ON THE PDF FRONTIER: NNPDF

Maria Ubiali

(University of Cambridge)

Stress testing of the SM at the LHC 23-27 May 2016 KITP, Santa Barbara, USA

NNPDF

The NNPDF collaboration:

J. I. Latorre (Barcelona) MU (Cambridge) S. Carrazza (CERN)

R. D. Ball, L. Del Debbio, P. G. Merrild (Edinburgh)

Z. Dim, S. Forte (Milano), A. Guffanti (Torino)

V. Bertone, N. Hartland, E. Nocera, J. Rojo, L. Rottoli (Oxford)

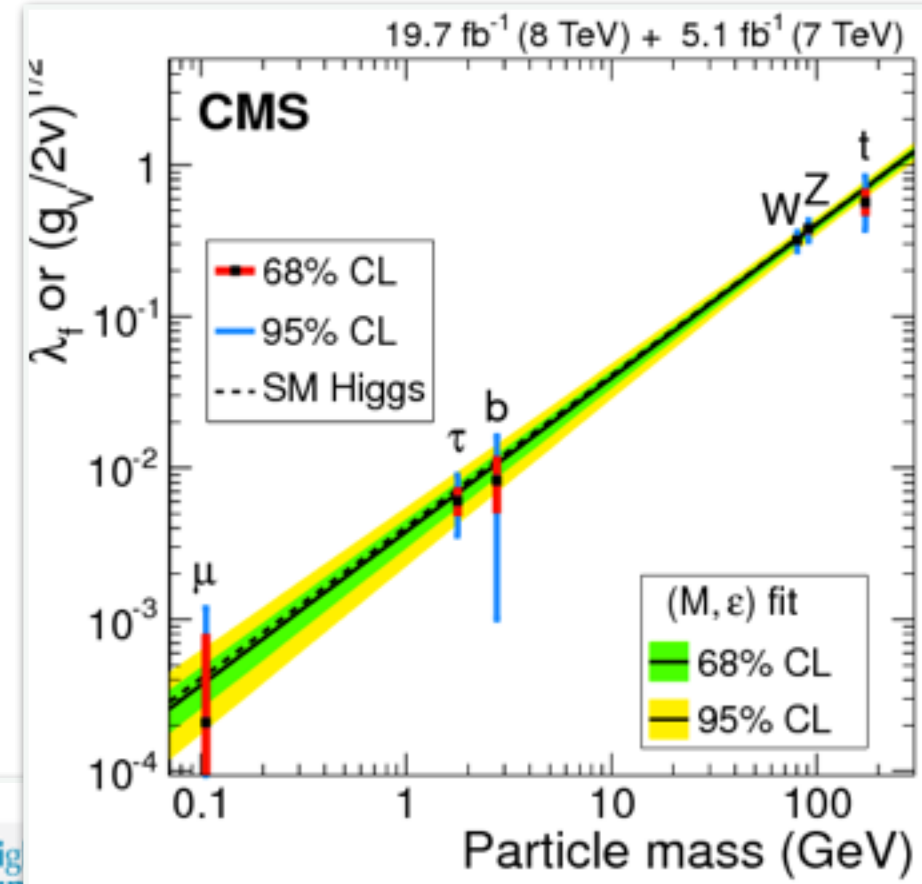
Outline of the talk

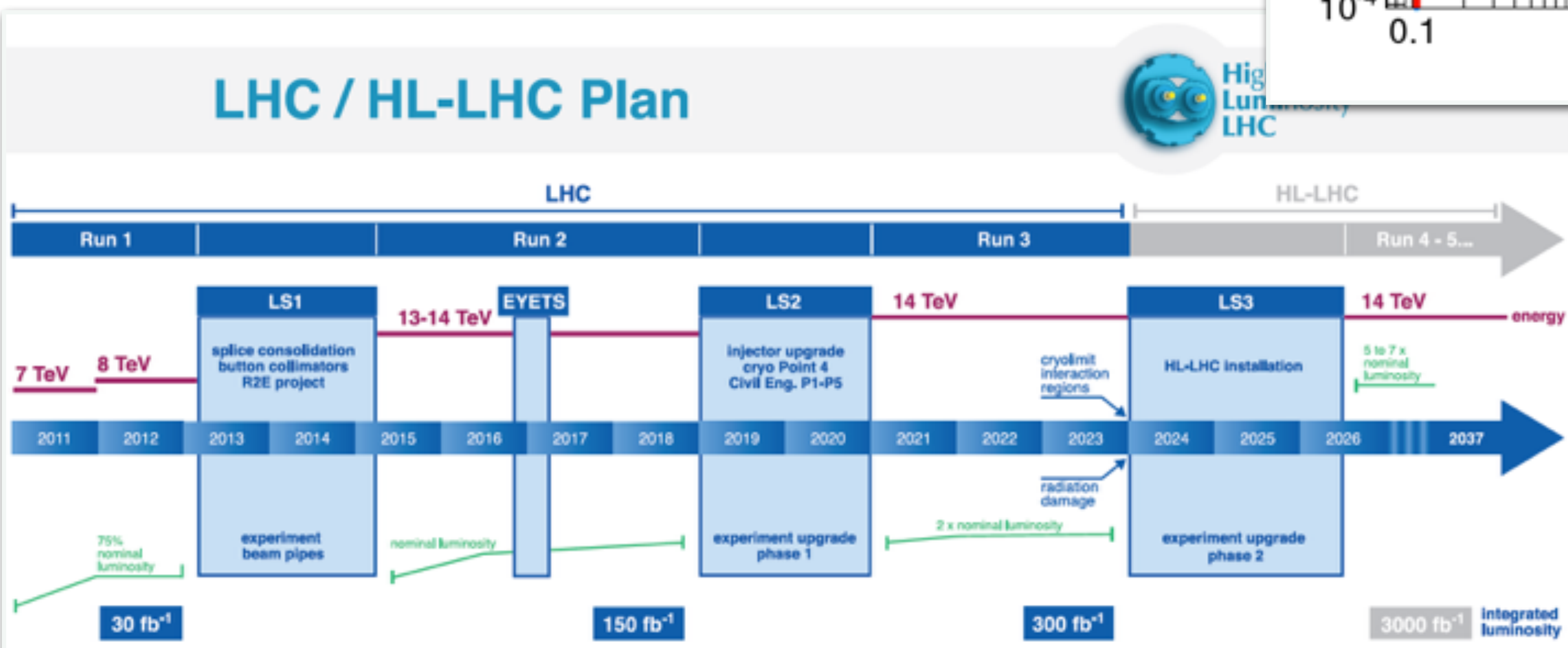
- Motivation
- The state of the art
    - The PDF extraction process
    - The NNPDF approach & comparison to traditional approaches
- Beyond the state of the art
    - New data & need for precision theory
    - Hidden uncertainties
- Conclusions and outlook

LHC physics at Run II

- Is precision physics possible/necessary at hadron colliders?
  At the LHC a paradigm shift is happening and theory has now to catch up with experimental precision

- Precise theoretical predictions are key to indirectly spot new physics signals and to characterise any “bump”

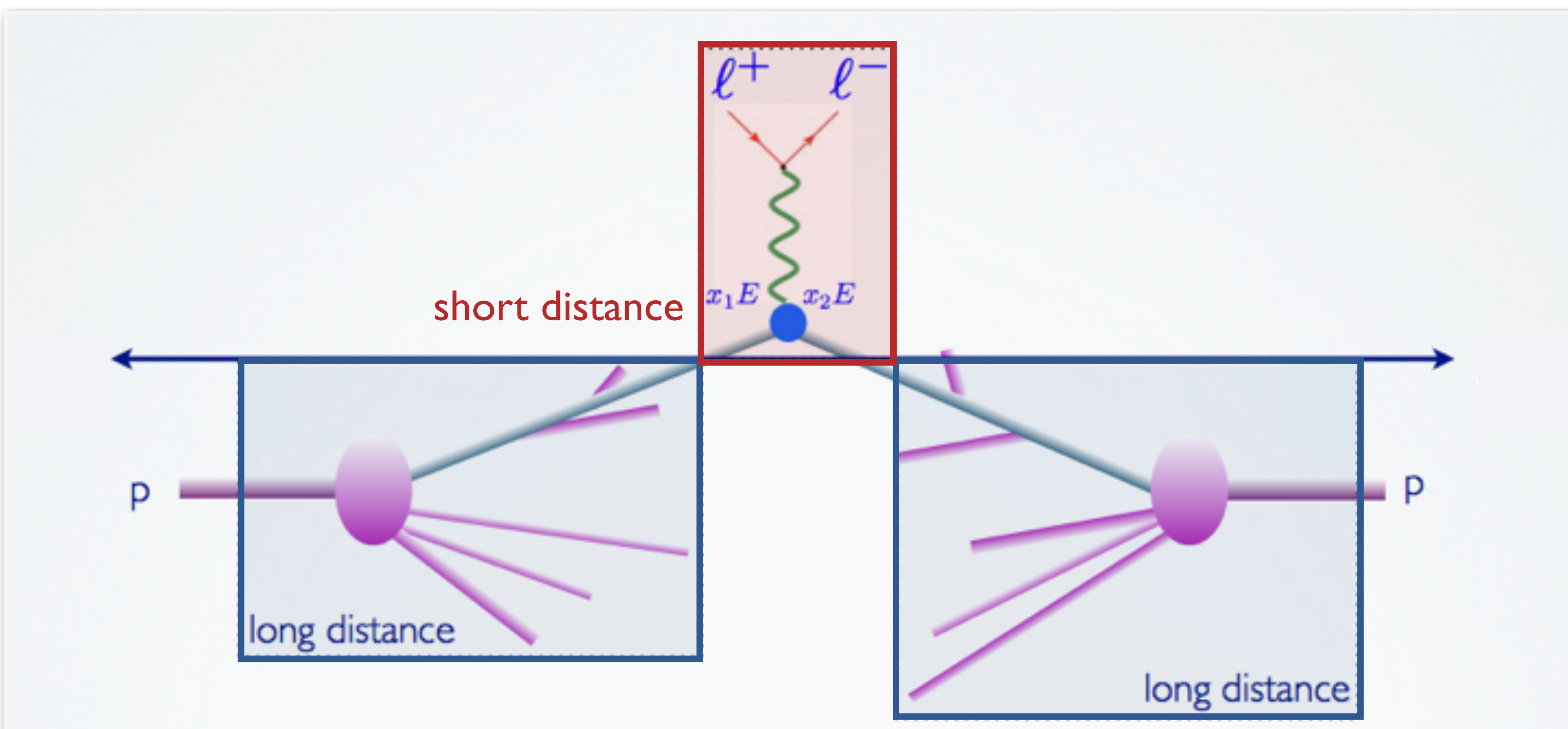




Why PDFs?

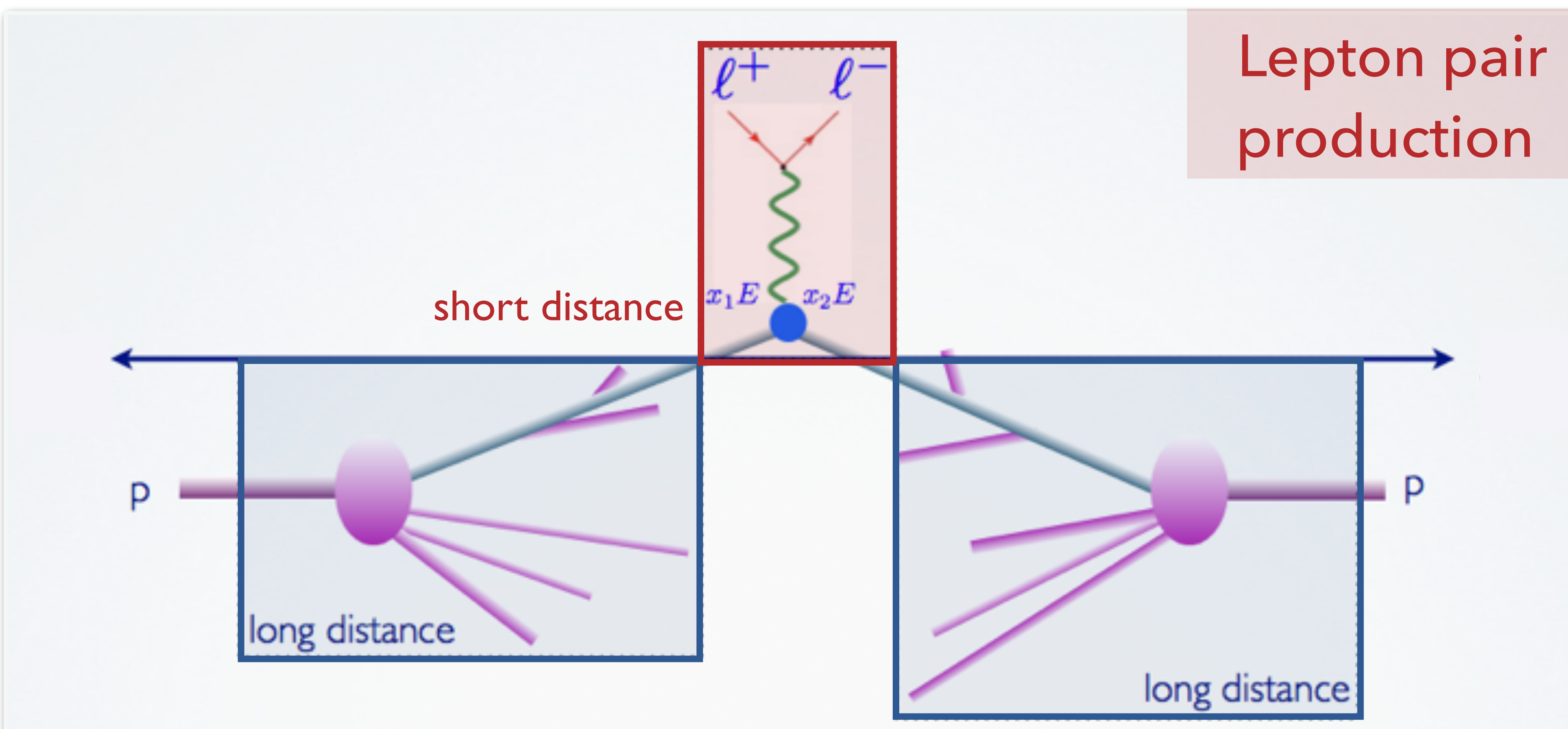
1) PDFs are ubiquitous

$$\frac{d\sigma_H^{pp\to ab}}{dX} = \sum_{i,j=1}^{N_f} f_i(x_1,\mu_F)\, f_j(x_2,\mu_F)\, \frac{d\sigma_H^{ij\to ab}}{dX}(x_1 x_2 S_{\rm had}, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda_{\rm QCD}^{2n}}{S_{\rm had}^n}\right)$$



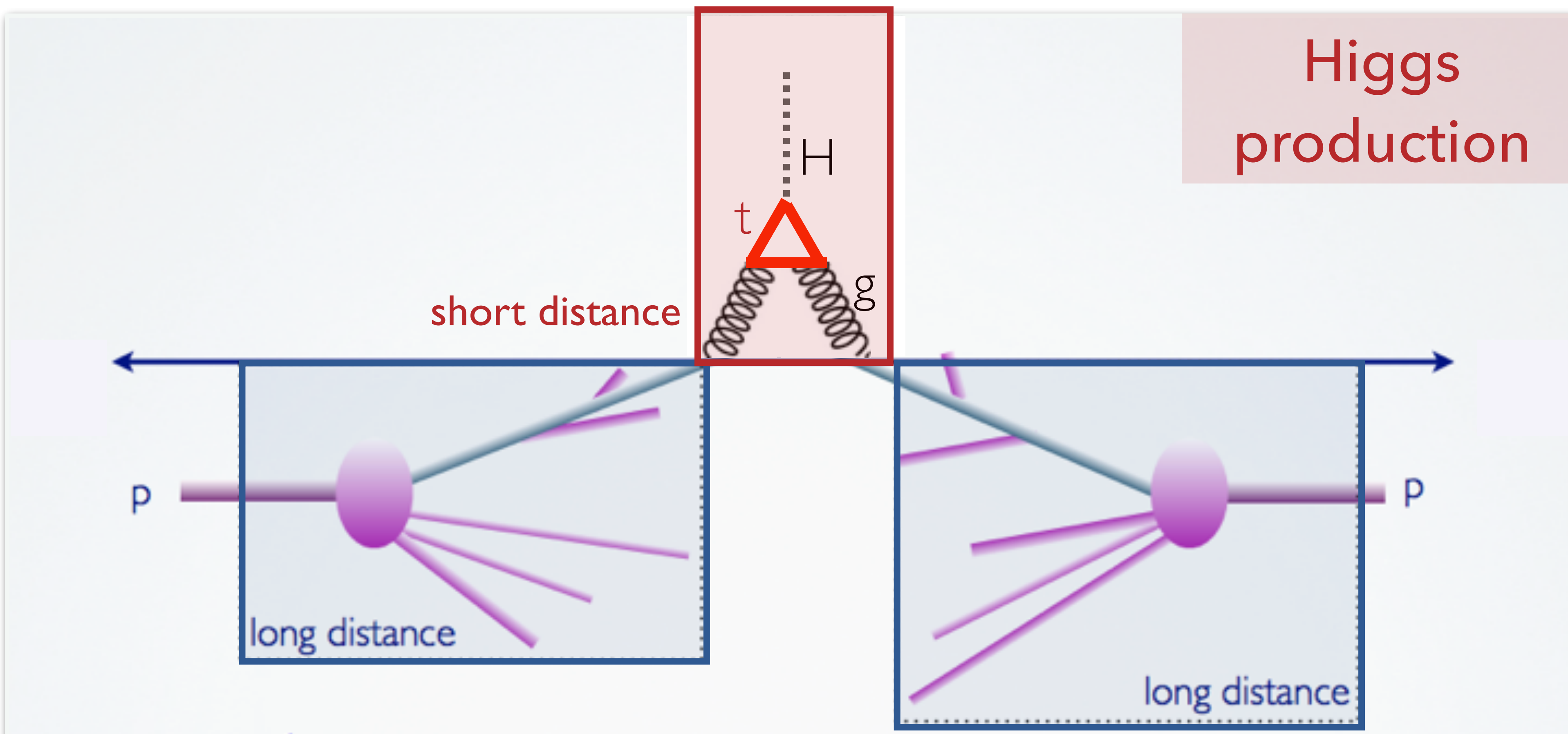
1) PDFs are ubiquitous

$$\frac{d\sigma_H^{pp\to ab}}{dX} = \sum_{i,j=1}^{N_f} f_i(x_1,\mu_F)\, f_j(x_2,\mu_F)\, \frac{d\sigma_H^{ij\to ab}}{dX}(x_1 x_2 S_{\rm had}, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda_{\rm QCD}^{2n}}{S_{\rm had}^n}\right)$$



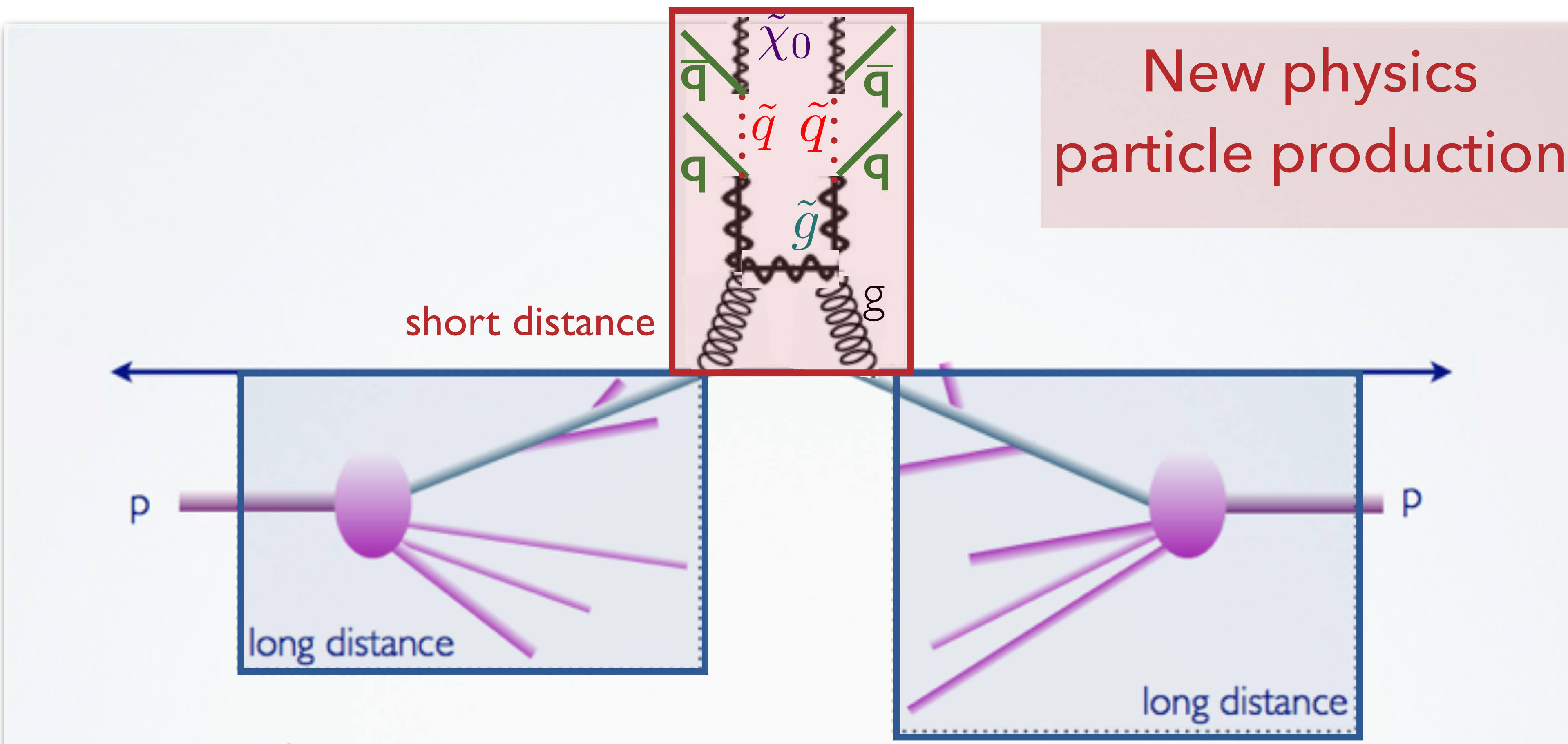
1) PDFs are ubiquitous

$$\frac{d\sigma_H^{pp\to ab}}{dX} = \sum_{i,j=1}^{N_f} f_i(x_1,\mu_F)\, f_j(x_2,\mu_F)\, \frac{d\sigma_H^{ij\to ab}}{dX}(x_1 x_2 S_{\rm had}, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda_{\rm QCD}^{2n}}{S_{\rm had}^{n}}\right)$$

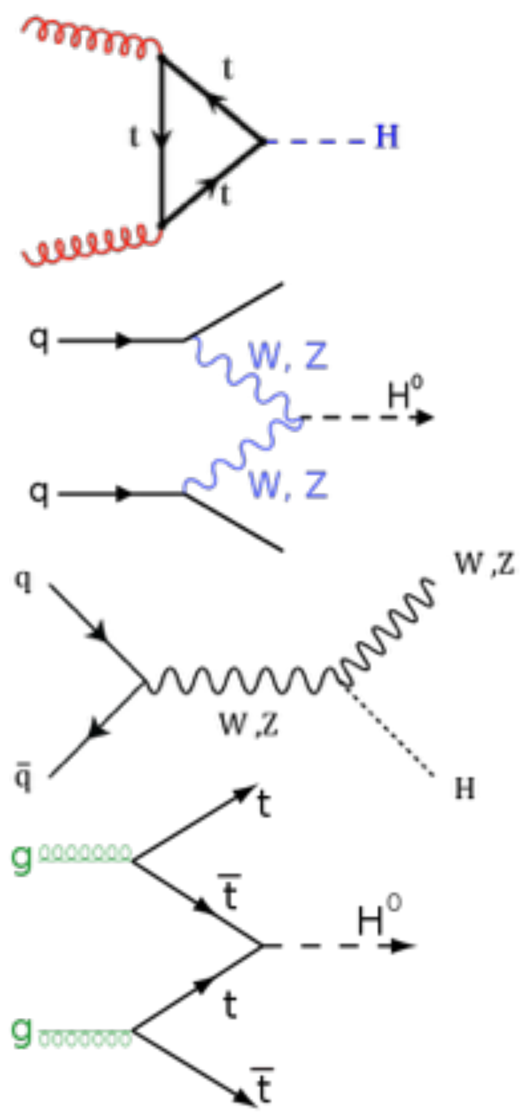


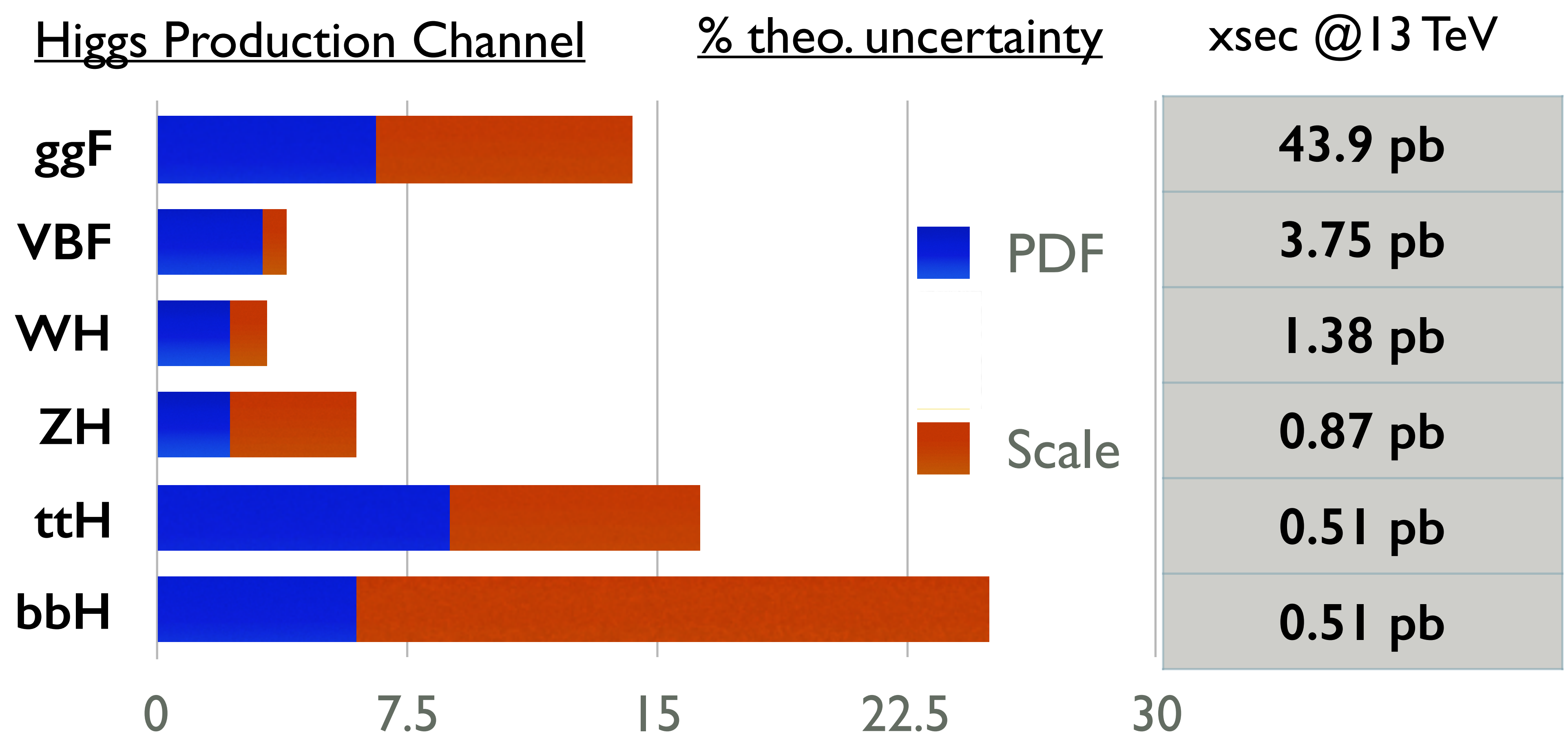
1) PDFs are ubiquitous

$$\frac{d\sigma_H^{pp\to ab}}{dX} = \sum_{i,j=1}^{N_f} f_i(x_1,\mu_F)\, f_j(x_2,\mu_F)\, \frac{d\sigma_H^{ij\to ab}}{dX}(x_1 x_2 S_{\rm had}, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda_{\rm QCD}^{2n}}{S_{\rm had}^n}\right)$$



2) The role of PDF uncertainty



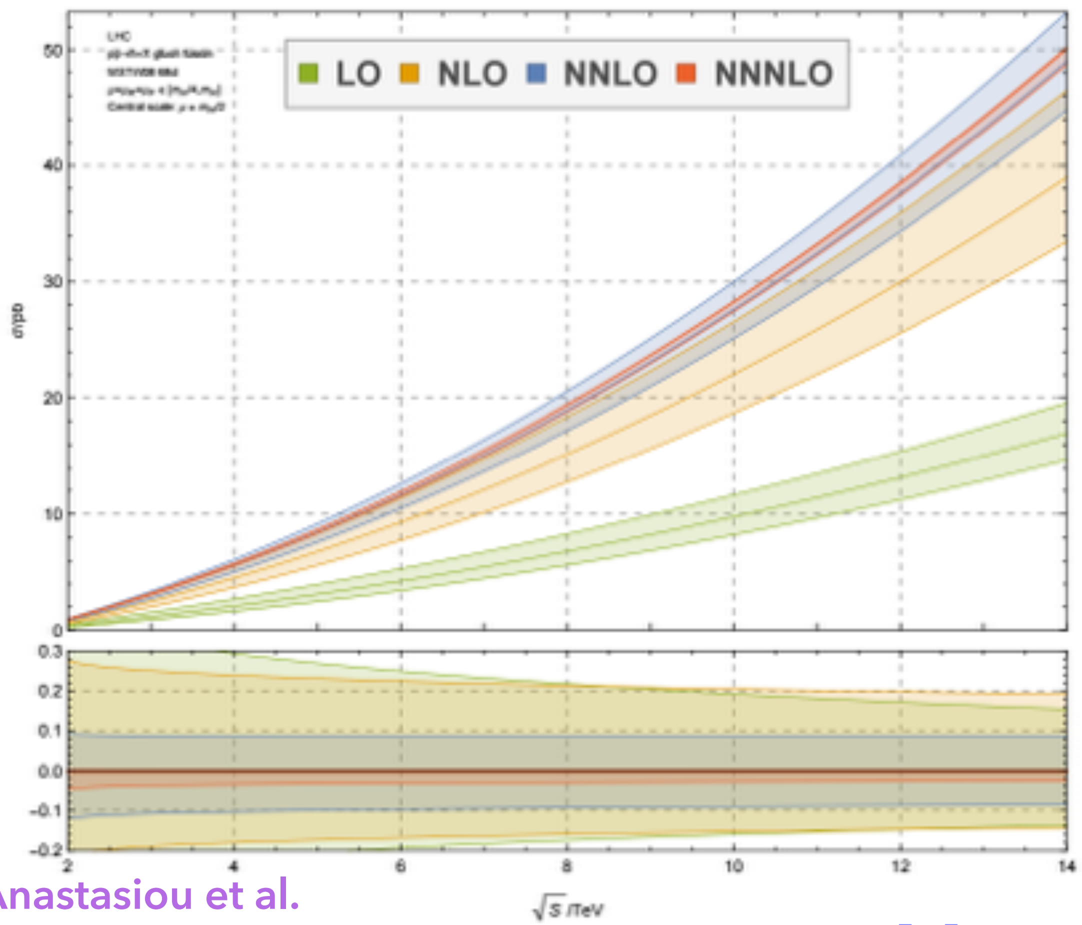


Pre-YR4 numbers from HXSWG Wiki for mH = 125 GeV

PDF uncertainties are a limiting factor in the accuracy of theoretical predictions, both within **SM** and **beyond**

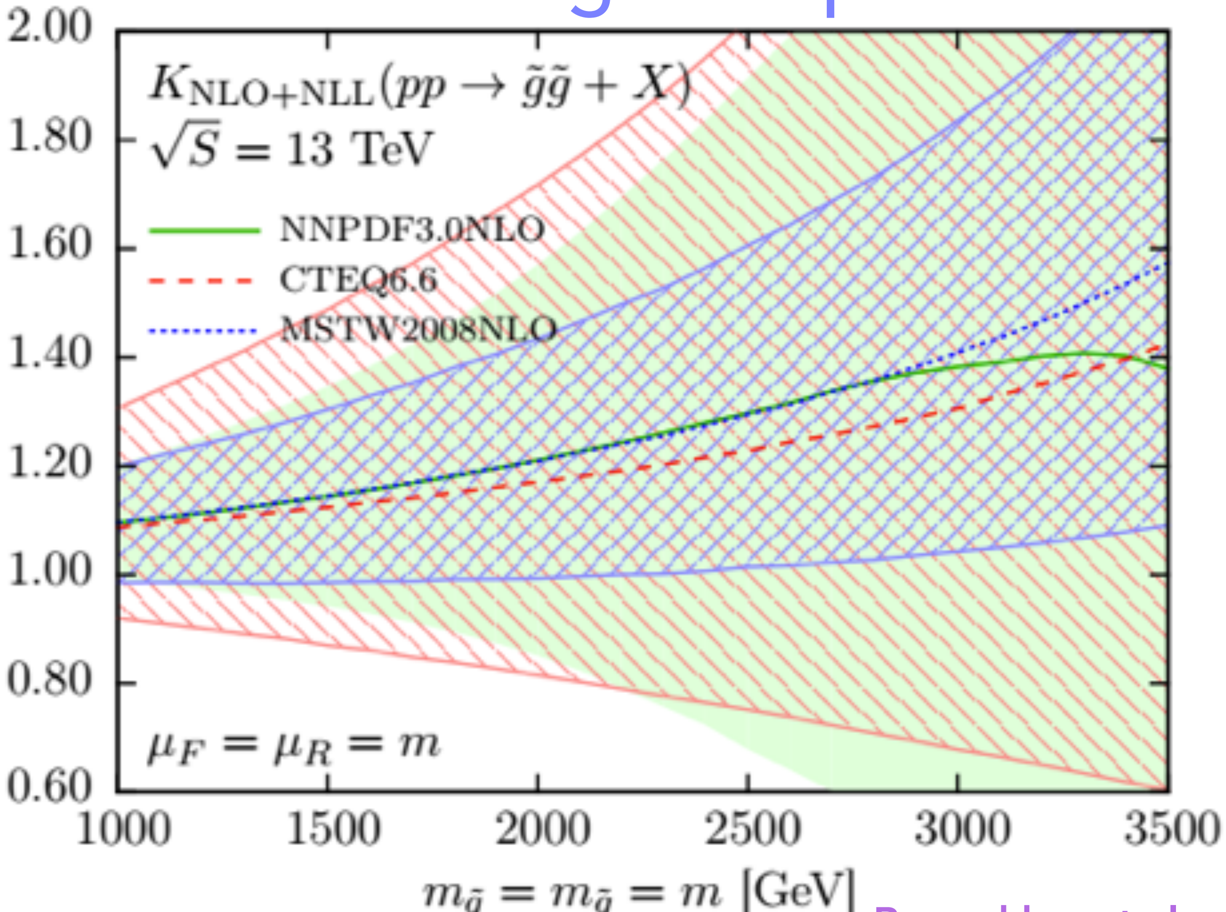
2) The role of PDF uncertainty

ggF @ NNNLO



Anastasiou et al.
PRL 114(2015) 212001

gluino production



Beenakker et al.
EPJC76 (2016)2, 53

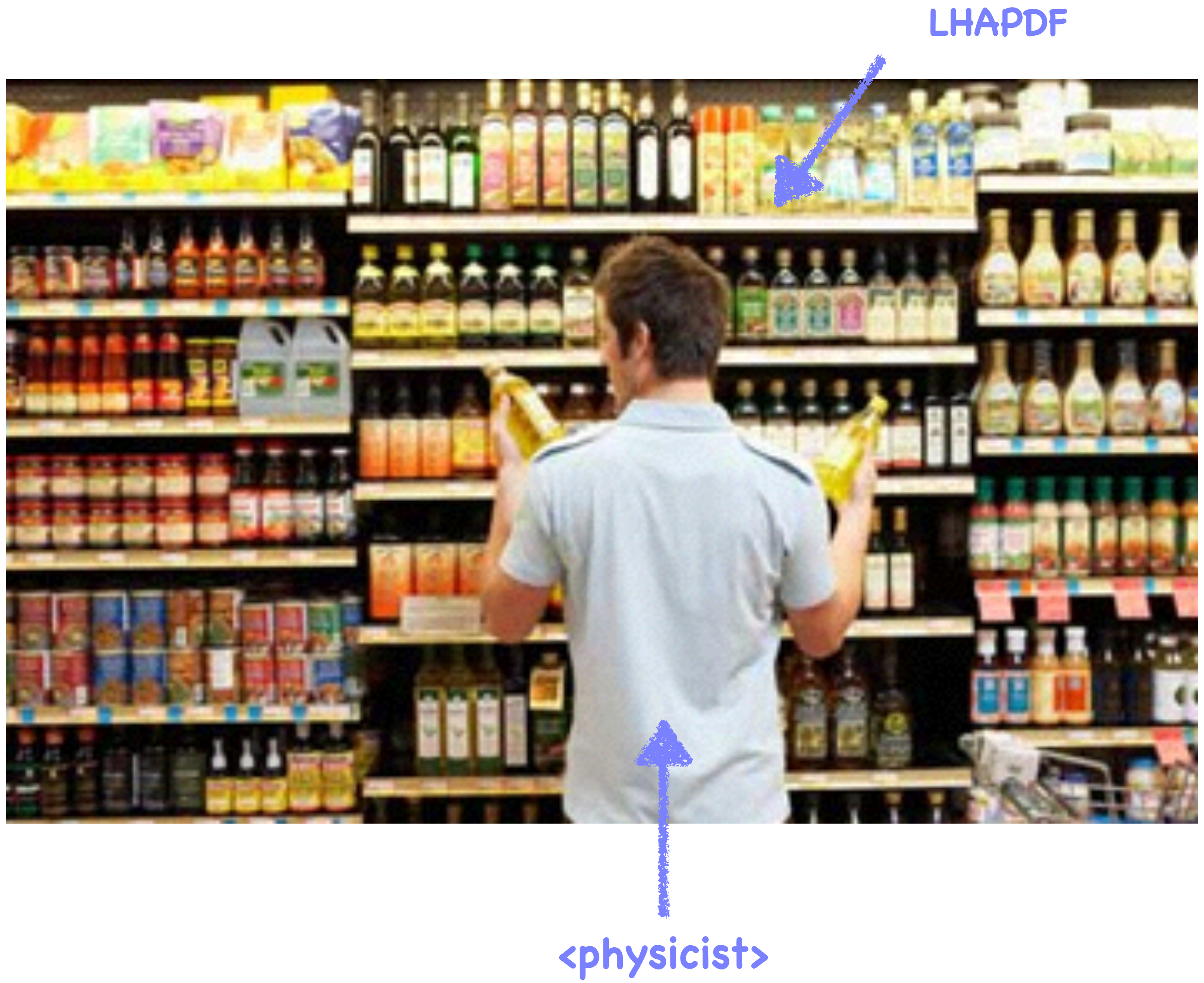
Mw determination

ΔM_W [MeV]	present	CDF	D0	combined	LHC		
$\mathcal{L}$[fb]	7.6	10	10	20	20 (8 TeV)	300	3000
PDF	10	5	5	5	10	5	3
QED rad.	4	4	3	3	4	3	2
$p_T(W)$ model	2	2	2	2	2	1	1
other systematics	9	4	11	4	10	5	3
W statistics	9	6	8	5	1	0.2	0
Total	16	10	15	9	15	8	5

D. Wackeroth's
talk at KITP

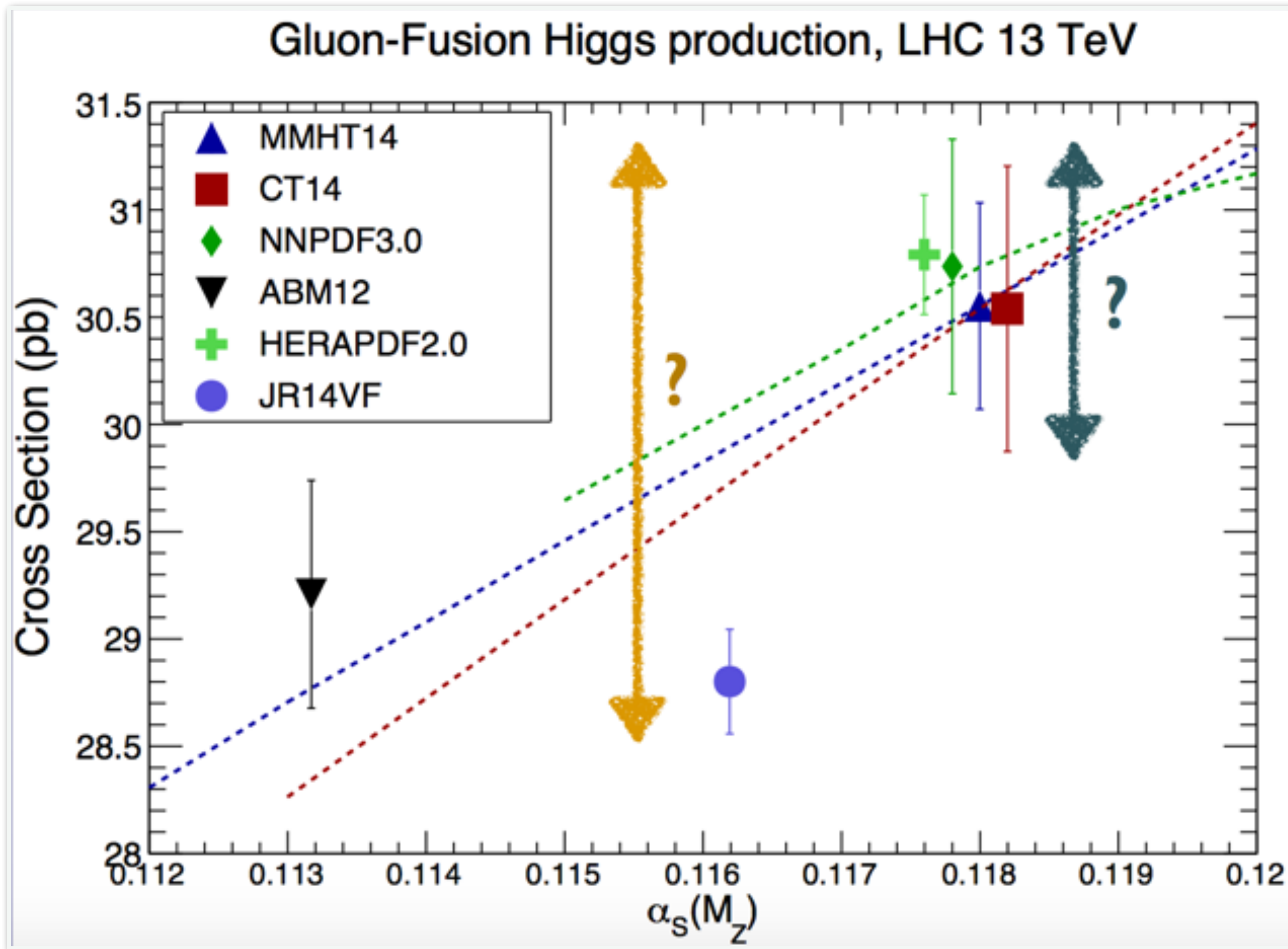
3) The choice of PDFs matters

- What does PDF uncertainty include? How reliable it is?
- How do we interpret the difference predictions using different PDF sets?
- Shall we just pick a set out of the PDFs "supermarket" shelf or take the envelope of ALL predictions?



3) The choice of PDFs matters

- What does PDF uncertainty include? How reliable it is?
- How do we interpret the difference predictions using different PDF sets?
- Shall we just pick a set out of the PDFs "supermarket" shelf or take the envelope of ALL predictions?

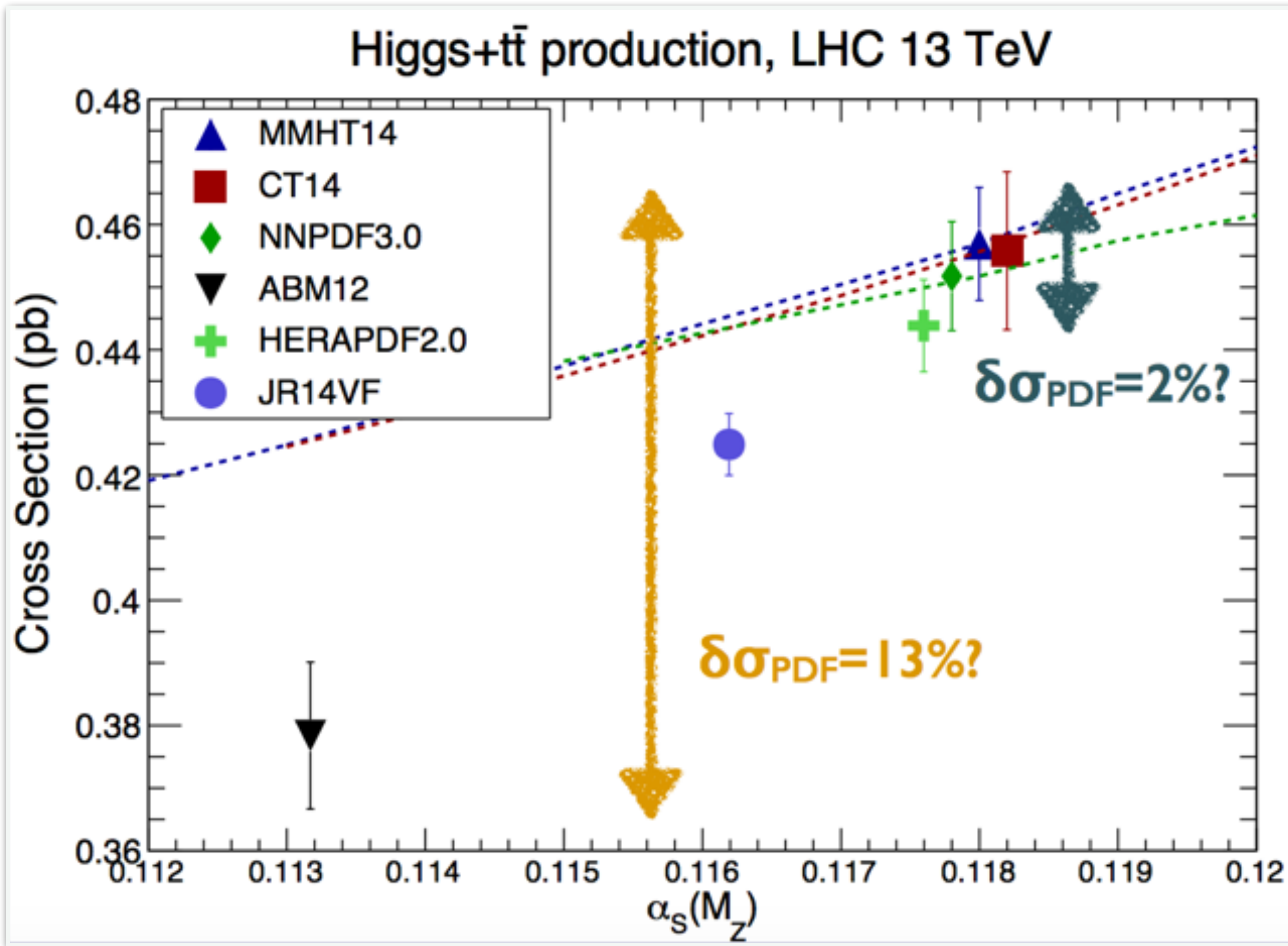


$\delta\sigma_{PDF}$ = 2% $\delta\sigma_{PDF}$ = 5%

J. Rojo's talk at DIS2016

3) The choice of PDFs matters

- What does PDF uncertainty include? How reliable it is?
- How do we interpret the difference predictions using different PDF sets?
- Shall we just pick a set out of the PDFs "supermarket" shelf or take the envelope of ALL predictions?



J. Rojo's talk at DIS2016

What PDFs are

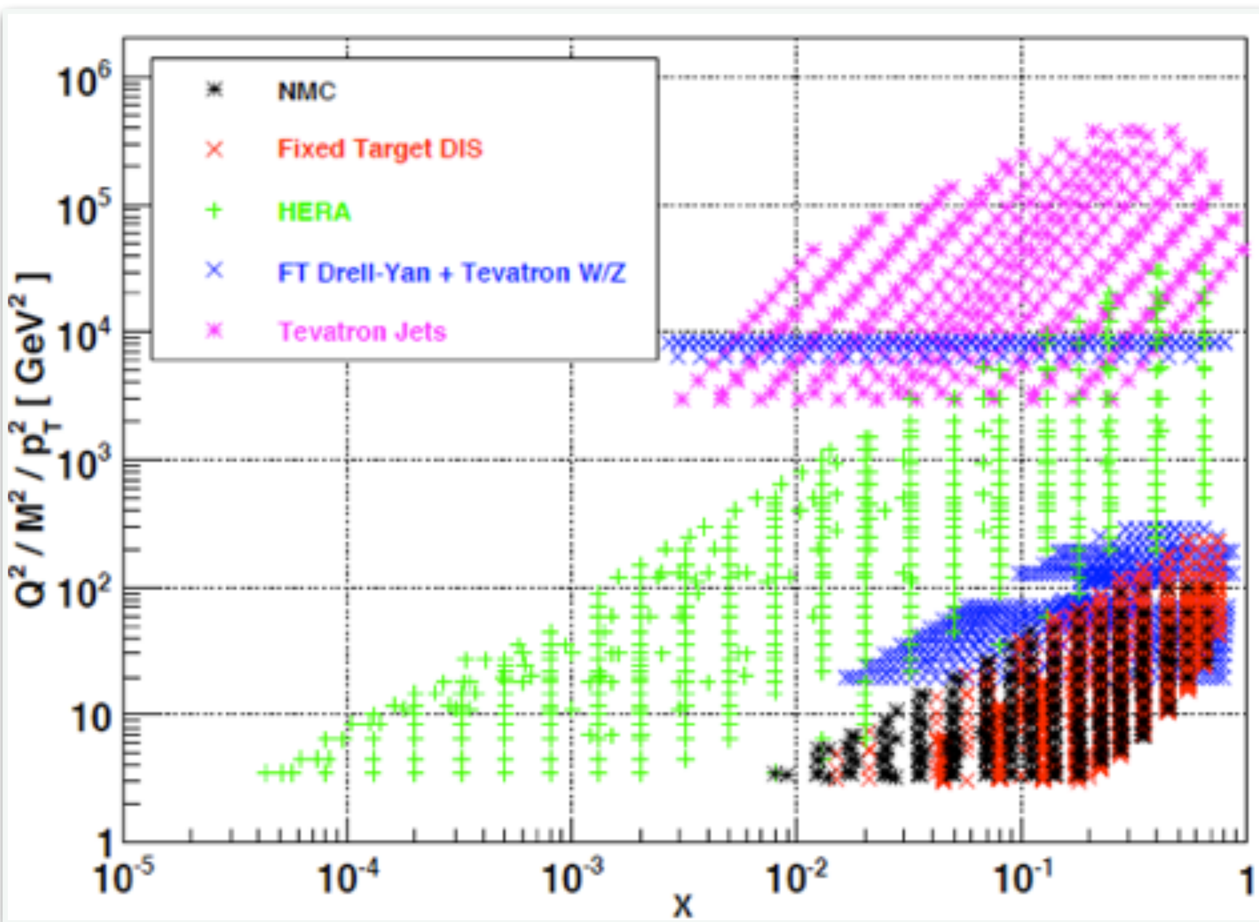
Collinear Factorisation Theorem

$$\frac{d\sigma_H^{pp\to ab}}{dX} = \sum_{i,j=1}^{N_f} f_i(x_1,\mu_F)\, f_j(x_2,\mu_F)\, \frac{d\sigma_H^{ij\to ab}}{dX}(x_1 x_2 S_{\rm had}, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda_{\rm QCD}^{2n}}{S_{\rm had}^n}\right)$$

$$\mu^2 \frac{\partial\, f(x,\mu^2)}{\partial\, \mu^2} = \int_z^1 \frac{dz}{z}\frac{\alpha_s}{2\pi} P(z)\, f\left(\frac{x}{z},\mu^2\right)$$

Dokshitzer, Gribov, Lipatov, Altarelli, Parisi renormalization group equations

Q-dependence: pert. theory



x-dependence: from data

LO - Dokshitzer; Gribov, Lipatov; Altarelli, Parisi, 1977

NLO - Floratos,Ross,Sachrajda; Floratos,Lacaze,Kounnas, Gonzalez-Arroyo,Lopez,Yndurain; Curci,Furmanski Petronzio, 1981

NNLO - Moch, Vermaseren, Vogt, 2004

The PDF extraction process

- Choose **experimental data** to fit and include all info on correlations
- **Theory settings**: perturbative order, heavy quark mass scheme, EW corrections, intrinsic heavy quarks, α_s, quark masses value and scheme
- Choose a starting scale Q_0 where pQCD applies
- **Parametrise** independent quarks and gluon distributions at the starting scale
- Solve **DGLAP equations** from initial scale to scales of experimental data and build up observables
- **Fit** PDFs to data
- Provide **error sets** to compute PDF uncertainties

PDF uncertainty

Hidden uncertainty

Parametric versus non-parametric approach

Hessian versus MC approach

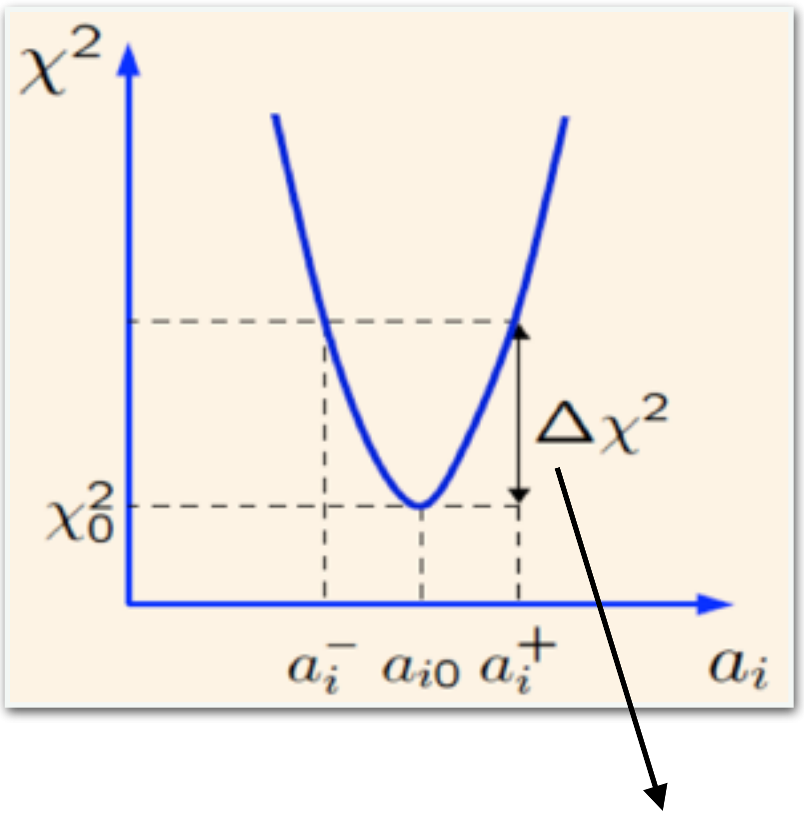
Hessian versus MC

Given a finite number of experimental point want a set of functions with error

$$\langle \mathcal{F}[f_{\{i\}}(x)] \rangle = \int [\mathcal{D}f]\, \mathcal{F}[f_{\{i\}}(x)]\, \mathcal{P}[f_{\{i\}}(x)]$$

Hessian approach: project into a N_{par}-dimensional space of parameters and use linear approximation around the minimum of the χ^2

$$F_0 = F(S_0), \quad \sigma_F = \sqrt{\sum_{i=1}^{N_{\rm par}} [F(S_i) - F(S_0)]^2}$$



Hessian versus MC

Given a finite number of experimental point want a set of functions with error

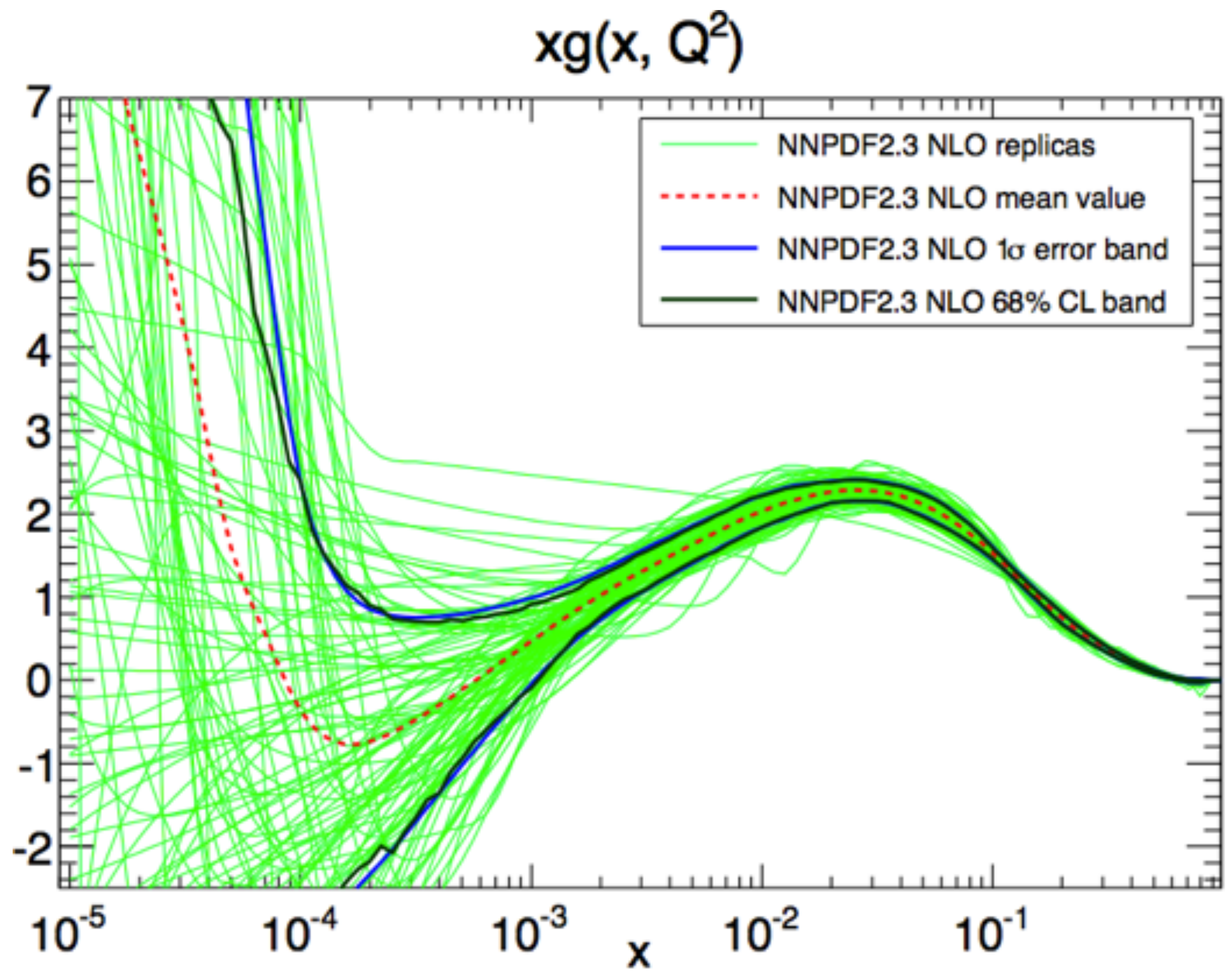
$$\langle \mathcal{F}[f_{\{i\}}(x)] \rangle = \int [\mathcal{D}f]\, \mathcal{F}[f_{\{i\}}(x)]\, \mathcal{P}[f_{\{i\}}(x)]$$

Monta Carlo (NNPDF) approach:

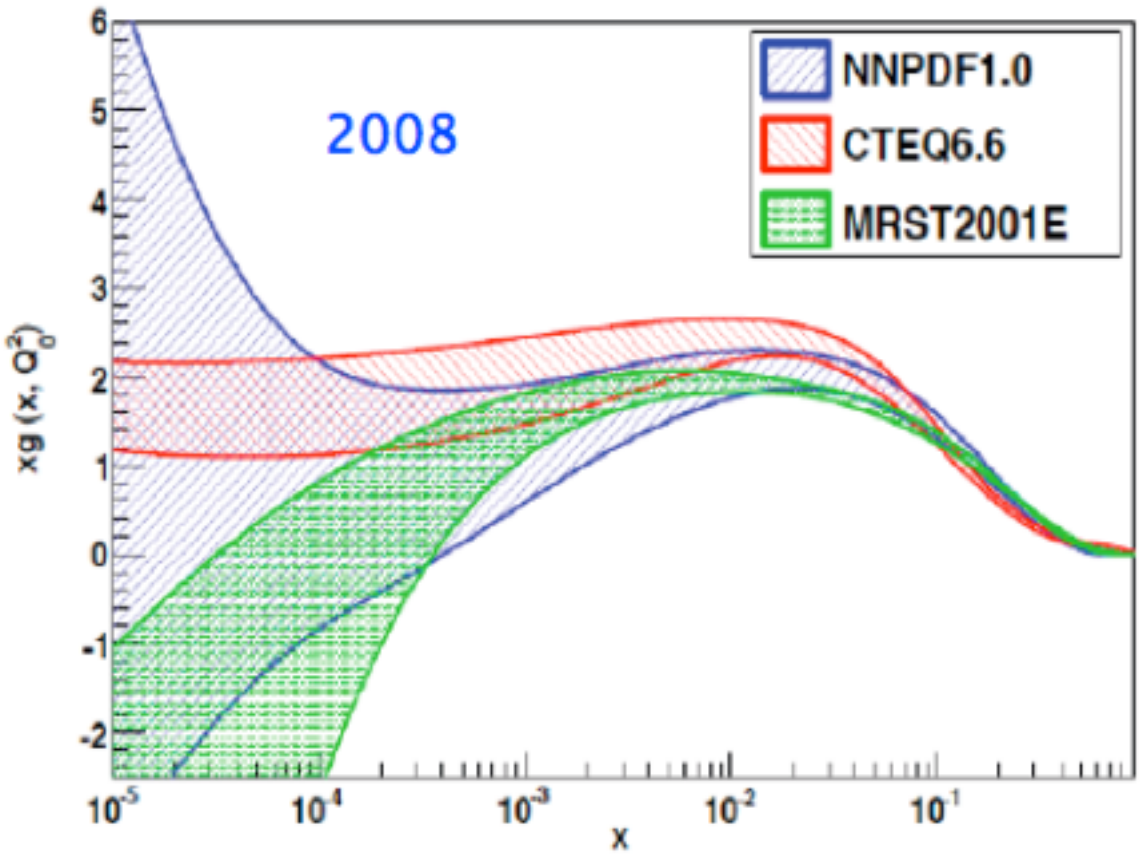
Sampling the probability measure in PDF space by projecting down from probability density in data space

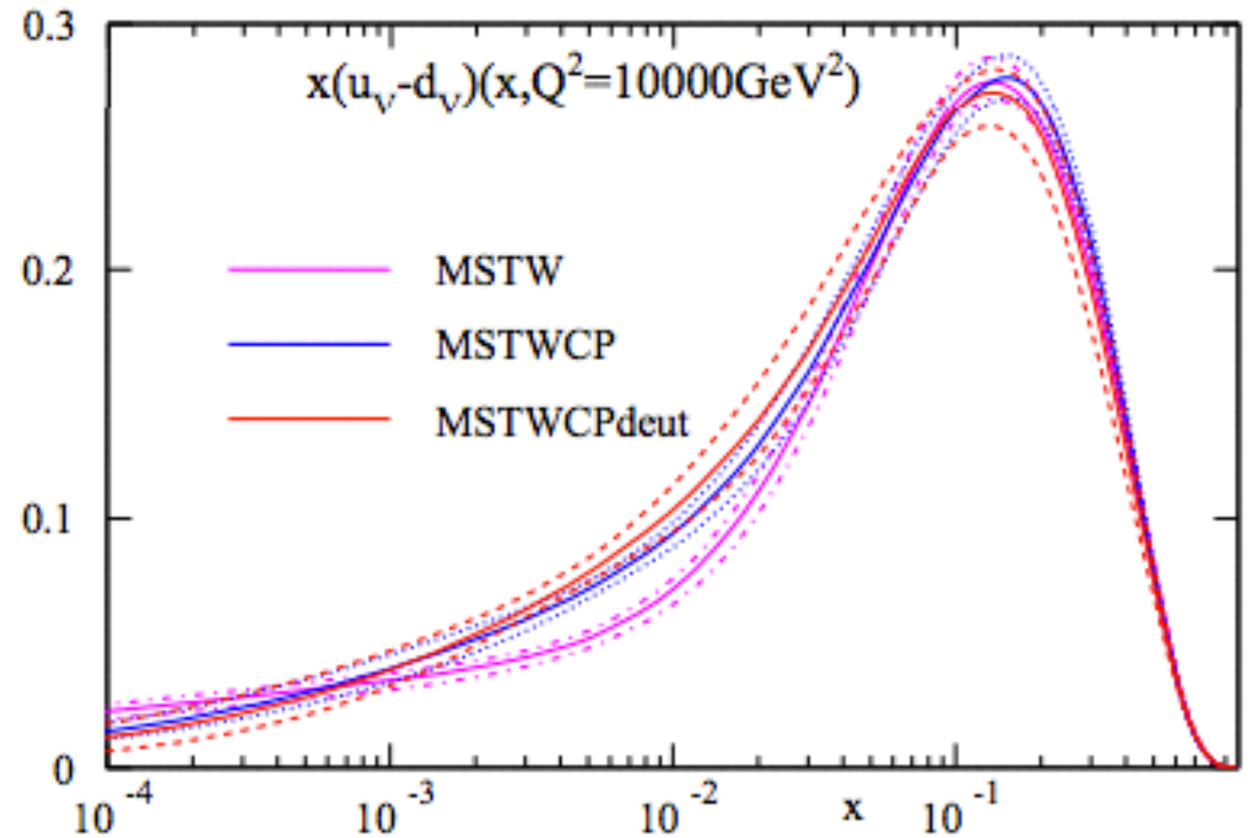
$$F_0 = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} F(S^k)$$

$$\sigma_F = \sqrt{\frac{1}{N_{\rm rep}-1} \sum_{k=1}^{N_{\rm rep}} [F(S^k) - F_0]^2}$$



Parametric vs non-parametric



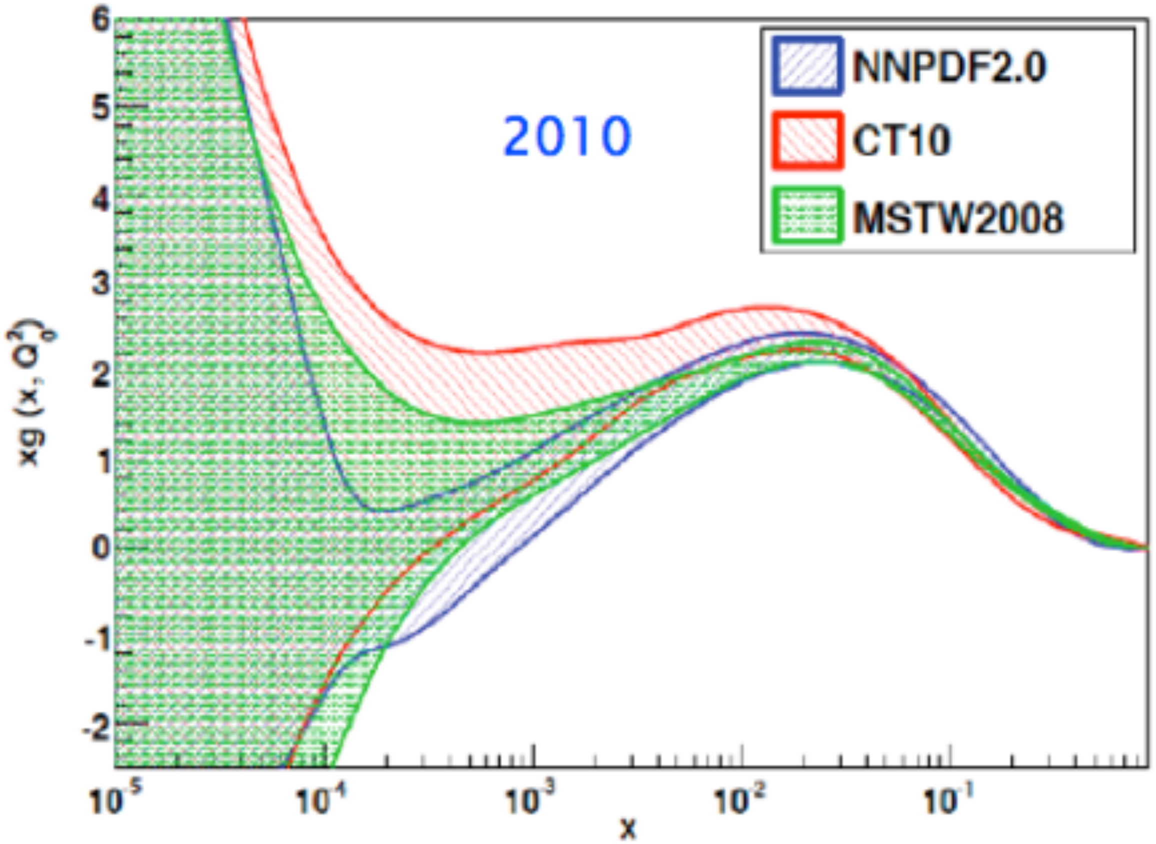


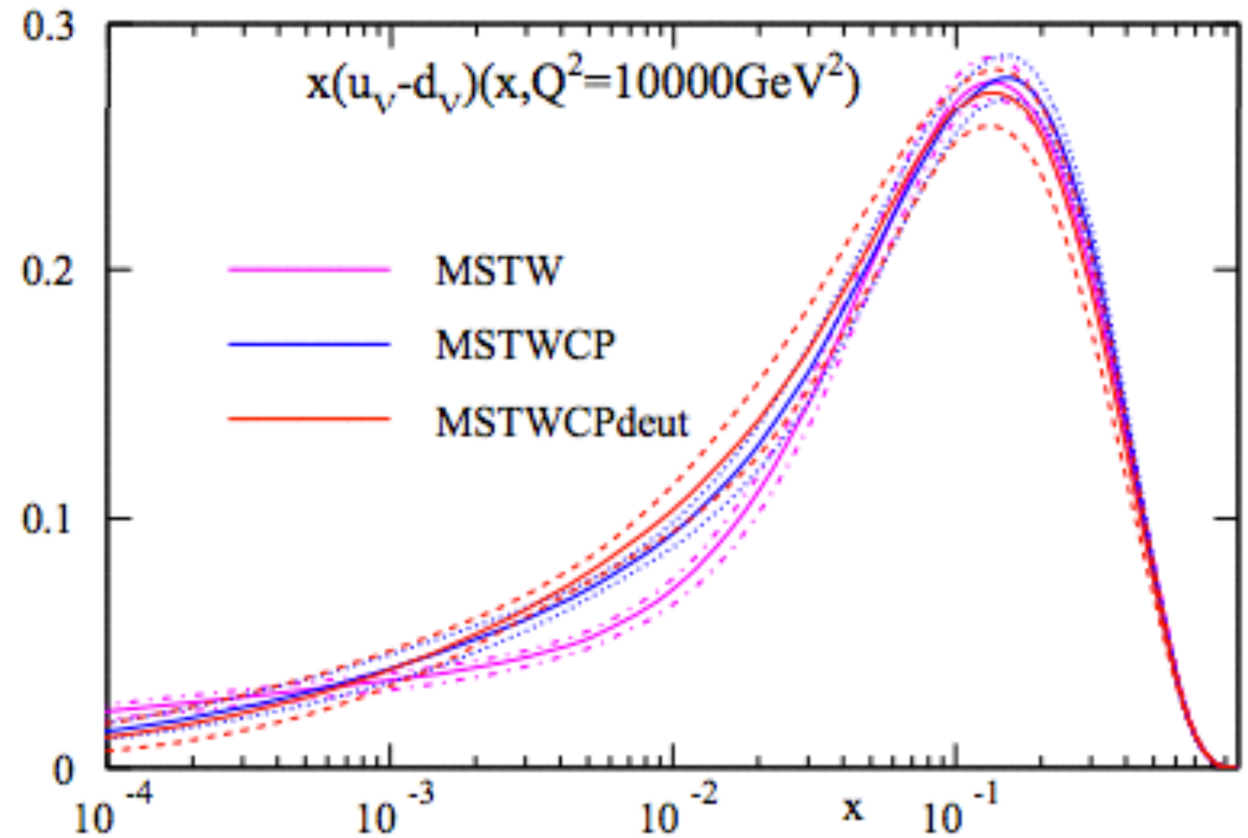
Martin et al
EPJC73 (2013) 2, 2318

- What is the error associated to a given choice of functional form?
  If too rigid PDFs may not adapt to new data or present small errors where data do not constrain PDFs

- Neural Networks: all independent PDFs are associated to an unbiased and flexible parametrization: O(300) parameters versus O(20) in polynomial parametrisation

Progress and convergence



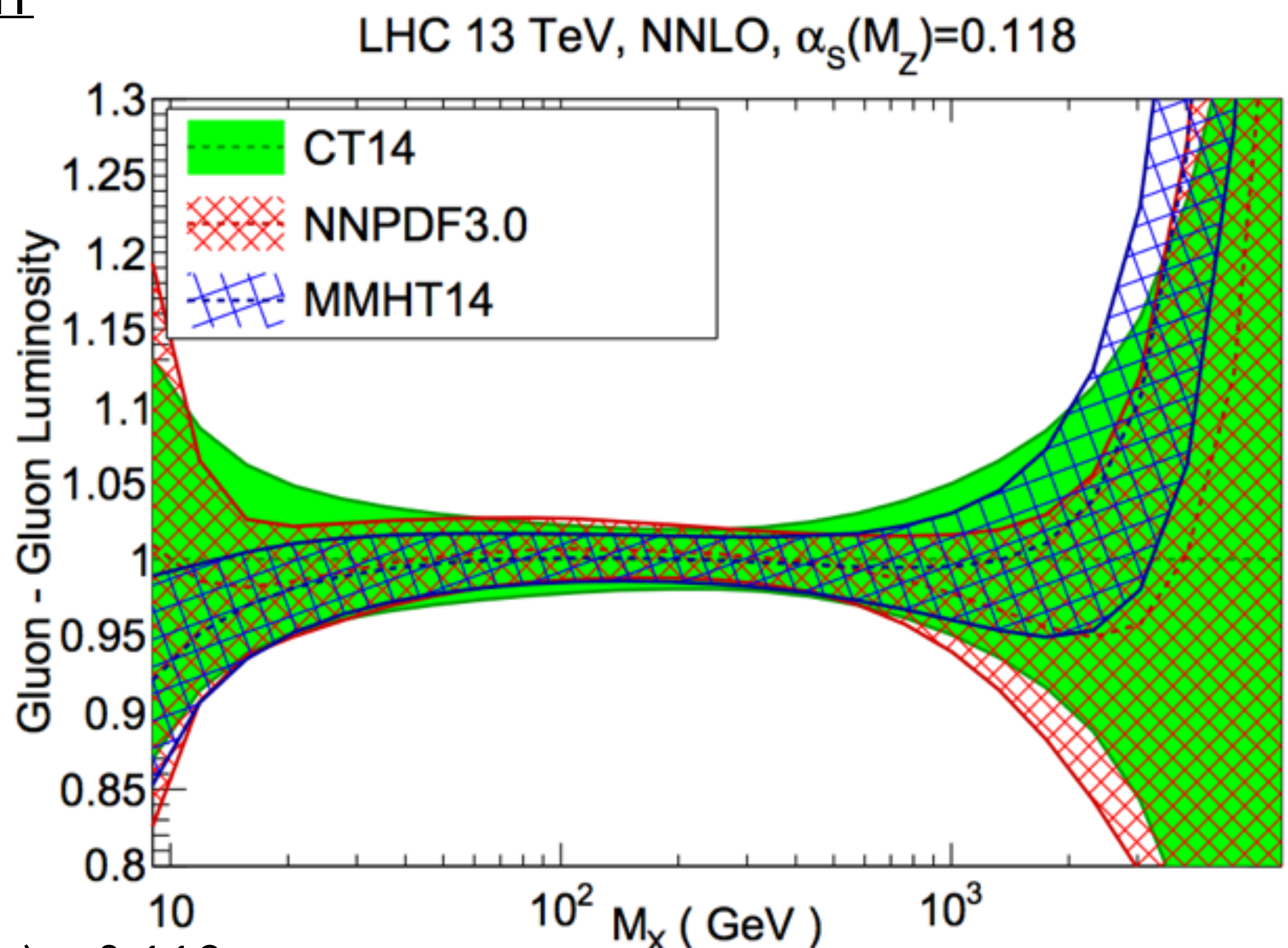


Martin et al
EPJC73 (2013) 2, 2318

- Parametric approach: lot of progress in achieving a minimally biased parametrisation form
- Non-parametric approach: methodology tested via closure test studies
- Hessian vs Monte Carlo: now possible to go from Hessian to MC and vice-versa and test deviations from Gaussianity

State of the art

NNPDF3.0 / CT14 / MMHT



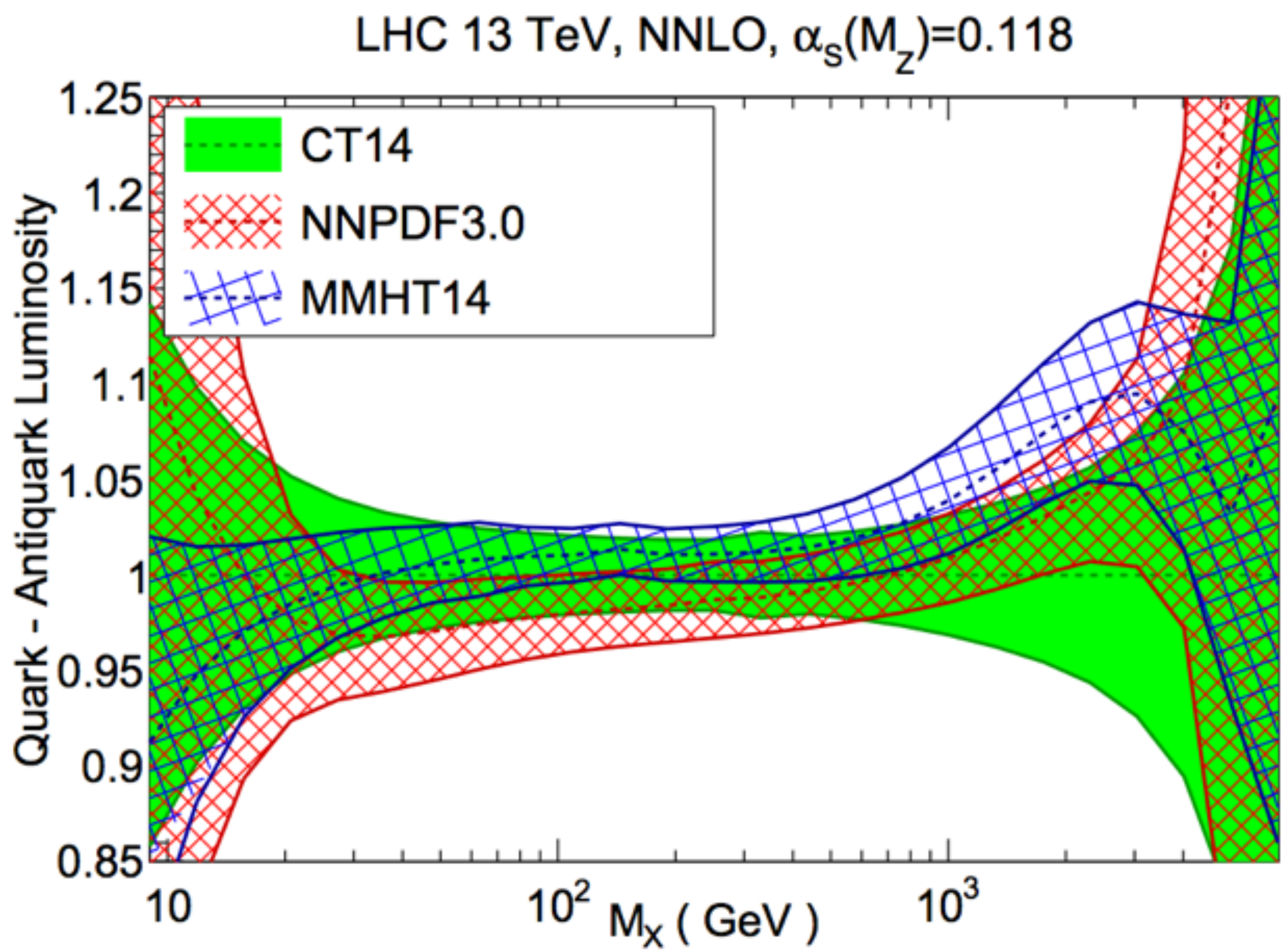
- common value of $\alpha_S(M_Z) = 0.118$
- comparable GM-VFN schemes for inclusion of HQ masses
- global sets: inclusion of O(4000) experimental data
- extensive benchmarking

J. Butterworth et al
J.Phys. G43 (2016) 023001

State of the art

NNPDF3.0 / CT14 / MMHT

See Pavel's talk



- common value of $\alpha_S(M_Z) = 0.118$
- comparable GM-VFN schemes for inclusion of HQ masses
- global sets: inclusion of O(4000) experimental data
- extensive benchmarking

J. Butterworth et al
J.Phys. G43 (2016) 023001

Current frontiers

Past frontiers

2016 First PDF set with fitted charm

2015 First PDF set with threshold resummation

2014 First PDF set with methodology validated with closure test

2013 First PDF set with fitted photon PDF

First PDF set with LHC data

2012 Reweighting PDFs

2011 Heavy quark mass effects

Determination of α_S from PDF fit

2010 First NNPDF global set

2009 Determination of the proton strangeness: solved NuTeV anomaly

2008 First NNPDF set - only DIS data

NNPDF

Present frontiers

THEORY

- pQCD loop revolution - PDF must keep up!
- Large invariant mass & large rapidity - EW and photon-initiated processes become important
- Closer to kinematic boundaries - resummation in PDFs?

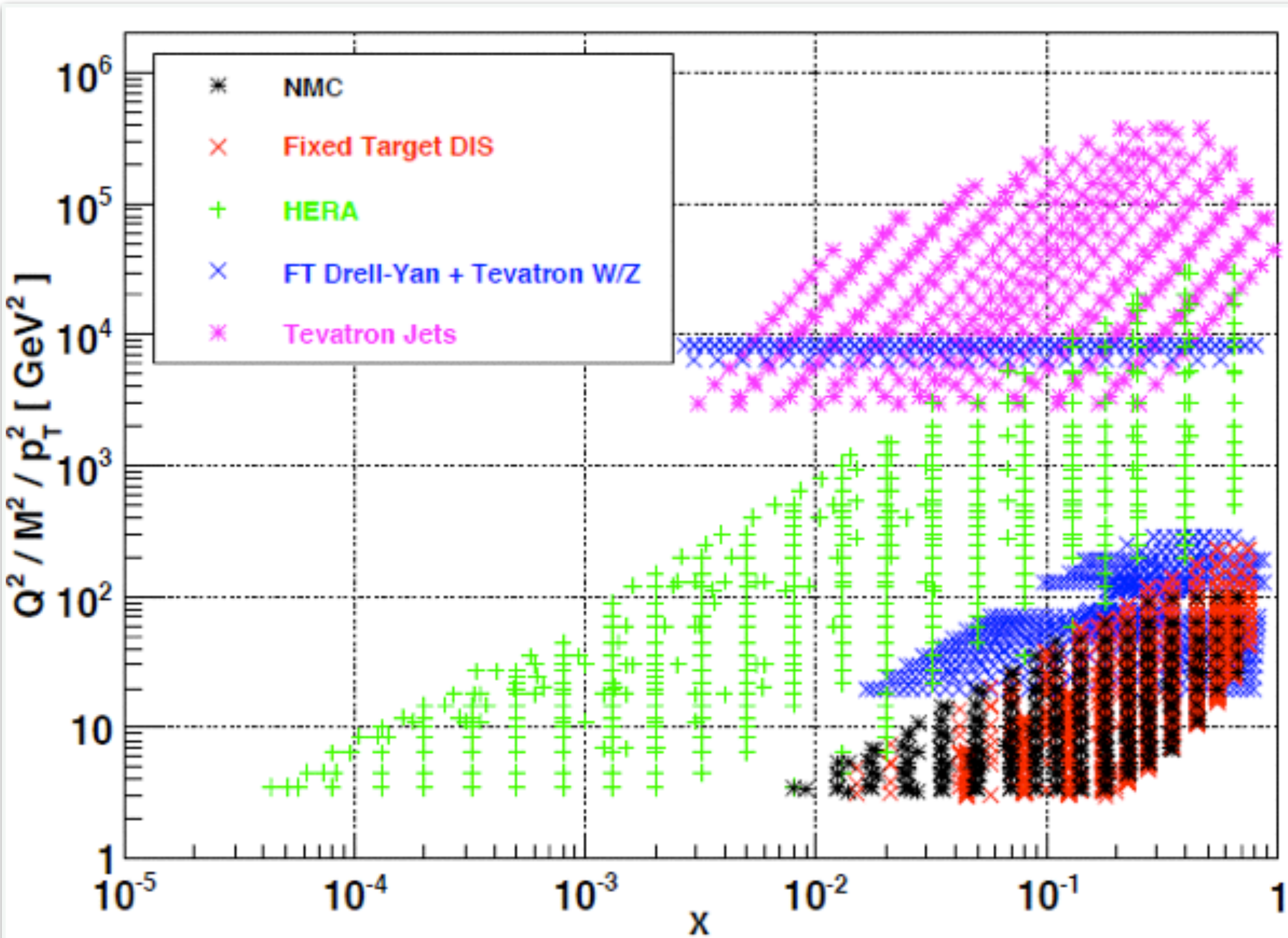
DATA

- Many new accurate LHC data - collider-only fit?
- Prospects for PDF determination at future colliders

METHODOLOGY

- Closure tests to establish methodology
- Combination of different PDF sets
- Inclusion of hidden uncertainties in PDF error bands (especially theory uncertainties)
- How not to absorb new physics in PDFs?

The data (before LHC)



large-x gluon

u/d u~/d~ separation

small/moderate-x gluon and light quarks

u/d separation & strangeness

The LHC data

GLUON

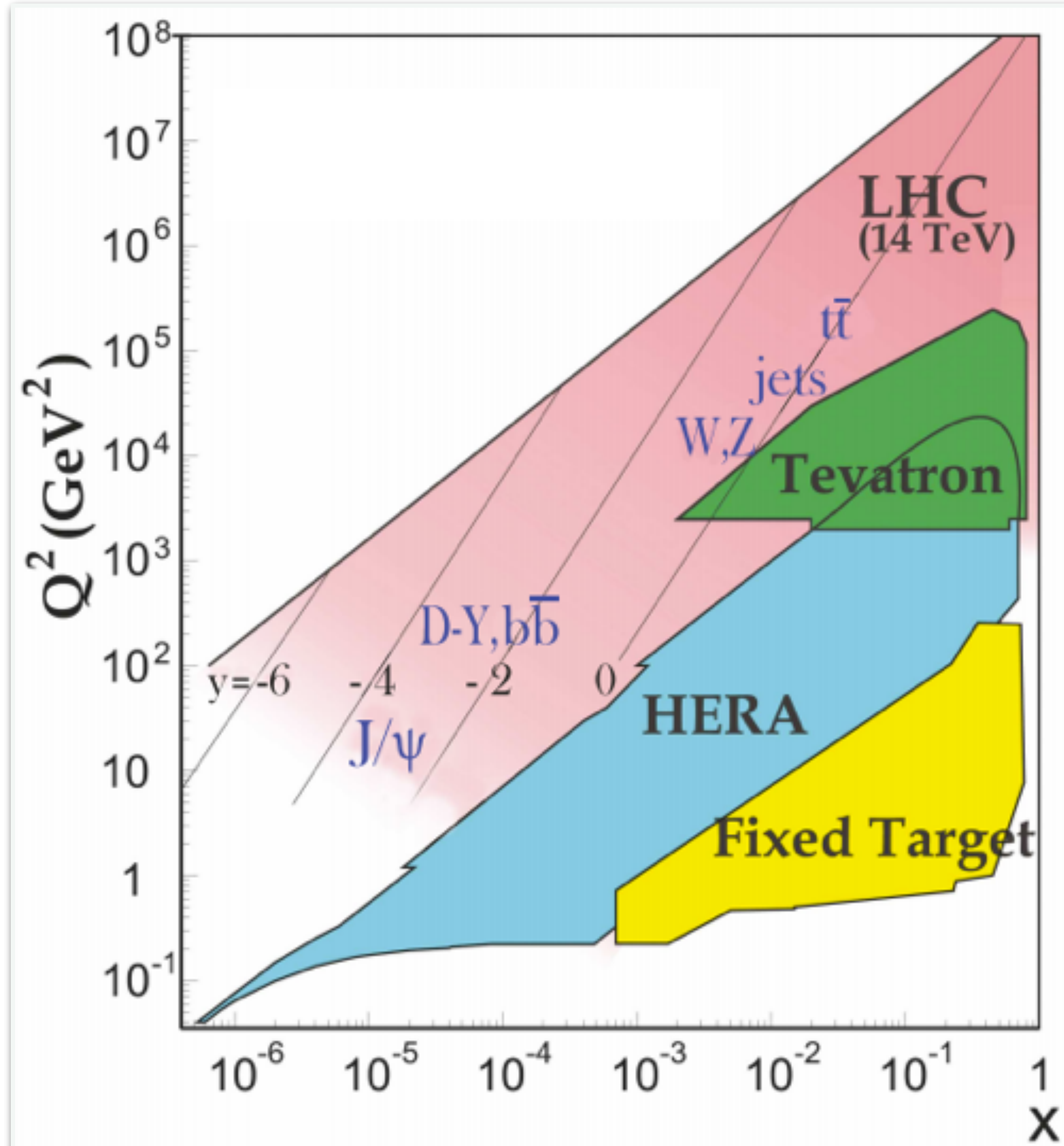
Inclusive jets and dijets
(medium/large x)
Isolated photon and γ+jets
(medium/large x)
Top pair production **(large x)**
High p_T V(+jets) distribution
(small/medium x)

QUARKS

High p_T W(+jets) ratios
(medium/large x)
W and Z production
(medium x)
Low and high mass Drell-Yan
(small and large x)
Wc **(strangeness at medium x)**

PHOTON

Low and high mass Drell-Yan
WW production



Effect of LHC data on PDFs

NNPDF3.0

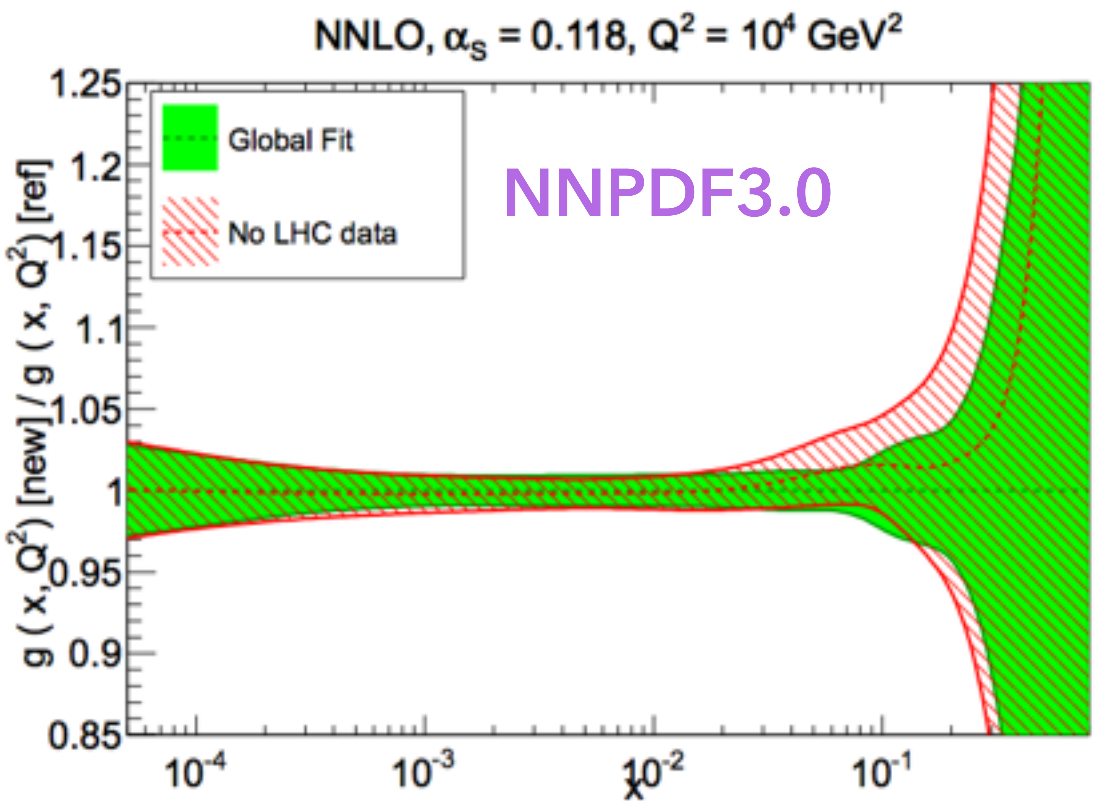
ATLAS jets 2.76 TeV and 7 TeV	gluon large x
ATLAS high-mass DY at 7 TeV	q/q~ separation
ATLAS W pT data at 7 TeV	g and q at moderate x
CMS (Y,M) double diff distributions 7 TeV	flavour separation
CMS jets at 7 TeV	gluon large x
CMS muon charge asymmetry at 7 TeV	quark separation
CMS W+c at 7 TeV	strangeness
LHCb Z rapidity distribution at 7 TeV	small/large x quarks
ATLAS+CMS tt total xsec at 7/8 TeV	gluon large x

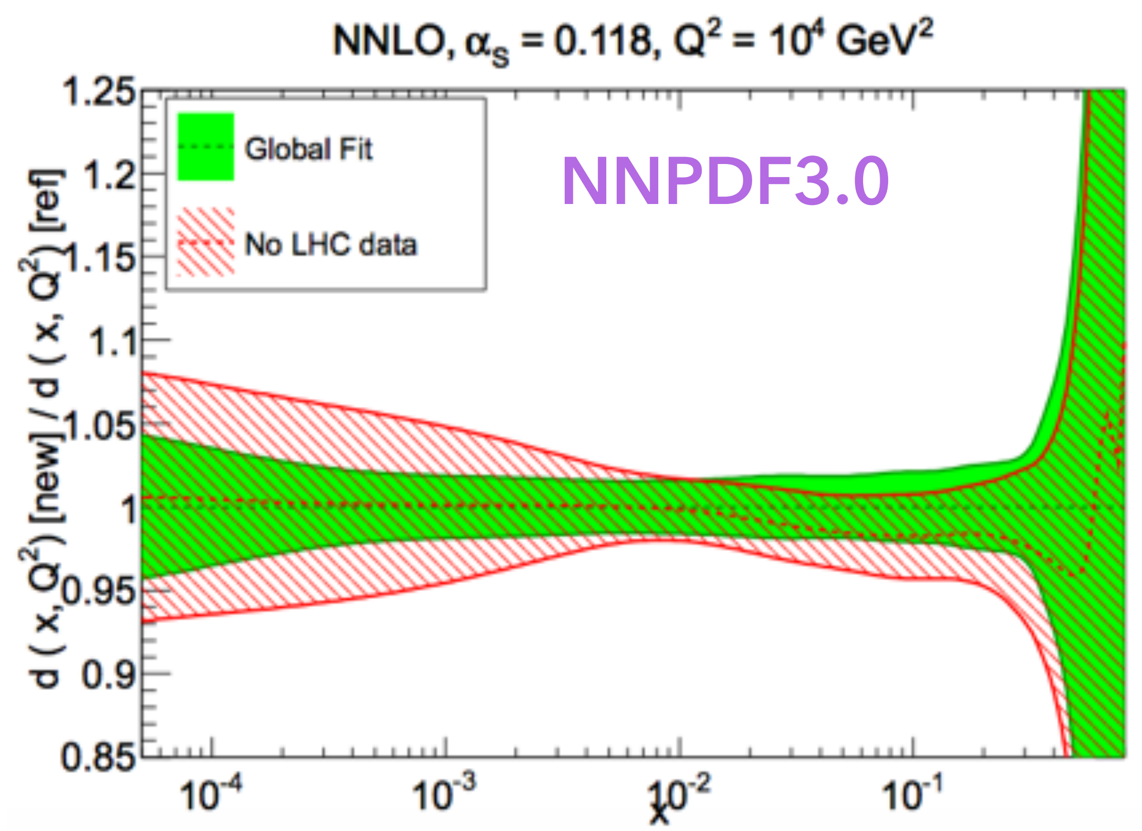
Effect of LHC data on PDFs

NNPDF3.1

ATLAS jets 2.76 TeV and 7 TeV **+ 2011 data 7 TeV**	gluon large x
ATLAS high-mass DY at 7 TeV **+ low mass**	q/q~ separation
ATLAS W pT data at 7 TeV **+ ATLAS & CMS double diff Z pT**	g and q at moderate x
CMS (Y,M) double diff distributions 7 TeV **+ 8 TeV**	flavour separation
CMS jets at 7 TeV **+ 2.76 and 8 TeV jet data**	gluon large x
CMS muon charge asymmetry at 7 TeV **+ 8 TeV**	quark separation
CMS W+c at 7 TeV	strangeness
LHCb Z rapidity distribution at 7 TeV **+ 8 TeV (legacy data)**	small/large x quarks
ATLAS+CMS tt total xsec at 7/8 TeV **+ differ. distributions**	gluon large x
D0 legacy W asymmetry data	q/q~ separation

Effect of LHC data on PDFs





Ball et al.
JHEP 1504 (2015) 040

- Data give increasingly stronger constraints in known and less-known kinematic regions => PDF experimental uncertainties reduced
- Are we keeping up with theory settings in PDF fits?

The NNLO frontier

- NNLO calculations are essential to reduce theoretical uncertainties in PDF analyses

- Stunning progress has been made on some key processes for PDF determination

- Great progress also in tools to interface NLO codes to PDF fitting code

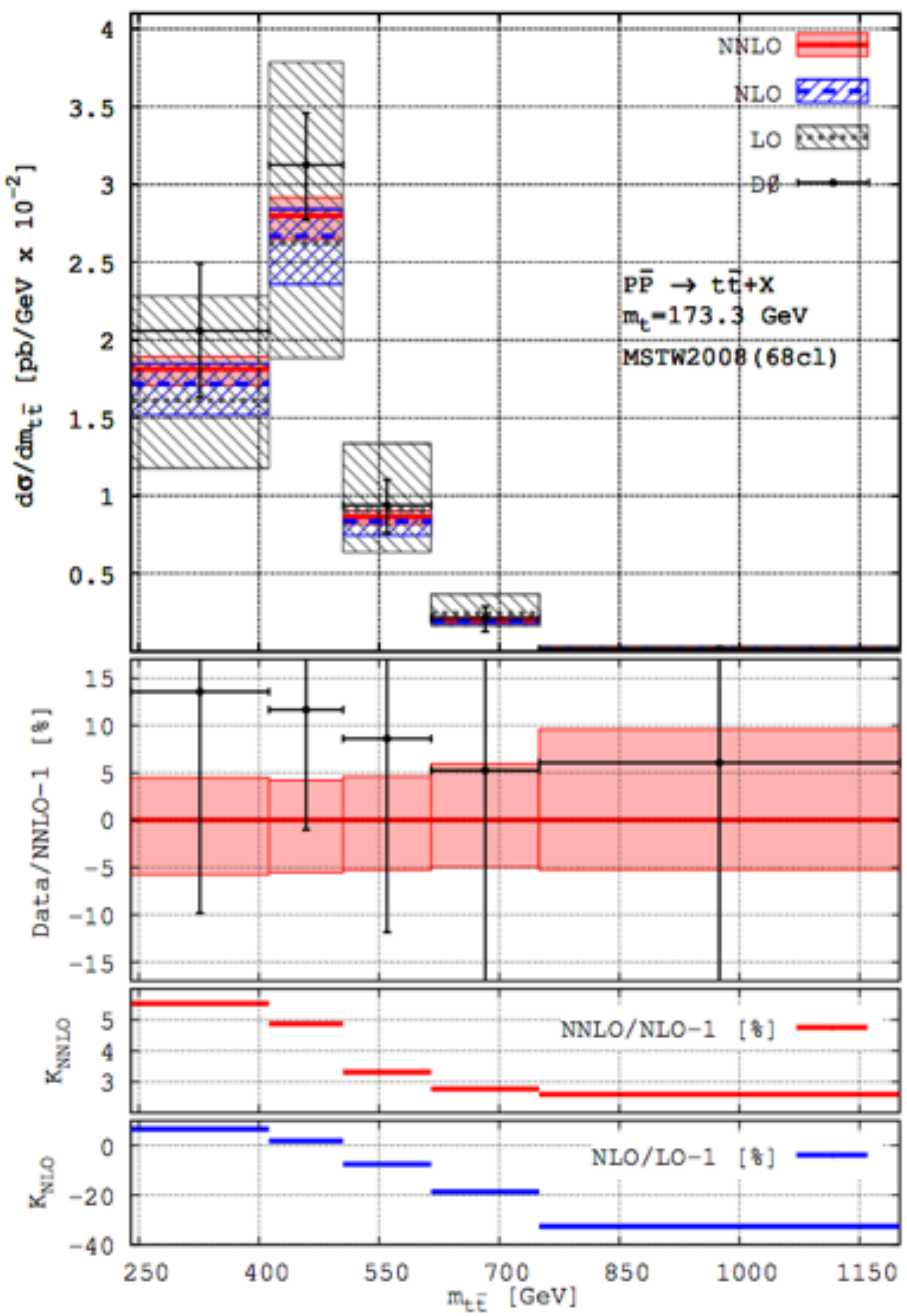
✓ NNLO top pair production
  Czakon, Fiedler, Mitov [PRL 116(2016) 082003]
  Czakon, Mitov [JHEP 1301(2015)]

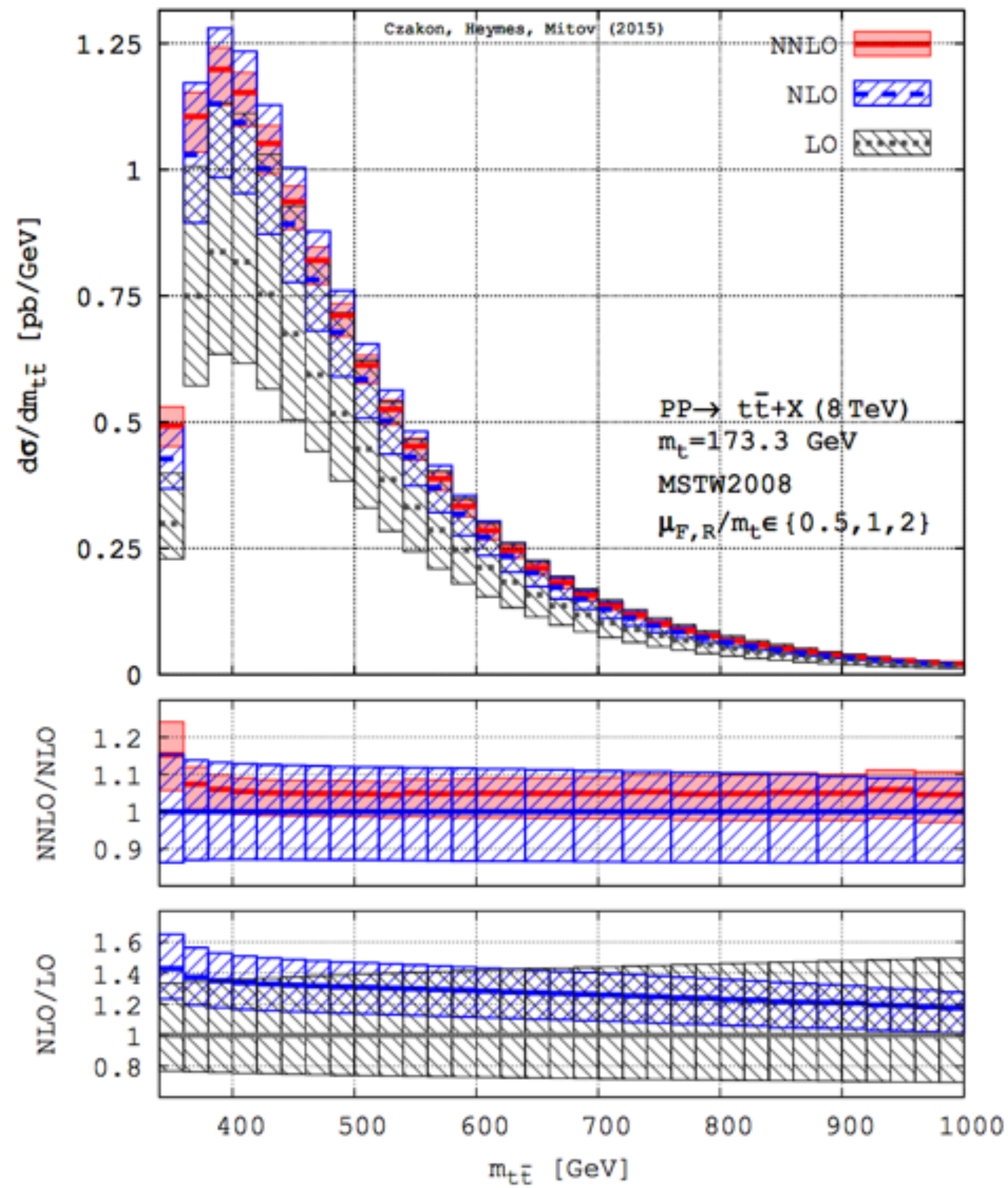
✓ W/Z+j and W/Z transverse momentum distributions
  Gehrmann-De Ridder et al [1605.04295]
  Boughezal, Liu, Petriello [1602.08140]
  Boughezal, Liu, Petriello [1602.06965]
  Boughezal et al [PRL 116(2016) 152001 & 062002]
  Gehrmann-De Ridder et al [1507.02850]

✓ Inclusive jet cross section
  Currie et al [JHEP 1401 (2014) 110]
  Gehrmann-De Ridder et al [PRL 110 (2016) 162003]

APFELgrid, Bertone et al 1605.02070
aMCfast, Berton et al JHEP 1408 (2014) 166
MCgrid, Del Debbio et al Comput.Phys.Commun. 185 (2014) 2115-2126
APPLgrid, Carli et al EPJC66 (2010) 503-524
FASTNLO, Kluge et al

The NNLO frontier - top data





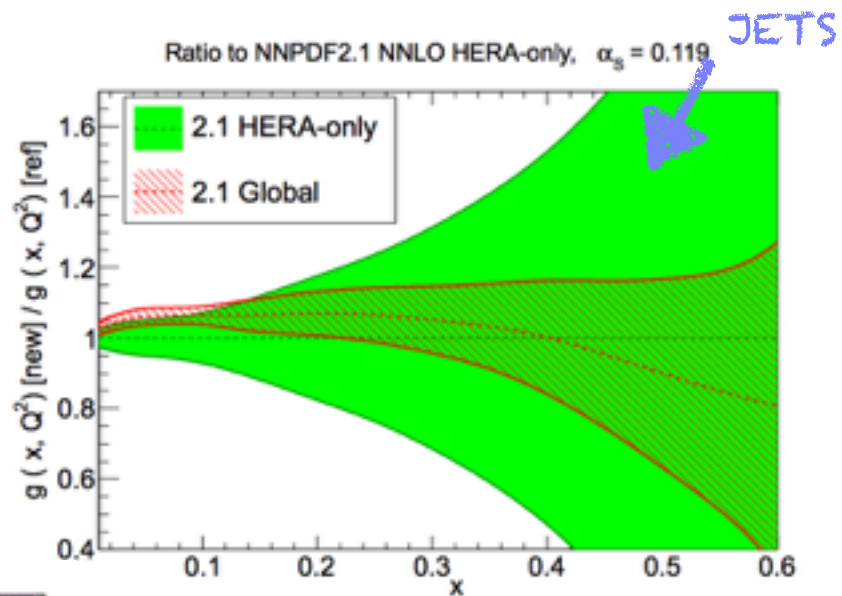
Czakon, Fiedler, Mitov [PRL 116(2016) 082003]

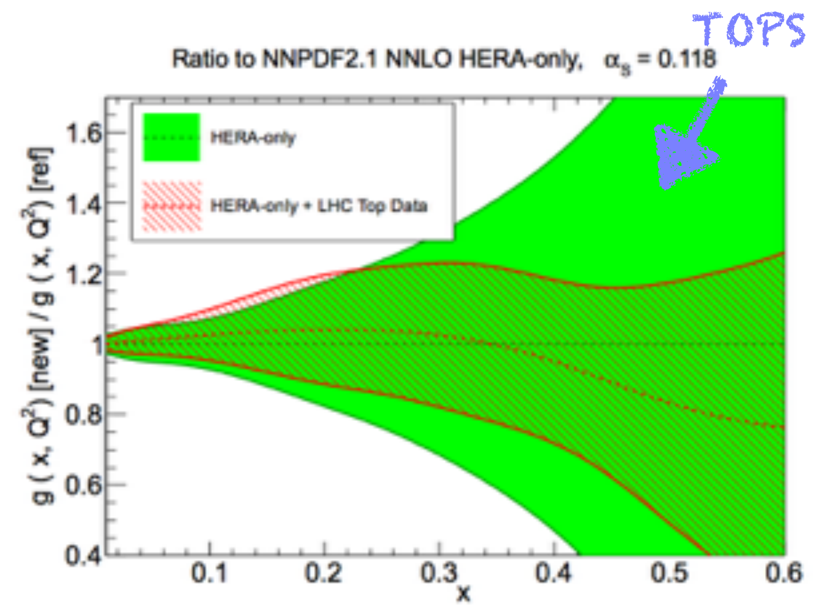
The NNLO frontier - top data

Total cross section →

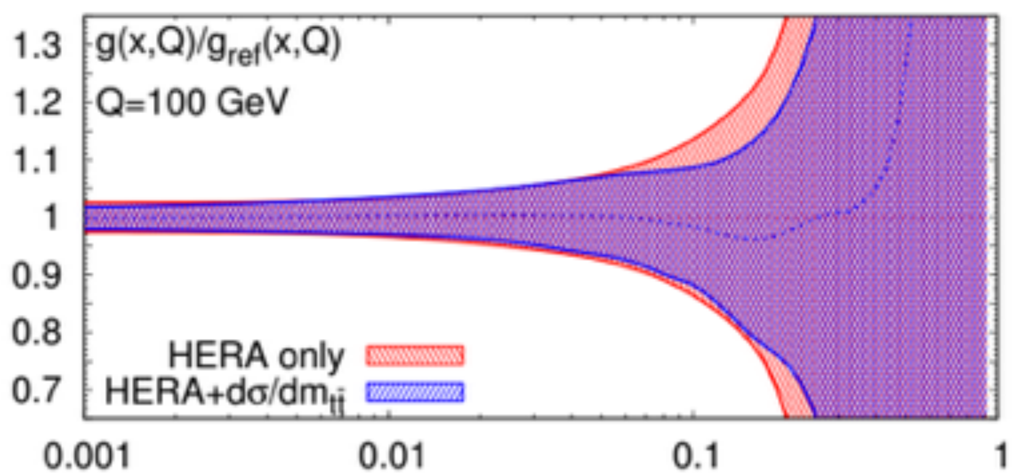
Differential cross section

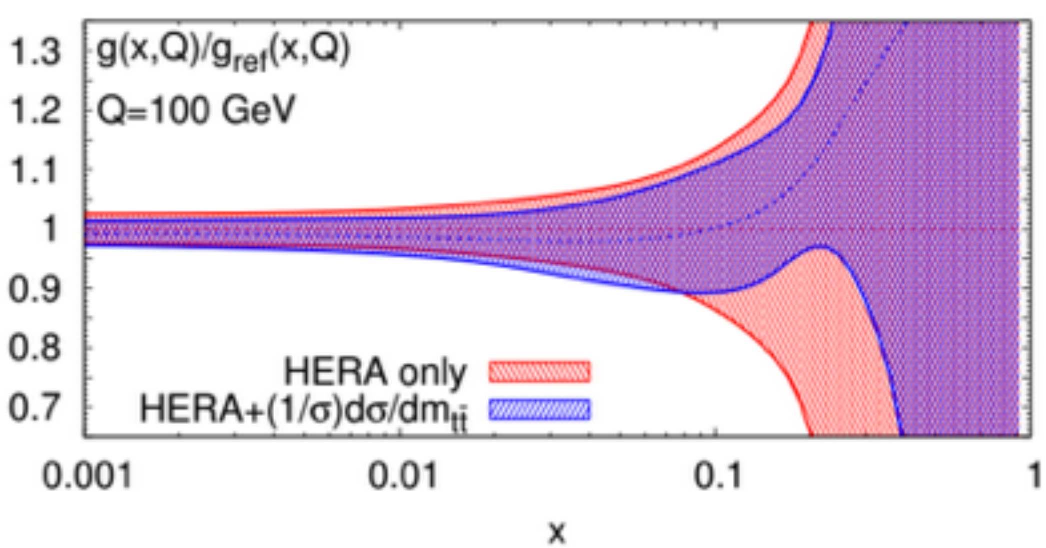
↓





Czakon et al [JHEP 1307 (2013) 167]
Beneke et al [JHEP 1207 (2012) 194]



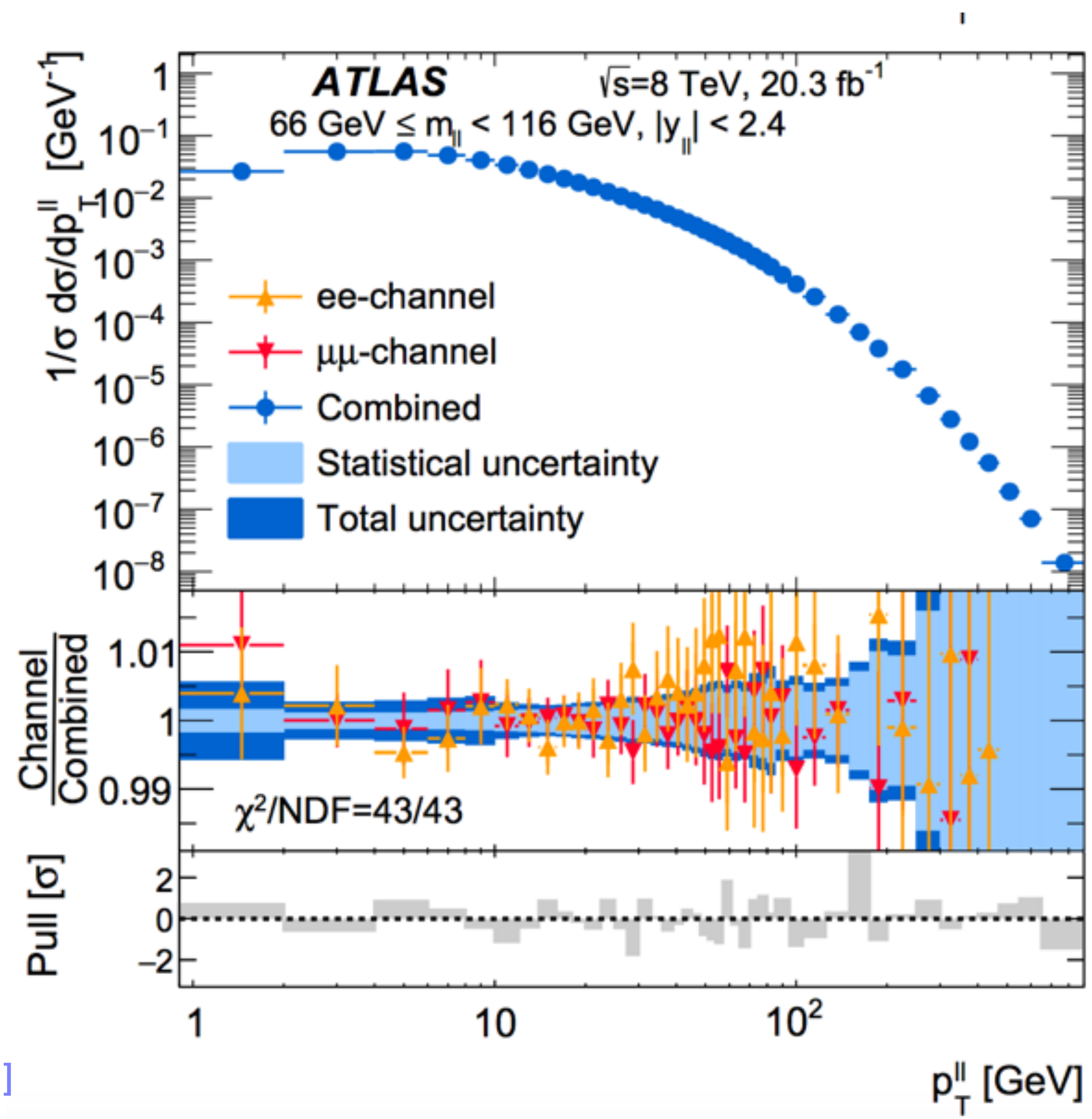


- Inclusion of top pair production data (total cross section and differential distributions) competitive to jets data and cleaner from non-perturbative effects

Courtesy of J. Rojo
Czakon, Hartland, Mitov, Nocera and Rojo, in preparation

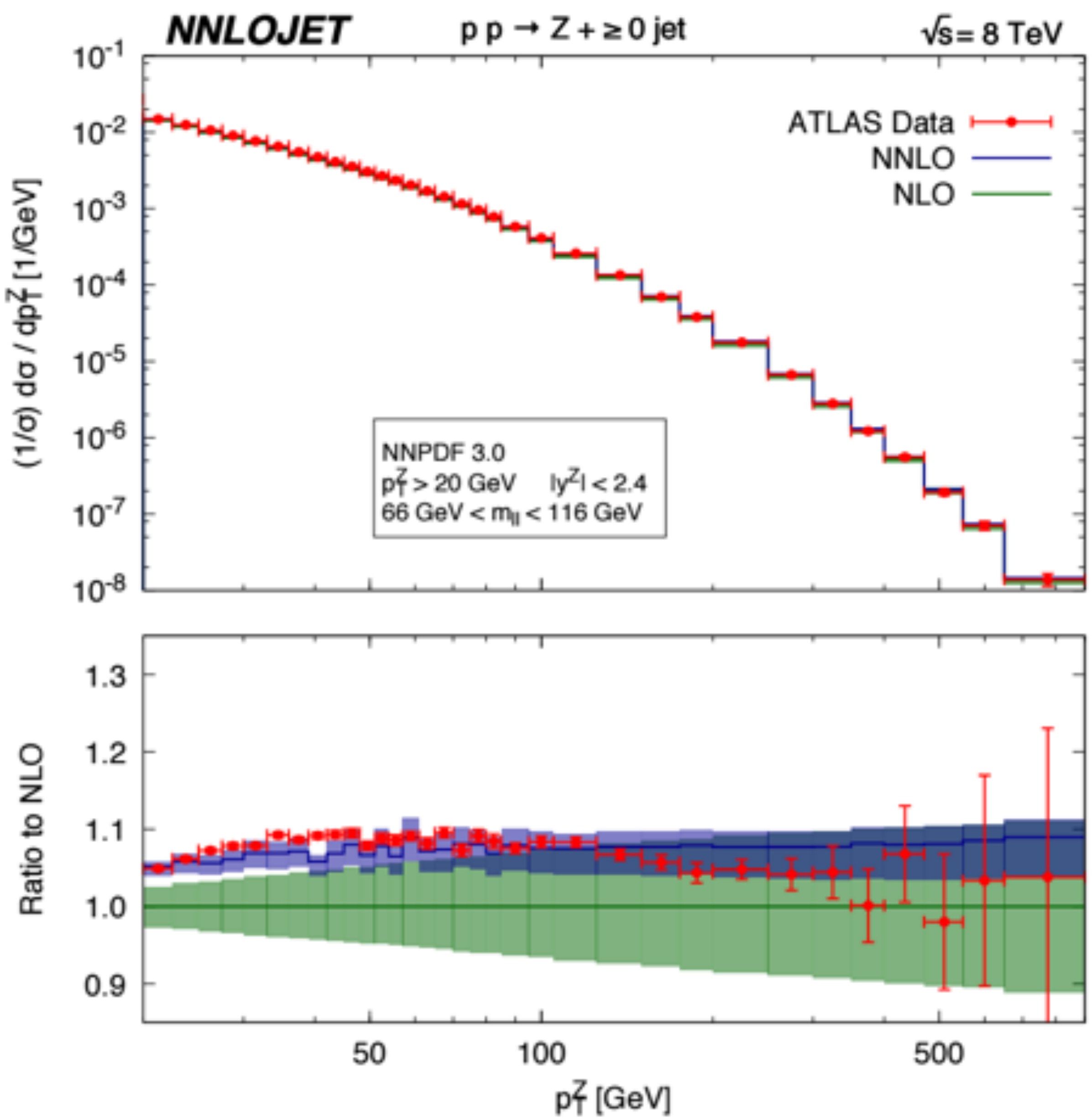
The NNLO frontier - Z pT

- Experimental precision < 1% up to pT~200 GeV
- Expect a great impact on the quark-gluon luminosity
- To fit the data NNLO corrections are needed, discrepancies in non-normalised distributions

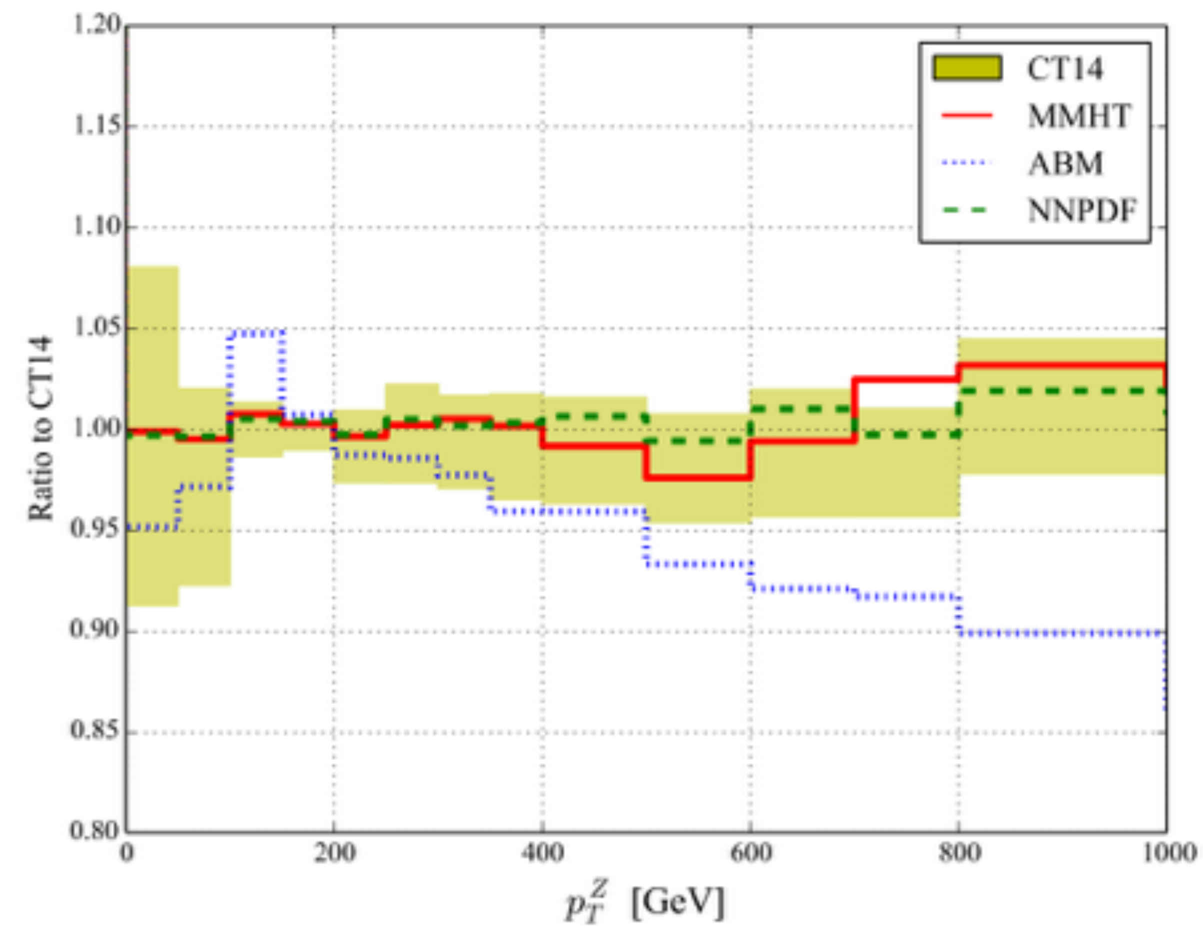


ATLAS corrections [1512.02192]

The NNLO frontier - Z pT



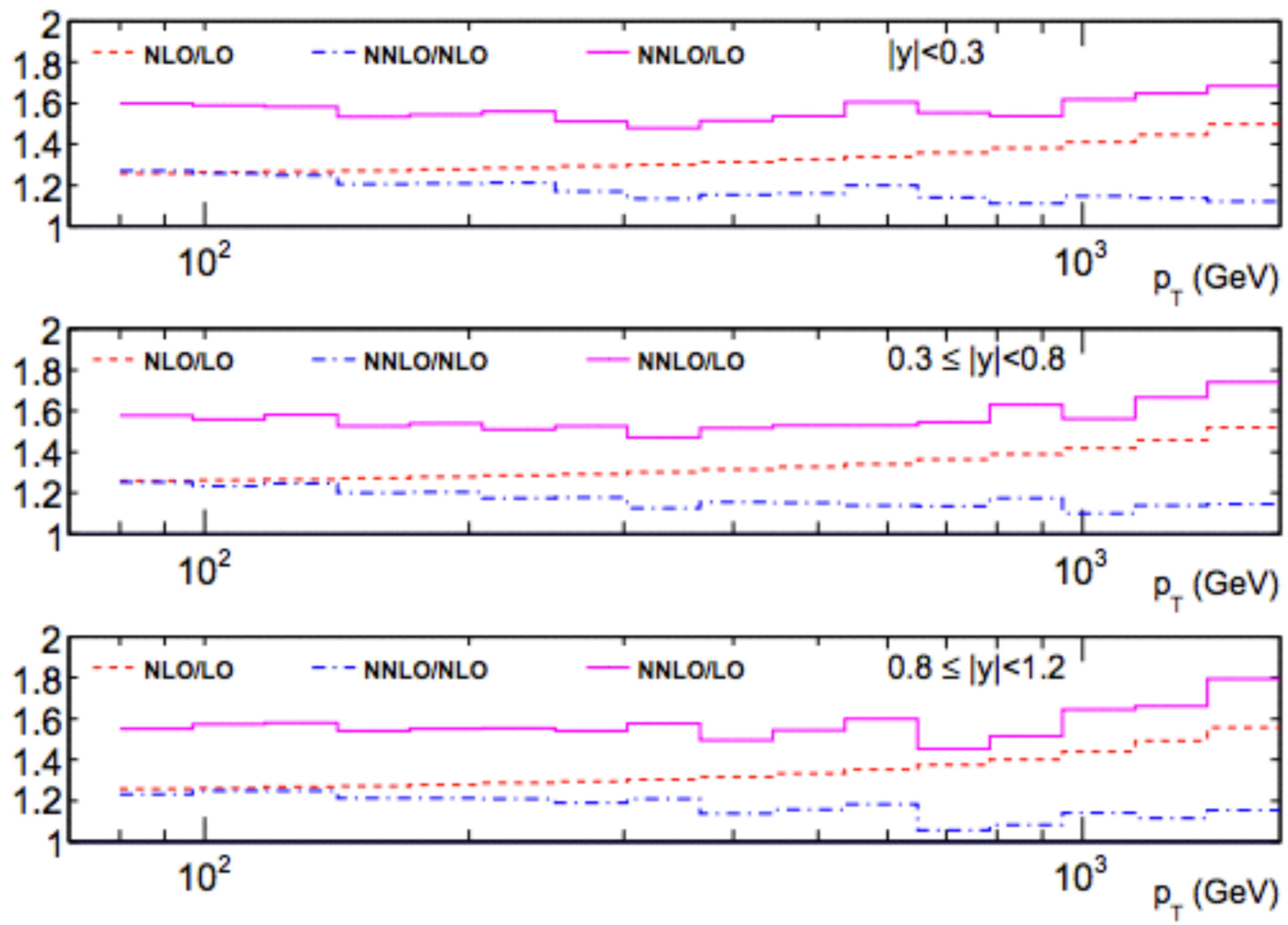
Gehrmann-De Ridder et al [1605.04295]



Boughezal, Liu, Petriello [1602.08140]

- Z + 0j less subject to QCD hadronization effects than Z + 1j process
- Study of effect of inclusive double differential Z pT distribution by ATLAS and CMS on NNLO PDF fit in progress

The NNLO frontier - jets data

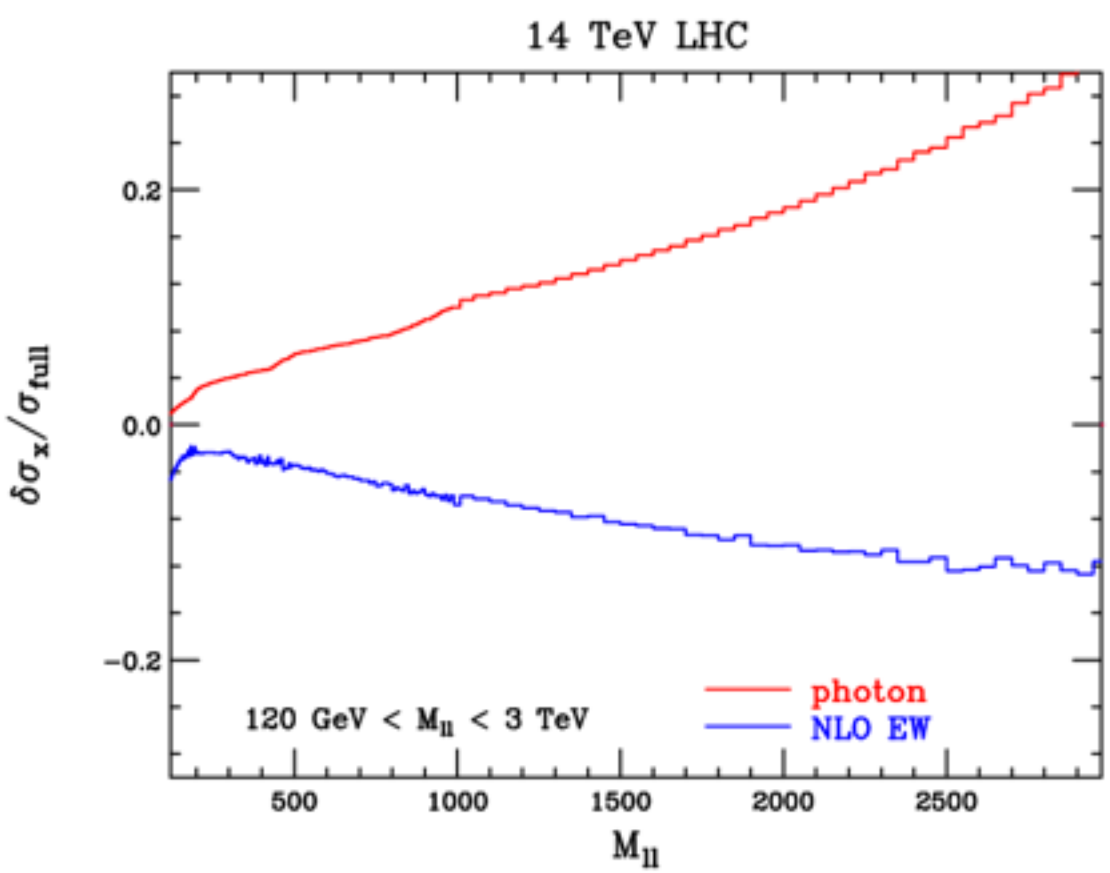


- NNLO corrections only partially known (gg channel)
- Several PDF groups make different choices: CT14 includes all jet data in NNLO fit assuming overall C-factor small, MMHT14 and ABM12 do not include LHC jet data at NNLO, NNPDF3.0 include some jet data based on goodness of threshold approximation
- These choices affect precision of the gluon, full NNLO calculation is very much needed

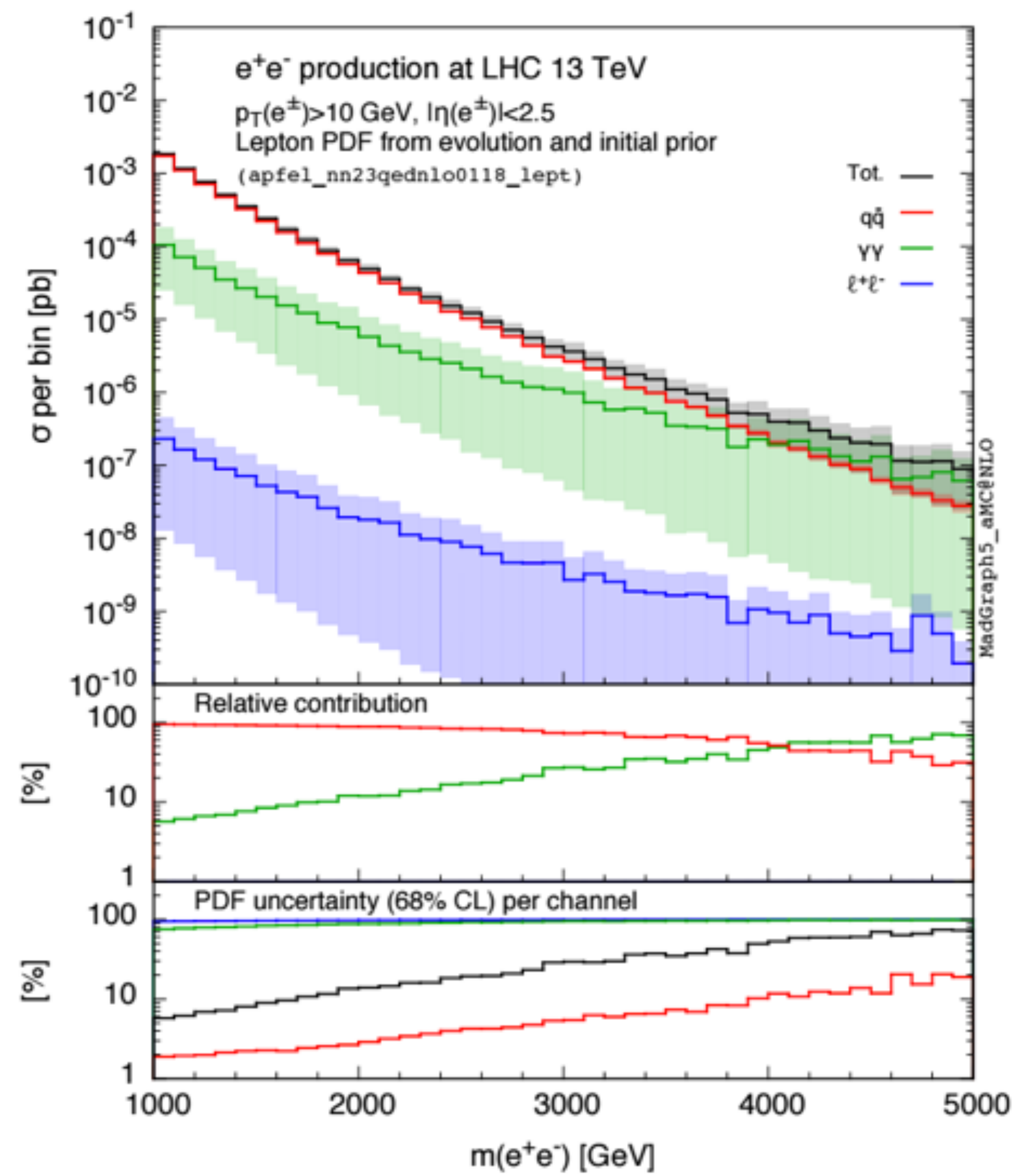
Gehrmann-De Ridder et al [PRL 110 (2016) 162003]

EW corrections matter

- EW corrections become relevant at the current precision level as are sizeable at large invariant mass
- Full inclusion of EW corrections requires initial γ PDF, which induces large uncertainty



Boughezal et al [Phys.Rev. D89 (2014)3, 034030]



Bertone et al [JHEP 1511 (2015) 194]

The photon PDF

- **NNPDF23QED** provides γ PDF and its uncertainty at (N)NLO QCD + LO QED, by reweighting photon PDF
  Ball et al [Nucl.Phys. B877 (2013)]

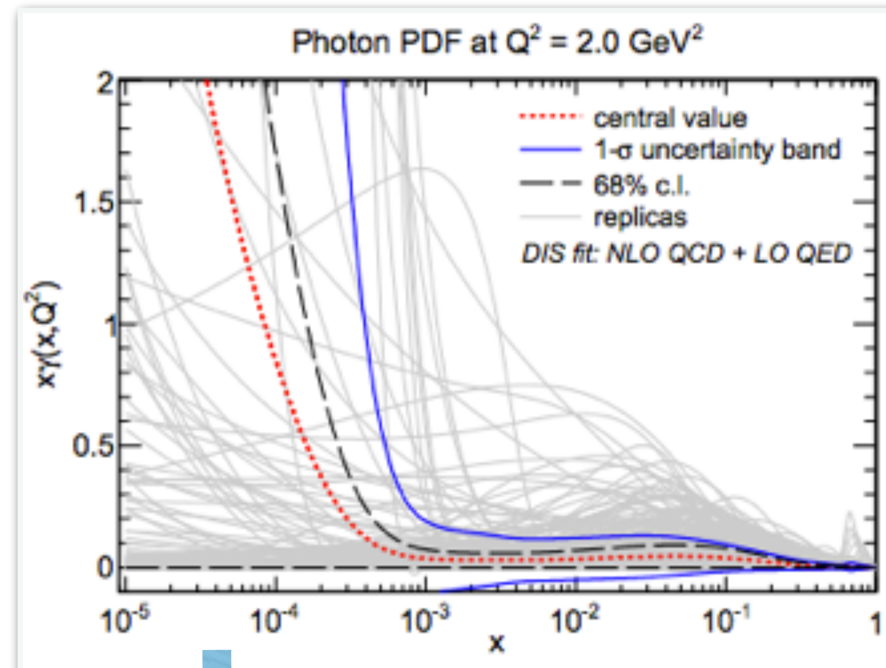
- **CT14QED** set based on two-parameter ansatz from model of photon radiate from valence quarks (extension to MRST2004QED model)
  Schmidt et al [1509.02905]

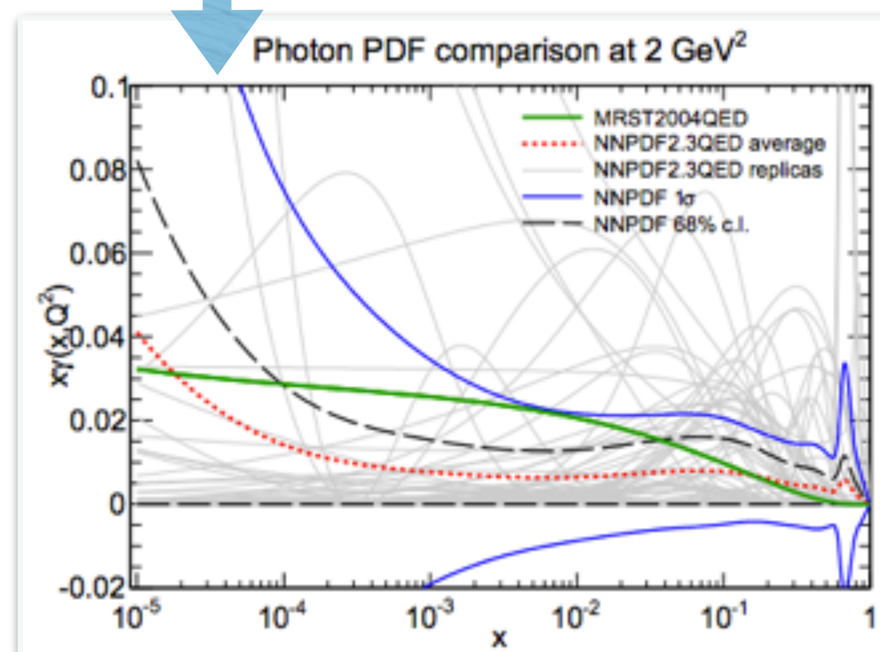
$$f_{\gamma/p}(x,Q_0) = \frac{\alpha}{2\pi}\left(A_u e_u^2 \tilde{P}_{\gamma q}\circ u^0(x) + A_d e_d^2 \tilde{P}_{\gamma q}\circ d^0(x)\right)$$
$$f_{\gamma/n}(x,Q_0) = \frac{\alpha}{2\pi}\left(A_u e_u^2 \tilde{P}_{\gamma q}\circ d^0(x) + A_d e_d^2 \tilde{P}_{\gamma q}\circ u^0(x)\right)$$

- γ PDF poorly determined by DIS data. Need hadron collider processes where γ contributes at LO (on-shell W,Z production and low/high mass DY)

- NNPDF plan: fit photon along with other PDFs (thanks to upgrade of APFEL - simultaneous diagonalization of QCD and QED evolution matrices - and APFELgrid - now includes photon-induced processes)

DIS





DIS+LHC

Resummed PDFs

- Multi-scale processes: log(Qi/Qj) = L arise, which may spoil perturbative expansion
- If (α_S * L) ~ O(1) fixed order perturbative QCD is no longer justified
- Resummation effectively rearranges perturbative series

fixed order		all order (L = some large logarithm)	
$\frac{\sigma}{\sigma_0} = 1$	LO	$\ln\frac{\sigma}{\sigma_0} = \alpha^n L^{n+1}$	LL
$+ c_1\alpha$	NLO	$+ \alpha^n L^n$	NLL
$+ c_2\alpha^2$	NNLO	$+ \alpha^n L^{n-1}$	NNLL
$+ \dots$		$+ \dots$	

- Various kinds of logs:

L = log (1-x) threshold (soft-gluon) resummation ⟵ Ball et al, JHEP09(2015)091

L = log (1/x) high-energy (small-x) resummation ⟵ See Simone's talk

L = log (pT/M) transverse momentum resummation

Threshold resummation

- Threshold resummation: initial energy just enough to produce final state with mass M, so emissions forced to be soft and logs at each order in PT are enhanced

$$x = \frac{M^2}{\hat{s}} \qquad \mathrm{NLO}: \; M^2 = z\hat{s} \qquad \left[\frac{\log^k(1-z)}{(1-z)}\right]_+$$

- Transform factorised cross section into Mellin space

$$\sigma(x,Q^2) = x\sum_{a,b}\int_x^1 \frac{dz}{z}\mathcal{L}_{ab}\left(\frac{x}{z},\mu_{\mathrm{F}}^2\right)\frac{1}{z}\hat{\sigma}_{ab}\left(z,Q^2,\alpha_s(\mu_{\mathrm{R}}^2),\frac{Q^2}{\mu_{\mathrm{F}}^2},\frac{Q^2}{\mu_{\mathrm{R}}^2}\right)$$

$$\sigma(N,Q^2) = \int_0^1 dx\, x^{N-2}\sigma(x,Q^2) = \sum_{a,b}\mathcal{L}_{ab}(N,Q^2)\hat{\sigma}_{ab}\left(N,Q^2,\alpha_s\right)$$

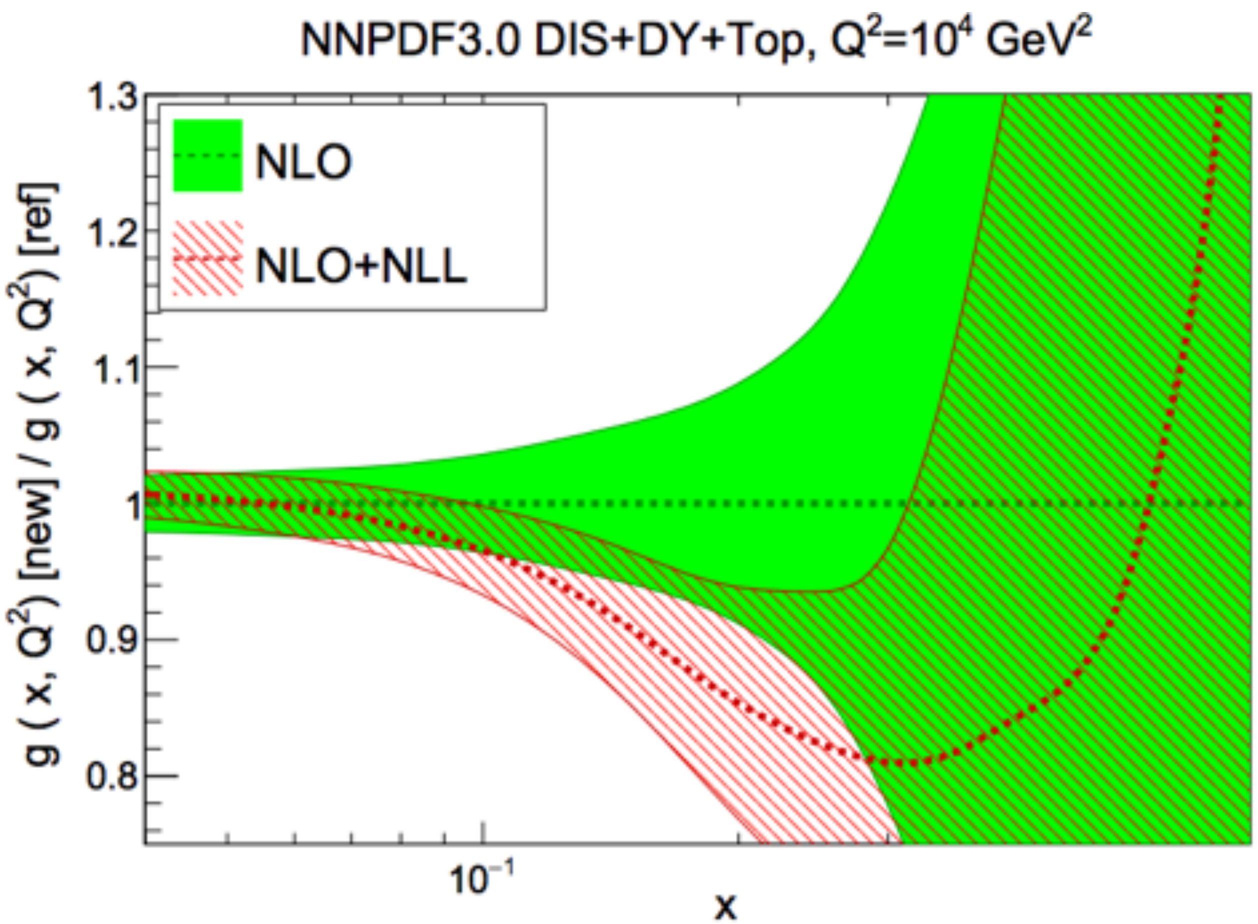
- In the MSbar scheme PDF evolution does not contain large-x logs and the effect of resummation can be included in resummed coefficient functions

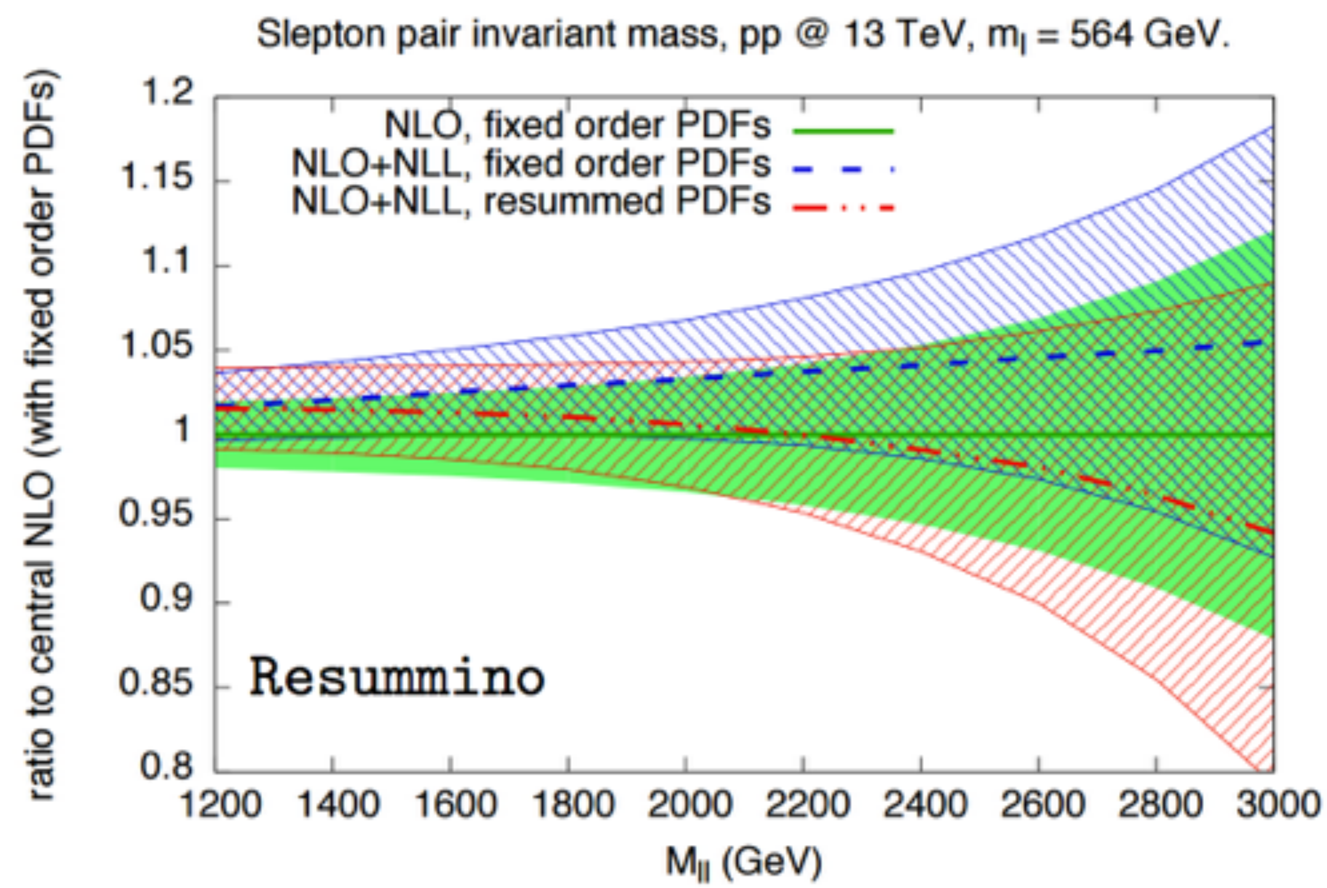
$$\hat{\sigma}_{ab}^{(\mathrm{res})}(N,Q^2,\alpha_s) = \sigma_{ab}^{(\mathrm{born})}(N,Q^2,\alpha_s)\,C_{ab}^{(\mathrm{res})}(N,\alpha_s)$$

$$C^{(N\text{-soft})}(N,\alpha_s) = g_0(\alpha_s)\exp\mathcal{S}(\ln N,\alpha_s),$$

$$\mathcal{S}(\ln N,\alpha_s) = \left[\frac{1}{\alpha_s}g_1(\alpha_s\ln N) + g_2(\alpha_s\ln N) + \alpha_s g_3(\alpha_s\ln N) + \ldots\right]$$

Threshold-resummed PDFs



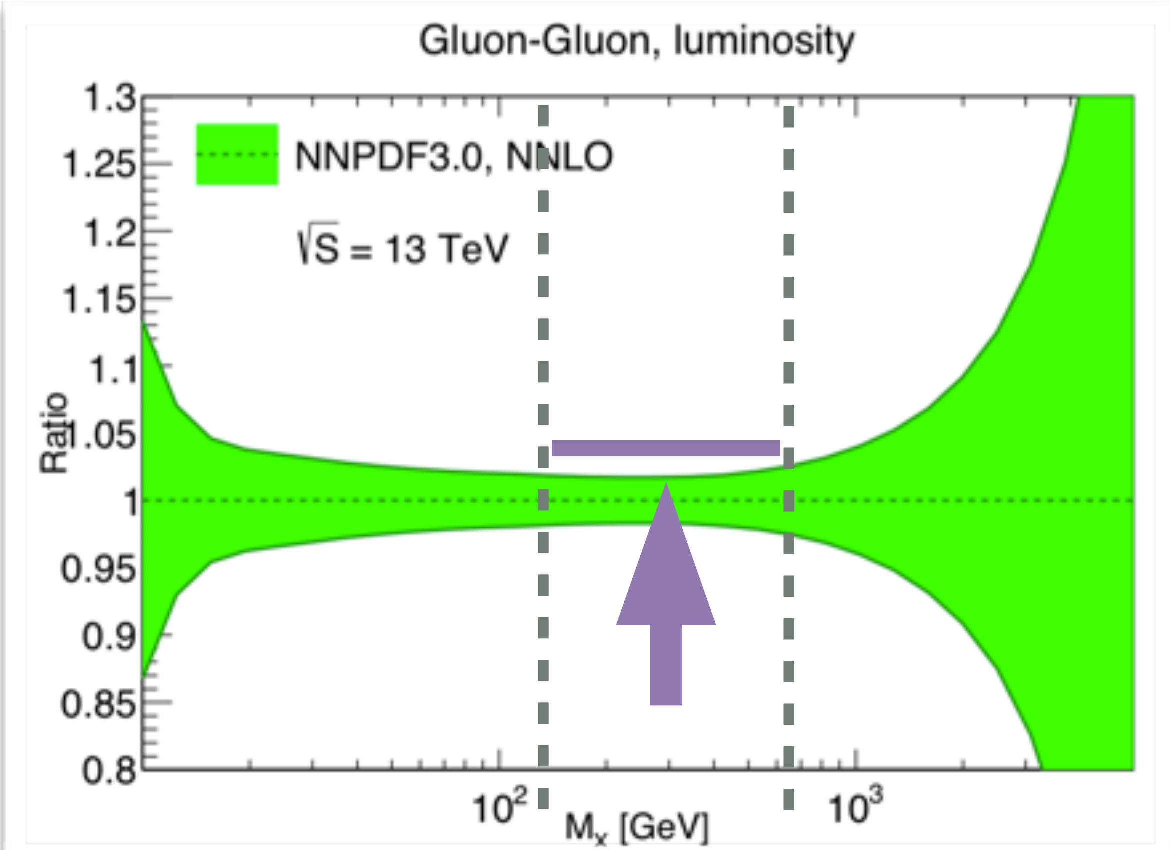


Bonvini et al, JHEP 1509 (2015) 191

- Threshold-resummed PDFs will be suppressed as compared to fixed-order PDFs
- Mostly due to enhancement of NLO+NLL xsecs used in the fit of DIS structure functions and DY distributions
- This suppression partially or totally compensates enhancements in partonic cross sections
- Phenomenologically relevant for new physics processes [Beenakker et al. EPJC76 (2016)2, 53]

Beyond the state of the art

Theoretical uncertainty



Theory boundaries

Precision Physics

Discovery region

Reduce **theoretical uncertainty** in PDF fits: resummation, EW effects, HQ masses, intrinsic HQ, parton shower

Introduce a way to measure residual **theoretical uncertainty** in PDF fits

- PDF fits are all made at a given theoretical accuracy (fixed order perturbative QCD)
- PDF uncertainties only reflect lack of information from data given the theory
- Changes in theory may cause shifts outside the error band: lack of accuracy!
- No longer an option!

Theoretical uncertainty

Exploit precise **LHC data** to reduce PDF uncertainties

Explore potential constraints from **future colliders**

Reduce **theoretical uncertainty** in PDF fits: resummation, EW effects, HQ masses, intrinsic HQ, parton shower

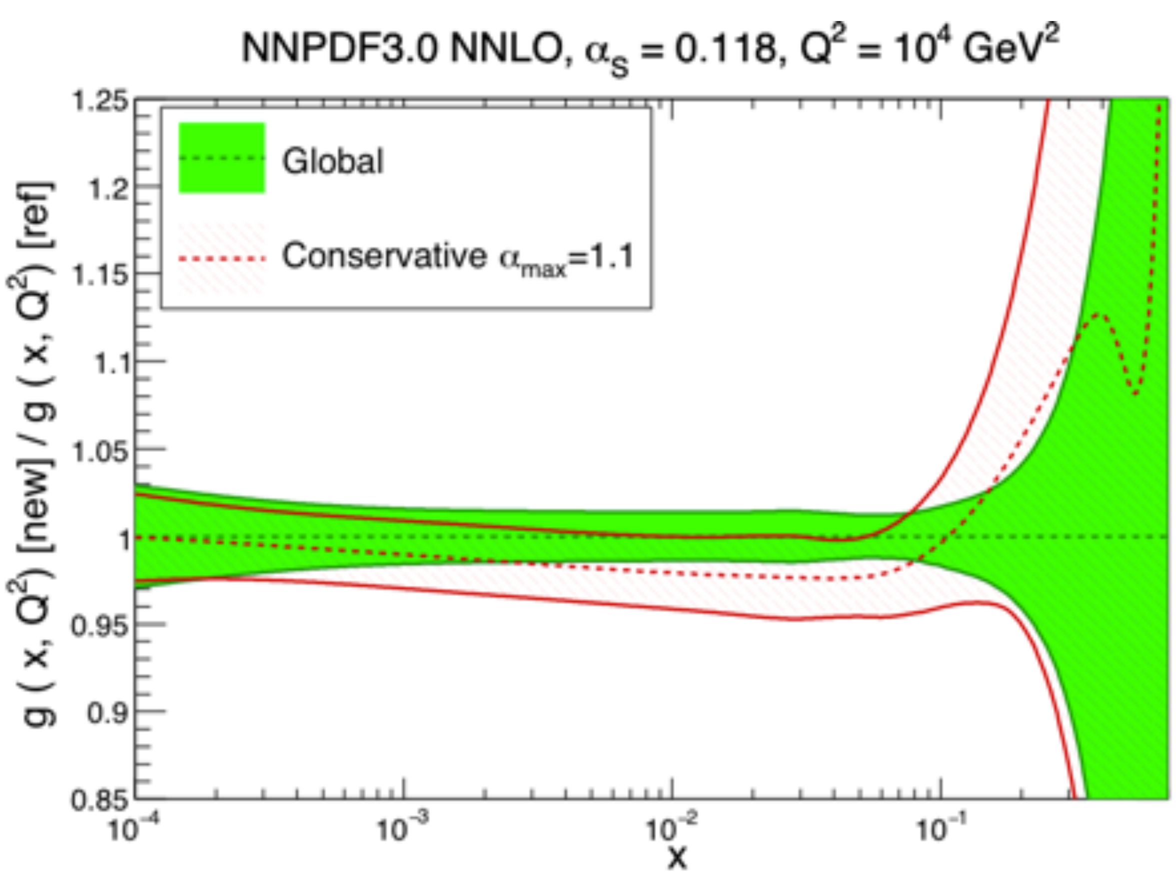
Introduce a way to measure residual **theoretical uncertainty** in PDF fits

The higher the energy regime, the more theory boundaries are probed
The smaller the experimental uncertainty, the more crucial is theory uncertainty

A final remark

Q: As more data at higher energy will be released, how can we make sure that one does not absorb new physics in PDFs?

- Inconsistencies between data that enter a global PDF analysis can distort statistical interpretation of PDF uncertainties
- Inconsistency of any individual dataset with the bulk of global fit may suggest that its understanding (theory or experiment) is incomplete
- Set of **conservative partons** based on measure of consistency are crucial for a systematic inclusion of new data



NNPDF collaboration, JHEP04(2015)040

Conclusions

- Parton Distribution Functions essential ingredient for LHC phenomenology
- Accurate PDFs are required for precision SM measurements, Higgs characterisation and New Physics
- NNPDF approach provides parton distributions based on a robust, unbiased methodology, the most updated theoretical information and most relevant hard scattering data including LHC data
- Good convergence of different approaches
- Frontiers:

  Bring the pQCD loop revolution & resummations into the PDF world
  Measure hidden uncertainties: theoretical uncertainty, intrinsic heavy quark distributions, photon
  How not to include effects that go beyond DGLAP/SM formalism into PDF fits?

- A challenging and exciting road ahead!

THANK YOU!