Mapping the sensitivity of future facilities to collinear PDFs with PDFSense

Pavel Nadolsky[1*]

with T. J. Hobbs,[1] B.-T. Wang,[1] S. Doyle,[1]
J. Gao,[2] T.-J. Hou,[3] M. Hamilton,[1]
and F. I. Olness[1]

[1] Southern Methodist University, Dallas, USA
[2] Shanghai Jiao Tong University, China
[3] Northeastern University, China
* presenter

SMU

CTEQ

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

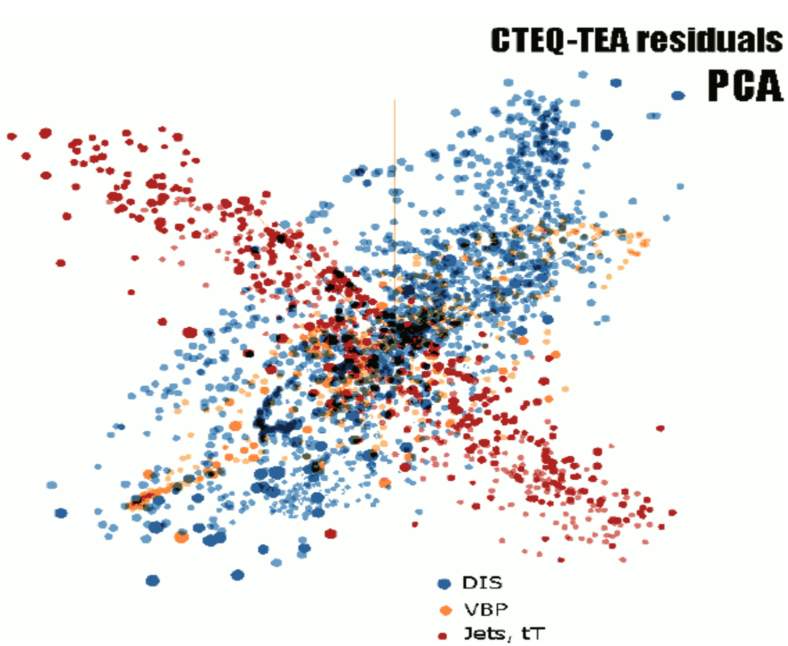
2. The coming synergy between lattice QCD and high-energy phenomenology
T.J. Hobbs, Bo-Ting Wang, Pavel Nadolsky, Fredrick Olness
arXiv:1904.00222

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

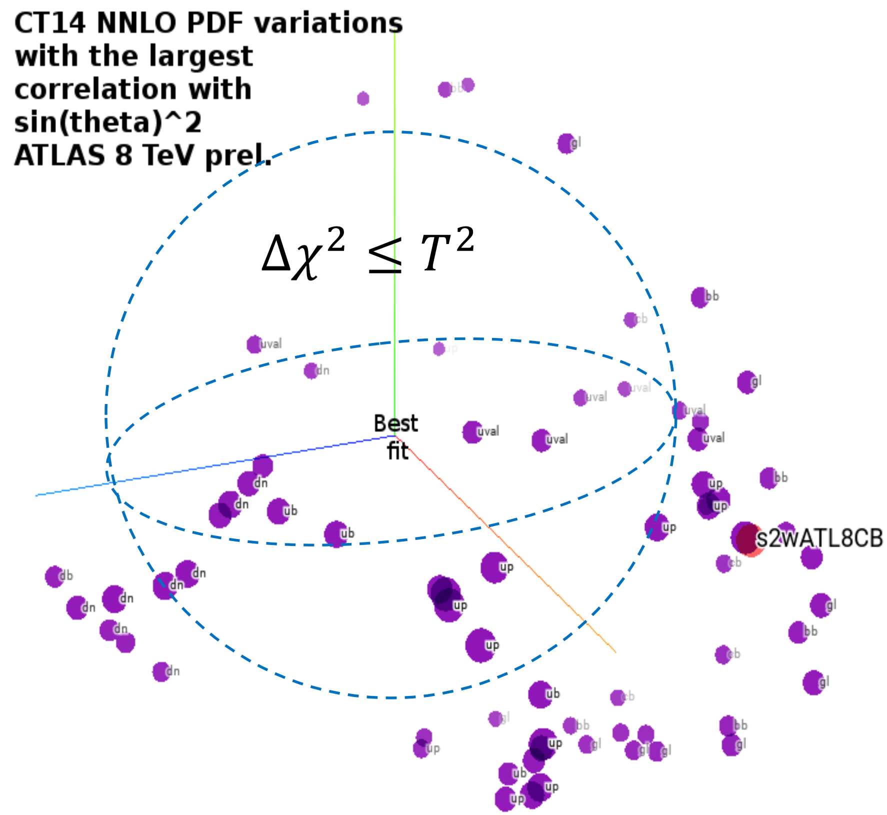
This talk

Vectors of data point residuals…

… carry detailed information about sensitivity of individual experimental data points to PDFs; can be studied using statistical packages (TensorFlow, Mathematica,…)

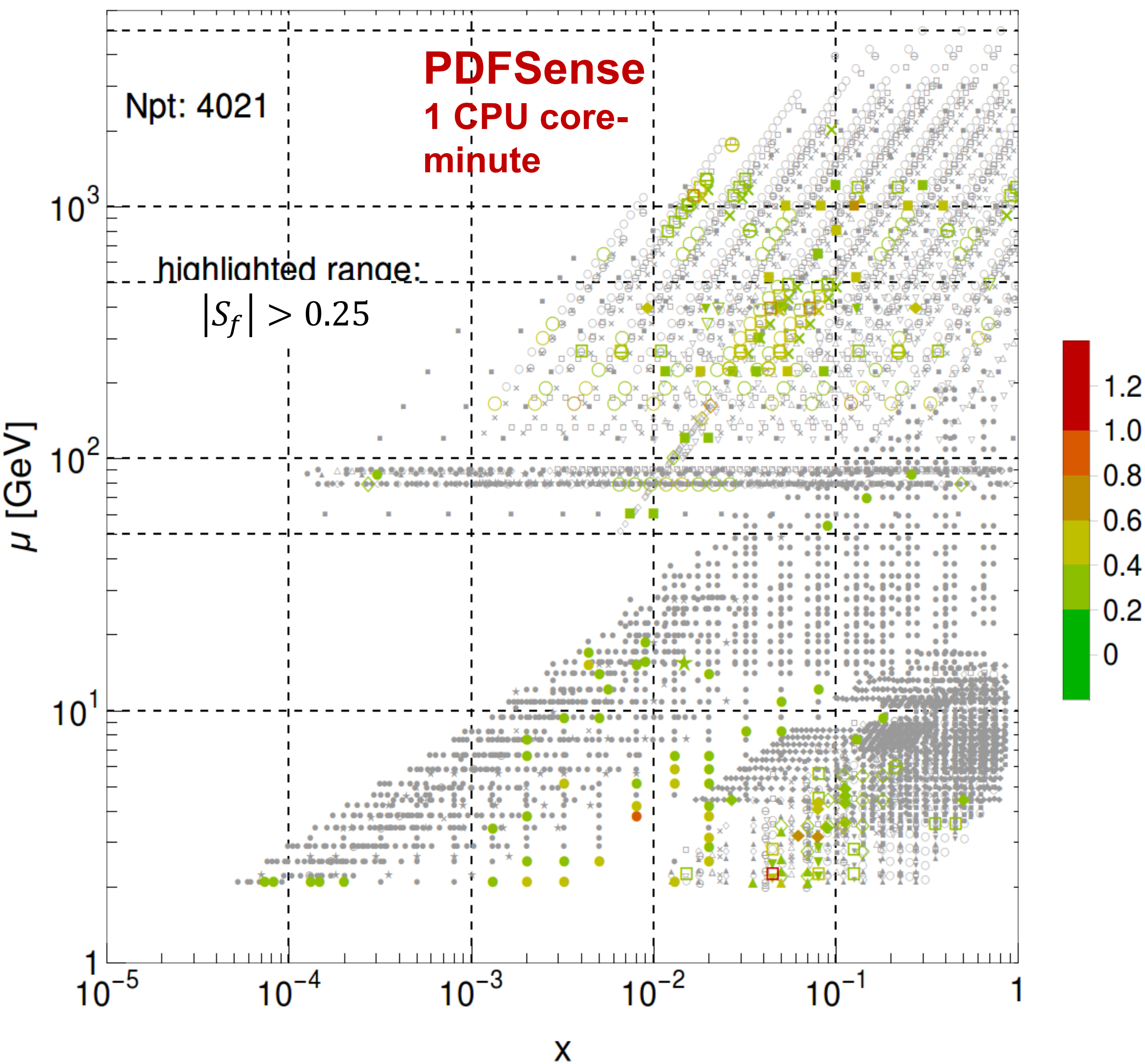


Principal Component Analysis (PCA) visualizes the 56-dim. manifold by reducing it to 10 dimensions (à la META PDFs)



Using Hessian PDFs

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



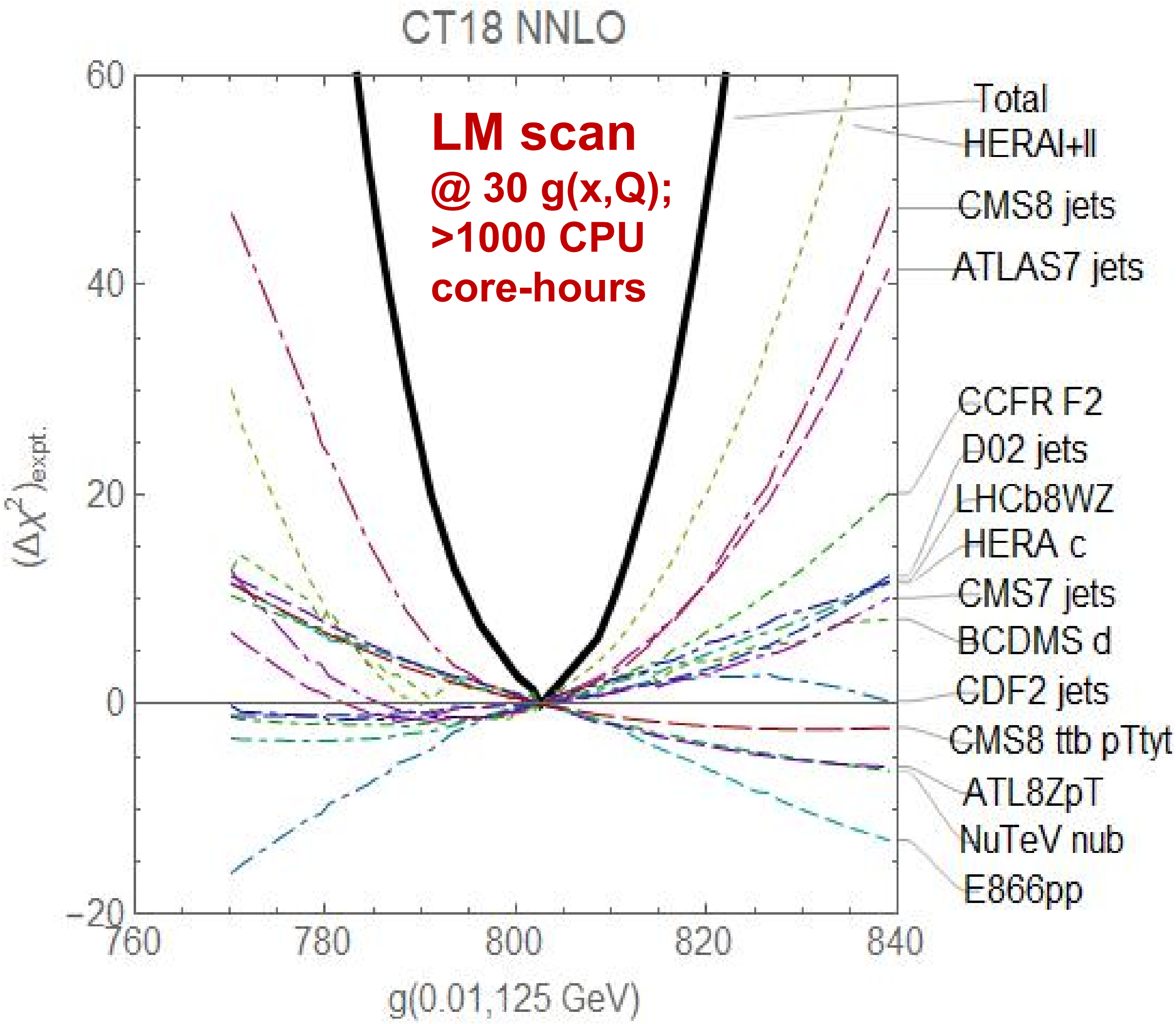
Higgs boson production

HERA DIS still has the dominant sensitivity!

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

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]



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 $g(0.01, 125\ \text{GeV})$

24

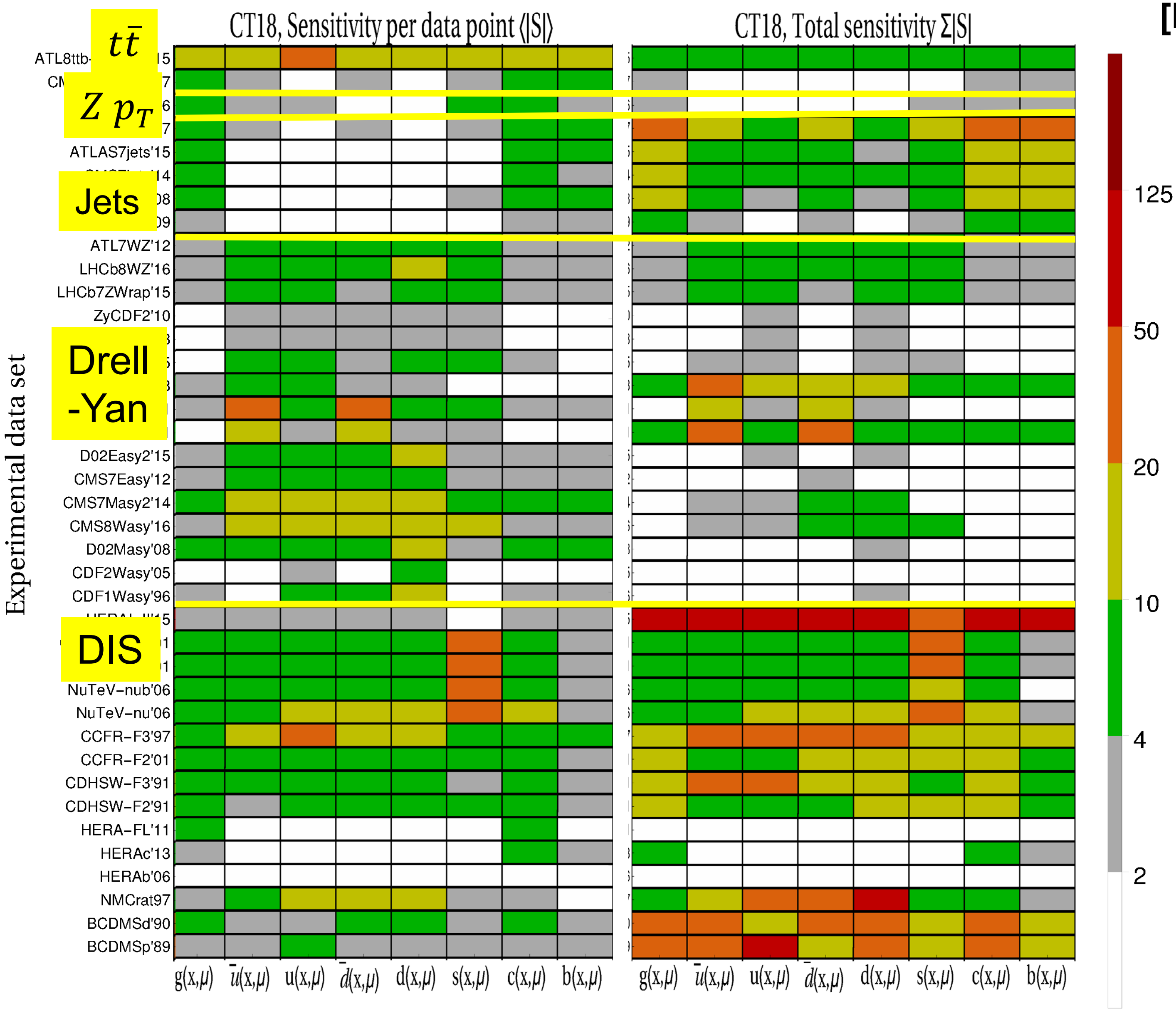
$d/u(x=0.1, \mu=1.3\ \text{GeV})$			$g(x=0.01, \mu=125\ \text{GeV})$		
PDFSense		LM scan	PDFSense		LM scan
CT14HERA2	CT18pre	CT18pre	CT14HERA2	CT18pre	CT18pre
HERAI+II'15	NMCrat'97	NMCrat'97	HERAI+II'15	HERAI+II'15	HERAI+II'15
BCDMSp'89	HERAI+II'15	CCFR-F3'97	CMS8jets'17	CMS8jets'17	CMS8jets'17
NMCrat'97	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]

Sensitivity of hadronic experiments to CT18 PDFs

[P.N., Tuesday, WG1]



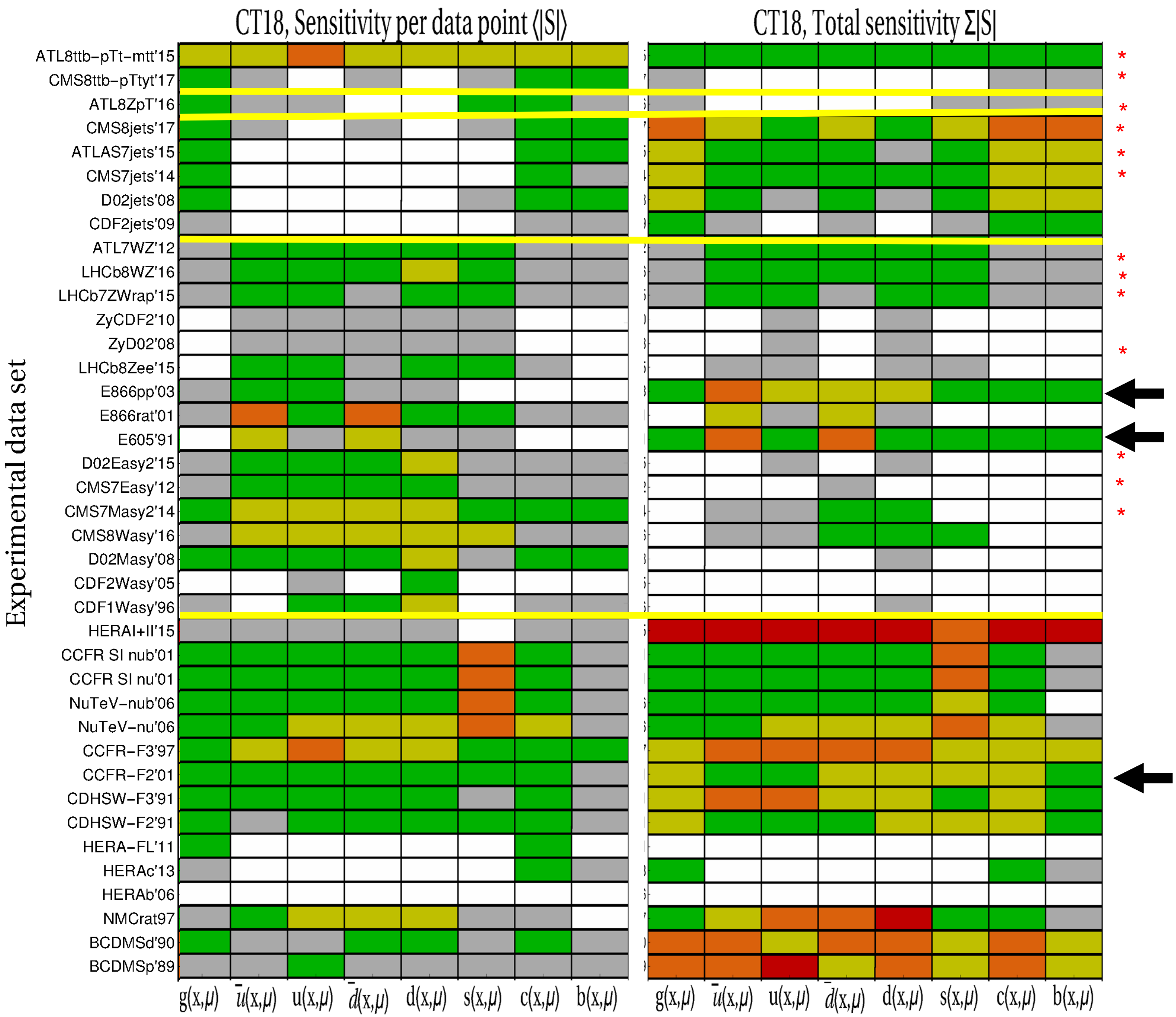
Per-data point and total sensitivity of experimental data sets in the CT18 NNLO analysis to indicated PDF flavors

Computed using the **PDFSense** code
[arXiv:0803.02777]

Red bars = most sensitive experiments

Sensitivity of hadronic experiments to CT18 PDFs

[P.N., Tuesday, WG1]



The LHC data sets (*) hold a great promise – if they agree

"Midsize" Drell-Yan experiments (E605, E866, SeaQuest, STAR…) continue to provide competitive constraints

HERA I+II, BCDMS, NMC, DIS data sets dominate present experimental constraints. Large numbers of data points matter!

What about *future experiments*
…like the EIC or LHeC?
especially, in the context of other measurements at HL-LHC

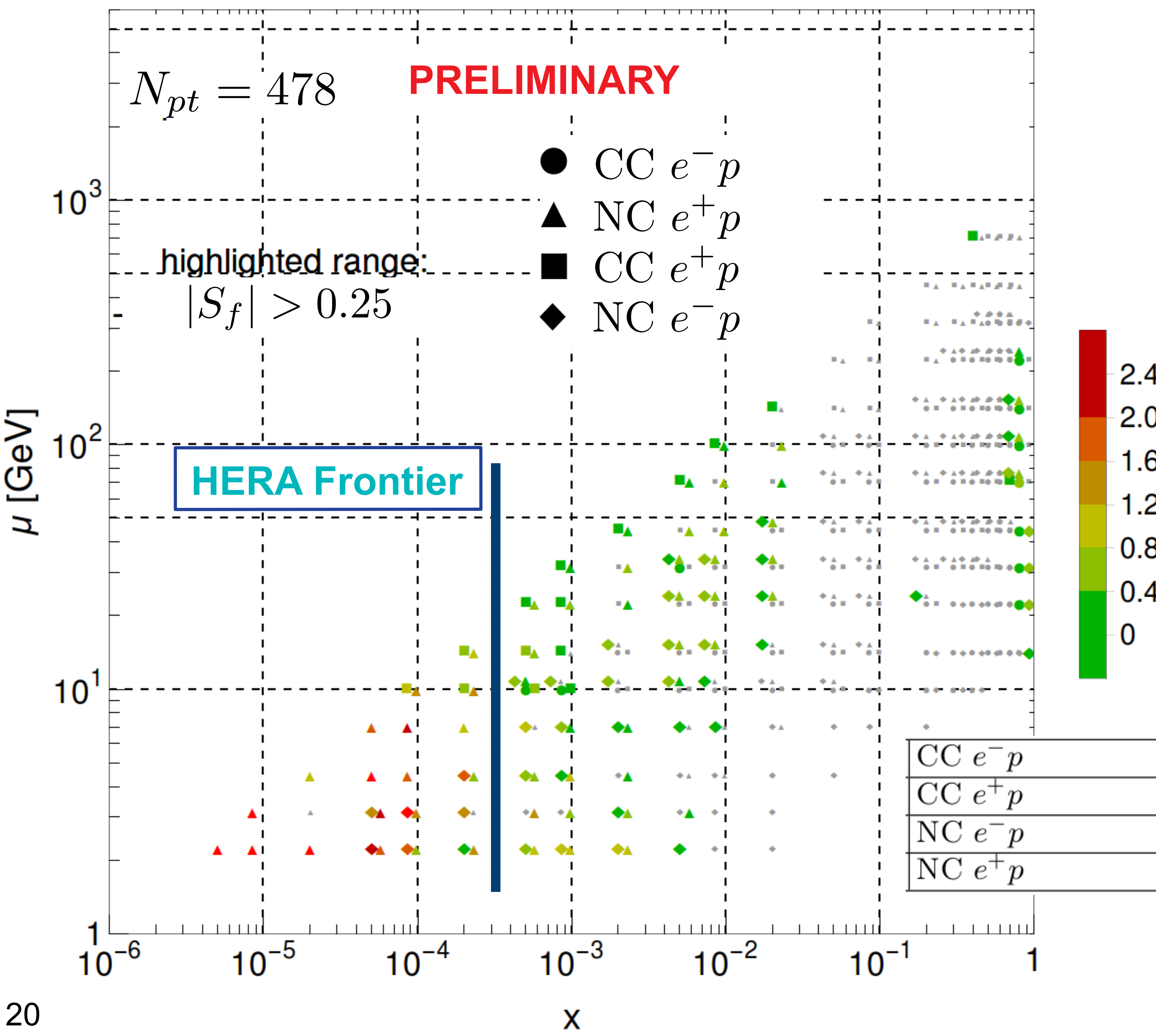
- **EIC and LHeC** PDFSense projections by Hobbs and Wang
- Compared to **HL-LHC** projections by Abdul Khalek, Bailey, Gao, Harland-Lang, Rojo **[arXiv:1810.03039]**

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

| S_f | for g(x,μ), PDF4LHC15 NNLO



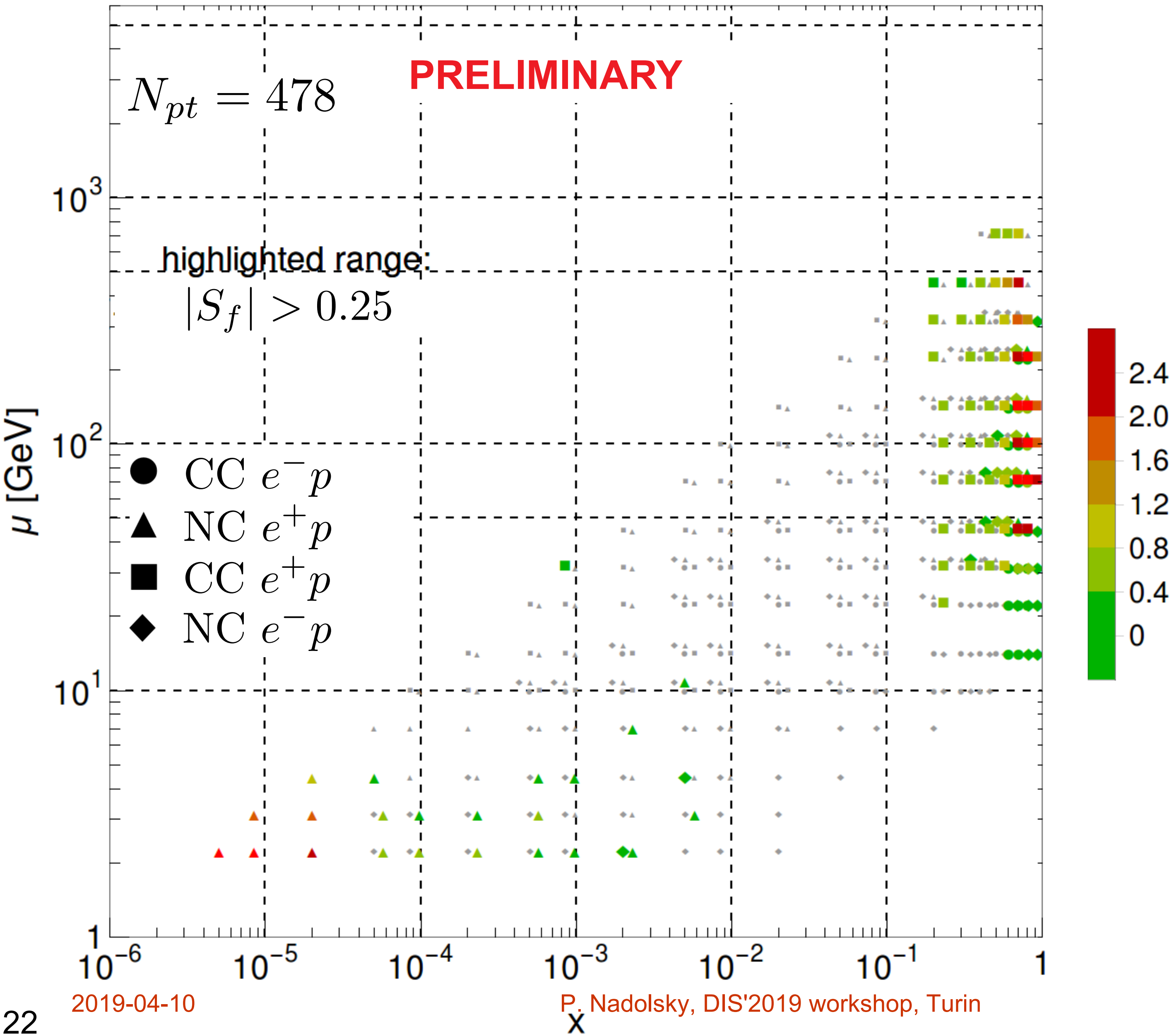
LHeC

- very strong sensitivities along the frontier of the HERA Run II data, $x \lesssim 10^{-4}$
- In addition, important constraints for $x \lesssim 10^{-2}$

…and high x, μ

	N_{pt}	
CC $e^- p$	93	circle
CC $e^+ p$	109	square
NC $e^- p$	128	diamond
NC $e^+ p$	148	triangle

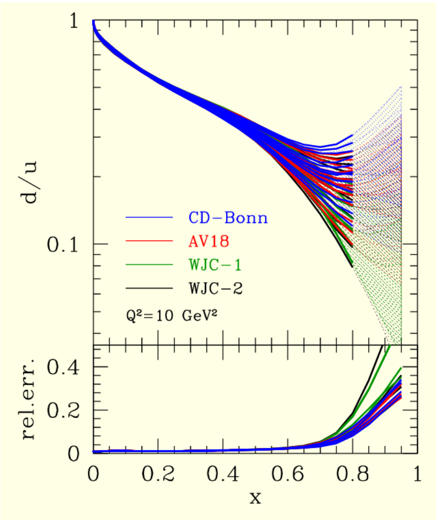
$|S_f|$ for d/u(x,mu), PDF4LHC15 NNLO



LHeC

- LHeC's high luminosity may give it a reach to high enough x to help resolve the stubborn d/u question

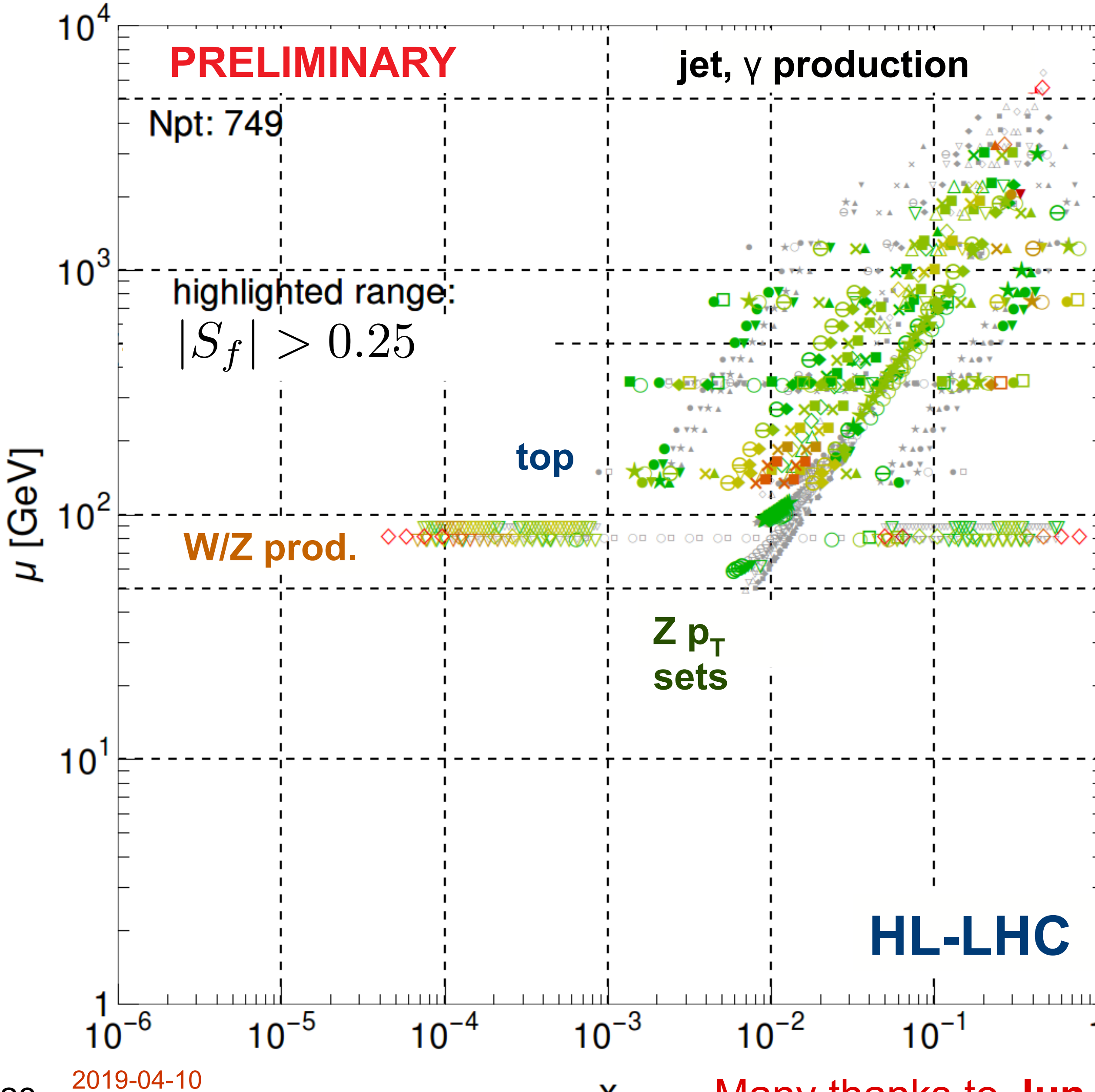
- …without a nuclear target…



J.F. Owens, PAVI11

| S_f | for $g(x,\mu)$, PDF4LHC15 NNLO

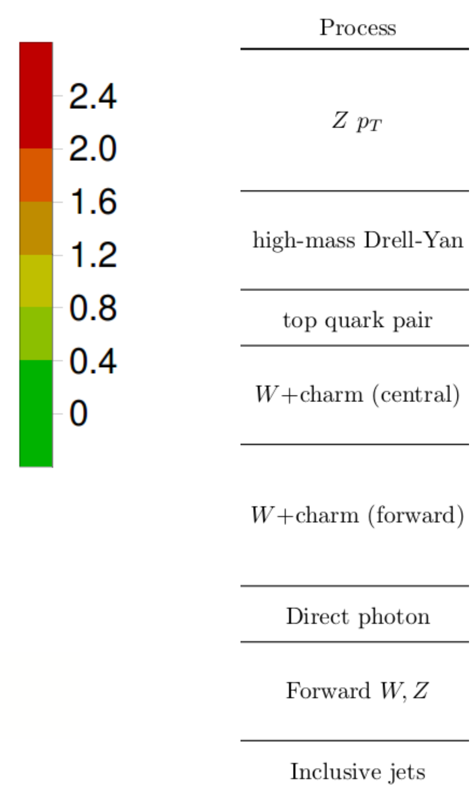
HL-LHC



- very <u>large</u> integrated luminosities:

$\mathcal{L} = 3\ ab^{-1}$ (CMS/ATLAS)

$\mathcal{L} = 0.3\ ab^{-1}$ (LHCb)



Many thanks to **Jun Gao** for the pseudodata!

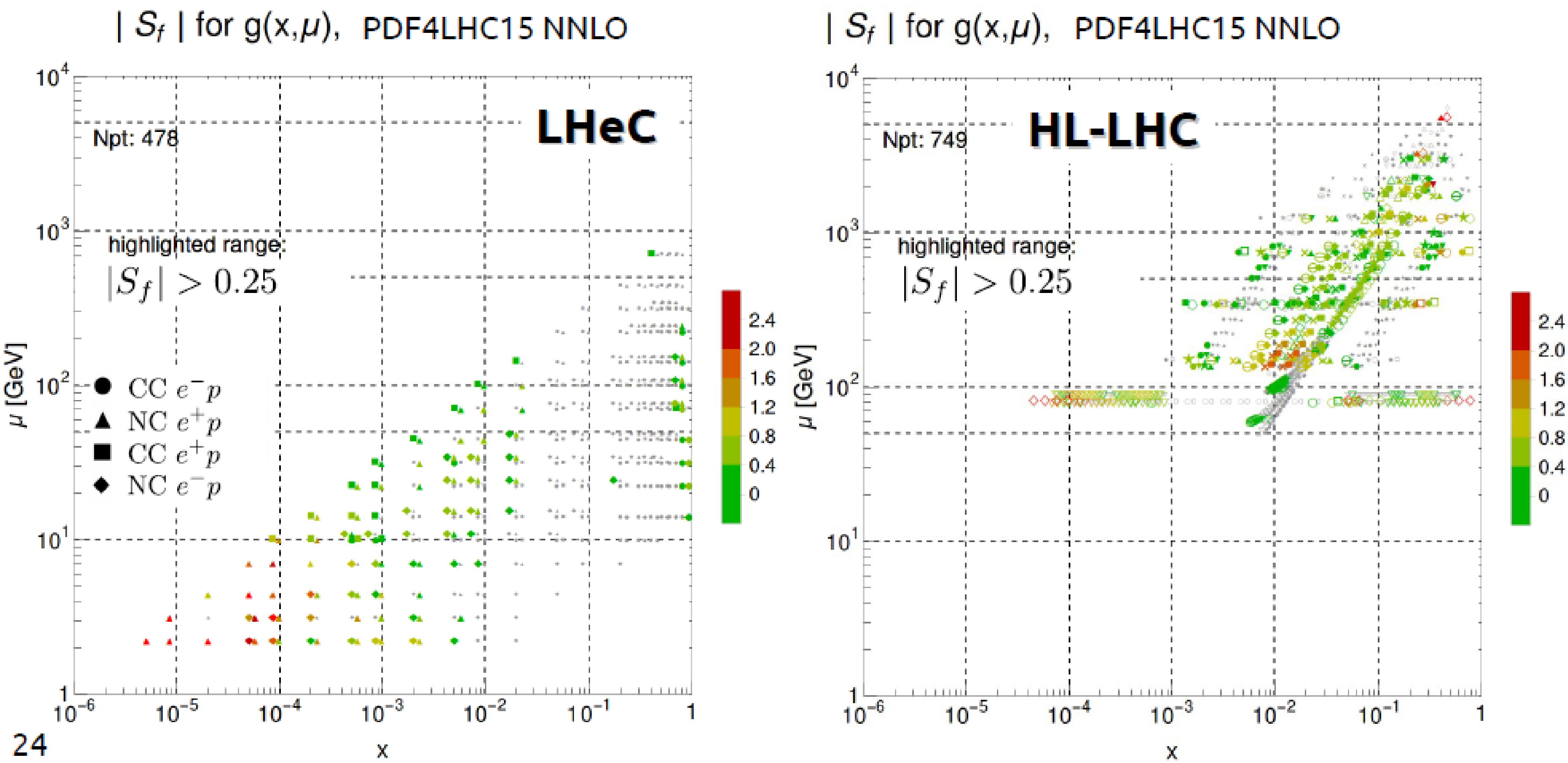
now directly compare the LHeC vs. HL-LHC flavor sensitivities*

$g(x, \mu)$

*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:

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

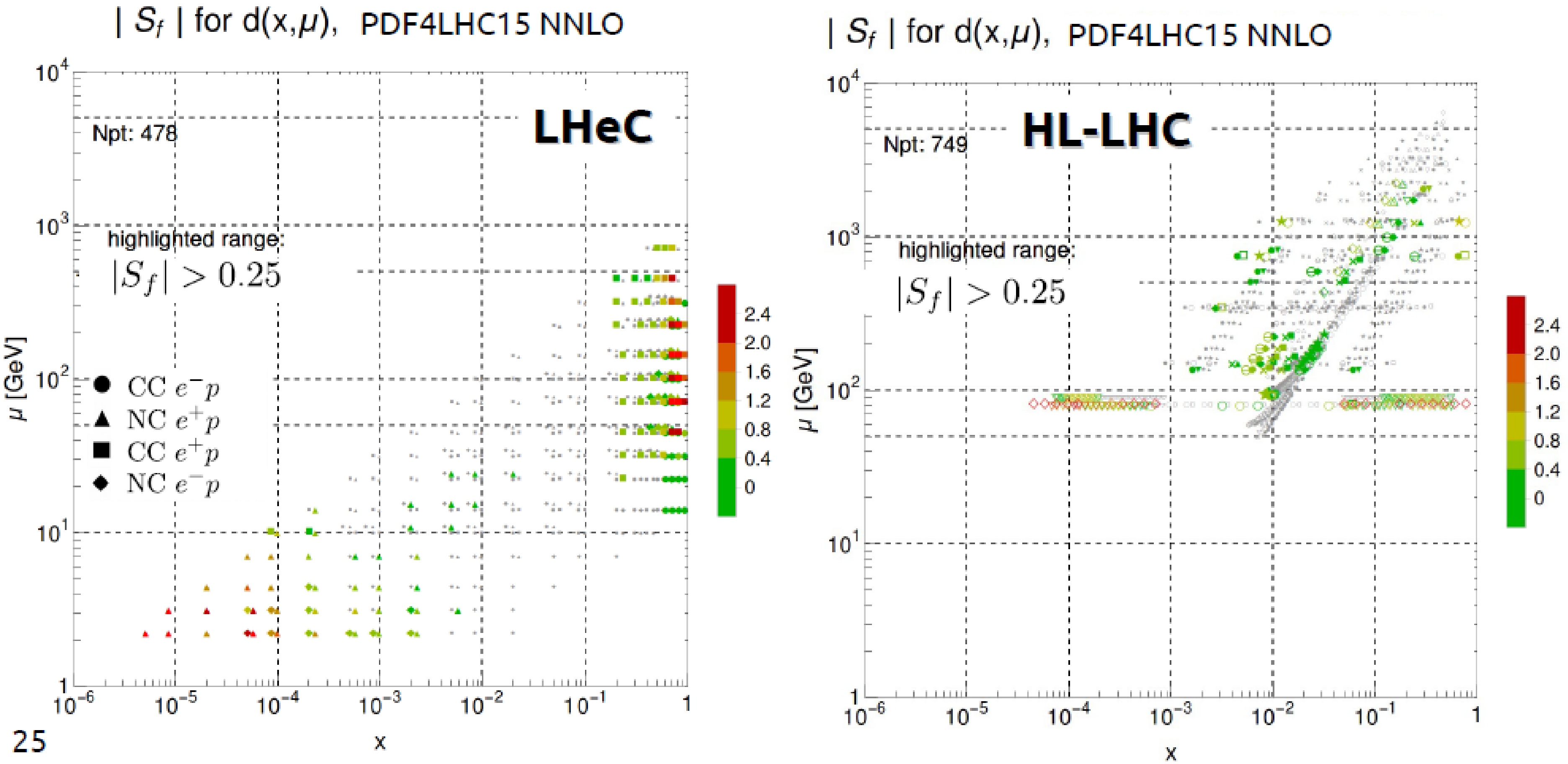


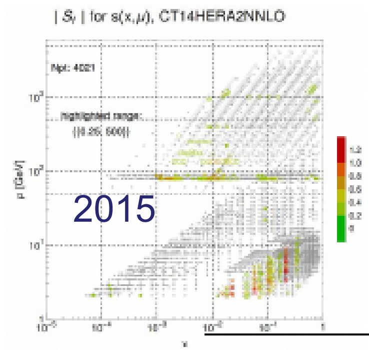
even with a small (pseudo)data set, LHeC enjoys strong sensitivity to down-type distributions!

$d(x,\mu)$

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

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



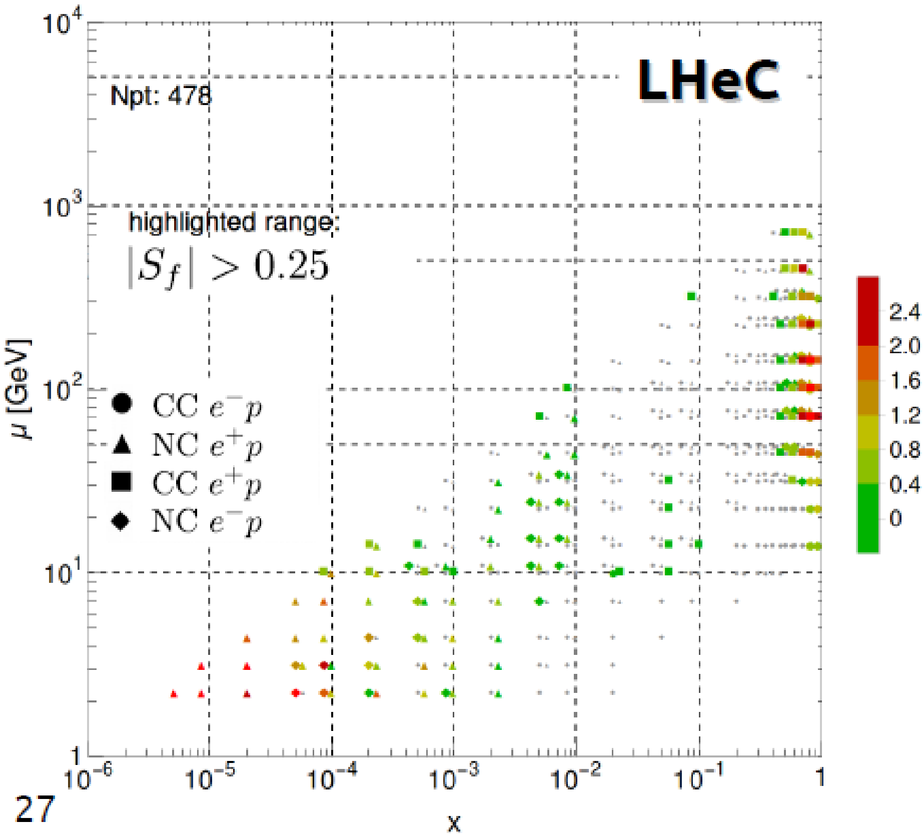


→ LHeC especially portends significantly heightened knowledge of nucleon strangeness

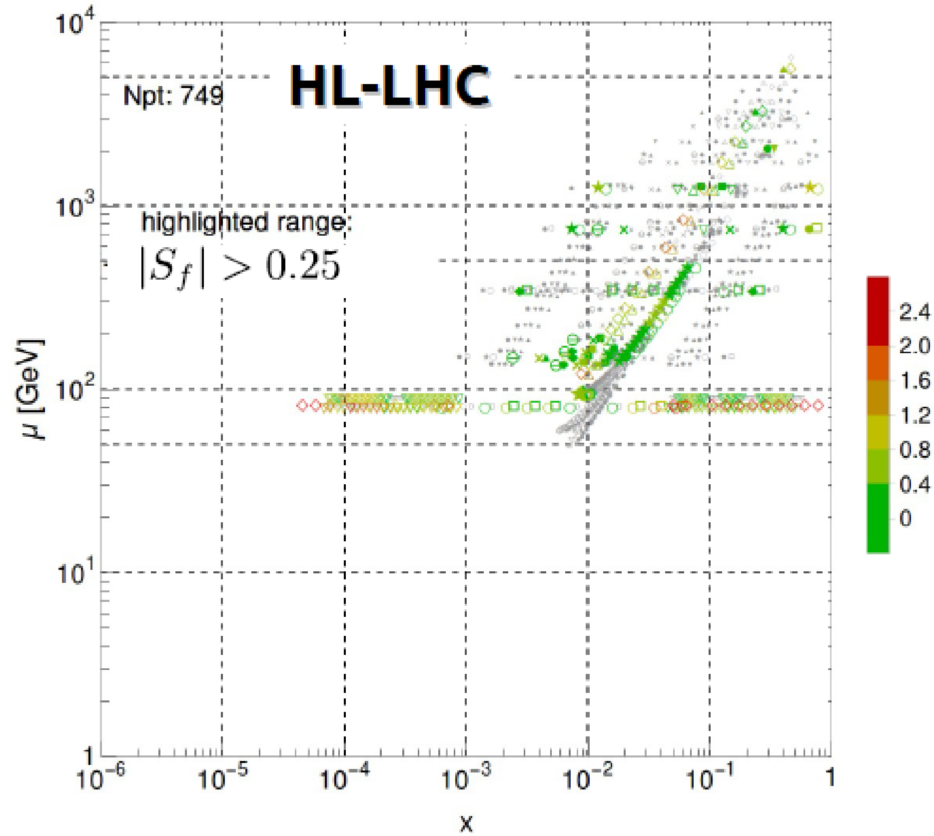
- CTEQ-TEA constraints come primarily through older fixed-target data and Tevatron data (and LHC Run I)

s(x, μ) $|S_s^{\mathrm{LHeC}}| = 214.1 \quad \gtrsim \quad |S_s^{\mathrm{HL-LHC}}| = 184.8$

| S_f | for s(x,μ), PDF4LHC15 NNLO



| S_f | for s(x,μ), PDF4LHC15 NNLO

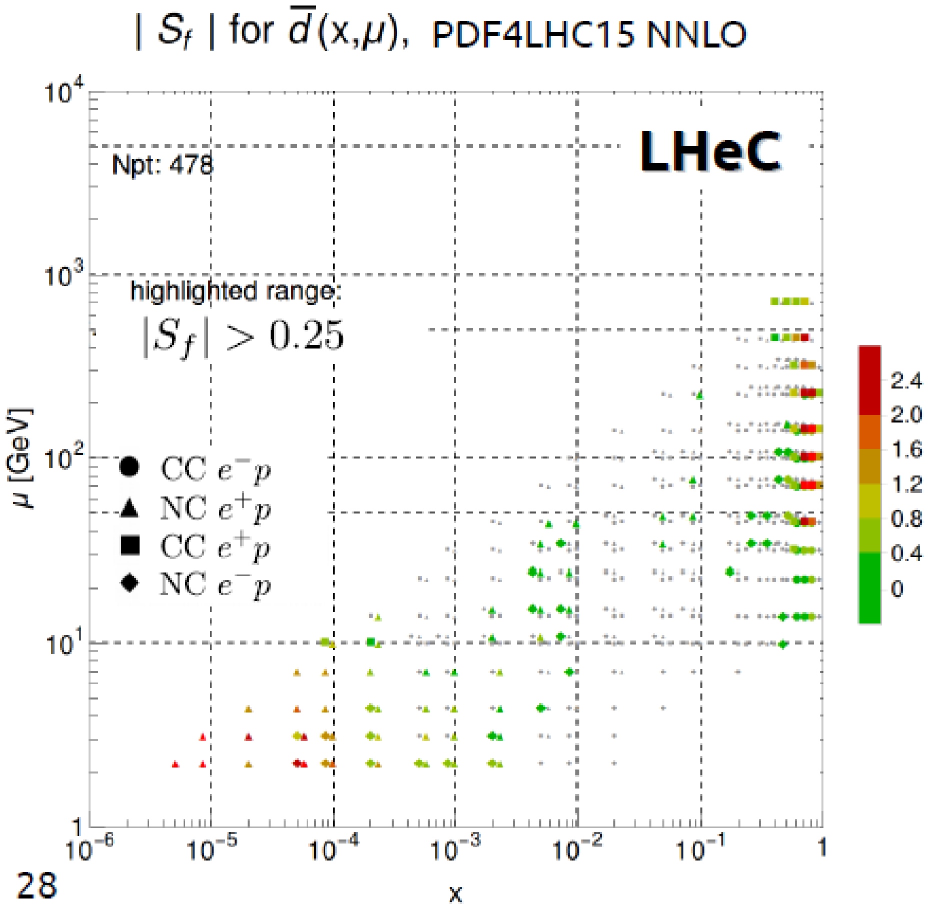


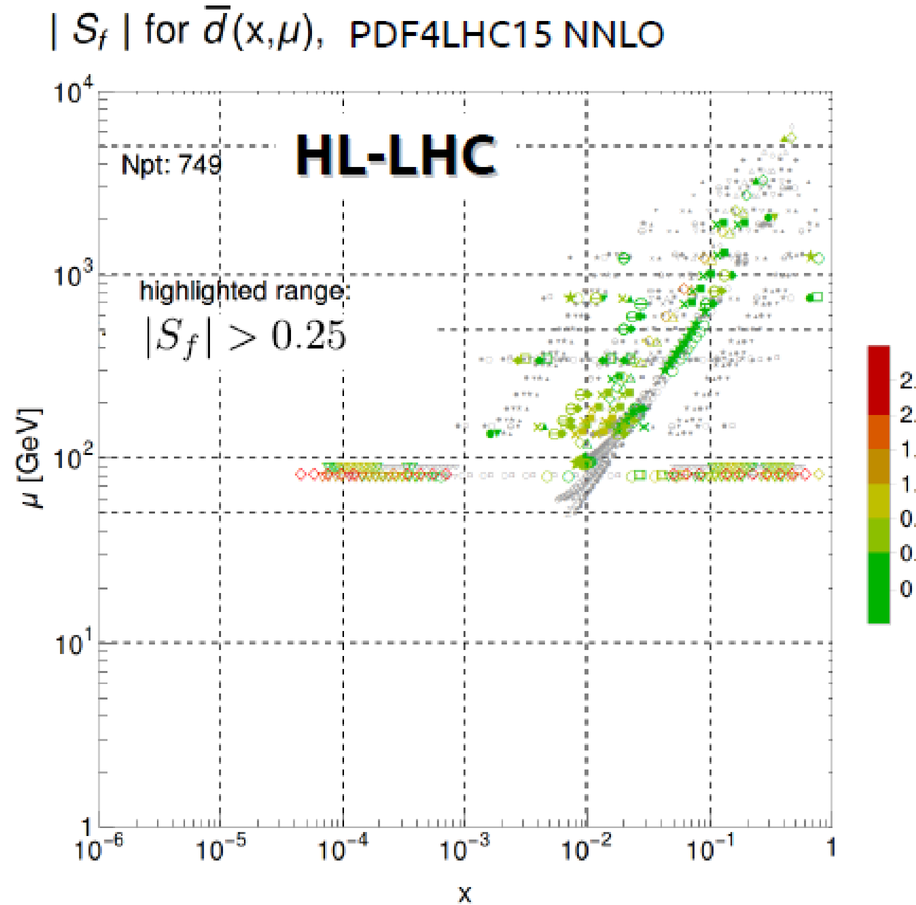
in the SU(2) quark sea, the LHeC 100 fb^{-1} set imposes constraints of magnitude comparable to HL-LHC

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

...these again predominate at the extrema of x

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





- **lattice QCD** calculations continue to improve and will be increasingly useful as inputs into QCD global analyses

PDF-Lattice whitepaper – Lin et al., PPNP100, 107 (2018); arXiv:1711.07916.

- the PDF-Lattice relationship is *synergistic* :

→ PDF phenomenologists deliver benchmarks to challenge the Lattice

↕

→ Lattice delivers theoretical priors for QCD global fits

PDFSense analysis – Hobbs, Wang, Nadolsky and Olness, arXiv:1904.00022.

---

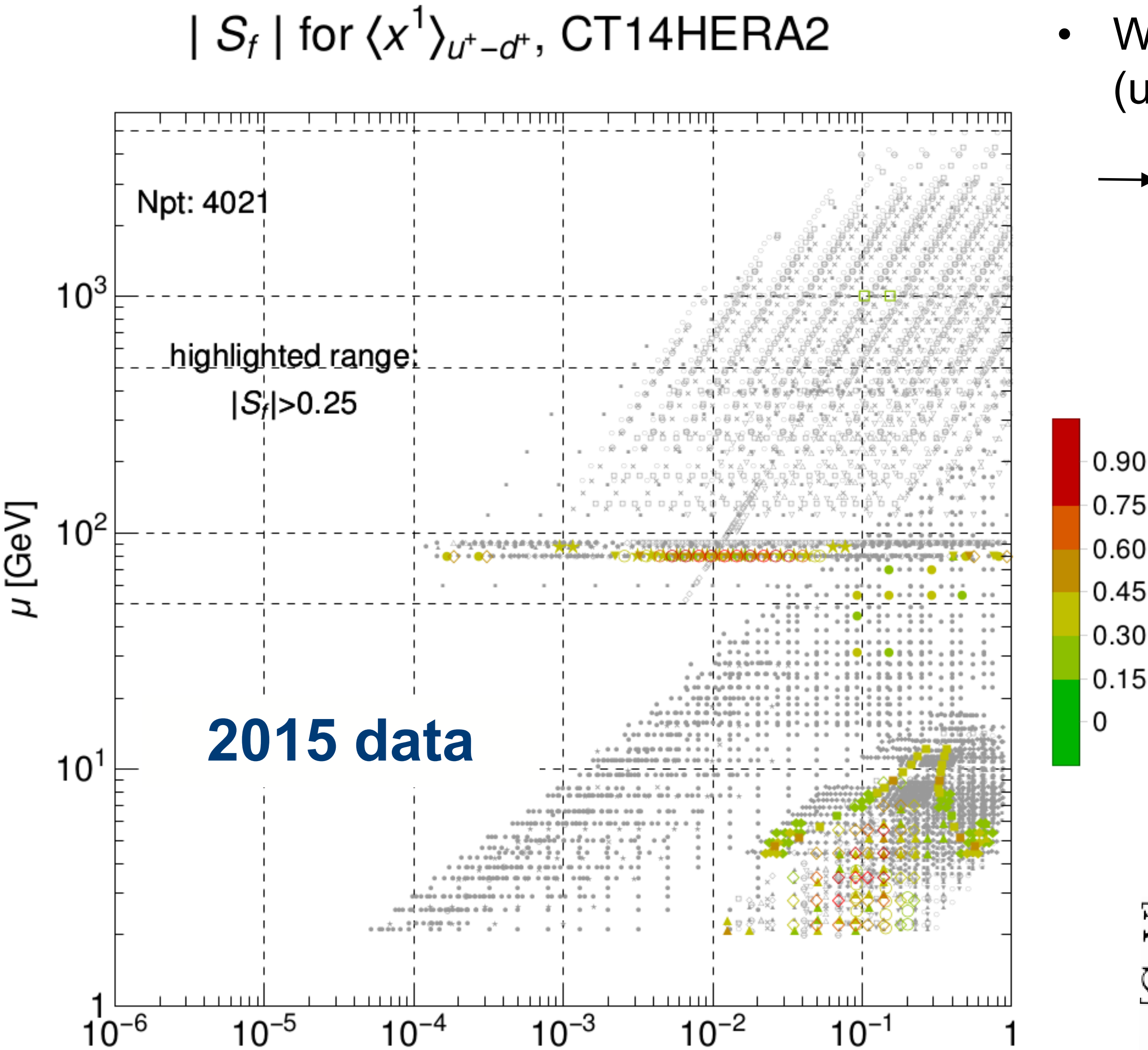
1. Mellin moments from lattice

$$\langle x^n \rangle_q = \int_0^1 dx\ x^n \left[q(x) + (-1)^{n+1}\, \overline{q}(x) \right] \qquad \rightarrow \langle x^{1,3,\dots} \rangle_{q^+},\ \langle x^{2,4,\dots} \rangle_{q^-}$$

2. Quasi-PDFs (qPDFs) from lattice

$$\widetilde{q}(x, P_z, \widetilde{\mu}) = \int_{-\infty}^{\infty} \frac{dz}{4\pi} e^{ixP_z z} \langle P | \overline{\psi}(z) \gamma^z U(z,0) \psi(0) | P \rangle$$

Sensitivity maps: isovector moments



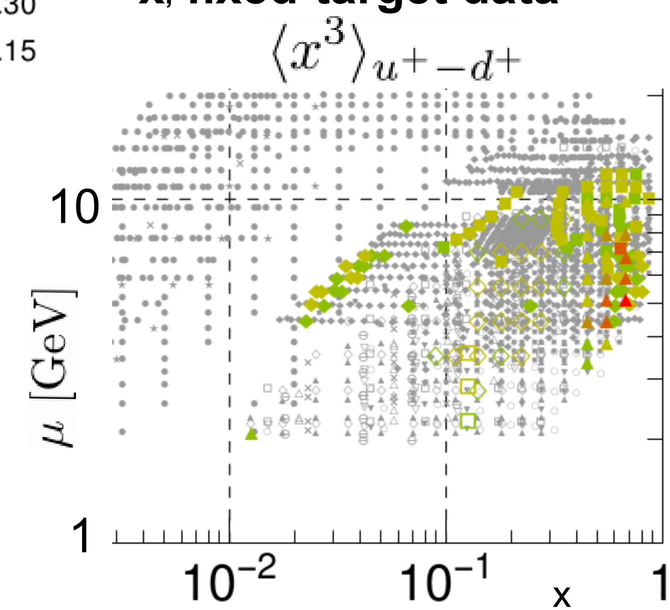
Hobbs et al., arXiv:1904.00222

2019-04-10

- We focus on **isovector** (u-d) PDF combinations

→ on the lattice, these are more readily computed, since flavor non-singlet combinations do not receive disconnected insertions

- **Higher-order moments increasingly constrained by higher x, fixed-target data**



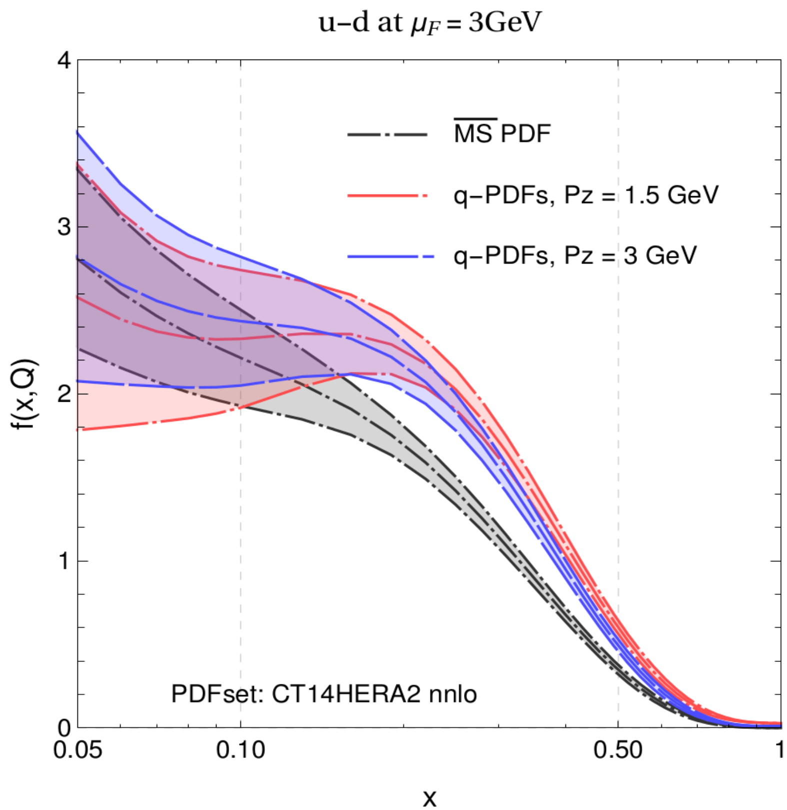
- Also, can **predict** the quasi-PDFs using phenomenological CT PDFs

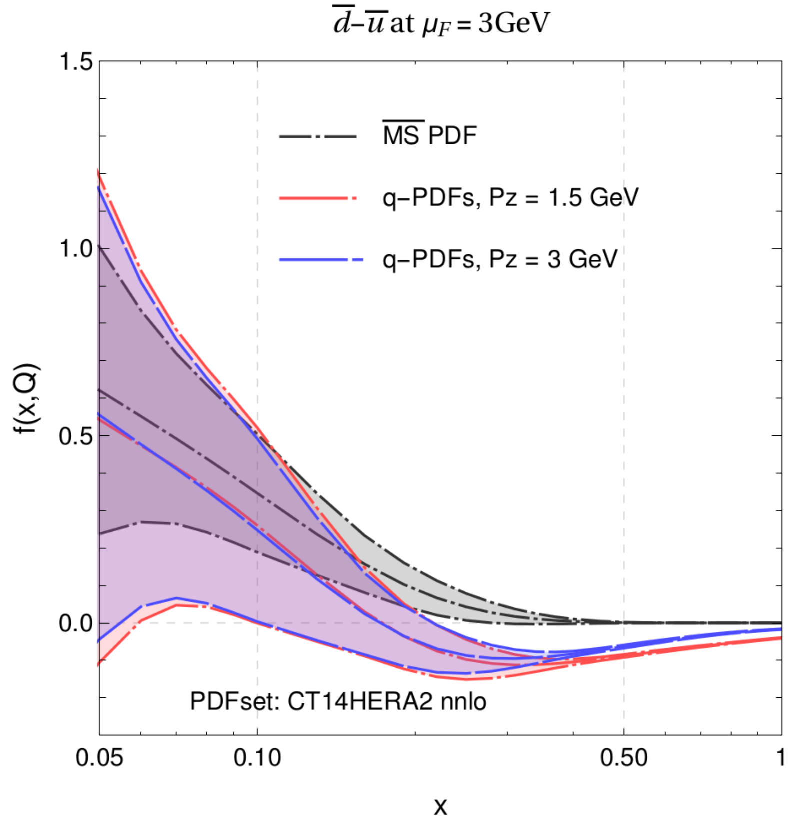
$$\widetilde{q}(x, P_z, \widetilde{\mu}) = \int dy\, Z\left(\frac{x}{y}, \frac{\Lambda}{P_z}, \frac{\mu}{P_z}\right) q(y, \mu) + \mathcal{O}\left(\frac{\Lambda^2}{P_z^2}, \frac{M^2}{P_z^2}\right)$$

Output qPDF;
To be compared against
lattice QCD

Input $\overline{MS}$ PDF (CT14, etc.)

The dependence of qPDFs on the nucleon boost P_z mostly saturates for P_z>1.5 GeV



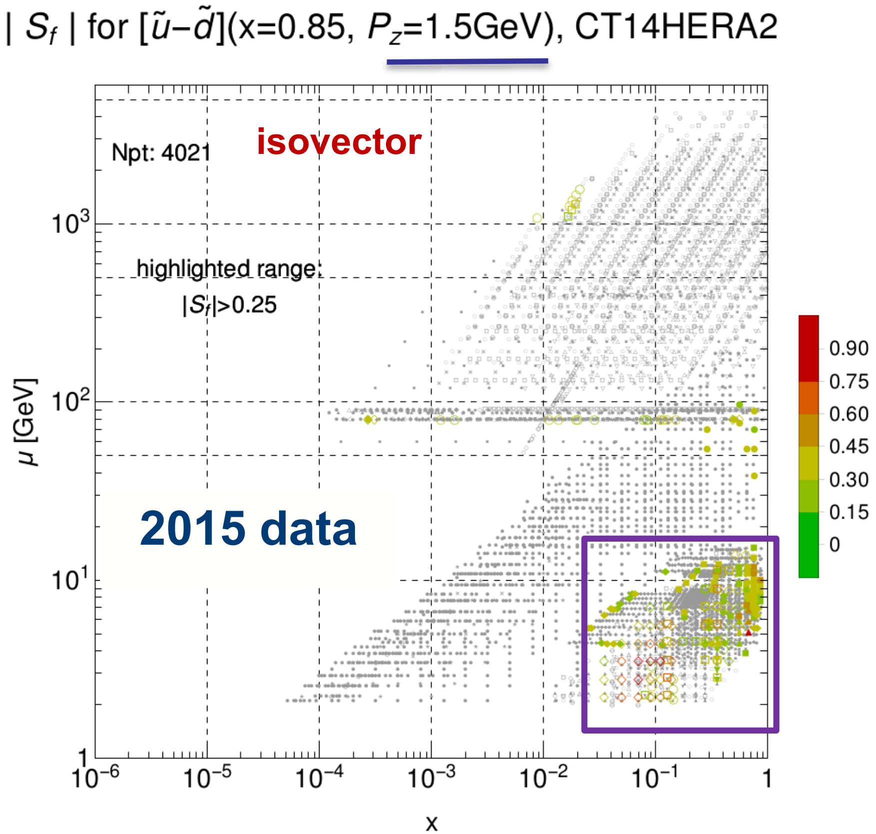


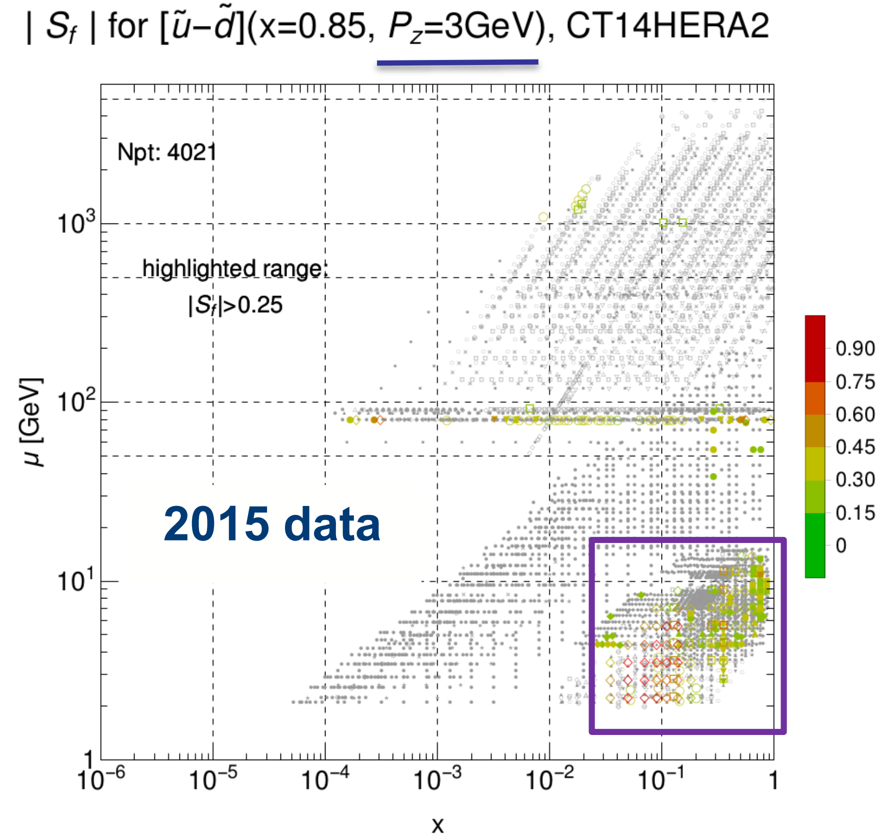
Sensitivity to the "derived" isovector q-PDF

$$\widetilde{q}(x, P_z, \widetilde{\mu}) = \int dy\, Z\left(\frac{x}{y}, \frac{\Lambda}{P_z}, \frac{\mu}{P_z}\right) q(y, \mu) + \mathcal{O}\left(\frac{\Lambda^2}{P_z^2}, \frac{M^2}{P_z^2}\right)$$

Output qPDF;
To be compared against
lattice QCD

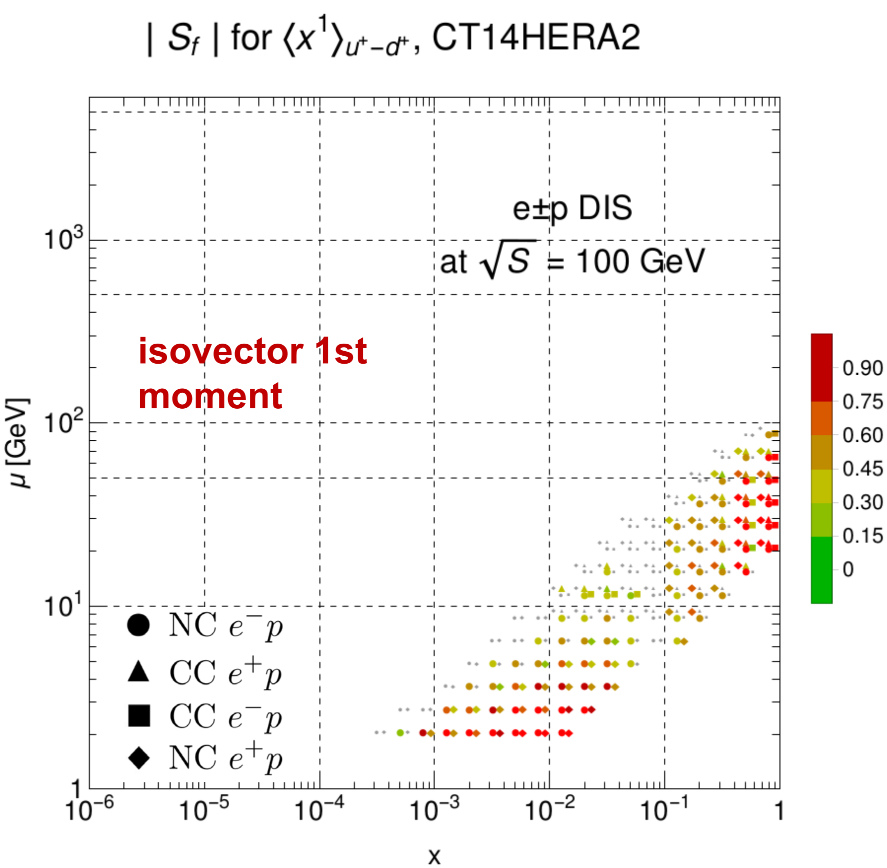
Input $\overline{MS}$ PDF (CT14, etc.)

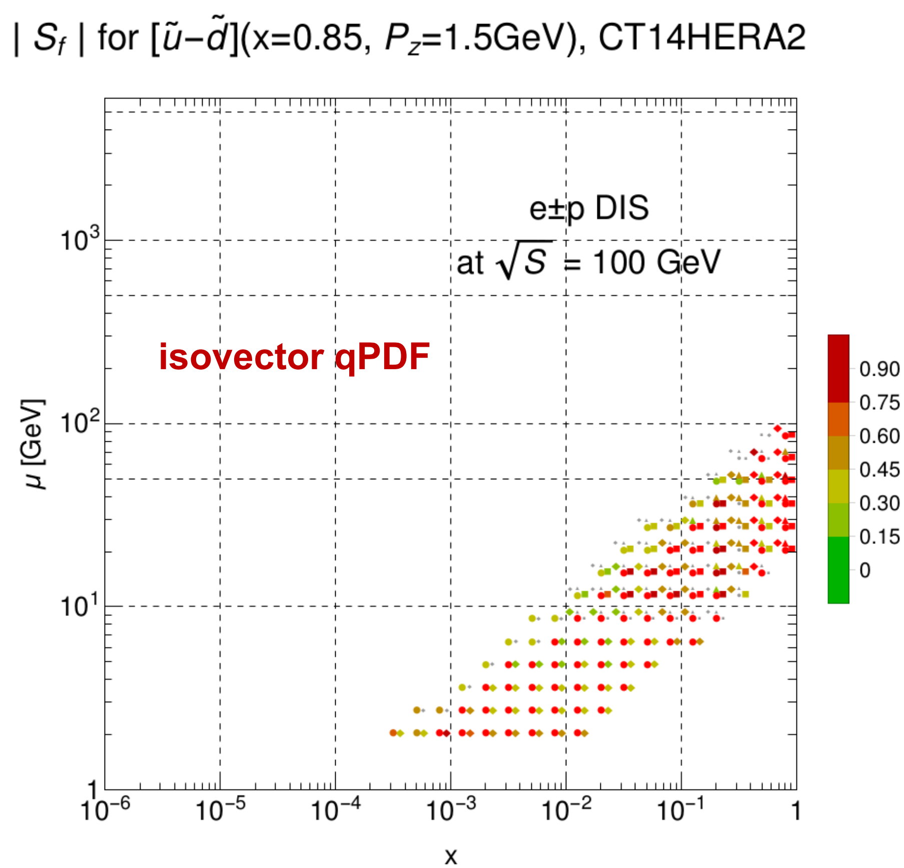


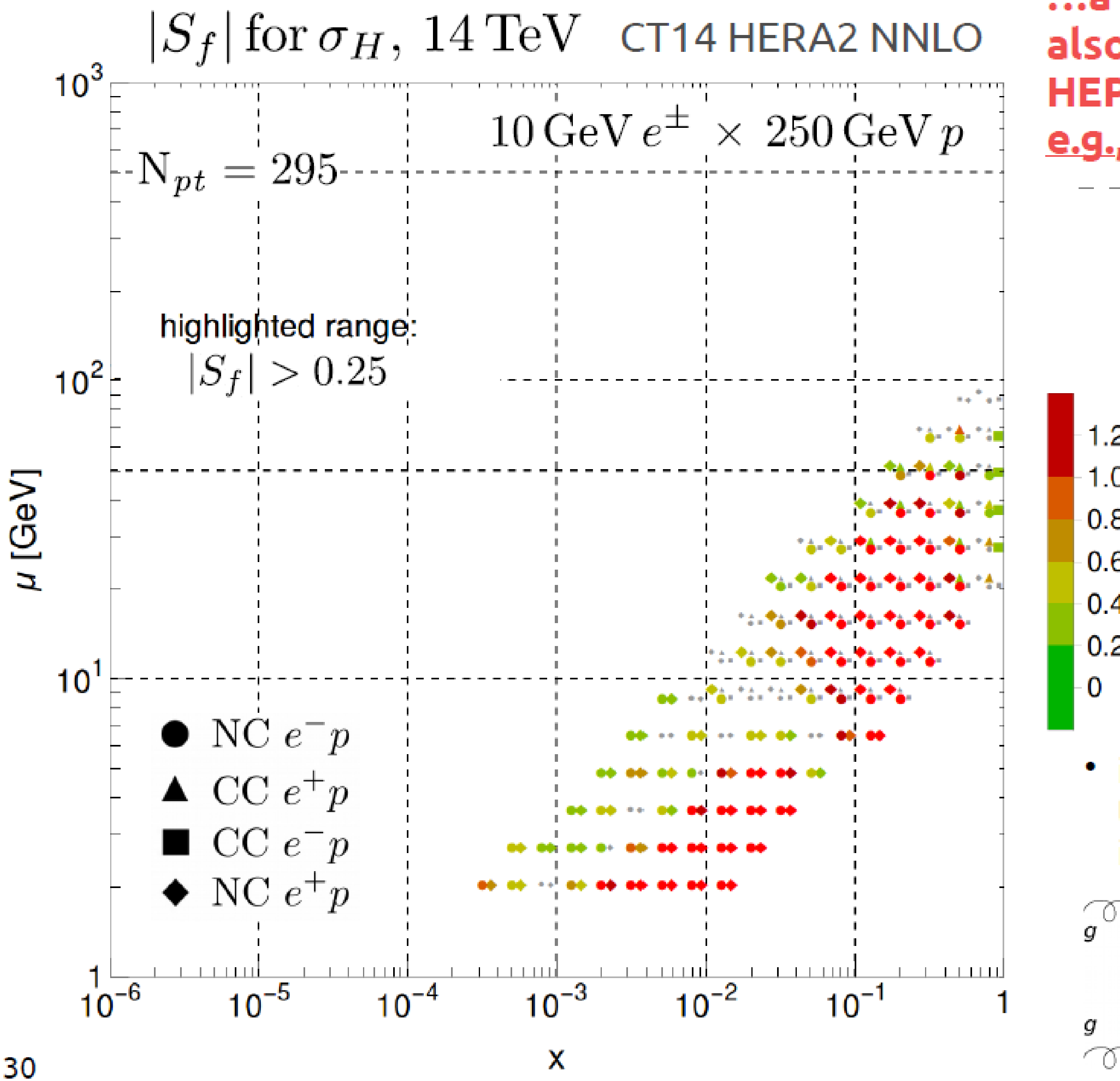


An EIC would drive lattice phenomenology

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



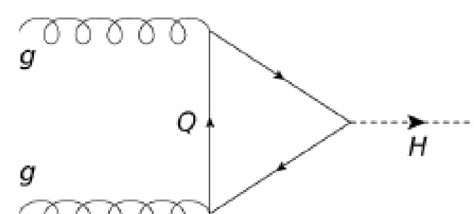




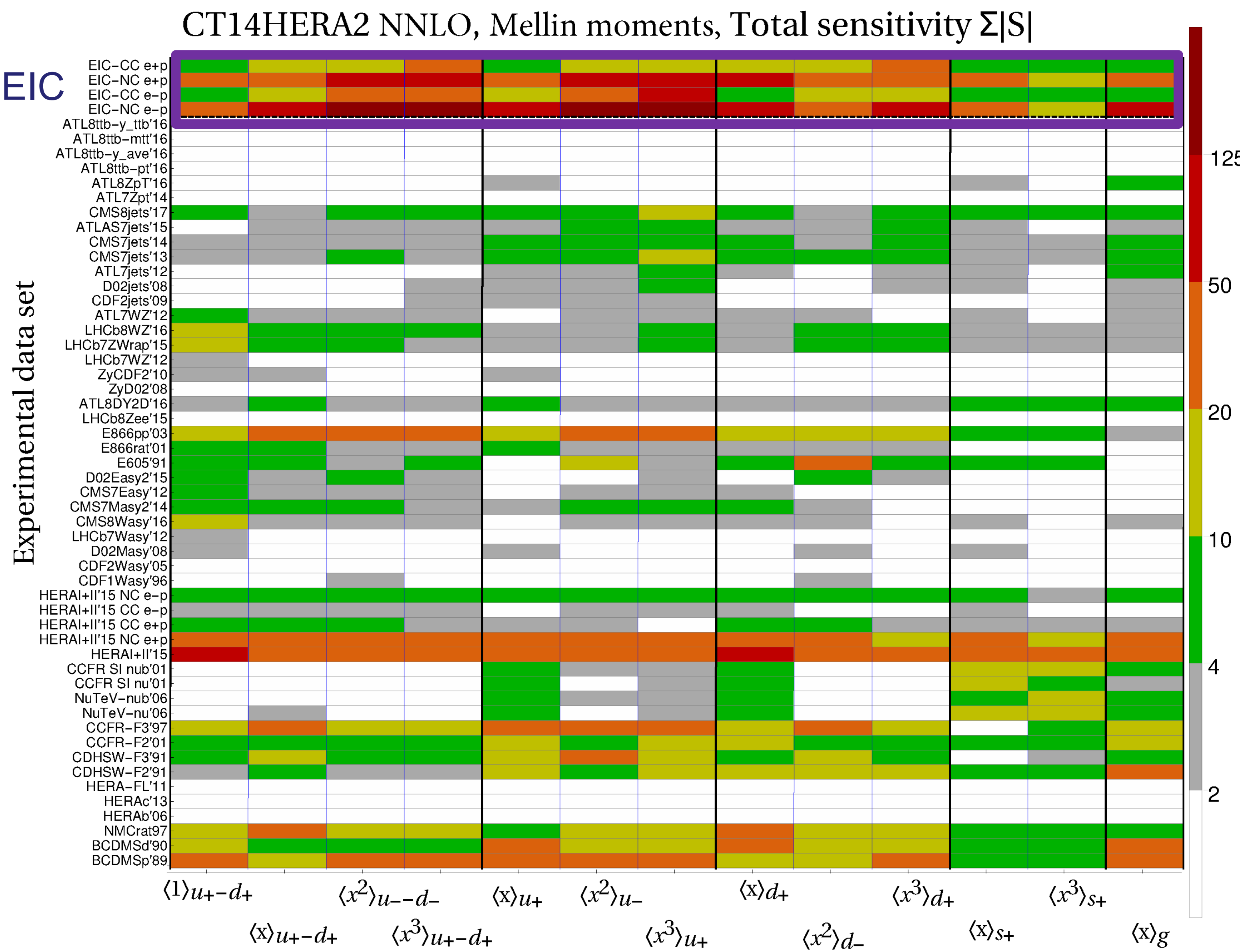
...a US-based EIC will also have important HEP consequences, e.g., on Higgs physics

- 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:



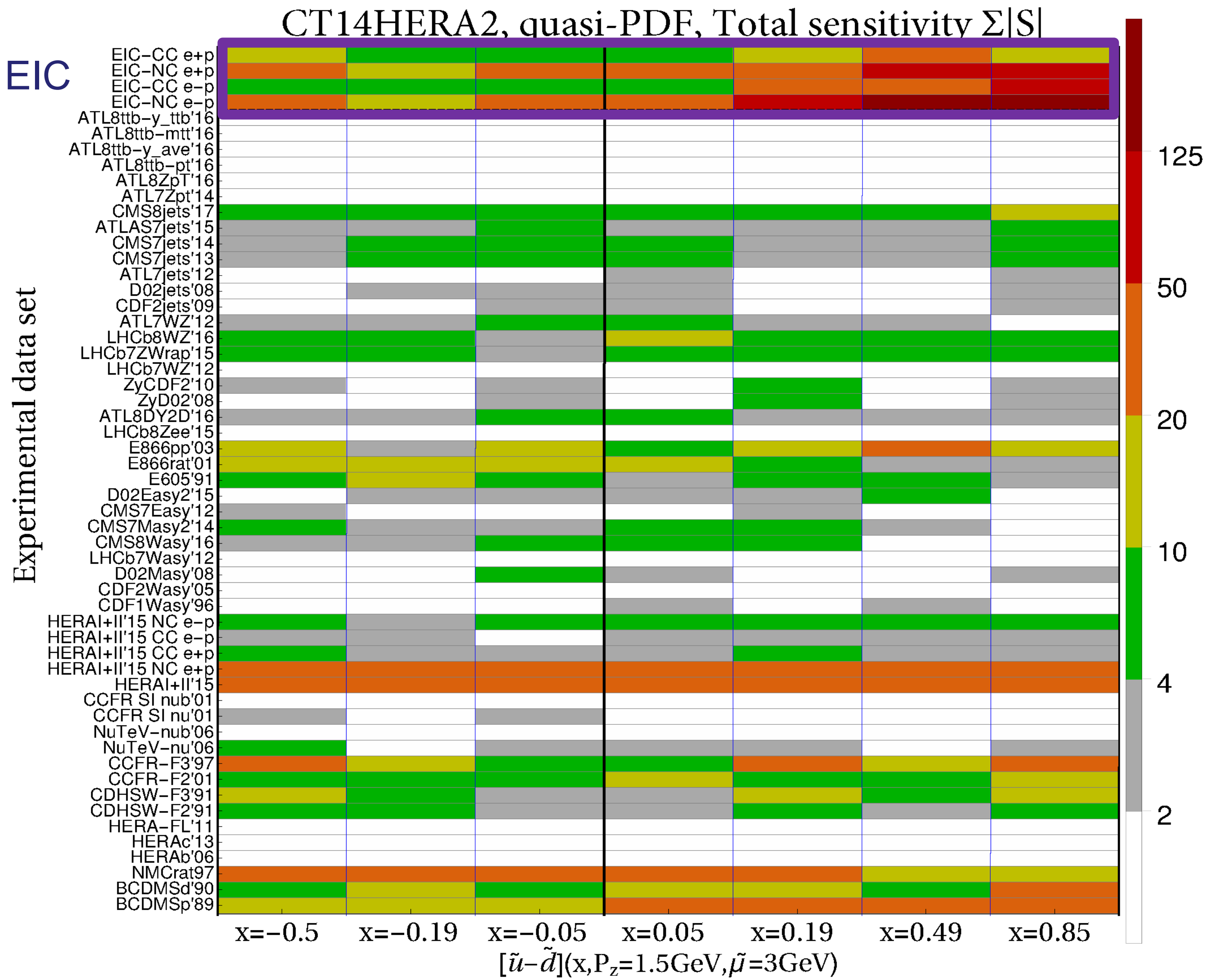
Total sensitivity to Mellin moments



We show Mellin moments computable on the lattice at the scale $\mu_F = 2\,\mathrm{GeV}$

Pre-2019 data: HERA, BCDMS, NMC, E866 DY pair production are most sensitive to the moments

Total sensitivity to lattice quasi-PDFs



$$\tilde{q}(x) \equiv \begin{cases} q(x), \\ x > 0; \\ -\bar{q}(|x|), \\ x < 0 \end{cases}$$

$Q = 3$ GeV,
$P_z = 1.5$ GeV

Key points: PDFSense

The PDFSense technique:

- Operates with the (theory-data) residuals evaluated with Hessian/MC error PDFs
- Accounts for PDF errors, uncorrelated and correlated experimental errors
- Produces kinematic maps and summary charts of data-point sensitivities to PDFs and PDF-dependent quantities
- Extendable to include new types of experiments and PDFs
- Fast; public code available at http://metapdf.hepforge.org/PDFSense

Key points: future experiments

A PDFSense study of kinematic reach and sensitivities of future pp and ep experiments

- Same metric for comparisons, independent of PDF tolerance
- Broadly agrees with the PDF reweighting studies
- The HL-LHC, EIC, and LHeC have **complementary potentials,** make a powerful physics case in combination. Highlights:
  - **HL-LHC:** reach to high Q, a variety of processes sensitive to PDFs; high sensitivity to the gluon, u and d antiquarks
  - **LHeC:** reach to $x < 10^{-6}$; high sensitivity to small-x gluon, as well as d and s quarks (esp. at $x \to 1$) in the clean $e^{\pm}p$ scattering environment; independent of BSM physics
  - **EIC:** supercedes the bulk of fixed-target DIS experiments; offers unique reach in the region of $x > 0.1$ necessary for tests of lattice QCD and nonperturbative QCD models; nucleon and nuclear beams; flavor separation via SIDIS; polarized beams

Thank you!

And many thanks to Bo-Ting, Tim, and Fred for the collaboration on these slides

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

Finding shifted residuals r_i from the covariance matrix

The CTEQ-TEA fit returns tables of $r_i(\vec{a})$ and $\bar{\lambda}_\alpha(\vec{a})$ for every i and α

Alternatively, they can be found from the covariance matrix:

$$r_i(\vec{a}) \;=\; s_i \sum_{j=1}^{N_{pt}} (\mathrm{cov}^{-1})_{ij}\; (T_j(\vec{a}) - D_j)\,, \qquad \overline{\lambda}_\alpha(\vec{a}) = \sum_{i,j=1}^{N_{pt}} (\mathrm{cov}^{-1})_{ij} \frac{\beta_{i\alpha}}{s_i} \frac{(T_j(\vec{a}) - D_j)}{s_j}$$

Vectors of data residuals

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

$$\vec{\delta}_i = \{\delta_{i,1}^{+}, \delta_{i,1}^{-}, \dots, \delta_{i,N}^{+}, \delta_{i,N}^{-}\} \qquad [\mathrm{N} = 28]$$

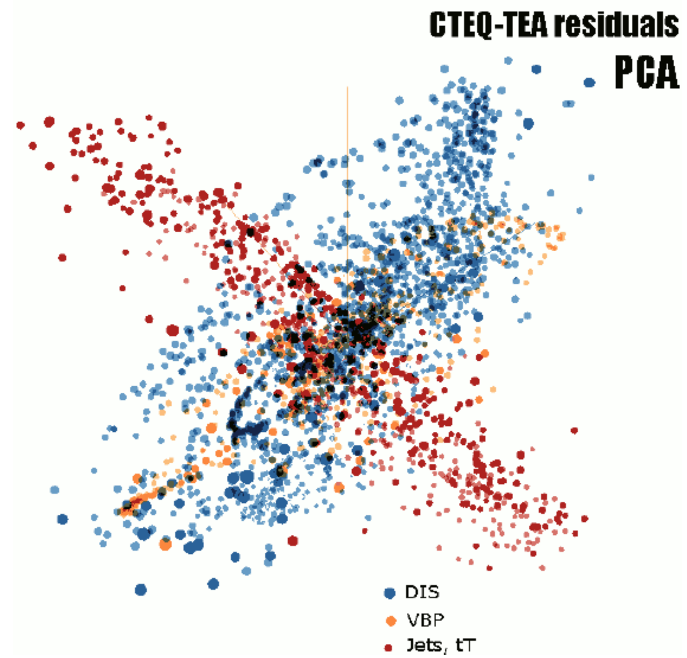
$$\delta_{i,k}^{\pm} \equiv \left(r_i(\vec{a}_k^{\pm}) - r_i(\vec{a}_0)\right) / \langle r_0 \rangle_E$$

-- 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

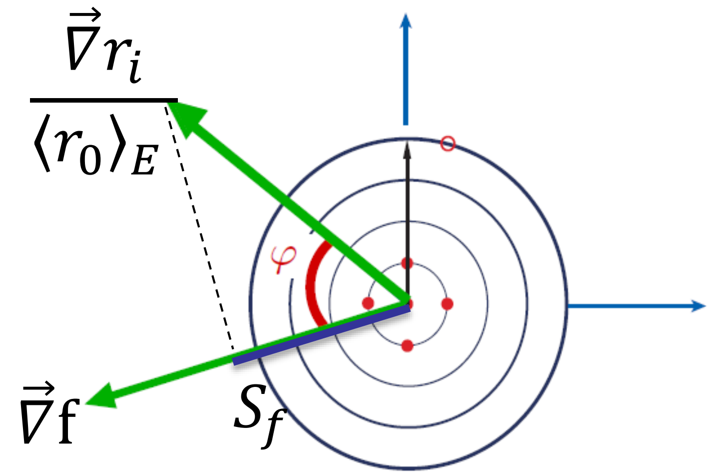
r_i is defined in the backup



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.

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

PDFSense maps, kinematical matchings

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

x_i	: parton mom. fraction
μ_i	: factorization scale

deeply-inelastic scattering:

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

hadron-hadron collisions:

$AB \to CX$

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

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