Understanding (PDF) Uncertainties

A personal overview of parton distribution function determinations

Alberto Guffanti
NBIA & Discovery Center, Niels Bohr Institute - University of Copenhagen

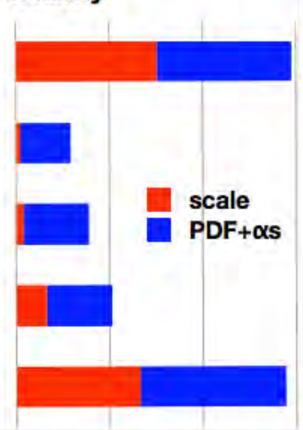
GRK Seminar
Freiburg (DE), 29.04.2015

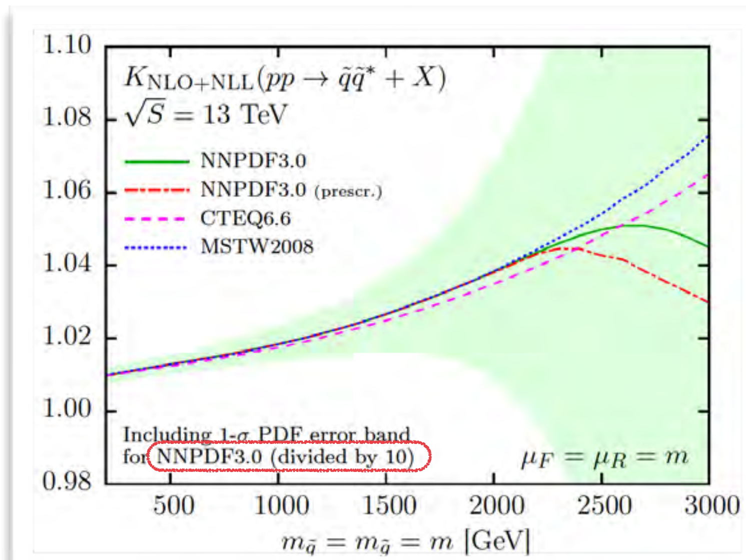
Motivation

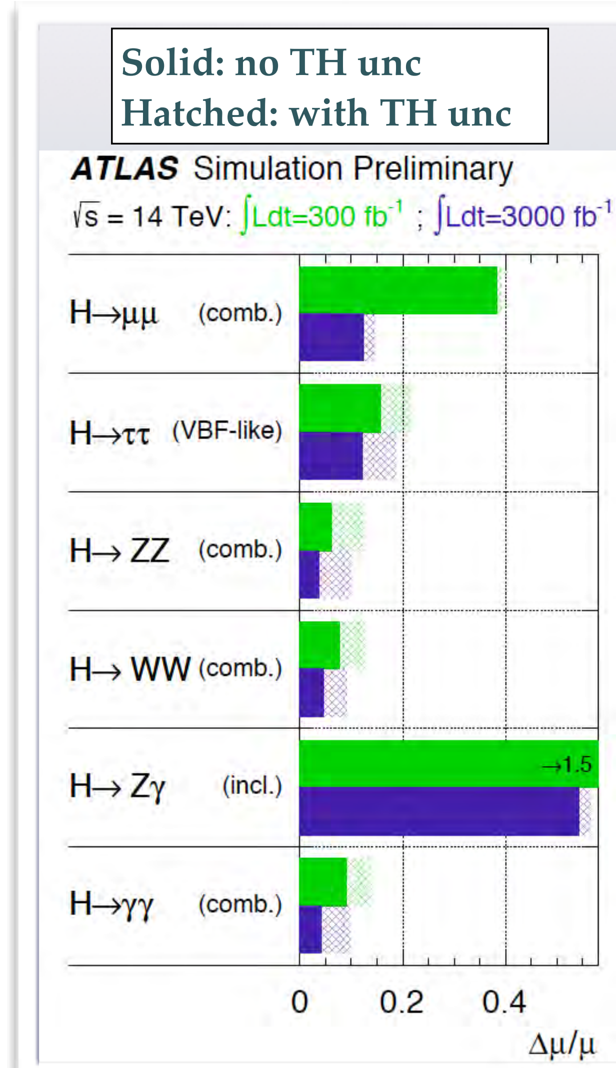
Experimental push

- Uncertainties on Parton Distribution Functions (PDFs) are often the limiting factor in precision Standard Model studies and New Physics searches

	σ (8 TeV)	uncertainty
gg→H	19.5 pb	14.7%
VBF	1.56 pb	2.9%
WH	0.70 pb	3.9%
ZH	0.39 pb	5.1%
ttH	0.13 pb	14.4%







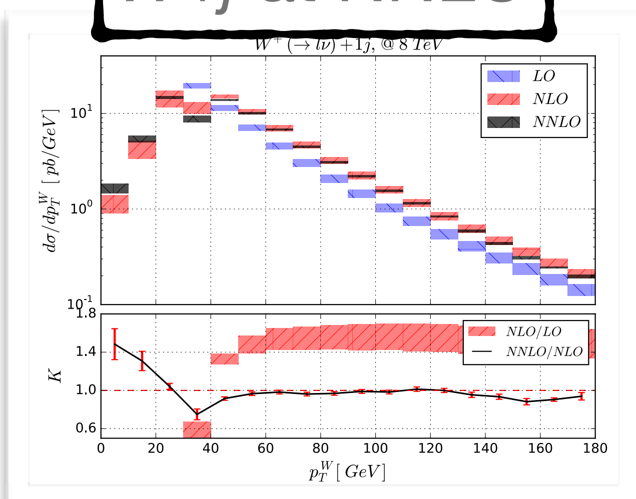
Accurate and reliable PDFs are crucial for exploiting the full potential of LHC experiments

Motivation

Theoretical push

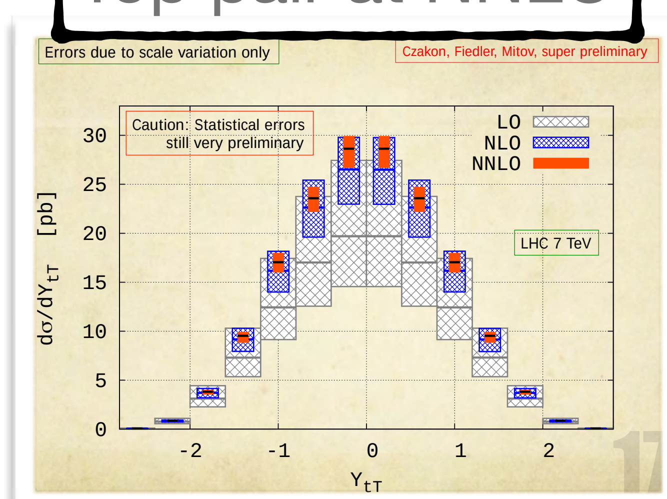
- Impressive progress in computation of higher order corrections (QCD & EW)

W+j at NNLO



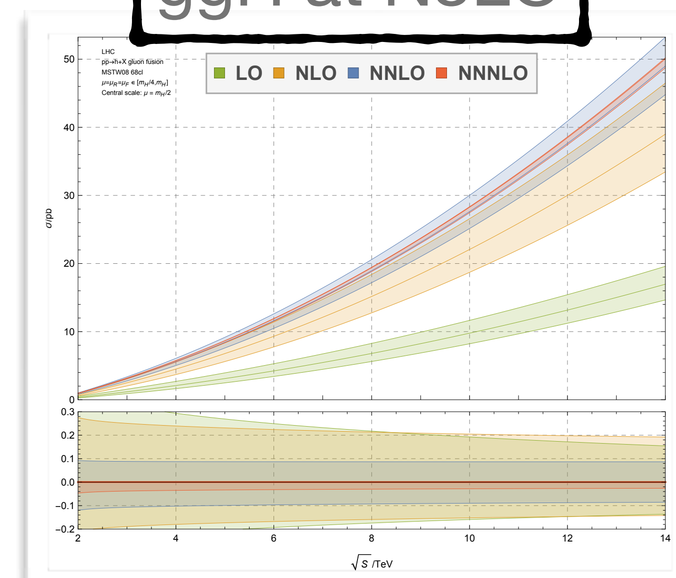
R. Boughezal et al, [arXiv:1504.02131]

Top pair at NNLO



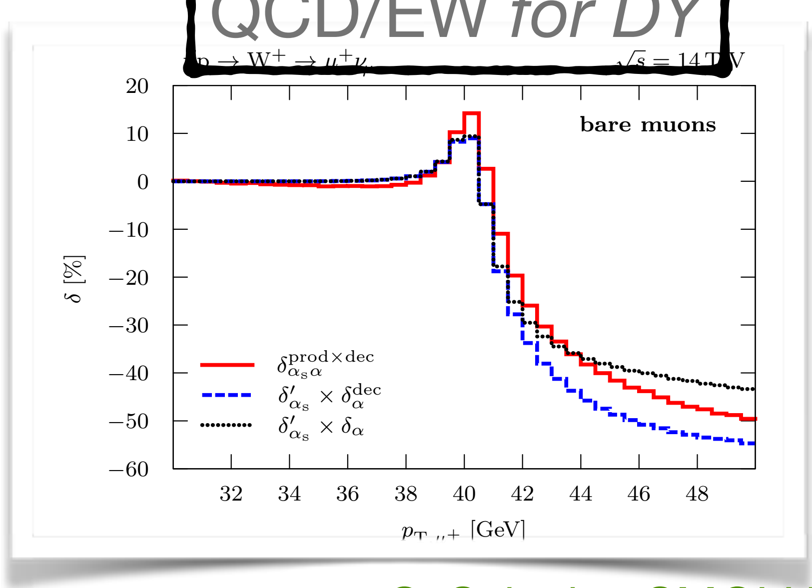
A. Mitov et al, Top2014

ggH at N3LO



C. Anastasiou et al, [arXiv:1503.06056]

QCD/EW for DY



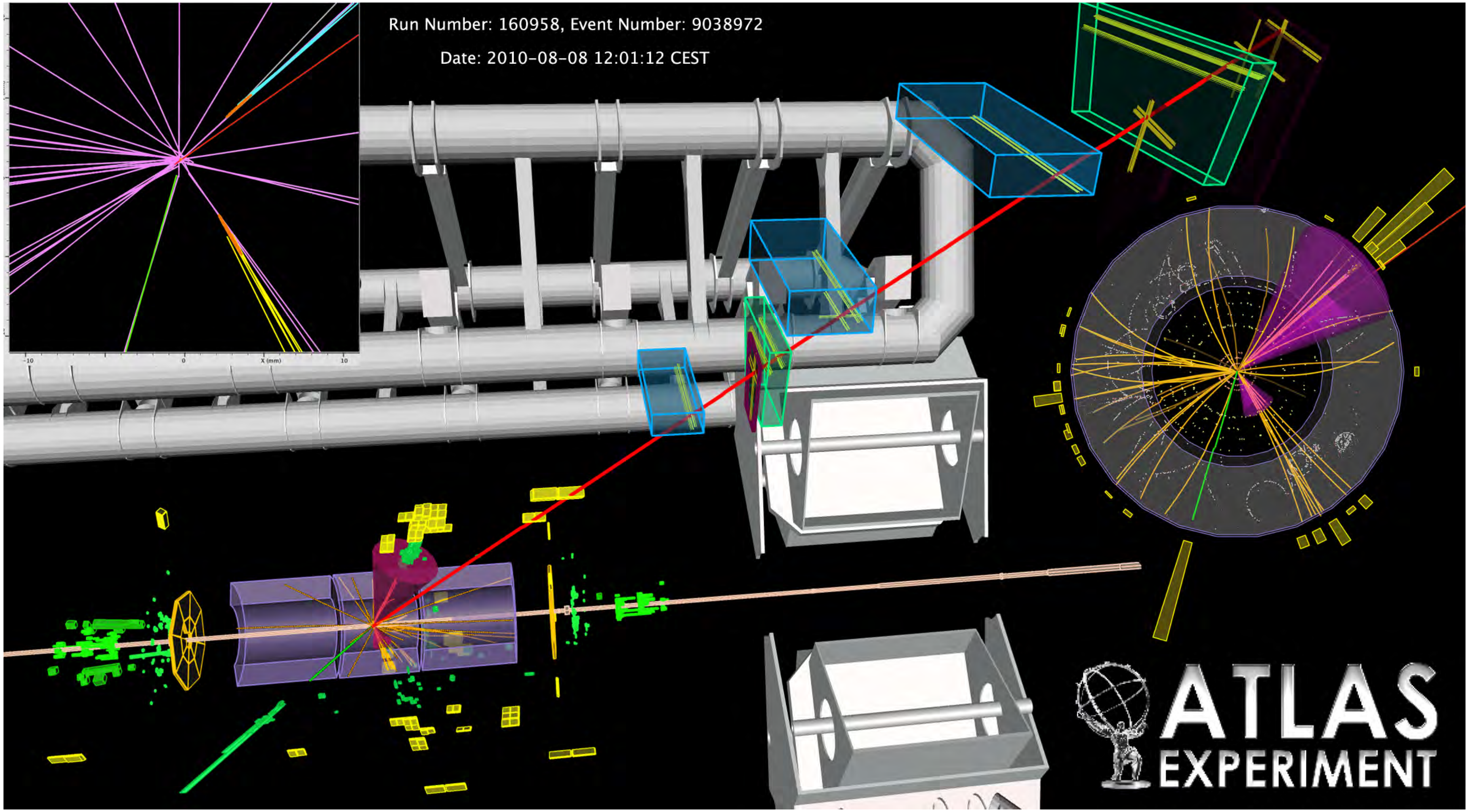
C. Schwinn,SM@LHC2015

Finally, the computation of the hadronic cross-section relies crucially on the knowledge of the strong coupling constant and the parton densities. After our calculation, the uncertainty coming from these quantities has become dominant. Further progress in the determination of parton densities must be anticipated in the next few years due to the inclusion of LHC data in the global fits and the impressive advances in NNLO computations, improving the theoretical accuracy of many standard candle processes.

Basics of PDF determinations

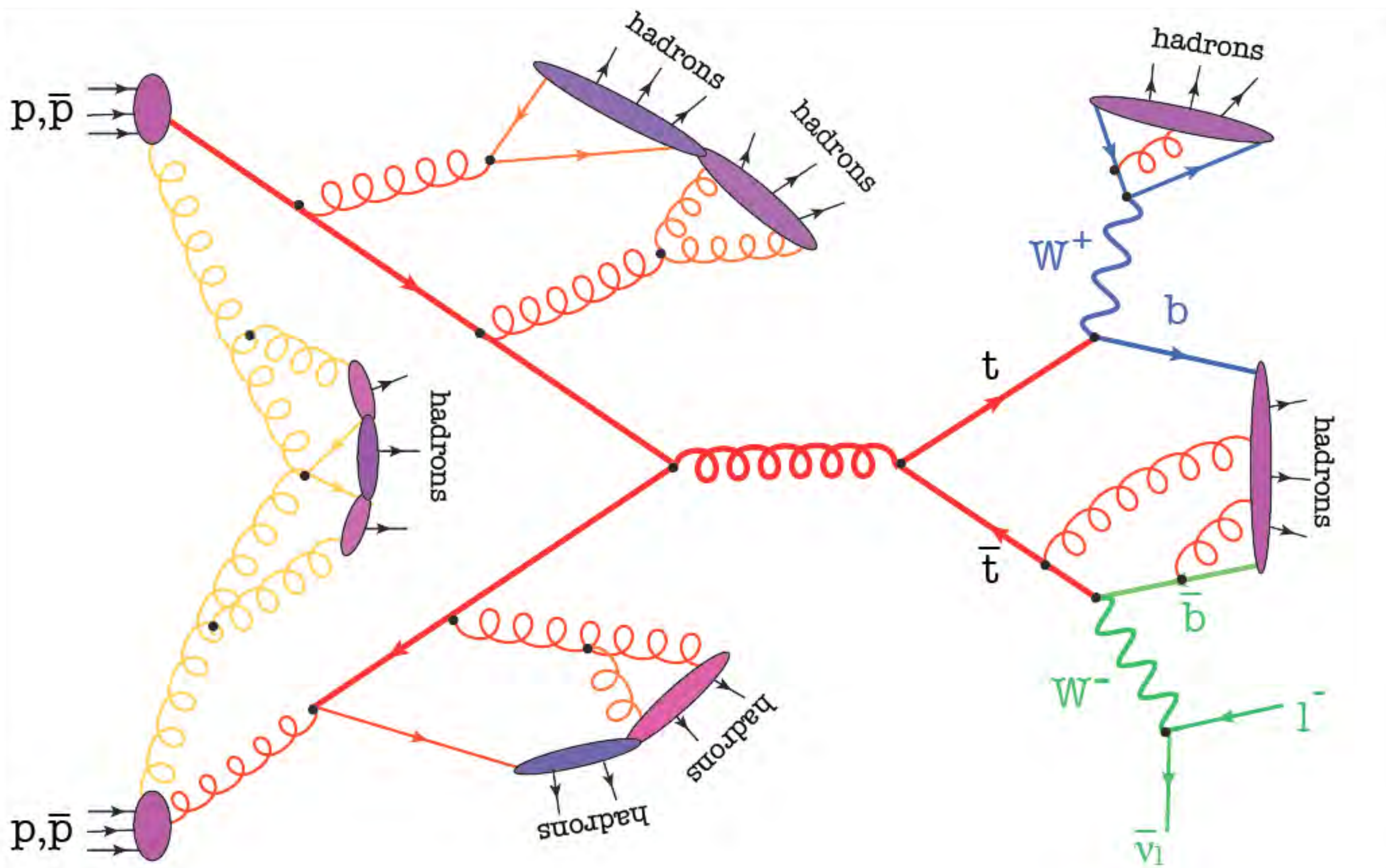
LHC events

An experimentalist view



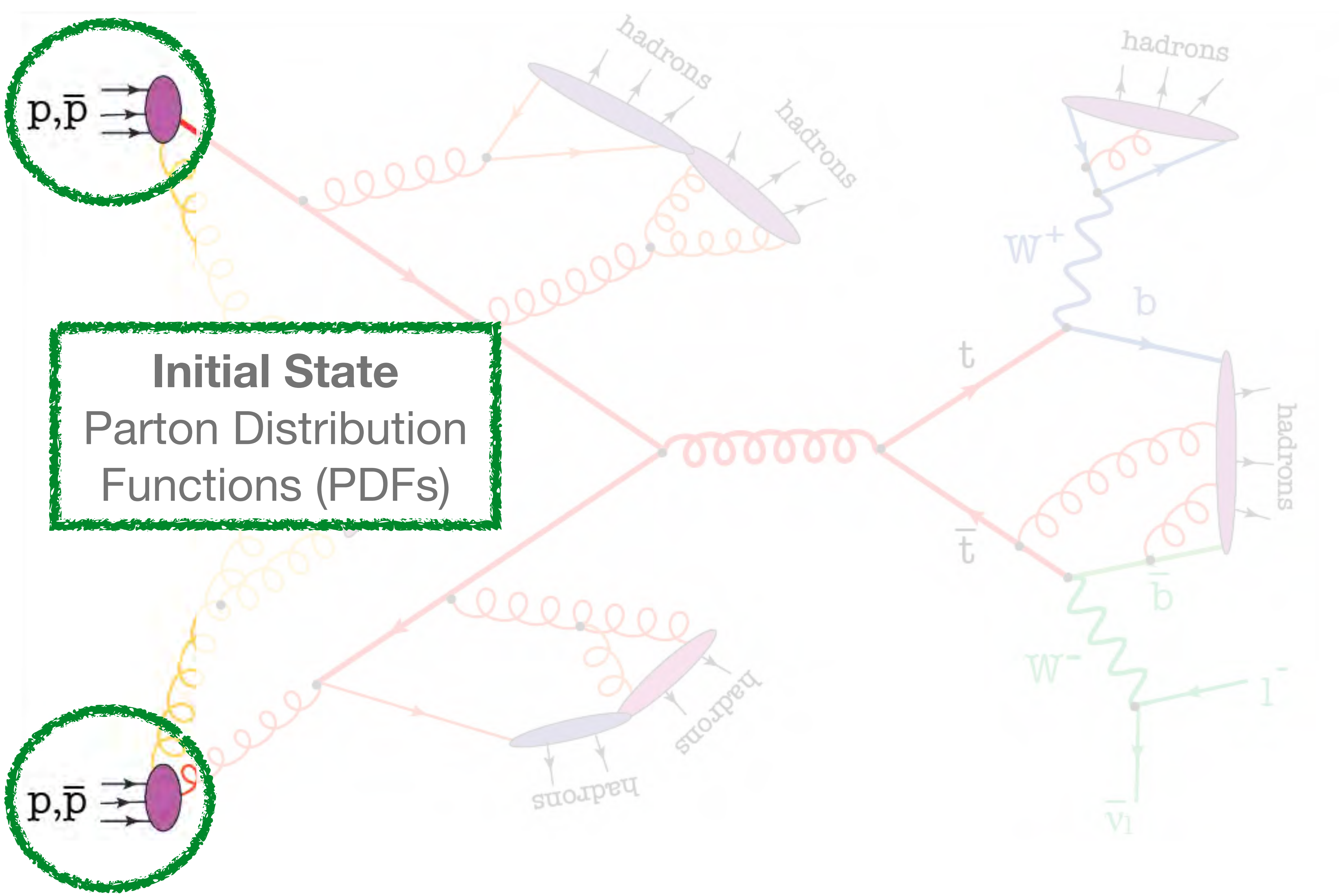
LHC events

A theorist view



LHC events

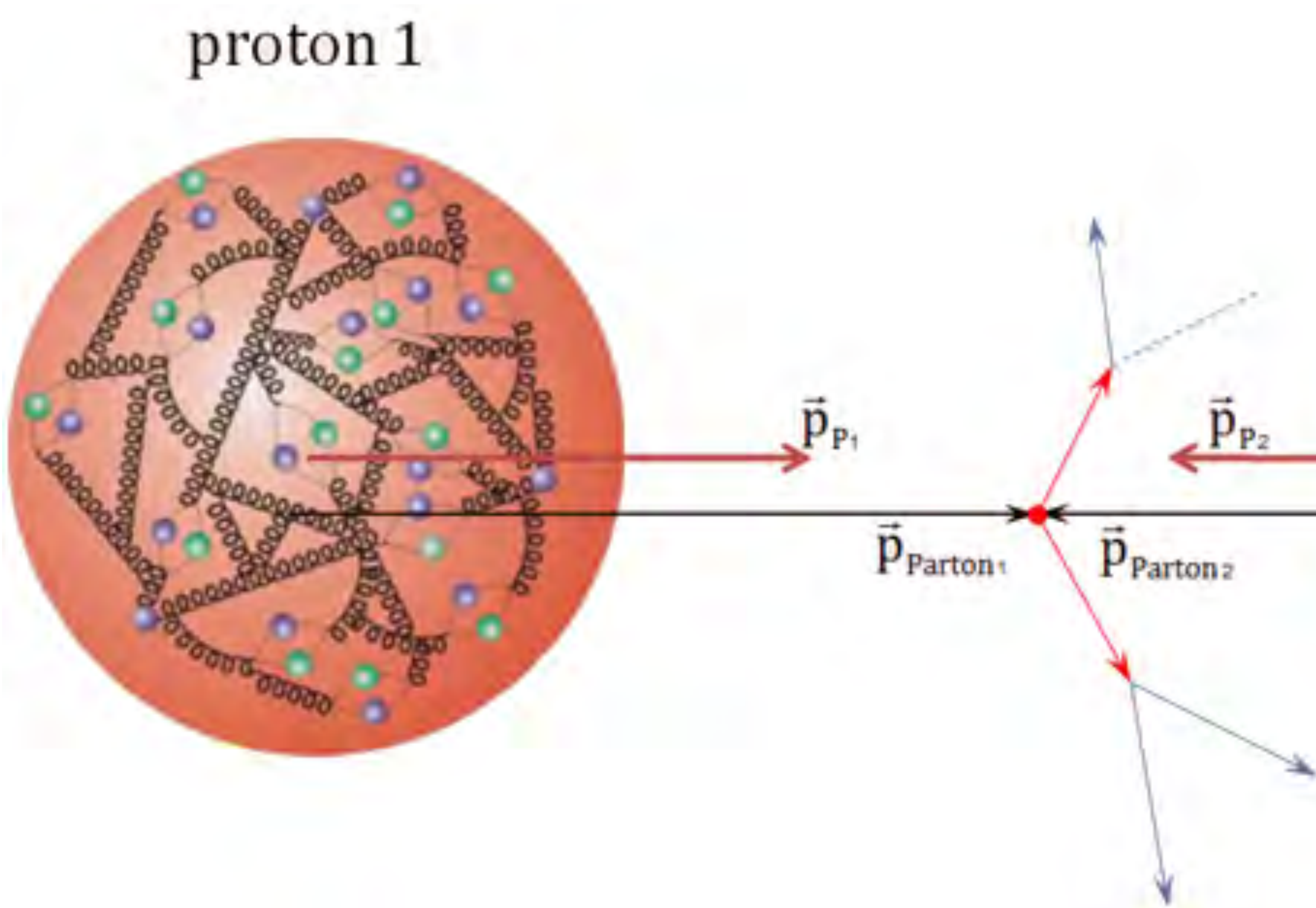
A theorist view

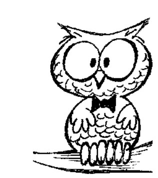


Initial State
Parton Distribution
Functions (PDFs)

Parton Distribution Functions at the LHC

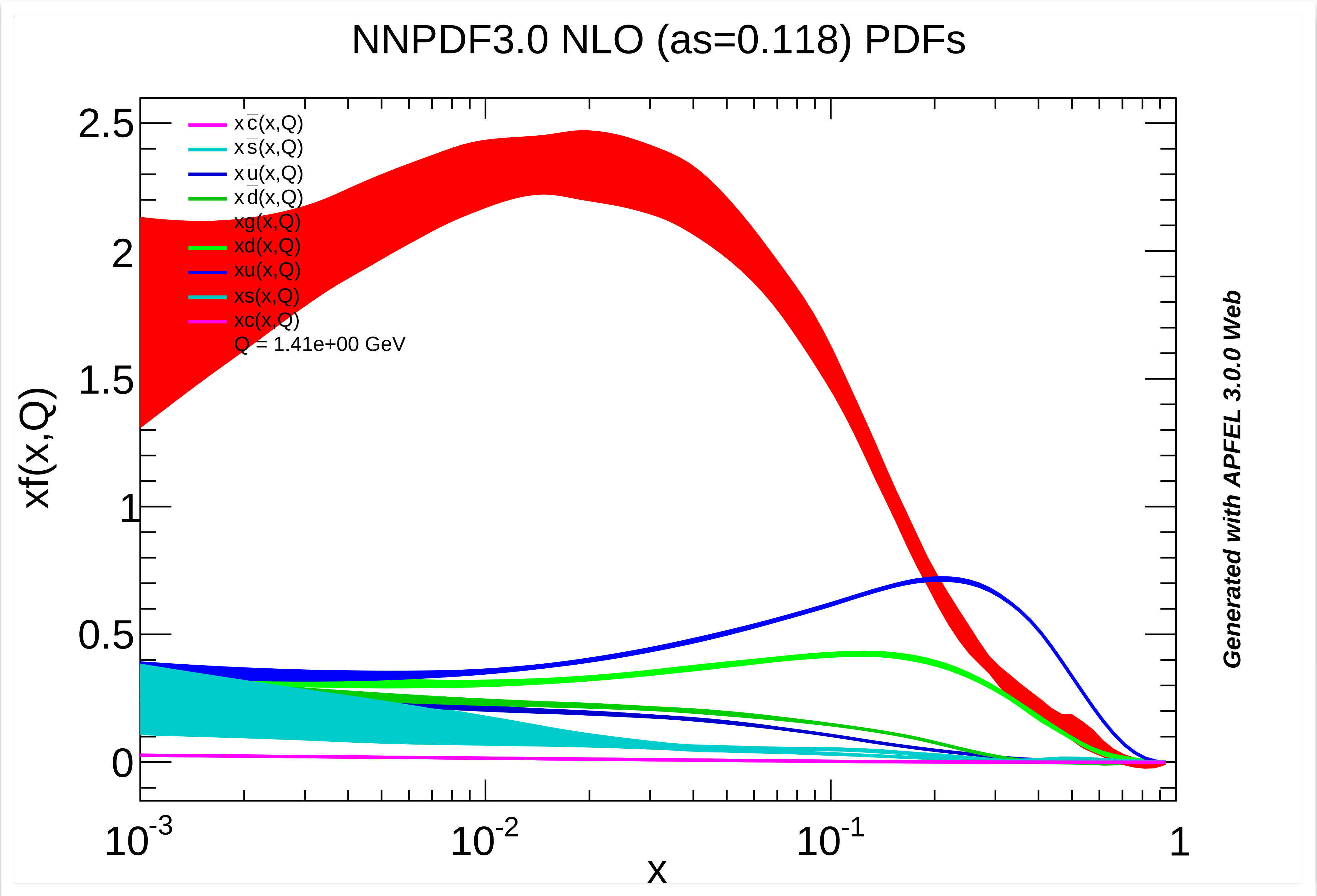
A phenomenologist view





Parton Distribution Functions at the LHC

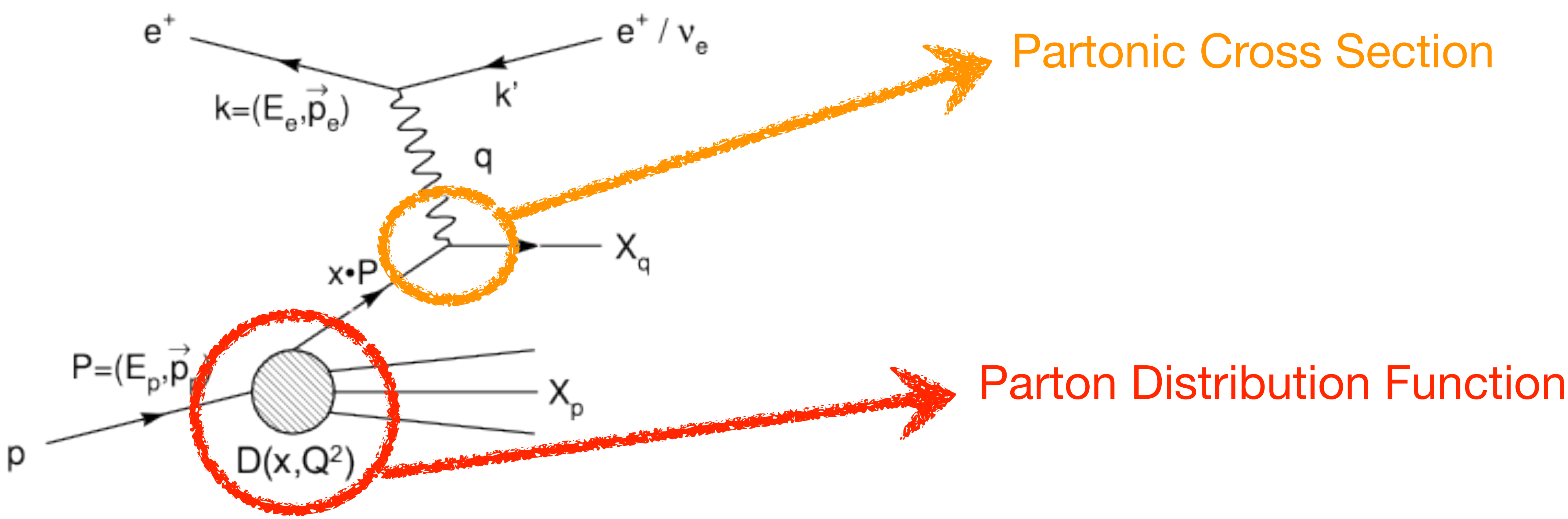
A PDF fitter view



Parton Distribution Functions

Collinear Factorization

- Consider a process with one hadron in the initial state (Deep Inelastic Scattering)



- The cross-section can be written as (**Factorization Theorem**)

$$d\sigma = \sum_a \int_0^1 \frac{d\xi}{\xi} D_a(x,\mu^2)\, d\hat{\sigma}_a\left(\frac{x}{\xi},\frac{\hat{s}}{\mu^2},\alpha_s(\mu^2)\right) + O\left(\frac{1}{Q^p}\right)$$

Parton Distribution Functions

DGLAP Evolution

- Parton Distribution Functions are non-perturbative objects and their numerical value at a given *x* and Q^2 cannot be computed in perturbative QCD (Lattice?)

- … but their scale dependence is described by evolution equations (DGLAP)

$$\frac{\partial q_i(x,\mu^2)}{\partial \ln \mu^2} = \frac{\alpha_s(\mu^2)}{2\pi}\left[P_{qq}(x) \otimes q_i(x,\mu^2)\right] + \frac{\alpha_s(\mu^2)}{2\pi}\left[P_{qg} \otimes g(x,\mu^2)\right]$$

$$\frac{\partial g(x,\mu^2)}{\partial \ln \mu^2} = \frac{\alpha_s(\mu^2)}{2\pi}\left[P_{gq}(x) \otimes \sum_i \left(q_i(x,\mu^2) + \bar{q}_i(x,\mu^2)\right)\right] + \frac{\alpha_s(\mu^2)}{2\pi}\left[P_{gg} \otimes g(x,\mu^2)\right]$$

- … where the splitting functions (P_{ij}) can be computed in QCD perturbation theory and are known up to NNLO

[**LO** - Dokshitzer; Gribov, Lipatov; Altarelli, Parisi (1977)]
[**NLO** - Floratos, Ross, Sachrajda; Gonzalez-Arroyo, Lopez, Yndurain; Curci, Furmanski, Petronzio (1981)]
[**NNLO** - Moch, Vermaseren, Vogt (2004)]

- Moreover, Parton Distributions Functions are **universal: determine** them from **lepton-hadron scttaring, use** them for predictions in **hadron-hadron collisions**

Parton Distribution Functions

DGLAP Evolution

- Parton Distribution Functions are non-perturbative objects and their numerical value at a given

HEP-EPS Prize 2015

- … but their scale dependence is described by evolution equations (DGLAP)

The 2015 High Energy and Particle Physics Prize of the European Physical Society has been awarded jointly to five theoretical physicists: James D. Bjorken (SLAC National Accelerator Laboratory, Stanford, USA) “for his prediction of scaling behaviour in the structure of the proton that led to a new understanding of the strong interaction” and to Guido Altarelli (University of Roma Tre, Rome, Italy and CERN, Geneva, Switzerland), Yuri Dokshitzer (Laboratory of Theoretical and High Energy Physics, Paris, France and St. Petersburg Nuclear Physics Institute, Gatchina, Russia), Lev N. Lipatov (National Research Center "Kurchatov Institute", Petersburg Nuclear Physics Institute, Gatchina, Russia) and Giorgio Parisi (University of Rome, La Sapienza, Rome, Italy) “for developing a probabilistic field theory framework for the dynamics of quarks and gluons, enabling a quantitative understanding of high-energy collisions involving hadrons”.

- … wher … theory and are known up to NNLO

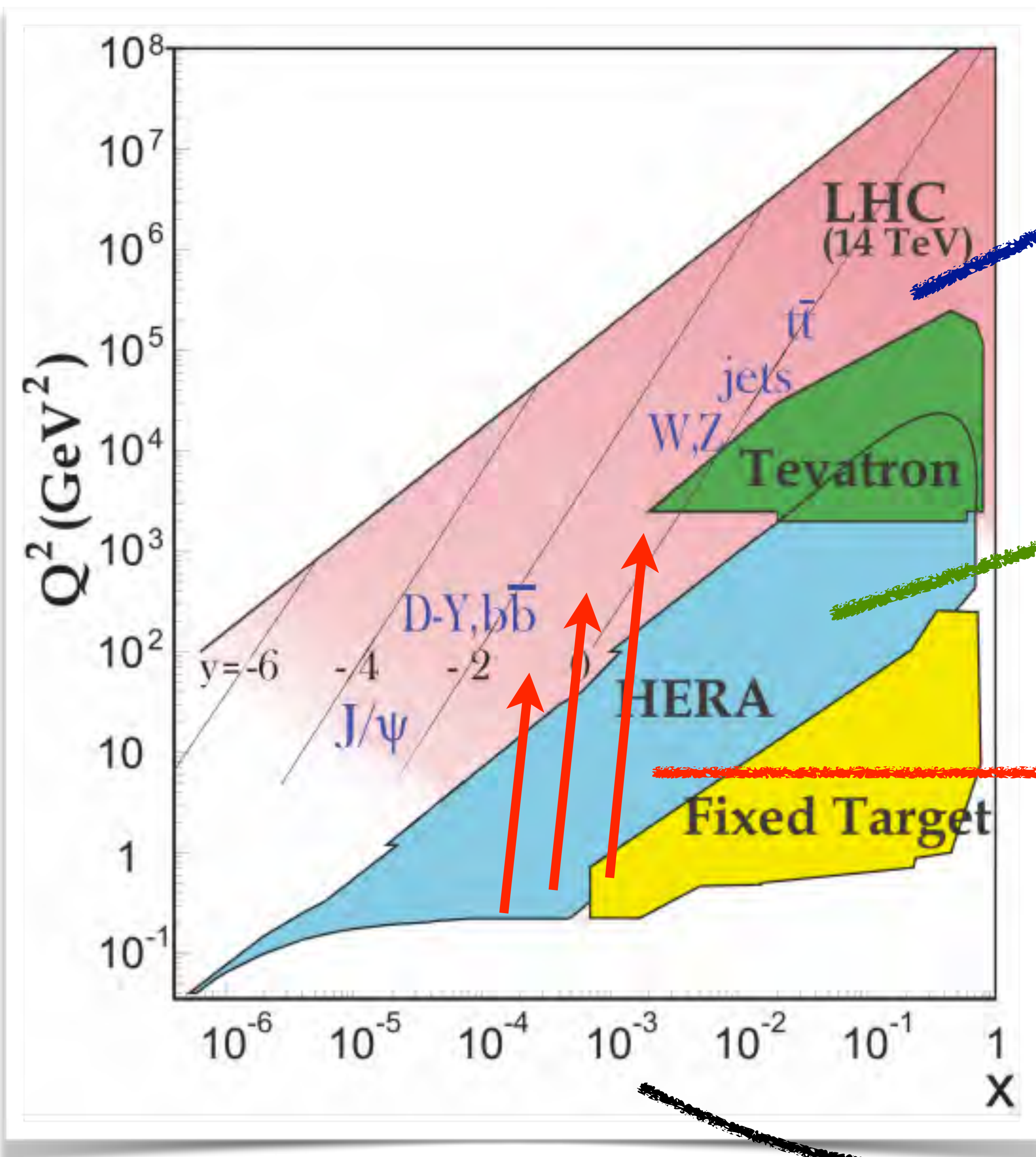
[**LO**

[**NLO** - Floratos, Ross, Sachrajda; Gonzalez-Arroyo, Lopez, Ynduraín; Curci, Furmanski, Petronzio (1981)]

[**NNLO**

Parton Distribution Functions

Where do we measure them, where do we use them



This is the region where we want to use PDFs for predictions at the LHC

This is the region where we measure PDFs, from DIS and Tevatron experiments

DGLAP evolution

x dependence of PDFs at the reference scale extracted from (global) fits to data

PDF determinations at the dawn of LHC Run II

State of the art, April 2015

	Dataset	Pert. Order	HQ Treatment	α	Param.	Uncert.
ABM12 [arXiv:1310.3059]	DIS Drell-Yan	NLO NNLO	FFN (BMSN)	Fit (multiple values available)	6 indep. PDFs Polynomial (25 param.)	Hessian (Δ
CT14 [preliminary]	**Global**	LO NLO NNLO	GM-VFNS (S-ACOT)	External (multiple values available)	6 indep. PDFs Polynomial (27 param.)	Hessian Tolerance
HERAPDF2.0 [preliminary]	DIS (HERA I+II)	NLO NNLO	GM-VFNS (TR)	External (multiple values available)	5 indep. PDFs Polynomial (14 param.)	Hessian (Δ
MMHT14 [arXiv:1410.3989]	**Global**	LO NLO NNLO	GM-VFNS (TR)	Fit (multiple values available)	7 indep. PDFs Polynomial (37 param.)	Hessian Dyn. Tolerance
NNPDF3.0 [arXiv:1410.8849]	**Global**	LO NLO NNLO	GM-VFNS (FONLL)	External (multiple values available)	7 indep. PDFs Neural Nets (259 param.)	Monte Carlo

[LHAPDF v6.1.5 - http://lhapdf.hepforge.org/]

NNPDF Methodology Interlude

R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte,
P. Groth-Merrild, N. P. Hartland, Z. Kassabov, J. I. Latorre, J. Rojo, M. Ubiali & AG

NNPDF Methodology
... in a Nutshell

- **Generate** N_{rep} **Monte Carlo replicas** of the experimental data, taking into account all experimental correlations

- **Fit** a set of Parton Distribution Functions, parametrized at the initial scale using Neural Networks, **to each replica**

- **Expectation values** for observables are then given by

$$\langle \mathcal{F}[f_i(x,Q^2)] \rangle = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} \mathcal{F}\left(f_i^{(net)(k)}(x,Q^2) \right)$$

- .... and corresponding formulae for the estimators of Monte Carlo samples are used to compute **uncertainties**, **correlations**, etc.

NNPDF Methodology
Main Ingredients

- **Monte Carlo** determination of **uncertainties**
  - **No** need to rely on **linear propagation** of errors
  - Possibility to test the impact of **non-gaussianly** distributed uncertainties
  - Possibility to test for **non-gaussian behaviour** of uncertainties of fitted PDFs
- Parametrization of PDFs using **Neural Networks**
  - Provide an **unbiased parametrization**
- Determine the **best fit** PDFs using **Cross-Validation**
  - Ensures **proper fitting**, avoiding overlearning

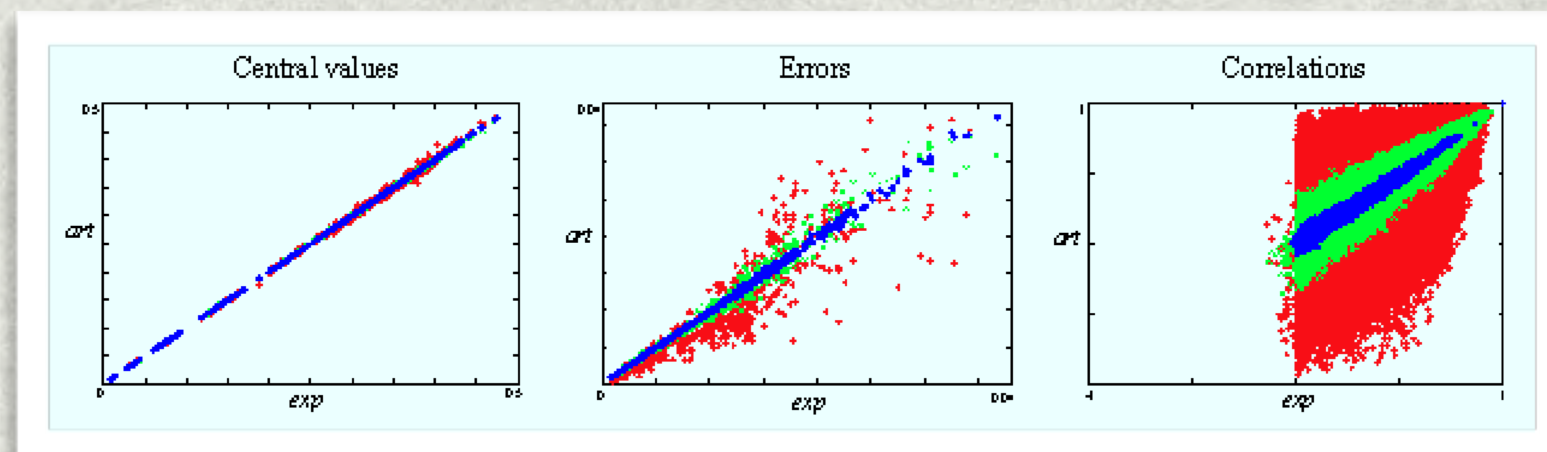
NNPDF Methodology
Monte Carlo replica generation

- **Monte Carlo replicas** are generated according to the distribution

$$O_i^{(art)\,(k)} = (1 + r_N^{(k)}\,\sigma_N)\left[O_i^{(exp)} + \sum_{p=1}^{N_{sys}} r_p^{(k)}\,\sigma_{i,p} + r_{i,s}^{(k)}\,\sigma_s^i\right]$$

  where r_i are (gaussianly distributed) random numbers

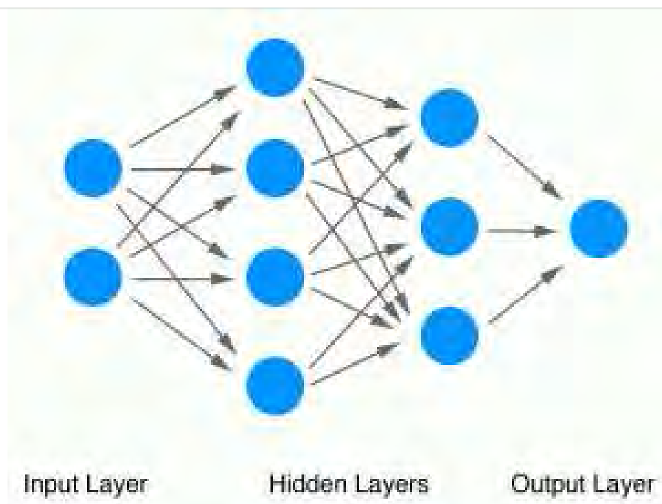
- **Validate** Monte Carlo replicas against experimental data



- O(1000) replicas needed to **reproduce correlations** in experimental data to percent accuracy

NNPDF Methodology

PDF parametrization using Neural Networks



$$\xi_i^{(l)} = g\left(\sum_j^{n_l-1} \omega_{ij}^{(l-1)} \xi_j^{(l-1)} - \theta_i^{(l)}\right)$$

$$g(x) = \frac{1}{1+e^{-x}}$$

- **Artificial Neural Networks** provide us with a parametrization for PDFs at the initial scale which is extremely **redundant** and **robust against variations**
- Very **efficient algorithms** are available which allow us to train NN (efficient fit to **large datasets** in a **very high dimensional parameter space**)
- ... but in the end they are **just another basis of functions**

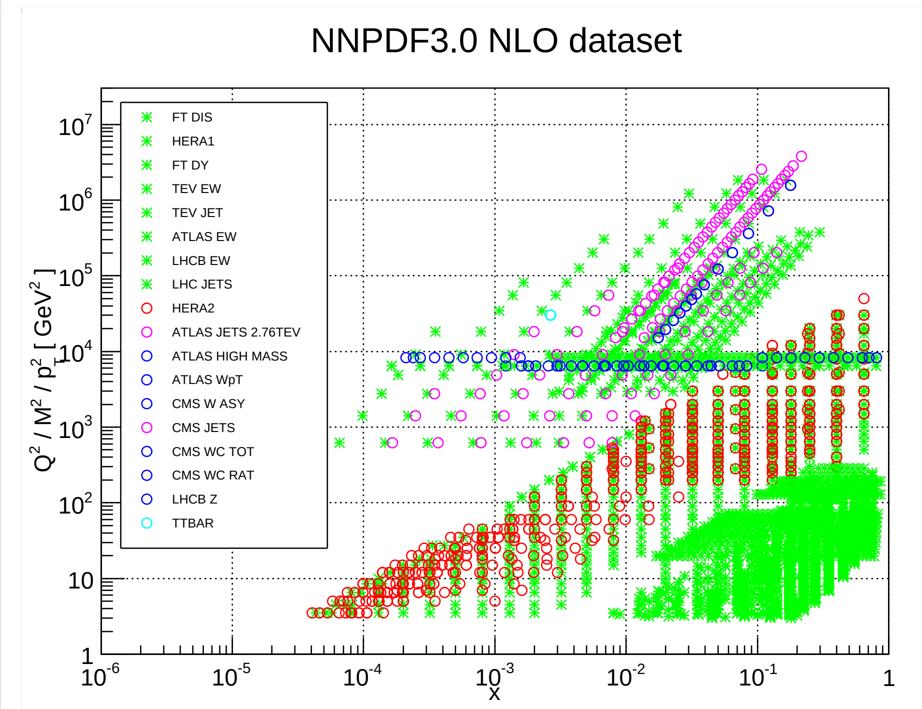
Parton Distributions for LHC Run II

… mostly based on [arXiv:1410.8849]

Parton Distributions for LHC Run II

Dataset

	NNPDF3.0	MMHT14	CT14(prel)
SLAC p,d DIS	✔	✔	✗
BCDMS p,d DIS	✔	✔	✔
NMC p,d DIS	✔	✔	✔
E665 p,d DIS	✗	✔	✗
CDHSW nu-DIS	✗	✗	✔
CCFR nu-DIS	✗	✔	✔
CHORUS nu-DIS	✔	✔	✗
CCFR dimuon	✗	✔	✔
NuTeV dimuon	✔	✔	✔
HERA I NC,CC	✔	✔	✔
HERA I charm	✔	✔	✔
H1,ZEUS jets	✗	✔	✗
H1 HERA II	✔	✗	✗
ZEUS HERA II	✔	✗	✗
E605 & E866 FT DY	✔	✔	✔
CDF & D0 W asym	✗	✔	✔
CDF & D0 Z rap	✔	✔	✔
CDF Run-II jets	✔	✔	✔
D0 Run-II jets	✗	✔	✔
ATLAS high-mass DY	✔	✔	✔
CMS 2D DY	✔	✔	✗
ATLAS W,Z rap	✔	✔	✔
ATLAS W pT	✔	✔	✗
CMS W asy	✔	✔	✔
CMS W +c	✔	✗	✗
LHCb W,Z rap	✔	✔	✔
ATLAS jets	✔	✔	✔
CMS jets	✔	✔	✔
ttbar tot xsec	✔	✔	✗
TOTAL NLO	4276	2996	3248
TOTAL NNLO	4078	2663	3045

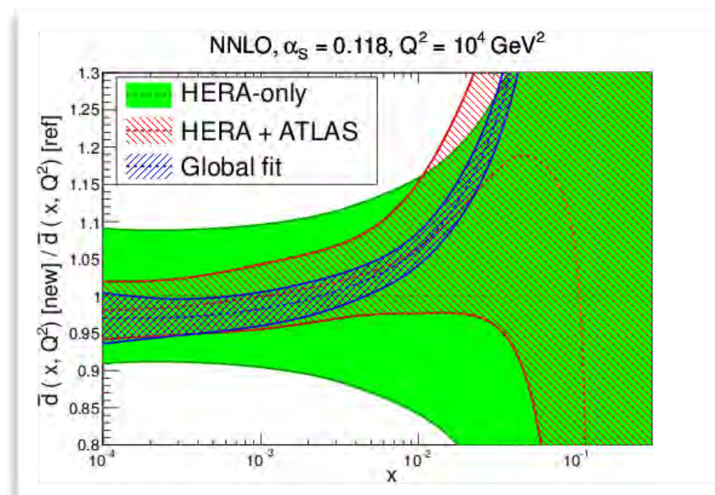


Modern PDF sets include a substantial number of data from the LHC experiments.

Parton Distributions for LHC Run II

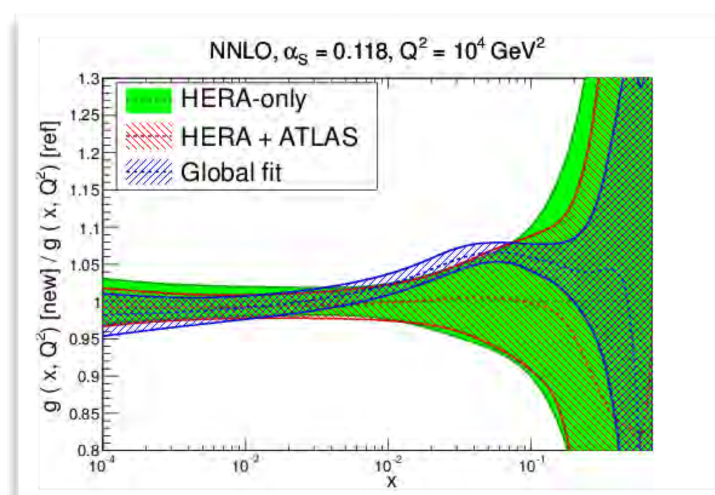
The reason for global fits

- Using a **wide range of measurements** coming from **different experiments** is **crucial to constrain** different **PDF combinations** over the whole kinematic range



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

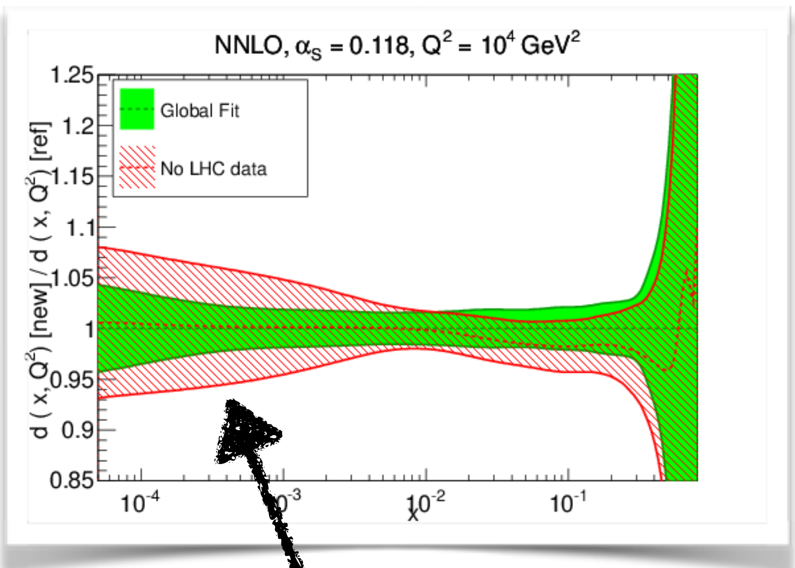
MSTW08, arXiv:0901.0002



Parton Distributions for LHC Run II

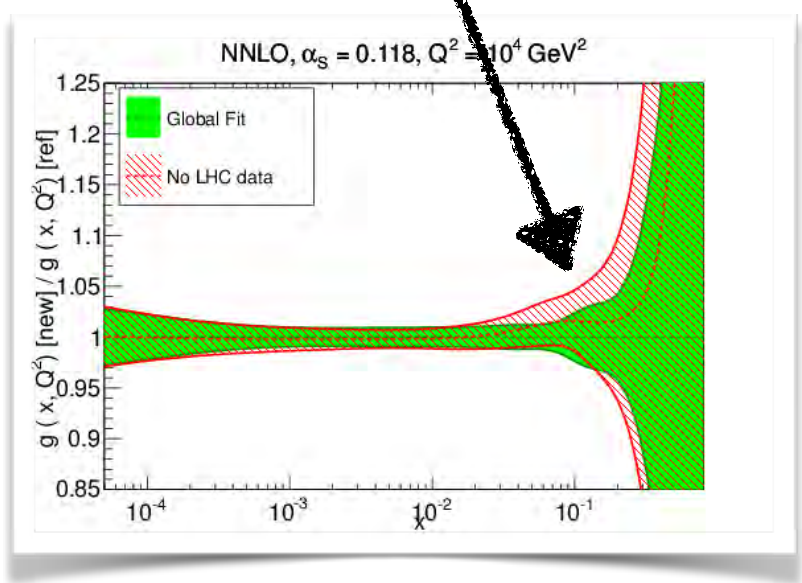
Impact of LHC data

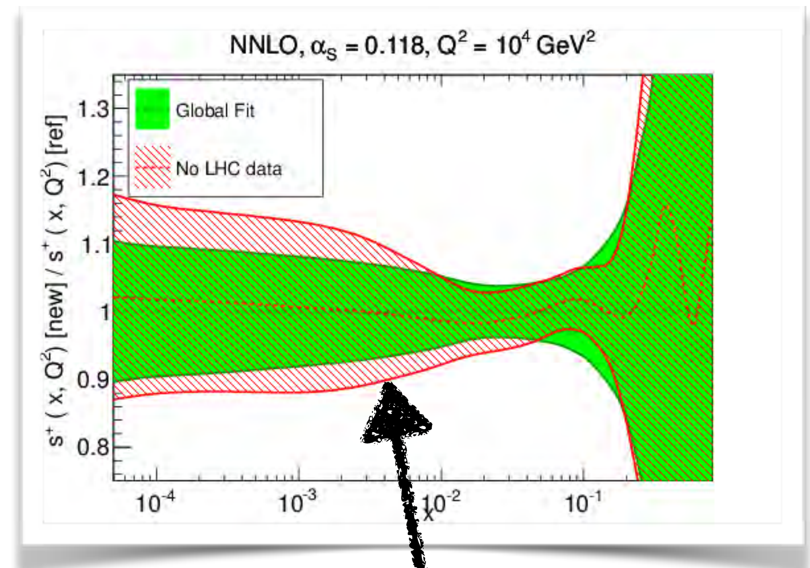
- Impact of LHC data (still) moderate but definitely noticeable



Light flavours
W asymmetry &
(2D) Drell-Yan

Large-*x* gluon
inclusive jet & top





Strangeness
W+charm

Methodology - Closure tests

- **Core idea**
  - **Assume** the **underlying PDFs** are known and **generate pseudodata** based on a **given theory set up** and the chosen PDF set.
  - Decide data uncertainties: **zero**, as in **real data**, study impact of **inconsistent datasets**, …
  - **Fit PDFs** to the generated **pseudodata** using the **same theory setup.**
  - **Check** if the **fitted PDFs** reproduce the **underlying truth**:
    - are true values gaussianly distributed around the fit?
    - are uncertainties a faithful reproduction of input experimental ones?
    - are the results stable upon variations of the fitting methodology?

NNPDF3.0 is the **first PDF determination** based on **Closure Tests**

Parton Distributions for LHC Run II

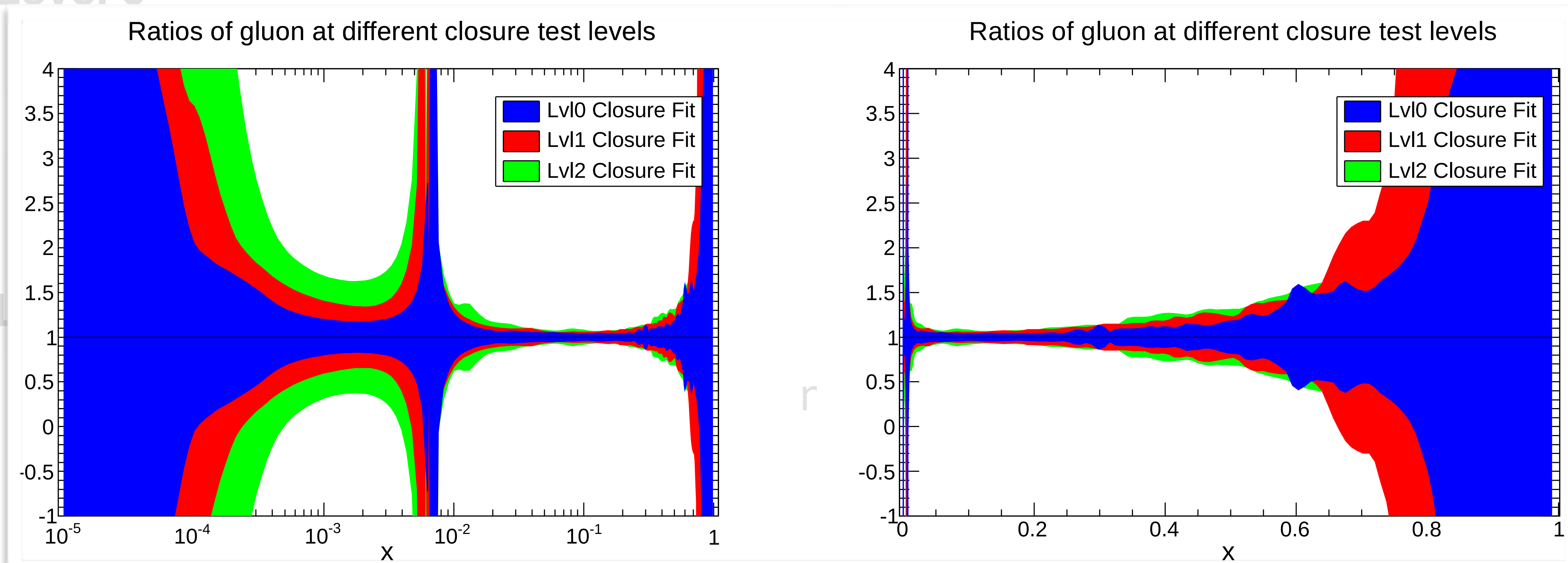
Three levels of Closure tests

- **Level 0**
  - **Fake data** are generated **without uncertainty**
  - Test for **efficiency and adequacy** of fitting **methodology**
  - Determine **interpolation** and **extrapolation uncertainty**
- **Level 1**
  - Fake data are generated with the same uncertainty of real data (correlations).
  - **No "data replicas"**, fit to the same data over and over.
  - Determine **functional uncertainty**, due to infinity of equivalent minima
- **Level 2**
  - Replicas are fitted to **fake data replicas**
  - Determine **data uncertainty**

Parton Distributions for LHC Run II

Three levels of Closure tests

- Level 0



- Level 2

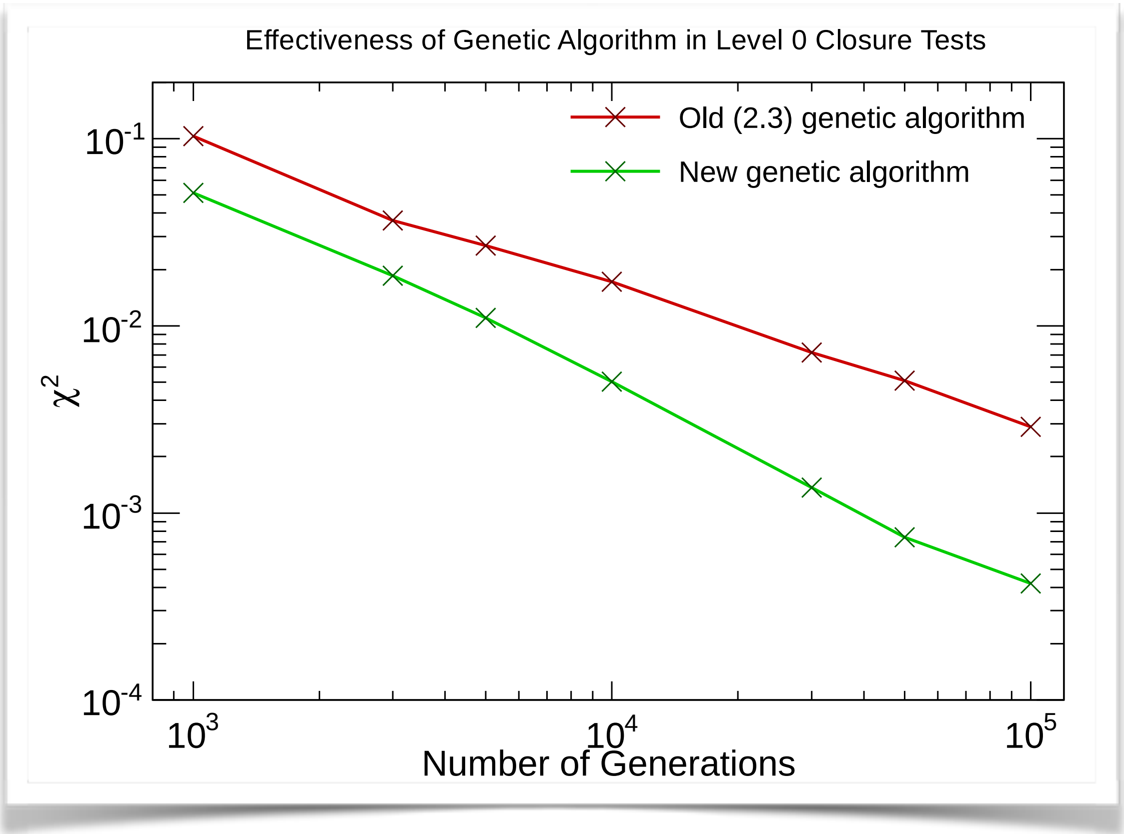
The three **sources of uncertainty** are **comparable** in the **data region**

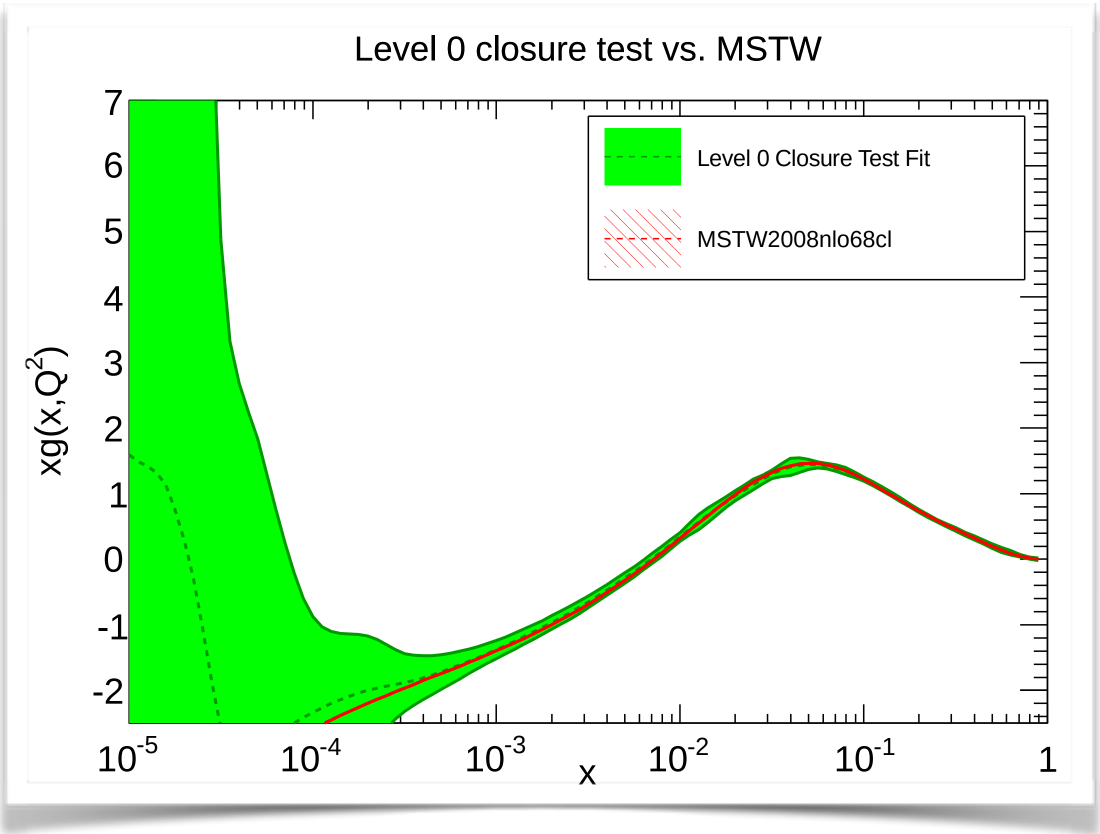
- Determine

Parton Distributions for LHC Run II

Closure test results - Level 0

- Assume **vanishing experimental uncertainty** on generated data
- **Perfect description of data** ($\chi^2 = 0$) must be **possible** with **adequate** fitting **methodology**





Parton Distributions for LHC Run II

Closure test results - Level 2

Central Values

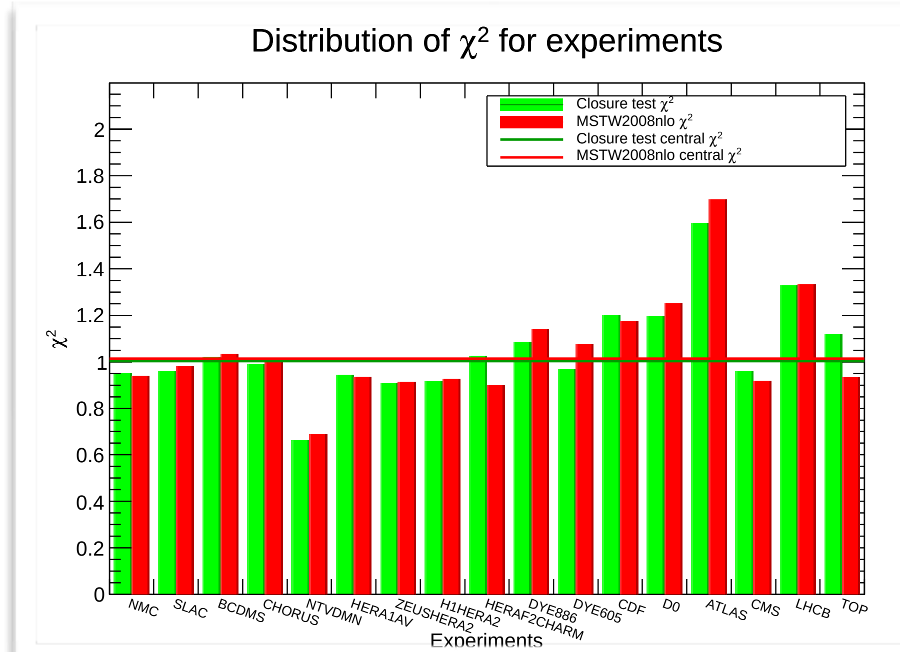
compare fitted vs. true χ^2 both for individual experiments and the entire dataset

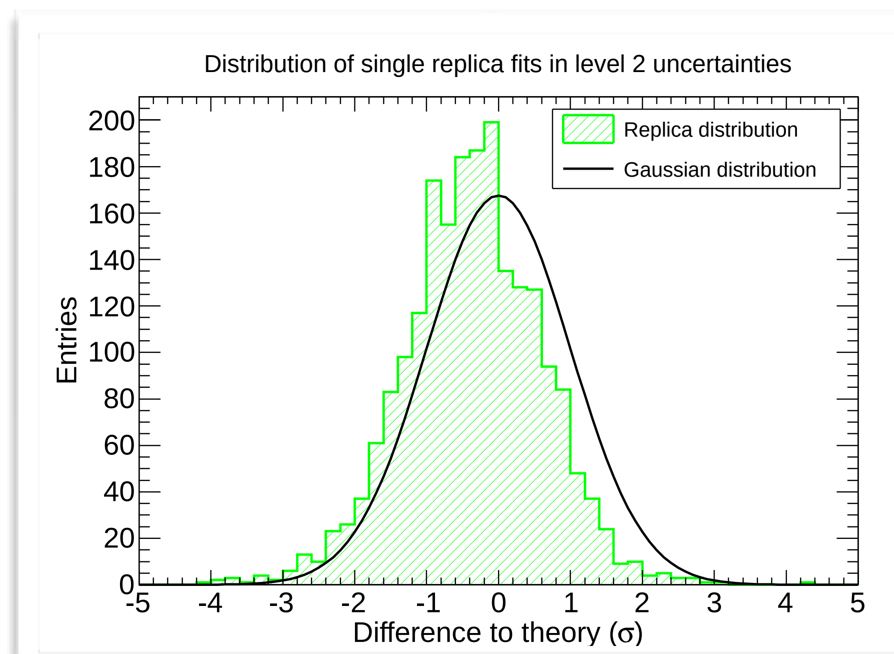
Total dataset

$$\Delta\chi^2 = 0.001 \pm 0.003$$

Uncertainties

compare the distribution of deviations between fitted and true PDFs to theoretical expectation (gaussian)



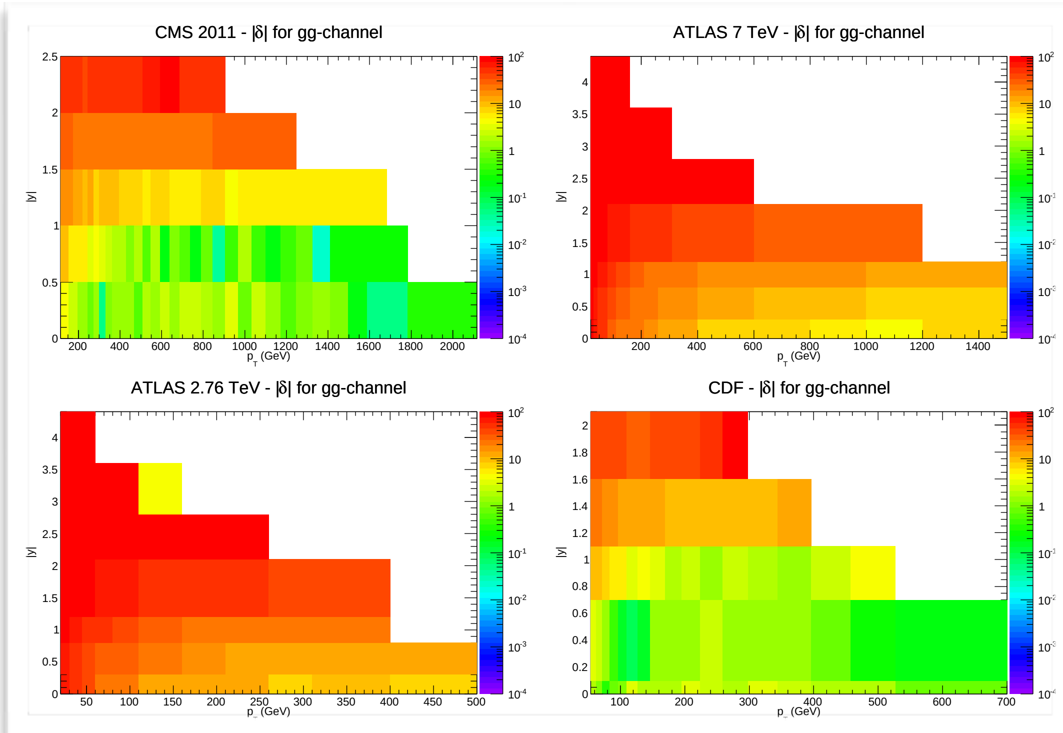


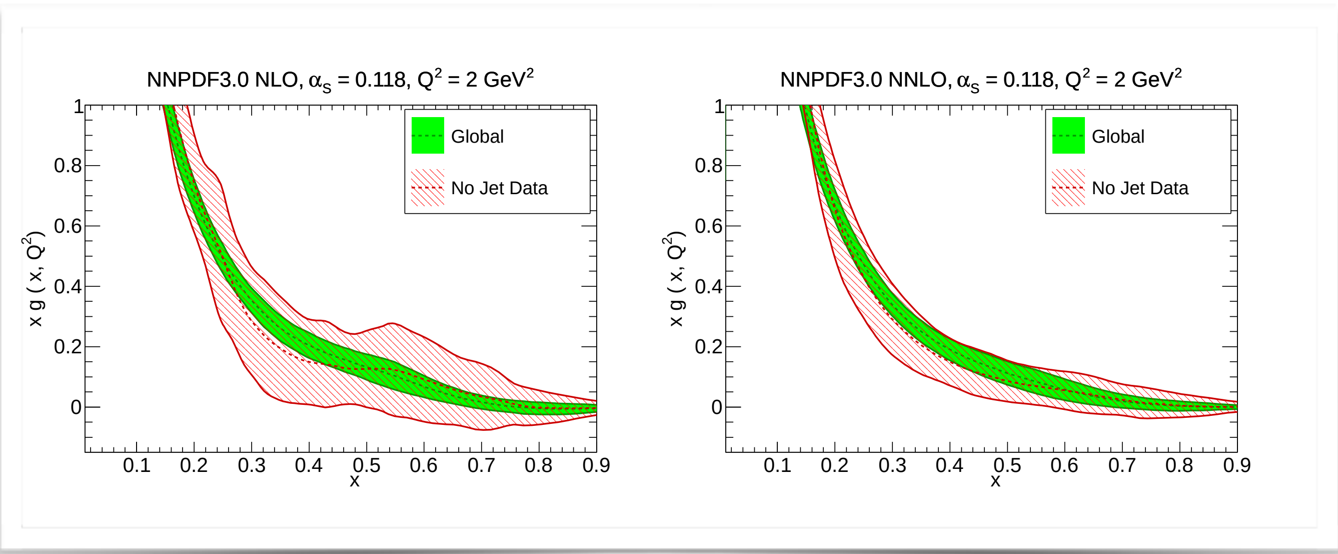
Parton Distributions for LHC Run II

Theory improvements - Inclusive Jet data

- Systematic comparison of exact NNLO computation for the *gg*-channel (Gehrmann et al, 2014) and threshold approximation (De Florian et al, 2014) to determine which data points to include in the NNLO fit.

- Threshold resummation only accurate in the central rapidity and high p_T region (Carrazza & Pires, 2014)





Impact of **jet data** on **large-x gluon** sizeable both at NLO at NNLO despite reduced number of data points

Parton Distributions for LHC Run II

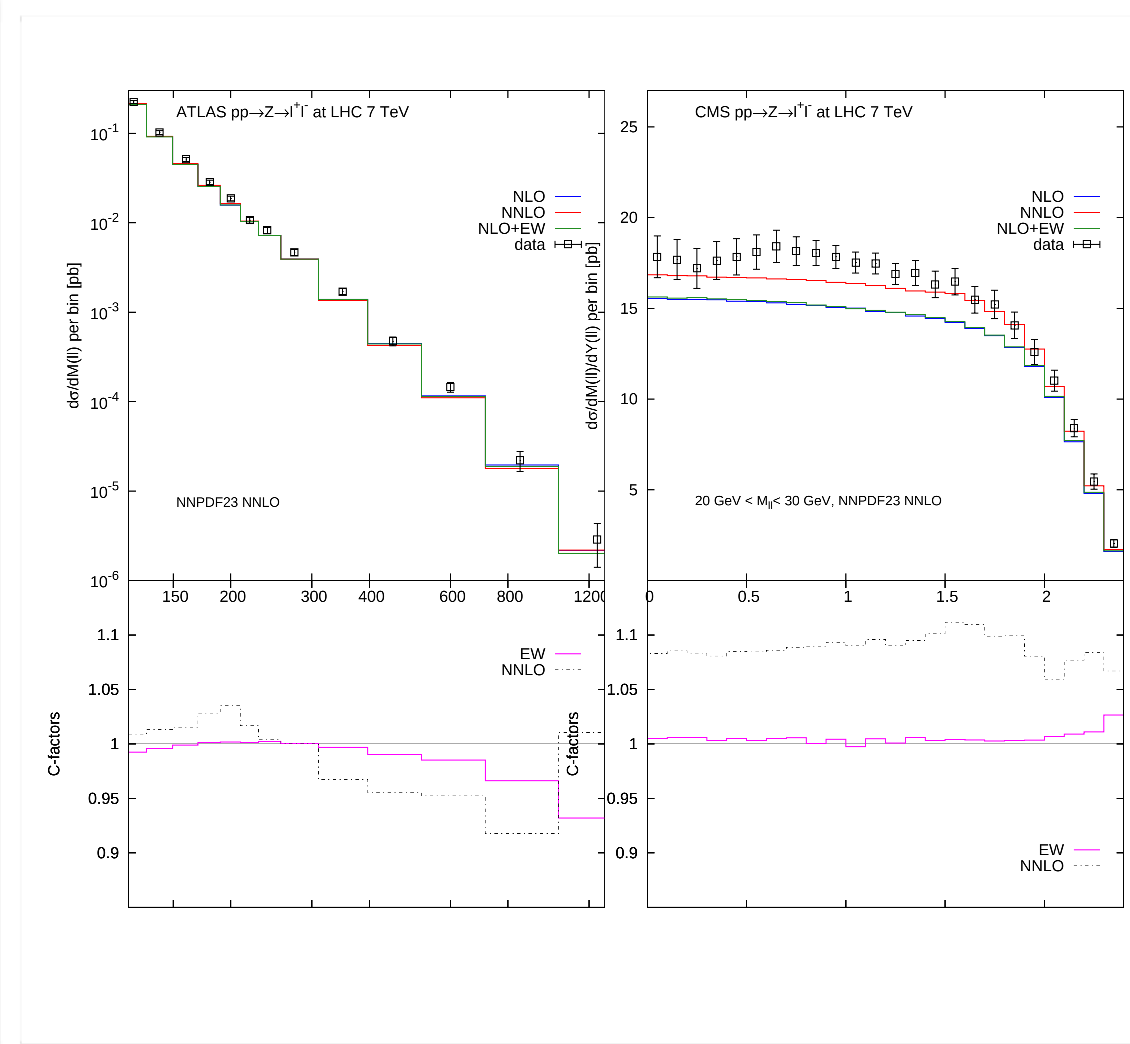
Theory improvements - Electroweak corrections

- Virtual **pure EW corrections** taken into account for all **neutral Drell-Yan** datasets (CMS double differential, ATLAS *Z*-peak and high-mass).

- Use data corrected for Final State Radiation

- **QED corrections** checked, **highest invariant mass bins** for ATLAS and CMS excluded from the fit because of **large corrections due to photon-initiated processes**

- Electroweak corrections still missing for DIS, *W* production, inclusive jet and top pair production



Parton Distributions for LHC Run II

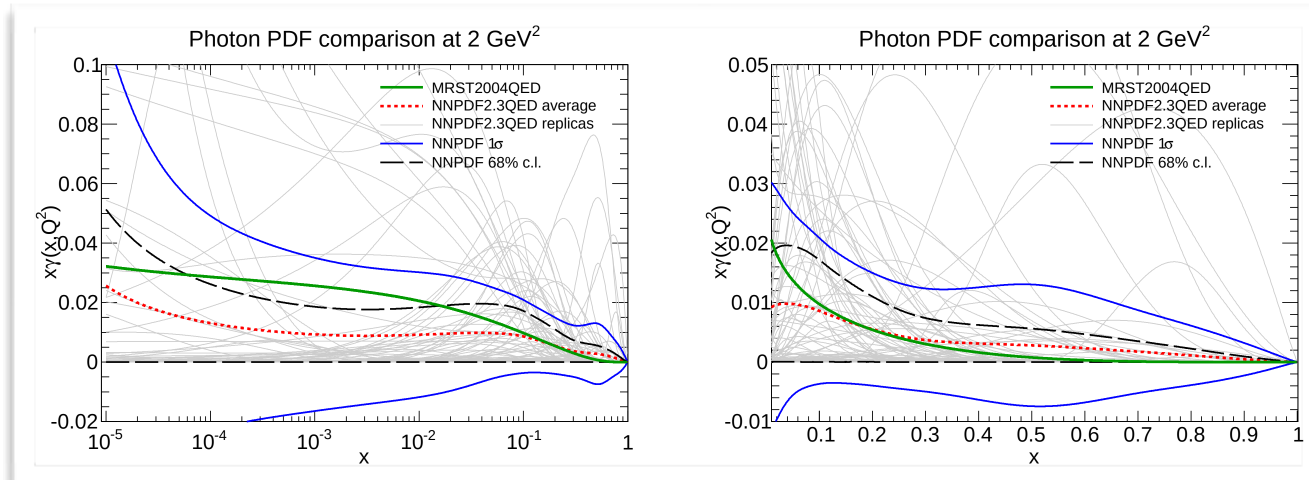
Parton Distributions with QED corrections (NNPDF2.3 QED)

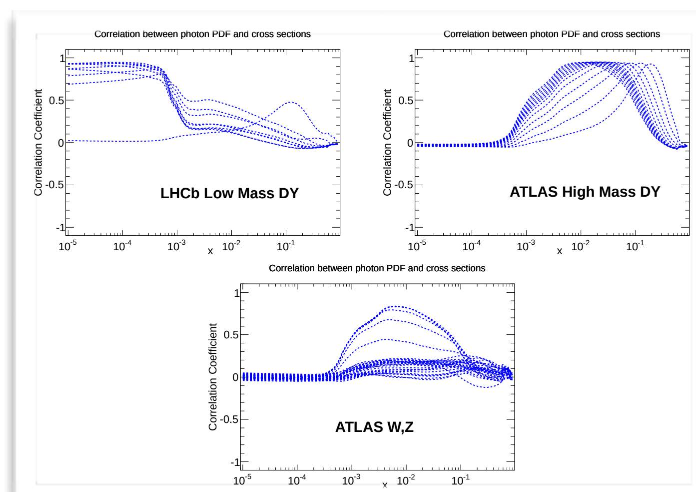
- Precision LHC phenomenology, including EW effects, requires **parton distributions with QED effects** included in the evolution and a **photon PDF**

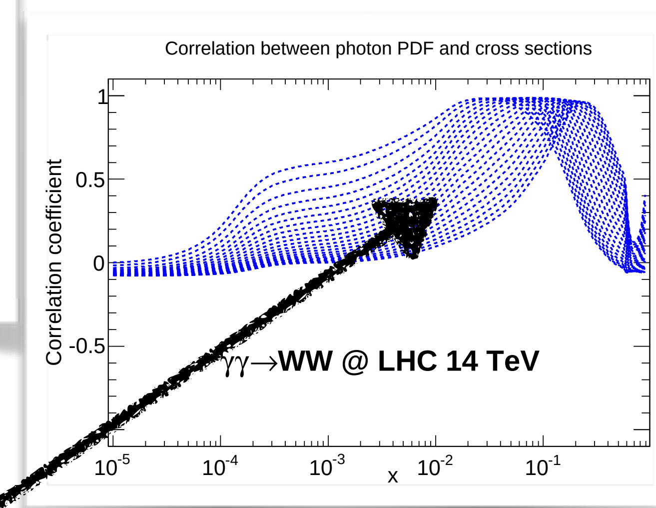
- **NNPDF2.3 QED** is the most recent PDF fit based on **(N)NLO QCD + LO QED** evolution and with a photon PDF determined from DIS and Drell-Yan (low-mass LHCb, *W* & *Z* peak and high-mass ATLAS) production

R. D. Ball et al, [arXiv:1308.0598]

- LHC data are crucial for a reliable determination of the photon PDF







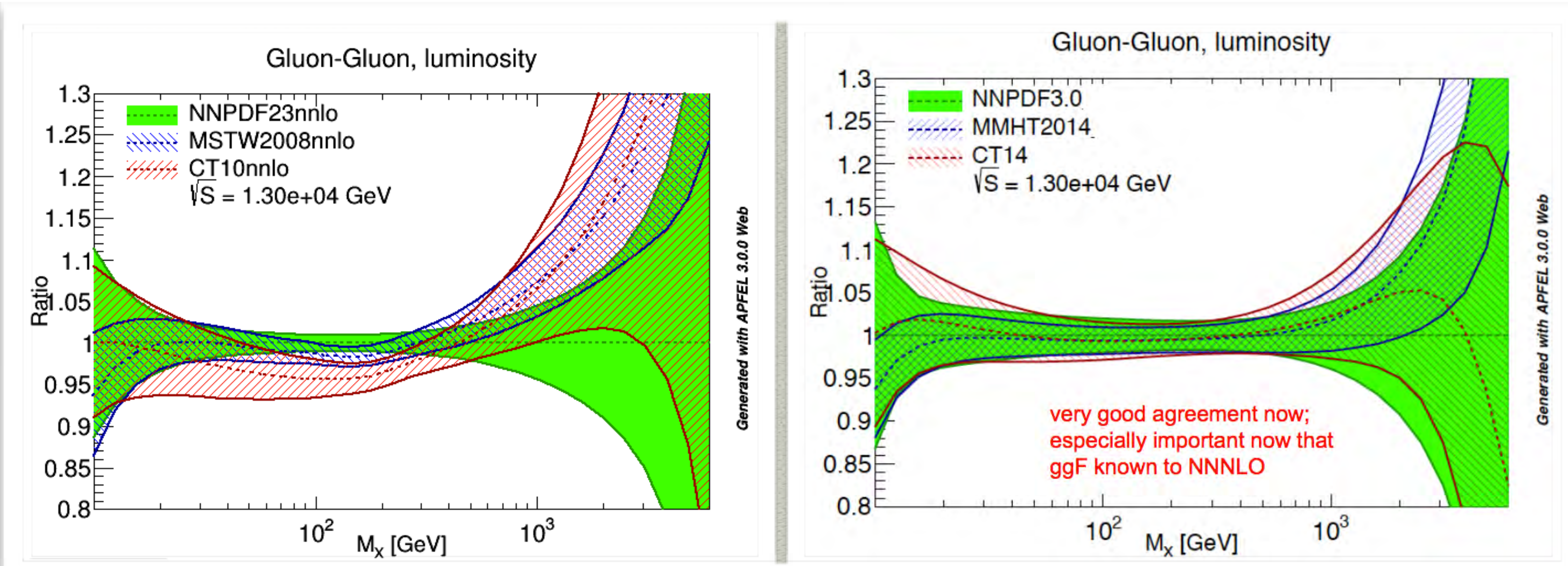
Photon PDF strongly correlated with **WW production** at the LHC

Parton Distributions for LHC Run II

Improved agreement among global PDF sets

"Progress in convergence between the parton distribution functions will also be needed in order to reduce the theoretical uncertainties below the experimental measurement uncertainties"

J. Ellis, [arXiv:1504.03654]



J. Houston, PDF4LHC 2015

Almost perfect agreement among the newest releases of global PDF sets for ggH

	CT14	MMHT2014	NNPDF3.0
8 TeV	18.66 pb -2.2% +2.0%	18.65 pb -1.9% +1.4%	18.77 pb -1.8% +1.8%
13 TeV	42.68 pb -2.4% +2.0%	42.70 pb -1.8% +1.3%	42.97 pb -1.9% +1.9%

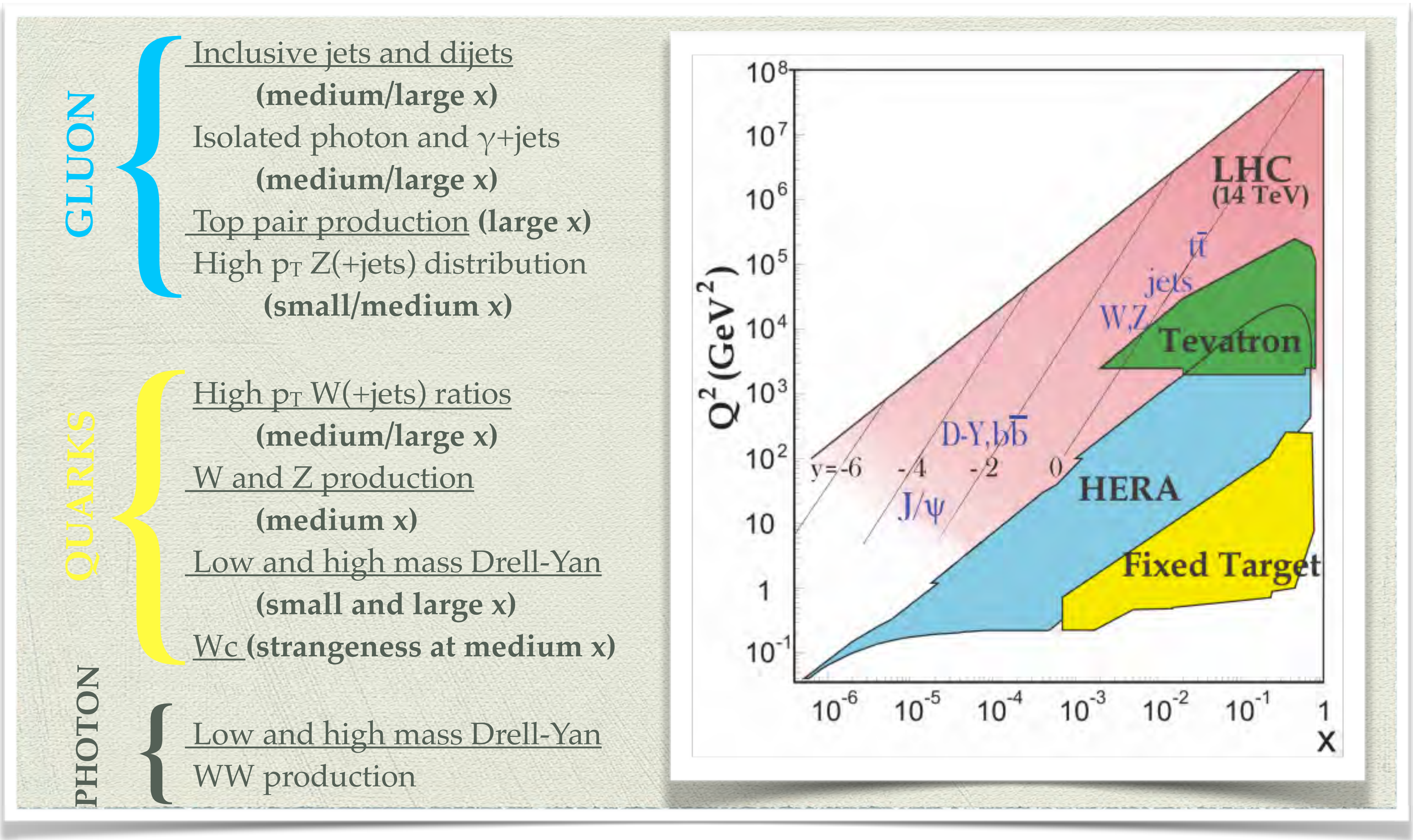
The road ahead…

The Future of PDF fits

More data

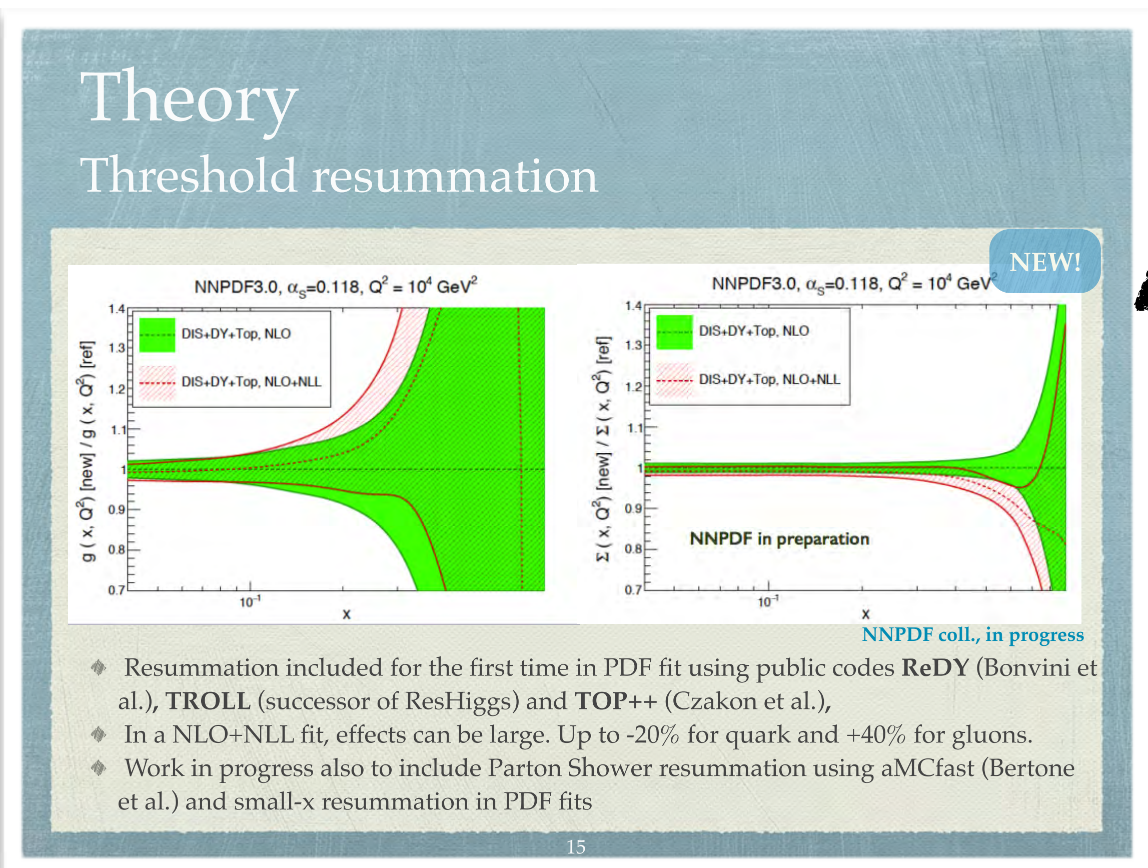
- **LHC data** (both from Run I and Run II) to provide **substantial constraints** on PDFs in the (near) future

M.Ubiali, SM@LHC 2015

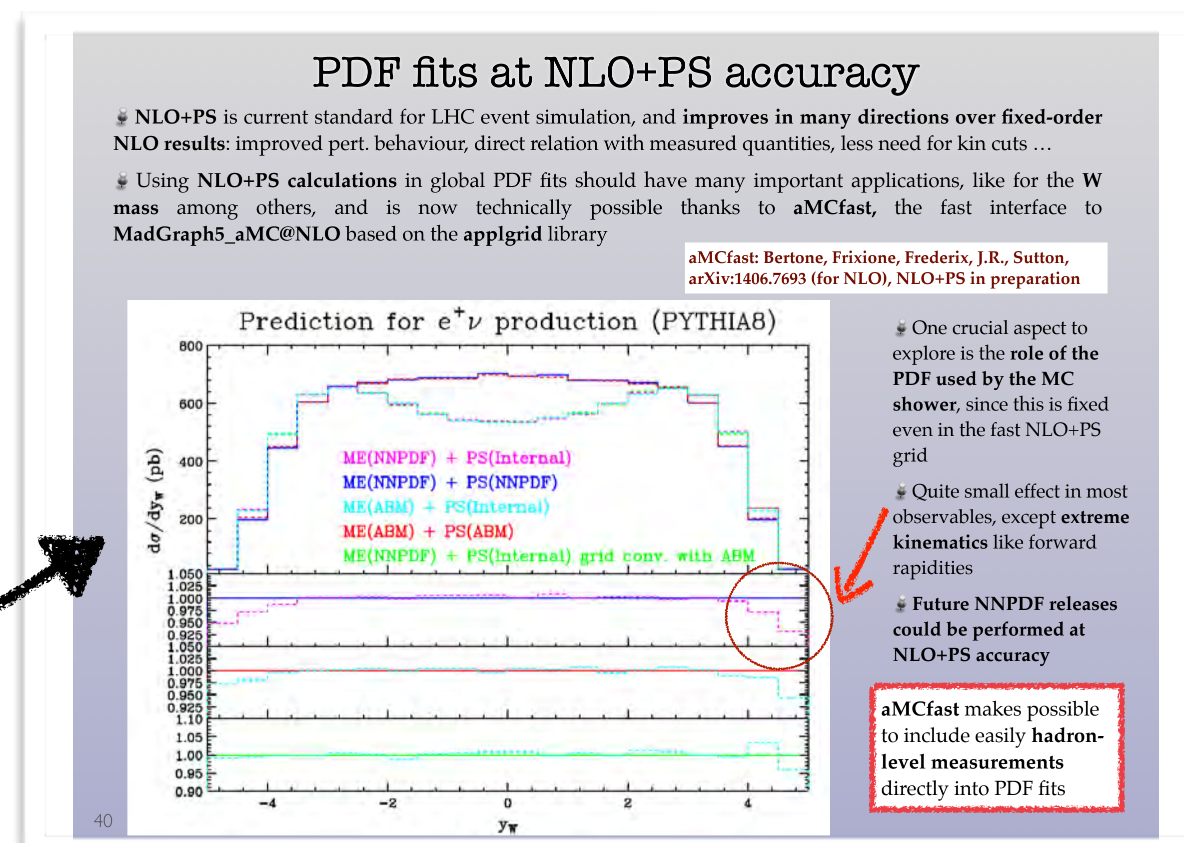


The Future of PDF fits

Theory refinements



PDFs including large-*x* resummation effects expected to improve predictions in regions important for high mass searches



PDFs including Parton Shower effects to be used in combination with NLO Monte Carlo codes

The Future of PDF fits

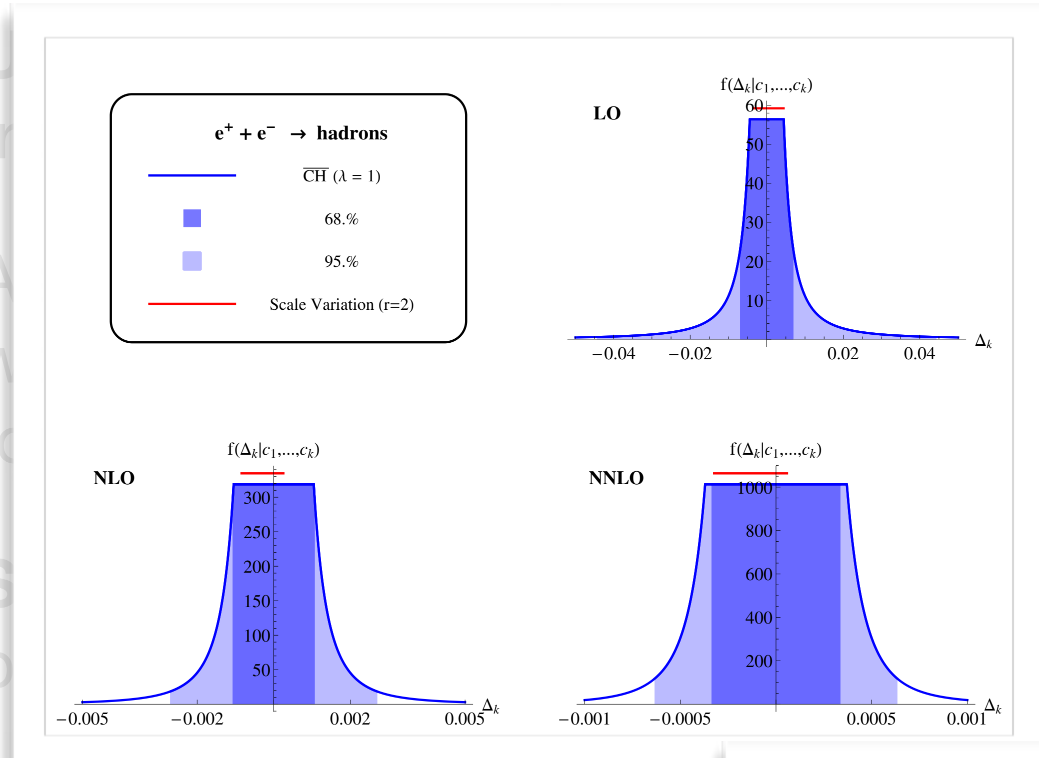
Theoretical uncertainties

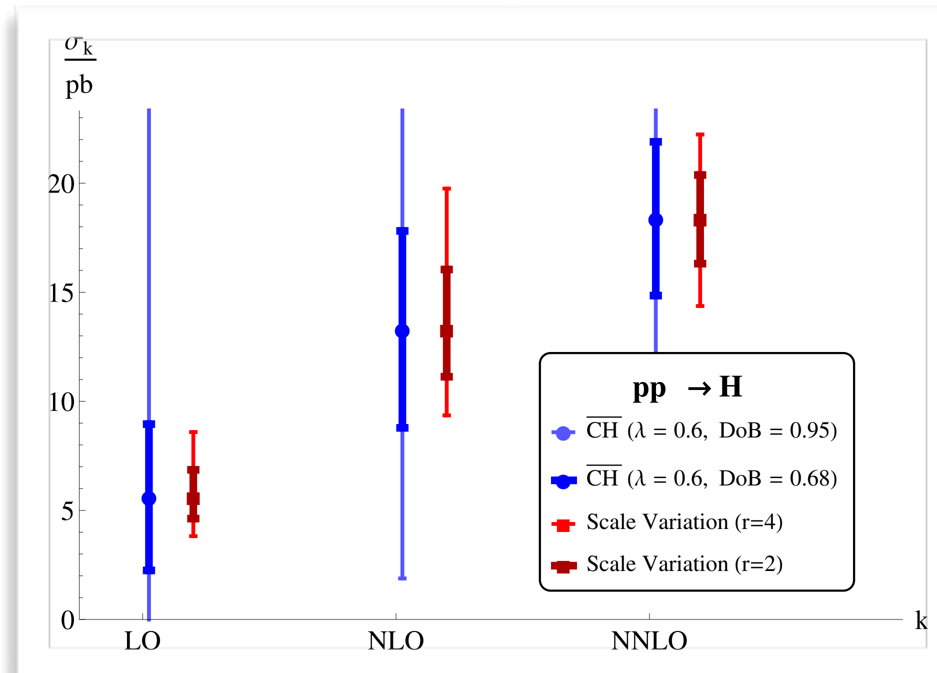
- **Uncertainties on PDFs** only reflect **experimental uncertainties** propagated from the data included in the fit
- As **data become more precise** and **constraining** we should probably begin to worry about the **uncertainties on the theoretical predictions** we use in the fits, for example from **missing higher orders**
- **Scale variations** is the **conventional way** to estimate theoretical uncertainties, but the **resulting uncertainty intervals** have **no statistical interpretation**
- **Inclusion in PDF fits** is at best **ambiguous**
- Recently developed **Bayesian framework** to estimate theoretical uncertainties might provide a way forward … **work in progress!**

[E. Bagnaschi, M. Cacciari, L. Jenniches & AG, arXiv:1409.5036]

The Future of PDF fits

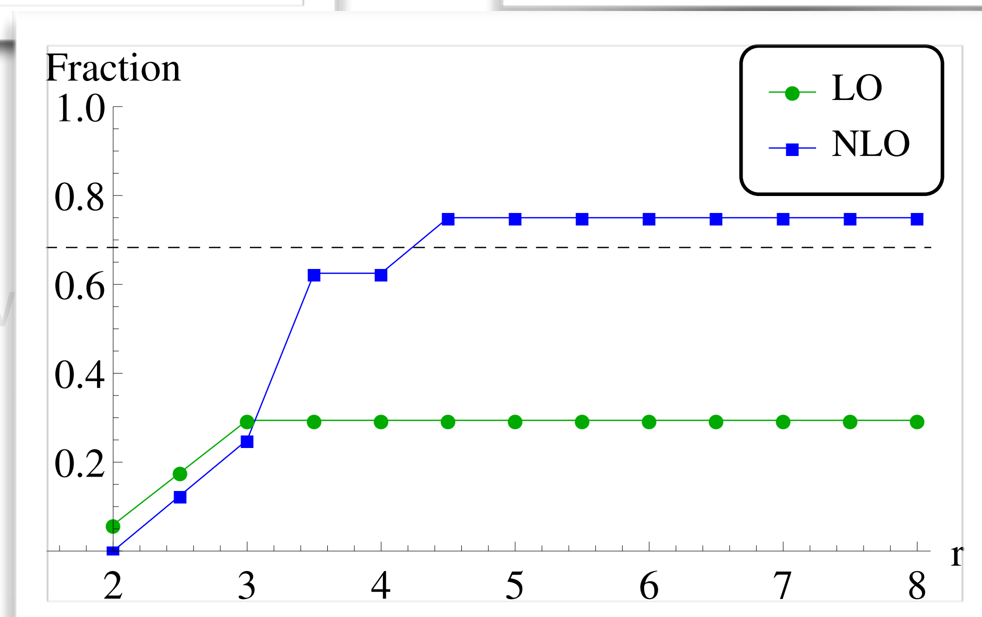
Theoretical uncertainties





- **Inclusion in PDF fits**

- Recently developed might provide a way forw



niches & AG, arXiv:1409.5036]

Conclusions & Outlook

Final thoughts

- An **accurate knowledge of Parton Distribution Functions** with a **reliable estimate of** their **uncertainties** is a **crucial ingredient** to exploit the potential of the **LHC experiments**
- Parton Distribution Function determinations evolved substantially in the last years, in every single aspect: data, theoretical input, fitting methodology …
- … agreement among global PDF fits (still based on different ingredients) for crucial observables like Higgs production in gluon-gluon fusion are a sign of progress …
- The future promises
  - More high precision data from the LHC providing constraints on PDFs
  - More refined theoretical predictions (EW effects, resummations, …)
  - … we might finally start thinking about theoretical uncertainties