


Pion fragmentation functions at high energy collidersIgnacio Borsa^{*} and Rodolfo Sassot[†]*Departamento de Física and IFIBA, Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 1 (1428) Buenos Aires, Argentina*Daniel de Florian[‡]*International Center for Advanced Studies (ICAS) and IFICI, UNSAM,
Campus Miguelete, 25 de Mayo y Francia (1650) Buenos Aires, Argentina*Marco Stratmann[§]*Institute for Theoretical Physics, University of Tübingen,
Auf der Morgenstelle 14, 72076 Tübingen, Germany* (Received 28 October 2021; accepted 21 January 2022; published 9 February 2022)

We revisit the description of pion production in proton-proton collisions in the light of the very precise data taken at the Large Hadron Collider (LHC) over the past decade. First attempts to include LHC results in next-to-leading order global QCD analyses of parton-to-pion fragmentation functions insinuated some conflict between data sets at different center-of-mass system energies. We show that the data can be well described within their uncertainties by a consistent set of pion fragmentation functions once the theoretical scale dependence is taken into account in the global QCD analysis.

DOI: [10.1103/PhysRevD.105.L031502](https://doi.org/10.1103/PhysRevD.105.L031502)**I. INTRODUCTION**

Pions are the most copiously produced particles in proton-proton collisions, and their different parton content and distinctive decay modes make them invaluable tools to analyze the breakdown of the colliding protons and the subsequent recombination of their constituents into new hadrons [1]. The precise measurement of these processes and their comparison with the most accurate predictions tests our understanding of hard scattering processes as described by the theory of strong interactions, quantum chromodynamics (QCD) [2].

Fragmentation functions (FFs) play a cardinal role in the perturbative QCD description of processes involving identified hadrons in the final state, as they connect the scattering of partons at short distance to the emerging hadrons observed in experiment [3]. From a phenomenological point of view, these functions parametrize the probabilities for the measured final state hadron to evolve from the partonic seed excited in the hard interaction [1].

In the field theoretical framework, the nonperturbative, scale-dependent FFs are used to factorize any information on the hadronization process happening at long distances from the calculable partonic cross sections, thereby canceling final state singularities at each order of perturbation theory [3].

The notion of FFs is in complete analogy to that of the more familiar parton distribution functions (PDFs). In fact, the first attempts to infer them from data date back to the 1970s, when combined PDF and FF extractions were first explored [1]. The experimental challenges to obtain precise data on hard processes with observed final state hadrons, not only well identified but also fully characterized kinematically, hindered the progress in the determination of FFs compared to that of PDFs. Until recently, FFs were obtained solely from the analysis of single inclusive electron-positron annihilation (SIA) data.

In 2007, a first global QCD analysis of FFs at next-to-leading order (NLO) accuracy was presented in Ref. [4]. It combined SIA data with pion production measurements performed in semi-inclusive deep inelastic scattering (SIDIS) and proton-proton collisions (PP) and demonstrated the anticipated universality and factorization properties of FFs within the precision available at that time. Similar analyses have been published since then [5–7].

The analysis of Ref. [4] was subsequently updated in [8] to account for the extremely precise SIA measurements from the Belle [9] and BABAR [10] collaborations, SIDIS

^{*}iborsa@df.uba.ar[†]sassot@df.uba.ar[‡]deflo@unsam.edu.ar[§]marco.stratmann@uni-tuebingen.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

data from COMPASS [11], and the first PP results from the LHC provided by the ALICE experiment [12].

The results hinted at some degree of tension between the data sets, most noticeable between the LHC and the lower energy hadroproduction experiments at BNL-RHIC. In fact, a slight tension was even observed between ALICE data taken at different center-of-mass system (c.m.s.) energies (0.9 TeV and 7 TeV, respectively), suggesting possible limitations in the assumptions made in the analysis, such as, for example, the accuracy of the NLO approximation employed, or even of the underlying factorization and universality property for FFs, the foundation of the theoretical description.

In Ref. [8] it was also shown that the tension between the experimental sets can somehow be alleviated by discarding PP data in the fit where the observed pion's transverse momentum p_T is less than about 5 GeV. In this region the NLO approximation is expected to be less and less reliable with decreasing p_T . In addition, a very conservative normalization uncertainty, comparable in size to the large factorization scale dependence exhibited by the proton-proton cross sections, was introduced. However, this strategy was unable to fully reconcile the p_T dependence as predicted by the NLO approximation with the trend of the data. Since then, both ALICE and the RHIC experiments have delivered additional, remarkably precise PP data on neutral and charged pion production at different c.m.s. energies. They not only confirmed their previous results but also clearly showed that the shortcomings of the FF analysis of Ref. [8] further deepen with increasing c.m.s. energy.

In the following, we revisit pion production up to LHC energies in the light of the most recent measurements and by making use of up-to-date information on PDFs. We will show that the NLO framework can, contrary to the findings of Ref. [8], provide an accurate description of the world data not only at different c.m.s. energies but also down to values of p_T around 1 GeV. To this end, we need to exploit the factorization scale uncertainty that is inherent to any perturbative QCD estimate. More specifically, we perform a NLO global analysis similar to those of Refs. [4,8], based on the most recent sets of hard scattering data with identified pions in the final state, which explores different choices for the, in principle, arbitrary factorization scale at different c.m.s. energies. At variance with Refs. [4,8], where the uncertainties of the obtained FFs were estimated using Lagrange multipliers and the improved Hessian technique, respectively, the present analysis implements a Monte Carlo sampling approach to produce a large sample of replicas for the FFs. In this way, estimates of the uncertainties inherited by any observable computed with our FFs can be obtained much more easily than in the Lagrange multipliers approach and without the complications related to choosing a particular tolerance criterion as in the Hessian framework.

II. GLOBAL ANALYSIS: SETUP AND SETS OF DATA

As just described, crucial to our new analysis are the very precise PP data produced by ALICE at different c.m.s. energies \sqrt{s} , comprising 2.76, 7, 13 TeV and 0.9, 2.76, 7, 8 TeV for charged and neutral pion production, respectively [12–15], as well as the STAR and PHENIX data at 0.2 and 0.51 TeV [16–21]. We also include the final π^\pm SIDIS data by COMPASS instead of the preliminary sets still used in Ref. [8]. In this way, our global analysis covers the wide range of c.m.s. energies spanned by the different PP experiments, the latest SIDIS data, as well as all the SIDIS and SIA data already available in Ref. [8].

The general strategy for our global analysis has been described in detail in Refs. [4,8] and need not be repeated here. It is based on the numerically efficient Mellin-moment technique that allows one to tabulate and store the computationally most demanding parts of a NLO calculation prior to the χ^2 minimization from which we infer the optimum shapes of our FFs that best suit the data. The FFs for a parton of flavor i into a positively charged pion are parametrized at our initial scale $Q_0 = 1$ GeV as

$$D_i^{\pi^+}(z, Q_0) = N_i z^{\alpha_i} \sum_{j=1}^3 \gamma_j (1-z)^{\beta_{ij}}, \quad (1)$$

where the free parameters N_i , α_i , β_{ij} , and γ_j are determined by the fit. As usual, we assume charge conjugation symmetry, i.e., $D_q^{\pi^+} = D_{\bar{q}}^{\pi^-}$, and isospin symmetry for the unfavored FFs of the light sea quarks $D_u^{\pi^+} = D_d^{\pi^-}$.

The computation of precise partonic cross sections is a key ingredient in the quest for extracting FFs from data. One of the main issues in this respect is the need to truncate the perturbative expansion at a given fixed order. While next-to-NLO (NNLO) QCD corrections are available in the case of SIA, only NLO results are at our disposal for SIDIS and PP. For the latter, estimates for the size of the missing higher orders have been obtained by means of all-order expansions around the threshold for the partonic reaction [22,23], finding rather large corrections along with a reduction of the scale dependence and, in the case of PP, a much better agreement with the experimental data. But since the full set of NNLO corrections is still unavailable, we perform our analysis at NLO accuracy. For the same reason, we restrict ourselves to the massless quark approximation, even though heavy quark mass effects have been shown to be sizable in SIA analyses of pions and kaons [24,25].

A possible way to get a rough estimate of the relevance of the missing higher order terms is by analyzing the dependence of the results on the unphysical renormalization scale, μ_R , and factorization scales for both the initial (PDFs) and final (FFs) states, μ_{FI} and μ_{FF} , respectively. It is customary to set the default value of those scales to be of

the same order as the typical physical scale \mathcal{E} of the process, i.e., in our case, \sqrt{s} for SIA, Q for SIDIS, and p_T for PP. Next, one varies the scales by a factor of 2 up and down; more sophisticated procedures can be found in [26,27]. In this paper, we follow the standard approach, choose all scales to be the same $\mu_R = \mu_{FI} = \mu_{FF} = \kappa\mathcal{E}$, and assume that any value of κ between 1/2 and 2 is equally acceptable. On the one hand, this choice allows us to define a scale uncertainty band for each observable; on the other hand, we can treat κ as a free parameter for each experiment, or group of experiments with similar kinematics, and let the fit select the values for κ that yield an optimum reproduction of the data. In this way, the longstanding issue with the description of the p_T distributions in hadronic collisions [28] can be solved.

Finally, at variance with the analysis of Ref. [8], where the factorization scale dependence was included as a theoretical error in the χ^2 minimization to which the Hessian method was applied, we first determine the optimum scale parameters κ for each group of experiments and then produce for those scales a large set of replicas of the FFs by fitting them to statistically equivalent replicas of the data used in the fit. The so-obtained replica ensemble provides a faithful representation of the uncertainties stemming from the FF parametrization. The optimum fit and its uncertainty estimates are assumed to be given by the statistical average of the obtained ensemble of replicas and their corresponding variance, respectively, thereby avoiding the arbitrariness of choosing a particular tolerance criterion. This Monte Carlo sampling approach with a fixed functional form has already been successfully implemented for a global analysis of helicity parton densities in Refs. [29,30].

III. RESULTS

Figure 1 confronts the precise charged pion PP data from ALICE with our results at NLO computed with the newly obtained FFs and the set of PDFs from Ref. [31]. For comparison, we also show the outcome of a similar fit with fixed $\kappa = 1$ and a calculation based on the DSS set of FFs [8]. Both results clearly fail in reproducing the p_T dependence of the data, with DSS overshooting most of them by a large amount. The discrepancies are more noticeable with increasing c.m.s. energy.

The fitted scale factors κ that optimize the agreement with the data—1.491, 1.165, and 0.912 for $\sqrt{s} = 13, 7$, and 2.76 TeV, respectively—are within the customary range and increase with increasing c.m.s. energy. The same values of κ reproduce ALICE π^0 data as can be inferred from Fig. 2. RHIC PP data on π^0 and π^\pm yields—shown in Figs. 2 and 3, respectively—prefer much smaller scale factors, $\kappa = 0.660$ and 0.705 for $\sqrt{s} = 0.2$ and 0.51 TeV, respectively, thus confirming the dependence on \sqrt{s} .

The right-hand sides of Figs. 1–3 provide for each set of data a detailed “(data-theory)/theory” comparison along

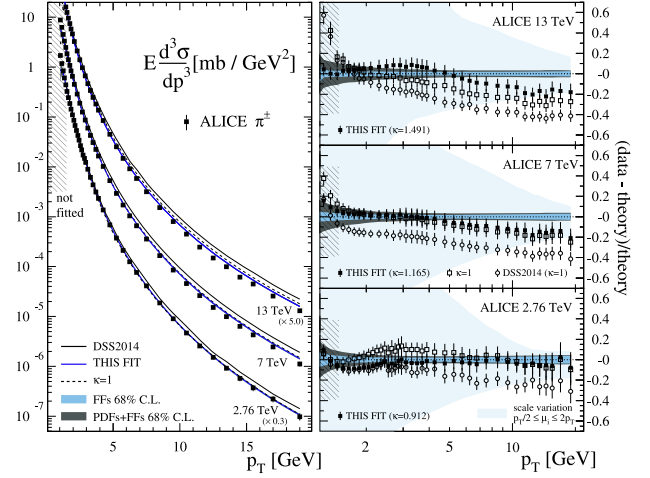


FIG. 1. Left-hand side: comparison of our best fit and other NLO results with ALICE π^\pm PP data. Right-hand side: “(data-theory)/theory” for each set of data and relevant sources of uncertainties for the new fit (shaded bands).

with estimates of the remaining relative uncertainties at NLO stemming from the FFs and, on top of the latter, from the PDFs (blue and grey bands, respectively). Thanks to the new sets of data and the lower cut in p_T possible in our analysis, the FF uncertainties of the present fit are significantly reduced compared to those found in Ref. [8]. The light blue bands indicate conservative estimates of the scale ambiguity relative to the $\kappa = 1$ fit for a “27 point,” independent variation of μ_R , μ_{FI} , and μ_{FF} within the standard range.

The much better agreement of the new FFs with the PP data compared to [8] does not spoil the accord with SIA or SIDIS data as can be seen in Figs. 4 and 5, respectively. In the fit of the SIA data we have again introduced scale factors κ (as indicated in the plot) for the different c.m.s. energies of the experiments in the range $\sqrt{s} = 10.52$ GeV

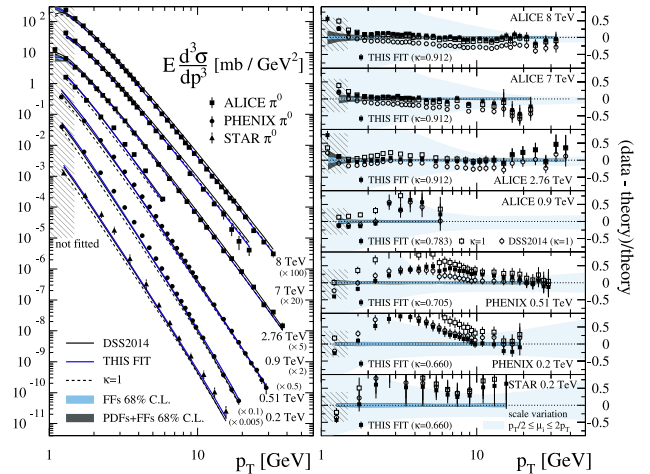


FIG. 2. Same as in Fig. 1 but for π^0 production data from ALICE, PHENIX, and STAR.

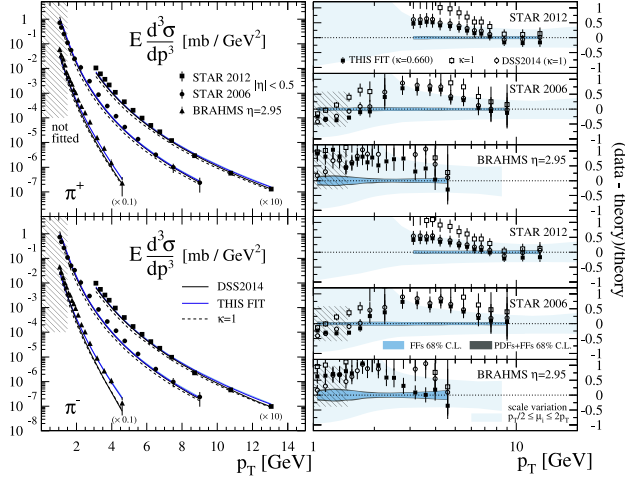


FIG. 3. Same as Fig. 1 but for π^\pm production data from STAR and BRAHMS at different ranges of pseudorapidity η .

to M_Z . Note that all available fits describe the SIA data equally well.

The present fit includes the final π^+ and π^- SIDIS multiplicities from COMPASS, which are superior both in number and precision to the preliminary set used in [8]. Our analysis includes data for $Q^2 \geq 1.5 \text{ GeV}^2$ and neglects deuteron nuclear corrections [32] throughout. The quality of the fit is illustrated in Fig. 5 in terms of a “(data-theory)/theory” comparison for each kinematic bin along with estimates for the relevant relative uncertainties of the fit (shaded bands). The data prefer a scale factor $\kappa = 1.872$, while the corresponding HERMES data [33] (not shown [34]) are best reproduced with $\kappa = 1.402$.

Figure 6 compares the shape of the newly obtained FFs for the different flavors to previous extractions by DSS and shows their absolute and relative uncertainties (shaded

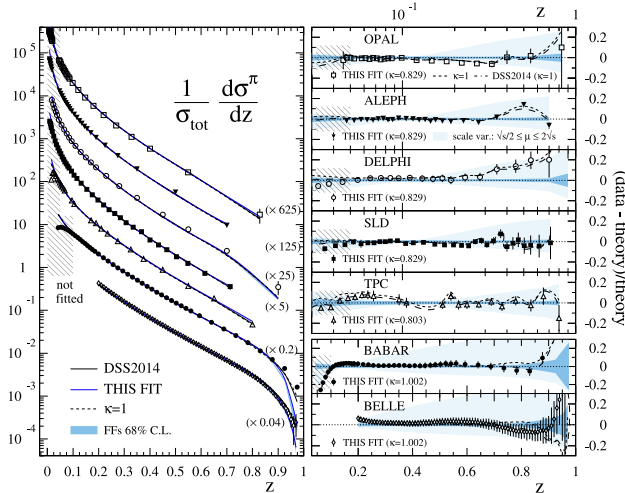


FIG. 4. Similar to Fig. 1 but for SIA data from various experiments.

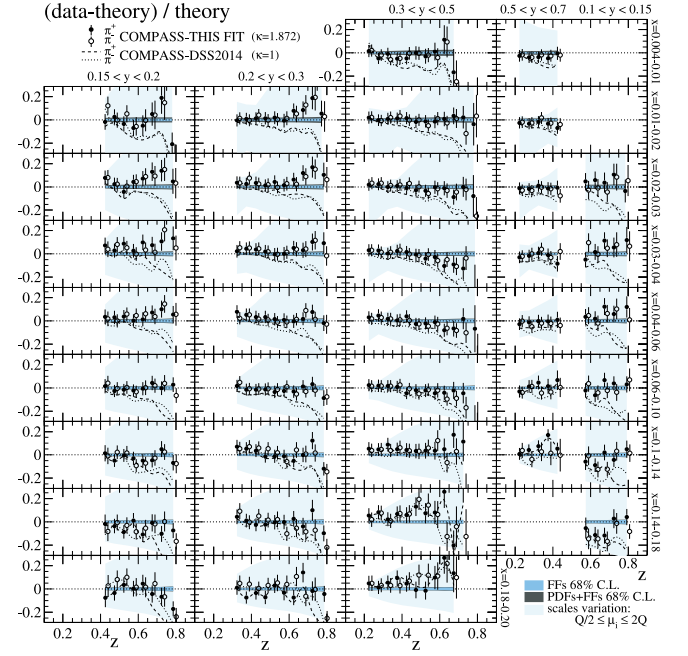


FIG. 5. “(Data-theory)/theory” comparison of the DSS and our best fit to the π^\pm SIDIS multiplicities from COMPASS along with various uncertainty estimates (shaded bands).

bands). None of the distributions is dramatically different than previous extractions, except in regions where they are still poorly constrained by data, but the remaining uncertainties are significantly reduced, especially for the gluon FF.

Finally, Table I summarizes the data sets used in our NLO global analysis, the computed normalization shifts N_i as defined in Eq. (6) of Ref. [8], and the χ^2 -values.

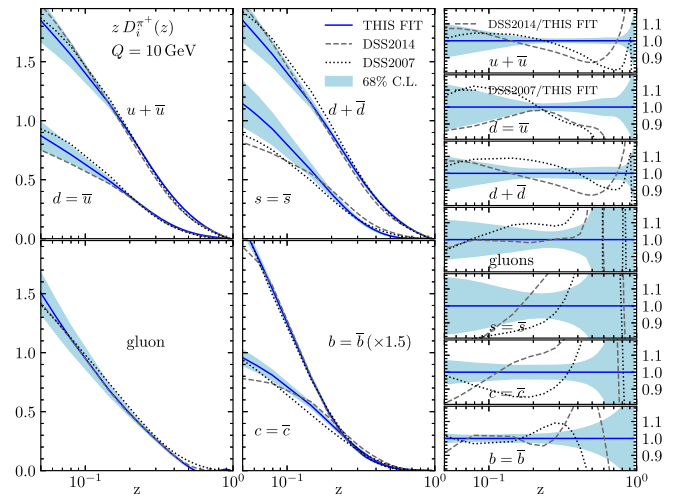


FIG. 6. Obtained individual FFs for π^+ at the scale $Q = 10 \text{ GeV}$ along with uncertainty estimates at 68% C.L. compared to previous analyses by DSS [4,8]. The panels on the right-hand side show the corresponding relative uncertainties.

TABLE I. Data sets, normalizations N_i as defined in Eq. (6) of [8], and the partial and total χ^2 values obtained in the fit.

Experiment		Data type	N_i	Number of data in fit	χ^2
TPC [35]	29 GeV	inclusive	1.038	17	22.0
		uds, c, b tag	1.038	27	16.6
TASSO [36]	34 GeV	inclusive	1.038	11	28.4
		44 GeV	inclusive	1.038	7
SLD [37]	91.2 GeV	inclusive	0.977	28	19.5
		uds, c, b tag	0.977	51	39.2
ALEPH [38]	91.2 GeV	inclusive	1.012	22	44.5
DELPHI [39]	91.2 GeV	inclusive	1.000	17	23.0
		uds, b tag	1.000	34	33.8
OPAL [40]	91.2 GeV	inclusive	1.000	21	31.5
		u, d, s, c, b tag	0.793	25	62.1
BABAR [10]	10.54 GeV	inclusive	1.060	45	142.4
BELLE [9]	10.52 GeV	inclusive	1.067	78	60.2
SIA data (sum)				378	544.1
HERMES [33]	π^+, π^-	(p- Q^2)	0.984	56	54.3
		(d- Q^2)	0.988	56	46.5
		(p-x)	1.007	56	159.5
		(d-x)	1.009	56	189.5
COMPASS [11]	π^+, π^-	(d-z)	1.004	510	302.1
SIDIS data (sum)				734	751.9
BRAHMS [41]	0.20 TeV	π^+, π^-	1.313	26	13.5
STAR [16–19]	0.20 TeV	π^0	1.190	12	8.2
		π^0	0.921	7	4.0
		π^+, π^-	1.029	26	37.3
PHENIX [20,21]	0.20 TeV	π^+, π^-	1.158	34	73.8
		π^0	1.177	22	13.8
		π^0	1.178	27	32.9
ALICE [12–15]	0.90 TeV	π^0	1.012	7	52.0
		π^0	1.002	24	17.4
	2.76 TeV	π^\pm	0.959	38	15.6
		π^0	1.016	25	30.6
	7 TeV	π^\pm	0.976	32	23.9
	8 TeV	π^0	1.048	36	34.5
13 TeV	π^\pm	0.981	32	56.2	
PP data (sum)				348	413.7
TOTAL:				1460	1709.7

IV. DISCUSSION AND CONCLUSIONS

We have shown that the same set of nonperturbative pion FFs that reproduces the world data on SIA, SIDIS, and PP at RHIC kinematics can also fully account for the recent, very precise measurements of pion production cross sections up to the highest energies available at the LHC. The obtained constraints on the FFs from this large set of data are significant.

The key ingredient of our new global analysis at NLO accuracy of QCD is to fully exploit the theoretical scale dependence that determines how much of a measured cross section is attributed to nonperturbative quantities such as FFs. This choice is, in principle, arbitrary when applied in a consistent fashion. Since the inclusion of different data sets taken at various energy scales in the global fit makes the choice

of scale not obvious, we let the data determine, within the conventional range, what optimum value is preferred for each set of data. The result is a new fit of parton-to-pion FFs with increased precision and significantly reduced uncertainties that turns one-particle-inclusive processes and FFs to a much better tool to unveil new aspects of hadron structure in the next generation of experiments such as the Electron-Ion Collider.

ACKNOWLEDGMENTS

I. B. wishes to thank the University of Tuebingen for hospitality during the final stages of the work. This work was supported in part by CONICET, ANPCyT, UBACyT, and by Deutsche Forschungsgemeinschaft (DFG) through the Research Unit FOR 2926 (Project No. 40824754).

- [1] R. D. Field and R. P. Feynman, *Phys. Rev. D* **15**, 2590 (1977).
- [2] J. C. Collins, D. E. Soper, and G. F. Sterman, *Adv. Ser. Dir. High Energy Phys.* **5**, 1 (1989).
- [3] J. C. Collins and D. E. Soper, *Nucl. Phys.* **B194**, 445 (1982).
- [4] D. de Florian, R. Sassot, and M. Stratmann, *Phys. Rev. D* **75**, 114010 (2007).
- [5] N. Sato, J. J. Ethier, W. Melnitchouk, M. Hirai, S. Kumano, and A. Accardi, *Phys. Rev. D* **94**, 114004 (2016).
- [6] V. Bertone *et al.* (NNPDF Collaboration), *Eur. Phys. J. C* **77**, 516 (2017).
- [7] R. A. Khalek, V. Bertone, and E. R. Nocera, *Phys. Rev. D* **104**, 034007 (2021).
- [8] D. de Florian, R. Sassot, M. Epele, R. J. Hernández-Pinto, and M. Stratmann, *Phys. Rev. D* **91**, 014035 (2015).
- [9] M. Leitgab *et al.* (Belle Collaboration), *Phys. Rev. Lett.* **111**, 062002 (2013).
- [10] J. P. Lees *et al.* (BABAR Collaboration), *Phys. Rev. D* **88**, 032011 (2013).
- [11] C. Adolph *et al.* (COMPASS Collaboration), *Phys. Lett. B* **764**, 1 (2017).
- [12] B. Abelev *et al.* (ALICE Collaboration), *Phys. Lett. B* **717**, 162 (2012).
- [13] S. Acharya *et al.* (ALICE Collaboration), *Eur. Phys. J. C* **77**, 339 (2017).
- [14] S. Acharya *et al.* (ALICE Collaboration), *Eur. Phys. J. C* **78**, 263 (2018).
- [15] S. Acharya *et al.* (ALICE Collaboration), *Eur. Phys. J. C* **81**, 256 (2021).
- [16] J. Adams *et al.* (STAR Collaboration), *Phys. Lett. B* **637**, 161 (2006).
- [17] B. I. Abelev *et al.* (STAR Collaboration), *Phys. Rev. D* **80**, 111108 (2009).
- [18] G. Agakishiev *et al.* (STAR Collaboration), *Phys. Rev. Lett.* **108**, 072302 (2012).
- [19] L. Adamczyk *et al.* (STAR Collaboration), *Phys. Rev. D* **89**, 012001 (2014).
- [20] A. Adare *et al.* (PHENIX Collaboration), *Phys. Rev. D* **76**, 051106 (2007).
- [21] A. Adare *et al.* (PHENIX Collaboration), *Phys. Rev. D* **93**, 011501 (2016).
- [22] M. Abele, D. de Florian, and W. Vogelsang, *Phys. Rev. D* **104**, 094046 (2021).
- [23] D. de Florian and W. Vogelsang, *Phys. Rev. D* **71**, 114004 (2005).
- [24] M. Epele, C. A. Garcia Canal, and R. Sassot, *Phys. Rev. D* **94**, 034037 (2016).
- [25] M. Epele, C. García Canal, and R. Sassot, *Phys. Lett. B* **790**, 102 (2019).
- [26] M. Bonvini, *Eur. Phys. J. C* **80**, 989 (2020).
- [27] C. Duhr, A. Huss, A. Mazeliauskas, and R. Szafron, *J. High Energy Phys.* **09** (2021) 122.
- [28] D. d'Enterria, K. J. Eskola, I. Helenius, and H. Paukkunen, *Nucl. Phys.* **B883**, 615 (2014).
- [29] D. De Florian, G. A. Lucero, R. Sassot, M. Stratmann, and W. Vogelsang, *Phys. Rev. D* **100**, 114027 (2019).
- [30] I. Borsa, G. Lucero, R. Sassot, E. C. Aschenauer, and A. S. Nunes, *Phys. Rev. D* **102**, 094018 (2020).
- [31] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, *Eur. Phys. J. C* **81**, 341 (2021).
- [32] L. N. Epele, H. Fanchiotti, C. A. Garcia Canal, and R. Sassot, *Phys. Lett. B* **275**, 155 (1992).
- [33] A. Airapetian *et al.* (HERMES Collaboration), *Phys. Rev. D* **87**, 074029 (2013).
- [34] Further results and the set of replicas of our FFs are available upon request from the authors.
- [35] H. Aihara *et al.* (TPC/TWO GAMMA Collaboration), *Phys. Lett. B* **184**, 299 (1987); *Phys. Rev. Lett.* **61**, 1263 (1988); X.-Q. Lu, Ph.D. thesis, Johns Hopkins University, UMI-87-07273, 1987.
- [36] W. Braunschweig *et al.* (TASSO Collaboration), *Z. Phys. C* **42**, 189 (1989).
- [37] K. Abe *et al.* (SLD Collaboration), *Phys. Rev. D* **59**, 052001 (1999).
- [38] D. Buskulic *et al.* (ALEPH Collaboration), *Z. Phys. C* **66**, 355 (1995).
- [39] P. Abreu *et al.* (DELPHI Collaboration), *Eur. Phys. J. C* **5**, 585 (1998).
- [40] R. Akers *et al.* (OPAL Collaboration), *Z. Phys. C* **63**, 181 (1994).
- [41] I. Arsene *et al.* (BRAHMS Collaboration), *Phys. Rev. Lett.* **98**, 252001 (2007).