The HEP implications of an EIC

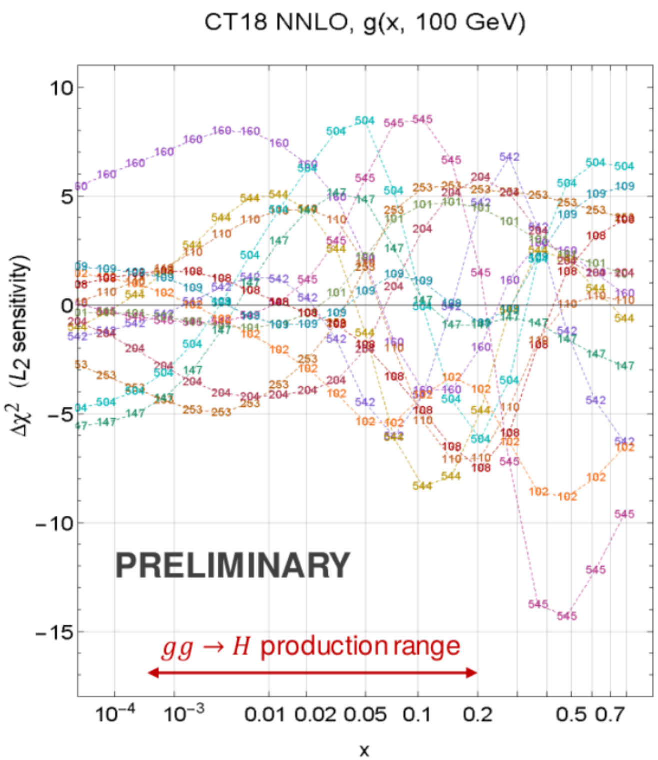
a theory/PDF analysis perspective.

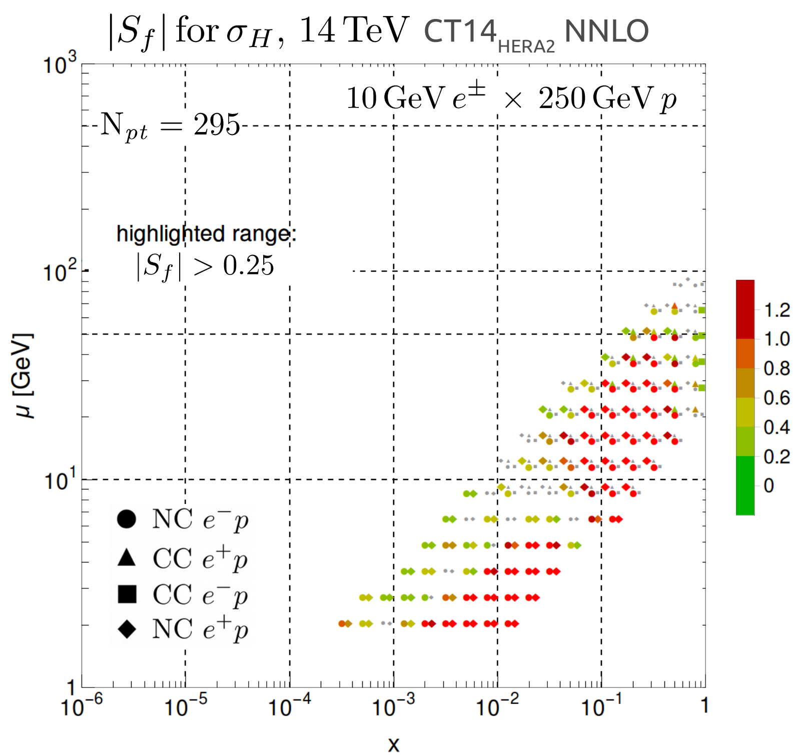
Tim Hobbs, CTEQ@SMU and JLab EIC Center

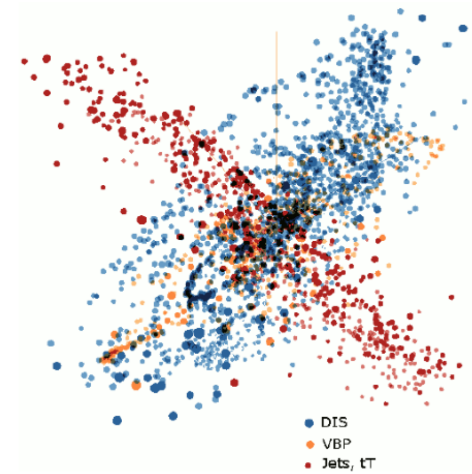
July 25th 2019

Collaboration with Bo-Ting Wang, Pavel Nadolsky, Fred Olness

...and CTEQ-TEA, EIC[2]@JLab...



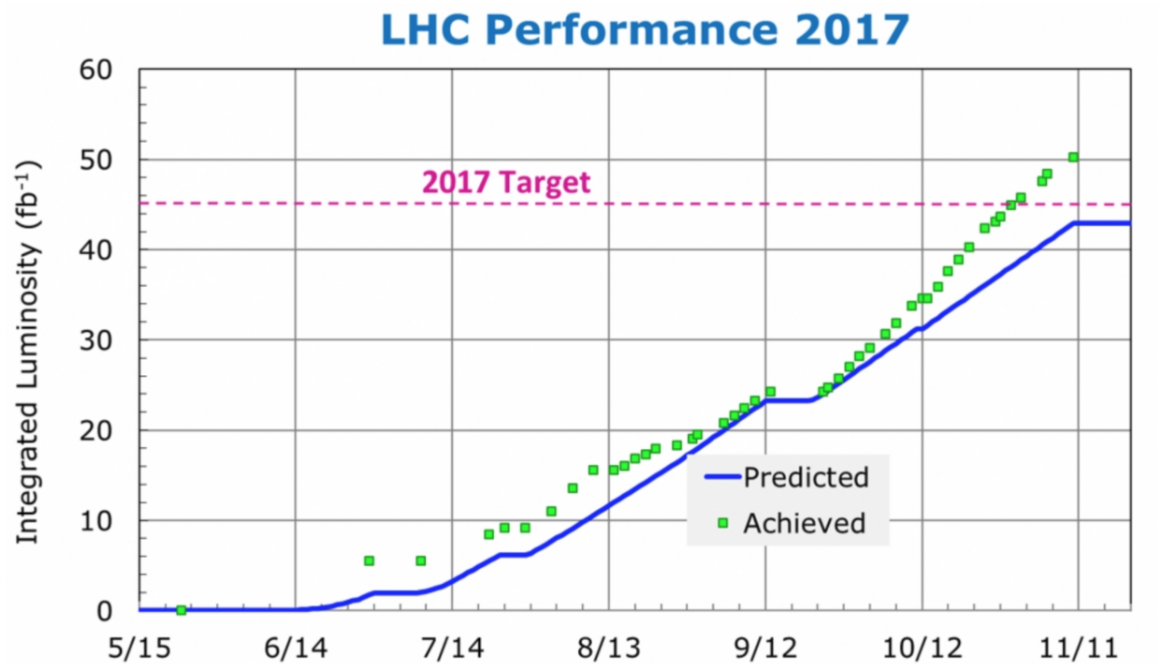




EIC User Group Meeting '19. Paris; 22-26 July 2019.

the view from particle phenomenology

- with the completion of Run-2, LHC has accumulated copious data



- this data is an opportunity, but also a challenge:

$\rightarrow$ many standard-candle measurements are crucially limited by PDF uncertainties, as are new physics searches

(See talk by Paul Newman, 22 July.)

$$\text{e.g., } \sigma_H\,,\ \sin^2\theta_W\,,\ M_W\,,\ \ldots$$

$\rightarrow$ to reach (sub)percent-level precision objectives for HL-LHC, PDF improvements are obligatory; must resolve tensions/pulls in modern PDF analyses

$\longrightarrow$ input from EIC will be essential

Higgs, g(x)

a typical example: σ_H and PDF, $\boldsymbol{\alpha}_s$ uncertainties

- there remains considerable dependence (as large as ~13%) upon PDF paramatrization and running coupling

→ the situation is such that precision in Higgs phenom. is significantly **PDF-limited**

Accardi et al., EPJC**76**, 471 (2016).

PDF sets	$\sigma(H)^{\mathrm{NNLO}}$ (pb) nominal $\alpha_s(M_Z)$	$\sigma(H)^{\mathrm{NNLO}}$ (pb) $\alpha_s(M_Z) = 0.115$	$\sigma(H)^{\mathrm{NNLO}}$ (pb) $\alpha_s(M_Z) = 0.118$
ABM12 [2]	39.80 ± 0.84	41.62 ± 0.46	44.70 ± 0.50
CJ15 [1][a]	$42.45^{+0.43}_{-0.18}$	$39.48^{+0.40}_{-0.17}$	$42.45^{+0.43}_{-0.18}$
CT14 [3][b]	$42.33^{+1.43}_{-1.68}$	$39.41^{+1.33}_{-1.56}$ (40.10)	$42.33^{+1.43}_{-1.68}$
HERAPDF2.0 [4][c]	$42.62^{+0.35}_{-0.43}$	$39.68^{+0.32}_{-0.40}$ (40.88)	$42.62^{+0.35}_{-0.43}$
JR14 (dyn) [5]	38.01 ± 0.34	39.34 ± 0.22	42.25 ± 0.24
MMHT14 [6]	$42.36^{+0.56}_{-0.78}$	$39.43^{+0.53}_{-0.73}$ (40.48)	$42.36^{+0.56}_{-0.78}$
NNPDF3.0 [7]	42.59 ± 0.80	39.65 ± 0.74 (40.74 ± 0.88)	42.59 ± 0.80
PDF4LHC15 [8]	42.42 ± 0.78	39.49 ± 0.73	42.42 ± 0.78

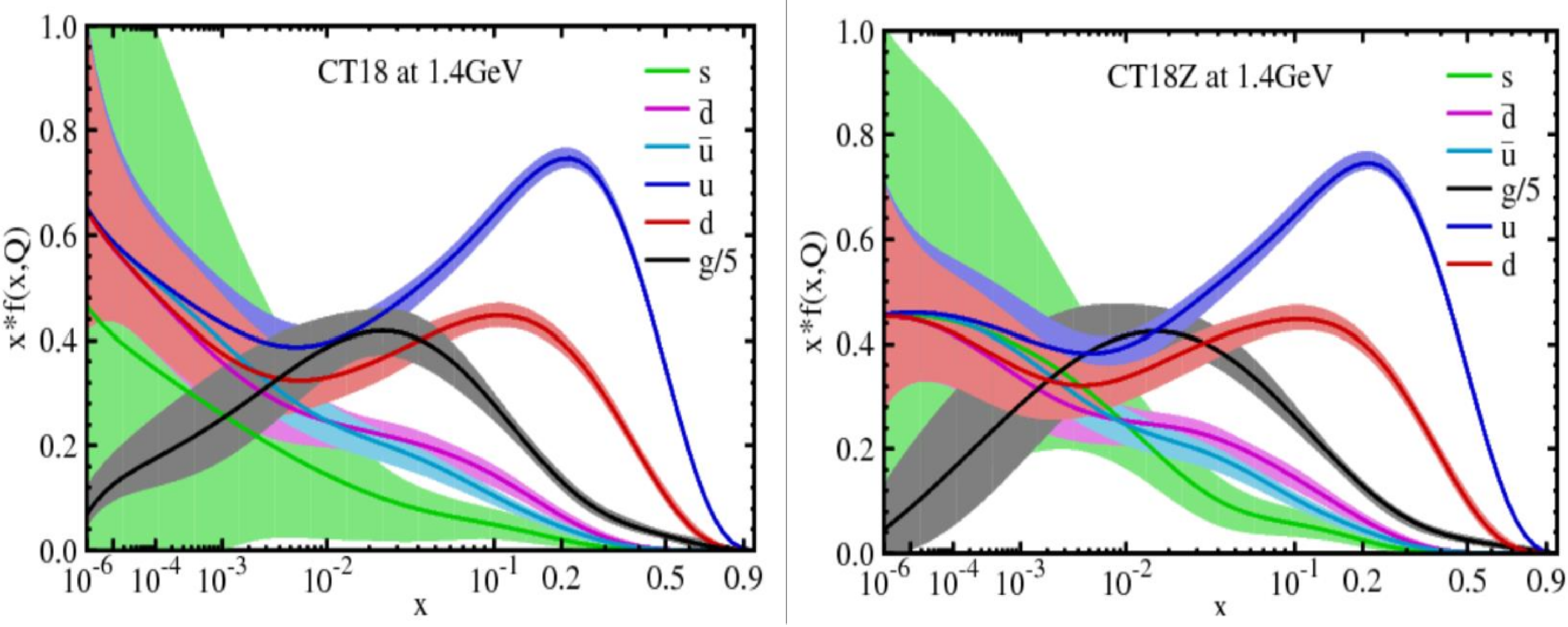
$$\sigma_H \text{ at NNLO and } \sqrt{s} = 13\,\text{TeV}; \;\; \mu_F = \mu_R = m_H$$

→ enhancing the discovery potential in the Higgs sector will require improving these uncertainties!

CT18 parton distributions

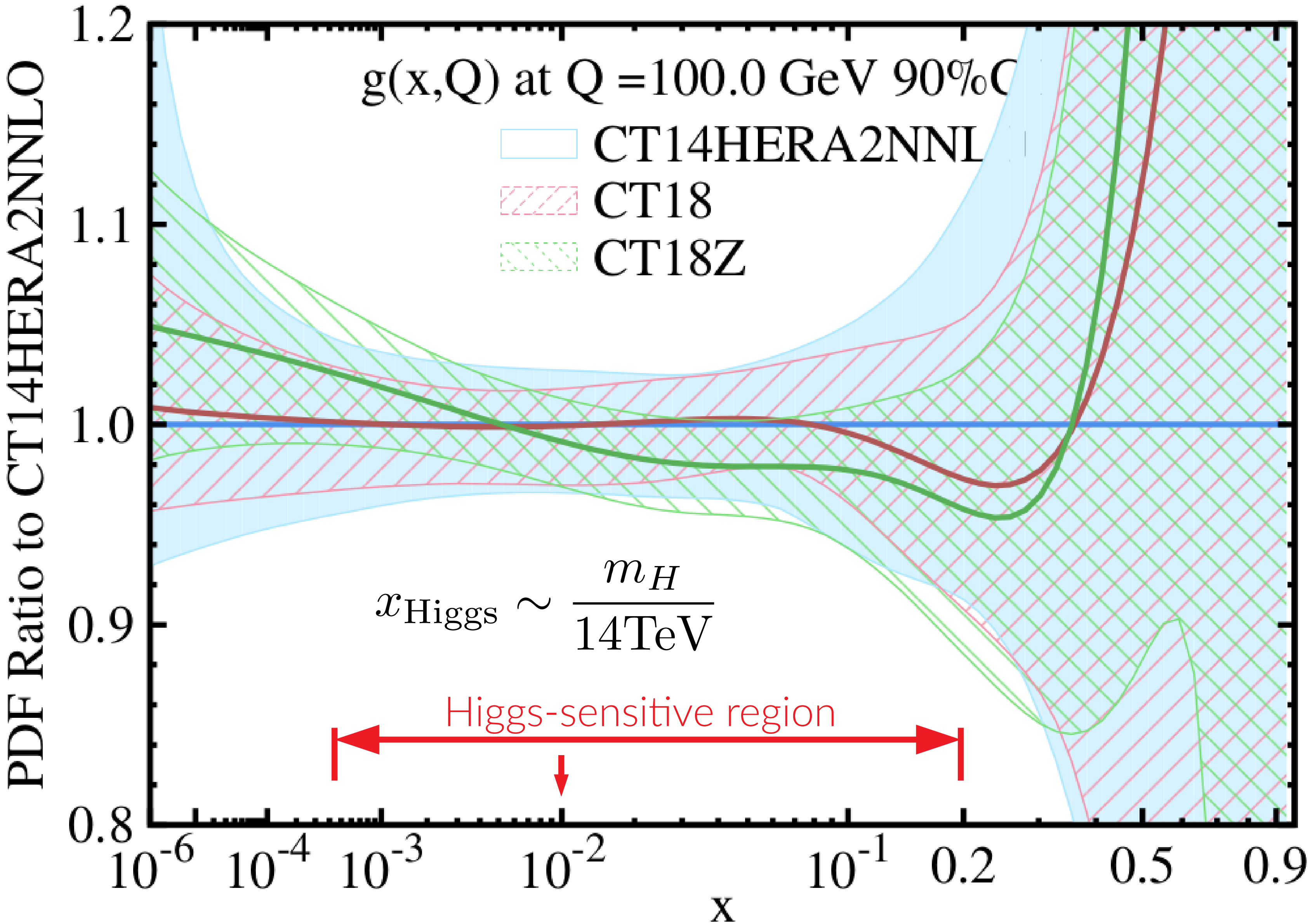
PDF analyses are challenging! (theoretically, computationally, statistically, …)

Four PDF ensembles: CT18 (default), A, X, and Z



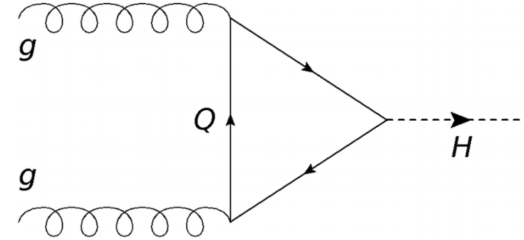
CT18Z has enhanced gluon and strange PDFs at $x \sim 10^{-4}$, and reduced light-quark PDFs at $x < 10^{-2}$. The CT18Z fit is performed so as to maximize the differences from CT18 PDFs, while preserving about the same goodness-of-fit as for CT18. CT18A and CT18X include some features of CT18Z

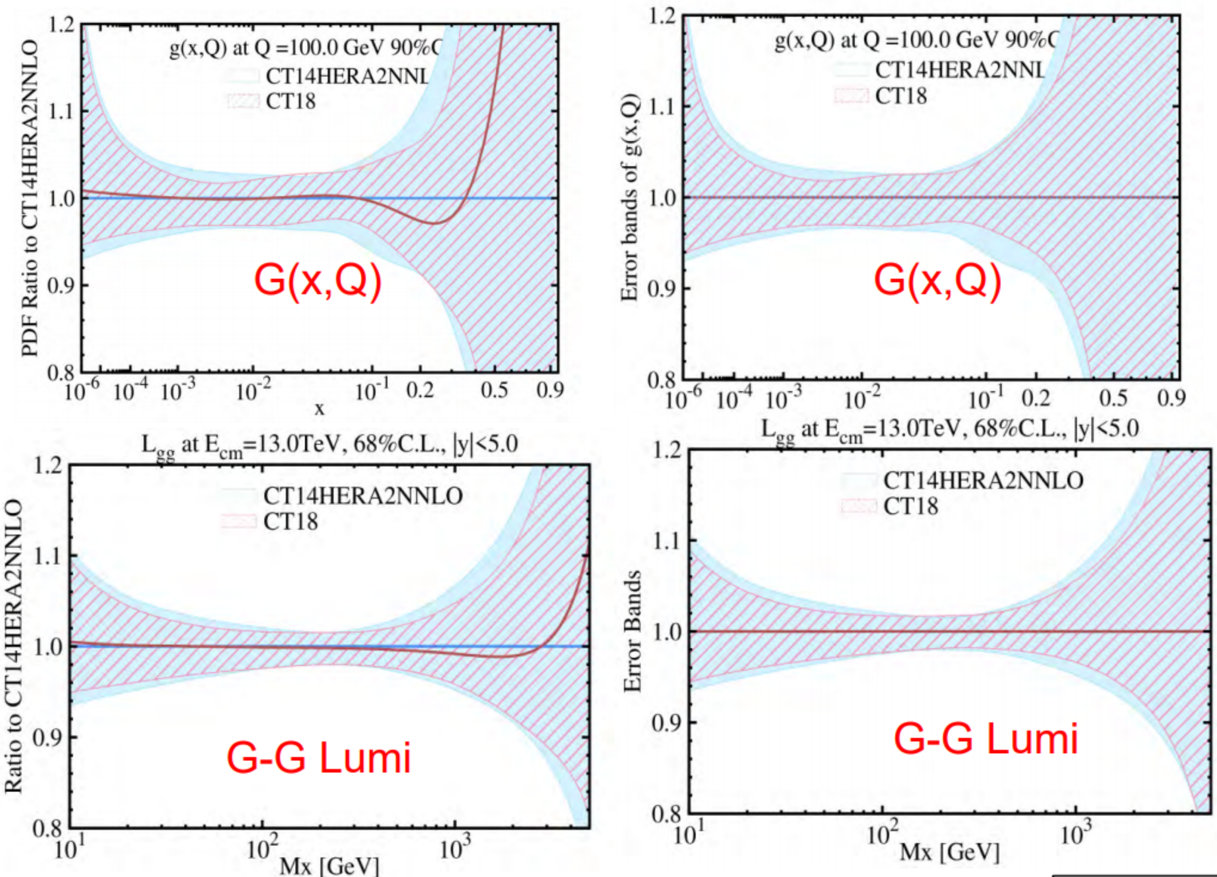
LHC Run-1 gluon PDF impact in CT14 → CT18(Z)



- while LHC Run-1 data drive important PDF improvements, including for the gluon at high-, low-x, the effect is relatively incremental

CT14 → CT18 modestly shifts Higgs cross sections and slightly reduces PDF uncertainties





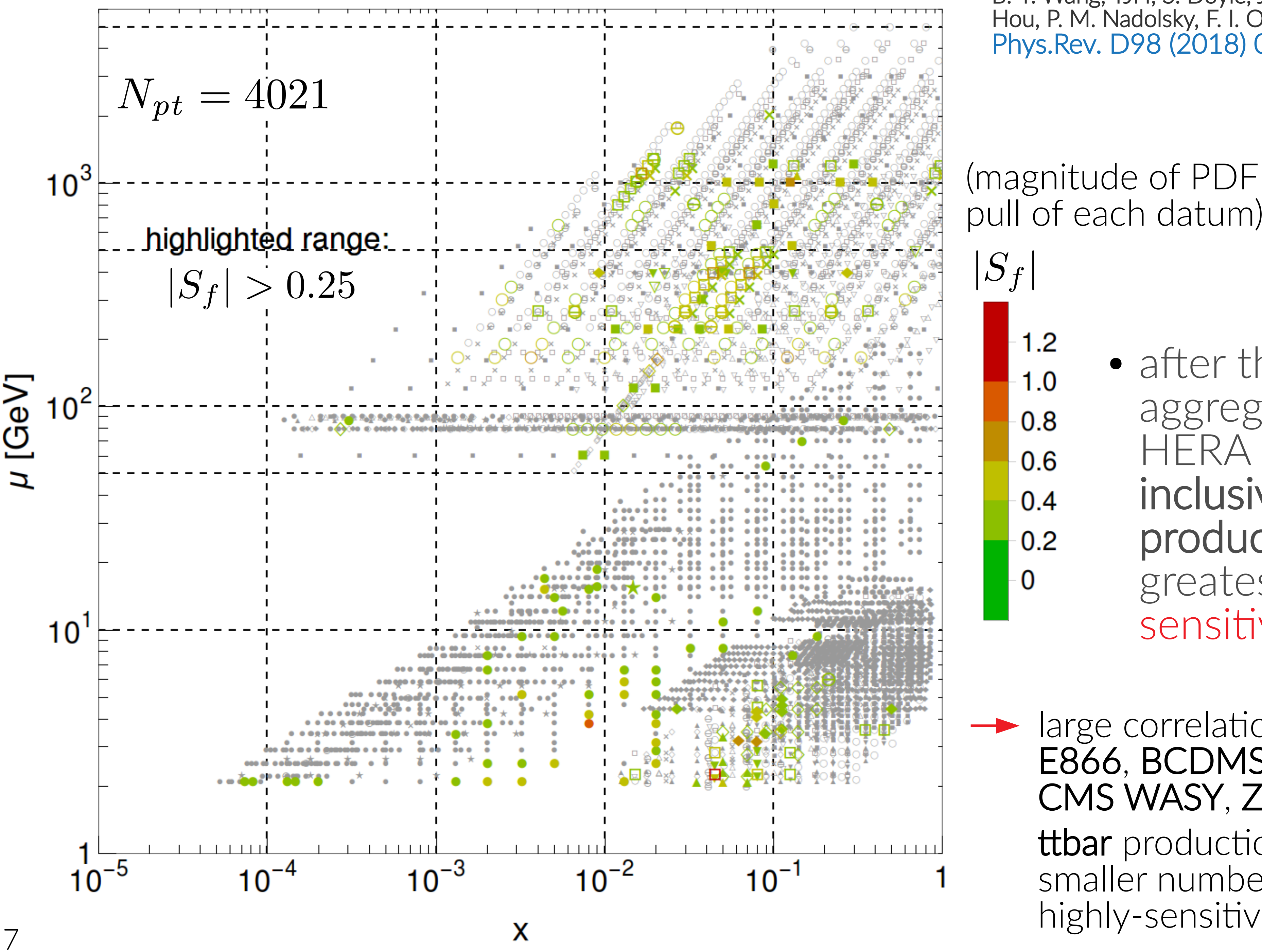
	σ(gg-h)	$\delta\sigma$ sym(90%C.L.)
7 TeV		
CT14NNLO	14.67	0.46
CT18	14.57	0.44
8 TeV		
CT14NNLO	18.70	0.57
CT18	18.45	0.55
13 TeV		
CT14NNLO	42.78	1.32
CT18	42.43	1.26
14 TeV		
CT14NNLO	48.23	1.50
CT18	47.91	1.42

PDF induced errors (at 90% CL) are reduced by about 5% as compared to CT14 predictions.

→ can we disentangle elements of the global analysis responsible for these improvements?

| S_f | for σ_{H^0} 14 TeV, CT14$_{\text{HERA2}}$ NNLO

B.-T. Wang, TJH, S. Doyle, J. Gao, T.-J. Hou, P. M. Nadolsky, F. I. Olness
Phys.Rev. D98 (2018) 094030

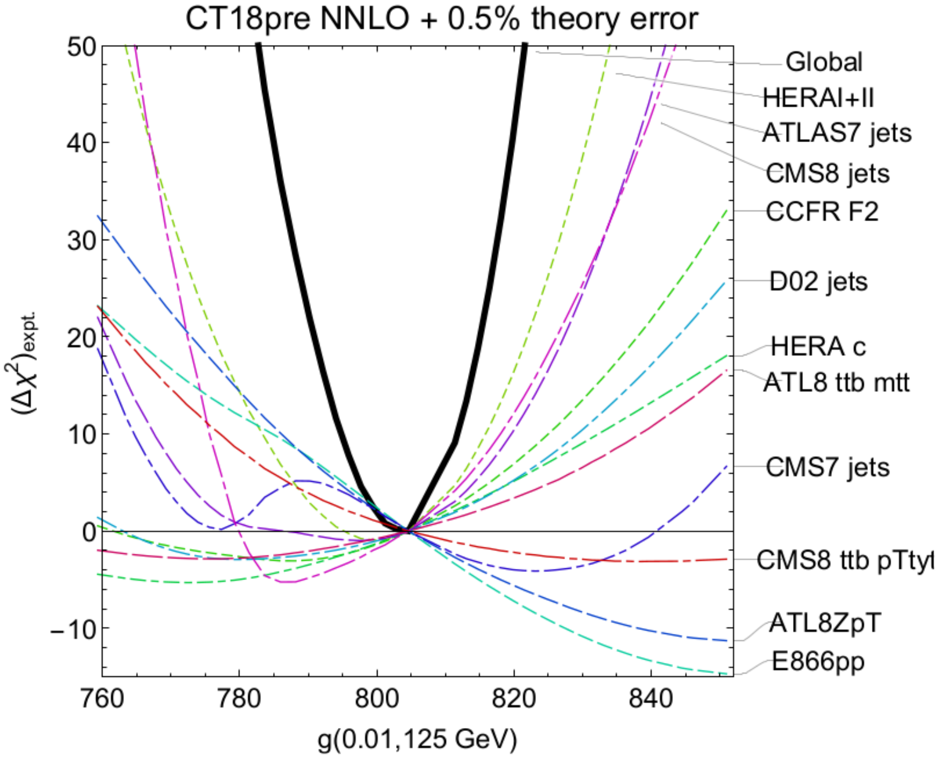


- after the aggregated HERA data, **inclusive jet production** – greatest total sensitivity!

→ large correlations for **E866**, **BCDMS**, **CCFR**, **CMS WASY**, **Z p_T** and **ttbar** production, but smaller numbers of highly-sensitive points

we use the Higgs region g(x) to validate `PDFSense`

…for the gluon PDF in the Higgs region, $g(0.01, m_H)$

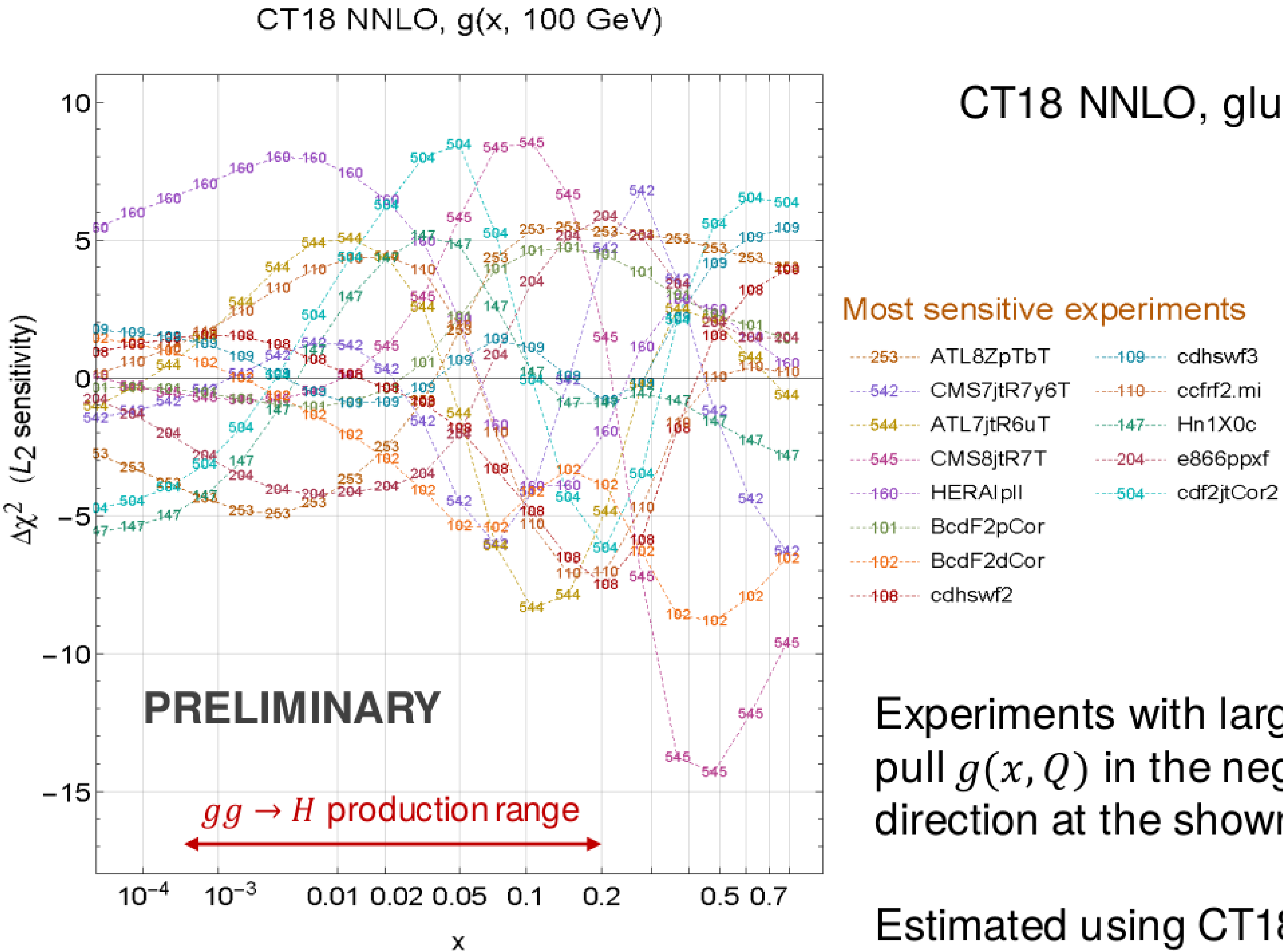


$g(x=0.01, \mu=125$ GeV)		
PDFSense		LM scan
CT14HERA2	CT18pre	CT18pre
HERAI+II'15	HERAI+II'15	HERAI+II'15
CMS8jets'17	CMS8jets'17	CMS8jets'17
CMS7jets'14	CMS7jets'14	ATL8ZpT'16
ATLAS7jets'15	E866pp'03	E866pp'03
E866pp'03	ATLAS7jets'15	ATLAS7jets'15
BCDMSd'90	BCDMSd'90	CCFR-F2'01
CCFR-F3'97	BCDMSp'89	D02jets'08
D02jets'08	D02jets'08	HERAc'13
NMCrat'97	NMCrat'97	NuTeV-nub'06
BCDMSp'89	CDHSW-F2'91	CCFR-F3'97

- PDFSense identifies the most sensitive experiments with high confidence and in accord with other methods such as the LM scans. It works the best when the uncertainties are nearly Gaussian, and experimental constraints agree among themselves [arXiv:1803.02777]

Estimated χ^2 pulls from experiments

(L_2 sensitivity, arXiv:1904.00222, v. 2)



CT18 NNLO, gluon at Q=100 GeV

stronger
(anti-)
correlation

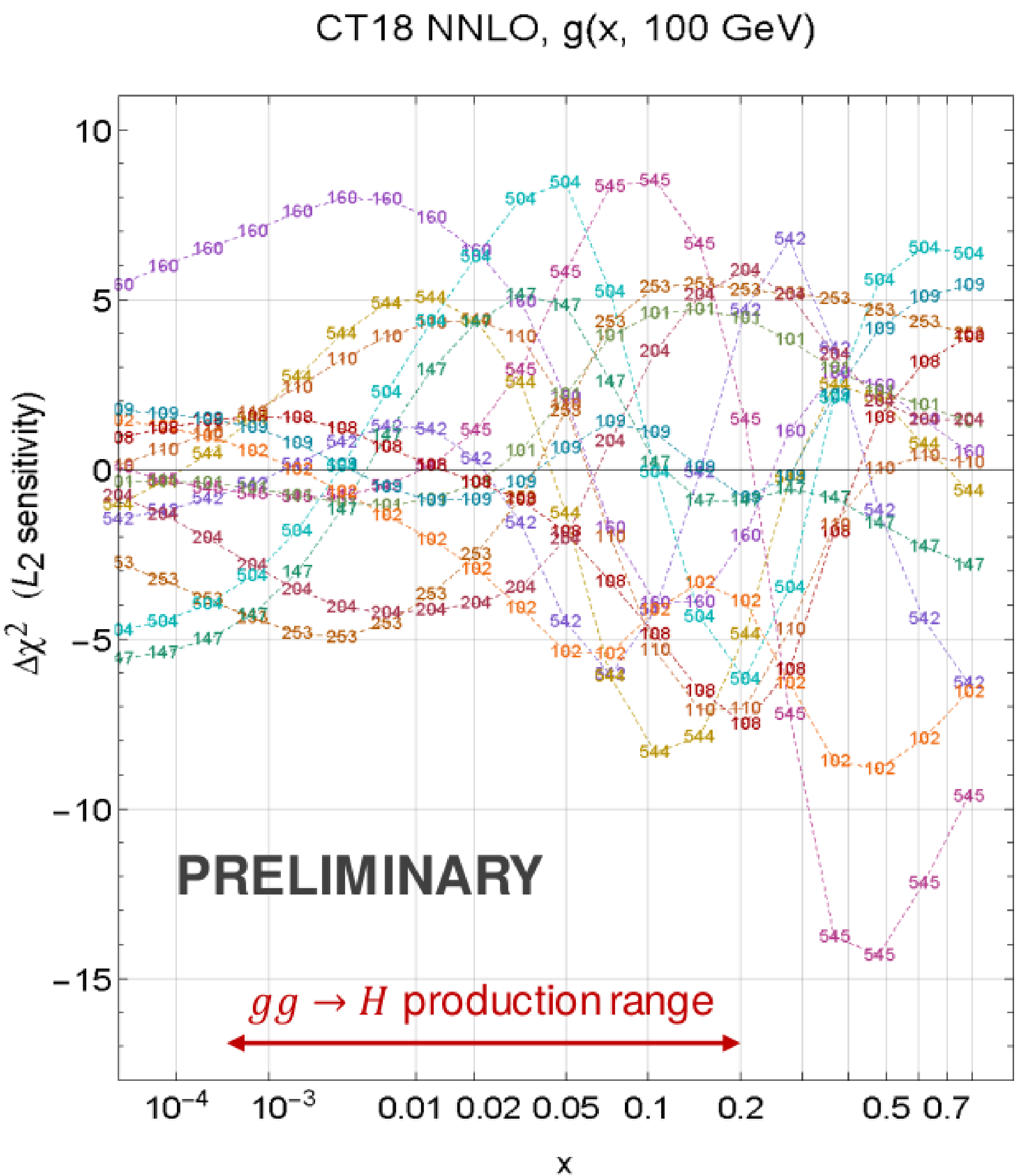
Experiments with large $\Delta\chi^2 > 0$ [$\Delta\chi^2 < 0$] pull $g(x, Q)$ in the negative [positive] direction at the shown x

Estimated using CT18 Hessian PDFs

$$S_{f,\,L_2} \sim \mathrm{Corr}[f_a, \chi^2_E]$$

Estimated χ^2 pulls from experiments

(L_2 sensitivity, arXiv:1904.00222, v. 2)



precise data from EIC sensitive to the gluon PDF Higgs region needed to help unravel the systematic tensions evident here

Most sensitive experiments

- 253 ATL8ZpTbT
- 542 CMS7jtR7y6T
- 544 ATL7jtR6uT
- 545 CMS8jtR7T
- 160 HERAIpII
- 101 BcdF2pCor
- 102 BcdF2dCor
- 108 cdhswf2
- 109 cdhswf3
- 110 ccfrf2.mi
- 147 Hn1X0c
- 204 e866ppxf
- 504 cdf2jtCor2

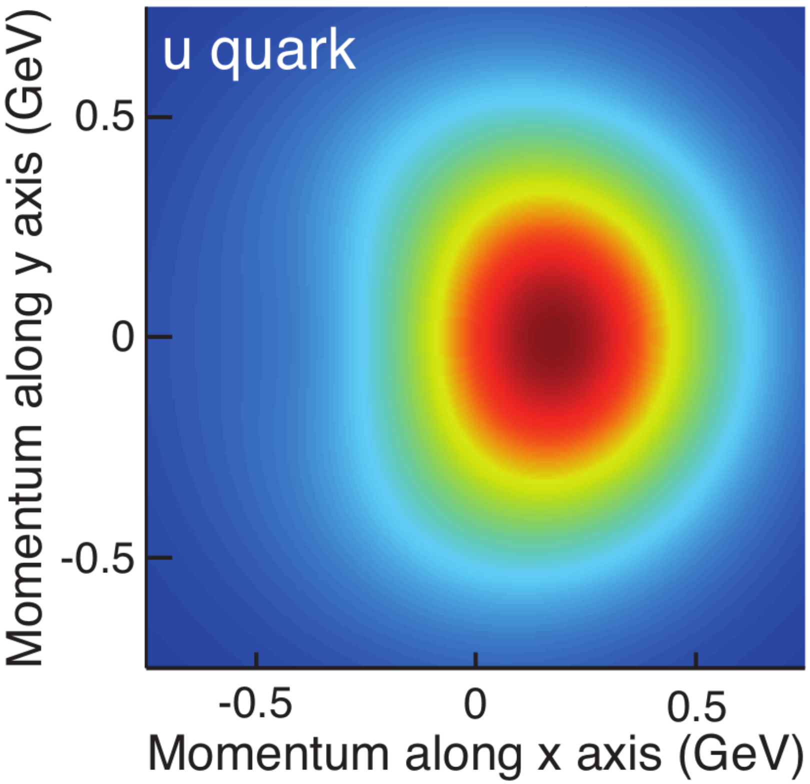
Note opposite pulls (tensions) in some x ranges between HERA I+II DIS (ID=160); CDF (504), ATLAS 7 (544), CMS 7 (542), CMS 8 jet (545) production; E866pp DY (204); ATLAS 8 Z pT (253) production; BCDMS and CDHSW DIS

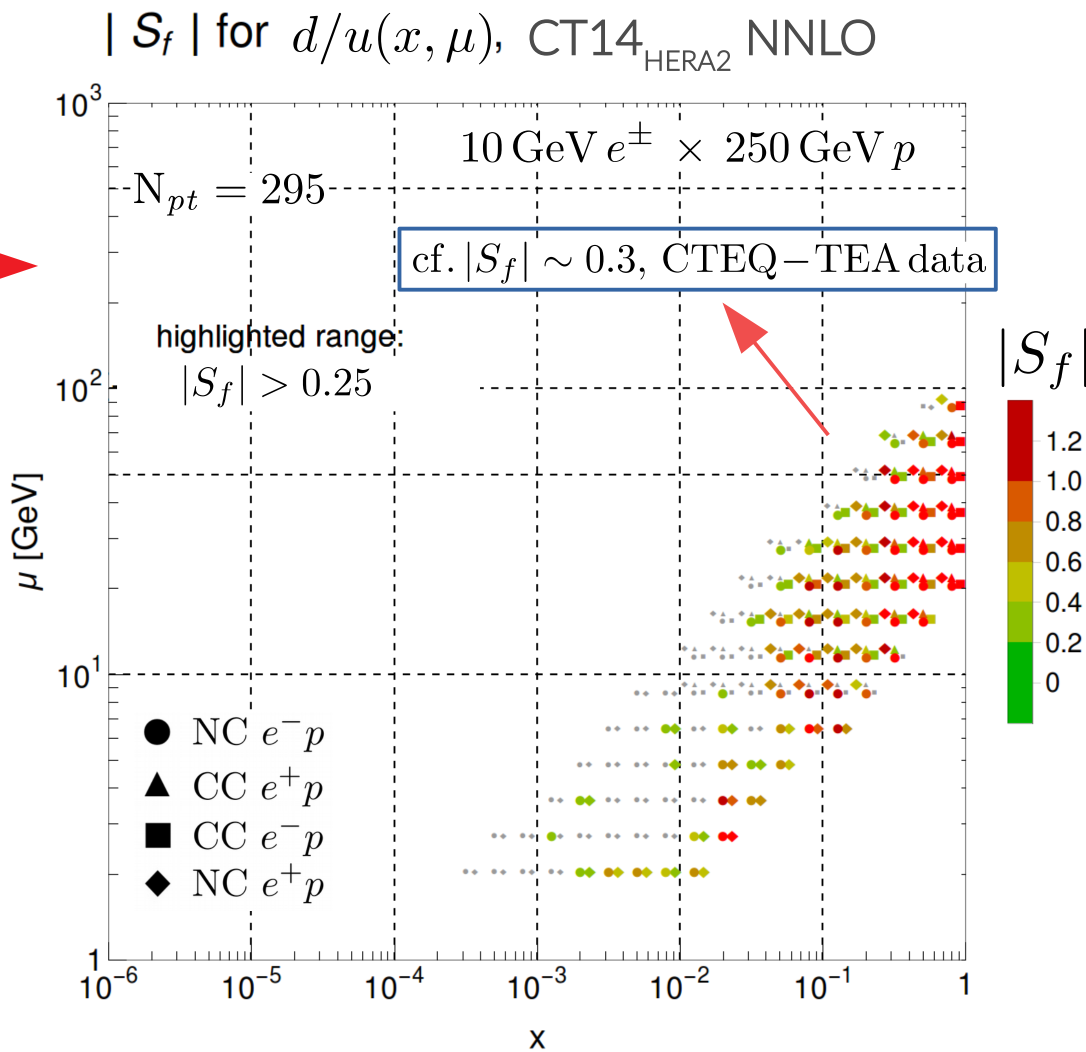
$$S_{f,\,L_2} \sim \mathrm{Corr}[f_a, \chi^2_E]$$

the EIC tomography program will deliver high-precision DIS

- by measuring the nucleon's multi-dimensional wave function with high precision, the EIC will hugely constrain proton collinear structure

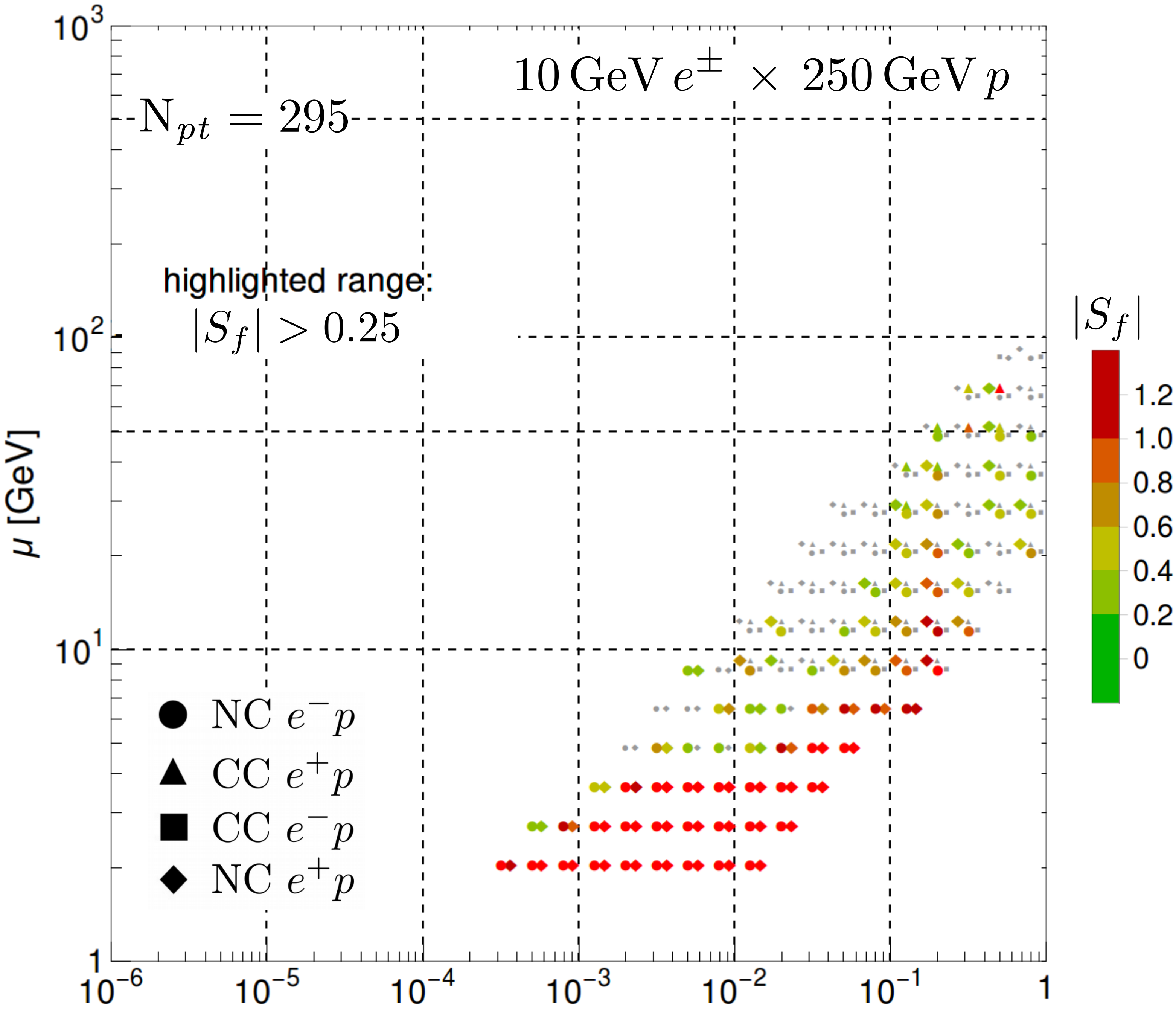
Accardi et al., EPJA52 (2016) no.9, 268.





- DIS cross sections from EIC will supercede the bulk of fixed-target information in contemporary QCD fits; provide an 'anchor-point' to resolve systematic PDF tensions

$|S_f|$ for $g(x, \mu)$ CT14 HERA2 NNLO

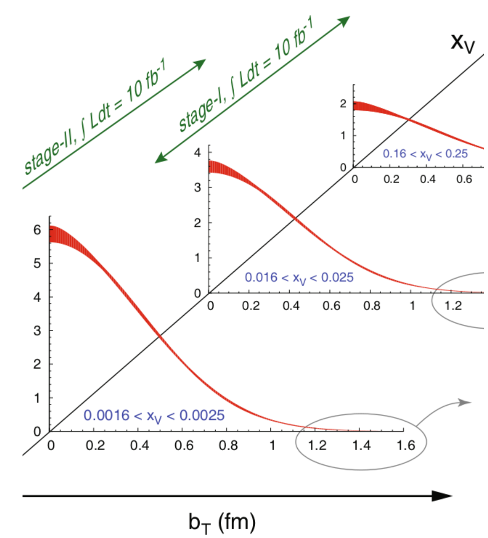


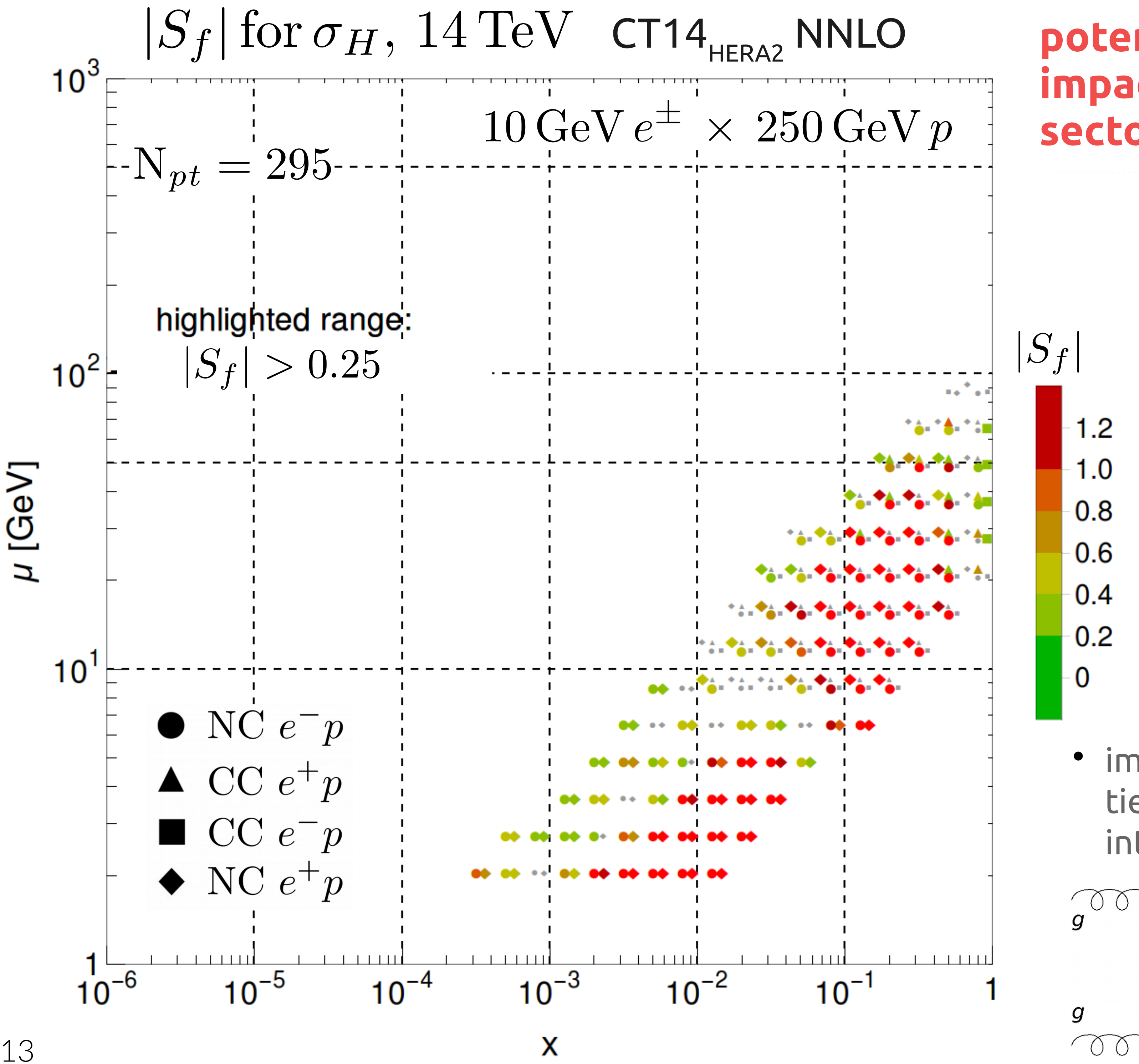
- an EIC will provide a sensitive probe to the gluon distribution – especially at low x

  $x \gtrsim 3 \times 10^{-4}$

- these constraints arise from high statistics neutral current data on $\sigma_{r,\mathrm{NC}}^{e^{\pm}p}$

→ complement to tomography:

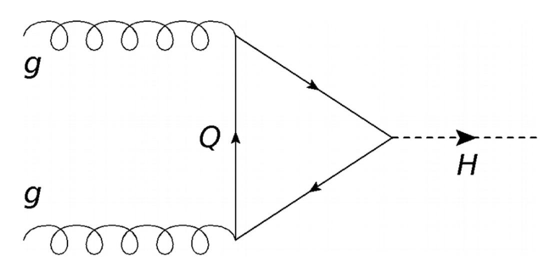




potentially strong impact on the Higgs sector

- the impact of an EIC upon the theoretical predictions for inclusive Higgs production arises from a very broad region of the kinematical space it can access

- impact rather closely tied to that of the integrated gluon PDF:



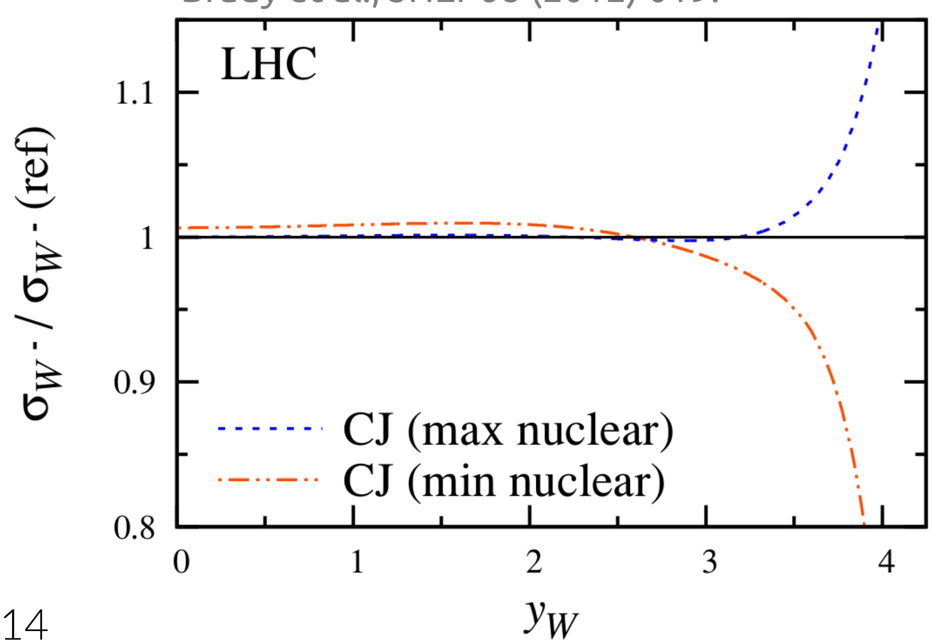
EIC and an era of (higher) precision **electroweak physics** (?)

- theory predictions for the production of gauge bosons are quite sensitive to the nucleon PDFs: *e.g.*, d(x) at x ~ 1, which is poorly constrained

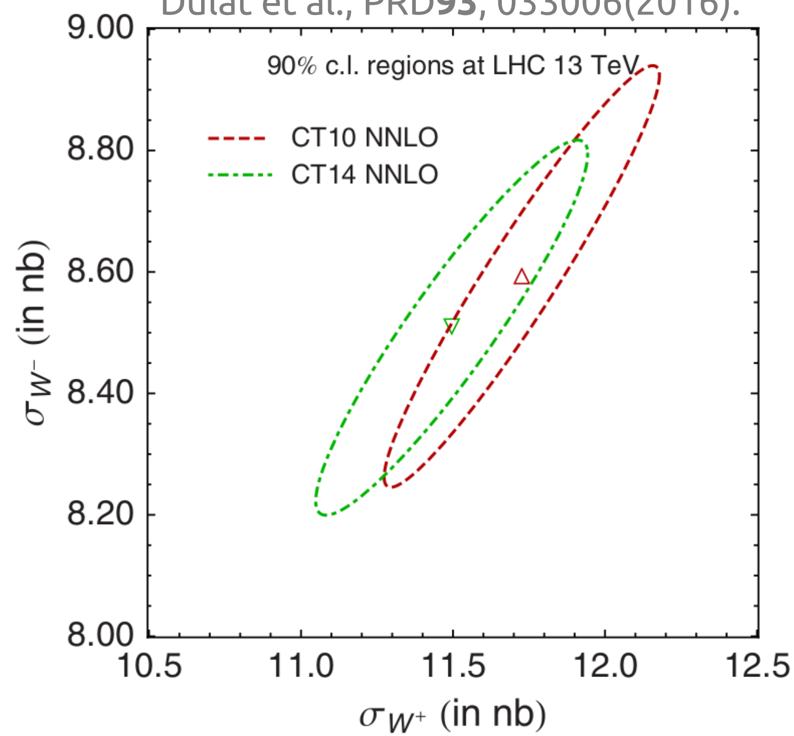
$$x_{1,2} = \frac{M}{\sqrt{s}} e^{\pm y}$$

$$\frac{d\sigma}{dy}(pp \to W^- X) = \frac{2\pi G_F}{3\sqrt{2}} x_1 x_2 \Big(\cos^2\theta_C \{ d(x_1)\bar{u}(x_2) + \bar{u}(x_1) d(x_2) \} + \sin^2\theta_C \{ s(x_1)\bar{u}(x_2) + \bar{u}(x_1) s(x_2) \} \Big)$$

Brady et al., JHEP**06** (2012) 019.



Dulat et al., PRD**93**, 033006(2016).



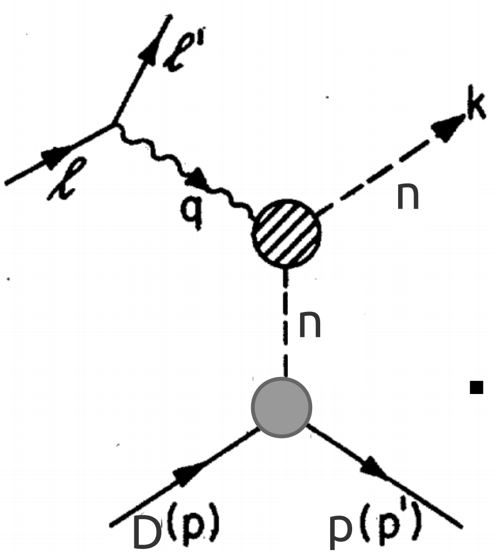
historically, extractions of $d(x)$, $x \to 1$ have depended on nuclear targets (and corrections!)

- in principle, a neutron target would allow the flavor separation needed to access $d(x, Q^2)$

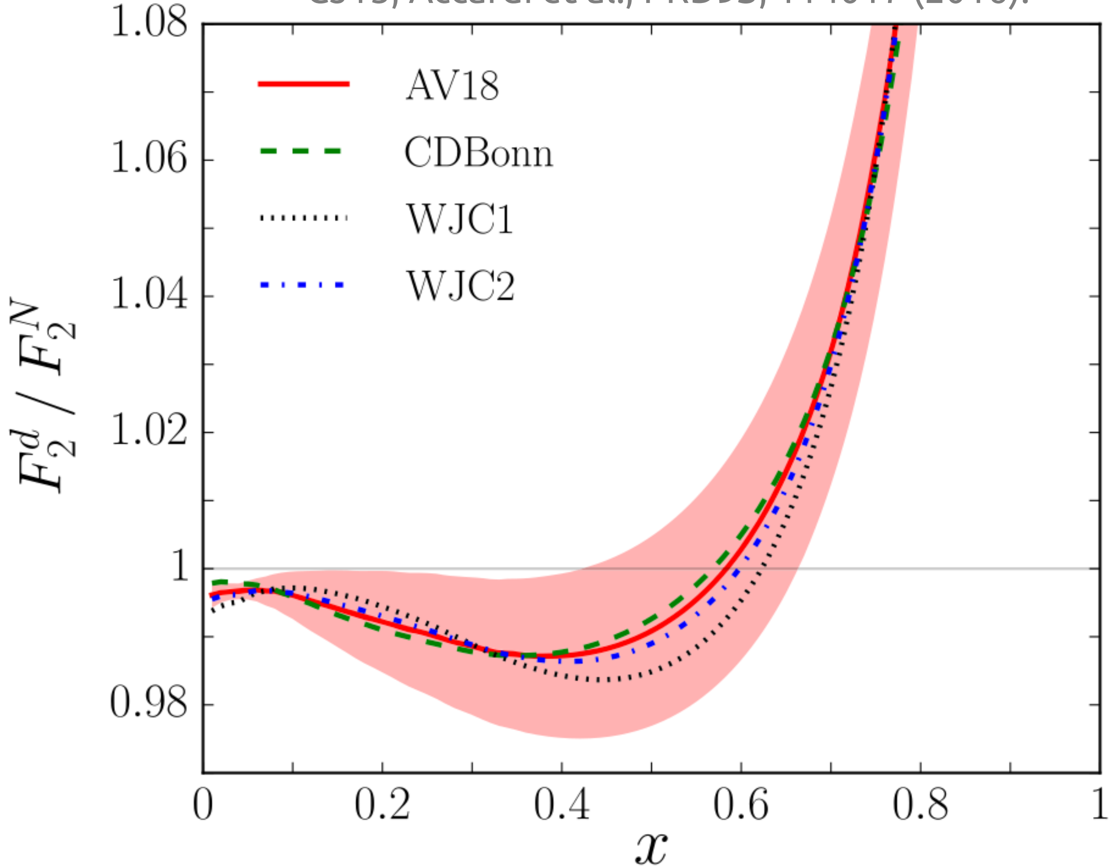
$$F_2^{e^- n} \sim x(4d + u)/9$$

$$vs$$

$$F_2^{e^- p} \sim x(4u + d)/9$$



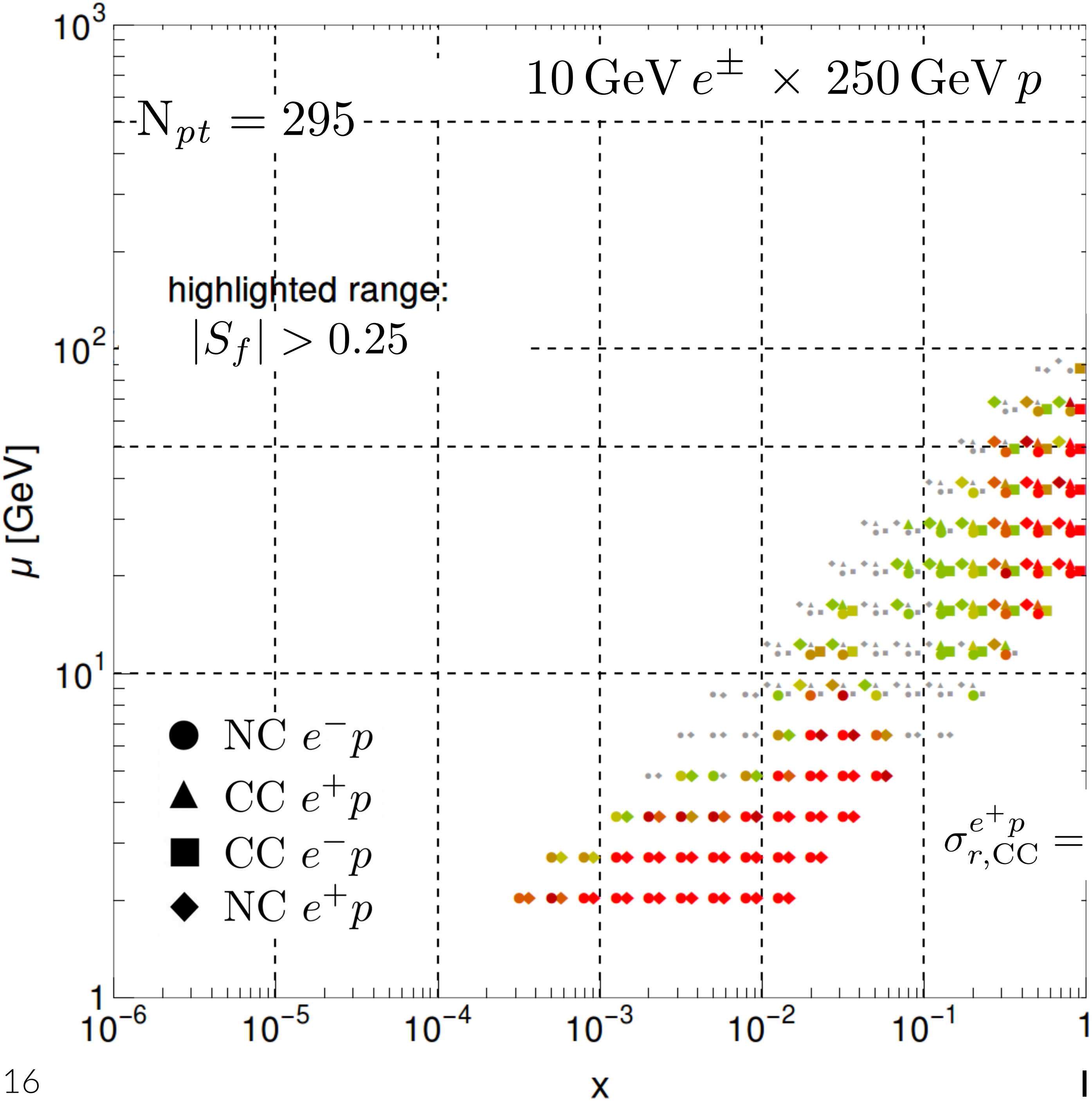
CJ15, Accardi et al., PRD**93**, 114017 (2016).



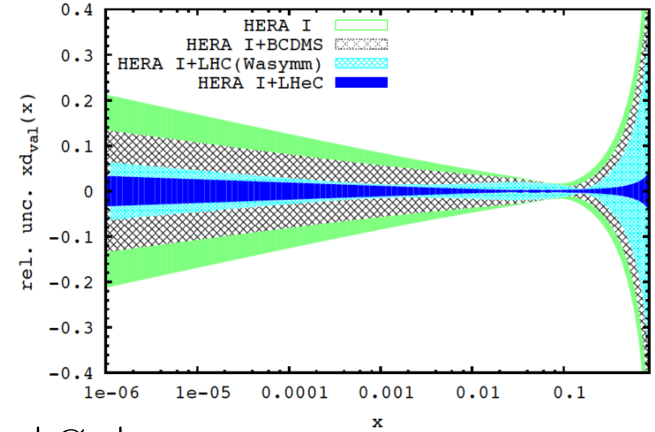
- **BUT**: in the absence of a free neutron target, scattering from nuclei (e.g., the deuteron) is necessary

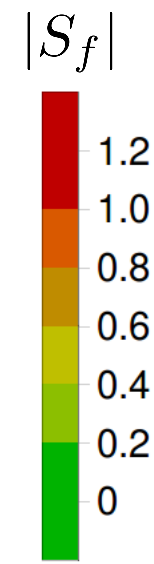
→ **nuclear corrections (Fermi motion) are sizable, especially for large x**

$|\,S_f\,|$ for $d(x,\mu)$ CT14 HERA2 NNLO



EIC Whitepaper:1206.2913





- an EIC affords **strong sensitivities** **without a nuclear target**; here, at both very high and very low x

for $x \to 1$

$$\sigma_{r,CC}^{e^+p} = \frac{Y_+}{2}W_2^+ \mp \frac{Y_-}{2}xW_3^+ - \frac{y^2}{2}W_L^+$$

$$\simeq [1-y]^2\, x(d+s)$$

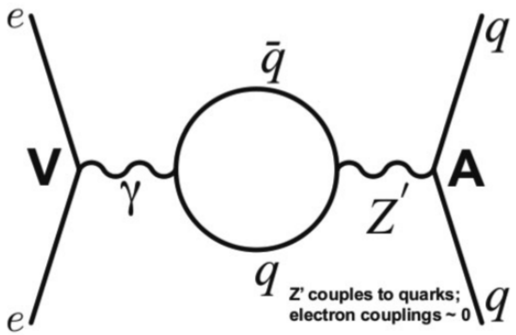
In the **LO quark-parton model**

the electroweak sector and **New Physics** searches at EIC

- if measured to sufficient precision, the quark-level electroweak couplings may be sensitive to an extended EW sector, e.g., Z'

$$\mathcal{L}^{\rm PV} = \frac{G_F}{\sqrt{2}} \left[\bar{e}\gamma^\mu\gamma_5 e \left(C_{1u}\bar{u}\gamma_\mu u + C_{1d}\bar{d}\gamma_\mu d\right) + \bar{e}\gamma^\mu e \left(C_{2u}\bar{u}\gamma_\mu\gamma_5 u + C_{2d}\bar{d}\gamma_\mu\gamma_5 d\right)\right]$$

$$C_{1u} = -\frac{1}{2} + \frac{4}{3}\sin^2\theta_W$$



- a unique strength of an EIC is its combination of very high precision and **beam polarization**, which allows the observation of **parity-violating helicity asymmetries**:

$$A^{\rm PV} = \frac{\sigma_R - \sigma_L}{\sigma_R + \sigma_L} \quad ({\rm R/L} : e^- \text{ beam helicities})$$

selects γ-Z interference diagrams!

TJH and Melnitchouk, PRD**77**, 114023 (2008).

$$A^{\rm PV} = -\left(\frac{G_F Q^2}{4\sqrt{2}\pi\alpha}\right)(Y_1\ a_1 + Y_3\ a_3)$$

$$a_1 = \frac{2\sum_q e_q\ C_{1q}\ (q+\bar{q})}{\sum_q e_q^2\ (q+\bar{q})} \qquad a_3 = \frac{2\sum_q e_q\ C_{2q}\ (q-\bar{q})}{\sum_q e_q^2\ (q+\bar{q})}$$

the electroweak sector and **New Physics** searches at EIC

- if measured to sufficient precision, the quark-level electroweak couplings may be sensitive to an extended EW sector, e.g., Z'

$$\mathcal{L}^{\mathrm{PV}} = \frac{G_F}{\sqrt{2}} \left[\bar{e}\gamma^\mu\gamma_5 e \left(C_{1u}\bar{u}\gamma_\mu u + C_{1d}\bar{d}\gamma_\mu d \right) + \bar{e}\gamma^\mu e \left(C_{2u}\bar{u}\gamma_\mu\gamma_5 u + C_{2d}\bar{d}\gamma_\mu\gamma_5 d \right) \right]$$

$$C_{1u} = -\frac{1}{2} + \frac{4}{3}\sin^2\theta_W$$

→ with sufficient precision, an EIC (which will be statistics-limited in these measurements) can extract $\sin^2\theta_W$

- this measurement is potentially sensitive to the TeV-scale in a complementary fashion to energy-frontier searches!

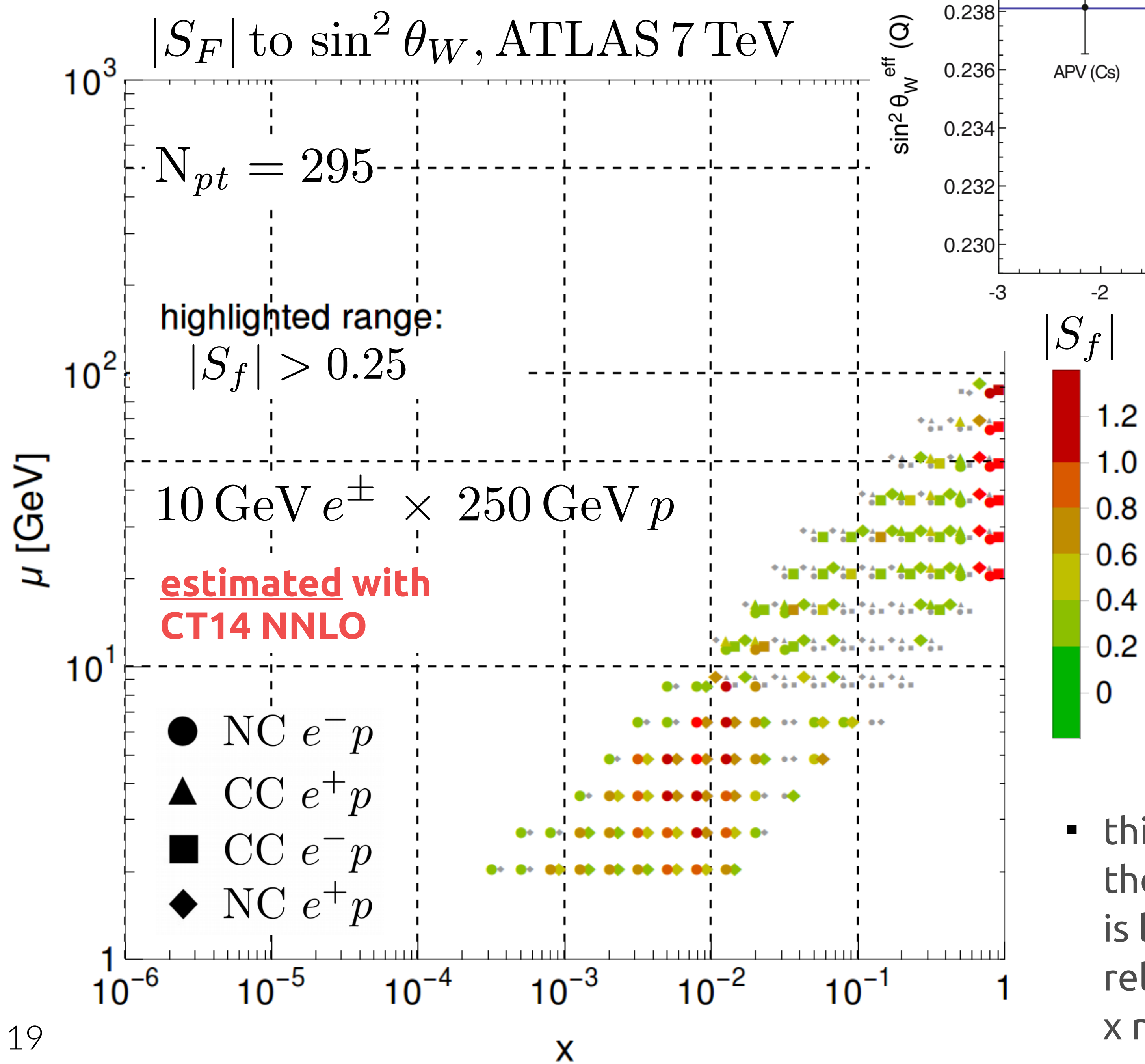
TJH and Melnitchouk, PRD**77**, 114023 (2008).

$$A^{\mathrm{PV}} = -\left(\frac{G_F Q^2}{4\sqrt{2}\pi\alpha} \right) \left(Y_1\, a_1 + Y_3\, a_3 \right)$$

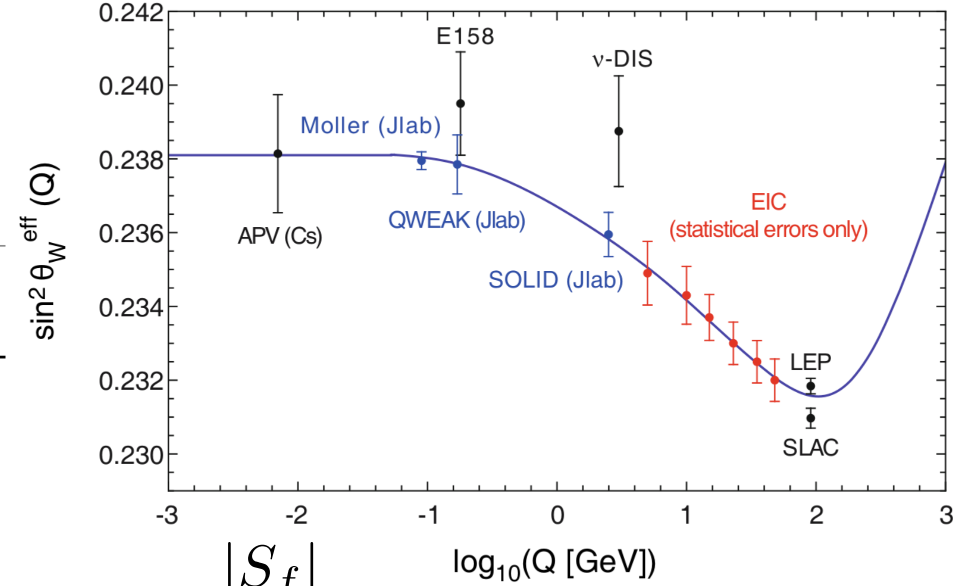
N.B.: extractions are dependent upon knowledge of the PDFs

$$a_1 = \frac{2\sum_q e_q\, C_{1q}\, (q+\bar{q})}{\sum_q e_q^2\, (q+\bar{q})} \qquad a_3 = \frac{2\sum_q e_q\, C_{2q}\, (q-\bar{q})}{\sum_q e_q^2\, (q+\bar{q})}$$

an EIC will probe EW parameters and New Physics!



Accardi et al., EPJA**52**, 268 (2016).



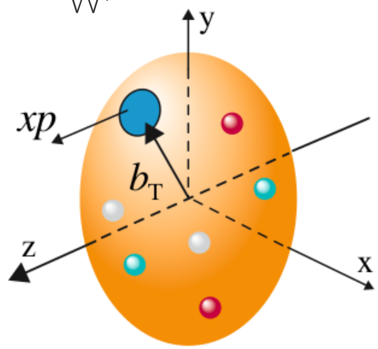
- observe a pronounced sensitivity to the Weinberg angle, especially low and high x, even at $\mathcal{L} = 100\mathrm{fb}^{-1}$

- this corresponds closely to the kinematics at which EIC is likely to measure A^{PV} — relatively large Q² and in the x range $0.2 \lesssim x \lesssim 0.5$

key points... ...and **the future.**

- numerous observables central to LHC's present/future discovery program at limited by uncertainties associated with nucleon structure

  → for the unpolarized PDFs, systematic tensions among modern world data are an impediment to higher precision for σ_H, M_W, ...



  → an EIC will be ideally suited to perform measurements with the ability to unravel such systematic issues

the EIC impact upon high-energy pheno will be pivotal

→ controlling PDFs/SM backgrounds; BSM searches; event generators

- confronting systematic PDF issues and exploring the HEP implications of the EIC require **community efforts**, esp. to optimize the output of the eventual program and its utility to HEP

  many areas on both sides of the medium-, high-energy divide in which input is needed.

THANKS!!

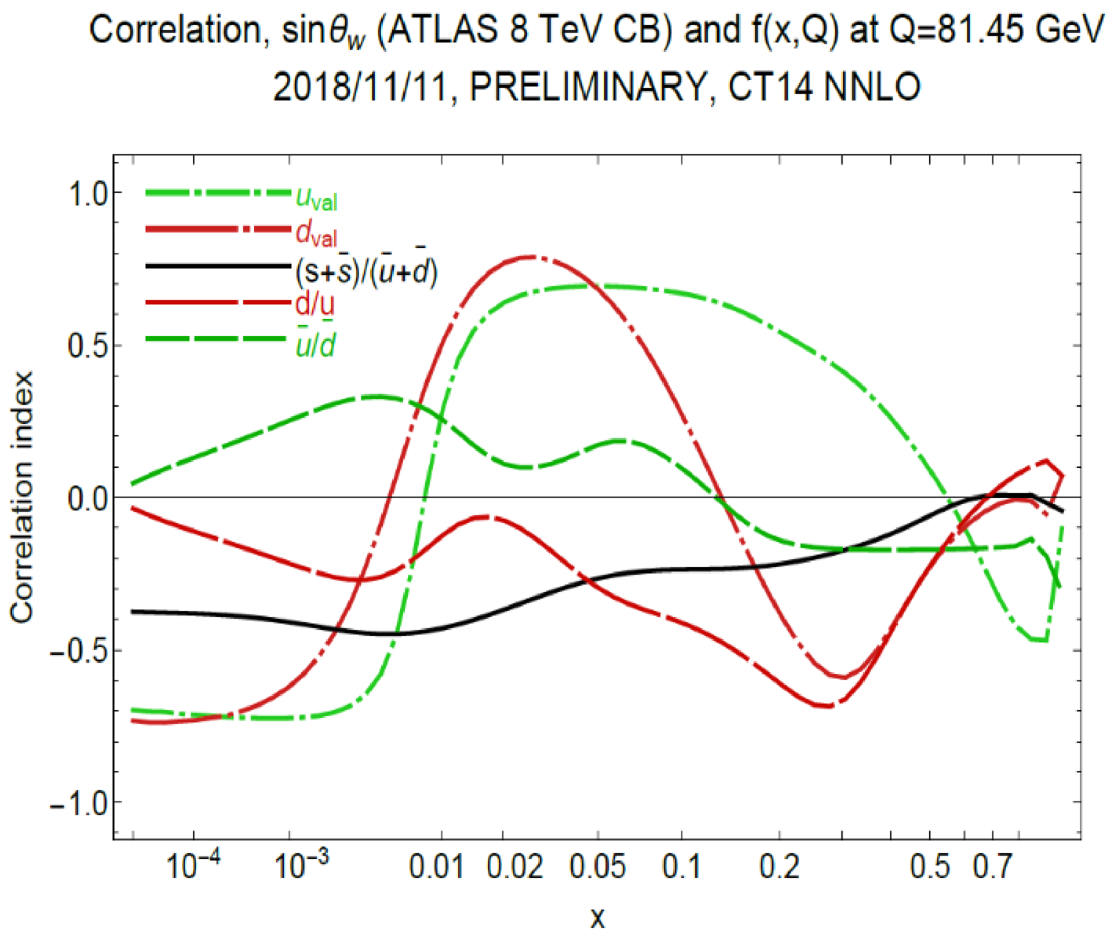
supplementary material

$\sin^2\theta_W$ (and, eventually, M_W)

...as a follow-on to Alesandro's EW-focused overview:

important PDF correlations for the ATLAS extraction of $\sin^2\theta_W$

Example: $\sin^2\theta_{weak} \equiv s2w$ measured by ATLAS 8 TeV



Strongest correlations of s2w with u_{val}, d_{val} at

$$0.005 \lesssim x \lesssim 0.2$$

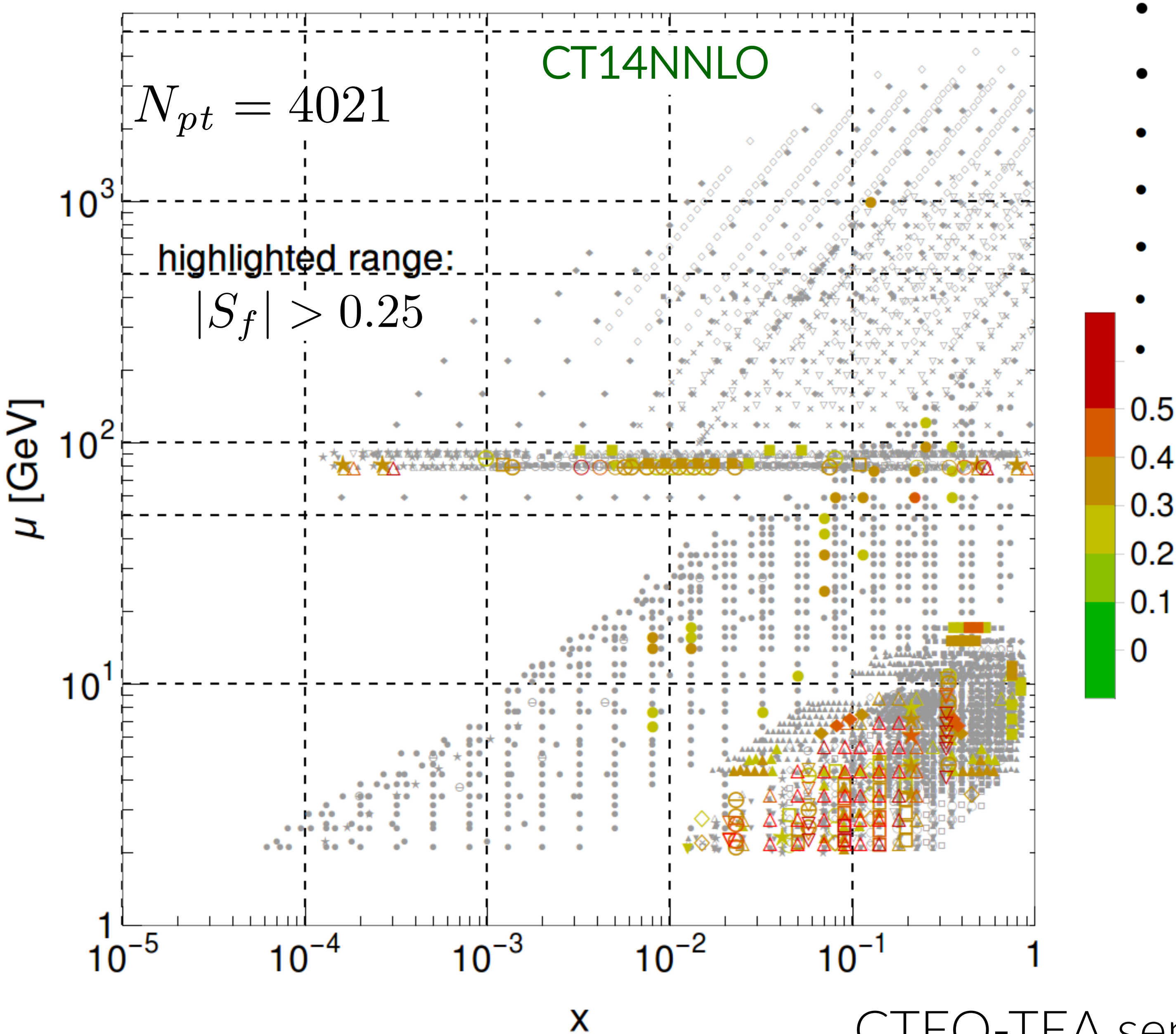
weak correlations with $\bar{u}$, $\bar{d}$, $\bar{s}$, g

u_{val}, d_{val} changed between CT10 and CT14 **[1506.07433, Sec. 2B]**

It is instructive to explore the data pulls on u_{val}, d_{val}

$\sin^2\theta_W$

PDF sensitivity of $\sin^2\theta_W$ from 7 TeV ATLAS data



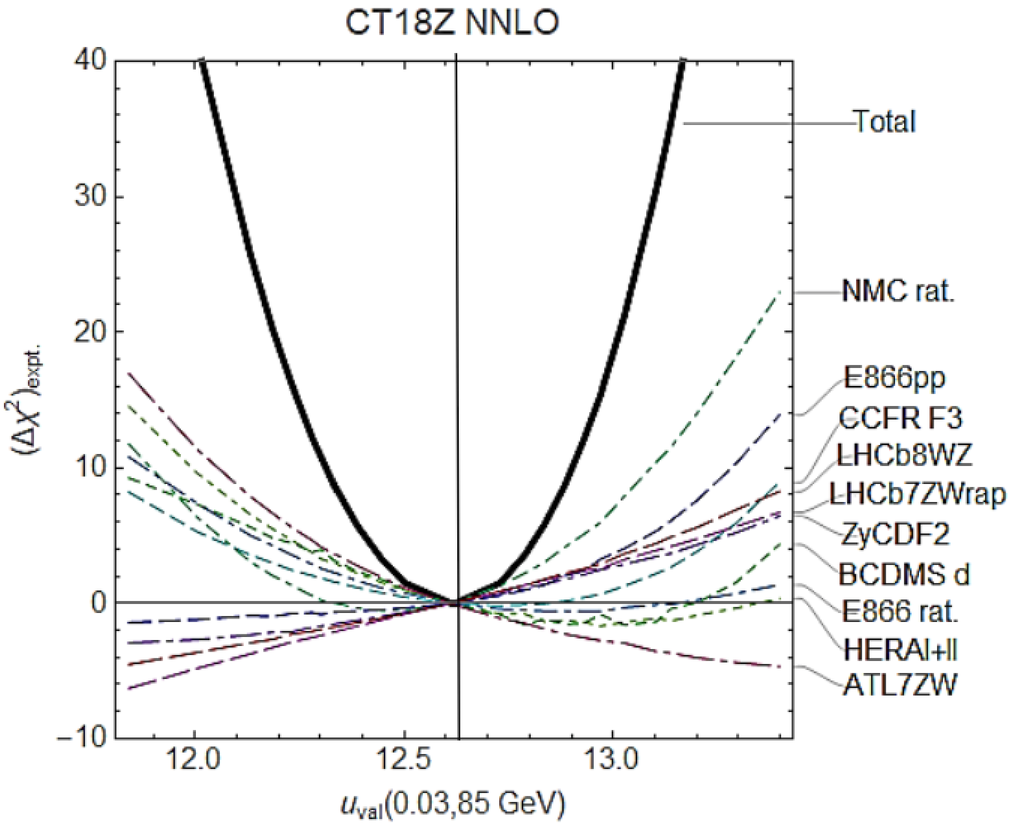
- **combined HERA1 DIS [most sensitive]**
- CCFR νp DIS $F_{3,2}$
- BCDMS $F_2^{p,d}$
- NMC ep, ed DIS
- CDHSW νA DIS
- NuTeV $\nu A \to \mu\mu X$
- CCFR $\nu A \to \mu\mu X$
- E866 $pp \to \ell^+\ell^- X$
- ATLAS 7 TeV W/Z (35 pb^{-1})

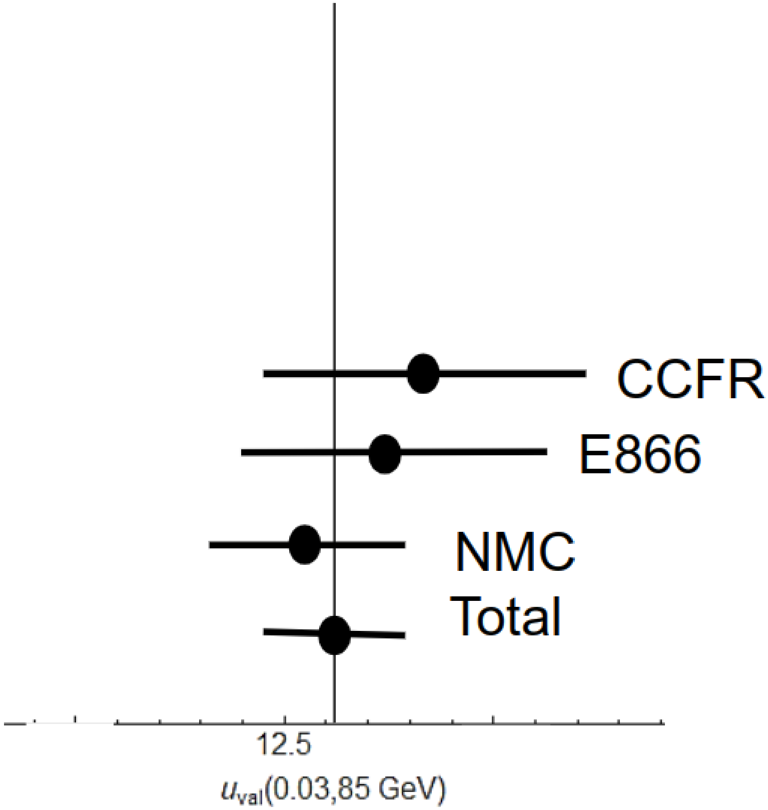
CTEQ-TEA sensitivities to $\sin^2\theta_W$

… using a Lagrange Multiplier scan…

… or using residuals for replicas

[errors and correlations; most replicas are not good fits]



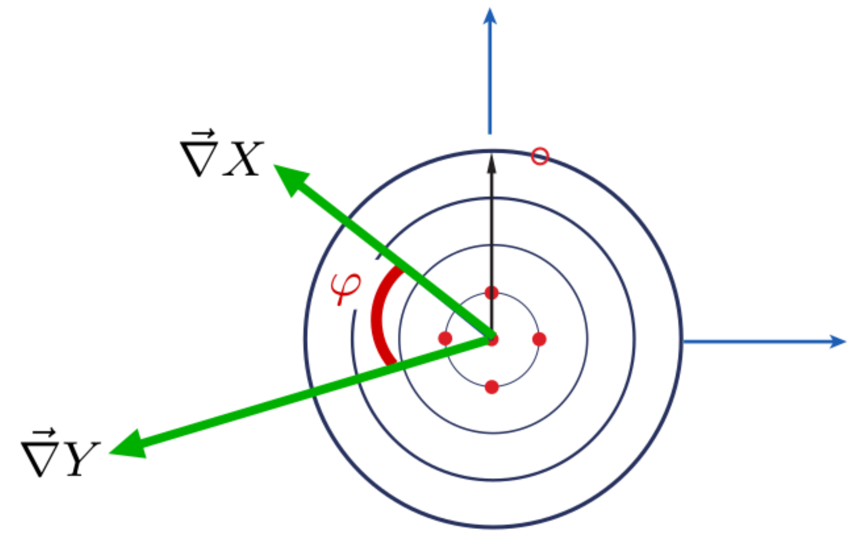


PRELIMINARY

Can repeat for s2w, M_W,…

rather than the costly LM scans, we can examine a “cheaper” measure which yields comparable information

the *L2 sensitivity*



L_2 sensitivity. Take $X = f_a(x_i, Q_i)$ or $\sigma(f)$; $Y = \chi^2_E$ for experiment E. Find $\Delta Y(\vec{z}_{m,X})$ for the displacement $|\vec{z}_{m,X}| = 1$ along the direction $\vec{\nabla}X / \left|\vec{\nabla}X\right|$ (corresponding to $\Delta\chi^2_{tot} = T^2$ and $X(\vec{z}) = X(0) + \Delta X$):

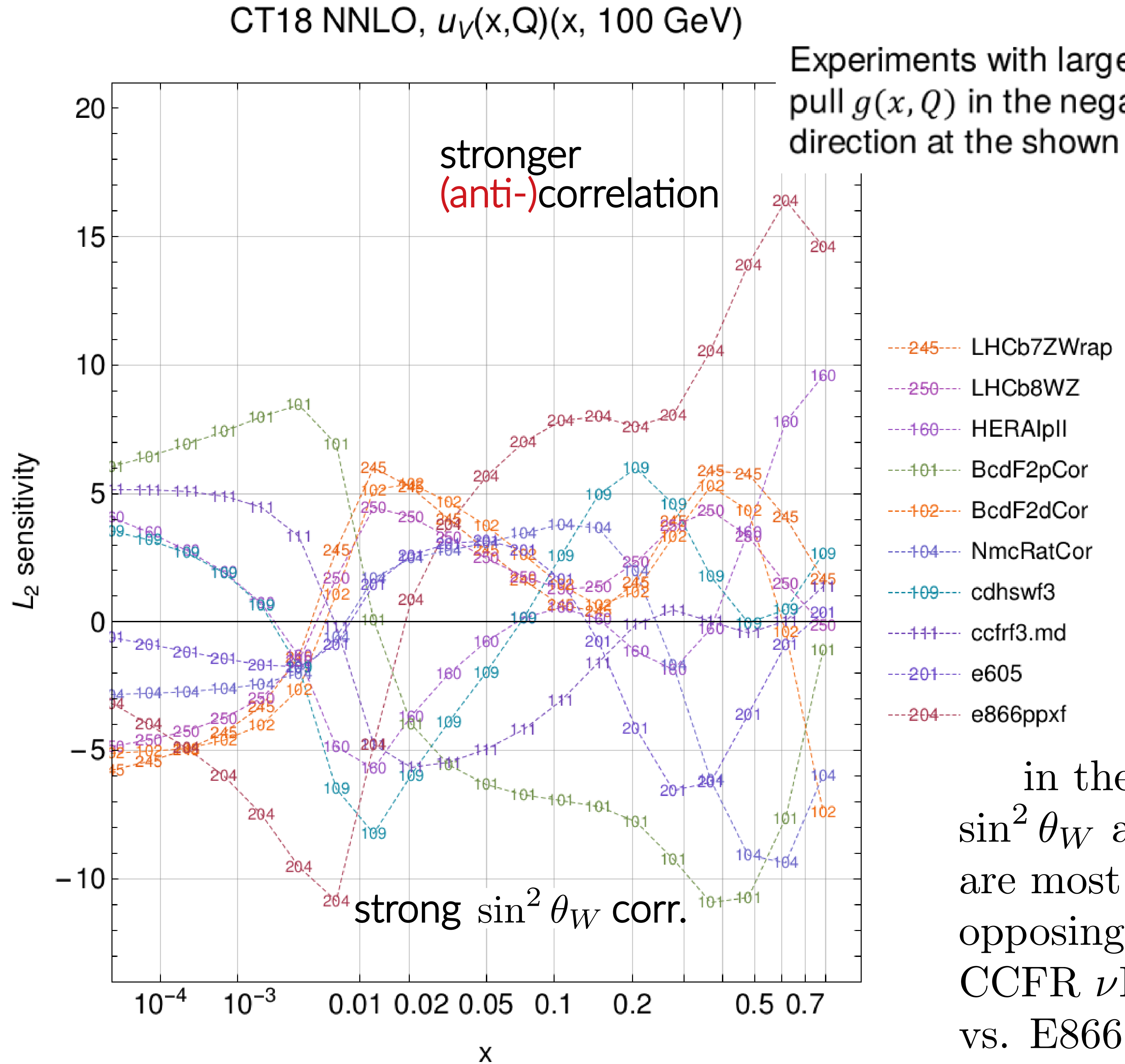
$$S_{f,L_2} \equiv \Delta Y(\vec{z}_{m,X}) = \vec{\nabla}Y \cdot \vec{z}_{m,X} = \vec{\nabla}Y \cdot \frac{\vec{\nabla}X}{|\vec{\nabla}X|}$$

$$= \Delta Y \, \cos\varphi$$

or, $\sim \mathrm{Corr}[f_a, \chi^2_E]$

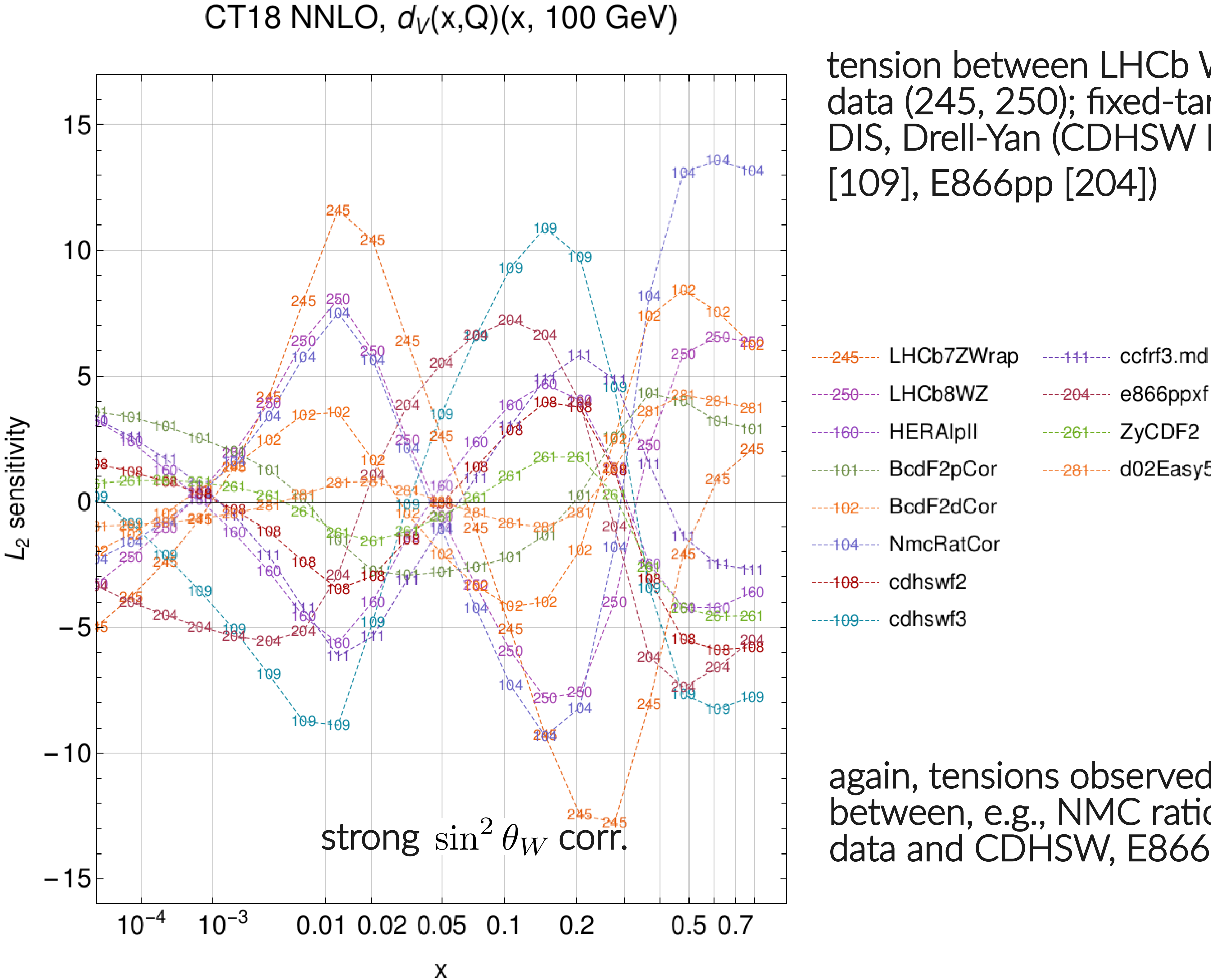
...extent to which total χ^2_E of specific expts. correlates with x-dep. of PDFs

Experiments with large $\Delta\chi^2 > 0$ [$\Delta\chi^2 < 0$] pull $g(x,Q)$ in the negative [positive] direction at the shown x



tension between LHCb W/Z data (245, 250); fixed-target DIS, Drell-Yan (CDHSW F_3 [109], E866pp [204])

in the region where $\sin^2\theta_W$ and u_v are most correlated, opposing pulls from CCFR νDIS, BCDMS vs. E866 pp, NMC rat.



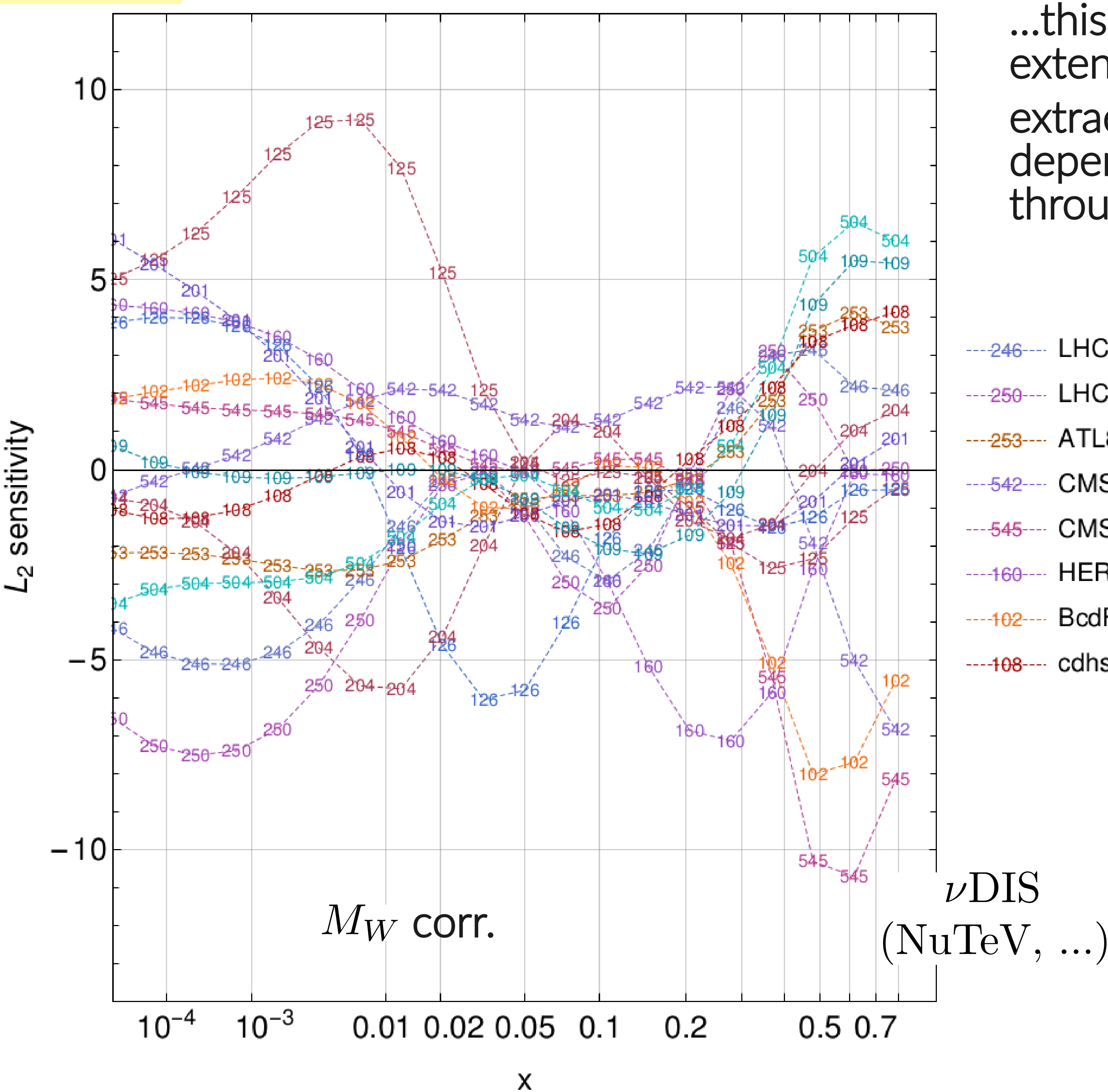
tension between LHCb W/Z data (245, 250); fixed-target DIS, Drell-Yan (CDHSW F_3 [109], E866pp [204])

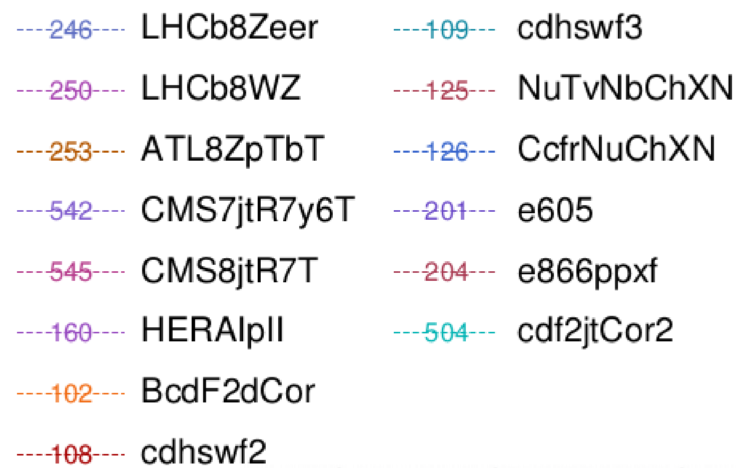
again, tensions observed between, e.g., NMC ratio data and CDHSW, E866pp

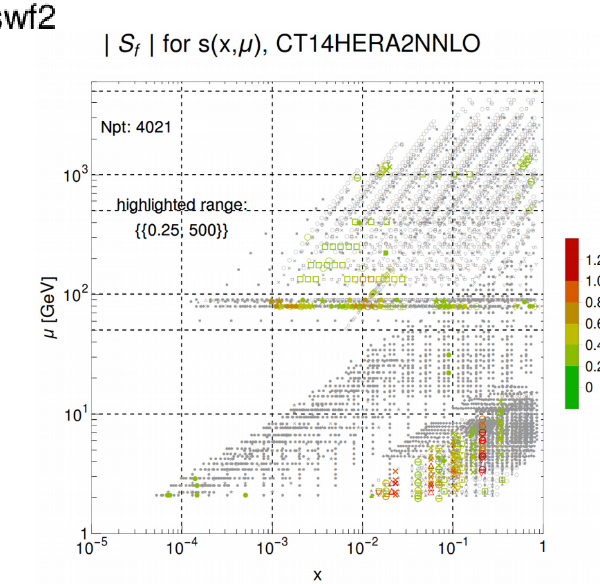
M_W

CT18 NNLO, s(x, 100 GeV)

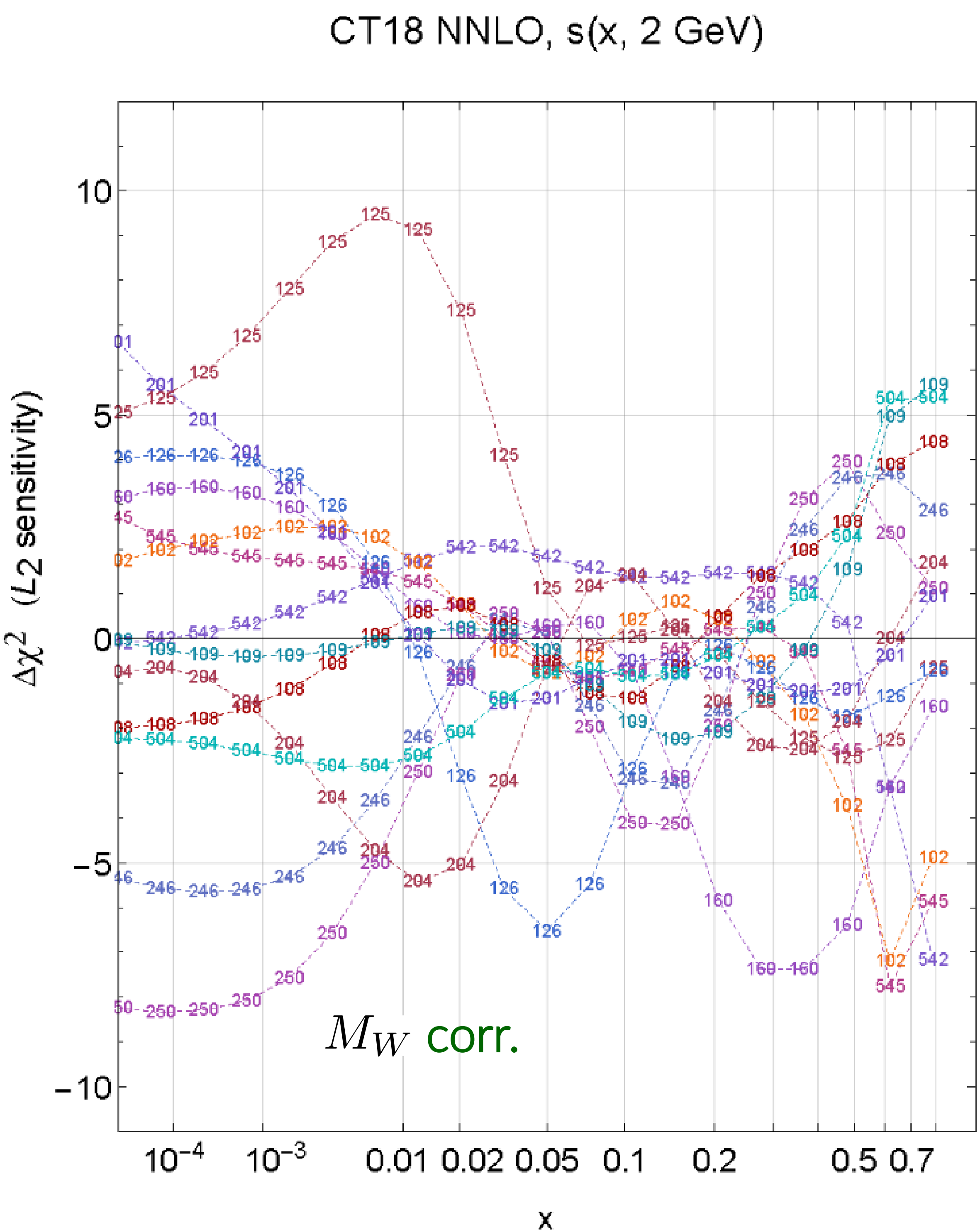
...this analysis can be extended to M_W, extractions of which are dependent upon s(x), through Z-calibration







L_2 sensitivity, strangeness: CT18

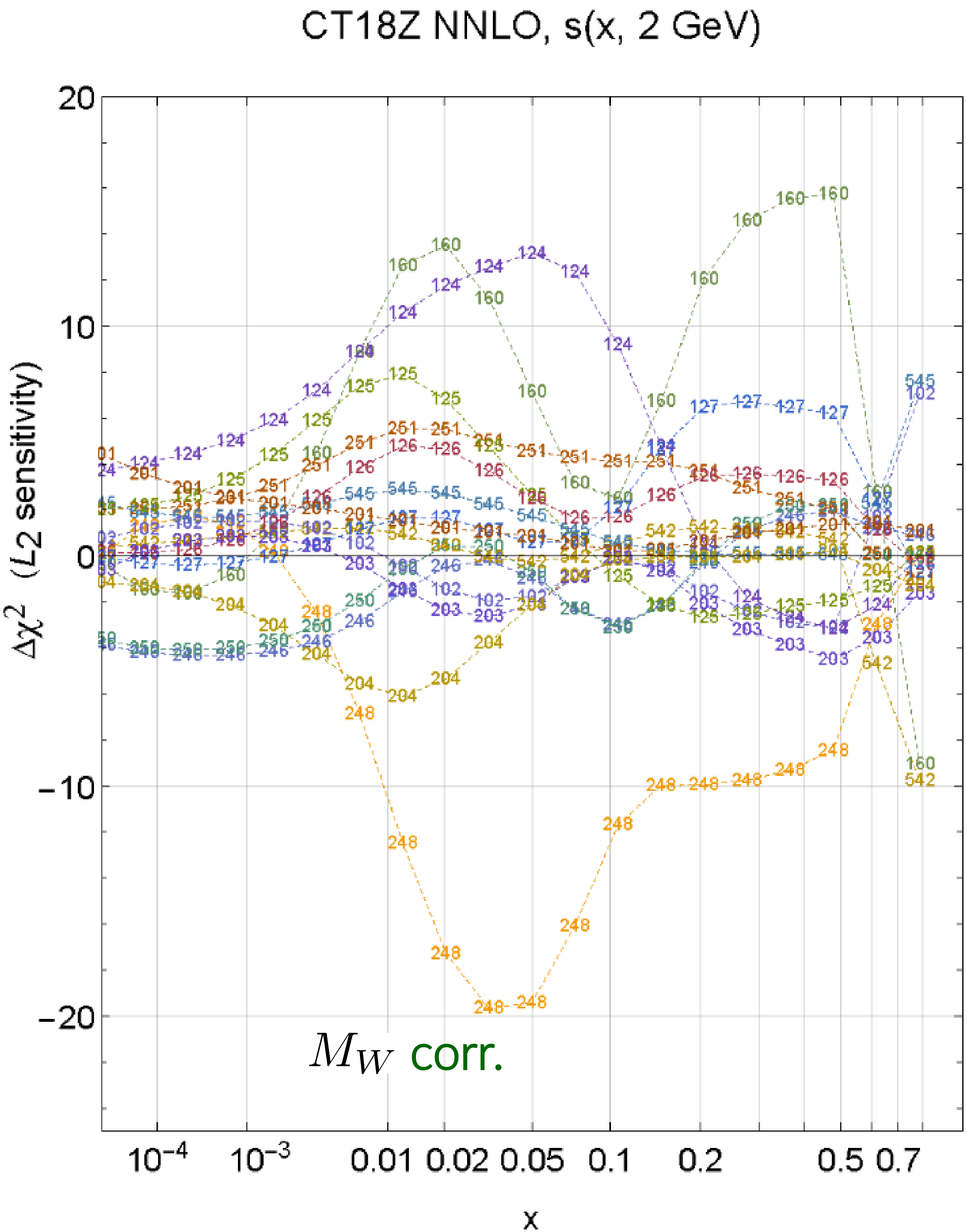


Most sensitive experiments

- 246 LHCb8Zeer
- 250 LHCb8WZ
- 542 CMS7jtR7y6T
- 545 CMS8jtR7T
- 160 HERAIpII
- 102 BcdF2dCor
- 108 cdhswf2
- 109 cdhswf3
- 125 NuTvNbChXN
- 126 CcfrNuChXN
- 201 e605
- 204 e866ppxf
- 504 cdf2jtCor2

A tension trend between DIS (HERA I+II, CCFR, NuTeV) and Drell-Yan (ATLAS 7 Z/W, LHCb W/Z, E866 pp, …) experiments

L_2 sensitivity, strangeness: CT18Z



Most sensitive experiments

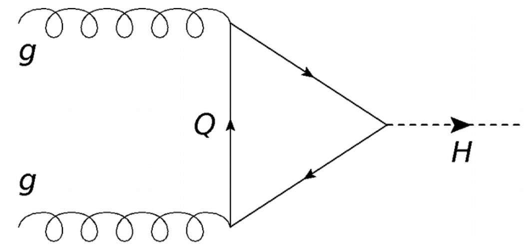
246	LHCb8Zeer	124	NuTvNuChXN
248	ATL7ZW.xF	125	NuTvNbChXN
250	LHCb8WZ	126	CcfrNuChXN
251	ATL8DY	127	CcfrNbChXN
542	CMS7jtR7y6T	201	e605
545	CMS8jtR7T	203	e866f
160	HERAIpII	204	e866ppxf
102	BcdF2dCor		

A tension trend between DIS (HERA I+II, CCFR, NuTeV) and Drell-Yan (ATLAS 7 Z/W, LHCb W/Z, E866 pp, ...) experiments

pronounced effect of ATLAS 7 TeV Z/W data!

QCD at high energies: an EIC and control over the gluon

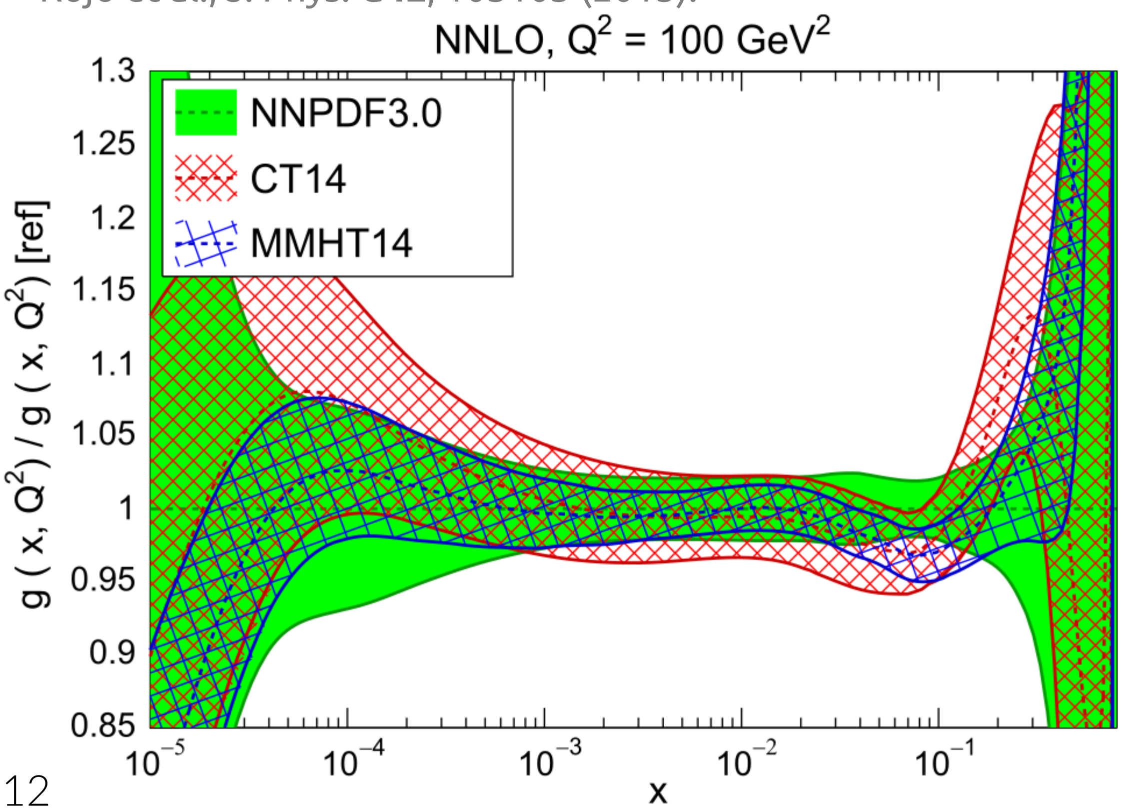
- the gluon is crucial to the mass of hadronic bound states, and gg → H is the dominant channel in Higgs production



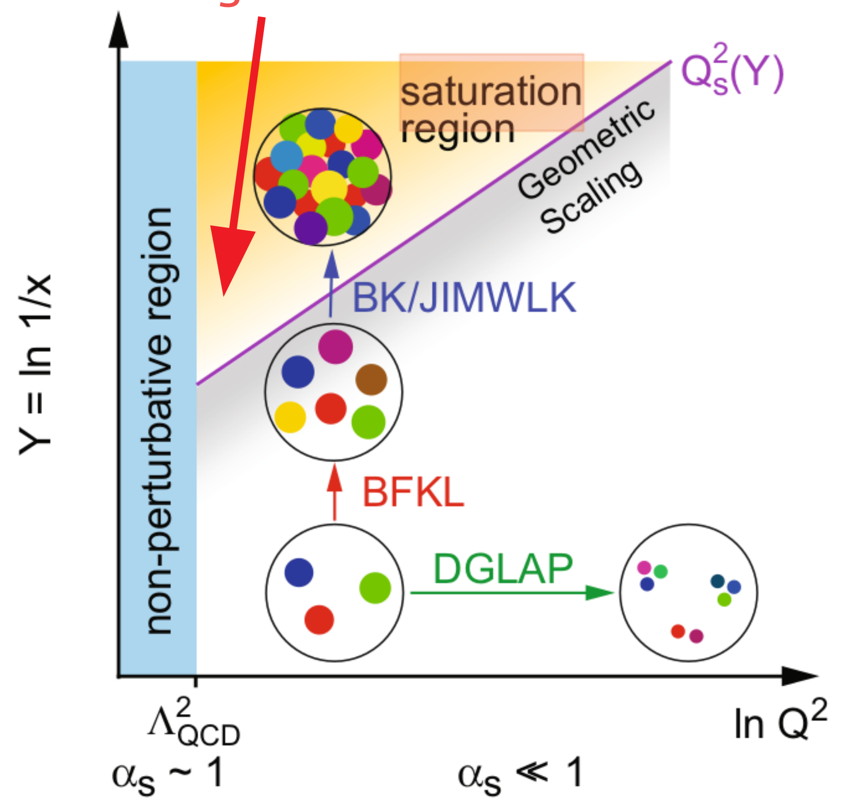
BUT

- while under better control at intermediate x, the collinear gluon PDF is poorly known toward the distribution endpoints, *i.e.*, $g(x, \mu)$ for $x \to 0,\ 1$

Rojo et al., J. Phys. G**42**, 103103 (2015).



can we begin to observe this transition?



the goal is to quantify the strength of the constraints placed on a particular set of PDFs by both individual and aggregated measurements *without direct fitting*

- for single-particle hadroproduction of gauge bosons at, e.g., LHC, factorization gives

$$\sigma(AB \to W/Z+X) \; = \; \sum_n \alpha_s^n(\mu_R^2) \sum_{a,b} \int dx_a dx_b$$

$$\times \, f_{a/A}(x_a, \mu^2) \, \hat{\sigma}^{(n)}_{ab\to W/Z+X} \left(\hat{s},\, \mu^2, \mu_R^2\right) f_{b/B}(x_b, \mu^2)$$

PDFs determined by fits to data; e.g., "CT14H2"

pQCD matrix elements – specified by theoretical formalism in a given fit

- *idea* : study the statistical <u>correlation</u> between PDFs and the quality of the fit at a measured data point(s); fit quality encoded in a (Theory) – (shifted Data) *residual* :

$$r_i(\vec{a}) = \frac{1}{s_i}\left(T_i(\vec{a}) - D_{i,sh}(\vec{a})\right)$$

s_i : uncorrelated uncert.

$\vec{a}$: PDF parameters

a brief statistical aside, i

- **the CTEQ-TEA global analysis relies on the Hessian formalism for its error treatment**

$$\chi^2_E(\vec{a}) = \sum_{i=1}^{N_{pt}} r_i^2(\vec{a}) + \sum_{\alpha=1}^{N_\lambda} \overline{\lambda}_\alpha^2(\vec{a})$$

nuisance parameters to handle correlated errors

$$r_i(\vec{a}) = \frac{1}{s_i}\left(T_i(\vec{a}) - D_{i,sh}(\vec{a})\right)$$

these result in systematic shifts to data central values:

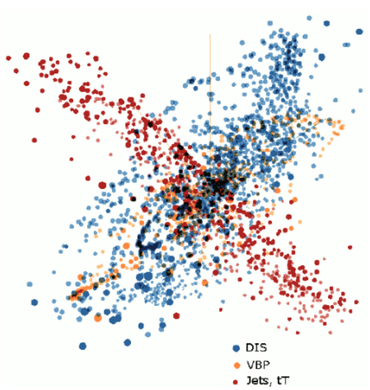
$$D_i \rightarrow D_{i,sh}(\vec{a}) = D_i - \sum_{\alpha=1}^{N_\lambda} \beta_{i\alpha}\overline{\lambda}_\alpha(\vec{a})$$

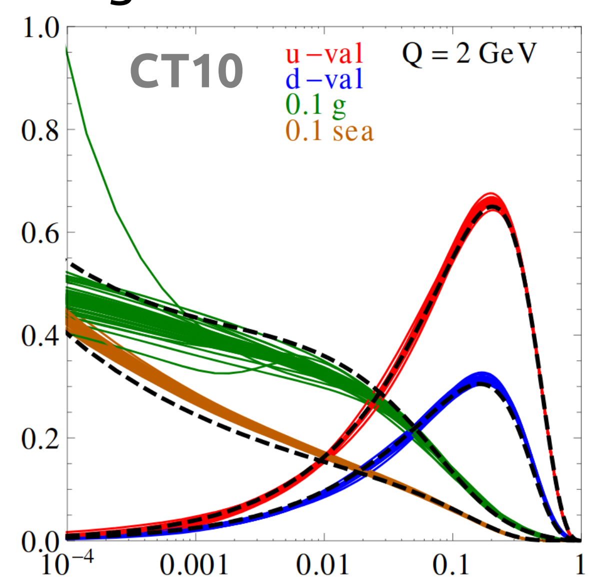
- a 56-dimensional parametric basis $\vec{a}$ is obtained by diagonalizing the Hessian matrix H determined from χ^2 (following a 28-parameter fit)

use this basis to compute 56-component "normalized" residuals :

$$\delta^{\pm}_{i,l} \equiv \left(r_i(\vec{a}^{\pm}_l) - r_i(\vec{a}_0)\right) / \langle r_0 \rangle_E$$

where $\langle r_0 \rangle_E \equiv \sqrt{\frac{1}{N_{pt}} \sum_{i=1}^{N_{pt}} r_i^2(\vec{a}_0)}$



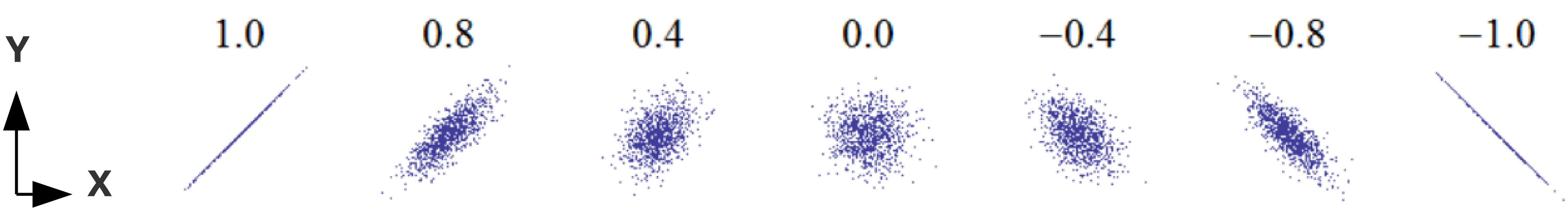


a brief statistical aside, ii

- **... but how does the behavior of these residuals relate to the fitted PDFs and their uncertainties?**

for example, how does the PDF uncertainty (at specific x, μ) correlate with the residual associated with a theoretical prediction at the same x, μ?

examine the Pearson correlation over the 56-member PDF error set between a PDF of given flavor and the residual



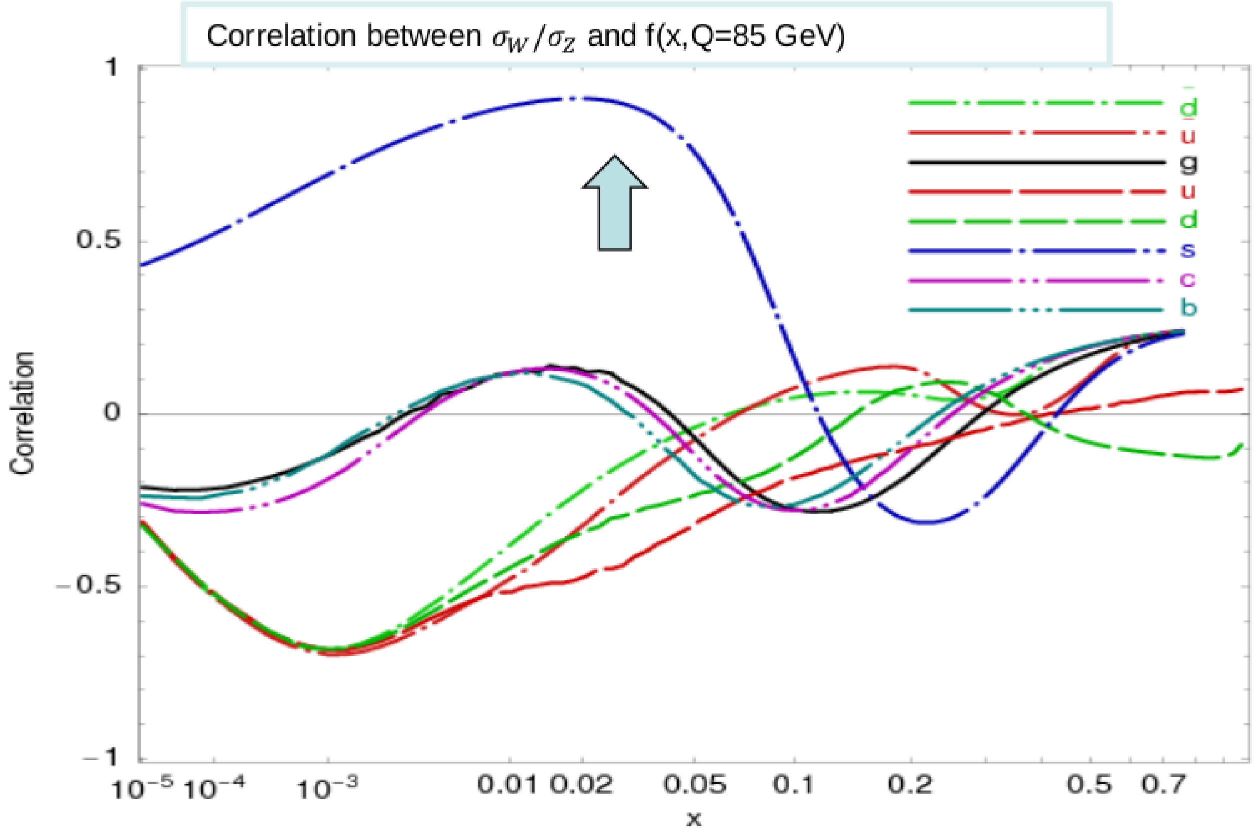
[X,Y] are exactly (anti-)correlated at the far (right) left above.

- we may then evaluate correlations between arbitrary PDF-derived quantities over the ensemble of error sets ([X,Y] may be PDFs, cross sections, residuals,...):

$$\mathrm{Corr}[X,Y] = \frac{1}{4\Delta X \Delta Y} \sum_{j=1}^{N} (X_j^+ - X_j^-)(Y_j^+ - Y_j^-) \qquad \Delta X = \frac{1}{2}\sqrt{\sum_{j=1}^{N} (X_j^+ - X_j^-)^2}$$

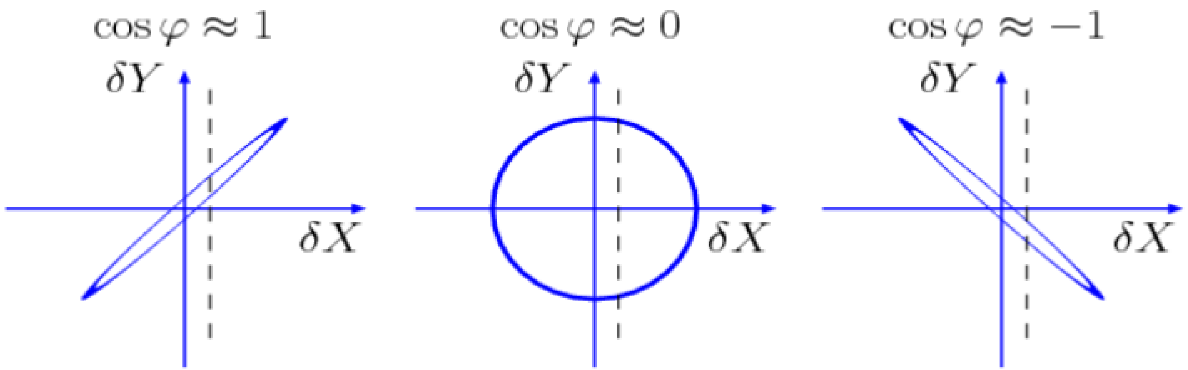
...we may turn to the Pearson correlations between PDFs and δ_i, but we first note

Correlations carry useful, but limited information



CTEQ6.6 [arXiv:0802.0007]: $\cos\varphi > 0.7$ shows that the ratio σ_W/σ_Z at the LHC must be sensitive to the strange PDF $s(x,Q)$

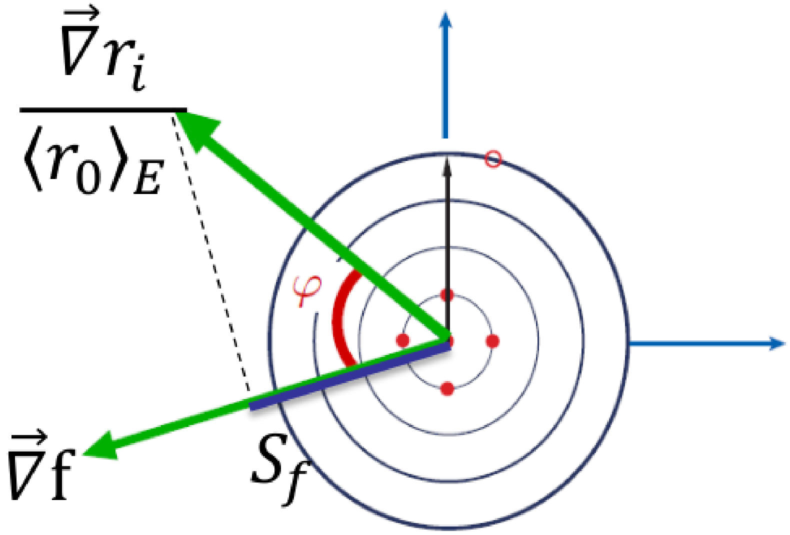
$\cos\varphi \approx \pm 1$ suggests that a measurement of X **may** impose tight constraints on Y



But, Corr[X,Y] between **theory** cross sections X and Y does not tell us about **experimental** uncertainties

Correlation C_f and sensitivity S_f

The relation of data point i on the PDF dependence of f can be estimated by:



- $C_f \equiv \text{Corr}[\rho_i(\vec{a})), f(\vec{a})] = cos\varphi$

$\vec{\rho}_i \equiv \vec{\nabla} r_i / \langle r_0 \rangle_E$ -- gradient of r_i normalized to the r.m.s. average residual in expt E;

$$\left(\vec{\nabla} r_i\right)_k = \left(r_i(\vec{a}_k^+) - r_i(\vec{a}_k^-)\right)/2$$

$$\text{Corr}[X, Y] = \frac{1}{4\Delta X \Delta Y} \sum_{j=1}^{N} (X_j^+ - X_j^-)(Y_j^+ - Y_j^-)$$

C_f is **independent** of the experimental and PDF uncertainties. In the figures, take $|C_f| \gtrsim 0.7$ to indicate a large correlation.

- $S_f \equiv |\vec{\rho}_i| cos\varphi = C_f \frac{\Delta r_i}{\langle r_0 \rangle_E}$ -- projection of $\vec{\rho}_i(\vec{a})$ on $\vec{\nabla} f$

S_f is proportional to $\cos\varphi$ and the ratio of the PDF uncertainty to the experimental uncertainty. We can sum $|S_f|$.
In the figures, take $|S_f| > 0.25$ to be significant.

2nd aside: **kinematical matchings**

- residual-PDF correlations and sensitivities are evaluated at parton-level kinematics determined according to leading-order matchings with physical scales in measurements

x_i : parton mom. fraction

μ_i : factorization scale

deeply-inelastic scattering:

$$\mu_i \approx Q\big|_i\,,\;\; x_i \approx x_B\big|_i$$

hadron-hadron collisions:

$AB \to CX$

$$\mu_i \approx Q\big|_i\,,\;\; x_i^{\pm} \approx \frac{Q}{\sqrt{s}}\,\exp(\pm y_C)\bigg|_i$$

single-inclusive jet production: $Q = 2p_{Tj}\,,\;\; y_C = y_j$

$t\bar{t}$ pair production: $Q = m_{t\bar{t}},\;\; y_C = y_{t\bar{t}}$

etc...

$d\sigma/dp_T^Z$ measurements: $Q = \sqrt{(p_T^Z)^2 + (M_Z)^2},\;\; y_C = y_Z$

Sensitivity ranking tables

… to assess the impact of separate experiments

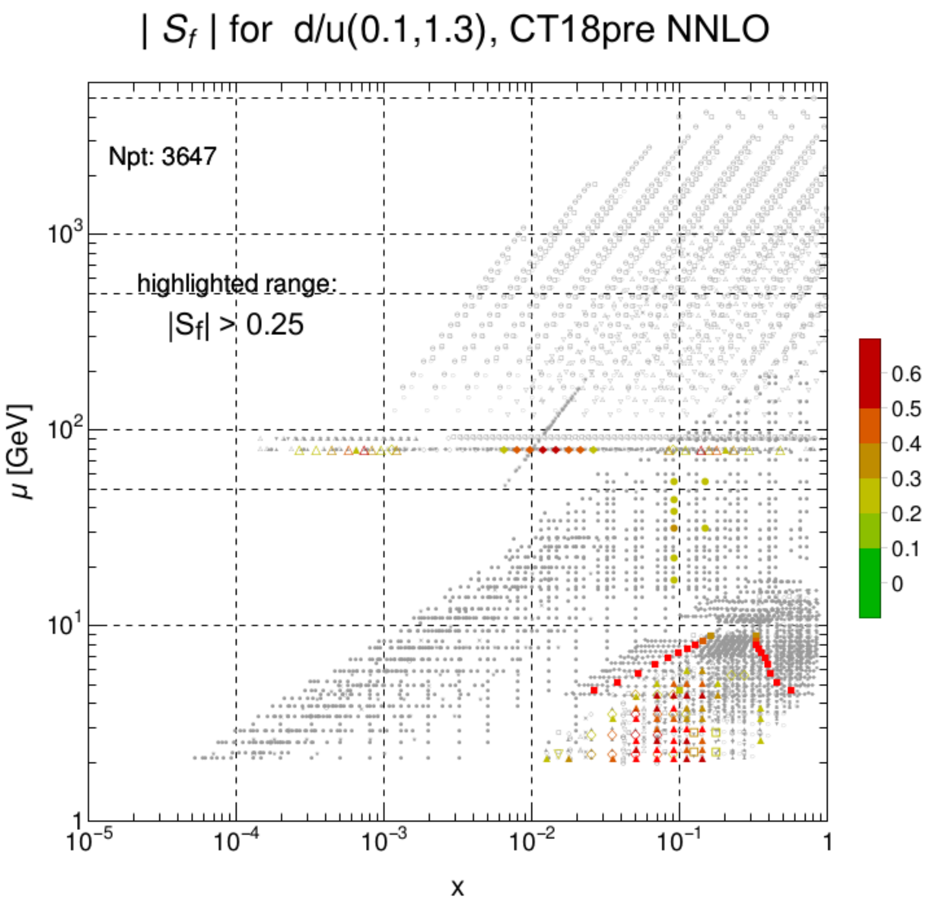
No.	Expt.	N_{pt}	$\sum_f \lvert S_f^E \rvert$	$\langle \sum_f \lvert S_f^E \rvert \rangle$	$\lvert S_{\bar{d}}^E \rvert$	$\langle \lvert S_{\bar{d}}^E \rvert \rangle$	$\lvert S_{\bar{u}}^E \rvert$	$\langle \lvert S_{\bar{u}}^E \rvert \rangle$	$\lvert S_g^E \rvert$	$\langle \lvert S_g^E \rvert \rangle$	$\lvert S_u^E \rvert$	$\langle \lvert S_u^E \rvert \rangle$	$\lvert S_d^E \rvert$	$\langle \lvert S_d^E \rvert \rangle$	$\lvert S_s^E \rvert$	$\langle \lvert S_s^E \rvert \rangle$
			Rankings, **CT14 HERA2 NNLO PDFs**													
1	HERAI+II'15	1120.	620.	0.0922	B		**A**	3	**A**	3	**A**	3	B		C	
2	CCFR-F3'97	86	218.	0.423	C	**1**	C	**1**		3	B	**1**	C	2		
3	BCDMSp'89	337	184.	0.0908			C		C		B	3	C			
4	NMCrat'97	123	169.	0.229	C	2					C	2	B	2		
5	BCDMSd'90	250	141.	0.0939	C				C	3	C	3	C	3		
6	CDHSW-F3'91	96	115.	0.199	C	2	C	2		3	C	2	C	3		
7	E605'91	119	113.	0.158	C	2	C	2				3				
8	E866pp'03	184	103.	0.0935		3	C	3			C	3				
9	CCFR-F2'01	69	89.1	0.215		3		3	C	2		3		2		3
10	**CMS8jets'17**	185	87.6	0.0789					C	3						
11	CDHSW-F2'91	85	82.4	0.162		3		3		3		3	C	3		
12	CMS7jets'13	133	63.8	0.0799					C	3						
13	NuTeV-nu'06	38	58.9	0.259		3		3				3		3	C	**1**
14	**CMS7jets'14**	158	57.5	0.0606					C	3						
15	CCFR SI nub'01	38	49.4	0.217		3		3				3		3	C	**1**
16	**ATLAS7jets'15**	140	48.2	0.0574						3						
17	CCFR SI nu'01	40	48.	0.2		3		3				3		3	C	**1**

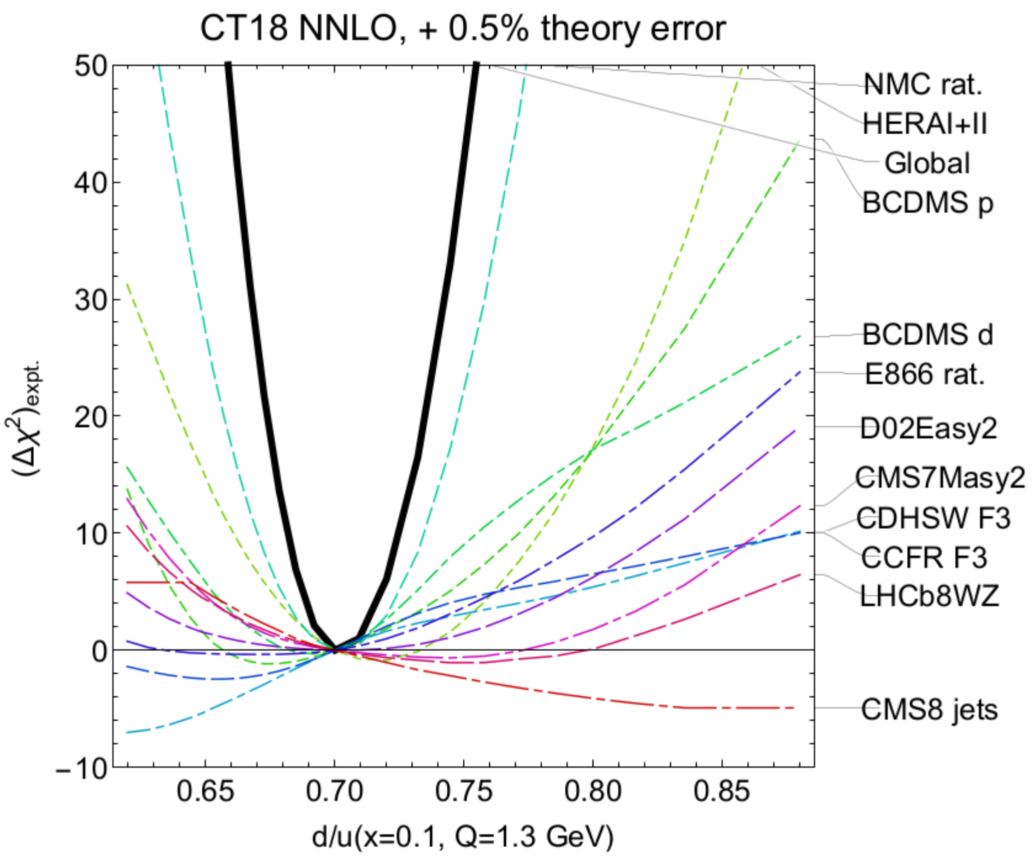
Experiments are listed in the descending order of the summed sensitivities to $\bar{d}, \bar{u}, g, u, d, s$

For each flavor, A and 1 indicate the strongest total sensitivity and strongest sensitivity per point

C and 3 indicate marginal sensitivities; low sensitivities are not shown

PDFSense predictions can be validated against actual fits





- PDFSense successfully predicts the highest impact data sets *before* fitting, as shown in this illustration for the large x PDF ratio d/u
- Lagrange Multiplier scans provide an independent test of which datasets most drive the global fit in connection with specific PDFs

HERA and fixed-target (BCDMS, NMC) data are dominant!