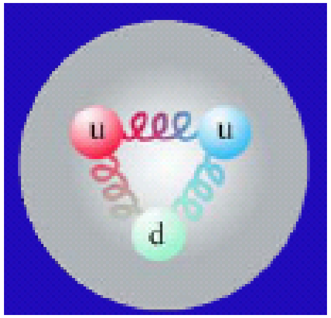
The Deep Structure of the Proton-

-what do we know and how can we do better

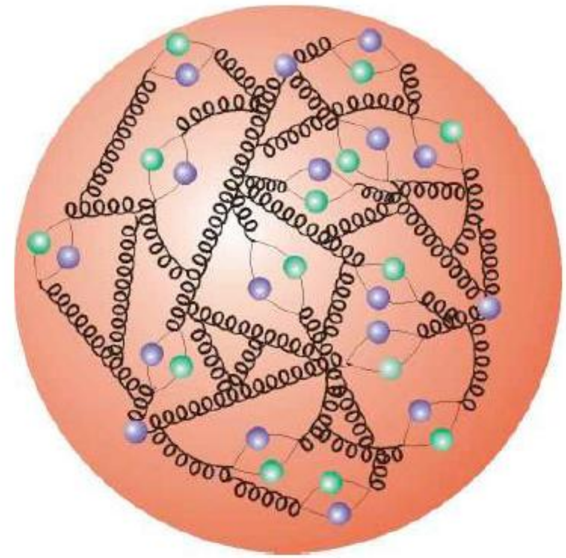
May 2025

A.M.Cooper-Sarkar

Oxford



Are there 3 quarks in the proton?



No - there are many more- there are quarks and anti-quarks and gluons- collectively known as ‘partons’

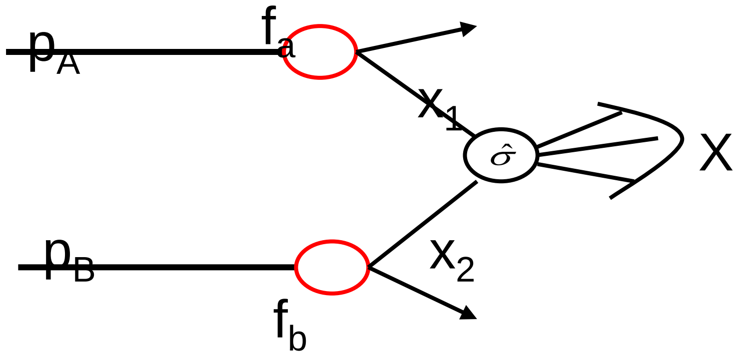
BUT

$$\int [q(x) - \bar{q}(x)].dx = 3$$

the net number of quarks is 3

3 valence quarks give the proton its flavour properties

This is known as the Gross-Llewellyn-Smith sum-rule



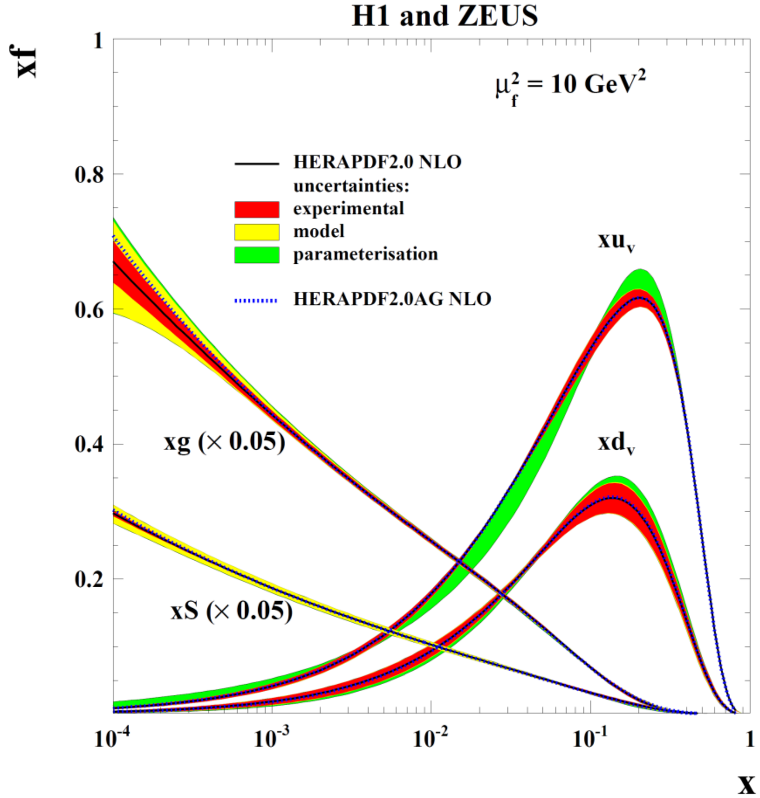
$$\sigma_X = \sum_{a,b} \int_0^1 dx_1 dx_2\, f_a(x_1, \mu_F^2)\, f_b(x_2, \mu_F^2)$$
$$\times\ \hat{\sigma}_{ab\to X}\left(x_1, x_2, \{p_i^\mu\}; \alpha_S(\mu_R^2), \alpha(\mu_R^2), \frac{Q^2}{\mu_R^2}, \frac{Q^2}{\mu_F^2}\right)$$

When a collision happens at the LHC a parton from one of the protons (A) takes a fraction x_1 of the momentum of this proton and a parton from the other proton (B) takes a fraction x_2 of the momentum of this proton, such that **the centre-of-mass energy squared of this collision is not s =(13 TeV)^2 it is x_1x_2s**

Thus the energy involved in each collision– its scale- is different

AND the probability of each collision depends on the joint probability that proton A contained a parton of momentum fraction x_1 , $f_a(x_1)$, and proton B contained a parton of momentum fraction x_2 , $f_b(x_2)$, and that these two partons were of the right type, or flavour to interact to make final state X (as embodied in the cross section for interaction $\sigma_{ab\to X}$).

The probabilities * momentum fractions: $x_1f_a(x_1)$, $x_2f_b(x_2)$, are the parton momentum distributions or PDFs



What is much more interesting than the numbers of quarks is their momentum distributions xq(x)- known as Parton Distributions Functions PDFs- where x is the fractional momentum that the struck quark takes of the proton’s momentum

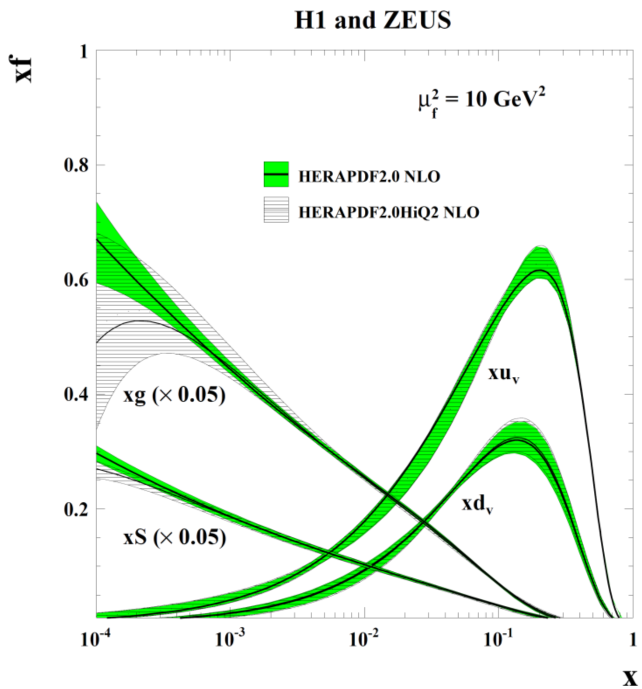
In this figure the total momentum taken by all the partons sum to unity

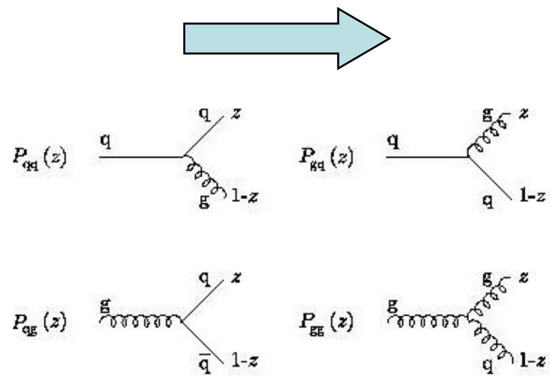
There are 2 u-valence quarks to each 1 d-valence quark- though their momentum distributions are NOT exactly the same

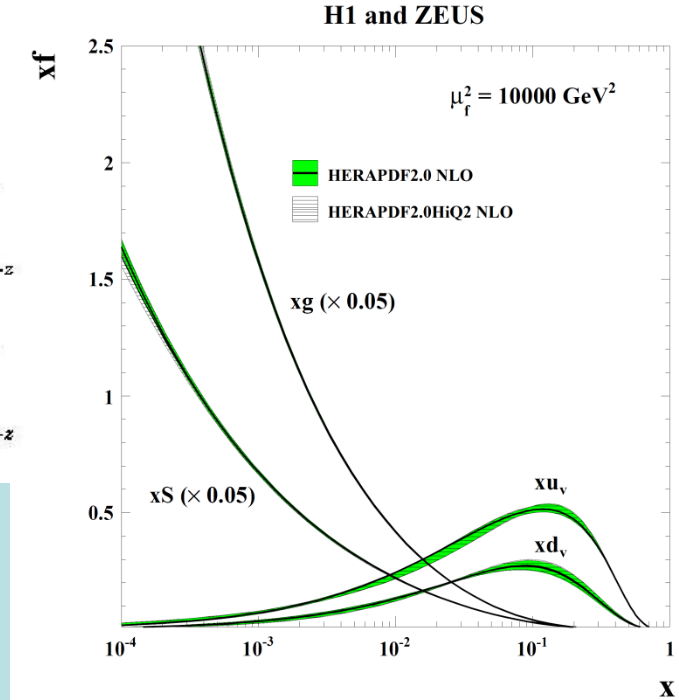
And you can see that there is a ‘SEA’ of quark-antiquark pairs as well as GLUONS at low-momentum

These momentum distributions depend on the dynamics of the interactions between quarks and gluons-

Quantum Chromo Dynamics







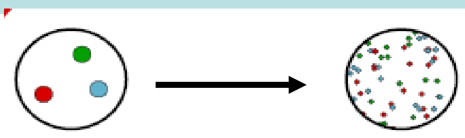
The valence quarks are radiating gluons and the gluons are splitting into quark-antiquark pairs

Parton momentum distributions change with the scale of the probe: $Q^2=p^2-E^2\sim 10\ \mathrm{GeV}^2$ is typical scale for low energy experiments

Whereas $Q^2=\sim 10^{4-6}\ \mathrm{GeV}^2$ are the scales that we are now probing at the Large Hadron Collider

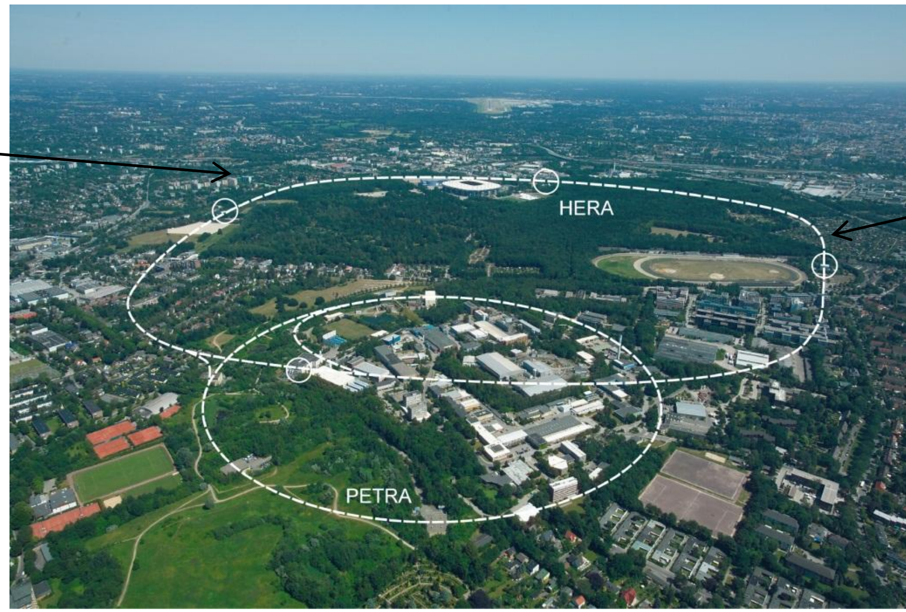
And at this scale the proton is pretty well all glue

And the harder we hit the more of this activity we see- rather than seeing further sub-structure



How did we come to know all this?

From Deep Inelastic Scattering of leptons on hadrons. The HERA e-p collider provides the most extensive data set today, but there are also earlier fixed target leptob-scattering results



The HERA the e-p collider at DESY, Hamburg.

~500pb^{-1} per experiment split ~equally between e^+ and e^- beams:
ARXIV:1506.06042

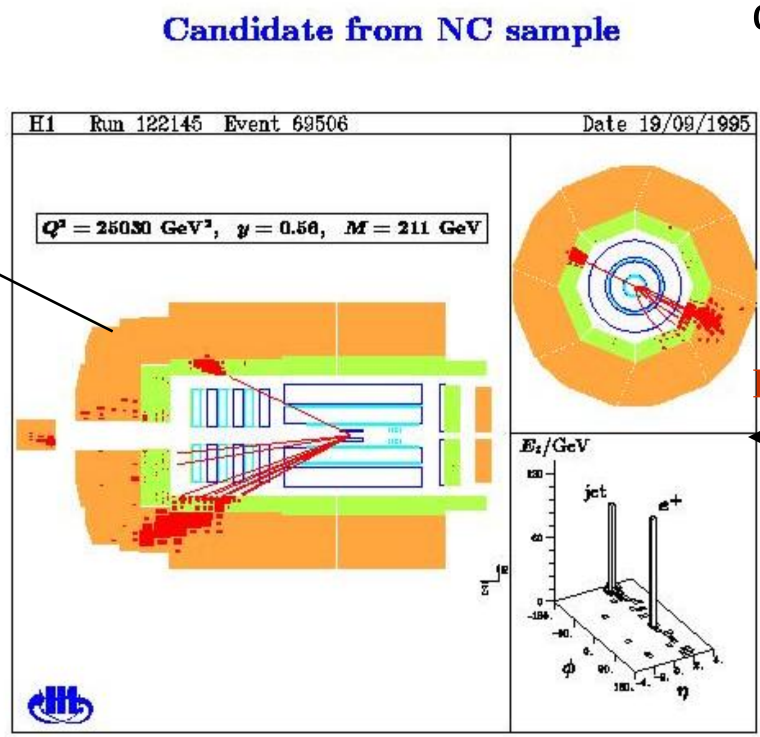
Running at Ep = 920, 820, 575, 460 GeV
$\sqrt{s}$ = 320, 300, 251, 225 GeV

From 1992-2007

PDFs were first investigated in deep inelastic lepton-nucleon scattering -DIS

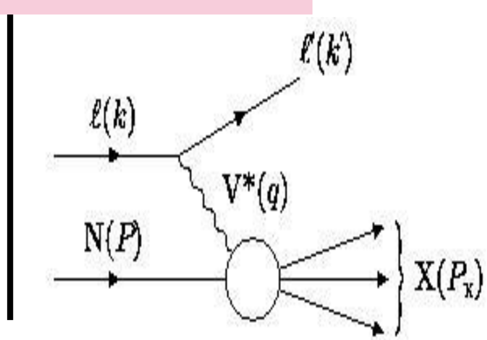
E'

E_e



E_p

dσ ~



^{2} $L^{\mu\nu} W_{\mu\nu}$

Leptonic tensor - calculable

Hadronic tensor- constrained by Lorentz invariance

$q = k - k'$, $Q^2 = -q^2$

This is the scale of the vector boson probe

$s = (p + k)^2$

$x = Q^2 / (2p.q)$

$y = (p.q)/(p.k)$

These are 4-vector invariants

$Q^2 = s\ x\ y$

$s = 4\ E_e\ E_p$
$Q^2 = 4\ E_e\ E'\ \sin^2\theta_e/2$
$y = (1 - E'/E_e\ \cos^2\theta_e/2)$
$x = Q^2/sy$

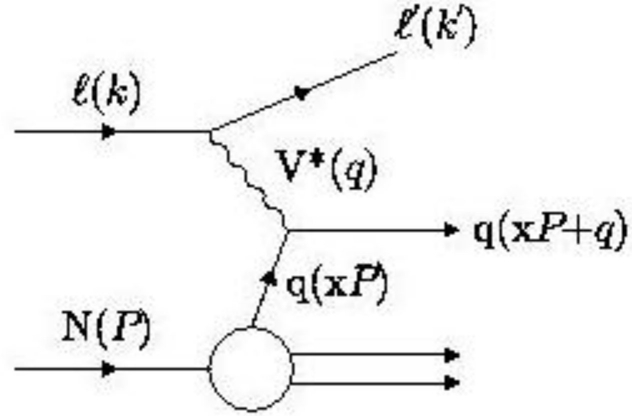
The kinematic variables are measurable

Without assumptions as to what goes on in the hadron the double differential cross-section for $e^{\pm}$ N scattering can be written as

$$\frac{d^2\sigma(e\pm N)}{dxdy} = \frac{2\pi\alpha^2 s}{Q^4}\left[Y_+ \mathbf{F_2}(x,Q^2) - y^2\, \mathbf{F_L}(x,Q^2) \pm Y_- x\mathbf{F_3}(x,Q^2)\right], \quad Y\pm = 1 \pm (1-y)^2$$

Leptonic part

hadronic part



F_2, F_L and xF_3 are structure functions

which express the dependence of the cross-section on the structure of the nucleon (hadron)

The Quark-Parton Model interprets these structure functions as related to the momentum distributions of point-like quarks or partons within the nucleon AND the measurable kinematic variable $x = Q^2/(2p.q)$ is interpreted as the FRACTIONAL momentum of the incoming nucleon taken by the struck quark

$$(xP+q)^2 = x^2p^2 + q^2 + 2xp.q \sim 0$$

for massless quarks and $p^2 \sim 0$ so

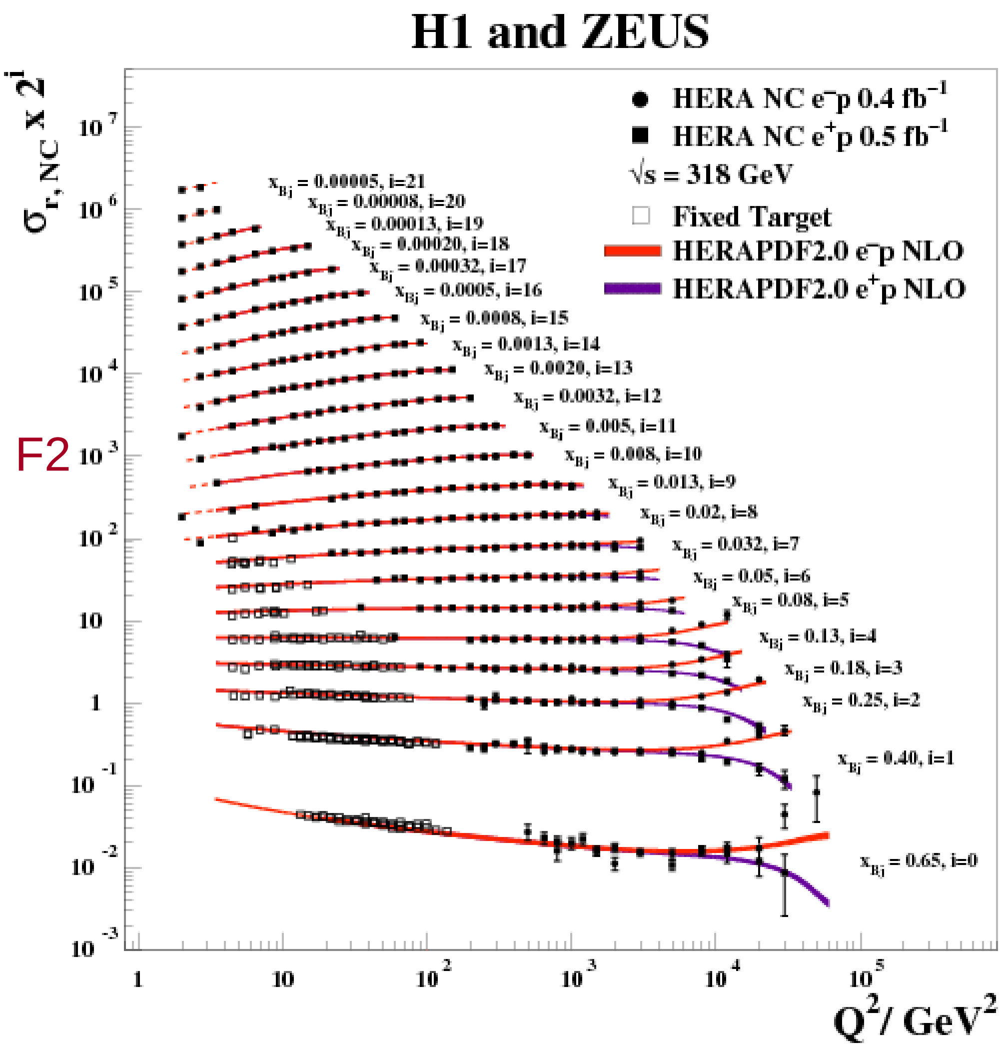
$$x = Q^2/(2p.q)$$

The FRACTIONAL momentum of the incoming nucleon taken by the struck quark is the MEASURABLE quantity x

We can extract all three structure functions experimentally by looking at the x, y, Q^2 dependence of the double differential cross-section- thus we can check out the parton model predictions

So the prediction was that the differential cross-section for lepton-proton scattering would depend only on the structure function F_2

F2



AND that

$F_2(x,Q^2) = \Sigma_i\, e_i^2(xq(x) + x\bar{q}(x))$

F_2 would be independent of the scale of the probe Q^2

How good is this?

Pretty good- this is a log plot

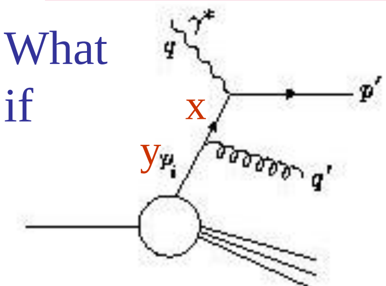
Non-point like structure would have ~$1/Q^2$ behaviour, we are sure there is no substructure down to 10^{-19} m now.

But it's clearly not perfect

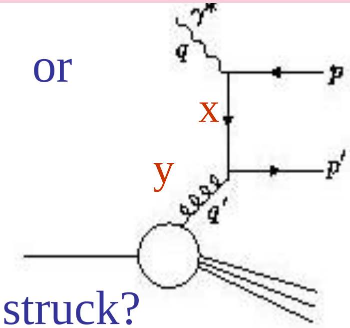
Before we leave this page NOTE the terrific kinematic reach of the HERA e-p scattering experiments – 5 decades in x and Q^2

The theory of Quantum Chromo-Dynamics (QCD) improves on the Quark Parton Model

What if

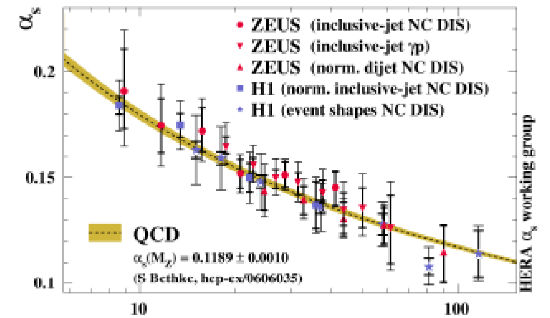


or



Before the quark is struck?

Note there is a minimum scale Q^2_0 below which we do not dare to tread – the calculation would become non-perturbative because the strong coupling $\alpha_S(Q^2_0)$ is too large



The strong coupling α_S decreases as the scale Q^2 increases

$y > x, \;\; z = x/y$

$$\frac{F_2(x)}{x} = \int dy\, dz\, \delta(x - zy)\, q(y) \left[\sigma^{POINT}(z) + \sigma(\gamma^* q \to qg)\right]$$

$$\sigma(\gamma^* q \to qg) = \frac{e_q^2 \alpha_s}{2\pi} P_{qq}(z) \ln\frac{Q^2}{Q_0^2}$$

$$\frac{F_2(x,Q^2)}{x} = e_q^2\left[q(x) + \Delta q(x,Q^2)\right] = e_q^2 q(x,Q^2)$$

$$\Delta q(x,Q^2) = \frac{\alpha_s}{2\pi} \ln\frac{Q^2}{Q_0^2} \int_x^1 \frac{dy}{y} q(y) P_{qq}\left(\frac{x}{y}\right)$$

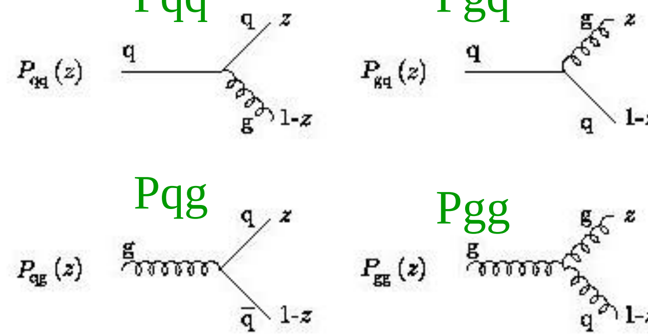
$$\frac{dq(x,Q^2)}{d\ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{dy}{y} \left[Pqq(z) q(y,Q^2) + Pqg(z) g(y,Q^2)\right]$$

$$\frac{dg(x,Q^2)}{d\ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{dy}{y} \left[\Sigma_q Pgq(z) q(y,Q^2) + Pgg(z) g(y,Q^2)\right]$$

The DGLAP parton evolution equations

Pqq Pgq

Pqg Pgg

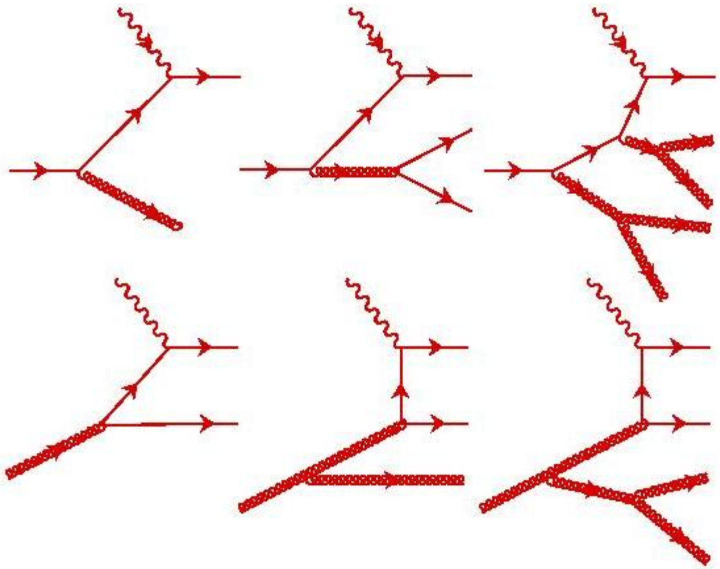


So $F_2(x,Q^2) = \Sigma_i e_i^2 (xq(x,Q^2) + x\bar{q}(x,Q^2))$

in Leading Order QCD

The theory predicts the rate at which the parton distributions (both quarks and gluons) evolve with Q^2- (the energy scale of the probe) -BUT it does not predict their shape at Q^2_0

But is Leading Order enough? What if higher orders are needed?



$Pqq(z) = P^0qq(z) + \alpha_s\, P^1qq(z) + \alpha_s^2\, P^2qq(z)$

LO NLO NNLO

This just means we need to calculate higher-order splitting functions

And the structure function F_2 is no longer so simply expressed in terms of partons -

convolution with coefficient functions is needed –

but these are calculable in QCD

$$\frac{F_2(x,Q^2)}{x} = \int_0^1 \frac{dy}{y}\left[\Sigma_i C_2(z,\alpha_s)\, q_i(x,Q^2) + C_g(z,\alpha_s)\, g(y,Q^2)\right]$$

$$C_2(z,\alpha_s) = e_i^2\,[\delta(1-z) + \alpha_s f_2(z)]$$

$$C_g(z,\alpha_s) = \alpha_s f_g(z)$$

$$F_L(x,Q^2) = \frac{\alpha_s}{\pi}\left[\frac{4}{3}\int_0^1 \frac{dy}{y} z^2 F_2(y,Q^2) + 2\Sigma_i e_i^2 \int_0^1 \frac{dy}{y} z^2(1-z)\, y g(y,Q^2)\right]$$

And the structure function F_L is no longer zero..

… it depends on the gluon (we are no longer scattering from purely spin-1/2 Quarks)

It remains true that we know how to evolve parton distributions from low to high scale and how to construct the measurable structure functions from them.

The only thing we don't know from THEORY is what the shapes of the parton distributions are at the starting scale- this is a non-perturbative problem that is not yet solved. Lattice gauge theory may one day help with this,but right now, we must induce it from data in the perturbative region.

The cleanest data theoretically are those from Deep Inelastic Scattering

So in more detail…

There is differing information according to which leptons and nucleons you

$$F_2(lp) = x(\frac{4}{9}(u+\bar{u}) + \frac{1}{9}(d+\bar{d}) + \frac{1}{9}(s+\bar{s}) + \frac{4}{9}(c+\bar{c})$$

$$F_2(lN) = \frac{5}{18}\,x\left[u+\bar{u}+d+\bar{d}+\frac{2}{5}(s+\bar{s})+\frac{8}{5}(c+\bar{c})\right]$$

Assuming u in proton = d in neutron – strong-isospin

Charged lepton proton and deuteron data gives different combinations of quark flavours.
(Low energy γ-exchange only formulae, HERA went beyond this to W,Z exchange)

$$F_2(\nu,\bar{\nu}N) = x(u+\bar{u}+d+\bar{d}+s+\bar{s}+c+\bar{c})$$

$$xF_3(\nu,\bar{\nu}N) = x(u-\bar{u}+d-\bar{d}) = x(u_v+d_v)$$

u,d,s quarks and antiquarks are intrinsic to the proton heavier quarks like c, b are generated in gluon to q-qbar splitting

Further information can be extracted from **neutrino beam data**

The **gluon** comes indirectly from QCD from the rate of change with Q^2

$$\frac{d\,q(x,Q^2)}{d\ln Q^2} = \frac{\alpha_s}{2\pi}\int_x^1 \frac{dy}{y}\left[P_{qq}\left(\frac{x}{y}\right)q(y,Q^2) + P_{qg}\left(\frac{x}{y}\right)g(y,Q^2)\right]$$

These expressions to give an idea of which flavours contribute. In practice there are higher-order corrections but it is all completely calculable

You just need to know what the PDFs are at a starting scale Q^2_0 – and QCD will tell you what they are for any scale $Q^2 > Q^2_0$

How do you know what the PDFs are at the starting scale?
You don't, you have to parametrise them at a starting scale Q^2_0

$$xq_i(x) = A_i x^{B_i}(1-x)^{C_i} P_i(x), \qquad q_i = \{u,\ \bar{u},\ d, \bar{d},\ s,\ \bar{s},\ c,\ g\},$$

Where $P_i(x)$ can be ordinary polynomials of x, or $\sqrt{x}$, or Chebyshevs, Bernstein polynomials, typically ~30 parameters---- or even a neural net

Some parameters are fixed through conservation of the total amount of momentum and the number of quarks of each flavour - but others are model choices-

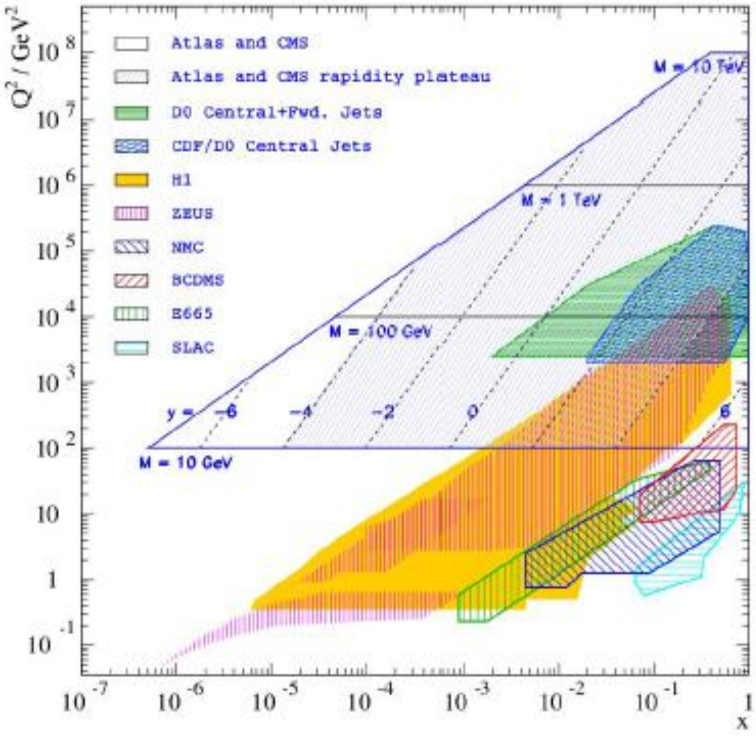
Model choices $\Rightarrow$Form of parametrization at Q^2_0, value of Q^2_0, which data are accepted for the fit, what kinematic cuts are applied to the data, 'heavy quark schemes', values of heavy quark masses

Use QCD to 'evolve' these PDFs to higher scale $Q^2 > Q^2_0$
Construct predictions for the measurable structure functions in terms of PDFs for ~4000 data points across the x,Q^2 plane
Perform χ2 fit to the data.

The fact that so few parameters allows us to fit so many data points established QCD as the THEORY OF THE STRONG INTERACTION and provided the first measurements of the strong interaction coupling, $\alpha_s(M_Z)$

NOW we evolve the PDFs from the HERA region up to the LHC region

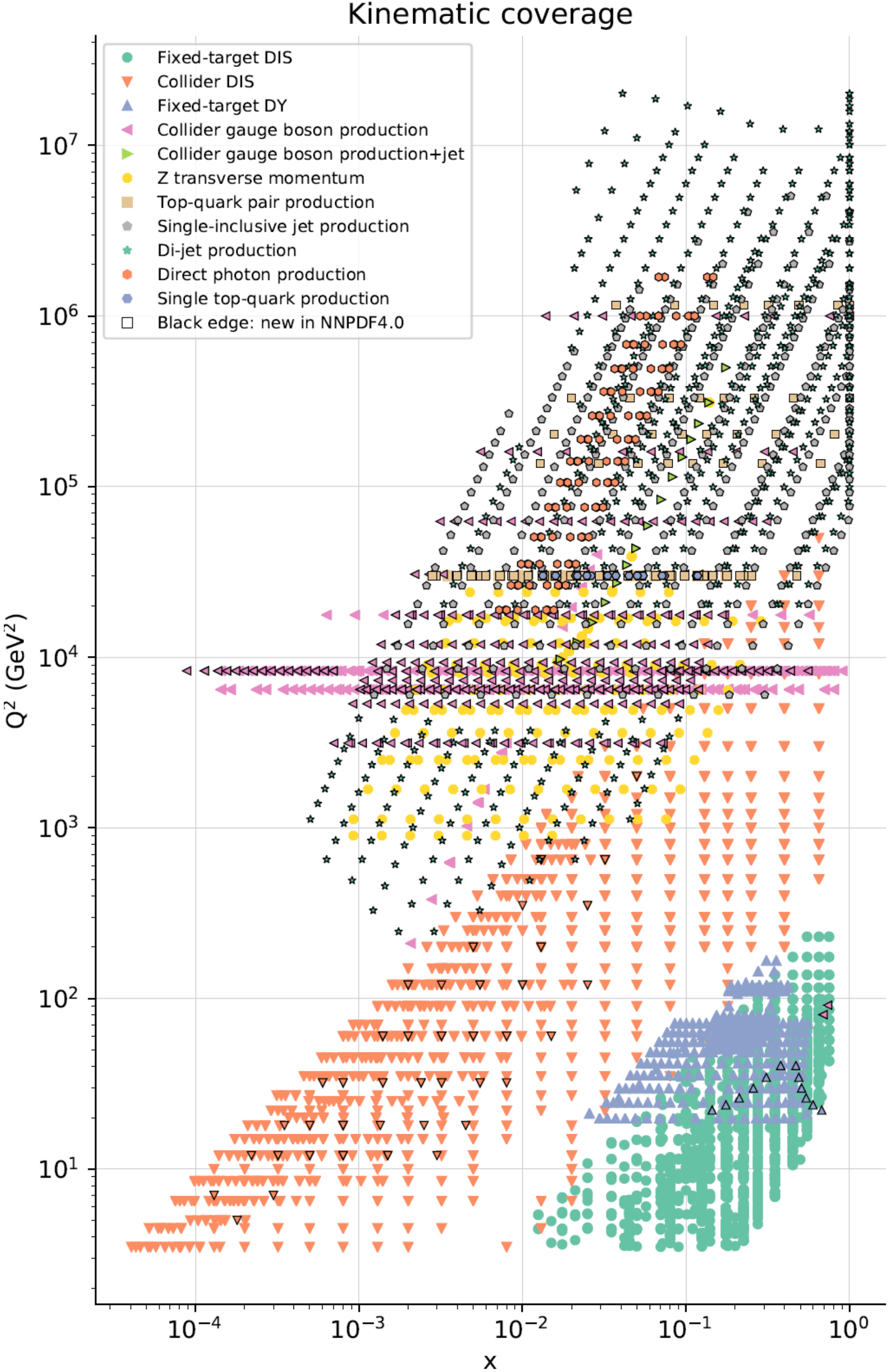
LHC kinematics



$$x_{1,2} = \frac{M}{\sqrt{S}} \exp(\pm y)$$

- To create an object with a mass M and rapidity y for the LHC center of mass energy S we need to have two partons with momentum fraction of corresponding proton of x_1 and x_2
- At electroweak scale, LHC needs PDFs with $x>10^{-4}$

And assuming the PDFs are universal—actually only proven for Drell-Yan– we predict LHC cross sections.
So we can also use SM LHC processes –those which can be reliable predicted to fixed order in QCD---to improve the PDFs….

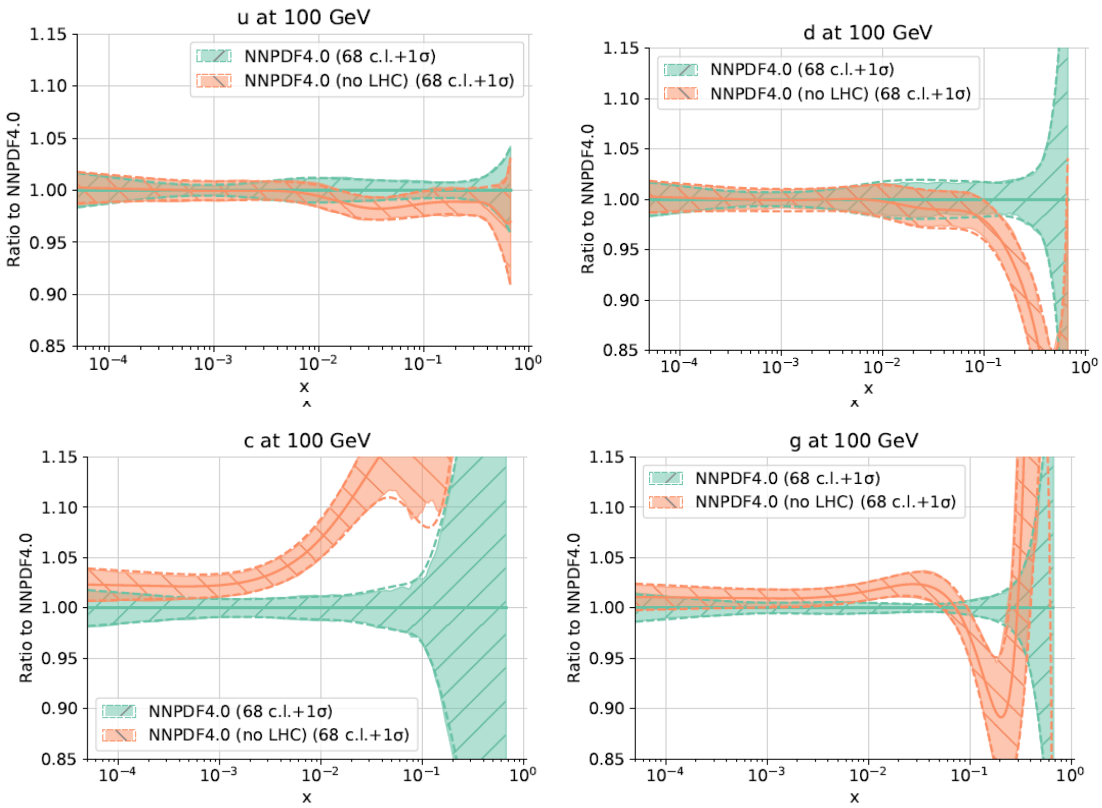


We now use many other processes than deep-inelastic scattering for the determination of PDFs

We use many SM processes that can be reliably calculated to NNLO in fixed order QCD

- Drell-Yan data from fixed targets and the Tevatron and LHC
- W,Z rapidity spectra from Tevatron and LHC
- Jet pT spectra from Tevatron and LHC
- Top-anti-top differential cross-sections
- W and Z +jet spectra, or W,Z pt spectra
- W and Z +heavy flavours
- Beware: there may be new physics at high scale that we ‘fit away’

How do the PDFs look before/after LHC data?



Some change in shape
Some reduction in uncertainty

Illustration from NNPDF

But several groups extract PDFs and there are significant differences between them because of slightly different model choices:

- Exact choice of data entering fit and cuts impose on them
- Choice of heavy quark masses, heavy quark schemes
- Choice of starting scale for QCD evolution, Value of $\alpha_S(M_Z)$

And differing methodology

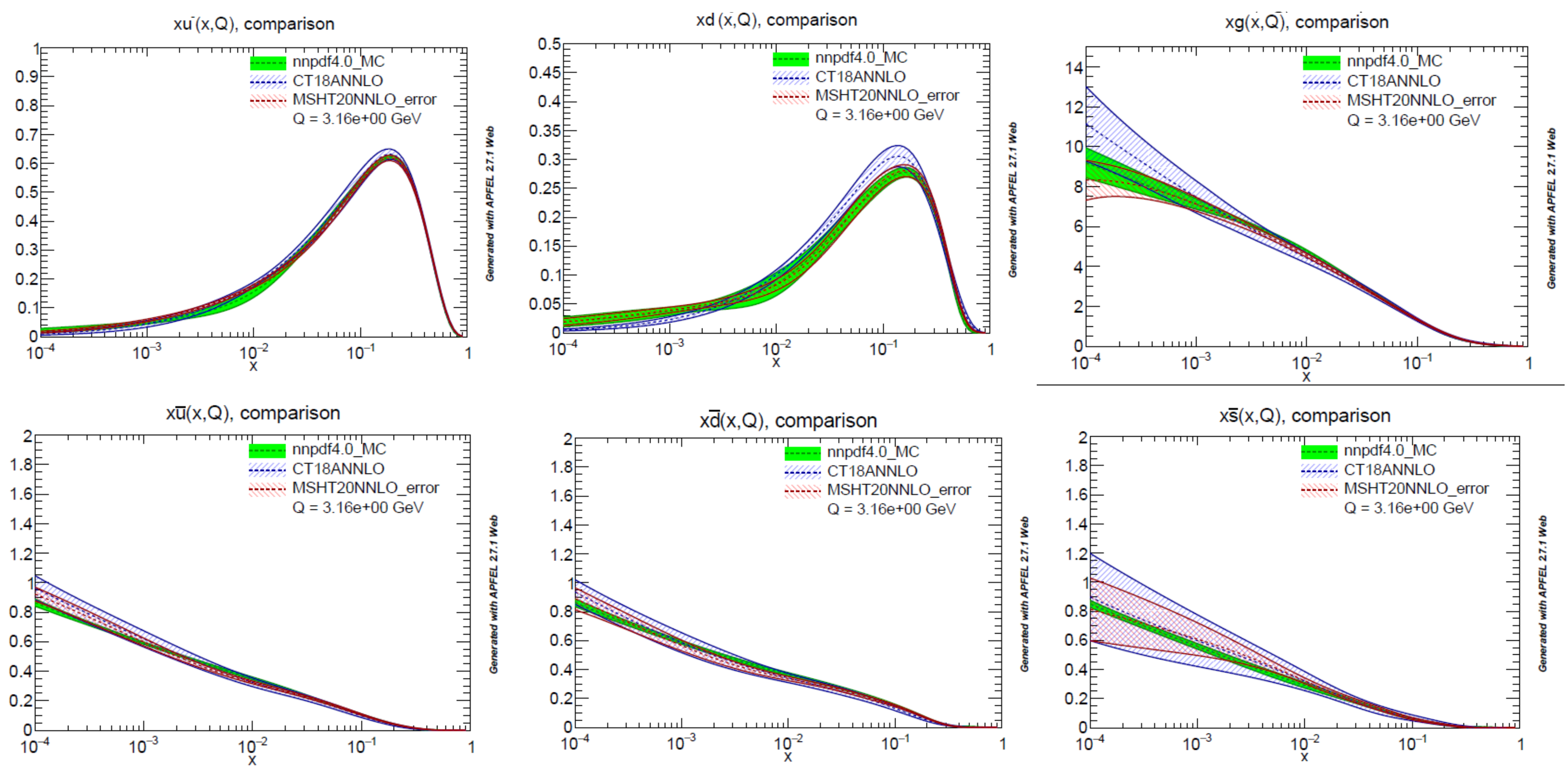
- Use of parametrization, what parametrization
- Use of NN, what architecture, training, stopping etc

The ‘big three’ of PDF fitting groups are CT, NNPDF and MSHT

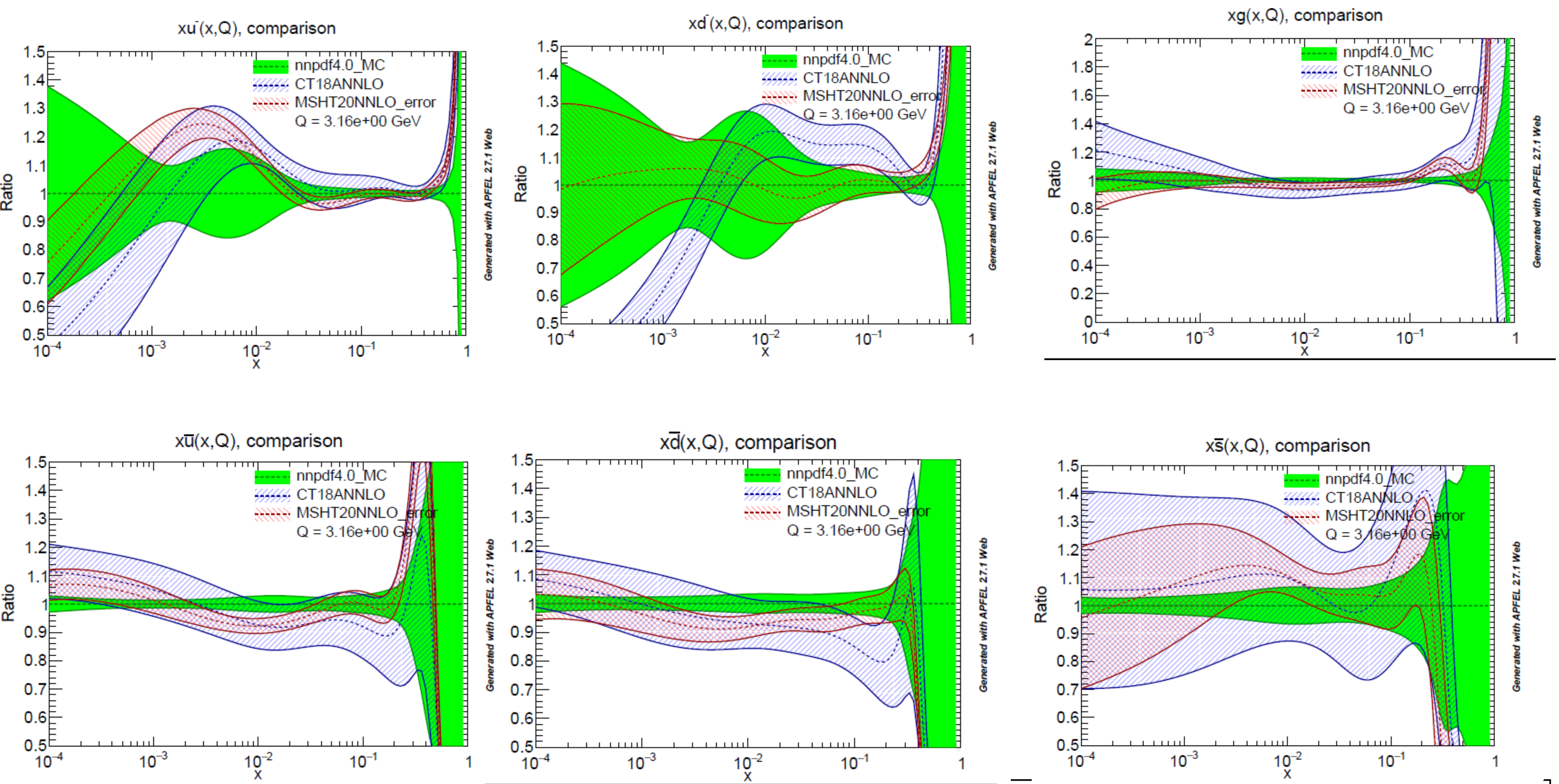
The latest CT18, NNPDF4.0, MSHT20. Be agnostic in the choice between these.

The one with the smallest uncertainties is not necessary the best.

Note CT18 actually came out end 2019 (1912.10053), MSHT came out end 2020 (2012.04684), NNPDF4.0 in 2021(2109.02653). NNPDF3.1 in 2017



This looks as if we are doing reasonably well –but look at ratios….



Differences are more obvious in ratio. They are large at small-x and at high-x, where there is less data. Differing model choices matter.

Different methodologies matter.

So also do theory choices, standard is NNLO, often using NNLO/NLO ‘k-factors’ but such calculations can differ. Even when using direct NNLO grids - treatment of grid uncertainties can differ.

PDFs also differ in how they evaluate their uncertainties some use enhanced $\chi 2$ tolerances -- closer to the hypothesis testing criterion– but this is a whole lecture series in itself.

One way to see the impact of the uncertainties on the parton distribution functions is in terms of parton-parton luminosities, which are the convolution of the purely partonic part of the sub-process cross-section.

The quark-antiquark and gluon-gluon luminosities for various PDFs are compared here for 13 TeV LHC running in terms of the centre of mass energy of the parton sub- process M_X

Small M_X corresponds to small x and

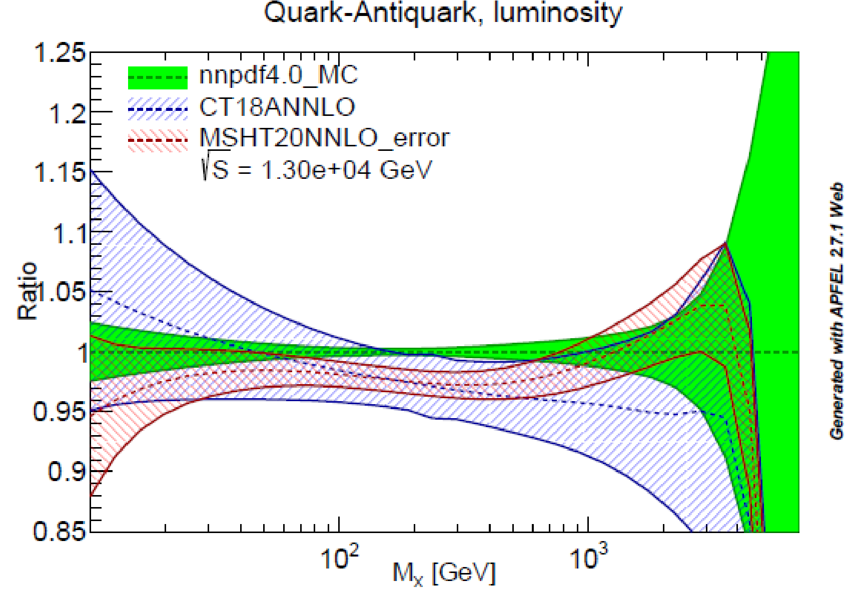
Large M_X to large x

So for quark-antiquark production of W or Z bosons ----at Mx ~80,90 GeV

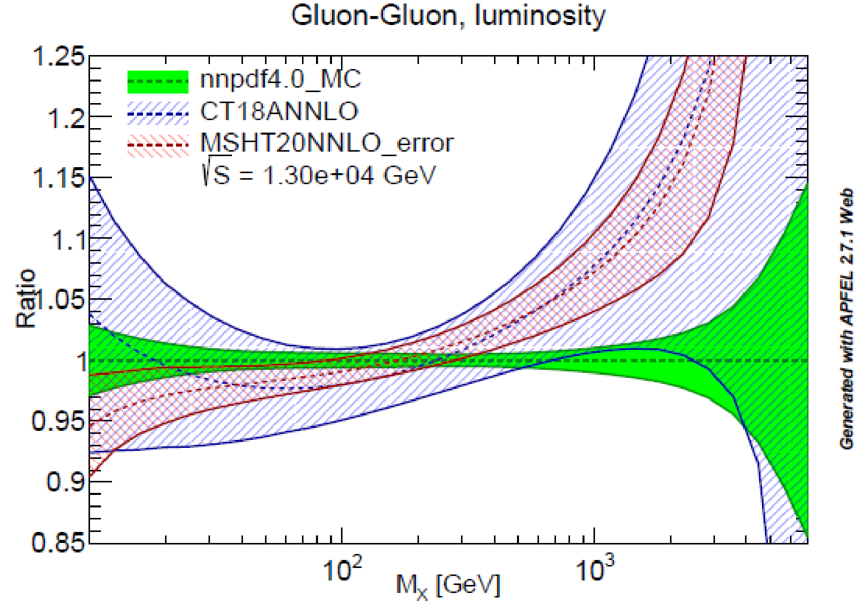
Or for gluon-gluon production of Higgs at ---Mx~125 GeV

the parton-parton luminosities are fairly well known- well this is as good as it gets--

This is not so for higher mass particles that could be produced by ‘Beyond’ Standard Model (BSM) physics



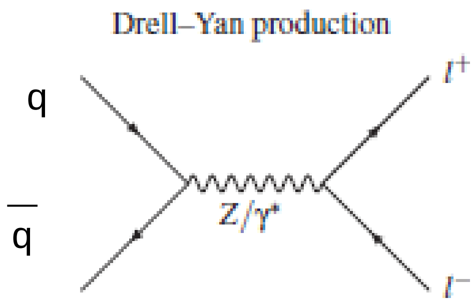
$$\frac{\partial \mathcal{L}_{\Sigma_q(q\bar{q})}}{\partial \hat{s}} = \frac{1}{s}\int_{\tau}^{1} \frac{dx}{x} \sum_{q=d,u,s,c,b} \left[f_q(x,\hat{s}) f_{\bar{q}}(\tau/x,\hat{s}) + f_{\bar{q}}(x,\hat{s}) f_q(\tau/x,\hat{s})\right],$$

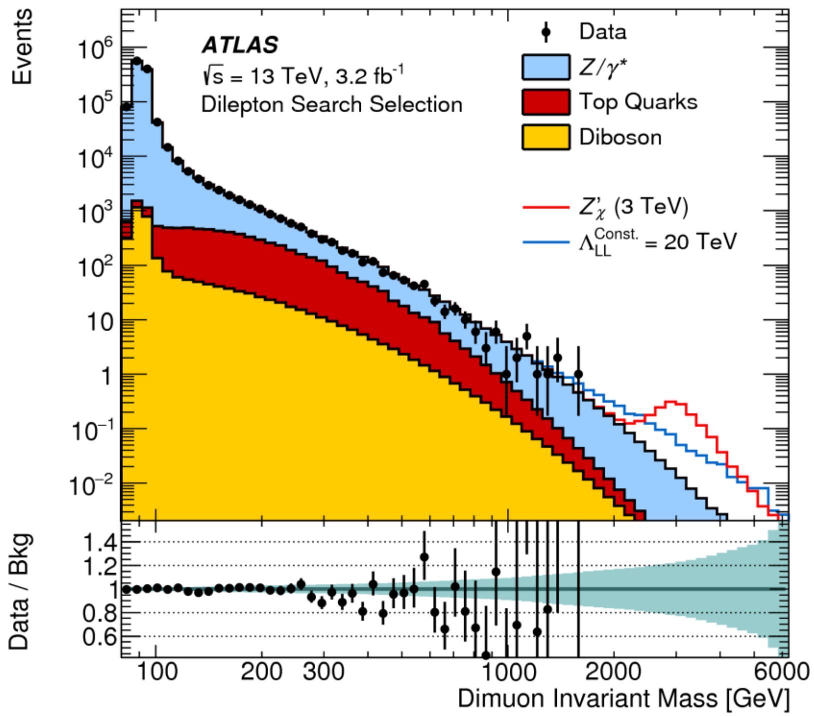


$$\frac{\partial \mathcal{L}_{gg}}{\partial \hat{s}} = \frac{1}{s}\int_{\tau}^{1} \frac{dx}{x} f_g(x,\hat{s}) f_g(\tau/x,\hat{s}),$$

Uncertainty in the high-x sea?-one example

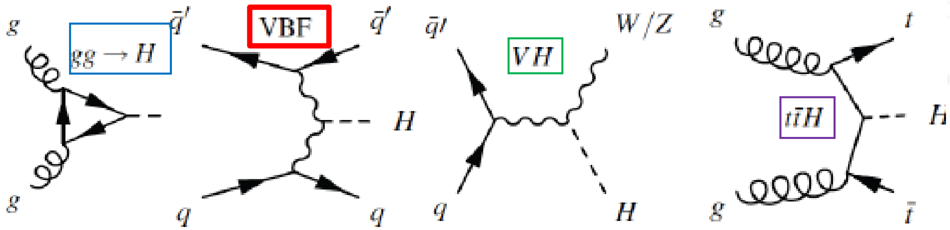
Current BSM searches in High Mass Drell-Yan are limited by high-x antiquark uncertainties as well as by high-x valence quark uncertainties





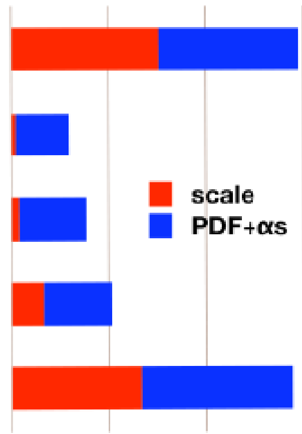
Drell-Yan is a term for q-qbar → $\mu^+ \mu^-$ collisions mediated by Z or virtual γ,Z bosons.
Some new theories predict higher mass Z' states, these have been excluded up to 2 TeV
The main reason we cannot do better is that the PDF uncertainty on the 'normal' Standard Model background is too big.

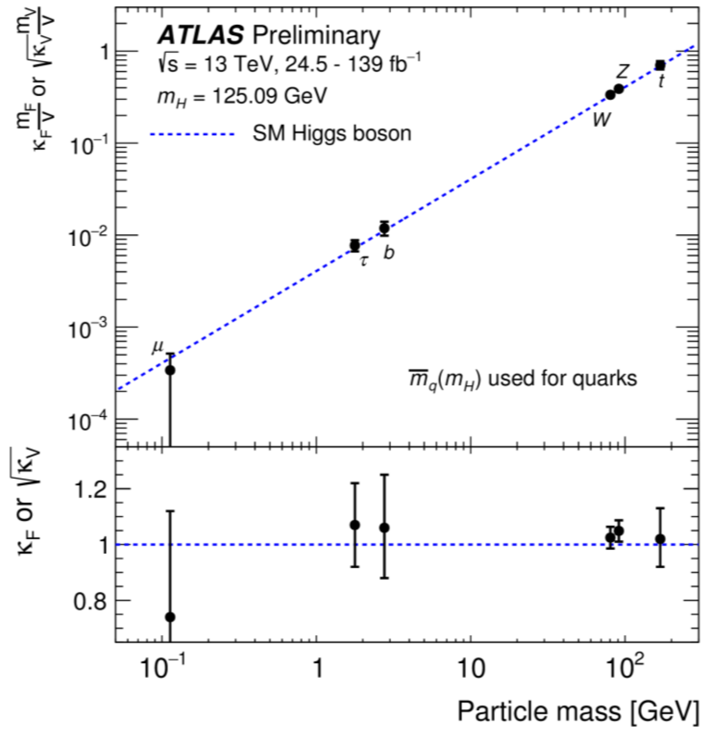
What about the Higgs? Gluon-gluon to Higgs is the dominant channel



PDFs and the Higgs

		σ (8 TeV)	uncertainty
NNLL QCD +NLO EW	gg→H	19.5 pb	14.7%
	VBF	1.56 pb	2.9%
NNLO QCD +NLO EW	WH	0.70 pb	3.9%
	ZH	0.39 pb	5.1%
NLO QCD	ttH	0.13 pb	14.4%



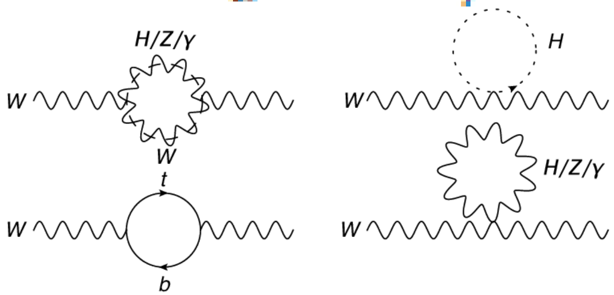


And the uncertainty on the calculation for this is ~50% from uncertainty on the gluon PDF

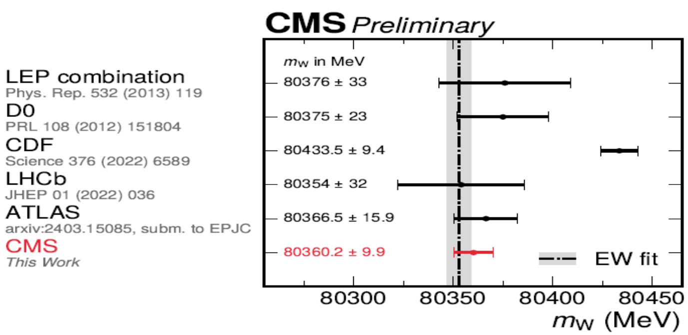
All indications are that the Higgs boson that we have seen is the Higgs boson of the Standard Model (SM) , but measurements are not as yet very accurate.
Beyond Standard Model (BSM) physics can manifest itself as small discrepancies from SM predictions. We need to reduce uncertainties on our predictions.

But as well as limiting our ability to identify BSM effects at high M_X, uncertainties on PDFs also limit indirect observations of new physics which we may hope to make by measuring discrepancies from the Standard Model (SM) values for fundamental parameters such as m_W – the W mass

$$m_W^2\left(1-\frac{m_W^2}{m_Z^2}\right)=\frac{\pi\alpha}{\sqrt{2}G_\mu}(1+\Delta r),$$



The W mass is predicted in the SM in terms of other SM parameters like the fine structure constant and the weak coupling G, but Δr represents higher order loops in the diagrams which are presently calculated with known particles like the top quark or Higgs, but could also contain BSM particles.



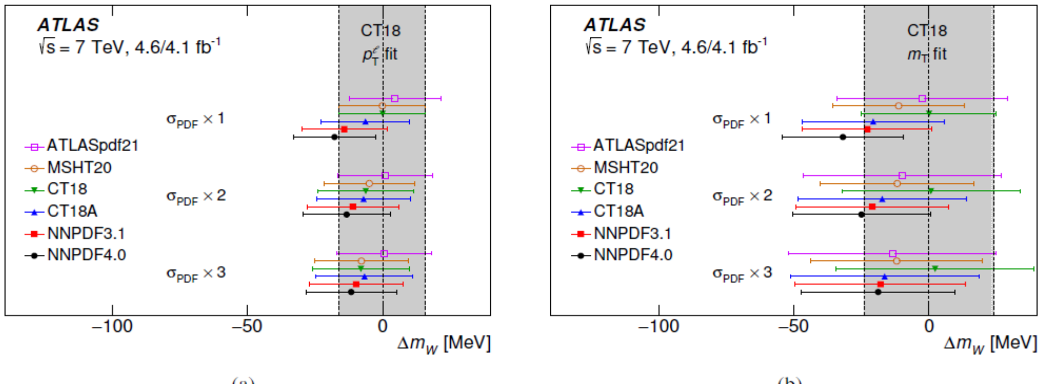
In that case the value of m_W would differ from its SM value
And indeed that is what we saw in the 2022 CDF measurement from FermiLab

However, this has been checked at the LHC.

The plot shows the latest measurements from ATLAS and CMS (2024).
A major contribution to the uncertainty of these measurements is the PDF uncertainty, which comprises half the uncertainty in both cases. LHC uses p-p not p-pbar and its kinematic reach is such that most collisions producing W are sea-quark collisions.

What is our problem?

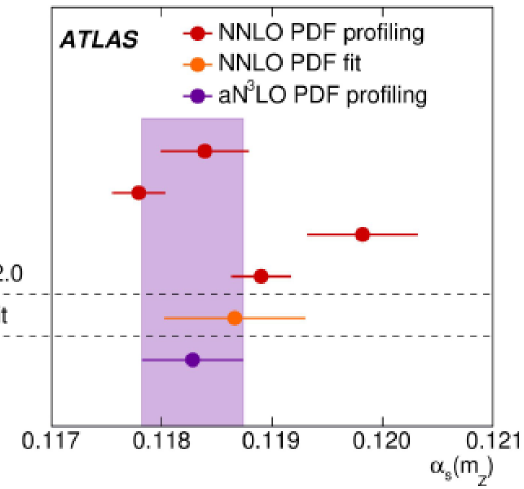
ATLAS Mw



CMS Mw

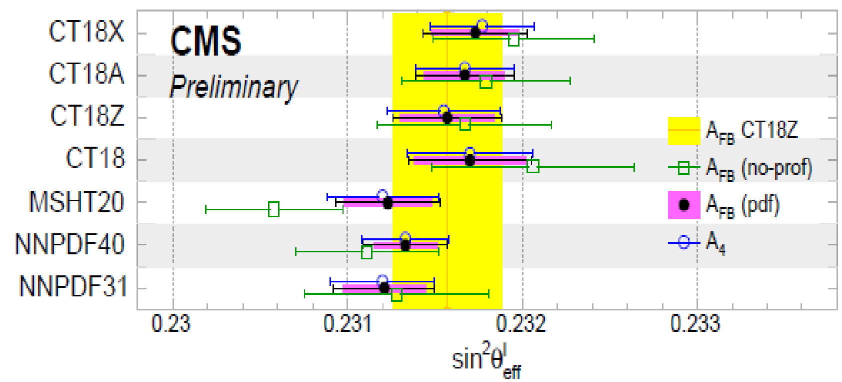
PDF set	Extracted m_W (MeV)	
	Original σ_{PDF}	Scaled σ_{PDF}
CT18Z	$80\,360.2 \pm 9.9$	
CT18	$80\,361.8 \pm 10.0$	
PDF4LHC21	$80\,363.2 \pm 9.9$	
MSHT20	$80\,361.4 \pm 10.0$	$80\,361.7 \pm 10.4$
MSHT20aN3LO	$80\,359.9 \pm 9.9$	$80\,359.8 \pm 10.3$
NNPDF3.1	$80\,359.3 \pm 9.5$	$80\,361.3 \pm 10.4$
NNPDF4.0	$80\,355.1 \pm 9.3$	$80\,357.0 \pm 10.8$

ATLAS $\alpha_S(M_Z)$



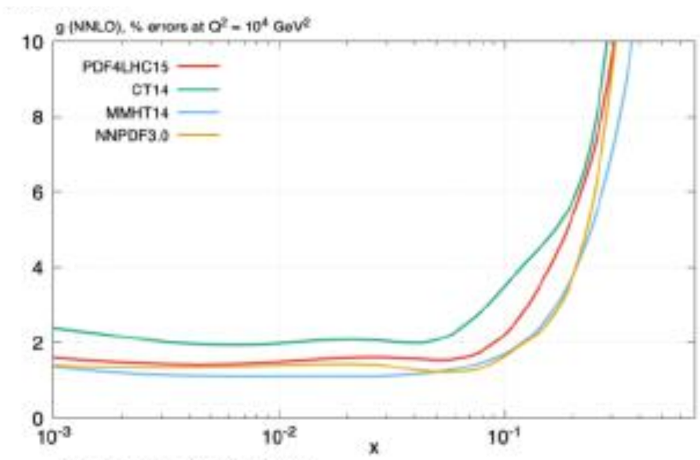
PDF set	$\alpha_s(m_Z)$	PDF uncertainty
MSHT20 [37]	0.11839	0.00040
NNPDF4.0 [84]	0.11779	0.00024
CT18A [29]	0.11982	0.00050
HERAPDF2.0 [65]	0.11890	0.00027

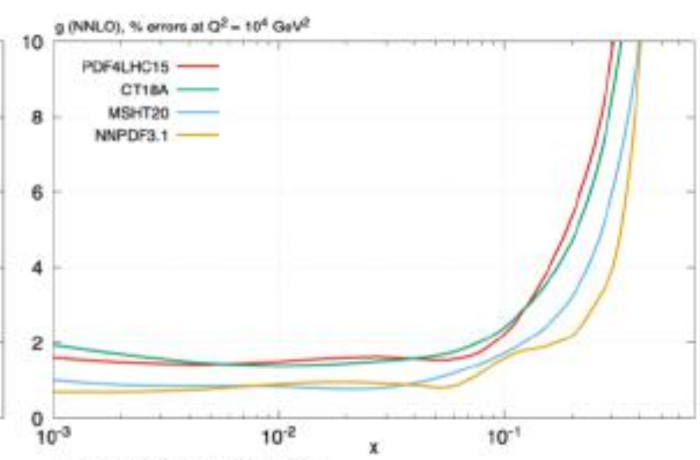
CMS $\sin^2\theta_W$



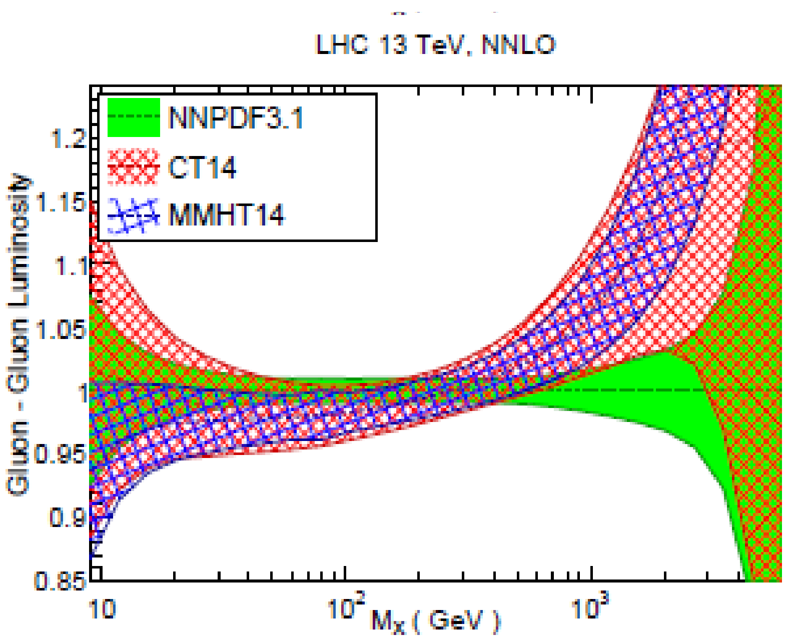
Has there been progress in recent years?

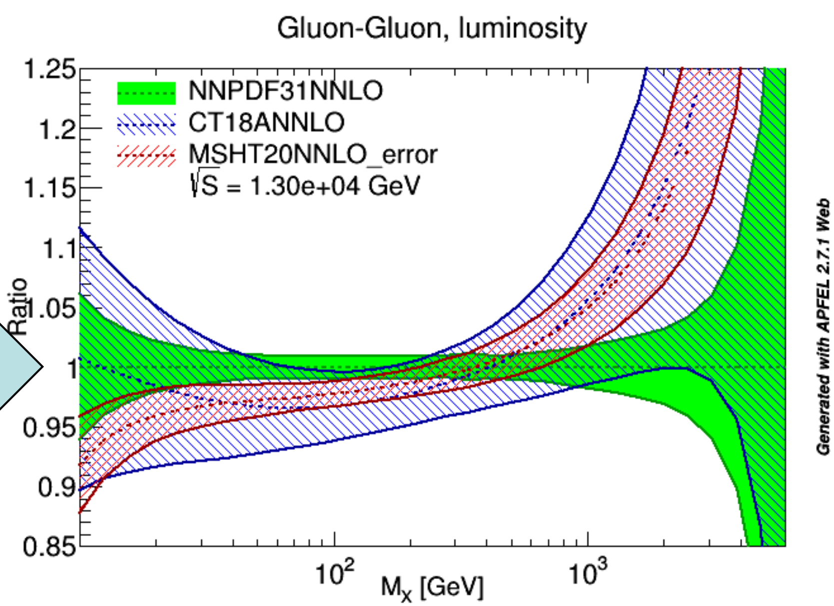
Gluon Uncertainties





As the uncertainties of each individual PDF decrease with the input of more information, the divergence of the PDFs from each other has increased

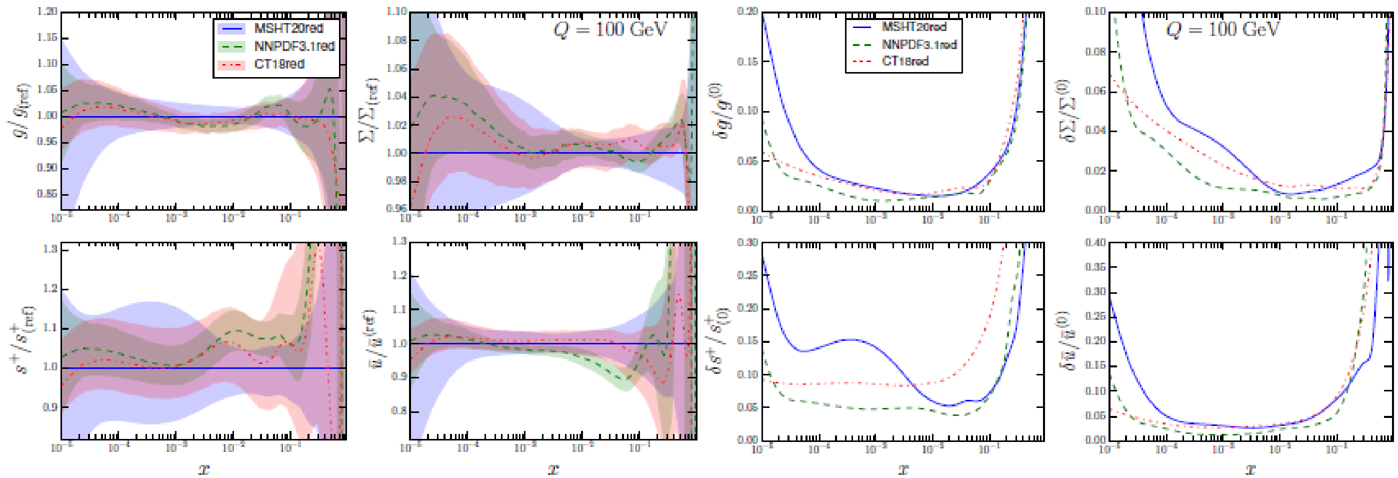




The PDF4LHC group makes combinations of the PDFs from the three main fitting groups NNPDF, CT and MSHT (arxiv:2203.05506)

First try to understand differences by using a common data set and common settings for heavy quark masses and alphas

- Use fits to reduced common datasets and common theory settings.



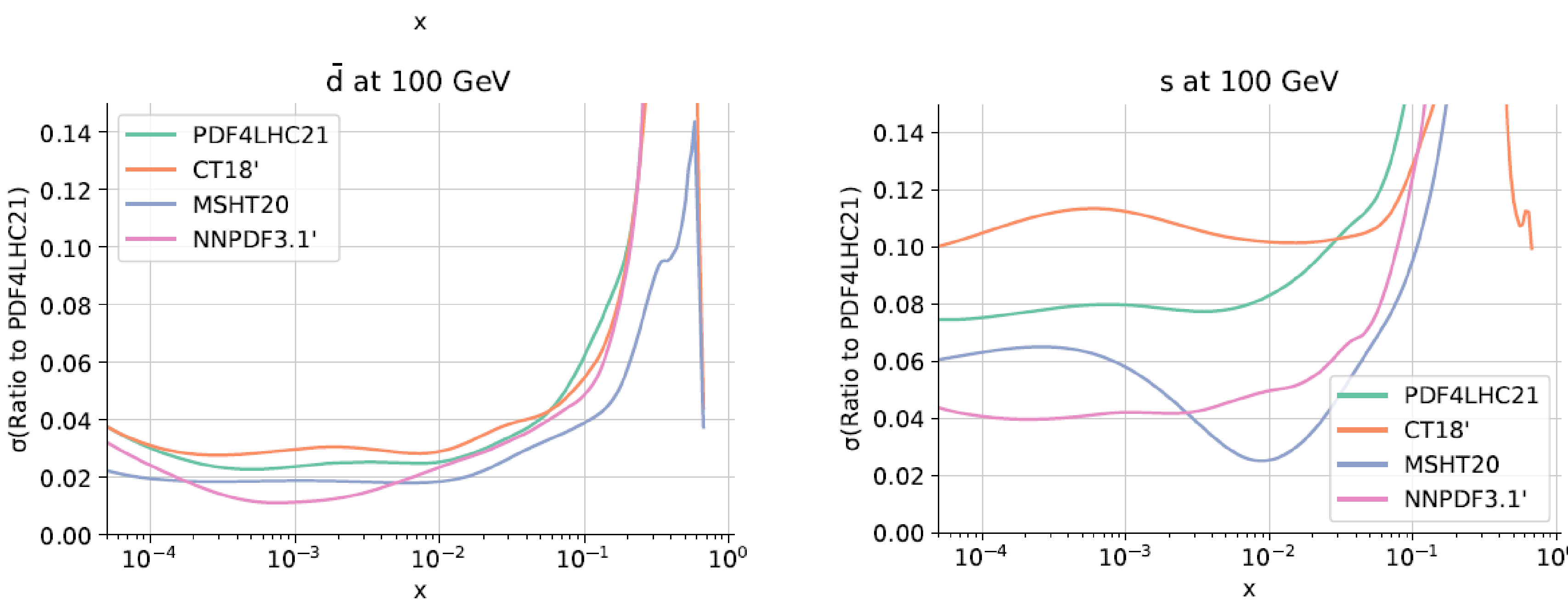
- Very good agreement within uncertainties, including gluon.
- Similar size uncertainties in data regions, differences outside this, reflecting remaining methodological and other choices.
- Agreement much improved relative to global PDFs.
- Same data and theory settings → consistent PDFs. Smaller remaining differences, e.g. in errors, reflect methodological choices.

BUT It is not recommended to use these reduced fits, greater consistency does not mean greater accuracy—the methodological differences in the main fits are there for a reason!

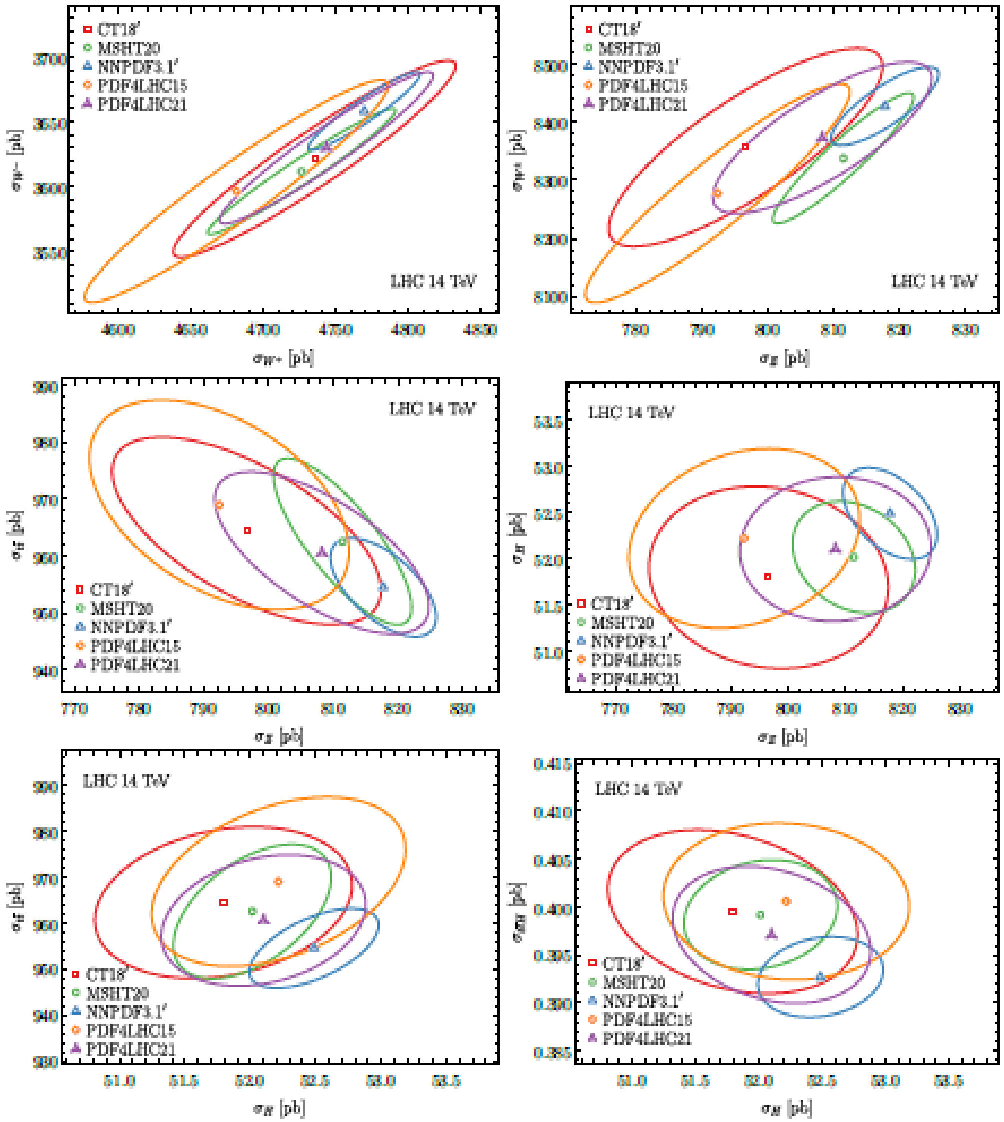
PDF4LHC21 compromise combination uses modified CT (HQ masses) and NNPDF3.1 (closer to common data set) but does not strive for common methodology, or to determine inter-PDF set correlations (eg as requested by LHC-EW group)

PDF4LHC21 (which was published in 2022)

PDF4LHC21 actually combines variants of CT18 and NNPDF3.1 with MSHT20. Variants set heavy quark masses to a common value and have a slight difference in input data sets for NNPDF3.1.
The combination is a statistical combination without correlations between the three input PDFs. Where the three input PDFs are consistent the resulting PDF4LHC uncertainty represents an average of the the PDF set uncertainties—generally closest to, though smaller than, the largest uncertainty of the three, namely CT18. But where there are discrepancies the PDF4LHC uncertainty can be larger than those of any of the individual sets since they include the spread in the central prediction



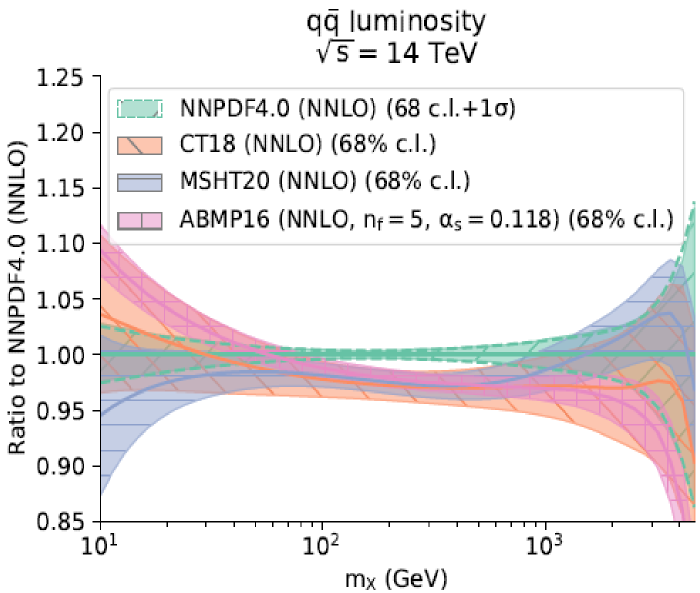
These plots represent ratios of uncertainties

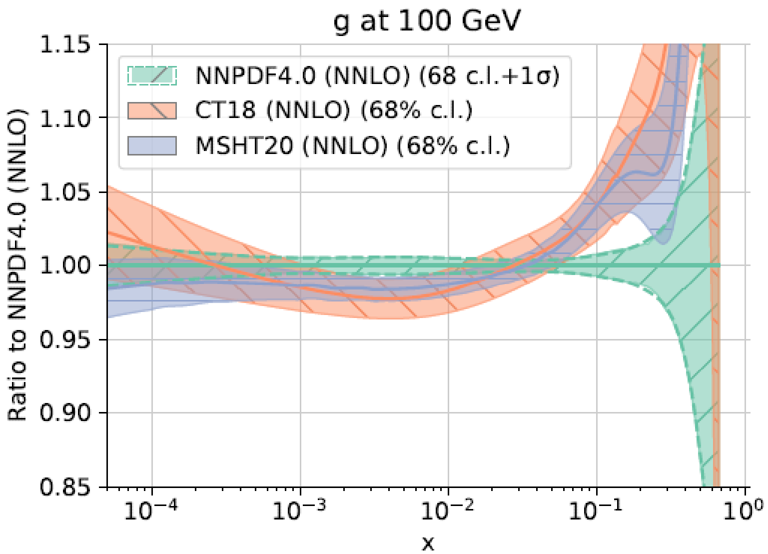


The PDF4LHC group makes combinations of the PDFs from the three main fitting groups NNPDF, CT and MSHT
The PDF4LHC15 combination has just been superseded by the PDF4LHC21 combination

There IS an improvement in uncertainty BUT this is not enough to reduce the PDF uncertainty on on LHC measurement of m_W , $\sin^2\theta_W$, $\alpha_S(M_Z)$ dramatically

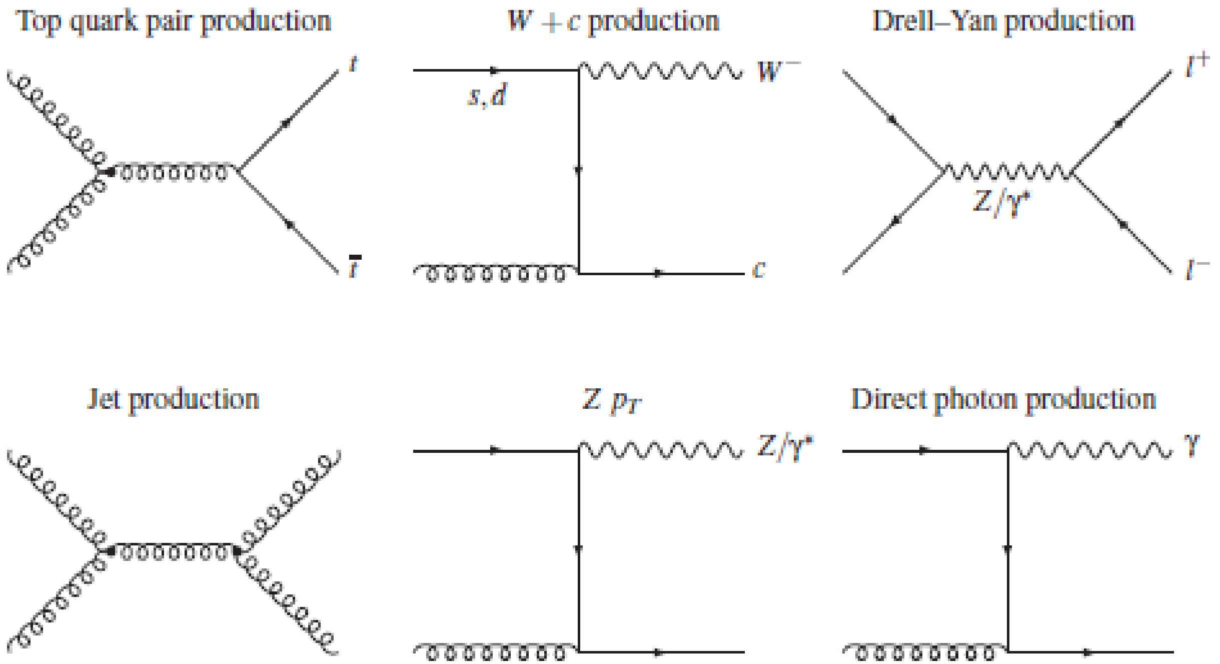
Since the issue of PDF4LHC21 there has been the new PDF set from NNPDF4.0
This has a lot of new data from the LHC and considerable decrease in uncertainty, with respect to NNPDF3.1.
BUT the improvements in uncertainty are **not so much due to the new data**, they are more **due to improvements in their procedure.**
Unfortunately decrease in uncertainty of a single PDF does not help much if there are discrepancies with other PDFs.
The uncertainty on combination of PDFs will remain higher than the uncertainty of any individual PDF set





And there are other PDFs eg ABMP, which does not use jet data
And ATLAS itself produces PDFs…

SO how can we improve PDFs?-A study of potential improvements has been made using processes which are now statistics limited, where the High-Luminosity LHC (HL-LHC) should help



Pseudo-data is generated for these processes assuming luminosity of 3 ab $^{-1}$ for CMS and ATLAS and 0.3 ab $^{-1}$ for LHCb

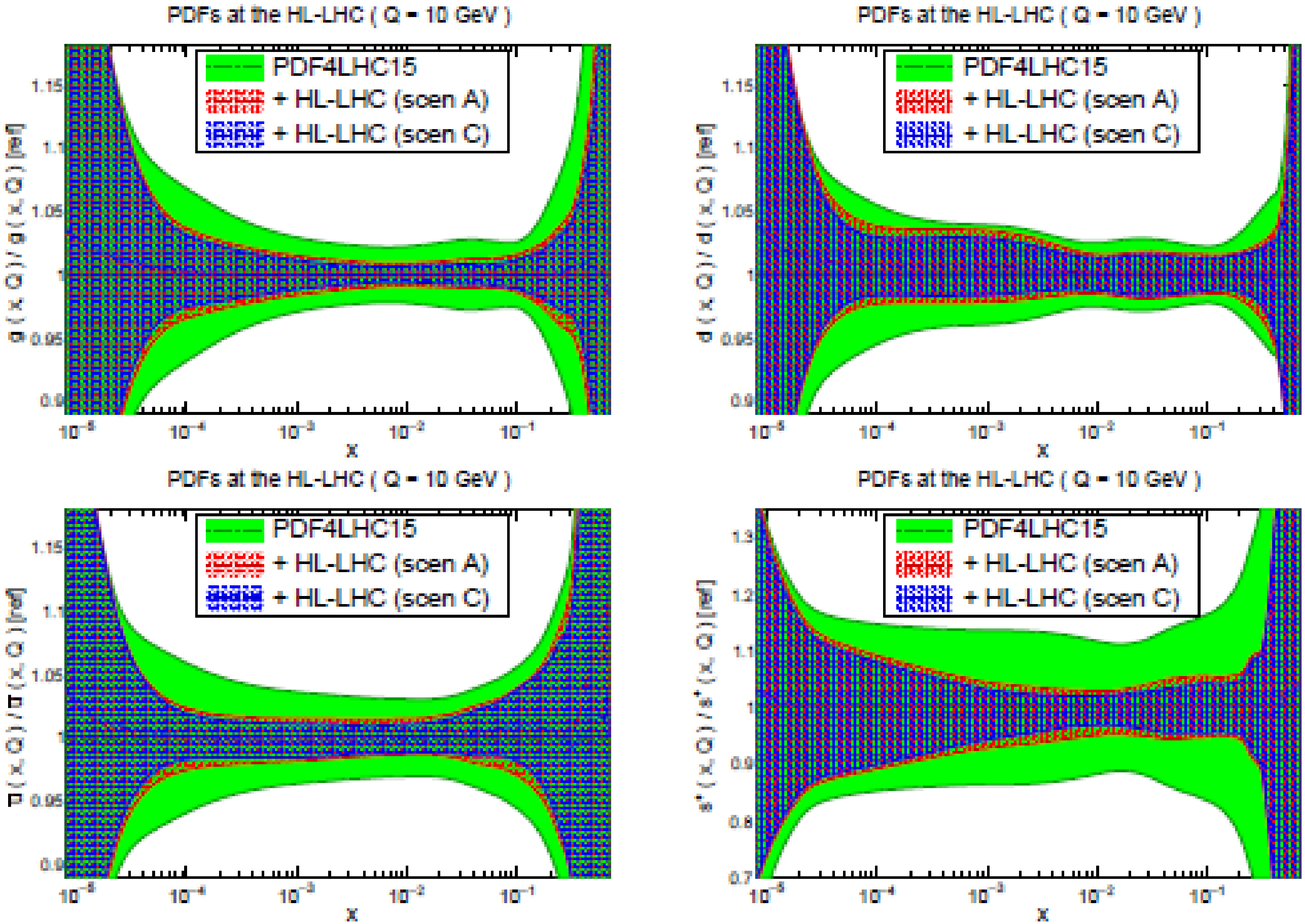
Pessimistic and Optimistic assumptions are made about systematic uncertainties based on experience with real data

Both about the effect of correlations-- typically, $f_{corr} = 1, 0.25$

And about possible reduction in uncertainty typically, $f_{red} = 1, 0.4$

This is about as good as you can do with pseudo-data but let's not forget that this is a somewhat ideal situation

So we see potential improvements in the PDFs at the HL-LHC

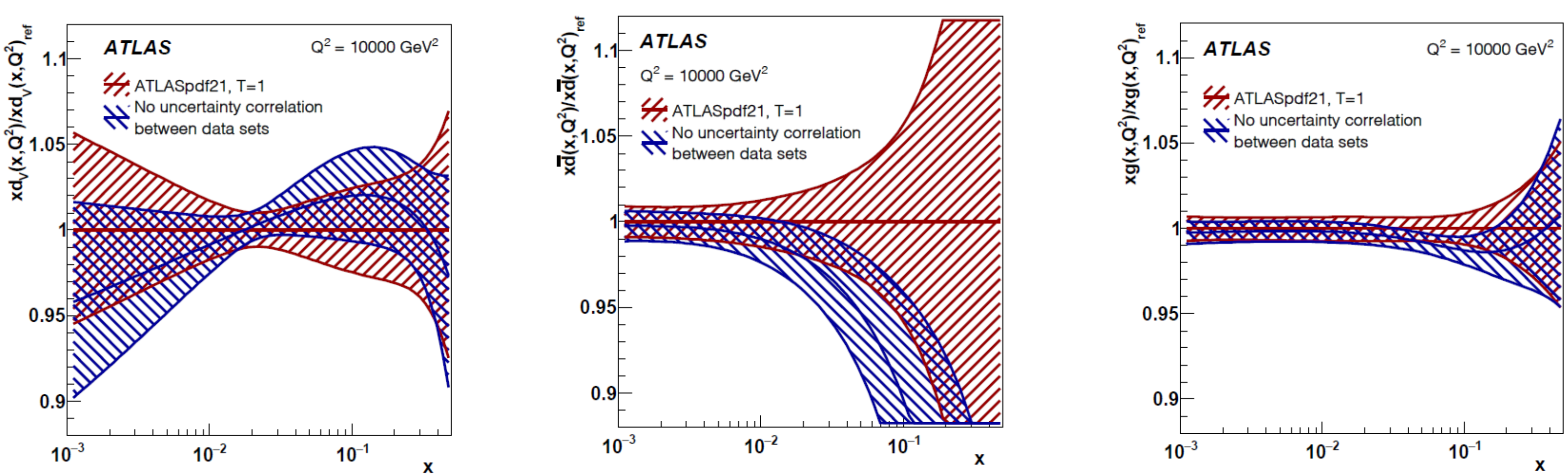


Where scenario A is pessimistic and scenario C is optimistic
--Such improvements could give up to a factor 2 improvement in the PDF uncertainty on something like m_W

But are we being a little too optimistic?

One of the issues with LHC data is that realistically it involves the combination of many data sets analysed by different groups and with differing procedures for the evaluation of systematic uncertainties, which makes cross-correlating them difficult. Such correlations are not usually known/applied

Recent work by ATLAS uses many different types of LHC data, evaluating the largest correlations (arXiv:2112.11266)



The larger correlations come between data sets such as: inclusive jets, W or Z boson +jets, t-tbar in lepton+jets mode

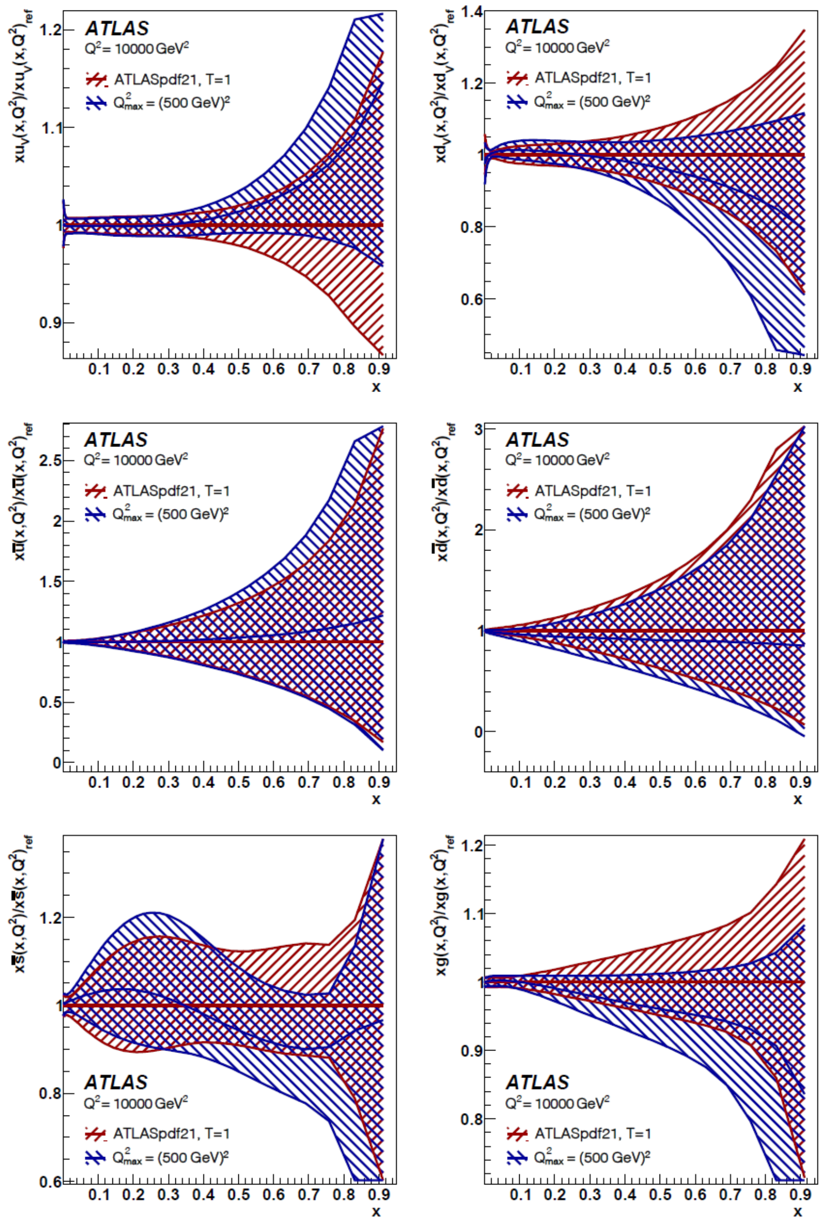
The difference between accounting for the correlations or not doing so is the shift from red to blue—shown in ratio here

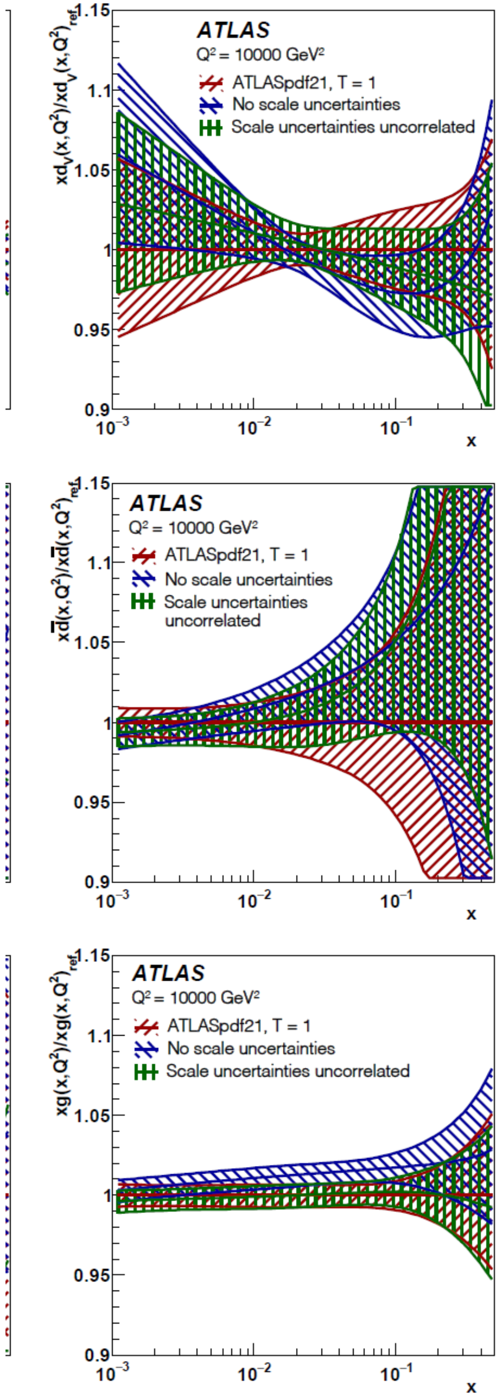
It is not a big effect, but if you want ~1% accuracy on PDFs then it matters

Also, there is a danger when fitting high scale data—such as high pT jet production– of ‘fitting away’ the very BSM effects you would like to look for, ie including the deviations BSM from SM in your PDF fit.

Thus ATLAS also cut all data at scale
$Q^2 > 250000\ \text{GeV}^2$
From the fit and re-evaluated the PDFs.

This time we can only see the difference if we look at very high-x.
Note the linear x scale of the plots.
Differences only exceed 5% for x>0.5
We have little data here, but we must be vigilant.





But just before we get to the new machines..

Another issue is that PDFs are extracted at finite order, the current state of the art is NNLO

How much difference does this make?
We use the variation of uncertainties on the choice of scale for the process as a measure of the missing higher order corrections.

The natural scale for W,Z boson production is the mass of the boson. This is varied by a factor of two to evaluate the scale uncertainty.
The plots show the change in the PDFs when including or not including scale uncertainty for W, Z boson production under two assumptions:

- Scale uncertainties correlated between W and Z and between data taken at 7 and 8 TeV
- Scale uncertainties correlated between W and Z but not between data taken at 7 and 8 TeV

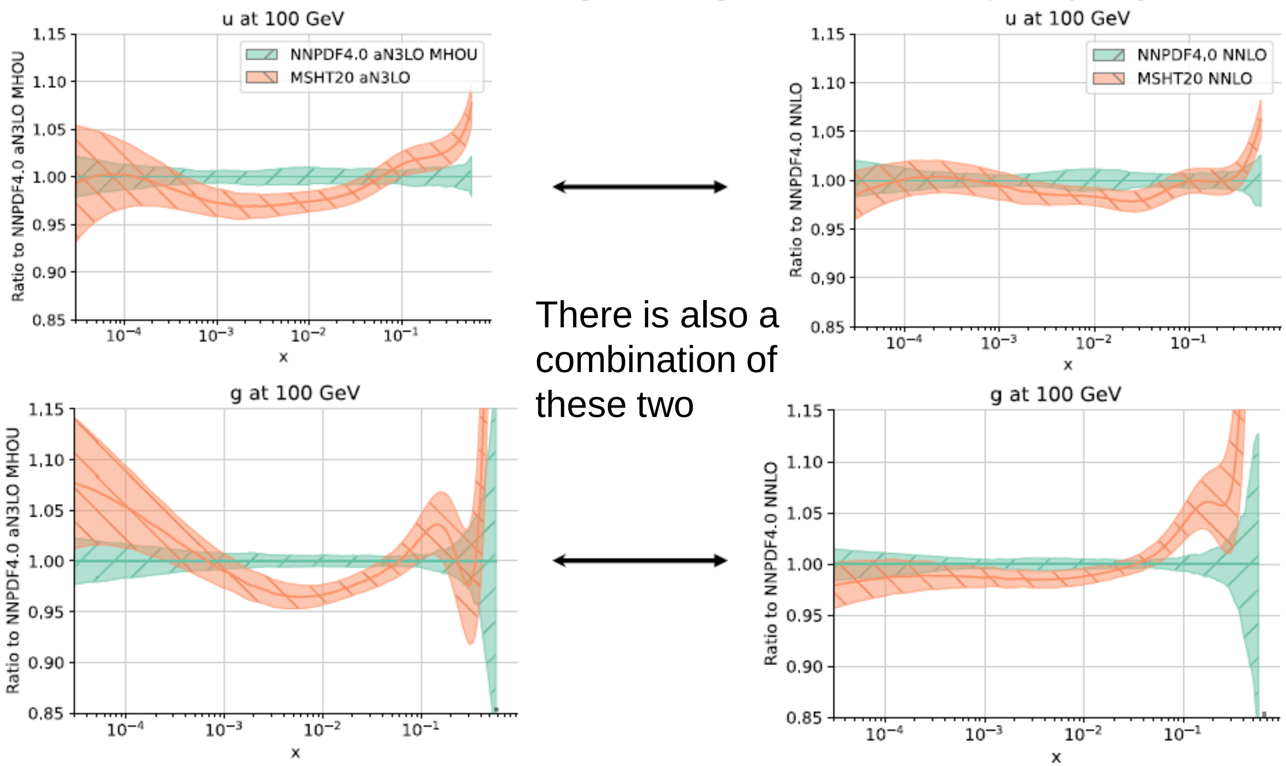
This is not a very big effect but it matters if we are striving for ultimate accuracy O(1%)

Another issue is that PDFs are extracted at finite order, the current state of the art is NNLO

We use the variation of uncertainties on the choice of scale for the process as a measure of the missing higher order corrections.. MHOUs are now included in some PDFs notably NNPDF

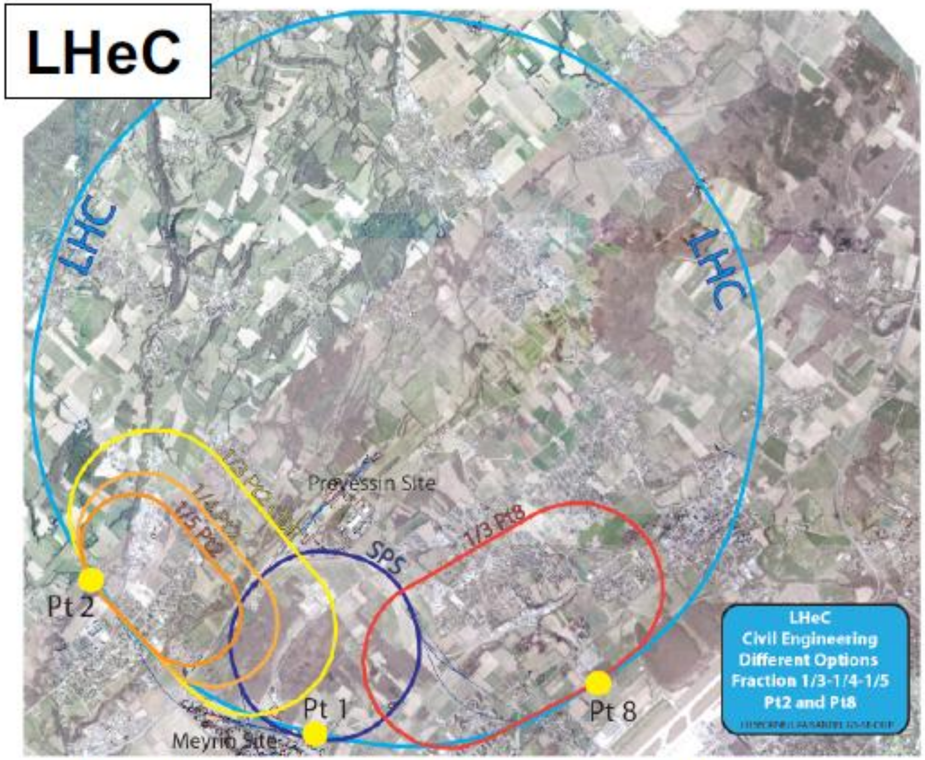
BUT recently there has been a move to N3LO..**which is not fully computed yet** so MSHT and NNPDF have given **approximate** PDFs.. The differences between NNLO and aN3LO are larger than our estimates from MHOUs led us to believe, but also the two groups do not agree very well with each other

Differences between PDFs at aN3LO are in general larger than at NNLO, especially for gluon



SO HOW MAY WE ACTUALLY DO MUCH BETTER?

One thing that has always made PDF fitters agree better with each other is better data. We need New DIS machines…



(M Klein, Rencontre du Vietnam, Sept 2017)

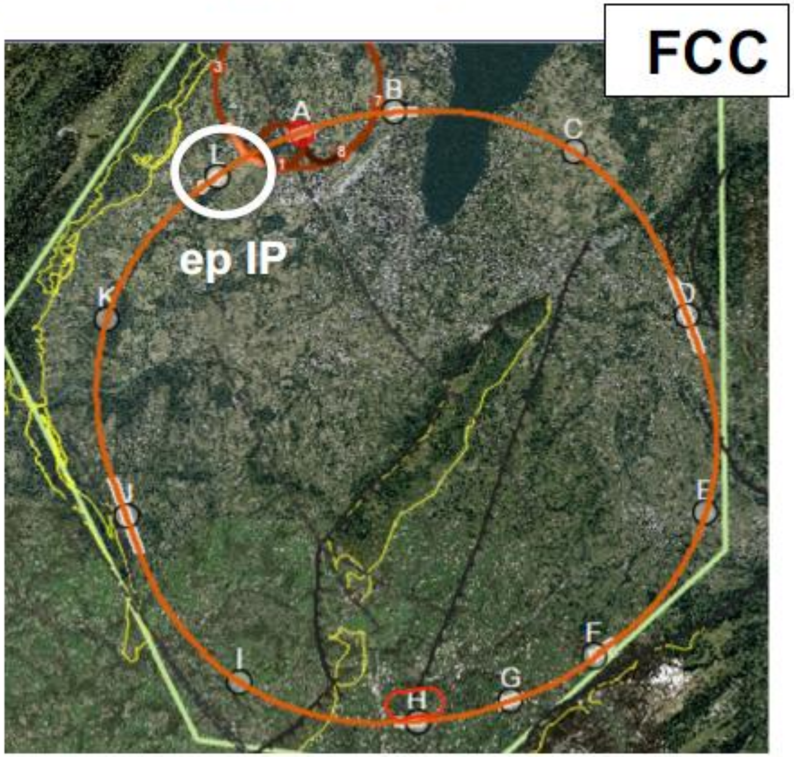
LHeC may be considered as a 'bridge' project to any possible future FCC arXiv:2503.17727

LHeC and **FCC-eh**

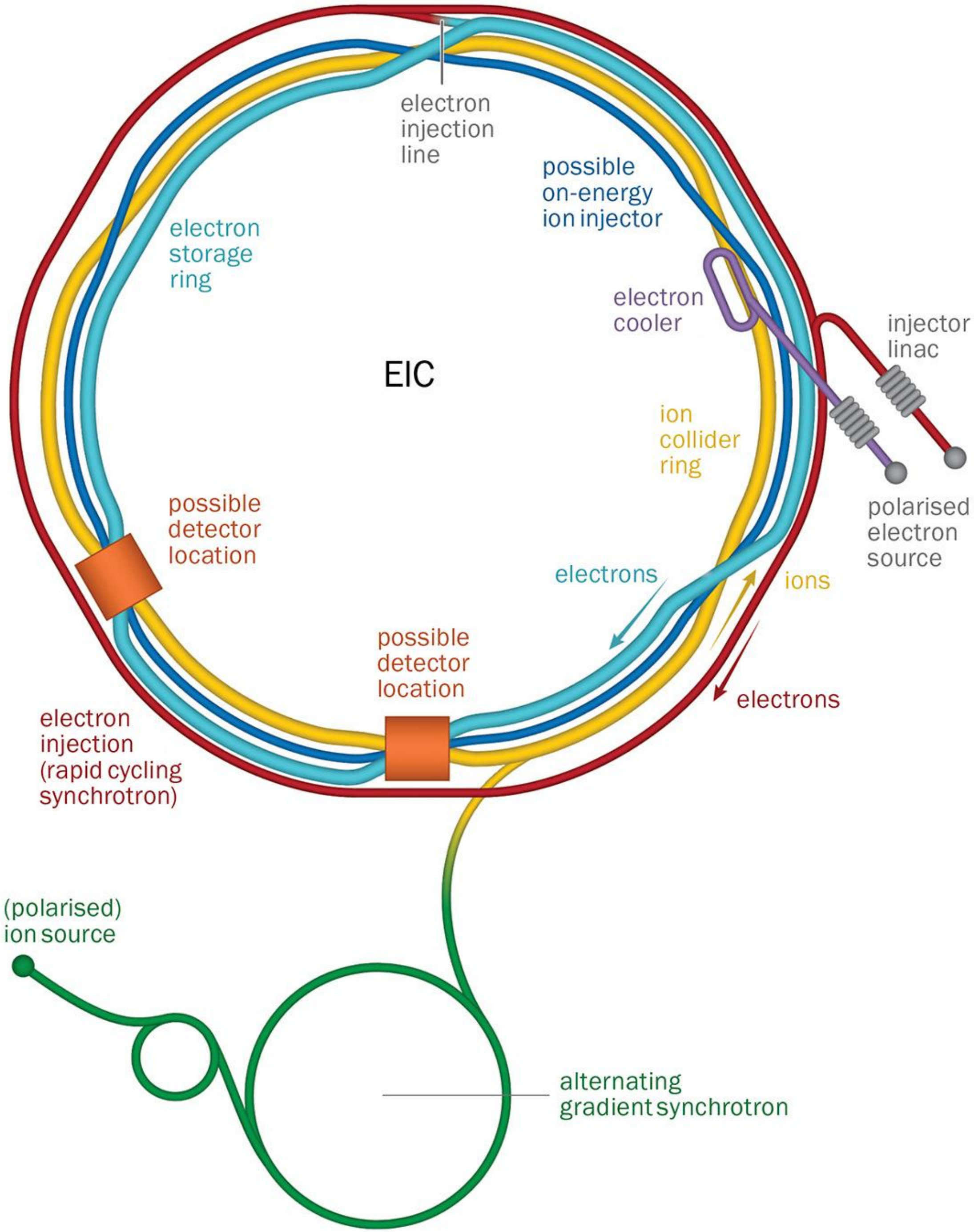
energy recovery LINAC

e-beam: 60 GeV

$L_{int} \to 1\ ab^{-1}$



LHeC (FCC-eh) complementary to, synchronous with, **HL-LHC** (**FCC**)



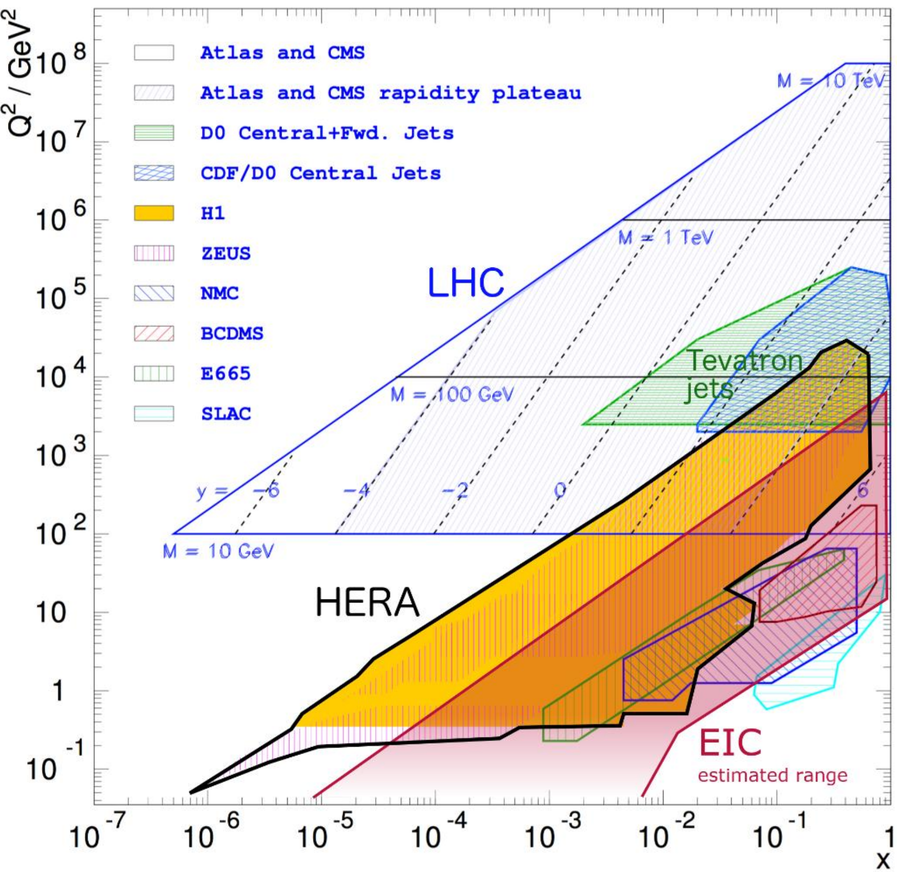
The Electron Ion Collider at Brookhaven

AND this one WILL definitely happen !

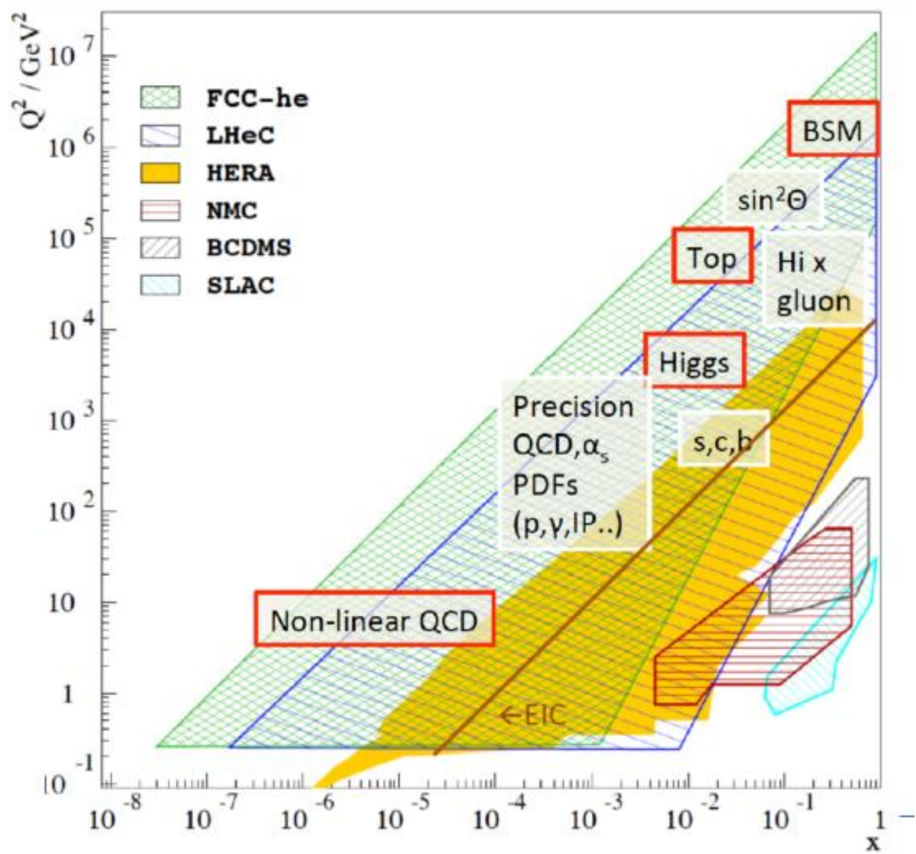
One of the issues with LHC data is that realistically it involves the combination of many data sets analysed by different groups and with differing procedures for the evaluation of systematic uncertainties, which makes cross-correlating them difficult.

A new DIS machine would instead give a consistent data set across a wide x, Q^2 range

Consider the kinematic reach of each of these



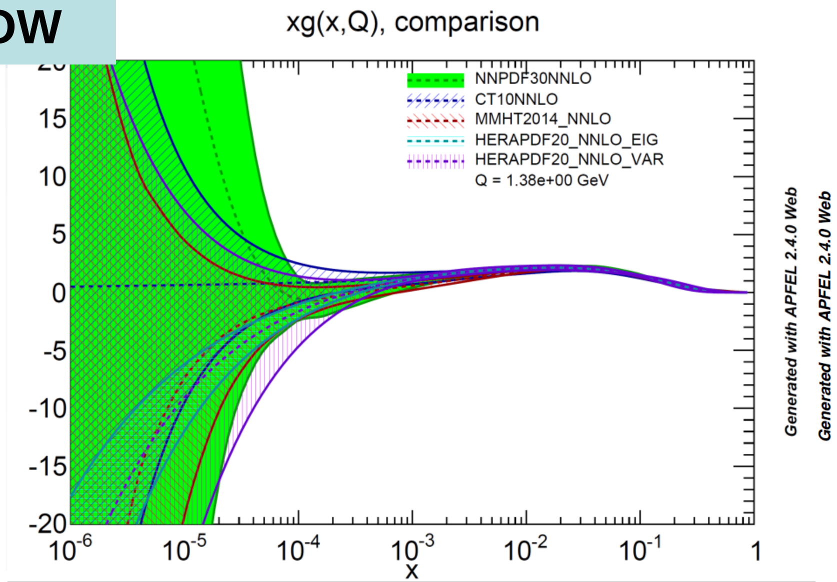
The EIC will reach higher x than HERA could reach



The proposed LHeC and FCC-eh machines reach lower x than HERA could reach

The LHeC would extend sensitivity to gluon and sea at low x

NOW



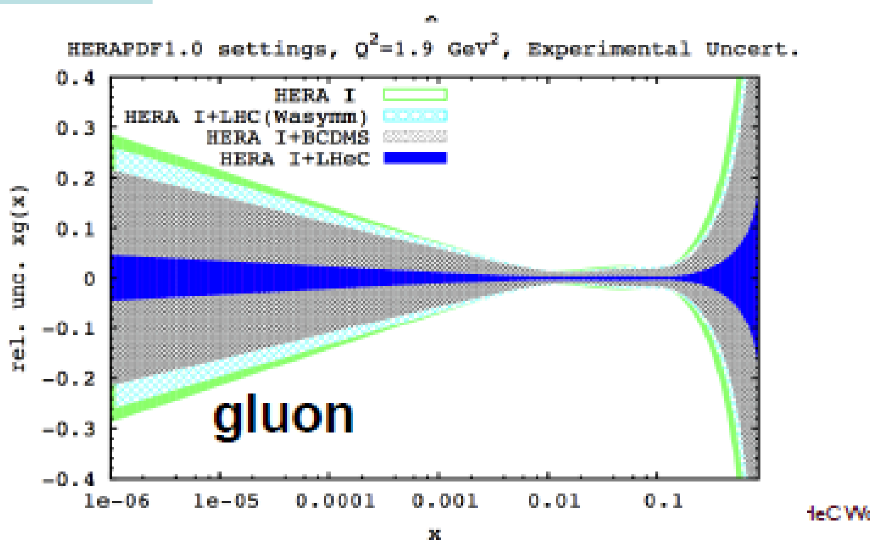
HERA sensitivity stops at x > 5 10^{-4}

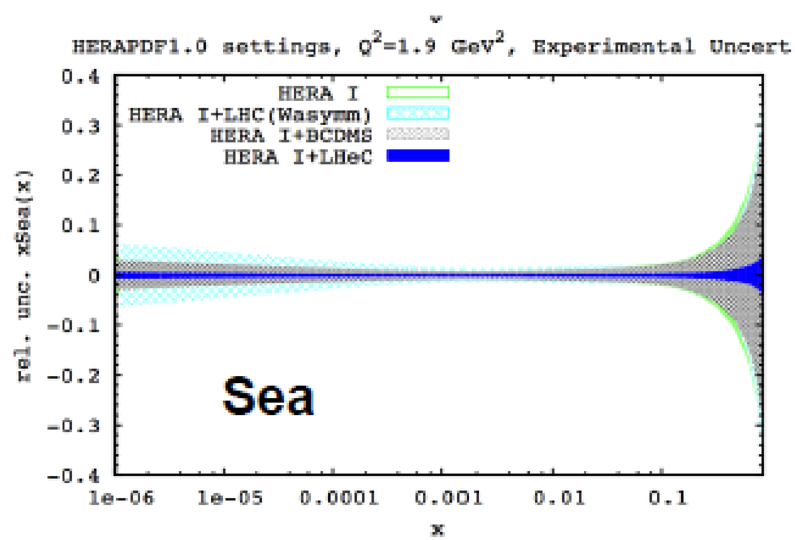
Below that uncertainties depend on the parametrisation

LHeC goes down to 10^{-6}

- FL measurement will also contribute
- Explore low-x QCD DGLAP vs BFKL or non-linear evolution
- Important for high energy neutrino cross sections – Auger etc.

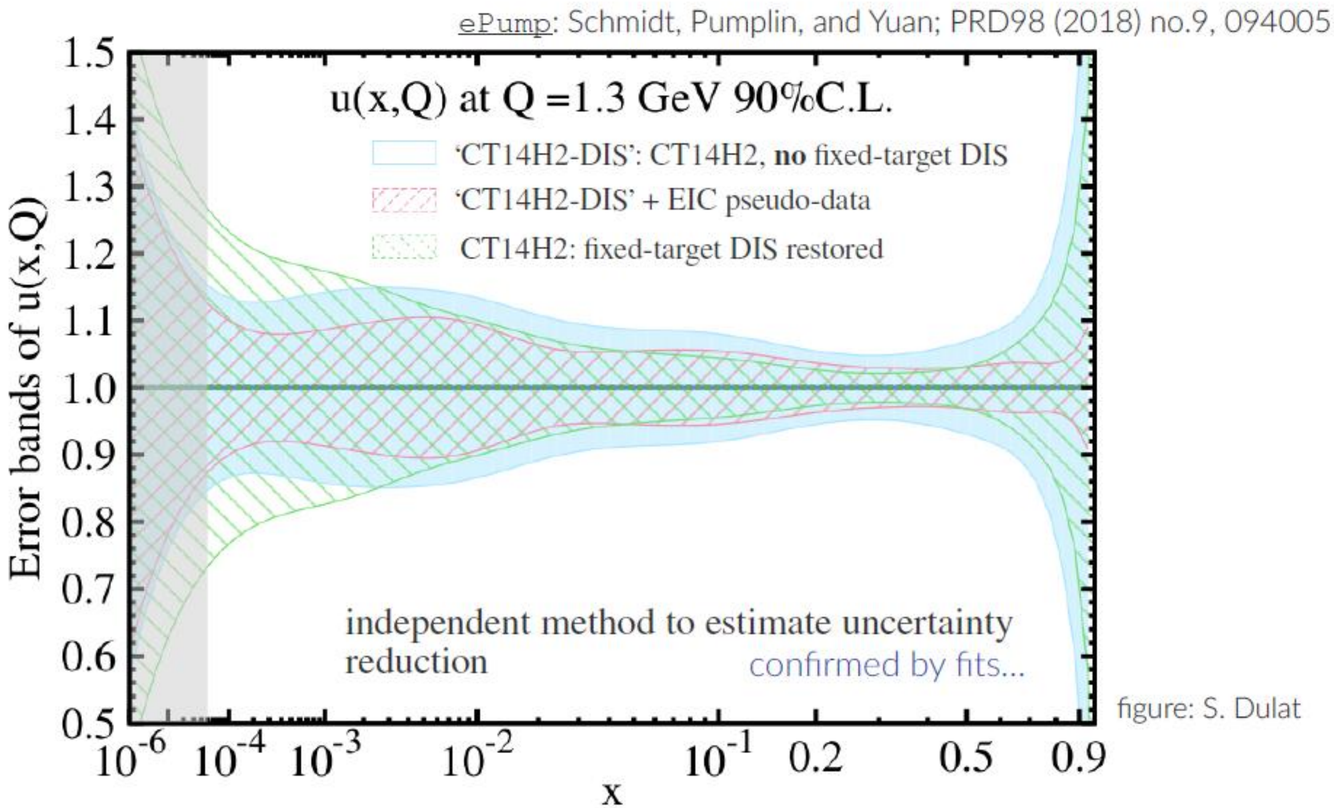
THEN





Consider the EIC

This time the big impact is at high x



- **inclusive EIC may surpass total impact of fixed-target DIS in modern fits**

And this is only to mention the impact on Parton Distribution Functions which are afunction of the fractional longitudinal momentum of the parton. The EIC is about much more, transverse momentum distributions, spin, generalized parton distribution functions etc

Summary

What have we learnt about/from the deep structure of the proton in the last 40 years?

•It's full of partons- quarks, antiquarks and gluons
•The harder you look the more of them you see
•Established Quantum-Chromo-Dynamics as the theory of the strong interaction
•Measurement of its essential parameters: Parton Distribution Functions, the strong coupling parameter, $\alpha_s(M_Z)$ and the running of α_s with scale
•Sets the background for discovery physics at the LHC

But the PDF uncertainties need to be reduced

•Precision PDFs are needed reduce the background in searches for BSM physics- both at the LHC and any FCC-hh

•They are also needed for precision measurements of SM parameters, where small deviations from SM values may indicate BSM physics

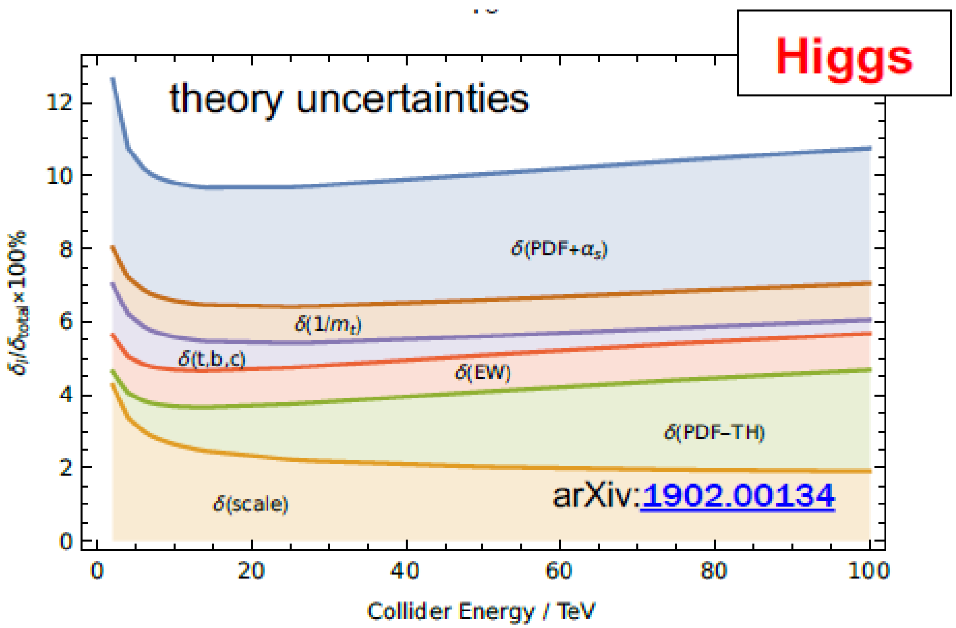
•The measurements from the High Luminosity –LHC should improve on our current knowledge

•But a dedicated Deep Inelastic Scattering machine such as an LHeC/FCCeh or EIC could do better --- and EIC will definitely happen!

- And I have only spoken about the longitudinal momemtum distributions of the partons, the EIC will also probe transverse momentum and spin distributions
- And we should learn more about QCD beyond the DGLAP formalism– gluon saturation, BFKL resummation.

But there is no time for that today!

extras



Higgs
theory uncertainties
$\delta(\mathrm{PDF}+\alpha_s)$
$\delta(1/m_t)$
$\delta(\mathrm{t,b,c})$
$\delta(\mathrm{EW})$
$\delta(\mathrm{PDF-TH})$
$\delta(\mathrm{scale})$
arXiv:1902.00134
$\delta_i/\delta_{\mathrm{total}}\times 100\%$
Collider Energy / TeV

Consequence of uncertainty in the high-x gluon?-one example

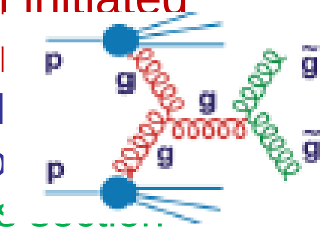
Many interesting processes at the LHC are gluon-gluon initiated

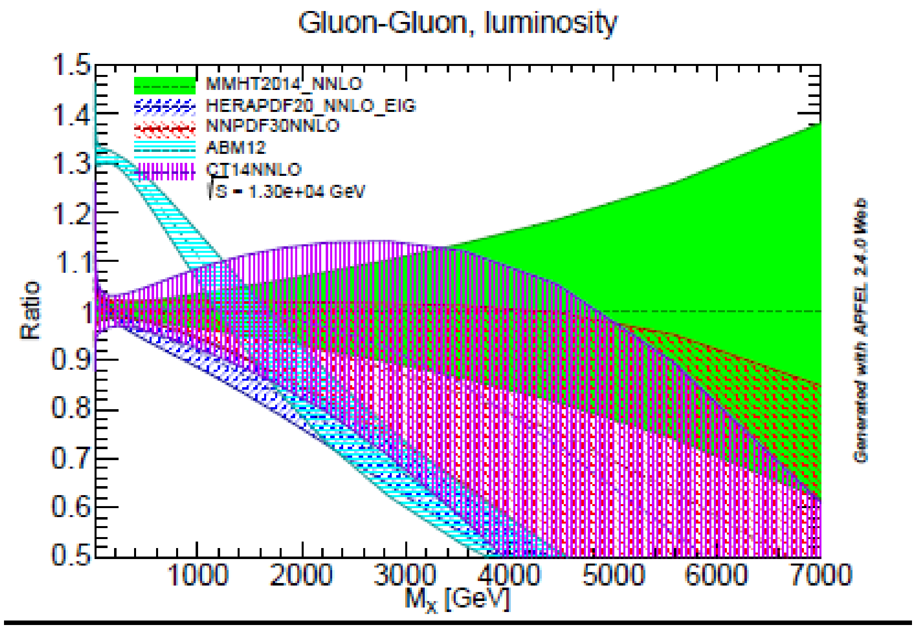
…BSM processes like gluon-gluon → gluino-gluii

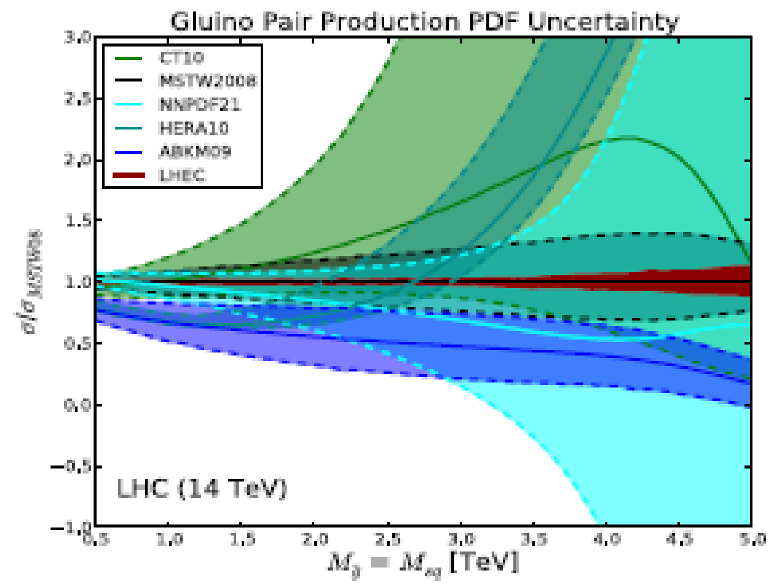
And the high-scale needed for this involves the high-x gl

The gluon-gluon luminosity at high-scale is not well-kno

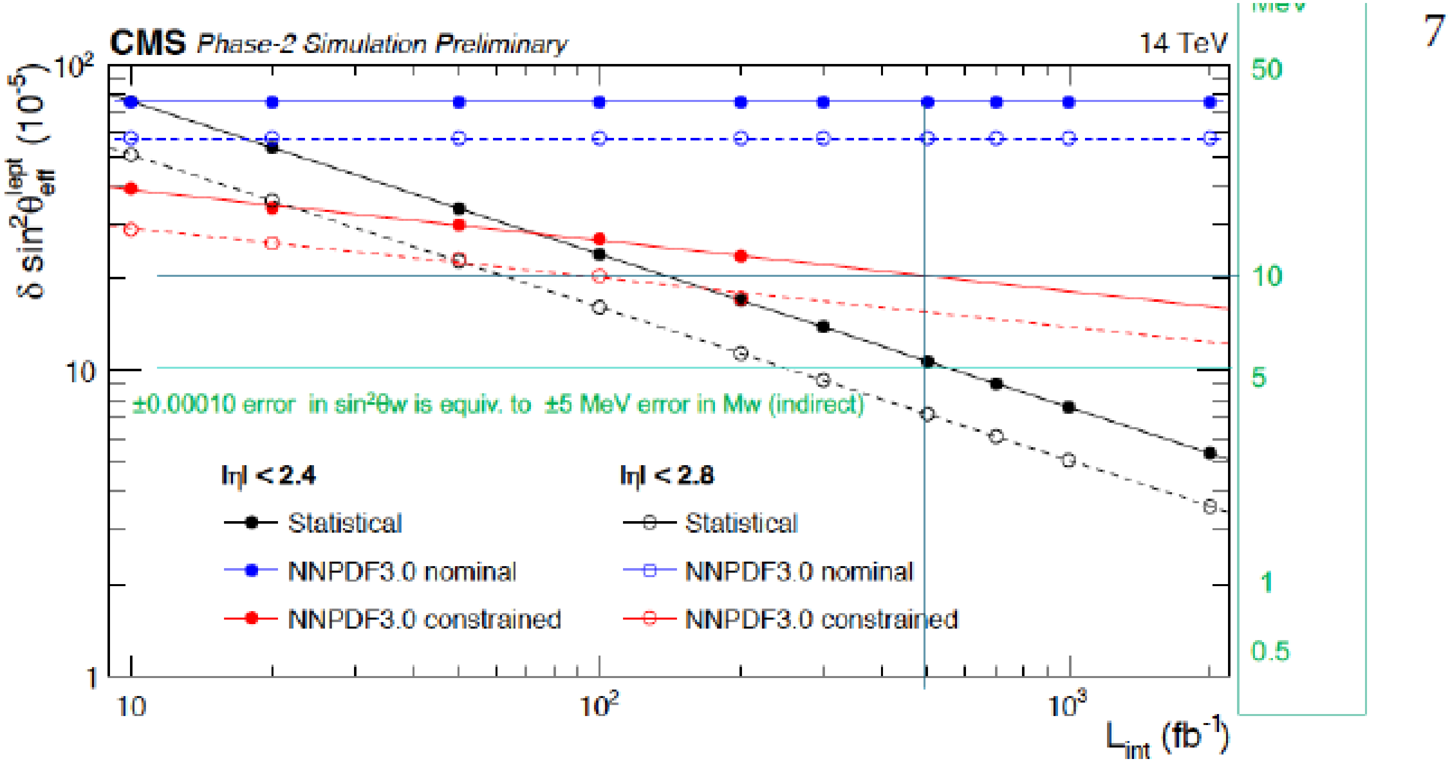
This leads to uncertainties on the gluino pair production cross







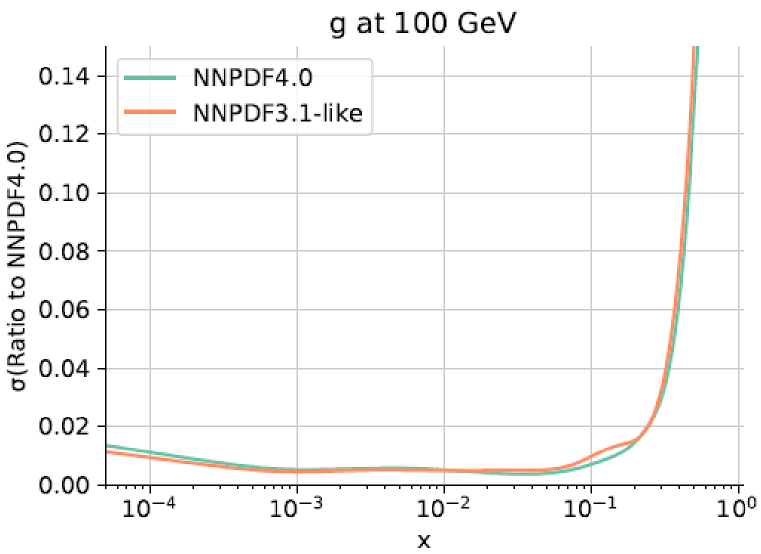
A further example is the uncertainty on the electro-weak mixing angle $\sin^2\theta_W$



The plot shows the projected decrease in the statistical uncertainty on $\sin^2\theta_W$ with future data

But the PDF uncertainty will not decrease much

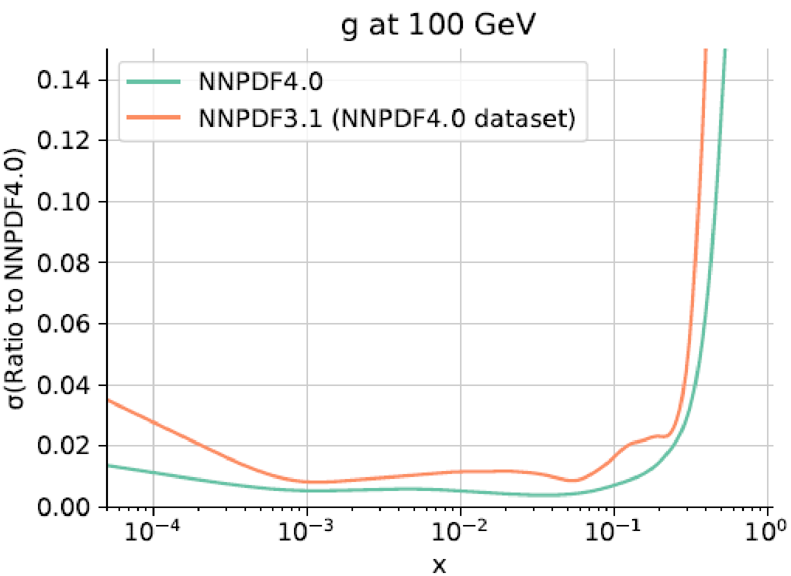
Unless some further constraints can be applied



Since the issue of PDF4LHC21 there has been a new PDF set from NNPDF4.0
This has a lot of new data from the LHC

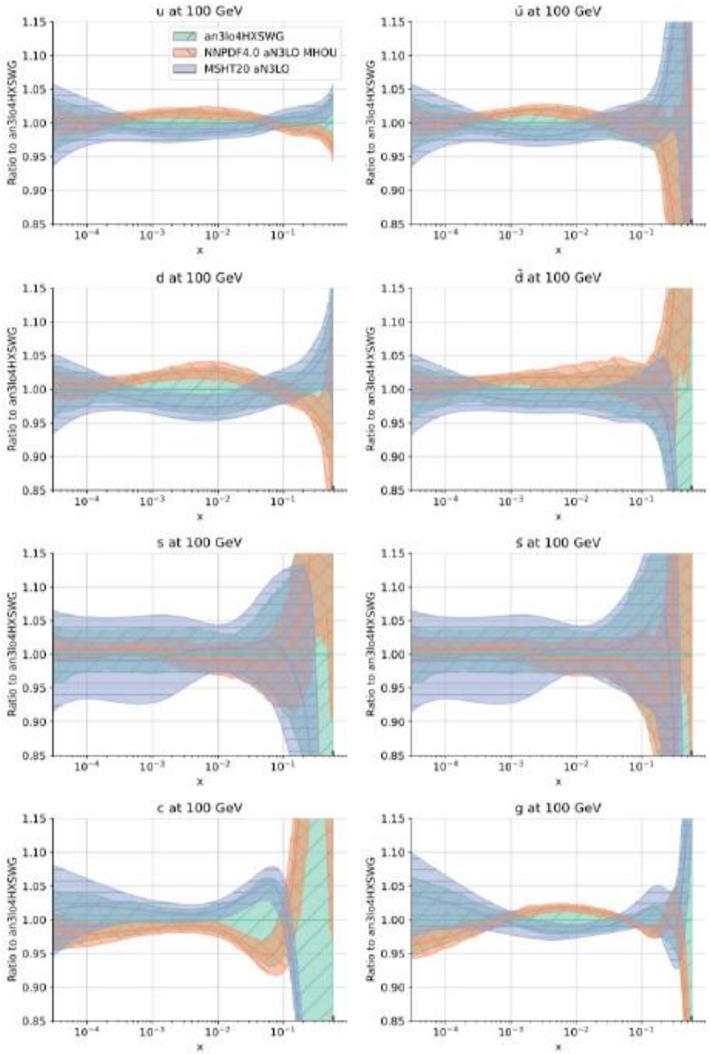
Nevertheless the improvements in uncertainty are not much due to these data, they are more **due to improvements in their procedure**

The top plot compares the uncertainties of NNPDF4.0 and 3.1 data sets using the **SAME new methodology**



The bottom plot shows the impact of the methodology on the **SAME new data** set
4.0 shows new methodology and 3.1 here shows old methodology on new data-set

There has been a lot of debate in the PDF community over the new methodology.
But if we just accept it this still does not help much if one is trying to combine with other PDFs MSHT20 and CT18 with different central values

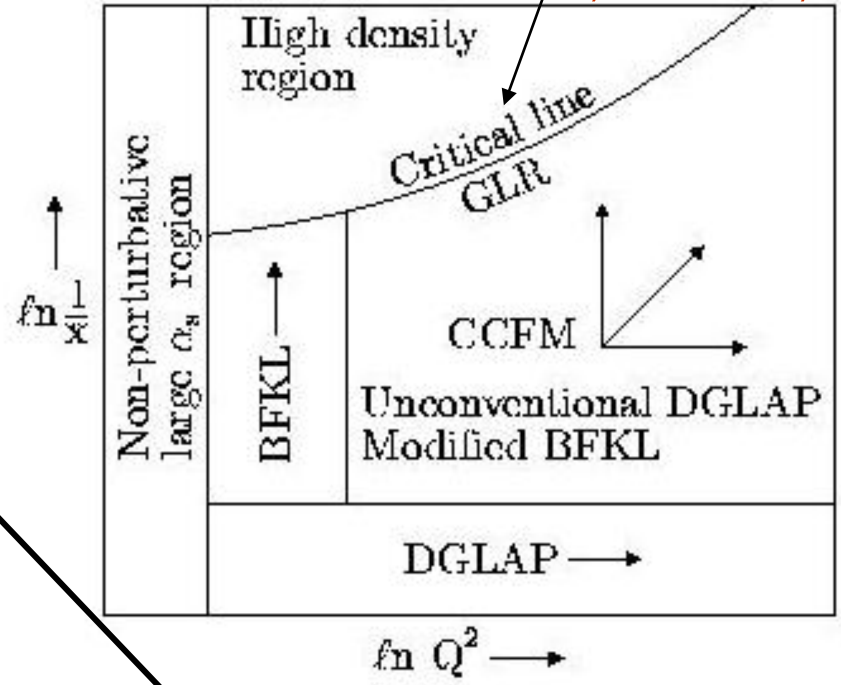


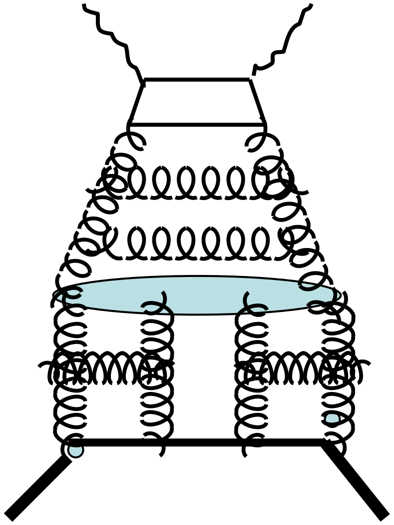
[2411.05373](https://arxiv.org/abs/2411.05373)

Paper provides a combination of the two aN3LO PDFs to produce aN3LOHXSWG; note this is not a benchmarked study, just a combination of the two aN3LO PDFs generating the same number of MC replicas for each.

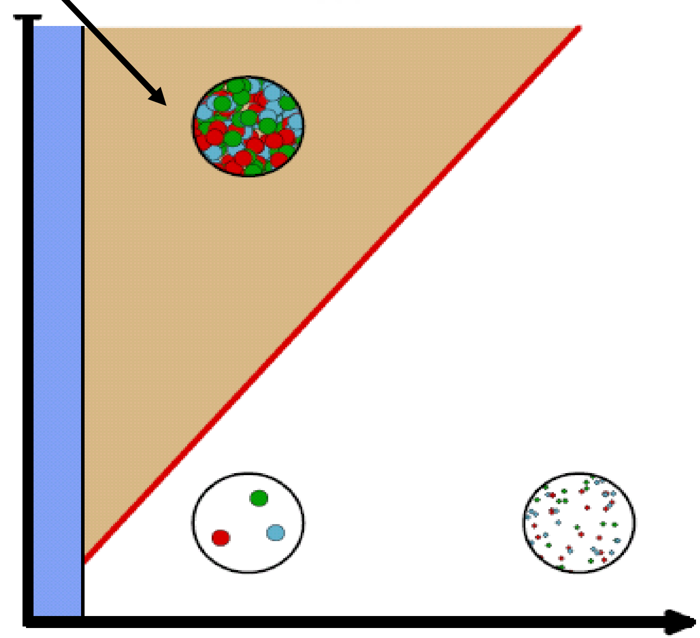
The strong rise in the gluon density at small-x leads to speculation that there may ALSO be a need for non-linear terms in the evolution equations?-
gluon recombination $gg \to g$ competes with gluon splitting $g \to gg$
And the gluon distribution might saturate?

Colour Glass Condensate, JIMWLK, BK





Non-linear fan diagrams could be imprtant at low x



But is there any experimental evidence for the need for such extensions to the formalism?

When you look at the sea and the gluon deduced from the DGLAP formalism at low Q^2 there are odd features

the gluon is no longer steep at small x – in fact its valence-like or even negative!

The problem is that we are deducing this from limited information

So far at low-x, we only use

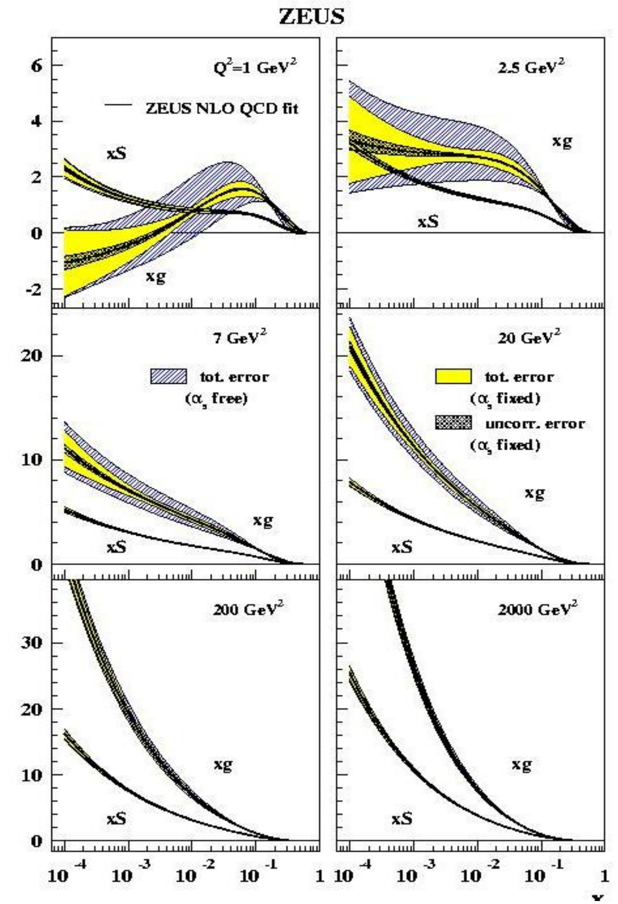
$F_2 \sim xq$ for the sea

$dF_2/d\ln Q^2 \sim Pqg\ xg$ for the gluon

Unusual behaviour of $dF_2/d\ln Q^2$ may come from

unusual gluon or from unusual Pqg- alternative BFKL evolution?. Non-linear effects?

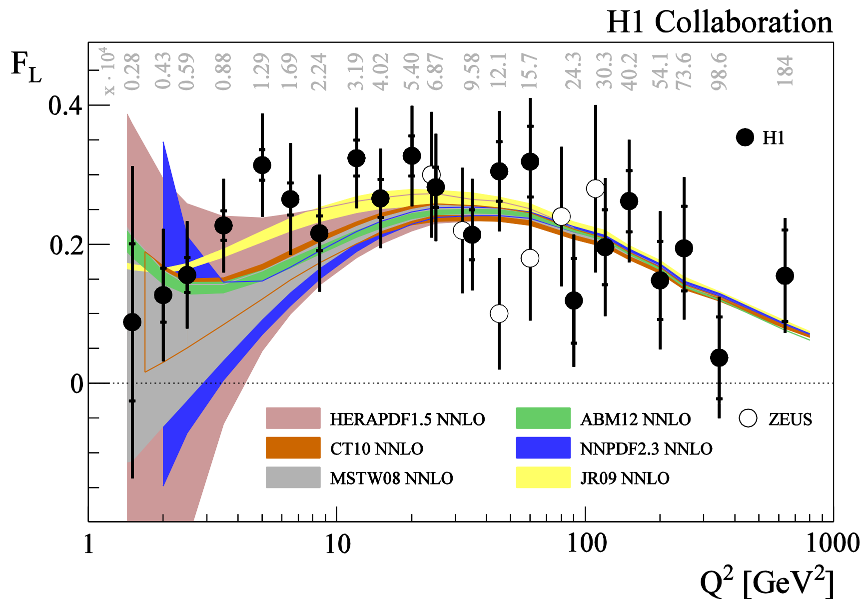
__**We need alternative ways to probe the gluon**__

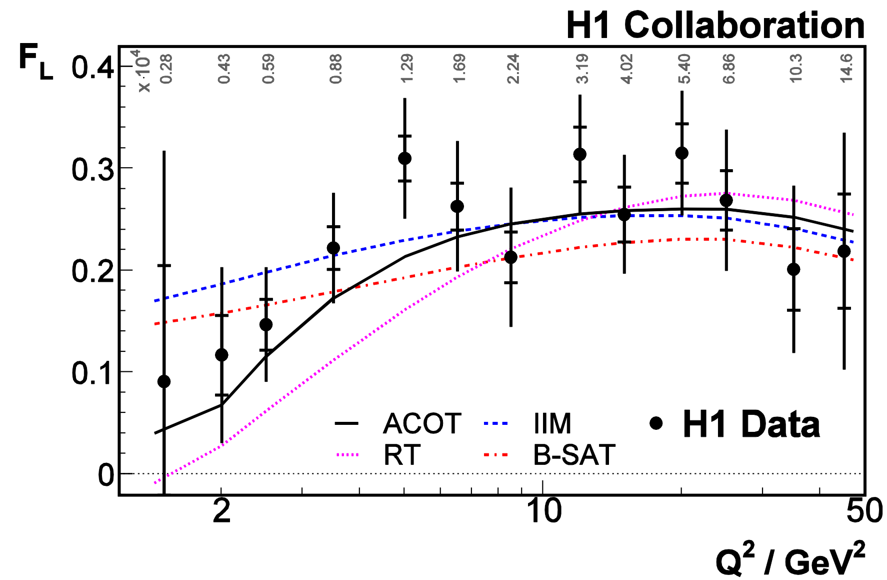


We need other gluon sensitive measurements like F_L: in NLO DGLAP FL is given by

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

And at low-x this becomes gluon dominated





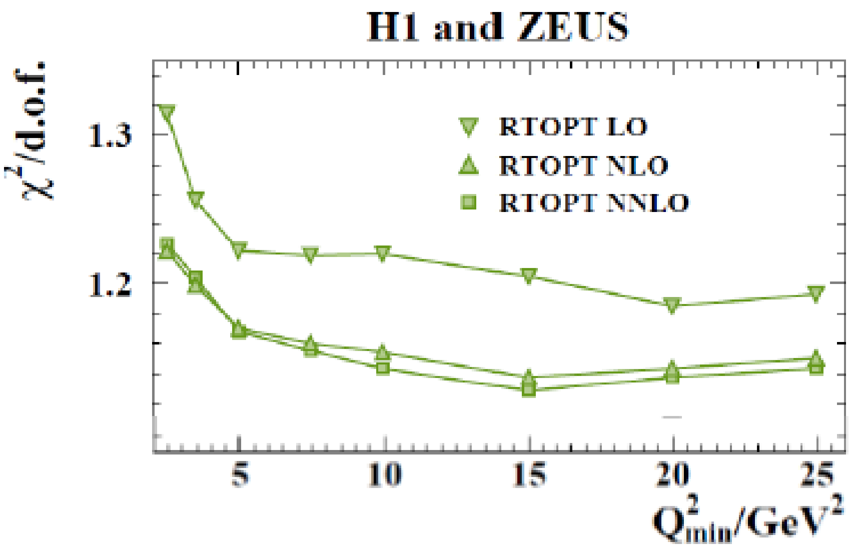
Now there are HERA measurements on F_L analysis completed in 2013: Compare to various NLO DGLAP fits

Unfortunately this is not conclusive

And compare to alternative theoretical predictions:

IIM and B-Sat are different dipole models which can accommodate non-linear effects/ saturation eg IIM colour glass condensate

There are small signs of strain in the low Q^2/low x HERA combined 2015 data

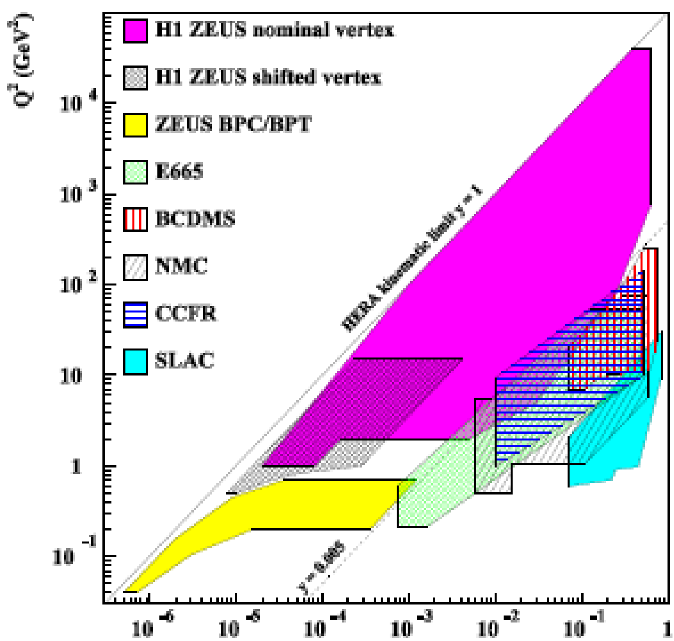


Fit χ2 deteriorates at low Q2
And NNLO is NOT better than NLO

Study two different ways of getting a better fit at low Q^2/ low-x

Adding higher twists or introducing ln(1/x) resummation

These work for $2.5 < Q^2 < 25$ GeV^2..for lower Q^2 you need something else



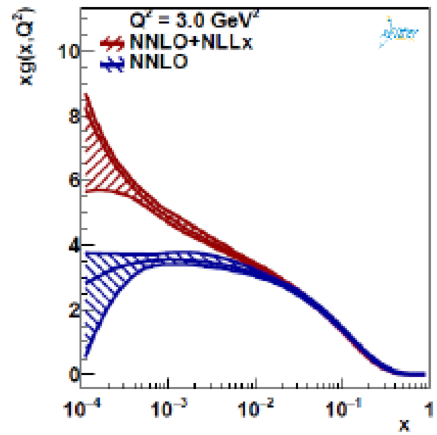
NOTE: HERA data at low Q^2 are also at low-x

7

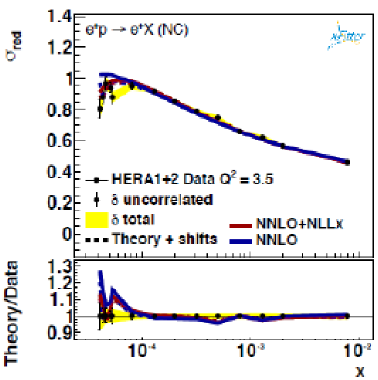
10 10 10 10 10 10 1 x

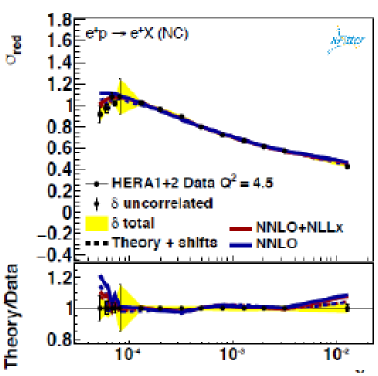
Now examine the alternative approach of ln(1/x) resummation adding terms to DGLAP splitting functions and coefficient functions – see talk of Bertone

The programme to do these High Energy Leading log resummation (HELL) has been implemented in xFitter



Comparison to data

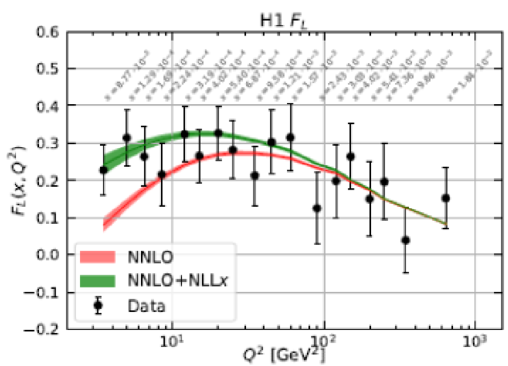


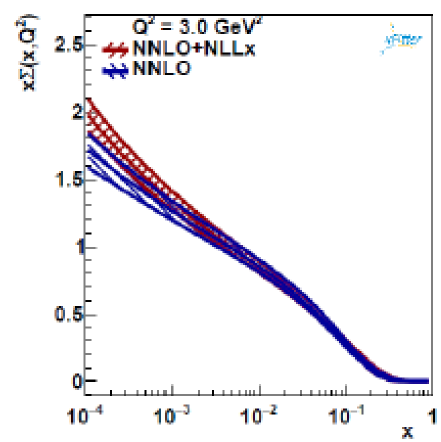


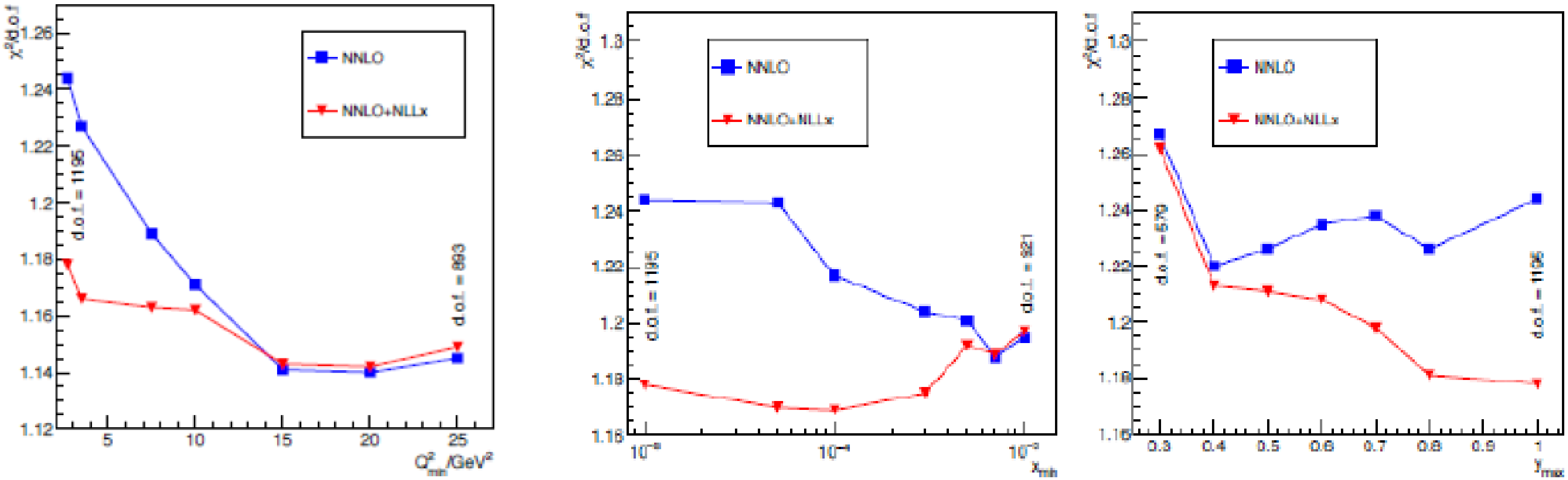
Comparison to data in the lowest Q^2 bins shows that the fit with low x resummation is much better able to follow the turn over of the data that happens at low-x, low Q^2, high-y due to the F_L term in the reduced cross section

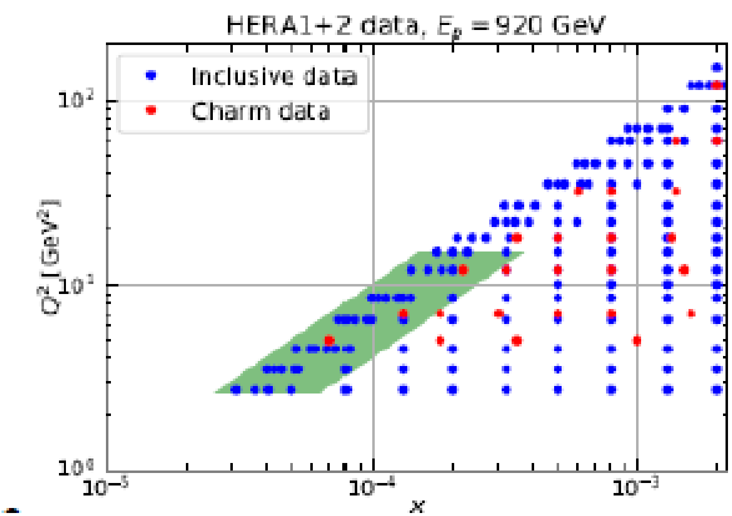
$$\sigma_{\rm red} = F_2 - \frac{y^2}{Y_+} F_L$$

Looking at H1 F_L data directly shows that FL is larger at low Q^2/x for the NLLx fit









Here we see the kinematic region where this can be important.

This only involves LHCb data at low-scale for the LHC
But for an FCC machine the kinematics moves further into the low-x region