Parton Distribution Functions for discovery physics at the LHC

Amanda Cooper-Sarkar
Corfu 2022

At the LHC we are colliding protons

But it is not the protons that are doing the interacting

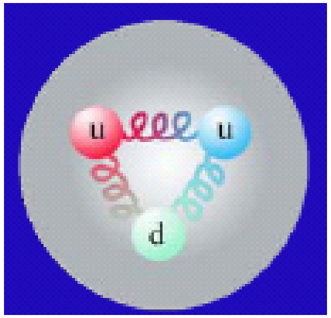
It is their constituents — quarks, antiquarks, gluons - collectively known as partons

We need to know how these partons are distributed in momentum, at the energy scales of the collisions, in order to understand LHC physics.

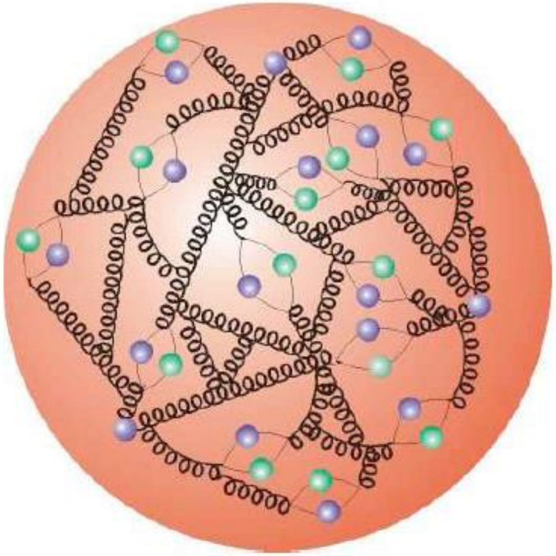
These parton momentum distributions are known as PDFs—Parton Distribution Functions– and are a field of study in their own right.

It is now the case that the uncertainties on Parton Distribution Functions are a major contributor to the background to discovery physics:

- Searches at the highest energy scales- of a few TeV – for physics Beyond our Standard Model (BSM) are limited by PDF uncertainty
- Precision measurements of standard model parameters, like m_W, $\sin^2\theta_W$, which can provide indirect evidence for BSM physics, are also limited by PDF uncertainty, even though PDFs are much better known at these scales, $m_W \sim 80$ GeV



Are there 3 quarks in the proton?



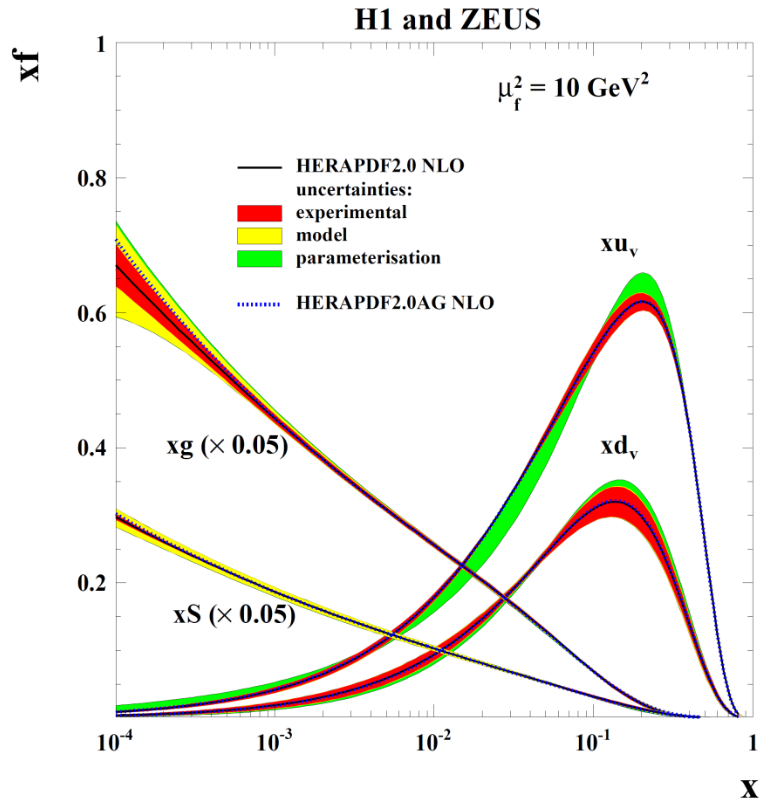
No - there are many more … there are quarks and anti-quarks and gluons - collectively known as ‘partons’

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



What is much more interesting than the numbers of quarks is their dynamics, ie the behaviour of their momentum distributions $xf(x)$ at different scales

The behaviour of these momentum distributions depends on the dynamics of the strong interaction between quarks and gluons-

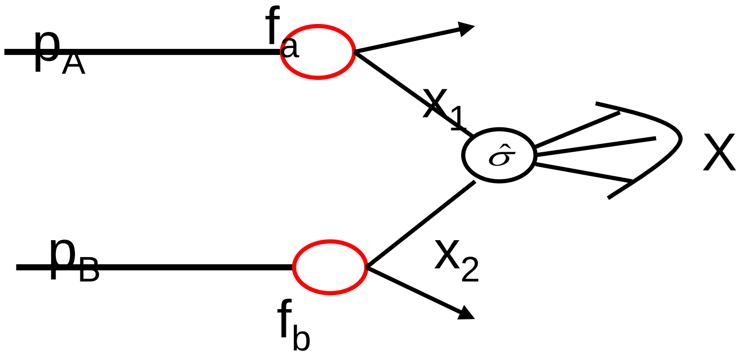
Quantum Chromo Dynamics

In this figure the momentum distributions of the valence quarks, xu_v and xd_v are shown

There are 2 u-valence quarks to each 1 d-valence quark- though the shapes of their momentum distributions are NOT exactly the same, so it is not 2:1 at every x

And you can also see that there is a ‘SEA’, xS, of quark-antiquark pairs and a rising GLUON distribution, xg, at low-momentum fractions

The total momentum taken by all the partons sums to unity



$$\sigma_{\mathbf{X}} = \sum_{\mathbf{a,b}} \int_0^1 \mathbf{dx_1 dx_2\, f_a(x_1, \mu_F^2)\, f_b(x_2, \mu_F^2)}$$
$$\times\ \hat{\sigma}_{\mathbf{ab \to X}}\left(\mathbf{x_1, x_2}, \{\mathbf{p_i^\mu}\}; \alpha_{\mathbf{S}}(\mu_{\mathbf{R}}^2), \alpha(\mu_{\mathbf{R}}^2), \frac{\mathbf{Q^2}}{\mu_{\mathbf{R}}^2}, \frac{\mathbf{Q^2}}{\mu_{\mathbf{F}}^2}\right)$$

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 $\sigma_{ab\to X}$).

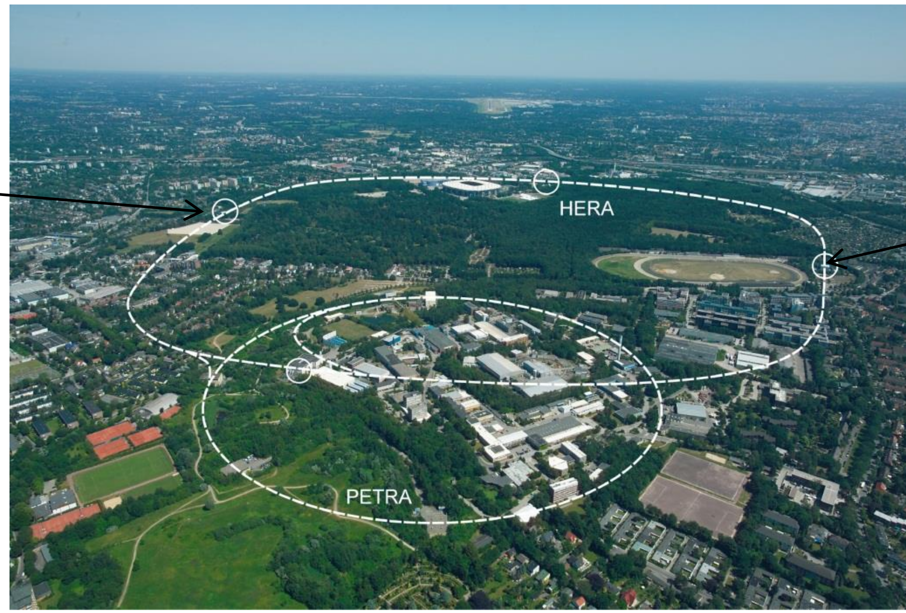
The momentum fractions: $x_1f_a(x_1)$, $x_2f_b(x_2)$, are the parton momentum distributions or PDFs

They depend (mildly) on the scale of the process μ_F which is a measure of the energy put into the sub-process collision

How did we come to know all this?

From Deep Inelastic Scattering

The most important experiment and backbone of all fits to determine PDFs was HERA



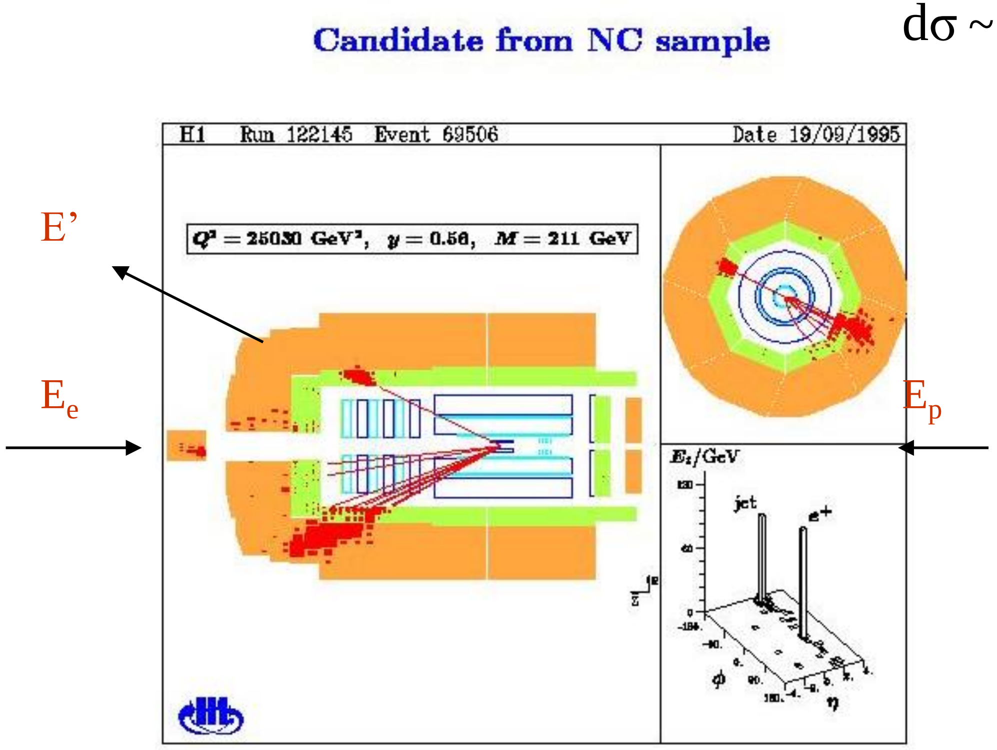
From HERA the e-p collider at DESY, Hamburg.

~500pb^{-1} per experiment split ~equally between e^{+} and e^{-} beams: DESY-15-039

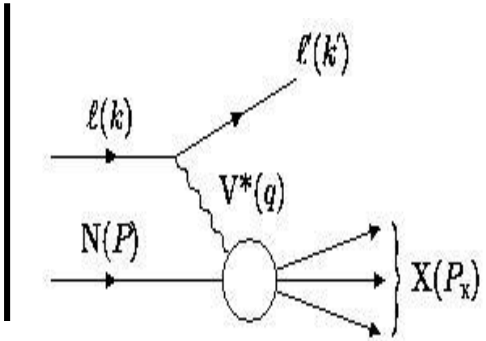
Running at Ep = 920, 820, 575, 460 GeV
$\sqrt{s}$ = 320, 300, 251, 225 GeV

From 1992-2007

Deep inelastic lepton-nucleon scattering -DIS



$d\sigma \sim$



^{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

$Q^2 = s\, x\, y$

$s = 4\, E_e\, E_p$

$Q^2 = 4\, E_e\, E' \sin^2\theta_e/2$

$y = (1 - E'/E_e \cos^2\theta_e/2)$

$x = Q^2/sy$

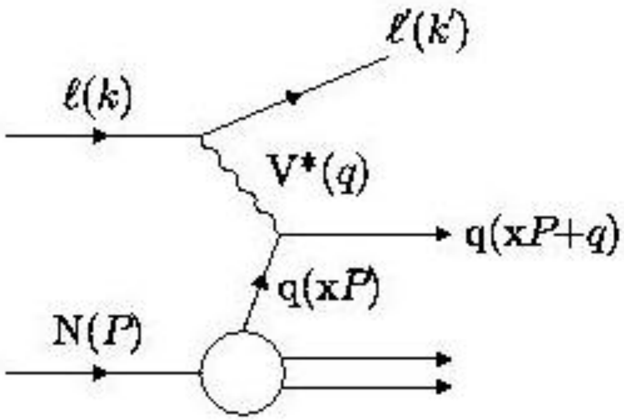
These kinematic variables are measurable

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

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

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 target 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

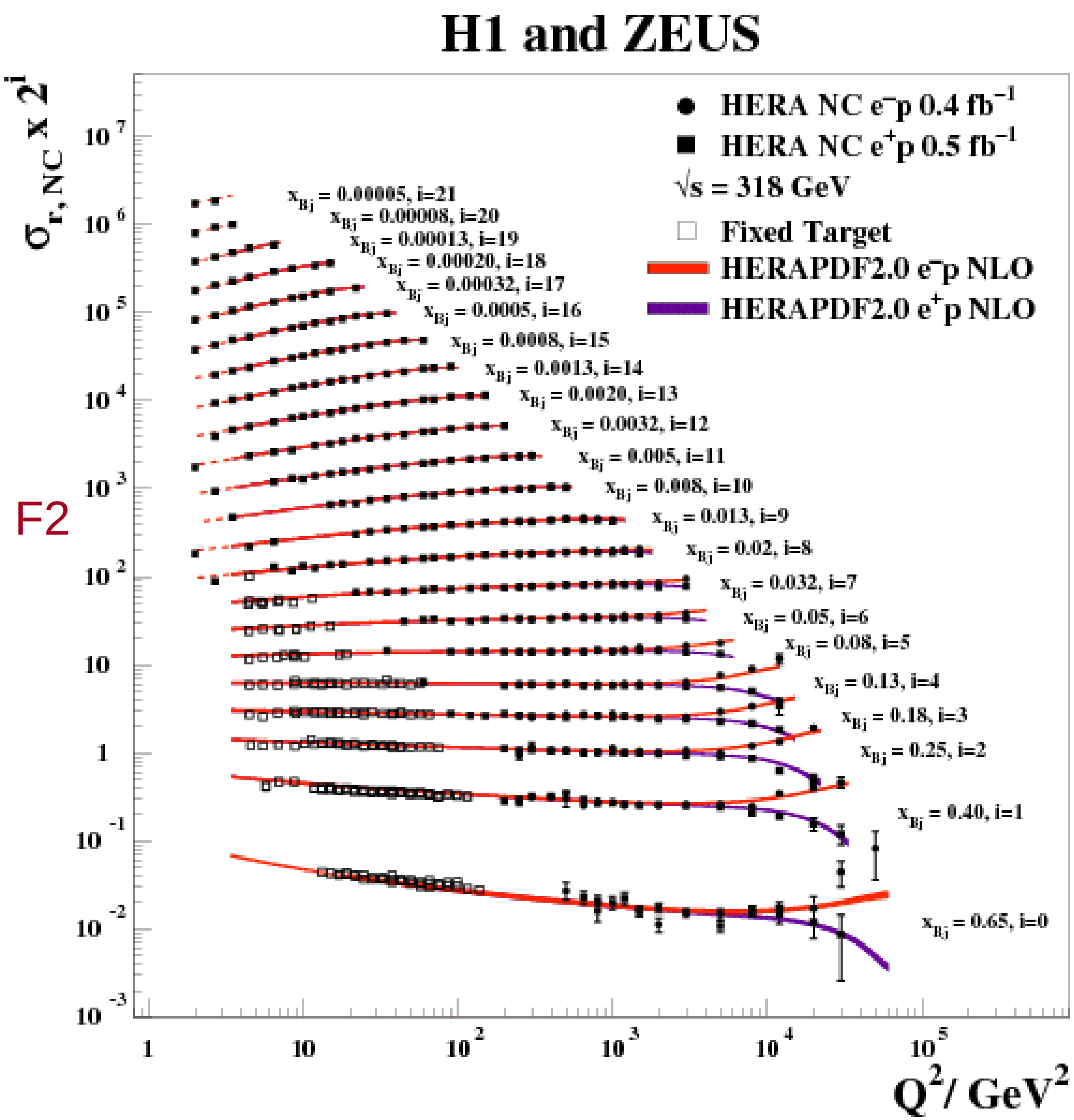
$(xP+q)^2 = x^2P^2 + q^2 + 2xP.q \sim 0$ for massless quarks, and given that te proton is also close to massles, $P^2 \sim 0$, so

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

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

The prediction of the quark-parton model was that the differential cross-section for lepton-proton scattering would depend only on the structure function F_2



AND that

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

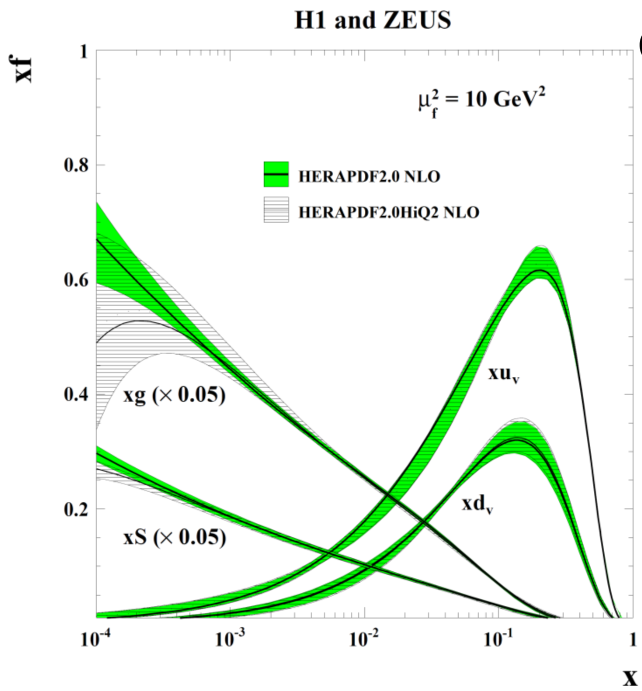
i.e. that F_2 would be independent of the scale of the probe Q^2

How good is this?

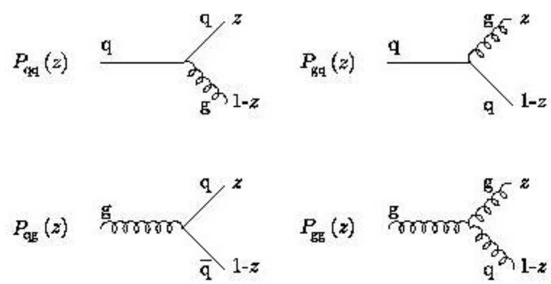
Pretty good – this is a log plot

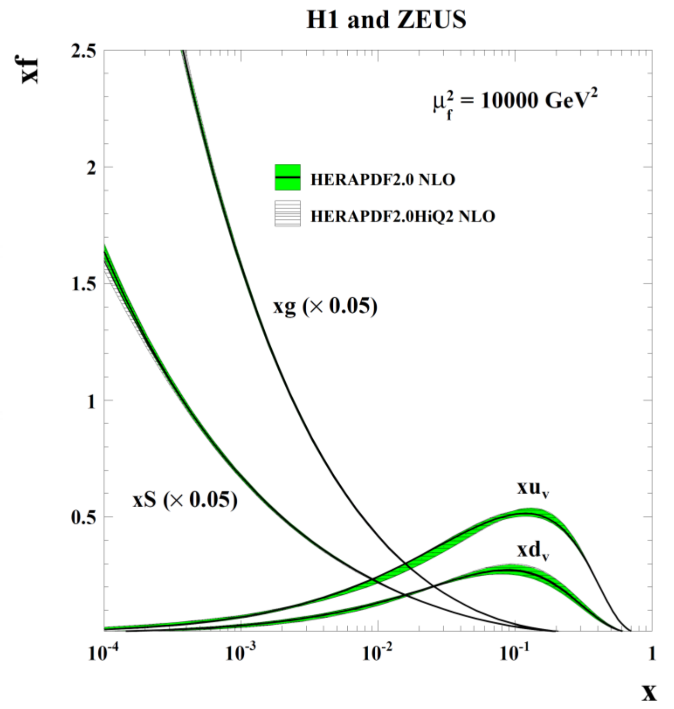
Non-point like structure would have ~1/Q^2 behaviour, we are sure there is no substructure down to 10^{-19} m now.

But it's clearly not perfect- that's because Quantum Chromo-Dynamics tells us that there is a slow change with log Q^2



Quantum Chromodynamics





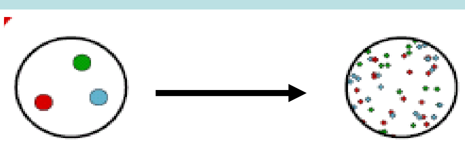
The valence quarks are radiating gluons and the gluons are splitting into quark-antiquark pairs

Parton momentum distributions change with the scale of the probe: Q^2 = ~10 GeV^2 is typical scale for low energy experiments

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

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

And at these scales the proton is pretty well all glue



There is information on parton flavour according to which leptons and nucleons you collide

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

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

Assuming:
u in proton = d in neutron
⇒ strong-isospin

Low energy γ-exchange only formulae (HERA went beyond this)

Charged lepton proton and deuteron data give different flavour combinations

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

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

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

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

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

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

The above 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

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

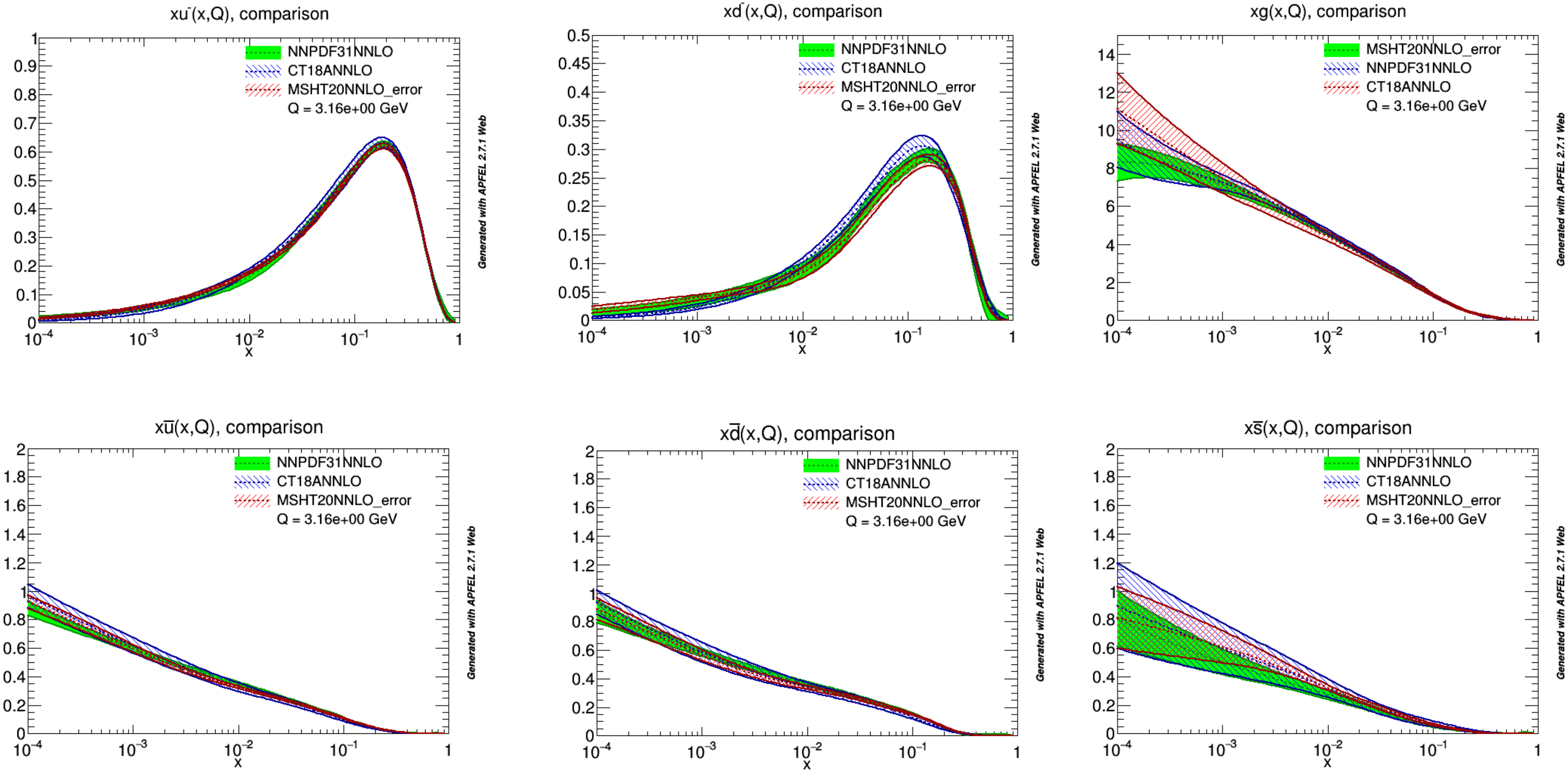
Where $P_i(x)$ can be ordinary polynomials of x, or $\sqrt{x}$, or Chebyshevs, Bernstein polynomials, typically ~20-30 parameters - or even a neural net

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'…

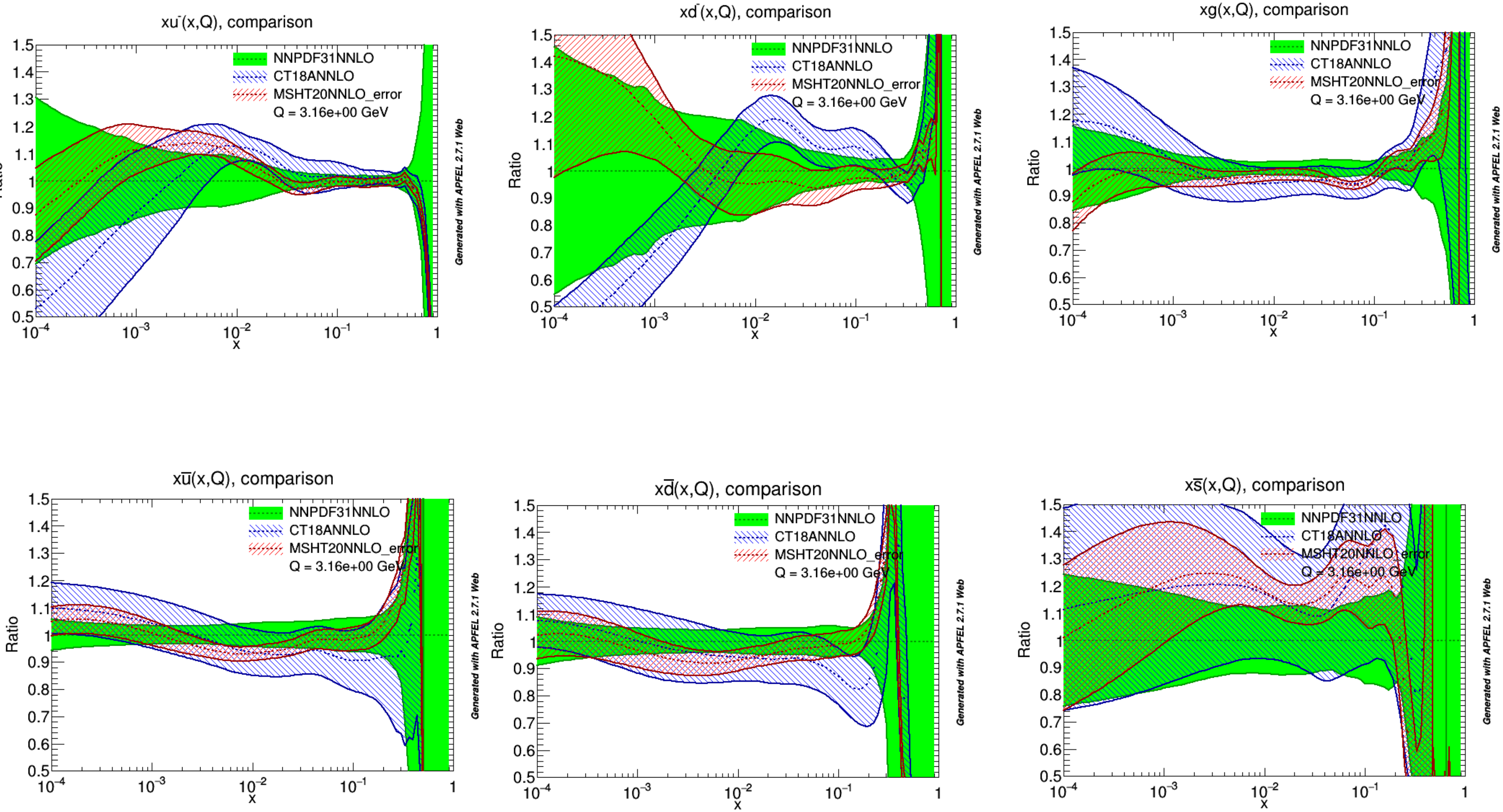
Use QCD to 'evolve' these PDFs to higher scale $Q^2 > Q^2_0$
Construct predictions for the measurable structure functions in terms of PDFs for ~3000 data points across the x, Q^2 plane
Perform $\chi 2$ fit to the data.

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, $\alpha_s(M_Z)$

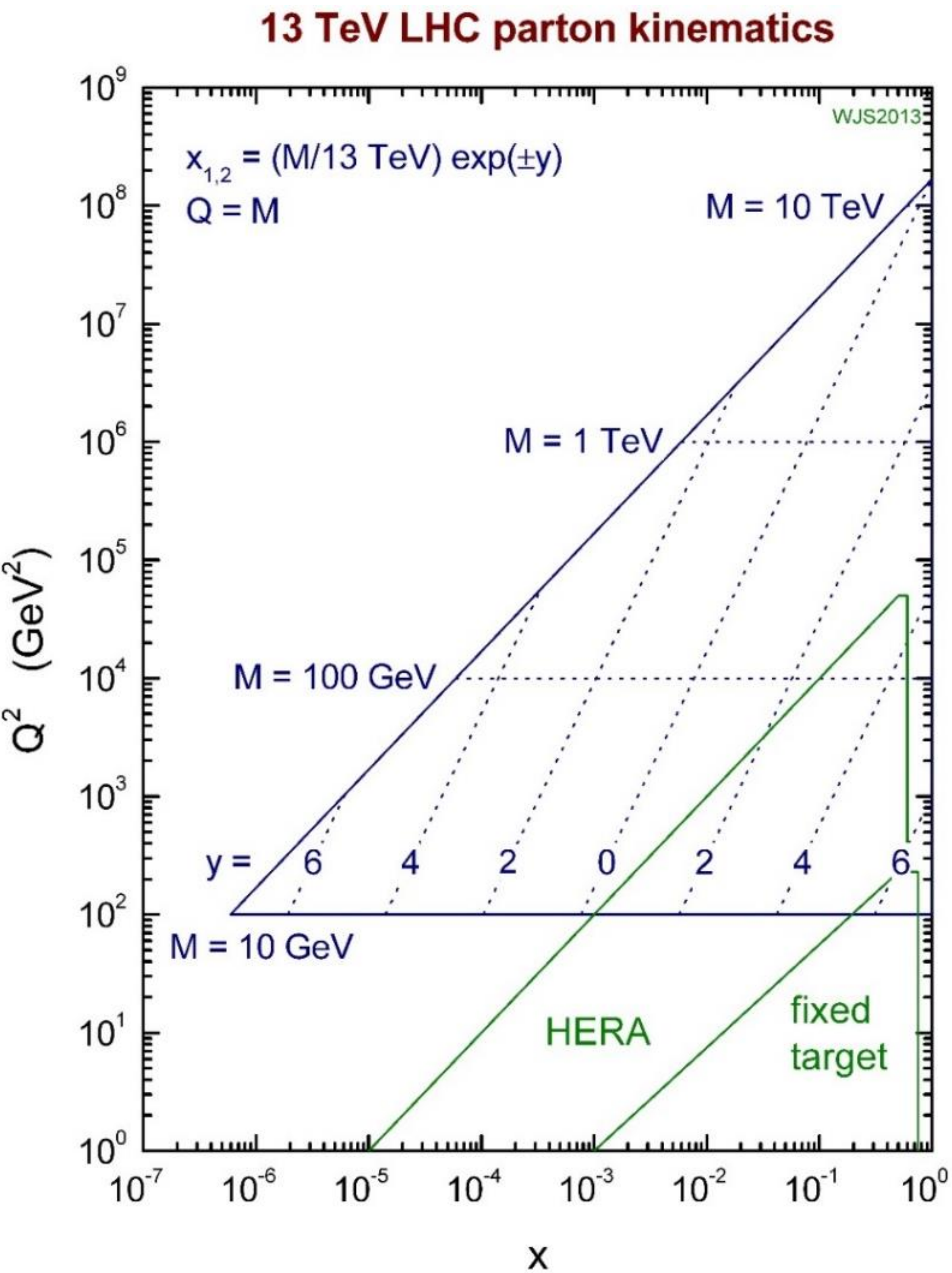


But several groups extract PDFs and there are significant differences because of slightly different model choices:

- Exact choice of data entering fit
- Choice of heavy quark masses
- Choice of starting scale for QCD evolution … etc, etc



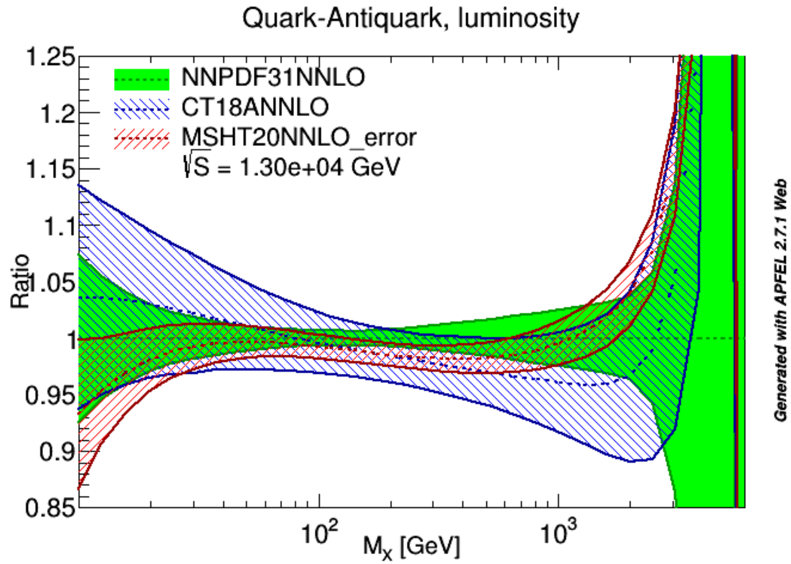
Differences are more obvious in ratio
They are large at small-*x* and at high-*x*
where *x* is the fractional momentum of the struck parton



Going to the LHC

Our knowledge of PDFs in the LHC kinematic region has come from evolving the results from HERA and other Deep Inelastic Scattering experiments in Q^2 using the QCD 'DGLAP' evolution

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

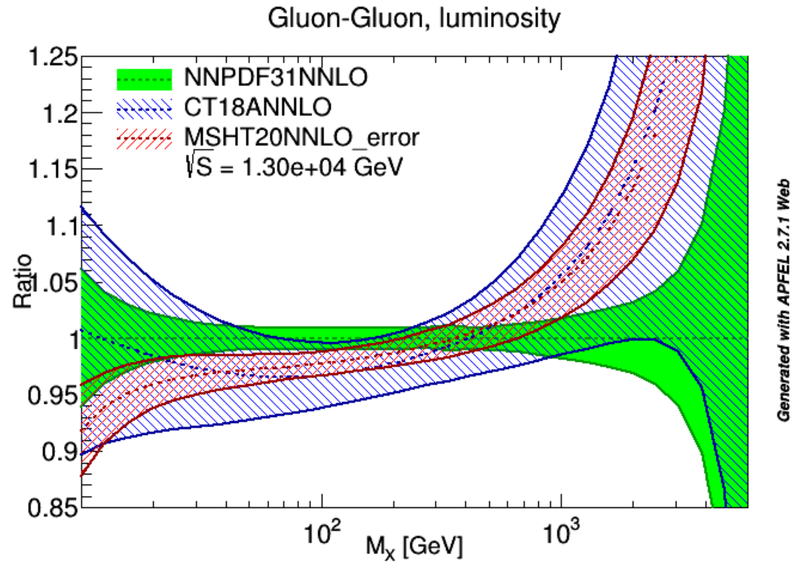


The quark-antiquark and gluon-gluon luminosities for various PDFs are compared here for 13 TeV LHC running in terms of the centre of mass energy of the parton sub- process M_X

Small M_X corresponds to small x and Large M_X to large x

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

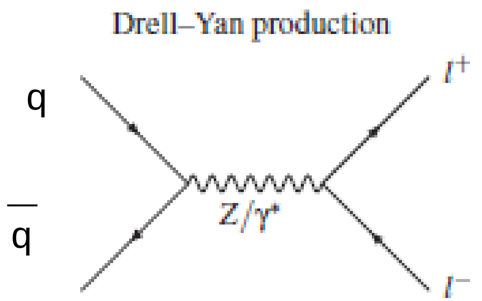
So for quark-antiquark production of W or Z bosons – at Mx ~ 80, 90 GeV
Or for gluon-gluon production of Higgs at – Mx ~ 125 GeV
the parton-parton luminosities are fairly well known *(but perhaps not well known enough…)*
This is not so for higher mass particles that could be produced by 'Beyond' Standard Model (BSM) physics

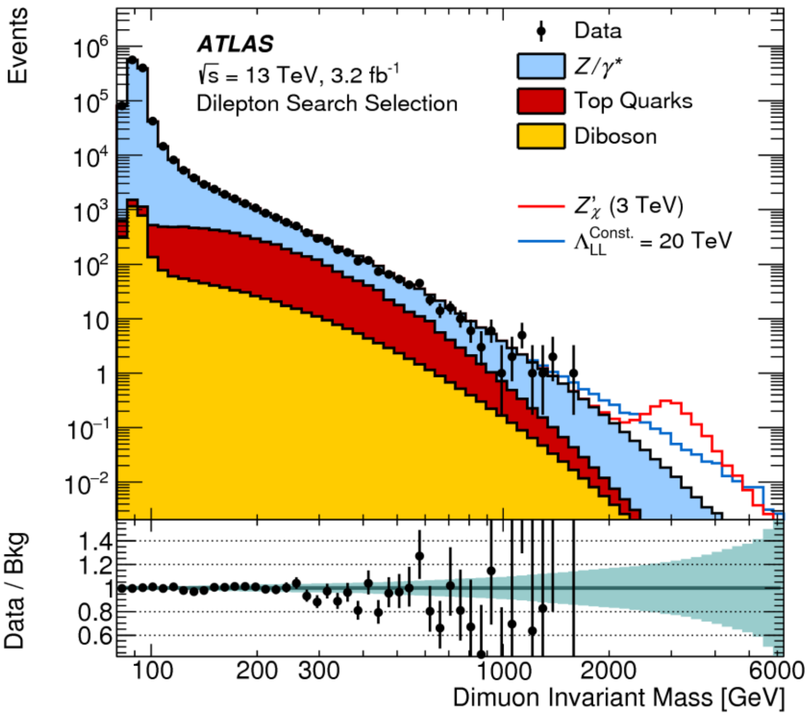


$$\frac{\partial \mathcal{L}_{gg}}{\partial \hat{s}} = \frac{1}{s}\int_\tau^1 \frac{dx}{x} f_g(x,\hat{s}) f_g(\tau/x,\hat{s}),$$

Consequence of uncertainty at high-x?- 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





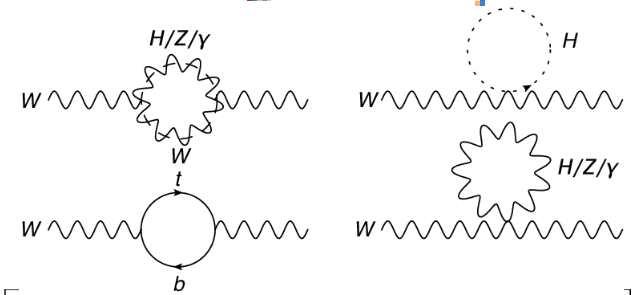
Drell-Yan is a term for q-qbar → $\mu^+ \mu^-$ collisions mediated by *Z* or virtual *γ*, *Z* bosons.

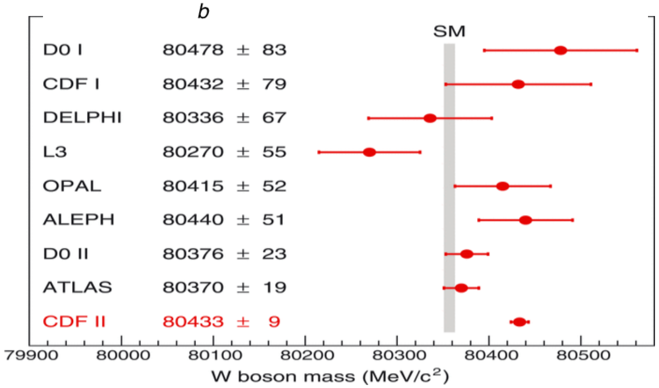
Some new theories predict higher mass *Z'* states, these have been excluded up to 2 TeV.

The main reason we cannot do better is that the PDF uncertainty on the 'normal' Standard Model background is too big.

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.

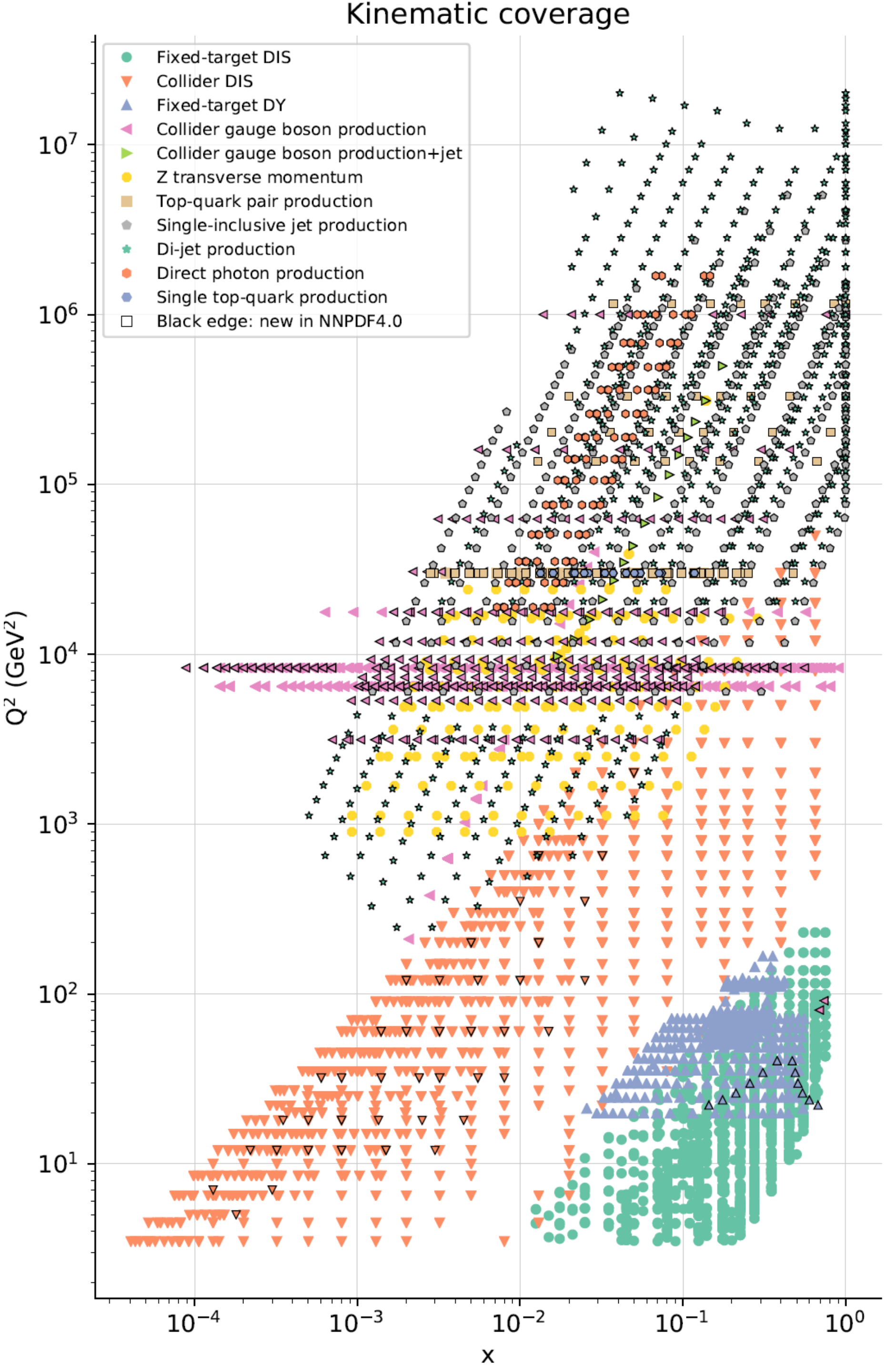
In that case the value of m_W would differ from its SM value

And indeed that is what we see in the latest Fermilab measurement!

BUT how can this be checked?

Well it can be checked at the LHC.

The most accurate LHC measurement to date is from ATLAS and is shown on the plot. A major contribution its uncertainty of 19 MeV is the PDF uncertainty of 10 MeV. LHC uses p-p not p-pbar and its kinematic reach is such that most collisions producing W are sea-quark –antiquark collisions. It is not clear that the PDF uncertainties can be improved quickly…….



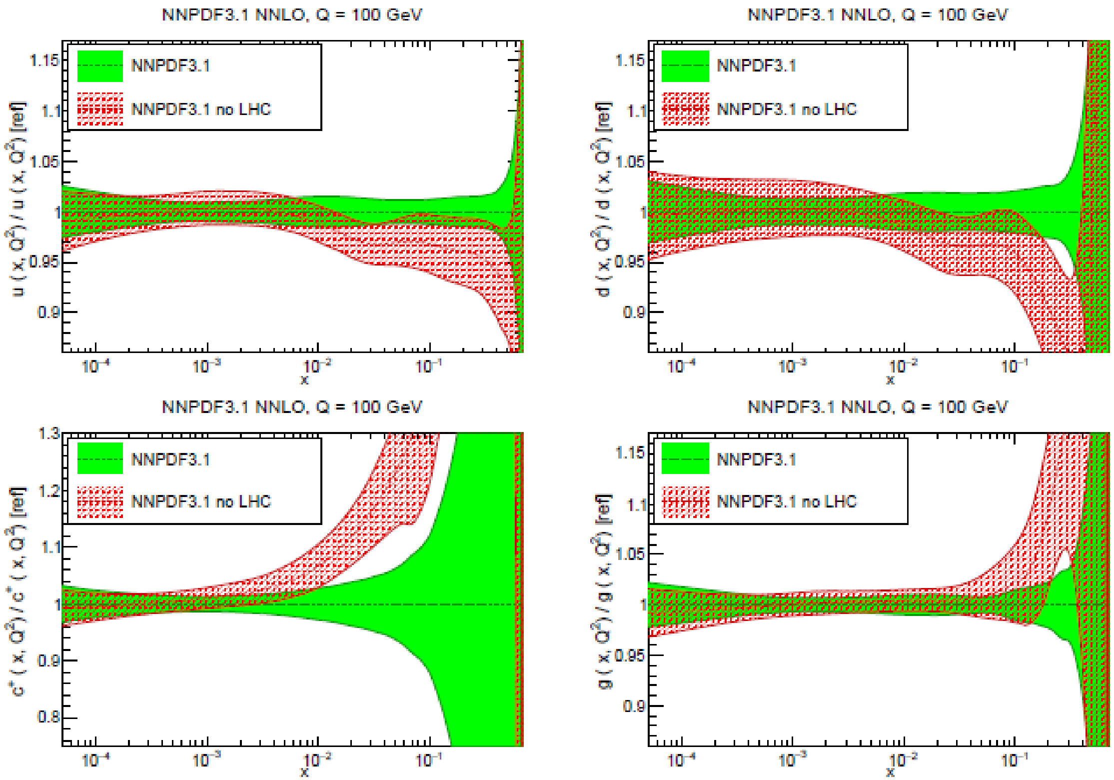
So how can we hope to improve matters in future?

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

- 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
- These are all processes that can be calculated at fixed order, currently NNLO
- But beware: there may be new physics at high scale that we 'fit away'

So let's see how much LHC data is improving PDFs

NNPDF3.1 includes modern LHC data on W,Z + jets + top + Zpt from 7 and 8 TeV running. Compare PDFs with and without LHC



Some of the data input to NNPDF3.1 –like the ATLAS W,Z data have already reached a limit of how accurate they could be.

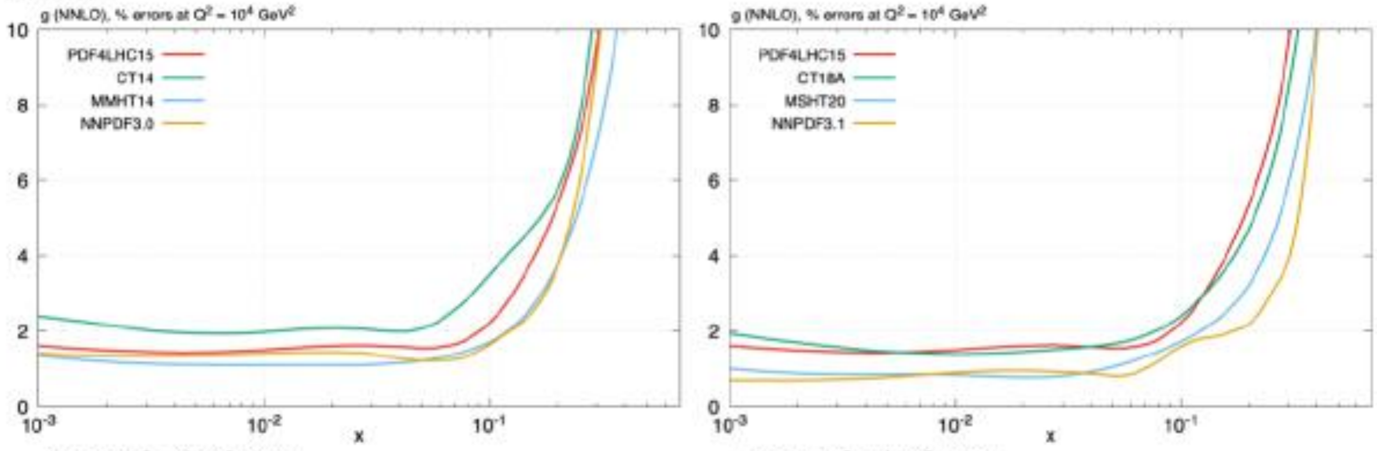
The uncertainties of *O*(1%) are limited by experimental systematics not by statistics. This will not get better in the foreseeable future with the High Luminosity LHC

FURTHERMORE, this looks good BUT specific choices were made by NNPDF e.g which top-quark differential distributions are used and of which jet data distributions are used etc.

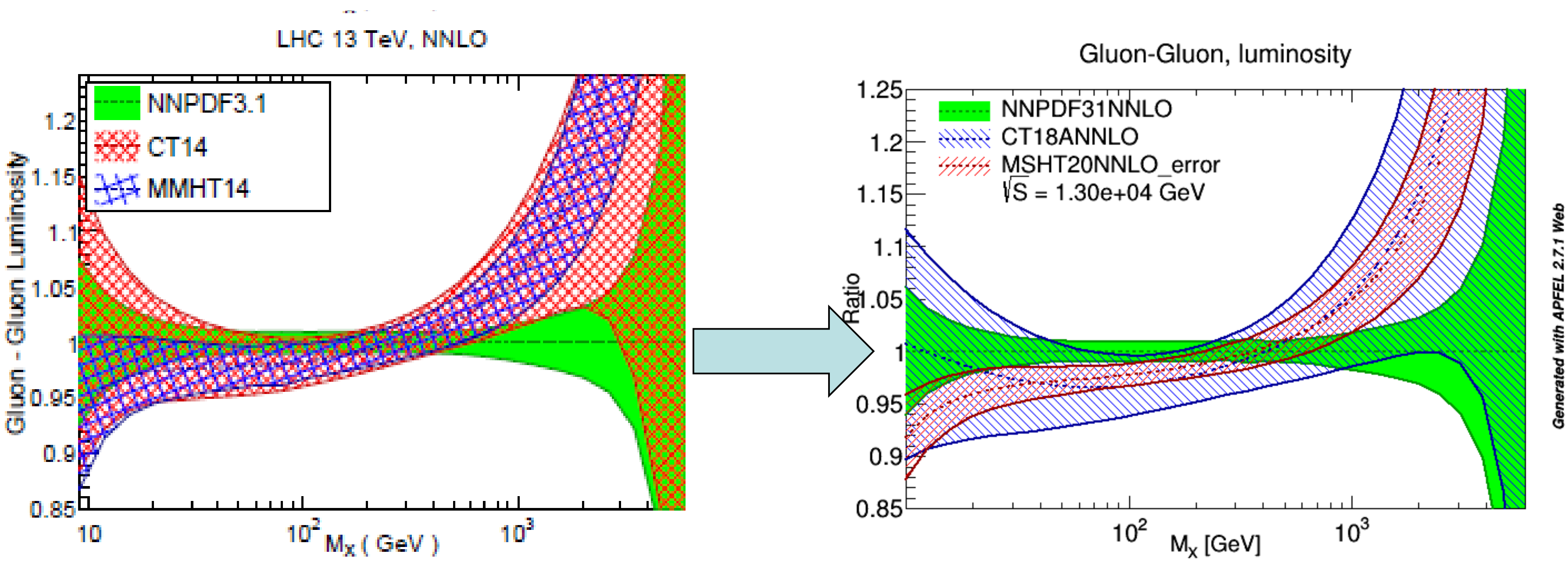
NNPDF4.0 improves this further… but this is not yet agreed by other groups.

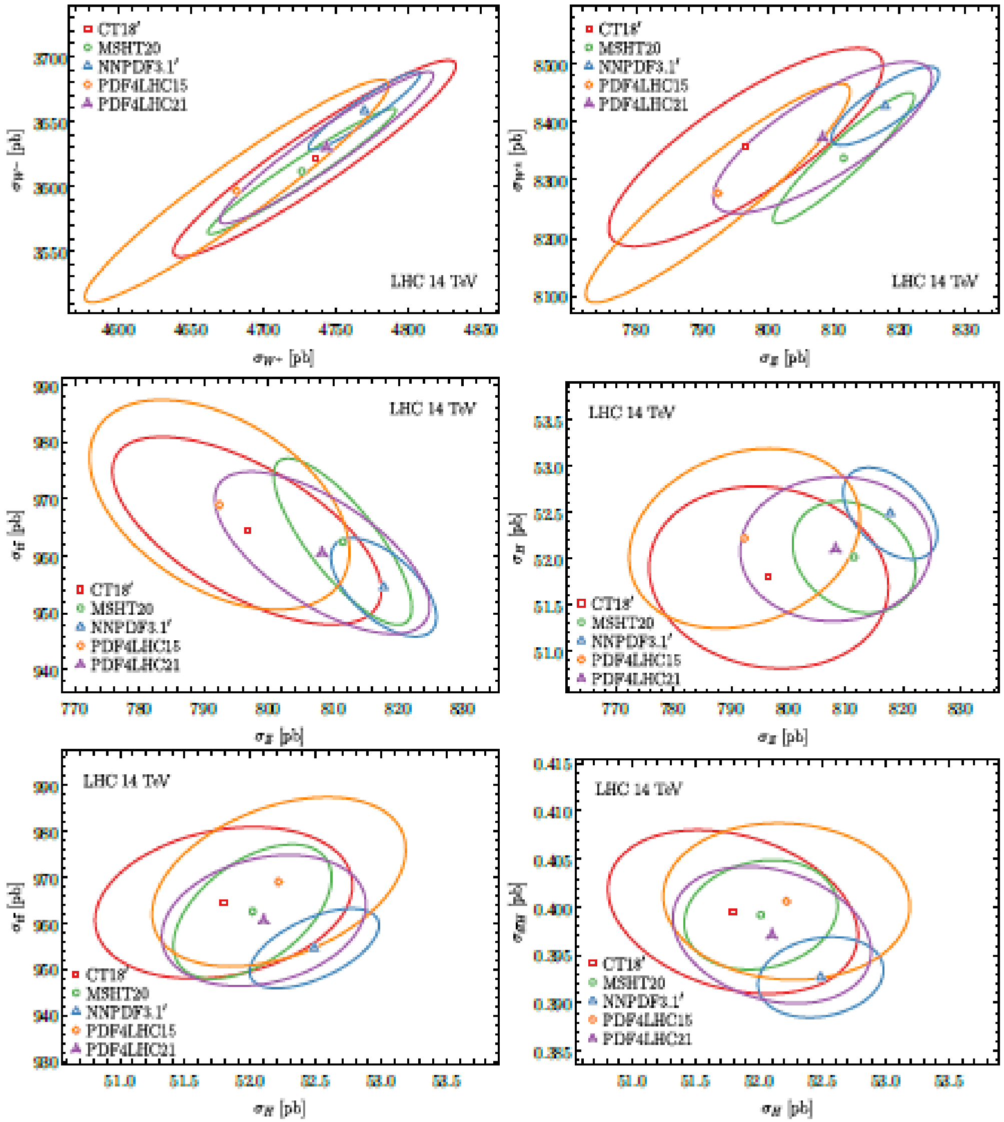
Other PDF groups are making other choices--this could even increase the total uncertainty due to differences between PDF sets

Gluon



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



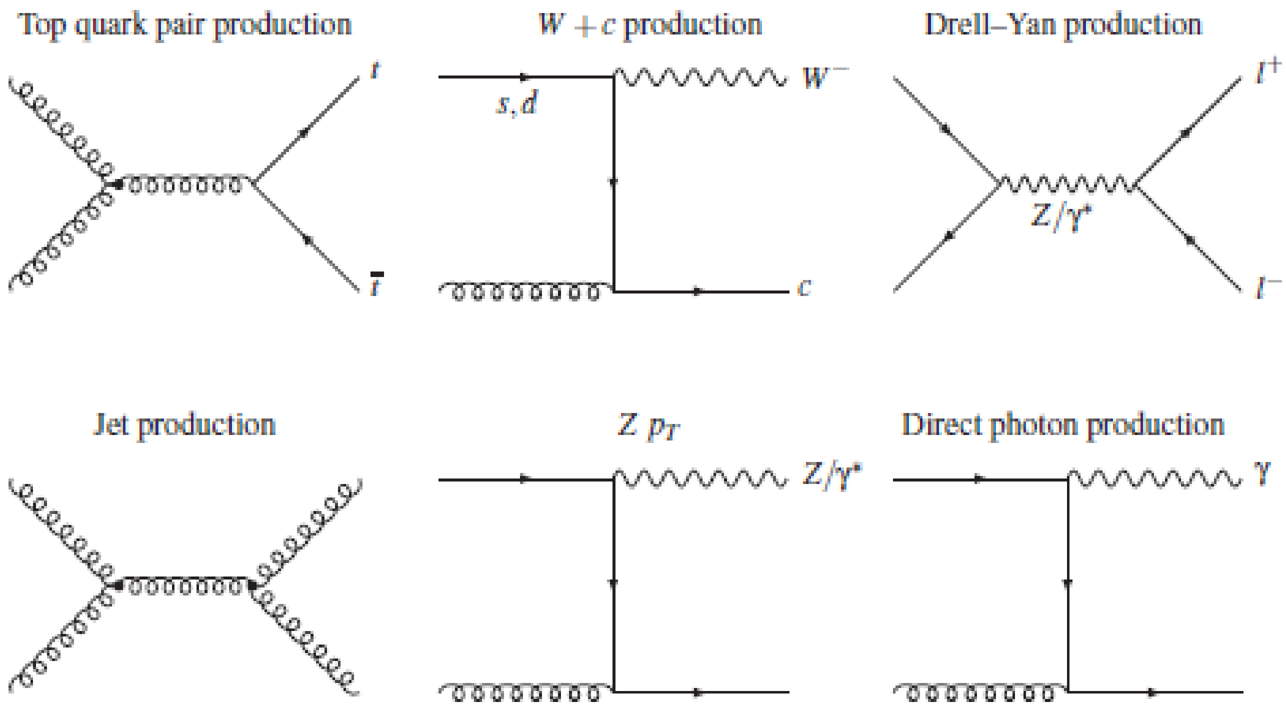


The PDF4LHC group makes combinations of the PDFs from the three main fitting groups NNPDF, CT and MSHT

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 sufficiently -10 MeV could decrease to ~8MeV – we need more than this …

A recent study of potential improvements has been made using processes for 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

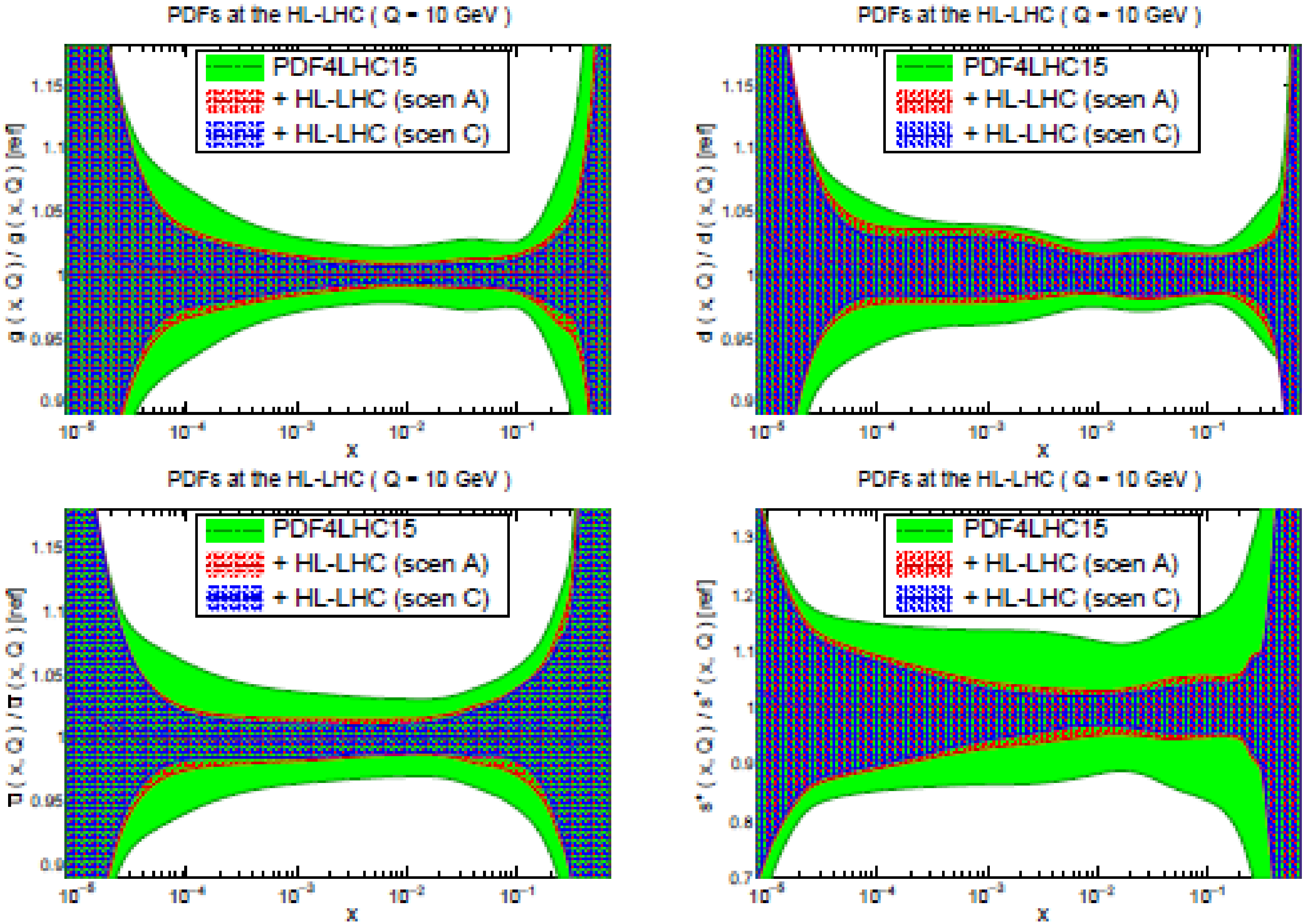
Pessimistic and Optimistic assumptions are made about systematic uncertainties based on experience with real data

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

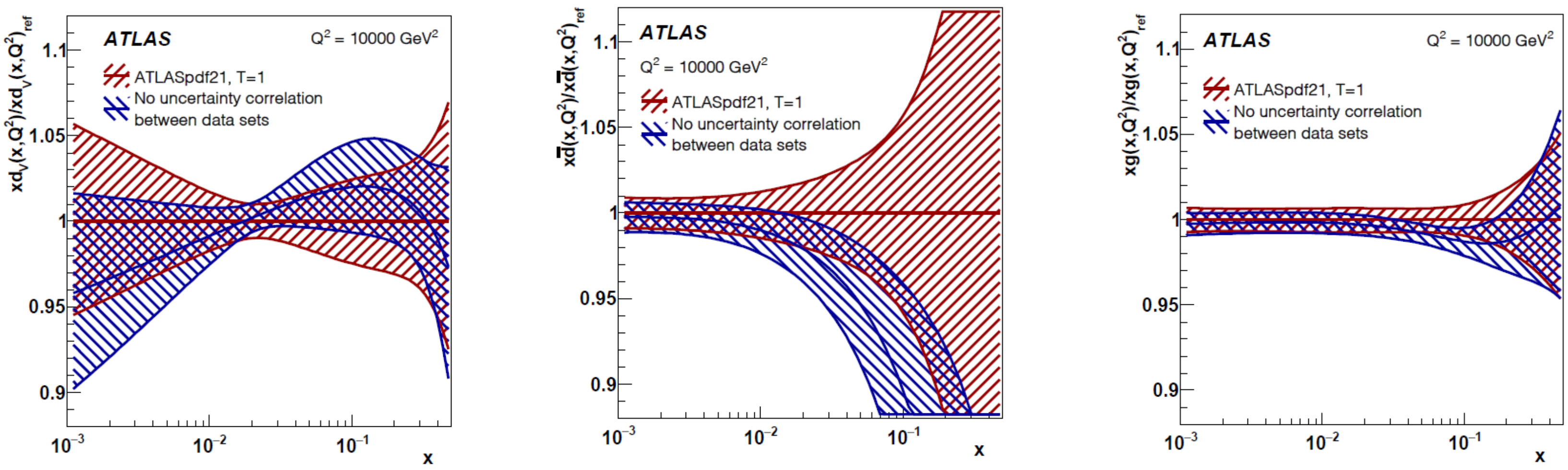


Where scenario A is pessimistic and scenario C is optimistic
- Such improvements could give up to a factor 2 improvement in the PDF uncertainty on something like m_W

But are we being a little too optimistic? YES…

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

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



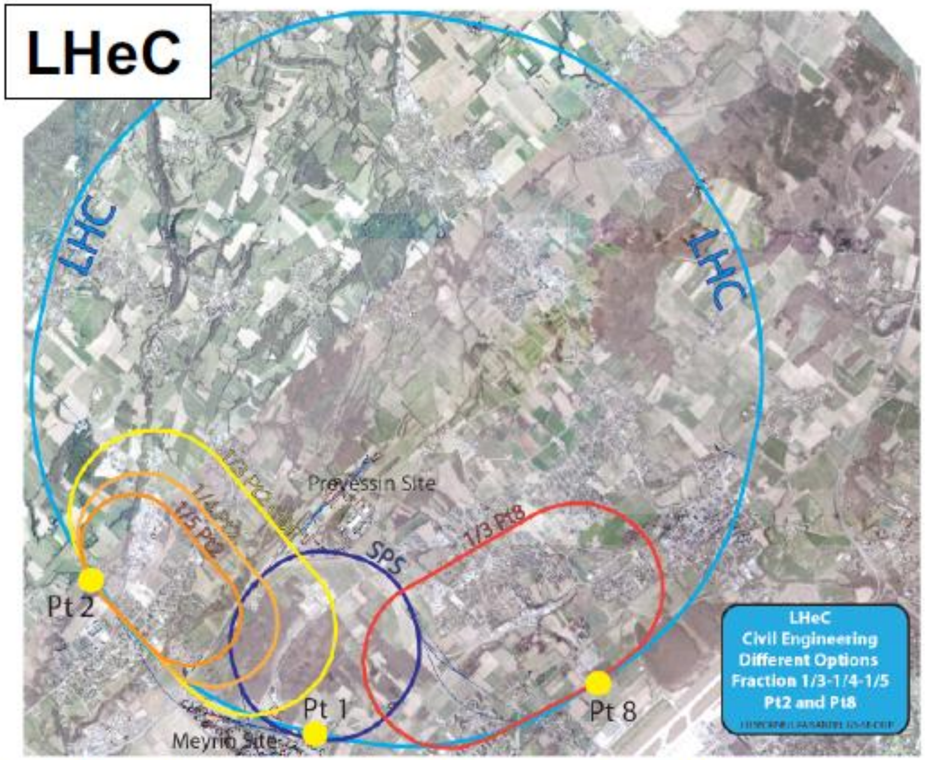
The larger correlations come between data sets involving the jet data such as: inclusive jets, *W* or *Z* boson +jets, t-tbar in lepton+jets mode

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

SO HOW MAY WE ACTUALLY DO MUCH BETTER?

New DIS machines



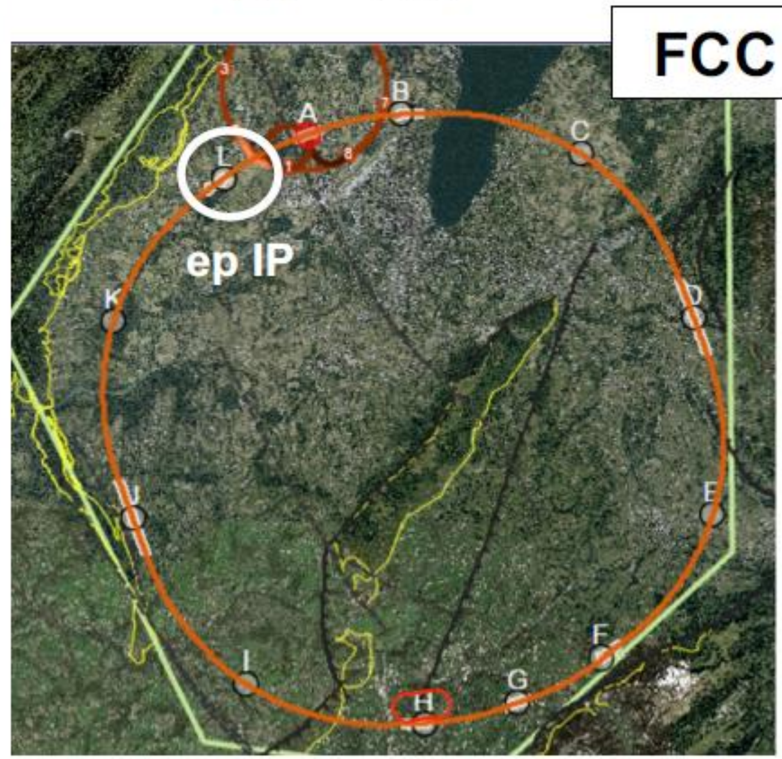
(M Klein, Rencontre du Vietnam, Sept 2017)

LHeC and **FCC-eh**

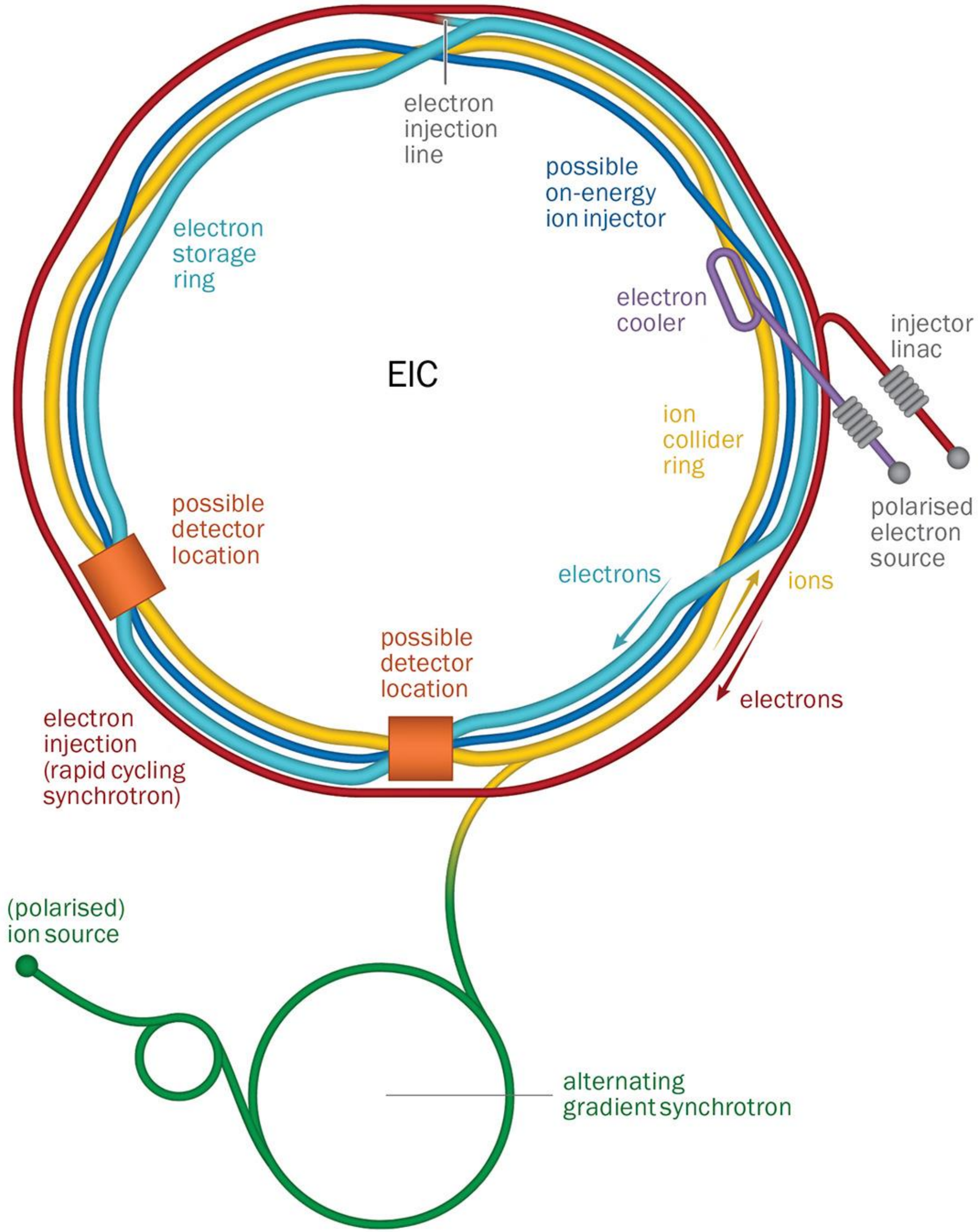
energy recovery LINAC

e-beam: 60 GeV

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

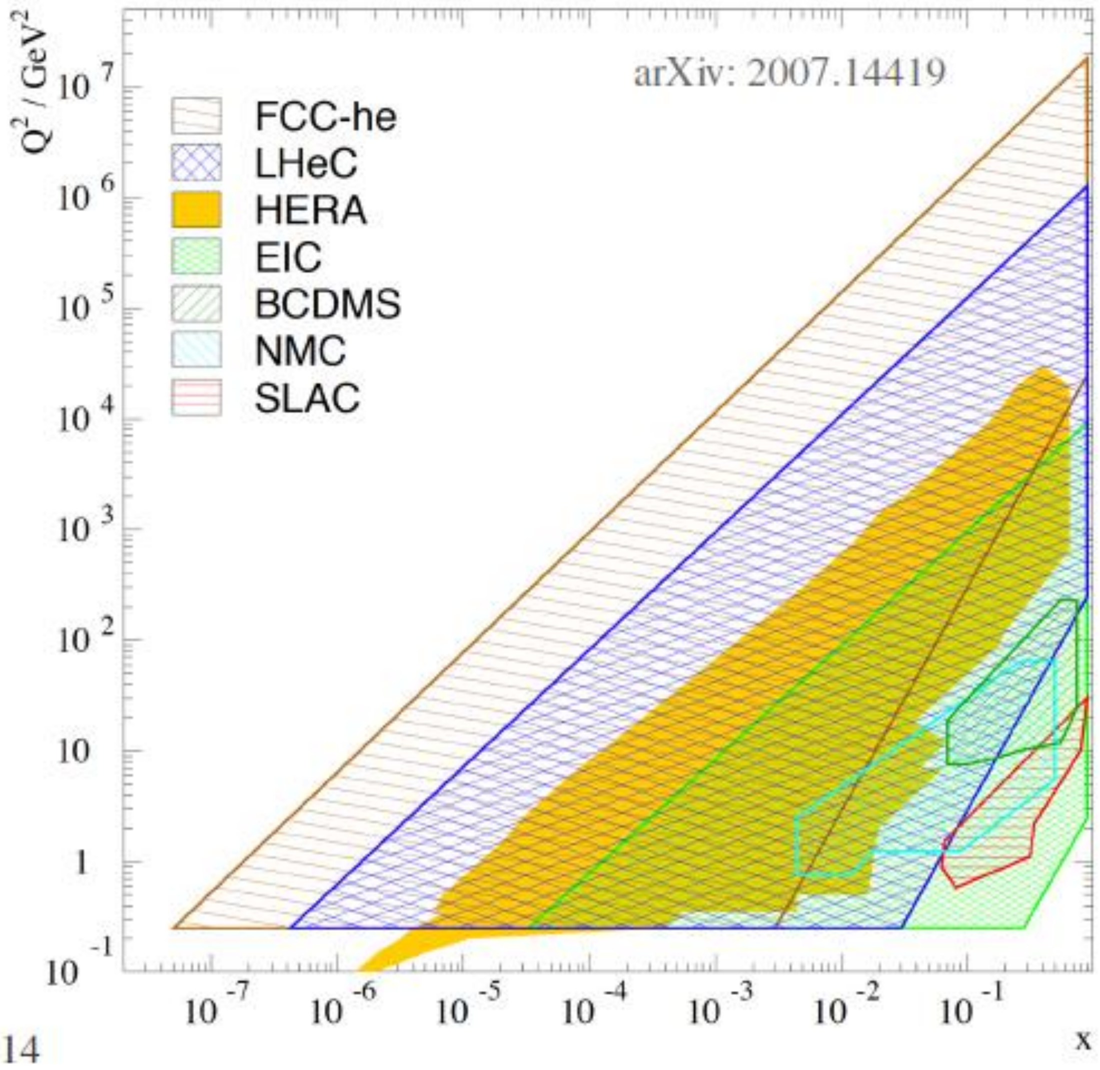


LHeC (**FCC-eh**) complementary to, synchronous with, **HL-LHC** (**FCC**)



OR the **Electron Ion Collider**
to be at Brookhaven—
And this one WILL be built

Consider the kinematic reach of each of these

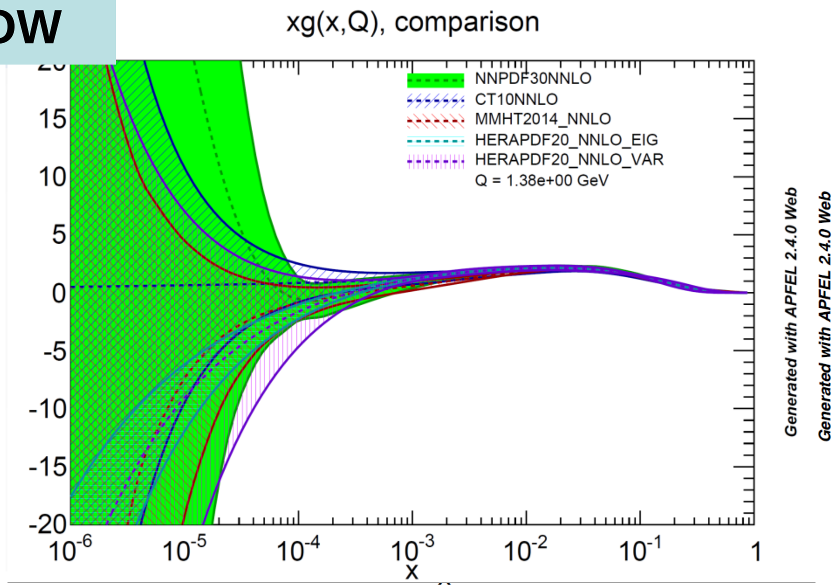


The proposed LHeC and FCC-eh machines reach lower x than HERA could reach

The EIC will reach higher x than HERA could reach

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

NOW



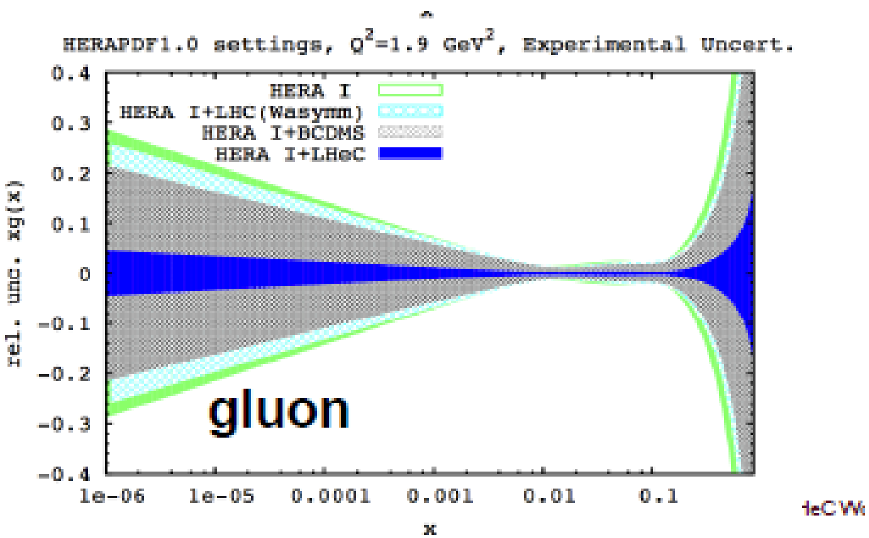
HERA sensitivity stops at $x > 5\ 10^{-4}$

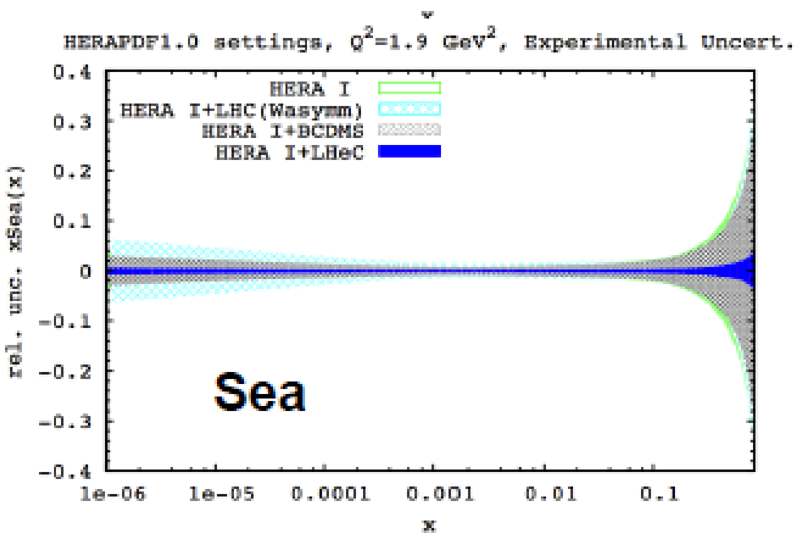
Below that uncertainties depend on the parametrisation

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 impact is at high x

ePump: Schmidt, Pumplin, and Yuan; PRD98 (2018) no.9, 094005

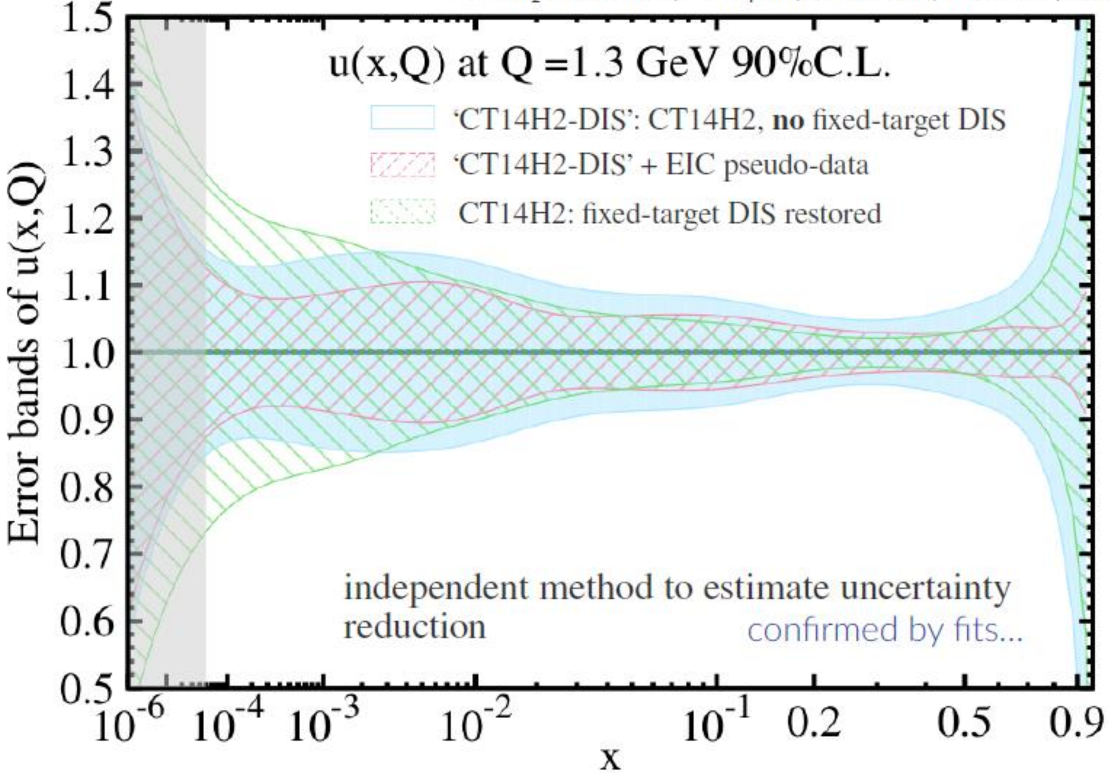


figure: S. Dulat

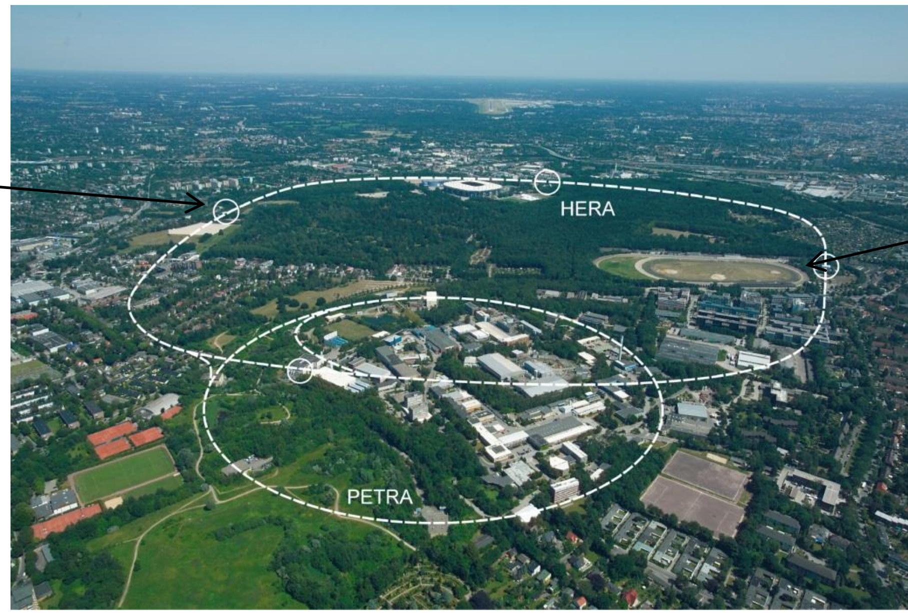
- **inclusive EIC may surpass total impact of fixed-target DIS in modern fits**

Summary

- 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!**

Back ups

How did we come to know all this?



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

~500pb^{-1} per experiment split ~equally between e$^+$ and e$^-$ beams: DESY-15-039

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

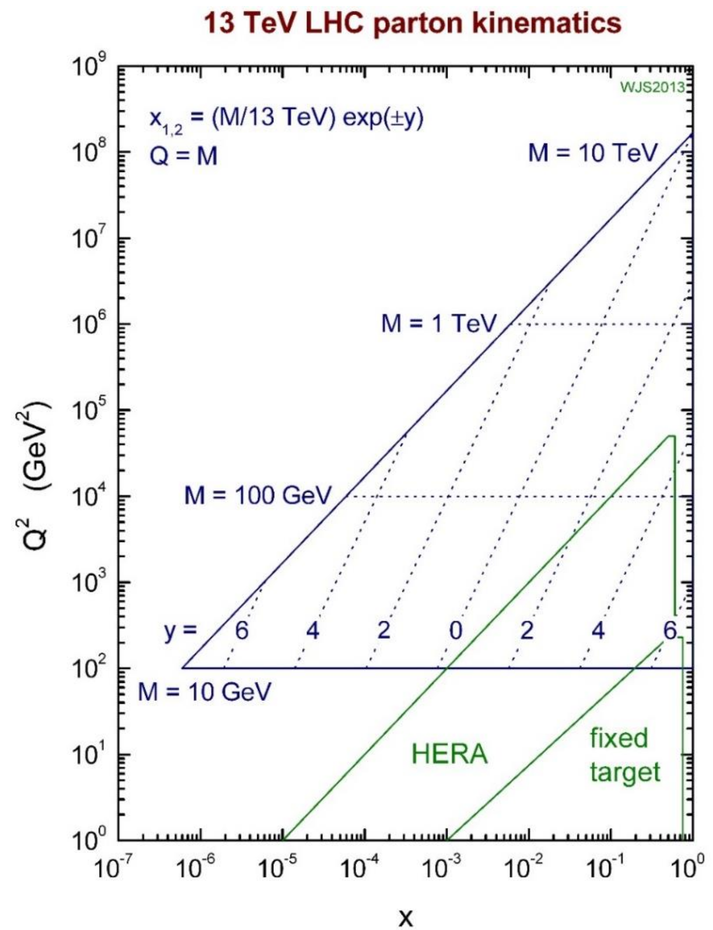
From 1992-2007

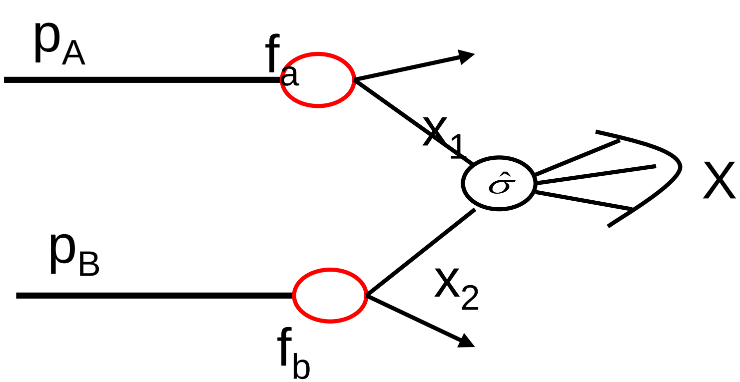
So the Standard Model is not as well known as you might think

In the strong interaction (QCD) sector the uncertainties on Parton Distribution Functions (PDFs) limit our knowledge of all cross sections- whether Standard Model or Beyond- because we are colliding protons But it is the partons doing the interacting

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

where X=W, Z, Higgs, high-E_T jets, or something new and σ is known in the Standard Model perturbative QCD and ElectroWeak theories, or it is a 'new-physics' cross section

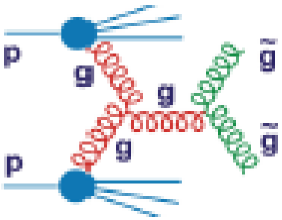




Our knowledge of PDFs in the LHC kinematic region has come from evolving the results from HERA and other Deep Inelastic Scattering experiments in Q^2 using the QCD 'DGLAP' evolution

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

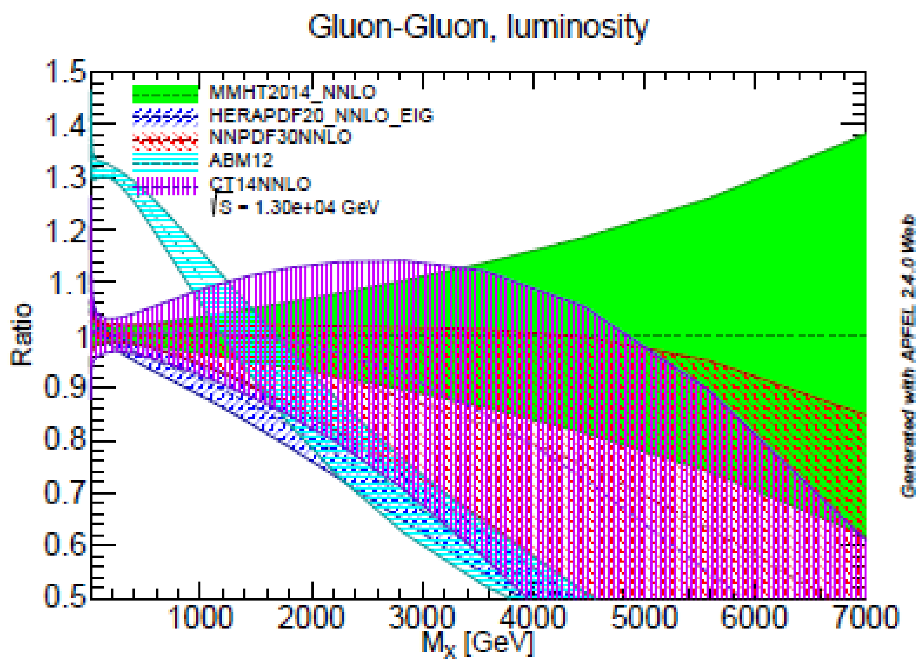
Many interesting processes at the LHC are gluon-gluon initiated
…BSM processes like gluon-gluon → gluino-gluino

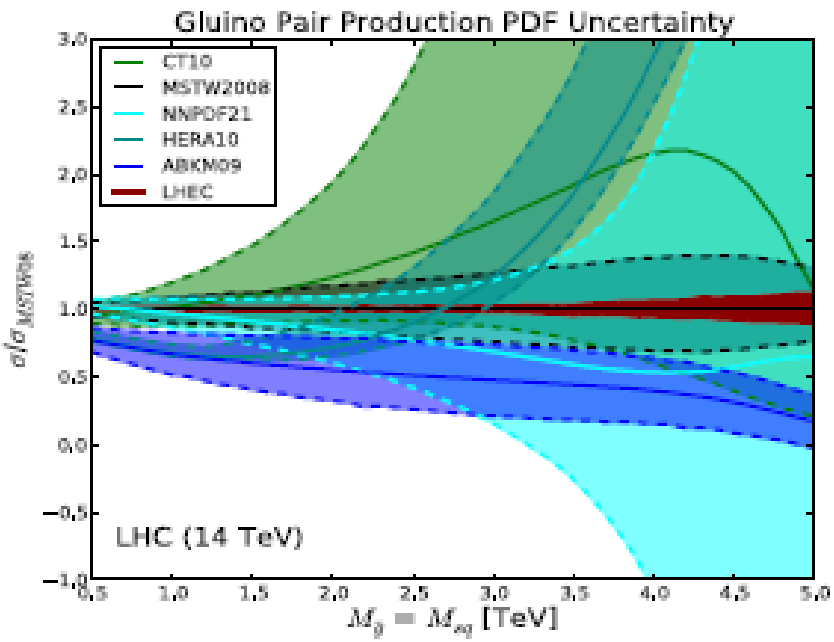


And the high-scale needed for this involves the high-x gluon

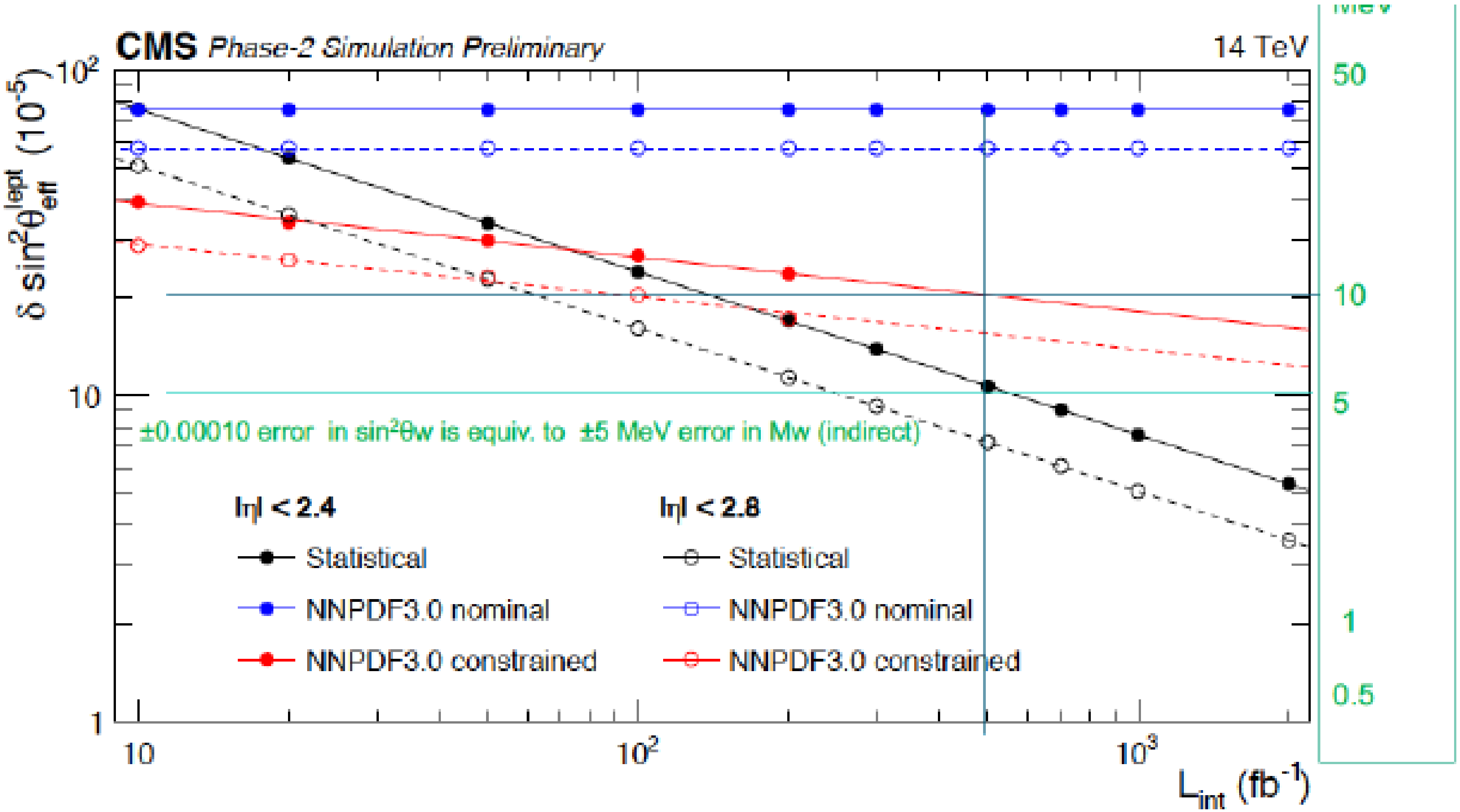
The gluon-gluon luminosity at high-scale is not well-known

This leads to uncertainties on the gluino pair production cross section





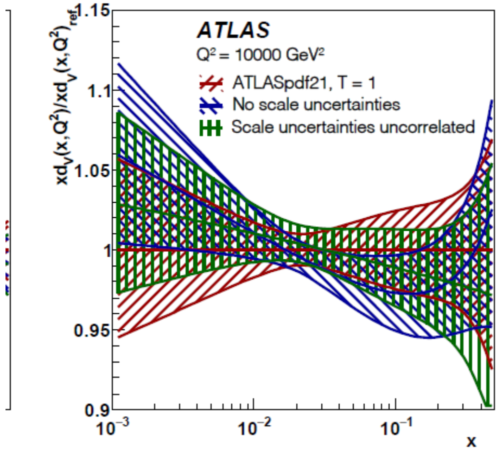
A further example is the uncertainty on the electro-weak mixing angle $\sin^2\theta_W$

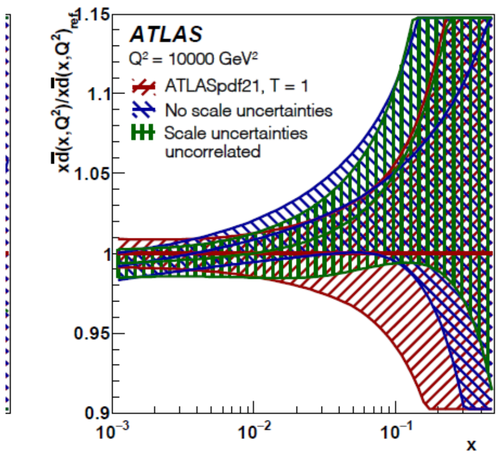


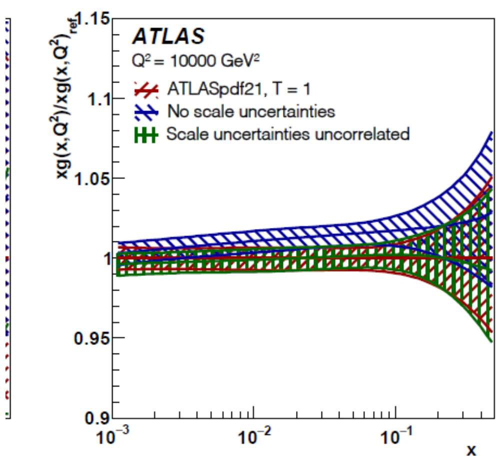
The plot shows the projected decrease in the statistical uncertainty on $\sin^2\theta_W$ with future data

But the PDF uncertainty will not decrease much

… Unless some further constraints can be applied







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

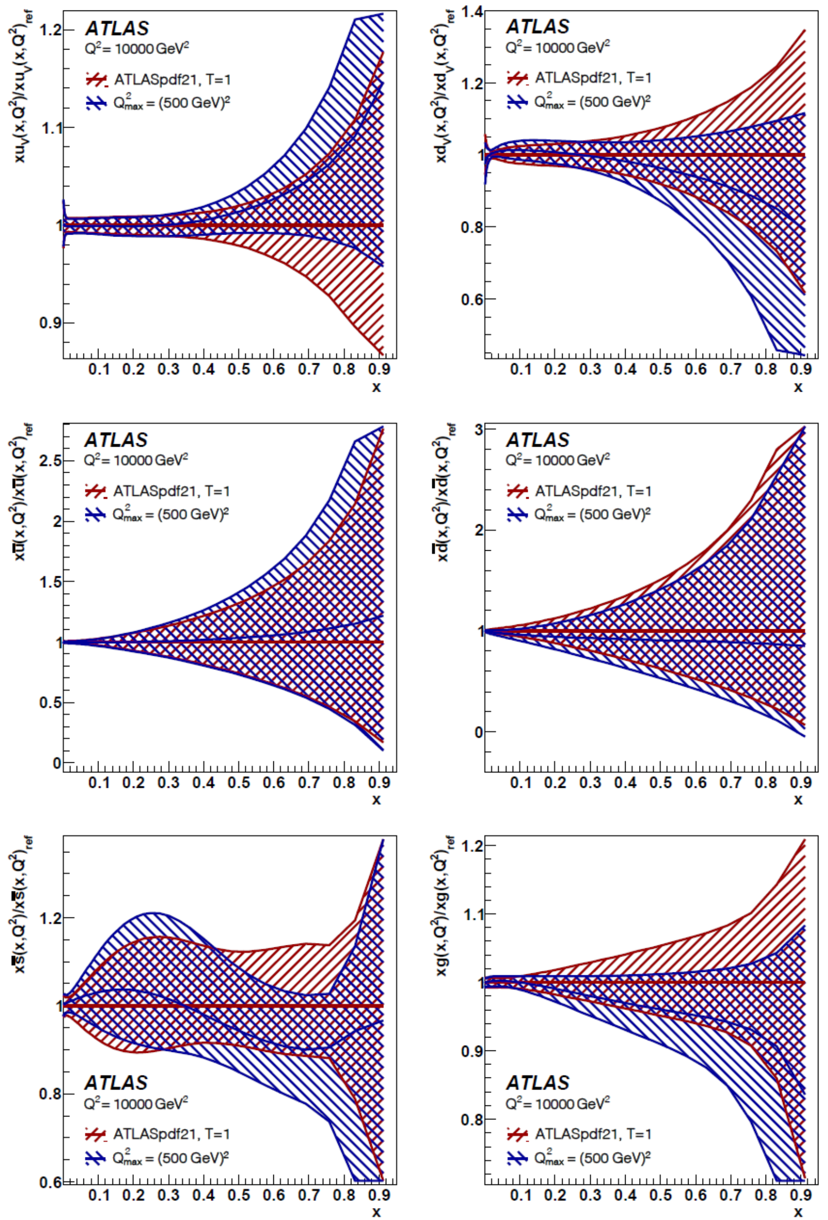
Again this is not a very big effect but it matters if we are striving for ultimate accuracy

Finally, 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.

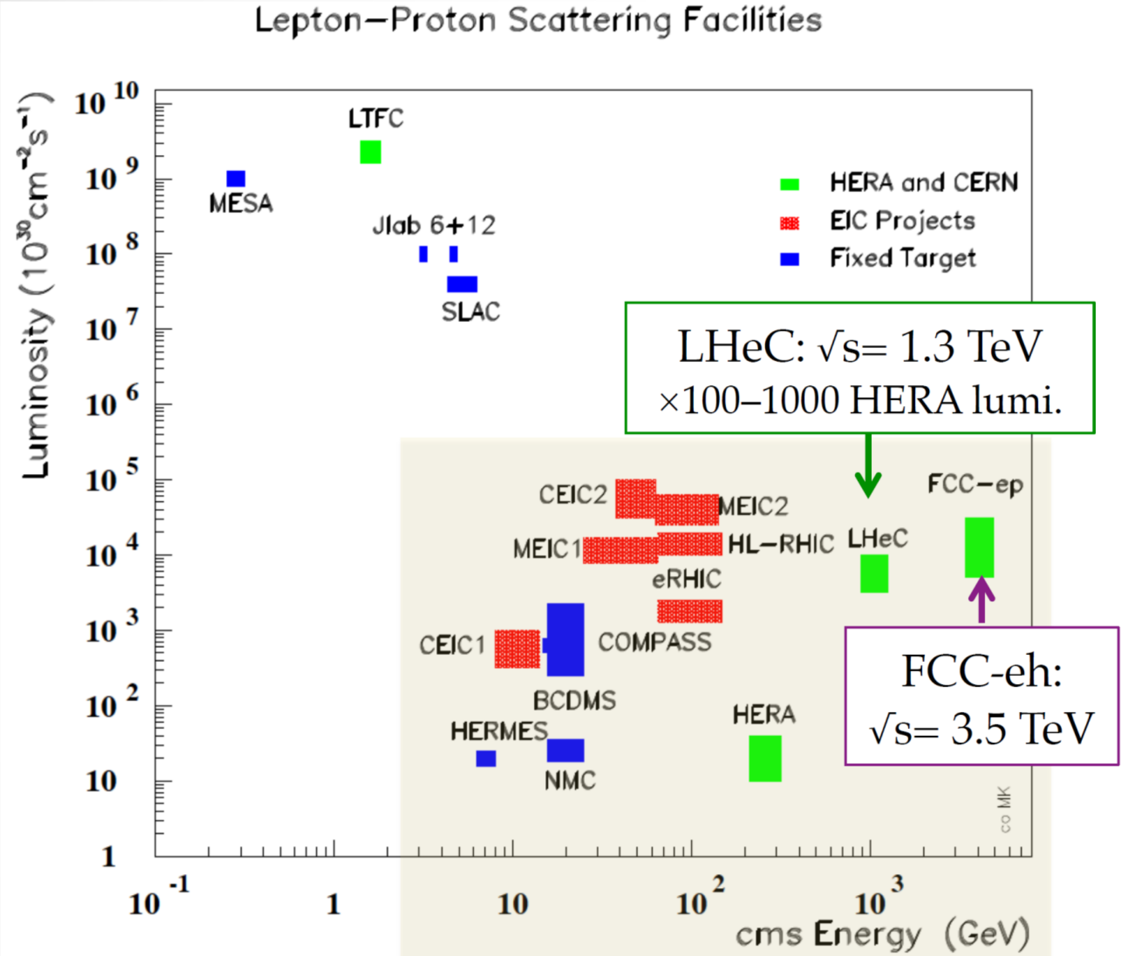
Thus ATLAS also cut all data at scale
$Q^2 > 250000\ \text{GeV}^2$
From the fit and re-evaluated the PDFs.

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.

SO HOW MAY WE ACTUALLY DO MUCH BETTER?
New DIS machines

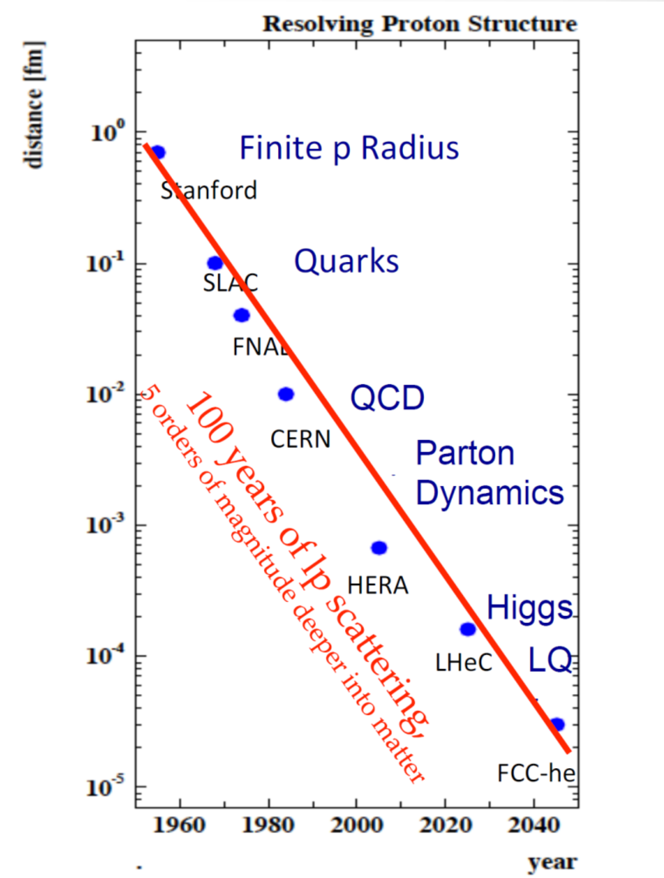


lepton-proton facilities



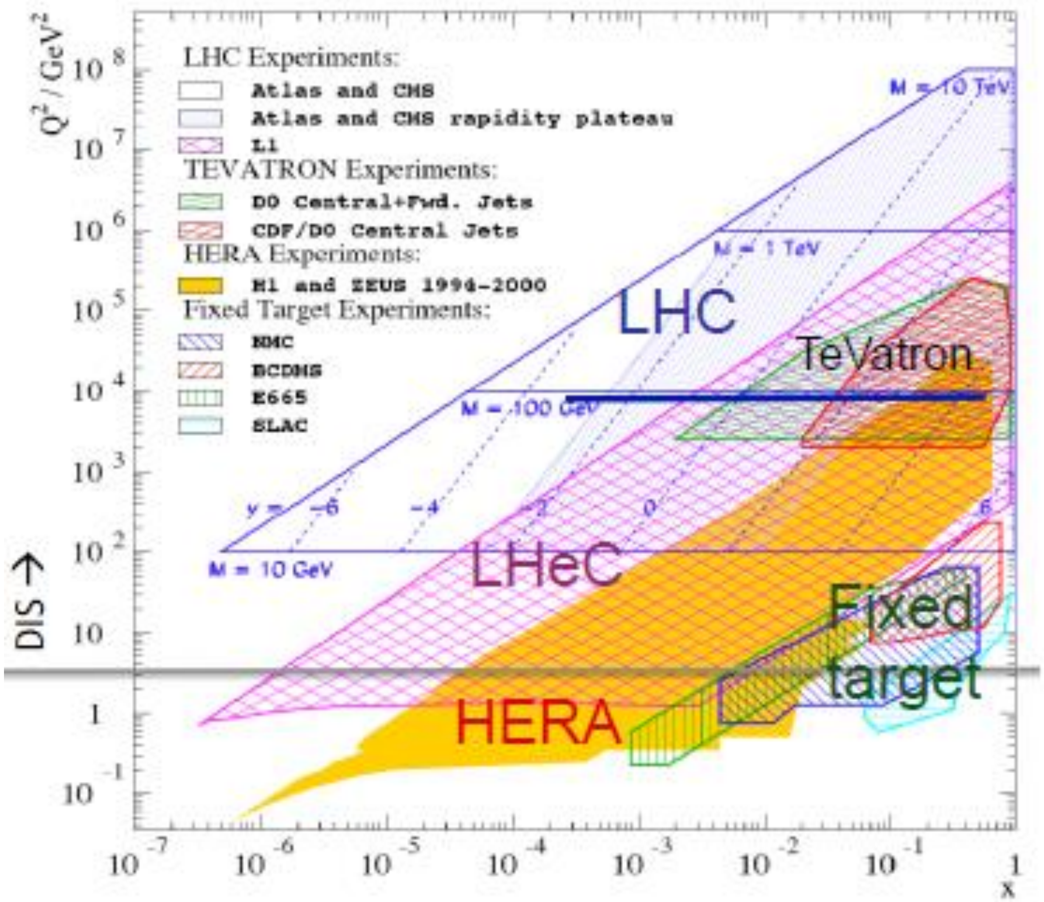
HERA–LHeC–FCC-eh:

finest microscopes, resolution as 1/Q



LHC (and other future machines eg. FCC-pp) is/will be main discovery machine

LHeC not a competitor to these; complementary; synchronous with HL-LHC; transforms them into high precision facilities



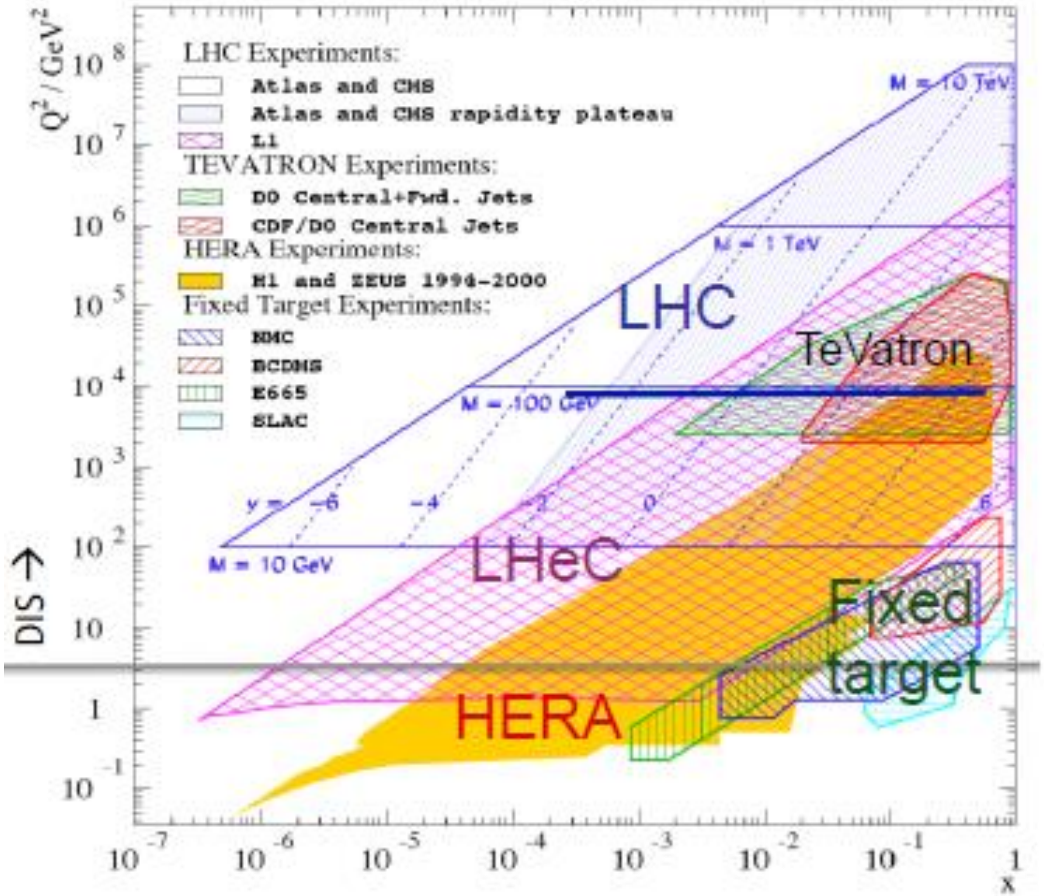
Consider the LHeC

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.

An LHeC would give a consistent data set across an enormous x, Q^2 range

The LHeC option represents an increase in the kinematic reach of Deep Inelastic Scattering and an increase in the luminosity.

- This represents a tremendous potential for the increase in the precision of Parton Distribution Functions
- **And the exploration of a kinematic region at low-x where we learn more about QCD - e.g. is there gluon saturation?— subject of another talk!!**
- Plus Precision PDFs that are needed for BSM physics



Consider the LHeC

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.

An LHeC would give a consistent data set across an enormous x, Q^2 range

The LHeC option represents an increase in the kinematic reach of Deep Inelastic Scattering and an increase in the luminosity.

- This represents a tremendous potential for the increase in the precision of Parton Distribution Functions
- **And the exploration of a kinematic region at low-x where we learn more about QCD - e.g. is there gluon saturation?— subject of another talk!!**
- Plus Precision PDFs that are needed for BSM physics

Why do PDF sets differ?

•Data sets included

•Cuts applied to remain in kinematic region of DGLAP evolution:Q^2 cut, W^2 cut, x cuts?

•Form of parametrization at Q^2_0, value of Q^2_0

•Assumptions on flavour structure of sea and valence

•heavy flavour scheme, heavy quark masses

•the value of $\alpha_S(M_Z)$ assumed, or fitted

PDFs also differ in how they evaluate their uncertainties:
some include variations of model and parametrisation assumptions
some use inflated $\chi2$ tolerances --closer to the hypothesis testing criterion–
but this is a whole lecture in itself