-14,3,18} - \frac{16}{3}S_{1,-14,3,19} - \frac{8}{3}S_{1,-14,3,20} + \frac{3}{2}S_{1,-14,3,21} + \frac{181}{12}S_{1,-14,3,22} - \frac{1}{2}S_{1,-14,3,23} \\
& - \frac{1}{3}S_{1,-14,3,24} + 43S_{1,-14,3,25} + [N_{-}, N_{+}] \left[ \frac{25}{3}S_{1,-14,3,26} - \frac{1}{3}S_{1,-14,3,27} - \frac{95}{2}S_{1,-14,3,28} + \frac{1}{2}S_{1,-14,3,29} - \frac{1}{3}S_{1,-14,3,30} \right. \\
& - 1625S_{1,-14,3,31} + \frac{5}{2}S_{1,-14,3,32} - 2S_{1,-14,3,33} - \frac{1}{2}S_{1,-14,3,34} - 4S_{1,-14,3,35} - [N_{-}, N_{+}] \left[ \frac{7}{2}S_{1,-14,3,36} - \frac{11}{2}S_{1,-14,3,37} \right. \\
& - 144S_{1,-14,3,38} + \frac{5}{2}S_{1,1,1,1,1} - 2S_{1,1,1,1,2} - \frac{1}{2}S_{1,1,1,1,3} - \frac{83}{108}S_{1,1,1,1,4} - \frac{3}{2}S_{1,1,1,1,5} - \frac{1}{2}S_{1,1,1,1,6} \\
& - S_{1,-1,2,-5S_{1,-1}} + [N_{-}, N_{+}] \left[ \frac{15137}{864}S_{1,-1} + \frac{49}{6}S_{1,-2} - \frac{107}{18}S_{1,-3} + \frac{19}{12}S_{1,-4} - \frac{5}{6}S_{1,-5} - 10S_{1,-6} - 45S_{1,-7} \right. \\
& - \frac{1}{2}S_{1,2,1,1} + S_{2,2,1,2} - \frac{155}{24}S_{2,2,1,3} + S_{2,2,1,4} - S_{2,2,1,5} \Big) + 16C_F^{-3} \left( 2N_{-} - 4N_{+} - N_{-} + 3 \right) \left[ S_{1,1,1,1,2,-2} \right. \\
& - 47S_{1,-1} - S_{1,-2} - \frac{7}{2}S_{1,-3} + 6S_{1,-4} - \frac{1}{16}S_{1,-5} + 6S_{1,\zeta_5} + 4S_{1,1,-1} - 6S_{1,1,1,-2} - 3S_{1,1,2,-2} - 3S_{1,1,3,-1} \\
& - 23S_{1,1,1,-1} - \frac{9}{4}S_{1,1,1,1,1} + 2S_{1,1,1,1,2} + 4S_{1,1,1,1,3} + \frac{1}{2}S_{1,1,1,1,4} + 2S_{1,1,2,-2} + 2S_{1,1,3,-1} - 2S_{1,1,2,1} \\
& - \frac{3}{2}S_{1,1,2,2} + [N_{-}, N_{+}] \left[ \frac{287}{32}S_{1,1,2,3} - 24S_{1,1,2,4} - 24S_{1,1,2,5} + S_{1,1,1,1,1} \right. \\
& - 12S_{1,-1} - 36S_{1,-2} + \frac{111}{8}S_{1,-3} + 16S_{1,-4} - \frac{9}{2}S_{1,-5} + \frac{1}{2}S_{1,-6} - 2S_{1,-7} + 2S_{1,-8} - 45S_{1,-9} - 45S_{1,-10} \\
& + \frac{91}{16}S_{1,-11} - 2S_{1,-12} - \frac{41}{30}S_{1,-13} - \frac{3}{2}S_{1,-14} - S_{1,1,1,1,1} - 2S_{1,1,1,1,2} + 2S_{1,1,2,1} - \frac{35}{36}S_{1,1,2,2} + 34S_{1,1,2,3} \\
& - 2S_{1,1,2,4} + [N_{-}, N_{+}] \left[ \frac{749}{64}S_{1,-1} + 20S_{1,-2} - \frac{141}{12}S_{1,-3} - \frac{433}{16}S_{1,-4} + \frac{17}{6}S_{1,-5} + 6S_{1,-6} + 1S_{1,-7} \right. \\
& - 30S_{1,-8} - S_{1,1,2,1,1} - 2S_{1,1,2,1,2} + 2S_{1,1,2,1,3} + \frac{1}{2}S_{1,1,2,1,4} - 2S_{1,1,2,1,5} + 2S_{1,1,2,1,6} - 2S_{1,1,2,1,7} \\
& + \frac{19}{4}S_{1,1,2,1,8} + \frac{3}{4}S_{1,1,2,1,9} + 2S_{1,1,2,1,10} - \frac{11}{2}S_{1,1,2,1,11} - \frac{485}{16}S_{1,1,2,1,12} + \frac{27}{4}S_{1,1,2,1,13} - \frac{9}{4}S_{1,1,2,1,14} \quad (3.12)
\end{aligned}$$

aN<sup>3</sup>LO (2020 - ongoing)

# Anomalous Dimensions: Scale Dependence at LO



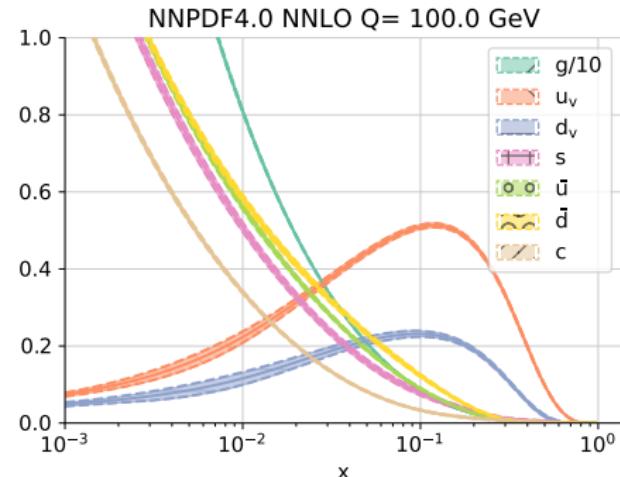
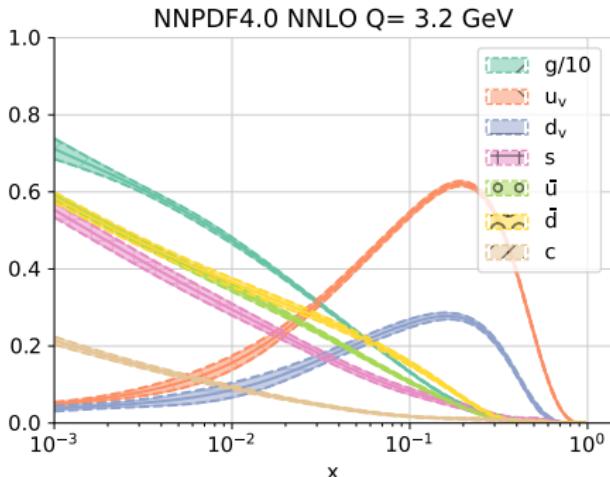
As  $Q^2$  increases, PDFs decrease at large  $x$  and increase at small  $x$  due to radiation

Gluon sector singular at  $N = 1$ , therefore the gluon grows faster 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  $\gamma_{gg}$  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  $f : [0, 1] \rightarrow \mathbb{R}$ .

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

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

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

## THE PROBLEM IS ILL-DEFINED

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

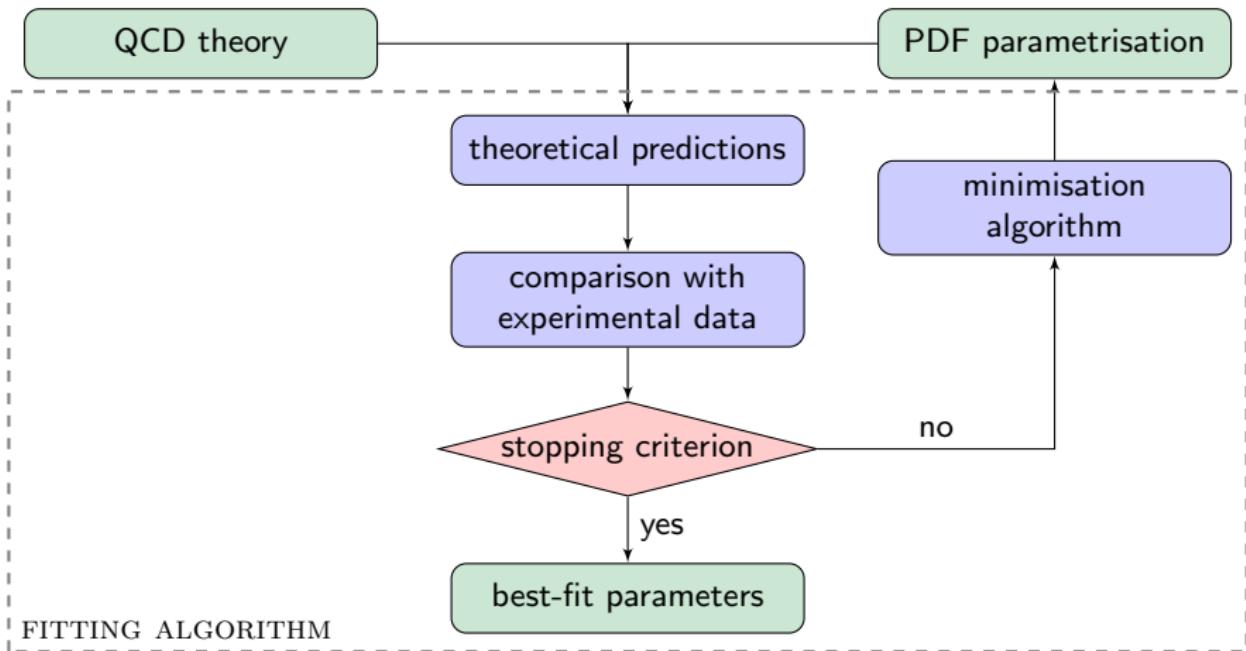
## Solution: parametric regression

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

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

# Determining PDFs from (LHC) experimental data

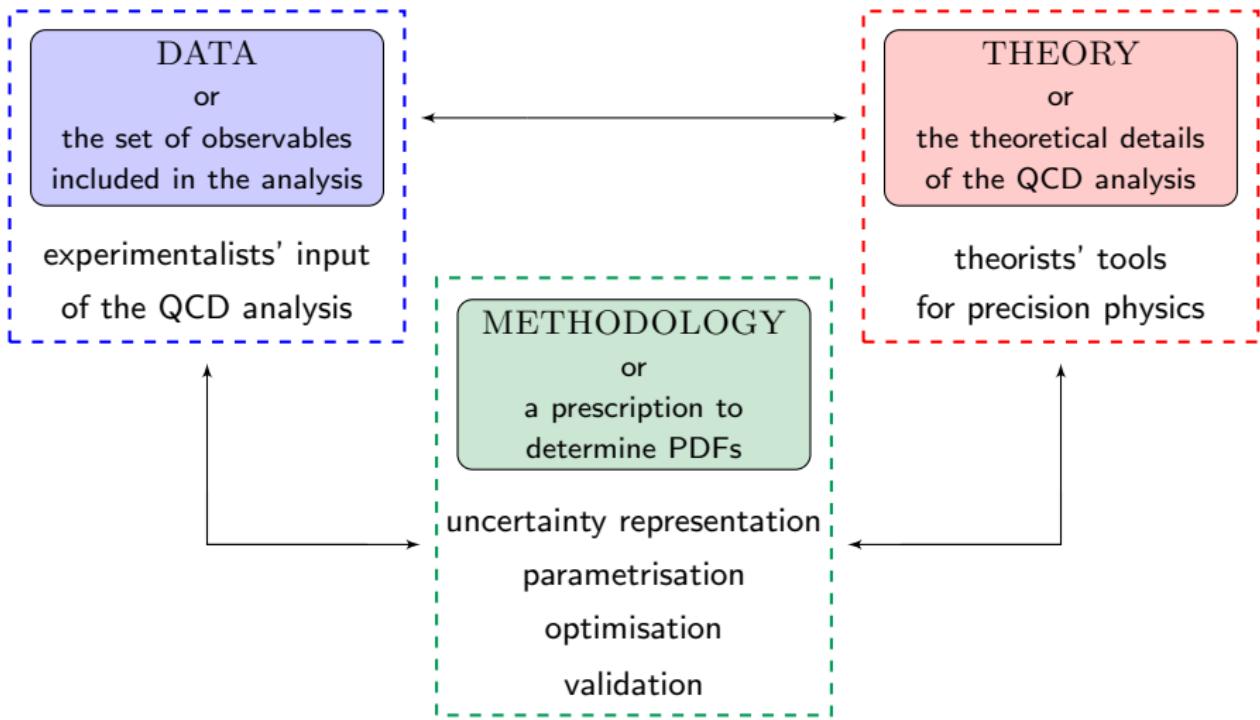
$$\sigma(\tau, Q^2 \mathbf{k}) = \sum_{ij} \int_\tau^1 \frac{dz}{z} \hat{\sigma}_{ij} \left( \frac{\tau}{z}, \alpha_s(Q^2), \mathbf{k} \right) \mathcal{L}_{ij}(z, Q^2) \quad \mathcal{L}_{ij}(z, Q^2) = (f_i^{h_1} \otimes f_j^{h_2})(z, Q^2)$$



FITTING ALGORITHM

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

# 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

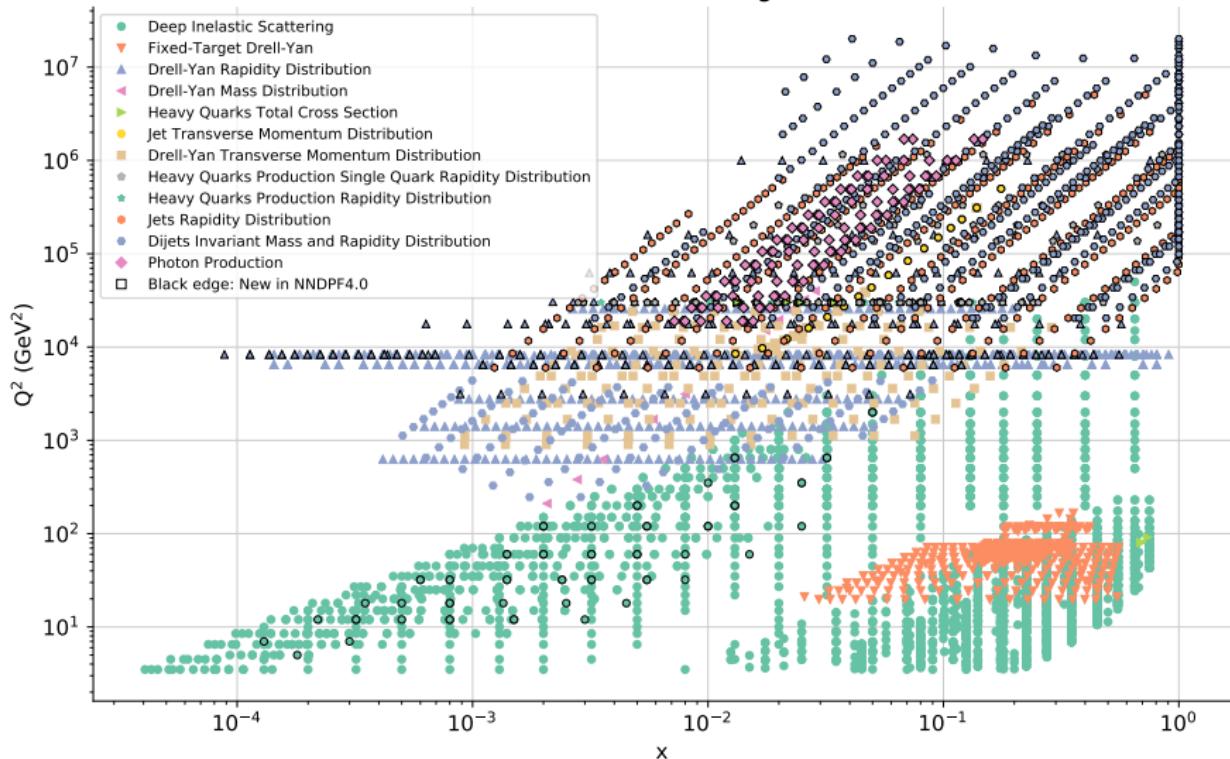
	NNPDF4.0	MSHT20	CT18	HERAPDF2.0	CJ22	ABMP16
Fixed-target DIS	✓	✓	✓	✗	✓	✓
JLAB	✗	✗	✗	✗	✓	✗
HERA I+II	✓	✓	✓	✓	✓	✓
HERA jets	✓	✗	✗	✓	✗	✗
Fixed target DY	✓	✓	✓	✗	✓	✓
Tevatron $W, Z$	✓	✓	✓	✗	✓	✓
LHC vector boson	✓	✓	✓	✗	✓	✓
LHC $W + c$ $Z + c$	✓	✗	✗	✗	✗	✗
Tevatron jets	✓	✓	✓	✗	✓	✗
LHC jets	✓	✓	✓	✗	✗	✗
LHC top	✓	✓	✗	✗	✗	✓
LHC single $t$	✓	✗	✗	✗	✗	✗
LHC prompt $\gamma$	✓	✗	✗	✗	✗	✗
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	Chebyshev pol.	Bernstein pol.	polynomial	polynomial	polynomial
HQ scheme	FONLL	TR'	ACOT- $\chi$	TR'	ACOT- $\chi$	FFN
accuracy	aN <sup>3</sup> LO	aN <sup>3</sup> LO	NNLO	NNLO	NLO	NNLO
latest update	EPJC82 (2022) 428	EPJC81 (2021) 341	PRD 103 (2021) 014013	EPJC82 (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

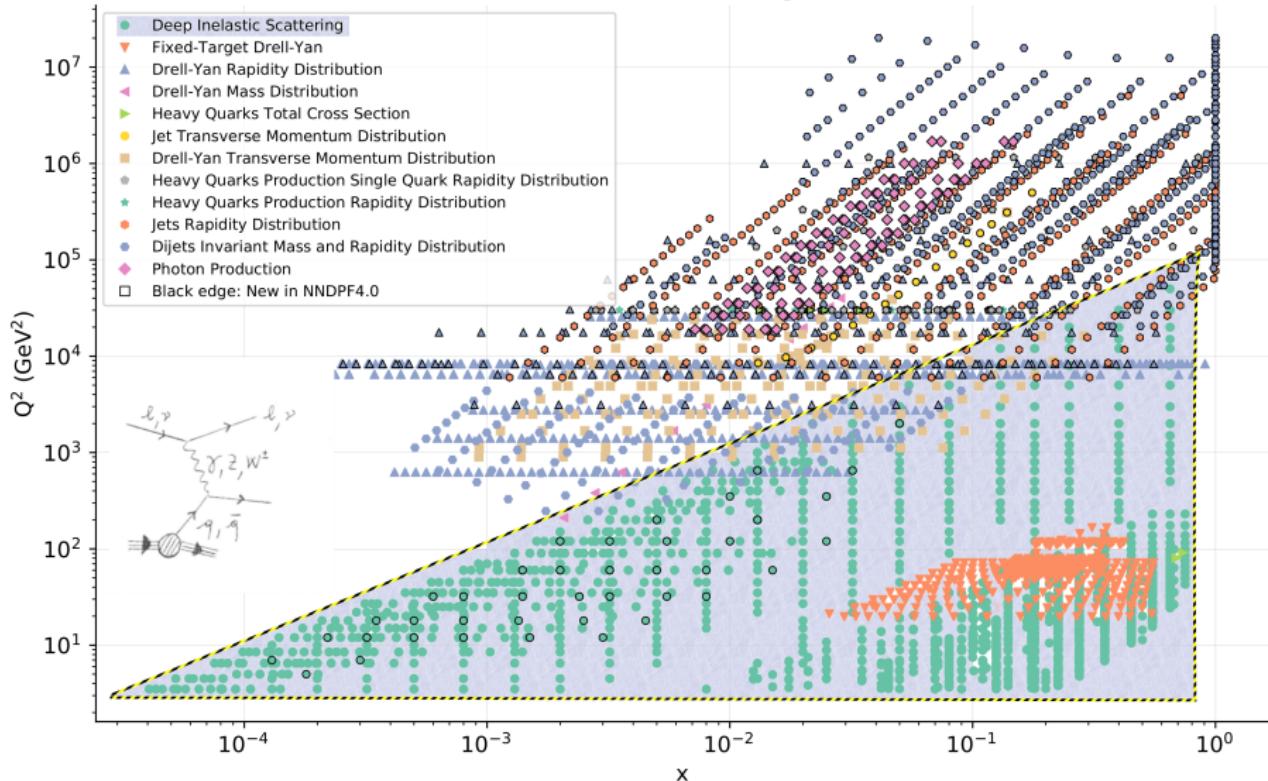
Kinematic coverage



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

# Deep Inelastic Scattering

## Kinematic coverage



# Deep Inelastic Scattering

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

$$\frac{d^2\sigma}{dxdy} \propto Y_+ F_2^i \mp Y_- x F_3^i - y^2 F_L^i \quad Y_{\pm} = 1 \pm (1-y)^2 \quad F_L^i = F_2^i - 2x F_1^i \quad i = \text{NC}(\gamma, Z, \gamma Z), \text{CC}(W^{\pm})$$

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

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

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

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

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

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

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

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

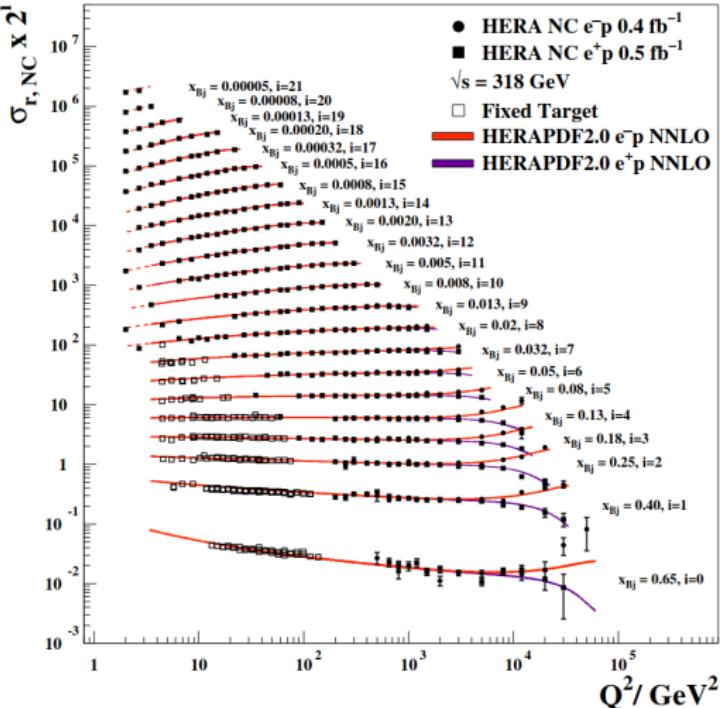
isospin ( $p \rightarrow n$ ):  $u^p = d^n \quad d^p = u^n \quad \bar{u}^p = \bar{d}^n \quad \bar{d}^p = \bar{u}^n$

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

$$Q^2 \geq Q_{\min}^2 \sim 1 \text{ GeV}^2 \quad W^2 = Q^2(1-x)/x \geq W_{\min}^2 \sim 10 \text{ GeV}^2$$

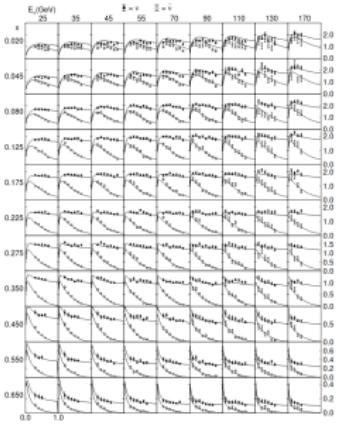
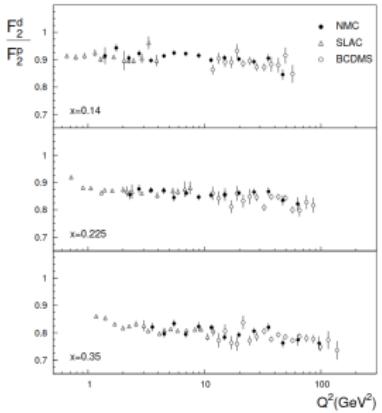
# Deep Inelastic Scattering

## H1 and ZEUS



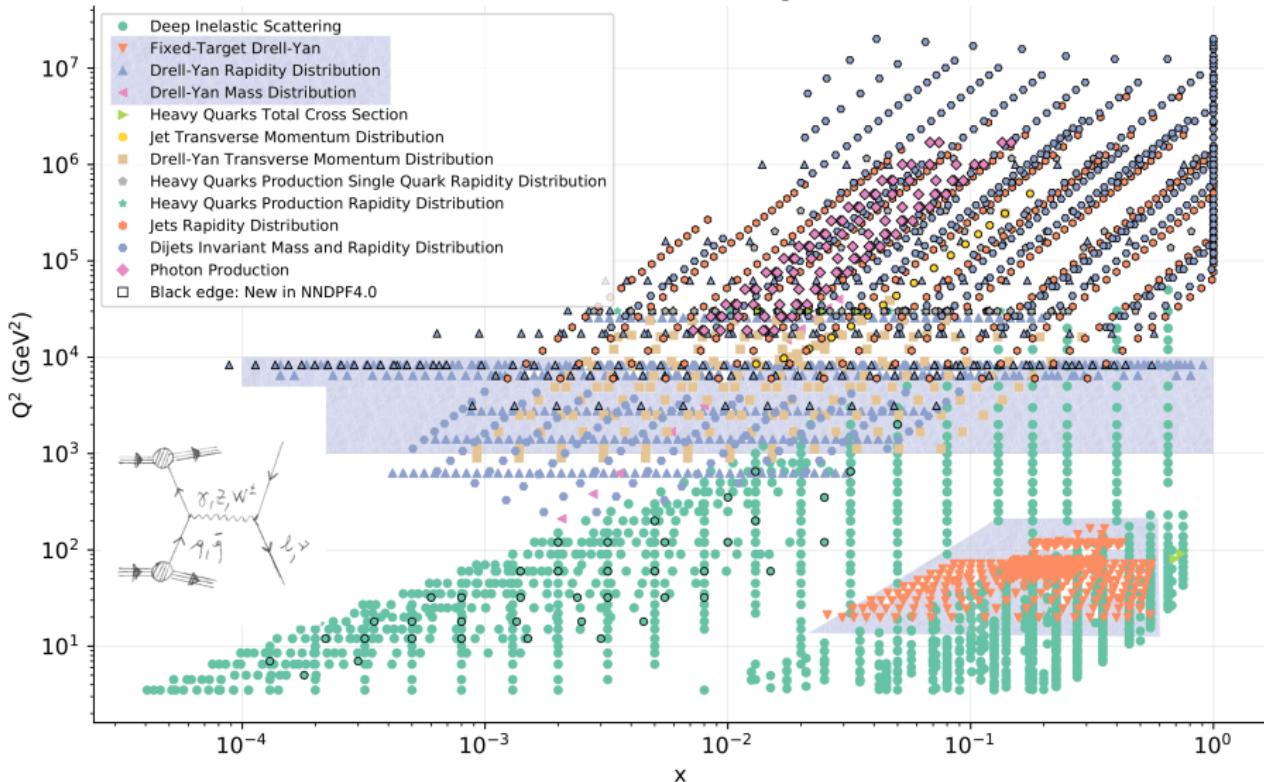
[HERA, EPJ C75 (2015) 580]

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



# Drell-Yan

## Kinematic coverage



# Drell-Yan

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

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

NC DY ( $\gamma, Z$ ) at LO

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

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

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

$$\text{isospin } (p \rightarrow n): u^p = d^n \quad d^p = u^n \quad \bar{u}^p = \bar{d}^n \quad \bar{d}^p = \bar{u}^n$$

deuteron target approximated as the average of one proton and one neutron  
different experiments measure different cross section combinations

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

DY  $p/d$  asymmetry (NuSea, SeaQuest)

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

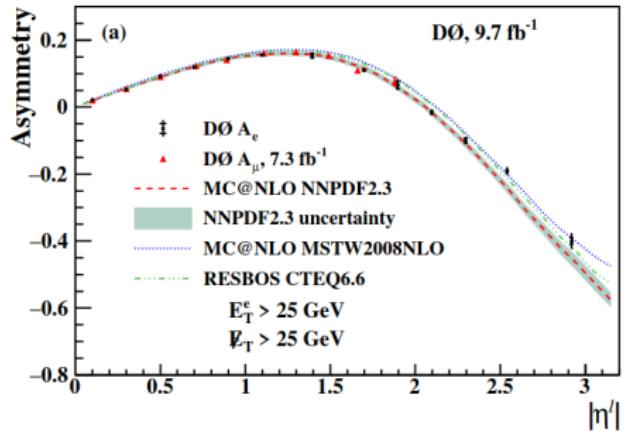
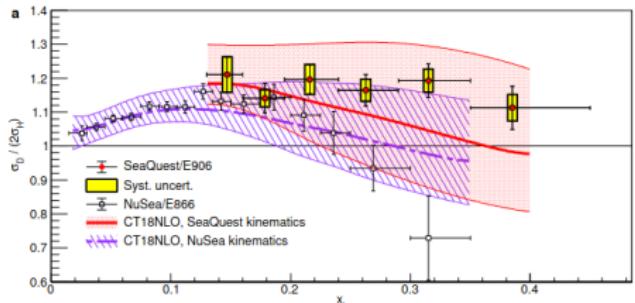
$W^\pm$  asymmetry (CDF, D0)

$$\sigma_{pp}^{W^+} \approx u\bar{d} + c\bar{s} \quad \sigma_{pp}^Z \approx u\bar{u} + d\bar{d} + s\bar{s} \rightarrow s, \bar{s} \quad W^\pm \text{ and } Z \text{ production (ATLAS, CMS, LHCb)}$$

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

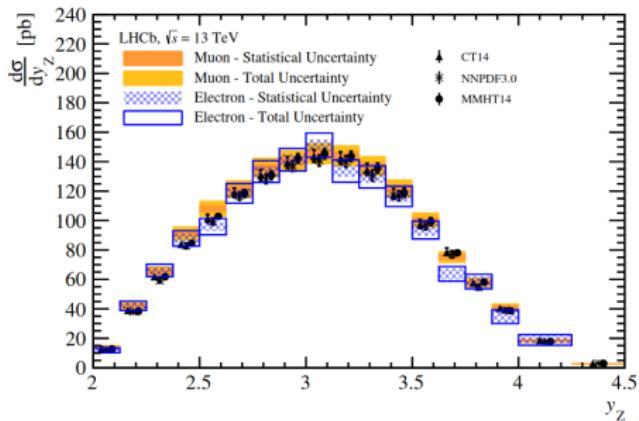
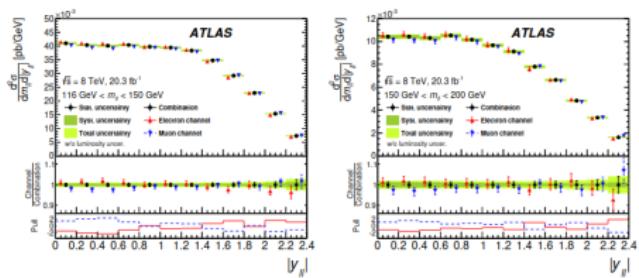
$W^\pm$  muon asymmetry (ATLAS, CMS)

# Drell-Yan



[SeaQuest, Nature 590 (2021) 7847]

[DØ, PRD 91 (2015) 032007]

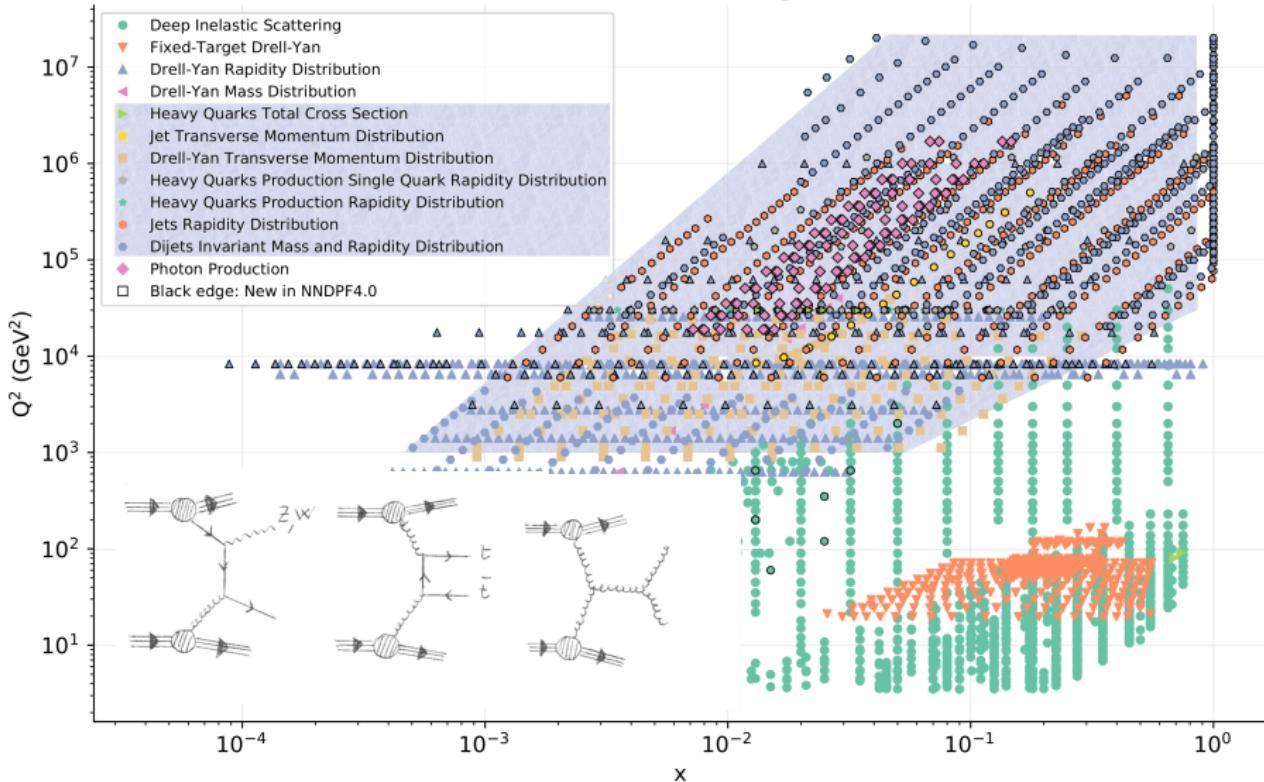


[ATLAS, JHEP 08 (2016) 009]

[LHCb, JHEP 09 (2016) 136]

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

## Kinematic coverage



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

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

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

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

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

process initiated by two gluons in the initial state

differential cross sections reconstructed at parton level (additional systematics)  
normalise by  $\sigma_{\text{tot}}^{t\bar{t}}$  (systematics largely cancel, but loose control on PDF shape)

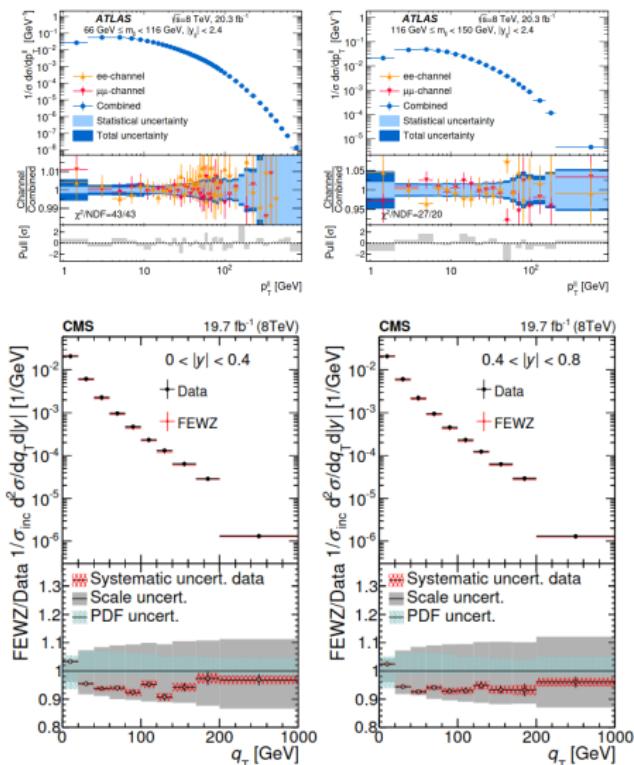
particle-level cross sections are theoretically more complicated  
wide rapidity range, constraints on the large- $x$  region

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

process initiated by two gluons in the initial state

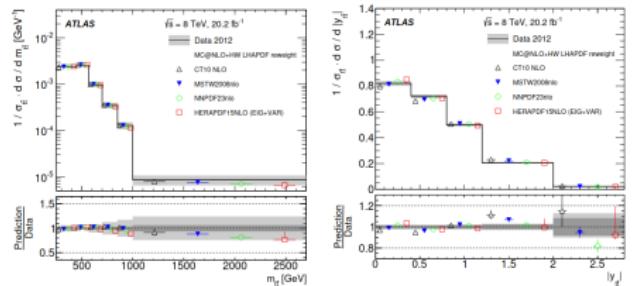
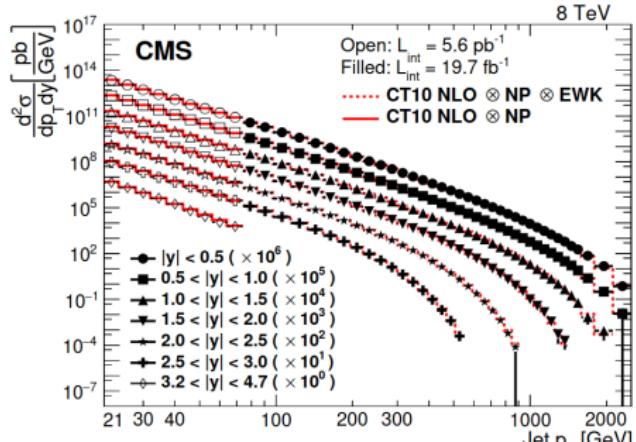
be careful with systematic uncertainties (mostly driven by jet energy reconstruction)  
can also be measured in DIS  
wide rapidity range, constraints on the large- $x$  region

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



[ATLAS, EPJ C76 (2016) 291]

[CMS PhysLB 749 (2015) 187]



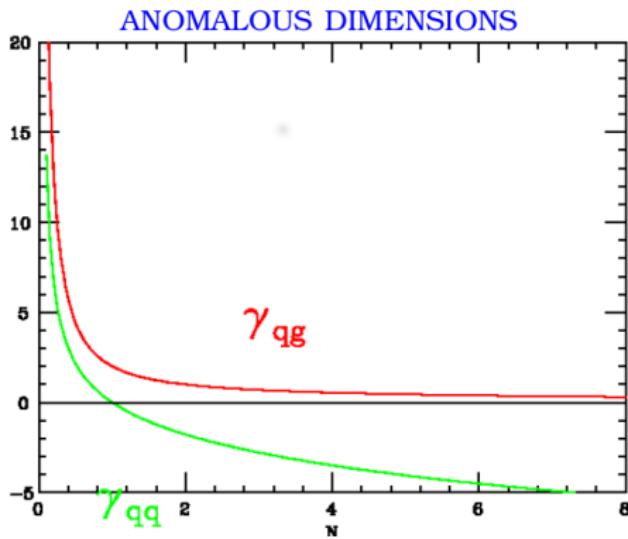
[CMS, JHEP 03 (2017) 156]

[ATLAS, EPJ C76 (2016) 538]

# Scaling Violations and Heavy Flavour Production

Scale dependence in DIS  
of singlet structure function

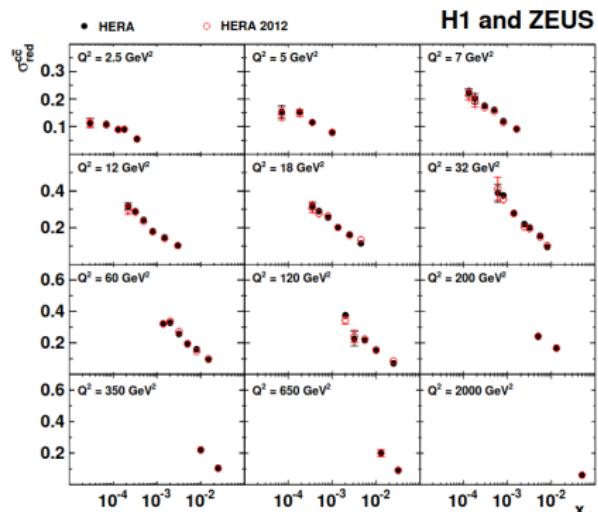
$$\frac{d}{d \ln(Q^2)} F_2^\Sigma \approx \frac{\alpha_s}{2\pi} [\gamma_{qq} f_\Sigma + 2n_f \gamma_{qg} f_g]$$



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

Heavy quark production in DIS  
initiated by gluons

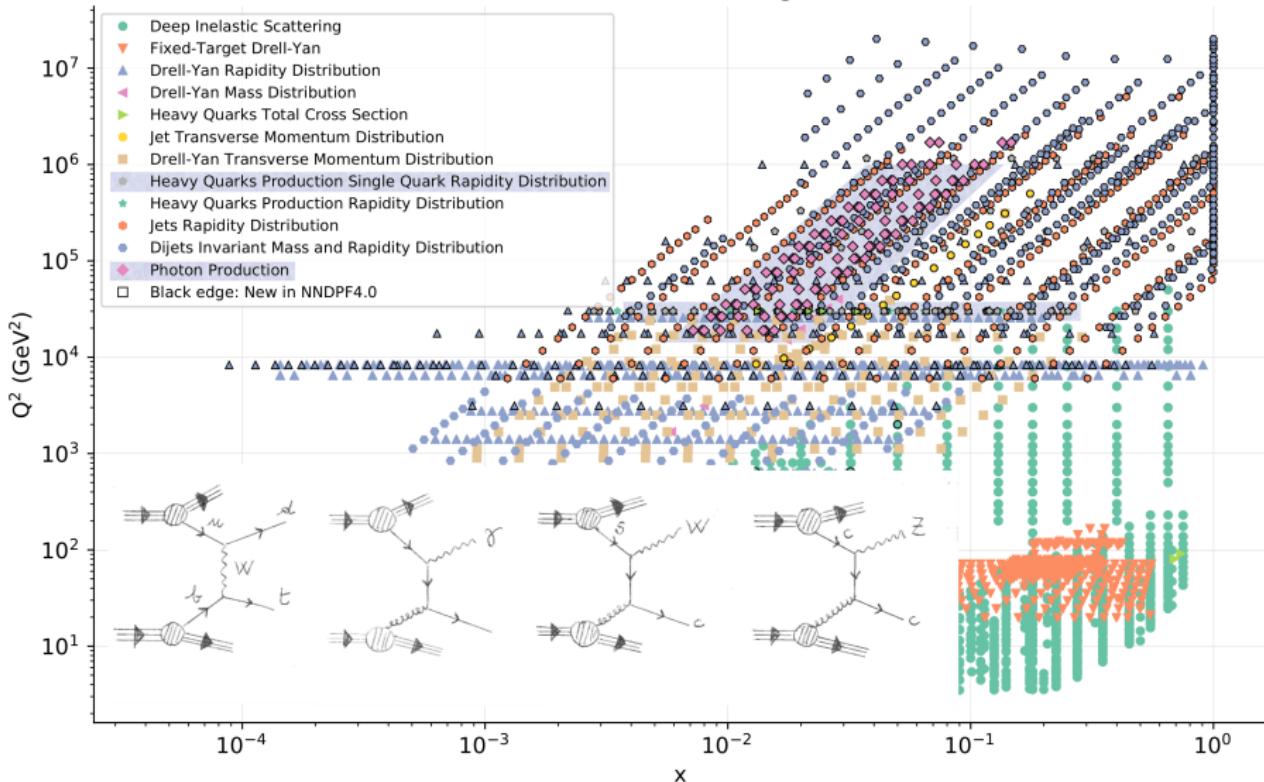
$$\sigma_{\text{red}}^{c,b} = F_2^{c,b} - \frac{y^2}{1 + (1-y)^2} F_L^{c,b}$$



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

# Single $t$ , Direct $\gamma$ , $W, Z+c$ -jets

## Kinematic coverage



# Single $t$ , Direct $\gamma$ , $W, Z+c$ -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  $\gamma$  production:  $\frac{d\sigma^2}{dE_T^\gamma dy^\gamma}$  (ATLAS and CMS)

gluon-quark-initiated Compton scattering  
potential sensitivity to the gluon PDF

potential currently limited by large experimental uncertainties

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

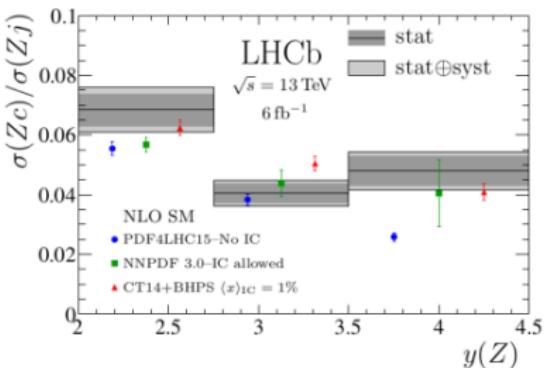
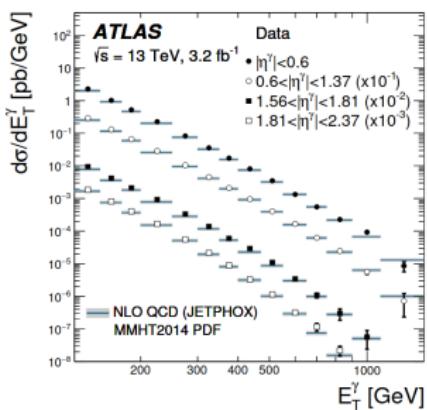
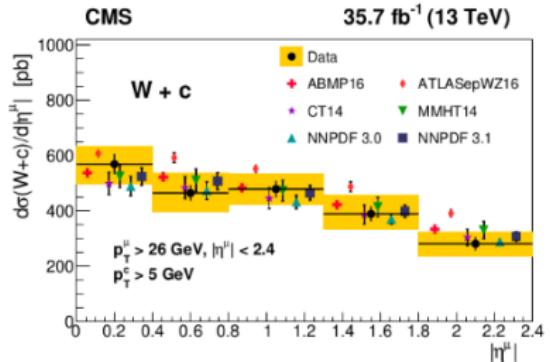
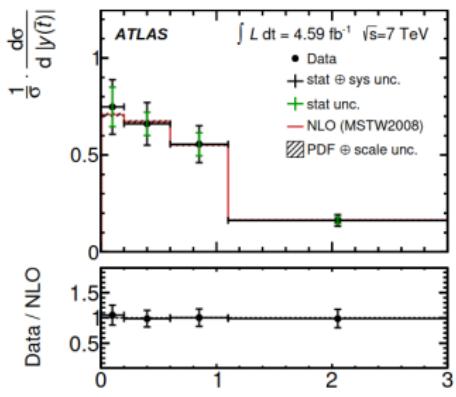
$W + c$ : sensitivity to strange PDF (and  $s - \bar{s}$  asymmetry)

$W + Z$ : 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 $\gamma$ , $W, Z+c$ -jets



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

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

# All the data together

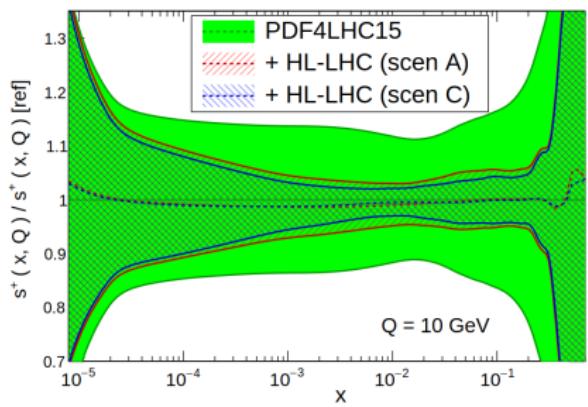
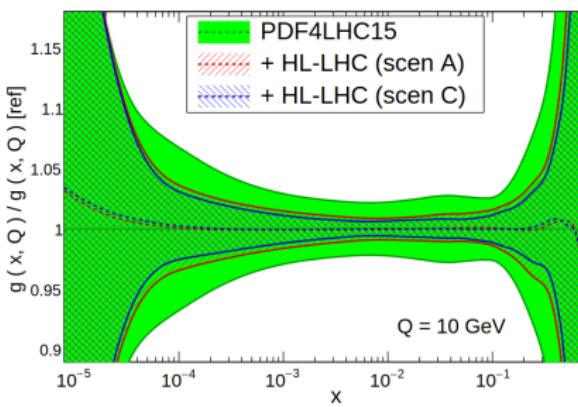
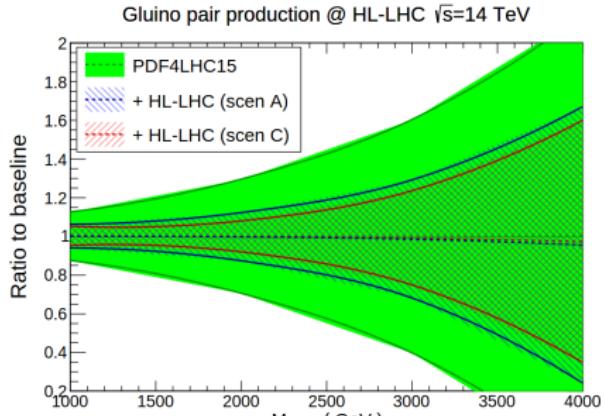
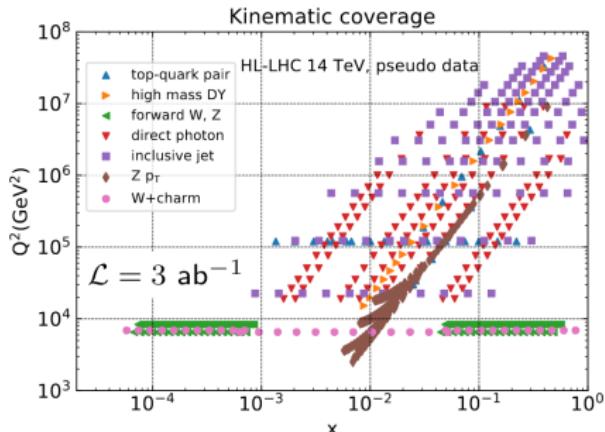
	Hadronic Process	Partonic Process	PDFs probed	$x$ coverage
Lepton-nucleon	$\ell^\pm \{p, n\} \rightarrow \ell^\pm + X$	$\gamma^* q \rightarrow q$	$q, \bar{q}, g$	$x \gtrsim 0.01$
	$\ell^\pm n/p \rightarrow \ell^\pm + X$	$\gamma^* d/u \rightarrow d/u$	$d/u$	$x \gtrsim 0.01$
	$\nu(\bar{\nu})N \rightarrow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

# 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 $\gamma$	✓	✗	✗	✗	✗	✗
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	Chebyshev pol.	Bernstein pol.	polynomial	polynomial	polynomial
HQ scheme	FONLL	TR'	ACOT- $\chi$	TR'	ACOT- $\chi$	FFN
accuracy	aN <sup>3</sup> LO	aN <sup>3</sup> LO	NNLO	NNLO	NLO	NNLO
latest update	EPJC82 (2022) 428	EPJC81 (2021) 341	PRD 103 (2021) 014013	EPJC82 (2022) 243	PRD 107 (2023) 113005	PRD 96 (2017) 014011

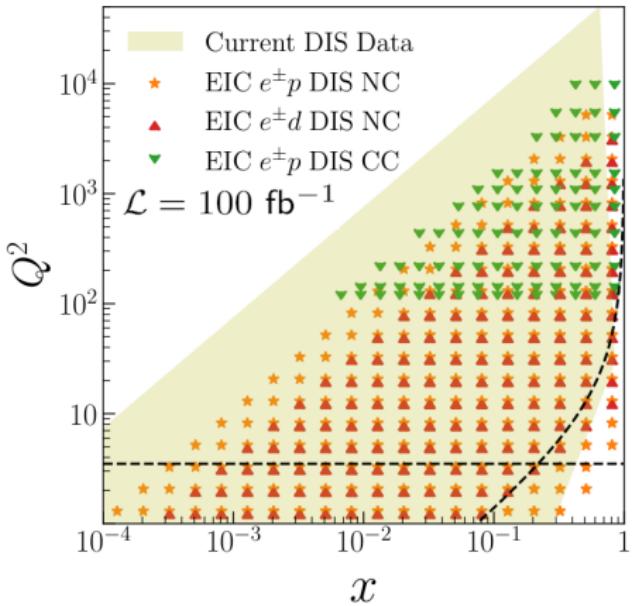
All PDF sets are available as  $(x, Q^2)$  interpolation grids through the LHAPDF library

# Impact of future data: HL-LHC



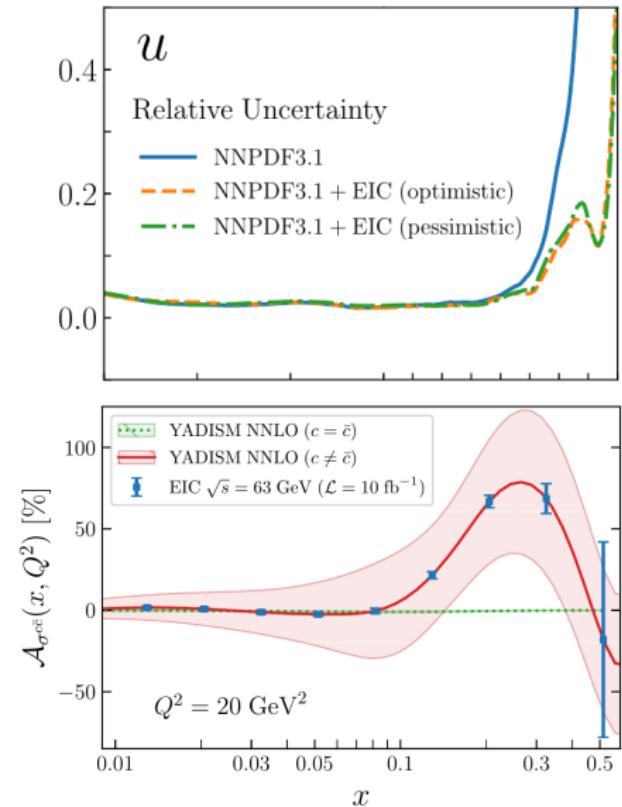
[EPJ C78 (2018) 962]

# Impact of future data: EIC



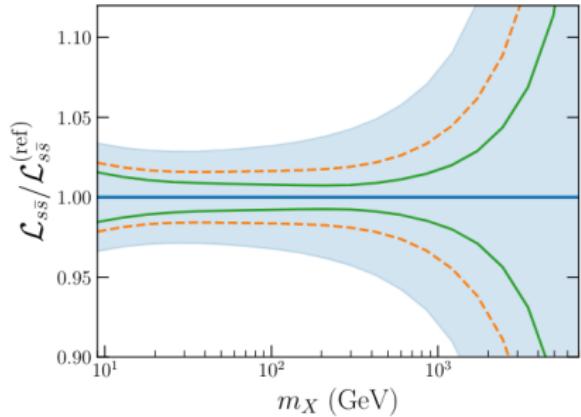
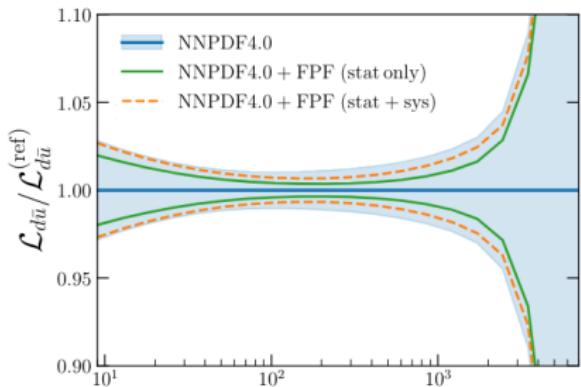
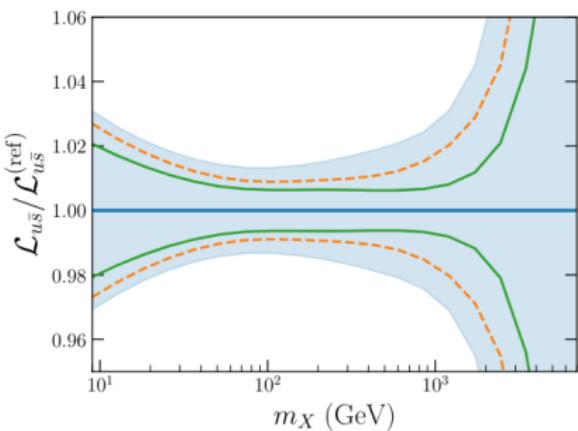
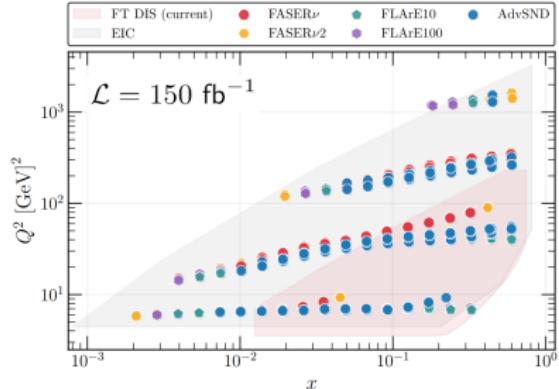
$E_\ell \times E_p$  [GeV]: 18 × 275; 10 × 100; 5 × 100

$$A_{\sigma^{c\bar{c}}} = \frac{\sigma_{\text{red}}^c - \sigma_{\text{red}}^{\bar{c}}}{\sigma_{\text{red}}^{c\bar{c}}}$$



[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

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

Lecture 2: Theoretical and methodological accuracy in PDF determination