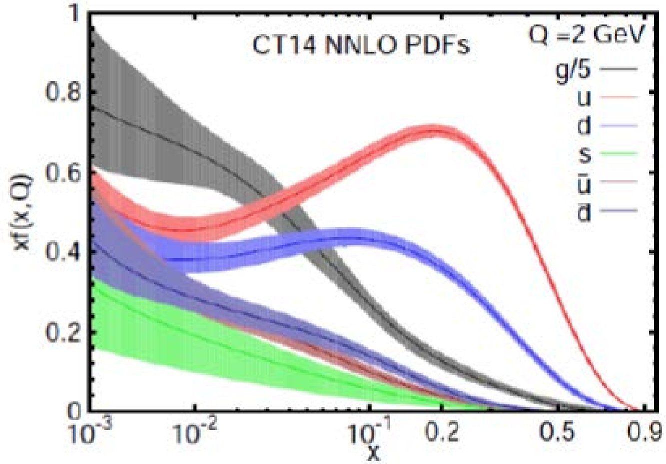
Parton distribution functions (PDFs) for Higgs boson studies



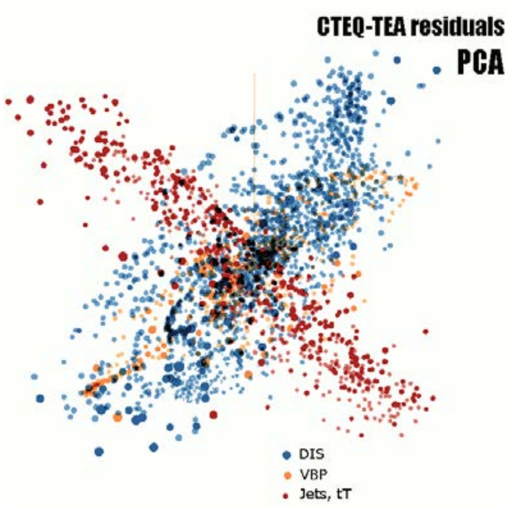
Pavel Nadolsky

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

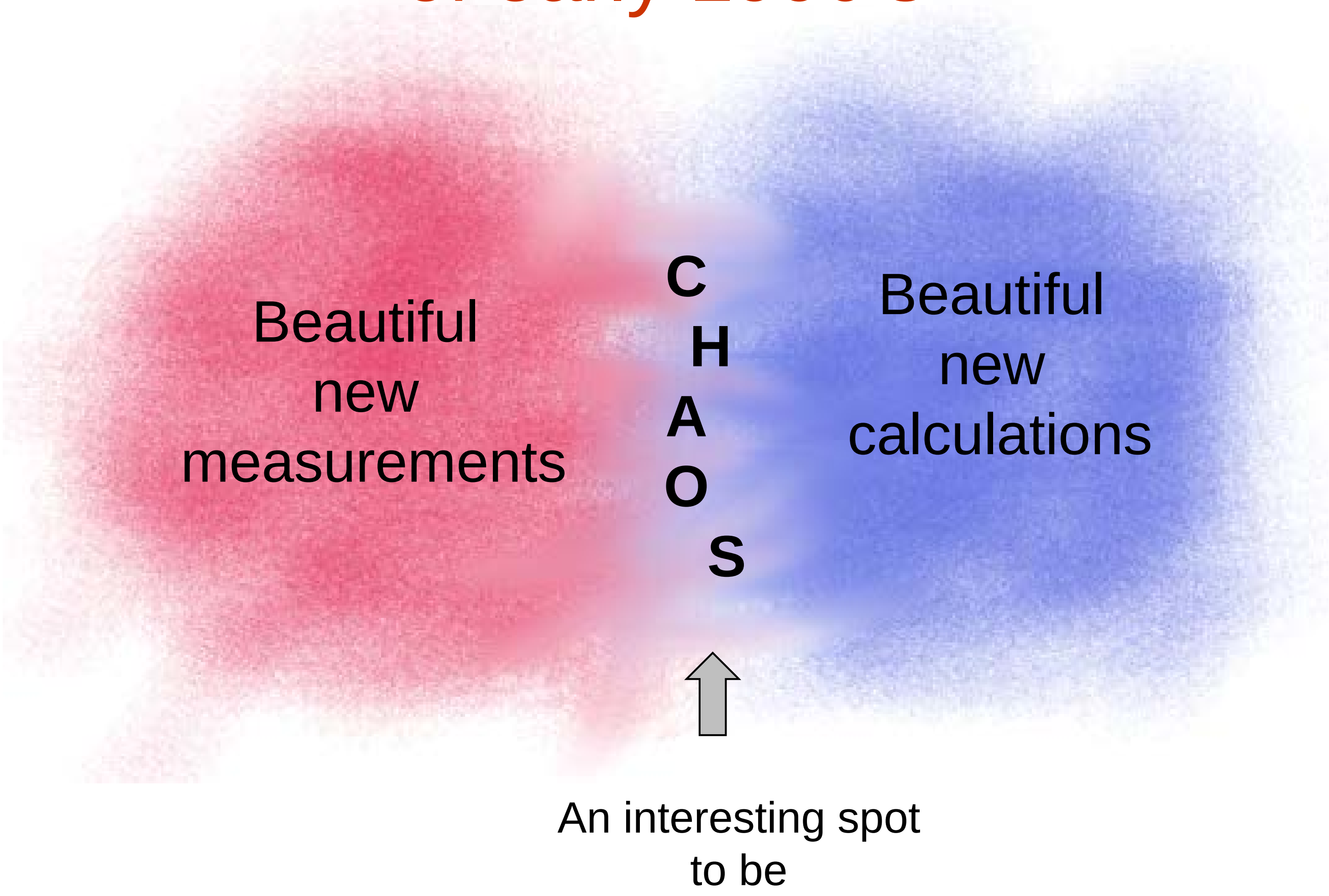
Southern Methodist University

and **CTEQ-TEA (Tung et al.) working group**

Kennesaw State U., Shanghai Jiao Tong U., Michigan State U., SMU, Xinjiang U.



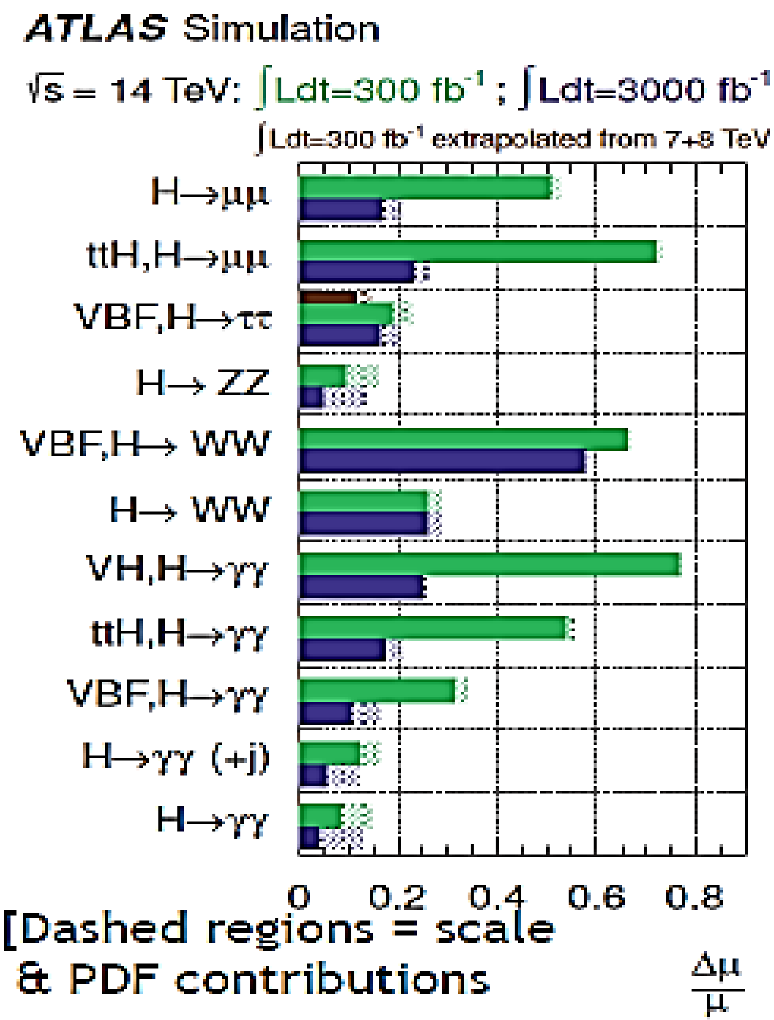
Today’s high-energy QCD is reminiscent of early 1990’s



QCD expectations for high-luminosity LHC

- New (N)NNLO calculations likely to be completed
- 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

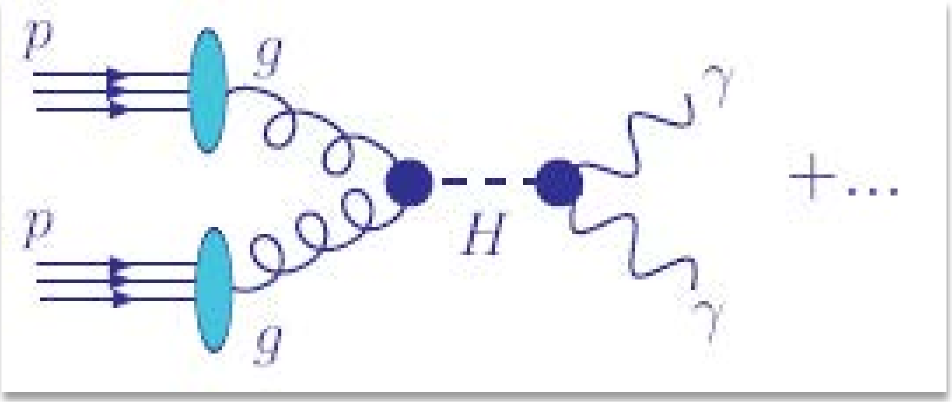
Projected Experimental Uncertainties



[Dashed regions = scale & PDF contributions]

P. Newman, DIS'2016

Example: total cross section for gg → Higgs → $\gamma\gamma$



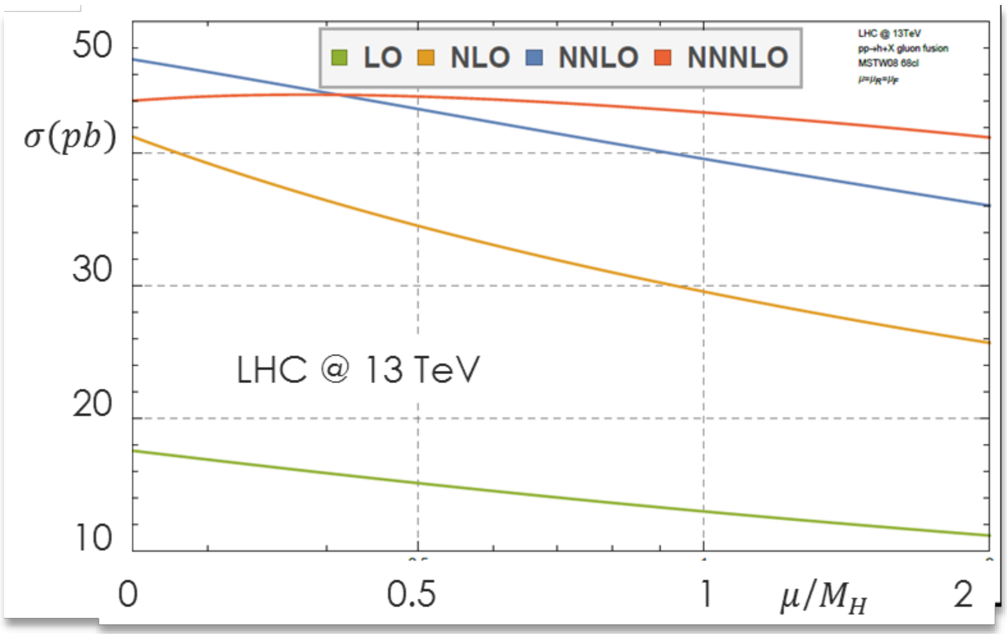
$$\sigma_{pp\to H\to\gamma\gamma X}(Q) = \sum_{a,b=g,q,\bar{q}} \int_0^1 d\xi_a \int_0^1 d\xi_b \hat{\sigma}_{ab\to H\to\gamma\gamma}\left(\frac{x_a}{\xi_a}, \frac{x_b}{\xi_b}, \frac{Q}{\mu_R}, \frac{Q}{\mu_F}; \alpha_s(\mu_R)\right)$$
$$\times f_a(\xi_a, \mu_F) f_b(\xi_b, \mu_F) + O\left(\frac{\Lambda_{QCD}^2}{Q^2}\right)$$

$\hat{\sigma}_{ab\to H\to\gamma\gamma}$ is the perturbative cross section to produce a Higgs boson from partons a and b; $a, b = g, u, \bar{u}, d, \bar{d}, \ldots$

$f_a(\xi, \mu)$ is a parton distribution function (PDF) associated with the probability for finding a parton a with the "+" momentum ξp^+ in a proton with the "+" momentum p^+ for $p^+ \to \infty$, at a factorization scale $\mu > 1$ GeV

Hard-scattering cross sections for $gg \to H \to \gamma\gamma$

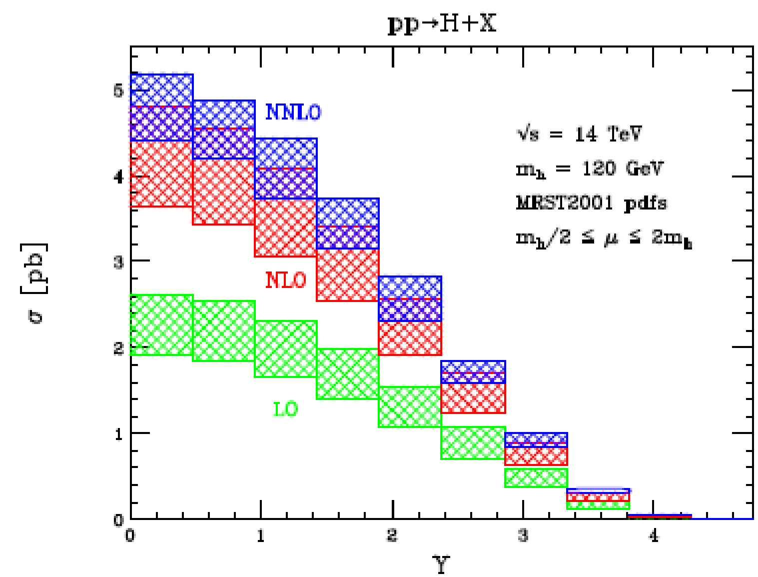
N3LO for total cross sections



Anastasiou, Duhr, Dulat, Herzog, Mistlberger, 1503.06056

N3LO corrections are of the order of +2.2%. The total scale variation at N3LO is 3%

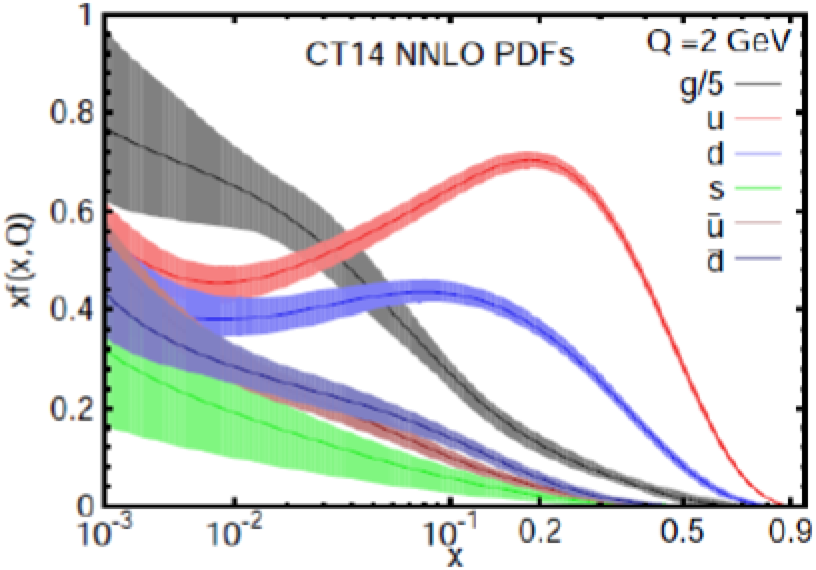
NNLO for differential distributions



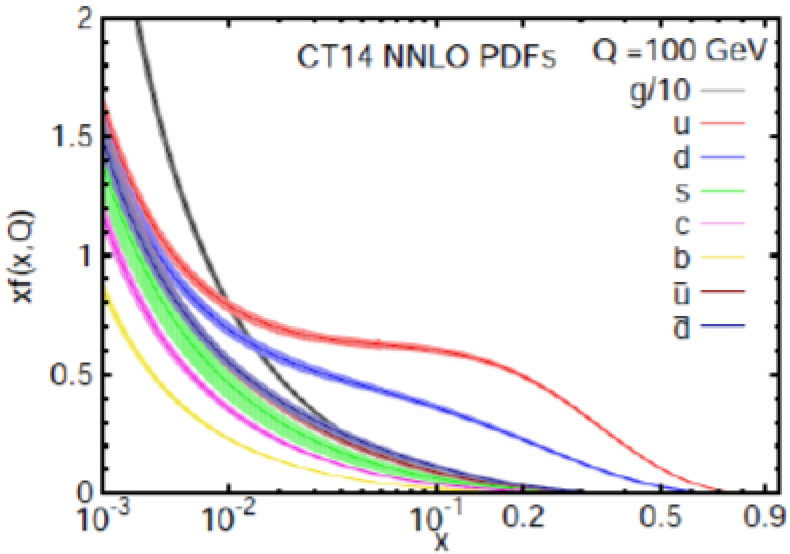
Anastasiou, Melnikov, Petriello, hep-ph/0409088, 0501130

Recent CT14 PDFs

(S. Dulat et al., arXiv:1506.07443)



Q= 2 GeV



Q= 100 GeV

Phenomenological parametrizations of PDFs are provided with estimated uncertainties of multiple origins (**uncertainties of measurement, theoretical model, parametrization form, statistical analysis**, …)

The shape of PDFs is optimized w.r.t. hundreds of **nuisance parameters**

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 (→ 18p)
MMHT'14 (→ 16)
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

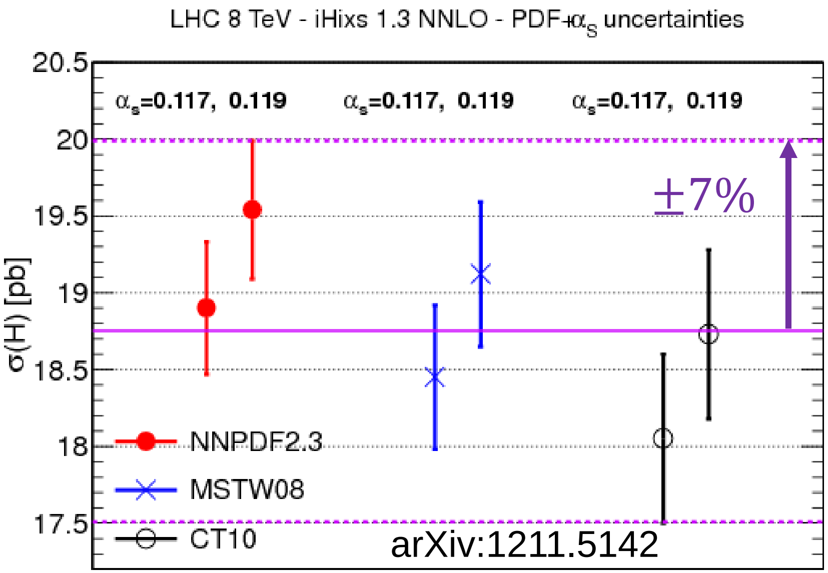
SMU theory contributions

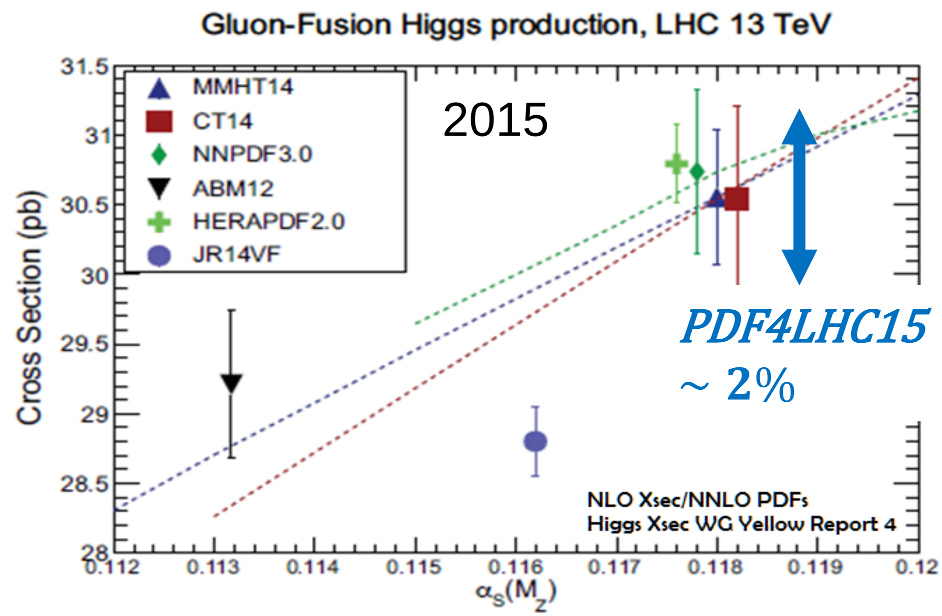
Toward a new generation of PDFs

[“CT18preliminary” PDFs]

What can the LHC do to constrain the PDFs?

Example: $gg \to H^0_{SM}$ at the LHC





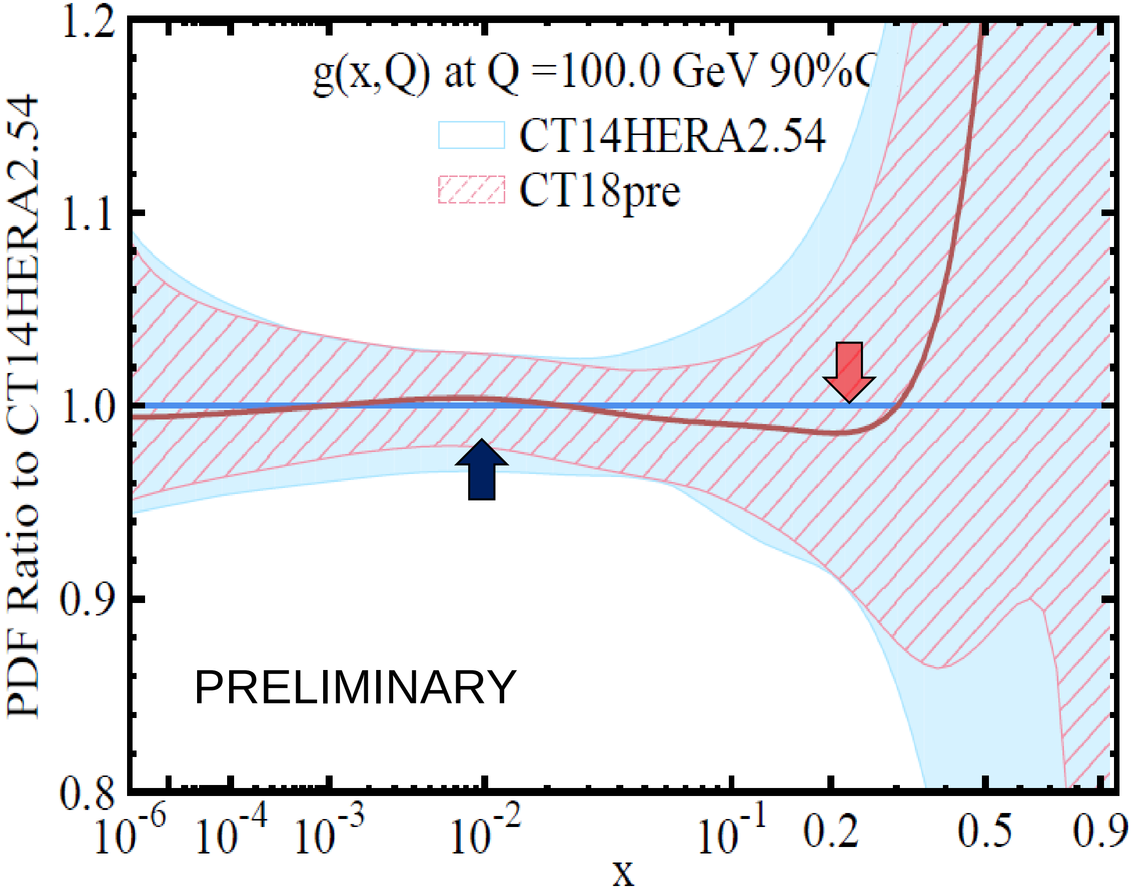
2012->2015: Uncertainty δ_{PDF} on Higgs cross sections based on 3 global fits has reduced from 7% to within 3%, i.e., the PDF uncertainty is now of order of N3LO QCD scale uncertainty

This improvement is due to benchmarking of general-mass factorization schemes and development of new PDF-averaging methods with active participation of SMU's Jun Gao **(2018 Altarelli Award)**.

Can we further improve on this?

Gluon PDF before and after including the LHC data

[CT14HERA2 vs. CT18pre NNLO]



$x \approx 0.01$: $g(x,Q)$ **mildly** increases within the uncertainty

⇒ **slightly** larger Higgs production rates at 14 TeV

Minor reduction in the gluon PDF uncertainty

$0.05 \lesssim x \lesssim 0.3$: $g(x,Q)$ **mildly** decreases; lower gg luminosities for $M_X > 700$ GeV

After the fit

CT18p: preliminary PDFs with new LHC data

Issues:

- Experimental, theoretical, and procedural **systematic** uncertainties dominate the PDF uncertainty in many cases
  - Tensions between some experimental data sets
  - Large QCD uncertainties in some kinematic regions (e.g., large y)

The CT18pre analysis examines how the PDFs depends on…

… settings of NNLO calculations

(SACOT-χ heavy-quark scheme, QCD scales, m_c, numerical codes,…)

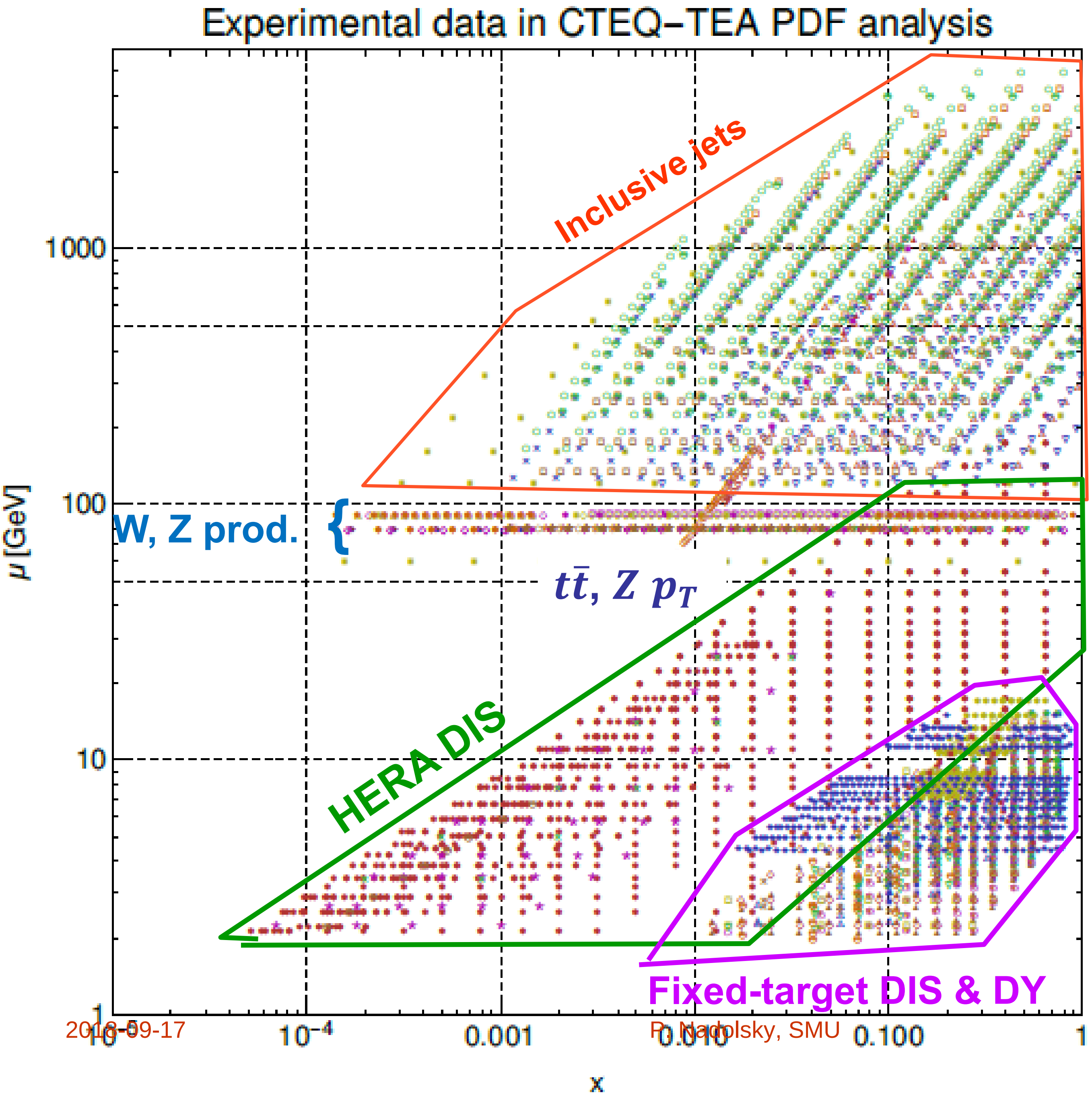
… selection of experiments and kinematic cuts

For instance, $g(x,Q)$ at $x > 0.05$ is already constrained in CT14/MMHT14 by D0 Run-2 incl. jet data which is not in NNPDF3.1. Disagreements exist within available ATLAS/CMS experiments and between some LHC and non-LHC experiments

…the fitting procedure

Definition of PDF uncertainties
Parametrization forms
PDF error analysis (Hessian vs. Monte-Carlo)

…



CT18pre analysis includes **new** LHC experiments on W/Z, high-p_T Z, jet, $t\bar{t}$ production

(25% more data)

Selected using fast statistical tools **PDFSense** and **ePump**

Experiments in the CT14 HERA2 fit

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, $Q^3 d^2\sigma_{pp}/(dQ dx_F)$	[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

New experiments in the CT17pre fit

1. LHCb 7 TeV Z/W muon rapidity 1505.07024
2. LHCb 8 TeV Z rapidity 1503.00963
3. CMS 8 TeV W lept. asymmetry 1603.01803
4. LHCb 8 TeV Z/W muon rapidity 1511.08039
5. ATLAS 7 TeV Z p_T 1512.02192
6. CMS incl. jet 7 TeV, R=0.7 1406.0324
7. ATLAS incl. jet at 7 TeV, R=0.6 1410.8857
8. CMS incl. jet at 8 TeV, R=0.7 1609.05331
9. ATLAS 8 TeV $t\bar{t}$ p_T 1511.04716
10. ATLAS 8 TeV $t\bar{t}$ $m_{t\bar{t}}$ 1511.04716
11. CMS 8 TeV $t\bar{t}$ $d^2\sigma/dp_{Tt}dy_t$ 1703.01630

N_d is the number of data points

How sensitive is an experiment to a PDF? Can we estimate it **before** doing the global fit?

Yes, using new statistical tools:

1. Generalized correlations (**sensitivities** S_f) comparing experimental and PDF uncertainties for fitted data points

PDFSense

(1803.02777)

http://metapdf.hepforge.org/PDFSense

2. PDF reweighting
3. Hessian profiling

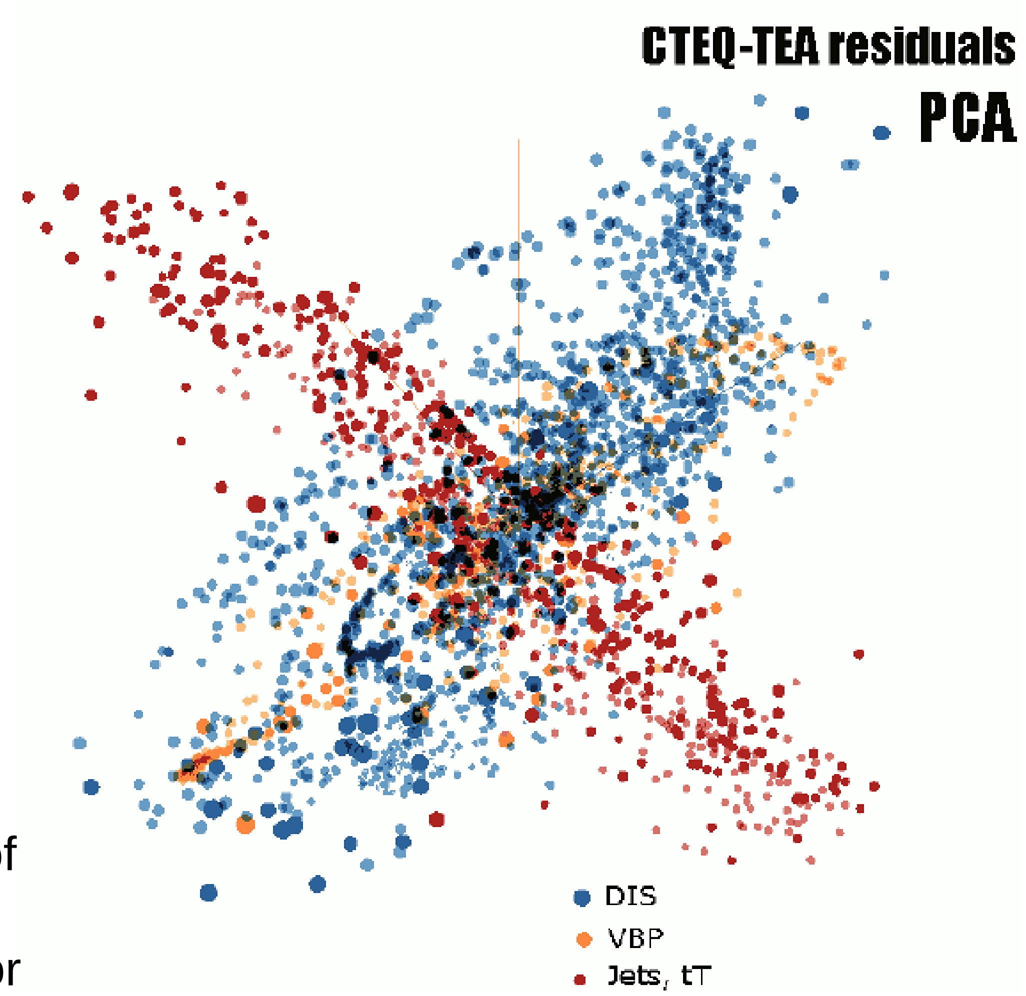
} $\Rightarrow$ mcgen (1607.06066)
ePump (1806.07950)

Manifolds of data residuals [TensorFlow]

Analysis flow:

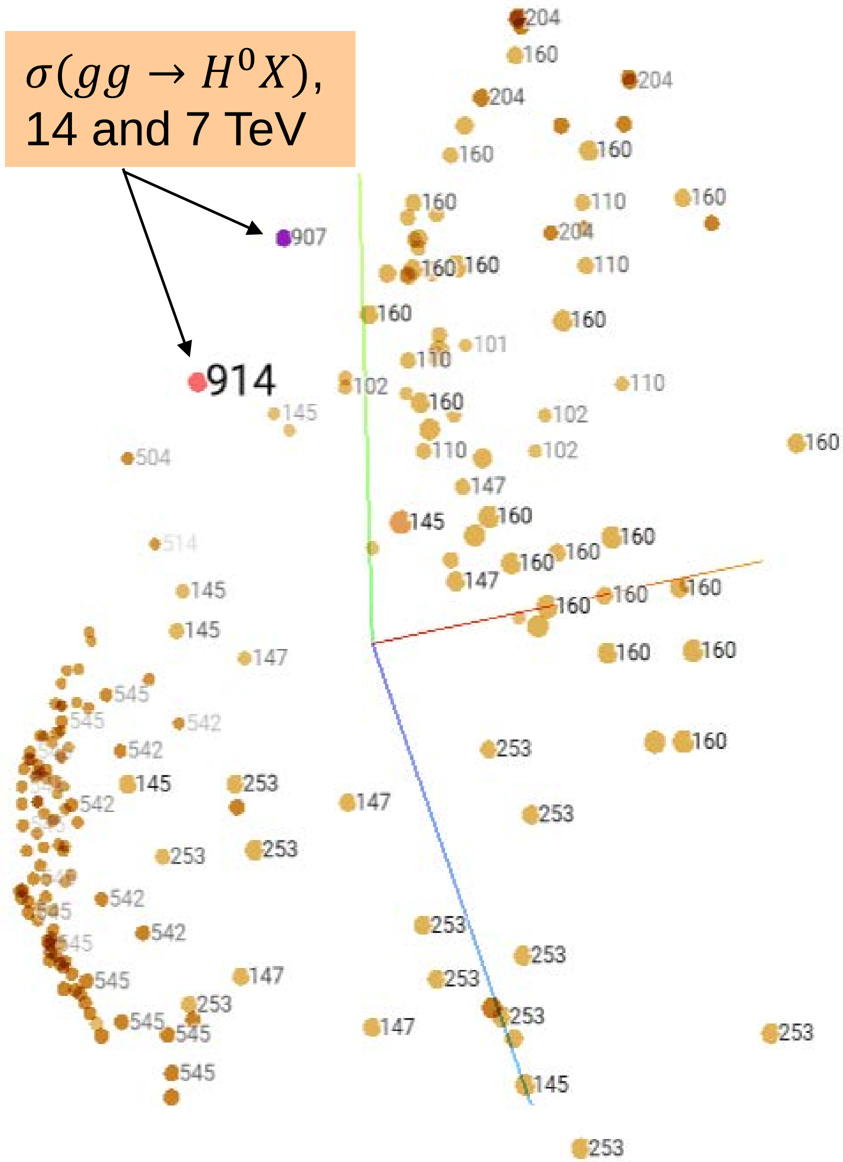
- We give you a table of N_{pt} normalized data point residuals $\vec{r}_i(\vec{a})$ for every CT14HERA2 error PDF **[on the PDFSense website]**

- You examine the 56-dim. distribution of $\vec{r}_i(\vec{a})$ in PDFSense or another data analysis software



Right: a sample 3-dim. projection of the 56-dim. manifold obtained with the TensorFlow Embedding Projector (http://projector.tensorflow.org)

A PDF-dependent quantity f, such as the Higgs cross section at 7 or 14 TeV (ID=907, 914), defines a direction $\vec{\delta}_f$ in the (2)N-dim space.



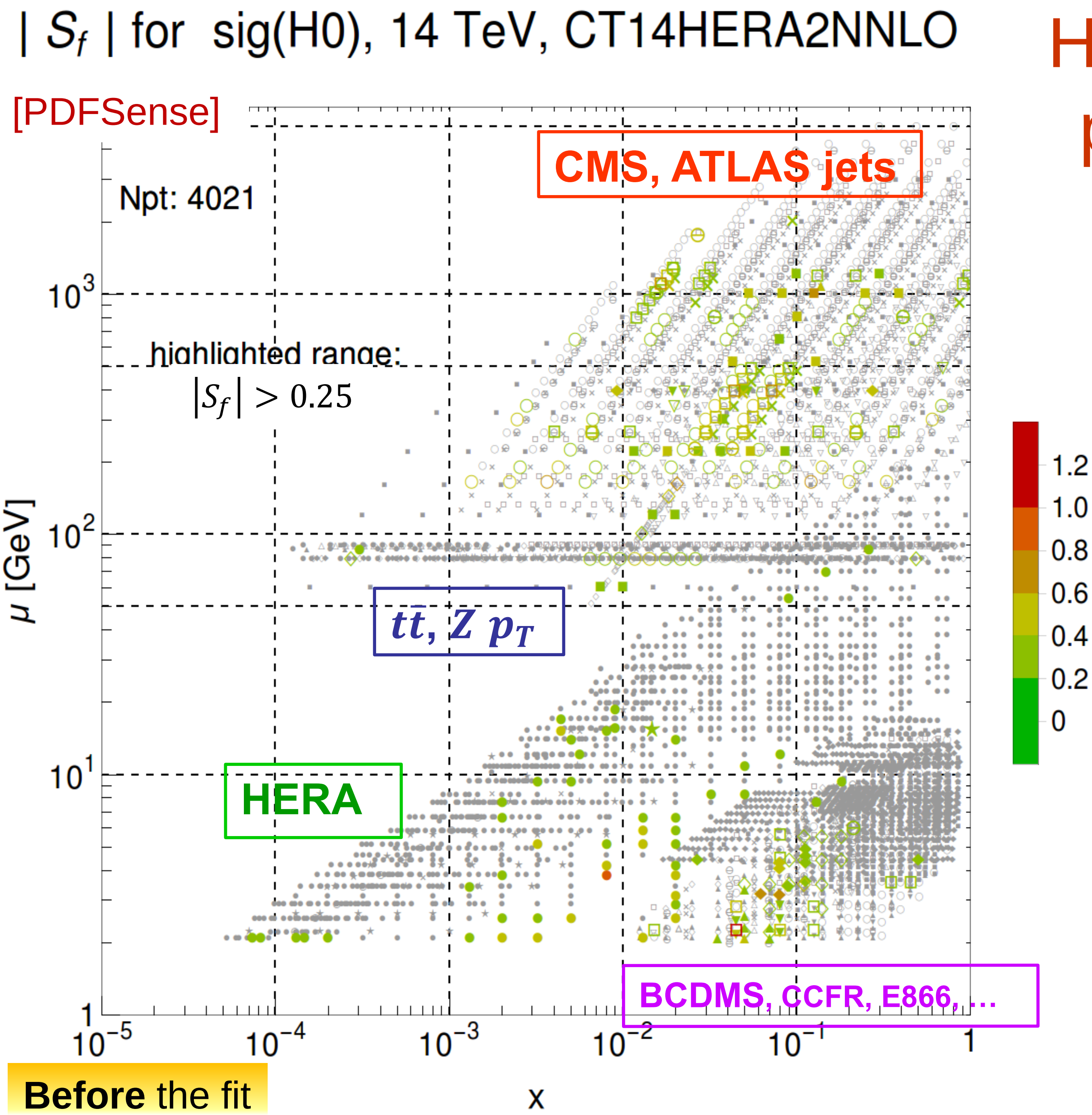
The net constraint of the i-th point on $\sigma(H)$, including systematic errors, is quantified by the projection of $\vec{r}_i$ on $\vec{\boldsymbol{\delta}}_{\boldsymbol{f}}[\sigma(H)]$, called the sensitivity $S_{f,i}$.

Right: 300 vectors $\vec{r}_i$ of the CT14HERA2 global set whose directions are closest to $\vec{\delta}_f(\sigma(H^0))$. **These vectors are given by the experiments:**

160=HERA I+II; 101, 102=BCDMS; 110=CCFR F2p; 147, 145=HERA I+II c, b; 204=E866 σ_{pp}; 253=Z p_T 8 TeV; 542, 545=CMS jets 7, 8 TeV; 504, 514=Tevatron jets

Sensitivity of expt E = sum of $S_{f,i}$ over data points in E

Higgs boson production

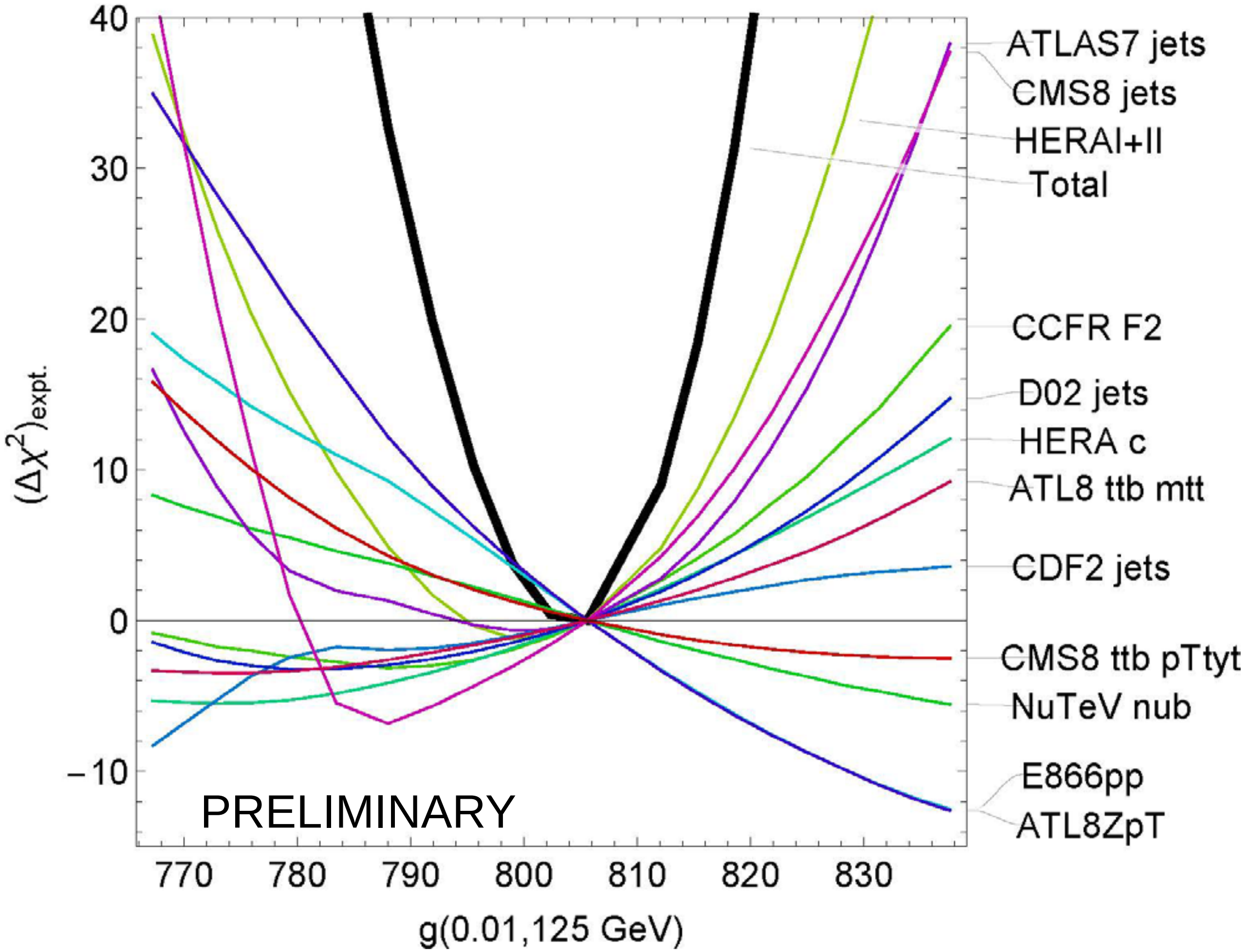


HERA DIS still has the dominant sensitivity!

CMS 8 TeV jets is the next expt. after HERA sensitive to $\sigma_H(14$ 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

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



A Lagrange multiplier scan [Stump et al., hep-ph/0101151] of

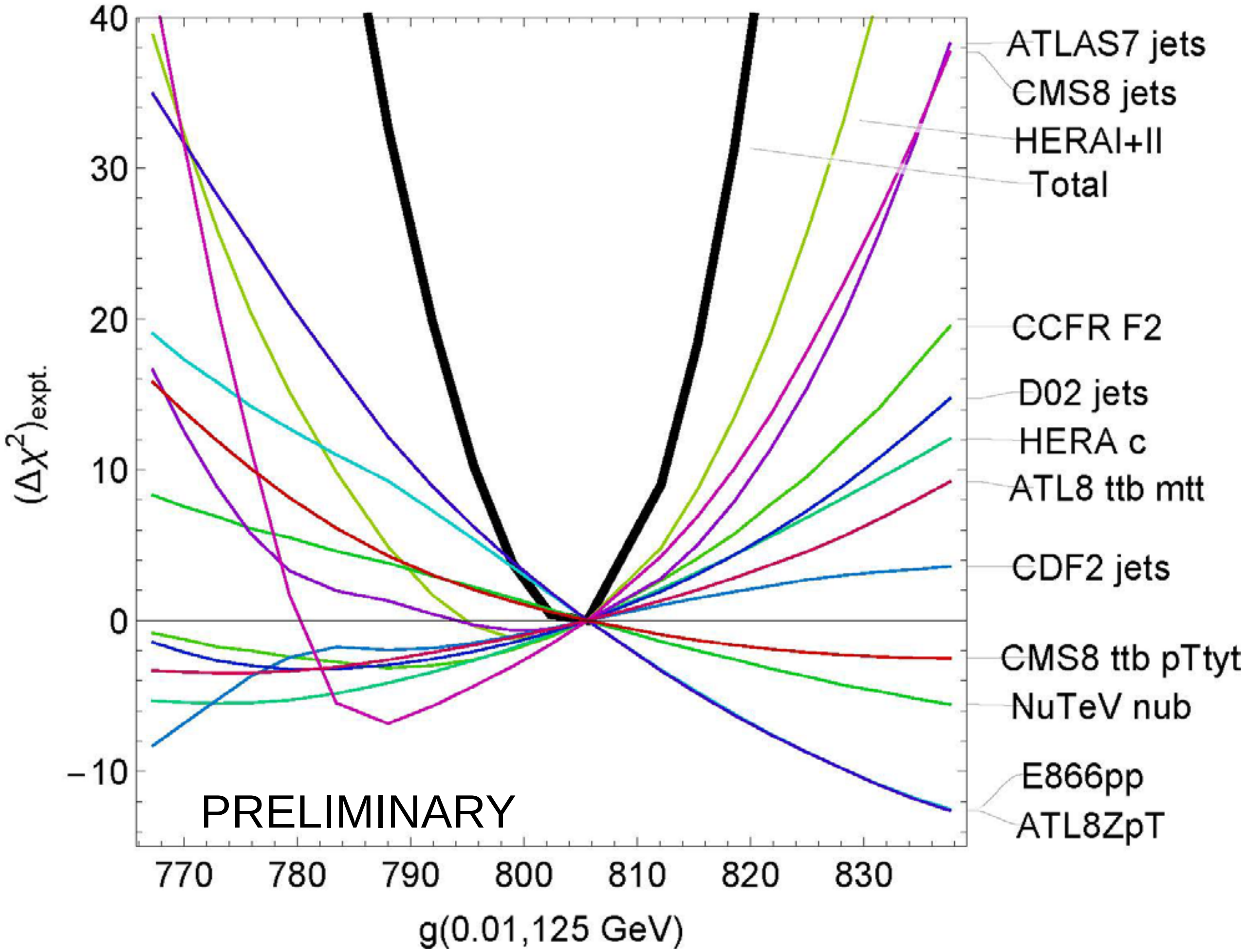
$$\Delta\chi^2 = \chi^2(g) - \chi^2_{best-fit}$$

for all (black line) and individual (colored lines) experiments

Best-fit $g(0.01{,}125\text{GeV})$=806

After the fit

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



The LM scans broadly confirm S_f estimates

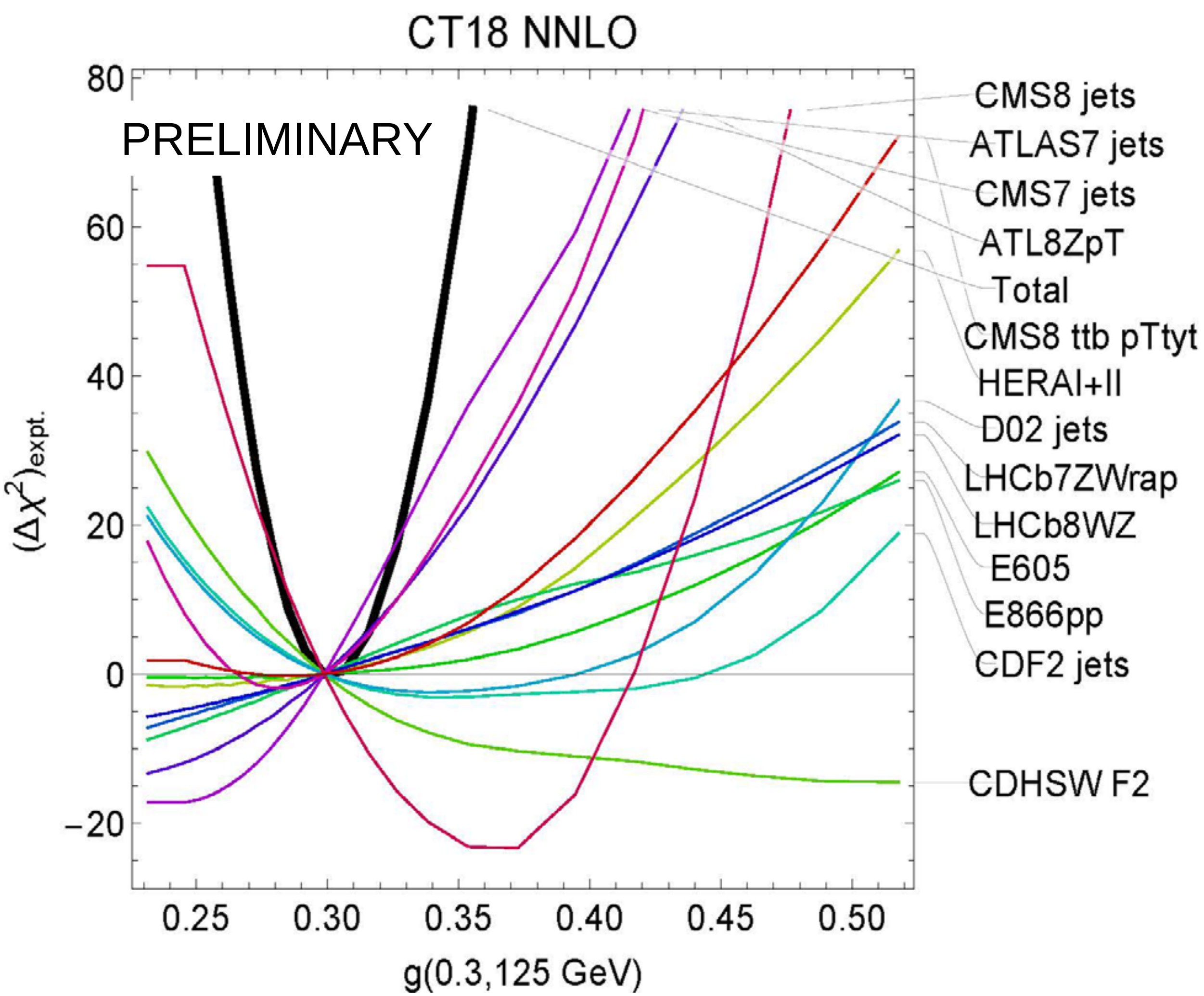
HERAI+II, ATLAS7 jets, CMS8 jets impose the tightest constraints; are in agreement

E866, ATLAS 8 Z p_T prefer higher gluon

After the fit

Which experiments constrain the gluon?

$x = 0.3, Q = 125$ GeV

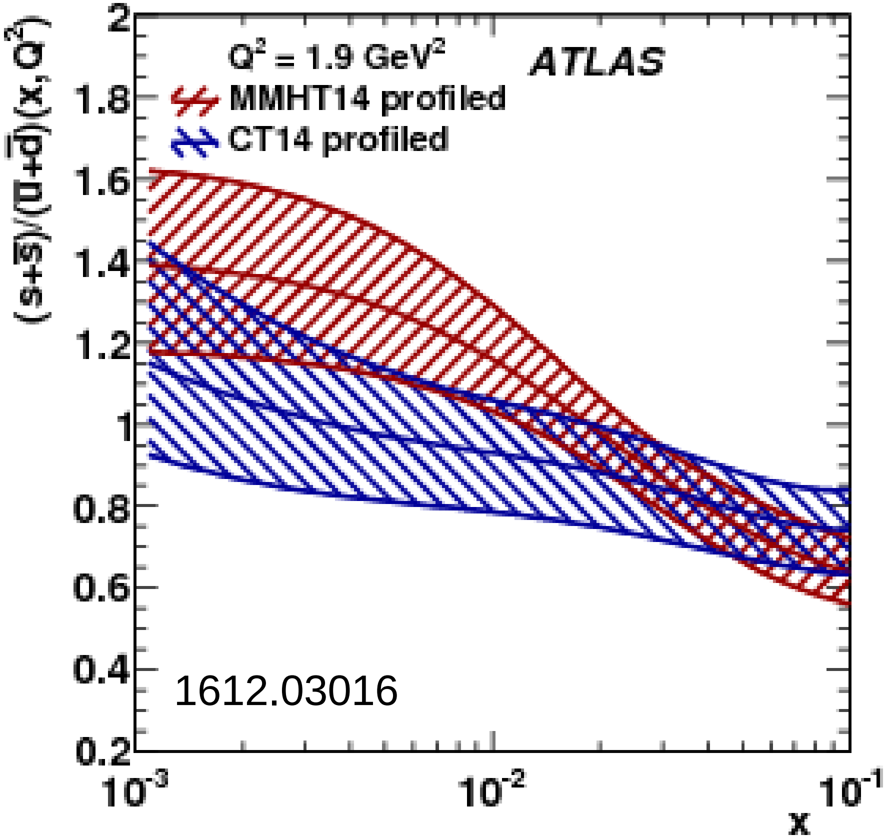


ATLAS 7 and CMS 7 TeV jets, ATLAS 8 Z p_T disagree with CMS8 jets

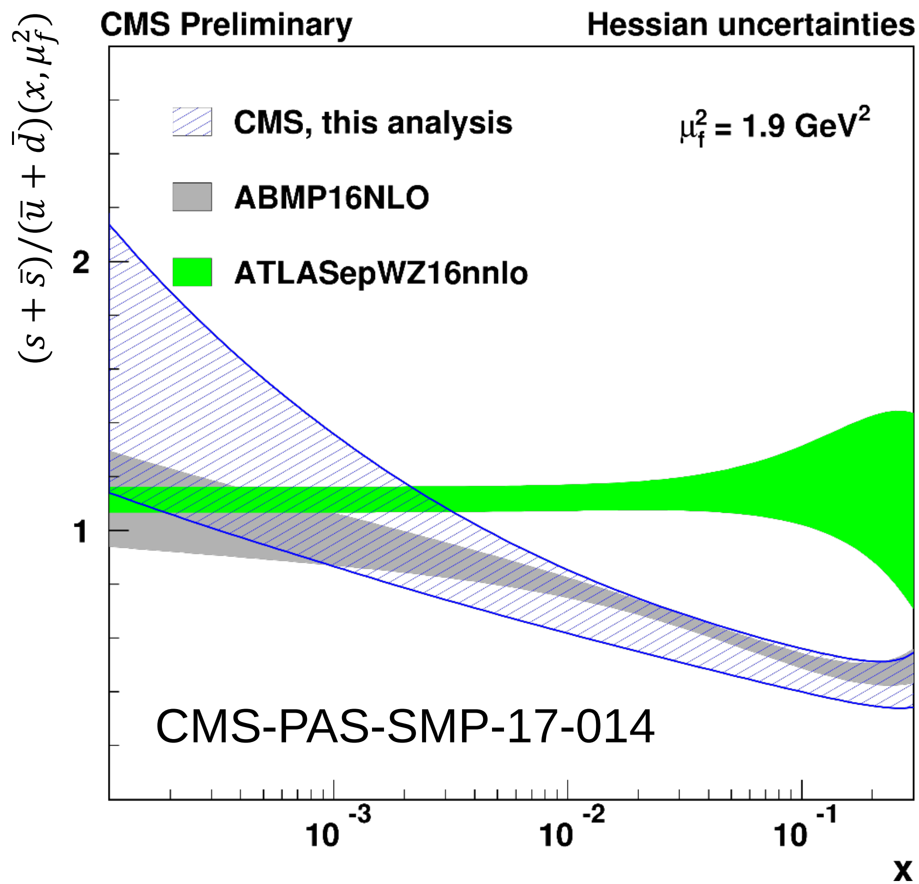
Weaker constraints from HERAI+II, E866, LHCb, Tevatron jets, CDHSW F2, $t\bar{t}$ production

After the fit

Inconsistent conclusions in literature about strangeness preferred by the LHC data

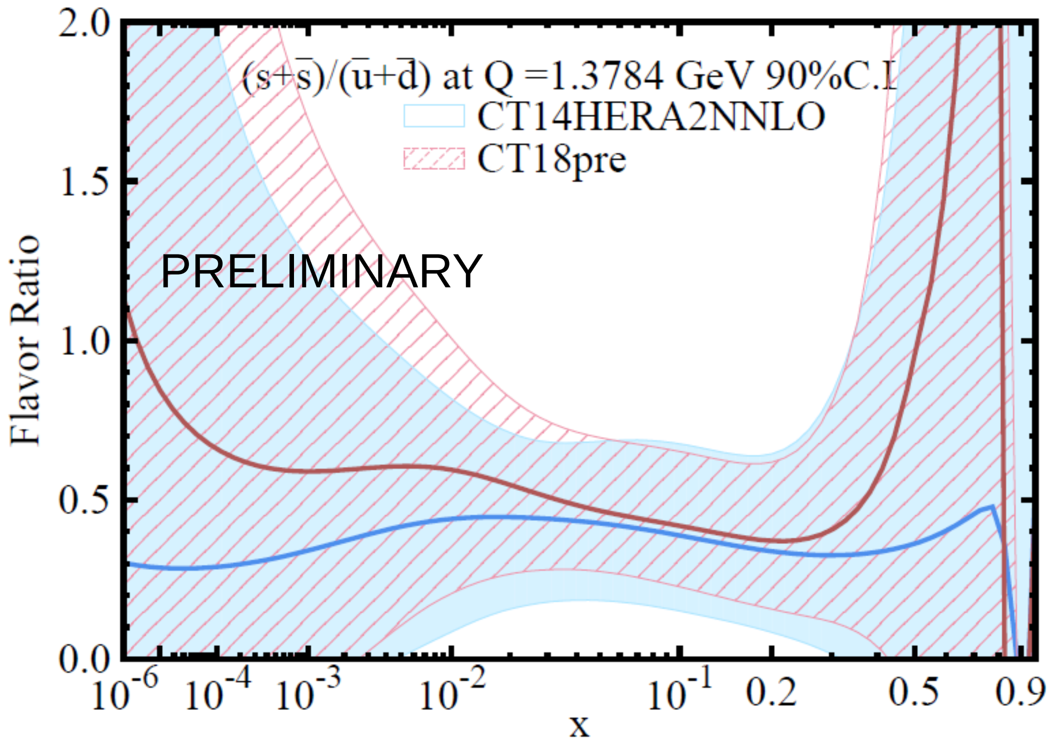


CT14 and MMHT14 NNLO PDFs profiled using ATLAS 7 TeV (4.6 fb^{-1}) $W^{\pm}$, Z xsecs prefer $s(x, 1.9\ GeV) \sim 1$ @ $x = 0.023$



CMS W,Z, CMS W+c (13 TeV) prefer smaller $s(x,\mu)$ than ATLAS for $x \gtrsim 10^{-3}$

Effect of LHC data on strangeness: the actual CT18pre fit

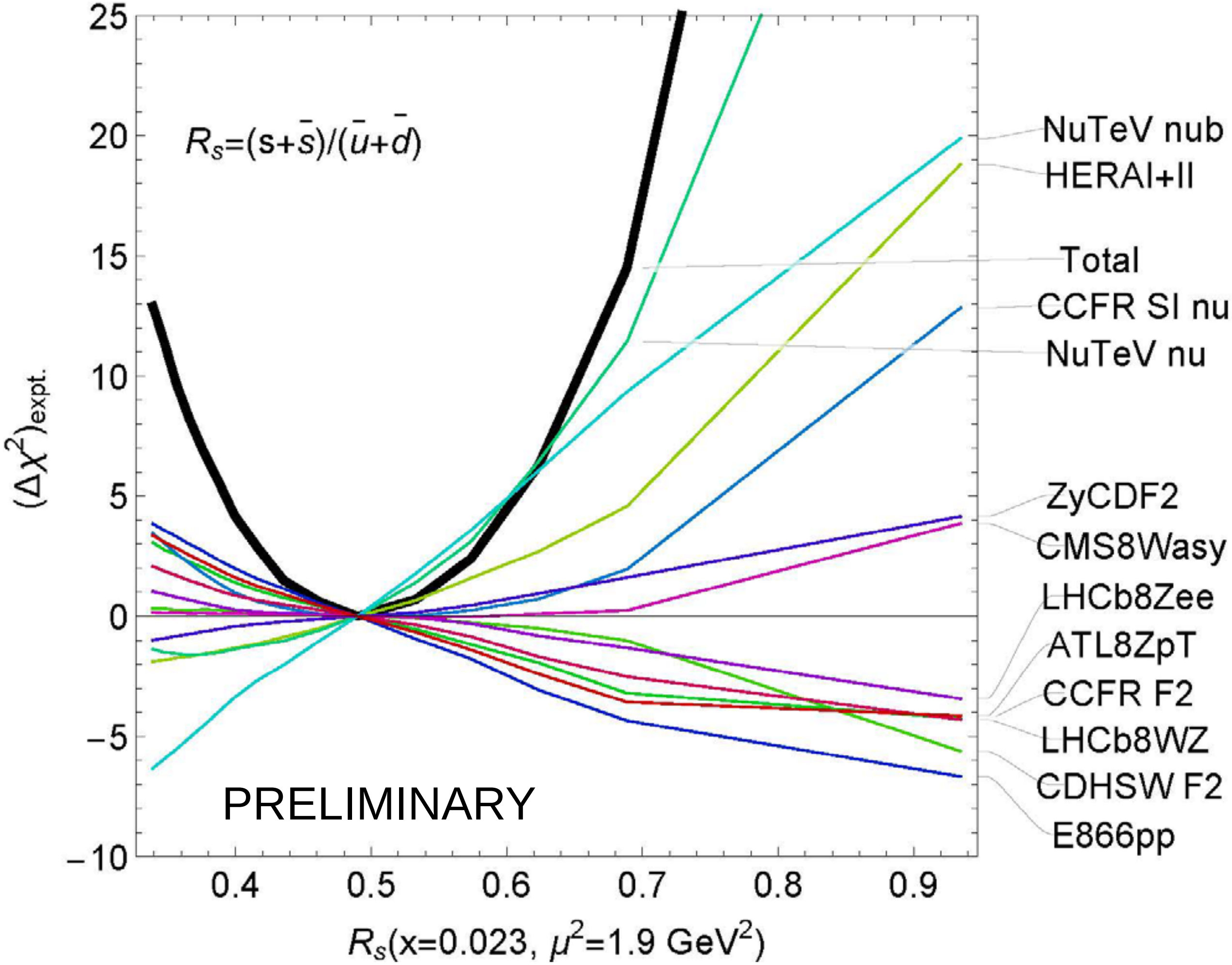


Mild upward pool on $s(x,Q)$ compared to CT14 HERA2

Small to no reduction on the PDF error on $s(x,Q)$

$R_s(CT18pre) = \frac{s+\bar{s}}{\bar{u}+\bar{d}} = 0.53 \pm 0.16$ (90% c.l.) at $x = 0.023$, $Q^2 = 1.9\ \mathrm{GeV}^2$

Effect of LHC data on strangeness: the actual CT18pre fit



Some tension between NuTeV, CCFR dimuon production, HERAI+II (preferring $R_s < 0.6$);

and vector boson production at the LHC and Tevatron (preferring $R_s > 0.6$)

However, still large uncertainties

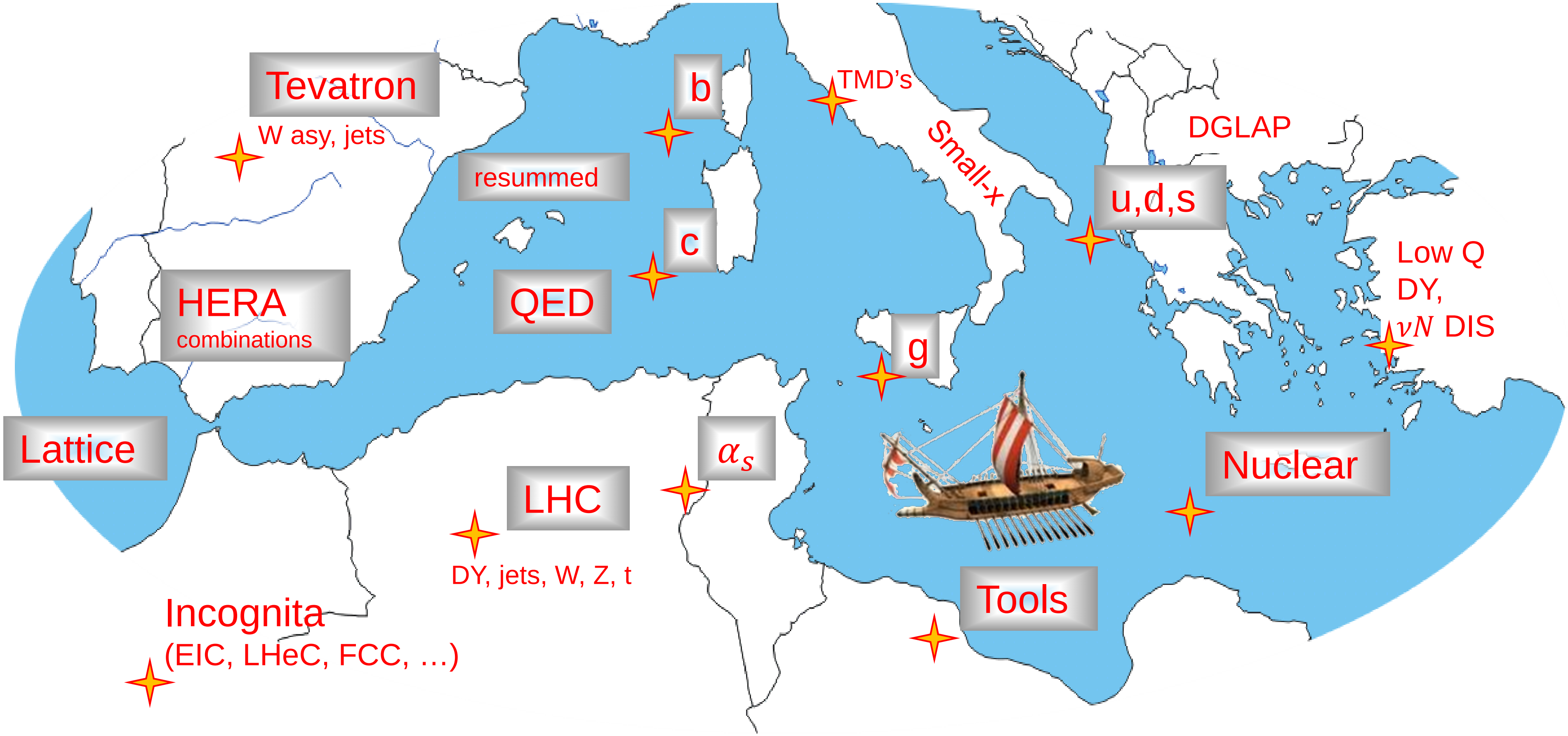
Outlook for CTEQ-TEA PDFs

- **Ongoing CTEQ-TEA PDF analysis**

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]**

Oekumene of the PDF universe



Recent SMU contributions

New lands for charting,
new tools for exploring

Extra details

Fast estimation of sensitivity S_f of experimental data to theory cross sections [PDFSense]

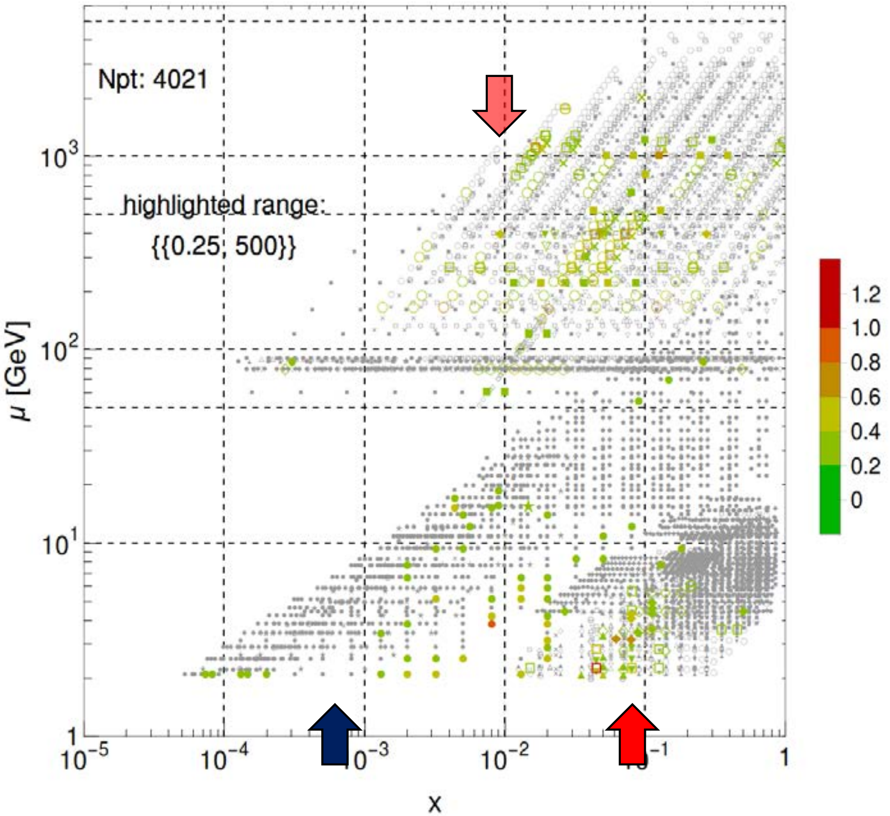
[B.T. Wang, T.J. Hobbs, et al., arXiv:1803.02777]

The sensitivity S_f compares full experimental and PDF uncertainties

S_f is not affected by limitations of PDF reweighting

If a **data point** is sensitive to a given PDF f, then the sensitivity $|S_f|$ of this data point to f is much larger than 0 [say, $|S_f| > 0.25$]. S_f is defined on the next slide.

Sensitivity to the PDF error on $\sigma(pp \to H^0 X)$ at 14 TeV



Before the fit

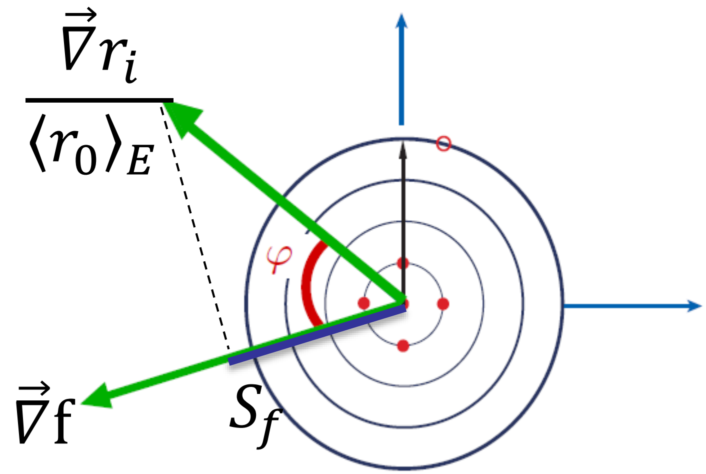
Correlation C_f and sensitivity S_f

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.

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