Nuclear effects in the deuteron and global QCD analyses

S. I. Alekhin[*]
II. Institut für Theoretische Physik, Universität Hamburg, Luruper Chaussee 149, D–22761 Hamburg, Germany

S. A. Kulagin[†]
Institute for Nuclear Research of the Russian Academy of Sciences, 117312 Moscow, Russia

R. Petti[‡]
Department of Physics and Astronomy, University of South Carolina, Columbia, South Carolina 29208, USA

(Received 18 March 2022; accepted 3 June 2022; published 24 June 2022)

We report the results of a new global QCD analysis, which includes deep-inelastic e/μ scattering data off proton and deuterium, as well as Drell-Yan lepton pair production in proton-proton and proton-deuterium collisions and $W^{\pm}/Z$ boson production data from pp and $p\bar{p}$ collisions at the LHC and Tevatron. Nuclear effects in the deuteron are treated in terms of a nuclear convolution approach with bound off-shell nucleons within a weak binding approximation. The off-shell correction is controlled by a universal function of the Bjorken variable x, describing the modification of parton distributions in bound nucleons, which is determined in our analysis along with the parton distribution functions of the proton. A number of systematic studies are performed to estimate the uncertainties arising from the use of various deuterium datasets, from the modeling of higher twist contributions to the structure functions, from the treatment of target mass corrections, as well as from the nuclear corrections in the deuteron. We obtain predictions for the ratios F_2^n/F_2^p, and d/u, focusing on the region of large x. We also compare our results with the ones obtained by other QCD analyses, as well as with the recent data from the MARATHON experiment.

DOI: 10.1103/PhysRevD.105.114037

I. INTRODUCTION

An accurate determination of the parton distribution functions (PDFs) in the proton and the neutron is of primary importance for modern high-energy physics, as PDFs determine the leading contribution to the cross sections of various high-energy processes. Since PDFs are not directly observable, they are usually extracted phenomenologically from global QCD analyses to experimental data at large momentum transfer, including lepton deep inelastic scattering (DIS), lepton-pair production (Drell-Yan process), jet production, and $W^{\pm}/Z$ boson production in hadron collisions (for a review, see, e.g., Ref. [1]). While the abundant data available from a hydrogen target allow a reliable determination of the PDF content of the proton, data from various nuclei—most notably deuterium—are required as effective neutron targets to constrain the parton content of the neutron. Furthermore, a combination of hydrogen and deuterium data has been commonly used to separate the u and d quark PDFs, in particular, at large values of Bjorken x.[1]

Precision studies require one to address the effects of nuclear environment at the parton level. While the nuclear effects in nuclear PDFs (nPDFs) analyses are usually treated empirically [4–7], a number of physics mechanisms are known to affect the PDFs and the structure functions (SFs) of the bound nucleons (for a review, see, e.g., [8–10]). In the region of large x, the relevant nuclear effects are related to the smearing of the cross sections with the nuclear momentum distribution [11] (Fermi motion), together with the nuclear binding correction [12]. In addition to these corrections, which have kinematical origin, nuclear effects related to the dynamical modification of the internal parton structure have to be addressed in bound nucleons. In Refs. [13,14], such a

[*]sergey.alekhin@desy.de
[†]kulagin.physics@gmail.com
[‡]roberto.petti@cern.ch

[1]For a recent discussion of the impact of deuterium data on global QCD analyses, see Refs. [2,3].

modification is related to the off-mass-shell effect, i.e., the dependence of bound nucleon SFs on its virtual mass squared $p^2 = p_0^2 - \boldsymbol{p}^2$, where p_0 and $\boldsymbol{p}$ are the nucleon energy and momentum, respectively. This dependence is treated in Refs. [13,14] as a perturbative correction in the nucleon virtuality $v = (p^2 - M^2)/M^2$, relying on the fact that the typical nucleon momentum and energy are small compared to the nucleon mass M in the nuclear ground state. Within this weak binding approximation, the corresponding nuclear correction is controlled by the SFs' derivative with respect to p^2, which is described in Ref. [14] in terms of a dimensionless function $\delta f(x)$. Additional effects related to the meson-exchange currents and the nuclear shadowing are relevant at intermediate and small x values. A model combining all of these effects has been successfully used to quantitatively explain the observed dependencies on x, invariant momentum transfer squared Q^2, nuclear mass number A of the nuclear DIS data in a wide range of targets from ^{3}He to ^{207}Pb [14–16]. The same model also demonstrates an excellent agreement with the magnitude, the x and mass dependence of the nuclear Drell-Yan (DY) data [17], as well as with the data on the differential cross sections for $W^{\pm}/Z$ boson production in proton-lead collisions at the LHC [18].

The off-shell effect is an important contribution to the full nuclear correction. The corresponding function δf was determined for the isoscalar nucleon from an analysis of nuclear DIS data on the cross-section ratios σ^A/σ^d [14]. The function $\delta f(x)$ was also independently extracted together with the proton PDFs in global QCD analyses of proton and deuterium DIS data [19,20]. The results of Ref. [19] on δf are consistent with the previous determination from nuclei with $A \geq 4$ [14]. However, Refs. [20] and [19] strongly disagree on both the values of the function δf and on the ratio of the d and u quark PDFs at large x. These observations motivate the present study, in which we perform a new global QCD analysis with updated sets of deuterium DIS data. We discuss a number of systematic studies aimed at understanding the uncertainties associated with a number of effects, including the consistency of various deuterium datasets, the treatment of target mass correction, and the modeling of higher-twist contributions and of the nuclear corrections in the deuterium. We also provide our predictions on the ratios F_2^n/F_2^p and d/u and compare them with the ones from the QCD analyses of Refs. [20–23], as well as with the recent data from MARATHON experiment [24].

The paper is organized as follows. In Sec. II, we outline the theory framework used in our analysis of the proton and deuterium DIS. In Sec. III, we discuss the data samples and the details of our analysis. In Sec. IV, we summarize our results, while in Sec. V, we discuss the uncertainties associated with the use of different deuterium datasets and with the modeling of the structure functions. In Sec. VI, we compare our predictions on F_2^n/F_2^p with MARATHON data and with the results of other QCD analyses, including the d/u ratio. In Appendix A, we show the pulls obtained in our fit from different deuterium datasets. In Appendix B, we discuss in detail the phase space in the nuclear convolution equations employed in our analysis.

II. THEORY FRAMEWORK

A. Nucleon structure functions

The inclusive spin-independent electron(muon)-nucleon inelastic cross section is described by two SFs, $F_T = 2xF_1$ and F_2, which depend on two independent variables, the invariant momentum transfer squared $Q^2 = -q^2$ and the dimensionless Bjorken $x = Q^2/(2p \cdot q)$, where p is the nucleon four-momentum and q is the four-momentum transfer.

A common framework to describe the DIS is the operator product expansion (OPE), which introduces the power series in Q^{-2} (twist expansion). To the first order, i.e., in the leading twist (LT), the SFs are fully determined by the PDFs. Corrections from the higher-twist (HT) quark-gluon operators should also be supplemented by those arising from the finite nucleon mass (target mass correction, or TMC) [25]. We also note that for the sake of computing the nuclear SFs (see Sec. II B), the nucleon SFs are required in the off-mass-shell region $p^2 < M^2$, where M is the nucleon mass. The unpolarized nucleon SFs in the DIS region can then be written as follows:

$$F_i(x, Q^2, p^2) = F_i^{\mathrm{TMC}}(x, Q^2, p^2) + H_i/Q^2, \qquad (1)$$

where $i = T, 2$ and F_i^{TMC} are the corresponding LT SFs corrected for the target mass effect, and H_i describe the dynamical twist-four contribution (for brevity, we suppress explicit notation to the twists higher than four). In this study, we consider two different phenomenological HT models: (1) the additive HT model, in which we assume $H_i = H_i(x)$ and (2) the multiplicative HT model [26], in which H_i is assumed to be proportional to the corresponding LT SF, $H_i = F_i^{\mathrm{LT}}(x, Q^2)h_i(x)$. The HT terms from both models are addressed in this study.

The LT SFs are computed using the nucleon PDFs and coefficient functions, which are subject to a power series in the QCD coupling constant. The neutron LT SFs are computed in terms of the proton PDFs relying on the isospin symmetry of u and d quark PDFs. The isospin relations for the HT terms are not so obvious. By default, we assume $H_i^p = H_i^n$.[2] We also consider the relation $h_i^p = h_i^n$ with the multiplicative HT model.

To account for the TMC, we follow the Georgi-Politzer OPE approach [25]. Since the calculation of the nuclear SFs

[2]We note, however, that a nonzero isovector component $H_2^p - H_2^n$ was obtained in a QCD fit [27], although with rather large fit uncertainties. The difference $H_T^p - H_T^n$ was consistent with 0 within uncertainties.

requires the nucleon SFs in the off-shell mass region, we analytically continue the equations of Ref. [25] into the off-shell region by replacing the nucleon mass squared M^2 with p^2. We have

$$F_T^{\mathrm{TMC}}(x,Q^2,p^2)=\frac{x^2}{\xi^2\gamma}F_T^{\mathrm{LT}}(\xi,Q^2,p^2)+\frac{2x^3p^2}{Q^2\gamma^2}\int_\xi^1\frac{du}{u^2}F_2^{\mathrm{LT}}(u,Q^2,p^2), \quad (2a)$$

$$F_2^{\mathrm{TMC}}(x,Q^2,p^2)=\frac{x^2}{\xi^2\gamma^3}F_2^{\mathrm{LT}}(\xi,Q^2,p^2)+\frac{6x^3p^2}{Q^2\gamma^4}\int_\xi^1\frac{du}{u^2}F_2^{\mathrm{LT}}(u,Q^2,p^2), \quad (2b)$$

where $\xi=2x/(1+\gamma)$ is the Nachtmann variable, and $\gamma=(1+4x^2p^2/Q^2)^{1/2}$. Note that in Eq. (2) we drop the terms of order x^4p^4/Q^4, which produce numerically small contributions in the considered region. It should be commented that Eqs. (2a) and (2b) lead to a nonzero SFs at $x\to 1$. However, in practice, this violation of the inelastic threshold behavior does not affect the DIS region, which is characterized by high values of the invariant mass W of the produced hadronic states.

In the off-mass-shell region, the SFs explicitly depend on the nucleon invariant mass squared p^2. This dependence has two different sources: (i) the terms p^2/Q^2 in Eq. (2), which lead to power terms at large values of Q^2 and (ii) nonpower terms from the off-shell dependence of the LT SFs. Following Refs. [13,14], we note that for computing the nuclear SFs, it would be sufficient to know the proton and the neutron SFs in the vicinity of the mass shell $p^2=M^2$. We then treat the nucleon virtuality $v=(p^2-M^2)/M^2$ as a small parameter and expand SFs in series in v. To the leading order in v, we have

$$F_i^{\mathrm{LT}}(x,Q^2,p^2)=F_i^{\mathrm{LT}}(x,Q^2,M^2)[1+\delta f_i(x,Q^2)v], \quad (3)$$

$$\delta f_i(x,Q^2)=M^2\partial_{p^2}\ln F_i^{\mathrm{LT}}(x,Q^2,p^2), \quad (4)$$

where F_i^{LT} on the right-hand side in Eq. (3) are the structure functions $i=T,2$ of the on-mass-shell nucleon, and ∂_{p^2} in Eq. (4) denotes the partial derivative with respect to p^2 taken on the mass shell $p^2=M^2$. According to Eq. (4), the function δf_i describes the relative modification of the nucleon LT F_i in the vicinity of the mass shell, which is related to the corresponding PDF modification.

In this study, we assume the function δf to be the same for F_T and F_2 motivated by the fact that $F_T\approx F_2$ in the region of large x. The function δf drives the nuclear correction associated with the modification of the bound nucleon in the nuclear environment [14]. Detailed studies of nuclear DIS, DY production of the lepton pair and W/Z boson in Refs. [14,16–19] are consistent with no significant scale and nucleon isospin dependencies of δf. We thus assume the same $\delta f=\delta f(x)$ function for the proton and the neutron.

Note that Eq. (3) holds in the vicinity of the mass shell where $|v|\ll 1$. In computing the nuclear SFs, we integrate over the bound nucleon momentum as discussed in Sec. II B. For kinematics reason, $p^2<M^2$ and $v<0$ for bound nucleons. Using the results of Ref. [14], we have $\delta f\sim 1$ at large $x>0.6$. Then the off-shell correction in Eq. (3) is large and negative for $v\sim -1$, and the off-shell SFs may be negative in this region. Since the values $|v|\gtrsim 1$ are outside of the region of applicability of the linear approximation in v, Eq. (3), we consider the following model in the full region of v:

$$F_i^{\mathrm{LT}}(x,Q^2,p^2)=F_i^{\mathrm{LT}}(x,Q^2,M^2)\exp[\delta f(x)v]. \quad (5)$$

This equation ensures the positivity of SFs in the off-shell region, and for a small off-shell correction, Eq. (5) is identical to Eq. (3). In the study of the deuteron SFs, we consider both Eqs. (3) and (5).

B. Deuteron structure functions

We assume that the nuclear DIS in the region $x>0.1$ is dominated by the incoherent scattering off the bound protons and neutrons and consider the process in the target rest frame. The deuteron structure functions can be written as follows [14,27]:

$$F_i^d(x,Q^2)=\int \mathrm{d}^3\boldsymbol{p}|\Psi_d(\boldsymbol{p})|^2K_{ij}[F_j^p(x',Q^2,p^2)+F_j^n(x',Q^2,p^2)], \quad (6)$$

where $i,j=T,2$, and we assume a summation over the repeated subscript j. The integration is performed over the bound nucleon momentum $\boldsymbol{p}$, and $\Psi_d(\boldsymbol{p})$ is the deuteron wave function in the momentum space, which is normalized as

$$\int \mathrm{d}^3\boldsymbol{p}|\Psi_d(\boldsymbol{p})|^2=1. \quad (7)$$

Because of the energy-momentum conservation, the four-momentum of the struck proton (neutron) is $p=(M_d-\sqrt{M^2+\boldsymbol{p}^2},\boldsymbol{p})$, where M_d is the deuteron mass, and M is the mass of residual nucleon [$M=M_n$ for the proton contribution and $M=M_p$ for the neutron contribution in Eq. (6)]. We use a coordinate system in which the momentum transfer $\boldsymbol{q}$ is antiparallel to the z axis, and p_z and $\boldsymbol{p}_\perp$ are the longitudinal and transverse component of the nucleon momentum, $p^2=p_0^2-\boldsymbol{p}^2$ and $x'=Q^2/(2p\cdot q)$ are the invariant mass and the Bjorken variable of the off-shell nucleon, respectively. The kinematic factors K_{ij} are [14]

$$K_{TT} = \left(1 + \frac{\gamma p_z}{M}\right), \qquad K_{T2} = 2\frac{x'^2 \boldsymbol{p}_\perp^2}{Q^2}, \tag{8a}$$

$$K_{2T} = 0, \quad K_{22} = \left(1 + \frac{\gamma p_z}{M}\right)\left(1 + \frac{x'^2(4p^2 + 6\boldsymbol{p}_\perp^2)}{Q^2}\right)\frac{1}{\gamma^2}, \tag{8b}$$

where $\gamma = (1 + 4x^2M^2/Q^2)^{1/2}$. Note that Eq. (6) and (8) are the result of a series expansion of relativistically covariant operators in the parameters $\boldsymbol{p}/M$ and $(p_0 - M)/M$ to order $\boldsymbol{p}^2/M^2$ (for more detail, see [13,14] and Appendices B and C of Ref. [28]). The factor $1 + \gamma p_z/M$ in Eq. (8) describes the change in the virtual photon flux for a bound nucleon with the momentum $\boldsymbol{p}$ compared to the corresponding flux for the nucleus at rest. Note also the term $K_{T2} \sim x^2\boldsymbol{p}_\perp^2/Q^2$ resulting from a mixing effect between the longitudinal and transverse structure functions at finite values of Q^2, which is due to the transverse motion of the bound nucleon.

Assuming no p^2 dependence of the nucleon structure functions, in the limit $Q \gg M$, Eq. (6) reduces to the standard convolution of the nucleon SFs with the nucleon distribution over the light-cone momentum $y = (p_0 + p_z)/M$ in the deuteron. In the presence of an off-shell p^2 dependence, we have a generalized convolution, which involves the integration over the light-cone momentum y and the nucleon virtuality p^2 [13]. The phase space at finite Q^2 used in Eq. (6) is discussed in more detail in Appendix B.

In the region $x < 0.1$, the corrections due to the meson-exchange currents and the nuclear shadowing, at even smaller values of $x \ll 0.1$, are relevant. In this study, while focusing on $x > 0.1$, we treat these effects following Refs. [14,17].

III. OFF-SHELL FUNCTION WITHIN GLOBAL QCD ANALYSIS

A. Data samples

The present study is an update of our former analysis [19] based on the data on the DIS of charged leptons off hydrogen and deuterium combined with the ones on $W^\pm/Z$ boson production at hadron colliders. The latter samples allow the separation of the u and d quark distributions in a wide range of x that, in turn, provides a basis for studying nuclear effects in the deuteron for the DIS structure functions. The deuterium datasets employed for this purpose are listed in Table I. They comprise the ones used in the analysis of Ref. [19] supplemented by additional datasets including the most recent results on σ^d/σ^p by the MARATHON experiment at Jlab [24]. Due to the increased energy of the upgraded Jlab beam, the MARATHON data cover a much wider kinematics as compared to the earlier JLab BoNuS experiment [29].[3] Besides, a dedicated study performed by MARATHON allowed one to reduce the normalization uncertainty in its measurements to unprecedented level of 0.55%. This guarantees a superior statistical significance of the MARATHON data over both the original BoNuS sample [29] and the results of the study [30] based on the BoNuS measurements. The results on F_2^d/F_2^N derived in Ref. [30] from the BoNuS data on F_2^n/F_2^d using a parametrization of F_2^p were employed in our earlier study [19]. However, since they are sensitive to model assumptions about the F_2^p shape, in the present study, we select the original BoNuS data in order to reduce the model dependence of the analysis. To provide a complete representation of the relevant data, we also add to the fit the DIS data collected in the Jlab-E00-116 [31] and DESY-HERMES [32] experiments. Finally, we replace the deuteron NMC data [33] used in Ref. [19] by more recent measurements [34] performed with a higher luminosity. In Fig. 1, we illustrate the accuracy of various datasets by plotting the ratio F_2^d/F_2^p in the range $0.1 < x < 0.5$ and for $Q^2 = 14x\ (\mathrm{GeV}^2) \pm 10\%$ from various measurements [24,32,34,35].[4] This selection of the Q^2 band is motivated by kinematics of the MARATHON experiment [24].

Whenever possible, we select the data on the cross sections and their ratios rather than on the structure functions F_2. This makes the modeling more involved, however, allows for a consistent account of the contribution from the structure function F_L since the various experiments do not follow a common convention on the shape of F_L when extracting F_2 from the cross-section measurements. As an exception, the BoNuS [29] and the NMC [34] data are taken in the form of the ratios F_2^n/F_2^d and F_2^d/F_2^p, respectively, as the cross-section results have not been released by these experiments. Such an inconsistency can be, however, justified since the F_L contribution to great extent cancels out in the ratios. To ensure a perturbative QCD description of the leading-twist terms in the DIS structure functions, we impose a general cut of $Q^2 > 2.5\ \mathrm{GeV}^2$ and $W^2 > 3\ \mathrm{GeV}^2$. For the BoNuS data [30] used in our earlier study [19], a relaxed cut of $Q^2 > 1.5\ \mathrm{GeV}^2$ was selected in order to increase the statistical significance of this sample. In the present analysis, which includes the precision MARATHON data, this exemption is not applied, and the BoNuS data [29] are considered within a common framework.

Information about the point-to-point correlation of systematic errors in the data is taken into account in the fit whenever available. In particular, a detailed breakdown of the systematic uncertainties over independent sources is provided for the SLAC, CERN-BCDMS, CERN-NMC, and JLab-BoNuS experiments. For the Jlab-E00-116 and

[3]Note that while the MARATHON nuclear data covers the region $0.19 < x < 0.85$, the measurement of the ratio σ^d/σ^p is available for a limited region $0.19 < x < 0.4$.

[4]We note that the reanalyzed SLAC data [35] given in Fig. 1 are somewhat different from the original data [36] due to updated radiative corrections and x-rebinning [37]. The normalization of the original SLAC data is more consistent with the MARATHON F_2^d/F_2^p data, as shown in Ref. [24].

TABLE I. The list of DIS data on the deuterium target employed in the present analysis alongside with the values of χ^2/NDP and normalization factors obtained in the fit in comparison with the experimentally determined normalization errors.

Facility	Experiment	Reference	Beam	Beam energy (GeV)	Observable	Normalization factor	Normalization error(s) (%)	$\frac{\chi^2}{\text{NDP}}$
SLAC	E49a	[36,38]	e	11 ÷ 19.5	$\frac{d^2\sigma^d}{dE'd\Omega}$	0.988(10)	2.1[a]	25/59
”	E49b	”	”	4.5 ÷ 18	”	0.996(10)	”	187/145
”	E87	”	”	8.7 ÷ 20	”	1.000(9)	”	114/109
”	E89b	[38,39]	”	10.4 ÷ 19.5	”	0.987(9)	”	52/72
”	E139	[38,40]	”	8 ÷ 24.5	”	1.002(9)	”	8/17
”	E140	[38,41]	”	3.7 ÷ 19.5	”	1	1.7	25/26
CERN	BCDMS	[42]	μ	100 ÷ 280	$\frac{d^2\sigma^d}{dxdQ^2}$	0.989(7)	3	273/254
”	NMC	[34]	”	90 ÷ 280	F_2^d/F_2^p	1	<0.15	155/165
DESY	HERMES	[32]	e	27.6	σ^d/σ^p	1	1.4	21/30
JLab	E00-116	[31]	”	5.5	$\frac{d^2\sigma^d}{dE'd\Omega}$	0.981(10)	1.75	208/136
”	BoNuS	[29]	”	4.2, 5.2	F_2^n/F_2^d	0.97(9)	7 ÷ 10	90/63
”	MARATHON	[24]	”	10.6	σ^d/σ^p	1	0.55	8/7
Total								1166/1083

[a]A general normalization uncertainty for the SLAC experiments derived from re-analysis of those data. The contributions of marginal size also apply to particular datasets [35].

MARATHON datasets, only the overall systematic uncertainty is published, and in the present fit, it is combined in quadrature with the statistical (uncorrelated) uncertainties. The systematic uncertainties of DESY-HERMES measurements are separated by sources; however, no information about their point-to-point correlation was provided. For this reason, we select for our fit the DESY-HERMES data on the ratio σ^d/σ^p, where the correlated uncertainties partially cancel. The remaining systematic uncertainty, except of the normalization one, are combined with the statistical uncertainty, in line with the approximation adopted in the DESY-HERMES analysis of their own data [32].

The normalization uncertainty, a peculiar case of systematic errors, often dominates the uncertainty of the datasets considered. Furthermore, the normalization factors for the available SLAC-E49a, E49b, E87, E89b, E139, and CERN-NMC datasets were estimated by comparing them to the measurements of the SLAC-E140 experiment with the normalization uncertainty of 1.7%. Following a similar approach, we release the normalization factors of those data and determine such factors from a fit simultaneously with other parameters. Furthermore, this procedure is also applied to the CERN-BCDMS and Jlab-E00-116 data, allowing for improvement of their instrumentally

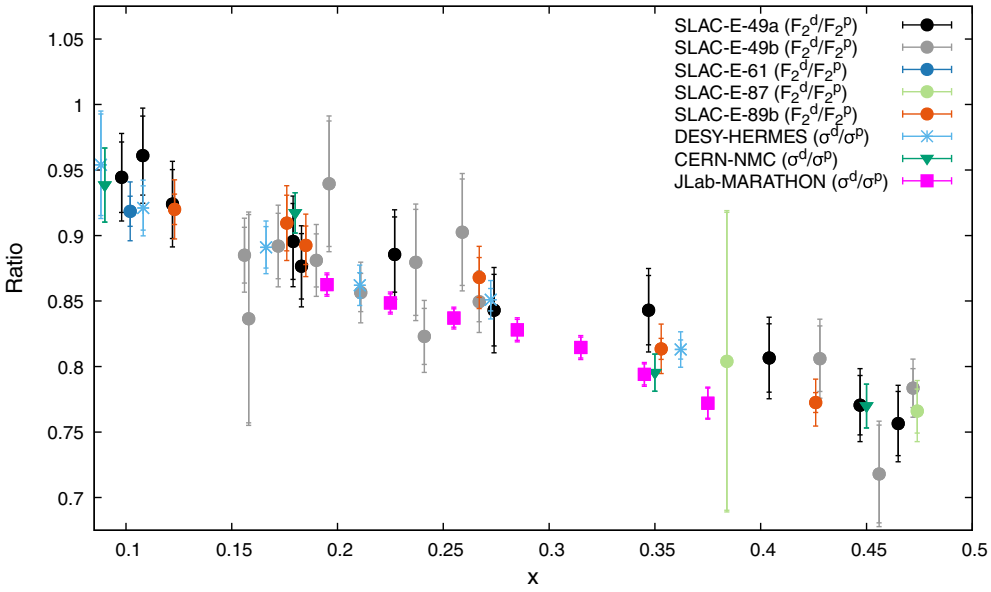

FIG. 1. Data on F_2^d/F_2^p (or σ^d/σ^p) for $0.1 < x < 0.5$ from various experiments. Data legend is shown in the plot. The data points were selected within interval $Q^2 = 14x$ (GeV^2) $\pm$ 10% to facilitate the comparison with Ref. [24]. The inner error bar shows the statistical and systematic error taken in quadrature, while the outer error bar in addition includes the normalization error listed in Table I.

determined normalizations. For the CERN-NMC and DESY-HERMES data, which were taken in the form of ratios, the impact of the normalization uncertainty is greatly reduced; therefore, their normalizations were kept fixed. A similar treatment was applied to the MARATHON data, which have a very accurate luminosity monitoring and for which we also avoid normalization tuning. This allows us to use those data for the calibration of the other datasets, in addition to the SLAC-E140 data set.

B. Analysis setup

The leading-twist PDFs, which are necessary for computing the nucleon SFs, Eq. (1), are parametrized using the shape employed in the ABMP16 fit [43] and in our earlier analysis [19]. The DIS SFs are treated as outlined in Sec. II A. The functions $H_T(x)$ and $H_2(x)$, which describe the HT contributions, are treated independently and are parametrized in a model-independent form of spline polynomials interpolating between the points $x = (0, 0.1, 0.3, 0.5, 0.7, 0.9, 1)$ with the values (HT coefficients) determined on this grid. We assume the HT terms to be independent of the nucleon isospin state. The Q^2 dependence of the LT component of the nucleon SFs was computed taking into account NNLO perturbative QCD (pQCD) corrections, while for the HT coefficients, possible pQCD effects have been neglected.

The nuclear effects in the deuteron are accounted by Eq. (6) with the off-shell correction governed by Eq. (4). The deuteron AV18 wave function is used [44,45]. The function $\delta f(x)$ is determined along with the proton PDFs and HTs in a fit to the deuterium data listed in Table I and the proton data from Table II in Ref. [19]. This function is parametrized as a polynomial:

$$\delta f(x) = c_0 + c_1 x + c_2 x^2. \tag{9}$$

Note that Eq. (6) describes the nuclear corrections driven by the momentum distribution, the nuclear binding, and the off-shell effect, which dominate in the present analysis. We also verified [19] that the other nuclear effects, such as the meson-exchange currents and the nuclear shadowing, are within experimental uncertainties, and for this reason, they are not considered in the present analysis.

IV. RESULTS

We simultaneously fit the parameters of the off-shell function δf with those of the PDFs and HTs in order to provide a consistent separation of the various contributions to the SFs by exploiting the broad x and Q^2 coverage of available data. The resulting data normalization factors and χ^2 values corresponding to various deuterium datasets are listed in Table I, and the parameters of the $\delta f(x)$ function are $c_0 = -0.16 \pm 0.11$, $c_1 = -2.04 \pm 0.73$, and $c_2 = 4.86 \pm 1.13$. This function is shown in Fig. 2, together with the results of other determinations of this quantity from Refs. [14,19,20].

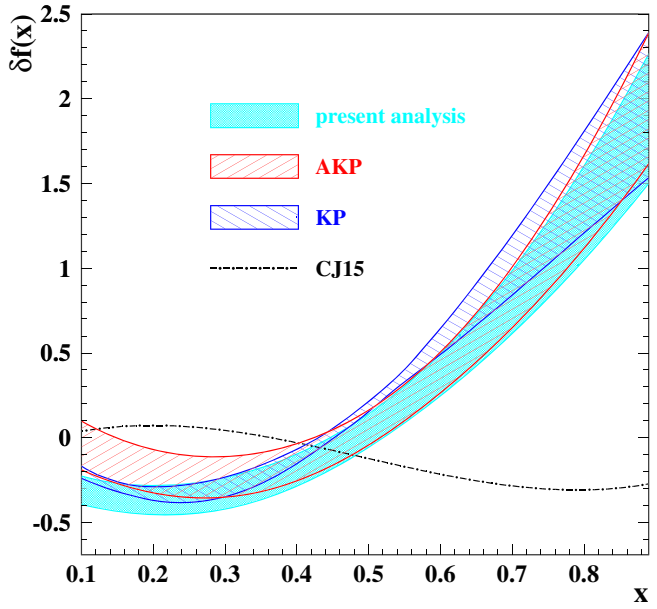

FIG. 2. The 1σ uncertainty band for the off-shell function δf as a function of x obtained in the present analysis (shaded cyan area) in comparison with the results of the earlier AKP [19] (right-tilted hash area), KP [14] (left-tilted hash area), and CJ15 [20] (dash-dotted curve) fits.

For all data points included in our fit, we have $\chi^2/\text{d.o.f.} = 4842/4044$. The present results are in a good agreement with both our former global QCD analysis [19] and the analysis of Ref. [14], in which the function $\delta f(x)$ was determined from a global fit to the data on the ratios σ^A/σ^d for the DIS cross sections off nuclear targets with the mass number $4 \le A \le 208$ using the proton and neutron SFs of Ref. [46]. However, our results are in a strong disagreement with those of Ref. [20]. Below we trace possible reasons of the discrepancy with Ref. [20] by verifying the differences in the experimental datasets and in the underlying model.

In Fig. 3, we compare our results with the recent measurement of σ^d/σ^p by the MARATHON experiment [24]. Also shown are the predictions from the CT18 [21], MSHT20 [22], NNPDF4.0 [23], and CJ15 [20] QCD analyses. Note that the CJ15 analysis was performed to the NLO approximation, while all others were done to the NNLO one. The CT18, MSHT20, and NNPDF4.0 structure functions are computed in the three-flavor scheme using the code OPENQCDRAD (version 2.1) [47] combined with the LHAPDF (version 6) PDF grids [48,49] CT18NNLO, MSHT20nnlo_nf3, and NNPDF40_nnlo_pch_as_01180_nf_3, respectively. We use $F_2^d = F_2^p + F_2^n$ for CT18 and NNPDF4.0 in Fig. 3, as those analyses do not account for the deuteron correction. Both MSHT20 and CJ15 account for the deuteron effect in their PDF fits. For MSHT20, we take $F_2^d = R_d(F_2^p + F_2^n)$ with the correction factor R_d obtained in the NNLO global QCD fit of Ref. [22]. For CJ15, we use their results obtained from Ref. [24]. All the predictions are in agreement with MARATHON σ^d/σ^p data within

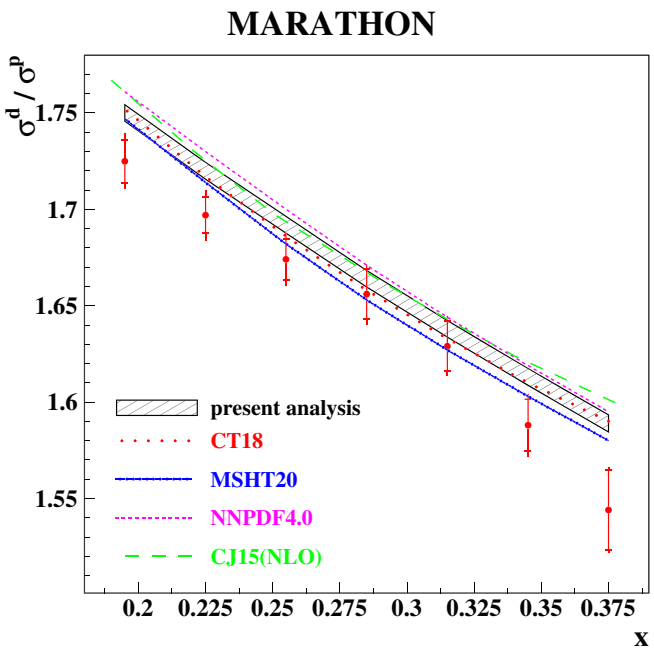

FIG. 3. Data on σ^d/σ^p from MARATHON measurement [24] compared with corresponding 1σ band of present analysis (left-tilted area). Also shown are the predictions of various global QCD analyses: CT18 [21] (dots), MSHT20 [22] (connected dots), NNPDF4.0 [23] (short dashes), and CJ15 [20] (long dashes).

uncertainties in the region about $x = 0.3$ and somewhat overshoot the data for x values about 0.2 and 0.4.[5]

The quality of our fit for the newly added BoNuS data on F_2^n/F_2^d [29], the cross-section measurements from JLab-E00-116 experiment [31], NMC data on F_2^d/F_2^p [34], and the HERMES data on σ^d/σ^p [32] are illustrated in detail in Figs. 8–11 in Appendix A. In general, we observe a good agreement of the fit and the data with no regular and/or statistically significant trend in the pulls.

V. SYSTEMATIC STUDIES

We verify the stability of our results on δf by performing a number of fits with a modified ansatz. Some of these modifications, like the parametrization of higher-twist terms (additive vs multiplicative form) and the off-shell correction [a linear dependence on v in Eq. (3) vs exponentiated form of Eq. (5)], reflect uncertainties in the theory framework of the fit. The other modifications are motivated by other studies in the field and aimed to facilitate a comparison with those studies. The impact of the modifications considered is summarized in Fig. 4, and their detailed description is given below in this section.

[5]Note that in Fig. 3 our predictions are for the ratio of cross sections σ^d/σ^p, while for the other groups, we compute the ratio F_2^d/F_2^p. The relation $\sigma^d/\sigma^p = F_2^d/F_2^p$ is justified by observation that the deuteron, and the nucleon have equal $R = \sigma_L/\sigma_T$ within experimental uncertainties [34].

Note that for all considered modifications of the fit, the data normalization factors are kept at the values of Table I, thus allowing us to avoid the interplay with the data shift.

A. NMC data choice

Our fit includes the NMC data on the ratio F_2^d/F_2^p [34]. These data are derived from the cross-section data assuming the same ratio $R = \sigma_L/\sigma_T$ for the proton and deuteron, which was verified experimentally by the NMC with a good accuracy. Alternatively, in our former analysis [19], the cross-section data for the deuterium target [33] have been employed instead. In order to verify the impact of the particular NMC data choice, we perform a variant of our nominal fit with the NMC data on F_2^d/F_2^p replaced with the cross-section data on the deuterium target [33]. This fit results in $\chi^2/\text{d.o.f.} = 4693/3988$. As can be seen in Fig. 4, the difference in the function δf obtained in these two versions of the fit is significant only for $x \lesssim 0.4$. In this region, the uncertainty in the value of δf extracted from the deuterium cross-section data is somewhat larger due to less statistical significance of this sample. Meanwhile, the error bands for the two determinations almost overlap with each other and with the determination based on the heavy-nuclear data [14] (see also Fig. 2). Note also that $\delta f(x)$ obtained in the fit with the NMC deuterium cross-section data is almost identical to our earlier result [19].

B. MARATHON data normalization

The normalization of the ratio σ^d/σ^p in the MARATHON experiment is determined experimentally with a very high accuracy of 0.55%. Nonetheless, the MARATHON data go somewhat lower than the other samples used in our analysis, cf. Fig. 1. To quantify this tension, we perform a variant of fit with the normalization of the MARATHON σ^d/σ^p data released and adjusted simultaneously with other fit parameters. The normalization factor of 1.014(4) obtained in this way is at about 2σ off the nominal value 1.0000(55). However, this renormalization of the MARATHON σ^d/σ^p ratio has a negligible impact on the value of δf extracted from the data, and the corresponding change in its value is well within 1σ uncertainty band. For this fit, we have $\chi^2/\text{d.o.f.} = 4834/4044$.

C. Higher-twist correction

In our analysis, we compute the structure functions following Eq. (1) with an additive model of the higher-twist (HT) terms motivated by the OPE. However, a multiplicative HT model is often used in the literature (see, e.g., Refs. [20,26,51]):

$$F_i(x, Q^2, p^2) = F_i^{\text{TMC}}(x, Q^2, p^2) + F_i^{\text{LT}}(x, Q^2) h_i(x)/Q^2, \tag{10}$$

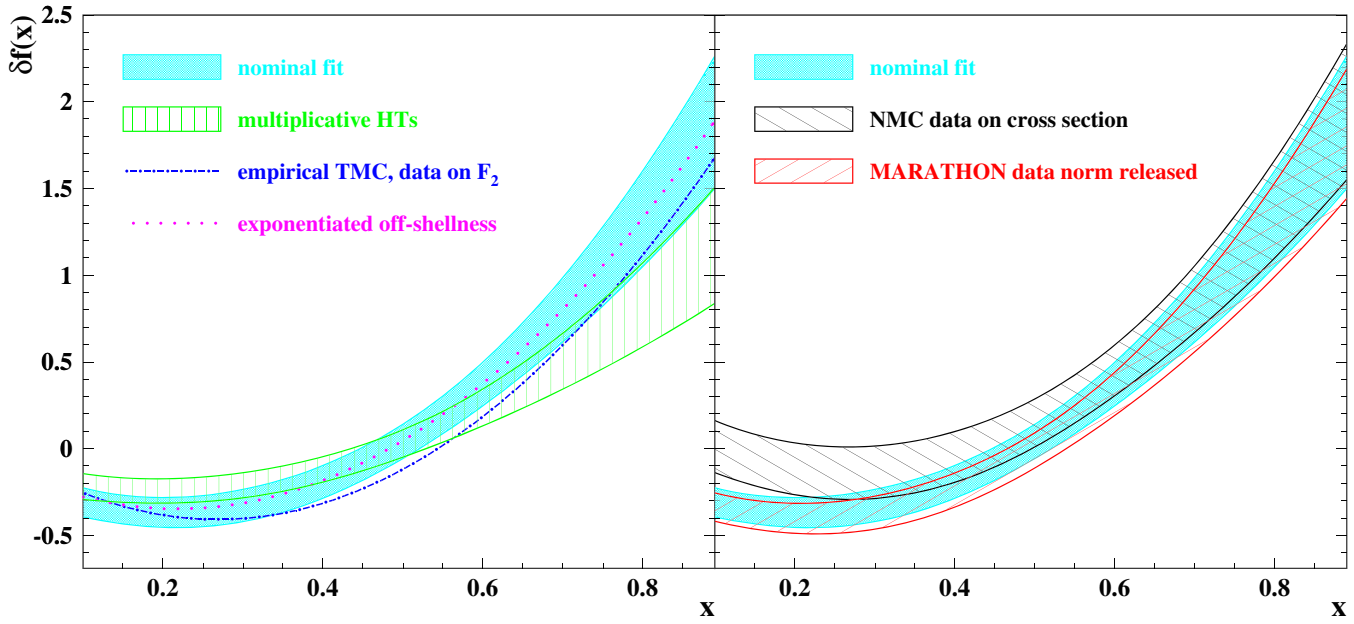

FIG. 4. The 1σ uncertainty band on δf from our nominal fit same as in Fig. 2 (shaded cyan area) compared with the results of a modified fit framework. Left panel: higher twists (HTs) parametrized in a multiplicative ansatz by Eq. (10) (dashed curve); using approximate TMC scheme by Eq. (11) (see Eq. (61) of Ref. [50]) and the data on F_2 (dash-dotted curve); exponential form of the off-shell correction by Eq. (5) (dotted curve). Right panel: using NMC cross-section data [33] instead of NMC F_2^d/F_2^p data [34] in the nominal fit (right-tilted hash area); released normalization of MARATHON σ^d/σ^p data (left-tilted hash area).

where $i = 2, T$. To compare the additive and multiplicative HT models, one should confront the coefficients H_i in Eq. (1) with the corresponding product $F_i^{\mathrm{LT}} h_i$ in Eq. (10). These terms have a different Q^2 dependence driven by assumptions about anomalous dimensions of the HT operators: For the additive form, they are neglected, and for the multiplicative one, they are similar to the leading-twist case. This difference is important at large x, where the leading-twist evolution and the TMC are most significant (for illustration, see Fig. 5). The same trend appears in the determination of the off-shell function, which is sensitive to the assumed HT model at $x \gtrsim 0.5$, although the shape of $\delta f(x)$ (left panel of Fig. 5) does not change essentially under the HT model variation. For the fit with multiplicative HT model, $\chi^2/\mathrm{d.o.f.} = 4798/4044$.

D. Target mass effect

The target mass effects are taken into account in our analysis using the Georgi-Politzer formalism in the off-shell region [see Eq. (2)]. In other global QCD fits, the TMC is either neglected or treated differently, assuming $p^2 = M^2$. For example, in the CJ15 analysis [20,52], the TMC is accounted using an approximation to Eq. (2) [see Eq. (61) in Ref. [50]]:

$$F_2^{\mathrm{TMC}}(x,Q^2) \approx \frac{(1+\gamma)^2}{4\gamma^3} F_2^{\mathrm{LT}}(\xi,Q^2)\left[1+\frac{3(\gamma-1)}{\gamma}(1-\xi)^2\right]. \tag{11}$$

To verify the sensitivity of our results to a particular TMC treatment, we perform a fit using Eq. (11) instead of Eq. (2). A similar approximation for F_T^{TMC} is not available in the formalism of Ref. [50]. For this reason, in our modified fit, we employ data on F_2, instead of cross-section data, for the SLAC, CERN-BCDMS, CERN-NMC, and JLab-E00-116 experiments. The DESY-HERMES and MARATHON data on the cross-section ratio are treated using the relation $\sigma^d/\sigma^p \approx F_2^d/F_2^p$ since $R = \sigma_L/\sigma_T$ is similar for proton and deuteron at moderate x [34]. Finally, for the HERA cross-section data, we take F_T according to Eq. (2a). Such an approach does not cause a serious inconsistency due to the HERA data being localized at small x, where the TMC is generally small. For this variant of the fit, we have $\chi^2/\mathrm{d.o.f.} = 4831/4050$. The best fit result on $\delta f(x)$ is shown by the dashed-dotted curve in Fig. 4, which is within the 1σ band of our nominal fit for almost all x values but a region around $x = 0.4$. In this region, Eq. (11) gives a rather poor approximation on Eq. (2b), as illustrated in Fig. 5 (right panel). The maximal difference between the two implementations is observed in the range $0.2 \le x \le 0.5$. Note that in the same region, we observe opposite deviations between the HT contributions (left panel of Fig. 5) obtained by the CJ15 analysis and by our fit using the same multiplicative HT form and the TMC from Eq. (2).

E. Exponential model of off-shell correction

In our present study, we compute the off-shell correction using Eq. (3). As noted in Sec. II A, at high nucleon

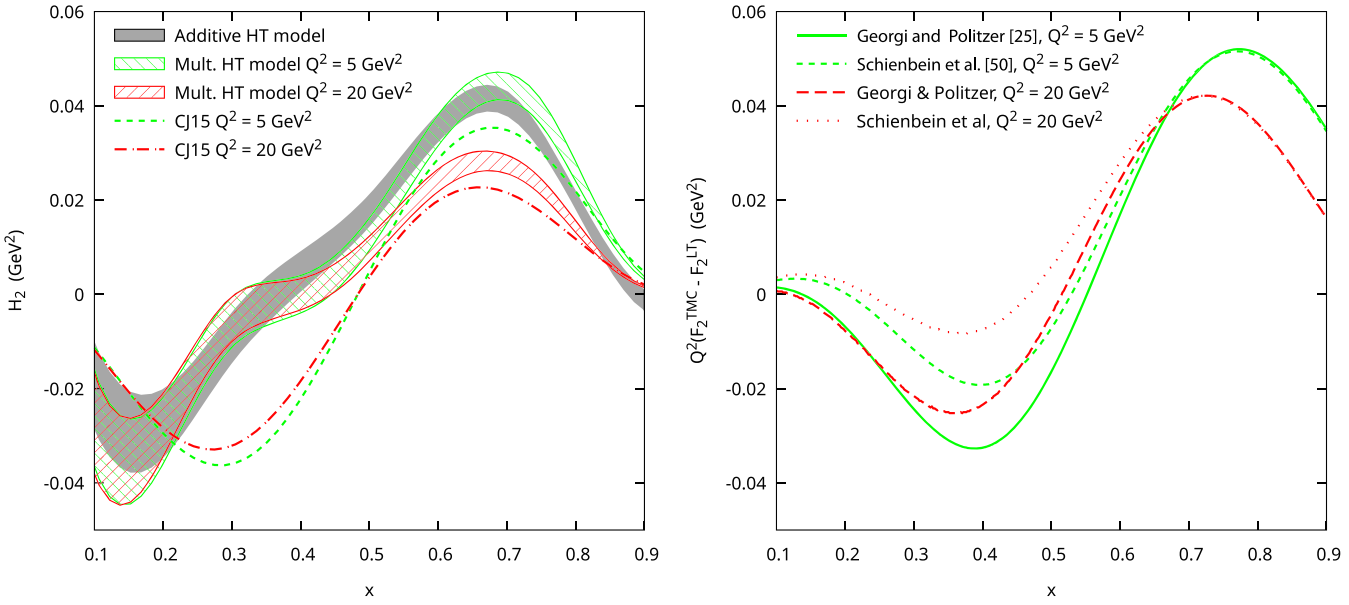

FIG. 5. Left panel: The 1σ bands of twist-four correction to the proton F_2 using the additive HT model [Eq. (1), gray band] and multiplicative HT model [Eq. (10)]. The multiplicative HT is shown for $Q^2 = 5$ (right-tilted band) and 20 GeV^2 (left-tilted band) in order to illustrate their residual Q^2 dependence. Also shown are the corresponding CJ best fit results [20] (dashed and dashed-dotted curves). Right panel: Target mass correction for the proton F_2 in terms of $Q^2[F_2^{\mathrm{TMC}}(x,Q^2) - F_2^{\mathrm{LT}}(x,Q^2)]$ suitable for comparison with HT in the left panel. Georgi-Politzer TMC [25] is shown for $Q^2 = 5$ and 20 GeV^2 (solid and long-dashed curves). The short-dashed and dotted lines show an empirical TMC by Eq. (11), which was used in analysis [20].

momentum $|\boldsymbol{p}| \gtrsim M$ in the nuclear convolution Eq. (6), the off-shell structure function from Eq. (3) may be negative, thus signaling a violation of the linear approximation in v. To verify the relevance of Eq. (3), we performed a fit using the exponential model of off-shell correction from Eq. (5), which gives a positive SF at any value of v. The resulting function $\delta f(x)$, shown in Fig. 4, is identical to our nominal fit result within the fit uncertainties and $\chi^2/\mathrm{d.o.f.} = 4847/4044$ for this variant of the fit.

VI. DISCUSSION

In Sec. IV and V, we discussed determination of the quantity $\delta f(x)$ from a global QCD analysis of the most recent DIS data off hydrogen and deuterium combined with the ones on W- and Z-boson production at hadron colliders. We reiterate that $\delta f(x)$ describes the modification of the nucleon PDFs in the off-shell region for bound nucleons and, as such, it is expected to be a universal quantity independent of the nucleus considered. This quantity has a considerable impact on the nuclear corrections obtained within the nuclear convolution approach and is required to describe available nuclear DIS [14,16] and Drell-Yan data [17,18]. The results from our current analysis are in good agreement with our previous analysis [19], as well as with the study of nuclear ratios of DIS cross sections for nuclear targets with nuclear number $A \geq 3$ [14,16] (Fig. 2). However, our results disagree with the ones of Ref. [20].

In our previous analysis [19], we evaluated the uncertainties associated with the modeling of the deuteron wave function and with the use of different datasets. In Sec. V, we performed additional systematic studies on both the input ansatz and the datasets in order to verify our results and to further investigate the observed discrepancies with the analysis of Ref. [20]. In all cases, our results on δf are stable against the modifications of the fit considered, and the corresponding variations are generally consistent with the quoted uncertainties. We find that while δf has some sensitivity to the implementation of the HT corrections (i.e., additive vs multiplicative) at large Bjorken x, its shape is essentially unchanged (Fig. 4). In general, the systematic uncertainties related to the use of different deuterium datasets, in particular NMC cross-section data vs NMC F_2^d/F_2^p ones, are comparable to the ones related to the input model assumptions.

We further verify our results by comparing our predictions on the ratio $R_{np} = F_2^n/F_2^p$ with the recent MARATHON data [24], which were not included in our fits. The calculations are performed for the MARATHON kinematics, which is roughly consistent with $Q^2 = 14x$ (GeV^2), and are shown in Fig. 6. Our independent predictions for R_{np} are in excellent agreement with the MARATHON measurement over the entire x range available.

In Fig. 6 (left panel), we show the predictions on R_{np} obtained from the QCD analyses CJ15, MSHT20, NNPDF4.0, CT18, and JAM21. The structure functions are computed as described in Sec. IV. The calculations include the target mass corrections of Ref. [25]. The CJ15 prediction is obtained from Fig. 3 of Ref. [24]. At large Bjorken x ($x > 0.6$), significant differences are observed.

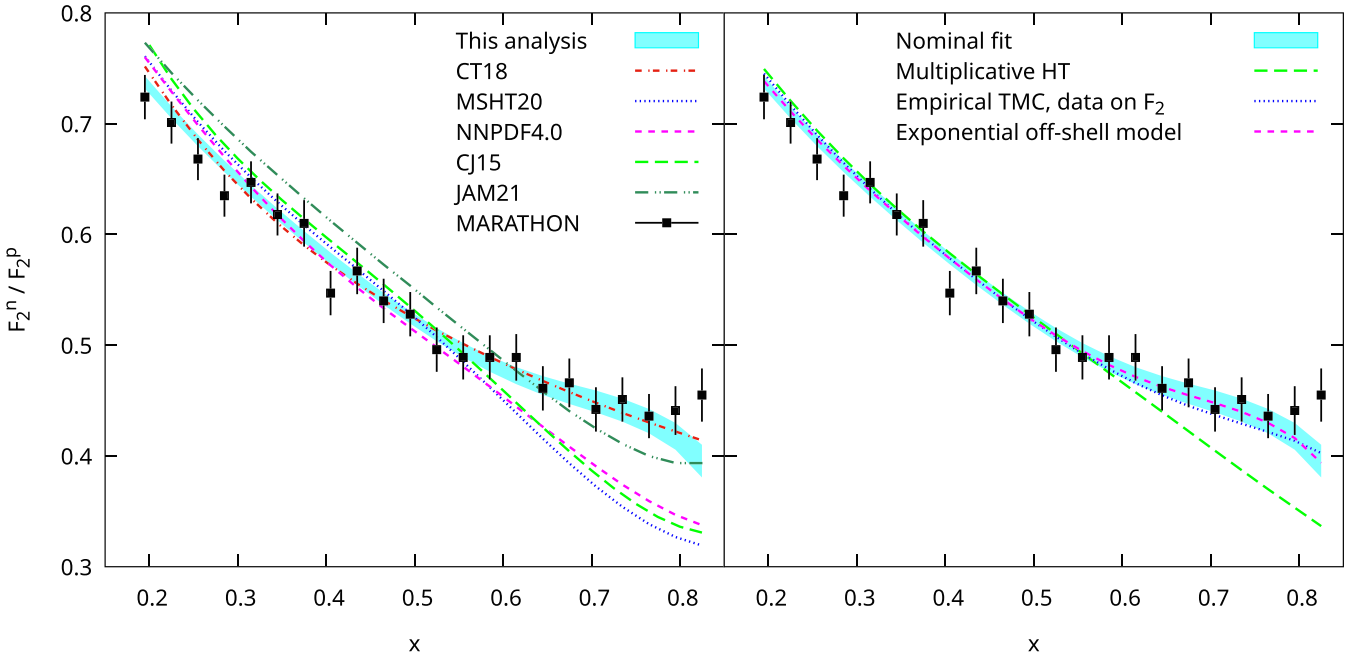

FIG. 6. Data from the MARATHON measurement of F_2^n/F_2^p [24] compared with the predictions of the present analysis. In the left panel, shown are the predictions from recent PDF analyses: CT18 [21] (dash-dotted line), MSHT20 [22] (dotted line), NNPDF4.0 [23] (short dashes), CJ15 [20] (long dashes), and JAM21 [51] (dash-double-dot). In the right panel, shown are our predictions based on modified fit results discussed in Sec. V.

While the predictions from CJ15, MSHT20, and NNPDF4.0 are consistent with each other, they differ substantially from CT18 and our results. Note that different assumptions on the HT contributions are used for the various calculations: additive HT for our results, multiplicative HT for CJ15, and no HT (only LT term) for the others. To verify the impact of the d-quark PDF on such differences, we compare the corresponding predictions for the $R_{du} = d/u$ ratio for the MARATHON kinematics using the PDFs from the LHAPDF library [49].[6] Figure 7 shows our 1σ uncertainty band together with the central values obtained for the other analyses. The uncertainties on these latter are relatively large for $x > 0.6$ due to the use of tight W cuts, which effectively exclude high-x DIS data from the QCD analysis.[7] Figure 6 indicates that the CJ15-, MSHT20-, and NNPDF4.0-based predictions are in clear disagreement with the MARATHON R_{np} data.

The effect of the variations of the model assumptions discussed in Sec. V on our predictions for R_{np} is also illustrated in Fig. 6 (right panel). Although in most cases, the results are consistent with the nominal one within the quoted uncertainties, a significant deviation is apparent for the variant with multiplicative HT contributions from Eq. (10). In this latter case, our predictions for R_{np} appear to be closer to the CJ15-, MSHT20-, and NNPDF4.0-based calculations. However, the corresponding R_{du} ratio shown in Fig. 7 (right panel) is consistent with our nominal fit. These observations indicate that the MARATHON data are sensitive to the HT contributions in the region $x > 0.6$.

The comparison with the CT18-based predictions is instructive, as they provide a good LT description of the MARATHON data on R_{np} (Fig. 6) without any HT contributions. This agreement is explained by a much larger value of the R_{du} ratio at $x > 0.6$ compared to the other QCD analyses, as illustrated in Fig. 7 (left panel).[8] By contrast, we obtain a good description of MARATHON data on R_{np} with $R_{du} \to 0$ as $x \to 1$ and a sizable HT contribution, which is maximal at $x \sim 0.7$ (Fig. 5). In general, $W^{\pm}$ boson production and the corresponding lepton asymmetries from D0 and LHCb data at high rapidity could help to clarify the differences observed on the d-quark distribution at large Bjorken x [53,54]. However, the calculation of the cross section for W-boson production in the NNLO pQCD approximation suffers from uncertainties in the available numerical codes [55].

As discussed in Sec. II A, we assume isoscalar HT terms in the additive HT model ($H^p = H^n$). We also assume

[6]As noted in Ref. [2], the nuclear correction at large x also affects the valence d quark PDF at small values of $x \approx 0.03$ because of fermion number conservation. While this effect is relevant for the electroweak studies at LHC, we leave its analysis for future, as in the present work, we focus on the region relevant for the MARATHON measurement.

[7]As an illustration, the CT18 uncertainty on R_{du} is about 100% as can be seen in Fig. 9 of Ref. [21].

[8]We note that the CT18 analysis utilizes neutrino data from heavy nuclear targets. The value of R_{du} is significantly reduced at large x if neutrino data are removed from the analysis [2].

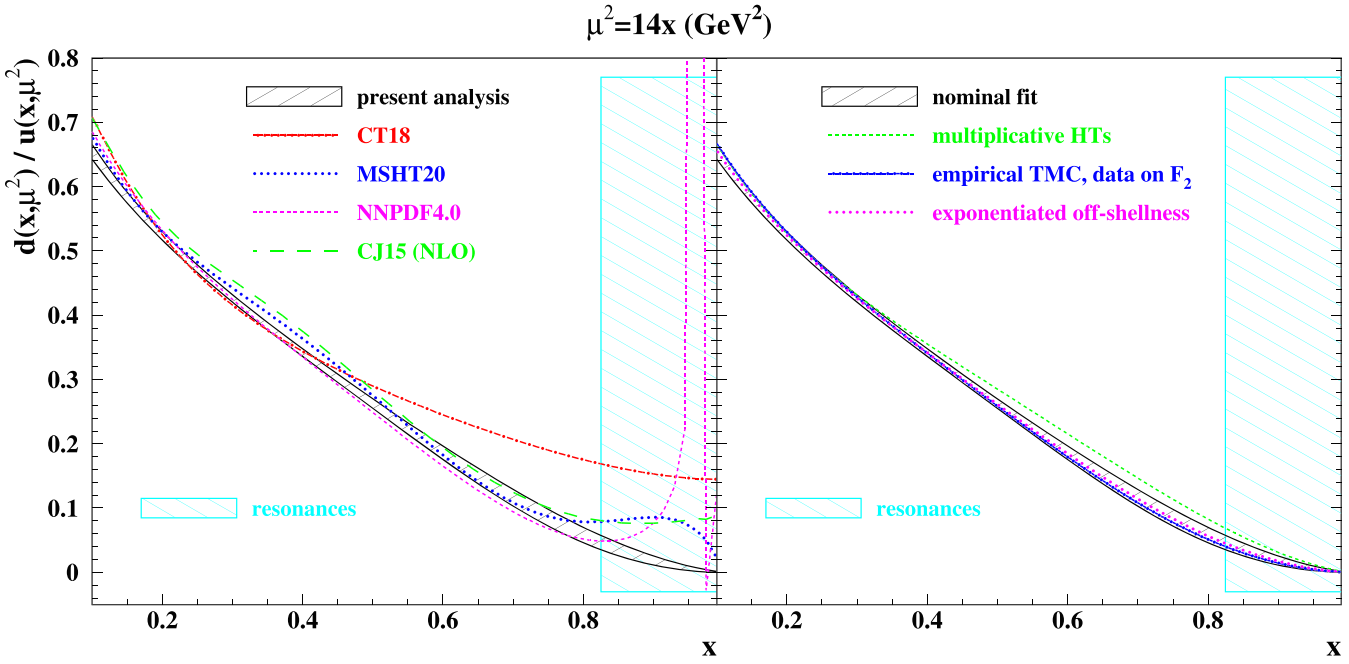

FIG. 7. The 1σ band of the PDF ratio d/u vs x computed at variable scale $\mu^2 = 14x$ (GeV^2) (left-tilted area). Also shown are the results from PDF analyses, same as in Fig. 6 (left panel). The right-tilted (yellow) area shows the nucleon resonance region of $W \lesssim 2$ GeV. All shown analyses but Ref. [20] are to NNLO pQCD order. Right panel shows our results from modified fits, same notations as in Fig. 6.

$h^p = h^n$ for the multiplicative HT model. However, in the latter case, the overall HT correction is different for protons and neutrons since h is multiplied by the corresponding LT structure functions. Note that for the multiplicative HT model, the contribution from h cancels out in the ratio R_{np}, which has a similar behavior as the corresponding LT approximation. Conversely, in the case of the additive HT model used in our nominal fit the ratio, R_{np} receives a finite HT correction. The MARATHON R_{np} data seem to prefer a common additive HT contribution (Fig. 6) for both the neutron and the proton. Although the disagreement observed for $x > 0.6$ with the multiplicative HT form may be mitigated by the introduction of an explicit isospin dependence in the h terms, such an effect could result in observable deviations for other DIS data sensitive to isospin effects.

A recent paper [51] reports the results of a global QCD analysis (JAM21) including MARATHON data on the cross-section ratios σ^d/σ^p and $\sigma^{3\mathrm{H}}/\sigma^{3\mathrm{He}}$ for the three-body nuclei, as well as the previous measurement of $\sigma^{3\mathrm{He}}/\sigma^d$ from E03-103 at JLab [56]. The study includes multiplicative HT corrections and a calculation of nuclear effects based on the convolution approach supplemented by off-shell corrections. However, the treatment of these latter corrections is rather different with respect to our implementation. In the JAM21 analysis, the off-shell correction depends on both the specific nucleus considered and on the isospin of the target nucleon (different for protons and neutrons), resulting in multiple functions that are extracted from data. In particular, the need of an explicit isovector contribution in the off-shell functions is advocated to describe the MARATHON data. Our analysis indicates that this result may be affected by the assumption of multiplicative HT terms. The result of the JAM21 fit on R_{np} is shown in the left panel of Fig. 6 and appears to be in disagreement with the MARATHON F_2^n/F_2^p data [24].

The determination of δf described in this paper is based on deuterium DIS data and is therefore only sensitive to the isoscalar combination $F_2^p + F_2^n$. Our results are in a good agreement with the study of the ratios of nuclear DIS cross sections σ^A/σ^d with nuclear number $A \geq 3$, in which the nuclear EMC effect was successfully described in terms of a nuclear convolution approach with a universal off-shell function $\delta f(x)$ independent of the nucleus. Although the model [14] could naturally incorporate an isospin dependence into the off-shell correction, the good agreement with data on nonisoscalar nuclei obtained using the same off-shell function $\delta f(x)$ for the proton and neutron [14,16] seems to indicate that potential isospin dependence of δf is small. Dedicated studies of nuclear effects using DIS data from mirror nuclei ^{3}H and ^{3}He [24] and upcoming DIS data on proton and deuterium from JLab12 [57] could provide new insights on the origin of modification of parton structure in bound nucleons, as well as improved constraints on nucleon d-quark distribution at large x and on the isospin dependence of HT corrections. A more comprehensive study of these effects would require future data from high-energy processes, which can provide a flavor selection like the hadronic Drell-Yan reaction or DIS using both the electron and (anti)neutrino charged-current (CC) process. To this end, the availability of precision measurements at future electron-ion collider [58] and of both neutrino

and antineutrino CC interactions off hydrogen and various nuclear targets [59,60] at the long-baseline neutrino facility could provide valuable insights.

ACKNOWLEDGMENTS

We would like to thank A. Accardi for useful discussions and communications on the CJ15 results, G. G. Petratos for useful communications on the MARATHON results, M. V. Garzelli and S.-O. Moch for useful discussions and a careful reading of the manuscript, and Fr. Kok for the help with manuscript preparation. The work of S. A. is supported by the DFG grants MO 1801/5-1, KN 365/14-1. R. P. is supported by Grant No. DE-SC0010073 from the Department of Energy, USA.

APPENDIX A: COMPARISONS OF THE FIT RESULTS WITH DATA

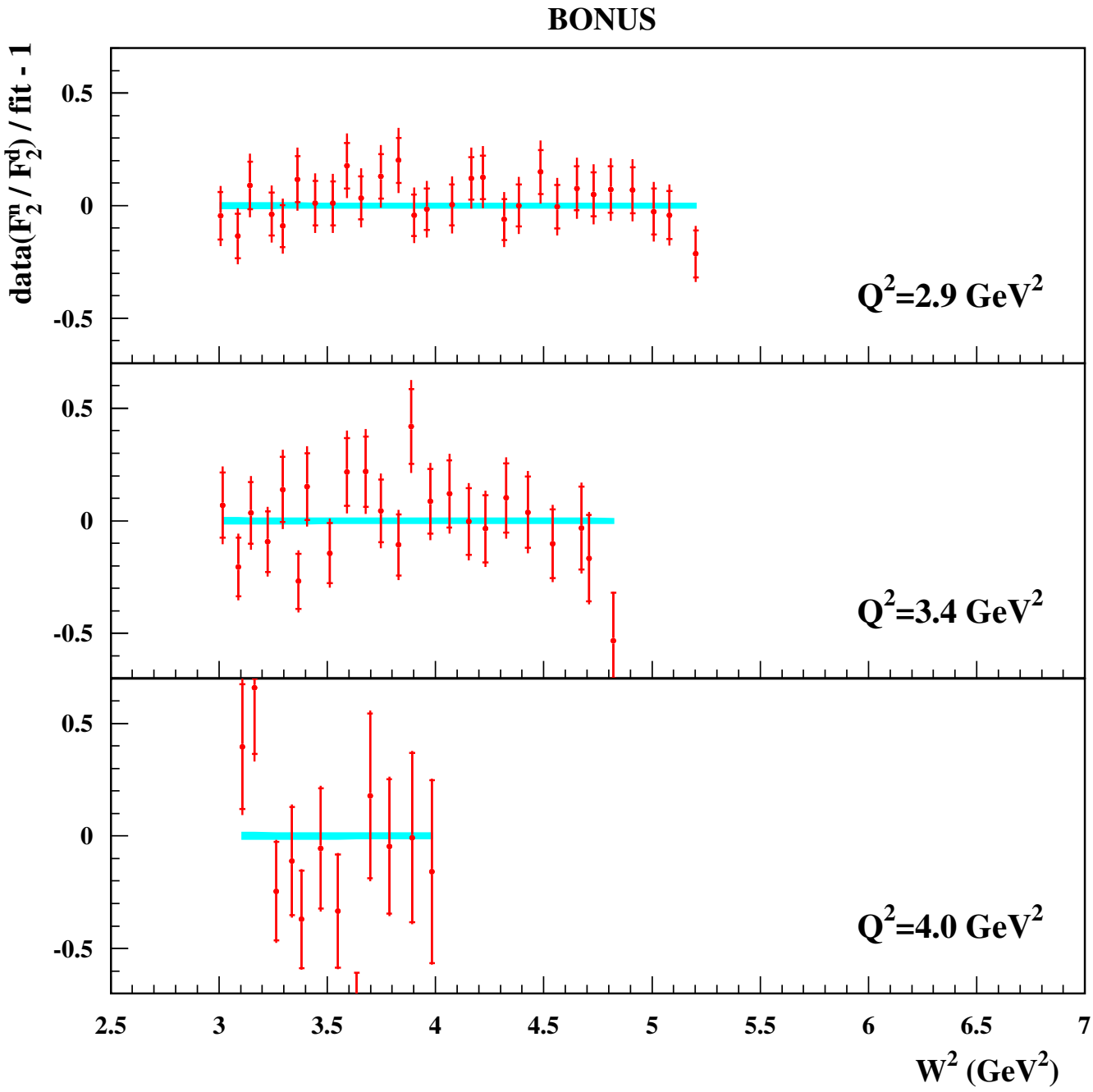

FIG. 8. The pulls of the present analysis for the BoNuS data on the ratio F_2^n/F_2^d [29] vs W^2 displayed in the panels of Q^2 bins. The inner error bars reflect the statistical and uncorrelated systematic uncertainties of data, while the outer bars are the total experimental error. The shaded region shows $\pm 1\sigma$ uncertainty of the fit.

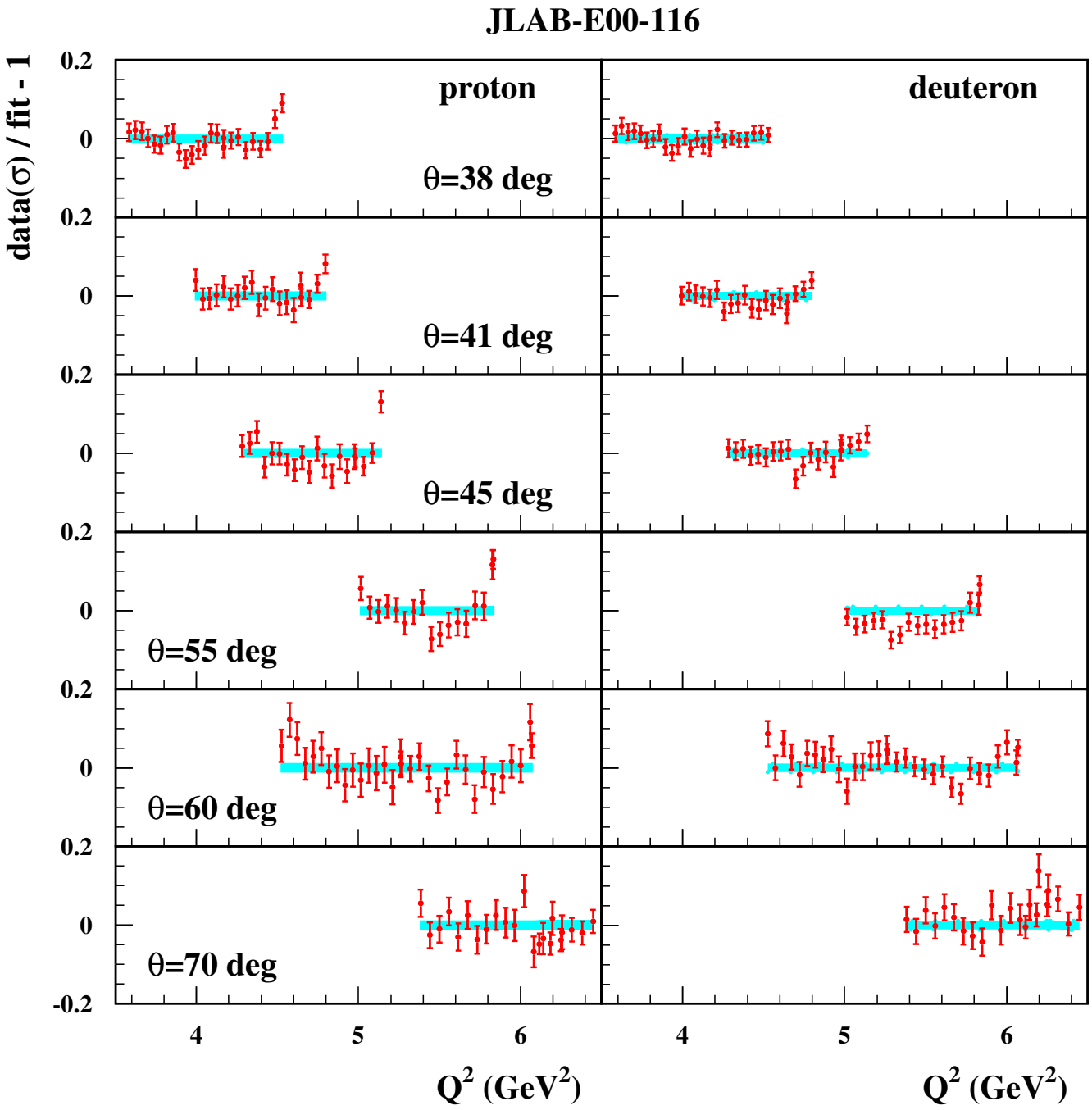

FIG. 9. The pulls of the JLab-E00-116 data [31] on $\frac{d^2\sigma}{dE'd\Omega}$ for the proton (left panels) and the deuteron (right panels) targets vs Q^2. The panels correspond to the bins of the electron scattering angle θ. Notations are similar to those in Fig. 8.

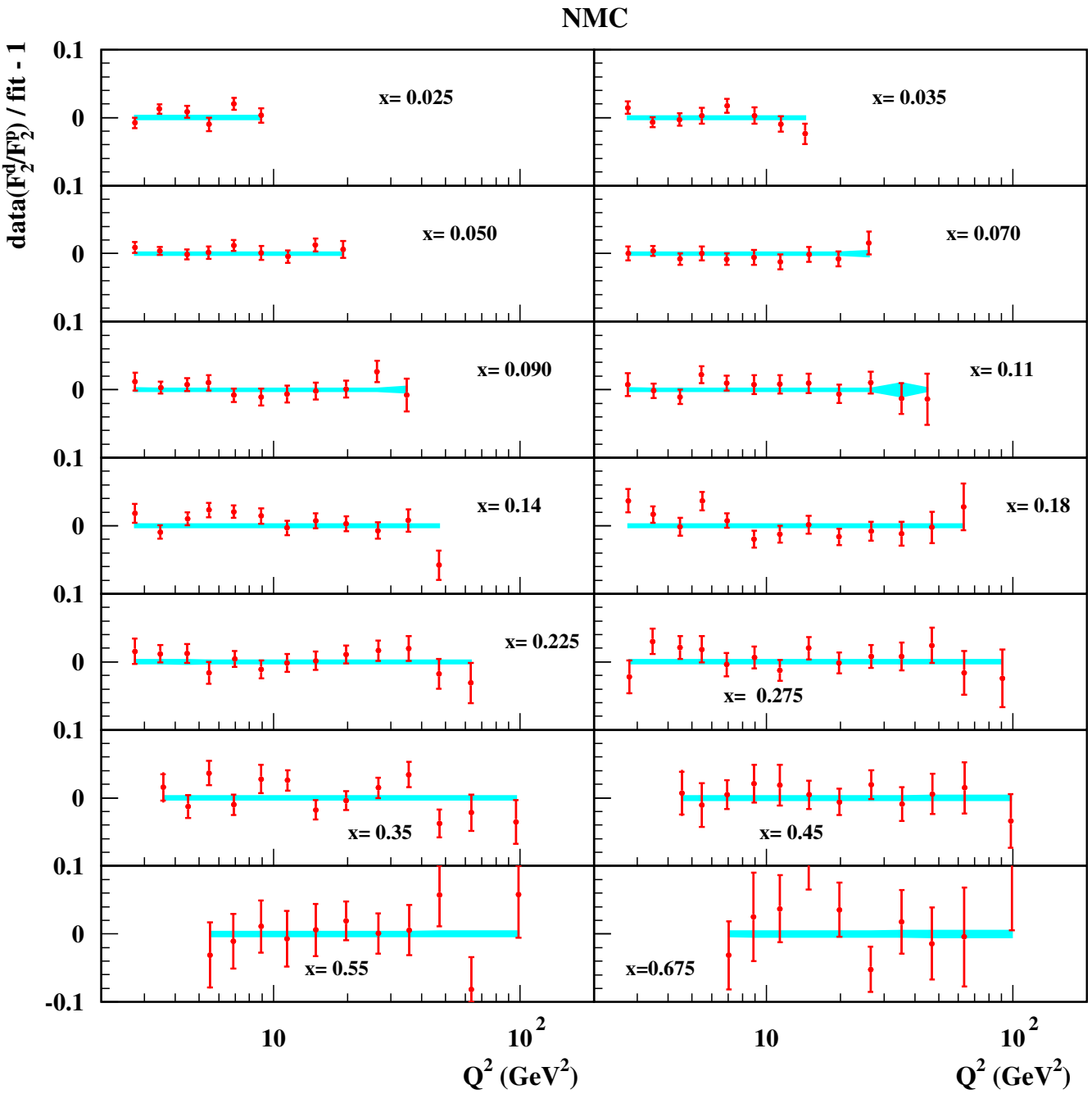

FIG. 10. The pulls of the NMC data [34] on F_2^d/F_2^p vs Q^2. The panels correspond to the bins of x. Notations are similar to those in Fig. 8.

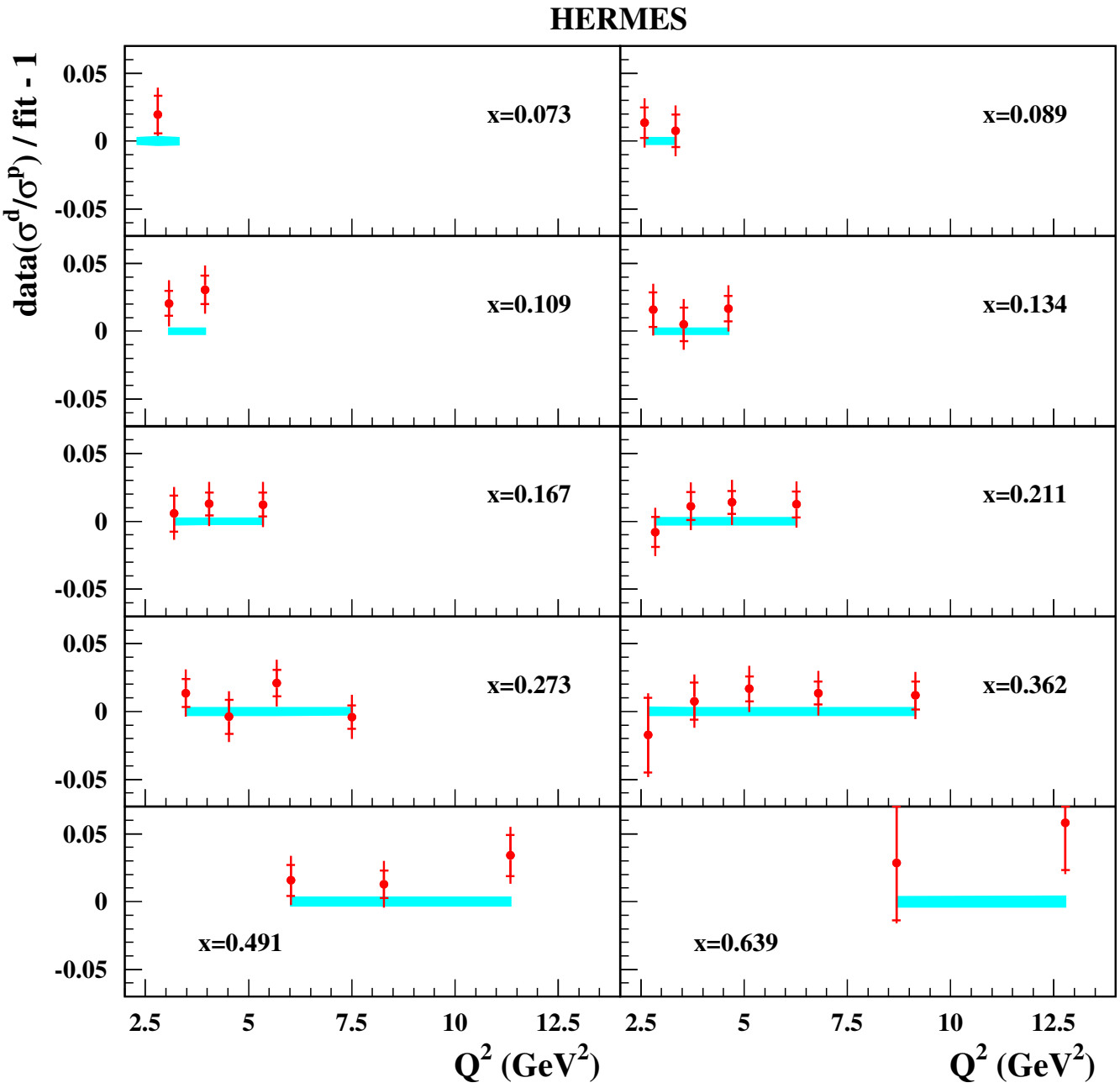

FIG. 11. The pulls of the DESY-HERMES data [32] on σ^d/σ^p vs Q^2. The panels correspond to the bins of x. Notations are similar to those in Fig. 10.

APPENDIX B: PHASE SPACE INTEGRATION IN THE NUCLEAR CONVOLUTION

Here, we discuss in more detail the integration in Eq. (6) for general kinematics of x and Q^2. Recall that the integration region is constrained by the condition,

$$W^2 \geq W_{\rm th}^2, \tag{B1}$$

where $W^2 = (p+q)^2$ and p is the four-momentum of the bound nucleon, and $W_{\rm th}$ is the threshold mass. The pion production threshold corresponds to $W_{\rm th} = M + m_\pi$, and by setting $W_{\rm th} = M$, we also include the elastic channel. Because of the energy-momentum conservation, $p = p_d - p_S$, where p_d is the deuteron four-momentum, and $p_S = (\sqrt{m_S^2 + \boldsymbol{p}^2}, -\boldsymbol{p})$, the four-momentum of the spectator system with the mass m_S.[9] In Appendix A of Ref. [14], the nuclear convolution integral with the constraint (B1) was considered for nonrelativistic spectator

[9]For scattering off the deuteron, $m_S = M$. In case of scattering off a nucleus of A nucleons, m_S is the mass of the residual nucleus of $A-1$ nucleons.

assuming $\sqrt{m_S^2+\boldsymbol{p}^2} = m_S + \boldsymbol{p}^2/(2m_S)$. This approximation makes sense as the deuteron is a weakly bound system, and most of the momentum distribution is in the nonrelativistic region. However, the high-momentum part with $|\boldsymbol{p}|$ of order of a few hundred MeV requires a relativistic analysis. Here, we discuss the fully relativistic case of spectator kinematics that would allow us to better describe the contribution from the high-momentum region of the spectator.

In terms of the four-vectors p_d, q, and p_S, we can write Eq. (B1) as follows:

$$(p_d+q)^2 + m_S^2 - 2(p_d+q)\cdot p_S \geq W_{\rm th}^2. \tag{B2}$$

In order to facilitate the discussion of Eq. (B2), we use the following notations:

$$S = (p_d+q)^2 = M_d^2 + Q^2(1/x_d - 1), \tag{B3a}$$

$$E = M_d + q_0 = \sqrt{S+\boldsymbol{q}^2}, \tag{B3b}$$

$$\alpha = (S + m_S^2 - W_{\rm th}^2)/(2Em_S), \tag{B3c}$$

$$\beta = |\boldsymbol{q}|/E, \tag{B3d}$$

where S and E are, respectively, the invariant mass squared and the energy of the virtual photon–deuteron system, and $x_d = Q^2/(2p_d\cdot q)$ is the natural Bjorken variable for the deuteron. As it follows from the definitions in Eq. (B3), $\alpha > 0$, and $0 < \beta < 1$ at any finite Q^2 value. In the limit $Q^2 \to \infty$, we have $\beta = 1$ and $\alpha = (1-x_d)M_d/m_S$. Using Eq. (B3), we can write Eq. (B2) as follows:

$$\alpha m_S - \sqrt{m_S^2+\boldsymbol{p}^2} + \beta p_z \geq 0. \tag{B4}$$

For completeness, we also give here Eq. (B4) for nonrelativistic spectator kinematics,

$$2(\alpha-1)m_S^2 - \boldsymbol{p}^2 + 2\beta m_S p_z \geq 0. \tag{B5}$$

Below we discuss the solution to Eqs. (B4) and (B5) in terms of both the spherical coordinates and the $(p_z,\boldsymbol{p}_\perp)$ basis for the momentum $\boldsymbol{p}$.

1. Convolution integral using spherical coordinates

We consider Eqs. (B4) and (B5) in spherical coordinates, in which $p_z = |\boldsymbol{p}|\cos\theta$ with θ the zenith angle. Both Eqs. (B4) and (B5) have two nodes, for which we will use the notation $p_\pm(\cos\theta)$. For Eq. (B4), we have

$$p_\pm(\cos\theta) = m_S\frac{\alpha\beta\cos\theta \pm \sqrt{\beta^2\cos^2\theta+\alpha^2-1}}{1-\beta^2\cos^2\theta}, \tag{B6}$$

while for Eq. (B5), we have

$$p_\pm(\cos\theta) = m_S(\beta\cos\theta \pm \sqrt{\beta^2\cos^2\theta + 2(\alpha-1)}). \tag{B7}$$

In solving the inequalities (B4) and (B5), it is convenient to consider the cases $\alpha \leq 1$ and $\alpha > 1$. As a result, the solution involves two different regions:

$$\begin{cases} -1 \leq \cos\theta \leq 1 \\ 0 \leq |\boldsymbol{p}| \leq p_+(\cos\theta) \end{cases}, \qquad \text{for } \alpha > 1, \tag{B8}$$

and

$$\begin{cases} c \leq \cos\theta \leq 1 \\ p_-(\cos\theta) \leq |\boldsymbol{p}| \leq p_+(\cos\theta) \end{cases}, \qquad \text{for } \alpha_0 \leq \alpha \leq 1. \tag{B9}$$

The parameters α_0 and c are different for Eqs. (B4) and (B5). For relativistic kinematics $c = \sqrt{1-\alpha^2}/\beta$ and $p_\pm(\cos\theta)$ are given by Eq. (B6), while for the nonrelativistic spectator, $c = \sqrt{2(1-\alpha)}/\beta$ and $p_\pm(\cos\theta)$ are given by Eq. (B7). The minimum value of α in Eq. (B9) is derived from the condition $c = 1$. We have $\alpha_0 = \sqrt{1-\beta^2}$ and $\alpha_0 = 1 - \frac{1}{2}\beta^2$ for Eqs. (B4) and (B5), respectively.

Also, the condition $c = 1$ determines the maximum allowed value of x_d for given Q^2, which is consistent with Eq. (B1). For $Q^2 \to \infty$, we have $x_d^{\max} = 1$ and $x_d^{\max} = 1 - M/(2M_d) \approx 3/4$ for the relativistic and nonrelativistic spectator kinematics, respectively. For finite values of Q^2, we have in case of Eq. (B4),

$$\begin{aligned} x_d^{\max} &= (1 + ((M+W_{\rm th})^2 - M_d^2)/Q^2)^{-1} \\ &\approx (1 + 4m_\pi M/Q^2)^{-1}. \end{aligned} \tag{B10}$$

We use Eqs. (B8) and (B9) to cast the momentum integral in Eq. (6) as follows:

$$\int \mathrm{d}^3\boldsymbol{p}|\Psi_d(\boldsymbol{p})|^2\theta(W^2 - W_{\rm th}^2)$$

$$= \begin{cases} \frac{1}{2}\int_{-1}^{1}\mathrm{d}\cos\theta\int_0^{p_+(\cos\theta)}\mathrm{d}p\,p^2(\psi_0^2(p)+\psi_2^2(p)), & \text{for } \alpha > 1, \\ \frac{1}{2}\int_{c}^{1}\mathrm{d}\cos\theta\int_{p_-(\cos\theta)}^{p_+(\cos\theta)}\mathrm{d}p\,p^2(\psi_0^2(p)+\psi_2^2(p)), & \text{for } \alpha \leq 1, \end{cases} \tag{B11}$$

where $p = |\boldsymbol{p}|$, and ψ_0 and ψ_2 are the deuteron orbital momentum wave functions for $l = 0$ and $l = 2$, respectively:

$$|\Psi_d(\boldsymbol{p})|^2 = (\psi_0^2(p) + \psi_2^2(p))/(4\pi). \tag{B12}$$

Following Eq. (7), the functions ψ_0 and ψ_2 are normalized as follows:

$$\int_0^\infty \mathrm{d}p p^2(\psi_0^2+\psi_2^2)=1. \tag{B13}$$

Note that in numerical applications, we apply a cut on the bound nucleon momentum $p_{\mathrm{cut}}\sim 1$ GeV in the nuclear convolution. For this reason, we replace the upper limit on the momentum in Eq. (B11) with $\min(p_{\mathrm{cut}}, p_+(\cos\theta))$. The integration region in Eq. (B11) is illustrated in Fig. 12 for both the relativistic and the nonrelativistic spectator and for a few fixed values of x and Q^2. The integration region systematically shrinks with rising x, and the allowed kinematical region is somewhat larger for the relativistic case, although the difference is only visible for high nucleon momenta $p>0.5$ GeV. As a somewhat extreme example of the deuteron kinematics, in the last panel of Fig. 12, we show the integration region for $x=1.3$, which is limited to high values of $\cos\theta$ and momentum $p>300$ MeV.

2. Convolution integral using $(y,\boldsymbol{p}_\perp)$ basis

Consider Eq. (6) and note that $x'=x/y$, where the dimensionless variable $y=p\cdot q/(Mq_0)=(p_0+\gamma p_z)/M$ is usually referred to as the nucleon light-cone momentum. The integral over the nucleon momentum in Eq. (6) can be cast in terms of integration over y and $p_\perp$:

$$F_i^d(x,Q^2)=\int \mathrm{d}y\mathrm{d}p_\perp^2 d_{ij}(y,p_\perp^2,\gamma)F_j^N(x/y,Q^2,\mu^2), \tag{B14}$$

where $i=T,2$, and we assume the sum over the repeated subscript $j=T,2$ and

$$d_{ij}(y,p_\perp^2,\gamma)=\pi\int \mathrm{d}p_z|\Psi_d(\boldsymbol{p})|^2K_{ij}\delta\left(y-\frac{p_0+\gamma p_z}{M}\right), \tag{B15}$$

and $\mu^2=p_0^2-\boldsymbol{p}^2$, and p_0 is the energy of the active nucleon, and the kinematical factors K_{ij} in Eq. (B14) are given by Eq. (8).

Note that the kernel d_{ij} in the convolution integral Eq. (B14) depends on the kinematic variables x and Q^2 through a dimensionless parameter $\gamma=(1+4x^2M^2/Q^2)^{1/2}$. We now briefly consider the $\gamma=1$ case corresponding to the

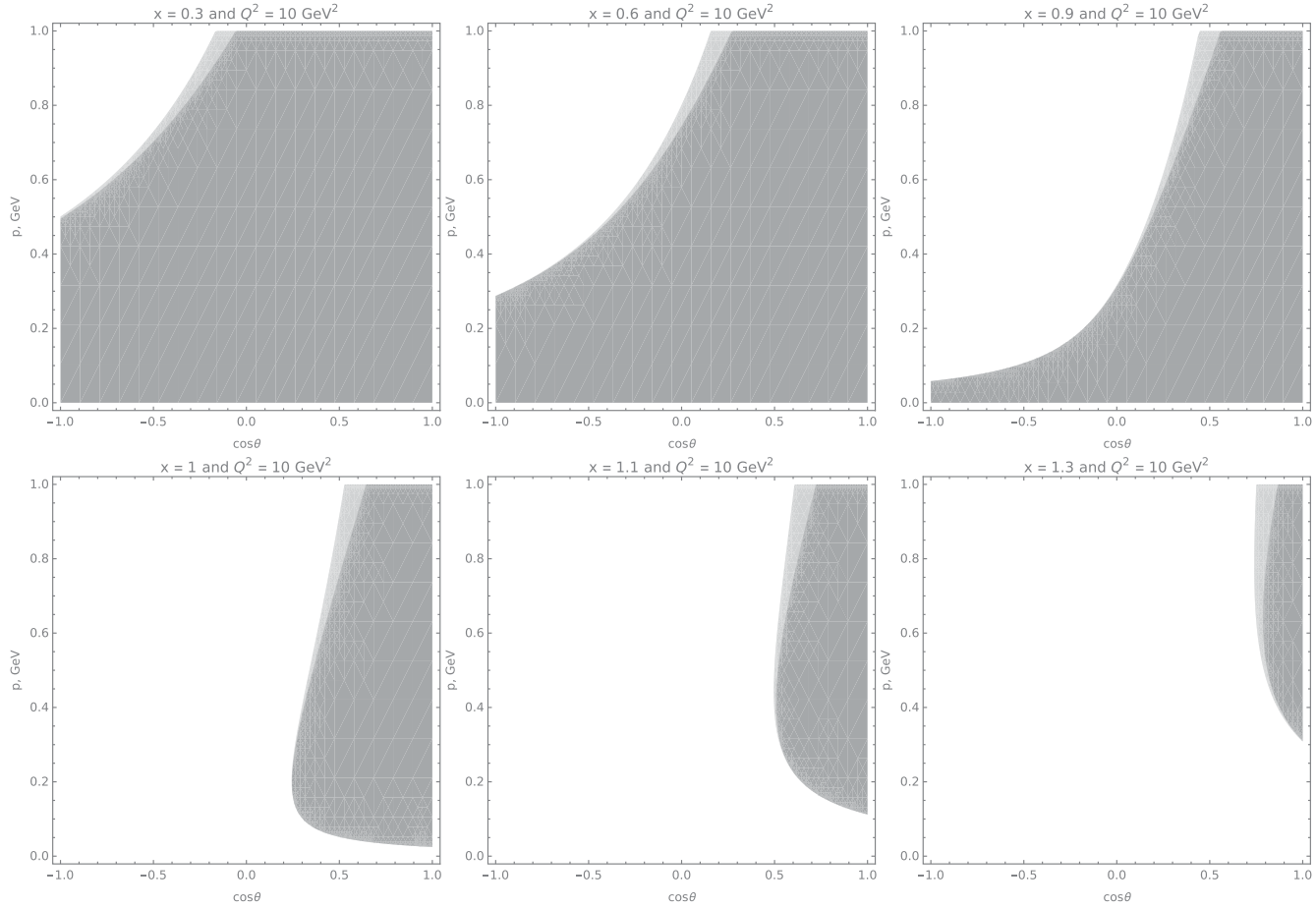

FIG. 12. The shaded area shows the region constrained by Eqs. (B8) and (B9) together with the cut $p<1$ GeV for a few fixed values of x and Q^2 shown in the panels. The light-gray region corresponds to a relativistic spectator, while the dark-gray is for the nonrelativistic one.

light-cone kinematics of $Q^2 \to \infty$. In this limit, the matrix K_{ij} has only the diagonal components, and $K_{TT} = K_{22} = 1 + p_z/M$. Then the nuclear convolution for F_T and F_2 have the same form with the kernel $d_{TT} = d_{22} = D(y, p_\perp^2)$, which has the meaning of distribution over the corresponding variables:

$$D(y, p_\perp^2) = \pi \int \mathrm{d}p_z |\Psi_d(\boldsymbol{p})|^2 \left(1 + \frac{p_z}{M}\right) \delta\left(y - \frac{p_0 + p_z}{M}\right). \tag{B16}$$

The distribution by Eq. (B16) is normalized to 1:

$$\int \mathrm{d}y \mathrm{d}p_\perp^2 D(y, p_\perp^2) = \int \mathrm{d}^3\boldsymbol{p} |\Psi_d(\boldsymbol{p})|^2 \left(1 + \frac{p_z}{M}\right) = 1. \tag{B17}$$

The term proportional to p_z vanishes after angular integration.

Note that in the off-shell region, the nucleon SF in Eq. (B14) depends on the virtual nucleon mass square μ^2. We use Eq. (3) in order to separate the off-shell dependence of the bound nucleon structure function and integrate over $p_\perp^2$. Then Eq. (B14) can be cast in terms of a one-dimensional convolution integral as follows:

$$F_2^d(x, Q^2) = \int_x^{y_{\max}} \mathrm{d}y [S_0(y) F_2^N(x/y, Q^2) + S_1(y) \delta f(x/y) F_2^N(x/y, Q^2)], \tag{B18}$$

where the light-cone smearing functions S_0 and S_1 are as follows:

$$S_0(y) = \int \mathrm{d}p_\perp^2 D(y, p_\perp^2), \tag{B19}$$

$$S_1(y) = \int \mathrm{d}p_\perp^2 D(y, p_\perp^2) v, \tag{B20}$$

where $v = (\mu^2 - M^2)/M^2$ is the nucleon virtuality. The function $S_0(y)$ makes sense of the nucleon light-cone distribution in the deuteron and normalized to unity according to Eq. (B17). Note that Eq. (B18) was derived for $\gamma = 1$, i.e., light-cone kinematics. In this limit, the constraint by Eq. (B1) reduces to $y > x$ and $y_{\max} = M_d/M$.

Below we discuss in more detail the nuclear convolution by Eqs. (B14) and (B15) for both the relativistic and nonrelativistic kinematics of the nucleon spectator while keeping finite Q^2 effects.

a. Nonrelativistic spectator

We first consider Eq. (B15) assuming the *nonrelativistic* nucleon with energy $p_0 = M + \varepsilon_d - \boldsymbol{p}^2/(2M)$, where $\varepsilon_d = M_d - 2M$ is the deuteron binding energy. Taking the integral in Eq. (B15), we have

$$\int \mathrm{d}p_z \delta\left(y - \frac{p_0 + \gamma p_z}{M}\right) = \frac{M^2}{\sqrt{t^2 - p_\perp^2}}, \tag{B21}$$

where

$$t^2 = 2M^2(y_{\max} - y), \tag{B22}$$

$$y_{\max} = 1 + \frac{\gamma^2}{2} + \frac{\varepsilon_d}{M}. \tag{B23}$$

Note also that by integrating the δ function in Eq. (B21), we have p_z as a function of y and $p_\perp^2$:

$$p_z = \gamma M - \sqrt{t^2 - p_\perp^2}. \tag{B24}$$

Note that t makes sense of the maximum $p_\perp$ for the given y. The condition $t^2 = 0$ determines the maximum value $y = y_{\max}$; see Eq. (B23). Note that for $\gamma = 1$ and neglecting a small correction due to the deuteron binding energy in Eq. (B23), we have $y_{\max} = 3/2$. This is different from the kinematical maximum $y_{\max} = M_d/M$ in the relativistic case which is discussed in Sec. B 2 b and illustrated in Fig. 13.

Using Eqs. (B21)–(B24), we cast Eq. (B14) as follows:

$$F_i^d(x, Q^2) = \frac{M^2}{4} \int_{y_{\min}}^{y_{\max}} \mathrm{d}y \int_0^{t^2} \mathrm{d}p_\perp^2 \frac{(\psi_0^2(p) + \psi_2^2(p))}{\sqrt{t^2 - p_\perp^2}} \times K_{ij} F_j^N(x/y, Q^2, \mu^2), \tag{B25}$$

where $p = \sqrt{p_z^2 + p_\perp^2}$ and p_z is given by Eq. (B24). Note the $p_\perp^2$ integration in Eq. (B25) has a singularity at $p_\perp^2 = t^2$. Although this is an integrable singularity, it may cause an instability in numerical applications. For this reason, it is convenient to change the integration variable in Eq. (B25) from $p_\perp^2$ to $u = \sqrt{t^2 - p_\perp^2}$. Then we have

$$F_i^d(x, Q^2) = \frac{M^2}{2} \int_{y_{\min}}^{y_{\max}} \mathrm{d}y \int_0^{t} \mathrm{d}u (\psi_0^2(p) + \psi_2^2(p)) K_{ij} F_j^N(x/y, Q^2, \mu^2), \tag{B26}$$

where $p_\perp^2 = t^2 - u^2$, $p_z = \gamma M - u$, $p = \sqrt{p_\perp^2 + p_z^2}$, and $\mu^2 = M^2 + 2M\varepsilon_d - 2p^2$. The lower limit of integration over the light-cone variable is $y_{\min} = x$. Note, however, that this integration region in the nuclear convolution is modified for finite Q^2. The corresponding region can be inferred from Eq. (B1). Unlike the case of spherical coordinates discussed in Sec. B 1, the analytic solution to the inequality (B1) in terms of $(y, p_\perp)$ is somewhat cumbersome and not shown here. The resulting integration region in the nuclear convolution Eq. (B25) is illustrated in Fig. 13.

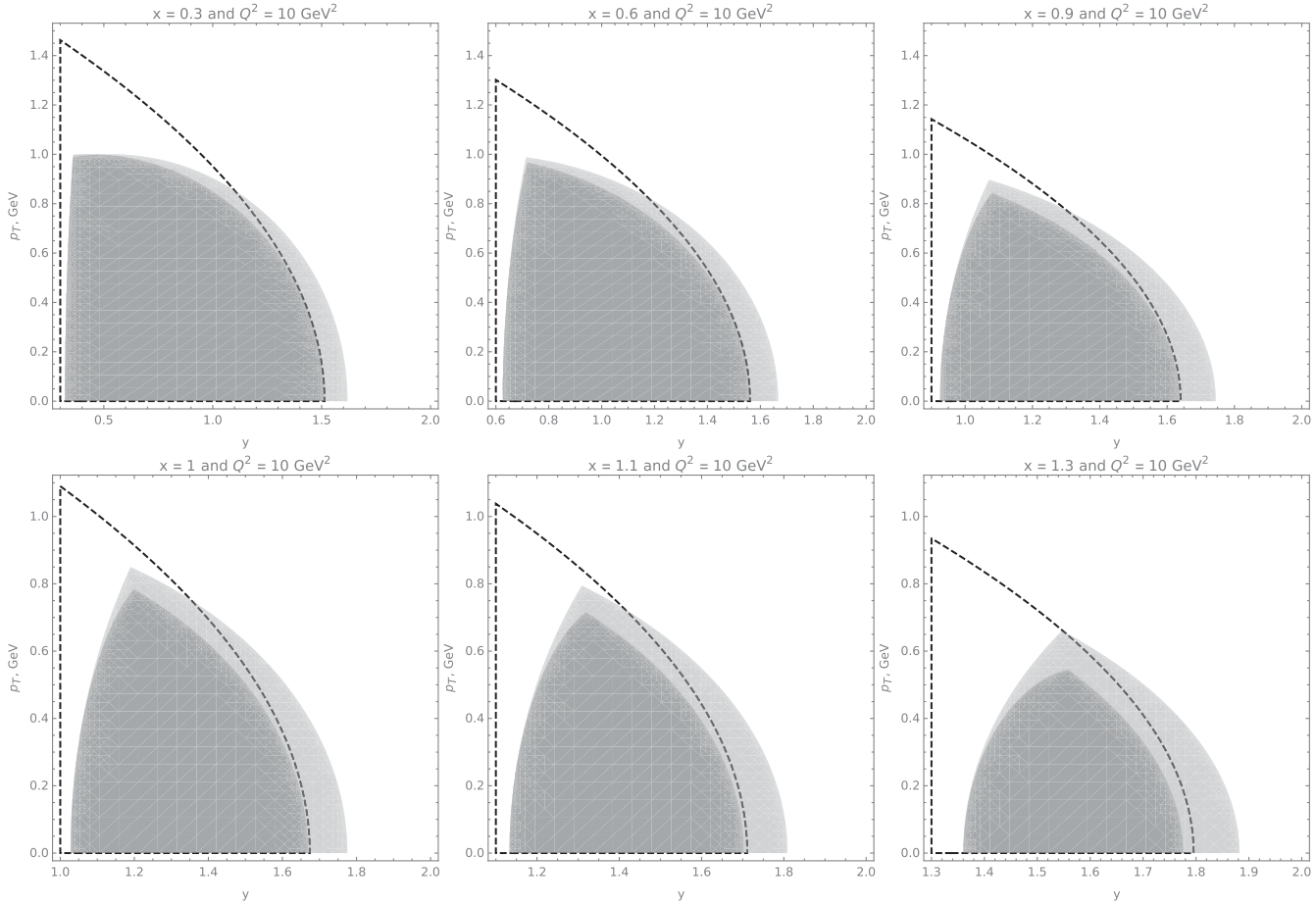

FIG. 13. The integration region in Eq. (B25) for different values of x and Q^2. The region $x < y < y_{\max}$ and $p_\perp^2 < t^2(y)$ is shown by dashed lines. The shaded area is a region restricted to $W^2 > (M + m_\pi)^2$ and $|\boldsymbol{p}| < 1$ GeV computed for both the relativistic and the nonrelativistic spectator, and for fixed values of x and Q^2 indicated in the panels. The light-gray region corresponds to the relativistic spectator, while the dark-gray region is for the nonrelativistic one.

In conclusion of this section, we present the explicit expressions for the light-cone smearing functions by Eqs. (B19) and (B20):

$$S_0(y) = \frac{M^2}{2}\int_0^t \mathrm{d}u(\psi_0^2(p) + \psi_2^2(p))\left(2 - \frac{u}{M}\right), \quad \text{(B27)}$$

$$S_1(y) = \frac{M^2}{2}\int_0^t \mathrm{d}u(\psi_0^2(p) + \psi_2^2(p))\left(2 - \frac{u}{M}\right)v, \quad \text{(B28)}$$

where $v = \mu^2/M^2 - 1$ and the other notations are similar to those in Eq. (B26). The function $S_0(y)$ and $S_1(y)$ computed for the AV18 deuteron wave function are plotted in Fig. 14 (left panel).

b. Relativistic spectator

For the *relativistic* kinematics, we have $p_0 = M_d - \sqrt{M^2 + \boldsymbol{p}^2}$. Integrating the δ function in Eq. (B15), we have

$$\int \mathrm{d}p_z \delta\left(y - \frac{p_0 + \gamma p_z}{M}\right) = \frac{\gamma M E}{aM + (\gamma^2 - 1)E}, \quad \text{(B29)}$$

where $a = M_d/M - y$, $E = \sqrt{p_z^2 + p_\perp^2 + M^2}$, and we should replace p_z with the solution of the following equation:

$$\gamma p_z = \sqrt{p_z^2 + p_\perp^2 + M^2} - aM. \quad \text{(B30)}$$

For $\gamma > 1$, this equation has a solution for any value of the parameter a:

$$p_z = \frac{-\gamma a M + \sqrt{a^2M^2 + (\gamma^2 - 1)(M^2 + p_\perp^2)}}{\gamma^2 - 1}. \quad \text{(B31)}$$

Note, that for $\gamma \to 1$ (or $Q^2 \to \infty$), only the region $a > 0$ is allowed, and we have

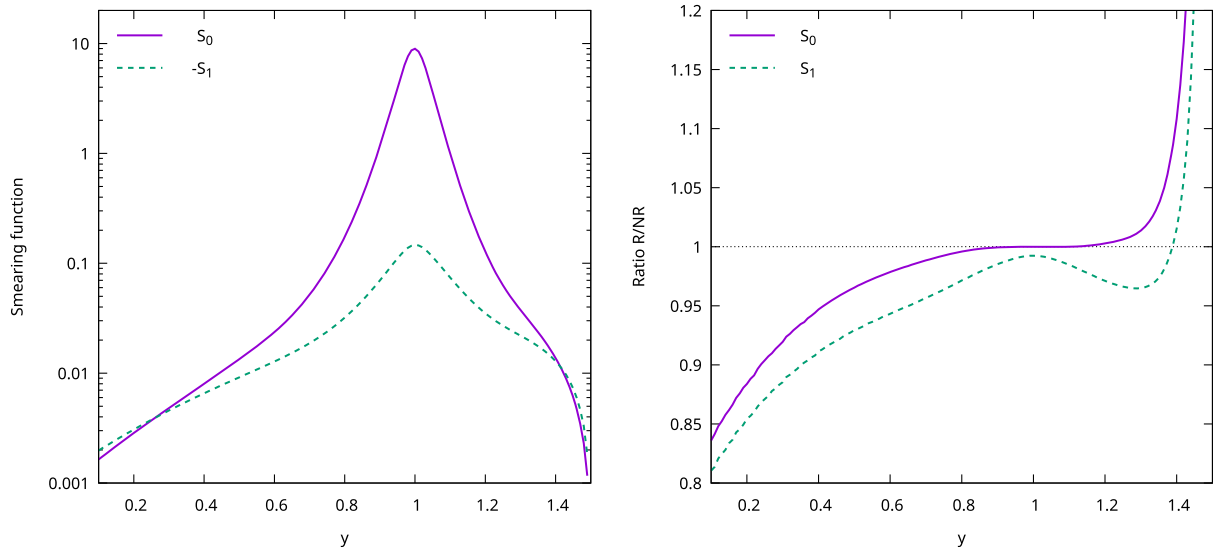

FIG. 14. Left panel shows the smearing functions $S_0(y)$ (solid line) and $-S_1(y)$ (dashed line) computed using Eqs. (B27) and (B28) for the AV18 deuteron wave function. The right panel illustrates the relativistic effects in the smearing functions (see also text).

$$p_z = \frac{p_\perp^2 + (1-a^2)M^2}{2aM}. \tag{B32}$$

In this case, the condition $a = 0$ determines the upper limit on y, $y_{\max} = M_d/M$. For finite values of Q^2, the integration region extends to $y > M_d/M$.

For the deuteron structure functions, we have

$$F_i^d(x,Q^2) = \frac{\gamma}{4}\int_{y_{\min}}^{y_{\max}} dy \int dp_\perp^2 \frac{ME}{aM+(\gamma^2-1)E} \times (\psi_0^2(p)+\psi_2^2(p)) K_{ij} F_j^N\left(\frac{x}{y},Q^2,\mu^2\right). \tag{B33}$$

The integration region in Eq. (B33) is limited by Eq. (B1). The resulting region is shown in Fig. 13, in which we also illustrate the impact of the momentum cut on the integration region.

The light-cone distributions by Eqs. (B19) and (B20) can be written as

$$S_0(y) = \frac{1}{4a}\int dp_\perp^2 (\psi_0^2(p)+\psi_2^2(p)) E\left(1+\frac{p_z}{M}\right), \tag{B34}$$

$$S_1(y) = \frac{1}{4a}\int dp_\perp^2 (\psi_0^2(p)+\psi_2^2(p)) E\left(1+\frac{p_z}{M}\right) \times \frac{M_d^2 - 2M_d E}{M^2}. \tag{B35}$$

Note that these functions have a pole at $a = 0$ that corresponds to $y = M_d/M$. However, this value of y requires an infinite nucleon momentum as $p_z \to \infty$ at $y = M_d/M$. Such configurations should be suppressed by the deuteron wave function. In practice, the region of large $y \sim M_d/M$, and therefore the singularity, can be avoided by applying a reasonable cut on the nucleon momentum in the convolution integral.

The effect of relativistic kinematics is illustrated in Fig. 14 (right panel), in which we show the ratio of the function S_0 computed with Eq. (B34) and (B27) and a similar ratio for S_1. For the most important region $|y-1| < 0.2$, which drives the nuclear convolution, this relativistic effect is negligible for S_0. For this reason, the relativistic effect has only a small impact on the deuteron structure function for $x < 1$. The relativistic correction is somewhat larger for S_1 but does not exceed 3% in this region. The region of large $|y-1|$ is driven by a high-momentum component of the deuteron wave function. For this reason, the effect of relativistic kinematics on the smearing functions is more important in this region, as illustrated by Fig. 14. As the region $y > 1$ drives the deuteron structure functions for $x \gtrsim 1$, one cannot ignore the effect of relativistic kinematics in this region of x.

3. Benchmarks of the convolution integral

In order to facilitate the comparison with the present approach, in Table II, we list our results for F_2^d computed for the test functions $F_2^N = (1-x)^3$ and $\delta f = x$.

TABLE II. The values of the deuteron structure function F_2^d computed by Eq. (6) using the AV18 deuteron wave function [45] and the test functions $F_2^N = (1-x)^3$ and $\delta f = x$.

	x						
Q^2	0	0.2	0.4	0.6	0.8	0.9	1.0
1.0	1.000	5.032E-01	2.086E-01	6.188E-02	9.986E-03	2.977E-03	8.851E-04
10.0	1.000	5.034E-01	2.095E-01	6.198E-02	8.711E-03	1.785E-03	2.359E-04
100.0	1.000	5.035E-01	2.096E-01	6.198E-02	8.550E-03	1.631E-03	1.649E-04

[1] A. Accardi, S. Alekhin, J. Blümlein, M. Garzelli, M. K. Lipka, W. Melnitchouk, S. Moch, J. Owens, R. Placakyte, E. Reya, N. Sato, A. Vogt, and O. Zenaiev, A critical appraisal and evaluation of modern PDFs, Eur. Phys. J. C **76**, 471 (2016).

[2] A. Accardi, T. J. Hobbs, X. Jing, and P. M. Nadolsky, Deuterium scattering experiments in CTEQ global QCD analyses: A comparative investigation, Eur. Phys. J. C **81**, 603 (2021).

[3] R. D. Ball, E. R. Nocera, and R. L. Pearson, Deuteron uncertainties in the determination of proton PDFs, Eur. Phys. J. C **81**, 37 (2021).

[4] M. Hirai, S. Kumano, and M. Miyama, Determination of nuclear parton distributions, Phys. Rev. D **64**, 034003 (2001).

[5] D. de Florian, R. Sassot, P. Zurita, and M. Stratmann, Global analysis of nuclear parton distributions, Phys. Rev. D **85**, 074028 (2012).

[6] K. Kovarik *et al.*, nCTEQ15—Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework, Phys. Rev. D **93**, 085037 (2016).

[7] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, EPPS21: A global QCD analysis of nuclear PDFs, Eur. Phys. J. C **82**, 413 (2022).

[8] M. Arneodo, Nuclear effects in structure functions, Phys. Rep. **240**, 301 (1994).

[9] D. F. Geesaman, K. Saito, and A. W. Thomas, The nuclear EMC effect, Annu. Rev. Nucl. Part. Sci. **45**, 337 (1995).

[10] P. R. Norton, The EMC effect, Rep. Prog. Phys. **66**, 1253 (2003).

[11] W. B. Atwood and G. B. West, The extraction of asymptotic nucleon cross-sections from deuterium data, Phys. Rev. D **7**, 773 (1973).

[12] S. V. Akulinichev, S. A. Kulagin, and G. M. Vagradov, The role of nuclear binding in deep inelastic lepton nucleon scattering, Phys. Lett. **158B**, 485 (1985).

[13] S. A. Kulagin, G. Piller, and W. Weise, Shadowing, binding and off-shell effects in nuclear deep inelastic scattering, Phys. Rev. C **50**, 1154 (1994).

[14] S. A. Kulagin and R. Petti, Global study of nuclear structure functions, Nucl. Phys. **A765**, 126 (2006).

[15] S. A. Kulagin and R. Petti, Neutrino inelastic scattering off nuclei, Phys. Rev. D **76**, 094023 (2007).

[16] S. A. Kulagin and R. Petti, Structure functions for light nuclei, Phys. Rev. C **82**, 054614 (2010).

[17] S. A. Kulagin and R. Petti, Nuclear parton distributions and the Drell-Yan process, Phys. Rev. C **90**, 045204 (2014).

[18] P. Ru, S. A. Kulagin, R. Petti, and B.-W. Zhang, Study of $W^{\pm}$ and Z boson production in proton-lead collisions at the LHC with Kulagin-Petti nuclear parton distributions, Phys. Rev. D **94**, 113013 (2016).

[19] S. I. Alekhin, S. A. Kulagin, and R. Petti, Nuclear effects in the deuteron and constraints on the d/u ratio, Phys. Rev. D **96**, 054005 (2017).

[20] A. Accardi, L. T. Brady, W. Melnitchouk, J. F. Owens, and N. Sato (CJ Collaboration), Constraints on large-x parton distributions from new weak boson production and deep-inelastic scattering data, Phys. Rev. D **93**, 114017 (2016).

[21] T.-J. Hou *et al.*, New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC, Phys. Rev. D **103**, 014013 (2021).

[22] T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne (MSHT Collaboration), An investigation of the α_S and heavy quark mass dependence in the MSHT20 global PDF analysis, Eur. Phys. J. C **81**, 744 (2021).

[23] R. D. Ball *et al.* (NNPDF Collaboration), The path to proton structure at 1% accuracy, Eur. Phys. J. C **82**, 428 (2022).

[24] D. Abrams *et al.* (MARATHON Collaboration), Measurement of the Nucleon F_2^n/F_2^p Structure Function Ratio by the Jefferson Lab MARATHON Tritium/Helium-3 Deep Inelastic Scattering Experiment, Phys. Rev. Lett. **128**, 132003 (2022).

[25] H. Georgi and H. D. Politzer, Freedom at moderate energies: Masses in color dynamics, Phys. Rev. D **14**, 1829 (1976).

[26] M. Virchaux and A. Milsztajn, A measurement of α_s and of higher twists from a QCD analysis of high statistics F_2 data on hydrogen and deuterium targets, Phys. Lett. B **274**, 221 (1992).

[27] S. I. Alekhin, S. A. Kulagin, and S. Liuti, Isospin dependence of power corrections in deep inelastic scattering, Phys. Rev. D **69**, 114009 (2004).

[28] S. A. Kulagin and W. Melnitchouk, Spin structure functions of ^{3}He at finite Q^2, Phys. Rev. C **78**, 065203 (2008).

[29] S. Tkachenko *et al.* (CLAS Collaboration), Measurement of the structure function of the nearly free neutron using spectator tagging in inelastic ^{2}H(e, e'p)X scattering with CLAS, Phys. Rev. C **89**, 045206 (2014); **90**, 059901(A) (2014).

[30] K. Griffioen *et al.*, Measurement of the EMC effect in the deuteron, Phys. Rev. C **92**, 015211 (2015).
[31] S. Malace *et al.* (Jefferson Lab E00-115 Collaboration), Applications of quark-hadron duality in F_2 structure function, Phys. Rev. C **80**, 035207 (2009).
[32] A. Airapetian *et al.* (HERMES Collaboration), Inclusive measurements of inelastic electron and positron scattering from unpolarized hydrogen and deuterium targets, J. High Energy Phys. 05 (2011) 126.
[33] M. Arneodo *et al.* (NMC Collaboration), Measurement of the proton and deuteron structure functions, F_2^p and F_2^d, and of the ratio σ_L/σ_T, Nucl. Phys. **B483**, 3 (1997).
[34] M. Arneodo *et al.* (NMC Collaboration), Accurate measurement of F_2^d/F_2^p and $R^d - R^p$, Nucl. Phys. **B487**, 3 (1997).
[35] L. Whitlow, Deep inelastic structure functions from electron scattering on hydrogen, deuterium, and iron at $0.6 \leq Q^2 \leq 30$ GeV^2, Ph.D. thesis, SLAC, 1990, https://www-public.slac.stanford.edu/sciDoc/docMeta.aspx?slacPubNumber=slac-r-357.
[36] A. Bodek *et al.*, Experimental studies of the neutron and proton electromagnetic structure functions, Phys. Rev. D **20**, 1471 (1979).
[37] G. G. Petratos (private communication).
[38] L. W. Whitlow, S. Rock, A. Bodek, E. M. Riordan, and S. Dasu, A precise extraction of $R = \sigma_L/\sigma_T$ from a global analysis of the SLAC deep inelastic ep and ed scattering cross-sections, Phys. Lett. B **250**, 193 (1990).
[39] M. D. Mestayer, W. B. Atwood, E. D. Bloom, R. L. A. Cottrell, H. DeStaebler, C. Y. Prescott, L. S. Rochester, S. Stein, R. E. Taylor, and D. Trines, The ratio σ_L/σ_T from deep inelastic electron scattering, Phys. Rev. D **27**, 285 (1983).
[40] J. Gomez *et al.* (SLAC E-139 Collaboration), Measurement of the A-dependence of deep inelastic electron scattering, Phys. Rev. D **49**, 4348 (1994).
[41] S. Dasu *et al.*, Measurement of kinematic and nuclear dependence of $R = \sigma_L/\sigma_T$ in deep inelastic electron scattering, Phys. Rev. D **49**, 5641 (1994).
[42] A. C. Benvenuti *et al.* (BCDMS Collaboration), A high statistics measurement of the deuteron structure functions $F_2(x, Q^2)$ and R from deep inelastic muon scattering at high Q^2, Phys. Lett. B **237**, 592 (1990).
[43] S. Alekhin, J. Blümlein, S. Moch, and R. Placakyte, Parton distribution functions, α_s, and heavy-quark masses for LHC Run II, Phys. Rev. D **96**, 014011 (2017).
[44] R. B. Wiringa, V. G. J. Stoks, and R. Schiavilla, An accurate nucleon-nucleon potential with charge independence breaking, Phys. Rev. C **51**, 38 (1995).
[45] S. Veerasamy and W. N. Polyzou, A momentum-space Argonne V18 interaction, Phys. Rev. C **84**, 034003 (2011).
[46] S. Alekhin, S. A. Kulagin, and R. Petti, Modeling lepton-nucleon inelastic scattering from high to low momentum transfer, in *Proceedings of the 5th International Workshop on Neutrino-Nucleus Interactions in the Few GeV Region (NUINT07): Batavia, USA, 2007* [*AIP Conf. Proc.* 967, 215 (2007)], 10.1063/1.2834481.
[47] https://www-zeuthen.desy.de/~alekhin/OPENQCDRAD/.
[48] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, LHAPDF6: Parton density access in the LHC precision era, Eur. Phys. J. C **75**, 132 (2015).
[49] https://lhapdf.hepforge.org/.
[50] I. Schienbein *et al.*, A review of target mass corrections, J. Phys. G **35**, 053101 (2008).
[51] C. Cocuzza, C. E. Keppel, H. Liu, W. Melnitchouk, A. Metz, N. Sato, and A. W. Thomas (JAM Collaboration), Isovector EMC Effect from Global QCD Analysis with MARATHON Data, Phys. Rev. Lett. **127**, 242001 (2021).
[52] A. Accardi (private communication).
[53] S. Alekhin, J. Blümlein, S. Moch, and R. Plačakytė, Isospin asymmetry of quark distributions and implications for single top-quark production at the LHC, Phys. Rev. D **94**, 114038 (2016).
[54] S. Alekhin, J. Blümlein, S. Kulagin, S.-O. Moch, and R. Petti, Strange and non-strange distributions from the collider data, Proc. Sci. DIS2018 (**2018**) 008 [arXiv:1808.06871].
[55] S. Alekhin, A. Kardos, S. Moch, and Z. Trócsányi, Precision studies for Drell–Yan processes at NNLO, Eur. Phys. J. C **81**, 573 (2021).
[56] J. Seely *et al.*, New Measurements of the EMC Effect in Very Light Nuclei, Phys. Rev. Lett. **103**, 202301 (2009).
[57] S. P. Malace *et al.* (E12-10-002 Collaboration), Precision measurements of the F_2 structure function at large x in the resonance region and beyond, https://misportal.jlab.org/mis/physics/experiments/viewProposal.cfm?paperId=622.
[58] R. Abdul Khalek *et al.*, Science requirements and detector concepts for the electron-ion collider: EIC yellow report, arXiv:2103.05419.
[59] R. Petti, Precision measurements of fundamental interactions with (anti)neutrinos, in *Proceedings of the 27th International Workshop on Deep Inelastic Scattering and Related Subjects (DIS 2019): Torino, Italy, 2019* (2019), arXiv:1910.05995.
[60] H. Duyang, B. Guo, S. R. Mishra, and R. Petti, A novel approach to neutrino-hydrogen measurements, arXiv:1809.08752.