Parton-to-Pion Fragmentation Reloaded

Daniel de Florian[*] and R. Sassot[†]

Departamento de Física and IFIBA, Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 1 (1428) Buenos Aires, Argentina

Manuel Epele[‡]

Instituto de Física La Plata, CONICET - UNLP,
Departamento de Física, Facultad de Ciencias Exactas,
Universidad de La Plata, C.C. 69, La Plata, Argentina

Roger J. Hernández-Pinto[§]

Departamento de Física and IFIBA, Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 1 (1428) Buenos Aires, Argentina and
Instituto de Física Corpuscular, Universitat de València-Consejo Superior de Investigaciones Científicas,
Parc Científic, E-46980 Paterna, Valencia, Spain

Marco Stratmann[¶]

Institute for Theoretical Physics, University of Tübingen,
Auf der Morgenstelle 14, 72076 Tübingen, Germany

We present a new, comprehensive global analysis of parton-to-pion fragmentation functions at next-to-leading order accuracy in QCD. The obtained results are based on the latest experimental information on single-inclusive pion production in electron-positron annihilation, lepton-nucleon deep-inelastic scattering, and proton-proton collisions. An excellent description of all data sets is achieved, and the remaining uncertainties in parton-to-pion fragmentation functions are estimated based on the Hessian method. Extensive comparisons to the results from our previous global analysis are performed.

PACS numbers: 13.87.Fh, 13.85.Ni, 12.38.Bx

I. INTRODUCTION AND MOTIVATION

The quantitative description of hard scattering processes involving identified light hadrons in the final-state requires a precise knowledge of how quarks and gluons hadronize. In the framework of perturbative QCD (pQCD), which we pursue in the following, this vital information is encoded in parton-to-hadron fragmentation functions (FFs) [1]. To match the increasing amount and precision of experimental results, the availability of reliable sets of FFs for a large variety of hadrons, in particular, for neutral and charged pions and kaons, as well as accurate estimates of their uncertainties is of the utmost relevance and the subject of this study.

Like parton distribution functions (PDFs), FFs are required in a pQCD calculation to consistently absorb certain classes of collinear parton-parton configurations related to long-distance physics, i.e., interactions happening a long time after the actual hard scattering process. As such, FFs are non-perturbative quantities, and any information about them needs to be gathered from data, preferably in a global QCD analysis combining results obtained in a large variety of processes. These fits of FFs are facilitated by assuming factorization [2], which allows one to compute the relevant short-distance hard scattering matrix elements perturbatively, and the fact that pQCD predicts the scale evolution of FFs very much in the same way as for PDFs. Fragmentation functions depend on the parton of flavor i which hadronizes, the fraction z of its four-momentum taken by the observed hadron H, and the scale Q at which they are probed in a hard-scattering process. In what follows, they will be denoted as $D_i^H(z, Q^2)$. All relevant ingredients for a global QCD analysis of FFs are fully known up to next-to-leading order (NLO) accuracy in the strong coupling α_s comprising the kernels governing the time-like scale evolution [3, 4], single-inclusive hadron production in electron-positron annihilation (SIA) [5–7] and proton-proton (pp) collisions [8], and hadron multiplicities in semi-inclusive deep-inelastic lepton-nucleon scattering (SIDIS) [7, 9].

A first global QCD analysis of data collected for all these hadron production processes in terms of pion (and kaon) FFs has been performed quite some time ago in Ref. [10], commonly known as "DSS analysis", followed by similar fits for eta mesons, protons, and unidentified charged hadrons [11, 12]. Compared to earlier studies based on SIA data [13, 14], the DSS fit [10] fully exploited the synergy of the complementary probes of FFs

[*]Electronic address: deflo@df.uba.ar
[†]Electronic address: sassot@df.uba.ar
[‡]Electronic address: manuepele@gmail.com
[§]Electronic address: roger@df.uba.ar
[¶]Electronic address: marco.stratmann@uni-tuebingen.de

(SIA, SIDIS, and pp) to arrive at a more detailed and data-driven separation of the individual FFs for different quark flavors than before. The use of charge separated, i.e., $H = \pi^{\pm}, K^{\pm}$, SIDIS multiplicities [15] was instrumental in this respect as they weight quite differently contributions of the various quark flavors i in the hadronization process. While SIA data are more straightforward to analyze and free of PDF uncertainties, the LEP [16–18] and SLAC [19] data used in the DSS fit only provided a handle on flavor-separated FFs when supplemented by corresponding flavor-tagged results [17, 19, 20], which have no unambiguous theoretical description [10, 13] and heavily rely on Monte Carlo simulations to extract them experimentally [20]. One peculiar finding of the DSS global analysis [10] was an unexpectedly large charge symmetry violation between the total u- and d-quark FFs for pions, within sizable uncertainties though. In addition, single-inclusive pion data from pp collisions at BNL-RHIC [21] provided a first constraint on the gluon-to-pion FFs, which, at that time, was impossible to determine otherwise as precise enough SIA data were only available from the LEP and SLAC experiments, i.e., at a fixed scale M_Z, the mass of the Z boson.

To implement the lengthy, exact NLO expressions for hadron production processes in SIDIS and pp collisions without any approximations into the theoretical framework of a global fit, DSS adopted the Mellin technique [22, 23]. The gist of this method is to pre-calculate all time-consuming NLO expressions for SIDIS and pp processes once, before the actual fit is performed, and to store the required information on look-up tables; for details, see [10, 23, 24]. The use of Mellin moments is also most appropriate in solving the QCD evolution equations.

Uncertainties of the extracted FFs were estimated based on the robust Lagrange multiplier (LM) technique [25], but only for a specific moment of the $D_i^H(z, Q^2)$ contributing to the momentum sum rule. More detailed studies were performed in Ref. [26], where, in addition, the applicability of the standard iterative Hessian (IH) approach [27] was explored. Comparisons with the results obtained with the LM technique revealed, however, some limitations of the IH method, mainly due to the lack of sufficiently precise experimental information at that time to warrant the assumption that any deviations from the optimum fit are quadratic in all the parameters specifying the FFs.

In the present paper we build upon the theoretical and conceptual framework developed for the DSS analysis [10] but make use of a wealth of newly available data sets, which will enable us to relax and scrutinize some of the constraints imposed on the parameter space in the DSS fit. The key assets of the new analysis are the recently published precise SIA data from BaBar [28] and Belle [29], which, in principle, should provide a novel handle on the gluon FF through QCD scaling violations of the SIA structure functions between the scale $Q = M_Z$, rel-

evant for the LEP and SLAC experiments, and the scale corresponding to the center-of-mass system (c.m.s.) energy of BaBar and Belle, $Q = \sqrt{S} \simeq 10.5\,\mathrm{GeV}$. In addition, since the electroweak couplings of up-type and down-type quarks to the Z boson become almost equal at $Q \approx M_Z$, LEP and SLAC data are mainly sensitive to the total quark singlet FF for any observed hadron H. At the lower $\sqrt{S}$ of BaBar and Belle, the quark-antiquark pairs in SIA are weighted according to their electrical charge, which in our global fit should allow for some partial flavor separation of FFs.

Another important and new ingredient to the current analysis are the final SIDIS data released by the Hermes Collaboration [30] which supersede the preliminary and much less precise data utilized in the DSS fit [15]. New, still preliminary data for pion multiplicities in SIDIS are also available, for the first time, from the Compass experiment at CERN [31], which are very precise despite exhibiting a fine binning in the relevant kinematic variables. Finally, first results on single-inclusive pion spectra at high transverse momenta p_T have become available from the LHC at c.m.s. energies of up to $7\,\mathrm{TeV}$ [32], which supplement the data from BNL-RHIC taken at $\sqrt{S} = 200\,\mathrm{GeV}$ that have been already used in the original DSS analysis [21]. We also include several recent results from the Star Collaboration for both neutral and charged pion production at $\sqrt{S} = 200\,\mathrm{GeV}$ [33–36]. We note that at variance with the original DSS analysis, we now determine the optimum normalization shifts for each data set in the fit analytically (see, e.g. Ref. [25] for a discussion of normalization shifts in PDF fits), which greatly facilitates the global fitting procedure.

The main goal of our new analysis is to extract an updated set of parton-to-pion FFs and to determine their uncertainties reliably based on the IH method [27] in light of all the newly available, precise experimental results in SIA, SIDIS, and pp collisions. This will allow us to scrutinize the consistency of the information on FFs extracted across the different hard scattering processes, i.e., to validate the fundamental notion of universality, which is at the heart of any pQCD calculation based on the factorization of short- and long-distance physics [2] sketched above.

For the time being, we have to limit ourselves to pion FFs as a similar level of improvements on the available experimental information is still lacking for kaons, most noticeable for the SIDIS process, which is crucial in determining flavor-separated FFs. Nevertheless, we strongly believe that our updated global analysis of parton-to-pion FFs is very timely for the reasons mentioned above and the fact that precise FFs are in high demand as input for global analyses of helicity PDFs [24, 37] and transverse momentum dependent PDFs [38], both of which heavily draw on data with identified pions in the final-state. Other applications involve to quantify and understand possible modifications of hadron production yields in the presence of a nuclear medium, as studied in heavy ion collisions both at RHIC and the LHC [39].

Since extractions of leading order (LO) FFs have yielded a much less satisfactory description of the available pion production data in the DSS analysis [10], we only perform our global QCD fit at NLO accuracy. In any case, the need for LO FFs (and PDFs) has greatly diminished in recent years with the advent of novel theoretical tools that allow one to compute NLO cross sections largely automatically. The obtained optimum NLO parton-to-pion FFs, including the Hessian eigenvector sets, are available upon request and enable one to straightforwardly propagate our obtained uncertainties to any observable of interest.

It should be noted that the necessary time-like evolution kernels for FFs are available even at next-to-NLO (NNLO) accuracy now, with the exception of one, presumably minor, detail for phenomenological applications [40]. However, the corresponding partonic hard scattering processes have been only computed for SIA so far [41]. Nevertheless, it might be an interesting future endeavor to perform a NNLO analysis of SIA data alone and, perhaps, to investigate the impact of also available all-order resummations of potentially large logarithmic corrections near the partonic threshold [42]. This is, however, well beyond the scope of the current analysis.

The remainder of the paper is organized as follows: in the next Section we briefly summarize the main aspects of our updated global analysis, including the choice for the functional form used to parametrize the FFs at the initial scale for the QCD evolution, the selection of data sets and cuts imposed on them, and the treatment of experimental normalization uncertainties. The outcome of the new fit is discussed in depth in Sec. III. The obtained parton-to-pion fragmentation functions and their uncertainties are shown and compared to the results of our previous global analysis. Detailed comparisons to the individual data sets are given to demonstrate the quality of the fit. Potential open issues and tensions among the different data sets will be discussed. We briefly summarize the main results in Sec. IV.

II. FRAMEWORK

In this Section we lay out the framework and key ingredients for our global QCD analysis of parton-to-pion FFs. We mainly focus on those aspects that differ from the original DSS analysis [10].

A. Functional Form and Fit Parameters

The functional form adopted in the DSS global analysis [10] is flexible enough to accommodate also the wealth of new experimental information included in the present fit. Therefore, we continue to parametrize the hadronization of a parton of flavor i into a positively charged pion at

an initial scale of $Q_0 = 1\,\text{GeV}$ as

$$D_i^{\pi^+}(z, Q_0) = \frac{N_i z^{\alpha_i}(1-z)^{\beta_i}[1+\gamma_i(1-z)^{\delta_i}]}{B[2+\alpha_i, \beta_i+1] + \gamma_i B[2+\alpha_i, \beta_i+\delta_i+1]} .$$

Here, $B[a,b]$ denotes the Euler Beta-function, and the N_i in (1) are chosen in such a way that they represent the contribution of $zD_i^{\pi^+}$ to the momentum sum rule.

Compared to our previous analysis, the improved experimental information now allows us to impose less constraints on the parameter space spanned by the input function in Eq. (1). More specifically, as before we still have to assume isospin symmetry for the unfavored FFs of light sea quarks, i.e.,

$$D_{\bar{u}}^{\pi^+} = D_d^{\pi^+} ,$$

and we need to relate the total u-quark and d-quark FFs by a global, z-independent factor $N_{d+\bar{d}}$,

$$D_{d+\bar{d}}^{\pi^+} = N_{d+\bar{d}} D_{u+\bar{u}}^{\pi^+},$$

which quantifies any charge symmetry violation found in the fit. The fragmentation of a strange quark into a pion is now related to the unfavored FFs in Eq. (2) by

$$D_s^{\pi^+} = D_{\bar{s}}^{\pi^+} = N_s z^{\alpha_s} D_{\bar{u}}^{\pi^+}$$

rather than just using a constant as in the DSS analysis.

The charm- and bottom-to-pion FFs no longer assume $\gamma_c = \gamma_b = 0$ in Eq. (1) but can now exploit the full flexibility of the ansatz. This is not due to new flavor-tagged data but helps the global fit to accommodate the recent, very precise results from BaBar [28] and Belle [29] in SIA and from Compass [31] and Hermes [30] in SIDIS, which now constrain both the total quark fragmentation, i.e., summed over all flavors, and the individual flavor-separated, light quark FFs much better than before. As in the DSS and all other analyses [10–14], we include heavy flavor FFs discontinuously as massless partons in the QCD scale evolution above their $\overline{\text{MS}}$ "thresholds", $Q = m_{c,b}$, with m_c and m_b denoting the mass of the charm and bottom quark, respectively. Conceptually, due to confinement, there has to be a heavy quark FFs present as soon as the heavy quark can be produced in the final-state of a hard scattering process. We leave it to dedicated future studies to explore and incorporate an improved theoretical framework for heavy quark-to-light hadron fragmentation functions into the global fitting procedure, following the rather elaborate schemes that have been developed for heavy quark parton densities [43] to properly include mass effects near threshold and to resum potentially large logarithms $\sim \ln m_{c,b}^2/Q^2$ for $Q^2 \gg m_{c,b}^2$. We note that a dynamical, parameter-free generation of the heavy flavor component to light meson fragmentation functions has been developed, for instance, in Ref. [44].

In total we now have 28 free fit parameters describing our updated FFs for quarks, antiquarks, and gluons

into positively charged pions, which are determined from data by a standard χ^2 minimization to be described below. The corresponding FFs for negatively charged pions are obtained by charge conjugation and those for neutral pions by assuming $D_i^{\pi^0} = [D_i^{\pi^+} + D_i^{\pi^-}]/2$. We note that none of the constraints imposed on the fit through Eqs. (2)-(4) has any impact on its overall quality.

B. Data Selection

We make use of all the currently available experimental information on single-inclusive charged and neutral pion production in SIA, SIDIS, and hadron-hadron collisions to determine the free fit parameters defined in Sec. II A.

Compared to the data sets already used in the DSS global analysis [10], we include the new results from BaBar [28] and Belle [29] in SIA at a c.m.s. energy of $\sqrt{S} \simeq 10.5\,\text{GeV}$. Both sets are very precise, with relative uncertainties of about $2-3\%$, and reach all the way up to pion momentum fractions z close to one, well beyond of what has been measured so far. We analyze both sets with $n_f = 4$ active, massless flavors using the standard expression for the NLO SIA cross section [5]. As customary, we limit ourselves to data with $z \geq 0.1$ to avoid any potential impact from kinematical regions where finite, but neglected, hadron mass corrections, proportional to $M_\pi/(Sz^2)$, might become of any importance [10, 13, 14]. For SIA data taken at higher $\sqrt{S}$ we use $n_f = 5$ and $z > 0.05$, following the original DSS analysis. Any incompatibility of the two new precise sets of data at $\sqrt{S} \simeq 10.5\,\text{GeV}$ with each other or with the old LEP and SLAC data at $\sqrt{S} \simeq 91.2\,\text{GeV}$ [16–19] has the potential to seriously spoil the quality of the global fit.

In case of SIDIS, we replace the preliminary multiplicity data from Hermes [15] by their recently released final results [30]. More specifically, we use the data for charged pion multiplicities as a function of momentum transfer Q^2 in four bins of z taken on both a proton and a deuteron target. The range of average values of Q^2 covered by the data is from about $1.1\,\text{GeV}^2$ to $7.4\,\text{GeV}^2$ and $0.2 \leq z \leq 0.8$. In addition, we include the still preliminary multiplicity data for $\pi^\pm$ from the Compass Collaboration [31], which are given as a function of z in bins of Q^2 and the initial-state momentum fraction x. The coverage in z is the same as for the Hermes data, but due to the higher $\sqrt{S}$ of the Compass experiment the reach in x and Q^2 is wider. Experimental information is available for $0.004 \leq x \leq 0.7$ and $1.2 \leq Q^2 \leq 22.4\,\text{GeV}^2$. We do not have to impose any cuts on both data sets to accommodate them in the global analysis. As for SIA, having now available two precise sets of multiplicity data in SIDIS, covering slightly different but partially overlapping kinematics, makes it very important to validate their consistency in a global fit.

Finally, we add a couple of new sets of data for inclusive high-p_T pion production in pp collisions to the results from the Phenix experiment [21] already included in the

DSS analysis. Most noteworthy are the first results for neutral pions from the Alice Collaboration at CERN-LHC [32], covering unprecedented c.m.s. energies of up to 7 TeV. In addition, we add Star data taken at $\sqrt{S} = 200\,\text{GeV}$ in various rapidity intervals for both neutral and charged pion production and for the π^-/π^+ ratio [33–36]. As we will demonstrate and discuss in more detail in Sec. III D below, it turns out that a good global fit of RHIC and LHC pp data, along with all the other world data, can only be achieved if one imposes a cut on the minimum p_T of the produced pion of about 5 GeV. Such a cut eliminates some of the pp data points included in the previous DSS analysis from the fit, in particular, all the Brahms [45] and Star [46] data at forward pseudo-rapidities and, hence, too small values of p_T.

C. Fit Procedure and Uncertainty Estimates

The 28 free parameters describing the updated parton-to-pion FFs in Eq. (1) at the chosen input scale of 1 GeV are again determined from a standard χ^2 minimization where

$$\chi^2 = \sum_{i=1}^{N} \left[\left(\frac{1 - \mathcal{N}_i}{\delta \mathcal{N}_i} \right)^2 + \sum_{j=1}^{N_i} \frac{(\mathcal{N}_i T_j - E_j)^2}{\delta E_j^2} \right], \qquad (5)$$

for $i = 1, \ldots, N$ data sets, each contributing with N_i data points. E_j is the measured value of a given observable, δE_j the error associated with this measurement, and T_j is the corresponding theoretical estimate for a given set of parameters in Eq. (1). Since the full error correlation matrices are not available for some of the data sets used in the fit, statistical and systematical errors are simply added in quadrature in δE_j as in all previous fits [10, 11, 13, 14].

At variance with the original DSS fit [10], where we have introduced several extra fit parameters to account for experimental normalization uncertainties $\mathcal{N}_i$ in (5), we now derive the optimum normalization shifts for each data set analytically from the condition $\partial \chi^2 / \partial \mathcal{N}_i = 0$, which yields

$$\mathcal{N}_i = \frac{\sum_{j=1}^{N_i} \frac{\delta \mathcal{N}_i^2}{\delta E_j^2} T_j E_j + 1}{1 + \sum_{j=1}^{N_i} \frac{\delta \mathcal{N}_i^2}{\delta E_j^2} T_j^2} . \qquad (6)$$

Here, $\delta \mathcal{N}_i$ denotes the quoted experimental normalization uncertainty for data set i. In Sec. III A we will list the so obtained normalizations $\mathcal{N}_i$ along with the individual χ^2 values for each data set included in the fit.

In the DSS analysis [10] we assessed uncertainties in the extraction of fragmentation functions with the help of the LM technique [25] by mapping out the maximum allowed range of variation in the fit of the truncated second moments of the fragmentation functions

$$\eta_i^{\pi^+}(x_{\min}, Q^2) \equiv \int_{x_{\min}}^{1} z D_i^{\pi^+}(z, Q^2) dz, \qquad (7)$$

for $x_{\min} = 0.2$ and $Q = 5\,\mathrm{GeV}$. While this method is very robust, even when some of the fit parameters are only loosely constrained by data, it has the disadvantage that uncertainties cannot be easily propagated to other observables of interest. In Ref. [26] we have therefore explored the applicability of the iterative Hessian approach [27] based on the original DSS choice of data sets, cuts, and parameters by comparing its outcome to uncertainty estimates obtained with the LM method. The main idea of the IH method is to assume a quadratic behavior of the χ^2 hyper-surface of parameter displacements and to express the χ^2 increment from its minimum value in terms of combinations of fit parameters that maximize the variation. Such an eigenvector representation of the Hessian matrix proves to be extremely suitable to compute the propagation of uncertainties to arbitrary observables in terms of a limited number of pre-calculated sets of FF functions (in fact, twice the amount of fit parameters). These sets correspond to fixed displacements along the eigenvector directions of the Hessian matrix.

With the much increased availability of precise data for the current analysis, the sole use of the computationally less demanding IH method to quantify uncertainties of FFs becomes viable and will be pursued in the following. The obtained eigenvector sets of FFs will be made available upon request from the authors along with a parametrization of the optimum fit. To define the eigenvector sets one has to choose a tolerance parameter $\Delta\chi^2$ for the increment in χ^2 which is still acceptable in the global fit. Here we proceed as follows: the tolerances for the eigenvector sets corresponding to 68% and 90% confidence level (C.L.) intervals are determined from the Gaussian probability density function for a χ^2 distribution with k degrees of freedom (d.o.f.):

$$P_k(x) = \frac{x^{k/2-1}e^{-x/2}}{\Gamma(k/2)2^{k/2}} \ . \tag{8}$$

The $\Delta\chi^2$ related to the 68th and 90th percentiles are then obtained by solving $\int_0^{\chi^2+\Delta\chi^2} d\chi^2 P_k(\chi^2) = 0.68$ and 0.90, respectively.

Finally, we choose the NLO set of PDFs from the MSTW group [47] and the corresponding uncertainty estimates in computations of the SIDIS and pp cross sections. For consistency, we also fix the strong coupling α_s to the values obtained in the MSTW fit. We note that in the x and Q^2 region relevant for our global analysis of FFs, the needed combinations of PDFs are relatively well constrained. A choice of PDFs other than the MSTW set would not alter the outcome of our fit in any significant way. We will illustrate in Sec. III D below that theoretical scale ambiguities are considerably larger than PDF uncertainties.

III. RESULTS

In this Section we present and discuss in depth the results of our global analysis of parton-to-pion FFs. First,

we present the obtained fit parameters, normalization shifts, and individual χ^2 values. Next, the obtained $D_i^{\pi^+}(z, Q^2)$ and their uncertainties are shown and compared to the results of the DSS fit. The quality of the fit to SIA, SIDIS, and pp data and potential open issues and tensions among the different sets of data are illustrated and discussed in Sec. III B, III C, and III D, respectively.

A. Parton-To-Pion Fragmentation Functions

TABLE I: Parameters describing the NLO FFs for positively charged pions, $D_i^{\pi^+}(z, Q_0)$, in Eq. (1) in the $\overline{\mathrm{MS}}$ scheme at the input scale $Q_0 = 1\,\mathrm{GeV}$. Results for the charm and bottom FFs refer to $Q_0 = m_c = 1.43\,\mathrm{GeV}$ and $Q_0 = m_b = 4.3\,\mathrm{GeV}$, respectively.

flavor i	N_i	α_i	β_i	γ_i	δ_i
$u + \overline{u}$	0.4465	-0.455	0.912	8.00	4.14
$d + \overline{d}$	0.4471	-0.455	0.912	8.00	4.14
$\overline{u} = d$	0.127	0.997	2.884	31.48	7.70
$s + \overline{s}$	0.378	0.546	2.884	31.48	7.70
$c + \overline{c}$	0.348	0.555	4.883	7.62	7.64
$b + \overline{b}$	0.401	-0.262	4.369	20.80	11.72
g	0.216	1.542	4.066	86.34	19.32

In Table I we list the obtained set of parameters specifying our updated, optimum parton-to-pion fragmentation functions at NLO accuracy at the input scale $Q_0 = 1\,\mathrm{GeV}$ for the light quark flavors and the gluon, and at their respective thresholds $Q_0 = m_{c,b}$ for the charm and bottom quarks.

Table I reveals already a notable difference to one of the findings of the DSS analysis which preferred an unexpectedly sizable breaking of the charge symmetry between $u + \bar{u}$ and $d + \bar{d}$ FFs of about 10% [10], within large uncertainties though. This was mainly driven by the preliminary $\pi^\pm$ multiplicities from HERMES [15] used in the fit at that time. Now, with much improved experimental information on charged pion multiplicities both from HERMES [30] and COMPASS [31] and new data on the ratio π^-/π^+ in pp collisions from STAR [35], the parameter $N_{d+\bar{d}}$ in Eq. (3) prefers to stay very close to unity, i.e., very little or no breaking.

As has been mentioned above, in case of the unfavored FFs, data now allow us to introduce some non-trivial z dependence, see Eq. (4), to parametrize a potential SU(3) breaking between the $\bar{u} = d$ and $s = \bar{s}$ FFs. This little extra freedom not only helps to accommodate all the different data sets used in the global analysis in a much better way, but it also leads to more realistic uncertainty estimates for both $D_s^{\pi^+}$ and $D_{\bar{u}}^{\pi^+}$. In particular, it now turns out that the uncertainties for $D_s^{\pi^+}$ are much bigger than for $D_{\bar{u}}^{\pi^+}$ as can be inferred from Fig. 1, where we present the individual parton-to-pion FFs $D_i^{\pi^+}(z, Q^2)$ at $Q^2 = 10\,\mathrm{GeV}^2$. The four leftmost panels show the opti-
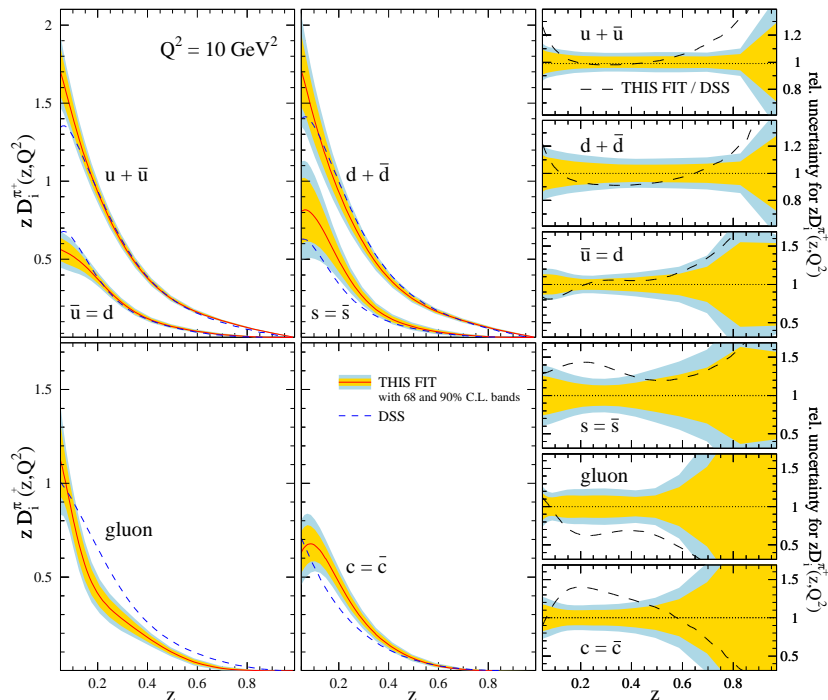
FIG. 1: The individual FFs for positively charged pions $zD_i^{\pi^+}(z,Q^2)$ at $Q^2 = 10\,\mathrm{GeV}^2$ along with uncertainty estimates at 68% and 90% C.L. indicated by the inner and outer shaded bands, respectively. The panels on the right-hand-side show the corresponding relative uncertainties. Also shown is a comparison to the previous global analysis by DSS [10] (dashed lines).

mum $zD_i^{\pi^+}$ at NLO accuracy for $i = u + \bar{u}$, $d + \bar{d}$, $\bar{u} = d$, $s = \bar{s}$, $c = \bar{c}$, and the gluon g (solid lines) along with our uncertainty estimates at 68% C.L. (inner bands) and 90% C.L. (outer bands), obtained as described in Sec. II C. For better visibility, the rightmost panels give the relative uncertainties for the same set of $zD_i^{\pi^+}$. The results of the previous NLO DSS fit are shown as dashed lines.

As can be inferred from Fig. 1, for the light quark flavors the old DSS results are either close to the updated fit or within its 90% C.L. uncertainty band. The best determined pion FFs is $D_{u+\bar{u}}^{\pi^+}$, where the relative uncertainties are below 10% at 90% C.L. throughout most of the relevant z range. Only for $z \gtrsim 0.8$ the errors rapidly increase because of the lack of experimental constraints in this region. The corresponding uncertainties for $D_{d+\bar{d}}^{\pi^+}$ turn out to be slightly larger as they also include possible violations of SU(2) charge symmetry through Eq. (3). We stress again, that at variance with the DSS analysis [10], the new fit does not favor any SU(2) breaking. For the unfavored FFs, $D_{\bar{u}}^{\pi^+} = D_d^{\pi^+}$ are determined well in a much more limited range of z, and uncertainties start to increase already for $z \gtrsim 0.5$. The corresponding ambiguities on $D_s^{\pi^+} = D_{\bar{s}}^{\pi^+}$ are about a factor of two larger and amount to at least 25% at 90% C.L. for $z \simeq 0.3$.

Bigger deviations from the DSS analysis are found for both the gluon and the charm FFs. In the latter case, this is driven by the greater flexibility of the functional form, five fit parameters rather than three, which helps

with the overall quality of the global fit and cannot be pin-pointed to a particular data set. In fact, there had been no new charm (or bottom) tagged data since the LEP and SLAC era. The significantly reduced $D_g^{\pi^+}$ as compared to the DSS fit is a result of the new ALICE pp data [32], which have a strong preference for less pions from gluon fragmentation for basically all values of z. We will discuss this finding, and possible tensions arising with the pp data from RHIC, in more detail in Sec. III D. The relative uncertainties on $D_g^{\pi^+}$ at $Q^2 = 10\,\mathrm{GeV}^2$ are about 20% at 90% C.L. up to $z \simeq 0.5$ and quickly increase towards larger z values.

We refrain from performing a detailed comparison to the uncertainty estimates based on the data sets available for the original DSS analysis [10, 26] as they can be viewed at best as a rough approximation. Only with the quality and variety of data sets available for the current global analysis one can arrive at a first meaningful determination of uncertainties for parton-to-pion FFs, which therefore constitutes as one of the main results of this study.

We note that the new very precise SIA data from BABAR [28] and BELLE [29] help to reliably constrain light quark FFs to much higher values of z than before, in particular, $D_{u+\bar{u}}^{\pi^+}$. In combination with the LEP and SLAC data, which, at $Q^2 = M_Z^2$, mainly constrain the total quark singlet fragmentation function, the new precise data at $\sqrt{S} \simeq 10.5\,\mathrm{GeV}$ also help to provide some partial
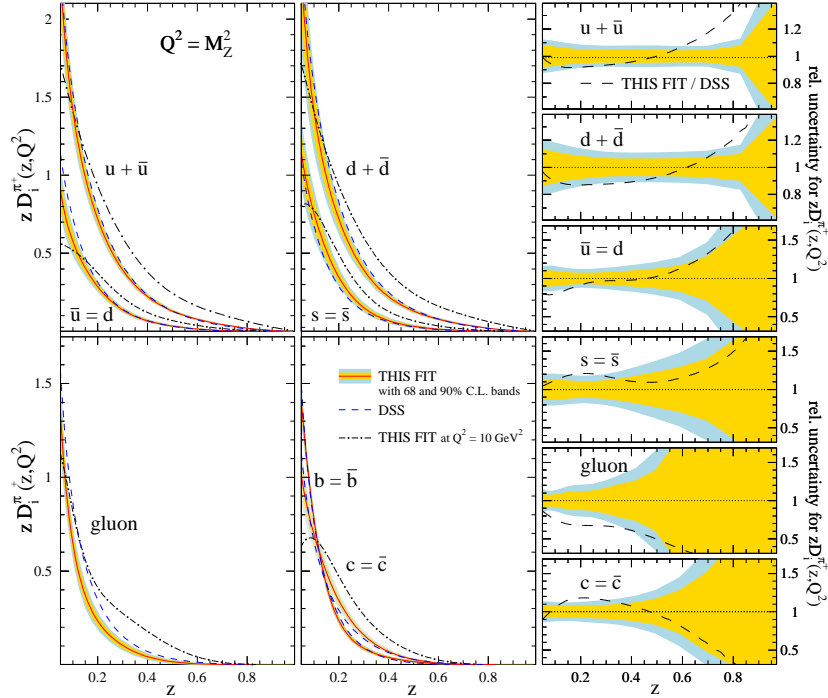
FIG. 2: As in Fig. 1 but now for $Q^2 = M_Z^2$. Note that here also a bottom-to-π^+ fragmentation function is shown.

flavor separation from SIA data alone, as they are sensitive to the electrical charge weighted sum of quark FFs. Multiplicities in SIDIS for identified charged pions provide further invaluable experimental input to address this question, see Sec. III C below. In case of $D_g^{\pi^+}$, for the first time, some constraint can be derived from QCD scaling violations in SIA thanks to having now available two precise sets of data at different energy scales $Q^2 \approx 110 \, \mathrm{GeV}^2$ and $Q^2 = M_Z^2$. However, scaling violations for FFs in the relevant medium-to-large z range are fairly mild and also the coverage of the LEP and SLAC data is much more sparse towards high z, which, to some extent, is reflected in the still relatively large uncertainties obtained for $D_g^{\pi^+}$ in Fig. 1.

To demonstrate the scale evolution of FFs, we show in Fig. 2 the same $zD_i^{\pi^+}$ as in Fig. 1 but now at $Q^2 = M_Z^2$. Since we are above the bottom threshold $Q = m_b$, we now include also our results for $zD_b^{\pi^+} = zD_{\bar{b}}^{\pi^+}$ in the middle panel of the lower row. To facilitate the comparison of the FFs computed at the two different scales, the dot-dashed lines in Fig. 2 repeat the results for the new, optimum fit at $Q^2 = 10 \, \mathrm{GeV}^2$ shown in Fig. 1. The FFs of the DSS fit are again denoted as dashed lines. As can be seen, evolution to larger Q^2 reduces the FFs for essentially all relevant z values above about $z \simeq 0.15$. This trend is reminiscent of the Q^2 evolution of PDFs at not to small values of x, which is not surprising as the LO evolution kernels are essentially the same for the time-like and space-like case. The increase of the FFs at small z is phenomenologically not relevant as their

range of applicability is anyhow restricted to $z \gtrsim 0.05$. The relative uncertainties, again given in the rightmost panels of Fig. 2, are largely similar to those obtained at $Q^2 = 10 \, \mathrm{GeV}^2$. Some of the sizable ambiguities at large z are pushed towards smaller z by evolution, most noticeable for $D_g^{\pi^+}$.

The overall quality of the fit is summarized in Tab. II where we list all data sets included in our global analysis, as discussed in Sec. II B, along with their individual χ^2 values and the analytically determined normalization shifts according to Eq. (6). We note that the quoted χ^2 values are based only on fitted data points, i.e., after applying the cuts mentioned in Sec. II B, and include the χ^2 penalty from the N_i, i.e., the first term in Eq. (5).

Firstly, it is worth mentioning that there is a more than twofold increase in the number of available data points as compared to the original DSS analysis [10]. Secondly, the quality of the global fit has improved dramatically from $\chi^2/\mathrm{d.o.f.} \simeq 2.2$ for DSS, see Tab. II in Ref. [10], to $\chi^2/\mathrm{d.o.f.} \simeq 1.2$ for the current fit. A more detailed comparison reveals that the individual χ^2 values for the SIA data [16–19, 48, 49], which were already included in the DSS fit, have, by and large, not changed significantly. The description of the fully flavor separated data from OPAL [20] in the fit favors a rather large normalization shift but has nevertheless deteriorated. Given that this set has only 25 data points, it is the biggest contributor to the total χ^2. However, in general, flavor-tagged results should not be taken too literally as they lack a proper interpretation and theoretical framework beyond the lowest

TABLE II: Data sets used in our NLO global analysis, their optimum normalization shifts N_i, cf. Sec. II C and Eq. (6), the individual χ^2 values (including the χ^2 penalty from the obtained N_i), and the total χ^2 of the fit.

experiment		data type	norm. N_i	# data in fit	χ^2
Tpc [48]		incl.	1.043	17	17.3
		uds tag	1.043	9	2.1
		c tag	1.043	9	5.9
		b tag	1.043	9	9.2
Tasso [49]	34 GeV	incl.	1.043	11	30.2
	44 GeV	incl.	1.043	7	22.2
Sld [19]		incl.	0.986	28	15.3
		uds tag	0.986	17	18.5
		c tag	0.986	17	16.1
		b tag	0.986	17	5.8
Aleph [16]		incl.	1.020	22	22.9
Delphi [17]		incl.	1.000	17	28.3
		uds tag	1.000	17	33.3
		b tag	1.000	17	10.6
Opal [18, 20]		incl.	1.000	21	14.0
		u tag	0.786	5	31.6
		d tag	0.786	5	33.0
		s tag	0.786	5	51.3
		c tag	0.786	5	30.4
		b tag	0.786	5	14.6
BaBar [28]		incl.	1.031	45	46.4
Belle [29]		incl.	1.044	78	44.0
Hermes [30]		π^+ (p)	0.980	32	27.8
		π^- (p)	0.980	32	47.8
		π^+ (d)	0.981	32	40.3
		π^- (d)	0.981	32	59.1
Compass [31] prel.		π^+ (d)	0.946	199	174.2
		π^- (d)	0.946	199	229.0
Phenix [21]		π^0	1.112	15	15.8
Star [33–36]	$0 \leq \eta \leq 1$	π^0	1.161	7	5.7
	$0.8 \leq \eta \leq 2.0$	π^0	0.954	7	2.7
	$|\eta| < 0.5$	$\pi^\pm$	1.071	8	4.3
	$|\eta| < 0.5$	$\pi^+,\pi^-/\pi^+$	1.006	16	17.2
Alice [32]	7 TeV	π^0	0.766	11	27.7
TOTAL:				973	1154.6

order as was already pointed out, e.g., in Refs. [10, 13].

The biggest improvement concerns the SIDIS multiplicities from Hermes which, in their recently published version [30], are now described very well by the updated fit. Also, the preliminary charged pion multiplicities from Compass [31] and the new SIA data from BaBar [28] and Belle [29] integrate nicely into the global analysis of parton-to-pion FFs.

Finally, and as we will illustrate in detail in Sec. III D below, there is some tension among the pp data sets from RHIC and the LHC, which forced us to introduce a cut $p_T > 5$ GeV on the pion's transverse momentum in the current fit to accommodate both of them. The obtained individual χ^2 values are all reasonable, as can be inferred from Tab. II, with the new Alice data [32] being on the high side, which largely stems from the penalty for the

still sizable normalization shift. This large shift reflects the preference of the new Alice data for a smaller gluon-to-pion FF than extracted by the original DSS fit based on RHIC Phenix data [21] alone. As a result of the p_T cut, the number of pp data in the fit for RHIC has decreased as compared to the DSS analysis. Both the Brahms [45] and Star [46] results at forward pseudo-rapidities do not pass the p_T cut anymore, and, hence, are excluded from the updated fit. Likewise, we do not consider the Alice data taken in pp collisions at $\sqrt{s} = 900$ GeV [32], where only a single point would survive the cut in p_T.

B. Electron-Position Annihilation Data

In Figs. 3 and 4 we present a detailed comparison of the results of our fit and its uncertainties at both 68% and 90% C.L. with the SIA data already included and newly added to the original DSS analysis [10], respectively. In general, the agreement of the fit with SIA data is excellent in the entire energy and z range covered by the experiments. For $Q^2 = M_Z^2$, the Delphi data [17] exhibit some mild tension with other sets at the same c.m.s. energy in the largest z bins, resulting in a somewhat higher individual χ^2 value, as can be gathered from Tab. II.

The Belle data [29], shown in the right-hand panel of Fig. 4, provide not only the finest binning in z but also reach the highest z values measured so far. Above $z \gtrsim 0.8$ one observes an increasing trend for the new fit to overshoot the data, but still within the estimated and growing theoretical uncertainties though. In this kinematic regime one expects large logarithmic corrections, which appear in each order of perturbation theory, to become more and more relevant. It is known how to resum such terms to all orders in the strong coupling [42], and it might be worthwhile to explore their relevance in a future dedicated analysis and whether they could further improve the agreement with data. Resummations also provide a window to non-perturbative contributions to the perturbative series so far little explored. The binning of BaBar data [28] is more sparse towards large z, and a similar trend as for the Belle data is not visible here.

For all the sets shown in Fig. 3, the new fit is able to follow the trend of the data even below the z values included in the analysis (the region indicated by the hatched area). Agreement with BaBar data below the cut $z = 0.1$ quickly deteriorates though. In this region, the data start to drop while the NLO SIA cross section continues to rise as can be seen in the left-hand panel of Fig. 4. Since the BaBar data are taken at the lowest c.m.s. energy, such an effect is not unexpected and signifies the onset of neglected hadron mass effects in the theoretical framework. In fact, this was the reason for us to choose a somewhat higher cut in z, $z > 0.1$, than for the other SIA data obtained at higher c.m.s. energies.
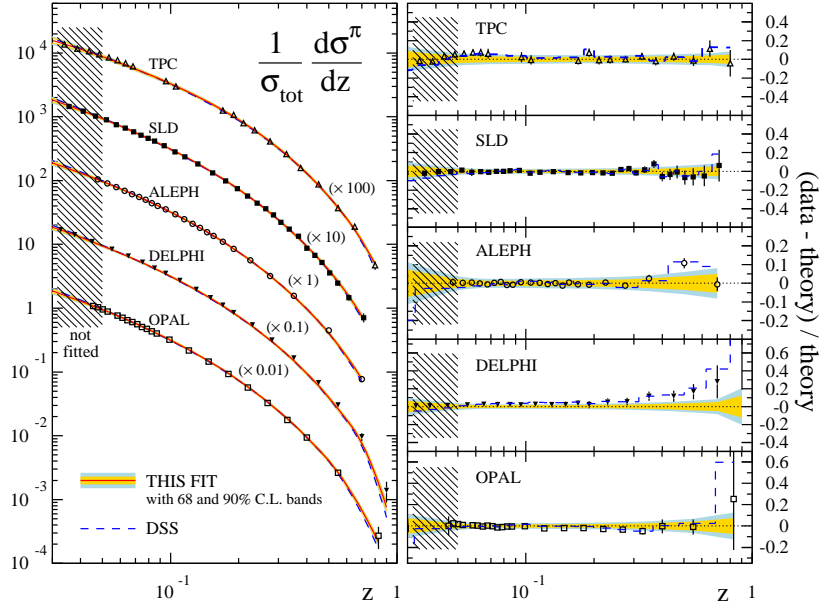
FIG. 3: Left-hand side: comparison of our new NLO results (solid lines) and the previous DSS fit [10] (dashed lines) with data sets for inclusive pion production in SIA used in both fits, see Tab. II. The inner and outer shaded bands correspond to new uncertainty estimates at 68% and 90% C.L., respectively. Right-hand side: "(data-theory)/theory" for each of the data sets w.r.t. our new fit (symbols) and the DSS analysis (dashed lines).
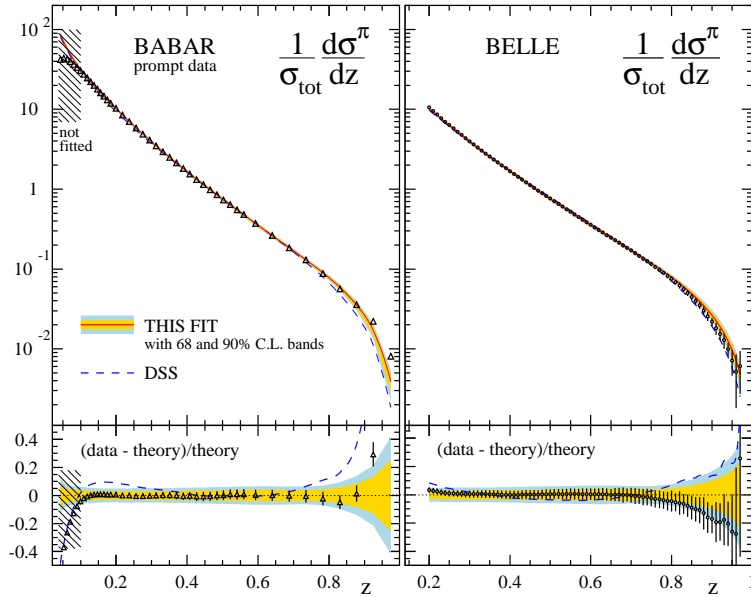

FIG. 4: Left-hand side: comparison of our new NLO results (solid line) with the new BaBar "prompt" data [28]; also shown is the result obtained with the DSS fit [10] (dashed line). Right-hand side: same, but now for the Belle data [29]. The lower panels show "(data-theory)/theory" for each of the data sets w.r.t. our new fit (symbols) and the DSS analysis (dashed lines). The inner and outer shaded bands correspond to the new uncertainty estimates at 68% and 90% C.L., respectively.

The BELLE experiment did not publish any data below $z = 0.2$ [29].

Also shown in Figs. 3 and 4 are the theoretical results obtained with the original DSS FFs (dashed lines), i.e.,
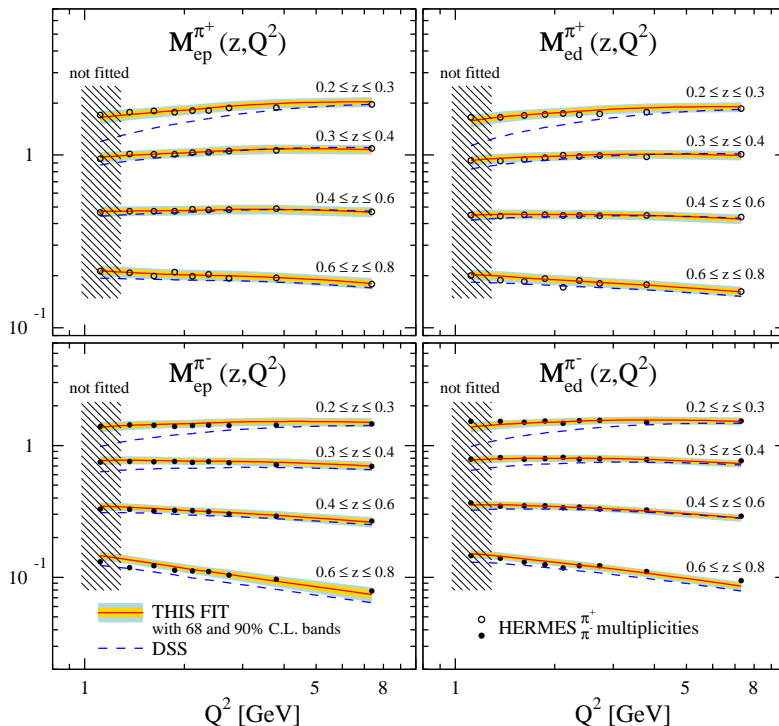
FIG. 5: Comparison of our NLO results for charged pion multiplicities in SIDIS off proton (left panels) and deuteron (right panels) targets with data from the HERMES Collaboration [30]. The inner and outer shaded bands correspond to uncertainty estimates at 68% and 90% C.L., respectively. Also shown are the results obtained with the DSS FFs (dashed lines).

without any refitting or adjusting normalization shifts. The agreement with SIA data is in general very good, except for some small deviations from the recent B factory data, most noticeable in the comparison to BABAR. Contrary to the new analysis, the original DSS fit undershoots both the BELLE and BABAR data at high z.

Our estimated uncertainty bands, also shown in Figs. 3 and 4, reflect the accuracy and kinematical coverage of the fitted data. They increase towards both small and large z, similar to the pattern observed for the individual $D_i^{\pi^+}$ in Figs. 1 and 2. One should keep in mind that the obtained bands are constrained by the fit to the global set of SIA, SIDIS, and pp data and do not necessarily have to follow the accuracy of each individual set of data.

As was already mentioned in Sec. III A, the SIA data from the LEP and SLAC experiments constrain mainly the total quark singlet fragmentation to pions as up-type and down-type quark couplings to the exchanged Z gauge boson are roughly equal at $Q \simeq M_Z$. The new BABAR and BELLE data are dominated by photon exchange and, hence, prefer up-type quark flavors. When combined, this leads to some partial flavor separation. QCD scale evolution between $Q^2 \simeq 110\,\mathrm{GeV}^2$ and $Q^2 = M_Z^2$ provides some additional constraints, in particular, also for the gluon FF. The flavor-tagged LEP and SLAC data, listed in Tab. II, are still the best "direct" source of information on the charm- and bottom-to-pion FFs.

Finally, we wish to remark that despite the excellent agreement with all SIA data there are still some issues which require further scrutiny and, perhaps, more detailed comparisons among the different experimental groups. One concern is the question to what extent "feed-down" pions from weak decays contribute to the individual data sets. Different treatments of QED radiative corrections, whose main effect is to lower the "true" c.m.s. energy $\sqrt{S}$ of the collisions, might be another source of potential tension. For instance, the BELLE Collaboration [29] provides only a measurement of the cross section $d\sigma/dz$, while all other experiments in SIA scale their quoted results by the total cross section σ_{tot} for $e^+e^- \to$ hadrons. Since BELLE cuts on radiative photon events if their energy exceeds a certain threshold, rather than attempting to unfold the radiative QED effects, one has to take this into account when normalizing the BELLE data to the conventional $1/\sigma_{\mathrm{tot}}\, d\sigma/dz$ in a global fit.

C. Semi-Inclusive DIS Multiplicities

The most powerful constraint of flavor-separated FFs comes from charged pion multiplicities in SIDIS. Contrary to SIA, which produces π^+ and π^- at equal rates, multiplicities are sensitive to the produced hadron's charge through the choice of the target hadron in DIS.
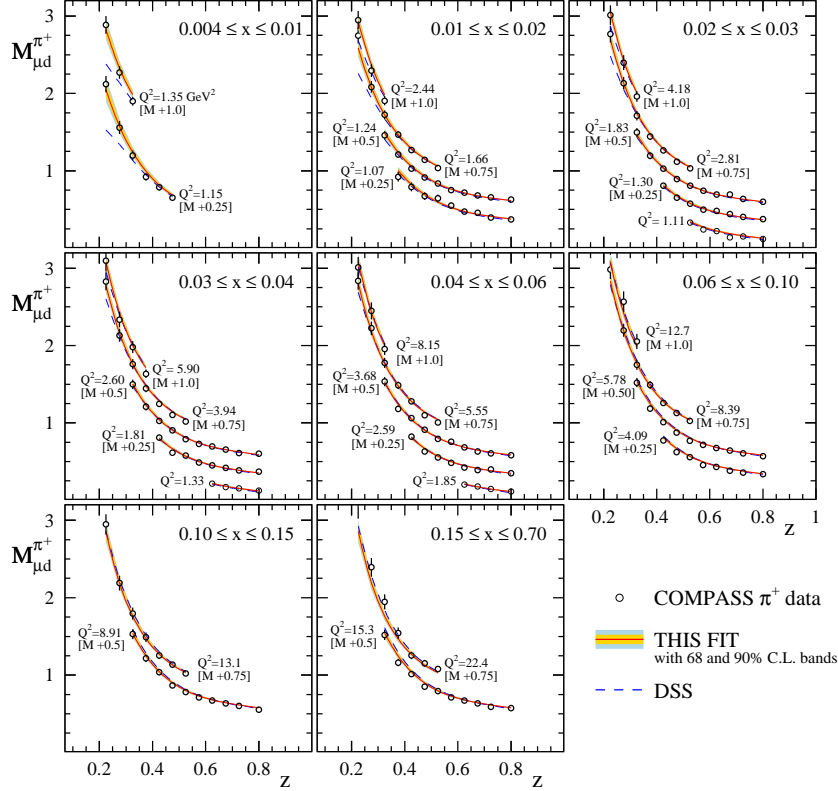
FIG. 6: Comparison of our NLO results for π^+ multiplicities in SIDIS of muons off a deuteron target with preliminary data from the COMPASS experiment [31] for various bins in x and Q^2. The inner and outer shaded bands correspond to our uncertainty estimates at 68% and 90% C.L., respectively. Also shown are the results obtained with the DSS FFs (dashed lines). As indicated in the plot, different constant factors are added to the multiplicities $M_{\mu d}^{\pi^+}$ to distinguish results for different values of Q^2 in the same x bin.

For instance, data taken on a proton target will produce more π^+ than π^-, since u-quarks are more abundant in a proton than d-quarks, and they are also preferred in their coupling to the probing virtual photon due to their larger electrical charge.

Compared to the DSS analysis, where we only had some preliminary set of pion multiplicities on a deuteron target from the HERMES Collaboration at our disposal [15], we can now use their recently published, final set of data for both proton and deuteron targets [30]. In Fig. 5 we illustrate the quality of the new fit with respect to the HERMES data. Shown are the charged pion multiplicities $M_{e,p(d)}^{\pi^\pm}$, which are defined as the ratio of the inclusive pion yield and the total DIS cross section at the same x and Q^2 values (bins) in electron-proton (ep) or electron-deuteron (ed) scattering:

$$M_{e,p(d)}^{\pi^\pm} \equiv \frac{d\sigma^{\pi^\pm}/dx\,dQ^2\,dz}{d\sigma/dx\,dQ^2} \quad . \tag{9}$$

The extraction of the FFs requires knowledge of the PDFs of the proton (deuteron) target for which we use the NLO parametrization of the MSTW Collaboration

[47] as was already mentioned above. In the fit we consider the projection of the three-dimensional multiplicity data onto the Q^2 dependence for four different bins of the pion's momentum fraction z, which is most sensitive to the quantities we are interested in, the parton-to-pion FFs. The x integrated ratio (9) is also least sensitive to the actual choice of PDFs. We use the standard Mellin technique [23] to pre-calculate look-up tables for each data point at NLO accuracy to speed up the fitting procedure and to facilitate the uncertainty analyses significantly. We recall that at NLO, the relevant hard scattering coefficient functions for SIDIS [7, 9] depend in a non-trivial way on both x and z, such that an often used naive approximation, where the x and the z dependence in Eq. (9) is assumed to completely factorize, is bound to fail. Even at LO accuracy such an assumption cannot work as soon as different quark flavors fragment differently into the observed hadron which they do for charged pions.

The agreement between the HERMES data [30] and the updated fit is remarkably better than with the preliminary results [15] used in the DSS analysis, see Fig. 4 in Ref. [10]. This is largely due to the much improved preci-
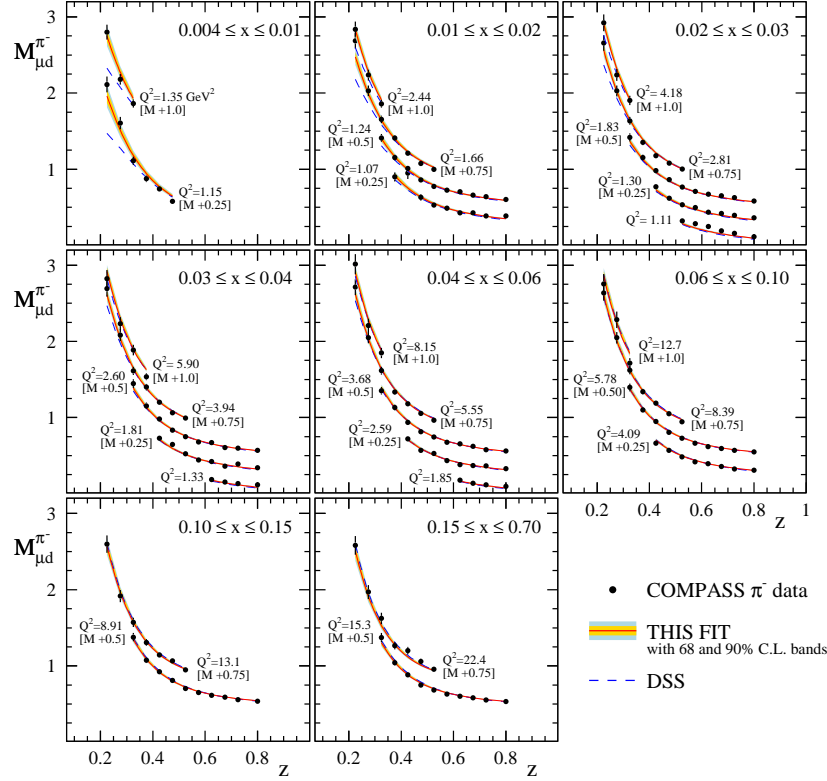
FIG. 7: As in Fig. 6 but now for π^- multiplicities.

sion of the final data, which also exhibit considerably less fluctuations from bin to bin, in particular, for π^-. This is also reflected in the total χ^2 for the HERMES data set, which reduces from 188.2 for 64 data points on a proton target in the DSS fit [10] to 175 for 128 data points in the current analysis. In Fig. 5 we compare again also to the result of a calculation based on the DSS FFs, without any re-fitting or adjusting normalizations. As can be seen, the agreement with data is not optimal, and the theory predictions fall short of the data in all bins. Most noticeable is the disagreement at $0.2 \leq z \leq 0.3$ and $0.6 \leq z \leq 0.8$ for both π^+ and π^- data. Here, the DSS result is well outside our current uncertainty estimates shown, as before, as shaded bands in Fig. 5.

The use of the HERMES multiplicity data as a means of providing a reliable flavor and charge separation for pion FFs in the DSS fit was often questioned in the past because of the smallish Q^2 values of some of the data points. New, still preliminary data from the COMPASS Collaboration [31], taken at a higher c.m.s. energy, will shed some light on the validity of using a standard, leading-twist pQCD framework at NLO accuracy [7, 9] to describe the HERMES multiplicity data for charged pions.

In the present fit we can use charged pion results from COMPASS obtained on a deuteron target [31]. More specifically, the data are presented as a function of z in 8 bins of x, each subdivided into various bins in Q^2. In total 199 data points pass our cuts for both π^+ and π^-.

The comparison of the COMPASS data to the results of our fit is presented in Figs. 6 - 8. A very satisfactory agreement is achieved in almost all bins across the entire kinematic regime covered by data, as can be best inferred from Fig. 8, where we show "(data-theory)/theory". The obtained χ^2/d.o.f. for both π^+ and π^- multiplicities is close to unity, see Tab. II, demonstrating that the low energy HERMES [30] and the COMPASS [31] data can be described simultaneously and without spoiling the agreement with SIA results. For comparison we show again theoretical results obtained with the DSS FFs (dashed lines), which also agree well with COMPASS data except for some of the bins corresponding to the lowest Q^2 values. This is in line with the observations for the HERMES data above, where the deviations with DSS were found to be largest at the smallest Q^2.

As was already mentioned in Sec. III A, the new SIDIS data now favor almost identical $u + \bar{u}$ and $d + \bar{d}$ FFs, i.e., very little or no charge symmetry breaking. This is also preferred by data on the π^-/π^+ ratio in pp collisions which we discuss next.

D. RHIC and LHC Data

The last of the three pillars of our global analysis of parton-to-pion FFs is the wealth of experimental information coming from hadron-hadron collisions, more
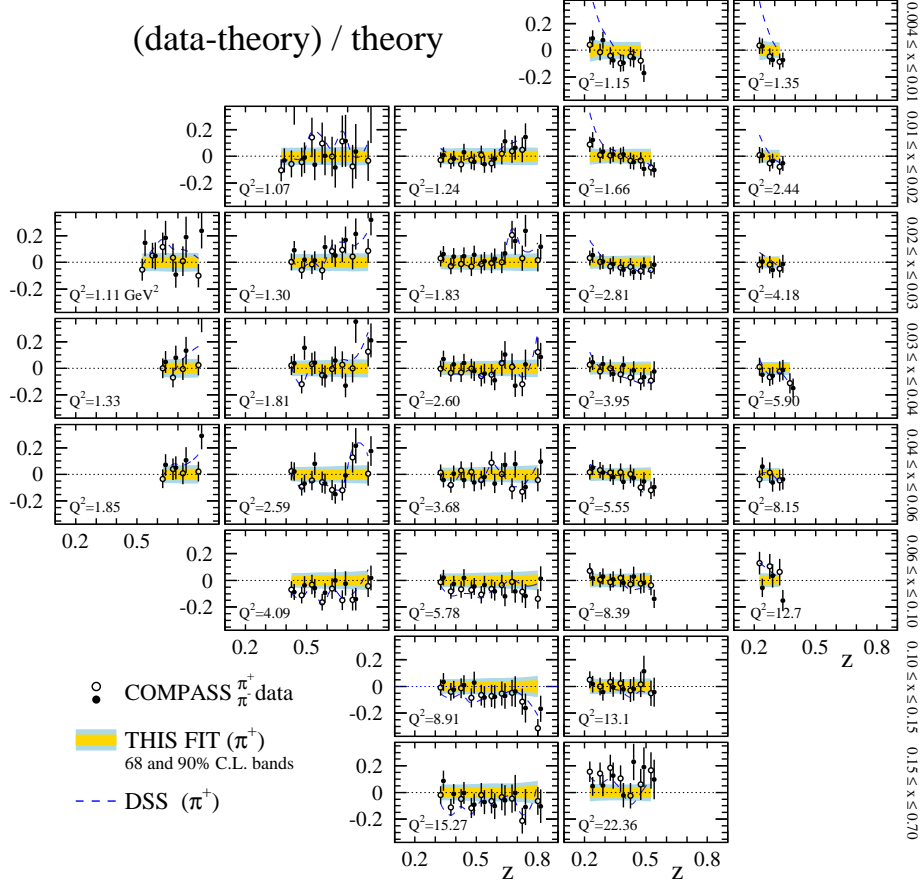
FIG. 8: As in Figs. 6 and 7 but now showing "(data-theory)/theory" for our new NLO fit (open and closed circles correspond to π^+ and π^- multiplicities, respectively) in each x and Q^2 bin. The shaded uncertainty bands and the results obtained with the DSS FFs (dashed lines) are for π^+ production.

specifically, single-inclusive high-p_T pion production in pp collisions at BNL-RHIC and CERN-LHC. Compared to the original DSS analysis [10], which mainly made use of the PHENIX data for π^0 production at mid rapidity [21], we now have, in addition, results from the STAR Collaboration for neutral and charged pions [33–36] as well as first data from the LHC [32].

Due to the complexity of the underlying hard-scattering processes at NLO accuracy [8], the use of a fast, grid-based method such as the Mellin technique, to implement the relevant expressions efficiently and without the need of any approximations is indispensable here. As in various previous analyses [10–12, 24], and for the implementation of the SIDIS multiplicities in NLO, we adopt the well-tested method based on Mellin moments as described in Ref. [23]. Since inclusive particle spectra at not too large values of p_T are dominated by gluon-induced processes in pp collisions [50], the RHIC and LHC data will provide invaluable information on the otherwise only weakly constrained gluon FF $D_g^{\pi^+}$.

One of the main results of our updated fits is to reveal a tension between the p_T spectra of neutral pions mea-

sured at the RHIC experiments and by the ALICE Collaboration. In some sense this was already anticipated by comparisons to expectations obtained with the previous DSS FFs, which are known to describe the RHIC data nicely down to $p_T \simeq 1.5\,\mathrm{GeV}$ [10, 21] but were found to grossly overshoot recent ALICE results at $\sqrt{s} = 7\,\mathrm{TeV}$ for essentially all p_T values [32, 51]. We have tried to accommodate both sets of pp data together by introducing additional freedom to our standard functional form in Eq. (1) but to no avail. In particular, at smallish p_T values, below about $5\,\mathrm{GeV}$, the two sets of data appear to be mutually exclusive in a global fit.

Since we do not want to remove either of the data sets from the analysis and, in any case, have no means of judging whether there is a potential experimental inconsistency among the different pp sets, we decided to introduce a cut on the p_T of the observed pion. Including only pp data with $p_T \geq 5\,\mathrm{GeV}$ largely resolves the observed tension between RHIC and LHC data. This is illustrated in Figs. 9 and 10 where we compare to the PHENIX [21] and ALICE [32] data, respectively. We note that the calculated normalization shift (6) for the ALICE
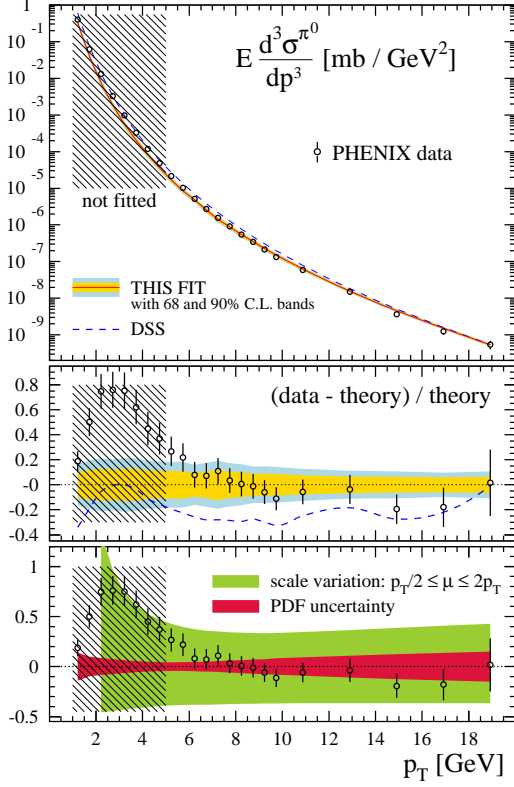
FIG. 9: Comparison of our NLO results for single-inclusive neutral pion production in pp collisions at $\sqrt{S} = 200\,\text{GeV}$ with the PHENIX data [21] (upper panel). The inner and outer shaded bands correspond to uncertainty estimates at 68% and 90% C.L., respectively. Also shown are the results obtained with the DSS FFs (dashed line). The middle panel shows corresponding results for "(data-theory)/theory". In the lower panel we illustrate the relevance of theoretical uncertainties due to scale and PDF variations.

FIG. 10: As in Fig. 9 but now for π^0 production in pp collisions at $\sqrt{S} = 7\,\text{TeV}$ as measured by the ALICE experiment [32].

7 TeV data results in a down-shift, outside the experimentally quoted normalization uncertainty, which contributes significantly to the quoted χ^2 value in Tab. II. As we have already hinted at in our discussion of the SIA data in Sec. III B, a different treatment of decay pions by the RHIC and LHC experiments might play some role for the tension observed at $p_T \lesssim 5\,\text{GeV}$.

As can be seen, both data sets are well described by the global fit above the introduced p_T cut which is indicated by the hatched area in both figures. One also notices the still sizable theoretical scale ambiguity at NLO accuracy, which is indicated in the lower panels of Figs. 9 and 10 and within which the data are consistent with the fit even below the imposed p_T cut. The PDF uncertainties, computed with the 90% CL NLO sets from MSTW [47] and also illustrated in the same panels, are much less significant than the scale ambiguities, in particular, for the ALICE data.

The resulting global fit is, as always, a compromise of all the data sets included in the analysis and, in particular, mediates between RHIC pp data preferring a larger
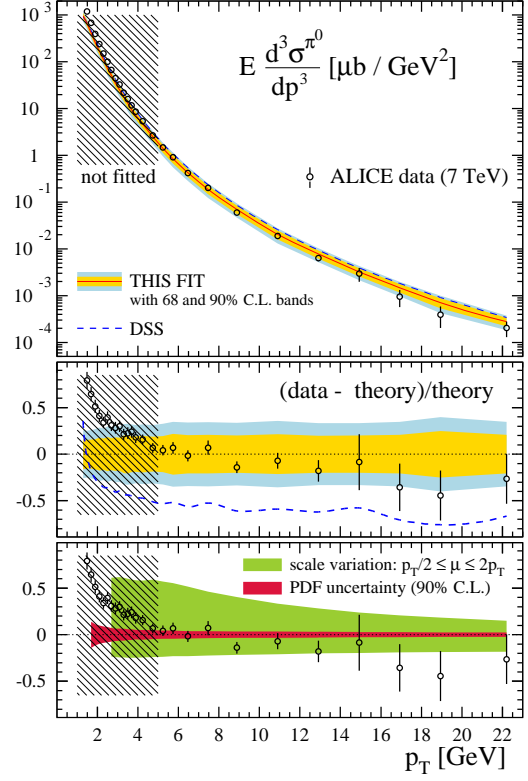
gluon-to-pion FF and LHC ALICE data favoring a smaller $D_g^{\pi^+}$. The net effect is a significantly reduced $D_g^{\pi^+}$ as compared to the DSS fit [10], as was already discussed in Sec. III A and illustrated in Fig. 1. Because of the remaining small tension, the estimated uncertainties on $D_g^{\pi^+}$ are sizable, despite the available amount of rather precise experimental data from pp collisions. If both RHIC and LHC data would point to a more similar $D_g^{\pi^+}$, the resulting uncertainties would likely to be somewhat smaller, however, the large theoretical scale ambiguities illustrated above still remain. We note that pp data at mid rapidity dominantly probe the gluon FF at medium-to-large z values as was, for instance, demonstrated in Ref. [50]

The last two figures give a similar comparison to the STAR data [33–36] for which we adopt, of course, the same p_T cut as for the other pp sets. None of these results was included in the DSS analysis. In Fig. 11 we focus on the π^-/π^+ ratio at mid rapidity [35], which is now much better described by the fit than with the DSS FFs. Scale ambiguities partially cancel in the ratio and are much less dramatic than for the individual cross sections, cf. Fig. 9. As was already mentioned, the ratio is sensitive to a potential charge asymmetry or SU(2) breaking, as parametrized by Eq. (3) in out fit. Like for the SIDIS multiplicities, the fit prefers little or no breaking, i.e.,
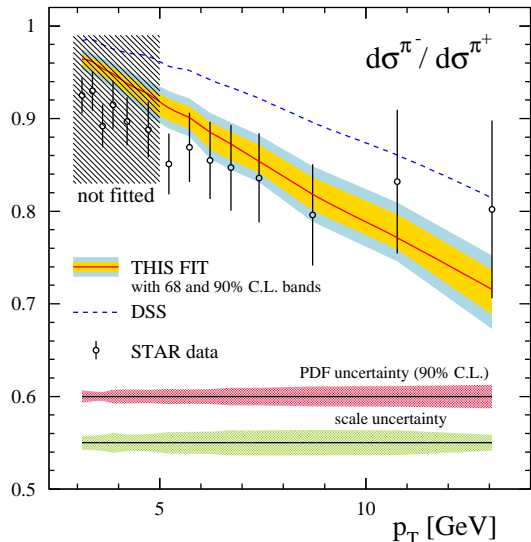
FIG. 11: Comparison of our NLO results for the π^-/π^+ cross section ratio in pp collisions at $\sqrt{S} = 200\,\text{GeV}$ with the STAR data [35]. The inner and outer shaded bands correspond to uncertainty estimates at 68% and 90% C.L., respectively. Also shown are the results obtained with the DSS FFs (dashed line). Scale and PDF uncertainties are indicated at the base of the plot.

$N_{d+\bar{d}}$ in (3) close to unity.

Figure 12 gives on overview of the three other sets of single-inclusive pion data from the STAR Collaboration used in the fit [33, 34, 36], which span different rapidity intervals. Since we fit to the π^-/π^+ ratio shown in Fig. 11 and π^+ data, we exclude results on the π^- cross section to avoid double-counting. The description of the data is very good, even below the p_T-cut of $5\,\text{GeV}$, indicating that there is a little bit less of a tension with ALICE results than for the PHENIX experiment. Calculations based on the DSS FFs (dashed lines) also provide a good description of data.

IV. SUMMARY AND OUTLOOK

We have presented a new, comprehensive global QCD analysis of parton-to-pion fragmentation functions at next-to-leading order accuracy including the latest experimental information. The analyzed data for inclusive pion production in semi-inclusive electron-positron annihilation, deep-inelastic scattering, and proton-proton collisions span energy scales ranging from about $1\,\text{GeV}$ up to the mass of the Z boson. The achieved, very satisfactory and simultaneous description of all data sets strongly supports the validity of the underlying theoretical framework based on pQCD and, in particular, the notion of factorization and universality for parton-to-pion fragmentation functions.

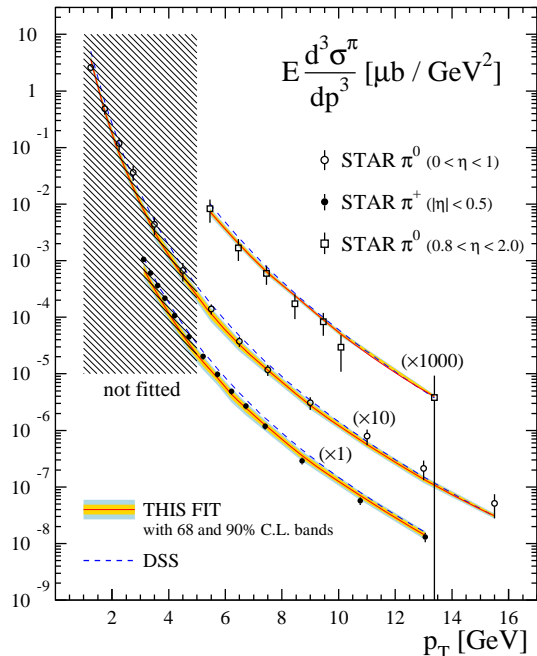Compared to our previous analysis, which was based

FIG. 12: Comparison of our NLO results for single-inclusive π^0 and π^+ production in different rapidity ranges in pp collisions at $\sqrt{S} = 200\,\text{GeV}$ with the corresponding STAR data [33, 34, 36]. The inner and outer shaded bands correspond to uncertainty estimates at 68% and 90% C.L., respectively. Also shown are the results obtained with the DSS FFs (dashed lines). Note that some of the results are rescaled for clarity.

on much less precise experimental input and to which we have made extensive comparisons throughout this work, we now obtained a significantly better fit, as measured in terms its the global χ^2, using the same functional form with only a few additional fit parameters. While most of the favored and unfavored quark-to-pion fragmentation functions are by and large similar to our previous results, the reduced amount of pions stemming for the hadronization of gluons is a noteworthy outcome of the new analysis. This finding was driven by first data from the CERN-LHC experiments, which, surprisingly, turned out to be mutually incompatible with previously available data obtained in lower center-of-mass system energy collisions at BNL-RHIC. To remedy this tension in our fit, we were forced to introduce a lower cut on the transverse momentum of the produced pions in proton-proton collisions. We have argued that it should be worthwhile for the experiments to compare in detail their procedures to determine pion yields as, for instance, different cuts for secondary pions from decays of other, heavier mesons perhaps have some numerical impact. We believe that such a contamination from feed-down pions might show up most prominently at small transverse momenta, where we currently observe the tension between RHIC and LHC data. We wish to mention that in the quark sector, the new data do not favor any charge symmetry violation be-

tween the total up- and down-quark fragmentation functions, contrary to our previous fit.

We have also performed a, what we believe, first reliable and trustworthy estimate of uncertainties for parton-to-pion fragmentation functions based on the standard iterative Hessian method. This was made possible by the wealth of new data included in our updated global analysis. The obtained uncertainties are still sizable and range at best from about ten percent to twenty-five percent for the total u-quark and gluon fragmentation function, respectively, in the kinematic regions covered by data and they quickly deteriorate beyond. A new asset of the current analysis is the analytic procedure to determine the optimum normalization shift for each data set in the fit, which greatly facilitated the global fitting procedure

The newly obtained pion fragmentation functions and their uncertainty estimates will be a crucial ingredient in future global analyses of both helicity and transverse-momentum dependent parton densities, which heavily draw on data with identified pions in the final-state. Our results will also serve as the baseline in heavy ion and proton-heavy ion collisions, where one of the main objectives is to quantify and understand possible modifications of hadron production yields by the nuclear medium. Also, the current analysis framework will be adopted for updates of the parton-to-kaon fragmentation functions, which we will pursue once all the promised sets of new data eventually become available. Since pions and kaons constitute by far the largest fraction in frequently measured yields of unidentified charged hadrons, a precise determination of their respective, optimum sets of fragmentation functions, including reliable uncertainty estimates, is critical to determine the room left for other hadrons, such as protons, in a future global analysis of charged hadron data.

Further improvements of parton-to-pion fragmentation functions from the theory side should include an improved treatment of heavy quark-to-pion fragmentation functions, likely along similar lines as for heavy flavor parton densities. Also, the impact of higher order corrections beyond the next-to-leading order accuracy should be explored. On the one hand, it is already possible to perform a next-to-next-to-leading order analysis of electron-positron annihilation data, and, on the other hand, the theoretical framework for all-order resummations of potentially large logarithmic corrections is available. On the experimental side, RHIC and the LHC will continue to provide new data on identified hadron spectra but it should be also worthwhile to explore the potential of future accelerator projects, such as an Electron-Ion Collider currently pursued in the U.S. [52], to further our knowledge of fragmentation functions and the physics behind hadronization.

Acknowledgments

[1] J. C. Collins and D. E. Soper, Nucl. Phys. B **193**, 381 (1981) [Erratum-ibid. B **213**, 545 (1983)]; Nucl. Phys. B **194**, 445 (1982).

[2] See, e.g., J. C. Collins, D. E. Soper, and G. F. Sterman, Adv. Ser. Direct. High Energy Phys. **5**, 1 (1988).

[3] G. Curci, W. Furmanski, and R. Petronzio, Nucl. Phys. B **175**, 27 (1980); W. Furmanski and R. Petronzio, Phys. Lett. B **97**, 437 (1980); L. Baulieu, E. G. Floratos, and C. Kounnas, Nucl. Phys. B **166**, 321 (1980).

[4] See also, M. Stratmann and W. Vogelsang, Nucl. Phys. B **496**, 41 (1997).

[5] G. Altarelli, R. K. Ellis, G. Martinelli, and S. Y. Pi, Nucl. Phys. B **160**, 301 (1979); W. Furmanski and R. Petronzio, Z. Phys. C **11**, 293 (1982).

[6] P. J. Rijken and W. L. van Neerven, Phys. Lett. B **386**, 422 (1996).

[7] D. de Florian, M. Stratmann, and W. Vogelsang, Phys. Rev. D **57**, 5811 (1998).

[8] F. Aversa, P. Chiappetta, M. Greco, and J. P. Guillet, Nucl. Phys. B **327**, 105 (1989); B. Jager, A. Schafer, M. Stratmann, and W. Vogelsang, Phys. Rev. D **67**, 054005 (2003); D. de Florian, Phys. Rev. D **67**, 054004 (2003).

[9] D. Graudenz, Nucl. Phys. B **432**, 351 (1994); L. Trentadue and G. Veneziano, Phys. Lett. B **323**, 201 (1994); D. de Florian, C. A. Garcia Canal, and R. Sassot, Nucl. Phys. B **470**, 195 (1996).

[10] D. de Florian, R. Sassot, and M. Stratmann, Phys. Rev. D **75**, 114010 (2007).

[11] D. de Florian, R. Sassot, and M. Stratmann, Phys. Rev. D **76**, 074033 (2007).

[12] C. A. Aidala, F. Ellinghaus, R. Sassot, J. P. Seele, and M. Stratmann, Phys. Rev. D **83**, 034002 (2011).

[13] S. Kretzer, Phys. Rev. D **62**, 054001 (2000);

[14] B. A. Kniehl, G. Kramer, and B. Potter, Nucl. Phys. B **582**, 514 (2000); S. Albino, B. A. Kniehl, and G. Kramer, Nucl. Phys. B **725**, 181 (2005); *ibid.* **803**, 42 (2008); M. Hirai, S. Kumano, T. -H. Nagai, and K. Sudoh, Phys. Rev. D **75**, 094009 (2007).

[15] A. Hillenbrand, "Measurement and Simulation of the Fragmentation Process at HERMES", Ph.D. thesis, Erlangen Univ., Germany, September 2005; private communications.

[16] D. Buskulic *et al.* [Aleph Collaboration], Z. Phys. C **66**, 355 (1995).

[17] P. Abreu *et al.* [Delphi Collaboration], Eur. Phys. J. C

5, 585 (1998).

[18] R. Akers *et al.* [Opal Collaboration], Z. Phys. C **63**, 181 (1994).

[19] K. Abe *et al.* [Sld Collaboration], Phys. Rev. D **59**, 052001 (1999).

[20] G. Abbiendi *et al.* [Opal Collaboration], Eur. Phys. J. C **16**, 407 (2000).

[21] A. Adare *et al.* [Phenix Collaboration], Phys. Rev. D **76**, 051106 (2007).

[22] D. Graudenz, M. Hampel, A. Vogt, and C. Berger, Z. Phys. C **70** (1996) 77.

[23] M. Stratmann and W. Vogelsang, Phys. Rev. D **64**, 114007 (2001).

[24] D. de Florian, R. Sassot, M. Stratmann, and W. Vogelsang, Phys. Rev. D **80**, 034030 (2009).

[25] D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung, Phys. Rev. D **65**, 014012 (2001)

[26] M. Epele, R. Llubaroff, R. Sassot, and M. Stratmann, Phys. Rev. D **86**, 074028 (2012).

[27] J. Pumplin, D. R. Stump, and W. K. Tung, Phys. Rev. D **65**, 014011 (2001); 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).

[28] J. P. Lees *et al.* [BaBar Collaboration], Phys. Rev. D **88**, 032011 (2013).

[29] M. Leitgab *et al.* [Belle Collaboration], Phys. Rev. Lett. **111**, 062002 (2013).

[30] A. Airapetian *et al.* [Hermes Collaboration], Phys. Rev. D **87**, 074029 (2013).

[31] N. Makke [Compass Collaboration], PoS DIS **2013**, 202 (2013).

[32] B. Abelev *et al.* [Alice Collaboration], Phys. Lett. B **717**, 162 (2012).

[33] J. Adams *et al.* [Star Collaboration], Phys. Lett. B **637**, 161 (2006).

[34] B. I. Abelev *et al.* [Star Collaboration], Phys. Rev. D **80**, 111108 (2009).

[35] G. Agakishiev *et al.* [Star Collaboration], Phys. Rev. Lett. **108** (2012) 072302.

[36] L. Adamczyk *et al.* [Star Collaboration], Phys. Rev. D **89**, 012001 (2014).

[37] D. de Florian, R. Sassot, M. Stratmann, and W. Vogelsang, Phys. Rev. Lett. **101**, 072001 (2008); Phys. Rev. Lett. **113**, 012001 (2014); E. Leader, A. V. Sidorov, and D. B. Stamenov, Phys. Rev. D **82**, 114018 (2010).

[38] For some recent TMD studies using FFs, see: M. G. Echevarria, A. Idilbi, Z. B. Kang, and I. Vitev, Phys. Rev. D **89**, 074013 (2014); M. Anselmino, M. Boglione, U. D'Alesio, S. Melis, F. Murgia, and A. Prokudin, Phys. Rev. D **89**, 114026 (2014); P. Sun, J. Isaacson, C.-P. Yuan, and F. Yuan, arXiv:1406.3073.

[39] See, e.g., R. Sassot, M. Stratmann, and P. Zurita, Phys. Rev. D **81**, 054001 (2010).

[40] A. Mitov, S. Moch, and A. Vogt, Phys. Lett. B **638**, 61 (2006); S. Moch and A. Vogt, Phys. Lett. B **659**, 290 (2008); A. A. Almasy, S. Moch, and A. Vogt, Nucl. Phys. B **854**, 133 (2012).

[41] P. J. Rijken and W. L. van Neerven, Nucl. Phys. B **487**, 233 (1997); A. Mitov and S. O. Moch, Nucl. Phys. B **751**, 18 (2006).

[42] M. Cacciari and S. Catani, Nucl. Phys. B **617**, 253 (2001); J. Blumlein and V. Ravindran, Phys. Lett. B **640**, 40 (2006); S. Moch and A. Vogt, Phys. Lett. B **680**, 239 (2009); D. P. Anderle, F. Ringer, and W. Vogelsang, Phys. Rev. D **87**, 034014 (2013).

[43] See, e.g., S. Alekhin, J. Blumlein, S. Klein, and S. Moch, Phys. Rev. D **81**, 014032 (2010); R. S. Thorne, Phys. Rev. D **86**, 074017 (2012) and references therein.

[44] See, e.g., M. Cacciari, P. Nason, and C. Oleari, JHEP **0510**, 034 (2005).

[45] I. Arsene *et al.* [Brahms Collaboration], Phys. Rev. Lett. **98**, 252001 (2007).

[46] J. Adams *et al.* [Star Collaboration], Phys. Rev. Lett. **97**, 152302 (2006).

[47] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, Eur. Phys. J. C **63**, 189 (2009).

[48] 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.

[49] W. Braunschweig *et al.* [Tasso Collaboration], Z. Phys. C **42**, 189 (1989).

[50] See, e.g., R. Sassot, M. Stratmann, and P. Zurita, Phys. Rev. D **82**, 074011 (2010).

[51] D. d'Enterria, K. J. Eskola, I. Helenius, and H. Paukkunen, Nucl. Phys. B **883**, 615 (2014).

[52] For information on the EIC project, see: D. Boer *et al.*, arXiv:1108.1713; A. Accardi *et al.*, arXiv:1212.1701; E. C. Aschenauer *et al.*, arXiv:1409.1633.