

# Proton structure at the precision frontier

S. Amoroso,<sup>1</sup> A. Apyan,<sup>2</sup> N. Armesto,<sup>3,\*</sup> R. D. Ball,<sup>4,\*</sup> V. Bertone,<sup>5,\*</sup> C. Bissolotti,<sup>6,\*</sup> J. Blümlein,<sup>1</sup>  
 R. Boughezal,<sup>6,\*</sup> G. Bozzi,<sup>7</sup> D. Britzger,<sup>8,\*</sup> A. Buckley,<sup>9,\*</sup> A. Candido,<sup>10,\*</sup> S. Carrazza,<sup>10,\*</sup>  
 F. G. Celiberto,<sup>11,12,13,\*</sup> S. Cerci,<sup>14</sup> G. Chachamis,<sup>15</sup> A. M. Cooper-Sarkar,<sup>16,\*</sup> A. Courtoy,<sup>17,\*</sup> T. Cridge,<sup>18,\*</sup> J.  
 M. Cruz-Martinez,<sup>10,\*</sup> F. Giuliani,<sup>19,\*</sup> M. Guzzi,<sup>20,\*</sup> C. Gwenlan,<sup>16,\*</sup> L. A. Harland-Lang,<sup>21,\*</sup> F. Hekhorn,<sup>10,\*</sup> T.  
 J. Hobbs,<sup>22,23,\*</sup> S. Hoeche,<sup>22</sup> A. Huss,<sup>19,\*</sup> J. Huston,<sup>24,\*</sup> J. Jalilian-Marian,<sup>25</sup> M. Klein,<sup>26</sup> G. K. Krintiras,<sup>27</sup>  
 H.-W. Lin,<sup>24</sup> C. Loizides,<sup>28</sup> G. Magni,<sup>29,30,\*</sup> B. Malaescu,<sup>31,\*</sup> B. Mistlberger,<sup>32,\*</sup> S. Moch,<sup>33,\*</sup> P.  
 M. Nadolsky,<sup>34,†</sup> E. R. Nocera,<sup>4,\*</sup> F. I. Olness,<sup>34,\*</sup> F. Petriello,<sup>35,6,\*</sup> J. Pires,<sup>15,36,\*</sup> K. Rabbertz,<sup>37,\*</sup>  
 J. Rojo,<sup>29,30,\*</sup> G. Schnell,<sup>38,39,\*</sup> C. Schwan,<sup>40,\*</sup> A. Siódmok,<sup>41,\*</sup> D. E. Soper,<sup>42,\*</sup> M. Sutton,<sup>43,\*</sup>  
 R. S. Thorne,<sup>18,\*</sup> M. Ubiali,<sup>44,†</sup> G. Vita,<sup>32,\*</sup> J. H. Weber,<sup>45,\*</sup> K. Xie,<sup>46,\*</sup> C.-P. Yuan,<sup>24</sup> and B. Zhou<sup>47,\*</sup>

<sup>1</sup>*Deutsches Elektronen-Synchrotron, DESY*

<sup>2</sup>*Brandeis University, Waltham, MA 02453, USA*

<sup>3</sup>*Instituto Galego de Física de Altas Enerxías IGFAE, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Galicia-Spain*

<sup>4</sup>*Higgs Centre, University of Edinburgh, JCMB, KB, Edinburgh EH9 3JZ, Scotland*

<sup>5</sup>*IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France*

<sup>6</sup>*High Energy Physics Division, Argonne National Laboratory, Argonne, IL 60439, USA*

<sup>7</sup>*Dipartimento di Fisica, Università di Cagliari and INFN, Sezione di Cagliari, I-09042 Monserrato (CA), Italy*

<sup>8</sup>*Max-Planck-Institut für Physik, München, Germany*

<sup>9</sup>*School of Physics & Astronomy, University of Glasgow, Glasgow G12 8QQ, Scotland, United Kingdom*

<sup>10</sup>*Dipartimento di Fisica, Università di Milano and INFN, Sezione di Milano, Milano, Italy*

<sup>11</sup>*European Centre for Theoretical Studies in Nuclear Physics and Related Areas (ECT\*), I-38123 Villazzano, Trento, Italy*

<sup>12</sup>*Fondazione Bruno Kessler (FBK), I-38123 Povo, Trento, Italy*

<sup>13</sup>*INFN-TIFPA Trento Institute of Fundamental Physics and Applications, I-38123 Povo, Trento, Italy*

<sup>14</sup>*Adiyaman University, Faculty of Arts and Sciences, Department of Physics, 02040 Adiyaman, Turkey*

<sup>15</sup>*Laboratório de Instrumentação e Física Experimental de Partículas (LIP), P-1649-003 Lisboa, Portugal*

<sup>16</sup>*Department of Physics, University of Oxford, UK*

<sup>17</sup>*Instituto de Física, Universidad Nacional Autónoma de México, Ciudad de México, Mexico*

<sup>18</sup>*Department of Physics and Astronomy, University College London, London, WC1E 6BT, UK*

<sup>19</sup>*CERN, EP Department, CH-1211 Geneva 23, Switzerland*

<sup>20</sup>*Kennesaw State University, Kennesaw, GA 30144, USA*

<sup>21</sup>*Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Oxford, OX1 3PU*

<sup>22</sup>*Fermi National Accelerator Laboratory, Batavia, IL 60510, USA*

<sup>23</sup>*Department of Physics, Illinois Institute of Technology, Chicago, IL 60616, USA*

<sup>24</sup>*Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824 US*

<sup>25</sup>*Baruch College, City University of New York, NY, USA*

<sup>26</sup>*Department of Physics, University of Liverpool, UK*

<sup>27</sup>*Department of Physics and Astronomy, University of Kansas, Lawrence, KS 66045*

<sup>28</sup>*ORNL, Physics Division, Oak Ridge, TN, USA*

<sup>29</sup>*Department of Physics and Astronomy, Vrije Universiteit, NL-1081 HV Amsterdam*

<sup>30</sup>*Nikhef Theory Group, Science Park 105, 1098 XG Amsterdam, The Netherlands*

<sup>31</sup>*LPNHE, Sorbonne Université, Université de Paris, CNRS/IN2P3, Paris, France*

<sup>32</sup>*SLAC National Accelerator Laboratory, Stanford University, Stanford, CA 94039, USA*

<sup>33</sup>*II. Institute for Theoretical Physics, Hamburg University, Luruper Chaussee 149, D-22761 Hamburg, Germany*

<sup>34</sup>*Department of Physics, Southern Methodist University, Dallas, TX 75275-0175, USA*

<sup>35</sup>*Department of Physics & Astronomy, Northwestern University, Evanston, IL 60208, USA*

<sup>36</sup>*Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal*

<sup>37</sup>*KIT, Karlsruhe, Germany*

<sup>38</sup>*Department of Physics, University of the Basque Country UPV/EHU, 48080 Bilbao*

<sup>39</sup>*IKERBASQUE, Basque Foundation for Science, 48013 Bilbao*

<sup>40</sup>*Universität Würzburg, Institut für Theoretische Physik und Astrophysik, Emil-Hilb-Weg 22, 97074 Würzburg, Germany*

<sup>41</sup>*Jagiellonian University, 31-007 Kraków, Poland*

<sup>42</sup>*Department of Physics, University of Oregon, Eugene, OR 97401, USA*

<sup>43</sup>*Department of Physics and Astronomy, University of Sussex, Brighton; United Kingdom*

<sup>44</sup>*DAMTP, University of Cambridge, Cambridge, CB3 0WA, United Kingdom*

<sup>45</sup>*Humboldt-Universität zu Berlin, D-12489 Berlin, Germany*

<sup>46</sup>*Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA 15260, USA*

<sup>47</sup>*Department of Physics and Astronomy, Johns Hopkins University, Baltimore, Maryland 21218, USA*

(Dated: March 24, 2022)

---

\* Leading author

† Leading author; Corresponding author; [nadolsky@smu.edu](mailto:nadolsky@smu.edu)

‡ Leading author; Corresponding author; [M.Ubiali@damtp.cam.ac.uk](mailto:M.Ubiali@damtp.cam.ac.uk)

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

# Proton structure at the precision frontier

## ABSTRACT

An overwhelming number of theoretical predictions for hadron colliders require parton distribution functions (PDFs), which are an important ingredient of theory infrastructure for the next generation of high-energy experiments. This whitepaper summarizes the status and future prospects for determination of high-precision PDFs applicable in a wide range of energies and experiments, in particular in precision tests of the Standard Model and in new physics searches at the high-luminosity Large Hadron Collider and Electron-Ion Collider. We discuss the envisioned advancements in experimental measurements, QCD theory, global analysis methodology, and computing that are necessary to bring unpolarized PDFs in the nucleon to the N<sup>2</sup>LO and N<sup>3</sup>LO accuracy in the QCD coupling strength. Special attention is given to the new tasks that emerge in the era of the precision PDF analysis, such as those focusing on the robust control of systematic factors both in experimental measurements and theoretical computations. Various synergies between experimental and theoretical studies of the hadron structure are explored, including opportunities for studying PDFs for nuclear and meson targets, PDFs with electroweak contributions or dependence on the transverse momentum, for incisive comparisons between phenomenological models for the PDFs and computations on discrete lattice, and for cross-fertilization with machine learning/AI approaches.

**Submitted to the US Community Study on the Future of Particle Physics (Snowmass 2021)**

## CONTENTS

I. Introduction	1
A. PDF analyses as a part of HEP theory infrastructure	1
B. Exploring PDFs in Snowmass community planning studies	2
C. New frontiers in PDF analyses in the HL-LHC era	2
D. Organization of the whitepaper	5
II. Modern PDFs and their applications	6
A. Comparisons of PDFs	6
B. Applications of PDFs to Higgs physics, BSM searches, SMEFT tests	10
1. PDFs and Higgs physics	11
2. PDFs and BSM searches	12
3. PDF fits and SMEFT fits interplay	12
III. Experiments	14
A. Measurements and applications of PDFs at the LHC	14
B. PDFs at the Electron-Ion Collider	16
1. Unpolarized Proton PDFs	16
2. Polarized Proton PDFs	17
3. Nuclear PDFs	18
C. The Large Hadron electron Collider (LHeC)	19
D. PDFs for neutrino phenomenology	22
E. Forward (and ultra-high energy) scattering processes	23
IV. Theory	24
A. PDF evolution at N <sup>3</sup> LO	25
B. N <sup>3</sup> LO Cross Sections And Perturbative Uncertainties	25

C. Electroweak Corrections in PDF fits	26
D. PDFs and resummations at extreme momentum fractions	27
1. Large $x$	27
2. Small $x$	29
E. Theoretical Developments beyond QCD and Electroweak at Fixed Order	29
F. Factorization schemes for event generators	30
V. Methodology	31
A. Experimental systematic uncertainties in PDF fits	31
B. Theoretical uncertainties in PDF fits	34
C. Machine learning/AI connections	37
1. PDF determination as a ML problem	37
2. Simultaneous fits of physics parameters and PDFs	38
3. Other PDF-related ML applications	39
D. Delivery of PDFs; PDF ensemble correlations in critical applications	39
VI. PDFs and the strong coupling from lattice QCD	40
A. Strong Coupling Calculations	40
B. Lattice calculations of PDFs	41
VII. Nuclear and meson PDFs	43
A. Impact on Proton PDFs	44
B. Exploring Nuclear $A$ Dependence	44
C. Collective Properties of QCD	44
D. Extreme Kinematics	45
E. Outlook	45
VIII. Transverse-momentum dependent distributions	45
A. Quark TMDs	46
B. Gluon TMDs	46
C. TMD evolution and matching to collinear PDFs	48
D. Status of unpolarized TMD extractions	49
E. Experimental prospects	50
IX. Computing needs and computing tools	51
A. The LHAPDF library and other user interfaces for PDFs	51
B. Public PDF fitting codes	51
1. <code>xFitter</code> : an Open Source QCD Analysis Framework	52
2. NNPDF: an open-source machine learning framework for global analyses of PDFs	52
C. Fast interfaces for pQCD computation	53
1. The APPLfast project	53
2. The <code>PineAPPL</code> interface	56
X. Benchmarking and combination of global PDF analyses; PDF4LHC21 recommendation	57
XI. Conclusion: precision PDFs in the United States	60
Acknowledgements	61
References	62

## I. INTRODUCTION

*Leading authors: M. Ubiali, P. Nadolsky*

Precision phenomenology at hadron colliders relies upon an accurate estimate of the uncertainty in Standard Model (SM) predictions. An overwhelming number of theoretical predictions for hadron colliders require parton distribution functions (PDFs) [1–11], the nonperturbative functions quantifying probabilities for finding quarks and gluons in hadrons in high-energy scattering processes.

In the decade since the Snowmass’2013 study, we witnessed a revolution in computing hard scattering cross sections in perturbative QCD to a high accuracy, achieved by including radiative contributions up to the second and third order in the strong coupling constant (N2LO and N3LO, respectively). A similar progress in understanding PDFs beyond the current level is critical for realizing the physics programs of the high-energy run (Run III) of the Large Hadron Collider (LHC) and of the high-luminosity runs (HL-LHC). Limitations in the knowledge of the PDFs constrain the accuracy of measurements of the Higgs boson couplings and electroweak parameters in key channels at the HL-LHC [12, 13]. By knowing the PDFs for the gluon and other quark flavors approximately to 1-2% accuracy, one greatly reduces the total uncertainties on the Higgs couplings in gluon-gluon fusion and electroweak boson fusion. The energy reach in searches for very heavy new particles at the HL-LHC can be extended to higher masses by better knowing the PDFs at the largest momentum fractions,  $x > 0.1$ , and by pinning down the flavor composition of the partonic sea [14]. As interest grows in hadron scattering at very small partonic momentum fractions,  $x < 10^{-5}$ , at hadron colliders (HL-LHC, LHeC, FCC-hh) as well as in the astrophysics experiments, one must include effects of small- $x$  resummation and saturation in QCD theory and, when warranted, in the PDFs [15].

PDFs contribute to precise measurements of the QCD coupling constant, heavy-quark masses, weak boson mass, and electroweak flavor-mixing parameters. This requires continuous benchmarking and improvements of the theoretical framework, particularly in the perturbative approach adopted for the computation of observables in a PDF fit [16, 17]. As lattice QCD techniques advance in computations of PDFs from the first principles, unpolarized phenomenological PDFs serve as important benchmarks for testing the lattice QCD methods [18, 19]. Namely, precisely determined phenomenological parametrizations for PDFs in the nucleon serve as a reference to validate lattice and nonperturbative QCD calculations. Methods of the precision PDF analysis are increasingly applied to explore the nuclear and meson structure, and they inspire related approaches in the studies of 3-dimensional hadron structure, including dependence on transverse momentum and spin.

In this Snowmass’2022 contribution, we emphasize importance of determination of parton distributions to accuracy that is comparable to those of N2LO/N3LO hard cross sections. Obtaining such accurate PDFs necessitates continued advancements in the areas of quantum field theory, experimental measurements, and statistical methods.

### A. PDF analyses as a part of HEP theory infrastructure

We emphasize that the accurate determination of PDFs constitutes a critical component of theory infrastructure for current and future hadronic experiments, together with the development of Monte-Carlo event generators and multi-loop calculations in QFT [20]. A typical computation of a cross section for a hadron-scattering process includes two parts, a hard cross section quantifying scattering rates for weakly interacting partons, and several functions quantifying probabilities for either finding partons in the initial-state hadrons or for partons fragmenting into final-state hadrons. While the hard cross section can be computed algorithmically using increasingly sophisticated techniques in perturbative quantum chromodynamics (QCD), the long-distance non-perturbative functions are found by other methods, most commonly using a large-scale, or global, analysis of hadronic scattering data. The PDFs are quintessential nonperturbative functions of the latter kind. They are ubiquitously used in hadron collider experiments.

N2LO and N3LO precision of hard cross sections requires equally accurate PDFs. The PDFs generally fall into two classes, *general-purpose* (suitable for the majority of applications) and *specialized* (suitable for certain applications or obtained using special techniques) ones. The PDF determination at the modern precision level is an exciting research area that incorporates advancements in the three frontiers illustrated in Fig. 1. First, new experimental measurements must be performed

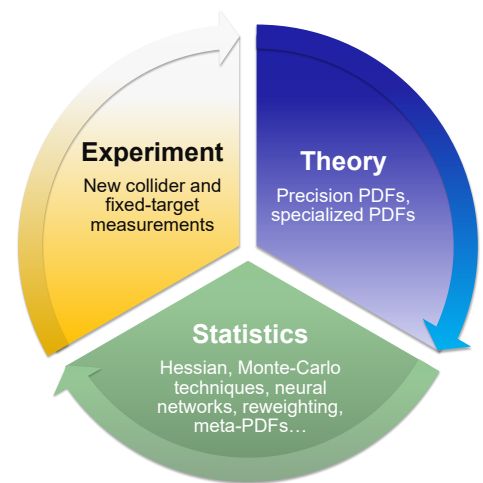


FIG. 1. Three constituent domains of the modern PDF analysis.

with consistent control of accuracy at all stages of the analysis. Second, new multi-loop theoretical cross sections must be computed and implemented in an optimal form in the global PDF fit. Third, the PDFs and their uncertainties consistent with the fitted data sets must be determined in a statistically robust way and delivered in a convenient format to a wide range of users.

There is a significant overlap and cross-talk among these three research areas. Progress toward the next generation of PDFs necessary to achieve the physics goals of the planned experiments, including the HL-LHC and EIC, should therefore happen at the intersection of state-of-the-art particle experiments, quantum field theory, multivariate data science and artificial intelligence, as well as high-performance computing. This research field presents ample opportunities for training of students and postdocs, who develop mathematical and theoretical skills applicable in many areas of science and industry.

## B. Exploring PDFs in Snowmass community planning studies

Already at the first Snowmass DPF Summer Studies in 1980's, theoretical issues and practical methods for determination of PDFs were in the focus of the attention of the HEP community, given the pivotal role of PDFs in predicting QCD processes. It was realized that progress in collider studies is impossible without trusted PDF parametrizations, as was exemplified by the seminal Eichten-Hinchliffe-Lane-Quigg (EHLQ) PDFs [21] published in the run-up to the Superconducting SuperCollider. The Snowmass community studies in the 1990's and 2000's have stimulated understanding of the hadron structure through increasingly precise experiments at FNAL, CERN, and DESY.

The 2021 Snowmass Community Planning Exercise has drawn a large group of participants to explore multi-faceted aspects of PDFs and their applications through collaborative meetings that took place in the Energy Frontier Topical Groups (EF06, as well as EF05, EF07, and other groups) over nearly two years. Many of these aspects are reviewed in this whitepaper, although the full coverage of all involved aspects would be prohibitively extensive. Table I illustrates the progress that has been made since the previous Snowmass Summer Study completed in 2013. Using the 2013 Working Group Report on Quantum Chromodynamics [22] as a reference, we compare the status of select topics in the 2013 study and in the current study. The main part of the whitepaper details these and many other topics in the order listed in Sec. ID. We also highlight the challenges and tasks that need to be addressed to advance our understanding.

## C. New frontiers in PDF analyses in the HL-LHC era

The bottom part of Table I summarizes some of the new tasks for the PDF analysis that emerge in the era of precision QCD. Several ingredients of the global fits are essential for robust modeling of the proton structure. Solutions for PDFs must reflect all allowed variations associated with statistical, systematical errors in the experiments, as well as with relevant error correlations. Needless to say, the most precise N2LO or even N3LO theoretical cross sections should be preferably used, when possible, as a prerequisite for achieving the highest accuracy. However, accuracy of the theoretical predictions used in the fit also depends on the other factors and must be properly estimated. At the same time, given the complexity of N2LO/N3LO calculations, their fast approximate implementations (such as *fast NNLO* interfaces) must be developed to allow efficient observable computations in the PDF analyses. Control of experimental and theoretical uncertainties requires, in particular, to either fit the experiments that are minimally affected by the unknown factors (for example, to include cross sections only on proton, rather than on nuclear targets to minimize the associated uncertainties in the most precise determinations), or to estimate the associated uncertainty of these unknown factors in the fit. Finally, the published range of solutions for the PDFs must account for acceptable variations in methodology, which encompasses such components as the functional forms adopted to parametrize PDFs at some initial energy scale, propagation of experimental uncertainties into the error associated with the fitted functions, the diverse statistical inference techniques, as well as implementation of physical constraints on the PDFs, such as QCD sum rules, positivity of physical observables, and integrability of relevant PDF combinations. Methodological advancements should also include development of practical standards for the delivery of PDFs to a wide range of users. The format of the PDF delivery must optimize for accuracy, versatility, and speed across a broad range of applications – a highly non-trivial task, given the ubiquity of the PDF uses by both experimentalists and theorists. The PDF4LHC working group [23] leads the development of such standards and delivery formats for the LHC community. In particular, the recently released 2021 recommendation of the PDF4LHC working group (PDF4LHC21 [24]) supersedes the previous recommendation issued in 2015 [16]. The PDF4LHC21 recommendation document stipulates guidelines for applications of PDFs and computation of PDF uncertainties at the LHC. With this document, the PDF4LHC working group also distributes combined N2LO PDF4LHC21 error sets (available in the LHAPDF library [25]) to streamline computations with PDFs across typical LHC studies, such as searches for new

## PDF-related topics in Snowmass'13 [arXiv:1310.5189] and '21 studies

Topic	Status, 2013	Status, 2022
Achieved accuracy of PDFs	N2LO for evolution, DIS and vector boson production	N2LO for all key processes; N3LO for some processes
PDFs with NLO EW contributions	MSTW'04 QED, NNPDF2.3 QED	LuXQED and other photon PDFs from several groups; PDFs with leptons and massive bosons
PDFs with resummations	Small $x$ (in progress)	Small- $x$ and threshold resummations implemented in several PDF sets
Available LHC processes to determine nucleon PDFs	$W/Z$ , single-incl. jet, high- $p_T$ $Z$ , $t\bar{t}$ , $W + c$ production at 7 and 8 TeV	+ $t\bar{t}$ , single-top, dijet, $\gamma/W/Z$ +jet, low- $Q$ Drell Yan pairs, ... at 7, 8, 13 TeV
Near-future experiments to probe PDFs	LHC Run-2 DIS: LHeC	LHC Run-3 DIS: EIC, LHeC, ...
Benchmarking of PDFs for the LHC	PDF4LHC'2015 recommendation in preparation	PDF4LHC'21 recommendation issued
Precision analysis of specialized PDFs		Nuclear, meson, transverse-momentum dependent PDFs

### NEW TASKS in the HL-LHC ERA:

Obtain complete N2LO and N3LO predictions for PDF-sensitive processes	Improve models for correlated systematic errors	Find ways to constrain large- $x$ PDFs without relying on nuclear targets
Develop and benchmark fast N2LO interfaces	Estimate N2LO theory uncertainties	New methods to combine PDF ensembles, estimate PDF uncertainties, deliver PDFs for applications

TABLE I. Top part: Some of the PDF-focused topics explored in Snowmass'2013 [22] and '2021 studies. Bottom part: a selection of new critical tasks for the development of a new generation of PDFs that achieve the objectives of the physics program at the high-luminosity LHC.

physics or theoretical simulations. However, comparisons to individual PDF ensembles from the groups, rather than combined ones, remain necessary in the most precise measurements, such as tests of electroweak precision symmetry breaking and Higgs boson physics.

The rest of the whitepaper discusses all these critical tasks of the precision PDF era in more detail. We wish to highlight some of the pertinent issues here.

Recent PDF analyses indicate that the LHC data is increasingly crucial in pinning down the parton densities, and its constraining power will become even more crucial in the HL-LHC run [26]. At the same time, new experiments on the deeply-inelastic scattering (DIS), in particular, at the Electron-Ion Collider planned at BNL in the USA, may be at least as instrumental as the LHC, and in some important cases more instrumental, in constraining the relevant PDF combinations [27]. Even more precise measurements of the PDFs in DIS may be obtained at the Large Hadron-Electron Collider (LHeC [28]) and Muon-Ion Collider (MuIC [29]).

To elevate the accuracy of PDFs in the next decade, it is critical that new experiments and theory calculations implement consistent error control at all stages, from experimental measurements to the distribution of final PDFs. In particular, while there is a reasonable overall agreement between the various experiments in the recent PDF fits [6, 7, 9, 10] in terms of their preferences for the PDFs, detailed testing with several methods reveals some disagreements (tensions) among the most precise experiments. The strength of these disagreements is about the



same in the three recent global fits.<sup>1</sup> The disagreements are sometimes stronger than would be expected simply based on random differences between theory and data [7, 11, 30]. Furthermore, the 2021 comparisons of PDF sets by the PDF4LHC working group [24] suggest that the differences among the global PDF ensembles have increased in some cases compared to the PDF4LHC15 combination [16]. The fitting groups regularly perform thorough benchmark studies [24] to understand the reasons for such behavior. In the course of such exercises, the various groups observe good agreement among their theoretical predictions for the most critical data sets when using the same PDF ensemble as the input. At the same time, when fitting the same data using freely varied PDF parametrizations, the groups arrive at mutually consistent, yet not exactly identical best-fit PDF parametrizations and especially PDF uncertainties. These exercises rule out “trivial” causes for the disagreements among the groups/data sets, such as an incorrect theoretical calculation or improper implementation of an experimental data set. Rather, sometimes the methodological differences, for example due to the fitting techniques, treatment of systematic uncertainties in the data sets, PDF functional forms, or the definition of the PDF errors, can be as large or even larger than the PDF uncertainties from the propagation of experimental errors. Increasing the precision of future PDFs must address such issues. Section X summarizes the ongoing efforts in this direction.

Assuming the possible tensions among the fitted data can be eliminated, as otherwise they may weaken the current HL-LHC projections, there is a hope to arrive at a situation in which, after years of trying to reduce PDF uncertainties, the parton luminosity uncertainties goes down to about 1% in the central rapidity region and for QCD scales around the  $Z$  pole. Nominal uncertainties may go down to 0.3-0.5% within a decade, provided we obtain consistent constraints from the near-future experiments. Can we really trust PDFs to that level of precision?

In such a situation, the precision versus accuracy challenge becomes crucial. In some cases, when a new PDF analysis including new data is released by a PDF-fitting collaboration, shifts from the previous to the new PDF set may be larger than the nominal PDF uncertainties. This does not undermine the accuracy of a PDF determination *per se*, as long as the origin of such shifts has been identified, and all aspects of the fit are kept under control. In the other cases, the uncertainties provided by the group may already include an estimated contribution from such behind-the-scene factors. This is to say that the span of the uncertainties may vary among the different PDF sets depending on how such situations are handled.

As far as experimental data are concerned, one of the key challenges has to do with the data sets which, as the luminosity increases, are more and more dominated by correlated systematics. These highly-correlated data sets may destabilize convergence of the fits if small changes in the data covariance matrix lead to dramatic changes in the fully correlated  $\chi^2$  to the data. Studies of covariance matrix stabilisation and of the effects of decorrelating the systematics are ongoing and will become increasingly vital. They require a strong synergy between theorists and experimentalists. See Sect. VD for a detailed discussion.

As far as the fitting methodology is concerned, several aspects are at play. With the traditional fitting technique based on the minimization of the log-likelihood  $\chi^2$ , the functional forms of the assumed PDF parametrizations are an important factor that must be carefully handled. The PDF parametrization must be flexible “just enough” to obtain good description of the data without overfitting. Significant progress has been made since the 2013 Snowmass study to understand the dependence of PDFs and their uncertainties on the parametrization form. Some examples of this progress include a more flexible parametrization introduced in the MSHT’20 study [9], which in particular results in a change in the down quark valence PDFs compared to the previous fits; a cross-validation test proposed to determine the optimal number of parameters for a given PDF parametrization form [30], similarly in its spirit to the cross-validation condition that prevents over-training of neural networks; a study of 250+ alternative functional forms for the PDFs to determine the component of the PDF uncertainty due to the parametrization in the CT18 analysis [7].

Various components of the fitting methodologies undergo continuous improvements and are subjected to increasingly incisive tests. Such tests subject the fits to closer scrutiny than in the past. Statistical closure tests [31] may become crucial for the modern PDF sets – they are already used to test the robustness of the latest NNPDF sets since NNPDF3.0 [32]. The idea of a closure test is that the PDFs determined from pseudodata generated from a known underlying law must correctly reproduce the statistical distributions expected on the basis of the assumed experimental uncertainties. While the closure test validates the performance of the fitting methodology with the idealized pseudodata, different kinds of tests have been developed for validation with the real-life data sets that are not perfectly consistent. This is the idea behind the strong goodness-of-fit tests that were developed in [30] and applied in the CT18 analysis [7]. The strong goodness-of-fit criteria demand internal consistency of the probability distribution in a global fit, in addition to requiring an excellent  $\chi^2$  describing the overall quality of the fit.

Another crucial element for the future progress is the theory framework. A large amount of work has been devoted to it in the latest analyses, focusing on the implementation of new theoretical calculations and techniques for estimating

---

<sup>1</sup> For example, the  $\chi^2/N_{pt}$  values for the LHC data sets and for all data sets tend to be elevated, as compared to the statistics expectation, in a similar fashion in the CT18, MSHT’20, and NNPDF3.1.1 NNLO global fits, cf. Tables 2.1-2.3 in [24].



theoretical uncertainties on the PDFs. While the theory error associated with the perturbative truncation of the theoretical predictions for the fitted processes was believed to be generally less important than the experimental uncertainties, it becomes significant at the present level of precision and must be taken into account. The effort towards the determination of theory uncertainties in fixed-order PDF fits (discussed in Sect. VB) and the multi-pronged work towards N3LO PDFs (discussed in Sect. IVA) will be paramount in the next few years.

Other sources of uncertainties will become crucial in the future. For example, as even more high-energy data from the LHC are included in PDF fits, the tails of the distributions that are used in PDF determination are potentially affected by new physics effects. To make sure that new physics is not absorbed or “fitted away” in the PDFs, one would either have to exclude this data, thus losing potentially important constraints, or carefully disentangle the Standard Model and new physics effects. More details are provided in Sect.IIB3.

If the advancements along the described directions are realized, the HL-LHC projections [33] will be extremely encouraging, with a foreseen reduction of PDF uncertainties by a factor of 2 to 3. Given the scope of the outstanding questions, this progress will require a broad effort from the HEP community to maintain elevated standards at all stages of the experimental measurements, theoretical computations, and global PDF fits themselves. Accomplishing this goal depends on a dedicated collaboration among experimentalists and theorists. Clearly the precision physics frontier opens up new fascinating challenges also for the exploration of hadron structure.

#### D. Organization of the whitepaper

Following this introduction, Section II compares the latest PDF parametrizations and partonic luminosities from various groups. It also discusses predictions for benchmark LHC measurements and applications of PDFs in studies of electroweak symmetry breaking, searches for new physics, and combined fits of the parameters in the Standard Model and its effective field theory (EFT) extensions.

Section III summarizes some of the main applications of PDFs in the experimental analyses. It reviews promising scattering processes at the LHC that can provide further constraints on the PDFs. Then, Section IIIB reviews prospects for obtaining incisive constraints on the unpolarized, spin-dependent, and nuclear PDFs at the planned Electron-Ion Collider at BNL. The potential for determination of PDFs at the Large Hadron-Electron Collider is explored in Sec. IIIC. Sections IIID and IIIE explore connections of the PDFs with the neutrino-scattering and forward-scattering experiments.

Section IV, dedicated to theoretical aspects of the PDF analyses, begins with a review of the progress toward achieving PDF evolution and computing hard cross sections at N3LO accuracy in Sections IVA and IVB, followed by a discussion of electroweak radiative contributions for the PDF fits in Section IV C. The role of all-order resummations at very small and very large partonic momentum fractions is addressed in Sec. IV D. It ends with a list of theoretical developments needed beyond fixed-order QCD and EW and with a discussion on the factorization schemes needed for event generators.

Section V addresses methodological aspects of global fits, starting with the pivotal role of the models for experimental systematic uncertainties for the future PDF fits, and proceeding to the various approaches for the estimate of theoretical uncertainties on the PDFs, machine learning applications in the context of PDF determinations, delivery of PDFs, and discussing the combination of PDF uncertainties without or with data-driven correlations.

Section VI presents an overview of the calculations of the QCD coupling strength and PDFs on the lattice – the rapidly growing field that holds the promise to predict the hadron structure, including the spin-independent and other types of the PDFs, starting from the first QCD principles. This is followed by a brief, yet informative summary of prospects for determination of nuclear and meson PDFs in Section VII, and then by an overview of the planned studies of transverse-momentum dependent PDFs in Section VIII.

Numerical computations constitute the essential part of the PDF analyses. Section IX reviews publicly available computer programs and resources to perform PDF fits and use PDFs in HEP applications. In this section, we discuss the LHAPDF library providing PDF parametrizations, the xFITTER and NNPDF open-source codes for global fits, as well as APPLGRID, FAST(N)NLO, and PINEAPPL interfaces for fast computations of QCD and EW radiative contributions.

Section X summarizes recent studies by the PDF4LHC working group to benchmark and combine PDF ensembles for LHC applications. This section also reviews the latest recommendation [24] from the PDF4LHC working group on using the NNLO PDFs in various LHC contests.

Conclusions for the whitepaper are provided in Section XI.

## II. MODERN PDFS AND THEIR APPLICATIONS

In this section we start by comparing the most recent PDF determinations presented by several PDF fitting collaborations. We then turn on discussing modern applications of PDFs, particularly focussing on the role of PDFs in Higgs physics, BSM searches and SMEFT global analyses.

### A. Comparisons of PDFs

*Leading authors: T. Cridge, F. Giuli, J. Huston, M. Ubiali, A. M. Cooper-Sarkar, K. Xie*

In this section we compare the most recent PDF sets released by modern PDF fitting collaborations on the market: the NNPDF4.0 set [10], the default CT18 set [7], the MSHT20 set [9], the ABMP16 set with  $\alpha_s(M_Z) = 0.118$  [5] and the ATLASpdf21 set [11].

We start with a comparison at the level of the PDFs themselves in Fig. 2, before looking at parton luminosities and phenomenological predictions. Beginning with the gluon, we see general agreement between the different groups over the range  $10^{-4} \lesssim x \lesssim 10^{-1}$  within uncertainties (albeit with the ATLASpdf21 set deviating slightly before this <sup>2</sup>), nonetheless outside of this range differences emerge. In particular the differences at high  $x$  reflect those observed in the gluon-gluon luminosity at high invariant masses, with the different quantities and treatments of high- $x$  gluon data in this region resulting in clear differences in the PDFs. The singlet  $\Sigma$ , representing the sum of all the quarks and antiquarks in the proton (up to 5 flavours), is in greater agreement, consistent within  $1\sigma$  between all the PDF groups, bar ATLASpdf21, until very low  $x$ , albeit with ABMP16 deviating earlier at around  $x \sim 10^{-3}$ . ATLASpdf21 is in agreement within  $2\sigma$  except at very low and very high  $x$ . The total strangeness has undergone notable changes over the past few years with the advent of new LHC precision Drell-Yan data, nonetheless MSHT20 and NNPDF4.0 are in excellent agreement until (very) high  $x$  and are raised relative to CT18 as a result of their inclusion of the ATLAS 7 TeV  $W, Z$  data [34] which is excluded by default from CT18 and is known to raise the strangeness in the intermediate to high  $x$  region. ABMP16 and ATLASpdf21 show differences at high and low  $x$  but also agree in the intermediate  $x$  range, ATLASpdf21 in particular agrees with NNPDF4.0 and MSHT20 very well from  $x \sim 10^{-4}$  to  $x \sim 0.1$ , reflecting the fact they are all constrained significantly by the high precision ATLAS  $W, Z$  data from LHC Run I. Meanwhile, the charm PDF shows substantial differences between NNPDF4.0 and the other groups, which are largely in agreement with one another. This reveals the difference between perturbative charm treatments where the charm PDF is generated purely from perturbative gluon radiation, and NNPDF4.0's fitted charm whereby the charm PDF is fitted [35] and therefore allowed a nonperturbative intrinsic component. This raises the charm of the latter at high  $x$  as seen here. Finally, we compare the valence PDFs, the up valence PDF shows reasonable agreement in the valence region and down to  $x \lesssim 10^{-2}$ , at this point there are some deviations between the groups, with MSHT20 and CT18 (and to a lesser extent ABMP16 and ATLASpdf21) preferring a different shape relative to NNPDF4.0. Such shape differences can arise as small differences in the valence region can impact the less well-constrained regions at low  $x$  through the valence sum rules. The down valence shows good consistency between NNPDF4.0, MSHT20 and ABMP16, however CT18 and ATLASpdf21 prefer a distinctly different shape, although this is consistent between them. For CT18 this reflects their differences in the strangeness in the  $10^{-2} \lesssim x \lesssim 10^{-1}$  region, with a relative reduction in strangeness compensated for by an increased down PDF here and with which there is some degeneracy. The valence sum rule then ensures this results in deficit in the down valence at low  $x$ . Nonetheless, despite several of these differences, the PDFs show broad agreement in the data regions and this results in the general accord in the PDF luminosities described below. However we first also briefly comment on the comparison of the PDF uncertainties shown in Fig. 3. As can be seen, differences in the uncertainty bands tend to be more significant than in the central values and these reflect a variety of factors, including methodologies applied as well as the data sets included. Even so, there is often closer agreement in the data regions. ATLASpdf21 generally have larger uncertainties at low and high  $x$  as a result of their reduced data sets relative to those of the global fitting groups, whilst ABMP16 often have smaller error bands as a result of their use of the  $\Delta\chi^2 = 1$  criterion to define their error bands without any tolerance (or effective tolerance). A full understanding of the origin of the differences between PDFs determined by different groups and their impact on LHC phenomenology is discussed in Section X in the context of the PDF4LHC21 benchmark studies [24, 36].

<sup>2</sup> The ATLASpdf21 applies a cut of  $Q^2 > 10 \text{ GeV}^2$  on HERA data because of doubts about the adequacy of N2LO DGLAP to describe low- $x$  data which fall below this cut in the HERA kinematic regime. However this means that these PDFs are designed for use at higher  $x$ ,  $x \gtrsim 10^{-4}$ . Indeed the deviation of the ATLASpdf21 from the others at lower  $x$  values indicates the region where low- $x$  physics effects may need to be considered.

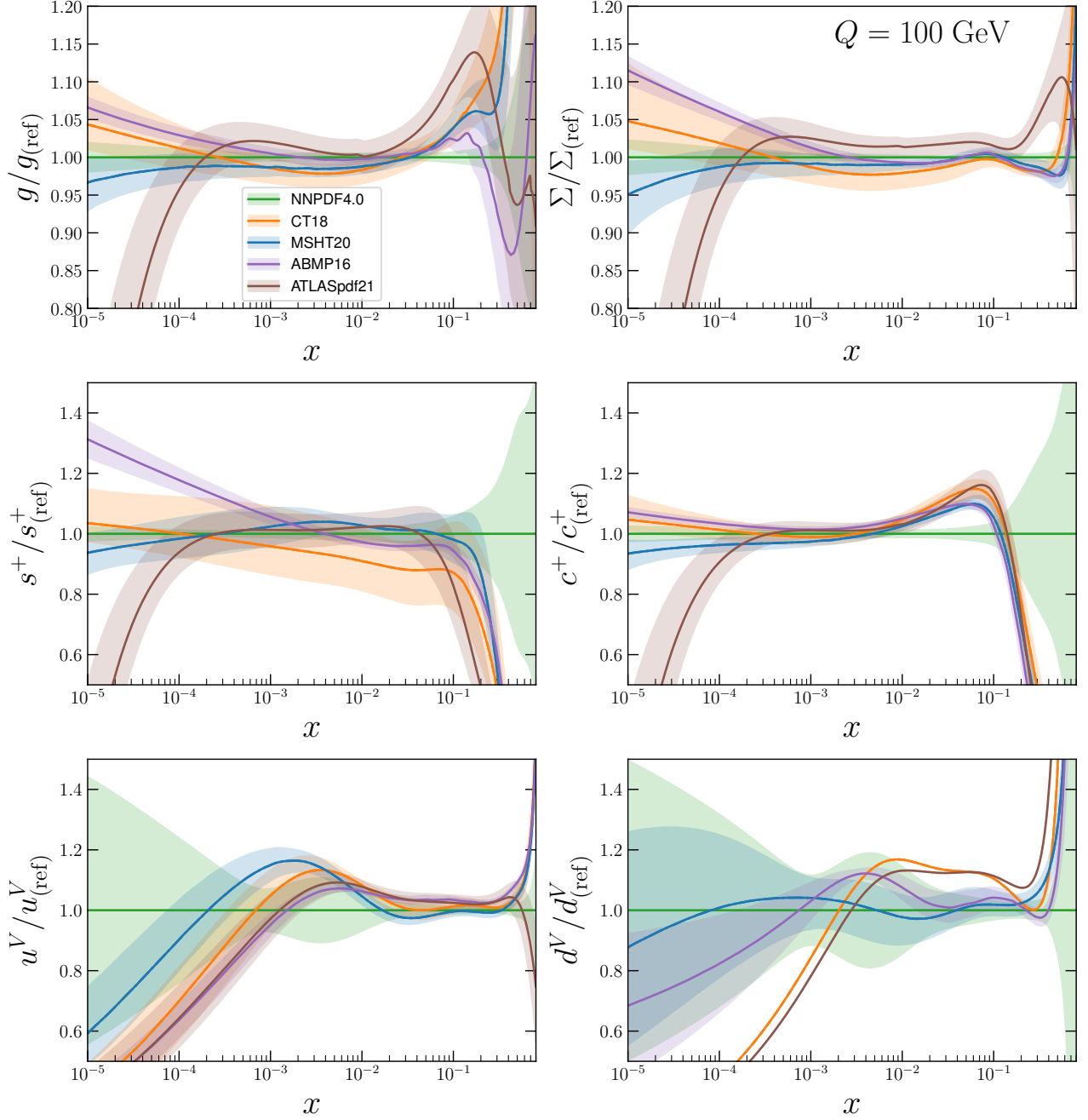


FIG. 2. Comparison of the PDFs at  $Q = 100$  GeV. The PDFs shown are the N2LO sets of NNPDF4.0, CT18, MSHT20, ABMP16 with  $\alpha_s(M_Z) = 0.118$ , and ATLASpdf21. The ratio to the NNPDF4.0 central value and the relative  $1\sigma$  uncertainty are shown for the gluon  $g$ , singlet  $\Sigma$ , total strangeness  $s^+ = s + \bar{s}$ , total charm  $c^+ = c + \bar{c}$ , up valence  $u^V$  and down valence  $d^V$  PDFs.

In Fig. 4 we compare, as a function of the invariant mass  $m_X$ , the parton luminosities at  $\sqrt{s} = 14$  TeV. All sets are N2LO PDF sets. The parton luminosities are defined as [37]:

$$\mathcal{L}_{ij}(m_X^2, \mu_F^2) = \frac{1}{1 + \delta_{ij}} \frac{1}{s} \int_{\tau}^1 \frac{dx}{x} [f_i(x, \mu_F^2) f_j(\tau/x, \mu_F^2) + (i \leftrightarrow j)] \quad (1)$$

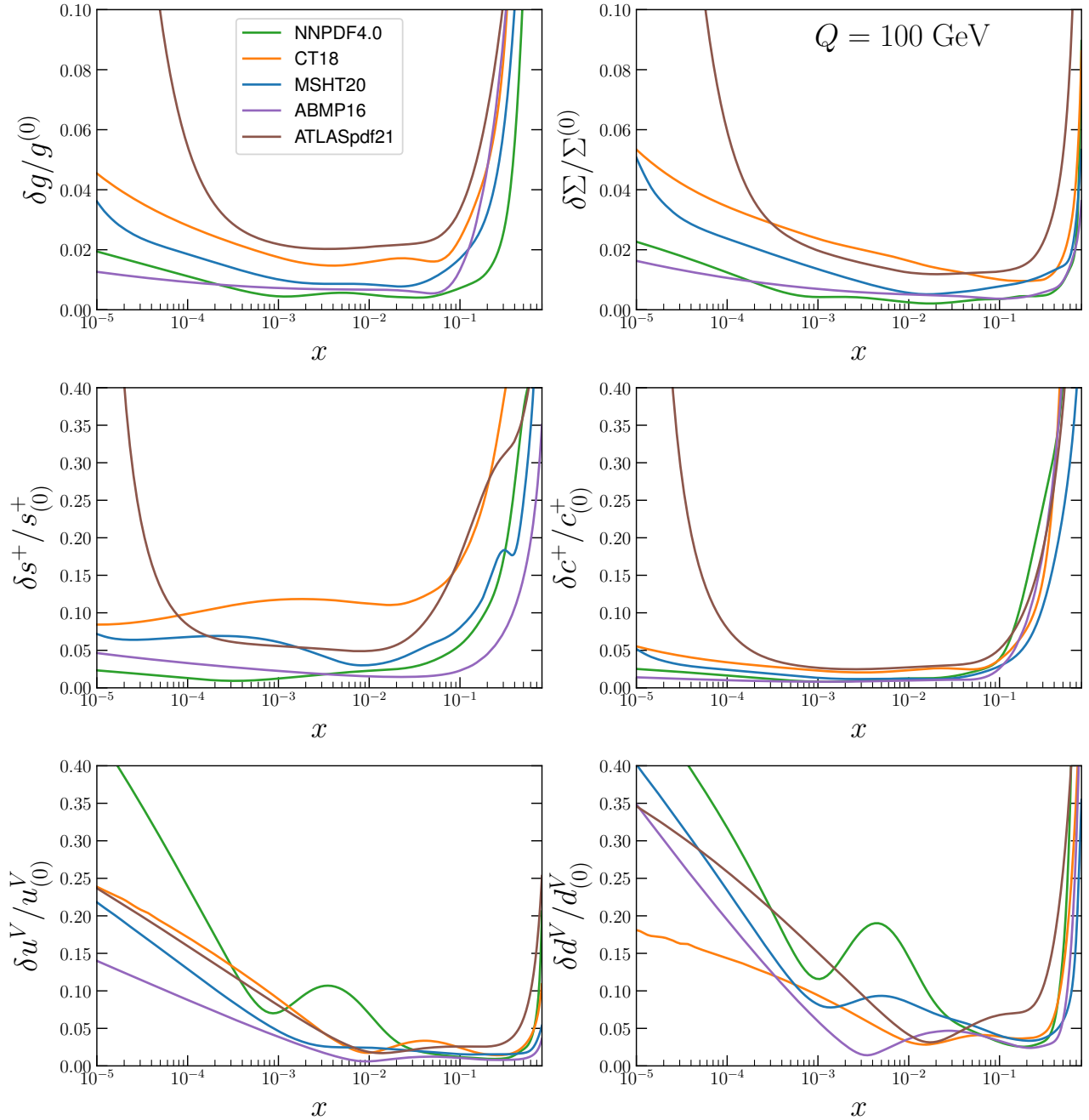


FIG. 3. Comparison of the symmetrized PDF uncertainties at  $Q = 100$  GeV for the gluon  $g$ , singlet  $\Sigma$ , total strangeness  $s^+ = s + \bar{s}$ , total charm  $c^+ = c + \bar{c}$ , up valence  $u^V$  and down valence  $d^V$  PDFs. The PDF sets shown are the N2LO sets of NNPDF4.0, CT18, MSHT20, ABMP16 with  $\alpha_s(M_Z) = 0.118$  and ATLASpdf21.

where  $\tau = m_X^2/s$  and a conventional choice  $\mu_F^2 = m_X^2$ . We have summed over flavors in combinations

$$\mathcal{L}_{q\bar{q}} = \sum_i \mathcal{L}_{q_i\bar{q}_i}, \quad \mathcal{L}_{qq} = \sum_i (\mathcal{L}_{q_i q_i} + \mathcal{L}_{\bar{q}_i \bar{q}_i}), \quad \mathcal{L}_{gq} = \sum_i (\mathcal{L}_{g q_i} + \mathcal{L}_{g \bar{q}_i}). \quad (2)$$

The ratio to the NNPDF4.0 central value and the relative  $1\sigma$  uncertainty are shown for each parton combination. All luminosities agree within uncertainties in the region around  $m_X \sim 100$  GeV, relevant e.g. for Higgs and gauge boson production. The ATLASpdf21 luminosities differ at low scale,  $m_X \lesssim 40$  GeV, because of the cut on low- $x$ ,  $Q^2$  HERA data as already remarked. The quark-quark and quark-antiquark luminosities are otherwise in reasonable

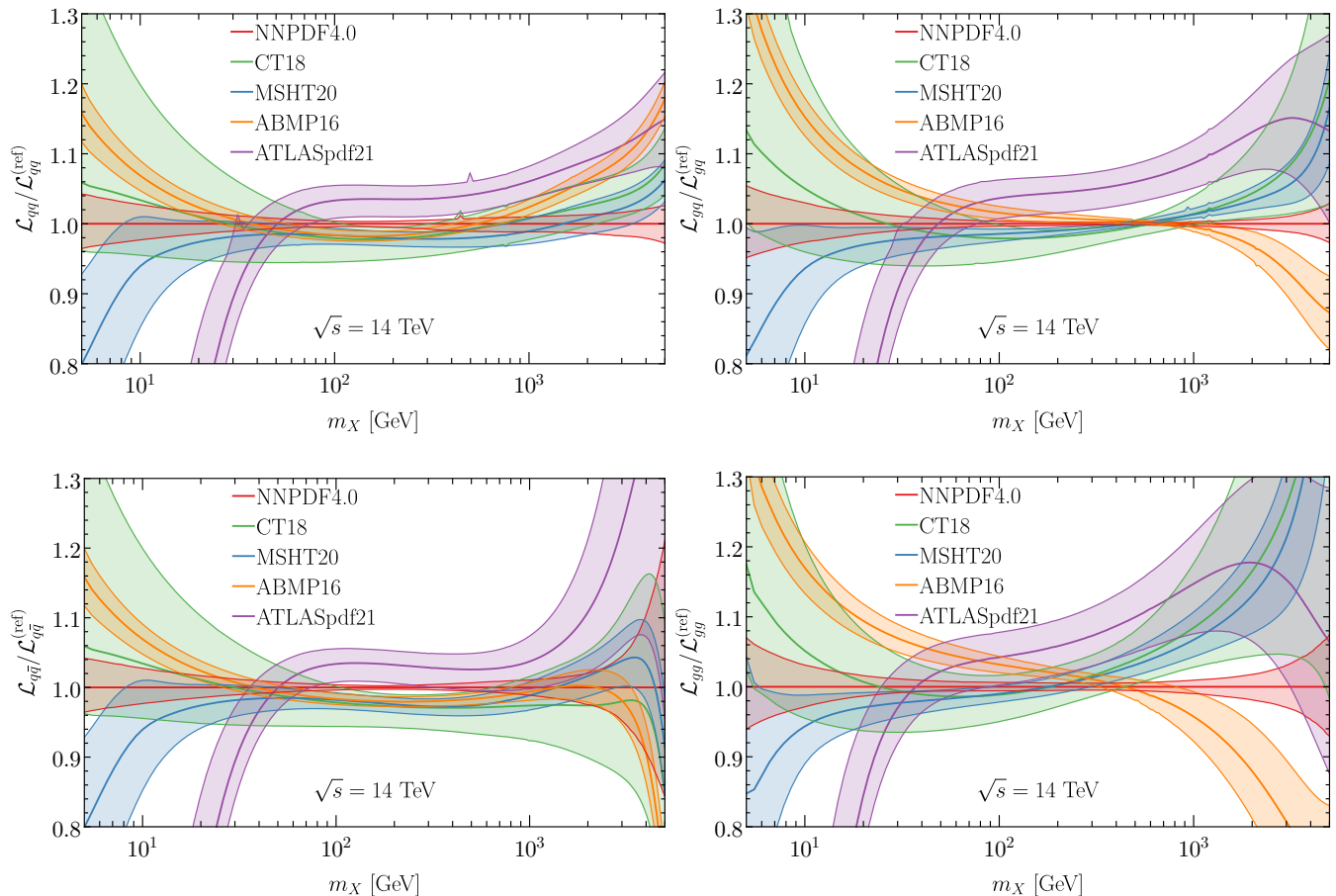


FIG. 4. Comparison, as a function of the invariant mass  $m_X$ , of the parton luminosities at  $\sqrt{s} = 14$  TeV, computed using N2LO NNPDF4.0, CT18, MSHT20, ABMP16 with  $\alpha_s(M_Z) = 0.118$ , and ATLASpdf21. The ratio to the NNPDF4.0 central value and the relative  $1\sigma$  uncertainty are shown for each parton combination.

agreement within  $2\sigma$  over the full mass range. For the gluon sector luminosities (gluon-gluon and gluon-quark), however, further differences are seen at large mass. Specifically, in the high-mass region,  $m_X \sim 1$  TeV the gluon-gluon and gluon-quark luminosities for NNPDF4.0 are rather smaller than MSHT20, CT18 and ATLASpdf21 while larger than ABMP16. These differences are possibly a consequence of both methodology and differences in data included, for example NNPDF4.0 include some data which are sensitive to the high- $x$  gluon that are not used by other groups, such as the di-jet cross-sections at 7 TeV and the  $t\bar{t}$  differential distributions from the LHC Run II. Nonetheless there are additional differences in this region in data inclusion and treatment, discussed in Sect. X. As for the luminosity uncertainties, NNPDF4.0 generally displays the smallest uncertainty in the luminosities, although there are exceptions, with ABMP16 smaller in some regions (such as the gluon-gluon luminosity for low invariant mass). These reflect the uncertainties seen in Fig. 3 where this general pattern also exists, nonetheless all groups, bar ATLASpdf21, show the smallest uncertainties for at least a portion of the  $x$  range across at least one PDF. The differences observed reflect methodological and other differences in approach and are the subject of further studies.

To conclude this section, we assess how the differences at the level of PDFs and parton luminosities displayed in Figs. 2-4 translate into differences in theoretical predictions for LHC cross-sections. These are displayed in Fig. 5, where we present a comparison of the  $2\sigma$  ellipses for pairs of inclusive cross sections among  $W^\pm$ ,  $Z$ ,  $t\bar{t}$ ,  $H$ ,  $t\bar{t}H$  production at the LHC at  $\sqrt{s} = 14$  TeV. The  $W^\pm/Z$  cross sections are defined in the ATLAS 13 TeV fiducial volume [38], while others correspond to the full phase space. The recent PDF4LHC21 combined PDF set [24], described in Section X, and the previous PDF4LHC15 combination [16] are also included in the plots and compared to the  $2\sigma$  ellipses obtained from the PDF sets displayed in Figs. 2-4, with the PDF uncertainties rescaled to the  $2\sigma$  prescription. There is a general agreement between the correlated predictions, with ATLASpdf21 predictions displaying larger uncertainties compared to the other sets and touching the PDF4LHC21  $2\sigma$  boundaries for the  $t\bar{t}$  and  $Z$  ellipses and ABMP16 giving lower predictions for  $H$  and  $t\bar{t}H$  cross sections. Generally, NNPDF4.0 predictions are at the boundary of the MSHT20

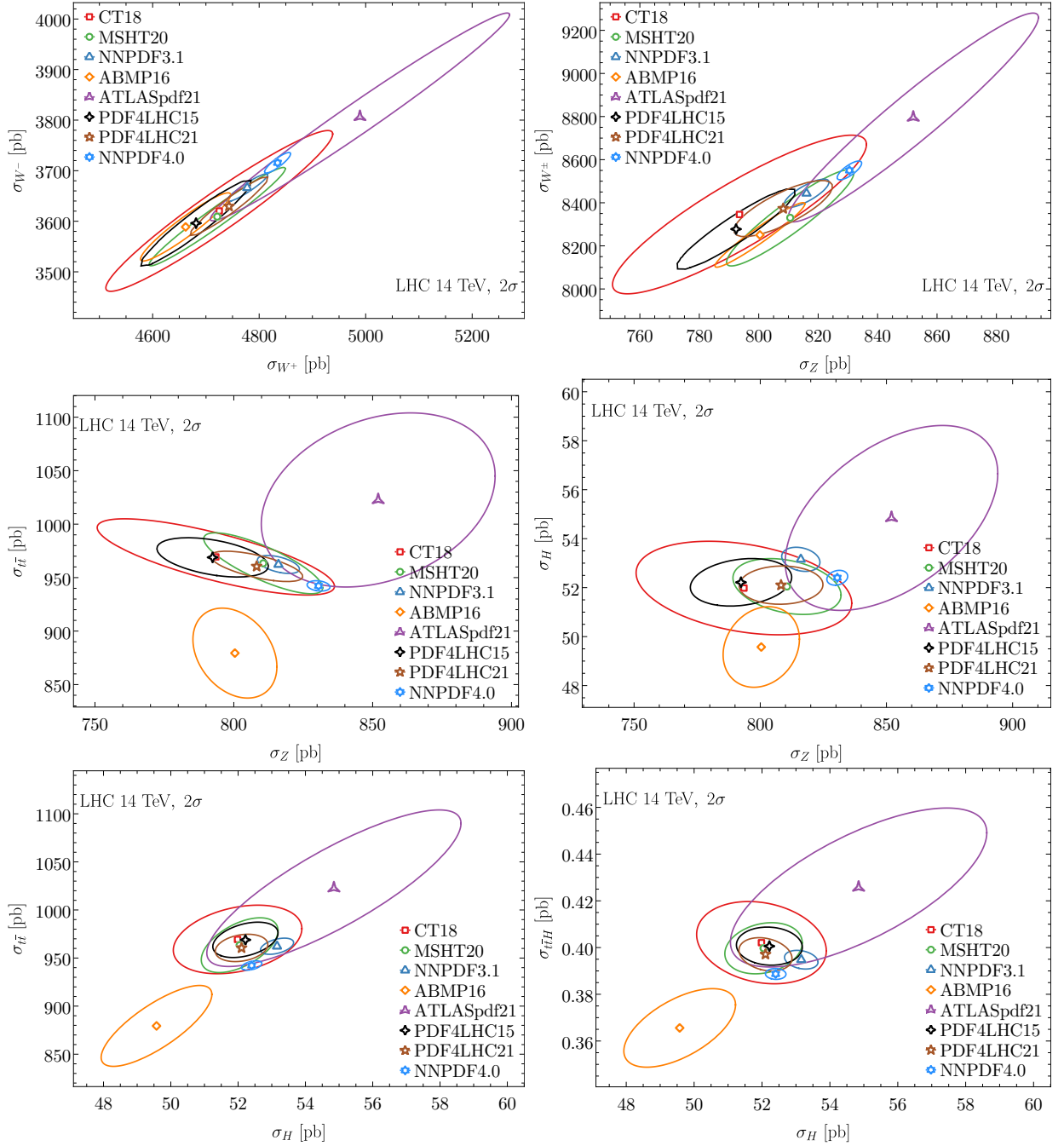


FIG. 5. Comparison between theoretical predictions for the  $2\sigma$  correlation ellipses for pairs of inclusive cross sections among the  $W^\pm, Z, t\bar{t}, H, t\bar{t}H$  production processes at the LHC 14 TeV, comparing the predictions based on PDF4LHC21 [24] with those from the previous combination PDF4LHC15 [16] and the individual NNPDF4.0, CT18, MSHT20, ABMP16 with  $\alpha_s(M_Z) = 0.118$ , and ATLASpdf21 releases.

ellipses, with smaller error bands. MSHT20 are also generally in agreement with CT18, albeit with the latter having notably larger error ellipses.

### B. Applications of PDFs to Higgs physics, BSM searches, SMEFT tests

PDFs are a crucial input at the LHC. Their uncertainty is a key component of theory uncertainties in Higgs physics, a limiting factor in the mass reach of experimental searches for heavy BSM particles and the treatment of



BSM sensitive data in PDF fits makes the interplay between PDFs and SMEFT tests significant. In what follows we briefly discuss each of these applications in turn and refer to a number of studies and new directions within each of these strands.

### 1. PDFs and Higgs physics

*Author: Maria Ubiali*

In the SM, once the Higgs mass  $M_H$  is measured, all other parameters of the Higgs sector, such as the strength of its coupling to fermions and vector bosons and its branching ratios, are uniquely determined [39]. Any deviation of the Higgs couplings with respect to the SM predictions would be a smoking gun for New Physics. Crucially, realising this program requires not only high precision experimental measurements of Higgs boson production and its decay in various channels, but also the calculation of the SM cross sections and decay rates with matching theoretical precision. Despite the progress in the precise determination of PDFs, PDF uncertainty is still one of the largest sources of theoretical uncertainty affecting the predictions for Higgs boson production [39, 40].

In Ref. [26] a study of the impact of HL-LHC pseudo-data for a number of PDF-sensitive processes was performed. Different scenarios are considered, from a conservative one with approximately the same systematics as the corresponding baseline measurements from Run I and a factor of 2 reduction for those from Run II, to an optimistic one with a reduction by a factor 2.5 as compared to Run I (II). It was found that the legacy HL-LHC measurements can reduce the uncertainties in the PDF luminosities by a factor between 2 and 5 in comparison to state-of-the-art fits, depending on the specific flavour combination of the initial state and the invariant mass of the produced final state. As an illustration, on the left panel of Fig. 6 we show a comparison of the PDF uncertainty for Higgs boson production in gluon fusion at  $\sqrt{s} = 14$  TeV and its reduction from predictions obtained with the PDF4LHC15 [16] baseline and the HL-LHC profiled sets in the conservative (scen A) and optimistic (scen C) scenarios.

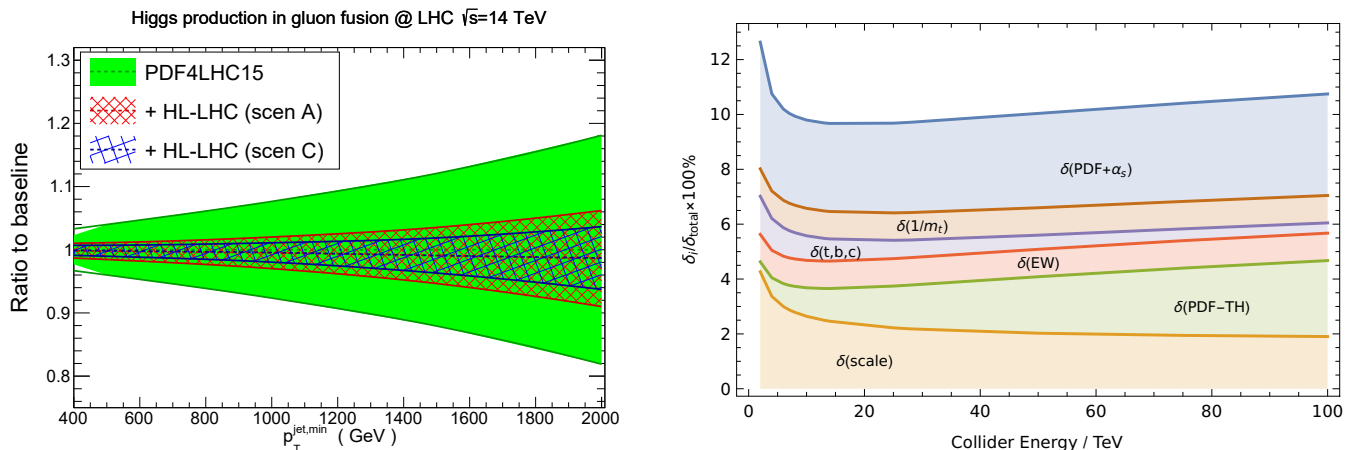


FIG. 6. **Left panel:** Comparison of the predictions for Higgs production via gluon fusion at  $\sqrt{s} = 14$  TeV at between the PDF4LHC15 baseline [16] and the HL-LHC profiled sets in the conservative and optimistic scenarios of Ref. [26]. Results are shown normalised to the central value of PDF4LHC15. Taken from [26], see reference for more details about the calculation. **Right panel:** linear sum of the different sources of relative uncertainties in the calculation of Higgs production via gluon fusion as a function of the collider energy. Each coloured band represents the size of one particular source of uncertainty. In particular, the component  $\delta(\text{PDFs} + \alpha_s)$  corresponds to the uncertainties due to our imprecise knowledge of the strong coupling constant and of PDFs combined in quadrature, while the  $\delta(\text{PDF} - \text{th})$  represents the mismatch in the perturbative order of the PDFs, evaluated at N2LO, and the perturbative QCD cross sections evaluated at N3LO, defined as in Eq. (3). Taken from [12].

However, the effect of the pure PDF uncertainty is not the end of the story. In [12], theoretical predictions for Higgs boson production through gluon fusion at  $pp$  collisions are provided as a function of the collider energy  $\sqrt{s}$ . As it can be observed from the right panel of Fig. 6, the theoretical uncertainty associated with the predictions is split into various components, including the missing higher order uncertainty  $\delta(\text{scale})$  (measured by the usual scale variation procedure) of the N3LO calculation of the  $gg \rightarrow H$  partonic cross section [41, 42]. Electroweak (EW) and approximated mixed QCD-EW corrections as well as effects of finite quark masses are also included in the  $\delta(\text{EW})$  component. Effects due to finite quark masses neglected in the QCD corrections are also accounted for in the  $\delta(1/m_t)$  and  $\delta(t, b, c)$  components. Finally, and most relevant for our discussions are the two components related with PDF

uncertainties. On the one hand, the usual component  $\delta(\text{PDFs} + \alpha_s)$  corresponding to the uncertainties due to our imprecise knowledge of the strong coupling constant  $\alpha_s(M_Z)$  and of the PDFs combined in quadrature. On the other hand, the  $\delta(\text{PDF} - \text{th})$  components, which represents the mismatch in the perturbative order of the PDFs, evaluated at N2LO, and the perturbative QCD cross sections evaluated at N3LO, defined as

$$\delta(\text{PDF-TH}) = \frac{1}{2} \left| \frac{\sigma_{\text{N2LO-PDF}}^{\text{N2LO}} - \sigma_{\text{NLO-PDF}}^{\text{N2LO}}}{\sigma_{\text{N2LO-PDF}}^{\text{N2LO}}} \right|. \quad (3)$$

As one may observe on the right panel of Fig. 6,  $\delta(\text{PDF-TH})$  leads to a significant uncertainty on N3LO cross section predictions, of the order of several percent in the case of Higgs via gluon fusion as well as in the case of other key LHC observables [41, 43, 44] and is comparable to the regular uncertainty associated with our current understanding of PDF themselves. Of course, the prescription of Eq. (3) is a very conservative estimate of the theory uncertainty due to the mismatch between the perturbative order of PDF evolution and partonic cross section, however it points to the need of devising a better procedure of estimating theory uncertainties in the now standard N2LO PDF fits (discussed in Sect. VB) and of moving towards N3LO PDFs (discussed in Sect. IVA).

## 2. PDFs and BSM searches

*Author: Marco Guzzi*

PDFs of the proton are a staple product of collinear factorization in QCD and are a fundamental ingredient for all theory predictions at hadron colliders. Their precise and accurate knowledge is critical not only to scrutinize the EW sector of the Standard Model (SM) and the properties of the Higgs boson, but also to search for New Physics (NP) interactions. NP interactions are currently searched for at the LHC, but are also important for a large variety of physics programs at future facilities (*e.g.*, HL-LHC, Future Circular Collider (FCC), Super proton proton Collider (SppC), Faser $\nu$  [45]) in order to explore beyond the SM (BSM) scenarios. The interplay between global PDF analyses, precision calculations of matrix elements, and BSM physics is therefore crucial to accomplish a wide range of physics goals in the future.

An example of this interplay is the impact of PDF uncertainties in searches for new vector bosons referred to as  $Z$ 's and  $W$ 's from BSM constructions that extend the gauge symmetry group of the SM.  $Z$ 's and  $W$ 's predicted by different models can have a mass that varies from few GeVs or less, to dozens of TeVs, and their fermion interactions share similar features to those of the  $Z$  and  $W$  from the SM. However, details of these interactions depend upon the complexity of the model considered. Models for  $W'/Z'$ s in Drell-Yan resonant dilepton production are currently being scrutinized at the LHC [46, 47]. At high energies,  $W'/Z'$ s can also be produced in association with another SM vector or scalar boson, or in association with a jet or single heavy quark [48, 49]. Current LHC bounds on mass disfavor extra vector bosons lighter than approximately 4-5 TeV. Therefore, BSM searches of  $W'/Z'$ s with larger mass, that can in principle be produced at large rapidity, are progressively more sensitive and are impacted by PDFs at large  $x$  where uncertainties are still large [14]. This impairs our ability to accurately calculate  $W'/Z'$  cross sections and to discriminate between BSM models. Constraining PDFs at large  $x$  is a very challenging task because there are many effects of comparable size that contribute and affect global PDF analyses in this kinematic region, see a related discussion in Sect. IVD 1. Examples of these are nuclear corrections, higher twist contributions, presence of intrinsic heavy-quark components, use of different general mass variable flavor number (GMVFN) schemes, etc. Future precision measurements at the next run of the LHC, at the HL-LHC, and at the high-luminosity EIC, will provide new insight and will be critical to shed light on these open issues. This motivates the search for new strategies to treat large- $x$  dynamics and the selection of new dedicated observables to constrain the large- $x$  region of collinear PDFs in global QCD analyses.

## 3. PDF fits and SMEFT fits interplay

*Leading authors: R. Boughezal, F. Petriello, M. Ubiali*

If the LHC experiments identify one or more significant deviations from the SM predictions, the most promising way to help to characterise their possible origin is via Effective Field Theories (EFTs). Even in the absence of any deviations, EFTs can be used to set lower bounds on the scales of a number of new physics scenarios and to steer the efforts of future searches [50]. Indeed, for a large class of BSM models, physics at energies well below the mass scale  $\Lambda$  of new physics can be parametrized by an EFT, by adding higher dimensional operators to the SM Lagrangian, whose

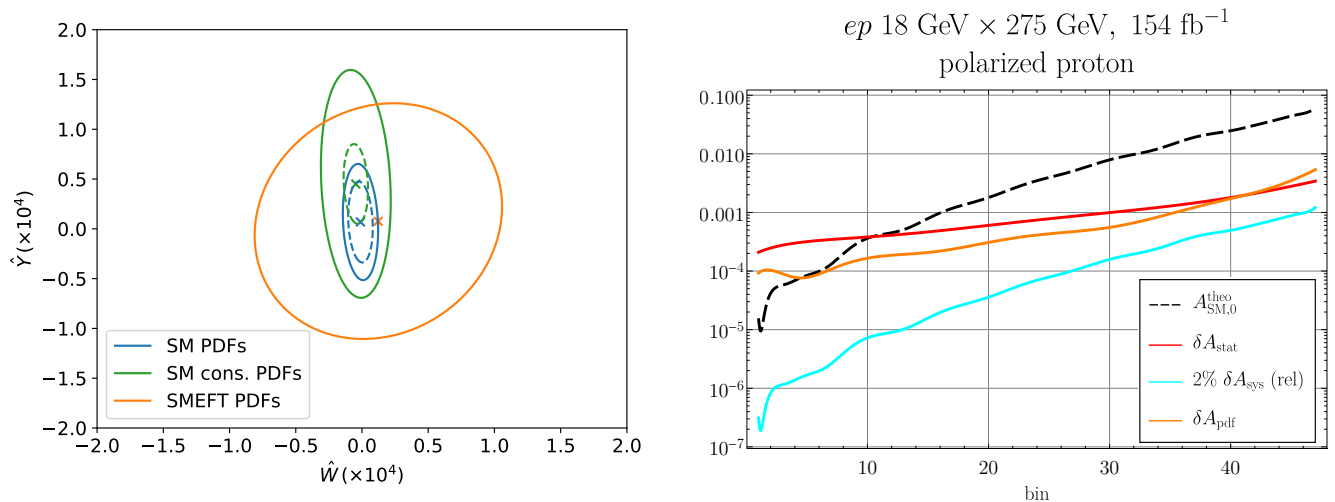


FIG. 7. **Left panel:** The 95% confidence level bounds on the plane of the Wilson coefficients considered in Ref. [52] obtained using either fixed SM PDFs (blue) or conservative SM PDFs that do not include high-energy data (green). PDF uncertainties are included in the solid lines and not included in the dashed lines. The results are compared to those obtained in a simultaneous fit of SMEFT and PDFs, when the PDFs are allowed to vary when varying the values of the Wilson coefficients (orange). **Right panel:** Error components for the polarization asymmetry at a future EIC as a function of bin number, adapted from Ref. [53]. The bins are ordered in the DIS momentum transfer  $Q^2$  and Bjorken- $x$ .

coefficients are suppressed by powers of  $\Lambda$ . Such extensions of the SM Lagrangian determine the effect of physics, that lives well above the energy scale probed by the LHC experiments.

The analysis of BSM effects via an EFT parametrization is a critical and increasingly active research area. A widely adopted EFT expansion is the Standard Model EFT (SMEFT) [51], which is built upon the assumption that all the known particles have the gauge transformation properties predicted by the SM, with their conventional dim-2 and dim-4 interactions being supplemented by new higher-dimensional interactions among all allowed combinations of the SM fields. Such interactions might be generated by massive particles exchanged at the tree-level or circulating in loop diagrams.

Although the proton structure parametrized by PDFs is intrinsically a low-energy quantity and, as such, it should in principle be separable from the high-energy new physics imprints, the complexity of the LHC environment might well intertwine them. Exploiting the full potential of current and future precision measurements at the LHC for indirect BSM searches requires the development of novel data interpretation frameworks that are able to account for hitherto ignored effects, such as the interplay with the PDFs in the high-energy tails of LHC distributions, that can no longer be neglected. Indeed, the very same data sets are being used both to determine the PDFs (assuming SM theoretical predictions) and, independently, to constrain the SMEFT Wilson coefficients (assuming SM PDFs). Given that these LHC processes provide significant information for both PDF and SMEFT fits, it is of paramount importance to ascertain the extent for which eventual BSM signals can be inadvertently reabsorbed into the PDFs, as well as how current bounds on the EFT coefficients are modified within a consistent simultaneous determination together with the PDFs.

Data sets that may contain information on new physics at high scales, such as inclusive jet production, also typically cover a wide dynamic range, both in terms of transverse momentum and of rapidity. If there is a PDF explanation for any variation from the SM prediction that is observed at high  $p_T$ , that explanation has to be universal, *i.e.* it also has to explain distributions at similar  $x$  values, but at lower transverse momentum, regions where new physics is not expected to produce any notable impacts. In this way, the separate rapidity regions serve as a cross-check, both for the PDF determinations themselves and the possible presence of new physics. Care must be taken, however, as tensions between rapidity regions may arise from an imperfect knowledge of the rapidity dependence of the experimental systematic errors as well.

The effects of a simultaneous determination of the Wilson coefficients of the SMEFT and of the proton PDFs has been pioneered in several recent studies performed by both theorists and experimentalists [52, 54–58]. These studies reveal that, while with current DIS and Drell-Yan data the interplay is already non negligible but can be kept under control, once High Luminosity LHC (HL-LHC) data are considered, neglecting the PDF interplay could potentially miss new physics manifestations or misinterpret them. This is illustrated on the left panel of Fig. 7, in which it can be observed that including high-mass data at the LHC both in a fit of PDFs and in a fit of SMEFT coefficients and

neglecting the interplay between them could result in a significant underestimate of the uncertainties associated to the EFT parameters. Indeed, the bounds on the Wilson coefficients considered in [52] become much looser once PDFs are allowed to vary as one varies the value of the Wilson coefficients. The broadening of the contour is evident even once PDF uncertainties are fully accounted for. Even using a conservative set of PDF was used in the analysis, a PDF set that does not include any of the high-mass Drell-Yan sets and compare the bounds obtained using this set of PDFs to those obtained consistently using SMEFT PDFs, then the size of the bounds obtained by keeping fixed SM PDFs is closer to the size obtained from the simultaneous fits, although still slightly underestimated [59].

These seminal studies need and deserve to be brought forward by exploring a broader number of operators and observables and at the same time building methodological developments that allow a robust simultaneous determination of the Wilson coefficients of an EFT expansion and the PDF parameters, such as those put forward in [54, 55].

We note that the interplay between PDF uncertainties and the determination of SMEFT parameters will play a crucial role at other future colliders such as the Electron-Ion Collider (EIC). The possibility of polarizing both electron and nucleon beams at an EIC provides unique probes of SMEFT operators complementary to those obtained at both the LHC and the HL-LHC [60, 61]. However, maximizing the potential of these measurements requires a precise determination of the polarized PDFs of the nucleons. A detailed study of the effect of PDF and other systematic errors on SMEFT parameter determinations with polarized deuteron and proton beams at an EIC was recently undertaken [53]. Particularly in the high-luminosity phase of the EIC, polarized PDF errors are expected to form by far the largest source of systematic error on determinations of SMEFT parameters from polarized proton beams. A summary of the anticipated errors at a high-luminosity EIC with polarized proton beams is shown below on the right panel of Fig. 7. The uncertainty from polarized PDFs on the polarization asymmetry is orders of magnitude larger than the expected beam polarization error and other systematic errors coming from background processes, and is nearly as large as the expected statistical error over most of the available  $(x, Q^2)$  parameter space. This provides additional motivation for joint PDF-SMEFT determinations from future polarized deep-inelastic scattering data from the EIC.

### III. EXPERIMENTS

In this section we first discuss the measurements that are relevant to constrain PDFs at the LHC and their applications. We then turn to the opportunities to constrain unpolarized, polarized and nuclear PDFs at the Electron Ion Collider (EIC). Afterwards we describe the opportunities to constrain PDFs at Large Hadron electron Collider (LHeC). Subsequently, we discuss the importance of PDFs in neutrino phenomenology and the experimental constraints that we expect from new neutrino experimental facilities. Finally, we focus on forward and ultra-high energy scattering processes.

#### A. Measurements and applications of PDFs at the LHC

*Leading authors: A. M. Cooper-Sarkar, J. Huston*

In this section we first review the measurements from the LHC which are most sensitive to PDFs and are commonly used by the global PDF fitting groups, CT, MSHT and NNPDF. Second we point out which measurements can be substantially improved at the HL-LHC. Third we consider measurements which may be most sensitive to new physics and for which PDF uncertainties are the dominant background/uncertainty.

The measurements which are most sensitive to PDFs are:

- Inclusive  $W$  and  $Z/\gamma^*$  boson differential measurements [62–70], as a function of pseudo-rapidity and rapidity respectively. For the  $Z/\gamma^*$  different ranges of dilepton mass are also considered (hence Drell-Yan production is included in this heading). Furthermore there is a triple differential  $Z/\gamma^*$  measurement in rapidity, mass and the Collins-Soper angle. The experimental precision of the mass peak data is  $\sim 0.5\%$ . These measurements have impact on the valence PDF distributions, and in the LHC kinematic range for ATLAS and CMS, they also have impact on the flavour structure of the sea. It is now the case that N2LO QCD analyses are needed to obtain good fits to these data. NLO-EW predictions are also standardly applied. When considering high-mass Drell-Yan the photon PDF in the proton is an essential part of the formalism. For the low-mass Drell-Yan and for the higher rapidity ranges probed by LHCb, one may need to move beyond DGLAP to  $\ln(1/x)$  resummation or non-linear evolution equations [15].
- Inclusive jet and dijet measurements as a function of  $p_T$  in rapidity bins [71–76]. For dijets the dijet mass  $m_{jj}$  can be considered instead of the average  $p_T$ . The data precision ranges from  $\sim 5 - 50\%$ . Jet measurements mostly

have impact on the high- $x$  gluon PDF. The current state of the art is N2LO QCD and NLO-EW but for jets nonperturbative corrections for hadronisation and underlying event) are also applied, and these differ according to jet-radius. Larger jet radii are preferred theoretically. The dijet data are not yet fully exploited in PDF fitting, although several dijet data sets are included in the latest NNPDF release [10]. The jet measurements probe the highest scale and hence may require consideration of new physics [77].

- $W$ - and  $Z$ -boson + jets measurements [78, 79] extend the kinematic reach of the inclusive  $W$  and  $Z$  data to higher scale and higher  $x$ . The measurements are, for example,  $p_T^W$  for the  $W$  +jets and  $y^{\text{jet}}$  in bins of  $p_T^{\text{jet}}$  for the  $Z$  +jets. The  $Zp_T$  spectrum [80] can also be used instead of  $Z$  + jets data. The data precision is  $\sim 15\%$ . These data have impact on the gluon and on the quark PDFs, both valence and higher- $x$  sea structure. Predictions are N2LO QCD and NLO-EW, with nonperturbative corrections also needed for the jets. Note that data at low- $p_T$  are usually cut out because of the need for nonperturbative modelling of low- $p_T$  resummation, but predictions can still be sensitive to this cut.
- $t\bar{t}$  both total and differential cross sections [81–88]. The measurements are typically of mass  $t\bar{t}$ , rapidity  $t\bar{t}$ , average rapidity and average  $p_T$  and can be double differential. The data precision is  $\sim 15\%$ . These data mostly have impact on the high- $x$  gluon PDF, although this is not as strong as the impact of the jet measurements. The measurements can be made in the lepton + jets, the di-lepton channel and the fully hadronic channels, although the latter have not been used for PDF fitting as yet. Predictions are N2LO QCD and NLO-EW, with nonperturbative corrections also needed for the jets.
- Direct photon production [89] is once again being considered as an input to PDF fits. They are measured as a function of  $E_T^\gamma$  in bins of the photon pseudorapidity and the data for 8 and 13 TeV have been combined as ratios, with an experimental precision of  $\sim 5\%$ . These data impact the high- $x$  gluon PDF although less strongly than either the  $t\bar{t}$  data and the jet data. Predictions are N2LO QCD and NLO-EW.

Many of the above measurements are already systematics limited so that improvement is not a matter of improved statistics at the HL-LHC. Although it should be noted that high statistics can lead to better systematic uncertainty estimates. Also note that much of the data at 13 TeV from the full statistics runs up to 2018 are not yet included in PDF fits. Processes which may bring improvement are:

- Inclusive  $W$  and  $Z$  production at high rapidity
- High-mass Drell-Yan production
- Low-mass Drell-Yan production, modulo the issue of extensions of the DGLAP formalism mentioned above
- For inclusive jet production information on correlations between data sets could allow us to exploit inclusive jet, dijet and even trijet information simultaneously
- $W$  and  $Z$  boson +jet data at higher  $p_T$
- $W$  and  $Z$  boson + heavy quark data, particularly  $W + c$ , which can constrain the strange quark, providing that theoretical calculations can be extended to N2LO. There is some progress in that direction [90] but at present experimental and theoretical jet algorithms are not fully consistent.
- More differential information on  $t\bar{t}$  production, again assuming that N2LO predictions are available for double, or even triple differential distributions.
- Single top distributions have been considered by MSHT [9] and by NNPDF [91]. The impact is small at present but with better data their use could be extended.
- Isolated photon production and isolated photon-jet correlations at forward rapidity are sensitive to gluon density and saturation effects. Isolated photons originate predominantly in quark-gluon Compton scattering which can be probed over a large range of  $x$  and  $Q^2$  [92].
- Open charm or beauty production at LHCb rapidities has also not been exploited to any extent in the global PDFs, although some PDF studies exist [93, 94]. These have impact on the low- $x$  gluon and the current NLO theory may need extension to include  $\ln(1/x)$  resummation or non-linear effects, as much as to N2LO in DGLAP. Use of ratios can help to reduce uncertainties.



However, data which extend to high-scale may be subject to new physics effects. For example, when looking for new physics in  $Z'$  production at very high mass [95], or in jet production at high-mass [96], we have found that the PDF uncertainty limits our ability to see any new-physics signal. Furthermore we may be ‘fitting away’ new physics effects in the tails of the distributions of the data that we input. For this reason ATLAS have considered PDF fits which exclude data at scale  $Q > 500$  GeV [11] (this is mostly the inclusive jet data) and CMS have considered fitting their inclusive jet data using PDF parameters and SMEFT parameters simultaneously [58]. No evidence for new physics has yet been found but such approaches will have to be pursued in future as the interplay between new physics and PDFs will become stronger at the HL-LHC [52, 54].

New physics can also manifest itself at lower scales by the deviation of Standard Model parameters from their SM values. For example in recent measurements of the mass of the W-boson,  $m_W$  [97] and the weak mixing angle,  $\sin^2 \theta_W$  [98], the uncertainty due to the PDF used in the extraction is now one of the largest uncertainties. Various strategies have been proposed. Since the PDFs used usually lag behind the new measurements, the PDF can be improved by profiling the same data that are used for the SM parameter measurement. Of course there can be correlations between the SM parameters and the PDF parameters, so ideally a new simultaneous fit should be performed. Another point is that the PDF uncertainty is usually evaluated from comparing the results using different global PDF sets, as well as from the uncertainty within any one PDF set. This ignores potential correlations between these sets. One may try to reduce this uncertainty by evaluating these correlations. This is explored further in Section V D.

## B. PDFs at the Electron-Ion Collider

*Leading authors: T. J. Hobbs, E. R. Nocera, and R. S. Thorne*

The construction of an Electron-Ion Collider (EIC) [27, 99, 100] has been recently approved by the United States Department of Energy at Brookhaven National Laboratory, and could record the first scattering events as early as 2030. By colliding (polarized) electron, and potentially positron, beams with proton or ion beams at a center-of-mass energy of up to  $\lesssim 140$  GeV, the EIC will perform key measurements to investigate QCD at the Intensity Frontier. These measurements will be fundamental to understand how partons are distributed in position and momentum spaces within a proton, how the proton spin originates from the spin and the dynamics of partons, how the nuclear medium modifies parton-level interactions and substructure, and whether gluons saturate within heavy nuclei. In addition, the EIC will be capable of a range of PDF-related precision measurements in fundamental QCD and electroweak phenomenology. These include new constraints on standard model inputs like  $\alpha_s$  and the heavy-quark masses; novel electroweak probes for beyond standard model (BSM) physics; precise tests of QCD factorization theorems; and subtleties in the transition between non/perturbative QCD dynamics. Many of these aspects have been investigated in a recent Yellow Report [27], and also form the subject matter of a series of dedicated Snowmass whitepapers. Here, we discuss the relevance of future EIC measurements to improve knowledge of the proton PDFs, both unpolarized and longitudinally polarized, and of nuclear PDFs. Below, we discuss each of these in turn. We note also that the potential for the EIC to furnish constraining information on the PDFs of other hadrons — particularly the light mesons — has been discussed elsewhere [101, 102]; for details, we refer interested readers to these documents.

### 1. Unpolarized Proton PDFs

A large quantity of EIC data sensitive to the unpolarized PDFs will be supplied through inclusive neutral-current (NC) and charged-current (CC) DIS cross section measurements involving electron-DIS collisions with protons and light nuclei like the deuteron,  $^3\text{He}$ ,  $^4\text{He}$ . In the latter case, measurements involving light nuclei could be used to determine proton PDFs, including a correction [9, 103] or an uncertainty [10, 104] that takes into account nuclear effects, or nuclear PDFs themselves as discussed in Sect. III B 3 below. The inclusive NC and CC DIS cross section measurements taken at the EIC are expected to cover a broad kinematic region significantly overlapping with that probed by HERA, although with instantaneous luminosities potentially three orders of magnitude larger. In comparison to their HERA counterparts, these probes of the  $x$ - $Q^2$  plane will stretch to much higher values of  $x$ , typically up to  $x \sim 0.6 - 0.7$ . At sufficiently high center-of-mass energies, this region is expected to be rather insensitive to higher-twist effects which can be significant at  $W^2 \leq 15 - 20 \text{ GeV}^2$ ; these measurements will therefore cleanly constrain PDFs at relatively large  $x$ . At the same time, EIC coverage will also extend to softer values of  $Q^2$ , allowing a rich phenomenological program to examine power-suppressed corrections like the higher-twist effects. In comparison to previous DIS experiments, systematic uncertainties will be small, possibly not exceeding 1%; statistical uncertainties



will be even smaller.

The impact of EIC NC and CC inclusive DIS cross section measurements on the proton’s unpolarized PDFs were investigated in dedicated studies (see Sect. 7.1.1 in [27] and [105, 106]), whereby EIC pseudo-data were included in a selection of PDF frameworks, namely CJ [4], CT [7], JAM [107] and NNPDF [108]. Pseudo-data were generated for realistic projections of statistical and systematic uncertainties. For proton beams an integrated luminosity  $\mathcal{L} = 100 \text{ fb}^{-1}$  and center-of-mass energies of  $\sqrt{s}=28.6, 44.7, 63.3$  and  $140.7 \text{ GeV}$  were used for NC DIS;  $\mathcal{L} = 100 \text{ fb}^{-1}$  and  $\sqrt{s}=140.7 \text{ GeV}$  were used for CC DIS; for deuteron beams only NC DIS was considered, with  $\mathcal{L} = 10 \text{ fb}^{-1}$  and  $\sqrt{s}=28.6, 66.3,$  and  $89.0 \text{ GeV}$ . Pseudo-data were found to have a potentially strong impact on the (large- $x$ ) valence PDF sector, where PDF uncertainties could decrease up to 80%. On the other hand, the sea PDF sector was predominantly modified in the small- $x$  region, with a decrease of PDF uncertainties up to 50%.

The EIC may also have at its disposal the ability to perform analogous measurements using positron beams — a possibility explored in Sect. 7.1.1 in [27] as an eventual program upgrade. The different charge of the exchanged  $W$  boson is such that positron-initiated CC DIS interactions are capable of probing a combination of flavor currents complementary to electron DIS. This potentially constrains the  $d$ -type PDFs, and, indirectly, the  $d/u$  ratio. Beyond this, the use of positron beams may also allow one to access other parity-violating effects, such as the breaking of the strange–antistrange symmetry or parton-level charge-symmetry violation [109].

Aside from purely inclusive measurements, the EIC will be capable of many other measurements to test fundamental QCD and which offer unique potential sensitivity to the unpolarized proton PDFs. For instance, the EIC will also measure semi-inclusive DIS processes. Tagged DIS (TDIS) data offer a way to probe the structure of a barely off-shell neutron via semi-inclusive tagging of a slow spectator proton in  $e + d \rightarrow e' + p + X$  events. The EIC electron DIS data augmented with TDIS data was shown [27] to improve the determination of all flavors over the whole  $x$  range, in particular, for the  $d/u$  ratio at large  $x$ . These measurements may complement those planned at JLab by extending their kinematic reach to higher energies. On the other hand, SIDIS data offer a way to access PDFs for individual quark flavors, given that the valence parton content of the hadron detected in the final state relates to the fragmenting parton flavor. By means of EIC pseudo-data (for  $\mathcal{L} = 10 \text{ fb}^{-1}$  and  $\sqrt{s}=45$  and  $140 \text{ GeV}$ ) it was shown [110] that the impact of pion production SIDIS data on up, down, anti-up and anti-down quark PDFs is moderate, as they are already very well determined. Conversely, the far less known strange PDFs could be constrained substantially by kaon production SIDIS data, particularly at low  $x$ . The analysis of SIDIS data requires the simultaneous knowledge of fragmentation functions (FFs) [111], whose determination will be concurrently improved at the EIC [107, 110, 112]. Finally, final-state tagging of a produced charm quark may also help discriminate among scenarios for the strange sea. In [113] an event-level variation in CC DIS production of charm jets was observed depending on the input strange PDFs. This dependence suggests that charm-jet production may be a sensitive channel to constrain nucleon strangeness and disentangle patterns of SU(3) symmetry breaking in the light-quark sea. Along with direct measurements of the proton’s charm structure function,  $F_2^{c\bar{c}}$  [114], the EIC’s charm-tagging ability may possibly constrain a nonperturbative component of the charm quark PDF.

## 2. Polarized Proton PDFs

The EIC will allow for the longitudinal polarization of both the colliding nucleon (and light nuclei) and lepton beams, *i.e.*, along their direction of motion. This is a feature unique to the EIC, specifically designed to probe the longitudinal spin structure of the proton. Key to this goal will be inclusive DIS measurements. Beside the parity-conserving longitudinal double-spin asymmetry, the EIC will access also the parity-violating asymmetry, see, *e.g.*, Sect. 18.2 in [115] for a definition. In the numerator of the latter observable, the parity-conserving contributions from the photon exchange and the vector–vector part of the  $Z$ -boson exchange cancel exactly, leaving the dominant contribution from the interference between the photon exchange and the axial-vector part of the  $Z$ -boson exchange. While the parity-conserving asymmetry probes the sum of polarized quark and antiquark distributions, the parity-violating asymmetry probes their difference. The combination of the two is one of the cleanest ways to separate quark and antiquark polarizations.

Parity-conserving and parity-violating polarized DIS asymmetries are expected to expand the kinematic coverage of current DIS measurements significantly, roughly by one order of magnitude or more, down to  $x \sim 10^{-4}$  and up to  $Q^2 \sim 1000 \text{ GeV}^2$ , see, *e.g.*, Fig. 1 in [116]. In addition to the increased sensitivity to quark, antiquark and gluon polarized PDFs at small values of  $x$ , the wide  $Q^2$ -coverage of the EIC will probe scaling violations in the  $g_1$  polarized structure function, offering significant additional constraints on the gluon polarized PDF.

The impact of parity-conserving and parity-violating longitudinal spin asymmetries was investigated in dedicated studies (see Sect. 7.1.2 in [27] and [117–121]), whereby EIC pseudo-data were included in a selection of polarized PDF frameworks, namely DSSV [122, 123], JAM [124], and NNPDF [125]. Pseudo-data were generated for realistic projections of statistical and systematic uncertainties, assuming an integrated luminosity  $\mathcal{L} = 10 \text{ fb}^{-1}$  and center-of-

mass energies of  $\sqrt{s}=45$  and 140 GeV. Pseudo-data were found to have a remarkable impact on the polarized PDF uncertainties. If one assumed SU(3) flavor symmetry for the axial-vector charges, the uncertainty on the first moment of the gluon polarized PDF reduced by 80–90%, depending on the behavior of the low- $x$  extrapolation of the structure function  $g_1$  used to generate the pseudo-data. The uncertainty reduction on the first moment of the sum of all quark and antiquark polarized PDFs was found to be similar (around 80%), however, it was more modest if one did not impose SU(3) symmetry. In this case, the uncertainty on the gluon moment decreased by about 60%, and no clear reduction in the uncertainty of the quark and antiquark moment was seen. Be that as it may, these results will test to which extent the small- $x$  dipole formalism [126–129] holds and which fraction of the proton spin cannot be ascribed to the spin of quarks and gluons [130].

Additional measurements will also be investigated at the EIC, which could further constrain sea-quark polarized PDFs and the gluon polarized PDF. Concerning sea quark polarized PDFs, measurements of SIDIS cross sections with polarized beams are expected to significantly reduce the uncertainties on up, down and strange antiquarks in comparison to current knowledge, see [131]. Identification of kaons in the final state may in particular shed light on the strange sea polarization, whose shape cannot be determined from parity-conserving DIS asymmetries and is usually constrained by assuming exact SU(3) flavor symmetry and relating its first moment to hyperon beta-decay constants. The EIC SIDIS data will possibly establish whether there is a non-zero strange polarization at  $x > 0.5 \times 10^{-5}$  [27]. The analysis of SIDIS data may require the simultaneous determination of fragmentation functions [124]. In a similar spirit, the use of DIS and SIDIS longitudinal spin asymmetries, instead of cross sections, may require the simultaneous determination of unpolarized PDFs [132]. Concerning the gluon polarized PDF, processes such as photon-gluon fusion in the production of back-to-back partonic jets with large transverse momentum have been shown to be feasible at the EIC [133]. Dijet longitudinal double-spin asymmetries could be measured with a moderate integrated luminosity; these could be used as a cross-check of the more stringent constraint on the gluon polarized PDF provided by the evolution of the polarized structure function  $g_1$ . Finally, additional constraints on the gluon polarized PDFs could come from a measurement of the heavy-quark contribution to the polarized structure function, in a manner similar to studies at HERA for the unpolarized case [134, 135], though theoretical precision is currently potentially slightly limited here.

### 3. Nuclear PDFs

The EIC will be capable of colliding (un)polarized light ion beams and unpolarized heavier ions with beams of electrons, and potentially, positrons. Inclusive NC DIS cross section measurements are envisioned using  $^4\text{He}$ , C, Ca, Au and Pb nuclei. Their kinematic coverage will roughly double that of currently-available data, both at low  $x$  and at high  $Q^2$ , see, *e.g.*, Fig. 7.66 in [27]. As with the DIS program involving proton collisions, systematic and statistical uncertainties are projected to be small in comparison with previous experiments. These measurements are therefore expected to constrain quark and gluon nuclear PDFs to unprecedented precision. The gluon nuclear PDF could be further constrained by heavy flavor cross section measurements, obtained by tagging the decay products of  $D$ -mesons originating from charm fragmentation.

The impact of NC DIS cross sections in electron–ion collisions on nuclear PDFs was studied in Sect. 7.3.3 of Ref. [27] and Ref.[106]. In [27], similarly to the case of unpolarized and polarized proton PDFs, pseudo-data were included in three different frameworks, namely EPPS [136], nCTEQ [137] and nNNPDF2.0 [138]. Pseudo-data were generated with an integrated luminosity of  $\mathcal{L} = 10 \text{ fb}^{-1}$  and three center-of-mass energies of  $\sqrt{s}=28.6$ , 66.3 and 89 GeV. It was found that EIC NC DIS cross section measurements could reduce the quark and gluon PDF uncertainties for nuclei in a wide range of atomic mass values both at small and large  $x$ , by up to a factor of two. The reduction is such that nuclear PDF uncertainties may no longer encompass the difference between predictions obtained with a free proton or with a proton bound in a nucleus, *e.g.*, as currently found to be the case when modelling the interactions of ultra-high energy cosmic neutrinos with matter [106]. The impact of heavy-flavor production was studied in [139], where a similar reduction of the gluon nuclear PDF uncertainty was found at large  $x$ .

Because the EIC will have the capability to operate with a range of nuclei, from deuterium to lead, the dependence of nuclear PDFs on the atomic mass number  $A$  could be investigated quantitatively. Current parametrizations assume that this dependence is continuous, and determine it by analyzing data for different nuclei at the same time. The abundance of EIC measurements will make it possible to determine nuclear PDFs independently for each nucleus; the dependence on  $A$  could therefore be studied *a posteriori*. Finally, because proton and ion beams will be used in a consistent experimental framework, the level of sophistication of PDF analyses may need to improve, in particular to allow for a combined, simultaneous determination of proton and nuclear PDFs. This may reduce inaccuracies that follow from using the former as input to the latter and *vice versa*: nuclear PDFs in the analysis of nuclear data included in proton PDF determinations; and proton PDFs as the boundary condition for analyses of nuclear PDFs.

### C. The Large Hadron electron Collider (LHeC)

Leading authors: N. Armesto, D. Britzger, C. Gwenlan, M. Klein, F. I. Olness

The proposed *Large Hadron electron Collider* experiment (LHeC) [28, 140–142] at CERN will provide a unique set of electron-proton/nucleus collision data. It will afford superior sensitivity to PDFs and related subjects through highly precise measurements of neutral-current and charged-current deep-inelastic scattering (NC and CC DIS) cross sections, jet-production cross sections in DIS, as well as heavy-flavor cross sections in NC and CC DIS. The LHeC experiment can (only) be realized in the 2030s at the HL-LHC, and it is also the *cleanest high-resolution microscope* that can be attained in the next decade due to its unprecedented resolution of the partonic constituents and dynamics in hadronic matter down to  $x$ -values as small as  $10^{-6}$ , and up to  $x \sim 0.9$ .

The LHeC experiment will add to the HL-LHC a new high-energy high-intensity electron accelerator based on an energy-recovery-linac (ERL) technology [142–144], which provides an electron beam energy,  $E_e$ , of 50 to 60 GeV. The electron beam will be collided with one of the proton beams from the HL-LHC, thus resulting in an  $ep$  center-of-mass energy of 1.3 TeV. Further running modes will provide positron-proton ( $e^+p$ ), lepton-nucleus ( $e^\pm A$ ), proton-proton or nucleus-nucleus collision data [142]. For  $ep$  collisions, the luminosity will reach  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ , so the LHeC could provide about  $50 \text{ fb}^{-1}$  during an initial 3-year run (which would be equivalent to 50 times the entire accumulated HERA data set), and will finally reach an integrated luminosity of about  $1 \text{ fb}^{-1}$  after the HL-LHC era. The data taking of  $ep$  collisions with the LHeC experiment will take place at LHC interaction point 2 (IP2) and will be performed concurrently to the  $pp$  data taking with ATLAS, CMS and LHCb at the other three IPs. Recently discussed new considerations on the accelerator, and particularly on the design of the interaction region, explore a unique three-beam interaction point, where lepton-hadron and hadron-hadron collisions can be recorded with a single experiment [142]. Together with a symmetric detector design, as is commonly used at hadron colliders, the LHeC physics programme could further comprise the physics of the ALICE3 programme, and that would even benefit considerably from the improved calibration that can be obtained from the  $ep$  collision data. The following discussion will focus on the LHeC; however, essentially all the results carry forward to the FCC-eh [145], which is designed to utilise the same ERL technology, and would further extend the rich physics program of the LHeC to even higher energies.

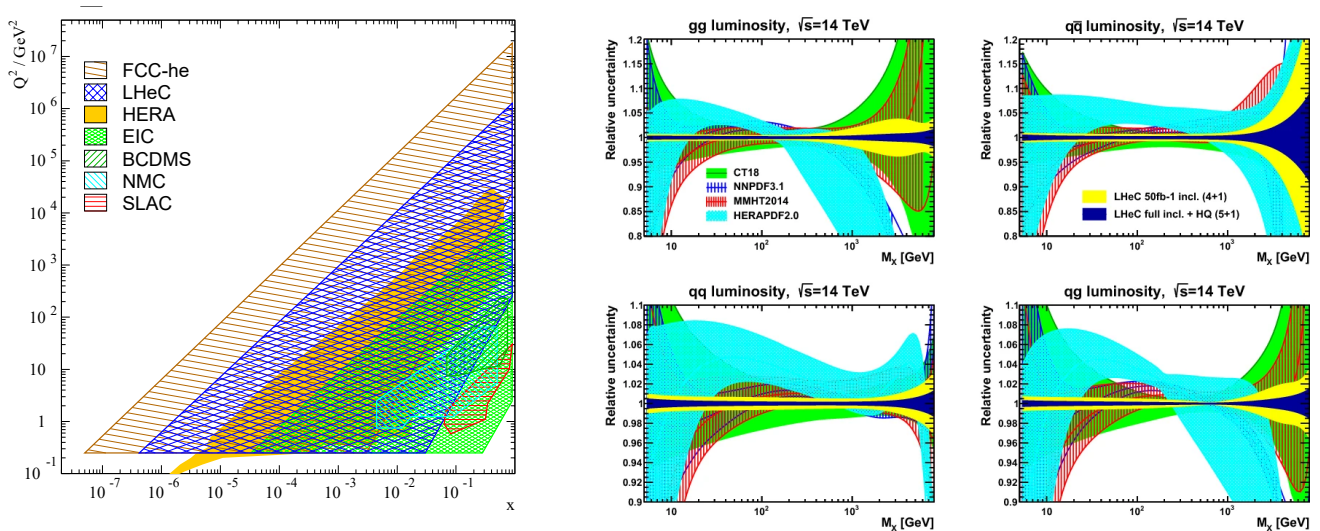


FIG. 8. Left: Coverage of the kinematic plane in deep inelastic lepton-proton scattering by some initial fixed target experiments (SLAC,NMS, BCDMS), and by the  $ep$  colliders: the EIC (green), HERA (yellow), the LHeC (blue) and the FCC-eh (brown). Figure from Ref. [28]. Right: Expected precision for the parton-parton luminosities as a function of  $M_X$  in Drell-Yan scattering at the LHC at  $\sqrt{s} = 14$  TeV for three recent PDFs (shaded areas) and for PDFs from LHeC (full areas) shown for an initial 3-year LHeC run (yellow) and the full LHeC data set (dark blue). Figure from Ref. [142].

Studies on the expected sensitivity of LHeC data on PDFs were presented in Refs. [28, 141, 146], where simulated NC and CC DIS data, including a full set of statistical and systematic uncertainties [28, 147], were investigated.

The coverage of the  $\{x, Q^2\}$  kinematic plane of the LHeC  $ep$  data is displayed in Fig. 8 (left) and compared to HERA, EIC, FCC-eh and fixed-target experiments. The data at the LHeC span a considerable kinematic range in  $Q^2$  up to  $10^6 \text{ GeV}^2$ , and  $x$  in the range of  $10^{-6} \lesssim x \lesssim 0.9$ . The measurements of inclusive NC and CC DIS cross sections at the LHeC benefit from the excellent calibration opportunities in  $ep$  collider experiments, from high-acceptance detectors with modern detector technologies [142], from sophisticated data analysis algorithms [148], and, of course, from high statistical precision. Typical total uncertainties in the bulk kinematic region will be of the order of 0.8 to 1.4%. Using such inclusive NC and CC DIS cross section measurements (which are expressed as combinations of the structure functions  $F_2$ ,  $xF_3$  and  $F_L$ ) as well as heavy quark production, the partonic structure of the proton (and nuclei) can, for the first time be completely resolved in a single experiment. The high energy collisions allow weak probes ( $W^\pm$ ,  $Z$ ) to dominate the interaction at larger  $Q^2$  values, which permits the up and down sea-quark PDFs and the valence quark distributions to be resolved in the full range of  $x$ . Data with different longitudinal electron-beam polarisation ( $P_e = -0.8, 0, \text{ or } +0.8$ ) also enhance the sensitivity. Obviously, independent data from the  $pp$  experiments will further improve the PDFs, but also introduce new theoretical challenges. The possibility to take positron-proton collision data greatly enhances the precision determination of the down-quark PDF. Dedicated data at different  $ep$  center-of-mass energies give access to the longitudinal structure functions  $F_L$  [28].

The gluon PDF is poorly known today, but is of crucial importance for precision Higgs, electroweak and top-quark physics at the (HL-)LHC [149], while large  $x$  is important for new physics searches. The LHeC constrains the gluon to percent accuracy for all  $x$ -values probed, by using a variety of measurements, primarily from scaling violations ( $\partial F_2 / \partial \log Q^2$ ) as well as the longitudinal structure function  $F_L$ . The measurement of jet cross sections in the Breit frame provides a further constraint on the gluon since jets are predominantly initiated in the boson-gluon fusion channel.

The size of the strange quark PDF is a long-standing puzzle. Measurements ranging from fixed-target to collider experiments have not resolved this important question, and the  $x$  dependence of  $xs(x, Q^2)$  is rather unknown, and it may differ from that of  $x\bar{d}$  or  $x(\bar{u} + \bar{d})$ . A direct measurement of  $xs(x, Q^2)$  and the resolution of the complete light-quark structure of the proton over a wide  $x$  range is a fundamental goal of the LHeC. To cite one example, the precise extraction of the strange PDF can be performed directly using the charm production process in CC DIS ( $Ws \rightarrow c$ ) [150].

The LHeC will provide unprecedented precision measurements on heavy quark production which can address a variety of outstanding questions, including the following: To what extent do the universality and factorization theorems work in the presence of heavy quarks?; Are the current theoretical tools sufficient to address the multi-scale paradigm we encounter when adding new heavy quark mass scales?; Are heavy quarks like charm and bottom radiatively generated, or is there also an intrinsic heavy quark component in the proton? Using charm and beauty tagging with high precision in NC  $ep$  scattering, the LHeC can completely resolve components of the proton by flavor, and extract  $F_2^c$  and  $F_2^b$ . In addition, one may also use DIS jets and low energy data to achieve a precision measurement of  $F_L$ . Finally, the LHeC provides the first direct access to top quark production in a DIS environment, and allows for single top production ( $Wb \rightarrow t$ ), top pair production ( $g \rightarrow t\bar{t}$ ), and even investigation of the top-quark PDF. The LHeC set of heavy quark measurements can offer unique and incisive information on the heavy quark content of hadrons, and adeptly address the outstanding theoretical questions concerning heavy quark mass effects.

The measurement of jet production cross sections in NC DIS in the Breit frame at the LHeC will not only enhance the sensitivity to the gluon PDFs (see above), but also provide high sensitivity to the strong coupling constant  $\alpha_s(m_Z)$  [28, 151], since jet cross sections are proportional to  $\mathcal{O}(\alpha_s)$  already in leading-order QCD. At the LHeC, jets with transverse momenta from 3 GeV up to 500 GeV will be recorded. Due to the over-constrained kinematics in NC DIS, the jet energy scale can be calibrated with high precision, and reaches an uncertainty of 0.3% to 0.5%, a value significantly smaller than present LHC experiments (in part also because of the absence of pile-up and underlying event). This translates into an uncertainty of about 1 to 5% on the jet cross sections in the Breit frame [28]. In a simultaneous PDF+ $\alpha_s$  fit, where inclusive DIS and jet pseudo-data were exploited, an uncertainty in the strong coupling constant of  $\delta\alpha_s(m_Z) = \pm 0.00018$  is achieved [28], which is a factor of 6 smaller than the present world average value, and it will be a challenge to match that experimental precision with equally accurate theoretical predictions (c.f. Sect. IV). Related measurements of the hadronic final state, like event shapes,  $n$ -jettiness observables, jet substructure observables or multi-jet cross sections, may all be included in PDF determinations and, commonly, provide predominantly a sensitivity to the gluon distribution, or to the valence quarks at high- $x$ . However, precision measurements of lepton-jet decorrelation observables, measured in the laboratory rest frame, may be further sensitive to TMD effects [152, 153].

Interestingly, the large luminosity of the LHeC provides high experimental precision at high  $x$ , where inclusive DIS data are sensitive to  $\alpha_s$  through scaling violations, as well as precision measurements of  $F_L$  at high  $y$ . Consequently,



the strong coupling constant can be determined together with the PDFs already from inclusive NC and CC DIS data alone, something, that was not possible with HERA data [154]. A PDF+ $\alpha_s$  analysis of inclusive DIS pseudo-data yields an uncertainty in  $\alpha_s$  of  $\pm 0.00022$ , which again imposes a real challenge to provide accurate theoretical predictions at N<sup>3</sup>LO or even beyond. These studies underline the extraordinary high precision of the inclusive DIS data from LHeC to QCD phenomena, which are otherwise inaccessible experimentally.

The physics of low  $x$  and high  $x$  phenomena can be studied at the LHeC, due to the high luminosity of the accelerators, the large acceptance of the LHeC detector, and the high  $ep$  center-of-mass energy. In addition, since only a single hadron is involved at the LHeC (in contrast to LHC  $pp$  data), the PDF determinations are free from low- $x$ –high- $x$  correlations, and the physics phenomena in these two extreme regions can therefore be studied separately, with high precision. The very high luminosity leads to ample statistics in the large  $x$  region at such a high  $Q^2$  that higher twist effects become negligible. This region is especially important for constraining BSM signatures with large mass scales at the LHC. At small  $x$  the gluon and sea quark densities, as discovered at HERA, rise so much that non-linear and possibly saturation effects may become manifest [28]. This can be studied for the first time reliably in  $ep$ , and  $eA$ , at the LHeC, at such a high  $Q^2$  that the strong coupling is small. This may replace the DGLAP evolution by BFKL type equations and/or non-linear evolution, with major consequences for future hadron collider physics at HL-LHC and beyond. With new measurements of diffractive DIS cross sections, the field of diffractive parton-distribution-functions will gain new interest [28, 155].

Beyond the collinear PDFs, semi-inclusive measurements of jets and vector mesons, and especially exclusive vector meson production and Deeply Virtual Compton Scattering (DVCS), the latter a process established at HERA, will shed light on the transverse structure of the proton in a new kinematic range. These measurements allow us to access the Wigner distribution  $W(x, k_T, b_T)$ ; one can think of it as the “master” parton distribution. When integrating the Wigner distribution over the transverse momentum ( $k_T$ ), one obtains a Generalized Parton Distribution (GPD)  $f_{\text{GPD}}(x, b_T)$ , while if we integrate over the impact parameter ( $b_T$ ), one obtains a Transverse Momentum Dependent (TMD) PDF  $f_{\text{TMD}}(x, k_T)$ . Due to the considerably higher  $ep$  center-of-mass energy, the LHeC will investigate both TMDs and GPDs down to much lower  $x$  and higher  $Q^2$  than the EIC, and thus provide a complementary perspective (c.f. Sect. VIII), and shed light on their evolution with  $x$  and  $Q^2$ .

While HERA inclusive NC and CC DIS data have relevant sensitivity to PDFs, their sensitivity to further parameters in so-called PDF+ $X$  fits is rather limited. For example,  $X$  could be  $\alpha_s$  [154] or electroweak parameters [156]. The high luminosity of the LHeC will change that picture significantly, and, as just discussed in the context of  $\alpha_s$ , above, the LHeC inclusive DIS data will have significant sensitivity to parameters other than PDFs. The sensitivity of inclusive DIS data to electroweak parameters was studied in Refs. [28, 157] using PDF+ $X$  fits, where NNLO QCD and NLO EW corrections were employed. It is found, that in the on-shell renormalization scheme, the mass of the  $W$ -boson can be determined with an uncertainty of  $\delta m_W = \pm 6$  MeV, which is at a level where EW theory uncertainties are significant. More interestingly, the leptonic effective weak mixing angle at the mass of the  $Z$ -boson can be determined with an uncertainty of  $\delta \sin^2 \theta_{\text{W},f}^{\text{eff}} = \pm 0.00015$  [157], which is of comparable size to the LEP+SLD combination [158] or the HL-LHC prospects [159]. Even parameters contributing beyond the leading-order formalism can be tested with LHeC inclusive DIS data, and, for example, the oblique parameters  $S$ ,  $T$  and  $U$  [160] can be studied, as well as modifications to the higher-order form factors  $\rho_{\text{NC,CC},f}$  or  $\kappa_{\text{NC},f}$  [157]. The prospects for the FCC-eh are, of course, even more promising due to increased  $\sqrt{s}$  and  $\mathcal{L}$  [161]. The sensitivity to other quantities can be considered as well, such as the proton radius, contact-interaction [141] or EFT parameters, and these will provide rich physics opportunities with PDF+ $X$  studies. Moreover, with the inclusion of HL-LHC  $pp$  data, combined fits of PDFs with SM/BSM parameters will gain a considerable attention in the 2030s.

The precise knowledge of the proton PDFs is an important prerequisite to achieve the goals of the HL-LHC physics programme, including Higgs-phenomenology, measurements of electroweak parameters ( $W$ -boson mass,  $\sin^2 \theta_W$ , ...), top-quark physics, and high-mass searches. Although PDFs can also be constrained from (HL-)LHC data themselves [26] (see also sect. III A), the importance of constraints from an independent experiment should not be underestimated. Since already today most of the measurements at the LHC are limited by systematic, rather than statistical uncertainties, the extracted PDFs represent the systematic uncertainties of those experiments. A valid application of such PDFs for LHC phenomenology is therefore non-trivial (c.f. Sect. V A), or requires comprehensive simultaneous PDF+ $X$  analyses (see Sect. V C 2). For the FCC-hh, the small- $x$  dynamics will affect production of particles with masses  $\mathcal{O}(100)$  GeV, including Higgs [28]. Consequently, an independent experiment to determine the proton PDFs, with completely uncorrelated systematic and even theoretical uncertainties, is therefore of crucial importance to achieve the physics goals of the HL-LHC  $pp$  programme. The expected uncertainties of the parton luminosities at the HL-LHC in  $pp$  collisions at  $\sqrt{s} = 14$  TeV of LHeC PDFs is displayed in Fig. 8 (right). It is clearly

seen that the uncertainties will reduce by an order of magnitude, compared to modern PDFs, and the improvement is particularly high at the electroweak scale. The improvement in PDF uncertainties afforded by the LHeC can also be predicted to exceed those of the HL-LHC PDFs, in particular for applications in SM phenomenology. As an example, it was studied in Ref. [159], that LHeC data would reduce the PDF uncertainty in the measurement of the  $W$ -boson by ATLAS to only  $\pm 1.6$  MeV, while it would be  $\pm 3.7$  to  $\pm 5.8$  MeV with HL-LHC data. In the small- $x$  region, the information and constraints that the LHeC could provide, is incomparable.

Additionally, the high luminosity ensures that proton data is sufficient to extract the flavor components without the use of fixed-target DIS data which typically involves nuclear corrections. While the LHeC can completely resolve the proton PDF flavors without using any nuclear data, the option of an LHC heavy ion beam allows exploration of individual nuclear PDFs.  $e^\pm A$  collisions at the LHeC [28, 141] will be performed at  $\sqrt{s} \simeq 0.8$  TeV per nucleon (for Pb) with per nucleon instantaneous luminosities  $\sim 7 \cdot 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$ . They will allow, as in  $ep$ , a complete unfolding of the PDFs of a single nucleus for the first time, without the use of fixed target or hadron-nucleus data. The corresponding uncertainties will be considerably smaller than those in present global fits due to the use of single nucleus data (therefore with no need of functional initial conditions depending on nuclear size) obtained in a single experiment (thus, large tolerances are not required). The data can also be used for global fits, and the single nucleus PDFs for precision checks of collinear factorization when used for predictions in proton-nucleus and nucleus-nucleus collisions.

Studies of diffraction on nuclei present, as in the case of the EIC, the challenge of forward instrumentation required to distinguish coherent from incoherent diffraction. If such separation can be achieved, diffractive nuclear PDFs will be measured in a large kinematic domain, comparable to that in  $ep$  collisions [28, 155]. Also nuclear GPDs and TMDs will be studied in the nuclear case, using the same observables employed in  $ep$ . Finally, the eventual discovery and verification of the current explanation of the non-linear saturation regime of QCD as a density effect requires both decreasing  $x$  and increasing  $A$ , making  $eA$  collisions essential.

As a final remark on  $eA$  collisions, a precise knowledge of the nuclear partonic structure in the collinear regime and beyond – in a kinematic region matching that of the corresponding hadronic colliders, and of the QCD dynamics at small  $x$  or high energies, is central for heavy ion collisions at the LHC or FCC-hh. The characterization of the hot dense medium produced in ion-ion collisions, the Quark-Gluon Plasma (QGP), suffers from large uncertainties derived from the present lack of knowledge on these aspects [28]. And many observables taken as signatures of QGP formation are found in smaller collisions systems, proton-proton and proton-nucleus, the most prominent of them being the ridge. Measurements at the LHeC and the FCC-eh can offer key information to solve these issues.

Undoubtedly, HERA had an outstanding impact on our present knowledge of the proton structure. The LHeC, with its 1000 times larger luminosity (and higher center-of-mass energy and kinematic reach), will equally advance the field and will provide the relevant experimental input data for precision PDF physics in the 2030s. Furthermore, such independent PDFs are of crucial importance to achieve the physics goal of the HL-LHC programme.

#### D. PDFs for neutrino phenomenology

*Leading authors: T. J. Hobbs, K. Xie, and B. Zhou*

PDFs play a crucial role in neutrino interactions above a few GeV, a kinematical region which is dominated by deep inelastic scattering (DIS) between neutrinos and the target nucleus [162–167]. The precision of theoretical predictions for the relevant cross sections is pivotal for medium-, high-, and ultra-high-energy neutrino physics and astrophysics. The relevant experiments include DUNE [168], Super-Kamionkande [169], Hyper-Kamionkande [170], IceCube [171], KM3NeT [172], Baikal-GVD [173], IceCube-Gen2 [174], ANITA [175], ARA [176], GRAND [177], etc.

Above  $\sim 100$  GeV and  $\sim 100$  TeV, charm- and top-quark production are especially important for DIS [178], but, even at lower energies, the precision of cross-section calculations depends in part on a proper accounting of heavy-quark mass effects [179]. At leading order, slow-rescaling and the modification light-cone momentum fraction in the PDFs must be included [180]. At higher orders, different formalisms have been developed [181–183]. However, depending on which approach is used, the contribution of charm or top production can be significantly different.

At PeV and EeV energies, DIS probes kinematical regions of very small  $x$  and large  $Q^2$ . PDFs have very limited data from colliders and fixed-target experiments [7] with direct sensitivity to this region; as a result, there is currently significant dependence on extrapolations to these largely unfitted regions. Moreover, at very small  $x$ , the perturbative expansion is not stable, such that resummation corrections must be included in the DGLAP formalism [167]. The color dipole approach [184–187] provides an efficient way to account for resummation and saturation effects. For



different formalisms and input PDFs, the predicted cross sections can differ by as much as a factor of a few in the ultra-high energy regime.

Nuclear effects on PDFs are also important, as DIS mostly happens on the nucleus for neutrino detection (water/ice, argon, lead, *etc.*) and via propagation through the Earth (iron, oxygen, silicon, *etc.*). An important effect at higher energies (related to the behavior of nuclear PDFs at low  $x$ ) is nuclear shadowing [188], *i.e.*, the relative depletion of nuclear structure functions as compared to their free-nucleon counterparts. In the past few years, nuclear PDFs have been published [136, 137, 189, 190] which explore this phenomenology systematically as discussed in Sec. VII. DIS cross sections calculated using nuclear PDFs show 5-15% suppression at PeV energies, but the uncertainty of the nuclear corrections is still large [167, 191, 192]. Another nuclear effect is isospin. For  $W$ -boson exchange, neutrinos interact differently with protons and neutrons because of their different quark content. Current works treat the Earth as an isoscalar target, which is not true for some nuclear targets in the Earth. It would be very interesting to examine the nuclear effect with various models or assumptions to understand its impact, as discussed in Sec. VII and III B 3.

On the other hand, the photon PDF is also important for neutrino phenomenology, as it is the most important input for calculating neutrino-nucleus  $W$ -boson production [193–196],  $\nu_\ell + A \rightarrow \ell^- + W^+ + X$  (see Fig. 1 of Ref. [195] for relevant diagrams), where a neutrino couples to a nuclear photon through the charged lepton or  $W$  boson split from the neutrino. The cross sections are the second largest for high-energy neutrinos, reaching 5–10% of DIS on water and 10–15% on iron [195, 196]. Therefore, a precise photon PDF is important to accurate determinations of the cross sections for this process, and nuclear effects on photon PDFs may also be important.

Dimuon events (involving two energetic muons emanating from one neutrino interaction) at accelerator-based neutrino experiments have been very important to measuring the strange-quark PDF in the region of  $x \gtrsim 0.01$  and  $Q \lesssim 10$  GeV. Recently, Ref. [197] proposed that high-energy neutrino telescopes like IceCube and IceCube-Gen2 can achieve a higher level of sensitivity in dimuon detection due to the small vertical spacing between the detector’s digital optical modules. This work further predicted that IceCube and IceCube-Gen2 can detect  $\simeq 1000$  dimuon events in 10 years and that these events can probe the strange-quark PDF in the region of  $x \gtrsim 0.01$  and  $10 \lesssim Q \lesssim 100$  GeV. These dimuons events can also be used to detect the production of  $W$  bosons as another means of probing the nuclear photon PDF as discussed above.

The FASER $\nu$  [45] experiment is designed to detect neutrinos produced in proton-proton collisions in the far-forward region of the ATLAS detector. Neutrinos produced from decays of heavy mesons such as the  $B, D$  *etc.*, can provide information on the heavy-flavor and gluon PDFs. Considering the far-forward kinematics, the FASER $\nu$  measurement is able to probe the PDFs down to  $x \sim 10^{-8}$  [198], a region which has not been probed by existing experiments. In this scenario, new data would provide fresh insights into how the PDFs behave at very low  $x$ , possibly informing a new understanding of QCD, especially with respect to the small- $x$  behavior as discussed in Sec. IV D 2.

Complementary to the considerations discussed above for neutrino scattering at the TeV scale and beyond, PDFs also play an important role at lower energies in GeV-scale experiments like the upcoming DUNE/LBNF effort [168] at Fermilab. Long-baseline experiments like DUNE depend upon precise control over the neutrino-nuclear interaction over a wide range of scattering energies,  $E_\nu$ , to achieve their target sensitivities to the neutrino-mass hierarchy and a possible CP-violating phase,  $\delta_{\text{CP}}$ , in the neutrino sector. In the case of DUNE, the anticipated neutrino flux will peak near  $E_\nu \sim 2.5$  GeV with a substantial tail to higher energies. In this region, the neutrino-nuclear cross section must be determined from a complicated mix of underlying processes, including quasi-elastic scattering, resonance excitation, and deeply-inelastic scattering. The latter of these dominates the cross section at successively higher values of  $E_\nu$ , but, in the few-GeV region, has an important dependence on various nonperturbative effects, including contributions from higher-twist (*i.e.*, twist-4) and target-mass corrections. These must be systematically assessed and controlled in the delicate resonance-to-DIS transition region in a context in which nuclear effects are also critical. For this reason, nuclear PDF studies and extrapolations to the lower  $W$  and  $Q^2$  values of greatest relevance to DUNE will be a priority for enhancing understanding of the DUNE neutrino-nuclear program.

### E. Forward (and ultra-high energy) scattering processes

*Leading authors: M. Guzzi, L. A. Harland-Lang, M. Hentchinski, K. Xie, with a contribution from C. Loizides*

The kinematic regime of scattering processes in the very forward region is outside of the range of genuine validity of DGLAP picture. One may ask up to which values of the momentum fractions or rapidities the DGLAP picture remains valid. Multiple experiments have been proposed to study forward production in the future. Here we focus on LHCb, FPF, and ALICE FoCal as examples of forward physics experiments in which calculations in the collinear factorization picture can provide useful guidance and in-depth tests of QCD.

The LHCb experiment is designed for precision physics in the forward region. Properties of final-state particles in the forward configuration can be used to probe PDFs in regions at both small and large  $x$  ( $x \sim (Q/\sqrt{s})e^{\pm y}$ ).

Measurements of heavy-flavor charm and bottom production at LHCb provide constraints on the gluon and heavy-flavor PDFs at  $x \sim 5 \times 10^{-6}$  [199]. Such a small value of  $x$  is currently not covered by other LHC experiments. The large  $x$  region has been recently studied [200] at LHCb where cross section measurements of  $Z$  bosons in association with a charm quark were used to probe the existence of an intrinsic charm (IC) component of the proton. In addition, a recent study from the CTEQ-TEA group [7], has shown high sensitivity of Drell-Yan dilepton production at LHCb to quark and antiquark PDFs at small  $x$ , especially strangeness. Future high-luminosity measurements at LHCb will be critical to set stronger constraints on PDFs [201] and to explore small- $x$  dynamics [202].

The ALICE detector at the LHC, equipped with a dedicated forward calorimeter system (FoCal) [92] will allow us to start a new program to investigate small- $x$  gluon distributions of hadrons and nuclei. FoCal is designed as a highly granular Si+W electromagnetic calorimeter combined with a conventional sampling hadronic calorimeter covering pseudorapidities of  $3.4 < \eta < 5.8$ . Its performance is optimized to measure isolated-photon spectra at forward rapidity in the range of about  $4 < p_T < 20$  GeV/ $c$  with high precision even at the lowest momenta. This kinematic reach with photons is equivalent to constraining the gluon distribution in Pb nuclei down to Bjorken- $x$  of about  $10^{-5}$  over a large range of  $Q^2$  [203]. In addition to the photon measurements, FoCal will allow us to measure photon-jet and jet-jet correlations, as well as  $J/\psi$  production in ultra-peripheral collisions. These processes are strongly sensitive to non-linear effects at small- $x$ .

The Forward Physics Facility (FPF) is a proposal at CERN to complement the existing experimental programme with a range of far-forward detectors that will be in particular be able to collect a significant data sample of neutrinos produced due to particle production in the central ATLAS detector. The FASER, FASER $\nu$  and SND experiments will begin taking data in 2022, while there is a dedicated proposal to extend these by creating space in the far-forward region for a suite of upgraded experiments that would run during the High Luminosity LHC (HL-LHC) era, see [204]. Along with the range of BSM and neutrino physics studies that this will permit, there is promising potential for information about the proton and nuclear PDFs to be provided by the FPF.

In more detail, the neutrino flux is produced via the very forward production of particles in proton-proton collisions at ATLAS, such as light hadrons or charmed mesons, which will therefore be sensitive to the proton PDFs at both rather low and high  $x$ . In the former case this provides the potential to probe the PDFs (e.g. the gluon) in a rather poorly constrained low  $x$  region, as well as being sensitive to a range of non-linear and BFKL resummation effects. In the latter case, there is a distinct sensitivity to possible intrinsic charm in the proton, which is theoretically expected to be enhanced in the high  $x$  region. Several studies have investigated the possible existence of this intrinsic charm, including the recent measurements of  $Z$ +charm production by the LHCb experiment [200], which hints at its presence. However, so far no firm evidence exists and hence the FPF could shed light on this unresolved question.

In addition, due to the interaction of the neutrino flux with the FPF detectors, it will effectively operate as a neutrino-induced deep-inelastic scattering experiment with TeV-scale neutrino beams. Measurements of the resulting DIS structure functions will provide a valuable handle on the partonic structure of both nucleons and nuclei, in particular concerning quark flavour separation. Of particular note is the potential for measurements of charm-tagged neutrino structure functions, which would provide further information about the possible tensions between existing such data and measurements from the LHC on  $W, Z$  production. Moreover, not only emulsion experiments, which allow several kinds of charmed baryons and mesons to be tagged by reconstructing in detail the topology of their decays, but also experiments which allow to charm tagging through dimuon events, will be present. Hence the measurements of both inclusive and charm-tagged neutrino structure functions should be feasible at the LHC.

#### IV. THEORY

*Leading authors: S.-O. Moch, B. Mistlberger, G. Magni, with a contribution from J. Blümlein*

LHC particle physics phenomenology at the percent level allows us to stringently test our understanding of fundamental interactions and it is experimentally feasible at the LHC. In particular, the precise measurement of observables involving highly energetic electroweak bosons, top quarks or jets of QCD radiation shed light on some of the most pressing questions of modern particle physics. This motivates a large effort to improve our theoretical capabilities to predict hadronic scattering cross sections at the enhanced level of precision required to extract the desired information from LHC data. PDFs are the backbone of such predictions. Theoretical developments will play a crucial role in future improvements of PDFs and consequently are of great importance to our aim of maximally utilising LHC data.

### A. PDF evolution at N3LO

QCD factorization allows to express observables in  $ep$  and  $pp$  hard scattering with large momentum transfer schematically as

$$O^{ep} = f_i \otimes c_i^o, \quad O^{pp} = f_i \otimes f_k \otimes c_{ik}^o, \quad (4)$$

where the PDFs of the proton with dependence on the momentum fraction  $x$  are denoted by  $f_i(x, \mu^2)$  and the process dependent partonic cross sections (coefficient functions) by  $c^o$ . QCD factorization holds up to power corrections and is performed at the (renormalization and factorization) scale  $\mu$ , which is taken to be of the order of a physical hard scale.

The scale dependence of the PDFs is governed by the well-known evolution equations [205–207]

$$\frac{\partial}{\partial \ln \mu^2} f_i(x, \mu^2) = [P_{ik}(\alpha_s(\mu^2)) \otimes f_k(\mu^2)](x), \quad (5)$$

where  $\otimes$  denotes the Mellin convolution with the evolution kernels, i.e. the splitting functions  $P_{ik}$ . The latter are calculable in QCD perturbation theory and, together with the coefficient functions  $c^o$ , can be expanded in powers of the strong coupling constant  $a_s \equiv \alpha_s(\mu^2)/(4\pi)$ ,

$$P = a_s P^{(0)} + a_s^2 P^{(1)} + a_s^3 P^{(2)} + a_s^4 P^{(3)} + \dots, \quad (6)$$

$$c_a^o = a_s^{n_o} [c_o^{(0)} + a_s c_o^{(1)} + a_s^2 c_o^{(2)} + a_s^3 c_o^{(3)} + \dots]. \quad (7)$$

$N^n$ LO parton distribution functions are extracted from hadronic scattering cross sections by fitting cross section predictions using  $N^n$ LO hadronic cross sections to data. N2LO PDFs represent the state of the art, where the first three terms in Eqs. (6) and (7) provide the N2LO predictions for the observables (4). Currently, this is standard approximation for many hard processes and for PDF determinations, see [208–215] for the corresponding N2LO splitting functions.

Tackling the next perturbative order - N3LO - requires significant improvements and requires an unified effort from the theoretical particle physics community. The benefit to the particle physics phenomenology program is nevertheless clear: a consistent extraction and application of N3LO PDFs will result in more reliable predictions of scattering cross sections and ultimately in a reduction of uncertainties due to our limited knowledge of PDFs. In particular, work on the four-loop splitting functions in Eq. (6) to ensure QCD evolution equations at N3LO accuracy is ongoing [216–218]. The massless and massive Wilson coefficients are known [219–221]. With these results, the flavor non-singlet N3LO contributions to DIS can be implemented already now, since all ingredients are known to sufficient accuracy in the relevant range of parton kinematics, following e.g., [219, 222].

Finally there is another piece that has to be taken into account when looking at N3LO PDF evolution: the matching conditions for different flavor number schemes. In fact if the number of active, light flavors that are participating in the DGLAP equation changes by one unit, the distributions do not behave in the same matter above and below the threshold: in particular the new quark distributions  $q_{n_f+1}(x, \mu) = h(x, \mu)$  did not take part in the evolution below the threshold, but above they do. The nontrivial contribution of these matching conditions are  $\mathcal{O}(a_s^2)$  [223] and have been computed almost completely also at N3LO [208, 221, 224–230], allowing for consistent N3LO PDF evolution of the heavy quarks.

In this context, there is number of programs able to solve DGLAP at N2LO [231–233], but not yet a N3LO evolution tool. The recently released EKO [234, 235] is able to perform the full evolution up to N2LO and it contains some ingredients needed to N3LO, such as  $a_s$  running and the matching conditions; however the implementation of  $\mathcal{O}(a_s^3)$  splitting function is still work in progress.

The work on the determination of all N3LO splitting functions and matching conditions, along with their implementation in public codes, is paramount and will be one of the most important development in the precision physics program of the next decade.

### B. N3LO Cross Sections And Perturbative Uncertainties

The current frontier in QCD perturbation theory is posed by third order - N3LO - predictions. Calculations for partonic scattering cross sections, Eq. (7), for DIS processes at N3LO are already readily available [208, 220, 221, 236–240]. Predictions for cornerstone LHC process at N3LO are a very active field of development and are available for key inclusive cross sections [42–44, 241–249] as well as some fully differential predictions [250–254]. The overall picture that emerges from these computations is that corrections at N3LO are of the order of a few percent and residual

uncertainties due to the truncation of the perturbative expansion in the strong coupling constant are comparable to, or subdominant with respect to, other sources of theoretical uncertainties. In particular, uncertainties on parton distribution functions often represent the largest of the residual theoretical uncertainties.

Currently, N3LO PDFs are not available and computations of hadronic cross sections at N3LO are consequently performed using N2LO PDFs as inputs. While this procedure is theoretically sound, it naturally leads to the question of the phenomenological impact of N3LO PDFs on such predictions. To quantify the answer to this question in terms of an uncertainty due to missing N3LO PDFs an ad-hoc procedure was introduced in Ref. [41]. The authors defined an uncertainty by the relative difference of computing an N2LO cross section once with N2LO and once with NLO PDFs and reducing the size of this uncertainty by half in order to account for a perturbative reduction of the impact of N3LO over N2LO corrections. The definition is given in Section II.B.1, Eq. (3). Albeit ad-hoc,  $\delta(\text{PDF-TH})$  leads to a significant uncertainty on N3LO cross section predictions at to order of several percent in the case of key LHC observables [41, 43, 44] and is consequently of comparable size as the regular uncertainty associated with our current understanding of PDFs themselves. Moving forward a better procedure of estimating the uncertainty due to the mismatch of the perturbative order of PDFs and partonic cross section calculations is very desirable.

Extracting N3LO parton distribution will require a large range of cross sections computed at this perturbative order in order to fit to input measurements. Some of these computations will not be available imminently. Consequently, it is necessary to develop a scheme to set perturbative uncertainties on PDFs that takes into account the effect of fitting cross sections to predictions based on a mix of N2LO and N3LO calculations. As the overall precision in the determination grows, it will be paramount to consistently treat uncertainties of the theoretical input cross sections. Theory uncertainties are important even in N2LO fits [208, 221, 255–257] and will be even more so as we progress towards N3LO PDFs. We refer to Section V for more details. Furthermore, fitting PDFs requires flexible and fast frameworks for the computation of N3LO cross sections. The development of such framework requires the collaboration of multiple research groups and should be supported by our field. Another key ingredient is the availability of high performance computing infrastructure that facilitate the complexity of PDF extractions, a problem that is only more complex when regarding N3LO PDFs and refers to Section IX.

### C. Electroweak Corrections in PDF fits

*Leading authors: L.A. Harland-Lang, T. J. Hobbs, E. R. Nocera, R. S. Thorne, and K. Xie*

The level of precision expected at the LHC and future colliders such as the EIC is now reaching the point where the inclusion of electroweak (EW) corrections in theoretical predictions is becoming necessary. This requires that EW corrections are applied both to the partonic cross sections and also to the PDFs. QED corrections form a significant part of this, particularly for the PDFs. These can be included by supplementing the DGLAP evolution of the PDFs to include QED parton splittings, automatically resulting in the photon becoming a constituent parton of the proton, i.e. to there being a photon distribution in the proton. This new distribution leads to photon-initiated (PI) sub-processes which enter as corrections to the purely QCD cross section for processes such as Drell–Yan [63], EW boson–boson scattering [258] and Higgs production with an associated EW boson [259]. In addition, semi-exclusive [260, 261] and exclusive PI production of states with EW couplings has significant potential as a probe of SM and BSM physics.

The inclusion of QED corrections in the DGLAP evolution equations, and the corresponding photon PDF goes back about two decades. MRST provided the first publicly available QED set [262], using splitting kernels at  $\mathcal{O}(\alpha)$  in QED and a model where the input photon was generated radiatively from the quarks below input. Subsequent sets either used similar phenomenological models [263], or constrained the photon by utilising the distinctly limited sensitivity of DIS and Drell–Yan data to the PI channel [264, 265]. This automatically led to photon PDF uncertainties of at least 10% and often considerably more. Moreover, the distinction between the elastic and inelastic photon emission was rarely considered. In [260, 266, 267] it was shown how a more accurate determination of the input photon distribution could be obtained by using the elastic form factors of the proton, which are experimentally well determined. However, as discussed long ago in e.g. [268], in fact the entire contribution from both elastic and inelastic emission to the photon PDF are directly related to the corresponding structure functions,  $F_{1,2}^{\text{el}}$ ,  $F_{1,2}^{\text{inel}}$ , as was also discussed in [269–272]. This basic idea has been provided with a precise theoretical framework by the LUXqed group [273, 274], and they were able to provide a publicly available photon PDF with uncertainties which are due overwhelmingly to those on the structure functions used as input. This approach improves the precision of the photon PDF to the level of a few percent. Moreover, QED DGLAP splitting kernels have now been calculated to  $\mathcal{O}(\alpha\alpha_S)$  [275] and  $\mathcal{O}(\alpha^2)$  [276]. These are implicit in the LUXqed approach, but also easily implemented in DGLAP evolution codes.

Hence, it is now possible to be far more precise and confident about the photon distribution, the related QED modifications to other partons, and the subsequent impact on cross section calculations. The first global PDF set



including a photon distribution based on the LUXqed approach was produced by the NNPDF group [277], and this was soon followed by QED corrected PDFs based on the MMHT14 PDFs [278]. More recently the CT group has also produced PDFs with QED corrections and a LUXqed-inspired photon distribution [279], and a set based on the MSHT20 PDFs, using an extremely similar approach to that in [278] has appeared very recently [280]. The photon distributions in these sets are all based on the same underlying principle, but have differences in the details of their methodology. They each now have uncertainties of a few percent, and are all broadly consistent with each other, despite the differences in their approach. This represents a huge improvement in the knowledge of the photon content of the proton. However, care must be taken when claiming equivalently high precision in the corresponding PI cross sections. Studies of this type are so far often based on calculations at LO in  $\alpha$ , in which case they will have significantly larger scale variation uncertainties than the percent level uncertainty due to the photon PDF. In practice, of the processes entering global PDF fits, the PI contributions to off-peak lepton pair production are by far the dominant ones. For these it is most accurate to follow the approach of [281, 282], which applies the structure function (SF) approach to directly calculate the dominant PI contribution to lepton pair production away from the  $Z$  peak. This provides percent level precision in the cross section prediction here, bypassing the issue of large LO scale variations. For many other processes, a standard EW K-factor approach can be taken (or fast interpolation grids, as presented in [283]), although in the majority of cases the impact of PI production is found to be marginal at the current level of precision.

As well as PI corrections there are other EW corrections relevant for processes used in PDF fits. For inclusive jet production it is possible to use K-factors evaluated from the calculation of [284] (see also [285]). These do not include QED corrections, and therefore PI production, arguing that the dominant contribution is from the pure weak corrections (a distinction that can be made in a gauge invariant way in this case), due to their Sudakov logarithmic enhancement. The size of the overall EW corrections, which is dominated by this source, can be as large as  $\sim 10\%$  at the highest jet  $p_\perp$  values. For  $Z$   $p_\perp$  data there is the calculation of [286]. This includes mixed  $\gamma q$  PI production, but these are found to enter at the per mille level and be significantly smaller than the other EW corrections. The total size of the EW corrections is as large as  $\sim 20\%$  at high  $p_\perp^H$  for current data, though is generally less than this [80]. For the precision  $W, Z$  data corrections can be derived from, for example, the MCSANC generator [287, 288]. The total size of the EW corrections is  $\sim 0.5\%$  at intermediate and high masses, but  $\sim 6\%$  in the lowest mass region. NLO EW corrections can also be calculated using FEWZ [289] and MG5\_aMC v3 [290]. For differential top quark pair production data EW corrections are calculated in [291] (based on the earlier study in [292]). These include the  $\gamma g$  initiated channel, calculated using the LUXqed [273] and CT18qed [279], although this contribution is found to be negligible. Top-quark pair production differential distributions which are more sensitive to EW corrections are  $p_{T,t}$   $p_{T,t\bar{t}}$ , especially at large  $p_T$ , and  $m_{t\bar{t}}$ . The rapidity distributions  $y_t$  and  $y_{t\bar{t}}$  are less sensitive although an effect can be seen at large rapidity values. Overall, the impact of EW corrections on these observables is negligible. For differential  $WH$  production, the EW correction is found to be enhanced at large invariant mass tail, mainly due to new channel initiated by  $\gamma q$  [279]. Hence, it is possible to include all significant EW corrections to processes currently involved in providing good constraints on PDFs. However, as precision and the range of processes considered both improve, further development of combined QCD and EW calculations will be necessary.

We also note that it can be useful to provide both the individual elastic,  $\gamma^{\text{el}}(x, Q^2)$ , and inelastic,  $\gamma^{\text{inel}}(x, Q^2)$ , photon PDF components, with  $\gamma(x, Q^2) = \gamma^{\text{el}}(x, Q^2) + \gamma^{\text{inel}}(x, Q^2)$ . An example is when making predictions for exclusive and semi-exclusive PI production [260, 261], although in this case care must be taken to also include the survival factor probability of no additional particle production due to multi-particle interactions (MPI). At high scales, e.g.  $Q^2 = 10^4$  GeV<sup>2</sup>, the inelastic component is dominant until very high  $x$ , while at lower scales, e.g.  $Q^2 = 10^2$  GeV<sup>2</sup>, the relative contribution from the elastic component is somewhat larger, due to the shorter evolution length for (inelastic)  $q \rightarrow q\gamma$  splitting. It is also important to consider a consistent set of QED corrected neutron PDFs as well as just those for the proton. Neutron PDFs are necessary for a consistent fit to deuteron and nuclear fixed target data from neutrino ( $\nu N$ ) DIS scattering experiments used to constrain the PDFs. The QED corrected neutron PDFs automatically provide isospin violating partons, with  $u_{(p)} \neq d_{(n)}$ , and these were seen in [262] to automatically reduce the NuTeV  $\sin^2 \theta_W$  anomaly [293]. The breaking of isospin symmetry may also have implications for the development of nuclear PDFs, in particular at the EIC. We also note that as well as considering the photon as a component of the proton it is also possible to include leptons [294], and electroweak bosons [295, 296]. The effect of the former is very small for almost all processes, while the latter may have more significance at a future very high energy collider.

## D. PDFs and resummations at extreme momentum fractions

### 1. Large $x$

*Leading authors: A. Courtoy, D. Soper, M. Ubiali*



The large- $x$  region is the least known PDF kinematic region, as the number of experimental data that constrain  $x \gtrsim 0.1$  are less in number and less precise than those that constrain the small and medium- $x$  regions. The resulting large- $x$  PDF uncertainties hamper the precision of both the signal and background theoretical predictions in the high energy tails, which are the focus of both direct and indirect searches for new physics (see Sect. II B 2). It is therefore crucial for hadron collider phenomenology to pin down the large- $x$  region.

Parton distributions at  $x > 0.1$  are also of a special interest for theoretical studies, as they can be increasingly connected to nonperturbative and lattice QCD approaches. These theoretical techniques can be assessed by comparisons against precisely known unpolarized collinear PDFs found from phenomenological analyses and then expanded to predict less experimentally accessible quantities such as spin-dependent PDFs.

The present data impose few little experimental constraints at  $x > 0.5$ , where various factors may introduce corrections to the simplest collinear factorization framework. Much of the relevant data lie at low  $Q$ , close to the lower boundary of the validity region for perturbation theory. Several groups develop frameworks to account for corrections (nuclear, target mass, higher twists) that affect extraction of nucleon collinear PDFs in the large- $x$  and low- $Q$  region, *e.g.* [4, 297]. Interplay between these corrections and determination of PDFs at large  $x$  will be increasingly relevant in near-future precision experiments [9, 103, 104]. It should be pointed out that, while these types of corrections are most pronounced at low  $Q$  and very large  $x$ , they propagate to smaller  $x$  at electroweak  $Q$  via DGLAP evolution and may affect percent-level phenomenology. Large- $x$  contributions might also matter for specific kinematics, see, *e.g.*, [24, 298].

In the realm of perturbative QCD, it is well-known that fixed-order perturbative calculations, even when computed at N2LO in  $\alpha_s$ , display classes of logarithmic contributions that become large in some kinematic regions, thus spoiling the perturbative expansion in the strong coupling constant  $\alpha_s$ . Among these enhanced logarithmic contributions, there are the high-energy (or small- $x$ ) contributions that will be discussed in the next section. Here we will focus on another type of logarithmic enhancement of higher order perturbative contributions that is relevant at large  $x$  [299]. This class of logarithms appears close to threshold for the production of the final states: this is the large- $x$  kinematic region, and the resummation of logarithms from this region is known as large- $x$ , soft gluon, or threshold resummation. The importance of these contributions varies significantly with both the type and the kinematic regime of the processes which enter PDF fits. Therefore, their omission can lead to a significant distortion of the PDFs, thereby reducing their theoretical accuracy. Some time ago, in Ref. [298] a set of PDFs was constructed in which fixed-order NLO and N2LO calculations were supplemented with soft-gluon (threshold) resummation up to next-to-leading-log (NLL) and next-to-next-to-leading-log (NNLL) accuracy respectively. This specialized set of PDFs was produced to be used in conjunction with any QCD calculation in which threshold resummation is included at the level of partonic cross sections. These resummed PDF sets, based on the old NNPDF3.0 analysis [300], were extracted from a restricted set of data, namely DIS, Drell-Yan, and top quark pair production data, for which resummed calculations were available in a usable format. The interesting result was that, close to threshold, the inclusion of resummed PDFs can partially compensate the enhancement in resummed matrix elements, leading to resummed hadronic cross-sections closer to the fixed-order calculation. On the other hand, far from threshold, resummed PDFs reduce to their fixed-order counterparts. This pointed to the need for a consistent use of resummed PDFs in resummed calculations.

Within the context of parton shower event generators, the need to sum threshold logarithms arises from a mismatch between the kinematic limits in the evolution of parton distribution functions and the evolution of the parton shower. In part, this means that one should use different PDFs within the splitting functions of the parton shower than the usual  $\overline{\text{MS}}$  PDFs used for fixed order perturbation theory [301–303]. The most practical way to do this is to transform the  $\overline{\text{MS}}$  PDFs, but a more ambitious solution would be to independently fit the PDFs in the needed scheme. For more details, see the discussion in Section IV F.

Once the corrections of perturbative QCD are properly accounted for, the obtained PDFs should be consistent with their field-theoretical definition. An interesting question is then to which degree the theoretical expectations, such as positivity, quark counting rules, or quark-hadron duality, must influence the shape of phenomenological PDFs [304, 305]. Should the allowed PDF solutions reflect these semi-quantitative constraints? This is a topic of the recent phenomenological work [306, 307] and exploration within global fits [7, 10]. While first principles of QCD need to be fulfilled, empirical testing of various hypotheses for the hadron structure must be mindful of biases introduced by such prior expectations. On the flip side, without the control of associated uncertainties, agreement between a theoretical model and phenomenological PDFs is not sufficient for validating the model; detailed studies of uncertainties in such tests are crucial both on the theoretical and phenomenological side [307]. Anticipated DIS and other measurements at higher  $x$  and  $Q$  values will advance our knowledge of large- $x$  dynamics [308, 309].

## 2. Small $x$

*Leading authors: R. D. Ball, M. Hentschinski, C. Royon, K. Xie*

For successful runs at any colliders, such as the LHC at CERN or the incoming EIC at BNL [27], and future projects such as FCC at CERN [145], it is fundamental to understand fully the complete final states. This obviously includes the central part of the detector that is used in searches for beyond standard model physics but also the forward part of the detector, the kinematic region close to the outgoing particles after collision. The detailed understanding of final states with high forward multiplicities, as well as those with the absence of energy in the forward region (the so-called rapidity gap), in elastic, diffractive, and central exclusive processes is of greatest importance. Some of these configurations originate from purely nonperturbative reactions, while others can be explained in terms of multi-parton chains or other extensions of the perturbative QCD parton picture such as the Balitsky-Fadin-Kuraev-Lipatov (BFKL) formalism [310–312]. Future progress in this fundamental area in high energy physics requires the combination of experimental measurements and theoretical work.

When the parton momentum fraction  $x$  becomes small, small- $x$  logarithms  $\log(x)$  become significant, and require all-order resummation to obtain a good convergence of the QCD theory. This can be achieved through the BFKL formalism at NLL [313–316], matched to collinear factorization at NLO or N2LO using either the ABF formalism [317–322], or the closely related CCS approach [323–329]. An efficient numerical implementation of the ABF results [330, 331] made it possible to perform a global PDF determination, based on the NNPDF3.1 data set, but also resumming small  $x$  logarithms in parton evolution and structure functions coefficients [15]. This analysis found significant evidence for BFKL resummation in the small  $x$  and low  $Q^2$  region of the HERA structure function data [154]. An analysis using the same ABF implementation and xFitter reached a similar conclusion [332]. Future work on using small- $x$  resummation to improve PDF fits will require the high energy resummation of the hadronic cross-sections [333–342] included in global fits, which while technically challenging is now perfectly feasible. The effects are most likely to be important in LHCb data, which can probe  $x$  as small as  $10^{-6}$ .

Eventually, at small enough  $x$  and low enough  $Q^2$ , we enter into the partonic saturation region [343]. The boundary to delineate the small- $x$  resummation region and the saturation one is ambiguous. In the latest round of the CTEQ-TEA global analysis [7], two alternative ensembles, CT18X and CT18Z, were released, in which an  $x$ -dependent DIS factorization scale was adopted. It is motivated by a partonic saturation model [344], and improves the QCD description of the HERA DIS data, obtaining a similar  $\chi^2$  for the same data set as the small- $x$  resummation treatment adopted in NNPDF [15] and xFitter [332]. Both approaches obtain an enhancement of gluon PDF at small  $x$ , which dies out with energy increasing [345]. However, the enhancement of the small- $x$  resummation is noticeably larger than the  $x$ -dependent scale approach, in which the small- $x$  growth is largely tamed toward  $x \rightarrow 10^{-6}$ . Some implications about the small- $x$  dynamics have been explored, such as the DIS structure functions [345]. As expected, both approaches have obtained similar predictions for the transverse structure functions  $F_2$ . But surprisingly, the longitudinal one  $F_L$  is pulled to different directions when Bjorken  $x$  below  $10^{-4}$ . It would be very interesting to see future measurements to discriminate between these two distinct approaches.

It is obvious that the PDF fits at small  $x$  will benefit from a better understanding of multi-gluon kinematics such as in the BFKL regime or the saturation phenomena that might appear at very small  $x$  especially in heavy-ion collisions. The understanding of diffractive events and their effects on PDFs is also fundamental. Some recent developments in the domain of small  $x$  physics, saturation and diffraction as well as future insights are presented in the dedicated white paper [346]. This document discusses first the occurrences of BFKL resummation effects in special final states, such as Mueller-Navelet jets, jet gap jets, and heavy quarkonium production. It further addresses TMD factorization at small  $x$  and the manifestation of a semi-hard saturation scale in (generalized) TMD PDFs. More theoretical aspects of low  $x$  physics, probes of the quark gluon plasma, as well as the possibility to use photon-hadron collisions at the LHC to constraint hadronic structure at low  $x$ , and the resulting complementarity between LHC and the EIC are also presented. We also briefly discuss diffraction at colliders as well as the possibility to explore further the electroweak theory in central exclusive events using the LHC as a  $\gamma\gamma$  collider.

### E. Theoretical Developments beyond QCD and Electroweak at Fixed Order

Beyond the aforementioned theoretical developments we identify briefly several key aspects that are required to further improve our knowledge of parton distribution functions.

- Effects due to non-zero quark masses become non-negligible at a certain level of precision and a consistent framework to take them into account in the extraction of PDFs is desirable. While several General-Mass-Variable-Flavor-Number-Scheme (GM-VFNS) calculations have been implemented in PDF global fits for DIS observables [181, 182, 347], all hadronic observables are computed in a Zero-Mass VFNS, thus ignoring the effects

associated with the finite mass of heavy quarks. This is currently justified by the larger energy scales associated with most LHC observables, as compared to the charm and bottom quark masses. However as the precision target increases and as lower energy regions are explored the implementation of  $pp$  GM-VFNS calculations might become necessary in global PDF fits.

- In the kinematic limits of a parton taking up almost all or almost none of the momentum of its hadron, parton distribution functions may be resummed to all orders in perturbation theory to a given logarithmic accuracy. Consistently including such resummation should be part of future research and ultimately the determination of PDFs. At small values of parton momentum fractions  $x$  the resummation of small- $x$  corrections to a given logarithmic accuracy to all orders has been considered. It has been shown [348, 349], however, that the yet unknown sub-leading small- $x$  terms are larger than the leading order terms and the entire tower of sub-leading terms is needed, at least to the fourth sub-leading logarithm, to consolidate this problem [211, 215]
- Fast interfaces of QCD, electroweak and resummation contributions are crucial for the extraction of PDFs and their development should be facilitated. This is partially discussed in Section IX.
- Studies of theoretical uncertainties and their propagation through the PDF extraction process should be encouraged.
- The distribution of final PDF parameterizations in a convenient form for applications, such as the LHAPDF format, is important to the usability of PDFs for the community.
- Beyond fixed order perturbation theory and the leading power expansion for observables in Eq. (4) several improvements of the theoretical description are compulsory, depending on the observable under consideration, and additional care has to be taken. For the kinematics range covered by currently available data from DIS  $ep$  scattering, higher twist effects become important. In the flavor non-singlet case these have been measured in [350, 351] and in the singlet case they were determined in [352].

## F. Factorization schemes for event generators

*Leading authors: S. Hoeche, A. Siodmok*

Defining a PDF requires the choice of a factorization scheme, which governs the allocation of finite terms between the PDFs and the hard, partonic cross sections. This choice is generally a matter of taste and convenience [353]. In practice, the majority of QCD hard process calculations and PDF sets adopt the  $\overline{\text{MS}}$ -scheme.

Recently there has been renewed interest in developing alternative factorization schemes [304, 354–359], including to investigate the positivity of  $\overline{\text{MS}}$  PDFs [304] and to simplify Monte Carlo calculations [356–359]. PDFs in different factorization schemes are related to each other, and to those in the  $\overline{\text{MS}}$  scheme, by a transition operator that mixes PDFs of different flavours,  $\mathbf{f}^{\text{FS}} = \mathbb{K}^{\overline{\text{MS}} \rightarrow \text{FS}} \otimes \mathbf{f}^{\overline{\text{MS}}}$ , so that for each flavour  $a$  we have

$$f_a^{\text{FS}}(x; \mu_{\text{F}}) = \sum_b \int_x^1 \frac{d\xi}{\xi} \mathbb{K}_{ab}^{\overline{\text{MS}} \rightarrow \text{FS}} \left( \frac{x}{\xi}; \mu_{\text{F}} \right) f_b^{\overline{\text{MS}}}(\xi; \mu_{\text{F}}),$$

where

$$\mathbb{K}_{ab}^{\overline{\text{MS}} \rightarrow \text{FS}}(x; \mu) \equiv \delta_{ab} \delta(1-x) + \frac{\alpha_s(\mu)}{2\pi} K_{ab}^{\text{FS}}(x; \mu) + \mathcal{O}(\alpha_s^2).$$

The transformation kernels are often further constrained to ensure that the transformed PDFs obey the same sum rules as the input PDFs (e.g. re-imposing momentum sum rules by modifying an end-point contribution  $\propto \delta(1-x)$  accordingly). The required independence of predictions from the choice of factorization scheme is achieved, to NLO accuracy, by a corresponding inverse transformation of the partonic cross-sections. The freedom to choose a factorization scheme therefore corresponds to a freedom to remove a common set of convolution terms  $K_{ab}^{\text{FS}}$  from all partonic cross-sections.

The **Krk** (formerly **MC**) factorization scheme [360] exploits this freedom to significantly simplify the matching of the parton shower Monte Carlo event generators to NLO calculations for the hard process by systematically removing the convolution terms  $K_{ab}^{\text{FS}}$  from the hard process. This may be conveniently done within the modified Catani-Seymour (CS) dipole subtraction method [361]. Within the **Krk** scheme the transition operator is therefore derived from the

finite and collinear part of the CS integrated-subtraction and collinear contributions, given by the  $\mathbf{P}$  and  $\mathbf{K}$  collinear operators, so that  $\mathbf{K}$  objects as

$$\mathbf{K}_{ab}^{\text{Krk}}(x; \mu) = \mathbf{K}^{ba}(x) + \mathbf{P}^{ba}(x; \mu).$$

These transition operators, in the  $\text{Krk}$  factorisation scheme are modified such that the NLO corrections to the heavy colour-neutral boson production in pp collisions (Drell-Yan type processes) and electron-proton scattering (DIS-type processes) are maximally simplified [356–358] and they have been applied to several public  $\overline{\text{MS}}$  PDF sets. These PDFs, uniquely, allow NLO-accurate calculations of any such process using CS dipole subtraction without requiring an on-the-fly convolution.

The  $\text{Krk}$  factorisation scheme has been employed in the  $\text{KrkNLO}$  parton-shower matching method, which has been implemented as a proof-of-concept for the Drell-Yan and (gluon-fusion) Higgs-production processes in both *Sherpa* and *Herwig*[358].

## V. METHODOLOGY

### A. Experimental systematic uncertainties in PDF fits

*Leading authors: A. M. Cooper-Sarkar, T. Cridge, F. Giuliani, J. Huston, R. S. Thorne*

The LHC has accumulated a large amount of data at 7,8 and 13 TeV, for a variety of processes. The data sets vary from purely inclusive processes, such as the  $W/Z$  cross sections, to differential measurements over a variety of kinematic variables, such as the Drell-Yan cross section as a function of the invariant mass, rapidity and transverse momentum of the final state leptons. Due to the large data samples, many of the measured distributions are limited by systematic uncertainties rather than statistics. Differential measurements over wide kinematic ranges (and over multiple detector regions) that are systematics-limited require detailed knowledge regarding the correlation of the systematic error components over these regions. Such error correlations are difficult to determine experimentally and their imperfect knowledge often results in tensions between rapidity ranges (for example for the case of the ATLAS inclusive jet cross section) or kinematic variables (for example for the ATLAS  $t\bar{t}$  rapidity and  $t\bar{t}$  mass distributions). Such tensions may mask or diminish the power of the data to determine PDFs and their uncertainties.

To be specific, many systematic uncertainties are point to point correlated within a kinematic distribution and between distributions of the same analysis. There can also be correlations between different analyses due to systematic uncertainties from the same sources. For example, inclusive jet production data are presented as functions of transverse momentum in several bins of jet rapidity and many systematic sources are correlated between rapidity bins. Alternatively  $t\bar{t}$  production data are presented in terms of several different variables such as the mass,  $m_{t\bar{t}}$ , or rapidity,  $y_{t\bar{t}}$ , of the  $t\bar{t}$  pair, and the average transverse momentum,  $p_T^t$ , or rapidity,  $y_t$ , of the  $t\bar{t}$  pair. There are both statistical and systematic correlations between all of these distributions. Finally since the  $t\bar{t}$  data are measured in the lepton+jets channel, there are potential correlations between the systematic uncertainties, from sources such as the jet energy scale, between the inclusive jet measurements and the  $t\bar{t}$  measurements.

Experimental correlated systematic uncertainties can be taken into account in PDF fits by using a covariance matrix provided by the experimental collaboration, but it is more informative if this information is given as a list of  $1\sigma$  uncertainties due to each source of systematic uncertainty for each data point. This information can then be applied to the fit using nuisance parameters which are common between the data points for the same source. It is the default to consider the same source of systematic uncertainty to be 100% correlated between the data points and this is the assumption used when constructing a covariance matrix, but this may not be realistic. The advantage of keeping the information split into separate sources is that one can trace the sources of uncertainty and one can change the degree of correlation in an informed manner by consulting the experimentalists.

Problems with the treatment of correlated systematic uncertainties first came to light in fits to the ATLAS 7 TeV jet data [71]. Whereas good fits could be found to the separate rapidity bins of the data, the fit to all rapidity bins taken together was very poor. The tensions result in  $\chi^2$  values that may be acceptable for individual rapidity regions (for the jet fit), but have vanishing probability when fit together. Such tensions may mask or diminish the power of the data to determine PDFs and their uncertainties. The information provided by the separate individual rapidity interval fits greatly reduce the discriminating power of the full data set, as (1) the  $x$ -range probed is reduced and (2) the systematic error shifts may differ significantly from distribution to distribution, a situation that does not reflect reality <sup>3</sup>

<sup>3</sup> Fits to a wider rapidity range for jet production also help to distinguish between PDF variations and the possible presence of new

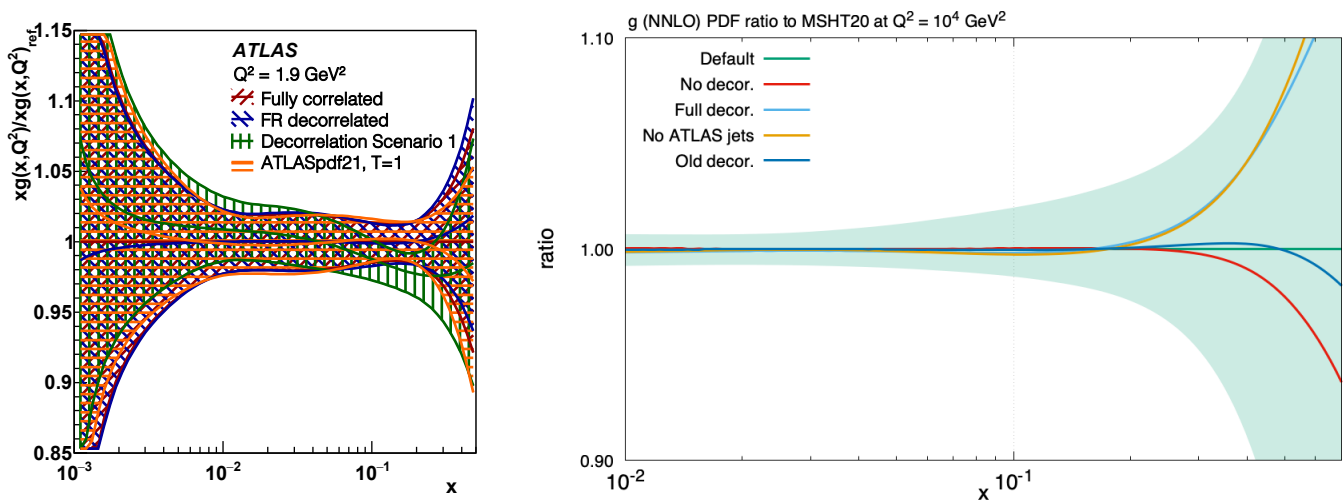


FIG. 9. Difference in the gluon PDF shown in ratio to the ATLASpdf21 (default) gluon(left). This default uses Decorrelation Scenario 2 and this is compared to the use of Full Correlation, Full decorrelation of the flavour response systematic and Decorrelation Scenario 1. The effect of no decorrelation, the default correlation of [9], the decorrelation in [362], and full decorrelation for the MSHT20 gluon (right).

A data set, such as the ATLAS jet cross section, could be divided into its individual rapidity intervals to determine any tensions that may exist, for example in the determination of the high- $x$  gluon distribution, and how these tensions and the constraining power on the PDFs change as decorrelation models are applied. In this context, a bad data set  $\chi^2$  may not necessarily represent a disappointing outcome, if the data set's constraining power is not reduced. In addition, it can be checked whether a decorrelation model that improves the global  $\chi^2$  affects the impact of this data set on the PDF fit.

This problem with fitting multiple rapidity intervals led NNPDF to fit only one rapidity bin [108]. Decorrelation models have been developed to reduce these tensions, but suffer from their somewhat ad hoc nature. MMHT made an alternative study of the effect of decorrelating some systematic sources between rapidity bins [362]. However, the most thorough study was made by the ATLAS collaboration, who studied the same problem in their ATLAS 8 TeV inclusive jet data [72]. Some of the systematic uncertainties pertaining to the jet energy scale are evaluated from the difference of two different Monte-Carlo estimates. Such "two-point systematics" are reasonable estimates of uncertainty, but they are far from being Gaussian distributed. These systematic sources are often the largest systematic uncertainties for analyses involving jet production. One may question the convention that these are 100% correlated between data points. However, it is vital to do this in collaboration with experimentalists with knowledge of which sources can be legitimately decorrelated. ATLAS developed some models for the decorrelation of such systematic sources as functions of rapidity and  $p_T$  [72]. These models were applied to several of the jet energy scale systematic sources and some favoured combinations of correlation model were suggested. These were then used in a PDF fit using these jet data [11]. The  $\chi^2/NDP$  for fits to these jet data with different levels of decorrelation are summarised in Table II. Whereas the  $\chi^2$  differ considerably, the difference in the resulting PDFs between the use of full correlation and extreme decorrelation is small, see Fig 9. The study on ATLAS 7 TeV jets associated with the MSHT20 PDFs [9] comes to a similar conclusion when decorrelating two of the jet energy scale systematic between rapidity bins (partial decorrelation) or indeed decorrelating all systematic sources between rapidity bins (full

ATLAS 8 TeV Jets R=0.6 Fully Correlated	FR Decorrelated	Decorrelation Scenario 1	Decorrelation Scenario 2
$\chi^2/NDP$	289/171	226/171	250/171
			248/171

TABLE II. Partial  $\chi^2$  for jet data entering the PDF fit, for different levels of decorrelation ranging from fully correlated to an extreme scenario of the jet flavour response (FR) decorrelated between rapidity bins. The Decorrelation Scenarios are described in ref. [11].



decorrelation), see Fig 9.

Although studying the impact of various experimental systematic decorrelations can very useful in general, such decorrelation models should be vetted by the original experimental collaboration, as they are the ones who best understand the origins of those systematics.

The second example of the need to consider some degree of decorrelation of 'two-point systematics' comes in fits to ATLAS 8 TeV  $t\bar{t}$  spectra [81]. When these data were first issued there was no information statistical correlations between the spectra such that only one spectrum could be fit at once [108]. However, such information was provided in ref. [363] and is now available in HEPDATA for Ref. [81], and a study of using both systematic and statistical correlations within a PDF fit was made. The  $\chi^2/NDP$  for separate fits to the the lepton+jets spectra are given in Table III, where it can be seen that the rapidity spectra cannot be fitted well. Further study of fitting the spectra simultaneously was then restricted to the  $m_{t\bar{t}}$  and  $p_T^t$  spectra. Since the  $\chi^2$  of the separate fits to  $p_T^t$  and  $m_{t\bar{t}}$  adds to 11.3 it was somewhat surprising that a fit with all sources of systematic uncertainty 100% correlated between these sources yields a joint  $\chi^2$  of 45. The answer lies in the correlation of the large two-point systematics related to the models for parton showering, hard scattering and initial/final state radiation. When the spectra are fitted separately, the nuisance parameters for these sources take very different values for the  $p_T^t$  and  $m_{t\bar{t}}$  spectra, see Table IV, but an assumption of 100% correlation forces them to be the same– this suits neither spectrum. A fit in which all three of these sources of systematics are decorrelated between the two spectra or a fit in which just the parton shower model sources is decorrelated produce considerably lower  $\chi^2$ , see Table V. The decorrelation of the parton shower systematic has been adopted for the ATLAS PDF fits and for the CT18 PDF fits. Indeed, it has been confirmed recently that as well as ATLAS, all of CT, MSHT and NNPDF find problems fitting all distributions simultaneously without some decorrelation [36]. The effect of this decorrelation on the gluon PDF is fortunately small as illustrated with the ATLASepWZtop18 fit [363] in Fig. 10. However, a study by MSHT [8] took the decorrelation further. In order to fit the  $y_{t\bar{t}}$  and  $y_t$  rapidity spectra, decorrelation of the parton shower systematic within these spectra is also considered. This decorrelation is done as a trigonometric function of rapidity. This reduces the  $\chi^2$  per point for this data set with all for distributions in the MSHT20 fit from 6.84 with no correlation to 1.69 with correlation between distributions to 1.04 for the additional decorrelation within spectra. In this case, although the difference in the gluon PDF between the fully correlated and uncorrelated case is still within uncertainties it is nevertheless comparable to the difference between and NLO and an N2LO analysis. This approach has been carried into the full MSHT20 analysis [9], see Fig. 10. It should also be noted that the more decorrelation is applied the less power the data have to constrain PDFs.

Finally, although some sources of systematics can be legitimately decorrelated between spectra of the same analysis, there are other systematic sources for which correlation between different analyses should be considered. This has been studied in a recent ATLAS PDF analysis [11] ATLASpdf21, where the correlations of various systematic sources have been considered between different analyses which use jet data: inclusive jet data [72],  $t\bar{t}$  data in the lepton+jets channel [81],  $W$  +jets data [78] and  $Z$ +jets [79] data. The details of the correlated systematic sources considered are given in ref [11]. Fig. 11 shows the difference in the resulting gluon and  $x\bar{d}$  PDFs when such correlations between the input data sets are considered and when they are not. Note that this figure is made for the scale  $Q^2 = 10,000\text{GeV}^2$  to illustrate that such differences are still visible at LHC scales.

In conclusion correlations of sources of systematic uncertainty both within and between data sets need to be carefully considered in PDF fits and although the difference between the resulting PDFs is not large in the best known kinematic region  $0.01 < x < 0.1$  (corresponding to mass scales  $\sim 100 \text{ GeV} \rightarrow 1 \text{ TeV}$  at the LHC) it can nevertheless be large enough to have impact if an ultimate precision of  $\sim 1\%$  is sought on PDFs. In the less well known regions, at smaller and larger mass scales, the impact can be considerably greater.

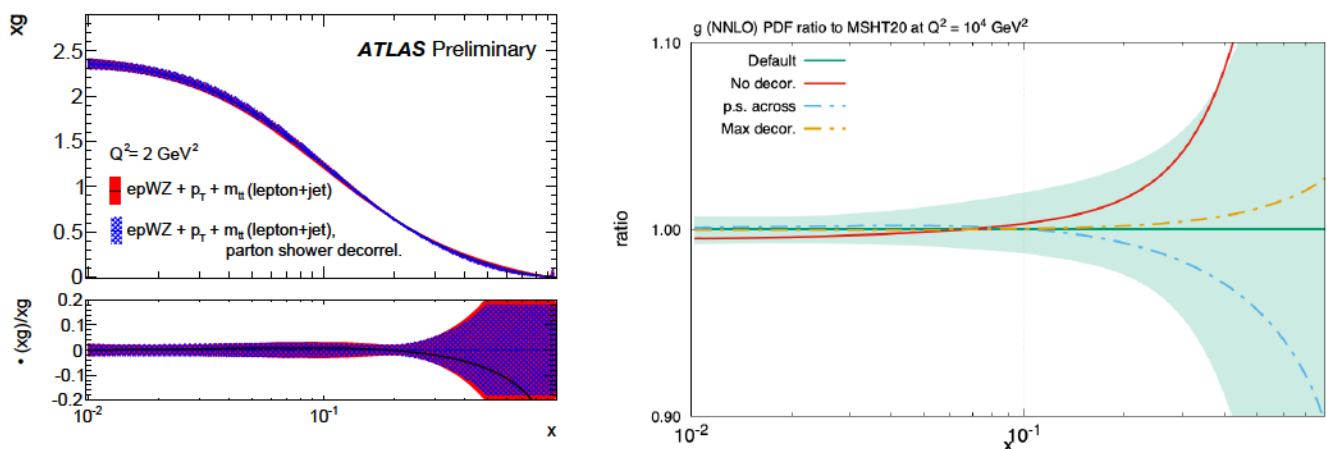
ATLAS 8 TeV $t\bar{t}$ lepton+jets spectrum	$m_{t\bar{t}}$	$p_T^t$	$y_{t\bar{t}}$	$y_t$
$\chi^2/NDP$	3.4/7	7.9/8	19.7/5	18.3/5

TABLE III. Partial  $\chi^2$  for data sets entering the PDF fit, for each of the top spectra separately.

ATLAS 8 TeV $t\bar{t}$ spectra	$p_T^t$	$m_{t\bar{t}}$
hard scattering model	+0.74	-0.43
parton shower model	-1.32	+0.39
ISR/fsr model	-0.47	+0.33

TABLE IV. Shifts of the named nuisance parameters, in units of standard deviations, for the fits to the top spectra separately

ATLAS 8 TeV $t\bar{t}$ spectra	$p_T^t$ and $m_{t\bar{t}}$	$p_T^t$ and $m_{t\bar{t}}$	$p_T^t$ and $m_{t\bar{t}}$
	fully correlated	3 sources decorrelated	parton shower decorrelated
$\chi^2$ /NDP	45/15	11.5/15	14.1/15

TABLE V. Partial  $\chi^2$  for data sets entering the PDF fit for simultaneous fits to the  $p_T^t$  and  $m_{t\bar{t}}$ .FIG. 10. Difference (left) in the gluon PDF shown in ratio to the ATLASepWZtop18 gluon (called  $epWZ+p_T^t+m_{t\bar{t}}$ ). The use of full correlation of all systematic sources (red) is compared to the result when decorrelating the parton shower systematic between the  $p_T^t$  and  $m_{t\bar{t}}$  spectra (blue). The difference (right) in the default, no decorrelation, decorrelation only across distributions and full decorrelation for the gluon PDF of [9].

## B. Theoretical uncertainties in PDF fits

*Leading authors: R.D. Ball, A. M. Cooper-Sarkar*

Over the last few years there has been considerable progress in developing new techniques for incorporating theoretical uncertainties into the determination of parton distribution functions (PDFs). This work centres on a Bayesian formalism, the “theory covariance matrix”, which can be simply added to the usual experimental covariance matrix used in the PDF fit [364]. While the experimental covariance matrix includes the statistical and systematic uncertainties in the measurement of a given cross-section, the theoretical covariance matrix incorporates all the various theoretical uncertainties, correlated across different experimental measurements, in the procedure which extracts the PDFs from a global data set. The main assumption is that the theoretical uncertainties are Gaussian, and independent of the experimental uncertainties (likewise also generally assumed Gaussian).

The theory covariance matrix formalism was first applied to the incorporation of nuclear uncertainties in PDF fits, firstly to data taken on heavy nuclear targets [365, 366], then to data taken on deuteron targets [104, 367]. The prior in these examples was determined empirically, through fits to nuclear data. These techniques were then used to incorporate nuclear uncertainties into the NNPDF4.0 fit [10, 368].

More challenging was to apply the new techniques to the incorporation of missing higher order uncertainties (MHOU). Here the prior was purely theoretical, determined using scale variations, taking great care to correctly

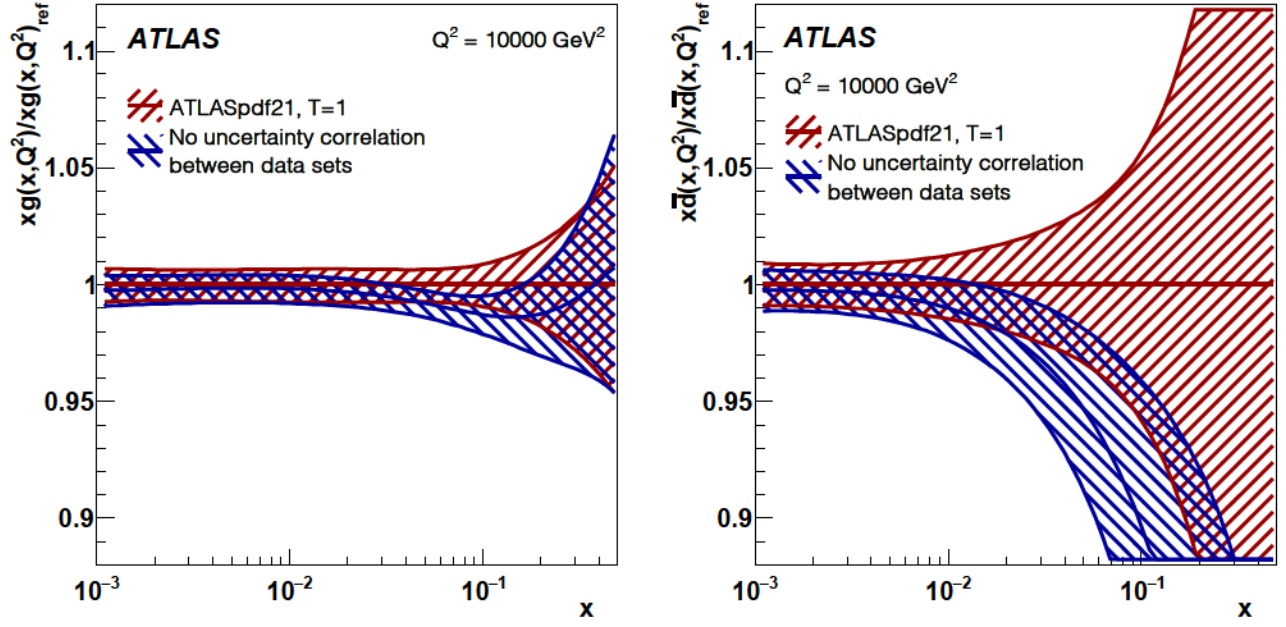


FIG. 11. Difference in the gluon and the  $x\bar{d}$  PDF shown in ratio to the ATLASpdf21 (default) PDFs. The default (red) analysis applies the full correlation of specified systematic sources between the data sets which use jet data and the alternative (blue) analysis does not apply any correlation of systematic sources between the data sets (apart from the luminosities).

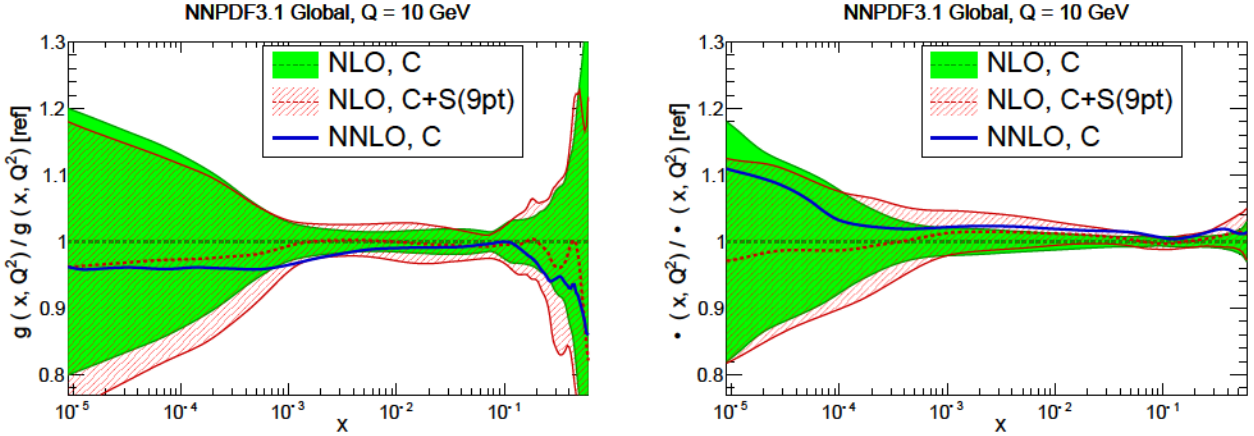


FIG. 12. Ratio of the gluon (left) and singlet (right) PDFs for the NLO NNPdf3.1 fit including factorization and renormalization scale uncertainties combined in a 9pt scheme (red), to a fit not including these scale uncertainties (green). Also shown (in blue) is the central value for the N2LO fit (without scale uncertainties).

incorporate the correlations between the different processes used in the PDF fit. As a test of principle the new formalism was applied to the global NLO NNPdf3.1 fit [369, 370], using a variety of scale variation schemes (5pt, 7pt and 9pt) for variations of the renormalization and factorization scales, and the result compared to the N2LO fit. The outcome was that while overall uncertainties in the NLO fit were only increased a little, interesting shifts were found in central values due to the rebalancing of the data sets in the fit (deweighting the data sets associated with large theoretical uncertainties in the PDF extraction), thereby taking the NLO results closer to the N2LO (see Fig. 12).

There was some debate concerning the appropriate way to use PDFs with MHOUs, since the MHOUs in the PDF might be correlated with the MHOUs in the matrix elements used in the prediction [255]. This issue was solved by computing the correlations explicitly [371], finding that though the correlations can often be ignored (as argued in [369, 370]), in some circumstances they could lead to significant improvements in the accuracy and precision of predictions (see Fig. 13). The same correlation machinery could be used to incorporate correlations with PDF uncertainties in extraction of physical parameters, such as  $\alpha_s$ ,  $m_W$  or  $\theta_W^{\text{eff}}$ : using PDF sets with a range of fixed

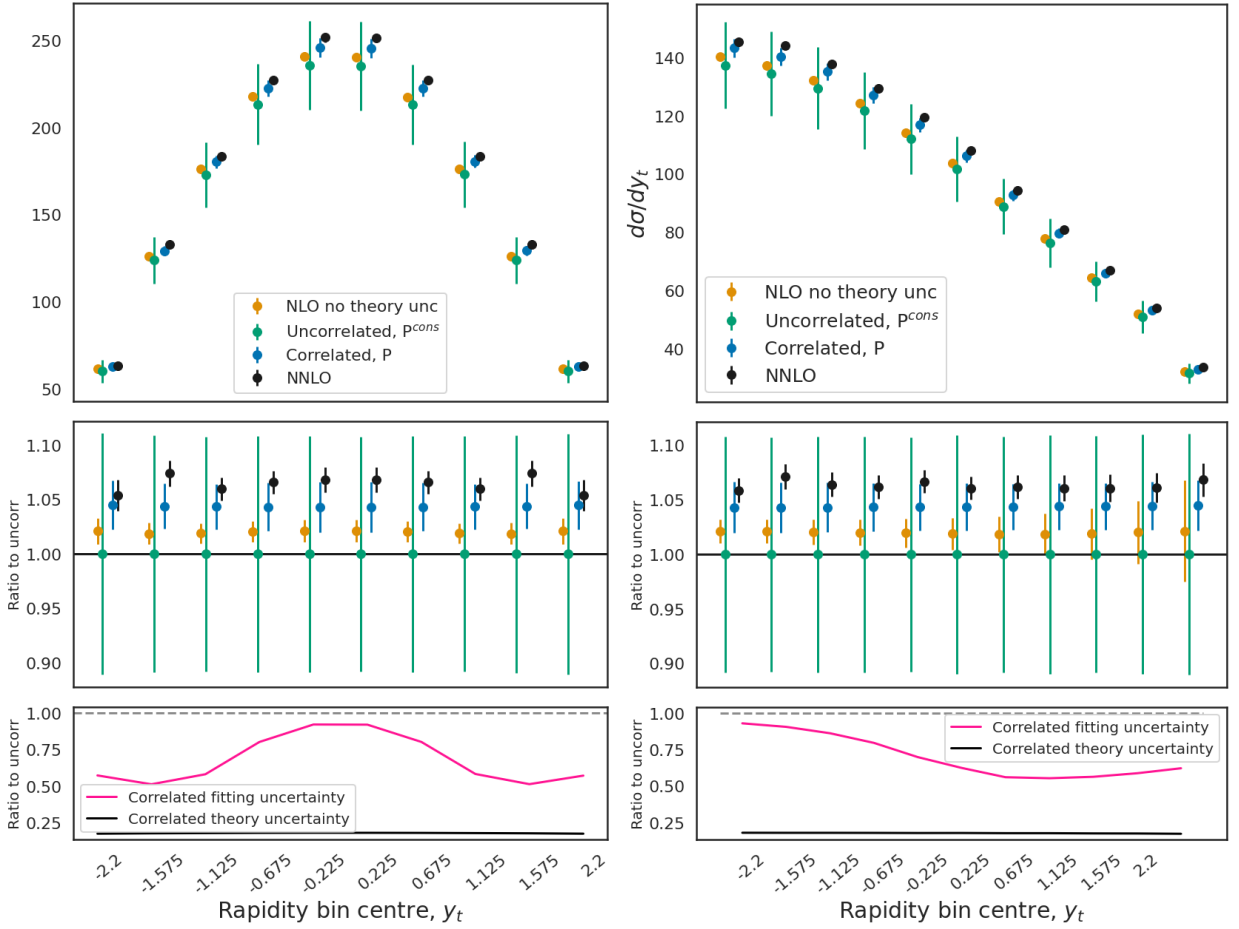


FIG. 13. Predictions for  $t\bar{t}$  unnormalized rapidity distribution data taken at 13 TeV by CMS, the dilepton rapidity distribution (left) and the lepton+jets distribution (right). The four predictions show: the NLO NNPDF3.1 fit with no MHOU, PDF error only; the NNPDF3.1 fit with MHOU, and with MHOU in the prediction, but ignoring correlations; the same, but including the correlations between MHOU in fit and prediction; and the N2LO result with no MHOU. In the middle panels the same results are shown, but normalized to the uncorrelated result. In the lower panels we show the fractional reduction in the PDF uncertainty and the theory in the prediction due to the inclusion of the correlations. Note that this the NNPDF3.1 global fits include data for the  $t\bar{t}$  total cross-section, but not the rapidity data, so in this example the correlations are particularly strong.

physical parameters is not sufficient for parameter extraction, since it does not include the correlation [372].

Theoretical MHOU were recently included in a N2LO PDF determination by the ATLAS collaboration [11]. Here fully correlated variations of renormalization and factorization scales in the calculation of  $K$ -factors for inclusive  $W$  and  $Z/\gamma^*$  production at 7 and 8 TeV were combined in a 5pt scheme, and added to the experimental uncertainty in order to estimate the MHOU. The impact of this is illustrated in Fig. 14.

In the near future, it is hoped that the next update of the NNPDF4.0 fit [10, 368], NNPDF4.1, will include MHOU, correlated across all processes in the global N2LO fit. This will use the newly developed PDF evolution code EKO [234, 235], which since the evolution kernel is in Mellin space, allows straightforward application of clean factorization scale variation, and comparison of different truncations of the evolution equations which are equivalent up to MHO corrections. One advantage of fits with MHOU is that it allows the inclusion of processes known only at NLO in an N2LO fit, the MHOU taking care of the appropriate deweighting. It will also facilitate the development of N3LO fits, in which the parton evolution will be at (approximate) N3LO, but the processes included in the fit will use a mixture of N2LO and N3LO predictions. These developments would be sufficient to give a global determination of the strong coupling at N2LO incorporating directly the MHOU, and then allow this determination to be repeated at N3LO.

There has also been progress recently in the development of methods which do not rely on scale variation to estimate MHOU [373–376]. These methods generally give rise to priors which are not Gaussian, making it difficult to incorporate them in the theory-covariance matrix formalism (although it should be noted that this problem also exists when estimating experimental systematics, so is perhaps not insurmountable).



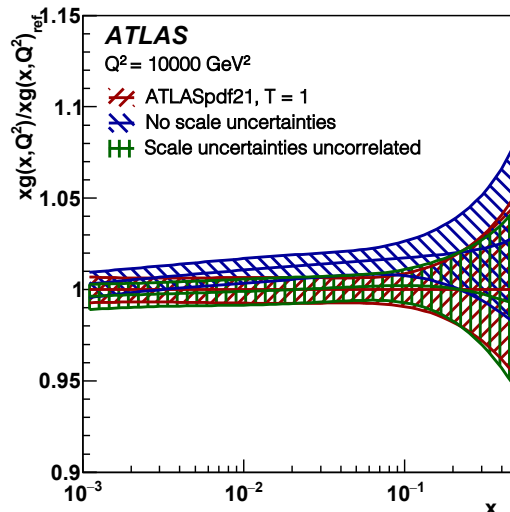


FIG. 14. Ratio of the gluon PDF for the ATLASpdf21 fit including N2LO scale uncertainties for the inclusive  $W, Z$  production at 7, 8 TeV (red) with scale uncertainties correlated between  $W$  and  $Z$  and between 7 and 8 TeV data, to a fit not including these scale uncertainties (blue) and a fit including them but not correlating them between 7 and 8 TeV data (green).

### C. Machine learning/AI connections

*Leading authors: S. Carrazza, J. Cruz-Martinez, M. Ubiali*

Machine learning (ML) methods are designed to exploit large data sets in order to reduce complexity and find new features in data. The current most frequently used ML algorithms in HEP are Boosted Decision Trees (BDTs) and Neural Networks (NN). Machine learning in particle physics is evolving fast and ML algorithms are already state of the art in many areas of particle physics and will likely be called on to take on a greater role in solving upcoming data analysis and event reconstruction challenges [377].

#### 1. PDF determination as a ML problem

Among various applications, ML techniques have contributed to a better understanding of the proton structure. PDFs are typically determined by means of a supervised regression model which compares a wide set of experimental data with theoretical predictions computed with a specific PDF parametrization. A truthful determination of PDFs and of their uncertainties are crucial when producing theoretical predictions for precision studies in high energy physics.

From a methodological point of view, the choice of a regression model and its uncertainty treatment is a crucial decision, which will impact the quality of PDFs and of theoretical predictions. The determination of PDFs is a problem very well suited for ML techniques: the functional form is not known and there is a well agreed upon a figure of merit to be minimized during the optimization procedure (the  $\chi^2$ ).

The NNPDF collaboration pioneered the usage of Neural Networks as universal approximators for a model independent determination of the structure function  $F_2$  [378, 379] and later for full-fledged PDF determinations [10, 380]. The NNPDF sets are based in multi-layer feed-forward architectures which are often also known as multi-layered perceptrons. In a perceptron each layer is feed information from the previous one in a sequential manner. Each parton in the proton is then parametrized as:

$$f_i(x, Q_0) = (1 - x)^{\beta_i} x^{\alpha_i} \text{NN}_i(x), \quad (8)$$

where the Neural Network plays the same role of the parametrical functional form used in other PDF determination methods. The index  $i$  represents the parton while  $\text{NN}_i$  refers to the  $i$ -th output of the Neural Network (note however a single-network for all partons can also be utilized [10]).

The fit is performed at a fixed scale,  $Q_0 \simeq 1.6$ . The pre-factor  $(1 - x)^{\beta} x^{\alpha}$ , while not strictly necessary [381], speeds up the convergence of the network in the extrapolation regions by providing a sensible prior. The optimization of the



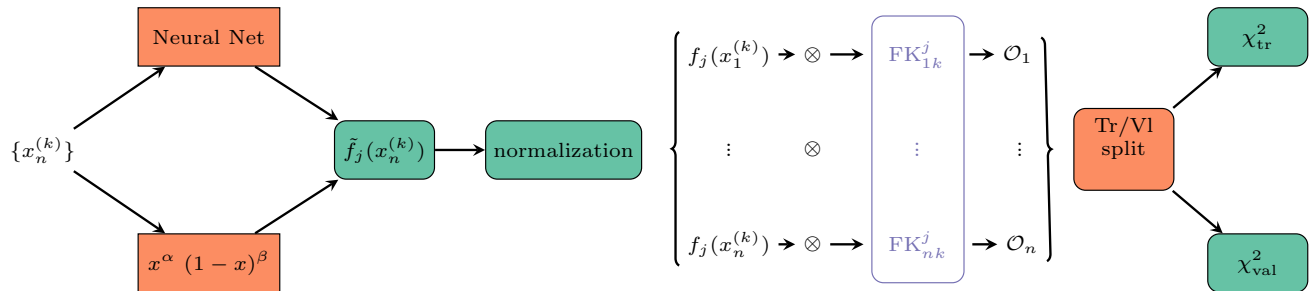


FIG. 15. Diagrammatic representation of the calculation of the  $\chi^2$  in the NNPDF fitting framework as a function of the values of  $\{x_n^{(k)}\}$  for the different data sets. Each block indicates an independent component.

parameters of the network has been performed using Genetic Algorithms until NNPDF3.1 [108]. Recently, Gradient Descent based algorithms have also been implemented within NNPDF framework [189, 382] and the NNPDF4.0 analysis [10] is based on them.

The fitting procedure described above requires a number of methodological choices: the exact architecture for the NN, the optimization algorithm -and its associated parameters if any-, stopping strategies, etc. Set of parameters defining the procedure are collectively known as hyperparameters. In previous versions of NNPDF these were determined by trial and error, most recently [368, 382] an automatic and systematic hyperparameter scan procedure has been introduced as a fundamental step of the NNPDF methodology. The faithfulness of the results (of the central value and corresponding uncertainties) are tested by statistical closure tests [31, 108]. These techniques, developed in the context of the NNPDF collaboration are not limited to partonic PDFs but are also used for the determination of parton distribution functions for nuclei collision [383] or the determination of fragmentation functions [384].

## 2. Simultaneous fits of physics parameters and PDFs

Despite the broad consensus on the need for precision, this need is often reduced to performing better measurements and improving the accuracy of theoretical predictions. However, it is equally important to have a robust framework that is able to globally interpret the LHC data, in particular to spot any subtle deviations from the SM predictions that might arise. While huge progress has been made in determining key ingredients of theoretical predictions from the data, such as the PDFs,  $\alpha_s(M_z)$ , EW parameters and the coefficients of a suitable parametrization of the effects of heavy new states via the addition of higher dimensional operators to the SM Lagrangian, such as the SMEFT, it is not yet evident how to combine all these partial fits into a global interpretation of the LHC data.

A very important step in this direction was done in [156], in which a full simultaneous fit of PDFs and EW parameters was done including full NNLO-QCD and full NLO-EW corrections. Future prospects for such simultaneous fits of PDF and EW precision parameters, using inclusive DIS data, were discussed in [157] – looking at projected LHeC data and [161] – looking at future electron-hadron Circular Collider (FCC-eh).

Also very recently, in [54] a new methodology, dubbed **SIMU**net, is presented which allows for a simultaneous determination of the PDFs alongside any physical parameter that enters theoretical predictions, whether a precision SM parameter, or the Wilson coefficients of some EFT expansion. The methodology is based on an extension of the **n3fit** methodology described in the previous section and the NNPDF4.0 neural network architecture, which treats both the PDFs and the parameters fitted alongside PDFs on a completely equal footing. The NNPDF fitting framework (see Section IX B 2) is extended to incorporate an extra layer of trainable edges to simultaneously determine the PDFs alongside an arbitrary number of such parameters. The capabilities of the new methodology are illustrated by simultaneously fitting PDFs with a subset of SMEFT Wilson coefficients and showing how the methodology extends naturally to larger subsets of parameters. For example, one could employ the methodology above to yield improved determinations of precision parameters along with the PDFs such as the strong coupling constant  $\alpha_s(M_z)$ , the EW parameters of the SM, the heavy quark masses, or a larger number of Wilson coefficients in the SMEFT or any other EFT expansion.

Concerning the simultaneous fit of PDFs and new physics parameters, in [55], a joint global fit including both PDFs and a single Wilson coefficient parametrizing a vector-current type lepton-quark contact interaction in the SMEFT is also successfully performed by means of a fast scan in the Wilson coefficient space and a study on how PDFs are

modified by the presence of a non-zero Wilson coefficients is performed. Also the ZEUS and CMS collaborations have performed similar studies in the context of DIS data and jets data [57, 58].

Moving towards simultaneous fits of PDFs and other parameters of the theory is certainly a very interesting direction that will receive a big deal of attention in the near future.

### 3. Other PDF-related ML applications

Beyond the determination of the PDFs themselves, ML techniques have also been proposed to power up PDF studies. In Ref. [385] genetic algorithms are utilized to compress a set of Monte Carlo replicas with the minimal loss of statistical information and in Ref. [386] this approach is further expanded with the usage of Generative Adversarial Model to enhance a PDF set, reducing finite-size artifacts that could be introduced by the compression procedure. In Ref. [387] PDFs are approximated using Quantum-ML algorithms in the context of the NNPDF fitting methodology. Another interesting application of ML to the study of PDFs was recently put forward [55]. Modern PDF analyses requires calculations of the log-likelihood functions from thousands of experimental data points, and scans of multidimensional parameter space with tens of degrees of freedom. To overcome the use of the Hessian method to approximate profile of the log-likelihood functions in neighborhood of the best-fit, and to estimate the PDF uncertainties, the authors put forward NNs and ML techniques to model profile of the log-likelihood functions or cross sections for such a multi-dimensional parameter space. The methodology is applied to the CT18 global analysis and to the study impact of the NOMAD dimuon data on constraining the strange content of the proton.

## D. Delivery of PDFs; PDF ensemble correlations in critical applications

*Leading authors: R. D. Ball, A. M. Cooper-Sarkar, T. Cridge, B. Malaescu, P. Nadolsky  
Contributions from S. Amoroso, A. Apyan, D. Froidevaux, S. Glazov, S.-O. Moch, R. Thorne*

An important component of the PDF methodology is the delivery of the PDFs to the users in the form that allows easy yet accurate estimation of a wide range of QCD cross sections and their PDF uncertainties. For this purpose, the PDFs are commonly distributed either as bundles of the central PDF parametrizations and error sets constructed as Hessian eigenvector sets [388] or as a Monte Carlo (stochastic) ensemble of replicas [389]. All error PDFs propagate the uncertainty from the fitted experimental data. Some PDF ensembles, e.g. [7, 11, 154, 370], include other sources of uncertainties implicitly or explicitly, such as those originating from the choice of input data, the methodology and parametrization, the choice of input theory and the related theoretical uncertainties. The error PDFs provide *approximations* to the full probability distributions explored in PDF fits. In practice, they reproduce the expectation values and key correlations in the fitted probabilities, while neglecting subleading features to some degree. There is a trade-off between the faithfulness of the reproduction of the full probability and the number of PDF error sets needed for this purpose. Several available methods can be further developed to compress [385, 390–392] or diagonalize [393, 394] the error PDFs to retain the relevant information with fewer PDF members/replicas.

As the field advances toward high precision in the LHC Run-3 and at the HL-LHC, more detailed models of PDFs may become necessary in experimental measurements. In particular, in the most precise cases it is observed that measurements performed with different PDF sets can differ by more than the expected PDF uncertainties, without an agreed-upon means to evaluate the degree of compatibility among the different results. A more rigorous and conservative quantitative approach would increase the PDF uncertainty of the measurement (e.g., following Ref. [395]) in the presence of statistically significant tension between results obtained using different PDF sets. A related question arises about the role of correlations among the PDF ensembles via the fluctuations in their shared fitted data sets.

In an effort taking place within the LHC Standard Model ElectroWeak and the PDF4LHC working groups, correlations between PDF sets obtained by different groups are being evaluated for the first time [396–398]. The study aims to gain precise knowledge of the degree of correlation between different PDF determinations, the essential missing ingredient to evaluate the degree of compatibility between different PDF sets and thereby derive realistic estimates of the overall PDF uncertainties for existing precision SM measurements. The correlations between different PDF determinations are studied by the means of fits to coherently generated pseudo-experiments, first by fitting a reduced ensemble of data sets, and then in a full-scale exercise [396] and accounting for fluctuating statistical and systematic correlated experimental uncertainties. The fluctuated data for each generated pseudoexperiment are shared by PDF fits of all participating groups, and hence the correlations between the fits via their shared data can be studied. A feasibility study of this kind has been already performed [399] by sharing the fluctuated pseudodata among PDF fits of different perturbative orders. Eventually, a follow-up study may be performed to understand the (de)correlations induced by the use of different parameterizations and fit methodologies, possibly performing comparisons using common

theoretical predictions and uncertainties [396–398].

As a related study to these proposals, a determination of the correlations between different PDF sets, using replica ensembles fitted to a common set of data replicas, has now been carried out [400]. It was found that even when fitted to identical data sets, using common theoretical predictions and parameter settings, different PDF sets are still only partially correlated, since the functional uncertainties arising from different methodologies (in this case, NNPDF3.1 and NNPDF4.0) are still treated as uncorrelated. This suggests it may be challenging to make use of the data or theory correlations to reduce uncertainties when combining different PDF sets, or making predictions obtained using different PDF sets, since the methodological correlations are unknown. Exploring this in more details is an interesting focal area for future studies.

## VI. PDFS AND THE STRONG COUPLING FROM LATTICE QCD

*Leading authors: H.-W. Lin, J. H. Weber*

### A. Strong Coupling Calculations

Traditionally or phenomenologically,  $\alpha_s(M_Z)$  is obtained by comparing experimental data involving a hard scale  $\nu_h$  to a function  $O(\alpha_s(n\nu_h), n)$  calculated in truncated perturbative QCD (pQCD); see Refs. [115, 401] for recent reviews. Nonperturbative lattice gauge theory (LGT) calculations – anchored to low-energy QCD by tuning the bare quark masses – provide numerical results for a wide range of quantities that may be utilized in place of experimental data. The basis of LGT is a regularization of the path integral on a discretized Euclidean space-time lattice with lattice spacing  $a$ , implicitly defined as a function of the bare gauge coupling  $g_0$ , that permits stochastic evaluation via MCMC simulations. Any LGT predictions are dimensionless ratios, e.g. of dimensionful quantities in units of the spacing  $a$ .  $a$  needs to be fixed in a somewhat arbitrary *scale setting* procedure that dictates the minimal errors for all dimensionful quantities. LGT is systematic, systematically improvable, and permits removing the regulator (continuum limit  $g_0 \rightarrow 0$ ). Moreover, just like pQCD, LGT is not restricted to the physical world; unphysical quantities are fair game, too. Hence, high precision LGT calculations of various  $O(\nu_h)$  now play a major role in determining  $\alpha_s(M_Z)$ . Functions  $O(\alpha_s(n\nu_h), n)$  entail unknown truncation errors (TE), which usually dominate the error budget at scales  $\nu_h \ll M_Z$ . As TE can only be estimated, e.g., by varying the scale (the number  $n \sim 1$ ), one should use as many quantities with unrelated TE as possible.  $\alpha_s(n\nu_h)$  is finally connected to  $\alpha_s(M_Z)$  by perturbative running and decoupling.

LGT calculations suffer from a window problem. The hierarchy  $\Lambda_{\text{QCD}} \ll \nu_h \ll 1/a$  is mandated when comparing to  $O(\alpha_s(n\nu_h), n)$ ; otherwise (former) there may be substantial nonperturbative effects, and the truncation introduces large uncertainties, or (latter) the hard scale is poorly resolved on the lattice, thus making continuum extrapolation challenging. On top of this, lattice simulations should maintain the hierarchy  $1/L \ll m_\pi \ll \Lambda_{\text{QCD}}$ , too, for a reliable connection to low-energy QCD. Unsatisfying realizations of the latter hierarchy are usually subleading in the error budgets. Topological freezing (incorrect sampling of the QCD vacuum’s topological sectors), which occurs on fine lattices, seems to have no significant impact on short distance quantities [402] used in determining  $\alpha_s(M_Z)$ . For  $< 5\%$  precision of the QCD Lambda parameter as required for reaching 1% accuracy of  $\alpha_s(M_Z)$  aimed at in the next decade, electroweak or isospin breaking effects can still be safely neglected.

There is a rich trove of literature on lattice determinations of the strong coupling constant; for modern reviews see Refs. [401, 403, 404]. There have been substantial albeit sometimes controversial efforts to formulate standardized quality criteria for lattice determinations of hadronic quantities or the strong coupling constant. The Flavor Lattice Averaging Group (FLAG) report [405–408] is the most impactful; FLAG reports its most recent global lattice average

$$\alpha_s(M_Z) = 0.1184(8) \quad (\text{FLAG global average}) [408]. \quad (9)$$

There is broad consensus in the LGT community that such quality criteria – if applicable – should be applied to phenomenological determinations, too. In the following, we summarize the most important conceptually different lattice methods. Similar to different classes of phenomenological determinations, these methods are thought to have unrelated TE; spread between or within these methods is usually rather narrow, and the error in Eq. (9) is taken to be the smallest among those of the individual methods instead of the much smaller naive error of a weighted average.

The *step scaling* method [409–412] allows to calculate  $\alpha_s(\nu_h = 1/L)$  at large energy scales while avoiding the window problem through a finite volume ( $V = L^3$ ) approach, and relies on the *Schrödinger functional* scheme [413–416]. Relevant pQCD expressions are known at N<sup>2</sup>LO resp.  $\mathcal{O}(\alpha_s^3(\nu_h))$  [417]. The most recent result [418] is widely regarded as the most reliable one for obtaining  $\alpha_s(M_Z)$ , and dominates the FLAG average.

Another lattice method utilizes short distance observables  $O(\nu_h)$  such as *small Wilson loops* [419–422]. The key difference to other lattice methods is that comparison to pQCD is performed at a finite lattice spacing, which is inversely proportional to the relevant hard scale  $\nu_h = d_O/a$ , with a coefficient  $d_O \sim \pi$  depending on the observable. Relevant pQCD expressions are known at N<sup>2</sup>LO resp.  $\mathcal{O}(\alpha_s^3(\nu_h))$  [419], which is the limiting factor for higher precision.

A third lattice method utilizes the *QCD static energy*, which can be studied via Wilson loops, or gauge-fixed Wilson or Polyakov line correlators, from which the static energy or the (singlet) free energy [423] are obtained. As reaching the perturbative regime at  $r \lesssim 0.15$  fm [424] presently requires fine lattices and distances  $r \lesssim 5a$  affected by non-smooth lattice artifacts, the continuum limit is still under investigation, see e.g. [404, 408]. No scheme change is required to obtain  $\alpha_s(\nu_h = 1/r)$ , and continuum pQCD expressions are known at N<sup>3</sup>LL or up to  $\mathcal{O}(\alpha_s^{4+n}(\nu_h) \ln^n(\alpha(\nu_h)))$ ,  $0 \leq n$  [425–435]. Using the singlet free energy at very short distances the error budget is dominated by statistics [436]. Otherwise, details of the resummation of ultra-soft logs  $\alpha_s^{4+n}(\nu_h) \ln^n(\alpha(\nu_h))$  [431, 437, 438] or scale variation generate the lion’s share of the error budget [436, 439–441].

A fourth lattice method (similar to quarkonium sum rules [442, 443]), uses *heavy-quark two-point correlators*. The valence heavy-quark mass serving as the hard scale  $\nu_h = xm_h$  can be varied across charm- and bottom-quark regions. The moments  $G_n$  are finite for  $n \geq 4$ , and known up to N<sup>3</sup>LO resp.  $\mathcal{O}(\alpha_s^3(\nu_h))$  for  $N_f$  massless and one massive flavor [444–446]. The large bare quark mass  $am_{h0}$  necessitates improved quark actions – usually HISQ [422, 447–451], or domain-wall fermions [452]. Reduced moments, e.g.  $R_4 = G_4^{\text{QCD}}/G_4^0$ , cancel the tree-level contribution and associated lattice artifacts [447]. The continuum limit turned out challenging, in particular for  $R_4$  at  $m_h \gtrsim 2m_c$ . Results at  $m_h = m_c$  are consistent [422, 447, 450–452] (up to known deficiencies), while reliable results up to  $m_h \leq 4m_c$  are recent [451]. The composition of the error budget for  $\alpha_s(M_Z)$  varies with  $m_h$ : at  $m_h \geq 2m_c$  statistical errors dominate and non-perturbative contributions are irrelevant, while at  $m_h < 2m_c$  truncation errors dominate. Omitting the scale variation led to underestimated error budgets in Refs. [422, 447–449], i.e. earlier than Refs. [443, 450, 451].

A fifth, somewhat similar method uses *light-quark two-point correlators* or the hadronic vacuum polarization. It can be computed via OPE in the isospin limit for Euclidean momenta  $Q^2 = -q^2 > 0$ , where the hard scale is  $\nu_h = \sqrt{Q^2}$ . This OPE is in terms of the Adler function, whose leading coefficient is known at N<sup>4</sup>LO resp.  $\mathcal{O}(\alpha_s^4(\nu_h))$  [453–456]; further terms are due to nonperturbative contributions. As the window problem is severe, these calculations are very challenging both in momentum space [457–459] or in position space [460].

The sixth widely used method – pioneered in Refs. [461, 462] – uses *QCD vertex functions* in a fixed gauge. Requiring nonperturbative dressing factors of only ghost and gluon two-point functions, the ghost-gluon vertex in Landau gauge is particularly simple [463]. The respective hard scale is  $\nu_h = \sqrt{q^2}$ , with  $q_\mu$  being the four-momentum of one ghost and the gluon. Nonperturbative contributions to the renormalized coupling  $\alpha_T(\nu_h)$  in an intermediate MOM Taylor scheme known at N<sup>4</sup>LO resp.  $\mathcal{O}(\alpha_s^4(\nu_h))$  [464] are suppressed for large  $\nu_h$ , while fundamental  $n$ -point functions themselves and the conversion to  $\overline{MS}$  are only known at N<sup>3</sup>LO resp.  $\mathcal{O}(\alpha_s^3(\nu_h))$  [465]. Neither calculations with twisted-mass Wilson fermions [464, 466–468] nor with domain-wall fermions [469] pass FLAG quality criteria.

A novel approach is the *decoupling method* [401, 470], in which  $N_f$  massive quark flavors with a large, common mass  $M$  serving as the hard scale  $\nu_h = M$ , are simultaneously decoupled to connect the running coupling to  $N_f = 0$ ; the decoupling relation is known at N<sup>4</sup>LO resp.  $\mathcal{O}(\alpha_s^4(\nu_h))$ . The matching between theories with different  $N_f$  is performed at the in principle arbitrarily high scale  $M$ , such that truncation errors are expected to be small, and volume effects are practically irrelevant.

A final, not yet widely used method relies on the *eigenvalues of the Dirac operator* that are known at N<sup>3</sup>LO resp.  $\mathcal{O}(\alpha_s^3(\nu_h))$  [471, 472]. While being a bit high, its large reported errors [473] overlap with the FLAG average.

## B. Lattice calculations of PDFs

There has been rapid progress calculating the Bjorken- $x$  dependence of PDFs on the lattice since the first proposal of Large-Momentum Effective Theory (LaMET, also called the “quasi-PDF” method) [474, 475]. LaMET relates equal-time spatial correlators, whose Fourier transforms, the quasi-PDFs become the lightcone PDFs in the limit of infinite hadron momentum. For large but finite momenta accessible on a realistic lattice, LaMET relates quasi-PDFs to physical ones through a factorization theorem, the proof of which was developed in Refs. [476–478]. Since the first lattice  $x$ -dependent PDF calculation [479], much progress has been made and many calculations done. Alternative approaches to lightcone PDFs in lattice QCD are “operator product expansion (OPE) without OPE” [480–487], “auxiliary heavy/light quark” [488–491], “hadronic tensor” [492–497], “good lattice cross sections” [476, 498–501] and the pseudo-PDF approach [502]. For recent reviews on these topics, we refer readers to Refs. [18, 19, 503–506] for more details.

The most studied  $x$ -dependent structure is the nucleon unpolarized isovector parton distribution function (PDF)  $u(x) - d(x)$ . Multiple collaborations have reported either direct lattice calculations at physical pion mass or extrapolations to physical pion mass using quasi-PDF [507–509] and pseudo-PDF methods [510, 511]. Ref. [512] presents



the first lattice-QCD calculation of the nucleon isovector unpolarized PDFs in the physical-continuum limit, using ensembles with multiple sea pion masses with the lightest around 135 MeV, three lattice spacings  $a \in [0.06, 0.12]$  fm, and multiple volumes with  $M_\pi L$  ranging 3.3 to 5.5. A simultaneous chiral-continuum extrapolation was performed to obtain RI/MOM-scheme renormalized nucleon matrix elements with various Wilson-link displacements and four physical-continuum matrix elements. Figure 16 shows the results of the lattice calculations using at least one near-physical pion mass. There are different systematics, some taken into account, some not. Overall there is a reasonable agreement after scaling up the systematics. However, the sea-flavor asymmetry in unpolarized PDFs, like the small- $x$  region, suffers large systematics and can only be removed when using a large value of  $P_z$  [509, 513], as predicted previously [507]. Increasing the boost momentum of the lattice calculations will be critical to extending the impact of future lattice PDF calculations at both large and small  $x$ .

The first lattice-QCD calculations of the strange and charm parton distributions using LaMET approach were reported in Ref. [514]. The calculation of light ( $M_\pi \approx 310$  MeV) and strange nucleon ( $M_\pi \approx 690$  MeV) two-point correlators includes 344,064 (57,344) measurements in total, allowing extrapolation to physical pion mass. They found that the renormalized real matrix elements are zero within the statistical errors for both strange and charm, supporting the strange-antistrange and charm-anticharm symmetry assumptions commonly adopted by most global PDF analyses. The imaginary matrix elements are proportional to the sum of the quark and antiquark distribution, and the strange contribution is about a factor of 5 or larger than the charm ones. They are consistently smaller than those from CT18 and NNPDF3.1, possibly due to the missing contributions from the mixing with gluon matrix elements in the renormalization. The later work by ETMC [515], which calculated both light and strange lattice matrix elements at 260-MeV pion mass, extracted individual quark-flavor PDFs with the mixing in the quark and gluon sectors neglected. Future calculations, to include the gluon mixing, will be crucial to the lattice lattice-flavor dependent PDFs.

There have been attempts in lattice QCD to constrain the notorious gluon PDFs, which usually require orders of magnitude higher statistics than their quark counterparts to get a nonzero result within the statistical uncertainty. The first exploratory study applying LQCD to gluon PDFs was done by MSULat group [516], using  $M_\pi^{\text{sea}} = 330$  MeV. The coordinate-space gluon quasi-PDF matrix-element ratios are compared to the corresponding gluon-PDF ones based on two global fits at NLO: the PDF4LHC15 combination [16] and the CT14 [2]. Up to perturbative matching and power corrections at  $O(1/P_z^2)$ , the lattice results are compatible with global fits within the statistical uncertainty at large  $z$ . Since then, there have been improvements to the operators for the gluon-PDF lattice calculations [517–519], which will allow us to take the continuum limit for the gluon PDFs in future lattice calculations. The followup work using “pseudo-PDF” method by MSULat group [520] attempts an extrapolation to physical pion mass of the gluon PDFs at a single lattice spacing. HadStruc collaboration [521] used a different numerical technique in extracting the gluon PDF at 360-MeV pion. Both results have limited  $x$  range access; the small- and large- $x$  regions can be improved further in the future calculations.

Early exploratory works have shown great promise in obtaining quantitative results for the helicity and transversity quark and antiquark distributions [522]. There have been two attempts to improve the helicity PDFs by removing the heavy pion-mass systematic by ETMC [508] and LP<sup>3</sup> [513]. Transversity has been dominated by quasi-PDF method carried out by LP<sup>3</sup> [523] and ETMC [524], whose transversity results at physical pion mass were reported in 2018. Recently, HadStruc Collaboration reported results from pseudo-PDF approach with a lattice spacing  $a = 0.094$  fm and 358-MeV pion mass. Excited-state systematics in above works are carefully studied using multiple source-sink separations; those including up to two excited-state-related matrix elements will result in larger statistical errors that embedded the systematics errors. Some results include the errors coming from varying the renormalization scale, the choice of  $zP_z$  in the Fourier transform, approximations made in the matching formula, lattice-spacing and finite-volume effects but not all. Using large boost momenta in the nucleon does make the statistical errors of the PDFs larger, even using high-statistics measurements.

LQCD calculations of the valence-quark dependence of PDFs for the pseudo-scalar mesons have also been done [500, 501, 525–531]. Reference [530] was the first study of lattice pion and kaon PDFs to take the continuum-physical limit of the matrix elements with a sufficient number of lattice spacings and light pion masses. There has also been a first next-to-next-to-leading order (N<sup>2</sup>LO) matching [532, 533] lattice calculation of the pion valence quark PDF [531] with 300-MeV pion mass. Both works take important steps toward precision PDFs from lattice QCD. Gluon PDFs of mesons suffer from significant signal-to-noise issues, so it is harder to get precise signals than for their valence-quark counterparts. Such calculations often require orders of magnitude increases of the computational resources to get nonzero results. Recently, there have been first attempts to study the gluon PDFs of the pion [520] and kaon with lightest pion masses of 220 and 310-MeV, respectively. However, the mixing with the quark PDFs was not directly calculated in the lattice calculations; rather, they use global fits to estimate their contribution. These studies show mild dependence of the pion gluon parton distribution on lattice spacing and pion mass. In the pion case, the results are comparable with the determination by JAM and xFitter global fits.

New experiments and facilities provide an opportunity to explore the transverse structure of hadrons, which is de-



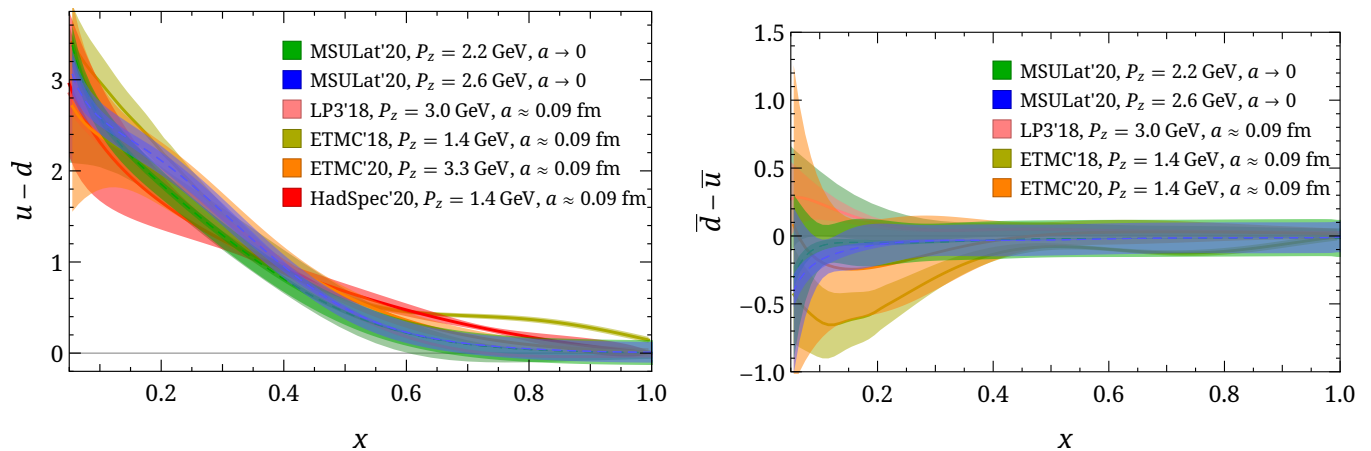


FIG. 16. The nucleon isovector unpolarized PDFs from a lattice calculation in the physical-continuum limit, “MSULat’20” [512], a single lattice spacing calculation at (or extrapolated to) physical pion mass using LaMET methods, “LP3’18” [509] and “ETMC’18” [508], and pseudo-PDF method, “ETMC’20” [510] and “HadSpec’20” [511], compared with past lattice quasi-PDF results from LP<sup>3</sup> and ETMC at one ensemble at physical pion mass.

scribed by the generalized parton distributions (GPDs) and transverse-momentum-dependent distributions (TMDs), providing three-dimensional structure of the hadron. GPDs provide hybrid momentum and coordinate-space distributions of partons and bridge the standard nucleon structure observables: form factors and collinear PDFs. Only recently have there been a few lattice calculations made for the pion GPDs at 310-MeV pion [534], and nucleon GPDs at 260-MeV [535] and 139-MeV pion [536, 537]. The calculations require an increase of at least an order of magnitude in computational expense relative to PDF calculations due to the additional boost momenta required for each transfer momentum studied. TMDs measure the parton transverse momentum  $k_T$  with longitudinal momentum fraction  $x$  and are nonperturbative inputs for processes that follow TMD factorization, such as Drell-Yan and semi-inclusive DIS (SIDIS). Early lattice studies attempted to compute the quasi-TMDs at heavy quark masses [538–542], and recent efforts in extract the Collins-Soper kernel, soft function and wavefunctions for TMDs [543–547]. A number of calculations pursue the latter, because the calculations can be done at much heavier pion mass, since these features are insensitive to pion mass. Like the PDF calculations, lattice precision calculations will require large hadron momentum to suppress the power corrections at  $\mathcal{O}(1/(P_z b_T)^2)$ . There have been rapid developments in lattice QCD toward the determination of parton physics. New theoretical advancements are opening doors to many previously unavailable quantities, from the Bjorken- $x$  dependence of collinear nucleon PDFs to higher-twist observables, and the three-dimensional structure of GPD/TMD, which no one envisioned as possible during the previous Snowmass process. There are remain challenges to be overcome: fighting the noise-to-signal for gluons, increasing the boost momenta used in the calculations to reduce the systematics, etc. Many of these are detailed in a separate Snowmass white paper focusing on lattice PDF calculations [548]. Computational resources are also significant limiting factor on what precision a lattice parton-physics calculation can achieve. A diverse workforce can bring in new ideas to solve some of the challenges that currently limit progress without an increase in computational time. With sufficient support, lattice QCD can fill in data gaps where the experiments are difficult (or not yet available), improve the precision of the global fits, and provide better Standard-Model inputs to aid new-physics searches in many HEP frontiers.

## VII. NUCLEAR AND MESON PDFS

*Leading authors: T. J. Hobbs, E. R. Nocera, F. I. Olness*

The QCD theory of the strong interactions is among the most complex and enigmatic, displaying both confinement of the quarks and gluons (at large distance scales) and asymptotic freedom (at short distance scales). These characteristics make QCD extremely challenging to formulate perturbatively, leading to intriguing non-linear collective effects. Even the structural details of stable nucleons and nuclei are not fully derivable from *ab initio* QCD theory. The extension of the quark-parton framework to include nuclear PDFs provides a pathway to describe some of these nuclear phenomena within a perturbative framework.

Progress in nuclear PDF (nPDF) [30, 116, 136, 138, 549] analyses was made rapidly in recent years due, in part, to new measurements from both fixed-target and collider (RHIC, LHC) experiments, see *e.g.* [137, 383, 550, 551] and references therein. As compared to the proton PDFs, the nuclear PDFs have an extra dimension to explore; the

nuclear  $A$  dimension represents both a challenge and an opportunity. It is a challenge because the size of the typical nuclear PDF data set is most/typically comparable to the proton data set, but it has the nuclear  $A$  as an extra degree of freedom. It is an opportunity because the freedom of the nuclear  $A$  dimension allows us to compare a variety of different nuclei as we look for patterns that may provide clues to a deeper understanding of QCD phenomenology.

Studies of nPDFs reciprocally leverage techniques from proton PDF analysis, and *vice versa*. As we explore different  $A$  values, we can move from the well-known limit of the proton ( $A=1$ ) up to the very heavy gold and lead nuclei. As such, free-nucleon global analyses generally serve at least as a baseline for attempts to extract the  $A$  dependence in nPDF fits, but can also be exploited as a constraining boundary of the  $A$  dependence at  $A=1$ . As discussed further below, nPDFs are also informative with respect to proton PDF determinations in the sense that they may illuminate nuclear effects in high-statistics or large volume data sets used in proton fits. This is an area where new approaches from machine learning, artificial intelligence, and lattice-QCD may be proven fruitful [19].

In addition to studying the PDFs of baryons, we can also explore meson PDFs such as the pion and kaon [101, 552–555]. These meson PDFs may exhibit a distinct structure and might thus provide additional clues to the character of the QCD theory. This research avenue is being pursued both with the QCD parton model and also QCD lattice calculations.

### A. Impact on Proton PDFs

Nuclear PDFs play an essential role in determining the proton PDF by helping to distinguish individual parton flavors. For example, a key data set is neutrino–nucleon ( $\nu A$ ) DIS structure functions  $\{F_{2,3}^\nu, F_{2,3}^{\bar{\nu}}\}$ ; these four independent structure functions provide unique information to help extract the individual parton PDFs. The heavy nuclear targets are required due to the small neutrino cross section; therefore, the nuclear correction ratios (e.g.,  $F_2^{\text{Fe}}/F_2^p$ ) are indispensable for translating the nuclear results to the proton PDFs. In heavy ion collisions, a superposition of these hot and cold nuclear matter effects is expected, and a quantitative evaluation of the latter is an important prerequisite for a detailed understanding of the former. As the precision of proton PDFs has steadily improved, it becomes critical to reduce the comparatively large uncertainties of the nuclear correction ratios. New data combined with improved theoretical analyses allow us to separately determine the nuclear correction factors and the flavor decomposition of the nuclei with high precision.

### B. Exploring Nuclear $A$ Dependence

As mentioned above, the nuclear  $A$  dimension represents the opportunity to explore a data set that is comparable and more diverse than the measurements limited to only the proton. These nuclear PDF fits do typically use a smooth parameterization in the nuclear  $A$  value, and hence make the implicit assumption that the nPDFs vary smoothly in this dimension. While this may be a reasonable assumption for the heavier nuclei, this can be problematic for light nuclei such as deuterium where few-body bound state effects may be more challenging to incorporate into a smooth parametrization; for instance, corrections associated with the deuteron can be sizable and significant [4, 103]. This is a topic that needs further investigation.

There are complementary approaches to the study of light nuclei using lattice-QCD calculations; improved PDF moment calculations and the quasi-PDF methods have proven beneficial for the proton analysis, and preliminary analysis of light nuclei are in progress [18, 19, 548]. For example, the NPLQCD collaboration computed the first moment of the unpolarized gluon distribution for the deuteron and  $^3\text{He}$  using a higher-than-physical quark mass; these investigations can serve as a starting point for future developments [556]. Although the lattice studies are limited to very light nuclei, within the nCTEQ parameterization it has been observed that some of the  $A$ -dependent parameters evolve quickly at low  $A$  values [137]. In particular, deuteron corrections have been studied extensively and these have been determined to be important in fitting nuclear data [4], especially since much of the nuclear structure function data is expressed as ratios of the form  $F_2^A/F_2^D$ . Thus, even additional insights on the first few nuclei may help us improve our description of the nPDFs in the low  $A$  region, and this provides a boundary condition that we can use as we extrapolate to larger  $A$  values.

### C. Collective Properties of QCD

Another important aspect of nuclear studies is the observation of collective effects. This can include jet quenching in nuclei–nuclei collisions [557, 558], long-range correlations (the ridge effect) in both proton–proton [559] and proton–lead [560] collisions, quark gluon plasma (GQP) [561], color glass condensate (CGC) [562], nuclear saturation [563], cold

nuclear matter effects [564], as well as others. These various phenomena may either induce characteristic signatures in the nPDFs themselves, or be phenomenologically limited by the present knowledge of the nPDFs.

#### D. Extreme Kinematics

Nuclei offer the opportunity to explore not only the large- $x$  region ( $x \lesssim 1$ ), but can also extend to  $x$  values beyond unity ( $x \gtrsim 1$ ). This is a region where target mass corrections are expected to be important. In the small- $x$  region we can explore the resummation of  $\ln(1/x)$  contributions in the BFKL framework. Additionally, parton saturation and recombination is expected to grow in this region with a  $A^{1/3}$  enhancement for heavy nuclei. Of course if gluon saturation regime is reached as expected at some value of  $x$ , then collinear factorization must break down and the concept of a PDF is not useful. While structure functions are physical observables, the concept of PDFs relies on collinear factorization. This becomes an even bigger issue for heavy nuclei where these small  $x$ -enhanced "highers twist" effects are further enhanced by  $A^{1/3}$ . In this kinematic regime all higher-point functions are of the same order as the two-point functions, i.e. the PDFs. Finally, we can extend analyses into the low- $Q^2$  region where the increase of  $\alpha_S(Q)$  pushes us into a nonperturbative regime. Preliminary investigations have examined the effects of relaxing the typical  $Q^2$  and  $W^2$  cuts [565, 566]. This work suggests that the characteristic  $x$  dependence of nuclear structure-function ratios persist into the resonance region at low- $W$  values and could be a manifestation of the quark-hadron duality phenomenon. If correct, this may permit a description of nuclear structure functions in terms of partonic degrees-of-freedom, even in kinematic regions where resonance excitation is the dominant effect.

#### E. Outlook

The QCD theory of strong interactions remains as one of the critical components of the Standard Model to be adequately understood. The large value of the  $\alpha_S$  coupling constant at soft momenta renders traditional small-coupling perturbation theory unavailable in the infrared, such that collective phenomena in nuclei are nonperturbative and therefore not amenable to such a computational framework. A concerted application of the QCD parton model for protons, nuclei and mesons, together with advances from Lattice QCD calculations, could yield a fundamentally improved understanding of the characteristics of nuclei and their interactions. Utilizing new precision measurements from JLab, RHIC, the LHC/HL-LHC and the future EIC and neutrino measurements, this combination of experimental and theoretical efforts could be the keystone to a deeper understanding of underlying nuclear dynamics, and represent a substantive step forward in an the ultimate understanding of the QCD theory based on first principles.

### VIII. TRANSVERSE-MOMENTUM DEPENDENT DISTRIBUTIONS

*Leading authors: V. Bertone, C. Bissolotti, F. G. Celiberto, G. Schnell, and G. Vita*

Collinear factorization and the ensuing collinear parton distribution functions (PDFs) have proven to be powerful tools for the study of high-energy collisions involving hadrons in the initial and/or final state. Nonetheless, the use of collinear factorization is limited to observables characterized by a single hard scale and cannot be applied to observables in which there are two or more widely separate hard scales. An example of observable that breaks collinear factorization is the transverse-momentum distribution  $\vec{q}_T$  of the lepton pair in Drell–Yan (DY) production at small values of  $q_T$ . In this regime, the presence of large logarithms of  $q_T$  in the perturbative calculation of the partonic cross sections spoils the perturbative convergence effectively invalidating collinear factorization. An appropriate description of DY at low  $q_T$  is instead achieved through *transverse-momentum-dependent* (TMD) factorization [567] that has the ability to resum the large logarithms of  $q_T$  to all orders in perturbation theory thus producing sensible results at low values of  $q_T$ .

A “byproduct” of TMD factorization is the introduction a TMD distributions (TMDs). TMDs can be regarded as a generalization of collinear distributions in that they provide information on the transverse momentum distribution of partons within hadrons. As a consequence, TMDs encode much more information on the structure of hadrons than PDFs. Therefore, their knowledge has the potential to shed light on the origin of basic properties of hadrons, such as their spin and mass decompositions. Precise studies of TMD fragmentation-functions (FF), counterparts of TMD distributions, doable at the Electron-Ion Collider (EIC) [27, 99] as well as at new-generation lepton-lepton machines [568] will certainly extend our knowledge of TMDs.

## A. Quark TMDs

When accounting for partonic transverse momentum, the interplay between hadron and parton polarizations gives rise to a much richer partonic structure of hadrons. It turns out that for a spin-1/2 hadron there exist eight independent leading-twist TMDs [569]. A second important feature of TMDs is that, as opposed to collinear distributions, they break naive universality; in other words they may depend on the process under consideration. This breaking of universality can be traced back to the presence of the Wilson line in the operator definition of TMDs necessary to guarantee gauge invariance. Specifically, partonic transverse-momentum effects introduce a dependence of the Wilson line on the integration path that in turn is determined by the process in which the TMDs are participating. In the case of quark TMDs, the possible Wilson-line configurations relevant to phenomenological applications are only two often referred to as future-pointing  $[+]$  and the past-pointing  $[-]$  staple links. For example the  $[-]$  configuration is to be used in DY while the  $[+]$  configuration enters in semi-inclusive deep-inelastic-scattering (SIDIS) production. The net effect of moving from one Wilson-line configuration to the other is a sign change for time-reversal-odd (T-odd) TMDs while time-reversal-even (T-even) TMDs remain unaffected. This is the origin of the now well-known Sivers effect [570] that predicts that the Sivers TMDs have opposite sign depending on whether they are being used to compute DY or SIDIS production.

The larger number of distributions associated to a lesser experimental accessibility as compared to collinear PDFs, makes the phenomenological exploration of TMD much more laborious. The consequence of this is that our current quantitative knowledge of TMDs is generally far less accurate than that of PDFs. In fact, many of the eight leading-twist quark TMDs are largely unknown with most of the effort being put into the study of the unpolarized quark TMDs,  $f_1^q$ , and the Sivers quark TMDs  $f_{1T}^{\perp q}$ .

In the past few years, the relevance of the unpolarized TMDs,  $f_1^q$ , has been fully recognized for questions related to the precision determination of the electroweak parameters of the Standard Model (SM) aimed at the search for physics beyond the SM. This triggered intense activities around these distributions that has led to fairly accurate determinations. In spite of pioneering determinations of  $f_1^q$  such as that of Ref. [571], the recent years have seen a steady acceleration mostly triggered by the data delivered by the LHC experiments and by important theoretical advances. See Sects. VIII C–VIII E for a more detailed discussion on the most recent progress concerning the determination of  $f_1^q$ , the relevant theoretical advances, and an overview of the experimental information.

A second category of TMDs that has seen raising interest in the past few years is the Sivers distribution,  $f_{1T}^{\perp q}$ . The recent determinations of Refs. [572–574] have brought the accuracy of these distributions to an unprecedented level. A point worth mentioning is that the collinear dynamics of the quark Sivers distributions is driven by the twist-3 quark Qiu–Sterman (QS) distributions [575]. At present, our knowledge of the QS distributions is very limited which left the floor open to different approaches to the determination of the quark Sivers TMDs that is a currently matter of debate.

## B. Gluon TMDs

The complete list of leading-twist gluon TMDs for a spin-1/2 target was first given in Ref. [576] (see also Refs. [577, 578]), where the polarization states of both of the parent nucleon and the struck gluon were accounted for. Gluon TMDs for a spin-1 target were listed Ref. [579], and this led to the emergence of 11 new distributions on top of the ones arising from the spin-1/2 case. The two gluon TMD functions that survive after the integration over transverse momentum are the distribution of unpolarized gluons inside an unpolarized nucleon,  $f_1^g$ , and of circularly-polarized gluons inside a longitudinally-polarized nucleon,  $g_1^g$ . They represent the TMD counterparts of the unpolarized and helicity gluon PDFs, in the collinear regime.

As it happens in the quark case, gluon TMDs are sensitive to the resummation of transverse-momentum logarithms which appear to all orders of the perturbative series. They constitute the perturbative contribution of the TMDs (see, *e.g.*, Refs. [580–582] and references therein). While our knowledge about the transverse-momentum resummation is quite solid, the nonperturbative component of gluon TMDs, relevant to understand the dynamics of intrinsic motion of partons inside nucleons, is poorly known.

Similarly to quark TMDs, different classes of reactions probe distinct gauge-link structures for gluon TMDs, each of them being given in term of a combination of two or more staple links. This leads to a more diversified kind of *modified universality*. Two major gluon structures exist: the  $f$ -type and the  $d$ -type one. They are also known in the context of small- $x$  studies as Weiszäcker–Williams and dipole links, respectively [583, 584]. The antisymmetric  $f_{abc}$  QCD color structure enters the analytic expression of the  $f$ -type T-odd gluon-TMD correlator, while the symmetric  $d_{abc}$  structure is part of the  $d$ -type T-odd one. The  $f$ -type gluon TMDs depend on  $[\pm, \pm]$  gauge-link combinations. The  $[+, +]$  structure is probed in reactions where the gluon interacts with a color-singlet initial particle (*e.g.*, a photon in a DIS process) and two colored objects (*e.g.*, two jets) are emitted in the final state. The  $[-, -]$  structure emerges in

processes where a gluon interacts with another gluon (color-octet state) and a color-singlet state (*e.g.*, a Higgs boson) is tagged in the final state. TMD factorization holds for all these reactions, and the following modified-universality relations for  $f$ -type distributions come from time-reversal invariance (T-symmetry)

$$\begin{aligned} f_1^g[+,+] &= f_1^g[-,-] && \text{(T-even),} \\ f_{1T}^g\perp[+,+] &= -f_{1T}^g\perp[-,-] && \text{(T-odd).} \end{aligned} \quad (10)$$

Here the unpolarized gluon TMD,  $f_1^g$ , is a representative of all the T-even distributions, while the Siver gluon TMD,  $f_{1T}^g$ , stands for all the T-odd functions. The  $d$ -type gluon TMDs depend on  $[\pm, \mp]$  gauge-link combinations and appear in reactions when a gluon interacts with a colored initial particle and a colored final-state system is produced (*e.g.*, when a photon is emitted together with a jet in proton-proton collisions). The  $d$ -type modified-universality relations are analogous to the  $f$ -type ones, given in Eq. (10). In this case, TMD factorization has not been proven and might be affected by issues connected with color entanglements [585]. More intricate gauge-link structures are involved in processes where multiple color states are present in both the initial and final state [586]. Here TMD factorization runs into even deeper issues.

Experimental information on gluon TMDs is very limited. First attempts at phenomenological analyses of the unpolarized gluon TMD have been presented in Refs. [587–591]. Experimental and phenomenological studies of the intrinsic motion of gluons in transversely-polarized protons via the Sivers function can be found in Refs. [592–595]. Thanks to its connection with the QCD Odderon, the gluon Sivers TMDs can also be studied in unpolarized electron-proton collisions [596].

In the *high-energy factorization* regime, where gluons are extracted from nucleons with a small longitudinal fraction  $x$  and a large transverse momentum, a relation can be established [584] between the unpolarized and linearly-polarized gluon distributions,  $f_1^g$  and  $h_1^{\perp g}$ , and the *unintegrated gluon distribution* (UGD), whose evolution is controlled by the Balitsky–Fadin–Kuraev–Lipatov (BFKL) equation [310–312, 597] (see Refs. [202, 598–606] for recent applications). A connection between the high-energy and the TMD factorization was recently highlighted in Refs. [607, 608]. The impact of embodying gluon-TMD inputs within high-energy factorization was recently assessed for vector-meson leptonproduction processes at the EIC [609–611].

The distribution of linearly polarized gluons in an unpolarized nucleon,  $h_1^{\perp g}$ , plays a crucial role on the dynamics underlying spin effects in collisions of unpolarized hadrons [612–617]. They are collectively known as the Boer–Mulders effect. Part of it is generated at large transverse momenta within perturbative QCD via the transverse-momentum resummation and it represents the perturbative part of  $h_1^{\perp g}$ . A genuine perturbative-QCD treatment would miss, however, the polarization effect generated by the *intrinsic* motion of gluons, which has a nonperturbative nature and cannot be caught by the resummation, but needs to be quantified via fits on global data that will be collected at new-generation colliders [27, 198, 204, 618–620].

With the aim of bridging the gap between theory and experiment, phenomenology-suited models are needed to perform exploratory studies of gluon TMDs. A recent calculation of all the unpolarized and polarized T-even gluon-TMD densities at twist-2 was done via an enhanced spectator-model approach [621] (see also Refs. [622–624], and Refs. [625, 626] for similar results in the quark case), where proton remnants after gluon emission are treated as a single on-shell effective fermion. Preliminary calculations of leading-twist T-odd functions were presented in Refs. [627–629].

Taking advantage of the link between TMD and collinear factorization, a consistent procedure was set up in Ref. [621] to simultaneously fit the unpolarized and helicity gluon TMD densities to the corresponding collinear PDFs obtained from NNPDF [15, 125] at the initial scale  $Q_0 = 1.64$  GeV. Predictions for the unpolarized and the linearly-polarized gluon TMD are presented in Fig. 17 as functions of the transverse momentum squared,  $\mathbf{p}_T^2$ , for  $x = 10^{-3}$  and at the initial scale  $Q_0$ , namely without switching TMD evolution on. Thus, initial-scale results precisely refer to the nonperturbative part of our TMD densities. Predictions are given as a set of 100 replicas, which are statistically equivalent and reproduce well the unpolarized and helicity collinear PDFs. Each red line in plots represents a single replica, with the black line corresponding to the most representative one (n. 11).

We note that each TMD exhibits a peculiar trend both in  $x$  and  $\mathbf{p}_T^2$ . The unpolarized TMD clearly shows a non-Gaussian pattern in  $\mathbf{p}_T^2$ , and goes to a small but non-vanishing value for  $\mathbf{p}_T^2 \rightarrow 0$ . The linearly-polarized gluon TMD is large at small  $\mathbf{p}_T^2$  and decreases very fast. Both of them are increasingly large at small  $x$ , and their ratio is constant in the asymptotic limit  $x \rightarrow 0$ . This is in line with the BFKL behavior of the small- $x$  UGD, which predicts an “equal number” of unpolarized and the linearly-polarized gluons up to higher-twist effects. This is a touch point between the TMD and the BFKL approach that could be explored via studies on processes featuring a *natural stability* of the high-energy resummation [203, 630–637]. Furthermore, even if all replicas reproduce similar collinear PDFs, they predict very different results for the TMDs in Fig. 17. Forthcoming data on gluon TMDs are expected to exclude many replicas and constrain parameters not yet so well constrained by collinear PDFs.



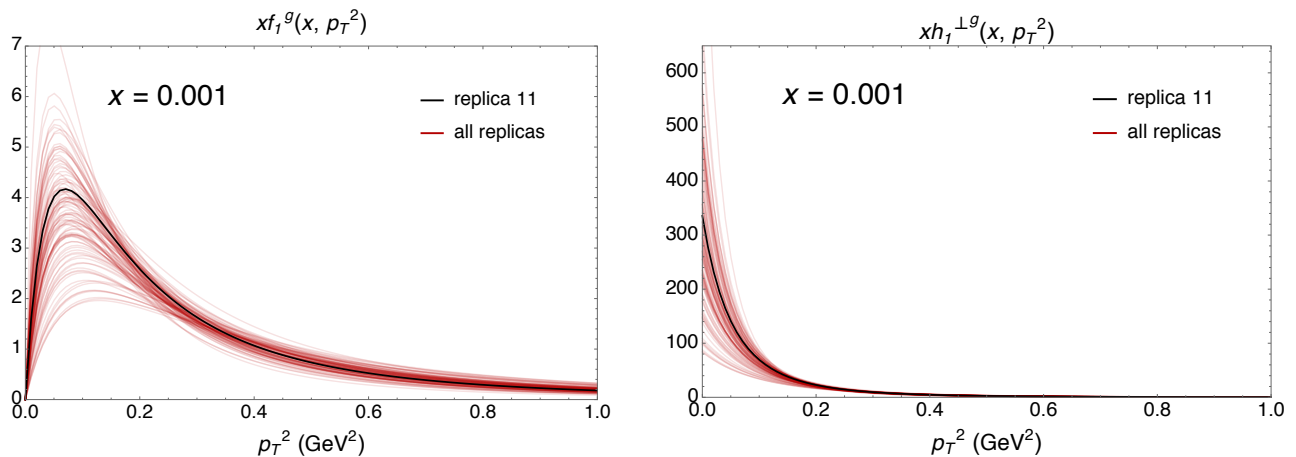


FIG. 17. Unpolarized (left) and linearly-polarized (right) gluon TMDs as functions of  $p_T^2$ , for  $x = 10^{-3}$  and at the initial scale  $Q_0 = 1.64$  GeV. Figures adapted from Ref. [621]

### C. TMD evolution and matching to collinear PDFs

Transverse momentum dependent PDFs allow for a 3D description of the internal dynamics of the proton. However, we also know that for high energy scattering the nonperturbative information regarding such dynamics can be typically described in terms of simply collinear PDFs. It is therefore interesting to understand how this transition happens and how we can improve theoretical predictions for TMD sensitive observables across the large spectrum of transverse momenta probed by current and future experiments. We start by noticing that for TMD observables at colliders there are typically 3 scales that characterize the physics at play: the scale of the hard scattering  $Q$ , the scale of transverse momenta measured for the observable of interest  $q_T$ , and  $\Lambda_{\text{QCD}}$ . For values of the transverse momentum that are perturbative, i.e. for  $q_T \gg \Lambda_{\text{QCD}}$ , it is possible to define an Operator Product Expansion (OPE) which matches TMDs onto collinear PDFs up to corrections of  $\mathcal{O}(\Lambda_{\text{QCD}}/q_T)$ . Schematically the OPE takes the form [638, 639]

$$f_i^{\text{TMD}}(z, \vec{q}_T, \mu, \nu) = \sum_j \int_z^1 \frac{dz'}{z'} \mathcal{I}_{ij}(z', \vec{q}_T, \mu, \nu) f_j\left(\frac{z}{z'}, \mu\right) \times [1 + \mathcal{O}(q_T/\Lambda_{\text{QCD}})], \quad (11)$$

where  $\mathcal{I}_{ij}(z, \vec{q}_T, \mu, \nu)$  is a perturbative matching kernel,  $f_j(x, \mu)$  is the standard collinear PDF for the flavor  $j$ , and  $\nu$  is a rapidity scale related to the presence of rapidity divergences in the renormalization of TMDs which can take a variety of different forms and notations depending on the renormalization procedure and scheme employed [640–650]. For an overview of different schemes for TMD definitions and rapidity regularization see for example App.B of ref. [651].

Throughout the years, significant progress has been made in the calculation of the matching kernels up to N2LO both for the quark [652–656] and for the gluon case [654, 655, 657, 658]. Recently, their calculation has been pushed to N3LO [659, 660]. It is important to note that, given the complexity of these analytic calculations, achieving such level of accuracy from the perturbative side required significant innovation in the way such objects are calculated. New methods for performing multiloop computation in the context of effective field theory have been developed, such as generalized integration by parts identities for the treatment of rapidity regulators [649, 659] and a framework for the collinear expansion of analytic cross sections [661]. These new tools, originally developed for the calculation of the TMDPDFs matching kernels, have been applied to obtain results at N3LO for a variety of different observables such as  $N$ -jettiness beam functions, time-like splitting functions, transverse momentum dependent fragmentation functions, and energy-energy correlators [662–666].

The evolution of the TMDPDFs is dictated by a coupled system of differential equations [640, 646, 649, 667] which becomes multiplicative in impact parameter space. Using the  $b_T$  as the conjugate variable of  $q_T$ , the RGEs take the form

$$\begin{aligned} \mu \frac{d}{d\mu} \tilde{f}_i(x, b_T, \mu, \nu/\omega) &= \tilde{\gamma}_\mu^i(\mu, \nu/\omega) \tilde{f}_i(x, b_T, \mu, \nu/\omega), \\ \nu \frac{d}{d\nu} \tilde{f}_i(x, b_T, \mu, \nu/\omega) &= -\frac{1}{2} \tilde{\gamma}_\nu^i(b_T, \mu) \tilde{f}_i(x, b_T, \mu, \nu/\omega), \end{aligned} \quad (12)$$

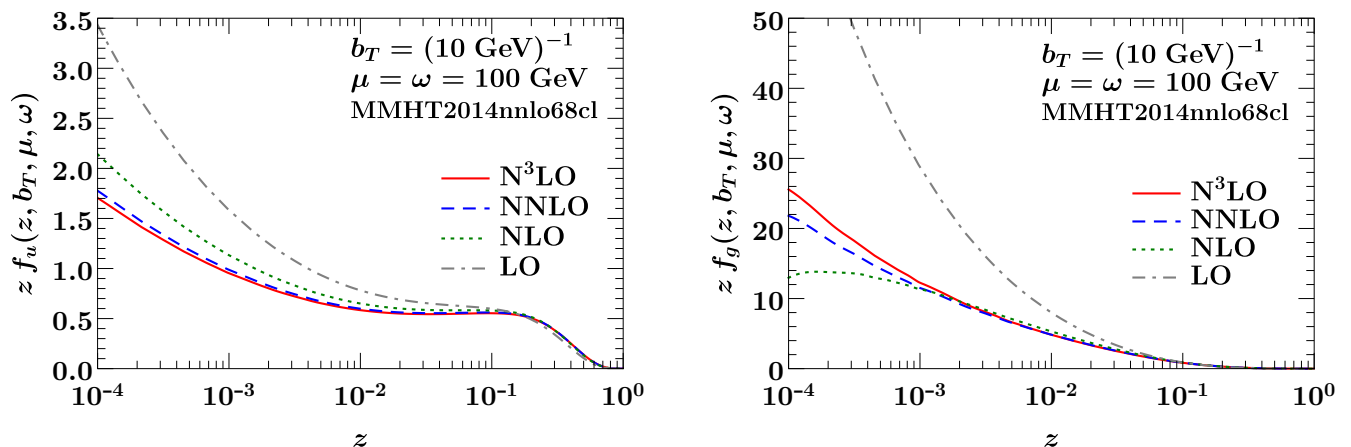


FIG. 18. The  $u$ -quark TMDPDF (left) and the gluon TMDPDF (right) as a function of  $z$  for fixed  $b_T = (10 \text{ GeV})^{-1}$  and  $\mu = \omega = 100 \text{ GeV}$ . We show the result at LO (which is equivalent to the PDF since the matching kernel is trivial at this order), NLO, N2LO, and N3LO. Plots taken from [660].

where  $\tilde{\gamma}_\mu^i(\mu, \nu/\omega)$  is related to the collinear and threshold anomalous dimensions and  $\tilde{\gamma}_\nu^i(b_T, \mu)$  is the so called rapidity anomalous dimension [646] which is closely related to the Collins-Soper kernel [640, 667] and has been obtained at N3LO in [668].

Progress in resummation accuracy is crucial for precise phenomenology. The determination of the complete singular analytic structure for various TMD observables at N3LO, obtained via the calculation of the N3LO TMD beam [659, 660], and fragmentation functions [664, 665], enabled the push of the TMD resummation accuracy to N3LL'. This was first applied for the description of the energy-energy correlator in the back-to-back limit [666], an event shape in electron-positron colliders, and then for transverse momentum distributions and fiducial cross sections at the LHC [254, 669]. In both cases, the reduction of the perturbative uncertainties is very significant as the resummation accuracy increases to N3LL'. A precise control of perturbative uncertainties on resummed cross sections thanks to calculation of anomalous dimensions and boundary functions to 3 loop and beyond will be even more relevant at future colliders, as the center of mass energies allowed for the hard scattering will be higher than it is at present colliders, therefore extending the region where resummation effects will be important.

#### D. Status of unpolarized TMD extractions

Information on the functional form of TMDs can be obtained from Drell–Yan (DY) and Semi-Inclusive Deep Inelastic Scattering (SIDIS). In fact, for these processes factorization theorems allow us to write the cross section in term of convolutions of TMDs. In particular, in the so-called *TMD factorization region*, where  $q_T \ll Q$ , the DY cross section is proportional to a convolution of two TMD PDFs and the SIDIS cross section can be expressed in terms of a convolution of one TMD PDF and one TMD FF. TMDs are partially computable by means of well-established perturbative methods that take into account soft and collinear radiation to all orders. However, calculations based on perturbative QCD become unreliable for values of transverse momentum close to the Landau pole ( $\Lambda_{\text{QCD}}$ ). In this regime, nonperturbative components have to be included and have to be determined through fits to experimental data.

Recent works directly performed extractions of TMDs from Drell–Yan data [670–672], Semi-Inclusive DIS data [673–675] or both [676–679]. At the present time, the best known quark TMD is the unpolarized TMD PDF  $f_1(x, k_\perp)$ , whose latest extractions reach the state-of-the-art perturbative accuracy of N3LL [680]. In Fig. 19 we show the results of the  $f_1(x, k_\perp)$  extraction performed in Ref. [680] and we compare the unpolarized TMD PDFs for the up and down quark at  $Q = 2 \text{ [GeV]}$  for different values of  $x$ , from  $x = 10^{-3}$  up to  $x = 0.3$ .

An accurate knowledge of the TMDs is useful not only to investigate the structure of the nucleon in greater detail, but also to improve the reliability of predictions involving TMDs. At high energies, the perturbative part of TMDs may be dominant, but when extreme precision is required, also the nonperturbative components become relevant (see, *e.g.*, Ref. [681]).

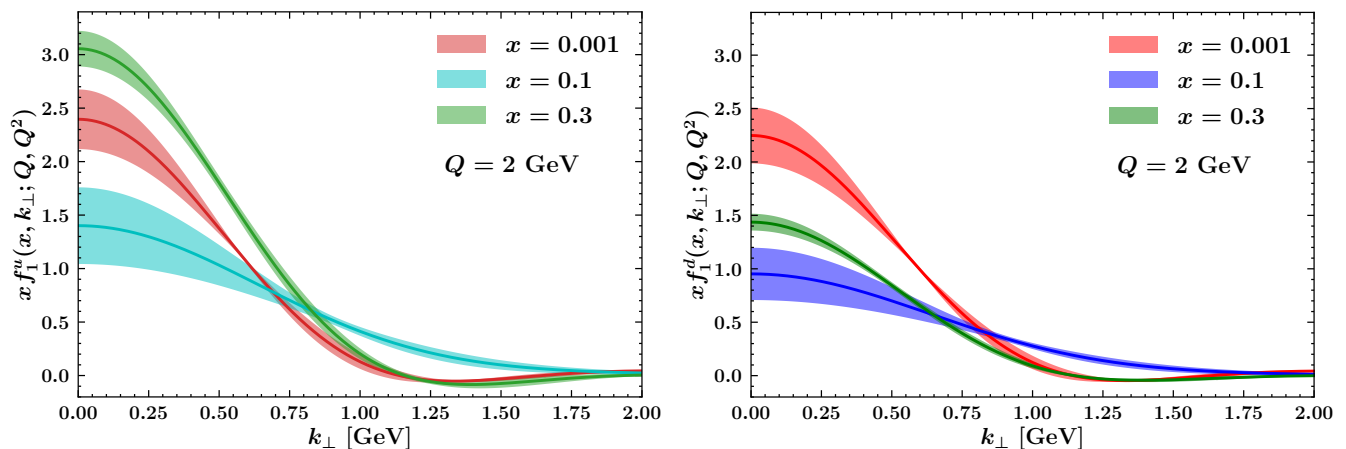


FIG. 19. The unpolarized TMD PDFs of the up and down quarks at  $Q = 2$  [GeV] as a function of the partonic transverse momentum  $k_{\perp}$  for different values of  $x$ . The bands give the  $1\text{-}\sigma$  uncertainty.

### E. Experimental prospects

While integrated PDFs have been extensively studied, especially in *inclusive* reactions, the knowledge of TMDs is still very limited as hardly accessible in such reactions. The preferred processes for TMDs are semi-inclusive DIS and Drell–Yan (or more general vector-boson production in hadron collisions) for TMD PDFs, and in addition  $e^+e^-$  annihilation into hadrons for TMD fragmentation functions. Lately, also hadron+jet production has been put forward as complementary observable, in particular for the study of such TMD fragmentation functions. Current data for all those processes cover a much smaller kinematic space as, e.g., for integrated PDFs (cf. fig. 5 of Ref. [679] vs. fig. 2.1 of Ref. [10]) or fragmentation functions. In particular, the region in  $x$  below 0.02 and below the scale of  $Z$  production, which in case of integrated PDFs is predominantly covered by collider DIS data, currently lacks suitable data for TMD studies.

Looking at the future of TMD physics, the EIC will be of paramount importance. At the EIC, the main access to TMDs will be through semi-inclusive DIS. What will make the EIC a unique accelerator is the fact that it will be colliding polarized electrons (and potentially positrons) with polarized protons and light nuclei at various energies and with unprecedented luminosity for a lepton-hadron collider. Together this will yield the precision and the spin degrees of freedom necessary to pursue an ambitious physics program driven by the exploration of the spin structure and the acquisition of new information on multi-dimensional tomographic images of protons and nuclei. [See the **”Hadronic Tomography at the EIC and the Energy Frontier” Snowmass whitepaper.**] In the future EIC era, the possibility to produce precise theoretical predictions is going to be crucial in the study of future experimental data, especially considering the foreseen luminosity of the new accelerator, coupled with its energy variability and reach.

The EIC will certainly contribute to better determine TMDs, as it will provide very precise measurements in kinematic regions where at the present moment no data are available: for example, EIC data will be able to cover the before-mentioned large gap between the low-energy fixed-target experiments and the high-energy experiments at the LHC.

The huge impact that the EIC will have on TMD extractions can be seen for instance by looking at the analyses reported in Sec. 7.2 of the EIC Yellow Report [27], where impact studies based on pseudo-data coming from PYTHIA simulations [682] have been performed. Impact studies carried out by fitting simultaneously both existing data and pseudo-data have shown a significant reduction (up to a factor of  $\sim 4$  in the kinematic regions not covered by present data) of the uncertainty bands for the unpolarized quark TMDs. Moreover, a reduction of the uncertainties of a factor of  $\sim 10$  is foreseen in the determination of the nonperturbative part of the evolution for the unpolarized TMDs  $f_1(x, k_{\perp})$  and  $D_1(z, P_{\perp})$ .

On the low-energy side, the Jefferson Lab 12 GeV program will continue to contribute on the precision frontier to the TMD mapping with orders of magnitude higher luminosity and a wide range of polarization and target configurations (see, e.g., Ref. [683]). CLAS12 [684], the SBS, and the future SoLID [685] experiments will explore the valence region with an unprecedented precision. These data will also provide crucial input in evolution studies of TMDs. Currently, ideas are put forward to expand the kinematic reach by increasing the beam energy to 24 GeV (cf. Appendix C of Ref. [683]).

The LHC will continue to provide crucial data especially on the high-energy end. Its importance lies also in the different processes used to study TMDs, allowing for tackling questions of factorization and universality. So far, the LHC is perceived as a machine for only unpolarized TMD studies. Installing polarized targets at the LHC would permit to also enter the domain of polarized TMDs, most notably the Sivers function that is expected to change sign when probed in Drell–Yan vs. semi-inclusive DIS. Indeed, such ideas have been put forward [686] and are extensively pursued within the [Physics Beyond Collider Study Group](#) at CERN [687]. Injecting polarized nucleons into a storage cell internal to the LHC ring in front of the LHCb detector [688], similar to what was done for the HERMES experiment at HERA [689], is currently the most promising avenue. Even the use of unpolarized gases in such fixed-target setup, as already foreseen for the LHC run 3 [690], opens up unique opportunities of studying nucleon TMDs at large scales and very large  $x$  [686, 691]. Dedicated Drell–Yan studies of TMDs are also foreseen by [SpinQuest](#) at Fermilab or are part of the remaining program at RHIC [692]. All these activities, using lepton-hadron as well as hadron-hadron reactions, will help tremendously to fill the currently existing gap in phase space (see, e.g., fig. 12 of [691]) that seriously limits precision studies of TMD physics. Last but not least, extraction of TMD fragmentation functions will crucially profit from the advent of Belle II [693]. Already its predecessor Belle has provided several unique and complementary TMD measurements of hadron production in  $e^+e^-$  annihilation [694–698]. They will profit from the 50 times increase in luminosity, while the latter will at the same time offer new avenues.

## IX. COMPUTING NEEDS AND COMPUTING TOOLS

### A. The LHAPDF library and other user interfaces for PDFs

*Leading authors: A. Buckley*

LHAPDF is the community standard resource for access to parton density fits across collider experiments and phenomenology, and in its current incarnation (v6, since 2013) contains more than 1150 PDF sets encoded in a uniform data format and interpolated with standard algorithms. While these have generally met or exceeded required precision for MC calculations, the expense of PDF interpolation is a non-trivial aspect of NLO calculations, and for calculations at N3LO order the default local-bicubic interpolation in  $\log x$ – $\log Q^2$  space has been found insufficiently stable [699–701].

Work in 2020-21 succeeded in both reducing the CPU cost (intrinsic to LHAPDF, as well as via optimised generator PDF-call strategies) and developing smoother Lagrange-based interpolators for stability in high-precision calculations. The latter, as well as support for GPU workflows (cf. Python-oriented tools like `PdfFlow` [702]) and more general error-set combination rules, will shortly appear in upcoming LHAPDF releases.

Longer-term requirements on PDFs, from precision hadron-collider studies,  $e^+e^-$  collider prospects, and  $ep$  physics at EIC, will require extension of the current nucleon-specific LHAPDF machinery and interface to support also resolved-photon and transverse-momentum dependent (TMD) PDFs. These extensions, while motivated by distinct physics processes, share the common feature of requiring interpolation in more than two variables: as the standard 2-variable  $x$ – $Q^2$  interpolation is implemented as composition of 1D interpolator functions, the extensions will be implemented by recursive strategies for higher-order composition. This generalisation may also be a useful opportunity to agree community-standard interfaces for PDF querying, to allow better interoperation of LHAPDF6 with PDF-fitting toolkits such as `ApfelGrid` [233] and `xFitter` [703].

### B. Public PDF fitting codes

Most PDF sets are made publicly available via the LHAPDF interface, described in the previous section. However, until recently only the outcomes of the global PDF fits, namely the LHAPDF interpolation grid files, were released, while the PDF fitting codes themselves remained private. This implied that results were not reproducible by external parties. Another limitation of private PDF codes is that benchmarking studies, such as those described in this document become more convoluted due to the challenge in disentangling the various components that determine the final outcome.

The open source QCD fit framework `xFitter` [703] was the first to make the HERAPDF fitting code publicly available, along with several other features that are essential to perform global QCD analyses. The NNPDF code [368] was also recently made available and it offers complementary functionalities as compared to those in `xFitter`, for example by offering machine learning tools for the PDF parametrization and the automated determination of the best minimisation algorithm and neural network parametrization that the data suggest [382], along with robust methods

to estimate the robustness of PDF analysis via closure tests [31], and an extensive suite of statistical validation and plotting tools. In what follows we describe the two available public frameworks in more details.

### 1. *xFitter: an Open Source QCD Analysis Framework*

*Leading authors: F. Giuli, F. I. Olness*

**xFitter** [703] is an open-source software package that provides a framework for the determination of the proton and pion PDFs, as well as fragmentation functions and related subjects.<sup>4</sup> **xFitter** version 2.2.0 has recently been released, and offers an expanded set of tools and options. It incorporates experimental data from a wide range of experiments including fixed-target, Tevatron, HERA, and LHC data sets. **xFitter** can analyze these data using predictions up to N2LO in perturbation theory with a variety of theoretical calculations including numerous methodological options for carrying out PDF fits and plotting tools which help visualize the results. While primarily based on the collinear factorization foundation, **xFitter** also provides facilities for fits of dipole models and transverse-momentum dependent (TMD) distributions.

First and foremost, **xFitter** provides a flexible open-source framework for performing PDF fits to data. **xFitter** can also automatically generate comparison plots of data vs. theory. There are a variety of options for the definition of the  $\chi^2$  function and the treatment of experimental uncertainties. Examples are presented in Ref. [703].

**xFitter** is able to perform PDF profiling and reweighting studies. The reweighting method allows **xFitter** to update the probability distribution of a PDF uncertainty set to reflect the influence of new data. For the PDF profiling, **xFitter** compares data and MC predictions based on the  $\chi^2$ -minimization, and then constrains the individual PDF eigenvector sets taking into account the data uncertainties. For example, it has been found that the forward-backward asymmetry in neutral current Drell-Yan production provides powerful constraints the valence quark PDFs, and this in turn can impact both SM and BSM physics [704].

The package can also be used to study the impact of new precise measurements from hadron colliders, and also assess the impact of future colliders. A typical study might be to use pseudo-data from a proposed experiment (e.g. LHeC or EIC) to constrain the relative uncertainty on the underlying PDFs. For example, Ref. [705] used LHeC pseudo-data to constrain the strange PDF with charged-current DIS charm production data. Additionally, it has been shown that measurements of lepton angular distributions can be used to improve the accuracy of theoretical predictions for Higgs boson production cross sections at the LHC [706]. The high-statistics determinations of the longitudinally polarized angular coefficient at the LHC Run III and high-luminosity HL-LHC improve the PDF systematic uncertainties of the Higgs boson cross section predictions by 50% over a broad range of Higgs boson rapidities. Moreover, the complementarity of the lepton-charge and forward-backward asymmetries in DY processes has been studied and the impact in reducing PDF uncertainties in observables relevant to both SM and BSM physics has been assessed [707].

**xFitter** can also study the impact of the  $\ln(1/x)$ -resummation corrections to the DGLAP splitting functions using DIS coefficient functions from the public code HELL [330, 331]; these effects are illustrated in Ref. [708]. In a related study [709], a more flexible PDF parametrization is used with **xFitter** which provides a better description of the combined inclusive HERA I+II data, especially at low- $x$ .

Another feature of **xFitter** is the ability to handle both pole masses and  $\overline{MS}$  running masses. While the pole mass is more closely connected to what is measured in experiments, the  $\overline{MS}$  mass has advantages on the theoretical side of improved perturbative convergence. **xFitter** was used to perform a high precision determination of the  $\overline{MS}$  charm mass in this new framework [710].

Finally, as many PDF analyses are now extended out to N2LO, the NLO QED effects can also become important. For example, including QED processes in the parton evolution will break the isospin symmetry as the up and down quarks have different couplings to the photon. **xFitter** is able to include NLO QED effects, and this is illustrated in Ref. [711] which computes the photon PDF as determined using a N2LO QCD and NLO QED analysis.

### 2. *NNPDF: an open-source machine learning framework for global analyses of PDFs*

*Leading author: M. Ubiali*

---

<sup>4</sup> The **xFitter** webpage is located at: <https://www.xfitter.org/xFitter/>



Along with the recent release of the NNPDF4.0 PDF set [10], in a companion paper [368] the public release of the complete software framework underlying the NNPDF4.0 global determination was presented<sup>5</sup>, together with user-friendly examples and an extensive documentation<sup>6</sup>. In addition to the NNPDF fitting code itself, the public repository includes the original and filtered experimental data, the fast NLO interpolation grids relevant for the computation of hadronic observables, and whenever available the bin-by-bin N2LO QCD and NLO electroweak K-factors for all processes entering the fit. Furthermore, the code comes accompanied by a battery of plotting, statistical, and diagnosis tools providing the user with an extensive characterisation of the PDF fit output.

These statistical analysis and plotting tools are provided by the validphys toolkit, which is at the heart of the NNPDF code base, bridging together the other components and providing basic data structures, compatibility interfaces, I/O operations and algorithms. The validphys code is in turn built on top reportengine [712], an user-friendly data analysis framework which provides a declarative interface that allows the user specifying the required analysis by providing a minimal amount of information in the form of a run card, making the analysis reproducible given the run card.

The availability of the NNPDF open-source code enables users to perform new PDF analyses based on the NNPDF methodology and modifications thereof. Some examples of potential applications include assessing the impact of new measurements in the global fit; producing variants based on reduced data sets, carrying out PDF determinations with different theory settings, such as different values of  $\alpha_s$ , heavy quark masses, electroweak parameters; estimating the impact on the PDFs of theoretical constraints and calculations e.g. from nonperturbative QCD models [306] or lattice calculations [18, 713]; and quantifying the role of theoretical uncertainties from missing higher orders to nuclear effects. One could also deploy the NNPDF code as a toolbox to pin down the possible effects of beyond the Standard Model physics at the LHC, such as Effective Field Theory corrections in high-pT tails [52, 56] or modified DGLAP evolution from new BSM light degrees of freedom [714]. Furthermore, while the current version of the NNPDF code focuses on unpolarised parton distributions, its modular and flexible infrastructure makes it amenable to the determination of closely related nonperturbative collinear QCD quantities such as polarised PDFs, nuclear PDFs, fragmentation functions, or even the parton distributions of mesons like pions and kaons.

### C. Fast interfaces for pQCD computation

Modern calculations of higher-order corrections in perturbative QCD for predictions of cross sections from collider experiments are computationally very demanding, particularly at N2LO where typically order  $\mathcal{O}(10^5)$  CPU hours are required due to the complicated singularity structure of the real-emission amplitudes and the delicate numerical cancellations they entail. The data for such cross sections at the LHC are becoming increasingly precise, and so for QCD analyses involving comparison with the N2LO predictions these calculations must be repeated thousands of times using different values for the strong coupling  $\alpha_s(M_Z)$ , different parametrizations for the PDFs, or different choices for the factorisation or renormalization scales. It is therefore computationally prohibitive to run the full calculation at N2LO for each phase space point as required in such an analysis.

Storing the perturbative coefficients on a grid, before the convolution with the parton luminosity and the strong coupling constant  $\alpha_s$ , allows the convolution with arbitrary PDFs to be performed later, with essentially no additional computational cost. Variation of  $\alpha_s(M_Z)$ , and the renormalization and factorisation scales is also possible. The grid technique, used in Ref. [715], is implemented independently in the APPLgrid [716, 717] and fastNLO [718, 719] packages. The technique works by using interpolation functions to distribute each single weight from the integration over the momentum fraction  $x$ , and hard scale  $\mu^2$  in the convolution. For pp collisions, a third dimension must be added to account for the momentum fraction  $x_2$ . The APPLfast project [720] implements an interface of fastNLO and APPLgrid with the NNLOJET program [721]. These programs and their APPLfast interface are described in Section IX C 1. In Section IX C 2 we describe the PineAPPL interface, that allows the inclusion of NLO EW corrections.

#### 1. The APPLfast project

*Leading authors: D. Britzger, C. Gwenlan, A. Huss, J. Pires, K. Rabbertz, M. R. Sutton*

The grid technique works by accurately interpolating the full behaviour of any function  $f(x)$  from at discrete nodes in  $a \equiv x^{[0]} < x^{[1]} < \dots < x^{[N]} \equiv b$  that partition the interval  $[x_{\min}, x_{\max}]$ . Interpolating polynomials,  $E_i(x)$ , of degree

<sup>5</sup> The NNPDF code repository can be downloaded at <https://github.com/NNPDF/>.

<sup>6</sup> The NNPDF code documentation webpage is located at <https://docs.nnpdf.science/>.

$n$  are used for each node  $i$ , such that  $f(x)$  can be approximated by

$$f(x) \simeq \sum_{i=0}^N f^{[i]} E_i(x) \quad \text{with} \quad f^{[i]} \equiv f(x^{[i]}). \quad (13)$$

To increase the accuracy of the interpolation with equally spaced nodes, a variable transformation  $x \mapsto y(x)$  is used to increase the density of nodes in regions where  $f(x)$  varies more rapidly. The corresponding interpolation functions are denoted by  $E_i^y(x)$ . The integration can then be approximated by a sum over the nodes  $i$ ,

$$\int_a^b dx f(x) g(x) \simeq \sum_{i=0}^N f^{[i]} g_{[i]} \quad \text{where} \quad g_{[i]} \equiv \int_a^b dx E_i(x) g(x), \quad (14)$$

and the time-consuming computation of the Monte Carlo integral in Eq.(14) is performed once and for all to produce a grid  $g_{[i]}$  ( $i = 0, \dots, N$ ) for subsequent use. The integral in Eq. (14) can then be *a posteriori* approximated for different functions  $f(x)$  using only the summation over the  $N$  grid nodes.

For DIS processes the different parton densities  $f_a(x, \mu_F)$  can be included directly using the grid technique. In this case, a two-dimensional grid in the two independent variables  $x$  and  $\mu_F$  is constructed. As described in detail elsewhere [720], for any value of  $x$  and  $\mu$ , both the PDFs and the running of the strong coupling can then be represented by a sum over the interpolation nodes,  $i$  and  $j$ ,

$$\alpha_s(\mu) f_a(x, \mu) \simeq \sum_{i,j} \alpha_s^{[j]} f_a^{[i,j]} E_i^y(x) E_j^\tau(\mu), \quad (15)$$

with  $\mu_R = \mu_F \equiv \mu$  for simplicity. The index  $a$  represents the different partons in the cross section and the calculation includes an implicit sum over these partons. In practice, the parton summations often reduce to simple factors and sums over the up-type and down-type quarks, and the gluons. The computationally expensive convolution with the PDFs in Eq. (14), for each order in the calculation,  $\alpha_s^p$ , can thus be approximated by a summation,

$$\sigma = \sum_p \int dx \left( \frac{\alpha_s(\mu)}{2\pi} \right)^{k+p} f_a(x, \mu) \hat{\sigma}_a^{(p)}(x, \mu) \approx \sum_p \sum_{i,j} \left( \frac{\alpha_s^{[j]}}{2\pi} \right)^{k+p} f_a^{[i,j]} \hat{\sigma}_{a[i,j]}^{(p)}$$

and where  $\hat{\sigma}_{a[i,j]}^{(p)} = \sum_{m=1}^{M_p} E_i^y(x_m) E_j^\tau(\mu_m) w_{a;m}^{(p)} \hat{\sigma}_{a;m}^{(p)}$  (16)

where the sum over  $i$  and  $j$  runs over the grid nodes  $x_i$  and  $\mu_j$ . In the interpolation of the product, one interpolation variable is needed per independent variable, such that with one scale, and one momentum fraction, only two are needed, and a separate grid is required for each parton contribution. Including the scale variation in the renormalisation and factorisation scales, the summation over the grids  $\hat{\sigma}_{a;m}^{(p)}$  is modified by the inclusion of additional terms in the logarithms of both scales. The full expression can be seen in Eq. (16) from [720]. A comprehensive study of the NNLO predictions, as well as an application in PDF and  $\alpha_s$  determinations was presented further in Ref. [151]. Grids for inclusive jet and dijet production at HERA in NNLO are available at [722].

For cross section predictions for hadron-hadron collisions the convolution over the underlying partonic hard scattering includes a separate PDFs for each target hadron and so requires an additional interpolation for the momentum fraction from the second hadron,  $x_2$ , resulting in the overall interpolation function to distribute the weights from the calculation onto a grid;

$$\alpha_s(\mu) f_a(x_1, \mu) f_b(x_2, \mu) \simeq \sum_{i,j,k} \alpha_s^{[k]} f_a^{[i,k]} f_b^{[j,k]} E_i^y(x_1) E_j^y(x_2) E_k^\tau(\mu), \quad (17)$$

with  $\mu_R = \mu_F \equiv \mu$  again for simplicity, and the transformations  $x_i \rightarrow y_i$ ,  $\mu^2 \rightarrow \tau$  have been included. The summation over  $i$ ,  $j$ , and  $k$  represents the summation over the nodes for  $x_1$ ,  $x_2$ , and  $\mu$  respectively.

A separate grid is needed for each parton contribution, so for  $pp$  collisions, 121 grids (excluding top as a parton) would be needed for summations over partons  $a$  and  $b$ . This would make the grid extremely large and potentially prohibitive for any practical application where many grids need to be stored in memory. To reduce the number of contributions in the grids, symmetries within the hard subprocesses should be exploited to produce a smaller set of unique contributions. This allows the summation over the full set of parton flavour combinations,  $a$  and  $b$ , to be replaced by a single summation over a significantly smaller set of contributions,  $F_\lambda(x_1, x_2, \mu)$ , such that

$$\sum_\lambda F_\lambda(x_1, x_2, \mu) h_\lambda(x_1, x_2, \mu) \equiv \sum_{a,b} f_a(x_1, \mu) f_b(x_2, \mu) h_{ab}(x_1, x_2, \mu), \quad (18)$$

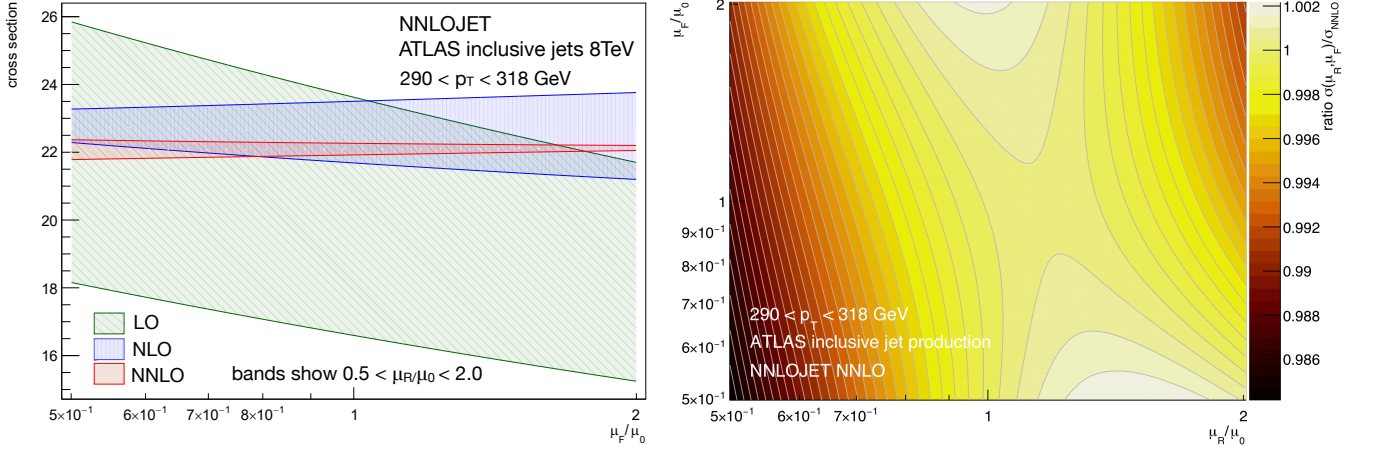


FIG. 20. The detailed scale variation for the ATLAS inclusive jet production cross section at 8 TeV in the range  $290 < p_T^{\text{jet}} < 318$  GeV. On the left with the factorisation scale factor on the  $x$ -axis and the bands showing the usual 0.5 to 2 scale variation for the renormalisation scale, and on the right, for the N2LO cross section as a function of the full two-dimensional variation in both the renormalisation and factorisation scales.

where the summations over the parton luminosities have been included explicitly on this occasion. For example, jet production in hadron–hadron collisions, can be represented using a smaller number of parton–parton luminosities – NNLOJET uses 13:

$$\begin{aligned}
 F_0(x_1, x_2, \mu) &\equiv f_g(x_1, \mu) f_g(x_2, \mu), \\
 F_1(x_1, x_2, \mu) &\equiv \sum_{i=1} f_g(x_1, \mu) f_{q_i}(x_2, \mu), \\
 &\dots \\
 F_{12}(x_1, x_2, \mu) &\equiv \sum_{i=1} f_{\bar{q}_i}(x_1, \mu) f_{\bar{q}_i}(x_2, \mu)
 \end{aligned} \tag{19}$$

reducing the number of separate contributions that must be stored in the grid from 121 down to 13. In this way, the interpolated cross section prediction can be written as

$$\sigma \simeq \sum_n \sum_{i,j,k} \left( \frac{\alpha_s^{[k]}}{2\pi} \right)^{p+n} F_\lambda^{[i,j,k]} \hat{\sigma}_{\lambda[i,j,k]}^{(n)}. \tag{20}$$

In the grid generation at N2LO, the process of reducing the number of separate parton luminosity contributions is performed automatically using the structure of the NNLOJET calculation, by mapping the separate parts of the calculation to the smaller number of unique parton–parton contributions. The final grid is then obtained by accumulating the weights according to

$$\hat{\sigma}_{\lambda[i,j,k]}^{(n)} = \xrightarrow{\text{MC}} \sum_{m=1}^{M_n} E_i^y(x_{1;m}) E_j^y(x_{2;m}) E_k^T(\mu_m) w_{\lambda;m}^{(n)} d\hat{\sigma}_{\lambda;m}^{(n)}, \tag{21}$$

where now the terms  $w_{\lambda;m}^{(n)}$  correspond to the weights  $w_{ab;m}^{(n)}$  associated with the individual terms for  $\lambda$ . As in the case of the DIS cross section, the full grid convolution including scale variations is significantly more complex and includes terms logarithmic in the scales.

In order to facilitate analyses up to N2LO in QCD with LHC data the authors are undertaking a campaign of high numerical precision grid production using NNLOJET for calculations of a number of jet production cross section data from the LHC. For such calculations, high statistics running is required to produce a stable cross section at higher orders, requiring typically hundreds of thousands of CPU hours. Typically grids are able to reproduce the reference cross section to within 0.1%. It is expected that a number of grids for LHC jet cross sections from both ATLAS and CMS will be made available to the wider community on the ploughshare web site [722] in the near future.

Figure 20 shows the variation of the LO, NLO and N2LO cross sections for ATLAS jet production [723] in the range  $290 < p_T^{\text{jet}} < 318$  GeV with variations of the renormalisation and factorisation scale factors. The renormalisation variation can in principle be determined *a posteriori* from the cross section alone, however, this is not the case for the factorisation scale variation, which requires the full calculation. The factorisation scale dependence is shown in the

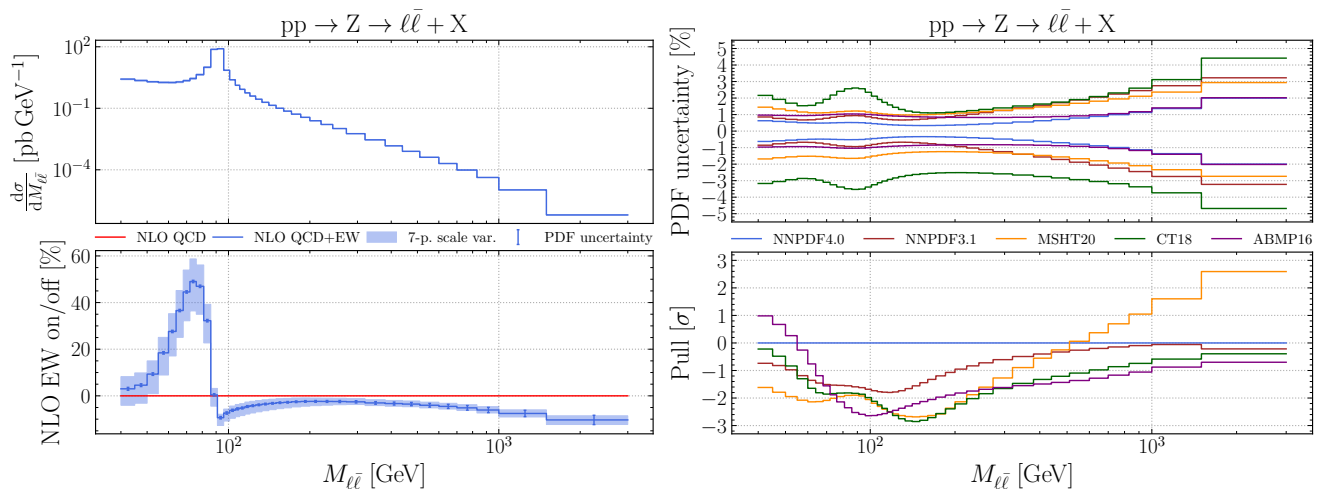


FIG. 21. NLO QCD+EW corrections for DY lepton-pair production at the LHC at 14 TeV, showing absolute predictions (top left) and relative size of the EW corrections (bottom left), PDF uncertainties (top right) and pulls (bottom right) for different PDF sets. See Ref. [10], Sec. 9.2 and 9.3 for full information; all plots have been generated with `PineAPPL`'s command-line interface.

leftmost panel, where the band illustrates the renormalisation scale uncertainty. The full structure of the factorisation scale dependence for the cross section can only be achieved in a realistic time using a fast grid technique to reproduce the scale variation. A more detailed illustration is seen in the rightmost panel which illustrates the full simultaneous variation of both the renormalisation and factorisation scales showing, with high granularity, the full scale plane for variations in the conventional range of factors from 0.5 to 2 for both the renormalisation and factorisation scales.

The full range of the cross section variation at N2LO is approximately 1.6% in total, with a saddle point close to the nominal cross section, much smaller than the 11% variation seen at NLO. As with the more significant cases of PDF fits, and fits to the strong coupling, such a detailed exploration of the behaviour of the cross section is only possible using a grid, and in this case evaluating the cross section only at the usual limits of the scale variations would not correctly determine the maximum of the cross section.

## 2. The `PineAPPL` interface

*Leading authors: A. Candido, F. Hekhorn, J. Cruz-Martinez, C. Schwan*

`PineAPPL` [283, 724] is the newest addition to the family of interpolation grid codes and was developed to support arbitrary coupling orders in  $\alpha_s$  and  $\alpha$ . In particular, this includes NLO EW corrections, but also mixed corrections like N2LO QCD–EW corrections, which are not supported by other interpolation grid libraries. Support for these corrections is needed to fit PDFs with these additional corrections [725].

`PineAPPL` is interfaced with `Madgraph5_aMC@NLO` [290] and `yadism` [726, 727], which we use to produce interpolation grids for hadron–hadron and hadron–lepton collider processes, respectively. Although `PineAPPL` is written in Rust, interfaces in C, C++, Fortran and Python are also provided so it can be easily integrated into any Monte Carlo (MC) generator. Interfaces to more MCs, including those with N2LO precision, are being worked on. Existing interpolation grids from `APPLgrid` and `fastNLO` can be converted into the `PineAPPL` format using one of the supplied programs. Finally, `PineAPPL` comes with a command-line interface, which allows the user to easily convolute grids with PDFs, and more operations such as: plot predictions and pulls (see Fig. 21), list the sizes of all partonic channels, show differences between two grids, show the size of the different coupling orders, calculate PDF uncertainties, calculate the pull between two PDF sets, etc. `PineAPPL` will be used in an updated version of `NNPDF`'s fitting code [368], for which a part that will be updated is shown in Fig. 22.

The interpolation grids generated by MCs are not directly used by the `NNPDF`'s fitting code, but instead they are first evolved into so-called Fast Kernel (FK) tables [380, 728, 729]

Using DGLAP equations the grids at factorization scale values set by the process are evolved to a (typically smaller) single scale, at which the PDFs are fitted. This procedure reduces the evaluation of theory predictions down to a simple linear algebra operation, which can be implemented very efficiently and easily parallelized.

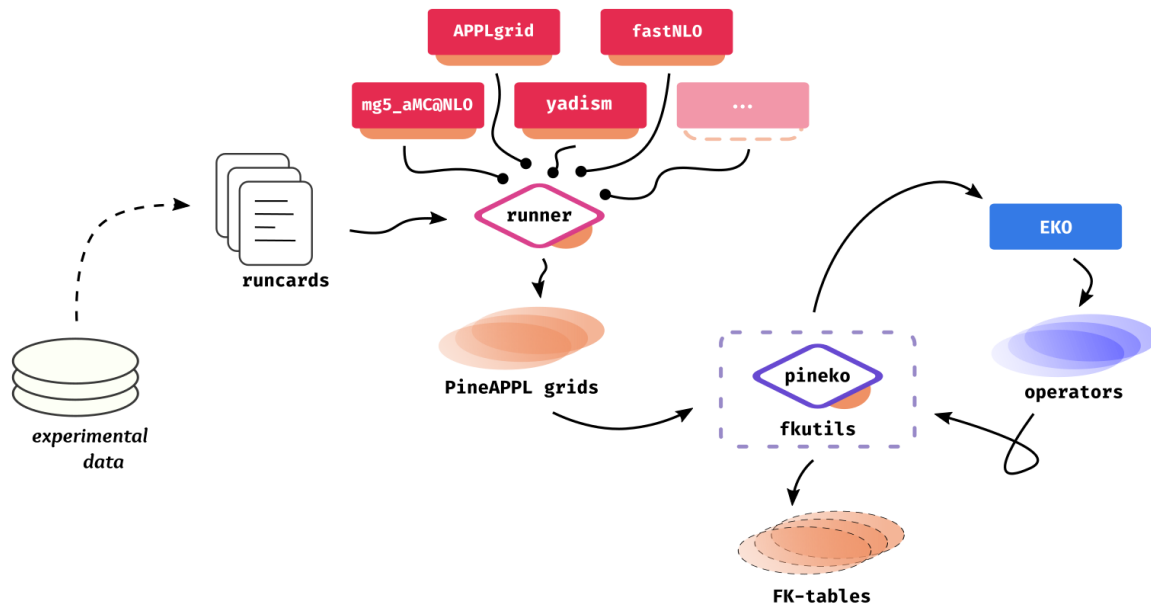


FIG. 22. Updated pipeline for NNPDF’s theory predictions. Runcards must be written specifying the parameters, phase-space cuts, scale choices, binning of observables, etc. matching the selected experimental data. Each runcard targets a specific MC generator, which is run by a dedicated runner tool (Madgraph5\_aMC@NLO, yadism, etc.) Alternatively, existing APPLgrids or fastNLO tables are converted to PineAPPL. In both cases the result are PineAPPL grids. Afterwards the interpolation grids are queried to generate suitable evolution operators with EKO, and finally the operators are consumed to produce the desired FK tables. All orange insets attached to the programs’ boxes represent usage of PineAPPL interface.

At the technical level these operations are shown in Fig. 22. First, a PineAPPL grid must be generated, either by converting existing APPLgrids and fastNLO tables, or by running programs that compute the PDF processes, for example yadism. In the second case runcards must be written, that specify how the process is calculated such that its predictions match the experimental measurements. Next, the PineAPPL grid is evolved into an FK table. This is performed by pineko, which instructs EKO [234, 235] to generate the evolution kernel operators (EKO), and subsequently uses the operators to perform the evolution. The program fkutils integrates this process for all the processes in NNPDF and finally provides the FK tables to the fitting code.

## X. BENCHMARKING AND COMBINATION OF GLOBAL PDF ANALYSES; PDF4LHC21 RECOMMENDATION

*Leading authors: A. M. Cooper-Sarkar, A. Courtoy, T. Cridge, J. Rojo, K. Xie*

The precise and accurate determination of the proton’s PDFs [30, 40] is a highly challenging endeavour, which requires tackling a number of issues from the limitations of fixed-order theory calculations, internal or external inconsistencies of the experimental measurements, ill-defined correlation models, choice of techniques for PDF error estimate and propagation, the choice of PDF parametrization, the implementation of theoretical constraints on the PDF shape like positivity, integrability, counting rules, or Regge theory behaviour, the treatment of the heavy quark PDFs, or the choice of SM parameters. Realizing an in-depth understanding of the differences and similarities between global PDF determinations requires carrying out dedicated benchmark exercises involving the close collaboration of the PDF fitting collaborations among them, as well as with the experimental groups.

In order to expedite progress in our understanding of proton structure, the PDF4LHC Working Group was established in 2008 [730] in order to coordinate scientific discussions and collaborative projects within the PDF theory and experimental LHC communities. The first PDF4LHC benchmarking exercise was performed in 2010 [731], resulting in an initial set of recommendations [732] for PDF usage at Run I of the LHC. Subsequently, several dedicated studies and benchmark exercises were carried out [733–736]. Then in 2015, following a year-long study, the PDF4LHC15 combined sets were released [16] together with updated set of recommendations for PDF usage and uncertainty estimate



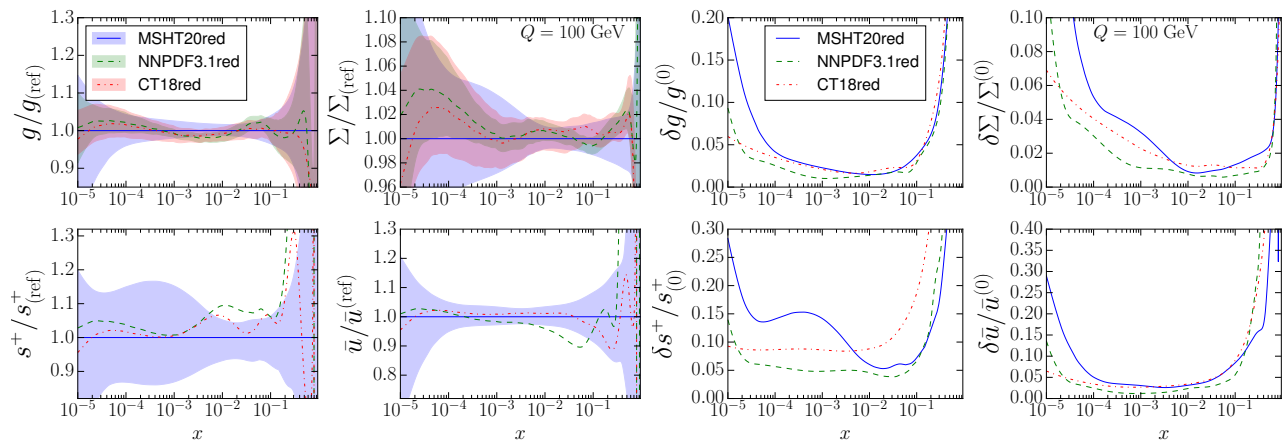


FIG. 23. Left: comparison between the reduced data set PDFs from the CT18, MSHT20, and NNPDF3.1 groups at  $Q = 100$  GeV (normalised to the central value of MSHT20). Right: same the corresponding one-sigma PDF uncertainties.

at the LHC Run II. PDF4LHC15 was based on the combination of the CT14 [2], MMHT2014 [1], and NNPDF3.0 [737] global analyses and was made possible thanks to developments regarding the transformation of Hessian PDF sets into their MC representation [738] and vice-versa [390–392] and the replica compression of MC sets [385].

Since the release of PDF4LHC15, several developments have taken place in topics of direct relevance for global PDF determinations. First of all, the availability of a large number of new data sets from the LHC, which provide significant constraints on the proton PDFs in a wide kinematic range and for many complementary flavour combinations. Second, the completion of crucial N2LO QCD calculations [739] for processes such as inclusive jet [740] and dijet [741] production to direct photon production [37], differential top quark pair production [742], and charged-current deep-inelastic scattering with heavy quark mass effects [743], of key relevance for global PDF fits [362, 744–748]. Third, steady progress in the developments of novel fitting methodologies, such as improved parameterisation strategies and machine learning techniques. An update of the PDF4LHC15 combination was both timely and relevant, especially taking into account the upcoming restart of data-taking at the LHC during its Run III and subsequently of its high-luminosity era [12, 159].

This state of affairs has motivated the very recent PDF4LHC21 study [24] based on the combination of three updated global PDF analyses, CT18 [7], MSHT2020 [9], and NNPDF3.1 [108], and the subsequent assessment of its implications for the phenomenology program of the LHC Run III. A requisite for this new combination has been an extensive set of benchmark studies [36] aiming to better pinpoint the origin of the differences between the three global PDF fits either in terms of the input data, the theory settings, or the fitting methodology. Special attention has also been paid in these benchmarking exercises to the role played by the assumptions underlying the experimental correlation models in the interpretation of high-precision LHC measurements, which are often limited by systematic uncertainties [8, 362, 363, 744, 749–754]. The PDF4LHC21 study also benefited from the lessons provided by PDF analysis carried out by ATLAS [11, 755] and CMS [58]. The results of the benchmarking studies carried out in the context of the PDF4LHC21 combination demonstrate that the differences observed between the three global PDF sets can be explained by genuinely valid choices related to the input data set, theory settings, and fitting methodology adopted in each case.

One of the main ingredients of the PDF4LHC21 benchmarking study has been the production and comparisons of variants of the CT18, MSHT20, and NNPDF3.1 fits based on a reduced, identical data set, and where furthermore one has striven to homogenize as much as possible the settings of the underlying theory calculations [36]. Fig. 23 compares the reduced-data set PDFs from the CT18, MSHT20, and NNPDF3.1 groups at  $Q = 100$  GeV, normalised to the central value of MSHT20, as well as the corresponding one-sigma PDF uncertainties. Good agreement between the three reduced fits is found, and in particular their agreement is improved as compared to the corresponding global fits based on the baseline data set from each group. This good agreement is clearly visible i.e. for the gluon and the total quark singlet PDFs across the whole range of  $x$ . Some differences observed in the baseline fits also persist in the reduced fits, such as in the magnitude of the PDF uncertainties. This observation indicates that the methodological choices adopted by each group, for example due to the parametrization form, tolerance, or fitting methodology, remain significant even when fitting the same data set and can be, in some cases, as large or even larger than the PDF uncertainties associated with the input fitted data.

Having established that the differences between CT18, MSHT20, and NNPDF3.1 are mostly dominated by differences associated to valid choices related to methodology and data set, the three global fits are combined in their MC

representation [738] with  $N_{\text{rep}} = 300$  replicas for each group, to form the PDF4LHC21 baseline set with  $N_{\text{rep}} = 900$  replicas. Fig. 24 displays the comparison between the partonic luminosities at 14 TeV of the resulting PDF4LHC21 baseline set and PDF4LHC15 (in this case, the compressed MC variant with  $N_{\text{rep}} = 100$  replicas). We display the quark-quark, gluon-gluon, and quark-antiquark luminosities at the LHC 14 TeV, normalised to the central value of PDF4LHC21 in the upper panels, and for their  $1\sigma$  relative uncertainty in the lower panels. Despite the many changes that the three constituent sets have undergone from the previous to the current combination, PDF4LHC21 does not only agree within uncertainties with PDF4LHC15 in the kinematic range relevant for the LHC, but also exhibits a moderate reduction of the PDF uncertainties in the gluon sector and for the quark luminosities in the invariant mass region  $m_X \leq 1$  TeV. Hence, theory predictions based on PDF4LHC21 will benefit from reduced PDF uncertainties for several precision LHC observables.

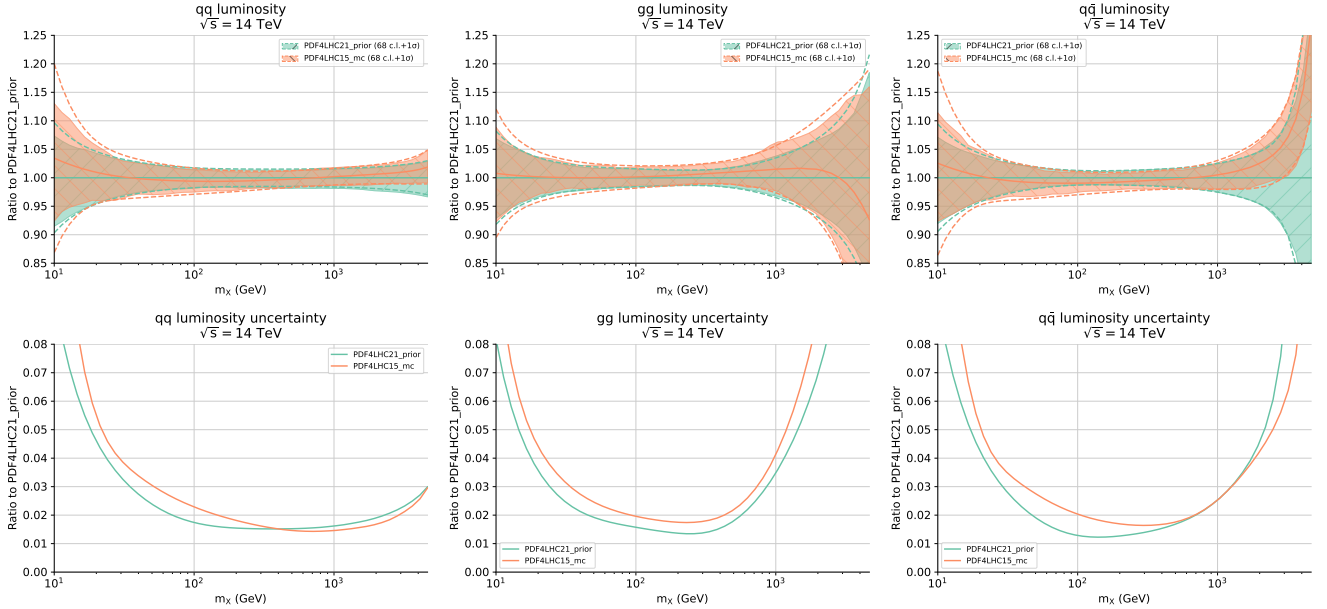


FIG. 24. Comparison between the partonic luminosities at the LHC 14 TeV of PDF4LHC21 (baseline combination with  $N_{\text{rep}} = 900$  replicas, labeled as “prior”) and of PDF4LHC15 (compressed set with  $N_{\text{rep}} = 100$  replicas). The upper panels display the ratio of the central value of PDF4LHC21, and the lower panels the relative  $1\sigma$  PDF uncertainty in each case.

As was the case of the previous combination [16], the  $N_{\text{rep}} = 900$  replicas of the PDF4LHC21 baseline are reduced down to a more manageable number for applications at the LHC. We have considered two different techniques to obtain a Hessian representation of PDF4LHC21, namely the META-PDF approach [390] and the MC2HESSIAN algorithm [391, 392]. The META-PDF method is based on constructing a common meta-parametrization of the replicas that constitute the baseline using Bernstein polynomials. All input replicas end up having associated the same parametric form, each with different numerical parameters of the Bernstein polynomials, the set of parameters and the degree of the polynomials define the meta-parametrization. Then, dimensionality reduction is performed in the space of meta-parameters by Principal Component Analysis (PCA). Within this method, it is also possible to impose the positive-definiteness of the central PDF member. The basic idea of MC2HESSIAN is to use the MC replicas of the prior themselves to construct a Hessian representation with the replicas’ linear expansion basis, and then to determine the numerical coefficients of the expansion to ensure that the mean, variance, and correlations of the baseline distribution are reproduced based on the combination of PCA and Singular Value Decomposition (SVD). After extensive comparison studies, the final deliverable Hessian set, PDF4LHC21\_40, was chosen to be the reduced set obtained through the updated META-PDF technique with  $N_{\text{eig}} = 40$ , that includes a feature to ensure positive-definite central PDFs of the resulting Hessian set.

The reduced Monte Carlo representation of PDF4LHC21 is also constructed by means of the replica compression algorithm [385, 386], whose goal is to extract the subset of the replicas that most faithfully reproduces the statistical properties of the prior distribution. The compression methodology relies on two main ingredients: a proper definition of a distance metric that quantifies the distinguishability between the baseline and the compressed probability distributions, and an appropriate minimization algorithm that explores the space of possible combinations of PDF replicas which leads to such a minima. The final deliverable MC set is the compressed MC to  $N_{\text{rep}} = 100$ , PDF4LHC21\_MC.

Fig. 25 compares the predictions for the  $1\sigma$  correlation ellipses at the LHC at  $\sqrt{s} = 14$  TeV for representative

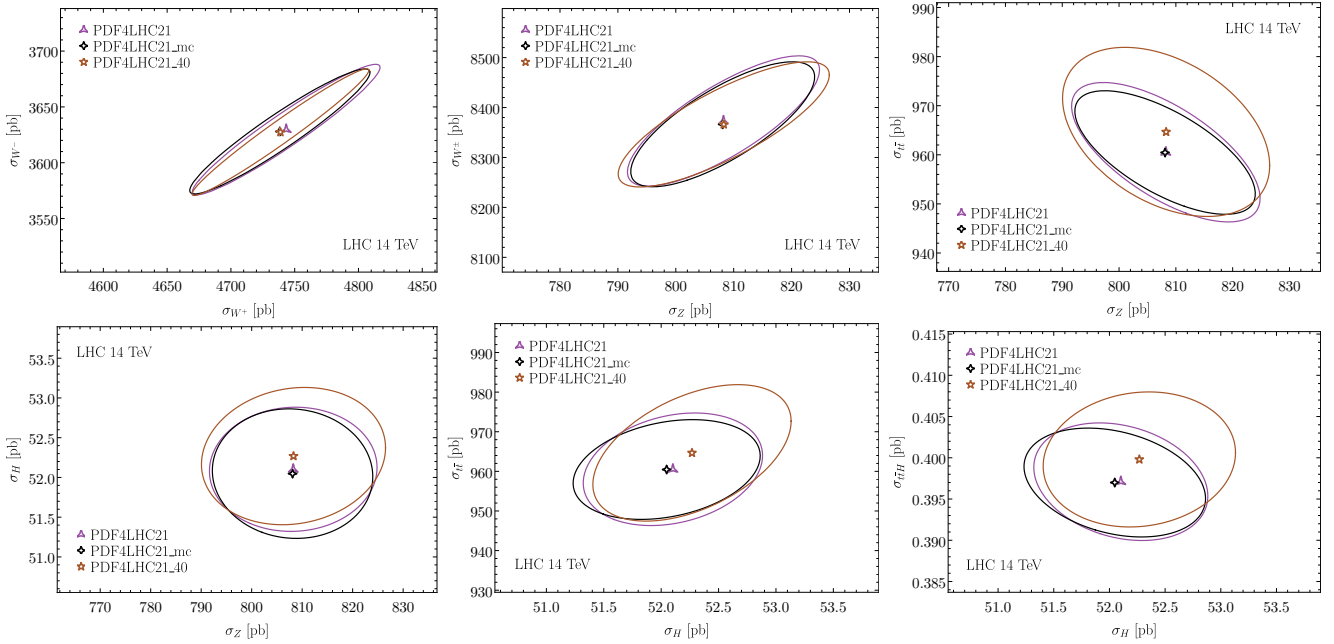


FIG. 25. Comparison between the predictions using the baseline PDF4LHC21 baseline set ( $N_{\text{rep}} = 900$  replicas) and those of its Hessian ( $N_{\text{eig}} = 40$ ) and compressed MC ( $N_{\text{rep}} = 100$ ) representations for the  $1\sigma$  correlation ellipses for pairs of inclusive cross sections among the  $W^\pm$ ,  $Z$ ,  $t\bar{t}$ ,  $H$ ,  $t\bar{t}H$  production processes at the LHC 14 TeV.

inclusive cross-sections between the PDF4LHC21 baseline combination and its Hessian, PDF4LHC21\_40 (with  $N_{\text{eig}} = 40$ ), and compressed Monte Carlo, PDF4LHC21\_MC (with  $N_{\text{rep}} = 100$ ), representations. We have considered the production of  $W^\pm$  and  $Z$  gauge bosons, top-quark pairs, Higgs bosons in gluon fusion, and  $t\bar{t}$  pairs associated with a Higgs boson. The  $W^\pm/Z$  cross sections correspond to the fiducial volume measured at ATLAS 13 TeV [38], while others refer to the full phase space. One finds in general good agreement between the baseline and its compressed MC and Hessian reduced sets. The small shift in central value in the Hessian set as compared to the baseline is related to imposing central PDF positivity in the former, and which effect is always contained within the uncertainties of the baseline set. Extensive comparisons for other LHC observables at the inclusive and differential level confirm that PDF4LHC21 is compatible with PDF4LHC15 while exhibiting a modest reduction of PDF uncertainties, and that furthermore the compressed MC and Hessian reduced sets provide an appropriate, user-friendly representation of the baseline combination. Additional studies of the phenomenological implications of PDF4LHC21 are reported in [24].

There are several directions in which the PDF4LHC21 studies could be expanded. To begin with, one could extend the analysis of PDF fits based on a common reduced data set by adding other measurements, since an even wider “reduced” data set could further highlight which of the differences observed between PDF groups can be traced back to the underlying methodological choices. One could also consider investigations of why the PDF uncertainties between various groups differ even when a similar input data set is considered. Also, the PDF4LHC21 combination will have to be eventually updated once new releases from the various PDF fitting collaborations are presented. Furthermore, future combinations will also have to account for not only the PDF contribution to the total uncertainty, but also other sources such as MHOUs which will be strongly correlated between the groups, as well as combinations between PDF sets including QED corrections and the photon PDF [263–265, 277, 278].

## XI. CONCLUSION: PRECISION PDFS IN THE UNITED STATES

Among several groups (ABM, CTEQ-TEA, HERAPDF, MMHT, NNPDF) working on the determination of general-purpose N2LO PDFs, one group (CTEQ-TEA [2, 7, 105, 752, 756–759]) is currently based in the US. Each general-purpose global analysis of PDFs is a major undertaking, involving significant investment in development, testing, and tuning of theoretical and computational frameworks. Recall that it took more than ten years from the publication of N2LO DGLAP equations [211, 215] to the release of N2LO PDF parametrizations with benchmarked accuracy [16]. Further advancements require support for the critical mass of the personnel with the specialized expertise. These advancements greatly benefit from the collaborations between experimentalists and theorists, and from international

collaborations.

Since the Electron-Ion Collider can provide powerful new constraints on large- $x$  PDFs [27], it makes sense to forge novel collaborations between the HEP and nuclear physics communities in the US. Unique data with high sensitivity to a wide range of PDF phenomena may be also collected with the LHeC experiment [28] at CERN in the 2030s. The US nuclear and particle physics community should feel encouraged to support CERN in its efforts to realize that unique experiment, and subsequently benefit from the open collaboration and access to these data. Looking even further into the future, the Muon-Ion Collider [29] in the US may become a factory of precision measurements of the hadron structure.

The precision physics frontier at the HL-LHC and EIC opens up new fascinating opportunities and challenges in the field of PDF determination. The HL-LHC projections are very encouraging, with a foreseen reduction of PDF uncertainties by factor 2-3. However, reaching this accuracy target requires coordinated advancements in experimental measurements, theoretical computations, and global analysis methodology. In particular, to be able to reduce the PDF uncertainties, the precision experiments that probe the PDFs should strive to reach better agreement among themselves than has been possible until now. We believe that, to reach such agreement, it is critical that new experiments and theory calculations implement consistent error control at all stages, from experimental measurements to the distribution of final PDFs. Efforts in this direction should go hand-in-hand with, and be as adequately supported as the investment into new conceptual advancements, such as the PDFs with electroweak constituents [273, 760, 761], as well as computations of new radiative contributions, such as those associated with N<sup>3</sup>LO QCD and NLO EW terms. As important is to continue development of the robust methodology for the global fits, including advanced statistical tests of the goodness of fit, methods for estimating theoretical uncertainties, novel statistical inference techniques inspired by the large-scale data science and artificial intelligence, and a new generation of computer programs for the global fits and fast multi-loop computations.

#### ACKNOWLEDGEMENTS

This work was supported by the U.S. Department of Energy under Contracts DE-AC02-76SF00515, DE-AC02-06CH11357, DE-FG02-91ER40684, DE-SC0010129, DE-SC0007914; by the U.S. National Science Foundation under Grants No. PHY-1820760, PHY-2112025, PHY-2013791; by the Deutsche Forschungsgemeinschaft (DFG) Research Unit FOR 292, project number 40824754; by the European Research Council under the European Union’s Horizon 2020 research and innovation Programme (grant agreements 740006, 824093, 950246); by the UK Royal Society grant DH150088; by the UK Science and Technology Facilities Council (STFC) grants ST/P000274/1, ST/L000377/1, ST/P000630/1, ST/T000600/1, ST/T000694/1, ST/T000856/1, ST/T000864/1.

N. Armesto acknowledges financial support by Xunta de Galicia (Centro singular de investigación de Galicia accreditation 2019-2022); the “María de Maeztu” Units of Excellence program MDM2016-0692 and the Spanish Research State Agency under project PID2020-119632GB-I00; European Union ERDF; the European Research Council under project ERC-2018-ADG-835105 YoctoLHC; MSCA RISE 823947 “Heavy ion collisions: collectivity and precision in saturation physics” (HIEIC). F. G. Celiberto acknowledges support from the INFN/NINPHA project and thanks the Università degli Studi di Pavia for the warm hospitality. A. M. Cooper-Sarkar wishes to thank the Leverhulme Trust. A. Courtoy is supported by the UNAM Grant No. DGAPA-PAPIIT IN111222 and CONACyT Ciencia de Frontera 2019 No. 51244 (FORDECYT-PRONACES). B. Malaescu gratefully acknowledges the continuous support from LPNHE, CNRS/IN2P3, Sorbonne Université and Université de Paris. J. Rojo is partially supported by the Dutch Science Council (NWO). A. Siódmok is supported by the National Science Centre, Poland, (Grant no. 2019/34/E/ST2/00457: A. Siódmok and J. Whitehead). The work of A. Siódmok was also funded by the Priority Research Area Digiworld under the program Excellence Initiative – Research University at the Jagiellonian University in Cracow. K. Xie was partly supported by PITT PACC. C.-P. Yuan is also grateful for the support from the Wu-Ki Tung endowed chair in particle physics. Bei Zhou was supported by the Simons Foundation.

- 
- [1] L. Harland-Lang, A. Martin, P. Motylinski, and R. Thorne, *Eur. Phys. J. C* **75**, 204 (2015), 1412.3989.
- [2] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C.-P. Yuan, *Phys. Rev. D* **93**, 033006 (2016), 1506.07443.
- [3] H. Abramowicz et al. (H1, ZEUS), *Eur. Phys. J. C* **75**, 580 (2015), 1506.06042.
- [4] A. Accardi, L. T. Brady, W. Melnitchouk, J. F. Owens, and N. Sato, *Phys. Rev. D* **93**, 114017 (2016), 1602.03154.
- [5] S. Alekhin, J. Blümlein, S. Moch, and R. Placakyte, *Phys. Rev. D* **96**, 014011 (2017), 1701.05838.
- [6] R. D. Ball et al. (NNPDF), *Eur. Phys. J. C* **77**, 663 (2017), 1706.00428.
- [7] T.-J. Hou et al., *Phys. Rev. D* **103**, 014013 (2021), 1912.10053.
- [8] S. Bailey and L. Harland-Lang, *Eur. Phys. J. C* **80**, 60 (2020), 1909.10541.
- [9] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, *Eur. Phys. J. C* **81**, 341 (2021), 2012.04684.
- [10] R. D. Ball et al. (2021), 2109.02653.
- [11] G. Aad et al. (ATLAS) (2021), 2112.11266.
- [12] M. Cepeda et al., *Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC* (2019), 1902.00134.
- [13] M. Ubiali (2020), a Letter of Interest submitted to the Snowmass 2021 Theory Frontier.
- [14] L. T. Brady, A. Accardi, W. Melnitchouk, and J. F. Owens, *JHEP* **06**, 019 (2012), 1110.5398.
- [15] R. D. Ball, V. Bertone, M. Bonvini, S. Marzani, J. Rojo, and L. Rottoli, *Eur. Phys. J. C* **78**, 321 (2018), 1710.05935.
- [16] J. Butterworth et al., *J. Phys. G* **43**, 023001 (2016), 1510.03865.
- [17] A. Accardi et al., *Eur. Phys. J. C* **76**, 471 (2016), 1603.08906.
- [18] H.-W. Lin et al., *Prog. Part. Nucl. Phys.* **100**, 107 (2018), 1711.07916.
- [19] M. Constantinou et al., *Prog. Part. Nucl. Phys.* **121**, 103908 (2021), 2006.08636.
- [20] C. Bauer et al. (2020), a Letter of Interest submitted to the Snowmass 2021 Theory Frontier.
- [21] E. Eichten, I. Hinchliffe, K. D. Lane, and C. Quigg, *Rev. Mod. Phys.* **56**, 579 (1984), [Addendum: *Rev. Mod. Phys.* **58**, 1065–1073 (1986)].
- [22] J. Campbell et al., in *Community Summer Study 2013: Snowmass on the Mississippi* (2013), 1310.5189.
- [23] The PDF4LHC working group, <https://www.hep.ucl.ac.uk/pdf4lhc/>.
- [24] R. D. Ball et al. (2022), 2203.05506.
- [25] The LHAPDF library, <https://lhapdf.hepforge.org/>.
- [26] R. Abdul Khalek, S. Bailey, J. Gao, L. Harland-Lang, and J. Rojo, *Eur. Phys. J. C* **78**, 962 (2018), 1810.03639.
- [27] R. Abdul Khalek et al. (2021), 2103.05419.
- [28] P. Agostini et al. (LHeC, FCC-he Study Group), *J. Phys. G* **48**, 110501 (2021), 2007.14491.
- [29] D. Acosta, E. Barberis, N. Hurley, W. Li, O. M. Colin, D. Wood, and X. Zuo, in *2022 Snowmass Summer Study* (2022), 2203.06258.
- [30] K. Kovařík, P. M. Nadolsky, and D. E. Soper, *Rev. Mod. Phys.* **92**, 045003 (2020), 1905.06957.
- [31] L. Del Debbio, T. Giani, and M. Wilson (2021), 2111.05787.
- [32] R. D. Ball et al. (NNPDF), *JHEP* **04**, 040 (2015), 1410.8849.
- [33] R. Abdul Khalek, S. Bailey, J. Gao, L. Harland-Lang, and J. Rojo, *Eur. Phys. J. C* **78**, 962 (2018), 1810.03639.
- [34] G. Aad et al. (ATLAS), *Phys. Rev. D* **85**, 072004 (2012), 1109.5141.
- [35] R. D. Ball, V. Bertone, M. Bonvini, S. Carrazza, S. Forte, A. Guffanti, N. P. Hartland, J. Rojo, and L. Rottoli (NNPDF), *Eur. Phys. J. C* **76**, 647 (2016), 1605.06515.
- [36] T. Cridge (PDF4LHC21 combination group), in *28th International Workshop on Deep Inelastic Scattering and Related Subjects* (2021), 2108.09099.
- [37] J. M. Campbell, R. K. Ellis, and C. Williams, *Phys. Rev. Lett.* **118**, 222001 (2017), [Erratum: *Phys. Rev. Lett.* **124**, 259901 (2020)], 1612.04333.
- [38] G. Aad et al. (ATLAS), *Phys. Lett. B* **759**, 601 (2016), 1603.09222.
- [39] D. de Florian et al. (LHC Higgs Cross Section Working Group), **2/2017** (2016), 1610.07922.
- [40] J. Gao, L. Harland-Lang, and J. Rojo, *Phys. Rept.* **742**, 1 (2018), 1709.04922.
- [41] C. Anastasiou, C. Duhr, F. Dulat, E. Furlan, T. Gehrmann, F. Herzog, A. Lazopoulos, and B. Mistlberger, *JHEP* **05**, 058 (2016), 1602.00695.
- [42] B. Mistlberger, *JHEP* **05**, 028 (2018), 1802.00833.
- [43] C. Duhr, F. Dulat, and B. Mistlberger, *Phys. Rev. Lett.* **125**, 051804 (2020), 1904.09990.
- [44] C. Duhr and B. Mistlberger (2021), 2111.10379.
- [45] H. Abreu et al. (FASER), *Eur. Phys. J. C* **80**, 61 (2020), 1908.02310.
- [46] A. M. Sirunyan et al. (CMS), *JHEP* **06**, 120 (2018), 1803.06292.
- [47] G. Aad et al. (ATLAS), *Phys. Lett. B* **796**, 68 (2019), 1903.06248.
- [48] J. Adelman, J. Ferrando, and C. D. White, *JHEP* **02**, 091 (2013), 1206.5731.
- [49] M. Guzzi and N. Kidonakis, *Eur. Phys. J. C* **80**, 467 (2020), 1904.10071.
- [50] A. V. Manohar, *Lect. Notes Phys.* **479**, 311 (1997), hep-ph/9606222.
- [51] I. Brivio and M. Trott, *Phys. Rept.* **793**, 1 (2019), 1706.08945.
- [52] A. Greljo, S. Iranipour, Z. Kassabov, M. Madigan, J. Moore, J. Rojo, M. Ubiali, and C. Voisey, *JHEP* **07**, 122 (2021), 2104.02723.
- [53] R. Boughezal, F. Petriello, K. Şimşek, X. Zheng, et al. (2022), 22xx.yyyy.



- [54] S. Iranipour and M. Ubiali (2022), 2201.07240.
- [55] D. Liu, C. Sun, and J. Gao (2022), 2201.06586.
- [56] S. Carrazza, C. Degrande, S. Iranipour, J. Rojo, and M. Ubiali, *Phys. Rev. Lett.* **123**, 132001 (2019), 1905.05215.
- [57] H. Abramowicz et al. (ZEUS), *Phys. Rev. D* **99**, 092006 (2019), 1902.03048.
- [58] A. Tumasyan et al. (CMS) (2021), 2111.10431.
- [59] M. Madigan and J. Moore, *PoS EPS-HEP2021*, 424 (2022), 2110.13204.
- [60] R. Boughezal, F. Petriello, and D. Wiegand, *Phys. Rev. D* **101**, 116002 (2020), 2004.00748.
- [61] R. Boughezal, F. Petriello, and D. Wiegand, *Phys. Rev. D* **104**, 016005 (2021), 2104.03979.
- [62] M. Aaboud et al. (ATLAS), *Eur. Phys. J. C* **77**, 367 (2016), 1612.03016.
- [63] G. Aad et al. (ATLAS), *JHEP* **08**, 009 (2016), 1606.01736.
- [64] M. Aaboud et al. (ATLAS), *JHEP* **12**, 059 (2017), 1710.05167.
- [65] G. Aad et al. (ATLAS), *Eur. Phys. J. C* **79**, 760 (2019), 1904.05631.
- [66] R. Aaij et al. (LHCb), *JHEP* **08**, 039 (2015), 1505.07024.
- [67] R. Aaij et al. (LHCb), *JHEP* **01**, 155 (2016), 1511.08039.
- [68] R. Aaij et al. (LHCb), *JHEP* **05**, 109 (2015), 1503.00963.
- [69] R. Aaij et al. (LHCb), *JHEP* **09**, 136 (2016), 1607.06495.
- [70] V. Khachatryan et al. (CMS), *Eur. Phys. J. C* **76**, 469 (2016), 1603.01803.
- [71] G. Aad et al. (ATLAS), *JHEP* **02**, 153 (2015), [Erratum: *JHEP* 09, 141 (2015)], 1410.8857.
- [72] M. Aaboud et al. (ATLAS), *JHEP* **09**, 020 (2017), 1706.03192.
- [73] M. Aaboud et al. (ATLAS), *JHEP* **05**, 195 (2018), 1711.02692.
- [74] S. Chatrchyan et al. (CMS), *Phys. Rev. D* **90**, 072006 (2014), 1406.0324.
- [75] V. Khachatryan et al. (CMS), *JHEP* **03**, 156 (2017), 1609.05331.
- [76] V. Khachatryan et al. (CMS), *Eur. Phys. J. C* **76**, 265 (2016), 1512.06212.
- [77] S. Alte, M. König, and W. Shepherd, *JHEP* **01**, 094 (2018), 1711.07484.
- [78] M. Aaboud et al. (ATLAS), *JHEP* **05**, 077 (2018), [Erratum: *JHEP* 10, 048 (2020)], 1711.03296.
- [79] G. Aad et al. (ATLAS), *Eur. Phys. J. C* **79**, 847 (2019), 1907.06728.
- [80] G. Aad et al. (ATLAS), *Eur. Phys. J. C* **76**, 291 (2016), 1512.02192.
- [81] G. Aad et al. (ATLAS), *Eur. Phys. J. C* **76**, 538 (2016), 1511.04716.
- [82] M. Aaboud et al. (ATLAS), *Phys. Rev. D* **94**, 092003 (2016), [Addendum: *Phys.Rev.D* 101, 119901 (2020)], 1607.07281.
- [83] G. Aad et al. (ATLAS), *Eur. Phys. J. C* **79**, 1028 (2019), [Erratum: *Eur.Phys.J.C* 80, 1092 (2020)], 1908.07305.
- [84] A. M. Sirunyan et al. (CMS), *Eur. Phys. J. C* **77**, 459 (2017), 1703.01630.
- [85] V. Khachatryan et al. (CMS), *Eur. Phys. J. C* **75**, 542 (2015), 1505.04480.
- [86] S. Chatrchyan et al. (CMS), *JHEP* **02**, 024 (2014), [Erratum: *JHEP* 02, 102 (2014)], 1312.7582.
- [87] A. M. Sirunyan et al. (CMS), *JHEP* **02**, 149 (2019), 1811.06625.
- [88] A. M. Sirunyan et al. (CMS), *Phys. Rev. D* **97**, 112003 (2018), 1803.08856.
- [89] M. Aaboud et al. (ATLAS), *JHEP* **04**, 093 (2019), 1901.10075.
- [90] M. Czakon, A. Mitov, M. Pellen, and R. Poncelet, *JHEP* **06**, 100 (2021), 2011.01011.
- [91] E. R. Nocera, M. Ubiali, and C. Voisey, *JHEP* **05**, 067 (2020), 1912.09543.
- [92] (2020).
- [93] M. V. Garzelli, S. Moch, O. Zenaiev, A. Cooper-Sarkar, A. Geiser, K. Lipka, R. Placakyte, and G. Sigl (PROSA), *JHEP* **05**, 004 (2017), 1611.03815.
- [94] O. Zenaiev, M. V. Garzelli, K. Lipka, S. O. Moch, A. Cooper-Sarkar, F. Olness, A. Geiser, and G. Sigl (PROSA), *JHEP* **04**, 118 (2020), 1911.13164.
- [95] M. Aaboud et al. (ATLAS), *JHEP* **10**, 182 (2017), 1707.02424.
- [96] G. Aad et al. (ATLAS), *Phys. Lett. B* **754**, 302 (2016), 1512.01530.
- [97] M. Aaboud et al. (ATLAS), *Eur. Phys. J. C* **78**, 110 (2018), [Erratum: *Eur.Phys.J.C* 78, 898 (2018)], 1701.07240.
- [98] A. M. Sirunyan et al. (CMS), *Eur. Phys. J. C* **78**, 701 (2018), 1806.00863.
- [99] A. Accardi et al., *Eur. Phys. J. A* **52**, 268 (2016), 1212.1701.
- [100] E. C. Aschenauer, S. Fazio, J. H. Lee, H. Mantysaari, B. S. Page, B. Schenke, T. Ullrich, R. Venugopalan, and P. Zurita, *Rept. Prog. Phys.* **82**, 024301 (2019), 1708.01527.
- [101] A. C. Aguilar et al., *Eur. Phys. J. A* **55**, 190 (2019), 1907.08218.
- [102] J. Arrington et al., *J. Phys. G* **48**, 075106 (2021), 2102.11788.
- [103] A. Accardi, T. J. Hobbs, X. Jing, and P. M. Nadolsky, *Eur. Phys. J. C* **81**, 603 (2021), 2102.01107.
- [104] R. D. Ball, E. R. Nocera, and R. L. Pearson, *Eur. Phys. J. C* **81**, 37 (2021), 2011.00009.
- [105] B.-T. Wang, T. Hobbs, S. Doyle, J. Gao, T.-J. Hou, P. M. Nadolsky, and F. I. Olness, *Phys. Rev. D* **98**, 094030 (2018), 1803.02777.
- [106] R. A. Khalek, J. J. Ethier, E. R. Nocera, and J. Rojo, *Phys. Rev. D* **103**, 096005 (2021), 2102.00018.
- [107] E. Moffat, W. Melnitchouk, T. C. Rogers, and N. Sato (Jefferson Lab Angular Momentum (JAM)), *Phys. Rev. D* **104**, 016015 (2021), 2101.04664.
- [108] R. D. Ball et al. (NNPDF), *Eur. Phys. J. C* **77**, 663 (2017), 1706.00428.
- [109] T. J. Hobbs, J. T. Londergan, D. P. Murdock, and A. W. Thomas, *Phys. Lett. B* **698**, 123 (2011), 1101.3923.
- [110] E. C. Aschenauer, I. Borsa, R. Sassot, and C. Van Hulse, *Phys. Rev. D* **99**, 094004 (2019), 1902.10663.
- [111] A. Metz and A. Vossen, *Prog. Part. Nucl. Phys.* **91**, 136 (2016), 1607.02521.
- [112] N. Sato, C. Andres, J. J. Ethier, and W. Melnitchouk (JAM), *Phys. Rev. D* **101**, 074020 (2020), 1905.03788.

- [113] M. Arratia, Y. Furetova, T. J. Hobbs, F. Olness, and S. J. Sekula, Phys. Rev. D **103**, 074023 (2021), 2006.12520.
- [114] T. J. Hobbs, M. Alberg, and G. A. Miller, Phys. Rev. D **96**, 074023 (2017), 1707.06711.
- [115] P. A. Zyla et al. (Particle Data Group), PTEP **2020**, 083C01 (2020).
- [116] J. J. Ethier and E. R. Nocera, Ann. Rev. Nucl. Part. Sci. pp. 1–34 (2020), 2001.07722.
- [117] E. C. Aschenauer, R. Sassot, and M. Stratmann, Phys. Rev. D **86**, 054020 (2012), 1206.6014.
- [118] E. C. Aschenauer, T. Burton, T. Martini, H. Spiesberger, and M. Stratmann, Phys. Rev. D **88**, 114025 (2013), 1309.5327.
- [119] E. C. Aschenauer, R. Sassot, and M. Stratmann, Phys. Rev. D **92**, 094030 (2015), 1509.06489.
- [120] R. D. Ball, S. Forte, A. Guffanti, E. R. Nocera, G. Ridolfi, and J. Rojo (NNPDF), Phys. Lett. B **728**, 524 (2014), 1310.0461.
- [121] Y. Zhou, C. Cocuzza, F. Delcarro, W. Melnitchouk, A. Metz, and N. Sato (Jefferson Lab Angular Momentum (JAM)), Phys. Rev. D **104**, 034028 (2021), 2105.04434.
- [122] D. de Florian, R. Sassot, M. Stratmann, and W. Vogelsang, Phys. Rev. Lett. **113**, 012001 (2014), 1404.4293.
- [123] D. De Florian, G. A. Lucero, R. Sassot, M. Stratmann, and W. Vogelsang, Phys. Rev. D **100**, 114027 (2019), 1902.10548.
- [124] J. J. Ethier, N. Sato, and W. Melnitchouk, Phys. Rev. Lett. **119**, 132001 (2017), 1705.05889.
- [125] E. R. Nocera, R. D. Ball, S. Forte, G. Ridolfi, and J. Rojo (NNPDF), Nucl. Phys. B **887**, 276 (2014), 1406.5539.
- [126] Y. V. Kovchegov, D. Pitonyak, and M. D. Sievert, JHEP **01**, 072 (2016), [Erratum: JHEP 10, 148 (2016)], 1511.06737.
- [127] Y. V. Kovchegov, D. Pitonyak, and M. D. Sievert, Phys. Rev. D **95**, 014033 (2017), 1610.06197.
- [128] Y. V. Kovchegov, D. Pitonyak, and M. D. Sievert, Phys. Rev. Lett. **118**, 052001 (2017), 1610.06188.
- [129] Y. V. Kovchegov and M. D. Sievert, Phys. Rev. D **99**, 054032 (2019), 1808.09010.
- [130] X.-D. Ji, Phys. Rev. Lett. **78**, 610 (1997), hep-ph/9603249.
- [131] I. Borsa, G. Lucero, R. Sassot, E. C. Aschenauer, and A. S. Nunes, Phys. Rev. D **102**, 094018 (2020), 2007.08300.
- [132] Y. Zhou, N. Sato, and W. Melnitchouk (2022), 2201.02075.
- [133] B. S. Page, X. Chu, and E. C. Aschenauer, Phys. Rev. D **101**, 072003 (2020), 1911.00657.
- [134] F. Hekhorn and M. Stratmann, Phys. Rev. D **98**, 014018 (2018), 1805.09026.
- [135] D. P. Anderle, X. Dong, F. Hekhorn, M. Kelsey, S. Radhakrishnan, E. Sichtermann, L. Xia, H. Xing, F. Yuan, and Y. Zhao, Phys. Rev. D **104**, 114039 (2021), 2110.04489.
- [136] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, Eur. Phys. J. C **77**, 163 (2017), 1612.05741.
- [137] K. Kovarik et al., Phys. Rev. D **93**, 085037 (2016), 1509.00792.
- [138] R. Abdul Khalek, J. J. Ethier, J. Rojo, and G. van Weelden, JHEP **09**, 183 (2020), 2006.14629.
- [139] E. C. Aschenauer, S. Fazio, M. A. C. Lamont, H. Paukkunen, and P. Zurita, Phys. Rev. D **96**, 114005 (2017), 1708.05654.
- [140] (1984).
- [141] J. L. Abelleira Fernandez et al. (LHeC Study Group), J. Phys. G **39**, 075001 (2012), 1206.2913.
- [142] K. D. J. André et al., Eur. Phys. J. C **82**, 40 (2022), 2201.02436.
- [143] D. Angal-Kalinin et al., J. Phys. G **45**, 065003 (2018), 1705.08783.
- [144] C. Adolphsen et al., **1/2022** (2022), 2201.07895.
- [145] A. Abada et al. (FCC), Eur. Phys. J. ST **228**, 755 (2019).
- [146] R. Abdul Khalek, S. Bailey, J. Gao, L. Harland-Lang, and J. Rojo, SciPost Phys. **7**, 051 (2019), 1906.10127.
- [147] M. Klein and V. Radescu (2013), URL <https://cds.cern.ch/record/1564929>.
- [148] M. Arratia, D. Britzger, O. Long, and B. Nachman, Nucl. Instrum. Meth. A **1025**, 166164 (2022), 2110.05505.
- [149] A. Dainese, M. Mangano, A. B. Meyer, A. Nisati, G. Salam, and M. A. Vesterinen, eds., *Report on the Physics at the HL-LHC, and Perspectives for the HE-LHC*, vol. 7/2019 of *CERN Yellow Reports: Monographs* (CERN, Geneva, Switzerland, 2019), ISBN 978-92-9083-549-3.
- [150] H. Abdolmaleki et al. (xFitter Developers' Team), Eur. Phys. J. C **79**, 864 (2019), 1907.01014.
- [151] V. Andreev et al. (H1), Eur. Phys. J. C **77**, 791 (2017), [Erratum: Eur.Phys.J.C 81, 738 (2021)], 1709.07251.
- [152] X. Liu, F. Ringer, W. Vogelsang, and F. Yuan, Phys. Rev. Lett. **122**, 192003 (2019), 1812.08077.
- [153] V. Andreev et al. (H1) (2021), 2108.12376.
- [154] H. Abramowicz et al. (H1, ZEUS), Eur. Phys. J. C **75**, 580 (2015), 1506.06042.
- [155] N. Armesto, P. R. Newman, W. Słomiński, and A. M. Staśto, Phys. Rev. D **100**, 074022 (2019), 1901.09076.
- [156] V. Andreev et al. (H1), Eur. Phys. J. C **78**, 777 (2018), 1806.01176.
- [157] D. Britzger, M. Klein, and H. Spiesberger, Eur. Phys. J. C **80**, 831 (2020), 2007.11799.
- [158] S. Schael et al. (ALEPH, DELPHI, L3, OPAL, SLD, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group), Phys. Rept. **427**, 257 (2006), hep-ex/0509008.
- [159] P. Azzi et al., CERN Yellow Rep. Monogr. **7**, 1 (2019), 1902.04070.
- [160] M. E. Peskin and T. Takeuchi, Phys. Rev. D **46**, 381 (1992).
- [161] D. Britzger, M. Klein, and H. Spiesberger, in *European Physical Society Conference on High Energy Physics 2021* (2022), 2203.06237.
- [162] R. Gandhi, C. Quigg, M. H. Reno, and I. Sarcevic, Astropart. Phys. **5**, 81 (1996), hep-ph/9512364.
- [163] R. Gandhi, C. Quigg, M. H. Reno, and I. Sarcevic, Phys. Rev. **D58**, 093009 (1998), hep-ph/9807264.
- [164] A. Cooper-Sarkar, P. Mertsch, and S. Sarkar, JHEP **08**, 042 (2011), 1106.3723.
- [165] A. Connolly, R. S. Thorne, and D. Waters, Phys. Rev. **D83**, 113009 (2011), 1102.0691.
- [166] C.-Y. Chen, P. S. Bhupal Dev, and A. Soni, Phys. Rev. **D89**, 033012 (2014), 1309.1764.
- [167] V. Bertone, R. Gauld, and J. Rojo, JHEP **01**, 217 (2019), 1808.02034.
- [168] R. Acciarri et al. (DUNE) (2015), 1512.06148.
- [169] <http://www-sk.icrr.u-tokyo.ac.jp/sk/index-e.html>.

- [170] K. Abe et al. (Hyper-Kamiokande) (2018), 1805.04163.
- [171] <https://icecube.wisc.edu/>.
- [172] S. Adrian-Martinez et al. (KM3NeT), J. Phys. **G43**, 084001 (2016), 1601.07459.
- [173] A. D. Avrorin et al. (Baikal-GVD), EPJ Web Conf. **191**, 01006 (2018), 1808.10353.
- [174] M. G. Aartsen et al. (IceCube-Gen2), J. Phys. G **48**, 060501 (2021), 2008.04323.
- [175] P. W. Gorham et al. (ANITA), Phys. Rev. D **82**, 022004 (2010), [Erratum: Phys.Rev.D 85, 049901 (2012)], 1003.2961.
- [176] P. Allison et al. (ARA), Phys. Rev. D **93**, 082003 (2016), 1507.08991.
- [177] J. Álvarez-Muñiz et al. (GRAND), Sci. China Phys. Mech. Astron. **63**, 219501 (2020), 1810.09994.
- [178] C. Quigg, M. H. Reno, and T. P. Walker, Phys. Rev. Lett. **57**, 774 (1986).
- [179] J. Gao, T. J. Hobbs, P. M. Nadolsky, C. Sun, and C. P. Yuan, Phys. Rev. D **105**, L011503 (2022), 2107.00460.
- [180] M. A. G. Aivazis, F. I. Olness, and W.-K. Tung, Phys. Rev. D **50**, 3085 (1994), hep-ph/9312318.
- [181] M. A. G. Aivazis, J. C. Collins, F. I. Olness, and W.-K. Tung, Phys. Rev. D **50**, 3102 (1994), hep-ph/9312319.
- [182] R. S. Thorne and R. G. Roberts, Phys. Rev. D **57**, 6871 (1998), hep-ph/9709442.
- [183] M. Cacciari, M. Greco, and P. Nason, JHEP **05**, 007 (1998), hep-ph/9803400.
- [184] M. Gluck, P. Jimenez-Delgado, and E. Reya, Phys. Rev. D **81**, 097501 (2010), 1003.3168.
- [185] V. P. Goncalves and P. Hepp, Phys. Rev. D **83**, 014014 (2011), 1011.2718.
- [186] J. L. Albacete, J. I. Illana, and A. Soto-Ontoso, Phys. Rev. D **92**, 014027 (2015), 1505.06583.
- [187] C. A. Argüelles, F. Halzen, L. Wille, M. Kroll, and M. H. Reno, Phys. Rev. D **92**, 074040 (2015), 1504.06639.
- [188] L. Frankfurt, V. Guzey, and M. Strikman, Phys. Rept. **512**, 255 (2012), 1106.2091.
- [189] R. Abdul Khalek, J. J. Ethier, and J. Rojo (NNPDF), Eur. Phys. J. C **79**, 471 (2019), 1904.00018.
- [190] M. Walt, I. Helenius, and W. Vogelsang, Phys. Rev. D **100**, 096015 (2019), 1908.03355.
- [191] A. Garcia, R. Gauld, A. Heijboer, and J. Rojo, JCAP **09**, 025 (2020), 2004.04756.
- [192] S. R. Klein, S. A. Robertson, and R. Vogt, Phys. Rev. C **102**, 015808 (2020), 2001.03677.
- [193] D. Seckel, Phys. Rev. Lett. **80**, 900 (1998), hep-ph/9709290.
- [194] I. Alikhanov, Phys. Lett. **B756**, 247 (2016), 1503.08817.
- [195] B. Zhou and J. F. Beacom, Phys. Rev. D **101**, 036011 (2020), 1910.08090.
- [196] B. Zhou and J. F. Beacom, Phys. Rev. D **101**, 036010 (2020), 1910.10720.
- [197] B. Zhou and J. F. Beacom (2021), 2110.02974.
- [198] L. A. Anchordoqui et al. (2021), 2109.10905.
- [199] O. Zenaiev et al. (PROSA), Eur. Phys. J. C **75**, 396 (2015), 1503.04581.
- [200] R. Aaij et al. (LHCb) (2021), 2109.08084.
- [201] Q. Deng, Q. Han, H. Yin, S. Dulat, T.-J. Hou, and C. P. Yuan, Chin. Phys. C **45**, 023110 (2021), 2009.03181.
- [202] F. G. Celiberto, D. Gordo Gomez, and A. Sabio Vera, Phys. Lett. **B786**, 201 (2018), 1808.09511.
- [203] M. Hentschinski et al., in *2022 Snowmass Summer Study* (2022), 2203.08129.
- [204] J. L. Feng et al., in *2022 Snowmass Summer Study* (2022), 2203.05090.
- [205] G. Altarelli and G. Parisi, Nucl. Phys. B **126**, 298 (1977).
- [206] Y. L. Dokshitzer, Sov. Phys. JETP **46**, 641 (1977).
- [207] V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. **15**, 438 (1972).
- [208] J. Ablinger, A. Behring, J. Blümlein, A. De Freitas, A. von Manteuffel, and C. Schneider, Nucl. Phys. B **890**, 48 (2014), 1409.1135.
- [209] J. Ablinger, A. Behring, J. Blümlein, A. De Freitas, A. von Manteuffel, and C. Schneider, Nucl. Phys. B **922**, 1 (2017), 1705.01508.
- [210] J. Blümlein, P. Marquard, C. Schneider, and K. Schönwald, Nucl. Phys. B **971**, 115542 (2021), 2107.06267.
- [211] S. Moch, J. A. M. Vermaseren, and A. Vogt, Nucl. Phys. **B688**, 101 (2004), hep-ph/0403192.
- [212] S. Moch, J. A. M. Vermaseren, and A. Vogt, Nucl. Phys. B **889**, 351 (2014), 1409.5131.
- [213] J. Blümlein, P. Marquard, C. Schneider, and K. Schönwald, JHEP **01**, 193 (2022), 2111.12401.
- [214] S. Moch, J. A. M. Vermaseren, and A. Vogt, Phys. Lett. B **748**, 432 (2015), 1506.04517.
- [215] A. Vogt, S. Moch, and J. A. M. Vermaseren, Nucl. Phys. **B691**, 129 (2004), hep-ph/0404111.
- [216] S. Moch, B. Ruijl, T. Ueda, J. Vermaseren, and A. Vogt, JHEP **10**, 041 (2017), 1707.08315.
- [217] S. Moch, B. Ruijl, T. Ueda, J. M. Vermaseren, and A. Vogt, Phys. Lett. B **782**, 627 (2018), 1805.09638.
- [218] S. Moch, B. Ruijl, T. Ueda, J. A. M. Vermaseren, and A. Vogt, Phys. Lett. B **825**, 136853 (2022), 2111.15561.
- [219] J. Blumlein, H. Bottcher, and A. Guffanti, Nucl. Phys. B **774**, 182 (2007), hep-ph/0607200.
- [220] J. Vermaseren, A. Vogt, and S. Moch, Nucl. Phys. B **724**, 3 (2005), hep-ph/0504242.
- [221] J. Ablinger, A. Behring, J. Blümlein, A. De Freitas, A. Hasselhuhn, A. von Manteuffel, M. Round, C. Schneider, and F. Wißbrock, Nucl. Phys. B **886**, 733 (2014), 1406.4654.
- [222] J. Blümlein and M. Saragnese, Phys. Lett. B **820**, 136589 (2021), 2107.01293.
- [223] M. Buza, Y. Matiounine, J. Smith, and W. L. van Neerven, Eur. Phys. J. C **1**, 301 (1998), hep-ph/9612398.
- [224] I. Bierenbaum, J. Blumlein, and S. Klein, Nucl. Phys. B **820**, 417 (2009), 0904.3563.
- [225] J. Ablinger, J. Blumlein, S. Klein, C. Schneider, and F. Wissbrock, Nucl. Phys. B **844**, 26 (2011), 1008.3347.
- [226] J. Ablinger, J. Blümlein, A. De Freitas, A. Hasselhuhn, A. von Manteuffel, M. Round, and C. Schneider, Nucl. Phys. B **885**, 280 (2014), 1405.4259.
- [227] A. Behring, I. Bierenbaum, J. Blümlein, A. De Freitas, S. Klein, and F. Wißbrock, Eur. Phys. J. C **74**, 3033 (2014), 1403.6356.

- [228] J. Blümlein, J. Ablinger, A. Behring, A. De Freitas, A. von Manteuffel, C. Schneider, and C. Schneider, PoS **QCDEV2017**, 031 (2017), 1711.07957.
- [229] J. Ablinger et al., Nucl. Phys. B **882**, 263 (2014), 1402.0359.
- [230] J. Blümlein, P. Marquard, C. Schneider, and K. Schönwald (2022), 2202.03216.
- [231] A. Vogt, Comput. Phys. Commun. **170**, 65 (2005), hep-ph/0408244.
- [232] G. P. Salam and J. Rojo, Comput. Phys. Commun. **180**, 120 (2009), 0804.3755.
- [233] V. Bertone, S. Carrazza, and J. Rojo, Comput. Phys. Commun. **185**, 1647 (2014), 1310.1394.
- [234] A. Candido, F. Hekhorn, and G. Magni (2022), 2202.02338.
- [235] A. Candido, F. Hekhorn, and G. Magni, *N3PDF/eko: Poetry rework, some N3LO, some ekobox*, <https://doi.org/10.5281/zenodo.5896965> (2022).
- [236] J. Ablinger, J. Blümlein, A. De Freitas, A. Hasselhuhn, C. Schneider, and F. Wißbrock, Nucl. Phys. B **921**, 585 (2017), 1705.07030.
- [237] S. Moch, J. A. M. Vermaseren, and A. Vogt, Phys. Lett. B **606**, 123 (2005), hep-ph/0411112.
- [238] S. Moch and M. Rogal, Nucl. Phys. B **782**, 51 (2007), 0704.1740.
- [239] S. Moch, M. Rogal, and A. Vogt, Nucl. Phys. B **790**, 317 (2008), 0708.3731.
- [240] S. Moch, J. A. M. Vermaseren, and A. Vogt, Nucl. Phys. B **813**, 220 (2009), 0812.4168.
- [241] C. Anastasiou, C. Duhr, F. Dulat, F. Herzog, and B. Mistlberger, Phys. Rev. Lett. **114**, 212001 (2015), 1503.06056.
- [242] A. Banfi, F. Caola, F. A. Dreyer, P. F. Monni, G. P. Salam, G. Zanderighi, and F. Dulat, JHEP **04**, 049 (2016), 1511.02886.
- [243] F. Dulat, B. Mistlberger, and A. Pelloni, Phys. Rev. D **99**, 034004 (2019), 1810.09462.
- [244] C. Duhr, F. Dulat, and B. Mistlberger (2020), 2007.13313.
- [245] C. Duhr, F. Dulat, and B. Mistlberger (2020), 2001.07717.
- [246] C. Duhr, F. Dulat, V. Hirschi, and B. Mistlberger, JHEP **08**, 017 (2020), 2004.04752.
- [247] L.-B. Chen, H. T. Li, H.-S. Shao, and J. Wang, JHEP **03**, 072 (2020), 1912.13001.
- [248] F. A. Dreyer and A. Karlberg, Phys. Rev. Lett. **117**, 072001 (2016), 1606.00840.
- [249] F. A. Dreyer and A. Karlberg, Phys. Rev. D **98**, 114016 (2018), 1811.07906.
- [250] L. Cieri, X. Chen, T. Gehrmann, E. W. N. Glover, and A. Huss, JHEP **02**, 096 (2019), 1807.11501.
- [251] X. Chen, T. Gehrmann, E. W. N. Glover, and A. Huss, JHEP **01**, 053 (2022), 2111.02157.
- [252] S. Camarda, L. Cieri, and G. Ferrera (2021), 2111.14509.
- [253] S. Camarda, L. Cieri, and G. Ferrera, Phys. Rev. D **104**, L111503 (2021), 2103.04974.
- [254] G. Billis, B. Dehnadi, M. A. Ebert, J. K. L. Michel, and F. J. Tackmann, Phys. Rev. Lett. **127**, 072001 (2021), 2102.08039.
- [255] L. A. Harland-Lang and R. S. Thorne, Eur. Phys. J. C **79**, 225 (2019), 1811.08434.
- [256] R. Abdul Khalek et al. (NNPDF), Eur. Phys. J. C, 79:838 (2019), 1905.04311.
- [257] R. Abdul Khalek et al. (NNPDF), Eur. Phys. J. C **79**, 931 (2019), 1906.10698.
- [258] A. Bierweiler, T. Kasprzik, J. H. Kühn, and S. Uccirati, JHEP **11**, 093 (2012), 1208.3147.
- [259] A. Denner, S. Dittmaier, S. Kallweit, and A. Muck, JHEP **03**, 075 (2012), 1112.5142.
- [260] L. A. Harland-Lang, V. A. Khoze, and M. G. Ryskin, Eur. Phys. J. **C76**, 255 (2016), 1601.03772.
- [261] L. A. Harland-Lang, M. Tasevsky, V. A. Khoze, and M. G. Ryskin, Eur. Phys. J. C **80**, 925 (2020), 2007.12704.
- [262] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, Eur. Phys. J. **C39**, 155 (2005), hep-ph/0411040.
- [263] C. Schmidt, J. Pumplin, D. Stump, and C. P. Yuan, Phys. Rev. D **93**, 114015 (2016), 1509.02905.
- [264] R. D. Ball, V. Bertone, S. Carrazza, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, and J. Rojo (NNPDF), Nucl. Phys. **B877**, 290 (2013), 1308.0598.
- [265] F. Giuliani et al. (xFitter Developers' Team), Eur. Phys. J. **C77**, 400 (2017), 1701.08553.
- [266] A. D. Martin and M. G. Ryskin, Eur. Phys. J. **C74**, 3040 (2014), 1406.2118.
- [267] L. A. Harland-Lang, V. A. Khoze, and M. G. Ryskin, Phys. Rev. **D94**, 074008 (2016), 1607.04635.
- [268] V. M. Budnev, I. F. Ginzburg, G. V. Meledin, and V. G. Serbo, Phys. Rept. **15**, 181 (1975).
- [269] H. Anlauf, H. D. Dahmen, P. Manakos, T. Mannel, and T. Ohl, Comput. Phys. Commun. **70**, 97 (1992).
- [270] J. Blümlein, G. Levman, and H. Spiesberger, J. Phys. **G19**, 1695 (1993).
- [271] A. Mukherjee and C. Pisano, Eur. Phys. J. **C30**, 477 (2003), hep-ph/0306275.
- [272] M. Luszczak, W. Schäfer, and A. Szczurek, Phys. Rev. **D93**, 074018 (2016), 1510.00294.
- [273] A. V. Manohar, P. Nason, G. P. Salam, and G. Zanderighi, JHEP20 **12**, 046 (2017), 1708.01256.
- [274] A. Manohar, P. Nason, G. P. Salam, and G. Zanderighi, Phys. Rev. Lett. **117**, 242002 (2016), 1607.04266.
- [275] D. de Florian, G. F. R. Sborlini, and G. Rodrigo, Eur. Phys. J. **C76**, 282 (2016), 1512.00612.
- [276] D. de Florian, G. F. R. Sborlini, and G. Rodrigo, JHEP **10**, 056 (2016), 1606.02887.
- [277] V. Bertone, S. Carrazza, N. P. Hartland, and J. Rojo (NNPDF), SciPost Phys. **5**, 008 (2018), 1712.07053.
- [278] L. A. Harland-Lang, A. D. Martin, R. Nathvani, and R. S. Thorne, Eur. Phys. J. **C79**, 811 (2019), 1907.02750.
- [279] K. Xie, T. J. Hobbs, T.-J. Hou, C. Schmidt, M. Yan, and C. P. Yuan (CTEQ-TEA) (2021), 2106.10299.
- [280] T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne (2021), 2111.05357.
- [281] L. A. Harland-Lang, V. A. Khoze, and M. G. Ryskin, Phys. Lett. B **761**, 20 (2016), 1605.04935.
- [282] L. A. Harland-Lang, Phys. Rev. D **104**, 073002 (2021), 2101.04127.
- [283] S. Carrazza, E. R. Nocera, C. Schwan, and M. Zaro, JHEP **12**, 108 (2020), 2008.12789.
- [284] S. Dittmaier, A. Huss, and C. Speckner, JHEP **11**, 095 (2012), 1210.0438.
- [285] R. Frederix, S. Frixione, V. Hirschi, D. Pagani, H.-S. Shao, and M. Zaro, JHEP **04**, 076 (2017), 1612.06548.
- [286] A. Denner, S. Dittmaier, T. Kasprzik, and A. Muck, JHEP **06**, 069 (2011), 1103.0914.
- [287] D. Bardin, S. Bondarenko, P. Christova, L. Kalinovskaya, L. Rymyantsev, A. Saproinov, and W. von Schlippe, JETP

- Lett. **96**, 285 (2012), 1207.4400.
- [288] S. G. Bondarenko and A. A. Sapronov, *Comput. Phys. Commun.* **184**, 2343 (2013), 1301.3687.
- [289] Y. Li and F. Petriello, *Phys. Rev. D* **86**, 094034 (2012), 1208.5967.
- [290] R. Frederix, S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro, *JHEP* **07**, 185 (2018), [Erratum: *JHEP* **11**, 085 (2021)], 1804.10017.
- [291] M. Czakon, D. Heymes, A. Mitov, D. Pagani, I. Tsinikos, and M. Zaro, *JHEP* **10**, 186 (2017), 1705.04105.
- [292] D. Pagani, I. Tsinikos, and M. Zaro, *Eur. Phys. J. C* **76**, 479 (2016), 1606.01915.
- [293] G. P. Zeller et al. (NuTeV), *Phys. Rev. Lett.* **88**, 091802 (2002), [Erratum: *Phys.Rev.Lett.* **90**, 239902 (2003)], hep-ex/0110059.
- [294] V. Bertone, S. Carrazza, D. Pagani, and M. Zaro, *JHEP* **11**, 194 (2015), 1508.07002.
- [295] C. W. Bauer, N. Ferland, and B. R. Webber, *JHEP* **08**, 036 (2017), 1703.08562.
- [296] B. Fornal, A. V. Manohar, and W. J. Waalewijn, *JHEP* **05**, 106 (2018), 1803.06347.
- [297] C. Cocuzza, C. E. Keppel, H. Liu, W. Melnitchouk, A. Metz, N. Sato, and A. W. Thomas (Jefferson Lab Angular Momentum (JAM)), *Phys. Rev. Lett.* **127**, 242001 (2021), 2104.06946.
- [298] M. Bonvini, S. Marzani, J. Rojo, L. Rottoli, M. Ubiali, R. D. Ball, V. Bertone, S. Carrazza, and N. P. Hartland, *JHEP* **09**, 191 (2015), 1507.01006.
- [299] G. F. Sterman, *Nucl. Phys. B* **281**, 310 (1987).
- [300] R. D. Ball et al., *Nucl. Phys. B* **867**, 244 (2013), 1207.1303.
- [301] Z. Nagy and D. E. Soper, *JHEP* **10**, 019 (2016), 1605.05845.
- [302] Z. Nagy and D. E. Soper, *Phys. Rev. D* **98**, 014035 (2018), 1711.02369.
- [303] Z. Nagy and D. E. Soper, *Phys. Rev. D* **102**, 014025 (2020), 2002.04125.
- [304] A. Candido, S. Forte, and F. Hekhorn, *JHEP* **11**, 129 (2020), 2006.07377.
- [305] J. Collins, T. C. Rogers, and N. Sato (2021), 2111.01170.
- [306] R. D. Ball, E. R. Nocera, and J. Rojo, *Eur. Phys. J. C* **76**, 383 (2016), 1604.00024.
- [307] A. Courtoy and P. M. Nadolsky, *Phys. Rev. D* **103**, 054029 (2021), 2011.10078.
- [308] I. Abt et al. (ZEUS), *Phys. Rev. D* **101**, 112009 (2020), 2003.08742.
- [309] A. Courtoy and P. M. Nadolsky, in *28th International Workshop on Deep Inelastic Scattering and Related Subjects* (2021), 2108.04122.
- [310] E. A. Kuraev, L. N. Lipatov, and V. S. Fadin, *Sov. Phys. JETP* **44**, 443 (1976).
- [311] E. Kuraev, L. Lipatov, and V. S. Fadin, *Sov. Phys. JETP* **45**, 199 (1977).
- [312] I. Balitsky and L. Lipatov, *Sov. J. Nucl. Phys.* **28**, 822 (1978).
- [313] V. S. Fadin and L. N. Lipatov, *Nucl. Phys. B* **477**, 767 (1996), hep-ph/9602287.
- [314] V. S. Fadin, R. Fiore, A. Flachi, and M. I. Kotsky, *Phys. Lett. B* **422**, 287 (1998), hep-ph/9711427.
- [315] V. S. Fadin, M. I. Kotsky, and L. N. Lipatov, *Phys. Lett. B* **415**, 97 (1997).
- [316] V. S. Fadin and L. N. Lipatov, *Phys. Lett. B* **429**, 127 (1998), hep-ph/9802290.
- [317] G. Altarelli, R. D. Ball, and S. Forte, *Nucl. Phys. B* **575**, 313 (2000), hep-ph/9911273.
- [318] G. Altarelli, R. D. Ball, and S. Forte, *Nucl. Phys. B* **599**, 383 (2001), hep-ph/0011270.
- [319] G. Altarelli, R. D. Ball, and S. Forte, *Nucl. Phys. B* **621**, 359 (2002), hep-ph/0109178.
- [320] G. Altarelli, R. D. Ball, and S. Forte, *Nucl. Phys. B* **674**, 459 (2003), hep-ph/0306156.
- [321] G. Altarelli, R. D. Ball, and S. Forte, *Nucl. Phys. B* **742**, 1 (2006), hep-ph/0512237.
- [322] G. Altarelli, R. D. Ball, and S. Forte, *Nucl. Phys. B* **799**, 199 (2008), 0802.0032.
- [323] G. P. Salam, *JHEP* **07**, 019 (1998), hep-ph/9806482.
- [324] M. Ciafaloni, D. Colferai, and G. P. Salam, *Phys. Rev. D* **60**, 114036 (1999), hep-ph/9905566.
- [325] M. Ciafaloni, D. Colferai, and G. P. Salam, *JHEP* **10**, 017 (1999), hep-ph/9907409.
- [326] M. Ciafaloni, D. Colferai, and G. P. Salam, *JHEP* **07**, 054 (2000), hep-ph/0007240.
- [327] M. Ciafaloni, D. Colferai, G. P. Salam, and A. M. Stasto, *Phys. Rev. D* **68**, 114003 (2003), hep-ph/0307188.
- [328] M. Ciafaloni and D. Colferai, *JHEP* **09**, 069 (2005), hep-ph/0507106.
- [329] M. Ciafaloni, D. Colferai, G. P. Salam, and A. M. Stasto, *JHEP* **08**, 046 (2007), 0707.1453.
- [330] M. Bonvini, S. Marzani, and T. Peraro, *Eur. Phys. J. C* **76**, 597 (2016), 1607.02153.
- [331] M. Bonvini, S. Marzani, and C. Muselli, *JHEP* **12**, 117 (2017), 1708.07510.
- [332] H. Abdolmaleki et al. (xFitter Developers' Team), *Eur. Phys. J. C* **78**, 621 (2018), 1802.00064.
- [333] R. D. Ball and R. K. Ellis, *JHEP* **05**, 053 (2001), hep-ph/0101199.
- [334] R. D. Ball, *Nucl. Phys. B* **796**, 137 (2008), 0708.1277.
- [335] S. Marzani and R. D. Ball, *Nucl. Phys. B* **814**, 246 (2009), 0812.3602.
- [336] G. Diana, J. Rojo, and R. D. Ball, *Phys. Lett. B* **693**, 430 (2010), 1006.4250.
- [337] S. Marzani, F. Caola, and S. Forte, in *19th International Workshop on Deep-Inelastic Scattering and Related Subjects* (2011), 1106.6297.
- [338] R. D. Ball, M. Bonvini, S. Forte, S. Marzani, and G. Ridolfi, *Nucl. Phys. B* **874**, 746 (2013), 1303.3590.
- [339] S. Marzani, *Phys. Rev. D* **93**, 054047 (2016), 1511.06039.
- [340] S. Forte and C. Muselli, *JHEP* **03**, 122 (2016), 1511.05561.
- [341] C. Muselli, M. Bonvini, S. Forte, S. Marzani, and G. Ridolfi, *JHEP* **08**, 076 (2015), 1505.02006.
- [342] M. Bonvini, *Eur. Phys. J. C* **78**, 834 (2018), 1805.08785.
- [343] A. H. Mueller, in *Cargese Summer School on QCD Perspectives on Hot and Dense Matter* (2001), pp. 45–72, hep-ph/0111244.



- [344] K. J. Golec-Biernat and M. Wusthoff, Phys. Rev. D **59**, 014017 (1998), hep-ph/9807513.
- [345] M. Guzzi et al., in *28th International Workshop on Deep Inelastic Scattering and Related Subjects* (2021), 2108.06596.
- [346] C. Royon et al. (2022).
- [347] S. Forte, E. Laenen, P. Nason, and J. Rojo, Nucl. Phys. B **834**, 116 (2010), 1001.2312.
- [348] J. Blümlein and A. Vogt, Phys. Lett. B **370**, 149 (1996), hep-ph/9510410.
- [349] J. Blümlein and A. Vogt, Phys. Rev. D **58**, 014020 (1998), hep-ph/9712546.
- [350] J. Blümlein and H. Böttcher, Phys. Lett. B **662**, 336 (2008), 0802.0408.
- [351] J. Blümlein and H. Böttcher, in *20th International Workshop on Deep-Inelastic Scattering and Related Subjects* (2012), pp. 237–241, 1207.3170.
- [352] S. Alekhin, J. Blümlein, and S. Moch, Phys. Rev. D **86**, 054009 (2012), 1202.2281.
- [353] R. Brock et al. (CTEQ) (1994).
- [354] E. G. de Oliveira, A. D. Martin, and M. G. Ryskin, JHEP **02**, 060 (2013), 1206.2223.
- [355] E. G. de Oliveira, A. D. Martin, and M. G. Ryskin, JHEP **11**, 156 (2013), 1310.8289.
- [356] S. Jadach, A. Kusina, W. Placzek, M. Skrzypek, and M. Sławińska, Phys. Rev. D **87**, 034029 (2013), 1103.5015.
- [357] S. Jadach, W. Placzek, S. Sapeta, A. Siódmok, and M. Skrzypek, JHEP **10**, 052 (2015), 1503.06849.
- [358] S. Jadach, G. Nail, W. Placzek, S. Sapeta, A. Siódmok, and M. Skrzypek, Eur. Phys. J. C **77**, 164 (2017), 1607.06799.
- [359] S. Jadach, Acta Phys. Polon. B **51**, 1363 (2020), 2004.04239.
- [360] S. Jadach, W. Placzek, S. Sapeta, A. Siódmok, and M. Skrzypek, Eur. Phys. J. C **76**, 649 (2016), 1606.00355.
- [361] S. Catani and M. H. Seymour, Nucl. Phys. B **485**, 291 (1997), [Erratum: Nucl.Phys.B 510, 503–504 (1998)], hep-ph/9605323.
- [362] L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, Eur. Phys. J. C **78**, 248 (2018), 1711.05757.
- [363] G. Aad et al. (ATLAS) (2018), ATL-PHYS-PUB-2018-017.
- [364] R. D. Ball and A. Deshpande, in *From My Vast Repertoire ...: Guido Altarelli's Legacy*, edited by A. Levy, S. Forte, and G. Ridolfi (World Scientific, 2019), pp. 205–226, 1801.04842.
- [365] R. D. Ball, E. R. Nocera, and R. L. Pearson (NNPDF), Eur. Phys. J. C **79**, 282 (2019), 1812.09074.
- [366] R. Pearson, R. Ball, and E. R. Nocera, PoS **DIS2019**, 027 (2019).
- [367] R. L. Pearson, R. D. Ball, and E. R. Nocera (2021), 2106.12349.
- [368] R. D. Ball et al. (NNPDF), Eur. Phys. J. C **81**, 958 (2021), 2109.02671.
- [369] R. Abdul Khalek et al. (NNPDF), Eur. Phys. J. C, 79:838 (2019), 1905.04311.
- [370] R. Abdul Khalek et al. (NNPDF), Eur. Phys. J. C **79**, 931 (2019), 1906.10698.
- [371] R. D. Ball and R. L. Pearson, Eur. Phys. J. C **81**, 830 (2021), 2105.05114.
- [372] S. Forte and Z. Kassabov, Eur. Phys. J. C **80**, 182 (2020), 2001.04986.
- [373] M. Cacciari and N. Houdeau, JHEP **09**, 039 (2011), 1105.5152.
- [374] E. Bagnaschi, M. Cacciari, A. Guffanti, and L. Jenniches, JHEP **02**, 133 (2015), 1409.5036.
- [375] M. Bonvini, Eur. Phys. J. C **80**, 989 (2020), 2006.16293.
- [376] C. Duhr, A. Huss, A. Mazeliauskas, and R. Szafron, JHEP **09**, 122 (2021), 2106.04585.
- [377] K. Albertsson et al., J. Phys. Conf. Ser. **1085**, 022008 (2018), 1807.02876.
- [378] S. Forte, L. Garrido, J. I. Latorre, and A. Piccione, JHEP **05**, 062 (2002), hep-ph/0204232.
- [379] L. Del Debbio, S. Forte, J. I. Latorre, A. Piccione, and J. Rojo (NNPDF), JHEP **03**, 080 (2005), hep-ph/0501067.
- [380] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, A. Piccione, J. Rojo, and M. Ubiali (NNPDF), Nucl. Phys. B **809**, 1 (2009), [Erratum: Nucl.Phys.B 816, 293 (2009)], 0808.1231.
- [381] S. Carrazza, J. M. Cruz-Martinez, and R. Stegeman (2021), 2111.02954.
- [382] S. Carrazza and J. Cruz-Martinez, Eur. Phys. J. C **79**, 676 (2019), 1907.05075.
- [383] R. A. Khalek, R. Gauld, T. Giani, E. R. Nocera, T. R. Rabemananjara, and J. Rojo (2022), 2201.12363.
- [384] V. Bertone, S. Carrazza, N. P. Hartland, E. R. Nocera, and J. Rojo (NNPDF), Eur. Phys. J. C **77**, 516 (2017), 1706.07049.
- [385] S. Carrazza, J. I. Latorre, J. Rojo, and G. Watt, Eur. Phys. J. C **75**, 474 (2015), 1504.06469.
- [386] S. Carrazza, J. M. Cruz-Martinez, and T. R. Rabemananjara, Eur. Phys. J. C **81**, 530 (2021), 2104.04535.
- [387] A. Pérez-Salinas, J. Cruz-Martinez, A. A. Alhajri, and S. Carrazza, Phys. Rev. D **103**, 034027 (2021), 2011.13934.
- [388] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung, Phys. Rev. D **65**, 014013 (2001), hep-ph/0101032.
- [389] W. T. Giele, S. A. Keller, and D. A. Kosower (2001), hep-ph/0104052.
- [390] J. Gao and P. Nadolsky, JHEP **07**, 035 (2014), 1401.0013.
- [391] S. Carrazza, S. Forte, Z. Kassabov, J. I. Latorre, and J. Rojo, Eur. Phys. J. C **75**, 369 (2015), 1505.06736.
- [392] S. Carrazza, S. Forte, Z. Kassabov, and J. Rojo, Eur. Phys. J. C **76**, 205 (2016), 1602.00005.
- [393] J. Pumplin, Phys. Rev. D **80**, 034002 (2009), 0904.2425.
- [394] S. Dulat, T.-J. Hou, J. Gao, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan, Phys. Rev. D **89**, 113002 (2014), 1310.7601.
- [395] R. Barlow, in *Conference on Advanced Statistical Techniques in Particle Physics* (2002), pp. 134–144, hep-ex/0207026.
- [396] A. Apyan and D. Froidevaux, *PDF benchmarking proposal for precision Drell-Yan*, PDF4LHC meeting, CERN, <https://tinyurl.com/4wcnd8xn> (2018).
- [397] D. Froidevaux, *PDF benchmarking discussion*, LHC EW Precision sub-group workshop, IPPP Durham UK, <https://tinyurl.com/2p8d8ba3> (2019).
- [398] B. Malaescu, *PDF benchmarking report*, LHC Electroweak WG meeting, CERN, <https://tinyurl.com/2p8tcn5b> (2019).
- [399] P. Belov et al. (HERAFitter developers' Team), Eur. Phys. J. C **74**, 3039 (2014), 1404.4234.

- [400] R. D. Ball, S. Forte, and R. Stegeman, *Eur. Phys. J. C* **81**, 1046 (2021), 2110.08274.
- [401] D. d'Enterria et al., in *2022 Snowmass Summer Study* (2022), 2203.08271.
- [402] J. H. Weber, A. Bazavov, and P. Petreczky, *PoS Confinement2018*, 166 (2019), 1811.12902.
- [403] M. Dalla Brida, *Eur. Phys. J. A* **57**, 66 (2021), 2012.01232.
- [404] J. Komijani, P. Petreczky, and J. H. Weber, *Prog. Part. Nucl. Phys.* **113**, 103788 (2020), 2003.11703.
- [405] S. Aoki et al., *Eur. Phys. J. C* **74**, 2890 (2014), 1310.8555.
- [406] S. Aoki et al., *Eur. Phys. J. C* **77**, 112 (2017), 1607.00299.
- [407] S. Aoki et al. (Flavour Lattice Averaging Group), *Eur. Phys. J. C* **80**, 113 (2020), 1902.08191.
- [408] Y. Aoki et al. (2021), 2111.09849.
- [409] M. Luscher, P. Weisz, and U. Wolff, *Nucl. Phys. B* **359**, 221 (1991).
- [410] M. Luscher, R. Sommer, P. Weisz, and U. Wolff, *Nucl. Phys. B* **413**, 481 (1994), hep-lat/9309005.
- [411] G. de Divitiis, R. Frezzotti, M. Guagnelli, M. Luscher, R. Petronzio, R. Sommer, P. Weisz, and U. Wolff (Alpha), *Nucl. Phys. B* **437**, 447 (1995), hep-lat/9411017.
- [412] K. Jansen, C. Liu, M. Luscher, H. Simma, S. Sint, R. Sommer, P. Weisz, and U. Wolff, *Phys. Lett. B* **372**, 275 (1996), hep-lat/9512009.
- [413] K. Symanzik, *Nucl. Phys. B* **190**, 1 (1981).
- [414] M. Luscher, *Nucl. Phys. B* **254**, 52 (1985).
- [415] M. Luscher, R. Narayanan, P. Weisz, and U. Wolff, *Nucl. Phys. B* **384**, 168 (1992), hep-lat/9207009.
- [416] S. Sint, *Nucl. Phys. B* **421**, 135 (1994), hep-lat/9312079.
- [417] A. Bode, P. Weisz, and U. Wolff (ALPHA), *Nucl. Phys. B* **576**, 517 (2000), [Erratum: *Nucl.Phys.B* 608, 481–481 (2001), Erratum: *Nucl.Phys.B* 600, 453–453 (2001)], hep-lat/9911018.
- [418] M. Bruno, M. Dalla Brida, P. Fritzsch, T. Korzec, A. Ramos, S. Schaefer, H. Simma, S. Sint, and R. Sommer (ALPHA), *Phys. Rev. Lett.* **119**, 102001 (2017), 1706.03821.
- [419] Q. Mason, H. D. Trottier, C. T. H. Davies, K. Foley, A. Gray, G. P. Lepage, M. Nobes, and J. Shigemitsu (HPQCD, UKQCD), *Phys. Rev. Lett.* **95**, 052002 (2005), hep-lat/0503005.
- [420] C. T. H. Davies, K. Hornbostel, I. D. Kendall, G. P. Lepage, C. McNeile, J. Shigemitsu, and H. Trottier (HPQCD), *Phys. Rev. D* **78**, 114507 (2008), 0807.1687.
- [421] K. Maltman, D. Leinweber, P. Moran, and A. Sternbeck, *Phys. Rev. D* **78**, 114504 (2008), 0807.2020.
- [422] C. McNeile, C. T. H. Davies, E. Follana, K. Hornbostel, and G. P. Lepage, *Phys. Rev. D* **82**, 034512 (2010), 1004.4285.
- [423] A. Bazavov, N. Brambilla, P. Petreczky, A. Vairo, and J. H. Weber (TUMQCD), *Phys. Rev. D* **98**, 054511 (2018), 1804.10600.
- [424] A. Bazavov, N. Brambilla, X. G. Tormo, I. P. Petreczky, J. Soto, and A. Vairo, *Phys. Rev. D* **90**, 074038 (2014), [Erratum: *Phys.Rev.D* 101, 119902 (2020)], 1407.8437.
- [425] T. Appelquist, M. Dine, and I. J. Muzinich, *Phys. Lett. B* **69**, 231 (1977).
- [426] W. Fischler, *Nucl. Phys. B* **129**, 157 (1977).
- [427] A. Billoire, *Phys. Lett. B* **92**, 343 (1980).
- [428] Y. Schroder, *Phys. Lett. B* **447**, 321 (1999), hep-ph/9812205.
- [429] N. Brambilla, A. Pineda, J. Soto, and A. Vairo, *Phys. Rev. D* **60**, 091502 (1999), hep-ph/9903355.
- [430] N. Brambilla, X. Garcia i Tormo, J. Soto, and A. Vairo, *Phys. Lett. B* **647**, 185 (2007), hep-ph/0610143.
- [431] N. Brambilla, A. Vairo, X. Garcia i Tormo, and J. Soto, *Phys. Rev. D* **80**, 034016 (2009), 0906.1390.
- [432] C. Anzai, Y. Kiyo, and Y. Sumino, *Phys. Rev. Lett.* **104**, 112003 (2010), 0911.4335.
- [433] A. V. Smirnov, V. A. Smirnov, and M. Steinhauser, *Phys. Rev. Lett.* **104**, 112002 (2010), 0911.4742.
- [434] R. N. Lee, A. V. Smirnov, V. A. Smirnov, and M. Steinhauser, *Phys. Rev. D* **94**, 054029 (2016), 1608.02603.
- [435] R. N. Lee and V. A. Smirnov, *JHEP* **10**, 089 (2016), 1608.02605.
- [436] A. Bazavov, N. Brambilla, X. Garcia i Tormo, P. Petreczky, J. Soto, A. Vairo, and J. H. Weber (TUMQCD), *Phys. Rev. D* **100**, 114511 (2019), 1907.11747.
- [437] A. Pineda and J. Soto, *Phys. Lett. B* **495**, 323 (2000), hep-ph/0007197.
- [438] N. Brambilla, A. Pineda, J. Soto, and A. Vairo, *Rev. Mod. Phys.* **77**, 1423 (2005), hep-ph/0410047.
- [439] H. Takaura, T. Kaneko, Y. Kiyo, and Y. Sumino, *Phys. Lett. B* **789**, 598 (2019), 1808.01632.
- [440] H. Takaura, T. Kaneko, Y. Kiyo, and Y. Sumino, *JHEP* **04**, 155 (2019), 1808.01643.
- [441] C. Ayala, X. Lobregat, and A. Pineda, *JHEP* **09**, 016 (2020), 2005.12301.
- [442] B. Dehnadi, A. H. Hoang, and V. Mateu, *JHEP* **08**, 155 (2015), 1504.07638.
- [443] D. Boito and V. Mateu, *JHEP* **03**, 094 (2020), 2001.11041.
- [444] C. Sturm, *JHEP* **09**, 075 (2008), 0805.3358.
- [445] Y. Kiyo, A. Maier, P. Maierhofer, and P. Marquard, *Nucl. Phys. B* **823**, 269 (2009), 0907.2120.
- [446] A. Maier, P. Maierhofer, P. Marquard, and A. V. Smirnov, *Nucl. Phys. B* **824**, 1 (2010), 0907.2117.
- [447] I. Allison et al. (HPQCD), *Phys. Rev. D* **78**, 054513 (2008), 0805.2999.
- [448] B. Chakraborty, C. T. H. Davies, B. Galloway, P. Knecht, J. Koponen, G. C. Donald, R. J. Dowdall, G. P. Lepage, and C. McNeile, *Phys. Rev. D* **91**, 054508 (2015), 1408.4169.
- [449] Y. Maezawa and P. Petreczky, *Phys. Rev. D* **94**, 034507 (2016), 1606.08798.
- [450] P. Petreczky and J. H. Weber, *Phys. Rev. D* **100**, 034519 (2019), 1901.06424.
- [451] P. Petreczky and J. H. Weber, *Eur. Phys. J. C* **82**, 64 (2022), 2012.06193.
- [452] K. Nakayama, B. Fahy, and S. Hashimoto, *Phys. Rev. D* **94**, 054507 (2016), 1606.01002.
- [453] K. G. Chetyrkin, A. L. Kataev, and F. V. Tkachov, *Phys. Lett. B* **85**, 277 (1979).

- [454] L. R. Surguladze and M. A. Samuel, Phys. Rev. Lett. **66**, 560 (1991), [Erratum: Phys.Rev.Lett. 66, 2416 (1991)].
- [455] S. G. Gorishnii, A. L. Kataev, and S. A. Larin, Phys. Lett. B **259**, 144 (1991).
- [456] P. A. Baikov, K. G. Chetyrkin, and J. H. Kuhn, Phys. Rev. Lett. **101**, 012002 (2008), 0801.1821.
- [457] E. Shintani, S. Aoki, T. W. Chiu, S. Hashimoto, T. H. Hsieh, T. Kaneko, H. Matsufuru, J. Noaki, T. Onogi, and N. Yamada (JLQCD, TWQCD), Phys. Rev. D **79**, 074510 (2009), 0807.0556.
- [458] E. Shintani, S. Aoki, H. Fukaya, S. Hashimoto, T. Kaneko, T. Onogi, and N. Yamada, Phys. Rev. D **82**, 074505 (2010), [Erratum: Phys.Rev.D 89, 099903 (2014)], 1002.0371.
- [459] R. J. Hudspith, R. Lewis, K. Maltman, and E. Shintani (2018), 1804.10286.
- [460] S. Cali, K. Cichy, P. Korcyl, and J. Simeth, Phys. Rev. Lett. **125**, 242002 (2020), 2003.05781.
- [461] B. Alles, D. Henty, H. Panagopoulos, C. Parrinello, C. Pittori, and D. G. Richards, Nucl. Phys. B **502**, 325 (1997), hep-lat/9605033.
- [462] P. Boucaud, J. P. Leroy, H. Moutarde, J. Micheli, O. Pene, J. Rodriguez-Quintero, and C. Roiesnel, JHEP **01**, 046 (2002), hep-ph/0107278.
- [463] P. Boucaud, F. De Soto, J. P. Leroy, A. Le Yaouanc, J. Micheli, O. Pene, and J. Rodriguez-Quintero, Phys. Rev. D **79**, 014508 (2009), 0811.2059.
- [464] B. Blossier, P. Boucaud, F. De Soto, V. Morenas, M. Gravina, O. Pene, and J. Rodriguez-Quintero (ETM), Phys. Rev. D **82**, 034510 (2010), 1005.5290.
- [465] K. G. Chetyrkin and A. Retey (2000), hep-ph/0007088.
- [466] B. Blossier, P. Boucaud, M. Brinet, F. De Soto, X. Du, M. Gravina, V. Morenas, O. Pene, K. Petrov, and J. Rodriguez-Quintero, Phys. Rev. D **85**, 034503 (2012), 1110.5829.
- [467] B. Blossier, P. Boucaud, M. Brinet, F. De Soto, X. Du, V. Morenas, O. Pene, K. Petrov, and J. Rodriguez-Quintero, Phys. Rev. Lett. **108**, 262002 (2012), 1201.5770.
- [468] B. Blossier, P. Boucaud, M. Brinet, F. De Soto, V. Morenas, O. Pene, K. Petrov, and J. Rodriguez-Quintero (ETM), Phys. Rev. D **89**, 014507 (2014), 1310.3763.
- [469] S. Zafeiropoulos, P. Boucaud, F. De Soto, J. Rodríguez-Quintero, and J. Segovia, Phys. Rev. Lett. **122**, 162002 (2019), 1902.08148.
- [470] M. Dalla Brida, R. Höllwieser, F. Knechtli, T. Korzec, A. Ramos, and R. Sommer (ALPHA), Phys. Lett. B **807**, 135571 (2020), 1912.06001.
- [471] K. G. Chetyrkin and J. H. Kuhn, Nucl. Phys. B **432**, 337 (1994), hep-ph/9406299.
- [472] J.-L. Kneur and A. Neveu, Phys. Rev. D **92**, 074027 (2015), 1506.07506.
- [473] K. Nakayama, H. Fukaya, and S. Hashimoto, Phys. Rev. D **98**, 014501 (2018), 1804.06695.
- [474] X. Ji, Phys. Rev. Lett. **110**, 262002 (2013), 1305.1539.
- [475] X. Ji, Sci. China Phys. Mech. Astron. **57**, 1407 (2014), 1404.6680.
- [476] Y.-Q. Ma and J.-W. Qiu, Phys. Rev. Lett. **120**, 022003 (2018), 1709.03018.
- [477] T. Izubuchi, X. Ji, L. Jin, I. W. Stewart, and Y. Zhao, Phys. Rev. D **98**, 056004 (2018), 1801.03917.
- [478] Y.-S. Liu, W. Wang, J. Xu, Q.-A. Zhang, J.-H. Zhang, S. Zhao, and Y. Zhao, Phys. Rev. D **100**, 034006 (2019), 1902.00307.
- [479] H.-W. Lin, J.-W. Chen, S. D. Cohen, and X. Ji, Phys. Rev. D **91**, 054510 (2015), 1402.1462.
- [480] U. Aglietti, M. Ciuchini, G. Corbo, E. Franco, G. Martinelli, and L. Silvestrini, Phys. Lett. B **441**, 371 (1998), hep-ph/9806277.
- [481] G. Martinelli, Nucl. Phys. B Proc. Suppl. **73**, 58 (1999), hep-lat/9810013.
- [482] C. Dawson, G. Martinelli, G. C. Rossi, C. T. Sachrajda, S. R. Sharpe, M. Talevi, and M. Testa, Nucl. Phys. B **514**, 313 (1998), hep-lat/9707009.
- [483] S. Capitani, M. Gockeler, R. Horsley, H. Oelrich, D. Petters, P. E. L. Rakow, and G. Schierholz, Nucl. Phys. B Proc. Suppl. **73**, 288 (1999), hep-lat/9809171.
- [484] S. Capitani, M. Gockeler, R. Horsley, D. Petters, D. Pleiter, P. E. L. Rakow, and G. Schierholz, Nucl. Phys. B Proc. Suppl. **79**, 173 (1999), hep-ph/9906320.
- [485] A. J. Chambers, R. Horsley, Y. Nakamura, H. Perlt, P. E. L. Rakow, G. Schierholz, A. Schiller, K. Somfleth, R. D. Young, and J. M. Zanotti, Phys. Rev. Lett. **118**, 242001 (2017), 1703.01153.
- [486] A. Hannaford-Gunn, R. Horsley, Y. Nakamura, H. Perlt, P. E. L. Rakow, G. Schierholz, K. Somfleth, H. Stüben, R. D. Young, and J. M. Zanotti (QCDSF-UKQCD-CSSM), PoS **LATTICE2019**, 278 (2020), 2001.05090.
- [487] R. Horsley, Y. Nakamura, H. Perlt, P. E. L. Rakow, G. Schierholz, K. Somfleth, R. D. Young, and J. M. Zanotti (QCDSF-UKQCD-CSSM), PoS **LATTICE2019**, 137 (2020), 2001.05366.
- [488] W. Detmold and C. J. D. Lin, Phys. Rev. D **73**, 014501 (2006), hep-lat/0507007.
- [489] W. Detmold, I. Kanamori, C. J. D. Lin, S. Mondal, and Y. Zhao, PoS **LATTICE2018**, 106 (2018), 1810.12194.
- [490] W. Detmold, A. V. Grebe, I. Kanamori, C. J. D. Lin, S. Mondal, R. J. Perry, and Y. Zhao, in *Asia-Pacific Symposium for Lattice Field Theory* (2020), 2009.09473.
- [491] V. Braun and D. Müller, Eur. Phys. J. C **55**, 349 (2008), 0709.1348.
- [492] K.-F. Liu and S.-J. Dong, Phys. Rev. Lett. **72**, 1790 (1994), hep-ph/9306299.
- [493] K. F. Liu, S. J. Dong, T. Draper, D. Leinweber, J. H. Sloan, W. Wilcox, and R. M. Woloshyn, Phys. Rev. D **59**, 112001 (1999), hep-ph/9806491.
- [494] K.-F. Liu, Phys. Rev. D **62**, 074501 (2000), hep-ph/9910306.
- [495] K.-F. Liu, PoS **LATTICE2015**, 115 (2016), 1603.07352.
- [496] K.-F. Liu, Phys. Rev. D **96**, 033001 (2017), 1703.04690.
- [497] K.-F. Liu, Phys. Rev. D **102**, 074502 (2020), 2007.15075.

- [498] G. S. Bali et al., Eur. Phys. J. C **78**, 217 (2018), 1709.04325.
- [499] G. S. Bali, V. M. Braun, B. Gläfle, M. Göckeler, M. Gruber, F. Hutzler, P. Korcyl, A. Schäfer, P. Wein, and J.-H. Zhang, Phys. Rev. D **98**, 094507 (2018), 1807.06671.
- [500] R. S. Sufian, J. Karpie, C. Egerer, K. Orginos, J.-W. Qiu, and D. G. Richards, Phys. Rev. D **99**, 074507 (2019), 1901.03921.
- [501] R. S. Sufian, C. Egerer, J. Karpie, R. G. Edwards, B. Joó, Y.-Q. Ma, K. Orginos, J.-W. Qiu, and D. G. Richards, Phys. Rev. D **102**, 054508 (2020), 2001.04960.
- [502] A. V. Radyushkin, Phys. Rev. D **96**, 034025 (2017), 1705.01488.
- [503] K. Cichy and M. Constantinou, Adv. High Energy Phys. **2019**, 3036904 (2019), 1811.07248.
- [504] Y. Zhao, PoS **LATTICE2019**, 267 (2020).
- [505] X. Ji, Y.-S. Liu, Y. Liu, J.-H. Zhang, and Y. Zhao, Rev. Mod. Phys. **93**, 035005 (2021), 2004.03543.
- [506] X. Ji (2020), 2007.06613.
- [507] H.-W. Lin, J.-W. Chen, T. Ishikawa, and J.-H. Zhang (LP3), Phys. Rev. D **98**, 054504 (2018), 1708.05301.
- [508] C. Alexandrou, K. Cichy, M. Constantinou, K. Jansen, A. Scapellato, and F. Steffens, Phys. Rev. Lett. **121**, 112001 (2018), 1803.02685.
- [509] J.-W. Chen, L. Jin, H.-W. Lin, Y.-S. Liu, Y.-B. Yang, J.-H. Zhang, and Y. Zhao (2018), 1803.04393.
- [510] M. Bhat, K. Cichy, M. Constantinou, and A. Scapellato, Phys. Rev. D **103**, 034510 (2021), 2005.02102.
- [511] B. Joó, J. Karpie, K. Orginos, A. V. Radyushkin, D. G. Richards, and S. Zafeiropoulos, Phys. Rev. Lett. **125**, 232003 (2020), 2004.01687.
- [512] H.-W. Lin, J.-W. Chen, and R. Zhang (2020), 2011.14971.
- [513] H.-W. Lin, J.-W. Chen, X. Ji, L. Jin, R. Li, Y.-S. Liu, Y.-B. Yang, J.-H. Zhang, and Y. Zhao, Phys. Rev. Lett. **121**, 242003 (2018), 1807.07431.
- [514] R. Zhang, H.-W. Lin, and B. Yoon, Phys. Rev. D **104**, 094511 (2021), 2005.01124.
- [515] C. Alexandrou, M. Constantinou, K. Hadjiyiannakou, K. Jansen, and F. Manigrasso, Phys. Rev. Lett. **126**, 102003 (2021), 2009.13061.
- [516] Z.-Y. Fan, Y.-B. Yang, A. Anthony, H.-W. Lin, and K.-F. Liu, Phys. Rev. Lett. **121**, 242001 (2018), 1808.02077.
- [517] I. Balitsky, W. Morris, and A. Radyushkin, Phys. Lett. B **808**, 135621 (2020), 1910.13963.
- [518] W. Wang, J.-H. Zhang, S. Zhao, and R. Zhu, Phys. Rev. D **100**, 074509 (2019), 1904.00978.
- [519] J.-H. Zhang, X. Ji, A. Schäfer, W. Wang, and S. Zhao, Phys. Rev. Lett. **122**, 142001 (2019), 1808.10824.
- [520] Z. Fan, R. Zhang, and H.-W. Lin, Int. J. Mod. Phys. A **36**, 2150080 (2021), 2007.16113.
- [521] T. Khan et al. (HadStruc), Phys. Rev. D **104**, 094516 (2021), 2107.08960.
- [522] J.-W. Chen, S. D. Cohen, X. Ji, H.-W. Lin, and J.-H. Zhang, Nucl. Phys. B **911**, 246 (2016), 1603.06664.
- [523] Y.-S. Liu, J.-W. Chen, L. Jin, R. Li, H.-W. Lin, Y.-B. Yang, J.-H. Zhang, and Y. Zhao (2018), 1810.05043.
- [524] C. Alexandrou, K. Cichy, M. Constantinou, K. Jansen, A. Scapellato, and F. Steffens, Phys. Rev. D **98**, 091503 (2018), 1807.00232.
- [525] J.-H. Zhang, J.-W. Chen, L. Jin, H.-W. Lin, A. Schäfer, and Y. Zhao, Phys. Rev. D **100**, 034505 (2019), 1804.01483.
- [526] T. Izubuchi, L. Jin, C. Kallidonis, N. Karthik, S. Mukherjee, P. Petreczky, C. Shugert, and S. Syritsyn, Phys. Rev. D **100**, 034516 (2019), 1905.06349.
- [527] B. Joó, J. Karpie, K. Orginos, A. V. Radyushkin, D. G. Richards, R. S. Sufian, and S. Zafeiropoulos, Phys. Rev. D **100**, 114512 (2019), 1909.08517.
- [528] C. Shugert, X. Gao, T. Izubuchi, L. Jin, C. Kallidonis, N. Karthik, S. Mukherjee, P. Petreczky, S. Syritsyn, and Y. Zhao, in *37th International Symposium on Lattice Field Theory* (2020), 2001.11650.
- [529] X. Gao, L. Jin, C. Kallidonis, N. Karthik, S. Mukherjee, P. Petreczky, C. Shugert, S. Syritsyn, and Y. Zhao, Phys. Rev. D **102**, 094513 (2020), 2007.06590.
- [530] H.-W. Lin, J.-W. Chen, Z. Fan, J.-H. Zhang, and R. Zhang, Phys. Rev. D **103**, 014516 (2021), 2003.14128.
- [531] X. Gao, A. D. Hanlon, S. Mukherjee, P. Petreczky, P. Scior, S. Syritsyn, and Y. Zhao (2021), 2112.02208.
- [532] L.-B. Chen, W. Wang, and R. Zhu, Phys. Rev. Lett. **126**, 072002 (2021), 2006.14825.
- [533] Z.-Y. Li, Y.-Q. Ma, and J.-W. Qiu, Phys. Rev. Lett. **126**, 072001 (2021), 2006.12370.
- [534] J.-W. Chen, H.-W. Lin, and J.-H. Zhang, Nucl. Phys. B **952**, 114940 (2020), 1904.12376.
- [535] C. Alexandrou, K. Cichy, M. Constantinou, K. Hadjiyiannakou, K. Jansen, A. Scapellato, and F. Steffens, Phys. Rev. Lett. **125**, 262001 (2020), 2008.10573.
- [536] H.-W. Lin, Phys. Rev. Lett. **127**, 182001 (2021), 2008.12474.
- [537] H.-W. Lin, Phys. Lett. B **824**, 136821 (2022), 2112.07519.
- [538] B. U. Musch, P. Hagler, J. W. Negele, and A. Schafer, Phys. Rev. D **83**, 094507 (2011), 1011.1213.
- [539] B. U. Musch, P. Hagler, M. Engelhardt, J. W. Negele, and A. Schafer, Phys. Rev. D **85**, 094510 (2012), 1111.4249.
- [540] M. Engelhardt, P. Hägler, B. Musch, J. Negele, and A. Schäfer, Phys. Rev. D **93**, 054501 (2016), 1506.07826.
- [541] B. Yoon, T. Bhattacharya, M. Engelhardt, J. Green, R. Gupta, P. Hägler, B. Musch, J. Negele, A. Pochinsky, and S. Syritsyn, in *33rd International Symposium on Lattice Field Theory* (SISSA, 2015), 1601.05717.
- [542] B. Yoon, M. Engelhardt, R. Gupta, T. Bhattacharya, J. R. Green, B. U. Musch, J. W. Negele, A. V. Pochinsky, A. Schäfer, and S. N. Syritsyn, Phys. Rev. D **96**, 094508 (2017), 1706.03406.
- [543] P. Shanahan, M. Wagman, and Y. Zhao, Phys. Rev. D **102**, 014511 (2020), 2003.06063.
- [544] Q.-A. Zhang et al. (Lattice Parton), Phys. Rev. Lett. **125**, 192001 (2020), 2005.14572.
- [545] M. Schlemmer, A. Vladimirov, C. Zimmermann, M. Engelhardt, and A. Schäfer, JHEP **08**, 004 (2021), 2103.16991.
- [546] Y. Li et al., Phys. Rev. Lett. **128**, 062002 (2022), 2106.13027.
- [547] P. Shanahan, M. Wagman, and Y. Zhao, Phys. Rev. D **104**, 114502 (2021), 2107.11930.

- [548] M. Constantinou et al. (2022), 2202.07193.
- [549] A. Kusina et al., *Eur. Phys. J. C* **80**, 968 (2020), 2007.09100.
- [550] I. Helenius, M. Walt, and W. Vogelsang (2021), 2112.11904.
- [551] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado (2021), 2112.12462.
- [552] P. C. Barry, N. Sato, W. Melnitchouk, and C.-R. Ji, *Phys. Rev. Lett.* **121**, 152001 (2018), 1804.01965.
- [553] I. Novikov et al., *Phys. Rev. D* **102**, 014040 (2020), 2002.02902.
- [554] Z. Fan and H.-W. Lin, *Phys. Lett. B* **823**, 136778 (2021), 2104.06372.
- [555] A. Freese, I. C. Cloët, and P. C. Tandy, *Phys. Lett. B* **823**, 136719 (2021), 2103.05839.
- [556] F. Winter, W. Detmold, A. S. Gambhir, K. Orginos, M. J. Savage, P. E. Shanahan, and M. L. Wagman, *Phys. Rev. D* **96**, 094512 (2017), 1709.00395.
- [557] S. Chatrchyan et al. (CMS), *Phys. Rev. C* **84**, 024906 (2011), 1102.1957.
- [558] A. Angerami (ATLAS), *Nucl. Phys. A* **910-911**, 12 (2013), 1210.0138.
- [559] V. Khachatryan et al. (CMS), *JHEP* **09**, 091 (2010), 1009.4122.
- [560] S. Chatrchyan et al. (CMS), *Phys. Lett. B* **718**, 795 (2013), 1210.5482.
- [561] I. Arsene et al. (BRAHMS), *Nucl. Phys. A* **757**, 1 (2005), nucl-ex/0410020.
- [562] S. Chatrchyan et al. (CMS), *Phys. Lett. B* **724**, 213 (2013), 1305.0609.
- [563] J. Adams et al. (STAR), *Phys. Rev. Lett.* **97**, 152302 (2006), nucl-ex/0602011.
- [564] J. Adam et al. (ALICE), *JHEP* **06**, 055 (2015), 1503.07179.
- [565] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, *Eur. Phys. J. C* **76**, 186 (2016), 1601.03413.
- [566] E. P. Segarra et al., *Phys. Rev. D* **103**, 114015 (2021), 2012.11566.
- [567] J. Collins, *Foundations of perturbative QCD*, vol. 32 (Cambridge University Press, 2013), ISBN 978-1-107-64525-7, 978-1-107-64525-7, 978-0-521-85533-4, 978-1-139-09782-6.
- [568] I. Adachi et al. (2022), 2203.07622.
- [569] P. J. Mulders and R. D. Tangerman, *Nucl. Phys. B* **461**, 197 (1996), [Erratum: *Nucl.Phys.B* 484, 538–540 (1997)], hep-ph/9510301.
- [570] J. C. Collins, *Phys. Lett. B* **536**, 43 (2002), hep-ph/0204004.
- [571] F. Landry, R. Brock, P. M. Nadolsky, and C. P. Yuan, *Phys. Rev. D* **67**, 073016 (2003), hep-ph/0212159.
- [572] A. Bacchetta, F. Delcarro, C. Pisano, and M. Radici, *Phys. Lett. B* **827**, 136961 (2022), 2004.14278.
- [573] M. G. Echevarria, Z.-B. Kang, and J. Terry, *JHEP* **01**, 126 (2021), 2009.10710.
- [574] M. Bury, A. Prokudin, and A. Vladimirov, *JHEP* **05**, 151 (2021), 2103.03270.
- [575] J.-W. Qiu and G. F. Sterman, *Phys. Rev. D* **59**, 014004 (1999), hep-ph/9806356.
- [576] P. J. Mulders and J. Rodrigues, *Phys. Rev. D* **63**, 094021 (2001), hep-ph/0009343.
- [577] S. Meissner, A. Metz, and K. Goeke, *Phys. Rev. D* **76**, 034002 (2007), hep-ph/0703176.
- [578] C. Lorce' and B. Pasquini, *JHEP* **09**, 138 (2013), 1307.4497.
- [579] D. Boer, S. Cotogno, T. van Daal, P. J. Mulders, A. Signori, and Y.-J. Zhou, *JHEP* **10**, 013 (2016), 1607.01654.
- [580] G. Bozzi, S. Catani, D. de Florian, and M. Grazzini, *Phys. Lett. B* **564**, 65 (2003), hep-ph/0302104.
- [581] S. Catani and M. Grazzini, *Nucl. Phys. B* **845**, 297 (2011), 1011.3918.
- [582] M. G. Echevarria, T. Kasemets, P. J. Mulders, and C. Pisano, *JHEP* **07**, 158 (2015), [Erratum: *JHEP*05,073(2017)], 1502.05354.
- [583] D. Kharzeev, Y. V. Kovchegov, and K. Tuchin, *Phys. Rev. D* **68**, 094013 (2003), hep-ph/0307037.
- [584] F. Dominguez, C. Marquet, B.-W. Xiao, and F. Yuan, *Phys. Rev. D* **83**, 105005 (2011), 1101.0715.
- [585] T. C. Rogers, *Phys. Rev. D* **88**, 014002 (2013), 1304.4251.
- [586] C. Bomhof, P. Mulders, and F. Pijlman, *Eur. Phys. J. C* **47**, 147 (2006), hep-ph/0601171.
- [587] W. J. den Dunnen, J. P. Lansberg, C. Pisano, and M. Schlegel, *Phys. Rev. Lett.* **112**, 212001 (2014), 1401.7611.
- [588] J.-P. Lansberg, C. Pisano, F. Scarpa, and M. Schlegel, *Phys. Lett. B* **784**, 217 (2018), [Erratum: *Phys.Lett.B* 791, 420–421 (2019)], 1710.01684.
- [589] A. Bacchetta, D. Boer, C. Pisano, and P. Taels, *Eur. Phys. J. C* **80**, 72 (2020), 1809.02056.
- [590] D. Gutierrez-Reyes, S. Leal-Gomez, I. Scimemi, and A. Vladimirov, *JHEP* **11**, 121 (2019), 1907.03780.
- [591] F. Scarpa, D. Boer, M. G. Echevarria, J.-P. Lansberg, C. Pisano, and M. Schlegel, *Eur. Phys. J. C* **80**, 87 (2020), 1909.05769.
- [592] C. Adolph et al. (COMPASS), *Phys. Lett. B* **772**, 854 (2017), 1701.02453.
- [593] U. D'Alesio, F. Murgia, C. Pisano, and P. Taels, *Phys. Rev. D* **96**, 036011 (2017), 1705.04169.
- [594] U. D'Alesio, C. Flore, F. Murgia, C. Pisano, and P. Taels, *Phys. Rev. D* **99**, 036013 (2019), 1811.02970.
- [595] U. D'Alesio, F. Murgia, C. Pisano, and P. Taels, *Phys. Rev. D* **100**, 094016 (2019), 1908.00446.
- [596] R. Boussarie, Y. Hatta, L. Szymanowski, and S. Wallon, *Phys. Rev. Lett.* **124**, 172501 (2020), 1912.08182.
- [597] V. S. Fadin, E. Kuraev, and L. Lipatov, *Phys. Lett. B* **60**, 50 (1975).
- [598] M. Hentschinski, A. Sabio Vera, and C. Salas, *Phys. Rev. Lett.* **110**, 041601 (2013), 1209.1353.
- [599] A. Besse, L. Szymanowski, and S. Wallon, *JHEP* **11**, 062 (2013), 1302.1766.
- [600] A. D. Bolognino, F. G. Celiberto, D. Yu. Ivanov, and A. Papa, *Eur. Phys. J. C* **78**, 1023 (2018), 1808.02395.
- [601] A. D. Bolognino, F. G. Celiberto, D. Yu. Ivanov, and A. Papa, *Frascati Phys. Ser.* **67**, 76 (2018), 1808.02958.
- [602] A. D. Bolognino, F. G. Celiberto, D. Yu. Ivanov, and A. Papa, *Acta Phys. Polon. Supp.* **12**, 891 (2019), 1902.04520.
- [603] A. D. Bolognino, A. Szczurek, and W. Schaefer, *Phys. Rev. D* **101**, 054041 (2020), 1912.06507.
- [604] F. G. Celiberto, *Nuovo Cim.* **C42**, 220 (2019), 1912.11313.
- [605] D. Brzeminski, L. Motyka, M. Sadzikowski, and T. Stebel, *JHEP* **01**, 005 (2017), 1611.04449.



- [606] A. Arroyo Garcia, M. Hentschinski, and K. Kutak, *Phys. Lett. B* **795**, 569 (2019), 1904.04394.
- [607] M. Nefedov, *Phys. Rev. D* **104**, 054039 (2021), 2105.13915.
- [608] M. Hentschinski, *Phys. Rev. D* **104**, 054014 (2021), 2107.06203.
- [609] A. D. Bolognino, F. G. Celiberto, D. Yu. Ivanov, A. Papa, W. Schäfer, and A. Szczurek, *Eur. Phys. J. C* **81**, 846 (2021), 2107.13415.
- [610] A. D. Bolognino, F. G. Celiberto, D. Yu. Ivanov, and A. Papa (2021), 2107.12725.
- [611] A. D. Bolognino, F. G. Celiberto, M. Fucilla, D. Yu. Ivanov, A. Papa, W. Schäfer, and A. Szczurek (2022), 2202.02513.
- [612] D. Boer, S. J. Brodsky, P. J. Mulders, and C. Pisano, *Phys. Rev. Lett.* **106**, 132001 (2011), 1011.4225.
- [613] P. Sun, B.-W. Xiao, and F. Yuan, *Phys. Rev. D* **84**, 094005 (2011), 1109.1354.
- [614] D. Boer, W. J. den Dunnen, C. Pisano, M. Schlegel, and W. Vogelsang, *Phys. Rev. Lett.* **108**, 032002 (2012), 1109.1444.
- [615] C. Pisano, D. Boer, S. J. Brodsky, M. G. Buffing, and P. J. Mulders, *JHEP* **10**, 024 (2013), 1307.3417.
- [616] W. J. den Dunnen, J. P. Lansberg, C. Pisano, and M. Schlegel, *Phys. Rev. Lett.* **112**, 212001 (2014), 1401.7611.
- [617] J.-P. Lansberg, C. Pisano, and M. Schlegel, *Nucl. Phys.* **B920**, 192 (2017), 1702.00305.
- [618] E. Chapon et al., *Prog. Part. Nucl. Phys.* **122**, 103906 (2022), 2012.14161.
- [619] A. Arbuzov et al., *Prog. Part. Nucl. Phys.* **119**, 103858 (2021), 2011.15005.
- [620] V. M. Abazov et al. (SPD proto) (2021), 2102.00442.
- [621] A. Bacchetta, F. G. Celiberto, M. Radici, and P. Tael, *Eur. Phys. J. C* **80**, 733 (2020), 2005.02288.
- [622] A. Bacchetta, F. G. Celiberto, M. Radici, and P. Tael, *Eur. Phys. J. C* **81**, 13446 (2021), 2107.13446.
- [623] F. G. Celiberto, *Nuovo Cim.* **C44**, 36 (2021), 2101.04630.
- [624] F. G. Celiberto (2022), 2202.04207.
- [625] A. Bacchetta, F. Conti, and M. Radici, *Phys. Rev. D* **78**, 074010 (2008), 0807.0323.
- [626] A. Bacchetta, M. Radici, F. Conti, and M. Guagnelli, *Eur. Phys. J. A* **45**, 373 (2010), 1003.1328.
- [627] A. Bacchetta, F. G. Celiberto, and M. Radici (2021), 2111.01686.
- [628] A. Bacchetta, F. G. Celiberto, and M. Radici (2021), 2111.03567.
- [629] A. Bacchetta, F. G. Celiberto, and M. Radici (2022), 2201.10508.
- [630] F. G. Celiberto, *Eur. Phys. J. C* **81**, 691 (2021), 2008.07378.
- [631] A. D. Bolognino, F. G. Celiberto, M. Fucilla, D. Yu. Ivanov, and A. Papa, *Eur. Phys. J. C* **79**, 939 (2019), 1909.03068.
- [632] F. G. Celiberto, D. Yu. Ivanov, M. M. A. Mohammed, and A. Papa, *Eur. Phys. J. C* **81**, 293 (2021), 2008.00501.
- [633] A. D. Bolognino, F. G. Celiberto, M. Fucilla, D. Yu. Ivanov, and A. Papa, *Phys. Rev. D* **103**, 094004 (2021), 2103.07396.
- [634] F. G. Celiberto, D. Yu. Ivanov, and A. Papa, *Phys. Rev. D* **102**, 094019 (2020), 2008.10513.
- [635] F. G. Celiberto, M. Fucilla, D. Yu. Ivanov, and A. Papa, *Eur. Phys. J. C* **81**, 780 (2021), 2105.06432.
- [636] F. G. Celiberto, M. Fucilla, D. Yu. Ivanov, M. M. A. Mohammed, and A. Papa, *Phys. Rev. D* **104**, 114007 (2021), 2109.11875.
- [637] F. G. Celiberto and M. Fucilla (2022), 2202.12227.
- [638] J. C. Collins and D. E. Soper, *Nucl. Phys. B* **194**, 445 (1982).
- [639] J. C. Collins, D. E. Soper, and G. F. Sterman, *Nucl. Phys. B* **250**, 199 (1985).
- [640] J. C. Collins and D. E. Soper, *Nucl. Phys. B* **193**, 381 (1981), [Erratum: *Nucl.Phys.B* 213, 545 (1983)].
- [641] X.-d. Ji, J.-p. Ma, and F. Yuan, *Phys. Rev. D* **71**, 034005 (2005), hep-ph/0404183.
- [642] M. Beneke and T. Feldmann, *Nucl. Phys. B* **685**, 249 (2004), hep-ph/0311335.
- [643] J.-y. Chiu, F. Golf, R. Kelley, and A. V. Manohar, *Phys. Rev. Lett.* **100**, 021802 (2008), 0709.2377.
- [644] T. Becher and G. Bell, *Phys. Lett. B* **713**, 41 (2012), 1112.3907.
- [645] J.-y. Chiu, A. Jain, D. Neill, and I. Z. Rothstein, *Phys. Rev. Lett.* **108**, 151601 (2012), 1104.0881.
- [646] J.-Y. Chiu, A. Jain, D. Neill, and I. Z. Rothstein, *JHEP* **05**, 084 (2012), 1202.0814.
- [647] J.-y. Chiu, A. Fuhrer, A. H. Hoang, R. Kelley, and A. V. Manohar, *Phys. Rev. D* **79**, 053007 (2009), 0901.1332.
- [648] M. G. Echevarria, A. Idilbi, and I. Scimemi, *JHEP* **07**, 002 (2012), 1111.4996.
- [649] Y. Li, D. Neill, and H. X. Zhu, *Nucl. Phys. B* **960**, 115193 (2020), 1604.00392.
- [650] M. A. Ebert, I. Moul, I. W. Stewart, F. J. Tackmann, G. Vita, and H. X. Zhu, *JHEP* **04**, 123 (2019), 1812.08189.
- [651] M. A. Ebert, I. W. Stewart, and Y. Zhao, *JHEP* **09**, 037 (2019), 1901.03685.
- [652] S. Catani, L. Cieri, D. de Florian, G. Ferrera, and M. Grazzini, *Eur. Phys. J. C* **72**, 2195 (2012), 1209.0158.
- [653] T. Gehrmann, T. Lubbert, and L. L. Yang, *Phys. Rev. Lett.* **109**, 242003 (2012), 1209.0682.
- [654] T. Gehrmann, T. Lubbert, and L. L. Yang, *JHEP* **06**, 155 (2014), 1403.6451.
- [655] M. G. Echevarria, I. Scimemi, and A. Vladimirov, *JHEP* **09**, 004 (2016), 1604.07869.
- [656] M.-X. Luo, X. Wang, X. Xu, L. L. Yang, T.-Z. Yang, and H. X. Zhu, *JHEP* **10**, 083 (2019), 1908.03831.
- [657] S. Catani and M. Grazzini, *Eur. Phys. J. C* **72**, 2132 (2012), [Erratum: *Eur.Phys.J.C* 72, 2132 (2012)], 1106.4652.
- [658] M.-X. Luo, T.-Z. Yang, H. X. Zhu, and Y. J. Zhu, *JHEP* **01**, 040 (2020), 1909.13820.
- [659] M.-x. Luo, T.-Z. Yang, H. X. Zhu, and Y. J. Zhu, *Phys. Rev. Lett.* **124**, 092001 (2020), 1912.05778.
- [660] M. A. Ebert, B. Mistlberger, and G. Vita, *JHEP* **09**, 146 (2020), 2006.05329.
- [661] M. A. Ebert, B. Mistlberger, and G. Vita, *JHEP* **09**, 181 (2020), 2006.03055.
- [662] M. A. Ebert, B. Mistlberger, and G. Vita, *JHEP* **09**, 143 (2020), 2006.03056.
- [663] H. Chen, T.-Z. Yang, H. X. Zhu, and Y. J. Zhu, *Chin. Phys. C* **45**, 043101 (2021), 2006.10534.
- [664] M.-x. Luo, T.-Z. Yang, H. X. Zhu, and Y. J. Zhu, *JHEP* **06**, 115 (2021), 2012.03256.
- [665] M. A. Ebert, B. Mistlberger, and G. Vita, *JHEP* **07**, 121 (2021), 2012.07853.
- [666] M. A. Ebert, B. Mistlberger, and G. Vita, *JHEP* **08**, 022 (2021), 2012.07859.
- [667] J. C. Collins and D. E. Soper, *Nucl. Phys. B* **197**, 446 (1982).

- [668] Y. Li and H. X. Zhu, Phys. Rev. Lett. **118**, 022004 (2017), 1604.01404.
- [669] T. Neumann, Eur. Phys. J. C **81**, 905 (2021), 2107.12478.
- [670] U. D'Alesio, M. G. Echevarria, S. Melis, and I. Scimemi, JHEP **11**, 098 (2014), 1407.3311.
- [671] I. Scimemi and A. Vladimirov, Eur. Phys. J. **C78**, 89 (2018), 1706.01473.
- [672] V. Bertone, I. Scimemi, and A. Vladimirov, JHEP **06**, 028 (2019), 1902.08474.
- [673] A. Signori, A. Bacchetta, M. Radici, and G. Schnell, JHEP **11**, 194 (2013), 1309.3507.
- [674] M. Anselmino, M. Boglione, J. O. Gonzalez Hernandez, S. Melis, and A. Prokudin, JHEP **04**, 005 (2014), 1312.6261.
- [675] A. Bermudez Martinez, P. Connor, H. Jung, A. Lelek, R. Žlebčák, F. Hautmann, and V. Radescu, Phys. Rev. D **99**, 074008 (2019), 1804.11152.
- [676] M. G. Echevarria, A. Idilbi, Z.-B. Kang, and I. Vitev, Phys. Rev. **D89**, 074013 (2014), 1401.5078.
- [677] P. Sun, J. Isaacson, C. P. Yuan, and F. Yuan, Int. J. Mod. Phys. **A33**, 1841006 (2018), 1406.3073.
- [678] A. Bacchetta, F. Delcarro, C. Pisano, M. Radici, and A. Signori, JHEP **06**, 081 (2017), [Erratum: JHEP06,051(2019)], 1703.10157.
- [679] I. Scimemi and A. Vladimirov, JHEP **06**, 137 (2020), 1912.06532.
- [680] A. Bacchetta, V. Bertone, C. Bissolotti, G. Bozzi, F. Delcarro, F. Piacenza, and M. Radici, JHEP **07**, 117 (2020), 1912.07550.
- [681] A. Bacchetta, G. Bozzi, M. Radici, M. Ritzmann, and A. Signori, Phys. Lett. **B788**, 542 (2019), 1807.02101.
- [682] T. Sjostrand, S. Mrenna, and P. Z. Skands, JHEP **05**, 026 (2006), hep-ph/0603175.
- [683] J. Arrington et al. (2021), 2112.00060.
- [684] V. D. Burkert et al., Nucl. Instrum. Meth. A **959**, 163419 (2020).
- [685] J. P. Chen, H. Gao, T. K. Hemmick, Z. E. Meziani, and P. A. Souder (SoLID) (2014), 1409.7741.
- [686] C. Hadjidakis et al., Phys. Rept. **911**, 1 (2021), 1807.00603.
- [687] J. Jaeckel, M. Lamont, and C. Vallée, Nature Phys. **16**, 393 (2020).
- [688] E. Steffens, PoS **PSTP2015**, 019 (2015).
- [689] A. Airapetian et al. (HERMES), Nucl. Instrum. Meth. A **540**, 68 (2005), physics/0408137.
- [690] A. Bursche et al. (2018).
- [691] A. Dainese et al. (QCD Working Group) (2019), 1901.04482.
- [692] E.-C. Aschenauer et al. (2016), 1602.03922.
- [693] W. Altmannshofer et al. (Belle-II), PTEP **2019**, 123C01 (2019), [Erratum: PTEP 2020, 029201 (2020)], 1808.10567.
- [694] K. Abe et al. (Belle), Phys. Rev. Lett. **96**, 232002 (2006), hep-ex/0507063.
- [695] R. Seidl et al. (Belle), Phys. Rev. D **78**, 032011 (2008), [Erratum: Phys.Rev.D 86, 039905 (2012)], 0805.2975.
- [696] Y. Guan et al. (Belle), Phys. Rev. Lett. **122**, 042001 (2019), 1808.05000.
- [697] H. Li et al. (Belle), Phys. Rev. D **100**, 092008 (2019), 1909.01857.
- [698] R. Seidl et al. (Belle), Phys. Rev. D **99**, 112006 (2019), 1902.01552.
- [699] F. Dulat, B. Mistlberger, and A. Pelloni, JHEP **01**, 145 (2018), 1710.03016.
- [700] R. Nagar, PoS **DIS2019**, 022 (2019), 1906.10059.
- [701] M. Diehl, R. Nagar, and F. J. Tackmann (2021), 2112.09703.
- [702] S. Carrazza, J. M. Cruz-Martinez, and M. Rossi, Comput. Phys. Commun. **264**, 107995 (2021), 2009.06635.
- [703] S. Alekhin et al., Eur. Phys. J. C **75**, 304 (2015), 1410.4412.
- [704] E. Accomando et al., JHEP **10**, 176 (2019), 1907.07727.
- [705] H. Abdolmaleki et al. (xFitter Developers' Team), Eur. Phys. J. **C79**, 864 (2019), 1907.01014.
- [706] S. Amoroso, J. Fiaschi, F. Giuli, A. Glazov, F. Hautmann, and O. Zenaiev, Phys. Lett. B **821**, 136613 (2021), 2012.10298.
- [707] J. Fiaschi, F. Giuli, F. Hautmann, and S. Moretti, Nucl. Phys. B **968**, 115444 (2021), 2103.10224.
- [708] H. Abdolmaleki et al. (xFitter Developers' Team), Eur. Phys. J. C **78**, 621 (2018), 1802.00064.
- [709] M. Bonvini and F. Giuli, Eur. Phys. J. Plus **134**, 531 (2019), 1902.11125.
- [710] V. Bertone et al. (xFitter Developers' Team), JHEP p. 050 (2016), 1605.01946.
- [711] F. Giuli et al. (xFitter Developers' Team), Eur. Phys. J. **C77**, 400 (2017), 1701.08553.
- [712] Z. Kassabov, *Reportengine: A framework for declarative data analysis* (2019), URL <https://doi.org/10.5281/zenodo.2571601>.
- [713] K. Cichy, L. Del Debbio, and T. Giani, JHEP **10**, 137 (2019), 1907.06037.
- [714] E. L. Berger, M. Guzzi, H.-L. Lai, P. M. Nadolsky, and F. I. Olness, Phys. Rev. D **82**, 114023 (2010), 1010.4315.
- [715] C. Adloff et al. (H1), Eur. Phys. J. C **19**, 289 (2001), hep-ex/0010054.
- [716] T. Carli, G. P. Salam, and F. Siegert, Contributed to HERA and the LHC: a Workshop on the Implications of HERA for LHC Physics p. 110 (2005), hep-ph/0510324.
- [717] T. Carli, D. Clements, A. Cooper-Sarkar, C. Gwenlan, G. P. Salam, F. Siegert, P. Starovoitov, and M. Sutton, Eur. Phys. J. **C66**, 503 (2010), 0911.2985.
- [718] T. Kluge, K. Rabbertz, and M. Wobisch, in *14th International Workshop on Deep Inelastic Scattering (DIS 2006)* (Tsukuba, Japan, April 20-24, 2006), p. 483, hep-ph/0609285.
- [719] D. Britzger, K. Rabbertz, F. Stober, and M. Wobisch, in *Proceedings, XX. International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2012)* (Bonn, Germany, March 26-30, 2012), p. 217, 1208.3641.
- [720] D. Britzger et al., Eur. Phys. J. C **79**, 845 (2019), [Erratum: Eur.Phys.J.C 81, 957 (2021)], 1906.05303.
- [721] T. Gehrmann et al., PoS **RADCOR2017**, 074 (2018), 1801.06415.
- [722] Ploughshare, <http://ploughshare.web.cern.ch>.
- [723] M. Aaboud et al. (ATLAS), JHEP **09**, 020 (2017), 1706.03192.

- [724] C. Schwan, A. Candido, F. Hekhorn, and S. Carrazza, *N3PDF/pineappl*, <https://doi.org/10.5281/zenodo.3890291>.
- [725] C. Schwan, in *28th International Workshop on Deep Inelastic Scattering and Related Subjects* (2021), 2108.05816.
- [726] A. Candido, F. Hekhorn, and G. Magni, *yadism: Yet Another DIS Module*, In preparation.
- [727] A. Candido, F. Hekhorn, and G. Magni, *N3PDF/yadism: NNLO*, <https://doi.org/10.5281/zenodo.5795955> (2021).
- [728] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali, *Nucl. Phys. B* **838**, 136 (2010), 1002.4407.
- [729] L. Del Debbio, S. Forte, J. I. Latorre, A. Piccione, and J. Rojo (NNPDF), *JHEP* **03**, 039 (2007), hep-ph/0701127.
- [730] A. De Roeck, in *HERA and the LHC: 4th Workshop on the Implications of HERA for LHC Physics* (2009), pp. 125–126.
- [731] S. Alekhin et al. (2011), 1101.0536.
- [732] M. Botje et al. (2011), 1101.0538.
- [733] R. D. Ball et al., *JHEP* **04**, 125 (2013), 1211.5142.
- [734] J. Rojo et al., *J. Phys. G* **42**, 103103 (2015), 1507.00556.
- [735] J. R. Andersen et al., in *9th Les Houches Workshop on Physics at TeV Colliders* (2016), 1605.04692.
- [736] J. R. Andersen et al. (2014), 1405.1067.
- [737] R. D. Ball et al. (NNPDF), *JHEP* **04**, 040 (2015), 1410.8849.
- [738] G. Watt and R. S. Thorne, *JHEP* **08**, 052 (2012), 1205.4024.
- [739] G. Heinrich, *Phys. Rept.* **922**, 1 (2021), 2009.00516.
- [740] J. Currie, E. W. N. Glover, and J. Pires, *Phys. Rev. Lett.* **118**, 072002 (2017), 1611.01460.
- [741] J. Currie, A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, A. Huss, and J. Pires, *Phys. Rev. Lett.* **119**, 152001 (2017), 1705.10271.
- [742] M. Czakon, D. Heymes, and A. Mitov, *Phys. Rev. Lett.* **116**, 082003 (2016), 1511.00549.
- [743] J. Gao, *JHEP* **02**, 026 (2018), 1710.04258.
- [744] R. Abdul Khalek et al., *Eur. Phys. J. C* **80**, 797 (2020), 2005.11327.
- [745] F. Faura, S. Iranipour, E. R. Nocera, J. Rojo, and M. Ubiali, *Eur. Phys. J. C* **80**, 1168 (2020), 2009.00014.
- [746] M. Czakon, N. P. Hartland, A. Mitov, E. R. Nocera, and J. Rojo, *JHEP* **04**, 044 (2017), 1611.08609.
- [747] J. M. Campbell, J. Rojo, E. Slade, and C. Williams, *Eur. Phys. J. C* **78**, 470 (2018), 1802.03021.
- [748] M. Czakon, S. Dulat, T.-J. Hou, J. Huston, A. Mitov, A. S. Papanastasiou, I. Sitiwaldi, Z. Yu, and C. P. Yuan, *J. Phys. G* **48**, 015003 (2020), 1912.08801.
- [749] R. S. Thorne, S. Bailey, T. Cridge, L. A. Harland-Lang, A. Martin, and R. Nathvani, *PoS DIS2019*, 036 (2019), 1907.08147.
- [750] R. Boughezal, A. Guffanti, F. Petriello, and M. Ubiali, *JHEP* **07**, 130 (2017), 1705.00343.
- [751] O. Amat et al. (2019), 1908.06441.
- [752] T.-J. Hou, Z. Yu, S. Dulat, C. Schmidt, and C.-P. Yuan, *Phys. Rev. D* **100**, 114024 (2019), 1907.12177.
- [753] G. Aad et al. (ATLAS), *Eur. Phys. J. C* **76**, 538 (2016), 1511.04716.
- [754] S. Amoroso et al., in *11th Les Houches Workshop on Physics at TeV Colliders: PhysTeV Les Houches* (2020), 2003.01700.
- [755] G. Aad et al. (ATLAS), *JHEP* **07**, 223 (2021), 2101.05095.
- [756] CTEQ-TEA collaboration, <https://ct.hepforge.org/>.
- [757] T.-J. Hou, S. Dulat, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan, *Phys. Rev. D* **95**, 034003 (2017), 1609.07968.
- [758] T.-J. Hou et al., *JHEP* **03**, 099 (2017), 1607.06066.
- [759] T. Hobbs, B.-T. Wang, P. M. Nadolsky, and F. I. Olness, *Phys. Rev. D* **100**, 094040 (2019), 1904.00022.
- [760] T. Han, Y. Ma, and K. Xie, *Phys. Rev. D* **103**, L031301 (2021), 2007.14300.
- [761] L. Buonocore, P. Nason, F. Tramontano, and G. Zanderighi, *JHEP* **08**, 019 (2020), 2005.06477.