Challenges and Strategies in Determining Longitudinal Unpolarized Proton PDFs: From the LHC to EIC Prospects

C.-P. Yuan
Michigan State University
Wu-Ki Tung Professorship in Particle Physics

June 16, 2025 @ Academia Sinica, Taiwan

Workshop on PDFs in the EIC era

In collaboration with CTEQ-TEA members

CTEQ – Tung et al. (TEA)
in memory of Prof. Wu-Ki Tung

CTEQ-TEA group

CTEQ: The Coordinated Theoretical-Experimental Project on QCD

- CTEQ – Tung Et Al. (TEA)

in memory of Prof. Wu-Ki Tung, who co-established CTEQ Collaboration in early 90's

- Current members:

China: Sayipjamal Dulat, Ibrahim Sitiwaldi, Alim Albet (Xinjiang U.), Tie-Jiun Hou (U. of South China), Liang Han, Minghui Liu, Siqi Yang (USTC) and other coauthors.

Mexico: Aurore Courtoy (Unam, Mexico)

USA: Marco Guzzi (Kennesaw State U.), Tim Hobbs (Argonne Lab), Pavel Nadolsky (Southern Methodist U.), Yao Fu, Joey Huston, Huey-Wen Lin, Max Ponce-Chavez, Dan Stump, Carl Schmidt, Keping Xie, C.-.P Yuan (Michigan State U.) and other coauthors.

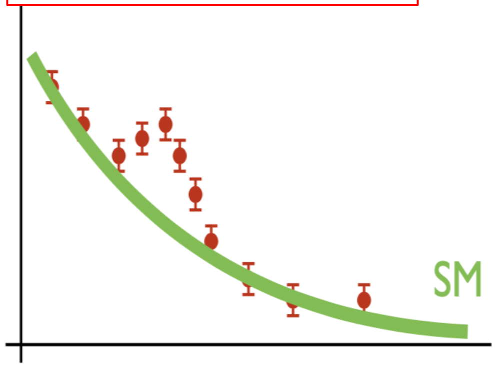
Some useful websites:

- CT18 PDFs — https://ct.hepforge.org/PDFs/ct18/
- L2 Sensitivity — https://ct.hepforge.org/PDFs/ct18/figures/L2Sensitivity/
- ePump — https://epump.hepforge.org/
- ResBos2 — https://gitlab.com/resbos2

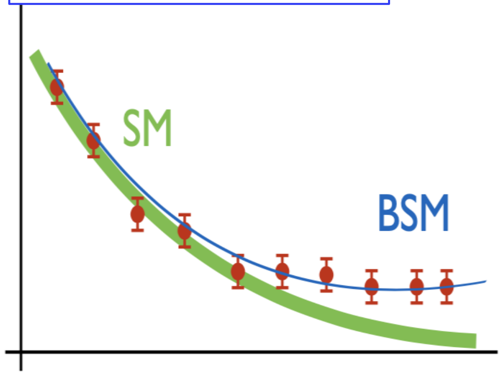
Physics Goals of the LHC

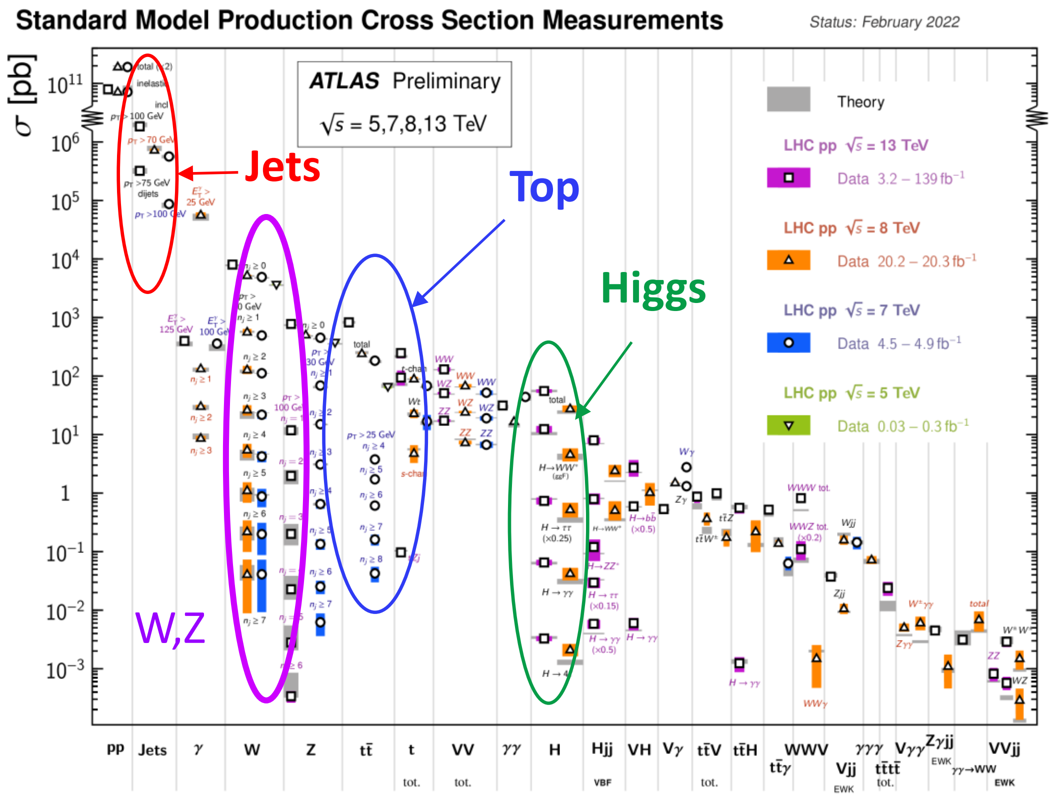
➢ **Goals:** 1. **Test Standard Model (SM)**
2. **Find New Physics (NP)**

Resonance search

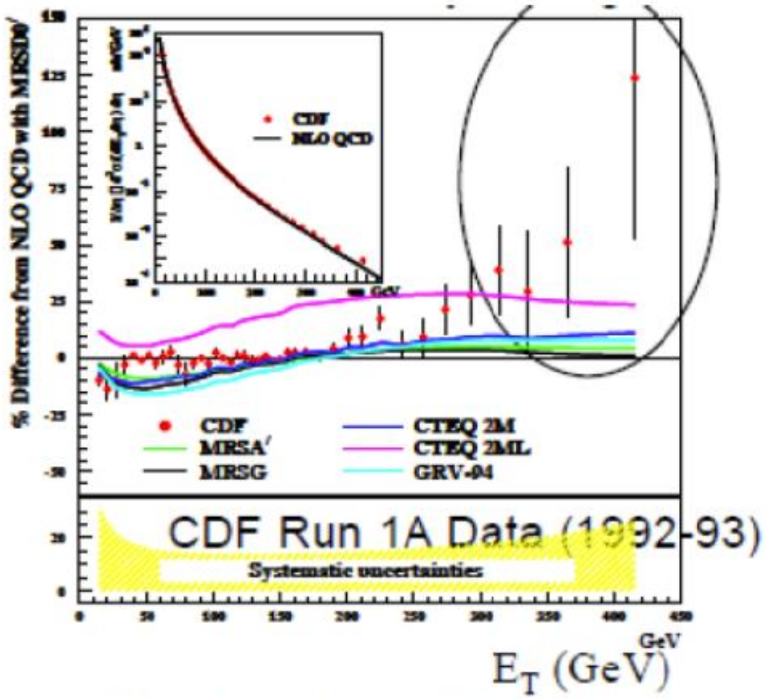


Deviation search





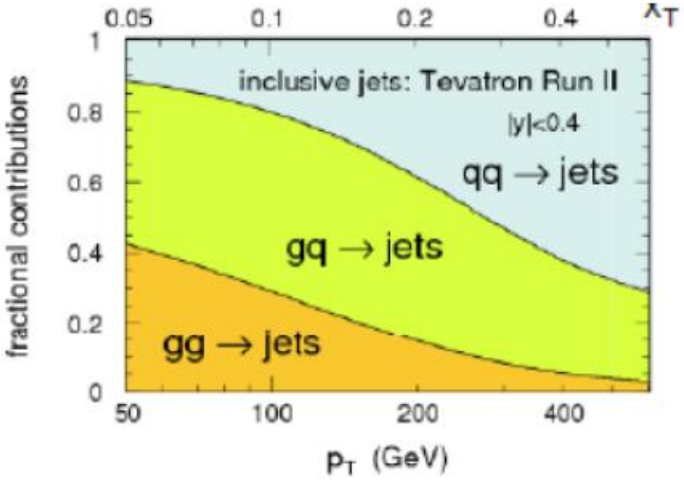
New Physics Found (in 1996) ?

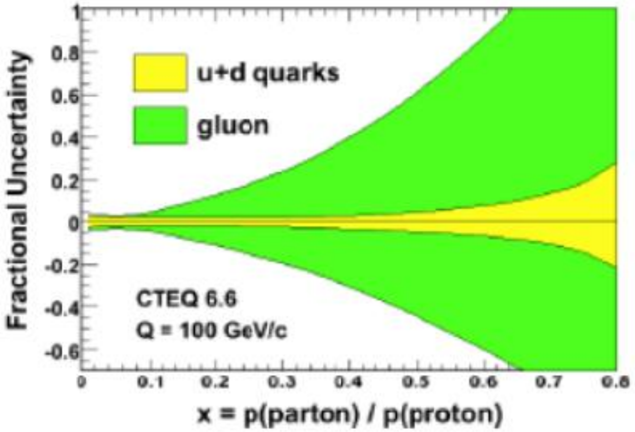


Phys. Rev. Lett. 77, 438 (1996)

High-x gluon not well known

...can be accommodated in the Standard Model





Explained by having better determined PDFs from global analysis;
no need for NP scenario yet.

J. Huston, E. Kovacs, S. Kuhlmann, J.L. Lai, J.F. Owens, D. Soper, W.K. Tung , Phys. Rev. Lett. 77 (1996) 444.

QCD Factorization Theorem and Parton Distribution Functions

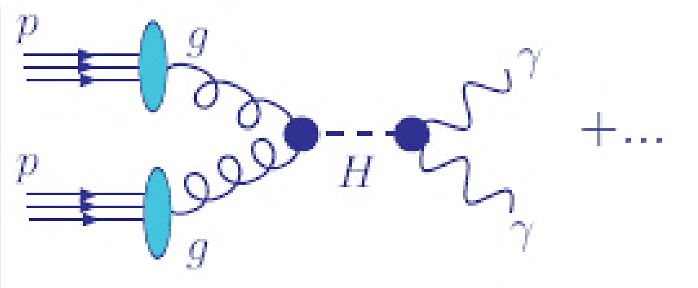
Hessian PDF eigenvector (EV) sets

vs

Monte Carlo (MC) PDF replicas

QCD Factorization Theorem and PDFs

$$\sigma_{pp\to H\to\gamma\gamma X}(Q) = \sum_{a,b=g,q,\bar{q}} \int_0^1 d\xi_a \int_0^1 d\xi_b \hat{\sigma}_{ab\to H\to\gamma\gamma}\left(\frac{x_a}{\xi_a}, \frac{x_b}{\xi_b}, \frac{Q}{\mu_R}, \frac{Q}{\mu_F}; \alpha_s(\mu_R)\right) \times f_a(\xi_a, \mu_F) f_b(\xi_b, \mu_F) + O\left(\frac{\Lambda_{QCD}^2}{Q^2}\right)$$



$\hat{\sigma}$ is the hard cross section; computed order-by-order in $\alpha_s(\mu_R)$
$f_a(x, \mu_F)$ is the distribution for parton a with momentum fraction x, at scale μ_F

$$f_{a/h}(x, Q)$$

Unpolarized collinear parton distribution functions (PDFs) $f_{a/h}(x, Q)$ are associated with probabilities for finding a parton a with the "+" momentum xp^+ in a hadron h with the "+" momentum p^+ for $p^+ \to \infty$, at a resolution scale $Q > 1$ GeV .

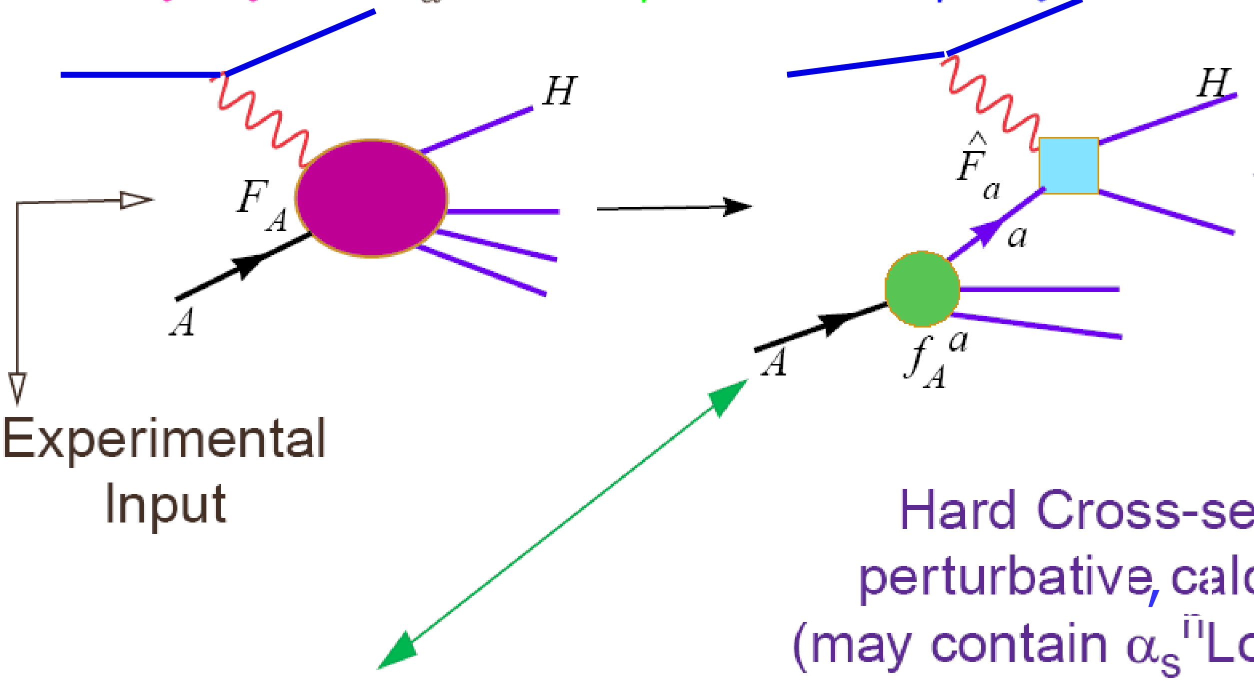
The (unpolarized) collinear PDFs describe long-distance dynamics of (single parton scattering) in high-energy collisions.

Lepton-hadron Sc.

C T E Q

Master Equation for QCD Parton Model – the Factorization Theorem

$$F_A^{\lambda}\left(x,\frac{m}{Q},\frac{M}{Q}\right) = \sum_a f_A^a\left(x,\frac{m}{\mu}\right) \otimes \hat{F}_a^{\lambda}\left(x,\frac{Q}{\mu},\frac{M}{Q}\right) + \mathcal{O}\left(\left(\frac{\Lambda}{Q}\right)^2\right)$$



Experimental Input

Theory Input

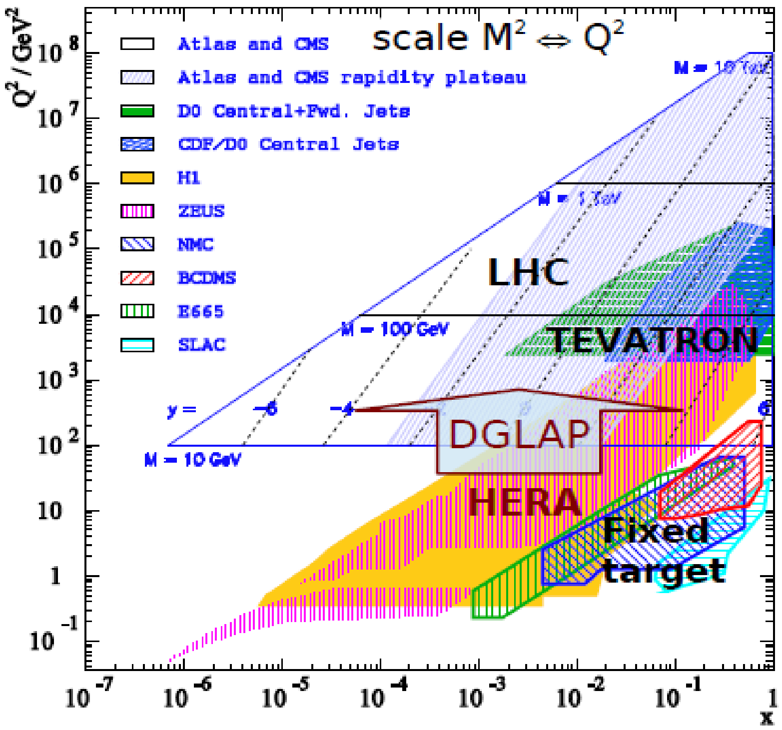
Hard Cross-section
perturbative, calculable
(may contain $\alpha_s^n \mathrm{Log}^n(M/Q)$)

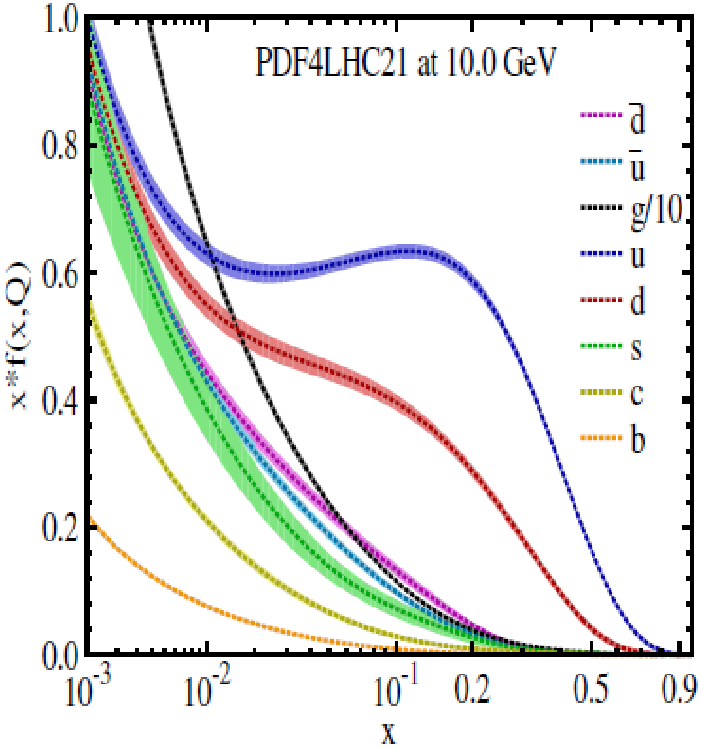
universal Parton Dist. Fn.
Non-Perturbative Parametrization at Q_0
DGLAP Evolution to Q

Extracted by global analysis

Global analysis of PDFs

✦ PDFs are usually extracted from global analysis on variety of data, e.g., DIS, Drell-Yan, jets and top quark productions at fixed-target and collider experiments, with increasing weight from LHC, together with SM QCD parameters **[see 1709.04922, 1905.06957 for recent review articles]**





parameter variations

$\alpha_S(M_Z)$

Mc, Mb, Mt

QCD/EW corrections

nuclear corrections

EW parameters

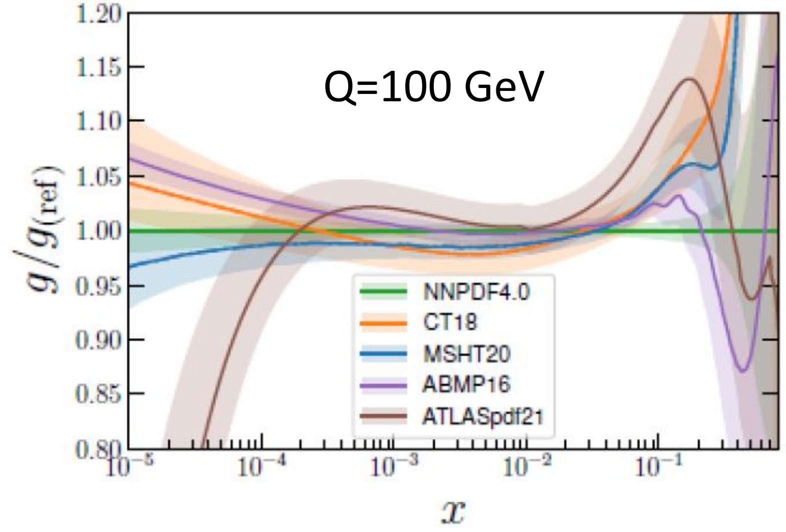
New Physics

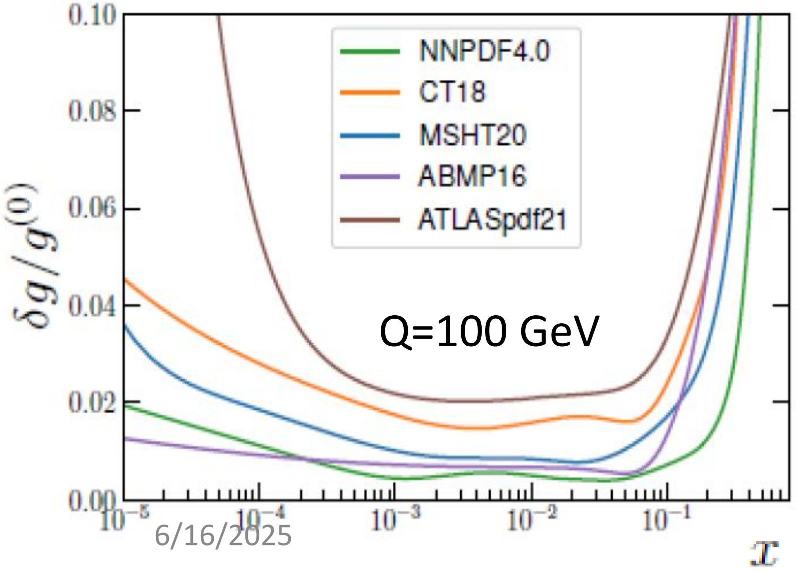
❖ diversity of the analysed data are important to ensure flavor separation and to avoid theoretical/experimental bias; possible extensions to include EW parameters and possible new physics for a self-consistent determination

❖ alternative approach from lattice QCD simulations, for various PDF moments or PDFs directly calculated in x-space with large momentum effective theory or pseudo-PDFs [2004.03543]

Comparing predictions from various QCD global analysis groups

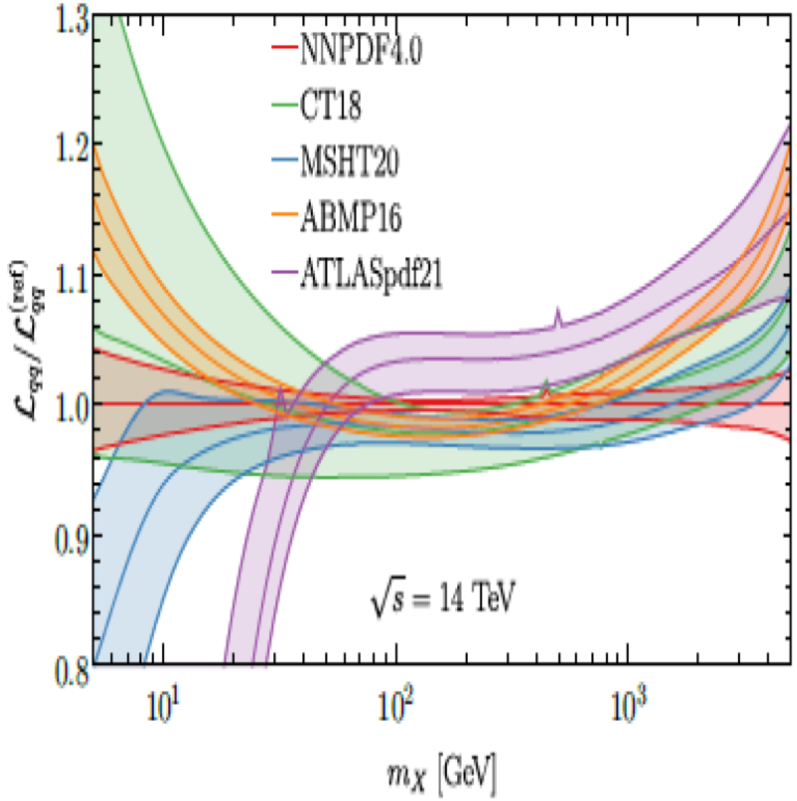
Snowmass 2021, 2203.13923

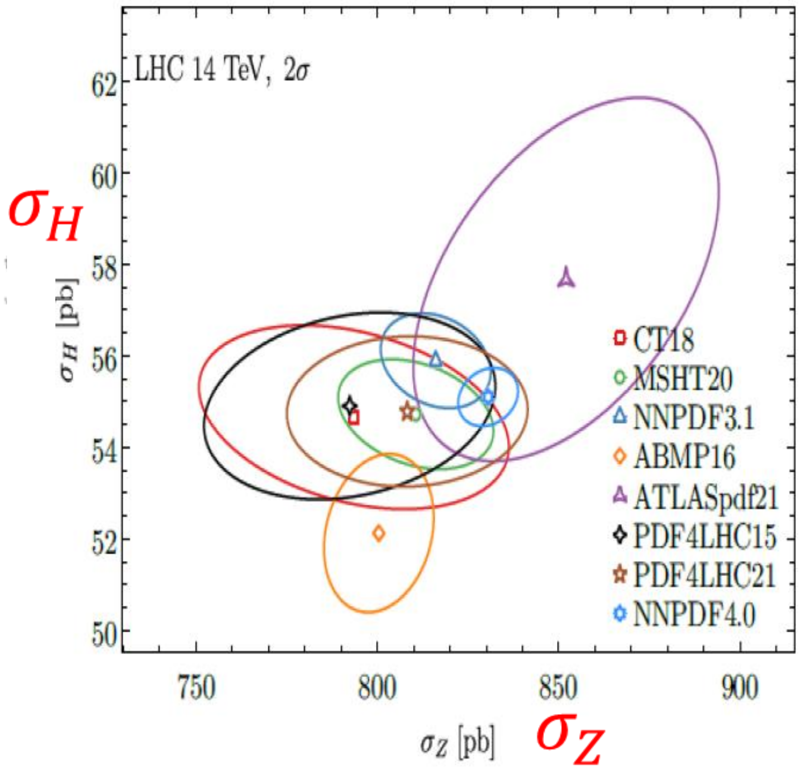




Smaller PDF errors lead to smaller

PDF luminosity errors, then smaller PDF-induced errors in cross sections.

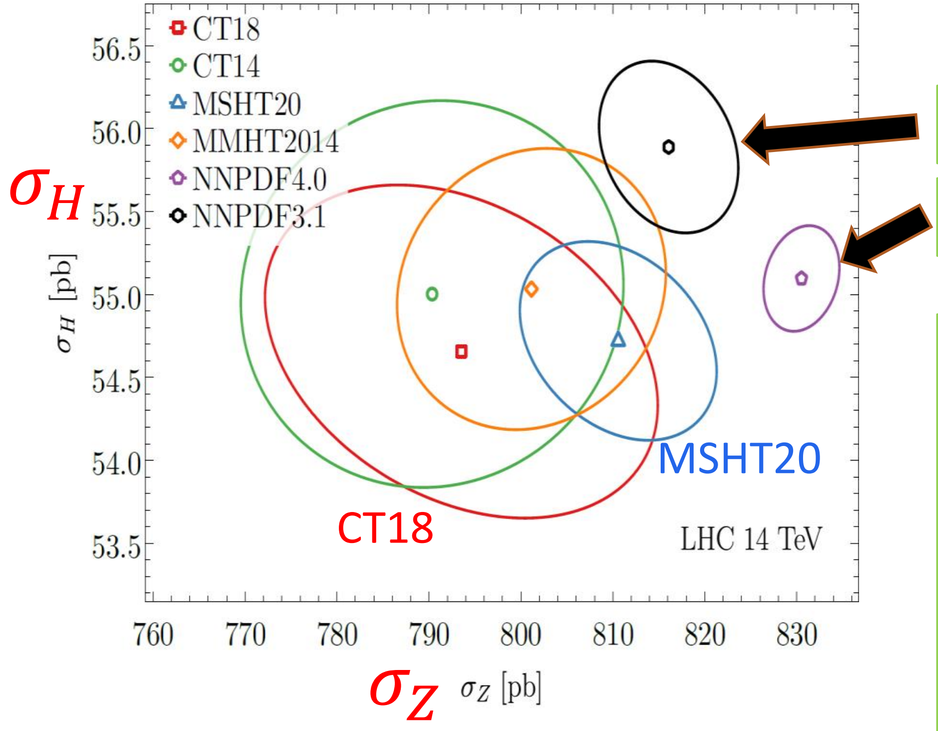




Comparing predictions from various QCD global analysis groups

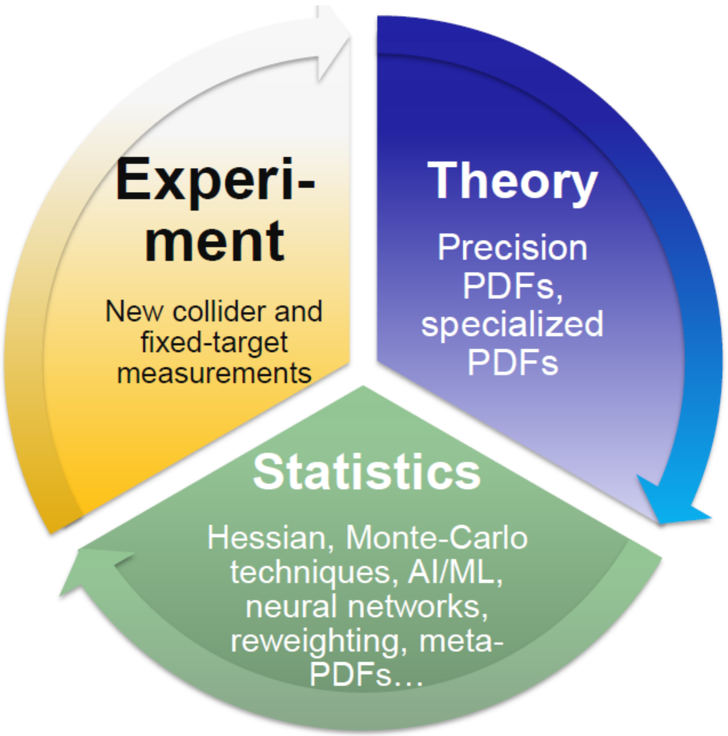
Snowmass 2021, 2203.13923

The PDF-induced errors @ 68% CL in $gg \to h$ and $q\,\bar{q} \to Z$ NNLO cross sections



Their predictions do not overlap at 1σ level.

Due to different choices of



Components of a global QCD fit

Different (though mostly consistent) predictions on

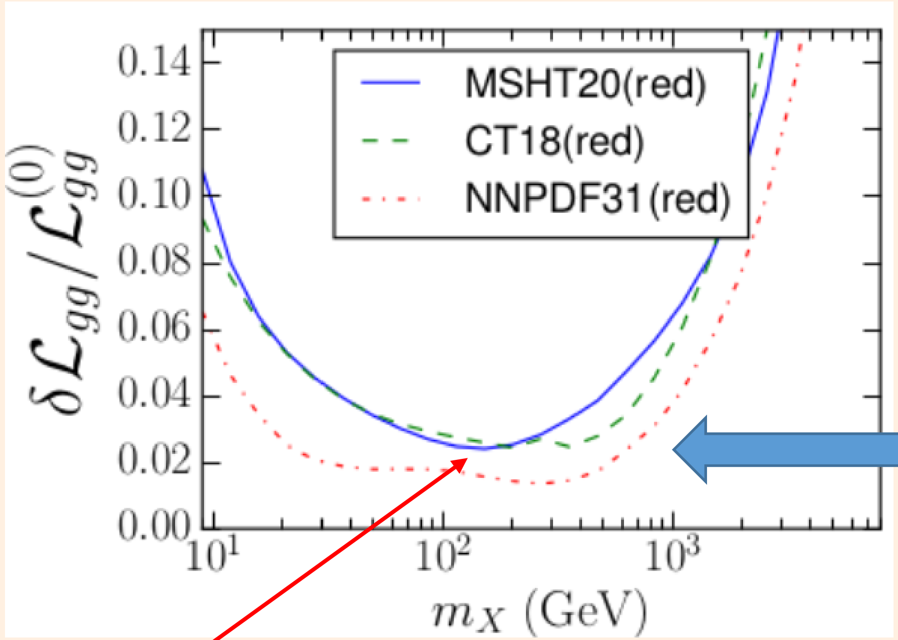
- central values and error estimates of PDFs,
- parton luminosities,
- physical cross sections, and
- various correlations among PDFs and data …

Benchmark Study: PDF4LHC21

arXiv:2203.05506

Relative PDF uncertainties on the gg luminosity at 14 TeV in three PDF4LHC21 fits to the **identical** reduced global data set

arXiv:2203.05506



× 1.5 – 2 difference

- Each analysis group (CT, MSHT, NNPDF) used the same (reduced) data sets and same theory predictions in the analysis

Smaller error size found by NNPDF

- NNPDF3.1' and especially 4.0 (based on the NN's+ MC technique) tend to give smaller uncertainties in data-constrained regions

The size of PDF error estimates depends on the methodology of global analysis adopted by the PDF fitting group.

Sources of PDF errors

Factorization Theorem:

Data = PDFs $\otimes$ Hard part cross sections (Wilson coeff.)

Experimental errors:

- Statistical
- Systematic
  - uncorrelated
  - correlated
- χ^2definition (experimental or t_0)
- Possible tensions among data sets

Extracted with errors, dependent of methodology of analysis

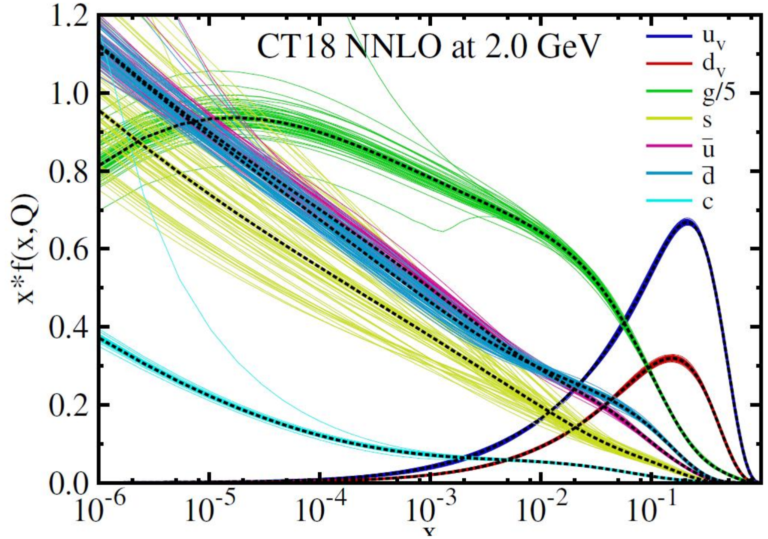
- Non-perturbative parametrization forms of PDFs
- Additional theory prior
- Choice of Tolerance (T^2) value

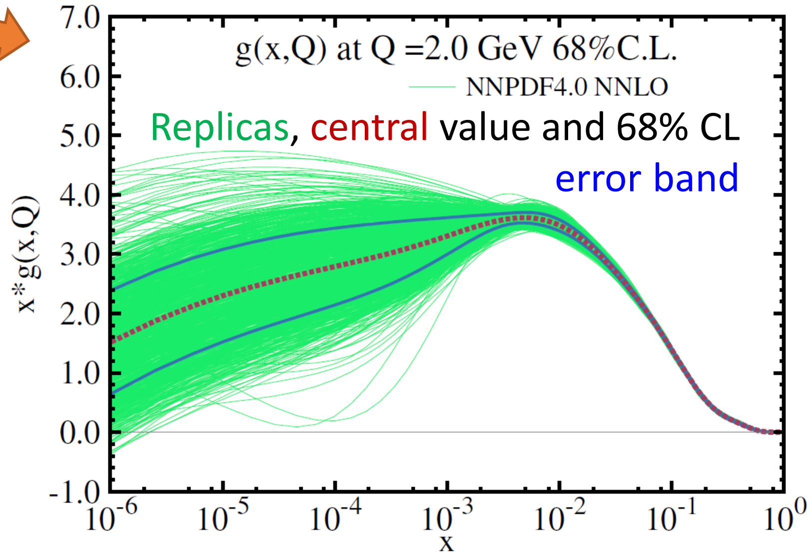
Theoretical errors:

- Which order: (NLO, NNLO, …, resummation – BFKL, qT, threshold)
- Which scale: (μ_F, μ_R)
- Which code: (antenna subtraction, sector decomposition,…, qT, N-jettiness,,…)
- Monte Carlo error: (most efficient implementation,…)

How to estimate PDF errors in QCD global analysis

- ➢ Error estimate is important.
- ➢ Two different methodology in global analysis
- ❖ Hessian PDF eigenvector (EV) sets,
  from analytic parametrizations of PDFs
  (ABM, CTEQ, HERA, MSHT, …)
- ❖ Monte Carlo (MC) PDF replicas,
  from Neural Network (NN) parametrizations
  (NNPDF)
- ➢ Both methods assume some non-perturbative input of PDFs at the initial Q_0 scale, around 1 GeV. (analytical parametrization vs. NN architecture)
- ➢ They are two powerful and complementary representations.
- ➢ Hessian PDFs can be converted into MC ones, and vice versa.





How to quantify PDF uncertainties

was first introduced in 2001 by
Jon Pumplin, Dan Stump and Wu-Ki Tung
@ Michigan State University

hep-ph/0101032

Uncertainties of predictions from PDFs:

The Hessian method

hep-ph/0101051

Uncertainties of predictions from PDFs:

The Lagrange multiplier method

$$\chi^2 = \chi_0^2 + \sum_{i,j} H_{ij} (a_i - a_i^0)(a_j - a_j^0)$$

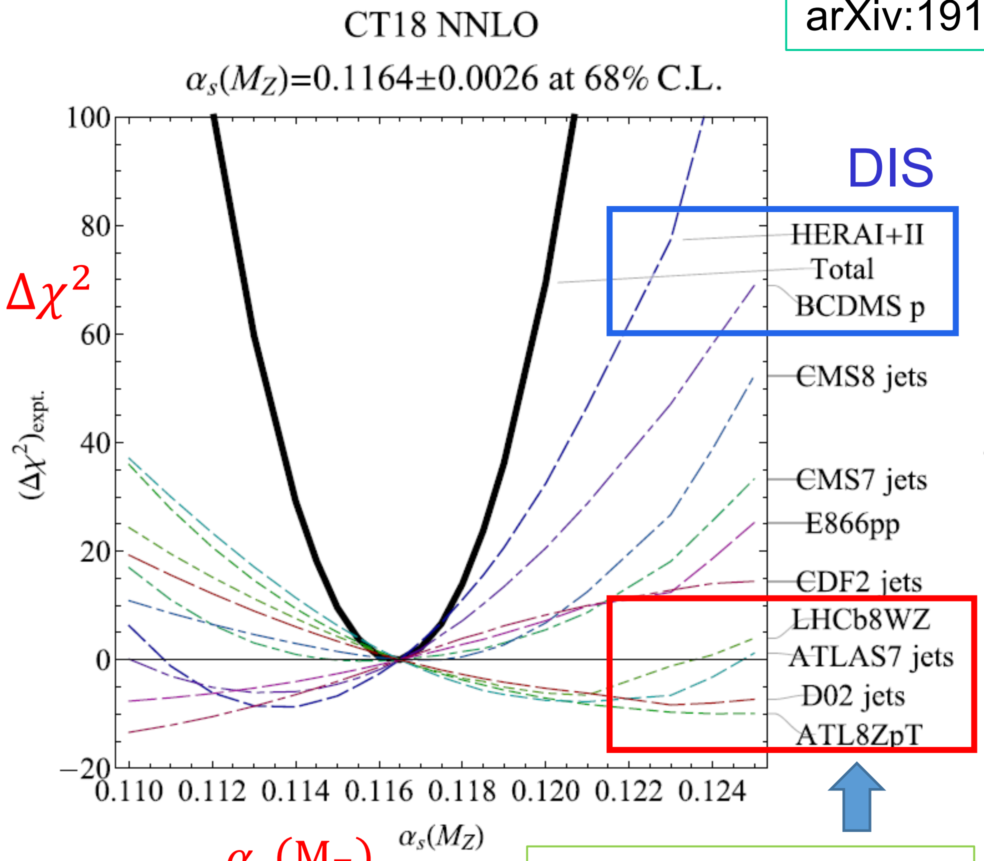
It was first implemented in CTQE6 PDFs.

They were used to determine uncertainty of PDFs, physical cross sections, α_s and m_t as well as exploring tensions among data sets in the CTEQ-TEA analysis.

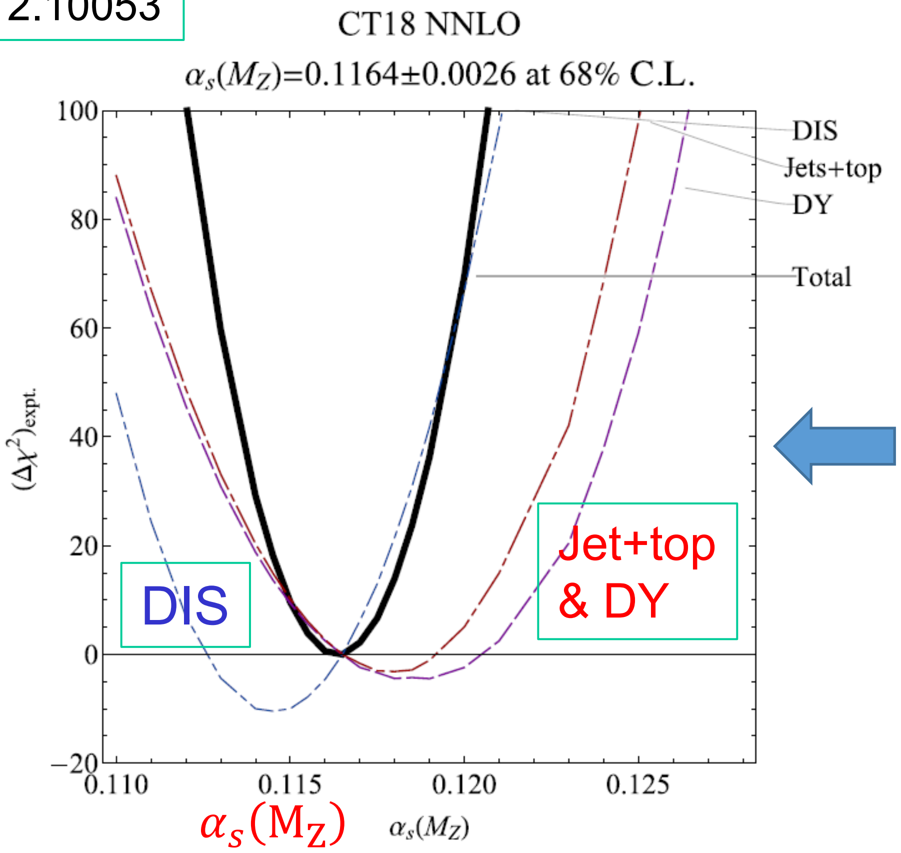
Lagrange Multiplier scan

To explore PDF-induced errors in the determination of $\alpha_s(\mathrm{M_Z})$ and tensions among data sets included in the fit

arXiv:1912.10053



Prefer larger $\alpha_s(M_Z)$



The scan of $\alpha_s(\mathrm{M_Z})$ values in CT18 NNLO PDF analysis.

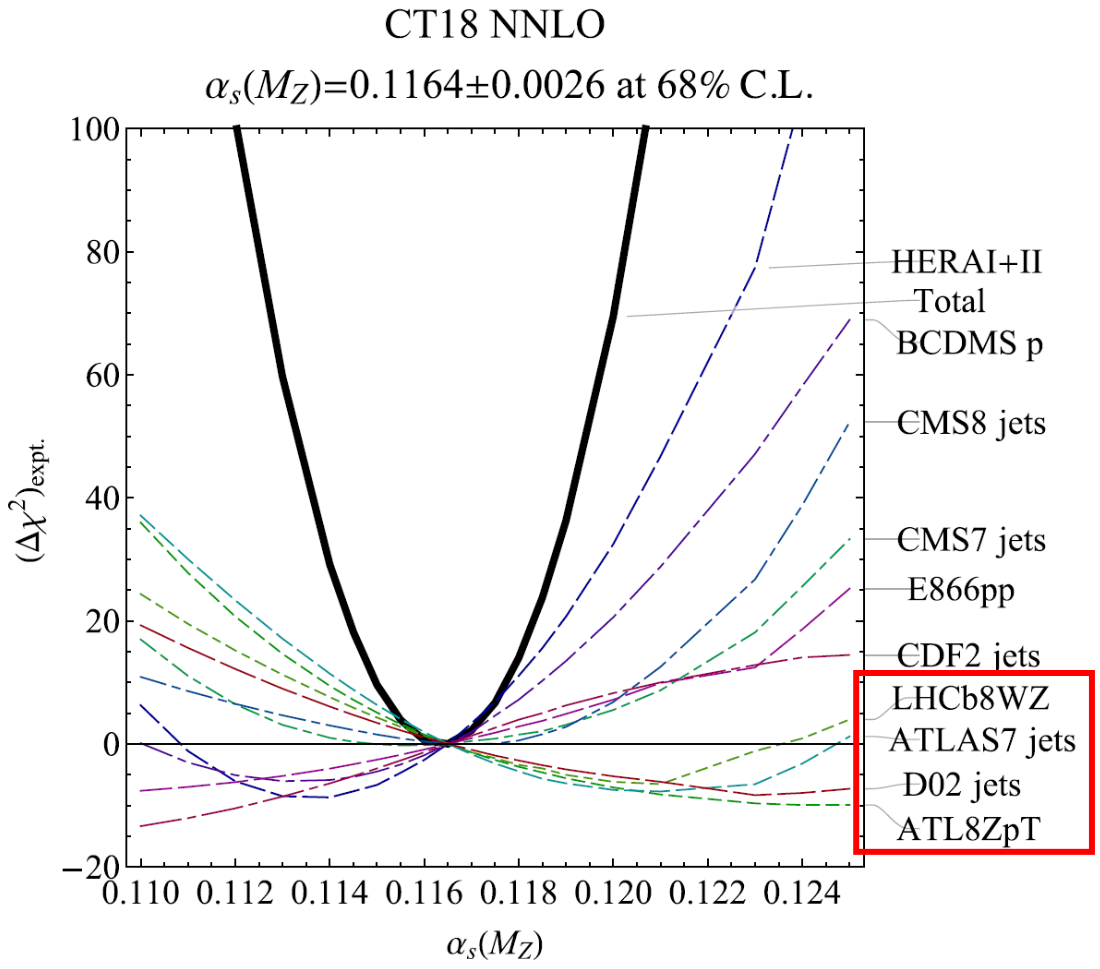
- The opposing pulls (i.e., tensions) of DIS and jet+top&DY experiments significantly exceed $\Delta\chi^2 = 1$ variation, as implied by the simplest statistical framework.
- Require a large value of Tolerance T^2, the maximum allowed total $\Delta\chi^2$, with $\Delta\chi^2 > 1$
- To agree with the error in $\alpha_s(\mathrm{M_Z})$, 0.007, as provided by PDG (in 2010), without including hadron collider data in the fit, it requires

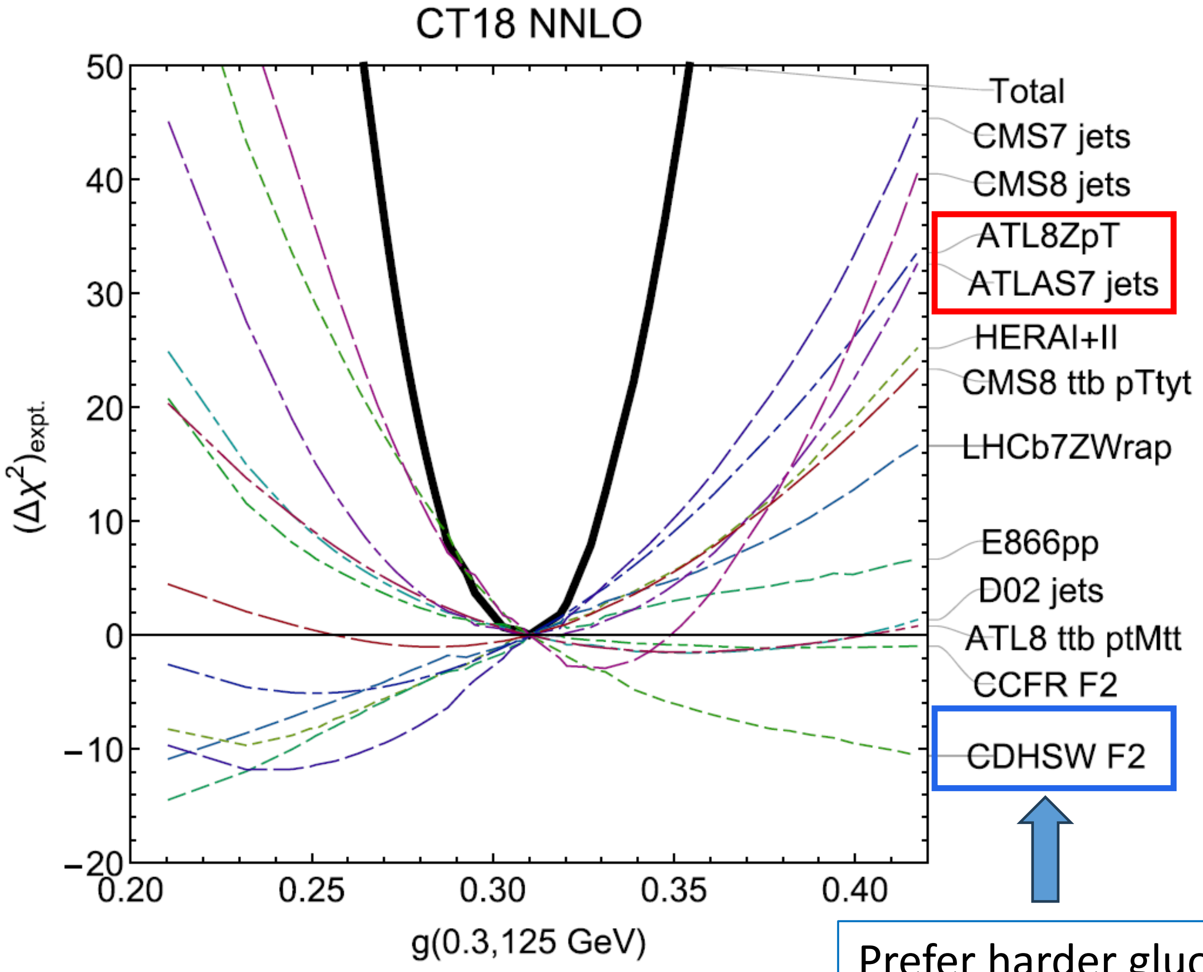
$$1 < \Delta\chi^2 \leq (5 - 10)$$

Possible tensions among experimental data sets

Require $\Delta\chi^2 > 1$

Tensions among experimental data sets





Prefer harder gluon PDF at large x.

Tolerance (T^2) values in various PDF analysis groups

- Tolerance T^2, the maximum allowed total $\Delta\chi^2$ value away from the best (or central) fit, was introduced to account for the sampling of
  - non-perturbative parametrization of PDFs (or NN architecture, smoothness, positivity) and
  - the allowed PDF variation due to various choices of data sets and theory calculations, etc.

- Roughly speaking, at the 68% CL,
  - CTEQ-TEA (CT) Tier-1 $T^2 \sim 30$
  - MSHT dynamical $T^2 \sim 10$
  - NNPDF effective $T^2 \sim 2$ (for MC replicas and their Hessian representation)

- A smaller T^2 value typically yields a smaller PDF error estimate.

CT tolerance includes both Tier-1 and Tier-2 contributions.

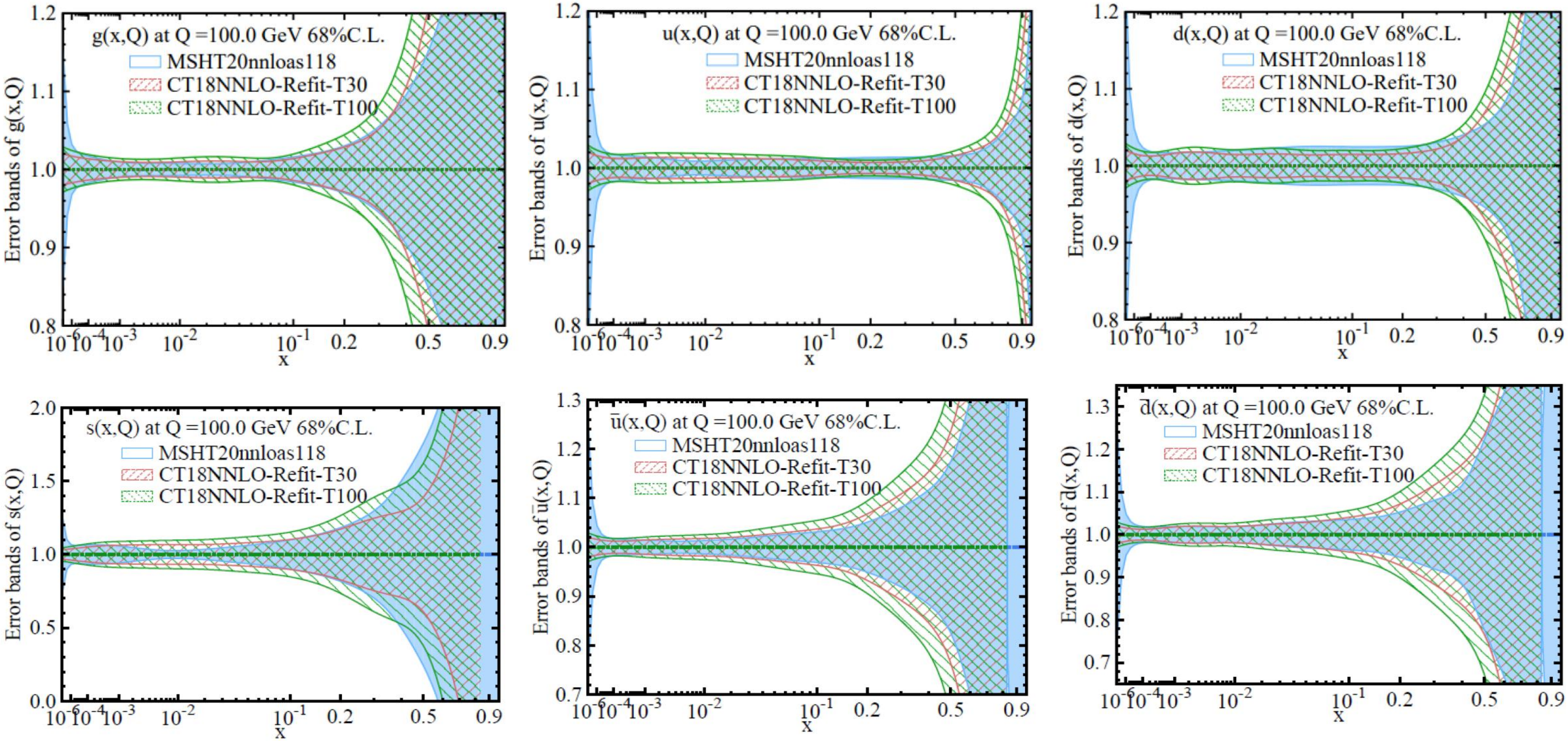
To reduce PDF uncertainty, one must maximize both

PDF fitting accuracy
(accuracy of experimental, theoretical and other inputs)

and

PDF sampling accuracy
(adequacy of sampling in space of possible solutions)

Compare PDF error bands with T = 37 or 10 (of CT18) and MSHT20, at 68% CL



The PDF errors of MSHT20 and CT18 (T=10) are alike in many cases.

Hessian profiling of CT and MSHT PDFs cannot use $\Delta\chi^2 = 1$

C T E Q

ATLAS-CONF-2023-015

The statistical analysis for the determination of $\alpha_s(m_Z)$ is performed with the xFitter framework [60]. The value of $\alpha_s(m_Z)$ is determined by minimising a χ^2 function which includes both the experimental uncertainties and the theoretical uncertainties arising from PDF variations:

$$\chi^2(\beta_{\rm exp},\beta_{\rm th}) = \sum_{i=1}^{N_{\rm data}} \frac{\left(\sigma_i^{\rm exp} + \sum_j \Gamma_{ij}^{\rm exp}\beta_{j,\rm exp} - \sigma_i^{\rm th} - \sum_k \Gamma_{ik}^{\rm th}\beta_{k,\rm th}\right)^2}{\Delta_i^2} + \sum_j \beta_{j,\rm exp}^2 + \sum_k \beta_{k,\rm th}^2 .$$

profiling of CT and MSHT PDFs requires to include a tolerance factor $T^2 > 10$ as in the ePump code

arXiv: 1907.12177

arXiv:1912.10053

- xFitter profiling uses $\Delta\chi^2 = 1$, by default.
- For CT (or MSHT) PDFs, using $\Delta\chi^2 = 1$ in profiling is equivalent to assigning a weight of about 30 (or 10) to the new data included in the fit. Hence, it will overestimate the impact of new data.
- CT: $T^2 \sim 30$; MSHT: $T^2 \sim 10$

When profiling a new experiment with the prior imposed on PDF nuisance parameters $\lambda_{\alpha,th}$:

$$\chi^2(\vec{\lambda}_{\rm exp},\vec{\lambda}_{\rm th}) = \sum_{i=1}^{N_{pt}} \frac{\left[D_i + \sum_\alpha \beta_{i,\alpha}^{\rm exp}\lambda_{\alpha,\rm exp} - T_i - \sum_\alpha \beta_{i,\alpha}^{\rm th}\lambda_{\alpha,\rm th}\right]^2}{s_i^2} + \sum_\alpha \lambda_{\alpha,\rm exp}^2 + \sum_\alpha T^2\lambda_{\alpha,\rm th}^2 .$$

$$\beta_{i,\alpha}^{\rm th} = \frac{T_i(f_\alpha^+) - T_i(f_\alpha^-)}{2},$$

new experiment

priors on expt. systematics and PDF params

Impact of SIDIS data

- Di-muon data
- Couple to final state fragmentation function and decay branching ratio

Impact of NuTeV and CCFR SIDIS dimuon data

arXiv:1907.12177

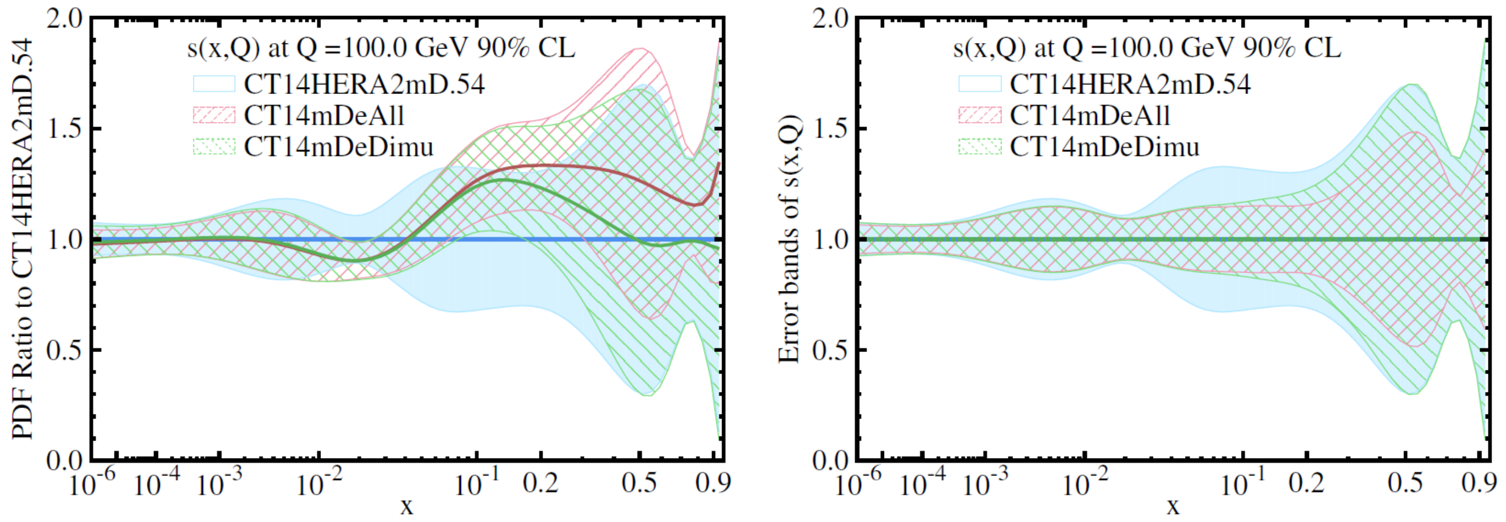
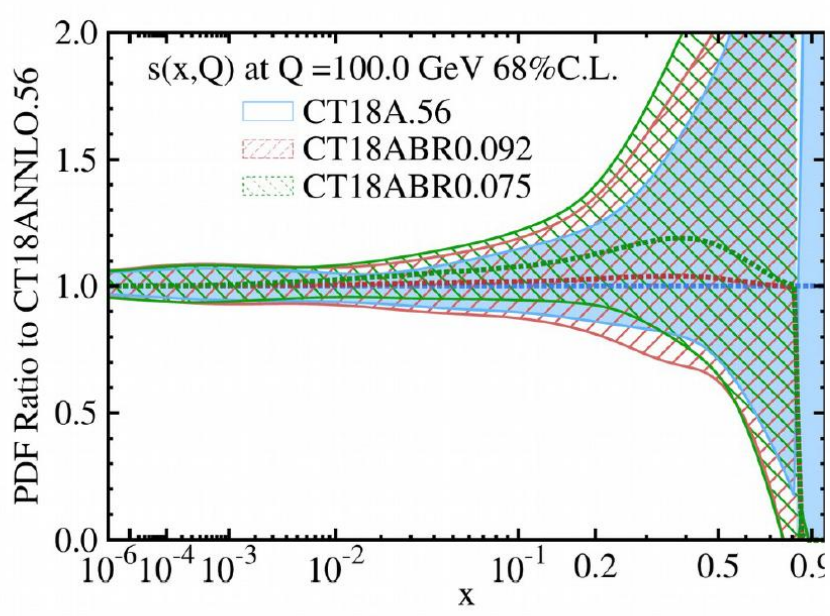


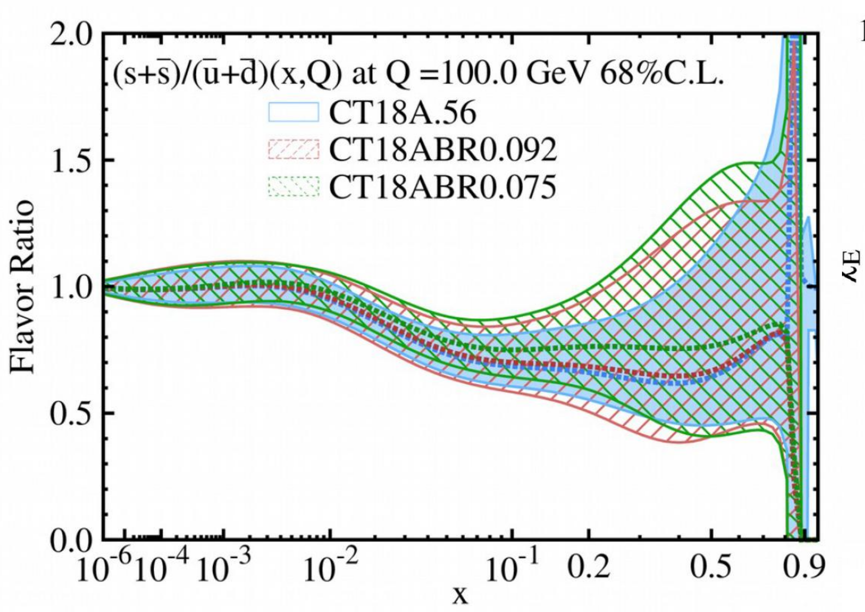
FIG. 16: Comparison of ePump-updated s-PDF, at $Q = 100$ GeV. CT14mDeDimu is obtained by adding only the DIS charged current dimuon data (NuTeV [18], and CCFR [19]) to CT14HERA2mD with ePump.

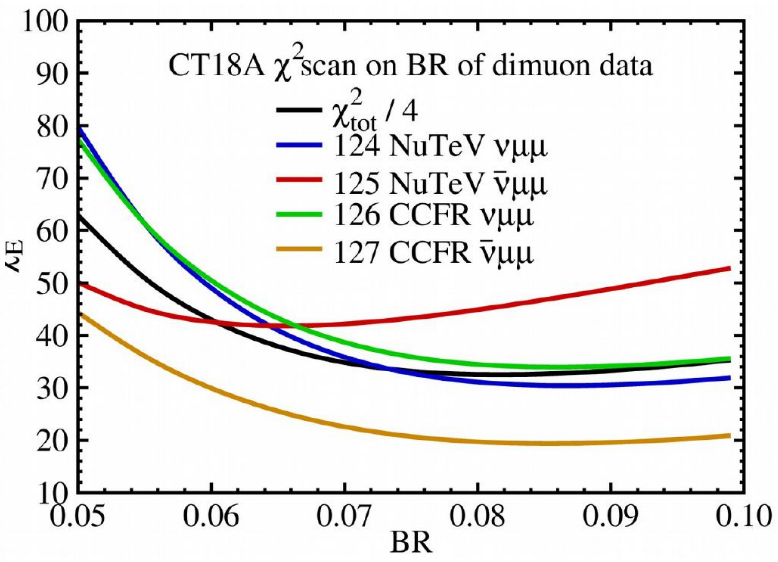
- NuTeV and CCFR di-muon data provide important constraints on s and $\bar{s}$ PDFs at large x.
- They are SIDIS data, so that constraints on PDFs depend on the modeling of final state fragmentation and the value of $c \rightarrow \mu$ decay branching ratio R.
- The LHC W and Z data can constrain s and $\bar{s}$ PDFs at $x \sim 10^{-2}$.

Impact of NuTeV and CCFR SIDIS dimuon data

arXiv:2211.11064







# fitname	248	124	125	126	127
CT18A	87.55	31.91	52.81	35.59	20.88
CT18ABR0.092	85.16	31.49	49.97	34.65	20.53
CT18ABR0.075	79.02	32.83	43.25	34.64	21.30

- In CT18A, the $c \to \mu$ decay branching ratio R is taken to be 0.099
- No noticeable changes by varying R from 0.099 to 0.092

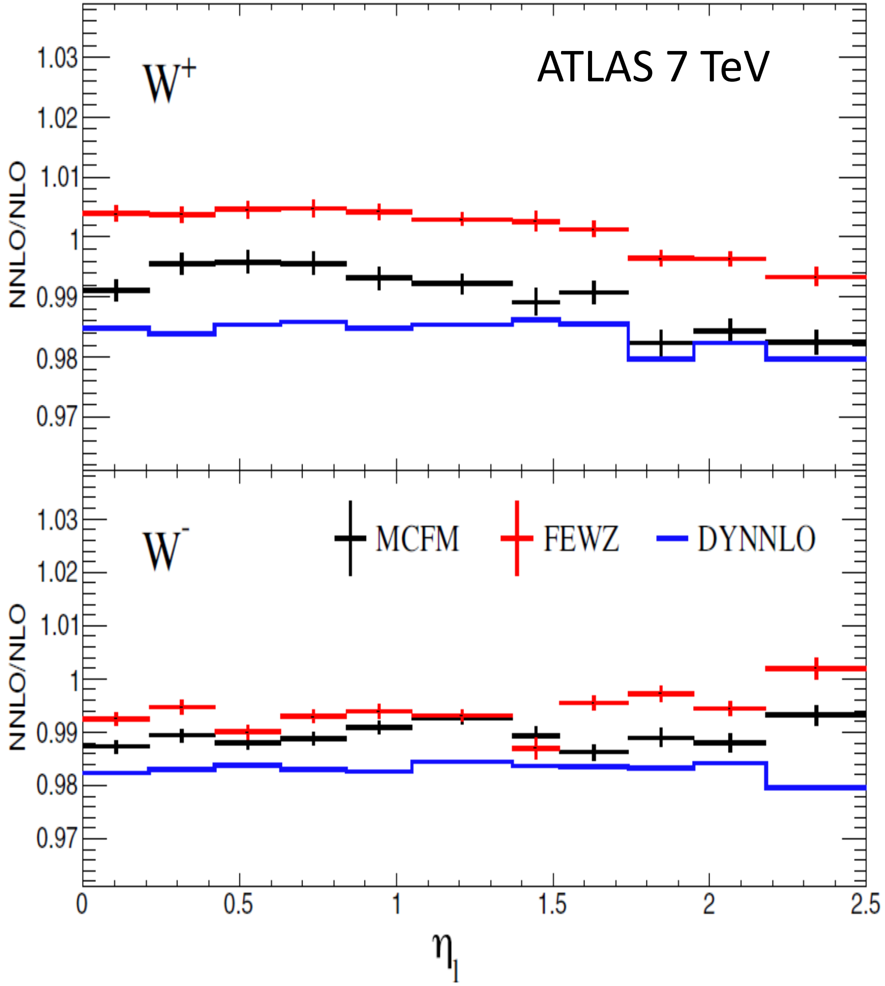
(ID=248 refers to ATLAS 7 TeV W/Z data.)

Impact of higher order theoretical predictions

- Theoretical errors can be larger than experimental errors, even at the NNLO in QCD interaction.

Different (NNLO) theory predictions from various codes; require $\Delta\chi^2 > 1$

arXiv:1912.10053



- Compare predictions of three different codes:
  - FEWZ (sector decomposition)
  - MCFM (N-jettiness)
  - DYNNLO (qT)
- Their predictions agree well at NLO.
- Their NNLO predictions agree well for inclusive cross sections (without imposing kinematic cuts).
- Their NNLO predictions for fiducial cross sections (with kinematic cuts) can differ at percent level, while the statistical error of the data is at the sub-percent level.

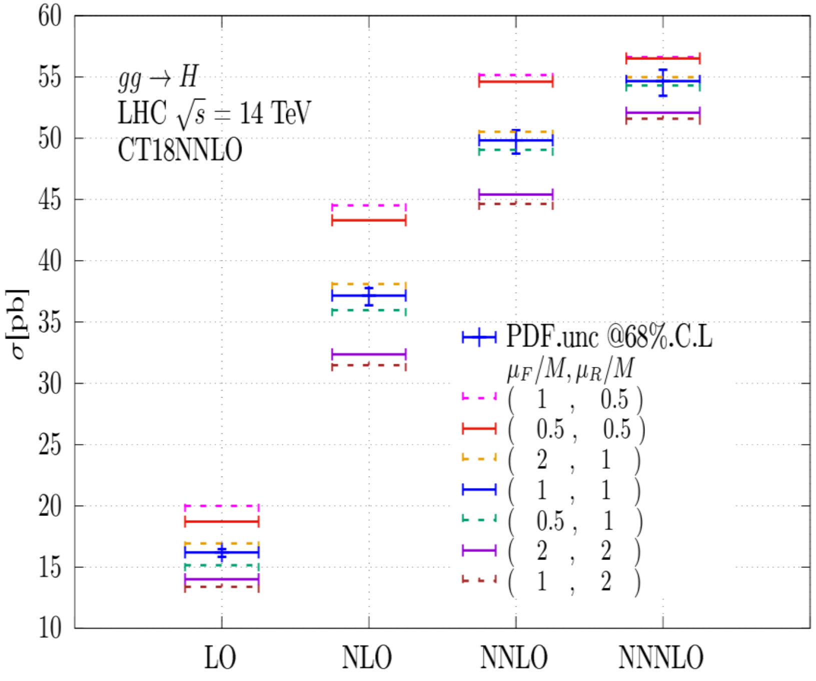
✓ The resulting PDFs from various theory predictions only differ slightly, when including this data in the CT18A fit.

✓ The kind of theory uncertainty is accounted for by choosing a larger Tolerance value than 1 (i.e., $\Delta\chi^2 > 1$) at the 68% CL.

Missing higher order (MHO) uncertainty estimated by scale variation

- General wisdom: Varying a “typical scale” by a factor of 2 (or 7-point scales) to estimate missing higher order (MHO) contribution.
- This wisdom does not always work. Namely, varying the factorization and normalization scales by a factor of 2 cannot accurately estimate MHO contribution.

The complete higher order calculations in QCD, EW, and the mixed QCD+EW are all very important for making precision theory prediction to compare to precision experimental data in order to extract precision PDFs.



Tools : ggHiggs(Marco Bonvini)

$\sigma(gg \to H)$ at 14 TeV LHC

7-point scale variation at N3LO in QCD for $m_t = 172.5$ GeV and $M = m_H = 125$ GeV

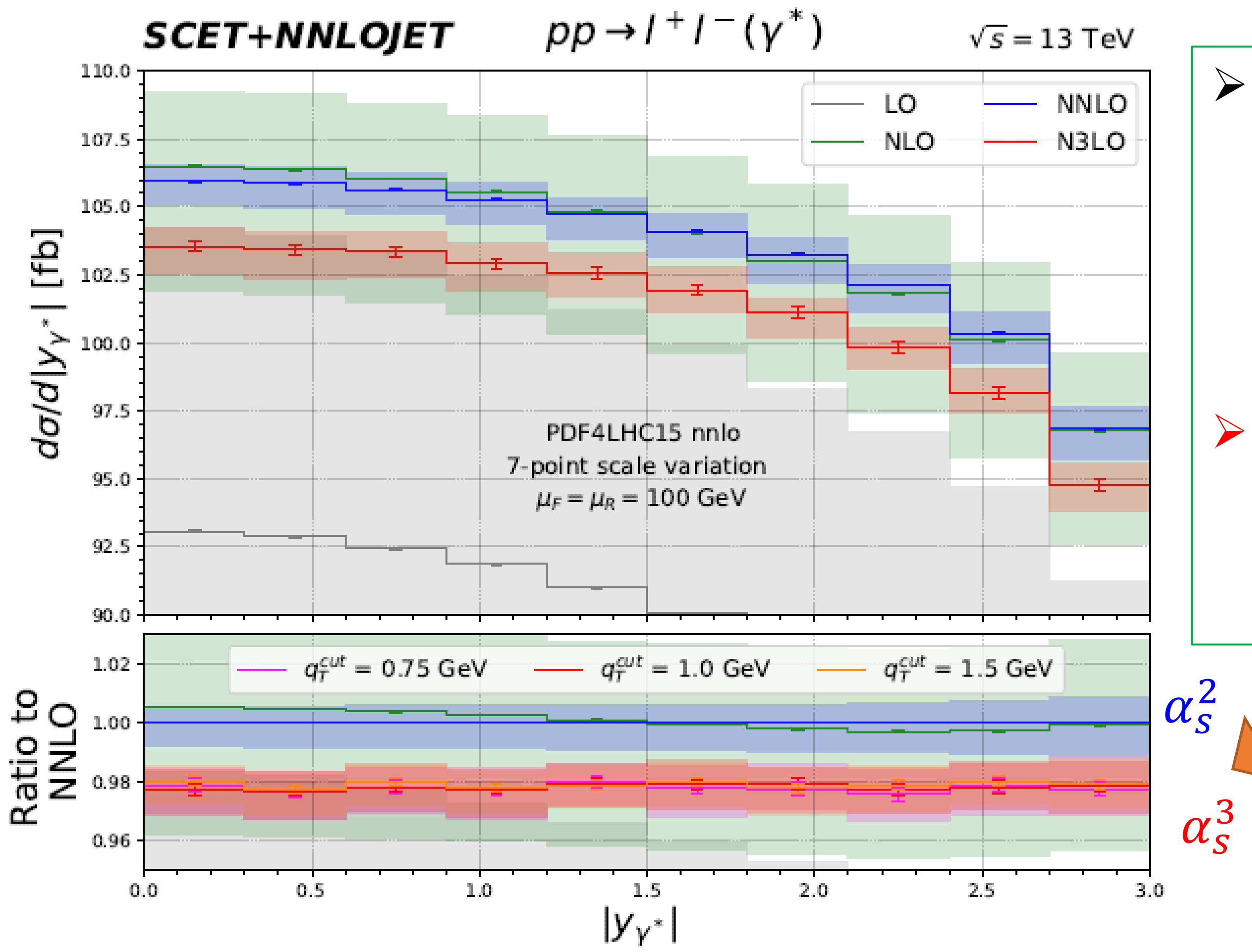
μ_R/M \ μ_F/M	0.5	1	2
0.5	3.4%	3.6%	-
1	-0.6%	-	0.6%
2	-	-5.6%	-4.7%

- The K-factor of electroweak (EW) correction is about 1.05
- The PDF uncertainty is about 2.8%

Estimating missing higher order contribution via varying μ_f and μ_R scales

CTEQ

arXiv:2107.09085



- Varying the factorization μ_f and renormalization μ_R scales by a factor of 2 around their nominal values (with 7-point scale variation) does not always lead to a good estimate of missing higher order (MHO) effect in the perturbative calculation.
- The N3LO correction is outside the scale variation band predicted at NNLO, due to accidental cancellation among various partonic subprocess contributions.

This comparison does not include PDF and α_s induced errors.

Some data requires all-order (resummation) calculations

- When applying a symmetric p_T cut (with same magnitude) on the decay leptons of inclusive W or Z boson production, the two leptons are almost back-to-back, decaying from a low p_T gauge boson.
- Fixed order predictions cannot correctly predict the low p_T distribution of W or Z.
- It requires a resummation calculation, such as ResBos, to resum all the large logs arising from multiple soft-gluon radiation.

What's QCD Resummation?

- Perturbative expansion

$$\frac{d\hat{\sigma}}{dq_T^2} \sim \alpha_s \left\{ 1 + \alpha_s + \alpha_s^2 + \cdots \right\}$$

- The singular pieces, as $\frac{1}{q_T^2}$ (1 or log's)

$$\frac{d\hat{\sigma}}{dq_T^2} \sim \frac{1}{q_T^2} \sum_{n=1}^{\infty} \sum_{m=0}^{2n-1} \alpha_s^{(n)} \ln^{(m)}\left(\frac{Q^2}{q_T^2}\right)$$

$$\sim \frac{1}{q_T^2} \{ \quad \alpha_s \left(\underline{L+1}\right) \qquad L \equiv \ln\left(\frac{Q^2}{q_T^2}\right)$$

$$+ \alpha_s^2 \left(\underline{L^3 + L^2} + \underline{L+1}\right)$$

$$+ \alpha_s^3 \left(\underline{L^5 + L^4} + \underline{L^3 + L^2} + \underline{L+1}\right)$$

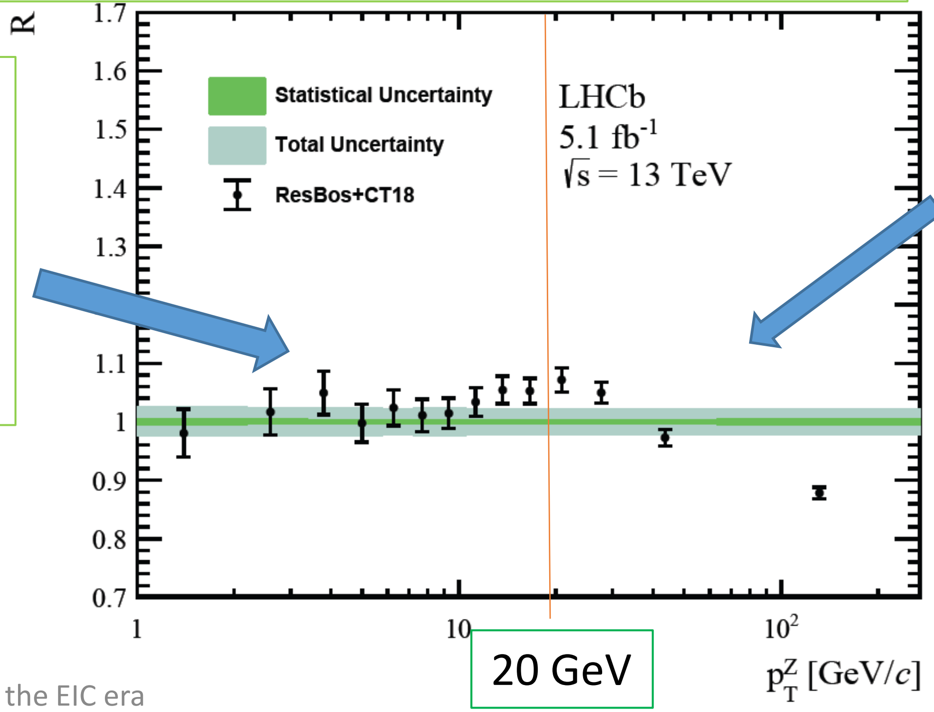
$$+ \cdots \quad \}$$

Resummation is to reorganize the results in terms of the large Log's.

$$\alpha_s \ln\left(\frac{Q^2}{q_T^2}\right) \sim 1$$

ResBos + CT18 can describe well low $p_T(Z)$ region, with $p_T(Z) < 20$ GeV

Compare to LHCb 13 TeV Z data; arXiv:2112.07458



20 GeV

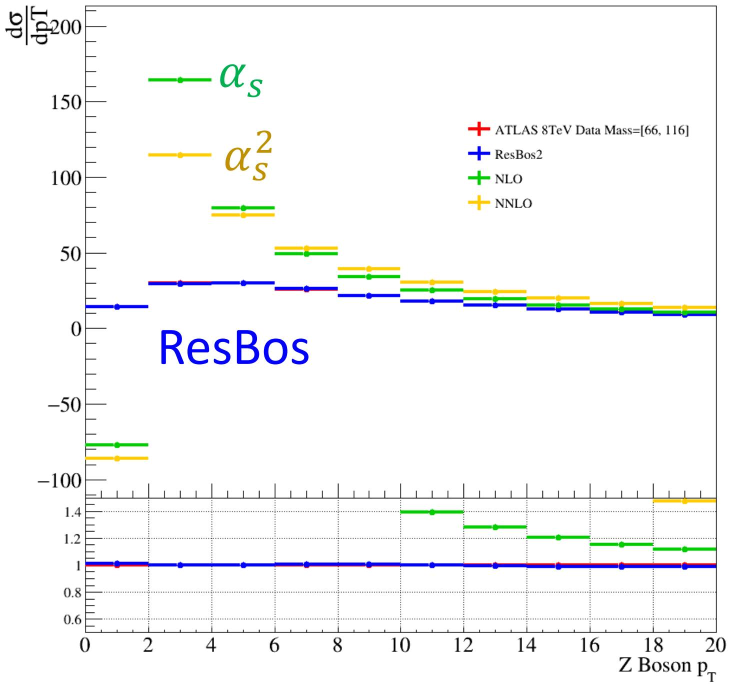
High $p_T(Z)$ region needs α_s^3 contrinution

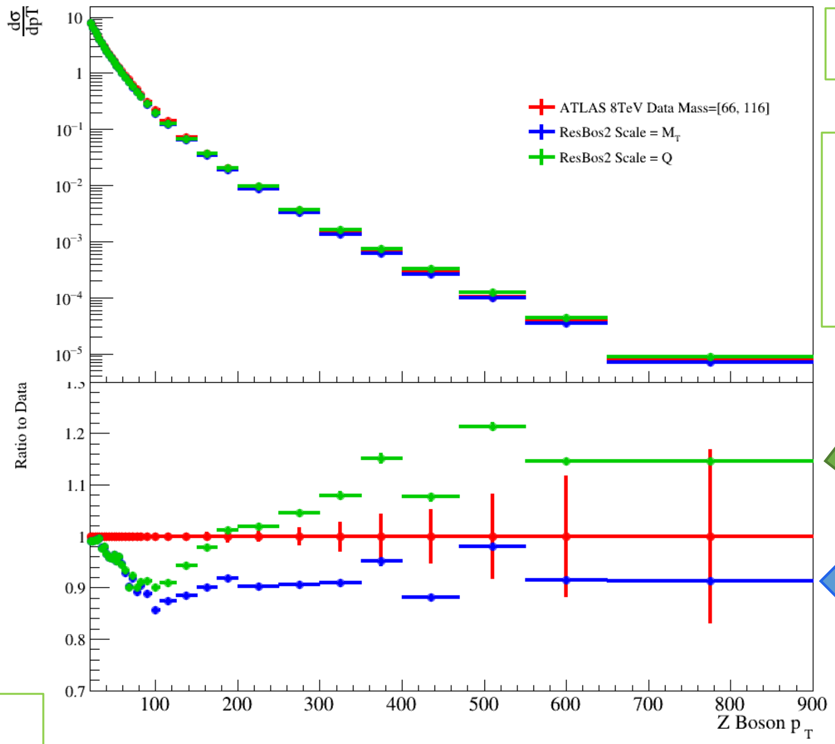
Some data requires all-order (resummation) calculations: ResBos

Compare to ATLAS 8 TeV Z data; arXiv:1606.00689

https://gitlab.com/resbos2

arXiv:2205.02788





- Sensitive to scale choices at α_s^2
- High $p_T(Z)$ region requires yet higher order (α_s^3) contribution.

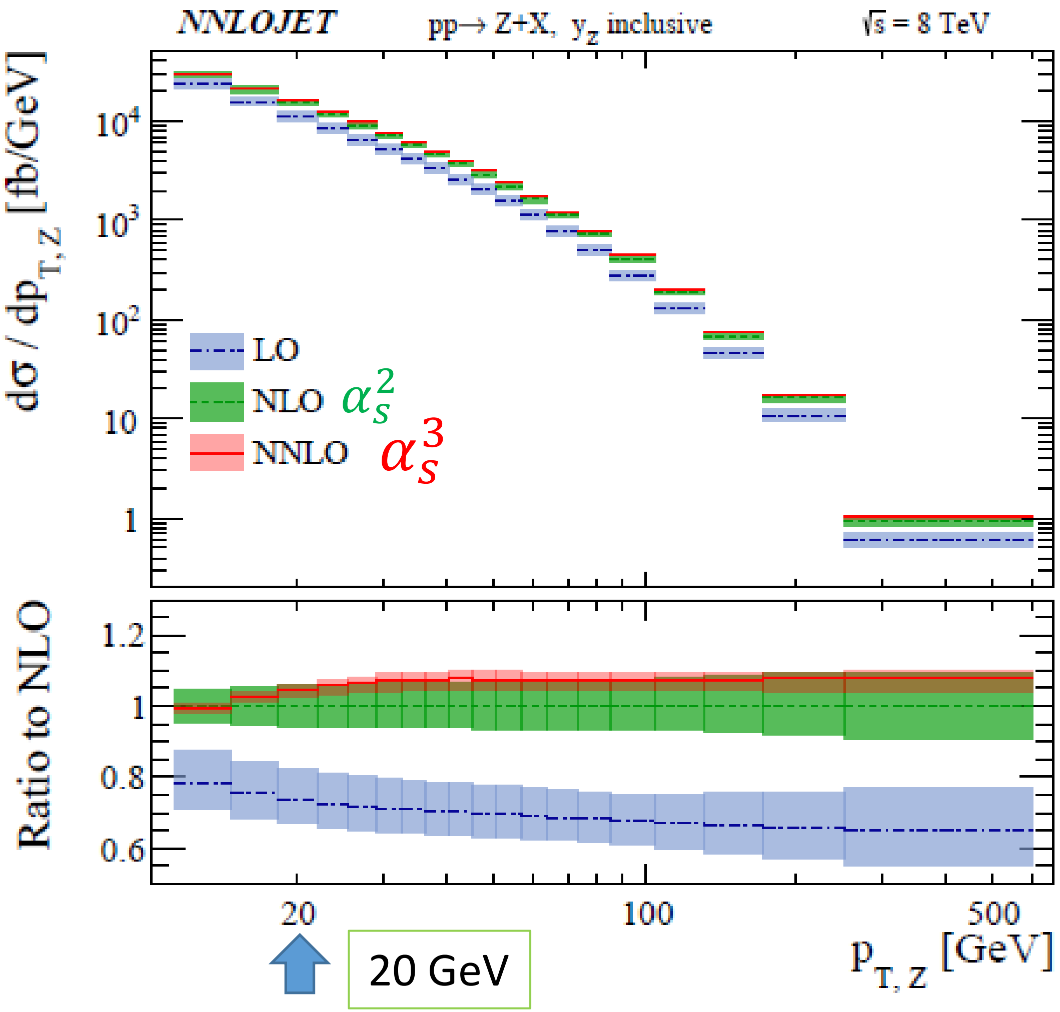
Use $\mu_F = \mu_R = \mathrm{Q}$
Invariant mass, at α_s^2

Use $\mu_F = \mu_R = \mathrm{m_T}$
where $m_T = \sqrt{Q^2 + p_T^2}$
Transverse mass, at α_s^2

The low p_T Z data, with $p_T(Z) < 20$ GeV, can be described well by ResBos, but not fixed order (NLO, NNLO,…) calculations which yield singular result as $p_T(Z) \to 0$.

Require higher (fixed) order calculations for $p_T(Z) > 20$ GeV; α_s^3 correction increases the rate by about 10% when using the scale m_T and renders a good agreement with data.

Higher order contributions are important



20 GeV

arXiv:1708.00008

- The α_s^3 prediction has much smaller scale variation as compared to α_s^2 calculation.
- For $p_{T(Z)} > 20$ GeV, the K-factor of α_s^3/a_s^2 is roughly a constant, about 1.1

Extensions of CT18 family PDFs: post-CT18

- CT18As: CT18A (a CT18 fit with the inclusion of ATLAS 7 TeV W, Z data), but with non-zero strangeness asymmetry $s_{-}(x, Q_0) = s(x, Q_0) - \bar{s}(x, Q_0)$ at Q_0 = 1.3 GeV.
- CT18As_Lat: CT18As, but including Lattice QCD data on strangeness asymmetry $s_{-}(x, Q_0) = s(x, Q_0) - \bar{s}(x, Q_0)$
- CT18FC: fitted charm PDF $c(x, Q_0) \neq 0$; for c(x, Q_0) = or $\neq \bar{c}(x, Q_0)$
- CT18qed: take photon as a parton of proton; $\gamma(x, Q_0) \neq 0$
- Machine Learning approach: A fast version of Lagrange Multiplier scan (for simultaneous fit to PDFs and SMEFT)
- CT18LO: LO PDF for event generators, e.g., PYTHIA
- NNLO-QCD+ NLO-QED PDFs for a neutron
- CT18MC: NLO PDFs for Monte Carlo event generators

Non-zero strangeness asymmetry at Q_0

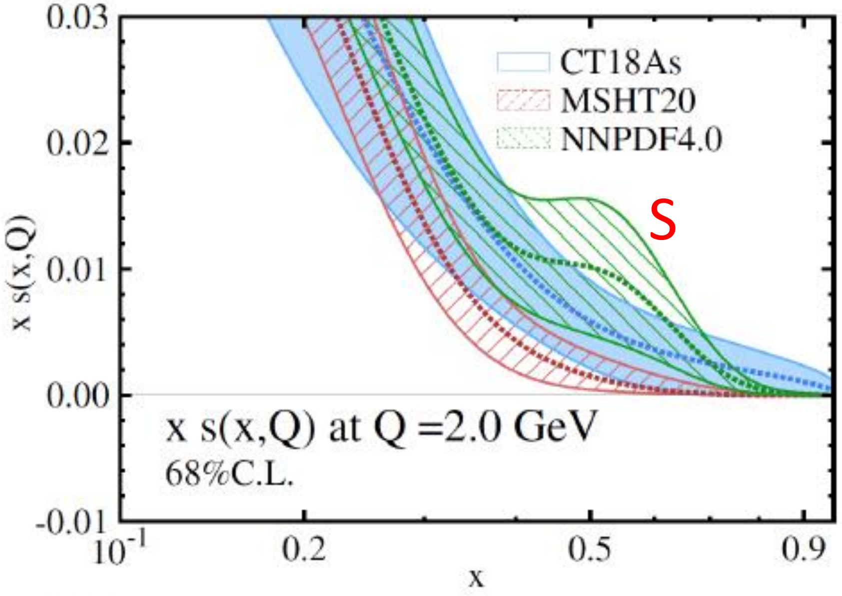
- CT18As: CT18A (a CT18 fit with the inclusion of ATLAS 7 TeV W, Z data), but with non-zero strangeness asymmetry $s_{-}(x, Q_0) = s(x, Q_0) - \bar{s}(x, Q_0)$ at Q_0 = 1.3 GeV.

- CT18As_Lat: CT18As, but including Lattice QCD data on strangeness asymmetry $s_{-}(x, Q_0) = s(x, Q_0) - \bar{s}(x, Q_0)$

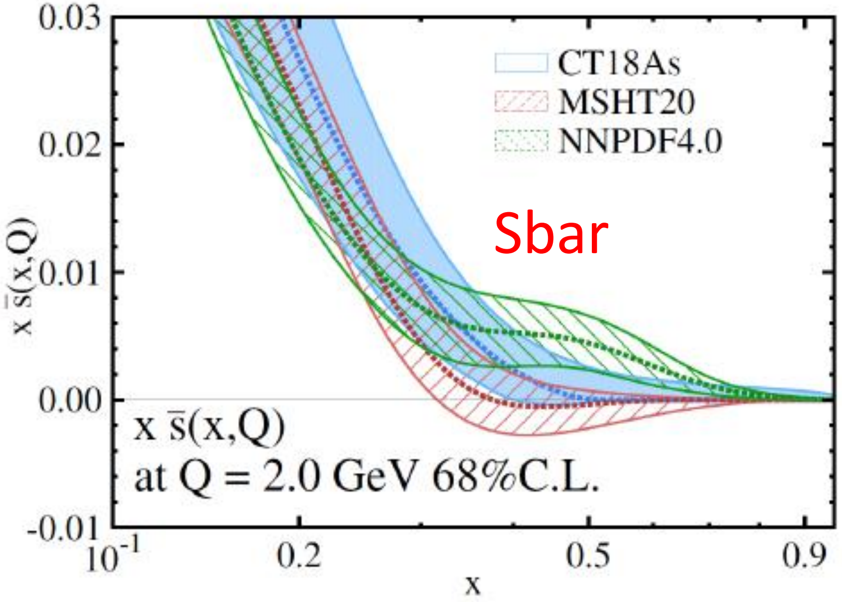
See talk by Huey-Wen Lin for lattice-QCD calculations.

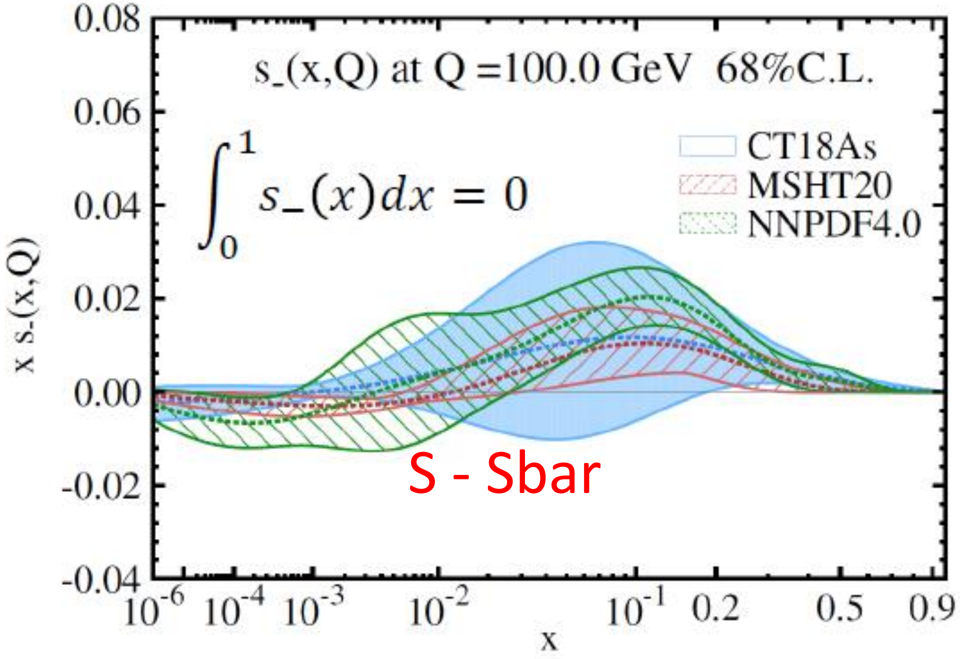
CT18As NNLO: Strangeness asymmetry

[arXiv:2204.07944]



New NNLO fits:
CT18As: CT18A with $s_- \equiv s - \bar{s} \neq 0$





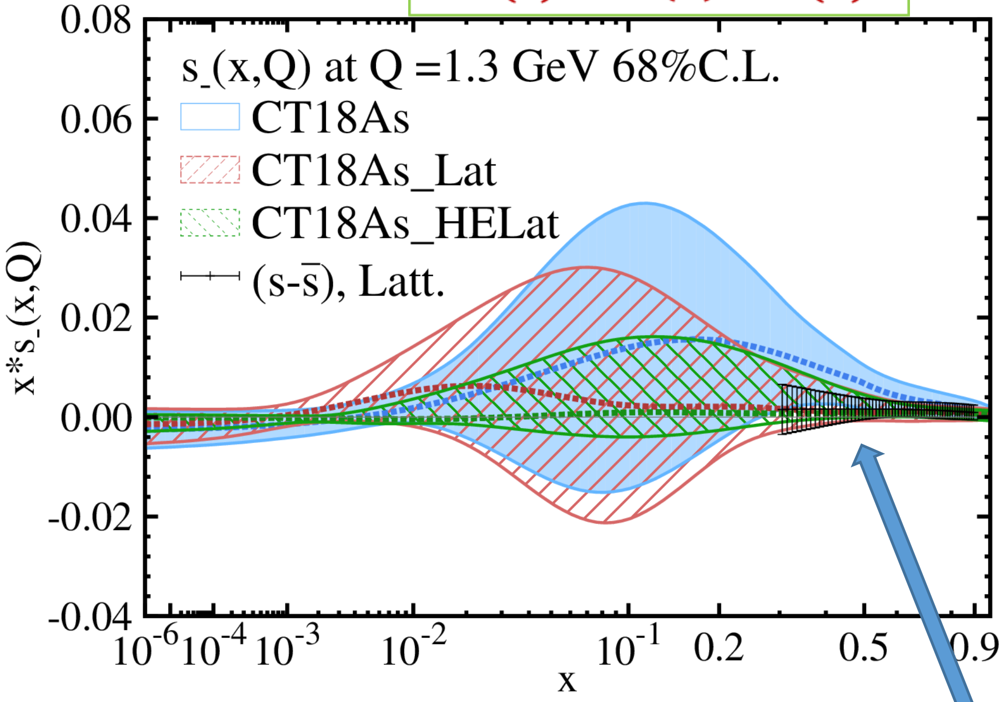
MSHT20 allows a negative sbar-PDF at low Q value.

Lattice QCD data as an input to PDF global analysis

C T E Q

arXiv: 2211.11064

$s_-(x) = s(x) - \bar{s}(x)$



- The uncertainties of PDFs can be further reduced by including Lattice QCD predictions in global analysis
- Complementarity of collider experimental data and lattice QCD data

CT18As: CT18A with non-zero strangeness asymmetry $s_-(x)$ at Q_0 = 1.3 GeV.
CT18As_Lat: CT18As PDFs with lattice input on $s_-(x)$
CT18As_HELat: CT18As_Lat with the lattice errors reduced by half.

CT18A = CT18 + ATLAS W,Z data

- Lattice QCD calculation provides prediction at 0.3 < x < 0.8, while NuTeV and CCFR SIDIS di-muon data constraint strangeness PDFs at 0.015 < x < 0.336.
- Lattice QCD data are consistent with $s(x) = \bar{s}(x)$ at large x.
- CT18 assumes $s(x, Q_0) = \bar{s}(x, Q_0)$; NNLO DGLAP evolution generates $s(x, Q) \neq \bar{s}(x, Q)$ at $Q > Q_0$

arXiv: 2005.12015

Fitted Charm
vs
Intrinsic Charm

- CT18: perturbative charm PDF $c(x, Q_0) = 0$
- CT18FC: fitted charm PDF $c(x, Q_0) \neq 0$; for $c(x, Q_0) =$ or $\neq \bar{c}(x, Q_0)$

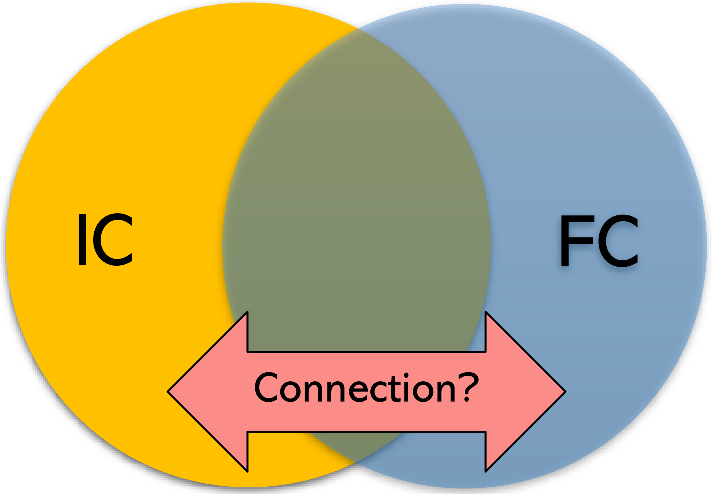
Intrinsic charm vs Fitting charm

IC is either process-dependent or scheme-dependent.

A persistent terminological and conceptual ambiguity:
Is IC a type of a QCD observable or a nonperturbative QCD function?

If an observable, it receives process-dependent radiative contributions.

⇒ **Process dependence**



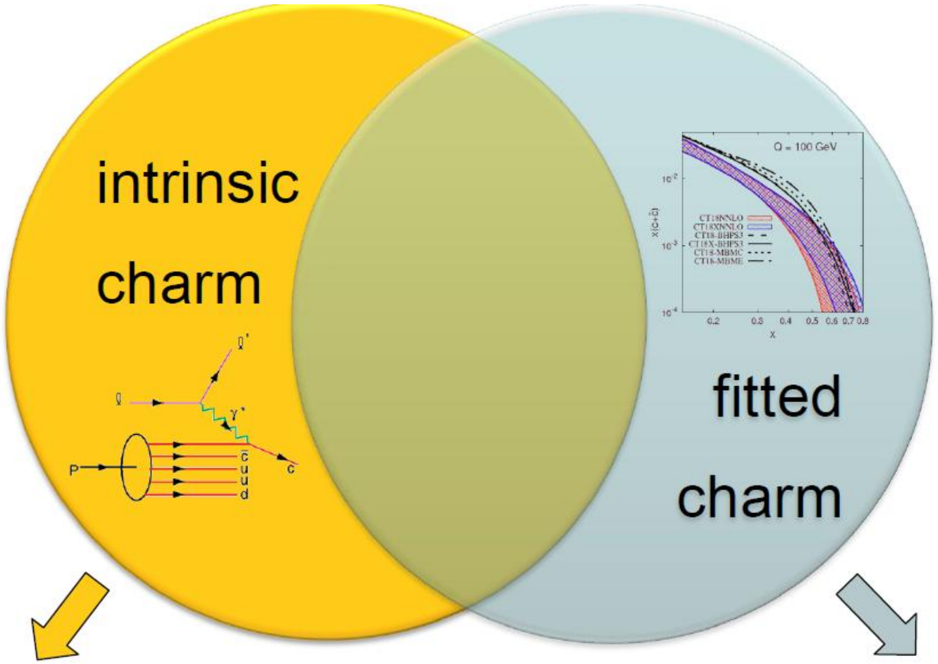
If a nonperturbative function, it can be defined in many ways.

⇒ **Scheme dependence**

Challenging to formulate a rigorous definition of Intrinsic Charm and its relation to Fitted Charm

CTEQ

arXiv:2211.01387



- The concept of nonperturbative methods
- Can refer to a component of the hadronic Fock state or the type of the hard process
- Predicts a typical enhancement of the charm PDF at $x \gtrsim 0.2$

Connection?

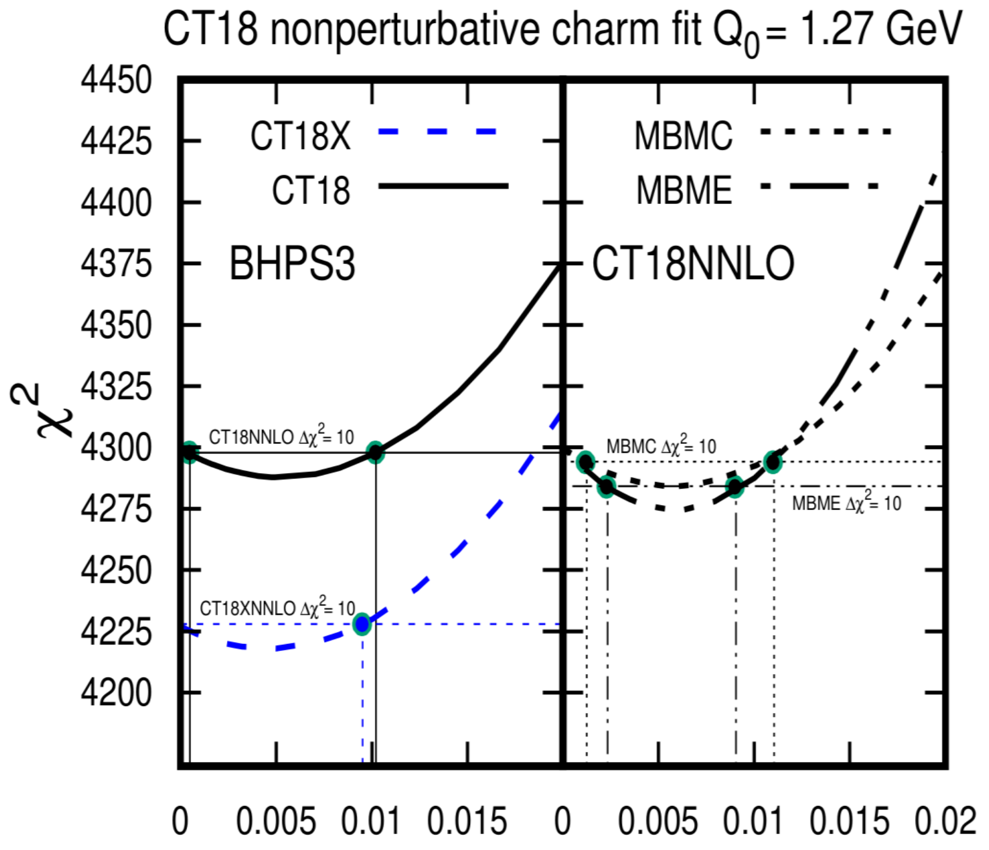
- A charm PDF parametrization at scale $Q_0 \approx 1$ GeV found by global fits [CT, NNPDF, …]
- Arises in perturbative QCD expansions over α_s and operator products
- May absorb process-dependent or unrelated radiative contributions

Nonperturbative (intrinsic) charm of proton CT18FC

arXiv:2211.01387

- Proton's intrinsic charm, a non-vanishing charm PDF at Q_0 (around 1 GeV) scale, remains indeterminate.
- Challenging to formulate a rigorous definition of intrinsic charm (IC) and its relation to fitted charm (FC).
- Need more NNLO and better showering calculations.
  - Z+c theory predictions have sizable uncertainties, e.g., flavor-tag jet definition, multi-parton interaction (MPI), showering effect.
- Need more sensitive data

arXiv: 2302.12844



- CT18FC study found no significant evidence for non-zero IC, as NNPDF4.0 IC, Nature 608 (2022) 7923, 483.
- FC in CT18FC study is currently consistent with zero, and with shallower $\Delta\chi^2$ than CT14IC.

$$\langle x\rangle_{\rm FC} \approx 0.5\% \ (\Delta\chi^2 \gtrsim -25) \text{ vs. } \langle x\rangle_{\rm FC} \approx 0.8-1\% \ (\Delta\chi^2 \gtrsim -40)$$ in CT14 IC

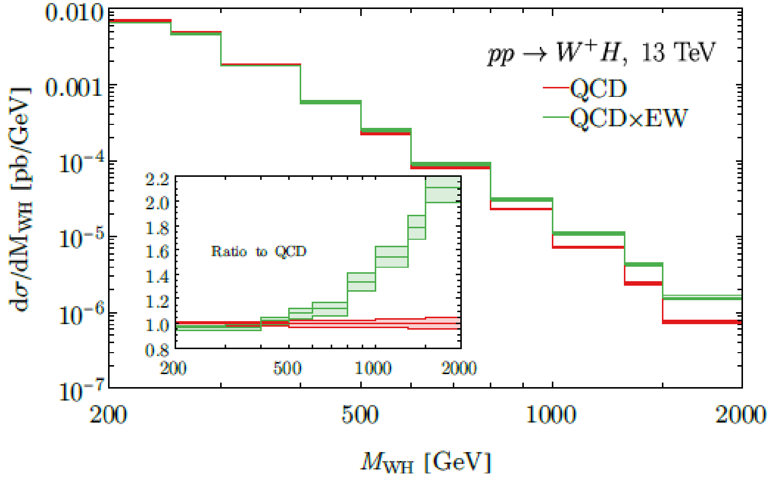
QED corrections added to NNLO QCD predictions

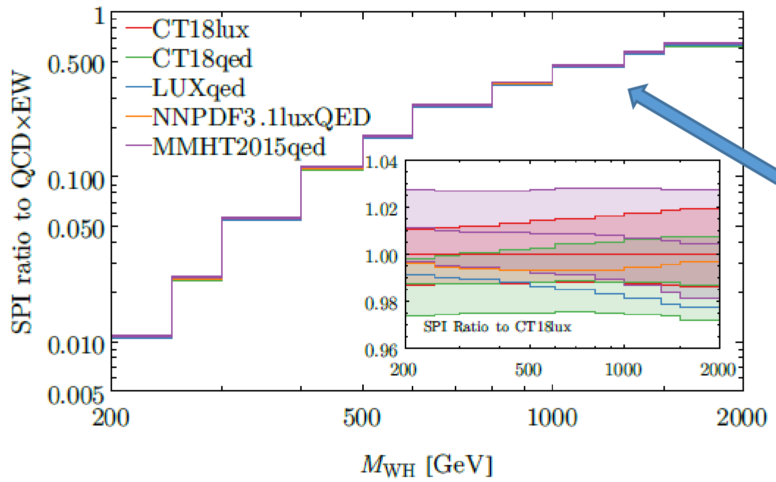
- CT18qed: take photon as a parton of proton; $\gamma(x, Q_0) \neq 0$

Photon PDF of proton: CT18qed

arXiv:2106.10299

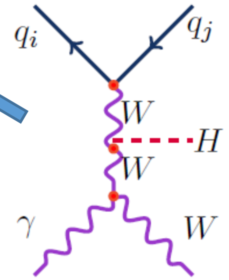
$\sigma(pp \to W^+H)$

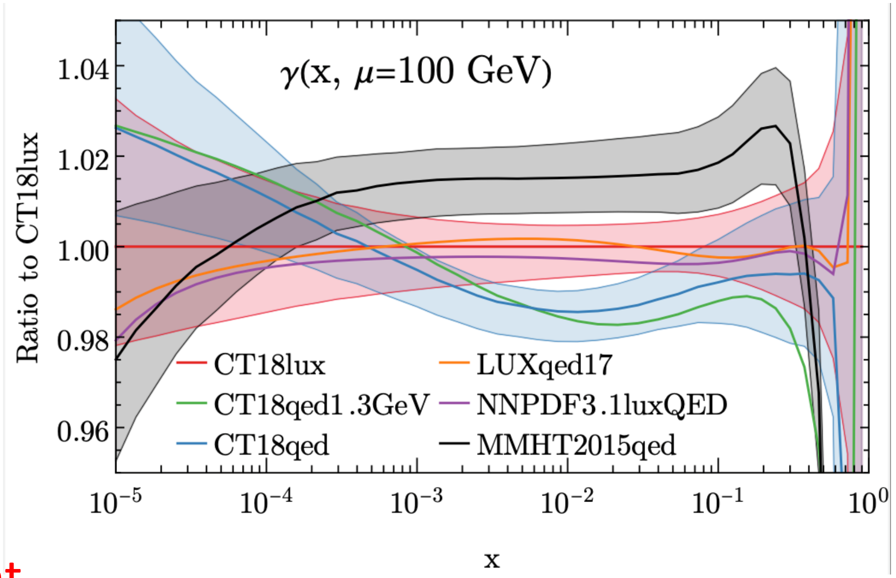




At α_s^2 accuracy, EW corrections and explicit photon PDF $\gamma(x, Q^2)$ are needed.

Single-photon-initiated (SPI) process; important at TeV scale



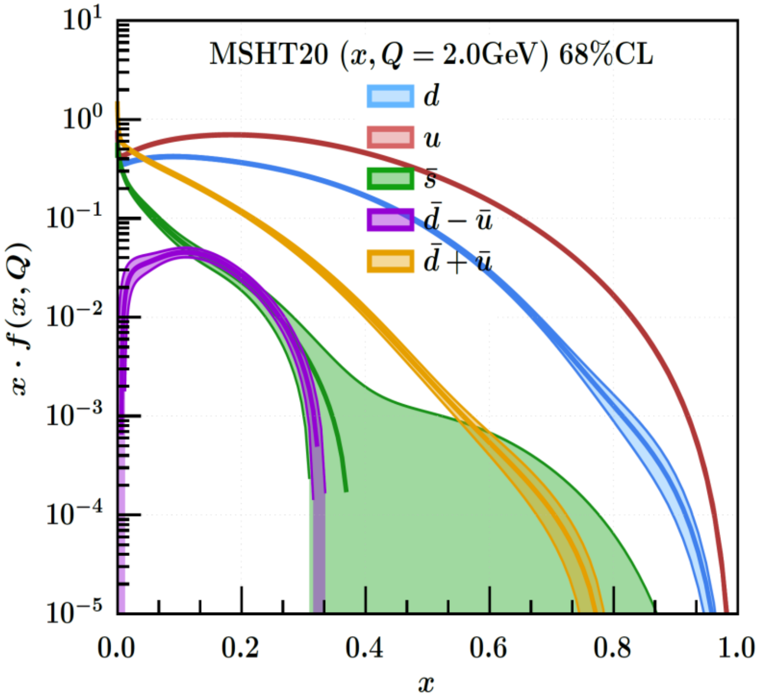


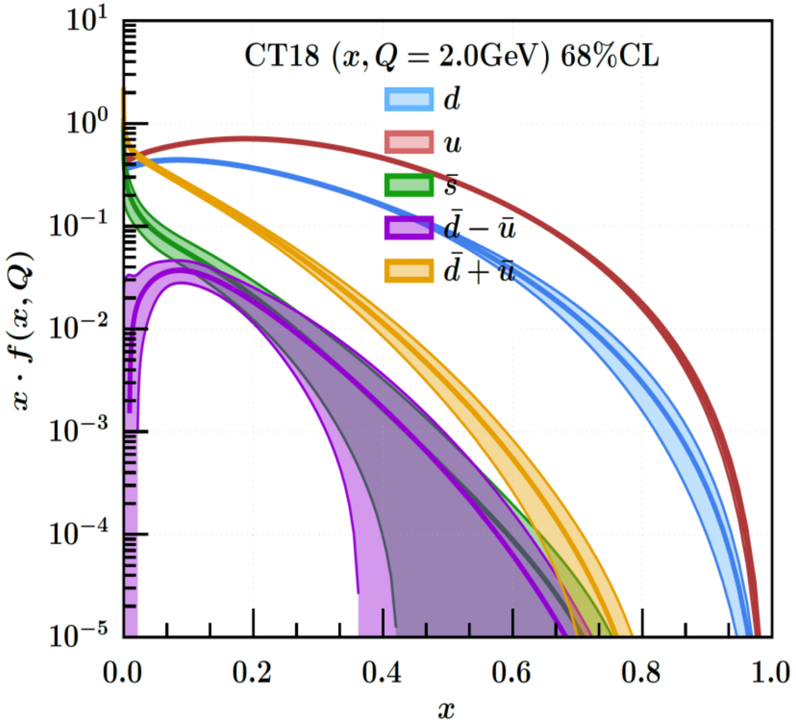
- CT18lux provides the photon PDF at all scales, μ.
- CT18qed initializes photon PDF at μ_0, and evolves to high scales.
- CT18lux gives the photon in between LUXqed(17) and MMHT2015qed, while CT18qed gives smaller photon.

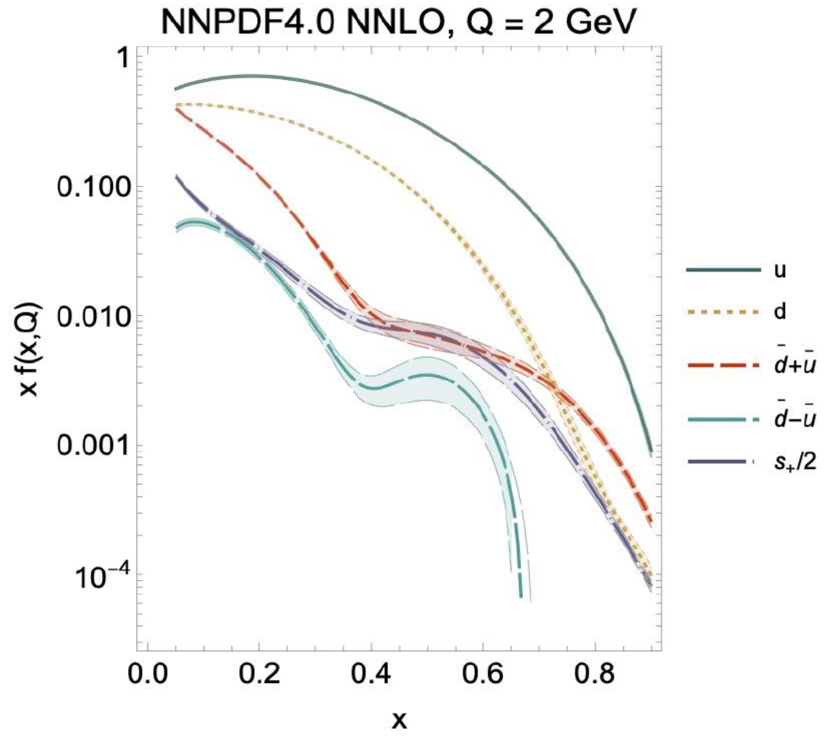
New experimental observables to further constrain PDFs

- Drell-Yan A_{FB} data for refining PDFs at large x

Compare MSHT20, CT18 and NNPDF4.0 PDFs







- MSHT20, CT18 and NNPDF4.0 predict very different sea quarks at large x.
- A_{FB} is sensitive to combinations of $\bar{u}/u$ and $\bar{d}/d$

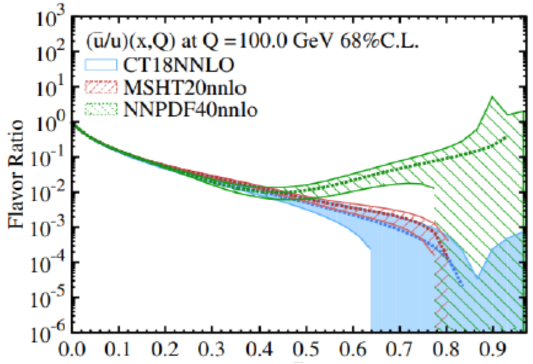
High-sea scenario with non-smooth light-sea quarks, with sea PDFs that can be larger than valence PDFs at large x.

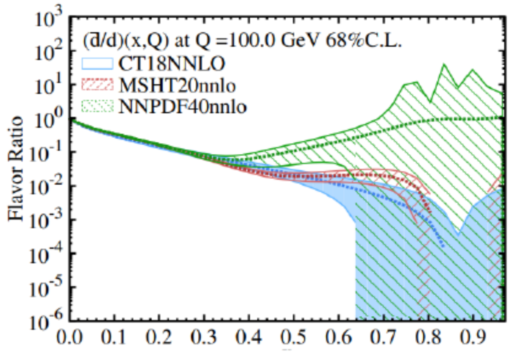
Smaller uncertainties.

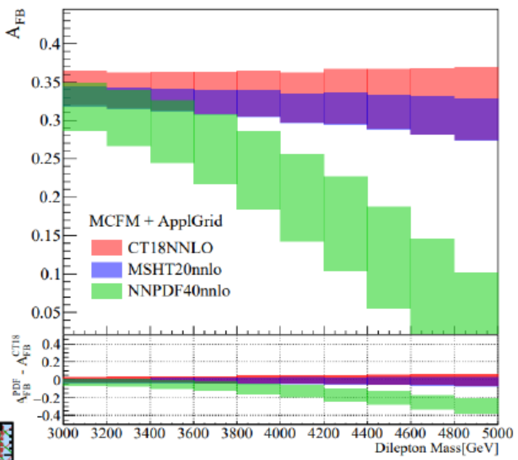
Impact of A_{FB} in the high-mass Drell-Yan process

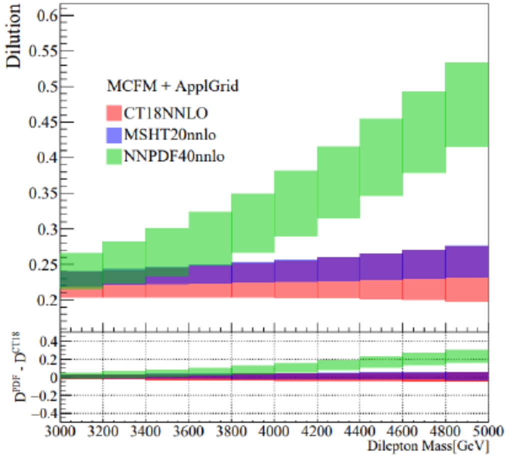
Y. Fu et al., 2307.07839
C. Willis et al., 1809.09481

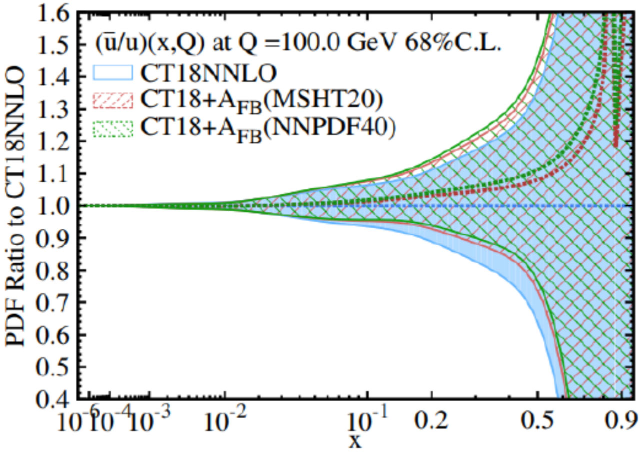
- A_{FB} at the LHC is sensitive to the energy dilution factor D (probability of $k_q^0 < k_{\bar{q}}^0$ in the Collins-Soper frame)
- $A_{FB}^h = \frac{N_F^h - N_B^h}{N_F^h + N_B^h} \approx (1-2D)A_{FB}^q$
- A_{FB} at high invariant mass region probes $\bar{u}/u$, $\bar{d}/d$ at $x > 0.2$

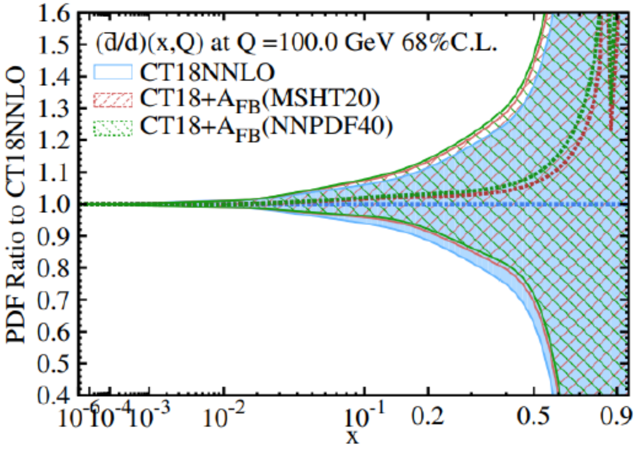












- CT18, MSHT20, and NNPDF4.0 predict very different $\bar{q}/q$ at $x > 0.2$
- The article quantified the potential effect of high-mass A_{FB} on large-x antiquarks

See also NNPDF (2209.08115), Fiaschi et al. (2211.06188)

Drell-Yan A_{FB} data for refining PDFs

Siqi Yang, et al, arXiv:2202.13628

For LHC's pp collision

$$C_u(x_1,x_2) = \frac{[u(x_1)\bar{u}(x_2) - \bar{u}(x_1)u(x_2)]\,\mathcal{N}_u}{\sum_{q=u,d,s,c,b}[q(x_1)\bar{q}(x_2) + \bar{q}(x_1)q(x_2)]\,\mathcal{N}_q}$$

$$C_d(x_1,x_2) = \frac{[d(x_1)\bar{d}(x_2) - \bar{d}(x_1)d(x_2)]\,\mathcal{N}_d}{\sum_{q=u,d,s,c,b}[q(x_1)\bar{q}(x_2) + \bar{q}(x_1)q(x_2)]\,\mathcal{N}_q}$$

For Tevatrons' $p\bar{p}$ collision

$$C_u(x_1,x_2) = \frac{[u(x_1)u(x_2) - \bar{u}(x_1)\bar{u}(x_2)]\,\mathcal{N}_u}{\sum_{q=u,d,s,c,b}[q(x_1)q(x_2) + \bar{q}(x_1)\bar{q}(x_2)]\,\mathcal{N}_q}$$

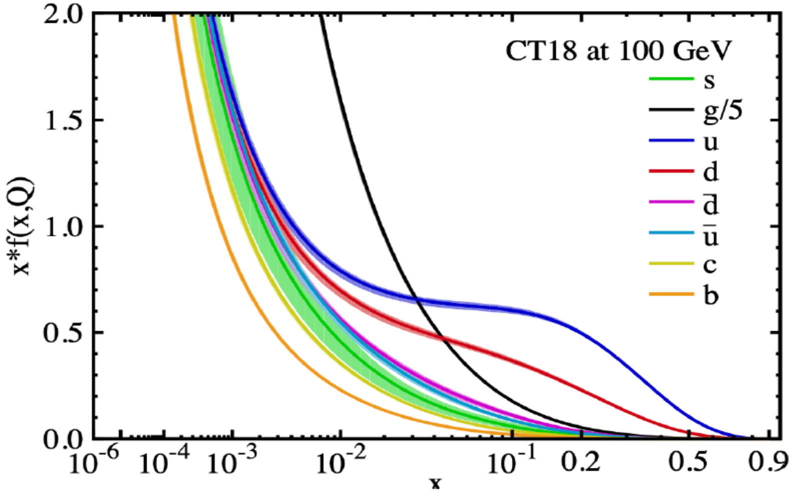
$$C_d(x_1,x_2) = \frac{[d(x_1)d(x_2) - \bar{d}(x_1)\bar{d}(x_2)]\,\mathcal{N}_d}{\sum_{q=u,d,s,c,b}[q(x_1)q(x_2) + \bar{q}(x_1)\bar{q}(x_2)]\,\mathcal{N}_q}$$

$$A^h_{FB} = C_u(Y,M,Q_T)\times A^u_{FB}(Y,M,Q_T) + C_d(Y,M,Q_T)\times A^d_{FB}(Y,M,Q_T)$$

Given that at hadron colliders x_2 is around 0.001 or even smaller, we have $u(x_2)\approx\overline{u}(x_2)\approx d(x_2)\approx\overline{d}(x_2)$ so they cancel out in the ratio parameter:

$$R = C_d/C_u = \frac{d-\bar{d}}{u-\bar{u}} = \frac{d_V}{u_V}$$

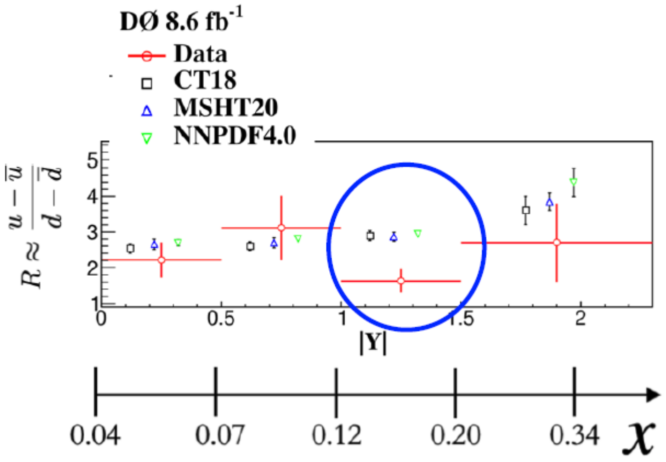
A perfect direct observable on valence quark ratio, without mix from anyother quarks !!



D0 and CMS measurements of A_{FB}

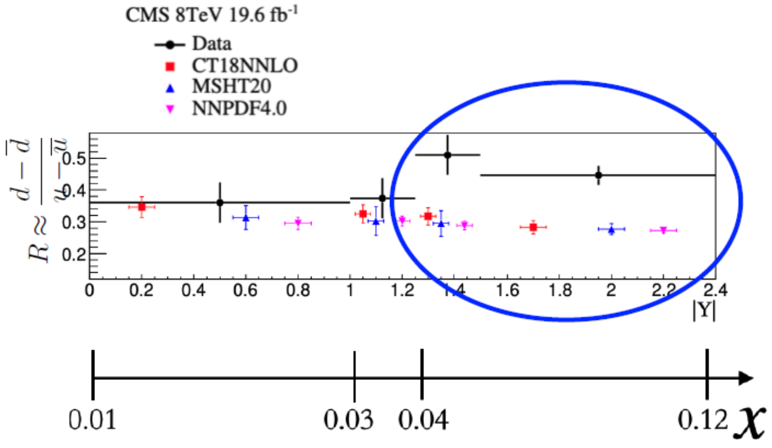
Expected to refine the existing PDF sets, as of 2025

- First direct measurements on valence quark ratio, DØ and CMS data
- The d_V/u_V is signifiantly higher than PDF predictions at $x \sim 0.1$
- The x-dependence is also different



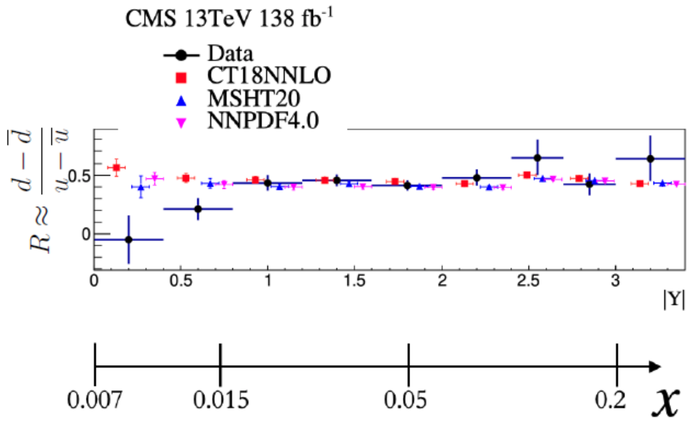
Measurement of $R = (u - \bar{u})/(d - \bar{d})$ at DØ, $p\bar{p}$ collision at 1.96 TeV

Phys. Rev. D 110, L091101 (2024)



Measurement of $R = (d - \bar{d})/(u - \bar{u})$ at CMS, pp collision at 8 TeV

arXiv: 2505.17608



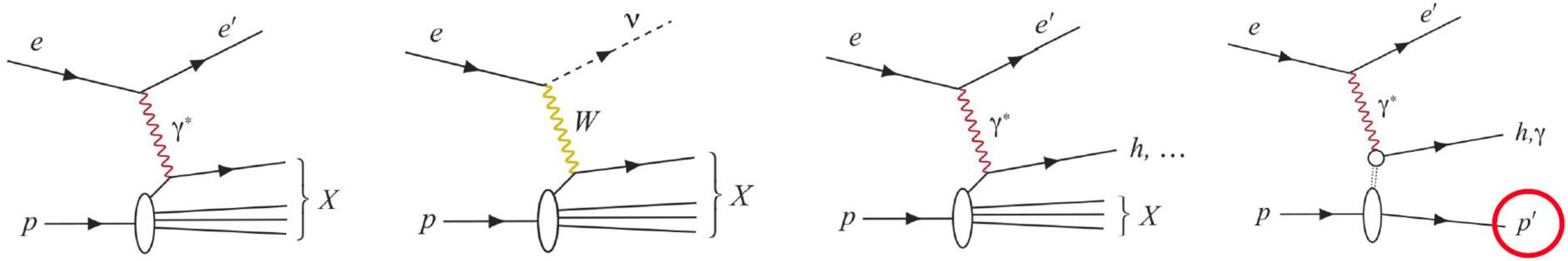
Measurement of $R = (d - \bar{d})/(u - \bar{u})$ at CMS, pp collision at 13 TeV

arXiv: 2505.17608

Challenges and Strategies in Determining Longitudinal Unpolarized Proton PDFs: From the LHC to EIC Prospects

- Scattering processes at the EIC
- Unpolarized PDFs at the EIC
- New Physics search opportunity at the EIC

Scattering processes at the EIC



Neutral Current DIS | **Charged Current DIS** | **Semi-Inclusive DIS** | **Exclusive Processes**

- Electron beam can be longitudinally polarized.
- Proton (Ion) beam can be longitudinally or transversely polarized.
- The measurements of exclusive processes are special at the EIC, as compared to the LHC.

ePIC performance: DIS kinematics with ePIC

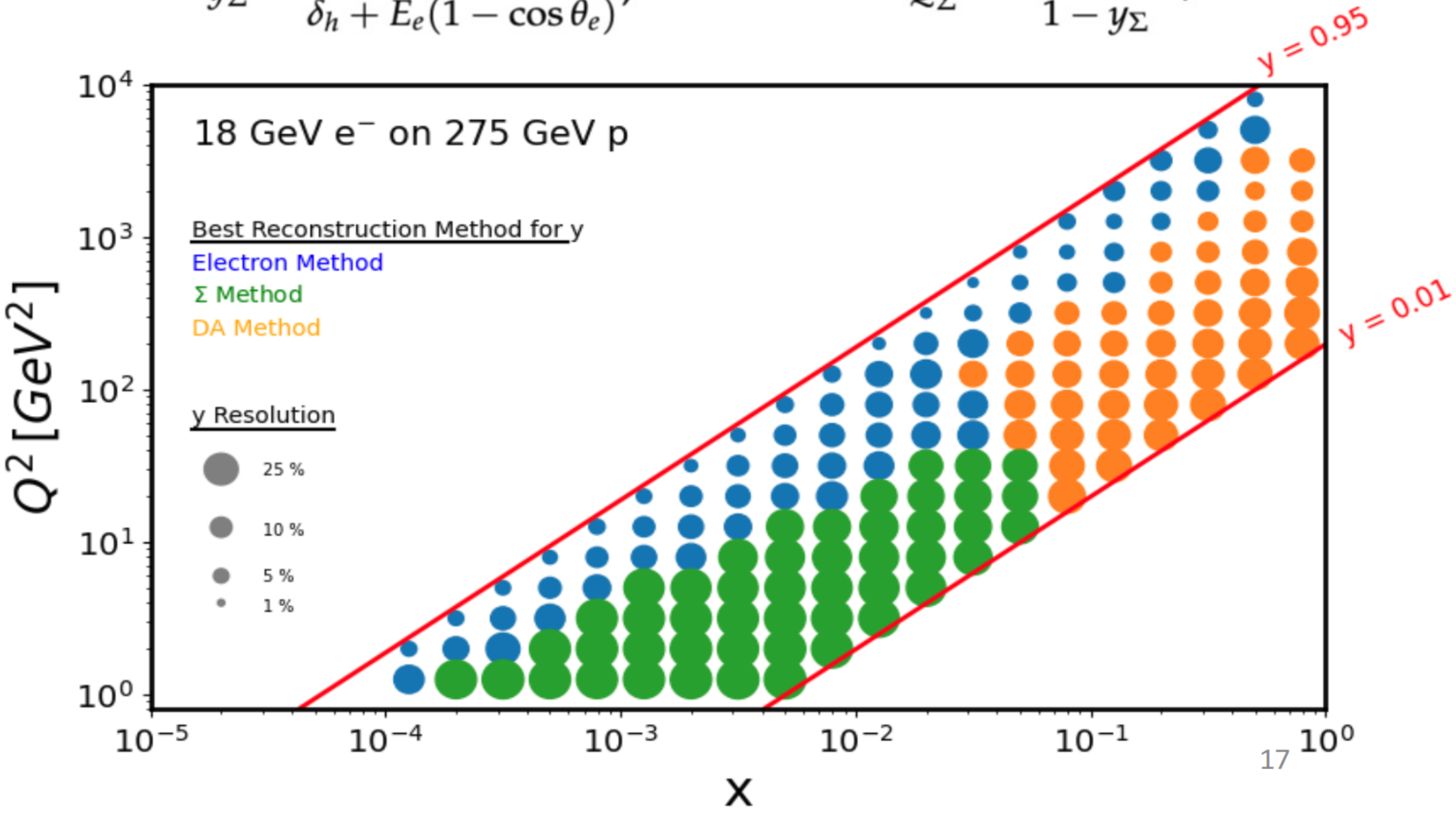
Kinematic Resolutions

$$y_e = 1 - \frac{E_e(1-\cos\theta_e)}{2E_0}, \qquad Q_e^2 = 2E_0E_e(1+\cos\theta_e).$$

$$y_{DA} = \frac{\alpha_h}{\alpha_e+\alpha_h}, \qquad Q_{DA}^2 = \frac{4E_0^2}{\alpha_e(\alpha_e+\alpha_h)}$$

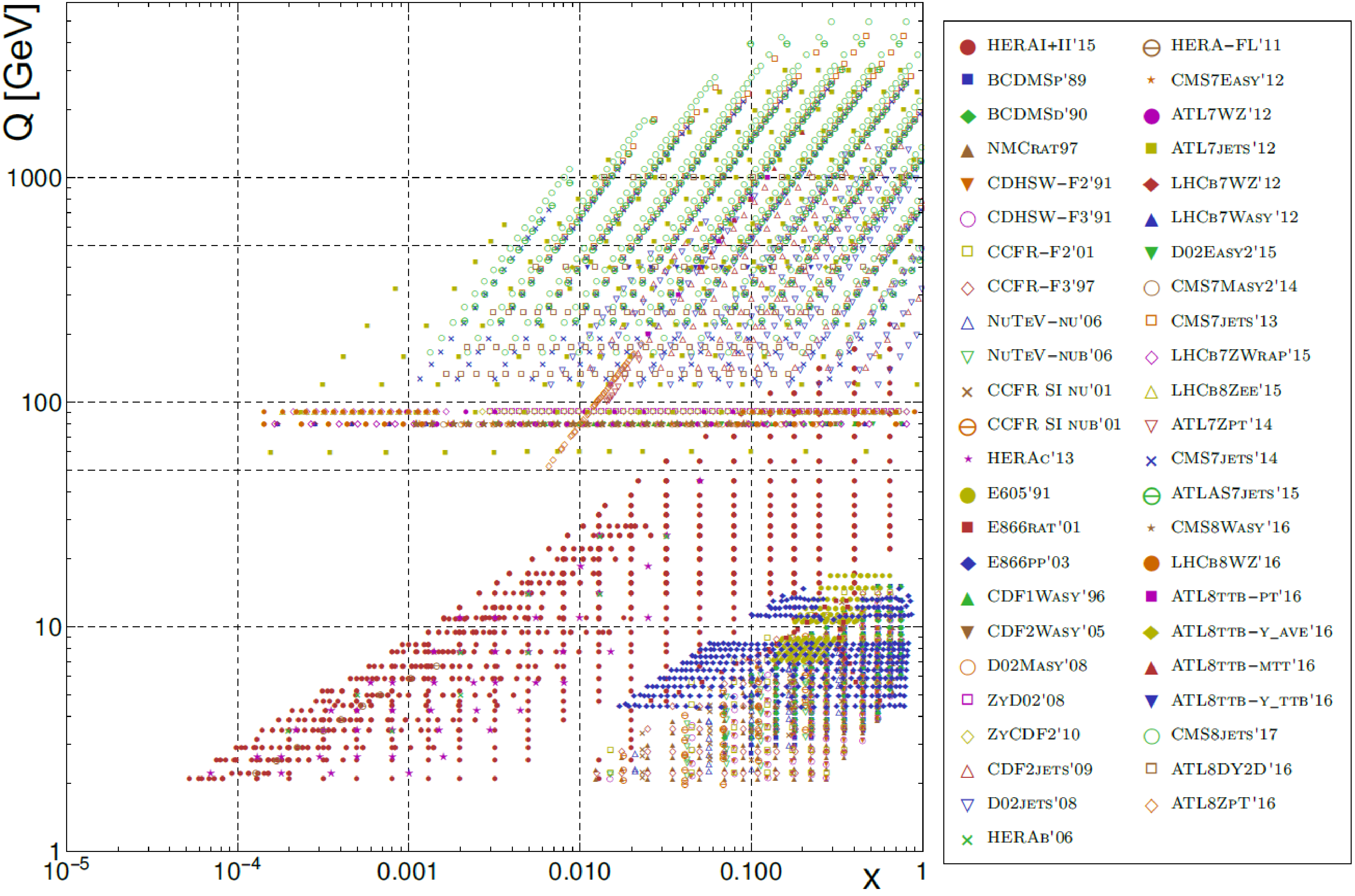
$$y_\Sigma = \frac{\delta_h}{\delta_h + E_e(1-\cos\theta_e)}, \qquad Q_\Sigma^2 = \frac{E_e'^2\sin^2\theta_e}{1-y_\Sigma}.$$

- Reconstruct inclusive kinematics using various methods
  → **compare reconstruction performance**
  - Color of point indicates best method for *y* (inelasticity)
  - Size of point indicates *y* resolution
- **~30% or better *y* resolution across $x - Q^2$ plane**



S Fazio, BNL-INT Joint Workshop, June 2025

Experimental data in CT18 PDF analysis

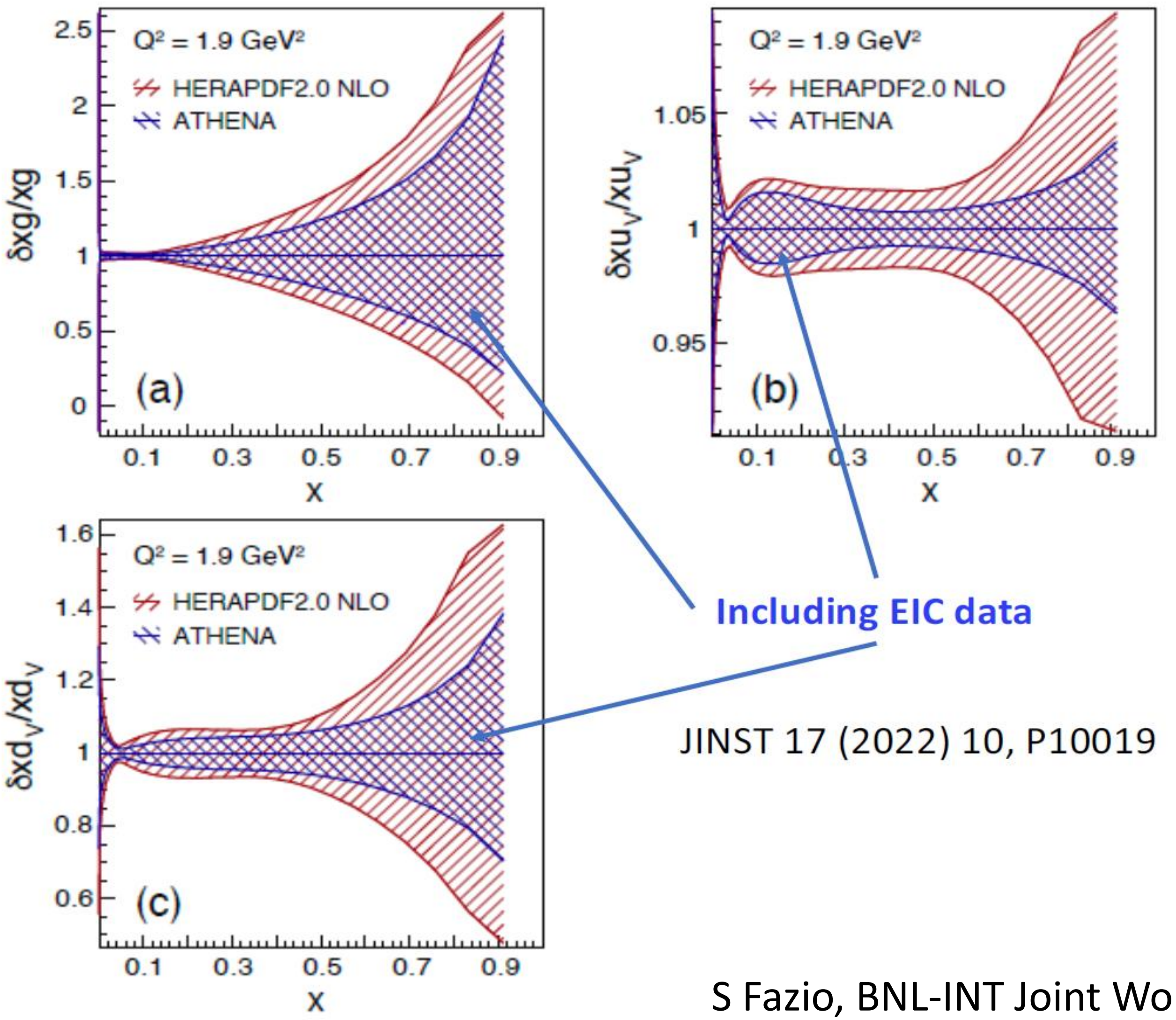


C T E Q

arXiv: 1912.10053

See talk by Tie-Jiun Hou for the Post-CT18 data sets

Scientific goals: proton PDFs



JINST 17 (2022) 10, P10019

Proton PDFs @ EIC

- ○ $F_2(x, Q^2)$ **largely studied at HERA**
- ○ Nevertheless, a better precision often needed for precise calculations!
  --> explore specific kinematics
- ○ EIC impact on HERA + LHC global fits, as estimated at the times of detector proposals in 2022 (ATHENA)
- ○ High-*x* region: constrain of both gluons and flavor-separated valence quarks

Key detector performance:

- Electron ID
- Fine *y* resolution over large phase space

S Fazio, BNL-INT Joint Workshop, June 2025

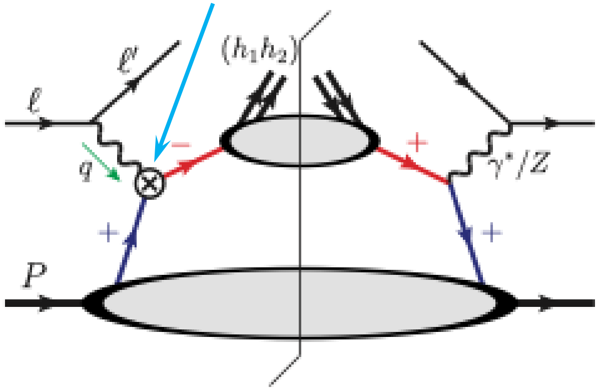
Transverse spin effects of quark @ EIC

Xin-Kai Wen, Bin Yan, Zhite Yu, C.-P. Yuan, 2408.07255

C T E Q

- **Quark dipole interactions can induce transversely polarized quarks**

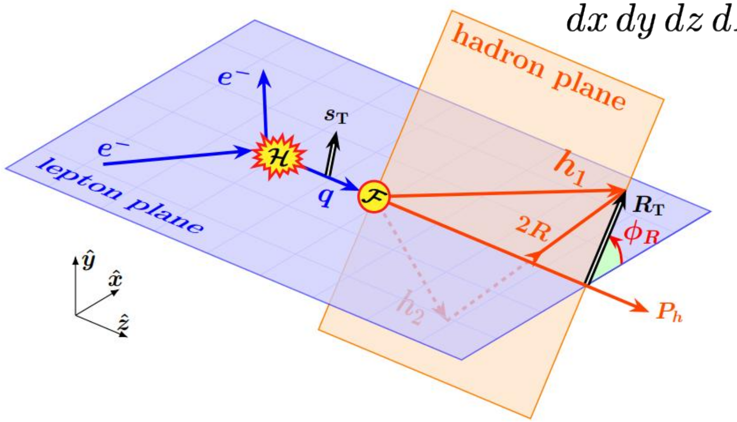
$\bar{q}_L \sigma_{\mu\nu} q_R A^{\mu\nu}, \bar{q}_L \sigma_{\mu\nu} q_R Z^{\mu\nu}$ **: Flips chirality from "Right" (+) to "Left" (−).**

$W^q_{-+} =$  $= \frac{C_q}{2}\left(s_q^x + i\, s_q^y\right) \propto \Gamma^q_{\gamma,Z}$

$$s_q^x = \frac{2}{C_q}\left(w_\gamma^q \,\mathrm{Re}\,\Gamma_\gamma^q + w_Z^q \,\mathrm{Re}\,\Gamma_Z^q\right)$$

$$s_q^y = \frac{2}{C_q}\left(w_\gamma^q \,\mathrm{Im}\,\Gamma_\gamma^q + w_Z^q \,\mathrm{Im}\,\Gamma_Z^q\right)$$

- **Measured via dihadron fragmentation function (DiFF)**



$$\frac{d\sigma}{dx\,dy\,dz\,dM_h\,d\phi_R} = \frac{N}{2\pi}\sum_q f_q(x)\otimes C_q(x,Q)\otimes D_q^{h_1h_2}(z,M_h)$$

$$\times\left[1 - \frac{H_q^{h_1h_2}(z,M_h)}{D_q^{h_1h_2}(z,M_h)}\left(s_q^x \sin\phi_R - s_q^y\cos\phi_R\right)\right]$$

Spin-analyzing power

Light-quark dipole interactions @ EIC

Xin-Kai Wen, Bin Yan, Zhite Yu, C.-P. Yuan, 2408.07255

$$s_q^x = \frac{2}{C_q}\left(w_\gamma^q \,\mathrm{Re}\,\Gamma_\gamma^q + w_Z^q \,\mathrm{Re}\,\Gamma_Z^q\right)$$

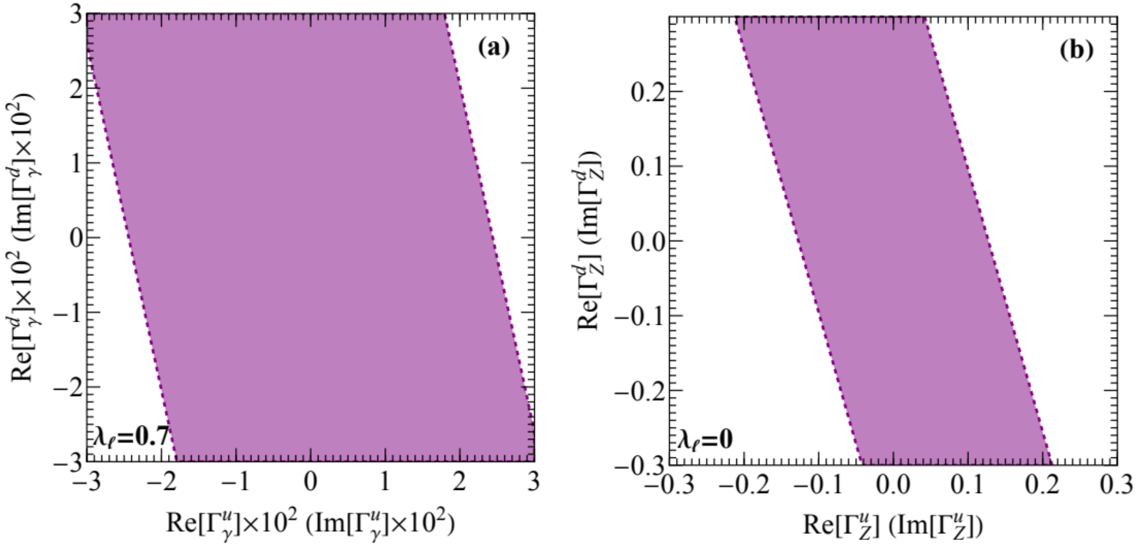
$$s_q^y = \frac{2}{C_q}\left(w_\gamma^q \,\mathrm{Im}\,\Gamma_\gamma^q + w_Z^q \,\mathrm{Im}\,\Gamma_Z^q\right)$$

(w_γ^q, w_Z^q) require parity-violating effects to be nonzero.

- **longitudinal polarization of the electron**
- **parity-odd Z interactions**

$$\frac{d\sigma}{d\phi_R} \propto 1 + A_R(\mathrm{Re}\,\Gamma_\gamma, \mathrm{Re}\,\Gamma_Z)\sin\phi_R + A_I(\mathrm{Im}\,\Gamma_\gamma, \mathrm{Im}\,\Gamma_Z)\cos\phi_R$$

$\gamma\gamma$ channel, $\lambda_e = 0.7$ **γZ channel, $\lambda_e = 0$**



$\sqrt{s} = 105\ \mathrm{GeV}, \mathcal{L} = 1\ \mathrm{ab}^{-1}$

- ☐ **γ dipole: $O(0.01)$**
- ☐ **Z dipole: $O(0.1)$**

Outperform current constraints by 1-2 orders of magnitude

However, only constrains a linear combination of Γ^u and Γ^d.

Conclusion: Key Takeaways from the Talk

- Addressed the critical challenges in determining the proton's longitudinal unpolarized parton distribution functions (PDFs).
- Identified key limitations arising from:
  - Perturbative and nonperturbative QCD interactions
  - Data analysis complexities
  - Statistical uncertainties

- Emphasized the importance of insights from LHC data in shaping future research at the Electron-Ion Collider (EIC) for PDF determination.
  - Discussed the role of lattice QCD calculations in enhancing our understanding of PDFs
  - Highlighted the potential of novel experimental observables to tighten constraints on PDFs and increase sensitivity in the search for New Physics

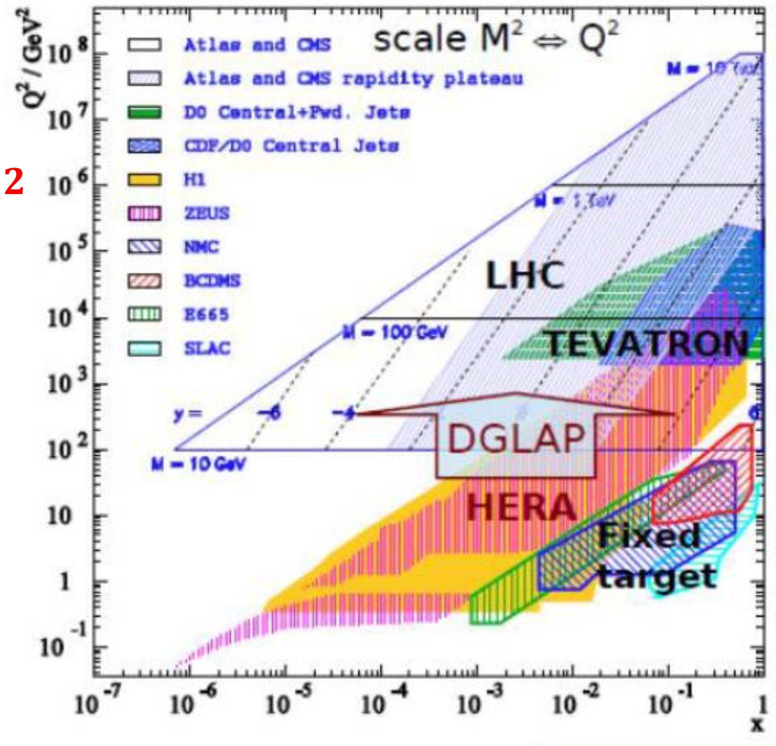
- Concluded that EIC data will be particularly valuable for further constraining PDFs of the proton at large x, paving the way for improved theoretical predictions and experimental tests.

Backup slides

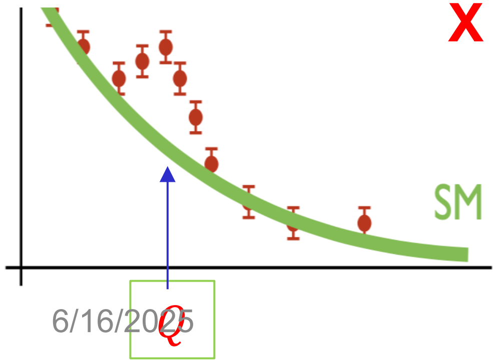
Some basics about PDFs: relevant kinematics in (x, Q^2)

CTEQ

Q^2



X



$$\sigma(Q) \simeq \sum_{i,j} f_{i/p}(x_1, Q) \otimes f_{j/p}(x_2, Q) \otimes \hat{\sigma}_{ij}(x_1, x_2; Q)$$

- Parton Distribution Function $f(x, Q)$
- Given a heavy resonance with mass Q produced at hadron collider with c.m. energy $\sqrt{S}$
- What's the typical x value?

$$<x>= \frac{Q}{\sqrt{S}}$$ at central rapidity (y=0)

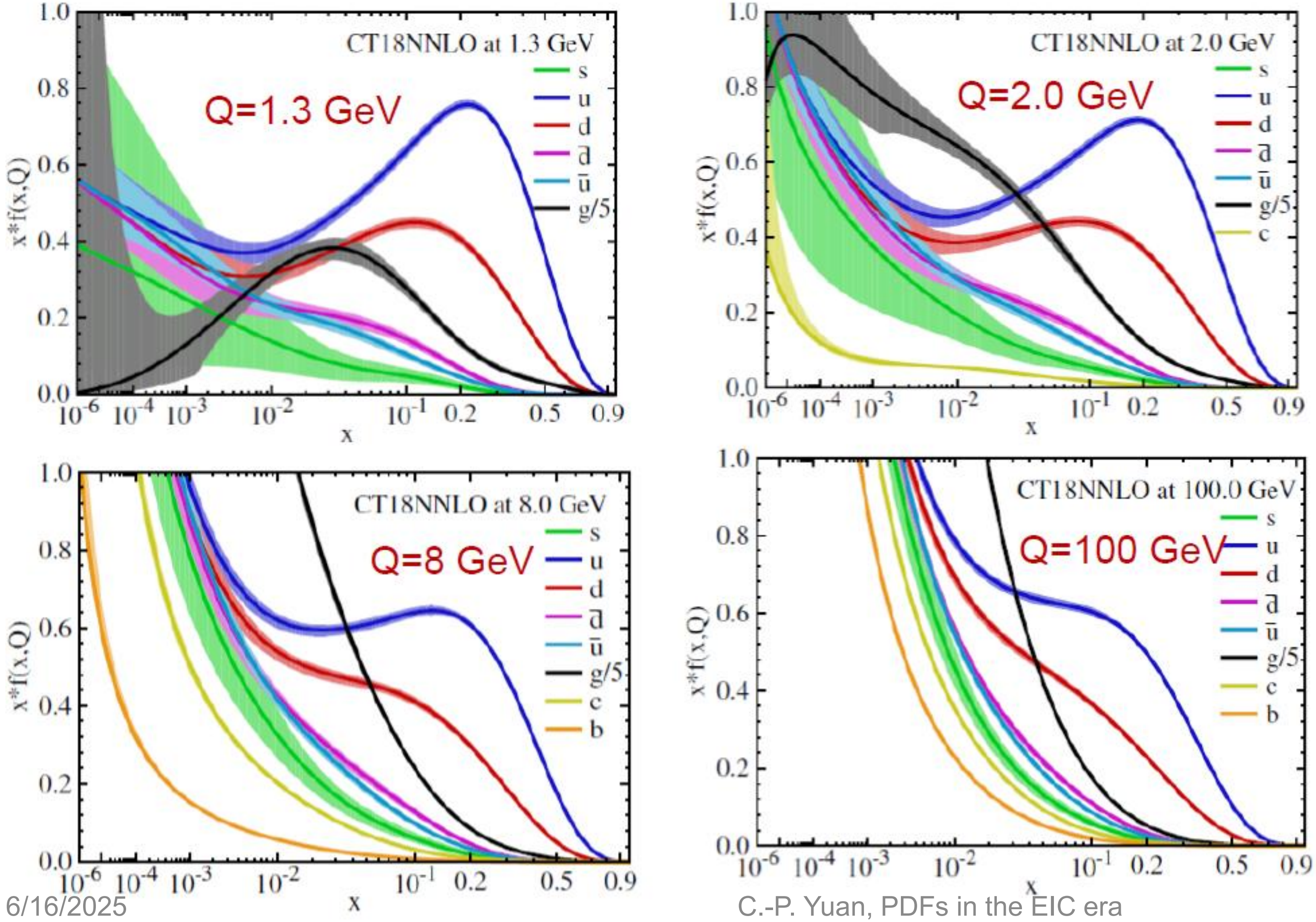
- Generally, $x_1 = \frac{Q}{\sqrt{S}} e^{y}$ and $x_2 = \frac{Q}{\sqrt{S}} e^{-y}$

$$x_1 + x_2 = 2\frac{Q}{\sqrt{S}}\cosh(y) \Rightarrow y_{max} : x_1 + x_2 = 1$$

PDF uncertainties vary as Q via DGLAP evolution

arXiv: 1912.10053

CT18 NNLO PDFs

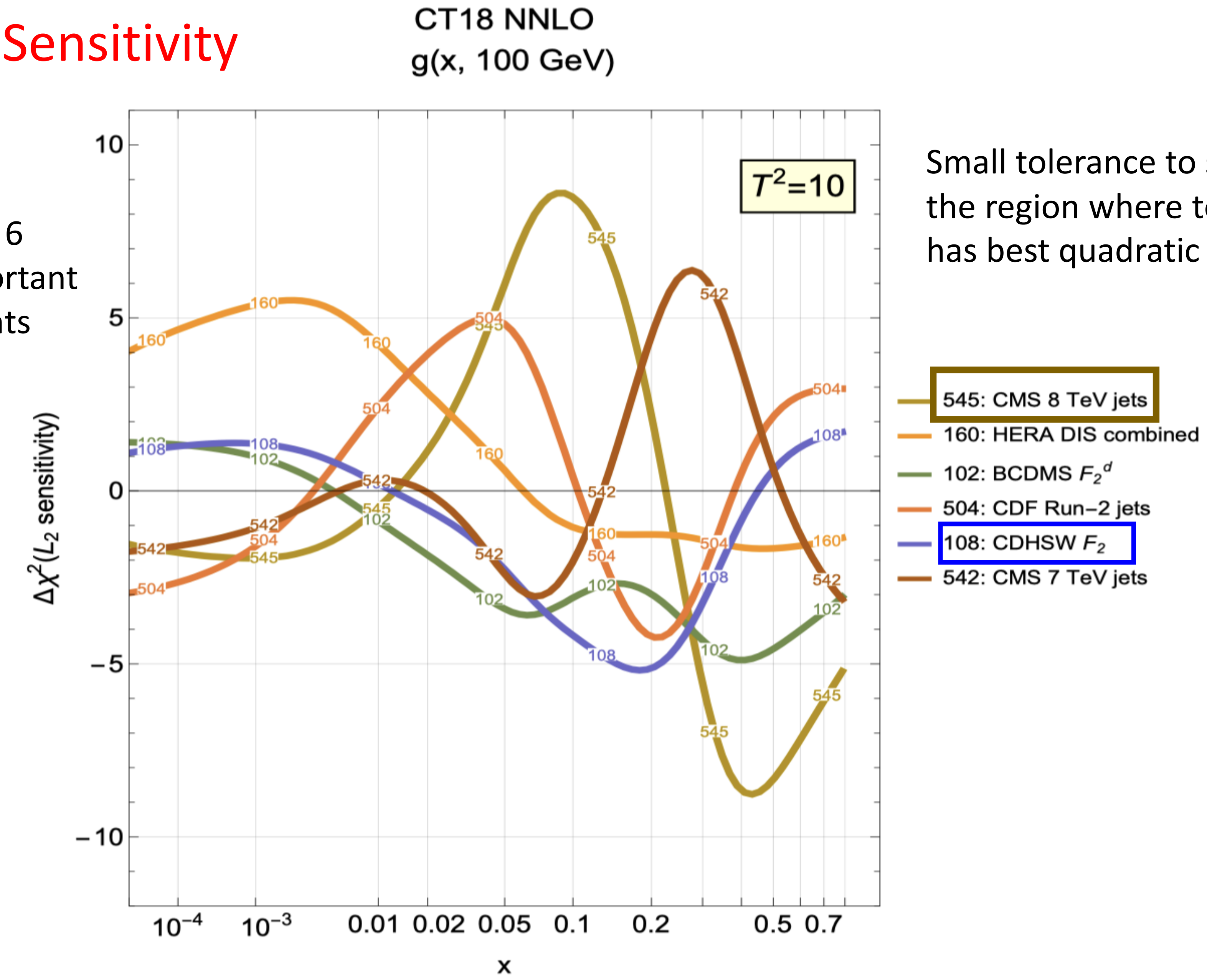


- Faster DGLAP evolution at low Q values.
- Smaller PDF error bands at higher Q values.
- At high Q, perturbaive contribution becomes more important than the non-perturbative part of PDF.

Relatively low energy data, such as HERA I+II, remain crucial for PDF global analysis.

L2 Sensitivity

show only 6 most important experiments



Small tolerance to stay in the region where total χ^2 has best quadratic behavior