

The Partonic Content of Nucleons and Nuclei

Juan Rojo

*Department of Physics and Astronomy, VU University, NL-1081 HV Amsterdam, and
Nikhef Theory Group, Science Park 105, 1098 XG Amsterdam, The Netherlands*

Abstract

Deepening our knowledge of the partonic content of nucleons and nuclei represents a central endeavour of modern high-energy and nuclear physics, with ramifications in related disciplines such as astroparticle physics. There are two main scientific drivers motivating these investigations of the partonic structure of hadrons. On the one hand, addressing fundamental open issues in our understanding in the strong interactions such as the origin of the nucleon mass, spin, and transverse structure; the presence of heavy quarks in the nucleon wave function; and the possible onset of novel gluon-dominated dynamical regimes. On the other hand, pinning down with the highest possible precision the substructure of nucleons and nuclei is a central component for theoretical predictions in a wide range of experiments, from proton and heavy ion collisions at the Large Hadron Collider to ultra-high energy neutrino interactions at neutrino telescopes. In this Article, I present a succinct non-technical overview of our modern understanding of the quark, gluon, and photon substructure of nucleons and nuclei, focusing on recent trends and results and discussing future perspectives for the field.

Invited Review Article to appear in the Oxford Research Encyclopedia of Physics

Contents

1	Summary	2
2	Introduction	3
3	From the quark model to the Higgs boson	4
4	The partonic structure of hadrons in the LHC era	11
5	Summary and outlook	18

1 Summary

Protons and neutrons, collectively known as nucleons, represent together with electrons the fundamental building blocks of matter and dominate the overall mass budget of the visible Universe. Nucleons, as well as all other hadrons, are characterised by a rich internal substructure, being composed by elementary particles, quarks and gluons, collectively known as *partons*. These partons are tightly held together within nucleons thanks to the properties of the quantum field theory of the strong interactions: Quantum Chromodynamics (QCD).

The study of the partonic content of nucleons is one of the central endeavors of modern high-energy and nuclear physics. It is motivated by a number of fundamental open questions in our understanding of the strong interactions such as the origin of the nucleon mass and spin, the three-dimensional profiling of hadron substructure, the role of heavy quarks in hadronic wave functions, and the potential onset of novel gluon-dominated dynamical regimes. In addition, deepening our knowledge of this partonic content of nucleons is of paramount importance for a wide range of theoretical predictions in high-energy process such as proton collisions at the Large Hadron Collider (LHC), heavy ion collisions at the Relativistic Heavy Ion Collider (RHIC), and high-energy neutrino interactions such as neutrino telescopes. Furthermore, the quark and gluon substructure of nucleons is modified once the latter become bound within heavy nuclei, leading to a remarkable pattern of nuclear modification effects whose study opens a novel window to the inner workings of the strong force in the nuclear environment.

The detailed investigation of the partonic content of nucleons and nuclei is however a challenging task. It requires the careful combination of state-of-the-art theoretical calculations and the widest possible range of experimental measurements by means of a statistically robust and efficient fitting methodology. This framework, the so-called *global QCD analysis* of the nucleon structure, has experienced remarkable progress in the recent years. Some important milestones include the assessment of the constraints on the proton structure provided by LHC measurements; the implementation of machine learning algorithms that speed up dramatically the analysis while minimising procedural biases; the exploitation of precision-frontier calculations in both the strong and electroweak interactions; the deployment of tailored advanced computational tools such as fast interfaces; the first-principles calculation of the photon content of the proton; an improved characterisation of the role of the initial state of heavy ion collisions for Quark-Gluon Plasma studies; and the development of novel methods to estimate and propagate relevant sources of uncertainties to the final theory predictions.

Despite all these achievements, a long road still lies ahead with pressing open questions, both of theoretical and of experimental nature, directly relevant for the full physics exploitation of current and future facilities. Some of these include puzzling results concerning the strange quark content of protons; how gluons behave at very small and very large momentum fractions; the impact of lattice QCD simulations; the specific pattern of modifications in the partonic structure of heavy nuclei induced by nuclear effects; and the interplay between Standard Model measurements and searches for New Physics at the high-energy frontier.

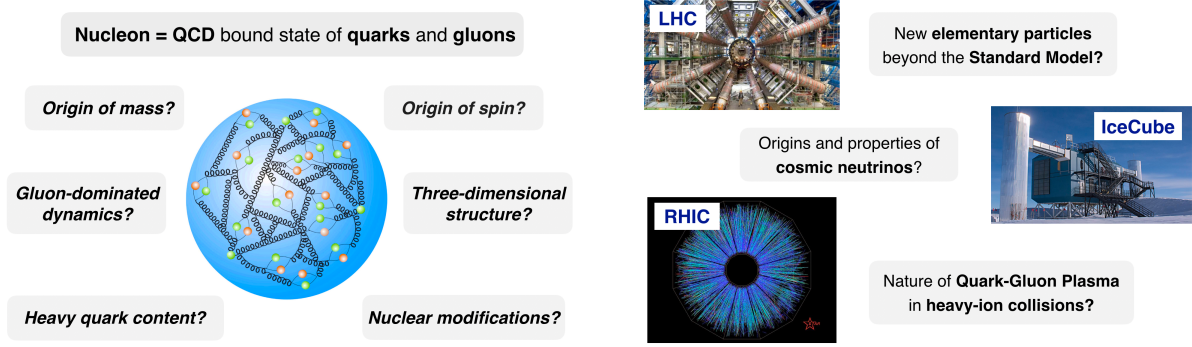


Figure 1. The investigation of the partonic content of nucleons and nuclei has a two-fold motivation: to address fundamental open issues in the strong interactions (left) and to provide precise theoretical predictions for experiments such as the LHC, IceCube, or RHIC (right).

2 Introduction

Elementary particles such as leptons (electrons and neutrinos) do not have any known substructure. As opposed to those, hadrons (particles that experience the strong nuclear force) turn out not to be elementary but rather bound states composed by quarks and gluons, collectively known as partons. These partons are tightly held together within hadrons by virtue of the properties of the quantum theory of strong interaction, Quantum Chromodynamics (QCD). Indeed, the mathematical structure of QCD implies that color-charged particles such as quarks (color being the analog of the electric charge in the strong interactions) cannot exist in isolation and need to be confined within hadrons.

In addition to the light quarks (up and down) and gluons, hadrons contain also heavier quarks (strange and charm quarks in particular) as well as a photon component. Furthermore, the behaviour of partons is modified once protons and neutrons (denoted as nucleons) become themselves the building blocks of heavy nuclei, reflecting a rich pattern of nuclear dynamics. Several of the most important properties related to the partonic constituents of hadrons, including their longitudinal and transverse momentum and spin distributions, are determined by the poorly understood non-perturbative regime of the strong force. In this regime, first principle calculations are challenging and most partonic properties need to be extracted from experimental data by means of a global QCD analysis.

Understanding the partonic content of hadrons and heavy nuclei plays a crucial role in modern particle, nuclear, and astroparticle physics from a two-fold perspective, summarised in Fig. 1. On the one hand, these partonic properties offer a unique window to address fundamental questions about the inner workings of the strong interaction and hadron structure, such as what is the origin of the nucleon mass and spin, how is the motion of quarks and gluons modified inside heavy nuclei, or what determines the onset of new states of matter where gluons dominate. On the other hand, a precise quantification of partonic substructure of nucleons is an essential input for any theoretical predictions for a variety of experiments, from proton-proton scattering at the high-energy frontier at the Large Hadron Collider (LHC) to the collisions between heavy ions and to the interpretation of the data provided by neutrino and cosmic ray telescopes. In this respect, the determination of partonic content of hadrons and nuclei offers a bridge between different areas of modern physics and of related disciplines such as advanced statistics, data analysis, and machine learning.

In this Article I present a succinct introduction to our modern understanding of the quark, gluon, and photon content of nucleons and nuclei, focusing on recent results and trends. I also outline and highlights possible future perspectives for the field of proton structure studies. The Article is deliberately non-technical, and for a detailed discussions on the various topics covered here and an extensive survey of the related literature the reader is encouraged to consult recent topical reviews [Forte and Watt, 2013, Rojo et al., 2015, Gao et al., 2018, Kovařík et al., 2019].

3 From the quark model to the Higgs boson

Pushing forward our understanding of the partonic structure of nucleons and nuclei has been at the forefront of fundamental research in particle and nuclear physics for more than five decades. The story of quarks and gluons starts around the early 1930s. By then, both the proton and neutron had been discovered, and thus the structure of atomic nuclei could be explained. At that time protons and neutrons were assumed to be point-like particles without further substructure, much in the same way as electrons. For a long time, protons and neutrons were the only known *hadrons*, that is, particles that interacted via the strong nuclear force, responsible in particular for keeping the atomic nucleus bound together.

The situation changed dramatically in the late 1940s with the discovery of neutral and charged pions in cosmic ray experiments - those were the main toolbox of particle physicists before accelerators became powerful enough. The discovery of the pions was followed by that of a plethora of other strongly interacting particles: kaons, rhos, lambdas, and omegas, just to name a few. Each of these new hadrons was characterised by different masses, electric charges, and spins (the intrinsic angular momenta of quantum particles). Physicists were confused and asked themselves how it was possible to establish some order in this chaos. In other words, what were the underlying laws of Nature that determined the properties of the large number of hadrons observed? Furthermore, these new particles did not seem to play any obvious role in the properties of everyday matter, with the possible exception of the pion that was thought to mediate the strong interactions.

The first breakthrough towards clarifying this confusing situation took place in the 1960s, when Gell-Mann and Zweig separately realised [Gell-Mann, 1964, Zweig, 1964] that the observed regularities in the hadron spectrum could be explained by assuming that they were not fundamental particles, but instead bound states composed by new hypothetical particles named *quarks*. These newly proposed elementary particles were characterised by being point-like, having a fractional charge of either $\pm 2/3$ or $\pm 1/3$ in units of the electron charge, and half-integer spin. Furthermore, this *quark model* assumed that quarks existed in three different types or “flavours”: the up, down, and strange quarks. It was then possible to show that by combining these three quarks in various ways one could reproduce the quantum numbers of most observed *baryons* (half-integer spin hadrons, assumed to be composed by three quarks) and *mesons* (integer-spin hadrons, composed by a quark-antiquark pair).

However, this quark model was purely phenomenological, and did not provide in particular a suitable mechanism to explain why quarks were tightly bound together within the hadrons. Furthermore, while the quark model was successful to describe hadron spectroscopy, it was unable to provide predictions for the observed behaviour of strongly interacting particles in high-energy scattering experiments. For these reasons, physicists were for some time sceptical of the concept of quarks: while certainly a useful mathematical framework to describe hadron structure, their actual existence as real elementary particles was far from widely accepted.

The road towards the acceptance of quarks as *bona fide* elementary particles was paved by two momentous discoveries that took place in the late 1960s and early 1970s. The first of those arose in a series of experiments at the Stanford Linear Accelerator Center (SLAC), where energetic electrons were used as projectiles and fired at protons and neutrons in atomic nuclei. By investigating the properties of the deflected electrons, physicists could probe for the first time the possible substructure of protons, in the same way as how Rutherford’s experiments at the beginning of the century had established the existence of the atomic nucleus.

The SLAC measurements of this process [Bloom et al., 1969], known as deep-inelastic scattering (DIS), were consistent with a model for the proton composed by point-like, non-interacting constituents of fractional electric charge, thereby providing the long-sought experimental evidence for the existence of quarks. Since those groundbreaking experiments until today, lepton-hadron DIS experiments such as the SLAC ones have provided the backbone for our investigations of the partonic structure of nuclei. However, from the theoretical point of view there were still important hurdles to be overcome before the quark model could be adopted universally. In particular, an explanation was needed for the fact that quarks appeared to be free within the nucleon, while strong interaction processes were characterised in general by higher rates than the electromagnetic and weak ones. The latter properties implied the presence of a large

coupling that determined the strength of the interaction, seemingly inconsistent with the SLAC measurements. How it was possible that the same interaction appeared to be either strong or weak depending on the specific process?

The second revolution that led to the modern understanding of the quark substructure of hadrons was provided by the realisation in 1973 by Gross, Politzer, and Wilczek that the strong nuclear force could be described in the framework of a renormalizable quantum field theory [Gross and Wilczek, 1973, Politzer, 1973] in the same way as the electromagnetic force was described by Quantum Electrodynamics (QED). This new theory, called Quantum Chromodynamics (QCD), was formulated in terms of quark fields interacting among them via the exchange of force mediators called *gluons* that transmitted color charges, therefore playing an analogous role as the photon in electromagnetism. A crucial prediction of this new theory was that the strength of the interaction decreased for small distances, or alternatively for high energies, thus explaining why in high-energy events quarks appeared to be essentially free particles while at low energies, where the coupling strength increases, quarks were bound within the hadrons.

Other predictions of QCD were subsequently confirmed, in particular evidence for the existence of gluons was obtained in electron-positron collisions in the late 1970s. In these events, the gluon appeared as a third stream of collimated hadrons (known as a jet) in addition to the two associated with a quark-antiquark pair [Brandelik et al., 1979]. In the following years, the new theory of the strong interactions was also extended with additional heavier quarks, up to a total of 6 different flavours, with the discovery of the charm (1972), bottom (1977), and top (1995) quarks. The charm and bottom quarks have masses around 1.5 and 5 times the proton mass respectively, and play an important role in the structure of heavy mesons and baryons. The top quark, of the other hand, is too heavy (175 times the proton mass) and decays before it can form bound states with other quarks.

The combination of the predictive power of the quark model, the formulation of Quantum Chromodynamics as the quantum field theory of the strong nuclear force, and the results of high-energy experiments such as SLAC’s deep-inelastic scattering confirmed beyond reasonable doubt that hadrons were bound states composed by quarks and gluons, tightly held together by the non-perturbative phenomena that dominate the strong nuclear force at low energies. It became customary to collectively denote quarks and gluons, as well as any eventual further component of hadrons, as *partons*. It was early on realised that the investigation of the partonic structure of nucleons¹ was going to be a challenging endeavor. Indeed, crucial properties of quarks and gluons, such as their contribution to the total momentum and spin of the parent proton, are determined by low-energy non-perturbative QCD dynamics. This non-perturbative character implied that first principle calculations of the partonic structure of nucleons could not be carried out within the framework of perturbation theory, where the Feynman diagrammatic expansion had led to striking successes in the case of weakly coupled theories such as Quantum Electrodynamics.

Already during the dawn of the studies of the partonic structure of hadrons it became clear that in order to make progress two main strategies could be pursued. The first one is based on a brute-force approach, where QCD is discretised on a space-time lattice allowing different non-perturbative quantities to be evaluated by means of powerful computer simulations. Until recently, the investigation of partonic properties using lattice QCD was however rather limited due to a number of both conceptual and computational bottlenecks. This approach has experienced significant progress in the recent years, as reviewed in the 2017 community White Paper [Lin et al., 2018], and is able now to provide valuable information on the properties of quark and gluons within hadrons.

The second strategy exploits a central property of QCD known as *factorisation*, whereby the total interaction cross-sections for processes involving hadrons in either the initial or the final state of the collision (or both) can be separated into two independent contributions. The first one is the short-distance hard-scattering partonic cross-section, calculable using Feynman diagrams in the same way as in QED. The second contribution encodes the information on the proton structure determined by long-distance, non-perturbative dynamics into objects called

¹The same considerations hold for any of other hadrons. In this Article I will concentrate on the partonic structure of protons, for which there is far more experimental information available as compared to other hadrons.

parton distribution functions (PDFs). Then experimental measurements of such cross-sections can be used, by virtue of their factorisable properties, to extract the PDFs and achieve this way powerful insight about the partonic content and properties of protons.

This latter approach is usually known as the *global QCD analysis*, and it has been successfully deployed in the last three decades to understand in ever-increasing detail the partonic structure of nucleons and nuclei, as will be explained in the remaining of this Article. There exist various different types of PDFs, for example one can consider either spin-dependent or spin-independent PDFs, collinear-integrated, or transverse momentum dependent PDFs. In this Article I will be focusing on the collinear unpolarised PDFs of nucleons and nuclei, which are in the majority of the cases the relevant quantities for the interpretation of the results of high-energy experiments such as the proton-proton collisions at the Large Hadron Collider (LHC). The interested reader can find further information about polarised PDFs in [Aidala et al., 2013, Nocera et al., 2014] and about transverse-momentum dependent PDFs in [Angeles-Martinez et al., 2015] and references therein.

In the following I will adopt the following notation for the collinear unpolarised parton distribution functions of the proton:

$$\{f_i(x, Q^2)\}, \quad i = 1, \dots, n_f, \quad (1)$$

where x stands for the fraction of the nucleon's longitudinal momentum carried by the i -th parton (often denoted as the Bjorken variable) and Q denotes the energy scale (which also corresponds to the inverse of the resolution length) at which the nucleon is being probed. In Eq. (1), the index i indicates the parton flavour, and n_f denotes the number of active partons at the scale Q . A quark is typically considered a massless parton when the momentum transfers involved are larger than its mass, $Q \gtrsim m_q$; else it is assumed to be a massive state that does not contribute to the partonic substructure of the proton. For instance, at $Q = 10$ GeV one would have $n_f = 11$, given that there are five quark PDFs (up, down, strange, charm, and bottom) and the corresponding antiquark counterparts, supplemented by the gluon PDF.

At leading order (LO) in the QCD perturbative expansion in the strong coupling α_s , the PDFs admit an interpretation as probability densities. This property implies that $g(x, Q = 10 \text{ GeV}) dx$ would correspond to the number of gluons in the proton at a scale of $Q = 10$ GeV that carry a momentum fraction between x and $x + dx$, and likewise for other quark flavour combinations. However, it is important to emphasize that this naive probabilistic picture is lost when higher order effects in QCD are taken into account. In particular, it can be shown that PDFs become dependent on the specific renormalisation scheme adopted, and thus cannot be associated to probabilities nor to any specific physical observable.

As mentioned above, the dependence of the PDFs on the momentum fraction x is determined by low-energy (long-distances) non-perturbative QCD dynamics, and therefore needs to be extracted from experimental data as follows. Consider the processes depicted in the left panel of Fig. 2, where a charged lepton such as a muon scatters at high energy off a proton. In such process, which is of the same family of deep-inelastic scattering measurements as those used in the pioneering SLAC experiments, the muon emits a virtual gauge boson (either a photon or a Z boson) which then interacts with one of the quarks in the proton, in the example here an up quark. The cross-section for this processes, which is proportional to the number of events of this type that will take place in a given period of time, can be schematically expressed by

$$\sigma_{\mu p \rightarrow \mu X}(Q) = \tilde{\sigma}_{\gamma^* u \rightarrow u}(Q) \otimes f_u(x, Q), \quad (2)$$

where Q is related to the momentum transfer between the muon and the proton and $\tilde{\sigma}_{\gamma^* u \rightarrow u}$ is the hard partonic cross-section for photon-quark scattering that can be computed in perturbation theory using Feynman diagrams. From Eq. (2) one can observe that if the hadronic cross-section $\sigma_{\mu p \rightarrow \mu X}$ is measured and the partonic one $\tilde{\sigma}_{\gamma^* u \rightarrow u}$ can be evaluated using perturbative QCD calculations, then one can extract the PDF of the up quark $u(x, Q)$ from the data.

At this point it is important to mention that, while the dependence of the PDFs on the momentum fraction x (the Bjorken variable) is indeed non-perturbative, their dependence on the momentum transfer Q can be evaluated in perturbation theory. This can be achieved using

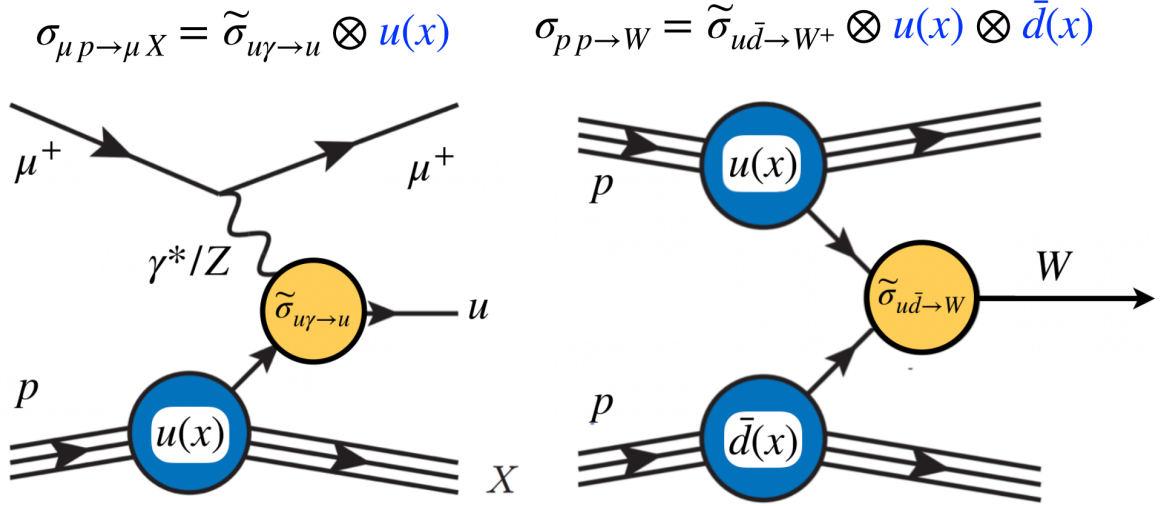


Figure 2. Left: the Feynman diagram associated to the deep-inelastic scattering process, where an energetic muon scatters off one of the quarks in the proton via the exchange of either a virtual photon or a Z boson. Right: the corresponding diagram now for W boson production in proton-proton collisions, known as the Drell-Yan process. By virtue of the QCD factorisation theorems and the PDF universality properties, it is possible to extract the proton PDFs from lepton-proton collisions (left) and then use the same PDFs to predict cross-sections in proton-proton collisions (right).

the DGLAP evolution equations, which take the schematic form

$$Q^2 \frac{\partial}{\partial Q^2} f_i(x, Q^2) = \sum_{j=1}^{n_f} P_{ij}(x, \alpha_s(Q^2)) \otimes f_j(x, Q^2), \quad i = 1, \dots, n_f, \quad (3)$$

where n_f indicates the number of active partons that participate in the evolution, and P_{ij} are perturbative kernels (the splitting functions) currently known up to $\mathcal{O}(\alpha_s^4)$. Thanks to these DGLAP equations, once we have determined the partonic structure of the nucleon at some low scale, say $Q_0 \simeq 1$ GeV, we can evaluate the behaviour of the PDFs for any other scale $Q > Q_0$.

Furthermore, one should also point out that there exist additional theory considerations restrict the shape of specific PDF combinations. In particular, energy conservation imposes the momentum sum rule,

$$\int_0^1 dx \, x \left(\sum_{i=1}^{n_f} (f_{q_i}(x, Q^2) + f_{\bar{q}_i}(x, Q^2)) + f_g(x, Q^2) \right) = 1, \quad (4)$$

which translate the property that the sum of the energies carried by all partons should add up to the total proton energy, while quark flavour number conservation leads to the valence sum rules,

$$\int_0^1 dx \, (u(x, Q^2) - \bar{u}(x, Q^2)) = 2, \quad \int_0^1 dx \, (d(x, Q^2) - \bar{d}(x, Q^2)) = 1, \quad (5)$$

which reflect that the proton contains two valence up quarks and one valence down quark. Note that these sum rules should hold for any value of the energy Q , and indeed one can verify that DGLAP evolution ensures that if they are satisfied at some scale Q_0 , they will also be satisfied for all other scales.

While the discussion so far has been kept somewhat schematic, it should provides an intuitive picture of the inner workings of the global PDF analysis paradigm: by combining a wide range of experimental measurements that involve proton targets with state-of-the-art theory calculations, it becomes possible to pin down the properties of the quark and gluon substructure of the proton. In Fig. 3 we present the results of a recent determination of the proton structure, the NNPDF3.0 global analysis [Ball et al., 2015]. We show the up and down quark valence PDFs, defined as $f_{uV} = f_u - \bar{u}$ and $f_{dV} = f_d - \bar{d}$, the sea quark PDFs $\bar{u}$, $\bar{d}$, f_s , f_c , f_b , as well as the gluon PDF f_g (divided by 10). The PDFs are displayed both at $Q^2 = 10$ GeV² (left) and at $Q^2 = 10^4$ GeV²

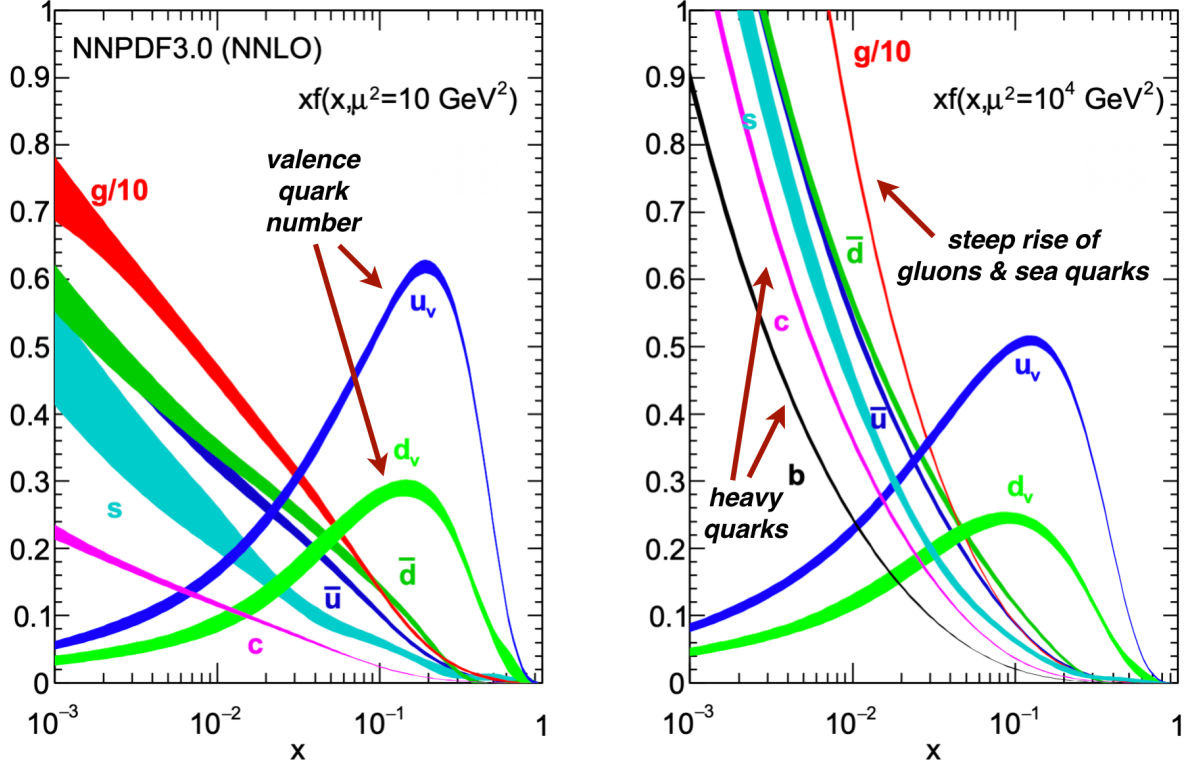


Figure 3. The results of the NNPDF3.0 NNLO global analysis for $Q^2 = 10 \text{ GeV}^2$ (left) and at $Q^2 = 10^4 \text{ GeV}^2$ (right panel) as a function of the partonic momentum fraction x . For each PDF, the bands provide an estimate of the associated uncertainty. The dependence of the PDFs with the scale Q is determined by the perturbative DGLAP evolution equations.

(right panel) as a function of the partonic momentum fraction x . Recall that the dependence of the PDFs with the scale Q is entirely fixed by the perturbative DGLAP evolution equations. For each of the PDFs shown in Fig. 3, the size of the bands provides an estimate of the associated uncertainty.

From Fig. 3 one can observe that the overall shape of the quark valence distributions, f_{u_v} and f_{d_v} , is consistent with the condition that the total number of valence quarks in the proton is two for the up and one for the down quarks. This is a direct consequence of the valence sum rules of Eq. (5). Furthermore, one sees that at large values of the momentum fraction x the valence quark PDFs are the largest, while at smaller x the proton is dominated by its sea quark and gluon components. Indeed, there is steep rise at small- x of the gluons and sea quarks, which quickly dominate over the valence distributions, specially at Q is increased. As we move from $Q^2 = 10 \text{ GeV}^2$ to $Q^2 = 10^4 \text{ GeV}^2$ as determined by the DGLAP evolution equations, one finds that the valence distributions f_{u_v} and f_{d_v} are relatively stable while the gluons and sea quarks rise strongly. This behaviour implies that the higher the energies at which one probes the internal structure of the proton, the larger its gluonic component will be.

The key property that allows the exploitation of the information contained by the PDFs in different experiments is their *universality*. Thanks to the factorisation theorems of the strong interaction, it is possible to determine the PDFs from a given type of processes, such as lepton-proton scattering, and then use the same PDFs to compute predictions for different types of processes, such as proton-proton collisions. Fig. 2 illustrated one important application of this PDF universality: by virtue of the QCD factorisation theorems, one can extract the proton PDFs from deep-inelastic scattering measurements (left) and then use the same PDFs to evaluate the production cross-section for electroweak gauge bosons in proton-proton collisions (right), the so-called Drell-Yan process. It is important to emphasize that this universality property is highly non-trivial: beyond the leading approximation, PDFs need to reabsorb into their definition soft and collinear divergences that appear in higher-order perturbative calculations. Fortunately, it

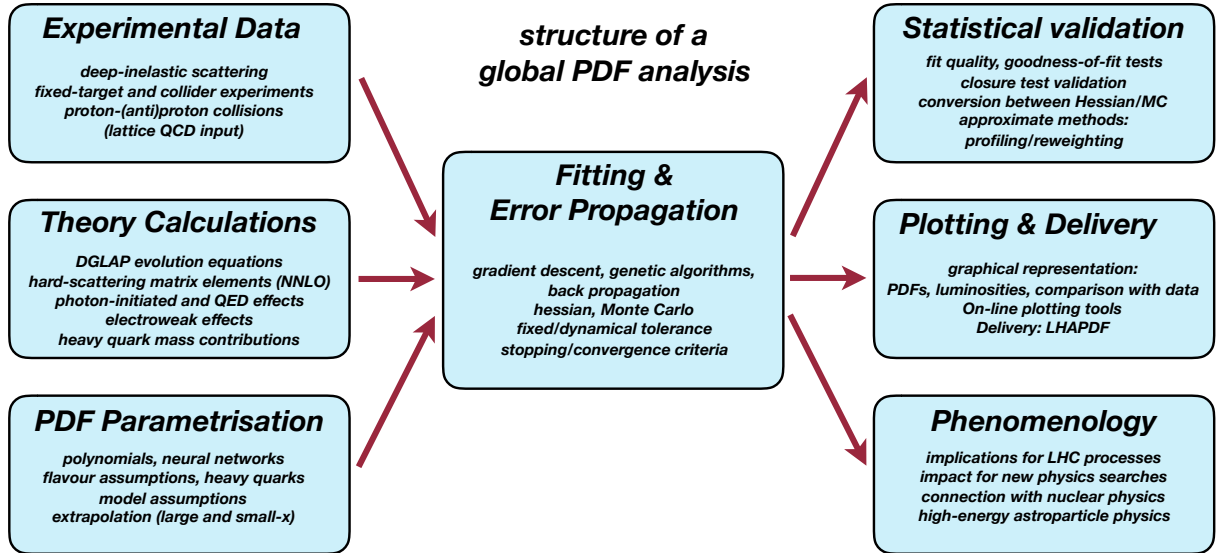


Figure 4. Schematic representation of the general structure of a global PDF analysis. It is based on three main groups of inputs: experimental data, theoretical calculations, and the PDF parametrisation and flavour assumptions. These inputs are processed through a fitting code that returns as output the most likely values of the PDFs and their associated uncertainties. Subsequently, the results of this global PDF analysis are statistical validated, made publicly available, and their phenomenological implications for the LHC and other experiments studied.

can be shown that these divergences are process-independent and thus do not compromise the universality of the PDFs.

The technical complexities that underlie a global PDF analysis are schematically summarised in Fig. 4. It is composed by three main types of inputs: experimental measurements, theoretical calculations, and the methodological assumptions such as those related to the PDF parametrisation and the quark flavour decomposition. This theory input involves higher-order perturbative calculations in both the strong and electroweak interactions for the DGLAP evolution kernels and for the hard-scattering matrix elements. In addition, these higher-order calculations need to be interpolated in the form of fast grids [Carli et al., 2010, Wobisch et al., 2011, Bertone et al., 2014] in order to satisfy the requirements of the CPU-time intensive PDF fits.

These inputs are processed through a fitting code that returns as output the most likely values of the PDFs and their associated uncertainties. Several methods have been proposed to estimate the latter, being the Hessian [Pumplin et al., 2001] and the Monte Carlo [Del Debbio et al., 2007] approaches the two most popular ones. In addition, approximate techniques have been constructed to emulate within certain approximations the results of a full PDF fit in a much faster way, in particular the Bayesian reweighting of Monte Carlo sets [Ball et al., 2012] and the profiling of Hessian sets [Paukkunen and Zurita, 2014]. Subsequently, the results of this global PDF analysis are statistical validated, made publicly available, and their phenomenological implications for the LHC and other experiments studied. It is beyond the scope of this Article to describe each of the components listed in Fig. 4, for further detail we point the reader to [Gao et al., 2018] and references therein.

Global fits of parton distributions, such as the one displayed in Fig. 3, have played a central role in the interpretation of experimental measurements in lepton-hadron and hadron-hadron collisions in the last three decades. As highlighted by Fig. 2, any theoretical predictions for processes at colliders that involve protons in the initial state, such as HERA and the LHC, necessarily require PDFs as input. Parton distributions in particular were one of the theoretical inputs that contributed to the discovery of the Higgs boson in 2012 by the ATLAS and CMS experiments [Aad et al., 2012, Chatrchyan et al., 2012], recognised with the Nobel Prize in Physics in 2013, and which heralded a new era for elementary particle physics. The Higgs boson can be rightly qualified as the most spectacular particle ever encountered. First of all, it is the only known particle which couples to anything that has mass. Second, it transmits a new type of fundamental force, which is completely different to all other interactions such

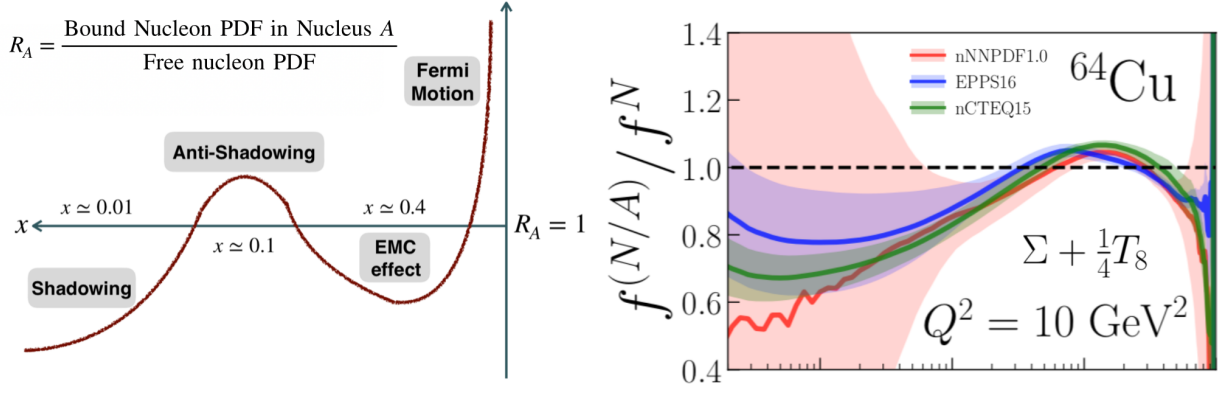


Figure 5. Left: schematic representation of nuclear effects that modify the PDFs of nucleons bound within heavy nuclei (nPDFs) as compared to their free nucleon counterparts. Right: comparison of the nuclear modifications for the $\Sigma + T_8/4$ quark combination in copper ($A = 64$) between three recent fits of nuclear PDFs at $Q^2 = 10 \text{ GeV}^2$, where the bands represent the associated nPDF uncertainties.

as electromagnetism. Thirdly, it is exquisitely sensitive to quantum effects taking place at the tiniest of the distances. Nowadays, improving our understanding of the quark and gluon structure of the proton makes possible scrutinizing the Higgs sector of the Standard Model in greater detail, and indeed several recent developments in the field of PDF fits aim to strengthen the robustness of theoretical predictions for Higgs boson production at the LHC [de Florian et al., 2016].

So far the discussion in this Article has been restricted to the partonic structure of free protons. A closely related field of research is the one which focuses on the study of the partonic structure of nucleons (protons and neutrons) which are bound within heavy nuclei. The quark and gluon structure of bound nucleons can be then parametrised by the so-called nuclear parton distribution functions (nPDFs). It was discovered by the EMC experiment in the 80s [Aubert et al., 1983], to great surprise of the community, that the parton distributions of bound nucleons were significantly different from those of their free-nucleon counterparts. Such behaviour was quite unexpected, since nuclear binding effects are characterised by MeV-scale momentum transfers, which should be negligible when compared to the typical momentum transfers ($Q \gtrsim 1 \text{ GeV}$) involved in deep-inelastic scattering.

In addition to this suppression of the nPDFs at $x \simeq 0.4$ (the EMC effect), other nuclear modifications that have been reported include an enhancement at large- x (Fermi motion) and a further suppression in the small- x region known as nuclear shadowing. In the left panel of Fig. 5 we display a schematic representation of how different nuclear effects modify the PDFs of nucleons bound within heavy nuclei as compared to those of free nucleons as a function of the partonic momentum fraction x . Note that in such comparisons the limit $R_A = 1$ corresponds to the absence of nuclear corrections.

From the conceptual and methodological points of view, global fits of nuclear PDFs proceed in a similar way as those of the free nucleon PDFs and summarised in Fig. 4. This said, there are two important differences as compared to the free proton case. First of all, the available experimental constraints are greatly reduced in the nuclear case, with a restricted kinematical coverage both in x and in Q^2 . Second, one needs to extend the fitting methodology to account also for the dependence on the nuclear mass number A . This implies that nPDFs will depend on three variables, $f_i^{(A)}(x, Q^2, A)$, two of which (x and A) are determined by non-perturbative dynamics and thus need to be parameterised and extracted from data. Several groups have presented fits of nPDFs with varying input datasets, theory calculations, and methodological assumptions. In the right plot of Fig. 5 we display a comparison of the nuclear modifications for the $\Sigma + T_8/4$ quark combination² in copper ($A = 64$) at $Q = 10 \text{ GeV}^2$ between the nNNPDF1.0 [Abdul Khalek et al., 2019d], nCTEQ15 [Kovarík et al., 2016], and EPPS16 [Eskola et al., 2017] nPDF sets,

²Where one has defined $f_\Sigma = \sum_i (f_{q_i} + f_{\bar{q}_i})$ and $f_{T_8} = f_{u_i} + f_{\bar{u}_i} + f_{d_i} + f_{\bar{d}_i} - 2(f_{s_i} + f_{\bar{s}_i})$.

where the bands indicate the associated uncertainties. While there is reasonable agreement in the central values, there are marked differences in the uncertainty estimates, specially in the small- and large- x regions.

In addition to shedding light on the inner workings of the strong force in the nuclear environment, the accurate determination of nuclear PDFs also represents an important ingredient for the interpretation of the results of the heavy ion programs of RHIC and the LHC, where one collides heavy nuclei such as lead ($A = 208$) among them. In such heavy ion collisions, it becomes possible to study the properties of the Quark Gluon Plasma (QGP), the hot and dense medium created first in the early Universe and that now we can replicate in the laboratory. Nuclear PDFs enter the initial state of heavy ion collisions whenever hard probes such as jets, weak bosons, or heavy quarks are produced [Abreu et al., 2008]. Therefore, improving our understanding of the nPDFs is important in order to tell apart the cold from the hot nuclear matter effects in those complex events, involving hundreds or even thousands of produced particles.

Following this general introduction to the topic of the partonic structure of nucleons and nuclei, in the second part of this Article we present an overview of the current state of (n)PDF determinations, emphasizing recent breakthroughs and important results in the field.

4 The partonic structure of hadrons in the LHC era

Global PDF fits. Several collaborations provide regular updates of their global proton PDF analyses, including ABMP [Alekhin et al., 2017], CT [Hou et al., 2019], MMHT [Harland-Lang et al., 2015], and NNPDF [Ball et al., 2017]. The differences among them stem from several of the aspects that define a PDF fit (see Fig. 4). To begin with, the input datasets are not the same, for example ABMP does not include jet production measurements, and not all groups include the same neutrino-included DIS experiments. Secondly, theoretical calculations can differ, for instance, part of the disagreement between ABMP and other groups can be traced back to the treatment of the heavy quark corrections in DIS structure functions. Furthermore, the treatment of the SM inputs in the PDF analysis is not homogeneous: while CT, MMHT, and NNPDF take external parameters such as $\alpha_s(m_Z)$ and m_c from their Particle Data Group averages [Patrignani et al., 2016], ABMP fits them simultaneously with the PDFs.

The third, and perhaps most important, source of differences between the PDF fitting groups are related to methodological assumptions such as the way to parametrise PDFs and to estimate the associated uncertainties. Here there are two main groups: ABMP, CT, and MMHT are based on the Hessian method (with or without tolerances) and a polynomial parametrisation, while NNPDF adopts the Monte Carlo method with artificial neural networks as universal unbiased interpolants. In Fig. 6 we compare the gluon and the down antiquark PDFs at $Q = 100$ GeV between the ABMP16, MMHT14, CT14, and NNPDF3.1 fits, normalised to the central value of the latter. For each fit, the bands represent the 68% confidence level uncertainties. Similar comparison plots (also for related quantities such as PDF luminosities) can be straightforwardly produced by means of the APFEL-Web online PDF plotter [Carrazza et al., 2015a].

Impact of LHC data. Traditionally, hadron colliders were considered the domain of discovery physics, while the cleaner lepton colliders had the monopoly of precision physics. However, the outstanding performance of the LHC and its experiments, complemented by the recent progress in higher order QCD and electroweak calculations, have demonstrated that despite this common lore the LHC is able to pursue a vigorous and exciting precision physics program. In the context of studies of the partonic content of nucleons and nuclei, this precision program aims to exploit the information contained in LHC measurements to pin down the quark and gluon PDFs and to reduce their uncertainties. In turn, this should lead to improved theory predictions for many important processes at the LHC and elsewhere from Higgs production to Dark Matter searches.

Until around 2012, all PDF fits were restricted to deep-inelastic scattering measurements (from both fixed-target experiments and from the HERA collider), fixed-target Drell-Yan cross-sections, and some jet production measurements from the Tevatron. Fortunately the situation has improved dramatically in the recent years. A large number of LHC processes are now staple components of global PDF analyses: inclusive and differential Drell-Yan measurements,

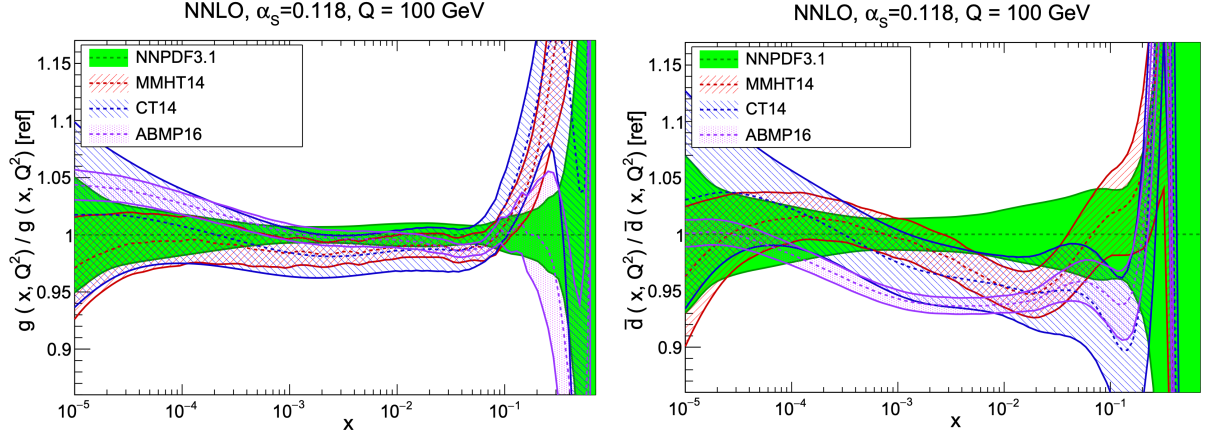


Figure 6. Comparison between the gluon (left) and down antiquark (right plot) PDFs at $Q = 100$ GeV between the ABMP16, MMHT14, CT14, and NNPDF3.1 fits, normalised to the central value of the latter. For each fit, the bands represent the 68% confidence level uncertainties.

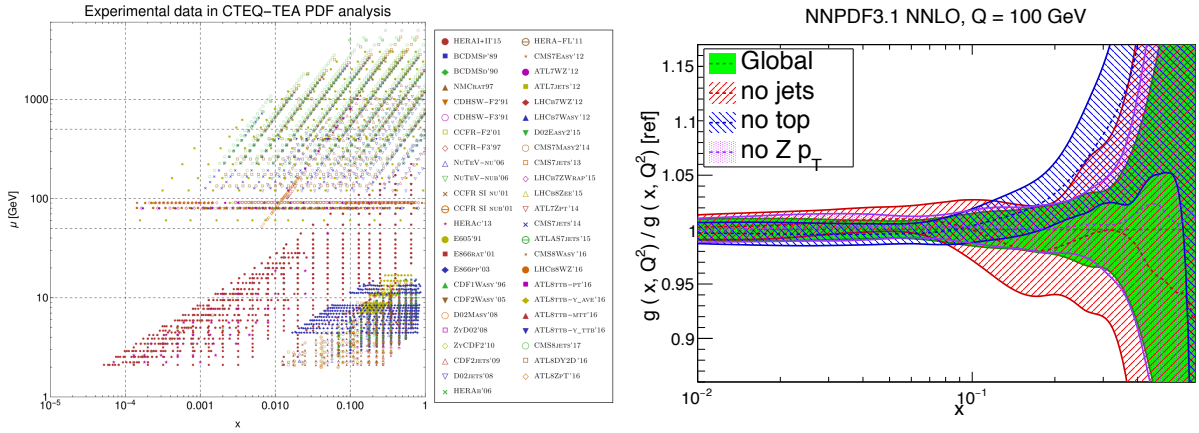


Figure 7. Left: the kinematic coverage in the (x, Q) plane of the experimental data included in a recent CT analysis. Right: comparison of the large- x gluon from the global NNPDF3.1 fit with the corresponding results in fits without the jet, top quark, or the $Z p_T$ measurements included.

the transverse momentum of W and Z bosons, top quark pair and single top production, and heavy meson production, among several others. In most cases, state-of-the-art NNLO QCD calculations are available, making possible carrying out fully consistent NNLO PDF fits based on a the widest possible range of fixed-target and collider measurements.

In order to illustrate the breadth of collider measurements included in a typical modern PDF fit, in the left panel of Fig. 7 we display the kinematic coverage in the (x, Q) plane of the experimental data included in the recent CT analysis of [Hobbs et al., 2019]. One can observe that this coverage extends down to $x \simeq 5 \times 10^{-5}$ and up to several TeV in Q , and furthermore that in most cases several processes constrain the PDFs in the same region of the (x, Q) plane. From the labels in the legend one can also read that around half of the datasets which constitute this fit correspond to LHC measurements. One important advantage of this rich dataset is to introduce elements of redundancy, meaning that a given PDF will typically be constrained by several datasets and resulting in a significantly more robust PDF extraction. This important property is illustrated in the case of the large- x gluon in the right panel of Fig. 7. There we display the results from the global NNPDF3.1 NNLO fit compared with the corresponding results in fits without any jet, top quark, or $Z p_T$ measurements included. One observes that in all cases the resultant gluons are consistent among them within uncertainties, highlighting the complementarity of the various gluon-sensitive measurements provided by the LHC.

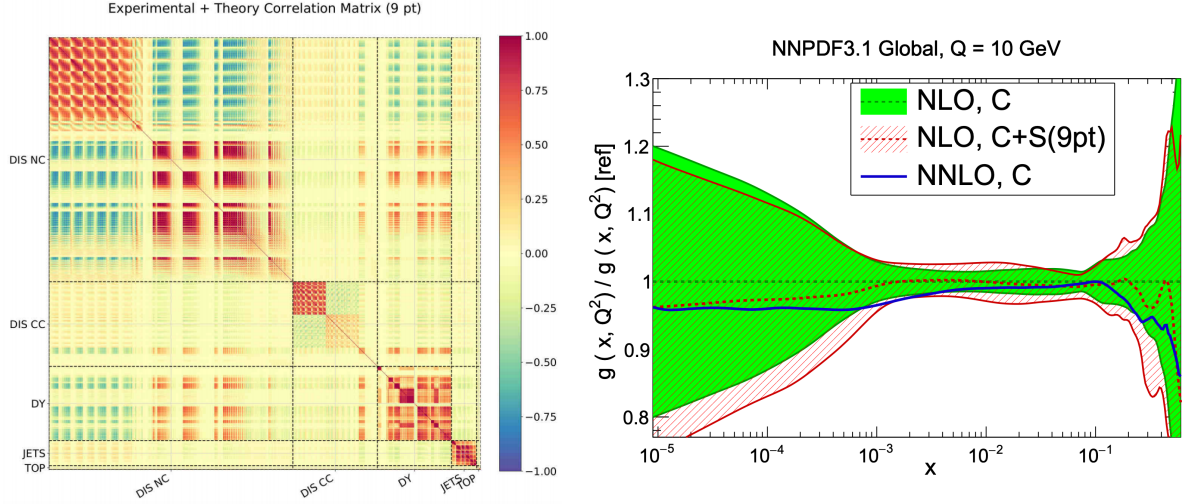


Figure 8. Left: the combined experimental and theoretical correlation matrix for the dataset used in the NNPDF3.1 analysis. Note how theory-induced correlations now connect different measurements, such as DIS and Drell-Yan, which are experimentally uncorrelated. Right: the results of the NNPDF3.1 NLO fit with theory uncertainties, compared with a baseline with only experimental errors in the covariance matrix and with the central value of the corresponding NNLO fit.

PDFs with theory uncertainties. An aspect of the global PDF fitting machinery that has attracted a lot of attention is the estimate and propagation of the associated uncertainties. In most PDF fits, what is provided as uncertainty estimate accounts only for the propagation of the uncertainties corresponding to the input experimental measurements, as well as for methodological components associated for example to the choice of functional form. PDF analysis however do not account in general for the theory uncertainties associated to the truncation of the perturbative expansions of the input QCD calculations. While neglecting these theory errors might have been justified in the past, the rapid pace of improvements in PDF fits, leading to ever smaller uncertainties, suggested that such assumption should be revisited.

In this context, an important recent development in global PDF analyses has been the formulation of a consistent framework to systematically include theoretical uncertainties [Abdul Khalek et al., 2019b, Abdul Khalek et al., 2019c], in particular those associated to the missing higher orders (MHOUs) in the perturbative QCD calculations used as input to the fit. The basic idea of this approach is to construct a combined covariance matrix that includes both experimental and theoretical components, with the latter estimated using the scale-variation method and validated whenever possible with existing higher order calculations.

One of the main consequences of the theory covariance matrix approach, beyond an overall increase of the total uncertainties, is that theory-induced correlations now connect different measurements, such as DIS and Drell-Yan or top quark production, which are experimentally uncorrelated. This feature is illustrated by the left panel of Fig. 8, which displays the combined experimental and theoretical correlation matrix for the dataset used in the NNPDF3.1 NLO analysis. There we see for example that theory uncertainties introduce a positive correlation between DIS and Drell-Yan in some kinematic regions, and a negative one in others.

This formalism has been used to carry out an updated version of the NNPDF3.1 NLO global analysis now accounting for the MHOUs. In the right panel of Fig. 8 we show the results of the NNPDF3.1 NLO fit with theory uncertainties (red hatched band) for the gluon PDF at $Q = 10$ GeV, compared with a baseline with only experimental errors in the covariance matrix (green solid band) and with the central value of the corresponding NNLO fit. We can observe that there are two main consequences of including the theory uncertainties in the fit covariance matrix. First of all, a moderate increase of the PDF uncertainties, showing that theory errors cannot be neglected when estimating the total PDF error budget, at least at NLO. Second, a shift in the central values due to the rebalancing effect of theory errors and their correlations, which is found in general to shift the fit results towards the corresponding NNLO result. The latter is a desirable feature of any method that estimates theory uncertainties, whose ultimate

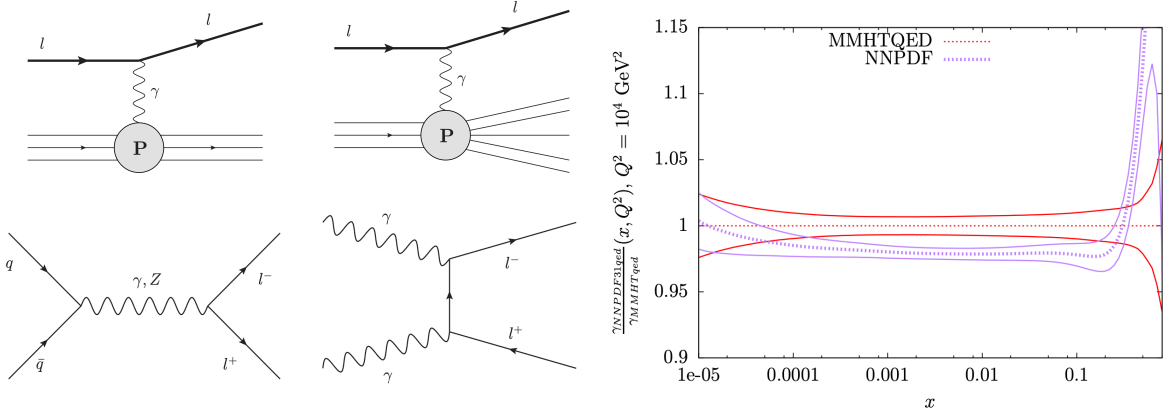


Figure 9. Upper left: the $F_{2,L}$ deep-inelastic structure functions have associated photon-induced contributions from elastic (where the proton remains intact) and inelastic processes (where the proton breaks down), from [Harland-Lang et al., 2019]. Bottom left: photon-induced processes contribute to the total cross-sections of many important hadron collider processes, in this case Drell-Yan lepton-pair production. Right: comparison between the photon PDFs $f_{\gamma}(x, Q = 100 \text{ GeV})$ between the NNPDF and MMHT QED analysis, normalised to the central value of the latter.

goal is to accurately predict the results of the next perturbative order.

The photon PDF. A perhaps unexpected observation whose implications were only fully appreciated recently is related to the fact that the partonic content of the proton should be composed not only by quarks and gluons, but also by photons. Indeed, it can be demonstrated that protons radiate quasi-real photons that induce photon-initiated scattering reactions in lepton-proton and in proton-proton collisions, and that these reactions can be described as arising from a photon PDF $f_{\gamma}(x, Q^2)$. This photon PDF receives two types of contributions, as schematically indicated in the upper left panel of Fig. 9: from elastic processes, where the proton is left intact, and from inelastic processes, where the proton breaks down after the scattering.

The presence of a photon component in the proton has two main phenomenological implications. First of all, the photon mixes with the quark and gluon PDFs via QED effects in the DGLAP evolution, and thus modifies the results of the latter as compared to the QCD-only case. In particular, the photon subtracts from the quarks and gluons a small amount of the total momentum of the proton, which is currently estimated to amount to a few permille. Second, photon-initiated contributions result in the opening of new channels for important hard-scattering processes. An important example of this is shown in the bottom left panel of Fig. 9: photon-initiated processes contribute to the total Drell-Yan cross-section (lepton pair production) on the same footing as the quark-antiquark annihilation reactions.

Until recently, the photon PDF was determined either from model assumptions [Martin et al., 2005, Schmidt et al., 2016] or freely parametrised and constrained by experimental data [Ball et al., 2013]. None of these options were satisfactory: the former due to the bias associated to the choice of model, the latter since the results were affected by large uncertainties due to limited available constraints. A major breakthrough was then the demonstration that the photon content of the proton does not need neither to be modeled nor to be fitted from data, but rather it can be computed from first principles and expressed in terms of the well-known inclusive deep-inelastic scattering structure functions F_2 and F_L [Manohar et al., 2016, Manohar et al., 2017], an approach that is known as the LUXqed formalism.

Since these pioneering analyses were presented, other global PDF fitting groups have provided QED variants of their PDF sets that include both QED corrections in the DGLAP evolution with a photon PDF determined by means of the LUXqed calculation or variations thereof [Bertone et al., 2018, Harland-Lang et al., 2019]. Given the relatively tight constraints imposed by the LUXqed framework, the resulting photon PDFs turn out to be quite similar. To illustrate this feature, in the right panel of Fig. 9 we display a comparison of the photon PDFs $f_{\gamma}(x, Q)$ at $Q = 100 \text{ GeV}$ between the NNPDF and MMHT QED analyses, normalised to the central value

of the latter. The two photon PDFs agree at the few percent level in most of the relevant x range, with the exception of the large- x regions where differences can be as large as 15%.

Implications for astroparticle physics. Another topic that has received ample attention in the recent years has been the interplay between proton structure studies in nuclear and particle physics with those in astroparticle physics. This interest was motivated by the realisation that improving our understanding of the partonic content of nucleons and nuclei was an important ingredient for those theoretical predictions relevant for the interpretation of high-energy astrophysics experiments. This connection is particularly important for neutrino telescopes, such as IceCube and KM3NET, which instrument large volumes of ice and water as effective detectors of energetic neutrinos, as well as for cosmic ray detectors such as Auger, which aim to detect the most energetic particles roaming through the Universe.

In the specific case of high-energy neutrino astronomy, QCD calculations are required for two different aspects of the data interpretation. First of all, to predict the expected event rates of neutrino-nucleus interactions at high energies (upper left panel of Fig. 10) via the neutral-current and charged current deep-inelastic scattering processes. Knowledge of these cross-sections is required both to predict the number of neutrinos that will interact with the nuclei in the ice or water targets, as well as the attenuation of the incoming neutrino flux as they traverse the Earth material [Vincent et al., 2017]. As indicated in the bottom left panel of Fig. 10, for a neutrino with an energy of $E_\nu = 5 \times 10^{10}$ GeV the charged-current DIS process probes the proton PDFs down to $x \simeq 10^{-8}$ [Cooper-Sarkar et al., 2011], a region far beyond present experimental constraints. This implies that the interpretation of the measurements of neutrino telescopes could be hindered by theoretical uncertainties related to QCD and the partonic content of protons.

Secondly, QCD processes are also relevant for the prediction of the dominant background for neutrino astronomy, the so called prompt neutrino flux. In this process, represented in the upper right panel of Fig. 10, an energetic cosmic ray, typically a proton, collides with an air nucleus and produces a charmed meson, that promptly decays into neutrinos. The flux of such neutrinos becomes at high energies larger than those from pion and kaon decays, since the short lifetime of D mesons implies that their energy is not attenuated before decay. As in the neutrino cross-section case, the calculation of the prompt neutrino flux involves knowledge of the proton structure for $Q \simeq m_c = 1.5$ GeV down to $x \simeq 10^{-6}$ [Zenaiev et al., 2015], where theoretical uncertainties are large.

The key ingredient that allows providing solid theoretical predictions for both types of processes is the exploitation of collider measurements sensitive to the small- x structure of the proton, such as the charm production data in the forward region from the LHCb experiment. Forward D meson production at LHCb, when Lorentz-boosted to the center of mass frame, covers the same kinematic region as that of charm production in energetic cosmic ray collisions, and therefore makes possible pinning down the gluon at small- x beyond the reach of the HERA collider data and in the kinematical region of relevance for neutrino astronomy. In the bottom right of Fig. 10 we display the significant reduction in the PDF uncertainties of the small- x gluon once the LHCb charm production measurements at $\sqrt{s} = 5, 7$ and 13 TeV are included in the NNPDF3.0 fit [Gauld and Rojo, 2017]. Thanks to this connection with LHC hard-scattering processes, it has become possible to provide robust state-of-the-art predictions for signal [Bertone et al., 2019] and background [Bhattacharya et al., 2016] processes at neutrino telescopes with reduced theoretical uncertainties. In the future, one can envisage a situation where measurements at high-energy astroparticle physics experiments can be used to constrain the small- x structure of the nucleon and the associated QCD dynamics in this regime.

PDFs and BSM searches. As more data is being collected and the precision and breadth of the LHC measurements improves, the information on the partonic structure of the proton that one can obtain from them will be consequently increased. In particular, the large integrated luminosities that will be accumulated in future runs of the LHC imply that extended measurements in the TeV region should make possible a more precise determination of poorly known PDFs such as the large- x sea quarks and the gluons, see also the left panel of Fig. 7. A potential

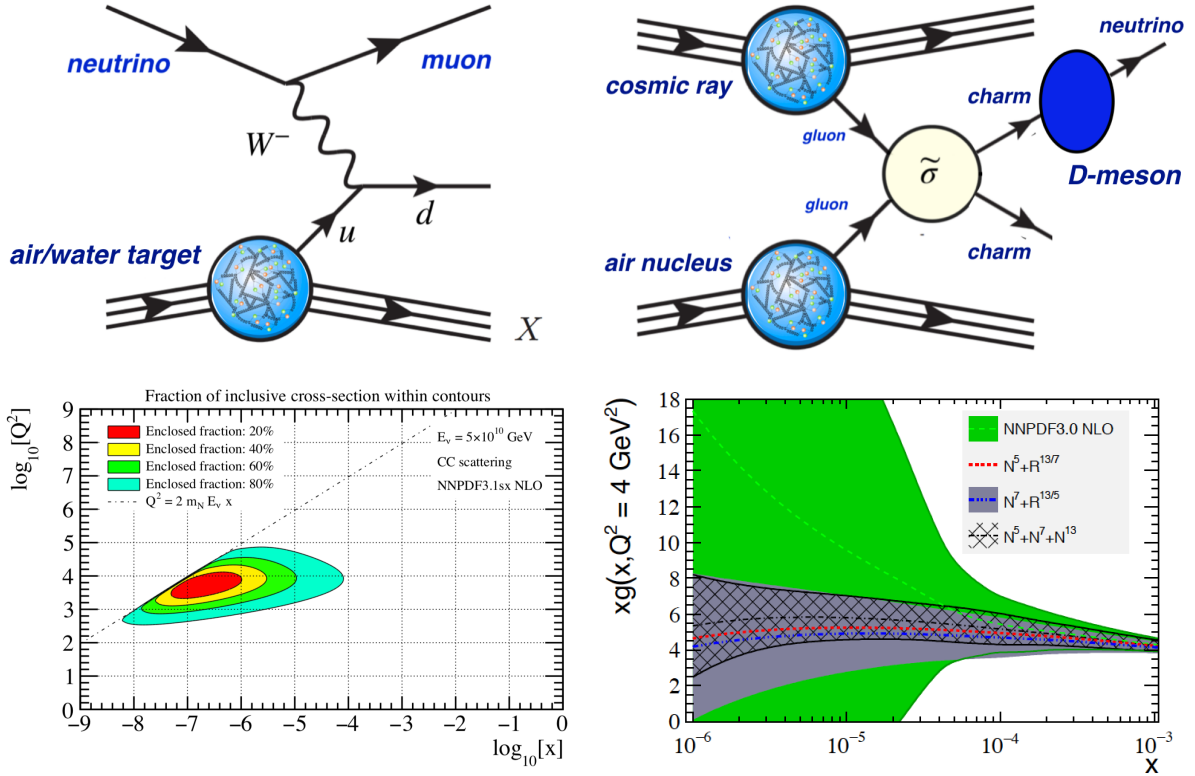


Figure 10. Upper plots: theoretical QCD calculations sensitive to the proton structure are required in high-energy neutrino astronomy both to predict signal event rates in neutrino detection (left panel) and the prompt neutrinos flux arising from cosmic ray collisions (right panel). Bottom left: the kinematic coverage in the (x, Q^2) plane for charged-current neutrino-nucleus scattering at an energy of $E_\nu = 5 \times 10^{10}$ GeV. Bottom right: the reduction in the PDF uncertainties of the small- x gluon at $Q^2 = 4 \text{ GeV}^2$ once the LHCb charm production measurements are included in the NNPDF3.0 fit.

worry in using LHC data in this high-energy region for PDF fits is that such PDF interpretation assumes the SM in its input theory calculations, and could thus be biased if deviations with respect to the SM were present. This concern is becoming more acute due to the lack of new particles or interactions detected so far at the LHC: this might suggest the presence of a band gap between the electroweak scale and the scale of new physics Λ . In turn, such a band gap could imply that Beyond the Standard Model (BSM) dynamics could very well manifest itself at the LHC only via subtle deviations in the tails of the measured distributions.

A powerful framework to interpret in a model-independent manner the results of the LHC is provided by the Standard Model Effective Field Theory (SMEFT) [Brivio and Trott, 2019]. Within the mathematical language of the SMEFT, the effects of BSM dynamics at high energies $\Lambda \gg v$ above the electroweak scale ($v = 246 \text{ GeV}$) can be parametrised at lower energies, $E \ll \Lambda$, in terms of higher-dimensional operators built up from Standard Model fields and satisfying its symmetries such as gauge invariance. Since several of the processes that are used at the LHC to constrain the SMEFT degrees of freedom are also used as inputs to PDF determinations, it is important to assess their interplay and to establish whether one could use the global PDF fits to disentangle BSM effects from QCD dynamics, as originally proposed in [Berger et al., 2010].

A first study in this direction has been recently presented in [Carrazza et al., 2019], where variants of the NNPDF3.1 DIS-only fit were carried out using theory calculations where the SM had been extended by specific subsets of dimension-6 four-fermion SMEFT operators. In the left panel of Fig. 11 we show the results of the gluon PDF at large- x for $Q = 10 \text{ GeV}$ in those fits. For benchmark points in the parameter space not already excluded by other experiments, one finds that the shift due to SMEFT corrections in the theory calculation is at most half of the PDF uncertainty. Crucially, one can exploit different scaling of the χ^2 with the process' energy to disentangle QCD effects from genuine BSM dynamics, as shown in the right panel of Fig. 11: the former are smooth as the energy increases, since DGLAP evolution effects are logarithmic in

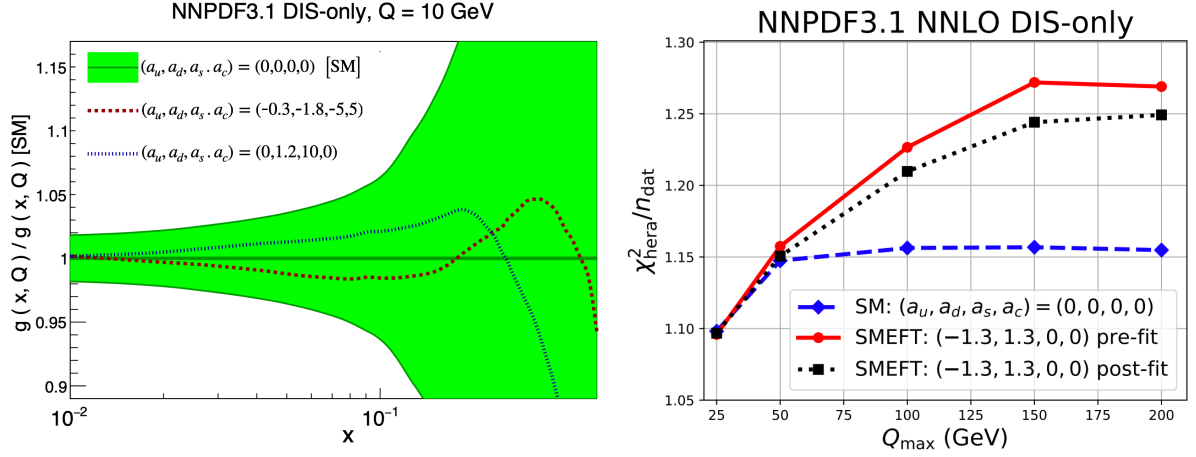


Figure 11. Left: the gluon PDF at large- x for $Q = 10$ GeV, comparing the results of fits based on SM calculations with those where the SM has been extended by specific combinations of SMEFT $d = 6$ operators. Right: the different scaling of the χ^2 with the energy of the process can be used to disentangle QCD effects from genuine BSM dynamics, in the case shown for a specific SMEFT benchmark point.

Q^2 , while the latter are much more marked since they scale as a power of Q^2 . While this study found that SMEFT-induced distortions are sub-dominant with respect to PDF uncertainties, it was restricted to DIS measurements and the picture could change significantly in the case of LHC measurements, specially the high-statistics ones from future LHC runs.

Constraints on nPDFs from LHC data. Concerning nuclear PDFs, perhaps the most important recent development has been the availability of a wide variety of hard probe measurements from proton-lead collisions at the LHC. In a similar way as in the proton PDF case, these measurements provide new and valuable constraints on various nPDF combinations, in particular for those that can only be loosely constrained from DIS data such as the gluon and the quark flavour separation. The information on the nuclear modifications of the gluon PDF is particularly valuable, since these are essentially unconstrained if only deep-inelastic scattering data is used in the nPDF fit. Furthermore, these hard-scattering LHC measurements offer novel opportunities to test the validity of both the nPDF universality and of the QCD factorisation properties in the nuclear environment. Of specific interest is the small- x regime, where eventual non-linear saturation dynamics are expected to be enhanced as compared to the free-nucleon case.

LHC measurements from proton-lead collisions that have demonstrated their constraining power for nPDFs include dijet production (sensitive to the gluon), D meson production in the forward region from LHCb (to pin down the small- x gluon nPDF), and electroweak gauge boson production (providing a handle on the quark flavour separation). To illustrate the information on the nPDFs provided by hard probes in proton-lead collisions from the LHC, in Fig. 12 we show in the left panel the nuclear ratio R_g^{Pb} for gluons in lead at $Q = 100$ GeV in the EPPS16 analysis before and after including the constraints from the CMS dijet measurements at $\sqrt{s} = 5.02$ TeV via Hessian profiling. One can observe how these data reduce the uncertainties of R_g^{Pb} for a wide range of values of x , and in particular they appear to suggest the presence of gluon shadowing at small- x . Then in the right panel of Fig. 12 we display the forward-backward asymmetry in W production in proton-lead collisions from CMS, compared with the nCTEQ15 predictions before and after including this measurement in their fit. The agreement between data and theory is clearly improved once the data on A_{FB} is added to the analysis, with χ^2 / n_{dat} decreasing from 4.03 to 1.31, and also the uncertainties in the theory prediction (which arise from the flavour decomposition of the quark nPDF) are likewise reduced.

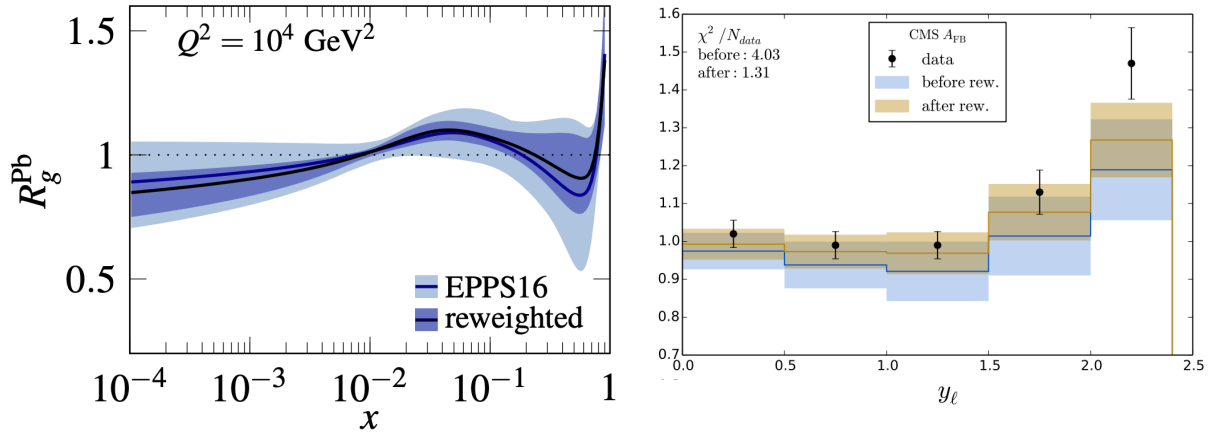


Figure 12. Left: the gluon nuclear ratio R_g^{Pb} in lead at $Q = 100$ GeV in the EPPS16 analysis before and after including the constraints from the CMS dijet measurements at $\sqrt{s} = 5.02$ TeV via Hessian profiling. Right: the forward-backward asymmetry in W production in proton-lead collisions from CMS, compared with the nCTEQ15 predictions before and after including this measurement in their fit.

5 Summary and outlook

As discussed through this Article, the study of the partonic structure of nucleons and nuclei is an exciting and rich field at the cross-roads between different aspects of particle, nuclear, and astroparticle physics. On the one hand, it allows us to shed light into the dynamics of the strong nuclear force, providing crucial input on pressing questions such as the origin of mass and spin of protons, the possible onset of new gluon-dominated regimes at small- x , the strange and heavy quark contribution to the nucleon’s wave function, or the dynamics of quark and gluons for nucleons bound within heavy nuclei. On the other hand, it provides essential input for the theoretical predictions in processes as diverse as the production of Higgs bosons at the LHC, the interactions of high-energy neutrinos at IceCube, and the collisions between lead nuclei at the Relativistic Heavy Ion Collider.

The field has witnessed in the recent years a number of important results which highlight its vitality and productivity. Just to mention a few of these: the demonstration of the impact on the proton structure of precision LHC measurements; the improved determinations of the strange, charm, and photon content of the proton; the first evidence for BFKL small- x dynamics in HERA data [Ball et al., 2018]; the calculation of x -space distributions with lattice QCD; the formulation of new frameworks to estimate and propagate theory uncertainties; the establishment of the connection with neutrino telescopes and cosmic ray physics; and the analysis of the interplay between proton structure and direct and indirect searches for New Physics at the high-energy frontier.

Many of these achievements in our understanding of the partonic structure of nucleons and nuclei have only become possible thanks to progress from the methodological side, from the development of new Machine Learning algorithms to parameterize and train the PDFs [Carrazza and Cruz-Martinez, 2019] to strategies to combine and compress different sets of parton distributions [Gao and Nadolsky, 2014, Carrazza et al., 2015b] and new methods to quantify and represent graphically the information provided by individual datasets [Wang et al., 2018]. Nevertheless, the lists above are necessarily limited due to space restrictions of this Article, and the interested reader is encouraged to turn to the more extended technical reviews mentioned in the introduction.

A fair overall assessment of the state of the field could be that, while great progress has been made in addressing long-standing questions in our knowledge of the partonic structure of the proton, important open questions still remain which should be addressed in the coming years. Therefore, to conclude this Article I wanted to mention two possible future directions for the field. The first one focuses on the exploitation of the constraints that will be provided by PDF-sensitive measurements at future facilities, some of them already approved, such as the High-Luminosity LHC, and some others still under discussion, such as the Electron Ion Collider

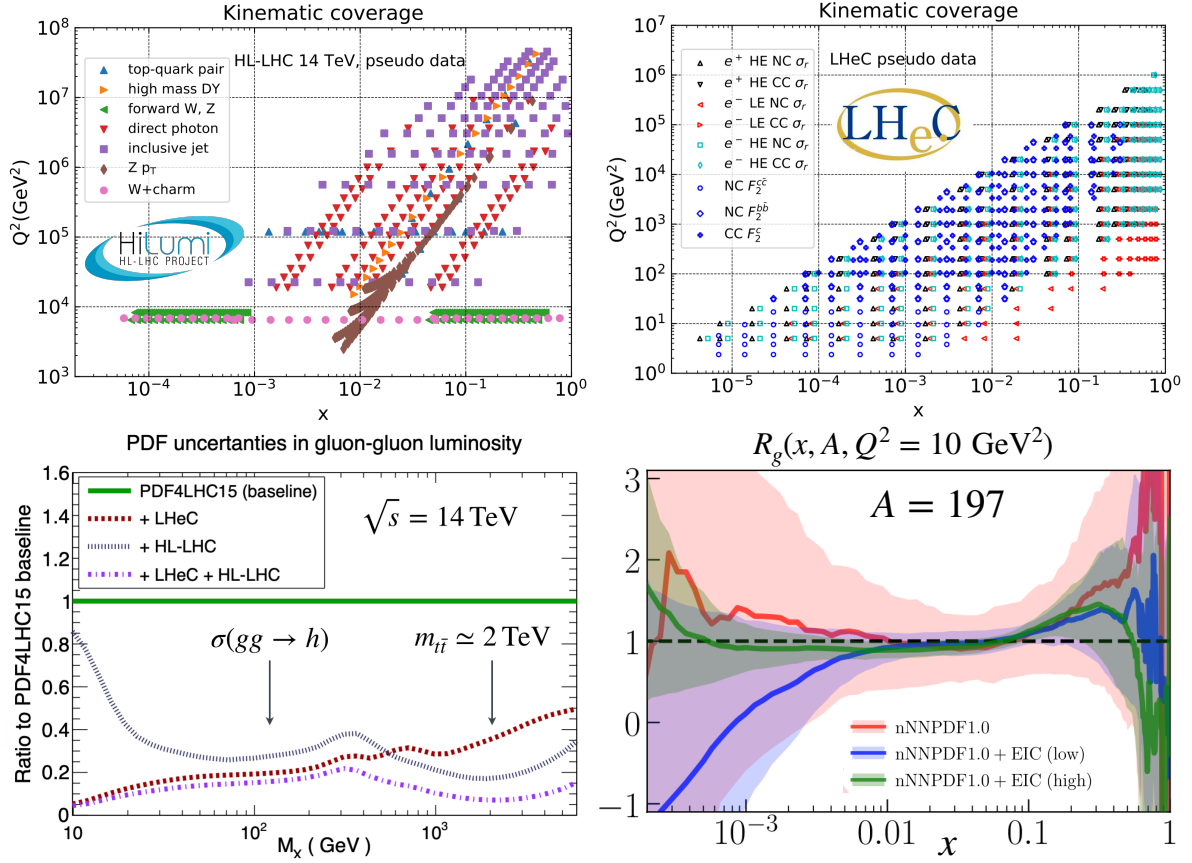


Figure 13. Upper plots: the kinematic coverage in the (x, Q^2) plane that will become available at the HL-LHC (left) and the one that would become available if the LHeC is approved (right). Bottom left: the reduction in the PDF uncertainties of the gluon-gluon luminosity at the LHC 14 TeV once the PDF information contained in the LHeC and HL-LHC projections are taken into account, compared with the PDF4LHC15 baseline. Bottom right: the improvement in the determination of the nuclear gluon PDF expected at the EIC, for two different scenarios for its center of mass energy.

or the Large Hadron electron Collider. The second direction points towards the unification of different aspects of the global QCD analysis paradigm into a unified framework that combines them in a fully consistent way. I will now discuss these two topics in turn.

In order to further delve in the mysteries of the proton structure, several future and proposed facilities should play a crucial role. To begin with, the high-luminosity upgrade of the LHC (HL-LHC), which will operate between 2025 and the late 2030s, will deliver a total integrated luminosity of more than $\mathcal{L} = 3 \text{ ab}^{-1}$ for ATLAS and CMS and $\mathcal{L} = 0.3 \text{ ab}^{-1}$ for LHCb. Such a gigantic dataset will provide ample opportunities for PDF studies, in particular with the measurements of cross-sections in the few TeV region for processes such as dijet, top quark pair, direct photon, and Drell-Yan production, or the transverse momentum distributions of weak gauge bosons. These measurements would constrain the large- x behaviour of the poorly known gluon and sea quarks, which in turn would lead to improved searches for new heavy particles predicted in scenarios of new physics beyond the Standard Model [Beenakker et al., 2016].

One proposed future facility that would provide unique input for both the proton and nuclear PDFs would be the Large Hadron electron Collider (LHeC) [Abeleira Fernandez et al., 2012]. The LHeC would extend the kinematical coverage achieved at HERA by more than one order of magnitude at small- x and at large- Q^2 , while operating both with proton and with light and heavy nuclear beams. In the upper panels of Fig. 13 we display the kinematic coverage in the (x, Q^2) plane of the HL-LHC and the LHeC. This comparison highlights their complementary, with the LHeC providing a superior handle in the small- x region (and in particular allowing for tests of novel QCD dynamics) with the HL-LHC offering unparalleled reach in the high-energy frontier. In the bottom left panel of Fig. 13 we display the expected reduction in the PDF uncertainties of the gluon-gluon luminosity at the LHC 14 TeV based on the LHeC and HL-LHC pseudo-

data projections presented in [Abdul Khalek et al., 2018, Abdul Khalek et al., 2019a]. In these forecasts, the baseline is taken to be the PDF4LHC15 set [Butterworth et al., 2016], which is assumed to represent our current knowledge of the proton PDFs. One can observe that in the most favorable scenario, where both HL-LHC and LHeC operate simultaneously, one might achieve an uncertainty reduction by up to an order of magnitude for several values of the final state invariant mass M_X .

The Electron Ion Collider (EIC) [Accardi et al., 2016], an US-based proposal, would also be able to scrutinise the properties of nucleons and nuclei with unprecedented detail. To illustrate its capabilities, in the bottom right panel of Fig. 13 we show the improvement in the determination of the nuclear gluon PDF expected at the EIC for two scenarios for its center of mass energy in the case of the nNNPDF1.0 analysis. These projections indicate that the EIC could pin down the nuclear gluon modifications down to $x \simeq 10^{-4}$, and would thus allow establishing the possible onset of new QCD dynamics such as non-linear (saturation) effects. Furthermore, the EIC would have the unique feature of being able to chart the proton spin structure in the small- x regime, as well as its three-dimensional structure for the first time in a wide kinematic range.

As mentioned above, another possible direction in which one should expect future progress in the field of nucleon structure is related to the combination of individual QCD fits. Let me illustrate what I mean by this with a specific example. Several proton PDF fits include data taken on nuclear targets, however they neglect to account for the effects of nuclear modifications. Furthermore, most nuclear PDF fits assume a proton PDF baseline (and its uncertainties) that has been extracted by other groups with in general different methodologies and input assumptions. Such a situation is far from optimal, since the assumption that the extraction of proton and nuclear PDFs can be decoupled from each other is not justified anymore given recent progress in experimental data, theory calculations, and methodological developments. Indeed, a more robust approach would consist on extracting simultaneously the free nucleon ($A = 1$) and the nuclear ($A > 1$) PDFs from a single QCD analysis, thus being able to keep track of their mutual interplay.

Similar considerations apply to other aspects of the QCD fitting paradigm. For instance, there is a natural cross-talk between the (un)polarised PDFs and the hadron fragmentation functions, which are connected by means of the semi-inclusive DIS (SIDIS) processes. SIDIS differs from standard DIS because one of the hadrons in the final state has been identified. This implies that the cross-section for this process depends both on the PDFs of the initial state proton as well as on the fragmentation functions of the final state hadron. Therefore, one should strive to determine simultaneously the (un)polarised PDFs and the fragmentation functions from the same joint QCD analysis, as done for example in [Ethier et al., 2017]. This approach can provide information on the partonic properties that is not available via other channels, for instance to constrain the strange content of the proton from unpolarised SIDIS measurements [Borsa et al., 2017].

In the long term, the community should aim to assemble a truly global analysis of non-perturbative QCD objects to extract simultaneously the unpolarised and polarised proton PDFs, the nuclear PDFs, and the hadron fragmentation functions. However, there are two main requirements that need to be satisfied in order to be able to carry out such ambitious goal. The first one is methodological progress in the fitting side that ensures that the large parameter space that arises once all QCD objects are jointly extracted can be explored efficiently. The second is the availability of new facilities that can provide suitable experimental measurements to constrain all the relevant non-perturbative QCD objects and their correlations. In this context, a machine such as the Electron Ion collider would be specially suited for the realisation of this ultimate “integrated” global QCD analysis.

Acknowledgments. J. R. has been supported by the European Research Council Starting Grant “PDF4BSM” as well as by the Netherlands Organization for Scientific Research (NWO).

References

- [Aad et al., 2012] Aad, G. et al. (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, arXiv:1207.7214.
- [Abdul Khalek et al., 2018] Abdul Khalek, R., Bailey, S., Gao, J., Harland-Lang, L., and Rojo, J. (2018). Towards Ultimate Parton Distributions at the High-Luminosity LHC. *Eur. Phys. J.*, C78(11):962, arXiv:1810.03639.
- [Abdul Khalek et al., 2019a] Abdul Khalek, R., Bailey, S., Gao, J., Harland-Lang, L., and Rojo, J. (2019a). Probing Proton Structure at the Large Hadron electron Collider. arXiv:1906.10127.
- [Abdul Khalek et al., 2019b] Abdul Khalek, R. et al. (2019b). A First Determination of Parton Distributions with Theoretical Uncertainties. arXiv:1905.04311.
- [Abdul Khalek et al., 2019c] Abdul Khalek, R. et al. (2019c). Parton Distributions with Theory Uncertainties: General Formalism and First Phenomenological Studies. arXiv:1906.10698.
- [Abdul Khalek et al., 2019d] Abdul Khalek, R., Ethier, J. J., and Rojo, J. (2019d). Nuclear parton distributions from lepton-nucleus scattering and the impact of an electron-ion collider. *Eur. Phys. J.*, C79(6):471, arXiv:1904.00018.
- [Abelleira Fernandez et al., 2012] Abelleira Fernandez, J. et al. (2012). A Large Hadron Electron Collider at CERN: Report on the Physics and Design Concepts for Machine and Detector. *J.Phys.*, G39:075001, arXiv:1206.2913.
- [Abreu et al., 2008] Abreu, S. et al. (2008). Proceedings, Workshop on Heavy Ion Collisions at the LHC: Last Call for Predictions. *J. Phys.*, G35:054001, arXiv:0711.0974.
- [Accardi et al., 2016] Accardi, A. et al. (2016). Electron Ion Collider: The Next QCD Frontier. *Eur. Phys. J.*, A52(9):268, arXiv:1212.1701.
- [Aidala et al., 2013] Aidala, C. A., Bass, S. D., Hasch, D., and Mallot, G. K. (2013). The Spin Structure of the Nucleon. *Rev. Mod. Phys.*, 85:655–691, arXiv:1209.2803.
- [Alekhin et al., 2017] Alekhin, S., Blümlein, J., Moch, S., and Placakyte, R. (2017). Parton distribution functions, α_s , and heavy-quark masses for LHC Run II. *Phys. Rev.*, D96(1):014011, arXiv:1701.05838.
- [Angeles-Martinez et al., 2015] Angeles-Martinez, R. et al. (2015). Transverse Momentum Dependent (TMD) parton distribution functions: status and prospects. *Acta Phys. Polon.*, B46(12):2501–2534, arXiv:1507.05267.
- [Aubert et al., 1983] Aubert, J. J. et al. (1983). The ratio of the nucleon structure functions F_2^n for iron and deuterium. *Phys. Lett.*, 123B:275–278.
- [Ball et al., 2018] Ball, R. D., Bertone, V., Bonvini, M., Marzani, S., Rojo, J., and Rottoli, L. (2018). Parton distributions with small- x resummation: evidence for BFKL dynamics in HERA data. *Eur. Phys. J.*, C78(4):321, arXiv:1710.05935.
- [Ball et al., 2012] Ball, R. D., Bertone, V., Cerutti, F., Del Debbio, L., Forte, S., et al. (2012). Reweighting and Unweighting of Parton Distributions and the LHC W lepton asymmetry data. *Nucl.Phys.*, B855:608–638, arXiv:1108.1758.
- [Ball et al., 2013] Ball, R. D. et al. (2013). Parton distributions with QED corrections. *Nucl.Phys.*, B877:290–320, arXiv:1308.0598.
- [Ball et al., 2015] Ball, R. D. et al. (2015). Parton distributions for the LHC Run II. *JHEP*, 04:040, arXiv:1410.8849.

- [Ball et al., 2017] Ball, R. D. et al. (2017). Parton distributions from high-precision collider data. *Eur. Phys. J.*, C77(10):663, arXiv:1706.00428.
- [Beenakker et al., 2016] Beenakker, W., Borschensky, C., Krämer, M., Kulesza, A., Laenen, E., Marzani, S., and Rojo, J. (2016). NLO+NLL squark and gluino production cross-sections with threshold-improved parton distributions. *Eur. Phys. J.*, C76(2):53, arXiv:1510.00375.
- [Berger et al., 2010] Berger, E. L., Guzzi, M., Lai, H.-L., Nadolsky, P. M., and Olness, F. I. (2010). Constraints on color-octet fermions from a global parton distribution analysis. *Phys. Rev.*, D82:114023, arXiv:1010.4315.
- [Bertone et al., 2018] Bertone, V., Carrazza, S., Hartland, N. P., and Rojo, J. (2018). Illuminating the photon content of the proton within a global PDF analysis. *SciPost Phys.*, 5(1):008, arXiv:1712.07053.
- [Bertone et al., 2014] Bertone, V., Frederix, R., Frixione, S., Rojo, J., and Sutton, M. (2014). aMCfast: automation of fast NLO computations for PDF fits. *JHEP*, 1408:166, arXiv:1406.7693.
- [Bertone et al., 2019] Bertone, V., Gauld, R., and Rojo, J. (2019). Neutrino Telescopes as QCD Microscopes. *JHEP*, 01:217, arXiv:1808.02034.
- [Bhattacharya et al., 2016] Bhattacharya, A., Enberg, R., Jeong, Y. S., Kim, C. S., Reno, M. H., Sarcevic, I., and Stasto, A. (2016). Prompt atmospheric neutrino fluxes: perturbative QCD models and nuclear effects. *JHEP*, 11:167, arXiv:1607.00193.
- [Bloom et al., 1969] Bloom, E. D. et al. (1969). High-Energy Inelastic e p Scattering at 6-Degrees and 10-Degrees. *Phys. Rev. Lett.*, 23:930–934.
- [Borsa et al., 2017] Borsa, I., Sassot, R., and Stratmann, M. (2017). Probing the Sea Quark Content of the Proton with One-Particle-Inclusive Processes. *Phys. Rev.*, D96(9):094020, arXiv:1708.01630.
- [Brandelik et al., 1979] Brandelik, R. et al. (1979). Evidence for Planar Events in e+ e- Annihilation at High-Energies. *Phys. Lett.*, 86B:243–249.
- [Brivio and Trott, 2019] Brivio, I. and Trott, M. (2019). The Standard Model as an Effective Field Theory. *Phys. Rept.*, 793:1–98, arXiv:1706.08945.
- [Butterworth et al., 2016] Butterworth, J. et al. (2016). PDF4LHC recommendations for LHC Run II. *J. Phys.*, G43:023001, arXiv:1510.03865.
- [Carli et al., 2010] Carli, T. et al. (2010). A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project. *Eur. Phys. J.*, C66:503, arXiv:0911.2985.
- [Carrazza and Cruz-Martinez, 2019] Carrazza, S. and Cruz-Martinez, J. (2019). Towards a new generation of parton densities with deep learning models. *Eur. Phys. J.*, C79(8):676, arXiv:1907.05075.
- [Carrazza et al., 2019] Carrazza, S., Degrande, C., Iranipour, S., Rojo, J., and Ubiali, M. (2019). Can New Physics hide inside the proton? *Phys. Rev. Lett.*, 123(13):132001, arXiv:1905.05215.
- [Carrazza et al., 2015a] Carrazza, S., Ferrara, A., Palazzo, D., and Rojo, J. (2015a). APFEL Web: a web-based application for the graphical visualization of parton distribution functions. *J. Phys.*, G42:057001, arXiv:1410.5456.
- [Carrazza et al., 2015b] Carrazza, S., Latorre, J. I., Rojo, J., and Watt, G. (2015b). A compression algorithm for the combination of PDF sets. *Eur. Phys. J.*, C75:474, arXiv:1504.06469.
- [Chatrchyan et al., 2012] Chatrchyan, S. et al. (2012). Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett.*, B716:30–61, 1207.7235.

- [Cooper-Sarkar et al., 2011] Cooper-Sarkar, A., Mertsch, P., and Sarkar, S. (2011). The high energy neutrino cross-section in the Standard Model and its uncertainty. *JHEP*, 08:042, arXiv:1106.3723.
- [de Florian et al., 2016] de Florian, D. et al. (2016). Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. arXiv:1610.07922.
- [Del Debbio et al., 2007] Del Debbio, L., Forte, S., Latorre, J. I., Piccione, A., and Rojo, J. (2007). Neural network determination of parton distributions: The nonsinglet case. *JHEP*, 03:039, hep-ph/0701127.
- [Eskola et al., 2017] Eskola, K. J., Paakkinen, P., Paukkunen, H., and Salgado, C. A. (2017). EPPS16: Nuclear parton distributions with LHC data. *Eur. Phys. J.*, C77(3):163, arXiv:1612.05741.
- [Ethier et al., 2017] Ethier, J. J., Sato, N., and Melnitchouk, W. (2017). First simultaneous extraction of spin-dependent parton distributions and fragmentation functions from a global QCD analysis. *Phys. Rev. Lett.*, 119(13):132001, arXiv:1705.05889.
- [Forte and Watt, 2013] Forte, S. and Watt, G. (2013). Progress in the Determination of the Partonic Structure of the Proton. *Ann.Rev.Nucl.Part.Sci.*, 63:291, arXiv:1301.6754.
- [Gao et al., 2018] Gao, J., Harland-Lang, L., and Rojo, J. (2018). The Structure of the Proton in the LHC Precision Era. *Phys. Rept.*, 742:1–121, arXiv:1709.04922.
- [Gao and Nadolsky, 2014] Gao, J. and Nadolsky, P. (2014). A meta-analysis of parton distribution functions. *JHEP*, 1407:035, arXiv:1401.0013.
- [Gauld and Rojo, 2017] Gauld, R. and Rojo, J. (2017). Precision determination of the small- x gluon from charm production at LHCb. *Phys. Rev. Lett.*, 118(7):072001, arXiv:1610.09373.
- [Gell-Mann, 1964] Gell-Mann, M. (1964). A Schematic Model of Baryons and Mesons. *Phys. Lett.*, 8:214–215.
- [Gross and Wilczek, 1973] Gross, D. J. and Wilczek, F. (1973). Ultraviolet Behavior of Non-abelian Gauge Theories. *Phys. Rev. Lett.*, 30:1343–1346. [271(1973)].
- [Harland-Lang et al., 2015] Harland-Lang, L. A., Martin, A. D., Motylinski, P., and Thorne, R. S. (2015). Parton distributions in the LHC era: MMHT 2014 PDFs. *Eur. Phys. J.*, C75:204, arXiv:1412.3989.
- [Harland-Lang et al., 2019] Harland-Lang, L. A., Martin, A. D., Nathvani, R., and Thorne, R. S. (2019). Ad Lucem: QED Parton Distribution Functions in the MMHT Framework. arXiv:1907.02750.
- [Hobbs et al., 2019] Hobbs, T. J., Wang, B.-T., Nadolsky, P. M., and Olness, F. I. (2019). The coming synergy between lattice QCD and high-energy phenomenology. arXiv:1904.00022.
- [Hou et al., 2019] Hou, T.-J. et al. (2019). Progress in the CTEQ-TEA NNLO global QCD analysis. arXiv:1908.11394.
- [Kovarik et al., 2016] Kovarik, K. et al. (2016). nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework. *Phys. Rev.*, D93(8):085037, arXiv:1509.00792.
- [Kovářík et al., 2019] Kovářík, K., Nadolsky, P. M., and Soper, D. E. (2019). Hadron structure in high-energy collisions. arXiv:1905.06957.
- [Lin et al., 2018] Lin, H.-W. et al. (2018). Parton distributions and lattice QCD calculations: a community white paper. *Prog. Part. Nucl. Phys.*, 100:107–160, arXiv:1711.07916.

- [Manohar et al., 2016] Manohar, A., Nason, P., Salam, G. P., and Zanderighi, G. (2016). How bright is the proton? A precise determination of the photon parton distribution function. *Phys. Rev. Lett.*, 117(24):242002, arXiv:1607.04266.
- [Manohar et al., 2017] Manohar, A. V., Nason, P., Salam, G. P., and Zanderighi, G. (2017). The Photon Content of the Proton. *JHEP*, 12:046, arXiv:1708.01256.
- [Martin et al., 2005] Martin, A. D., Roberts, R. G., Stirling, W. J., and Thorne, R. S. (2005). Parton distributions incorporating QED contributions. *Eur. Phys. J.*, C39:155, hep-ph/0411040.
- [Nocera et al., 2014] Nocera, E. R., Ball, R. D., Forte, S., Ridolfi, G., and Rojo, J. (2014). A first unbiased global determination of polarized PDFs and their uncertainties. *Nucl. Phys.*, B887:276, arXiv:1406.5539.
- [Patrignani et al., 2016] Patrignani, C. et al. (2016). Review of Particle Physics. *Chin. Phys.*, C40(10):100001.
- [Paukkunen and Zurita, 2014] Paukkunen, H. and Zurita, P. (2014). PDF reweighting in the Hessian matrix approach. *JHEP*, 12:100, arXiv:1402.6623.
- [Politzer, 1973] Politzer, H. D. (1973). Reliable Perturbative Results for Strong Interactions? *Phys. Rev. Lett.*, 30:1346–1349. [,274(1973)].
- [Pumplin et al., 2001] Pumplin, J. et al. (2001). Uncertainties of predictions from parton distribution functions. 2. The Hessian method. *Phys. Rev.*, D65:014013, hep-ph/0101032.
- [Rojo et al., 2015] Rojo, J. et al. (2015). The PDF4LHC report on PDFs and LHC data: Results from Run I and preparation for Run II. *J. Phys.*, G42:103103, arXiv:1507.00556.
- [Schmidt et al., 2016] Schmidt, C., Pumplin, J., Stump, D., and Yuan, C. P. (2016). CT14QED parton distribution functions from isolated photon production in deep inelastic scattering. *Phys. Rev.*, D93(11):114015, arXiv:1509.02905.
- [Vincent et al., 2017] Vincent, A. C., Argüelles, C. A., and Kheirandish, A. (2017). High-energy neutrino attenuation in the Earth and its associated uncertainties. *JCAP*, 1711(11):012, arXiv:1706.09895. [JCAP1711.012(2017)].
- [Wang et al., 2018] Wang, B.-T., Hobbs, T. J., Doyle, S., Gao, J., Hou, T.-J., Nadolsky, P. M., and Olness, F. I. (2018). Mapping the sensitivity of hadronic experiments to nucleon structure. *Phys. Rev.*, D98(9):094030, arXiv:1803.02777.
- [Wobisch et al., 2011] Wobisch, M., Britzger, D., Kluge, T., Rabbertz, K., and Stober, F. (2011). Theory-Data Comparisons for Jet Measurements in Hadron-Induced Processes. arXiv:1109.1310.
- [Zenaiev et al., 2015] Zenaiev, O. et al. (2015). Impact of heavy-flavour production cross sections measured by the LHCb experiment on parton distribution functions at low x. *Eur. Phys. J.*, C75(8):396, arXiv:1503.04581.
- [Zweig, 1964] Zweig, G. (1964). An SU(3) model for strong interaction symmetry and its breaking. Version 2. In Lichtenberg, D. and Rosen, S. P., editors, *DEVELOPMENTS IN THE QUARK THEORY OF HADRONS. VOL. 1. 1964 - 1978*, pages 22–101.