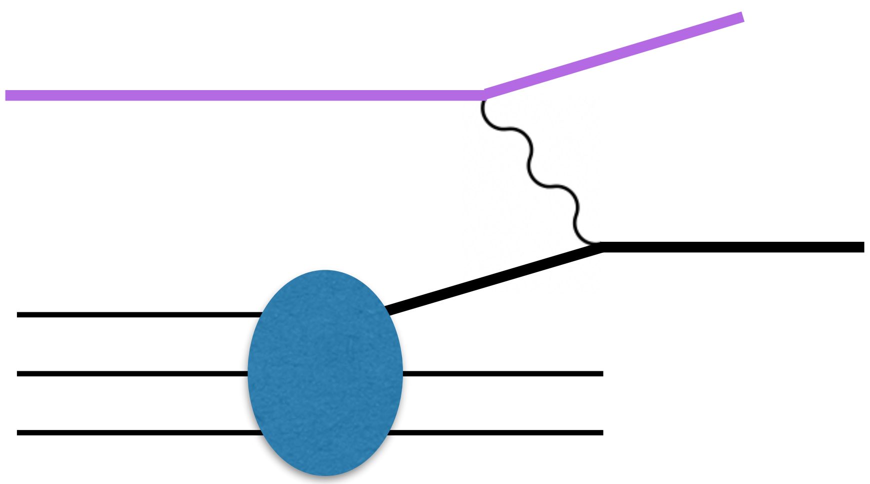
GLOBAL FITS OF PARTON DISTRIBUTION FUNCTIONS (II+III)

Maria Ubiali, University of Cambridge

18th September 2025

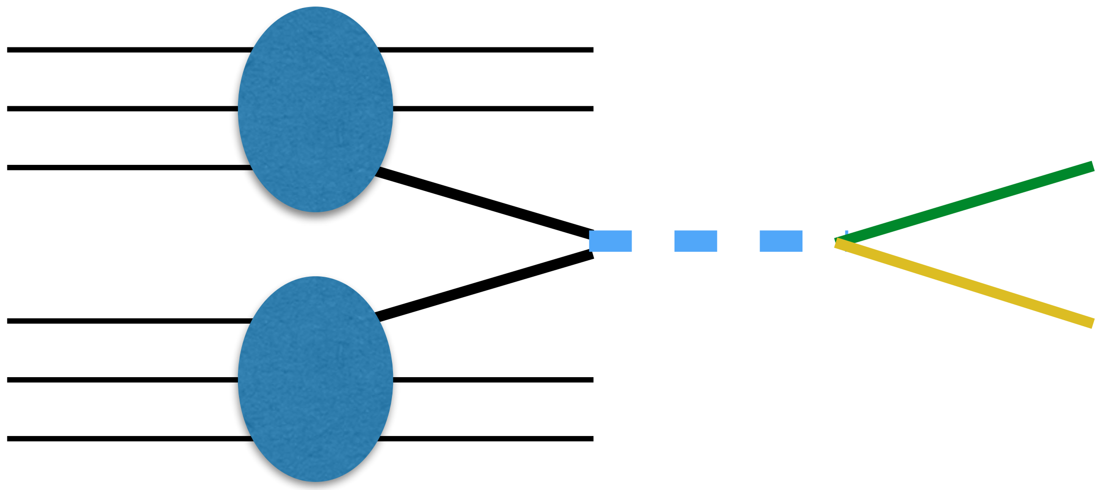
MPA summer school 2025

Highlights from yesterday



- Collinear factorisation picture
- Universality of PDFs
- DGLAP evolution of PDFs predicted by perturbative QCD

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



- All terms in master equation have an uncertainty associated that is a component of the uncertainty of theory predictions

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

Outline

- First two lectures (yesterday)
    - Motivation:
      the high energy big picture
    - Parton Model and QCD
    - Collinear Factorisation
- **Second and third lectures (today)**
    - Ingredients of a PDF global fits
    - Experimental input
    - Methodological aspects
    - Theoretical aspects
    - Frontiers and challenges

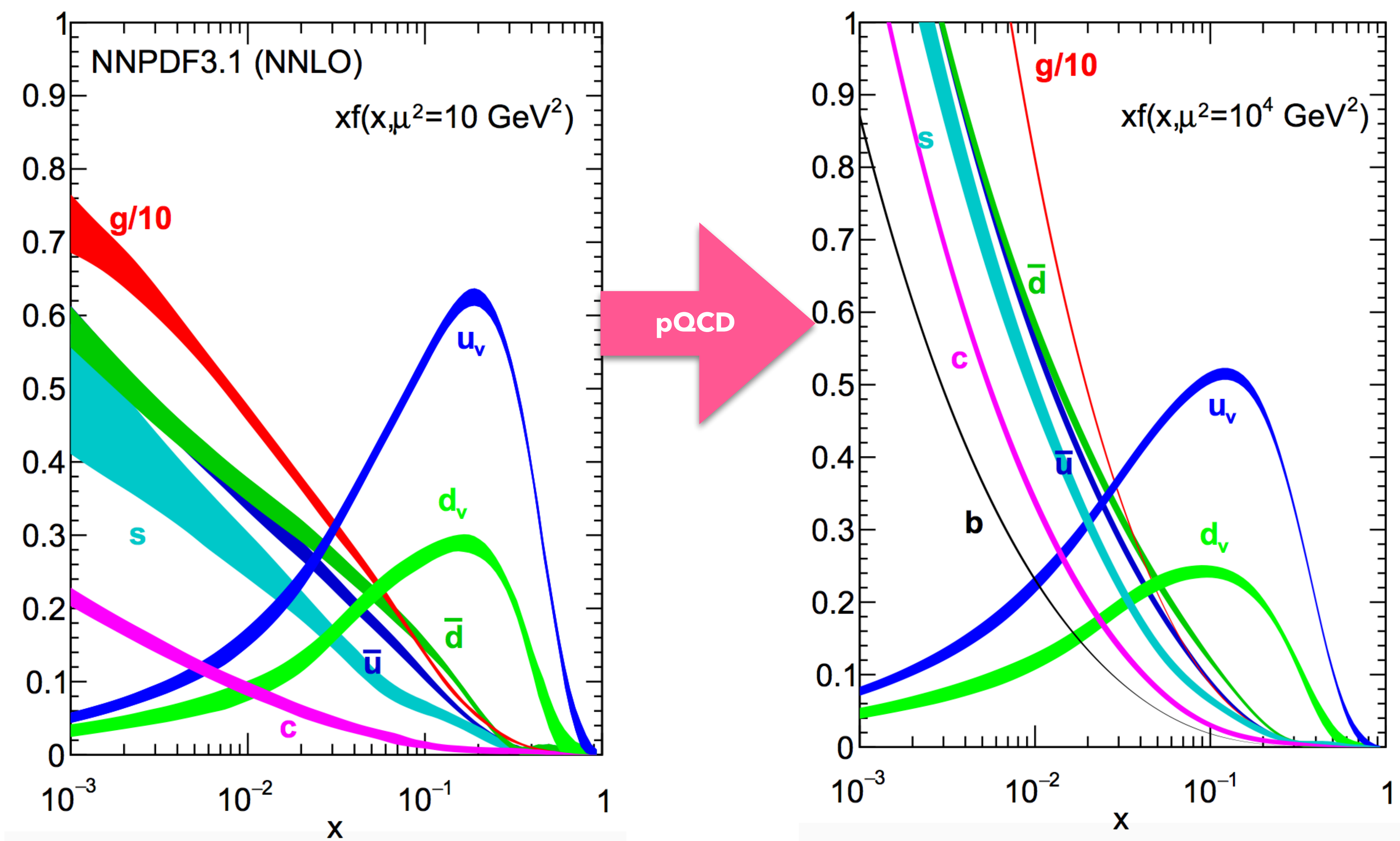
PDF determination

$$f_i(x, \mu)$$

Data

Perturbative QCD

Hadronic scale: global fit of PDFs



High scale: input to the LHC

How to determine initial conditions: PDF global fits

The ingredients

- 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 ingredients

- 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 ingredients

- 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 ingredients

- 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

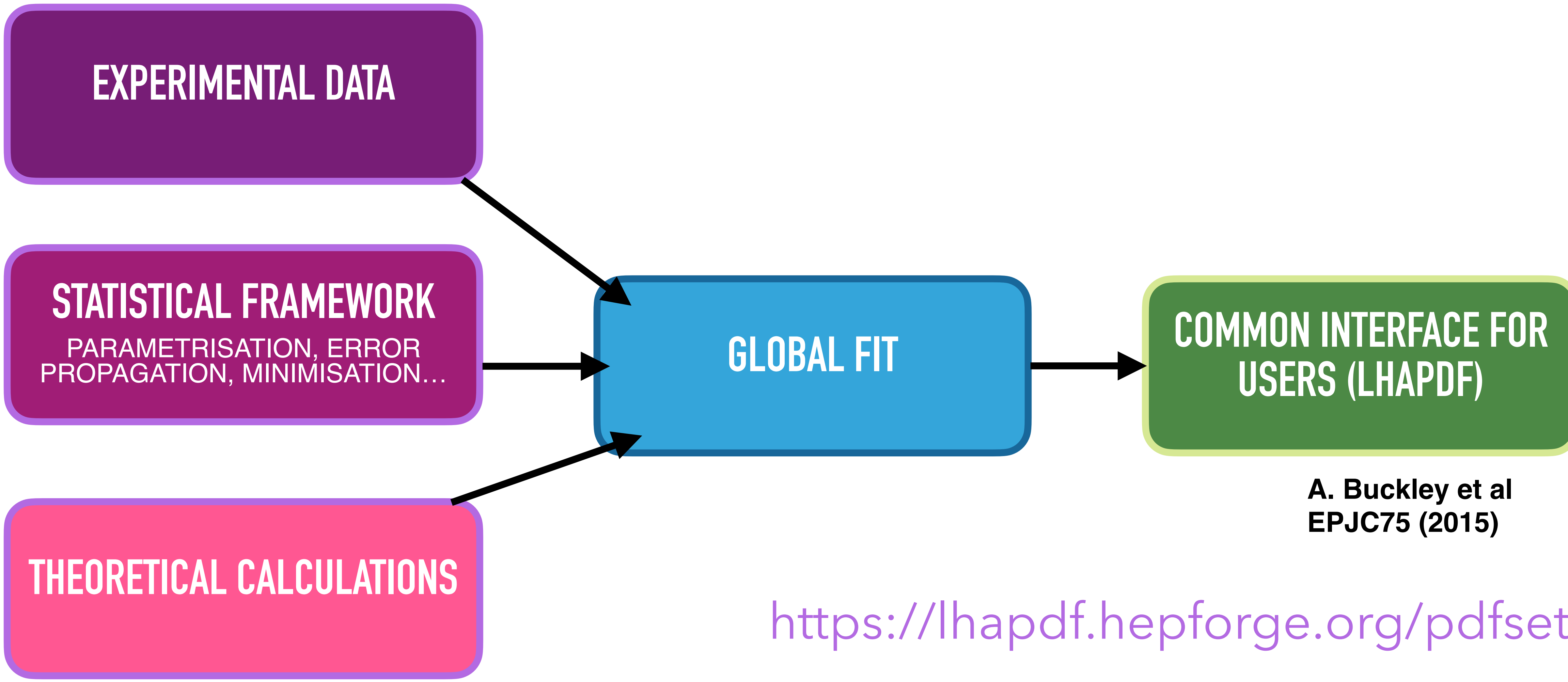
The ingredients

- 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 ingredients

- 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 PDF **error sets** to compute PDF uncertainties

A complex machinery



A. Buckley et al
EPJC75 (2015)

https://lhapdf.hepforge.org/pdfsets

1. Experimental input

Experimental data

- PDFs are not measurable, we measure observables that convolute PDFs with partonic cross sections

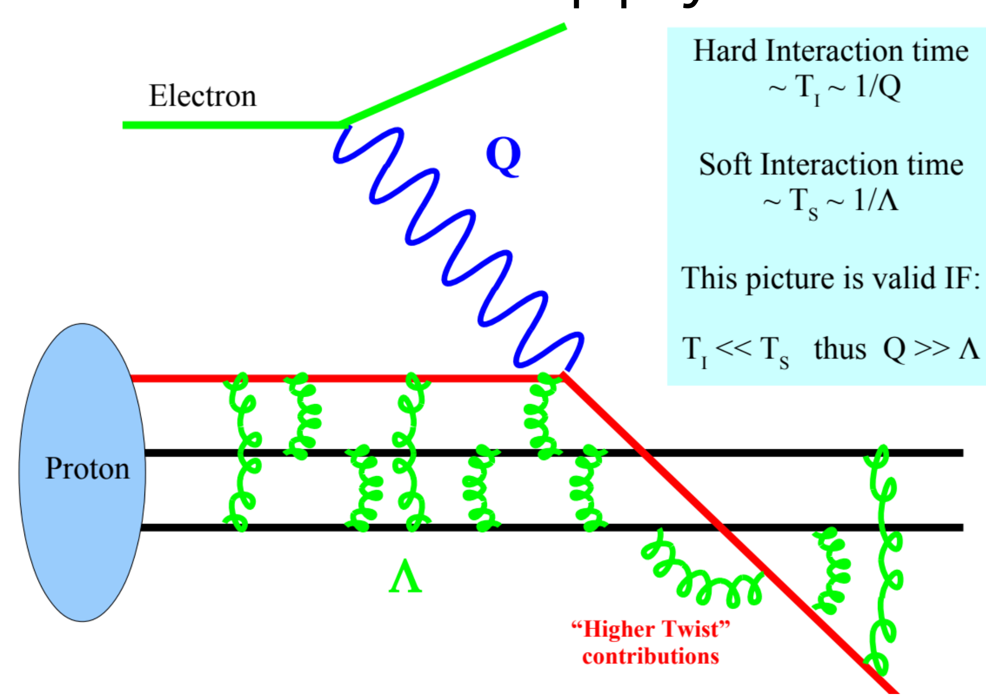
$$\frac{d\sigma_H^{ep\to ab}}{dX} = \sum_{i=-n_f}^{+n_f} \int_{x_B}^{1} \frac{dz}{z}\, f_i(z,\mu_F)\, \frac{d\hat{\sigma}_i^{ei}}{dX}(zS, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda^n}{S^n}\right)$$

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

- In global fits exclude regions where factorisation fails to apply (low Q^2 and large x). Typically

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

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



Experimental data

- PDFs are not measurable, we measure observables that convolute PDFs with partonic cross sections

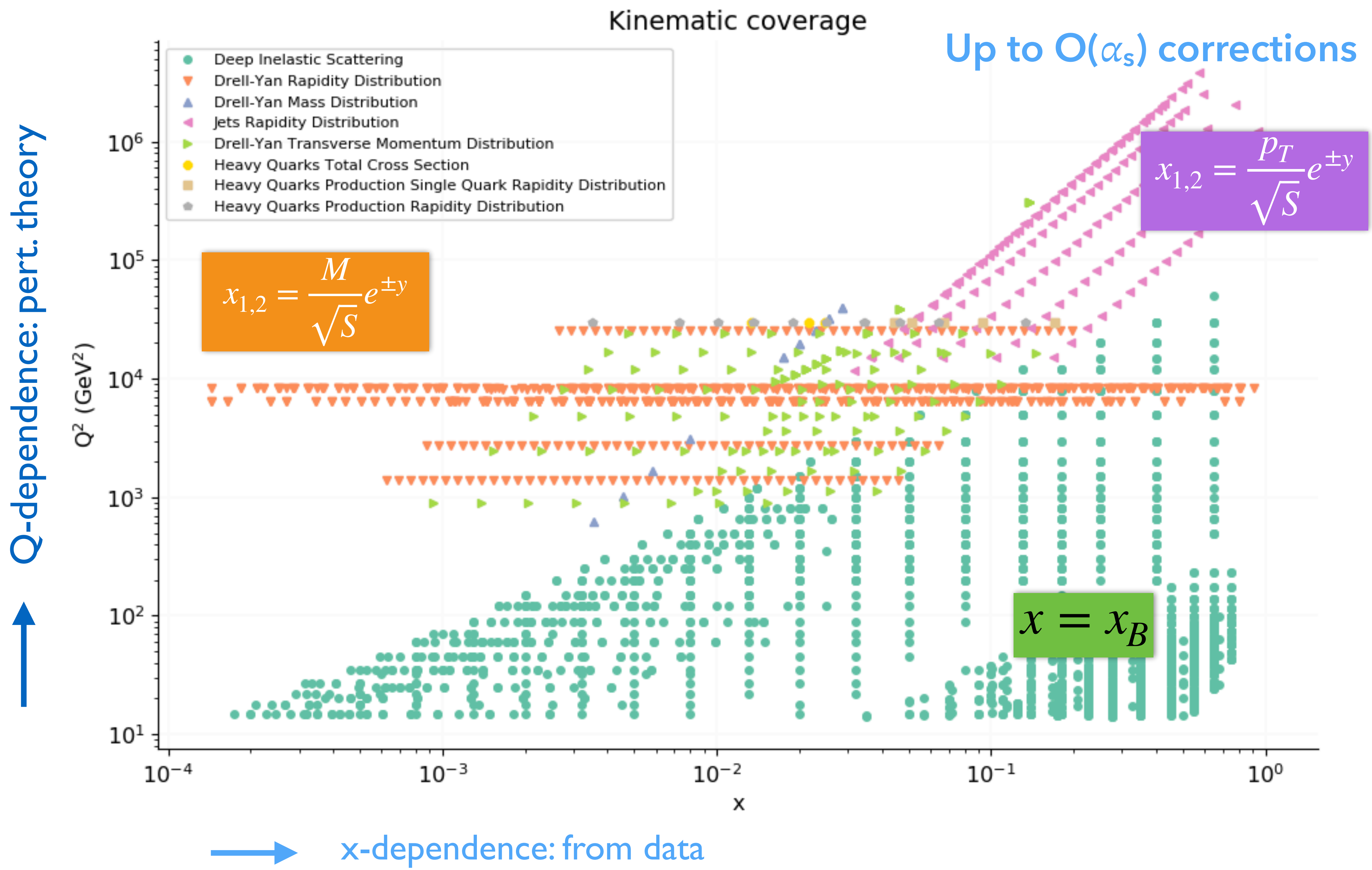
$$\frac{d\sigma_H^{ep\to ab}}{dX} = \sum_{i=-n_f}^{+n_f} \int_{x_B}^{1} \frac{dz}{z}\, f_i(z,\mu_F)\, \frac{d\hat{\sigma}_i^{ei}}{dX}(zS, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda^n}{S^n}\right)$$

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

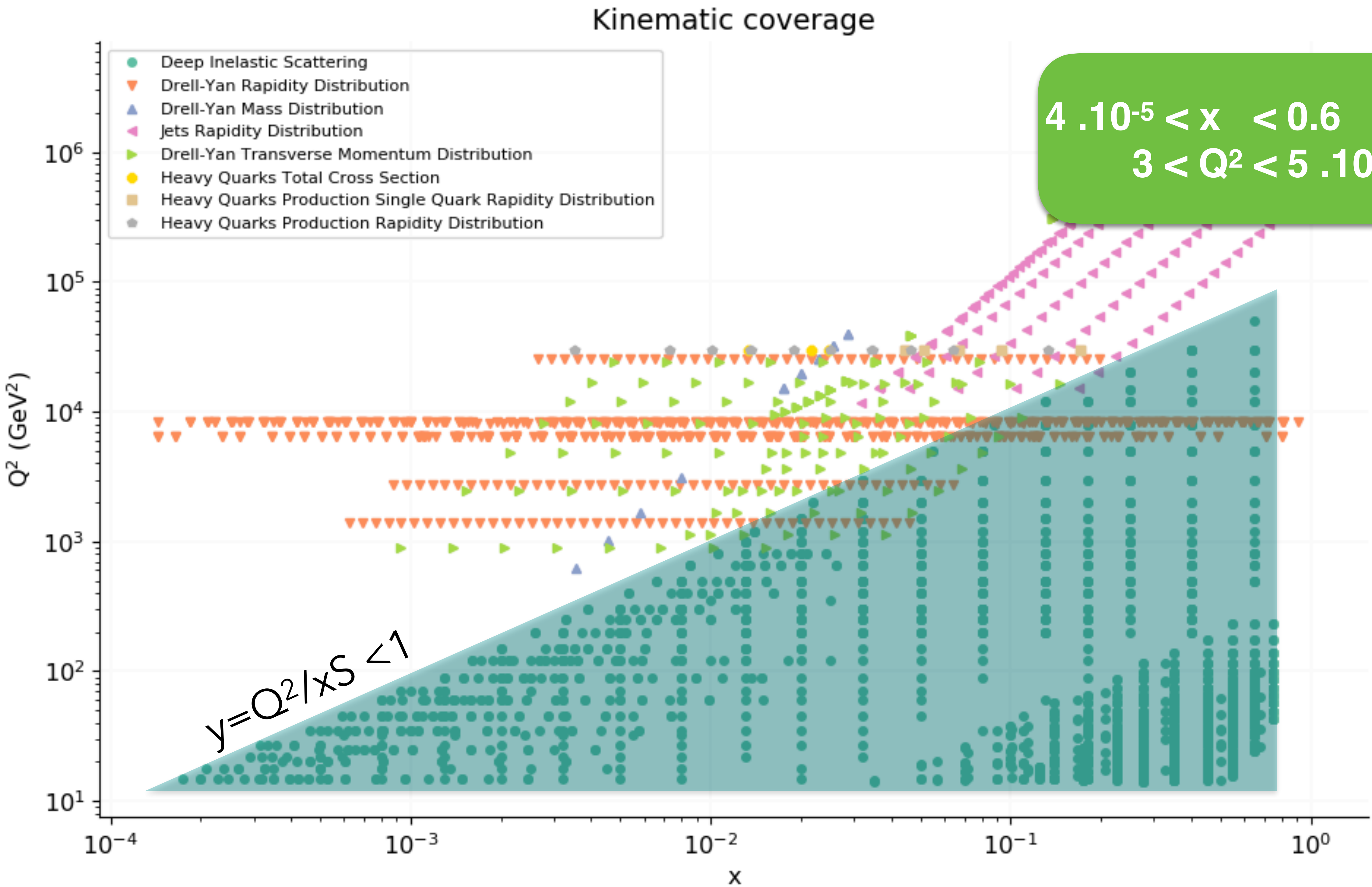
- Different data constrain different PDF combinations in different regions

  ➡ DIS data on proton abundant and precise (HERA)
  ➡ In principle F_2, F_3 CC provide 4 light quark combinations
  F_2, F_3 NC provide 2 extra light quark combinations
  ➡ HERA data only determine four combinations of PDFs
  ➡ Old DIS and Drell-Yan data still used because of isospin symmetry
  ➡ W,Z boson final state provide lot of information, gluon from scale dependence
  ➡ Processes with jets and/or heavy quark in final states direct handle on the gluon

Disentangling PDFs



HERA data

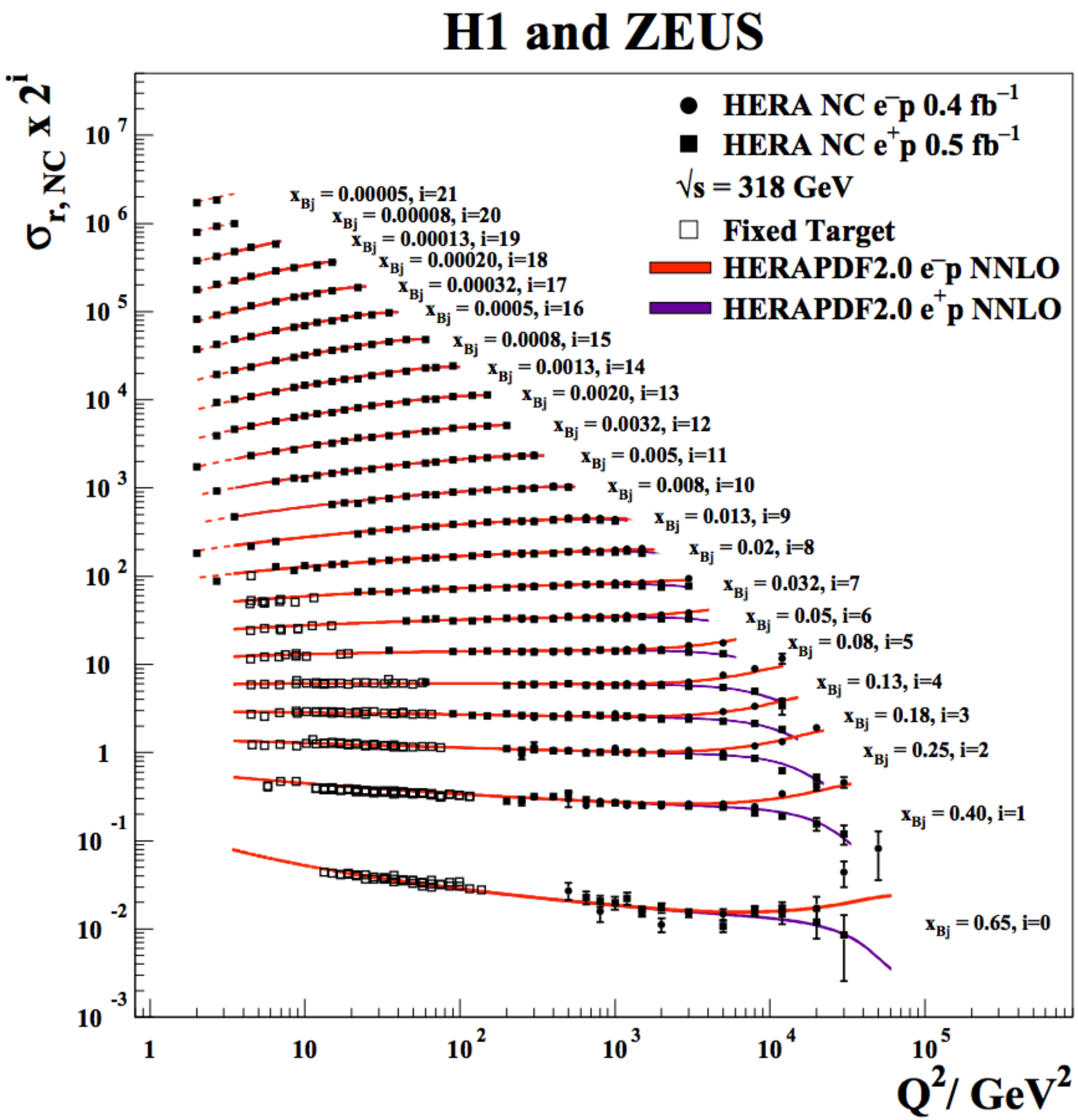


$4 \cdot 10^{-5} < x < 0.6$

$3 < Q^2 < 5 \cdot 10^4$ GeV^2

HERA data

- Combination of Run I + Run II data led to very precise measurements of reduced xsec
- F3 contribution for NC (parity violating) associated with Z exchange visible at larger x and Q~Mz



arXiv: 1506.06042

HERA data

Neutral Current

$$\left[F_2^\gamma, F_2^{\gamma Z}, F_2^Z\right] = x \sum_{i=1}^{n_f} \left[e_i^2, 2e_i g_V^i, (g_V^i)^2 + (g_A^i)^2\right](q_i + \bar{q}_i)$$

③ ④

$$\left[F_3^\gamma, F_3^{\gamma Z}, F_3^Z\right] = x \sum_{i=1}^{n_f} \left[0, 2e_i g_A^i, 2g_V^i g_A^i\right](q_i - \bar{q}_i)$$

Charged Current

①
$$F_2^{W^-} = 2x\left(u + \bar{d} + \bar{s} + c\right),$$
$$F_3^{W^-} = 2x\left(u - \bar{d} - \bar{s} + c\right),$$

②
$$F_2^{W^+} = 2x\left(d + \bar{u} + \bar{c} + s\right),$$
$$F_3^{W^+} = 2x\left(d - \bar{u} - \bar{c} + s\right),$$

$$\frac{d^2\sigma}{dx\,dQ^2} \propto Y_+ F_2(x, Q^2) \mp Y_- x F_3(x, Q^2) - y^2 F_L(x, Q^2)$$

Longitudinal Structure function

$$F_L(x, Q^2) = \frac{\alpha_s(Q^2)}{\pi}\left[\frac{4}{3}\int_x^1 \frac{dy}{y}\left(\frac{x}{y}\right)^2 F_2(y, Q^2) + 2\sum_i e_i^2 \int_x^1 \frac{dy}{y}\left(\frac{x}{y}\right)^2\left(1 - \frac{x}{y}\right) g(y, Q^2)\right]$$

Altarelli, Martiinelli Phys. Lett. 76B (1978)

HERA data

Neutral Current

$$\left[F_2^{\gamma}, F_2^{\gamma Z}, F_2^{Z}\right] = x \sum_{i=1}^{n_f} \left[e_i^2, 2e_i g_V^i, (g_V^i)^2 + (g_A^i)^2\right](q_i + \bar{q}_i)$$

③ ④

$$\left[F_3^{\gamma}, F_3^{\gamma Z}, F_3^{Z}\right] = x \sum_{i=1}^{n_f} \left[0, 2e_i g_A^i, 2g_V^i g_A^i\right](q_i - \bar{q}_i)$$

Charged Current

①

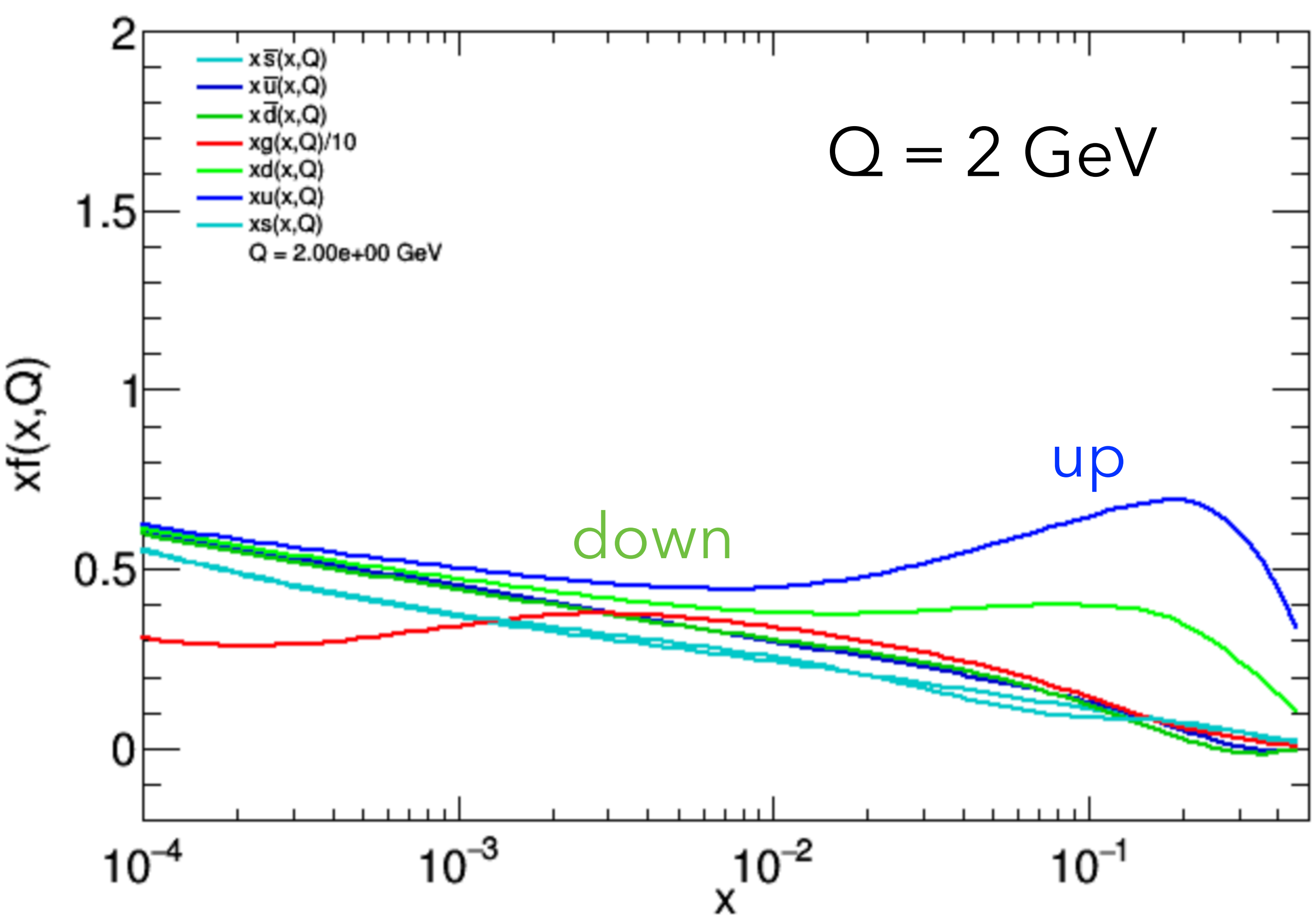
$$F_2^{W^-} = 2x\left(u + \bar{d} + \bar{s} + c\right),$$

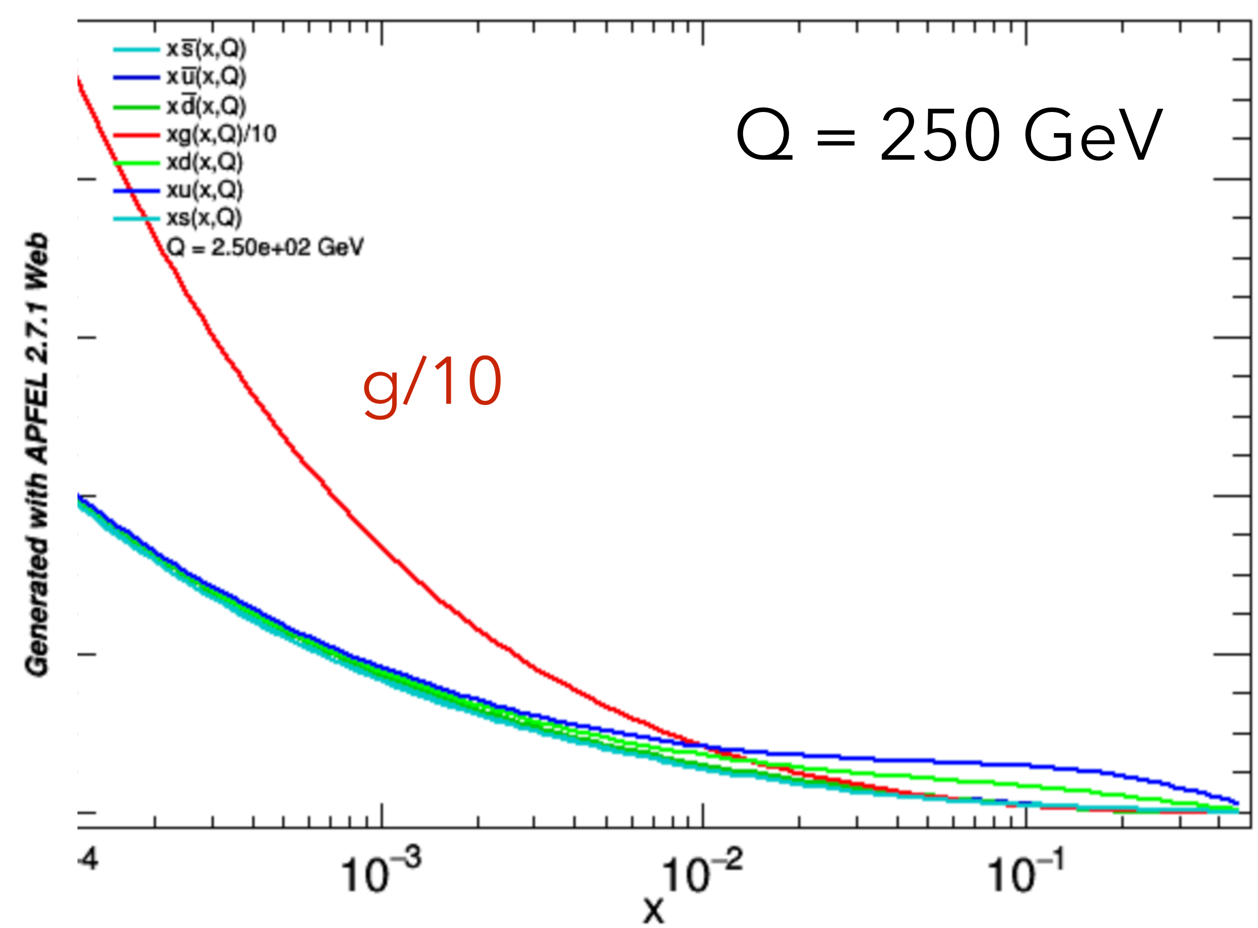
$$F_3^{W^-} = 2x\left(u - \bar{d} - \bar{s} + c\right),$$

②

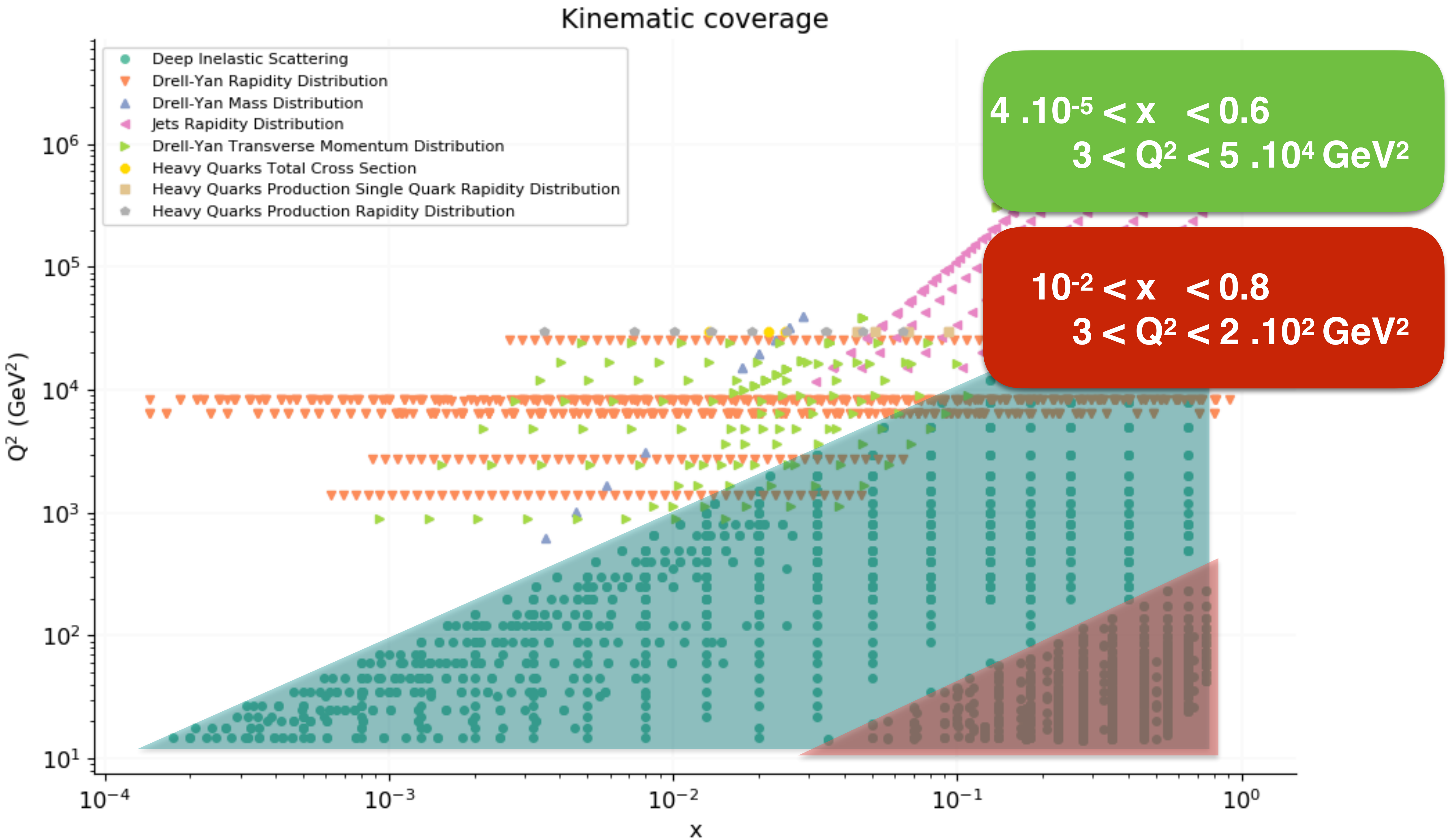
$$F_2^{W^+} = 2x\left(d + \bar{u} + \bar{c} + s\right),$$

$$F_3^{W^+} = 2x\left(d - \bar{u} - \bar{c} + s\right),$$





Fixed target DIS data



Fixed target DIS data

- Experimentally measured deuteron structure function

$$F_2^d = \frac{F_2^p + F_2^n}{2}$$

- Assumption SU(2) isospin symmetry: neutron is just like proton with u ⇔ d

proton = uud
neutron = ddu

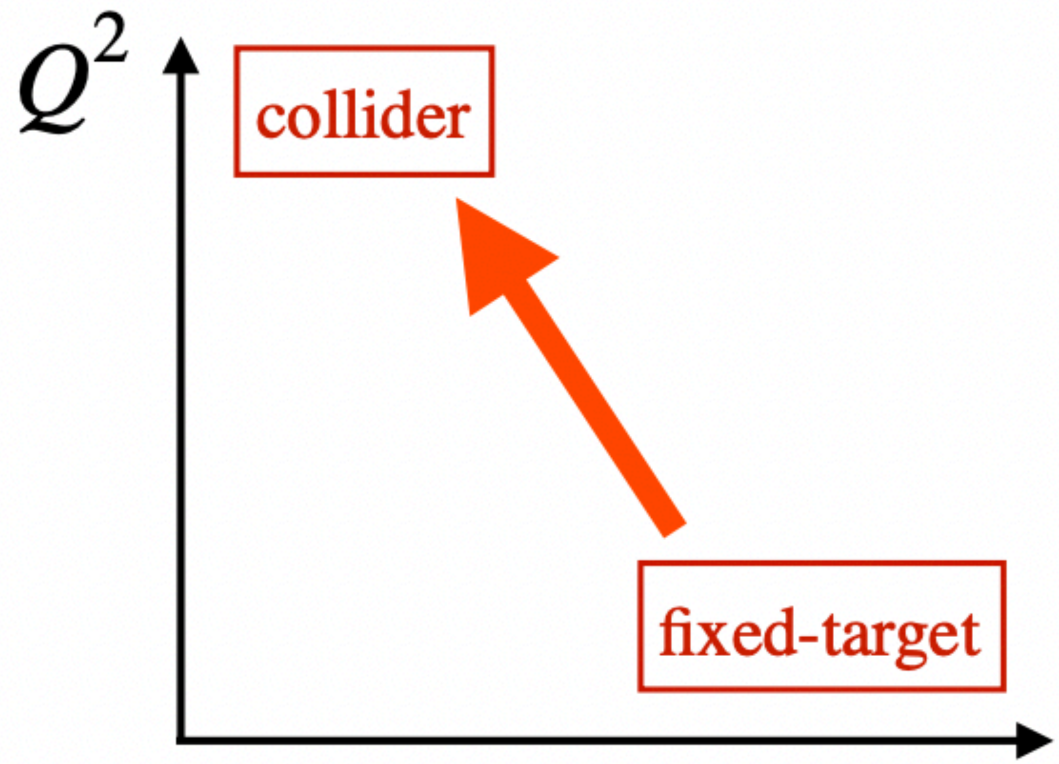
⇒ **$u_n(x)=d_p(x)$** and **$d_n(x) = u_p(x)$**

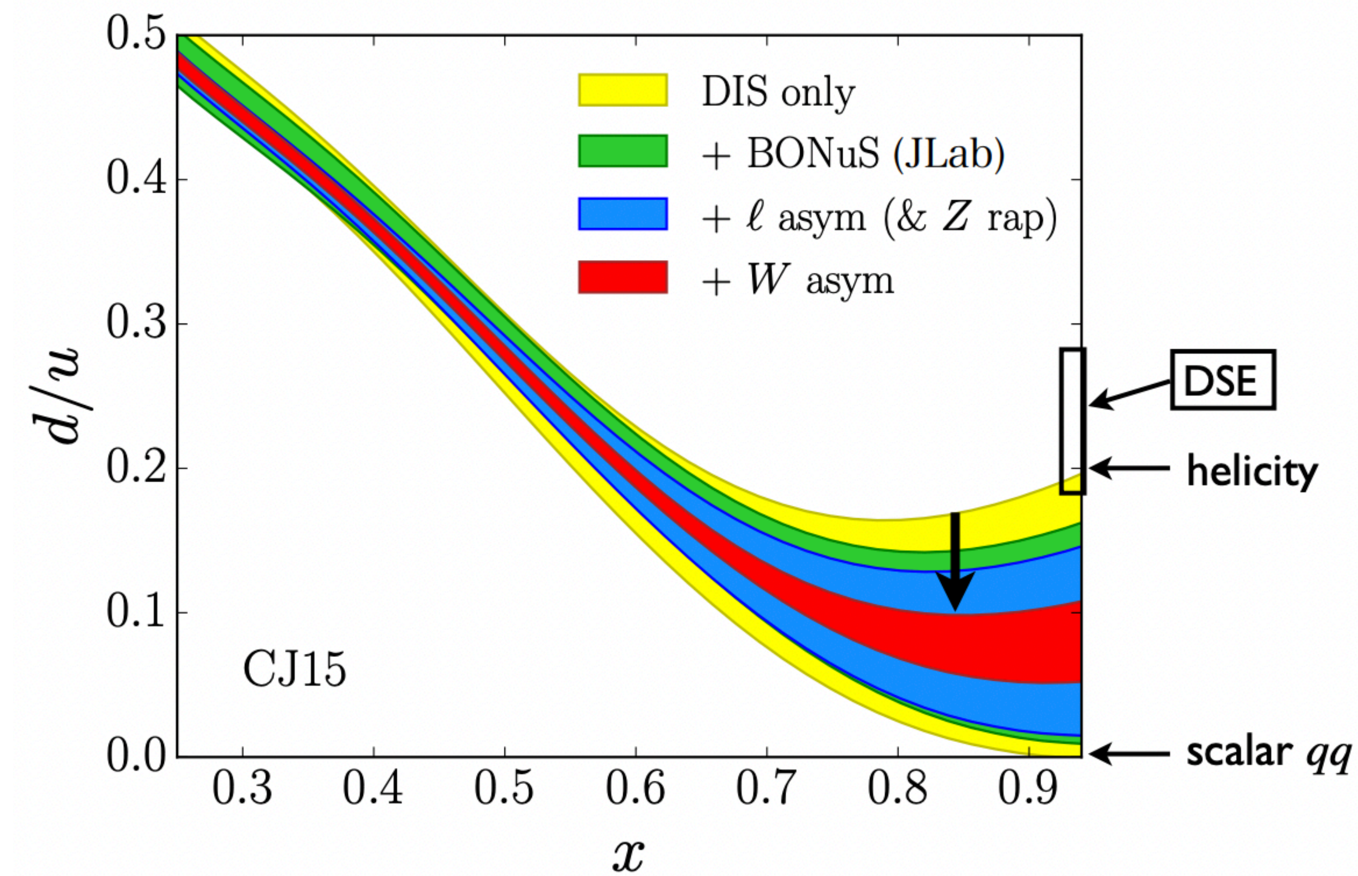
- Linear combinations of F_2^p and F_2^n give separately $u_p(x) \equiv u(x)$ and $d_p(x) \equiv d(x)$,

$$F_2^p(x,Q^2) - F_2^d(x,Q^2) = \frac{1}{3}(u + \bar{u} - d - \bar{d})$$

$$\frac{F_2^d(x)}{F_2^p(x)} \sim \frac{u}{d}$$

Fixed target DIS data





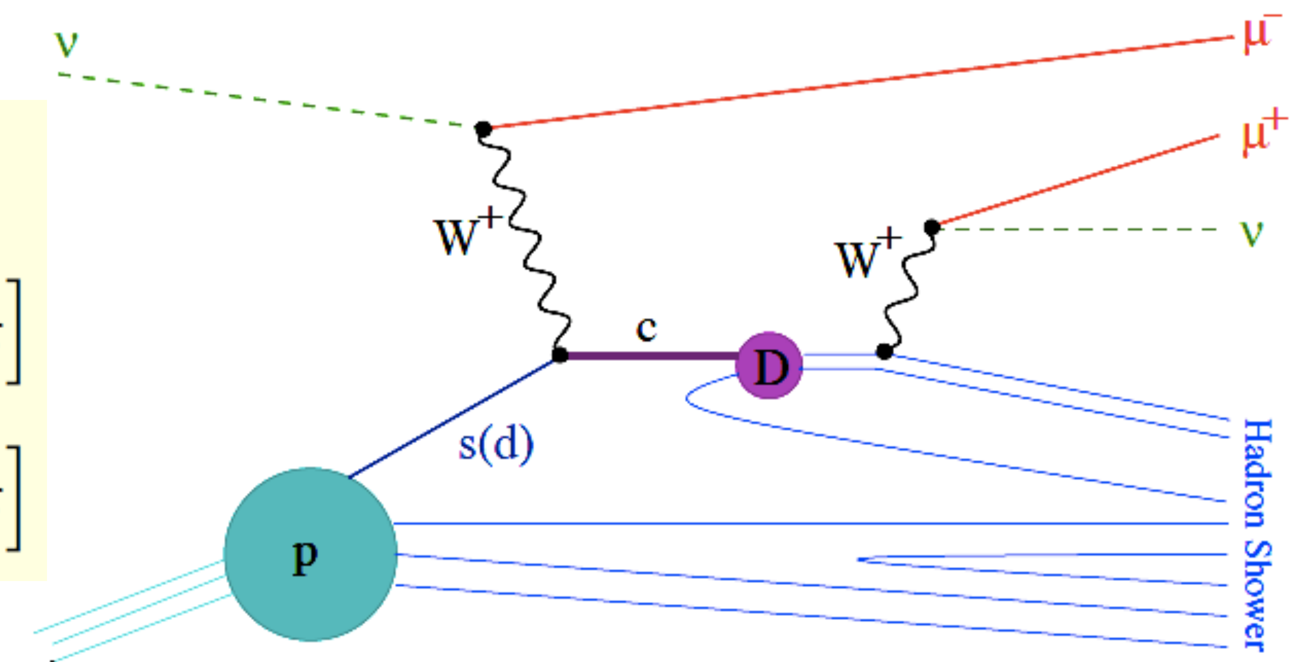
- Valence d/u ratio at high-x accessible thanks to low-Q2 fixed-target experiments (SLAC, BCDMS, NMC, CHORUS, NuTeV, JLAB)
- Testing ground for nucleon models in the x → 1 limit
- At high-x nuclear corrections are very important (deuteron target)

Fixed target DIS neutrino

$$\tilde{\sigma}^{\nu(\bar{\nu}),c} \propto (F_2^{\nu(\bar{\nu}),c}, F_3^{\nu(\bar{\nu}),c}, F_L^{\nu(\bar{\nu}),c})$$

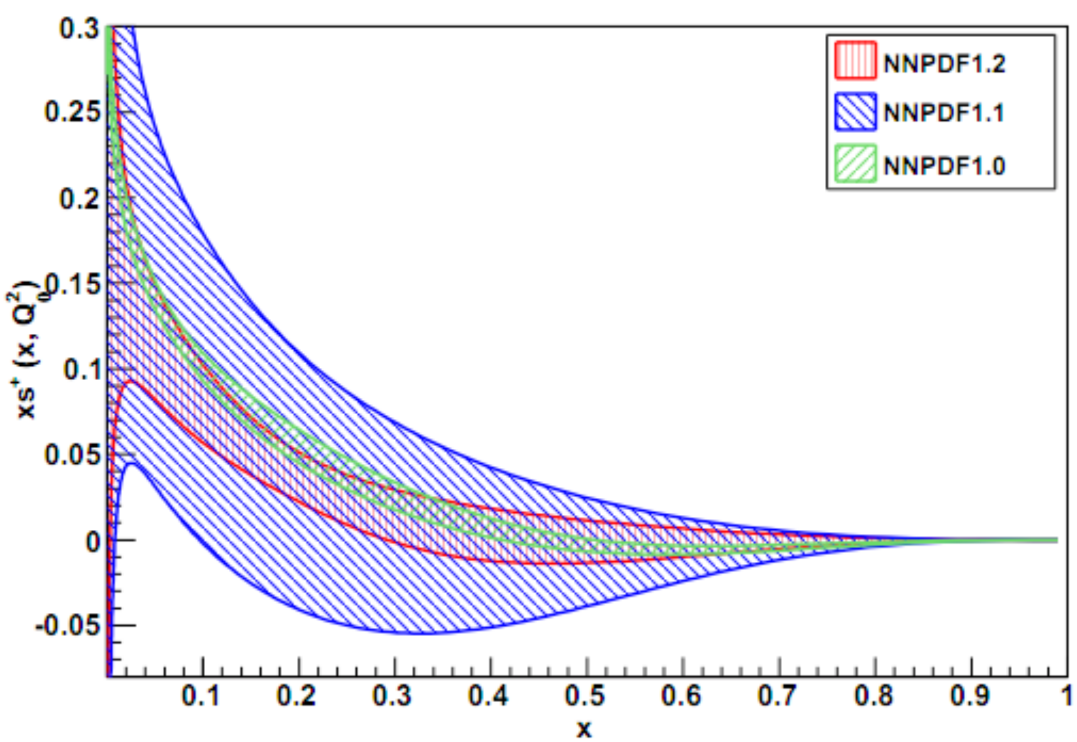
$$F_2^{\nu,c} = x\left[C_{2,q} \otimes 2|V_{cs}|^2 s + \frac{1}{n_f} C_{2,g} \otimes g\right]$$

$$F_2^{\bar{\nu},c} = x\left[C_{2,q} \otimes 2|V_{cs}|^2 \bar{s} + \frac{1}{n_f} C_{2,g} \otimes g\right]$$



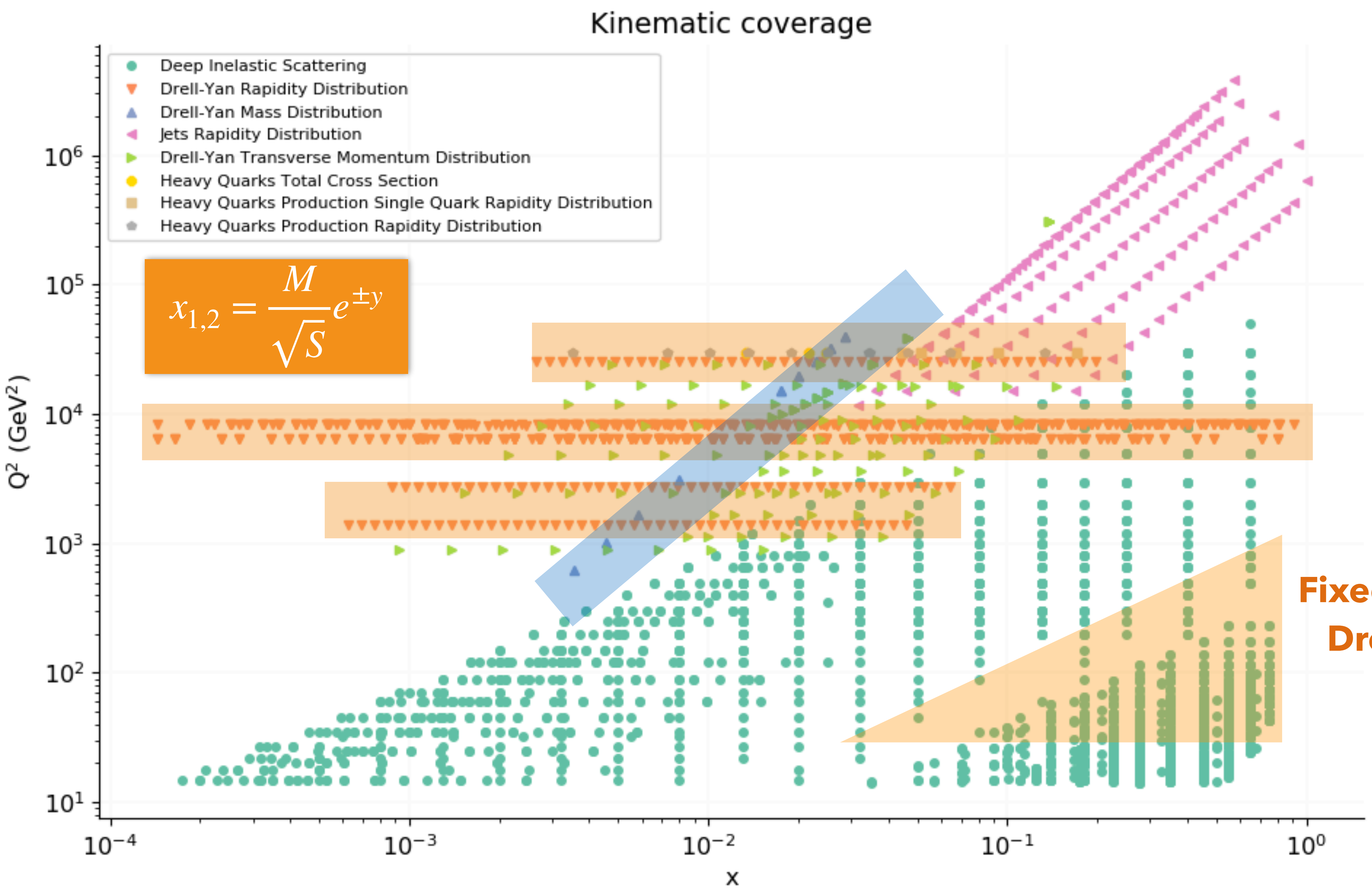
V_{cs} enhancement

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



- NuTeV structure function data on iron target provide strong constraints on strangeness inside the proton
- Some (mild) tension between fixed target data and W+c data at the LHC

Drell-Yan/V production data



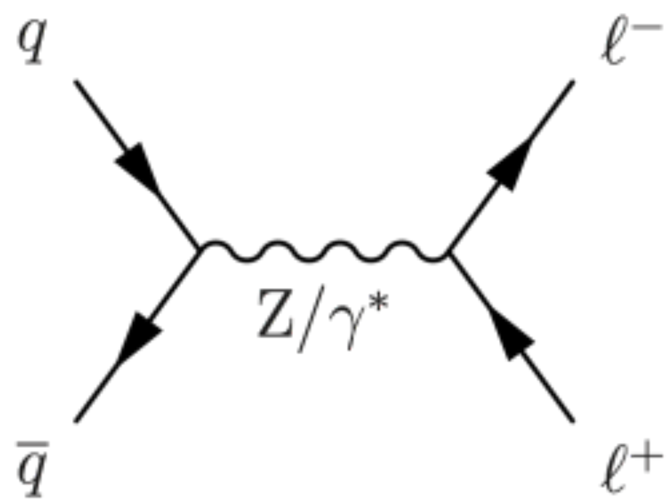
$$x_{1,2} = \frac{M}{\sqrt{S}} e^{\pm y}$$

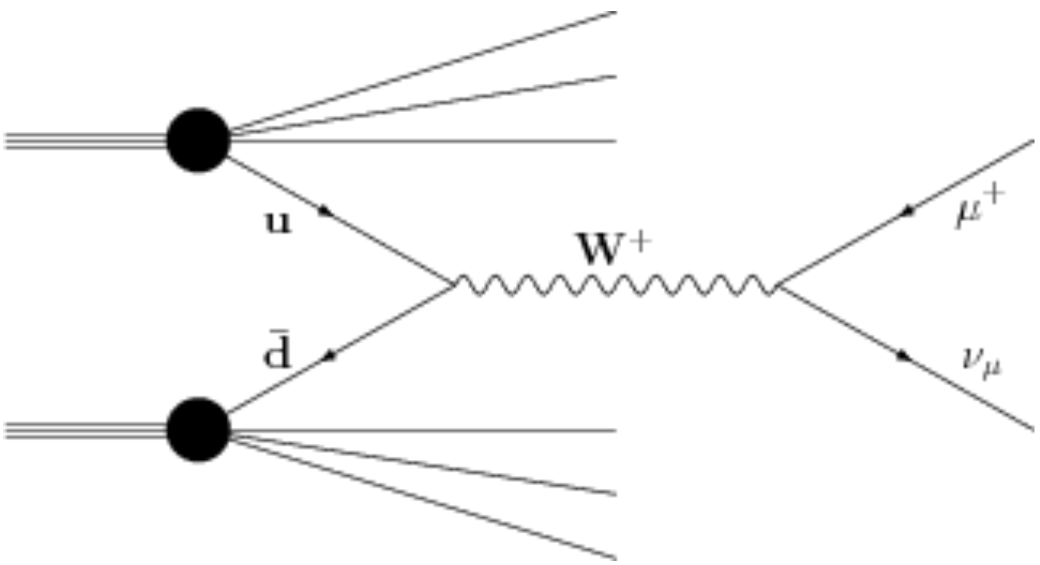
Fixed-target Drell-Yan

Rapidity and invariant mass distribution

Tevatron + LHC

Drell-Yan/V production data





$$\hat{\sigma}_{q\bar{q}} \propto \delta(x_1 x_2 S - M^2) \quad \Rightarrow \quad x_{1,2} = \frac{M}{\sqrt{S}} e^{\pm y}$$

LO in QCD

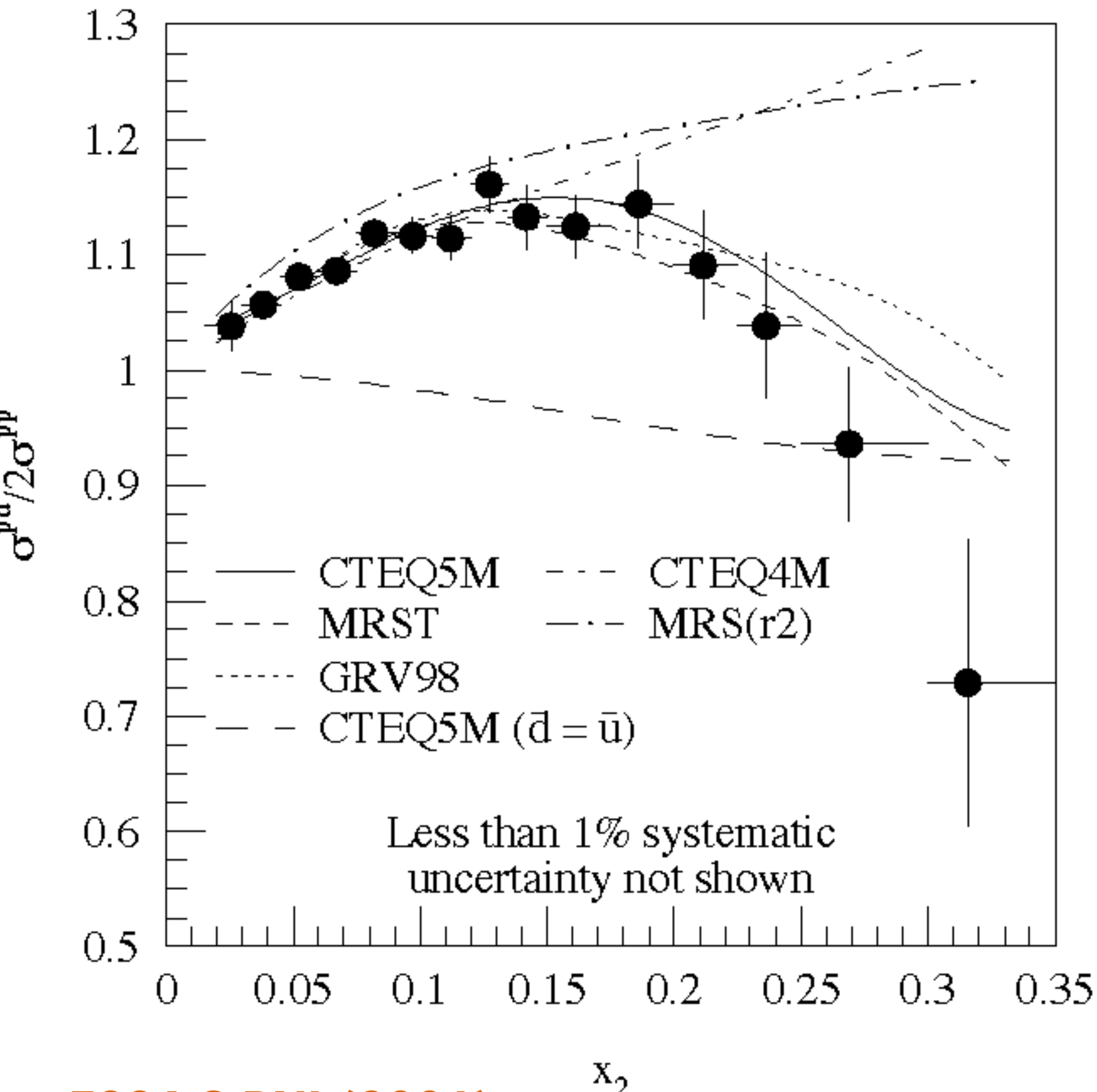
$$L_{ij}(x_1, x_2) = q_i(x_1)\bar{q}_j(x_2)$$

$$\gamma^* : \quad \frac{d\sigma}{dy dM^2} = \frac{4\pi\alpha^2}{9M^2 S} \sum_i e_i^2 L_{ij}(x_1, x_2)$$

$$Z : \quad \frac{d\sigma}{dy} = \frac{\pi G_F M_V^2 \sqrt{2}}{3S} \sum_i (v_{iZ}^2 + a_{iZ}^2) L_{ij}(x_1, x_2)$$

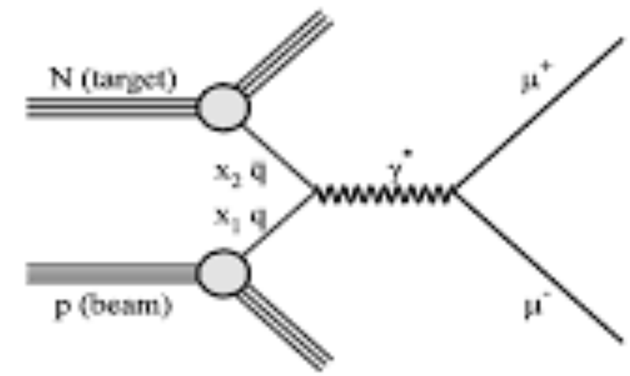
$$W : \quad \frac{d\sigma}{dy dM^2} = \frac{\pi G_F M_V^2 \sqrt{2}}{3S} \sum_{ij} |V_{ij}^{\mathrm{CKM}}|^2 L_{ij}(x_1, x_2)$$

Fixed target Drell-Yan



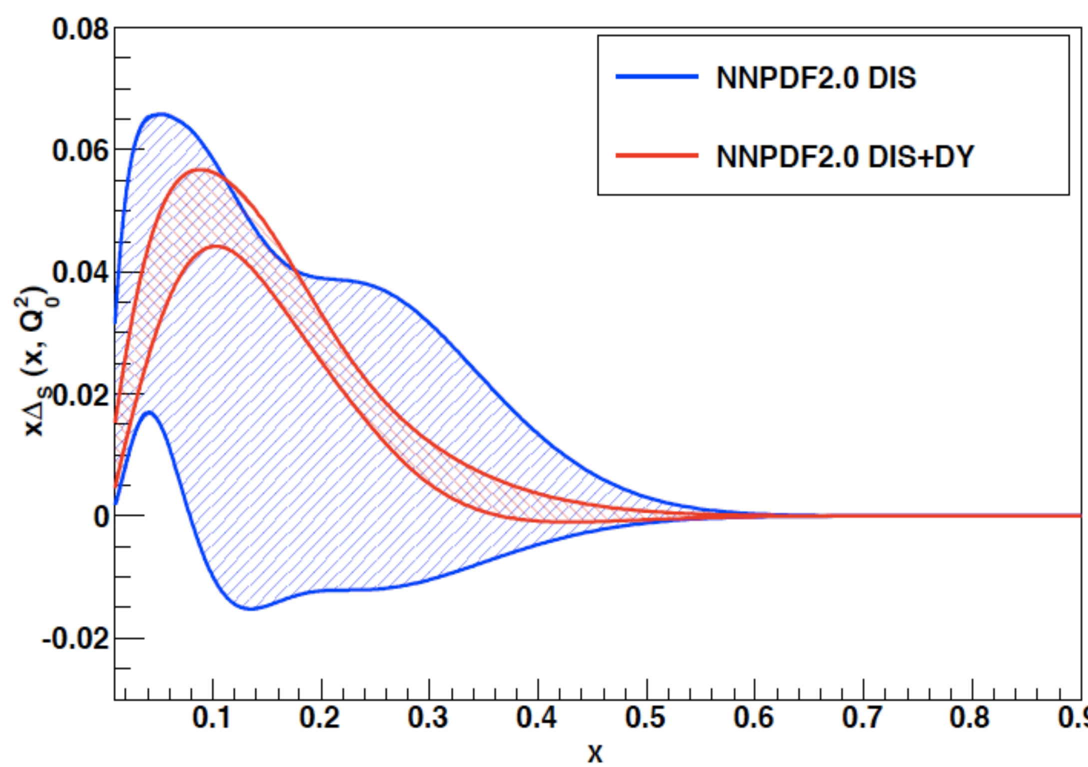
E886 @ BNL (2001)

The Drell-Yan Process: $pN \rightarrow \mu^+\mu^- X$

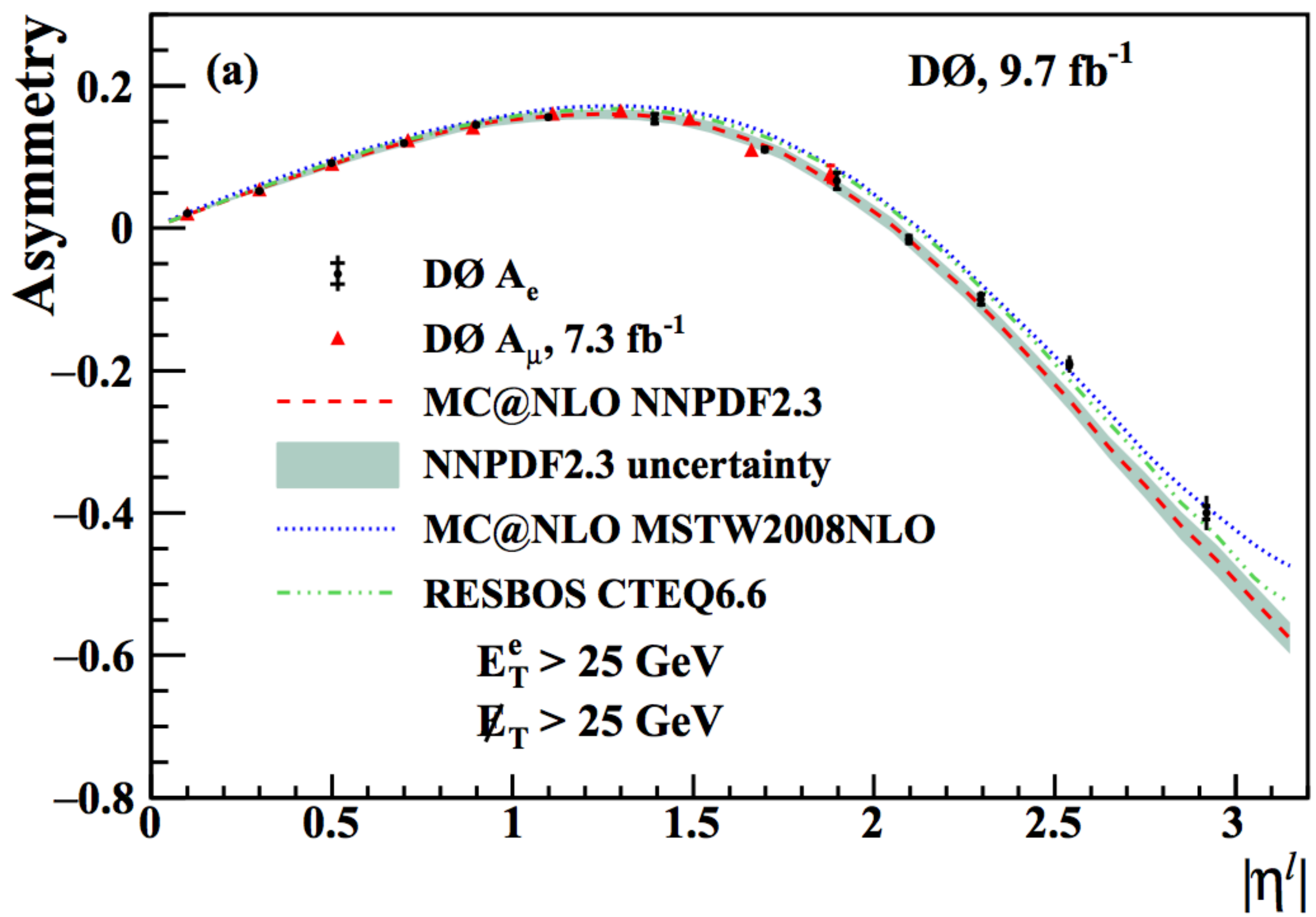


Fixed-target Drell-Yan

$$\frac{\sigma(pd \rightarrow \mu^+\mu^-)}{\sigma(pp \rightarrow \mu^+\mu^-)} = \frac{\frac{4}{9}u\bar{d} + \frac{1}{9}d\bar{u}}{\frac{4}{9}u\bar{u} + \frac{1}{9}d\bar{d}} \sim \frac{\bar{d}}{\bar{u}}$$



Z/W production at Tevatron



W asymmetry at Tevatron

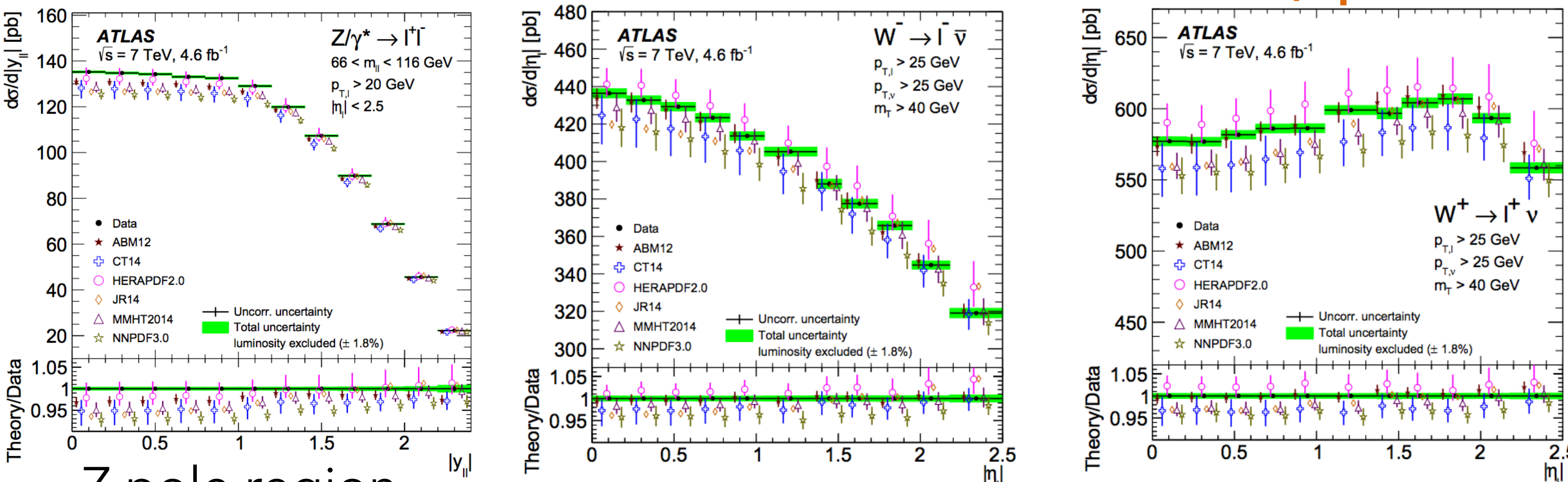
$$u^{\bar{p}} = \bar{u}^p$$
$$d^{\bar{p}} = \bar{d}^p$$

Charge conjugation

$$\frac{\sigma(p\bar{p} \to W^+)}{\sigma(p\bar{p} \to W^-)} = \frac{u(x_1)d(x_2) + \bar{u}(x_1)\bar{d}(x_2)}{d(x_1)u(x_2) + \bar{d}(x_1)\bar{u}(x_2)} \sim \frac{u}{d}(x_1)\frac{u}{d}(x_2)$$

Z/W production at the LHC

ATLAS W,Z production 7 TeV



Z pole region

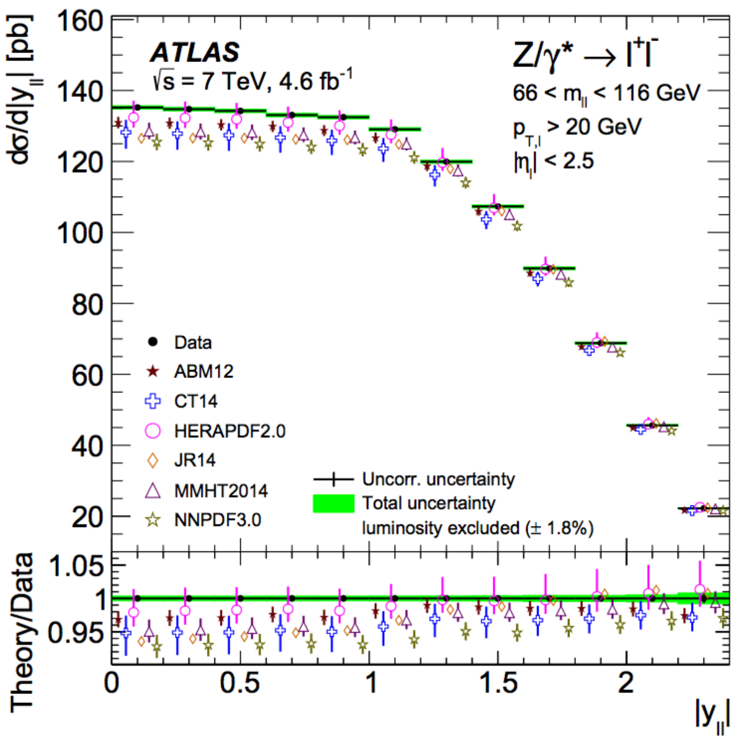
$$\sigma(pp \to Z) = u\bar{u} + d\bar{d} + s\bar{s}$$

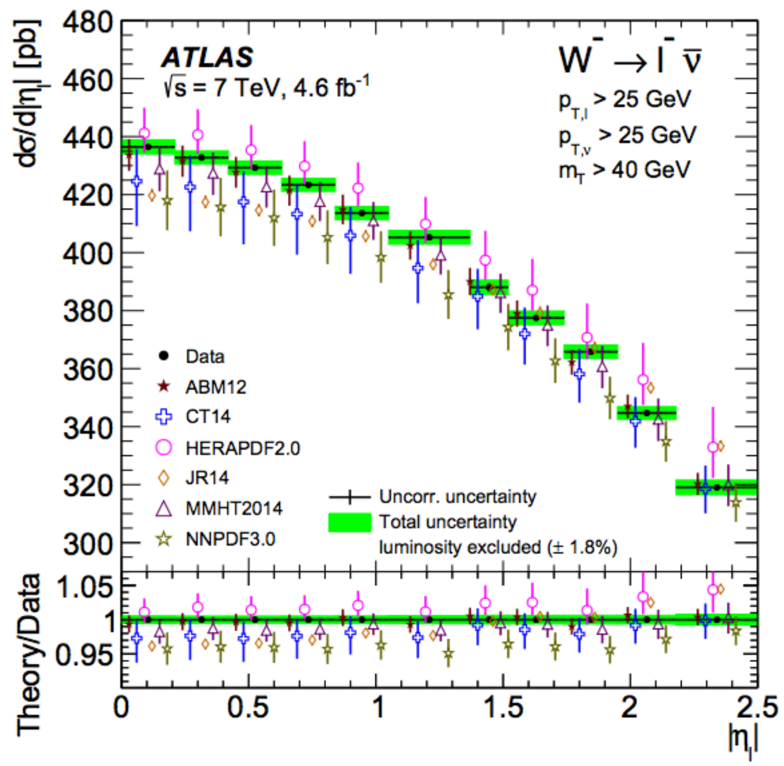
$$\sigma(pp \to W^+) = u\bar{d} + c\bar{s}$$

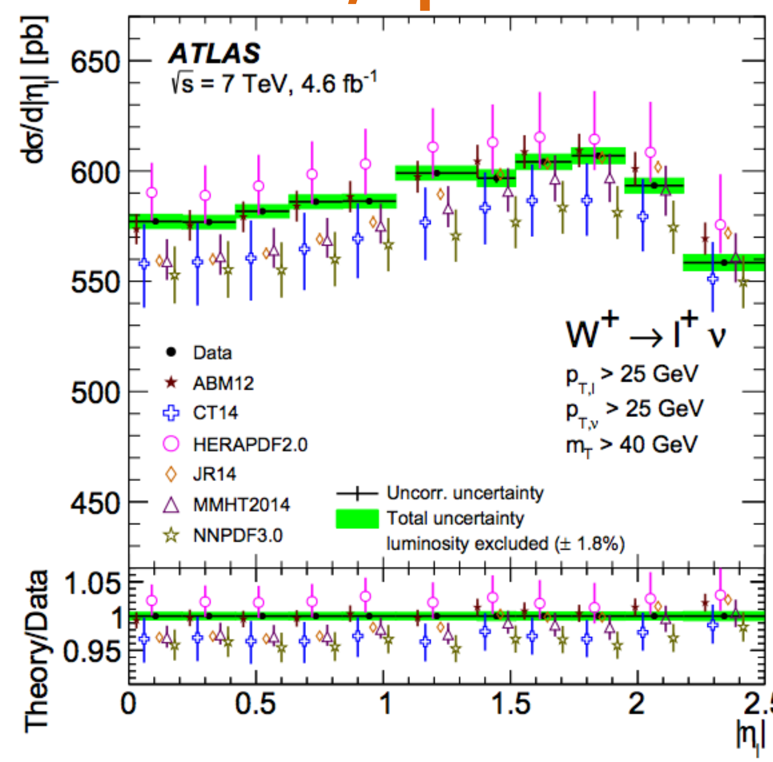
$$\sigma(pp \to W^-) = d\bar{u} + s\bar{c}$$

Z/W production at the LHC

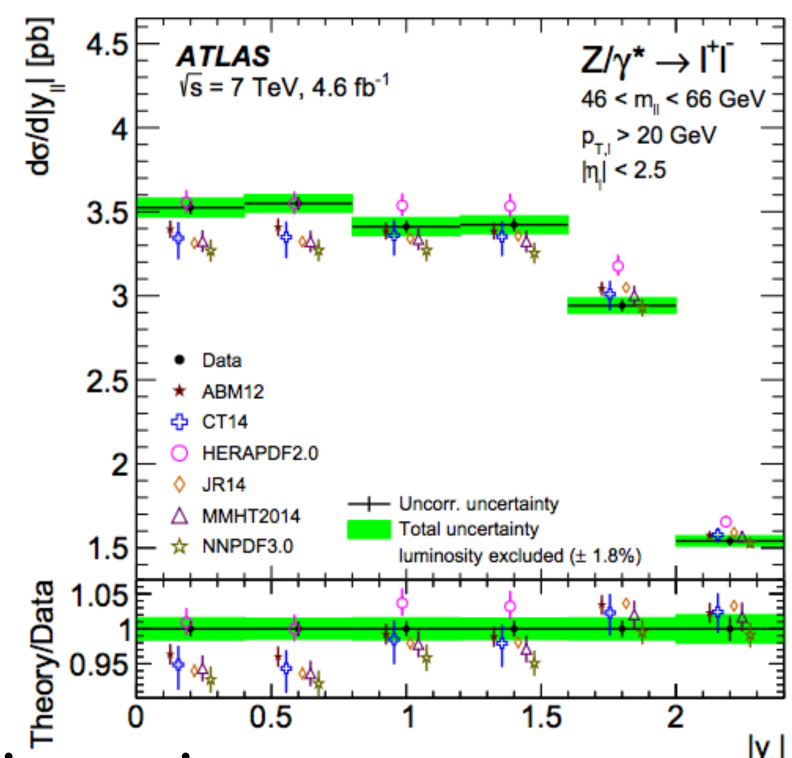
ATLAS W,Z production 7 TeV

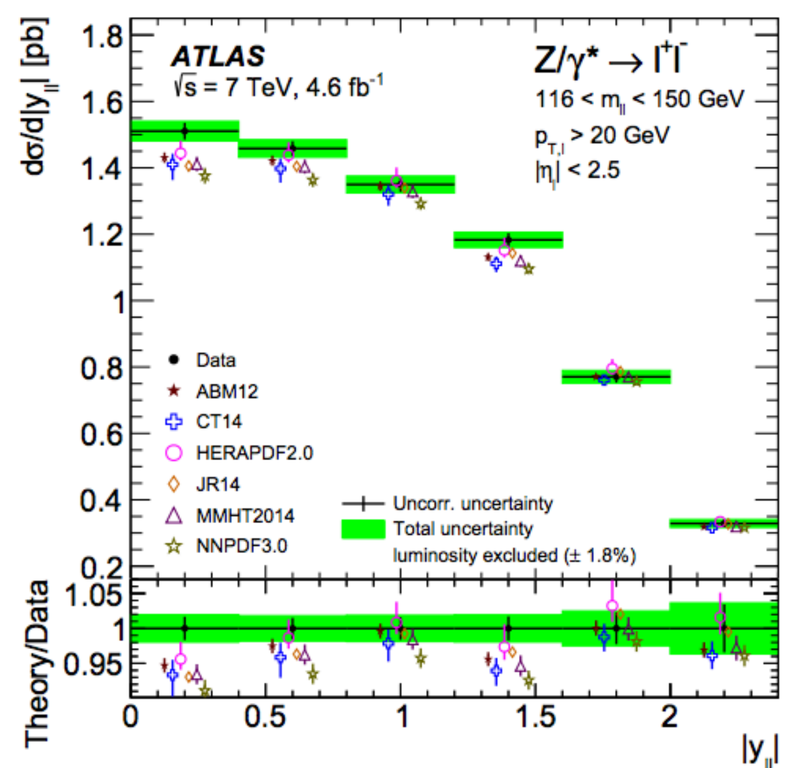






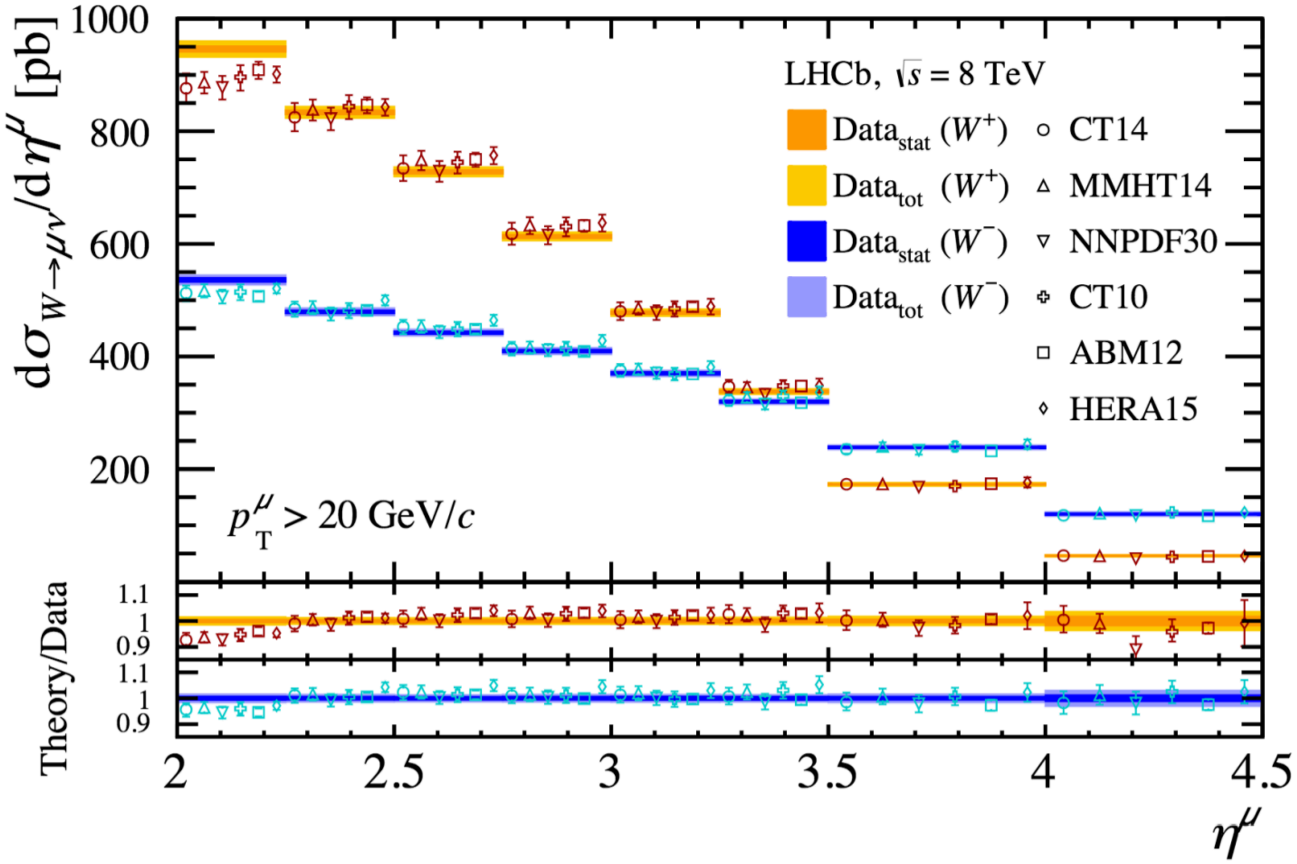
$$x_{1,2} = \frac{M}{\sqrt{S}} e^{\pm y}$$

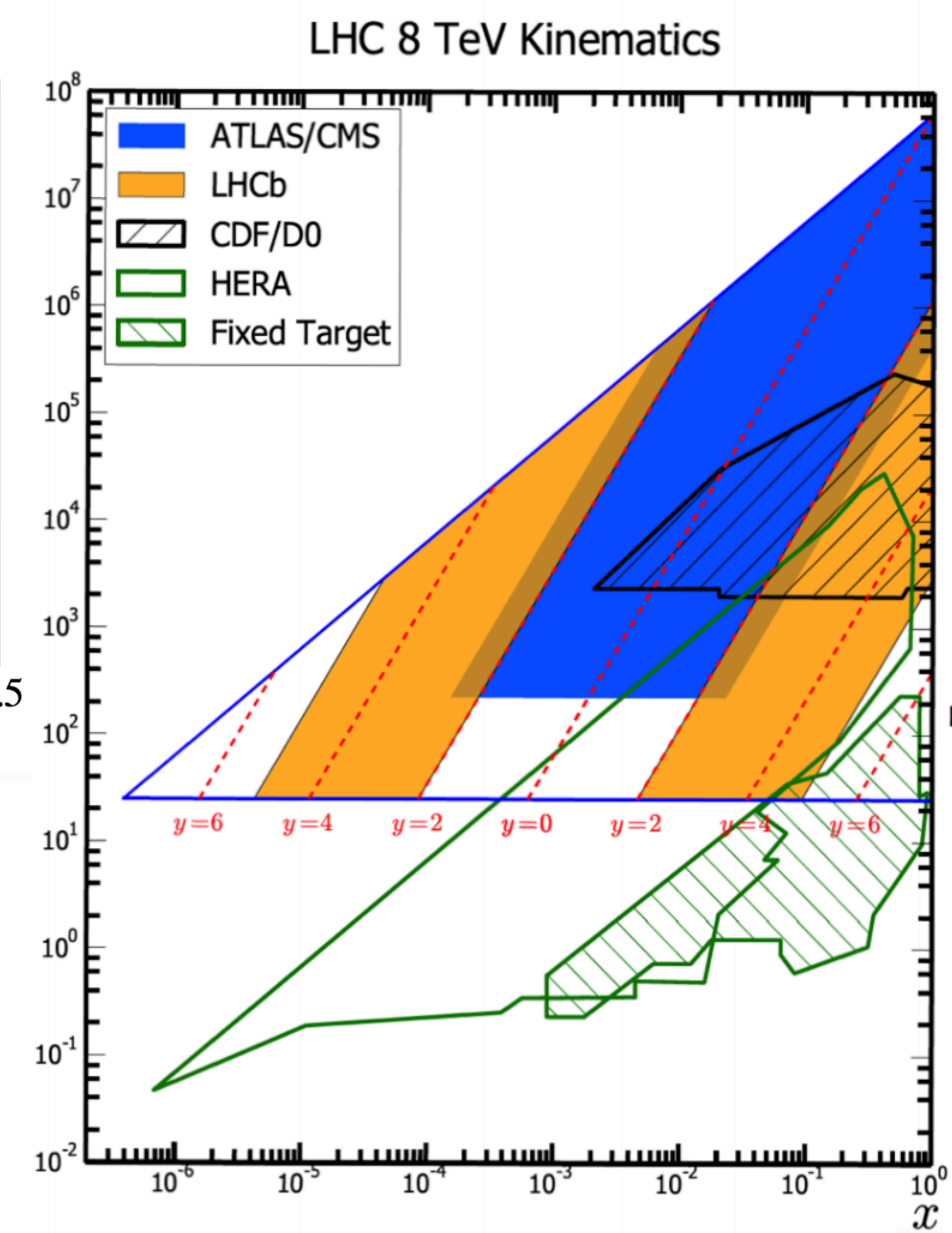




high/low invariant mass

Z/W production at the LHC

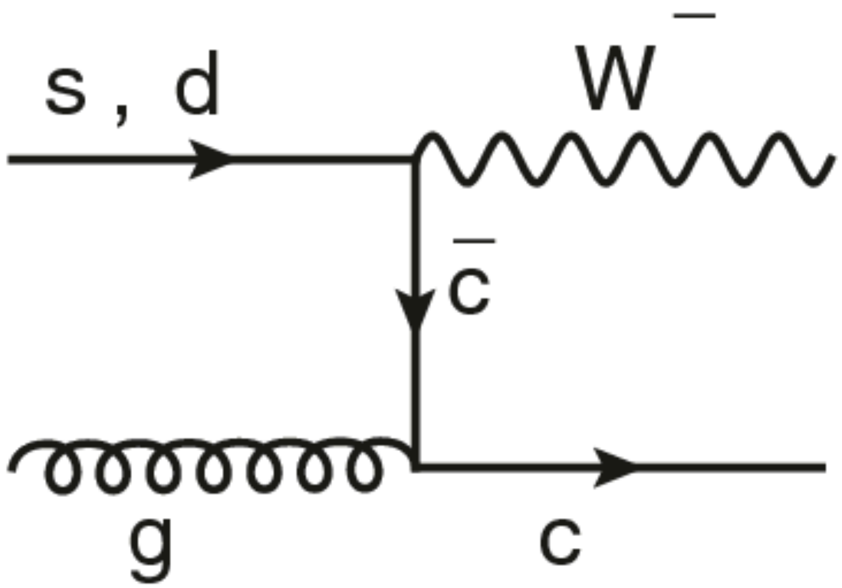


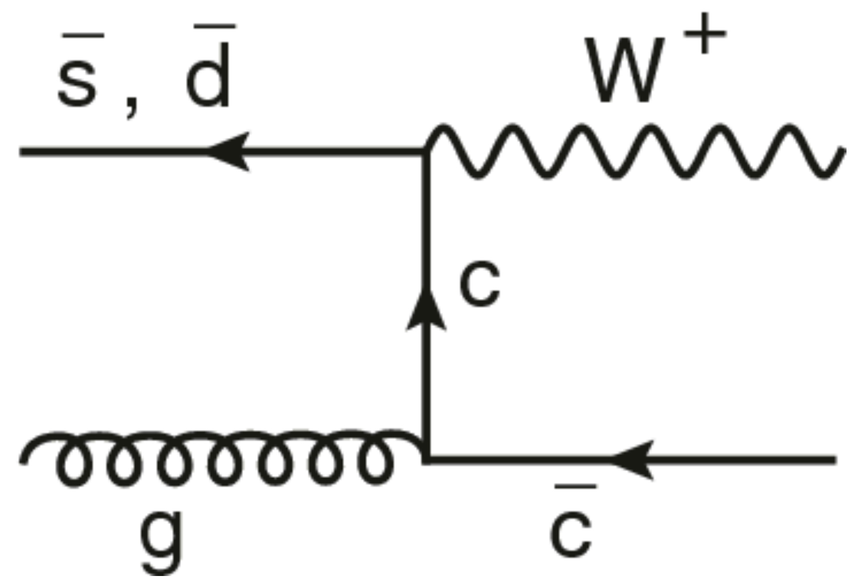


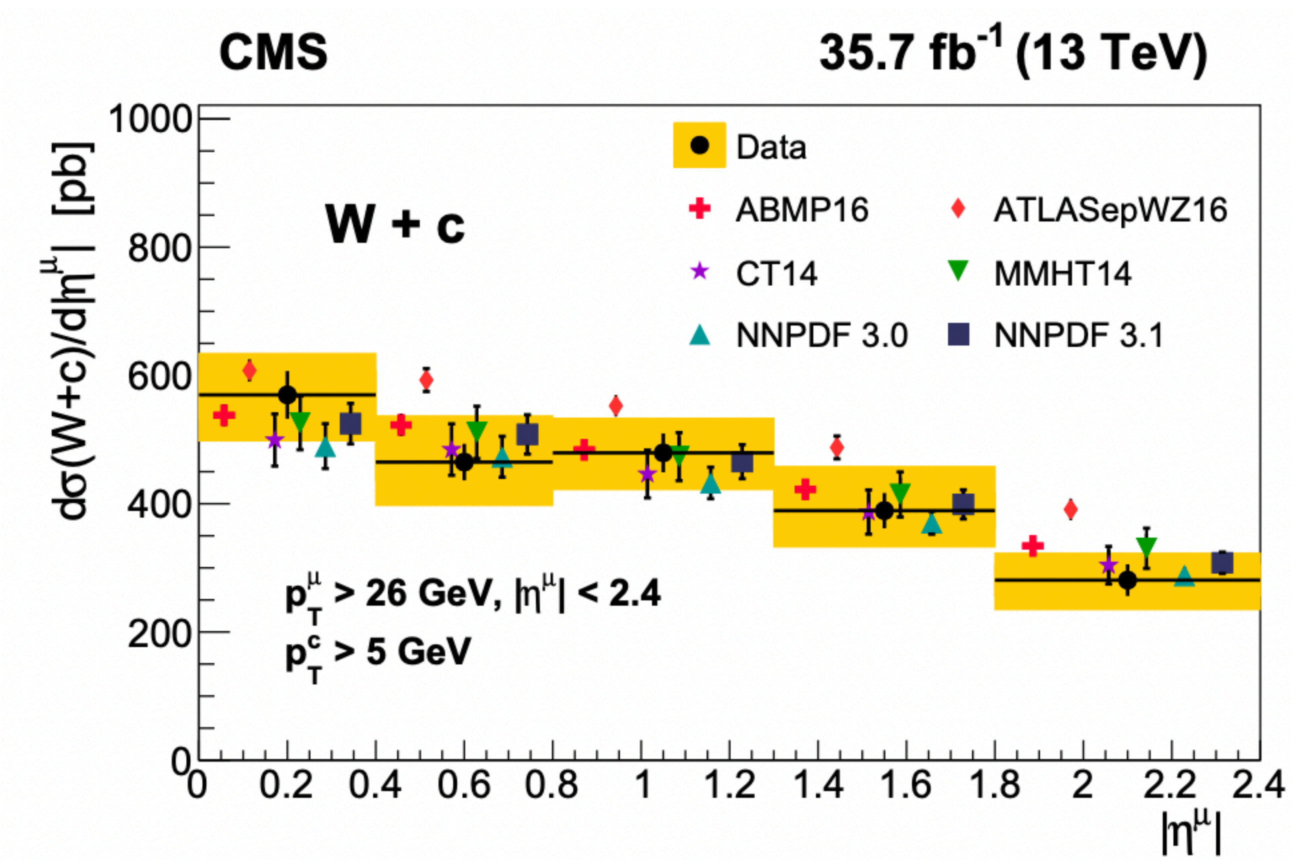
$$x_{1,2} = \frac{M}{\sqrt{S}} e^{\pm y}$$

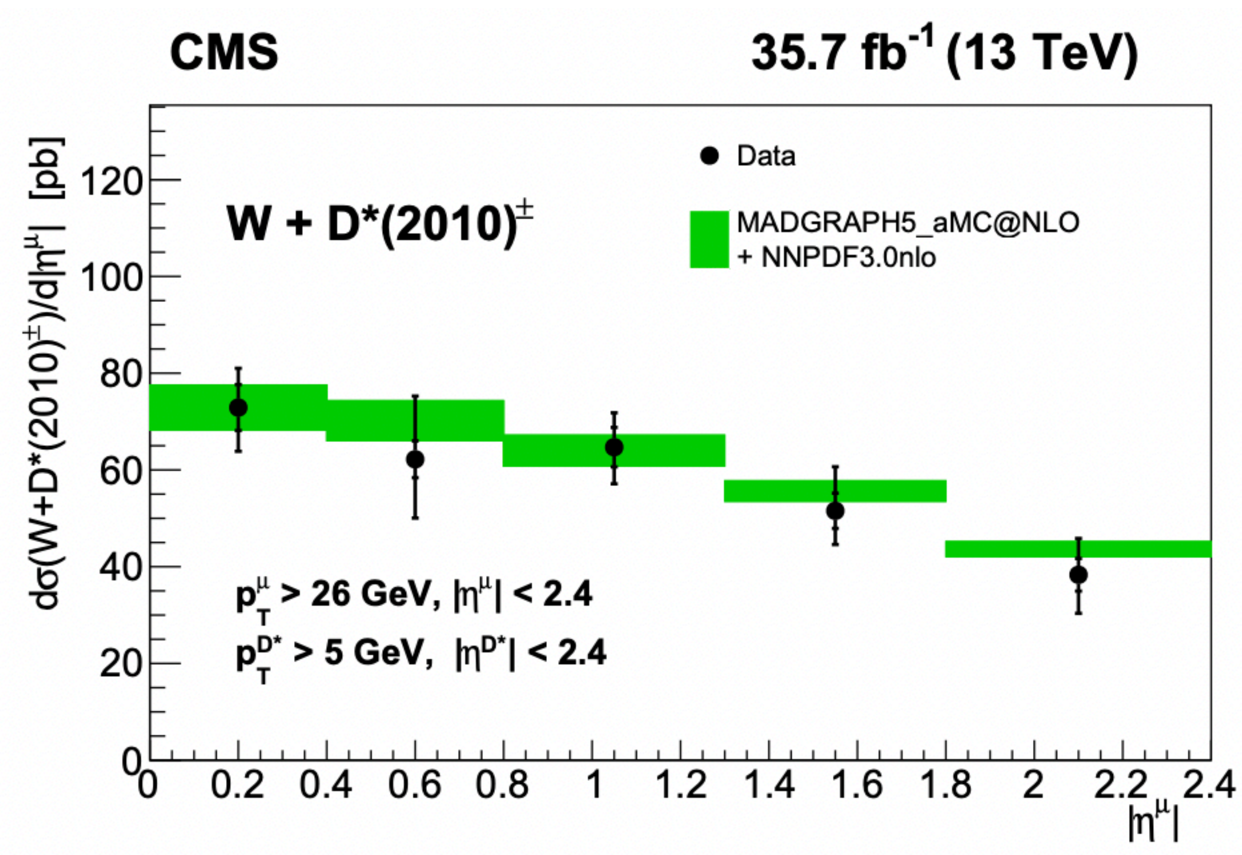
LHCb: forward region

W+charm data



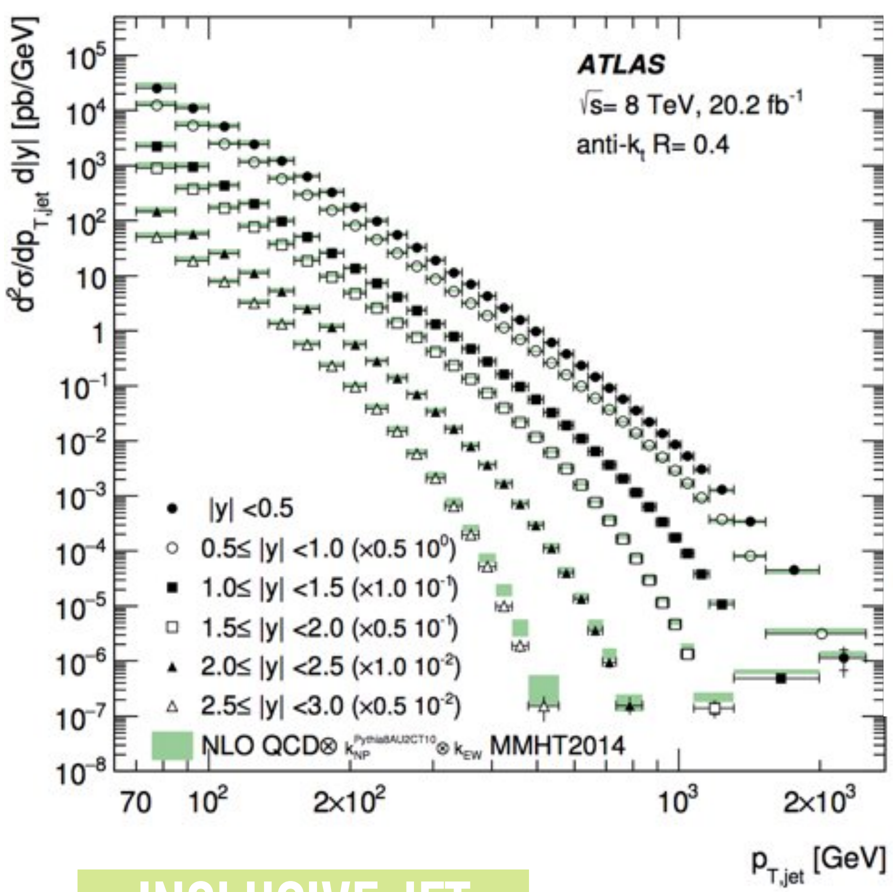




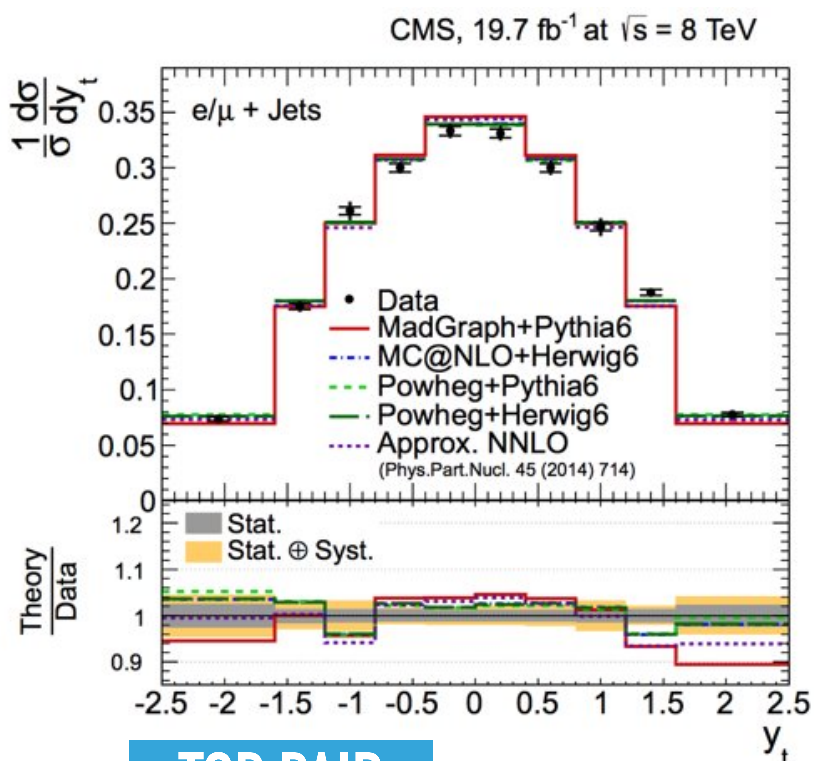


CMS collaboration, arXiv:1811.10021

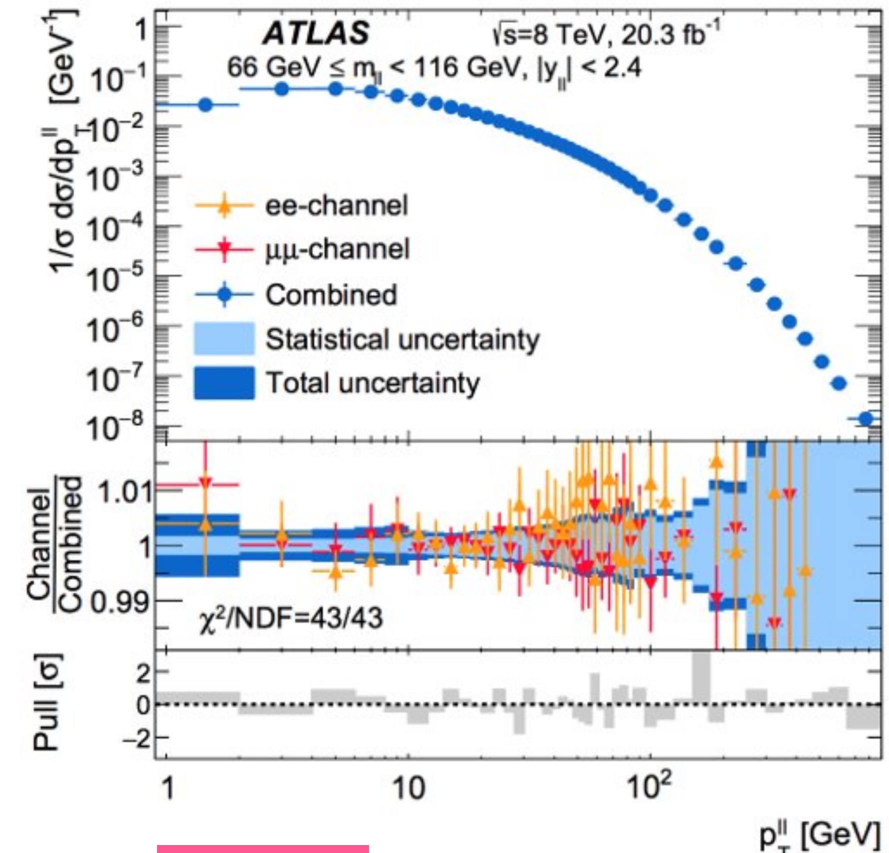
Gluon: direct handle



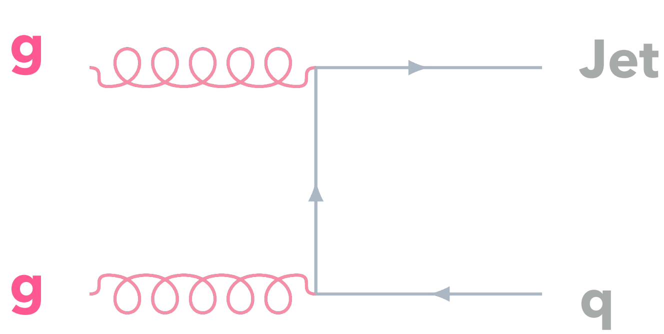
INCLUSIVE JET

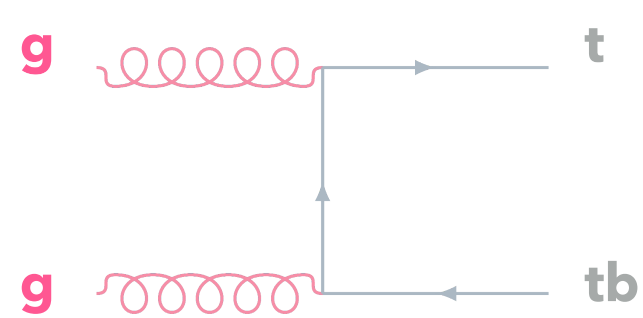


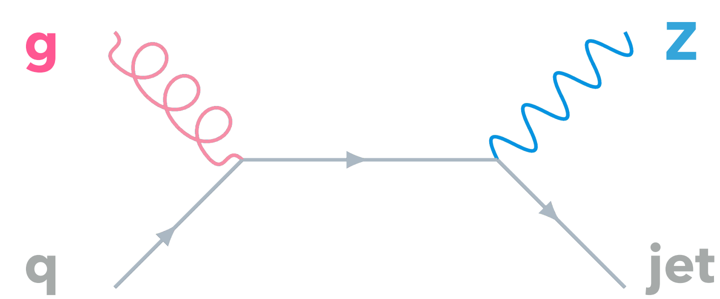
TOP PAIR



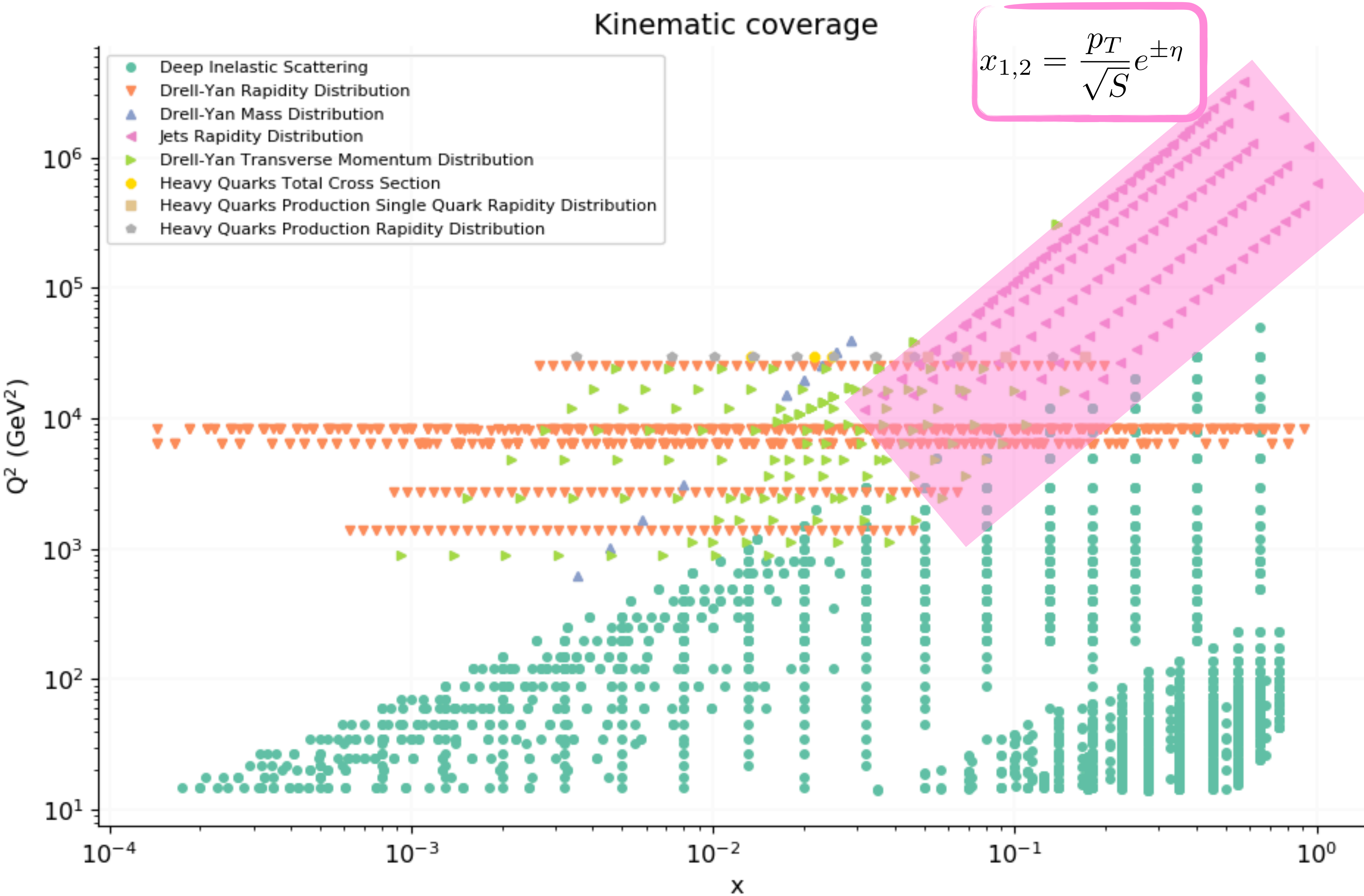
Z P_T



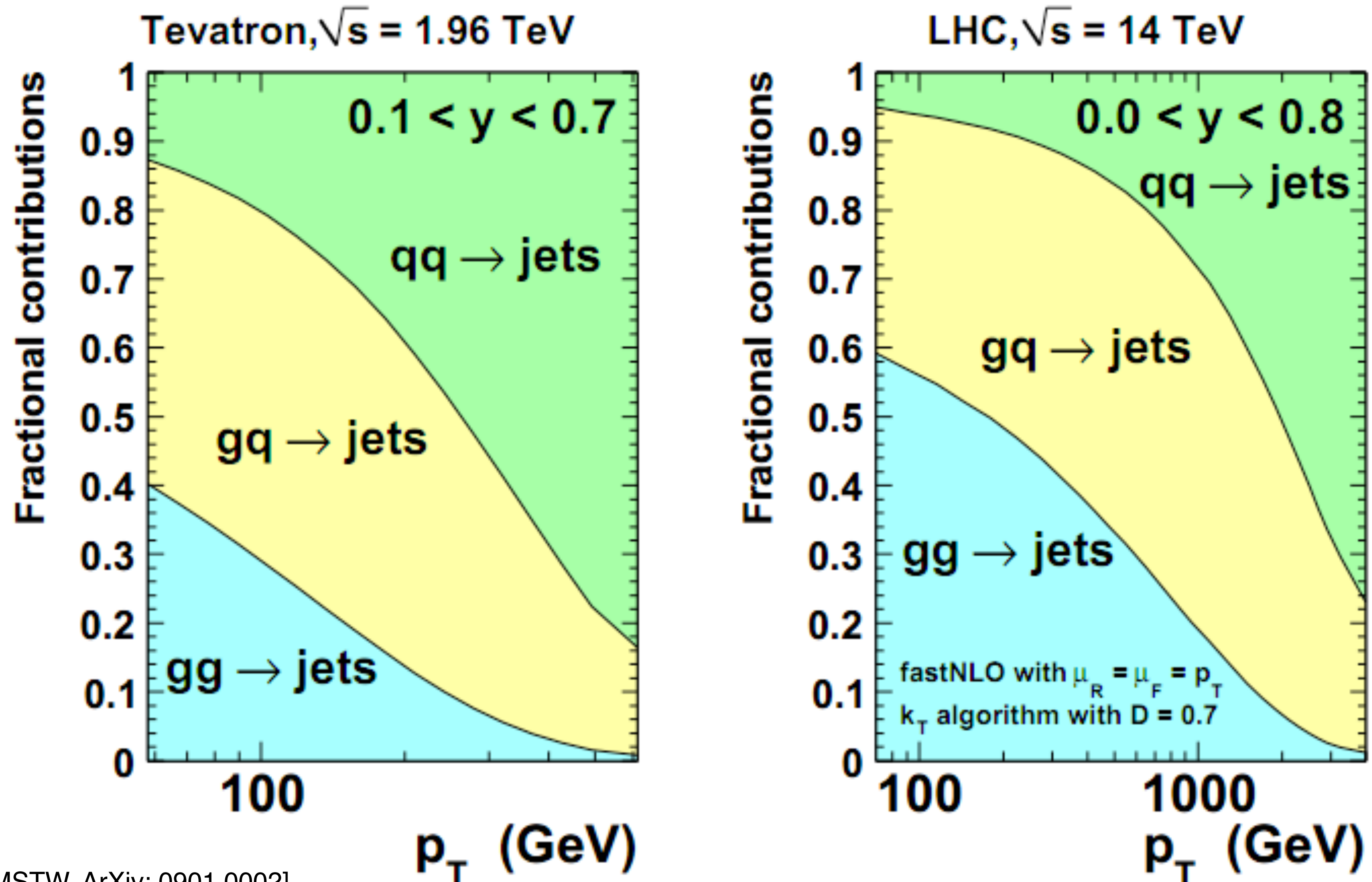




Gluon: jets data



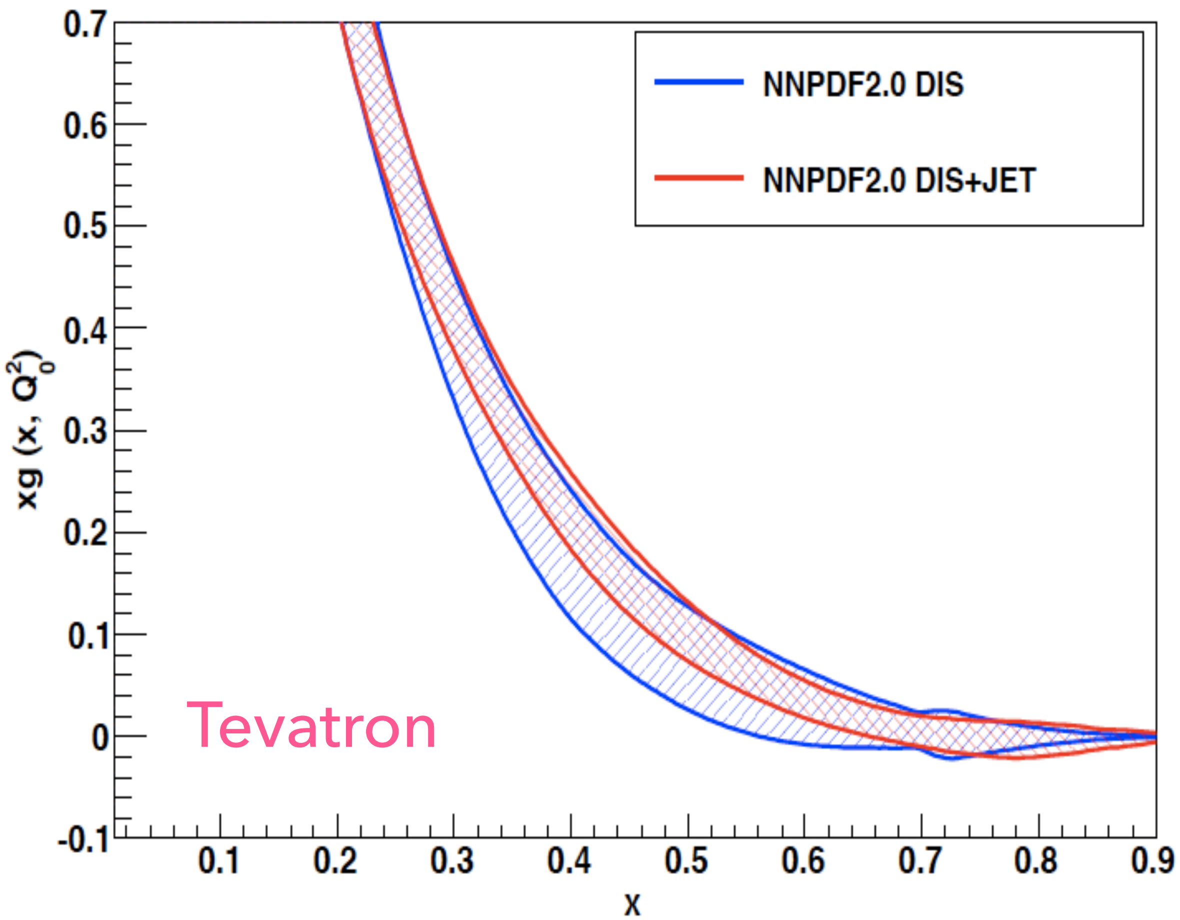
Gluon: jets data



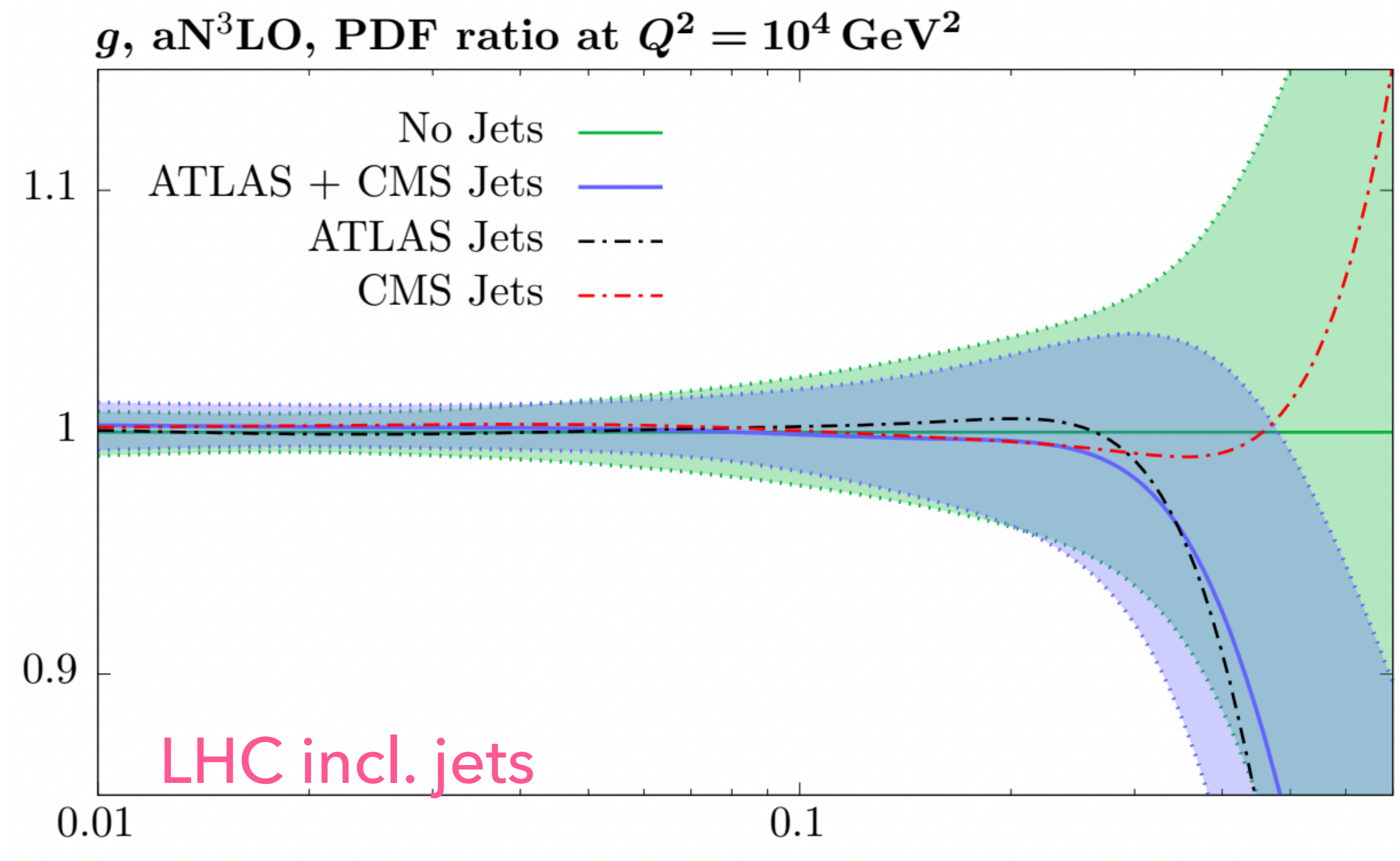
[MSTW, ArXiv: 0901.0002]

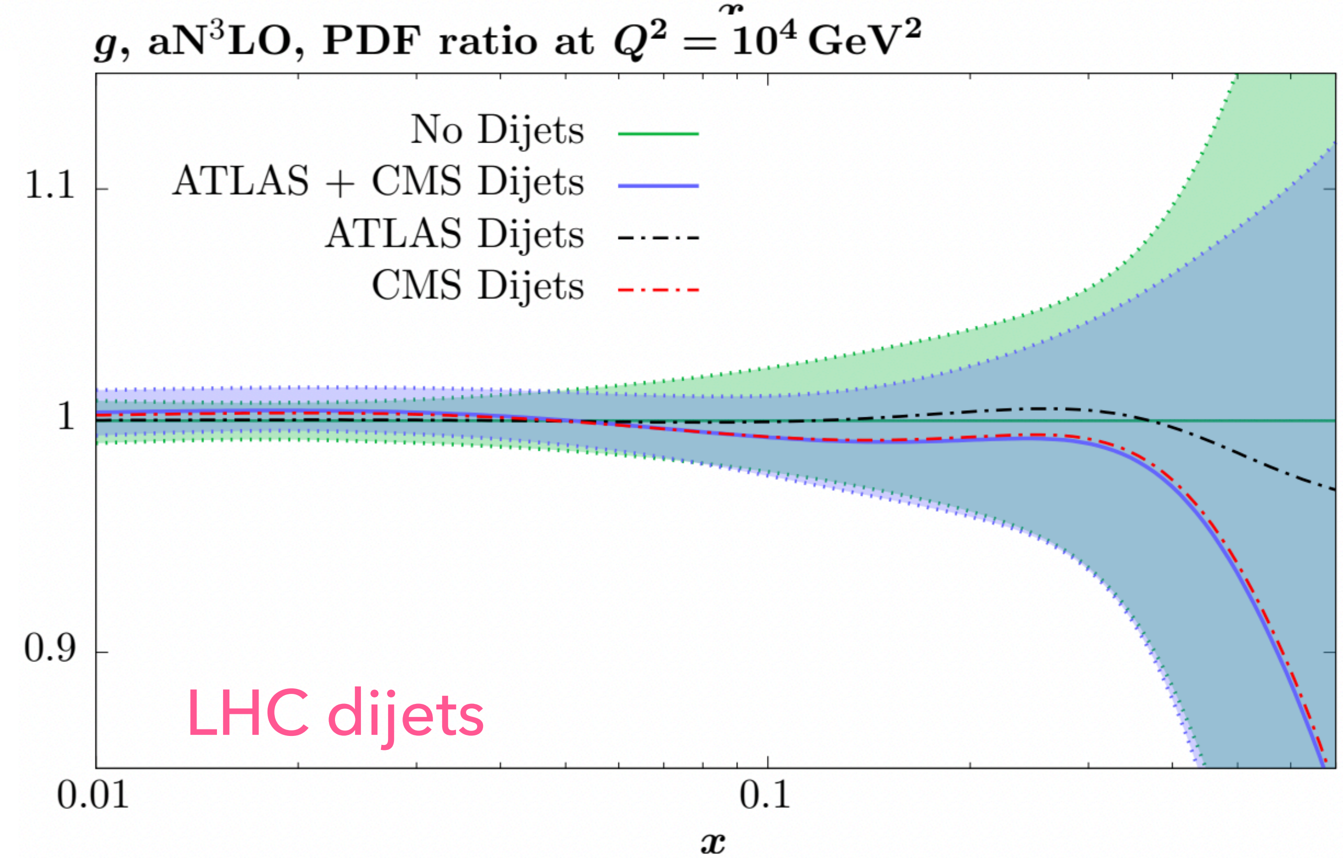
Gluon: jets data

MSHT, arXiv:2312.12505

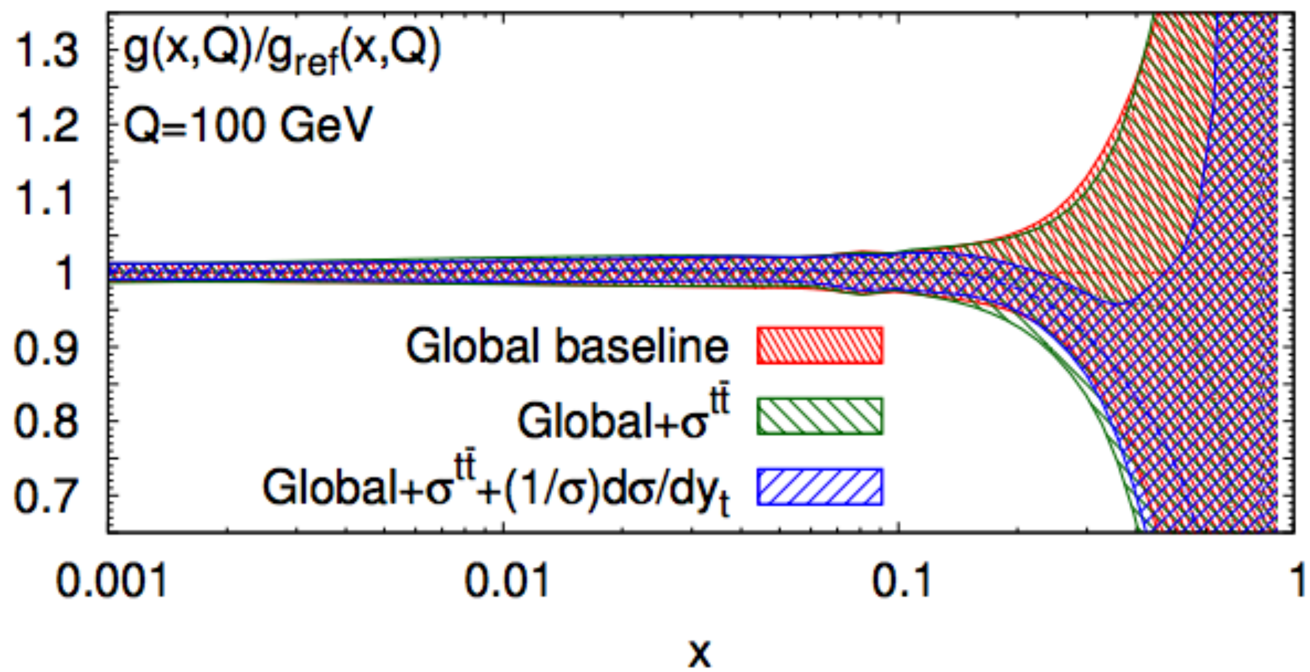


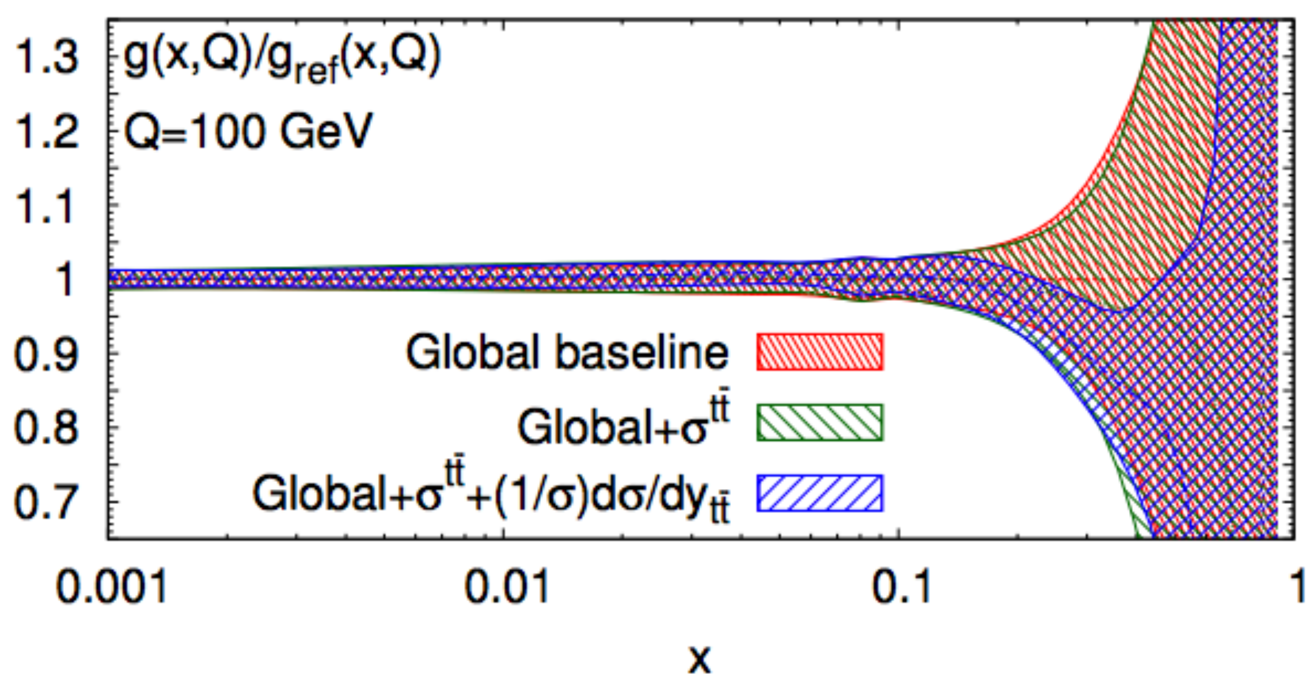
NNPDF, arXiv:1002.4407



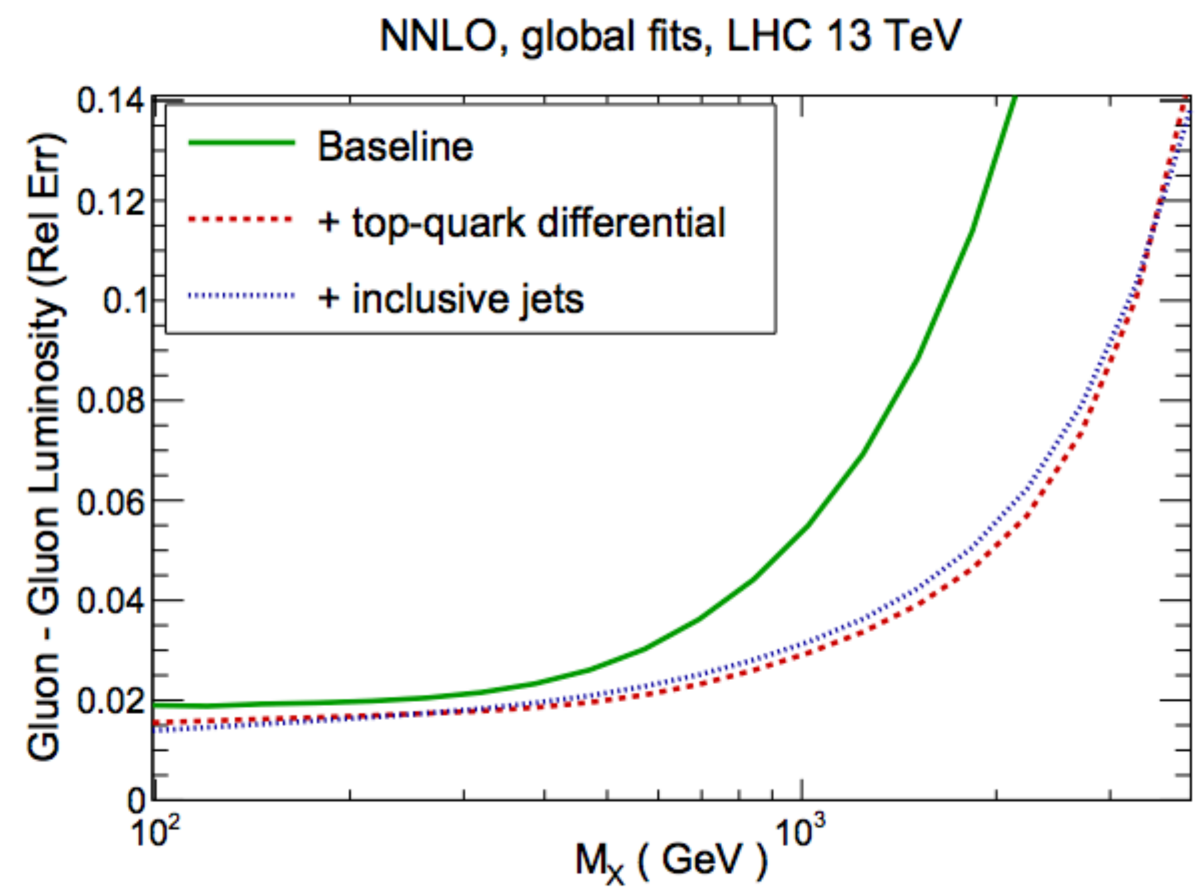


Gluon: top pair production





Czakon, Hartland, Mitov, Nocera and Rojo, arXiv: 1611.08609



- Most constraining is inclusion of y_t list from ATLAS and y_{tt} from CMS jointly with total xsec
- Competitive reduction of gluon uncertainty with jets measurement
- Slight tension between ATLAS and CMS ($\chi^2_{ATLAS} \sim 1.6$, $\chi^2_{CMS} \sim 0.9$)

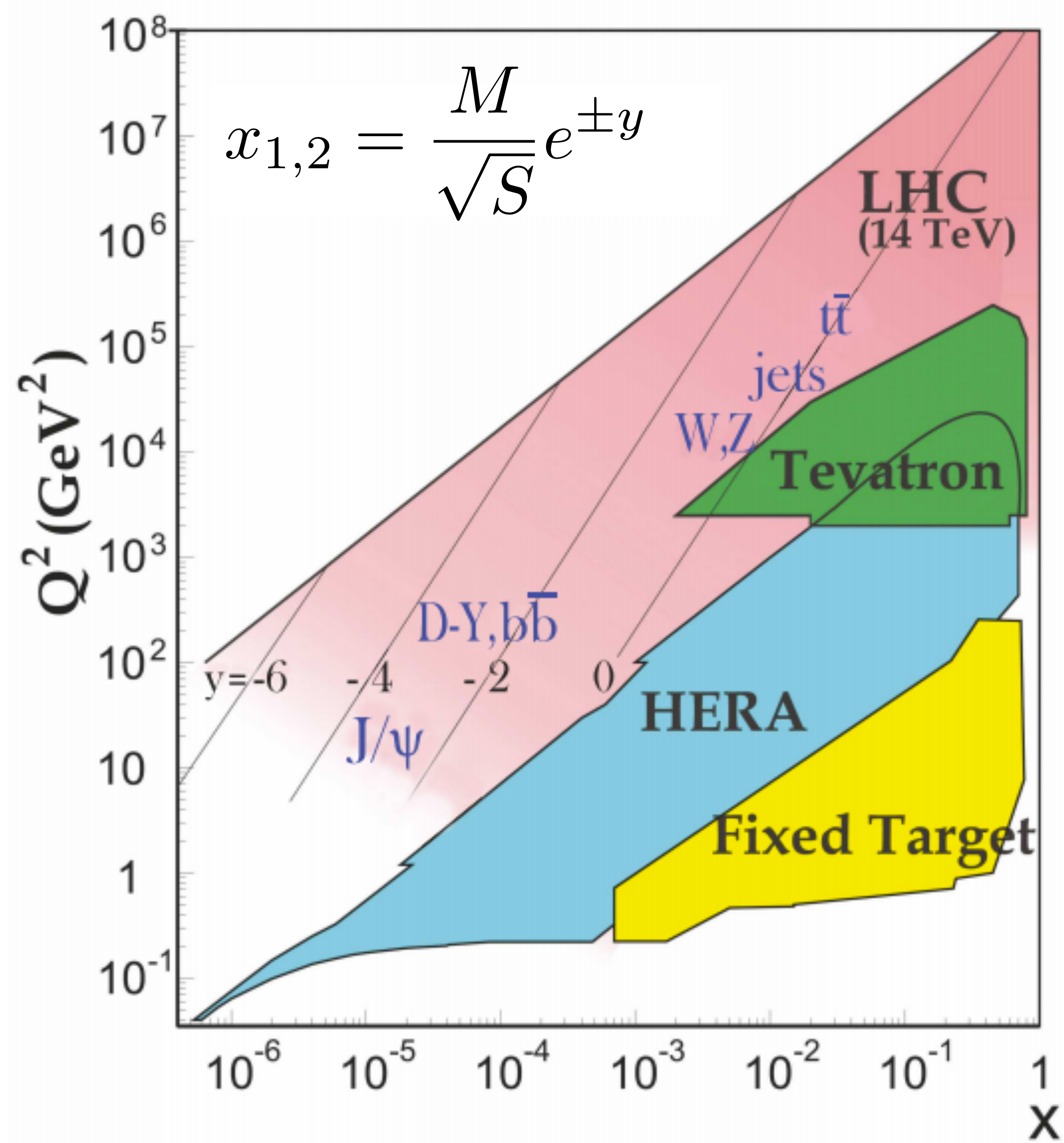
To summarise

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 **(medium x)**

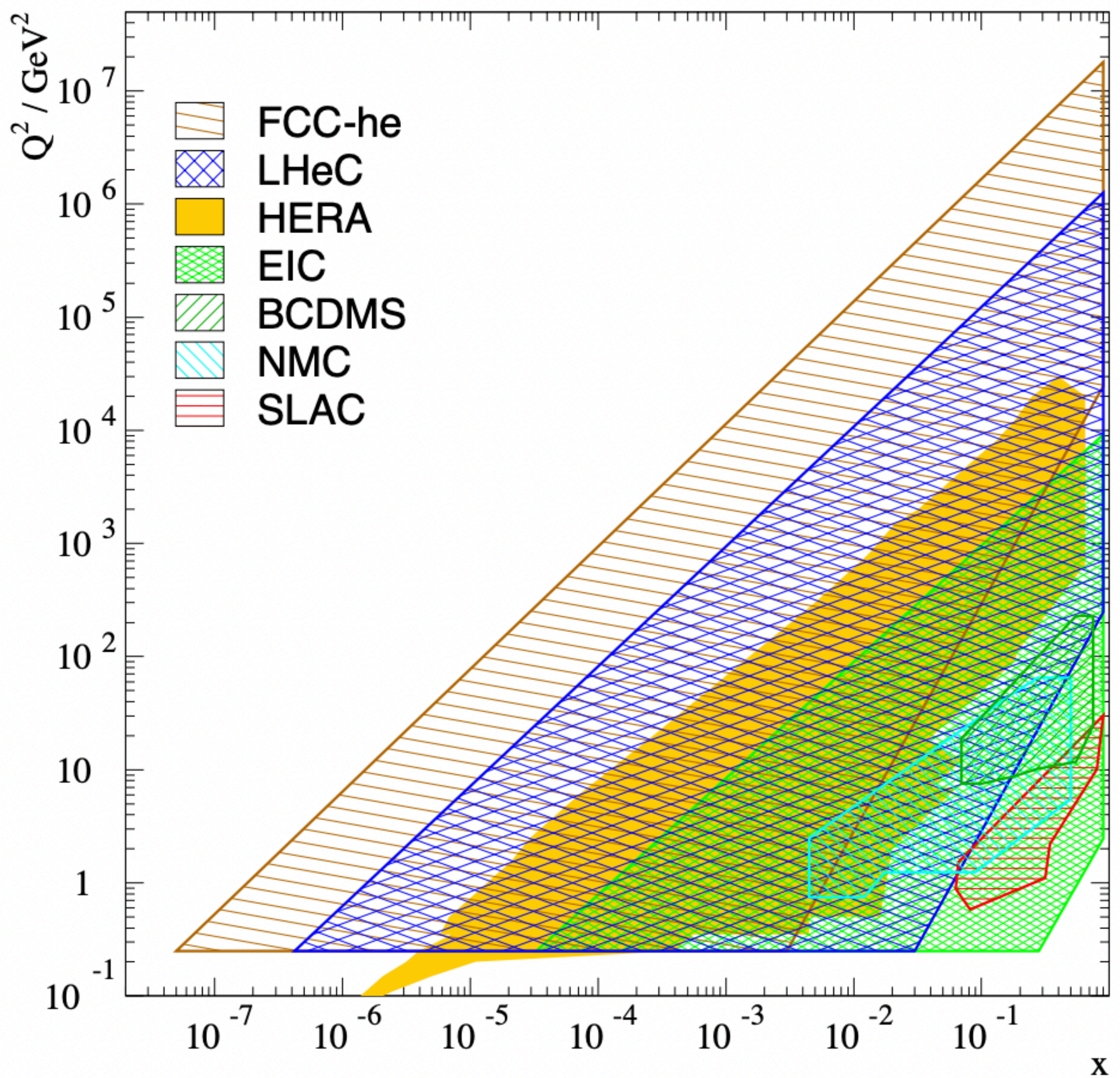
QUARKS

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



PDG 2016

Looking forward



T. Hobbs, SciPost(2022)

2. Methodological aspects

A quite complicated game

- A single quantity: 1σ error
- Multiple quantities: 1σ contours
- Functions: 1σ "error band" in the space of functions
  = find the probability density in the space of functions f(x)
  Expectation values are functional integrals

Not as simple as it may look…

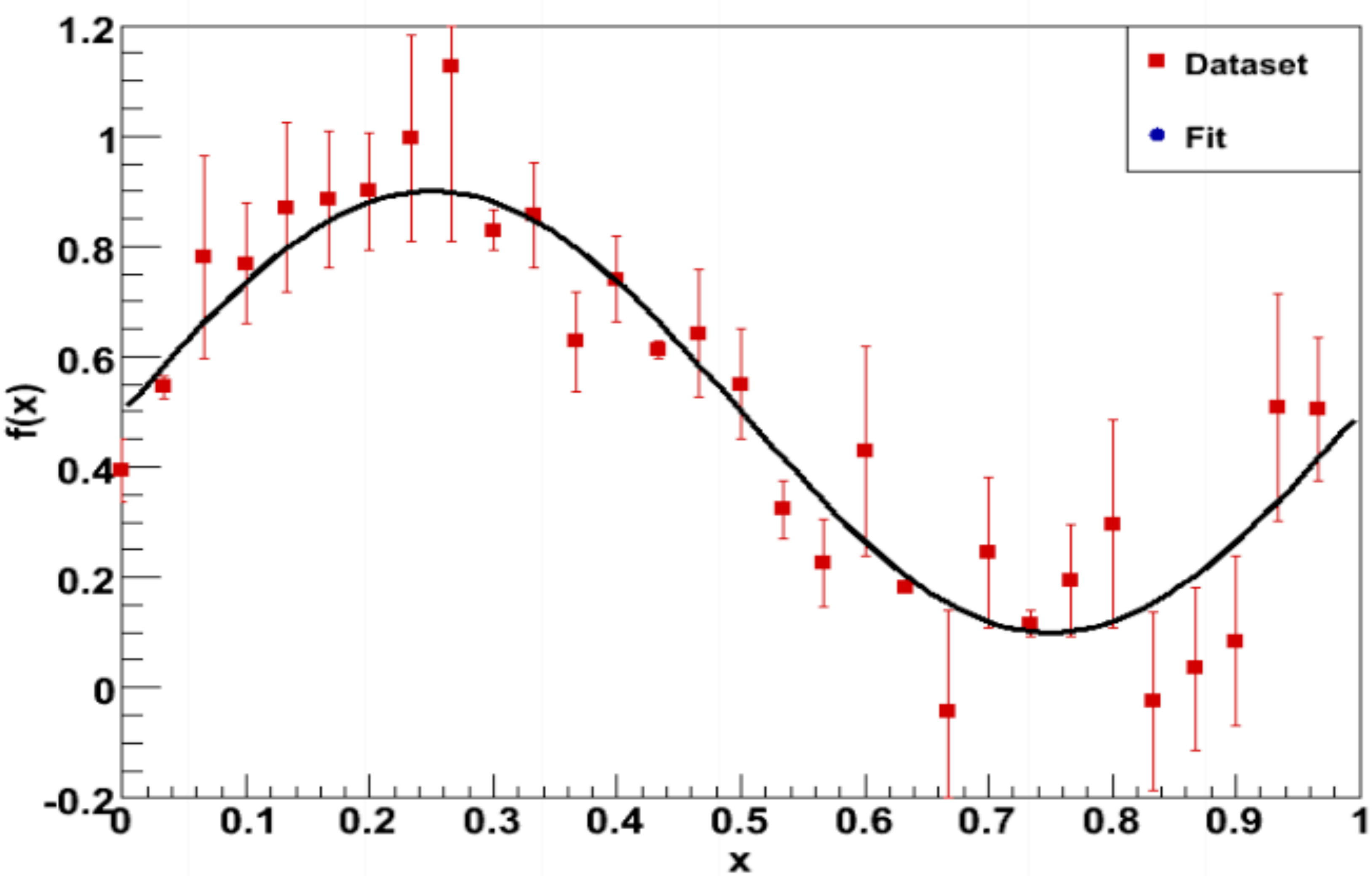
$$\langle \mathcal{O}[\{f\}] \rangle = \int [\mathcal{D}f]\, \mathcal{O}[\{f\}] \mathcal{P}[\{f\}]$$

- 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 toy model

A toy-model:

1) Imagine that we have a set of uncorrelated measurements of a quantity f(x) at different x. The underlying law that Nature established for this quantity is a sinusoidal, but we don't know anything about that and try to guess it with a fit.

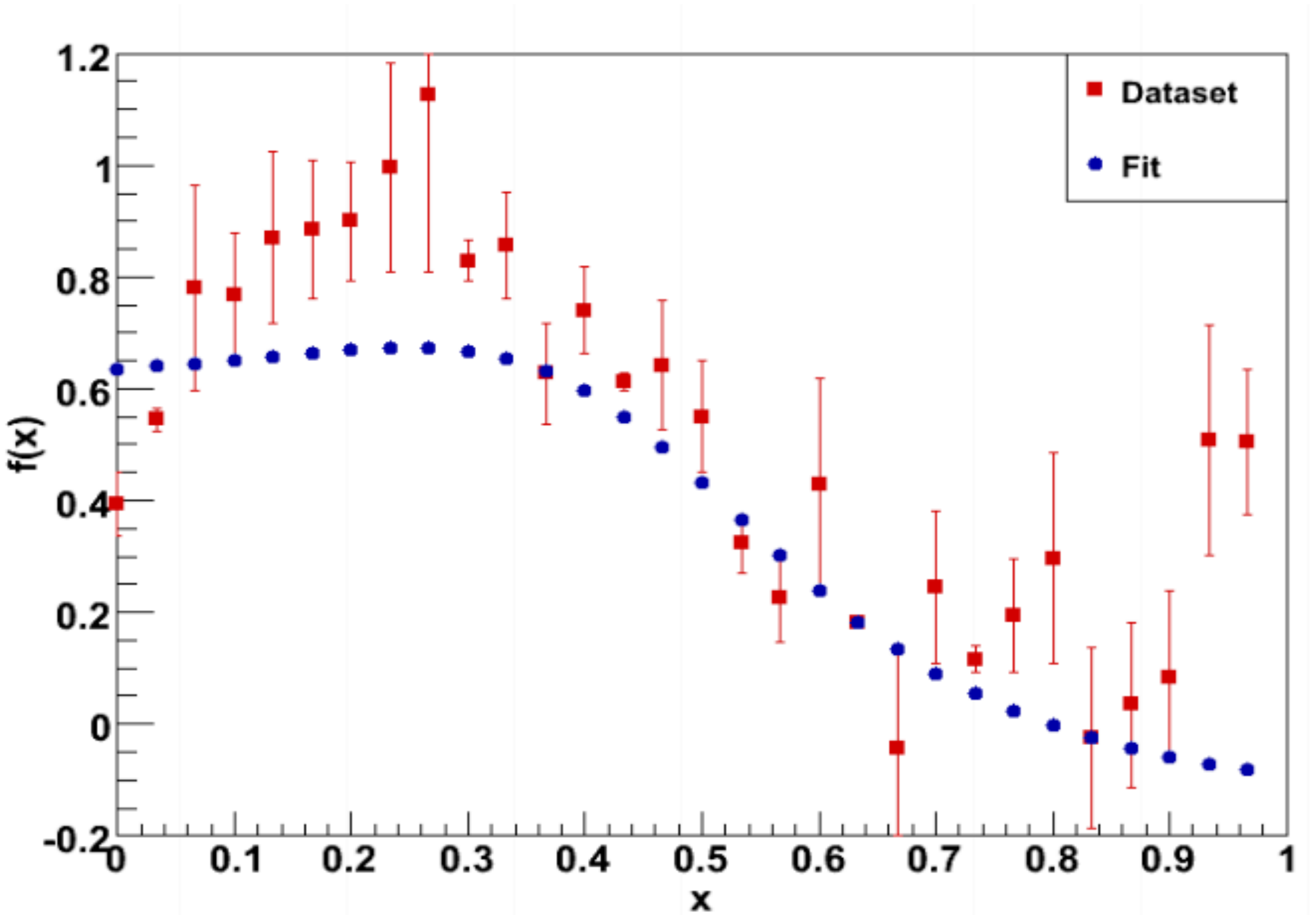


A toy model

A toy-model:

2) Choose a parametrisation for f(x) and perform a fit by minimising a loss function, a figure of merit, like the χ^2

$$\chi^2 = \sum_{i=1}^{N_{\mathrm{data}}} \frac{(D_i - T_i)^2}{\sigma_i^2}$$



χ^2/d.o.f. ≫ 1

We are not quite there…

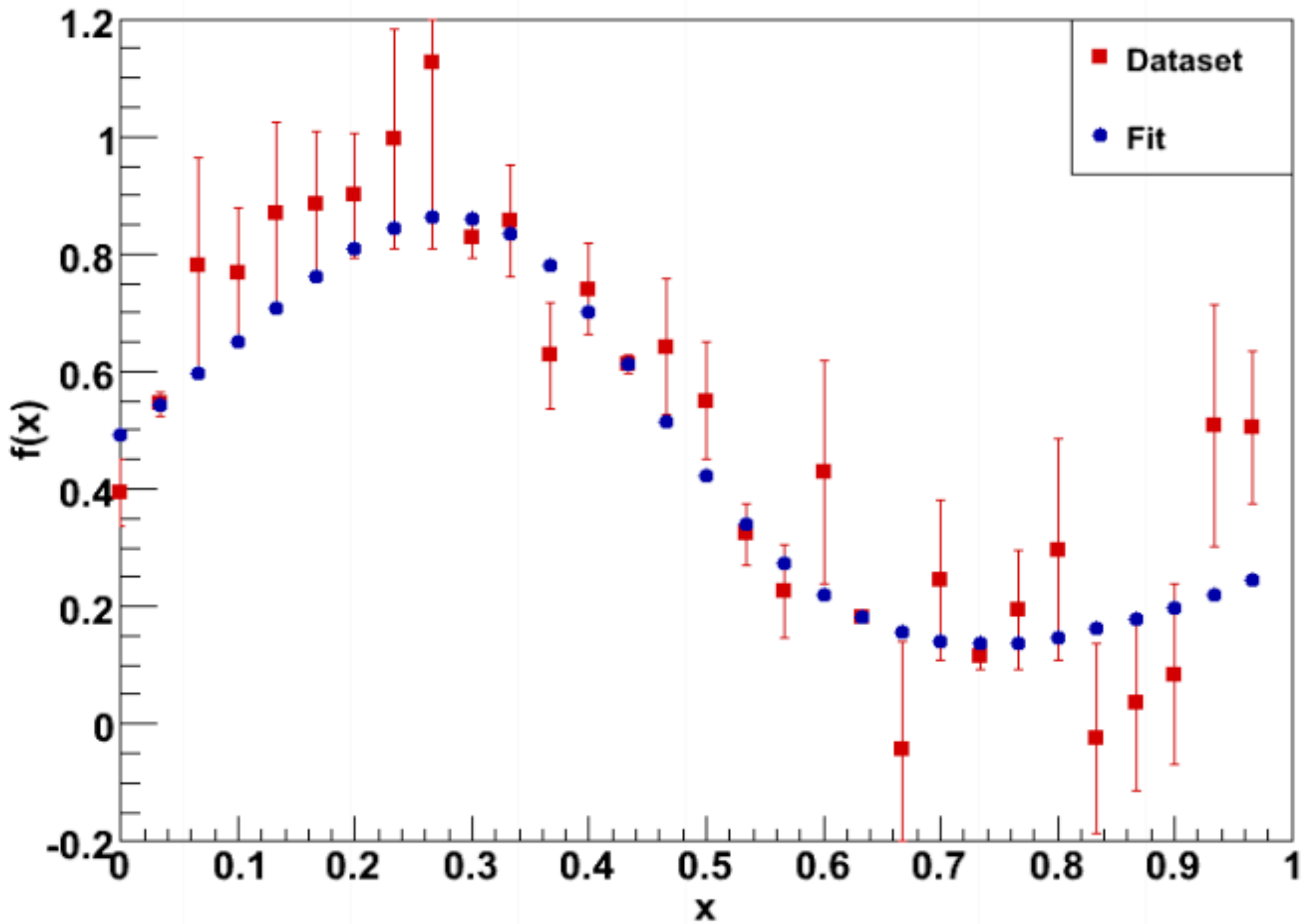
under-learning

A toy model

A toy-model:

2) Choose a parametrisation for f(x) and perform a fit by minimising a loss function, a figure of merit, like the χ^2

$$\chi^2 = \sum_{i=1}^{N_{\rm data}} \frac{(D_i - T_i)^2}{\sigma_i^2}$$



χ^2/d.o.f. ~ 1

We are there…

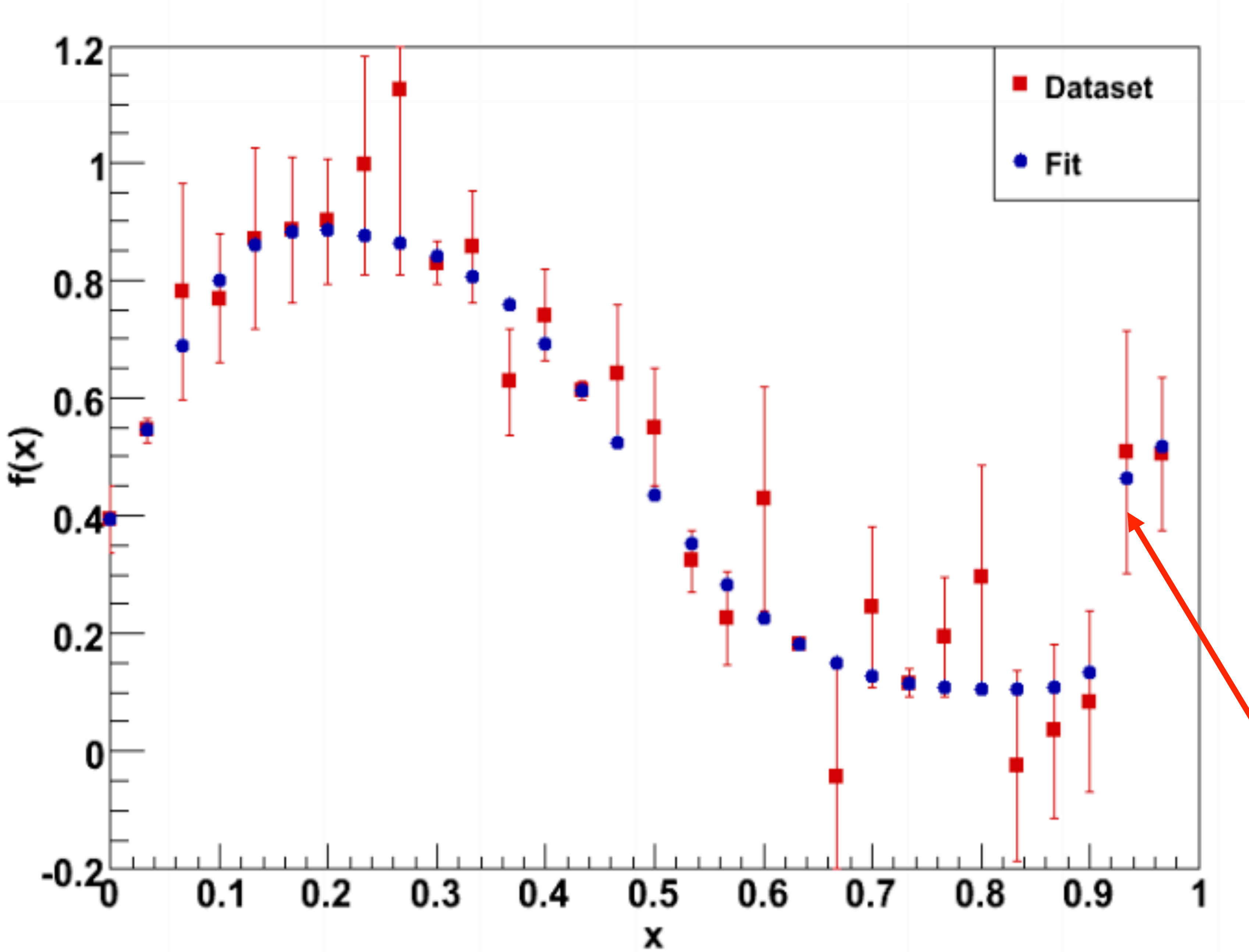
proper learning

A toy model

A toy-model:

2) Choose a parametrisation for f(x) and perform a fit by minimising a loss function, a figure of merit, like the χ^2

$$\chi^2 = \sum_{i=1}^{N_{\rm data}} \frac{(D_i - T_i)^2}{\sigma_i^2}$$



χ^2/d.o.f. → 0

We went too far…

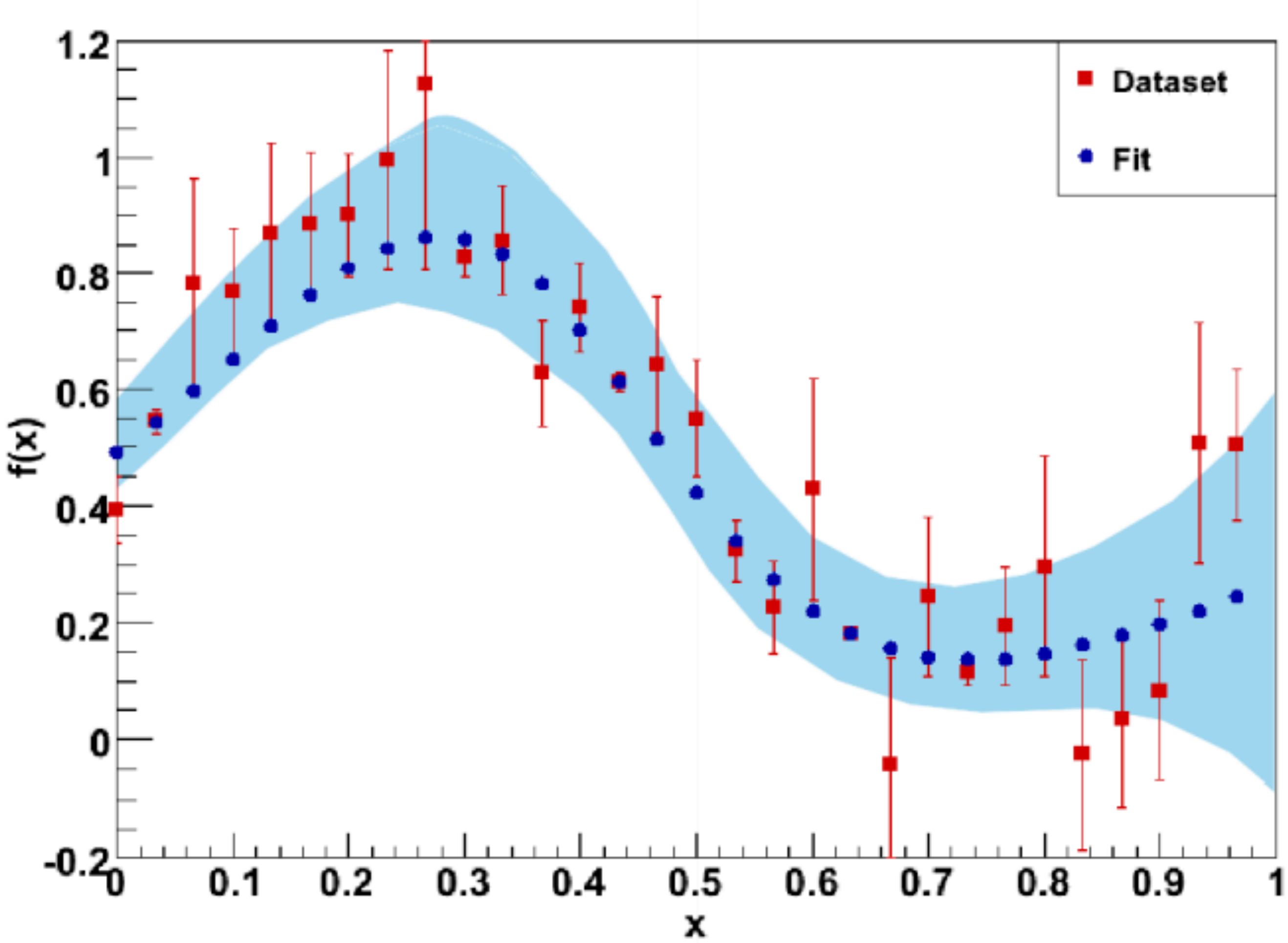
over-learning

start fitting the statistical noise

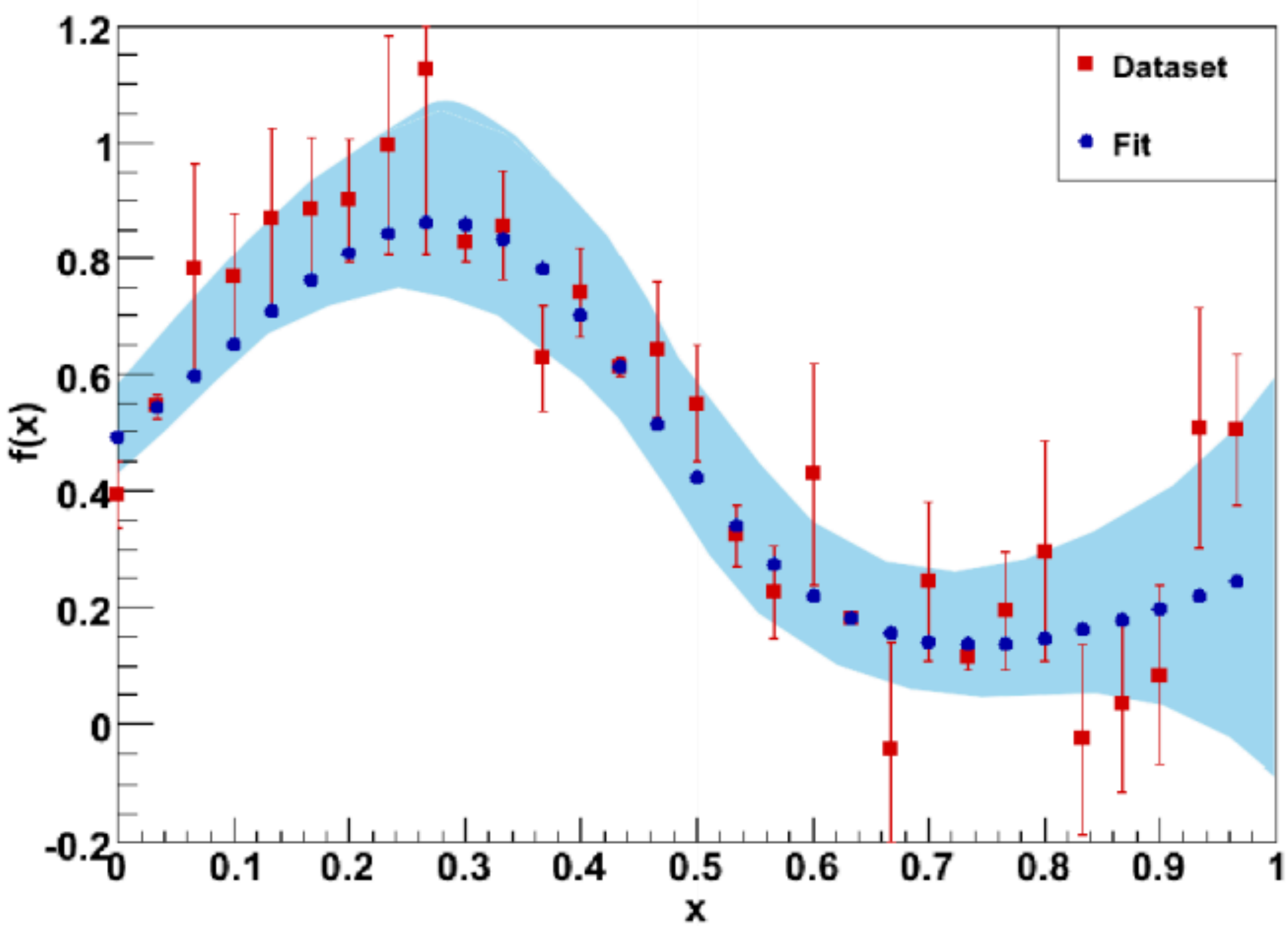
A toy model

A toy-model:

3) Determine the error of our fit, which corresponds to the lack of information that the data provide. In the limit of infinite and infinitely-precise and compatible data, the error band tends to 0



The actual game



The actual games is more complicated since we have 6+6+1 functions (actually 3+3+1+1) and errors to determine which are not directly measured. They enter in the measured observables according to different combinations. But still…

- ✓ Need to choose a clever and flexible parametrisation
- ✓ Need a way to stop the fit before over-learning sets in to avoid fitting statistical noise
- ✓ Need a reliable error estimate

Choice of parametrisation

Usually one parametrises independently the gluon, light quarks and anti-quarks, strange and anti-strange (+ intrinsic charm), while heavy quarks are generated perturbatively from light quarks and gluons*

The ideal parametrisation

- Too rigid

Global fit might not have flexibility to describe data or inadequate small uncertainties where there are no data

- Too flexible

Difficult minimisation and it might develop artefacts driven by statistical fluctuations of the data

Theory constraints

- Integrability, positivity of observables, must go to 0 as x → 1

- From baryon number conservation → Valence Sum Rules

$$\int_0^1 dx\,(u(x,Q^2) - \bar{u}(x,Q^2)) = 2$$

$$\int_0^1 dx\,(d(x,Q^2) - \bar{d}(x,Q^2)) = 1$$

$$\int_0^1 dx\,(s(x,Q^2) - \bar{s}(x,Q^2)) = 0$$

- From momentum conservation → Momentum Sum Rule

$$\int_0^1 dx\,(x\Sigma(x,Q^2) + xg(x,Q^2)) = 1$$

$$\text{with } \Sigma = \sum_{i=1}^{n_F} q_i + \bar{q}_i$$

Traditional (parametrical) approach

- Introduce a simple functional form with enough free parameters

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

- Typically about 20-25 free parameters for 7 independent functions

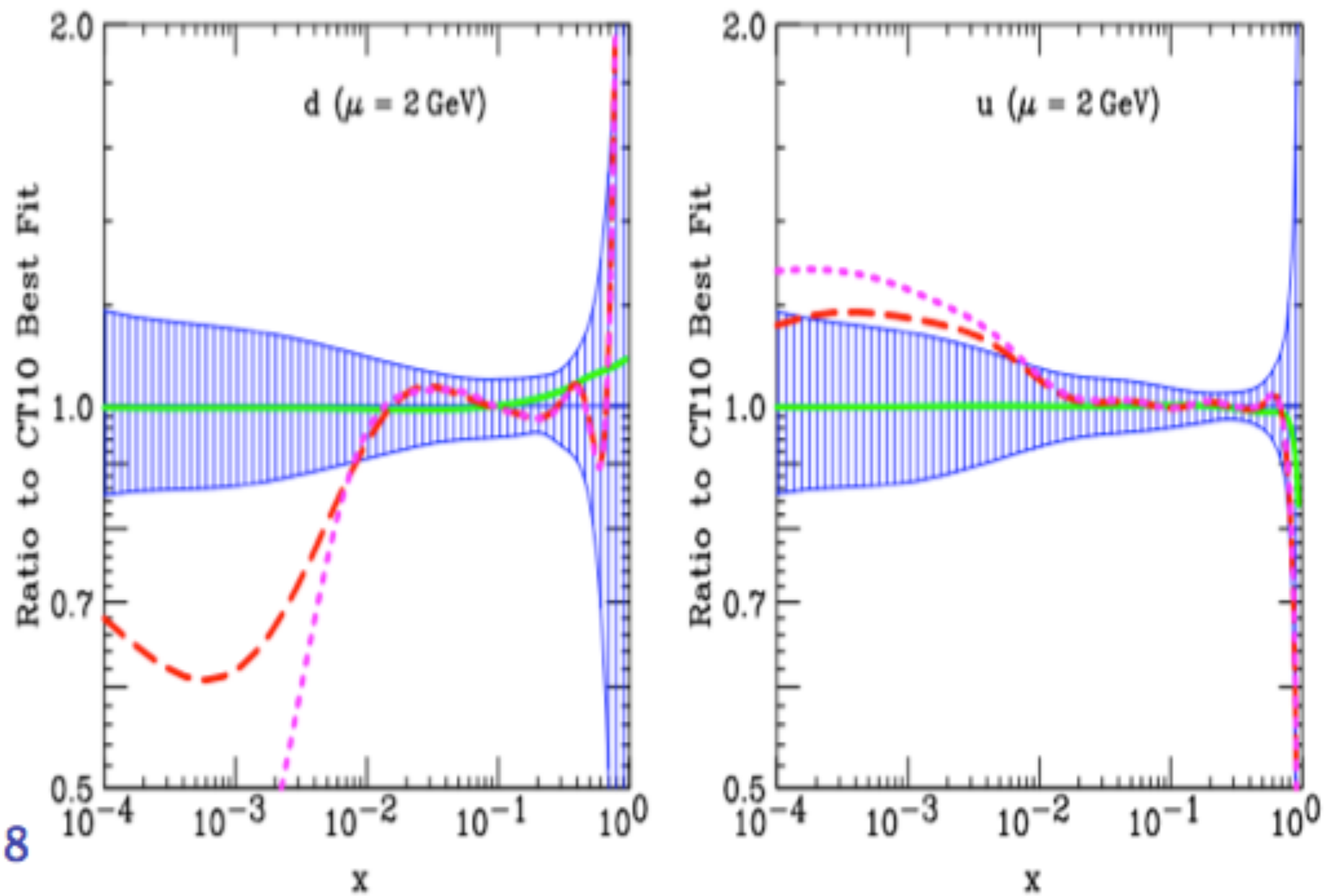
MSTW2008

20 free parameters

$$\begin{aligned}
xu_v(x, Q_0^2) &= A_u\, x^{\eta_1}(1-x)^{\eta_2}(1+\epsilon_u\sqrt{x}+\gamma_u\, x), \\
xd_v(x, Q_0^2) &= A_d\, x^{\eta_3}(1-x)^{\eta_4}(1+\epsilon_d\sqrt{x}+\gamma_d\, x), \\
xS(x, Q_0^2) &= A_S\, x^{\delta_S}(1-x)^{\eta_S}(1+\epsilon_S\sqrt{x}+\gamma_S\, x), \\
x\Delta(x, Q_0^2) &= A_\Delta\, x^{\eta_\Delta}(1-x)^{\eta_S+2}(1+\gamma_\Delta\, x+\delta_\Delta\, x^2), \\
xg(x, Q_0^2) &= A_g\, x^{\delta_g}(1-x)^{\eta_g}(1+\epsilon_g\sqrt{x}+\gamma_g\, x) + A_{g'}\, x^{\delta_{g'}}(1-x)^{\eta_{g'}}, \\
x(s+\bar{s})(x, Q_0^2) &= A_+\, x^{\delta_S}\,(1-x)^{\eta_+}(1+\epsilon_S\sqrt{x}+\gamma_S\, x), \\
x(s-\bar{s})(x, Q_0^2) &= A_-\, x^{\delta_-}(1-x)^{\eta_-}(1-x/x_0),
\end{aligned}$$

Traditional (parametrical) approach

- Possible issues:
  What is the error associated to a given functional form?

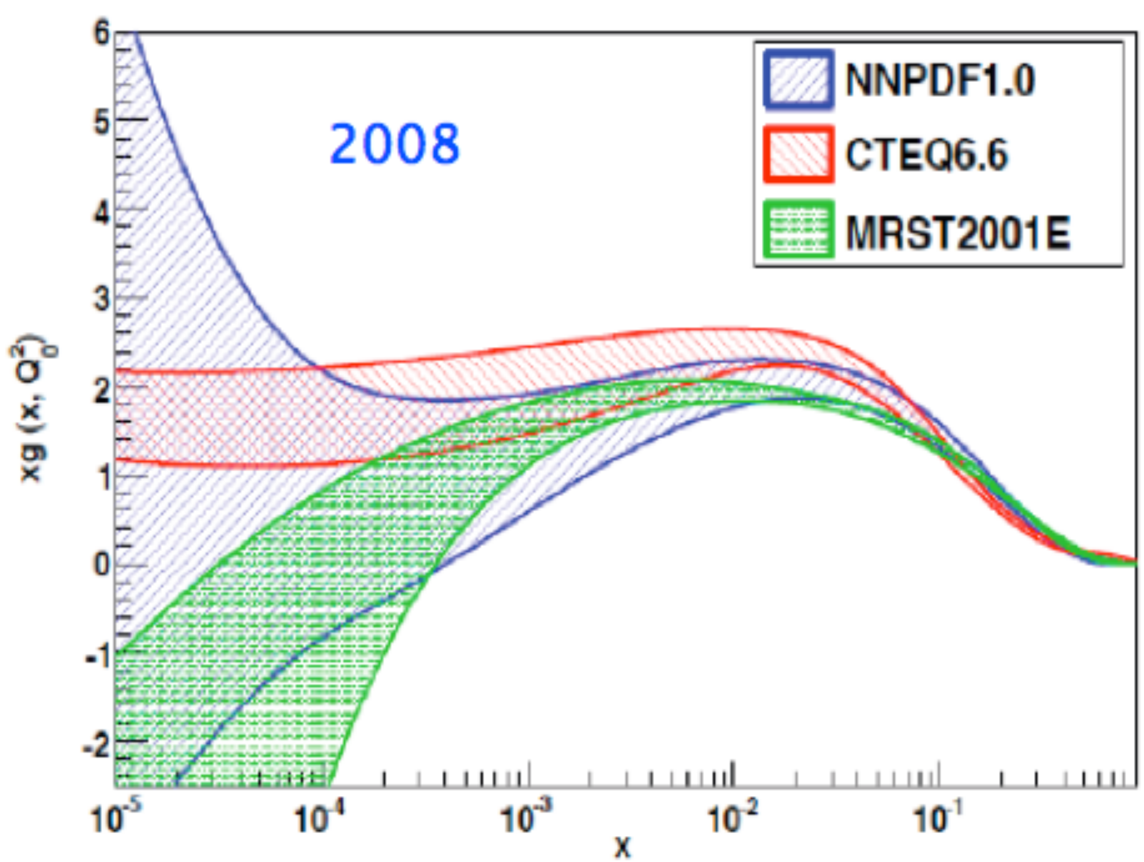


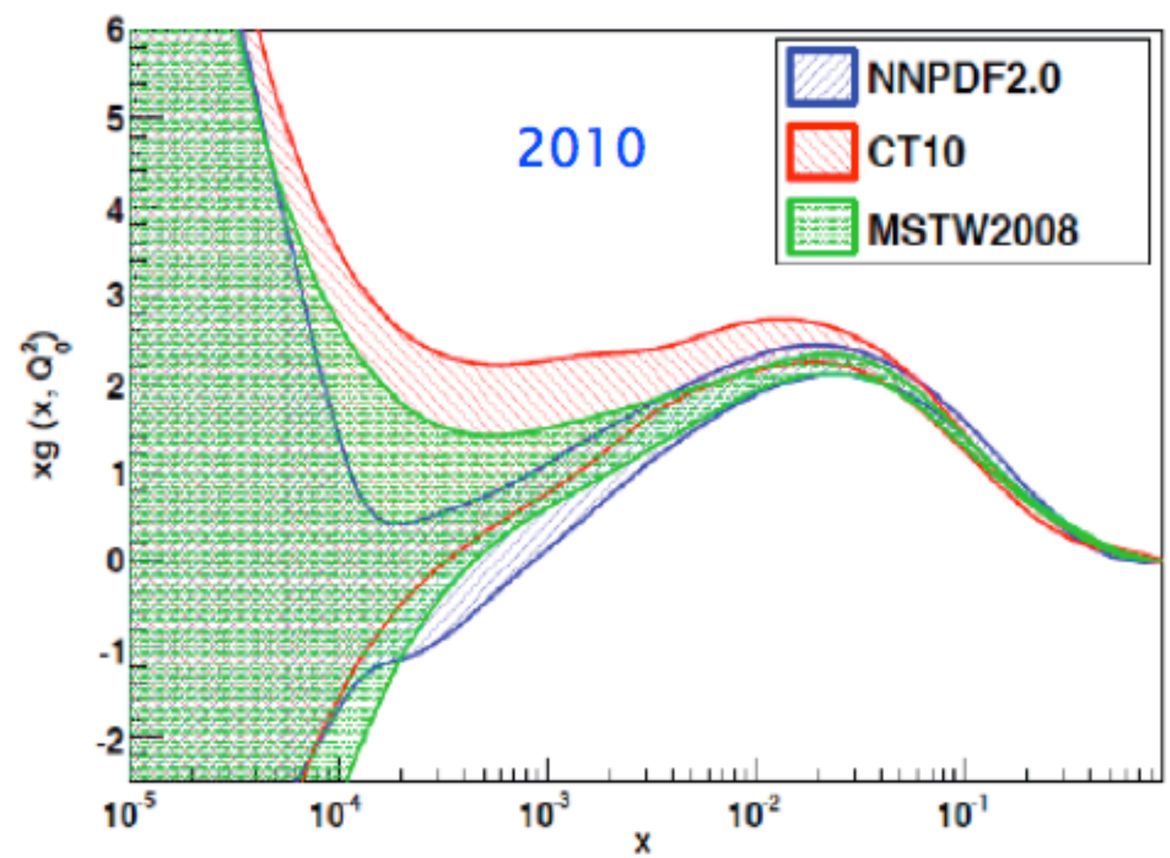
Pink and red curves give same good description of data but outside error bar

J. Pumplin ArXiv:0909.0268

Traditional (parametrical) approach

- Possible issues:
  If functional form not flexible enough PDFs may present unrealistically small errors where data do not constrain PDF uncertainties

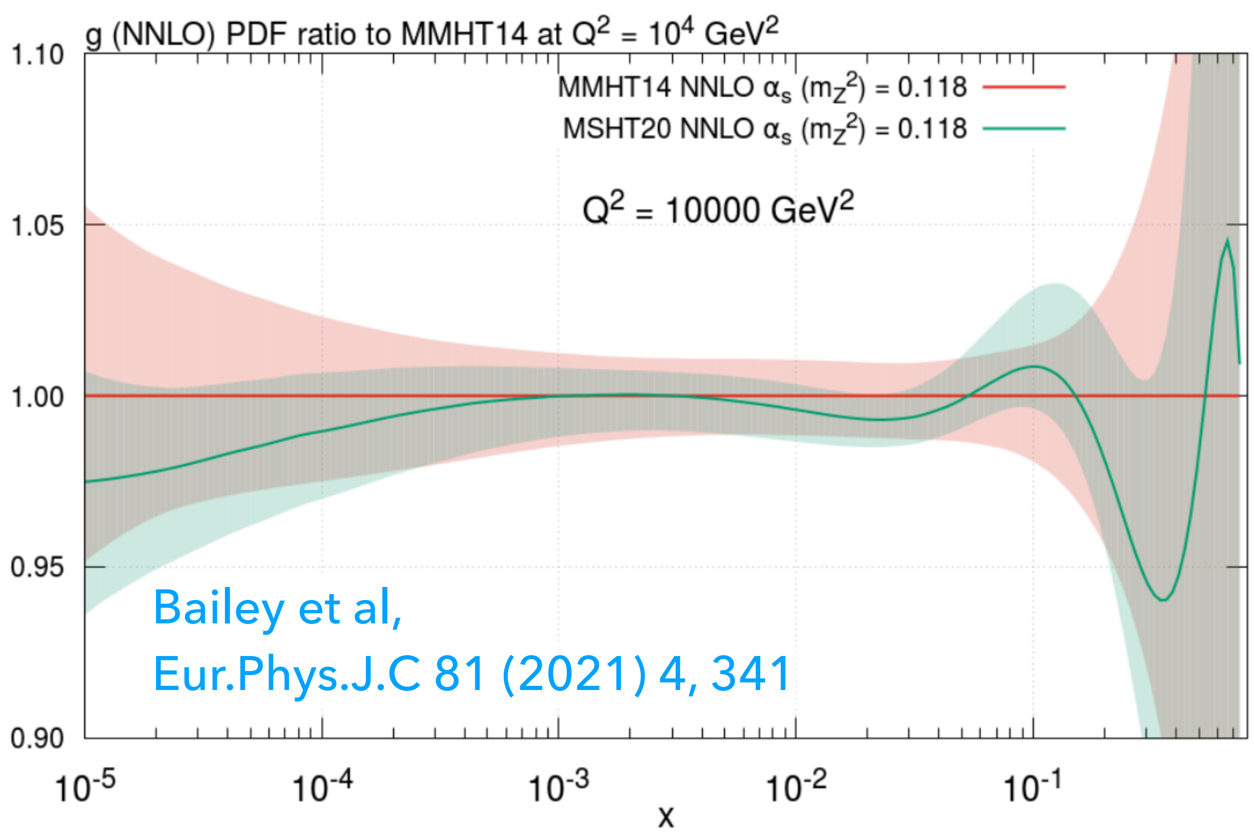


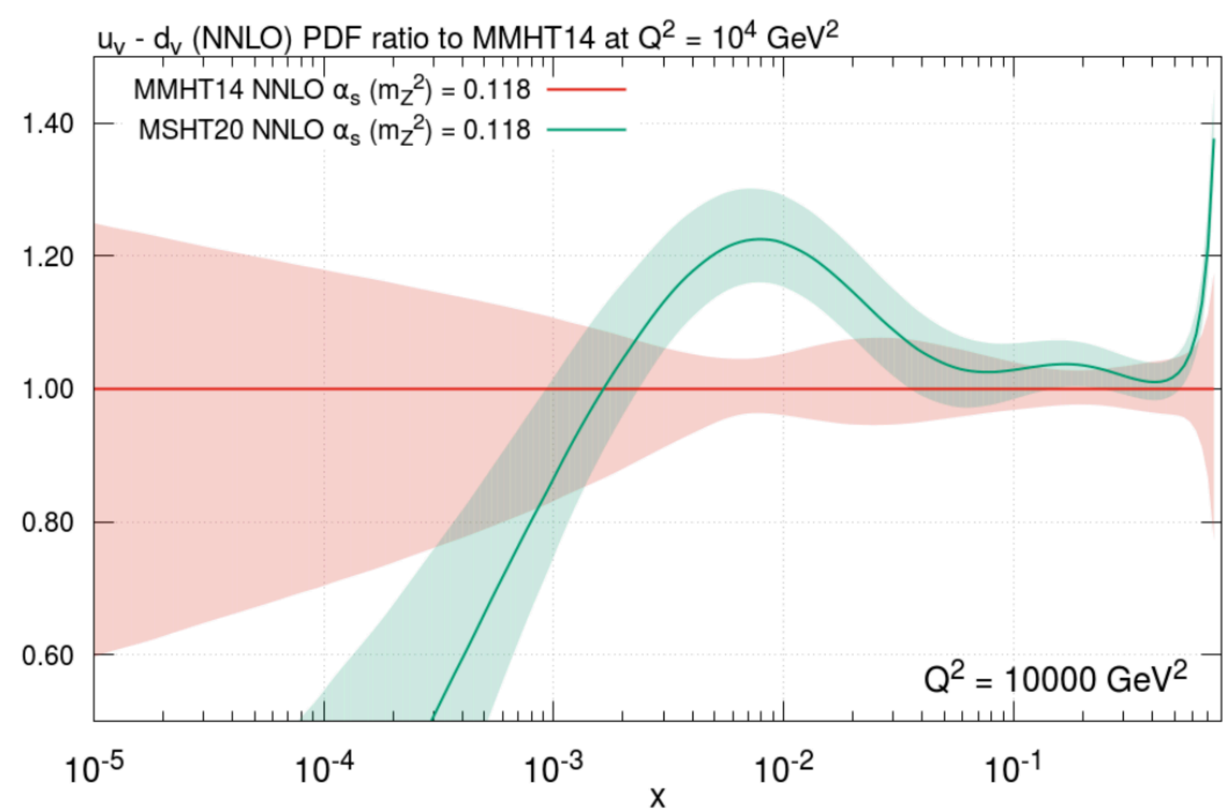


$$xg = A_g\, x^{\delta_g}(1-x)^{\eta_g}(1+\epsilon_g\sqrt{x}+\gamma_g\, x) + A_{g'}\, x^{\delta_{g'}}(1-x)^{\eta_{g'}}$$

Traditional (parametrical) approach

- Possible issues:
  If functional form not flexible enough PDFs may be not able to adapt to new data

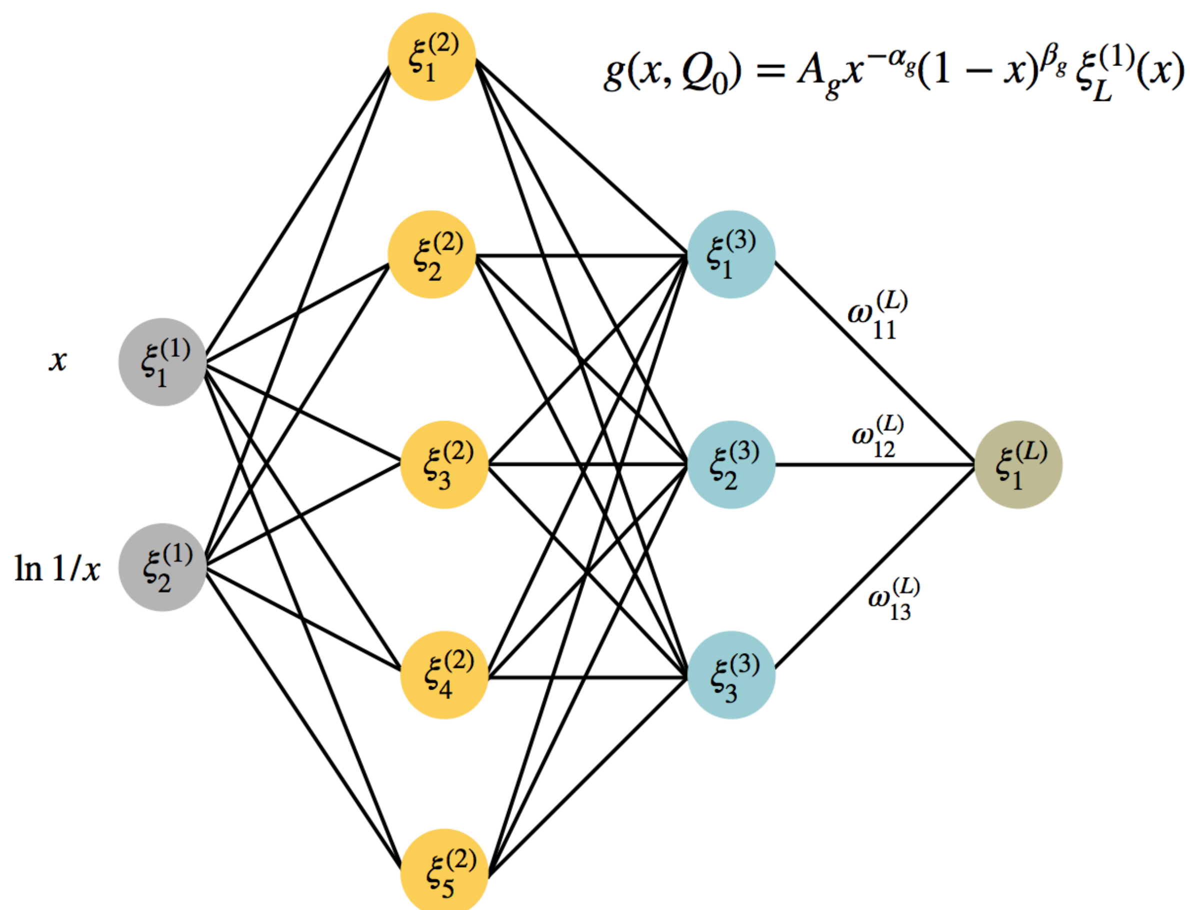




- In recent updates from a global PDF fitting collaborations (MSHT20) the effect of LHC data required big change in the parametrization which makes PDF uncertainty increase (data-driven parametrization)

Neural network parametrisation

Fully connected multi-layer perceptron



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

- NNPDF (2006) A 2-5-3-1 Neural network associated to each independent PDF (gluon, up, anti-up, down, anti-down, strange, anti-strange and charm)

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

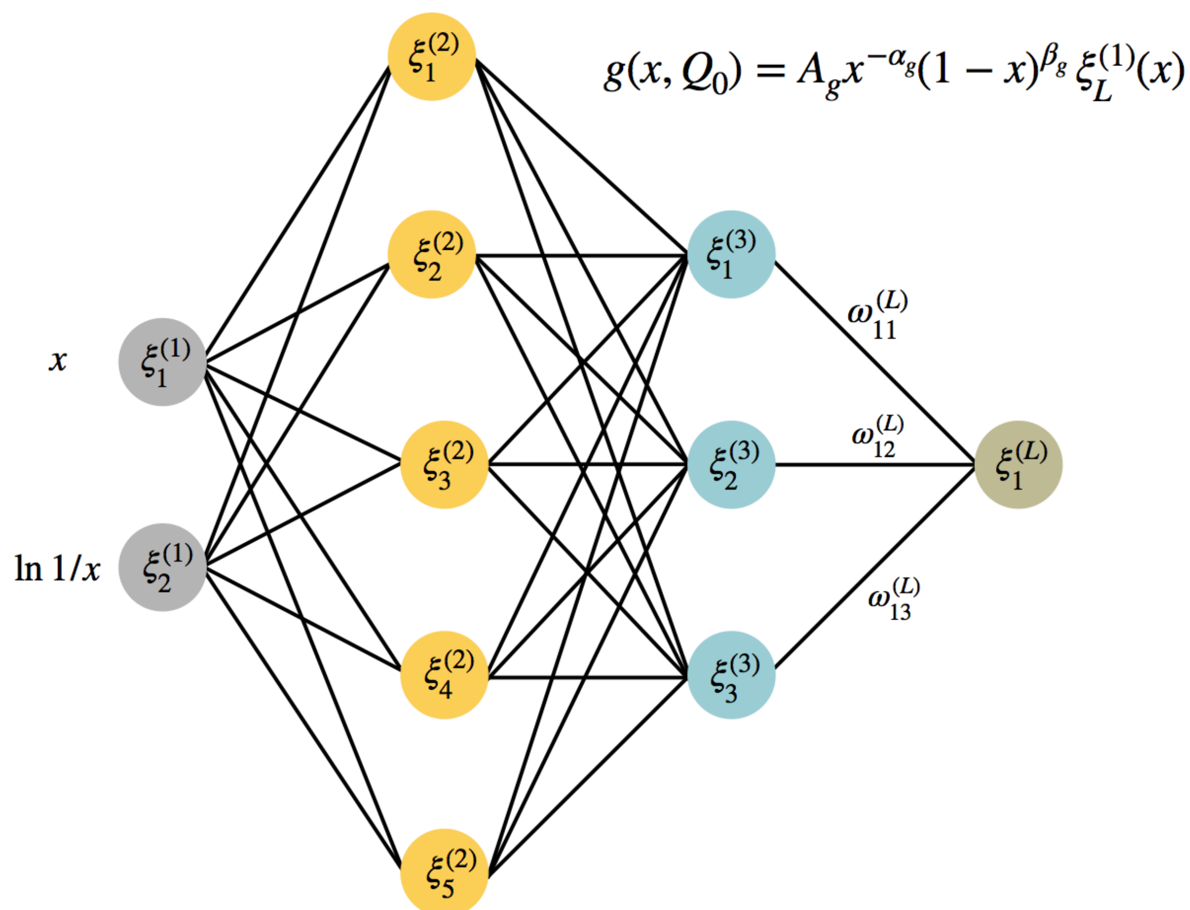
$$\xi_1^{(3)}(\xi_1^{(1)}) = \frac{1}{1 + e^{\theta_1^{(3)} - \frac{\omega_{11}^{(2)}}{1+e^{\theta_1^{(2)} - \xi_1^{(1)}\omega_{11}^{(1)}}} - \frac{\omega_{12}^{(2)}}{1+e^{\theta_2^{(2)} - \xi_1^{(1)}\omega_{21}^{(1)}}}}}$$

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

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

Neural network training

Fully connected multi-layer perceptron



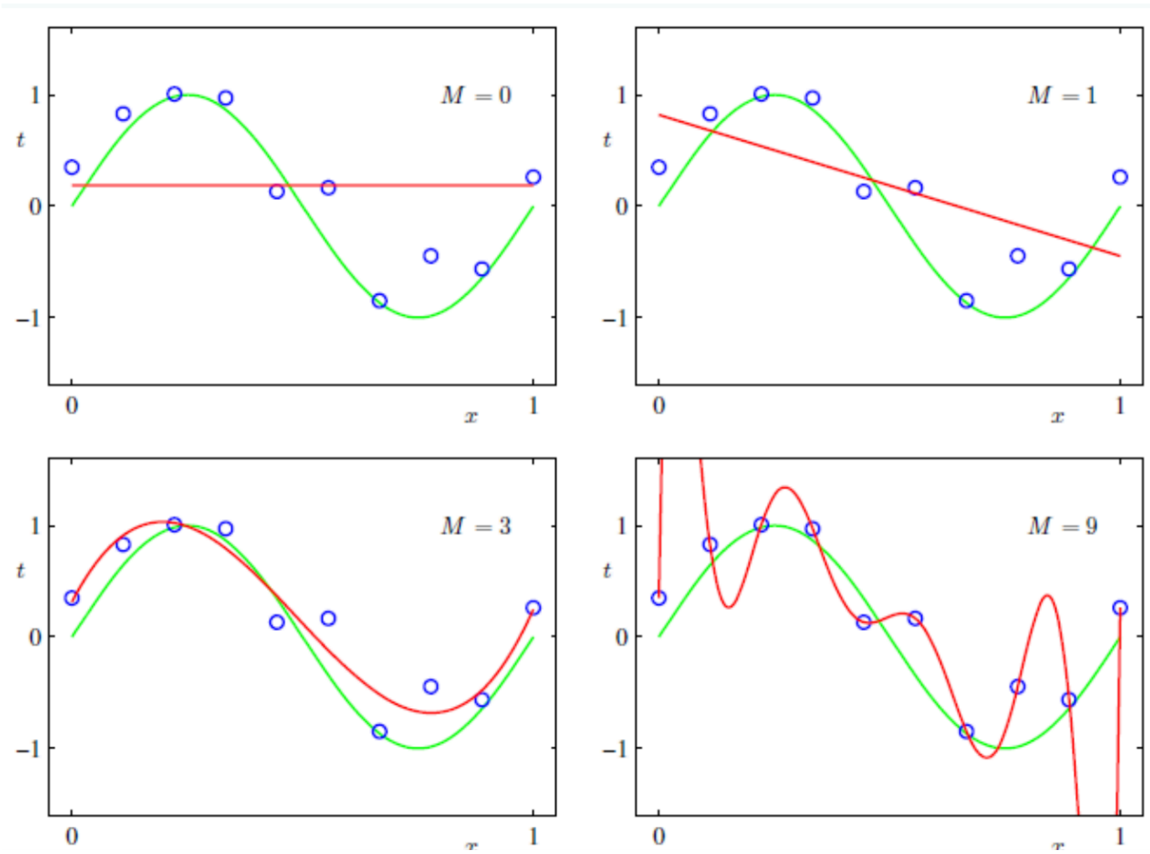
How do we train the 7(8) independent NN?

Minimise the loss function:

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

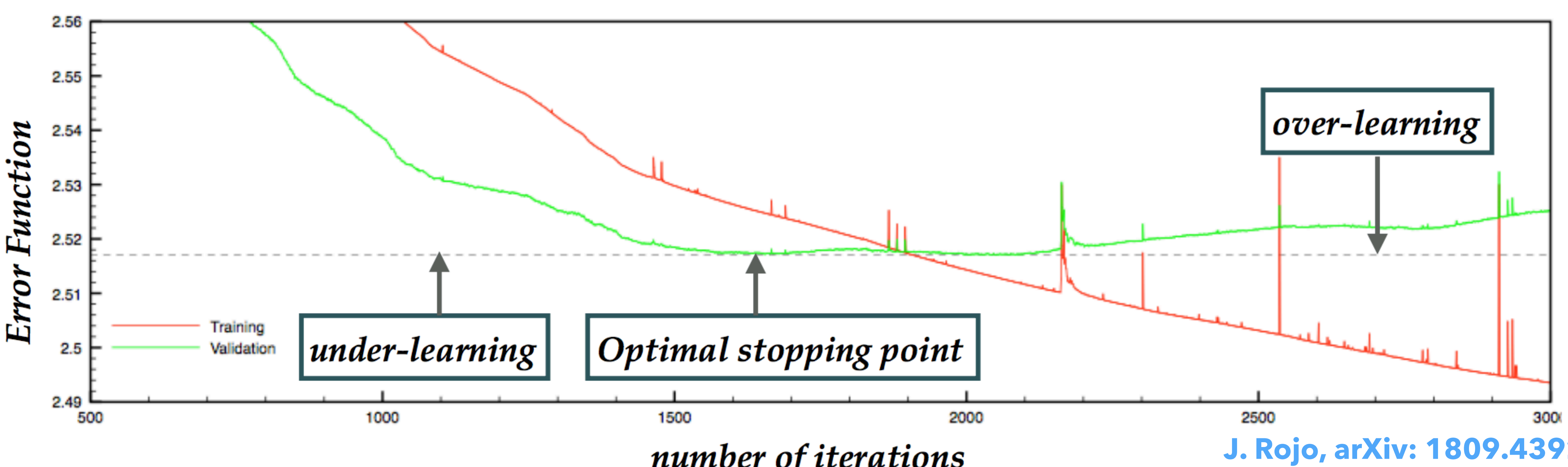
- D_i experimental measurement for the point i
- T_i theoretical prediction for the point i (depending on PDF parameters $\sigma_{DIS} = \sigma \otimes f_1$ - $\sigma_h = \sigma_{12} \otimes f_1 \otimes f_2$)
- $(\mathrm{cov})_{ij}$ is the covariance matrix between point i and j with corrections for normalisation uncertainties
- Supplemented by additional penalty for positive observables

Neural network training



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

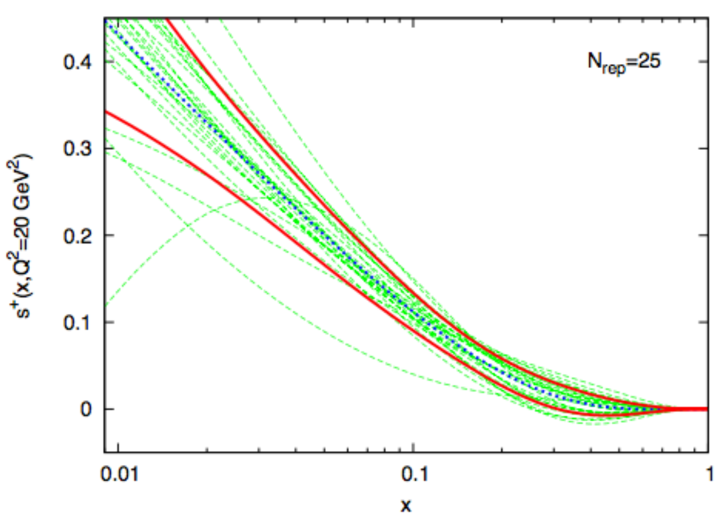
- Redundant parametrization: risk of over-fitting. Cross-validation necessary.



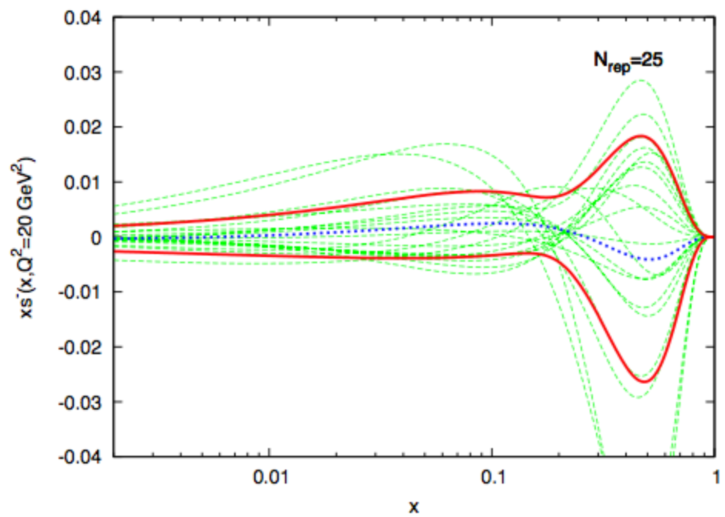
J. Rojo, arXiv: 1809.4392

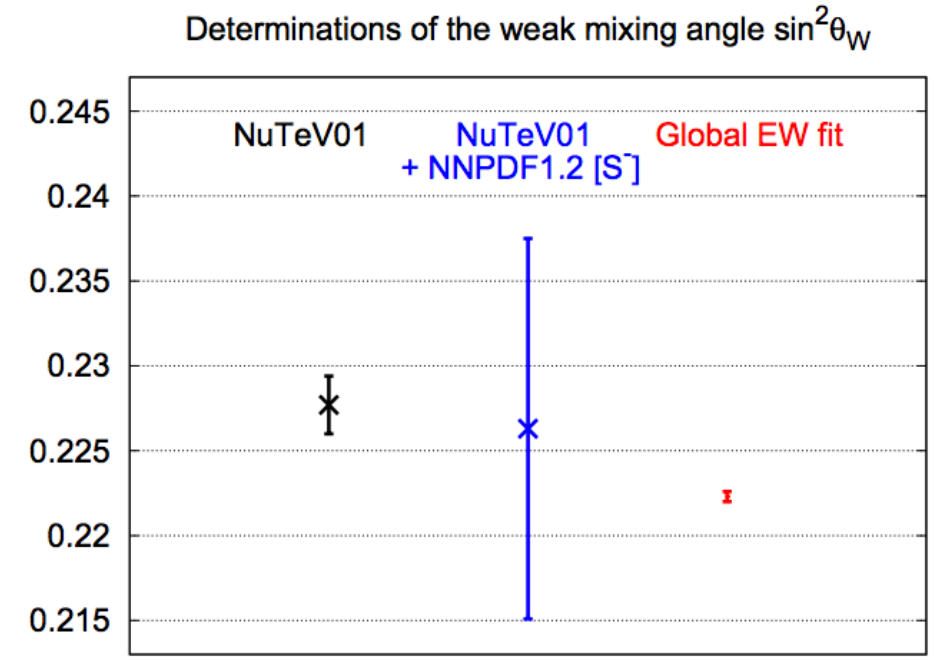
E.g. the NuTeV anomaly

Total strangeness ↓



Strange valence ↓





EW fit

$\sin^2\theta_W = 0.2223 \pm 0.0002$

NuTeV

$\sin^2\theta_W = 0.2276 \pm 0.0014$

$$\sin^2\theta_W\Big|_{\rm NuTeV} - \sin^2\theta_W\Big|_{\rm EW} = 0.0053$$

$$[F] = \int_0^1 dx\, x\, f(x, Q^2).$$

- $>3\sigma$ discrepancy between EW fits and NuTeV measurements
- Unbiased parametrisation of strangeness (2010) solved NuTeV anomaly

$$\delta_s \sin^2\theta_W \sim -0.240\frac{[S^-]}{[Q^-]}$$

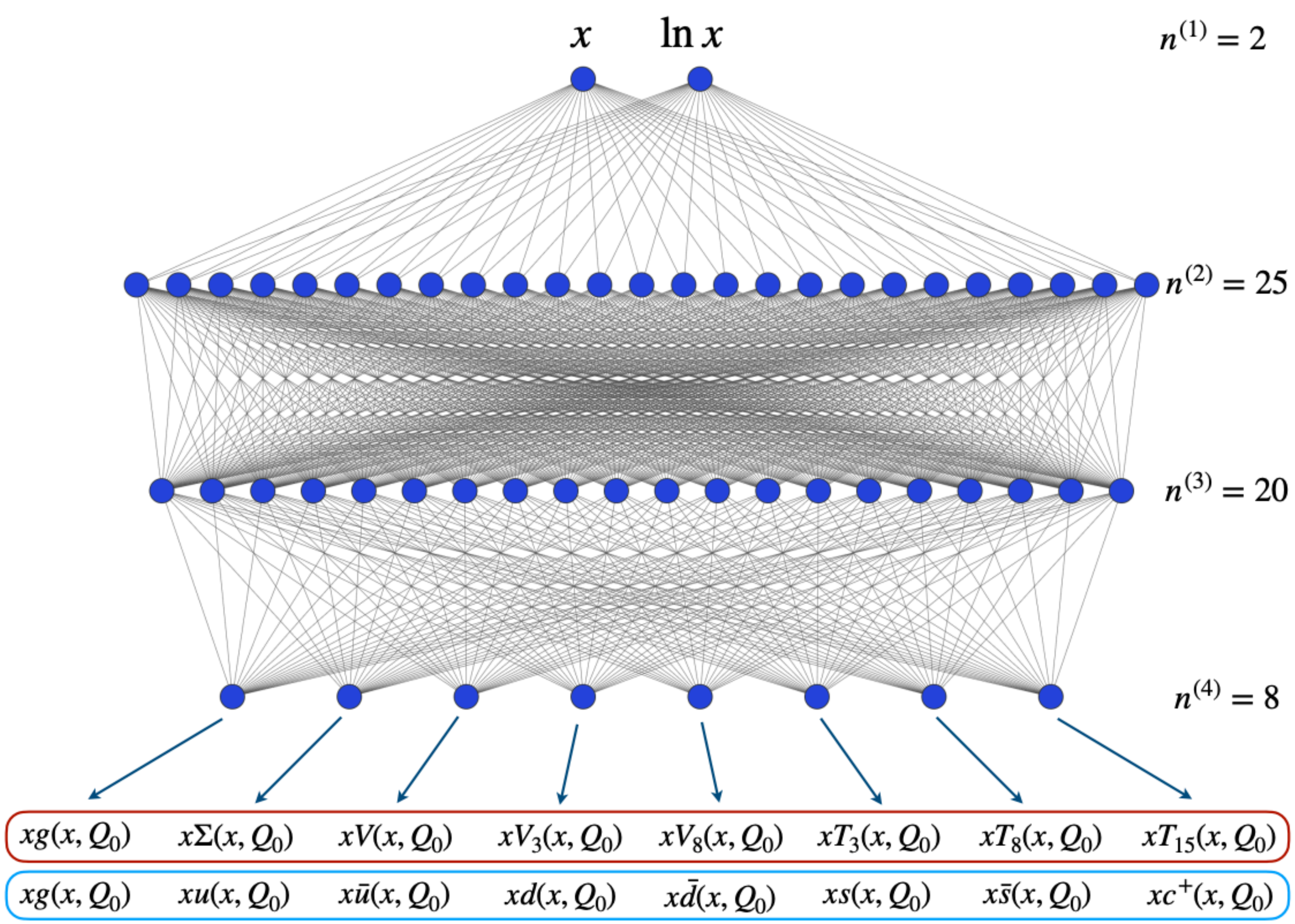
$$\delta_s \sin^2\theta_W = -0.0005 \pm 0.0096^{\rm PDFs} \pm sys$$

Ball et al, 0906.1958

A deep-learning based fit

- NNPDF4.0: updated methodology: single neural network to parametrise 8 independent PDF combinations (g, u, d, s, u~, d~, s~, c=c~)
- New optimisation strategy based on gradient descent rather than genetic algorithm
- Hyper-optimised methodology: scan of the hyper parameter space to find optimal minimisation settings (optimiser, initialiser, stopping patience, number of layers, learning rate, epochs, activation function) by using K-fold procedure

[Carrazza et al, Eur.Phys.J.C 79 (2019) 8, 676]



NNPDF4.0, arXiv: 2109.02653

Error propagation

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

- 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

Option a) Project into a n-dimensional space of parameters which parametrise PDFs and use linear approximation around minimum χ^2

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

Hessian Method

Option b) Choose a parametrisation and perform a Monte Carlo sampling of probability density in functional space

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

Monte Carlo Method

Hessian method

➡ Used by most PDF fitters (CTEQ/TEA, MSTW/MMHT, HERAPDF, ABMP

➡ Determine best fit values of parameters $\{\vec{a}_0\}$

➡ Shift $\vec{a} \rightarrow \vec{a} - \vec{a}_0$

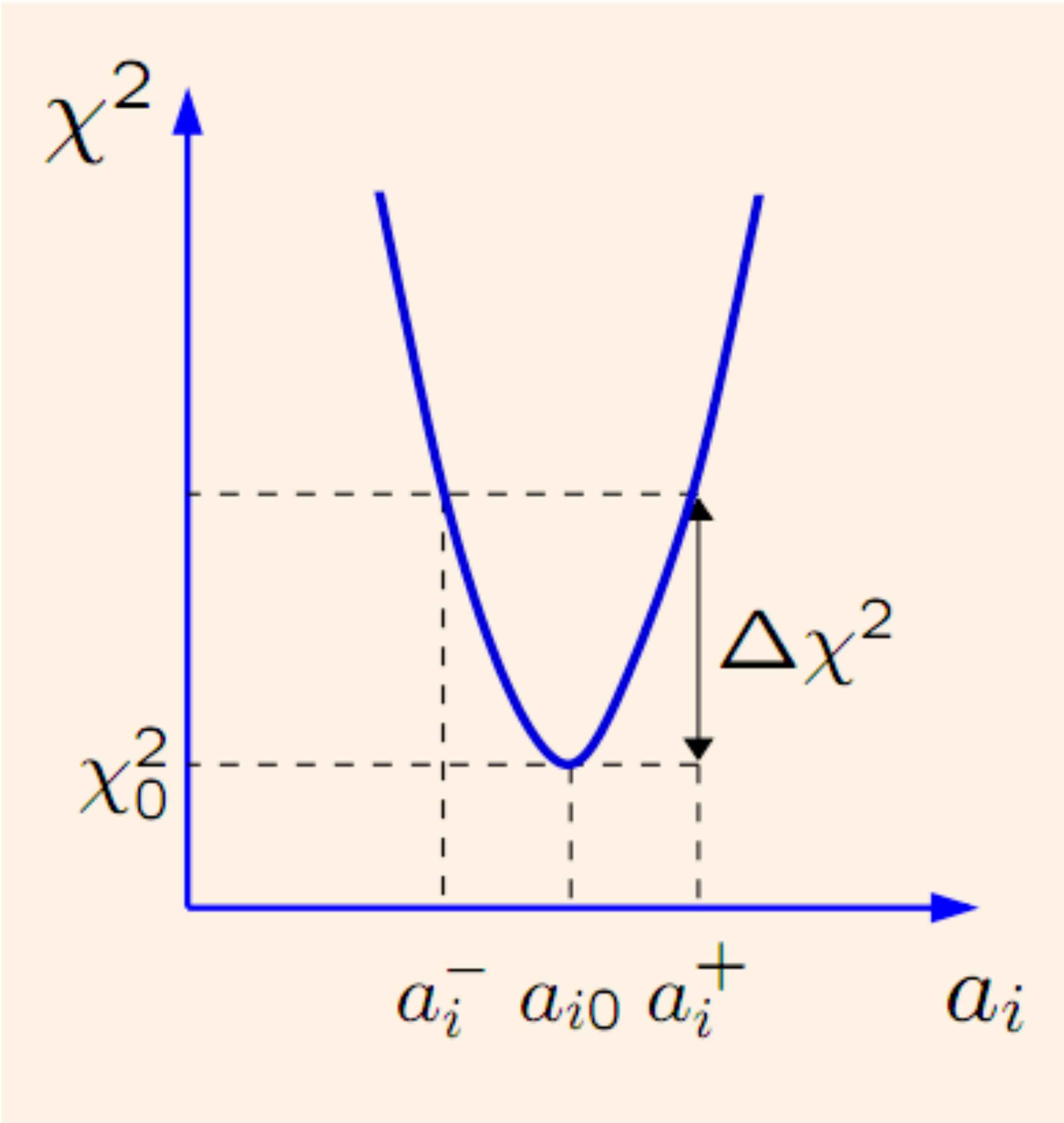
➡ Determine error on PDFs and any observable depending on PDFs (all denoted by X) by propagation of the error in the parameter space

Assuming linear prop: $X(\vec{a}) \simeq X(\vec{0}) + a_i \partial_i X(\vec{a})\big|_{\vec{a}=\vec{0}}$

Variance: $\sigma_X^2 = (\mathrm{cov})_{ij}\, \partial_i X \partial_j X$ $(\mathrm{cov})_{IJ}$ covariance matrix in param, space

Maximum likelihood: $(\mathrm{cov})_{ij} = (H)_{ij} = \left.\dfrac{\partial^2 \chi^2(\vec{a})}{\partial_i a \partial_j a}\right|_{\vec{a}=\vec{0}}$ cov $\Longleftrightarrow$ Hessian at the minimum of χ^2

Hessian method



P. Nadolsky, CTEQ summer school 2009

- According to textbook statistics, the 1σ contour in parameter space is given by

$$\Delta\chi^2 = 1$$

- Projection of the radius one sphere would give the uncertainty on parameters and on the PDFs, observables...
- The textbook statistics should work in case of perfectly compatible Gaussian errors
- But in practice, for global fits a tolerance is introduced T
- NB: introducing a tolerance corresponds to blow up uncertainties by a factor $\sqrt{\Delta\chi^2}$

Monte Carlo method

- ➡ First idea by Giele Keller Kosover (hep-ph/0104052)
- ➡ Monte Carlo in parameter space

$$X(\vec{a})$$

MC sampling

$$\langle X \rangle = \int d\vec{a} X[\vec{a}]\mathcal{P}[\vec{a}]$$

P probability of parameter values

MC sampling in **parameter** space

$$\langle X \rangle \sim \frac{1}{N_{\mathrm{rep}}} \sum_{i=1}^{N_{\mathrm{rep}}} X(\vec{a}_i)$$

$$\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2$$

Problem

How many replicas are needed?

Three bins per parameter ⇒ 3^{Npar} bins

E.g. for 23 parameters need more than 10^{11} replicas!!!

Monte Carlo method

- ➡ Forte, J. I Latorre, Piccione (hep-ph/0701127)
- ➡ First applied to structure functions then to PDFs

$$X(\vec{a})$$

MC sampling ➡

$$\langle X \rangle = \int d\vec{a} X[\vec{a}] \mathcal{P}[\vec{a}]$$

P probability of parameter values

MC sampling in **data** space

$$\langle X \rangle \sim \frac{1}{N_{\rm rep}} \sum_{i=1}^{N_{\rm rep}} X(\vec{a}_i)$$

$$\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2$$

Idea

Choose parameters along ∇X ⟺ Choose replicas of the data, i.e. work in the space of data and project back into PDF space

How many replicas does one need? 1-dim average of N_{rep} converges to true average with standard deviation $\sigma/\sqrt{N_{rep}}$

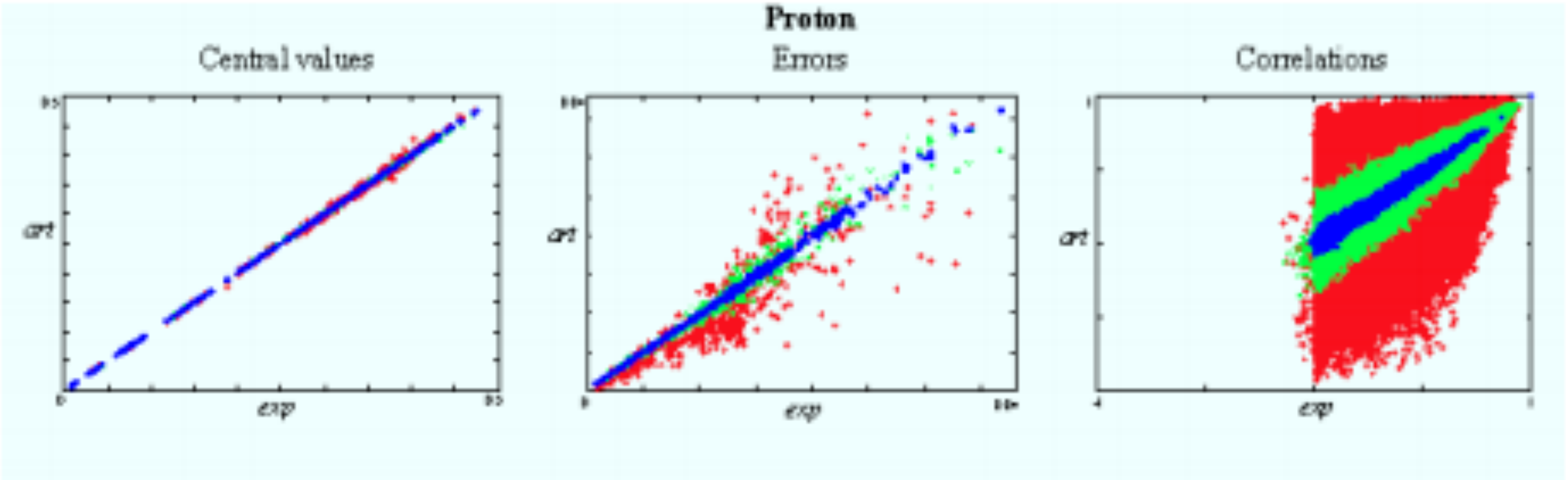
E.g. 10 replicas are enough for getting “true” central value with σ/3 accuracy

Monte Carlo method

- Generate artificial data according to distribution

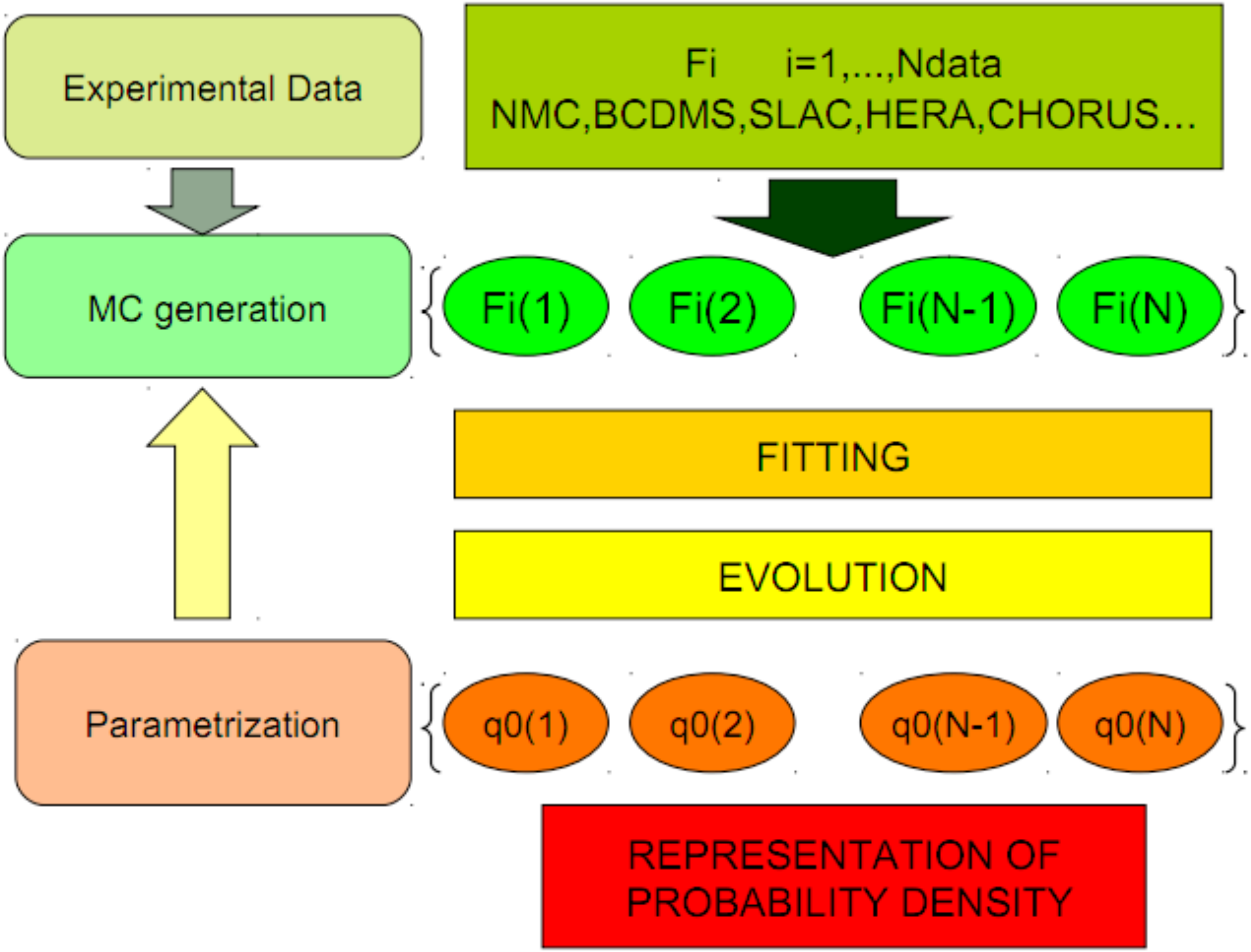
$$F_p^{(\mathrm{art})(k)} = S_{p,N}^{(k)} F_p^{(exp)} \left(1 + \sum_{l=1}^{N_c} r_{p,l}^{(k)} \sigma_{p,l} + r_p^{(k)} \sigma_{p,s} \right)$$

- r_i are univariate Gaussian random numbers such that if two points have correlated systematic uncertainties, they oscillate in the same directions
- S normalisation factors
- Validate Monte Carlo replicas against experimental data

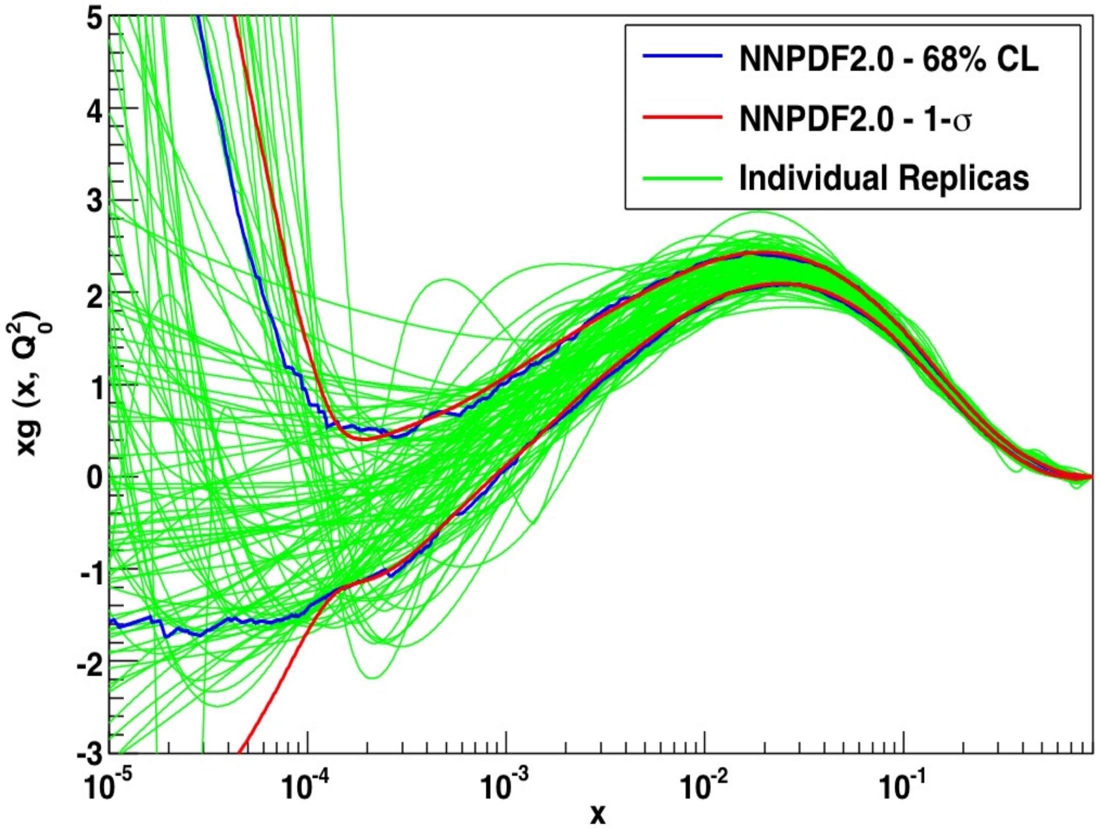


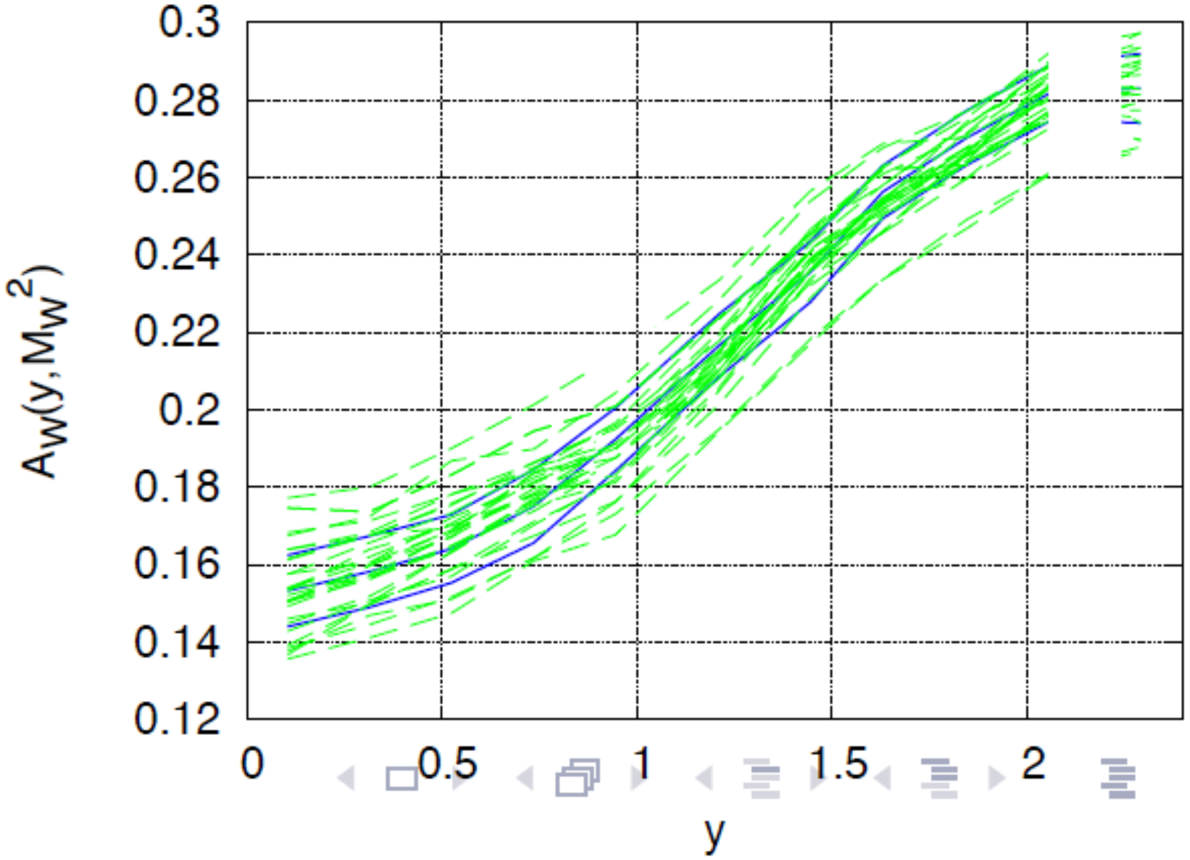
- Convergence rate increases with N_{rep}
- Correlations reproduced to % accuracy with 1000 reps

Monte Carlo method



Monte Carlo method





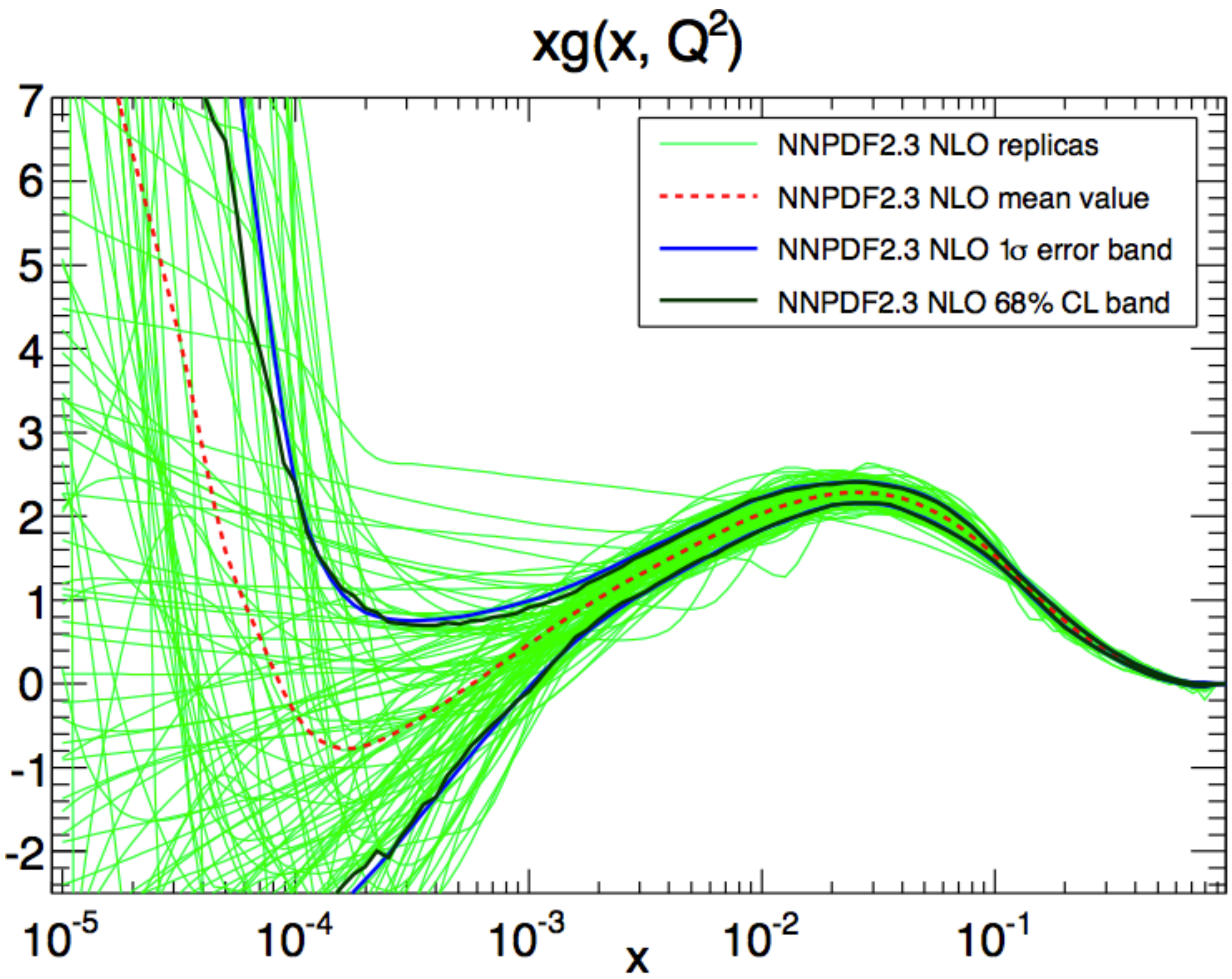
$$\langle f_J \rangle = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} f_J^{(k)}$$

$$\langle A(\{f\}) \rangle = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} A(\{f^{(k)}\})$$

$$\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2$$

Individual replicas may fluctuate significantly, average quantities such as central values and 1σ error bands are smooth inasmuch as stability is reached due to the dimension of the ensemble increasing

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 single NN

Summary for the user

Hessian method (CT, CJ, MSTW, ABKM, HERAPDF)

$$\langle \mathcal{F} \rangle = \mathcal{F}[q^{(0)}]$$

$$\sigma_{\mathcal{F}}^{\mathrm{Hess}} = \frac{1}{2}\left(\sum_{k=1}^{N_{\mathrm{set}}/2}\left(\mathcal{F}[\{q^{(2k-1)}\}] - \mathcal{F}[\{q^{(2k)}\}]\right)^2\right)^{1/2}$$

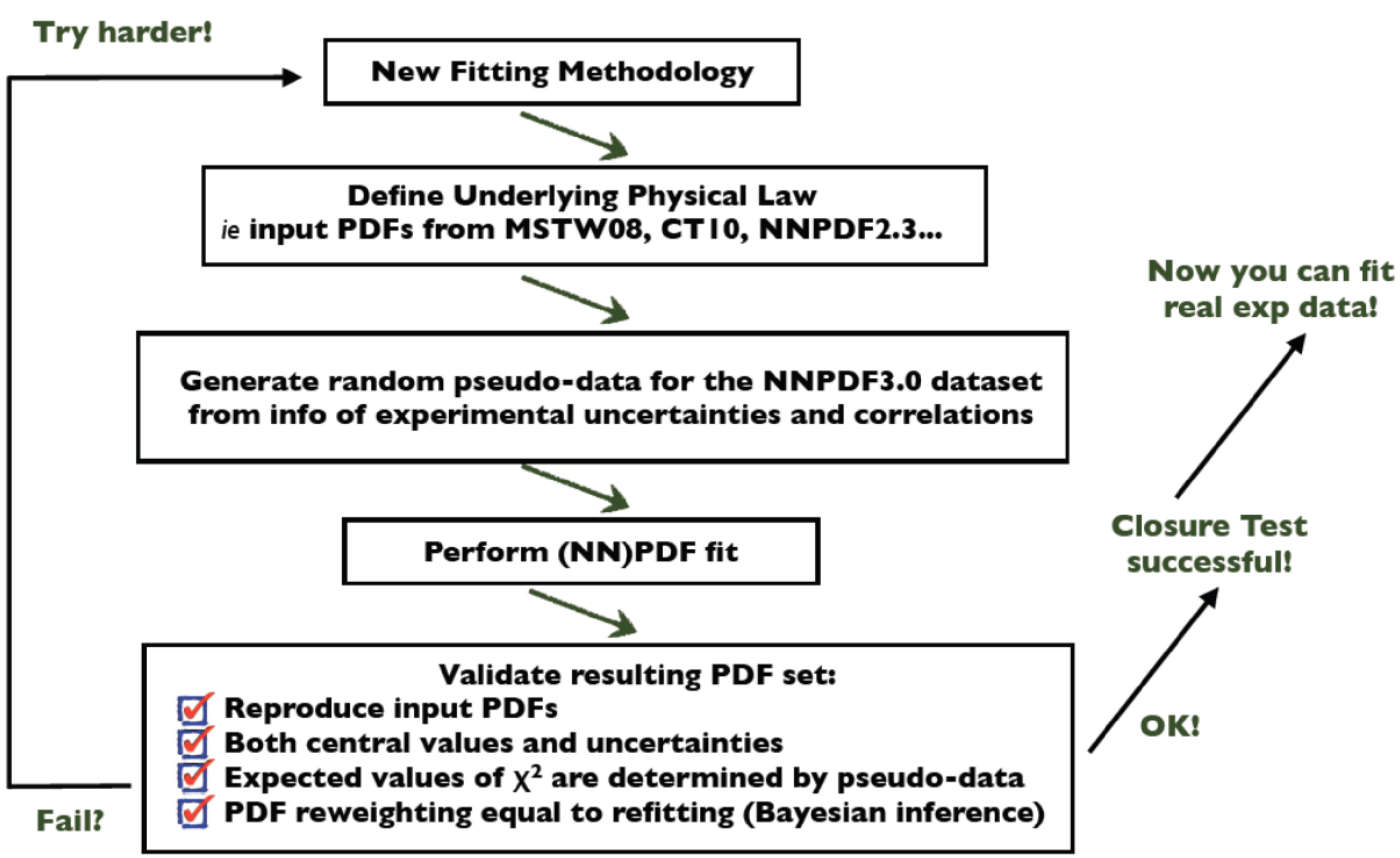
Monte Carlo method (NNPDF, JAM)

$$\langle \mathcal{F} \rangle = \frac{1}{N_{\mathrm{set}}}\sum_{i=1}^{N_{\mathrm{set}}} \mathcal{F}[q^{(i)}]$$

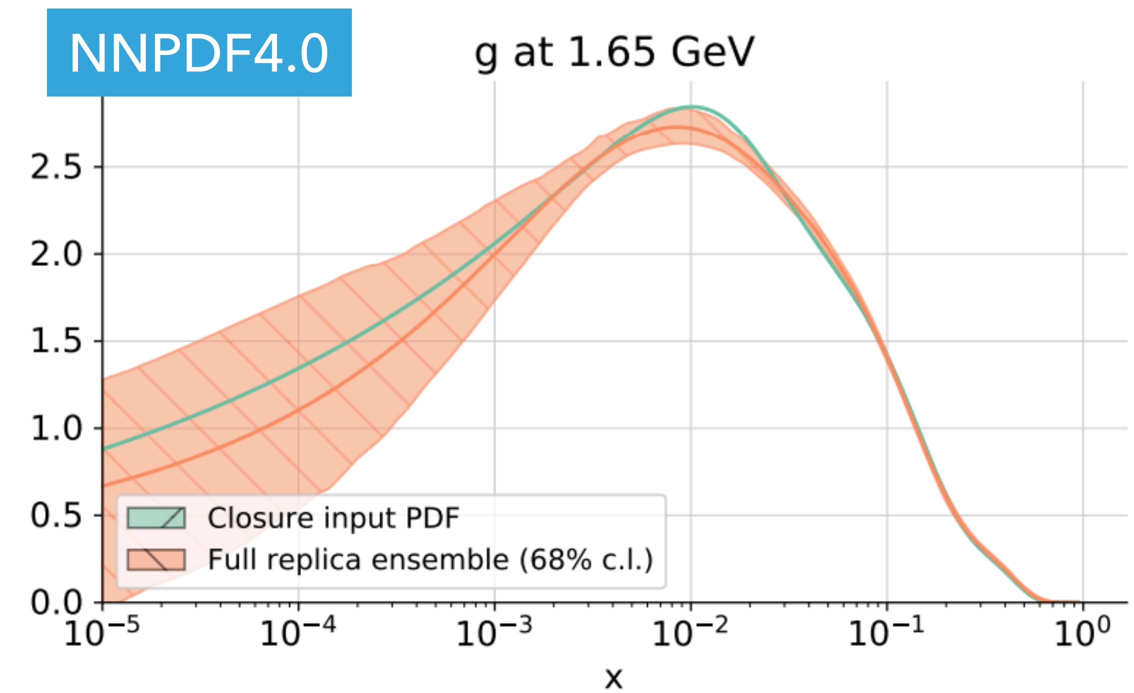
$$\sigma_{\mathcal{F}}^{\mathrm{MC}} = \left(\frac{1}{N_{\mathrm{N_{set}}}}\sum_{k=1}^{N_{\mathrm{set}}}\left(\mathcal{F}[\{q^{(k)}\}] - \langle \mathcal{F}[\{q\}] \rangle\right)^2\right)^{1/2}$$

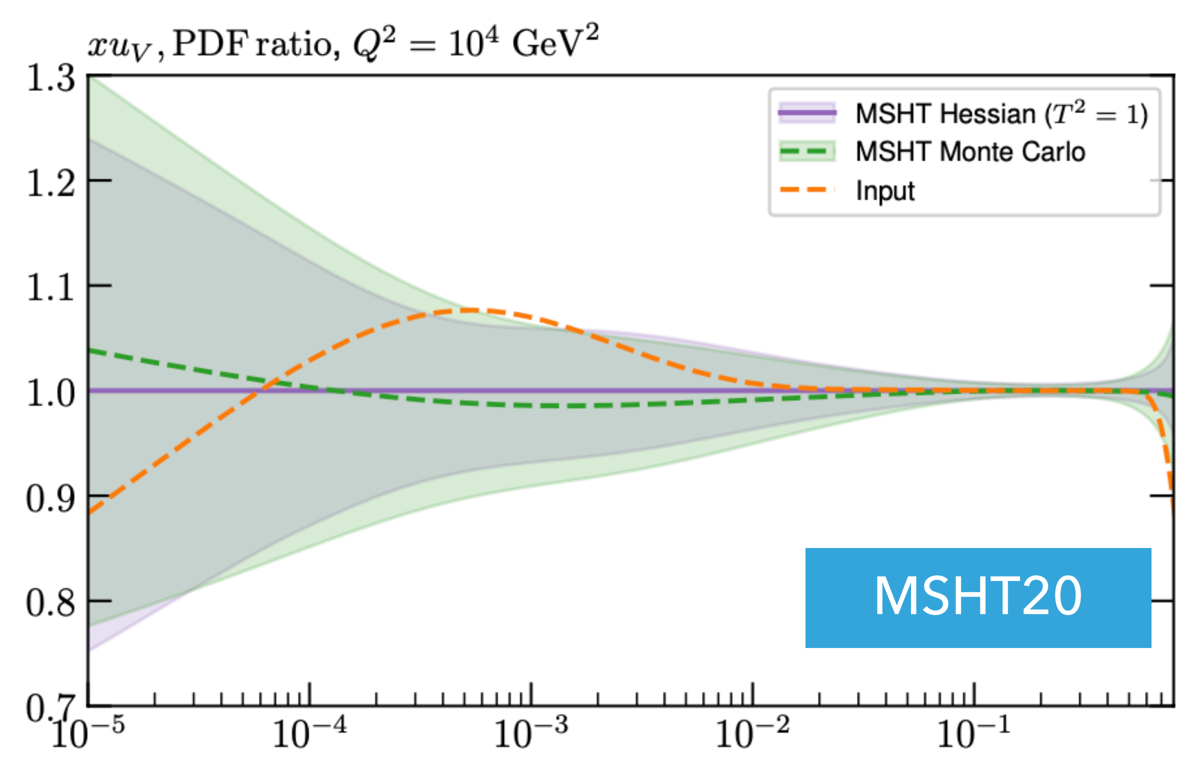
Statistical validation

- Assume PDFs known: generate artificial data with them and th. predictions
- Fit PDFs to artificial data
- Check whether fit reproduces the underlying "truth", and whether true values are Gaussianly distributed



Del Debbio et al, [arXiv: 2111.05787]





Harland-Lang et al, arXiv:2407.07944

Statistics and methodology summary

➡ PDF determination: Hessian Method
- Simple linear error propagation
- Tolerance required for realistic uncertainties
- Parametrisation bias possible

➡ PDF determination: Monte Carlo method
- Two-step procedure: data MC -> PDF MC
- Very general parametrisation allowed
- Need optimal fit determination method (cross-validation)

➡ PDF representation: Hessian vs Monte Carlo
- Conversion possible either way
- Compression method available either way
- MC very flexible, Hessian very efficient

➡ PDF validation: the closure test
- Performed by both NNPDF and MSHT
- Interpolation and functional uncertainties significant

3. Theoretical aspects & theory frontiers

Theory predictions in PDF fits

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

PDF parameters determined by minimising figure of merit

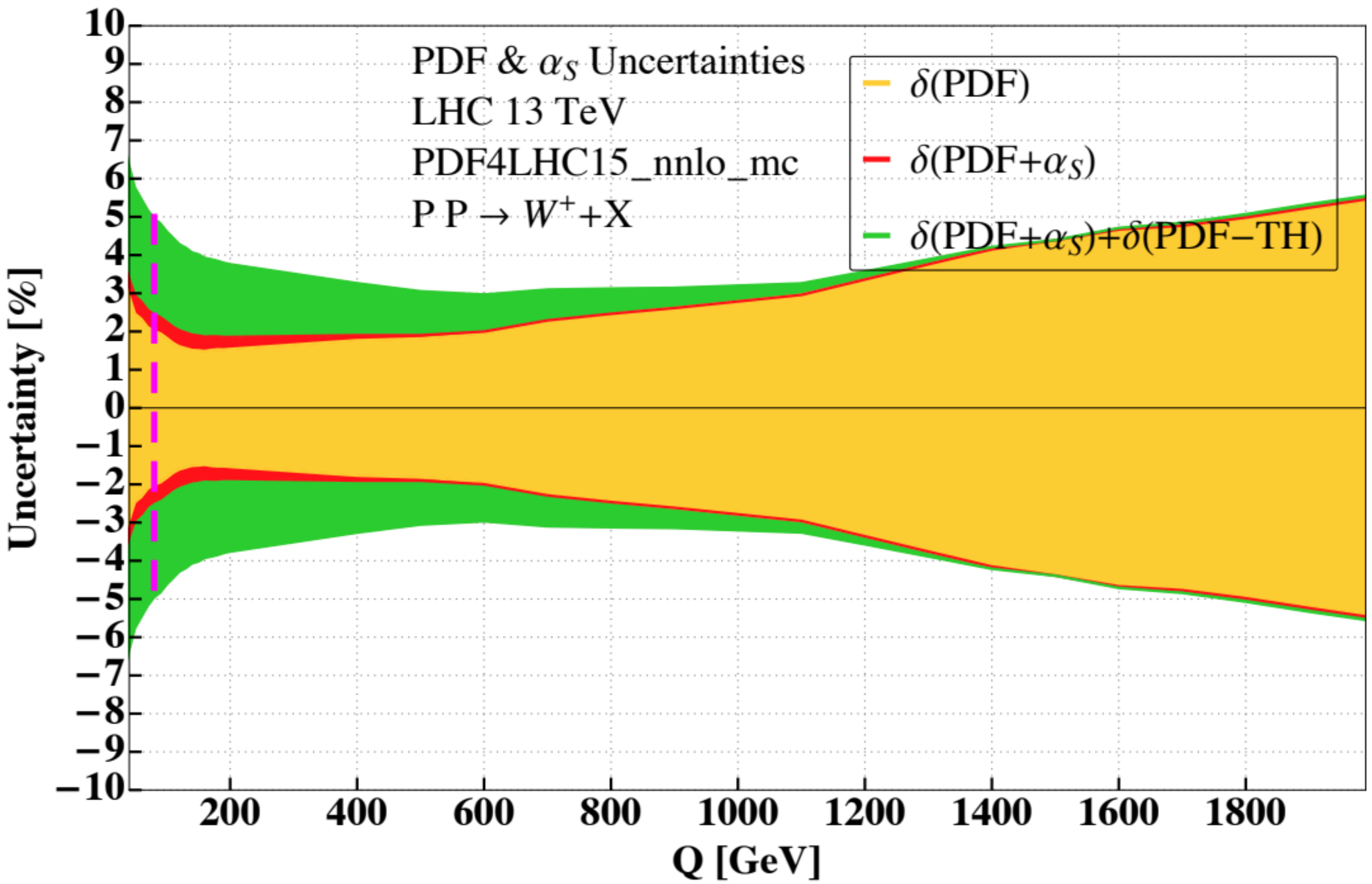
$$T = f_1\,(\otimes f_2)\,\otimes\,\hat{\sigma}$$

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

- Standard global PDF fits based on fixed-order NNLO QCD calculations with two PDF collaborations now producing aNNNLO PDFs (MSHTaN3LO and NNPDF40aN3LO)
- Standard global PDF fits set specific values for
  - $\alpha_s(M_z)$, M_W, M_z, $\alpha_e(M_z)$
  - CKM matrix elements
  - Heavy quark mass thresholds
  - Branching ratios…
- MHOUs in perturbative expansion, the uncertainty on the value of the parameters that enter a PDF fit are **NOT** included in default PDF error bars

Mismatch between pert. orders

- PDF fits performed at given perturbative order - NNLO or aN3LO
- PDF uncertainties only reflect lack of information from data
- Theoretical uncertainties (dominated by MHOU) ignored until recently
- Mismatch between perturbative order of partonic cross section and PDFs becoming significant source of uncertainty

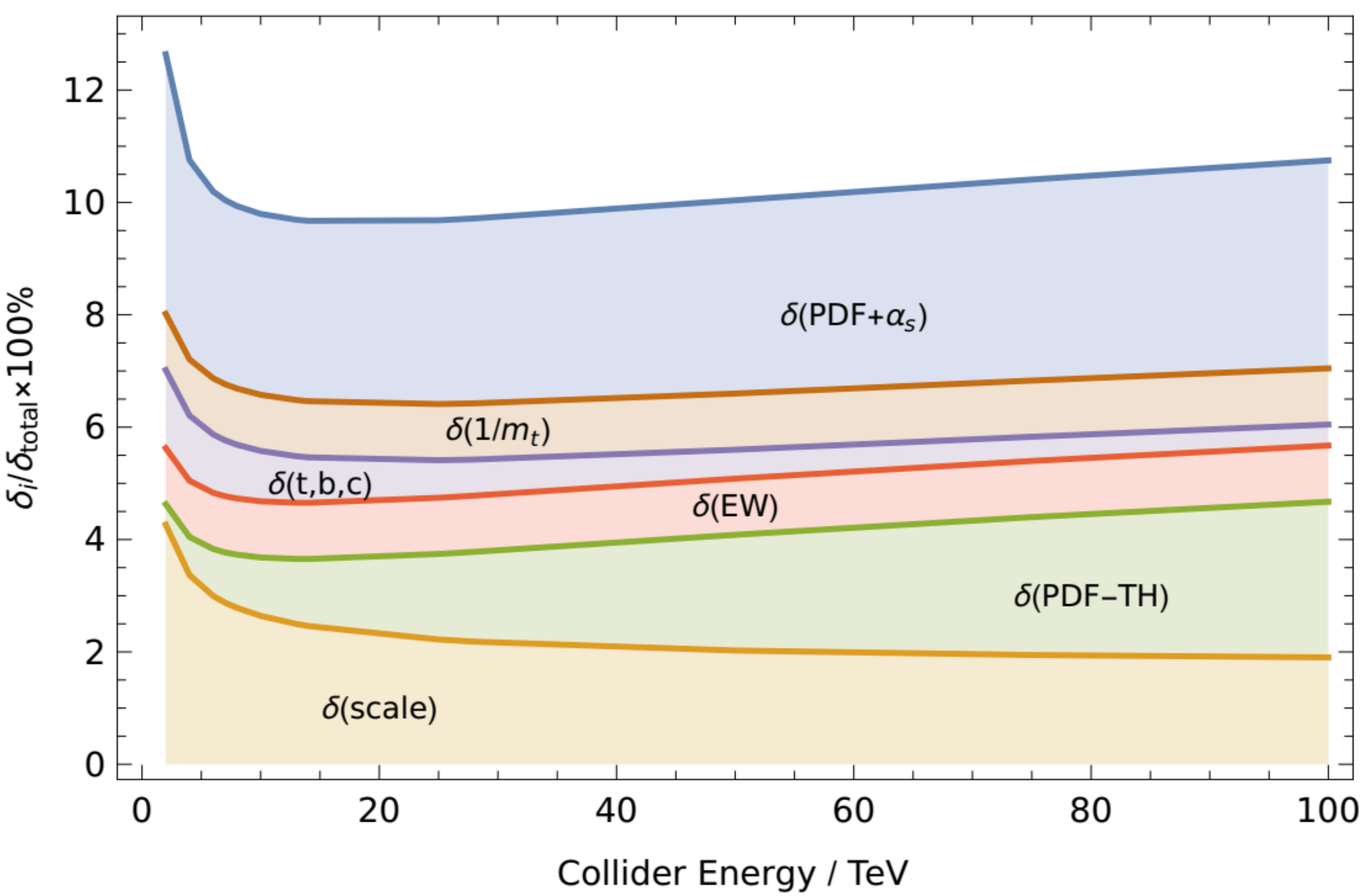


$$\delta(PDF-TH) = \frac{1}{2}\left|\frac{\sigma^{(2)}_{\mathrm{NNLO-PDFs}} - \sigma^{(2)}_{\mathrm{NLO-PDFs}}}{\sigma^{(2)}_{\mathrm{NNLO-PDFs}}}\right|$$

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

Mismatch between pert. orders

- PDF fits performed at given perturbative order - NNLO or aN3LO
- PDF uncertainties only reflect lack of information from data
- Theoretical uncertainties (dominated by MHOU) ignored until recently
- Mismatch between perturbative order of partonic cross section and PDFs becoming significant source of uncertainty

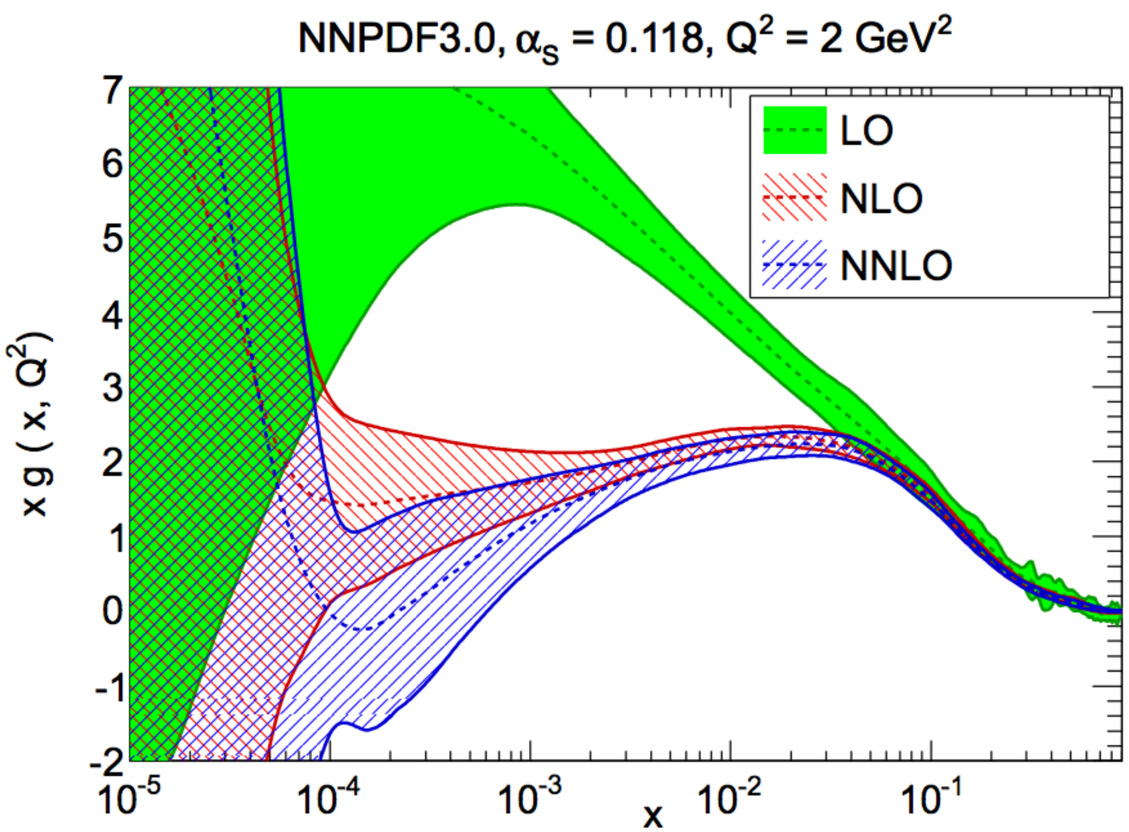


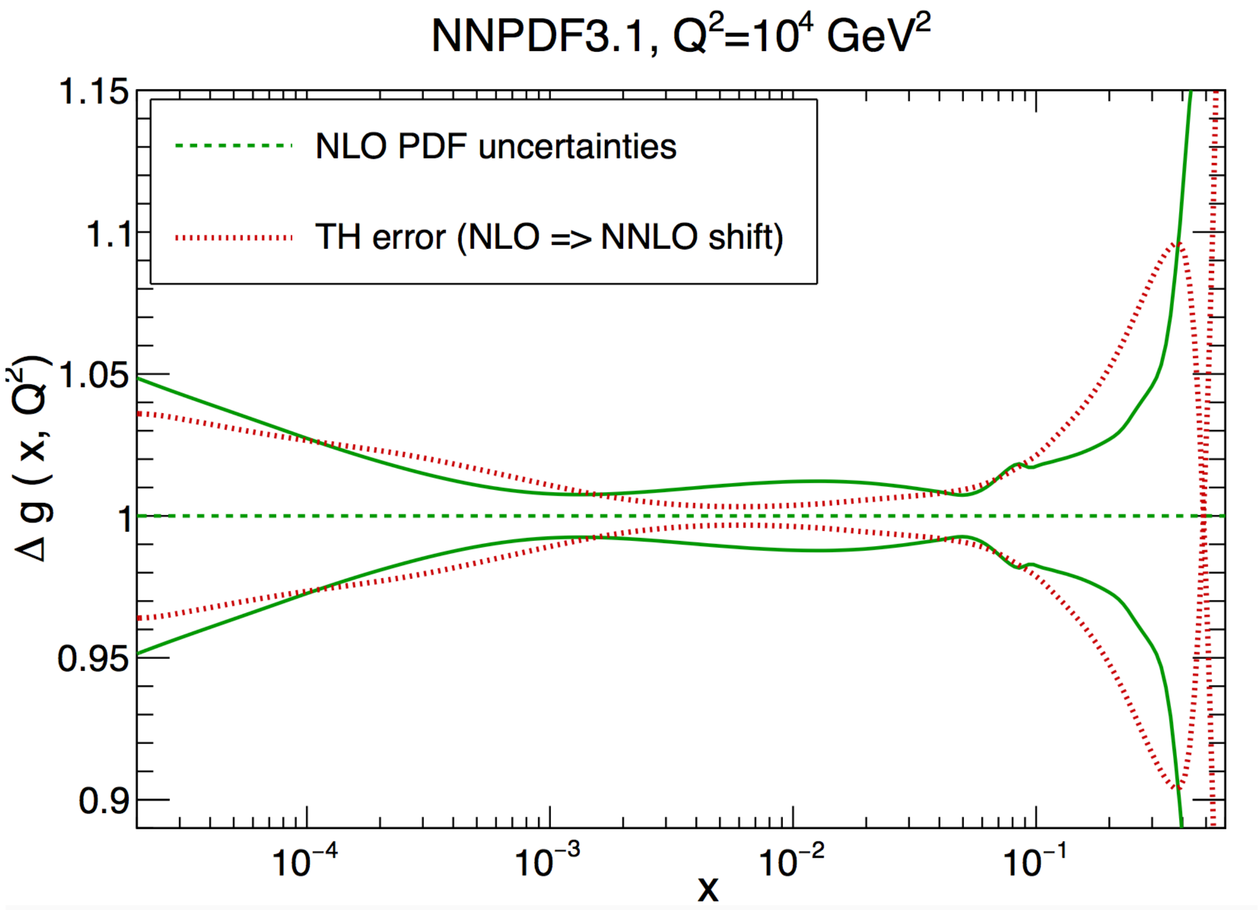
$$\delta(PDF - TH) = \frac{1}{2}\left|\frac{\sigma^{(2)}_{\text{NNLO-PDFs}} - \sigma^{(2)}_{\text{NLO-PDFs}}}{\sigma^{(2)}_{\text{NNLO-PDFs}}}\right|$$

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

MHOU in PDF fits

- In a fit based on NLO theoretical predictions the theory error is already comparable to experimental error. What about a NNLO fit?
- How to include MHOUs in PDF error bands at NNLO?





Ball et al, EPJC 77 (2017)

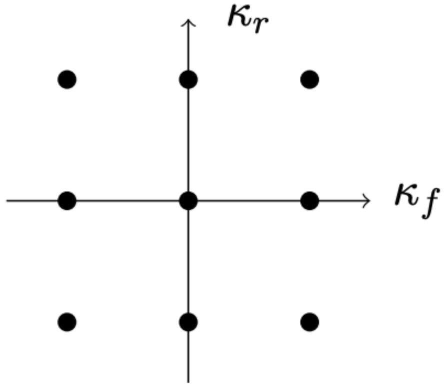
MHOU in PDF fits

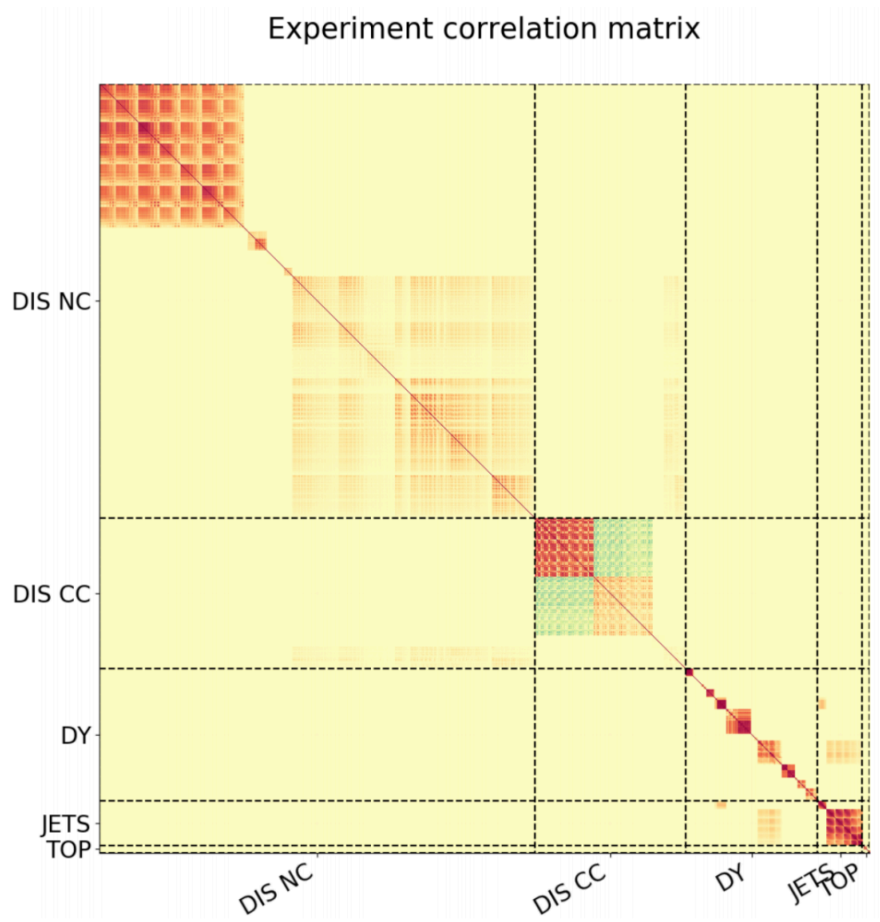
Option 1 - theory covmat [NNPDF: 1906.10698, 2401.10319]

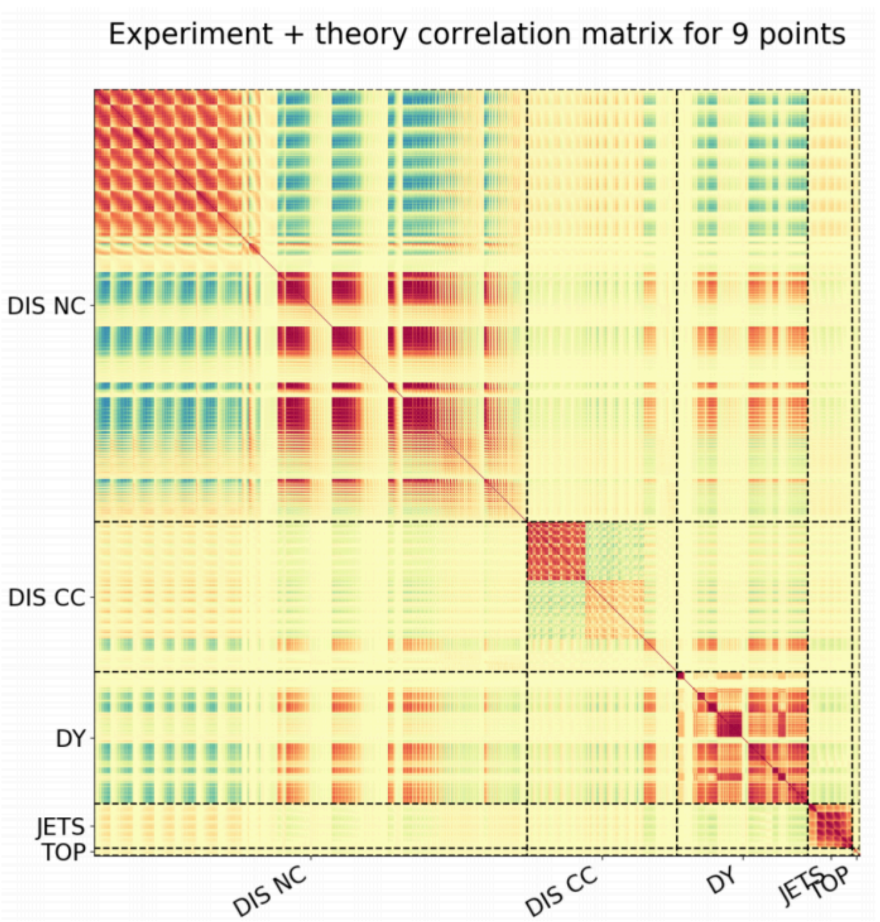
Construct a theory covariance matrix from scale-varied cross sections and combine it with the experimental covariance matrix

$$\chi^2 = \sum_{i,j=1}^{N_{\text{dat}}} (T_i - D_i) \left(\text{cov}_{\text{exp}} + \text{cov}_{\text{th}}\right)^{-1}_{ij} (T_j - D_j)$$

- Assumptions: experimental and theoretical errors independent and Gaussian
- Assumptions on correlation of scales and scale ratio will determine the specific form of the covariance matrix







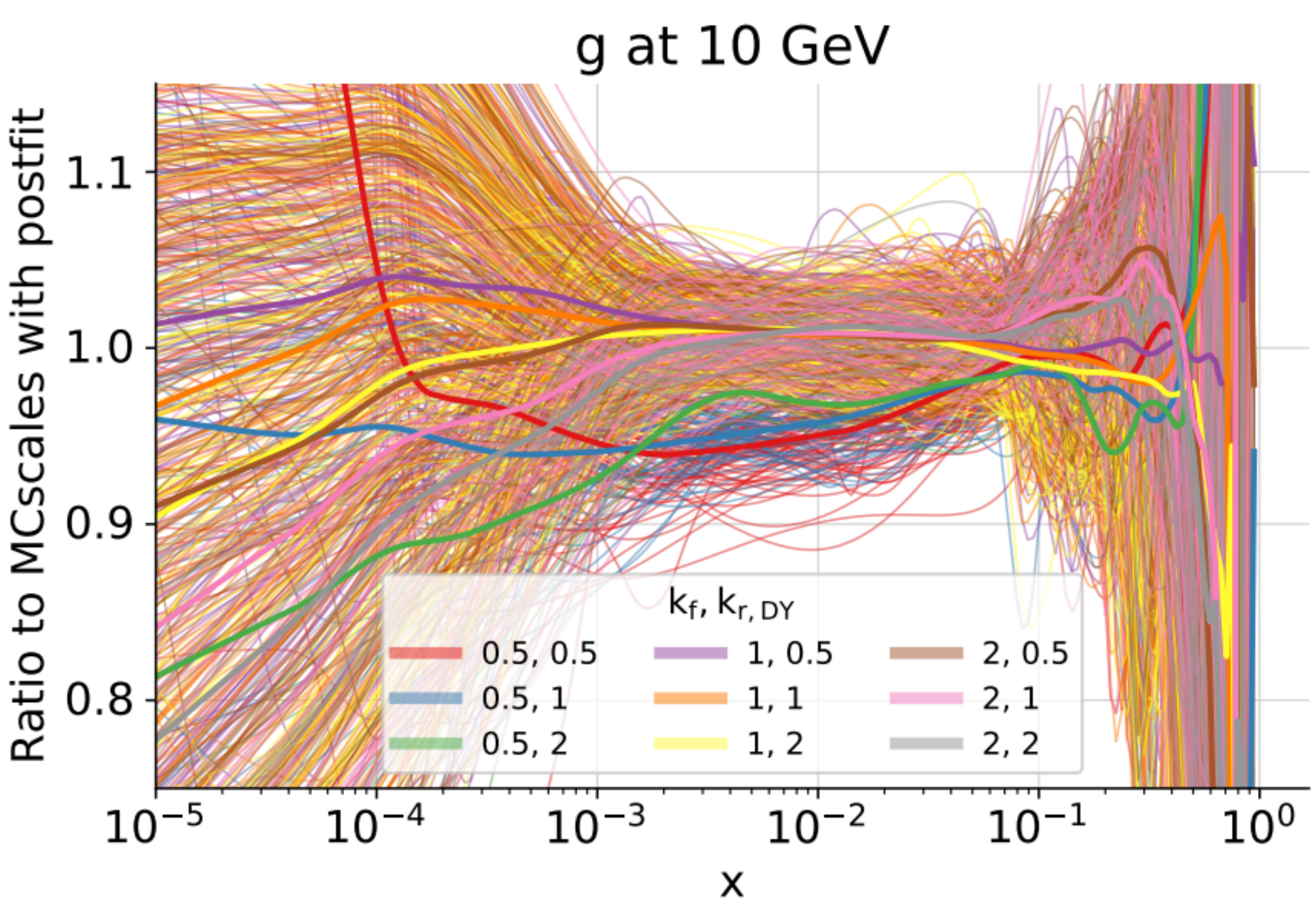
MHOU in PDF fits

Option 2 - MCscales [Kassabov et al: 2207.07616]

Main idea: renorm. and fact. scales are free parameters of the fixed-order theory, that induce an uncertainty on theory predictions included in a PDF fit & need to be propagated

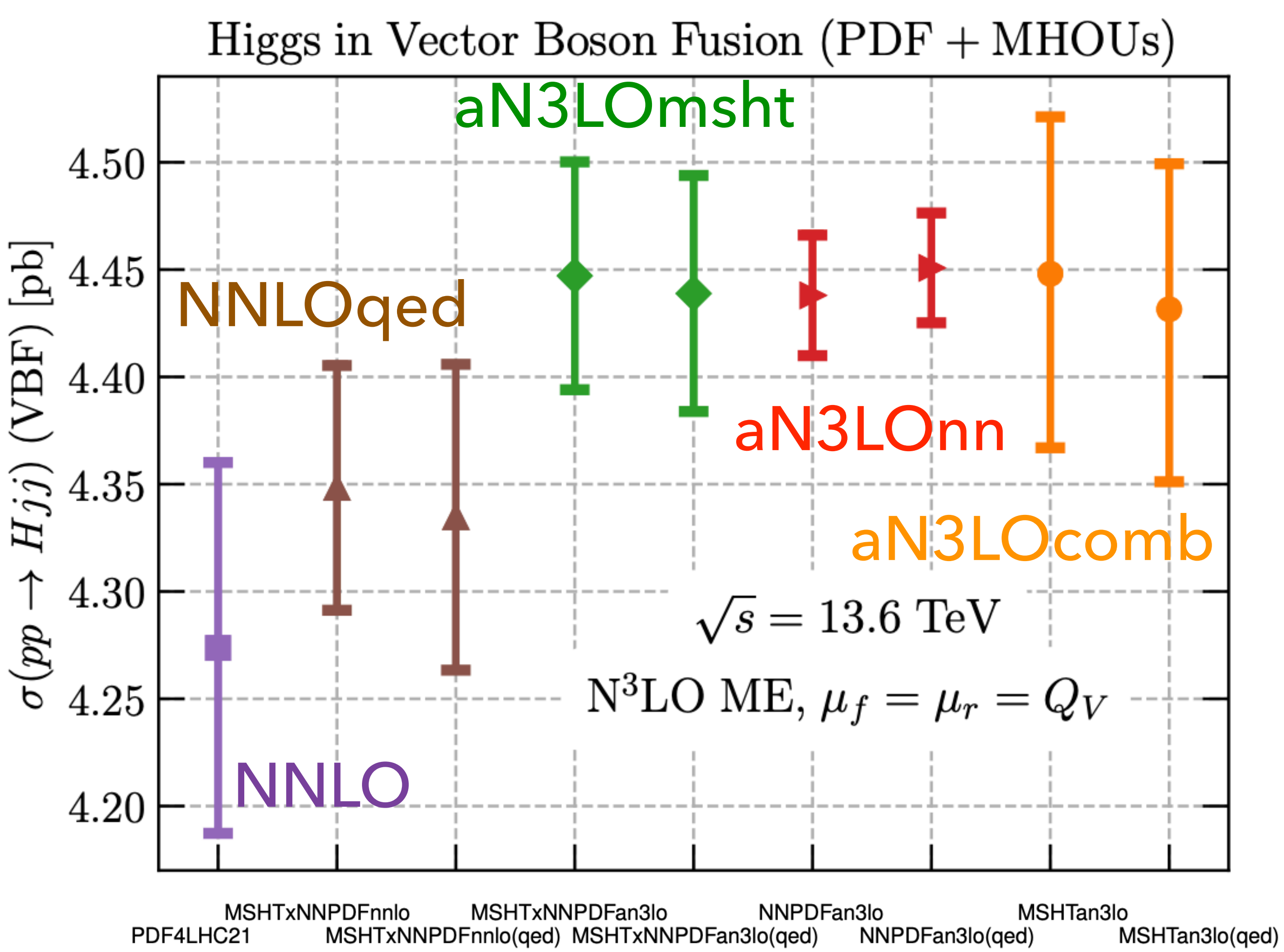
Option 3 - theory nuisance parameters [MMHT: 2207.04739]

Main idea: add MHOU as nuisance parameters and fit nuisance parameters from data.



The N3LO frontier

- MSHT & NNPDF: inclusion of available theoretical ingredients at N3LO (non-singlet splitting functions, singlet splitting function in the large nf limit, small-x limit, large-x limit, Mellin moments + DIS structure functions in the massless limit and approximate heavy flavour structure functions between known limits + hadronic N3LO K-factors)



arXiv:2411.05373

- MSHT: MHOU and IHOU (incomplete higher order uncertainty) added as nuisance parameters and fitted from the data

- NNPDF: MHOU added via theory covariance matrix, IHOU added as extra additional theory uncertainty

The N3LO frontier

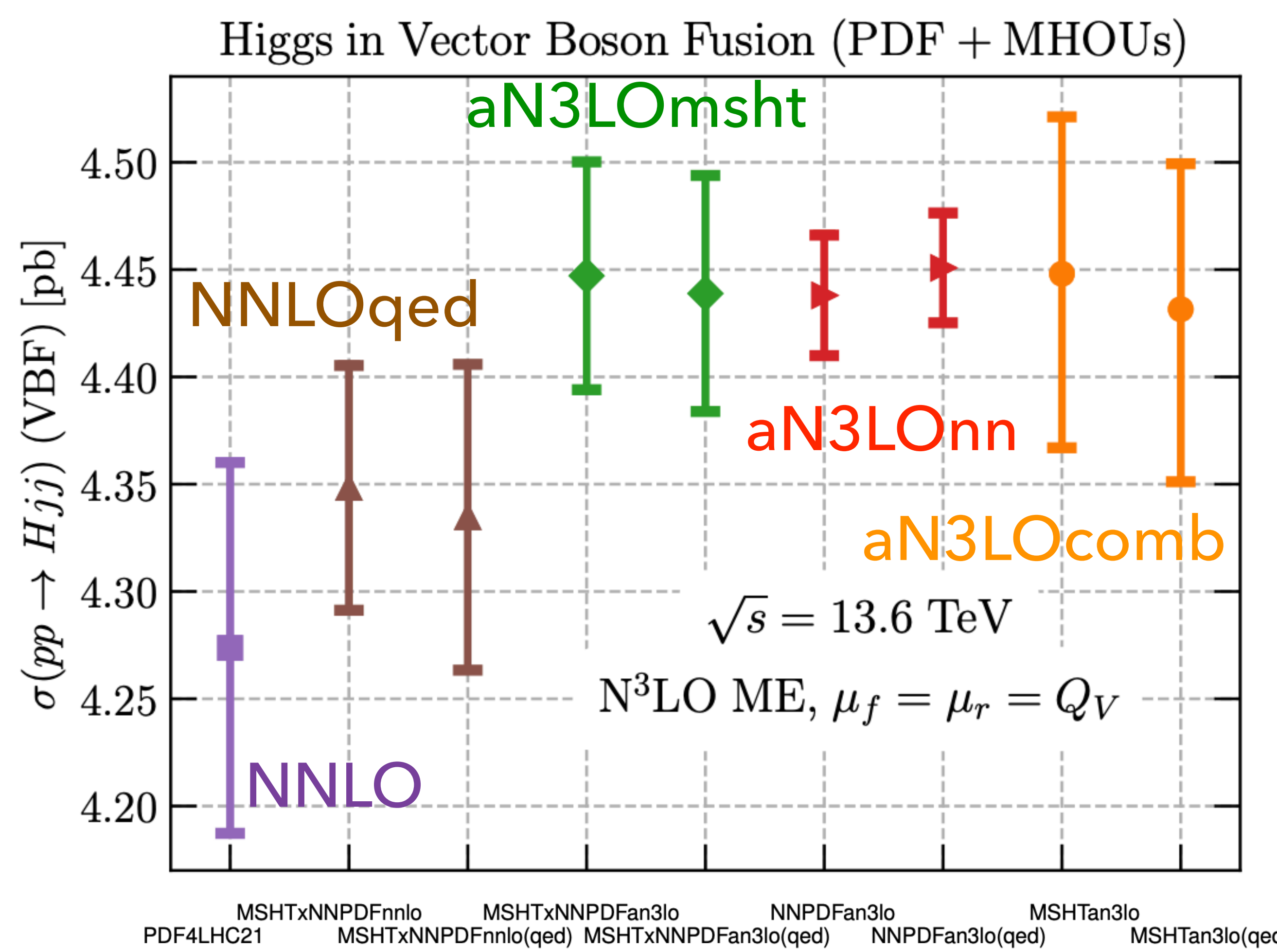
Error estimate before aN3LO PDFs

$$\Delta^{\text{app}}_{\text{NNLO}} \equiv \frac{1}{2}\left|\frac{\sigma^{\text{NNLO}}_{\text{NNLO-PDF}} - \sigma^{\text{NNLO}}_{\text{NLO-PDF}}}{\sigma^{\text{NNLO}}_{\text{NNLO-PDF}}}\right|$$

Actual error using aN3LO PDFs

$$\Delta^{\text{exact}}_{\text{NNLO}} \equiv \left|\frac{\sigma^{\text{N}^3\text{LO}}_{\text{N}^3\text{LO-PDF}} - \sigma^{\text{N}^3\text{LO}}_{\text{NNLO-PDF}}}{\sigma^{\text{N}^3\text{LO}}_{\text{N}^3\text{LO-PDF}}}\right|$$

	ggF	VBF
Δ_{NNLO} (app)	1.6%	0.5%
Δ_{NNLO} (actual)	3.3%	2.3%

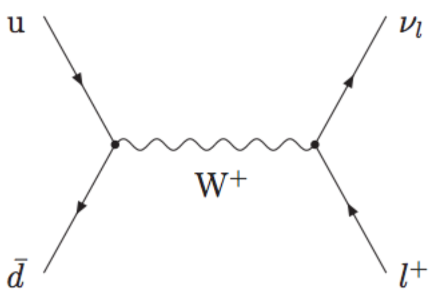


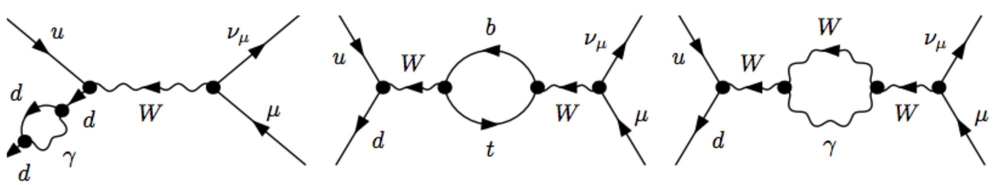
arXiv:2411.05373

Previous estimates of the effect of NNLO/N3LO mismatch were **optimistic**

Electroweak corrections

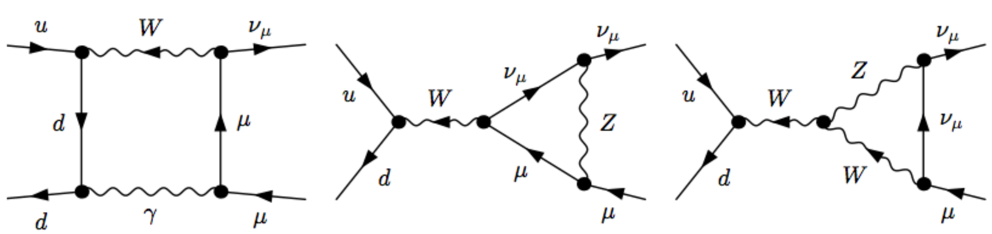
- Given that $\alpha(Mz) \sim \alpha_S(Mz)/10 \Longrightarrow$ NLO EW corrections ~ NNLO QCD corrections

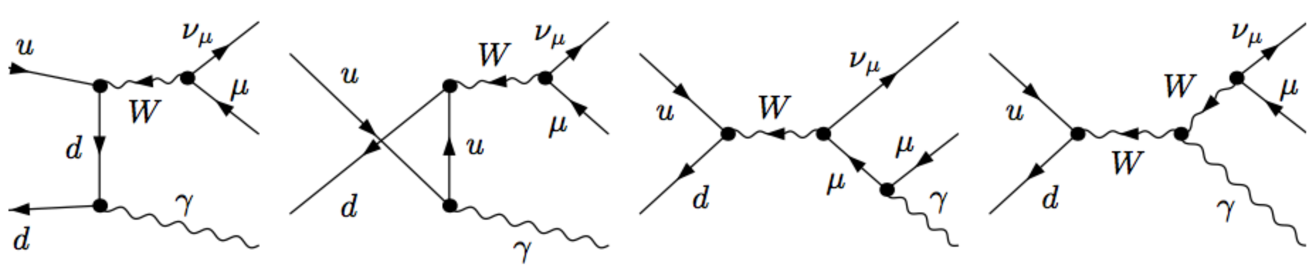




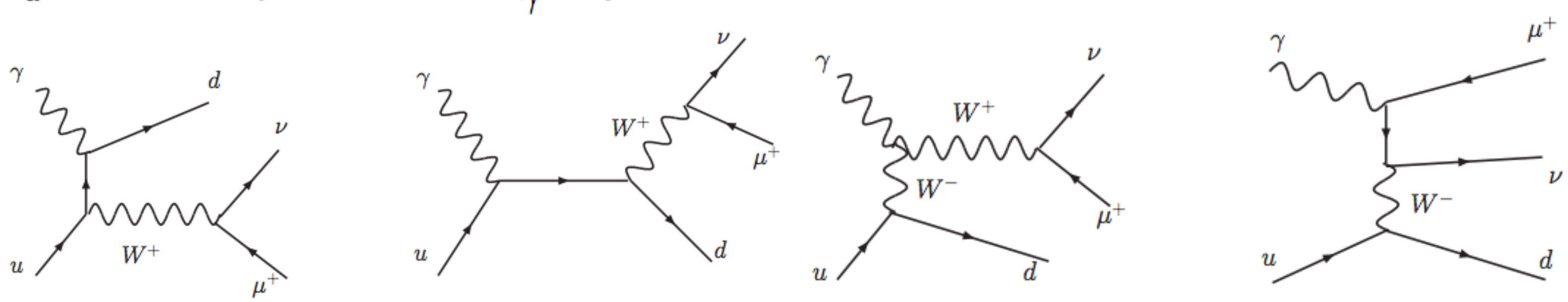
(a) (b) (c)

Virtual EW corrections





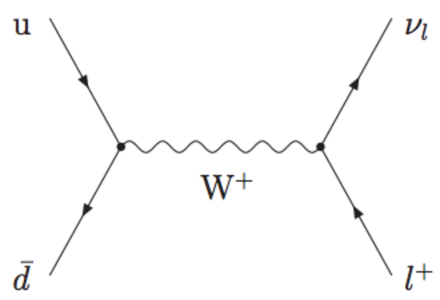
Real EW corrections - quark initiated

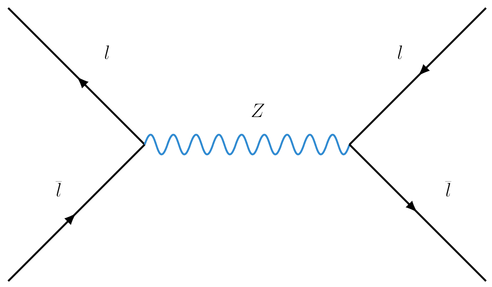


Real EW corrections - photon initiated

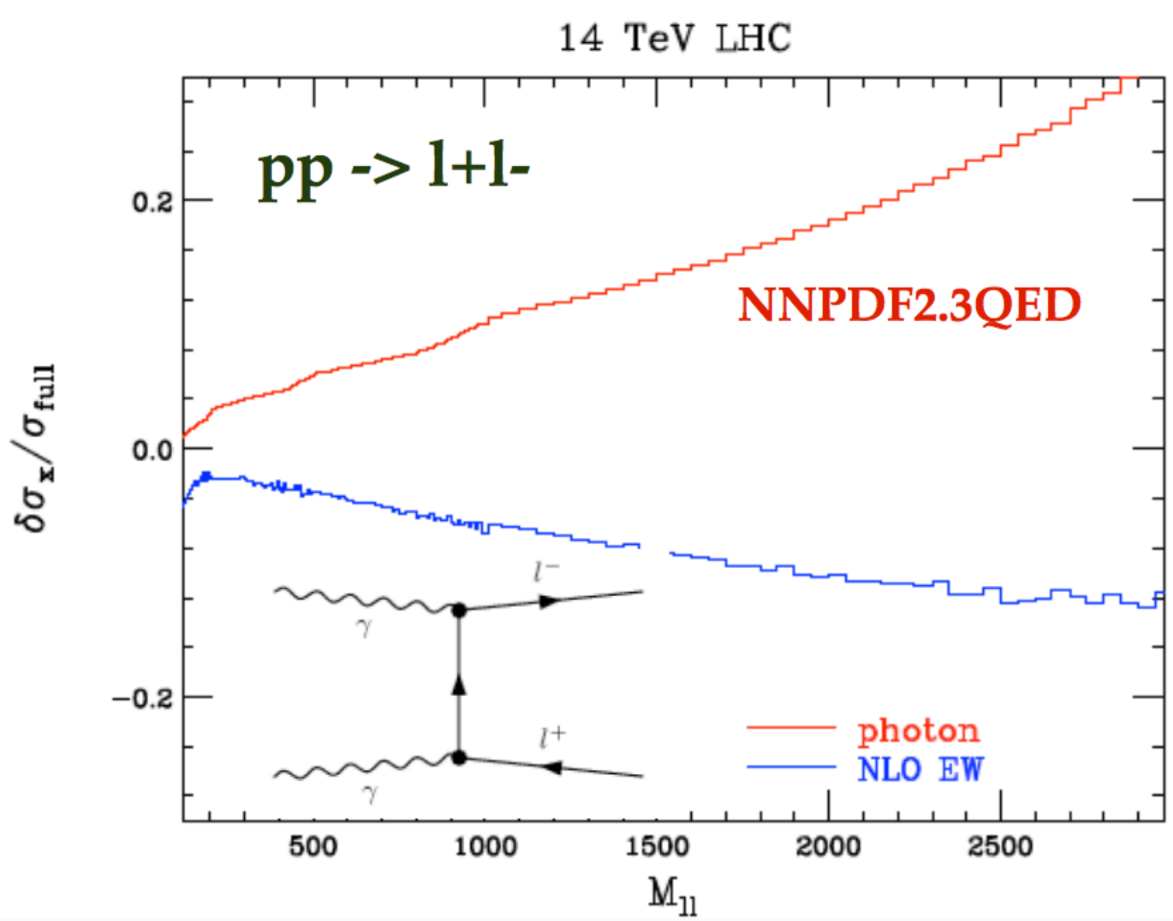
Electroweak corrections

- Given that $\alpha(M_Z) \sim \alpha_S(M_Z)/10 \Longrightarrow$ NLO EW corrections ~ NNLO QCD corrections





- NLO virtual EW corrections become large in the large p_T region of lepton but partially compensated by photon-initiated real corrections



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

Photon-modified DGLAP

- How are PDFs modified by inclusion of initial photon PDF?

$$Q^2 \frac{\partial}{\partial Q^2} g(x, Q^2) = \sum_{q,\bar{q},g} P_{ga}(x, \alpha_s(Q^2)) \otimes f_a(x, Q^2) + P_{g\gamma}(x, \alpha_s(Q^2)) \otimes \gamma(x, Q^2),$$

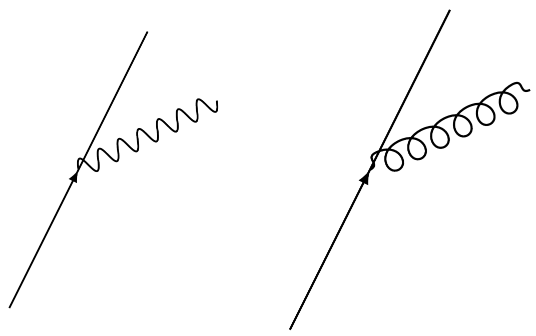
$$Q^2 \frac{\partial}{\partial Q^2} q(x, Q^2) = \sum_{q,\bar{q},g} P_{qa}(x, \alpha_s(Q^2)) \otimes f_a(x, Q^2) + P_{q\gamma}(x, \alpha_s(Q^2)) \otimes \gamma(x, Q^2),$$

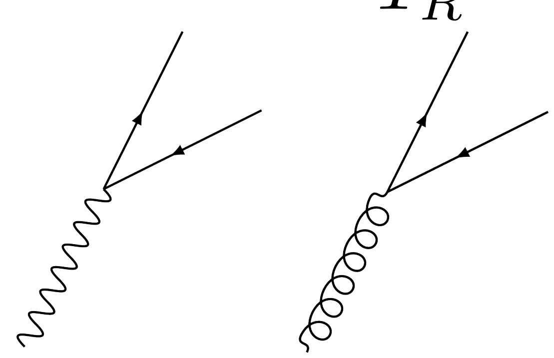
$$Q^2 \frac{\partial}{\partial Q^2} \gamma(x, Q^2) = P_{\gamma\gamma} \otimes \gamma(x, Q^2) + \sum_{q,\bar{q},g} P_{\gamma a}(x, \alpha_s(Q^2)) \otimes f_a(x, Q^2).$$

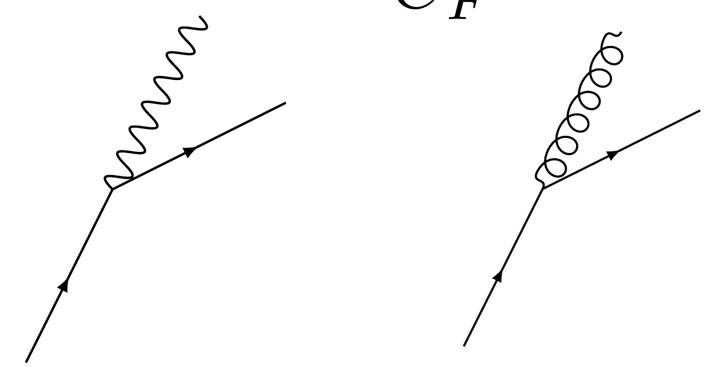
- DGLAP splitting functions expanded in powers of α_s and $\boldsymbol{\alpha}$

$$P_{ij} = \sum_{m,n} \left(\frac{\alpha_S}{2\pi}\right)^m \left(\frac{\alpha}{2\pi}\right)^n P_{ij}^{(m,n)}$$

$$P_{qq}^{(0,1)} = \frac{e_q^2}{C_F} P_{qq}^{(1,0)} \qquad P_{q\gamma}^{(0,1)} = \frac{e_q^2}{T_R} P_{qg}^{(1,0)} \qquad P_{\gamma q}^{(0,1)} = \frac{e_q^2}{C_F} P_{gq}^{(1,0)}$$

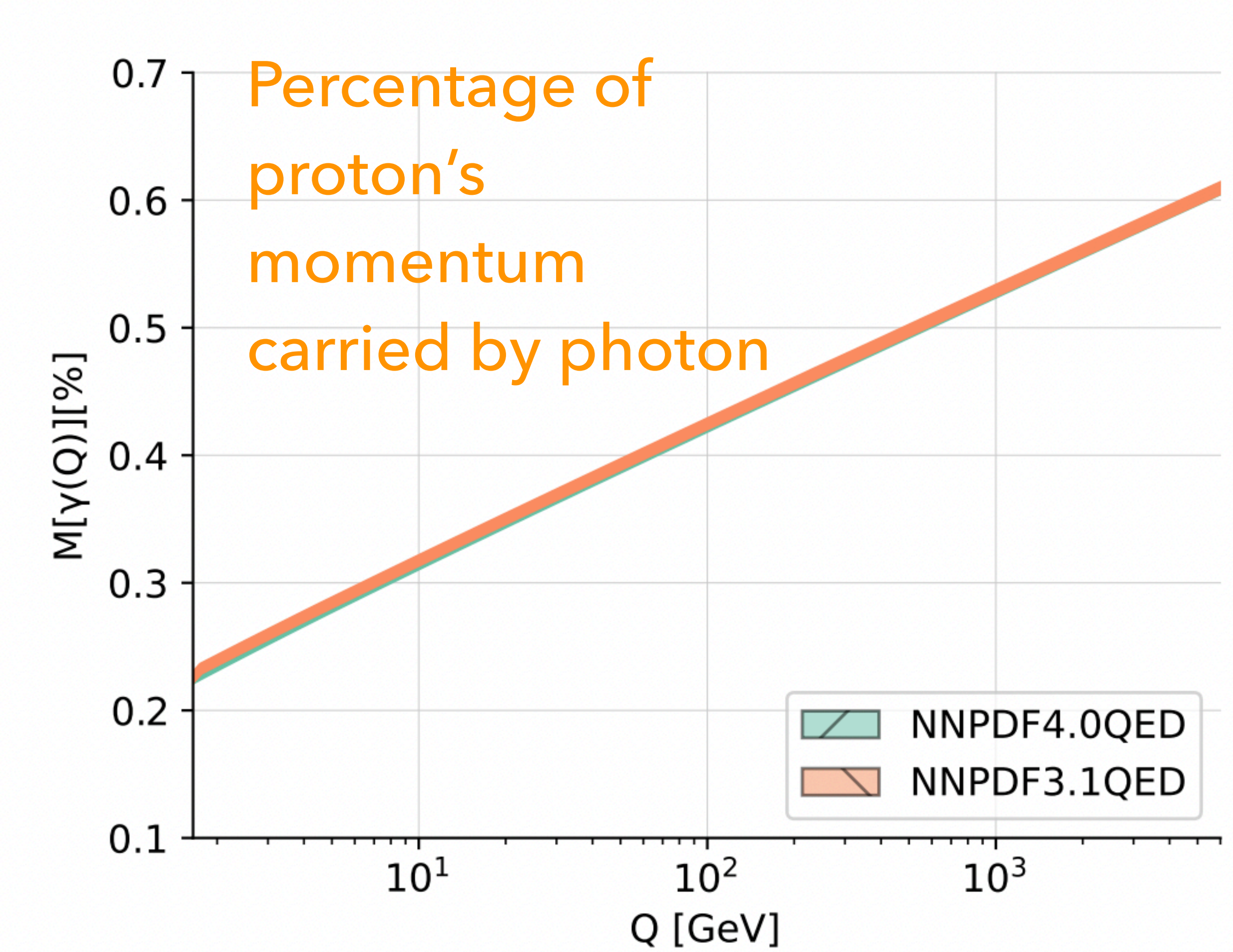


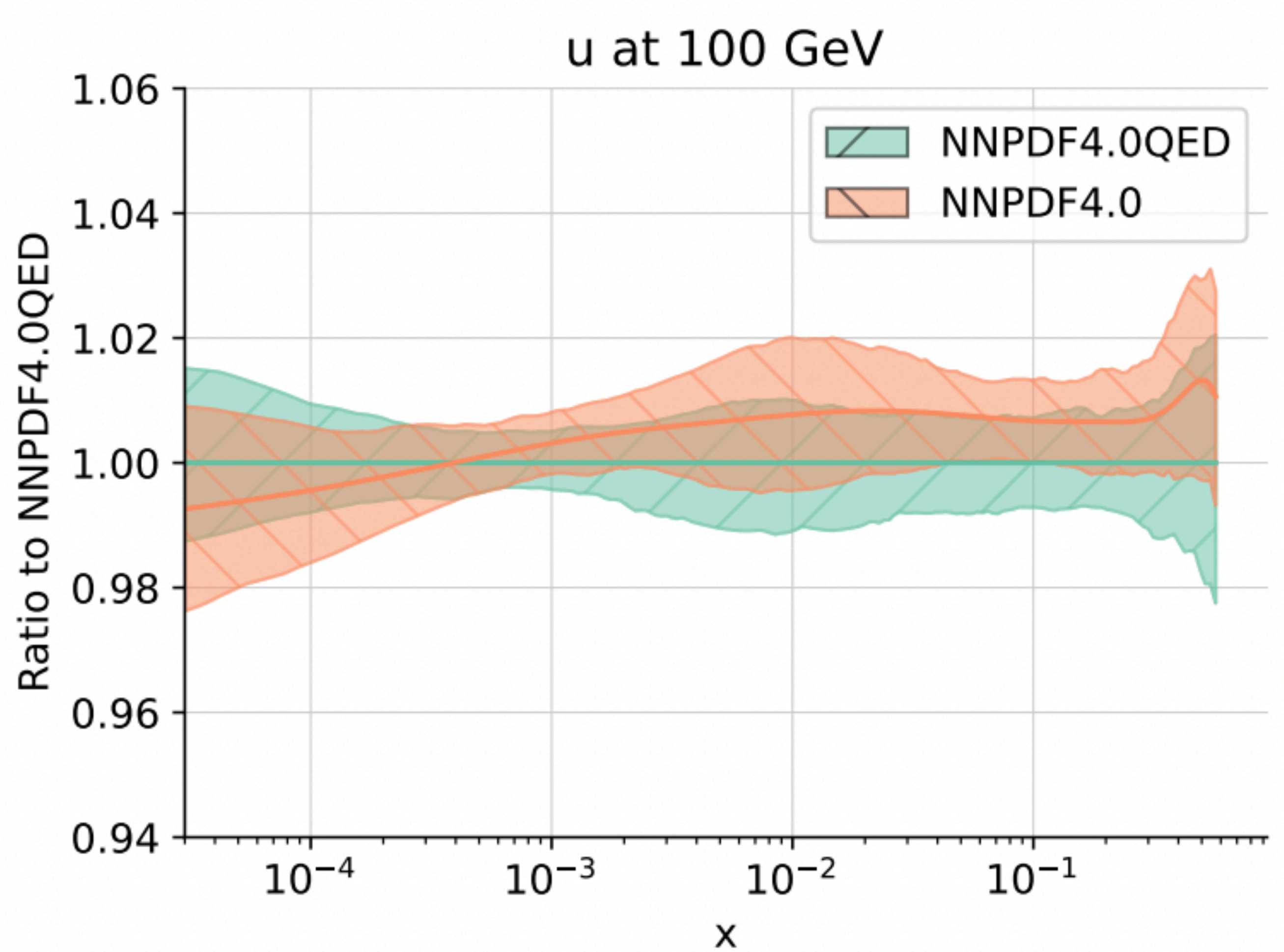




Photon-modified DGLAP

- Quark and gluon PDFs change up to 1% at large x
- How do we determine the photon PDF?
- In the best possible world: theory input and data input together
  Manohar, Nason, Salam, Zanderighi,1607.04266





NNPDF arXiv:2401.08749

State-of-the-art PDFs

The choice of PDFs matters

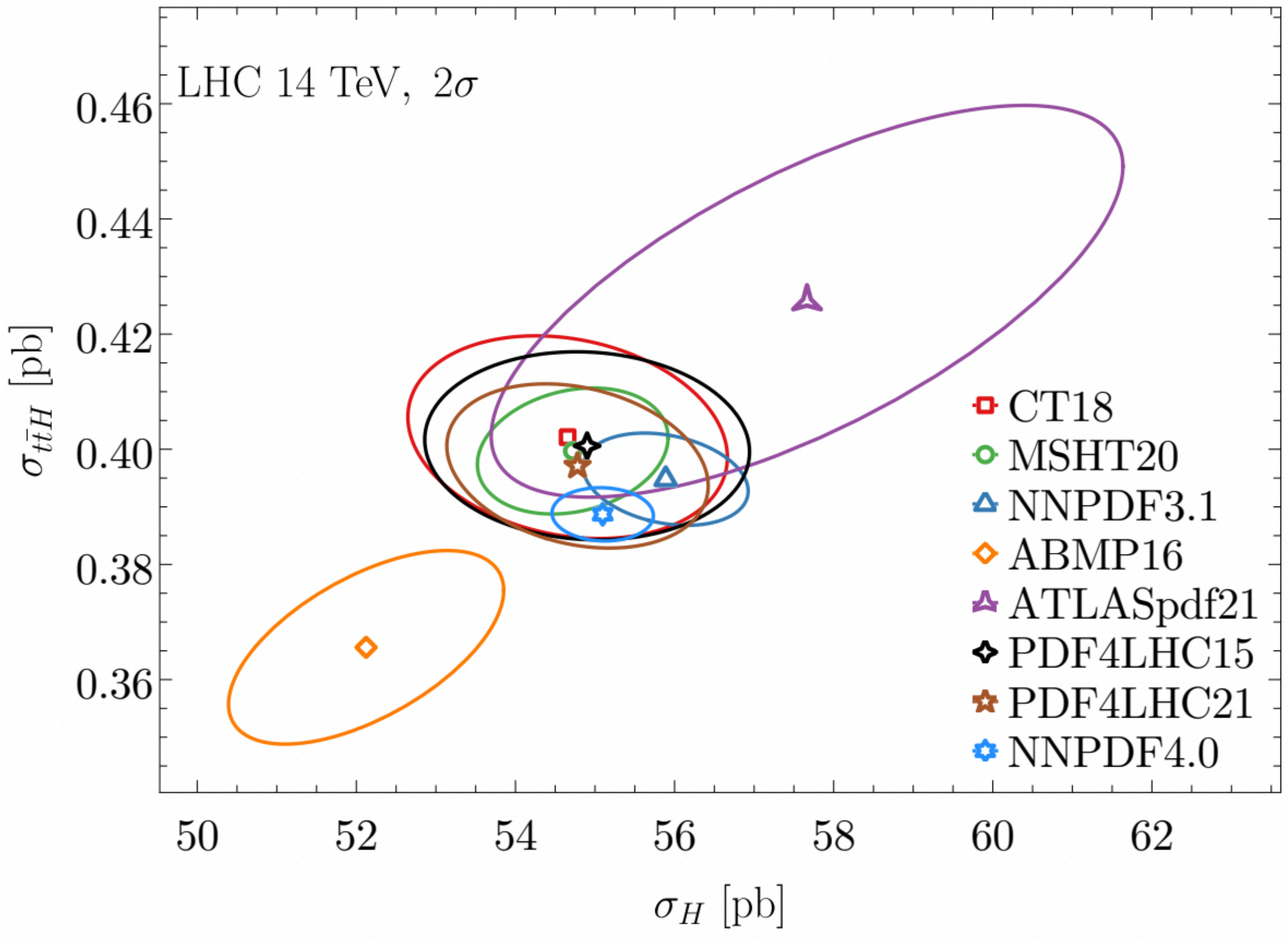
- What does PDF uncertainty include? How reliable it is?

- How do we interpret the difference predictions using different PDF sets?

- Shall we just pick a set out of the PDFs "supermarket" shelf or take the envelope of ALL predictions?

The choice of PDFs matters

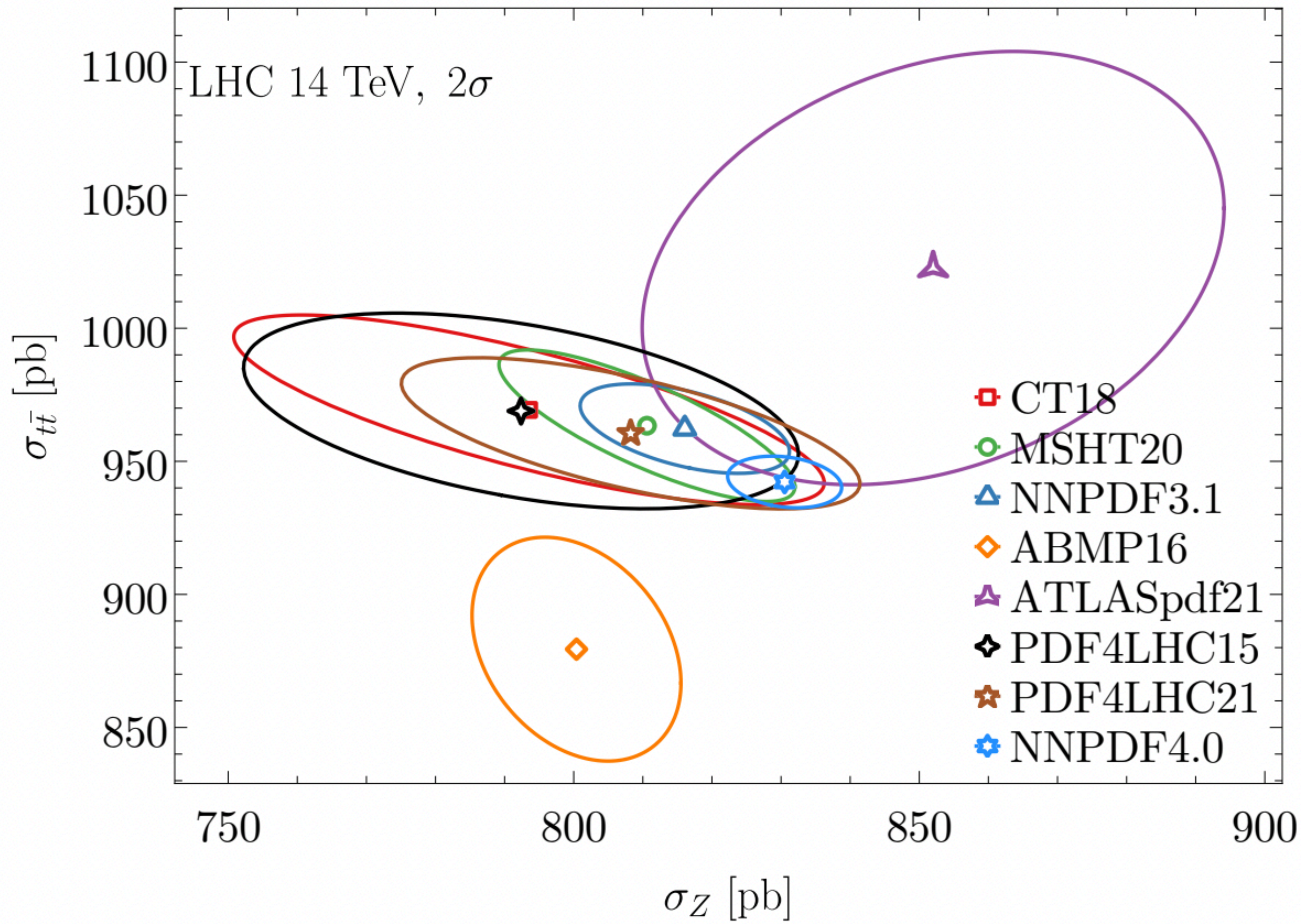
- What does PDF uncertainty include? How reliable it is?
- How do we interpret the difference predictions using different PDF sets?
- Shall we just pick a set out of the PDFs "supermarket" shelf or take the envelope of ALL predictions?



Snowmass 2022 white paper, arXiv: 2203.13923

The choice of PDFs matters

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



Snowmass 2022 white paper, arXiv: 2203.13923

The players

May 2025	NNPDF4.0	MSHT20	CT18	ABMP16	CJ22	JAM	ATLAS
Fixed Target DIS	✔	✔	✔	✔	✔	✔	✘
HERA I+II	✔	✔	✔	✔	✔	✔	✔
HERA jets	✔	✔	✘	✘	✘	✘	✘
Fixed Target DY	✔	✔	✔	✔	✔	✔	✘
SIDIS	✘	✘	✘	✘	✘	✔	✘
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 $\Delta\chi^2=1$	Hessian $\Delta\chi^2=1,10$	Monte Carlo	Hessian
Parametrization	Neural Networks	Polynomial (Chebyshev)	Polynomial (Bernstein)	Polynomial	Polynomial	Polynomial (50 pars)	Polynomial (25 pars)
HQ scheme	FONLL	TR’	ACOT-χ	FFN (+BMST)	ACOT	ACOT	TR’
Order	NNLO/aN3LO	NNLO/aN3LO	NNLO	NNLO	NLO	NLO	NNLO

Parton Luminosities

- A quick and easy way to assess the mass and the collider dependence of production cross sections at hadron-hadron colliders is to use Parton Luminosities
- At leading order in QCD (parton model)

$$\hat{\sigma}_{ab\to X} = C_X \delta(x_a x_b S - M^2)$$

$$\sigma_{pp\to X} = \int_0^1 dx_a dx_b f_a(x_a, M^2) f_b(x_b, M^2) \hat{\sigma}_{ab\to X}$$

- Thus

$$\sigma_{pp\to X} = C_X \int_0^1 dx_a dx_b f_a(x_a, M^2) f_b(x_b, M^2) \delta(x_a x_b S - M^2)$$
$$= \frac{C_X}{S} \frac{\partial \mathcal{L}_{ab}}{\partial \tau}$$

with

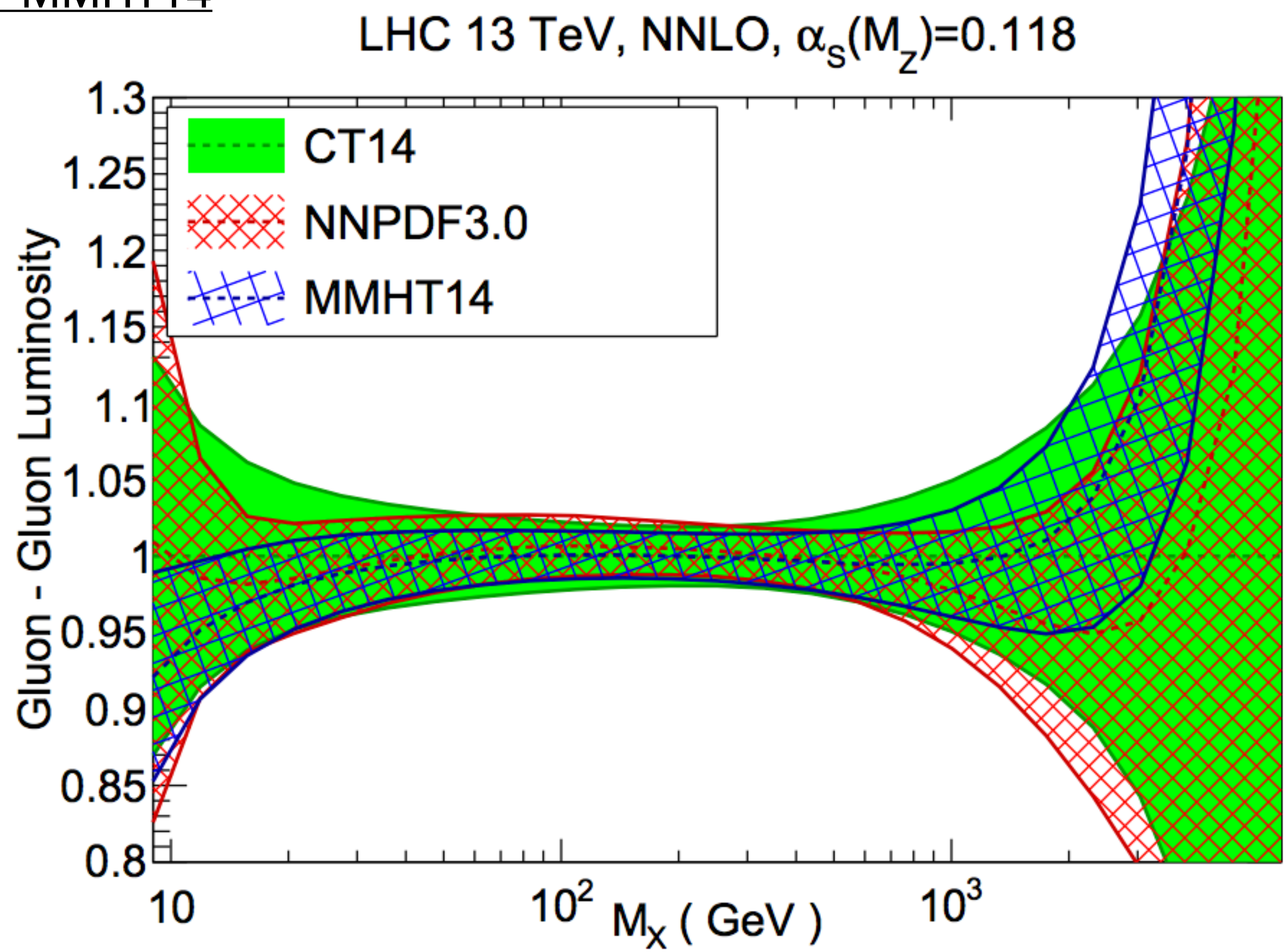
$$\tau = \frac{M^2}{\tau}$$

- Define

$$\Phi_{ab}(M^2) = \frac{\partial \mathcal{L}_{ab}}{\partial \tau}$$
$$= \int_0^1 dx_a dx_b f_a(x_a, M^2) f_b(x_b, M^2) \delta(x_a x_b - \tau)$$
$$= \frac{1}{S} \int_\tau^1 \frac{dy}{y} f_a(y, M^2) f_b\left(\frac{\tau}{y}, M^2\right)$$

Gluon luminosity

NNPDF3.0 / CT14 / MMHT14

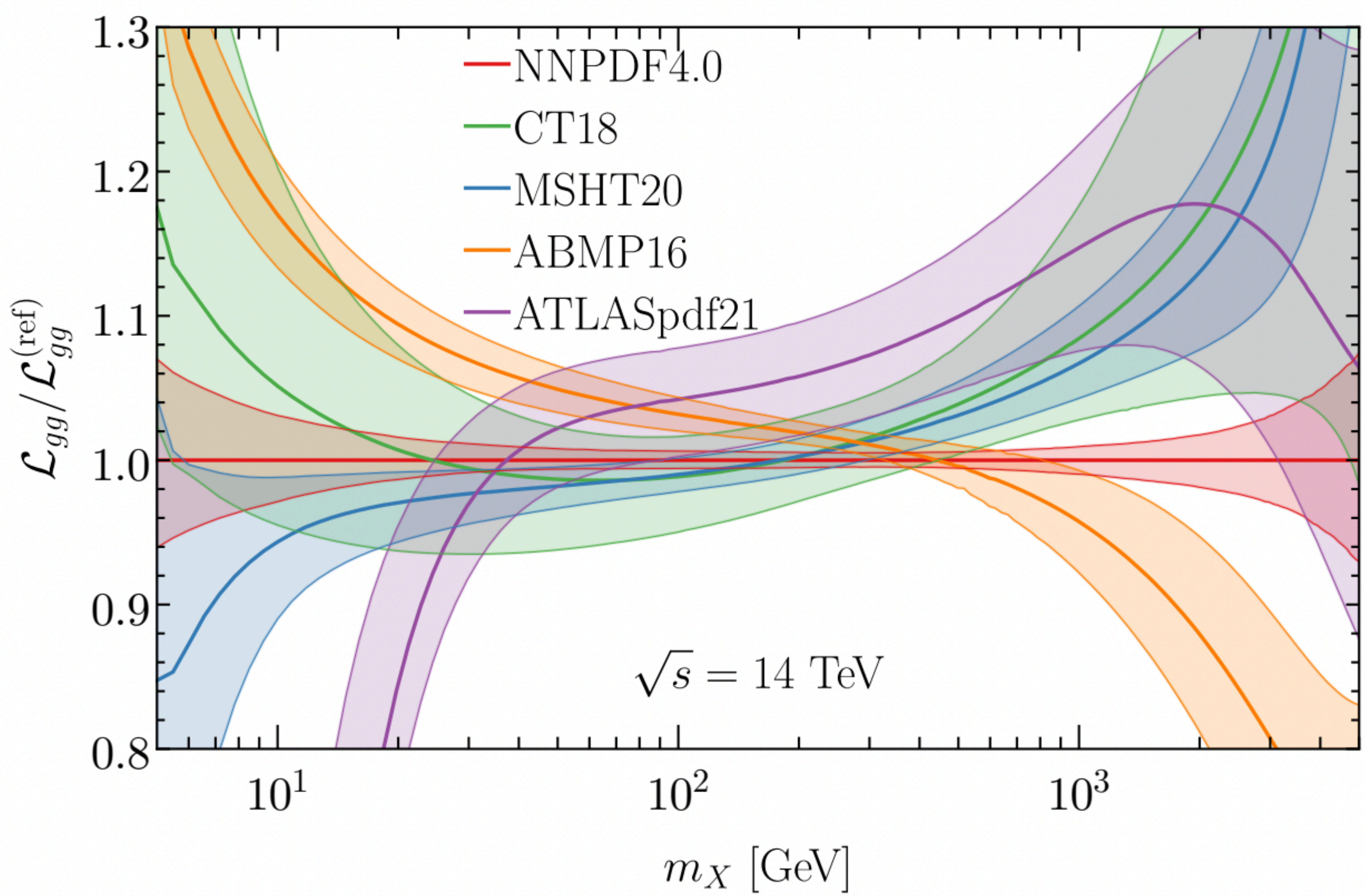


(2016)

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

Gluon luminosity

NNPDF4.0/ CT18/MSHT20/ABMP16/ATLASpdf21

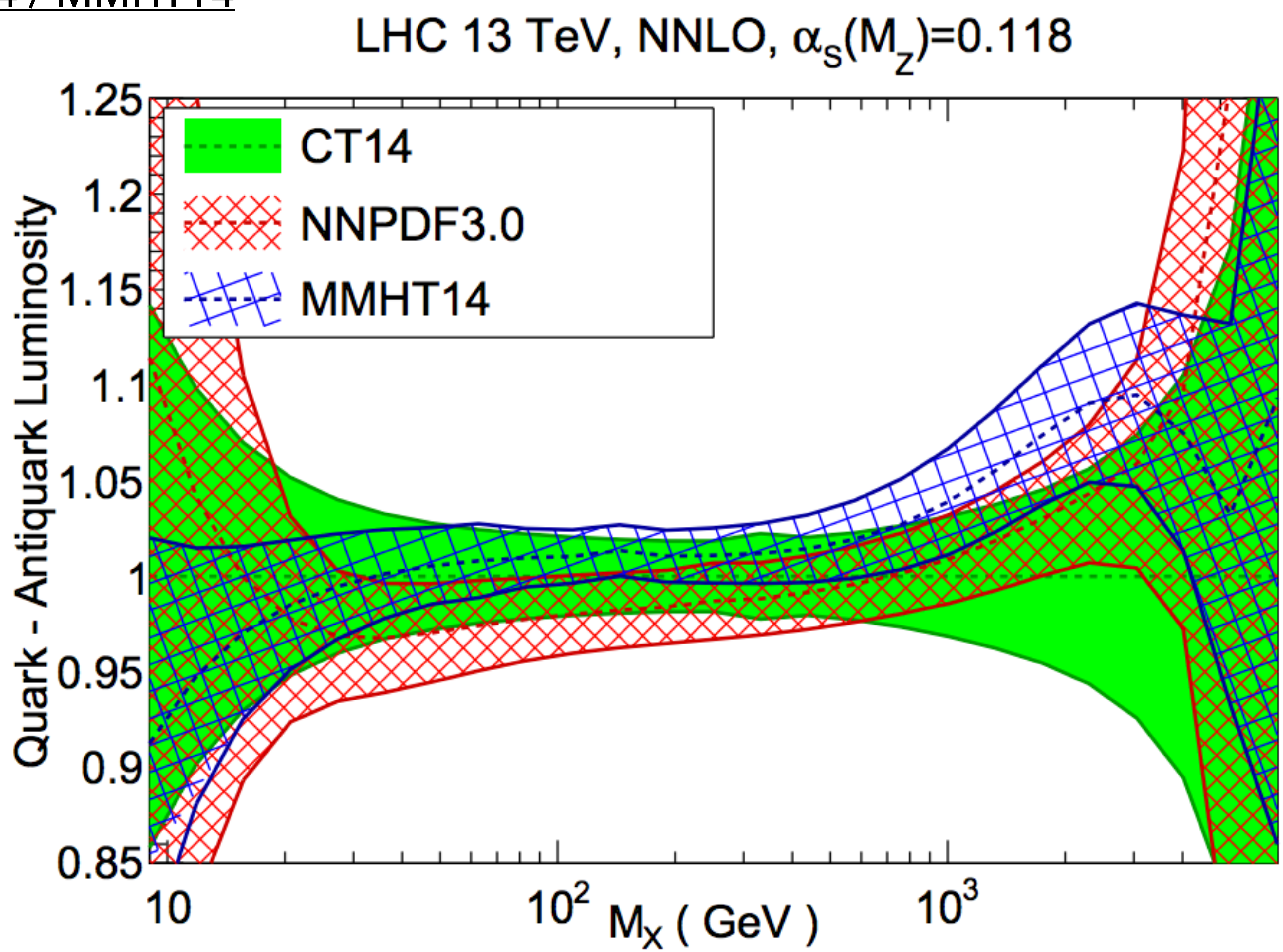


(2022)

Snowmass 2022 white paper, arXiv: 2203.13923

Quark-Antiquark luminosity

NNPDF3.0 / CT14 / MMHT14

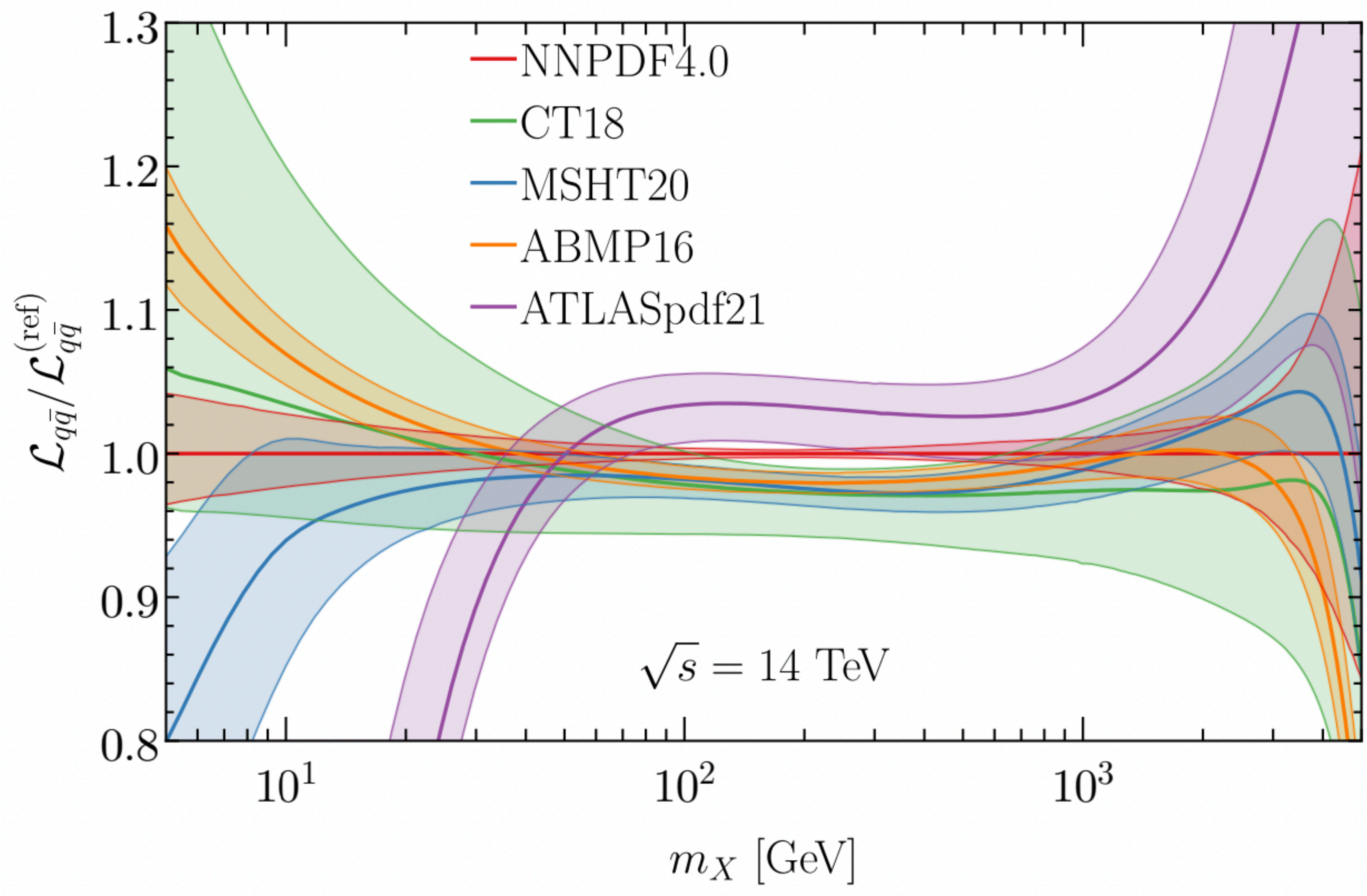


(2016)

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

Quark-Antiquark luminosity

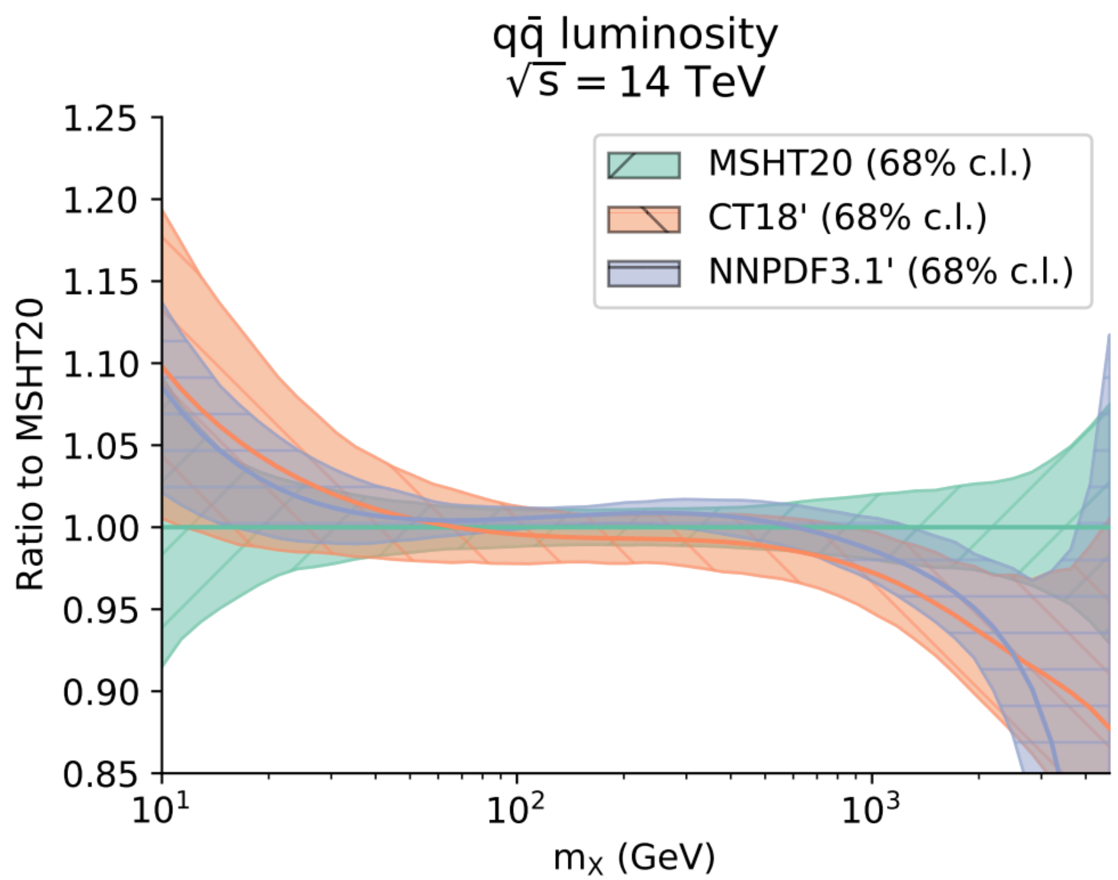
NNPDF4.0/ CT18/MSHT20/ABMP16/ATLASpdf21

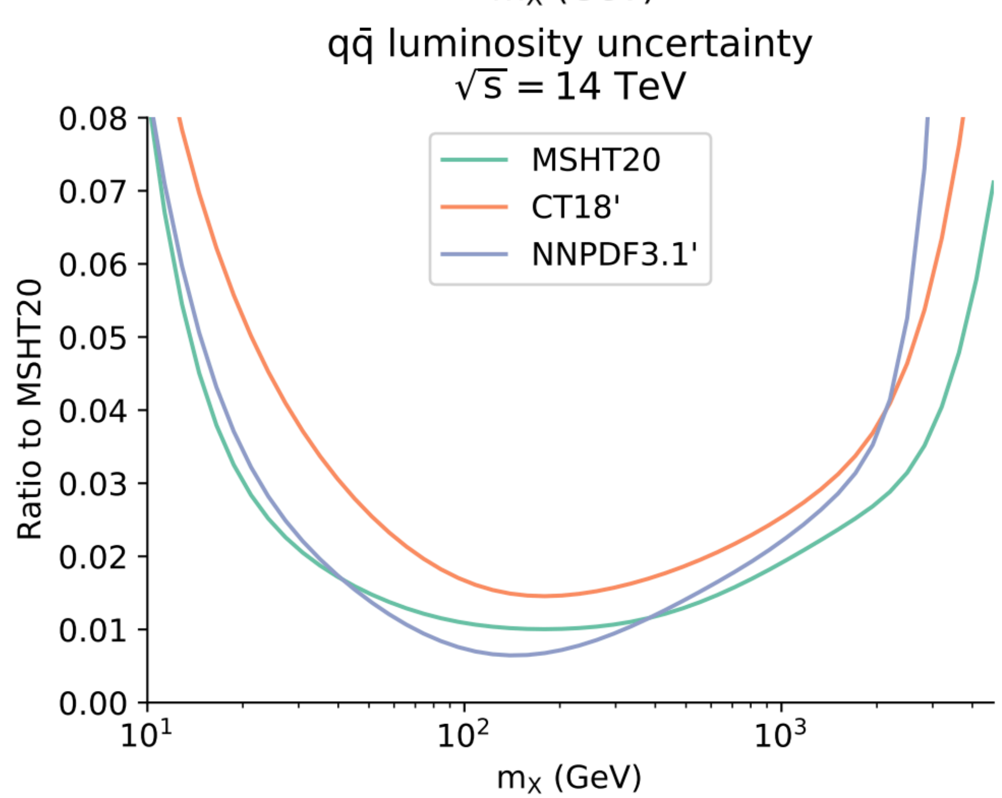


(2022)

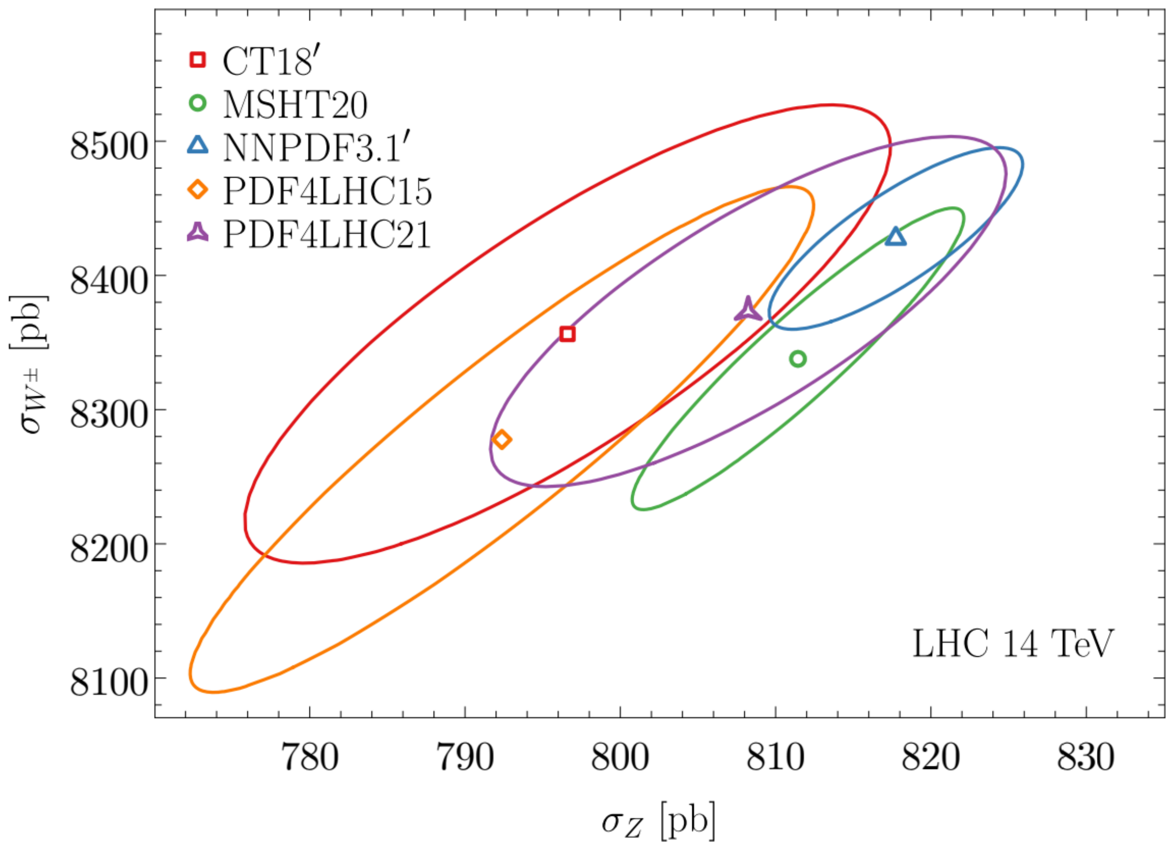
Snowmass 2022 white paper, arXiv: 2203.13923

Benchmarks

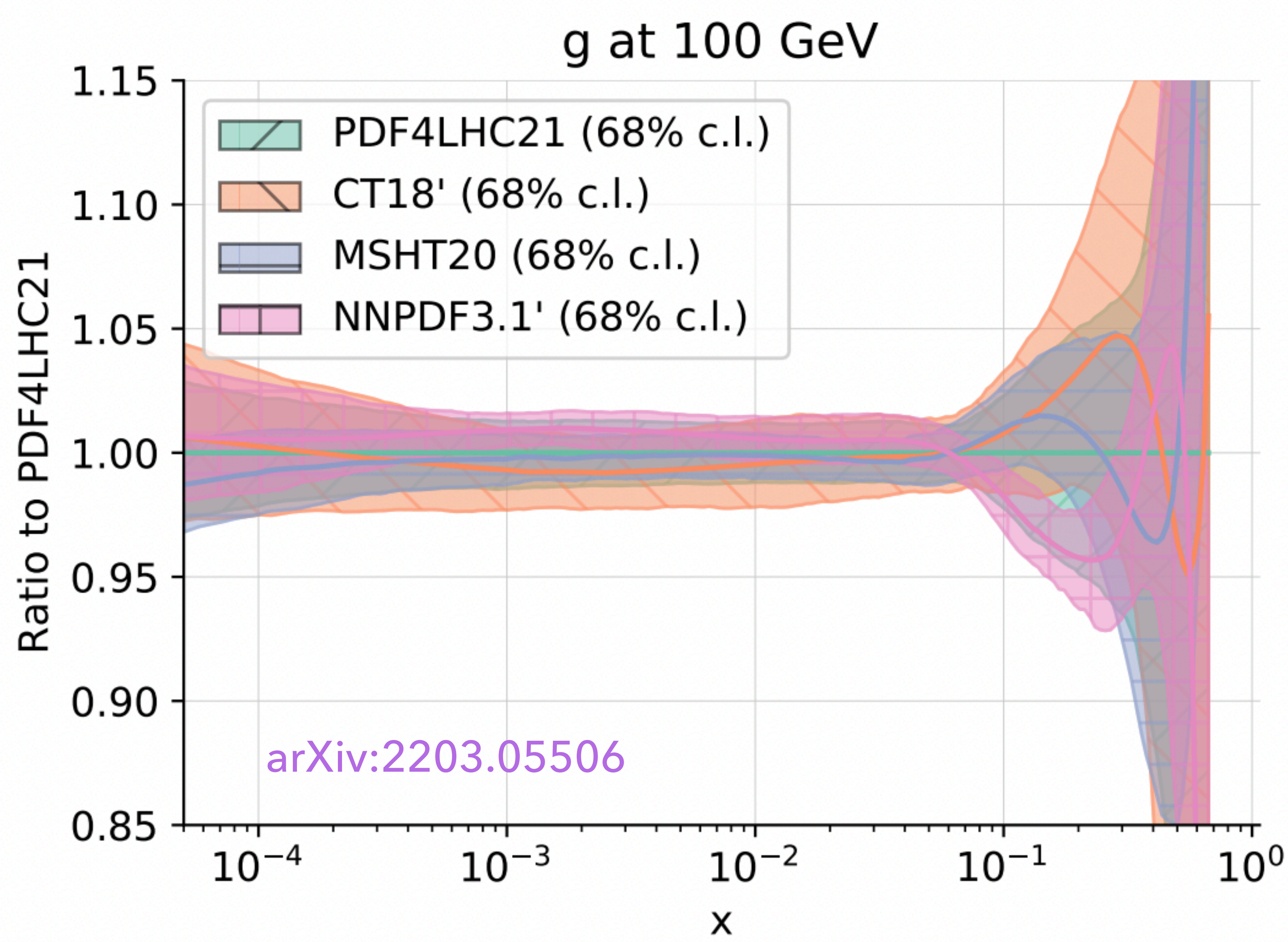




- Benchmark exercise among NNPDF3.1, MSHT20 and CT18 at the basis of PDF4LHC combination
- Overall agreement, which improves once common dataset is used, differences in uncertainties with ΔCT ≳ ΔMHST ≳ ΔNN due to methodology

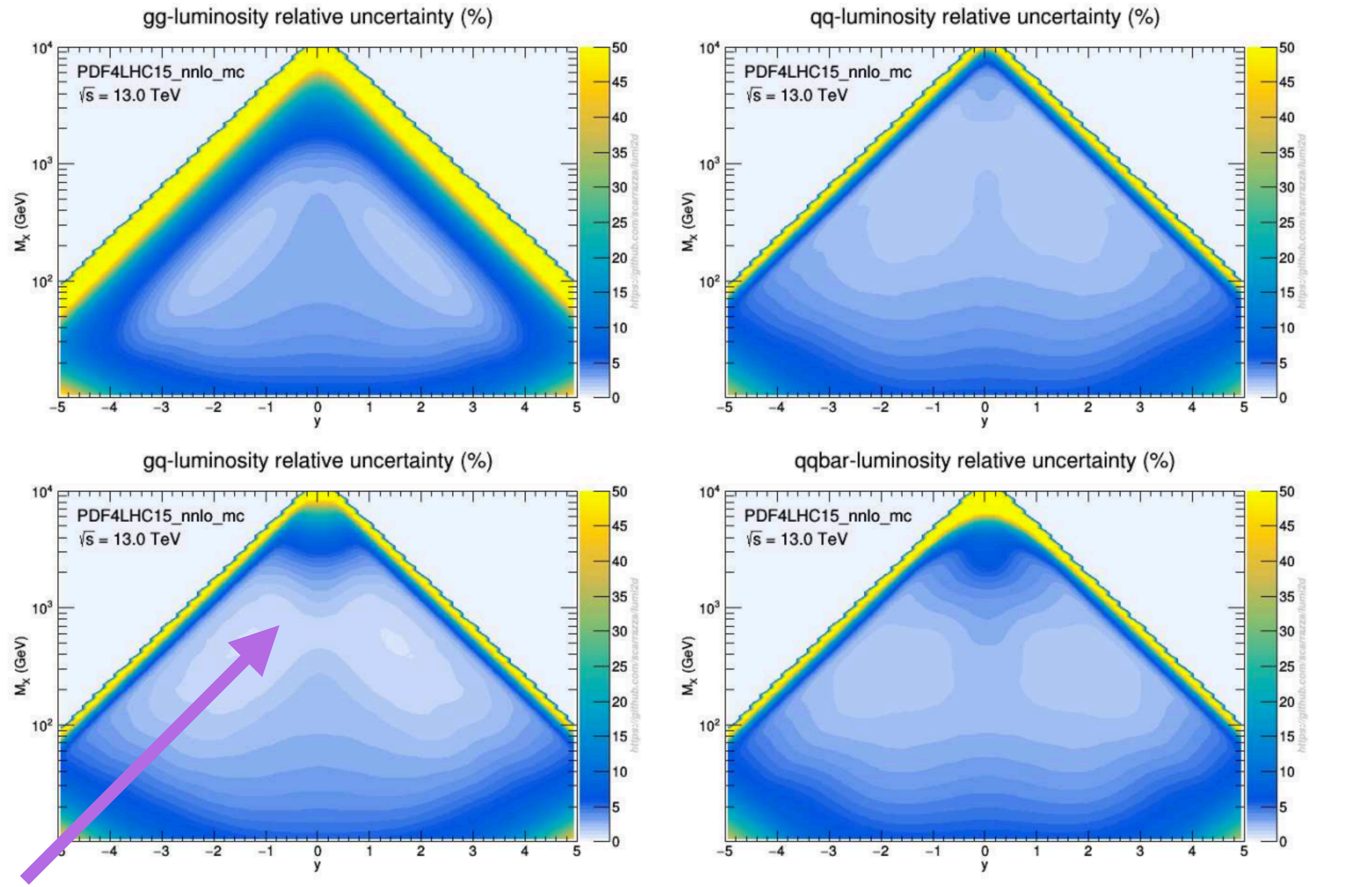


PDF4LHC21 combination



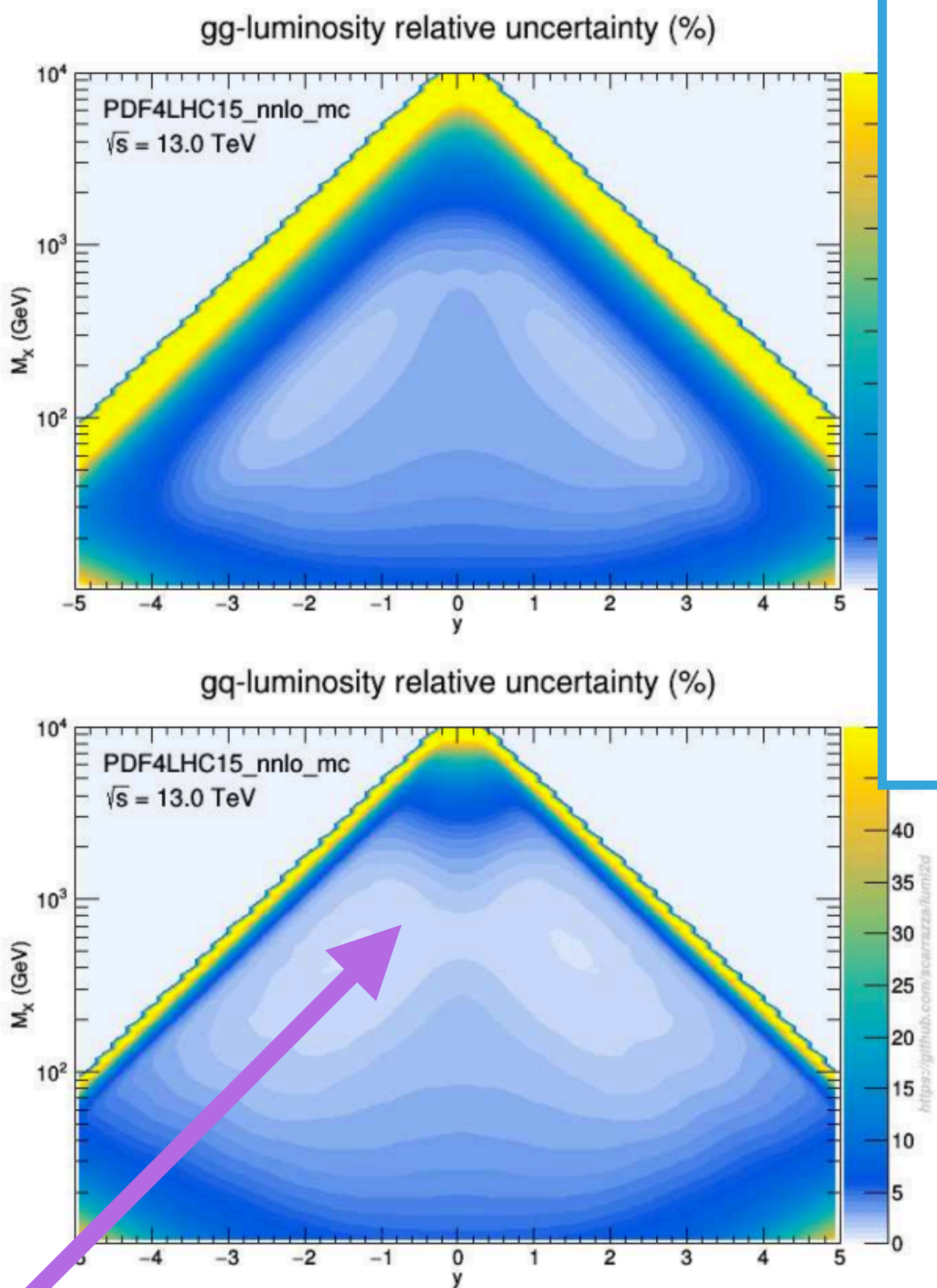
- Combination of PDF fits done on same dataset
- Generate 300 replicas of each PDF set
- Take them together
- Reduce 900 replicas to O(50) by standard compression method

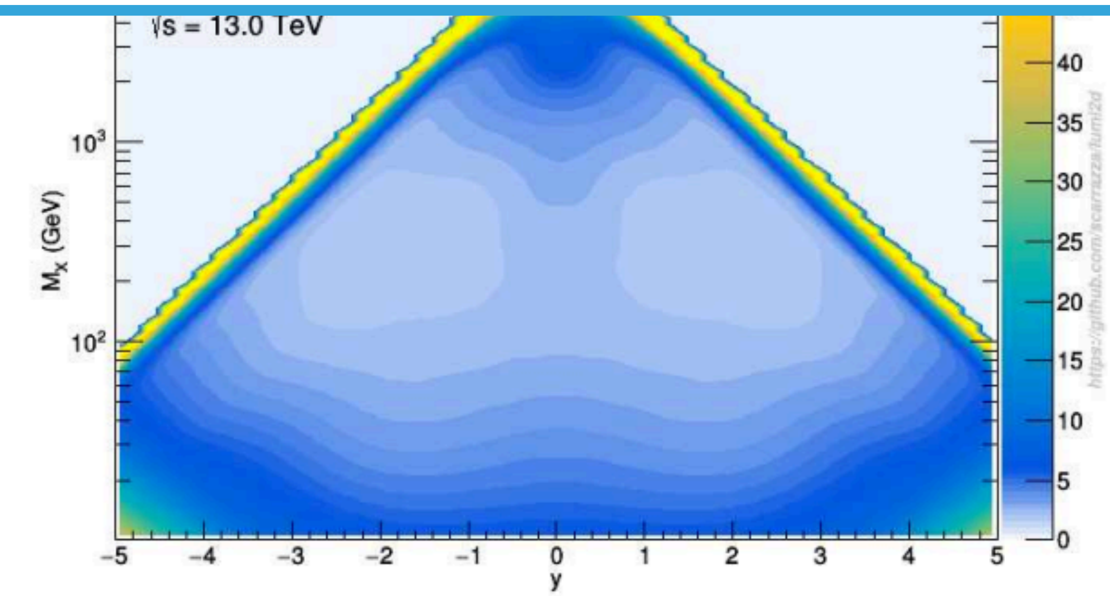
The precision frontier

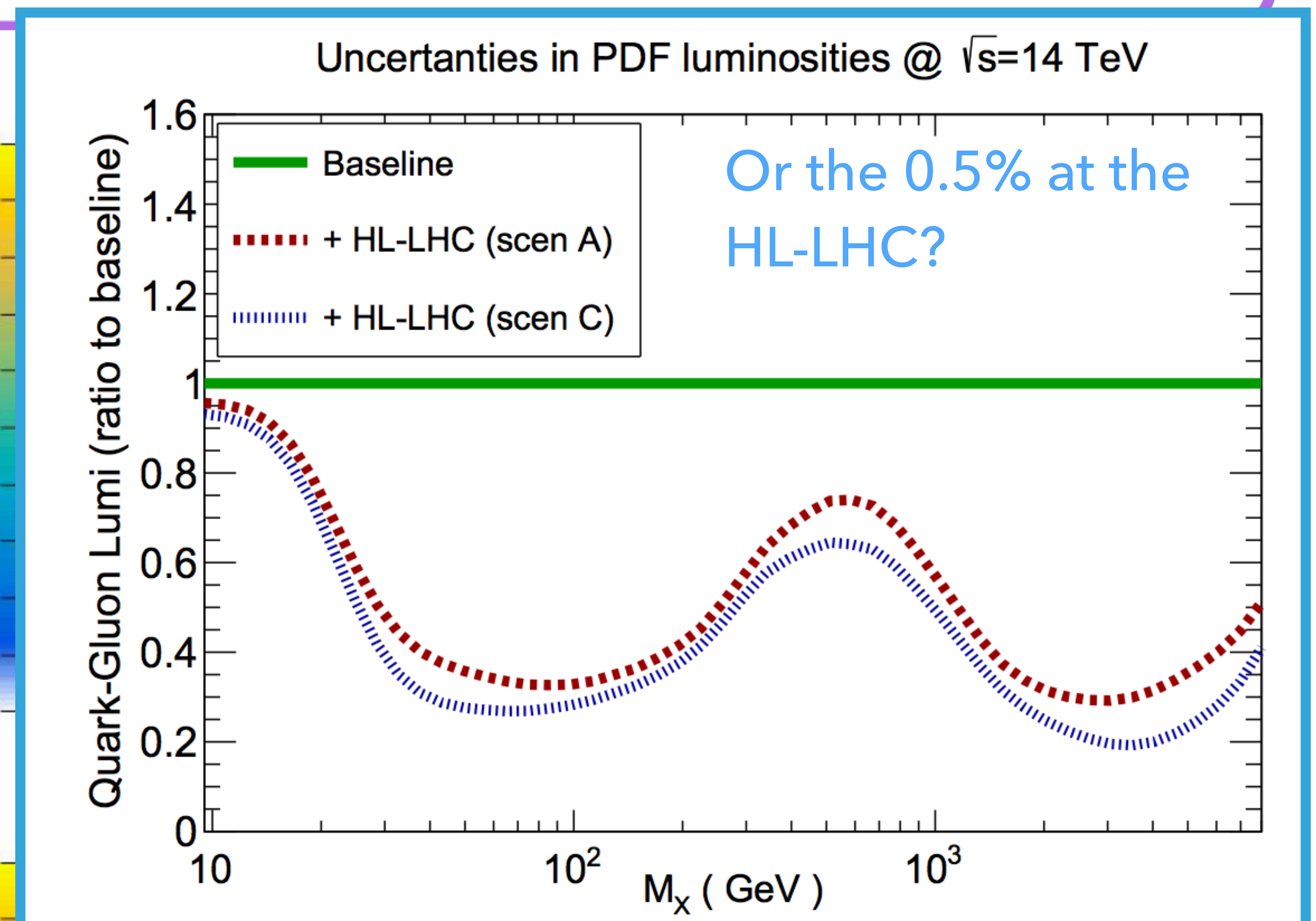


Can we trust 1% accuracy?

The precision frontier



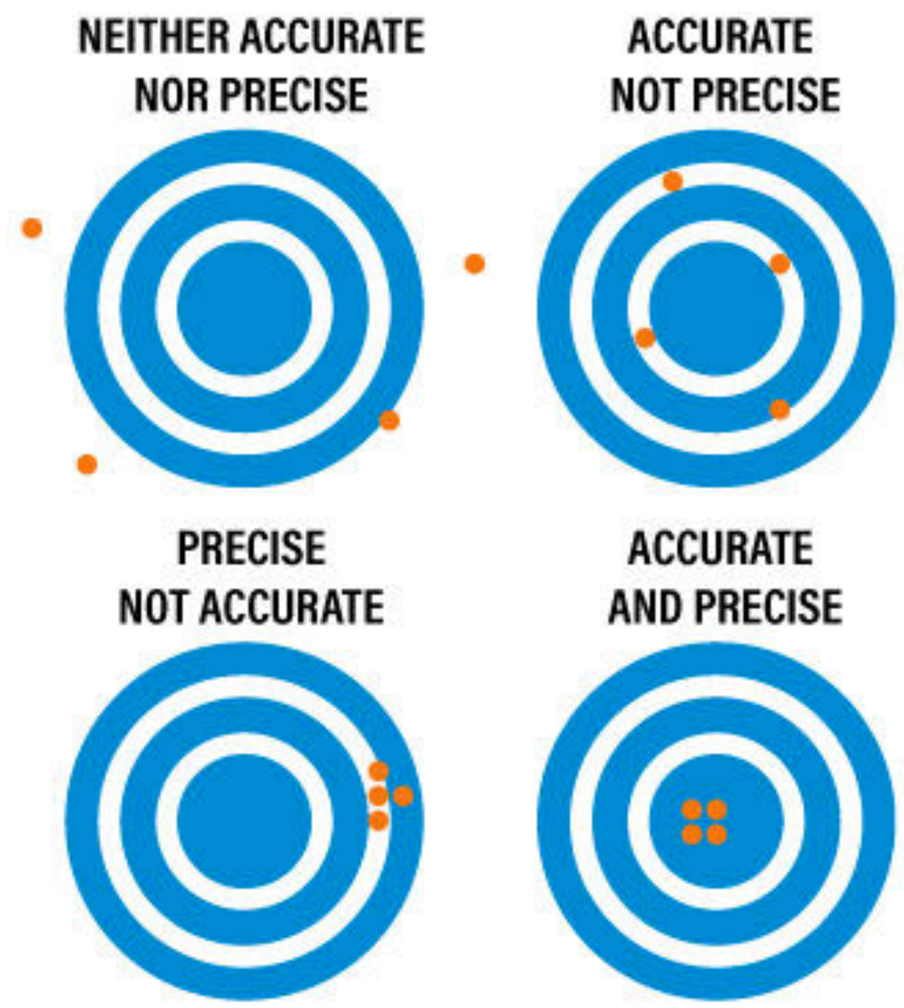


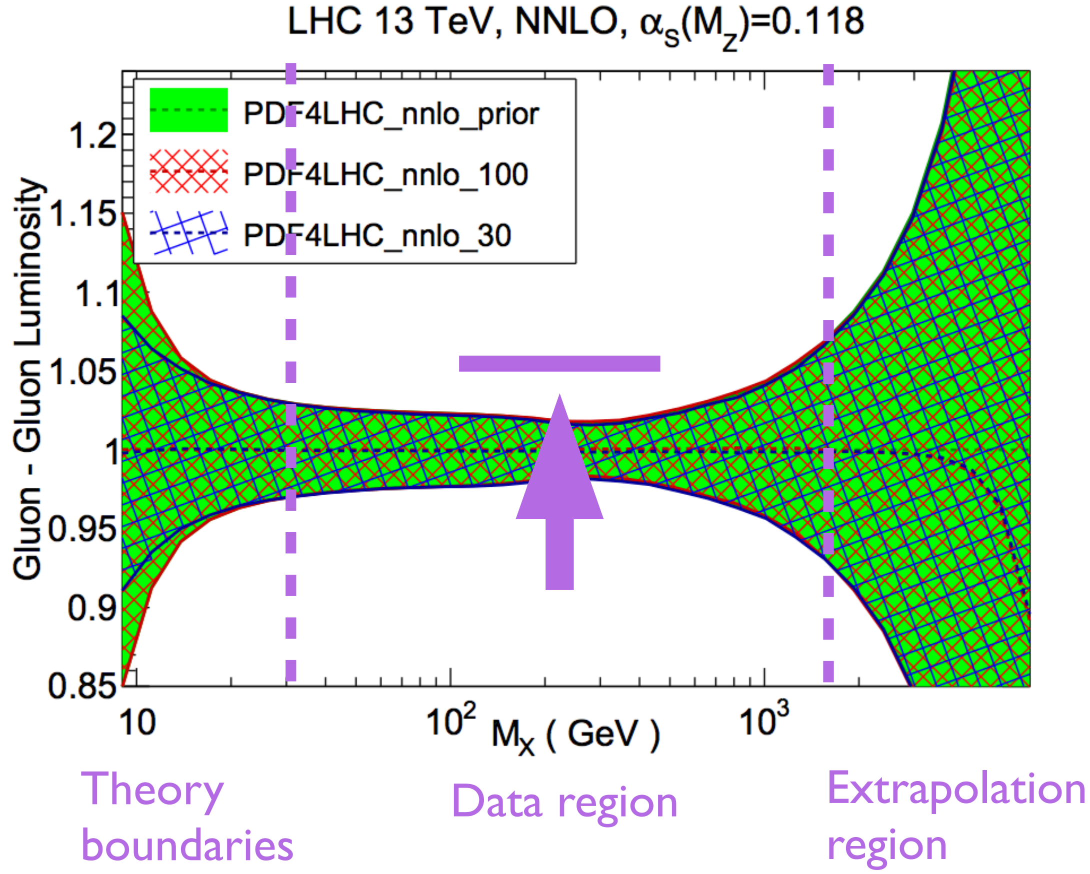


Can we trust 1% accuracy?

Theory uncertainties

On top of benchmarking different PDF sets, each set must deal with inconsistencies in updated determinations.

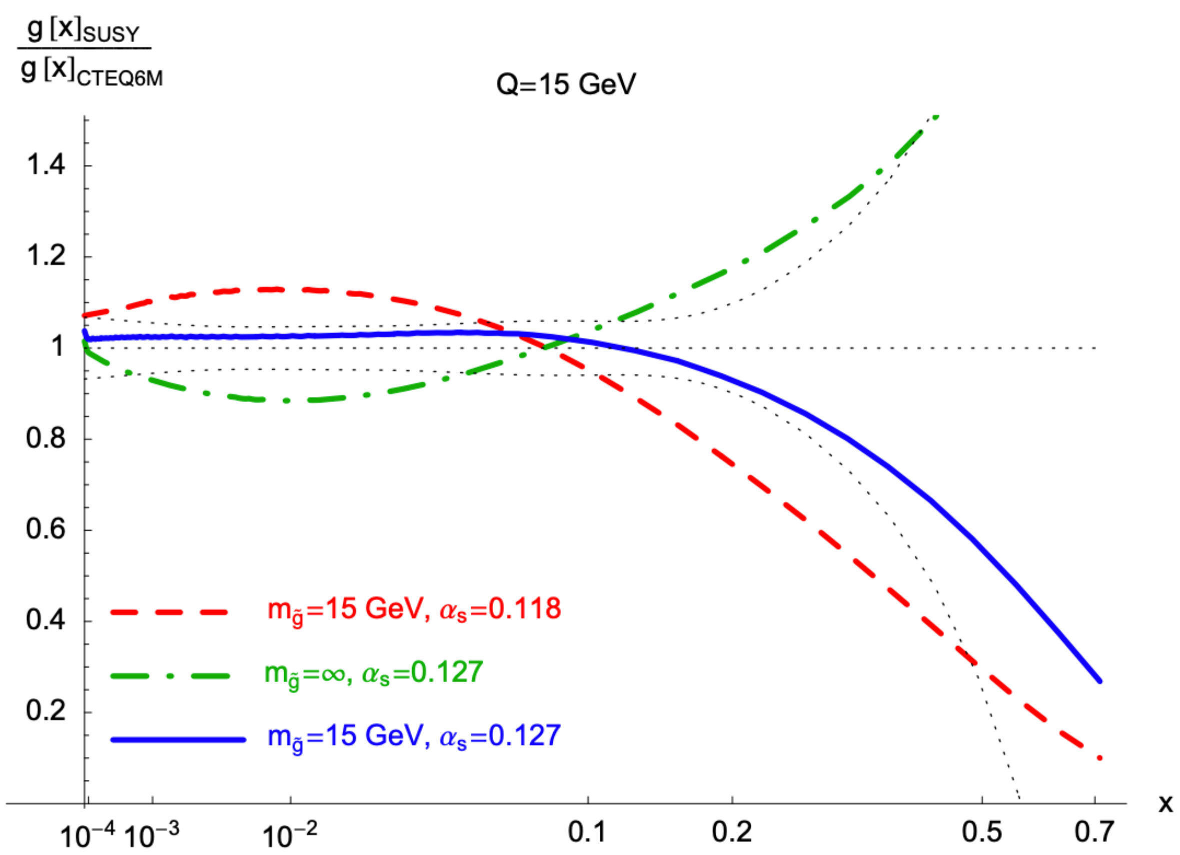




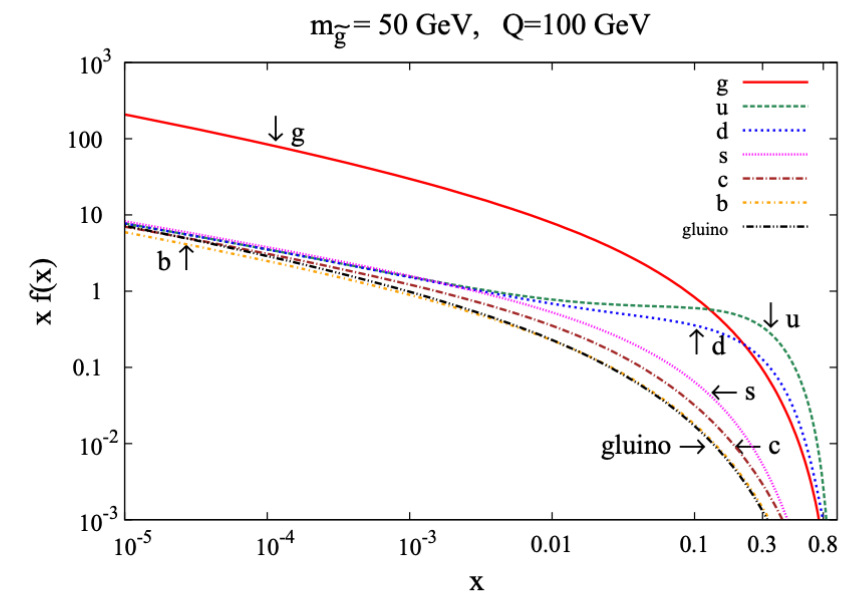
- Inconsistent data → Tolerance/ Statistical estimators
- Updated parametrization
- Differences in fitting methodology/minimisation? → Closure Test
- Changes in theory? → MHOU/New Physics

PDF fits and New Physics interplay

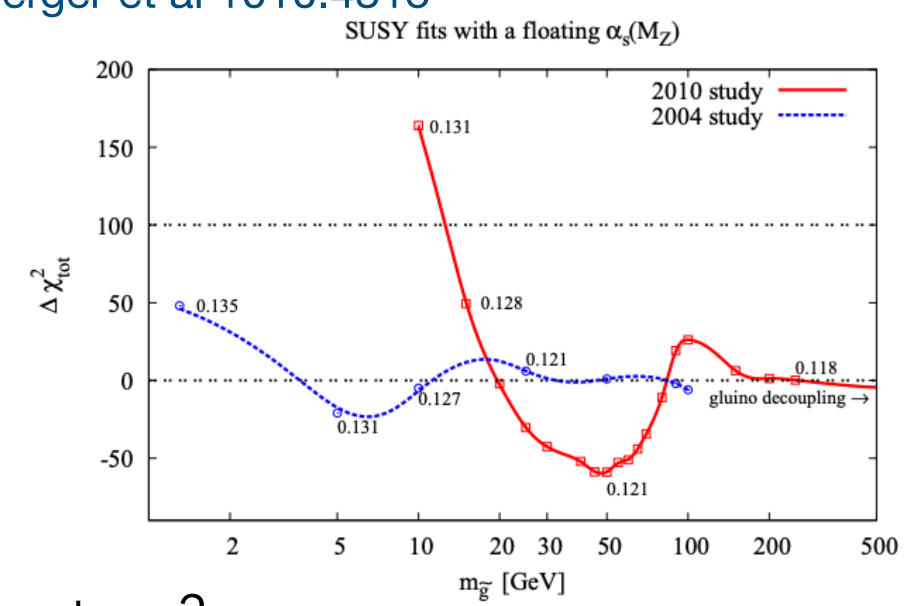
1. New partons



Berger et al hep-ph/0406143



Berger et al 1010.4315



- Pre-LHC studies: what is there was a light SUSY coloured partner?
- A light SUSY Parton would modify DGLAP equation and running of α_s
- Comparison to data excludes any light coloured parton on increasing mass range as more (and more precise) data are included in the global PDF fit

1. New partons

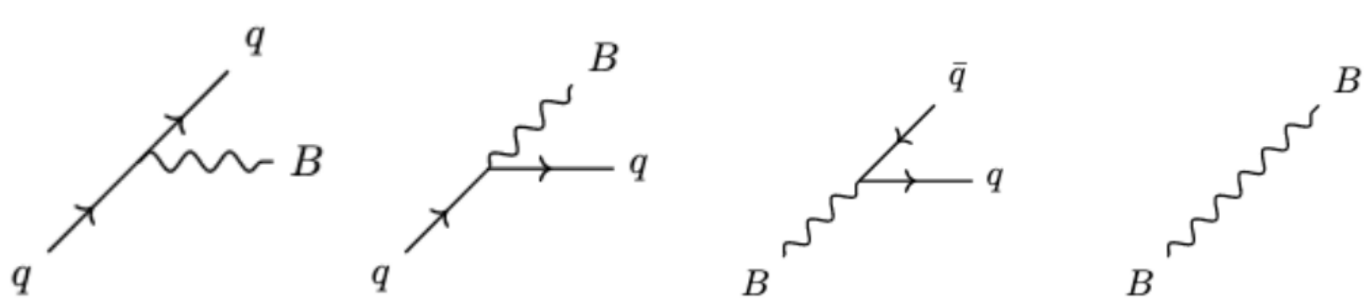
M. McCullough, J. Moore, MU, arXiv:2203.12628

- Idea: now PDFs are known very precisely, and their uncertainties will continue to reduce in the near future with the HL-LHC, could we do the same for a colourless particle too?
- If there was a lepto-phobic dark photon weakly coupled to quarks via effective Lagrangian

$$\mathcal{L}_{\text{int}} = \frac{1}{3} g_B \bar{q} \not{B} q \qquad m_B \in [2{,}80]\ \text{GeV}$$

it would appear among the partons of the proton.

- To include the dark photon as a constituent of the proton: compute the dark photon splitting functions, and add them to DGLAP evolution. Starting from an appropriate initial-scale ansatz (dark photon generated dynamically off quarks and antiquarks at threshold) and a reference PDF set, evolve using the modified DGLAP equations

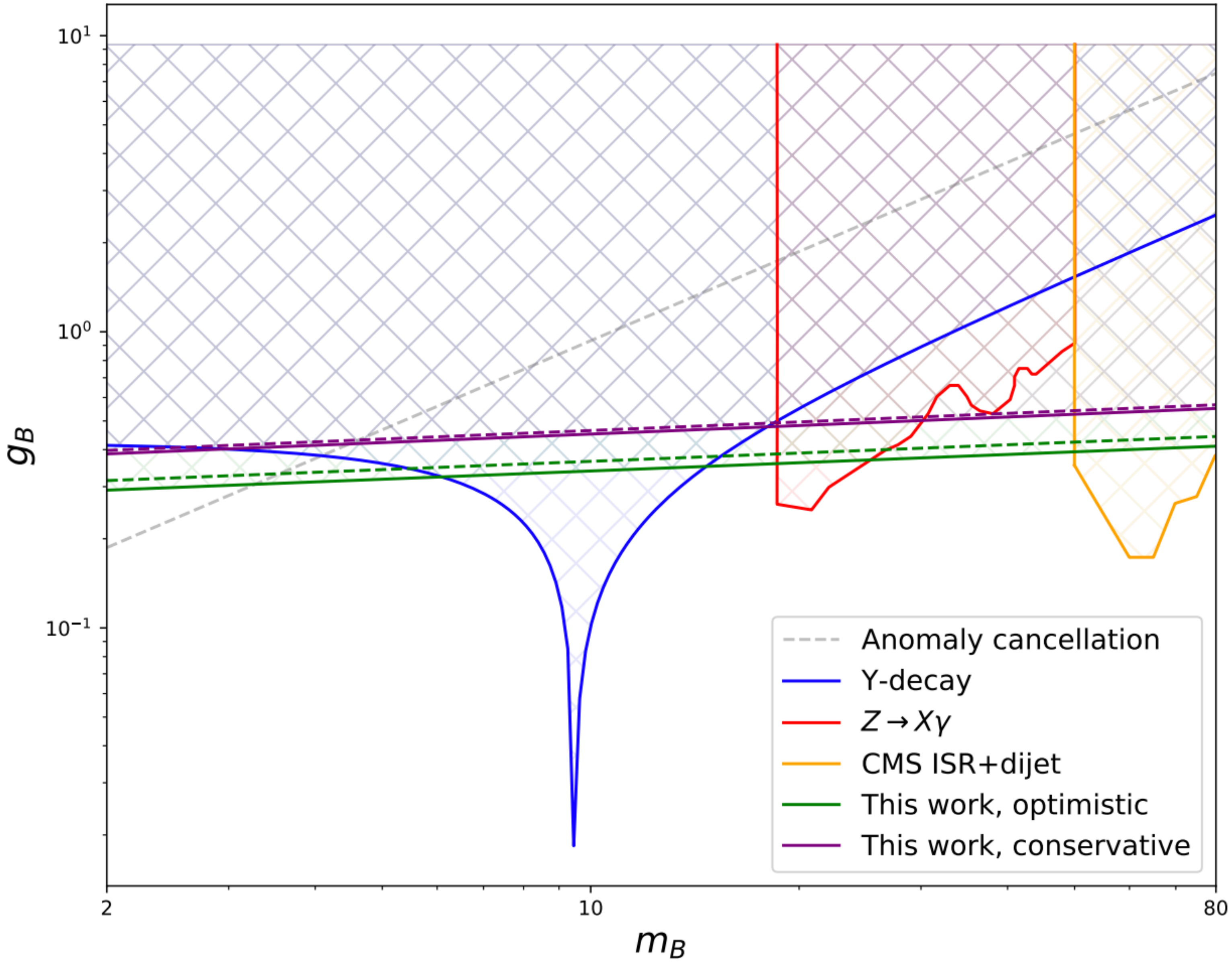


$$P_{ij} = \left(\frac{\alpha_s}{2\pi}\right) P_{ij}^{(1,0,0)} + \left(\frac{\alpha_s}{2\pi}\right)^2 P_{ij}^{(2,0,0)} + \left(\frac{\alpha_s}{2\pi}\right)^3 P_{ij}^{(3,0,0)} + \left(\frac{\alpha}{2\pi}\right) P_{ij}^{(0,1,0)} + \left(\frac{\alpha_s}{2\pi}\right)\left(\frac{\alpha}{2\pi}\right) P_{ij}^{(1,1,0)} + \left(\frac{\alpha}{2\pi}\right)^2 P_{ij}^{(0,2,0)} + \left(\frac{\alpha_B}{2\pi}\right) P_{ij}^{(0,0,1)} + \cdots ,$$

$$\alpha_B \sim 0.001$$

1. New partons

M. McCullough, J. Moore, MU, arXiv:2203.12628



- The presence of the dark Parton would modify the evolution of standard quarks and gluon.

Precise LHC data can indirectly constrain parameter space of the dark photon in a competitive way compared to direct searches

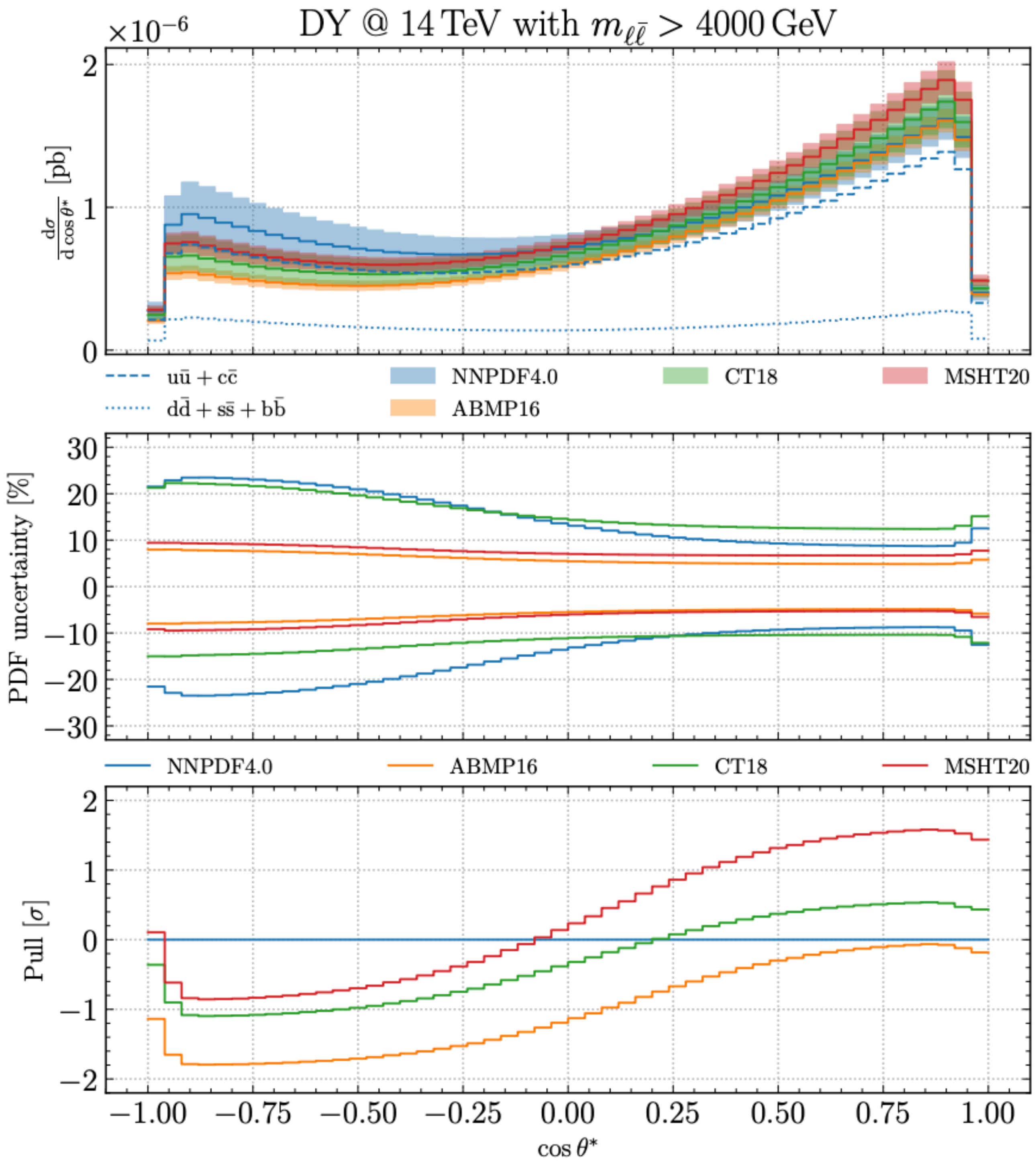
2. Large-x PDFs and new physics

✓ High mass Drell-Yan tails affected by large PDF uncertainties

✓ This affects searches for new physics, for example in forward-backward asymmetry

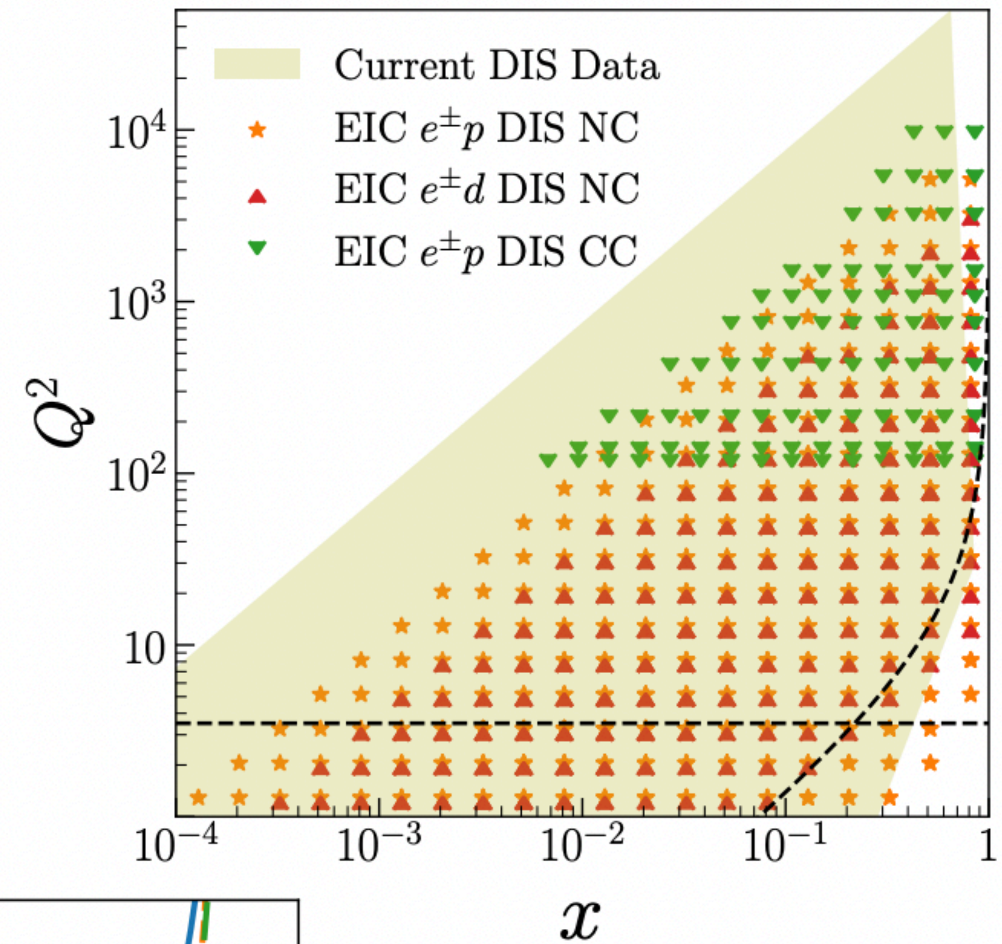
✓ Need data constraining large-x to see and characterise new physics
(at the LHC high energy - high-x)

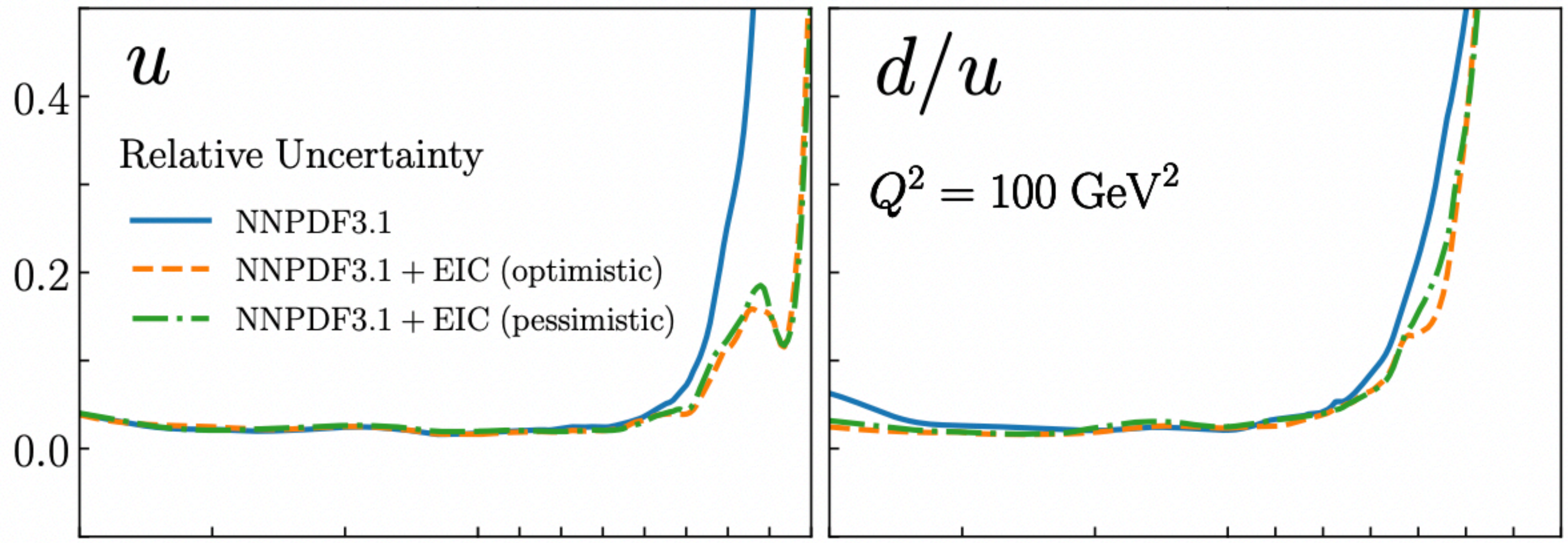
$$\frac{d\sigma}{d\cos\theta^*} = \int_{m_{\ell\bar{\ell}}^{\min}}^{\sqrt{s}} dm_{\ell\bar{\ell}} \int_{\ln(m_{\ell\bar{\ell}}/\sqrt{s})}^{\ln(\sqrt{s}/m_{\ell\bar{\ell}})} dy_{\ell\bar{\ell}} \frac{d^3\sigma}{dm_{\ell\bar{\ell}}\, dy_{\ell\bar{\ell}}\, d\cos\theta^*}$$



2. Large-x PDFs and new physics

- √ High mass Drell-Yan tails affected by large PDF uncertainties
- √ This affects searches for new physics, for example in forward-backward asymmetry
- √ Need data constraining large-x to see and characterise new physics (at the LHC high energy - high-x)
- √ EIC crucial to give complementary constraints





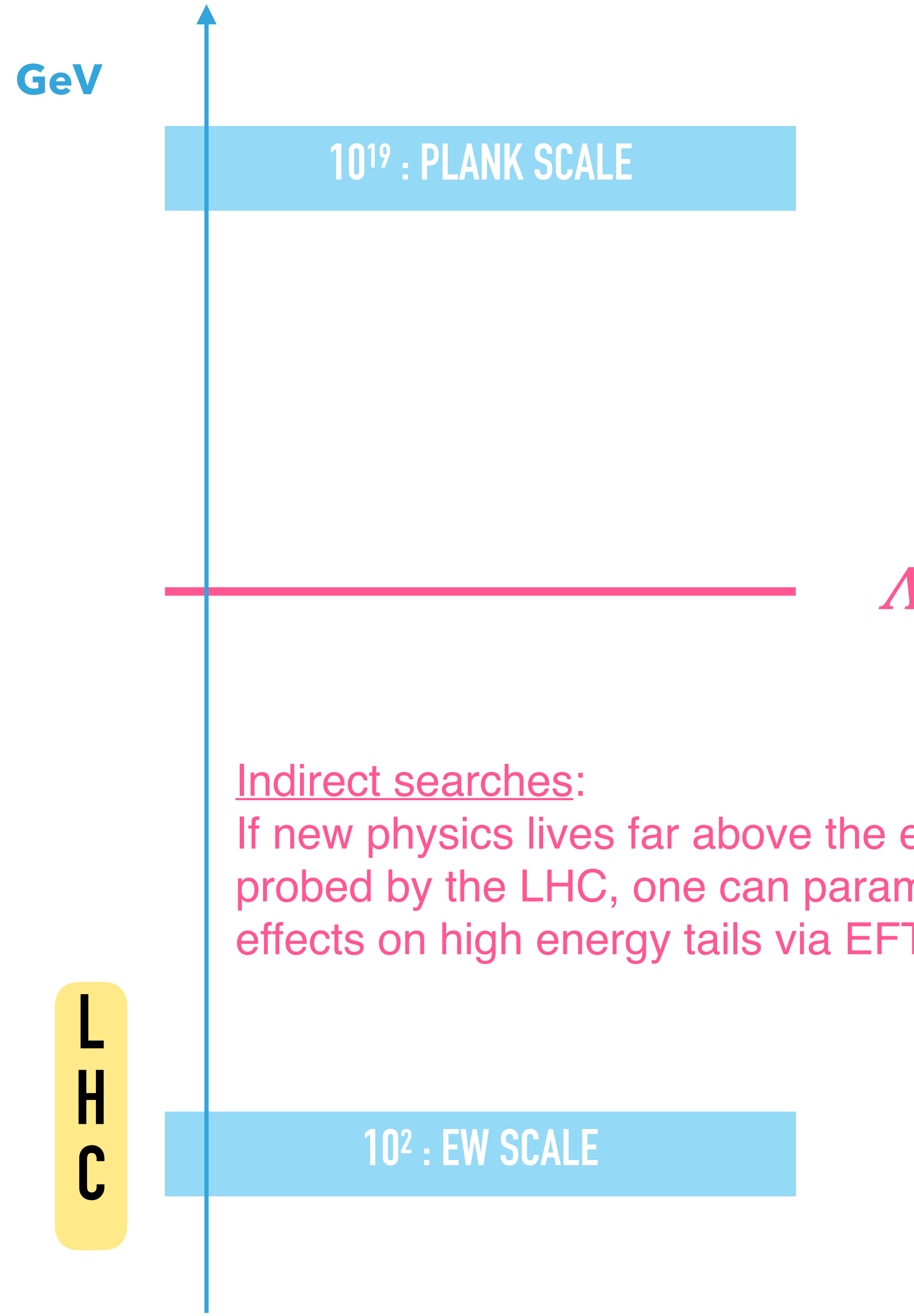
Rabah Abdul Khalek et al, arXiv:2102.00018

3. Indirect searches for New Physics

- EFT is a well-defined theoretical approach for indirect searches
- Assumption: new physics states are heavy
- Write the Lagrangian with only light SM particles
- BSM effects can be incorporated as a momentum expansion
- SMEFT: assume SM field content and gauge symmetries (apart from accidental)

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_{i}^{N_{d6}} \frac{c_i}{\Lambda^2}\mathcal{O}_i^{(6)} + \sum_{j}^{N_{d8}} \frac{b_j}{\Lambda^4}\mathcal{O}_j^{(8)} + \dots$$

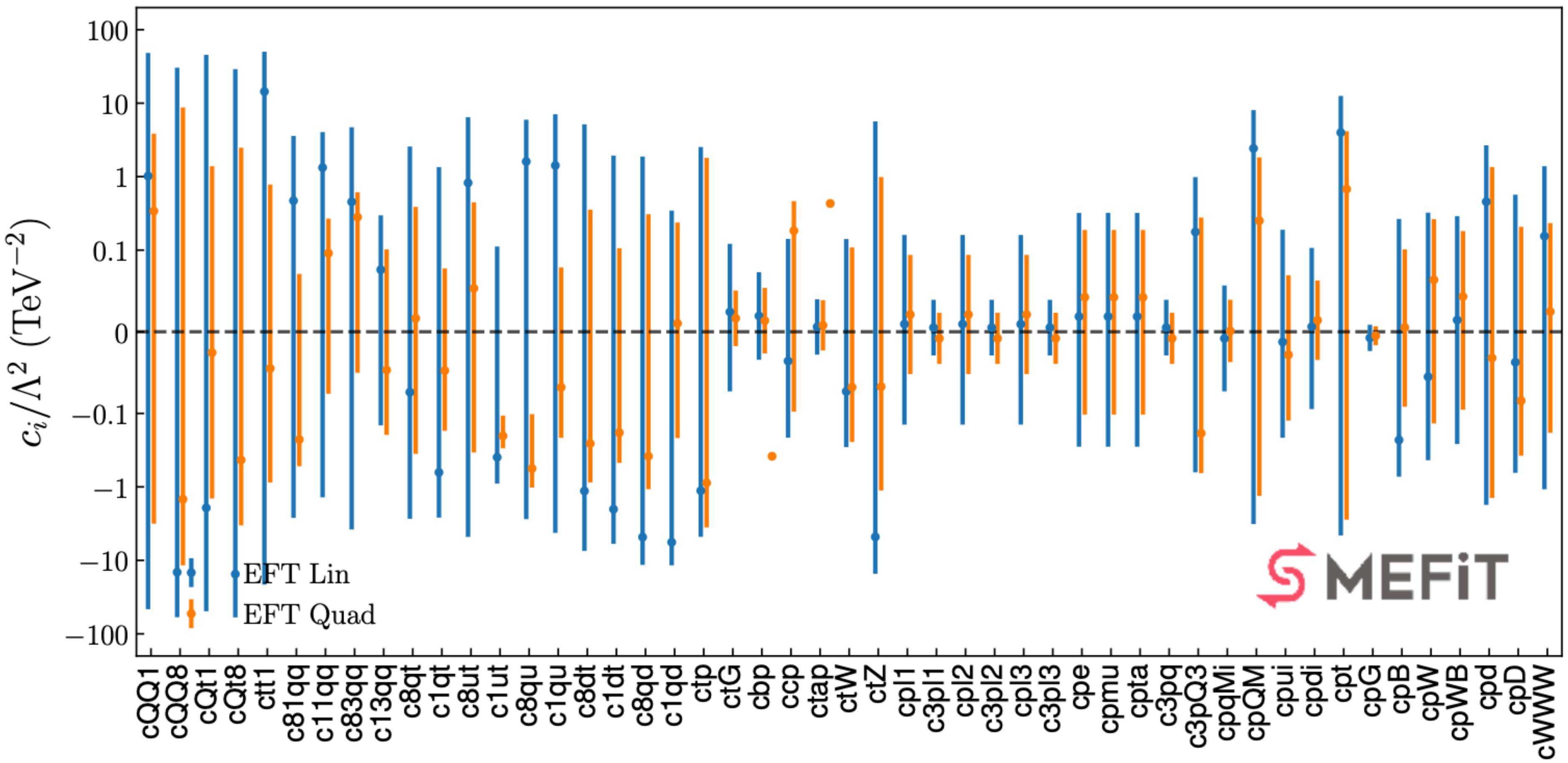
- Full dim-6 basis of operators under SMEFT assumptions includes **2499** operators **[Grzadkowski et al, arXiv:1008.4884]**



Indirect searches:
If new physics lives far above the energy probed by the LHC, one can parametric effects on high energy tails via EFT

3. Indirect searches for New Physics

- Current SMEFT fits make flavour assumptions and restricted to a few observables/sectors & reduce the number of operators.
- **SMEFiT**: SMEFT fit based on Monte Carlo technique for propagation of experimental uncertainty **[Hartland et al, arXiv:1901.05965]**
- Global dim-6 SMEFT fit of Higgs, diboson, and top quark production and decay measurements (36 independent Wilson coefficients, including linear and quadratic contributions and NLO QCD corrections to SMEFT) **[Either et al, arXiv:2105.00006]**



Either et al, arXiv:2105.00006

3. Indirect searches for New Physics

$$\chi^2 = \frac{1}{N_{\rm dat}} \sum_{i=1}^{N_{\rm dat}} (T_i(\{\theta\},\{c\}) - D_i)\, {\rm cov}_{ij}^{-1}\, (T_j(\{\theta\},\{c\}) - D_j)$$

$$T_i(\{\theta\},\{c\}) = {\rm PDFs}(\{\theta\},\{c\}) \otimes \hat{\sigma}_i(\{c\})$$

(B)SM parameters: $\alpha_s(M_Z)$, M_W, θ_W ,**SMEFT WCs**……

Parameters determining PDFs at initial scale

✓ In a PDF fit typically

$$T_i(\{\theta\}) = {\rm PDFs}(\{\theta\},\{c=0\}) \otimes \hat{\sigma}_i(\{c=0\})$$

✓ In a fit of SMEFT Wilson Coefficients

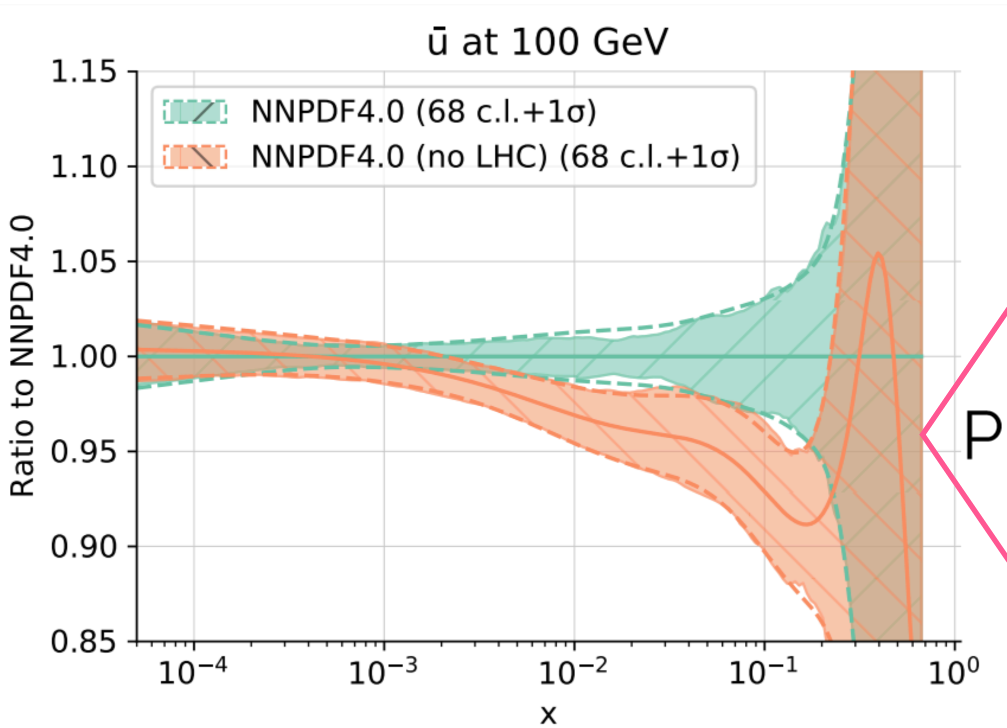
$$T_i(\{c\}) = {\rm PDFs}(\{\theta=\bar{\theta}\},\{c=0\}) \otimes \hat{\sigma}_i(\{c\})$$

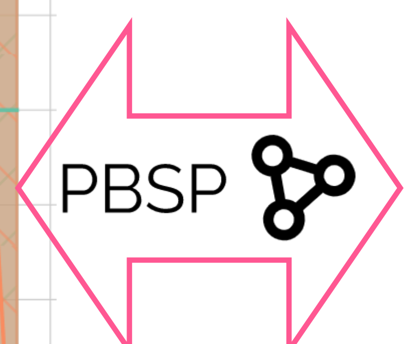
3. Indirect searches for New Physics

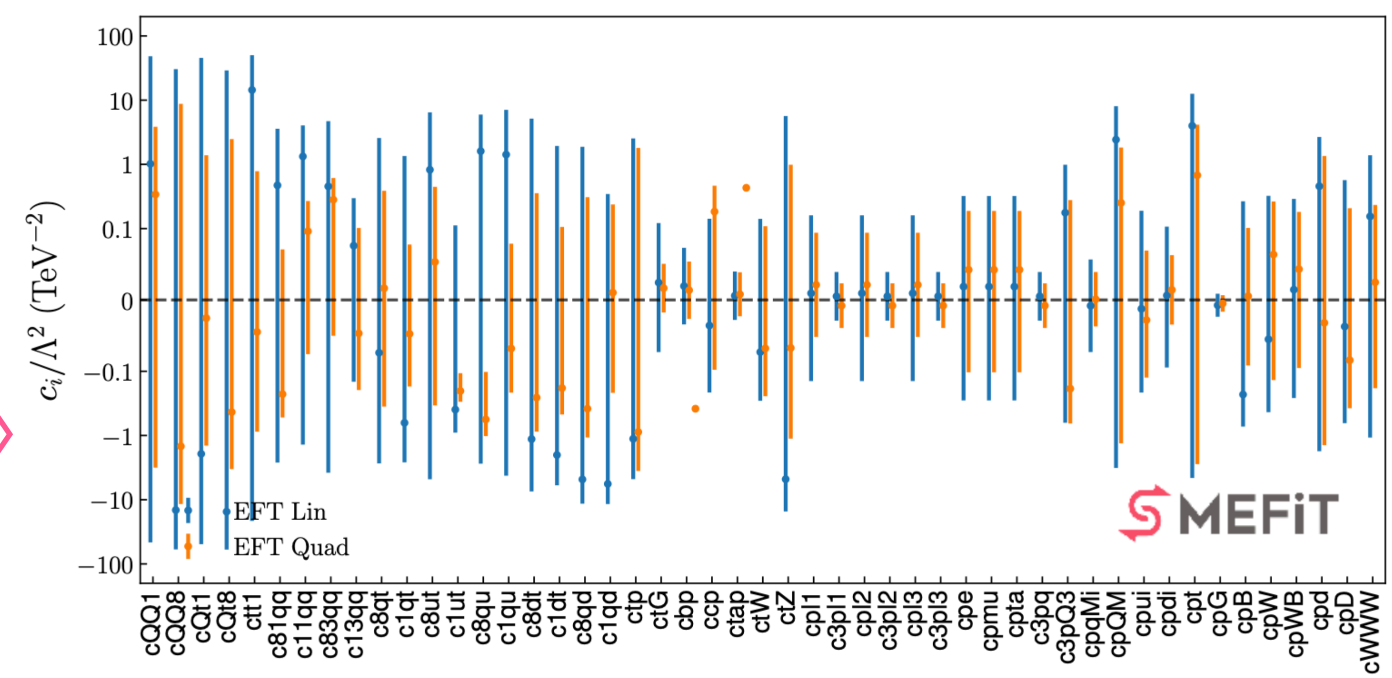
- In principle low-scale physics is separable from high-scale physics, BUT the complexity of the LHC environment might well intertwine them.

- PDFs are low-scale quantities extracted from experimental data at all scales, without considering any potential high-scale contamination due to new physics.

- (SM)EFT fits are performed by assuming a priori that PDFs are SM-like.





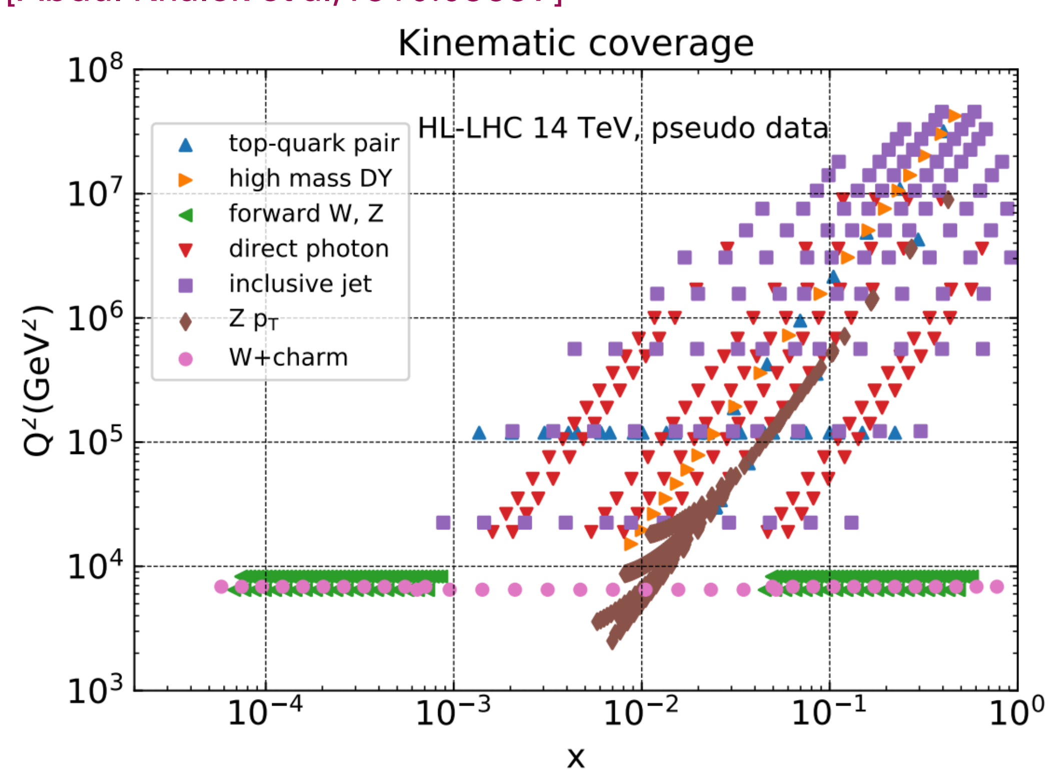


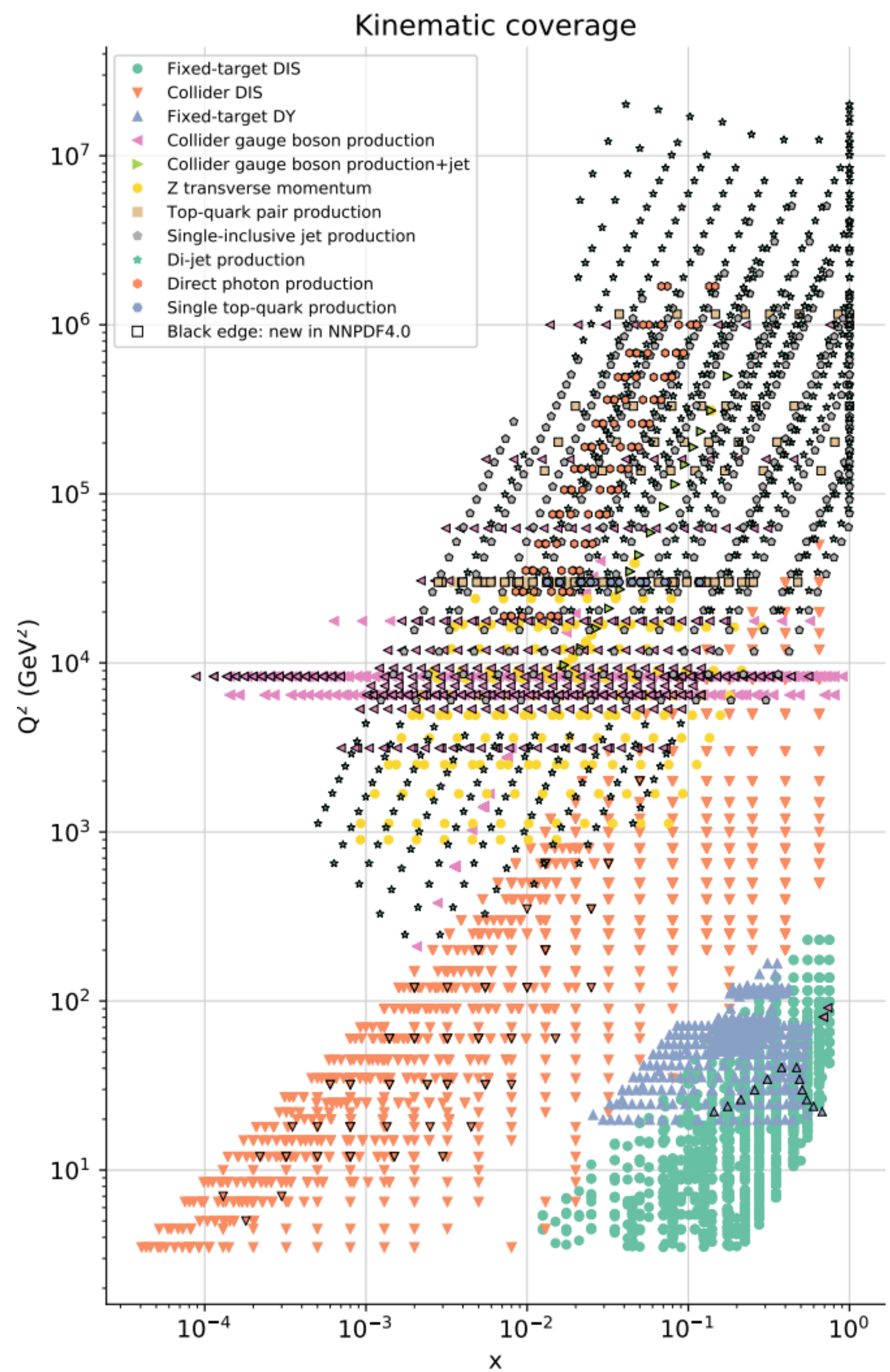
Ethier et al, arXiv: 2105.00006

Ball et al, arXiv:2109.02653

Data overlap

- ➡ Top pair production and single top data included in SMEFT analysis [Hartland et al 1901.05965] [Ellis et al 2012.02779]
- ➡ Dijets data [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 al,1810.03639]





3. Indirect searches for New Physics

- From the point of view of PDF fits:
  - ➡ How to make sure that new physics effects are not inadvertently fitted away in a PDF fit?

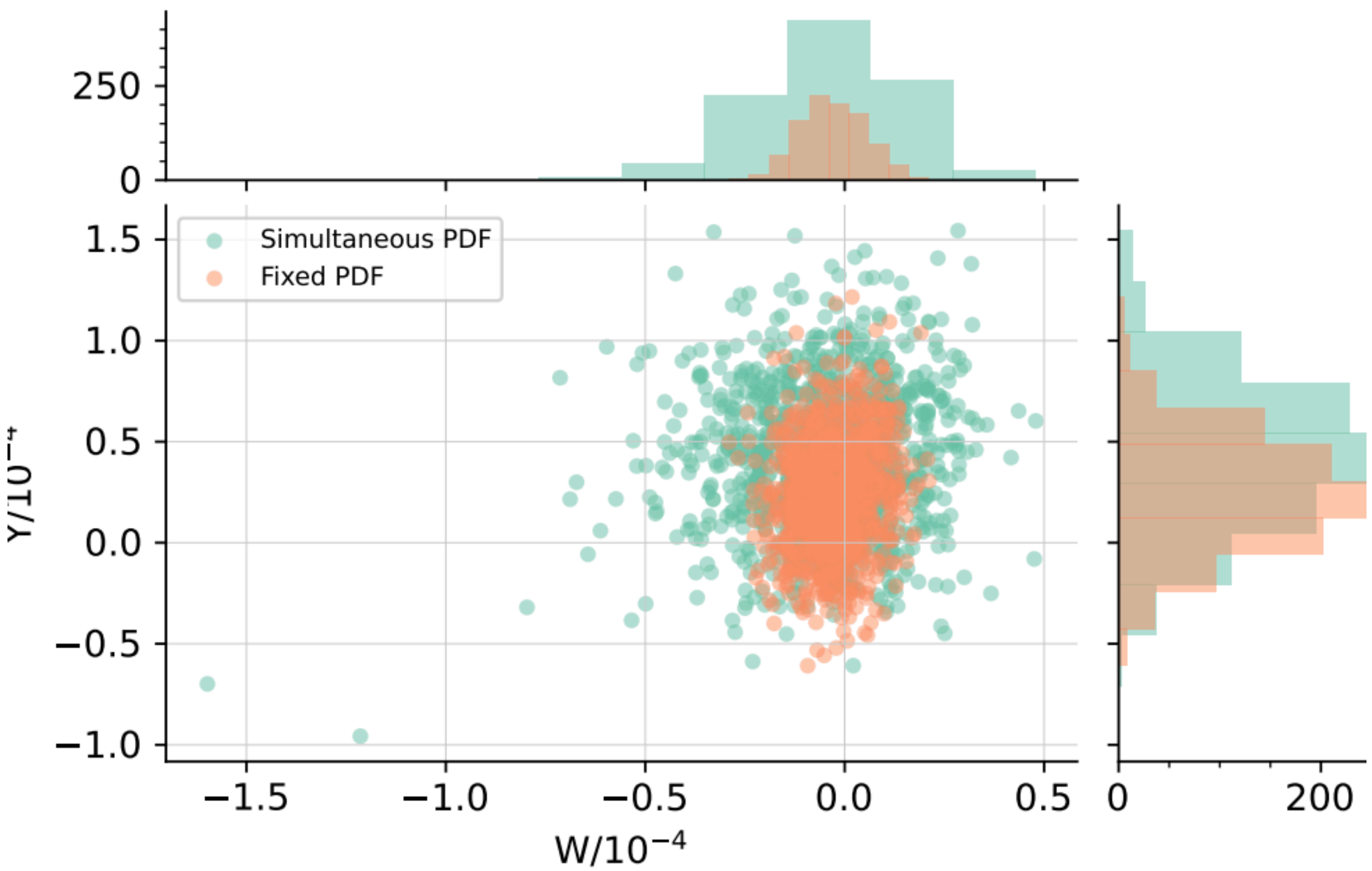
- From the point of view of SMEFT fits:
  - ➡ Should I make sure I am using a clean set of PDFs in a SMEFT analysis? How to define it? Is it enough?
  - ➡ How would the bounds change if I was consistently using PDFs that include in the fit theory predictions computed adding the same operators that I am fitting?

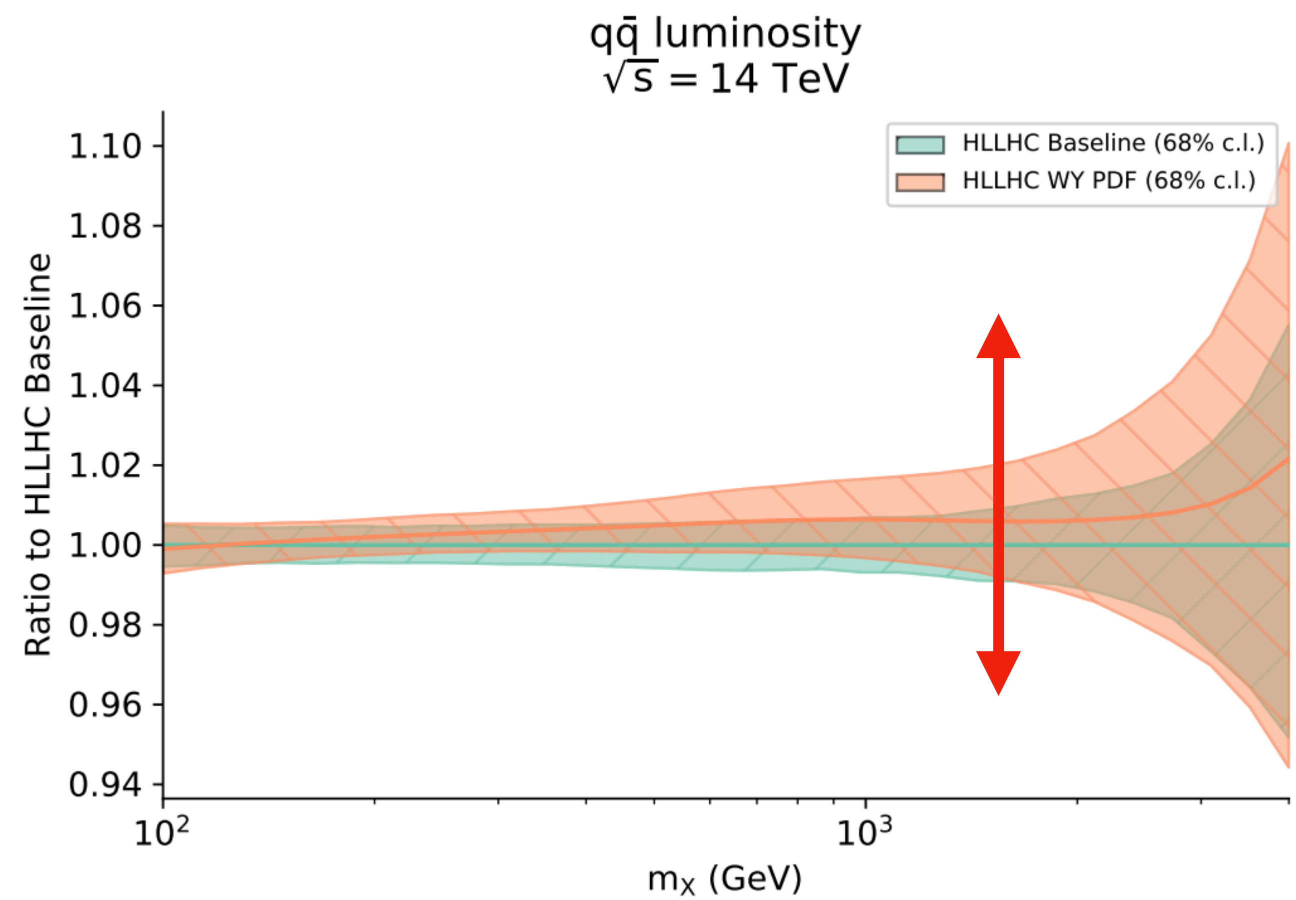
$$T_i(\{\theta\},\{c\}) = \mathrm{PDFs}(\{\theta\},\{c\}) \otimes \hat{\sigma}_i(\{c\})$$

Simultaneous fits can shed light on their interplay

3. Test-ground: DY data at HL-LHC

x 2.3 broadening of bounds for W
x 1.3 broadening of bounds for Y



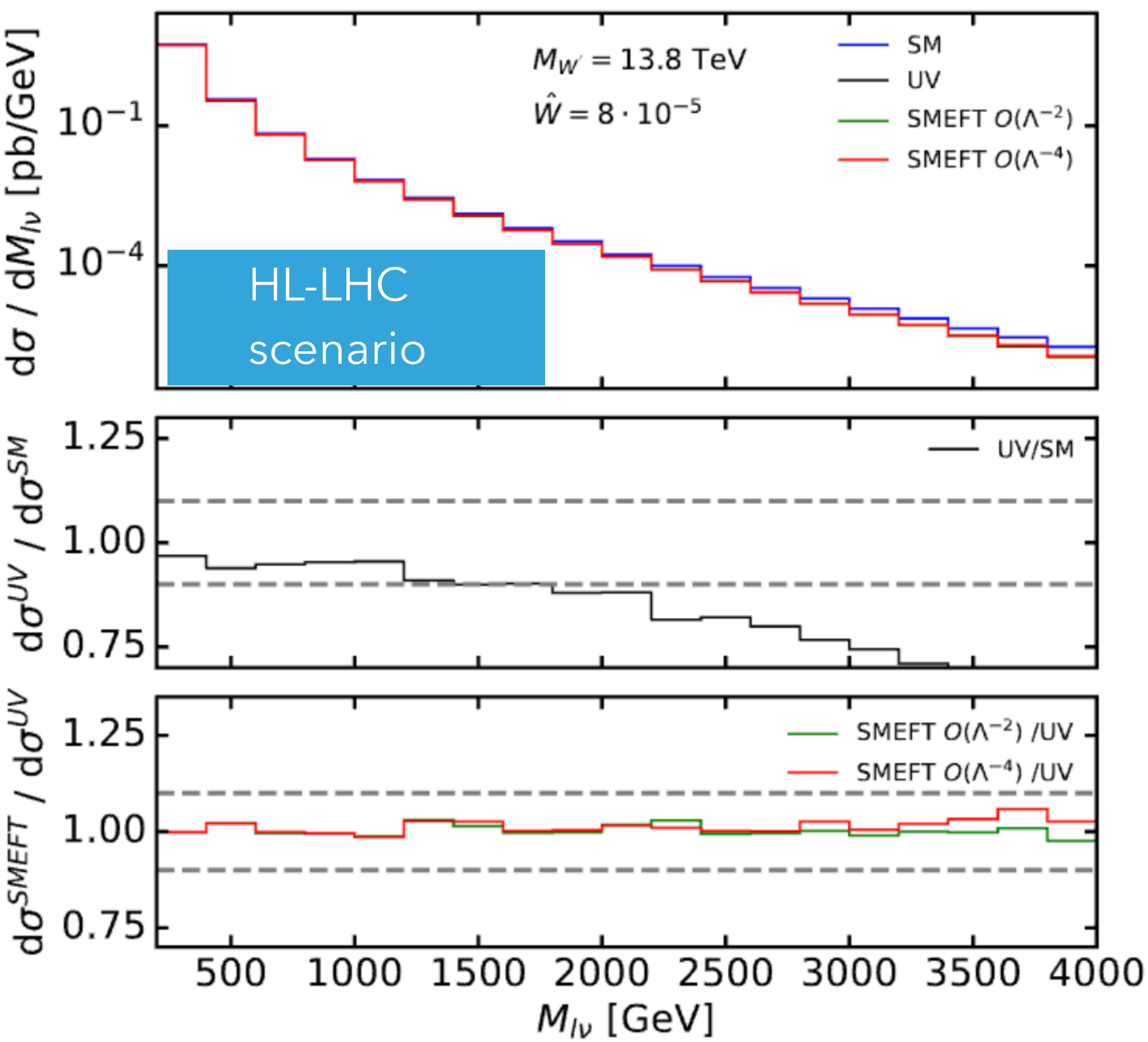


✓ Simultaneous fit shows that at HL-LHC the effect of interplay between SMEFT fits and PDF fits becomes important as bounds on Wilson Coefficients that affect high-mass invariant tails broaden

✓ Also PDF uncertainties broaden significantly once SMEFT effects allowed in theory predictions entering PDF fit

S. Iranipour, MU - arXiv: 2201.07240

3. Can PDFs absorb New Physics?



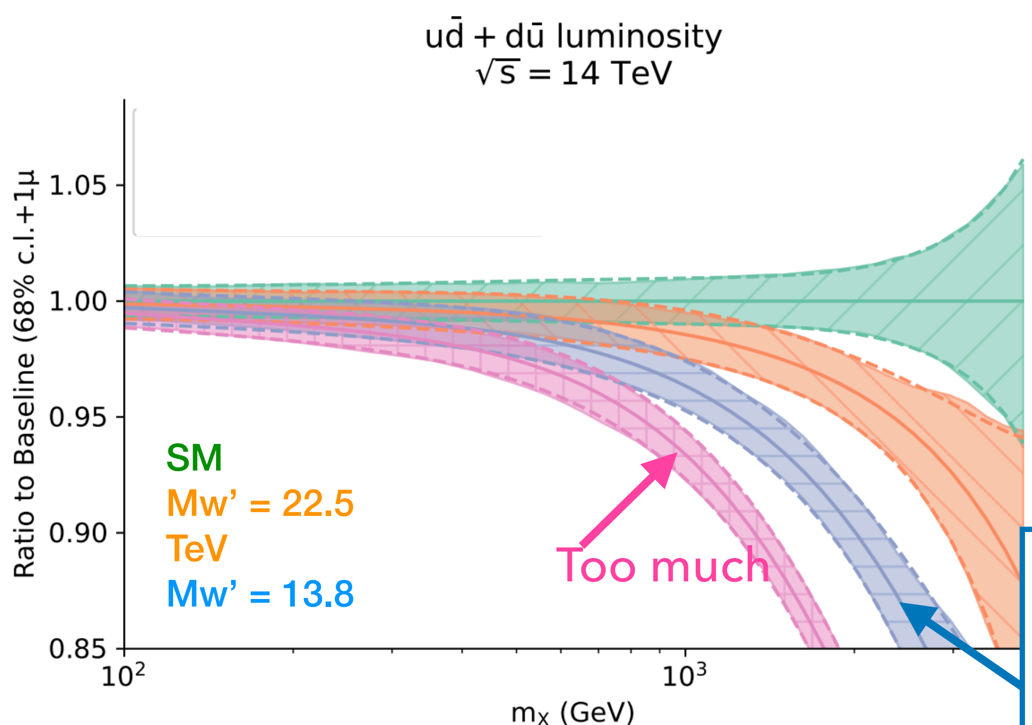
- Imagine instead to generate artificial data injecting a given underlying law:
  "True" law of nature =
  "True" PDFs + "True" BSM model
- Can inject directly UV model or W/Y SMEFT parametrisation in the region in which SMEFT approximation is good
- Fit PDFs **assuming SM**
- If fit quality of BSM (data) with SM (theory) does not deteriorate with respect to SM (data) with SM (theory) and PDFs deviate from SM PDFs then the BSM signal is absorbed by the PDFs!

✔ $\mathscr{L}^{W'}_{SMEFT} = \mathscr{L}_{SM} - \frac{g^2_{W'}}{2M^2_{W'}} J^{a,\mu}_L J^a_{L,\mu}$ $\qquad J^{a,\mu}_L = \sum_{f_L} \bar{f}_L T^a \gamma^\mu f_L$

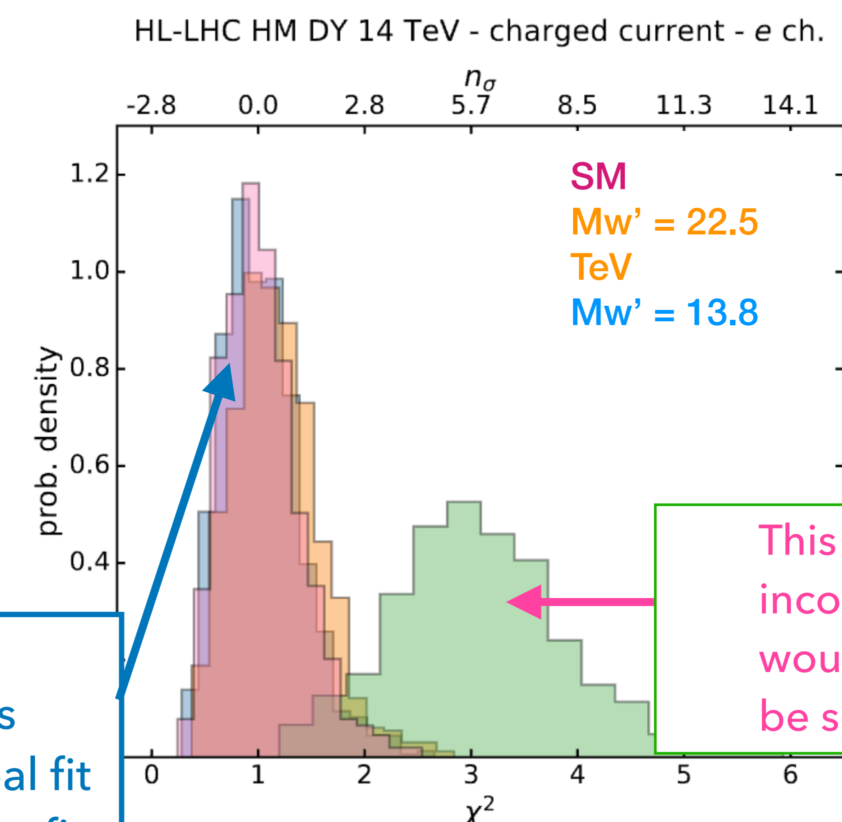
✗ $\mathscr{L}^{Z'}_{SMEFT} = \mathscr{L}_{SM} - \frac{g^2_{Z'}}{2M^2_{Z'}} J^{\mu}_Y J_{Y,\mu}$ $\qquad J^{\mu}_Y = \sum_{f} Y_f \bar{f} \gamma^\mu f$

3. Can PDFs absorb New Physics? Yes

- The fit-quality of the global fit is unchanged even with signal from Mw' = 13.8 TeV injected in all data (mostly visible in HL-LHC NC and CC Drell-Yan data)
- Once we go beyond this point ,the fit-quality deteriorates due to the HL-LHC neutral current and charged current Drell-Yan artificial data.
- Already for Mw' = 13.8 TeV the qq~ luminosity shifts far beyond the PDF uncertainties because anti-quark PDFs at large-x compensate or "fit away" the effect of New Physics and we would not know in a real fit.



Max BSM bias allowed for this model by global fit without spoiling fit quality



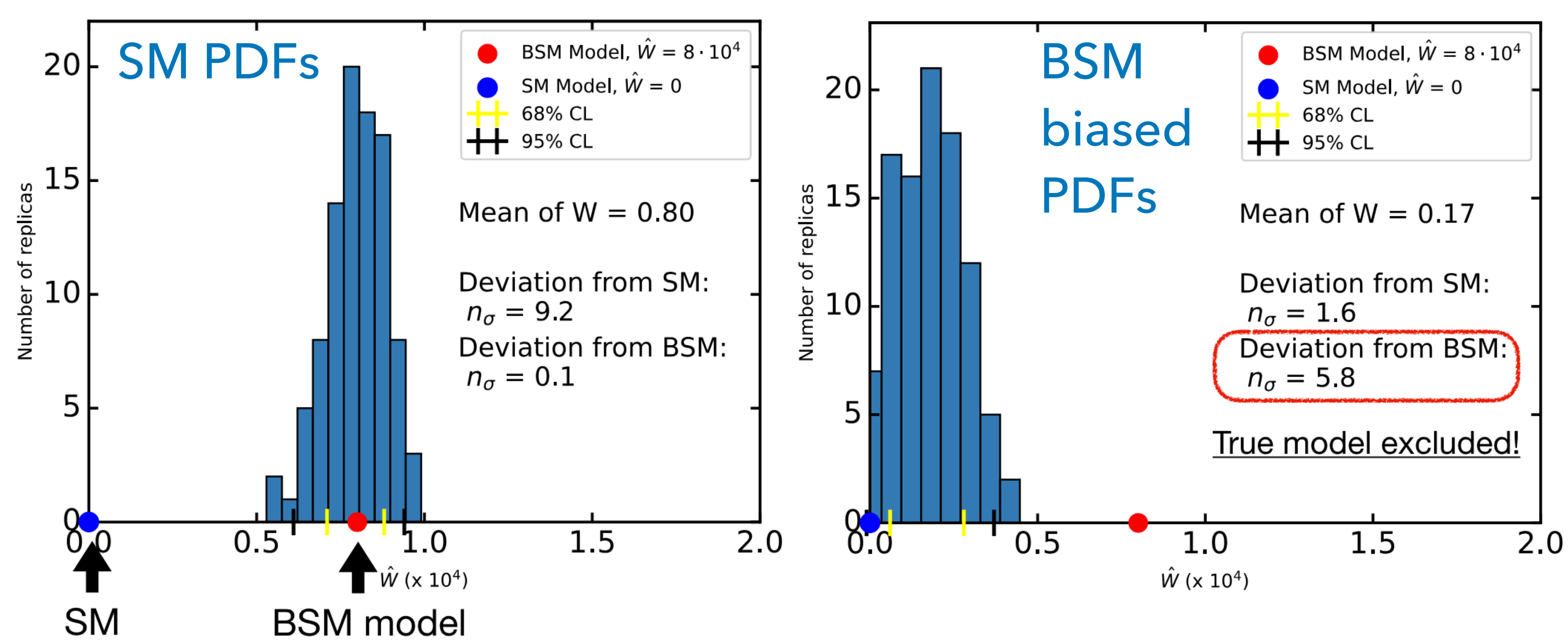
This NP-induced inconsistency would be spotted

E. Hammou, et al 2307.10370

3. Can PDFs absorb New Physics? Yes

- The fit-quality of the global fit is unchanged even with signal from Mw′ = 13.8 TeV injected in all data (mostly visible in HL-LHC NC and CC Drell-Yan data)
- Once we go beyond this point ,the fit-quality deteriorates due to the HL-LHC neutral current and charged current Drell-Yan artificial data.
- Already for Mw′ = 13.8 TeV the qq~ luminosity shifts far beyond the PDF uncertainties because anti-quark PDFs at large-x compensate or “fit away” the effect of New Physics and we would not know in a real fit.

Consequence #1:
Would not see indirect effect of new physics as we would find SMEFT bounds compatible with the SM!

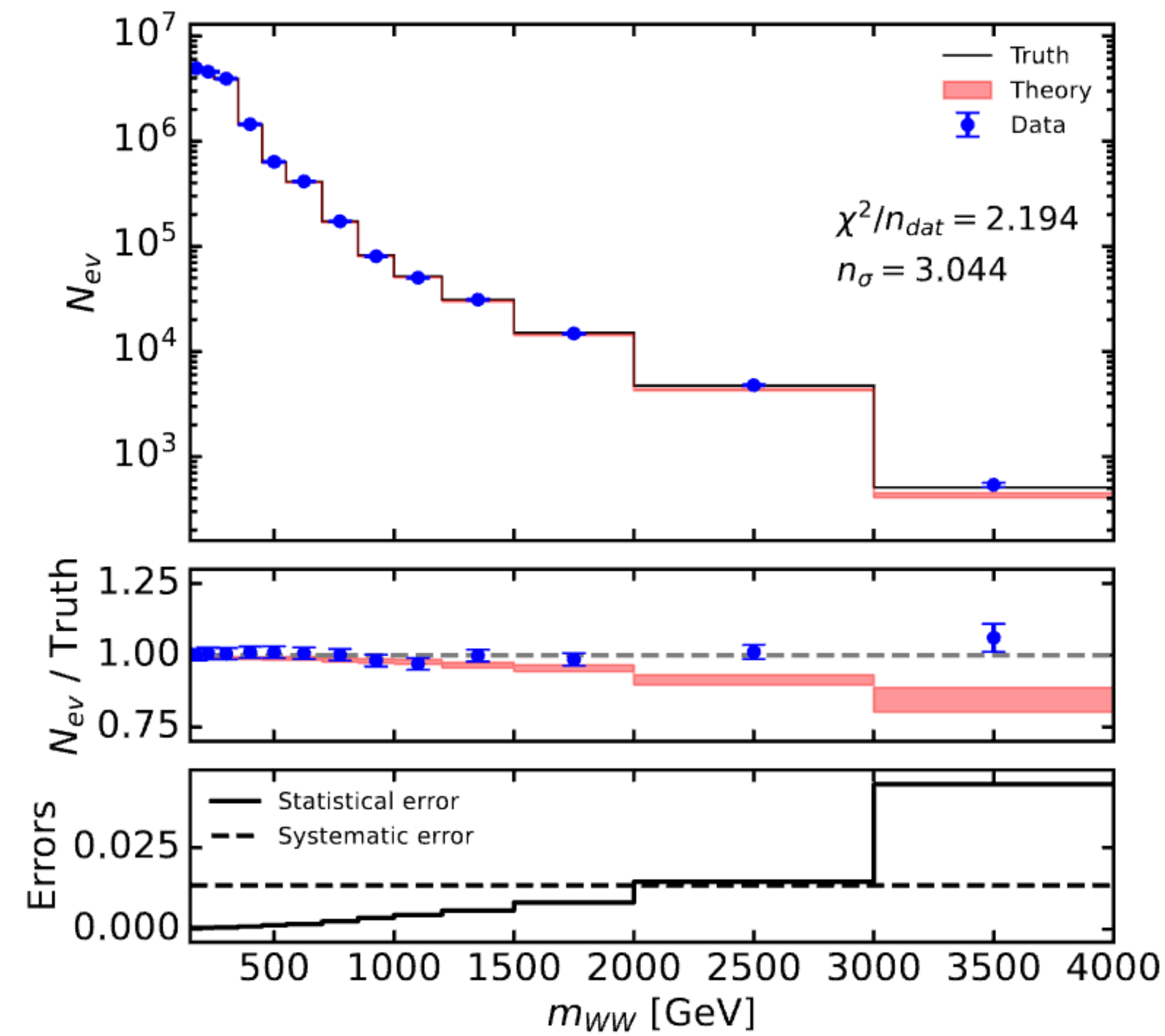


E. Hammou @HEFT2024

3. Can PDFs absorb New Physics? Yes

- The fit-quality of the global fit is unchanged even with signal from Mw' = 13.8 TeV injected in all data (mostly visible in HL-LHC NC and CC Drell-Yan data)
- Once we go beyond this point ,the fit-quality deteriorates due to the HL-LHC neutral current and charged current Drell-Yan artificial data.
- Already for Mw' = 13.8 TeV the qq~ luminosity shifts far beyond the PDF uncertainties because anti-quark PDFs at large-x compensate or "fit away" the effect of New Physics and we would not know in a real fit.

Consequence #2:
Would see New Physics effects where there are none (for example in WW)



E. Hammou, et al 2307.10370

4. How to avoid BSM bias in PDFs?

- LHCb on-shell at high rapidity data do not help as quark probed at large x, antiquark at small x
- If we fit giving more weight (artificially increasing statistics) of fixed-target DY data on proton/deuteron the BSM-induced inconsistency would be flagged
- Need more accurate low-energy/large-x constraining measurements to really disentangle such effects

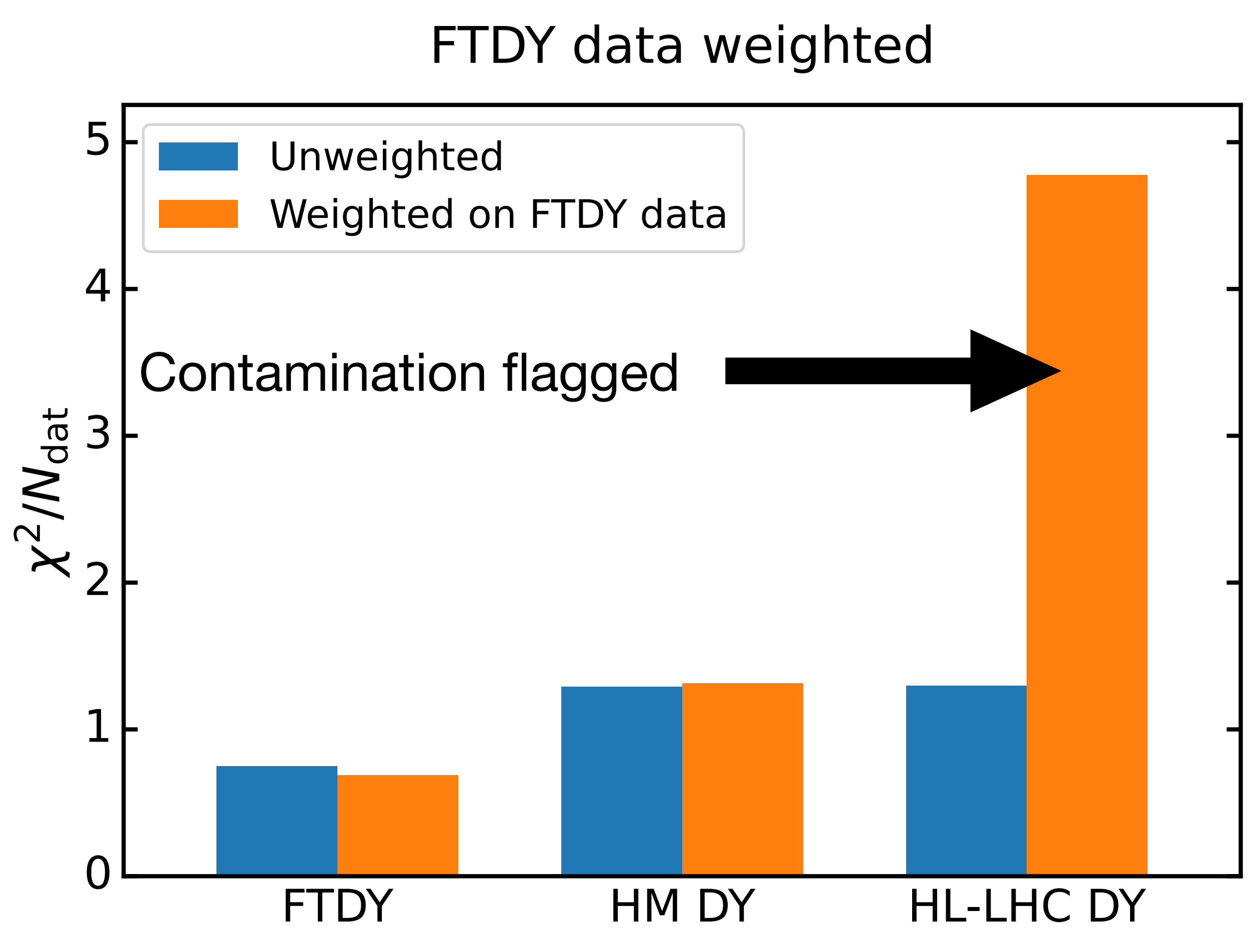
Flat direction in the anti-quark at large-x:

➡ Accommodates PDFs to artificial data with BSM injected

➡ Allows for BSM bias in qqbar luminosity

Including lower-energy large-x data:

- Constrain large-x region
- Safe from BSM contamination



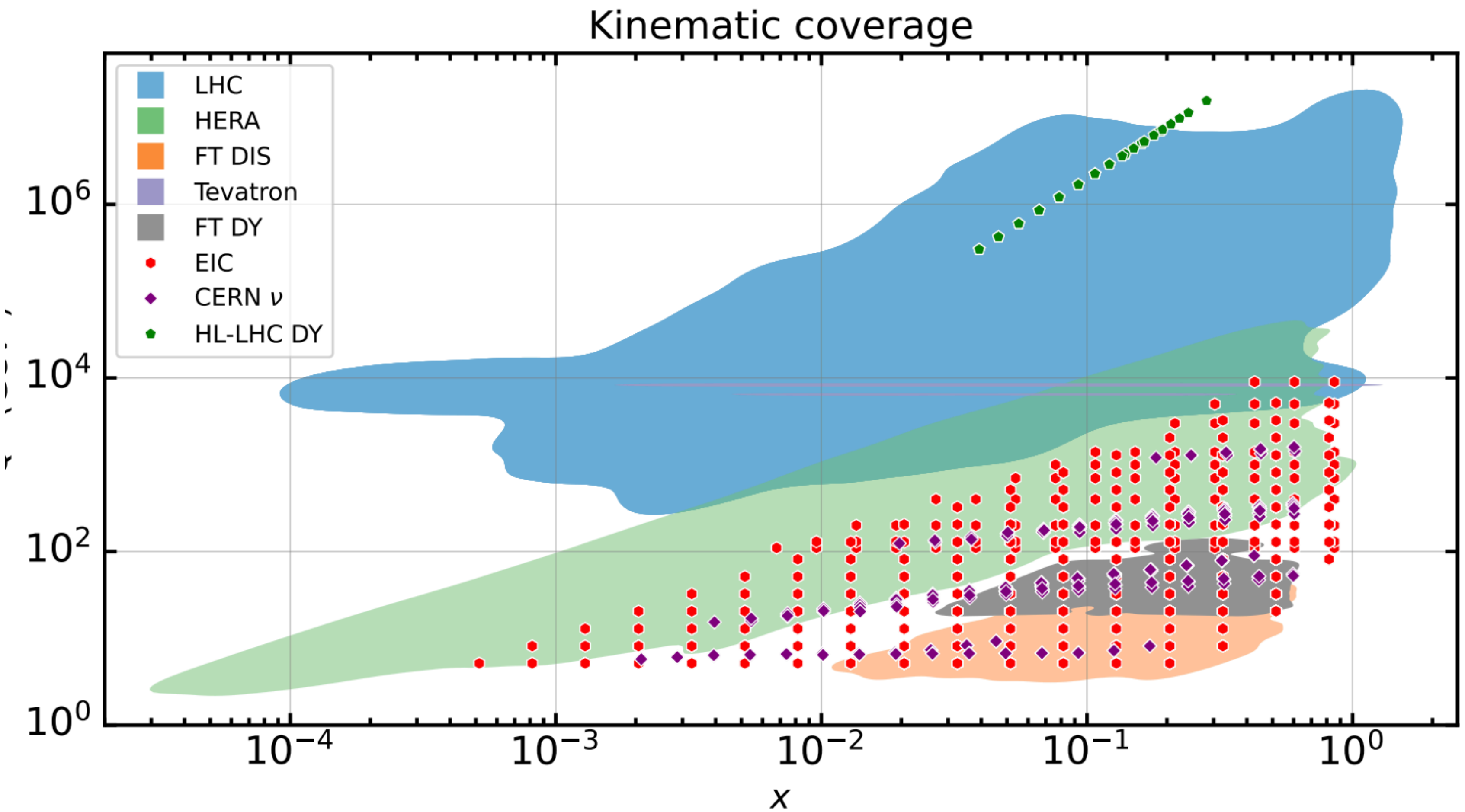
[Hammou, MU, 2410.00963]

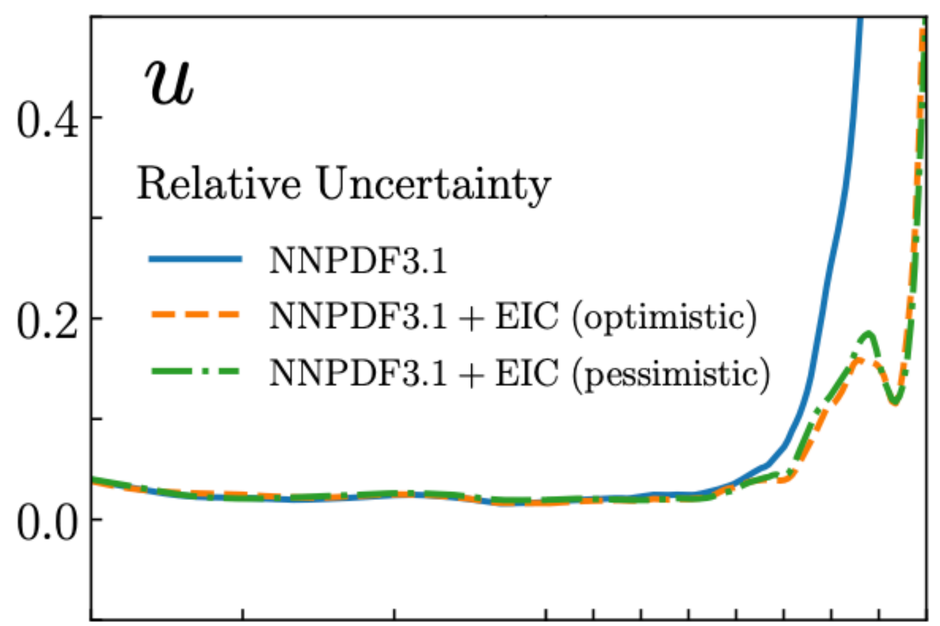
4. How to avoid BSM bias in PDFs?

Observable	N_{dat}	$\sqrt{s}$ [GeV]	$\mathcal{L}$ [fb$^-$1]
Charged Current			
$\tilde{\sigma}(e^- + p \to \nu + X)$	89 (89)	140.7	100
$\tilde{\sigma}(e^+ + p \to \bar{\nu} + X)$	89 (89)	140.7	10
Neutral Current (proton)			
$\tilde{\sigma}(e^- + p \to e^- + X)$	181 (131)	140.7	100
	181 (131)	63.2	100
	126 (91)	44.7	100
	87 (76)	28.6	100
$\tilde{\sigma}(e^+ + p \to e^+ + X)$	181 (131)	140.7	10
	181 (131)	63.2	10
	126 (91)	44.7	10
	87 (76)	28.6	10
Neutral Current (deuteron)			
$\tilde{\sigma}(e^- + d \to e^- + X)$	116 (116)	89.0	10
	107 (107)	66.3	10
	76 (76)	28.6	10
$\tilde{\sigma}(e^+ + d \to e^+ + X)$	116 (116)	89.0	10
	107 (107)	66.3	10
	76 (76)	28.6	10

Optimistic scenario

Pessimistic scenario



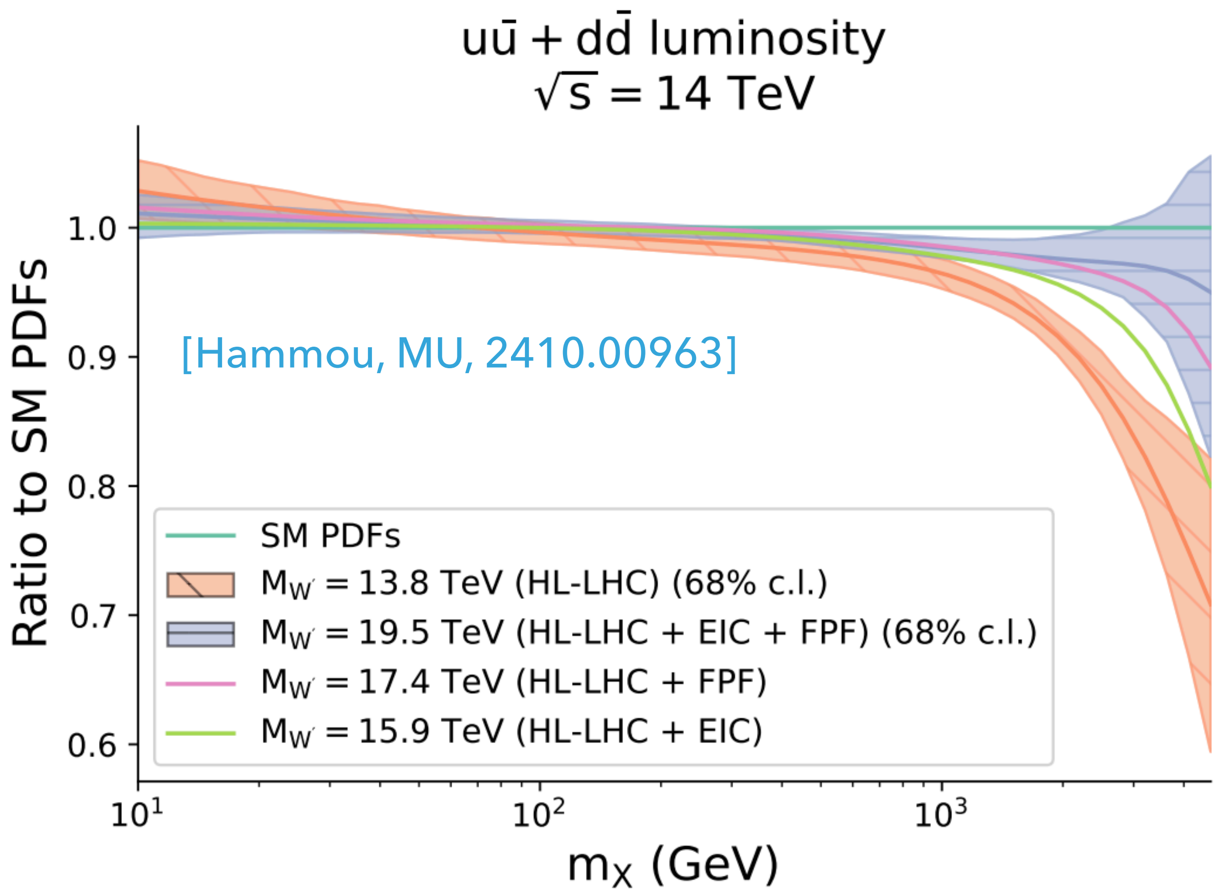


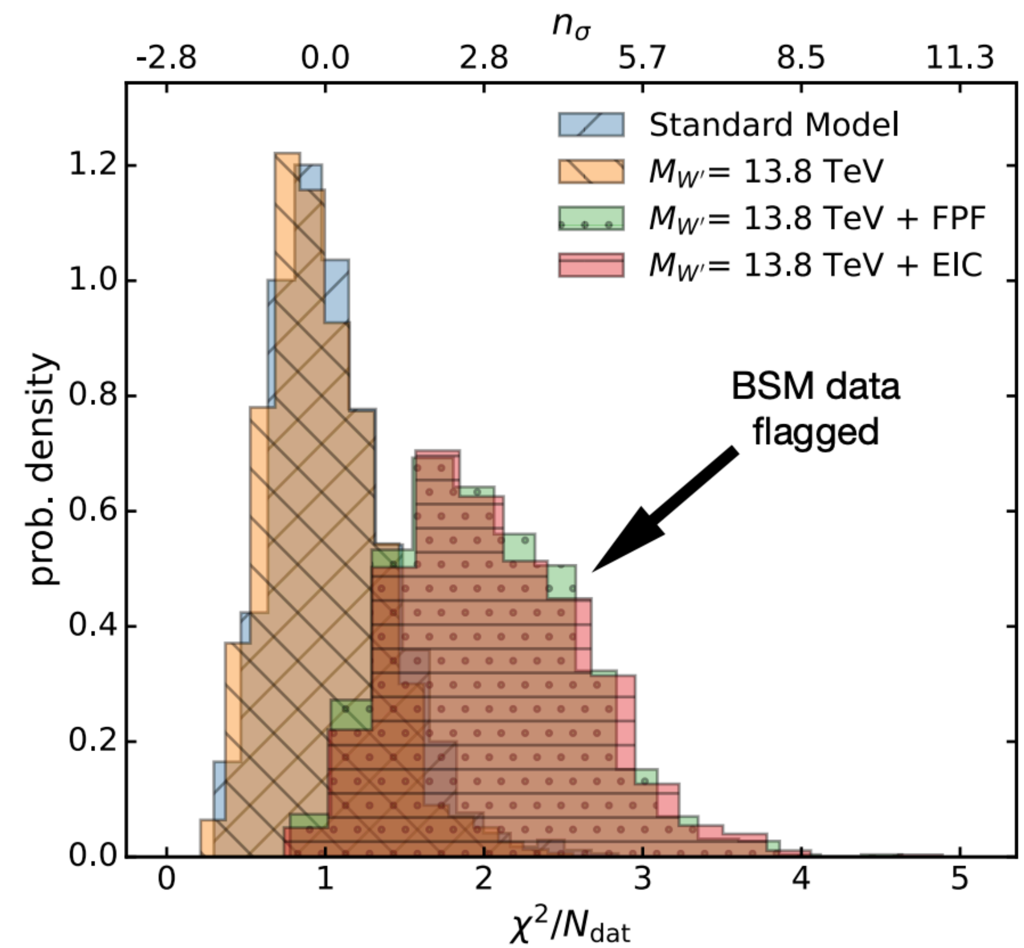
- NC and CC DIS measurements at the EIC would provide handles on up and down quarks at medium-large-x

EIC to the rescue

R. A. Khalek et al, 2102.00018

4. How to avoid BSM bias in PDFs?





- Including the EIC artificial data alongside the HL-LHC ones reduces the BSM-bias in PDF luminosity
- The HL-LHC data with Mw' = 13.8 TeV injected would be excluded from the fit as their theoretical description with the SM would be too poor and the data would be considered inconsistent
- EIC inclusion would push the threshold of BSM bias up to Mw'~20 TeV and would make the effect on PDFs much smaller

4. How to avoid BSM bias in PDFs?

✓ Potential BSM effects, which might otherwise be absorbed into the PDFs, can be disentangled in high-energy measurements by incorporating EIC measurements in a global PDF analysis

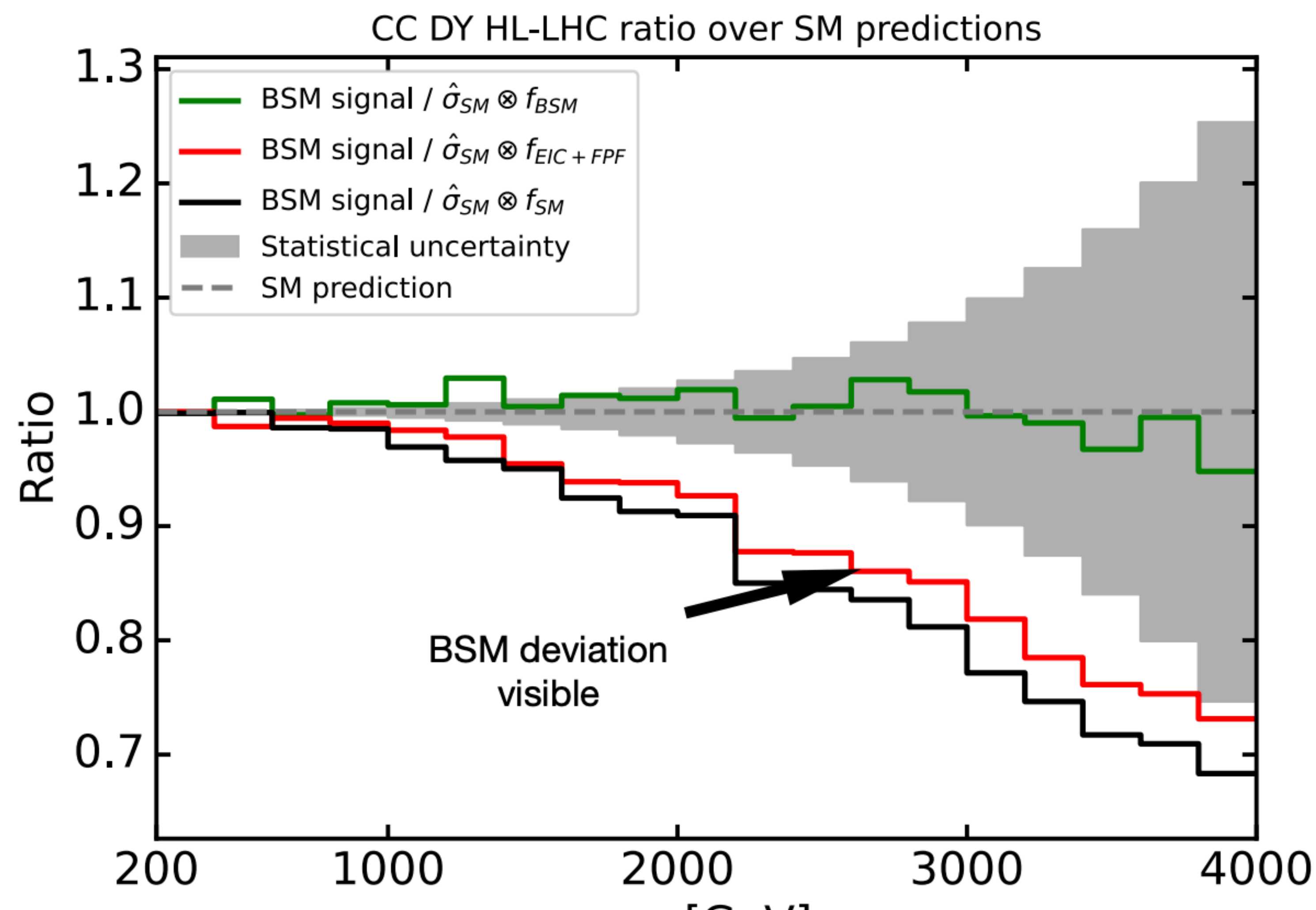
✓ This minimises the risk of BSM-induced bias in PDF fits and allow for more consistent identification of BSM effects in high energy data, which can then be analysed separately.

$$\hat{\sigma}_{BSM} \otimes \mathscr{L}_{SM} \approx \hat{\sigma}_{SM} \otimes \mathscr{L}_{BSM}$$

↓

$$\hat{\sigma}_{BSM} \otimes \mathscr{L}_{SM} \neq \hat{\sigma}_{SM} \otimes \mathscr{L}_{EIC+FPF}$$

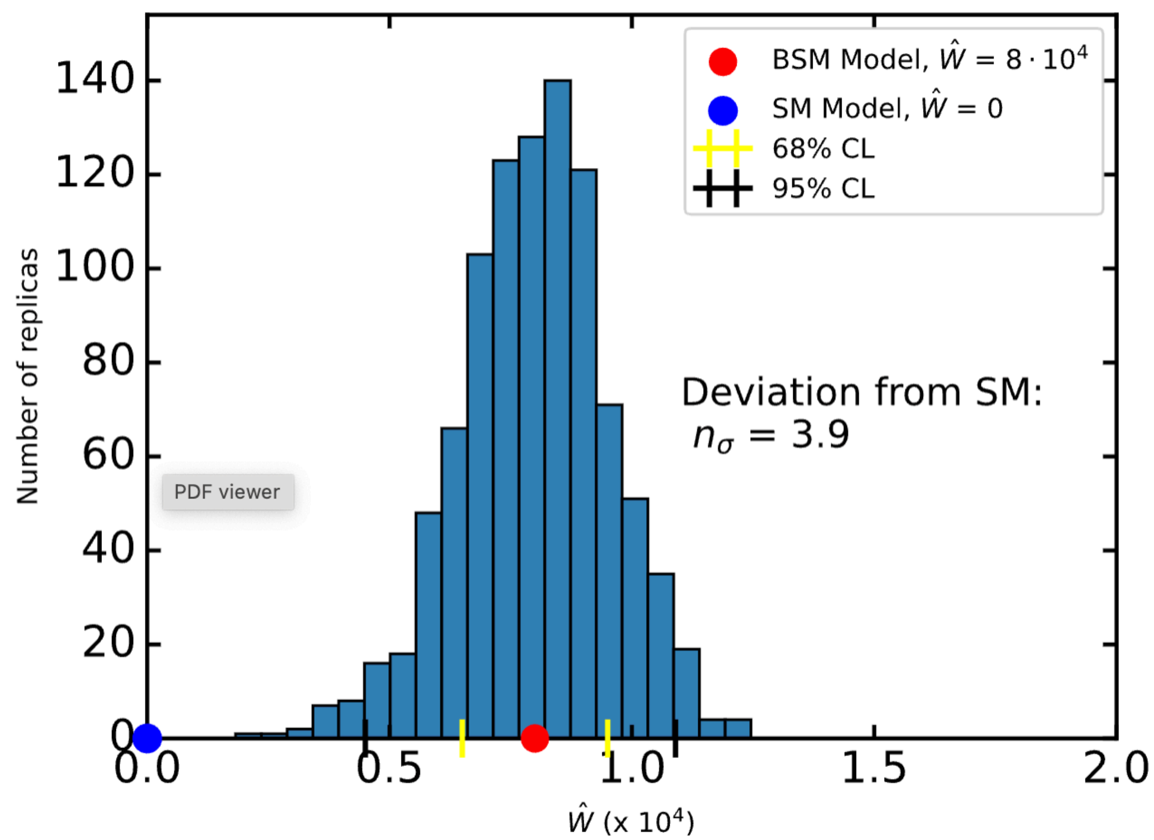
$M_{W'}$: 13.8 TeV

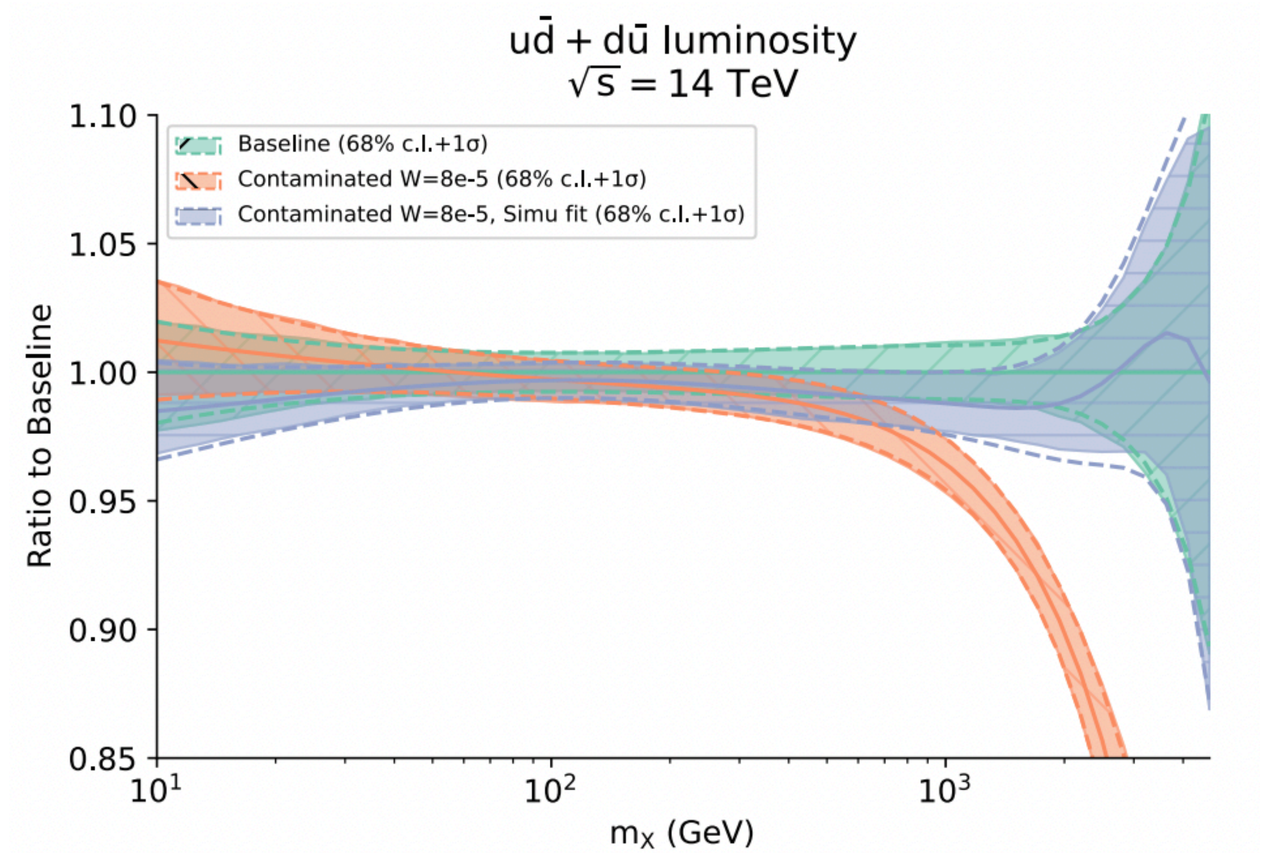


[Hammou, MU, 2410.00963]

4. How to avoid BSM bias in PDFs?

Simultaneous fits to the rescue



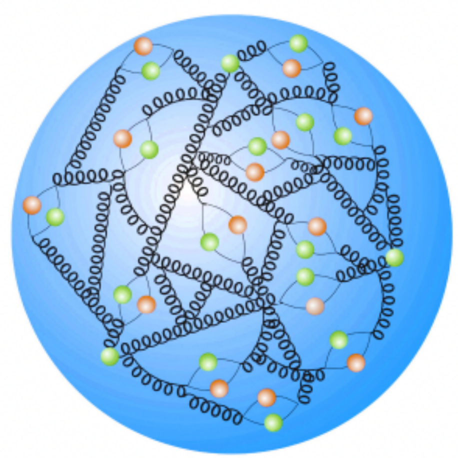


E. Hammou @HEFT2024

✓ Simultaneous analysis of PDFs and **Drell-Yan sector** Wilson coefficient in the context of universal parameters of DIS + DY (including HL-LHC projections) using simuNET method shows that if HL-LHC projections were generated by using a W' BSM model, the simultaneous fit would be able to find the "true" SM PDFs and the "true" SMEFT value!

[Costantini, Hammou et al - in preparation]

Conclusions



✓ The fits of the subnuclear structure of the proton involve wealth of ingredients from low to high energy: non-perturbative effects, perturbative QCD, experimental measurements, statistical and mathematical problems, higher order predictions, phenomenology tools, machine learning

✓ We reached a precision that was unthinkable and the very same precision opens deeper problems

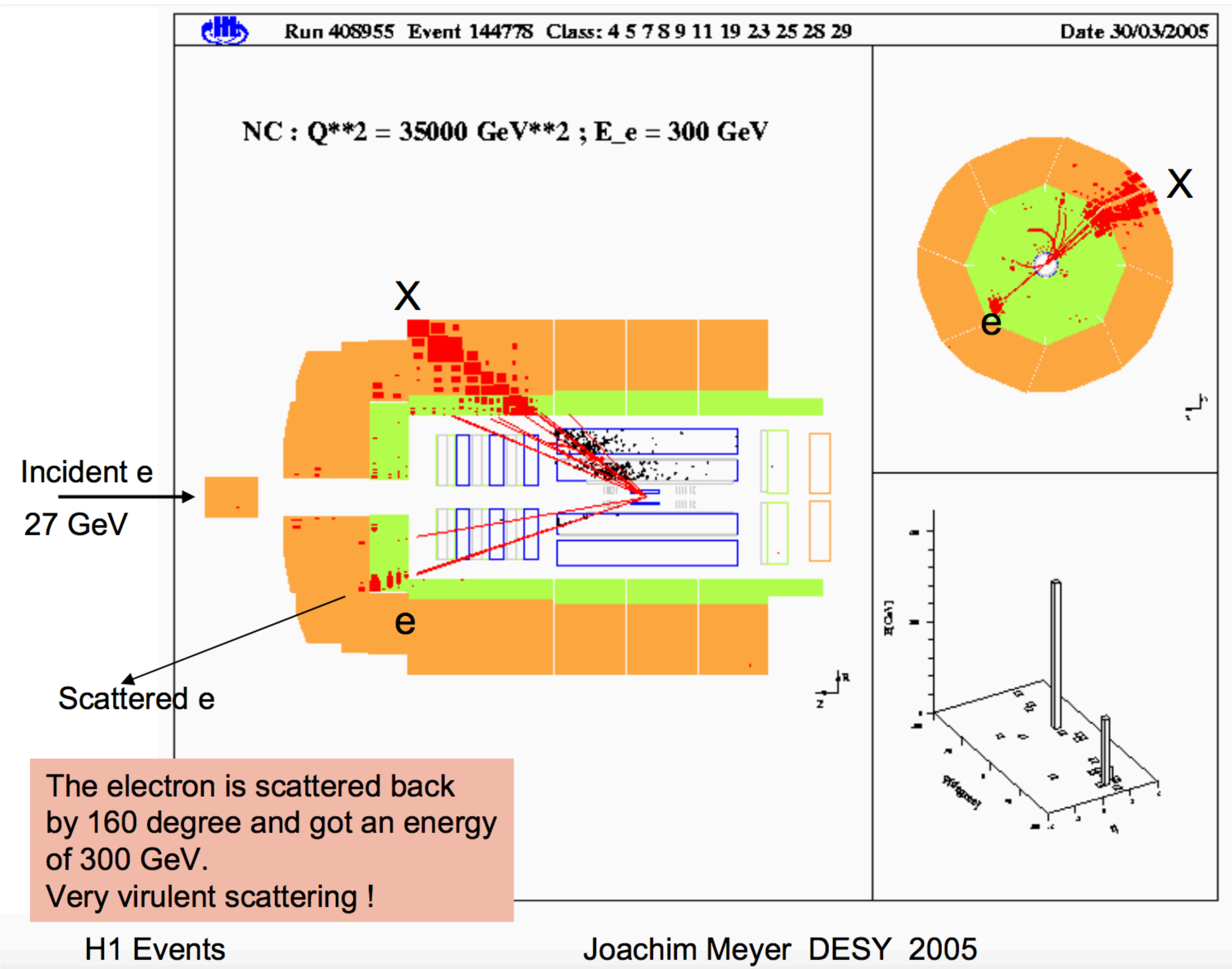
✓ After nearly 30 years spent looking into the proton …

"The same thrill, the same awe and mystery, come again and again when we look at any problem deeply enough. With more knowledge comes deeper, more wonderful mystery, luring one on to penetrate deeper still.
Never concerned that the answer may prove disappointing, but with pleasure and confidence we turn over each new stone to find unimagined strangeness leading on to more wonderful questions and mysteries -- certainly a grand adventure! "

R. P. *Feynman*

Extra material

HERA data



Neutral Current event

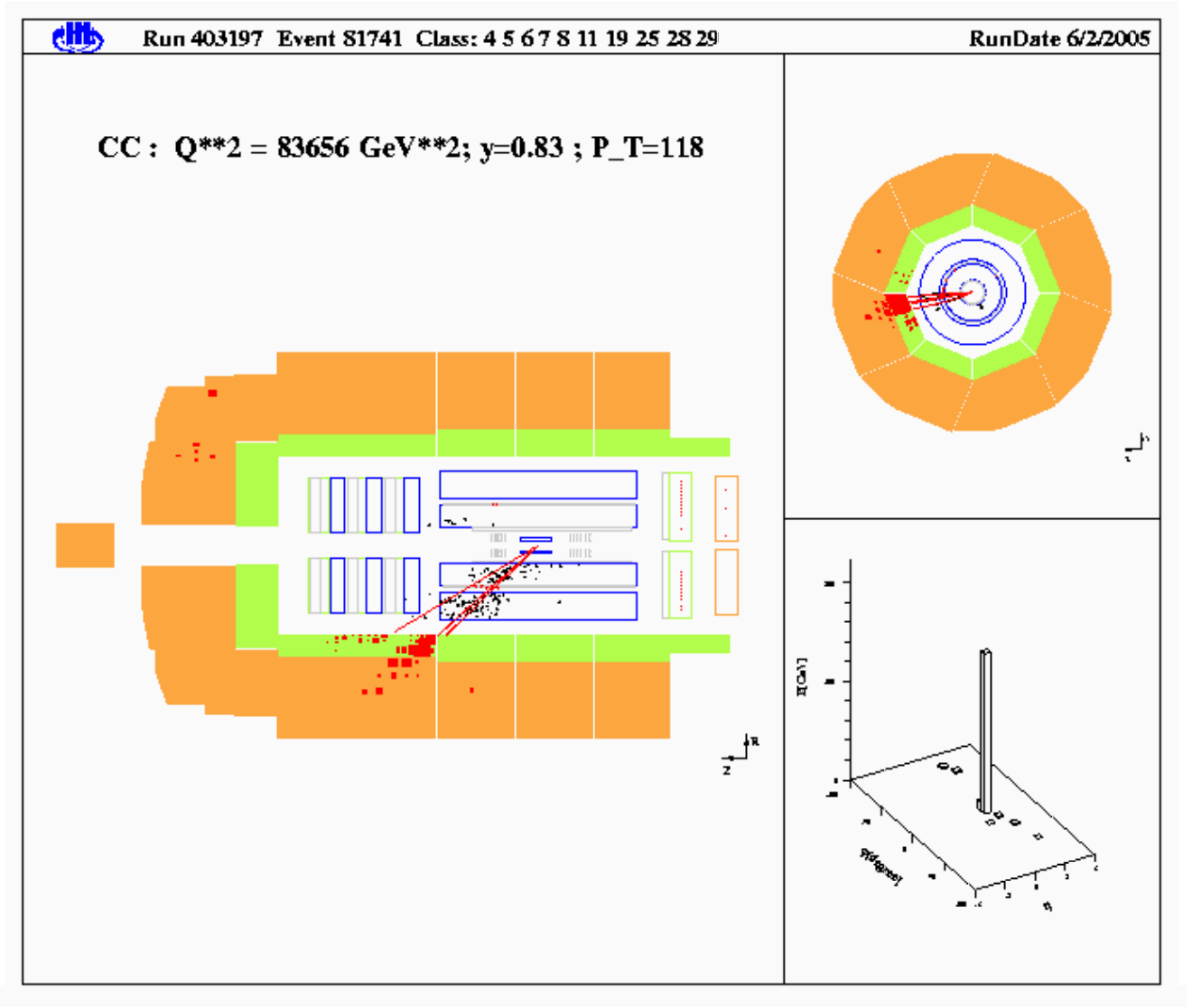
$$ep \to e\,X$$

$$Q = 180\,\mathrm{GeV}$$

$$y = 0.66$$

$$x_B = 0.47$$

HERA data



Charged Current event

$$ep \to \nu X$$

$$Q = 289\,\mathrm{GeV}$$

$$y = 0.83$$

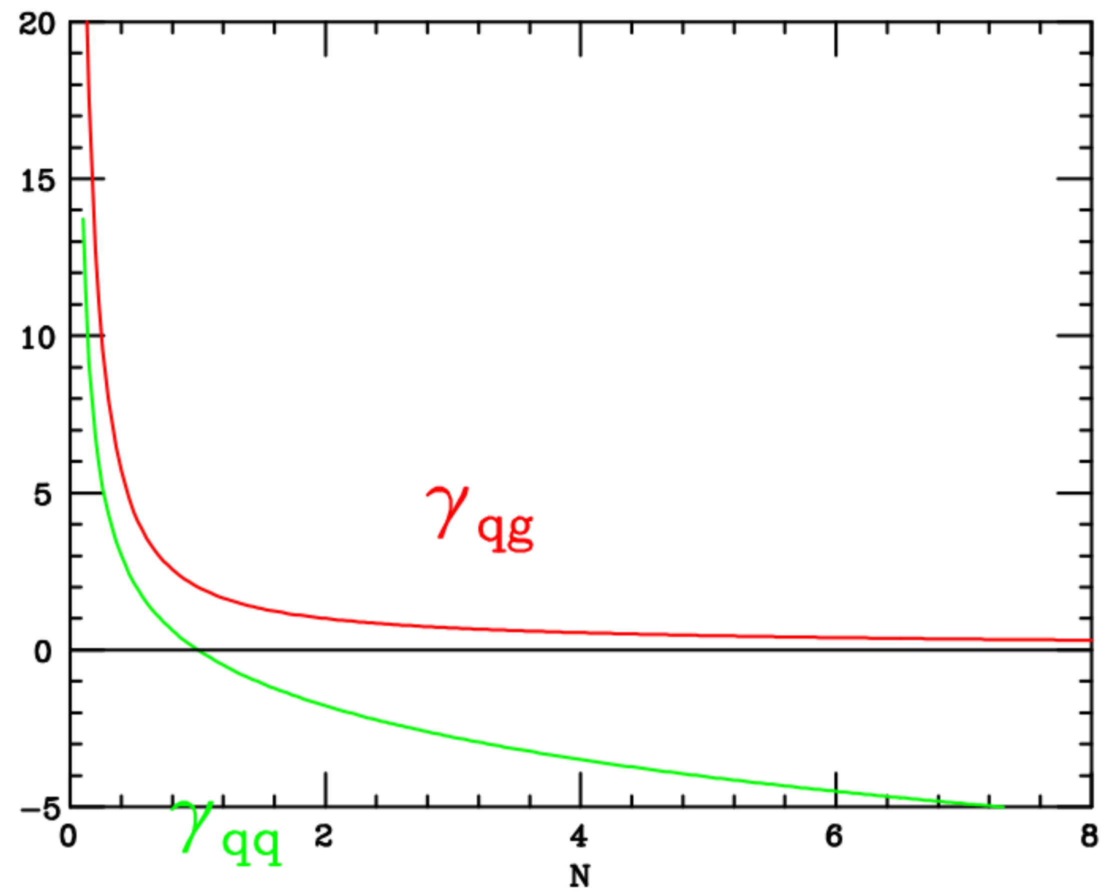
$$x_B = 0.91$$

Joachim Meyer DESY 2005

Gluon: indirect handle

- Gluon is partially determined by scale dependence of DIS structure functions and Drell-Yan/Vector Boson production

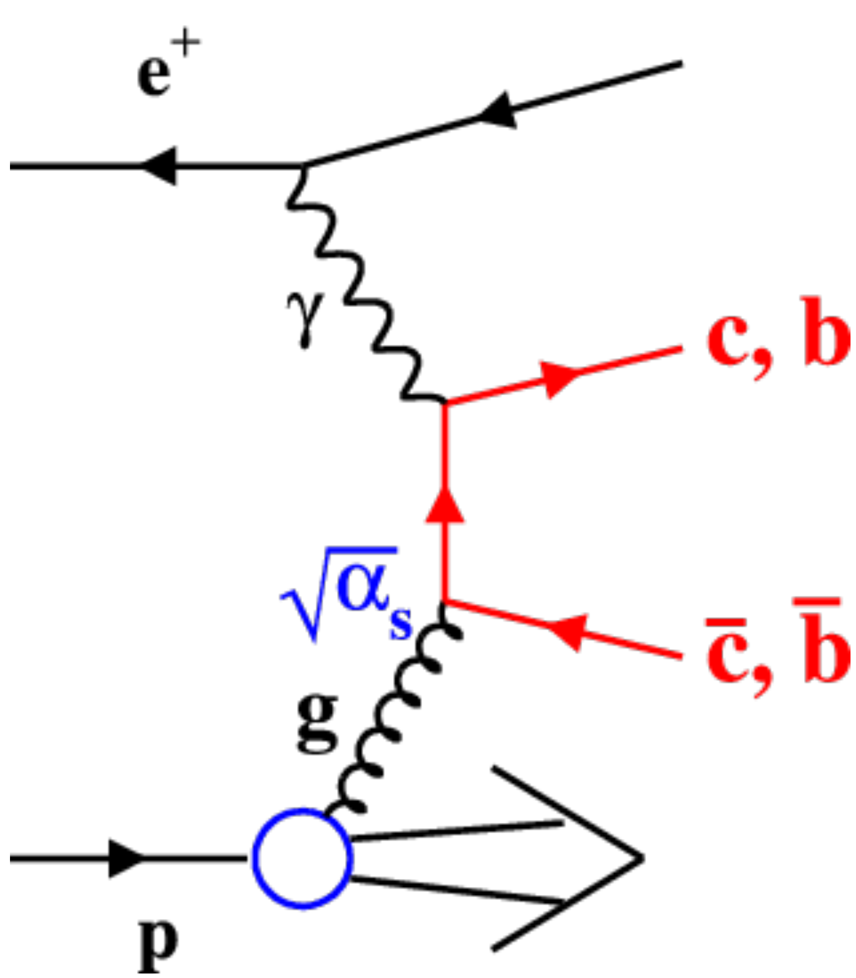
$$\frac{d}{d\log\mu^2}F_2(x,\mu^2) = \frac{\alpha_s(\mu^2)}{2\pi}\left[P_{qq}\otimes F_2(x,\mu^2) + 2n_f P_{qg}\otimes g(x,\mu^2)\right]$$



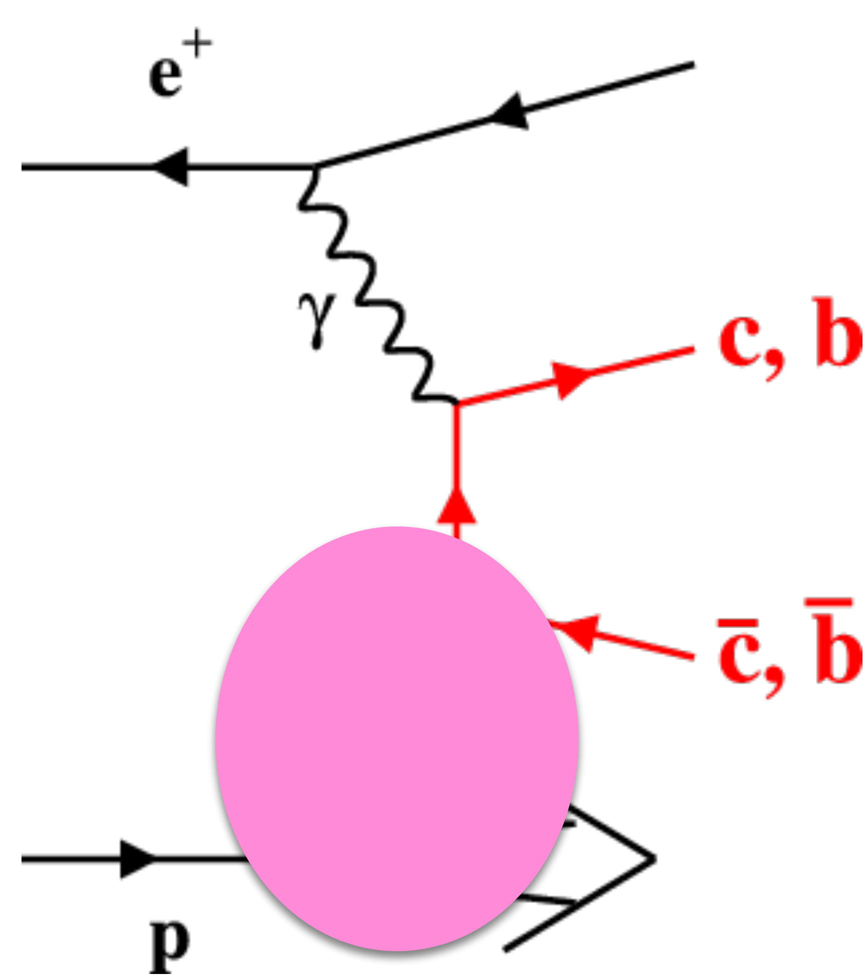
- Mostly determine small-x gluon, large-x gluon hard to determine from DIS+DY only data

Gluon: indirect handle

- Heavy quarks are produced at threshold inside proton
- Heavy quark production process (at ep and pp colliders) probe gluon
- Dependence on heavy flavour scheme adopted in PDF fitting



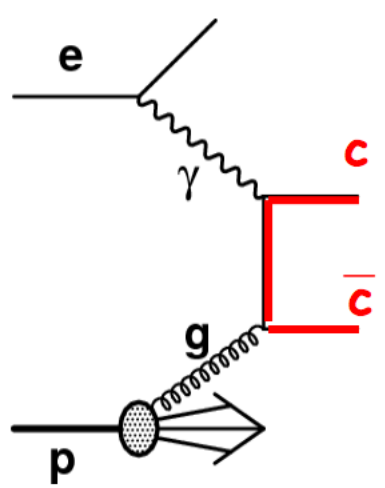
Nf = 3,4

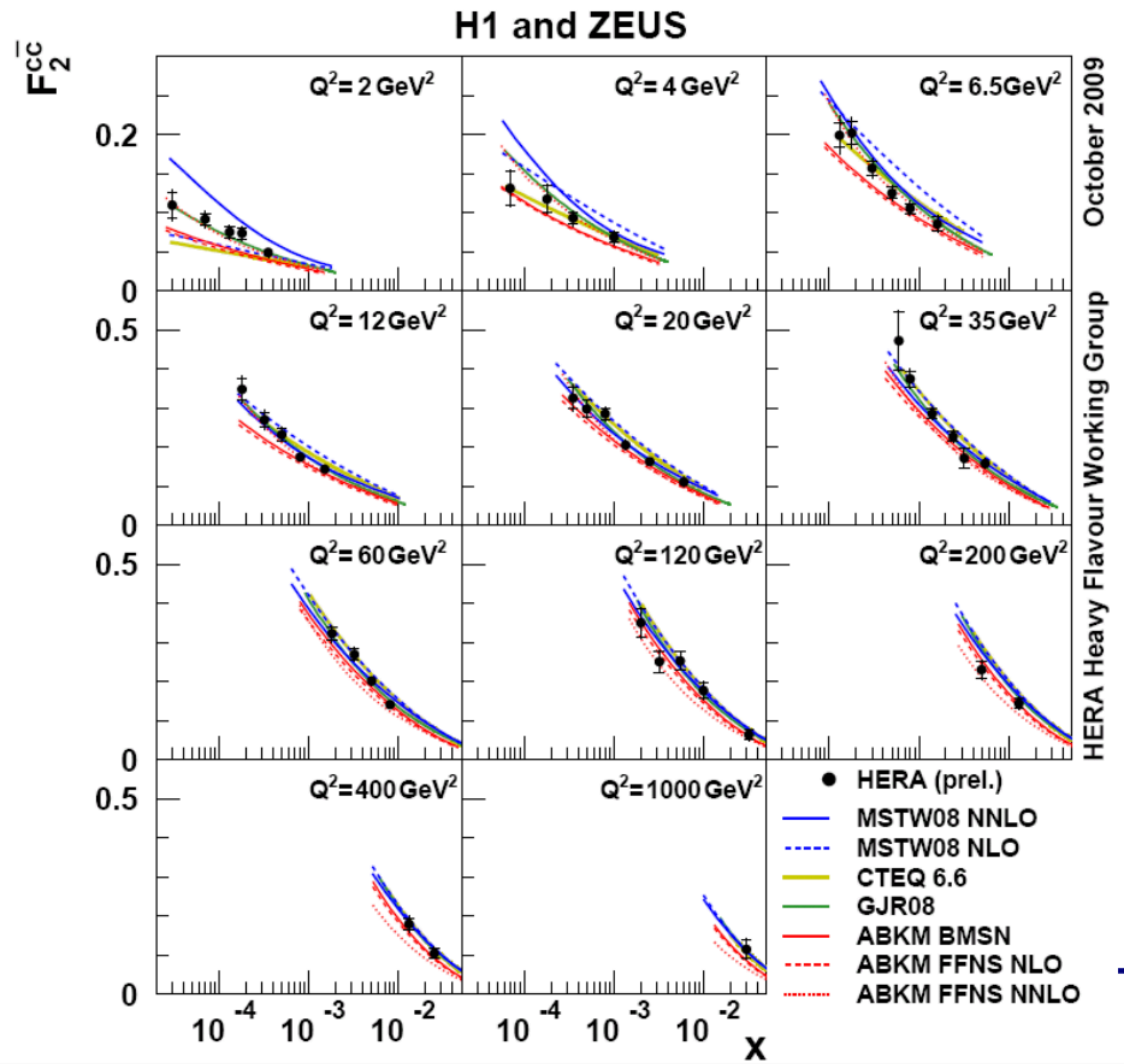


Nf = 5

Gluon: indirect handle

- Heavy quarks are produced at threshold inside proton
- Heavy quark production process (at ep and pp colliders) probe gluon
- Dependence on heavy flavour scheme adopted in PDF fitting





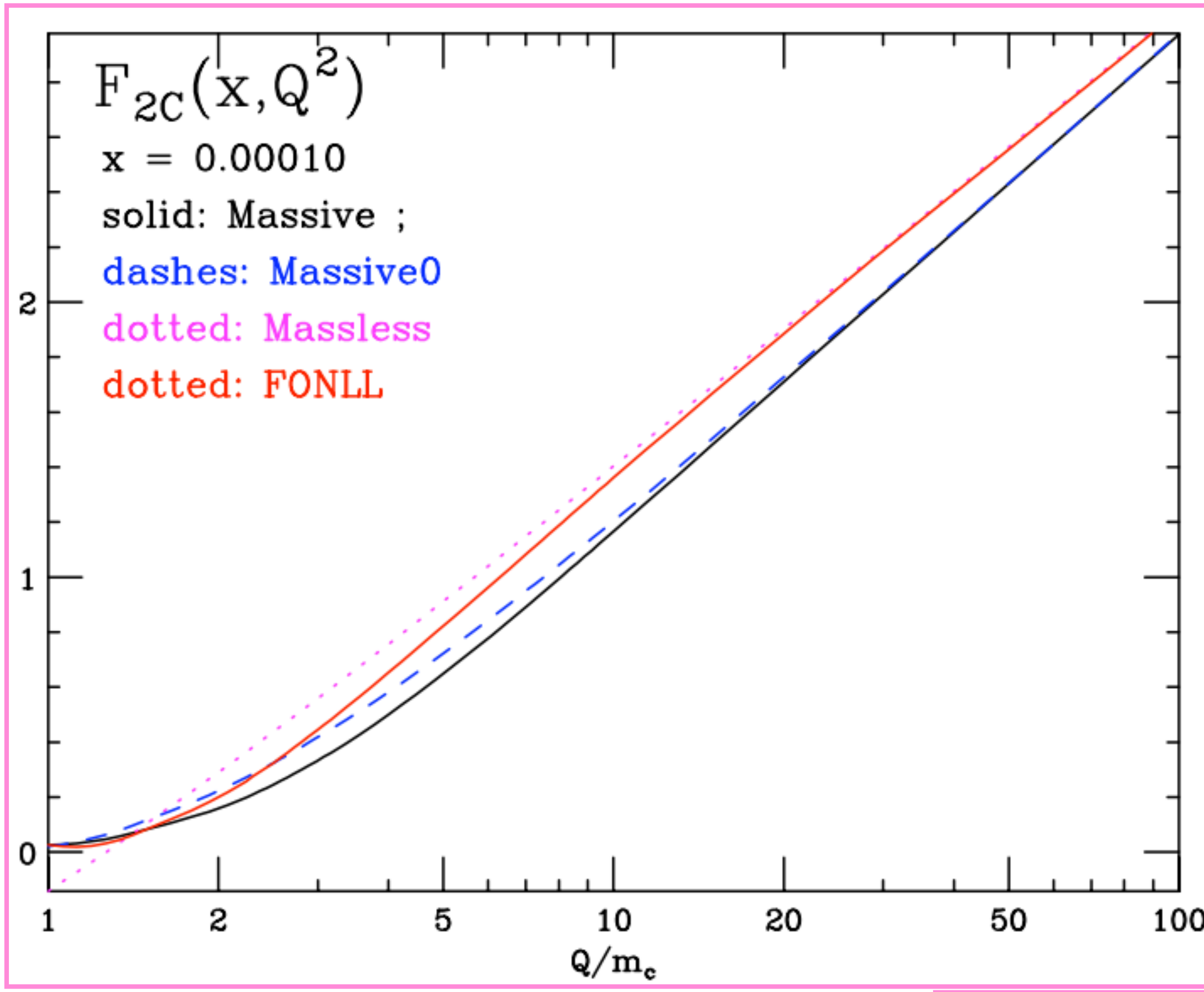
Intermission: heavy flavour schemes

- Charm, Bottom and Top have mass >> Λ_{QCD} - heavy quarks (HQ)
- The presence of a new scale, m_Q, makes pert QCD calculations more challenging
- Two well understood schemes:
    - Assume heavy quark effectively massless for $Q > m_Q$
      HQ becomes active massless parton above threshold
    - Heavy quarks retain their mass for all Q
      HQ is not a parton, it is a final state particle
- However in PDF fits we have all scales. General-Mass Variable-Flavor-Number schemes allow to match between the zero-mass and the massive scheme
- Many schemes available

e.g. FONLL

$$\begin{aligned}\sigma^{(FONLL)} &= \sigma^{(4)} + \sigma^{(5)} - \text{double counting}\\ &= \mathcal{L}_{ij}(x_1, x_2, \mu^2) \otimes \sum_{p}^{N} \left(\alpha_S^{(5)}(\mu^2)\right)^p\\ &\times \left\{ \mathcal{B}_{ij}^{(p)}\left(x_1, x_2, \frac{\mu^2}{m_b^2}\right) + \sum_{k=0}^{\infty} \mathcal{A}_{ij}^{(p),(k)}(x_1, x_2)\left(\alpha_S^{(5)}(\mu^2) L\right)^k \right\}\\ &- \text{double counting}\end{aligned}$$

Intermission: heavy flavour schemes



- heavy quarks (HQ)

D calculations more challenging

:less for Q > mQ
above threshold

Q2
article

l-Mass Variable-Flavor-Number
and the massive scheme

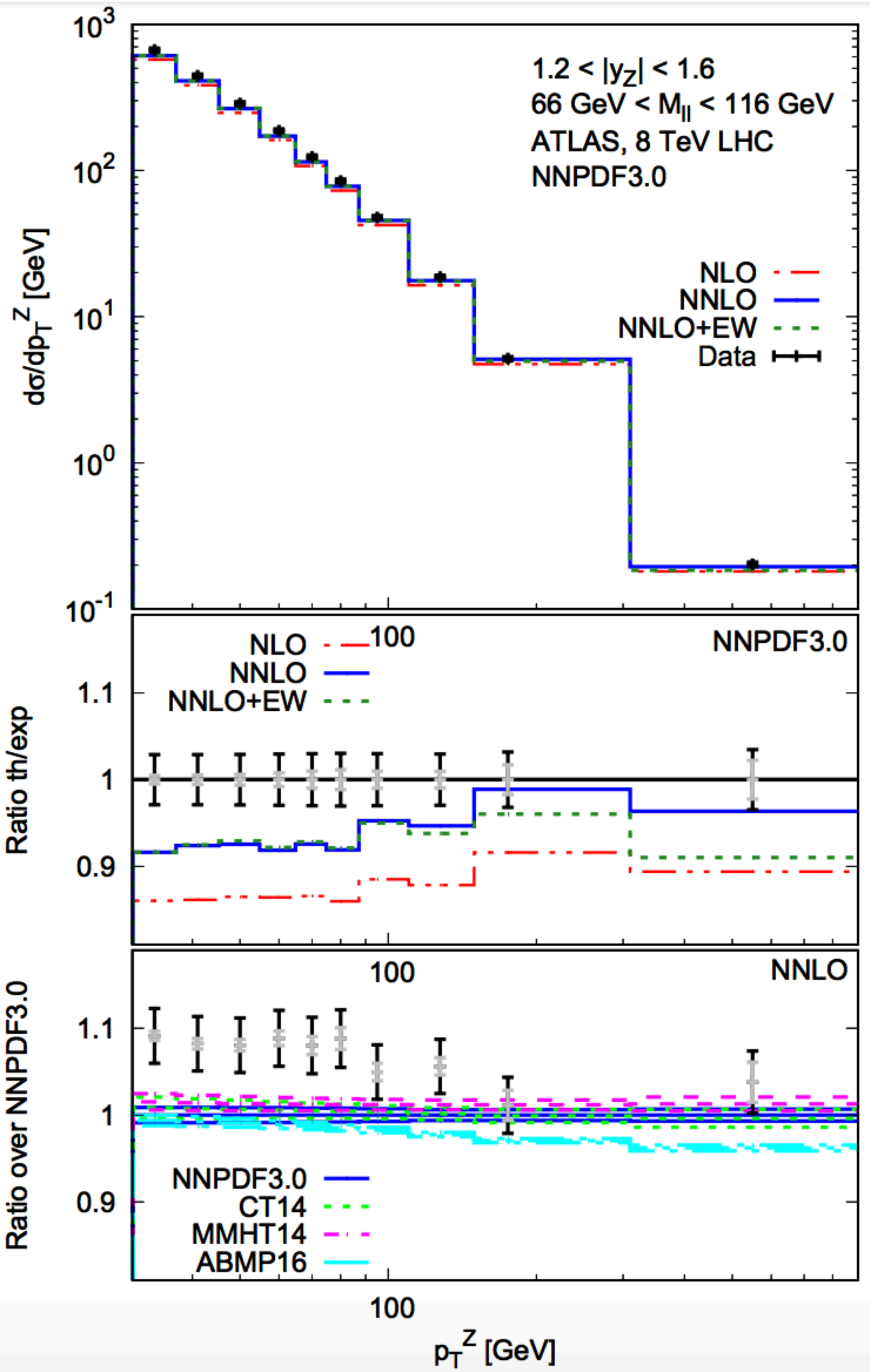
- Many schemes available

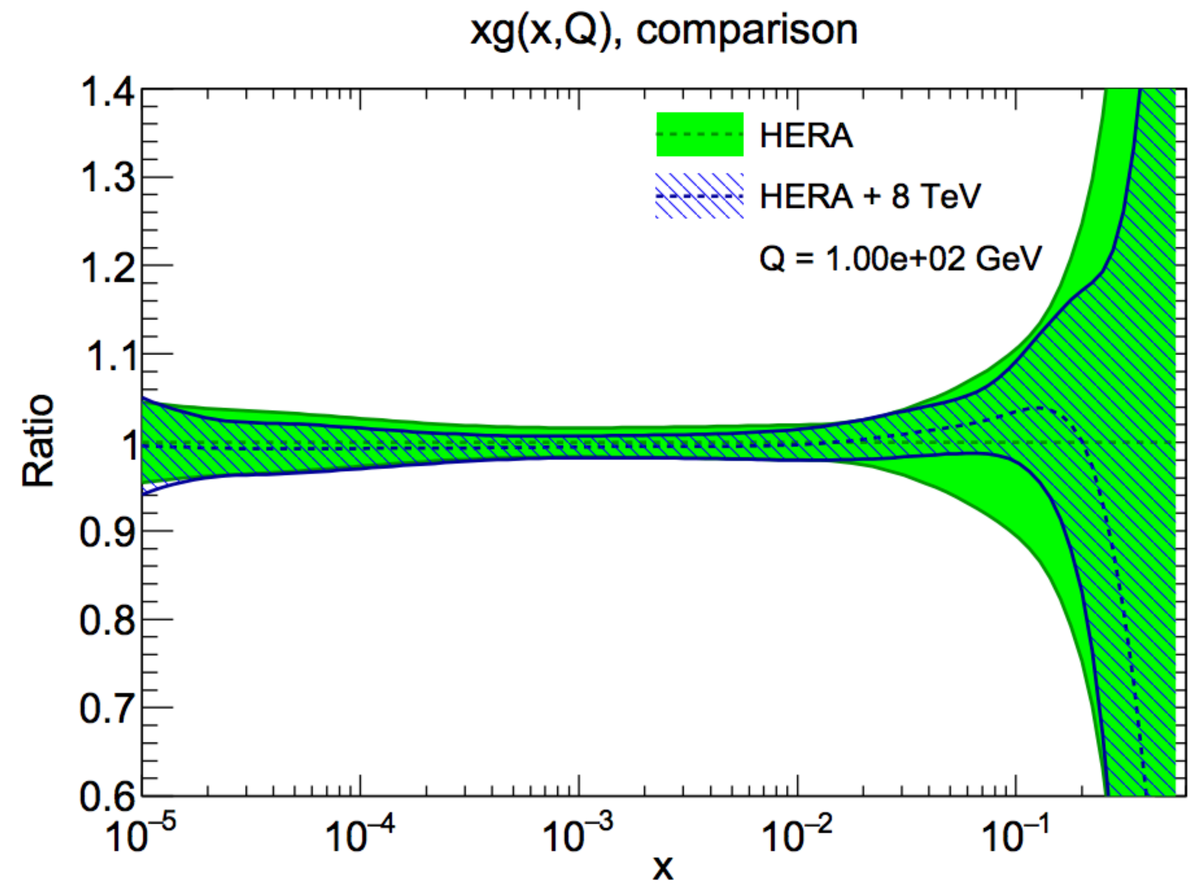
e.g. FONLL

$$
\begin{aligned}
\sigma^{(FONLL)} &= \sigma^{(4)} + \sigma^{(5)} - \text{double counting} \\
&= \mathcal{L}_{ij}(x_1, x_2, \mu^2) \otimes \sum_{p}^{N} \left(\alpha_S^{(5)}(\mu^2)\right)^p \\
&\times \left\{ \mathcal{B}_{ij}^{(p)}\left(x_1, x_2, \frac{\mu^2}{m_b^2}\right) + \sum_{k=0}^{\infty} \mathcal{A}_{ij}^{(p),(k)}(x_1, x_2)\left(\alpha_S^{(5)}(\mu^2) L\right)^k \right\} \\
&- \text{double counting}
\end{aligned}
$$

Gluon: Z transverse momentum

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

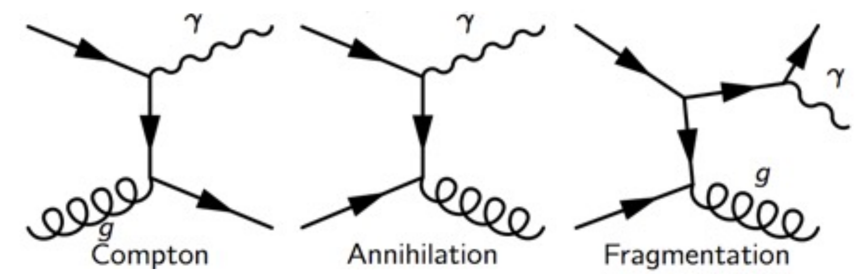


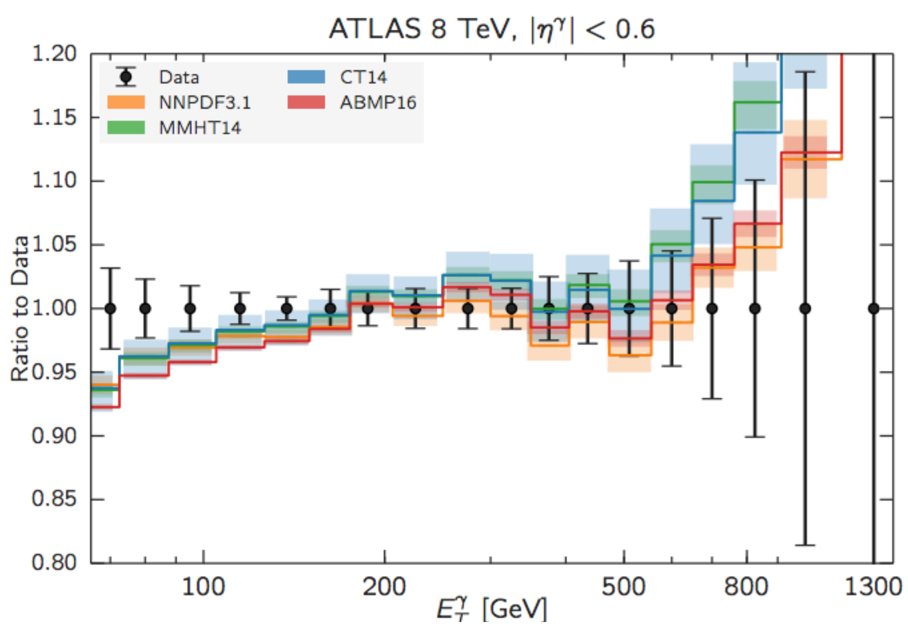


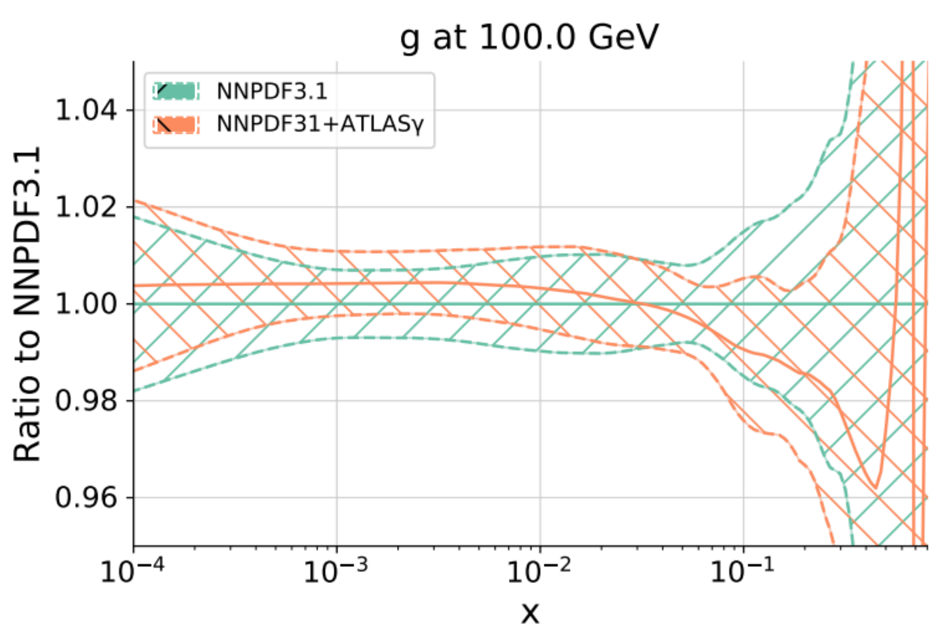
- Data/Theory comparison not so intuitive for correlation-dominated data
- Fluctuation in NNLO predictions (0.5 - 1%) had to be accounted for as extra nuisance parameter to get a good fit of such precise data

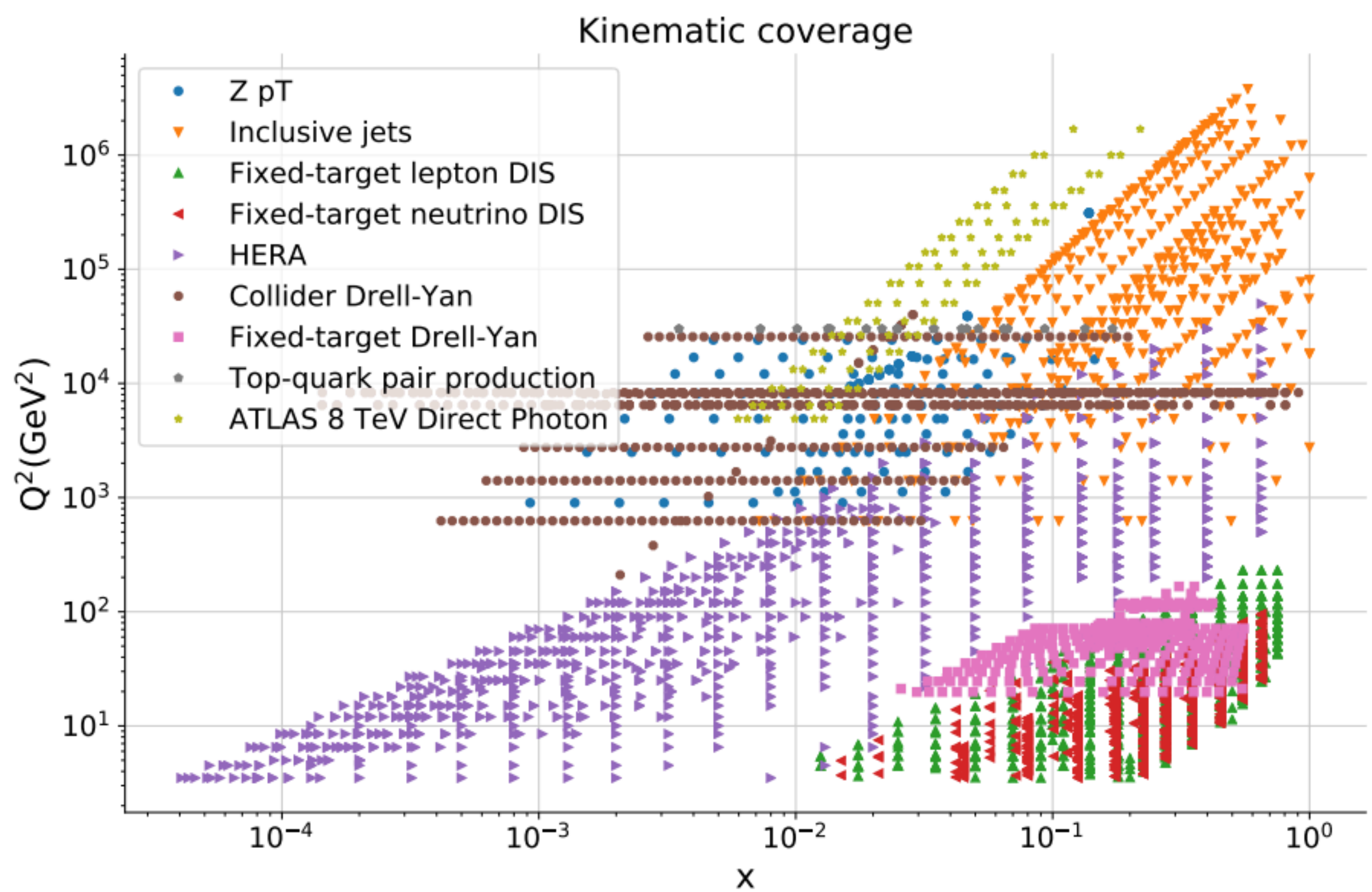
Gluon: direct photon production

- Prompt photon production directly sensitive to the gluon-quark luminosity via Compton scattering
- Isolated prompt photon data known at NNLO [Campbell et al 1612.04333] and accurately measured by ATLAS



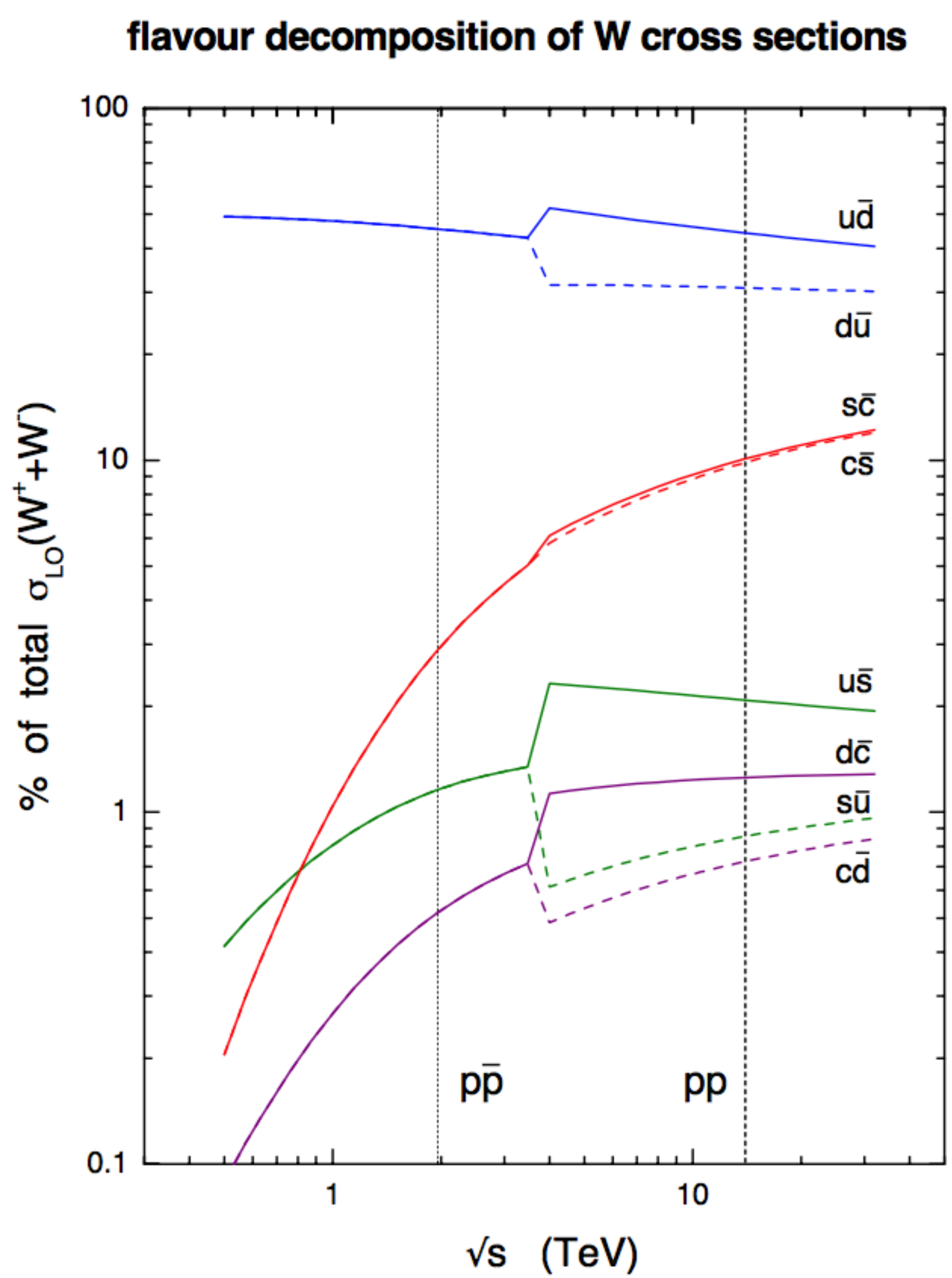


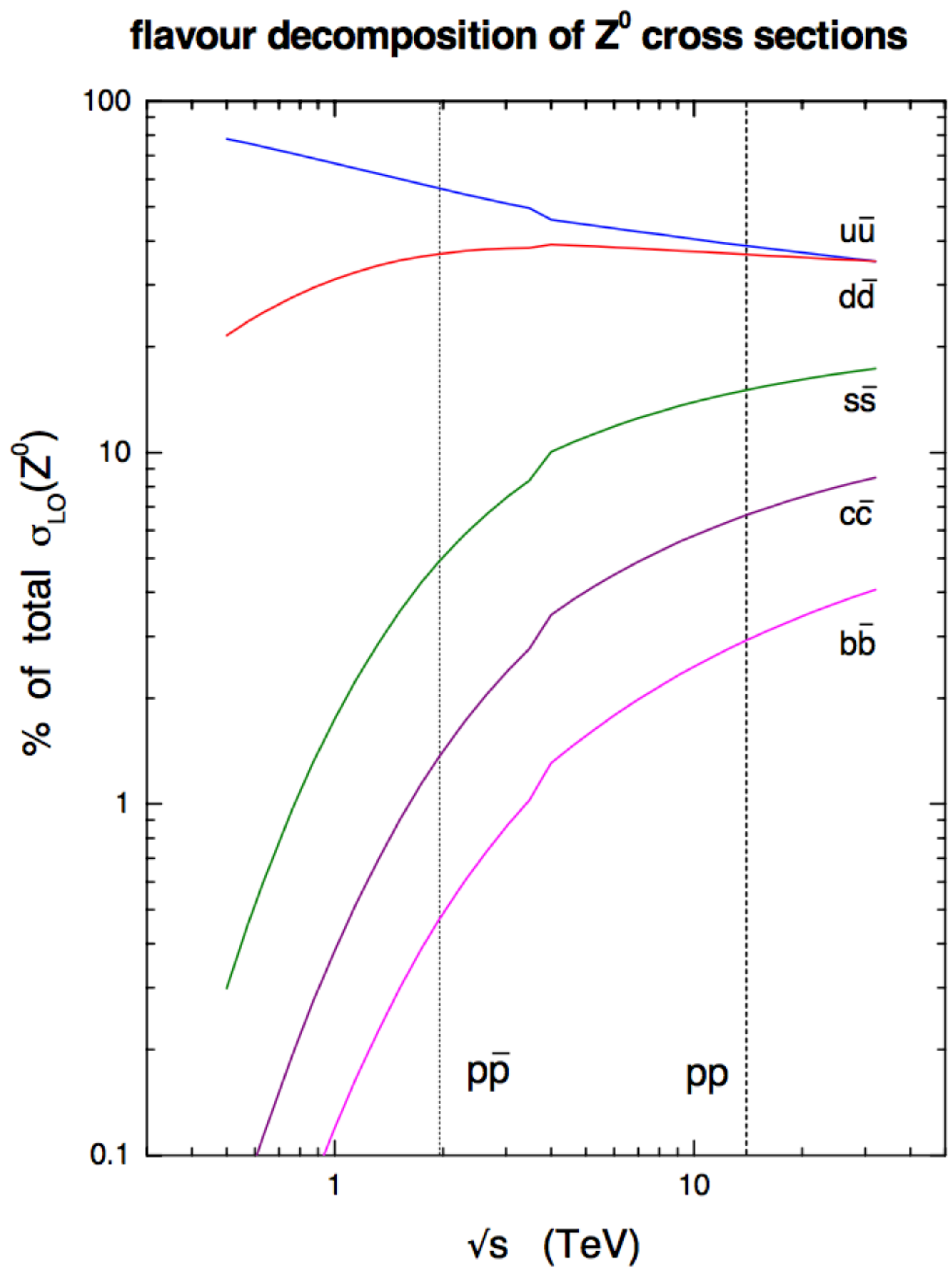




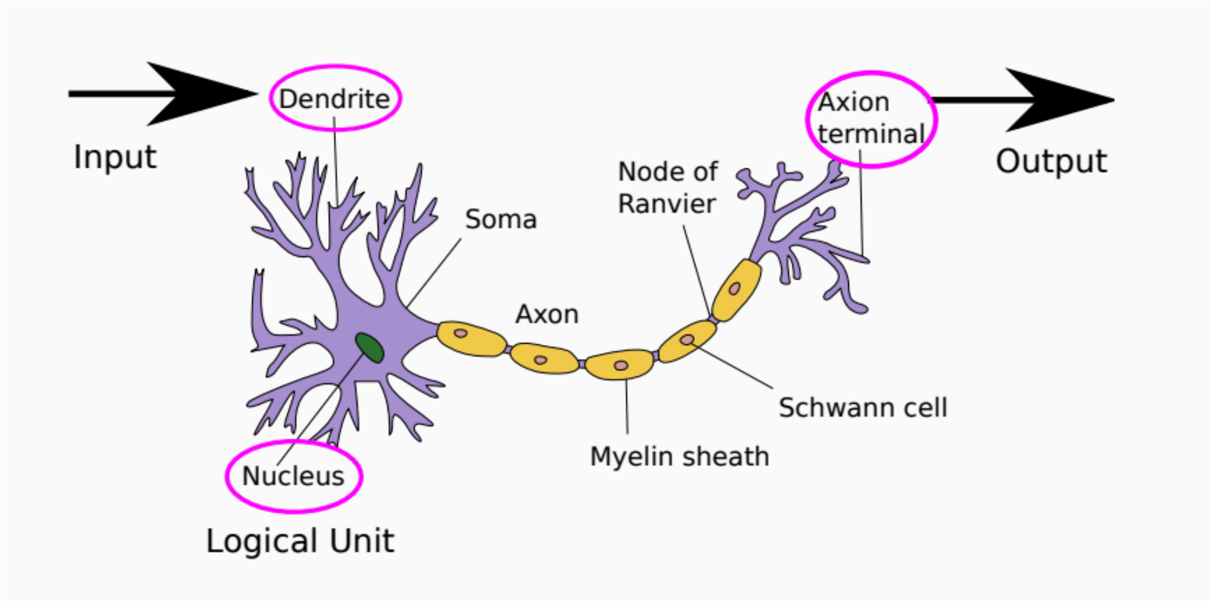
Campbell et al 1802.03021

Parton Luminosities





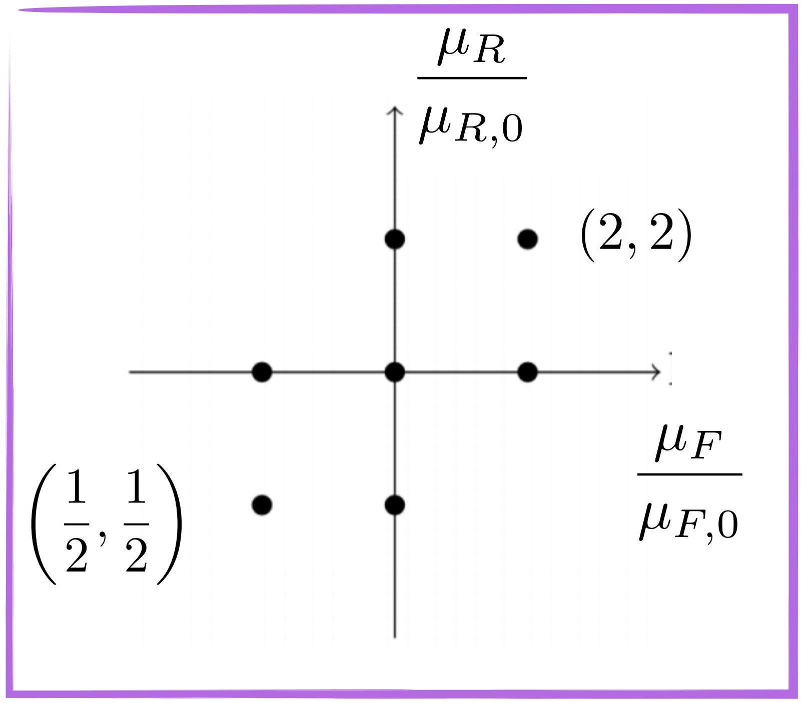
Neural networks and ML



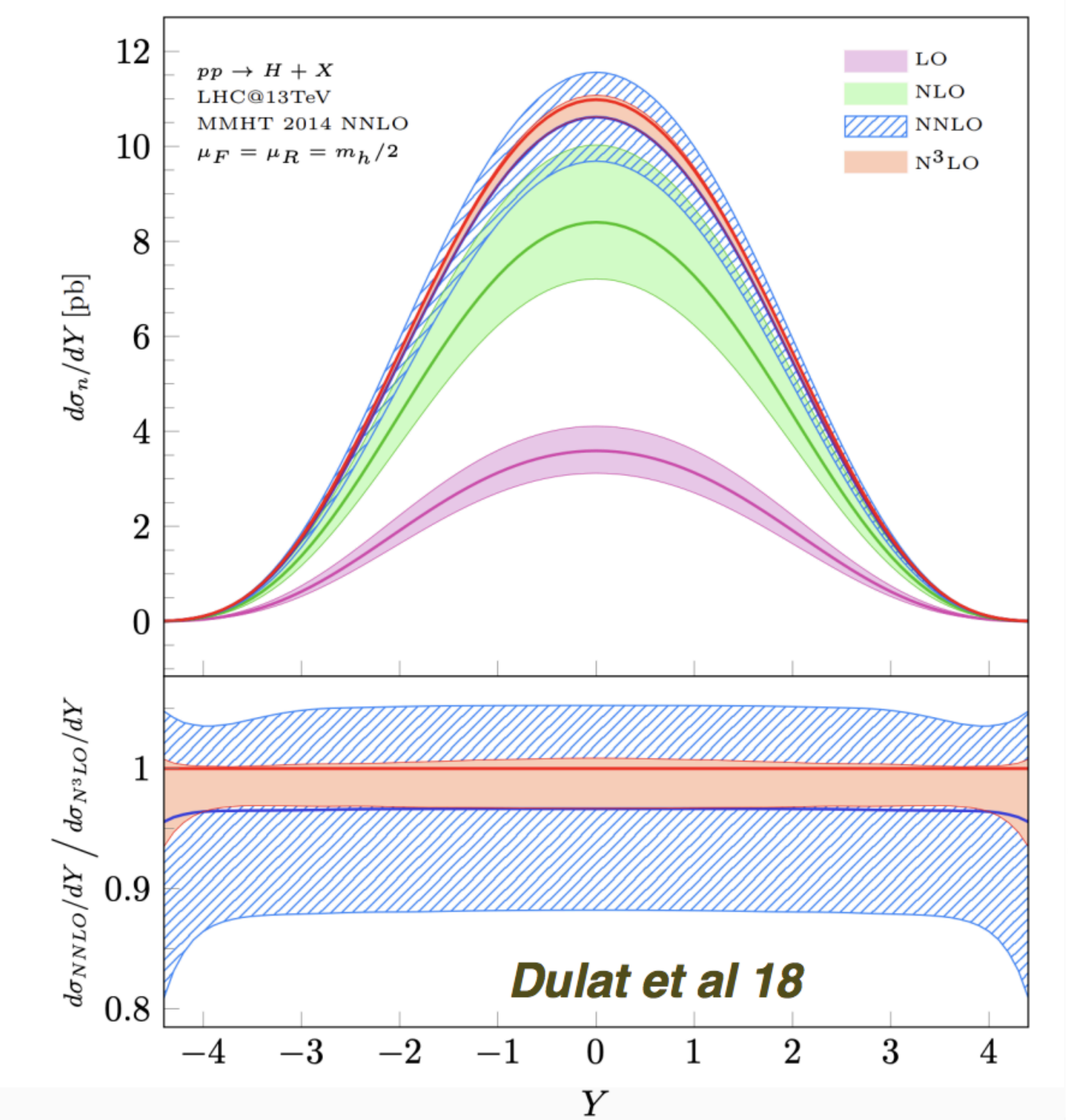
S. Carrazza, Colloquium, S. Paolo, N3PDF

- Artificial neural networks are computer systems inspired by the biological neural networks in the brain
- Data communication pattern
- Currently state-of-the-art for several Machine Learning Applications

MHOU in theoretical predictions

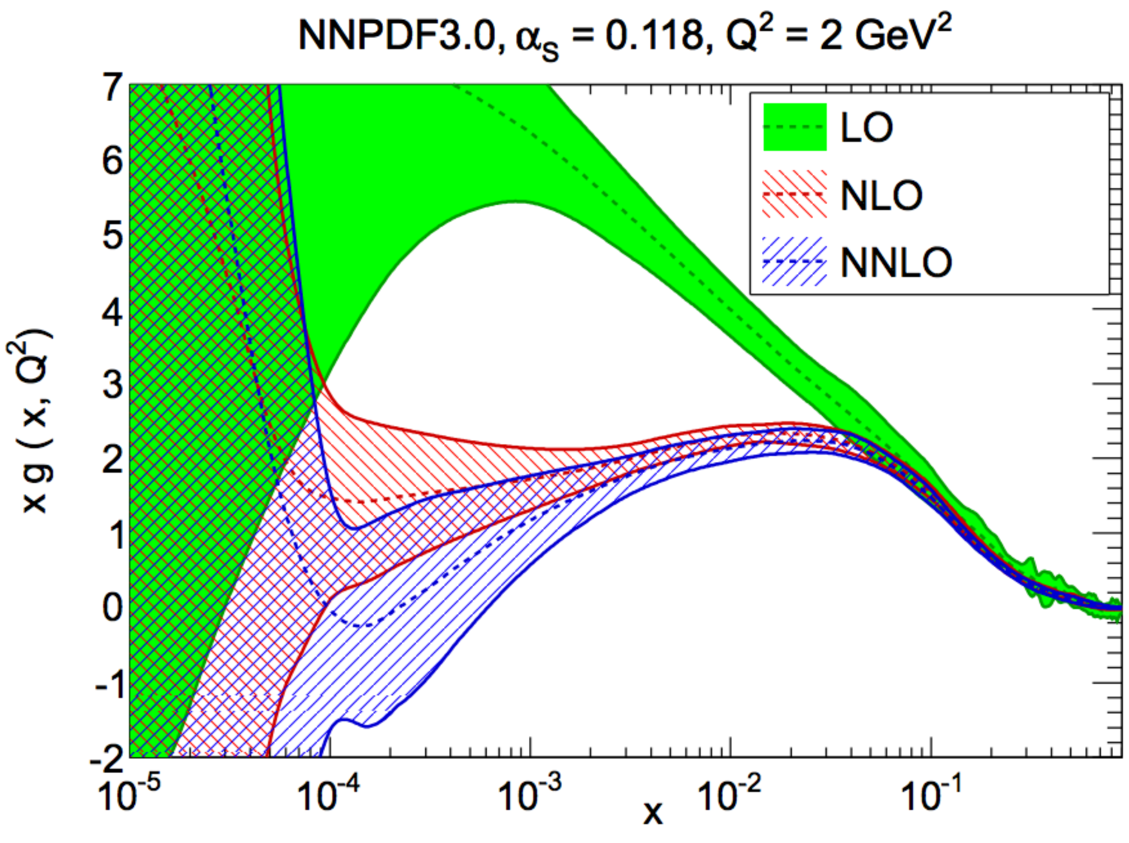


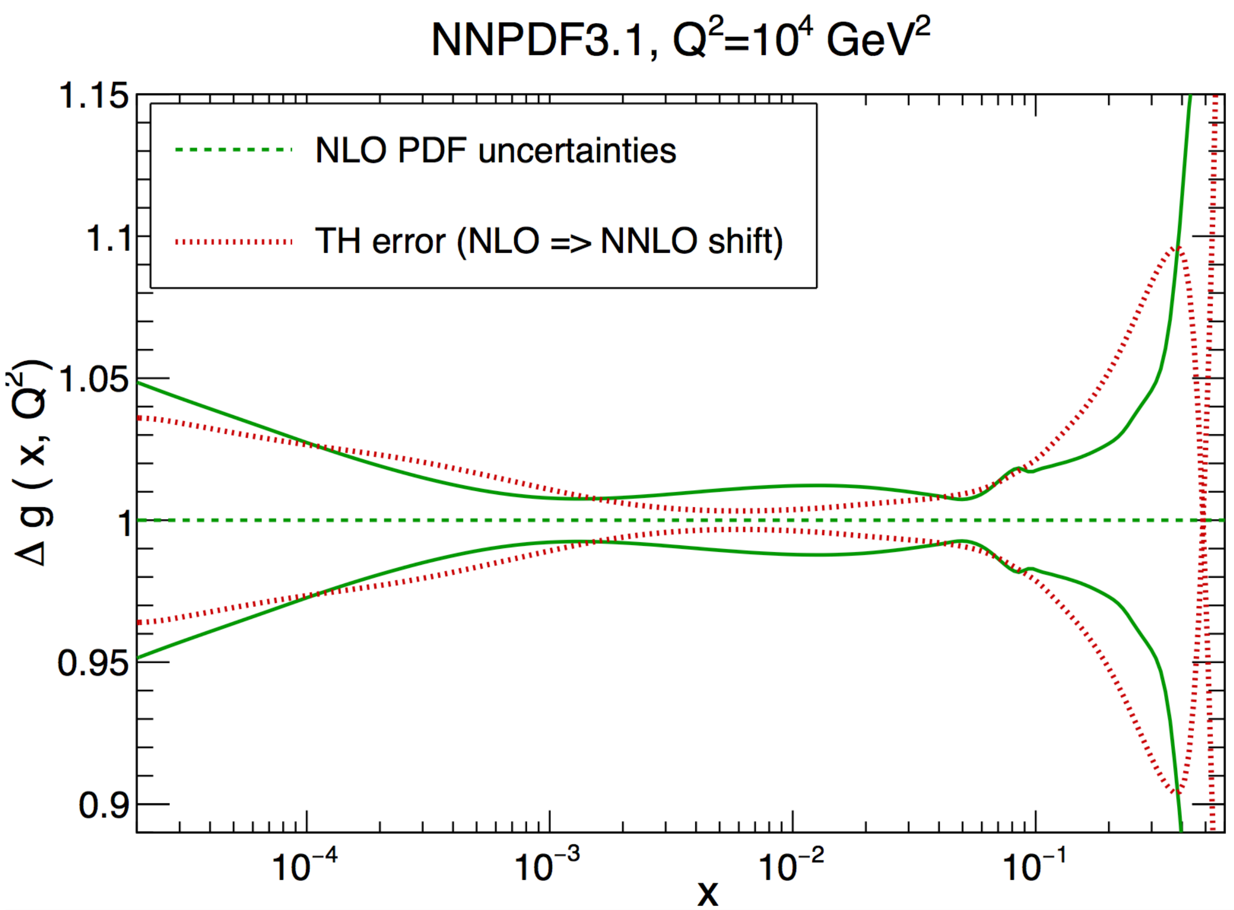
Increasing order in perturbation theory reduced "scale" uncertainty (or MHOU) in theoretical predictions



MHOU in PDF fits

- 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

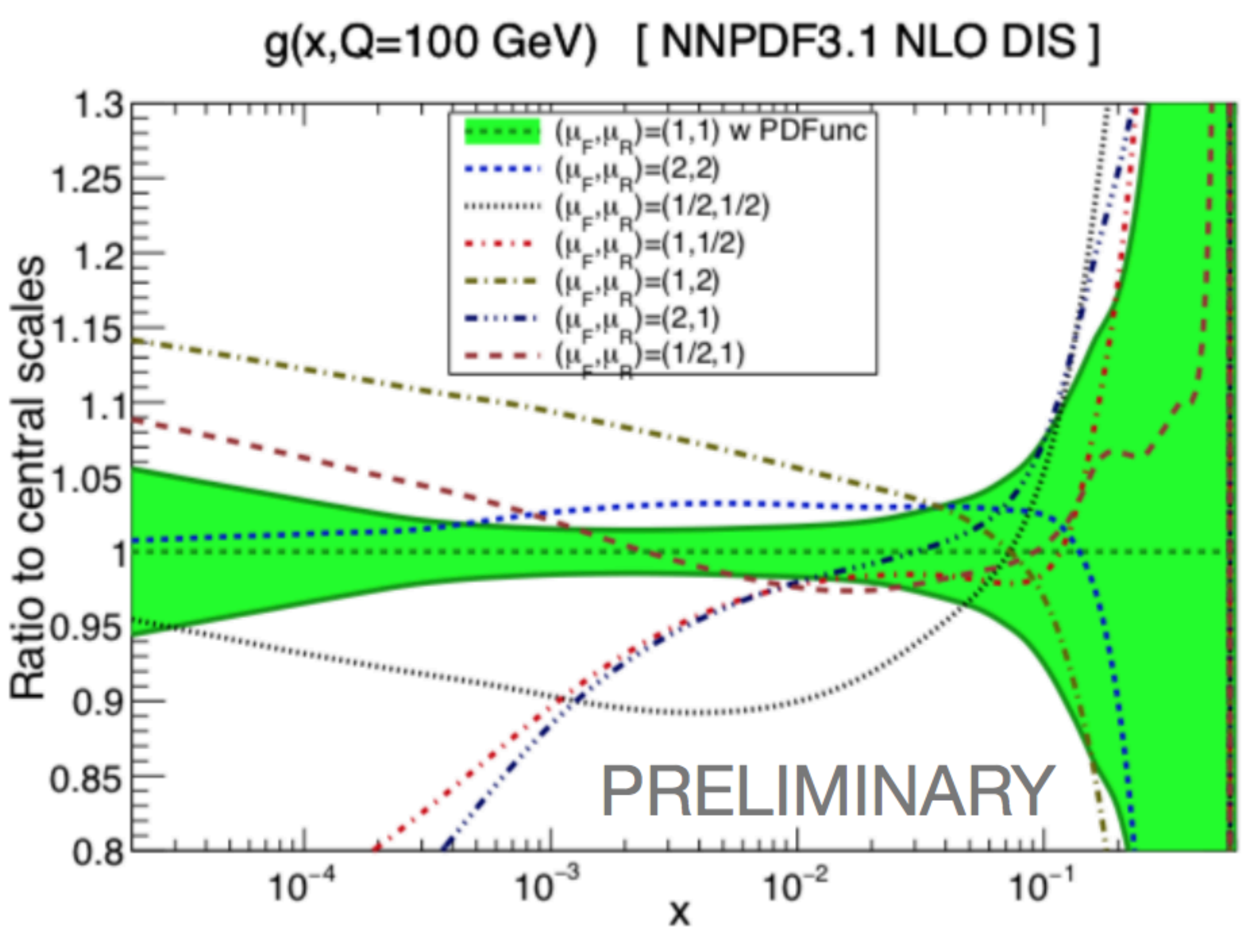


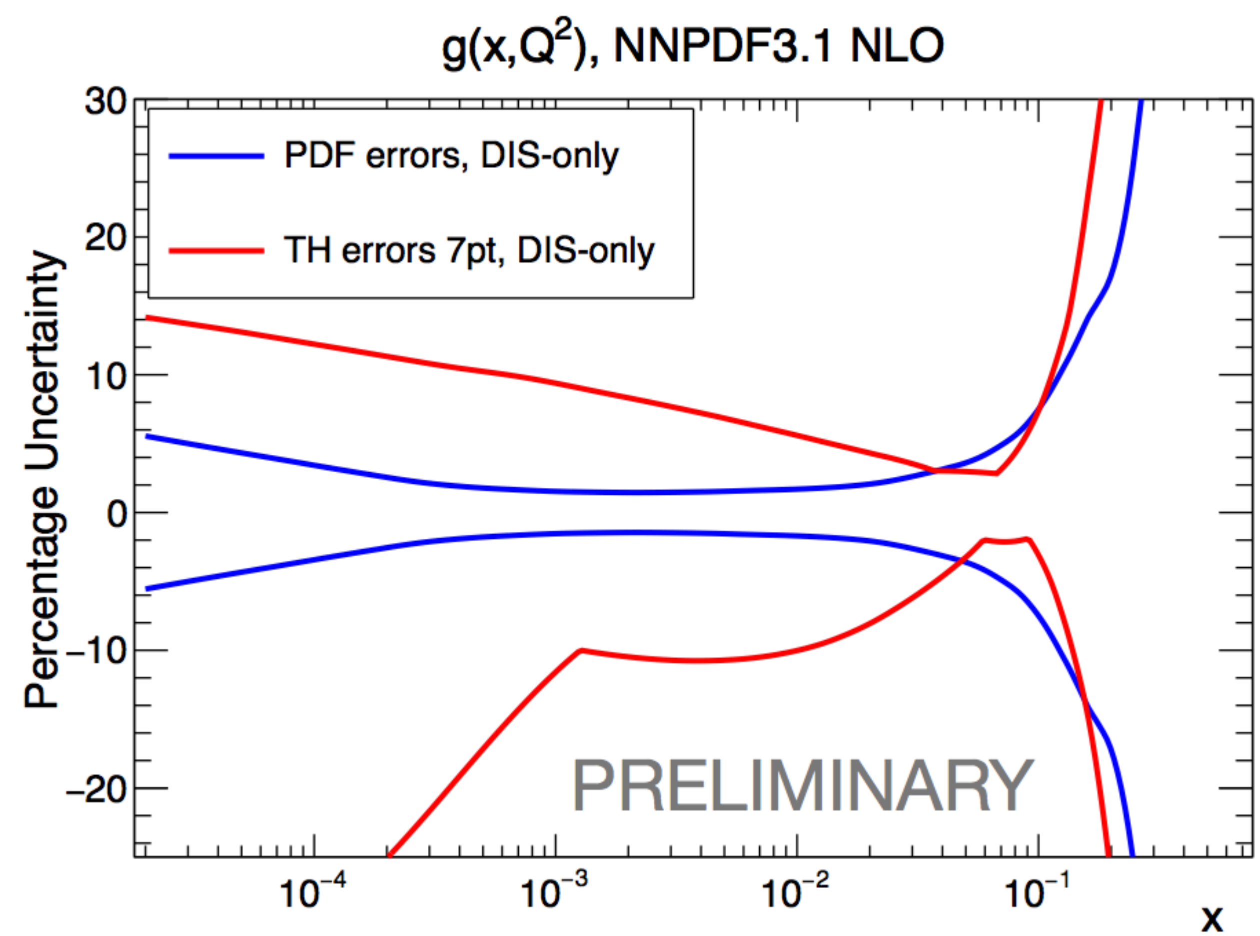


Ball et al, EPJC 77 (2017)

MHOU in PDF fits

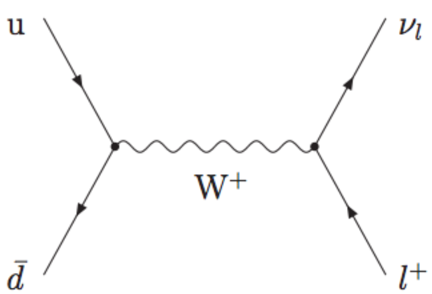
- How to estimate MHOU in PDF fits?
- Compare fits with varied scales
- Useful to have indication on the size of MHOU in PDFs
- A posteriori combination?
- How to include them in the fitting methodology along with other sources of theoretical uncertainty? - see tomorrow's lecture



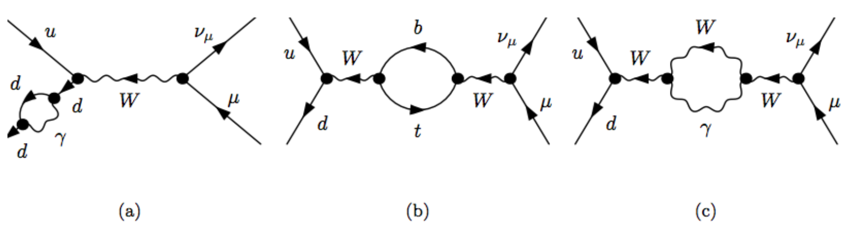


Electroweak corrections

- Because α(Mz) ~ α_S(Mz)/10 ⟹ NLO EW corrections ~ NNLO QCD corrections



$$\mathcal{O}(\alpha^2)\,[1]$$

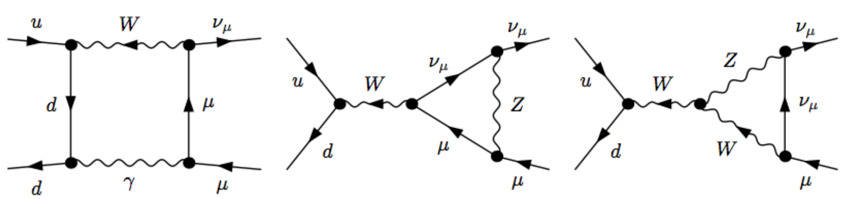


Virtual corrections

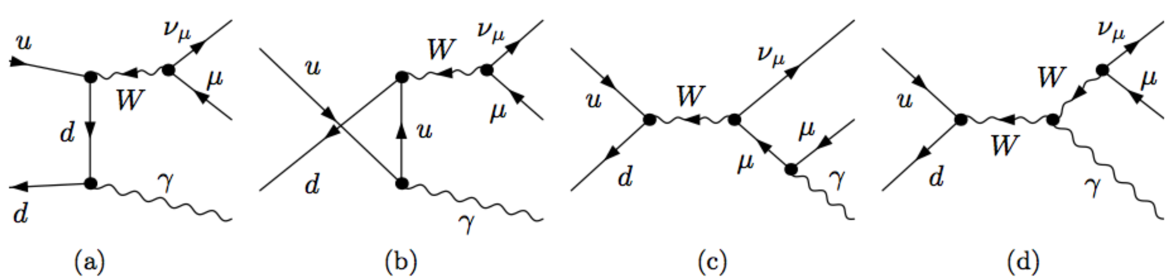
$$\mathcal{O}(\alpha^2)\,[1+\mathcal{O}(\alpha)]$$

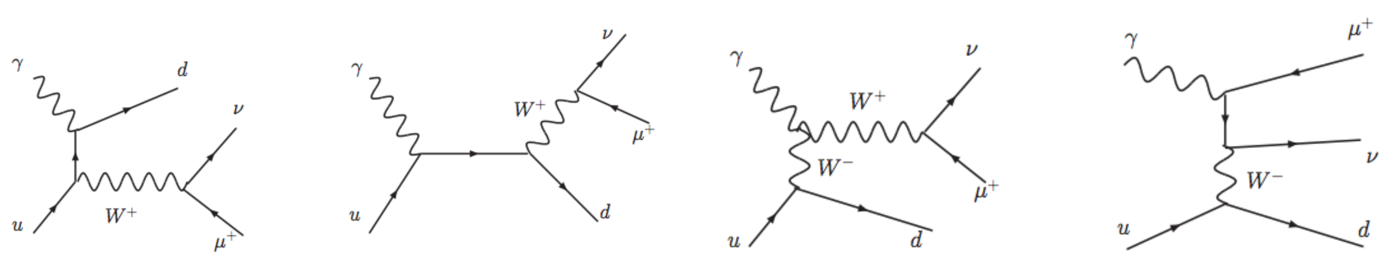
$$+\frac{3\alpha}{\pi s_W^2}\ln\left(\frac{s}{M_W^2}\right)$$

At large s these logs can become large



Real corrections - quark initiated

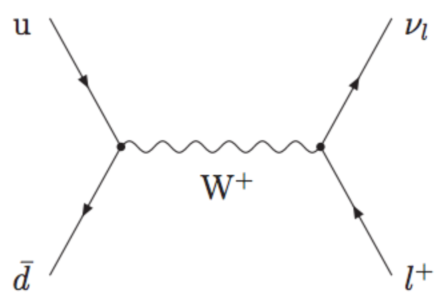




Real corrections - photon initiated

Electroweak corrections

- Because $\alpha(M_Z) \sim \alpha_S(M_Z)/10 \Longrightarrow$ NLO EW corrections ~ NNLO QCD corrections



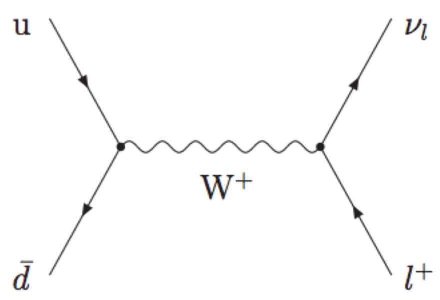
- NLO EW corrections become large in the large pT region of lepton but partially compensated by photon-initiated real corrections

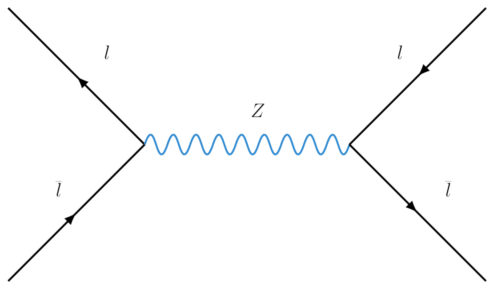
$p_{T,l}$/GeV	25–∞	50–∞	100–∞	200–∞	500–∞	1000–∞
$\delta_{e^+\nu_e}/\%$	−5.19(1)	−8.92(3)	−11.47(2)	−16.01(2)	−26.35(1)	−37.92(1)
$\delta_{\mu^+\nu_\mu}/\%$	−2.75(1)	−4.78(3)	−8.19(2)	−12.71(2)	−22.64(1)	−33.54(2)
$\delta_{rec}/\%$	−1.73(1)	−2.45(3)	−5.91(2)	−9.99(2)	−18.95(1)	−28.60(1)
$\delta_{\gamma q}/\%$	+0.071(1)	+5.24(1)	+13.10(1)	+16.44(2)	+14.30(1)	+11.89(1)

Dittmaier, Krämer

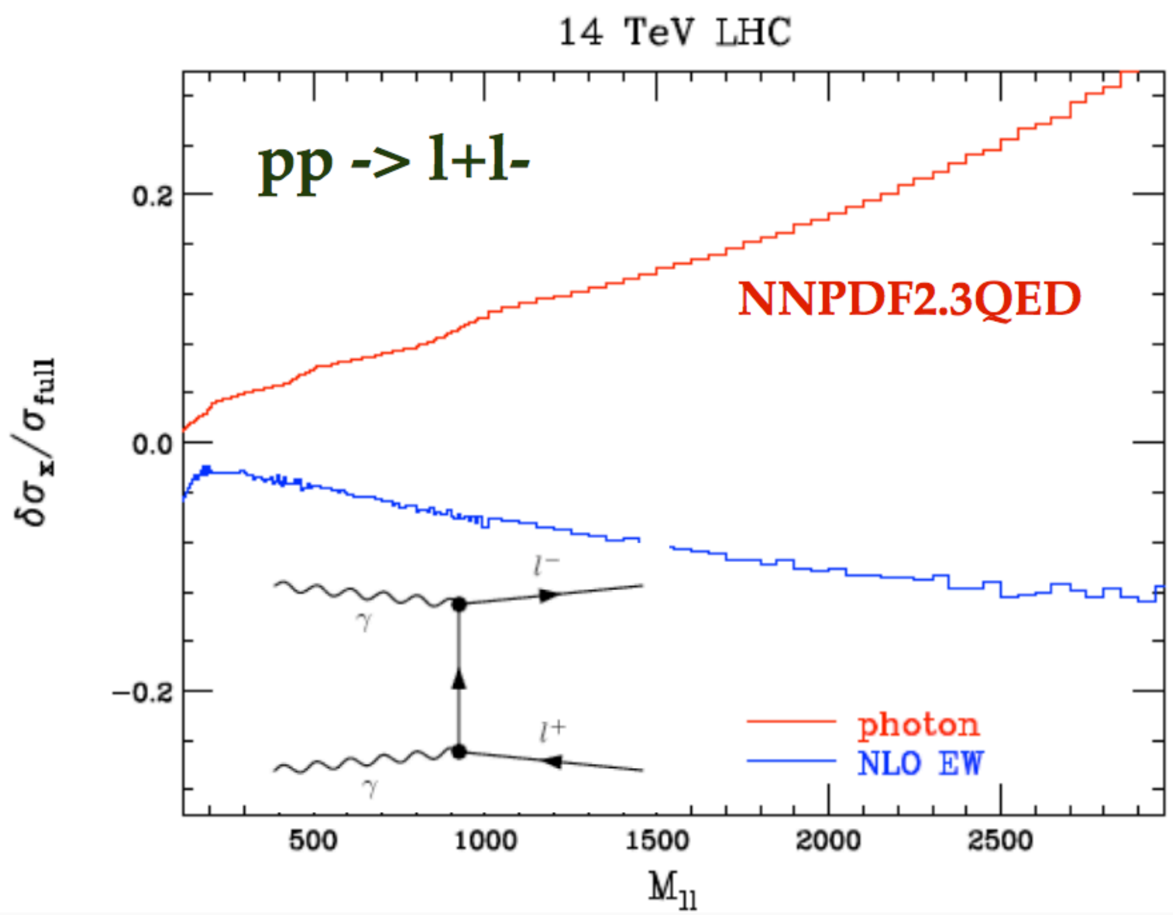
Electroweak corrections

- Because $\alpha(M_Z) \sim \alpha_S(M_Z)/10 \Longrightarrow$ NLO EW corrections ~ NNLO QCD corrections





- NLO EW corrections become large in the large pT region of lepton but partially compensated by photon-initiated real corrections



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

Modified DGLAP

- How are PDFs modified by inclusion of initial photon PDF?

$$Q^2 \frac{\partial}{\partial Q^2} g(x, Q^2) = \sum_{q,\bar{q},g} P_{ga}(x, \alpha_s(Q^2)) \otimes f_a(x, Q^2) + P_{g\gamma}(x, \alpha_s(Q^2)) \otimes \gamma(x, Q^2),$$

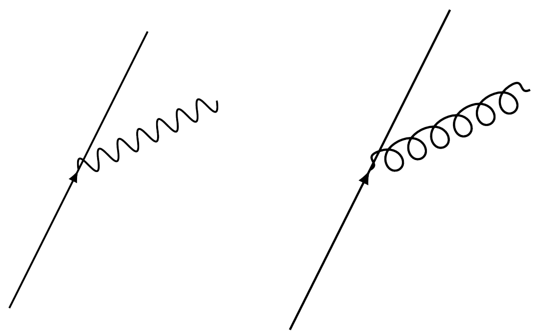
$$Q^2 \frac{\partial}{\partial Q^2} q(x, Q^2) = \sum_{q,\bar{q},g} P_{qa}(x, \alpha_s(Q^2)) \otimes f_a(x, Q^2) + P_{q\gamma}(x, \alpha_s(Q^2)) \otimes \gamma(x, Q^2),$$

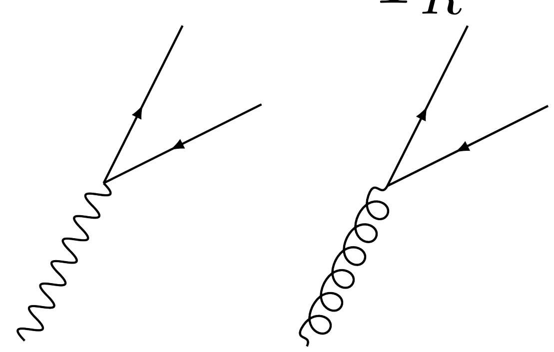
$$Q^2 \frac{\partial}{\partial Q^2} \gamma(x, Q^2) = P_{\gamma\gamma} \otimes \gamma(x, Q^2) + \sum_{q,\bar{q},g} P_{\gamma a}(x, \alpha_s(Q^2)) \otimes f_a(x, Q^2).$$

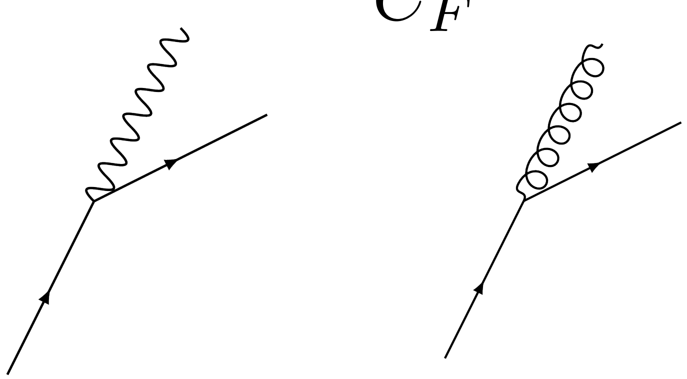
- DGLAP splitting functions expanded in powers of α_s and $\boldsymbol{\alpha}$

$$P_{ij} = \sum_{m,n} \left(\frac{\alpha_S}{2\pi}\right)^m \left(\frac{\alpha}{2\pi}\right)^n P_{ij}^{(m,n)}$$

$$P_{qq}^{(0,1)} = \frac{e_q^2}{C_F} P_{qq}^{(1,0)} \qquad P_{q\gamma}^{(0,1)} = \frac{e_q^2}{T_R} P_{qg}^{(1,0)} \qquad P_{\gamma q}^{(0,1)} = \frac{e_q^2}{C_F} P_{gq}^{(1,0)}$$

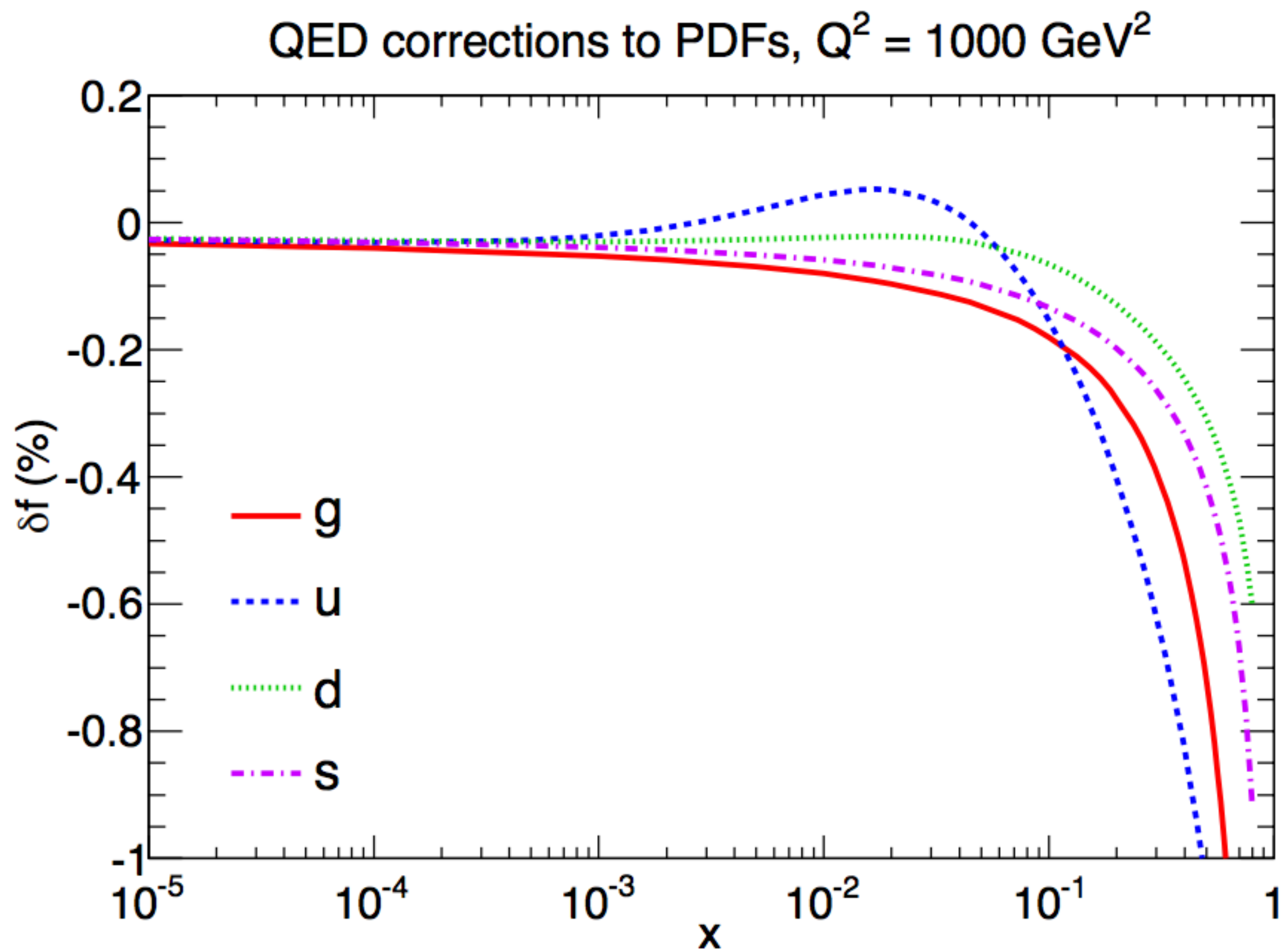






Modified DGLAP

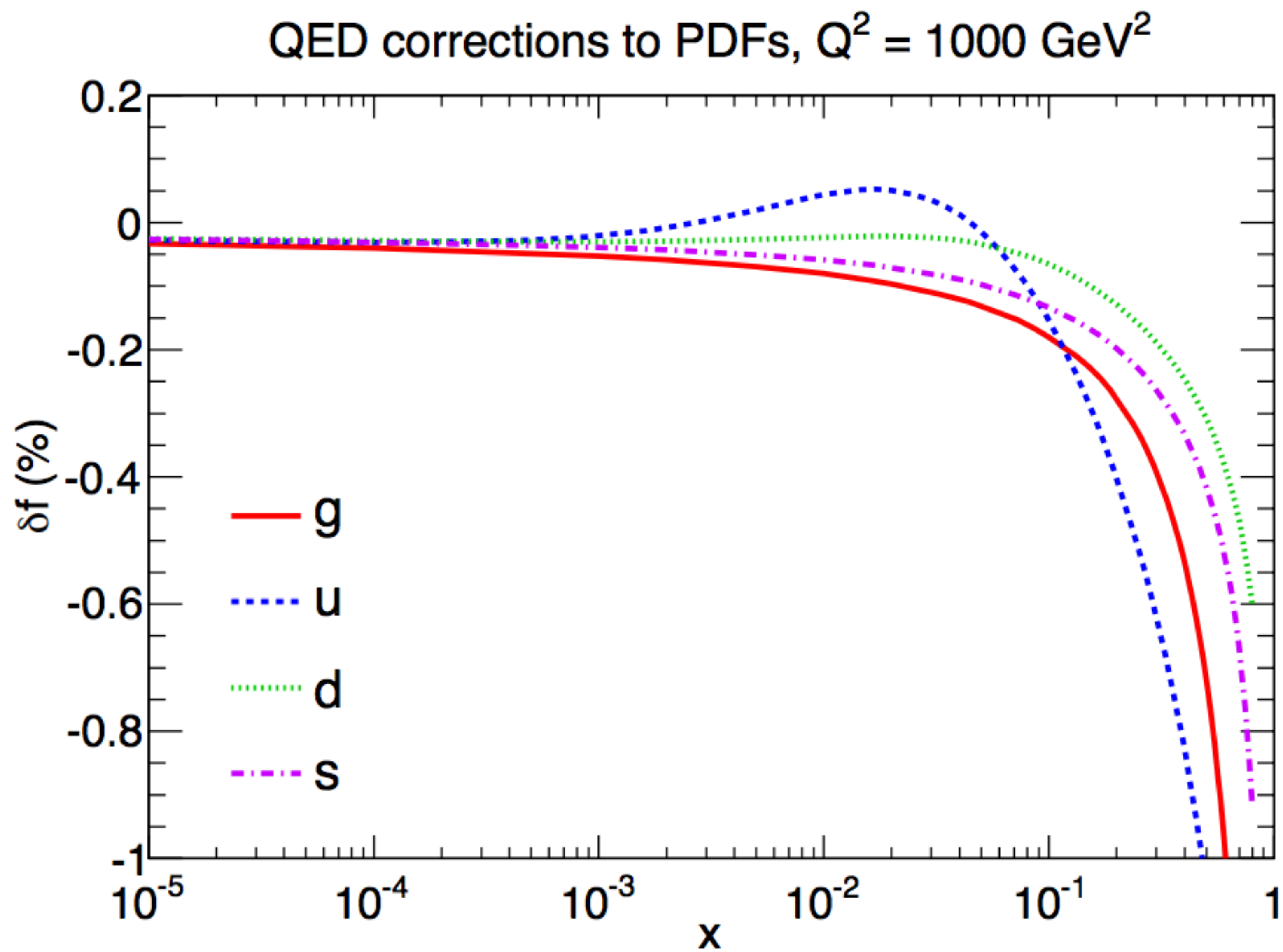
- Quark and gluon PDFs change up to 1% at large x



Ball et al, Nucl.Phys. B877 (2013) 290-320

Modified DGLAP

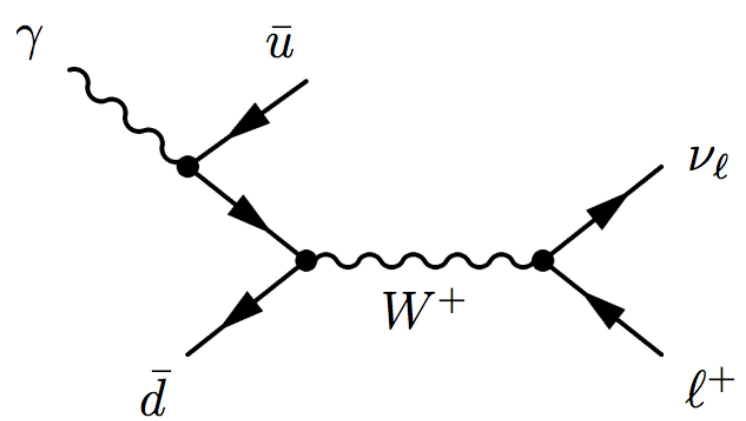
- Quark and gluon PDFs change up to 1% at large x
- How do we determine the photon PDF?
- Two ways in the next slides: from data or from theory
- In the best possible world: theory input and data input together

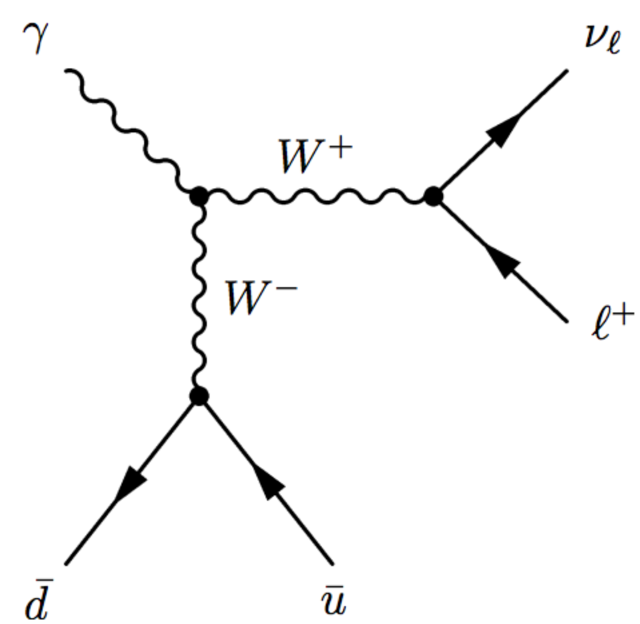


Ball et al, Nucl.Phys. B877 (2013) 290-320

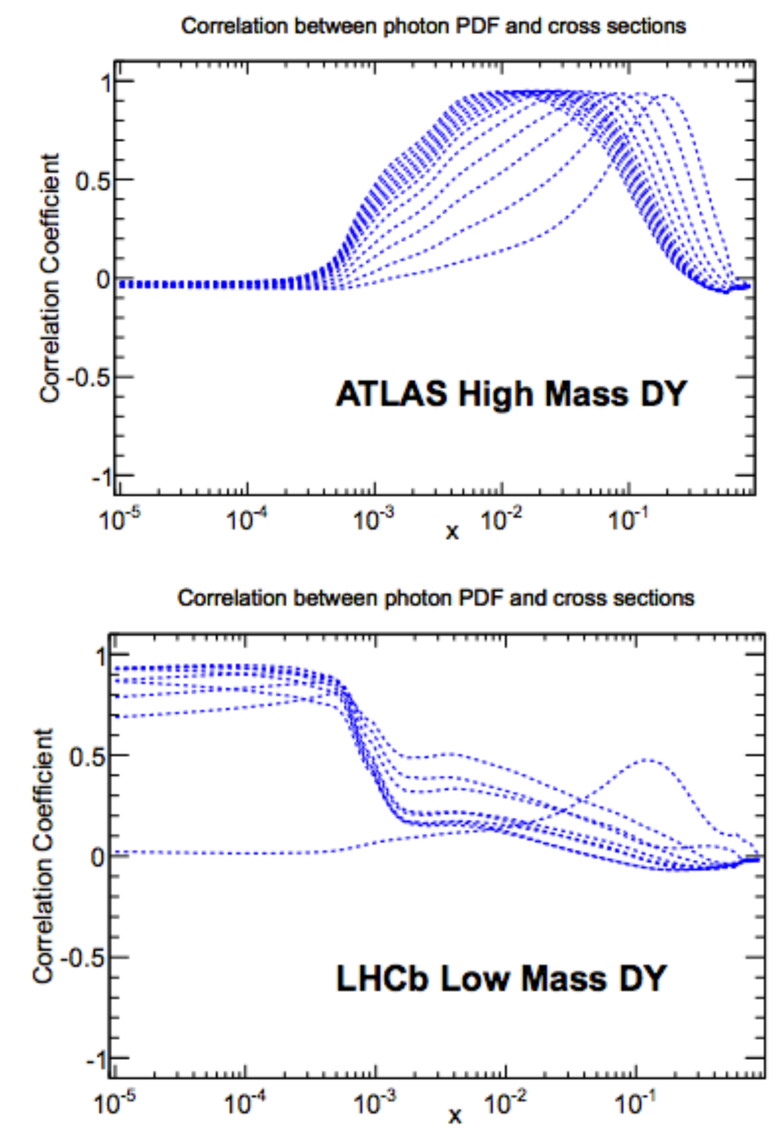
Photon PDFs

- Largest correlations between photon PDFs and pp cross sections are for Drell-Yan processes, but also for top pair production and VV production





$Q = 100\,\mathrm{GeV}$

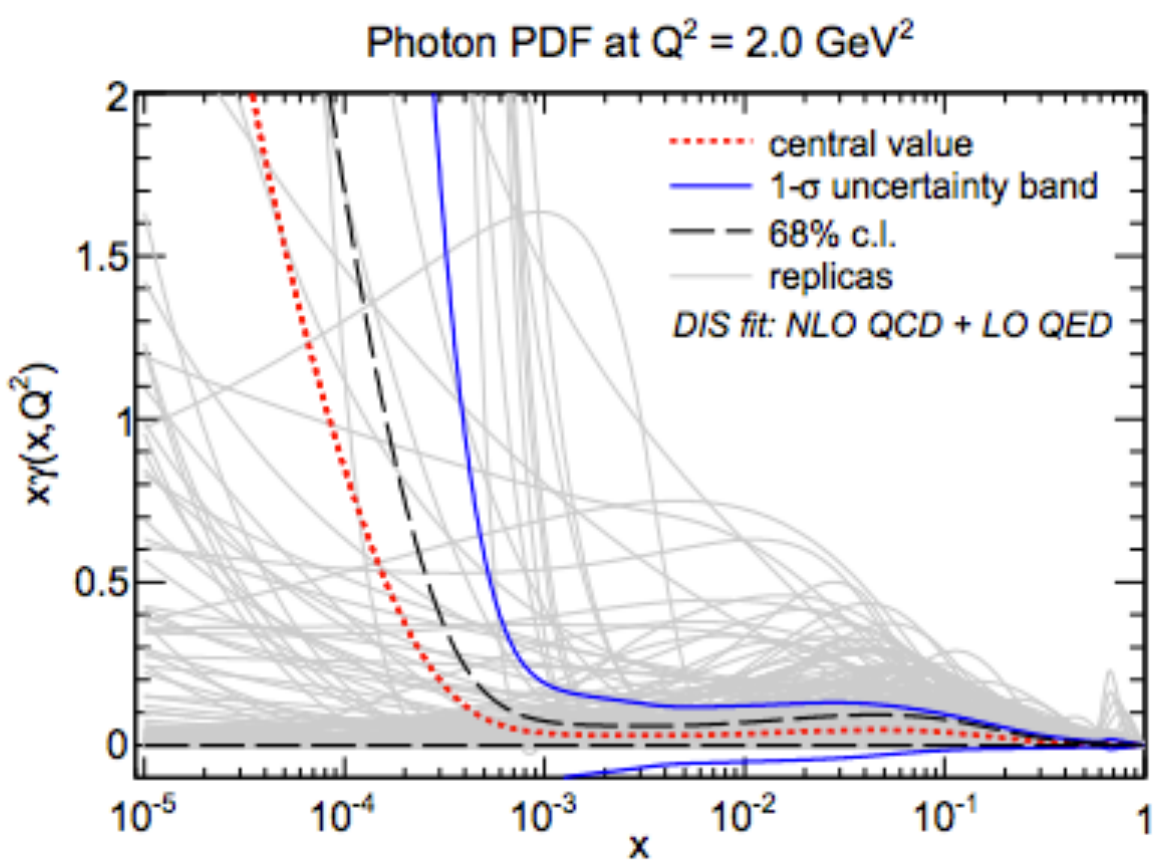


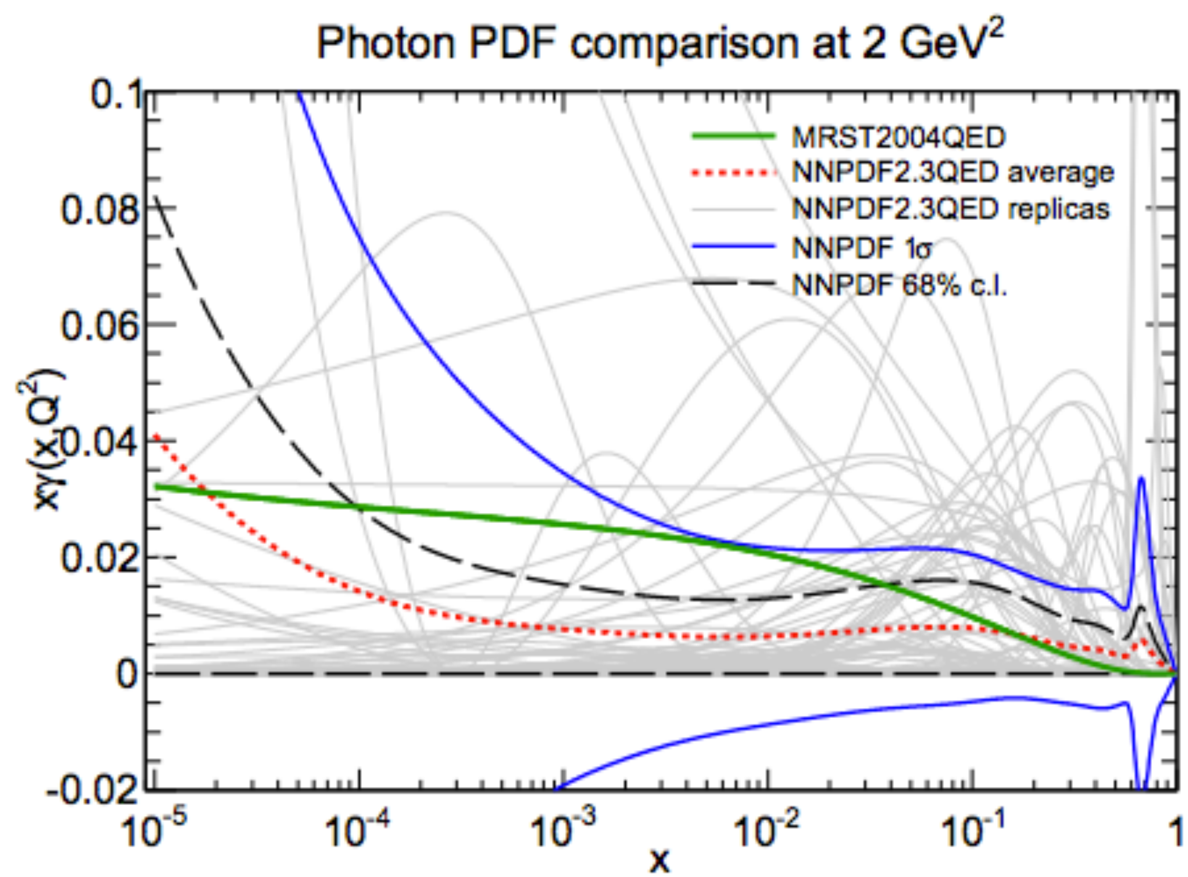
<u>Photon-induced Drell-Yan</u>

Ball et al, Nucl.Phys. B877 (2013) 290-320

Photon PDFs

Data-driven knowledge

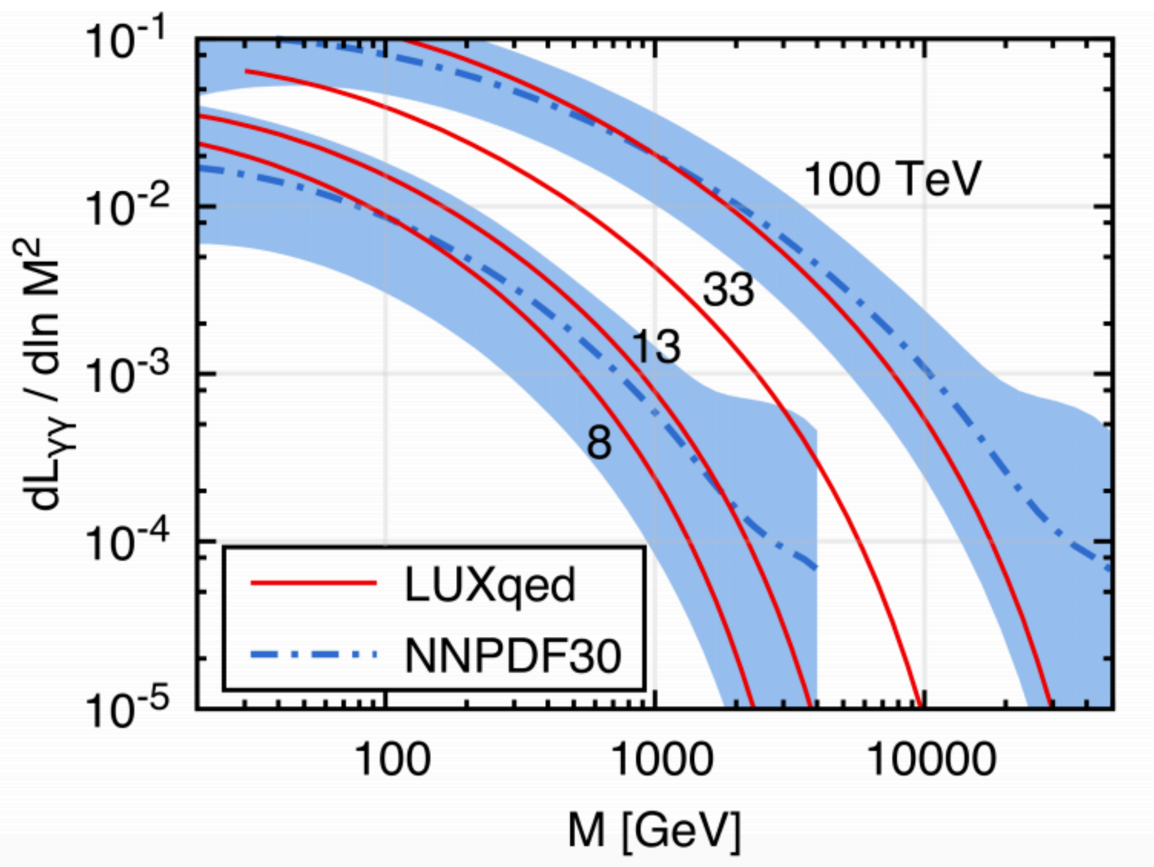


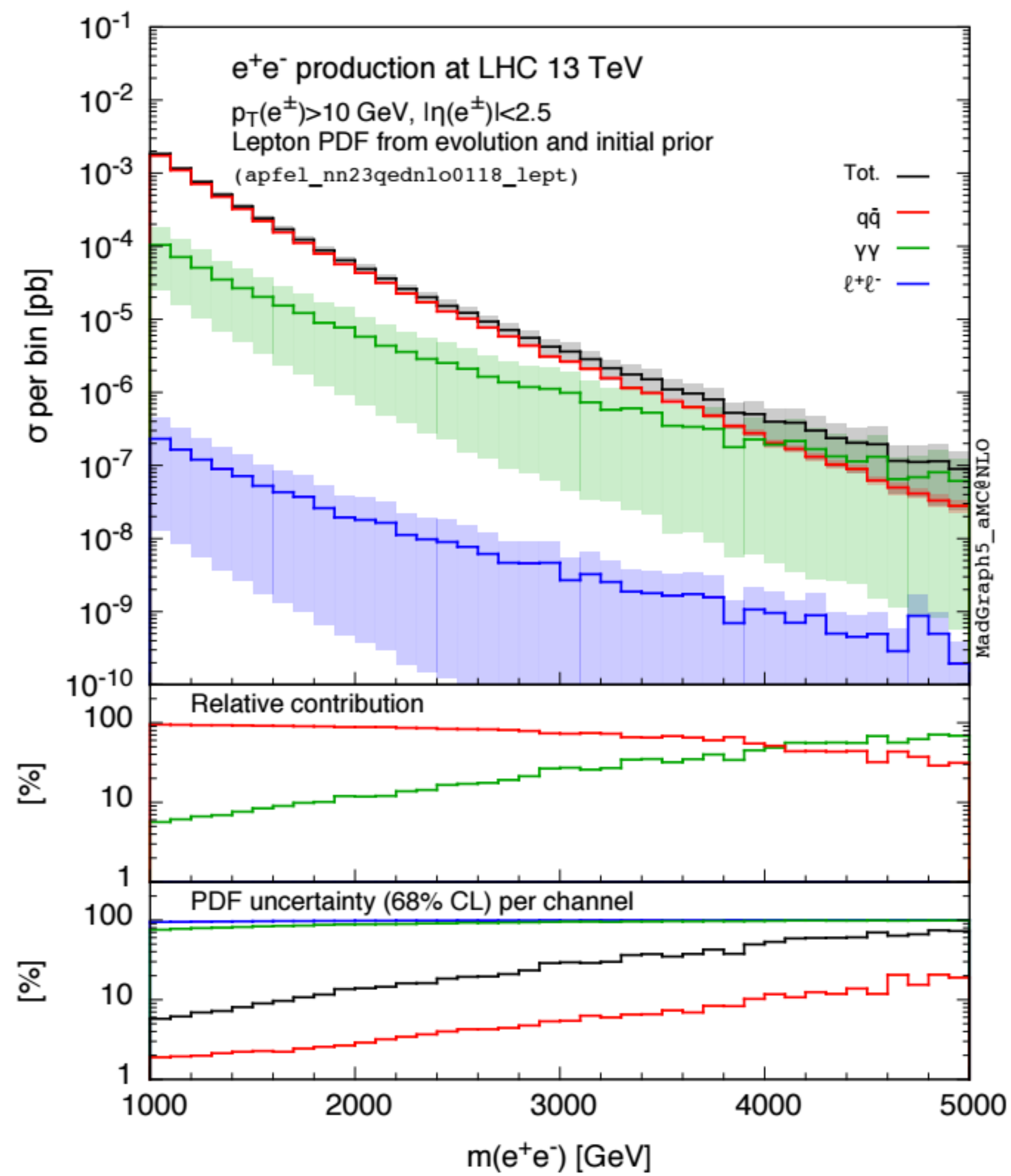


Ball et al, Nucl.Phys. B877 (2013) 290-320

Photon PDFs

- Data-driven approach associated with a large uncertainty on photon PDF
- Theory breakthrough: LUX PDF [Manohar, Nason, Salam, Zanderighi,1607.04266]





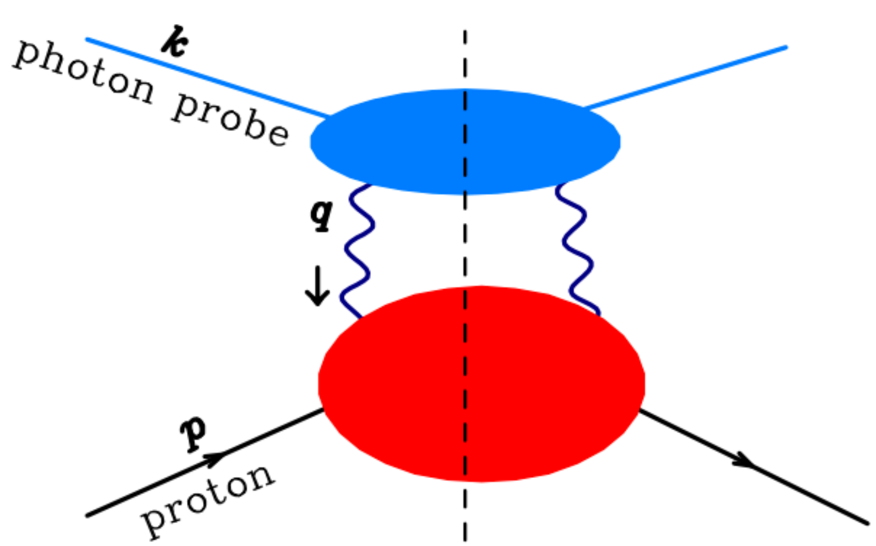
Bertone et al, 1508.07002

Photon PDFs

- QED is perturbative down to low scales ⟹ The photon must be computable is the input mark substructure is known
- Manohar et al: write the cross section for a chosen BSM process, e.g. production of heavy supersymmetric lepton L in ep collision (Drees, Zeppenfeld 1989)

$$\sigma = \frac{1}{4p\cdot k}\int \frac{d^4q}{(2\pi)^4 q^4} e^2_{\rm ph}(q^2)\left[4\pi W_{\mu\nu}(p,q)\, L^{\mu\nu}(k,q)\right]\; 2\pi\delta((k-q)^2 - M^2)$$

$$l(k) + p(p) \to L(k') + X$$



$$\sigma = c_0 \sum_a \int_x^1 \frac{dz}{z}\,\hat\sigma_a(z,\mu^2)\,\frac{M^2}{zs} f_{a/p}\left(\frac{M^2}{zs},\mu^2\right)$$

$$\sigma = \frac{c_0}{2\pi}\int_x^{1-\frac{2xm_p}{M}} \frac{dz}{z}\int_{Q^2_{\rm min}}^{Q^2_{\rm max}} \frac{dQ^2}{Q^2}\alpha^2_{\rm ph}(-Q^2)\Bigg[\Bigg(2-2z+z^2 + \frac{2x^2m_p^2}{Q^2} + \frac{z^2Q^2}{M^2} - \frac{2zQ^2}{M^2} - \frac{2x^2Q^2m_p^2}{M^4}\Bigg)F_2(x/z,Q^2) + \left(-z^2 - \frac{z^2Q^2}{2M^2} + \frac{z^2Q^4}{2M^4}\right)F_L(x/z,Q^2)\Bigg], \quad (3)$$

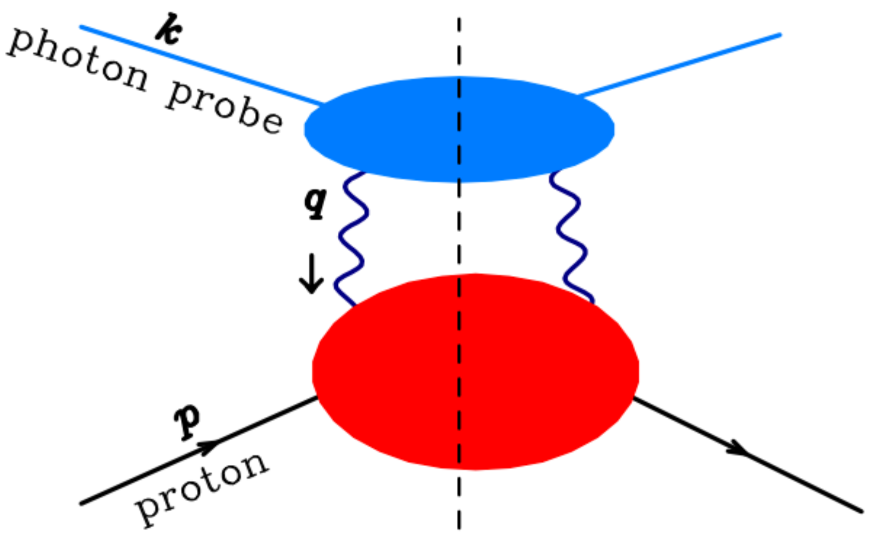
Manohar et al 1607.04266

<u>Theory-driven knowledge</u>

Photon PDFs

- QED is perturbative down to low scales ⟹ The photon must be computable is the input mark substructure is known
- Manohar et al: write the cross section for a chosen BSM process, e.g. production of heavy supersymmetric lepton L in ep collision (Drees, Zeppenfeld 1989)
- Equate the two expressions and find analytically the PDF of the photon

⇒ PDFs expressed in terms of the structure functions integrated over all scales, including elastic form factors (in the x →1 region)

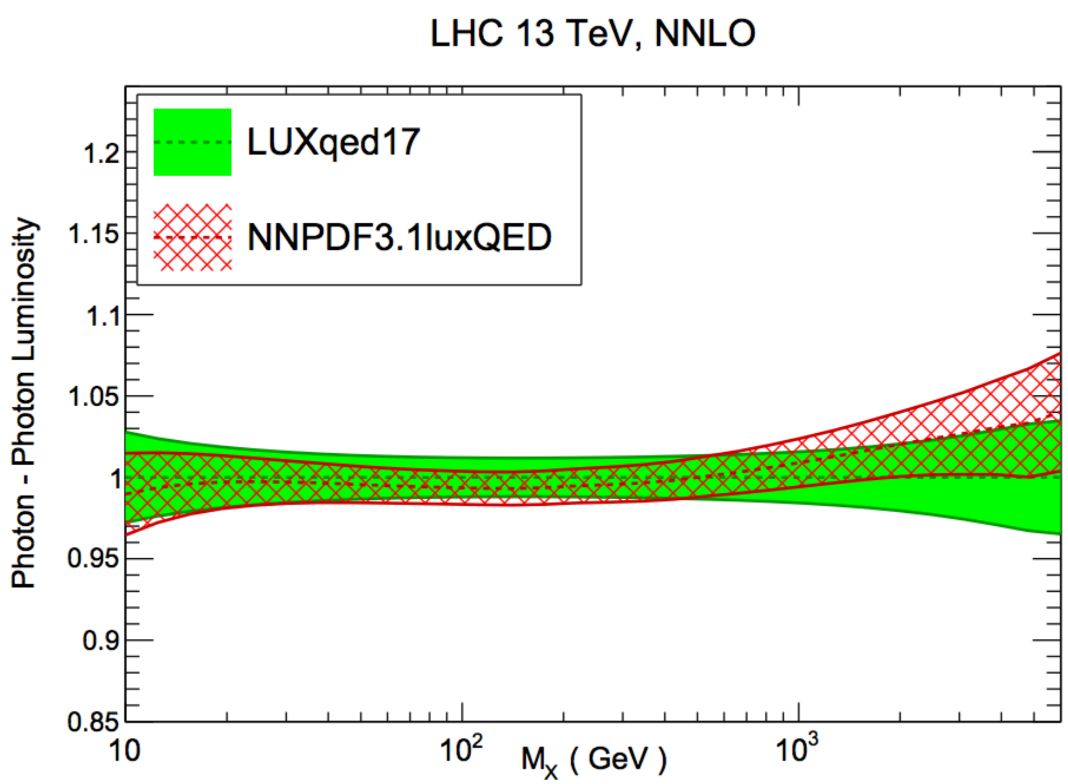


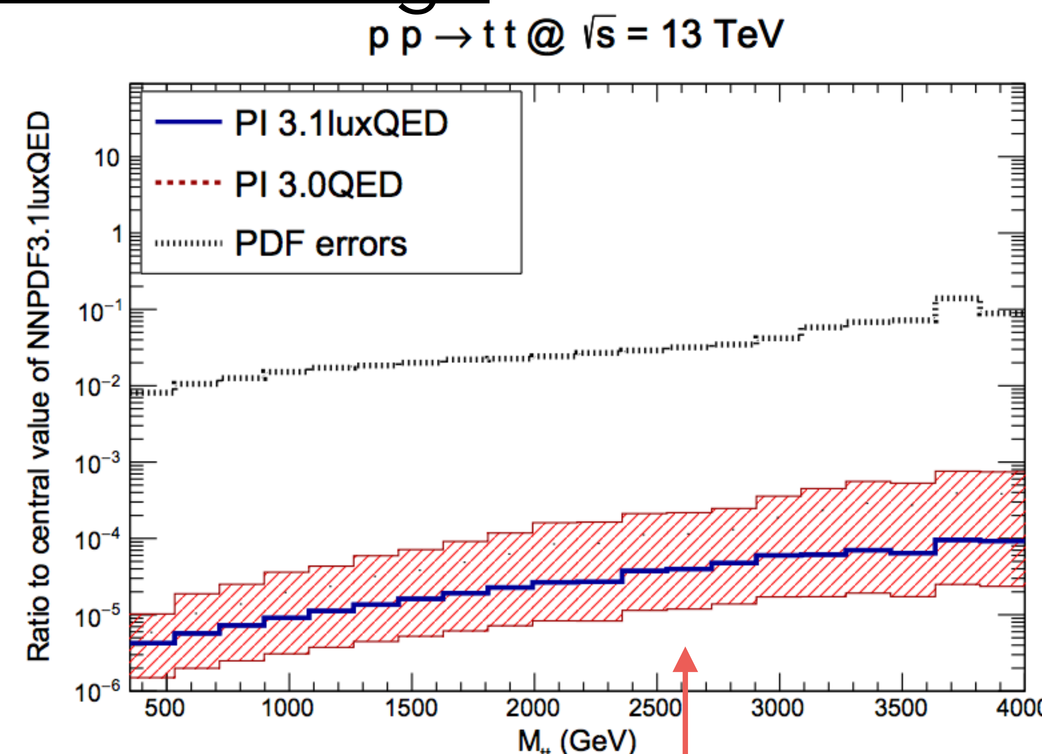
$$x f_{\gamma/p}(x,\mu^2) = \frac{1}{2\pi\alpha(\mu^2)}\int_x^1 \frac{dz}{z}\left\{ \int_{\frac{x^2 m_p^2}{1-z}}^{\frac{\mu^2}{1-z}} \frac{dQ^2}{Q^2}\alpha^2(Q^2) \left[\left(zp_{\gamma q}(z)+\frac{2x^2m_p^2}{Q^2}\right)F_2(x/z,Q^2) - z^2F_L\left(\frac{x}{z},Q^2\right)\right] - \alpha^2(\mu^2)z^2F_2\left(\frac{x}{z},\mu^2\right)\right\},$$

<u>Theory-driven knowledge</u>

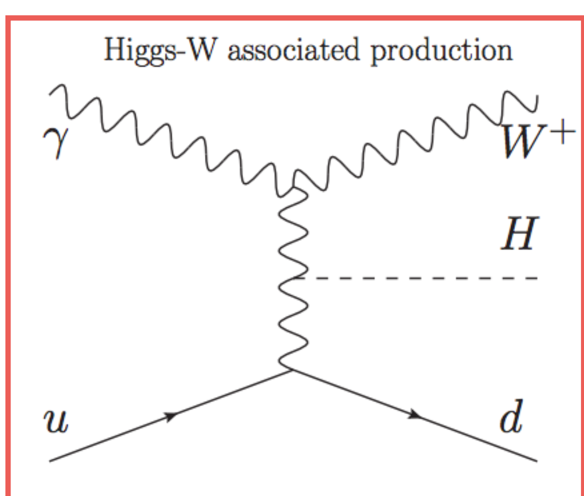
Photon PDFs

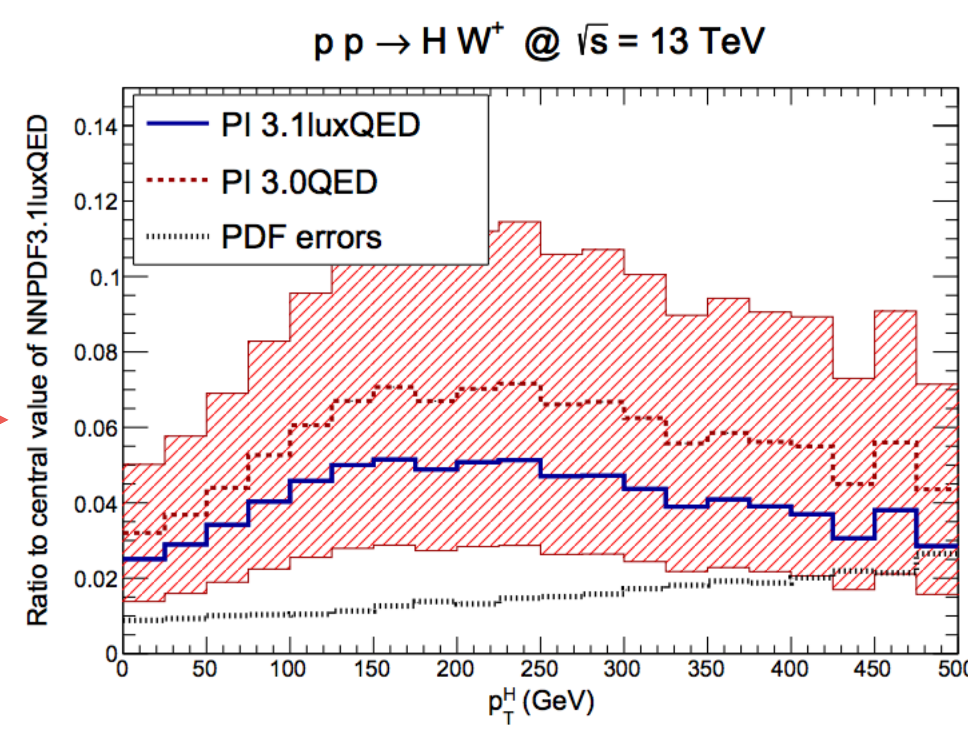
(Data+Theory)-driven knowledge

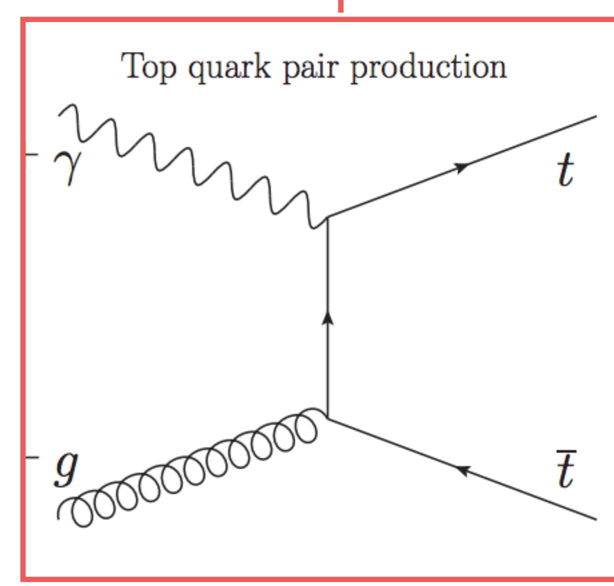




Bertone et al, 1712.07053

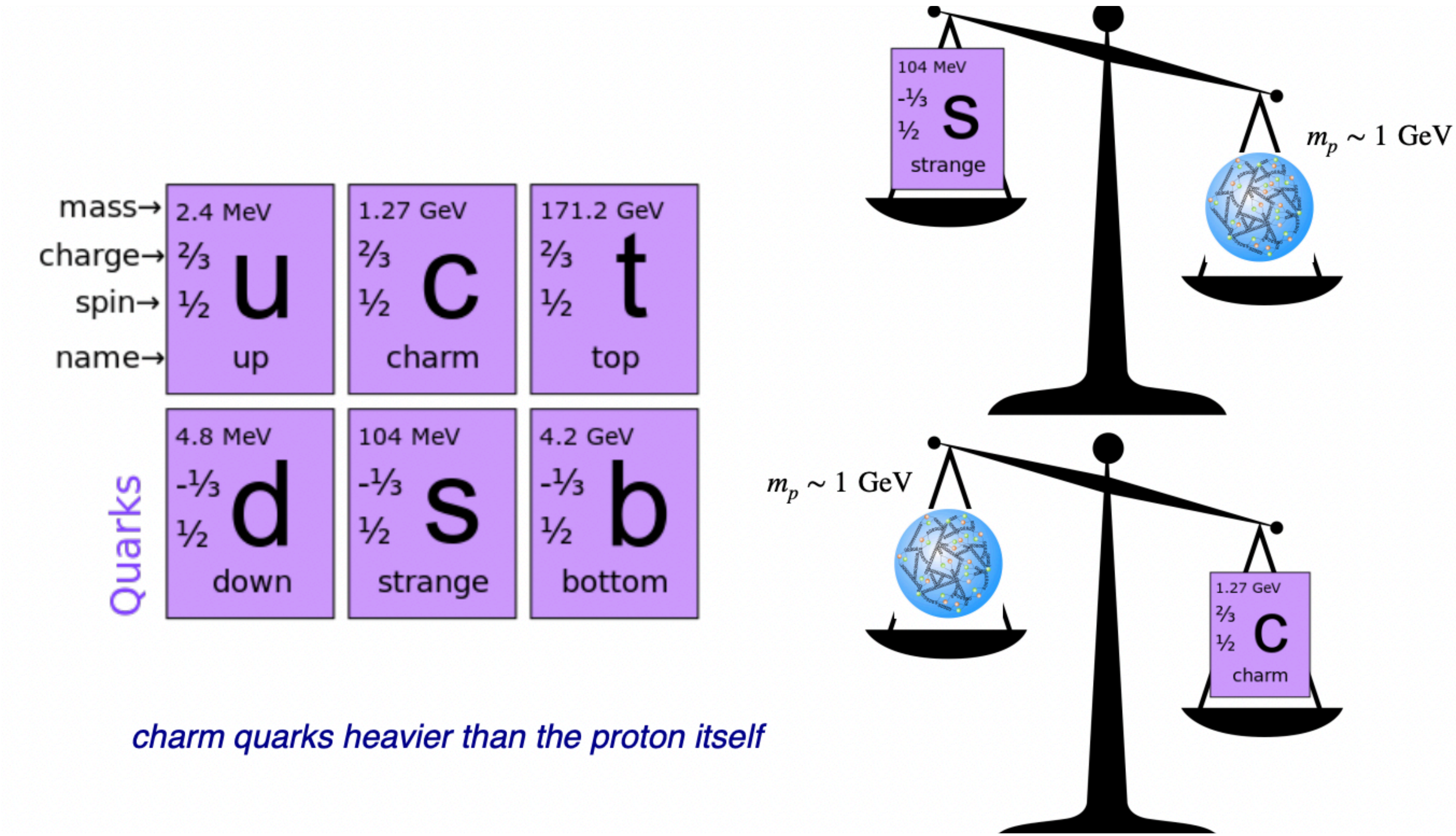






The charm of the proton

✓ Common assumption in PDF fits: the static proton wavefunction does not contain charm quarks: the proton contains intrinsic up, down, strange quarks and anti quarks but no intrinsic charm quarks



From J. Rojo's talk at KIT

The charm of the proton

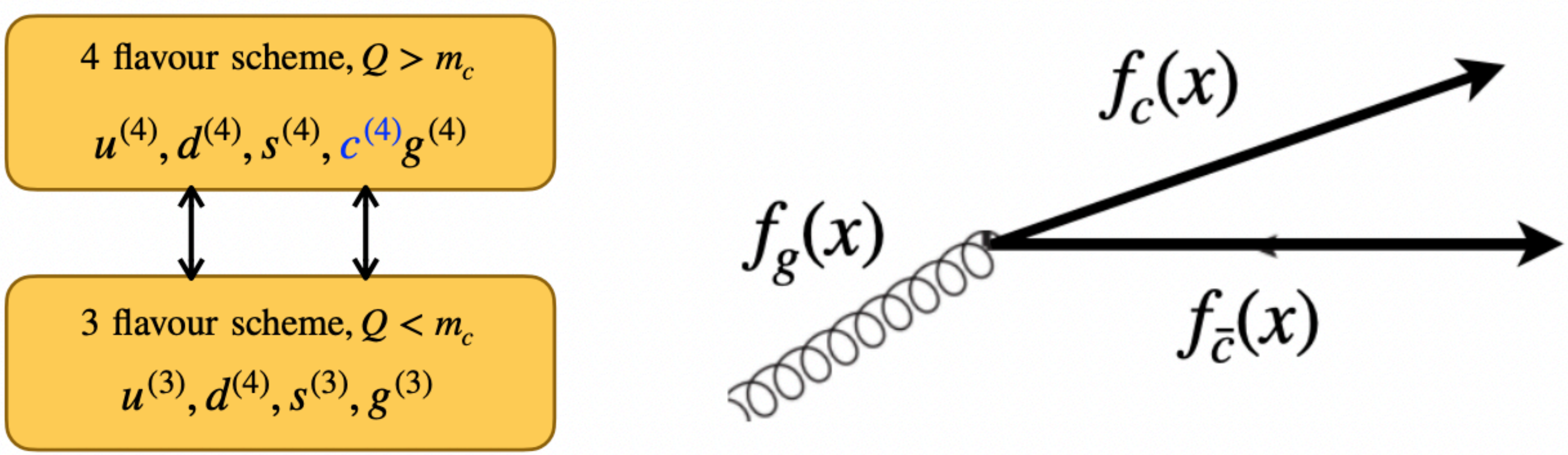
✓ Common assumption in PDF fits: the static proton wavefunction does not contain charm quarks: the proton contains intrinsic up, down, strange quarks and anti quarks but no intrinsic charm quarks

✓The charm PDF is generated perturbatively (DGLAP evolution from radiation off gluons and quarks)

$$f_c^{(n_f)} = 0 \quad \rightarrow \quad f_c^{(n_f+1)} \propto \alpha_s \ln \frac{Q^2}{m_c^2} \left(P_{qg} \otimes f_g^{(n_f+1)} \right) + \mathcal{O}\left(\alpha_s^2\right)$$

NLO matching

3FNS charm — 4FNS charm — 4FNS gluon



If charm is **perturbatively generated**, the charm PDF is **trivial**

From J. Rojo's talk at KIT

The charm of the proton

- ✓ Common assumption in PDF fits: the static proton wavefunction does not contain charm quarks: the proton contains intrinsic up, down, strange quarks and anti quarks but no intrinsic charm quarks
- ✓ It does not need to be so, as an intrinsic charm component is predicted in many models.

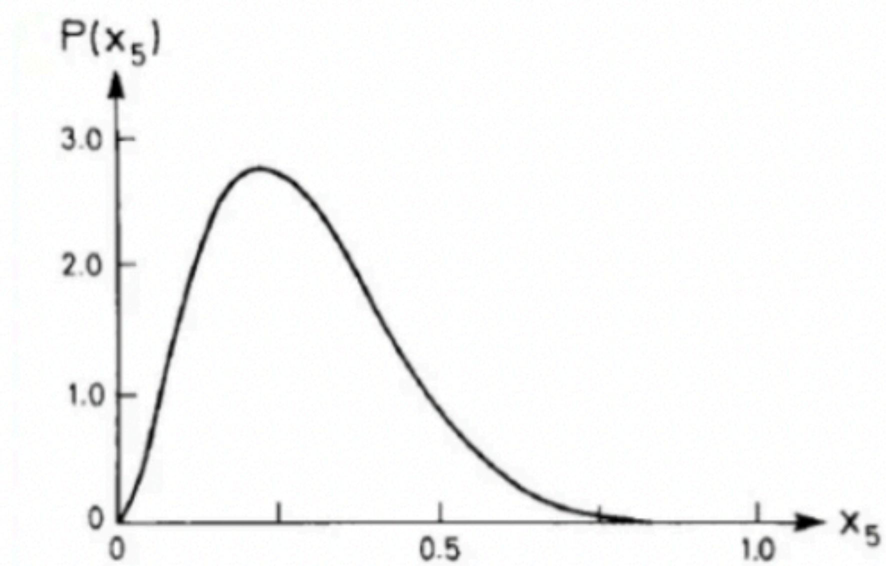
THE INTRINSIC CHARM OF THE PROTON

S.J. BRODSKY [1]
Stanford Linear Accelerator Center, Stanford, California 94305, USA

and

P. HOYER, C. PETERSON and N. SAKAI [2]
NORDITA, Copenhagen, Denmark

Received 22 April 1980



$$|p\rangle = \mathscr{P}_{3q}|uud\rangle + \mathscr{P}_{5q}|uudc\bar{c}\rangle + \ldots$$

Recent data give unexpectedly large cross-sections for charmed particle production at high x_F in hadron collisions. This may imply that the proton has a non-negligible $uudc\bar{c}$ Fock component. The interesting consequences of such a hypothesis are explored.

The charm of the proton

- Common assumption in PDF fits: the static proton wavefunction does not contain charm quarks: the proton contains intrinsic up, down, strange quarks and anti quarks but no intrinsic charm quarks
- It does not need to be so, as an intrinsic charm component is predicted in many models.

in this scenario, the charm PDF extracted from data in the global fit is the combination of the **perturbative** (DGLAP) and the **intrinsic** components

$$c^{(n_f=4)}(x,Q) \simeq c^{(n_f=4)}_{(\mathrm{pert})}(x,Q) + c^{(n_f=4)}_{(\mathrm{intr})}(x,Q)$$

Extracted phenomenologically from data

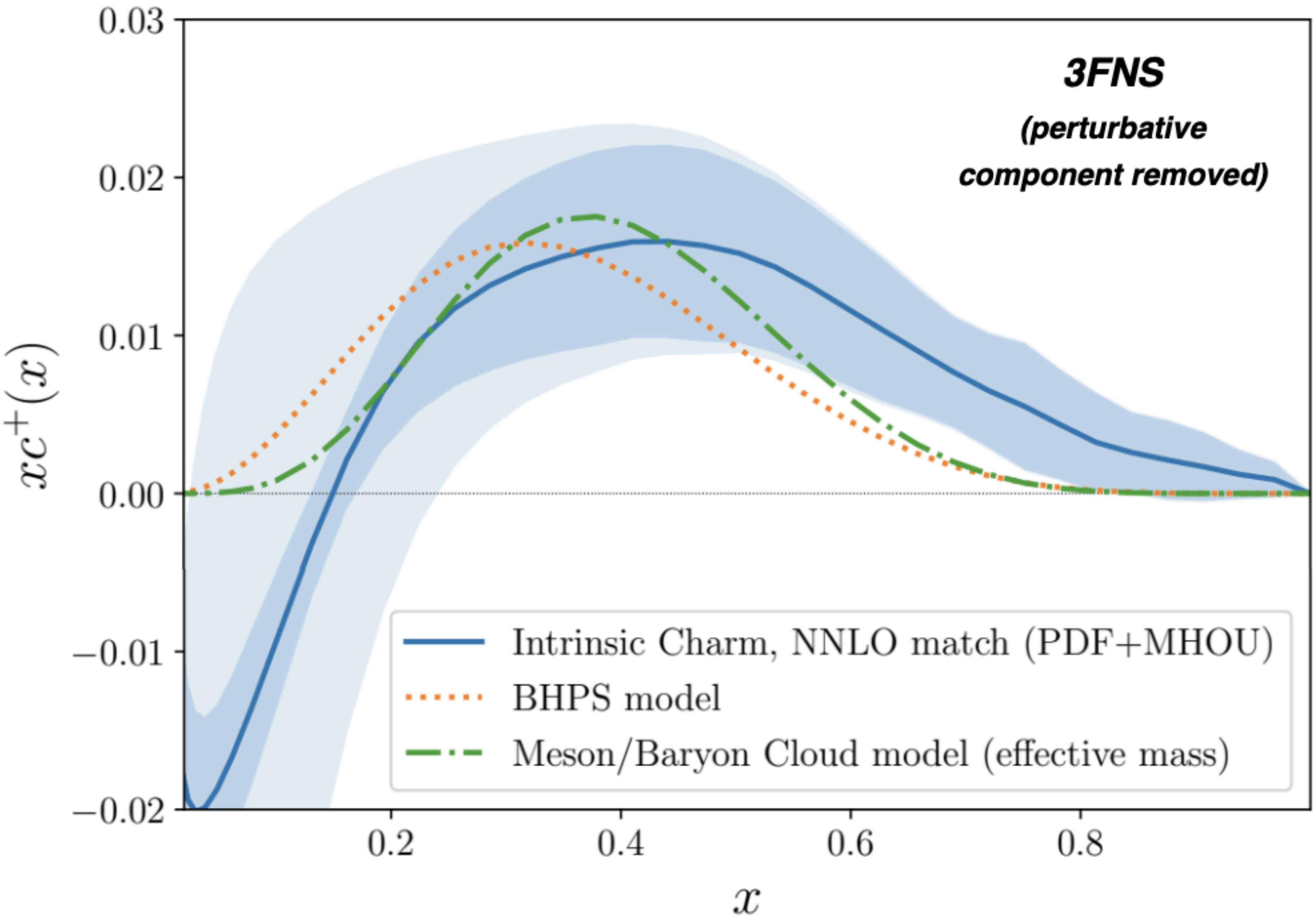
from QCD evolution and matching

from intrinsic component $c^{(n_f=3)}_{(\mathrm{intr})}(x) \neq 0$

How to **disentangle perturbative** from **intrinsic components?**

*nb we **define IC** as the charm PDF once know perturbative component is removed*

The charm of the proton



R. D. Ball, A. Candido, J. Cruz-Martinez, S. Forte, T. Giani, F. Hekhorn, K. Kudashkin, G. Magni & J. Rojo, Nature 608 (2022)

Covariance matrix

- Theory is perturbative expansion to some order : $$t_p = \sum_{m=0}^{p} c_m$$
- Standard case: $$P(d|t_p) \propto \exp\Big(-\frac{1}{2}(d-t_p)^T \mathrm{cov}_{\mathrm{exp}}^{-1}(d-t_p)\Big)$$ χ^2_{exp}
- Bayes' theorem: $$P(t_p|d) = \frac{P(d|t_p)P(t_p)}{P(d)} \propto P(d|t_p)P(t_p)$$
- Assume Gaussian theory prior:

$$P(t_p) = \prod_{m=0}^{p} P(c_m) \quad \text{where} \quad P(c_m) \propto \exp\Big(-\frac{1}{2} c_m^T \mathrm{cov}_{\mathrm{th},m}^{-1} c_m \Big)$$ χ^2_{th}

- Assume MHOUs due to $\mathcal{O}(\alpha^{p+1})$ terms only → marginalise these terms:

$$P(t_p|d) \propto \int dc_{p+1} P(d|c_{p+1}) P(t_{p+1})$$

$$\propto \exp\Big(-\frac{1}{2}(d-t_p)^T (\mathrm{cov}_{\mathrm{exp}} + \mathrm{cov}_{\mathrm{th}})^{-1}(d-t_p)\Big)$$ χ^2_{tot}

- Include higher order terms by induction

Covariance matrix

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

➡ How to build correlations between different points?

$$(\mathrm{cov}_{\mathrm{th}})_{mn} = \langle (t_p(\mu_R, \mu_F) - t_p(\mu_R^0, \mu_F^0))_m (t_p(\mu_R, \mu_F) - t_p(\mu_R^0, \mu_F^0))_n \rangle$$

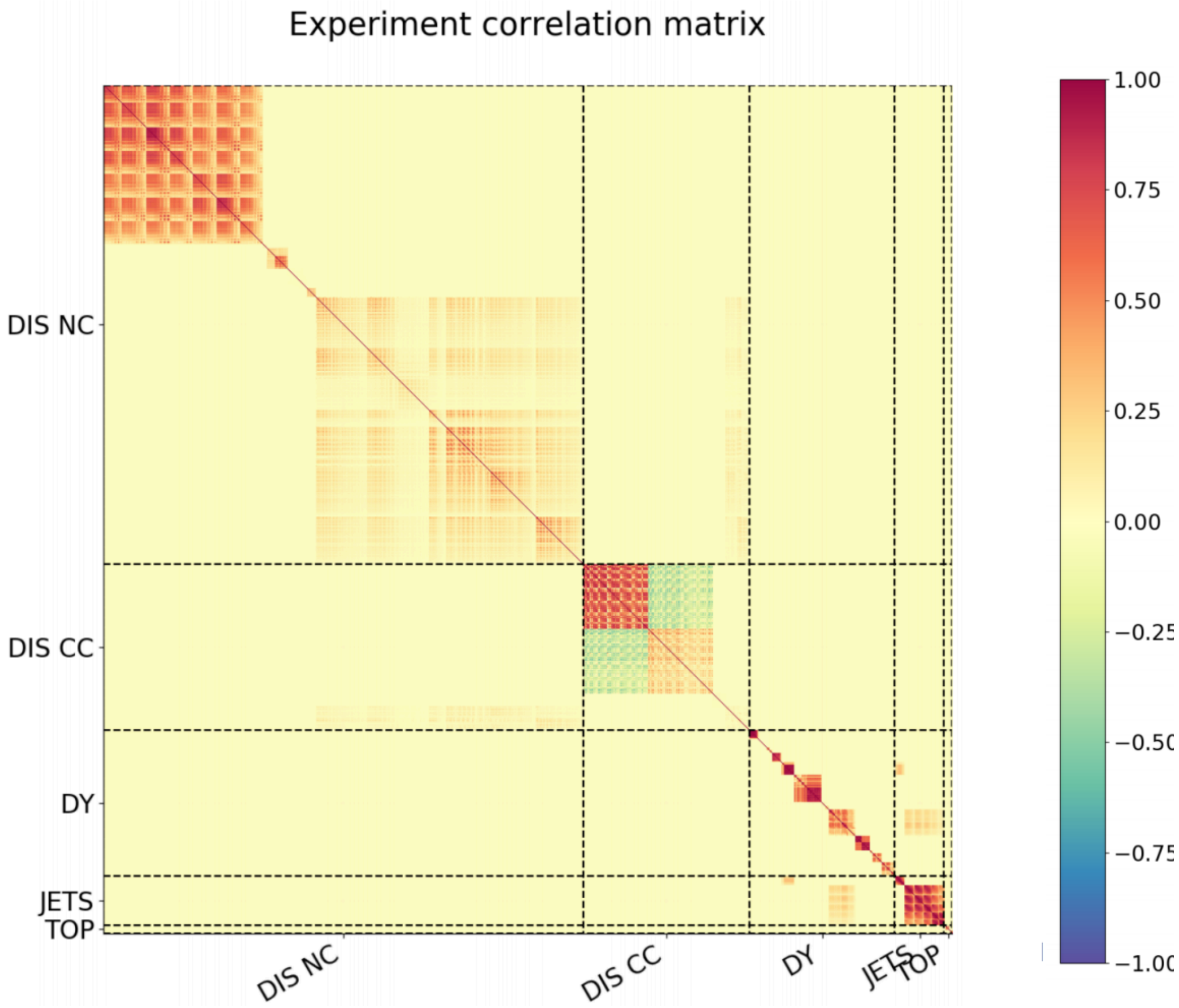
- $\boldsymbol{\mu}_F$ variations correlated across all processes by PDF evolution
- $\boldsymbol{\mu}_R$ variation correlated by process (hard cross section)

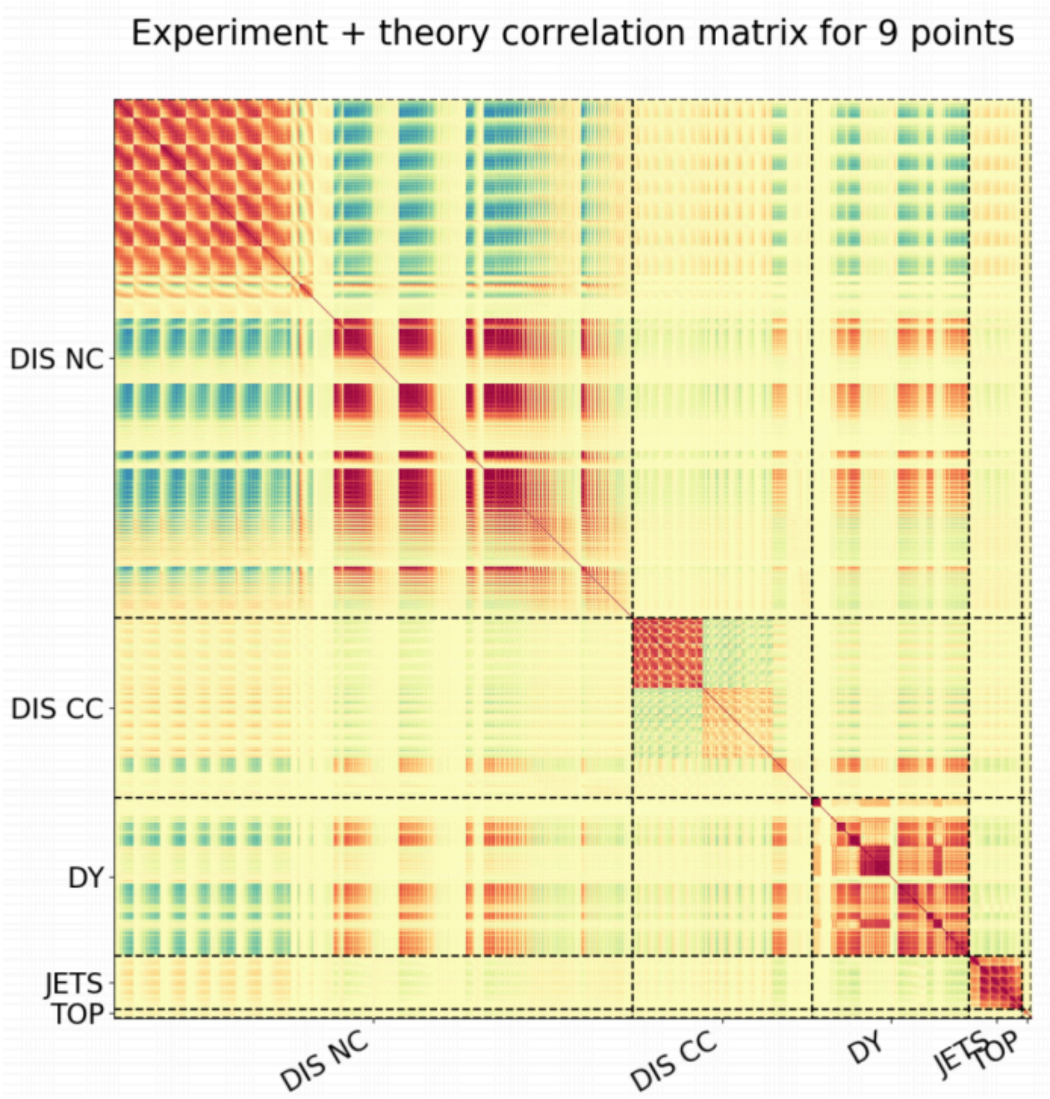
➡ Several recipes possible (3-points prescriptions, 7-points…)

➡ Details of correlations are also important

➡ A lot to be investigated

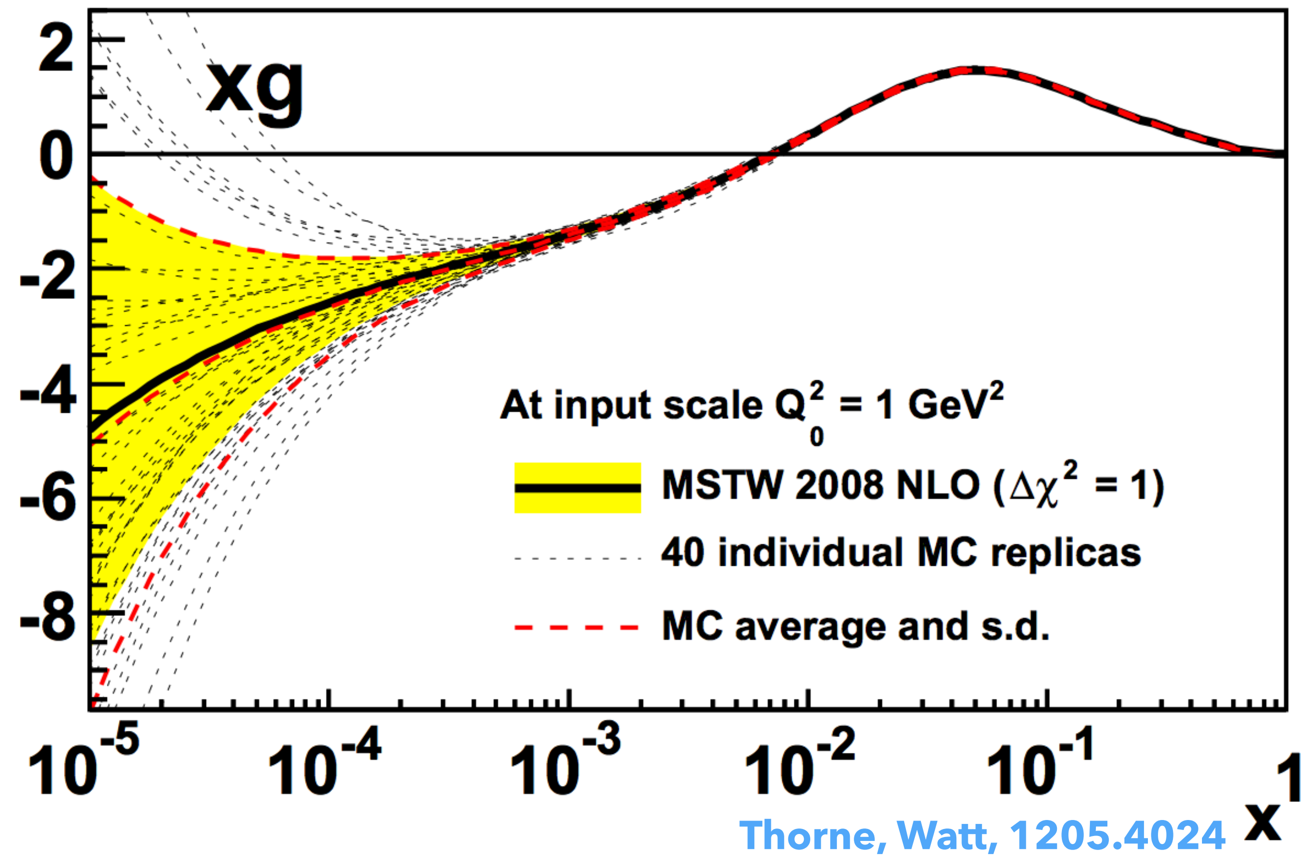
Covariance matrix



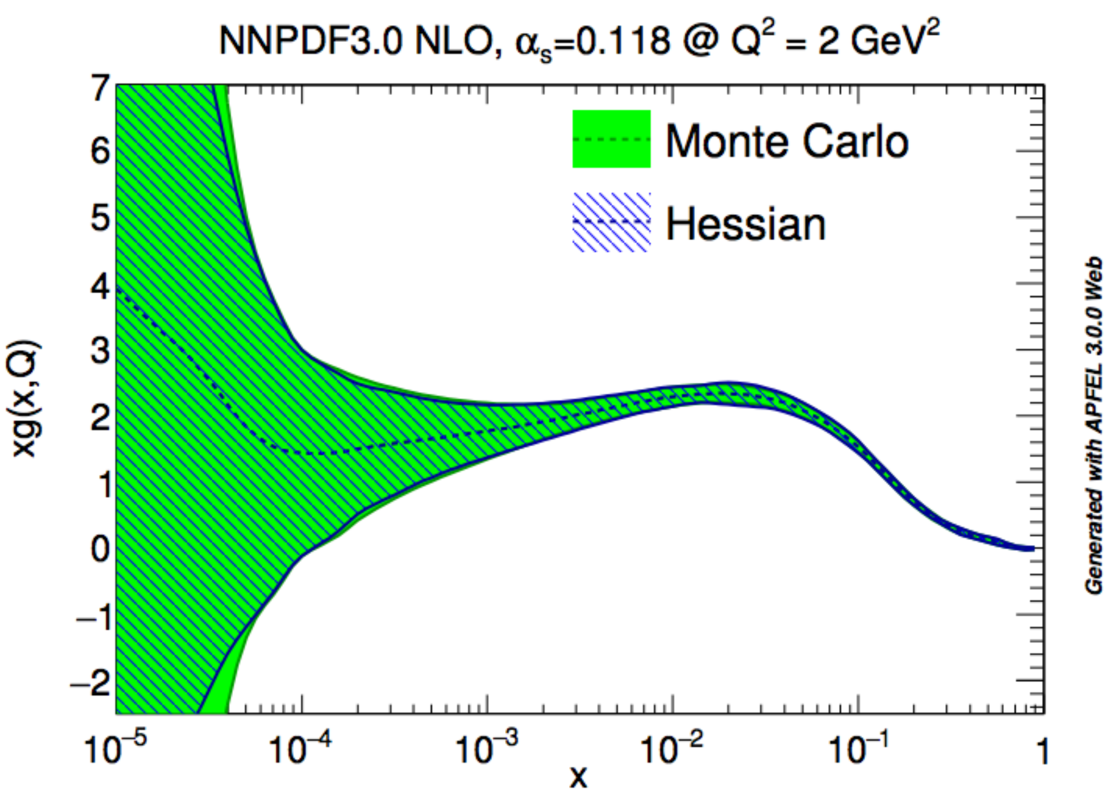


Hessian ⟺ Monte Carlo

- To convert Hessian into Monte Carlo, generate multi-gaussian replicas in the fitted parameters space
- Accurate when the number of replicas similar to that that reproduces the data



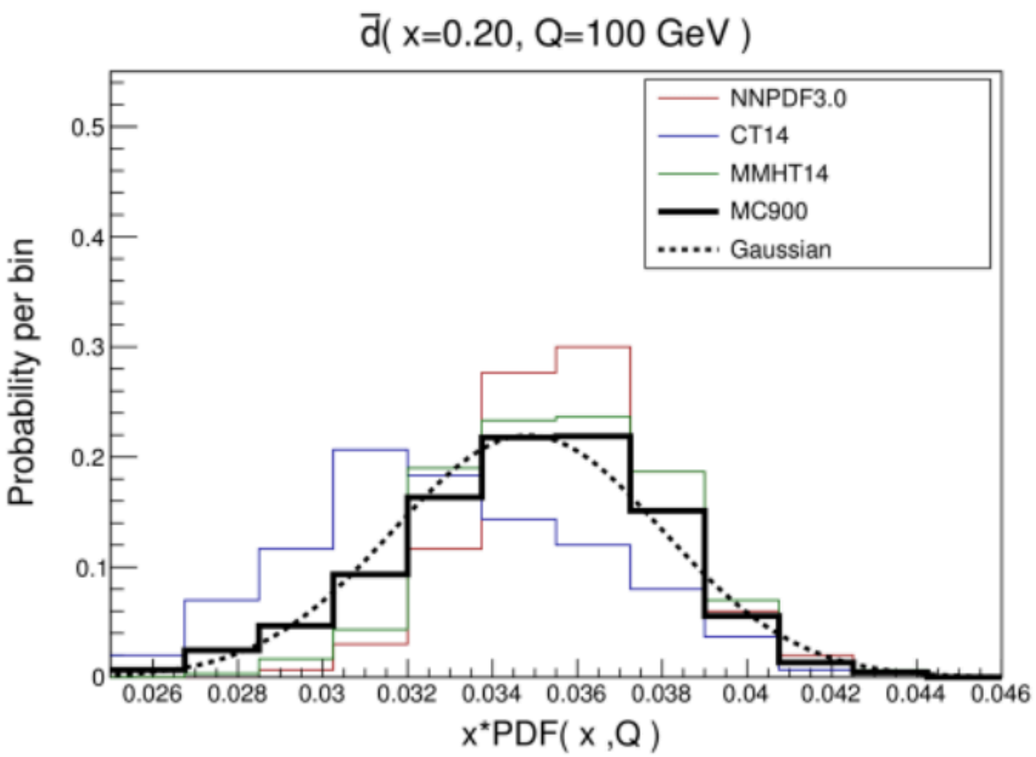
Thorne, Watt, 1205.4024



Carrazza et al 1505.06736

- To convert Monte Carlo into Hessian, sample the replicas f(x) at discrete set of points and construct the ensuing covariance matrix
- Eigenvectors of the covariance matrix as a basis in the vector space spanned by the replicas by the singular-value decomposition
- Number of dominant eigenvectors similar to numbers of replicas for accurate representation

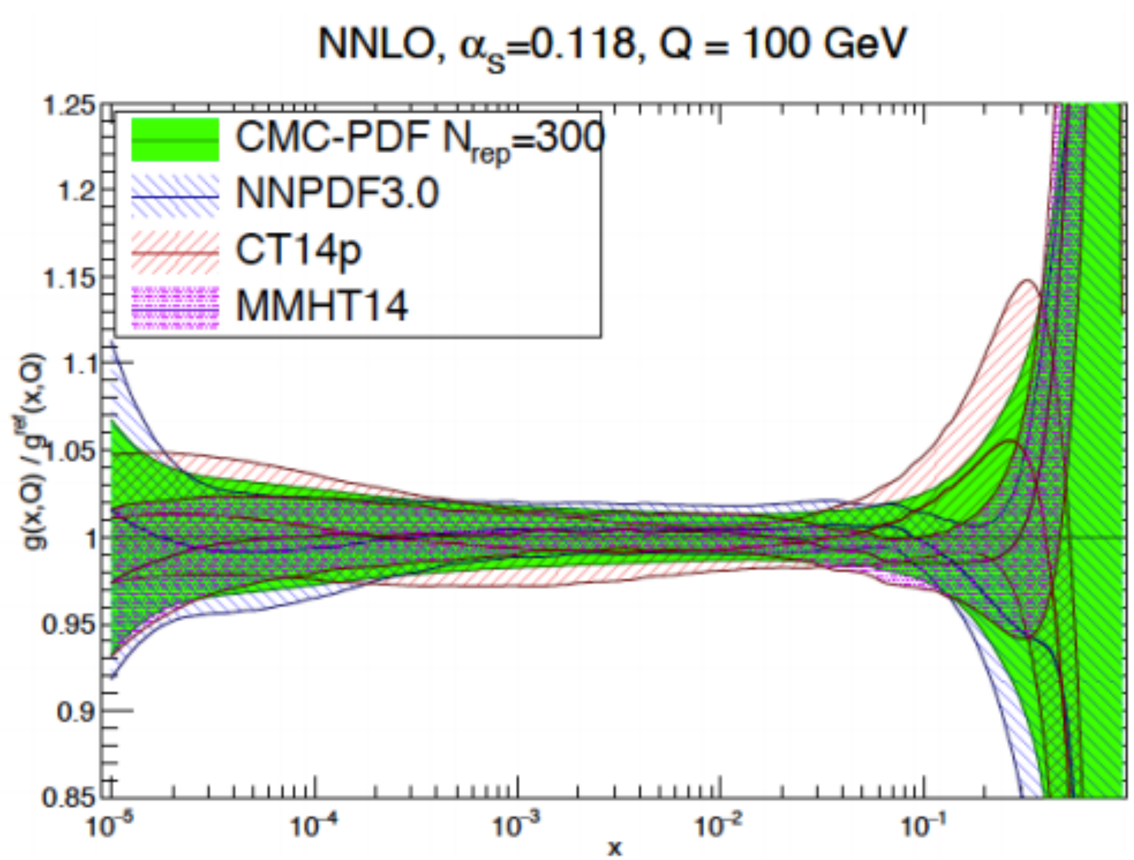
Hessian ⟺ Monte Carlo



- Using Monte Carlo conversion of Hessian sets, can combine different PDF sets, combining MC replicas into a single set
- Useful for conservative estimate
- Combined set approximatively Gaussian

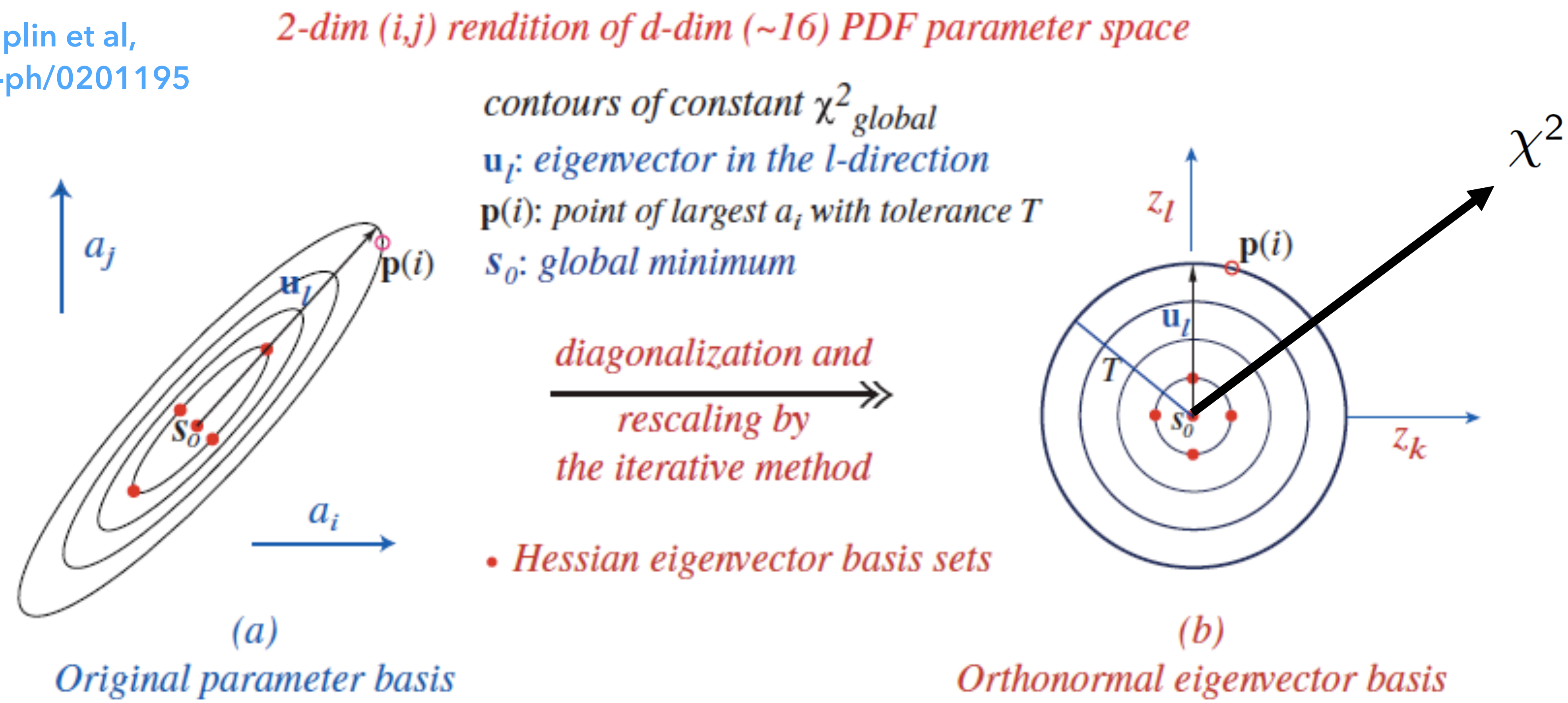
PDF4LHC15 recipe

- Monte Carlo combination of most recent global PDF sets **[Forte, Watt]**
- Each replica receives the same weight: uncertainty smaller than in the envelope, as in the latter outliers are given a larger weight
- New compression studies: N=40 replicas are virtually identical to the original 300 replicas from the point of view of correlation, standard deviation, observables **[Carrazza et al.]**



Hessian method

Pumplin et al, hep-ph/0201195



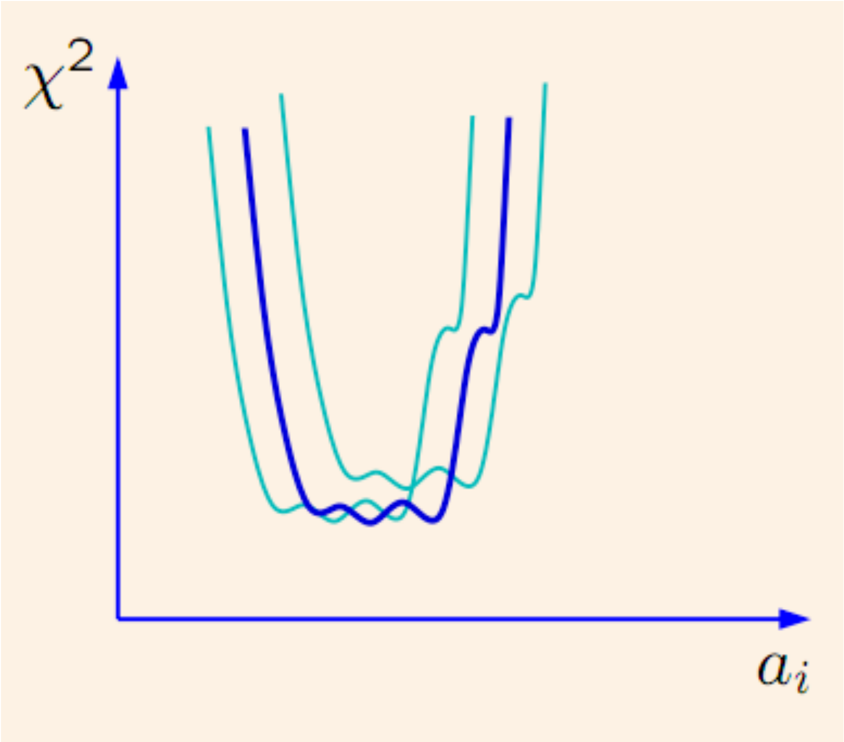
$$\sigma_X^2 = (H)_{ij}\,\partial_i X\,\partial_j X \quad \xrightarrow{\text{diagonalisation}} \quad \sigma_X^2 = |\vec{\nabla} X|^2$$

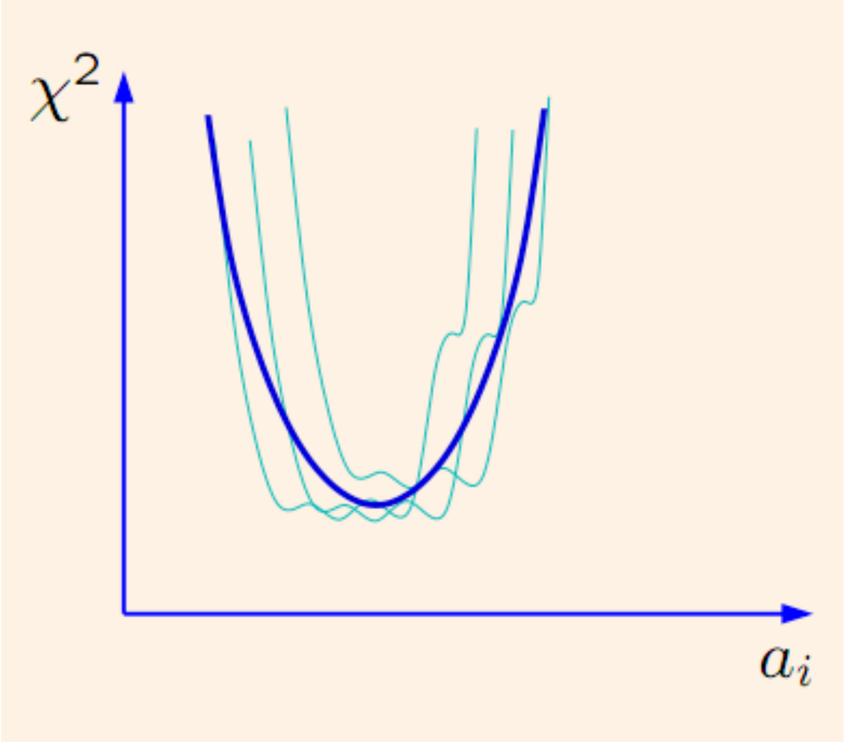
z_i eigenvectors of H with unit eigenvalues

➡ The total uncertainty is the sum in quadrature of the uncertainties due to each parameter

➡ $\Delta\chi^2 = \sum z_i^2$ the surfaces of constant χ^2 are spheres in the z space of radius $\sqrt{\Delta\chi^2}$

Hessian method





The actual χ^2 function displays

- A well pronounced global minimum χ_0^2
- Some tensions between datasets in the vicinity of the minimum
- Some dependence on assumptions about flat directions (= unconstrained combinations of PDF parameters)

The likelihood is approximately described by a quadratic χ^2 with a revised tolerance condition

$$\Delta\chi^2 \leq T^2$$

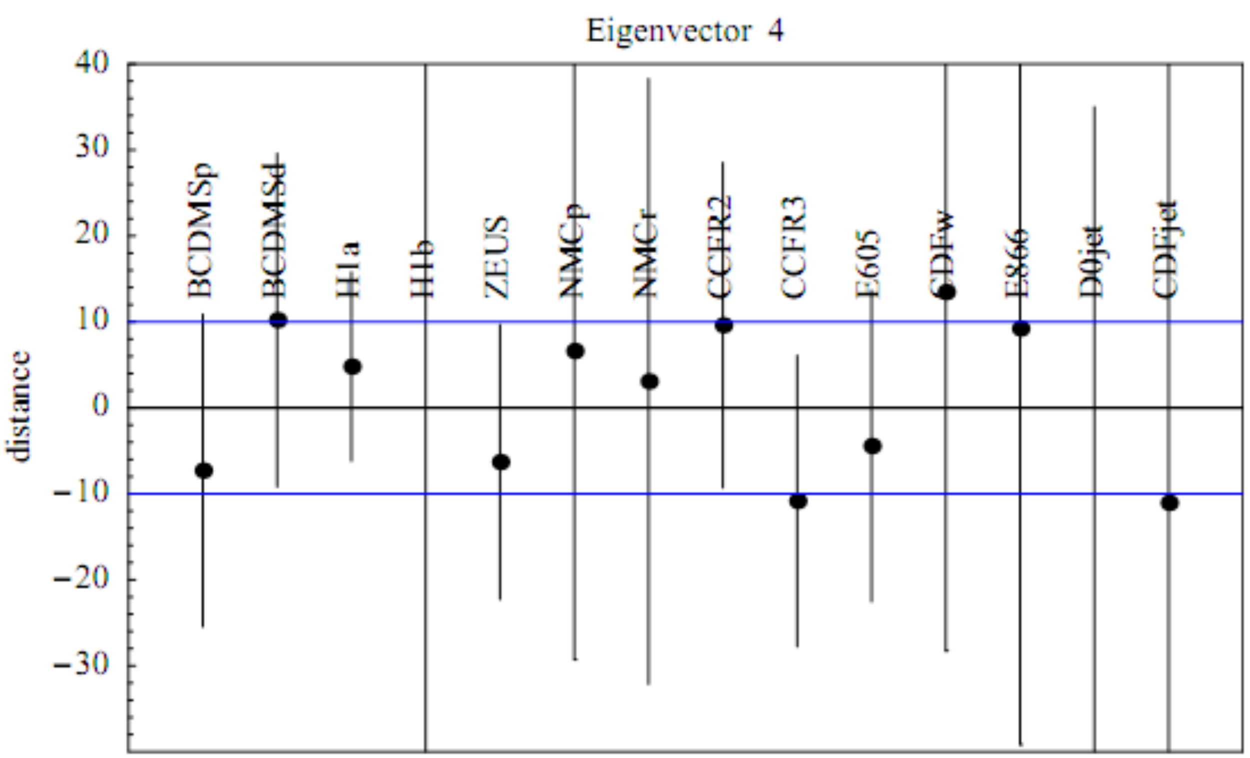
P. Nadolsky, CTEQ summer school 2009

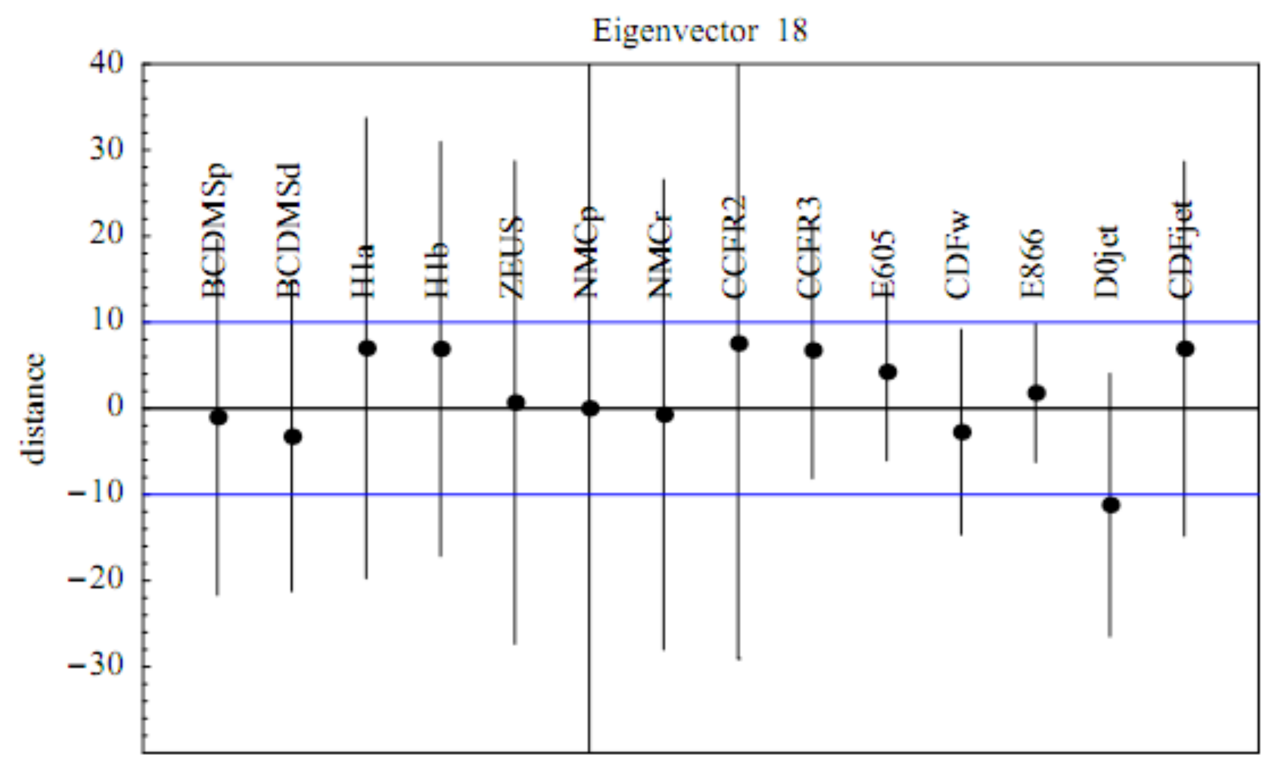
Hessian method

CTEQ6 tolerance criterion

- Acceptable values of PDF parameters must agree at ~ 90% C.L. with all experiments included in the fit, for a plausible range of assumptions about the PDF parametrisation, scale dependence, systematic uncertainty

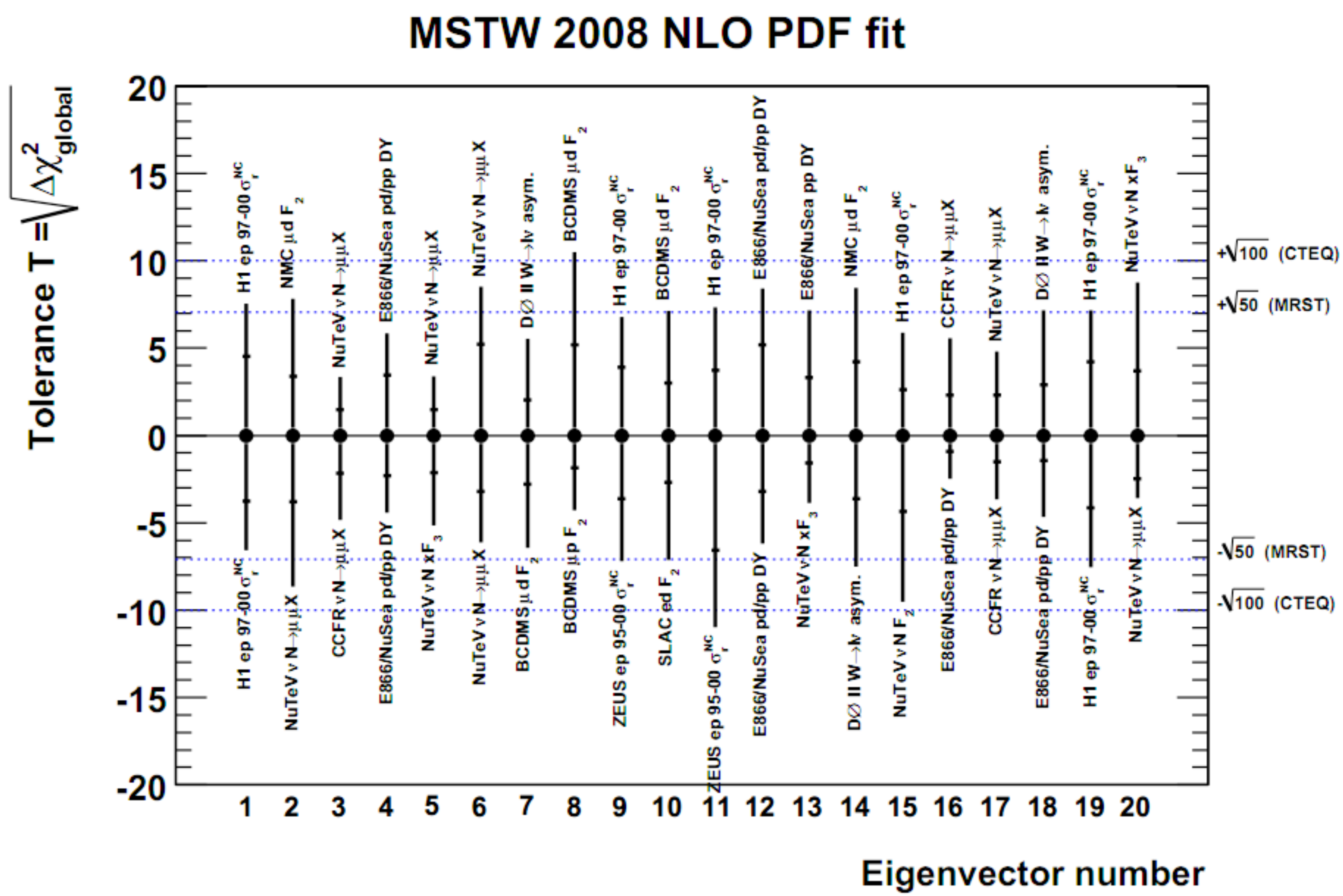
- Can be crudely approximated by assuming T ~ 10 for all PDF parameters





Hessian method

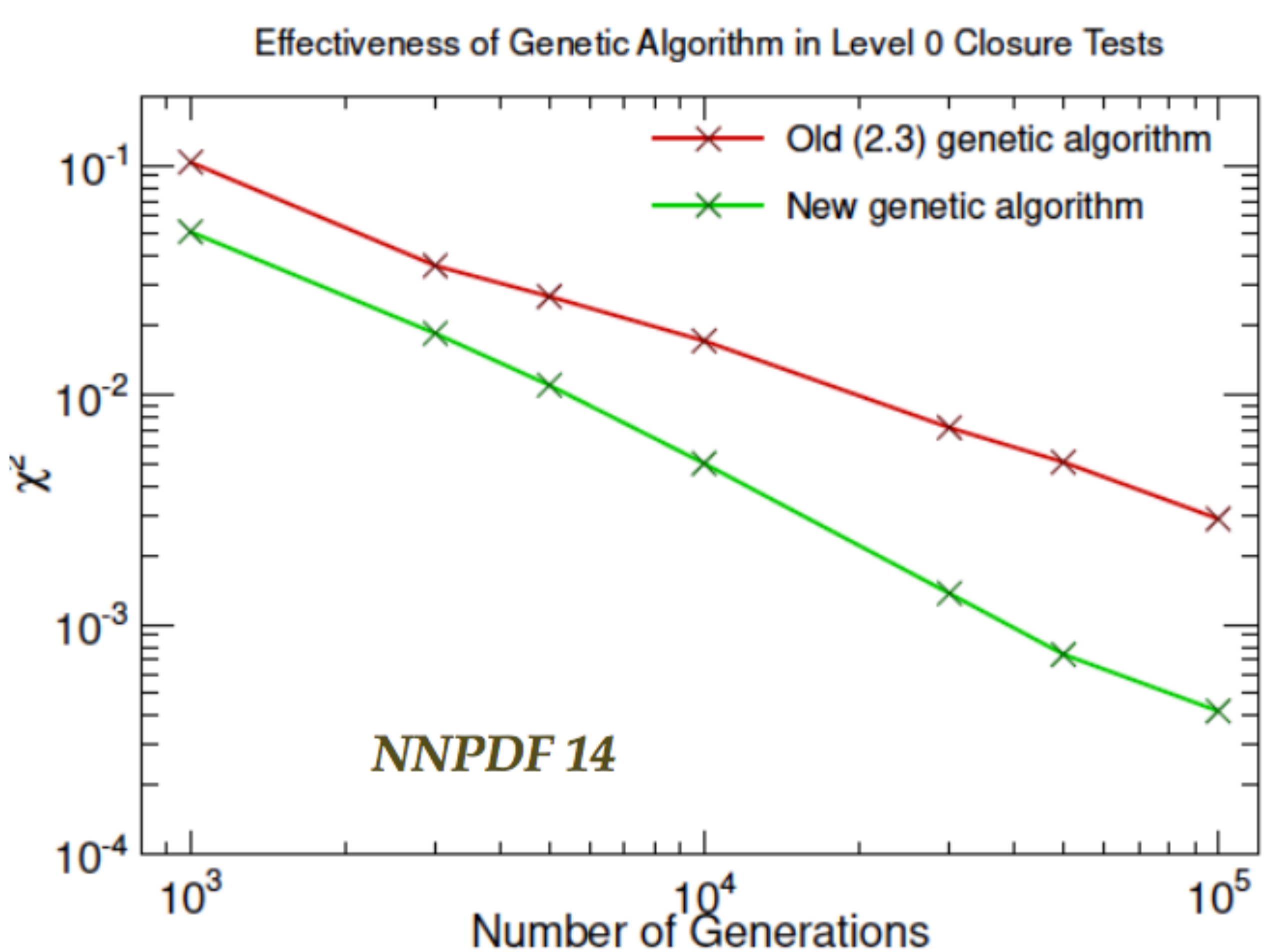
MSTW08 tolerance criterion

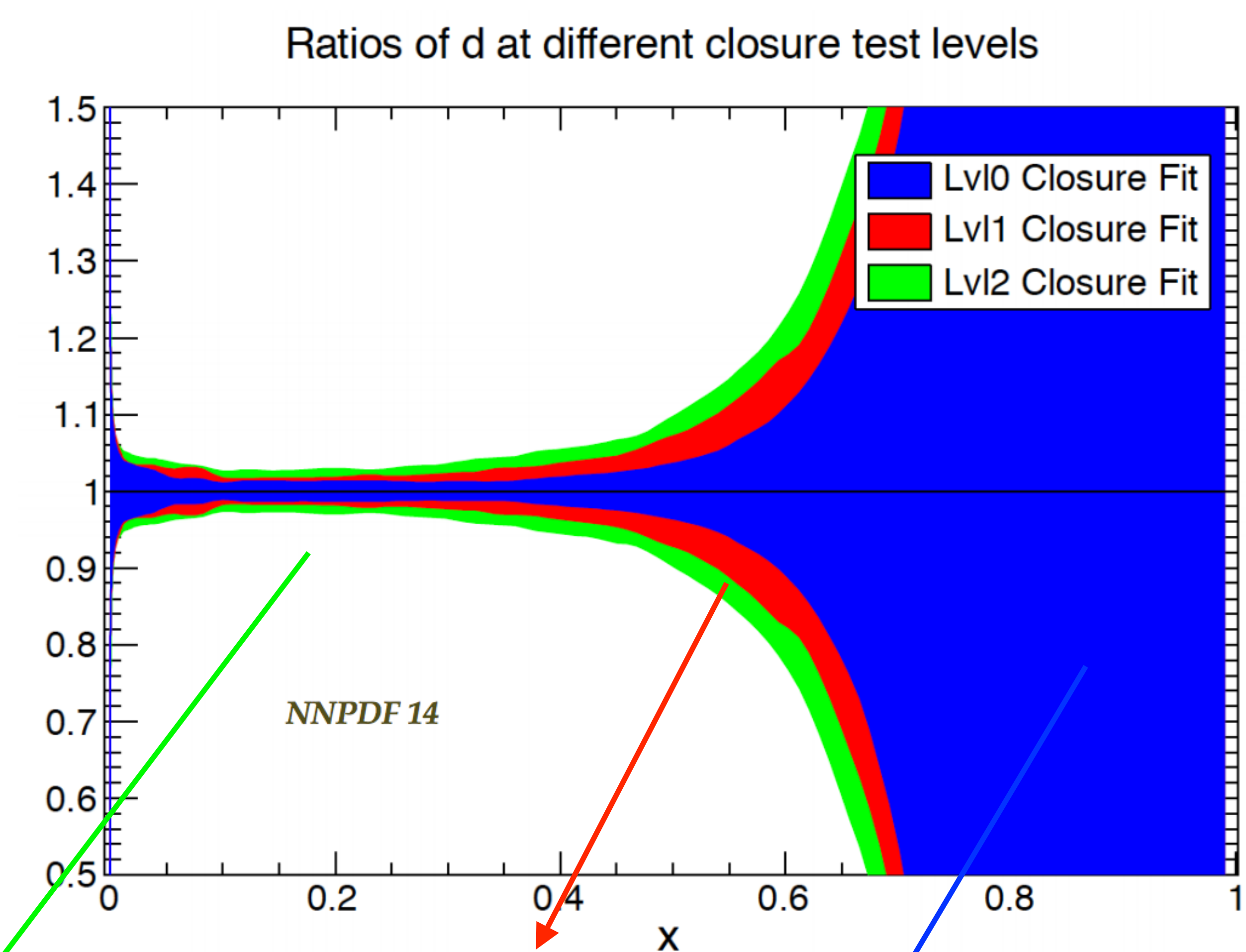


A dynamical tolerance, which varies according to the considered parameter

Statistical validation

- **Level-0:** if pseudo-data are identical to the input theory, then agreement with theory should be arbitrarily good, i.e. $\chi^2 \rightarrow 0$ but PDF uncertainty $\rightarrow 0$ only in the region where there are enough data
- **Level-1:** add uncertainty to pseudo data equal to actually experimental uncertainties: replicas fit same data over and over again, then $\chi^2 \rightarrow 1$ and test equivalent minima (parametrisation Δ)
- **Level-2:** generate Monte Carlo replicas of pseudo-data with fluctuations, then $\chi^2 \rightarrow 2$ (data Δ)





3 Δs comparable in data region

data uncertainty

parametrisation uncertainty

extrapolation uncertainty

Beyond fixed order

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

fixed order

$$\frac{\sigma}{\sigma_0} = 1 \quad \text{LO}$$
$$+ c_1\alpha \quad \text{NLO}$$
$$+ c_2\alpha^2 \quad \text{NNLO}$$
$$+ \ldots$$

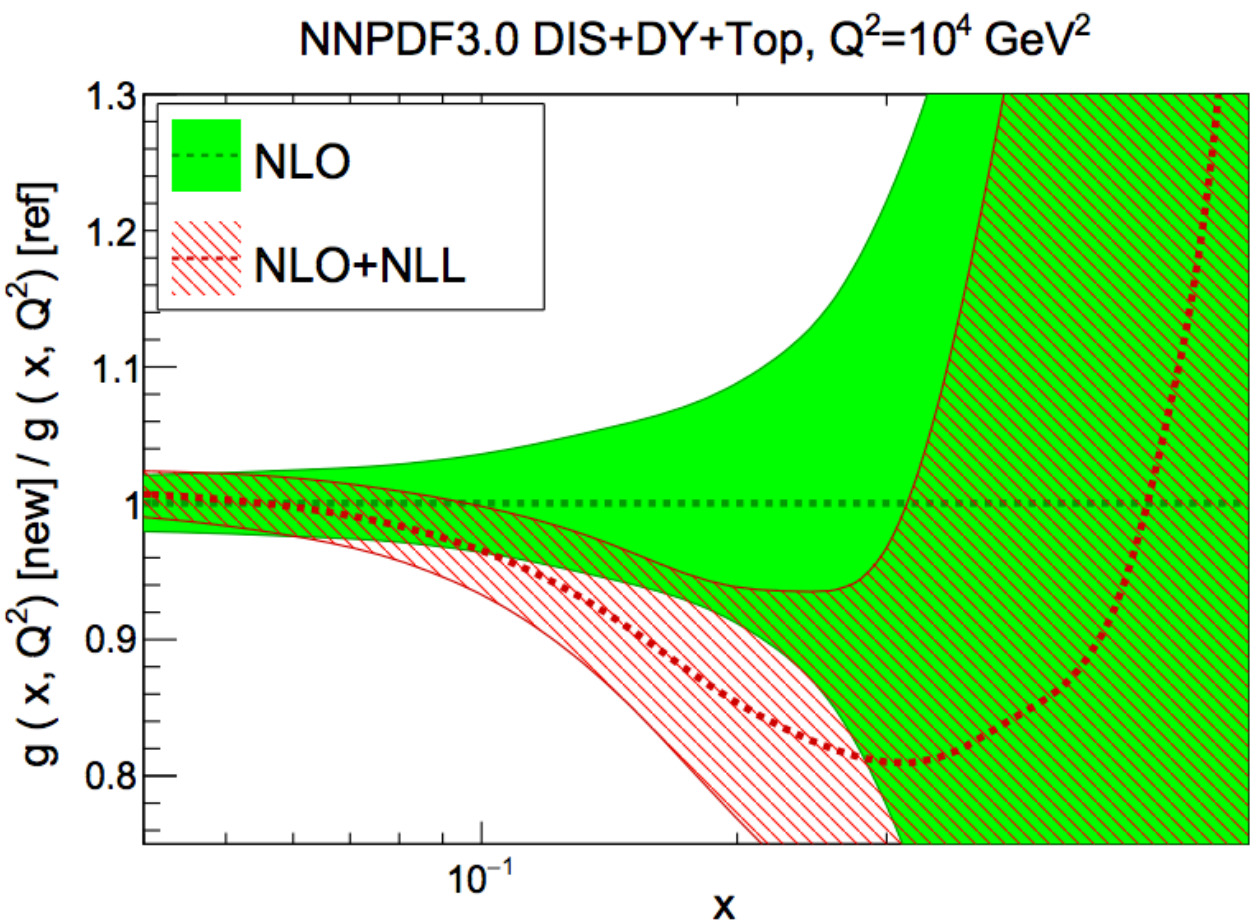
all order (L = some large logarithm)

$$\ln\frac{\sigma}{\sigma_0} = \alpha^n L^{n+1} \quad \text{LL}$$
$$+ \alpha^n L^n \quad \text{NLL}$$
$$+ \alpha^n L^{n-1} \quad \text{NNLL}$$
$$+ \ldots$$

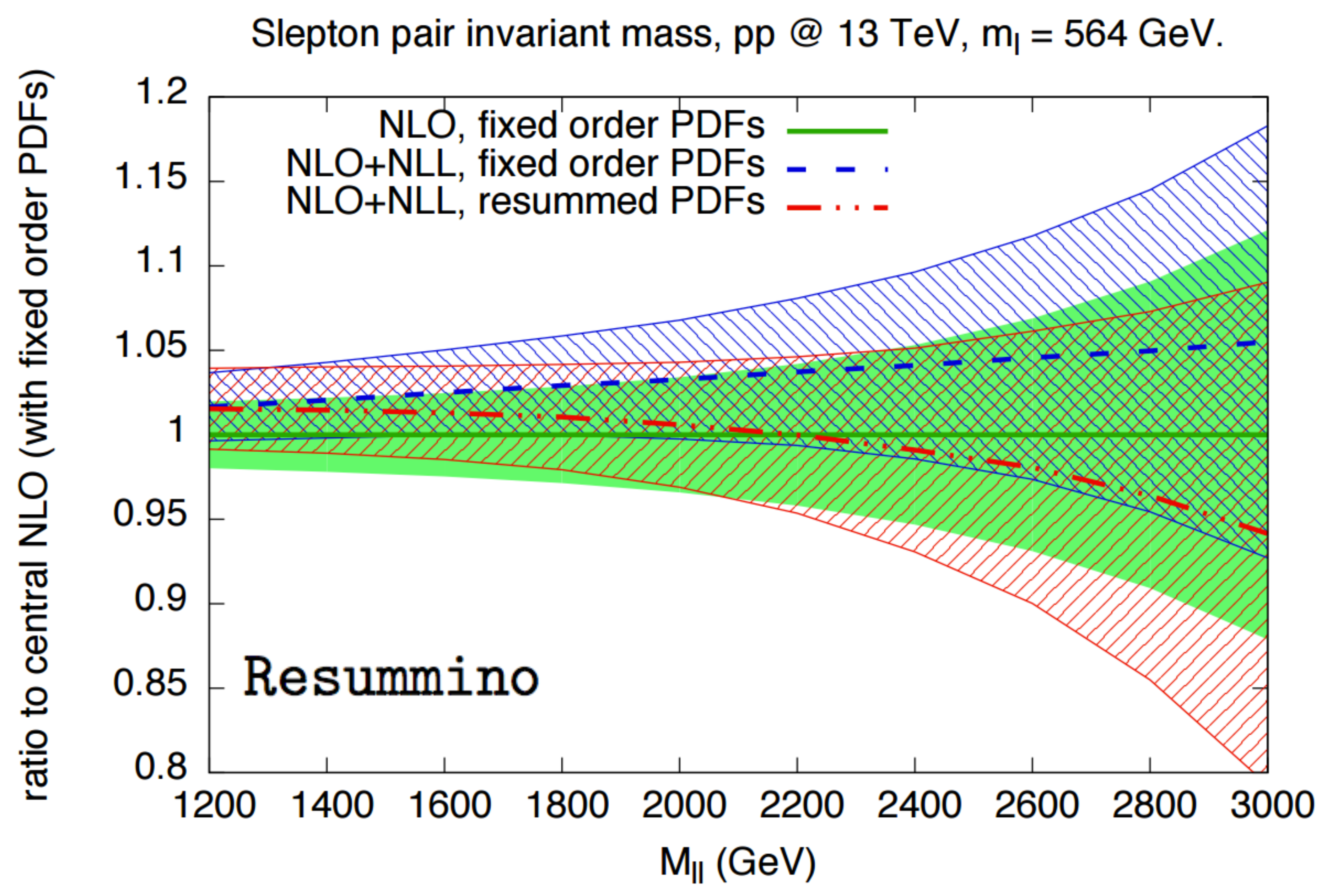
- Various kinds of logs:

L = log (1-x) threshold (soft-gluon) resummation
L = log (1/x) high-energy (small-x) resummation
L = log (pT/M) transverse momentum resummation

Threshold resummation



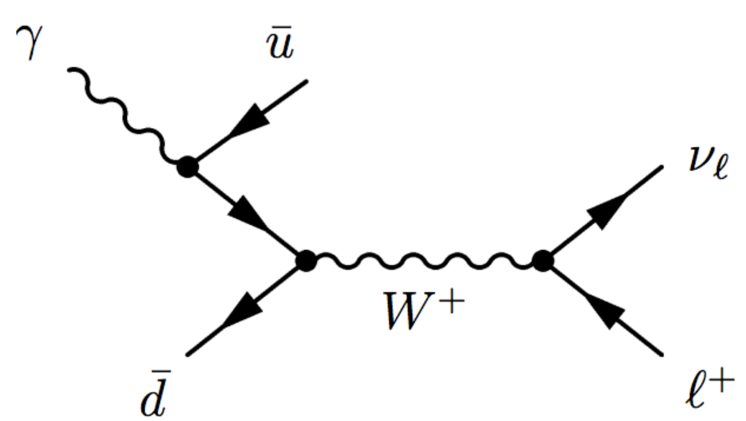
Bonvini et al, JHEP 1509 (2015) 191

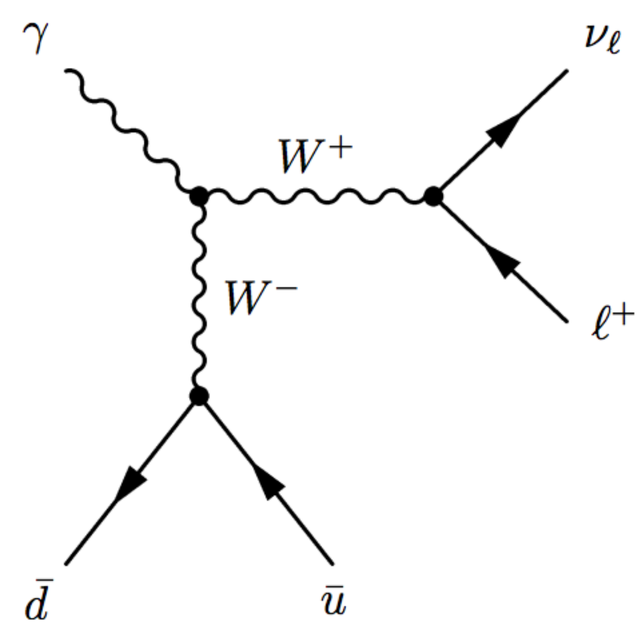


- In the MSbar scheme PDF evolution does not contain large-x logs and the effect of resummation can be included in resummed partonic cross sections
- Threshold-resummed PDFs will be suppressed as compared to fixed-order PDFs
- Mostly due to enhancement of NLO+NLL partonic xsecs used in the fit of DIS structure functions and DY distributions
- Phenomenologically relevant for new physics processes [Beenakker et al. EPJC76 (2016)2, 53]

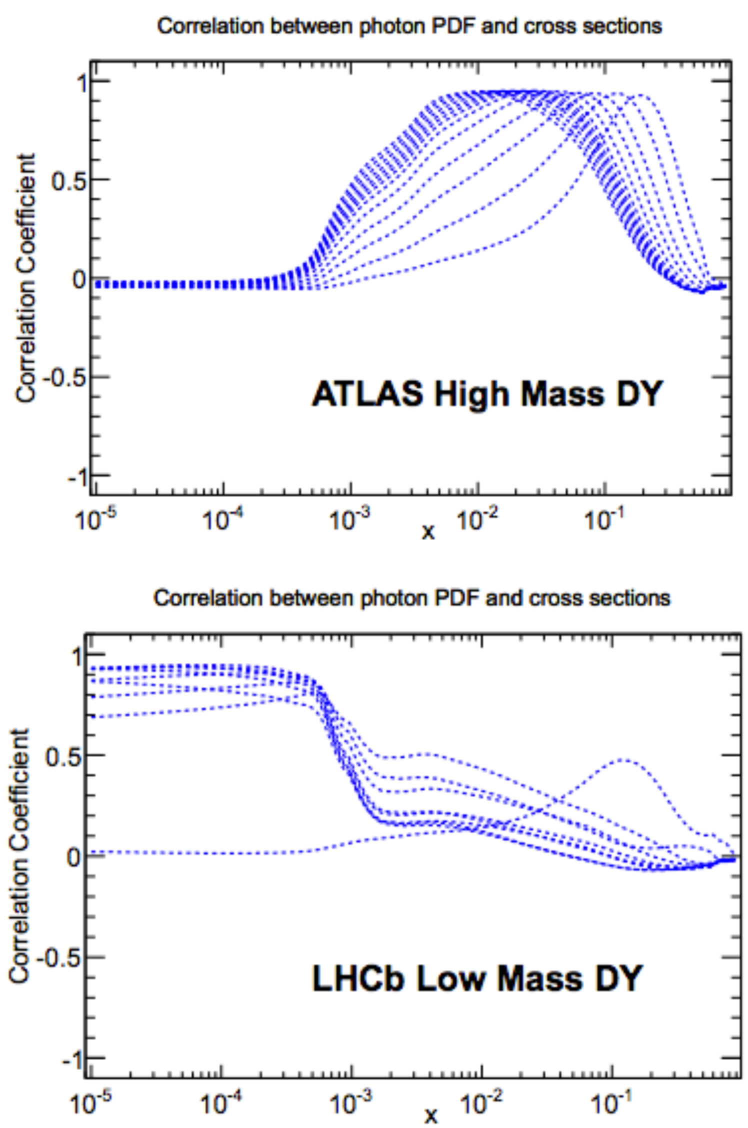
Photon PDFs

- Largest correlations between photon PDFs and pp cross sections are for Drell-Yan processes, but also for top pair production and VV production





$Q = 100\,\mathrm{GeV}$

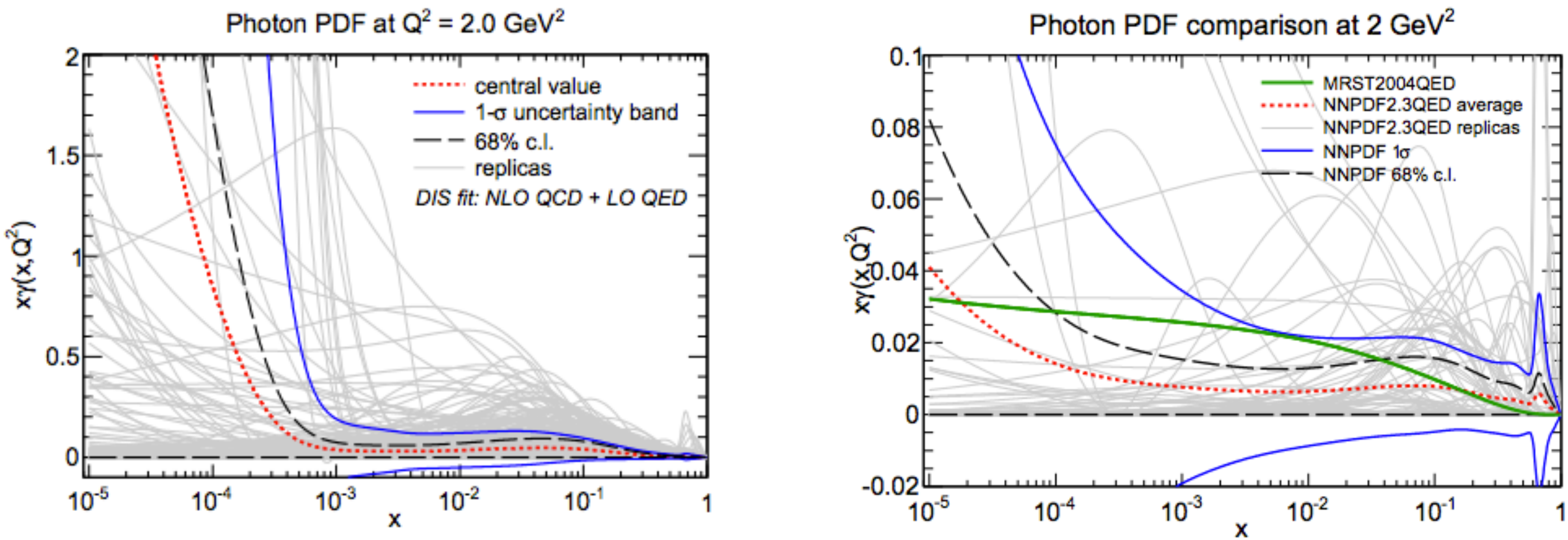


Photon-induced Drell-Yan

Ball et al, Nucl.Phys. B877 (2013) 290-320

Photon PDFs

Data-driven knowledge



Ball et al, Nucl.Phys. B877 (2013) 290-320

- Data-driven approach associated with a large uncertainty on photon PDF
- Theory breakthrough: LUX PDF [Manohar, Nason, Salam, Zanderighi,1607.04266]

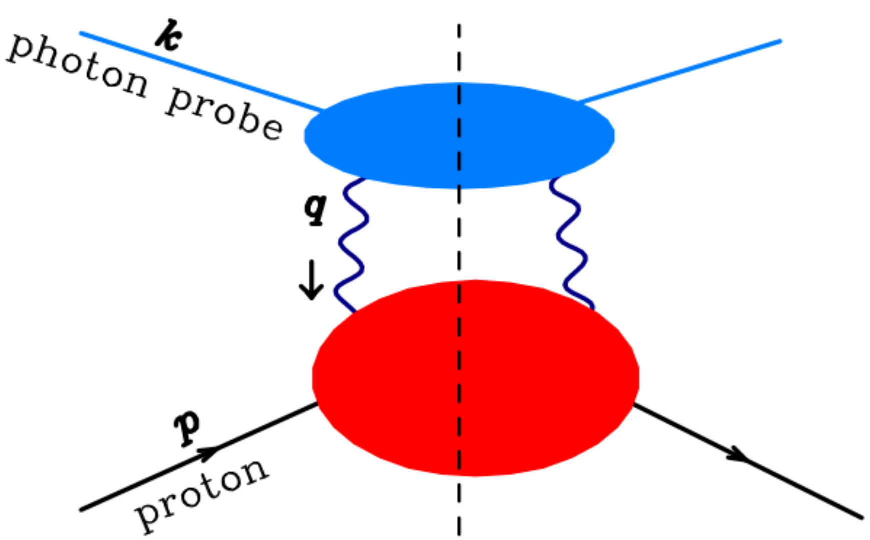
Photon PDFs

- QED is perturbative down to low scales ⟹ The photon must be computable if the input substructure is known
- Manohar et al: write the cross section for a chosen BSM process, e.g. production of heavy supersymmetric lepton L in ep collision (Drees, Zeppenfeld 1989)

$$\sigma = \frac{1}{4p\cdot k}\int \frac{d^4q}{(2\pi)^4 q^4} e^2_{\rm ph}(q^2)\left[4\pi W_{\mu\nu}(p,q)\,L^{\mu\nu}(k,q)\right]\; 2\pi\delta((k-q)^2 - M^2)$$

$$\sigma = c_0 \sum_a \int_x^1 \frac{dz}{z}\,\hat\sigma_a(z,\mu^2)\,\frac{M^2}{zs} f_{a/p}\left(\frac{M^2}{zs},\mu^2\right)$$

$$l(k) + p(p) \to L(k') + X$$



$$\sigma = \frac{c_0}{2\pi}\int_x^{1-\frac{2xm_p}{M}} \frac{dz}{z}\int_{Q^2_{\rm min}}^{Q^2_{\rm max}} \frac{dQ^2}{Q^2}\alpha^2_{\rm ph}(-Q^2)\Bigg[\Bigg(2 - 2z + z^2 + \frac{2x^2m_p^2}{Q^2} + \frac{z^2Q^2}{M^2} - \frac{2zQ^2}{M^2} - \frac{2x^2Q^2m_p^2}{M^4}\Bigg) F_2(x/z,Q^2) + \left(-z^2 - \frac{z^2Q^2}{2M^2} + \frac{z^2Q^4}{2M^4}\right) F_L(x/z,Q^2)\Bigg], \quad (3)$$

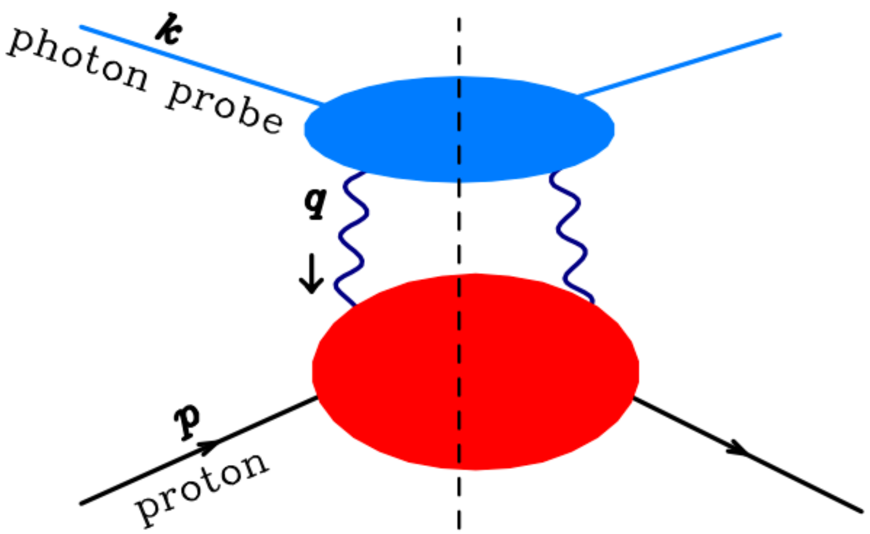
Manohar et al 1607.04266

Theory-driven knowledge

Photon PDFs

- QED is perturbative down to low scales ⟹ The photon must be computable if the input substructure is known
- Manohar et al: write the cross section for a chosen BSM process, e.g. production of heavy supersymmetric lepton L in ep collision (Drees, Zeppenfeld 1989)
- Equate the two expressions and find analytically the PDF of the photon

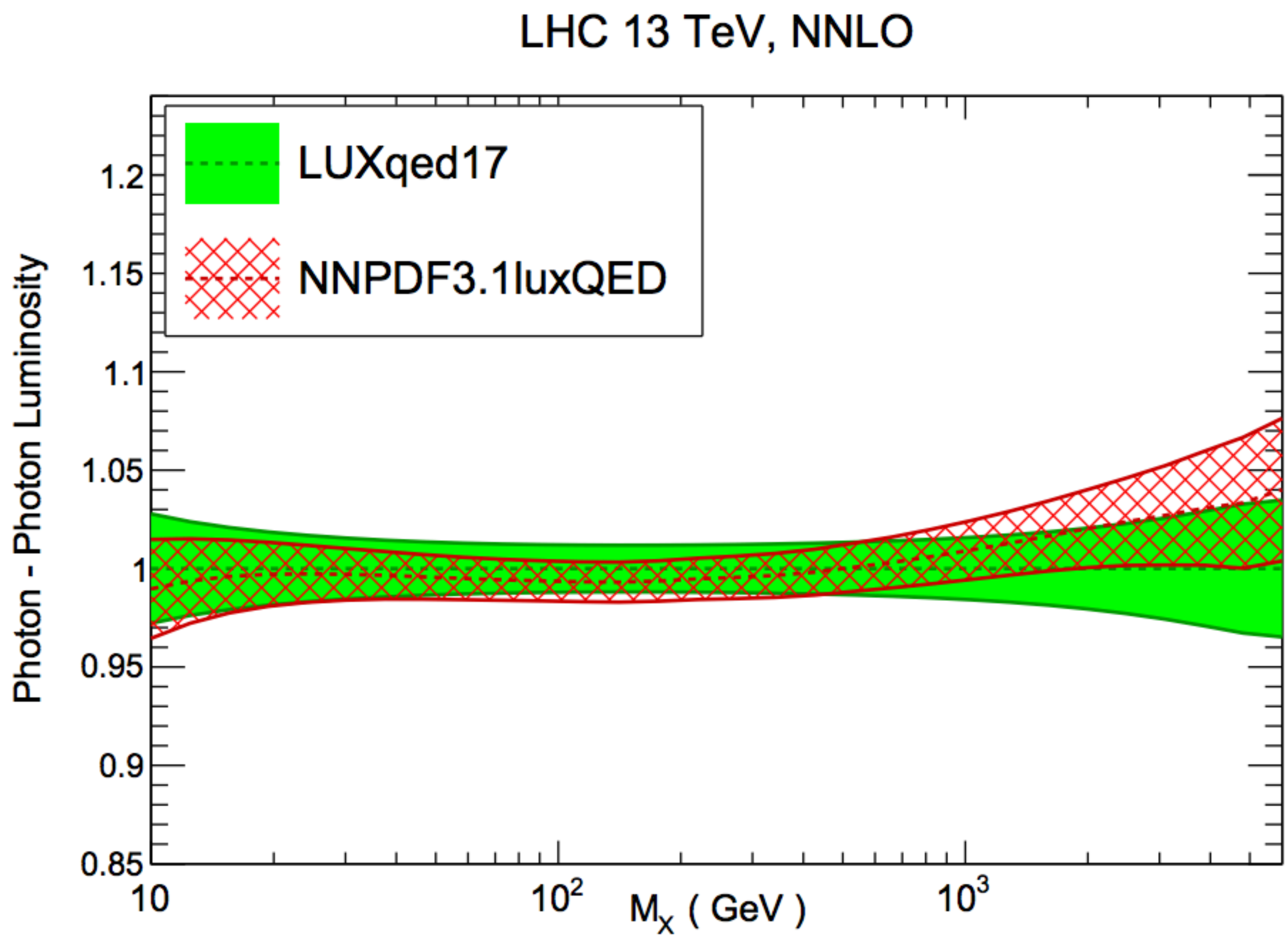
⇒ PDFs expressed in terms of the structure functions integrated over all scales, including elastic form factors (in the x →1 region)



$$x f_{\gamma/p}(x,\mu^2) = \frac{1}{2\pi\alpha(\mu^2)} \int_x^1 \frac{dz}{z} \left\{ \int_{\frac{x^2 m_p^2}{1-z}}^{\frac{\mu^2}{1-z}} \frac{dQ^2}{Q^2} \alpha^2(Q^2) \left[\left(z p_{\gamma q}(z) + \frac{2x^2 m_p^2}{Q^2} \right) F_2(x/z, Q^2) - z^2 F_L\left(\frac{x}{z}, Q^2\right) \right] - \alpha^2(\mu^2) z^2 F_2\left(\frac{x}{z}, \mu^2\right) \right\},$$

<u>Theory-driven knowledge</u>

Photon PDFs



Bertone et al, 1712.07053

(Data+Theory)-driven knowledge