# Nucleon PDFs:
# what we know and what is missing?

**Juan Rojo,** VU Amsterdam & Nikhef

# Why Nucleon Structure?

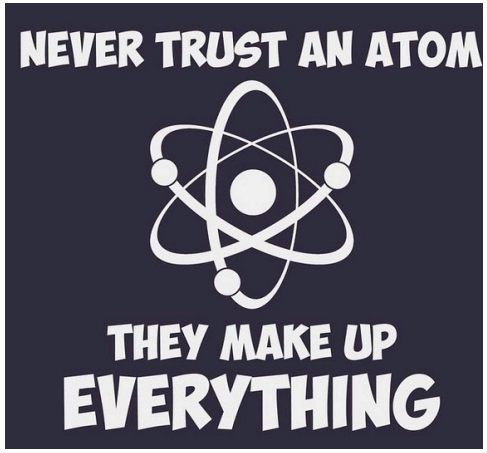
# Why Nucleon Structure?

One may claim that the nucleon is a rather ``boring'' particle, surely after one century of studying it, we know everything about the proton?

nothing farther from reality: the proton is a beautiful example of the richness of quantum mechanics: what a **proton is** depends on the **resolution with which we examine it**!

# Why Nucleon Structure?

nothing farther from reality: the proton is a beautiful example of the richness of quantum mechanics: what a **proton is** depends on the **resolution with which we examine it**!

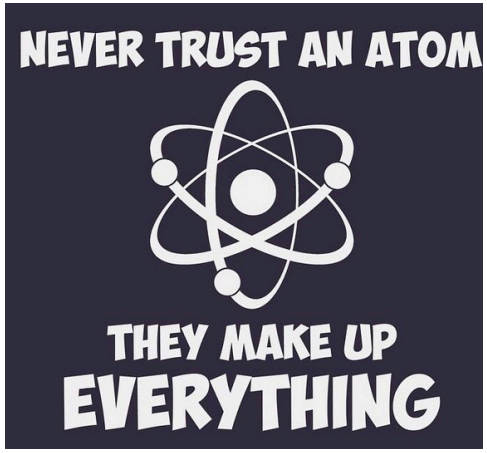


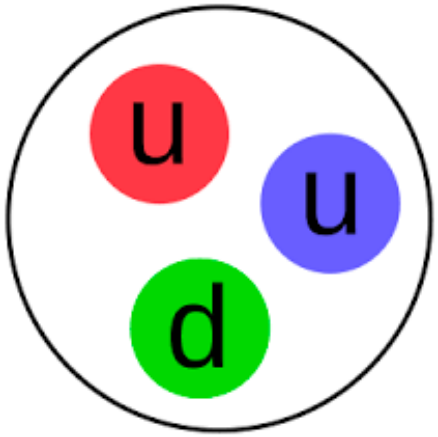

*long distances / low energies* → *short distances / high energies*

**a point particle**

# Why Nucleon Structure?

nothing farther from reality: the proton is a beautiful example of the richness of quantum mechanics: what a **proton is** depends on the **resolution with which we examine it**!



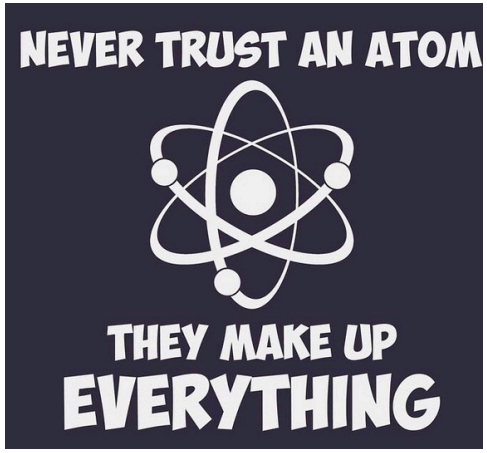

*long distances / low energies* → *short distances / high energies*

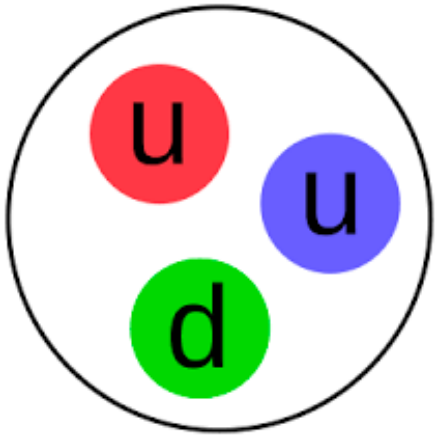
a point particle

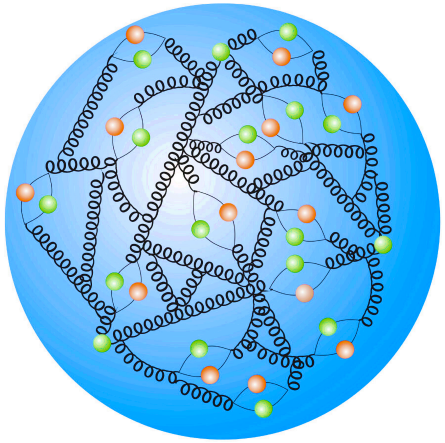
3 valence quarks

# Why Nucleon Structure?

nothing farther from reality: the proton is a beautiful example of the richness of quantum mechanics: what a **proton is** depends on the **resolution with which we examine it**!



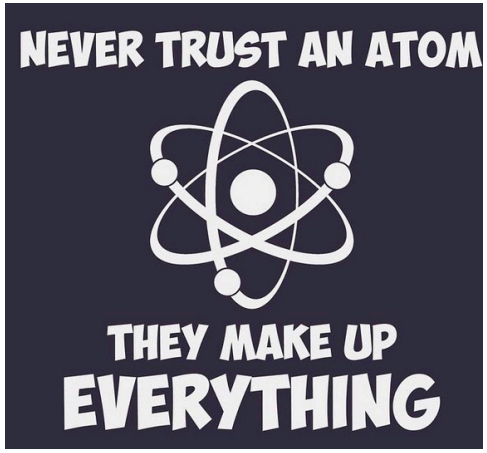



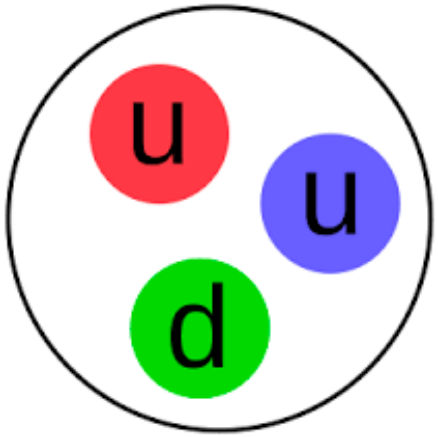
*long distances / low energies* → *short distances / high energies*

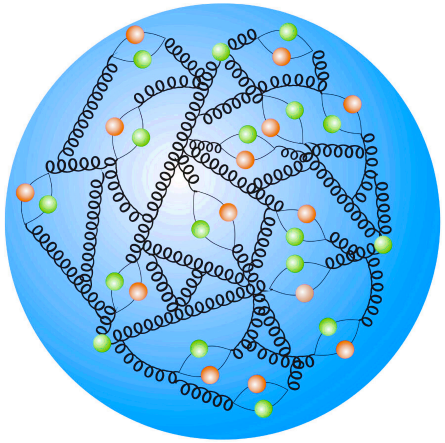
**a point particle** | **3 valence quarks** | **sea quarks, gluons**

# Why Nucleon Structure?

nothing farther from reality: the proton is a beautiful example of the richness of quantum mechanics: what a **proton is** depends on the **resolution with which we examine it**!



*long distances / low energies* → *short distances / high energies*

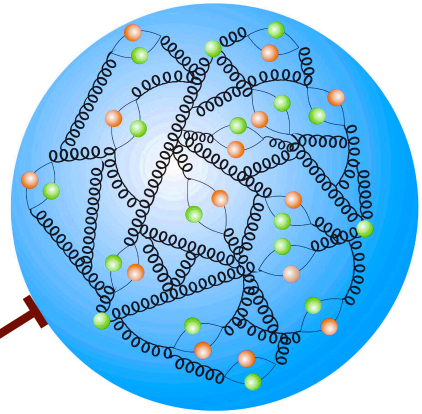
a point particle

3 valence quarks

sea quarks, gluons

heavy quarks, photons, leptons, gauge and Higgs bosons …

# Why Nucleon Structure?



***Key component of predictions for particle, nuclear, and astro-particle experiments***

- **pp: ATLAS, CMS, LHCb, ALICE**
- ep: fixed target DIS, HERA
- neutrinos: IceCube, KM3NET, Forward Physics Facility @ LHC
- heavy ions: LHC Pb, LHC O, RHIC
- *pp (future): HL-LHC, FCC, SppS*
- ep (future): **Electron-Ion Collider,** LHeC, FCC-eh

***LHC master formula***

*hard cross-section*

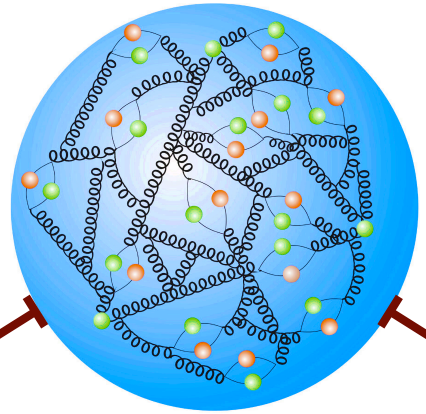
$$\sigma(M,s) \propto \sum_{ij=u,d,g,\ldots} \int_{M^2}^{s} d\hat{s}\ \mathscr{L}_{ij}(\hat{s},s)\,\widetilde{\sigma}_{ij}(\hat{s},\alpha_s(M))$$

*partonic luminosities*

*parton distributions*

$$\mathscr{L}_{ij}(Q,s) = \frac{1}{s}\int_{Q^2/s}^{1} \frac{dx}{x} f_i\left(\frac{Q^2}{sx},Q\right) f_j\left(x,Q\right)$$

# Why Nucleon Structure?



***Key component of predictions for particle, nuclear, and astro-particle experiments***

- **pp: ATLAS, CMS, LHCb, ALICE**
- ep: fixed target DIS, HERA
- neutrinos: IceCube, KM3NET, Forward Physics Facility @ LHC
- heavy ions: LHC Pb, LHC O, RHIC
- *pp (future): HL-LHC, FCC, SppS*
- ep (future): **Electron-Ion Collider,** LHeC, FCC-eh

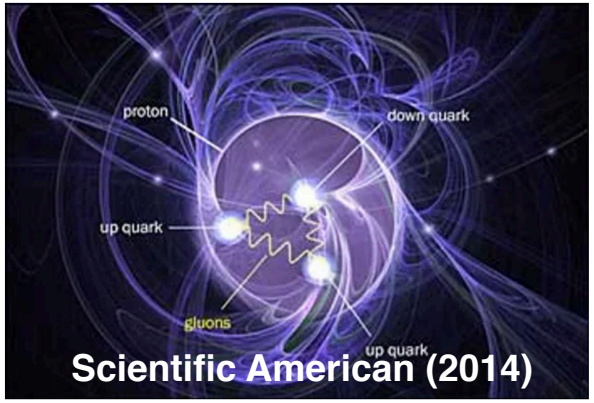
***Address fundamental questions about Quantum Chromodynamics***

- origin of mass & spin
- **heavy quark** & **antimatter content**
- 3D imaging
- gluon-dominated matter
- **nuclear modifications**
- Interplay with BSM e.g. via ``**SMEFT PDFs**”

# PDFs: a gateway to unravelling QCD

THE SCIENCES

**Proton Spin Mystery Gains a New Clue**




*Non-zero gluon polarisation*


*Intrinsic Charm*

The proton keeps surprising us as an endless source of **fundamental discoveries!**

QUANTUM PHYSICS

**Decades-Long Quest Reveals Details of the Proton's Inner Antimatter**

27

*Twenty years ago, physicists set out to investigate a mysterious asymmetry in the proton's interior. Their results, published today, show how antimatter helps stabilize every atom's core.*

*Antimatter asymmetry*


**After 40 years of studying the strong nuclear force, a revelation**

*BFKL dynamics*

**This was the year that analysis of data finally backed up a prediction, made in the mid 1970s, of a surprising emergent behaviour in the strong nuclear force**

# Nucleon Structure: what do we know

*Here I focus on unpolarised nucleon structure. Dedicated talks will cover polarised PDFs (Christine Aidala), nuclear PDFs (Pia Zurita), and TMDs (Zhongbo Kang). Also for lattice QCD calculations of PDFs, I refer to dedicated talks in this workshop*

Edinburgh 2021/12
Nikhef-2021-013
TIF-UNIMI-2021-11

## The Path to Proton Structure at One-Percent Accuracy

**The NNPDF Collaboration:**

Richard D. Ball,[1] Stefano Carrazza,[2] Juan Cruz-Martinez,[2] Luigi Del Debbio,[1] Stefano Forte,[2] Tommaso Giani,[1,8] Shayan Iranipour,[3] Zahari Kassabov,[3] Jose I. Latorre,[4,5,6] Emanuele R. Nocera,[1,8] Rosalyn L. Pearson,[1] Juan Rojo,[7,8] Roy Stegeman,[2] Christopher Schwan,[2] Maria Ubiali,[3] Cameron Voisey,[9] and Michael Wilson[1]

[1] *The Higgs Centre for Theoretical Physics, University of Edinburgh, JCMB, KB, Mayfield Rd, Edinburgh EH9 3JZ, Scotland*
[2] *Tif Lab, Dipartimento di Fisica, Università di Milano and INFN, Sezione di Milano, Via Celoria 16, I-20133 Milano, Italy*
[3] *DAMTP, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, United Kingdom*
[4] *Quantum Research Centre, Technology Innovation Institute, Abu Dhabi, UAE*
[5] *Center for Quantum Technologies, National University of Singapore, Singapore*
[6] *Qilimanjaro Quantum Tech, Barcelona, Spain*
[7] *Department of Physics and Astronomy, Vrije Universiteit, NL-1081 HV Amsterdam*
[8] *Nikhef Theory Group, Science Park 105, 1098 XG Amsterdam, The Netherlands*
[9] *Cavendish Laboratory, University of Cambridge, Cambridge, CB3 0HE, United Kingdom*



### Abstract

We present a new set of parton distribution functions (PDFs) based on a fully global dataset and machine learning techniques: NNPDF4.0. We expand the NNPDF3.1 determination with 44 new datasets, mostly from the LHC. We derive a novel methodology through hyperparameter optimisation, leading to an efficient fitting algorithm built upon stochastic gradient descent. We use NNLO QCD calculations and account for NLO electroweak corrections and nuclear uncertainties. Theoretical improvements in the PDF description include a systematic implementation of positivity constraints and integrability of sum rules. We validate our methodology by means of closure tests and "future tests" (i.e. tests of backward and forward data compatibility), and assess its stability, specifically upon changes of PDF parametrization basis. We study the internal compatibility of our dataset, and investigate the dependence of results both upon the choice of input dataset and of fitting methodology. We perform a first study of the phenomenological implications of



## An open-source machine learning framework for global analyses of parton distributions

**The NNPDF Collaboration: Richard D. Ball · Stefano Carrazza · Juan Cruz-Martinez · Luigi Del Debbio · Stefano Forte · Tommaso Giani · Shayan Iranipour · Zahari Kassabov · Jose I. Latorre · Emanuele R. Nocera · Rosalyn L. Pearson · Juan Rojo · Roy Stegeman · Christopher Schwan · Maria Ubiali · Cameron Voisey · Michael Wilson**

the date of receipt and acceptance should be inserted later

**Abstract** We present the software framework underlying the NNPDF4.0 global determination of parton distribution functions (PDFs). The code is released under an open source licence and is accompanied by extensive documentation and examples. The code base is composed by a PDF fitting package, tools to handle experimental data and to efficiently compare it to theoretical predictions, and a versatile analysis framework. In addition to ensuring the reproducibility of the NNPDF4.0 (and subsequent) determination, the public release of the NNPDF fitting framework enables a number of phenomenological applications and the production of PDF fits under user-defined data and theory assumptions.

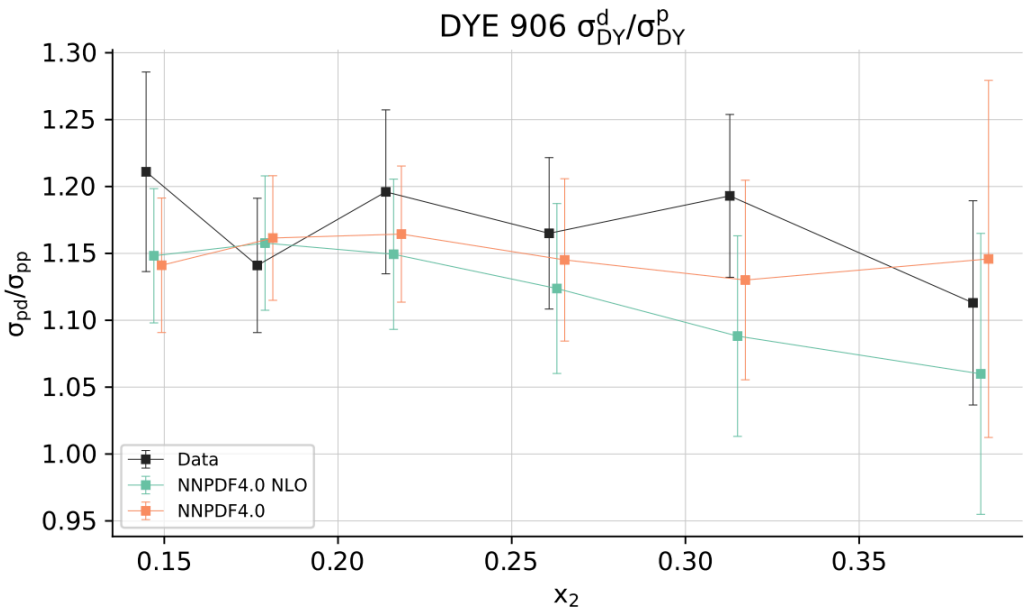
Edinburgh 2021/13
Nikhef-2021-020
TIF-UNIMI-2021-12

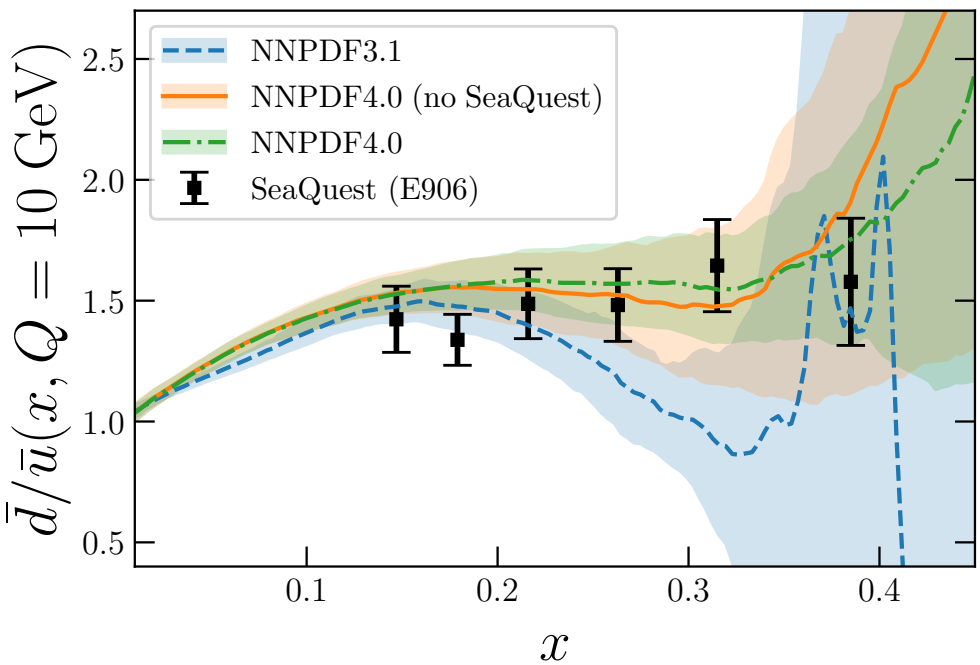
### Contents



***most results presented here based on the recent NNPDF4.0 global determination of proton structure - note that fitting framework is now available open source***

# Antimatter asymmetry in nucleons




☑ The recent SeaQuest data presented as evidence for **quark sea (``proton antimatter”) asymmetry**

$$\frac{\sigma_{\mathrm{DY,deuterium}}}{\sigma_{\mathrm{DY,hydrogen}}} \approx 1 + \frac{\bar{d}_p(x_t)}{\bar{u}_p(x_t)}$$

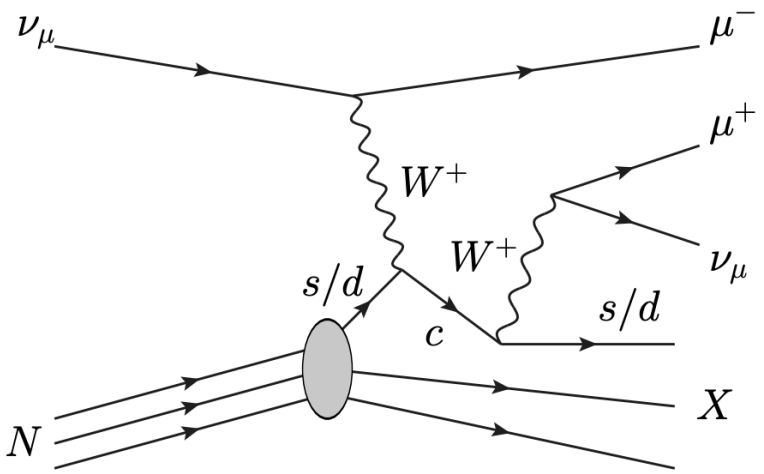
***with several approximations***

☑ The NNPDF4.0 fit prediction (without SeaQuest) agrees with these measurements, **mild impact**

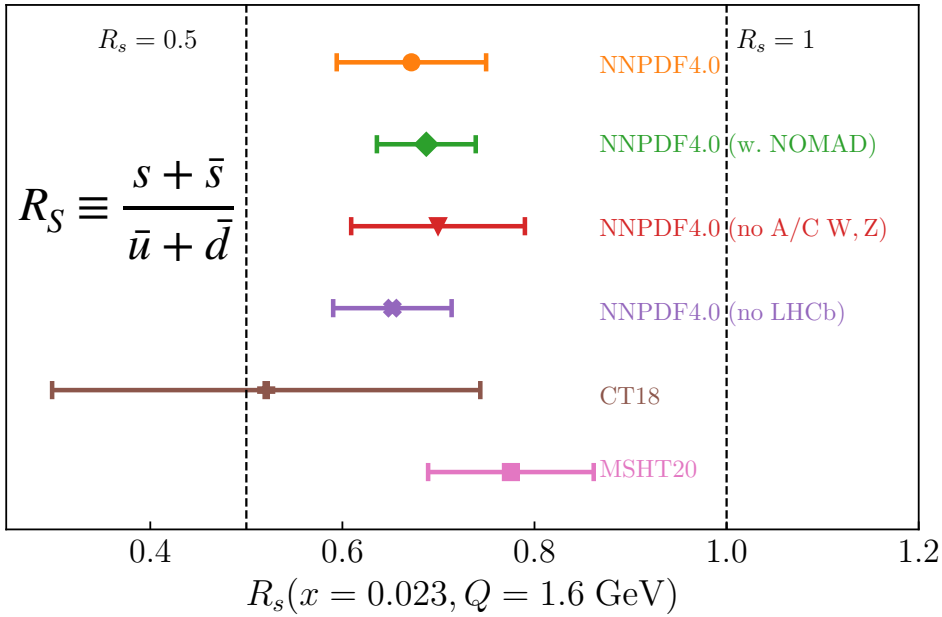
The nucleon exhibits a marked **light sea quark asymmetry** $\bar{d}(x)/\bar{u}(x) \geq 1\,, \quad x \gtrsim 10^{-2}$

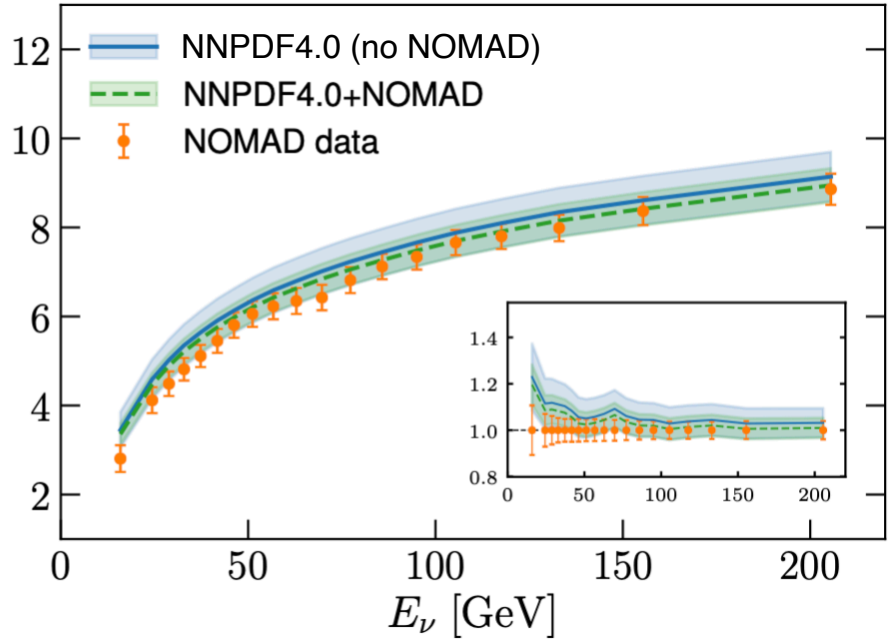
***specially sizeable at large-x***

# The strangest proton




- Is the proton strangeness **suppressed** or **symmetric?**
- NNPDF4.0 strangeness **stable** upon removal of subsets of LHC data (e.g. *W, Z* distributions)
- Dimuon DIS data sensitive to **strangeness**: NNPDF4.0 predictions consistent with the **NOMAD dimuon data**
- **No tension** between LHC and DIS neutrino data!



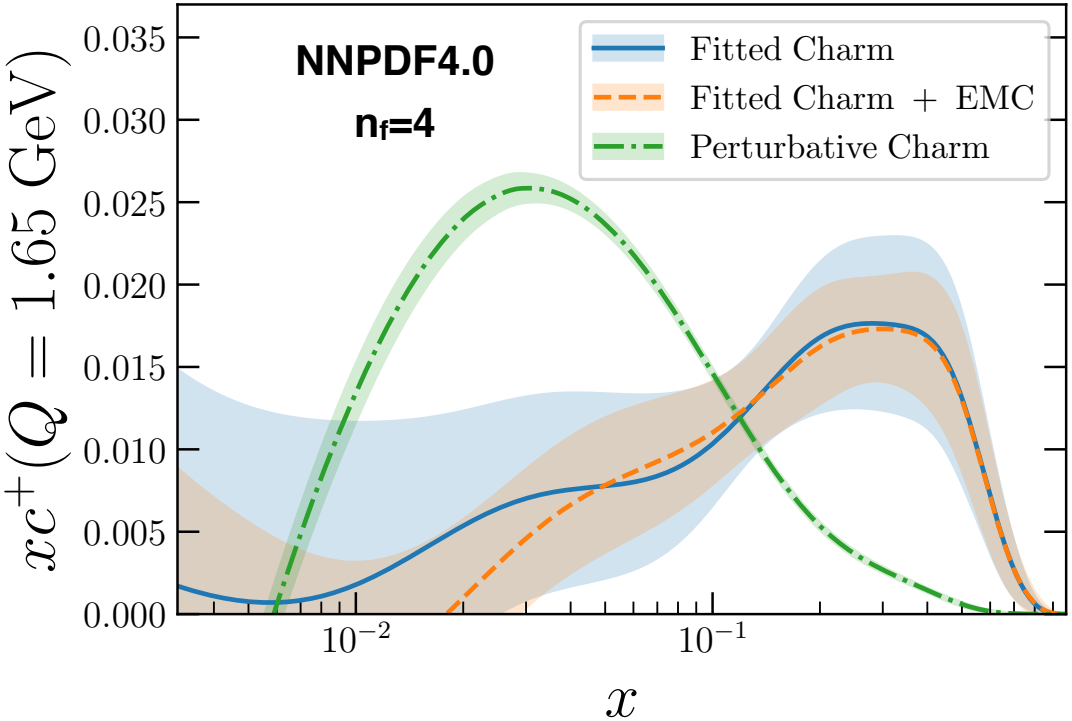



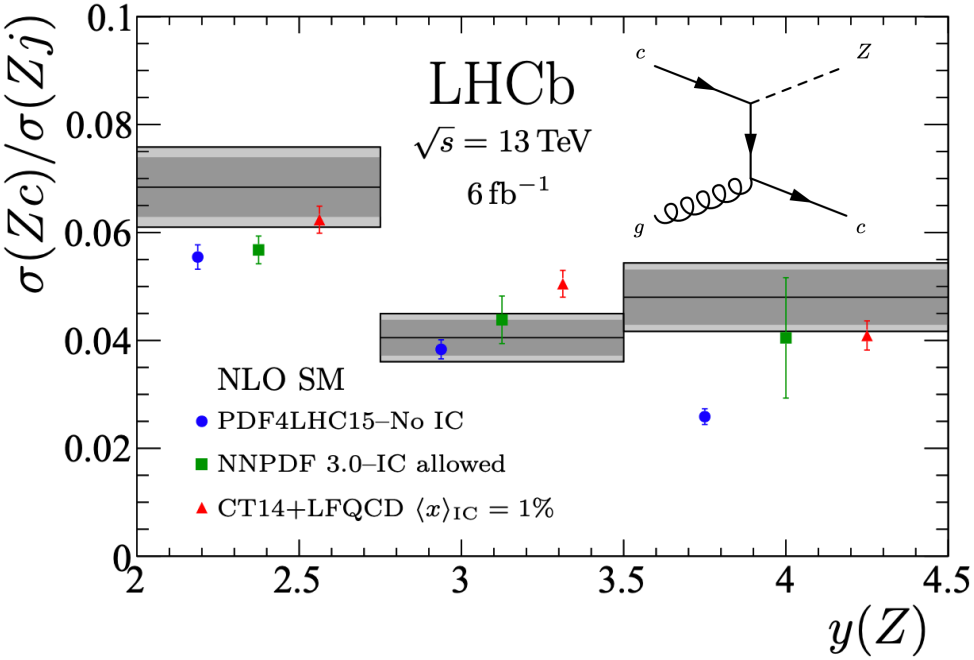

The proton strange content is **moderately suppressed** with respect to up and down sea

$$R_S(0.023, 1.6\ \text{GeV}) = 0.68 \pm 0.05$$

***including NOMAD data***

# Intrinsic Charm




- ☑ Increasing evidence for **non-perturbative charm component** within the proton
- ☑ Bulk of constraints provided by new **precision LHC data**, complemented by fixed-target DIS

  ***and consistent with EMC charm DIS measurements***
- ☑ Consistent with recent LHCb measurement of **forward *Z+D* production**, directly sensitive to the (large-*x*) charm content of the nucleon

**agreement between indirect (global PDF fit) and direct (LHCb Z+D) constraints on the charm PDF**

# Intrinsic Charm

**Disentangle** the perturbative from the intrinsic component of the charm PDF

**4FNS**

*Extract from global PDF fit*

$$c^{(n_f=4)}(x, Q_0) \simeq \left( c^{(n_f=4)}_{(\mathrm{pert})}(x, Q_0) + c^{(n_f=4)}_{(\mathrm{intr})}(x, Q_0) \right), \quad Q_0 = 1.65 \text{ GeV}$$

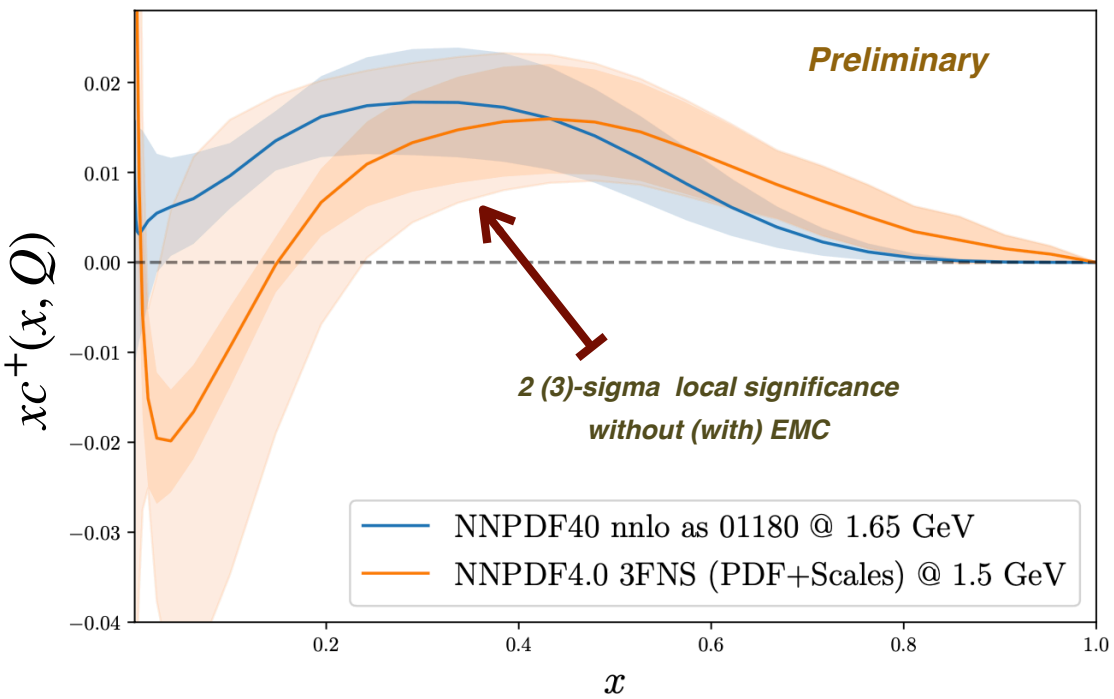
$$\mu_c = m_c = 1.51 \text{ GeV}$$

*Backwards DGLAP + matching conditions*

**3FNS**

*Intrinsic component of the charm PDF*

$$c^{(n_f=3)}(x, Q) \simeq c^{(n_f=3)}_{(\mathrm{intr})}(x), \quad Q = 1.5 \text{ GeV}$$




A **non-zero intrisic charm** a la BHPS favoured by both global PDF fit and LHCb Z+D data

**3-sigma** local significance in 3FNS

Charm momentum fraction:

$$(4\mathrm{FNS}) \quad [c] = 0.86 \pm 0.15_{\mathrm{pdf}}$$

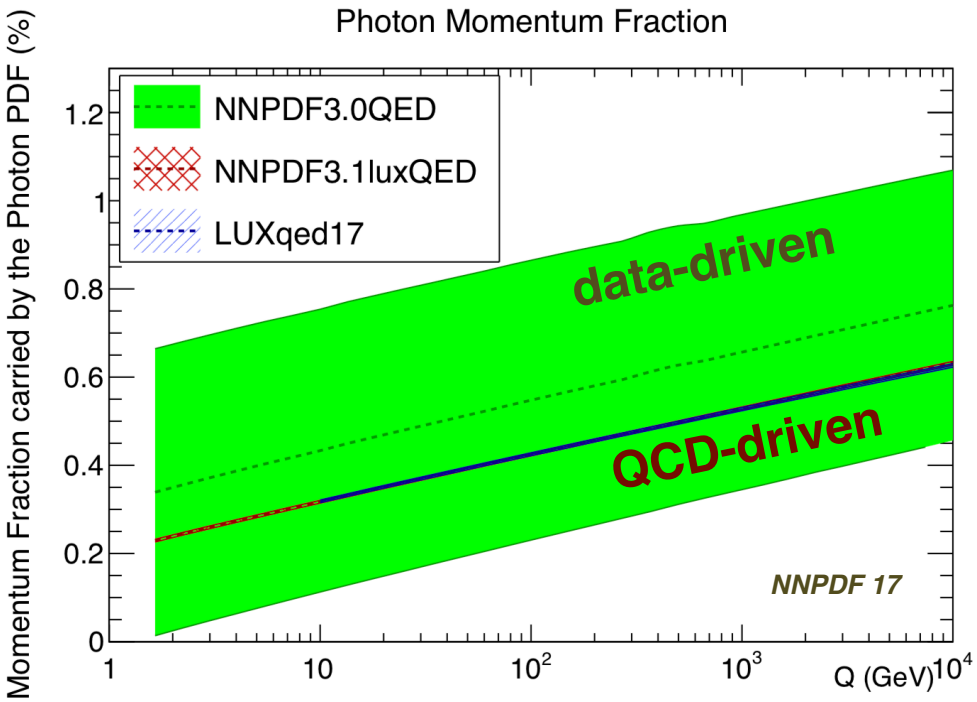
$$(3\mathrm{FNS}) \quad [c] = 0.60 \pm 0.55_{\mathrm{pdf+mhou}}$$

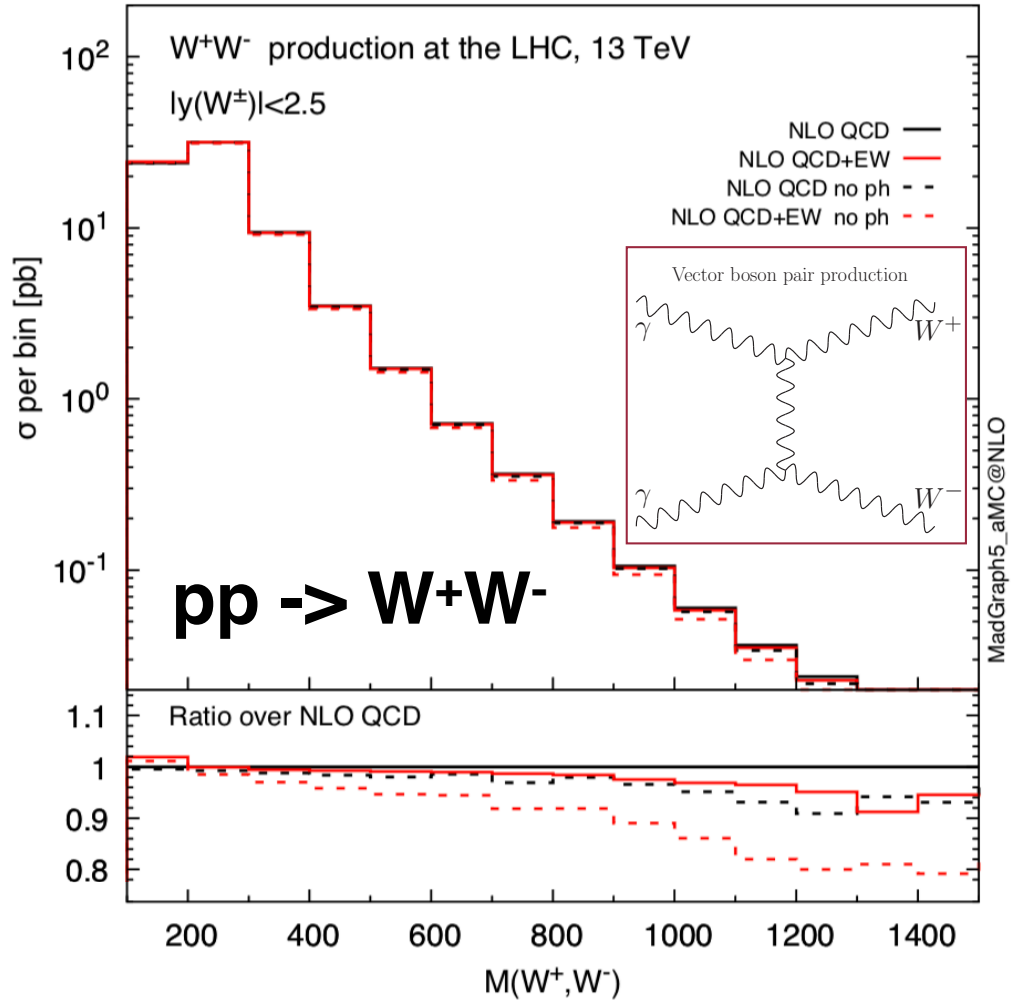
# The photon content of the proton

- The proton contains not only quark and gluons as constituents, but also **photons**
- The photon PDF can be evaluated from deep-inelastic **structure functions $F_2$ and $F_L$**

*LuxQED: Manohar et al 17*

- Required for consistent implementation of **electroweak corrections** at the LHC

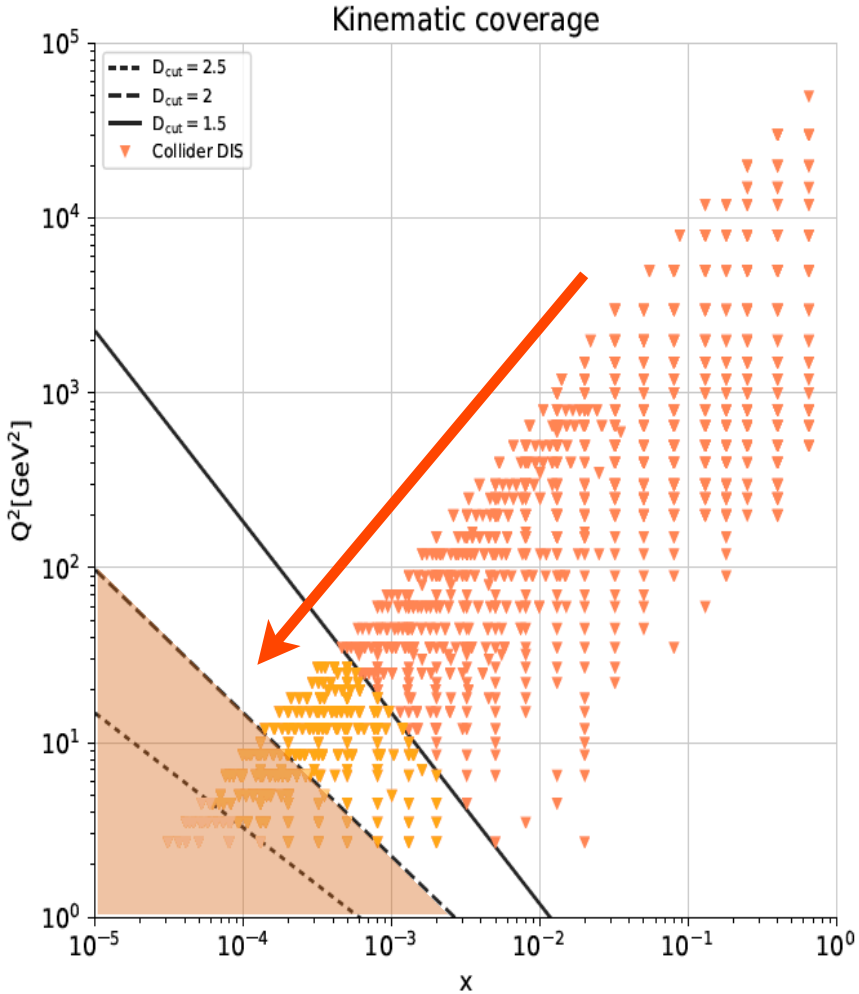


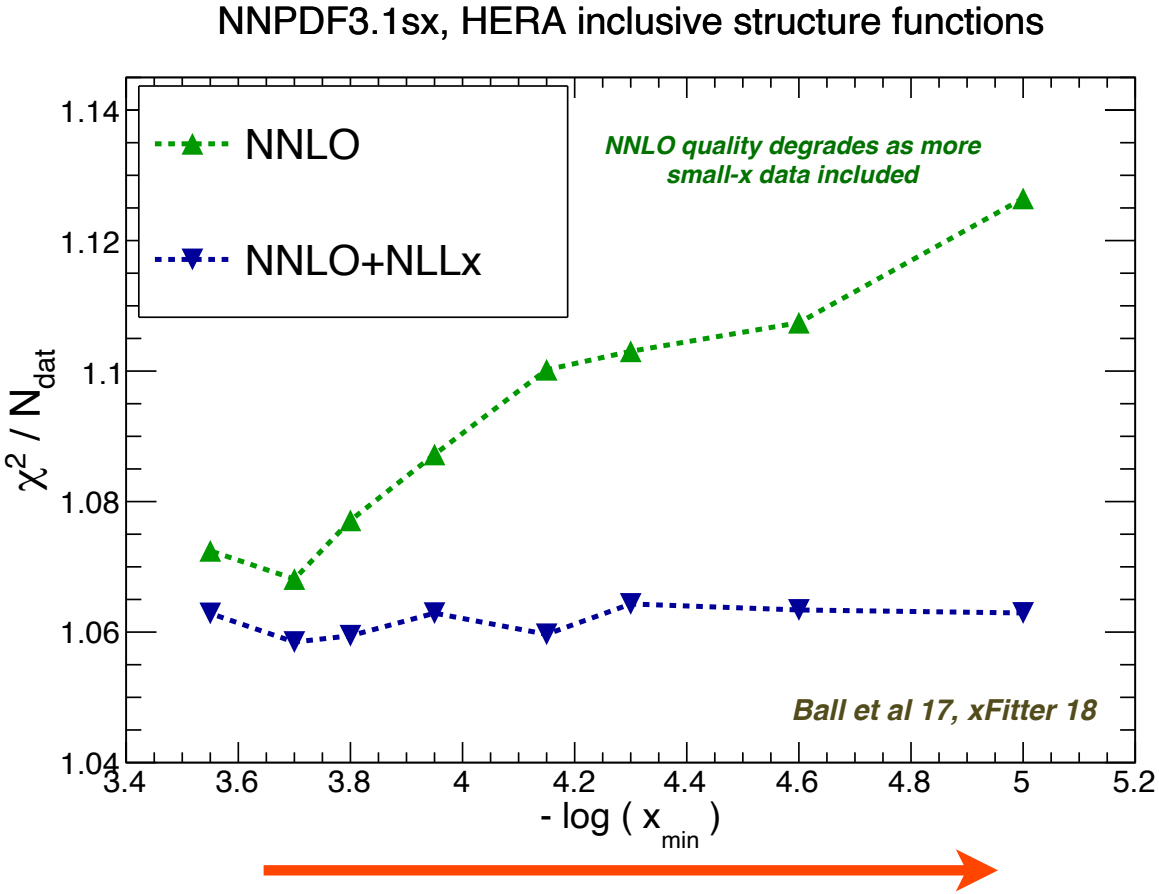

*Up to **0.6% of proton's momentum** is carried by photons*

# Evidence for BFKL dynamics

NNPDF3.1 fits based on **fixed order** (NNLO) and **small-*x* resumed** (NNLO+NLLx) theory




Monitor the fit quality in small-*x* region




Best description of **small-*x* HERA data: BFKL resummation**

**BFKL dynamics stablished**: crucial for description of low-x physics (also @ EIC, HL-LHC, LHeC …)

# Nucleon Structure: what are we missing

*a few representative examples only!*

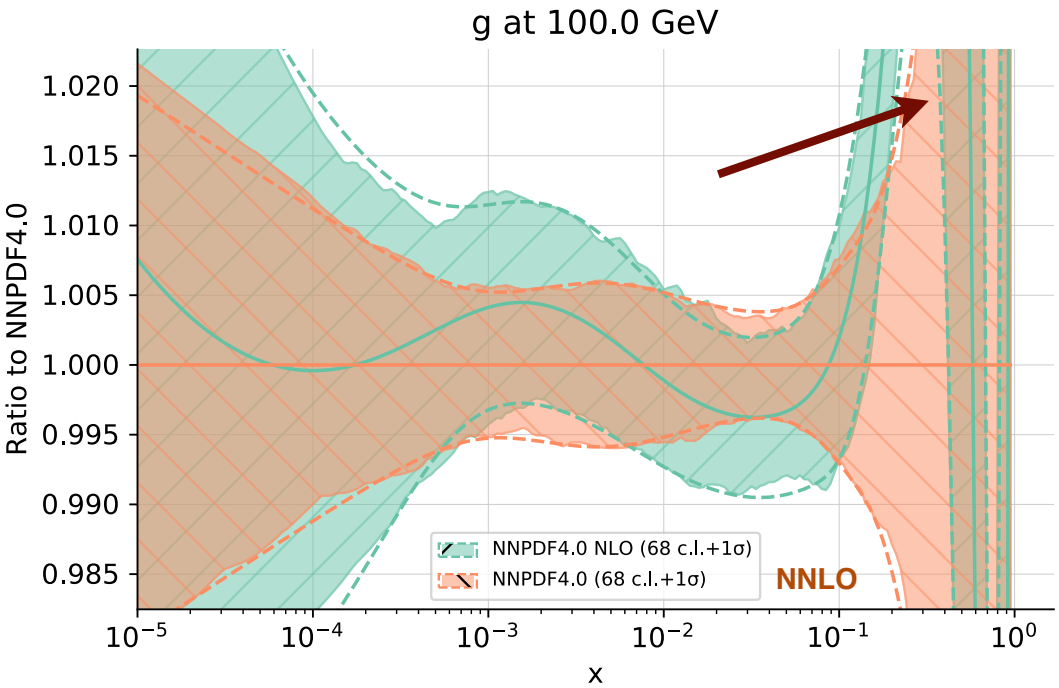
# QCD uncertainties in PDF fits

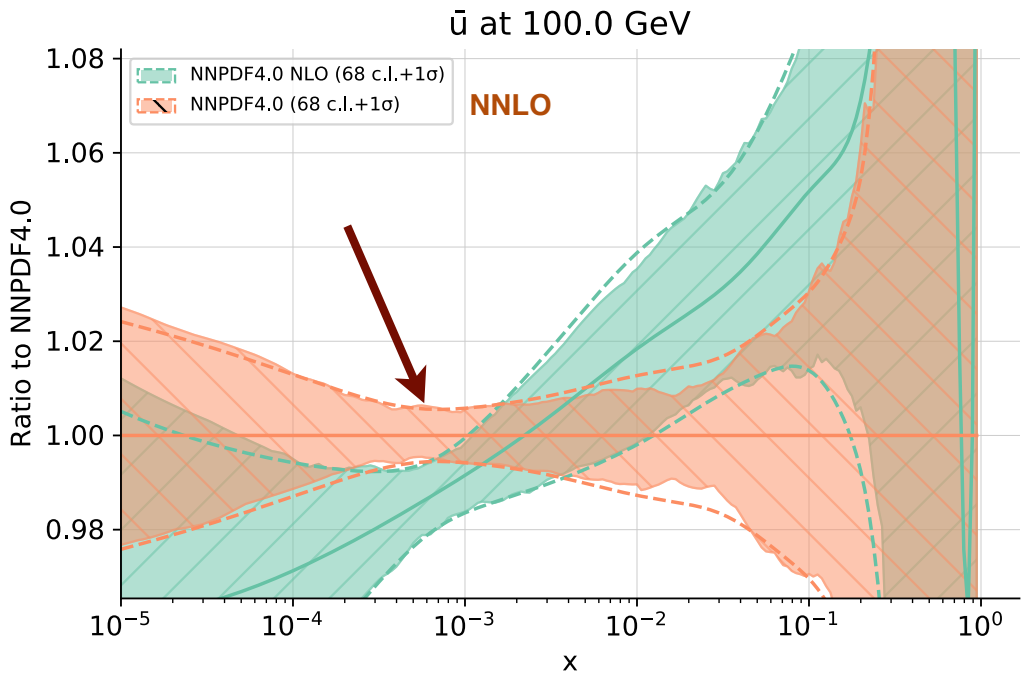
Standard global PDF fits are based on **fixed-order QCD calculations**

$$\sigma = \alpha_s^p \sigma_0 + \alpha_s^{p+1} \sigma_1 + \alpha_s^{p+2} \sigma_2 + \mathcal{O}(\alpha_s^{p+3})$$

The truncation of the perturbative series has associated a theoretical uncertainty: **Missing Higher Order (MHO)** uncertainty

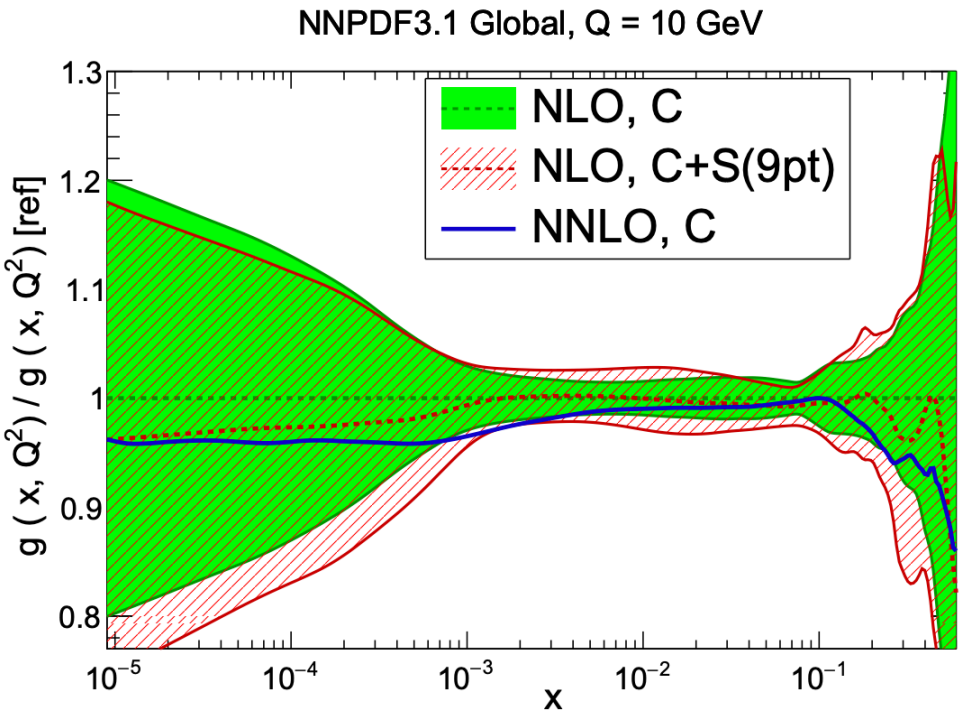
How severe is **ignoring MHOUs** in modern global PDFs fits?

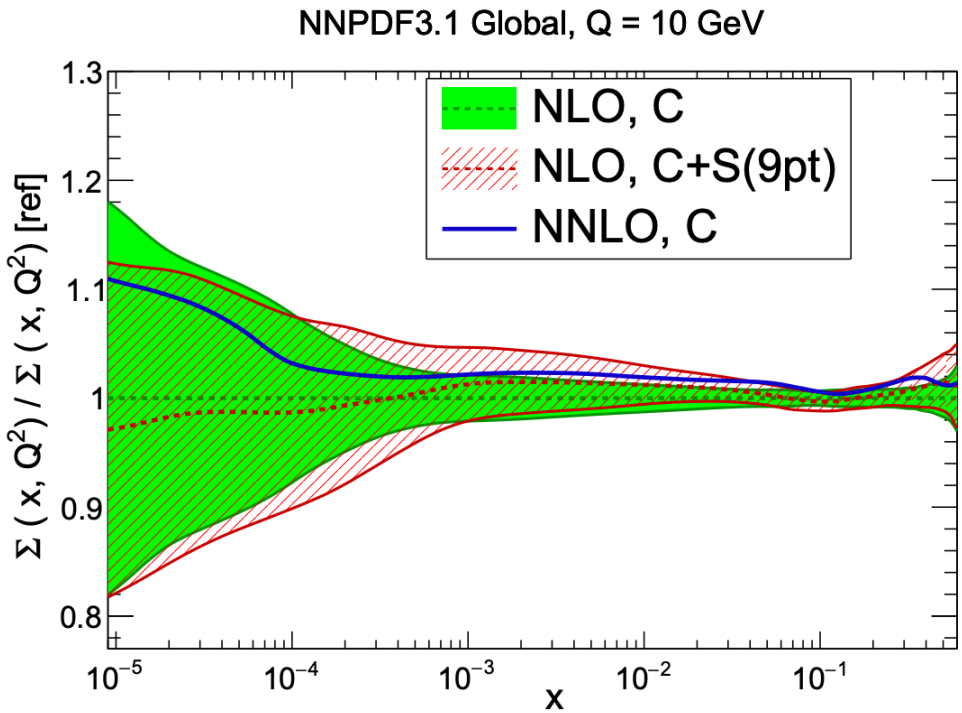



**MHOUs are comparable with PDF errors certainly at NLO, likely also at NNLO**

# QCD uncertainties in PDF fits




$$\chi^2_{\mathrm{C+S}} = \frac{1}{n_{\mathrm{dat}}} \sum_{ij} \left(T_i - D_i\right) \left(\mathrm{cov}^{\mathrm{exp}} + \mathrm{cov}^{\mathrm{mhou}}\right)^{-1}_{ij} \left(T_j - D_j\right)$$

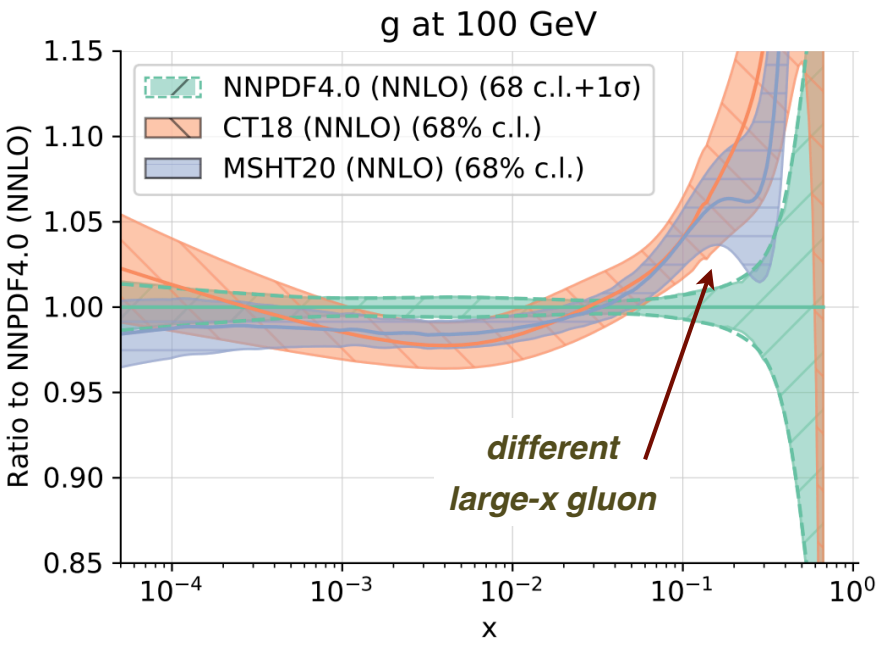
**experimental uncertainties** **theory uncertainties from MHOUs**

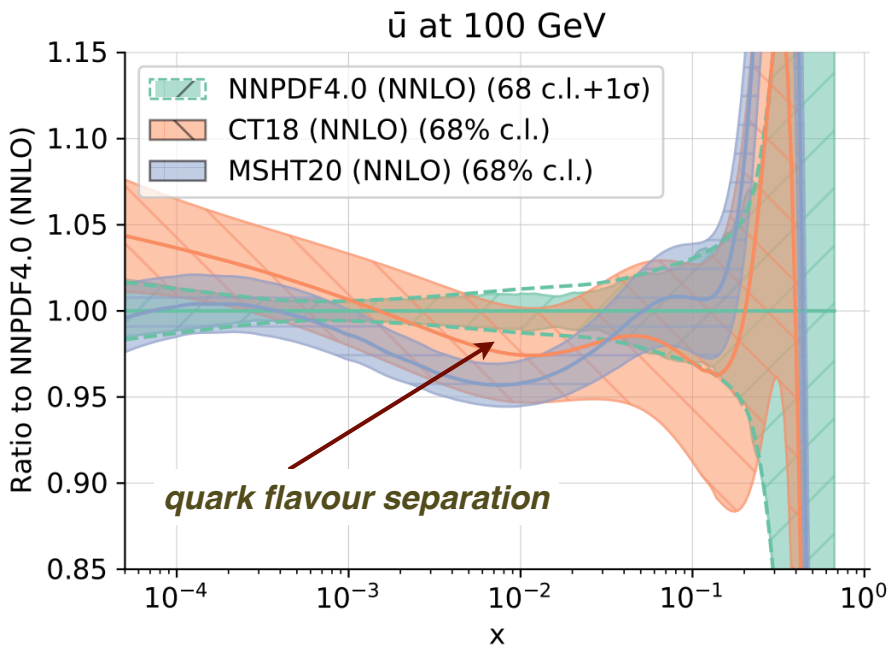
A NLO global fit with MHOUs highlights how these cannot be neglected, both in terms of **accuracy** and of **precision**

**Ongoing work towards a NNPDF4.0 NNLO global fit with MHOUs**

# (dis-)agreements in global fits

reasonable agreement with CT18 and MSHT20 in most cases, but some important **differences**




also some distinct patterns concerning the magnitude of the **PDF uncertainties**

- Need dedicated measurements on processes sensitive to PDF regions where **global PDF fits disagree,** such as inclusive jets & dijets & top quark to constrain the **large-*x* gluon**
- Crucial information will also be provided on measurements that are now **limited by statistics:** large-$p_T$ tails, rare processes, ….
- Several statistical challenges in interpreting **LHC measurements limited by systematics**

# (dis-)agreements in global fits

**pattern of PDF uncertainties** follows (qualitatively) the **dataset size**

$$\delta_{\rm PDF}(\text{CT18}) \gtrsim \delta_{\rm PDF}(\text{MSHT20}) \gtrsim \delta_{\rm PDF}(\text{NNPDF4.0})$$

$$N_{\rm data}(\text{CT18}) \lesssim N_{\rm data}(\text{MSHT20}) \lesssim N_{\rm data}(\text{NNPDF4.0})$$

| Data set | Ref. | NNPDF3.1 | NNPDF4.0 | ABMP16 | CT18 | MSHT20 |
|---|---|---|---|---|---|---|
| ATLAS $W, Z$ 7 TeV ($\mathcal{L} = 35$ pb$^{-1}$) | [51] | ✓ | ✓ | ✓ | ✓ | ✓ |
| ATLAS $W, Z$ 7 TeV ($\mathcal{L} = 4.6$ fb$^{-1}$) | [52] | ✓ | ✓ | ✗ | (✓) | ✓ |
| ATLAS low-mass DY 7 TeV | [53] | ✓ | ✓ | ✗ | (✓) | ✗ |
| ATLAS high-mass DY 7 TeV | [54] | ✓ | ✓ | ✗ | (✓) | ✓ |
| ATLAS $W$ 8 TeV | [79] | ✗ | (✓) | ✗ | ✗ | ✓ |
| ATLAS DY 2D 8 TeV | [78] | ✗ | ✓ | ✗ | ✗ | ✓ |
| ATLAS high-mass DY 2D 8 TeV | [77] | ✗ | ✓ | ✗ | (✓) | ✓ |
| ATLAS $\sigma_{W,Z}$ 13 TeV | [81] | ✗ | ✓ | ✓ | ✗ | ✗ |
| ATLAS $W$+jet 8 TeV | [93] | ✗ | ✓ | ✗ | ✗ | ✓ |
| ATLAS $Z$ $p_T$ 7 TeV | [259] | (✓) | ✗ | ✗ | (✓) | ✗ |
| ATLAS $Z$ $p_T$ 8 TeV | [63] | ✓ | ✓ | ✗ | ✓ | ✓ |
| ATLAS $W + c$ 7 TeV | [83] | ✗ | ✓ | ✗ | (✓) | ✗ |
| ATLAS $\sigma_{tt}^{\rm tot}$ 7, 8 TeV | [65] | ✓ | ✓ | ✓ | ✗ | ✗ |
| ATLAS $\sigma_{tt}^{\rm tot}$ 7, 8 TeV | [260–265] | ✗ | ✗ | ✓ | ✗ | ✗ |
| ATLAS $\sigma_{tt}^{\rm tot}$ 13 TeV ($\mathcal{L} = 3.2$ fb$^{-1}$) | [66] | ✓ | ✗ | ✓ | ✗ | ✗ |
| ATLAS $\sigma_{tt}^{\rm tot}$ 13 TeV ($\mathcal{L} = 139$ fb$^{-1}$) | [134] | ✗ | ✓ | ✗ | ✗ | ✗ |
| ATLAS $\sigma_{tt}^{\rm tot}$ and $Z$ ratios | [266] | ✗ | ✗ | ✗ | ✗ | (✓) |
| ATLAS $t\bar{t}$ lepton+jets 8 TeV | [67] | ✓ | ✓ | ✗ | ✓ | ✓ |
| ATLAS $t\bar{t}$ dilepton 8 TeV | [89] | ✗ | ✓ | ✗ | ✗ | ✓ |
| ATLAS single-inclusive jets 7 TeV, R=0.6 | [73] | ✓ | (✓) | ✗ | ✓ | ✓ |
| ATLAS single-inclusive jets 8 TeV, R=0.6 | [86] | ✗ | ✓ | ✗ | ✗ | ✗ |
| ATLAS dijets 7 TeV, R=0.6 | [148] | ✗ | ✓ | ✗ | ✗ | ✗ |
| ATLAS direct photon production 8 TeV | [100] | ✗ | (✓) | ✗ | ✗ | ✗ |
| ATLAS direct photon production 13 TeV | [101] | ✗ | ✓ | ✗ | ✗ | ✗ |
| ATLAS single top $R_t$ 7, 8, 13 TeV | [94, 96, 98] | ✗ | ✓ | ✓ | ✗ | ✗ |
| ATLAS single top diff. 7 TeV | [94] | ✗ | ✓ | ✗ | ✗ | ✗ |
| ATLAS single top diff. 8 TeV | [96] | ✗ | ✓ | ✗ | ✗ | ✗ |

| Data set | Ref. | NNPDF3.1 | NNPDF4.0 | ABMP16 | CT18 | MSHT20 |
|---|---|---|---|---|---|---|
| CMS $W$ asym. 7 TeV ($\mathcal{L} = 36$ pb$^{-1}$) | [267] | ✗ | ✗ | ✗ | ✗ | ✓ |
| CMS $Z$ 7 TeV ($\mathcal{L} = 36$ pb$^{-1}$) | [268] | ✗ | ✗ | ✗ | ✗ | ✓ |
| CMS $W$ electron asymmetry 7 TeV | [55] | ✓ | ✓ | ✗ | ✓ | ✓ |
| CMS $W$ muon asymmetry 7 TeV | [56] | ✓ | ✓ | ✓ | ✓ | ✗ |
| CMS Drell-Yan 2D 7 TeV | [57] | ✓ | ✓ | ✗ | (✓) | ✓ |
| CMS Drell-Yan 2D 8 TeV | [269] | (✓) | ✗ | ✗ | ✗ | ✗ |
| CMS $W$ rapidity 8 TeV | [58] | ✓ | ✓ | ✓ | ✓ | ✓ |
| CMS $W, Z$ $p_T$ 8 TeV ($\mathcal{L} = 18.4$ fb$^{-1}$) | [270] | ✗ | ✗ | ✗ | (✓) | ✗ |
| CMS $Z$ $p_T$ 8 TeV | [64] | ✓ | ✓ | ✗ | (✓) | ✗ |
| CMS $W + c$ 7 TeV | [76] | ✓ | ✓ | ✗ | (✓) | ✓ |
| CMS $W + c$ 13 TeV | [84] | ✗ | ✓ | ✗ | ✗ | (✓) |
| CMS single-inclusive jets 2.76 TeV | [75] | ✓ | ✗ | ✗ | ✗ | ✓ |
| CMS single-inclusive jets 7 TeV | [147] | ✓ | (✓) | ✗ | ✓ | ✓ |
| CMS dijets 7 TeV | [74] | ✗ | ✓ | ✗ | ✗ | ✗ |
| CMS single-inclusive jets 8 TeV | [87] | ✗ | ✓ | ✗ | ✓ | ✓ |
| CMS 3D dijets 8 TeV | [149] | ✗ | (✓) | ✗ | ✗ | ✗ |
| CMS $\sigma_{tt}^{\rm tot}$ 5 TeV | [88] | ✗ | ✓ | ✗ | ✗ | ✗ |
| CMS $\sigma_{tt}^{\rm tot}$ 7, 8 TeV | [146] | ✓ | ✓ | ✗ | ✗ | ✗ |
| CMS $\sigma_{tt}^{\rm tot}$ 8 TeV | [271] | ✗ | ✗ | ✗ | ✗ | ✓ |
| CMS $\sigma_{tt}^{\rm tot}$ 5, 7, 8, 13 TeV | [68, 272–280] | ✗ | ✗ | ✓ | ✗ | ✗ |
| CMS $\sigma_{tt}^{\rm tot}$ 13 TeV | [69] | ✓ | ✓ | ✓ | ✗ | ✗ |
| CMS $t\bar{t}$ lepton+jets 8 TeV | [70] | ✓ | ✓ | ✗ | ✗ | ✓ |
| CMS $t\bar{t}$ 2D dilepton 8 TeV | [90] | ✗ | ✓ | ✗ | ✓ | ✓ |
| CMS $t\bar{t}$ lepton+jet 13 TeV | [91] | ✗ | ✓ | ✗ | ✗ | ✗ |
| CMS $t\bar{t}$ dilepton 13 TeV | [92] | ✗ | ✓ | ✗ | ✗ | ✗ |
| CMS single top $\sigma_t + \sigma_{\bar{t}}$ 7 TeV | [95] | ✗ | ✓ | ✓ | ✗ | ✗ |
| CMS single top $R_t$ 8, 13 TeV | [97, 99] | ✗ | ✓ | ✓ | ✗ | ✗ |
| CMS single top 13 TeV | [281, 282] | ✗ | ✗ | ✗ | ✗ | (✓) |

| Data set | Ref. | NNPDF3.1 | NNPDF4.0 | ABMP16 | CT18 | MSHT20 |
|---|---|---|---|---|---|---|
| LHCb $Z$ 7 TeV ($\mathcal{L} = 940$ pb$^{-1}$) | [59] | ✓ | ✓ | ✗ | ✗ | ✓ |
| LHCb $Z \to ee$ 8 TeV ($\mathcal{L} = 2$ fb$^{-1}$) | [61] | ✓ | ✓ | ✓ | ✓ | ✓ |
| LHCb $W$ 7 TeV ($\mathcal{L} = 37$ pb$^{-1}$) | [283] | ✗ | ✗ | ✗ | ✗ | ✓ |
| LHCb $W, Z \to \mu$ 7 TeV | [60] | ✓ | ✓ | ✓ | ✓ | ✓ |
| LHCb $W, Z \to \mu$ 8 TeV | [62] | ✓ | ✓ | ✓ | ✓ | ✓ |
| LHCb $W \to e$ 8 TeV | [80] | ✗ | (✓) | ✗ | ✗ | ✗ |
| LHCb $Z \to \mu\mu, ee$ 13 TeV | [82] | ✗ | ✓ | ✗ | ✗ | ✗ |

| | |
|---|---|
| ✓ | in baseline dataset |
| ✗ | not considered |
| (✓) | impact assessed but excluded from baseline |

# Can New Physics hide inside the proton?

*``How can you be sure you are not reabsorbing BSM physics into your PDFs?"*

***perhaps most frequent question I am asked in talks!***

Assuming the **SM**, the theory calculations that enter a global PDF fit are:

$$\sigma_{\rm LHC}(\boldsymbol{\theta}) \propto \sum_{ij=u,d,g,\ldots} \int_{M^2}^{s} d\hat{s}\ \underbrace{\mathscr{L}_{ij}(\hat{s}, s, \boldsymbol{\theta})}_{\textit{SM PDFs}}\, \widetilde{\sigma}_{{\rm SM},ij}(\hat{s}, \alpha_s(M))$$

However in the case of BSM physics, here parametrised by the **SMEFT,** the correct expression is:

$$\sigma_{\rm LHC}\left(\boldsymbol{c}, \Lambda, \boldsymbol{\theta}\right) \simeq \left( \int_{M^2}^{s} d\hat{s}\ \overbrace{\mathscr{L}_{ij}(\hat{s}, s, \boldsymbol{\theta})}^{\textit{SMEFT PDFs}}\, \widetilde{\sigma}_{{\rm SM},ij}(\hat{s}, \alpha_s(M)) \right) \times \left( 1 + \sum_{m=1}^{N_6} c_m \frac{\kappa_m}{\Lambda^2} + \sum_{m,n=1}^{N_6} c_m c_n \frac{\kappa_{mn}}{\Lambda^4} \right),$$

***PDF parameters***

***SMEFT coefficients***

How different are ``SM PDFs" & ``SMEFT PDFs"? Can we quantify the risk of **fitting away BSM** in PDFs?

# Can New Physics hide inside the proton?

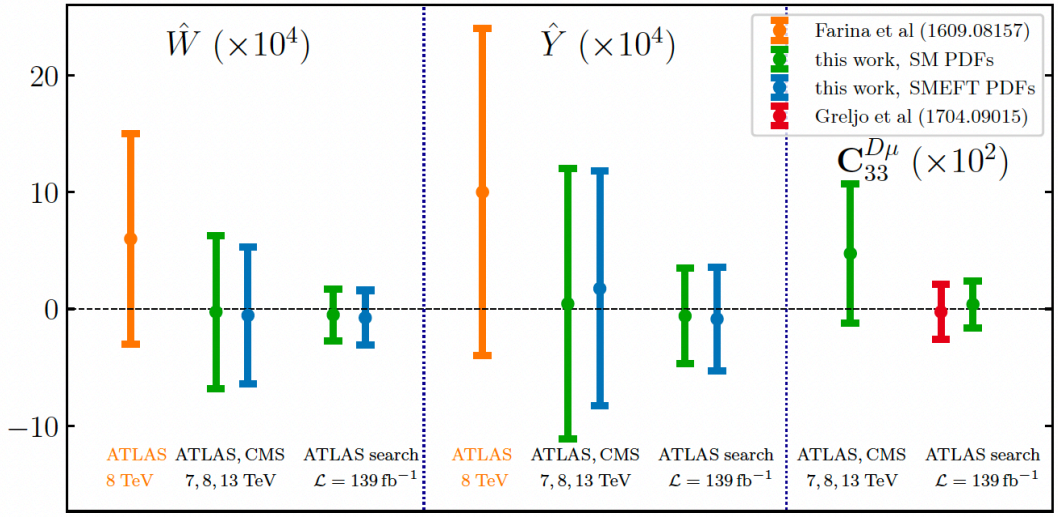
| Exp. | $\sqrt{s}$ (TeV) | Ref. | $\mathcal{L}$ (fb$^{-1}$) | Channel | 1D/2D | $n_{\rm dat}$ | $m_{\ell\ell}^{\rm max}$ (TeV) |
|---|---|---|---|---|---|---|---|
| ATLAS | 7 | [120] | 4.9 | $e^-e^+$ | 1D | 13 | [1.0, 1.5] |
| ATLAS (*) | 8 | [86] | 20.3 | $\ell^-\ell^+$ | 2D | 46 | [0.5, 1.5] |
| CMS | 7 | [121] | 9.3 | $\mu^-\mu^+$ | 2D | 127 | [0.2, 1.5] |
| CMS (*) | 8 | [87] | 19.7 | $\ell^-\ell^+$ | 1D | 41 | [1.5, 2.0] |
| CMS (*) | 13 | [122] | 5.1 | $e^-e^+$, $\mu^-\mu^+$<br>$\ell^-\ell^+$ | 1D | 43, 43<br>43 | [1.5, 3.0] |
| **Total** | | | | | | **270 (313)** | |

Extract PDFs from global fit where **high-mass DY cross-sections** account for EFT effects in two benchmark scenarios

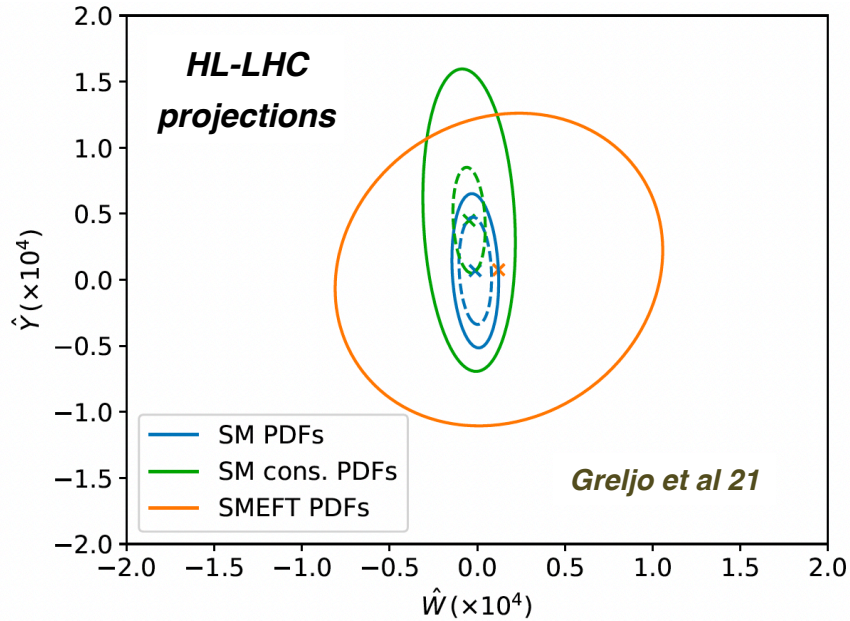
$$d\sigma_{\rm SMEFT} = d\sigma_{\rm SM} \times K_{\rm EFT}$$

$$K_{\rm EFT} = 1 + \sum_{n=1}^{n_{\rm op}} c_n R^{(n)}_{\rm SMEFT} + \sum_{n,m=1}^{n_{\rm op}} c_n c_m R^{(n,m)}_{\rm SMEFT}$$

Available data: limited interplay between PDF and EFT fits, best constraints from **searches**




**HL-LHC**: EFT effects, if present, would be **reabsorbed into PDFs**




**SMEFT PDFs** required to exploit the PDF/BSM interplay in future LHC runs!

# Universal QCD fits

Pushing the **precision frontier** of **QCD fits** requires accounting for **cross-talk** between different **non-perturbative QCD** quantities

| | standard approach | universal QCD analysis |
|---|---|---|
| fragmentation functions | $N(p+p\to\pi)\propto\sum_{i,j,k}\widetilde{\sigma}_{ij\to k}\otimes f_i^{(p)}\otimes f_j^{(p)}\otimes f_{k\to\pi}$ | $N(p+p\to\pi)\propto\sum_{i,j,k}\widetilde{\sigma}_{ij\to k}\otimes f_i^{(p)}\otimes f_j^{(p)}\otimes f_{k\to\pi}$ |
| nuclear PDFs | $N(p+\mathrm{Pb}\to W)\propto\sum_{i,j}\widetilde{\sigma}_{ij\to W}\otimes f_i^{(p)}\otimes f_j^{(A)}$ | $N(p+\mathrm{Pb}\to W)\propto\sum_{i,j}\widetilde{\sigma}_{ij\to W}\otimes f_i^{(p)}\otimes f_j^{(A)}$ |
| proton PDFs | $N(\nu+\mathrm{Pb}\to\mu^+\mu^-)\propto\sum_{i}\widetilde{\sigma}_{i\to\mu^+\mu^-}\otimes f_i^{(p)}\times\left(\frac{f_i^{(A)}}{f_i^{(p)}}\right)$ | $N(\nu+\mathrm{Pb}\to\mu^+\mu^-)\propto\sum_{i}\widetilde{\sigma}_{i\to\mu^+\mu^-}\otimes f_i^{(p)}\times\left(\frac{f_i^{(A)}}{f_i^{(p)}}\right)$ |

theory fitted fixed

theory fitted

***e.g.* proton PDFs fit data on heavy nuclear targets, while nuclear PDF fits reduce to proton baseline for A=1**

Ideally we'd like to combine unpolarised and polarised proton PDF determinations with those of nuclear PDFs and fragmentation functions into an **integrated, universal QCD fit**

***key to fully exploit EIC potential!***

***e.g. ongoing efforts from JAM group***

# Summary and outlook

- Recent progress in unpolarised nucleon structure has led to a **deeper understanding of its content**, from strangeness and photons to intrinsic charm and antimatter asymmetry

- This progress benefits from a **feedback look with LHC analyses** (tests/predictions/ constraints) anticipating how the **Electron Ion Collider will scrutinise the nucleon**

- Still plenty of work (and room for surprises): PDFs with theory errors, interplay between PDF and BSM searches, integrated/universal QCD analyses ….


***``The simple hydrogen atom nucleus appears to be surprisingly charming"***

# Extra Material