European Research Council

Established by the European Commission

MARIA UBIALI
UNIVERSITY OF CAMBRIDGE

NEW FRONTIERS IN THE DETERMINATION OF THE PROTON STRUCTURE

X SERIES OF MAJORANA LECTURES, PHYSICS DEPT, UNIVERSITY OF NAPLES “FEDERICO II” - 20 APRIL 2021

NEW PHENO CHALLENGES AT THE LHC PRECISION FRONTIER

- **Lecture I:**
  "LHC Run III and the precision frontier"
- **Lecture II:**
  "New frontiers in the determination of the proton structure"
- **Lecture III:**
  "Hints for new physics from precision physics"

AIM: GIVE A PERSONAL PERSPECTIVE ABOUT WHAT I IDENTIFY AS MOST EXCITING CHALLENGES THAT MODERN COLLIDER PHENOMENOLOGY FACES.

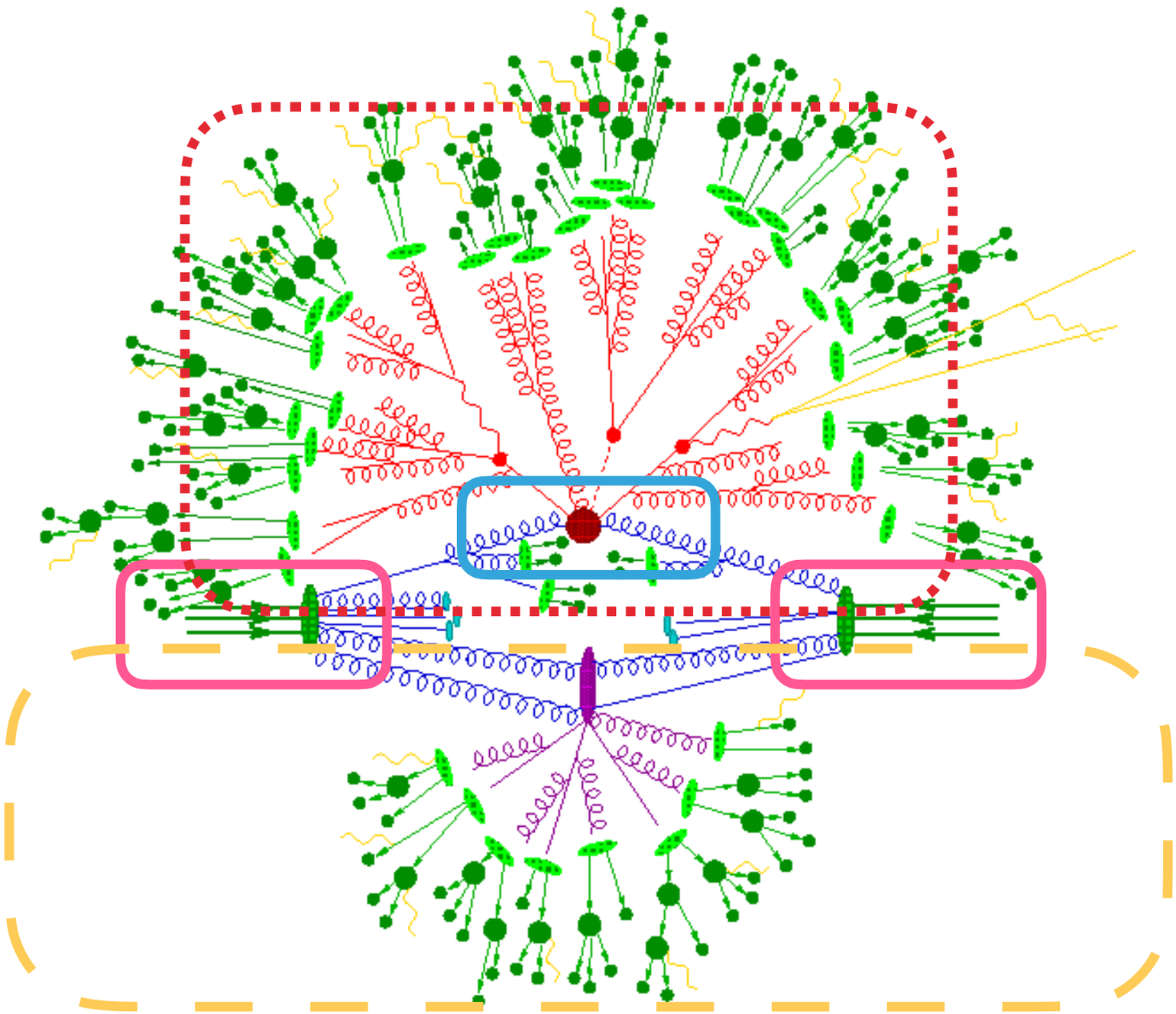
DISCLAIMER: IMPOSSIBLE TO GO INTO ALL DETAILS THAT TOPICS DESERVE NOR TO COVER ALL RELEVANT TOPICS.

OUTLINE

- Motivation
- The determination of the proton structure
  - ➡ Deep Inelastic Scattering and collinear factorisation
  - ➡ How PDFs are extracted from data
  - ➡ State of the art
- New frontiers at LHC Run III
  - ➡ The impact of the LHC data on PDFs
  - ➡ The precision vs accuracy challenge
- Conclusions and outlook

1. MOTIVATION

EVENTS AT HADRON COLLIDERS

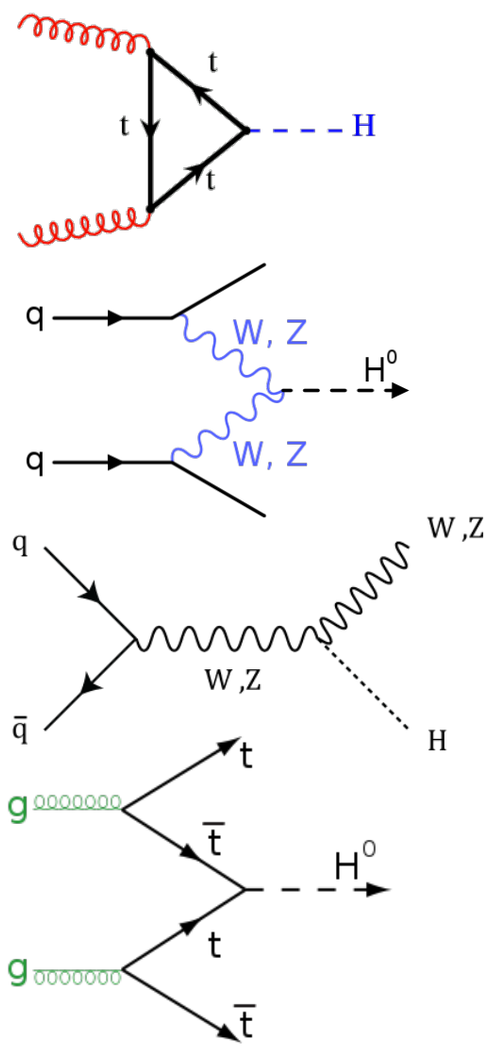


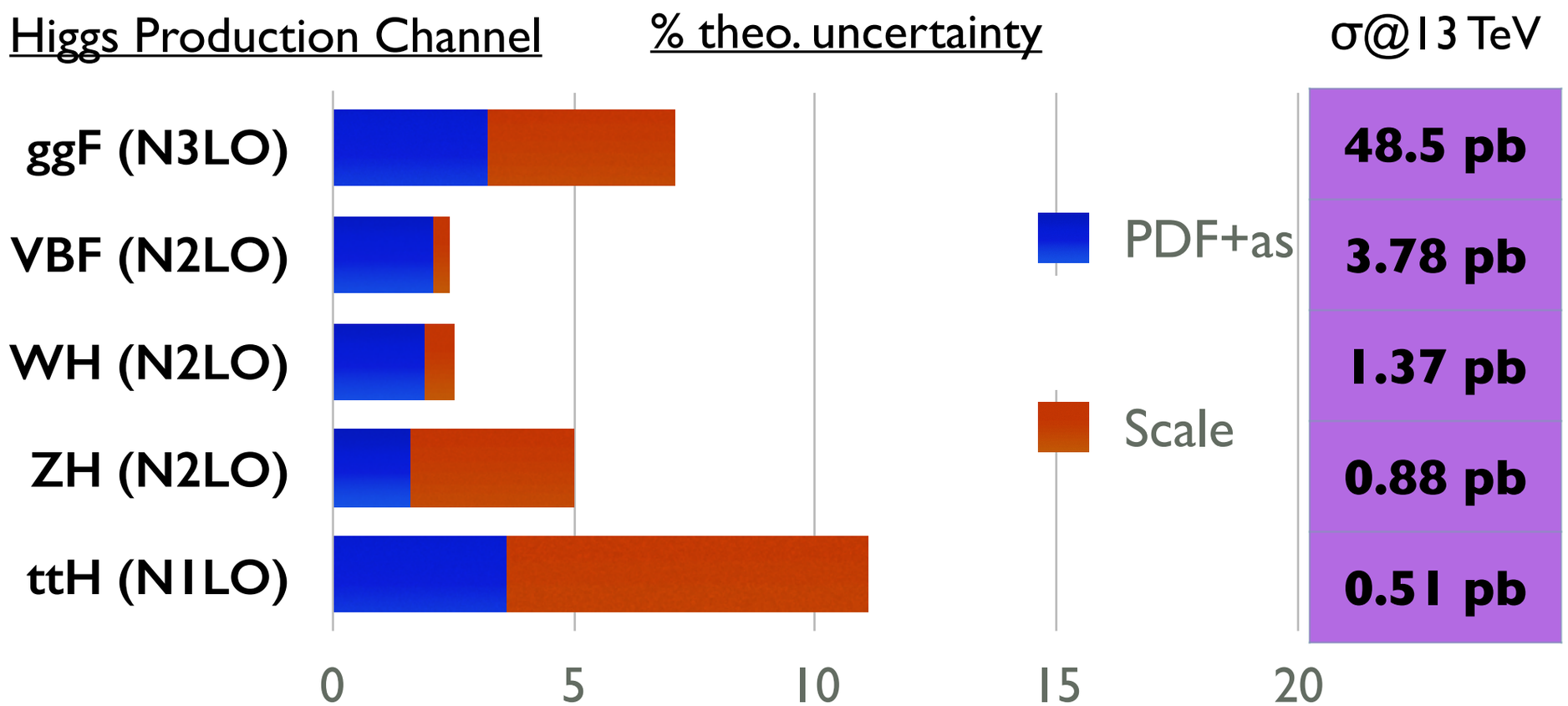
- Hard scattering of partons (Perturbative QCD+EW)
- The structure of the proton: parton distribution functions
- Parton Showering and Hadronization
- Multiple Parton Interaction, Underlying Events

WHY ARE PDFS SO IMPORTANT?

Example1: Higgs physics

Yellow Report 4 (2016)

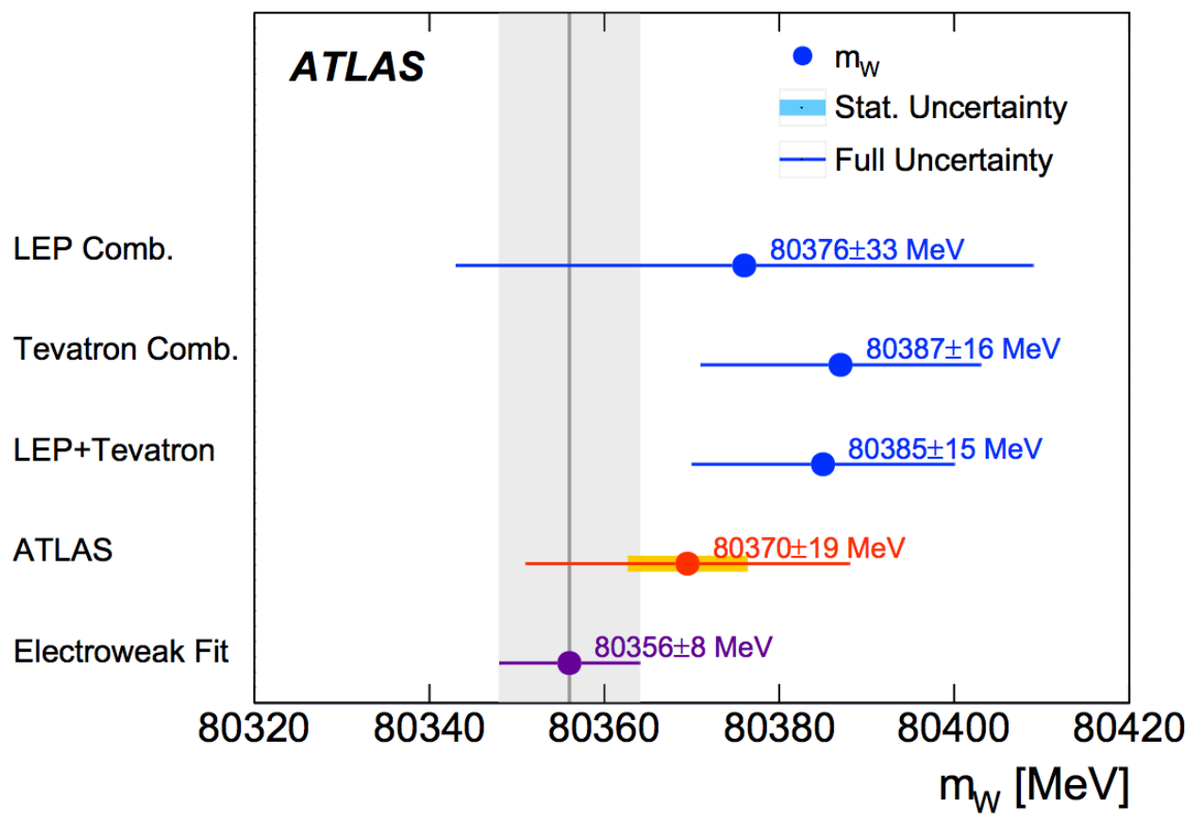




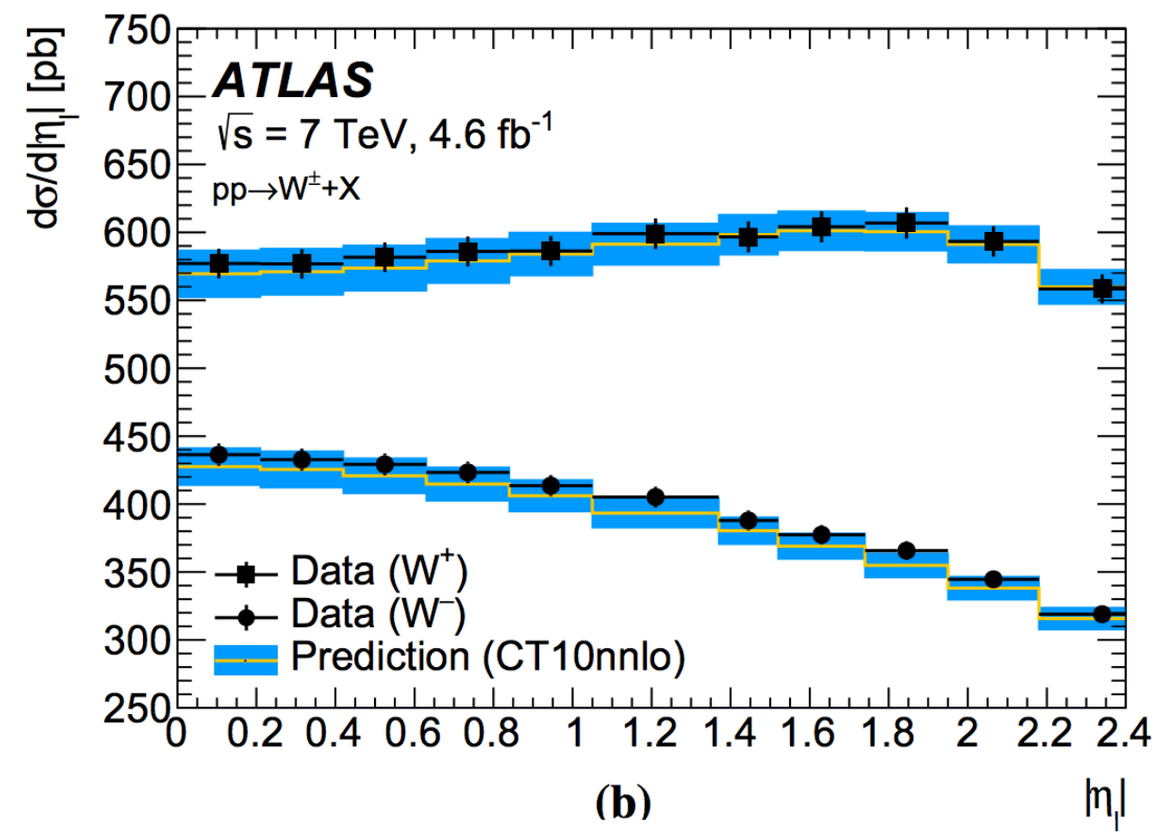
PDF uncertainties limiting factor in the accuracy of theoretical predictions

WHY ARE PDFS SO IMPORTANT?

Example 2: Precise determination of SM parameters



ATLAS collaboration, EPJC 78 (2018) 110



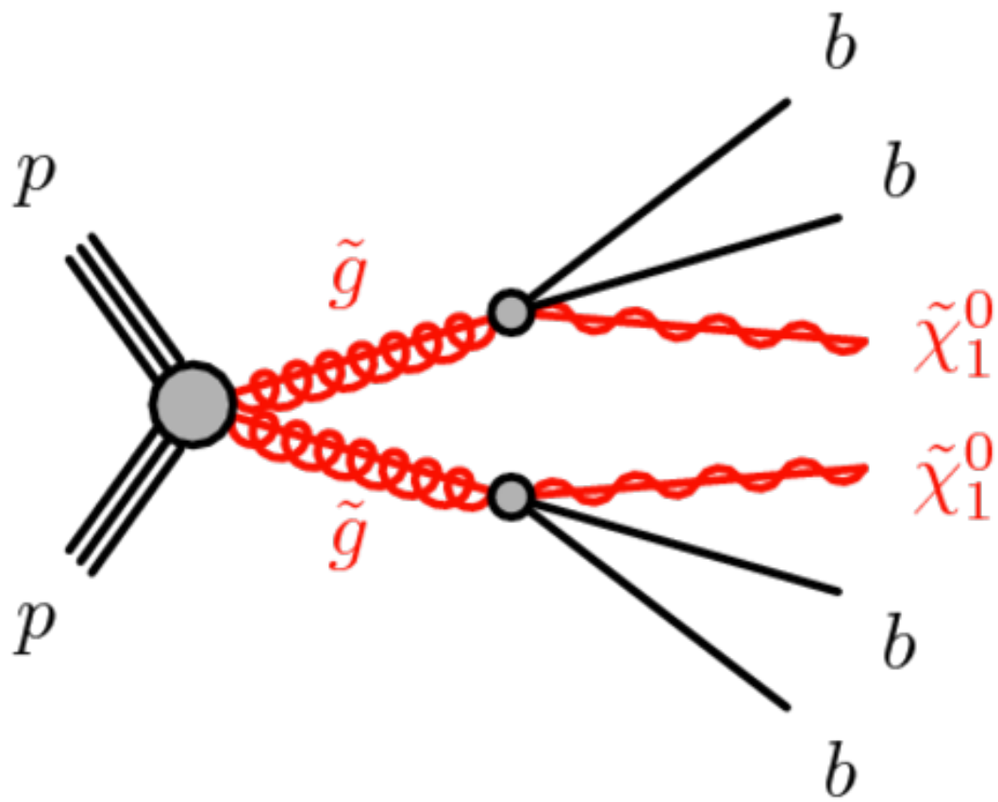
(b)

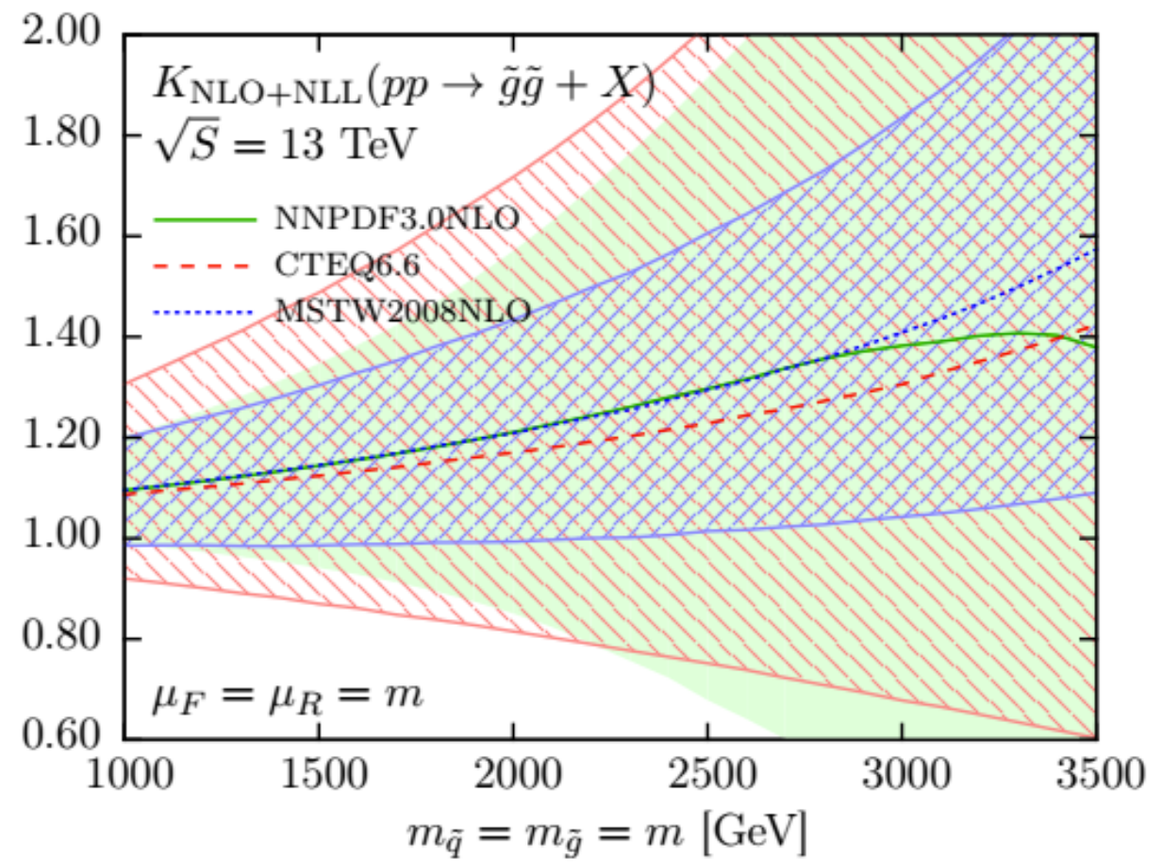
$$\eta = -\ln\tan(\theta/2)$$

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

WHY ARE PDFS SO IMPORTANT?

Example 3: characterisation of new physics

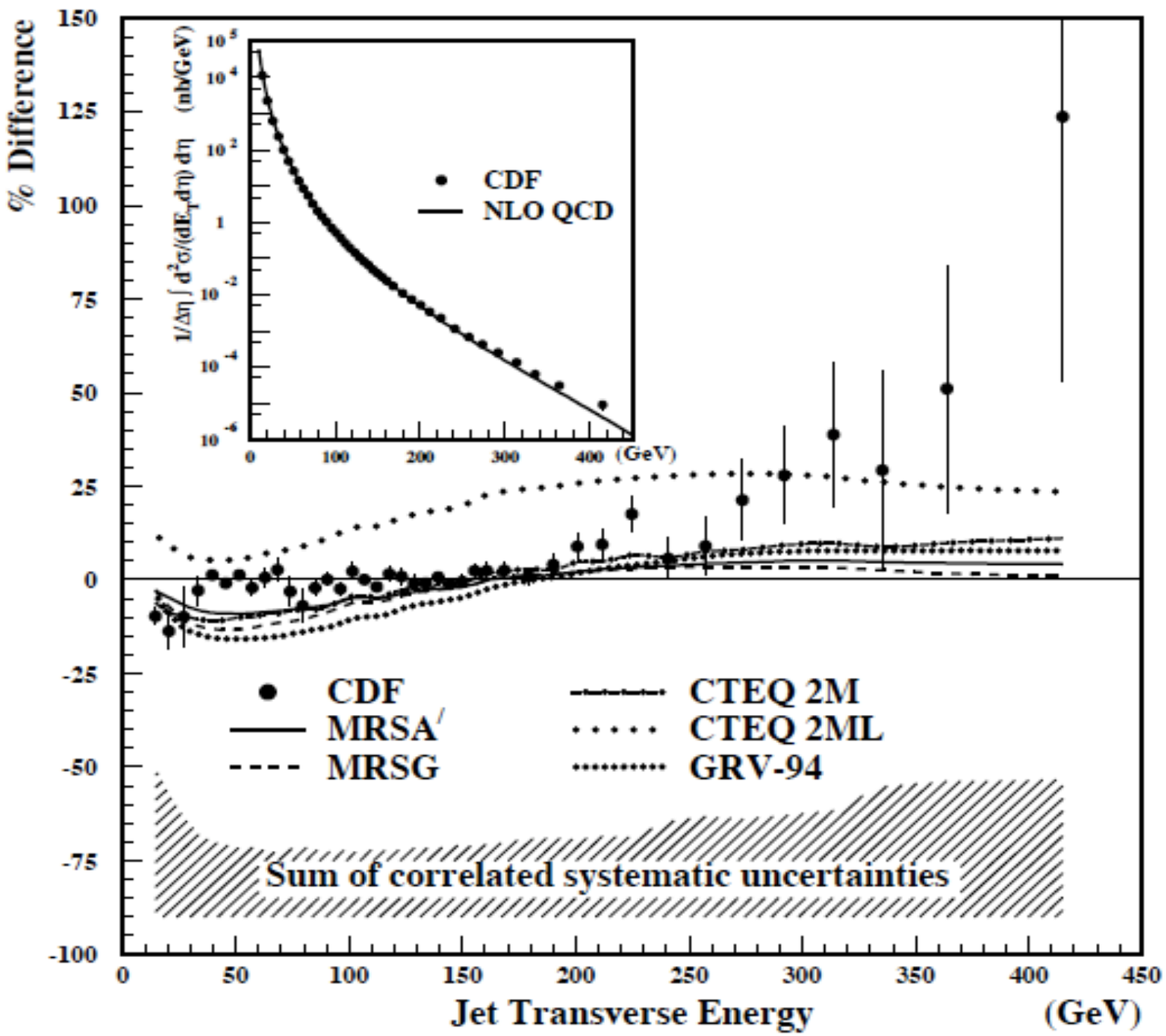




Beenakker et al. EPJC76 (2016)2, 53

PDF uncertainties are a limiting factor in the accuracy of theoretical predictions for production of heavy particles

WHY ARE PDFS SO IMPORTANT?



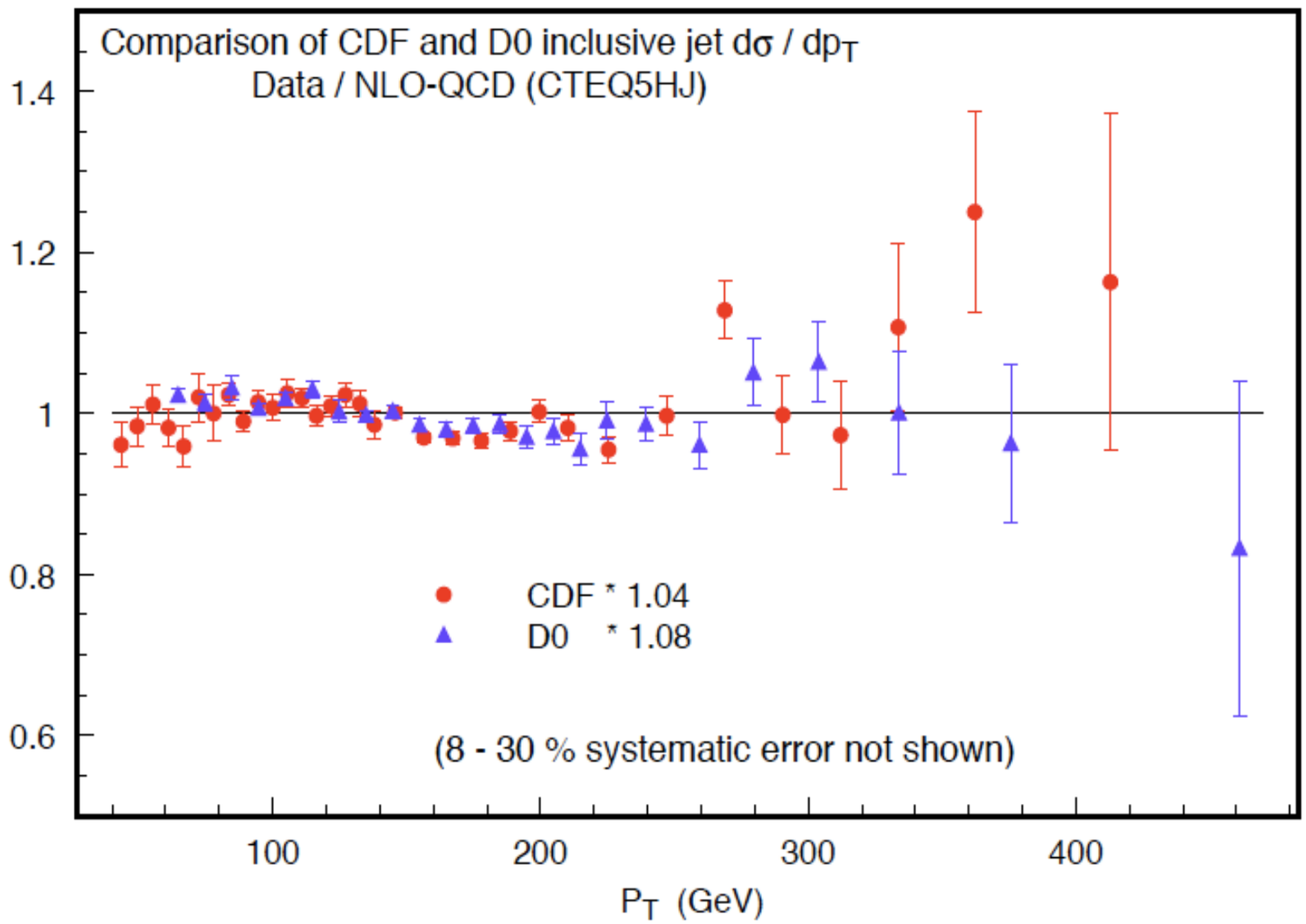
Historia magistra vitae est

Discrepancy between QCD calculations and CDF jet data (1995)

At that time there was no information on PDF uncertainties
and the theoretical prediction strongly depends on gluon shape at x>0.1

WHY ARE PDFS SO IMPORTANT?

FINAL CTEQ FIT (1998)



Historia magistra vitae est

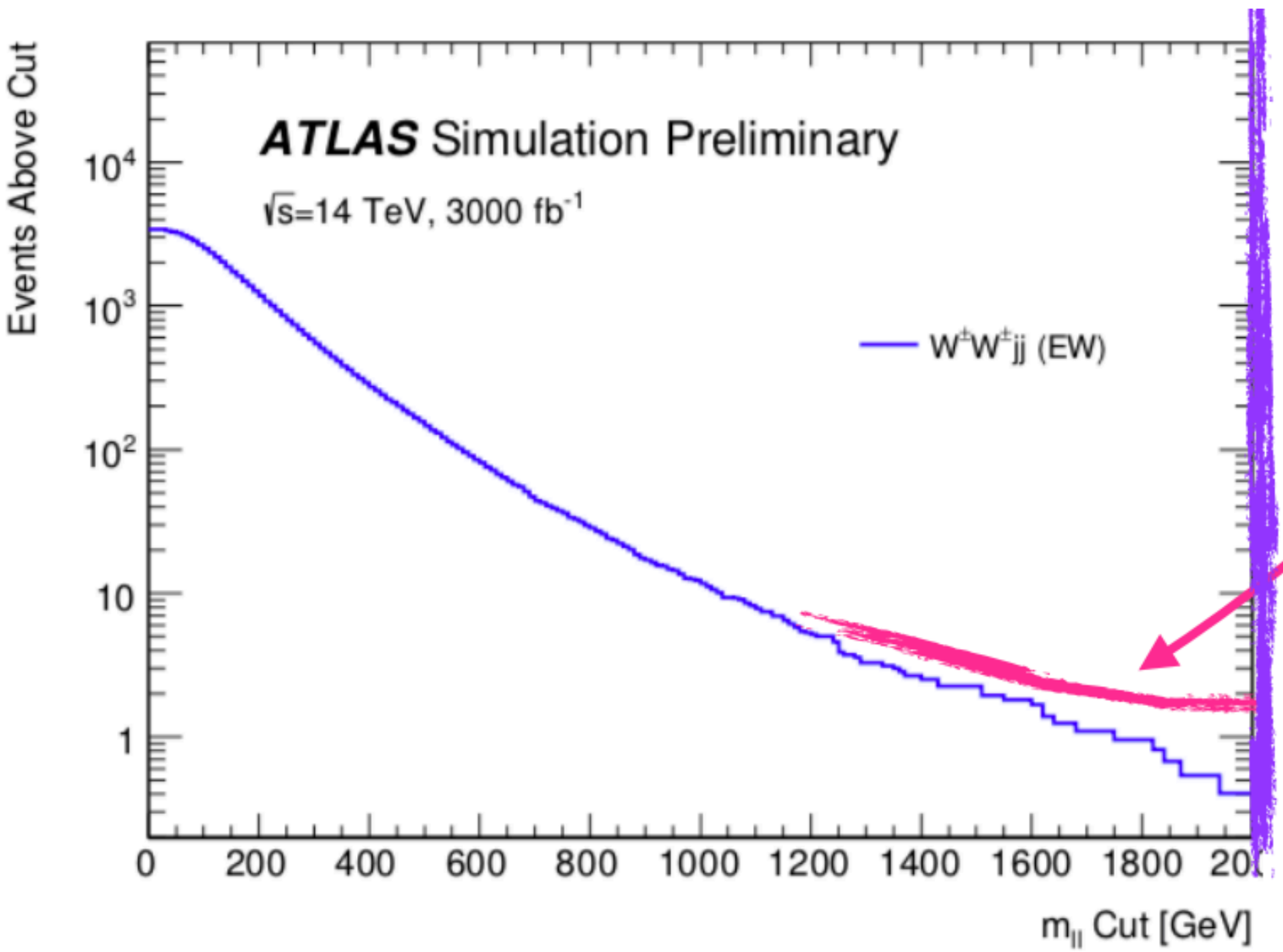
Discrepancy between QCD calculations and CDF jet data (1995)

At that time there was no information on PDF uncertainties
and the theoretical prediction strongly depends on gluon shape at x>0.1

CTEQ re-performed the parton fit by including the jet data and the discrepancy was removed.

WHY ARE PDFS SO IMPORTANT?

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



Current experimental reach

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

II. THE DETERMINATION OF PDFS

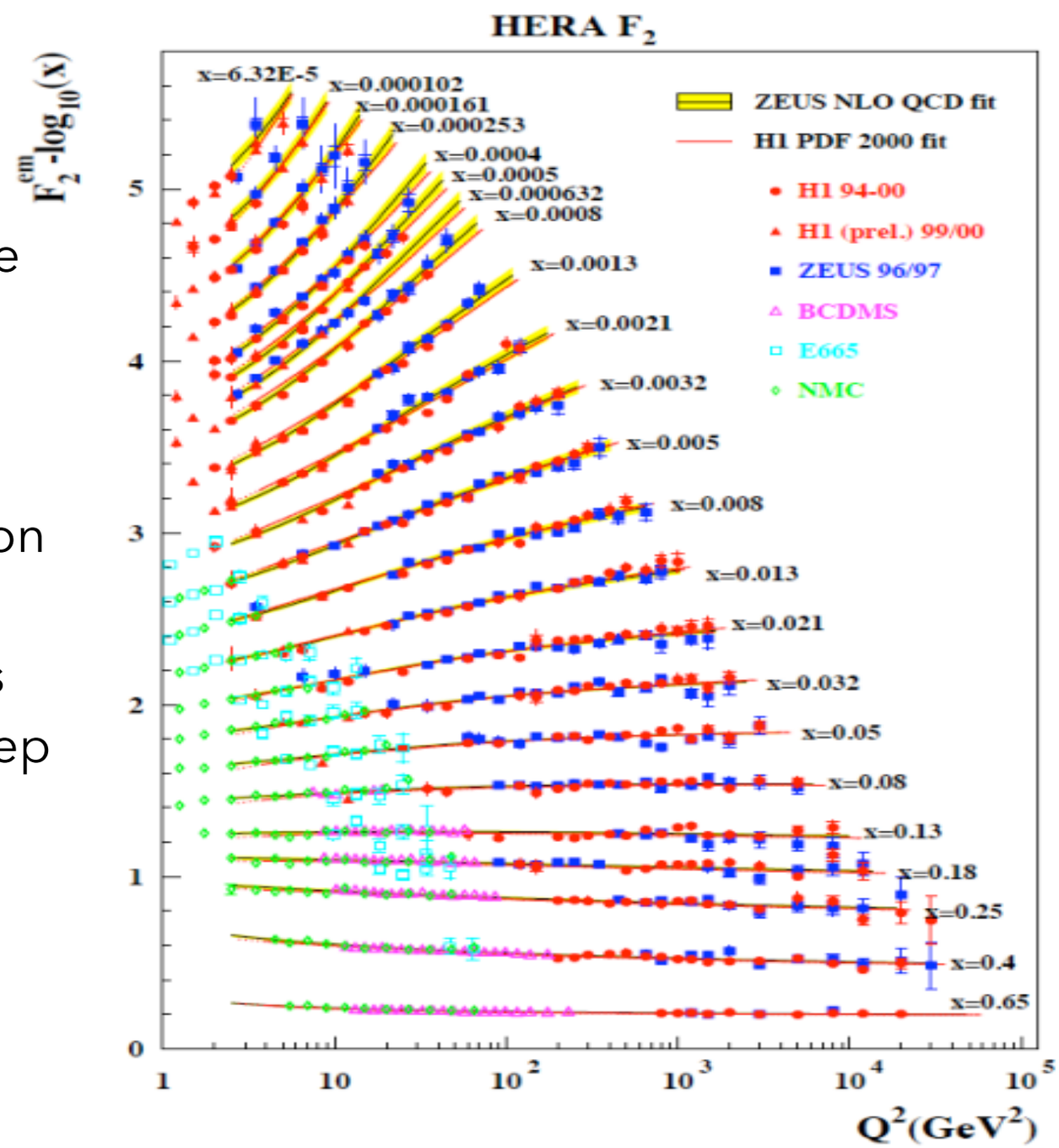
III

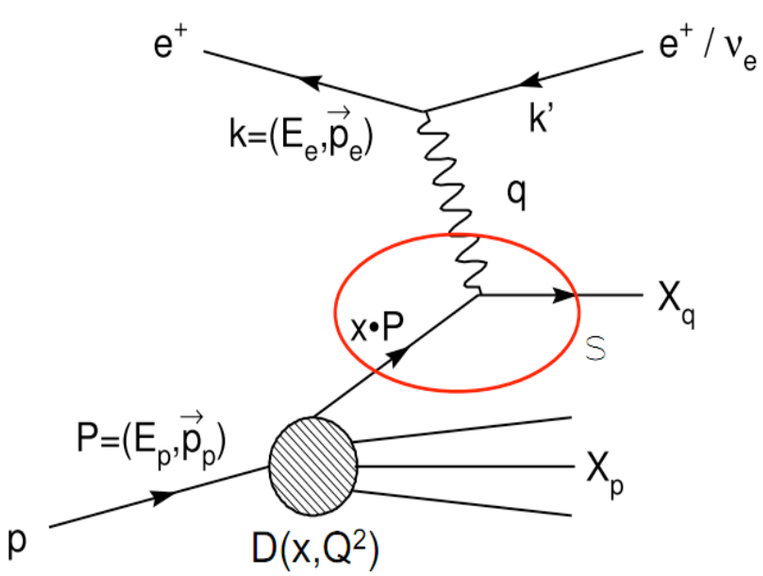
DEEP INELASTIC SCATTERING AND COLLINEAR FACTORIZATION

DEEP INELASTIC SCATTERING

First suprise about the proton structure was given by DIS data, which were interpreted the parton model (Feynman 1969)

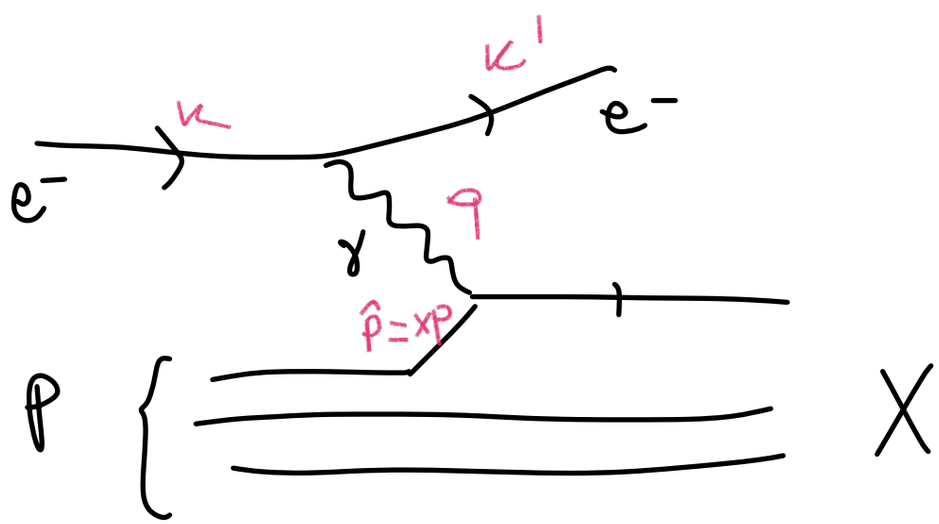
- Photon scatters incoherently off massless, free, point like, spin 1/2 particles (quarks)
- The functions q(x) are the Parton Distribution Functions (PDFs) encode probability that a quark carries a fraction x of parent proton's
- They encode information about proton deep structure





QCD not only explains the approximate scaling F~const at large x but also the violation of the scaling at smaller x and PDFs acquiring dependence on factorisation scale μ_F

THE QCD-IMPROVED PARTON MODEL



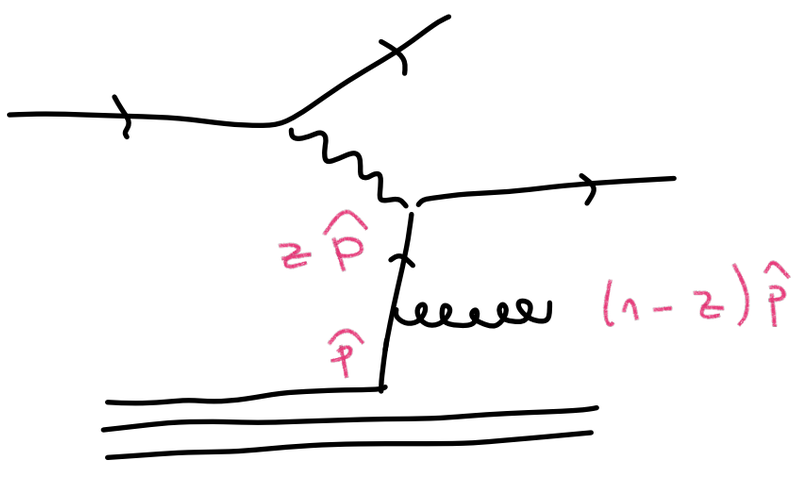
PARTON MODEL

$$\sigma(e^- p \to e^- X) = \sum_i \int_0^1 dx\, f_i(x)\, \hat{\sigma}(\underbrace{xp}_{\hat{p}})$$

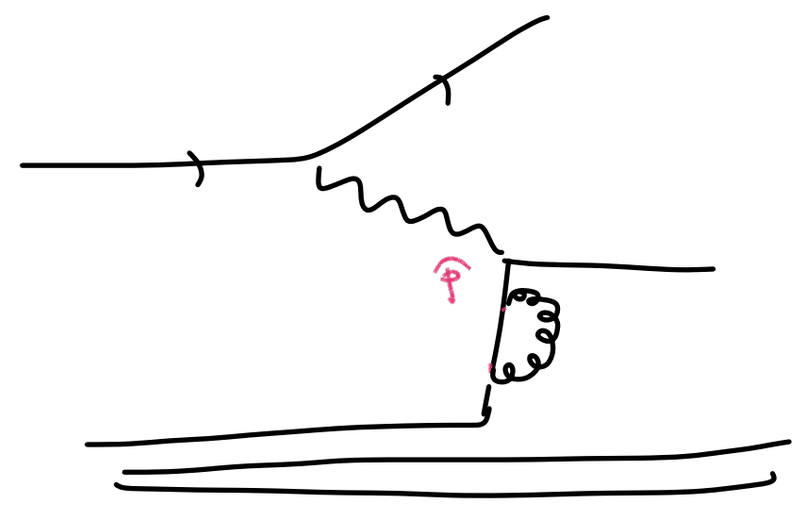
$\hat{\sigma}^{(0)}$

ADD QCD CORRECTIONS

REAL



−



VIRTUAL

THE QCD-IMPROVED PARTON MODEL

$$\hat{\sigma}^{(1)}_{q,R}(\hat{p}) = \frac{\alpha_S}{2\pi} C_F \int_0^1 \frac{dz}{1-z} \int_0^{|k_T^2|_{max}} d|k_T^2| \; \frac{1+z^2}{|k_T^2|} \; \hat{\sigma}_q^{(0)}(z\hat{p})$$

The partonic cross section at NLO in α_S (associated with real emission of a gluon off a quark displays two singularities:

- **SOFT** singularity ($z \to 1$), which we regulate with parameter ε ($\to 0$)
- **COLLINEAR** singularity ($|k^2_T| \to 0$), which we regulate with parameter λ^2 ($\to 0$)

$$\hat{\sigma}^{(1)}_{q,R}(\hat{p}) = \frac{\alpha_S}{2\pi} C_F \int_0^{1-\varepsilon} \frac{dz}{1-z} \int_{\lambda^2}^{|k_T^2|_{max}} d|k_T^2| \; \frac{1+z^2}{|k_T^2|} \; \hat{\sigma}_q^{(0)}(z\hat{p})$$

THE QCD-IMPROVED PARTON MODEL

Add virtual corrections, which also have soft and collinear singularities

$$\hat{\sigma}_v^{(1)} = -\hat{\sigma}_q^{(0)}(\hat{p})\,\frac{\alpha_s}{2\pi}\,C_F \int_0^{1-\varepsilon} \frac{dz}{1-z} \int_{\lambda^2}^{|k_T^2|_{max}} \frac{d|k_T^2|}{|k_T^2|}\,(1+z^2)$$

$$\hat{\sigma}_{q,R+V}^{(1)}(\hat{p}) = \frac{\alpha_s}{2\pi}\,C_F \int_0^1 \frac{dz}{1-z} \int_{\lambda^2}^{|k_T^2|_{max}} \frac{d|k_T^2|}{|k_T^2|}\,(1+z^2)\left[\hat{\sigma}_q^{(0)}(z\hat{p}) - \hat{\sigma}_q^{(0)}(\hat{p})\right]$$

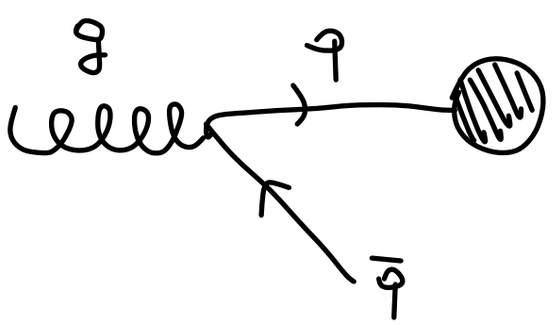
The soft singularity cancels between real and virtual contributions but the collinear singularity still there. Trick: split integration

$$\int_{\lambda^2}^{|k_T^2|_{max}} \frac{d|k_T^2|}{|k_T^2|} \rightarrow \underbrace{\int_{\lambda^2}^{\mu_F^2} \frac{d|k_T^2|}{|k_T^2|}}_{\text{SINGULAR}} + \underbrace{\int_{\mu_F^2}^{|k_T^2|_{max}} \frac{d|k_T^2|}{|k_T^2|}}_{\text{FINITE}}$$

THE QCD-IMPROVED PARTON MODEL

$$\hat{\sigma}_q(\hat{p}) = \hat{\sigma}_q^{(0)}(\hat{p}) + \hat{\sigma}_q^{(1)}(\hat{p})$$

$$= \hat{\sigma}_q^{(0)}(\hat{p}) + \frac{\alpha_s(Q^2)}{2\pi}\int_0^1 dz\, P_{qq}(z)\,\hat{\sigma}_q^{(0)}(z\hat{p})\,\boxed{\log\frac{M_F^2}{\lambda^2}} + \text{NON-SINGULAR TERMS}$$



Add LO and NLO correction from initial quark + NLO contribution from gluon-initiated process with gluon splitting into quark-antiquark pair

$$\hat{\sigma}_g(\hat{p}) = \frac{\alpha_s(Q^2)}{2\pi}\int_0^1 dz\, P_{qg}(z)\,\hat{\sigma}_q^{(0)}(z\hat{p})\,\boxed{\log\frac{M_F^2}{\lambda^2}} + \text{NON-SINGULAR TERMS}$$

$P_{qq}(z)$ and $P_{qg}(z)$ are universal functions associated to quark and gluon splittings, called splitting functions

THE QCD-IMPROVED PARTON MODEL

Plug into the parton model equation:

$$\sigma(p) = \int_0^1 dx\, f_q(x)\, \hat{\sigma}_q(\underbrace{xp}_{\hat{p}}) + f_g(x)\, \hat{\sigma}_g(xp)$$

$$\Downarrow$$

$$\sigma(p) = \int_0^1 dx \Bigg[f_q(x) \Bigg(\hat{\sigma}_q^{(0)}(xp) + \frac{\alpha_s}{2\pi} \int_0^1 dz\, P_{qq}(z)\, \hat{\sigma}_q^{(0)}(xzp) \log\frac{\mu_F^2}{\lambda^2} \Bigg)$$

$$+ f_g(x) \Bigg(\frac{\alpha_s}{2\pi} \int_0^1 dz\, P_{qg}(z)\, \hat{\sigma}_q^{(0)}(xzp) \log\frac{\mu_F^2}{\lambda^2} \Bigg) \Bigg]$$

$$+ \int_0^1 dx \Big[f_q(x)\, \hat{\sigma}^{(1)}_{q,\,reg}(xp, \mu_F^2) + f_g(x)\, \hat{\sigma}^{(1)}_{g,\,reg}(xp, \mu_F^2) \Big]$$

THE QCD-IMPROVED PARTON MODEL

Plug into the parton model equation:

$$\sigma(p) = \int_0^1 dx\, f_q(x)\, \hat{\sigma}_q(\underbrace{xp}_{\hat{p}}) + f_g(x)\, \hat{\sigma}_g(xp)$$

$$\Downarrow$$

$$\sigma(p) = \int_0^1 dx \left[f_q(x) \left(\hat{\sigma}_q^{(0)}(xp) + \frac{\alpha_s}{2\pi} \int_0^1 dz\, P_{qq}(z)\, \hat{\sigma}_q^{(0)}(xzp) \log \frac{\mu_F^2}{\lambda^2} \right) + f_g(x) \left(\frac{\alpha_s}{2\pi} \int_0^1 dz\, P_{qg}(z)\, \hat{\sigma}_q^{(0)}(xzp) \log \frac{\mu_F^2}{\lambda^2} \right) \right]$$

$$+ \int_0^1 dx \left[f_q(x)\, \hat{\sigma}^{(1)}_{q,\,reg}(xp, \mu_F^2) + f_g(x)\, \hat{\sigma}^{(1)}_{g,\,reg}(xp, \mu_F^2) \right]$$

Absorb collinear divergences into a redefinition of the parton distribution functions, which now depend on the factorisation scale

$$f_q(x, \mu_F^2) = \int_x^1 \frac{dy}{y} \left\{ f_q(y) \left[\delta\left(1 - \frac{x}{y}\right) + \frac{\alpha_s}{2\pi} P_{qq}\left(\frac{x}{y}\right) \log \frac{\mu_F^2}{\lambda^2} \right] + f_g(y) \left[\frac{\alpha_s}{2\pi} P_{qg}\left(\frac{x}{y}\right) \log \left(\frac{\mu_F^2}{\lambda^2}\right) \right] \right\}$$

THE QCD-IMPROVED PARTON MODEL

With this redefinition of the PDFs, both PDFs and partonic cross section are finite and they both acquired dependence on arbitrary factorisation scale

$$\Rightarrow \quad \sigma(p) = \int_0^1 dx\, f_q(x, \mu_F^2)\, \hat{\sigma}_{q,\mathrm{reg}}(xp, \mu_F^2) + f_g(x, \mu_F^2)\, \hat{\sigma}_{g,\mathrm{reg}}(xp, \mu_F^2)$$

From PDF redefinition (similar to renormalisation) note that the dependence of PDFs on the scale is totally fixed by perturbation theory.

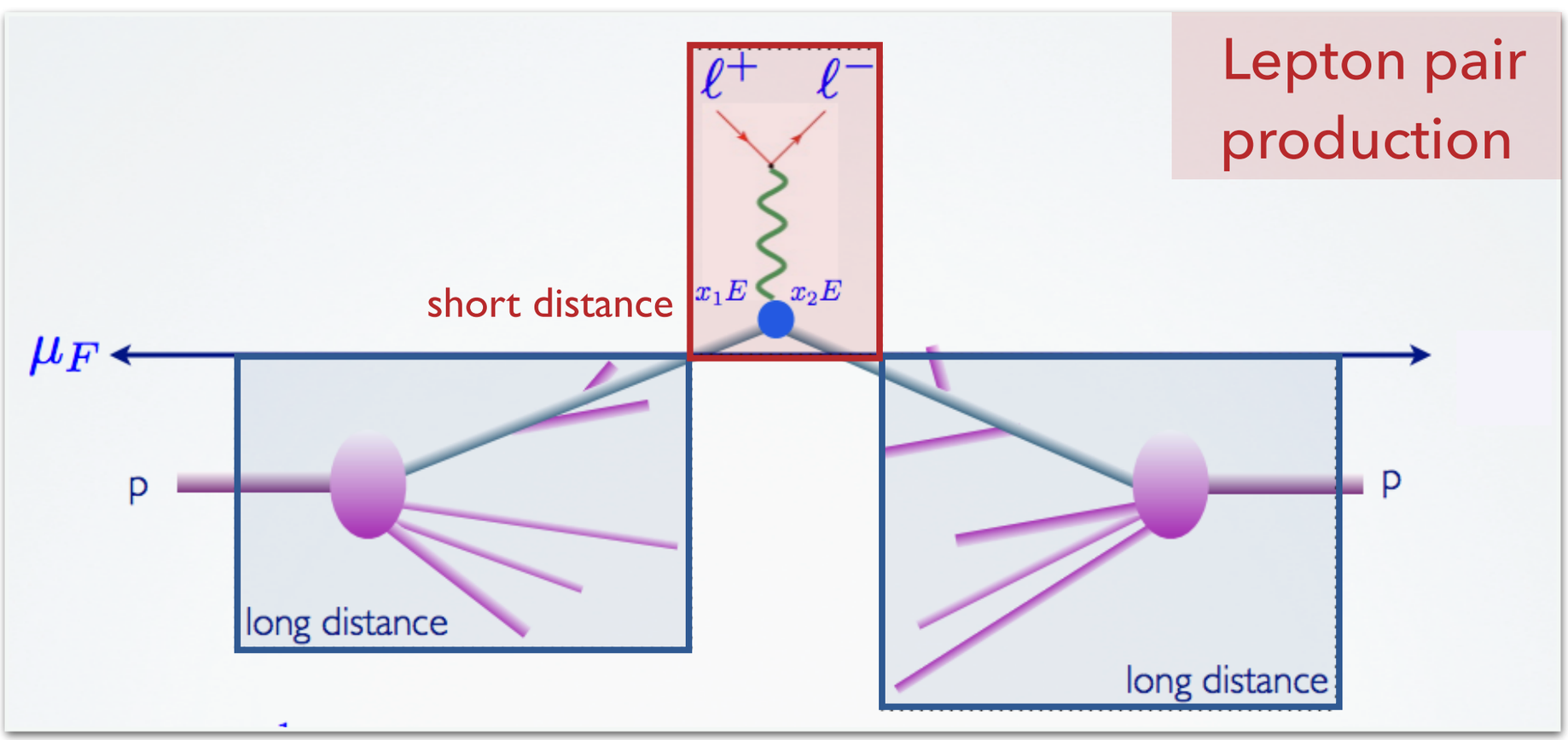
DGLAP evolution equation, similar to renormalisation group equations for $\boldsymbol{\alpha}_S$

$$\mu^2 \frac{\partial f_q(x, \mu^2)}{\partial \mu^2} = \frac{\alpha_s}{2\pi} \int_x^1 \frac{dy}{y}\, P_{qq}\left(\frac{x}{y}\right) f_q(y, \mu^2) + P_{qg}\left(\frac{x}{y}\right) f_g(y, \mu^2)$$

COLLINEAR FACTORISATION

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

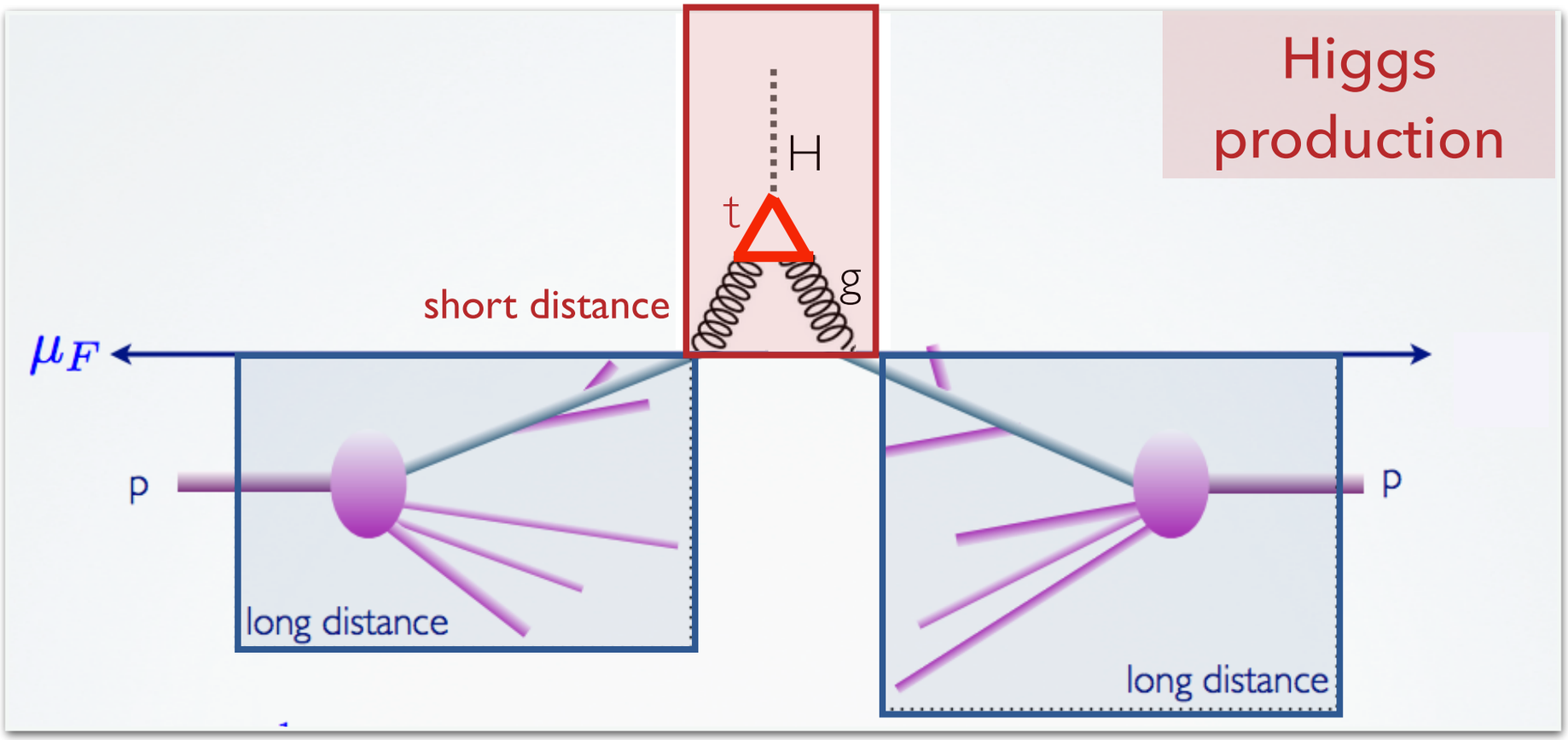
$$\sigma^{pp\to ab} = \sum_{i,j=-n_f}^{n_f} \int dz_1\, dz_2\, f_i(z_1,\mu_F)\, f_j(z_2,\mu_F)\, \hat{\sigma}^{ij\to ab}(z_1 z_2 S, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda^n}{S^n}\right)$$



COLLINEAR FACTORISATION

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

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



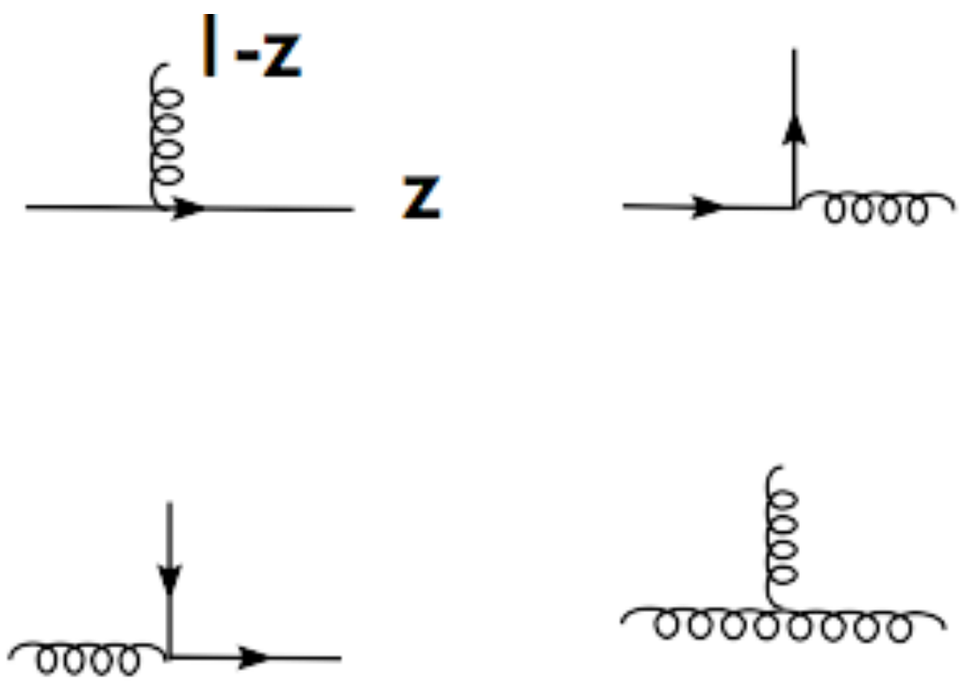
DGLAP EVOLUTION EQUATIONS

PDF dependence on the scale μ_F determined in pQCD by solving DGLAP evolution eqns

$$t = \log \frac{Q^2}{\mu_F^2}$$

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

Dokshitzer, Gribov, Lipatov, Altarelli, Parisi equations



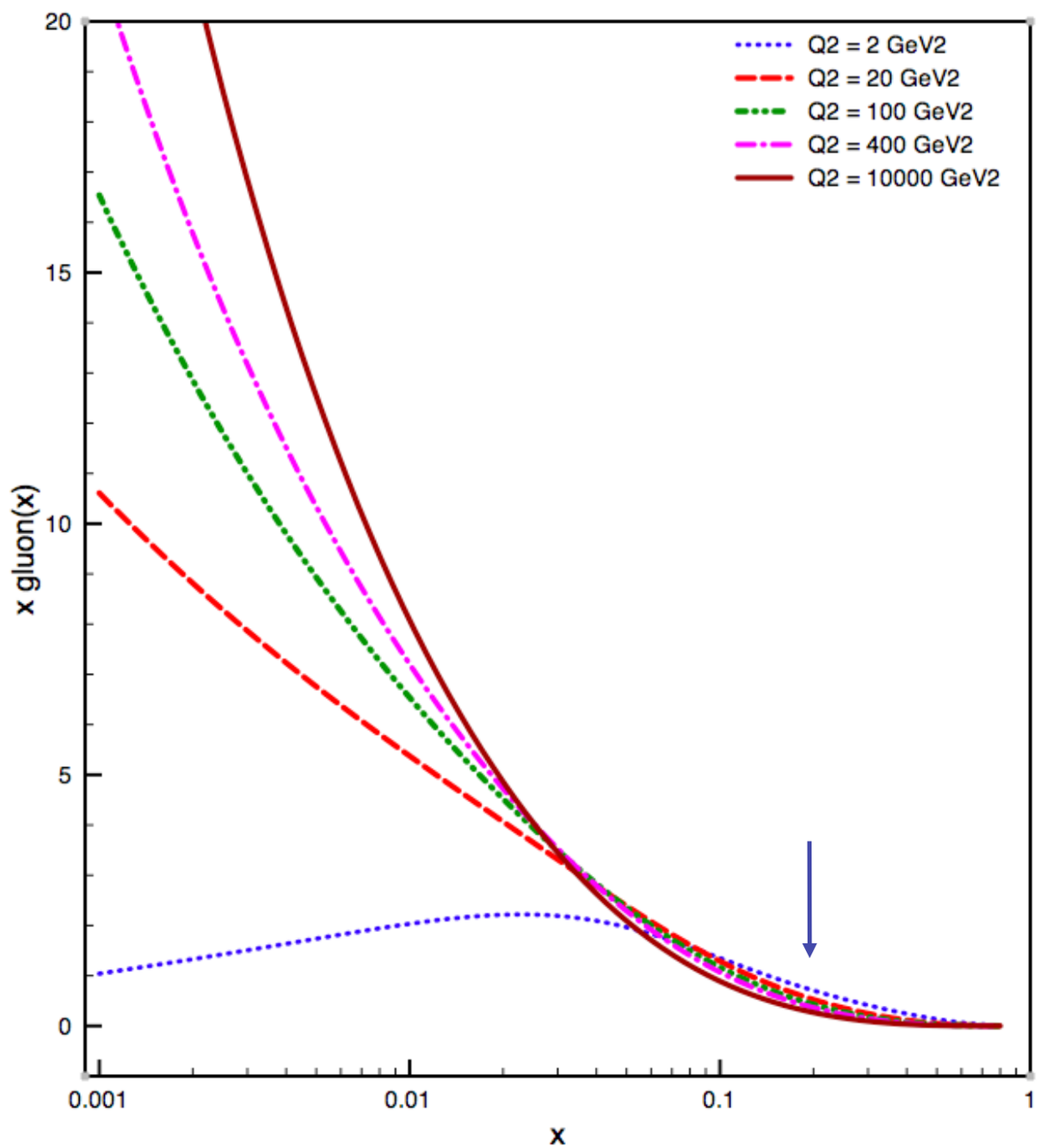
- 13 coupled integro-differential equations, which can be reduced to 11 decoupled and 2 coupled equation (with a change of basis in the space of PDFs)
- Splitting functions known up to NNLO:

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



Gluon evolution

$$g(x,\mu^2) = \Gamma_{gq} \otimes \Sigma(x,\mu_0^2) + \Gamma_{gg} \otimes g(x,\mu_0^2)$$

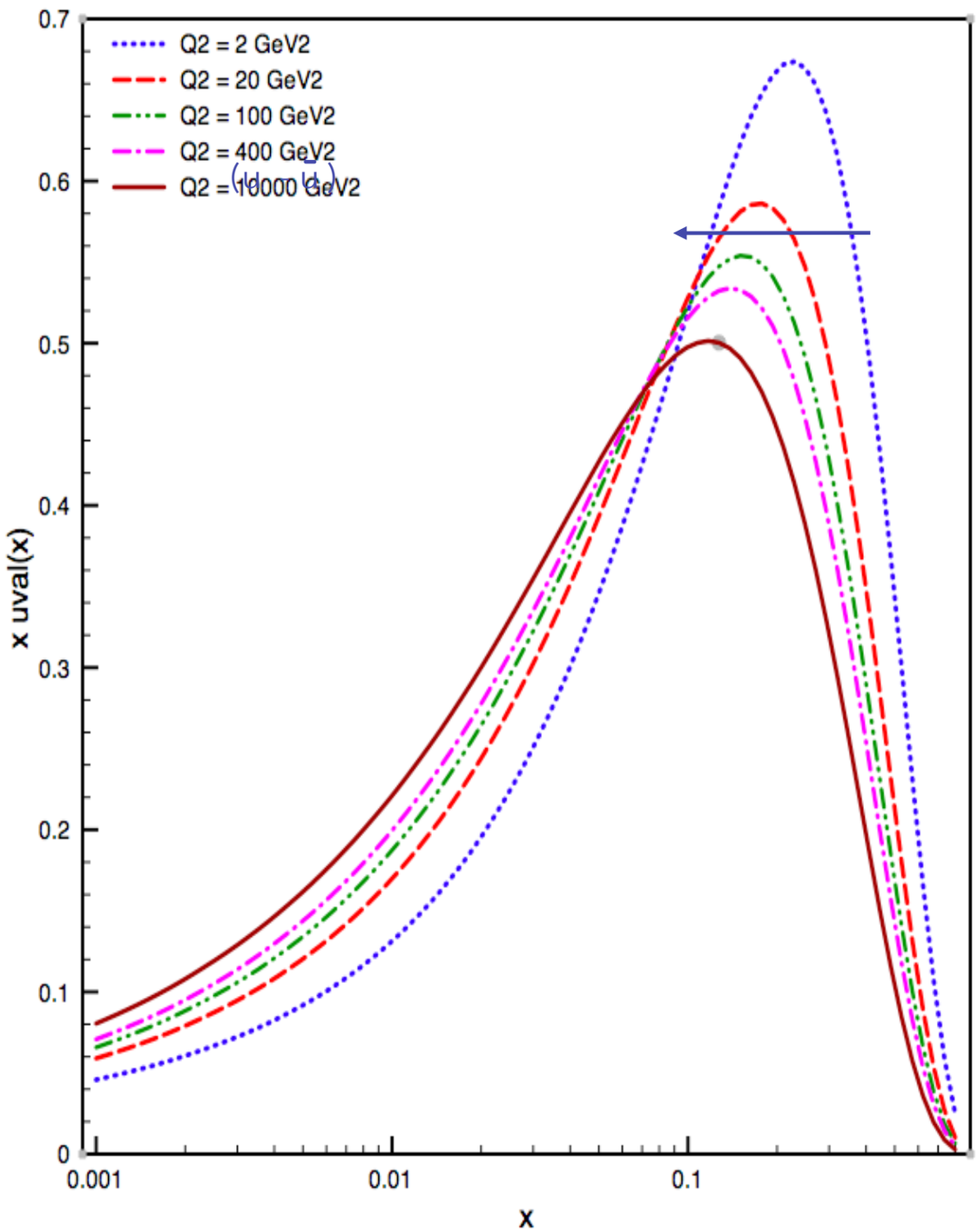
$$P_{gq}^{(0)}(x) = C_F \left[\frac{1+(1-x)^2}{x}\right]$$

$$P_{gg}^{(0)}(x) = 2N \left[\frac{x}{(1-x)_+} + \frac{1-x}{x} + x(1-x)\right] + \delta(1-x)\frac{(11N - 4n_f T_R)}{6}$$

- Both P_{gg} and P_{gq} diverge for $x \to 0$

- Gluon is depleted at large x

DGLAP EVOLUTION EQUATIONS



Non-singlet valence evolution

$$u_v(x,\mu^2) = \Gamma^v_{NS} \otimes u_v(x,\mu_0^2)$$

$$P^{(0),v}_{NS} = P^{(0)}_{qq}(x) = C_F\left[\frac{(1+x^2)}{(1-x)_+} + \frac{3}{2}\delta(1-x)\right]$$

- As Q^2 increases partons lose longitudinal momentum; distributions all shift to lower x

- Gluons can be seen because they help drive the quark evolution

11.11 HOW PDFS ARE EXTRACTED FROM DATA

DGLAP EVOLUTION EQUATIONS

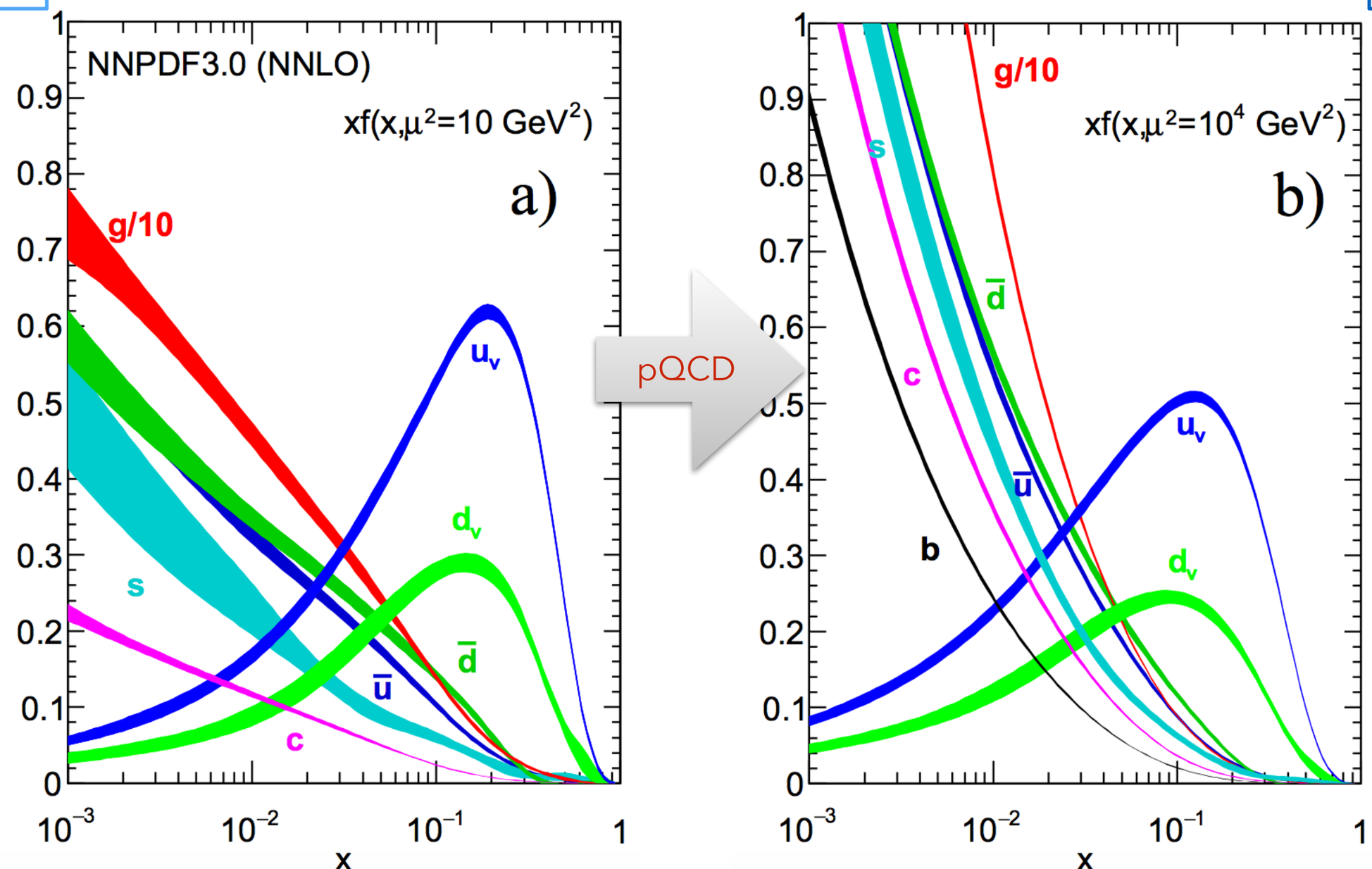
$$f_i(x, \mu)$$

Hadronic scale: global fit of PDFs

Data

Perturbative QCD

High scale: input to the LHC



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

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

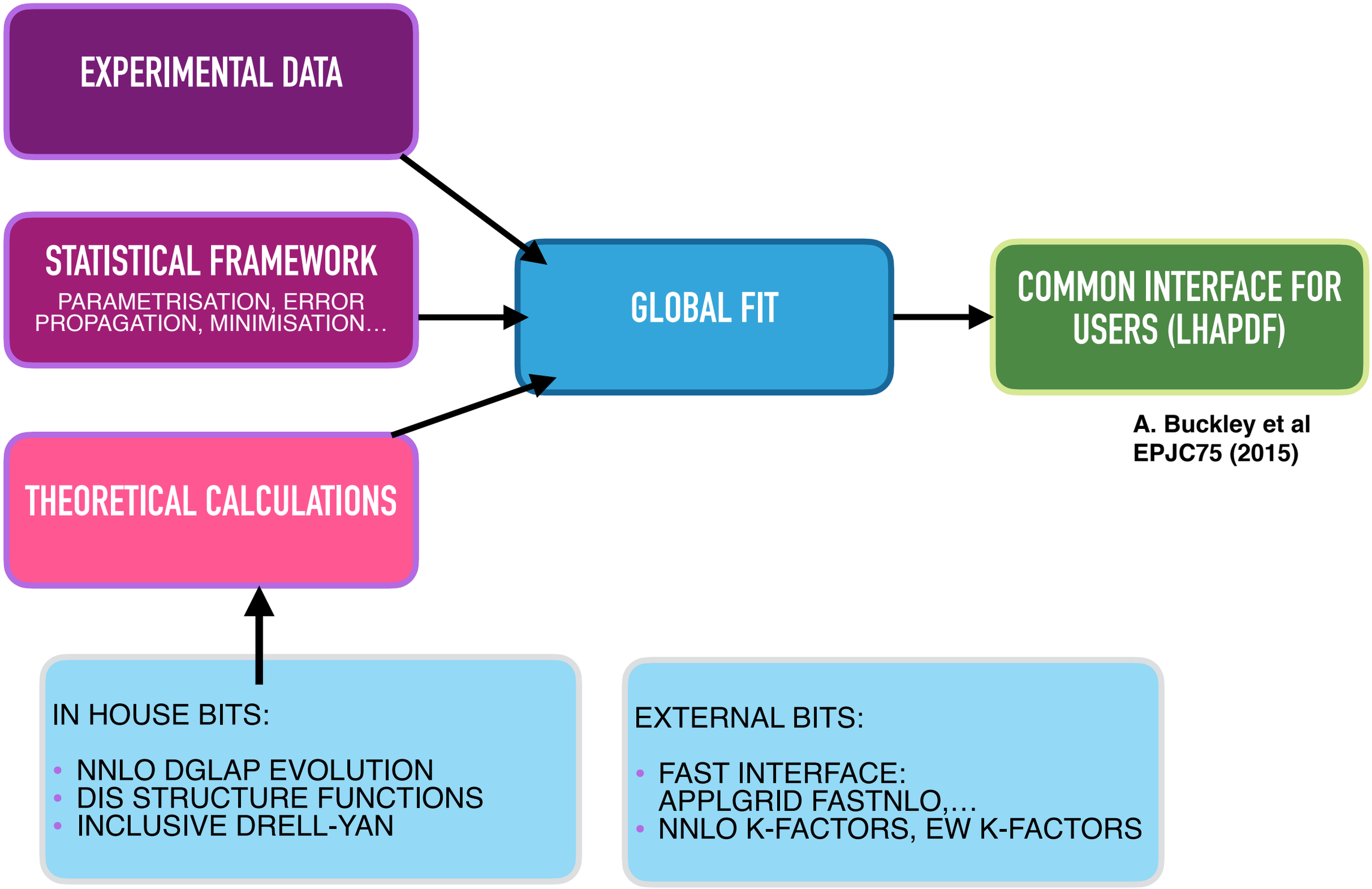
THE NAME OF THE GAME

- 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

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

A COMPLEX MACHINERY



DISENTANGLING PDFS WITH EXPERIMENTAL DATA

- PDFs are not measurable quantities, we measure observables that are convolutions of PDFs and partonic cross sections

D_i

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

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

T_i

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

- Must exclude regions where factorisation fails to apply (low Q2 and large x)

$$Q^2_{\rm min} = 2\,\mathrm{GeV}^2$$

$$W^2_{\rm min} = \left(Q^2\frac{1-x}{x}\right)_{\rm min} = 12.5\,\mathrm{GeV}^2$$

DISENTANGLING PDFS WITH EXPERIMENTAL DATA

- PDFs are not measurable quantities, we measure observables that are convolutions of PDFs and partonic cross sections

D_i

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

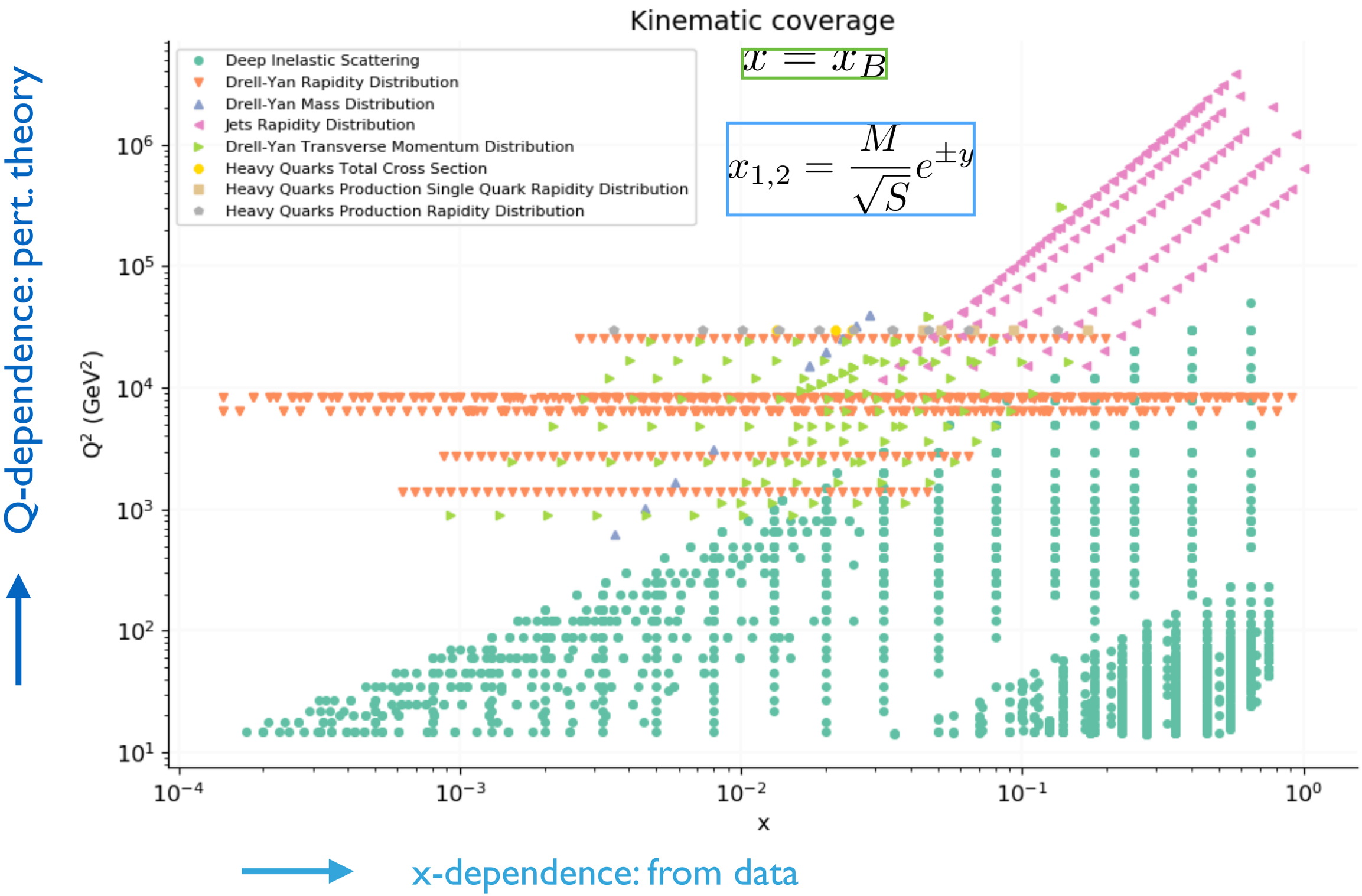
$$\frac{d\sigma_H^{pp\to ab}}{dX} = \sum_{i,j=-n_f}^{+n_f} \int_{\tau_0}^{1} \frac{dz_1}{z_1}\frac{dz_2}{z_2} f_i(z_1,\mu_F) f_j(z_2,\mu_F) \frac{d\hat{\sigma}_i^{ij}}{dX}(zS,\alpha_s(\mu_R),\mu_F) + \mathcal{O}\left(\frac{\Lambda^n}{S^n}\right)$$

T_i

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

- Different data constrain different PDF combinations in different regions
  - ➡ DIS data on proton abundant and precise (HERA)
  - ➡ In principle F_2, F_3 CC provide 2 light quark combinations
    F_2, F_3 NC provide 2 extra light quark combinations ➜ 4 combinations of PDFs
  - ➡ Old DIS and Drell-Yan data still used because of nuclear targets (isospin symmetry)
  - ➡ W,Z boson final state provide lot of information on quarks
  - ➡ Processes with jets and/or heavy quark in final states direct handle on the gluon

DISENTANGLING PDFS WITH EXPERIMENTAL DATA



ERROR PROPAGATION: THE HESSIAN SOLUTION

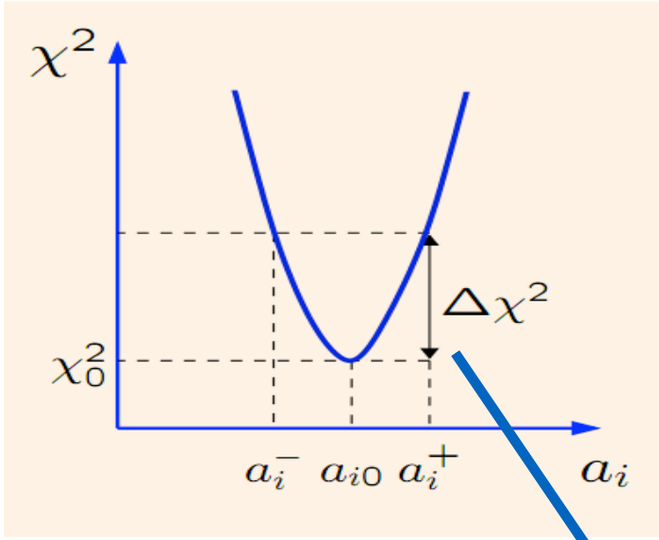
$$\langle \mathcal{O}[\{f\}] \rangle = \int [\mathcal{D}f]\, \mathcal{O}[\{f\}] \mathcal{P}[\{f\}]$$

- 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

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

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



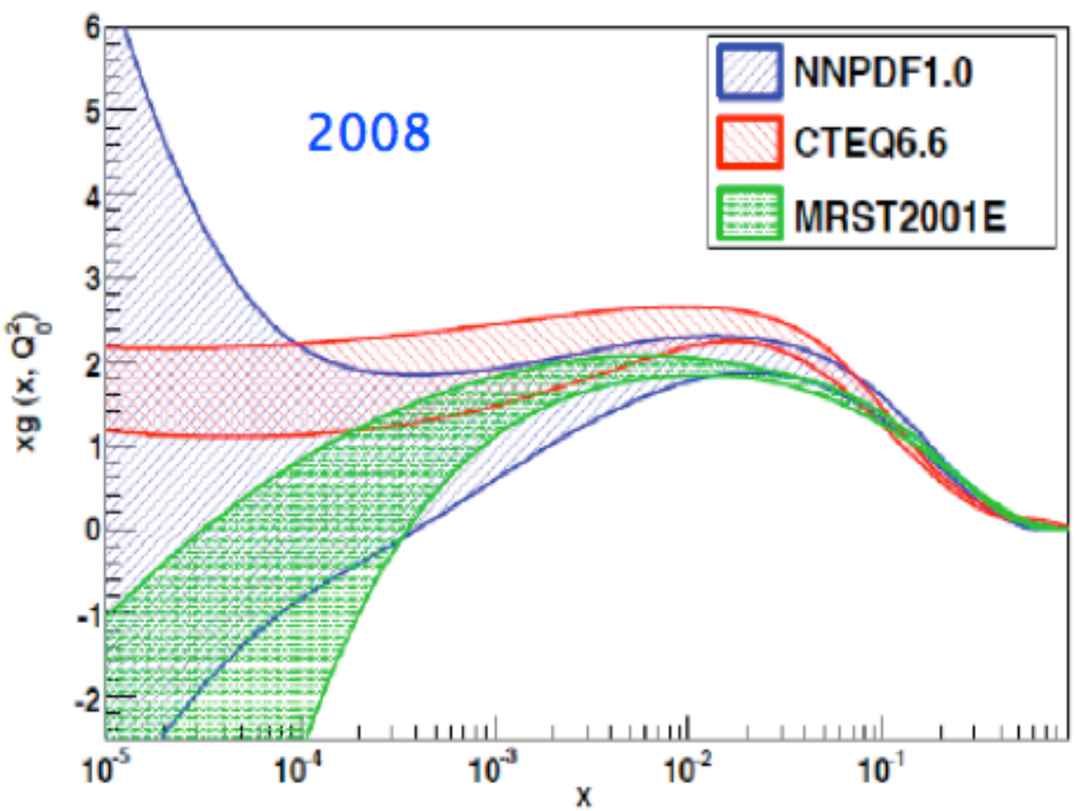
Tolerance

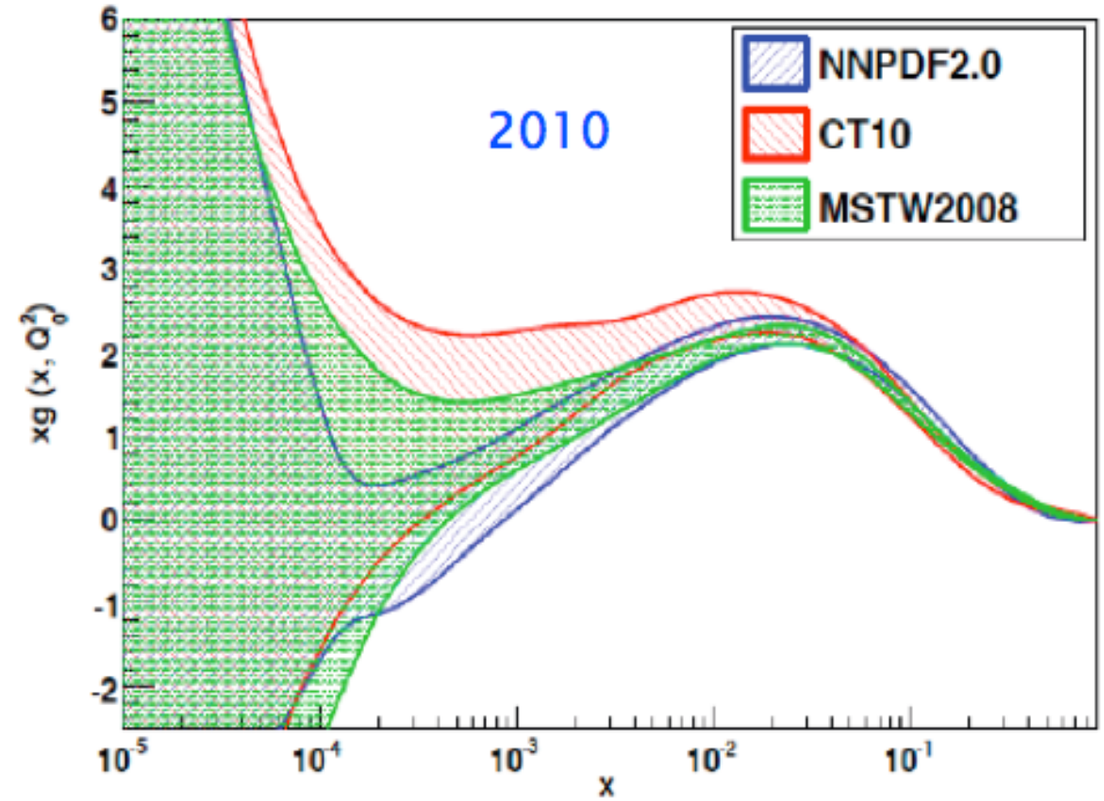
Parametrisation

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

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

ERROR PROPAGATION: THE HESSIAN SOLUTION



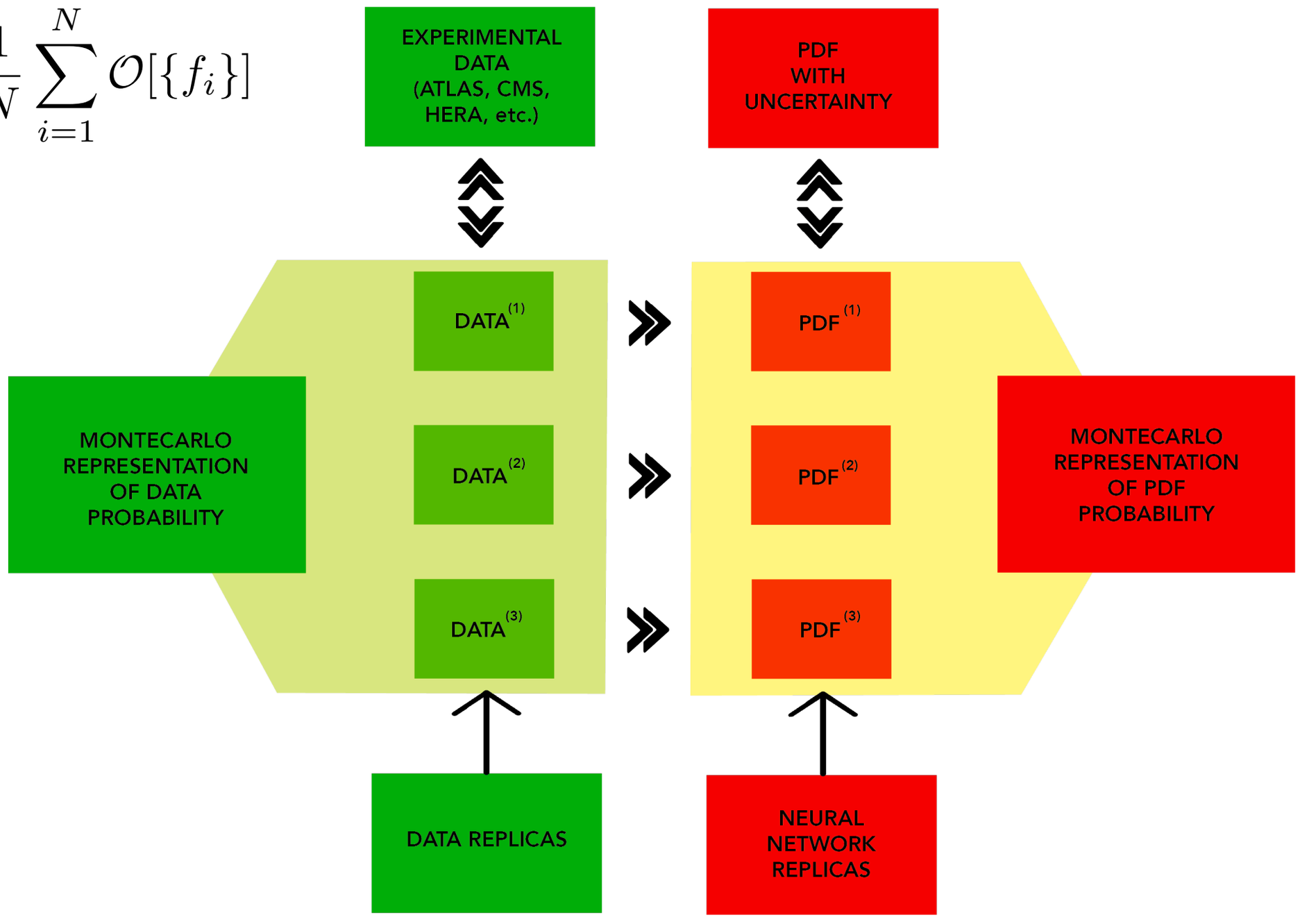


$$xg = A_g\, x^{\delta_g} (1-x)^{\eta_g} (1 + \epsilon_g \sqrt{x} + \gamma_g\, x) + A_{g'}\, x^{\delta_{g'}} (1-x)^{\eta_{g'}}$$

- 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)

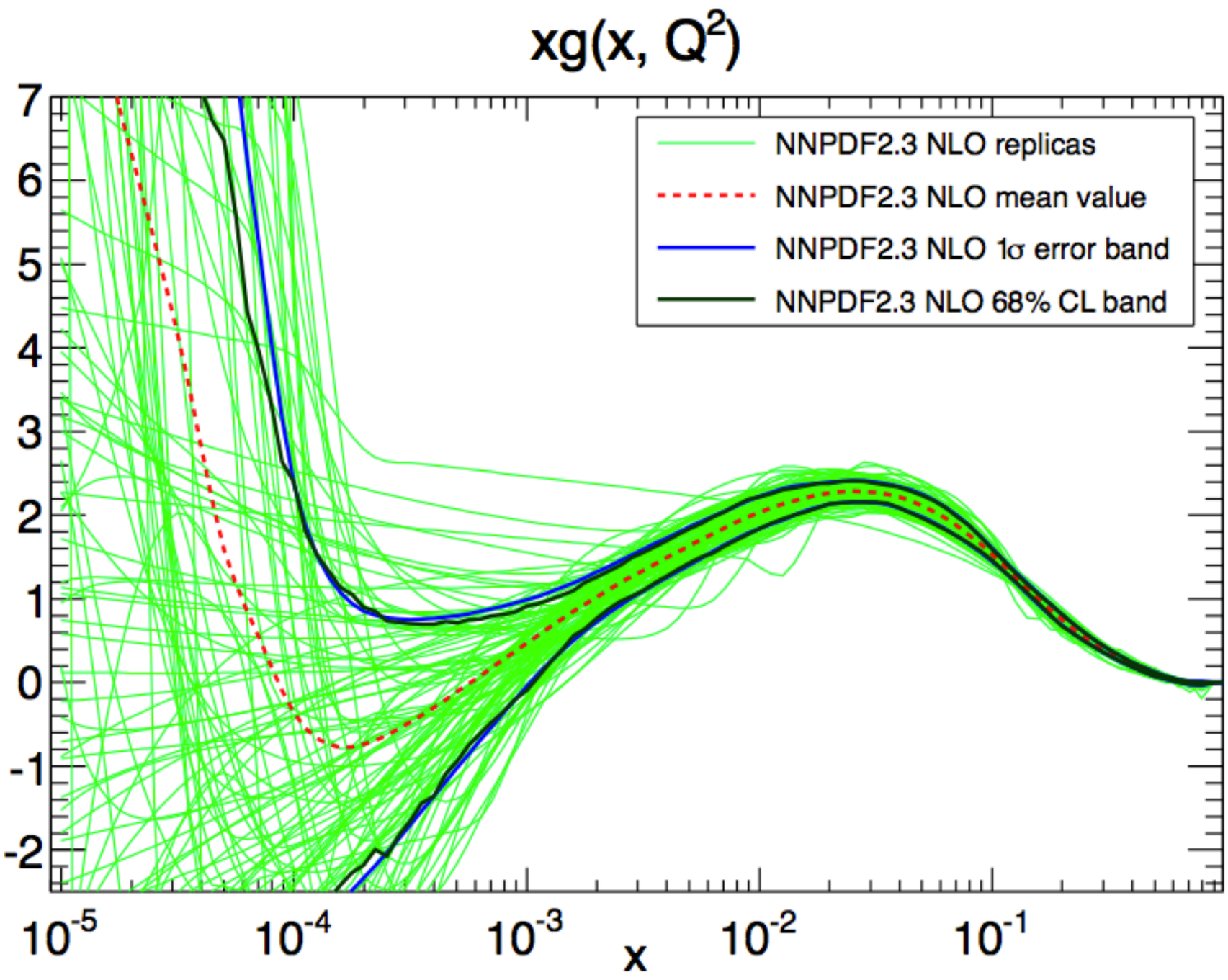
ERROR PROPAGATION: THE MONTE CARLO SOLUTION

$$\langle \mathcal{O}[\{f\}] \rangle \simeq \frac{1}{N} \sum_{i=1}^{N} \mathcal{O}[\{f_i\}]$$



Ball, Del Debbio, Forte, Guffanti, Latorre, Rojo, MU (2008)…
The NNPDF collaboration, EPJC 77 (2017)

THE NNPDF APPROACH



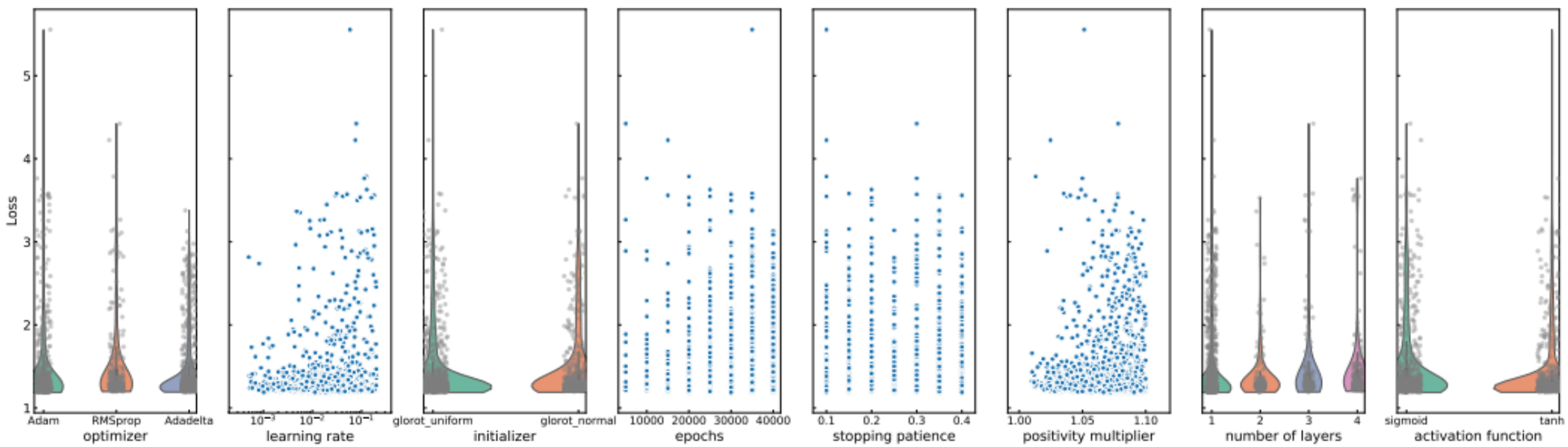
Monte Carlo technique: sampling the probability measure in PDF functional space by projecting from sampling in data space

Neural Networks: all independent PDFs are associated to an unbiased and flexible parametrization: O(300) parameters versus O(30) in polynomial parametrization

Ball, Del Debbio, Forte, Guffanti, Latorre, Rojo, MU (2008)...
The NNPDF collaboration, EPJC 77 (2017)

A NEW MACHINE-LEARNING DRIVEN METHODOLOGY

Each blue dot corresponds to a fit of a different set of hyperparameters:



Thousands of fits for the hyperoptimization algorithm to choose:

- ✓ Optimizer
- ✓ Initializer
- ✓ Stopping Patience
- ✓ Number of Layers
- ✓ Learning Rate
- ✓ Epochs
- ✓ Positivity Multiplier
- ✓ Activation Function

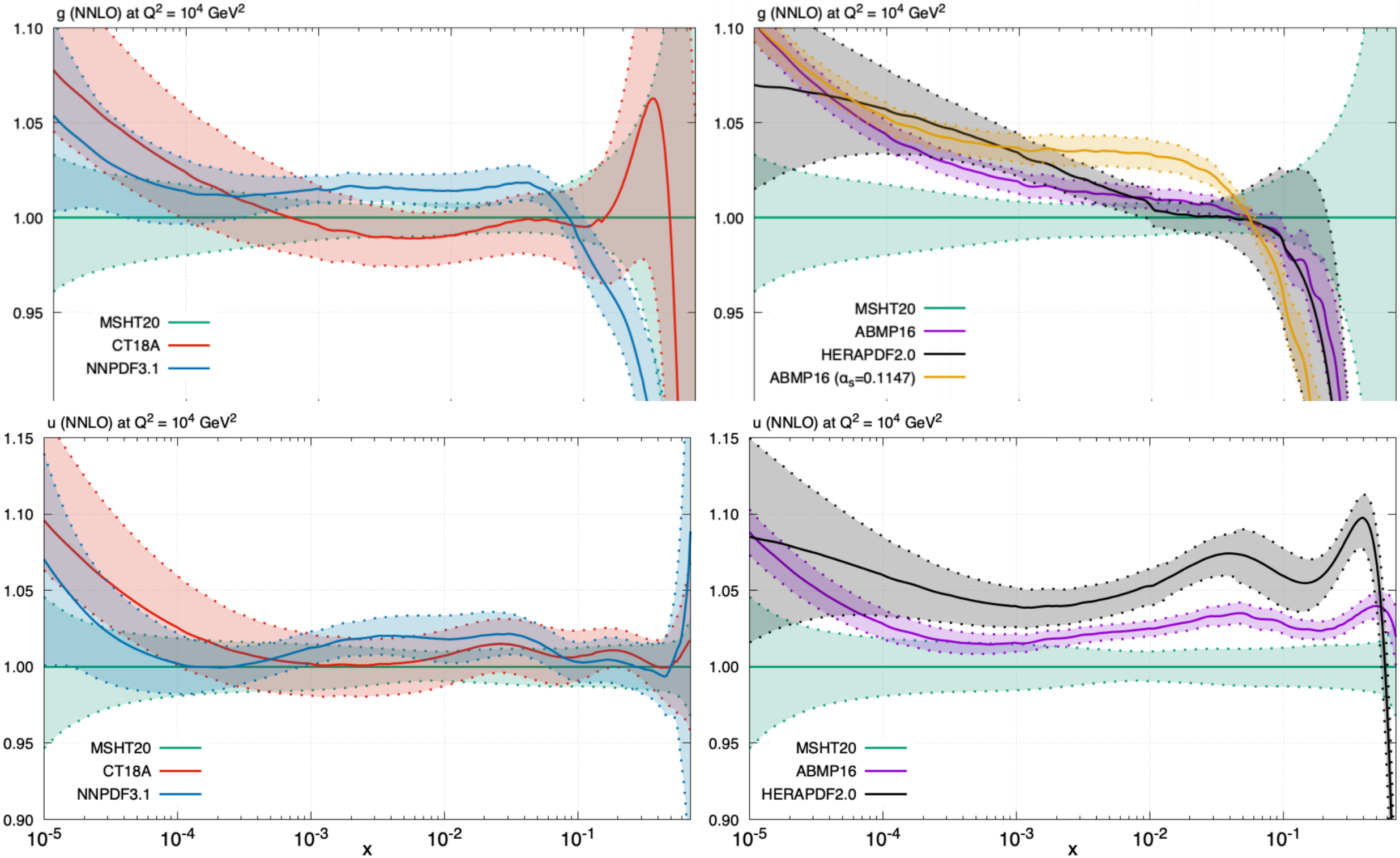
J. Cruz-Martinez slide from PDF4LHC2020

II.III STATE OF THE ART

STATE OF THE ART

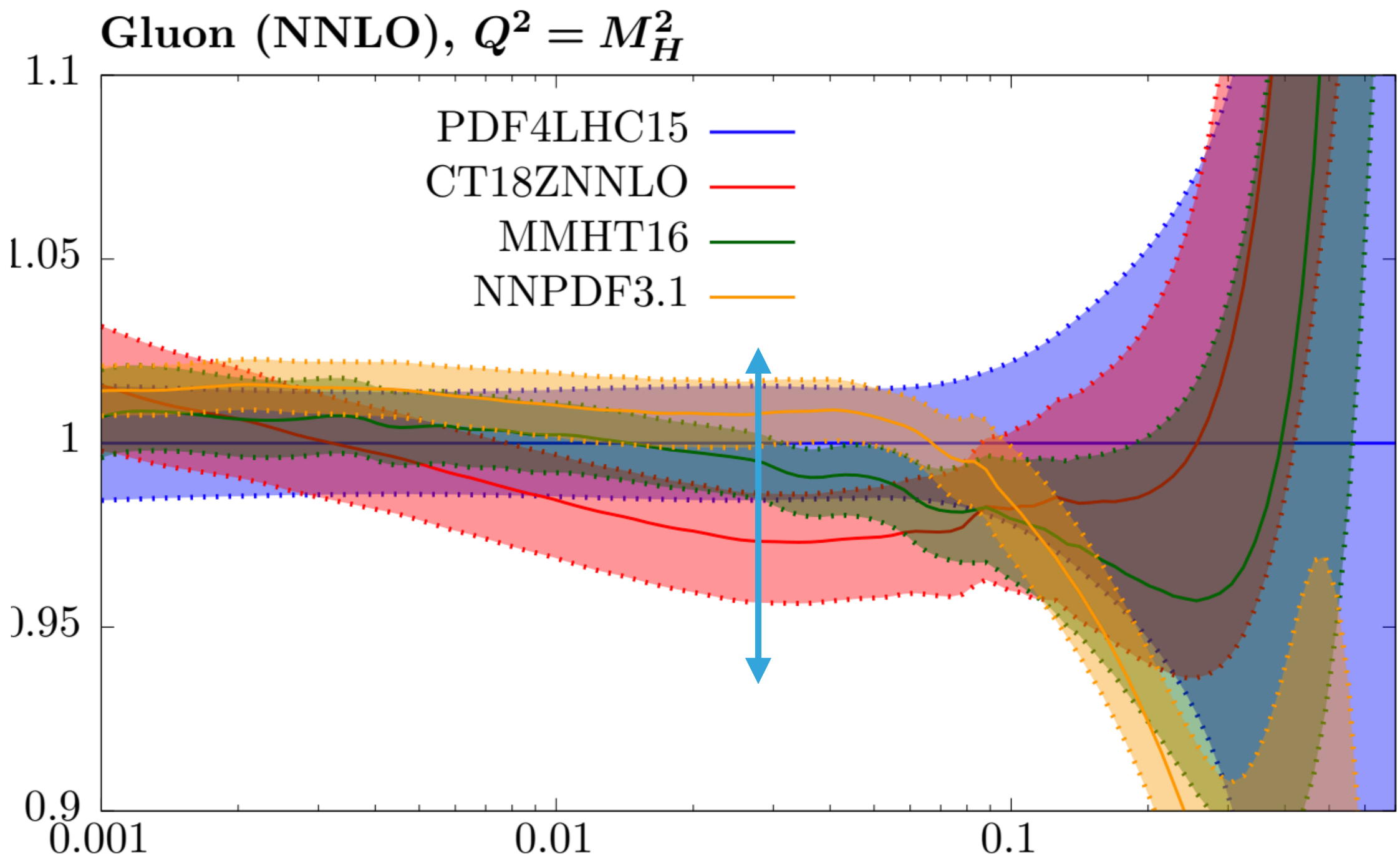
January 2021	NNPDF3.1	MSHT20	CT18	ABMP16
Fixed Target DIS	✔	✔	✔	✔
HERA I+II	✔	✔	✔	✔
Fixed Target DY	✔	✔	✔	✔
Tevatron W,Z	✔	✔	✔	✔
Tevatron jets	✔	✔	✔	✘
LHC jets	✔	✔	✔	✘
LHC vector boson	✔	✔	✔	✔
LHC top	✔	✔	✔	✔
Stat. treatment	Monte Carlo	Hessian $\Delta\chi^2$ dynamical	Hessian $\Delta\chi^2$ dynamical	Hessian $\Delta\chi^2=1$
Parametrization	Neural Networks (259 pars)	Chebyshev (36 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

STATE OF THE ART



T. Cridge et al, arXiv:2012.04684

STATE OF THE ART



III. NEW FRONTIERS AT LHC RUN III

IMPACT OF THE LHC DATA ON PDFS

EXPERIMENTAL DATA

- Effect of and issues with correlated and precise LHC data
- How to deal with data incompatibilities?

STATISTICAL FRAMEWORK

PARAMETRISATION, ERROR PROPAGATION, MINIMISATION…

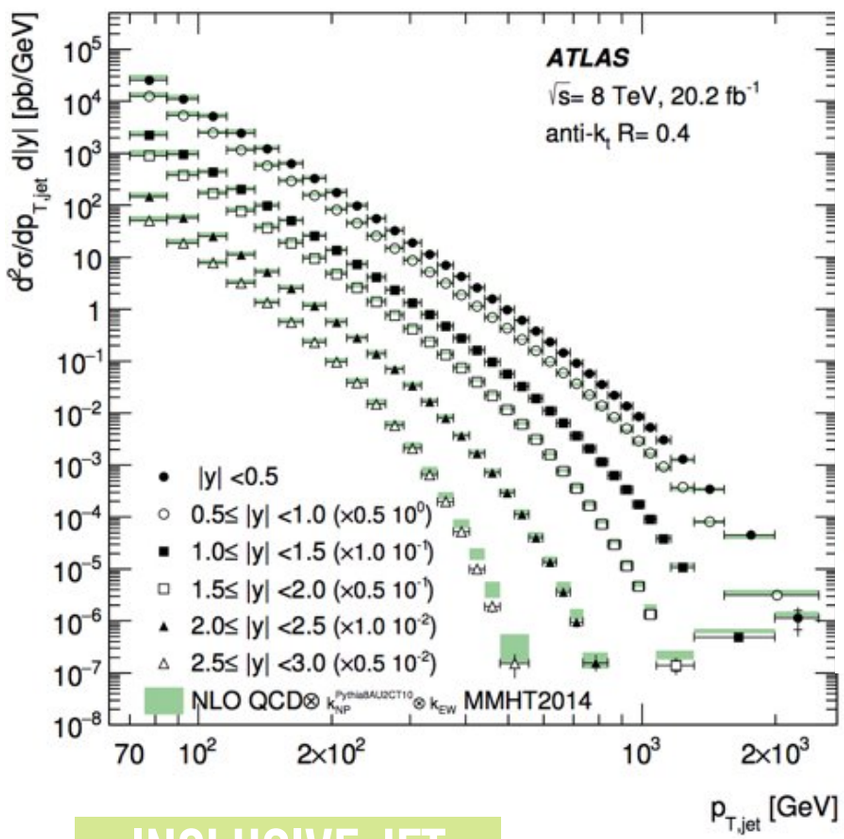
- Flexible parametrisation
- Robust minimisation

THEORETICAL CALCULATIONS

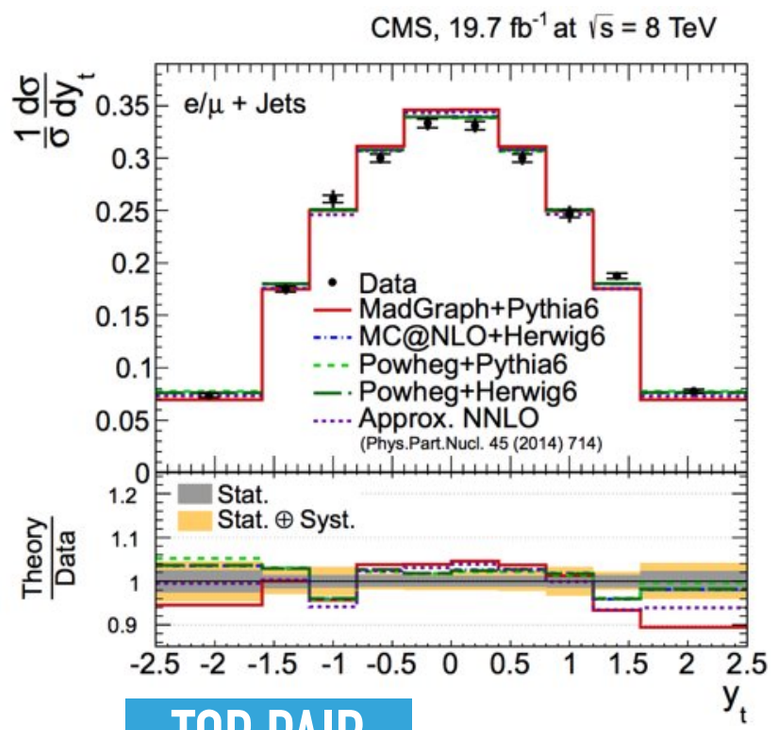
- N3LO
- Theory uncertainties in PDF fits
- Simultaneous fits
- Interplay between new physics and PDFs in high energy tails

III.1 IMPACT OF LHC DATA

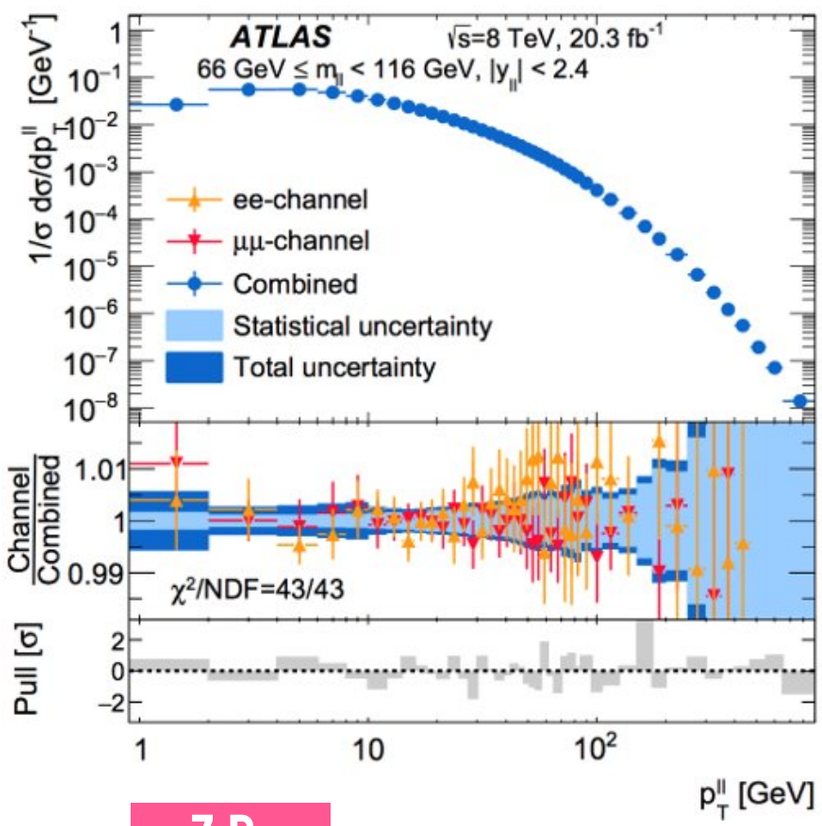
IMPACT OF THE LHC DATA ON PDFS



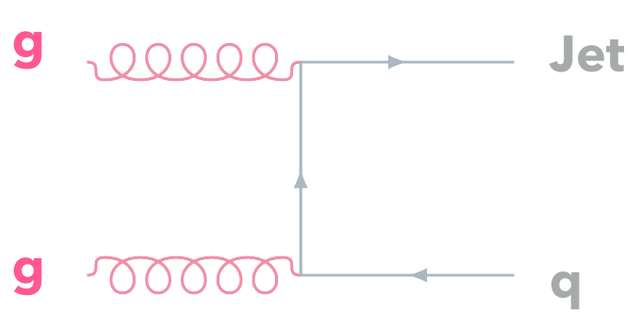
INCLUSIVE JET

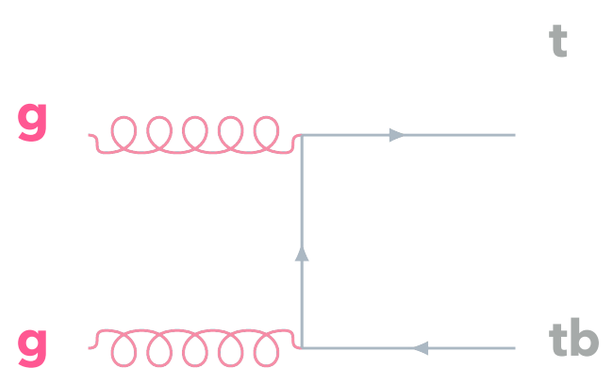


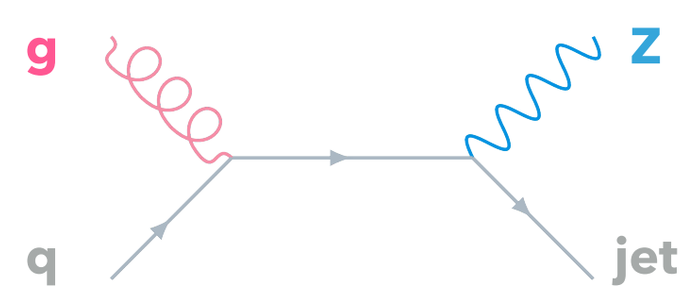
TOP PAIR



Z P_T







Ball et at, EPJC 77 (2017)

IMPACT OF THE LHC DATA ON PDFS

- NNLO calculations are essential to reduce theoretical uncertainties in PDF analyses
- Stunning progress has been made on some key processes for PDF determination

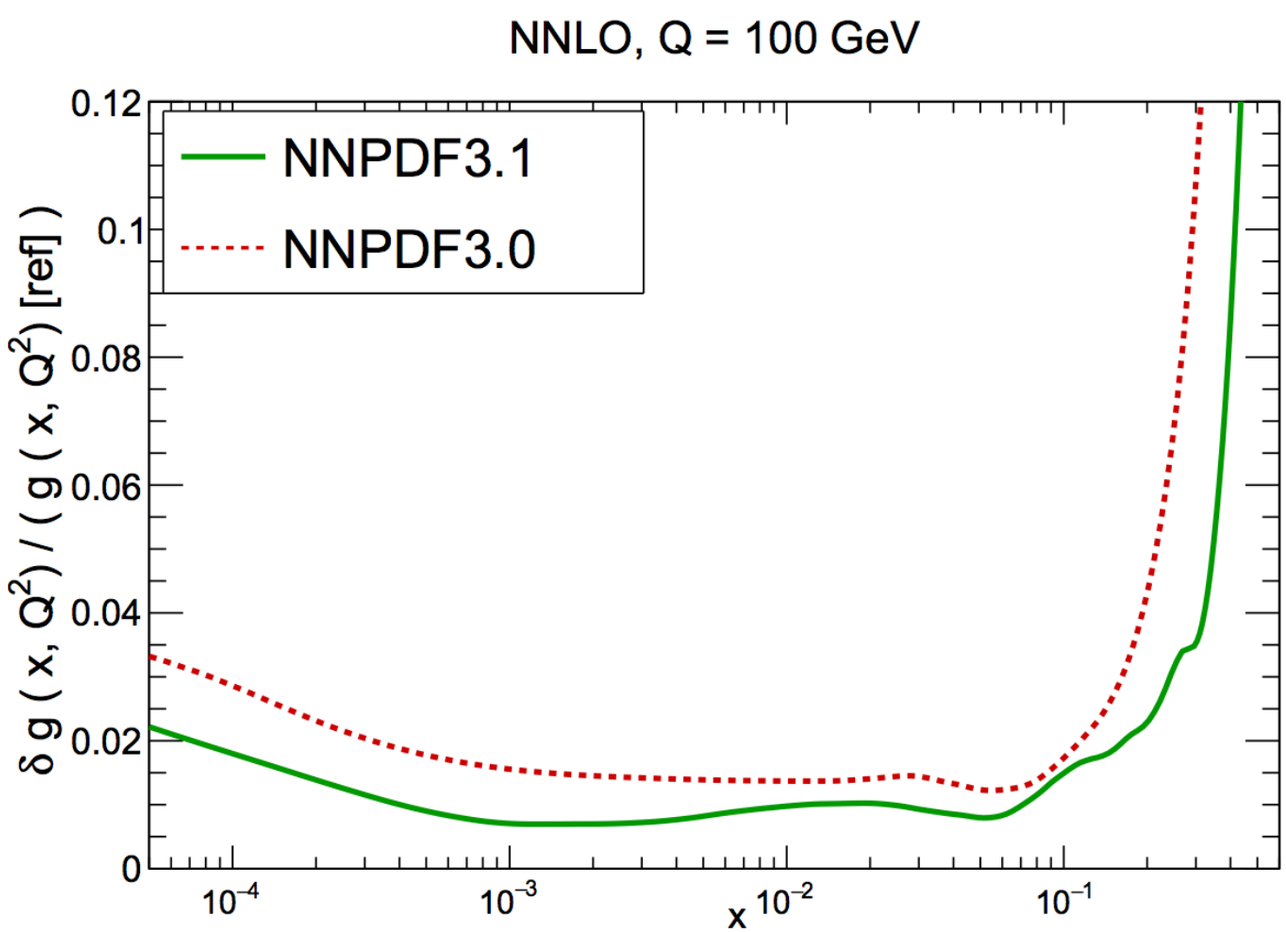
✓ NNLO top pair production
Czakon, Fiedler, Mitov [PRL 116(2016) 082003]
Czakon, Mitov [JHEP 1301(2015)]

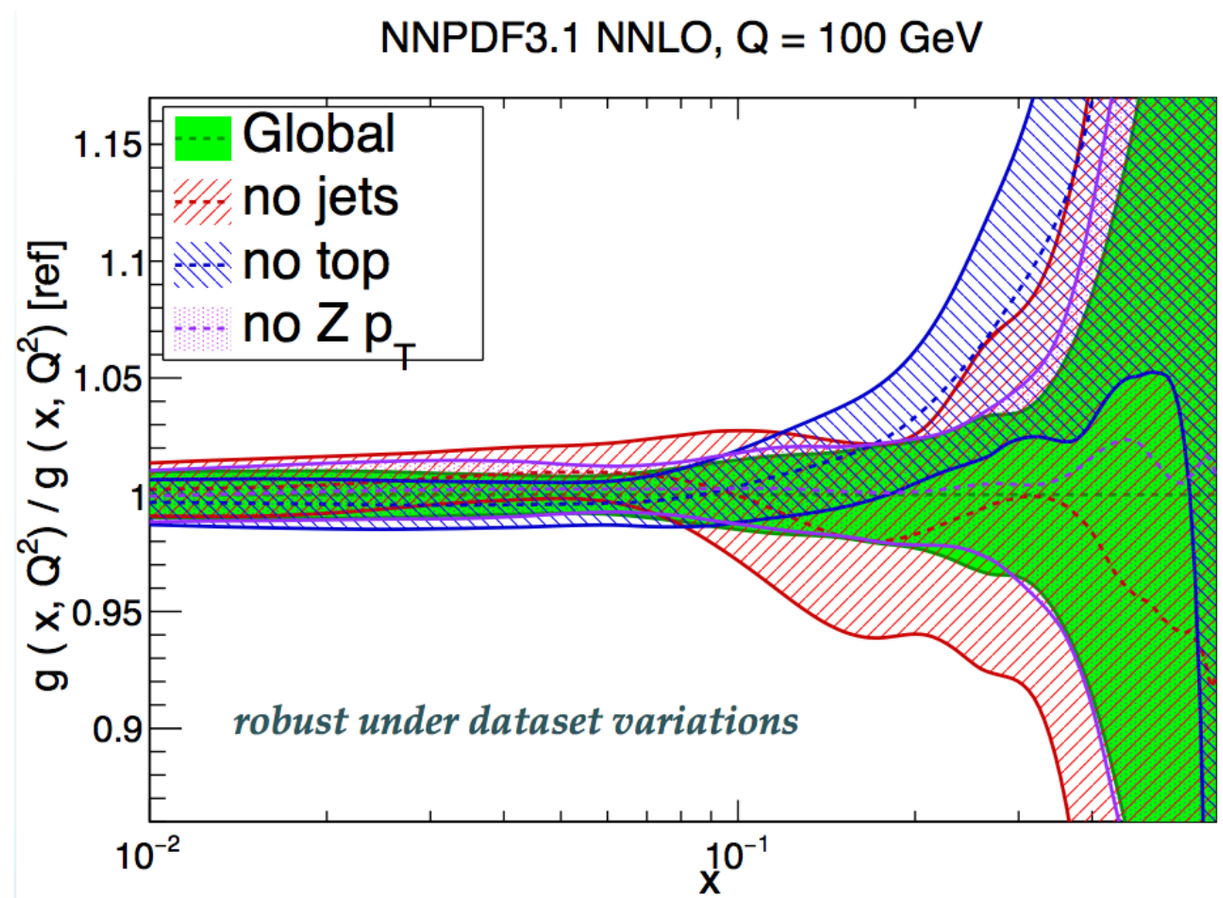
✓ W/Z+j and W/Z transverse momentum distributions
Gehrmann-De Ridder et al [1605.04295]
Boughezal, Liu, Petriello [1602.08140]
Boughezal, Liu, Petriello [1602.06965]
Boughezal et al [PRL 116(2016) 152001 & 062002]
Gehrmann-De Ridder et al [1507.02850]

✓ Inclusive jet cross section
Currie et al. [1611.01460]
Gehrmann-De Ridder et al [PRL 110 (2016)
Currie et al [JHEP 1401 (2014) 110]

IMPACT OF THE LHC DATA ON PDFS

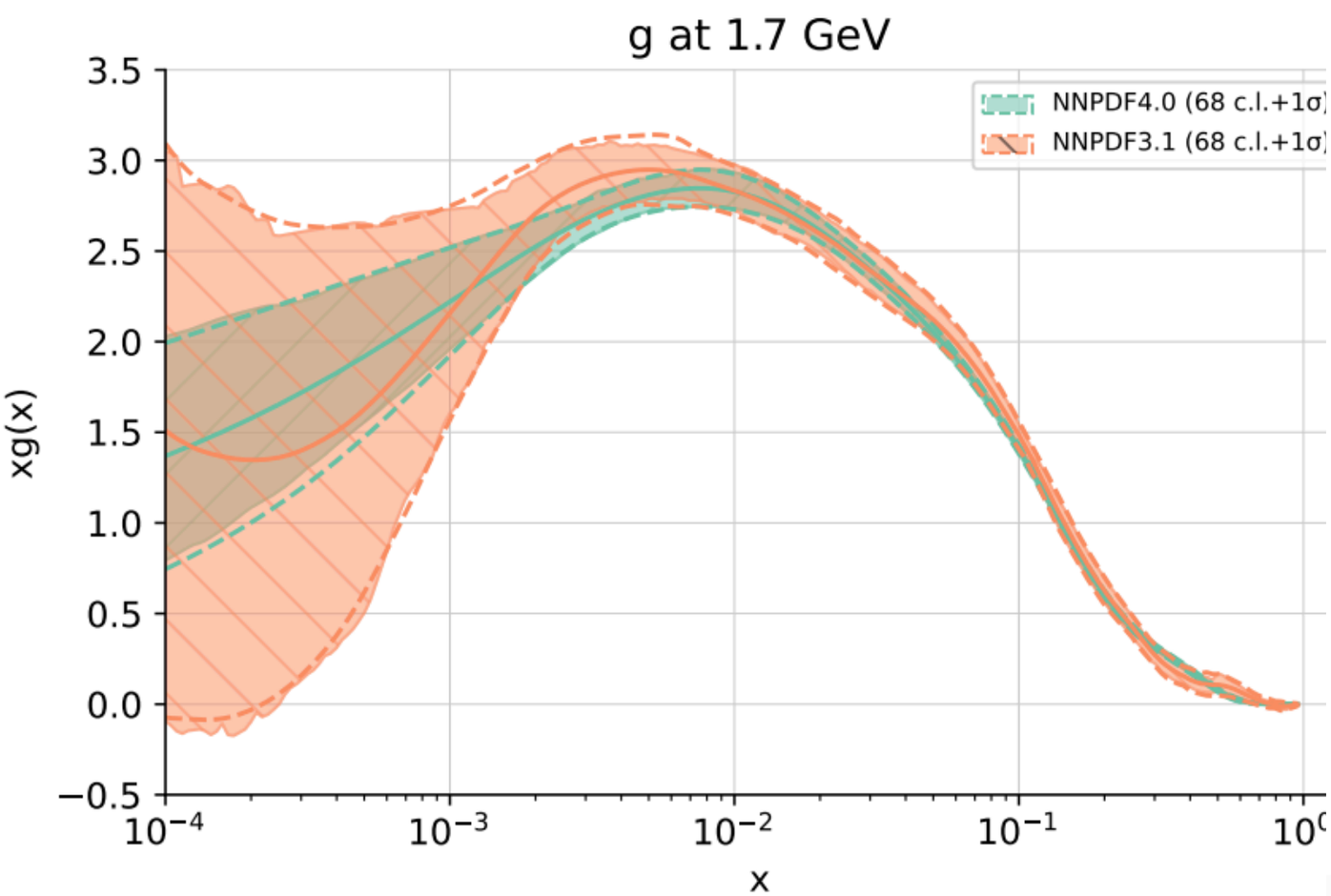
- Large-x gluon constrained by three independent processes
- Consistent picture and analysis robustness





LATEST RESULTS

- Increasingly important the inclusion of the LHC data and new observables
- Compatible with previous results within uncertainties!



Data set	$N_{\rm dat}$	$\chi^2/N_{\rm dat}$
Fixed-target DIS	1881	1.10
HERA	1208	1.21
σ_c	37	2.11
σ_b	26	1.48
Fixed-target Drell-Yan	189	1.00
CDF	28	1.31
D0	37	1.00
ATLAS	621	1.18
Drell-Yan, 7, 8, 13 TeV	153	1.32
W+jet, 8 TeV	32	1.15
single top, 7, 8, 13 TeV	14	0.36
di-jets, 7 TeV	90	1.93
jets, 8 TeV	171	0.61
top pair, 7, 8, 13 TeV	16	2.30
Zp_T, 8 TeV	92	0.86
direct photon, 13 TeV	53	0.72
CMS	411	1.40
Drell-Yan, 7, 8 TeV	154	1.34
single top, 7, 8, 13 TeV	3	0.43
di-jets, 7 TeV	54	1.67
di-jets, 8 TeV	122	1.50
top pair, 5, 7, 8 TeV	29	0.84
top pair, 13 TeV	21	0.67
Zp_T, 8 TeV	28	1.42
LHCb	116	1.53
Total	4491	1.17

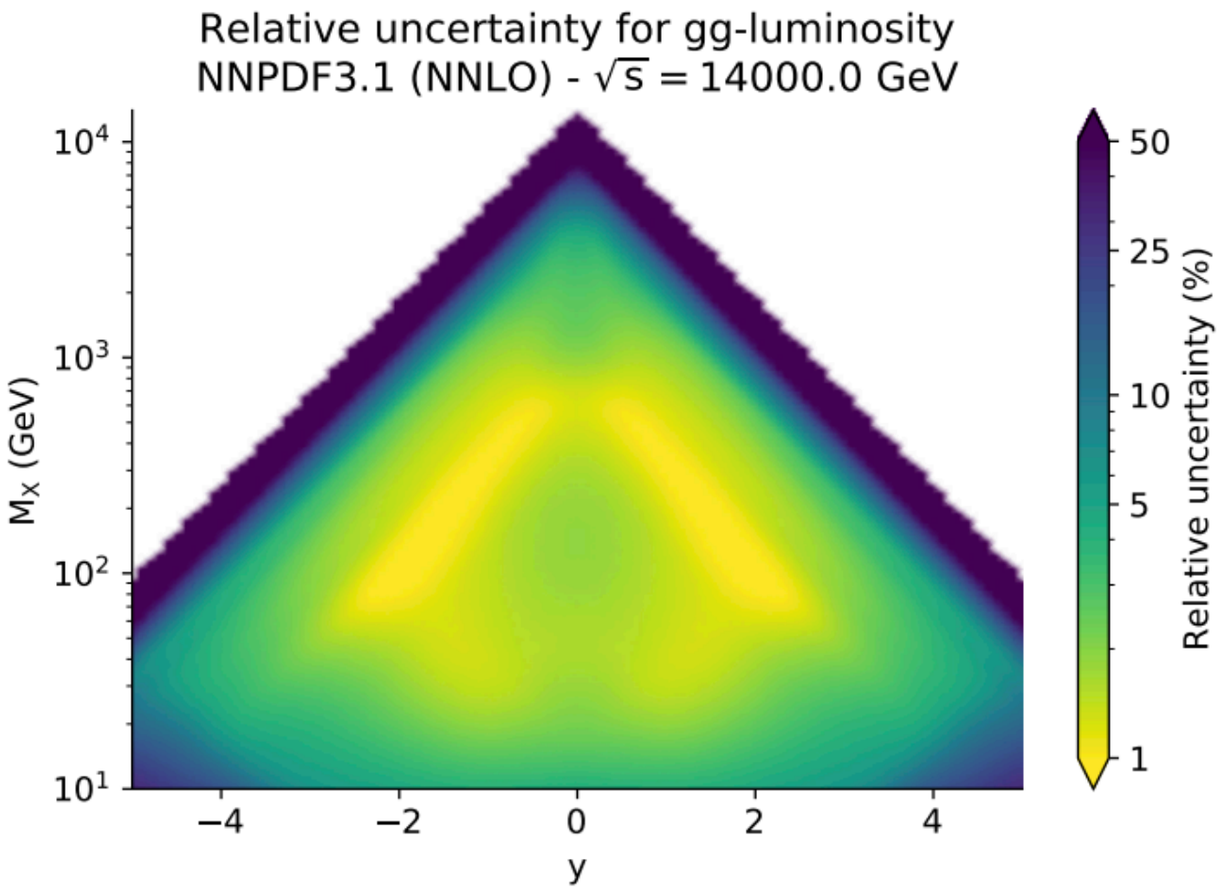
THE PRECISION VERSUS ACCURACY CHALLENGE

PRECISION CHALLENGE

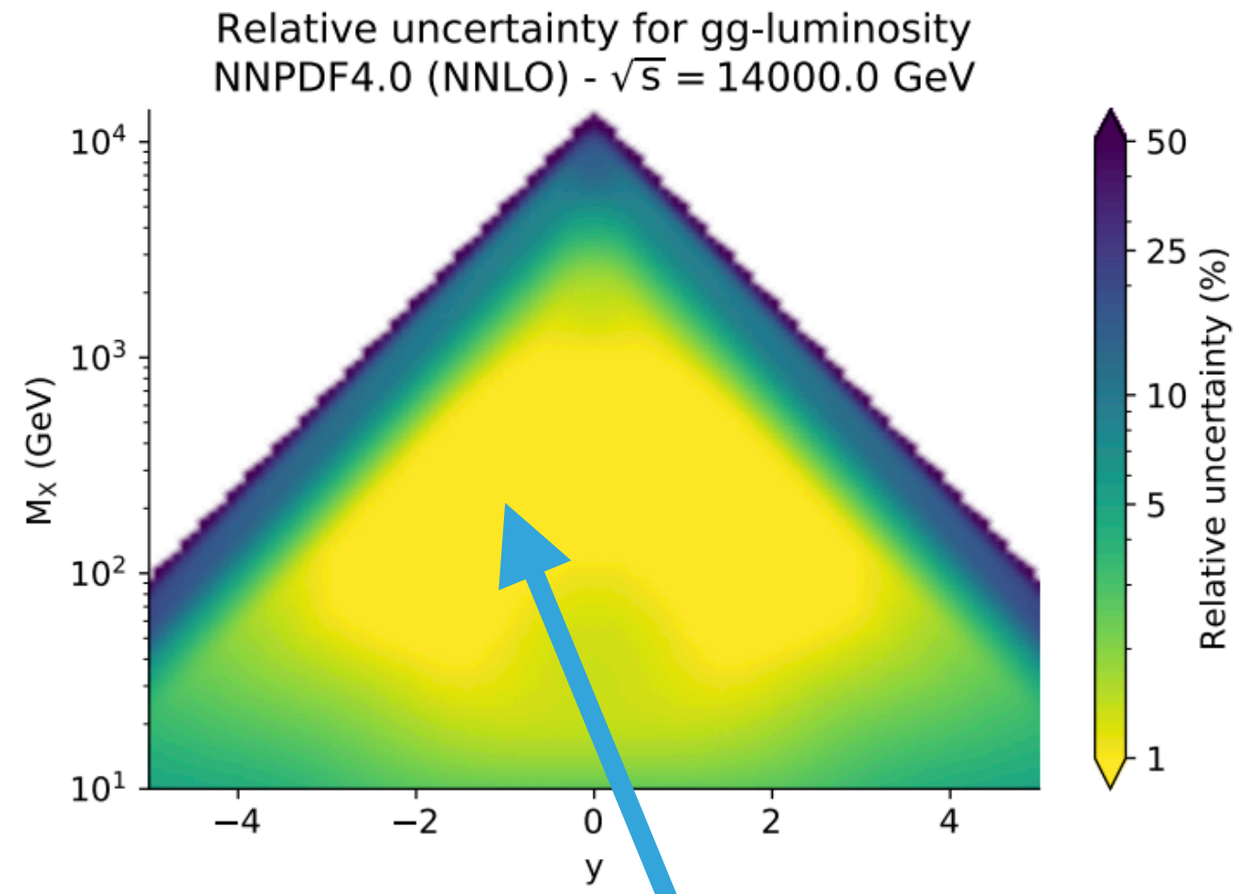
$$\mathcal{L}_{ij}(M_X, y, \sqrt{s}) = \frac{1}{s}\sum_{i,j} f_i\left(\frac{M_X e^{y}}{\sqrt{s}}, M_X\right) f_j\left(\frac{M_X e^{-y}}{\sqrt{s}}, M_X\right)$$

GLUON

NNPDF3.1 (NNLO)



NNPDF4.0 (NNLO)

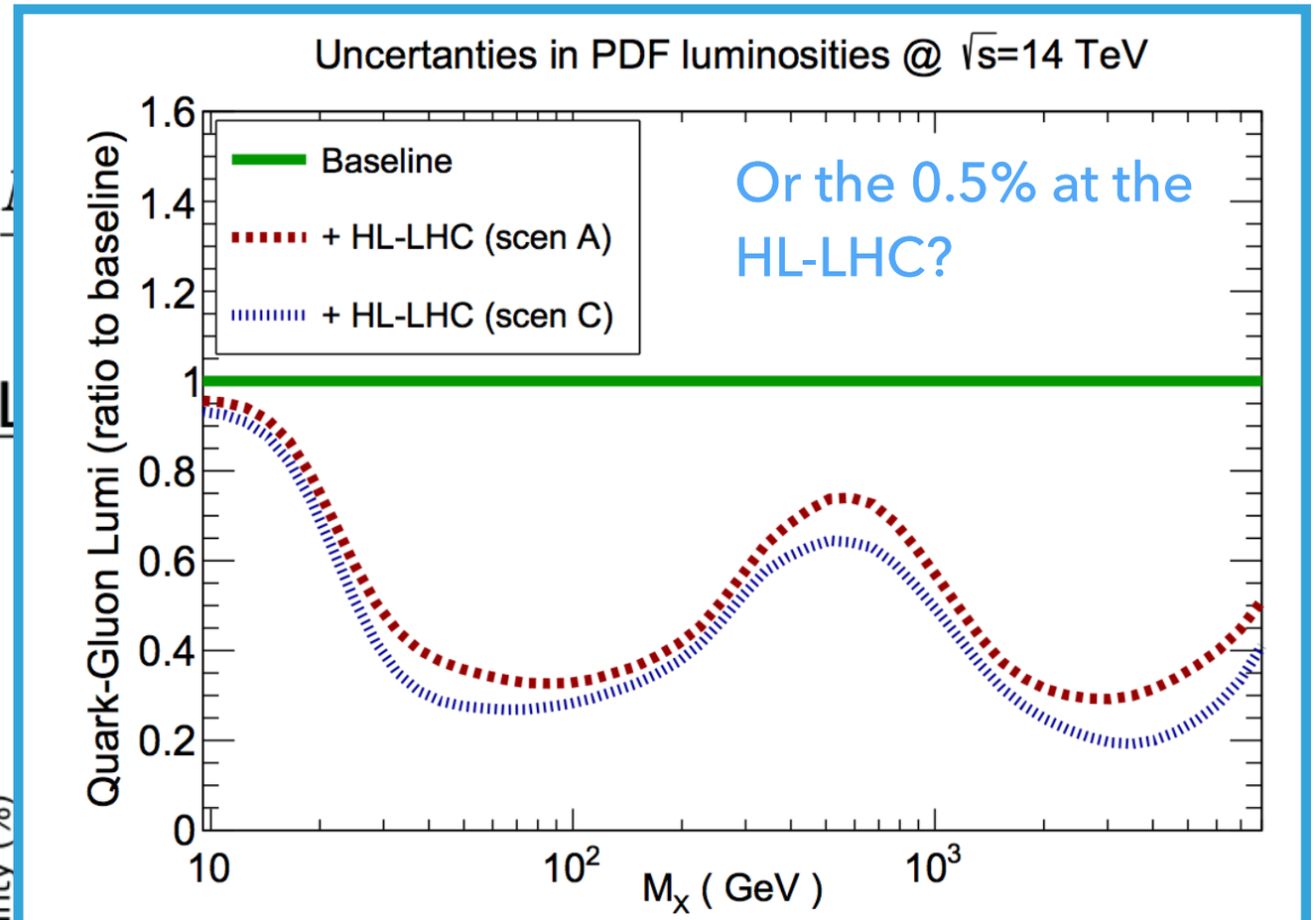


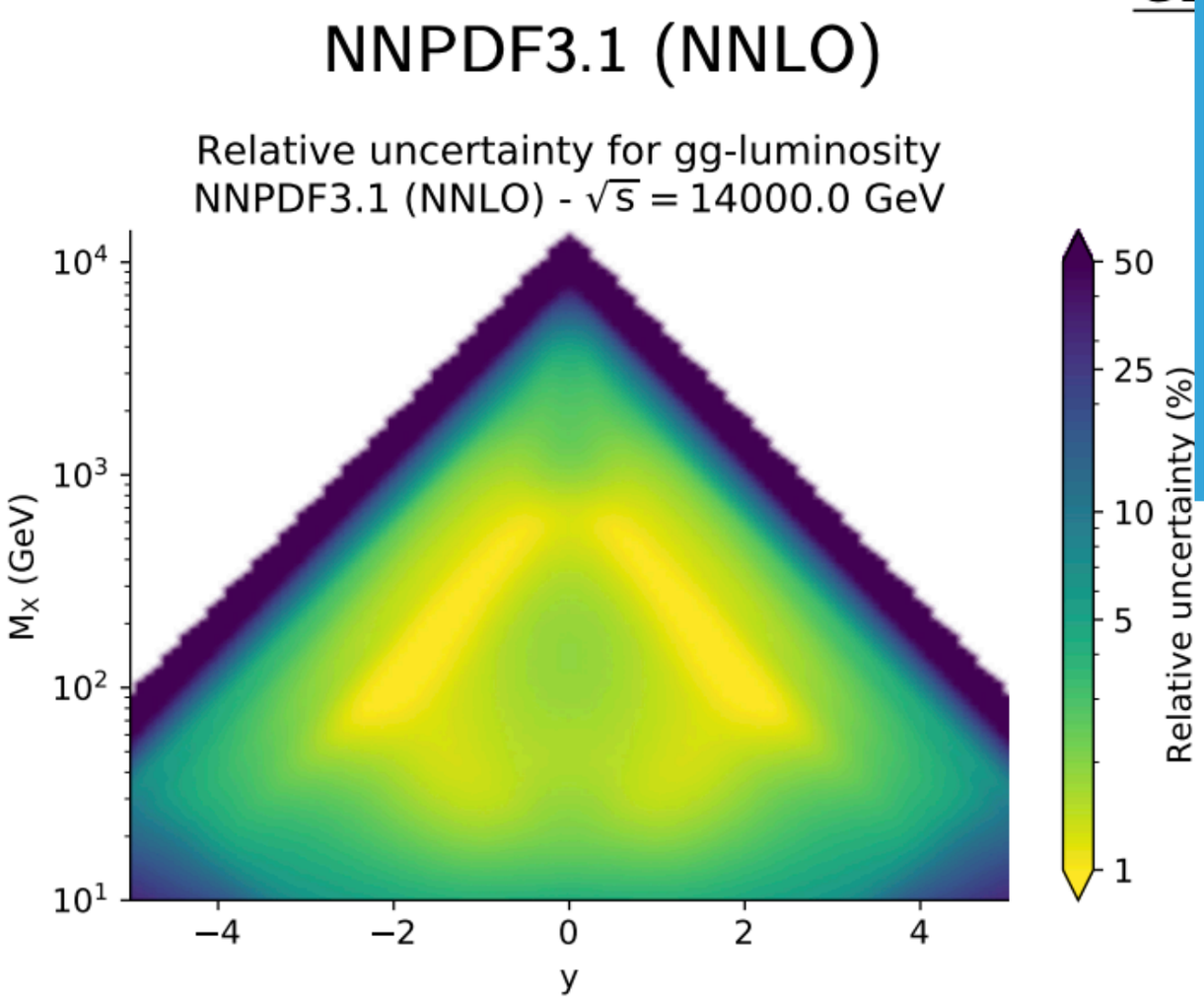
Can we trust 1% accuracy?

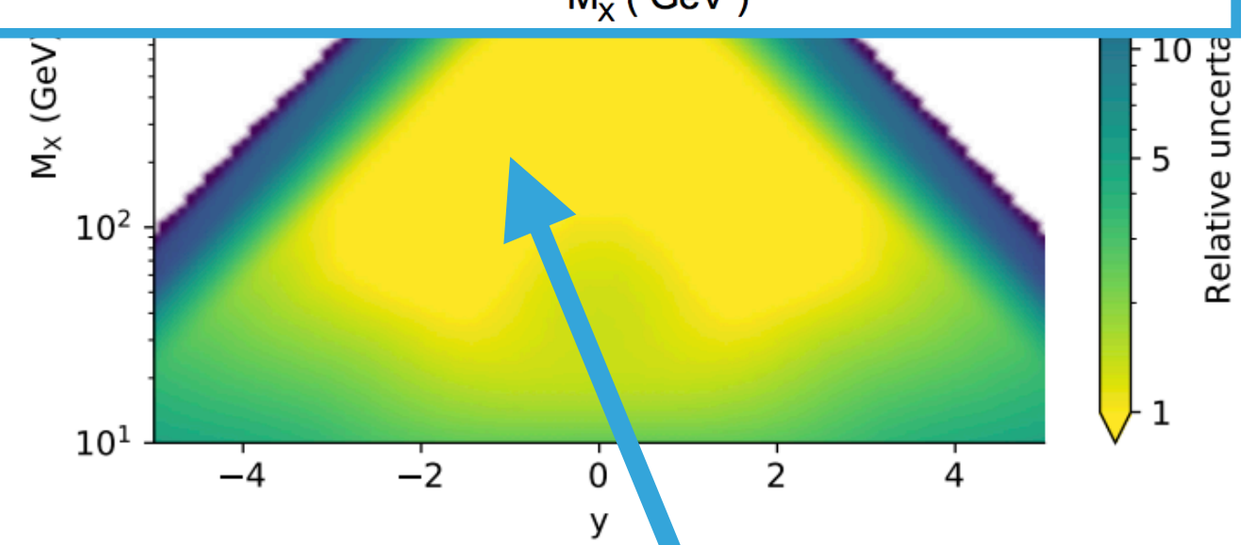
PRECISION CHALLENGE

$$\mathcal{L}_{ij}(M_X, y, \sqrt{s}) = \frac{1}{s}\sum_{i,j} f_i \left(\right.$$

GL



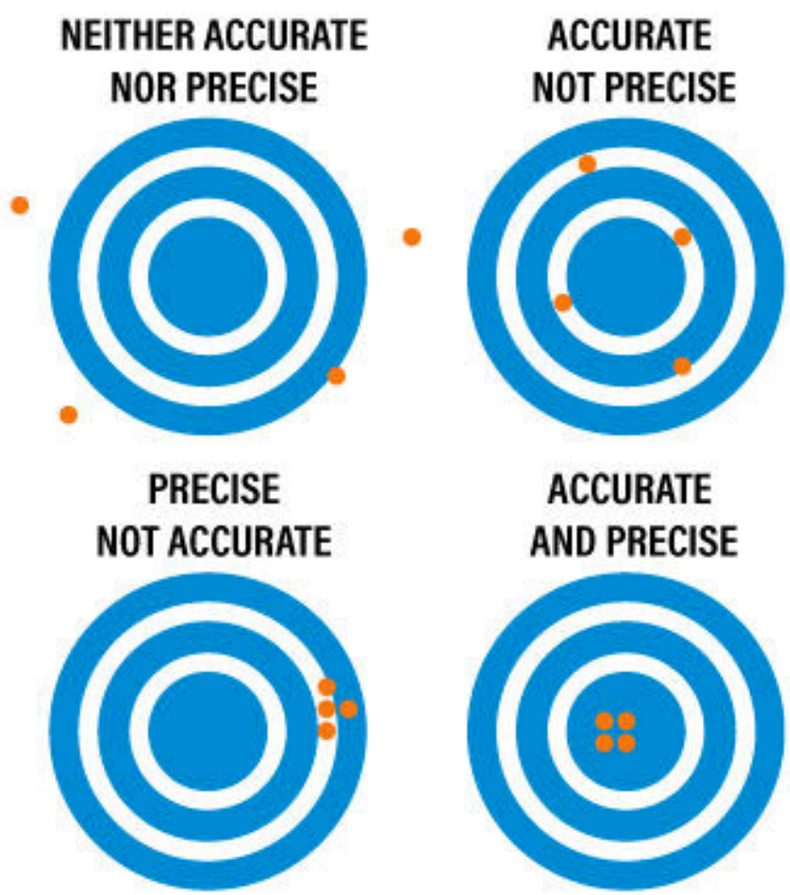


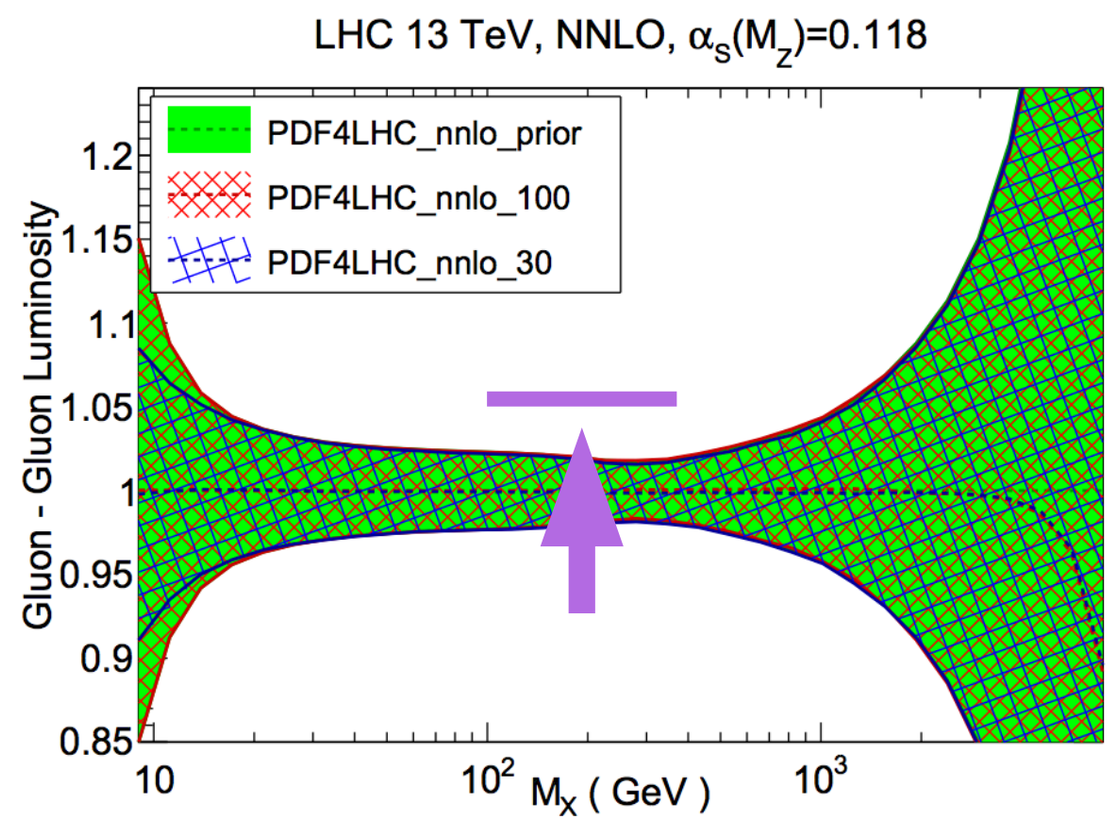


Can we trust 1% accuracy?

PRECISION VS ACCURACY

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

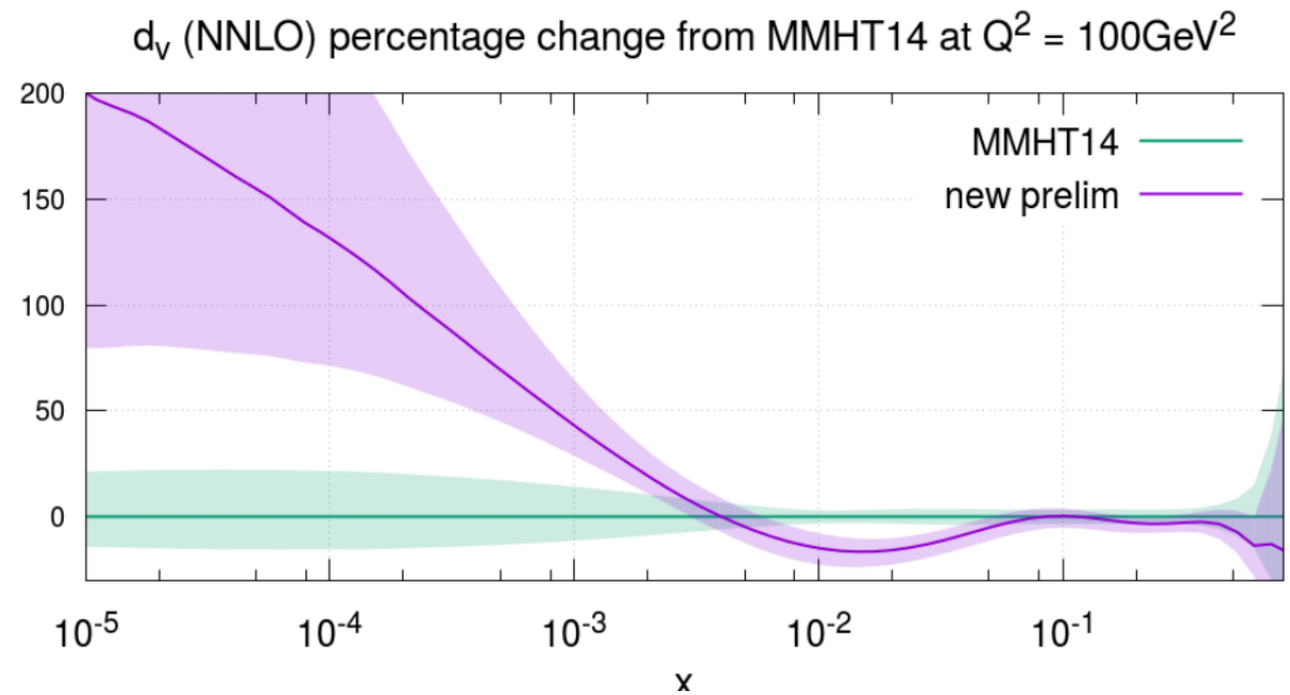




- Inconsistency or tension in data
- Data-driven parametrisation change
- Improvements in fitting methodology/minimisation?
- Theoretical framework?
  - ➡ Missing higher order uncertainties
  - ➡ Other corrections (nuclear)
  - ➡ BSM effects

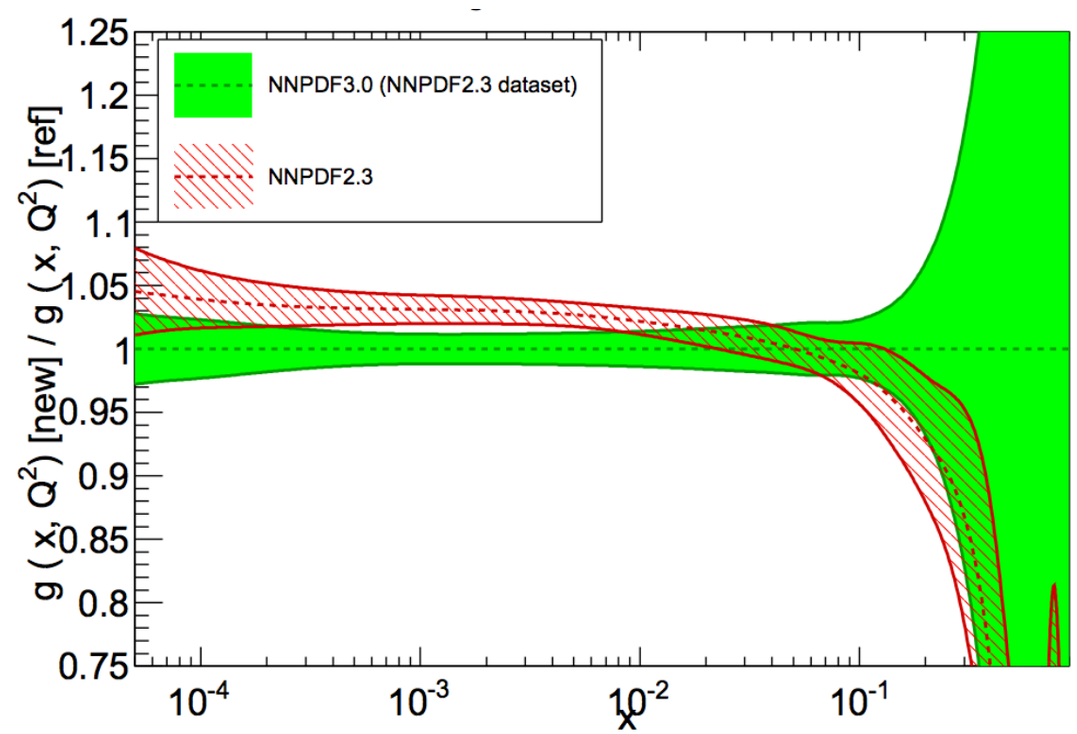
PRECISION VS ACCURACY

New MSHT20 parametrisation of valence quark based on more flexible Chebyshev parametrisation. Down quark changes reduce tension in the data

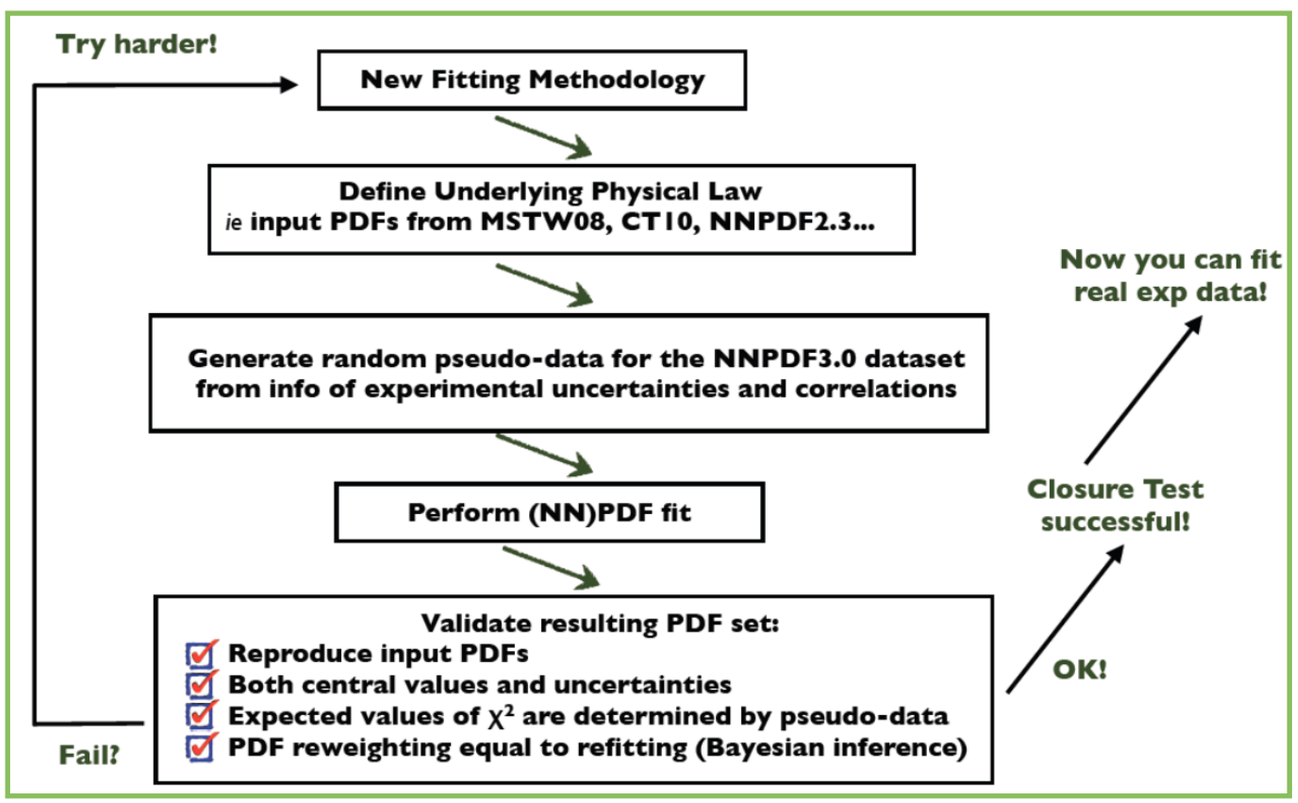


- Inconsistency or tension in data
- Data-driven **parametrisation change**
- Improvements in fitting methodology/minimisation?
- Theoretical framework?
  - ➡ Missing higher order uncertainties
  - ➡ Other corrections (nuclear)
  - ➡ BSM effects

PRECISION VS ACCURACY



NNPDF2.3 to 3.0 shift due to change in methodology!
≥ NNPDF3.0 validated by closure test



- Inconsistency or tension in data
- Data-driven parametrisation change
- Improvements in **fitting methodology/minimisation**?
- Theoretical framework?
  - ➡ Missing higher order uncertainties
  - ➡ Other corrections (nuclear)
  - ➡ BSM effects

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
- N3LO is now the precision frontier for theoretical predictions in PDF fits

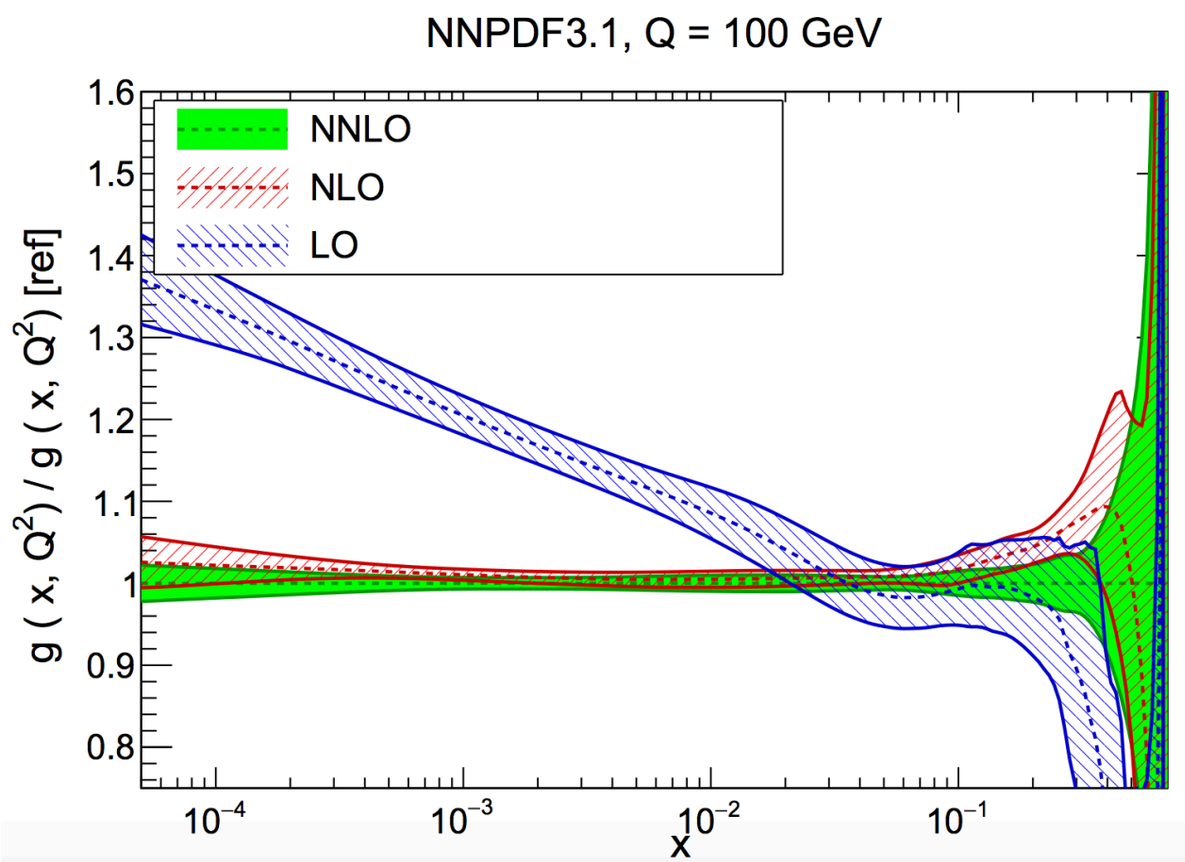
- Some N3LO splitting functions and DIS coefficient functions available but not yet implemented in evolution codes
- N3LO partonic cross sections not yet available for most pp processes (apart from Drell-Yan) but some approximations available

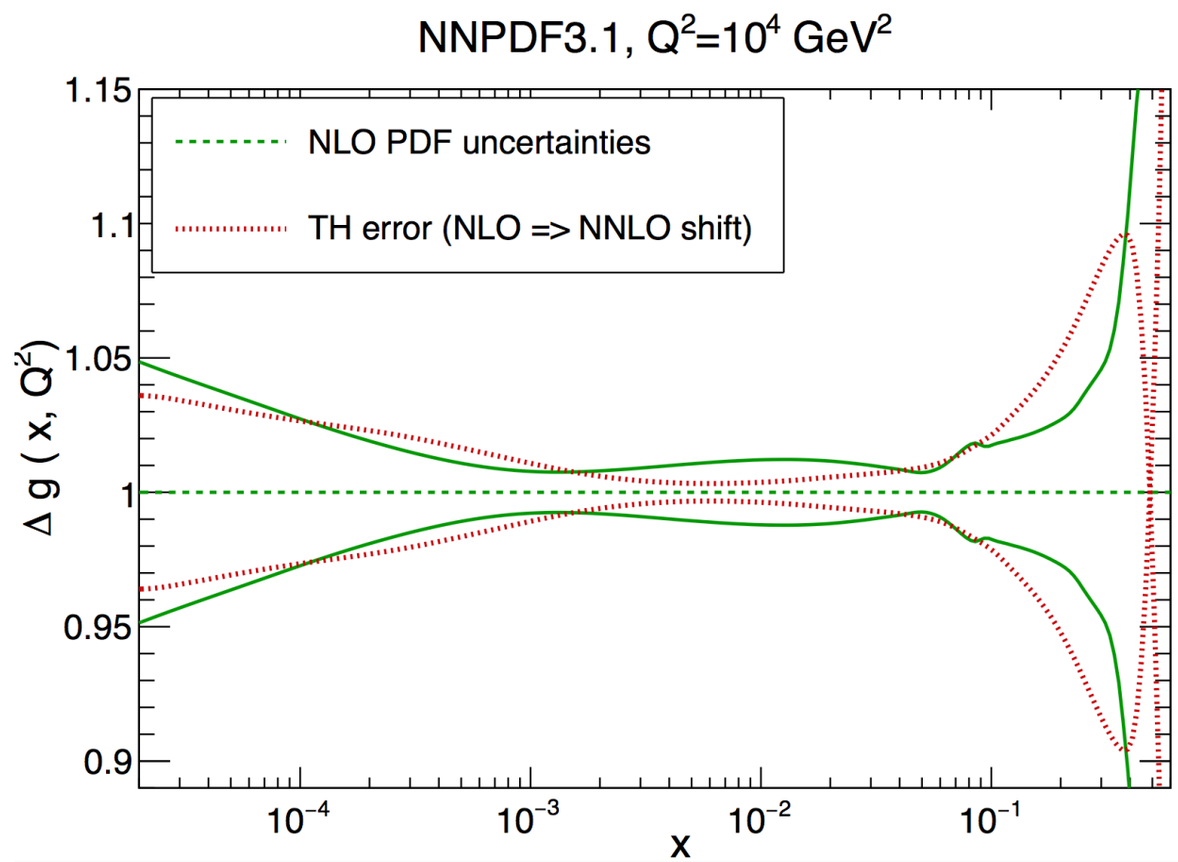
- While this is a long way, can we estimate the theoretical uncertainty in PDFs associated with missing higher order uncertainty?

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
- Shift between PDF fit at NNLO and NLO gives an idea of the size of missing higher order uncertainties in a NLO fit.



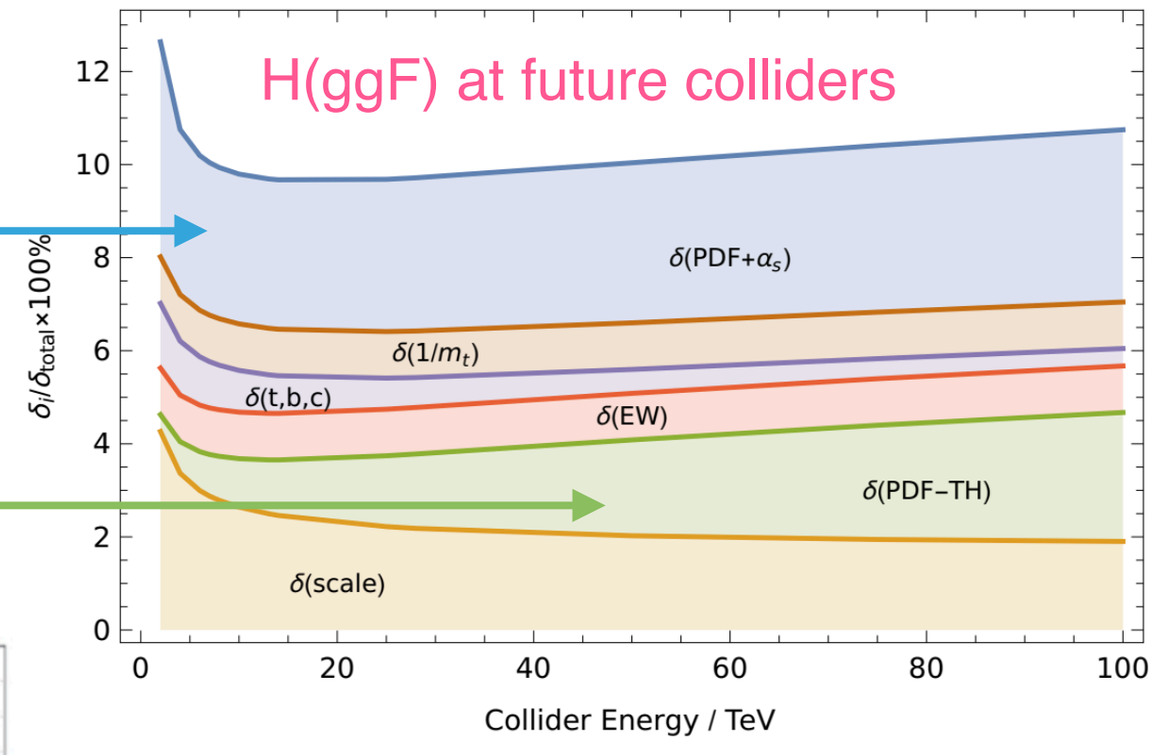


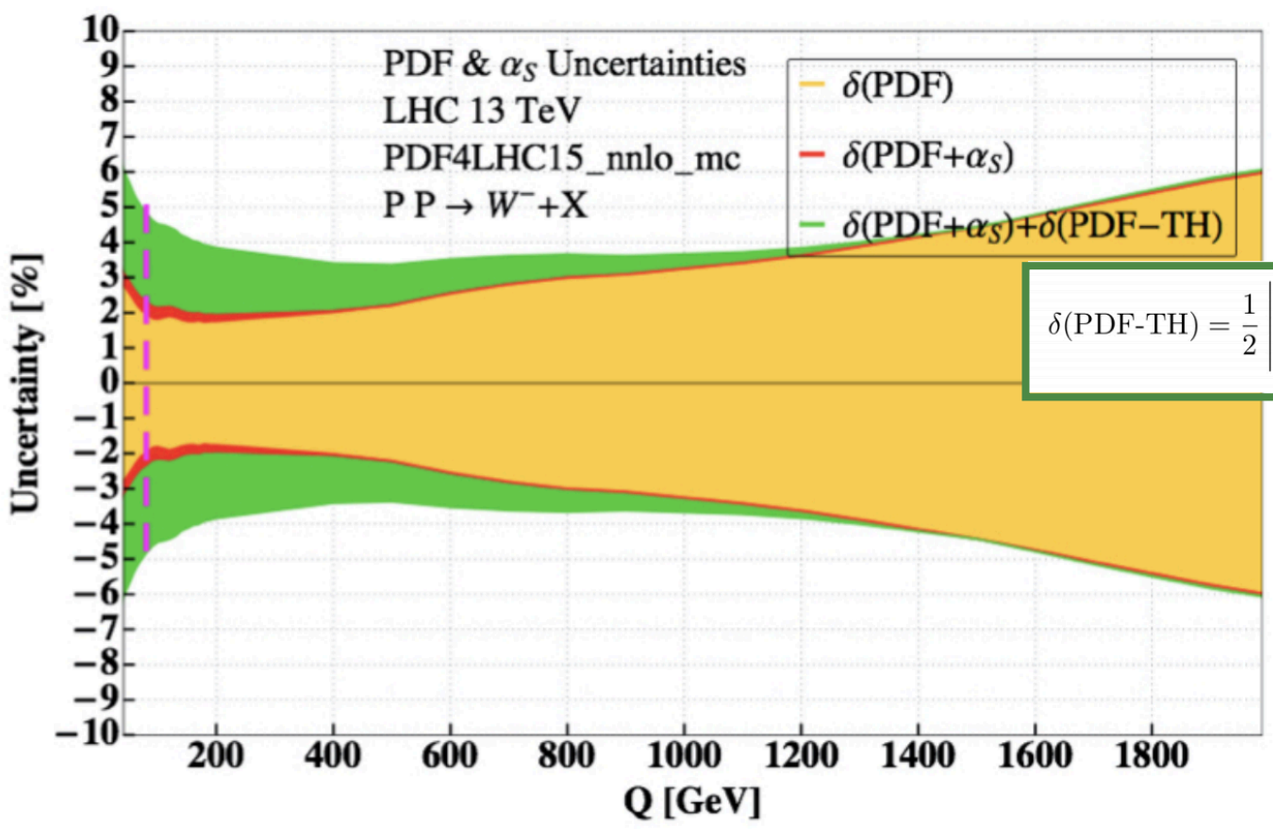
THEORY UNCERTAINTIES IN PDF FITS

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

PDF+as uncertainty

Th. mismatch (N3LO vs NNLO PDFs)





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

Th. mismatch (N3LO vs NNLO PDFs)

THEORY UNCERTAINTIES IN PDF FITS

- Construct a theory covariance matrix from scale-varied cross sections and combine it with the experimental covariance matrix

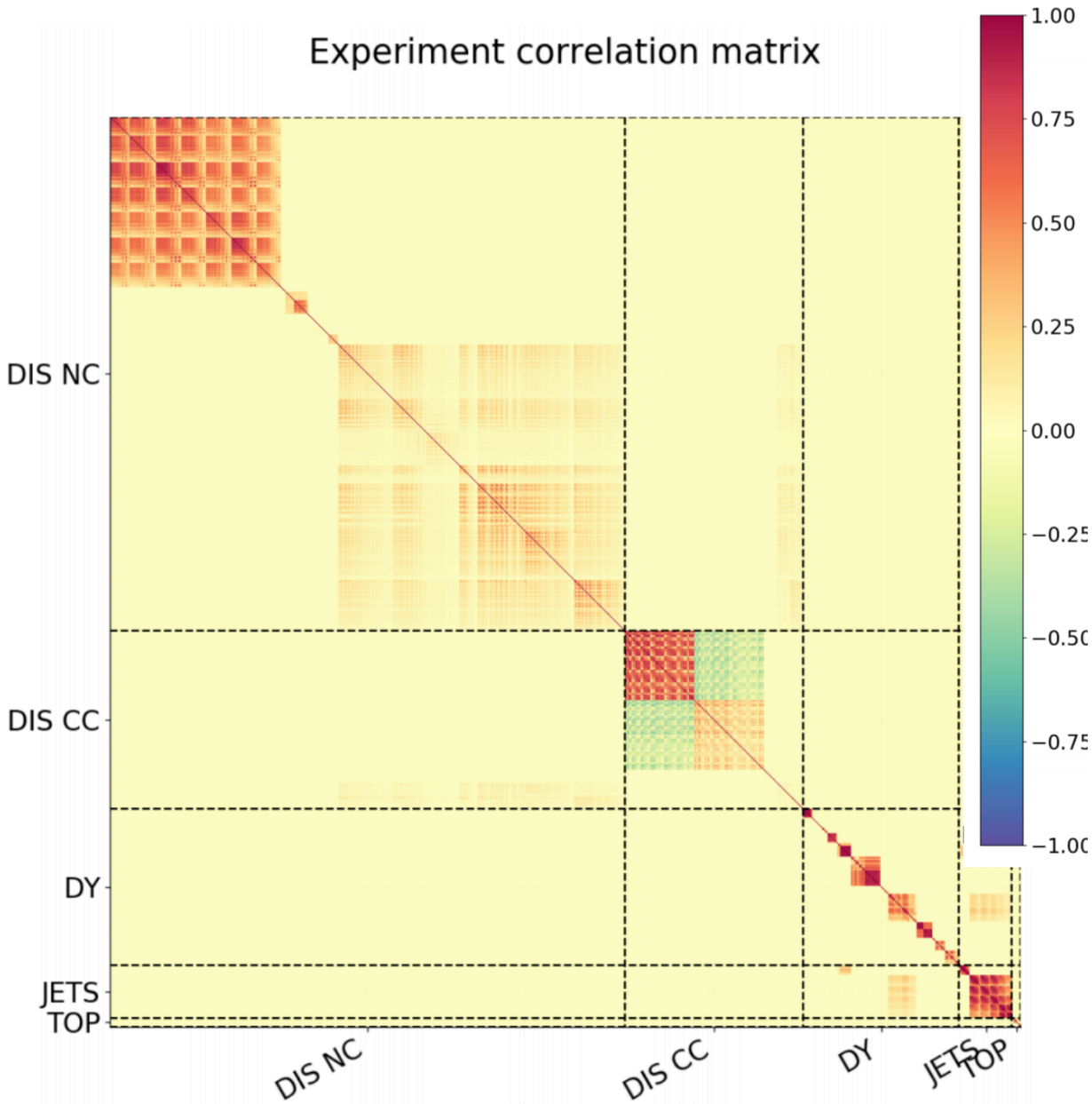
$$\chi^2 = \sum_{m,n=1}^{N} (d_m - t_m)(\mathrm{cov}_{\mathrm{exp}} + \mathrm{cov}_{\mathrm{th}})^{-1}_{mn}(d_n - t_n)$$

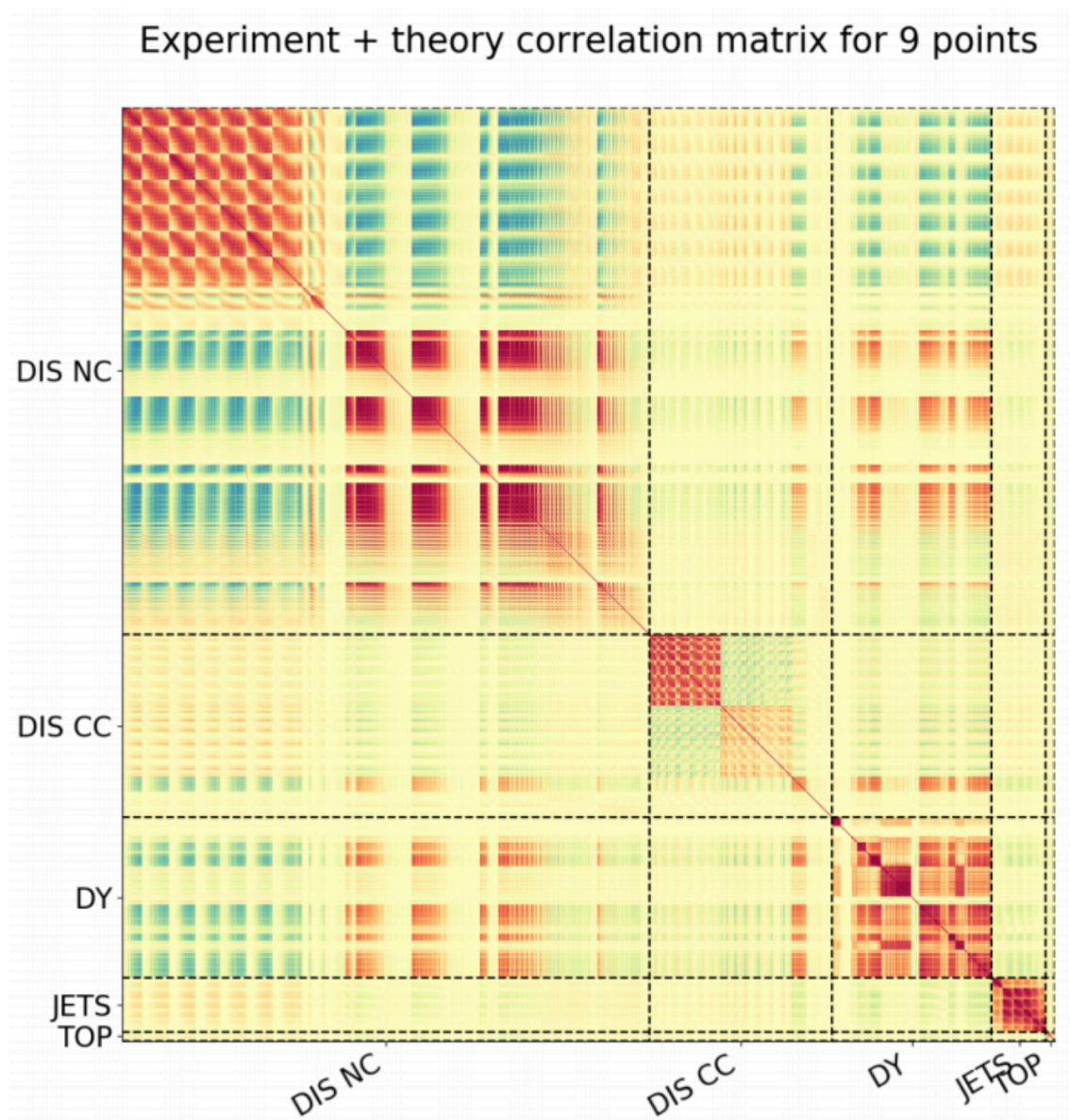
- Assumptions: experimental and theoretical errors independent and Gaussian
- Accounting for the theory covariance matrix will modify the relative weight that each of the datasets carries in the global fit: processes with higher MHOUs will be ``de-weighted''

- How to build it?
  - Specific form will depend on number of variation and assumption on correlations
  - μ_F variations correlated across all processes by PDF evolution
  - μ_R variation correlated by process (hard cross section)

The NNPDF collaboration, arXiv:1906.10698

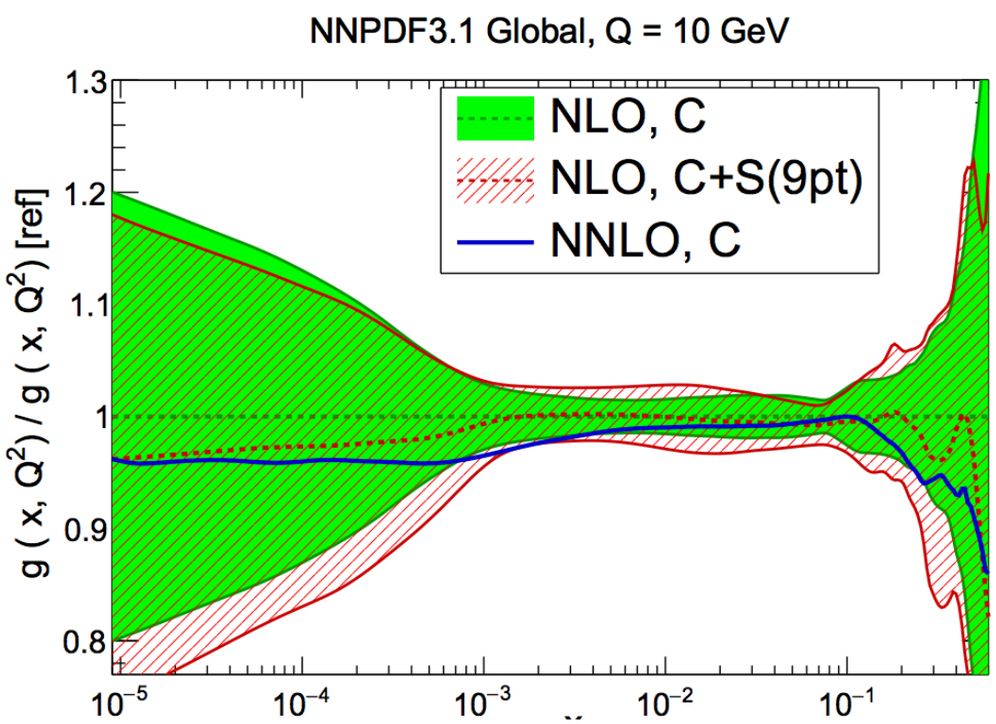
THEORY UNCERTAINTIES IN PDF FITS

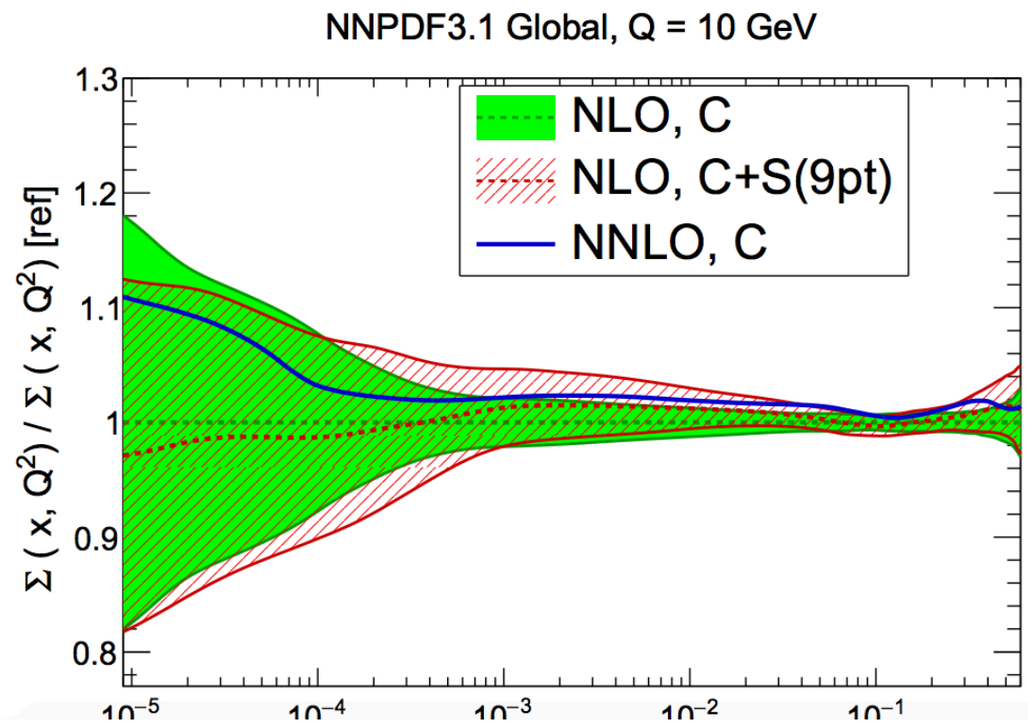


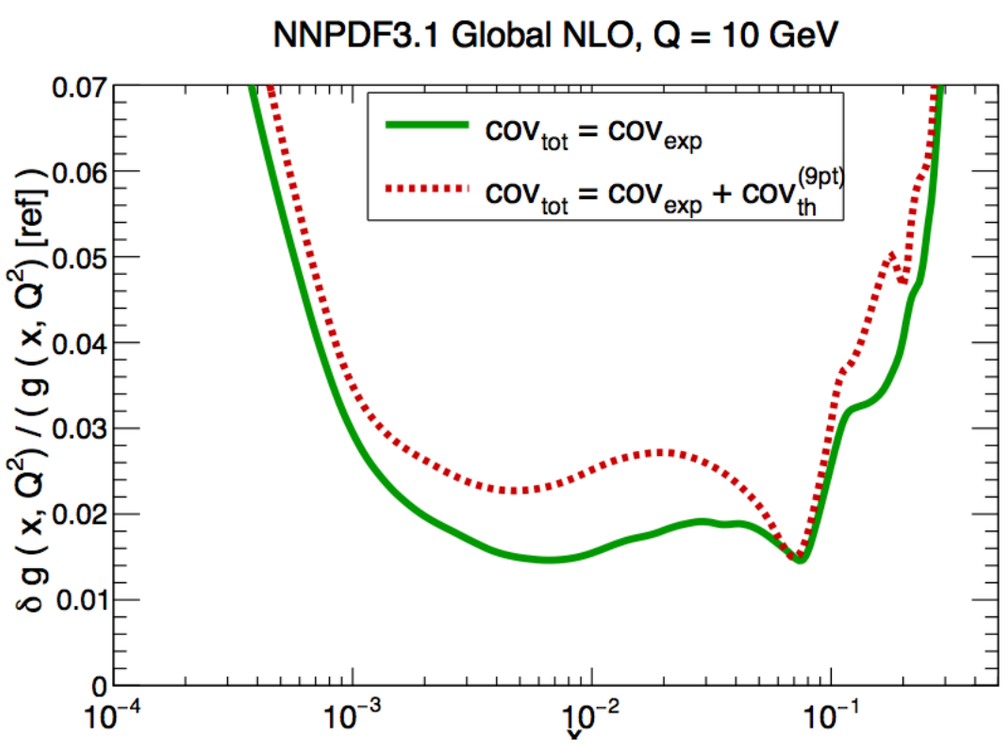


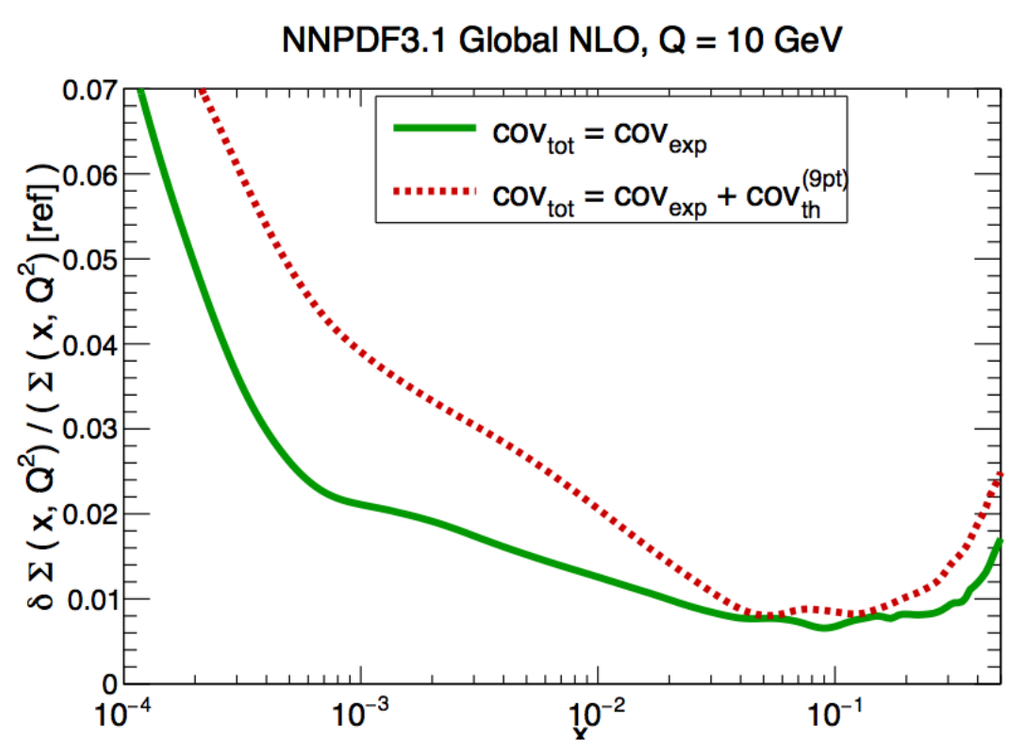
The NNPDF collaboration, arXiv:1906.10698

THEORY UNCERTAINTIES IN PDF FITS









The NNPDF collaboration, arXiv:1906.10698

(PARTIAL) CONCLUSIONS

- Precision physics opens up new fascinating challenges
- Precise and accurate understanding of the proton structure is key to make progress in comparison of standard model predictions to experimental data
- Need: robust methodology, estimate of theoretical uncertainties and robust statistics to deal with data inconsistency
- Aspects that could be previously neglected are now becoming extremely relevant
  - ➡ How to deal with MHOU (or any other uncertainties like nuclear corrections, errors on parameters used in fits) of theoretical predictions entering PDF fits?
  - ➡ Whenever parameters are determined from LHC data (for example Wilson coefficients of EFT expansion, or EW parameters or αS) simultaneous fits of PDFs and parameters is way to shed light on their interplay