European Research Council

Established by the European Commission

MARIA UBIALI

UNIVERSITY OF CAMBRIDGE

NEW FRONTIERS IN THE DETERMINATION OF PDFS AT THE LHC

OUTLINE

- Introduction

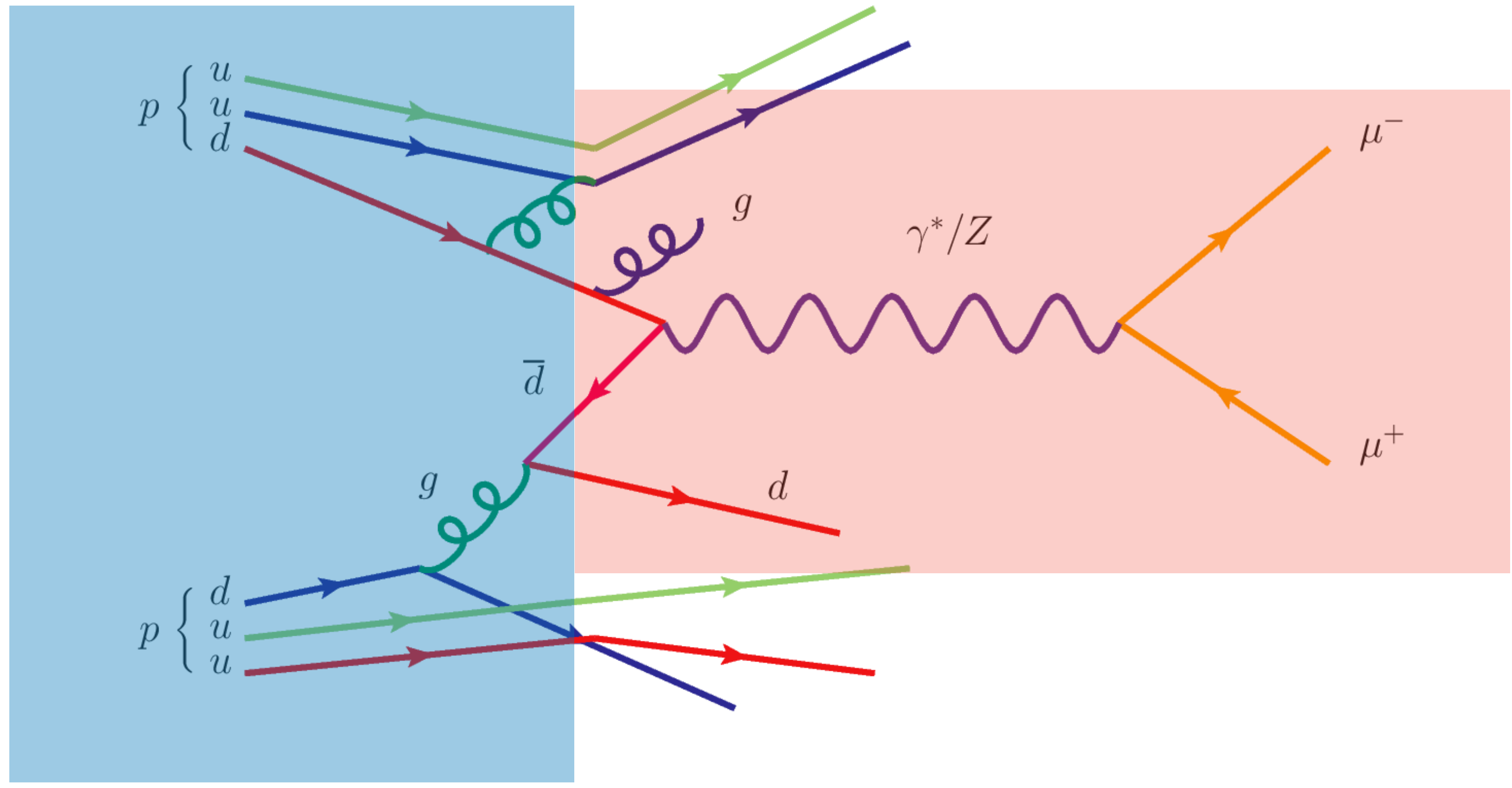
- Towards 1% precision in PDFs
  - ➡ Progress in PDF determination
  - ➡ Experimental data & Methodological advances

- New challenges
  - ➡ Theory: missing higher order uncertainties & electroweak corrections
  - ➡ Interpretation: Interplay between PDFs and indirect New Physics searches in high mass tails

- Conclusions and outlook

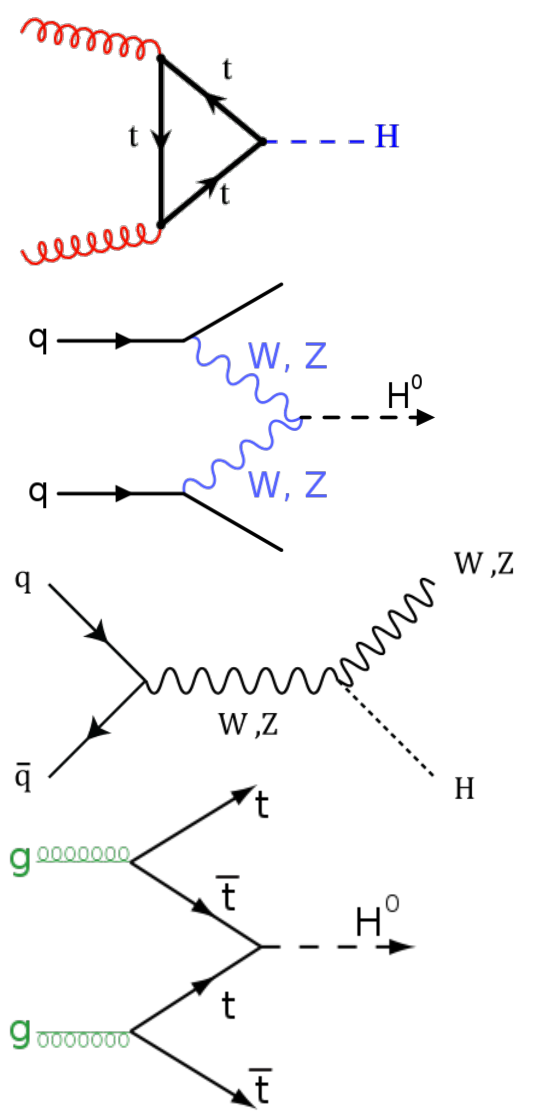
INTRODUCTION

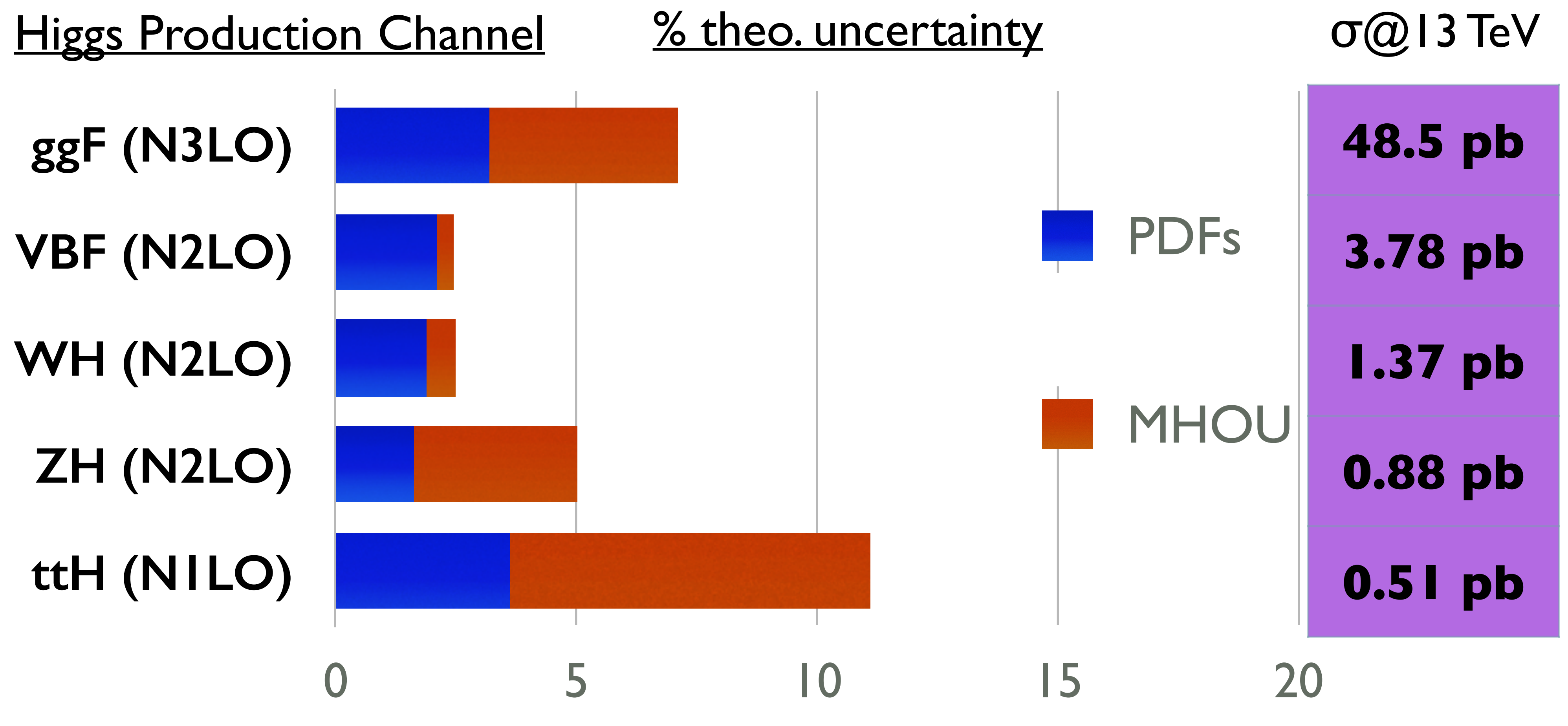
WHY ARE PDFS SO IMPORTANT FOR PRECISION PHYSICS …



$$d\sigma^{pp\to ab} = \sum_{i,j} f_i \otimes f_j \otimes d\hat{\sigma}^{ij\to ab} + \dots$$

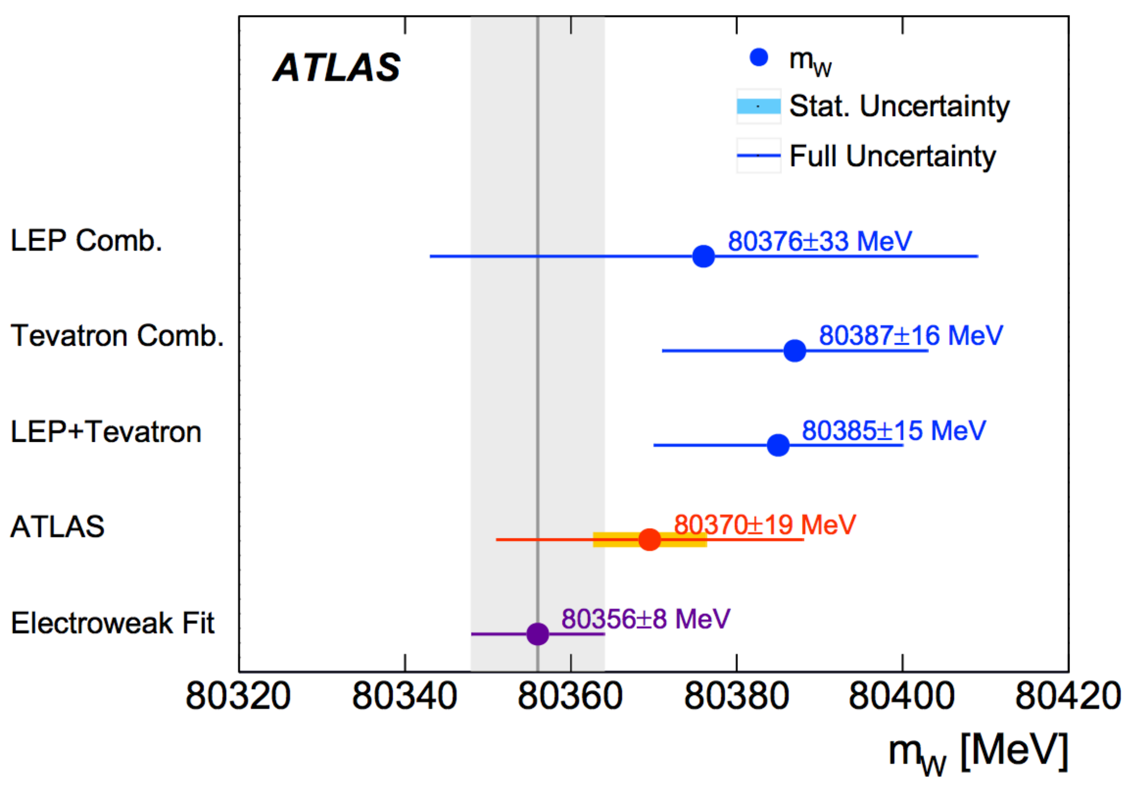
#1: Higgs physics

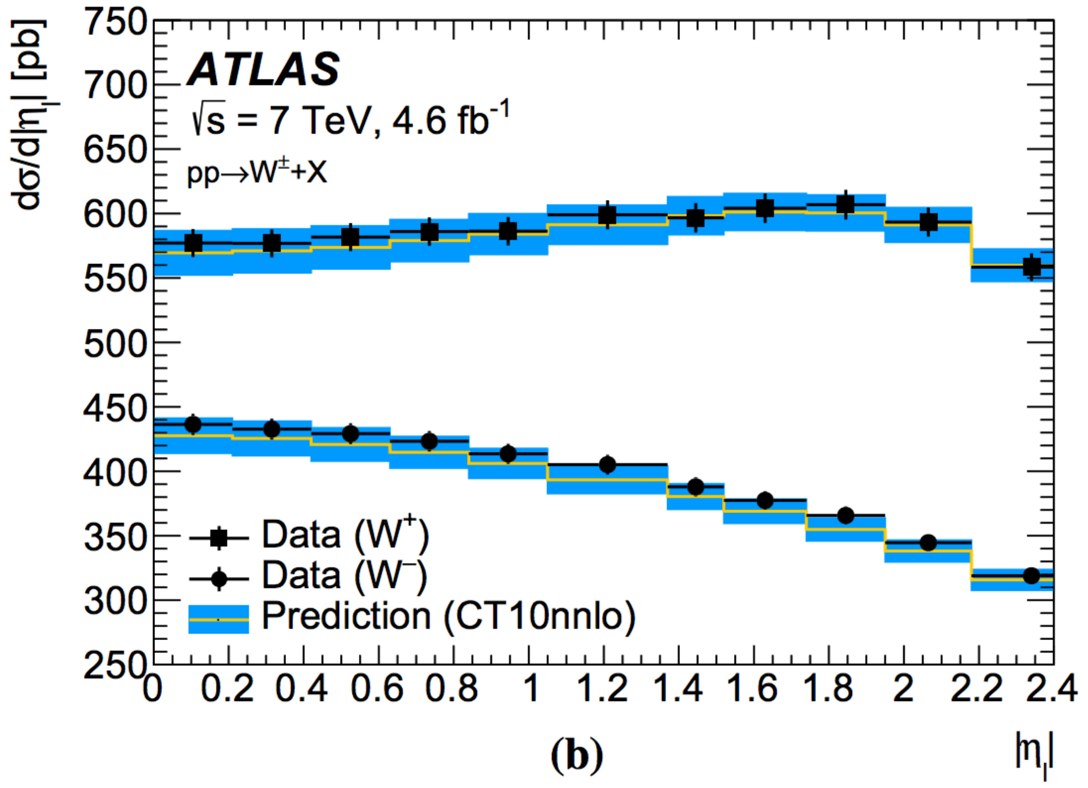




HXSWG Yellow report (2016)

#2: Determination of SM parameters



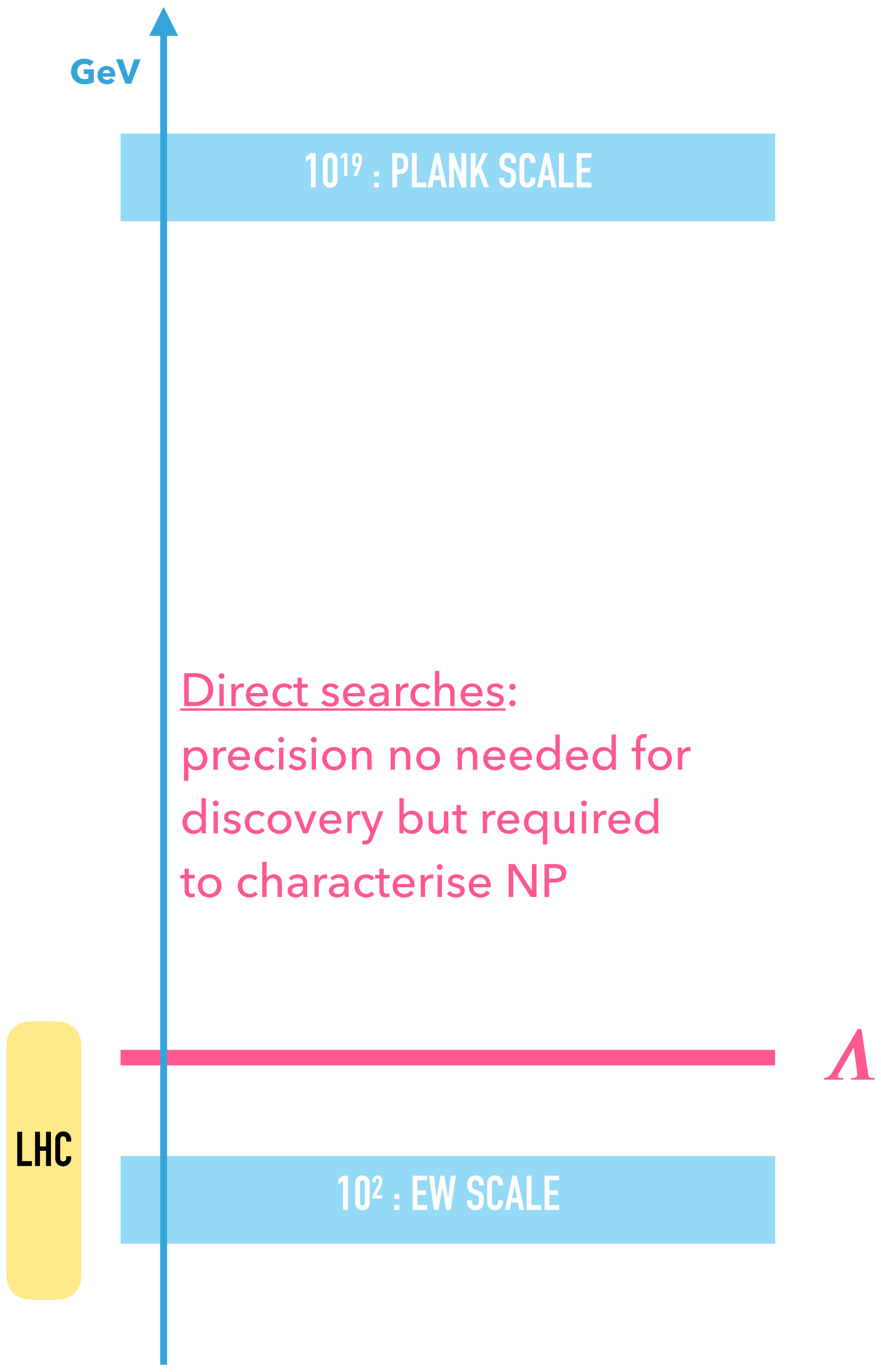


ATLAS collaboration, EPJC 78 (2018) 110

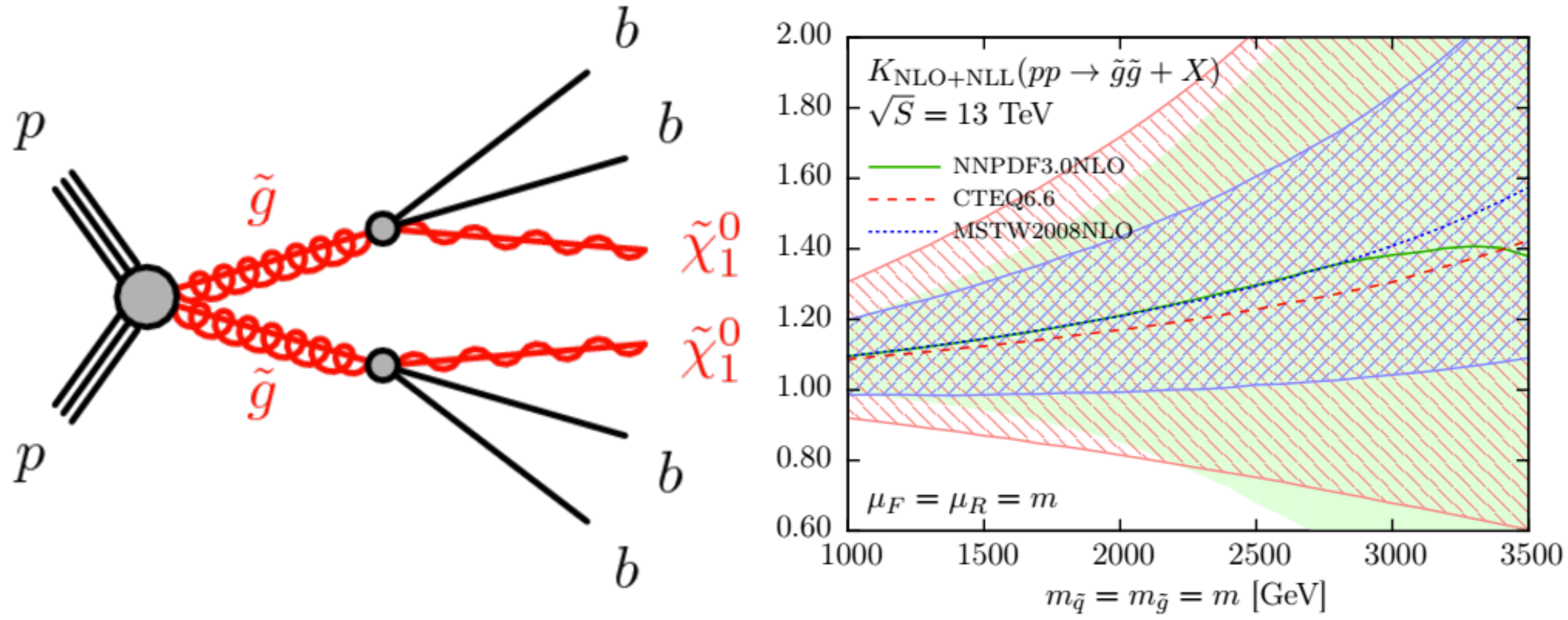
$\eta = -\ln\tan(\theta/2)$

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

… AND FOR NEW PHYSICS SEARCHES



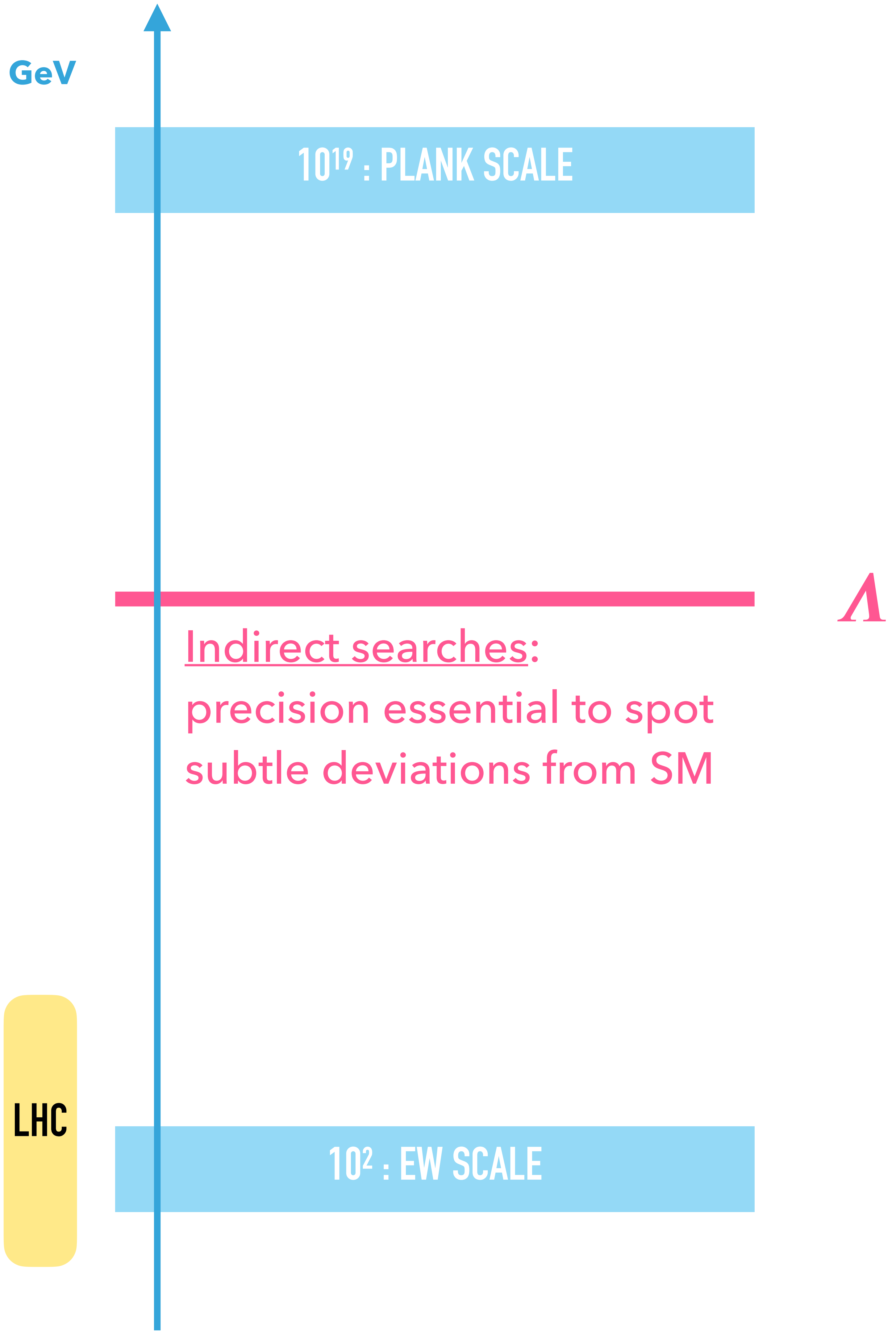
Direct searches:
precision no needed for discovery but required to characterise NP



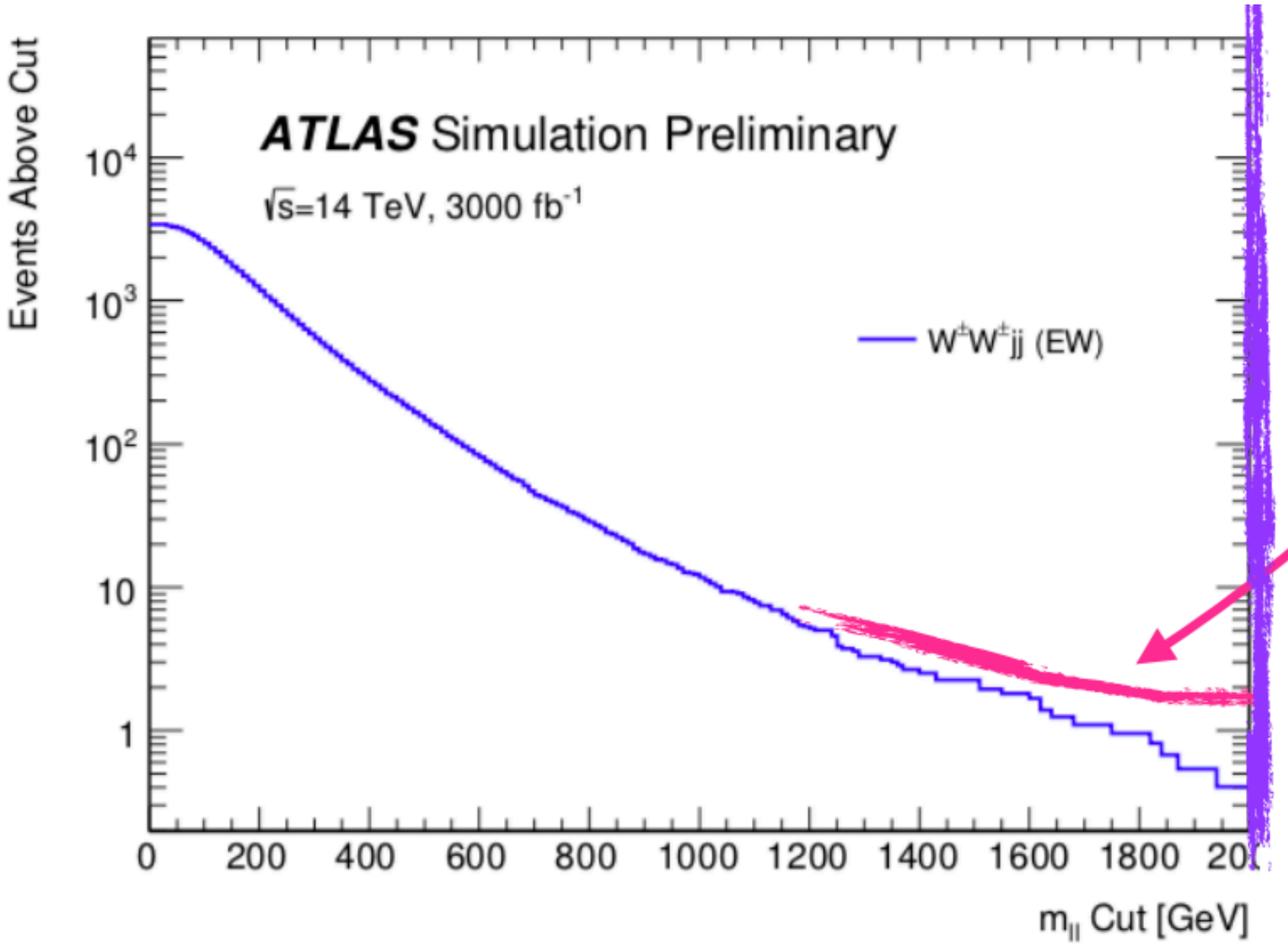
Beenakker et al.
EPJC76 (2016)2, 53

#1: Direct searches: heavy particles / large-x region

. . . AND FOR NEW PHYSICS SEARCHES



Indirect searches:
precision essential to spot subtle deviations from SM



CA Lee, HL/HE-LHC Jamboree, 1 March 2019

Deviations from SM predictions in high energy tails (Run III): new physics or limited understanding of proton structure?

#2: Indirect searches

TOWARDS 1% PRECISION IN PDF DETERMINATION

DETERMINATION OF PDFS

$$f_i(x, \mu)$$

Perturbative QCD

- DGLAP evolution equations fully known up to NNLO and partially at N3LO

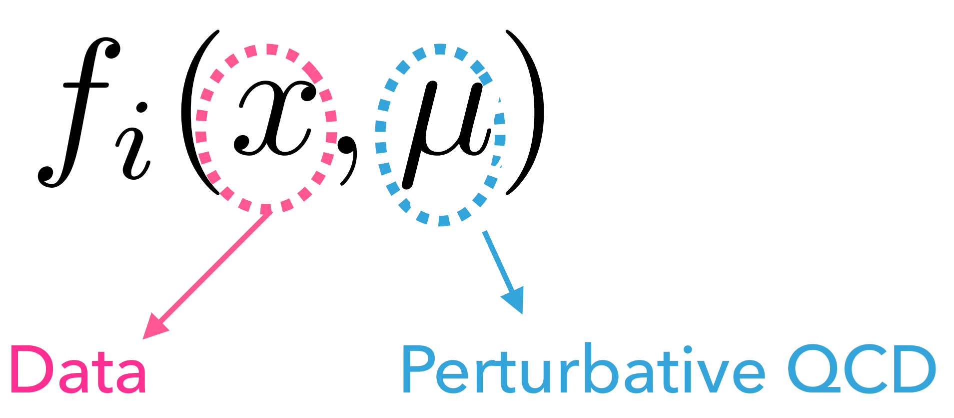
[Moch, Vermaseren, Vogt (2004)]

[Herzog, Moch, Vermaseren, Vogt, Bonvini…]

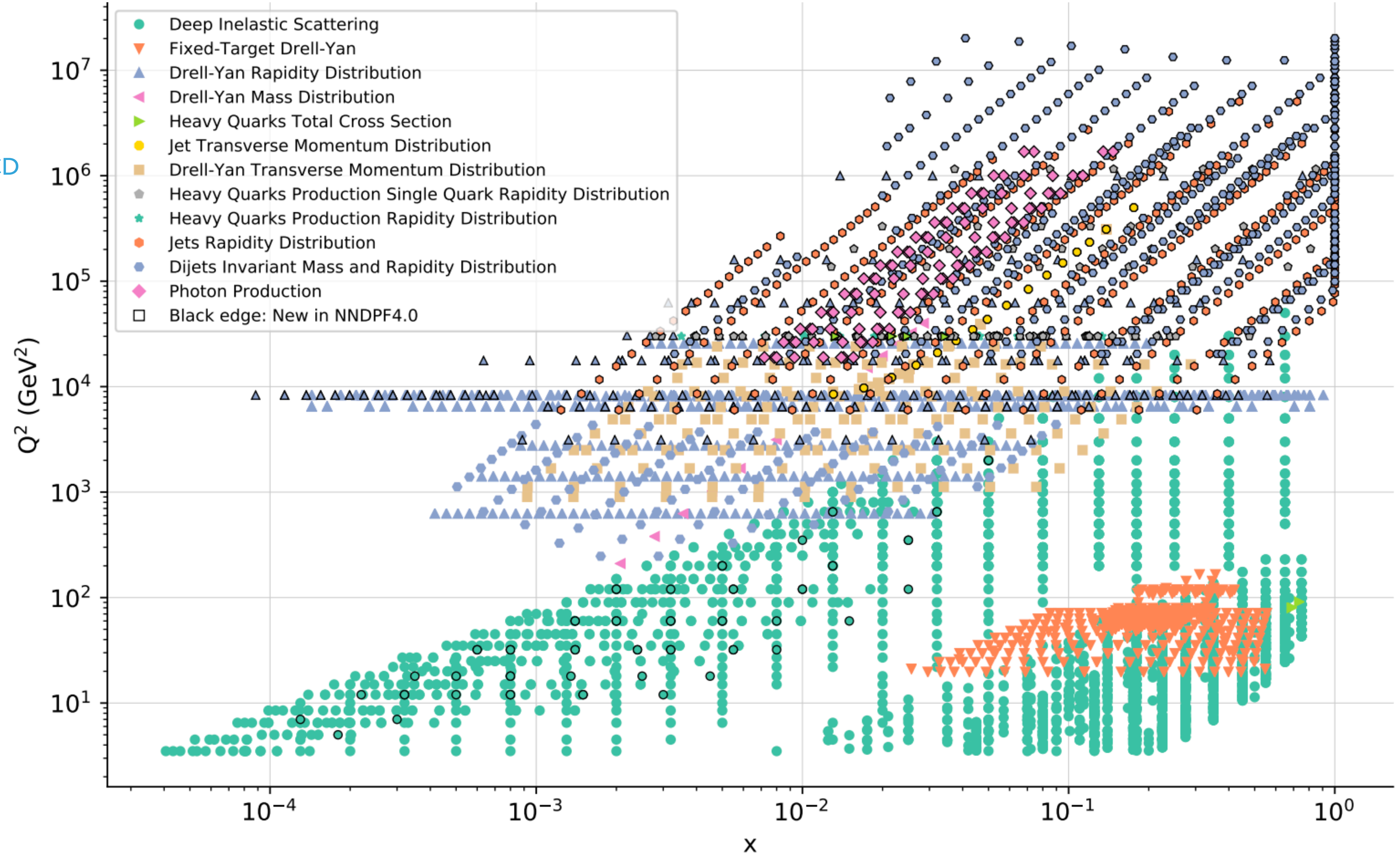
- Most theory predictions for processes in PDF fits known at NNLO

DETERMINATION OF PDFS

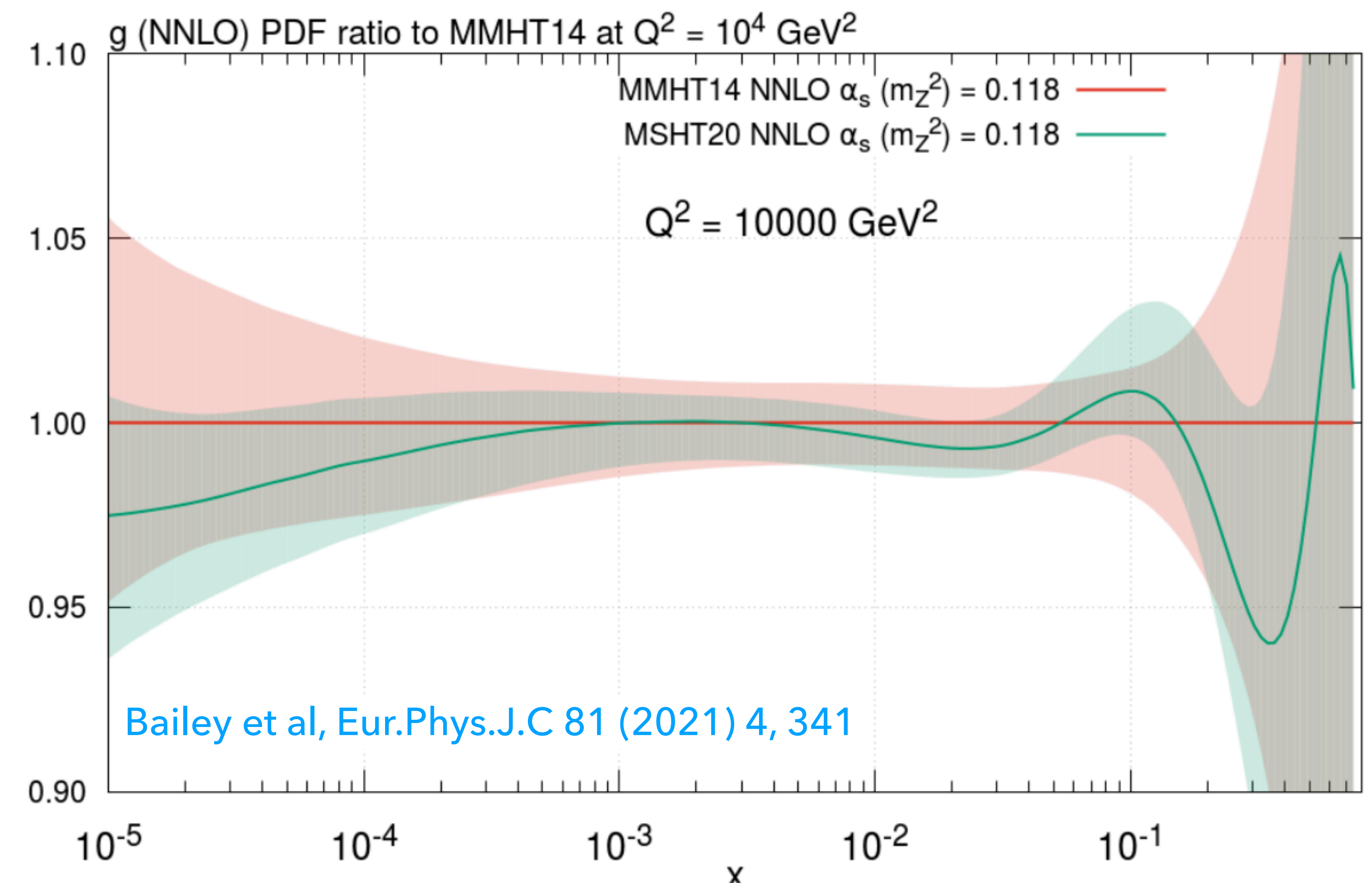
$$\chi^2 = \sum_{i,j=1}^{N_{\text{dat}}} (D_i - T_i(\theta_{\text{PDF}}))(\text{cov})^{-1}_{ij}(D_j - T_j(\theta_{\text{PDF}}))$$

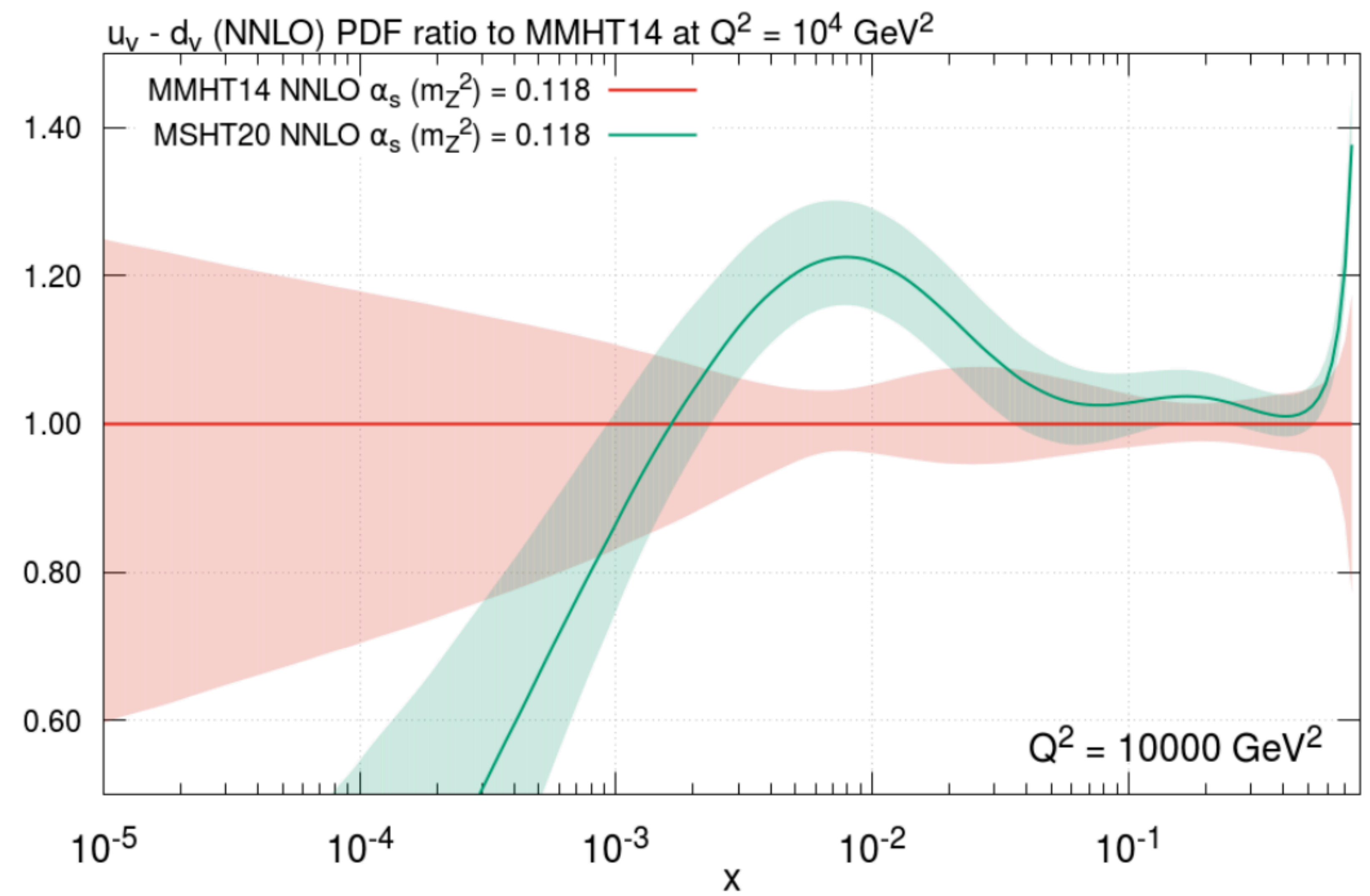


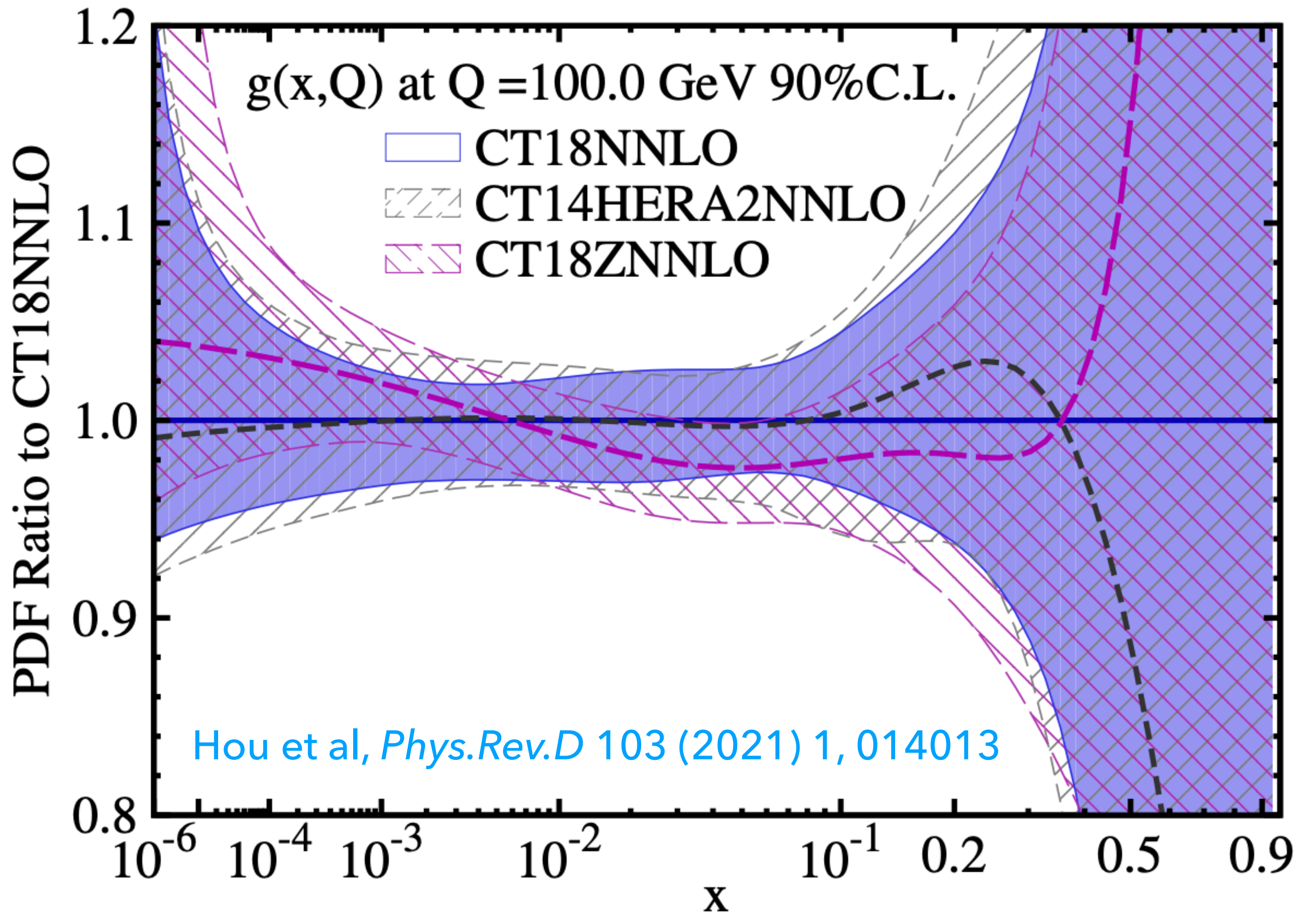
- DGLAP evolution equations fully known up to NNLO and partially at N3LO
- Most theory predictions for processes in PDF fits known at NNLO
- Precision of the most data of the order of a few percents
- Mostly from correlated systematic uncertainties

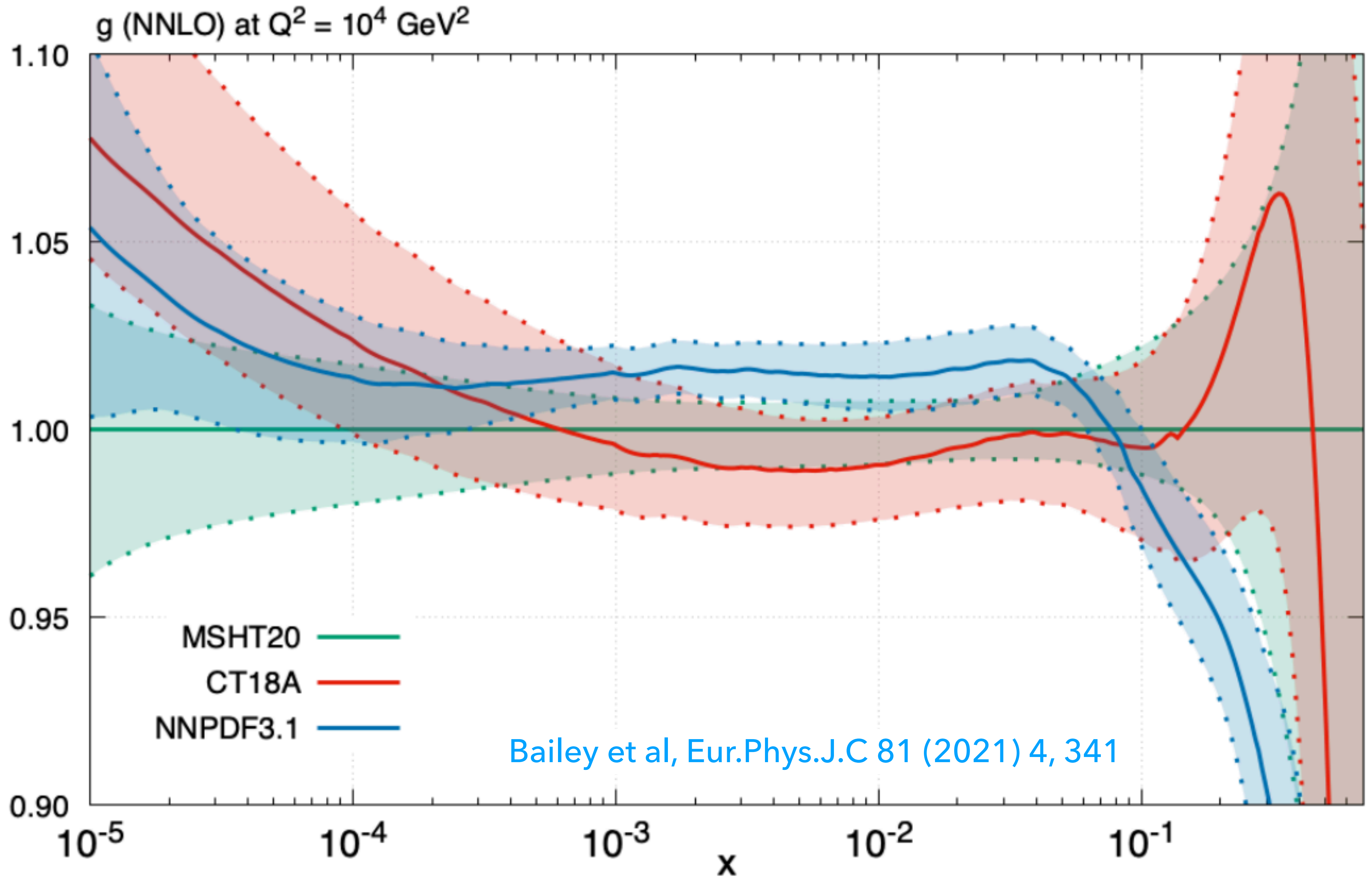


A LOT OF RECENT PROGRESS





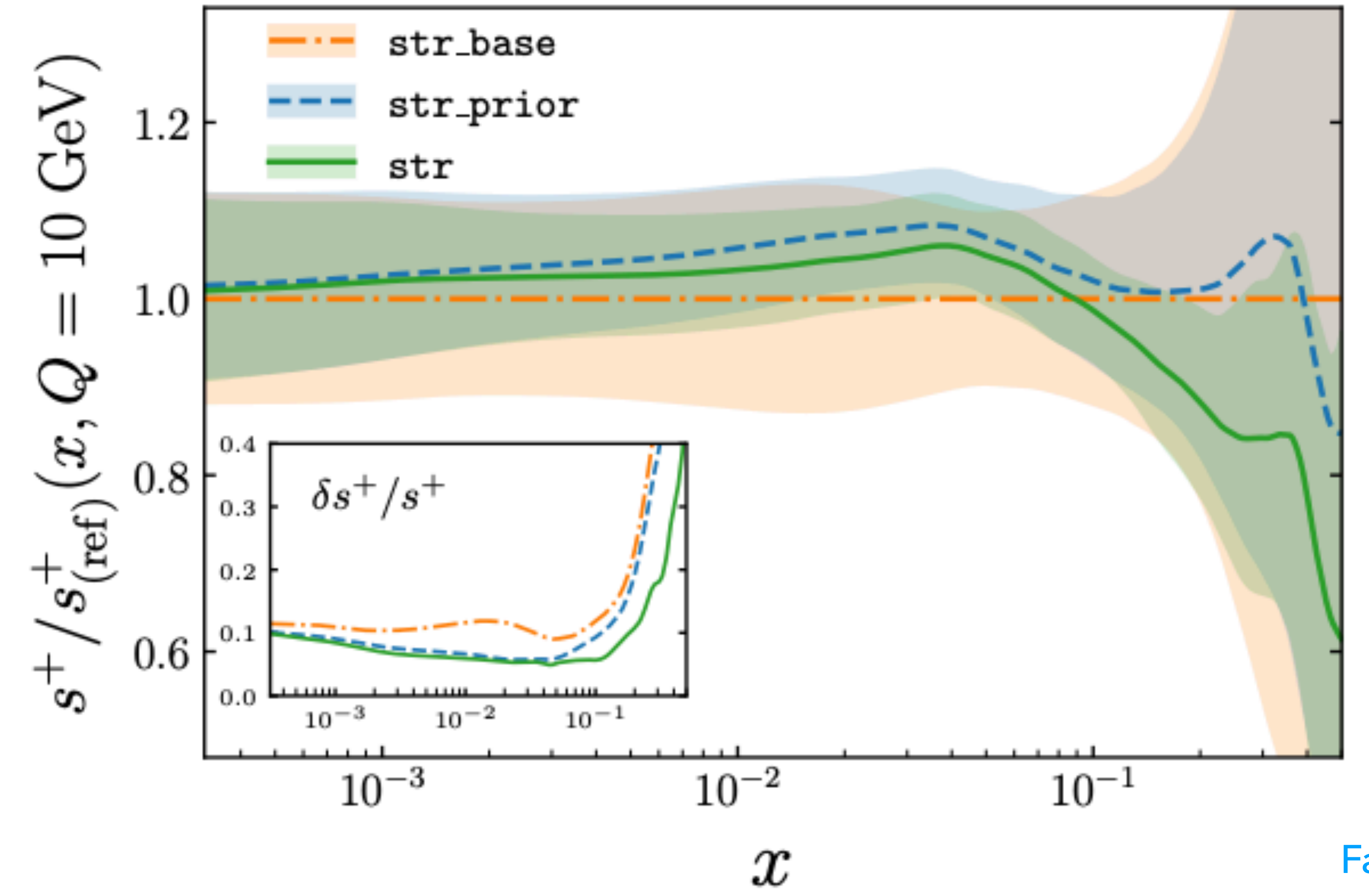




- In recent updates from global PDF fitting collaborations (CT18, MSHT20) the effect of LHC data is driving the PDF uncertainties down, although highly correlated data provide big challenge
- The increased precision of the data and their strong correlation demands methodological improvements (more flexible parametrisation, more refined statistical techniques)
- Some tension among data observed (CT18 vs CT18Z)
- Ongoing benchmark among NNPDF3.1, MSHT20 and CT18 (see J. Houston's talk on Friday)

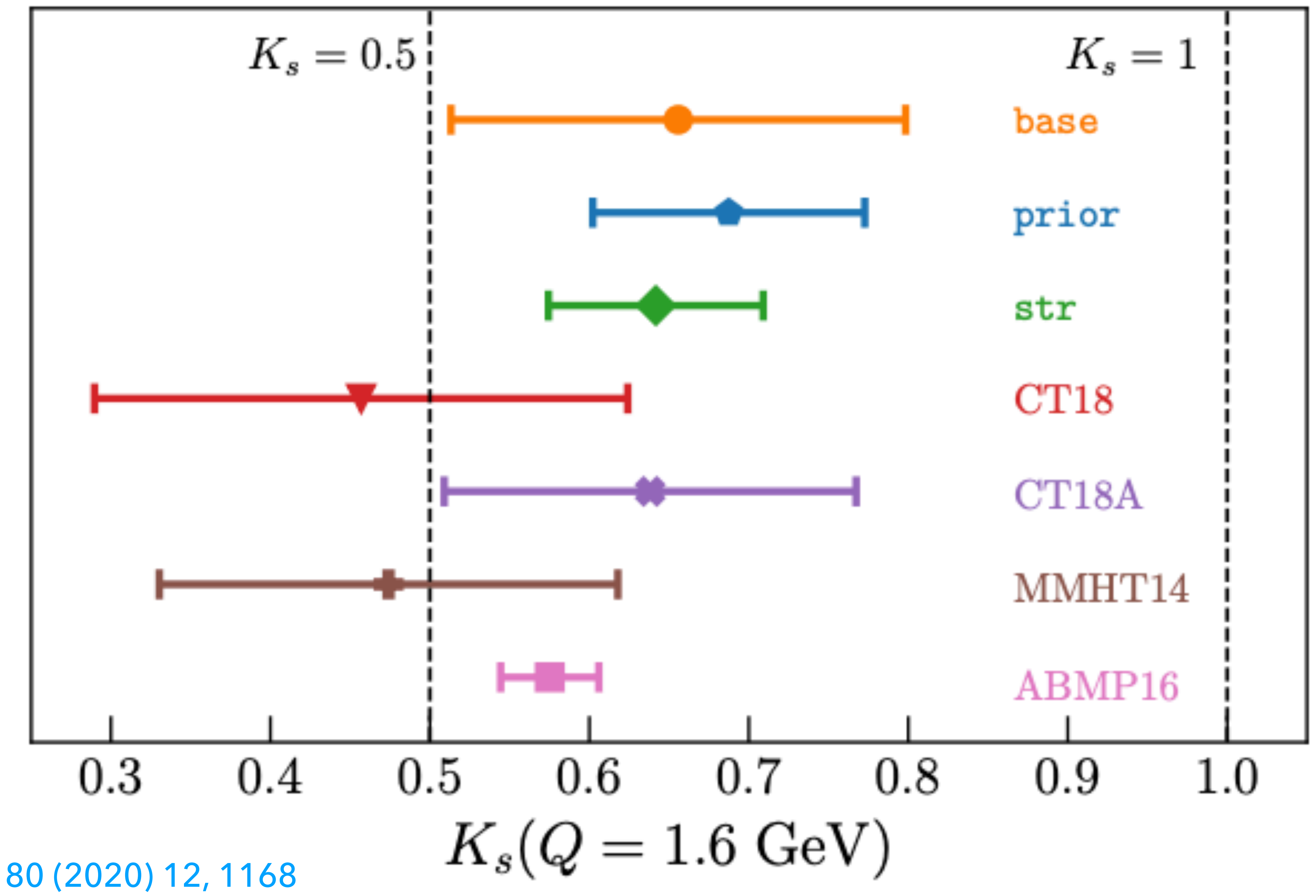
DATA COMPATIBILITY AND STRANGENESS

Process	Dataset	$n_{\rm dat}$	$\chi^2_{\rm base}$	$\chi^2_{\rm pr}$	$\chi^2_{\rm str}$
νDIS ($\mu\mu$)		76/76/95	0.70	0.71	0.53
	NuTeV	76/76/76	0.70	0.71	0.53
	NOMAD	—/—/19	[9.0]	[8.8]	0.55
W, Z (incl.)		327/418/418	1.38	1.40	1.40
	ATLAS	—/61/61	3.22	1.65	1.67
$W+c$		—/37/37	[0.76]	0.68	0.60
	CMS	—/15/15	[1.10]	0.98	0.96
	ATLAS	—/22/22	[0.53]	0.48	0.42
W+jets	ATLAS	—/32/32	[1.58]	1.18	1.18
Total		3917/4077/4096	1.17	1.17	1.17



- Long standing tension between fixed-target and LHC data in the strange sector much accommodated
- Satisfactory description of all datasets (no evidence of tension)
- Sizeable constrains from NOMAD data compatible with collider data
- Moderate suppression of strange PDFs
- Good consistency of Ks across PDF sets

$$K_s(Q^2) = \frac{\int_0^1 dx[s(x,Q^2)+\bar{s}(x,Q^2)]}{\int_0^1 dx[\bar{u}(x,Q^2)+\bar{d}(x,Q^2)]}$$



Faura, Iranipour, Nocera, Rojo, MU, Eur.Phys.J.C 80 (2020) 12, 1168

NNPDF4.0: THE ROLE OF LHC DATA

Data set	NNPDF4.0	NNPDF3.1	ABMP16	CT18	MSHT20
ATLAS W, Z 7 TeV (2010)	✓	✓	✓	✓	✓
ATLAS W, Z 7 TeV (2011)	✓	✓	✗	✓	✓
ATLAS low-mass DY 7 TeV	✓	✓	✗	✗	✗
ATLAS high-mass DY 7 TeV	✓	✓	✗	✗	✓
ATLAS W 8 TeV	✓	✗	✗	✗	✓
ATLAS DY 2D 8 TeV	✓	✗	✗	✗	✓
ATLAS high-mass DY 2D 8 TeV	✓	✗	✗	✗	✓
ATLAS $\sigma_{W,Z}$ 13 TeV	✓	✗	✓	✗	✗
ATLAS W^++jet 8 TeV	✓	✗	✗	✗	✓
ATLAS Z p_T 8 TeV	✓	✓	✗	✓	✓
ATLAS $\sigma_{tt}^{\rm tot}$ 7, 8 TeV	✓	✓	✓	✗	✗
ATLAS $\sigma_{tt}^{\rm tot}$ 13 TeV	✓	✓	✓	✗	✗
ATLAS $t\bar{t}$ lepton+jets 8 TeV	✓	✓	✗	✓	✓
ATLAS $t\bar{t}$ dilepton 8 TeV	✓	✗	✗	✗	✓
ATLAS single-inclusive jets 7 TeV, R=0.6	✗	✓	✗	✓	✓
ATLAS single-inclusive jets 8 TeV, R=0.6	✓	✗	✗	✗	✗
ATLAS dijets 7 TeV, R=0.6	✓	✗	✗	✗	✗
ATLAS direct photon production 13 TeV	✓	✗	✗	✗	✗
ATLAS single top R_t 7, 8, 13 TeV	✓	✗	✓	✗	✗
ATLAS single top diff. 7, 8 TeV	✓	✗	✗	✗	✗
ATLAS single top diff. 8 TeV	✓	✗	✗	✗	✗

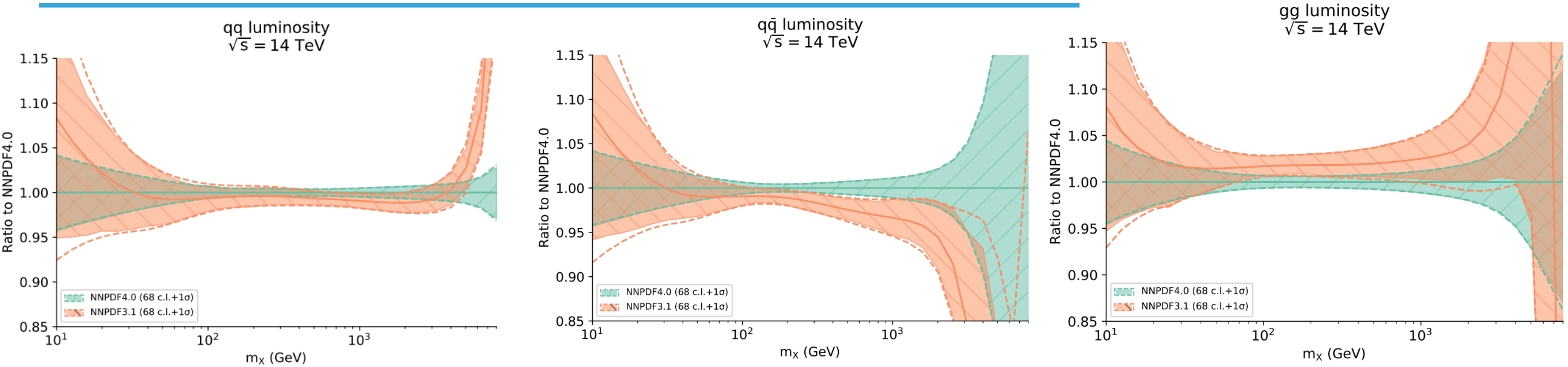
Data set	NNPDF4.0	NNPDF3.1	ABMP16	CT18	MSHT20
LHCb Z 940 pb	✓	✓	✗	✗	✓
LHCb $Z \to ee$ 2 fb	✓	✓	✓	✓	✓
LHCb $W, Z \to \mu$ 7 TeV	✓	✓	✓	✓	✓
LHCb $W, Z \to \mu$ 8 TeV	✓	✓	✓	✓	✓
LHCb $Z \to \mu\mu, ee$ 13 TeV	✓	✗	✗	✗	✗

Data set	NNPDF4.0	NNPDF3.1	ABMP16	CT18	MSHT20
CMS W electron asymmetry 7 TeV	✓	✓	✗	✓	✓
CMS W muon asymmetry 7 TeV	✓	✓	✓	✓	✗
CMS Drell-Yan 2D 7 TeV	✓	✓	✗	✗	✓
CMS W rapidity 8 TeV	✓	✓	✓	✓	✓
CMS Z p_T 8 TeV	✓	✓	✗	✓	✗
CMS $W + c$ 7 TeV	✓	✓	✗	✗	✓
CMS $W + c$ 13 TeV	✓	✗	✗	✗	✗
CMS single-inclusive jets 2.76 TeV	✗	✓	✗	✗	✓
CMS single-inclusive jets 7 TeV	✗	✓	✗	✓	✓
CMS dijets 7 TeV	✓	✗	✗	✗	✗
CMS single-inclusive jets 8 TeV	✗	✗	✗	✓	✓
CMS 3D dijets 8 TeV	✓	✗	✗	✗	✗
CMS $\sigma_{tt}^{\rm tot}$ 5 TeV	✓	✗	✓	✗	✗
CMS $\sigma_{tt}^{\rm tot}$ 7, 8 TeV	✓	✓	✓	✗	✓
CMS $\sigma_{tt}^{\rm tot}$ 13 TeV	✓	✓	✓	✗	✗
CMS $t\bar{t}$ lepton+jets 8 TeV	✓	✓	✗	✗	✓
CMS $t\bar{t}$ 2D dilepton 8 TeV	✓	✗	✗	✓	✓
CMS $t\bar{t}$ lepton+jet 13 TeV	✓	✗	✗	✗	✗
CMS $t\bar{t}$ dilepton 13 TeV	✓	✗	✗	✗	✗
CMS single top $\sigma_t + \sigma_{\bar{t}}$ 7 TeV	✓	✗	✓	✗	✗
CMS single top R_t 8, 13 TeV	✓	✗	✓	✗	✗

- NNPDF4.0 is upcoming new set from NNPDF based on very large set of data from LHC Run I and Run II (see T.Giani's talk on Friday)
- O(5000) data points and first time inclusion of di-jets preferred over inclusive jets based on [Khalek et al, Eur.Phys.J.C 80 (2020) 8, 797]

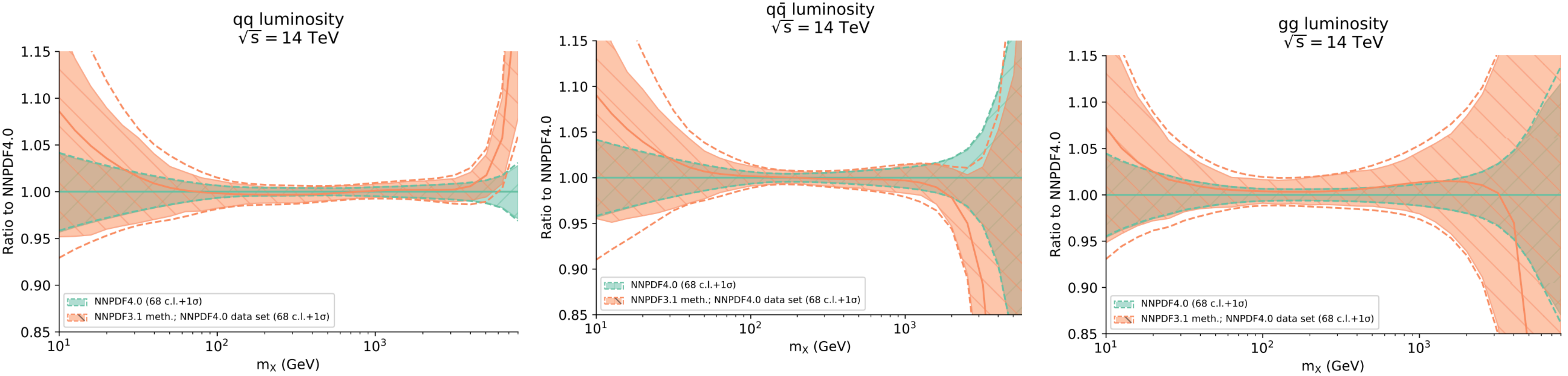
NNPDF4.0: THE ROLE OF METHODOLOGY

NNPDF3.1 vs NNPDF4.0



DATA \ METHODOLOGY	NNPDF3.1	NNPDF4.0
NNPDF3.1 (4093)	1.19	1.12
NNPDF4.0 (4491)	1.25	1.17

NNPDF4.0 data - NNPDF3.1 vs NNPDF4.0 methodology



DATA \ METHODOLOGY	NNPDF3.1	NNPDF4.0
NNPDF3.1 (4093)	1.19	1.12
NNPDF4.0 (4491)	1.25	1.17

- ➡ Shift in parton luminosities mostly due to inclusion of O(500) more data points
- ➡ Parton luminosities based on same dataset are consistent with each other but 4.0 methodology displays smaller uncertainty than 3.1 methodology: NNPDF4.0 more accurate and superior to previous methodology

THEORY FRONTIERS

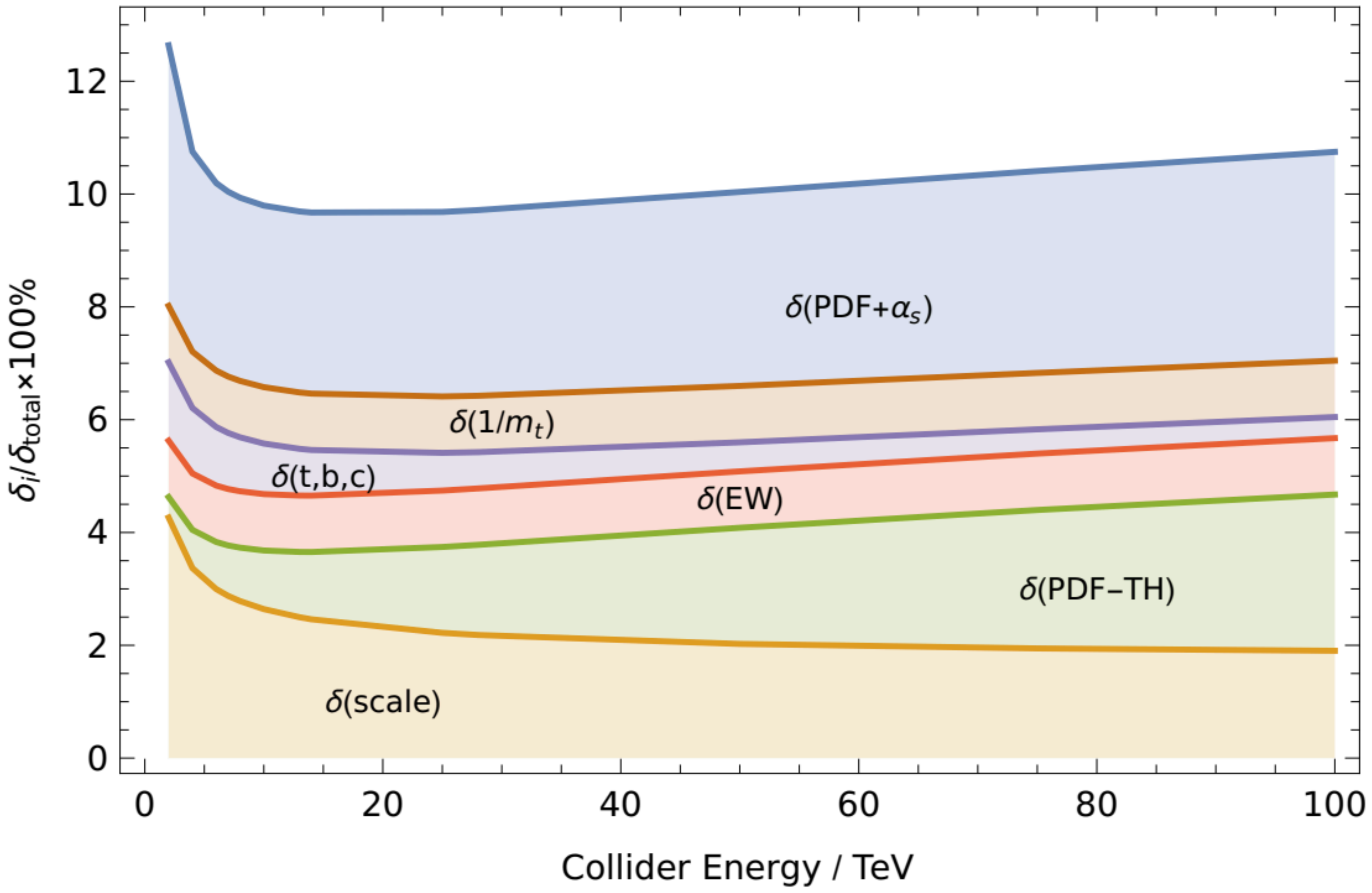
THEORY UNCERTAINTIES IN PDF FITS

$$\sigma = \alpha_s^p \sigma_0 + \alpha_s^{p+1}\sigma_1 + \alpha_s^{p+2}\sigma_2 + \mathcal{O}(\alpha_s^{p+3})$$

- Standard global PDF fits based on fixed-order NNLO QCD calculations (using fast interpolation grid for NLO predictions accompanied by local K-factors for NNLO). PDF uncertainty reflects experimental uncertainty.
- N3LO is now the precision frontier for partonic cross sections (N3LO splitting functions partially known in Nc→∞ limit)
- Mismatch between perturbative order of partonic cross section and PDFs becoming significant source of uncertainty

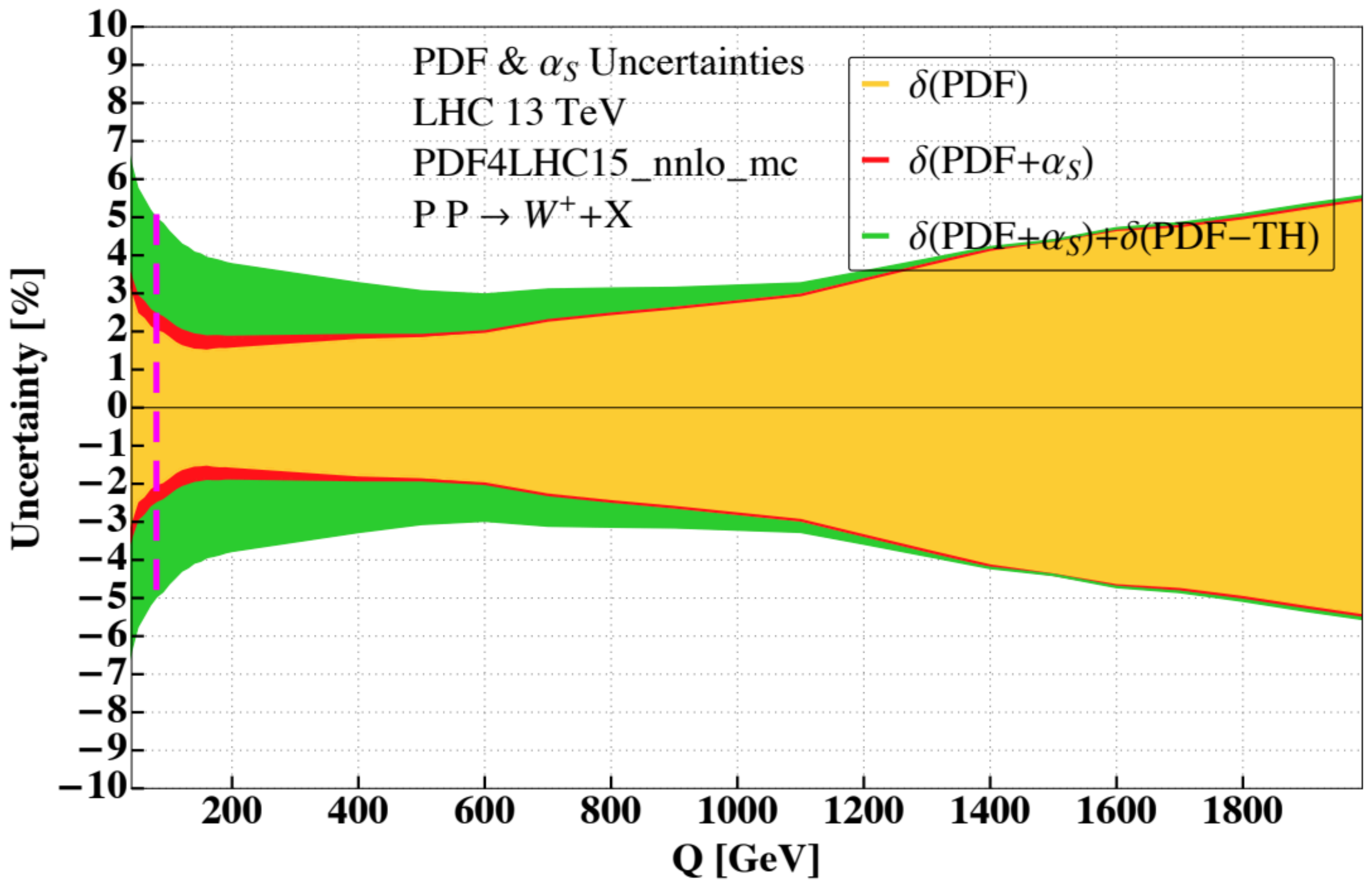
$$\delta(PDF-TH) = \frac{1}{2}\left|\frac{\sigma^{(2)}_{\text{NNLO-PDFs}} - \sigma^{(2)}_{\text{NLO-PDFs}}}{\sigma^{(2)}_{\text{NNLO-PDFs}}}\right|$$

Gluon-gluon fusion into Higgs



M. Cepeda et al. [HL/HE WG2 group], arXiv:1902.00134

Drell-Yan



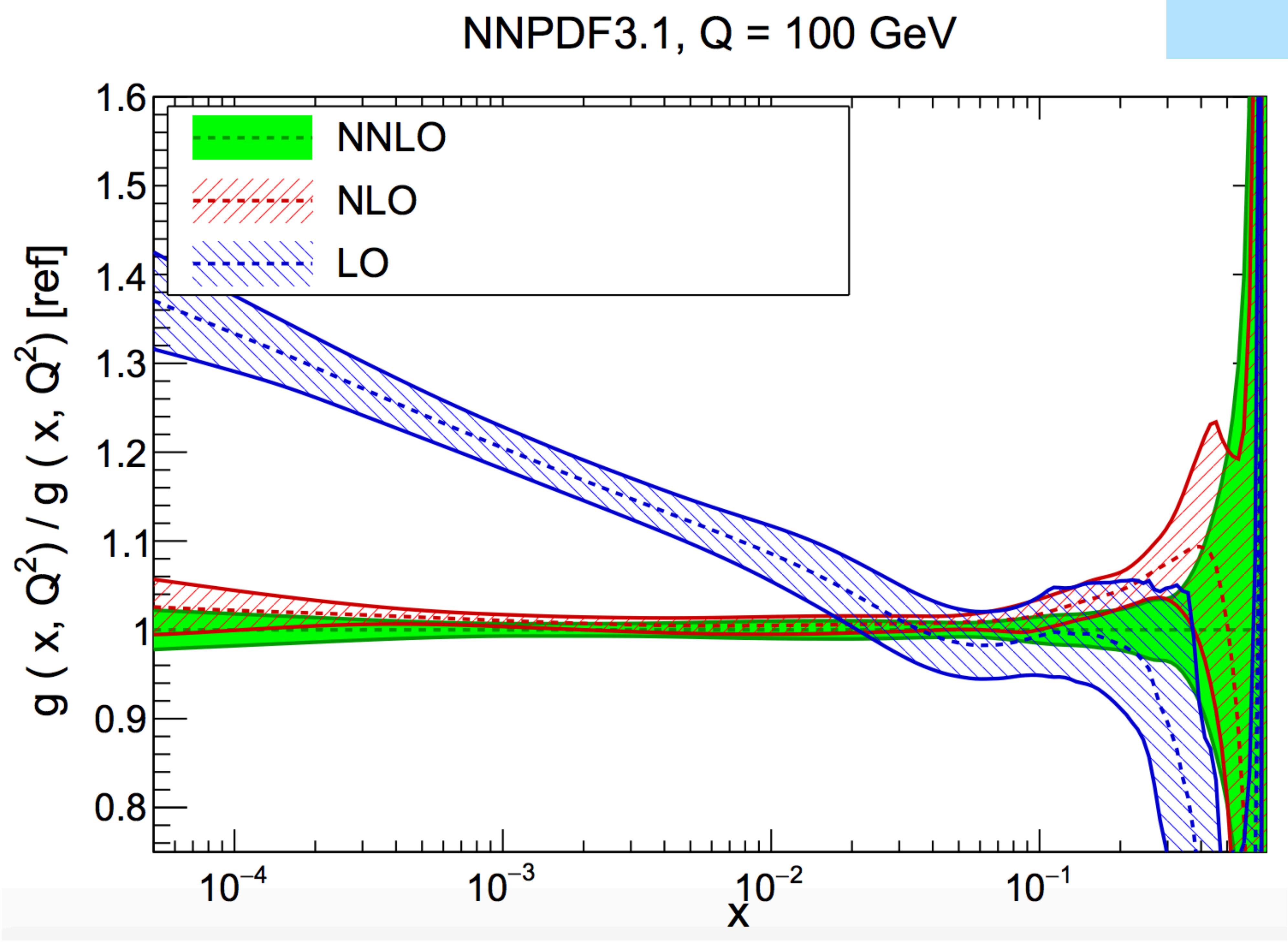
C. Duhr et al, *JHEP* 11 (2020) 143

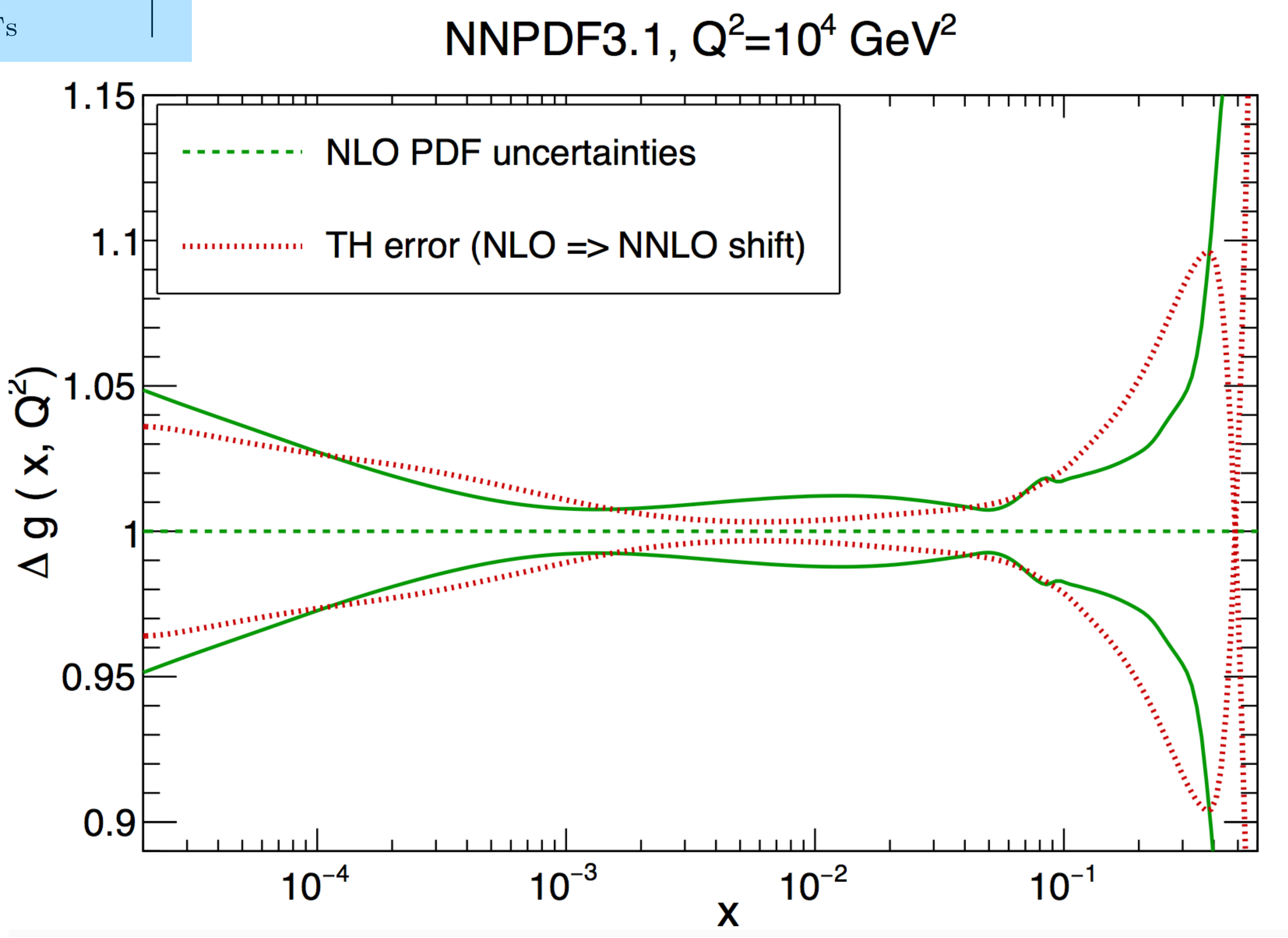
THEORY UNCERTAINTIES IN PDF FITS

$$\sigma = \alpha_s^p \sigma_0 + \alpha_s^{p+1}\sigma_1 + \alpha_s^{p+2}\sigma_2 + \mathcal{O}(\alpha_s^{p+3})$$

- Standard global PDF fits based on fixed-order NNLO QCD calculations (using fast interpolation grid for NLO predictions accompanied by local K-factors for NNLO). PDF uncertainty reflects experimental uncertainty.
- N3LO is now the precision frontier for partonic cross sections (N3LO splitting functions partially known in Nc→∞ limit)
- Mismatch between perturbative order of partonic cross section and PDFs becoming significant source of uncertainty

$$\delta(PDF-TH) = \frac{1}{2}\left|\frac{\sigma^{(2)}_{\rm NNLO-PDFs} - \sigma^{(2)}_{\rm NLO-PDFs}}{\sigma^{(2)}_{\rm NNLO-PDFs}}\right|$$





R.D. Ball et al, EPJ C77 (2017) 663

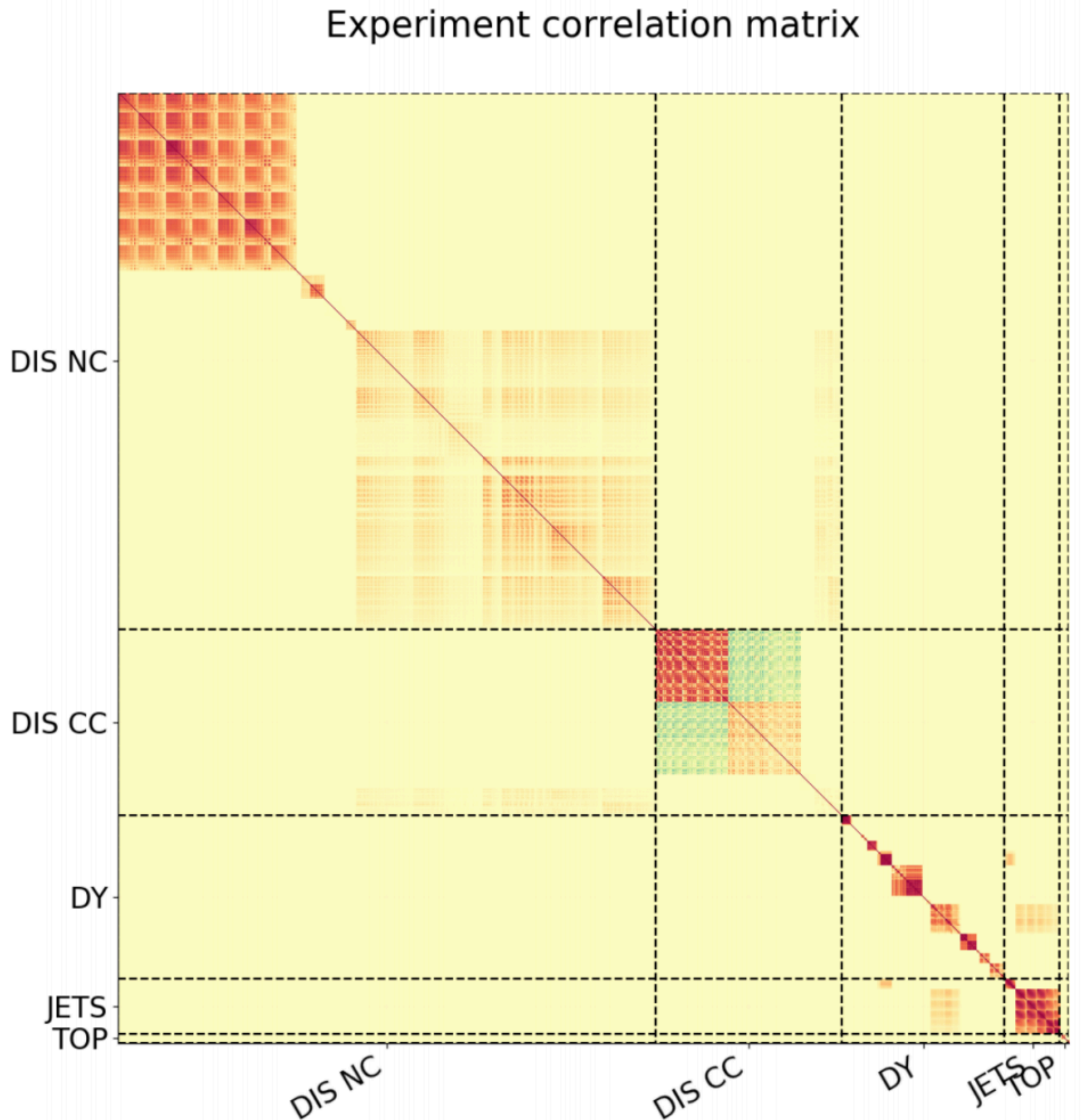
THEORY UNCERTAINTIES IN PDF FITS

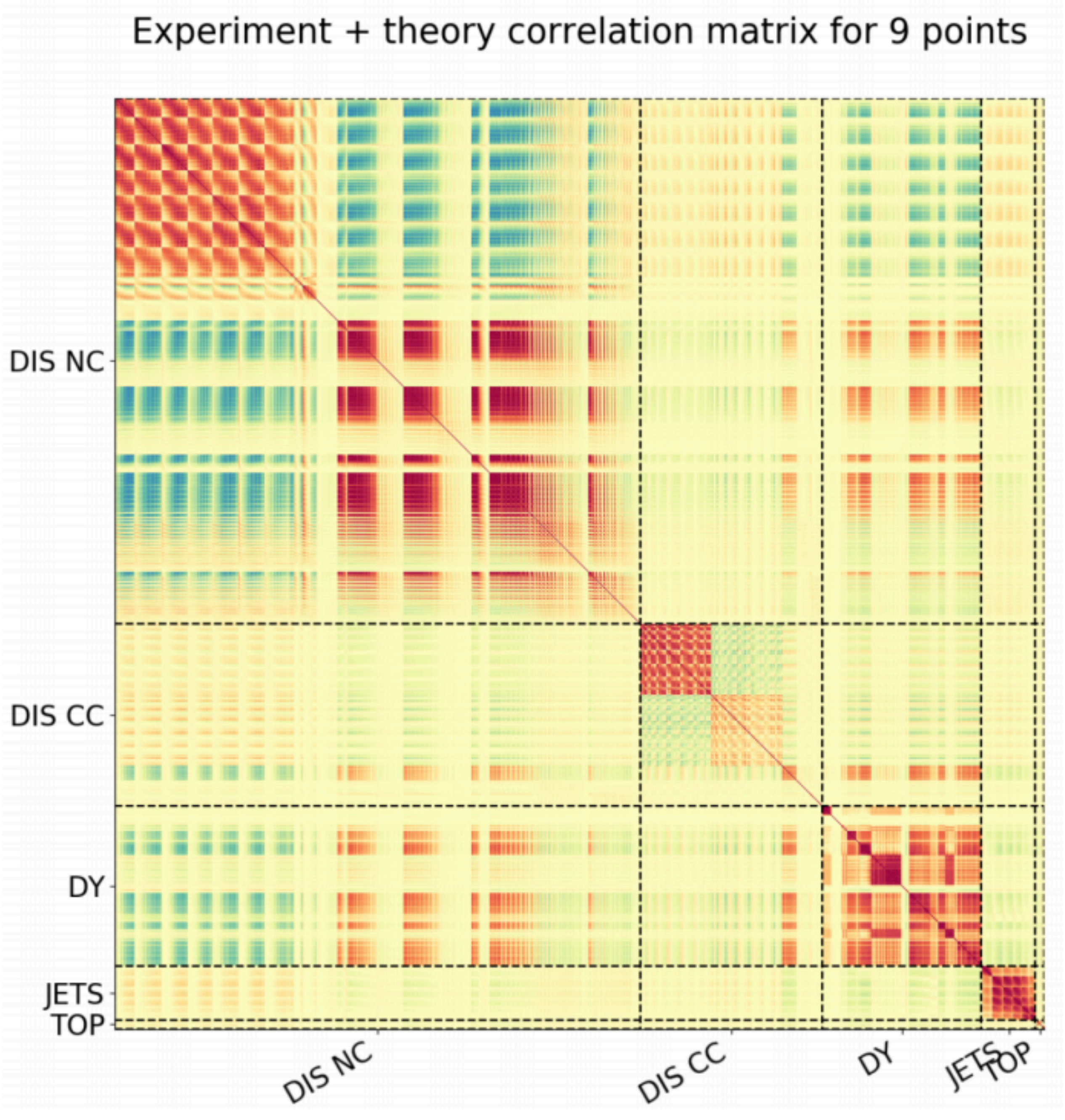
- ➡ How to include Missing Higher Order Uncertainties in a PDF fit? So far tested only at NLO but work in progress for NNLO
- ➡ Construct a theory covariance matrix from scale-varied cross sections and combine it with the experimental covariance matrix

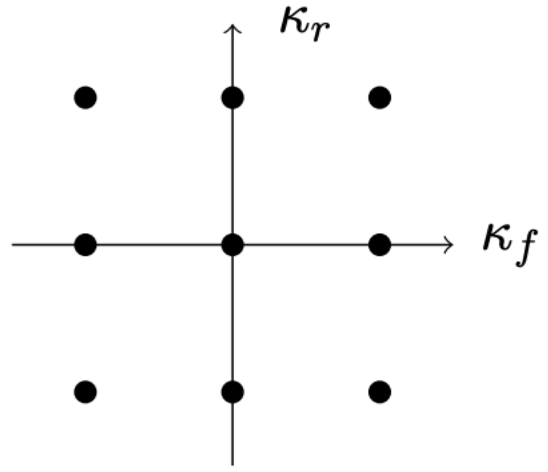
[R.D. Ball et al, Eur.Phys.J.C 79 (2019) 11, 931, Eur.Phys.J. C (2019) 79:838]

$$\chi^2 = \sum_{m,n=1}^{N} (d_m - t_m)(\mathrm{cov}_{\mathrm{exp}} + \mathrm{cov}_{\mathrm{th}})^{-1}_{mn}(d_n - t_n)$$

- ➡ Assumptions: experimental and theoretical errors independent and Gaussian
- ➡ Accounting for the theory covariance matrix will modify the relative weight that each of the datasets carries in the global fit: processes with higher MHOUs will be ``de-weighted''
- ➡ Assumptions on correlation of scales and scale ratio will determine the specific form of the covariance matrix



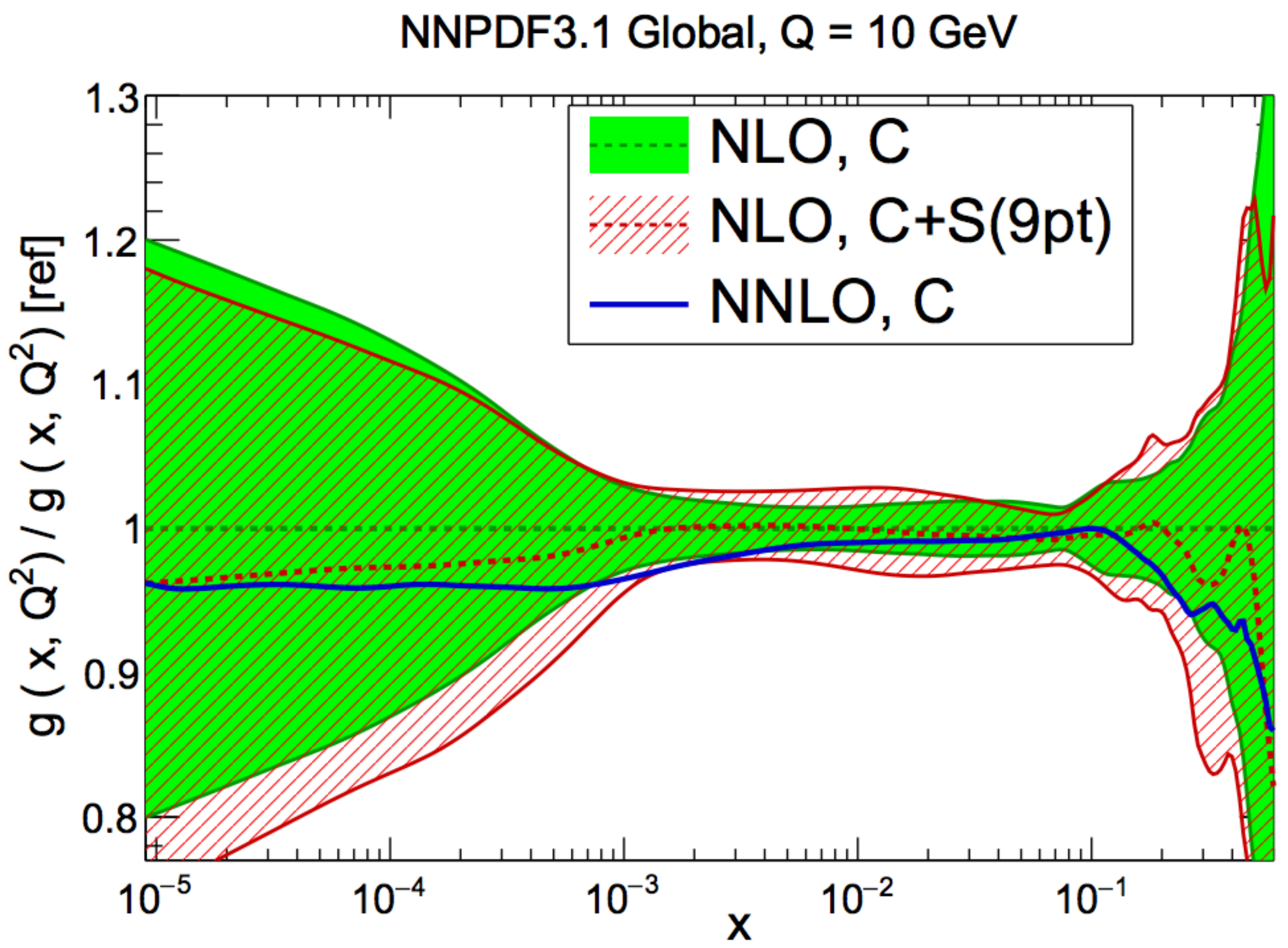


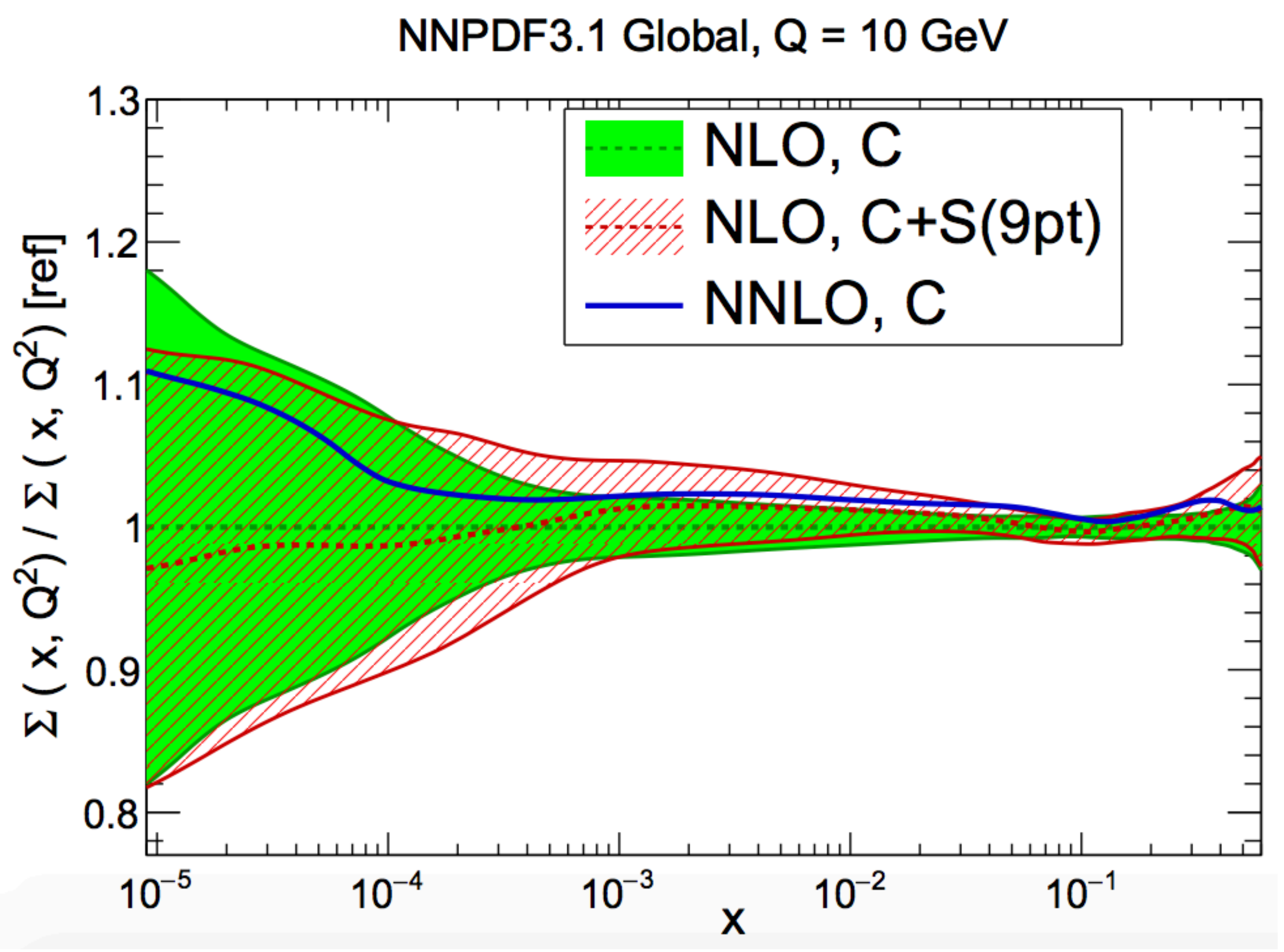


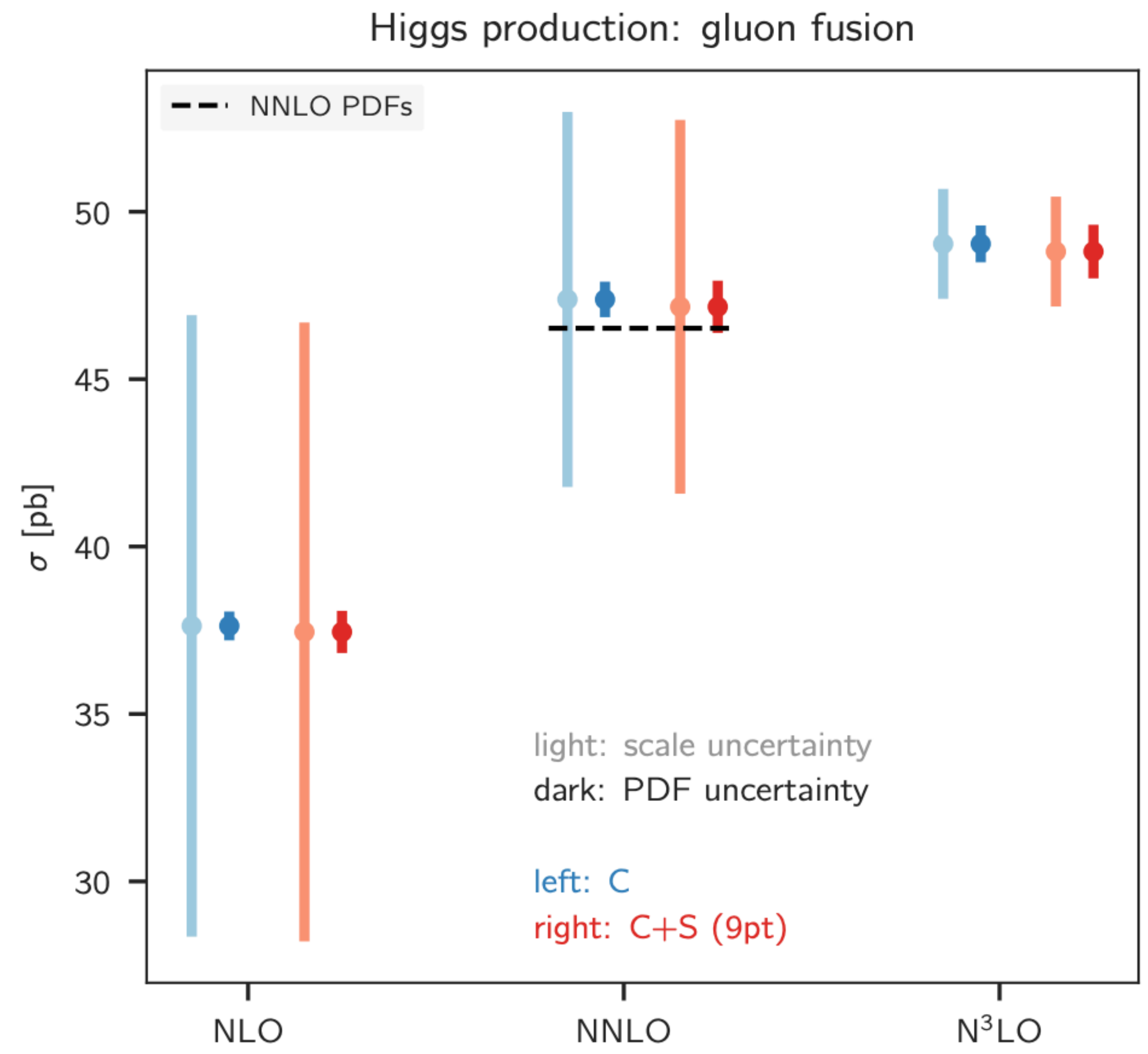
9-pts variation

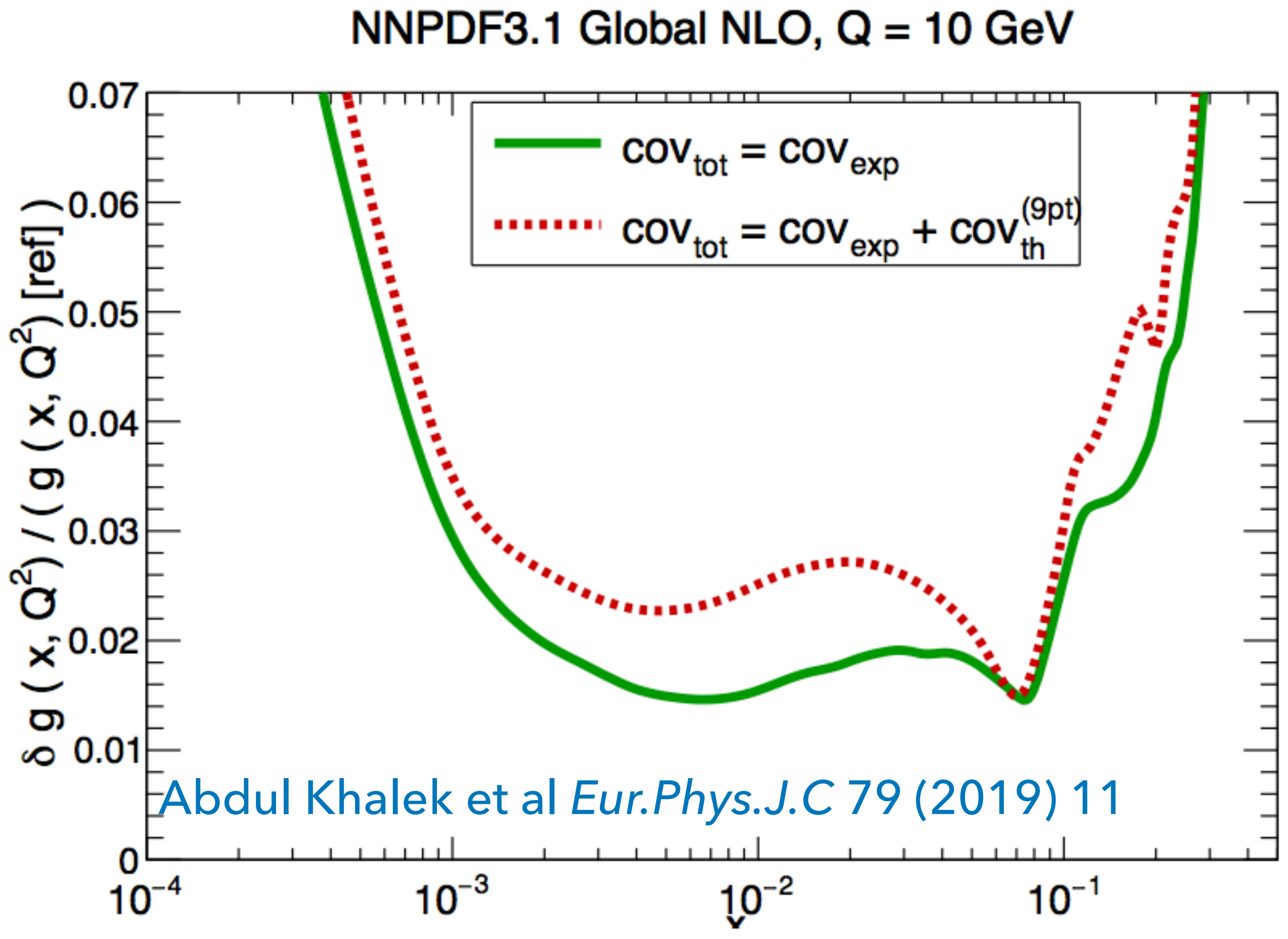
Abdul Khalek et al
Eur.Phys.J.C 79 (2019) 11

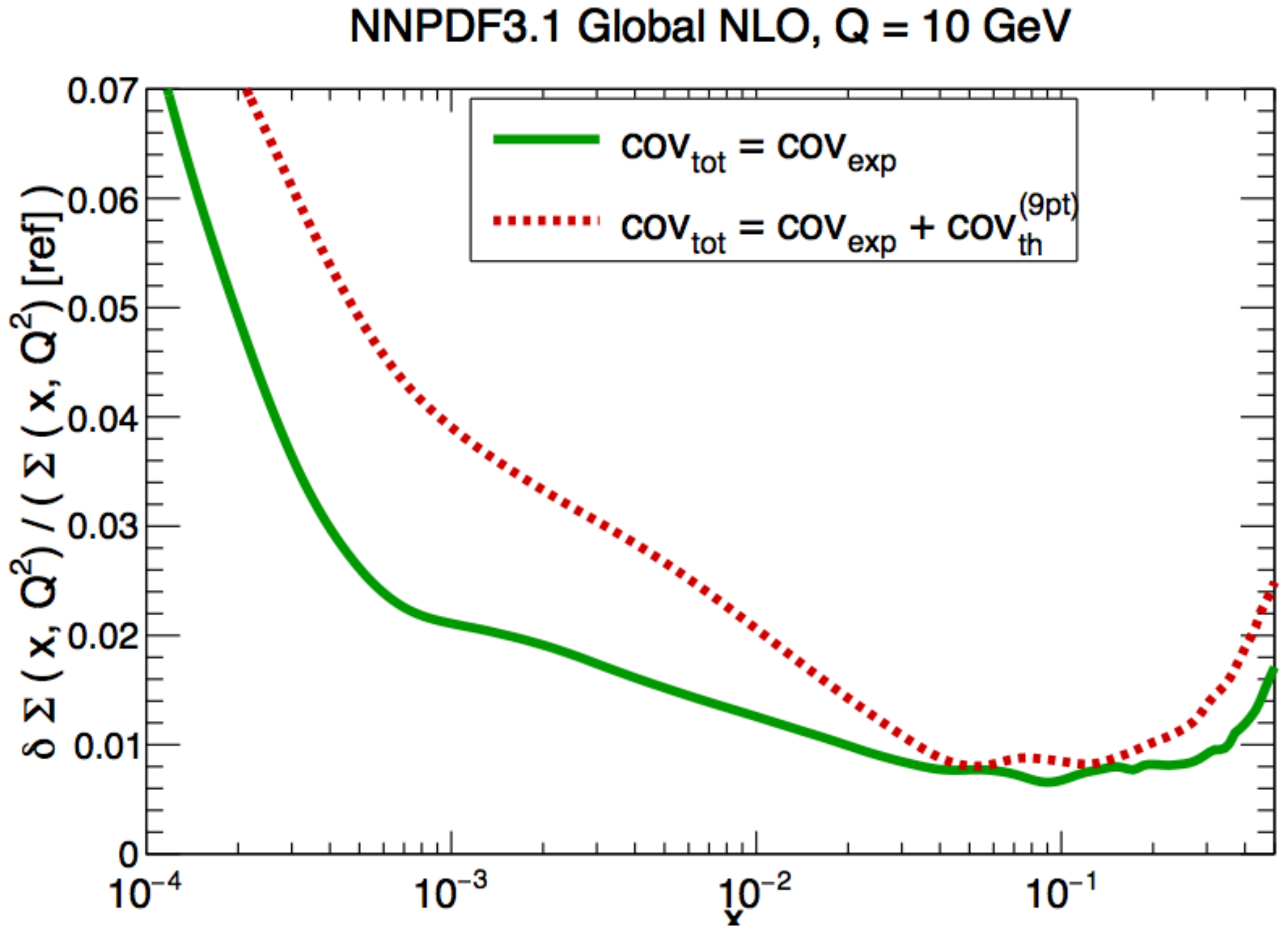
THEORY UNCERTAINTIES IN PDF FITS

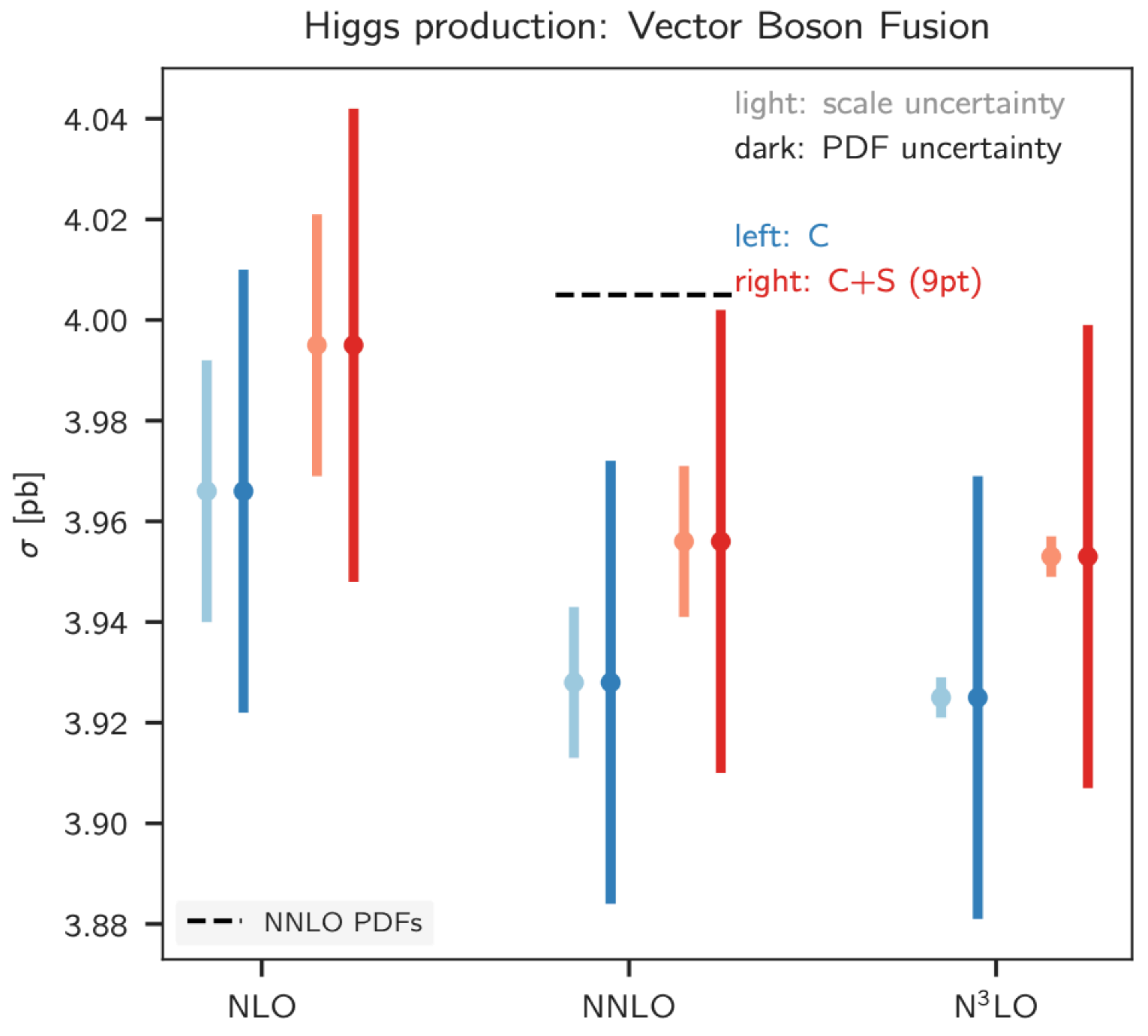












➡ Missing correlation between scale variation in PDF fits and hard cross sections

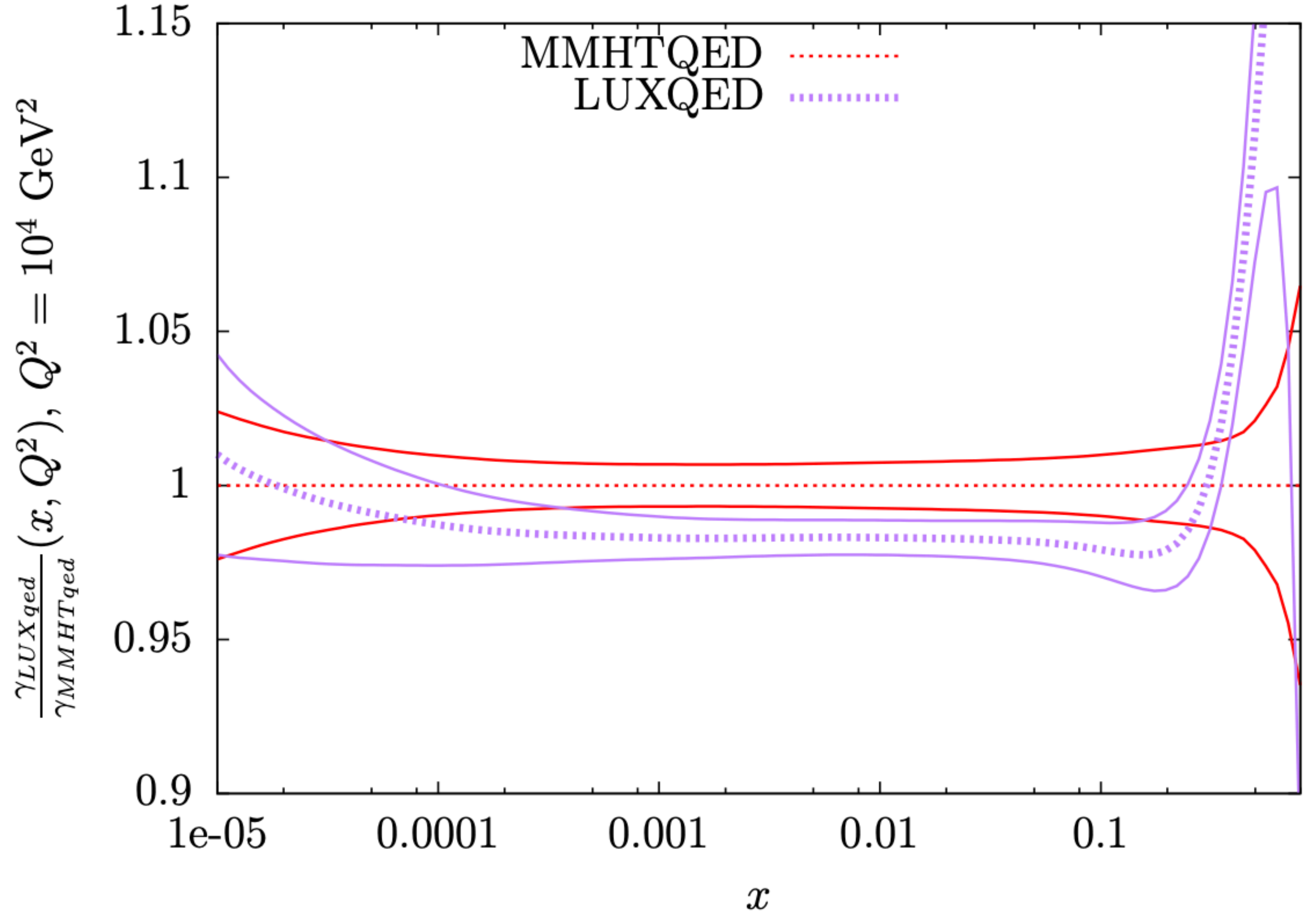
[Harland-Lang, Thorne Eur.Phys.J.C 79 (2019) 3, 225]

NLO EW CORRECTIONS FOR PDF FITS

- If we aim to 1% accuracy NLO EW corrections do matter.
- As the LHC reaches higher invariant mass regions and pT regions EW corrections increasingly relevant
- Currently NLO EW corrections included for some observables in MSHT20 (via K-factors) [Bailey et al 2012.04684]
- NLO EW corrections not included by default in NNPDF4.0 nor CT18 (large invariant mass regions, large Z pT excluded => less constraints to large-x PDFs)

What is needed

- ➡ QED corrections in DGLAP evolution [Bertone et al, APFEL, Comput.Phys.Commun. 185 (2014)]
- ➡ Photon PDFs [Manohar et al Phys.Rev.Lett. 117 (2016), Bertone et al SciPost Phys. 5 (2018) 1, 008, Harland-Lang JHEP 79 (2019) 10]
- ➡ NLO EW corrections for all processes included in PDF fits [Frederix et al,JHEP 07 (2018) 185]



Harland-Lang JHEP 79 (2019) 10

NLO EW CORRECTIONS FOR PDF FITS

- If we aim to 1% accuracy NLO EW corrections do matter.
- As the LHC reaches higher invariant mass regions and pT regions EW corrections increasingly relevant
- Currently NLO EW corrections included for some observables in MSHT20 (via K-factors) [Bailey et al 2012.04684]
- NLO EW corrections not included by default in NNPDF4.0 nor CT18 (large invariant mass regions, large Z pT excluded => less constraints to large-x PDFs)

What is needed

➡ QED corrections in DGLAP evolution [Bertone et al, APFEL, Comput.Phys.Commun. 185 (2014)]

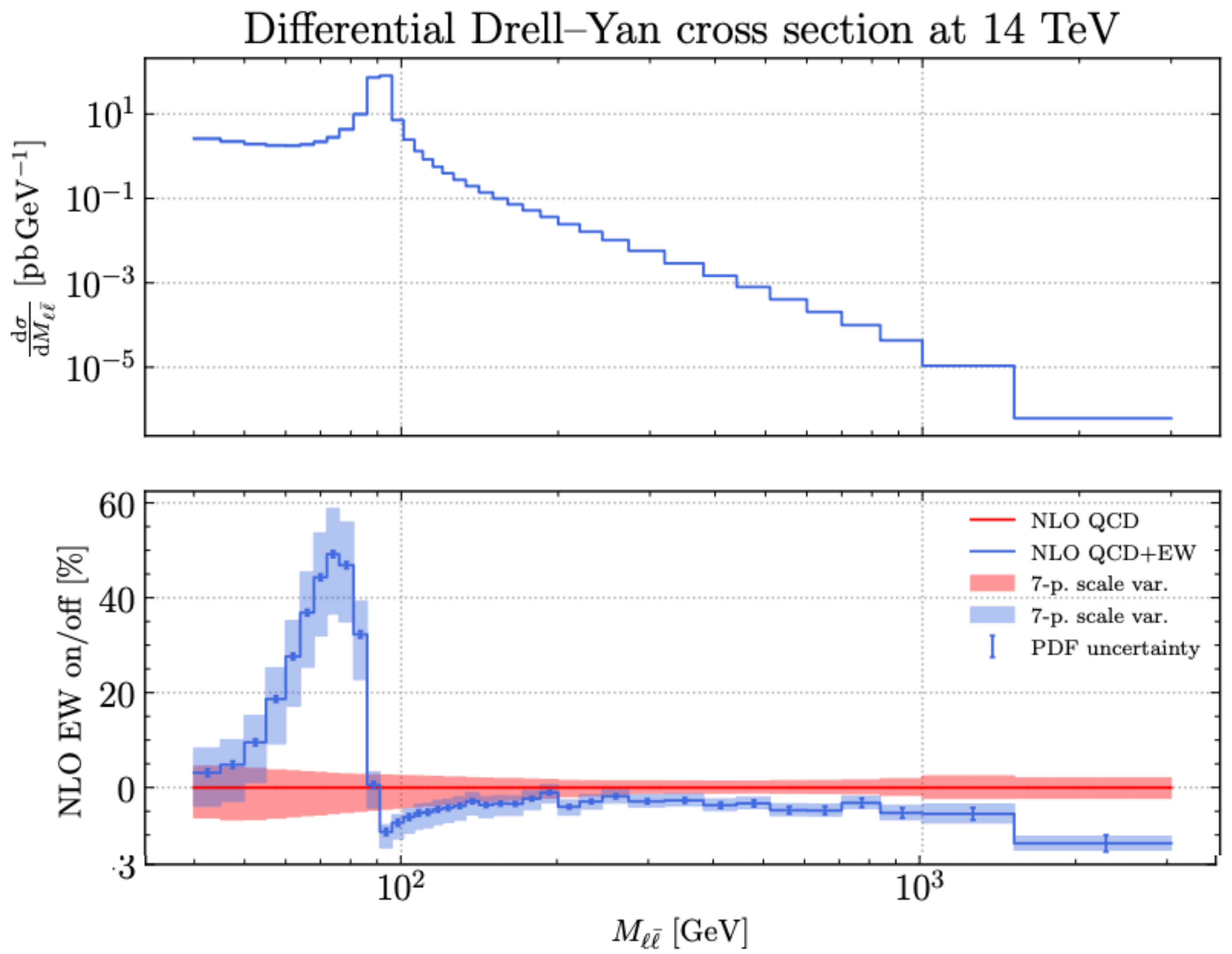
➡ Photon PDFs [Manohar et al Phys.Rev.Lett. 117 (2016), Bertone et al SciPost Phys. 5 (2018) 1, 008, Harland-Lang JHEP 79 (2019) 10]

➡ NLO EW corrections for all processes included in PDF fits [Frederix et al,JHEP 07 (2018) 185]

➡ Fast interpolating grids: **PineAPPL** [Carrazza, Nocera, Schwan, Zaro JHEP 12 (2020) 108]

➡ Careful scrutiny of data (no FSR nor photon-initiated subtraction & observable definition in experimental data)

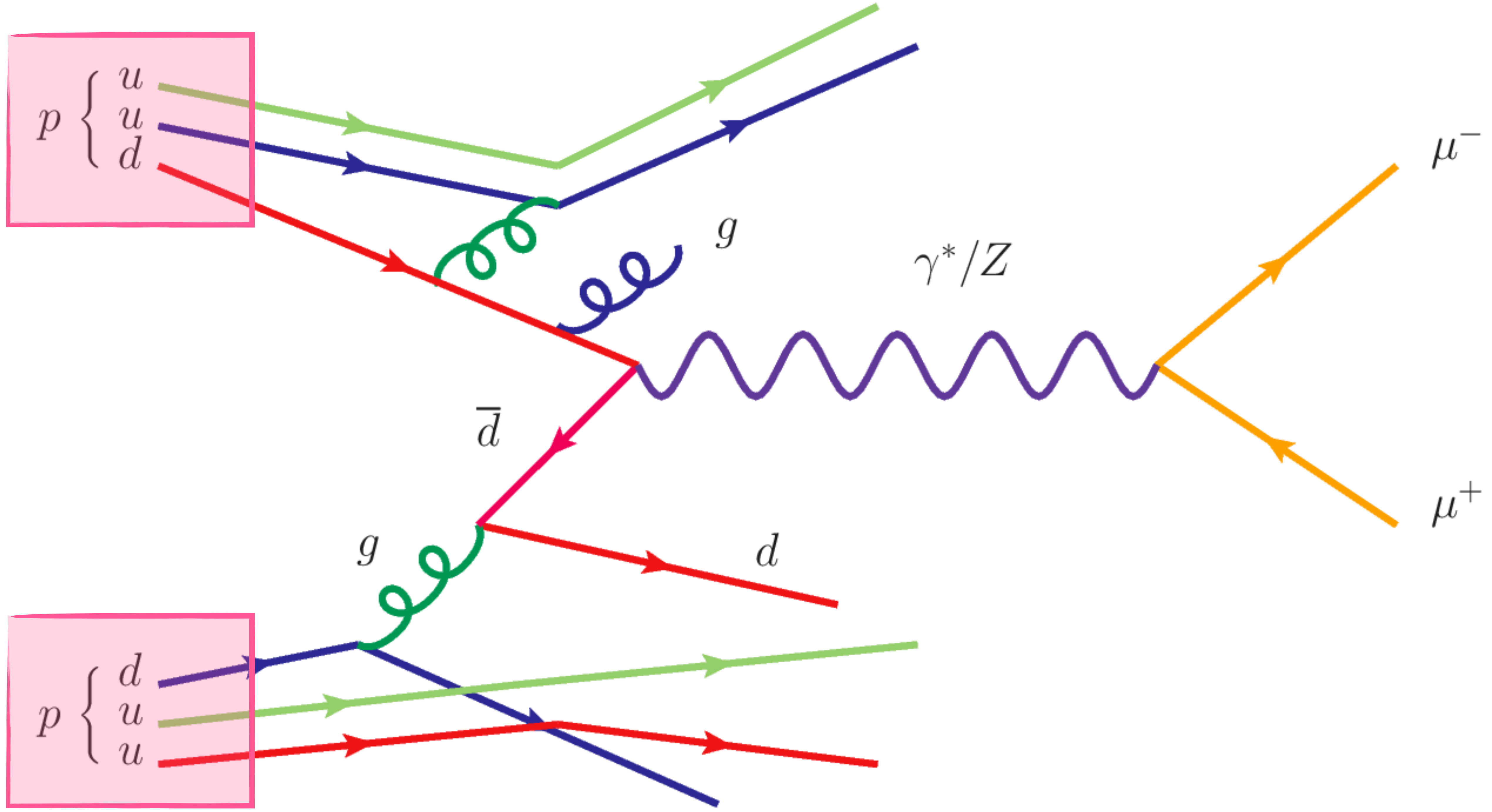
(see C. Schwan's talk on Friday)



Carrazza, Nocera, Schwan, Zaro JHEP 12 (2020) 108

PDFS AND NEW PHYSICS INTERPLAY IN HIGH ENERGY TAILS

PDF AND SMEFT FIT INTERPLAY



$$d\sigma^{pp\to ab} = \sum_{i,j} f_i \otimes f_j \otimes d\hat{\sigma}^{ij\to ab} + \ldots$$

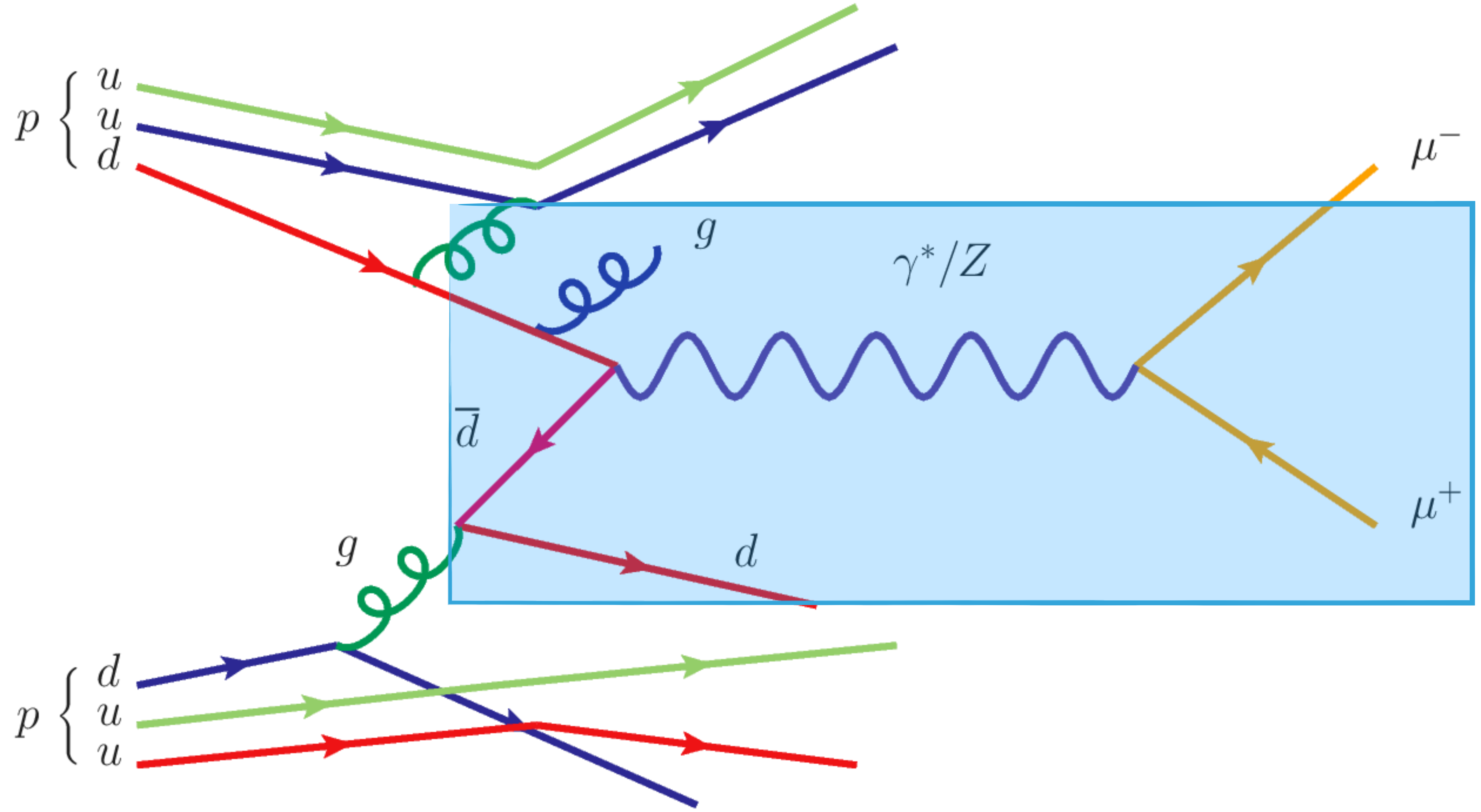
Fit PDF parameters $\{\theta_l\}$

LHC Data

$f_i(\{\theta_k\})$

Assume SM in theory predictions

PDF AND SMEFT FIT INTERPLAY



$$d\sigma^{pp\to ab} = \sum_{i,j} f_i \otimes f_j \otimes d\hat{\sigma}^{ij\to ab} + \ldots$$

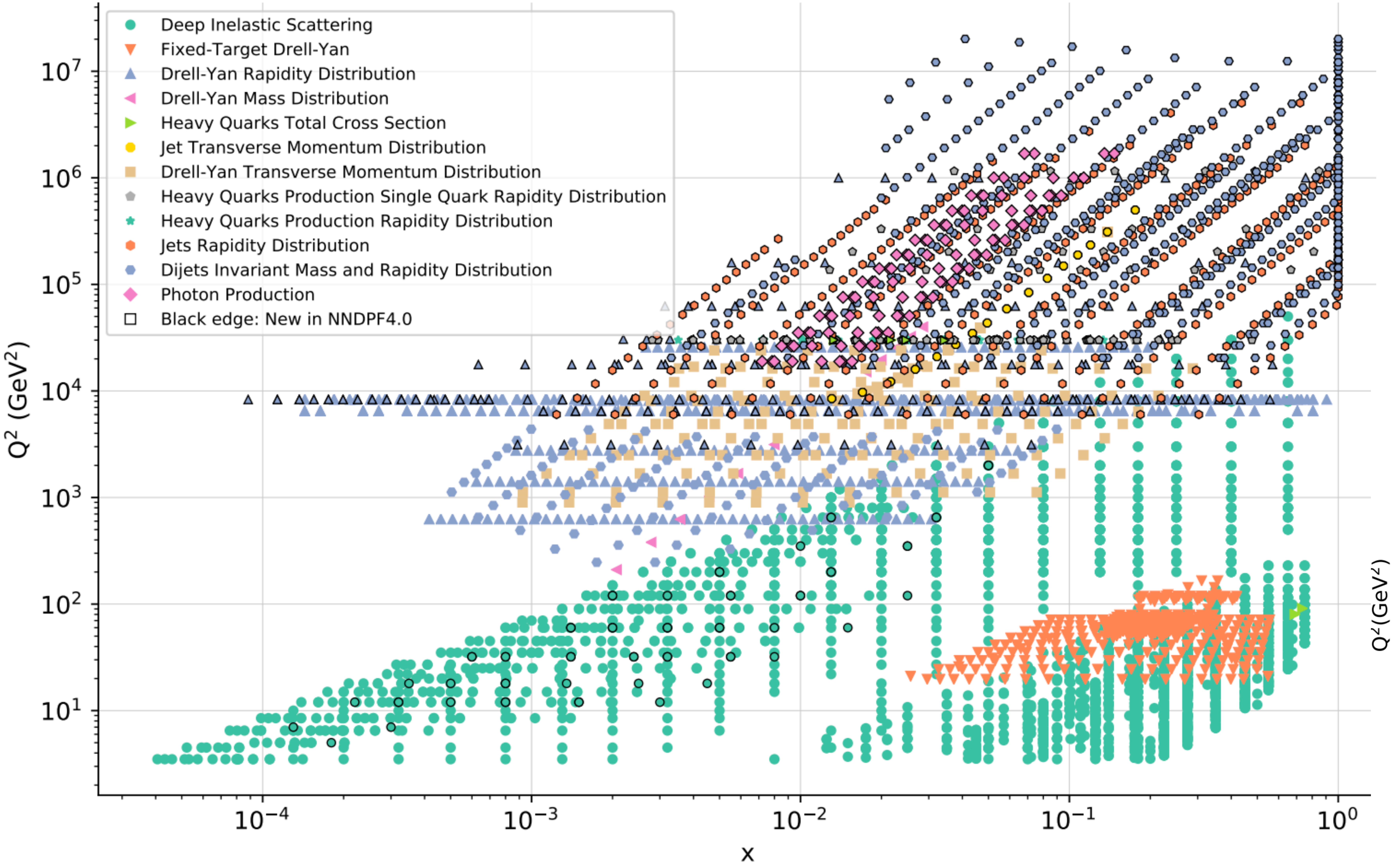
Fit of Wilson Coefficients $\{c_i\}$

LHC Data

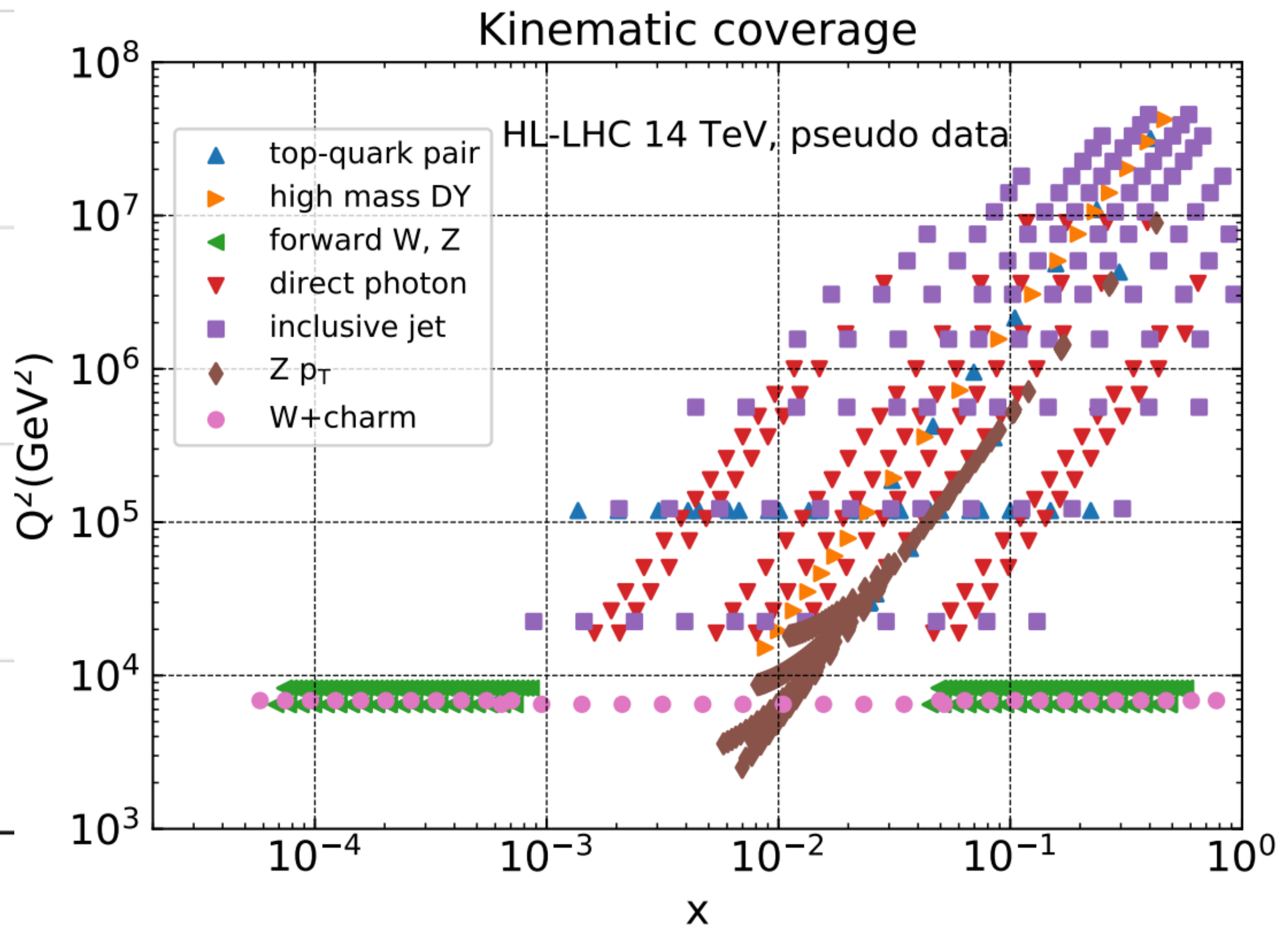
Assume SM PDFs

$$\mathcal{L}_{\mathrm{SMEFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_i \frac{c_i}{\Lambda^2}\mathcal{O}_i^{(6)} + \cdots$$

KINEMATIC OVERLAP



- ➡ Top pair production and single top data included in [Hartland et al 1901.05965]
- ➡ Dijets data in [Bordone et al 2103.10332] [Alioli et al 1706.03068]
- ➡ Drell-Yan data in [Farina et al 1609.08157, Torre et al 2008.12978]
- ➡ Inclusive jets in [Alte et al 1711.07484]
- ➡ Overlap enhanced in HL-LHC projections [Abdul Khalek et al1810.03639]

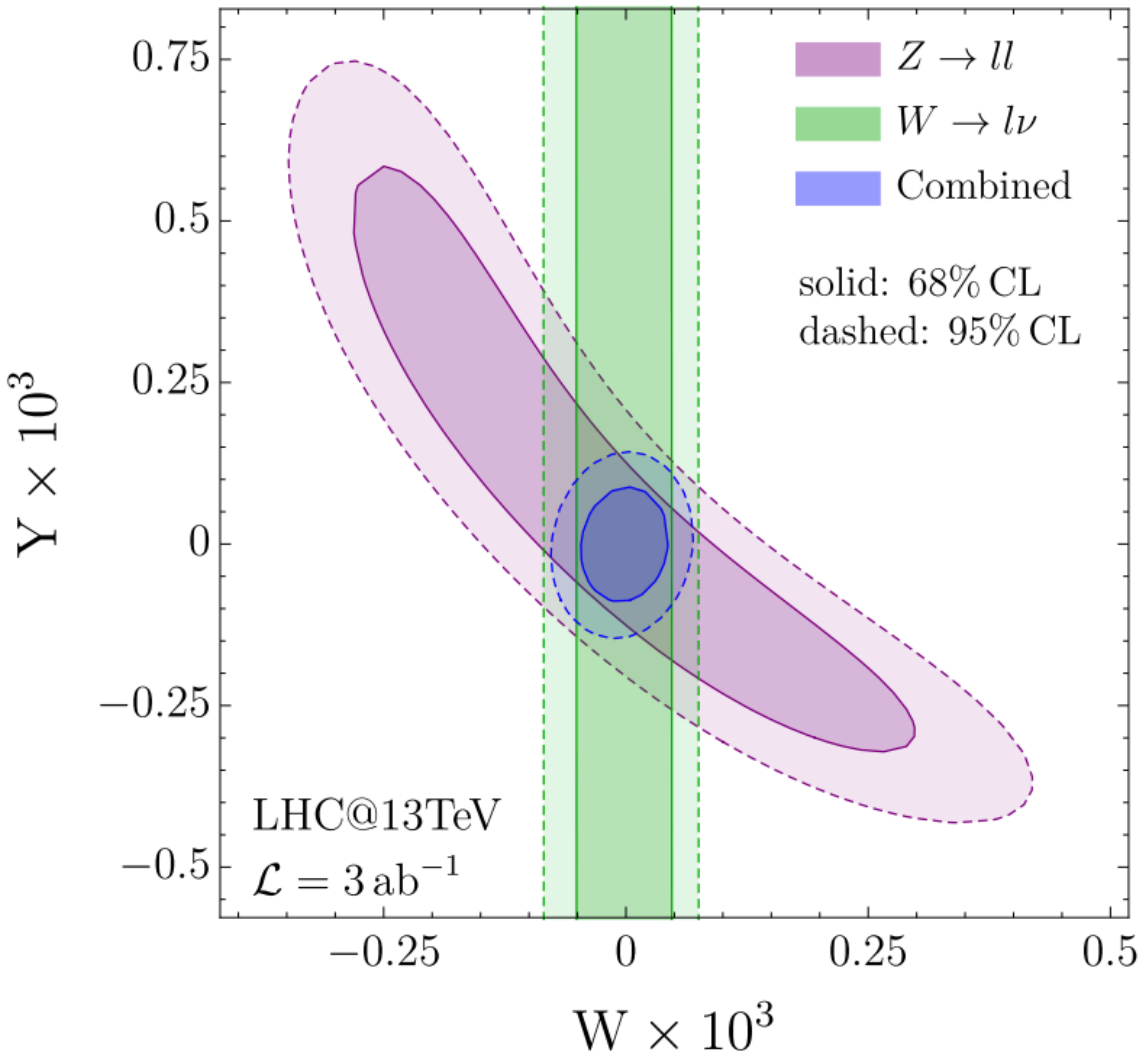


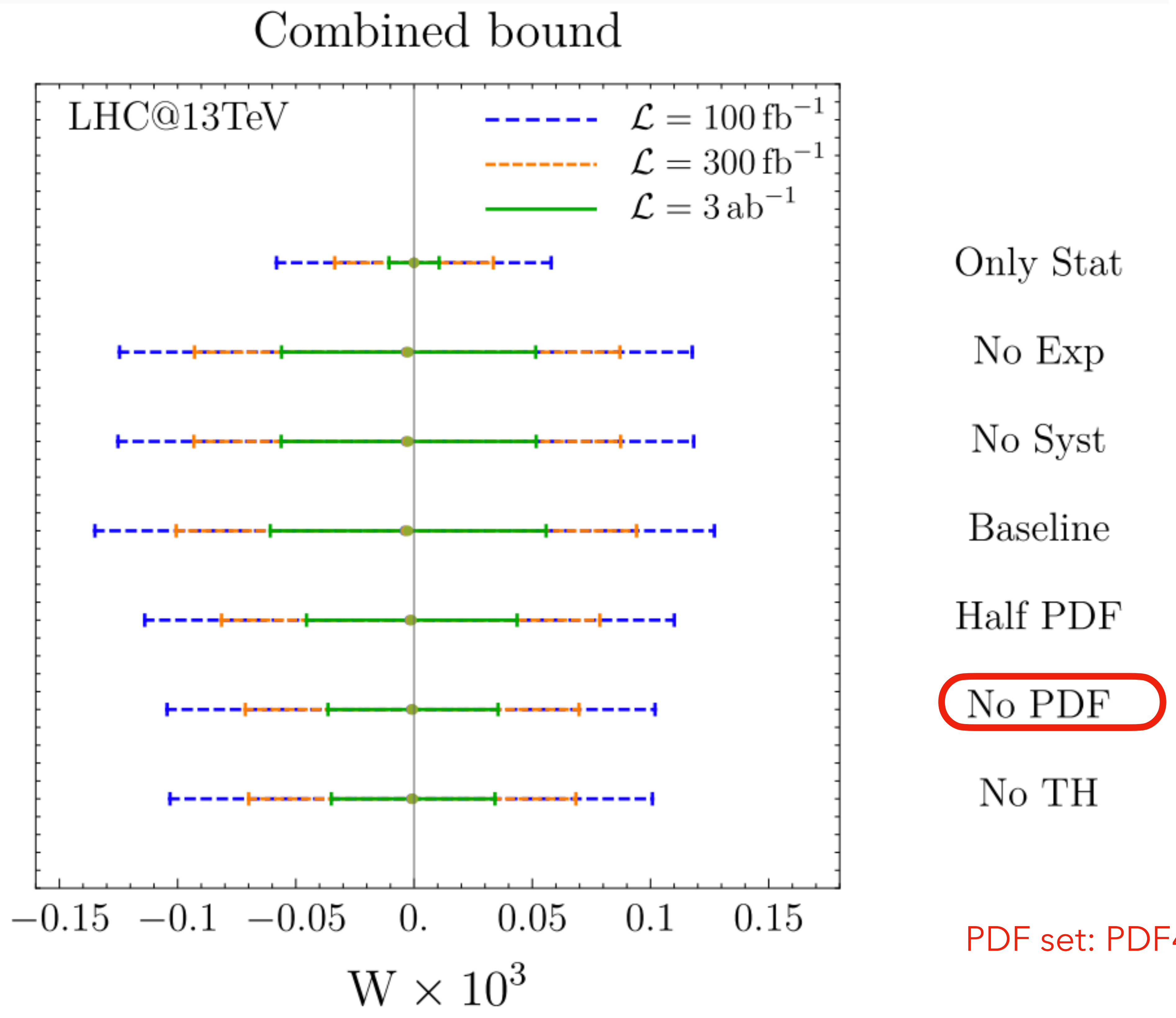
SHOULD WE WORRY ABOUT PDF AND SMEFT INTERPLAY?

- First study of interplay in case of DIS data
  [Carrazza et al Phys.Rev.Lett. 123 (2019) 13, 132001]
- Case study: EW oblique corrections in high-mass NC and CC Drell-Yan tails
- They are powerful probes of quark-lepton contact interactions that produce growing-with-energy effects. [Torre et al, 2008.12978]

$$\mathscr{L}_{\rm SMEFT} \supset -\frac{\hat{W}}{4m_W^2}(D_\rho W^a_{\mu\nu})^2 - \frac{\hat{Y}}{4m_W^2}(\partial_\rho B_{\mu\nu})^2$$

Torre et al, 2008.12978

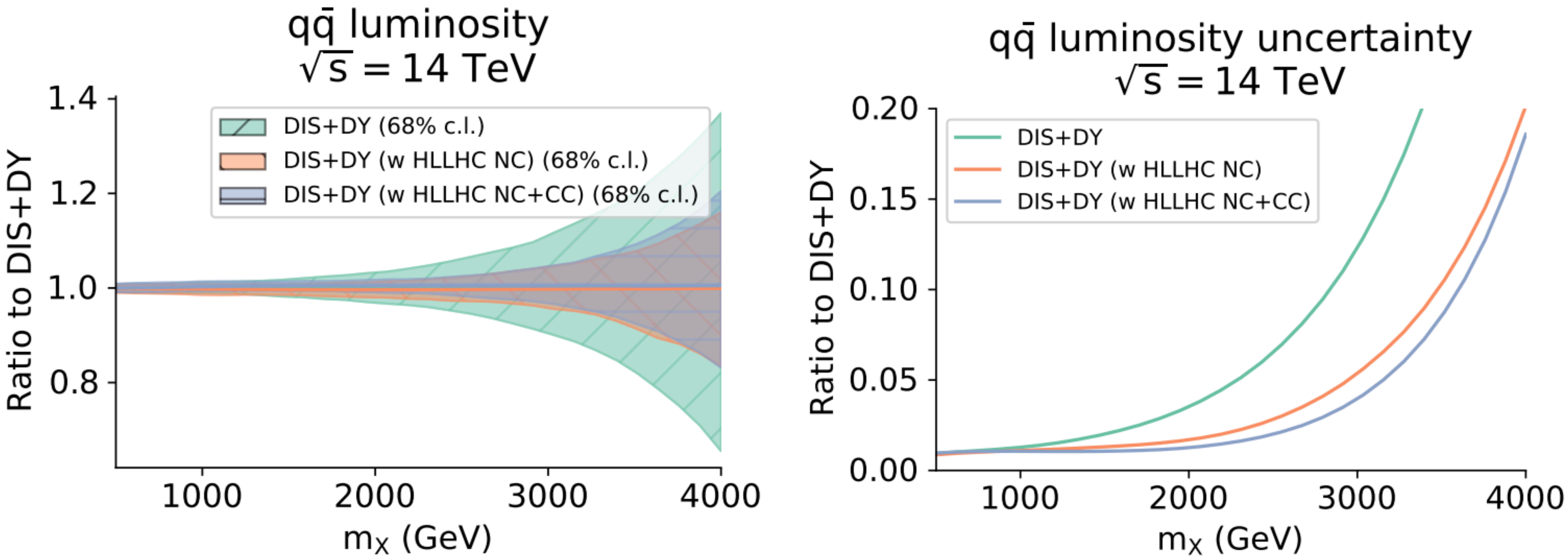




PDF set: PDF4LHC15_nlo_30_pdfas

SHOULD WE WORRY ABOUT PDF AND SMEFT INTERPLAY?

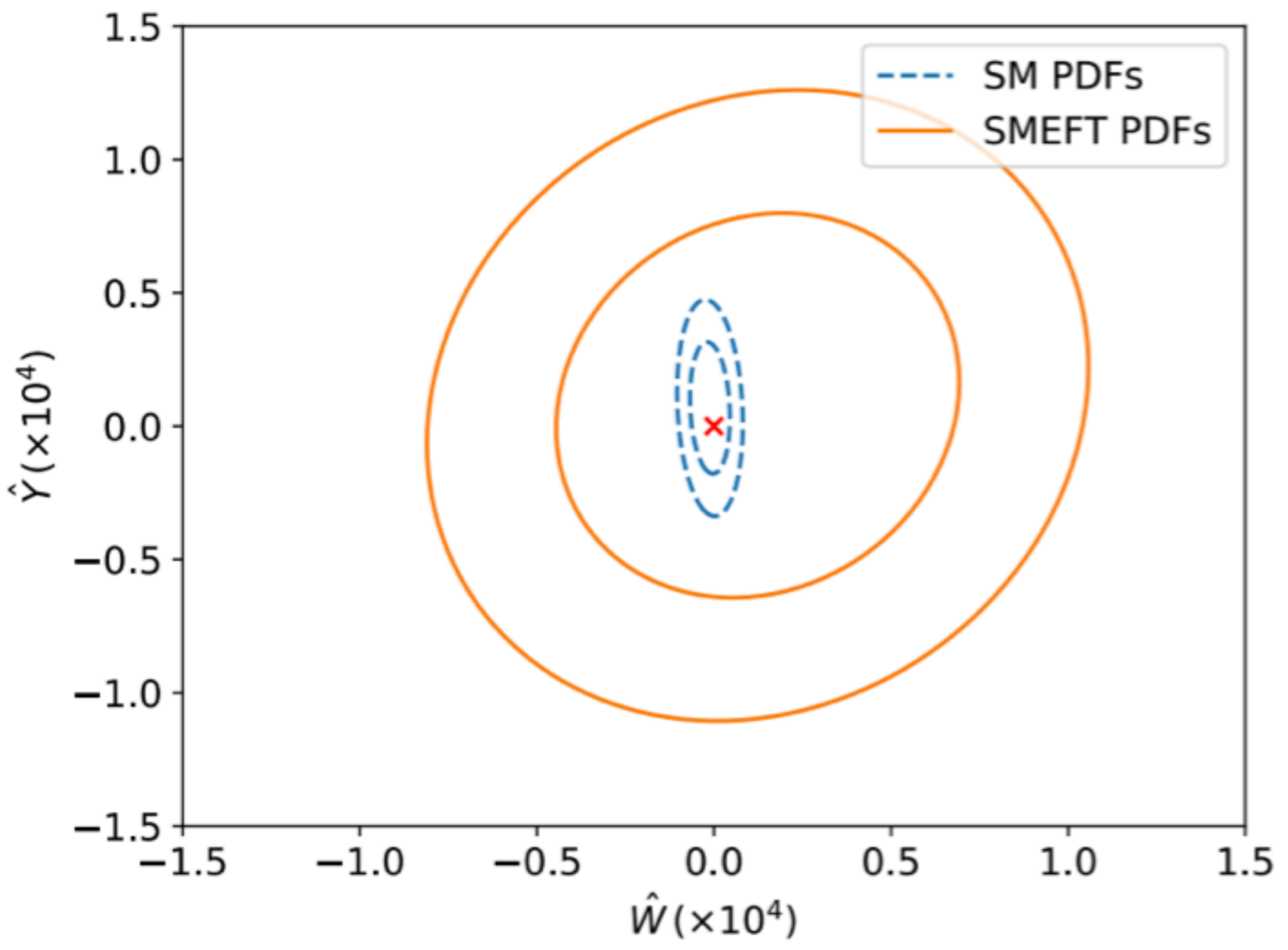
- A similar analysis has been performed, now with emphasis on PDF and their interplay with bounds on oblique operators + flavour-specific 4-fermion operator coupling muons and b-quarks in [Greljo et al, 2104.02723]
- Settings:
  - ➡ PDF fit based on DIS, Drell-Yan on-shell and low-mass data from ATLAS, CMS and LHCb
  - ➡ + Run I and II ATLAS and CMS high mass NC Drell-Yan data
  - ➡ + HL-LHC projections for NC and CC Drell-Yan data
  - ➡ SM predictions at NNLO QCD + NLO EW and SMEFT corrections added via local K-factors

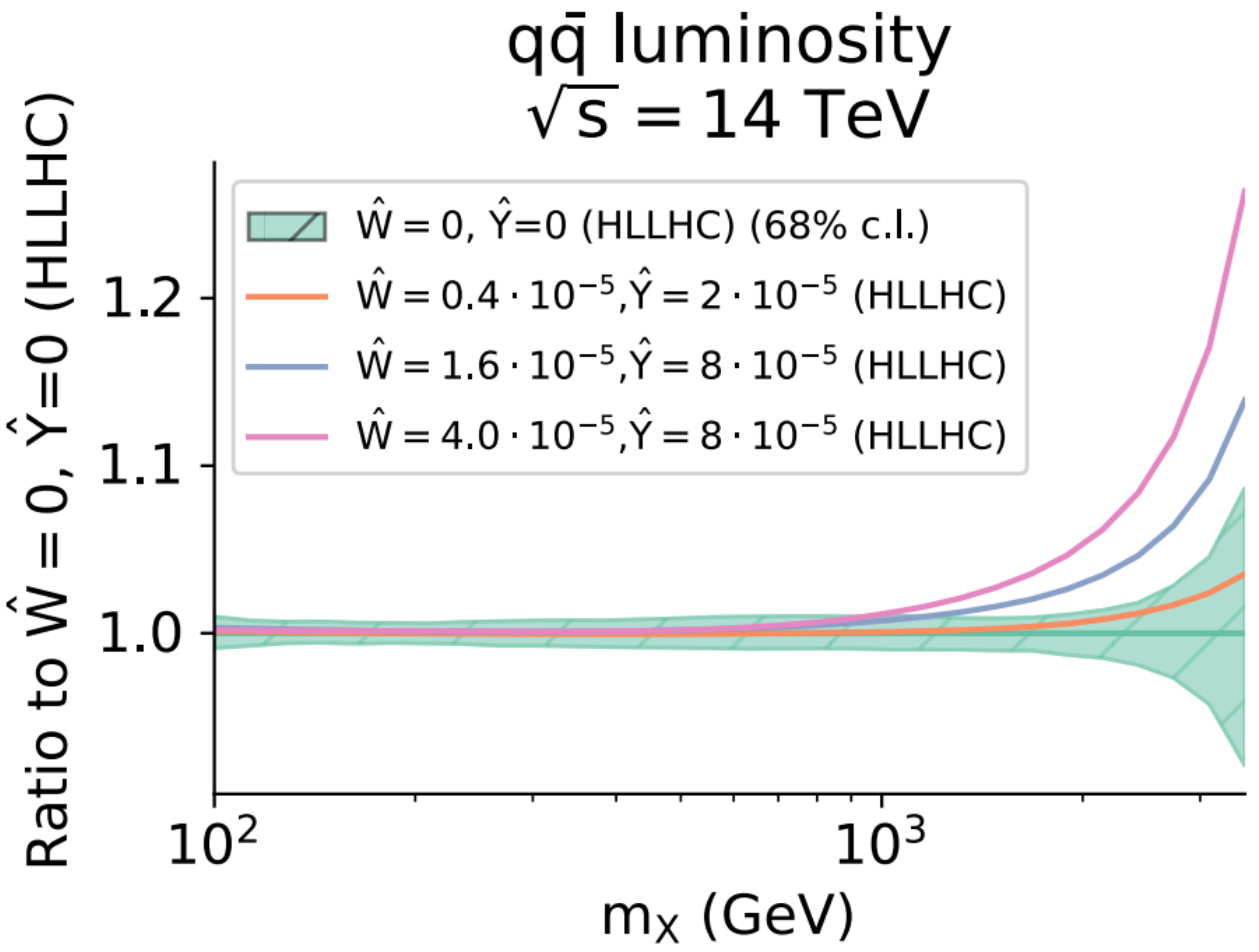


SHOULD WE WORRY ABOUT PDF AND SMEFT INTERPLAY?

- Non negligible interplay with Run I and Run II data, strong interplay at HL-LHC
- Compare Wilson coefficients bounds from HL-LHC projections assuming SM PDFs (that include NC+CC data) to the bounds on the same Wilson coefficients obtained from a simultaneous fit of PDFs and Wilson coefficients
- Not accounting for interplay (using PDFs as a black box) leads to over-constrained bounds
- Note that HL-LHC projections based on pseudo-data. Simultaneous fits of SMEFT and PDFs could point to a different minimum, and not only to larger uncertainties!

Greljo et al, 2104.02723





CONCLUSIONS

- The path towards 1% accurate PDFs opens up new fascinating challenges
- Precise and accurate determination of the proton structure is key to make progress in precision phenomenology
- Need: robust methodology & solid statistical tests of methodology
- Benchmark effort would benefit of public releases of PDF codes and inputs!

- Current uncertainties partially underestimated due to theory uncertainty not accounted for at NNLO & other effects (fixed-order calculations [Alekhin et al 2104.02400], full account of electroweak corrections, resummations, nuclear corrections [Ball et al 2105.05114]…)
- Proposal for inclusion of missing higher order uncertainties has been assessed at NLO. To go at NLO, need development of evolution code with scale variations at NNLO

- Time to study the interplay between indirect new physics searches and PDFs
- The preferred avenue ahead is to be able to perform simultaneous fits (like for PDFs and α_s)
- In parallel a more careful investigation of definition of “safe” PDF sets & account for PDF uncertainties

THANK YOU FOR YOUR ATTENTION!