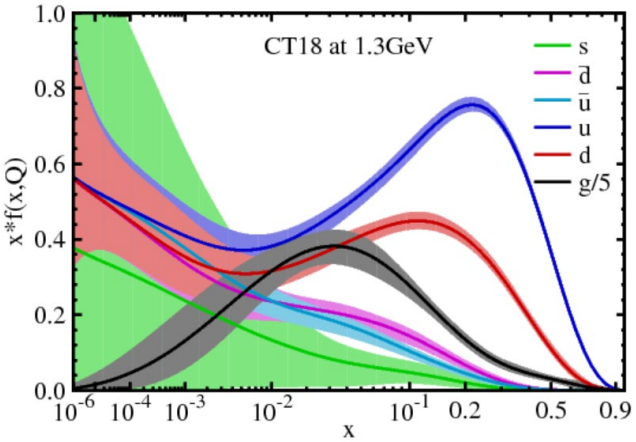
Precise analysis of hadron structure for the LHC era

Pavel Nadolsky

with Fred Olness, Sean Doyle, Madeline Hamilton, Tim Hobbs, Bo-Ting Wang, Keping Xie

Southern Methodist University



CTEQ-TEA (Tung et al.) working group

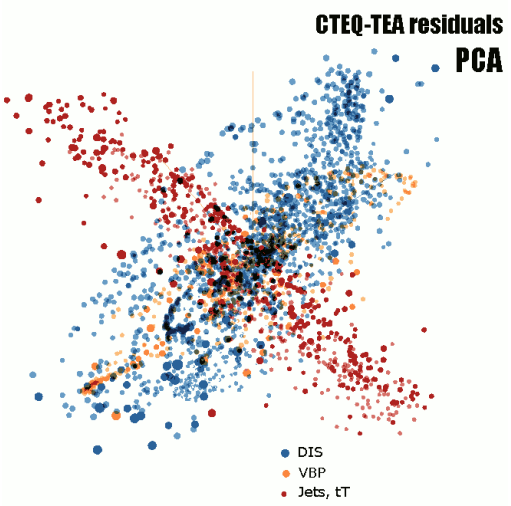
China Northeastern University: T.-J. Hou

Kennesaw State University: M. Guzzi

Michigan State U.: J. Huston, J. Pumplin, D. Stump, C. Schmidt, J. Winter, C.-P. Yuan

Shanghai Jiao Tong University: J. Gao

Xinjiang University: S. Dulat, I. Sitiwaldi

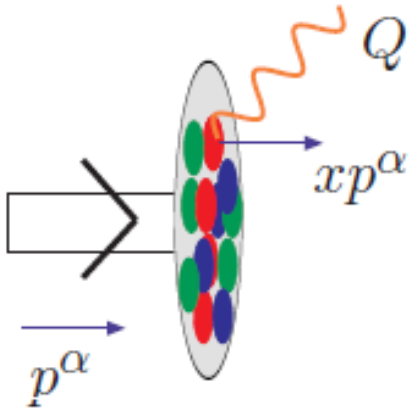


SMU

CTEQ

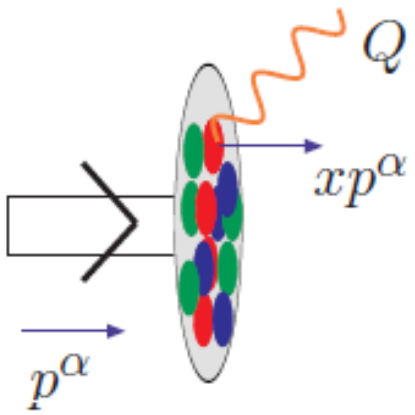
The inner world of a hadron

A short-distance probe (virtual photon, heavy boson, gluon) resolves increasingly small structures inside the nucleon.



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

Unpolarized collinear parton distributions $f_{a/h}(x, Q)$ 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



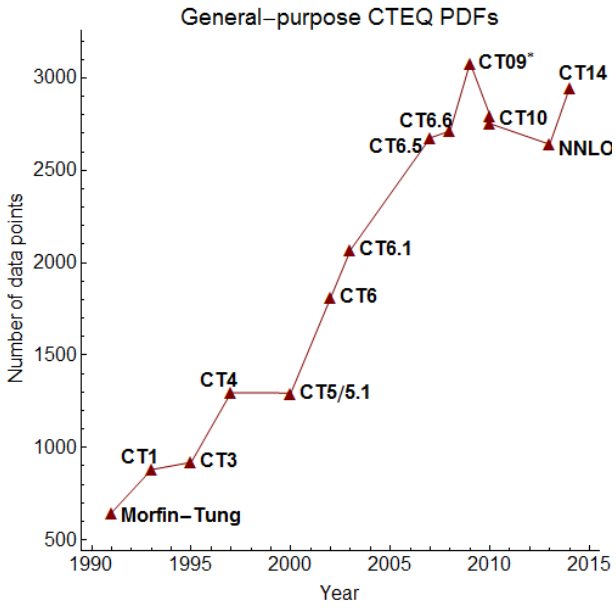
Coordinated Theoretical Experimental study of QCD

Initiated around 1990 to stimulate interactions between

- Experimentalists and theorists, especially at the newly built Tevatron
- High-energy physics and hadronic physics communities

This is achieved by various initiatives:

- **Global analysis** *(the term coined by J. Morfin and Wu-Ki Tung)* constrains PDFs or other nonperturbative functions with data from diverse hadronic experiments
- **Workshops and summer schools**
- **Annual Wu-Ki Tung award** for junior researchers working on intersections of experiment and theory **[nominate by August 15 each year]**

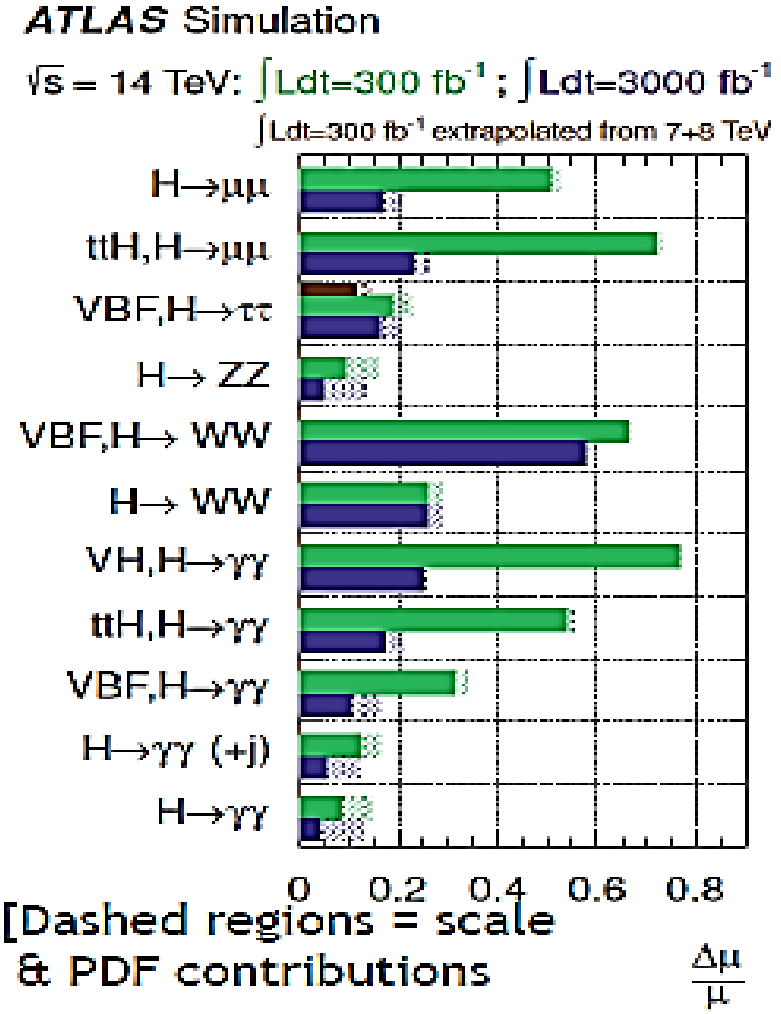


2019: new experiments (LHC, EIC, LHeC,…)! New objectives!

QCD expectations for high-luminosity LHC

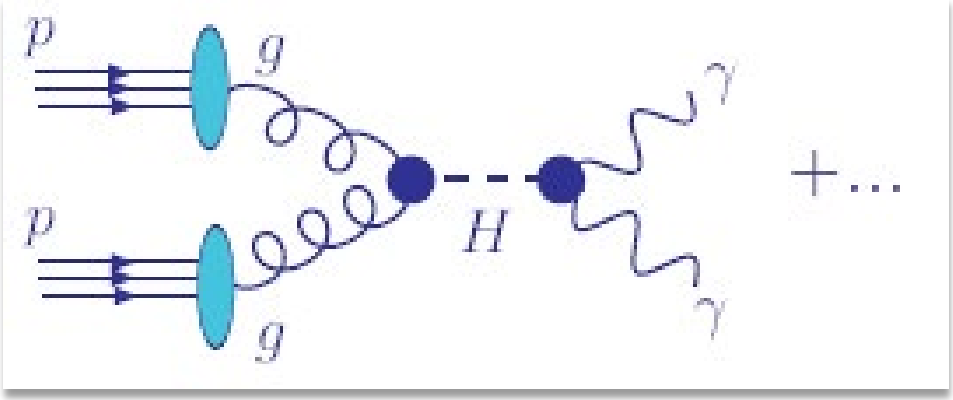
- Measurements of Higgs cross sections/couplings become limited by PDFs in the HL-LHC era
- Searches for non-resonant production in TeV mass range will demand accurate predictions for **sea** PDFs at $x > 0.1$
- The target is to obtain PDFs that "achieve 1% accuracy for LHC predictions" within about a decade

Projected Experimental Uncertainties



P. Newman, DIS'2016

Parton distributions describe long-distance dynamics in high-energy collisions



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

$\hat{\sigma}$ is the hard cross section

$f_a(x, \mu_F)$ is the distribution for parton a with momentum fraction x, at scale μ_F

Operator definition of PDFs; evolution equations

To all orders in α_s, PDFs are **defined** as matrix elements of certain correlator functions:

$$f_{q/p}(x,\mu) = \frac{1}{4\pi}\int_{-\infty}^{\infty} dy^- e^{iy^- p^+} \langle p \left| \overline{\psi}_q(0,y^-,\vec{0}_T)\gamma^+ \psi_q(0,0,\vec{0}_T) \right| p\rangle, \text{ etc.}$$

The exact form of $f_{a/p}$ is not known; but its μ dependence is described by **D**okshitzer-**G**ribov-**L**ipatov-**A**ltarelli-**P**arisi (**DGLAP**) equations

$$\mu \frac{df_{i/p}(x,\mu)}{d\mu} = \sum_{j=g,u,\bar{u},d,\bar{d},....} \int_x^1 \frac{dy}{y} P_{i/j}\left(\frac{x}{y}, \alpha_s(\mu)\right) f_{j/p}(y,\mu)$$

$P_{i/j}(x,\mu)$ are known up to N3LO

⇒ Starting from parametrizations of $f_{i/p}(x,\mu_0)$ at $\mu_0 \approx 1\ GeV$, DGLAP equations predict $f_{i/p}(x,\mu)$ at $\mu \geq \mu_0$

Perturbative QCD loop revolution

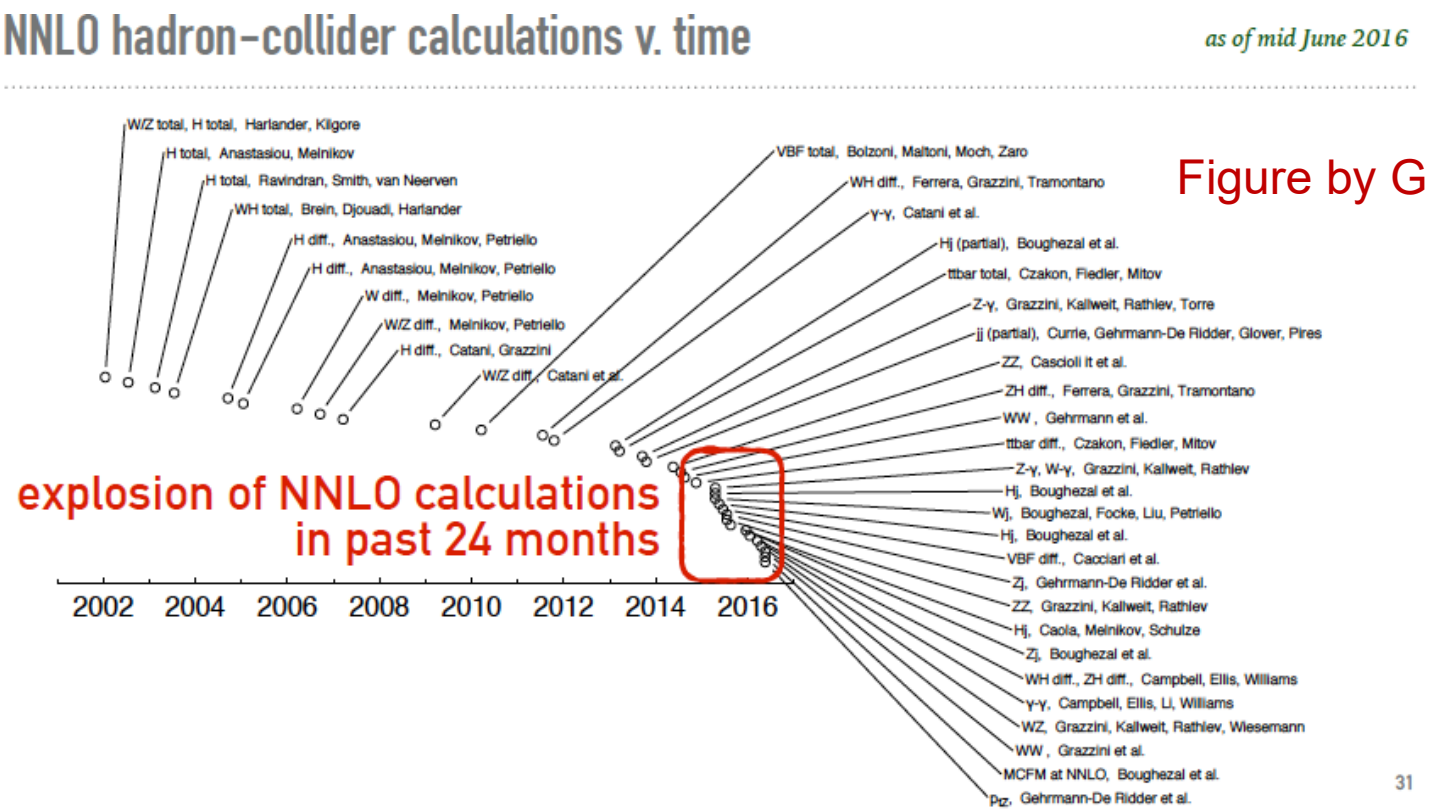
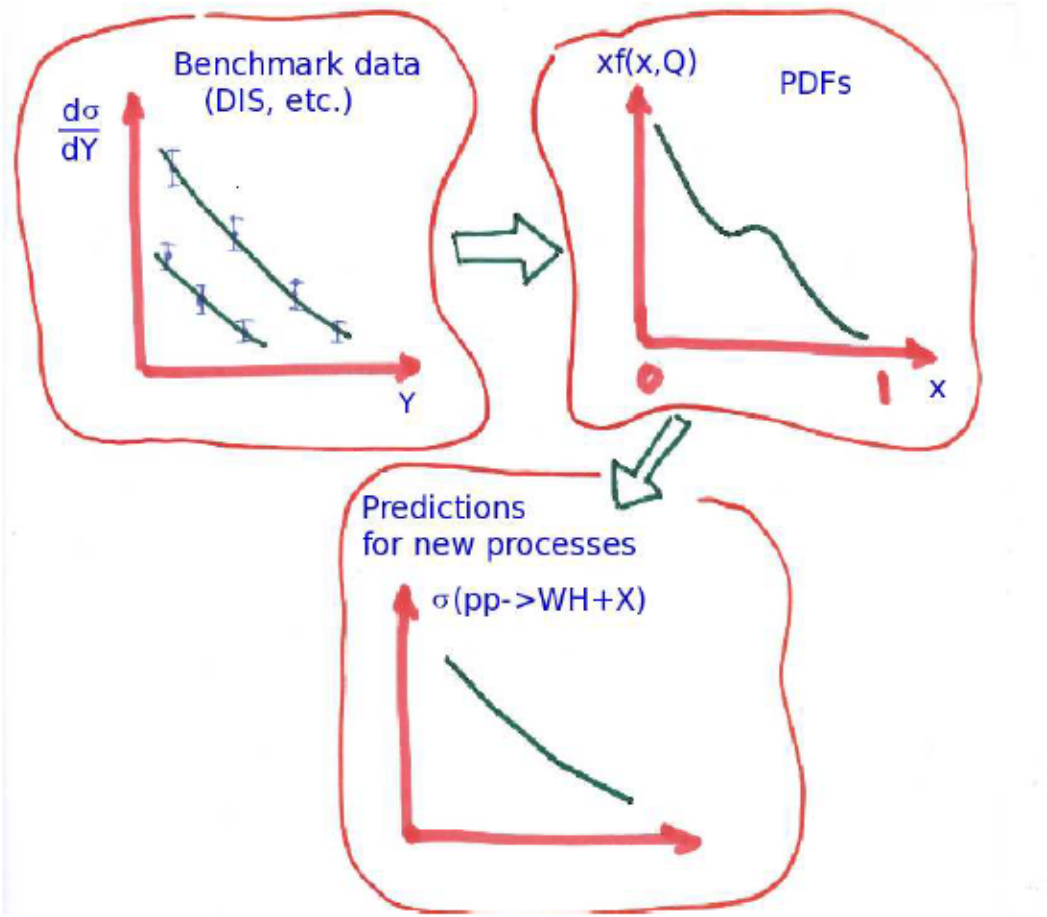


Figure by G. Salam

Since 2005, generalized unitarity and related methods dramatically advanced the computations of **perturbative** NLO/NNLO/N3LO hard cross sections $\hat{\sigma}$.

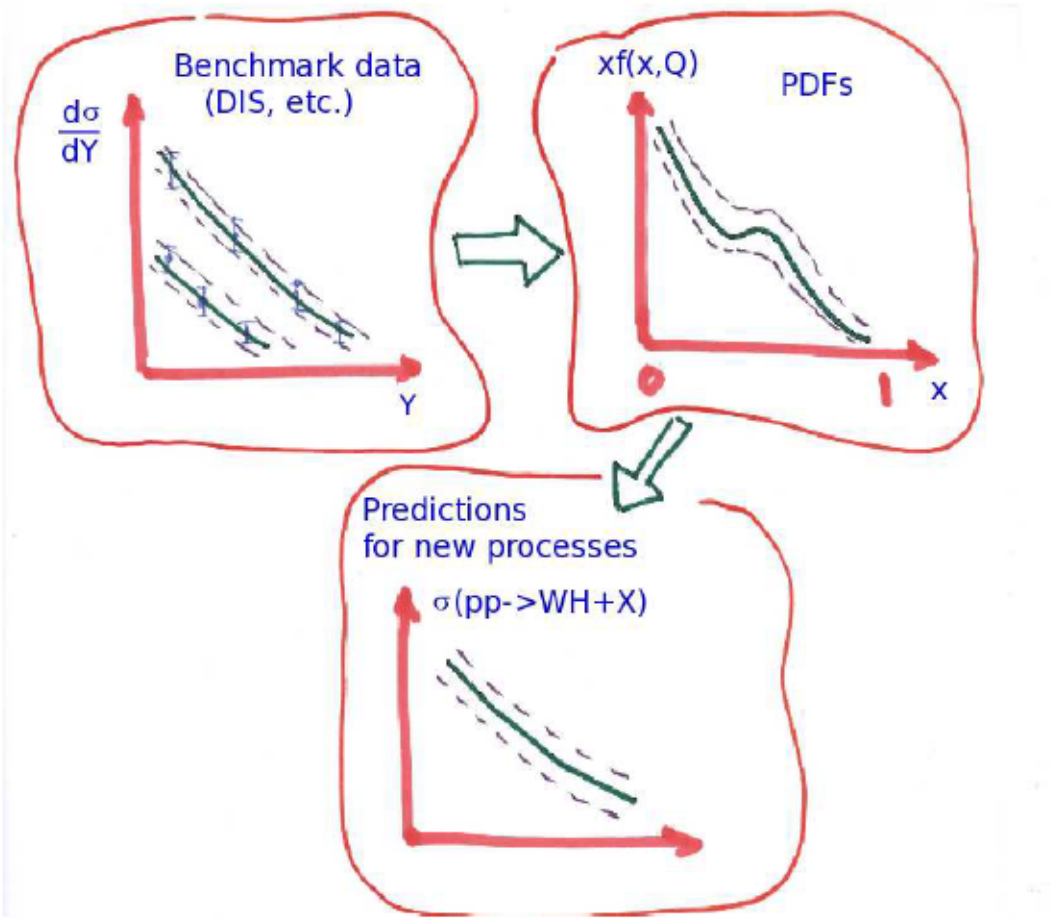
To make use of it, accuracy of PDFs $f_{a/p}(x,\mu)$ must keep up

The flow of the global analysis



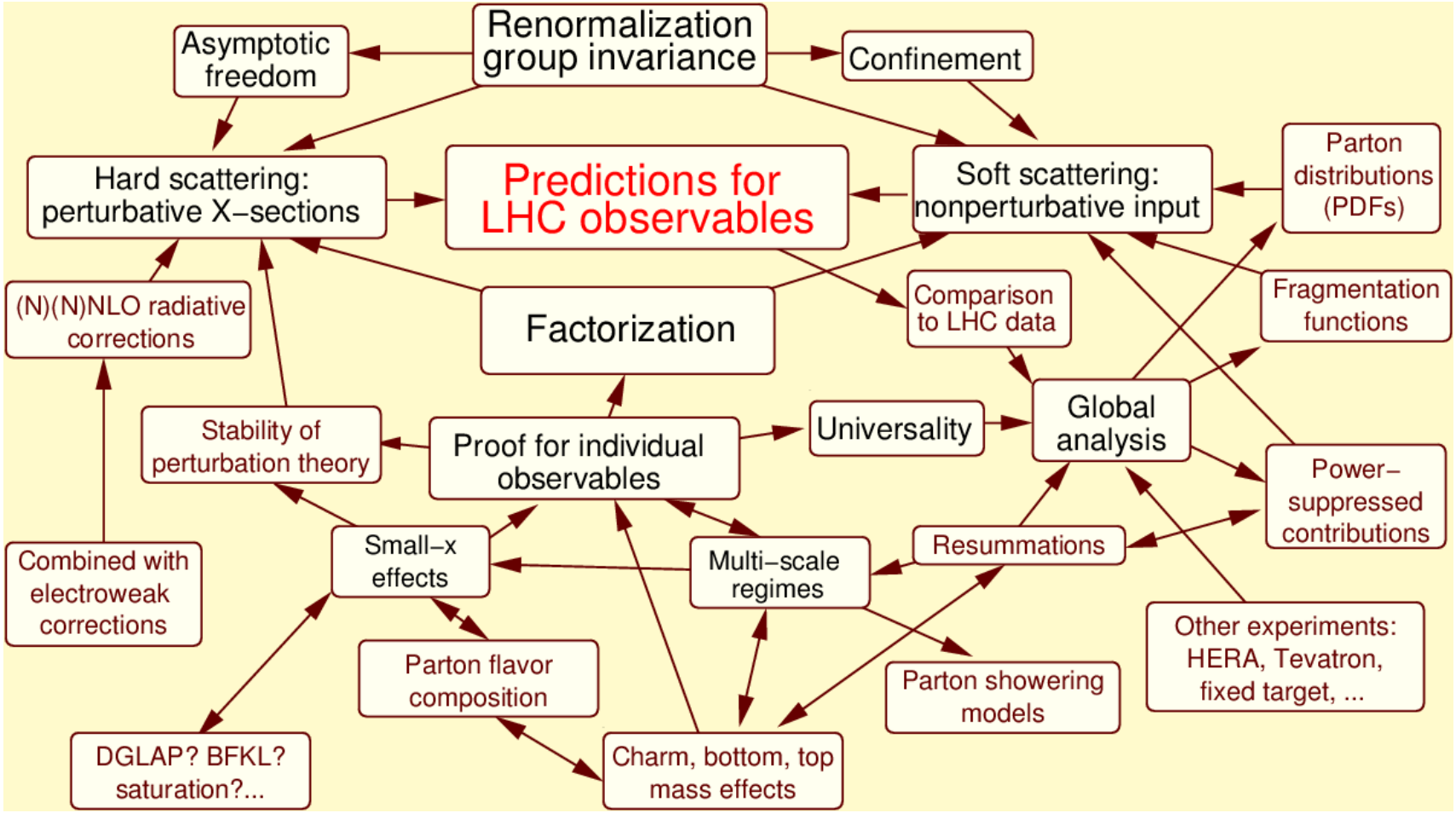
PDFs are not measured directly, but some data sets are sensitive to specific combinations of PDFs. By constraining these combinations, the PDFs can be disentangled in a combined (global) fit.

The flow of the global analysis



We are interested not just in one best fit, but also in the uncertainty of the resulting PDF parametrizations and theoretical predictions based on them. This will be covered in Lecture 2

Concepts of perturbative QCD at the LHC



At the (N)NNLO accuracy level, multiple aspects affect the PDF behavior

Classes of PDFs

General-purpose

For (N)NLO calculations with $N_f \leq 5$ active quark flavors

From several groups:
ABMP'16
CTEQ-Jlab (CJ'2015)
HERA2.0
CT14 (→ CT18)
MMHT'14
NNPDF3.1

Specialized

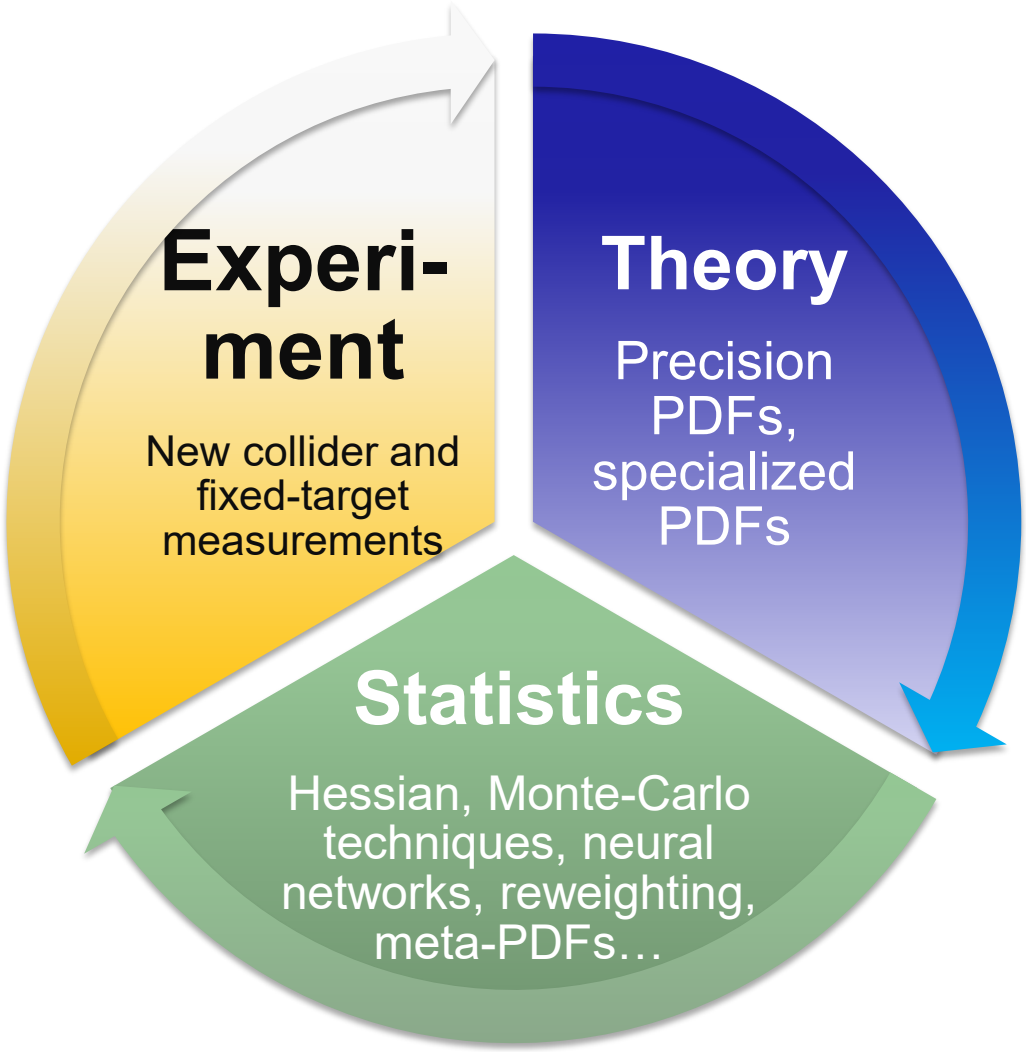
For instance, for CT14:
CT14 LO
CT14 $N_f = 3, 4, 6$
CT14 HERA2 **[arXiv:1609.07968]**
CT14 Intrinsic charm **[1707.00065]**
CT14 QCD+QED **[1509.02905]**
CT14 Monte-Carlo **[1607.06066]**

ATLAS & CMS exploratory

Combined **[1509.03865]**

PDF4LHC'15=CT14+MMHT'14+NNPDF3.0

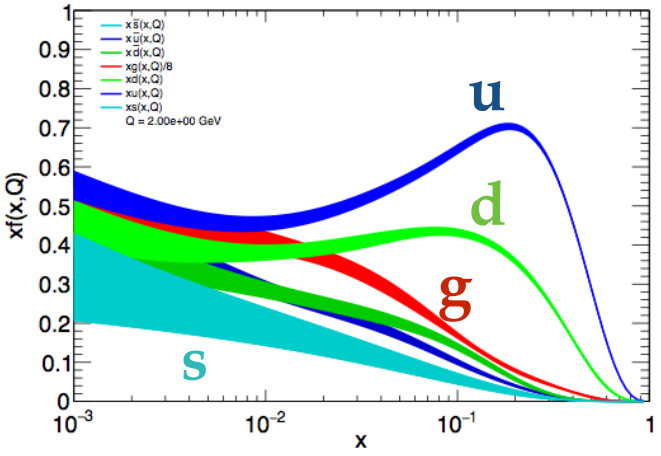
Frontiers of the PDF analysis



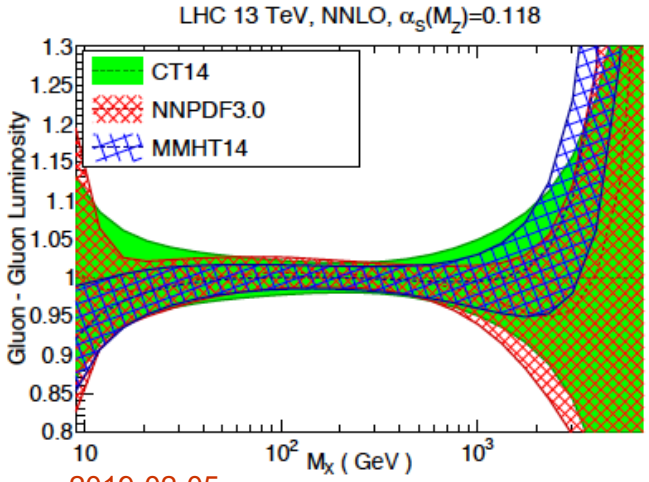
Significant advances on all frontiers will be necessary to meet the targets of the HL-LHC program

Previous generation: CT14 parton distributions

CT14 NNLO PDFs



gluon-gluon luminosity



- 2015 release of general-purpose PDFs, NNLO/NLO sets, alternative α_s series and $N_f = 3,4,6$ **[1506.07443];**
- update with HERA I+II DIS data **[1609.07968]**

ID#	Experimental dataset		N_d
101	BCDMS F_2^p	[47]	337
102	BCDMS F_2^d	[48]	250
104	NMC F_2^d/F_2^p	[49]	123
108	CDHSW F_2^p	[50]	85
109	CDHSW F_3^p	[50]	96
110	CCFR F_2^p	[51]	69
111	CCFR xF_3^p	[52]	86
124	NuTeV $\nu\mu\mu$ SIDIS	[40]	38
125	NuTeV $\bar{\nu}\mu\mu$ SIDIS	[40]	33
126	CCFR $\nu\mu\mu$ SIDIS	[41]	40
127	CCFR $\bar{\nu}\mu\mu$ SIDIS	[41]	38
145	H1 σ_r^b (57.4 pb^{-1})	[53][54]	10
147	Combined HERA charm production (1.504 fb^{-1})	[39]	47
160	HERA1+2 Combined NC and CC DIS (1 fb^{-1})	[6]	1120
169	H1 F_L (121.6 pb^{-1})	[55]	9

ID#	Experimental dataset		N_d
201	E605 DY	[56]	119
203	E866 DY, $\sigma_{pd}/(2\sigma_{pp})$	[57]	15
204	E866 DY, $Q^3 d^2\sigma_{pp}/(dQdx_F)$	[58]	184
225	CDF Run-1 $A_e(\eta^e)$ (110 pb^{-1})	[59]	11
227	CDF Run-2 $A_e(\eta^e)$ (170 pb^{-1})	[60]	11
234	DØ Run-2 $A_\mu(\eta^\mu)$ (0.3 fb^{-1})	[61]	9
240	LHCb 7 TeV W/Z muon forward-η Xsec (35 pb^{-1})	[62]	14
241	LHCb 7 TeV W $A_\mu(\eta^\mu)$ (35 pb^{-1})	[62]	5
260	DØ Run-2 Z $d\sigma/dy_Z$ (0.4 fb^{-1})	[63]	28
266	CMS 7 TeV $A_\mu(\eta)$ (4.7 fb^{-1})	[64]	11
267	CMS 7 TeV $A_e(\eta)$ (0.840 fb^{-1})	[65]	11
268	ATLAS 7 TeV W/Z Xsec, $A_\mu(\eta)$ (35 pb^{-1})	[66]	41
281	DØ Run-2 $A_e(\eta)$ (9.7 fb^{-1})	[67]	13
504	CDF Run-2 incl. jet $(d^2\sigma/dp_T^j dy_j)$ (1.13 fb^{-1})	[36]	72
514	DØ Run-2 incl. jet $(d^2\sigma/dp_T^j dy_j)$ (0.7 fb^{-1})	[37]	110
535	ATLAS 7 TeV incl. jet $(d^2\sigma/dp_T^j dy_j)$ (35 pb^{-1})	[68]	90
538	CMS 7 TeV incl. jet $(d^2\sigma/dp_T^j dy_j)$ (5 fb^{-1})	[69]	133

N_d is the number of data points

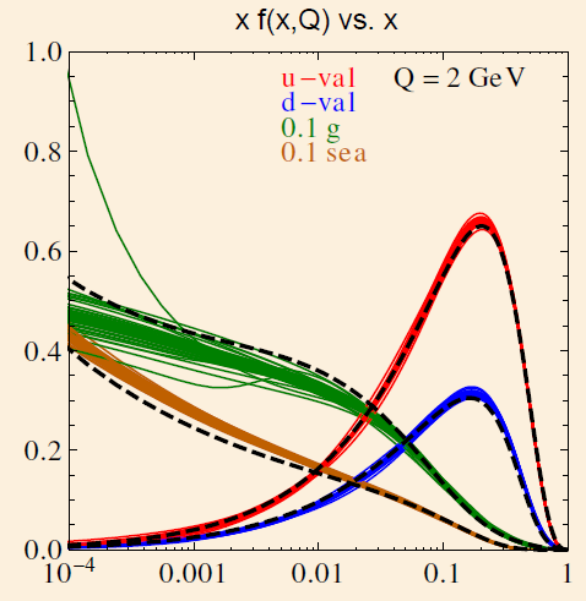
http://hep.pa.msu.edu/cteq/public/index.html

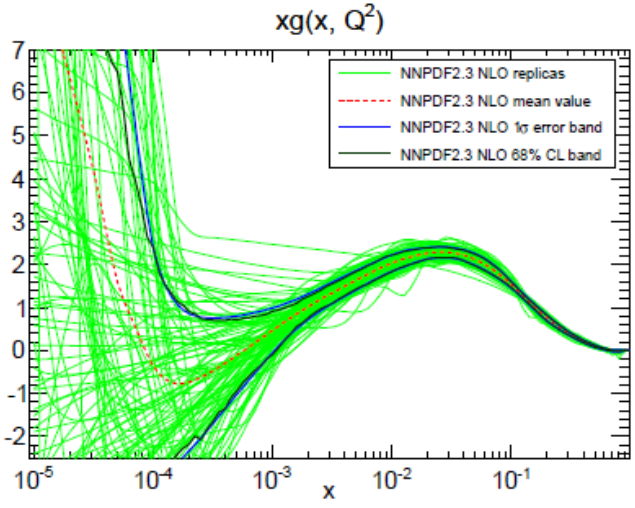
“Error sets” for computing PDF uncertainties

1. Based on diagonalization of the Hessian matrix
   - singular value decomposition of the covariance matrix in the Gaussian approximation
   - Default representation by CTEQ, MMHT, ABM, HERAPDF

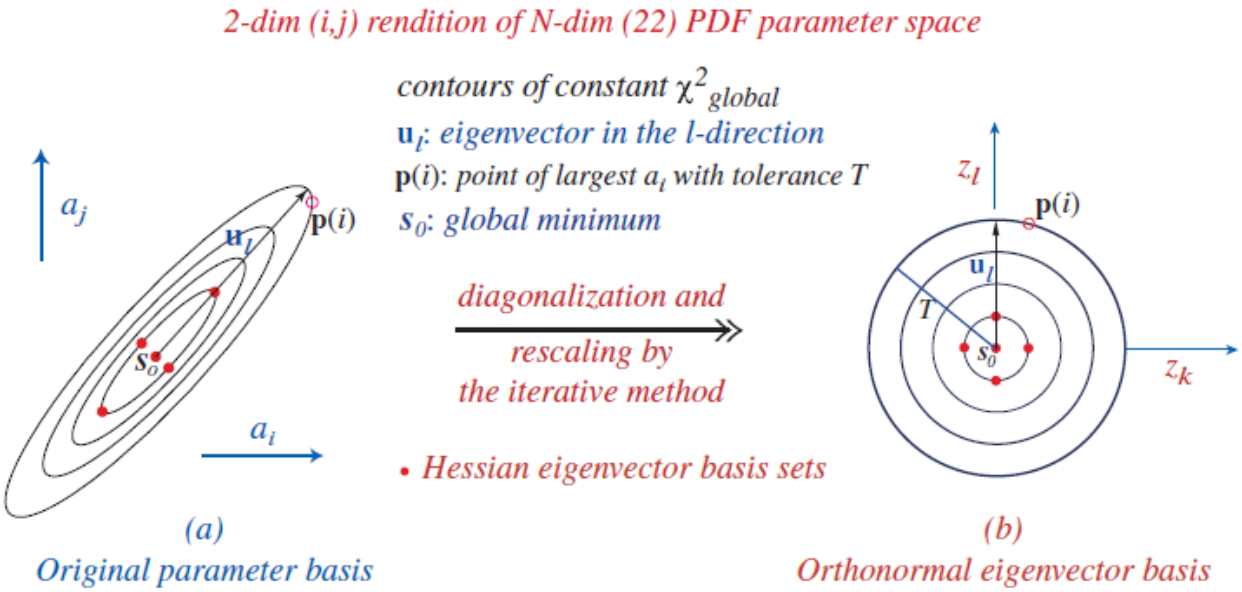
2. Based on Monte-Carlo sampling of probability
   - default representation by Neural Network PDF (NNPDF) collaboration

Available in the LHAPDF library





Tolerance hypersphere in the PDF space



A hyperellipse $\Delta\chi^2 \leq T^2$ in space of N physical PDF parameters $\{a_i\}$ is mapped onto a filled hypersphere of radius T in space of N orthonormal PDF parameters $\{z_i\}$

Tolerance hypersphere in the PDF space

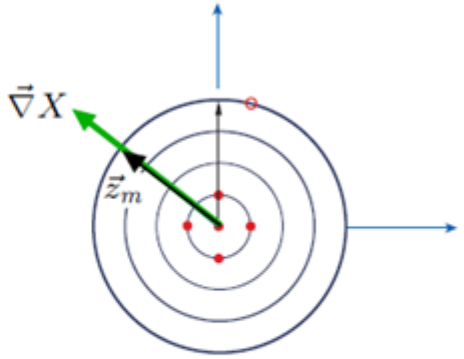
2-dim (i,j) rendition of N-dim (26) PDF parameter space

Hessian method: Pumplin et al., 2001

A symmetric PDF error for a physical observable X is given by

$$\Delta X = \vec{\nabla} X \cdot \vec{z}_m = \left| \vec{\nabla} X \right|$$

$$= \frac{1}{2} \sqrt{\sum_{i=1}^{N} \left(X_i^{(+)} - X_i^{(-)} \right)^2}$$

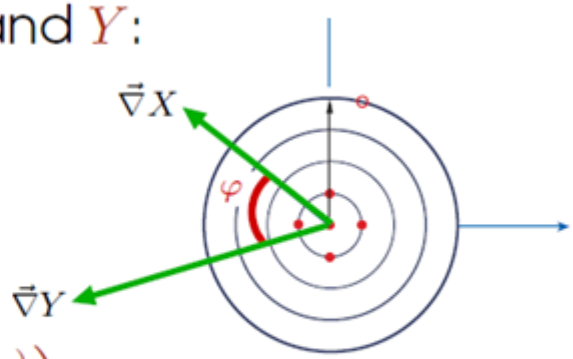


(b)
Orthonormal eigenvector basis

Correlation cosine for observables X and Y:

$$\cos\varphi = \frac{\vec{\nabla} X \cdot \vec{\nabla} Y}{\Delta X \Delta Y} =$$

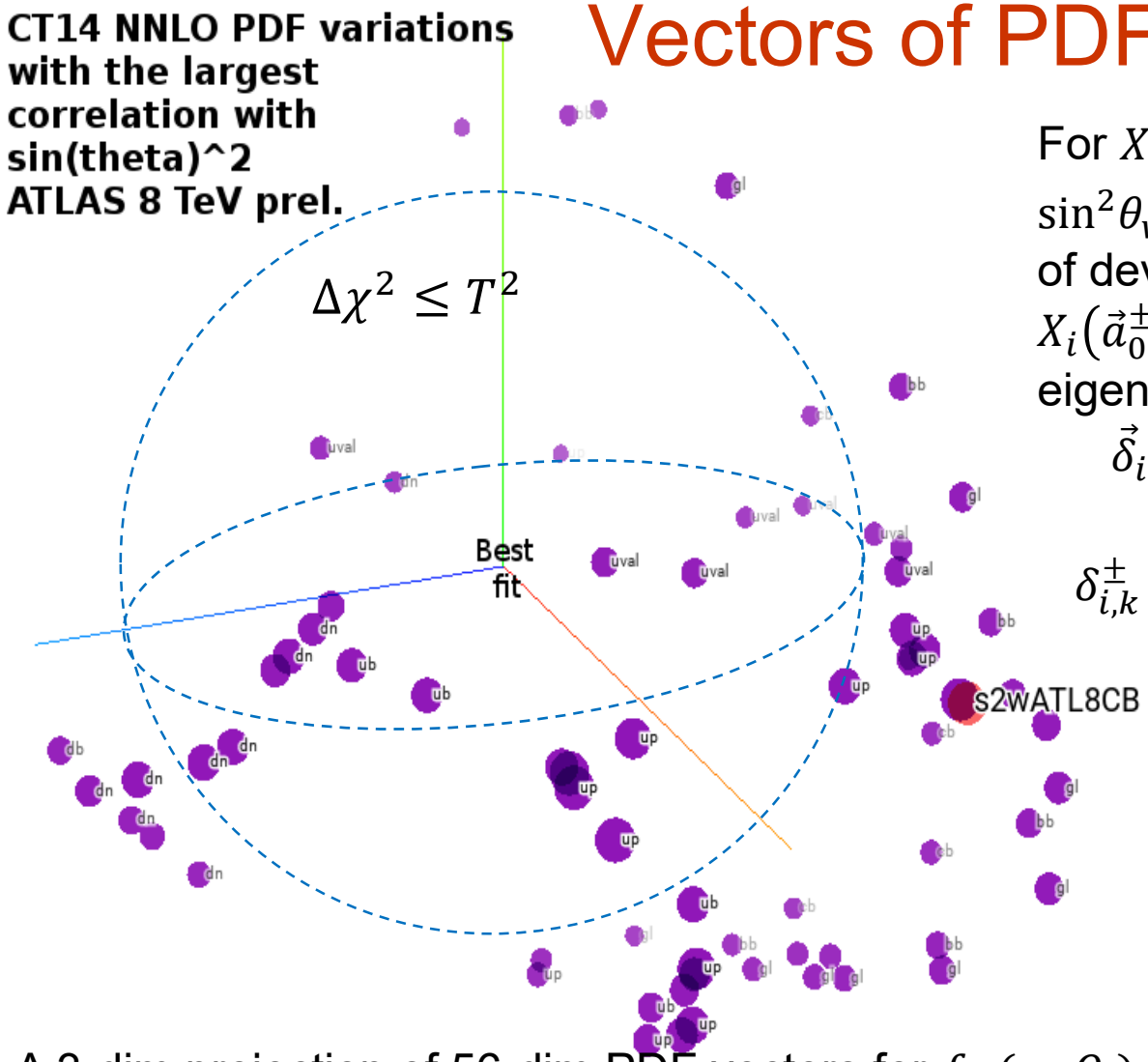
$$\frac{1}{4 \Delta X \Delta Y} \sum_{i=1}^{N} \left(X_i^{(+)} - X_i^{(-)} \right) \left(Y_i^{(+)} - Y_i^{(-)} \right)$$



(b)
Orthonormal eigenvector basis

Example: examine the PDF uncertainty of $\sin^2\theta_{weak} \equiv s2w$ measured by ATLAS 8 TeV

Vectors of PDF uncertainties



For $X_i(\vec{a}_i^{\pm}) \equiv f_{A_i}(x_i, Q_i; \vec{a}_i^{\pm})$ or $\sin^2\theta_w(\vec{a}_i^{\pm})$, construct a vector $\vec{\delta}_i$ of deviations from the best fit $X_i(\vec{a}_0^{\pm})$ for 2N Hessian eigenvectors.

$$\vec{\delta}_i = \{\delta_{i,1}^+, \delta_{i,1}^-, \ldots, \delta_{i,N}^+, \delta_{i,N}^-\}$$

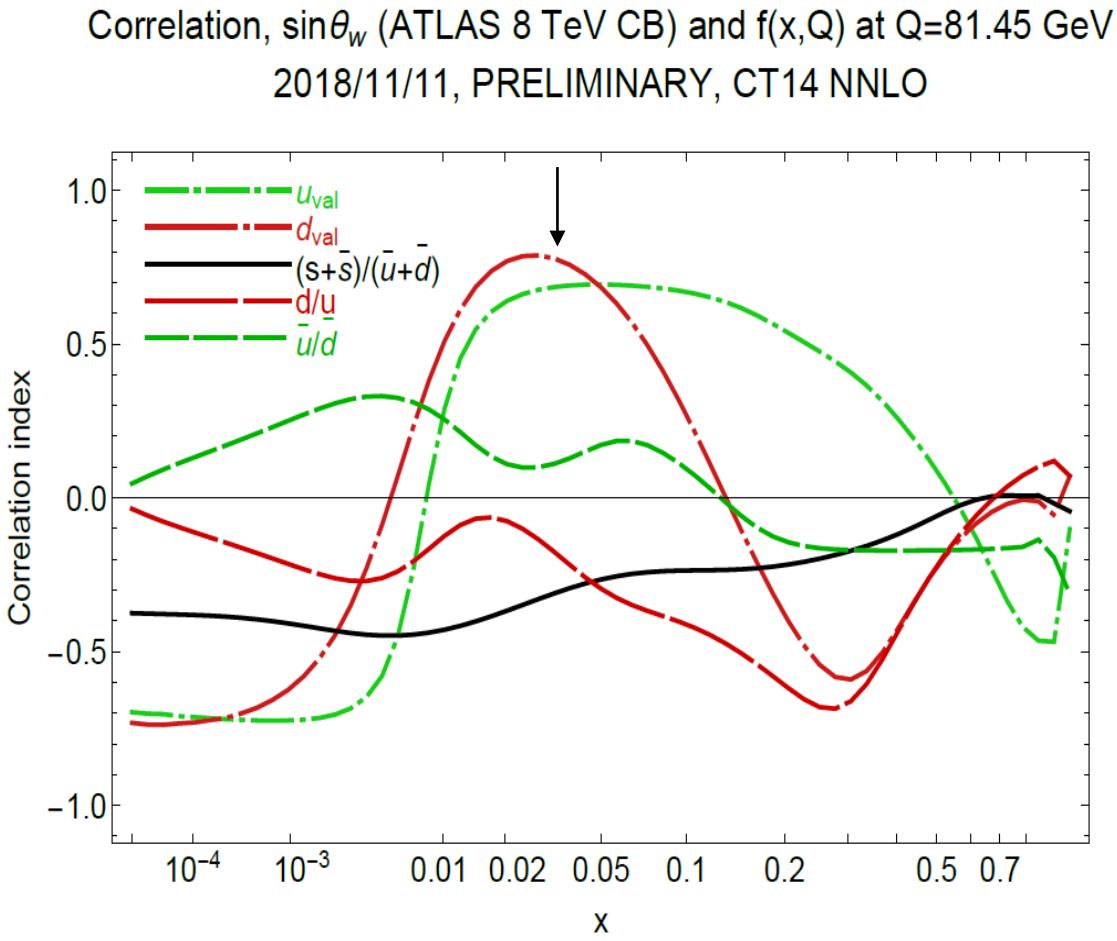
$$[\mathrm{N} = 28 \text{ for CT14 NNLO}]$$

$$\delta_{i,k}^{\pm} \equiv \left(X_i(\vec{a}_k^{\pm}) - X_i(\vec{a}_0)\right)/X_i(\vec{a}_0)$$

A 3-dim projection of 56-dim PDF vectors for $f_{A_i}(x_i, Q_i)$ with the smallest angular distance from the $\sin^2\theta_w(\vec{a}_i^{\pm})$ vector; $10^{-5} \leq x_i \leq 0.8$; $Q_i = 100$ GeV

Hessian correlation for $\sin^2\theta_w$ at 8 TeV

Presented at the EW precision subgroup meeting, Nov. 13, 2018



Strongest correlations of s2w with u_{val}, d_{val} at $x = 0.01 - 0.2$

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

PDFSense program: fast surveys of QCD data using a vector data technique

Estimates the sensitivity variable S_f ("correlation 2.0"): an easy-to-compute indicator of data point sensitivity to PDFs in the presence of experimental errors

References

1. Mapping the sensitivity of hadronic experiments to nucleon structure
B.-T. Wang, T.J. Hobbs, S. Doyle, J. Gao, T.-J. Hou, P. M. Nadolsky, F. I. Olness
Phys.Rev. D98 (2018) 094030

2. The Coming synergy between lattice QCD and high-energy phenomenology
T.J. Hobbs, Bo-Ting Wang, Pavel Nadolsky, Fredrick Olness
Preprint SMU-HEP-19-02 [available at **https://tinyurl.com/SMUpreprints]**

3. Sensitivity of future lepton-hadron experiments to nucleon structure
Paper in preparation

Vectors of data residuals

For every data point i, construct a vector of residuals $r_i(\vec{a}_k^{\pm})$ for 2N Hessian eigenvectors. $\mathrm{k} = 1, \ldots, N$, with $N = 28$ for CT14 NNLO:

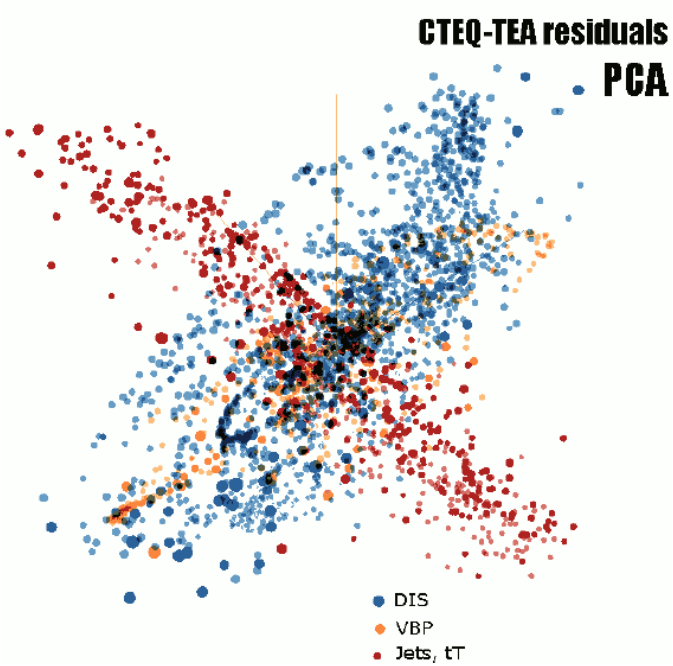
$$\vec{\delta}_i = \{\delta_{i,1}^+, \delta_{i,1}^-, \ldots, \delta_{i,N}^+, \delta_{i,N}^-\} \quad [\mathrm{N} = 28]$$
$$\delta_{i,k}^{\pm} \equiv \left(r_i(\vec{a}_k^{\pm}) - r_i(\vec{a}_0)\right) / \langle r_0 \rangle_E$$

-- a 56-dim vector normalized to $\langle r_0 \rangle_E$, the root-mean-squared residual for the experiment E for the central fit $\vec{a}_0$

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

$\langle r_0 \rangle_E \approx 1$ in a good fit to E

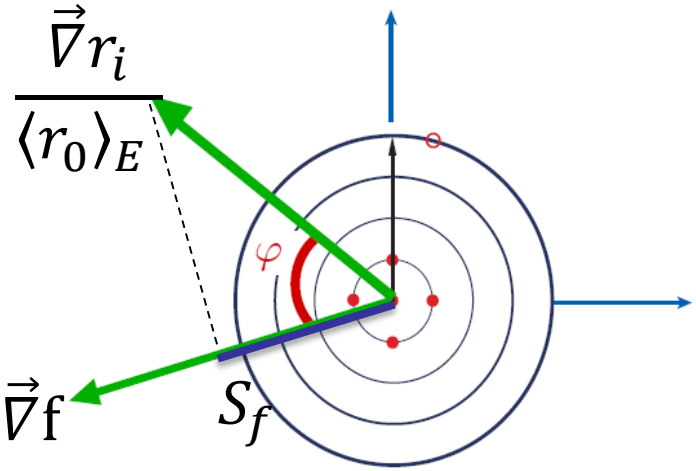
r_i is defined in the backup



The TensorFlow Embedding Projector (http://projector.tensorflow.org) represents CT14HERA2 $\vec{\delta}_i$ vectors by their 10 principal components indicated by scatter points. A sample 3-dim. projection of the 56-dim. manifold is shown above. A symmetric 28-dim. representation can be alternatively used.

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;

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

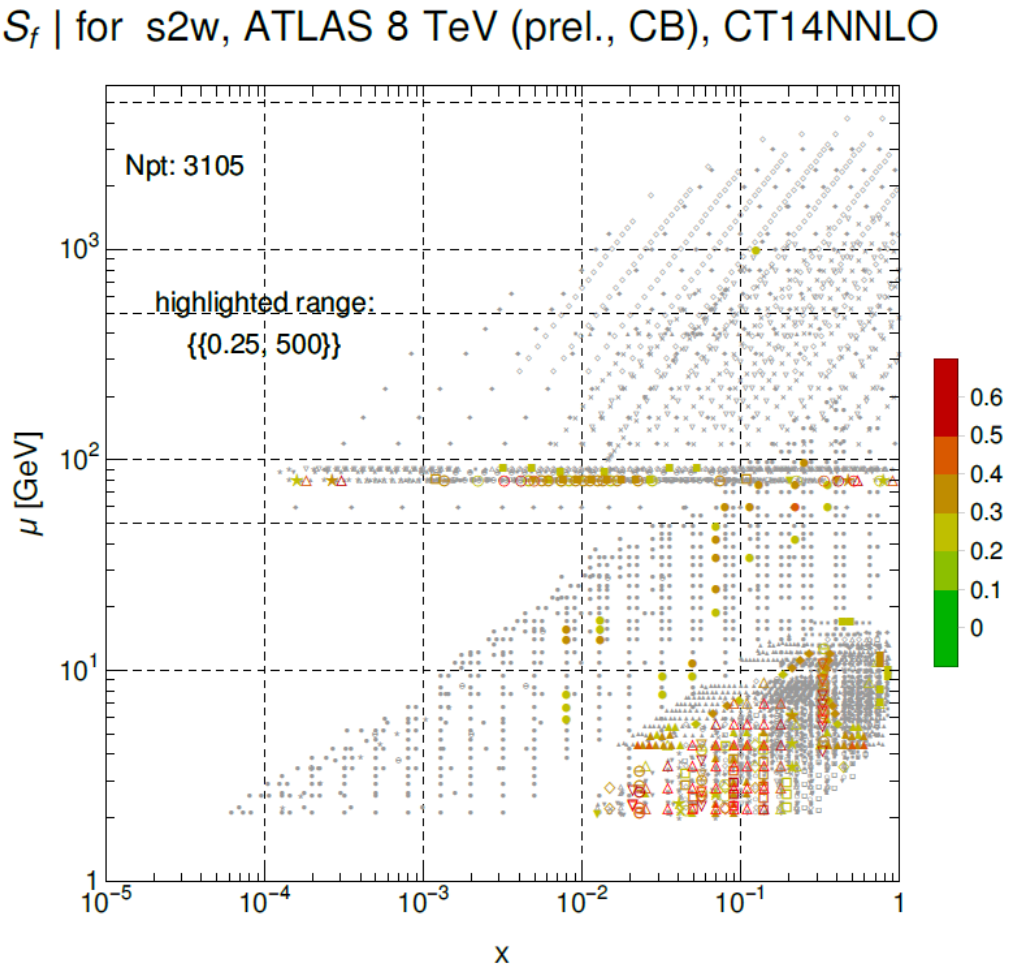
C_f is **independent** of the experimental and PDF uncertainties. In the figures, take $|C_f| \gtrsim 0.7$ to indicate a large correlation.

- $S_f \equiv |\vec{\rho}_i| cos\varphi = C_f \dfrac{\Delta r_i}{\langle r_0 \rangle_E}$ -- 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.

Sensitivity of CT14 experiments to $\sin^2\theta_{weak}$

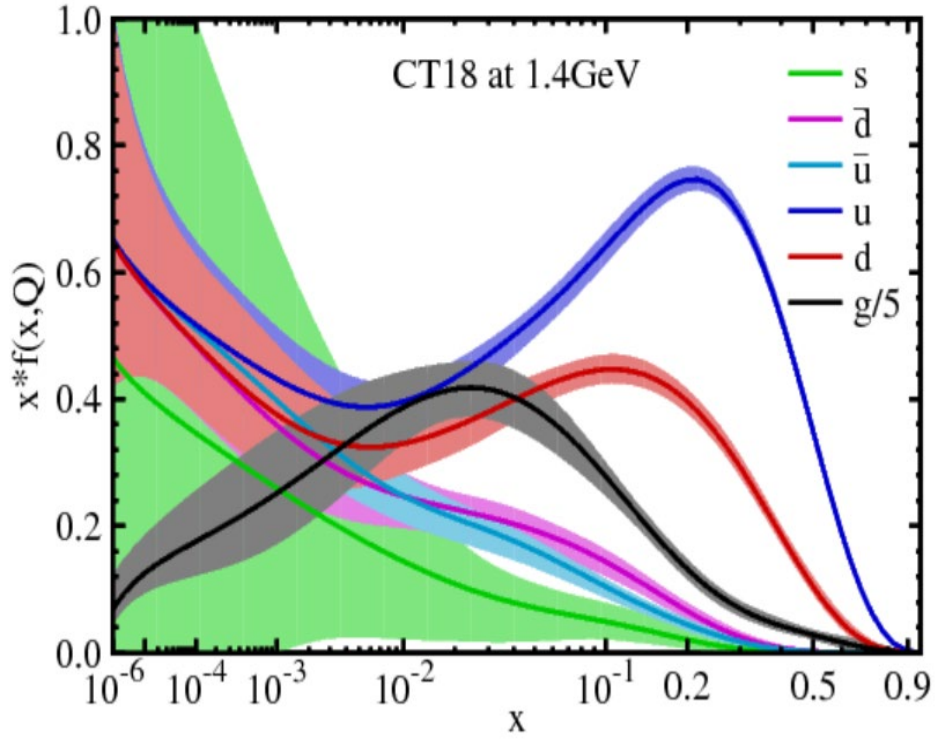


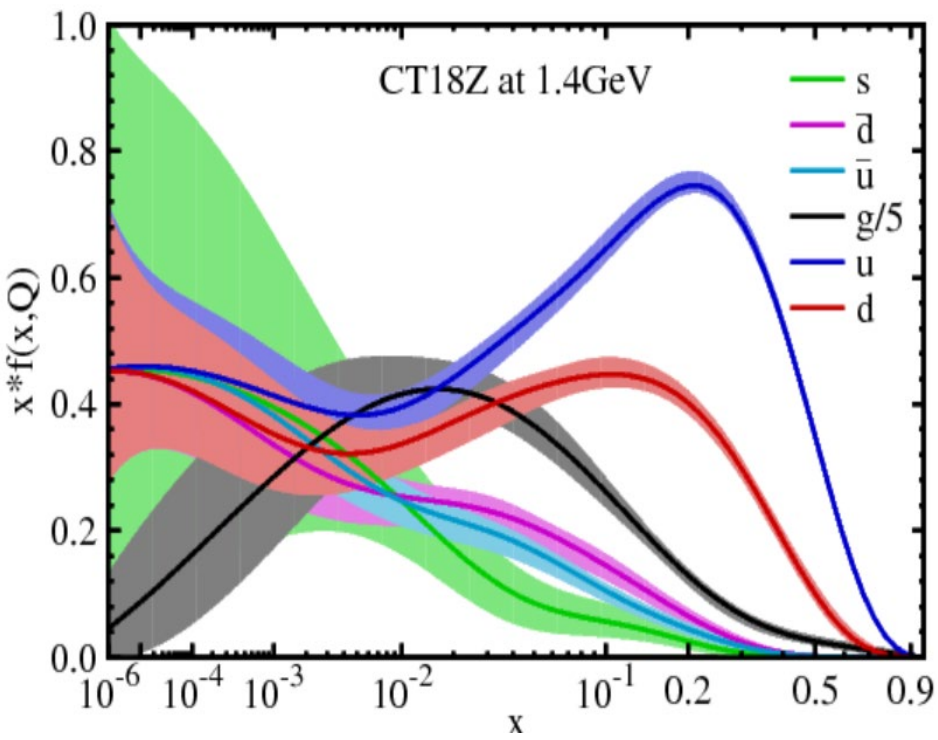
Based on the **PDFSense** [arXiv:1803.02777] analysis, the most sensitive CT14 data sets to $\sin^2\theta_{weak}$ ≡s2w measured by ATLAS are

- **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})
- …

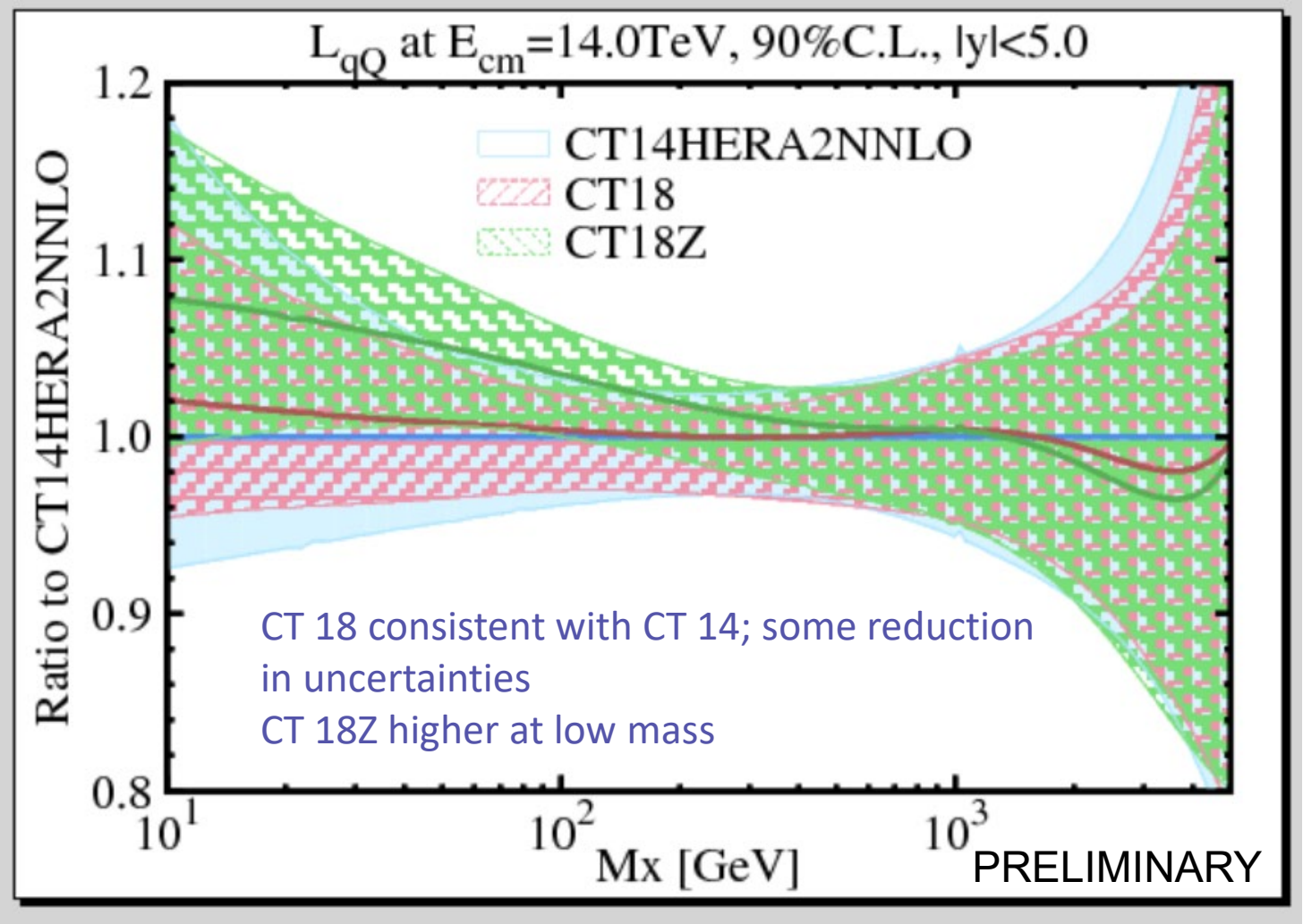
Toward a new generation of PDFs

[CT18/CT18Z PDFs]

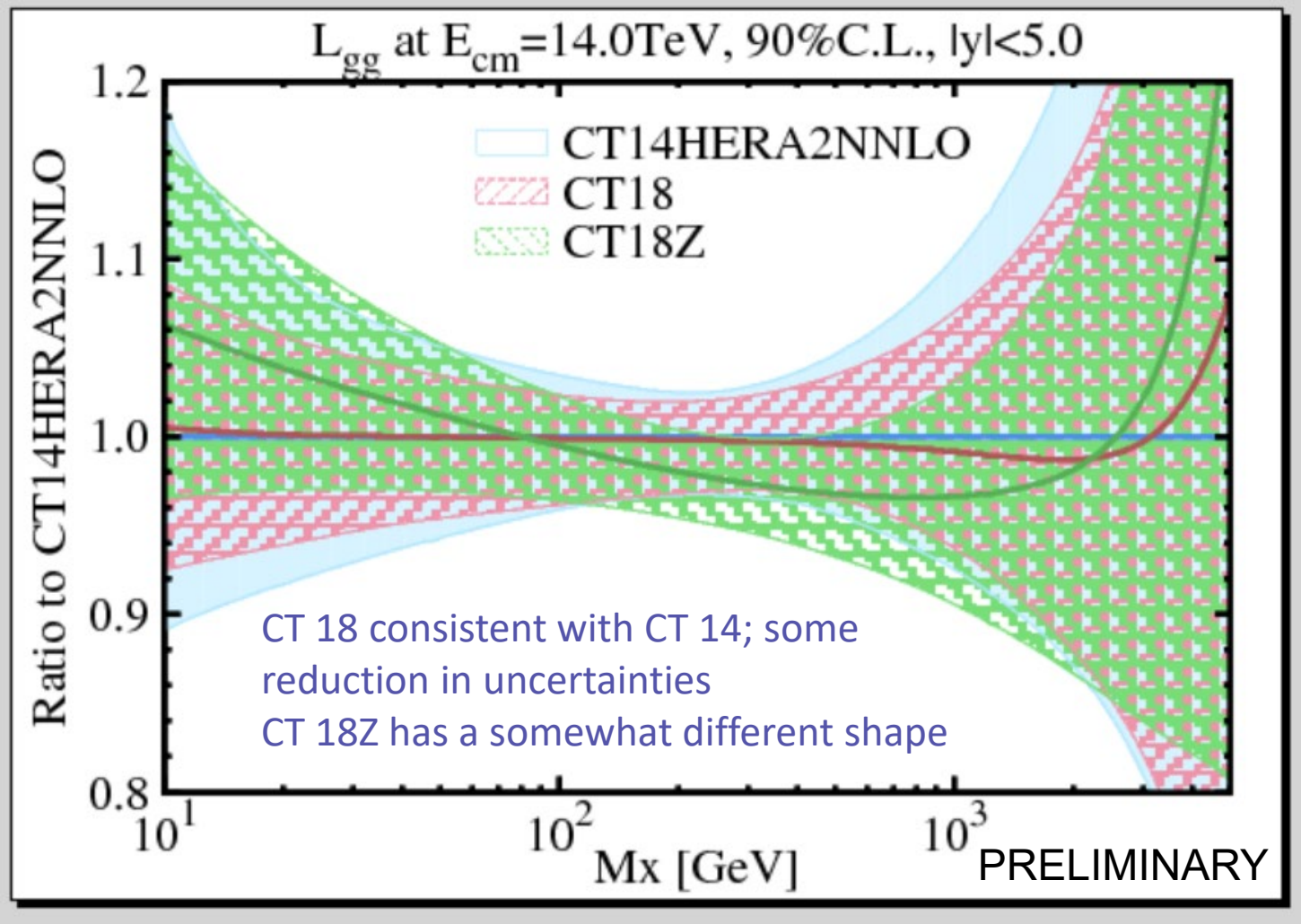




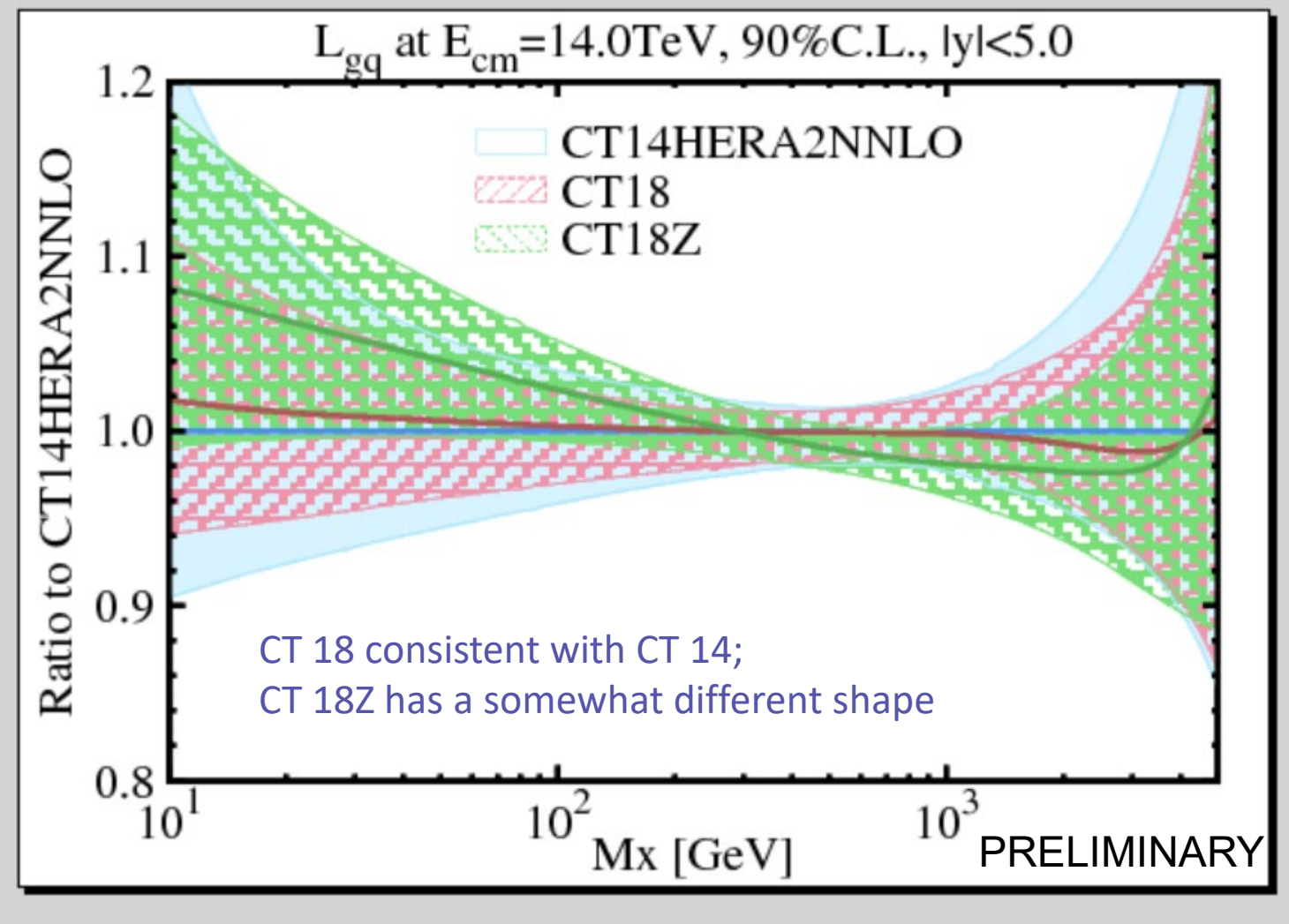
Quark-antiquark luminosities



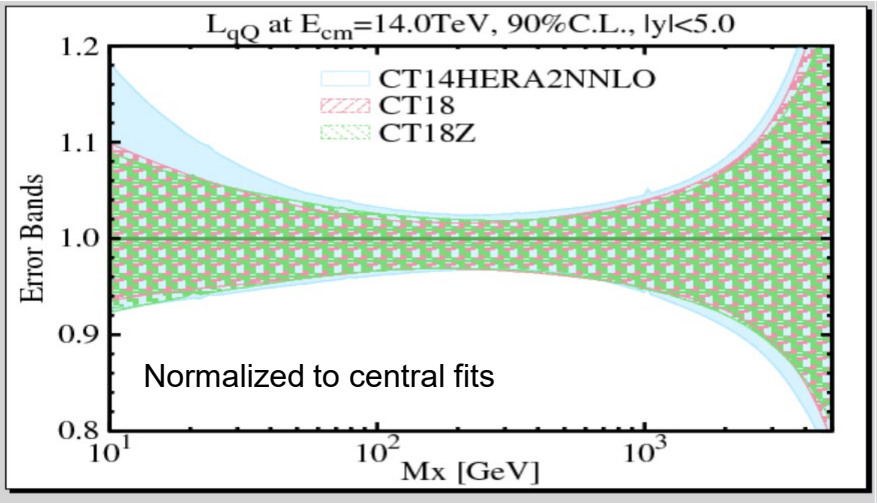
Gluon-gluon luminosities

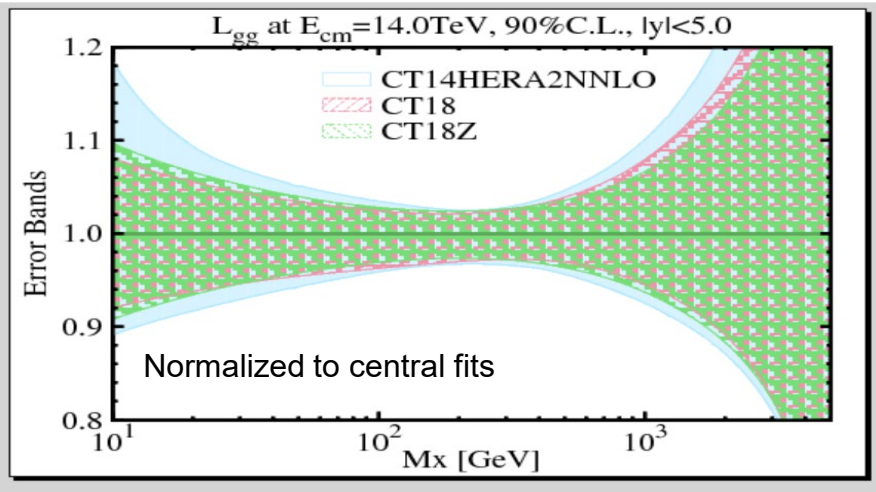


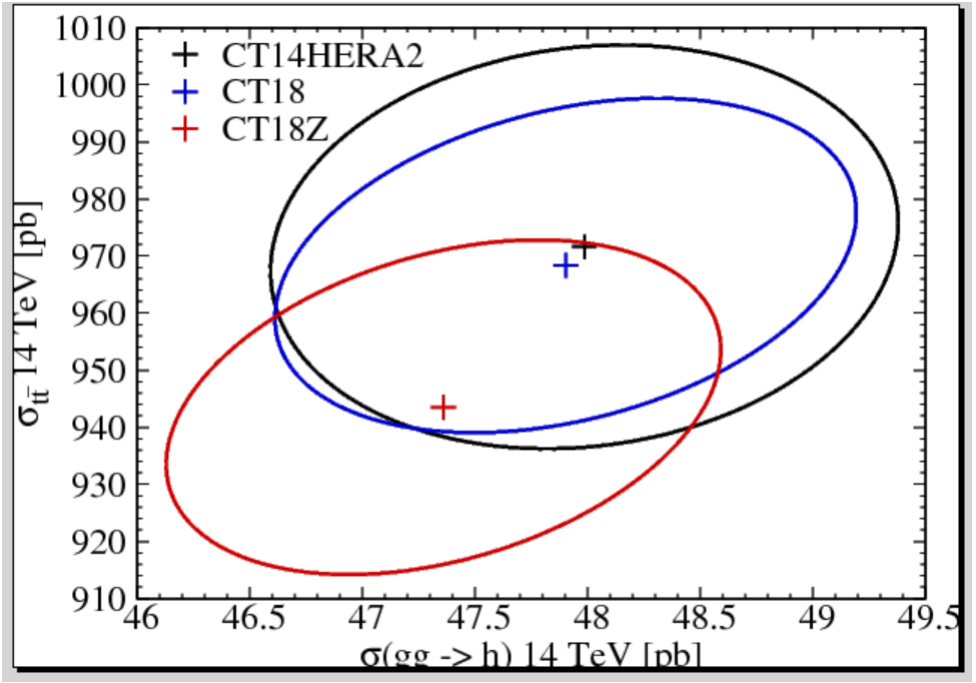
Gluon-quark luminosities



Mild reduction in nominal PDF error bands and cross section uncertainties



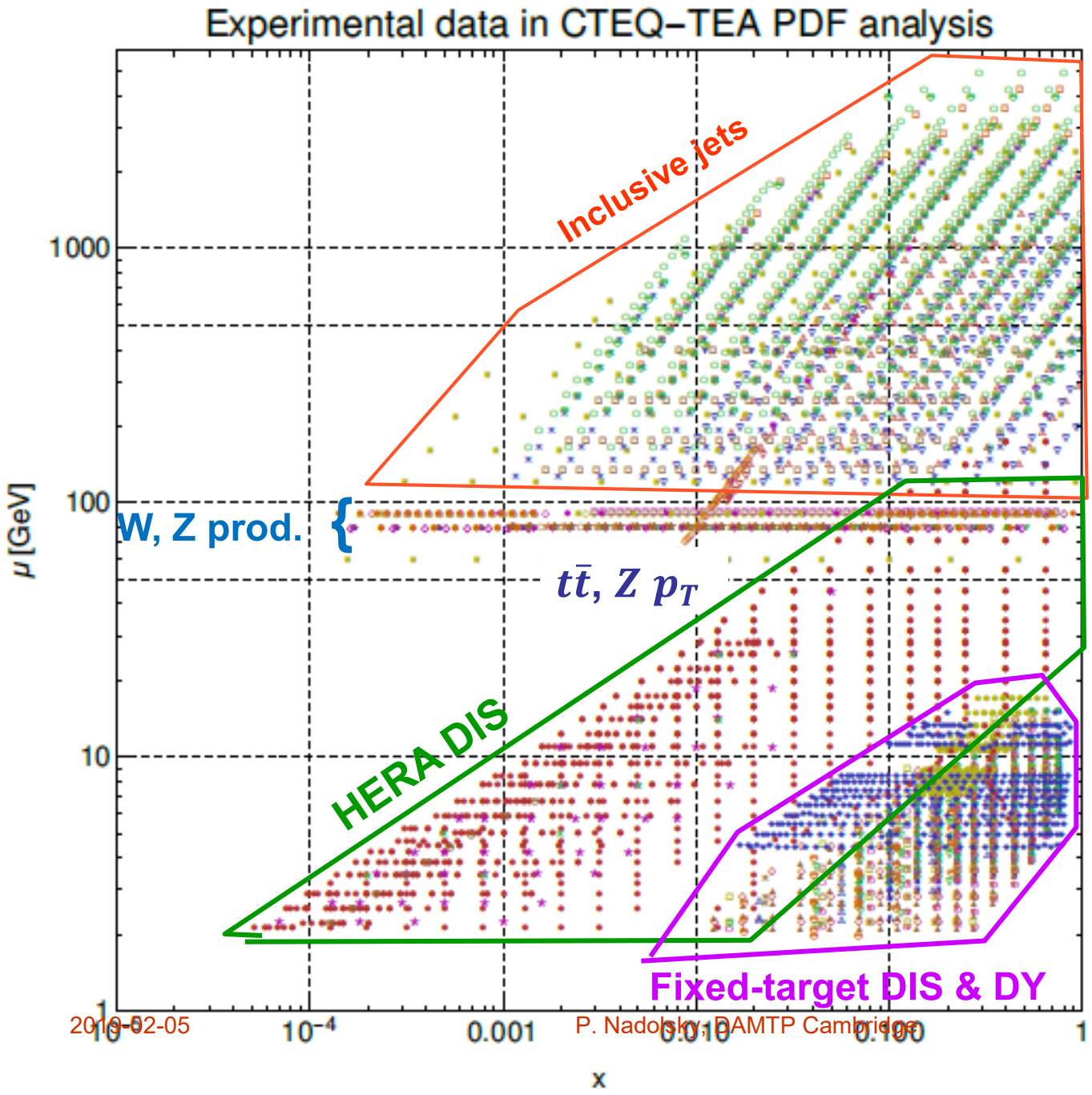




PRELIMINARY

CT18 in a nutshell

- Start with CT14-HERA2 (HERAI+II combined data released after publication of CT14)
- Examine a wide range of PDF parameterizations
- Use as much relevant LHC data as possible using applgrid/fastNLO interfaces to data sets, with NNLO/NLO K-factors, or fastNNLO tables in the case of top pair production
- **PDFSense** (Hobbs, Wang, et al., arXiv:1803.02777) to determine quantitatively which data will have impact on global PDF fit
- **ePump** (Schmidt, Pumplin, Yuan, arXiv:1806.07950) to quickly explore the impact of data prior to global fit using the Hessian reweighting
  - good agreement between PDFSense, ePump results, and global fit
- Implement a parallelization of the global PDF fitting to allow for faster turn-around time
- Lagrange Multiplier studies to examine constraints of specific data sets on PDF distributions, and (in some cases) the tensions



CT18 analysis includes **new** LHC experiments on W/Z, high-p_T Z, jet, $t\bar{t}$ production

Up to 30 candidate LHC data sets available

The challenge is to select and implement relevant and consistent experiments

CT18: advancements in theoretical and statistical methodology

- In-house development of fast ApplGrid/FastNLO calculations
- Parallelization of CTEQ fitting code
- Studies of QCD scale dependence and other theory uncertainties for DIS, high-p_T Z, jet production
- Studies of PDF functional forms

Theory input

Obs.	Expt.	fast table	NLO code	K-factors	R,F scales
Inclusive jet	ATL 7 CMS 7/8	APPLgrid fastNLO	NLOJet++	NNLOJet	p_T, p_T^1
p_T^Z	ATL 8	APPLgrid	MCFM	NNLOJet	$\sqrt{Q^2 + p_{T,Z}^2}$
W/Z rapidity W asymmetry	LHCb 7/8 ATL 7 CMS 8	APPLgrid	MCFM/aMCfast	FEWZ/MCFM	$M_{W,Z}$
DY (low,high mass)	ATL 7/8 CMS 8	APPLgrid	MCFM/aMCfast	FEWZ/MCFM	Q_{ll}
$t\bar{t}$	ATL 8 CMS 8	fastNNLO			$\frac{H_T}{4}$, $\frac{m_T}{2}$

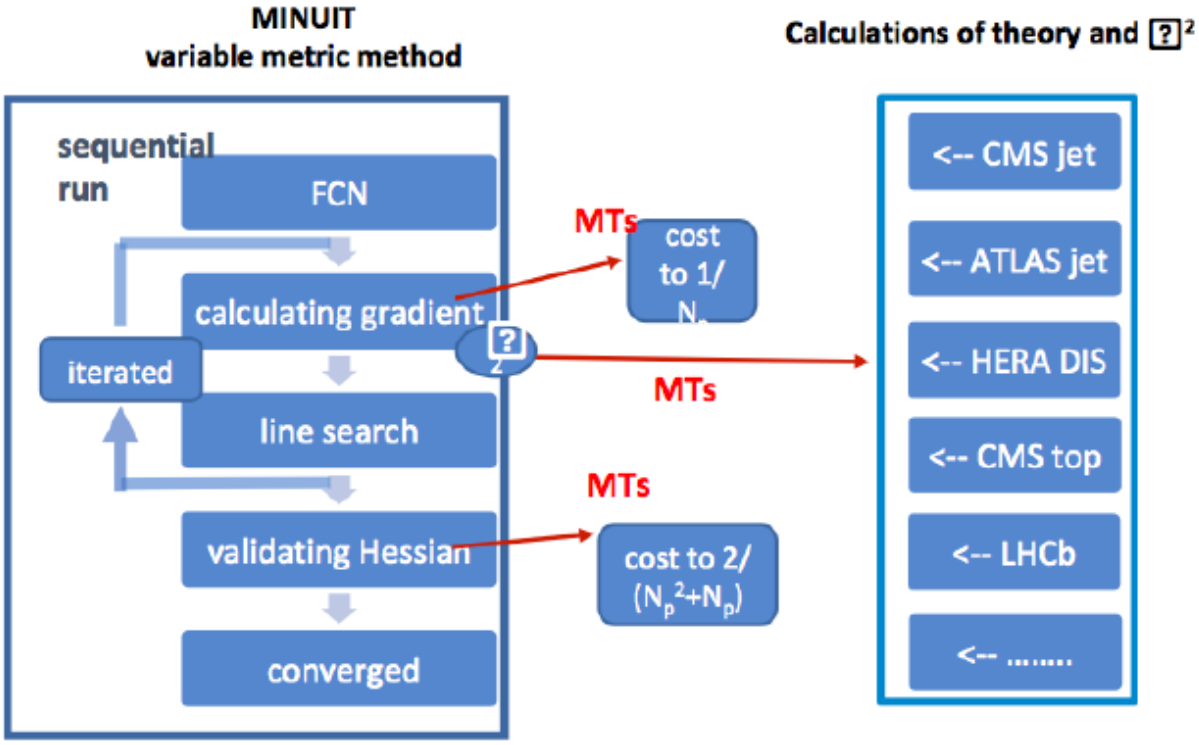
when justified, a small Monte-Carlo error (typically 0.5%) added for NNLO/NLO K-factors

Theoretical calculations for vector boson production

ID	Obs.	Expt.	fast table	NLO code	K-factors	$\mu_{R,F}$
245 246 250 249	$y_{\mu\mu}, \eta_\mu$ y_{ee} $y_{\mu\mu}, \eta_\mu$ $A(\mu)$	LHCb7ZW LHCb8Z LHC8ZW CMS8W	APPLgrid	MCFM/aMCfast	MCFM/FEWZ	$M_{Z,W}$
253	p_T^{ll}	ATL8Z	APPLgrid	MCFM	NNLOJet	M_T^{ll}
201 203 204	$\sqrt{\tau}, y$ $\sigma_{pd}/\sigma_{pp}, x_F$ Q, x_F	E605 E866 E866	CTEQ		FEWZ	Q_{ll}
225 227 234 281	$A(e)$ $A(e)$ $A(\mu)$ $A(e)$	CDF1Z CDF2W D02W D02W	CTEQ		ResBos	Q_{ll} M_W
260 261	y_{ll} y_{ll}	D02 CDF2	CTEQ		VRAP	Q_{ll}
266 267 268	$A(\mu)$ $A(e)$ $y_{ll}, \eta_l, A(l)$	CMS7W CMS7W ATL7ZW(2012)	CTEQ		ResBos	M_W $M_{Z,W}$
248	y_{ll}, η_l	ATL7ZW(2016)	APPLgrid	MCFM/aMCfast	MCFM/FEWZ	$M_{Z,W}$

Fitting code parallelization with multi-threads

upgrade to a parallelized version of the fitting code, two-layer parallelization: 1. through rearrangement of the minimization algorithm; 2. via redistribution of the data sets



Layer 1: after all a factor of 4~5 improvement on speed is achieved!

Layer 2: further speed up by a factor of 2, depending on data sets included

Functional forms of PDFs

Evolving PDF models

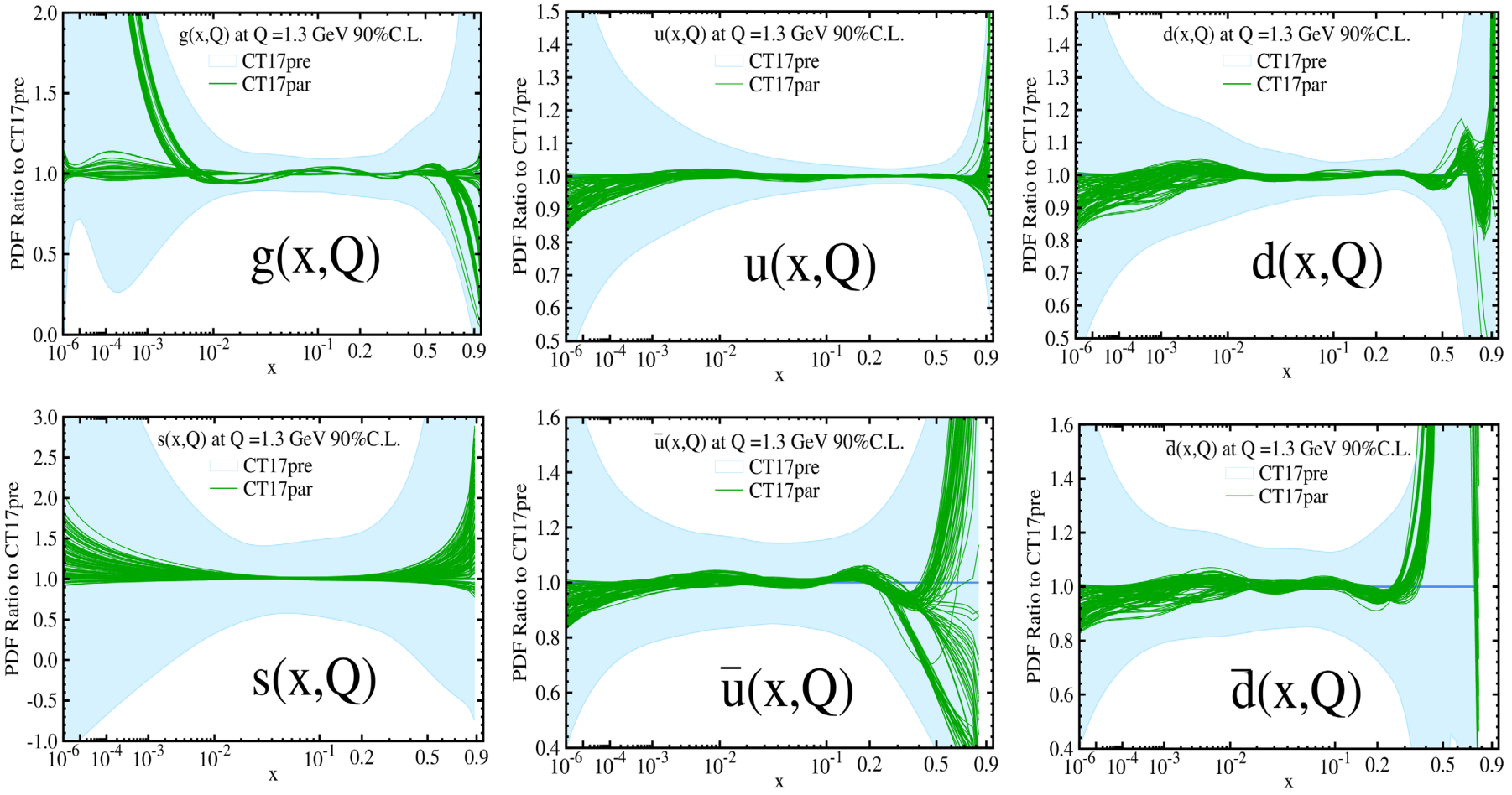
- EW precision fits and PDF fits are fundamentally different.
  - In an EW fit ("ZFITTER program"), the Standard Model parameters are found by fitting a **fixed** theoretical model.
  - In a PDF fit ("XFITTER program"), the theoretical model (PDF parametrization) **evolves** when more data are added.

⇒ A PDF model can change its functional form within some limits to evade falsification by a new data set

- The uncertainty due to the PDF functional form contributes as much as 50% of the total PDF uncertainty in CT fits. The CT18 analysis estimates this uncertainty using 100 trial functional forms. This part of analysis requires significant human intervention.

Carefully crafted PDF functional forms with >20-30 free parameters

Explore various non-perturbative parametrization forms of PDFs



- CT17par – sample result of using various non-perturbative parametrization forms.
- No data constrain very large x or very small x regions.

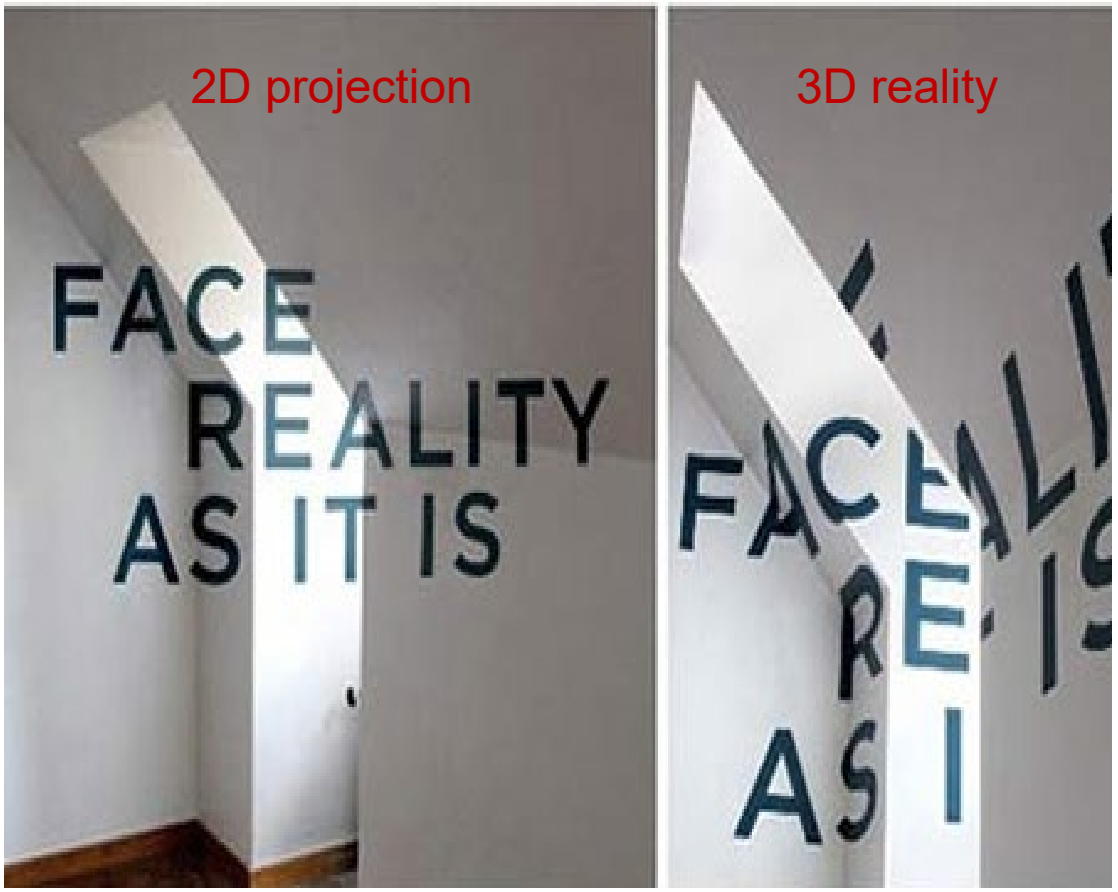
CT14: parametrization forms

- CT14 relaxes restrictions on several PDF combinations that were enforced in CT10. [These combinations were not constrained by the pre-LHC data.]
  - The assumptions $\frac{\bar{d}(x,Q_0)}{\bar{u}(x,Q_0)} \to 1$, $u_v(x,Q_0) \sim d_v(x,Q_0) \propto x^{A_{1v}}$ with $A_{1v} \approx -\frac{1}{2}$ at $x < 10^{-3}$ are relaxed once LHC W/Z data are included
  - CT14 parametrization for $s(x,Q)$ includes extra parameters
- Candidate CT14 fits have 30-35 free parameters
- In general, $\mathrm{f_a}\,(\mathrm{x},\mathrm{Q_0}) = \mathrm{A}x^{a_1}(1-\mathrm{x})^{\mathrm{a_2}}\mathrm{P_a}(\mathrm{x})$
- CT10 assumed $P_a(x) = \exp(a_0 + a_3\sqrt{x} + a_4 x + a_5\, x^2)$
  - exponential form conveniently enforces positive definite behavior
  - but power law behaviors from a_1 and a_2 may not dominate
- In CT14, $\mathrm{P_a(x)} = \mathrm{G_a(x)F_a(z)}$, where $\mathrm{G_a(x)}$ is a smooth factor
  - $\mathrm{z} = 1 - 1(1-\sqrt{\mathrm{x}})^{\mathrm{a_3}}$ preserves desired Regge-like behavior at low x and high x (with a_3>0)
- Express $F_a(z)$ as a linear combination of Bernstein polynomials:

$$z^4, 4z^3(1-z), 6z^2(1-z)^2\,, 4z(1-z)^3, (1-z)^4$$

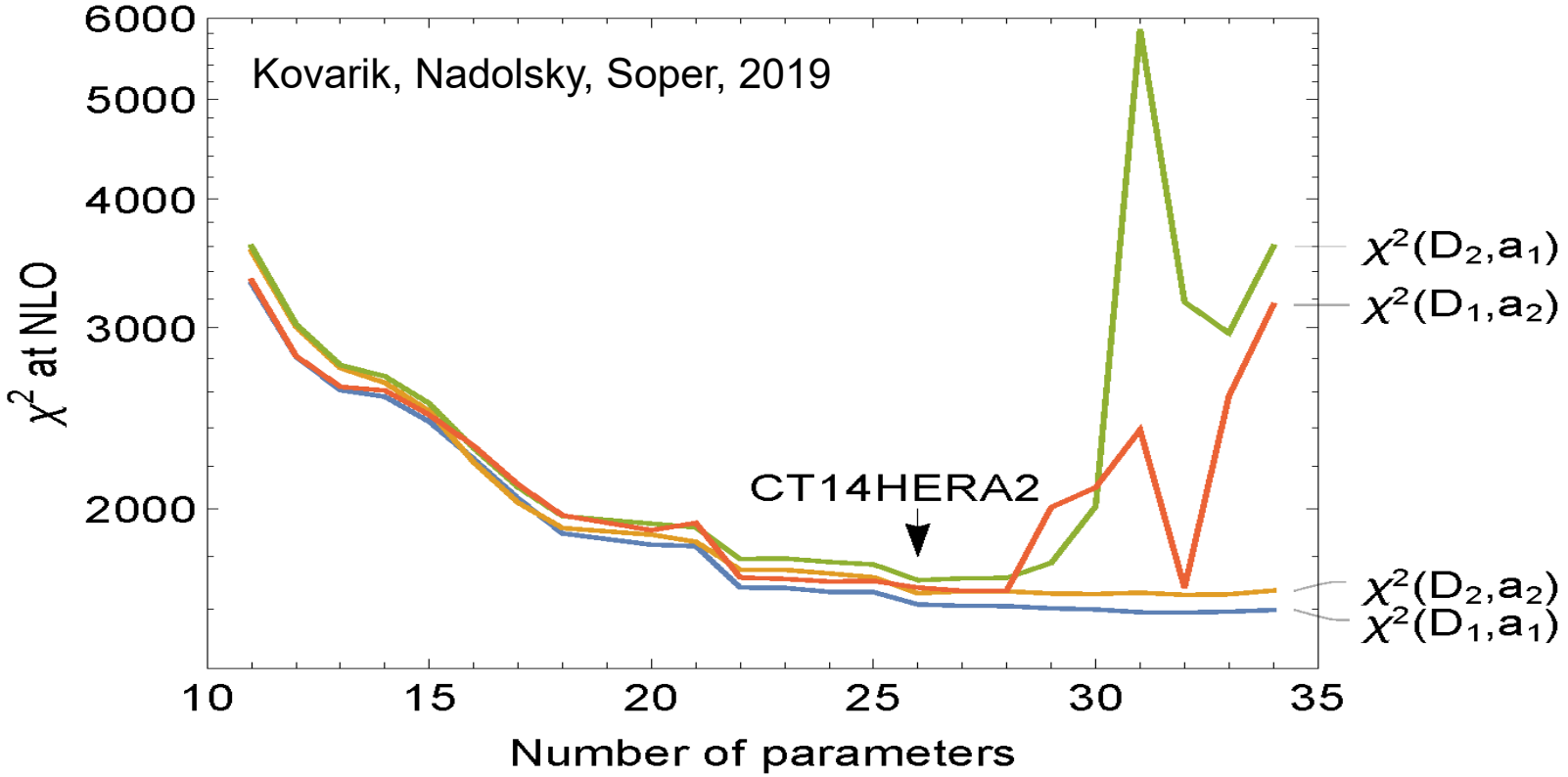
  - each basis polynomial has a single peak, with peaks at different values of z; reduces correlations among parameters

If too few parameters



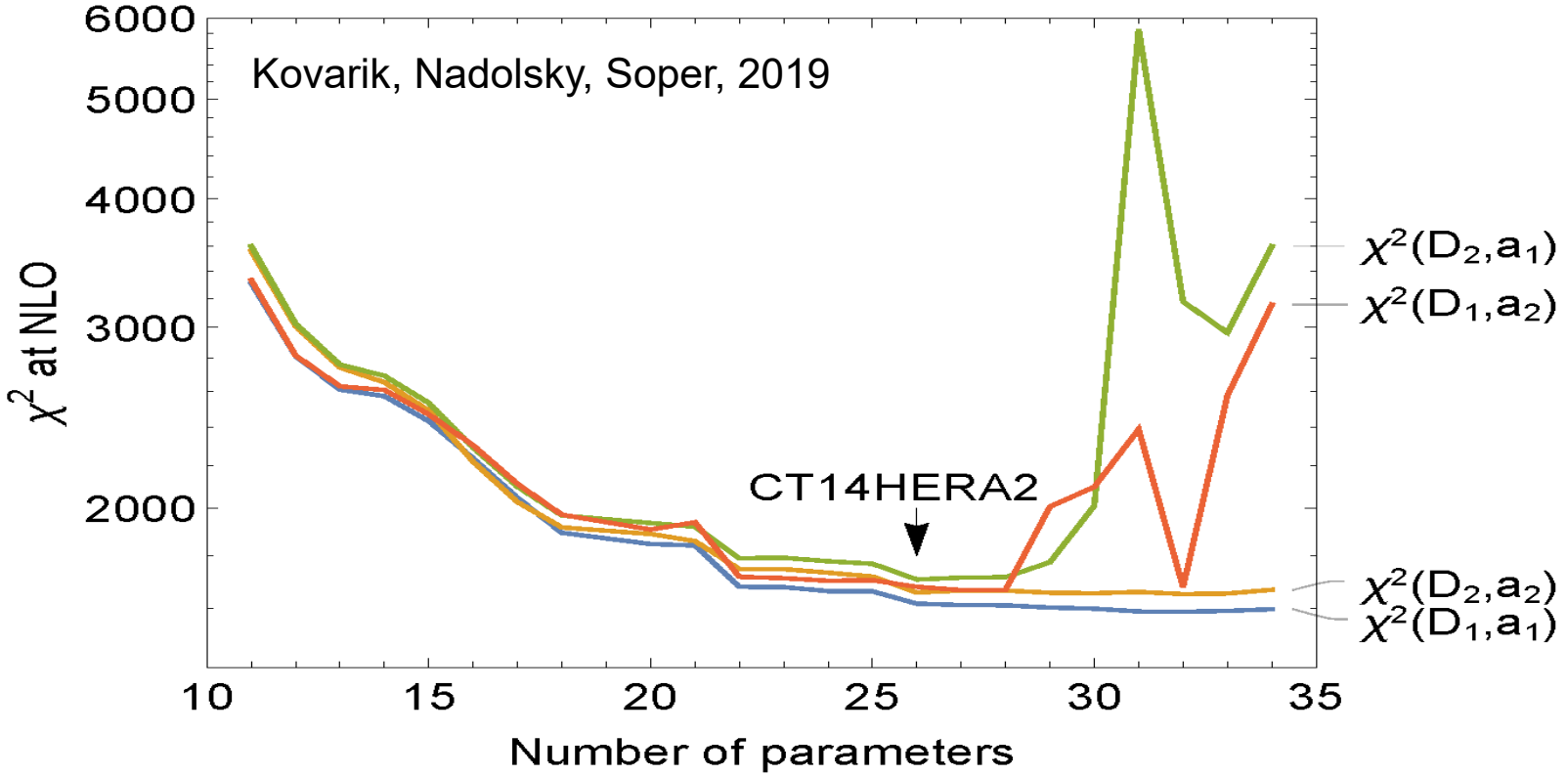
The solution can be consistent and false

If too many parameters



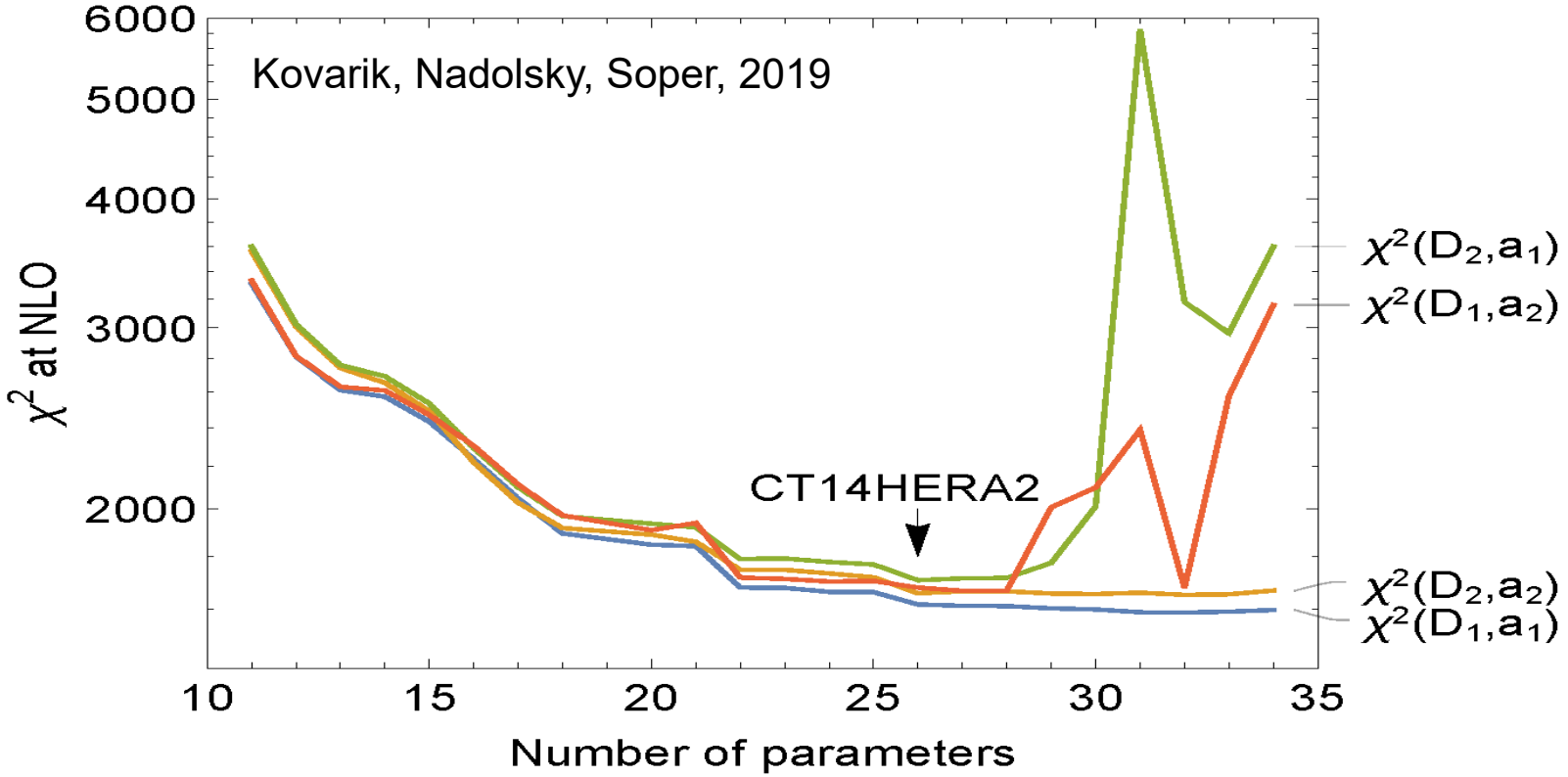
- Randomly split the CT14HERA data set into two halves, D_1 and D_2
- Find parameter vectors a_1 and a_2 from the best fits for D_1 and D_2, respectively

If too many parameters



- **Fitted samples:** $\chi^2(D_1, a_1)$ and $\chi^2(D_2, a_2)$ uniformly decrease with the number of parameters
- **Control samples:** $\chi^2(D_2, a_1)$ and $\chi^2(D_1, a_2)$ fluctuate when the number of parameters is larger than about 30

If too many parameters



$\lesssim 30$ parameters (26 in CT14HERA2) is optimal for describing the CT14HERA2 data set

Selection of experimental data sets

New LHC datasets for CT18

- 245 1505.07024 LHCb Z (W) muon rapidity at 7 TeV(applgrid)
- 246 1503.00963 LHCb 8 TeV Z rapidity (applgrid);
- 249 1603.01803 CMS W lepton asymmetry at 8 TeV (applgrid)
- 250 1511.08039 LHCb Z (W) muon rapidity at 8 TeV(applgrid)
- 253 1512.02192 ATLAS 7 TeV Z pT (applgrid)
- 542 1406.0324 CMS incl. jet at 7 TeV with R=0.7 (fastNLO)
- 544 1410.8857 ATLAS incl. jet at 7 TeV with R=0.6 (applgrid)
- 545 1609.05331 CMS incl. jet at 8 TeV with R=0.7 (fastNLO)
- 565 1511.04716 ATLAS 8 TeV tT pT diff. distributions (fastNNLO)
- 567 1511.04716 ATLAS 8 TeV tT mtT diff. distributions (fastNNLO)
- 573 1703.01630 CMS 8 TeV tT (pT , yt) double diff. distributions (fastNNLO)

- 248 1612.03016 ATLAS 7 TeV Z and W rapidity (applgrid)->CT18Z
  - also uses a special small-x factorization scale, charm mass m_c=1.4 GeV
  - serious changes in PDFs, so warrants a separate PDF

CT14 PDFs with HERA1+2 (=HERA2) combination

Phys.Rev. D95 (2017) 034003

Separate the four HERA2 DIS processes;
($Q_{cut} = 2$ GeV)

	N_{pts}	$\chi^2_{red.} / N_{pts}$
NC e^+p	880	1.11
CC e^+p	39	1.10
NC e^-p	159	1.45
CC e^-p	42	1.52
totals		
[reduced χ^2] /N	1120	1.17
χ^2 / N	1120	1.25
R^2 / N	1120	0.08

e^+p data are fitted fine

e^-p data are fitted poorly

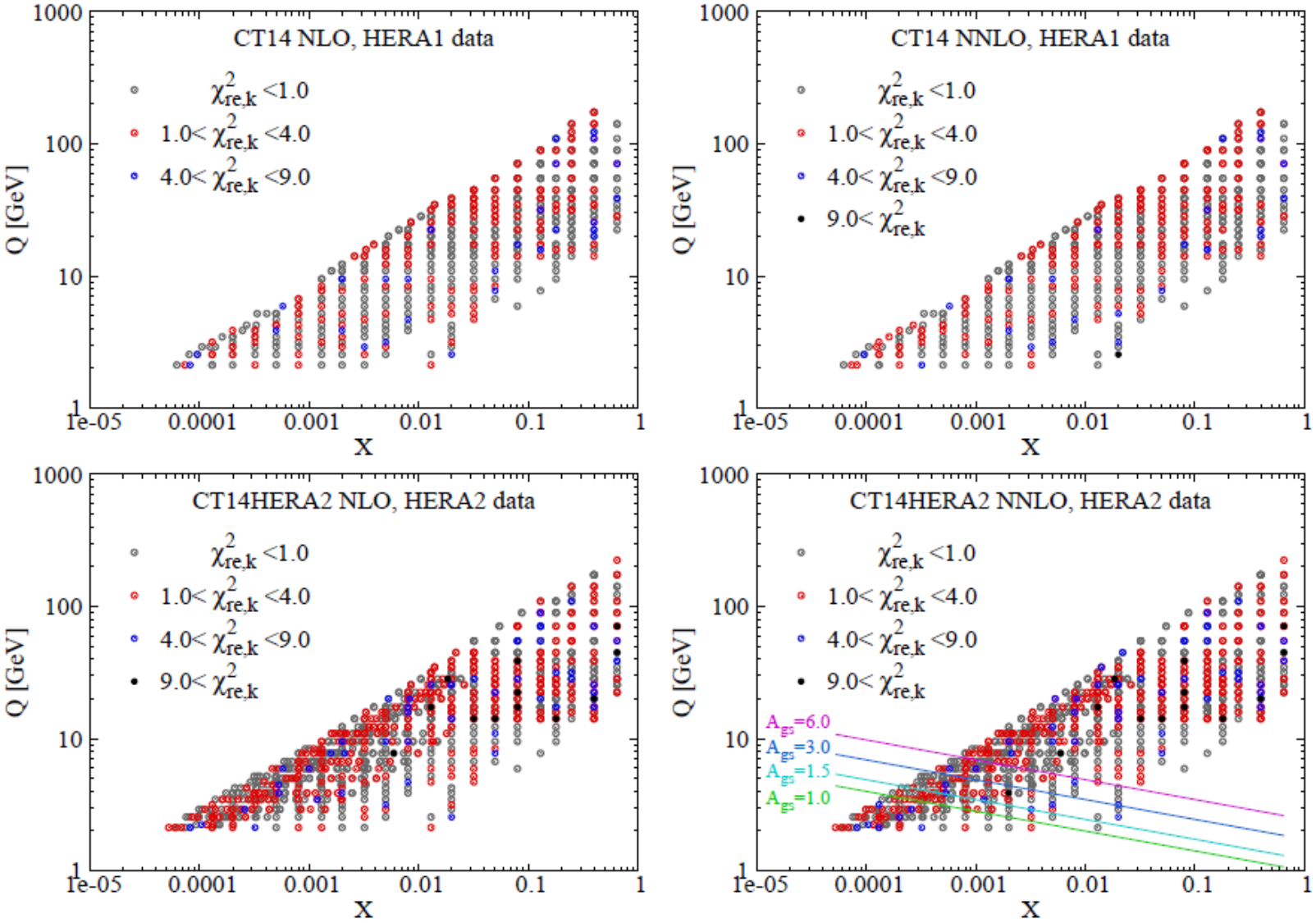
reduced χ^2 values

χ^2 = [reduced χ^2] + R^2

The quadratic penalty for 162 systematic errors = 87.5

Fair (not perfect) agreement

CT14 PDFs with HERA1+2 (=HERA2) data



Points with excessive χ^2 are randomly scattered in the $\{x, Q\}$ plane

Theoretical uncertainty in DIS

Mild NNLO theoretical uncertainties in **large** DIS data sets have a non-negligible overall effect on the global χ^2

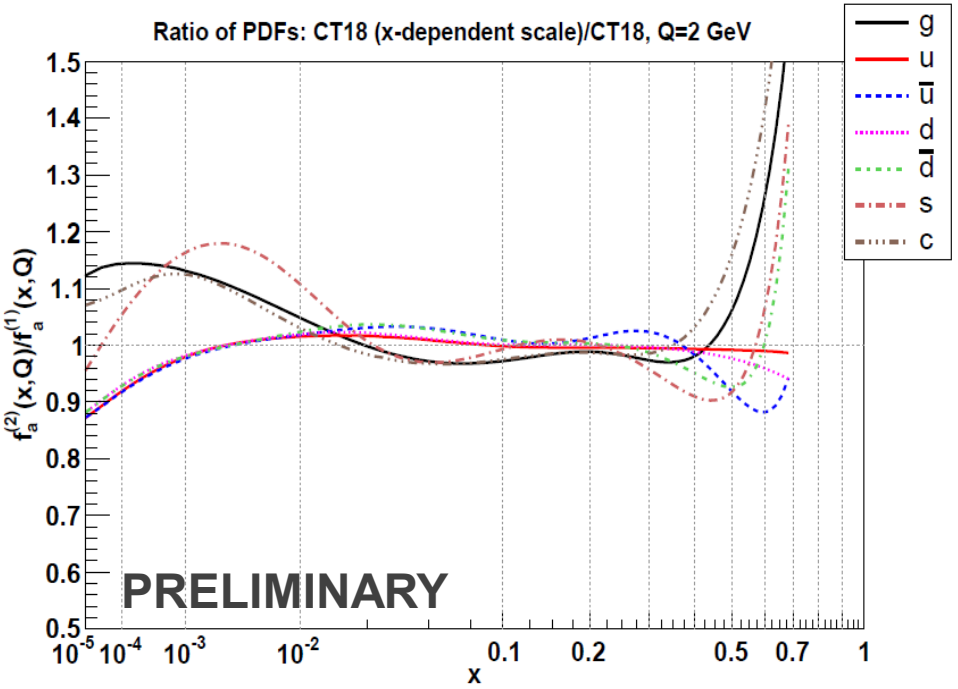
The following **x-dependent** factorization scale at NNLO improves description of CTEQ-TEA DIS data sets by mimicking

- missing N3LO terms at x>0.001
- small-x/saturation terms at x<0.001

$$\mu^2_{DIS,X} = 0.8^2 \left(Q^2 + \frac{0.3\ GeV^2}{x^{0.3}} \right)$$

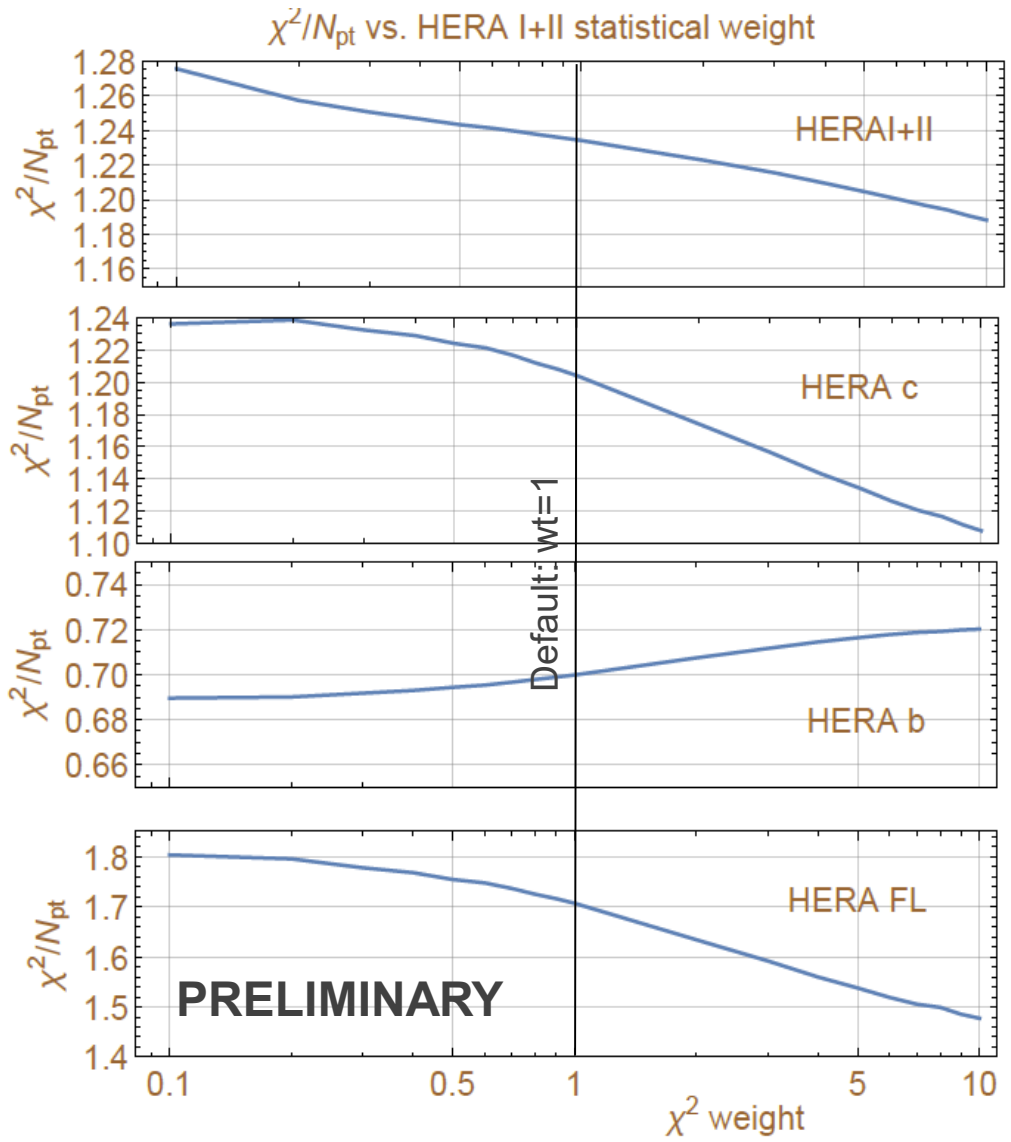
CT18Z uses a combination of $\mu_{DIS,X}$ (preferred by DIS) and increased $m_c^{pole} = 1.4$ GeV (preferred by LHC vector boson production, disfavored by DIS)

X-dependent DIS scale, effect on PDFs



Using $\mu_{DIS,X}$ in a **fixed-order** NNLO cross section bears similar effect to small-x resummation/saturation. In particular, the gluon and strange PDFs are enhanced at $x < 10^{-2}$

Varied statistical weight for HERA I+II DIS set

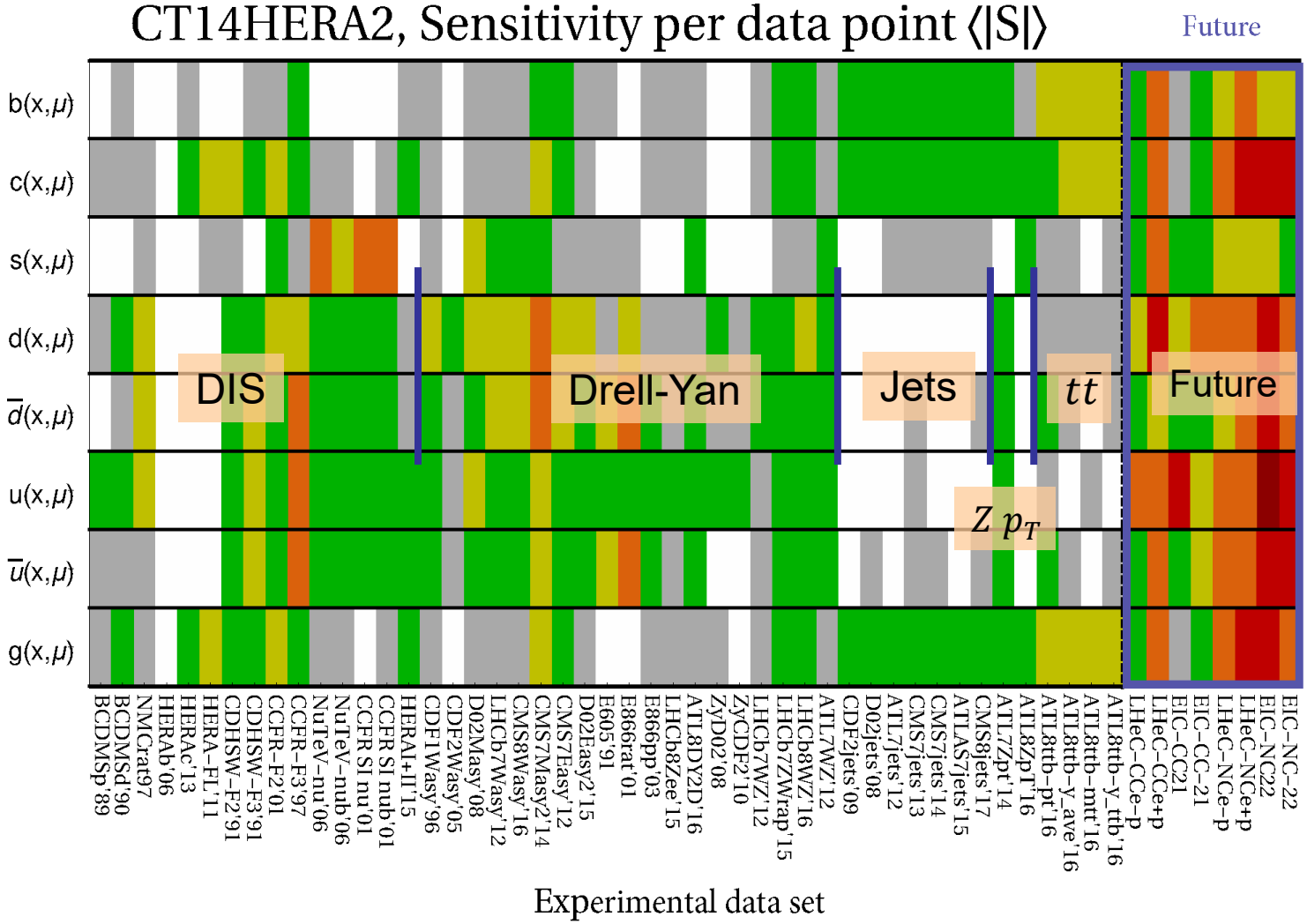


The CT18Z fits using the $\mu_{DIS,X}$ scale reproduce many features of NNLO-NNLx fits with $\ln(1/x)$ resummation by the NNPDF [arXiv:1710.05935] and xFitter [1802.0064] groups.

Left: when the statistical weight for the HERA I+II data set is increased to $wt = 10$ to suppress pulls from the other experiments, χ^2_{CT18Z}/N_{pt} for HERA I+II DIS **and** HERA heavy-quark production decreases to about the same levels as in NNLO+NNLx fits to HERA DIS only by NNPDF and xFitter.

Still, some residual tension

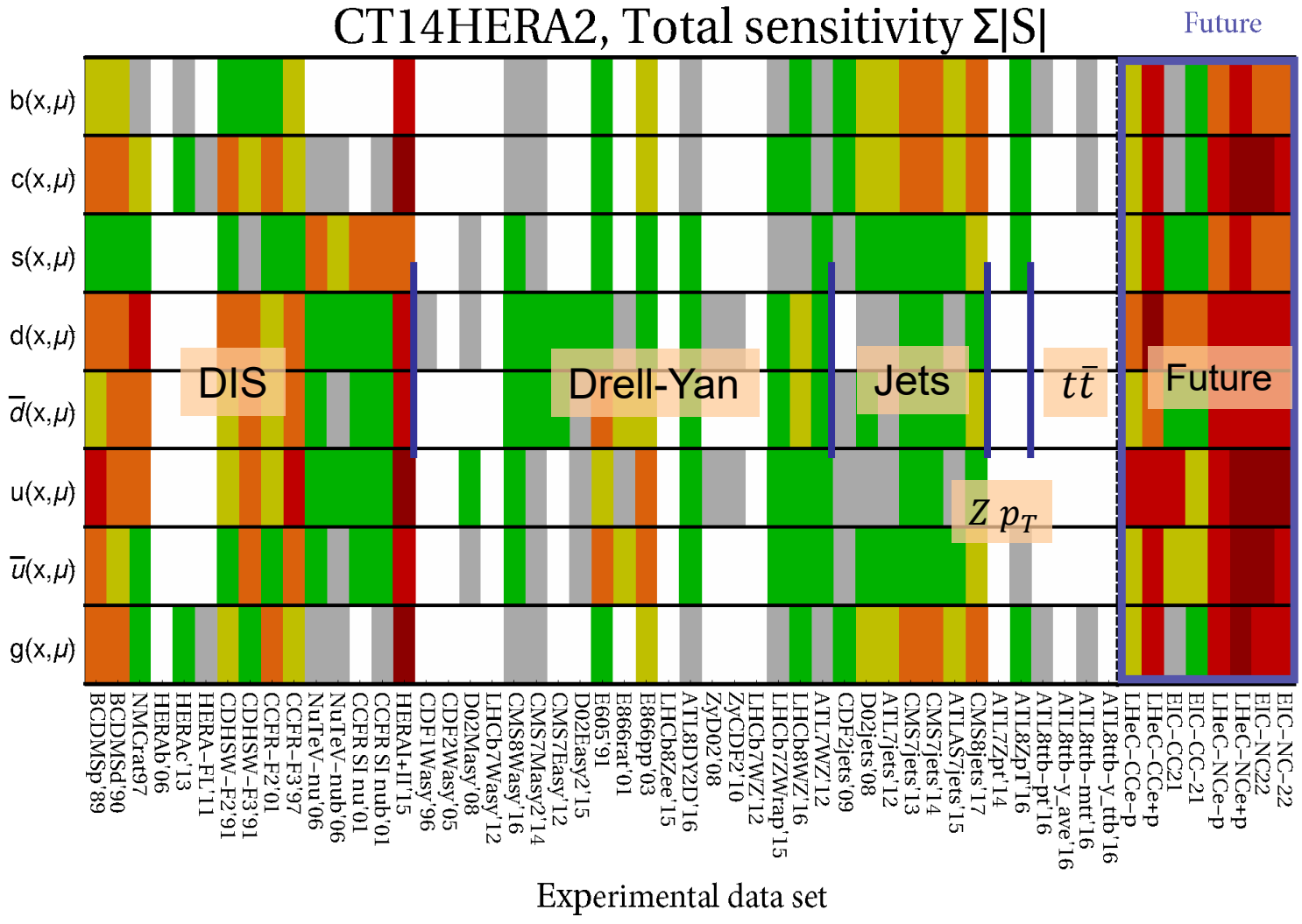
Sensitivity of hadronic experiments to PDFs



Average sensitivity to $f_a(x_i, \mu_i)$ **per data point**

Computed using the **PDFSense** code

Sensitivity of hadronic experiments to PDFs

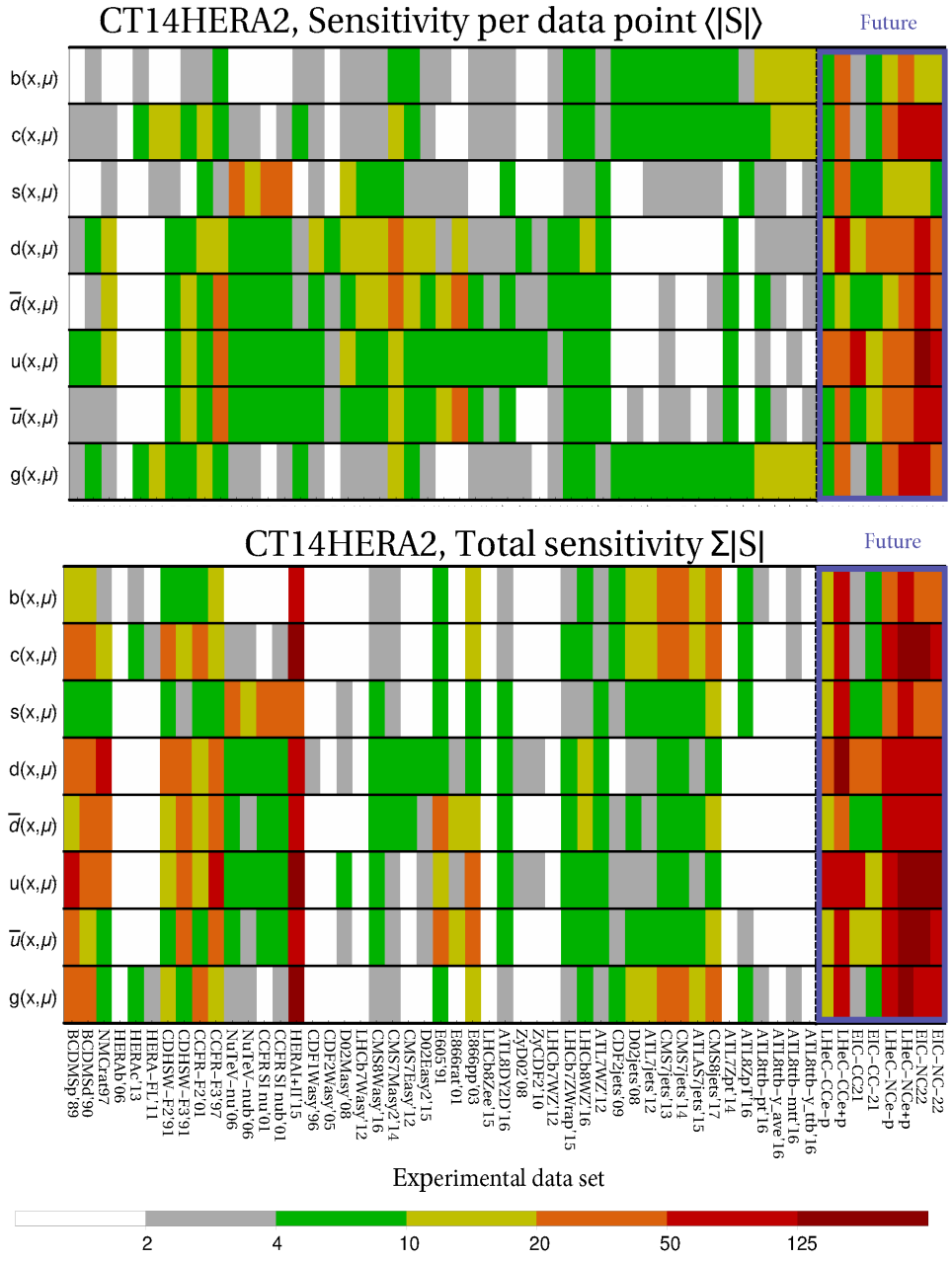


Total sensitivity to $f_a(x_i, \mu_i)$, **summed over data points**

$$\sum_{\boldsymbol{points}} |\boldsymbol{S_{f,i}}|$$

Computed using the **PDFSense** code

Sensitivity of hadronic experiments to PDFs



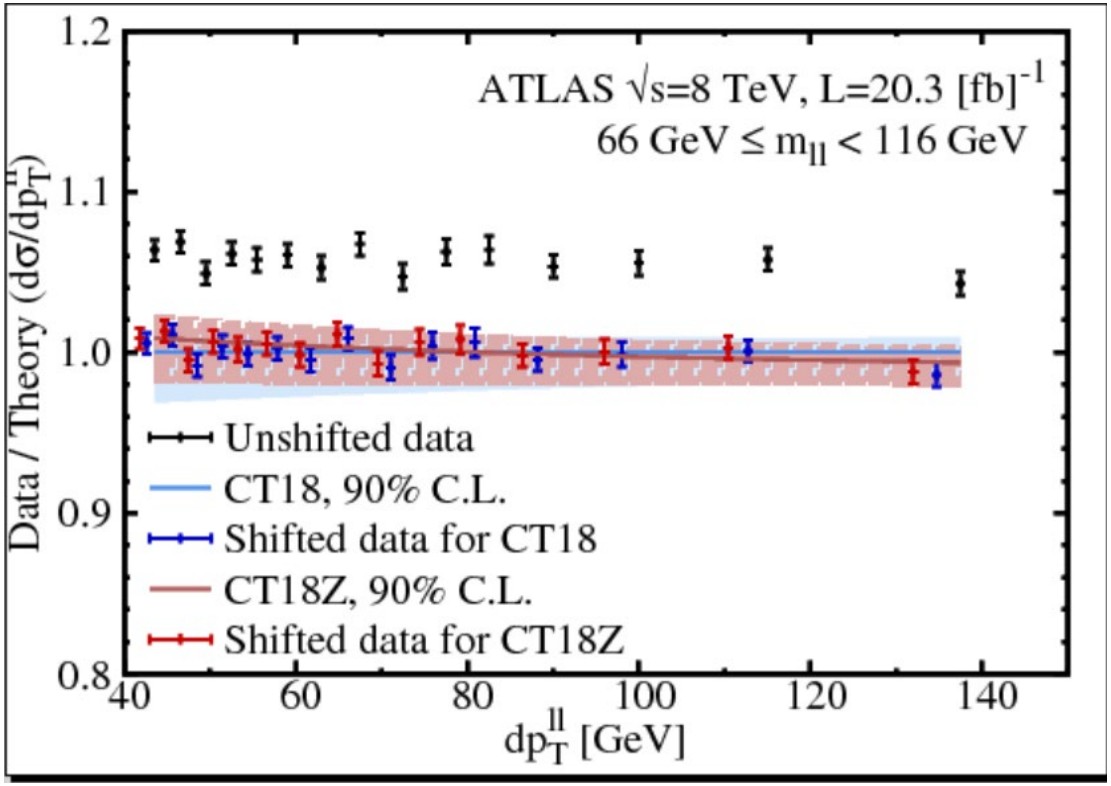
HERA I+II, BCDMS, NMC, DIS data sets dominate experimental constraints

Among the LHC data sets:
CMS 7 & 8 TeV single-inclusive jet production have highest total sensitivity ($N_{pt} > 100$), modest sensitivity per data point

$t\bar{t}$, high-$\mathrm{p_T}$ Z production have high sensitivity per data point, smaller total sensitivity ($N_{pt} \sim 10 - 20$)

Future DIS experiments (LHeC, EIC) will have dramatically higher sensitivity

High-p_T Z boson production

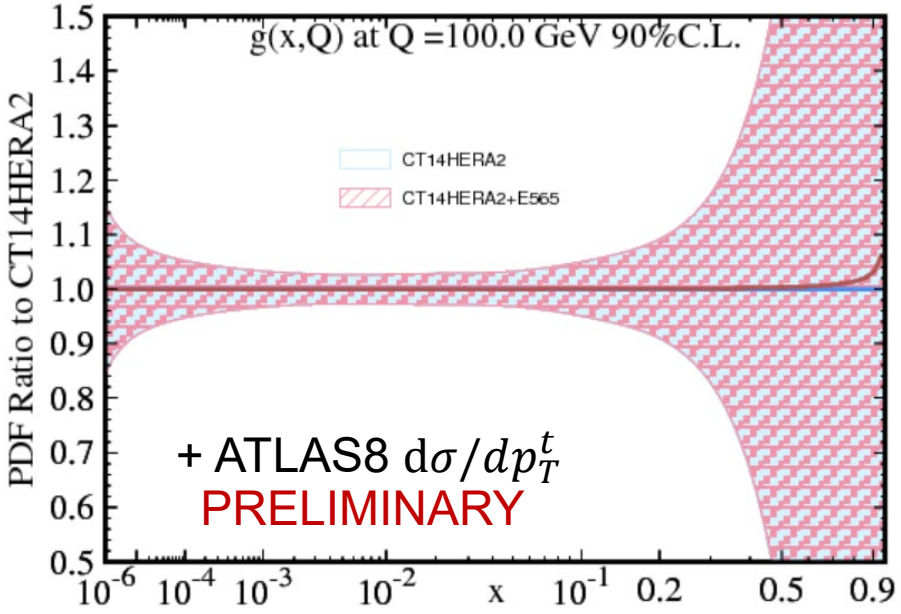


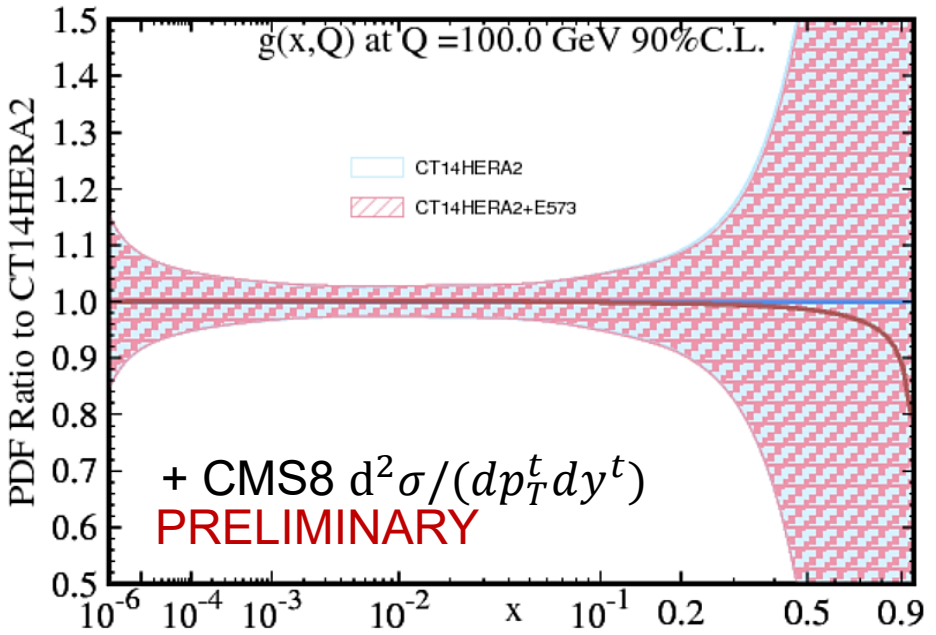
- Include the ATLAS 8 TeV data on $d\sigma/dQ^2 dp_T^{\ell\bar{\ell}}$ at $50 \leq p_T^{\ell\bar{\ell}} \leq 150$ GeV (in the region free from resummation effects)
- Fast ApplGrid $O(\alpha_s^2)$ calculation with point-by-point $O(\alpha_s^3)$ corrections from NNLOJet++

- To reach $\chi^2/N_{pt} \approx 1$, the scale $\mu_{R,F}^2 = M_{\ell\bar{\ell}}^2 + {p_T^{\ell\bar{\ell}}}^2$ and Monte-Carlo integration error of 0.5% must be used; non-neglible dependence on QCD scales at N3LO
- Agreement with other experiments, NNPDF3.0red study **[Boughezal, Guffanti, Petriello, Ubiali, 1705.00343]**

ePump estimates of impact of $t\bar{t}$ data

When the $t\bar{t}$ data were added to the CT14HERA2 NNLO data set according to the **ePump** Hessian reweighting method (CT14nn+EXXX), no significant change in the PDF uncertainty was observed



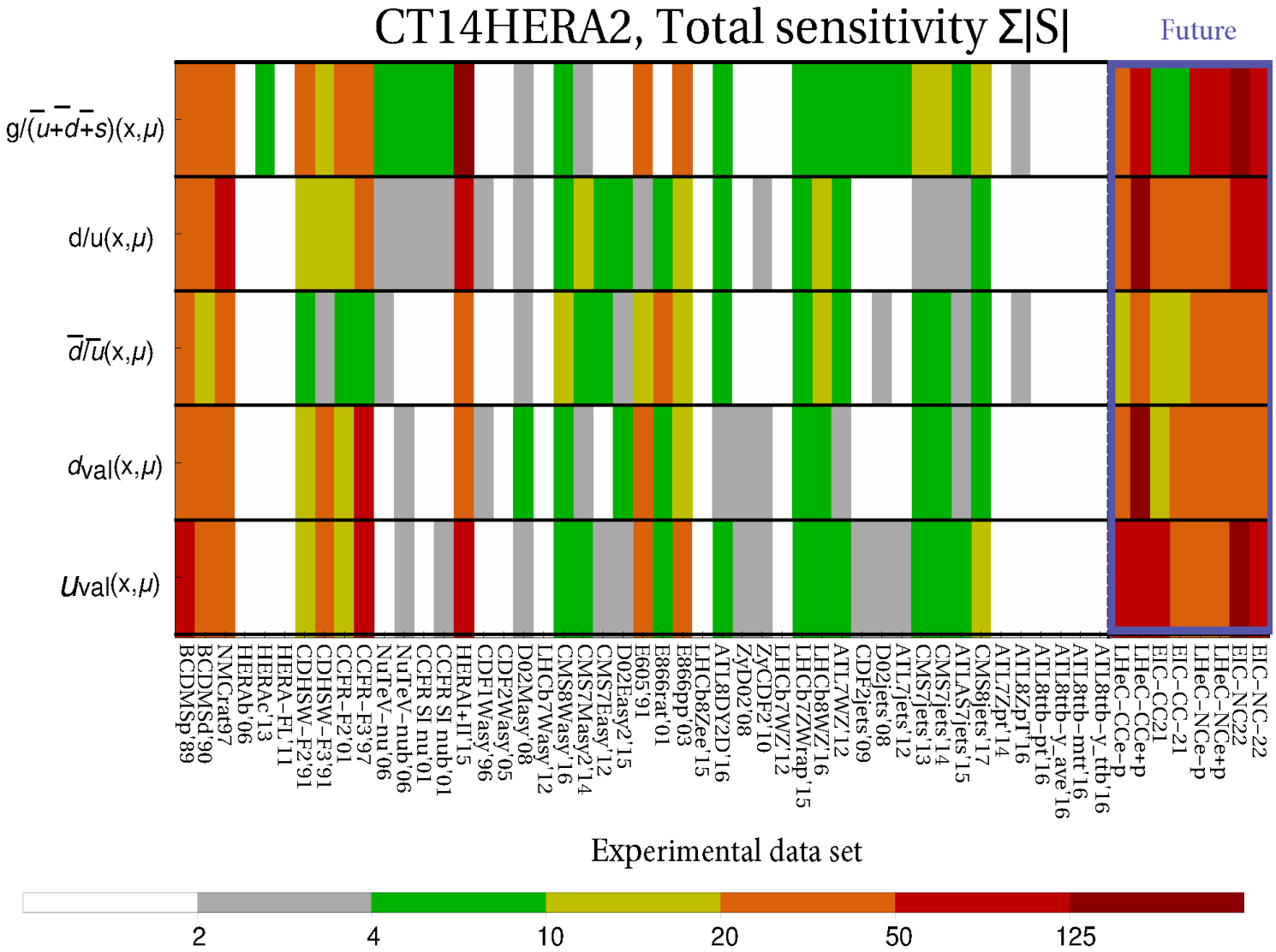


FastNNLO calculation from **Csakon, Mitov, et al**.

Slight reduction in gluon PDF uncertainty at $x \sim 0.2$; weaker than for the other groups because of also including jet production data

Top-antitop production experiments have **strong** sensitivity per data point, offer a **novel** independent measurement of the gluon PDF in several channels

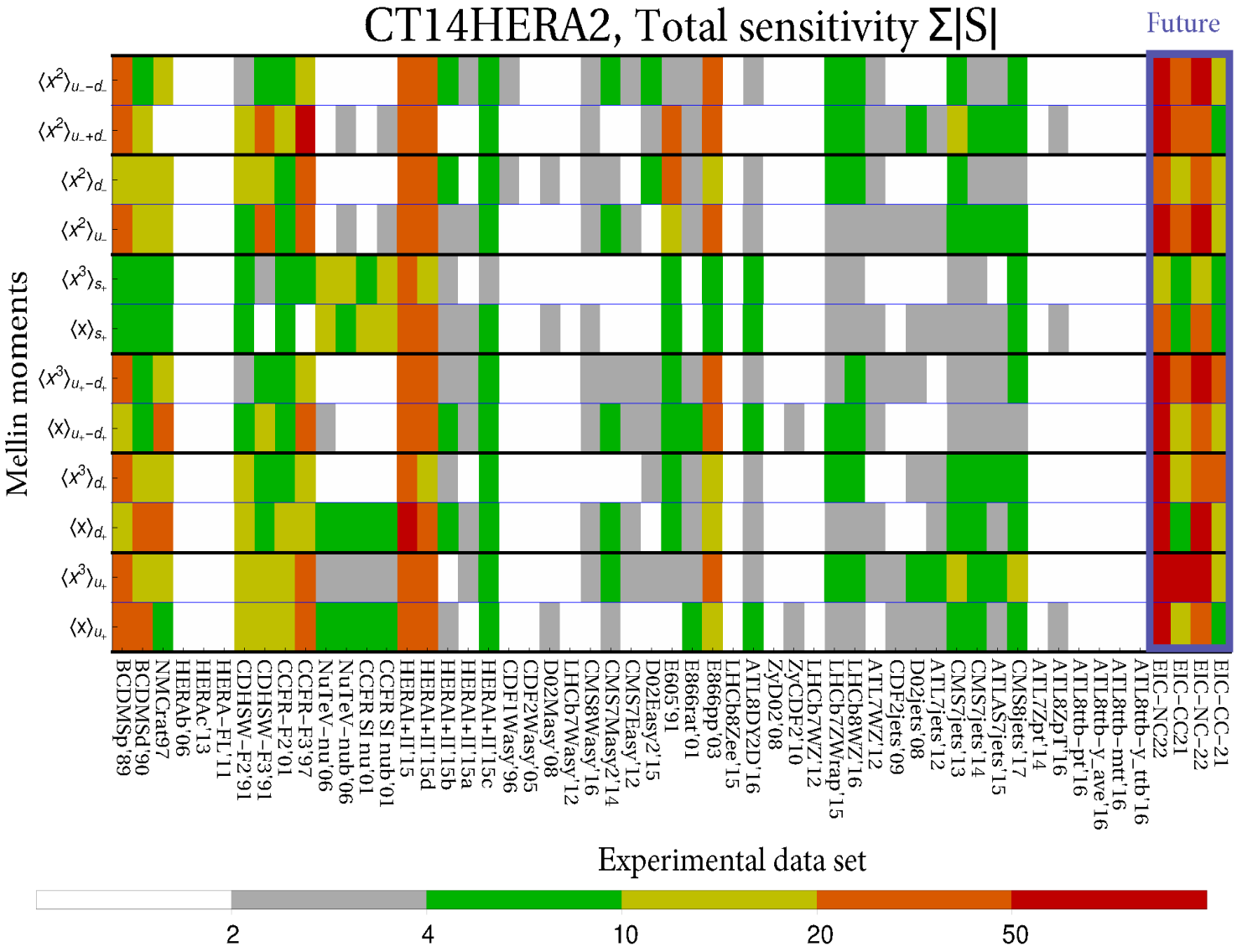
Sensitivity to PDF ratios



Total sensitivity to $f_a(x_i, \mu_i)$, **summed over data points**

Average sensitivity per data point in the backup

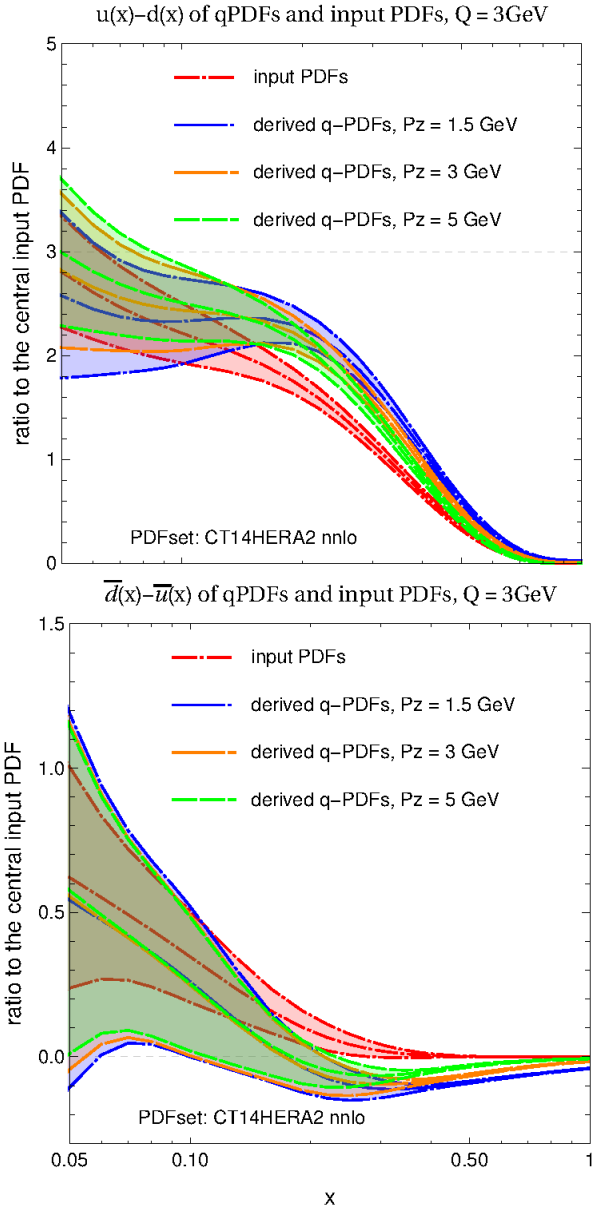
Sensitivity to Mellin moments



We show Mellin moments computable on the lattice

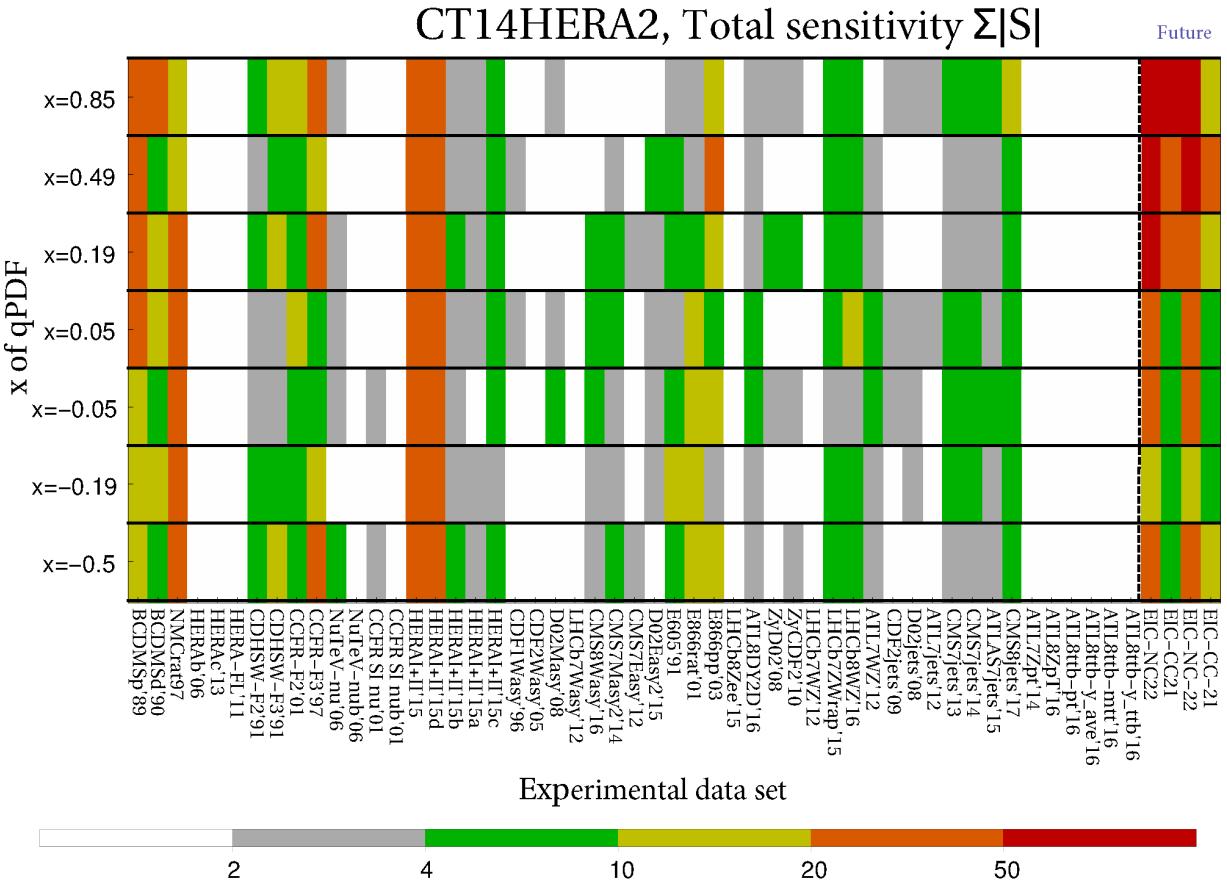
HERA, BCDMS, NMC, E866 DY pair production are most sensitive to the moments

Sensitivity to lattice quasi-PDFs



$$q(x) \equiv \begin{cases} u(x) - d(x), & x > 0 \\ \bar{d}(|x|) - \bar{u}(|x|), & x < 0 \end{cases}$$

$q(x, Q, P_Z)$, $Q = 3$ GeV, $P_z = 1.5$ GeV



How well do the PDF fits describe experimental data?

Weak and strong goodness-of-fit criteria
Kovarik, P. N., Soper, in preparation

Weak (common) goodness-of-fit (GOF) criterion

Based on the global χ^2

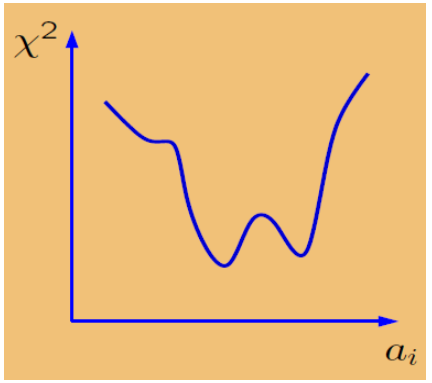
A fit of a PDF model to N_{exp} experiments with N_{pt} points ($N_{pt} \gg 1$) is good at the probability level p if $\chi^2_{global} \equiv \sum_{n=1}^{N_{exp}} \chi^2_n$ satisfies

$$P\left(\chi^2 \geq \chi^2_{global}, N_{pt}\right) \geq p; \qquad e.g.$$

$$\left|\chi^2_{global} - N_{pt}\right| \lesssim \sqrt{2N_{pt}} \text{ for } p = 0.68$$

Even when the weak GOF criterion is satisfied, parts of data can be poorly fitted

Then, **tensions between experiments** may lead to **multiple solutions** or **local χ^2 minima** for some PDF combinations



Strong GOF criterion

(From Kovarik, P.N., Soper, paper in preparation)

Shatter the global data set into N_{part} partitions with $N_{pt,n}$ points each

$$1 \leq N_{part} \leq N_{pt}$$

$$\sum_{n=1}^{N_{part}} N_{pt,n} = N_{pt}$$

A fit is good for this arrangement iff the weak GOF criterion is satisfied for every partition. That is, for each partition n:

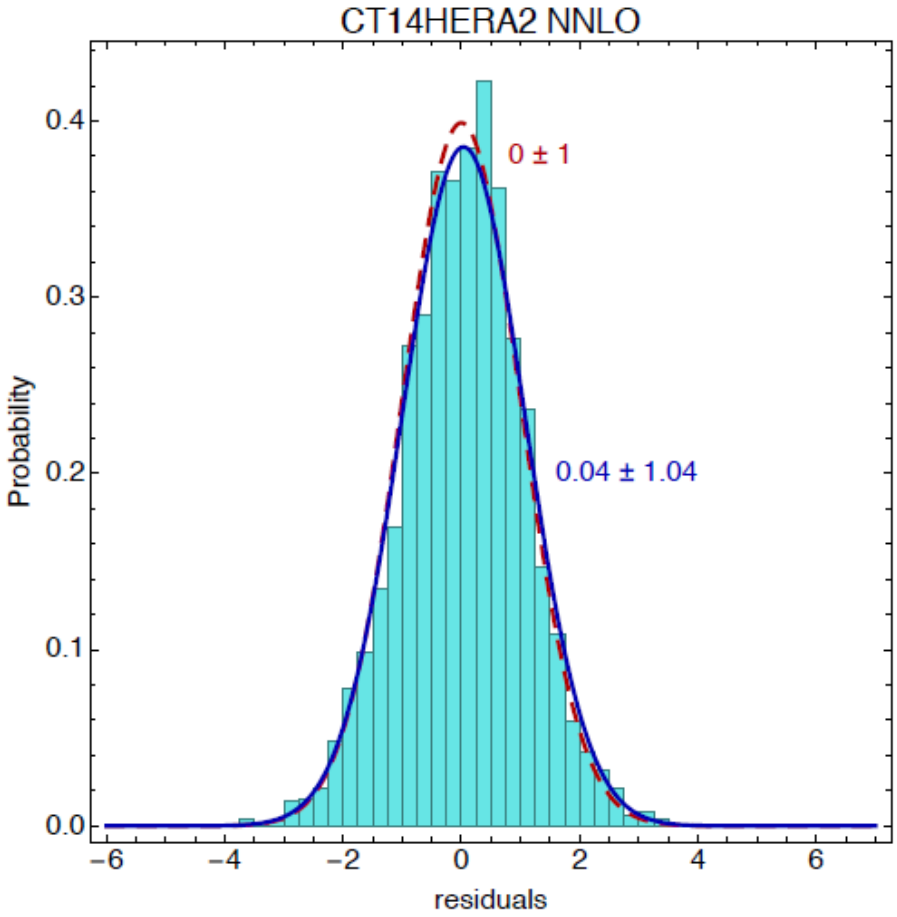
- differences between theory and data are indistinguishable from random fluctuations
- $P(\{\chi_n^2\}) \geq 0.68$ for the distribution of χ_n^2 over N_{part} partitions

A fit is close to the ideal when this condition is satisfied for many shattering arrangements

How good are our PDF fits?

Note: It is convenient to define $S_n(\chi^2, N_{pt})$ that approximately obeys the standard normal distribution (mean=0, width=1) independently of N_{pt}

Example: $N_{part} = N_{pt}$, data residuals r_n

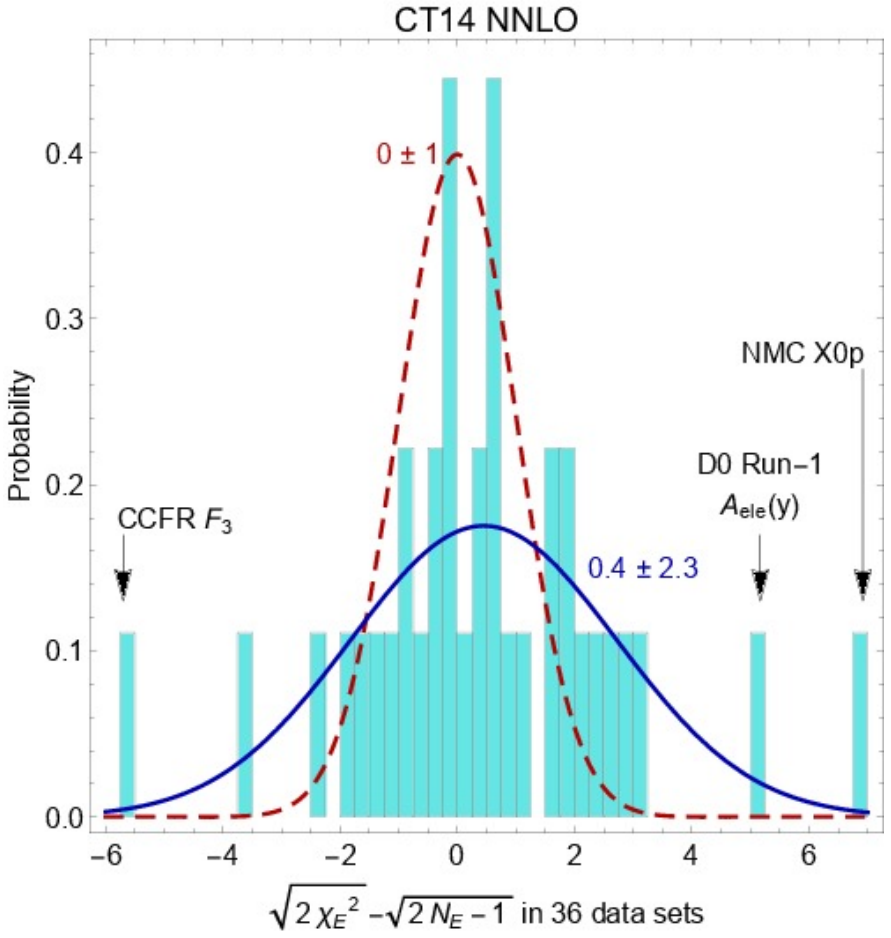


$$r_n \equiv \frac{T_n(\{a\}) - D_n^{shifted}(\{a\})}{\sigma_n^{uncorrelated}}$$

The distribution of residuals is consistent with the standard normal distribution

Full definition of r_n in the backup slides

Example: $N_{part} = N_{exp}$, individual experiments



Define

$$S_n(\chi^2, N_{pt}) \equiv \sqrt{2\chi^2} - \sqrt{2N_{pt} - 1}$$

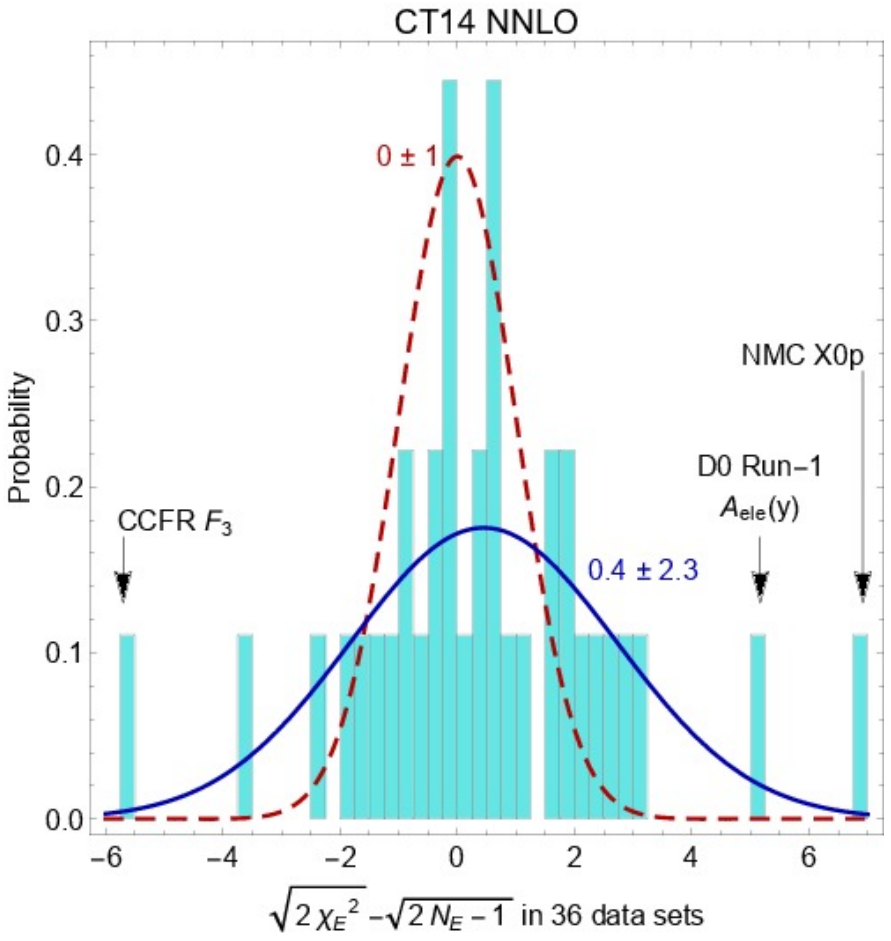
$S_n(\chi_n^2, N_{pt,n})$ are Gaussian distributed with mean 0 and variance 1 for $N_{pt,n} \geq 10$

[R.A.Fisher, 1925]

Even more accurate (χ^2, N_{pt}):
T.Lewis, 1988

An empirical S_n distribution can be compared to $\mathcal{N}(0,1)$ visually or using a statistical (KS or related) test

Example: $N_{part} = N_{exp}$, individual experiments

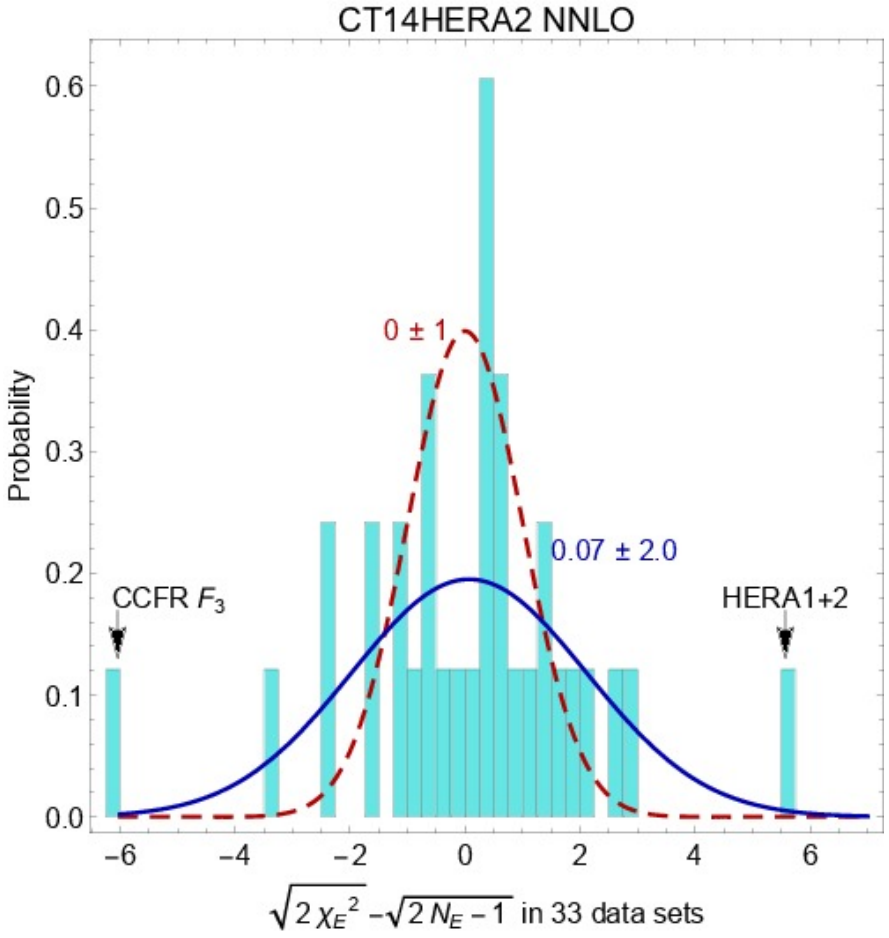


Some S_n are too big or too small in a global fit

CT14 NNLO:

- $S_n > 4$ for NMC DIS ep cross section and D0 Run-1 electron charge asymmetry
- These data sets are eliminated in CT14HERA2/CT18 fits
- The rest of CT14 experiments are reasonably consistent; $S_n \sim \mathcal{N}(0.3, 1.6)$

Example: $N_{part} = N_{exp}$, individual experiments



CT14 HERA2 NNLO:

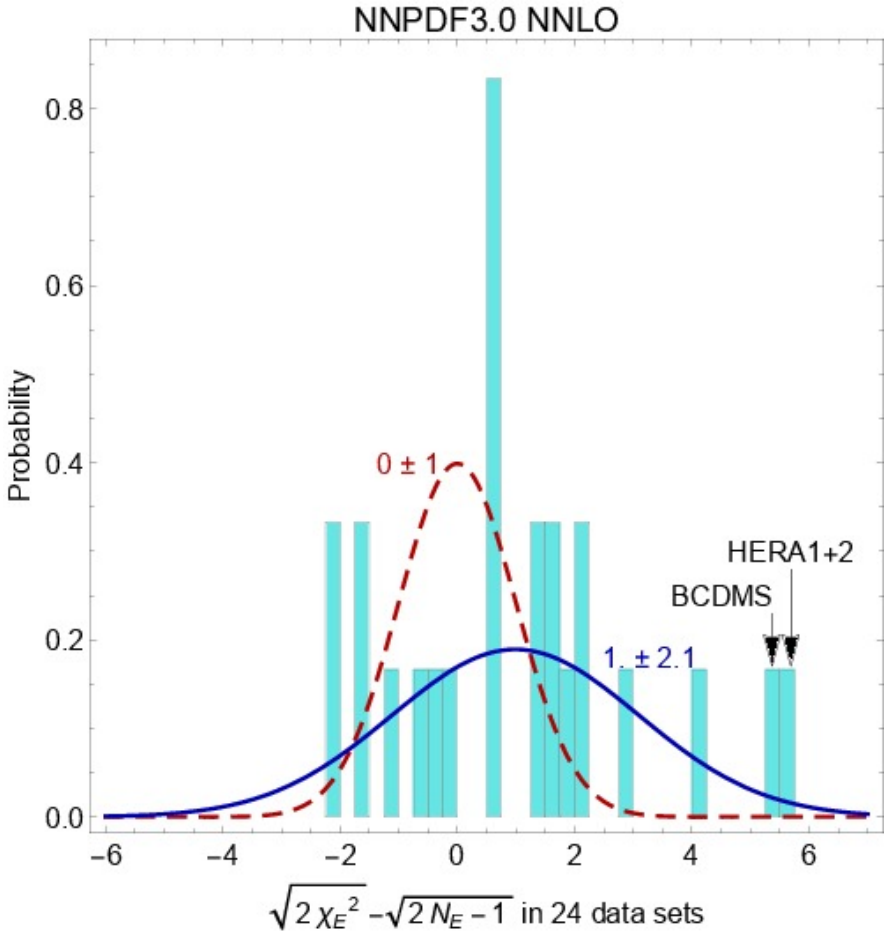
For HERA 1+2 inclusive DIS data

- $\frac{\chi^2}{N_{pt,n}} > 1.15$: not good for $N_{pt,n} = 1120$

$$\Rightarrow S_n^{HERA\ I+II} = 5.89$$

- Tensions between e^+p and e^-p DIS channels
- Partly improved by the x-dependent factorization scale (CT18Z) or small-x resummation

Example: $N_{part} = N_{exp}$, individual experiments



Similar tensions observed in other global fits

NNPDF3.0 NNLO:

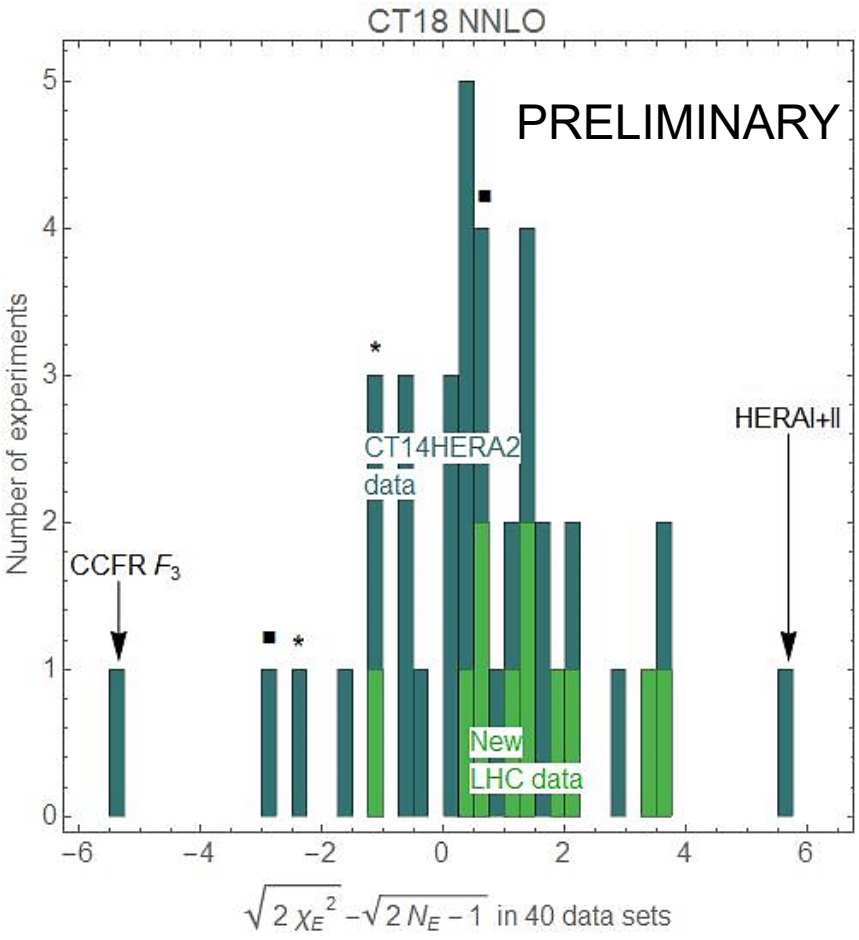
$S_n > 5$ for HERA I+II,
also BCDMS DIS

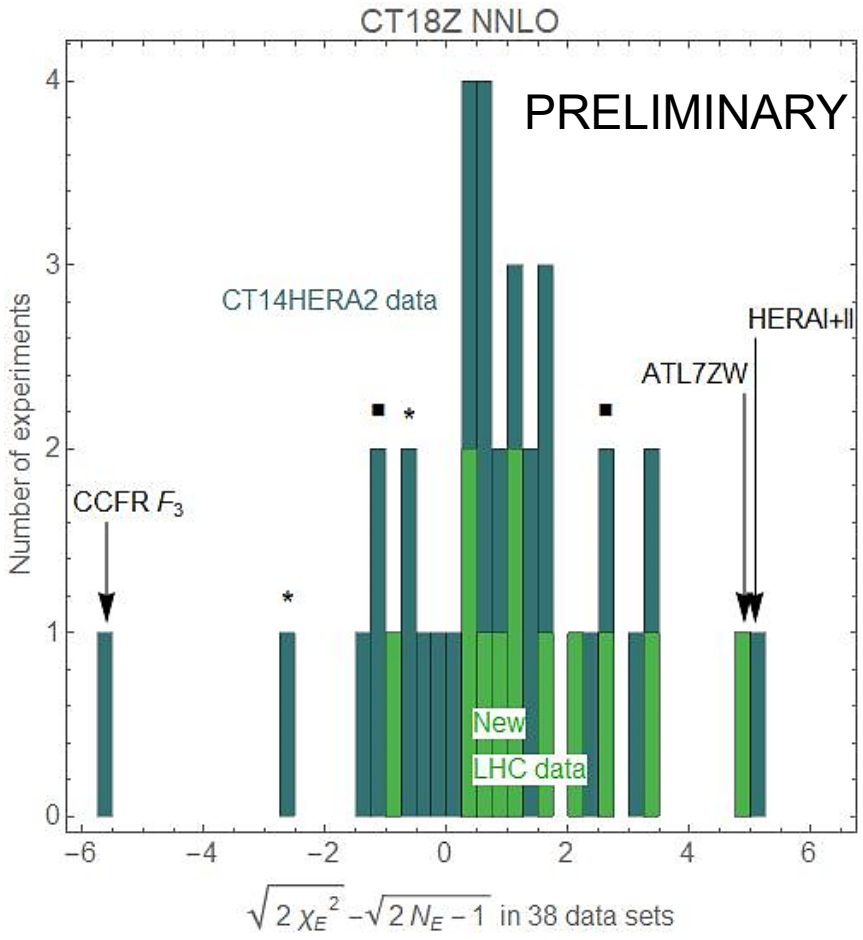
In NNPDF3.1, S_n(HERA I+II) is improved to ≈ 3 by using fitted charm and/or small-x resummation

CT18 (CT18Z) NNLO

13 (14) new LHC experiments with 665 (711) data points

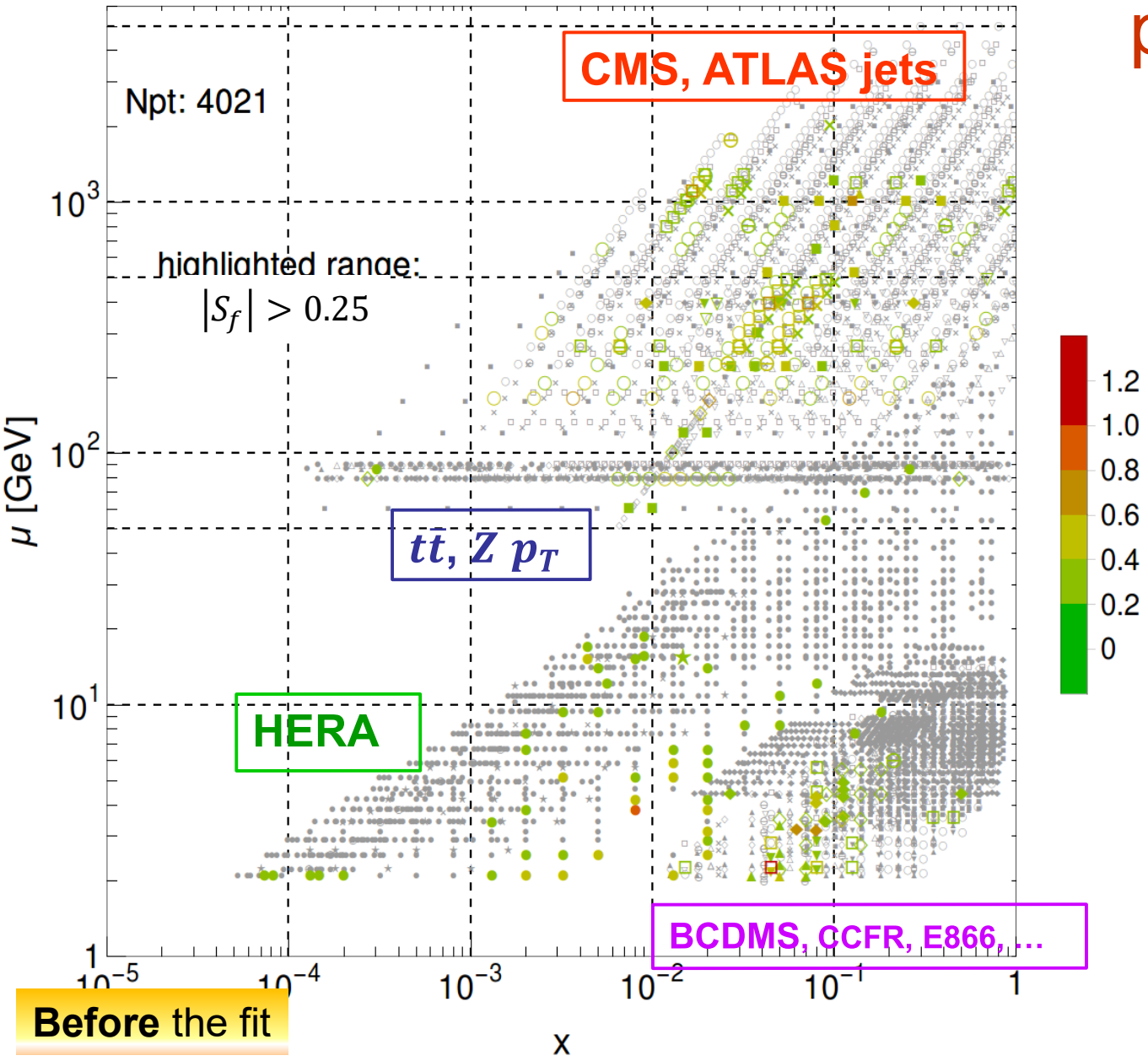
- New LHC experiments tend to have larger S_n
- ATLAS 7 TeV Z, W production has $S_n \approx 5.2$, included in CT18Z fit only





Studying the tensions of data sets in detail

$|S_f|$ for $\sigma(H^0)$, 14 TeV, CT14HERA2NNLO



Before the fit

Higgs boson production

HERA DIS still has the dominant sensitivity!

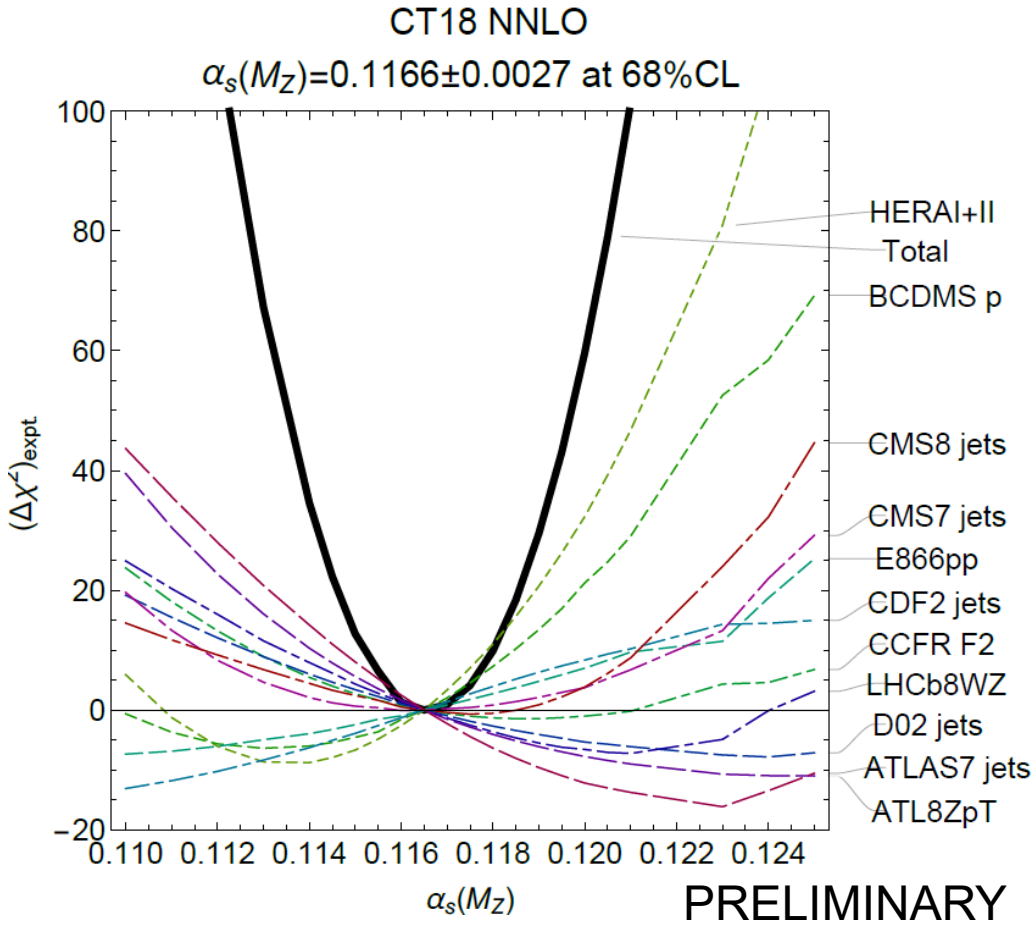
CMS 8 TeV jets is the next expt. after HERA sensitive to $\sigma_H(14\ \text{TeV})$; jet scale uncertainty dampens $|S_f|$ for jets

Good correlations C_f with some points in E866, BCDMS, CCFR, CMS WASY, Z p_T and $t\bar{t}$ production; but not as many points with high $|S_f|$ in these processes

Lagrange Multiplier (LM) Scans: $\alpha_s(M_Z)$

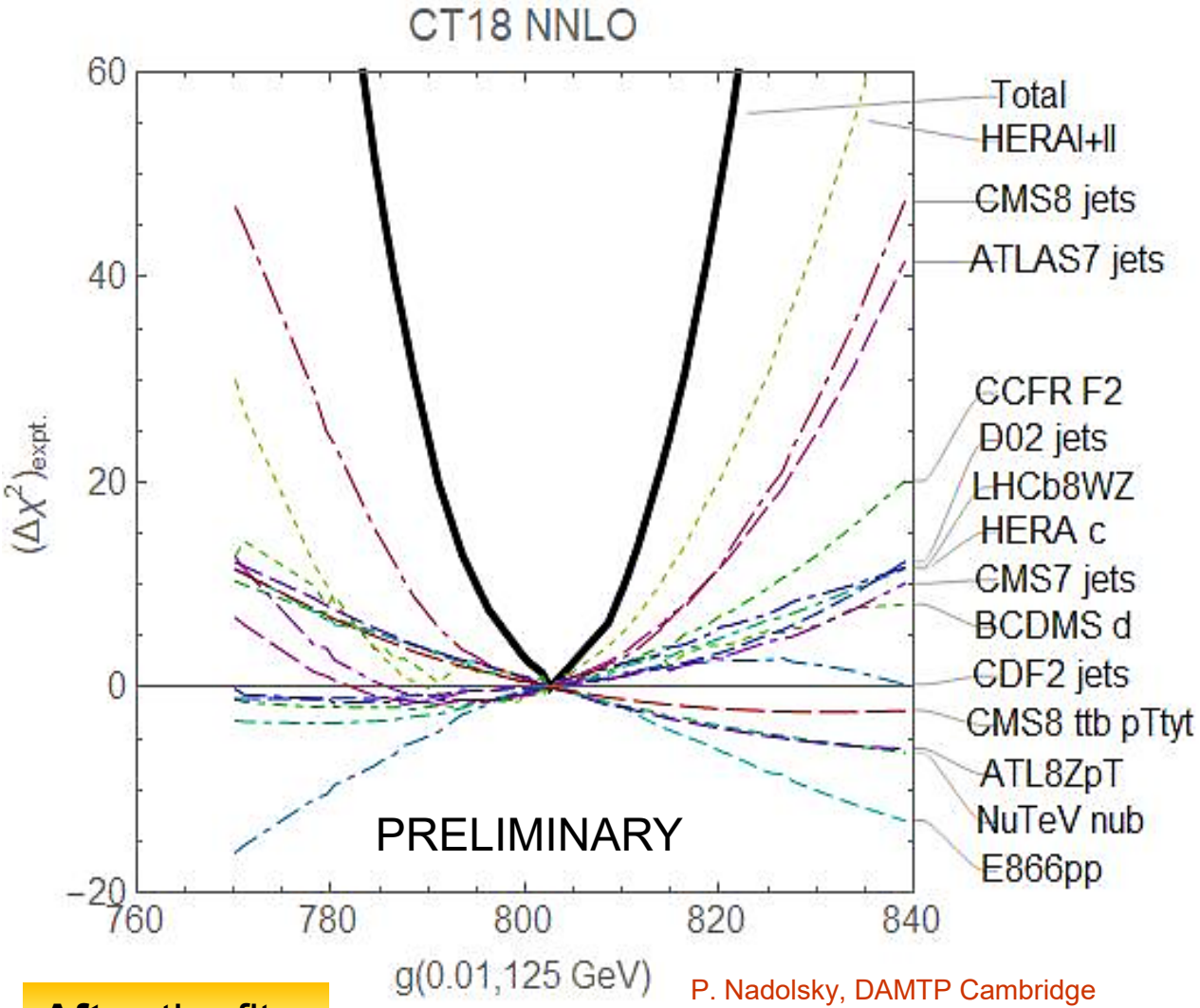
The LM scan technique is introduced in **Stump et al., Phys.Rev. D65 (2001) 014012**

☺ **Detailed dependence of χ^2** ☹ **slow; refitting on a supercomputing cluster**



$\alpha_s(m_Z)$ from global fit closer to 0.117 than to 0.118

Which experiments constrain the gluon? $x = 0.01, Q = 125$ GeV [Higgs region]



After the fit

The LM scans broadly confirm S_f estimates

HERAI+II, ATLAS7 jets, CMS8 jets impose the tightest constraints; are in agreement

E866, ATLAS 8 Z p_T prefer higher gluon

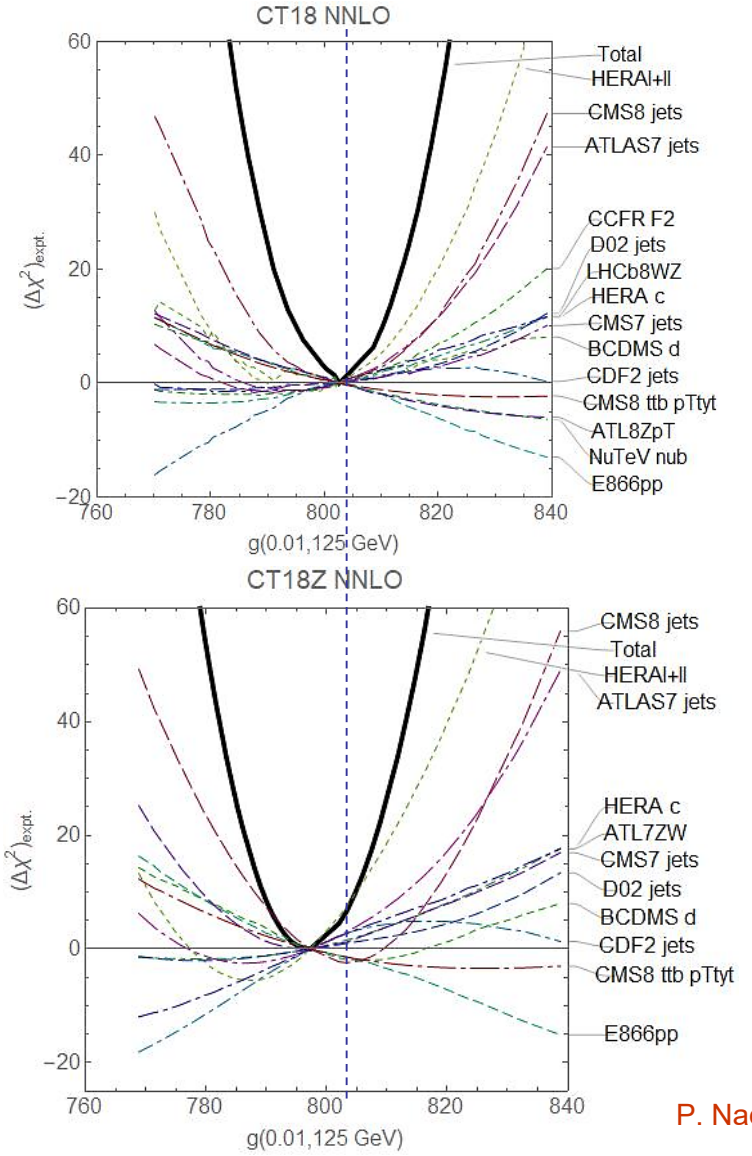
Rankings of experiments most sensitive to $g(0.01,125\ \mathrm{GeV})$

$d/u(x=0.1,\mu=1.3\ \mathrm{GeV})$			$g(x=0.01,\mu=125\ \mathrm{GeV})$		
PDFSense		LM scan	PDFSense		LM scan
CT14HERA2	CT18pre	CT18pre	CT14HERA2	CT18pre	CT18pre
HERAI+II'15	NMCrat'97	NMCrat'97	HERAI+II'15	HERAI+II'15	HERAI+II'15
BCDMSp'89	HERAI+II'15	CCFR-F3'97	CMS8jets'17	CMS8jets'17	CMS8jets'17
NMCrat'97	BCDMSp'89	HERAI+II'15	CMS7jets'14	CMS7jets'14	ATL8ZpT'16
CCFR-F3'97	CCFR-F3'97	BCDMSd'90	ATLAS7jets'15	E866pp'03	E866pp'03
E866pp'03	BCDMSd'90	BCDMSp'89	E866pp'03	ATLAS7jets'15	ATLAS7jets'15
BCDMSd'90	E605'91	CDHSW-F3'91	BCDMSd'90	BCDMSd'90	CCFR-F2'01
CDHSW-F3'91	E866pp'03	E866rat'01	CCFR-F3'97	BCDMSp'89	D02jets'08
CMS8jets'17	E866rat'01	CMS7Masy2'14	D02jets'08	D02jets'08	HERAc'13
E866rat'01	CMS8jets'17	NuTeV-nu'06	NMCrat'97	NMCrat'97	NuTeV-nub'06
LHCb8WZ'16	CDHSW-F3'91	CMS8jets'17	BCDMSp'89	CDHSW-F2'91	CCFR-F3'97

TABLE I: We list the top 10 experiments predicted to drive knowledge of the d/u PDF ratio and of the gluon distribution in the Higgs region according to PDFSense and LM scans. For both, we list the PDFSense evaluations based both on the CT14HERA2 fit and on a preliminary CT18pre fit in the first and second columns on either side of the double-line partition.

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, v.3]

Lagrange Multiplier scan: $g(0.01, 125\ GeV)$



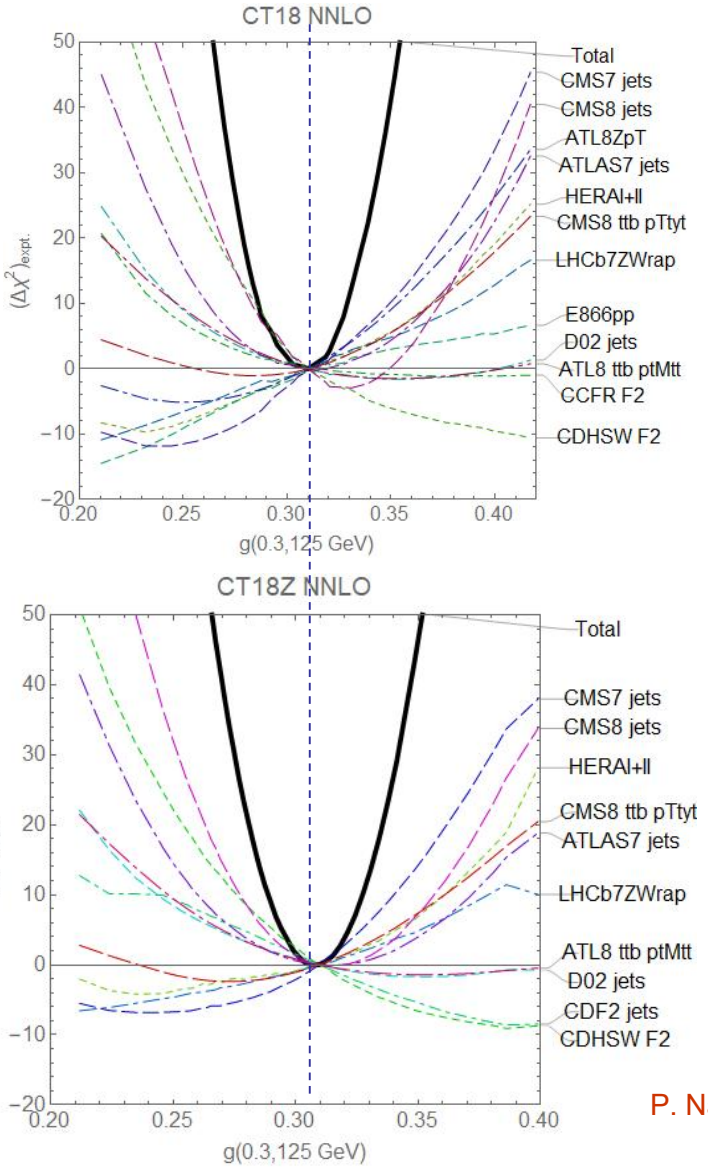
Upper row: CT18

- HERAI+II data set provides the dominant constraint, followed by ATLAS, CDF2, CMS, D02 jet production, HERA charm,…
- $t\bar{t}$ double-diff. cross sections provide weaker constraints

Lower row: CT18Z

- CT18Z: a 1% lower NNLO gluon in the Higgs production region than for CT14/CT18

Lagrange Multiplier scan: $g(0.3, 125\ GeV)$



Upper/lower rows: CT18/CT18Z

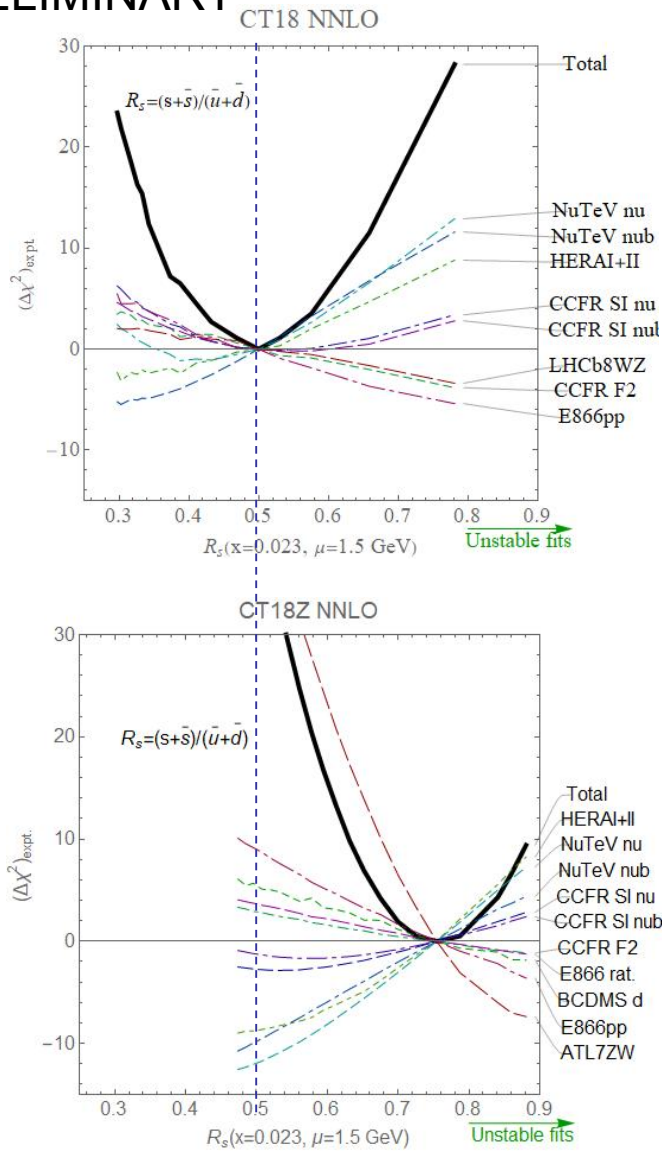
Good overall agreement. But observe opposite pulls from ATLAS7/CMS7 jet production and CMS8 jet production

Similarly, ATLAS $t\bar{t}$ distributions $\mathrm{d}^2\sigma/(dp_{T,t}dm_{t\bar{t}})$ and CMS $t\bar{t}$ distributions $\mathrm{d}^2\sigma/(dp_{T,t}dy_{t,ave})$ at 8 TeV impose weak opposite pulls

Constraints from ATLAS 8 Z p_T production data are moderate and still affected by NNLO scale uncertainty

Lagrange Multiplier scan: $\mathrm{R_s}(x = 0.023, \mu = \ 1.5\ GeV)$

PRELIMINARY

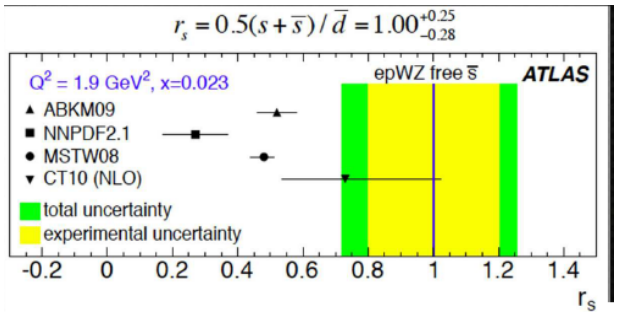


$$R_s(x,\mu) \equiv \frac{s(x,\mu) + \bar{s}(x,\mu)}{\bar{u}(x,\mu) + \bar{d}(x,\mu)}$$

Upper/lower rows: CT18/CT18Z

The CT18Z strangeness is increased primarily as a result of including the ATLAS 7 TeV W/Z production data (not in CT18), as well as because of using the DIS saturation scale in $m_c^{pole} = 1.4$ GeV

In either CT18 or CT18Z fit, observe instability in the fits for $R_s > 1$ at $x = 0.01 - 0.1$



What about *future experiments* **…like the LHeC?**

especially, in the context of other measurements at HL-LHC

- **LHeC** PDFSense projections by Tim Hobbs and Bo-Ting Wang
- Compared to **HL-LHC** projections by Gao, Harland-Lang, Rojo **[arXiv:1902.00134]**

a high-energy Electron-Ion Collider, Large Hadron-electron Collider

- an ep (eA) collider to achieve **high luminosities** > 1000 times that of HERA
  - access a wide range of x, including $x \sim 10^{-6}$
  - explore the dynamics of gluon saturation; greatly improve PDF precision; perform SM tests; and many other physics goals
- can perform a sensitivity analysis of Monte Carlo generated reduced NC/CC cross sections **[Klein & Radescu, LHeC-Note-2013-002 PHY]**

  60 GeV $\mathrm{e}^{\pm}$ on 1 or 7 TeV p
- pseudodata generated by randomly fluctuating about the PDF4LHC15 NNLO prediction according to putative LHeC uncorrelated errors – based on **100 fb-1** of data

now directly compare the LHeC vs. HL-LHC flavor sensitivities*

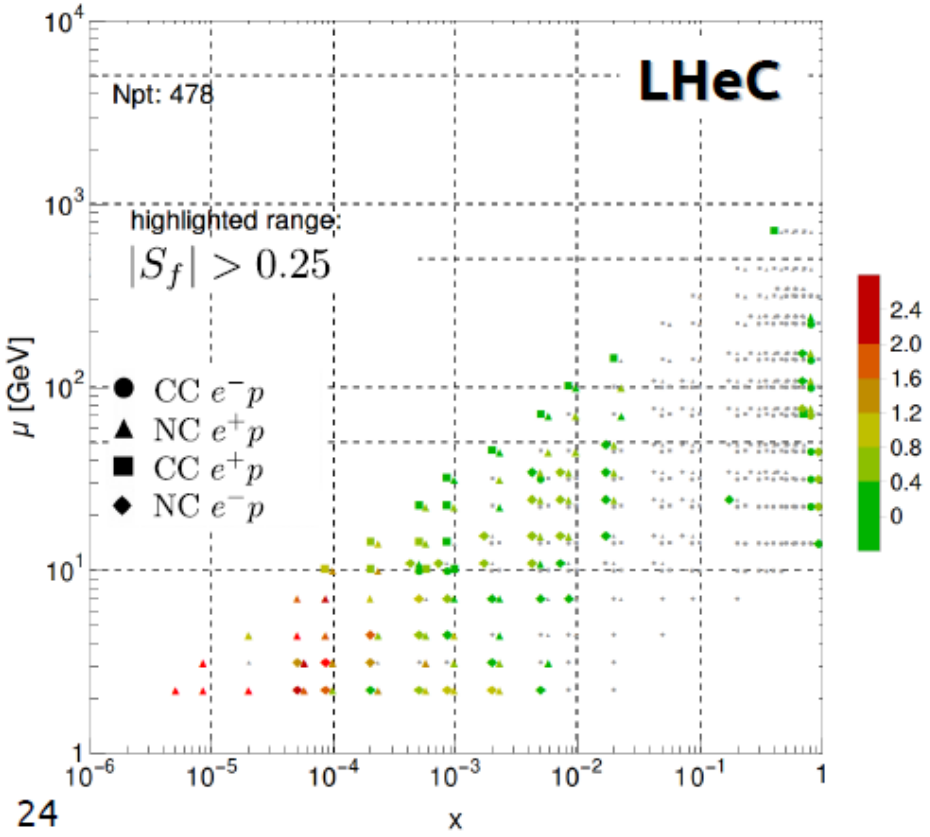
$g(x,\mu)$

*noting the much larger integrated luminosity of the HL-LHC pseudo-data

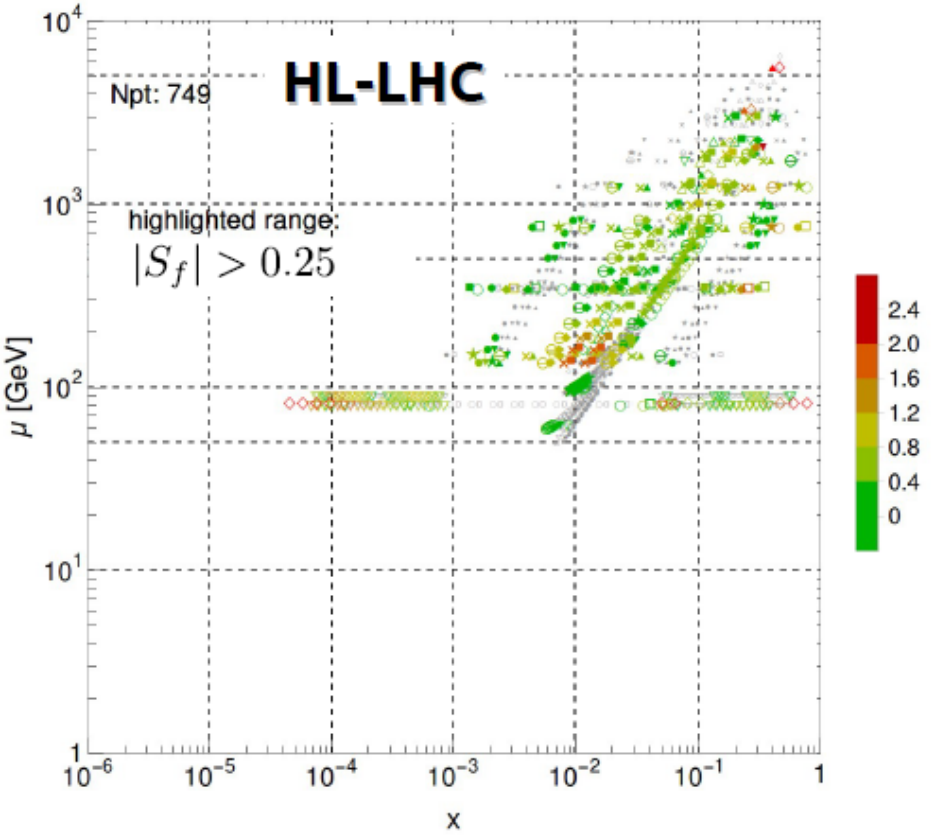
→ we compare both kinematic regions of especially strong sensitivity, and the aggregated impact of each experiment:

$$|S_g^{\mathrm{LHeC}}| = 151.4 \quad < \quad |S_g^{\mathrm{HL-LHC}}| = 244.6$$

| S_f | for g(x,μ), PDF4LHC15 NNLO



| S_f | for g(x,μ), PDF4LHC15 NNLO



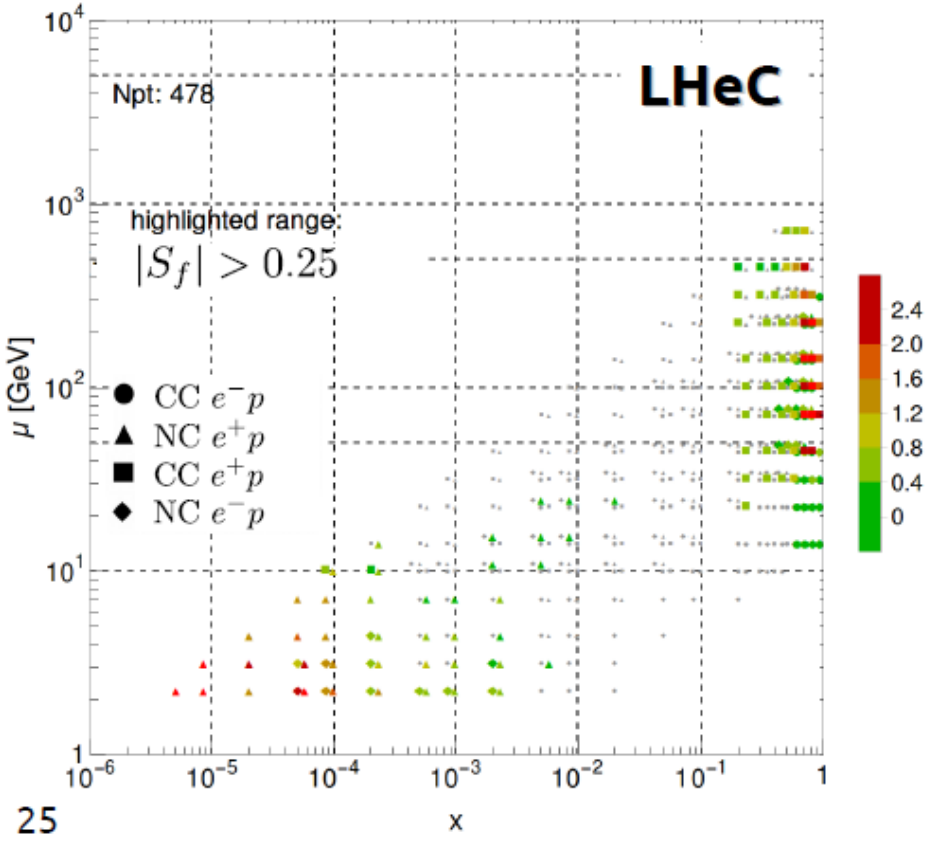
even with a small (pseudo)data set, LHeC enjoys strong sensitivity to down-type distributions!

$\mathrm{d}(x,\mu)$

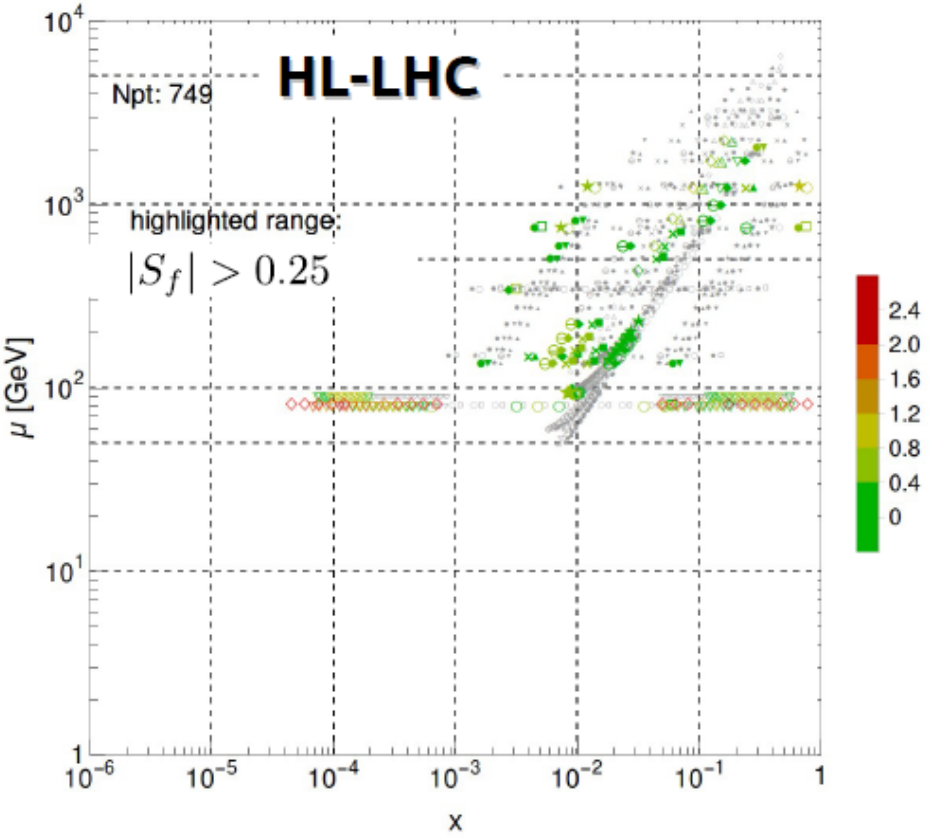
→ especially in a fashion complementary to HL-LHC, at very high/low x

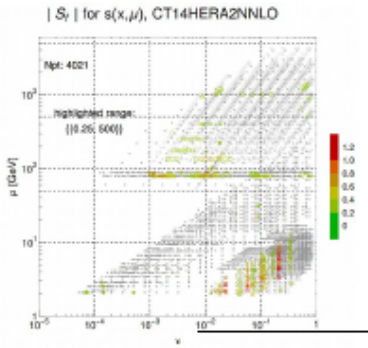
$$|S_d^{\mathrm{LHeC}}| = 214.3 \quad > \quad |S_d^{\mathrm{HL-LHC}}| = 170.8$$

| S_f | for d(x,μ), PDF4LHC15 NNLO



| S_f | for d(x,μ), PDF4LHC15 NNLO



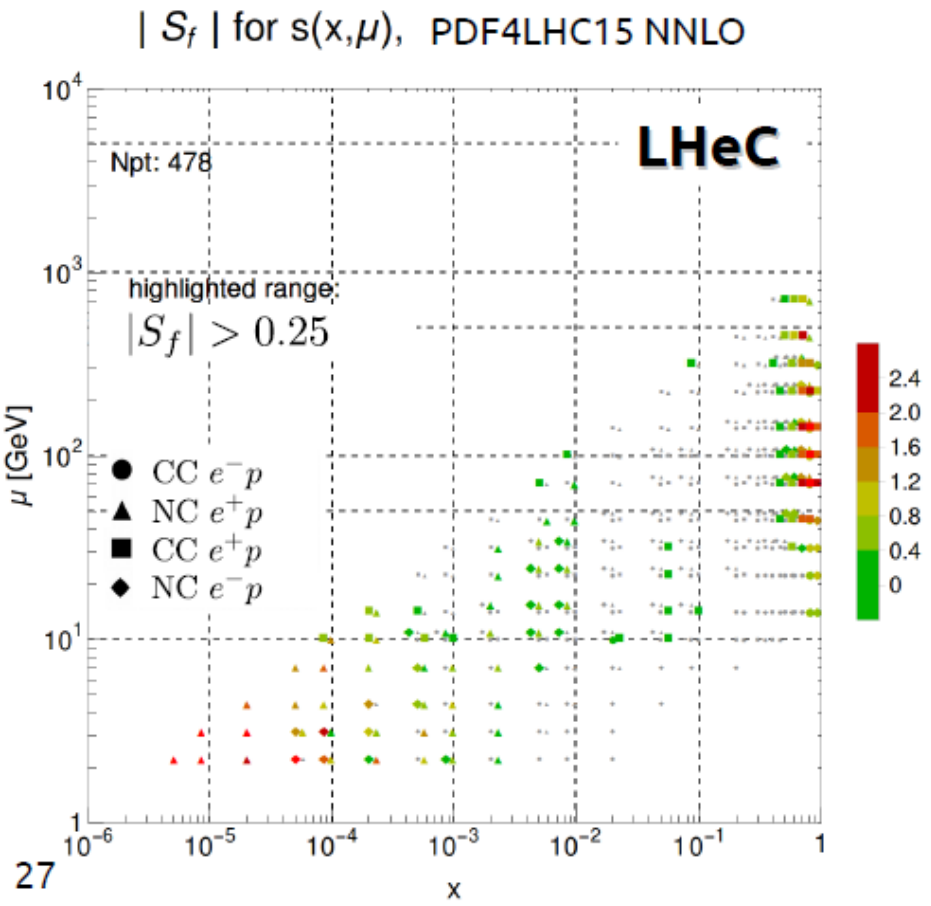


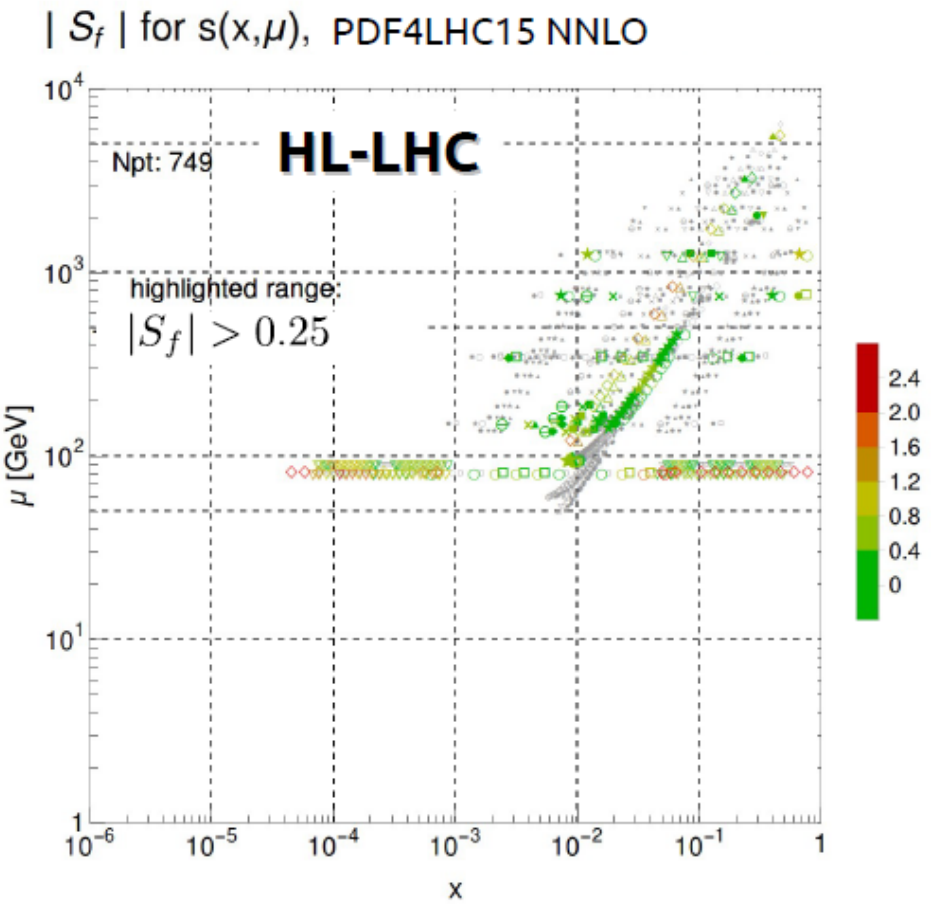
→ LHeC especially portends significantly heightened knowledge of nucleon strangeness

- CTEQ-TEA constraints come primarily through older fixed-target data and Tevatron data (and LHC Run I)

s(x,μ)

$$|S_s^{\mathrm{LHeC}}| = 214.1 \gtrsim |S_s^{\mathrm{HL-LHC}}| = 184.8$$





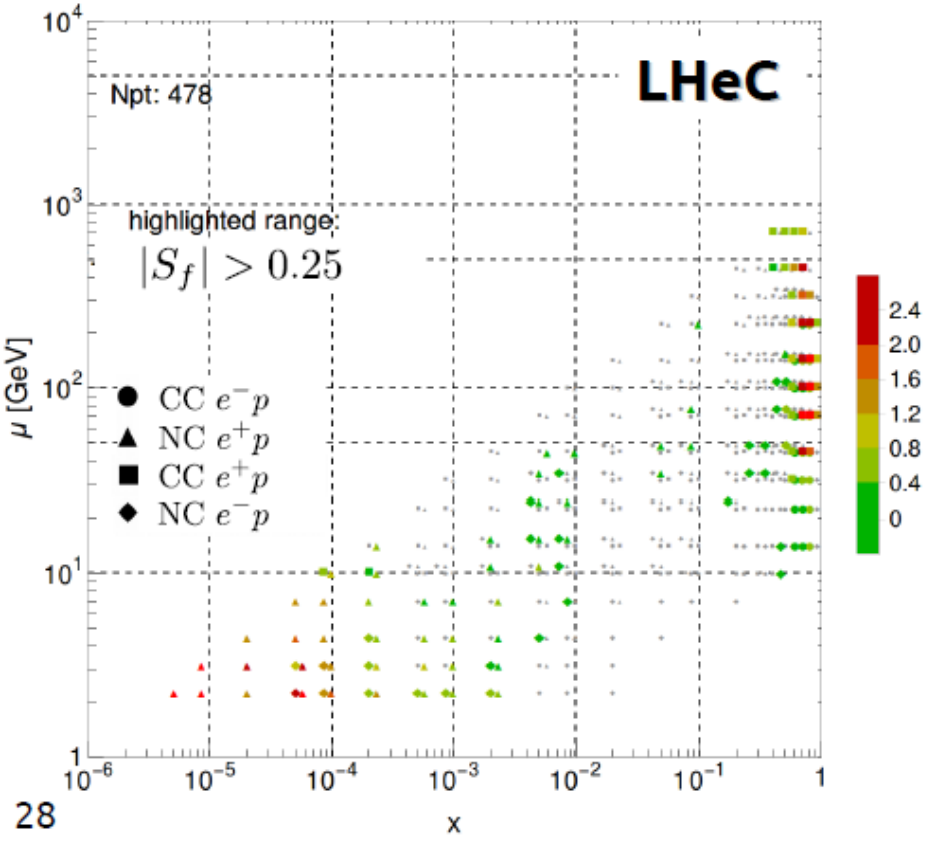
in the SU(2) quark sea, the LHeC 100 fb^{-1} set imposes constraints of magnitude comparable to HL-LHC

$\bar{d}(x,\mu)$

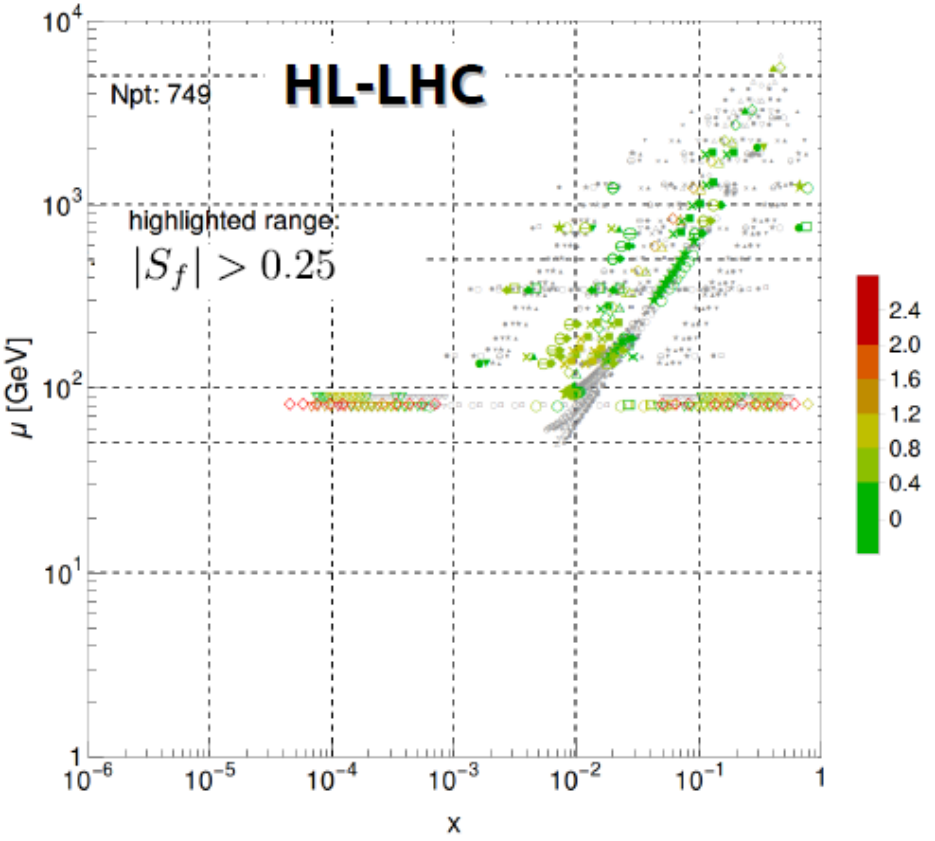
…these again predominate at the extrema of x

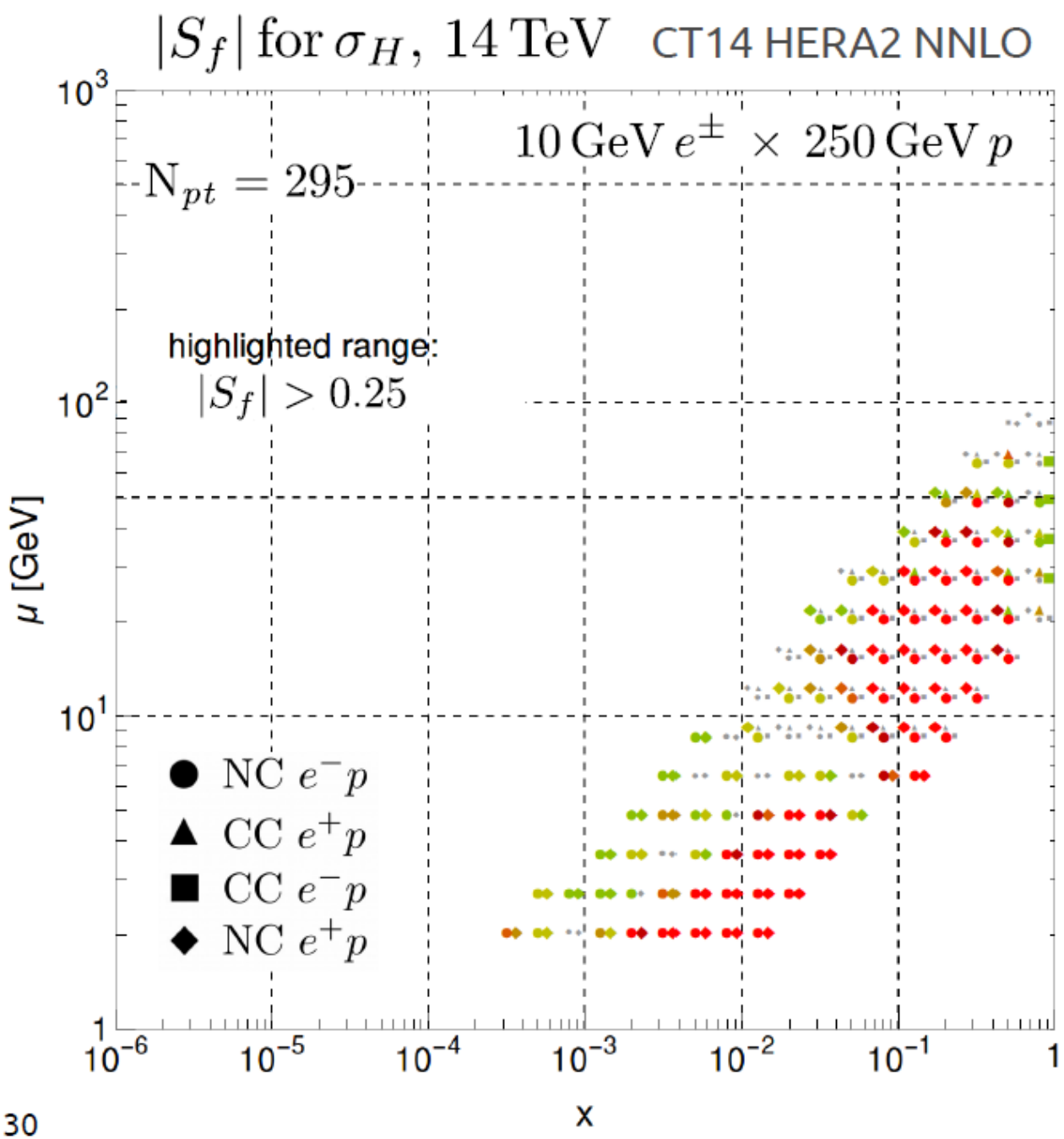
$$|S_{\bar{d}}^{\mathrm{LHeC}}| = 192.7 \quad \sim \quad |S_{\bar{d}}^{\mathrm{HL-LHC}}| = 199.4$$

| S_f | for $\bar{d}$(x,μ), PDF4LHC15 NNLO



| S_f | for $\bar{d}$(x,μ), PDF4LHC15 NNLO

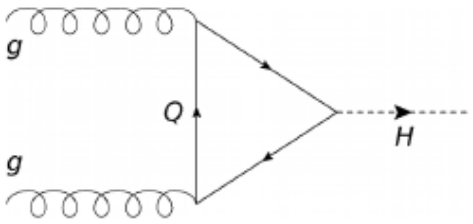




…a US-based EIC will also have important HEP consequences, e.g., on Higgs physics

- 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:



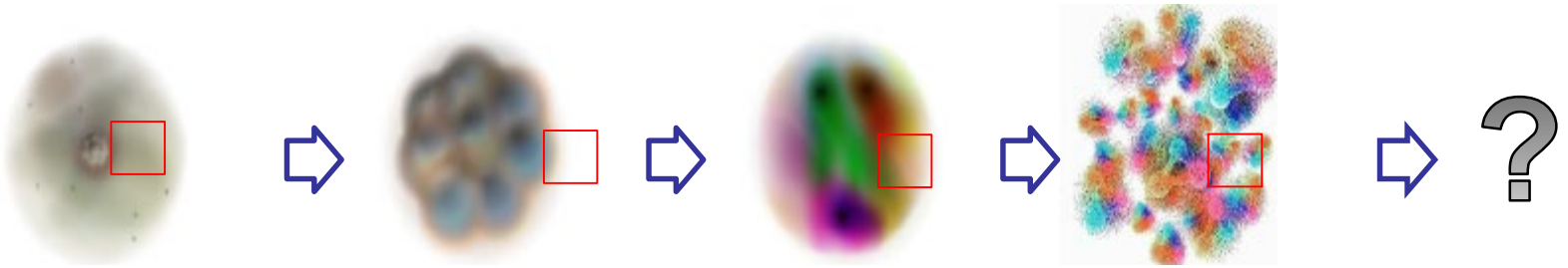
Outlook for CTEQ-TEA PDFs

- **CT18 PDF analysis is practically finished**

Detailed investigation of the LHC 7 and 8 TeV vector boson, jet, $t\bar{t}$ production data suggests mild changes in the central fits, PDF uncertainties, and precision EW observables, as compared to the CT14HERA2 NNLO data set. We notice the potential of the future ATLAS/CMS jet data, together with other LHC processes, for strengthening the constraints on the g, s, $\bar{u}$, and $\bar{d}$ PDFs with modest improvements in experimental systematics and full implementation of NNLO jet cross sections

- **CT14 PDFs** with photon PDFs **[1509.02905]**, intrinsic/fitted charm **[1706.00657]**, and Monte-Carlo error PDFs **[1607.06066]**
- NLO calculation for c, b **production at LHCb, ATLAS** in the S-ACOT-χ
- scheme using MCFM/Applgrid **[Campbell, P. N., Xie, in pre-publication]**
- Further development of programs for fast survey **[PDFSense]** and Hessian reweighting of the data **[ePump]**

The spirit of **meticulous** exploration

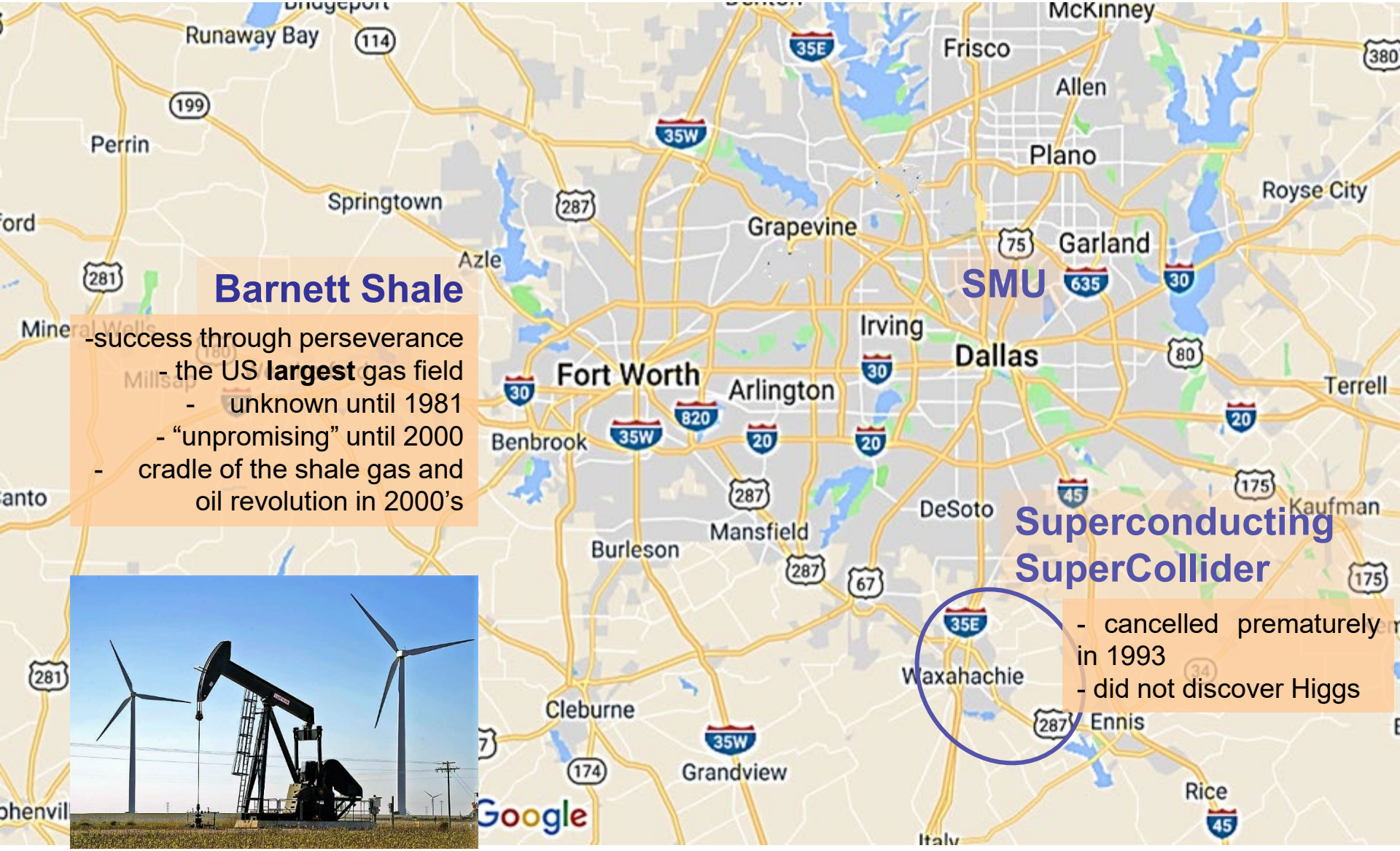


The global QCD analysis is a part of the unfolding scientific, social, and economical success of the HERA, Tevatron, and LHC colliders

Consistency of experimental measurements at the new (N)NNLO level of accuracy will be crucial for charting the TeV world at the high-luminosity LHC

Plentiful opportunities for testing new theoretical and statistical ideas. A unique project with multiple measurements of the same complex observables. **Unexpected** riches may lie even under a barren land ⇒ **Example: Barnett Shale**

Dallas-Fort Worth Metroplex



The future of the global QCD analysis is bright

Thank you!

Discriminating between two PDF fits based on the Bayes theorem

Given a new data set D ["$A_\ell(\mathrm{y}_\ell)$"], we can determine the ratio ρ of posterior likelihoods $P(T_i|D)$ of two fits T_1 and T_2:

$$\rho = \frac{P(T_2|D)}{P(T_1|D)} = \frac{P(D|T_1)\cdot P(T_1)}{P(D|T_2)\cdot P(T_2)}$$

- $P(D|T) \propto e^{-\chi^2(D,T)/2}$ is determined from the fit to D
- The prior $P(T)$ is determined by many theoretical considerations and past experimental measurements

Discriminating between two PDF fits based on the Bayes theorem

Given a new data set D ["$A_\ell(\mathrm{y}_\ell)$"], we can determine the ratio ρ of posterior likelihoods $P(T_i|D)$ of two fits T_1 and T_2:

$$\rho = \frac{P(T_2|D)}{P(T_1|D)} = \frac{P(D|T_1)\cdot P(T_1)}{P(D|T_2)\cdot P(T_2)}$$

- T_1 is very unlikely compared to T_2 iff $\rho \ll 1$
- ρ can be used to establish the PDF uncertainty, keeping in mind that
  - $P(D|T)$ need not be Gaussian for 1 experiment
  - modifications in the PDF functional form affect both $P(D|T)$ and $P(T)$
  - **Smoothness:** Models T_i must smoothly depend on model parameters (**"be natural"**) to reliably estimate P and ρ. [Unnatural models are not predictive.]

A shifted residual r_i

$r_i(\vec{a}) = \frac{T_i(\vec{a}) - D_i^{sh}(\vec{a})}{s_i}$ are N_{pt} **shifted residuals** for point i, PDF parameters $\vec{a}$

$\bar{\lambda}_\alpha(\vec{a})$ are N_λ **optimized nuisance parameters** (dependent on $\vec{a}$)

The $\chi^2(\vec{a})$ for experiment E is

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

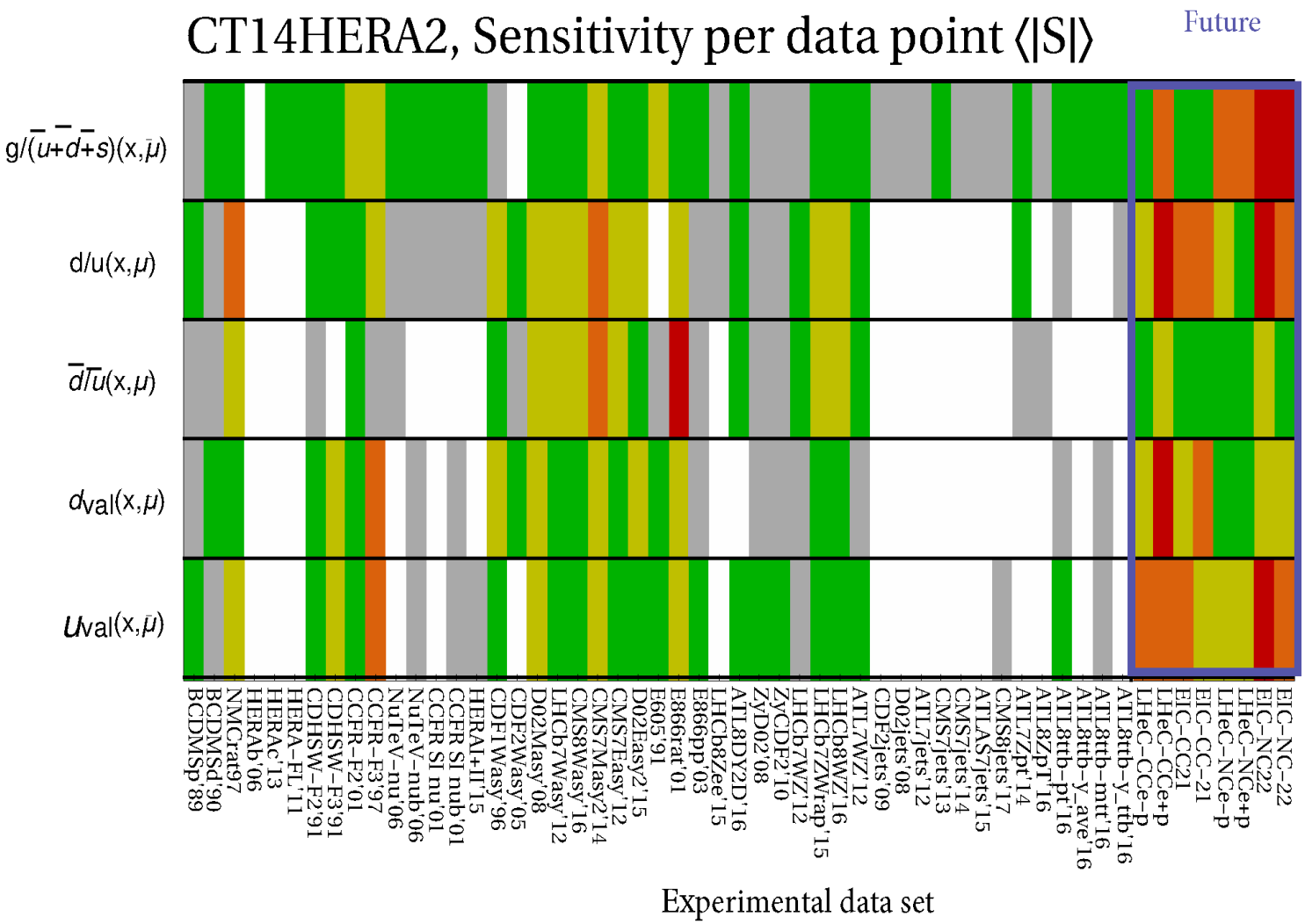
$T_i(\vec{a})$ is the theory prediction for PDF parameters $\vec{a}$
D_i^{sh} is the data value **including the optimal systematic shift**

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

s_i is the uncorrelated error

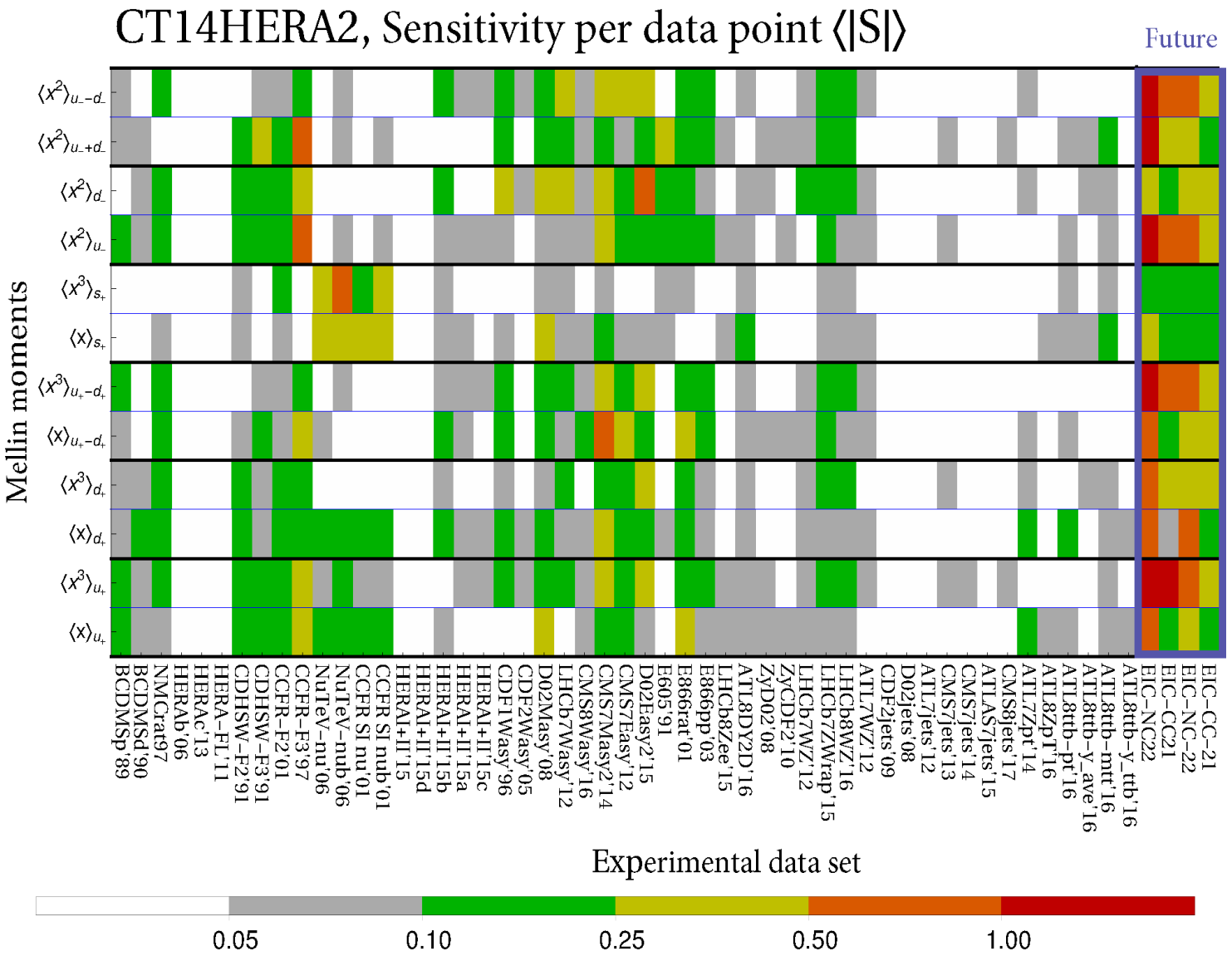
$r_i(\vec{a})$ and $\bar{\lambda}_\alpha(\vec{a})$ are tabulated or extracted from the cov. matrix

Sensitivity to PDF ratios

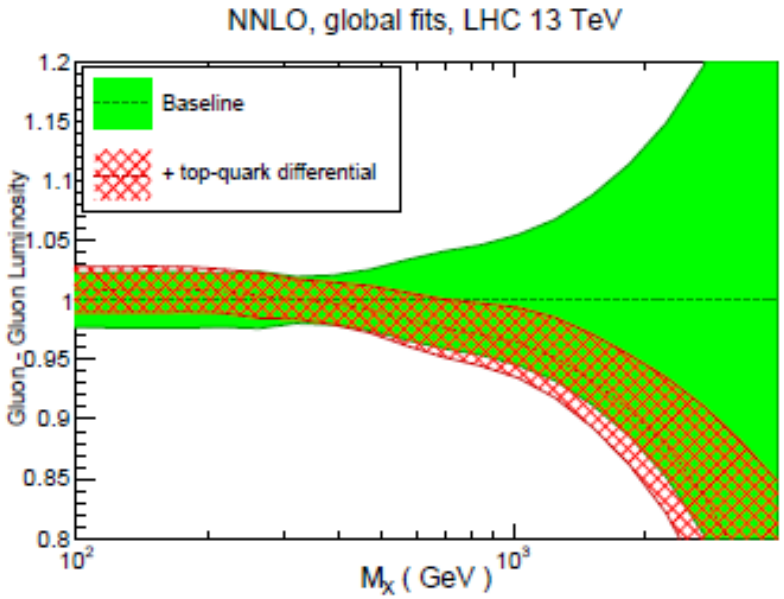


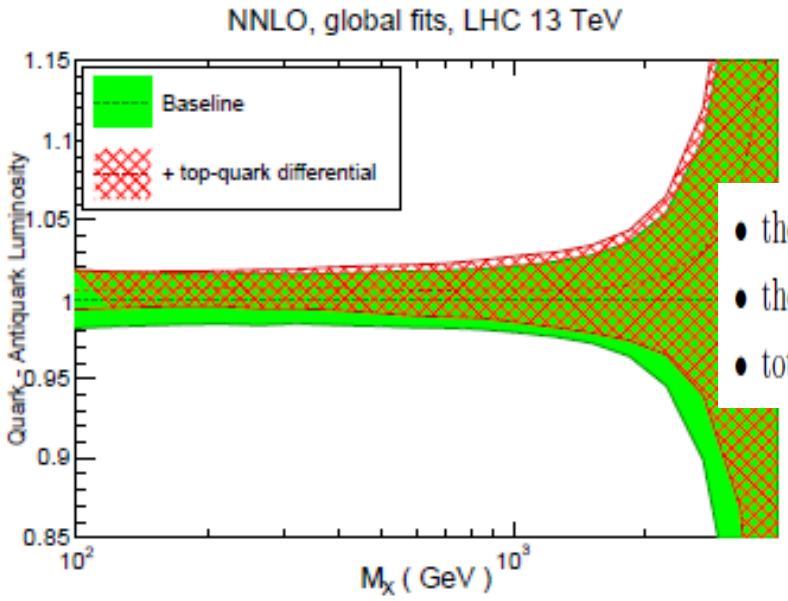
Sensitivity to $f_a(x_i, \mu_i)$, **per data point**

Sensitivity to Mellin moments



Sensitivity to $f_a(x_i, \mu_i)$, **per data point**





Pinning down the large-x gluon with NNLO $t\bar{t}$ differential distributions

Czakon, Hartland, Mitov, Nocera, Rojo, 1611.08609

Baseline global fit: no $t\bar{t}$ data, no inclusive jet data

"+top-quark differential" fit: add

- the normalized y_t distribution from ATLAS at $\sqrt{s} = 8$ TeV (lepton+jets channel),
- the normalized $y_{t\bar{t}}$ distribution from CMS at $\sqrt{s} = 8$ TeV (lepton+jets channel),
- total inclusive cross-sections at $\sqrt{s} = 7$, 8 and 13 TeV (all available data).