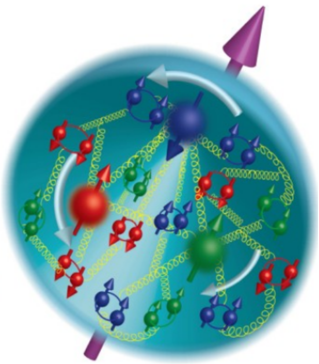
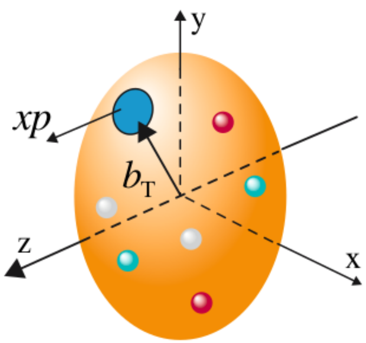
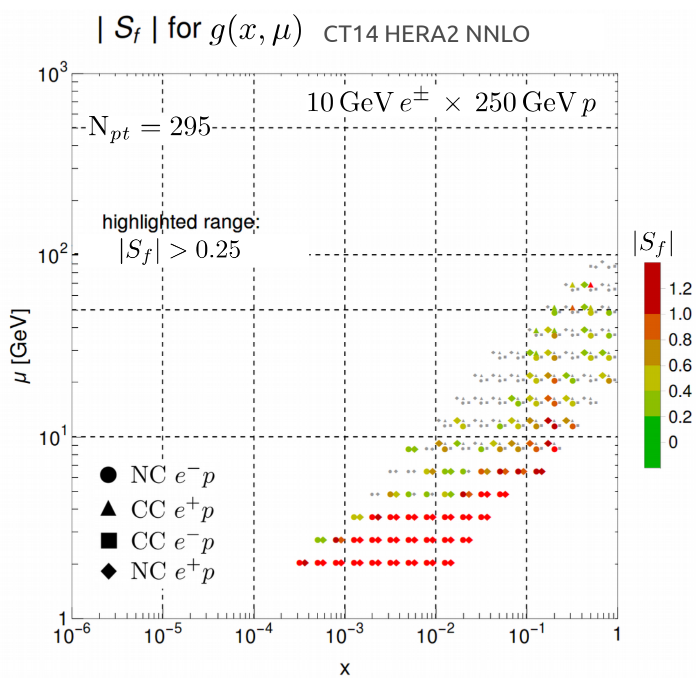
# First-round studies of the EIC’s PDF implications

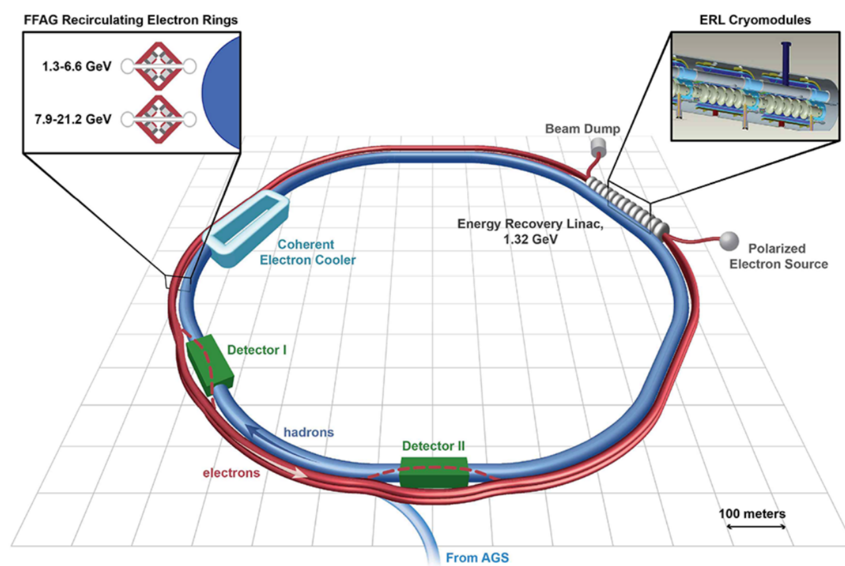
Tim Hobbs, JLab EIC Center & CTEQ@SMU

March 20th 2020






CTEQ

EIC² EIC Center at Jefferson Lab

# Unraveling PDFs' flavor dependence is challenging; multiple EIC channels/processes needed

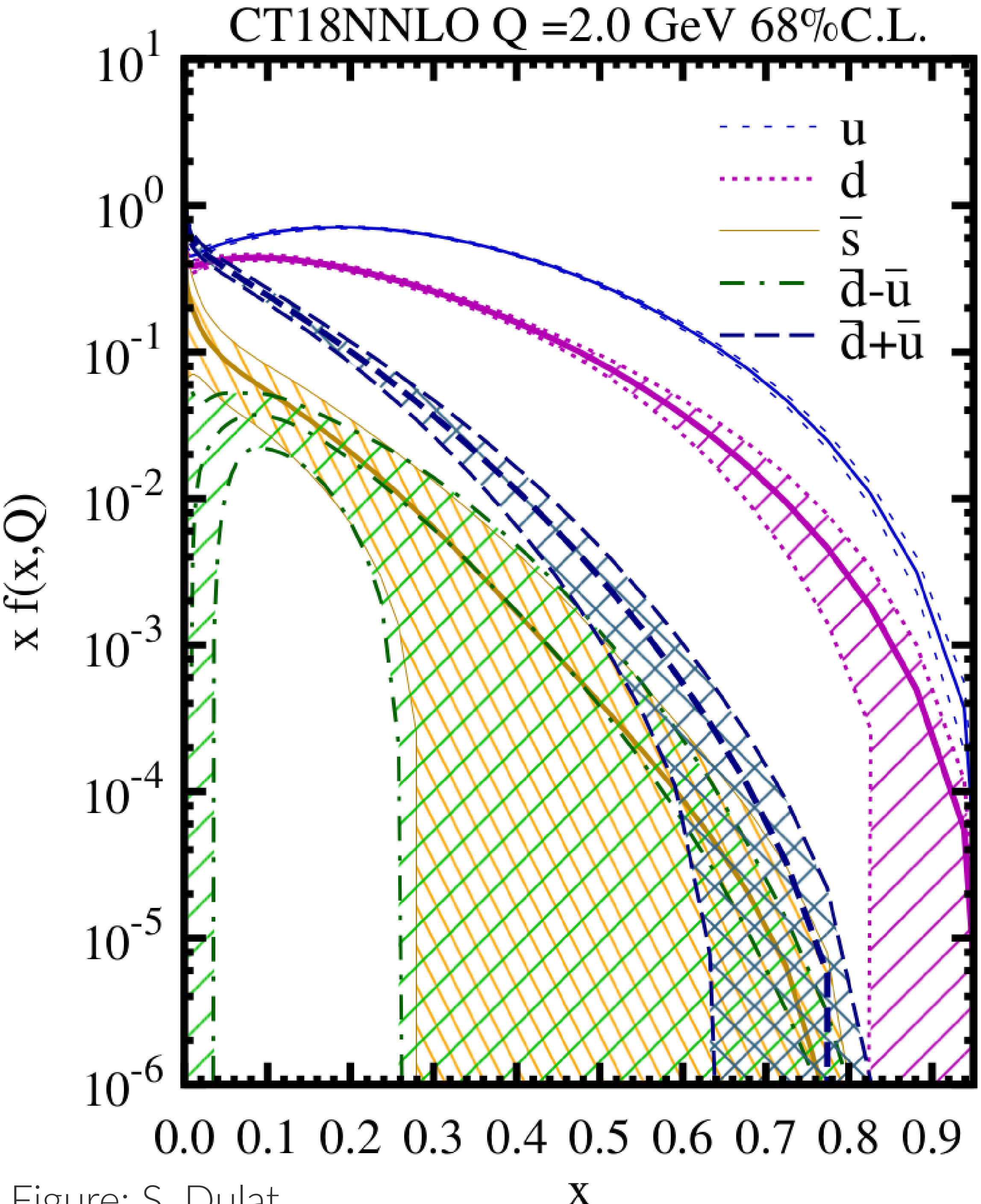



Figure: S. Dulat

**note PDFs' different orders-of-mag.!**

NC DIS: sensitivity to $d$-type quarks ¼ that of $u$-type

$$\sigma \propto \frac{4}{9}(u_+ + c_+) + \frac{1}{9}(d_+ + s_+ + b_+)$$

CC DIS: lower accuracy (1/10 lumi.)

high $x$ (>0.1) inclusive DIS

→ $u$-quark dominates

→ $d$-quark ½ of $u$, but harder to access in NC DIS (above)

→ $\bar{d} + \bar{u} \;\sim$ few percent of $u$

…1% error on $u$ → 50-100% error on $\bar{d} + \bar{u}$

→ for $x$~0.1,

$$s \approx \bar{s} \approx \bar{d} - \bar{u} < 0.1(\bar{d} + \bar{u})$$

→ at $x$>0.5, no separation for $\bar{u}, \bar{d}, \bar{s}$

# we have a dedicated effort to explore PDF impacts of the EIC

Collaborators and consultants

| | |
|---|---|
| Pavel Nadolsky<br>Fred Olness<br>Bo-Ting Wang | Southern Methodist |
| Sayipjamal Dulat | Xinjiang Univ. |
| C.-P. Yuan | Michigan State |
| Alberto Accardi<br>Yulia Furletova | Jefferson Lab |

- the goal: use recently-developed tools for PDF global analyses to examine the PDF pulls of EIC pseudodata

1803.02777
1806.07950
1904.00022
1907.00988
2001.07862

tools and EIC apps

- needed for the Yellow Report Initiative: quick, unambiguous PDF impact metrics

→ in turn, these can be incorporated into the YR workflow:

iteratively, **machine design → simulation → physics**

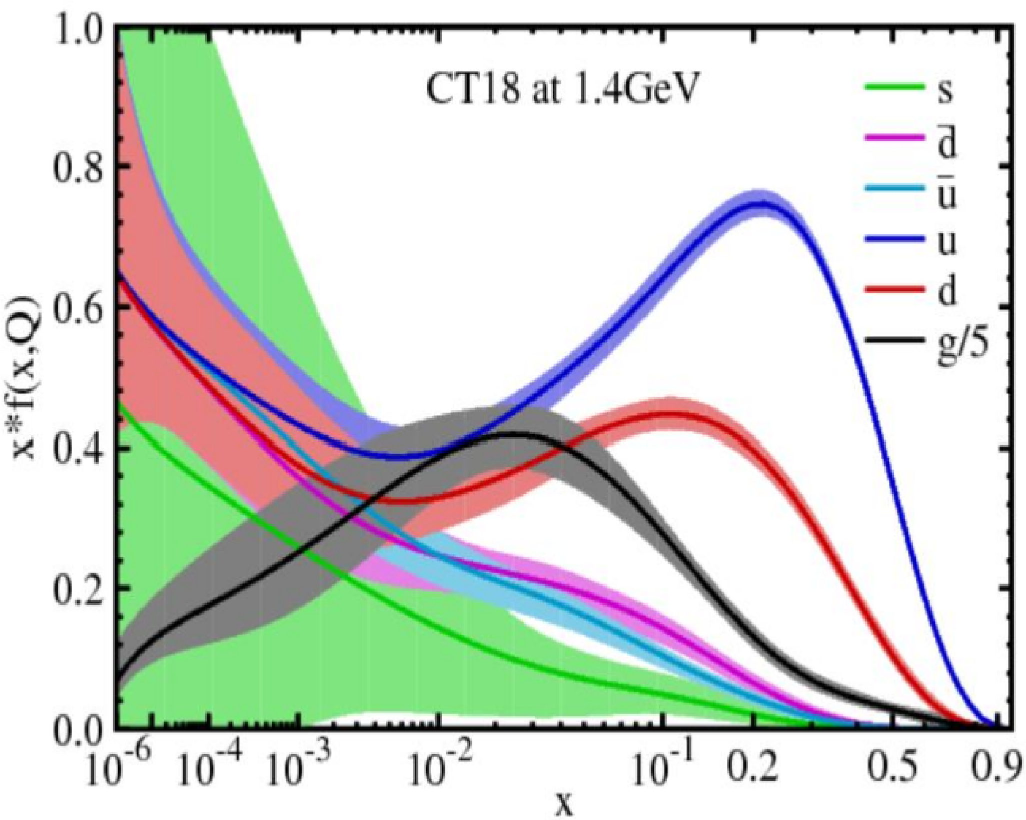
→ speed, simplicity, clarity can help ensure robust and timely convergence

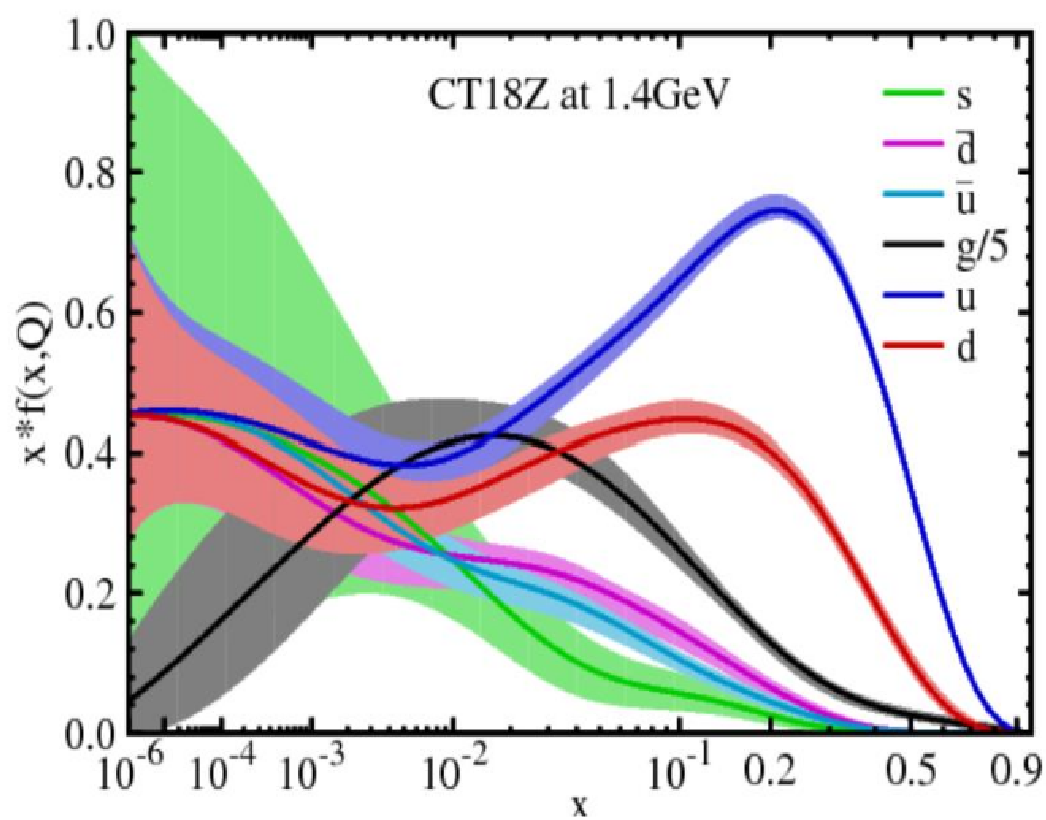
C T E Q

# CT18 parton distributions

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

CT18 main analysis → arxiv:1912.10053.




- have developed fast methods to guide global fits toward highest-impact data & calculations

**[here, Hessian profiling with (I) ePump and (II) PDFSense]**

→ can be quickly applied to EIC pseudodata to supplement fitting

# high-energy EIC pseudodata

- reach in center-of-mass energy, $20 \leq \sqrt{s} \leq 140\,\mathrm{GeV}$

→ luminosities 2-3 decades greater than at HERA

→ á la HERA, the combination of precision & kinematic coverage provide constraining 'lever arm' on QCD evolution

→ QCD evolution: (**high $x$, low Q**) ↔ (**low $x$, high Q**)

- as a generic scenario, we consider here the simulated impact of a machine with: $10\,\mathrm{GeV}\; e^{\pm}$ on $250\,\mathrm{GeV}\; p \quad (\sqrt{s} = 100\,\mathrm{GeV})$

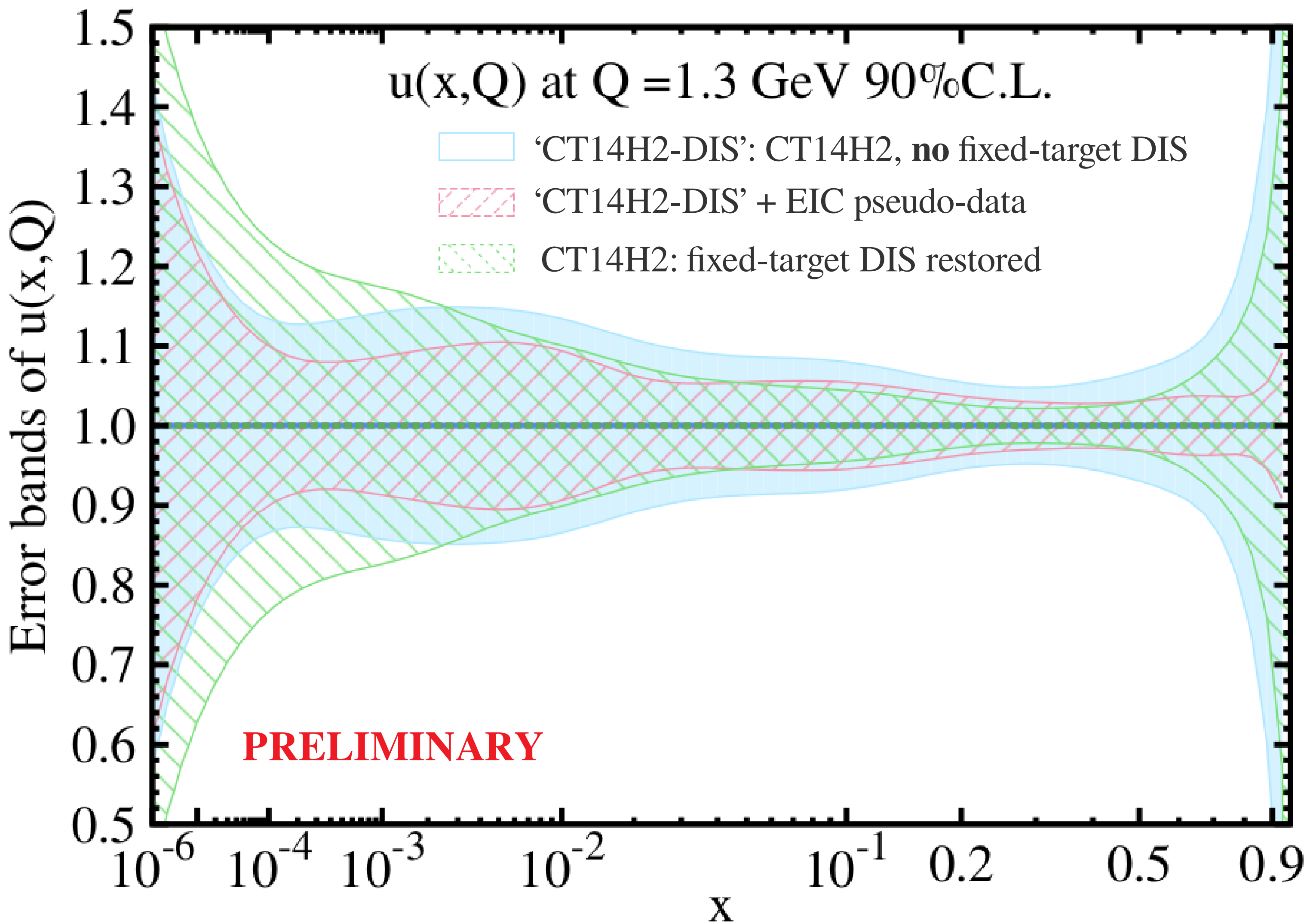
~year of data-taking

$$\left\{ \begin{array}{l} \mathcal{L} = 100\,\mathrm{fb}^{-1}\; e^{-}\ \text{pseudodata} \\ \mathcal{L} = 10\,\mathrm{fb}^{-1}\; e^{+}\ \text{pseudodata} \end{array} \right. \rightarrow \text{NC/CC}$$

→ generated based on CT14 HERA2 NNLO PDF fit

# I) Hessian profiling [`ePump`] for EIC impacts on PDF errors

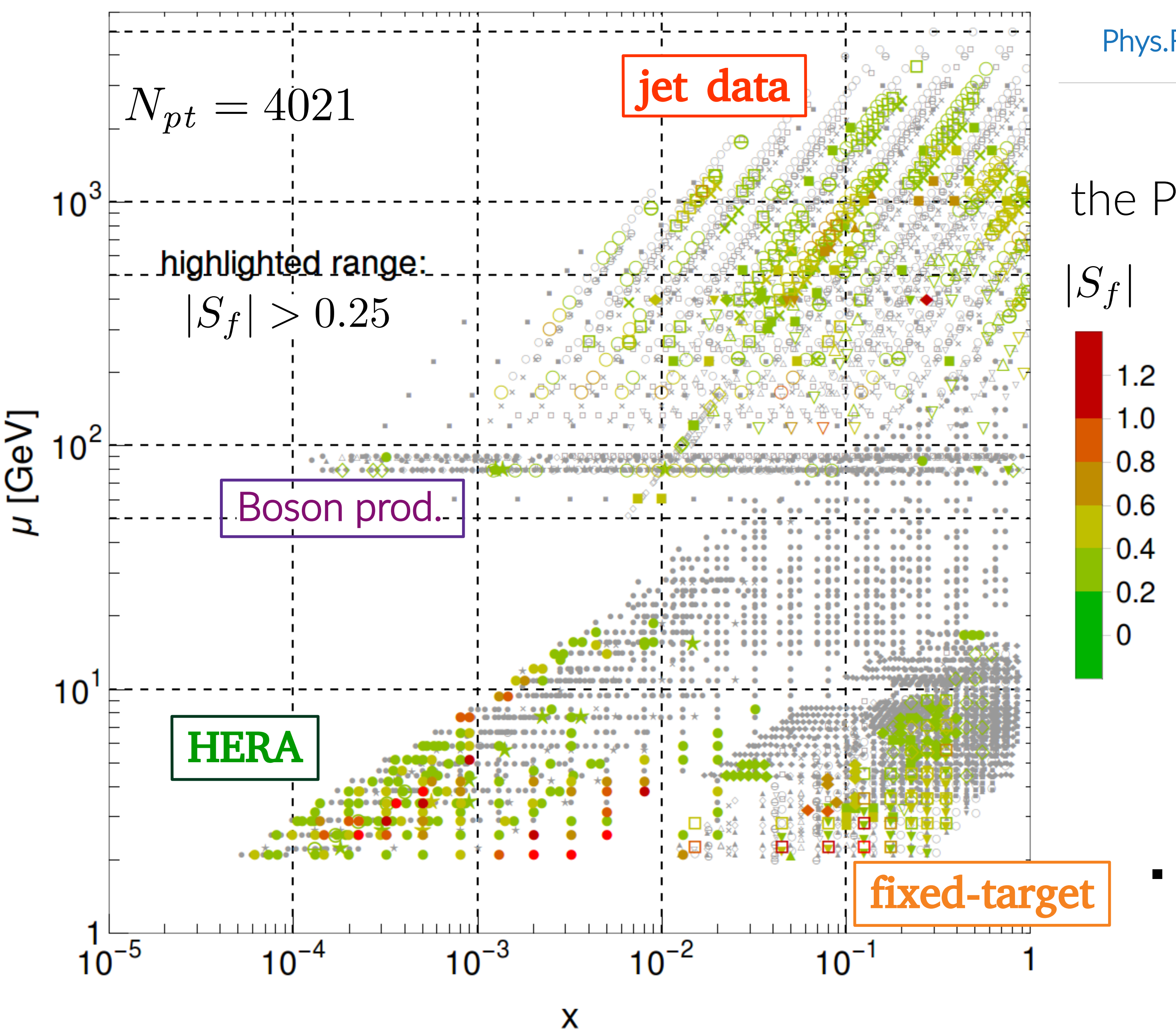
<u>`ePump`</u>: Schmidt, Pumplin, and Yuan; PRD98 (2018) no.9, 094005




- EIC pseudodata supersede fixed-target DIS information in CT fits
- reweighting strongly depends on parametrizations; other ambiguities

→ complementary approaches welcome!

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




II) visualizing impacts with PDFSense

Phys.Rev. D98 (2018) 094030

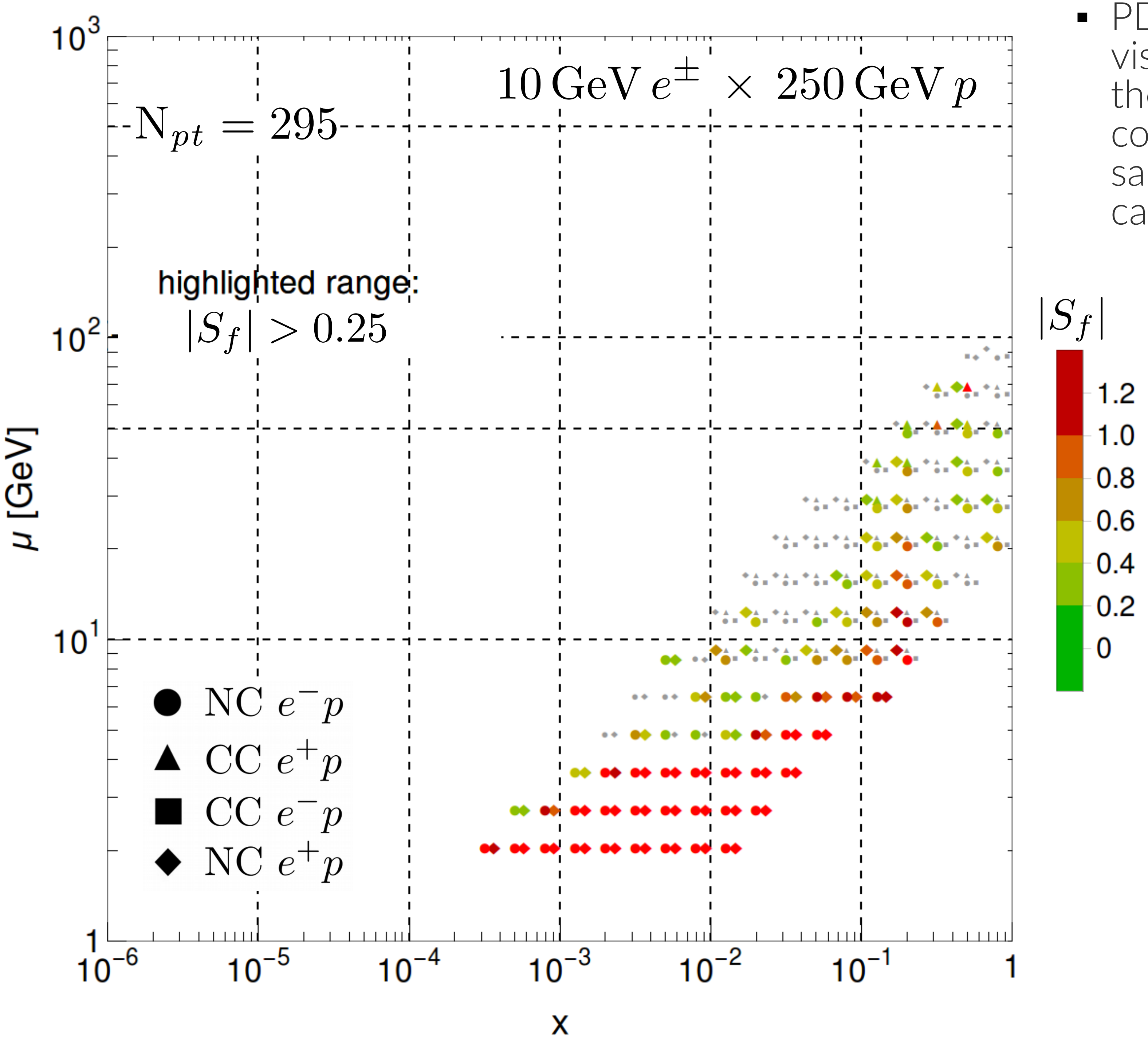
the PDF *sensitivity*

(magnitude of PDF pull of each datum)

→ measurements with strong PDF correlations <u>AND</u> high precision have high $|S_f|$

- used to identify high-impact data for CT18

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

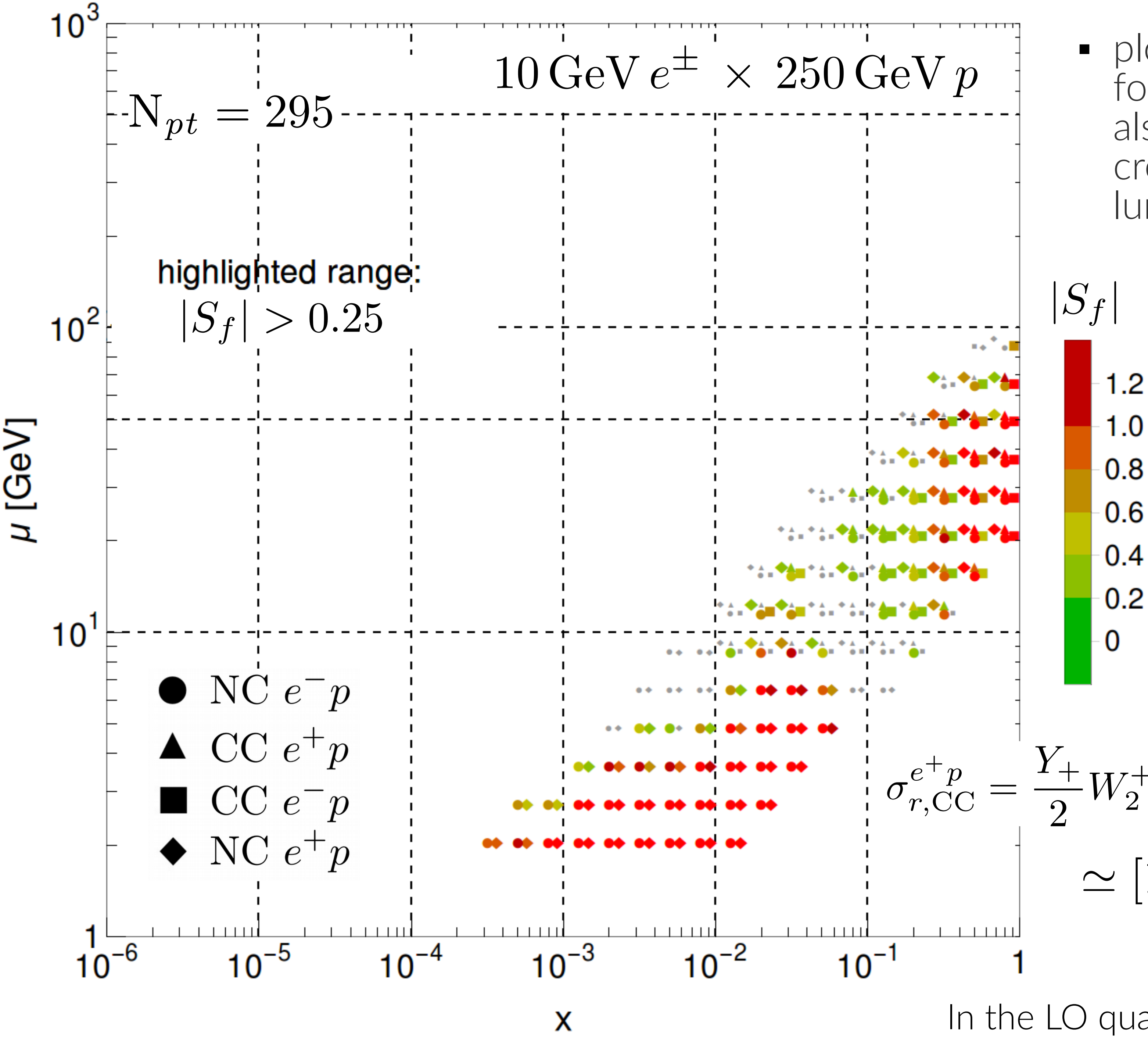



- PDFSense quickly visualizes the pulls of the EIC pseudodata, comparing on the same scale as for the candidate CT18 exps.

- 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 $\sigma_{r,\mathrm{NC}}^{e^{\pm}p}$

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




- plots easily calculable for other PDF flavors; also, Mellin moments, cross sections, parton luminosities, ...

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

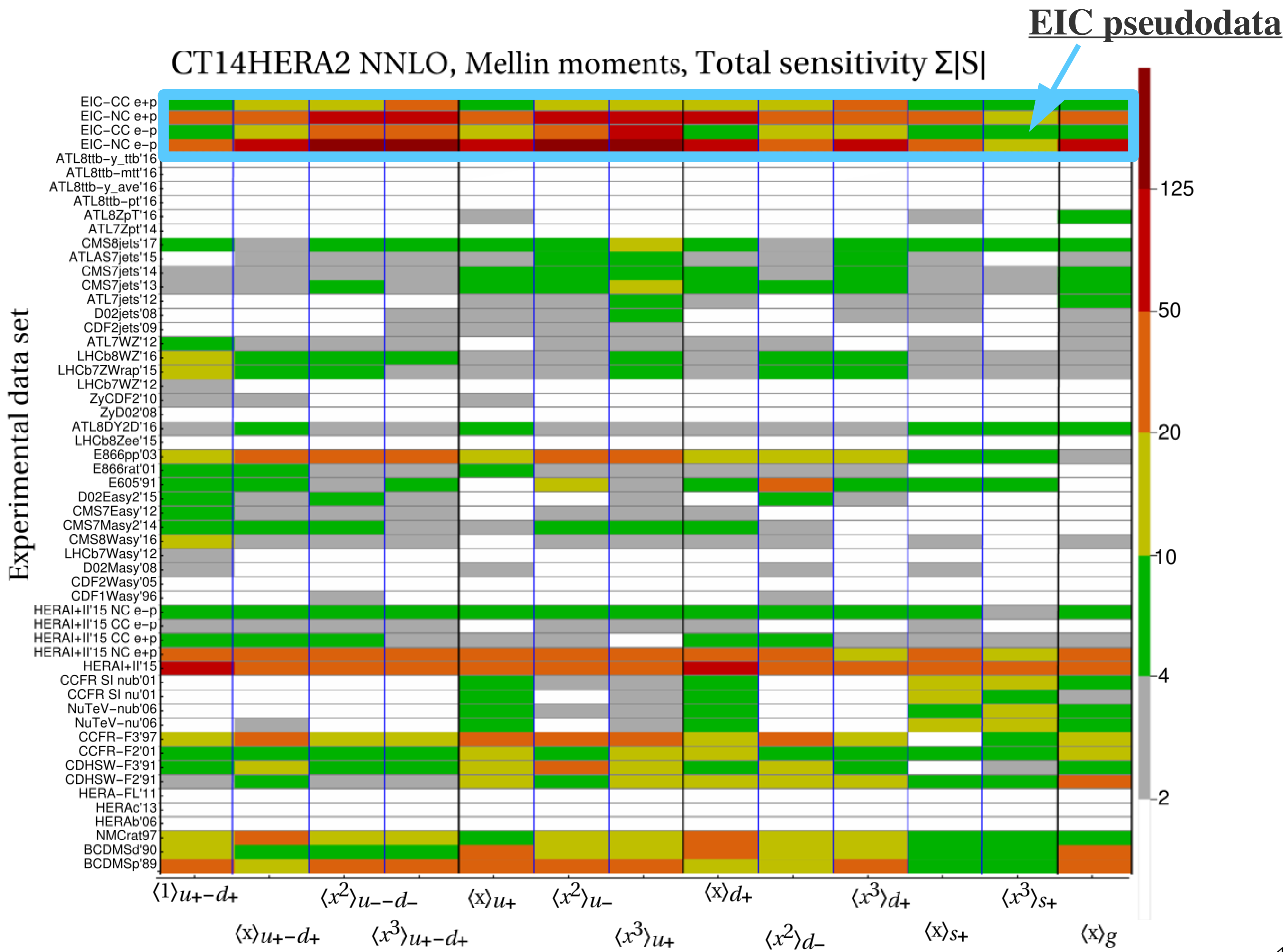
for $x \to 1$

$$\sigma_{r,\mathrm{CC}}^{e^+p} = \frac{Y_+}{2} W_2^+ \mp \frac{Y_-}{2} x W_3^+ - \frac{y^2}{2} W_L^+$$

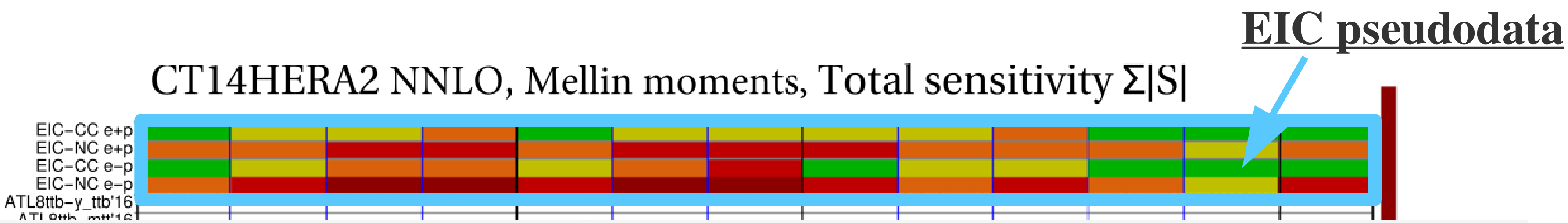
$$\simeq [1-y]^2\, x(d+s)$$

In the LO quark-parton model

# sensitivities can be aggregated for direct comparisons of exps



# sensitivities can be aggregated for direct comparisons of exps



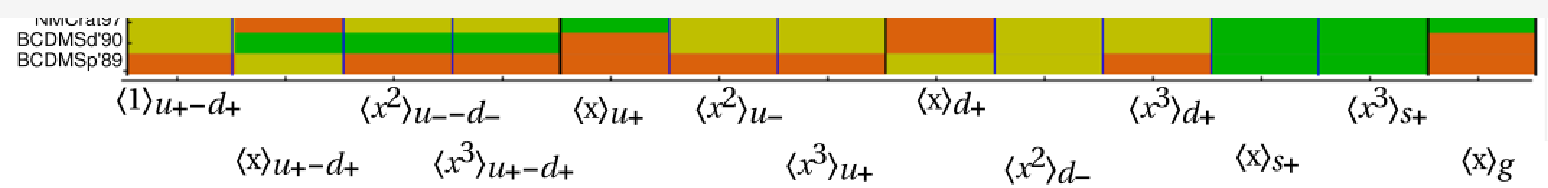

PDFSense* furnishes:

→ quick evaluations of pulls of individual data

→ info. less tied to PDF parametrization(s)

→ aggregate scores to compare experiments

*powerful in combination with direct fitting/reweighting



# conclusions, recommendations

- **the Yellow Initiative has an aggressive timeline; necessitates quick action**

→ need: an efficient, effective workflow to connect EIC design scenarios to physics output

→ in conjunction with fitting, fast methods here can inform/integrate into this workflow

→ we need guidelines on EIC operation as initial input

- performance in energy, lumi, systs (baseline ↔ optimistic scenarios)

→ identify 'lumi-hungry' vs. syst-limited measurements; explore error scenarios

- inclusive DIS will constrain: glue (scaling violations); $d$-PDF (high-lumi CC); lattice-calculable quantities; …

→ not sufficient alone for flavor sep.; need input from SIDIS/HQ production

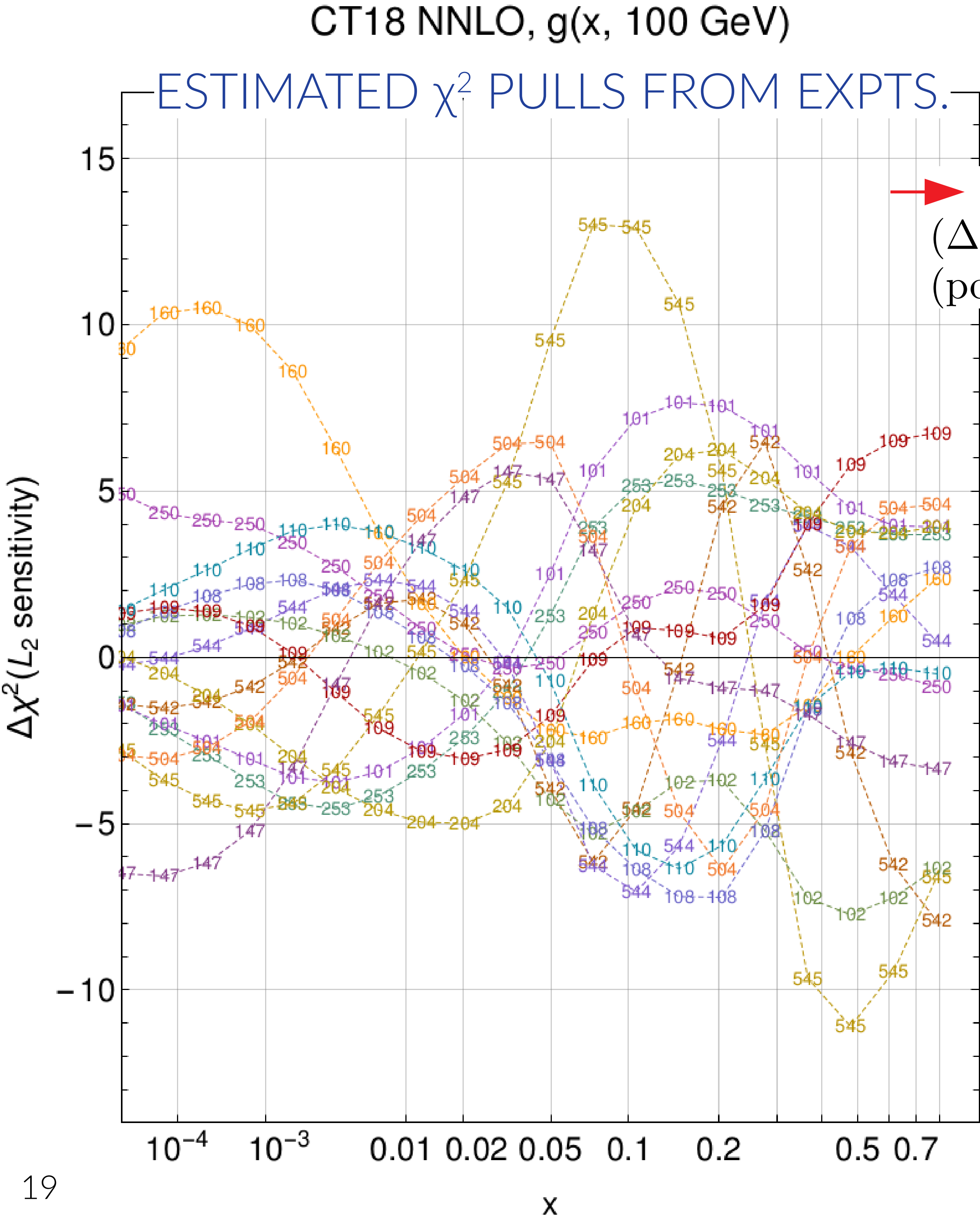
**… IRG activity must coordinate with other WGs!**

THANKS!

# supplementary material

# ESTIMATED $\chi^2$ PULLS FROM EXPTS.

TJH, Wang, Nadolsky, Olness;
PRD**100** (2019) 094040




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

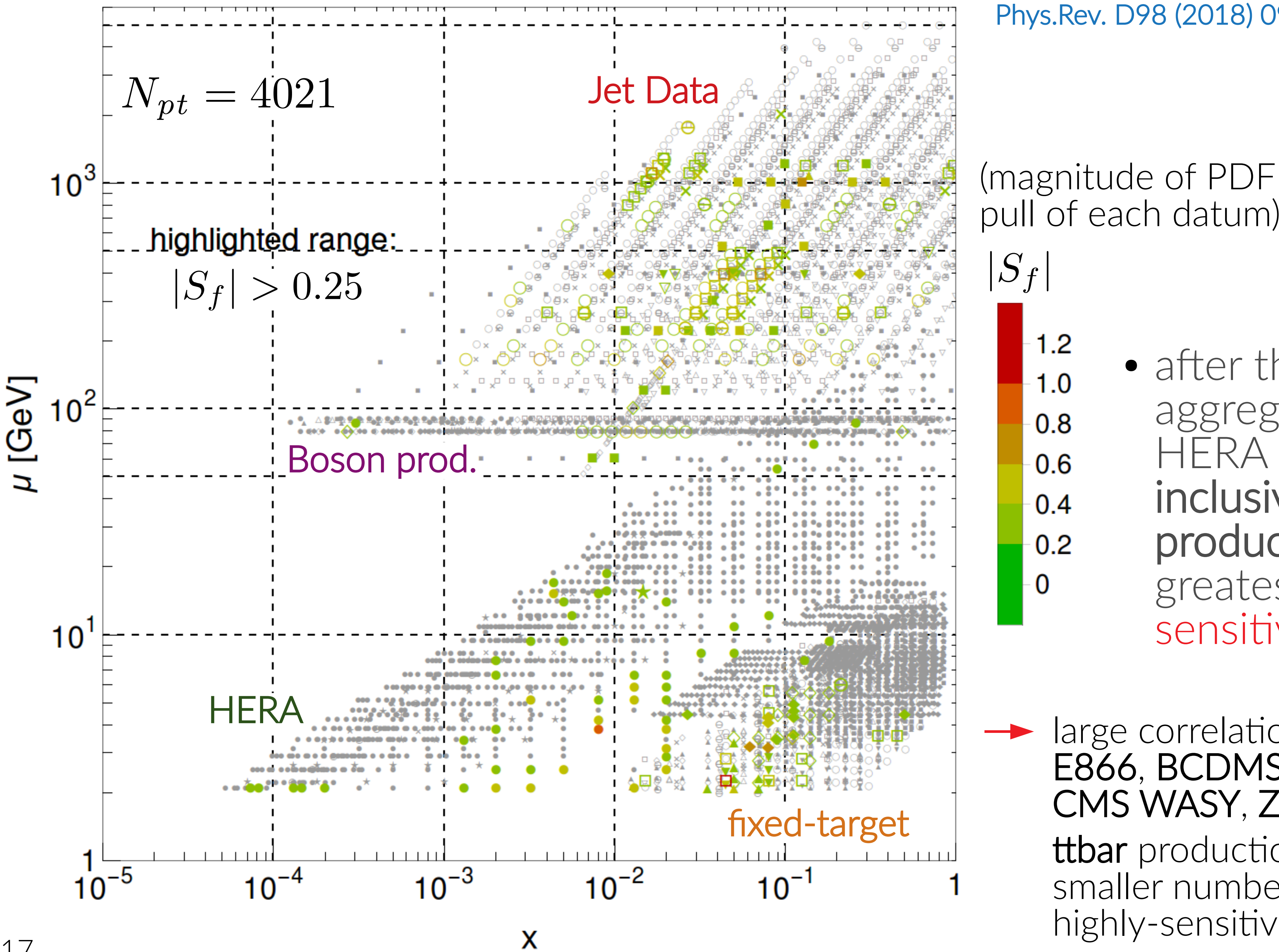
stronger (anti-) correlation

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

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

B.-T. Wang, TJH, S. Doyle, J. Gao, T.-J. Hou, P. M. Nadolsky, F. I. Olness

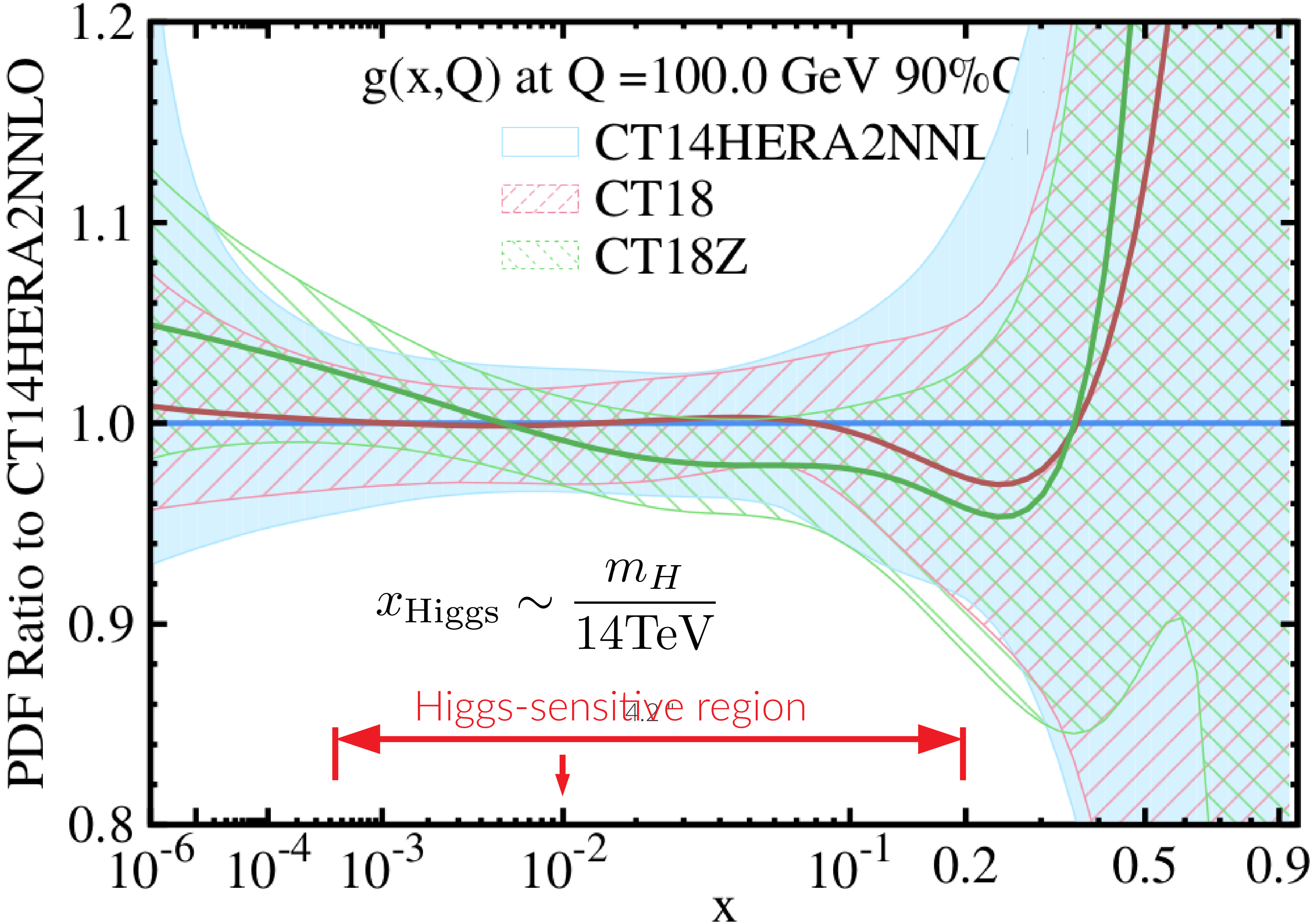
Phys.Rev. D98 (2018) 094030




- 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

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

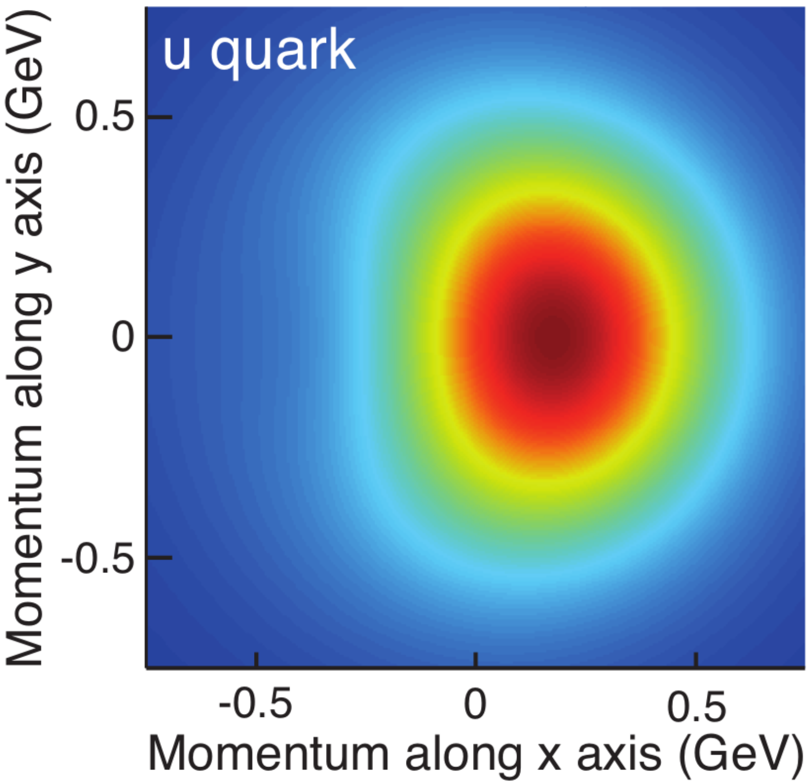


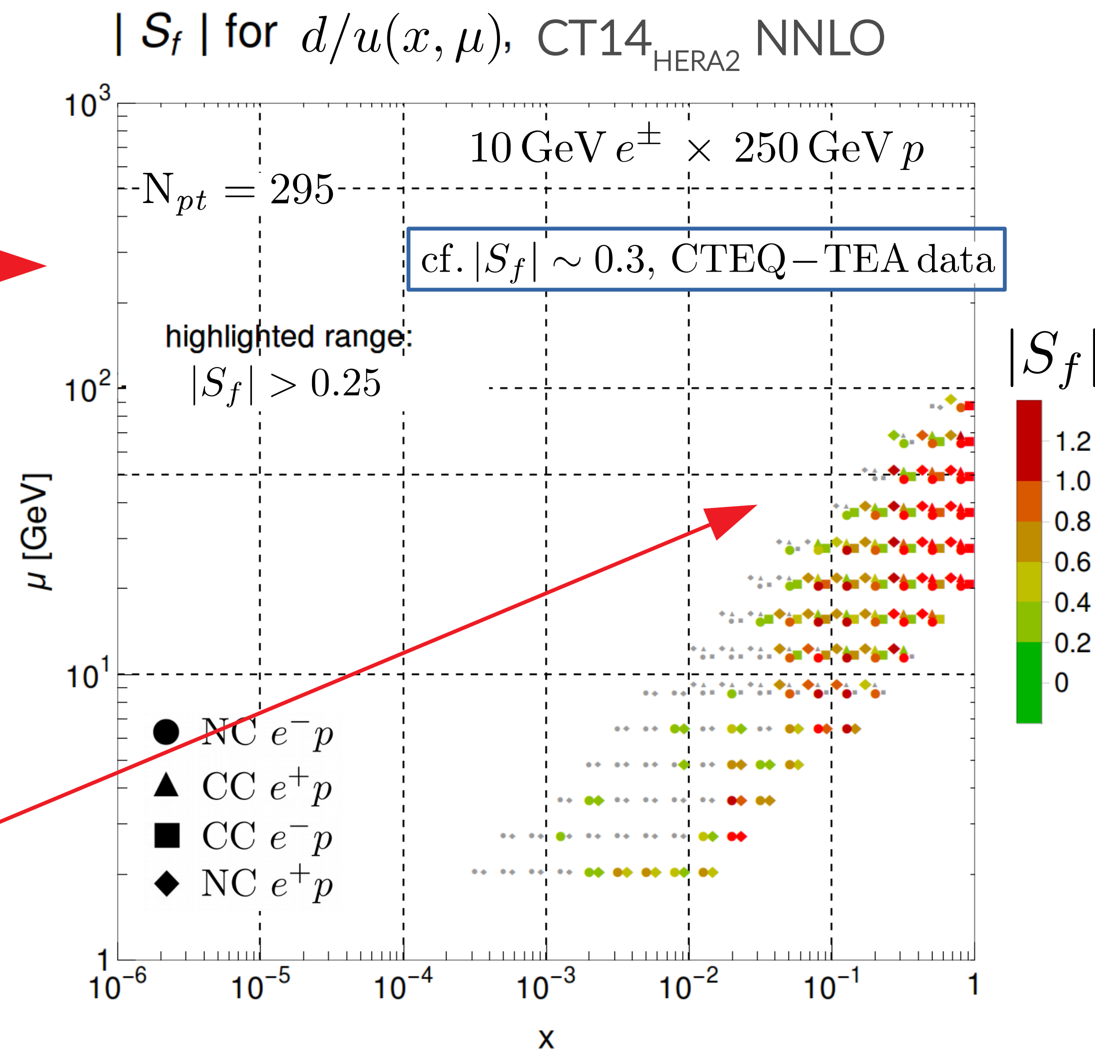

- while LHC Run-1 data drive important PDF improvements, including for the gluon at high-, low-$x$, the effect is relatively incremental

# 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

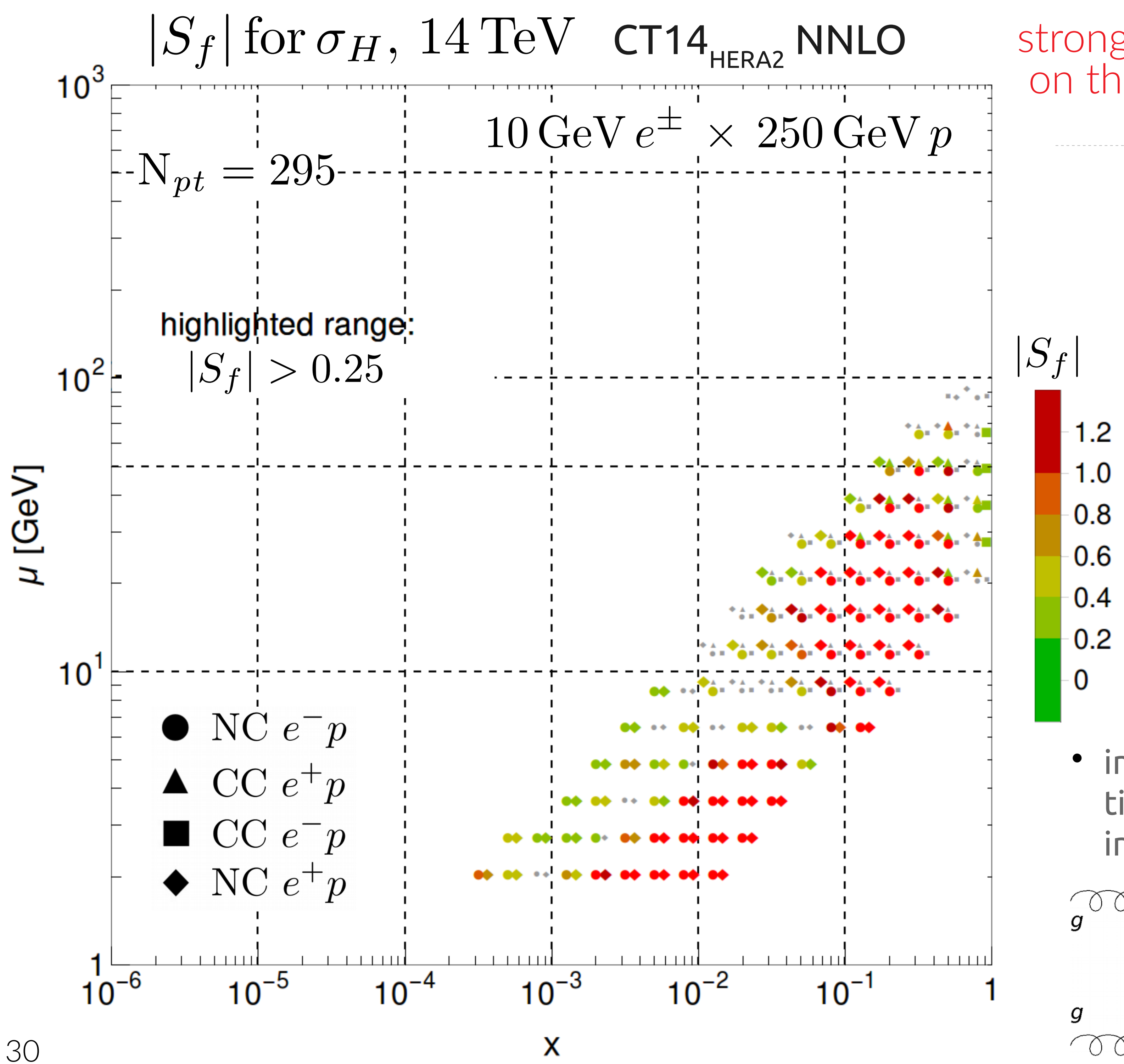
Accardi et al., EPJA52 (2016) no.9, 268.




PROJECTED IMPACT OF EIC PSEUDO-DATA VERY LARGE – RED SIMULATED MEASUREMENTS

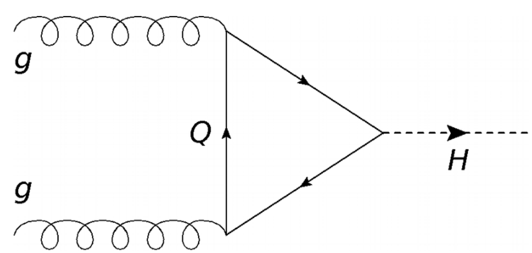
- DIS cross sections from EIC will supercede the bulk of fixed-target information in contemporary QCD fits; provide an 'anchor-point' to resolve systematic PDF tensions



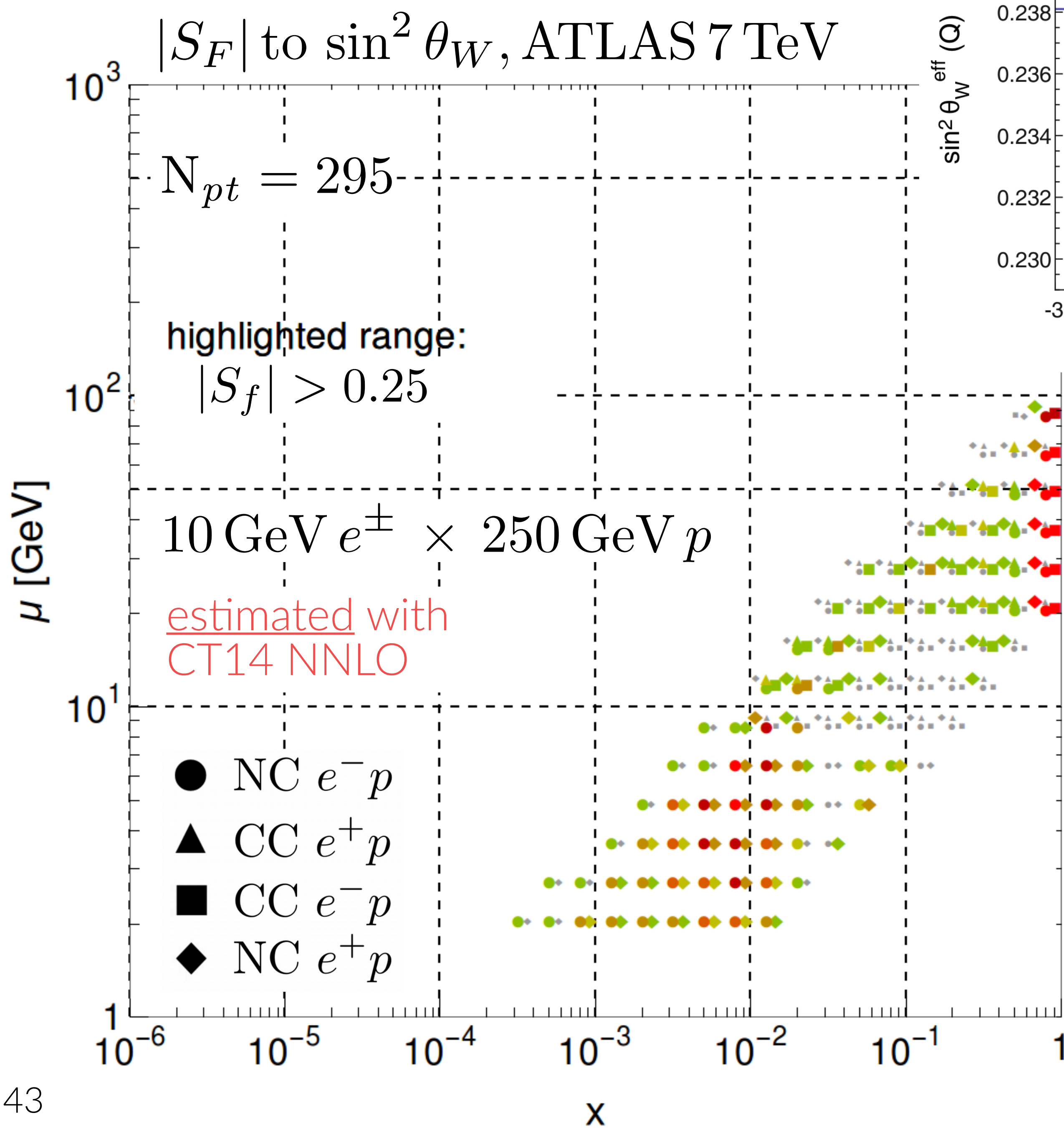

strong predicted impact on the Higgs sector

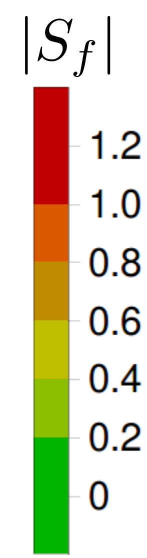
- the impact of an EIC upon the theoretical predictions for inclusive Higgs production arises from a very broad region of the kinematical space it can access

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

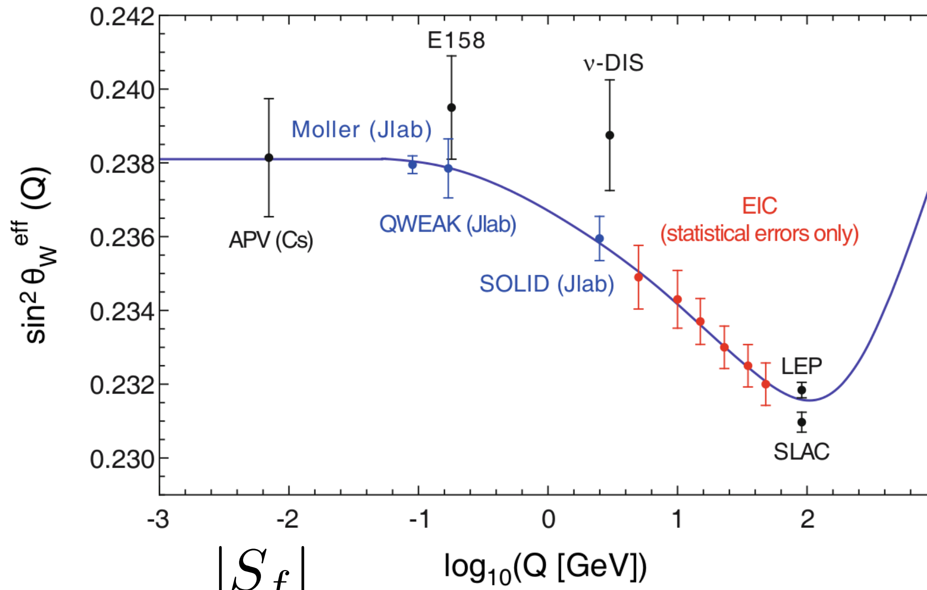


# an EIC will probe EW parameters and New Physics!




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 $A^{\mathrm{PV}}$ — relatively large $Q^2$ and in the x range $0.2 \lesssim x \lesssim 0.5$

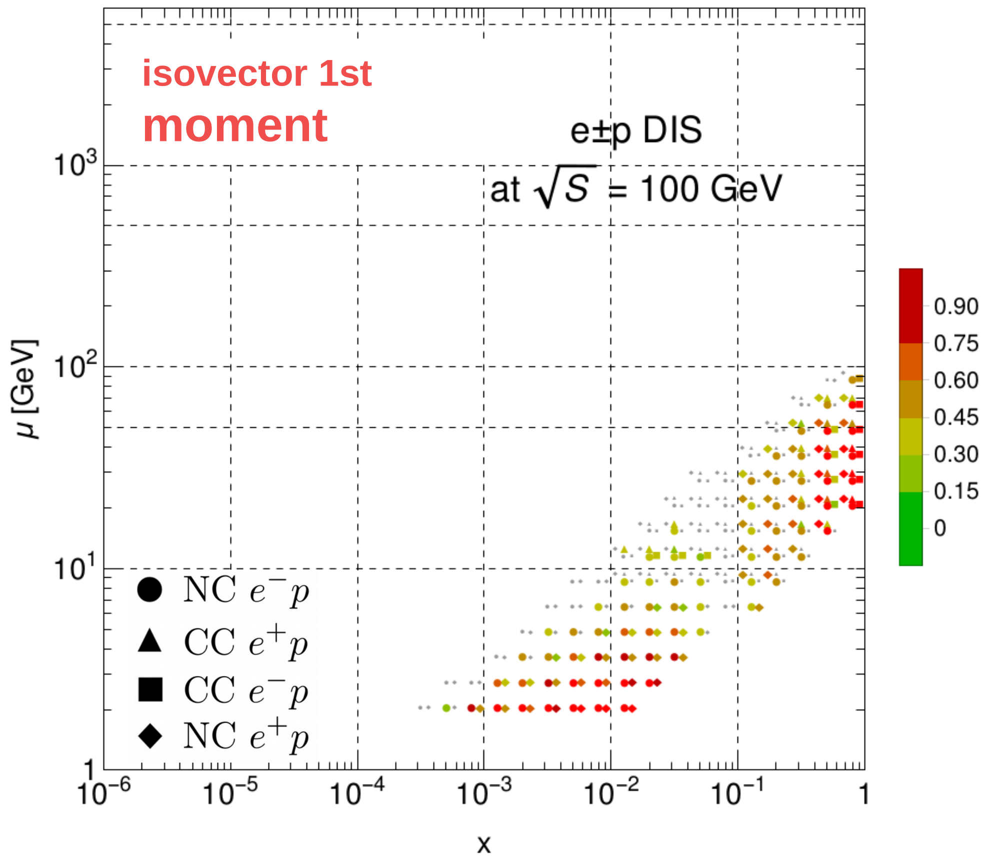
# An EIC would drive lattice phenomenology

arXiv:1904.00022 [hep-ph]

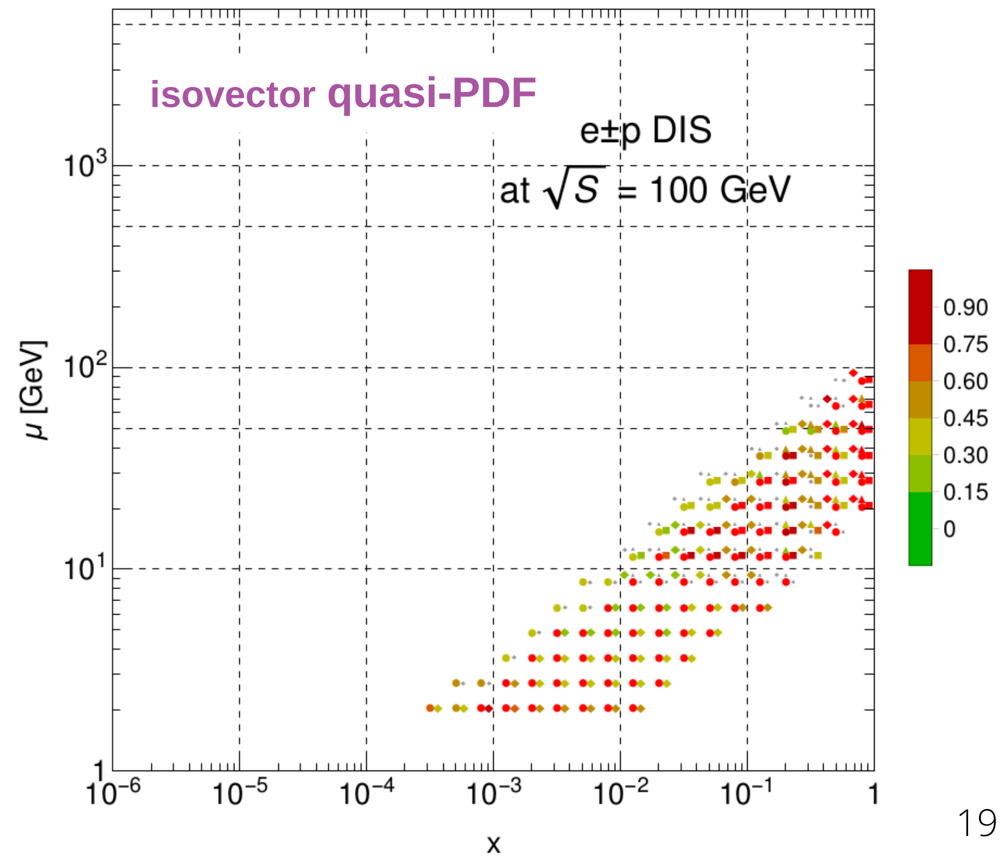
- A high-luminosity lepton-hadron collider will impose very tight constraints on many lattice observables; below, the isovector first moment and qPDF; this is crucial for benchmarking!
- Many of the experiments most sensitive to PDF Mellin moments and qPDFs involve nuclear targets $\longrightarrow$ **eA data from EIC would sharpen knowledge of nuclear corrections**

$$\langle x^n \rangle_{q,g} = \int dx\, x^n f_{q,g}(x, \mu = 2\,\mathrm{GeV}) \qquad \widetilde{q}(x, P_z, \widetilde{\mu}) = \int dy\, Z\left(\frac{x}{y}, \frac{\Lambda}{P_z}, \frac{\mu}{P_z}\right) q(y,\mu) + \mathcal{O}\left(\frac{\Lambda^2_{\rm QCD}}{P_z^2}, \frac{M^2}{P_z^2}\right)$$

$| S_f |$ for $\langle x^1 \rangle_{u^+-d^+}$, CT14HERA2

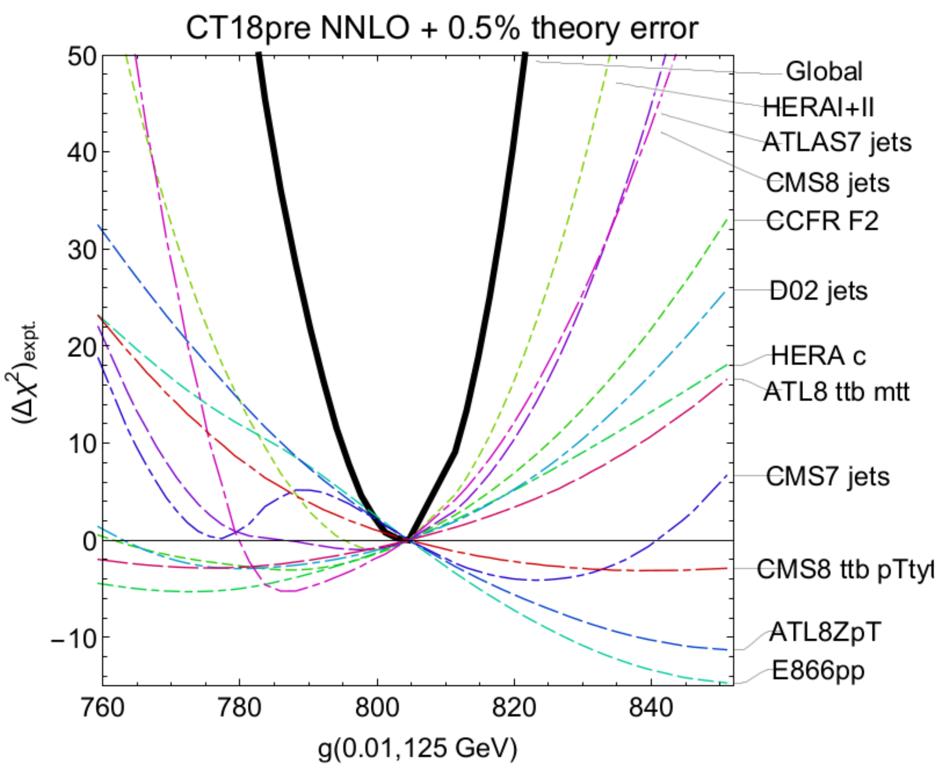



$| S_f |$ for $[\tilde{u}-\tilde{d}]$(x=0.85, $P_z$=1.5GeV), CT14HERA2



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

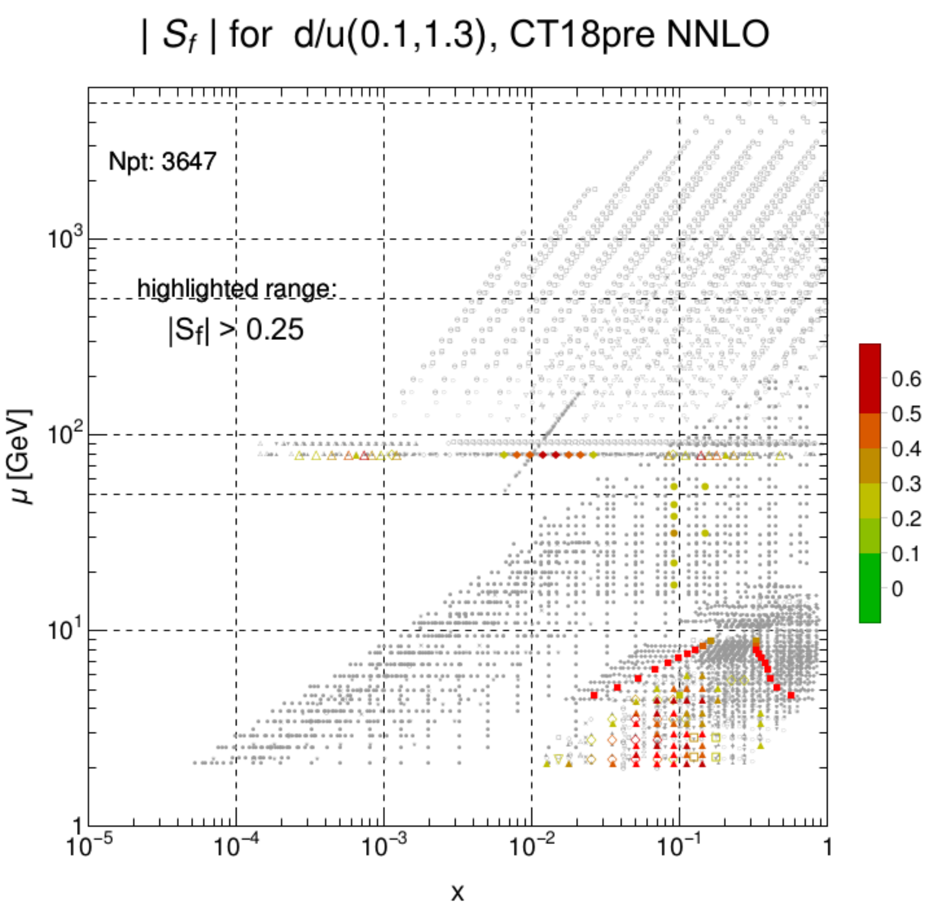
...for the gluon PDF in the Higgs region, $g(0.01, m_H)$

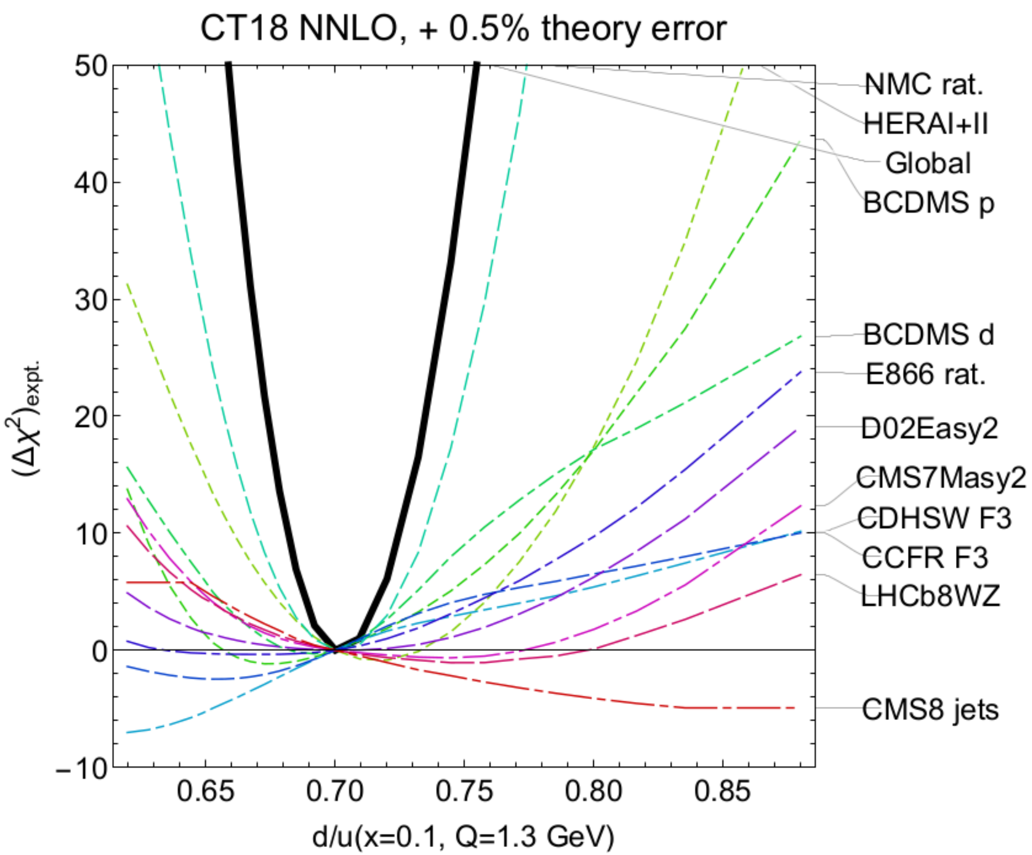



| $g(x=0.01, \mu=125\ \mathrm{GeV})$ | | |
|---|---|---|
| PDFSense | | LM scan |
| CT14HERA2 | CT18pre | CT18pre |
| HERAI+II'15 | HERAI+II'15 | HERAI+II'15 |
| CMS8jets'17 | CMS8jets'17 | CMS8jets'17 |
| CMS7jets'14 | CMS7jets'14 | ATL8ZpT'16 |
| ATLAS7jets'15 | E866pp'03 | E866pp'03 |
| E866pp'03 | 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]

# 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

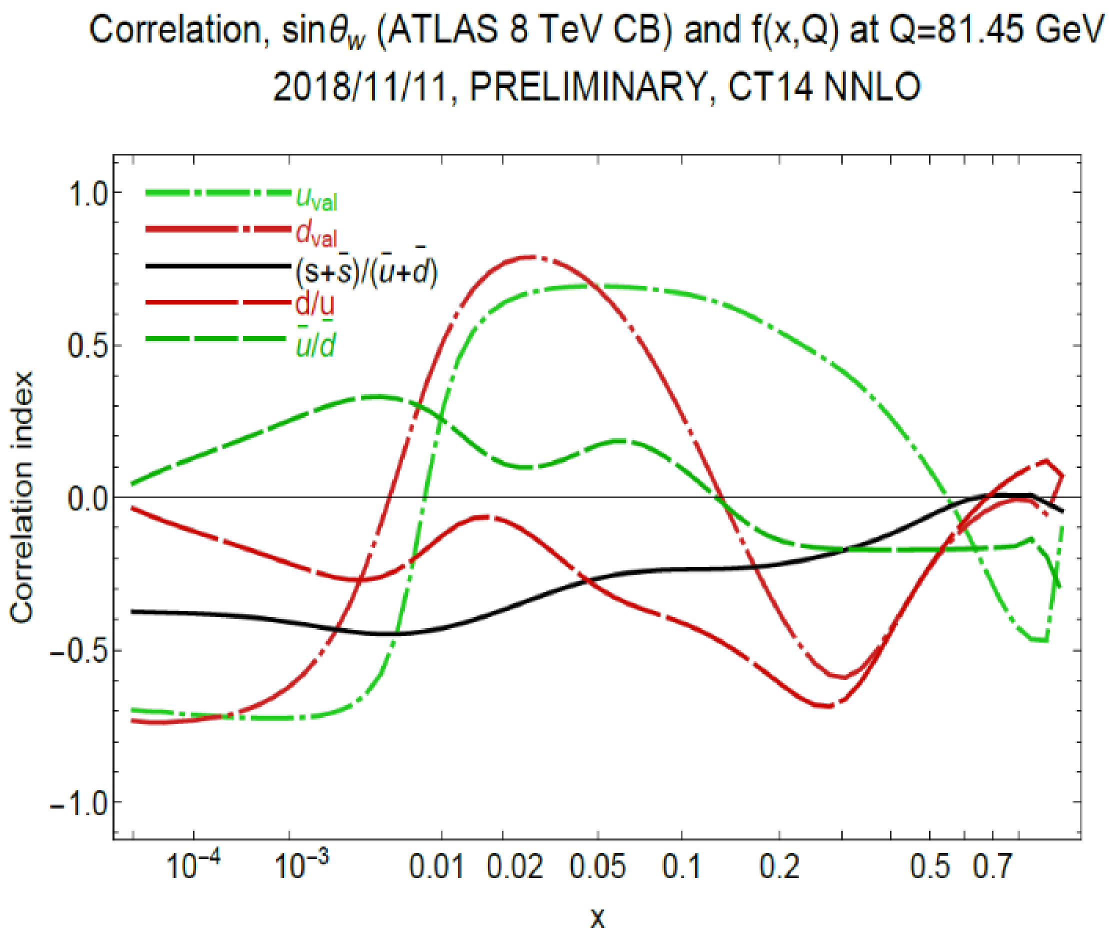
HERA and fixed-target (BCDMS, NMC) data are dominant!

$\sin^2\theta_W$ (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 $0.005 \lesssim x \lesssim 0.2$

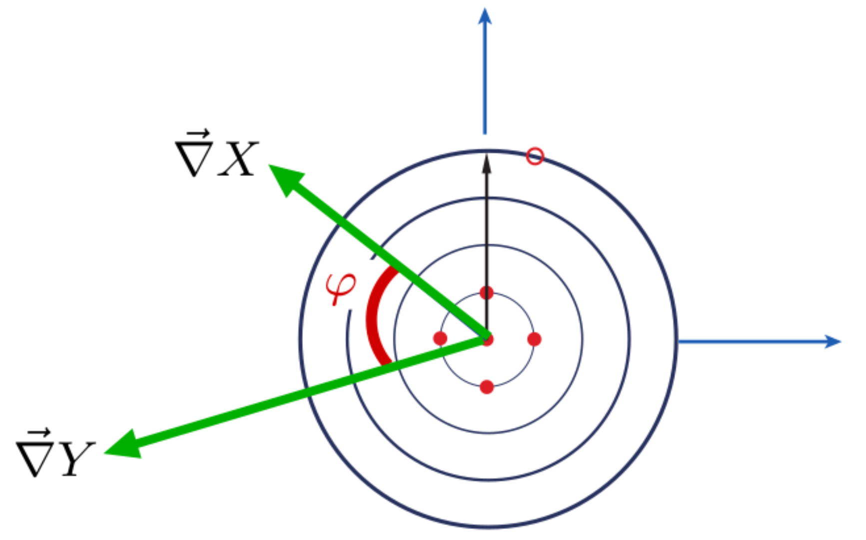
weak correlations with $\bar{u}$, $\bar{d}$, $\bar{s}$, $g$

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

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

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

the *L2 sensitivity*



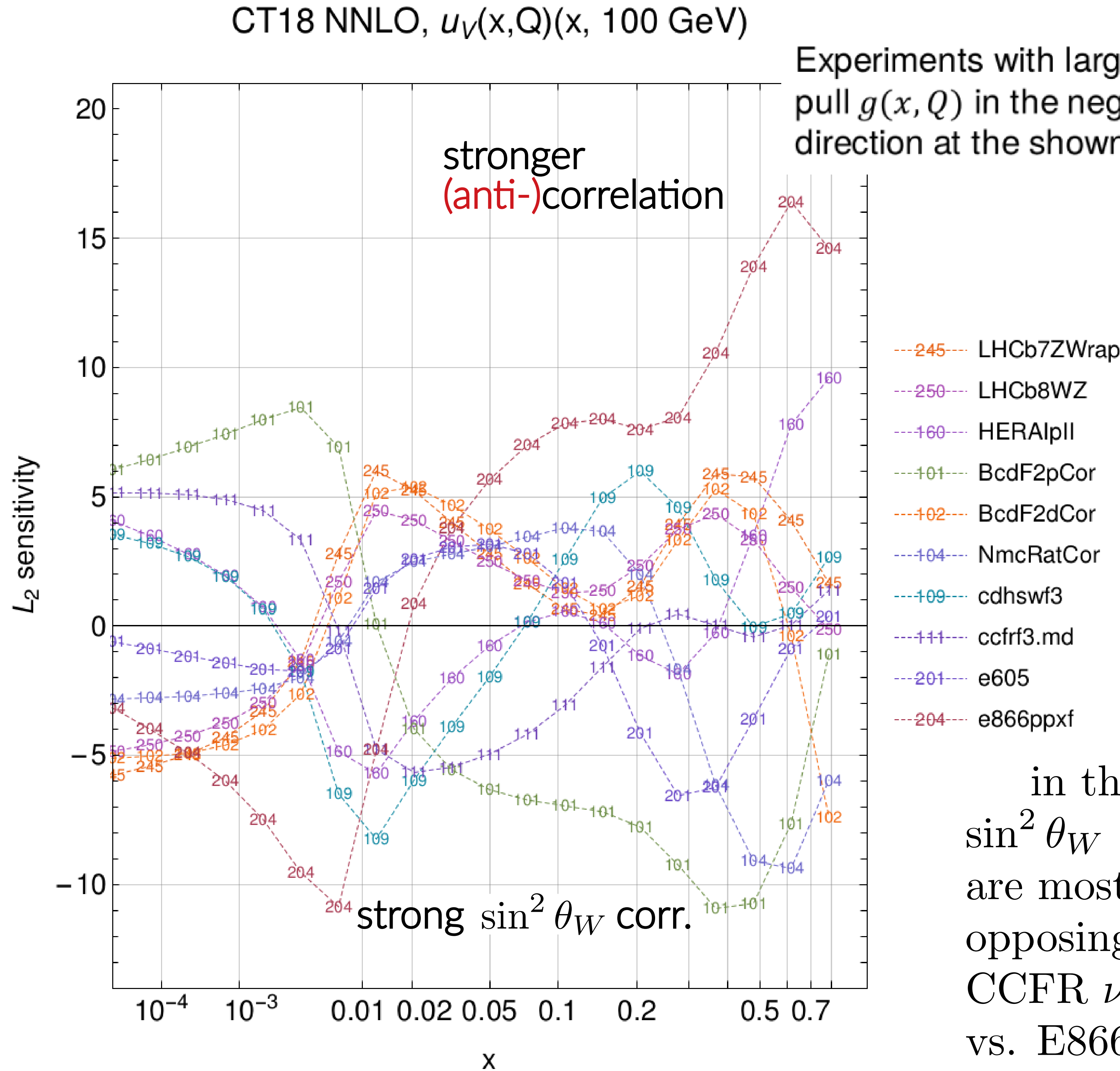

$L_2$ sensitivity. Take $X = f_a(x_i, Q_i)$ or $\sigma(f)$; $Y = \chi^2_E$ for experiment $E$. Find $\Delta Y(\vec{z}_{m,X})$ for the displacement $|\vec{z}_{m,X}| = 1$ along the direction $\vec{\nabla} X / \left|\vec{\nabla} X\right|$ (corresponding to $\Delta\chi^2_{tot} = T^2$ and $X(\vec{z}) = X(0)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

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

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

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

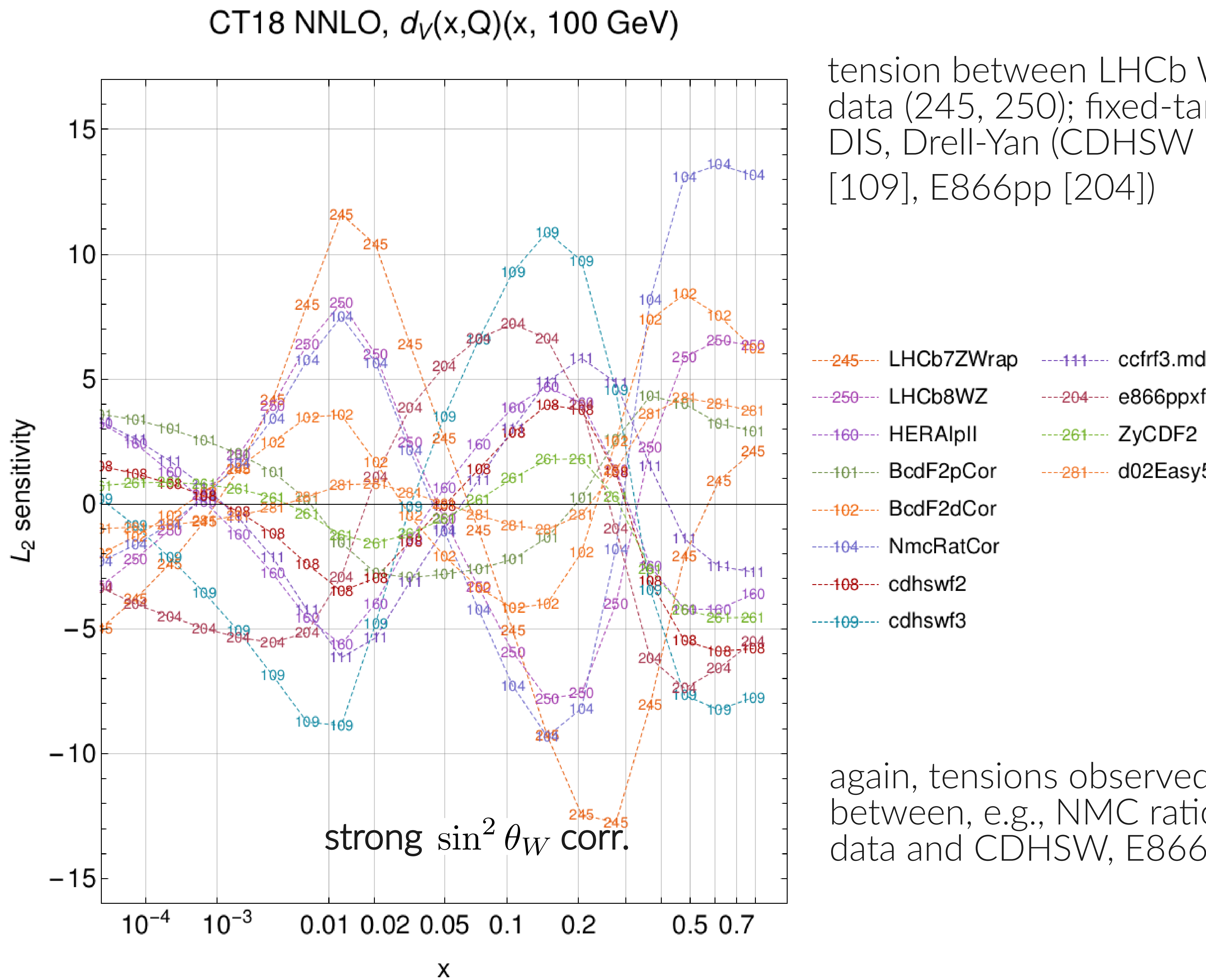
...extent to which total $\chi^2_E$ of specific expts. correlates with $x$-dep. of PDFs




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])

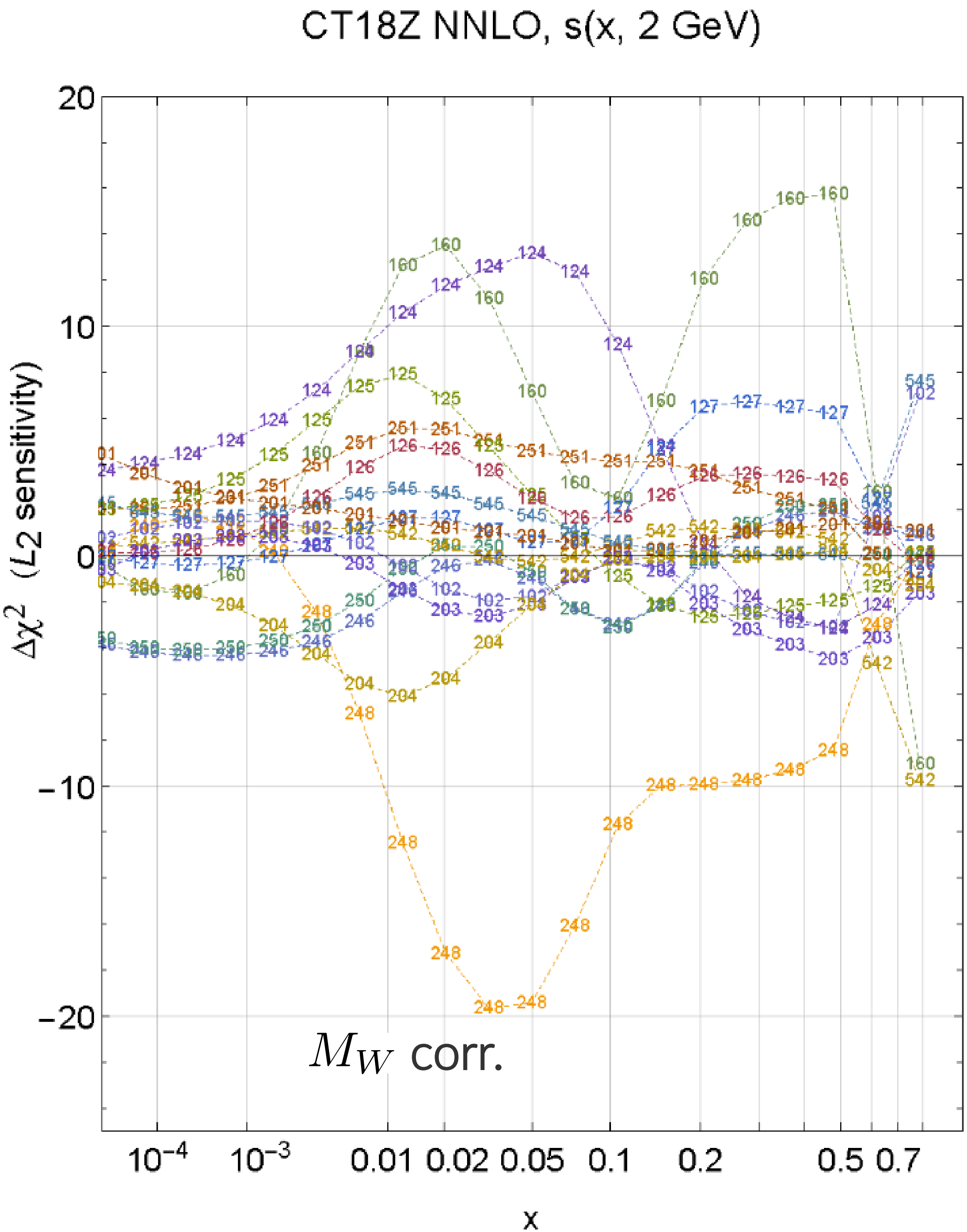
in the region where $\sin^2\theta_W$ and $u_v$ are most correlated, opposing pulls from CCFR $\nu$DIS, BCDMS vs. E866 $pp$, NMC rat.




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

# $L_2$ sensitivity, strangeness: CT18Z




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!

# 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:

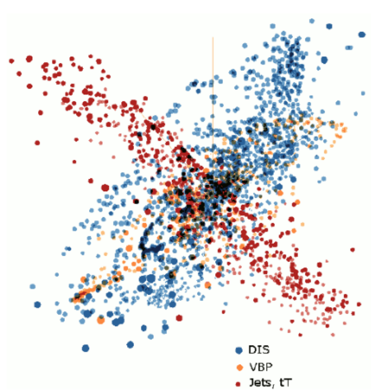
$$D_i \to D_{i,sh}(\vec{a}) = D_i - \sum_{\alpha=1}^{N_\lambda} \beta_{i\alpha}\overline{\lambda}_\alpha(\vec{a})$$

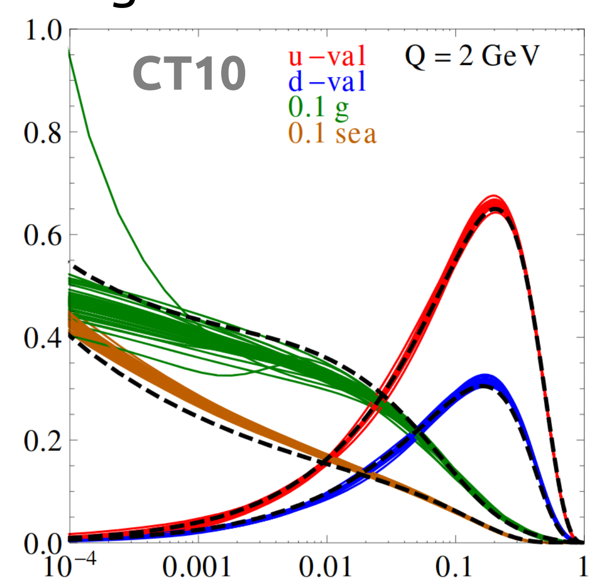
- a 56-dimensional parametric basis $\vec{a}$ is obtained by diagonalizing the Hessian matrix H determined from $\chi^2$ (following a 28-parameter fit)

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



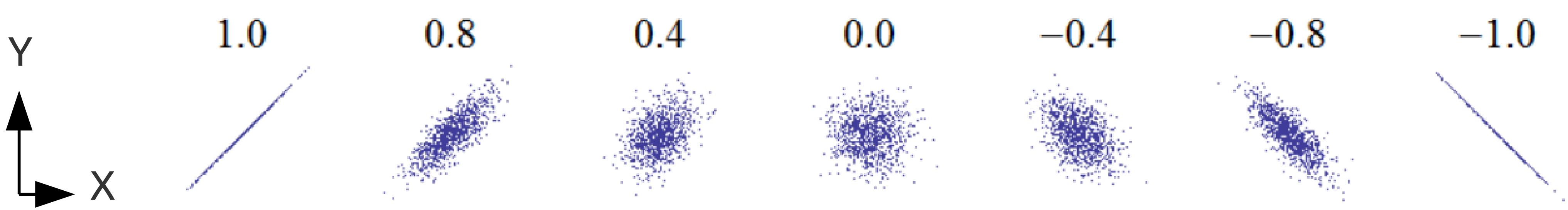



# a brief statistical aside, ii

- **… 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**



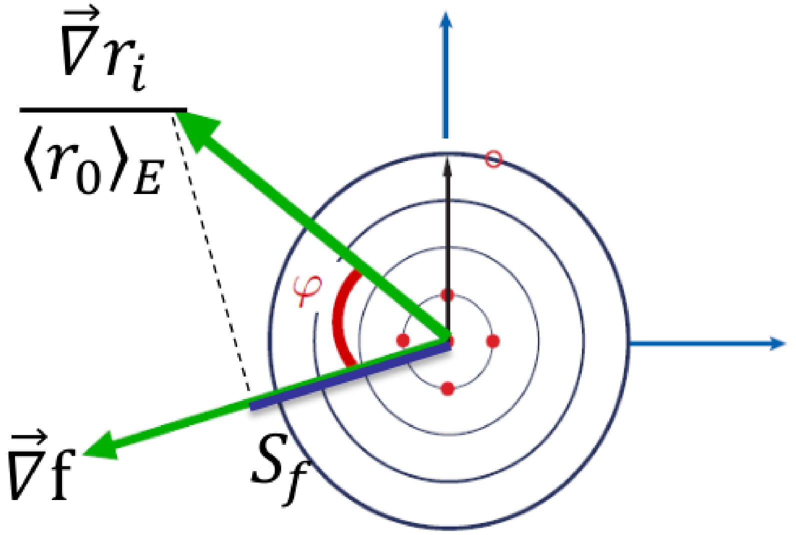

[X,Y] are exactly (anti-)correlated at the far (right) left above.

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

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

# Correlation $C_f$ and sensitivity $S_f$

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




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

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

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

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

$C_f$ 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.

# $2^{nd}$ 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

$\mu_i$ : factorization scale

deeply-inelastic scattering:

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

hadron-hadron collisions:

$AB \to CX$

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

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