



The path to proton structure at 1% accuracy

Juan Rojo, VU Amsterdam & Nikhef



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**!

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

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

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

a point particle



short distances / high energies

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

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**!



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

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**!



fundamental open questions about the proton



These (and many more) fascinating questions to be discussed during this workshop!

The Nucleon in the Spotlight

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 @ SeaQuest





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



knowledge of quark and gluon substructure of protons also essential for



New elementary particles beyond the Standard Model?

Origins and properties of **cosmic neutrinos**?





Nature of Quark-Gluon Plasma in heavy-ion collisions?

PDFs for precision LHC physics

PDF uncertainties are limiting factor in theoretical interpretation for many LHC analysis





PDFs for precision LHC physics

PDF uncertainties are limiting factor in theoretical interpretation for many LHC analysis





Novel experimental constraints

Collider Drell-Yan

Dijet production



Novel experimental constraints



Global fits benefit from redundancy: a given PDF combination is constrained by many processes

Improved theory



Improved theory



Certainly NLO, but also likely NNLO PDFs, underestimate uncertainties without MHOUs State-of-the-art LHC pheno demands both **NNLO PDFs with MHOUs** and **N3LO PDFs**: WIP!

Methodological developments



NNPDF4.0

The path to NNPDF4.0

Collaborative progress towards extending data, theory and methodology

06/2017	NNPDF3.1	[EPJ C77 (2017) 663]
10/2017	NNPDF3.1sx : PDFs with small- x resummation	[EPJ C78 (2018) 321]
12/2017	NNPDF3.1luxQED: consistent photon PDF à la luxQED	SciPost Phys. 5 (2018) 008
02/2018	NNPDF3.1+ATLASphoton: inclusion of direct photon data	[EPJ C78 (2018) 470]
12/2018	NNPDF3.1alphas: α_s from a correlated-replica method	[EPJ C78 (2018) 408]
12/2018	NNPDF3.1nuc: heavy ion nuclear uncertainties in a fit	[EPJ C79 (2019) 282]
05/2019	NNPDF3.1th: missing higher-order uncertainties in a fit	[EPJ C79 (2019) 838; ibid. 931]
07/2019	Gradient descent and hyperoptimisation in PDF fits	[EPJ C79 (2019) 676]
12/2019	NNPDF3.1singletop: inclusion of single top <i>t</i> -channel data	[JHEP 05 (2020) 067]
05/2020	NNPDF3.1dijets: comparative study of single- and di-jets	[EPJ C80 (2020) 797]
06/2020	Positivity of $\overline{\mathrm{MS}}$ PDFs	[JHEP 11 (2020) 129]
08/2020	PineAPPL: fast evaluation of EW×QCD corrections	[JHEP 12 (2020) 108]
08/2020	NNPDF3.1strangeness: assessment of strange-sensitive data	[EPJ C80 (2020) 1168]
11/2020	NNPDF3.1deu: deuteron uncertainties in a fit	[EPJ C81 (2021) 37]
03/2021	Future tests	[arXiv:2103.08606]
2021	NNPDF4.0	September 2021

The NNPDF4.0 dataset



 $\mathcal{O}(50)$ data sets investigated; $\mathcal{O}(400)$ data points more in NNPDF4.0 than in NNPDF3.1

From NNPDF1.0 to NNPDF4.0

Tevatron



A ML open-source QCD fitting framework



The full **NNPDF machine learning fitting framework** has been publicly released open source, together with extensive documentation and user-friendly examples

A ML open-source QCD fitting framework

* The NNPDF collaboration

Search docs

Getting started

Buildmaster

Theory

Servers

Tutorials

External codes

Fitting code: n3fit

Code for data: validphys Handling experimental data:

Storage of data and theory predictions

Continuous integration and deployment

Adding to the Documentation

View page source

The NNPDF collaboration

The NNPDF collaboration performs research in the field of high-energy physics. The NNPDF collaboration determines the structure of the proton using contemporary methods of artificial intelligence. A precise knowledge of the so-called **Parton Distribution Functions** (**PDFs**) of the proton, which describe their structure in terms of their quark and gluon constituents, is a crucial ingredient of the physics program of the Large Hadron Collider of CERN.

The NNPDF code

The scientific output of the collaboration is freely available to the publi through the arXiv, journal repositories, and software repositories. Along with this online documentation, we release the NNPDF code used to produce the latest family of PDFs from NNPDF, NNPDF4.0. The code is made available as an open-source package together with the user-friendly examples and an extensive documentation presented here.

The code can be used to produce the ingredients needed for PDF fits, to run the fits themselves, and to analyse the results. This is the first framework used to produce a global PDF fit made publicly available, enabling for a detailed external validation and reproducibility of the NNPDF4.0 analysis. Moreover, the code enables the user to explore a number of phenomenological applications, such as the assessment of the impact of new experimental data on PDFs, the effect of changes in theory settings on the resulting PDFs and a fast quantitative comparison between theoretical predictions and experimental data over a broad range of observables.

If you are a new user head along to Getting started and check out the Tutorials.

Opportunities for many studies within the ``**Nucleon Structure'' community**: looking forward to suggestions and starting new collaborations!

Improved fitting methodology

Stochastic Gradient Descent via TensorFlow for NN training

Automated model hyperparameter optimisation: NN architecture, minimiser, learning rates ...

Validation with future tests (forecasting new datasets) and closure tests (data based on known PDFs)

 $\begin{array}{l} \text{Loss (``average'')}\\ \text{ML model}\\ \text{hyperparams} \ \boldsymbol{\hat{\theta}} = \operatorname*{arg min}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left(\frac{1}{n_{\mathrm{fold}}} \sum_{k=1}^{n_{\mathrm{fold}}} \chi_k^2(\boldsymbol{\theta}) \right) \end{array}$

Loss (``max")

$$L = \max\left(\chi_1^2, \chi_2^2, \chi_3^2, \dots, \chi_{n_{ ext{fold}}}^2
ight)$$





Stability wrt hyperopt loss function

Improved fitting methodology

epoch 3



Illustrating the outcome of SGD minimisation (band: standard deviation over the MC replicas)

Parametrisation basis independence



$$xT_{3}(x, Q_{0}) \propto NN_{T_{3}}(x)$$
 ultimate test of parametrisation independence
lavour basis PDF parametrisation:
$$xV(x, Q_{0}) \propto \left(NN_{u}(x) - NN_{\bar{u}}(x) + NN_{d}(x) - NN_{\bar{d}}(x) + NN_{s}(x) - NN_{\bar{s}}(x)\right)$$

Radically different strategies to parametrize the quark

PDF flavour combinations lead to identical results:

flavour basis PDF para

 $xV(x, Q_0) \propto NN_V(x)$

 $xV(x, Q_0) \propto$ $xT_3(x, Q_0) \propto \left(NN_u(x) + NN_{\bar{u}}(x) - NN_d(x) - NN_{\bar{d}}(x) \right)$

Comparison with NNPDF3.1



- Good agreement within uncertainties, with NNPDF4.0 being more precise
- Differences can be traced back to the impact of specific datasets (e.g. dijets for large-x gluon) or improvements in theory calculations (e.g. NNLO corrections in dimuon DIS for strangeness)

Antimatter asymmetry



Mark SeaQuest measurement claims evidence for quark sea (``proton antimatter'') asymmetry

$$\frac{\sigma_{\rm DY,deuterium}}{\sigma_{\rm DY,hydrogen}} \approx 1 + \frac{\bar{d}_p(x_t)}{\bar{u}_p(x_t)} \qquad \text{with many caveats!}$$

Actually, SeaQuest further confirms the global fit prediction, which agrees with it even when not included

Already well described by NNPDF3.1 within uncertainties

Intrinsic charm



Increasing evidence for non-perturbative charm component within the proton, robust upon conversion to the 3FNS via backwards evolution and matching conditions (WIP)

Mathematical States and States a

✓ As opposed to previous studies, impact of the EMC charm measurements mild now. Information provided by EMC F₂^c consistent with latest collider data

Intrinsic charm



The strangest proton



- ☑ NOMAD dimuon DIS data sensitive to strangeness via charged-current scattering
- Fitting NOMAD had large impact on the strangeness in NNPDF3.1, now in NNPDF4.0 the no-NOMAD fit is already spot on the data

Excellent consistency of global dataset



Global PDF fits & LHC phenomenology

Comparison between global fits

reasonable agreement with CT18, and MSHT20, different pattern of PDF uncertainties



Comparison between global fits



LHC phenomenology

extensive comparisons between global PDF fits for inclusive and differential LHC cross-sections



Summary and outlook

The global NNPDF4.0 fit achieves high accuracy in an unprecedentedly broad kinematic range, thanks so its extensive dataset combined with deep-learning optimisation models

Its faithfulness in representing PDF uncertainties is completely validated by closure tests, future tests, and parametrisation basis independence

In addition to implications for LHC precision physics, NNPDF4.0 sheds light on aspects of proton structure from light antiquark asymmetries to strangeness and intrinsic charm

The current level of PDF uncertainties challenges the accuracy of theoretical predictions and demand an increased effort towards the systematic inclusion in the fit of theoretical uncertainties (nuclear, higher orders, SM parameters, ...) and higher-order QCD (including N3LO) and EW corrections