QCD: NNPDF

Maria Ubiali

University of Cambridge

09th January 2019

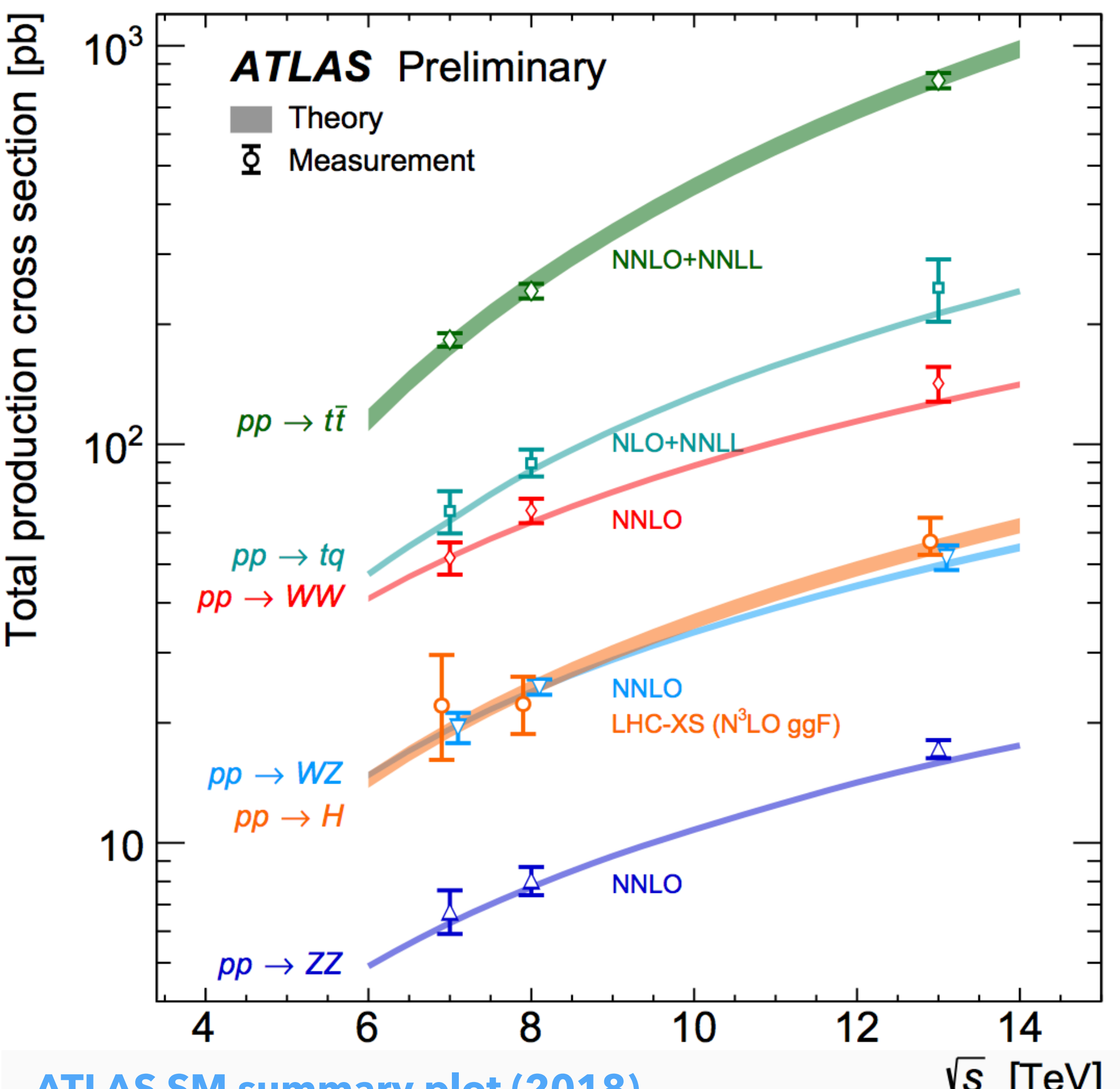
YETI 2019

A new precision era in particle physics

- Is precision physics possible/necessary at hadron colliders?

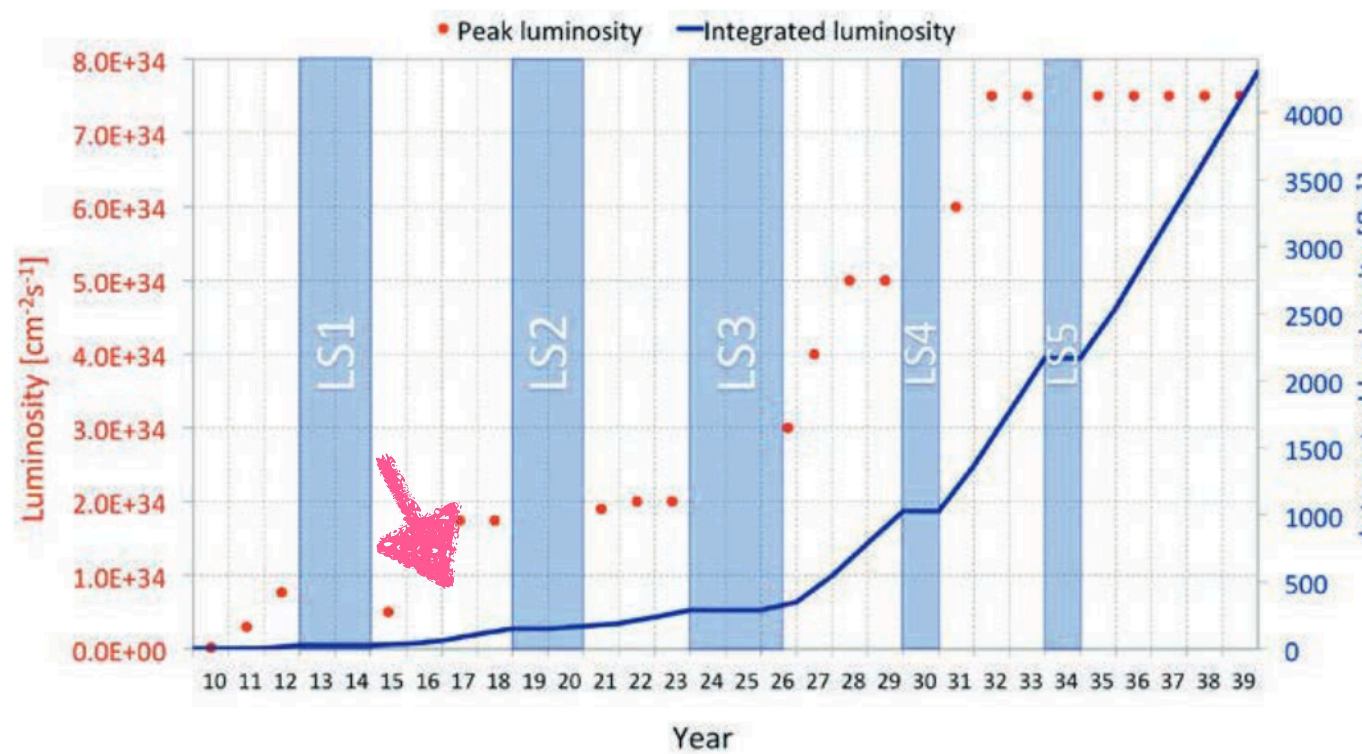
At the LHC a paradigm shift took place: discovery ➜ discovery through precision

- Theoretical predictions to catch up with precision of experimental data



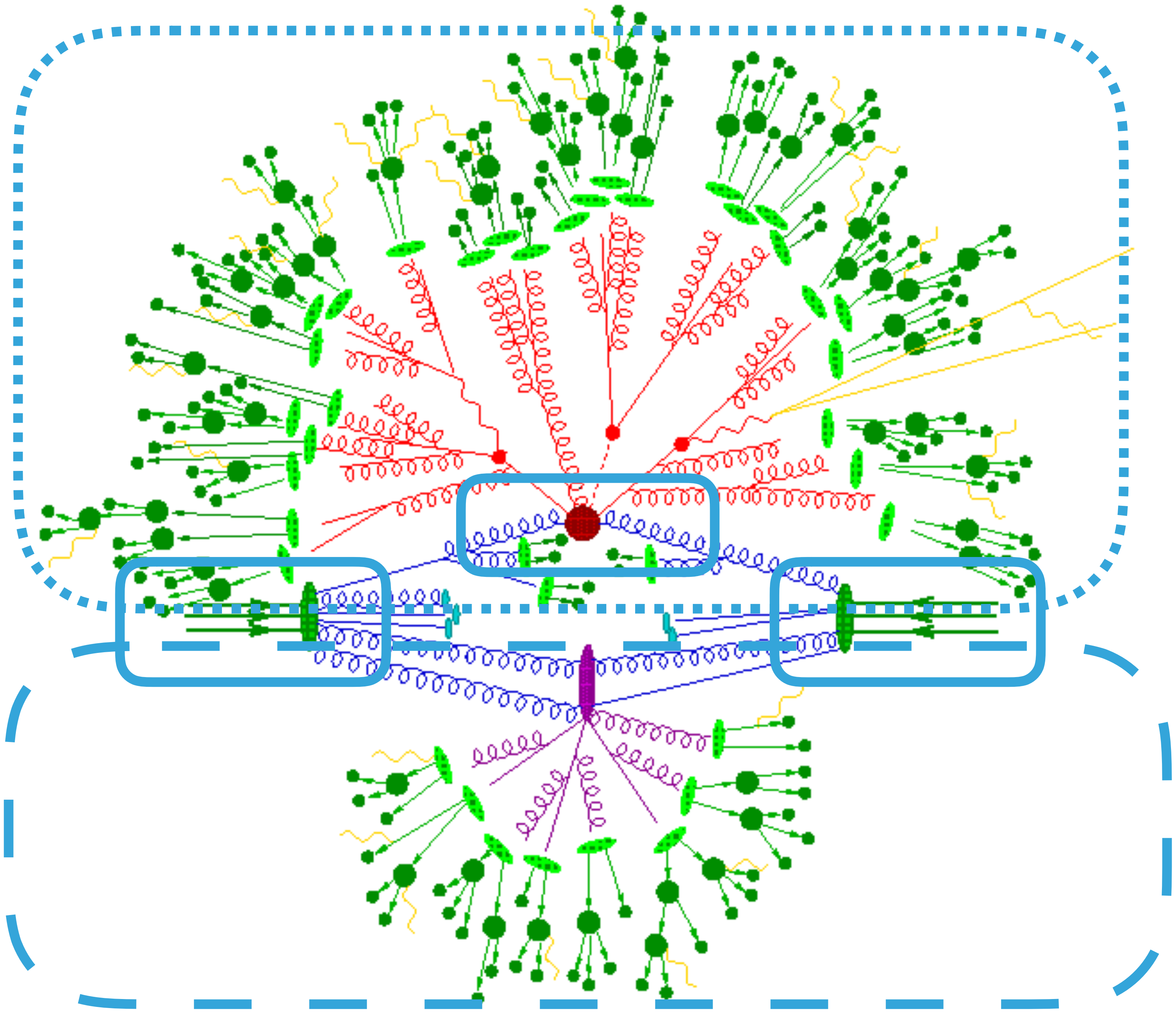
ATLAS SM summary plot (2018)

- Precise theoretical predictions are key not to miss unique opportunity given by precise data in the next 20+ years



Apollinari et al., CERN Yellow Report 4 (2017)

The precision ingredients

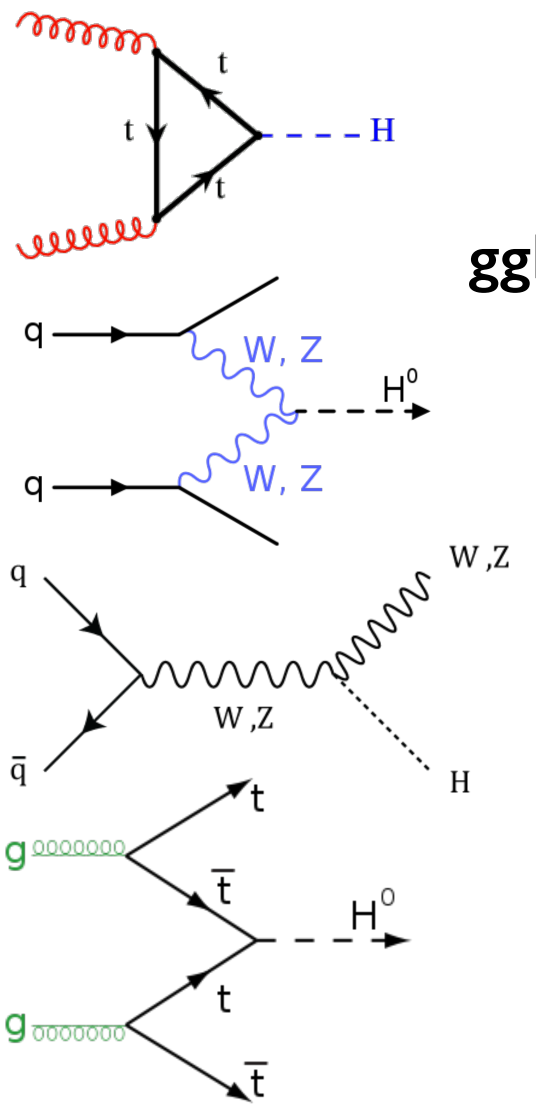


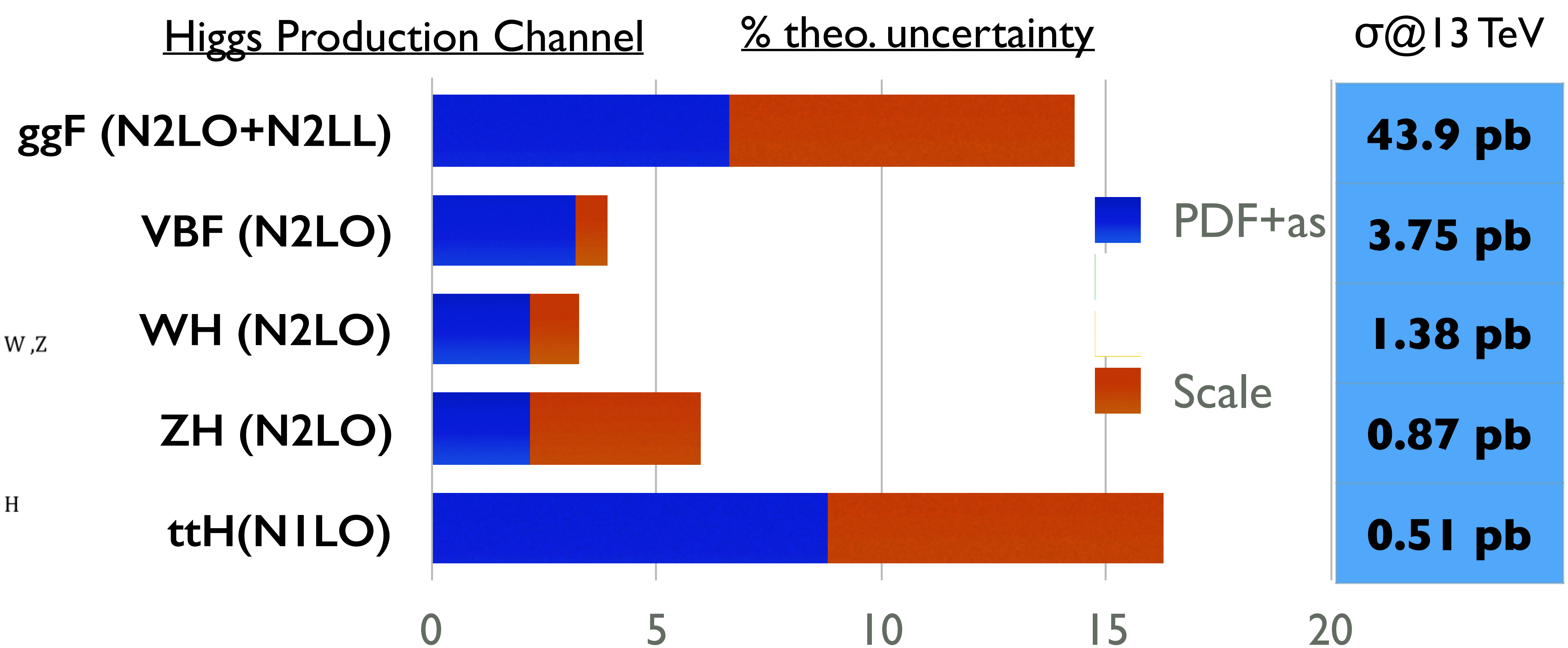
- Hard scattering of partons (Perturbative QCD+EW)
- Parton Distribution Functions
- Parton Showering and Hadronization
- Multiple Parton Interaction, Underlying Events

SHERPA artist

The role of PDF uncertainties

Yellow Report 3 (2013)



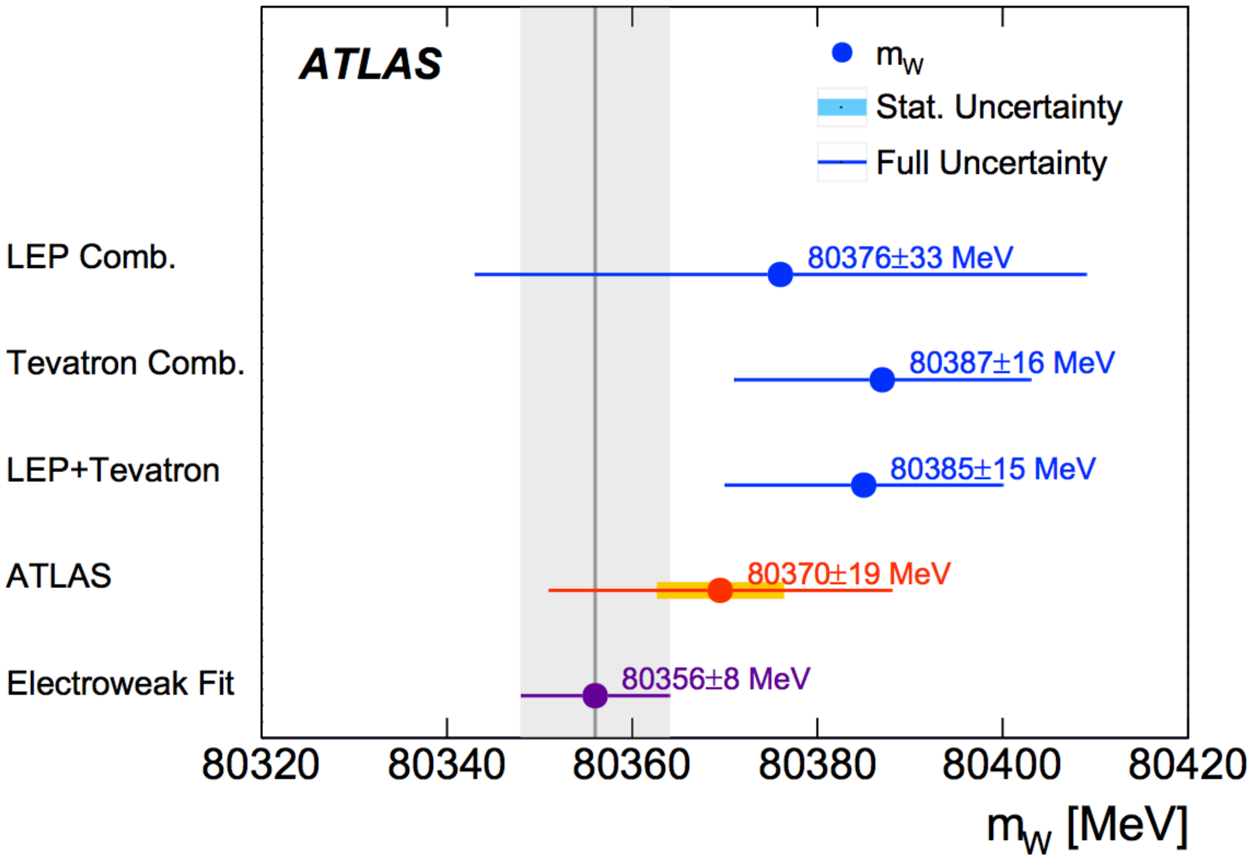


PDF uncertainties limiting factor in the accuracy of theoretical predictions

Higgs physics

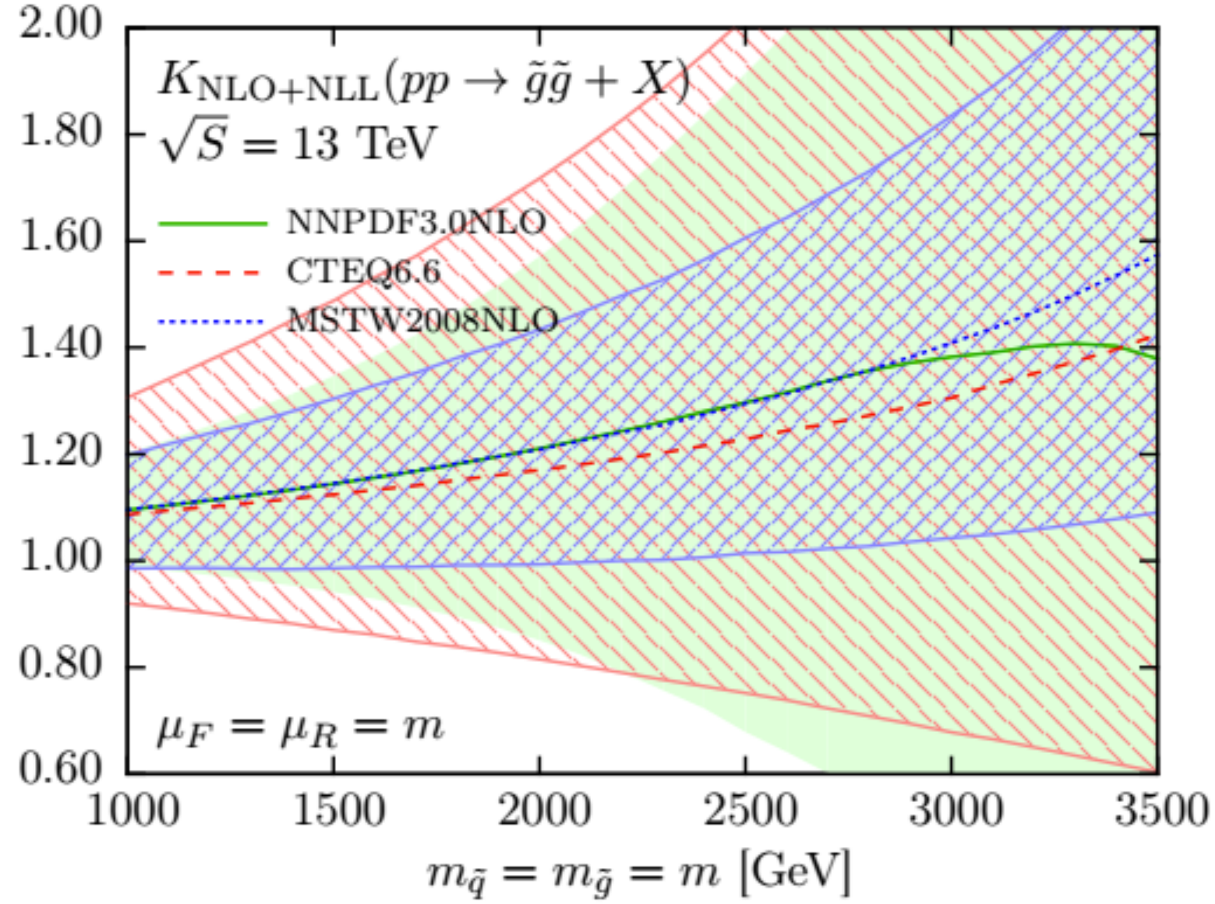
The role of PDF uncertainties

Determination of SM parameters



ATLAS collaboration, EPJC 78 (2018) 110

New Physics



Beenakker et al.
EPJC76 (2016)2, 53

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

Outline

- Introduction
  - Parton Model
  - Collinear Factorization
- PDFs determination
  - Experimental Data
  - Fits and methodology
  - The NNPDF approach
- State of the art and frontiers
  - Current status
  - New frontiers
- Conclusions

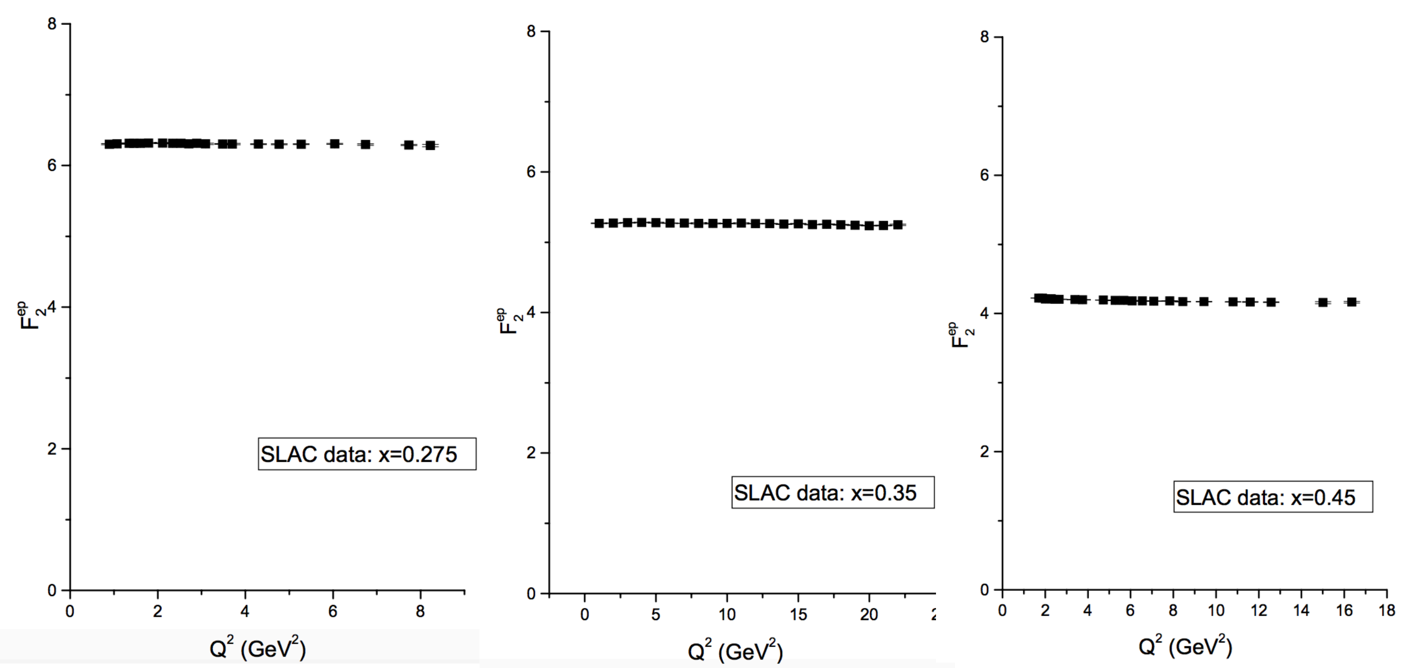
Introduction

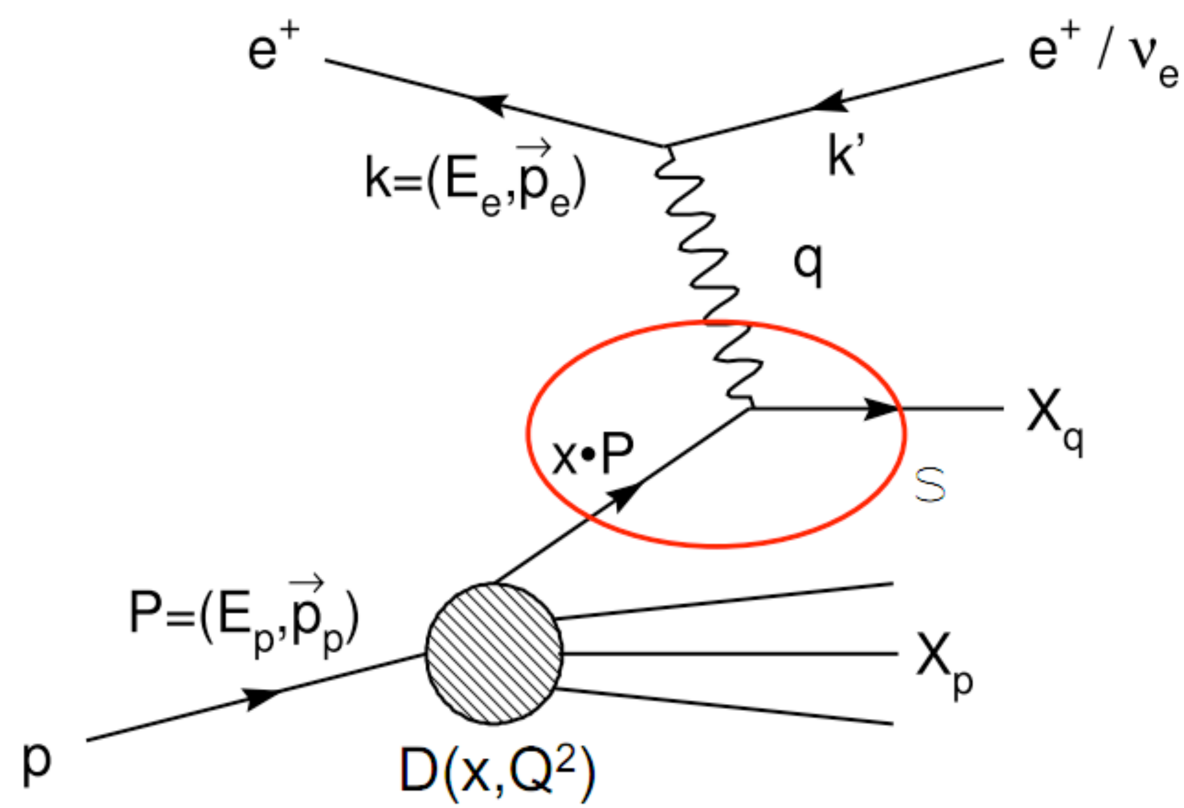
Parton Model

The parton model **(Feynman 1969)**

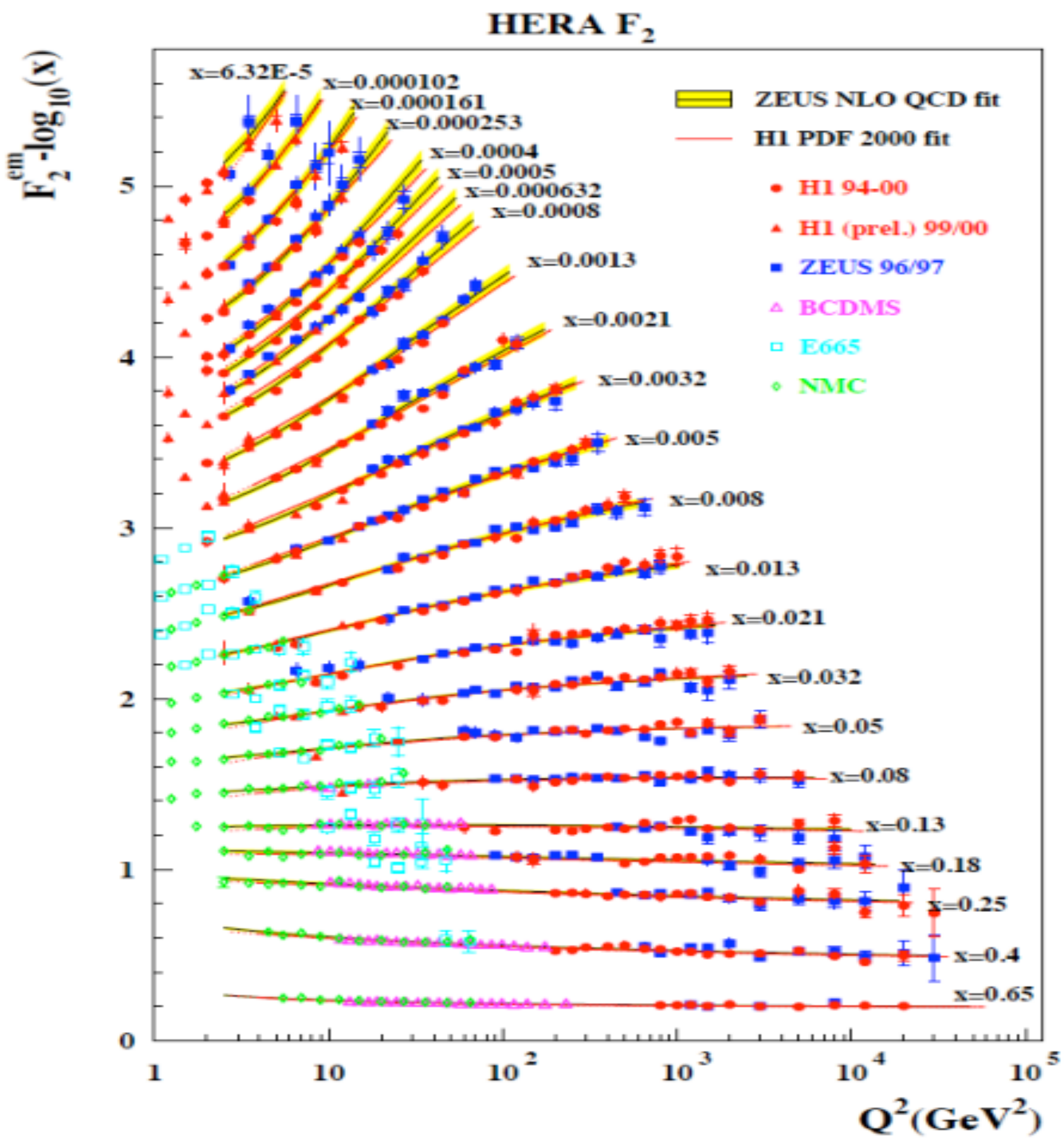
- Photon scatters incoherently off massless, free, point like, spin 1/2 quarks
- The functions q(x) are the Parton Distribution Functions encode probability that a quark carries a fraction x of parent proton's
- They encode information about proton deep structure

$$F_2^{\gamma,Z}(x) = x\sum_{i=1}^{n_f} c_i\left[q_i(x)+\bar{q}_i(x)\right]$$





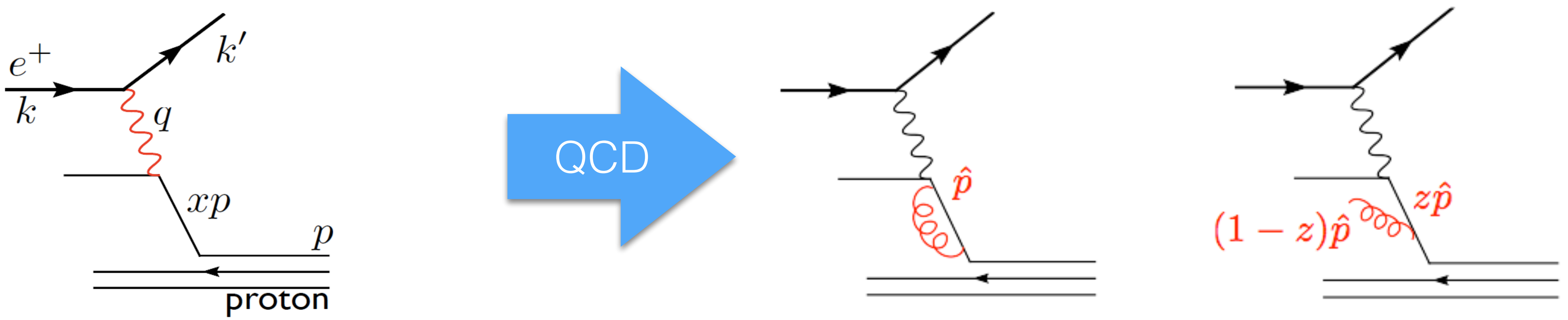
Parton Model



scaling violation

approx. scaling

QCD and improved parton model



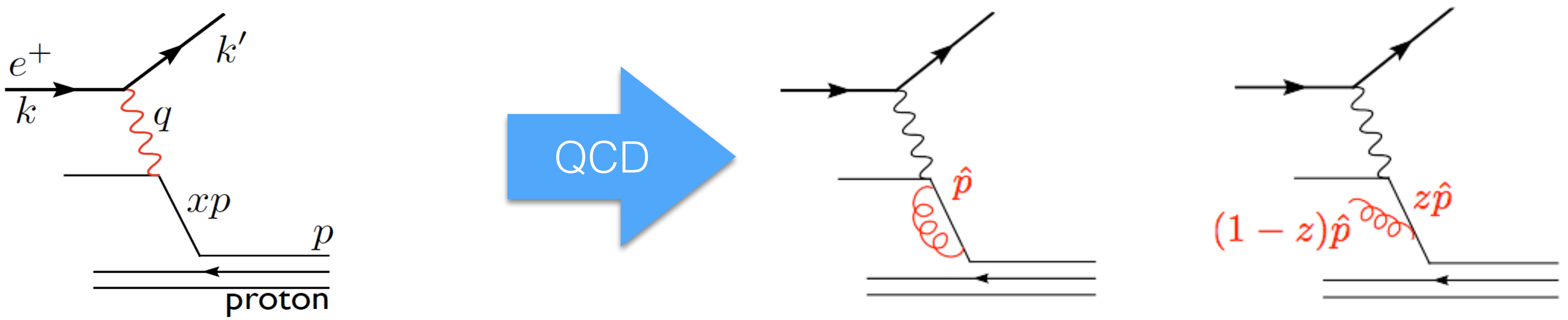
Adding **real emission** and **virtual correction**

$$\sigma^{(1)} = C_F \frac{\alpha_s}{2\pi} \int_0^{Q^2} \frac{dk_\perp^2}{dk_\perp^2} \int_0^1 dz P(z) \left(\sigma^{(0)}(z\hat{p}) - \sigma^{(0)}(\hat{p}) \right)$$

→ ∞!

- Soft limit (z→ 1): singularity cancels between real and virtual terms
- For z < 1 a collinear singularity appears for $k_\perp$ → 0
- Naive parton model does not survive QCD radiative corrections
- Need "renomalization" of collinear divergences as in the case of UV divergences

QCD and improved parton model



Adding **real emission** and **virtual correction**

$$\sigma^{(1)} = C_F \frac{\alpha_s}{2\pi} \int_{\lambda^2}^{Q^2} \frac{dk_\perp^2}{dk_\perp^2} \int_0^1 dz P(z) \left(\sigma^{(0)}(z\hat{p}) - \sigma^{(0)}(\hat{p}) \right)$$

$$\sigma^{(1)} = C_F \frac{\alpha_s}{2\pi} \int_{\lambda^2}^{Q^2} \frac{dk_\perp^2}{dk_\perp^2} \int_0^1 dz [P(z)]_+ \sigma^{(0)}(z\hat{p})$$

With $\int_0^1 dz f(z)_+ g(z) = \int_0^1 dz f(z)(g(z) - g(1))$ Plus prescription

QCD and improved parton model

The trick is to factorize the collinear singularities by splitting the log

$$\sigma^{(1)} = C_F \frac{\alpha_s}{2\pi} \int_{\lambda^2}^{Q^2} \frac{dk_\perp^2}{dk_\perp^2} \int_0^1 dz [P(z)]_+ \sigma^{(0)}(z\hat{p})$$

μ_{F}: factorisation scale

$$\log\left(\frac{Q^2}{\lambda^2}\right) = \log\left(\frac{Q^2}{\mu_F^2}\right) + \log\left(\frac{\mu_F^2}{\lambda^2}\right)$$

$$\sigma = \sigma^{(0)} + \sigma^{(1)} = \left(1 + \frac{\alpha_s}{2\pi}\log\frac{\mu_F^2}{\lambda^2}P_+\right) \times \left(1 + \frac{\alpha_s}{2\pi}\log\frac{Q^2}{\mu_F^2}P_+\right)$$

$$\sigma(p, \mu_F) = \left(1 + \frac{\alpha_s}{2\pi}\ln\frac{Q^2}{\mu_F^2}P_+\right) \times \sigma^{(0)}(p)$$

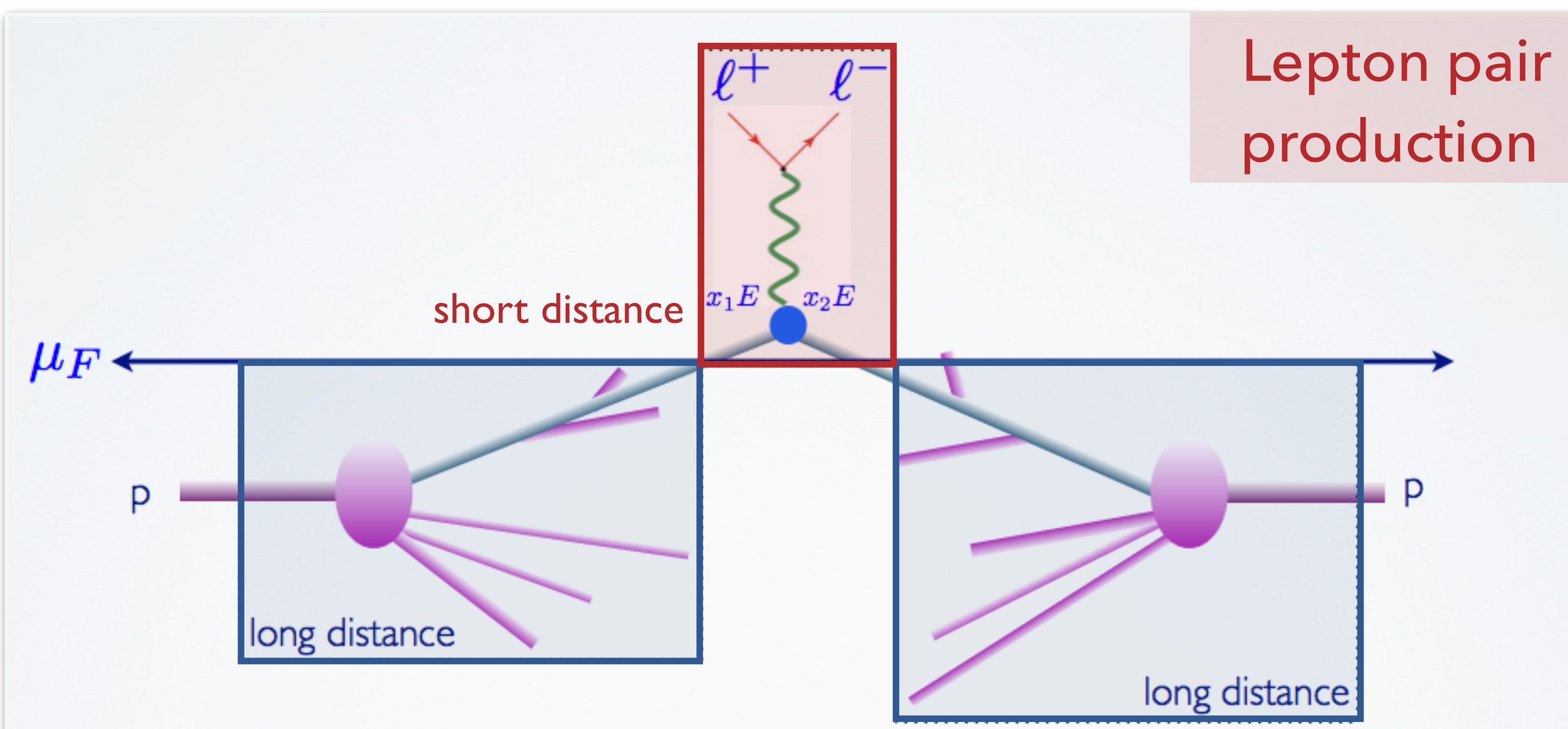
$$f(x, \mu_F) = f(x) \times \left(1 + \frac{\alpha_s}{2\pi}\ln\frac{\mu_F^2}{\lambda^2}P_{qq}^{(0)}\right)$$

Both PDFs and partonic cross section acquire dependence on arbitrary scale μ_F

Collinear Factorisation Theorem

$$\frac{d\sigma_H^{ep\to ab}}{dX} = \sum_{i=-n_f}^{+n_f} \int_{x_B}^{1} \frac{dz}{z} f_i(z,\mu_F)\, \frac{d\hat{\sigma}_i^{ei}}{dX}(zS, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda^n}{S^n}\right)$$

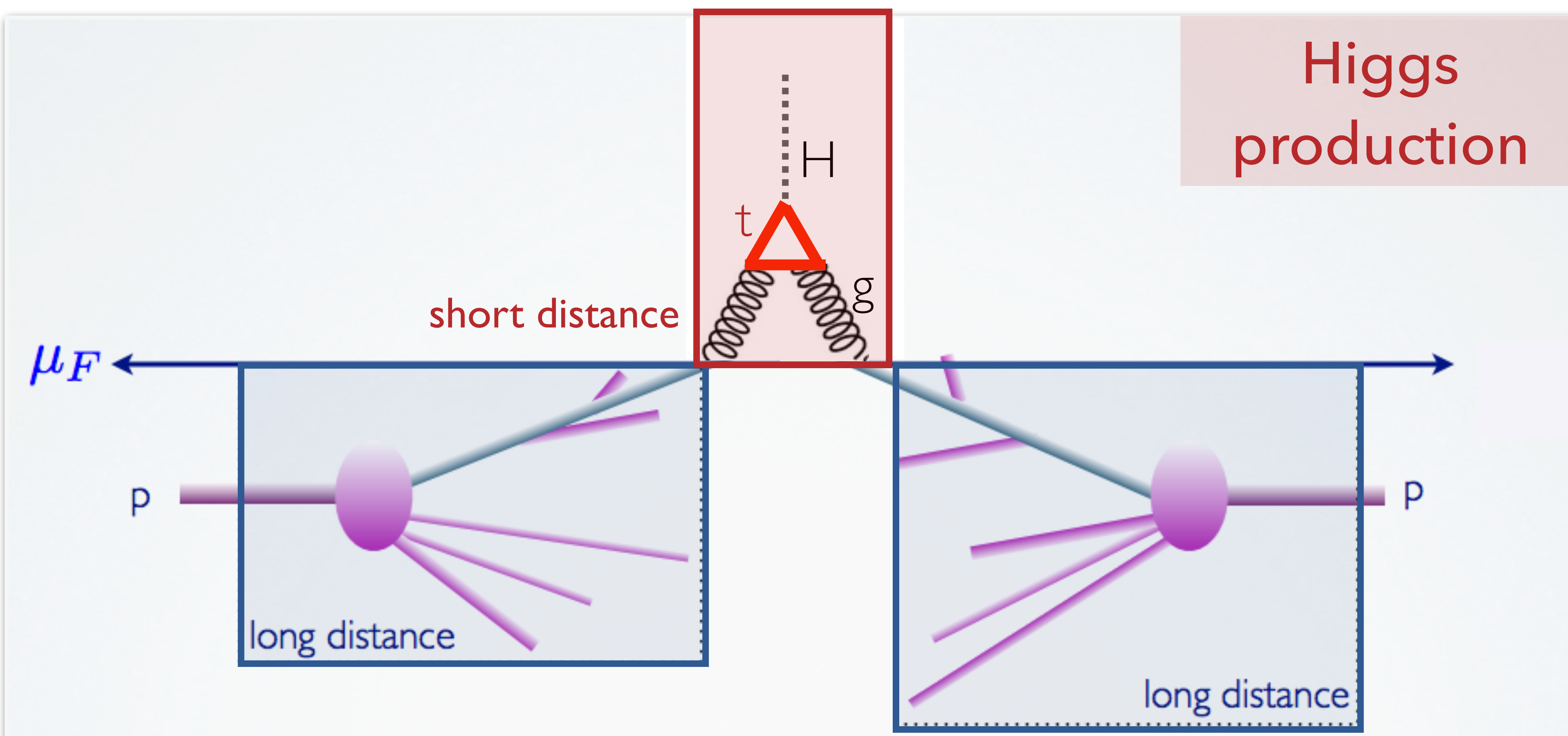
$$\frac{d\sigma_H^{pp\to ab}}{dX} = \sum_{i,j=-n_f}^{+n_f} \int_{\tau_0}^{1} \frac{dz_1}{z_1} \frac{dz_2}{z_2} f_i(z_1,\mu_F)\, f_j(z_2,\mu_F)\, \frac{d\hat{\sigma}_i^{ij}}{dX}(zS, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda^n}{S^n}\right)$$



Collinear Factorisation Theorem

$$\frac{d\sigma_H^{ep\to ab}}{dX} = \sum_{i=-n_f}^{+n_f} \int_{x_B}^{1} \frac{dz}{z}\, f_i(z,\mu_F)\, \frac{d\hat{\sigma}_i^{ei}}{dX}(zS, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda^n}{S^n}\right)$$

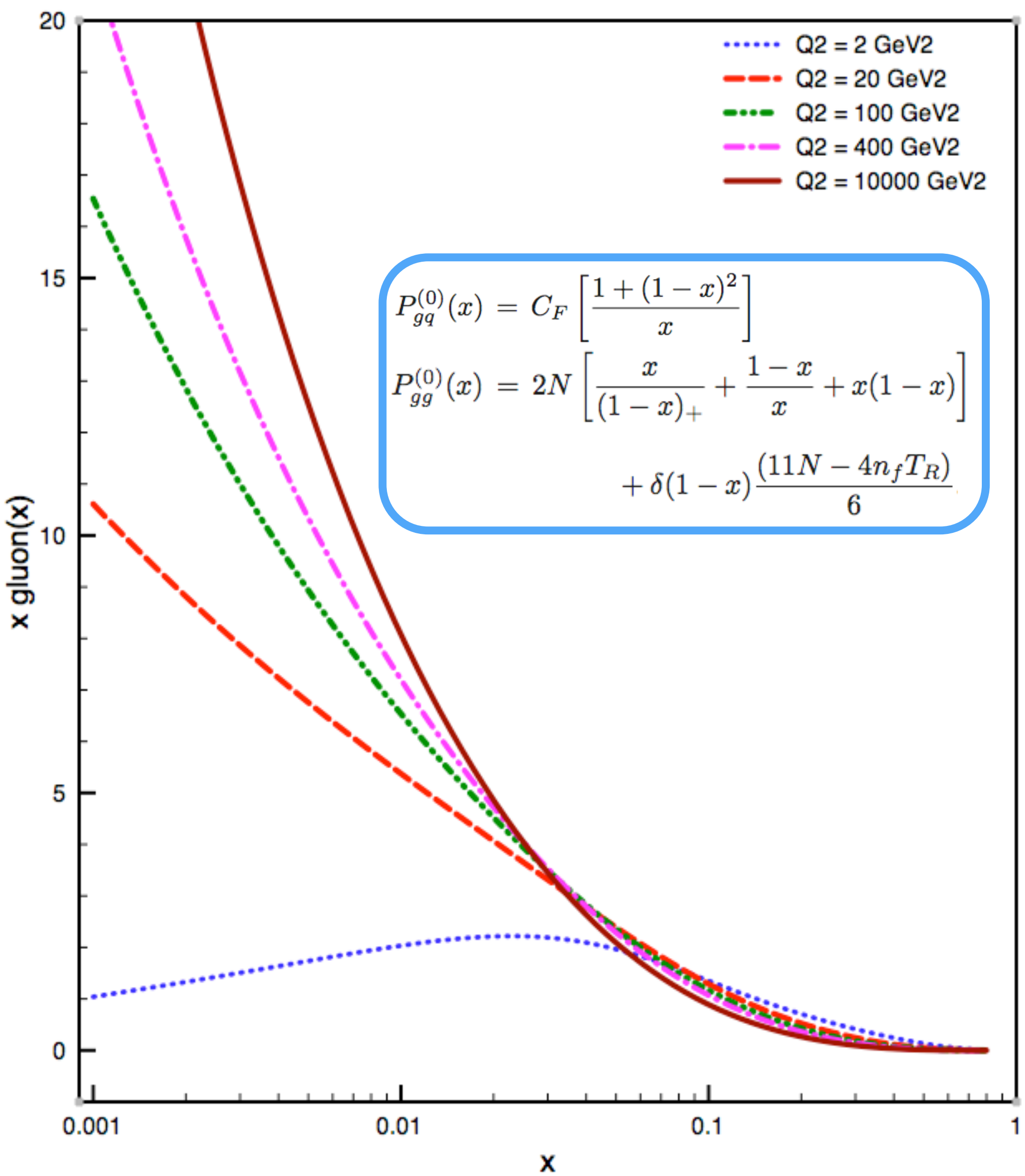
$$\frac{d\sigma_H^{pp\to ab}}{dX} = \sum_{i,j=-n_f}^{+n_f} \int_{\tau_0}^{1} \frac{dz_1}{z_1}\frac{dz_2}{z_2}\, f_i(z_1,\mu_F)\, f_j(z_2,\mu_F)\, \frac{d\hat{\sigma}_i^{ij}}{dX}(zS, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda^n}{S^n}\right)$$



DGLAP evolution equations

In analogy with running coupling, imposing that cross section does not depend on arbitrary scale **μ**, get renormalisation group equations for PDFs

$$t = \log \frac{Q^2}{\mu_F^2} \qquad \frac{d}{dt}\begin{pmatrix} q_i(x,t) \\ g(x,t) \end{pmatrix} = \frac{\alpha_s(t)}{2\pi}\int_x^1 \sum_{j=q,\bar{q}} \frac{d\xi}{\xi}\begin{pmatrix} P_{ij}\left(\frac{x}{\xi},\alpha_s(t)\right) & P_{ig}\left(\frac{x}{\xi},\alpha_s(t)\right) \\ P_{gj}\left(\frac{x}{\xi},\alpha_s(t)\right) & P_{gg}\left(\frac{x}{\xi},\alpha_s(t)\right) \end{pmatrix} \otimes \begin{pmatrix} q_j(\xi,t) \\ g(\xi,t) \end{pmatrix}$$



Dokshitzer, Gribov, Lipatov, Altarelli, Parisi equations

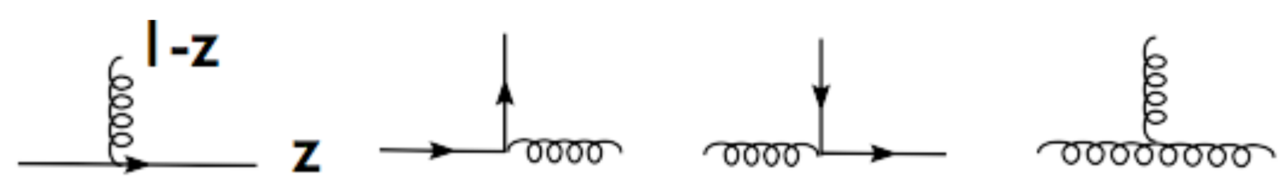
- Functional dependence on μ^2 is totally predicted by solving DGLAP evolution eqns

- Splitting functions known up to NNLO:

  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

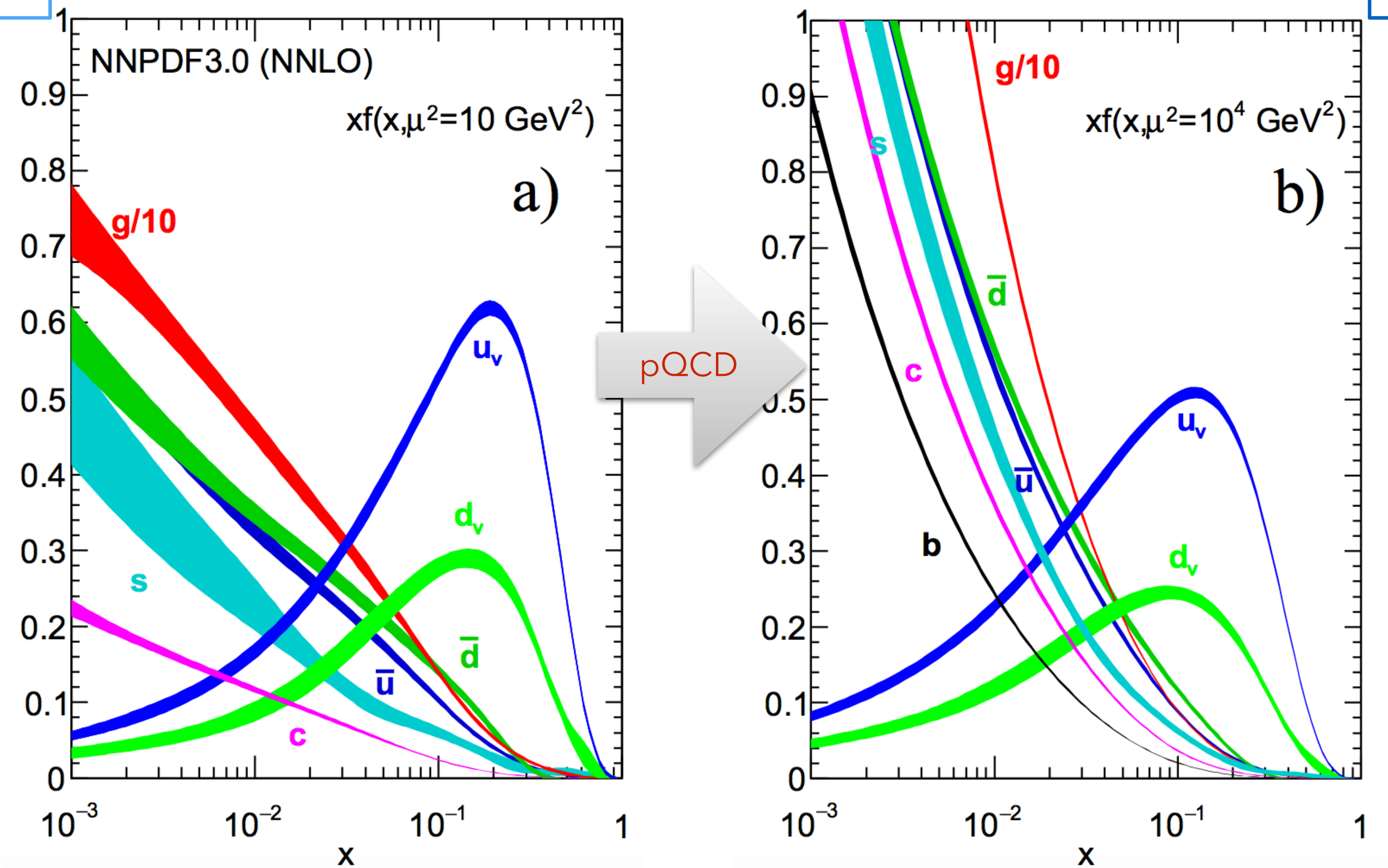


DGLAP evolution equations

Functional dependence of PDFs on the scale is totally predicted up to NNLO accuracy by solving DGLAP evolution equations

Hadronic scale: global fit of PDFs

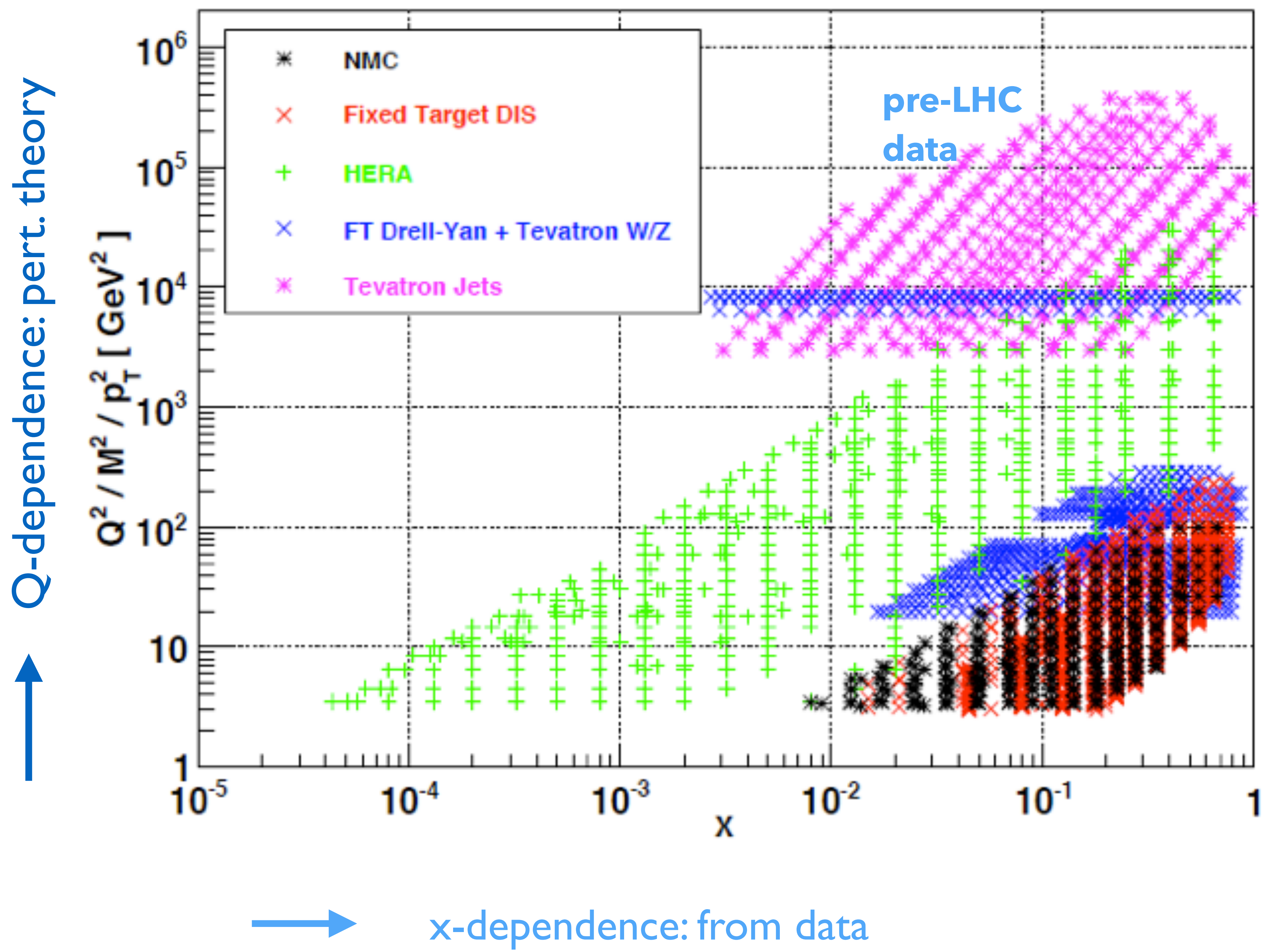
High scale: input to the LHC



PDG 2016

PDF determination

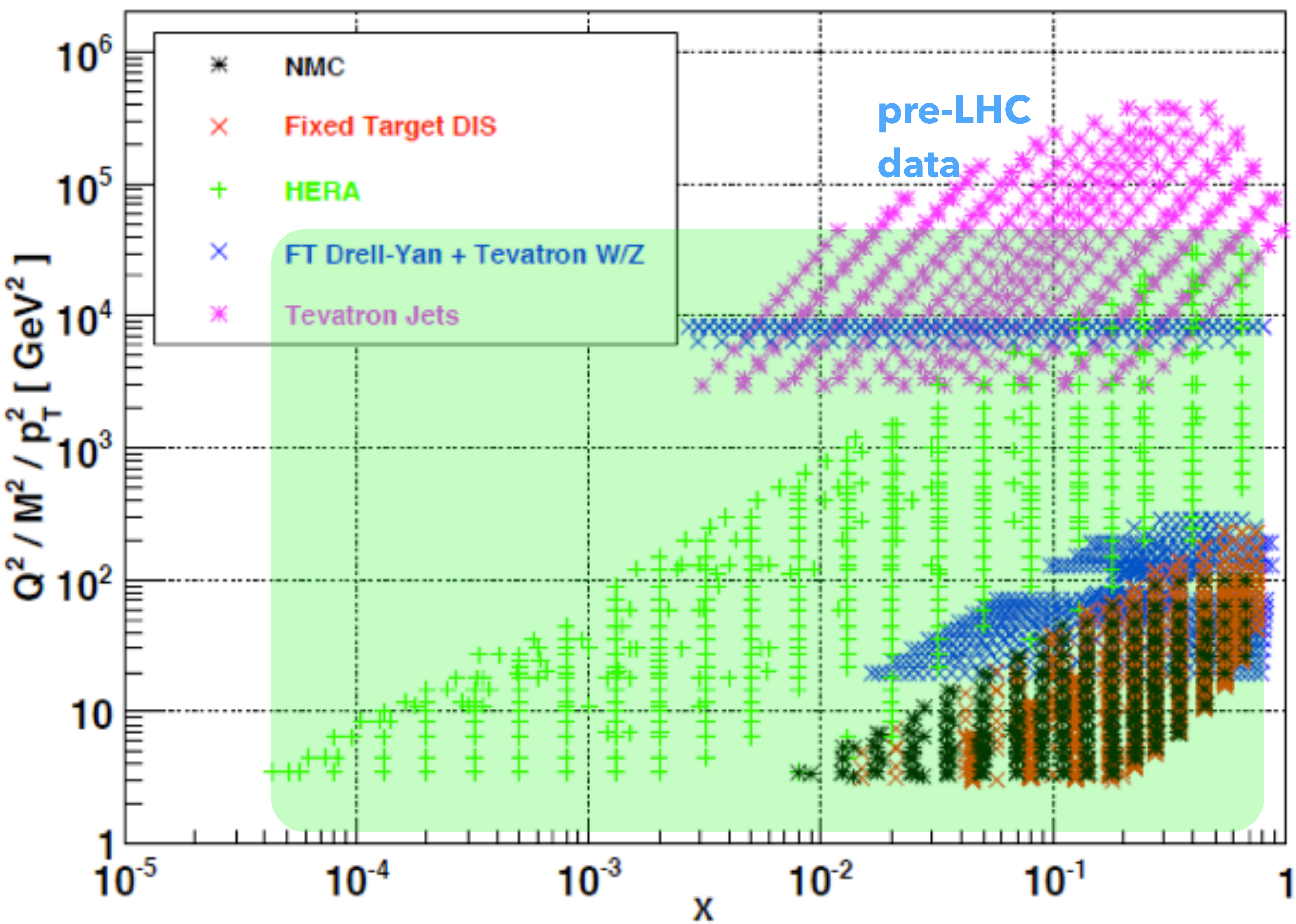
PDF determination



Different data constrain different PDF combinations in different kinematic regions.

PDF determination

Q-dependence: pert. theory



x-dependence: from data

NC $F_1^{\gamma, Z} = \sum_i e_i^2 (q_i + \bar{q}_i)$

CC $F_1^{W^+} = \bar{u} + d + s + \bar{c}$

CC $-F_3^{W^+}/2 = \bar{u} - d - s + \bar{c}$

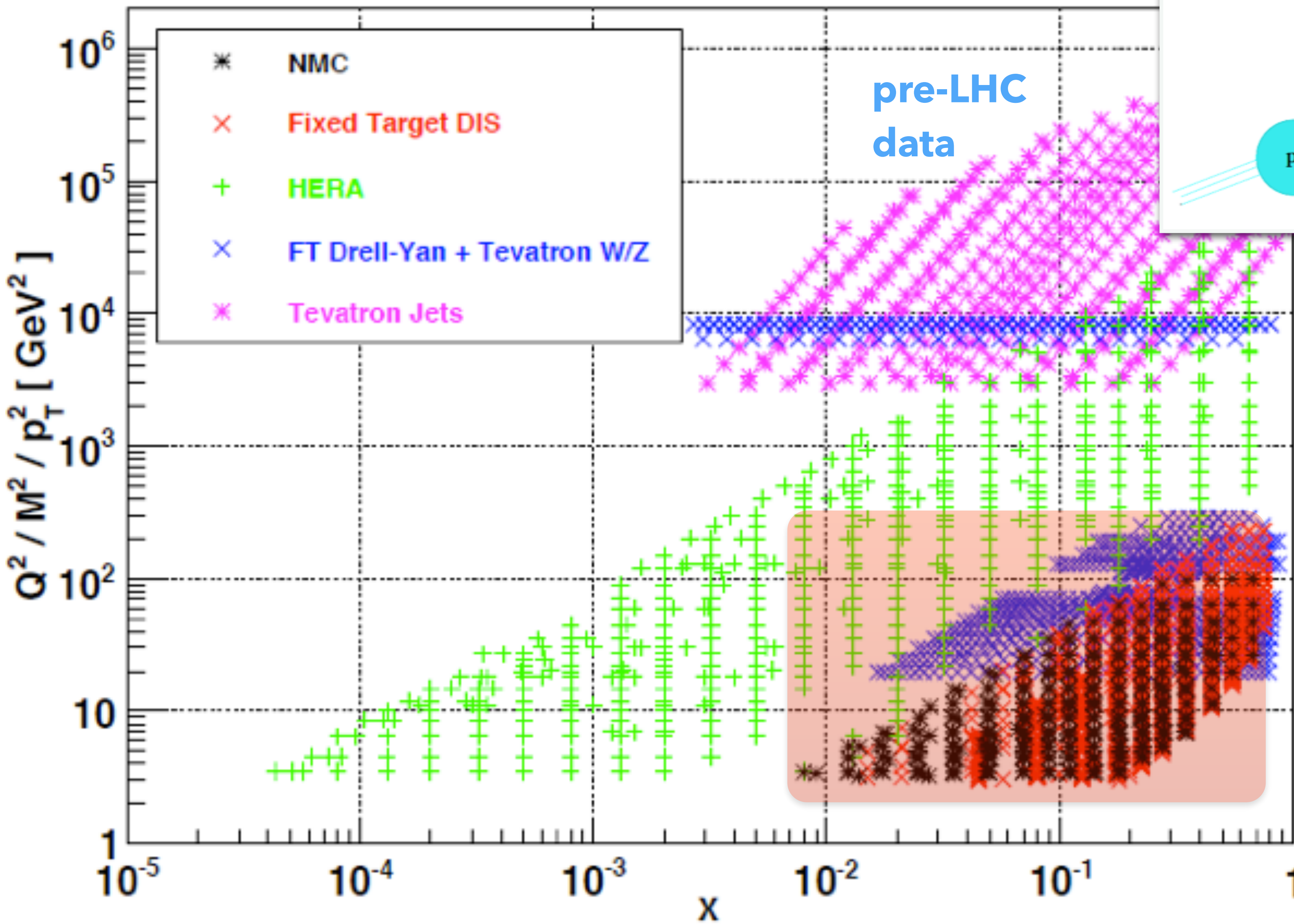
$F_2 = 2xF_1$

$$x \equiv \frac{Q^2}{2p_2.q}$$

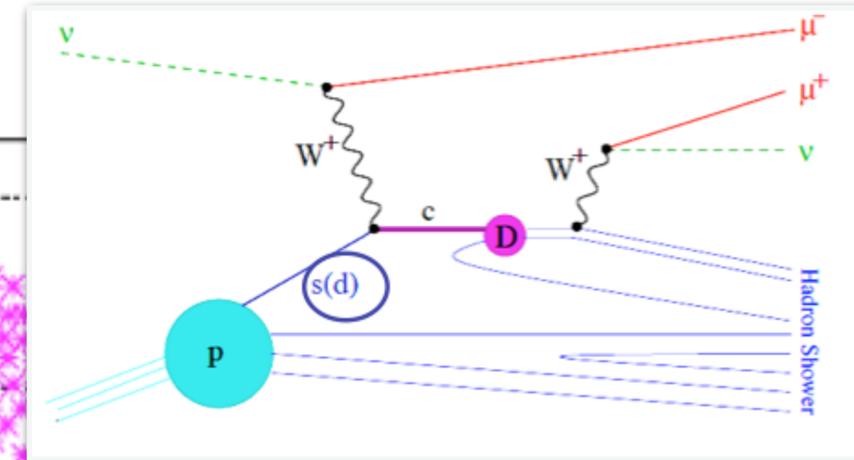
- Backbone of any PDF fit
- Structure functions known up to $O(\alpha_S^3)$
- Constrain singlet and uv dv combinations
- Constrain g at small and moderate x

PDF determination

Q-dependence: pert. theory



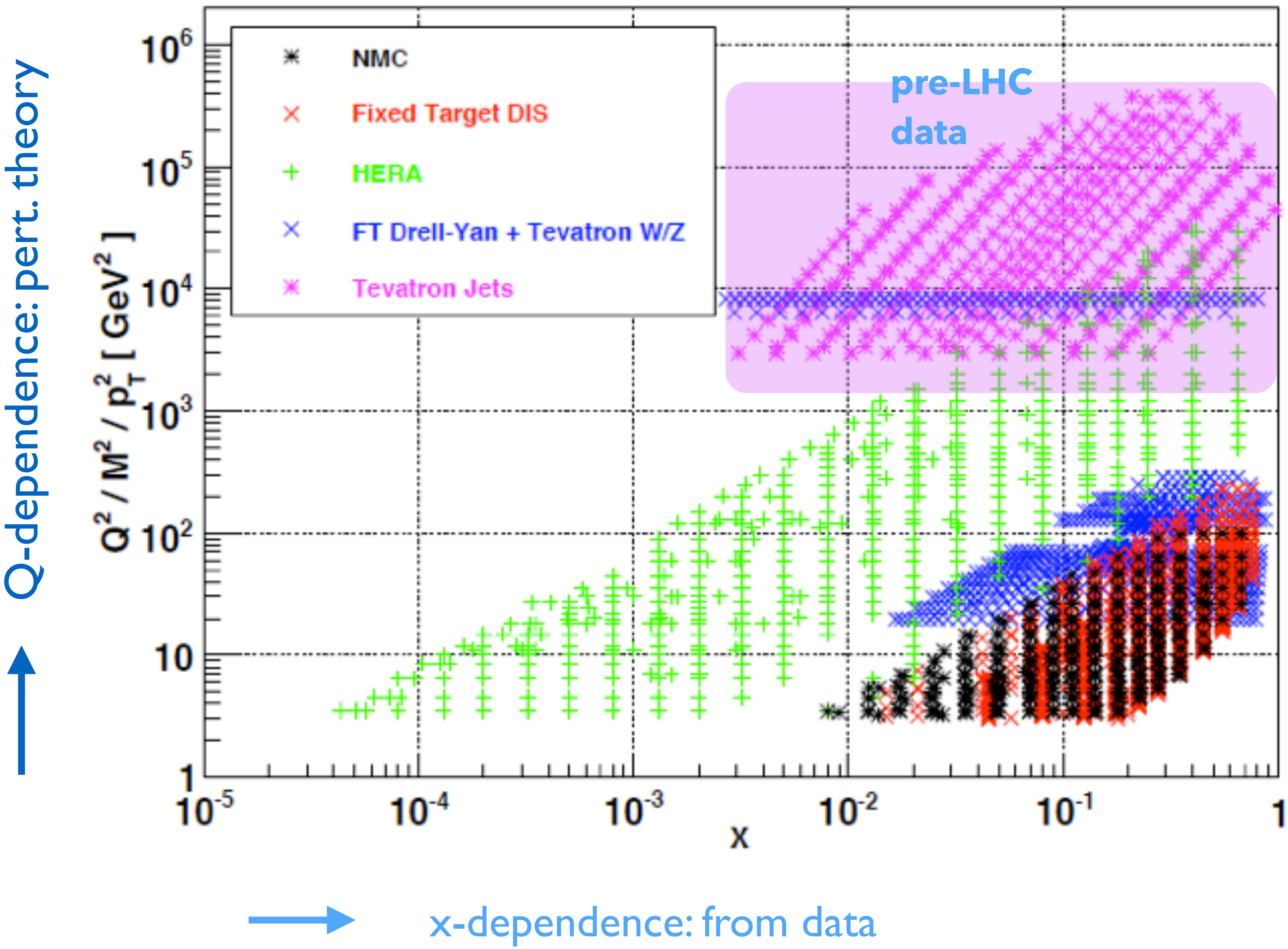
x-dependence: from data

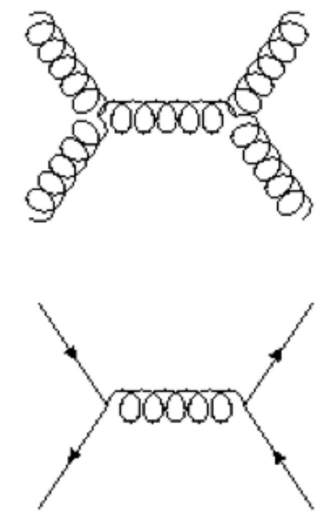


$$x = \frac{Q^2}{2M_n E_\nu y}$$

- Deep inelastic neutrino production of charm
- Constrain strange and anti-strange at moderate x > 10^{-2}

PDF determination





$$x_{1,2} = \frac{p_T}{\sqrt{S}} e^{\pm\eta}$$

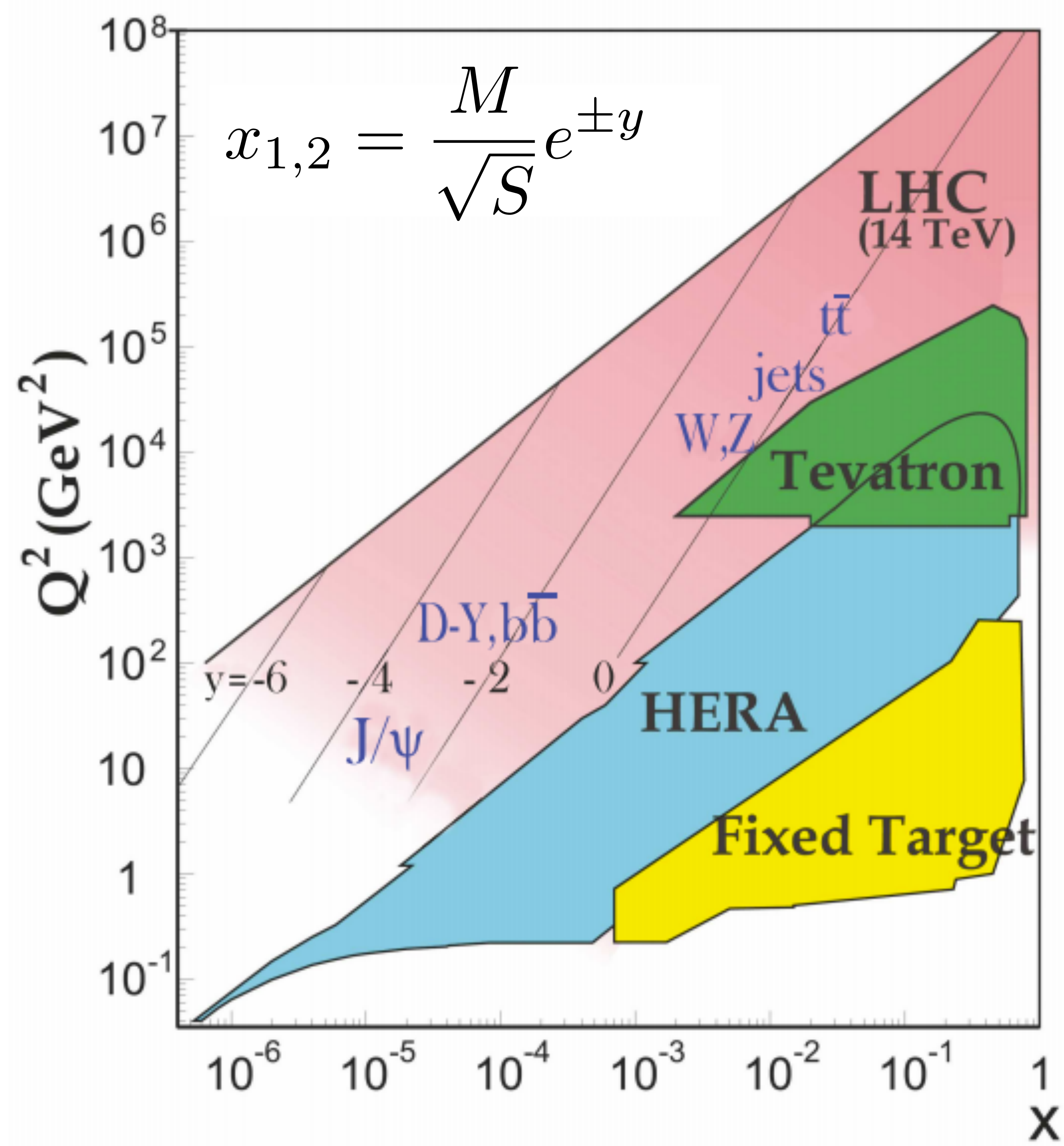
- Jet data
- Direct handle on quarks and gluons at large x

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



PDG 2016

The name of the game

- 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**

The name of the game

- 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**

Not as simple as it may look...

$$\langle \mathcal{O}[\{f\}] \rangle = \int [\mathcal{D}f]\, \mathcal{O}[\{f\}] \mathcal{P}[\{f\}]$$

- Given a finite number of experimental data points want a set of functions
- Want to find a infinite-dimensional object from a finite number of information

A quite complicated game

- 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**

Must propagate data uncertainty into PDF uncertainty. How to deal with inconsistencies?

Hidden uncertainty: still an option?

Parametric versus non-parametric approach

Methodology

Hessian versus MC approach

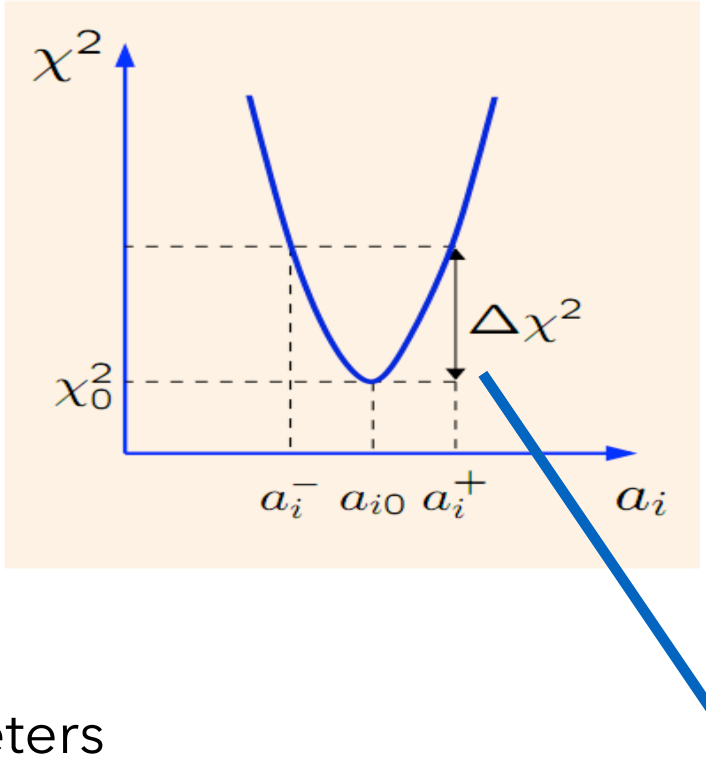
Standard solution

$$\langle \mathcal{O}[\{f\}] \rangle = \int [\mathcal{D}f]\, \mathcal{O}[\{f\}] \mathcal{P}[\{f\}]$$

- Given a finite number of experimental data points want a set of functions with errors
- Want to find a infinite-dimensional object from a finite number of information

Propagation of experimental uncertainty

$$\langle \mathcal{O}[\{f\}] \rangle \simeq \int da_1 da_2 ... da_{N_{par}}\, \mathcal{O}[\vec{a}]\, \mathcal{P}[\vec{a}]$$



Tolerance

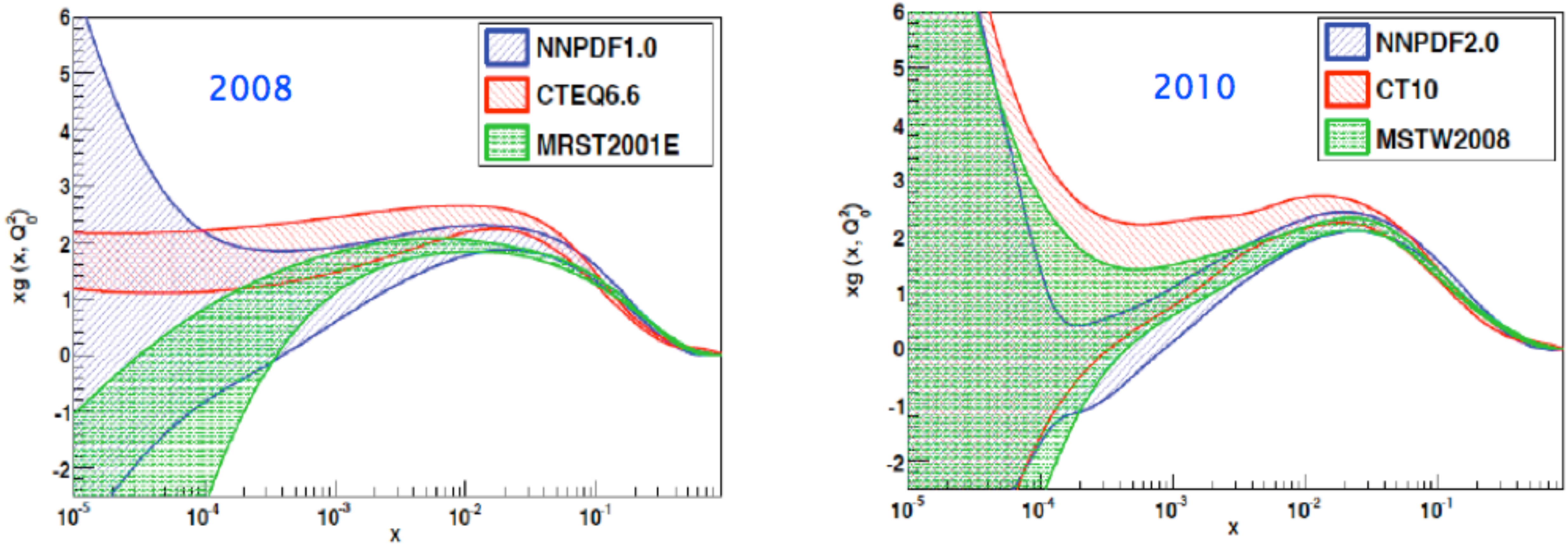
- Hessian approach: Project into a n-dimensional space of parameters and use linear approximation around minimum χ^2

Parametrisation

- Introduce a simple functional form with enough free parameters
- Typically about 20-40 free parameters for 7 independent functions

$$f_i(x, Q_0^2) = a_0\, x^{a_1}\, (1-x)^{a_2} P(x, a_3, a_4, ...),$$

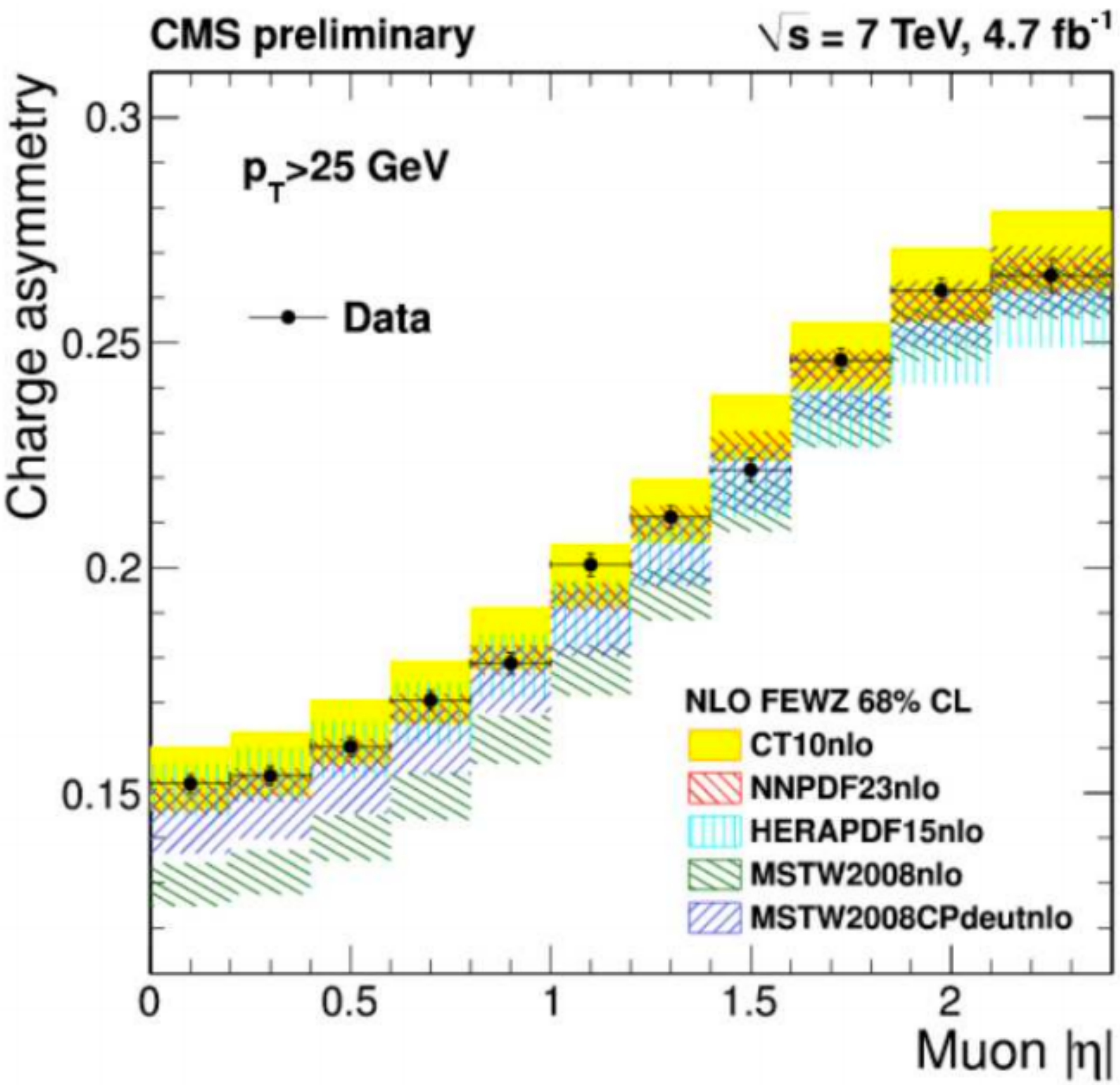
Data-driven progress

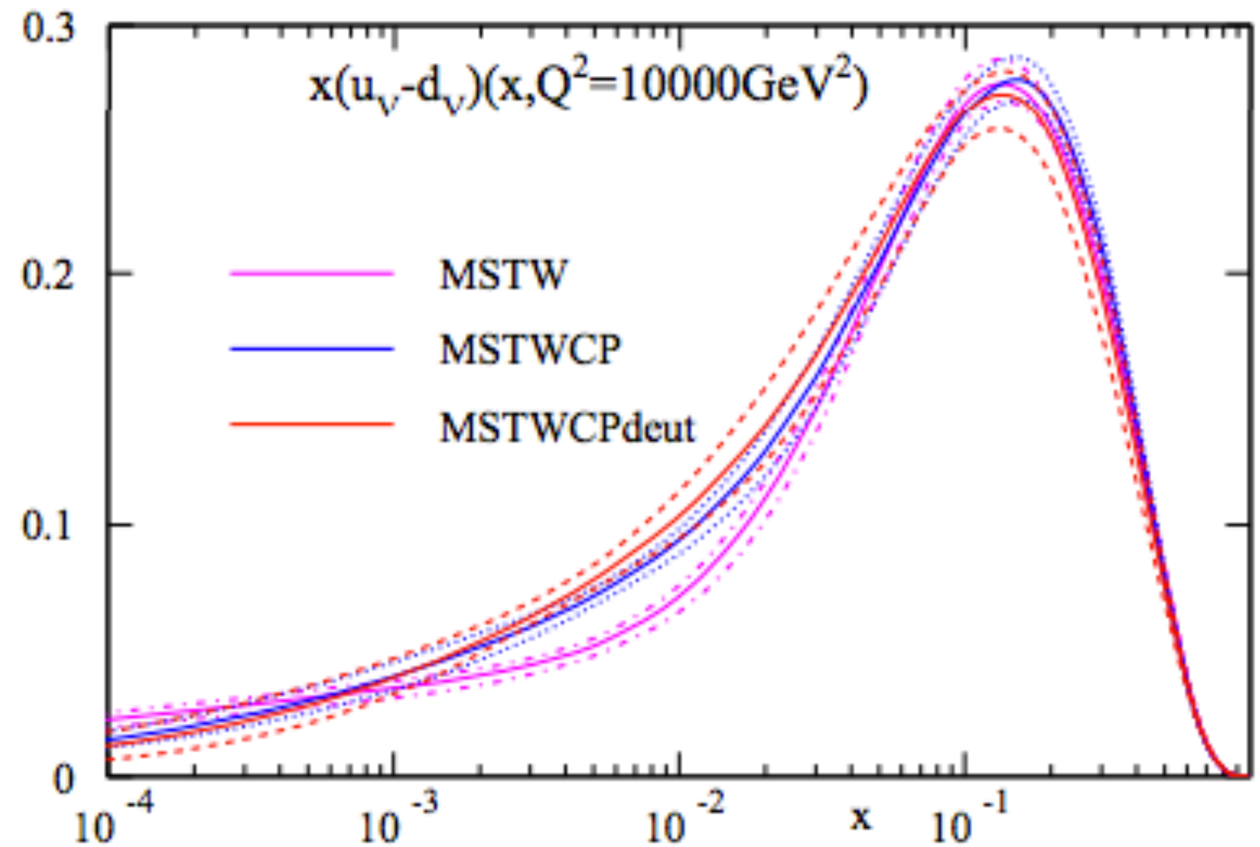


$$xg = A_g\, x^{\delta_g} (1-x)^{\eta_g} (1 + \epsilon_g \sqrt{x} + \gamma_g\, x) + A_{g'}\, x^{\delta_{g'}} (1-x)^{\eta_{g'}}$$

- PDF uncertainties tuned to data (tolerance $\Delta\chi^2 > 1$ - many studies/improvements)
- Fixed parametrisation was forced to be more flexible by new data => less biased parametrisation form (a posteriori data-driven progress)

Data-driven progress





Martin et al
EPJC73 (2013) 2, 2318

- PDF uncertainties tuned to data (tolerance $\Delta\chi^2 > 1$ - many studies/improvements)
- Fixed parametrisation was forced to be more flexible by new data => less biased parametrisation form (a posteriori data-driven progress)

The NNPDF idea

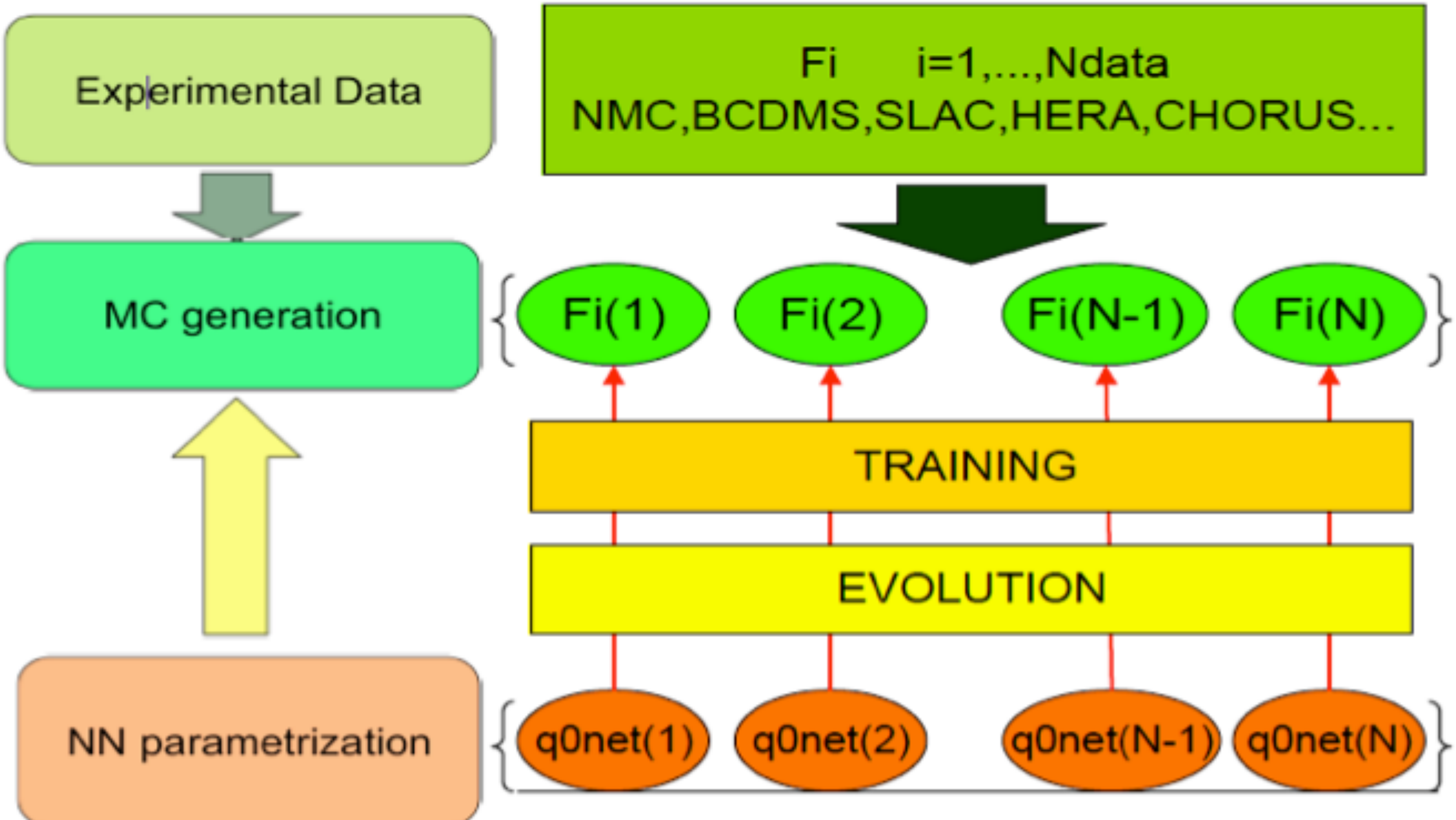
Monte Carlo sampling

Neural Network

http://nnpdf.mi.infn.it

- Fit of structure function **(2005)**
- DIS-only fit of PDFs **(2008)**
- First NNPDF global fit **(2010)**
- First fit including LHC data **(2013)**
- Closure test **(2016)**
- Fitted charm **(2018)**
- …

The NNPDF approach



$$\langle \mathcal{O}[\{f\}] \rangle \simeq \frac{1}{N_{\text{rep}}} \sum_{i=1}^{N_{\text{rep}}} \mathcal{O}[f_i],$$

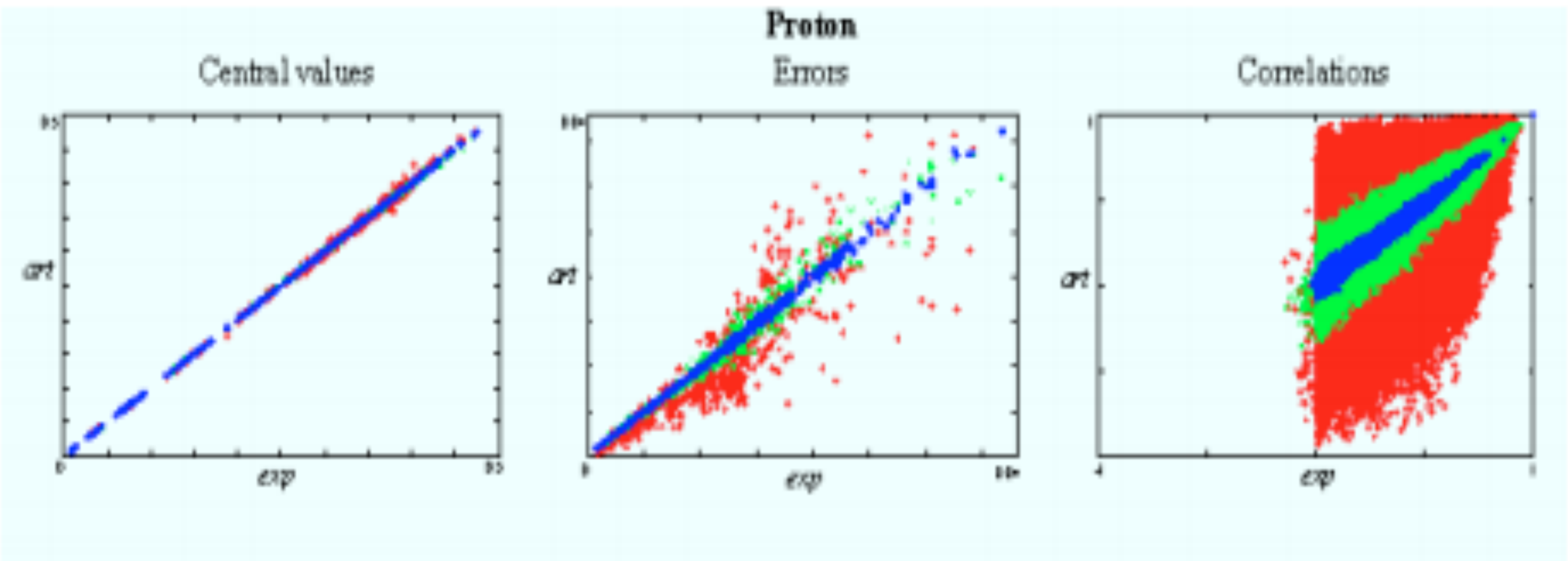
Ball, Del Debbio, Forte, Guffanti, Latorre, Rojo, MU (2008)

Monte Carlo sampling

$$F_p^{(\mathrm{art})(k)} = S_{p,N}^{(k)} F_p^{(exp)} \left(1 + \sum_{l=1}^{N_c} r_{p,l}^{(k)} \sigma_{p,l} + r_p^{(k)} \sigma_{p,s} \right)$$

Normalisation

- Monte Carlo techniques: sampling the probability measure in PDF functional space projecting from space of experimental data

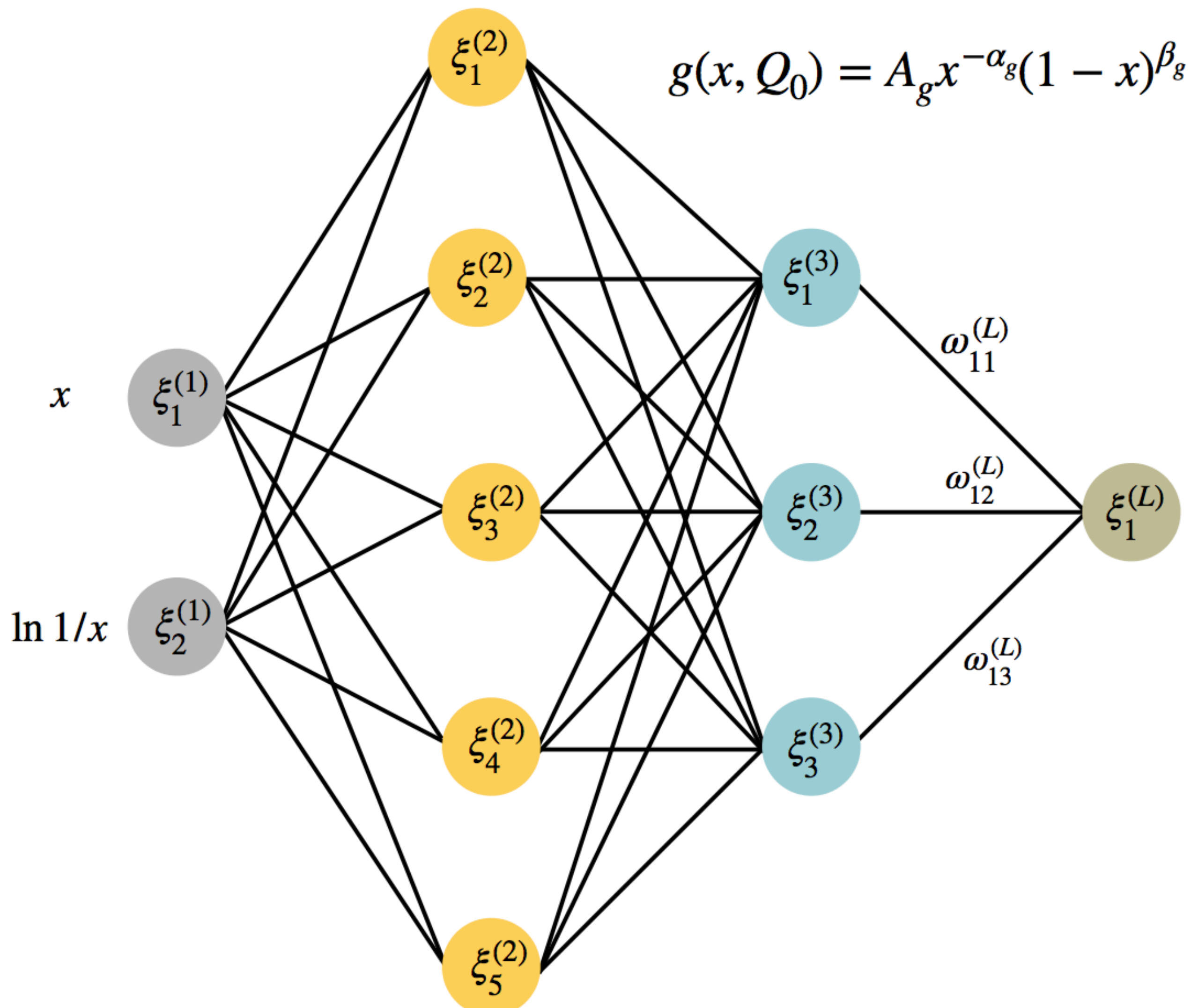


$$\langle F \rangle = \frac{1}{N_{\mathrm{rep}}} \sum_{i}^{N_{\mathrm{rep}}} F^{(k)},$$

$$\sigma^2[F] = \frac{1}{N_{\mathrm{rep}} - 1} \sum_{i=1}^{N_{\mathrm{rep}}} (F^{(k)} - \langle F \rangle)^2,$$

Neural network parametrisation

$$g(x, Q_0) = A_g x^{-\alpha_g}(1-x)^{\beta_g}\,\xi_L^{(1)}(x)$$



For a 1-2-1 feedforward neural network can write explicitly functional form

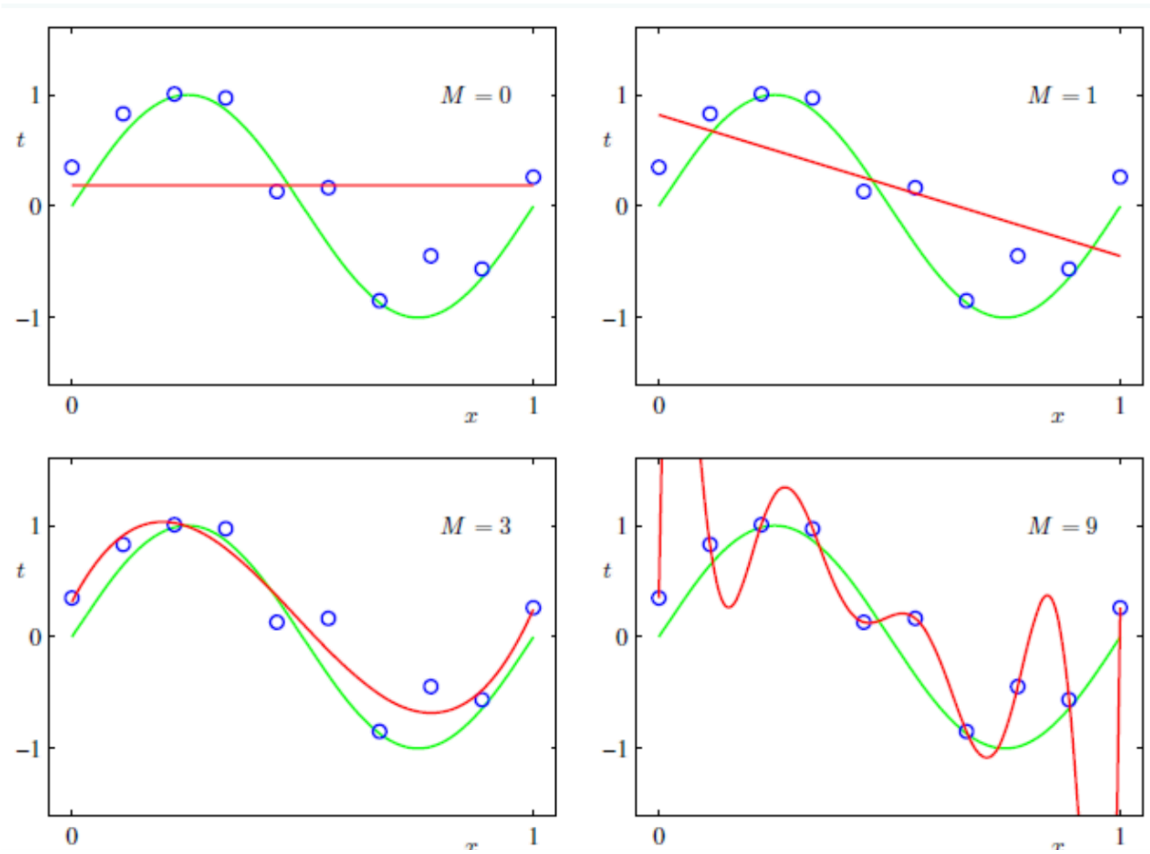
$$\xi_1^{(3)}(\xi_1^{(1)}) = \frac{1}{1+e^{\theta_1^{(3)} - \frac{\omega_{11}^{(2)}}{1+e^{\theta_1^{(2)} - \xi_1^{(1)}\omega_{11}^{(1)}}} - \frac{\omega_{12}^{(2)}}{1+e^{\theta_2^{(2)} - \xi_1^{(1)}\omega_{21}^{(1)}}}}}$$

- Neural Networks: all independent PDFs are associated to an unbiased and flexible parametrization: O(300) parameters versus O(30) in polynomial parametrization
- 2-5-3-1 Neural network associated to each independent PDF (gluon, up, anti-up, down, anti-down, strange, anti-strange and charm)

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

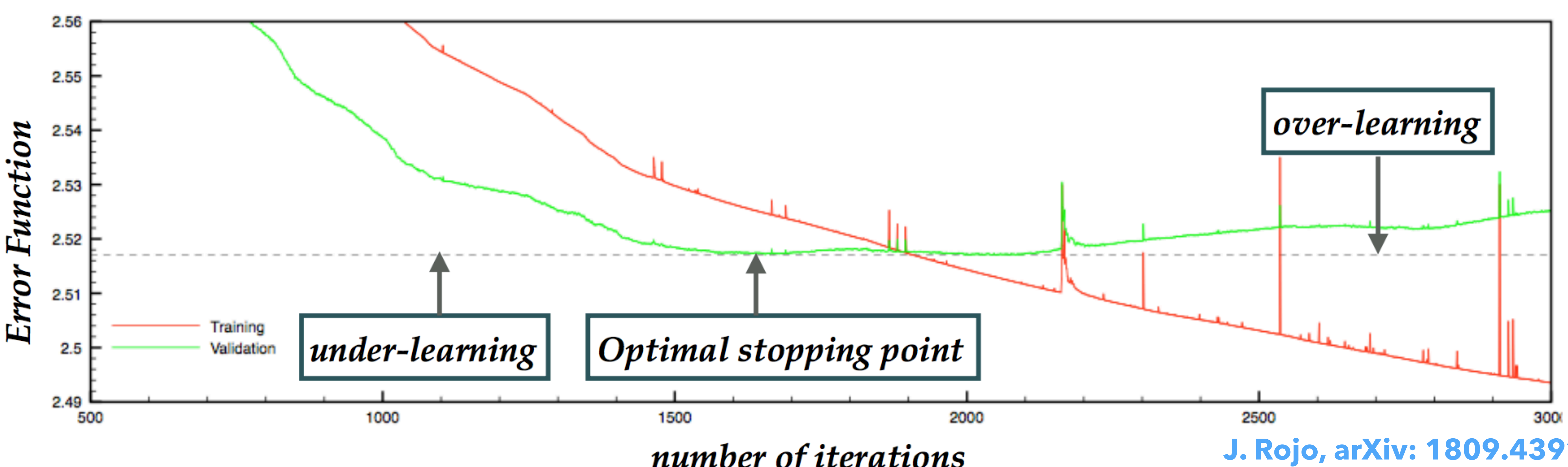
$$g(x) = \frac{1}{1+e^{-x}}$$

Minimisation



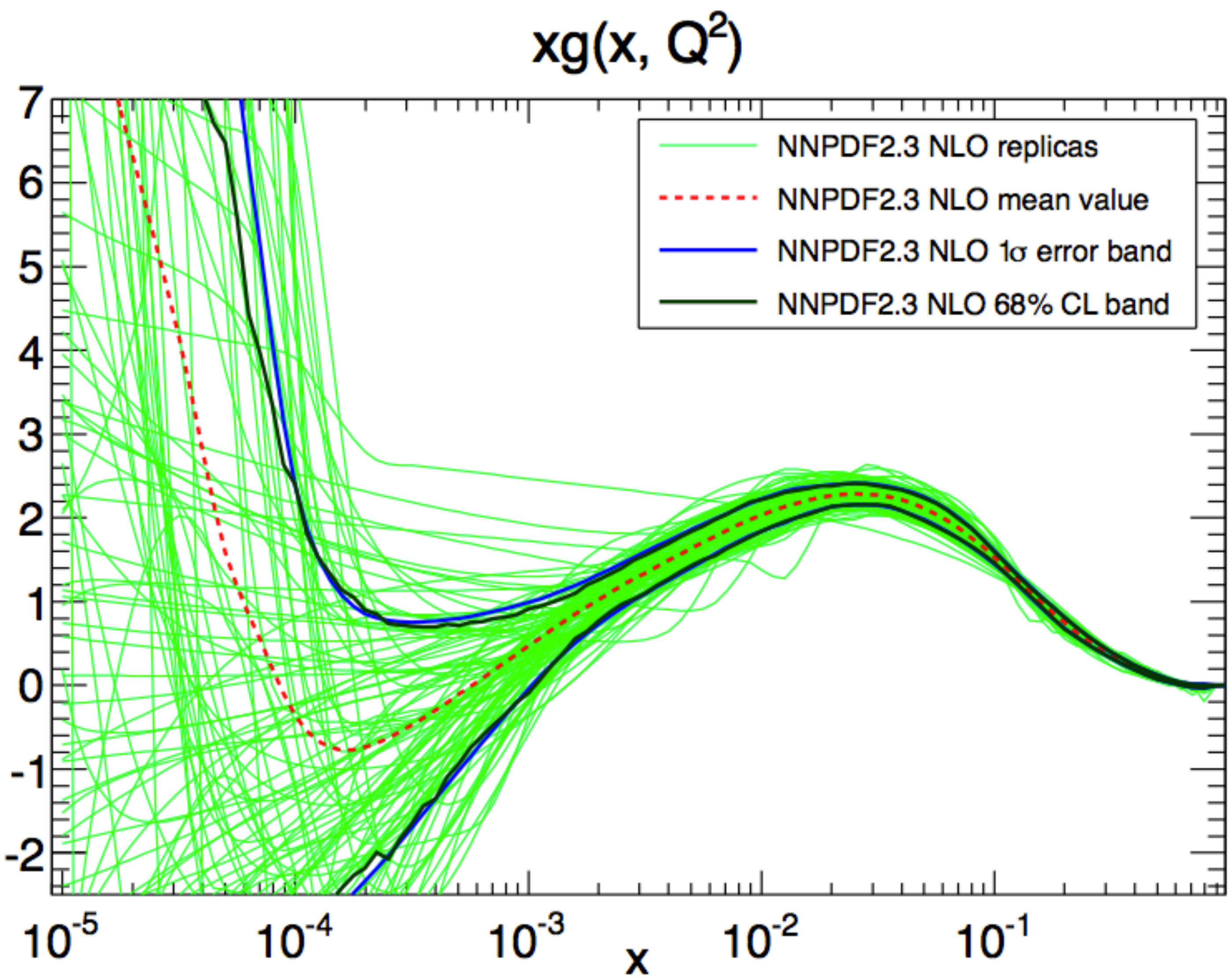
- Large parameter space: need an algorithm that is able to explore it without getting trapped in local minima such as genetic algorithm

- Redundant parametrization: risk of over-fitting. Cross-validation necessary.



J. Rojo, arXiv: 1809.4392

The NNPDF solution

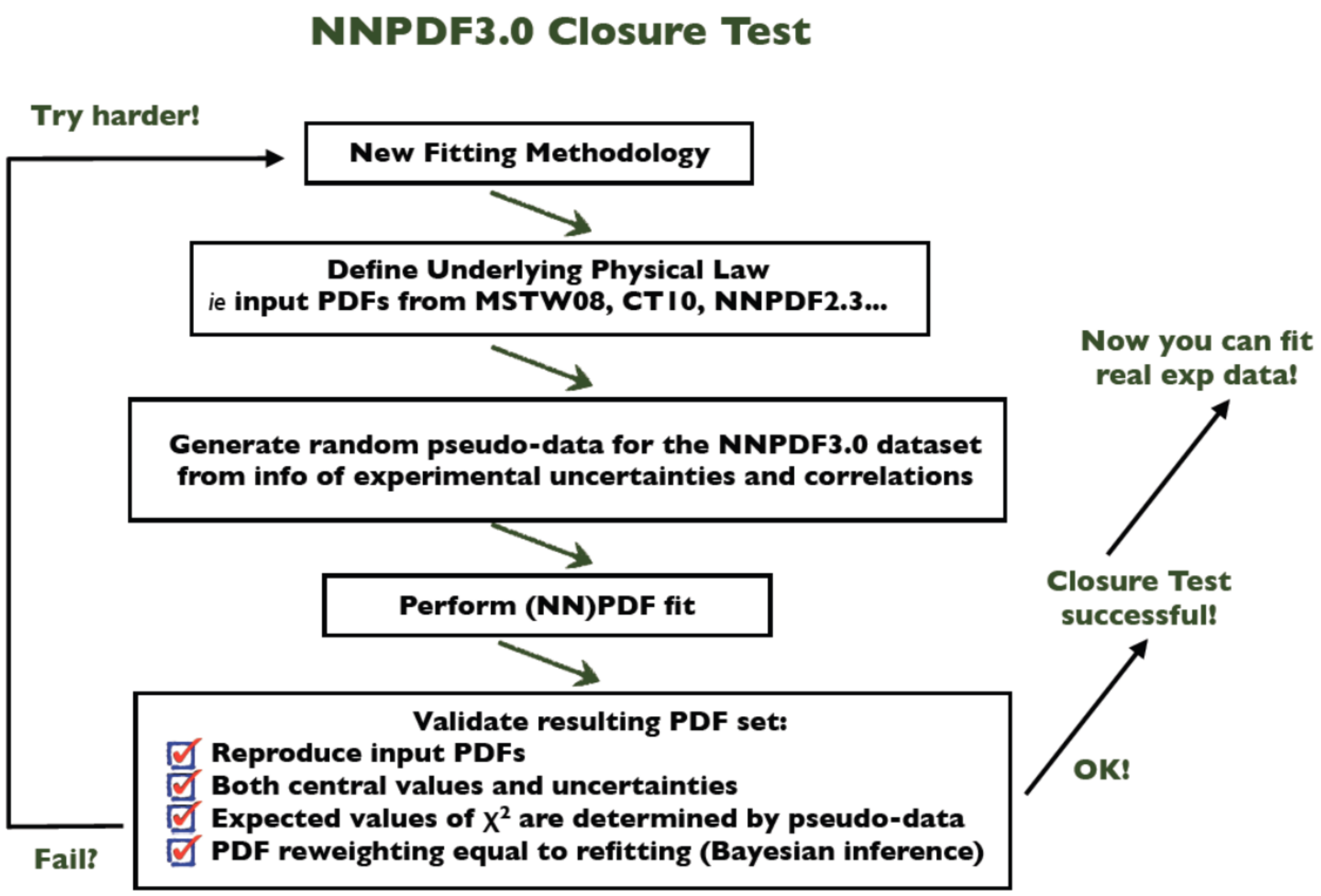


The N(eural)N(etwork)PDFs:

- Monte Carlo techniques: sampling the probability measure in PDF functional space
- Neural Networks: all independent PDFs are associated to an unbiased and flexible parametrization: O(300) parameters versus O(30) in polynomial parametrization
- Genetic algorithm and cross-validation methods

✓ Precise error estimate not driven by theoretical prejudice
✓ No need to add new parameters when new data are included
✓ Statistical interpretation of uncertainty bands
✓ Possibility to include data via re-weighting: no need to refit

Validation


NNPDF3.0 Closure Test
Try harder!
New Fitting Methodology
Define Underlying Physical Law
ie input PDFs from MSTW08, CT10, NNPDF2.3...
Generate random pseudo-data for the NNPDF3.0 dataset
from info of experimental uncertainties and correlations
Perform (NN)PDF fit
Validate resulting PDF set:
Reproduce input PDFs
Both central values and uncertainties
Expected values of χ² are determined by pseudo-data
PDF reweighting equal to refitting (Bayesian inference)
Fail?
OK!
Closure Test
successful!
Now you can fit
real exp data!

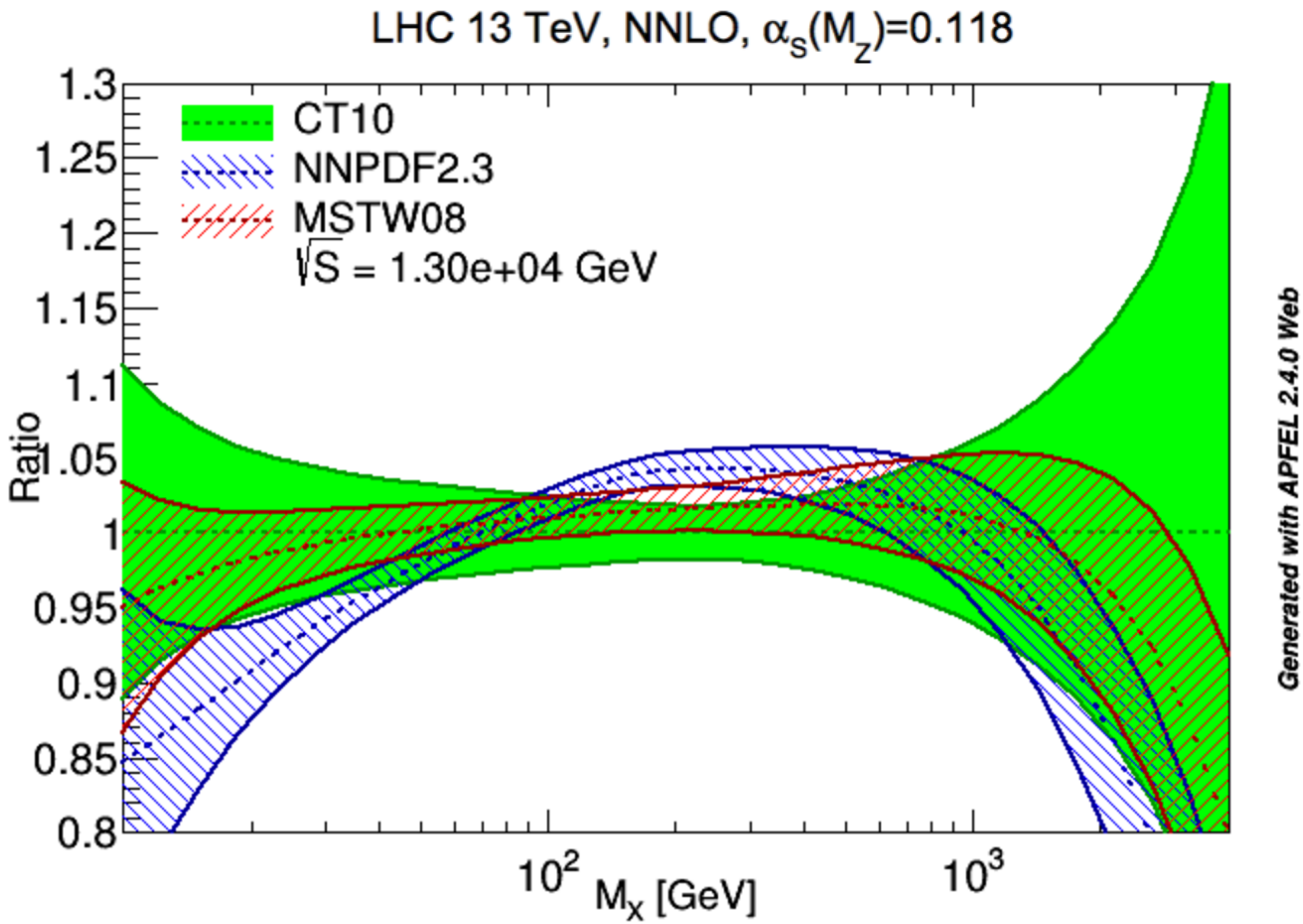
State of the art and frontiers

The players

June 2017	NNPDF3.0 ➜ 3.1	MMHT2014	CT14	ABMP16
Fixed Target DIS	✔	✔	✔	✔
HERA I+II	✔	✔	✔	✔
HERA jets	✘	✔	✘	✘
Fixed Target DY	✔	✔	✔	✔
Tevatron W,Z	✔	✔	✔	✔
Tevatron jets	✔	✔	✔	✘
LHC jets	✔	✔	✔	✘
LHC vector boson	✔	✔	✔	✔
LHC top	✔	✘	✘	✔
Stat. treatment	Monte Carlo	Hessian $\Delta\chi^2$ dynamical	Hessian $\Delta\chi^2$ dynamical	Hessian $\Delta\chi^2=1$
Parametrization	Neural Networks (259 pars)	Chebyshev (37 pars)	Bernstein (30-35 pars)	Polynomial (15 pars)
HQ scheme	FONLL	TR'	ACOT-χ	FFN (+BMST)
Order	NLO/NNLO	NLO/NNLO	NLO/NNLO	NLO/NNLO

Gluon luminosity

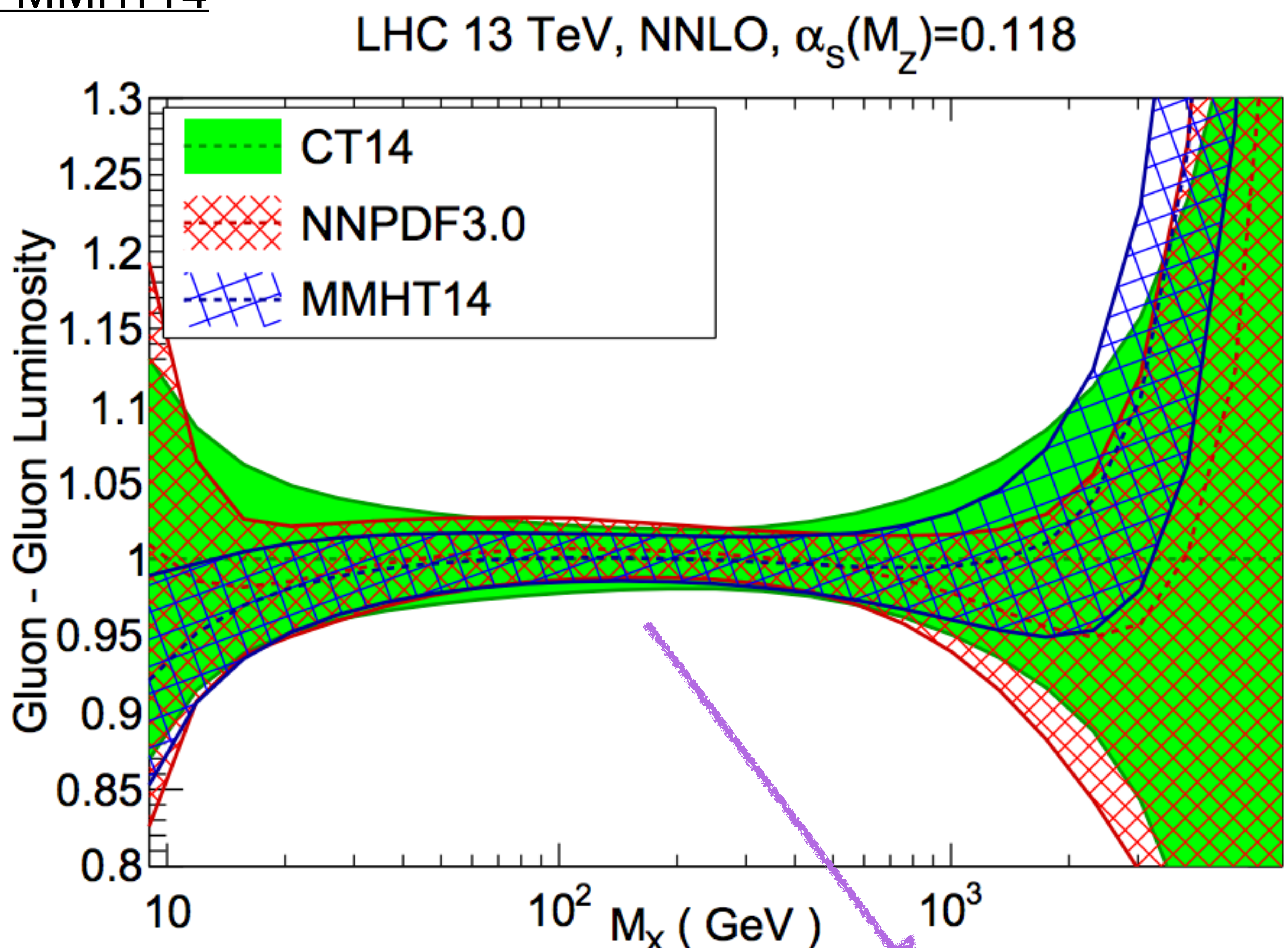
NNPDF2.3 / CT10 / MSTW2008



(2014)

Gluon luminosity

NNPDF3.0 / CT14 / MMHT14

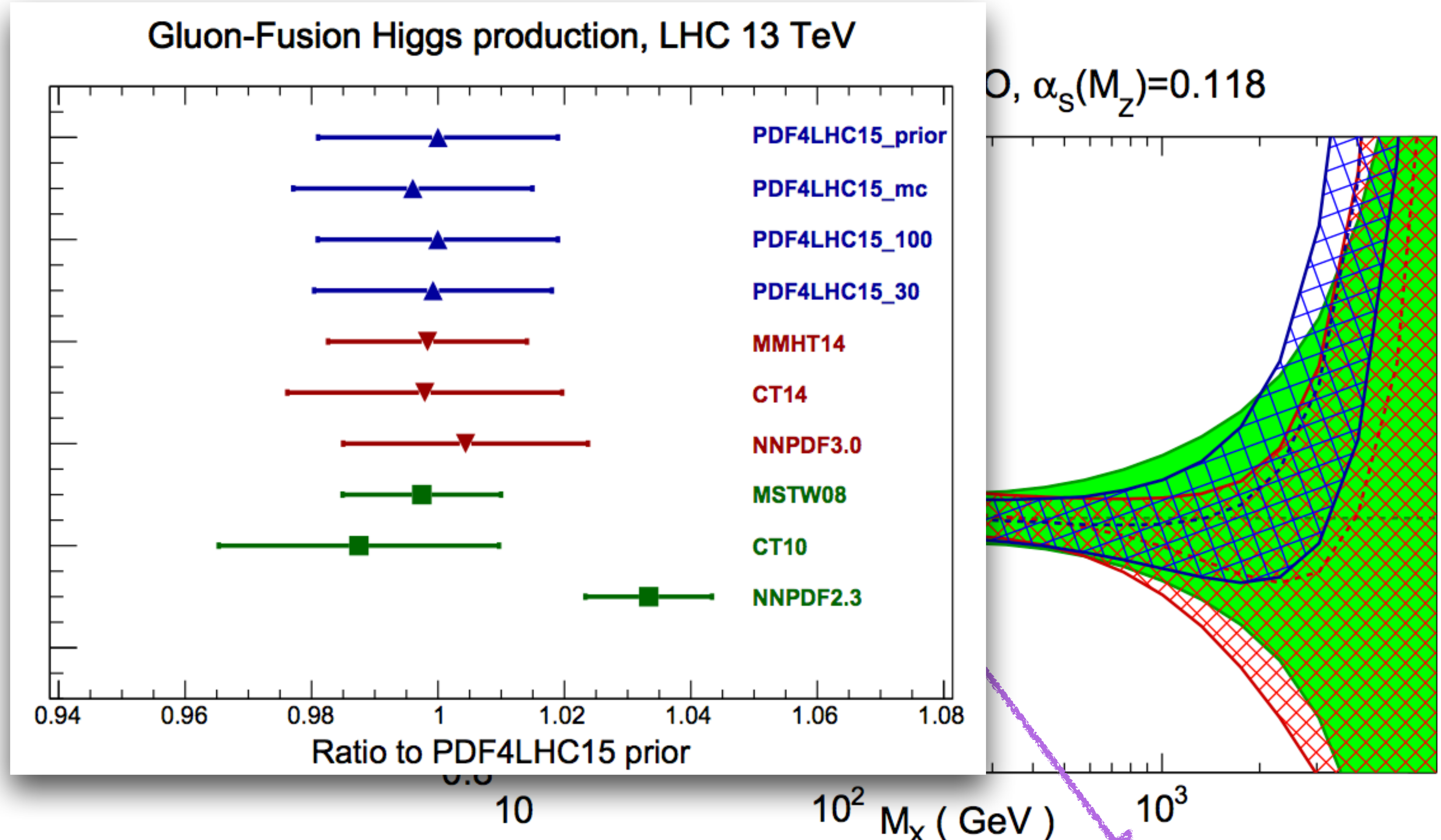


Impact on Higgs physics

(2016)

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

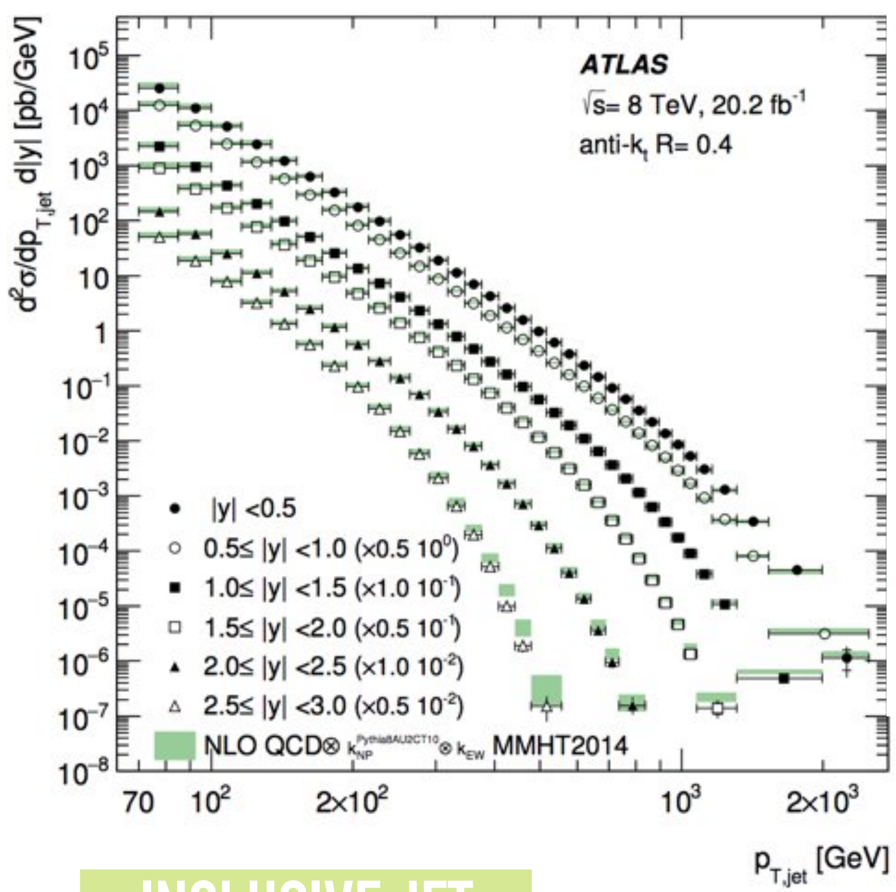
Consequence: Higgs physics



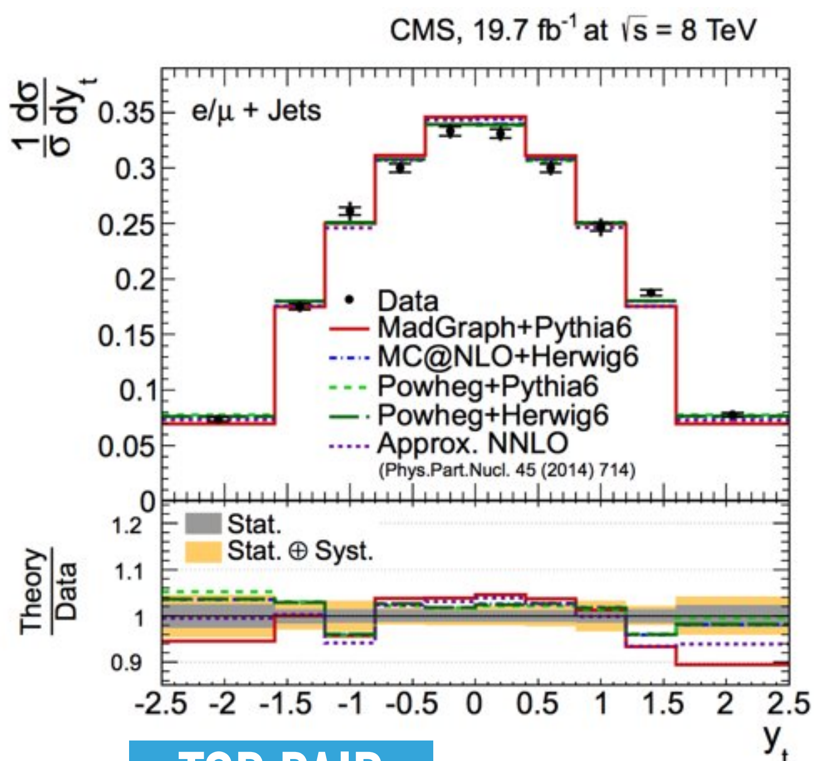
Impact on Higgs physics

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

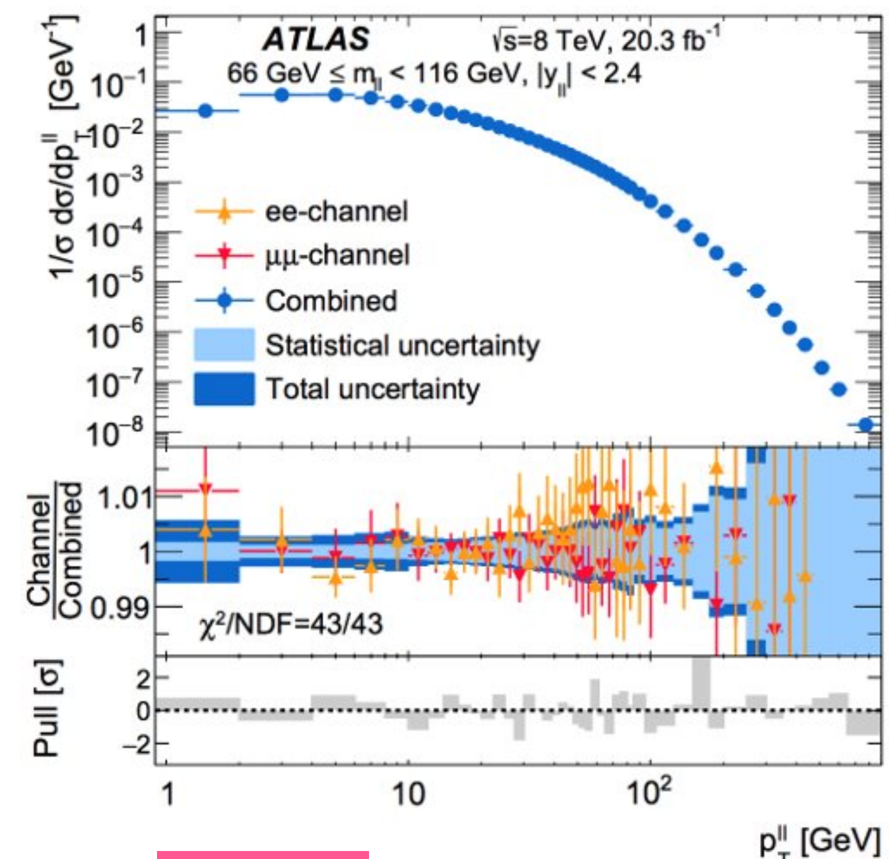
Update: LHC Run I data



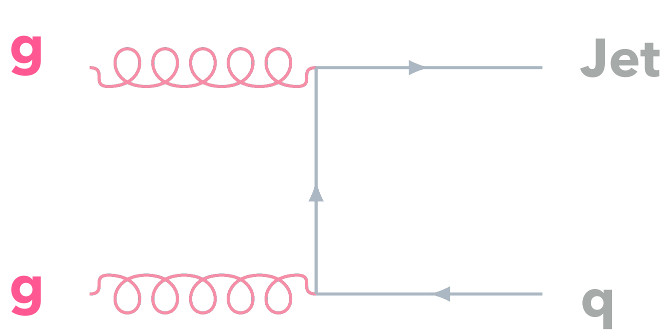
INCLUSIVE JET

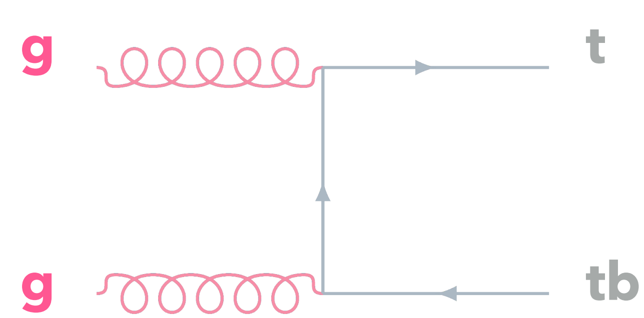


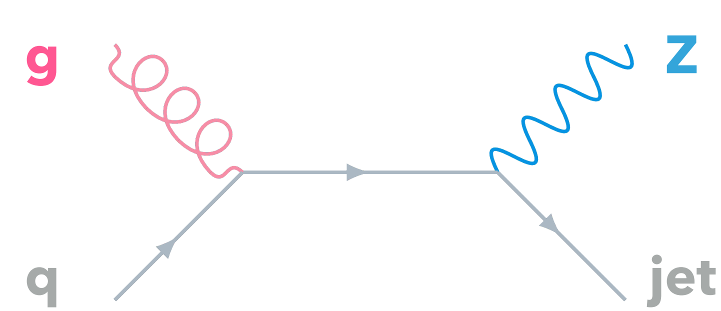
TOP PAIR



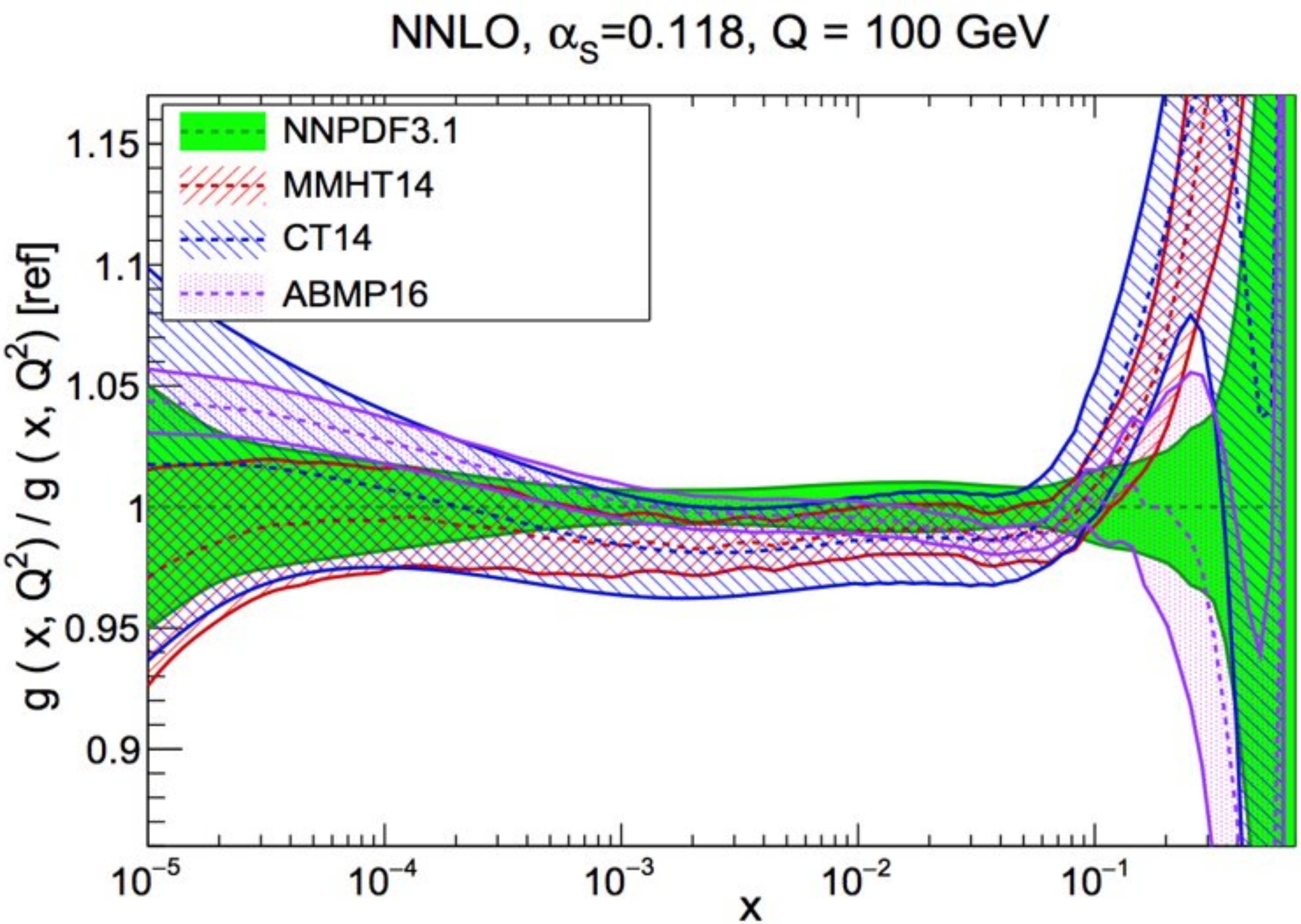
Z P$_T$







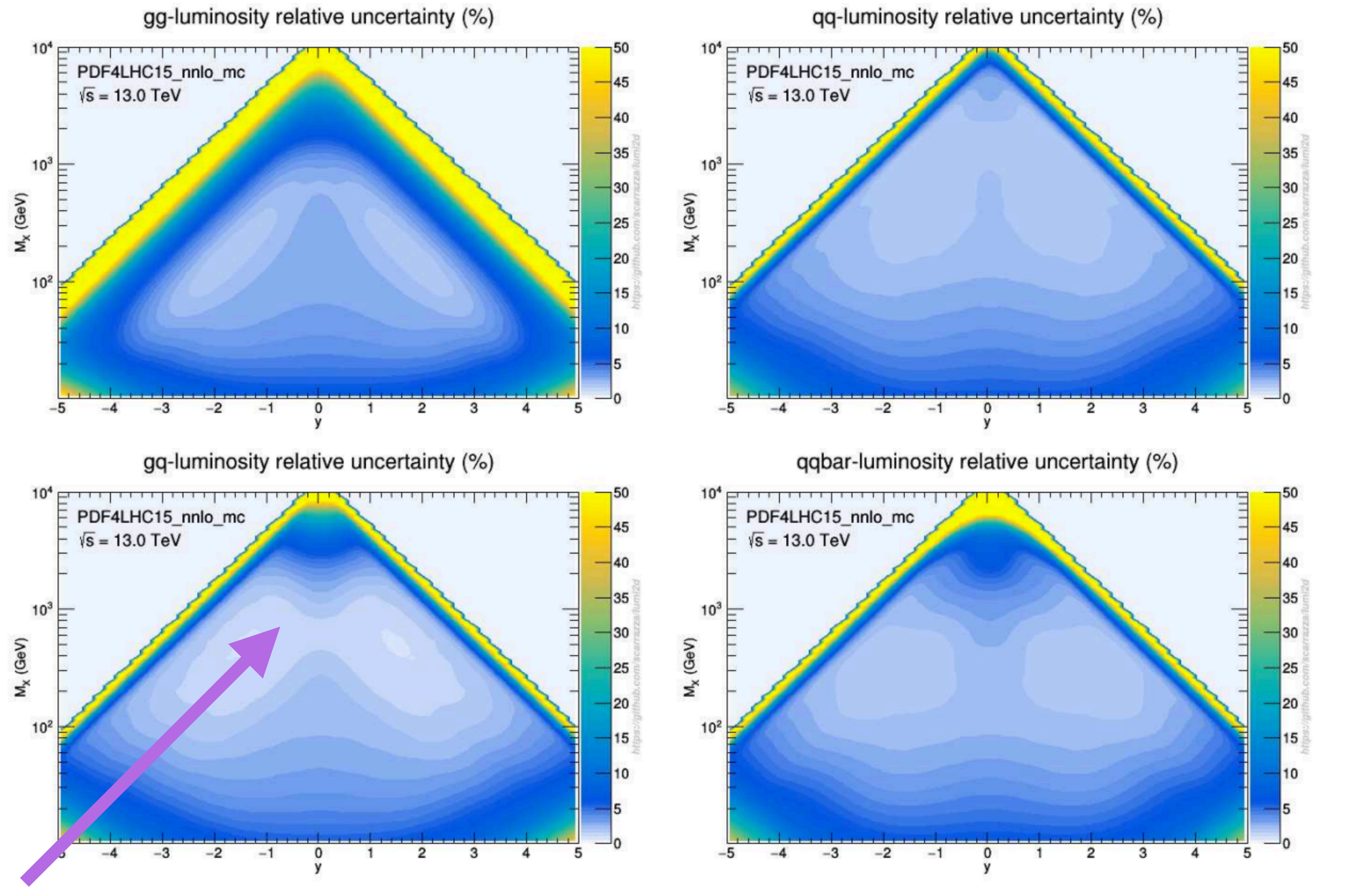
Update: LHC Run I data



Gao et al. Phys.Rept. 742 (2018) 1-121

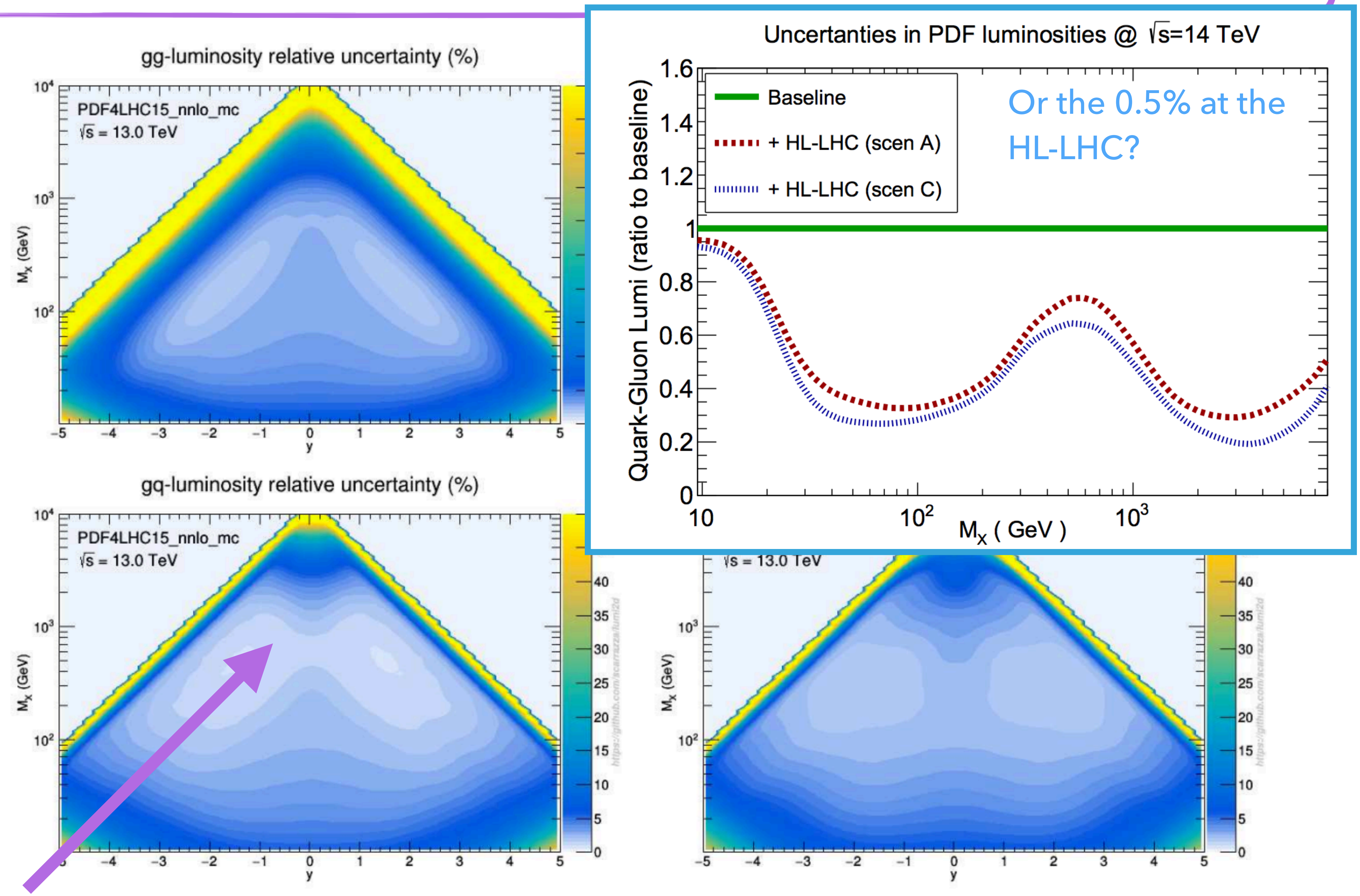
NNPDF3.1 gluon shifted upwards and smaller uncertainty due to LHC Run I data
New combination of PDF sets foreseen later this year

The precision frontier



Can we trust 1% accuracy?

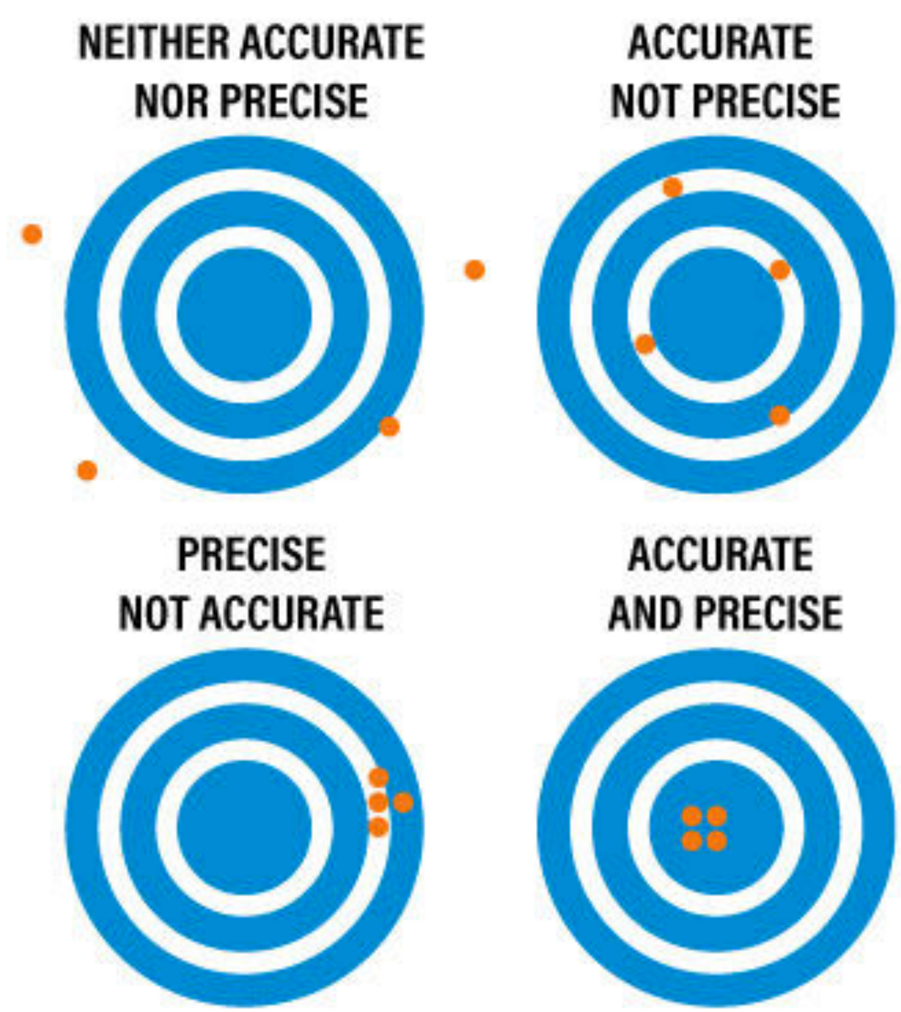
The precision frontier

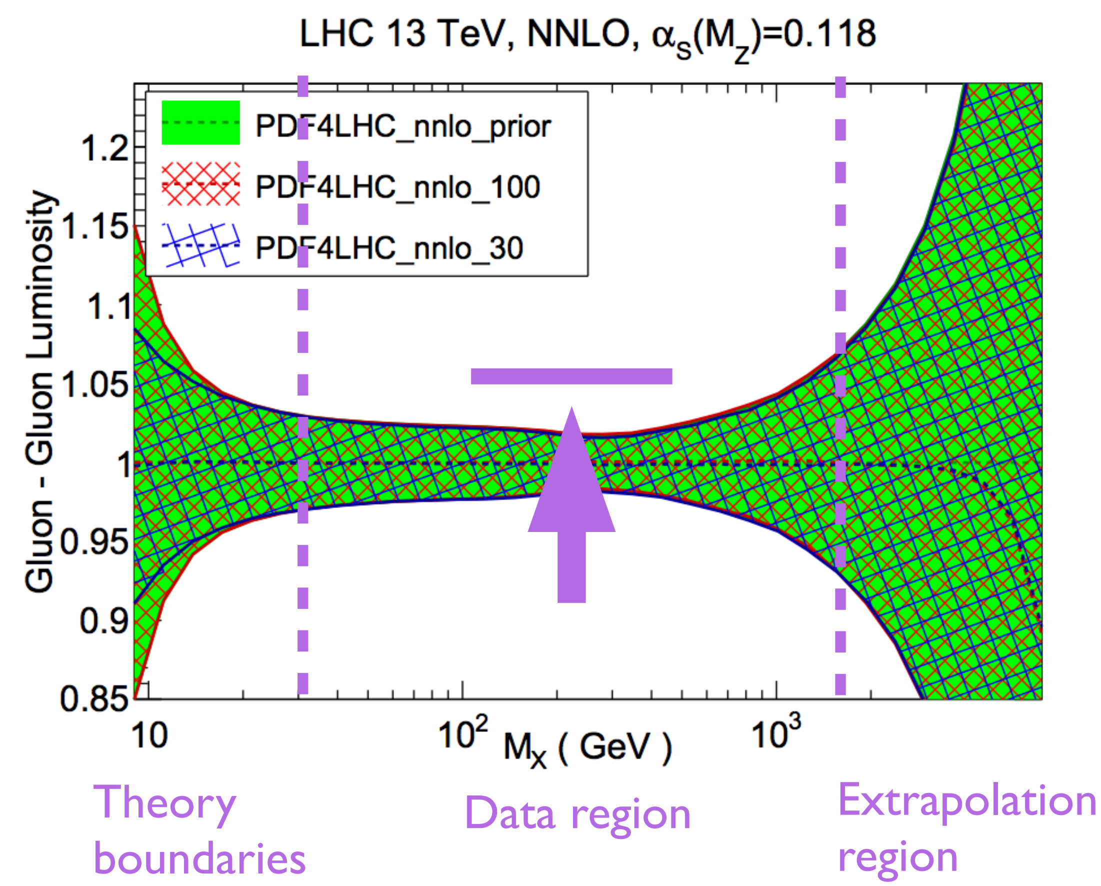


Can we trust 1% accuracy?

Theory uncertainties

In updated PDF analysis, shift between old and new set may be larger than PDF uncertainties

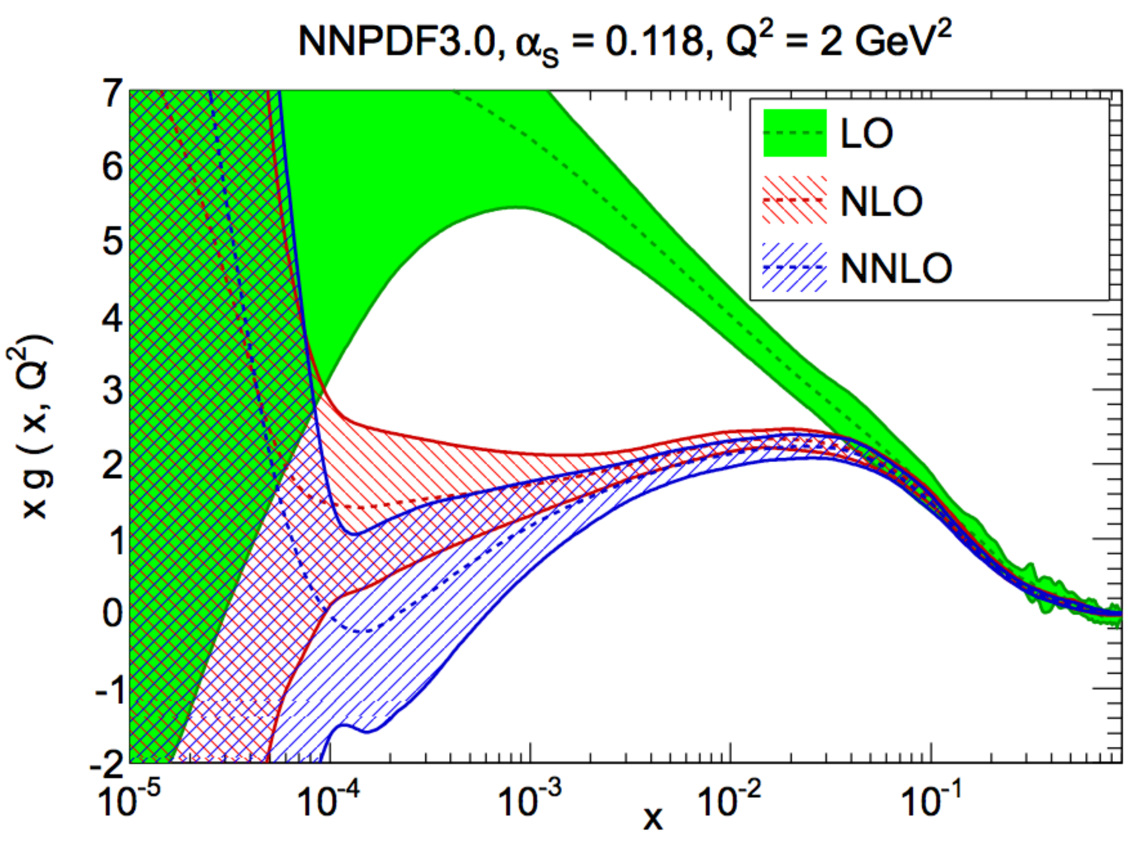


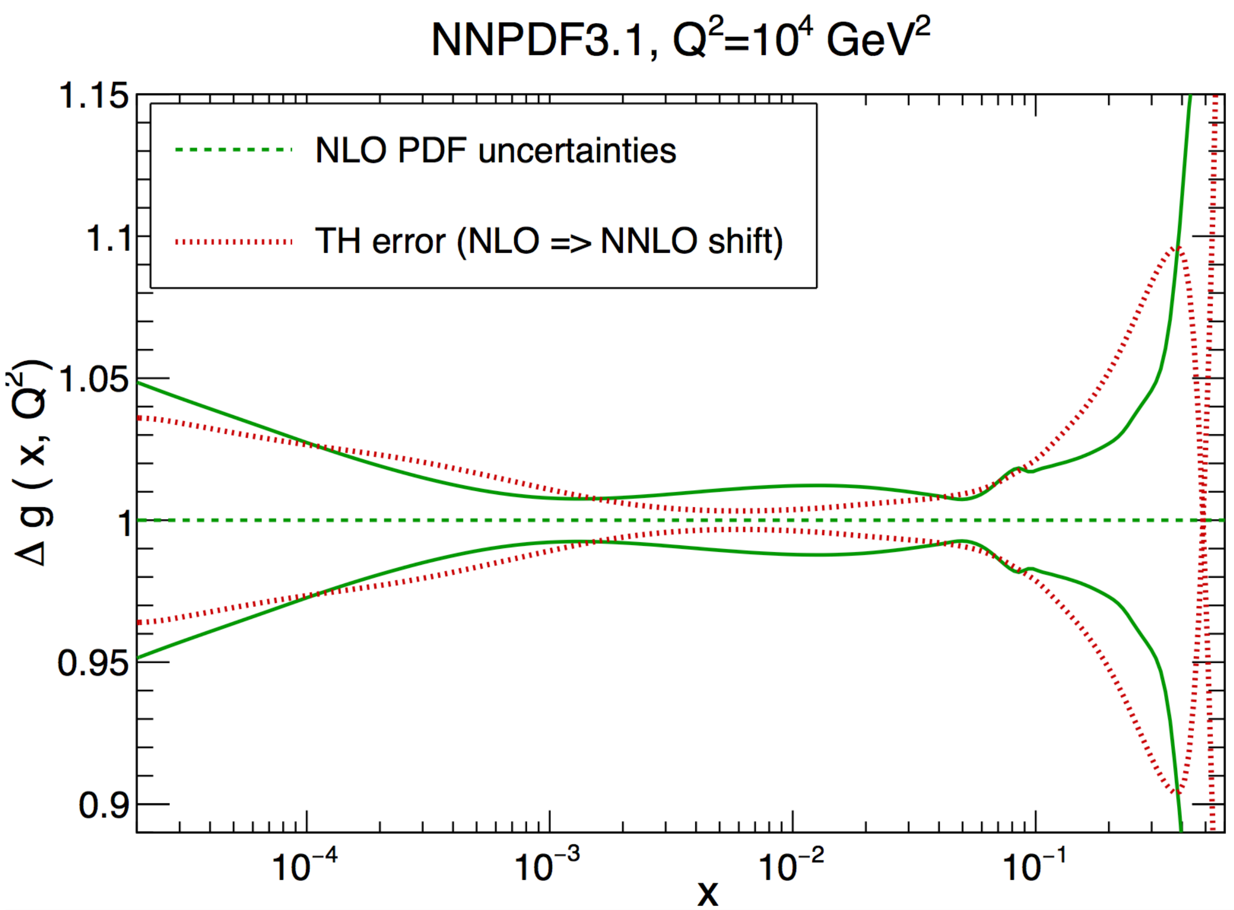


- Inconsistent data → Tolerance/ Statistical estimators
- Updated parametrization
- Differences in fitting methodology/minimisation? → Closure Test
- <u>**Changes in theory?**</u>

Theory uncertainties

- PDF fits performed at given perturbative order
- PDF uncertainties only reflect lack of information from data
- Theoretical uncertainties (dominated by MHOU) ignored so far
- At NLO PDF uncertainties and MHOU comparable
- Near future: NNLO PDF uncertainties will go down to level of MHOU
- Inclusion of theory uncertainties is the next frontier



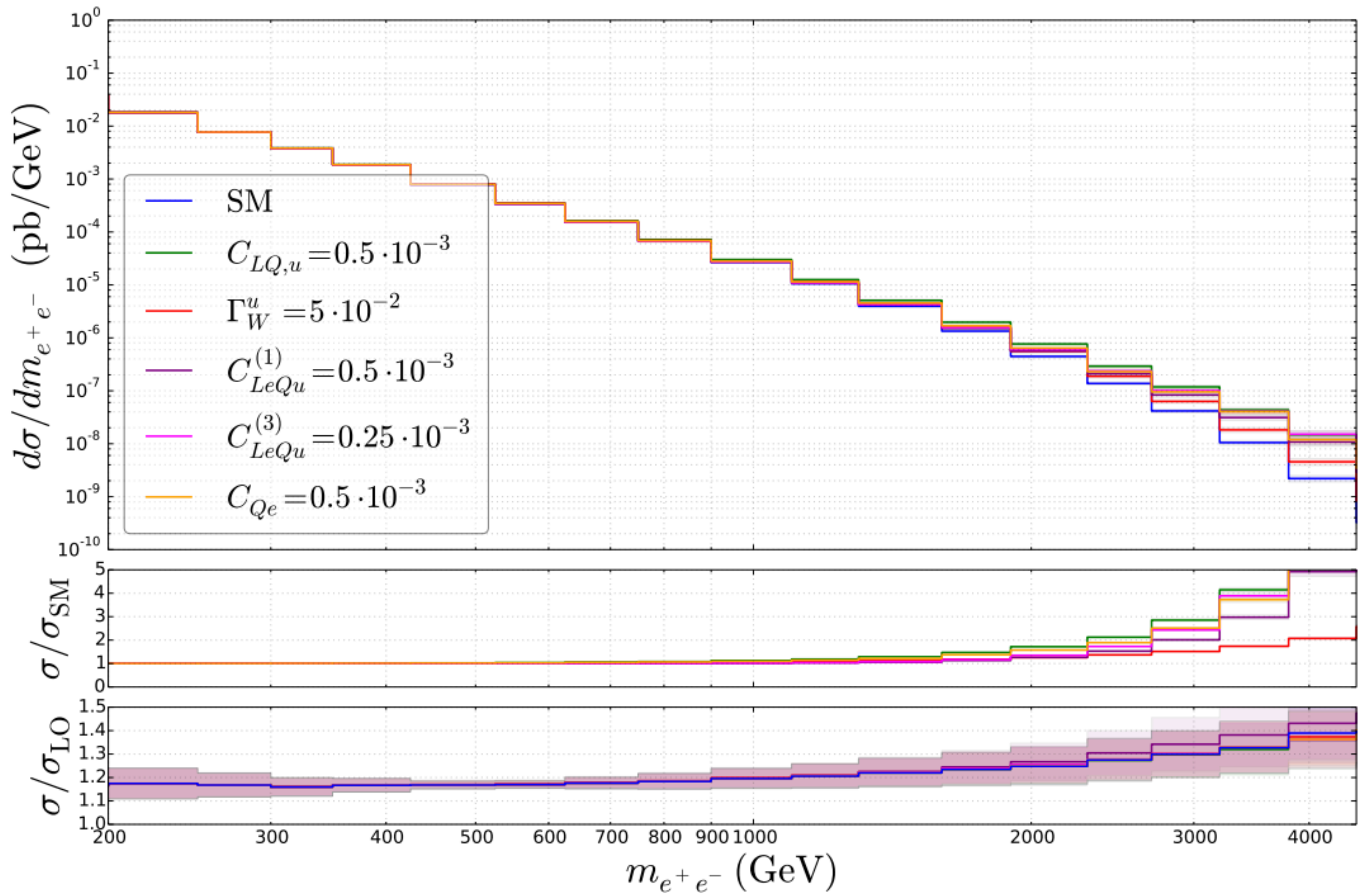


Ball et al, EPJC 77 (2017)

PDFs and new physics

$$\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \mathcal{L}^{(5)} + \mathcal{L}^{(6)} + \mathcal{L}^{(7)} + ..., \qquad \mathcal{L}^{(d)} = \sum_{i=1}^{n_d} \frac{C_i^{(d)}}{\Lambda^{d-4}} Q_i^{(d)} \quad \text{for } d > 4$$

- Many studies analyse effect of higher-dimensional operators on observables measured at the LHC
- Extract constraints on dim-6 operators that contribute to NC and CC Drell-Yan production at the LHC, H+V production and VFB



Alioli et al, 1804.07407

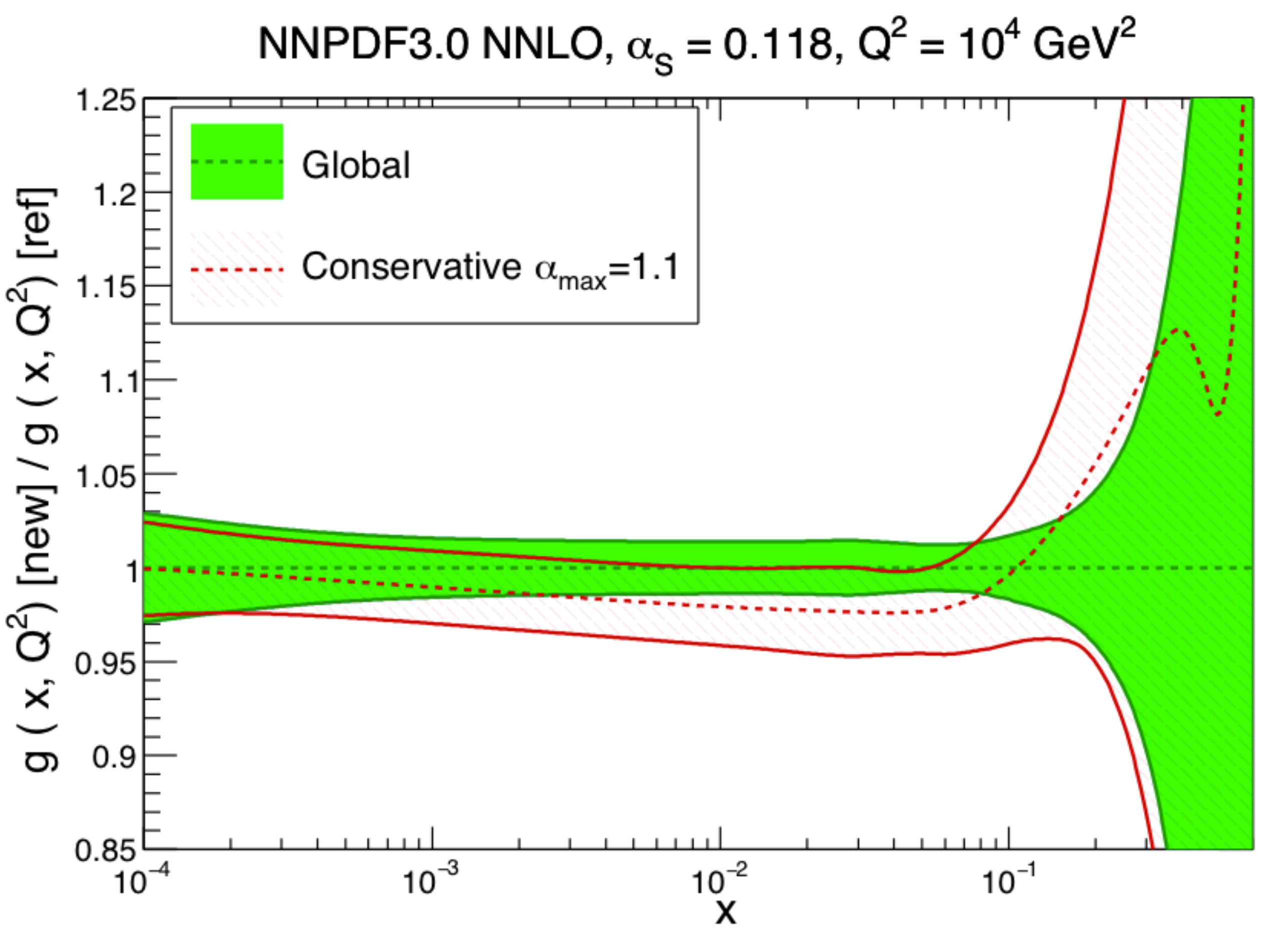
PDFs use some of these data and are determined within SM Framework

PDFs and new physics

- As more data at higher energy will be released, how can we make sure that new physics effects are not absorbed in the PDFs?

- If effects were big we would have bad chi2 and signs of inconsistency but probably would show up as mild inconsistencies

- Inconsistency of any individual dataset with the bulk of global fit may suggest its understanding is incomplete but might be due to many factors



NNPDF3.0, JHEP 1504 (2015) 040

Are conservative partons the answer?
- Not really: simultaneous fits of EFT coefficients and PDFs is the new frontier

Conclusions

- ➡ Precise parton distribution functions are key to reach precision in theory predictions at the LHC
- ➡ After years of effort to reduce PDF uncertainties, thanks to inclusion of many precise new data, we now worry that uncertainties might be <u>under</u>estimated
- ➡ New fascinating challenges open up
- ➡ Neglecting theory uncertainties in PDF fits will not be an option quite soon, many ongoing developments and tests
- ➡ Time to make PDF and SM-EFT talk to each other

Conclusions

➡ Precise parton distribution functions are key to reach precision in theory predictions at the LHC

➡ After years of effort to reduce PDF uncertainties, thanks to inclusion of many precise new data, we now worry that uncertainties might be <u>under</u>estimated

➡ New fascinating challenges open up

➡ Neglecting theory uncertainties in PDF fits will not be an option quite soon, many ongoing developments and tests

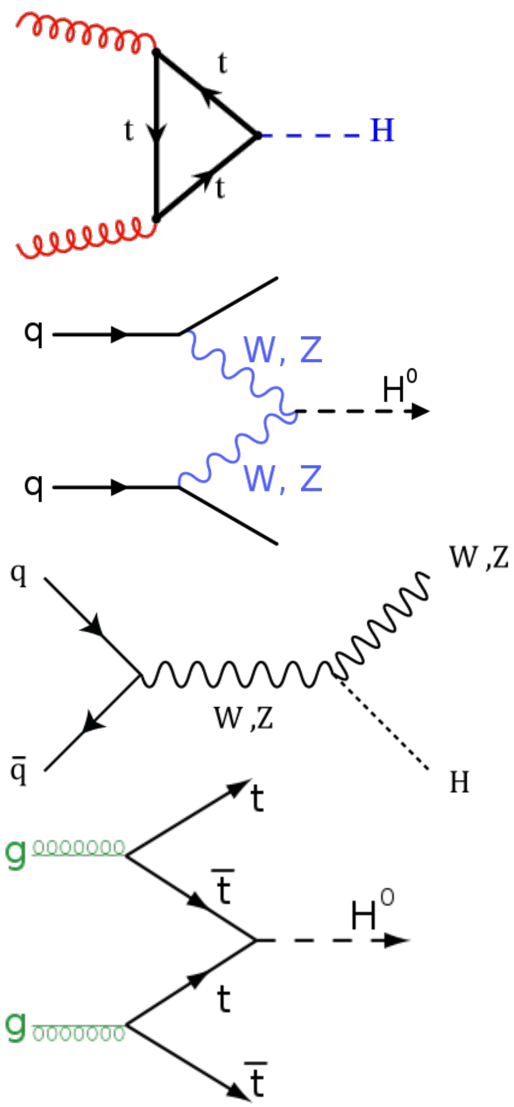
➡ Time to make PDF and SM-EFT talk to each other

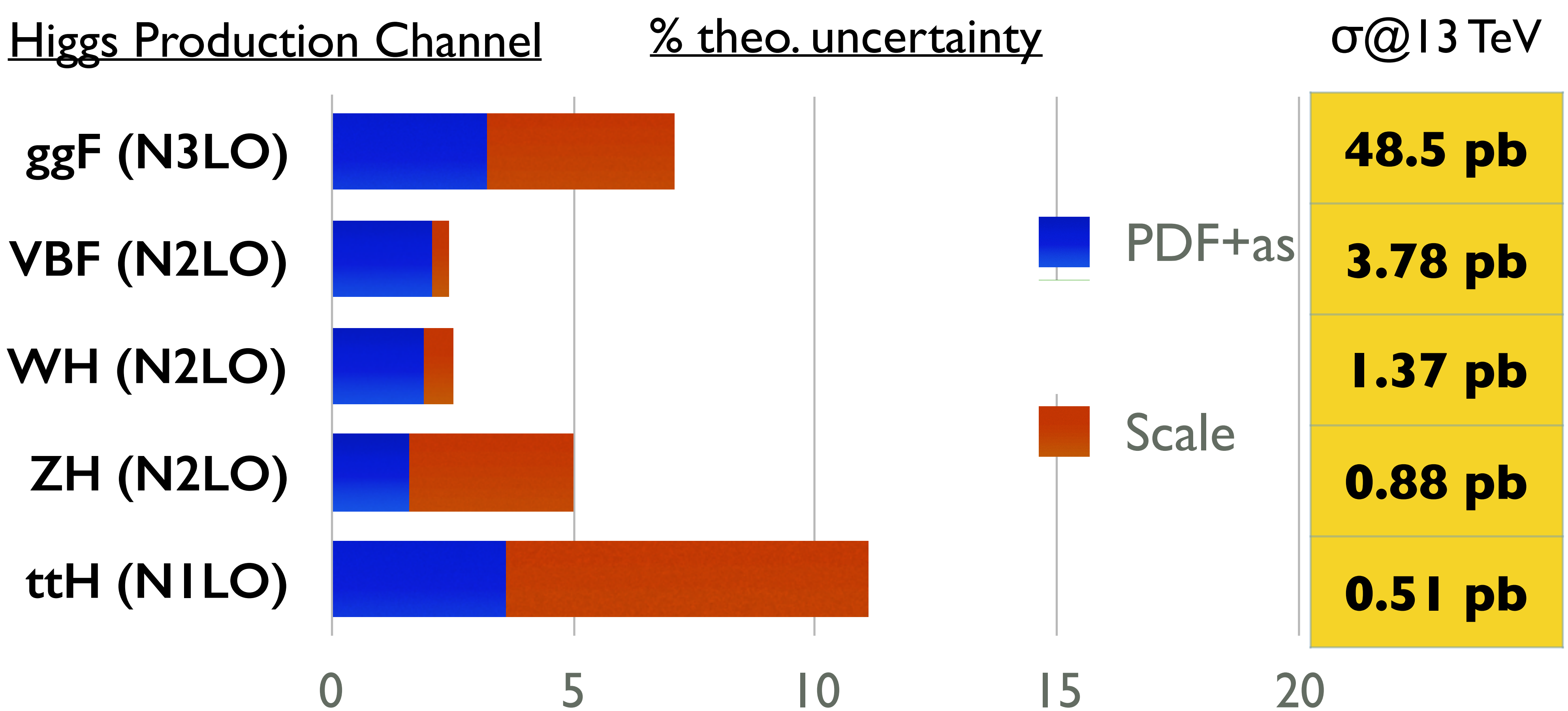
Thank you for your attention!

Extra material

The role of PDF uncertainties

Yellow Report 4 (2016)

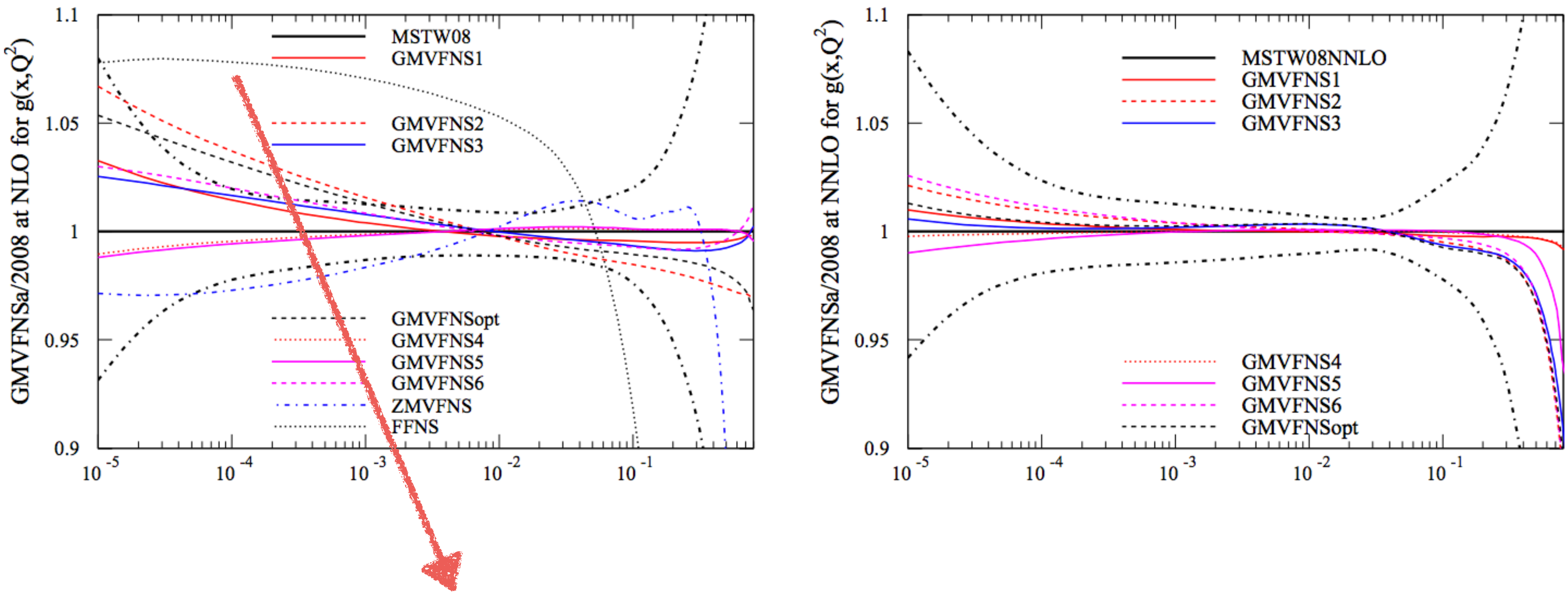




Reduced (still often dominant) PDF uncertainties

Theory settings

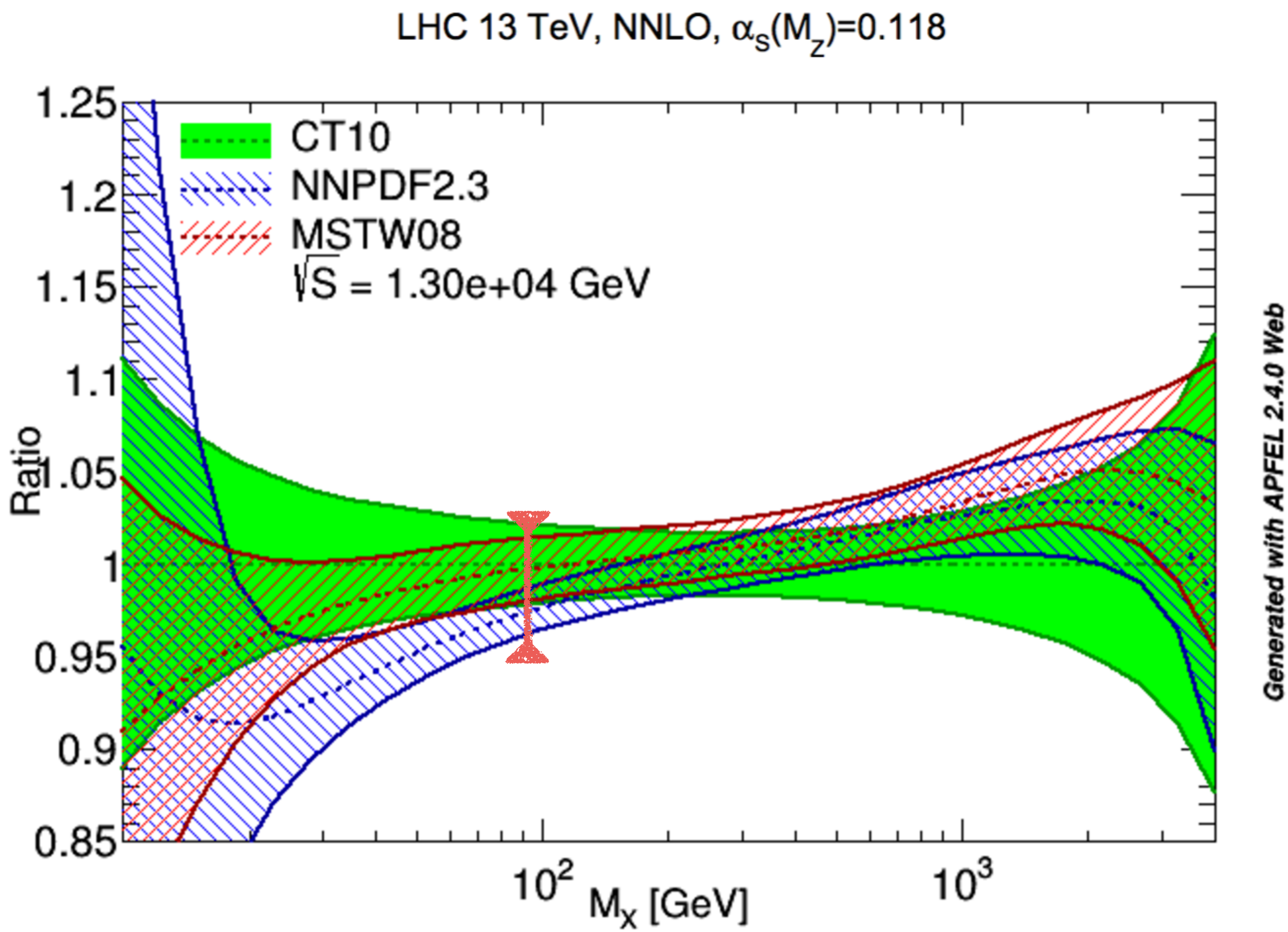
- Comparable GM-VFN schemes for inclusion of HQ masses (sub-leading differences less important at NNLO)
- Common $\alpha_S(M_Z) = 0.118$ (external parameter)
- NNLO (although with some caveat - especially concerning jets data)
- Extensive benchmarking



Compensate by lower $\alpha_S(M_Z)$ in structure function scaling

Quark-Antiquark luminosity

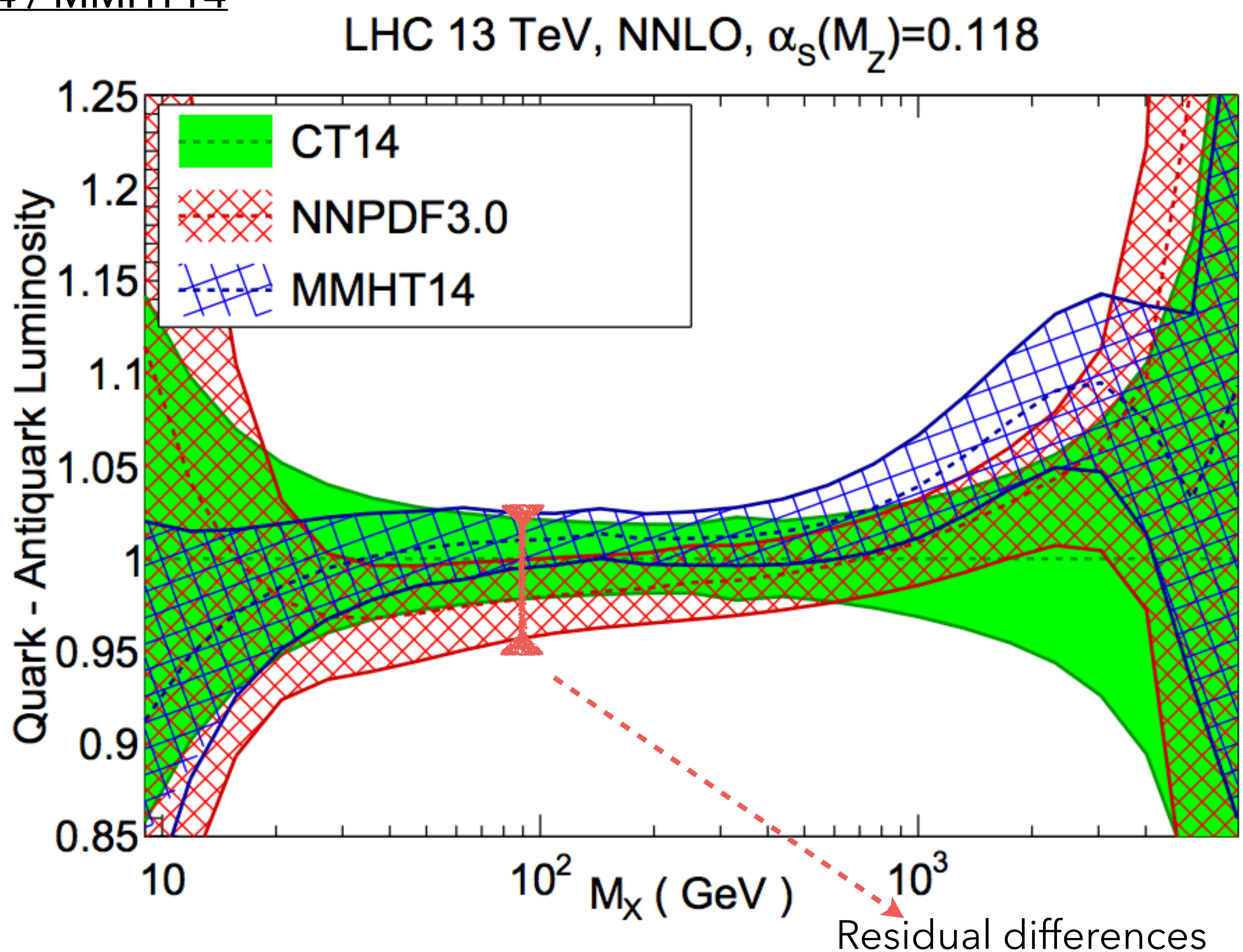
NNPDF2.3 / CT10 / MSTW2008



(2014)

Quark-Antiquark luminosity

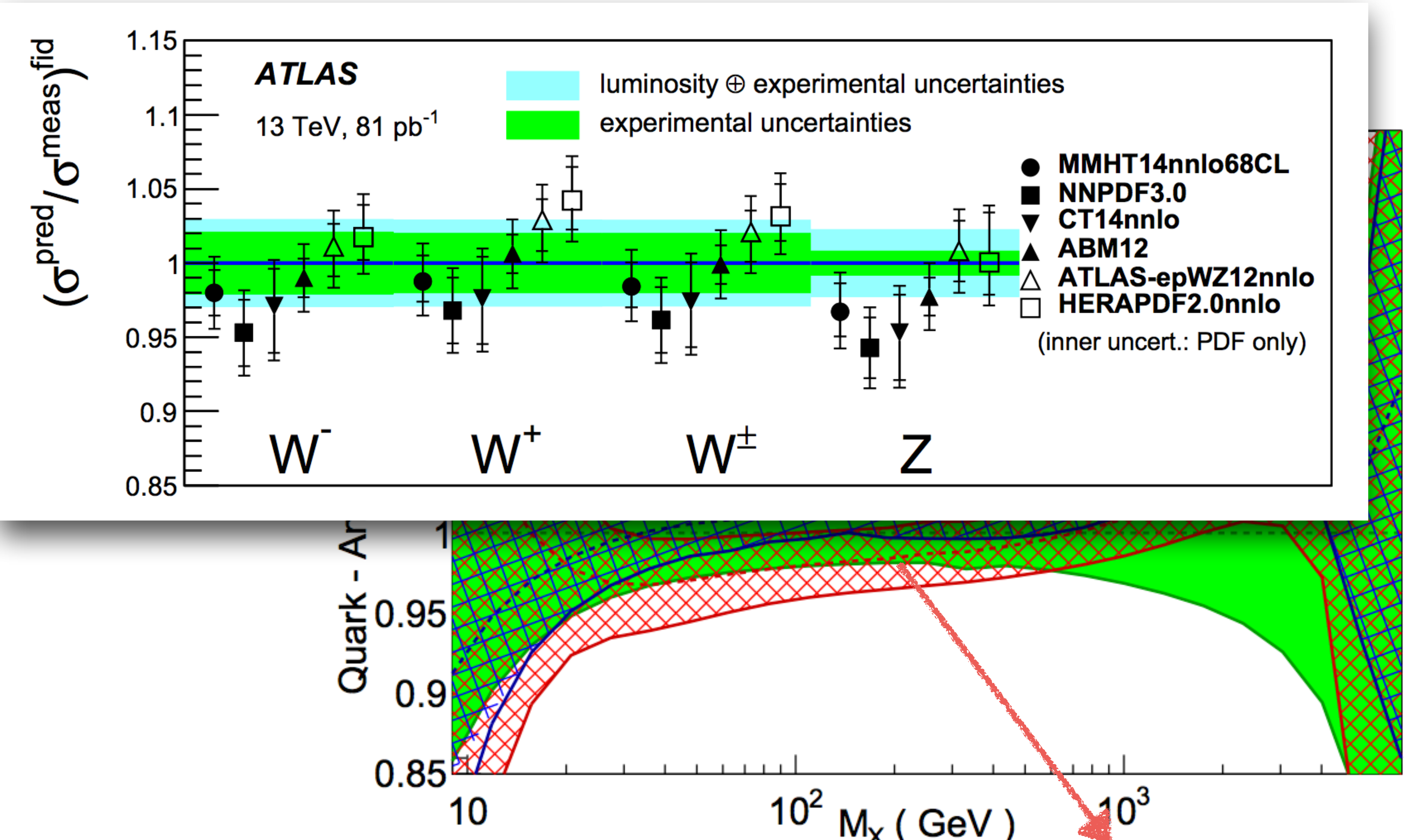
NNPDF3.0 / CT14 / MMHT14



(2016)

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

Quark-Antiquark luminosity



Residual differences

ATLAS, 1603.09222

i - New data from the LHC

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

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

✓ 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. [1611.01460]

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

Currie et al [JHEP 1401 (2014) 110]

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 (NNLO?) codes to PDF fitting code

$$\sigma = \sum_{i,j}^{n_f} \sum_{\alpha,\beta}^{n_x} W_{ij\alpha\beta}\, f_i(x_\alpha, Q_0^2) f_j(x_\beta, Q_0^2)$$

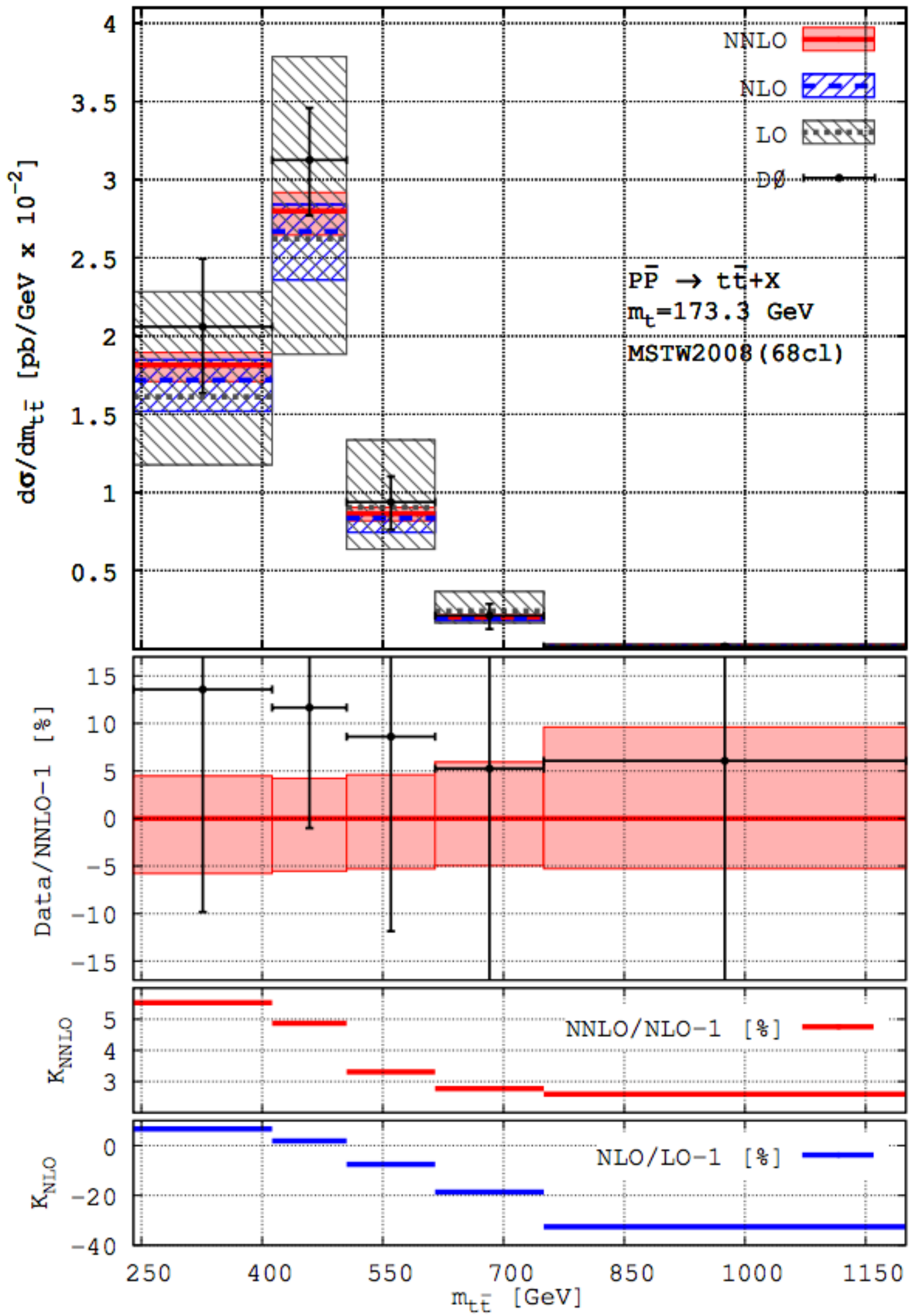
APFEL

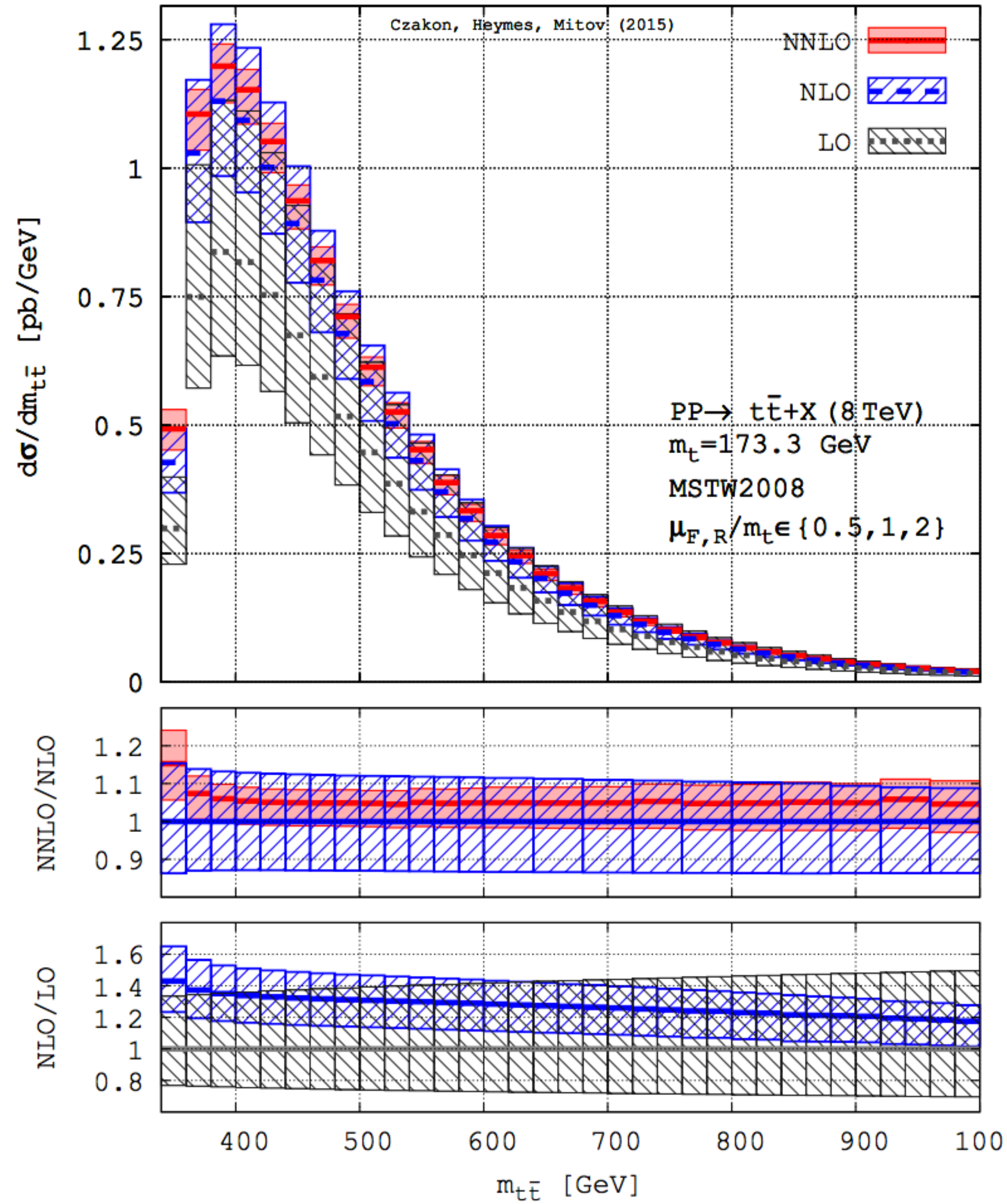
the APPLgrid project

Observable	APPLgrid	**APFELcomb**
W^+ production	1.03 ms	0.41 ms (2.5x)
Inclusive jet production	2.45 ms	20.1 μs (120x)

APPLgrid, Carli et al EPJC66 (2010) 503-524 & FASTNLO, Kluge et al
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

The NNLO frontier - top data

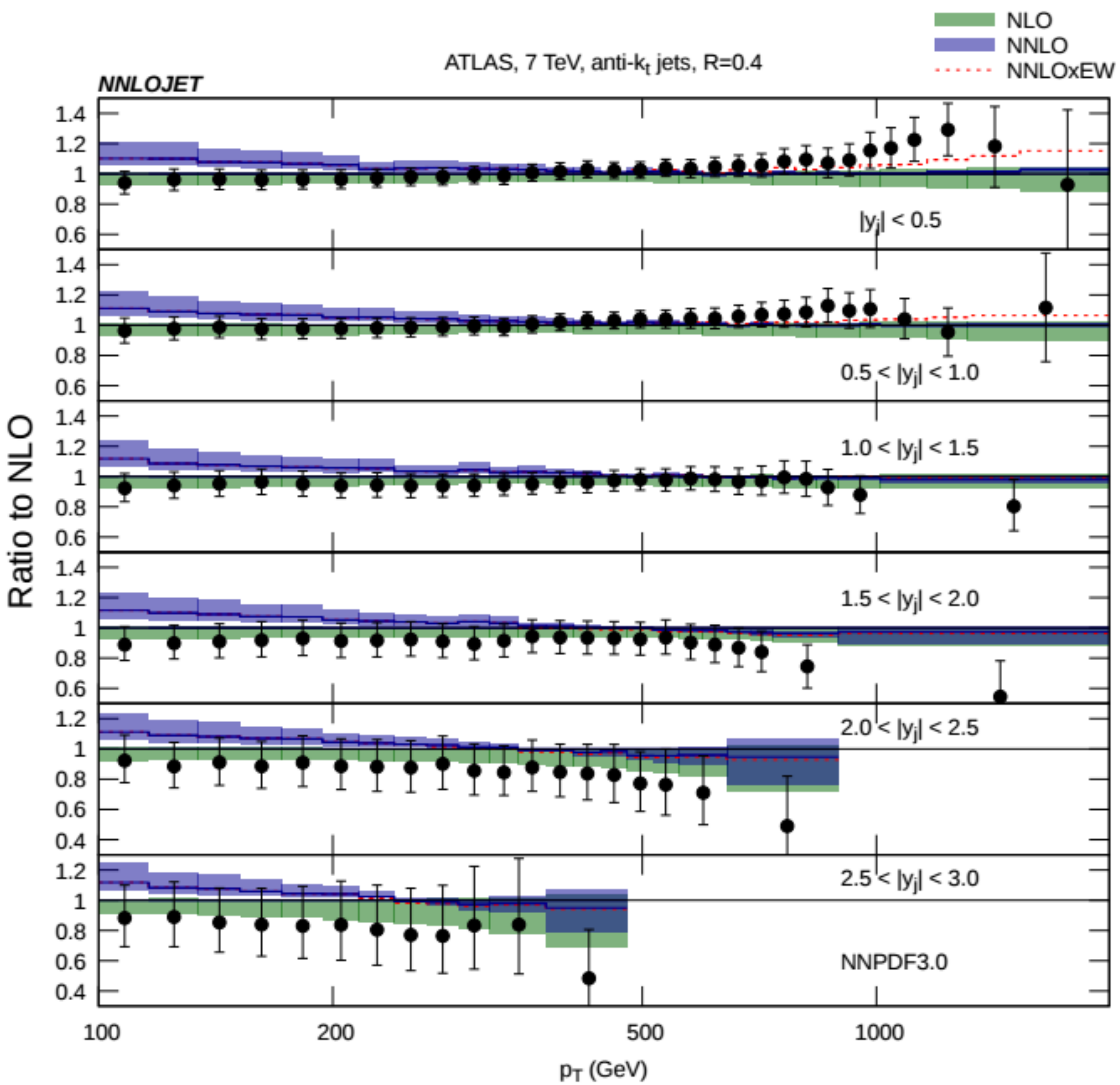




Czakon, Fiedler, Mitov [PRL 116(2016) 082003]

\+ Czakon, Hartland, Mitov, Nocera and Rojo, in preparation

The NNLO frontier - jets data



- NNLO corrections now known for all partonic channels (leading colour contribution only)

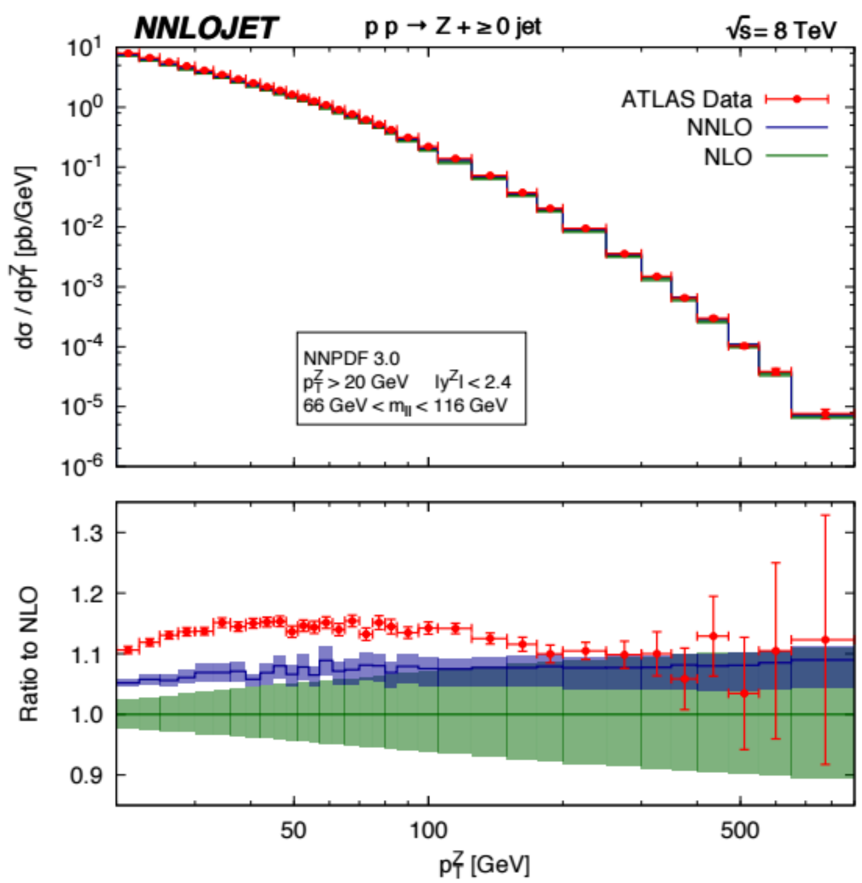
- So fare several PDF groups made different choices: CT14 includes all jet data in NNLO fit assuming overall C-factor small, MMHT14 and ABM12 do not include LHC jet data at NNLO, NNPDF3.0 include some jet data based on goodness of threshold approximation

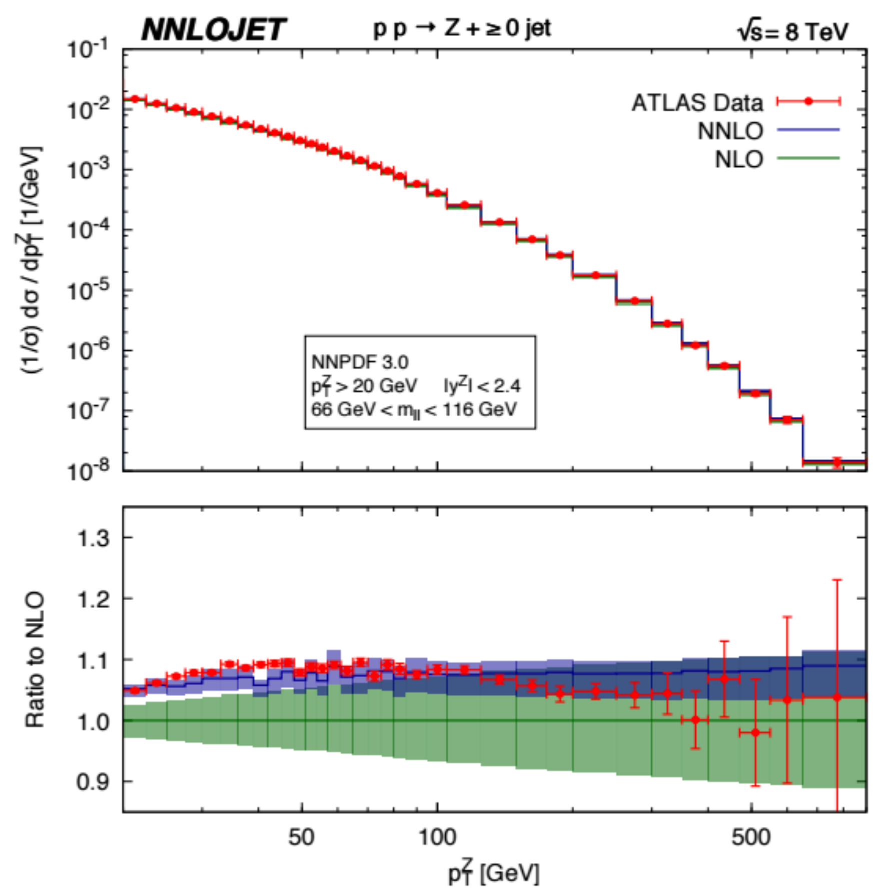
- These choices affect precision of the gluon!

Currie et al [arXiv:1611.01460]

The NNLO frontier - Z pT data

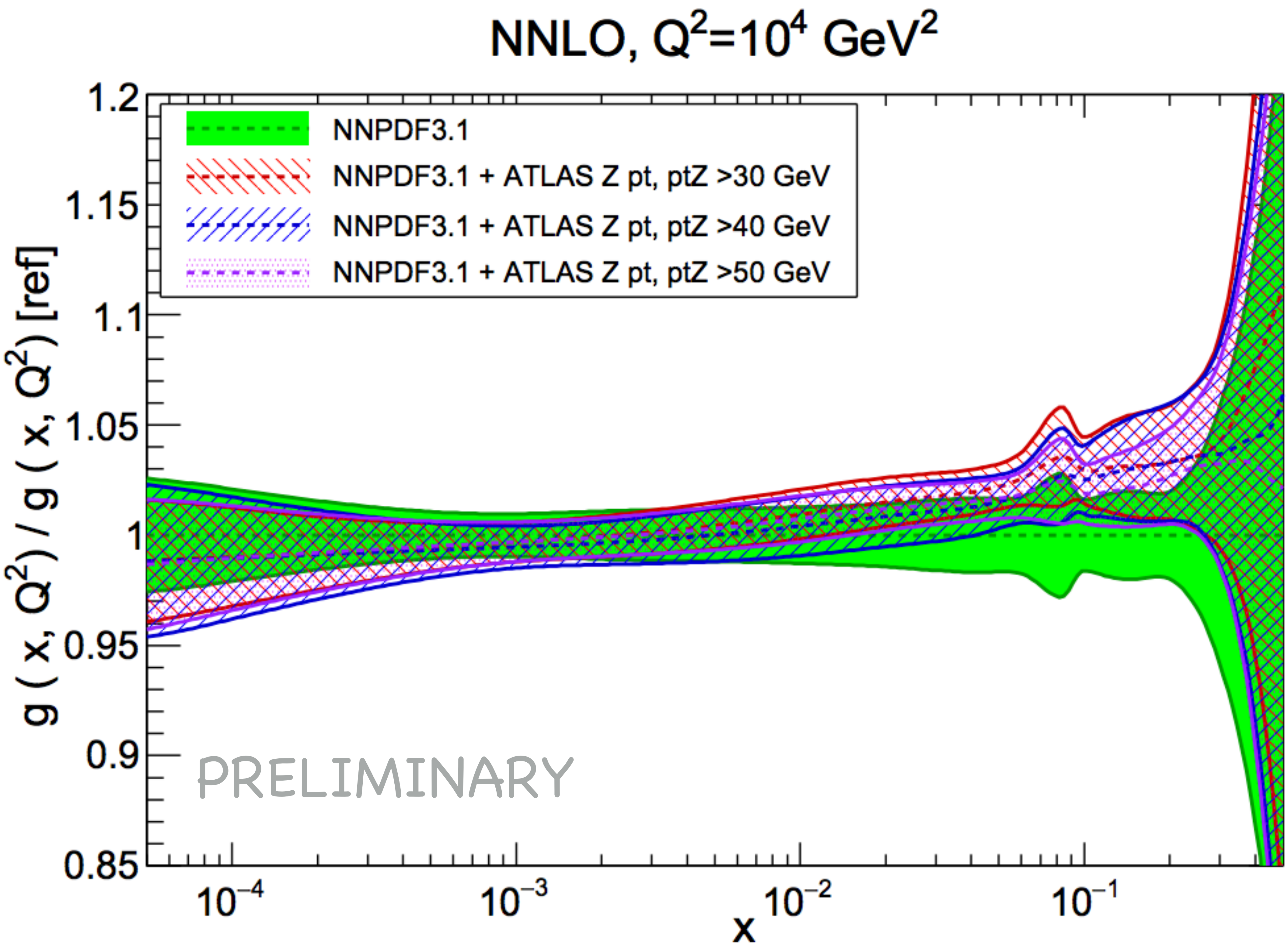
- Experimental precision < 1% up to pT~200 GeV
- Data hugely dominate by correlated systematic uncertainties
- Interesting case-study to probe current theory-experiment frontier





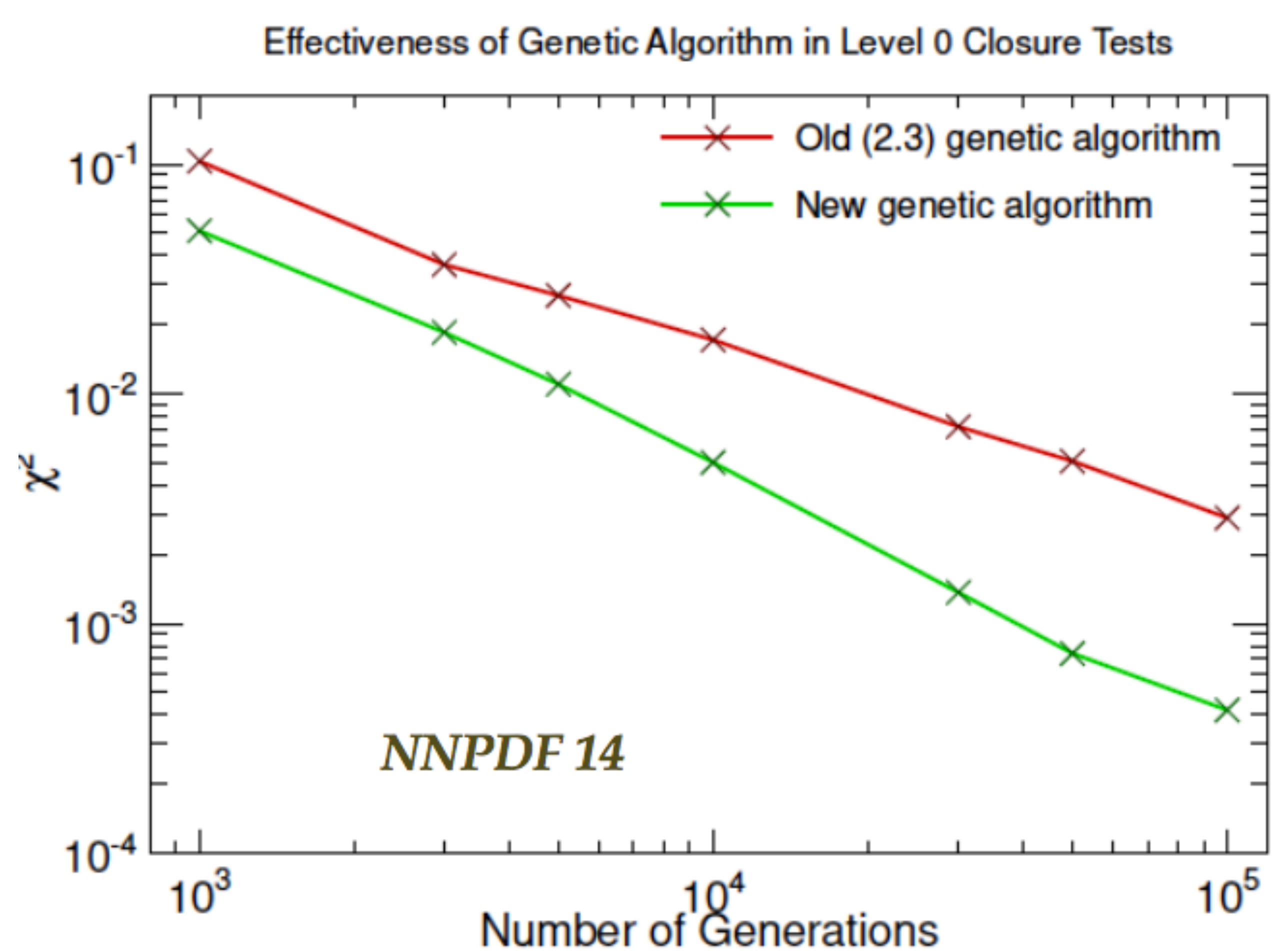
- ☑ ATLAS Z pT @LHC7, normalised distributions, 3 rapidity bins (0<Y<1, 1<Y<2, 2<Y<2.5)
  ~50 data in perturbative region pT > 30 GeV
- ☑ ATLAS Z pT @LHC8, normalised/unnormalised distributions, 6 rapidity bins in Z peak + low/high M
  ~150 data in perturbative region pT > 30 GeV
- ☑ CMS Z pT @LHC8, normalised/unnormalised distributions, 5 rapidity bins in Z peak
  ~50 data in perturbative region pT > 30 GeV

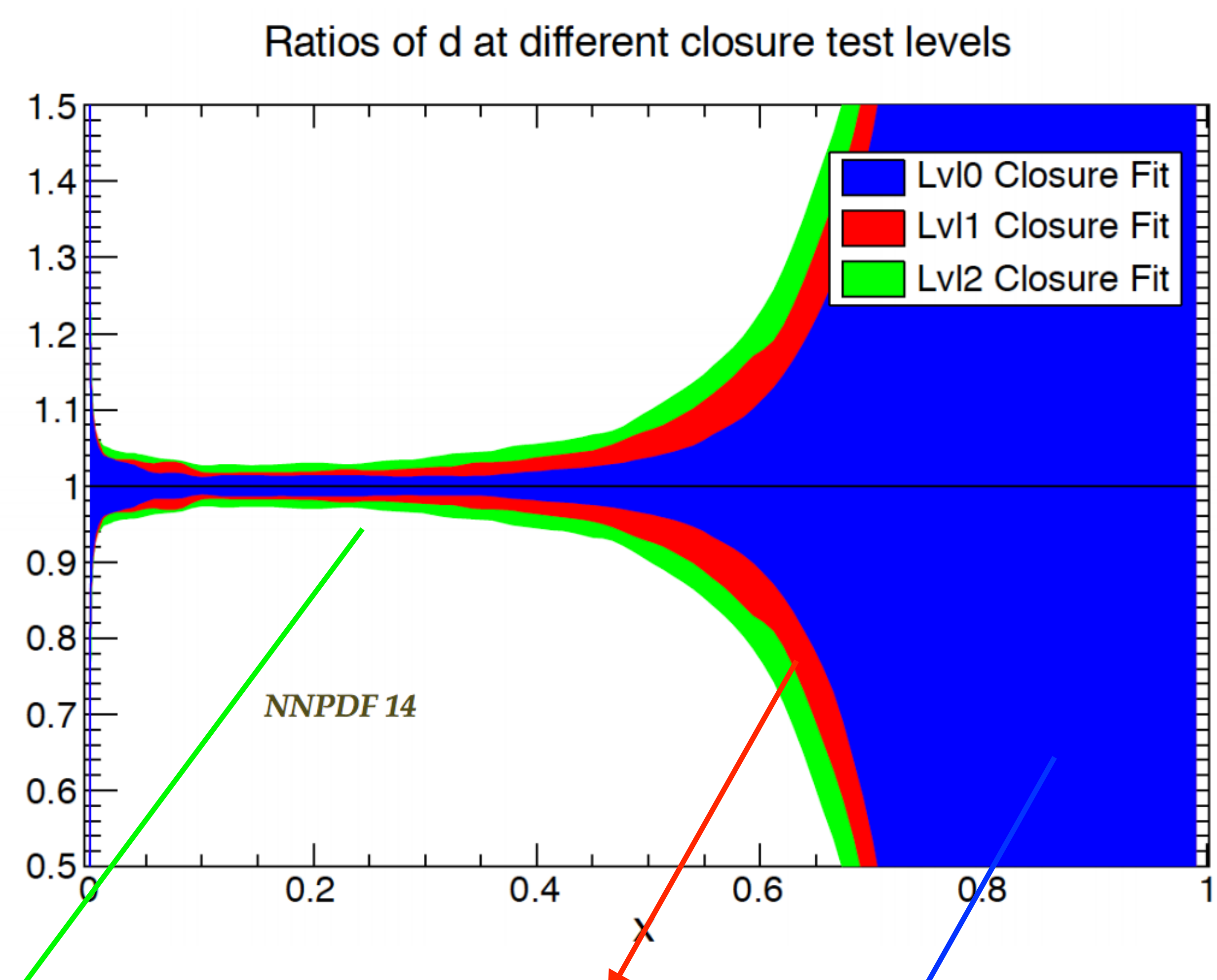
The NNLO frontier - Z pT data



Closure test

- Level-0: if pseudo-data are identical to the input theory, then agreement with theory should be arbitrarily good, i.e. $\chi^2 \to 0$
- Level-1: let pseudo-data fluctuate about their central values within data uncertainty, then $\chi^2 \to 1$
- Level-2: generate Monte Carlo replicas of pseudo-data with fluctuations, then $\chi^2 \to 2$





data uncertainty

parametrisation uncertainty

extrapolation uncertainty