Extraction of the neutron F_2 structure function from inclusive proton and deuteron deep-inelastic scattering data

S. Li[1,2] A. Accardi[3,4] M. Cerutti,[3,4] I. P. Fernando[5] C. E. Keppel,[4] W. Melnitchouk[4] P. Monaghan,[6]
G. Niculescu,[7] M. I. Niculescu[7] and J. F. Owens[8]

[1]*University of New Hampshire, Durham, New Hampshire 03824, USA*
[2]*Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
[3]*Hampton University, Hampton, Virginia 23668, USA*
[4]*Jefferson Lab, Newport News, Virginia 23606, USA*
[5]*University of Virginia, Charlottesville, Virginia 22904, USA*
[6]*Christopher Newport University, Newport News, Virginia 23606, USA*
[7]*James Madison University, Harrisonburg, Virginia 22801, USA*
[8]*Florida State University, Tallahassee, Florida 32306, USA*

The available world deep-inelastic scattering (DIS) data on proton and deuteron structure functions F_2^p, F_2^d, and their ratios are leveraged to extract the free neutron F_2^n structure function, the F_2^n/F_2^p ratio, and associated uncertainties using the latest nuclear effect calculations in the deuteron. Special attention is devoted to the normalization of the proton and deuteron experimental datasets and to the treatment of correlated systematic errors, as well as the quantification of procedural and theoretical uncertainties. The extracted F_2^n dataset is utilized to evaluate the Q^2 dependence of the Gottfried sum rule and the nonsinglet $F_2^p - F_2^n$ moments. To facilitate replication of our study, as well as for general applications, we provide a comprehensive DIS database including all recent Jefferson Lab 6 GeV measurements, the extracted F_n^2, a modified CTEQ-JLab global parton distribution function fit named CJ15nlo_mod, and grids with calculated proton, neutron, and deuteron DIS structure functions.

I. INTRODUCTION

Experimental data from deep-inelastic scattering (DIS) of leptons from nucleons have been pivotal in the study of nucleon structure for decades. In particular, DIS structure function measurements using charged lepton and neutrino beams have been used to extract parton distribution functions (PDFs) and investigate scaling and scaling violation, and they are generally the ubiquitous tools for studying quantum chromodynamics (QCD). An experimental difficulty in precision DIS nucleon structure studies is the use of nuclear targets, in particular, employment of deuterium as a surrogate neutron target. Here, Fermi motion, binding, and other nuclear effects render precision neutron measurements difficult, since these may change the shape of the free nucleon structure function. Yet, it is important to have experimental results for both protons and neutrons to separately determine u- and d-quark distributions at large parton momentum fractions x in order to fully understand the partonic structure of nucleons.

Deuteron nuclear corrections have been a challenge to experiments since the first DIS measurements at the Stanford Linear Accelerator Center (SLAC) in the 1970s [1–3]. Recently, a new testing ground has been provided for these corrections through the CTEQ-Jefferson Lab (CJ) global PDF analysis effort [4–7], where the neutron structure information extracted from DIS data on protons and deuterons using state-of-the-art nuclear corrections has been assessed simultaneously with other high-energy scattering datasets. The latter include, in particular, large rapidity W-boson asymmetries in proton-antiproton collisions at the Tevatron [8,9], which are sensitive to the d- to u-quark PDF ratio at large momentum fractions, and tagged DIS structure functions from the BONuS experiment [10] at Jefferson Lab (JLab), which select an effective neutron target. From agreement among the different datasets and nuclear versus nucleon targets in this more global context [11], confidence may be placed in the nuclear corrections procedure applied and constraints put on the nuclear correction model itself. This effort culminated in the

CJ15 simultaneous PDF and nuclear correction model extraction [7]. Global analyses similar in spirit to CJ15 have subsequently been performed by Alekhin, Kulagin, and Petti (AKP) [12,13] and the Jefferson Lab Angular Momentum (JAM) Collaboration [14], with some important differences that will be discussed further in the next section. Deuterium data have thus entered an era where the neutron structure function can be determined more accurately.

In fact, knowledge of both proton and neutron structure as measured by the corresponding structure functions F_2^p and F_2^n, respectively, and associated nucleon PDFs at large x, is central to a broad range of current scientific endeavors. Many present and planned experiments at JLab and other facilities seek to understand how the structure of bound objects such as nucleons can be described in terms of their fundamental partonic constituents. For example, the limit as $x \to 1$ is a testing ground for multiple perturbative and nonperturbative QCD predictions for the x dependence of PDFs [15–21]. In the unpolarized sector, this can be accessed by studying the ratio of d-quark to u-quark PDFs, which is directly constrained by the ratio F_2^n/F_2^p of neutron-to-proton structure functions.

From a different perspective, reliable knowledge of PDFs at large x is important for searches for new physics signals in collider experiments [22]. This is especially true if the search involves a region where the rapidity of the measured final state is large or the produced mass is large, and the reaction is sensitive to convolutions of two PDFs evaluated with one value of x being small and the other one large [23,24]. Similarly, in searches for novel high-mass resonances [25], both partons may be evaluated at large x. This issue has been discussed, for example, in the context of the forward-backward asymmetry for large mass dilepton states in processes with Drell-Yan-type kinematics [26]. There it was pointed out that such processes can be sensitive to PDFs evaluated at large values of x. For some PDFs, the relevant region of x can be beyond the region where they are constrained by data. This then involves extrapolations that can depend critically on the parametrization used. In such cases, disagreement between data and predictions may be a result of insufficient knowledge of the PDFs rather than new physics. Furthermore, uncertainties in PDFs at large x and low four-momentum transfer squared Q^2 percolate through QCD evolution to affect cross sections at smaller values of the Bjorken-x scaling variable (which at leading order in the strong coupling coincides with the parton momentum fraction x) and larger Q^2. In fact, nuclear corrections applied to DIS on deuteron targets, which are most prominent in the valence region at large x, also impact sea quarks and gluon PDFs over a wider range of x, and the achievement of precision in tests of the Standard Model in the electroweak sector will partly depend upon the successful treatment of nuclear corrections [27].

From yet another viewpoint, better control of nuclear corrections and a precise knowledge of the F_2^n/F_2^p ratio at large x for nonisoscalarity (the proton and neutron number imbalance) corrections will have a direct, measurable impact on the interpretation of current and future neutrino experiments [28], such as NOνA [29], DUNE [30], and MINERνA [31], as a large part of the theoretical uncertainty is from the lack of knowledge of the neutrino-nucleus interaction in the large-x valence regime.

On the theoretical front, lattice QCD simulations are providing increasingly precise predictions for moments of PDFs [32], as well as first results for the x dependence of PDFs [33], along with structure functions themselves [34,35], particularly in the valence quark regime. While calculations can nowadays be performed at physical quark masses, control over a range of systematic uncertainties, such as lattice discretization and volume, is still being improved. An important benchmark for these calculations is provided by the first moment of the isovector u-d-quark PDF, which can be readily compared to moments calculated with PDFs extracted in global QCD analyses. It is also interesting to more directly compare the lattice moments to experimental data, which can be accomplished by measuring moments of the isovector $F_2^p - F_2^n$ structure functions and using the operator product expansion to remove target mass corrections (TMCs). Phenomenological efforts in this direction have been restricted so far to specific experimental data choices [36,37], and the full power of the global DIS dataset has not yet been leveraged. Similar to the $p - n$ moments, neutron structure function data come into play in the experimental determination of the Gottfried sum rule (GSR) [38,39], through which constraints on the light antiquark distributions in the nucleon may be estimated as a complement to recent data from the SeaQuest experiment at Fermilab [40]. Here, again, previous extractions of the GSR have been limited to the analysis of specific datasets [37,41,42], and the statistical power of the global DIS dataset has not been fully exploited.

Historically, extractions of the neutron structure function have relied on inclusive lepton-proton and lepton-deuteron DIS data, using models for the nuclear wave function and an iterative extraction prescription [1,43]. Typically, these assume that the deuteron structure function can be written as a convolution of a smearing function (or light-cone momentum distribution of nucleons in the deuterium nucleus), constructed from a nonrelativistic deuteron wave function, and the free nucleon structure functions [44,45]. Relativistic and nucleon off-shell corrections, which generally cannot be expressed in convolution form [46,47], were included in subsequent analyses [15] of SLAC data on protons and deuterons, illustrating the impact of nuclear corrections in the deuteron on the extraction of the neutron structure function at large values of x. Other types of additive corrections to the convolution approximation, such as nuclear shadowing and meson exchange currents, as well

as limitations of smearing factor methods, were considered in Ref. [48]. A rather different approach [49] assumes scaling with nuclear density in which the nuclear effects in the deuteron are extrapolated from those in heavy nuclei such as ^{56}Fe [50,51]. Problems with application of the nuclear density model to light nuclei, including ambiguities in defining a nuclear density for deuterium, were discussed in Ref. [52].

In the context of global QCD analysis, nuclear corrections in the deuteron have been extensively studied by the CJ Collaboration [4–6], exploring different nuclear wave functions and models of nucleon off-shell corrections. While earlier CJ analyses relied on model calculations for the off-shell effects, more recent analyses [7] simultaneously fitted the PDF parameters, for given deuteron wave functions, together with the off-shell parameters. A similar approach was also adopted by Alekhin *et al.* [12,13]. Instead of using deuteron wave functions, Martin *et al.* [53] fitted the entire nuclear effect in the deuteron, including nuclear smearing, phenomenologically in terms of a suitable parametrization describing the deuteron corrections in terms four free parameters. Interestingly, the required deuteron corrections were of the expected general form with a large positive correction at very large x and a dip for $x \sim 0.5$. Most recently, Ball *et al.* [54] proposed an approach agnostic to the nuclear dynamics where the deuteron corrections were treated as an independent source of uncertainty in the determination of the proton PDFs.

In contrast to the PDF-level analyses, Arrington *et al.* [2,3] performed a study of the neutron structure function entirely at the structure function level, without any reference to partons or QCD. Surveying various models of nuclear corrections in the literature, including the dependence of the extraction procedure, they assessed the uncertainties from the deuteron wave function, off-shell effects, and different nuclear smearing models used to compute the nuclear corrections, correlating deuterium data with data on heavier nuclei.

The goal of the present work is to apply the CJ deuteron corrections [7] to the large global DIS dataset, including measurements on proton and deuteron targets and their ratios, and to provide an extensive neutron dataset. The extraction of the neutron structure function will be data driven, as much as possible, and the model dependence of the procedure confined to the deuteron to free nucleon ratio,

$$R_{d/N}(x, Q^2) = \frac{F_2^d(x, Q^2)}{F_2^N(x, Q^2)}, \tag{1}$$

where the nucleon structure function $F_2^N = \frac{1}{2}(F_2^p + F_2^n)$ and the per-nucleon structure function of the deuteron F_2^d will be calculated with the PDFs and nuclear correction model simultaneously fitted in the CJ15 next-to-leading order (NLO) global analysis [7]. Special attention will be devoted to the normalization of the experimental data and the treatment of correlated systematic errors, as well the quantification of procedural uncertainties.

The obtained neutron F_2^n and neutron-to-proton ratio $F_2^{n/p}$ datasets provide a data level representation of the CJ15 nuclear correction model, which may then be used for comparison to other current and future experimental analyses and as a convenient way to address deuteron nuclear corrections in other PDF extractions or phenomenological analyses. In this paper, the obtained neutron dataset will be utilized in a new data-driven extraction of the GSR, for which we will provide for the first time the Q^2 dependence, and an extraction of nonsinglet structure function moments for comparison with recent lattice QCD calculations. We will also discuss neutron excess corrections in DIS on heavy nuclear targets. Of course, other PDF sets, including ones fitted with even higher-order perturbative corrections, could, in principle, be used instead of the CJ15 PDFs; however, since our ultimate goal is to determine the neutron structure function and not PDFs, any set of PDFs can be used as an effective interpolation grid for the data, as long as it gives a reliable representation of the data and uses an internally consistent set of nuclear corrections.

The extraction of a neutron structure function dataset is particularly timely given that an increasing number of experiments at JLab and elsewhere are or will soon be providing additional data using a variety of experimental techniques to remove or minimize the need for theoretical deuteron nuclear corrections. The BONuS12 experiment [55] at JLab, for instance, uses a spectator tagging technique pioneered in the 6 GeV era to create an effective, essentially free, neutron target [10,56]. The recently published F_2^n/F_2^p from the MARATHON experiment [57], instead, used a theoretical superratio of the EMC effects in the ^{3}H and ^{3}He mirror nuclei from the AKP nuclear model [12,13] to extract F_2^n/F_2^p with reduced nuclear correction uncertainties. A more general analysis by the JAM Collaboration [14] fitted simultaneously both the nucleon PDFs and the nuclear (off-shell) effects in the $A = 3$ nuclei, which, in contrast to the AKP analysis, accounted for differences in the off-shell effects in the proton and neutron. Nevertheless, having determined the off-shell corrections from a global fit and the EMC effect in ^{3}H and ^{3}He, one can use these corrections to extract from the MARATHON data the free neutron structure function. Care should be taken, however, to ensure that correlations between nucleon off-shell and higher-twist (HT) effects and the relevant PDF parameters are properly taken into account [58].

In addition to these experiments that are directly sensitive to the neutron structure function, precise data on inclusive deuteron and proton F_2 structure functions are expected from the JLab E12-10-002 experiment [59], while the future SoLID program [60] will measure parity-violating electron scattering, which is sensitive to the d/u ratio using only a proton target. More proton and neutron

tagging and weak current DIS structure function measurements are discussed at the Electron-Ion Collider [61]. Moreover, W asymmetry measurements from RHIC at Brookhaven National Laboratory [62] and from the LHCb experiment at CERN [63–67] are also sensitive to the d quark and will thus further help constrain the neutron structure function when included in a global QCD analysis.

The rest of this paper is organized as follows. In Sec. II we discuss how deuteron corrections are studied and constrained in the context of global QCD analyses, with particular attention to the CJ15 analysis that will be leveraged in the extraction of the neutron F_2 dataset. We also analyze differences with the recent AKP and JAM analyses and briefly discuss their possible sources. In Sec. III we present our data-driven neutron extraction strategy and results, with a careful emphasis on experimental and theoretical uncertainties. The data selection criteria and cross-normalization procedure, which is of paramount importance to obtain precise neutron extraction, is also discussed in detail. In Sec. IV we provide several applications of the neutron data obtained in our analysis, including a new extraction of the Gottfried sum with the first-ever evaluation of its Q^2 dependence and a new extraction of nonsinglet moments for comparison to lattice QCD calculations. We also provide a new extraction of the F_2^n/F_2^p ratio for use, for example, in neutron excess corrections. In Sec. V we summarize our results.

A significant effort has also been made to construct a public database of world DIS data on protons and deuterons, whose usefulness goes beyond the application presented in this paper, which is supplemented by the extracted neutron structure function data and by the bin-centered data used in our phenomenological studies (Appendix A). We also make public the modified set of CJ15 PDF named CJ15nlo_mod used in our analysis, along with structure function grids in Les Houches accord parton density function (LHAPDF) format for proton, neutron, and deuteron targets calculated at NLO (Appendix B). We provide instructions for using the DIS structure function grids to correct for nuclear (isoscalar) effects in nuclear cross sections in Appendix C.

II. DEUTERON CORRECTIONS IN GLOBAL QCD ANALYSIS

Parton flavor separation can be most robustly performed through global QCD analysis [68–71], which utilizes large collections of data from high-energy collisions sensitive to different underlying PDF combinations. Conversely, the extracted PDFs can be utilized in a perturbative QCD calculation of the unmeasured free neutron F_2^n structure function, or, as we will exploit in this article, of the deuteron to isoscalar nucleon structure function ratio $R_{d/N}$. This program necessitates, however, knowledge of the effects on PDFs of the nucleon binding forces in the

deuteron in order to properly extract parton-level information from experiments on deuteron targets [11].

Interest in deuteron corrections has recently increased in the global QCD analysis community because these not only impact the d/u ratio at large momentum fractions, but also have important secondary effects on, e.g., the gluon or sea-quark PDFs at smaller x values [27]. In a global fit, deuteron corrections can be applied at different levels of sophistication. For example, one can account for the additional uncertainty associated with the deuteron effects by iteratively fitting deuteron PDFs in addition to the proton PDFs [54], parametrize and fit the nuclear deformation of the deuteron structure function [53,72,73], or utilize a dynamical model of nuclear interactions to calculate deuteron observables as a double convolution of (off-shell) parton distributions and nucleon wave functions [7,12,14]. In the CJ approach adopted here, we focus on the latter class of corrections.

The CJ Collaboration [74] has performed a series of global QCD analyses of unpolarized PDFs [4–7] with the aim of maximally utilizing DIS data at the highest x values amenable to perturbative QCD analysis. Special attention has been devoted to deuteron target dynamics, relevant at all energy scales, and to power corrections, such as HTs and TMCs, that become relevant for fixed target experiments probing low values of Q^2 and invariant final state mass squared, W^2. To separate the u- and d-quark PDFs, the CJ analysis fits DIS data from both hydrogen and deuterium targets. The theoretical description of the latter also requires a careful treatment of nuclear interactions, which modify the bound nucleon structure, particularly at large values of x at all Q^2 scales. Since the u-quark PDF is well constrained by a variety of ep, pp, and $p\bar{p}$ scattering data included in the global fit, the d-quark PDF extraction is, in practice, rather sensitive to the neutron structure function, and corrections for nuclear effects in deuterium become a major factor in its accurate determination above $x \approx 0.5$. This extraction has, therefore, historically suffered from large uncertainties due to the model dependence of these nuclear effects [18,19].

Beyond deuterium DIS data, the CJ15 PDF analysis [7] studied the impact on the determination of NLO PDFs and their uncertainties from large rapidity charged-lepton [75–77] and W-boson asymmetry data [8,9] from proton-antiproton collision at Fermilab's Tevatron collider. It also included, for the first time in a global QCD analysis, the novel JLab BONuS data on the free neutron structure function obtained from backward-spectator proton-tagged DIS on a deuterium target [10]—a technique utilized to effectively create a free neutron target from deuterium and significantly reduce the nuclear uncertainties that have afflicted previous neutron extractions. These two datasets provide the global fit with critical sensitivity to large-x d quarks and neutron structure. However, each data type presents positives and negatives within the global QCD

fitting context: the W-boson asymmetry data exist in the perturbative regime and are directly sensitive to the d quark without the kinematical smearing that affects the decay lepton measurements; however, the dataset is small and statistical reconstruction of the W-boson kinematics is needed to maximize the large-x reach of the measurements. The tagged DIS data, on the other hand, are more numerous and provide clean neutron data relatively free of deuteron nuclear corrections, but exist in a nonperturbative regime where HT and other effects are of concern. The deuteron DIS data, conversely, are rich in kinematic breadth, formed from numerous accurate measurements spanning decades in x and Q^2 from multiple laboratories and experiments. These data alone would be sufficient to access the neutron structure and d-quark PDF—as can successfully be done to high precision for the u quark with proton data—were it not for the theoretical nuclear uncertainty associated with extracting neutron structure from deuterium.

These three different datasets can then be simultaneously leveraged in the context of global QCD analysis of PDFs to extract the d-quark distribution and, at the same time, constrain the deuteron corrections. In the CJ15 analysis [7], for example, the nuclear corrections in the deuteron are performed using a smearing formalism in the weak binding approximation [78,79] that accounts for nuclear Fermi motion and binding through deuteron wave functions calculated in many-body theory and accounts for bound nucleon deformations of the nucleon PDFs by a Taylor expansion in the nucleon off-shellness, with the expansion coefficient fitted to experimental data. The free and bound nucleon structure functions are also corrected for TMC and HT effects. The interplay of d-quark sensitive observables on free protons (such as W-boson production in $p\bar{p}$ collisions) and of DIS data on deuteron targets allows one to disentangle nuclear dynamics and nucleon PDFs in the latter and to simultaneously constrain both the PDF and off-shell parameters, especially in the large-x region where these are poorly known. Tagged deuteron DIS measurements from BONuS then serve to verify the correctness of the deuteron model and, in the future with more statistics expected from the BONuS12 experiment, can put additional constraints on it.

The agreement of data from these three disparate reactions, simultaneously contributing to, and well described within, the fundamental QCD framework of the CJ15 global PDF fit thus provides one not only with an improved knowledge of the d-quark PDF, but also with confidence in the ability to access the neutron using the deuteron nuclear corrections deployed in that global analysis. If the nuclear corrections were inadequate beyond expected uncertainties, the three data types would exhibit significant tension within the global fitting framework. The fact that they do not gives confidence in the applied nuclear corrections and facilitates the neutron dataset extraction provided in this present work, with the global analysis
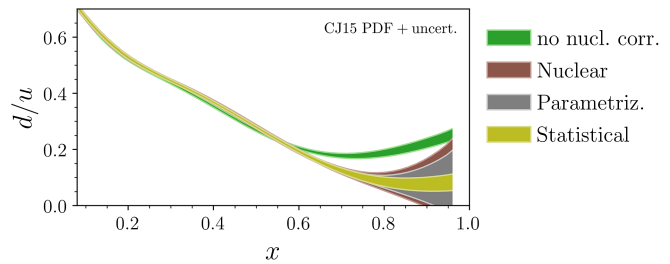

FIG. 1.　d/u PDF ratio from the CJ15 global QCD analysis [7]. The lower three bands show the size of the PDF and off-shell parameter variations compatible with the experimental uncertainties of the fitted experimental data (mustard colored band), as well as the impact of theoretical uncertainties from an extended d-quark parametrization (gray band) and the choice of different deuteron wave functions in the calculation of the deuteron structure function (maroon band). The upper green band, with only statistical uncertainties displayed, shows the result of a fit where nuclear corrections are neglected entirely; this fit is strongly disfavored over the nuclear corrected CJ15 fit, with a χ^2 increase of 262 units, largely driven by tensions between deuteron DIS and DØ W asymmetry data.

framework enabling the necessary evaluation of associated procedural uncertainties. As an example, the d/u PDF ratio obtained in the CJ15 analysis is displayed in Fig. 1, showing the importance of including nuclear corrections when fitting deuteron target DIS data. The "statistical" PDF uncertainties propagated from the experimental uncertainties are supplemented by an estimate of the theoretical uncertainties due to PDF parametrization choice and the deuteron many-body wave function.

In contrast to the CJ15 analysis, the global fits performed by Alekhin *et al.* in 2017 (AKP17) [12] and revisited in 2021 (AKP21) [13] resulted in marked differences in the fitted off-shell function and the d/u PDF ratio, which is softer than the CJ15 result at $x \gtrsim 0.7$ and (by construction) tends to 0 as $x \to 1$. At the same time, the AKP17/21 neutron structure function, which is indirectly extracted from the measured proton and deuteron data, also differs from the CJ15 result, especially at large x where it is actually harder than in the CJ15 case. This can be appreciated from Fig. 2, which compares the calculated $R_{d/N}$ ratio to the experimental values obtained by the BONuS Collaboration combining their tagged, quasi-free neutron-to-deuteron ratio F_2^n/F_2^d measurements with the world data on proton F_2^p and deuteron F_2^d structure functions. The recent JAM21 global analysis [14], on the other hand, which considered a dataset similar to those used in the CJ15 and AKP17/21 analyses, along with the ^{3}He/^{3}H cross section ratio from the MARATHON experiment [57], finds the d/u, F_2^n/F_2^p, and $R_{d/N}$ ratios that are similar to the present results. The extracted ratios and the isospin-averaged off-shell functions were also consistent with the results of the CJ15 analysis [7], but were less compatible with the AKP off-shell corrections.
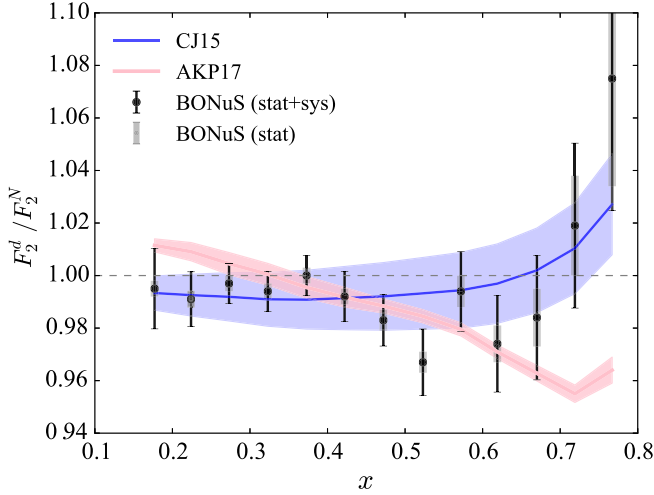
FIG. 2. Deuteron-to-nucleon F_2 structure function ratio from the CJ15 [7] and AKP17 [12] global QCD analyses compared to the BONuS experimental extraction [80]. The bands display the fit uncertainty at the 90% CL (the published 68% CL AKP band was rescaled by a factor 1.646). Data at $x \gtrsim 0.5$ have an average invariant mass squared $2 < \langle W^2 \rangle < 3$ GeV^2, and the theoretical calculations here should be understood in the spirit of quark-hadron duality [81].

The discrepancy between the deuteron-to-nucleon ratio $R_{d/N}$ calculated in independent global QCD analyses can have multiple sources; among these are the choice of fitted datasets and PDF parametrization, as well as the phenomenological implementation of HT effects, TMCs, and nuclear corrections [13,58,82]. Tracking down and disentangling from each other their effects is a complex task since they combine nonlinearly. In fact, HT corrections and their interplay with off-shell effects might be of particular relevance to the extraction of the neutron structure function [82]. That this may be the case can be appreciated by noticing that the AKP d-quark PDF is softer than the corresponding CJ15 distribution at large values of x, so that larger HTs are needed to bring the AKP $R_{d/N}$ ratio below the CJ15 ratio in Fig. 2.

An alternative approach to deuteron corrections that is agnostic to the nuclear dynamics was recently proposed by Ball *et al.* [54], where these are treated as an independent source of systematic error on the determination of the proton PDFs. In their approach, deuteron PDFs are separately fitted to deuteron target data, and the uncertainties on the deuteron PDFs are utilized to deweight deuteron data when utilizing these alongside the rest of the global dataset in a proton PDF fit. In contrast, in the approach followed by the CJ, AKP, and JAM Collaborations, the deuteron wave function is utilized to calculate Fermi motion and binding effects on the nucleons in the deuteron, and off-shell PDF deformations are constrained by the global dataset without the need of a separate fit of deuteron PDFs. The $R_{d/N}$ ratio found in Fig. 7 of Ref. [54] is

nonetheless very similar to the CJ15 ratio displayed in Fig. 2, as well as to the $R_{d/N}$ ratio fitted alongside the proton PDFs to the global data in the Martin–Motylinski–Harland-Lang–Thorne (MMHT2014) analysis [72].

A comprehensive investigation of the role of various theoretical corrections is outside of the scope of this article, but remains an important objective for future studies. These will be facilitated by the development of new statistical analysis methods, such as the recently developed L_2 sensitivity method [83], which was utilized in Ref. [27] to distinguish the data-driven features of a global fit from methodological differences in the CJ15 and CT18 PDF global analyses.

III. NEUTRON STRUCTURE FUNCTION EXTRACTION

The central aim of this paper is to perform a data-driven extraction of a cross-normalized set of neutron F_2^n structure functions from a set of experiments that measured F_2 on both proton and deuteron targets. A secondary aim is to extract neutron-to-proton F_2^n / F_2^p ratios from deuteron-to-proton ratio data, as well as from spectator tagging experiments, such as BONuS, that provide neutron-to-deuteron ratio measurements.

This program can be accomplished with minimal sensitivity to the theoretical inputs by systematically applying to the data the theoretically calculated correction ratio $R_{d/N}^{\text{th}}(x, Q^2)$, as in Eq. (1), across a range of x and Q^2, effectively converting the measured deuteron F_2^d structure function (measured alone or as part of a d/p or tagged n/d structure function ratio) into a superposition of free proton and free neutron structure functions. The neutron F_2^n structure function can then be extracted by suitably subtracting the measured free proton contribution from the observable under consideration. For consistency, the same theoretical calculations utilized for $R_{d/N}^{\text{th}}$ will also be used to cross-normalize data from different experiments and from different targets within one experiment in order to control relative fluctuations in their systematic shifts.

In the present analysis, we perform the needed theoretical calculations utilizing the PDFs and deuteron correction model simultaneously fitted in the CJ15 global QCD analysis [7]. The choice of CJ15 as a model to calculate the F_2 structure functions of free nucleons and to evaluate the nuclear effects in the deuteron affects the neutron data extraction in two places: first, the cross-normalization of the experimental proton and deuteron data and, second, the calculation of the nuclear correction ratio $R_{d/N}^{\text{th}}$. As mentioned earlier, a global analysis similar in spirit to CJ15 was performed by AKP in Refs. [12,13], but resulted in marked differences in the d/u PDF ratio and, more importantly for this analysis, in the obtained $R_{d/N}^{\text{th}}$ ratio. These differences would result in a systematic shift of the extracted neutron data if the AKP model were to be adopted instead of the
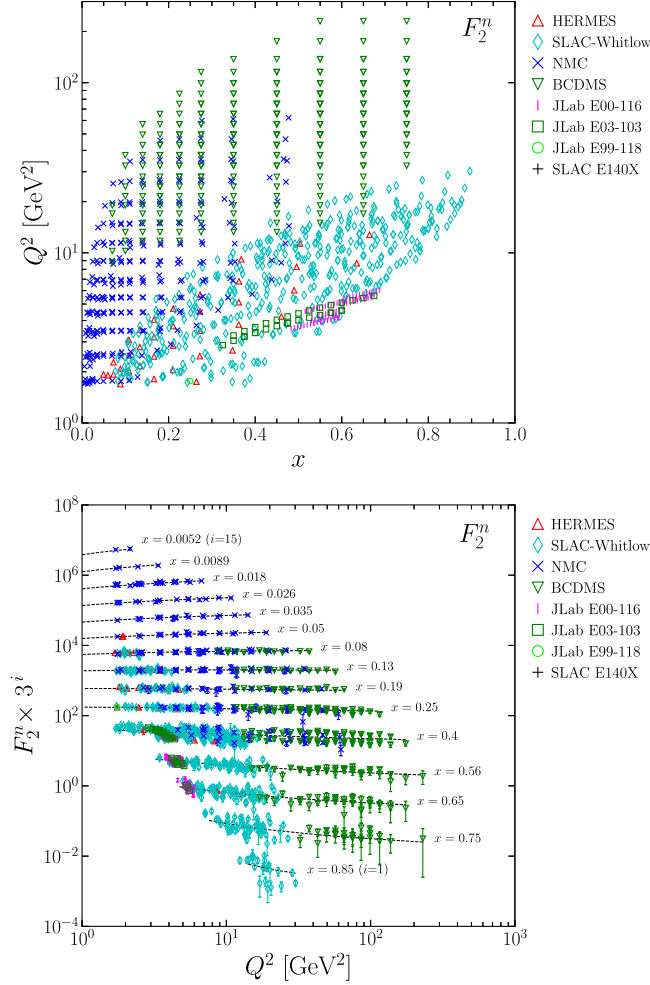
FIG. 3. "Matched" kinematic range (upper) and F_2 value binned in x (lower) of the neutron structure functions extracted from world data (see also Table I). The F_2 values are cut at $W^2 = 3.5$ GeV^2.

CJ15 model. However, a full evaluation of this systematic effect would require use of the AKP fitting framework and goes beyond the scope of this paper.

A. Proton and deuteron data selection and cross-normalization

In order to provide a complete neutron dataset over the widest possible kinematic range, we extended the DIS database utilized in the CJ15 global QCD analysis [7] to include all the relevant inclusive measurements from the JLab 6 GeV experimental program. We also revisited correlated errors in the existing CJ15 F_2 data collection [such as in the NMC [84] and SLAC [85] datasets] and included data from the SLAC-140x [86] experiment. This extended and up-to-date DIS database is discussed further in Appendix A.

Our extraction of the neutron structure function F_2^n (see Sec. III B below) requires us to simultaneously

manipulate proton and deuteron structure functions measured at the same values of x and Q^2 in order to minimize the size of systematic uncertainties. To this purpose, we select pairs of proton and deuteron data according to the following criteria:

(1) proton and deuteron F_2 data were measured within the same experiment and at the same beam energy;
(2) the corresponding Bjorken-x values agree to within an interval $\Delta x = 0.01$; and
(3) the Q^2 values agree to within 1%.

To be consistent with the CJ15 analysis [7], we also require that all selected data satisfy the cut $Q^2 > 1.69$ GeV^2, which marks the starting scale for QCD evolution of the PDFs, and $W^2 > 3.5$ GeV^2 to select DIS events. Note that we increased the W^2 cut from 3 GeV^2 (used in the CJ15 analysis) to 3.5 GeV^2 to better reject the resonance contamination in the large-x and low-Q^2 data. Overall, we obtain 1192 matched proton and deuteron F_2 data points covering the range $x = 0.005$–0.896 and $Q^2 = 1.69$–230 GeV^2, as shown in Fig. 3 and summarized in Table I. Conversely, no matching of d/p and tagged n/d data is required to extract the neutron-to-proton structure function ratio,

$$R_{n/p}(x, Q^2) \equiv \frac{F_2^n(x, Q^2)}{F_2^p(x, Q^2)}. \qquad (2)$$

The available ratio data, listed in Table II, have a smaller multiplicity, but also smaller statistical and systematic

TABLE I. Matched proton and deuteron structure function datasets used in the current F_2^n extraction. The number N_{match} of matched data points is listed, along with the experimental normalization uncertainties δn. The nuisance parameter λ^{norm} is needed to normalize the data to the CJ15-based theoretical calculations, and the size of its procedural uncertainty $\delta^{\text{CJ}}\lambda$ is discussed in Sec. III C. Datasets marked with "∗" have previously been included in the CJ15 analysis (except the NMC d/p ratio was used instead of the d structure function in the PDF fit). For consistency with the CJ15 analysis [7], eight data points from HERMES have been excluded in this work. The E99-118 data come with no normalization uncertainties, and thus no λ^{norm} are reported, although we note that a normalization shift of 1.55(37)% and 2.41(50)% would bring the p and d datasets, respectively, to good agreements with CJ15.

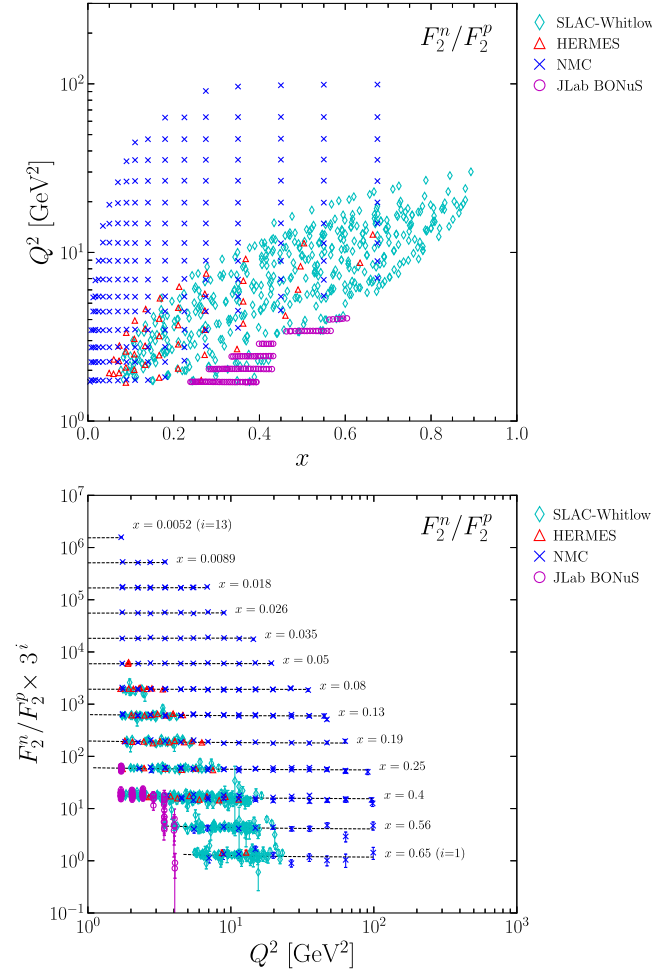| $F_2^{p,d}$ experiment	N_{match}	$\delta n_{p,d}$ (%)	$\lambda_{p,d}^{\text{norm}}$
*SLAC Whitlow [85]	478	2.1, 1.7	0.38(14), 0.65(18)
SLAC E140x [86]	9	1.73, 1.73	−0.12(6), −0.32(9)
*NMC [87]	275	2.5, 2.5	0.40(16), 0.04(12)
*BCDMS [88,89]	254	3.0, 3.0	−0.90(7), −0.23(7)
*HERMES [90]	37	7.5, 7.5	−0.03(3), −0.27(4)
JLab E99-118 [91]	2		
JLab E03-103 [92,93]	32	2.5, 2.5	−0.6(4), −0.88(16)
*JLab E00-116 [94]	97	1.75, 1.75	−0.59(17), −1.41(25)

FIG. 4. Kinematic range (upper) and F_2 value binned in x (lower) of the n/p ratio extracted from world data (see also Table II). The F_2 ratios are cut at $W^2 = 3.5$ GeV^2.

uncertainties. Their kinematic coverage is illustrated in Fig. 4.

Each experimental dataset typically provides three groups of uncertainties:

(1) A multiplicative overall relative normalization uncertainty δn. This allows all points in a given dataset to be scaled by a common normalization factor,

$$n = 1 + \lambda^{\mathrm{norm}} \delta n, \quad (3)$$

where λ^{norm} is a normally distributed stochastic variable with zero mean and unit standard deviation.

(2) Point-to-point uncorrelated statistical $\delta_{\mathrm{stat}} D_i$ and systematic $\delta_{\mathrm{syst}} D_i$ uncertainties, for the set's ith data point D_i. These are summed in quadrature to obtain the total uncorrelated uncertainty δD_i.

(3) A number K of point-to-point correlated additive systematic uncertainties $\vec{\beta}_k = (\beta_{k,1}, \beta_{k,2}, \ldots, \beta_{k,N_{\mathrm{data}}})$, with $k = 1, \ldots, K$ labeling the uncertainty sources, and N_{data} as the number of data points in that

TABLE II. Structure function ratio data used in the neutron-to-proton structure function ratio extraction. As for Table I, δn is the experimental normalization uncertainty, and λ^{norm} is the nuisance parameter needed for cross-normalization through the CJ15 global QCD analysis. Datasets marked with "∗" were included in the CJ15 global PDF fit.

Ratio experiment	N_{data}	δn (%)	λ^{norm}
*BONuS F_2^n/F_2^d [10]	115	4.5	−0.36(9)
*NMC F_2^d/F_2^p [95]	189	0.3	−1(1)
HERMES F_2^d/F_2^p [90]	45	1.4	−0.57(21)
SLAC-Whitlow F_2^d/F_2^p [85]	487	1.0	0.2(3)

dataset. The ith data point can then be shifted by an amount

$$\Delta_i = \sum_k \lambda_k \beta_{k,i}, \quad (4)$$

where λ_k are normally distributed stochastic variables with zero mean and unit standard deviation. The "nuisance parameters" λ^{norm} and λ_k from each experiment allow data points from proton and deuteron datasets to independently fluctuate within their reported normalization uncertainty and also, when available, within their correlated uncertainties [70]. Note that the NMC experiment also provides information on the cross-correlations between their measurements on proton and on deuteron targets; however, in this analysis we have conservatively treated these two targets as fully uncorrelated, as for all other datasets.

Since we aim at extracting the neutron structure function by subtraction of the proton component from its matched deuteron measurement, it is vital to determine an optimal set of nuisance parameters for all experiments or, in other words, to cross-normalize the datasets. In particular, cross-normalization of the proton and deuteron data is needed to avoid large fluctuations in the neutron extraction due to their relative systematic shifts. To accomplish this, we choose to fix λ^{norm} and λ_k of each experiment by comparing the experimental data D_i to a corresponding theoretical value T_i calculated in perturbative QCD at the measured kinematics of each data point. The following χ^2 function is then minimized with respect to λ^{norm} and the λ_k parameters of each experiment (expt.),

$$\chi^2 = \sum_{\mathrm{expt}} \left[\sum_{i=1}^{N_{\mathrm{data}}} \left(\frac{D_i + \Delta_i - T_i/n}{\delta D_i} \right)^2 + (\lambda^{\mathrm{norm}})^2 + \sum_{k=1}^{K} \lambda_k^2 \right]_{\mathrm{expt}}. \quad (5)$$

Note that measurements on different targets in a given experiment are here labeled and treated as separate experiments. (Only the NMC reported correlations between different target measurements, but we conservatively consider these as statistically independent.)

The theoretical value T_i for each kinematic point is calculated in perturbative QCD at NLO using the PDFs and the deuteron correction model determined in the CJ15 global analysis. While these are kept fixed, the nuisance parameters λ^{norm} and λ_k of all experiments are simultaneously fitted. (Note that, in fact, the CJ15 PDFs were fitted to a dataset that also included measurements of jet production and Drell-Yan lepton pair production in hadron-hadron collisions, besides the DIS measurements highlighted in Tables I and II, thus imposing tighter constraints on the nuisance parameters than allowed by DIS data alone.) The obtained normalization parameters λ^{norm} are listed in Tables I and II, alongside their fit uncertainty, to be discussed in Sec. III C. Finally, we define the central value of the cross-normalized F_2 data, denoted by the superscript (0), as

$$\hat{D}_i^{(0)} = n^{(0)}\left(D_i + \Delta_i^{(0)}\right), \qquad (6)$$

where $n^{(0)} = 1 + \lambda^{\text{norm}(0)}\delta n$ and $\Delta_i^{(0)} = \sum_k \lambda_k^{(0)}\beta_{k,i}$ are the fitted normalization factors and correlated data shifts.

B. Neutron F_2^n and neutron-to-proton ratio extraction

The cross-normalized $\hat{F}_2^p$ and $\hat{F}_2^d$ data can now be used to extract the neutron structure function $\hat{F}_2^n$. As noted above, in order to do this, we use the deuteron correction factor $R_{d/N}^{\text{CJ}}(x, Q^2)$ as in Eq. (1) to remove the nuclear effects from the deuteron data, with the proton and deuteron structure functions computed at NLO in perturbative QCD using the PDFs and nuclear correction model from the CJ15 global QCD analysis [7]. The central value of the neutron structure function can then be obtained as

$$\hat{F}_2^{n(0)}(x, Q^2) = \frac{2\hat{F}_2^{d(0)}(x, Q^2)_{\text{expt}}}{R_{d/N}^{\text{CJ}}(x, Q^2)} - \hat{F}_2^{p(0)}(x, Q^2)_{\text{expt}}. \qquad (7)$$

This formula first converts the experimentally measured deuteron data into the sum of free proton and free neutron structure functions, then subtracts from this the experimentally measured proton contribution to obtain the neutron structure function. An alternative approach defining $F_2^n = F_{2,\text{expt}}^d[F_2^n/F_2^d]_{\text{CJ}}$ was discarded because it trades a smaller amount of experimental information for a larger amount of theoretical input with its associated uncertainties. Conversely, Eq. (7) minimizes the use of the theoretical model in accordance with our goal of performing a data-oriented neutron extraction. The result is displayed in Fig. 3 as a function of Q^2 in selected x bins and in Fig. 5 as a function of x in selected Q^2 bins. The uncertainties will be discussed in Sec. III C and bin centering will be addressed in Sec. III D. The x and Q^2 kinematic coverage of the
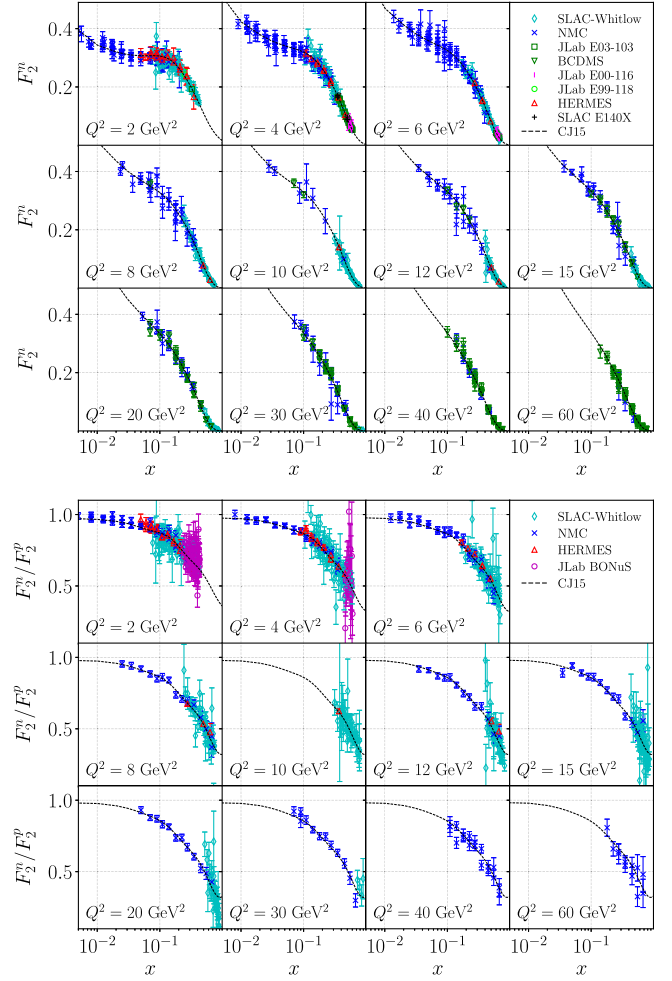

FIG. 5. Extracted F_2 neutron structure function (upper) and n/p ratio (lower) as a function of x in selected Q^2 bins, compared with the CJ15 fit (dashed black line). The data are bin centered in Q^2 and x for clarity, and the experimental (δ^u) and procedural (δ^{CJ}) uncertainties are added in quadrature.

extracted data corresponds to that of the matched p and d data, as pictured in Fig. 3.

This procedure cannot be directly benchmarked against experimental data for the neutron structure function. However, as a consistency check, we can combine the extracted F_2^n with the measured deuteron structure function and compare their ratio to the $R_{n/d} = F_2^n/F_2^d$ ratio experimentally measured by the 6 GeV BONuS experiment at JLab via spectator proton tagging [10]. As shown in Fig. 6, the neutron-to-deuteron ratio extracted as described above agrees with BONuS data and, conversely, validates the 6 GeV BONuS analysis.

Analogous to the neutron F_2^n extraction, the central values of the neutron-to-proton ratio $R_{n/p}$ can be obtained from the experimental d/p or n/d ratio data, utilizing the deuteron correction factor from Eq. (1) in analogy with the neutron extraction of Eq. (7),
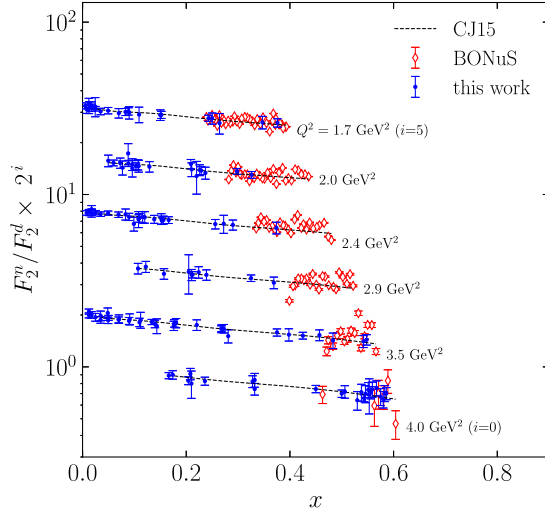
FIG. 6. Neutron-to-deuteron ratios F_2^n/F_2^d from this extraction (blue circles) compared with the JLab BONuS data (red diamonds) and the ratio calculated from CJ15 (dashed black lines). The F_2^n/F_2^d values extracted in this paper are selected within a 0.1 GeV^2 slice of each quoted Q^2 value.

$$\hat{R}_{n/p}^{(0)} \equiv \frac{2\hat{R}_{d/p}^{\exp,(0)}}{R_{d/N}^{CJ}} - 1, \tag{8}$$

and

$$\hat{R}_{n/p}^{(0)} \equiv \frac{\hat{R}_{n/d}^{\exp,(0)} R_{d/N}^{CJ}}{2 - \hat{R}_{n/d}^{\exp,(0)} R_{d/N}^{CJ}}. \tag{9}$$

The bin-centered version of the extracted n/p ratio is displayed as a function of x in Fig. 4 and as a function of Q^2 in Fig. 5.

C. Uncertainties

The extracted $\hat{F}_2^n$ structure function (and the $\hat{R}_{n/p}$ neutron-to-proton ratio) come with uncorrelated experimental uncertainties δD_i propagated from the experimental proton and deuteron data (or n/d and d/p ratio data) and with a procedural uncertainty δ^{CJ} due to data cross-normalization and the treatment of nuclear corrections. More specifically, the procedural uncertainty is due to the determination of the normalization factor λ^{norm}, the systematic nuisance parameters λ_k, and the use of the deuteron correction factor $R_{d/N}^{CJ}(x, Q^2)$. These are calculated utilizing the fitted CJ15 PDFs, nuclear corrections, and $1/Q^2$ power corrections, which in turn depend on a number of fitted parameters reflecting the uncertainty of the analyzed experimental data.

The uncertainties in the PDF and other QCD fitting parameters can be propagated into any observable using the Hessian method discussed in Refs. [6,96]. This involves calculating the quantity of interest utilizing a set of "error PDF sets" and estimating the uncertainty by comparing the obtained results with the calculation that utilizes the best-fit, or "central," PDF set. For the purpose of the current analysis, we have used a modified version of the published sets of CJ15 error PDFs that include the higher-twist correction parameters needed for F_2 structure function calculation, in addition to the PDFs and off-shell correction parameters, for a total of 24 fitted parameters [7]. Each error set was then scaled along its eigendirection to ensure a more faithful determination of the $\Delta\chi^2 = 1.646$ error band than provided by a straightforward use of the customary Gaussian approximation. This modified PDF set, named CJ15nlo_mod, along with corresponding calculated DIS structure functions are publicly available, see Appendix B.

With the modified PDF set fixed, we have repeated the fit of λ^{norm} and λ_k and the calculation of $R_{d/N}^{CJ}(x, Q^2)$ with each error PDF set in turn, to obtain a set of 49 ($24 \times 2 + 1$) values of $n^{(j)}$, $\lambda_k^{(j)}$, and $R_{d/N}^{CJ,(j)}(x, Q^2)$, with $j = 0$ representing the values obtained with the central CJ15 set and $j = 1, \ldots, 48$ corresponding to the fits obtained with each CJ15 error set. With these, we can evaluate the uncertainties on any quantities of interest. For the nuisance parameters λ, the symmetric CJ uncertainty is defined as

$$\delta^{CJ}\lambda = \frac{1}{2}\sqrt{\sum_{j=1}^{24}[\lambda^{(2j-1)} - \lambda^{(2j)}]^2} \tag{10}$$

and nominally produces 90% confidence level uncertainties that correspond to a $t = 1.646$ "tolerance factor," see Refs. [7,24] and Appendix B. Other confidence levels can be obtained by scaling the result to the desired value of the tolerance. The resulting $\delta^{CJ}\lambda^{\mathrm{norm}}$ uncertainties on the normalization nuisance parameters, listed in Tables I and II, correspond to normalization uncertainties $\delta^{CJ}n$ ranging from 0.1% to 0.5% and are subleading compared to the uncorrelated experimental uncertainties. The $\delta^{CJ}\Delta_i$ uncertainties on the fitted systematic data shifts are similarly subleading.

The discussed sources of procedural uncertainties can be simultaneously accounted for by repeating the data cross-normalization procedure 48 times with each pair of error PDFs and defining the procedural uncertainty on the cross-normalized data as

$$\delta^{CJ}\hat{D}_i = \frac{1}{2}\sqrt{\sum_{j=1}^{24}[\hat{D}^{(2j-1)} - \hat{D}^{(2j)}]^2}. \tag{11}$$

This can be (somewhat conservatively) considered to be point-to-point uncorrelated. Finally, the cross-normalized data can be quoted as

$$\hat{D}_i = \hat{D}_i^{(0)} \pm \delta\hat{D}_i \pm \delta^{CJ}\hat{D}_i, \tag{12}$$

with the uncorrelated uncertainty $\delta\hat{D}_i = n^{(0)}\delta D_i$ obtained by propagating the experimental value. In the results presented here, these two uncertainties are summed in quadrature.

The evaluation of the procedural uncertainty on F_2^n follows in a similar manner, with the neutron extraction repeated 48 times in addition to the determination of its central value, discussed in Sec. III B,

$$\delta^{CJ}\hat{F}_2^n = \frac{1}{2}\sqrt{\sum_{j=1}^{24}[\hat{F}_2^{n,(2j-1)} - \hat{F}_2^{n,(2j)}]^2}. \tag{13}$$

As discussed for the data cross-normalization case, this uncertainty is point-to-point uncorrelated and quoted in addition to the uncorrelated δF_2^n experimental errors obtained by straightforward propagation of the uncorrelated δF_2^p and δF_2^d uncertainties. We can therefore represent the extracted F_2^n data points and their uncertainties as

$$\hat{F}_2^n = \hat{F}_2^{n(0)} \pm \delta\hat{F}_2^n \pm \delta^{CJ}\hat{F}_2^n, \tag{14}$$

with the F_2^n results shown in Fig. 3 as a function of Q^2 in selected bins of x. The treatment of the experimental and procedural uncertainties for the neutron-to-proton ratio $R_{n/p}$ (see Fig. 4) is analogous to that discussed above for the neutron structure function F_2^n. These extracted data are publicly available in the database discussed in Appendix A.

D. Bin-centered data

Bin centering of the F_2 data in Q^2 (or x) is not only useful graphically, as in Fig. 5, but also for applications such as the analysis of the Gottfried sum rule or the evaluation of structure function moments, to be discussed in Sec. IV, which are typically performed at fixed values of Q^2. We perform bin centering in Q^2, utilizing the CJ15 structure functions as the underlying model and defining the bin-centering ratio

$$R_{bc}(Q_0^2, Q^2) \equiv \frac{F_2(x, Q_0^2)}{F_2(x, Q^2)}\bigg|_{CJ}, \tag{15}$$

where Q_0 is the nominal center of the bin for any given structure function or structure function ratio. The bin-centered structure functions can then be calculated by multiplying the $\hat{F}_2$ values by the bin-centering ratio,

$$\tilde{F}_2(x, Q_0^2) \equiv R_{bc}(Q_0^2, Q^2)\hat{F}_2(x, Q^2). \tag{16}$$

The experimental and procedural uncertainties are propagated from $\hat{F}_2$ to $\tilde{F}_2$ for both the proton and neutron

structure functions. The bin-centering procedural uncertainty δ_{bc}^{CJ} is less than 0.1% and can be safely neglected. Bin centering in x can be performed in an analogous way if needed.

The Q^2 bin-centered datasets

$$\tilde{F}_2(x, Q_0^2) \pm \delta\tilde{F}_2 \pm \delta^{CJ}\tilde{F}_2, \tag{17}$$

for the proton, neutron, and deuteron structure functions and the d/p, n/d, and n/p ratios with $Q_0^2 = 2, 4, 6, 8, 10, 12, 15, 20, 30, 40,$ and 60 GeV^2 as in Fig. 5 are publicly available in the database discussed in Appendix A, alongside their values at the original experimental kinematics.

IV. APPLICATIONS OF NEUTRON DATA

In this section we discuss several applications of the extracted neutron dataset to the determination of the isovector nucleon structure function $F_2^p - F_2^n$ and its lowest moments, including the Gottfried sum rule.

A. Gottfried sum rule

The Gottfried sum is given by the integral over x of the isovector nucleon structure function $F_2^p - F_2^n$ scaled by the factor $1/x$ [38]. Since we will be interested also in the shape of the integrand of the Gottfried sum, as well as its saturation as $x \to 0$, it is convenient to define the truncated Gottfried integral

$$I_G(x_{min}, x_{max}; Q^2) = \int_{x_{min}}^{x_{max}} \frac{dx}{x}[F_2^p(x, Q^2) - F_2^n(x, Q^2)], \tag{18}$$

so that the Gottfried sum can be expressed as

$$S_G(Q^2) \equiv I_G(0, 1; Q^2). \tag{19}$$

At leading order (LO) in the strong coupling α_s and leading power in $1/Q^2$, the F_2 structure function can be written in terms of a sum of quark and antiquark PDFs, $F_2^{(LO)}(x, Q^2) = x\sum_q e_q^2(q + \bar{q})(x, Q^2)$. In this case, the Gottfried sum can be written as

$$S_G(Q^2) = \frac{1}{3}\int_0^1 dx(u + \bar{u} - d - \bar{d})(x, Q^2)$$
$$= \frac{1}{3} - \frac{2}{3}\Delta(Q^2), \tag{20}$$

where the integrated antiquark asymmetry is

$$\Delta(Q^2) = \int_0^1 dx[\bar{d}(x, Q^2) - \bar{u}(x, Q^2)]. \tag{21}$$

The constant term in the Gottfried integral (20) arises from the normalization of the valence quark distributions, $q_v \equiv q - \bar{q}$, equal to 2 or 1 for $q = u$ or $q = d$ quarks in

the proton, respectively. If one further assumes that the $\bar{d} - \bar{u}$ difference integrates to zero, one arrives at the canonical ("naive") Gottfried sum rule prediction of $S_G^{\text{naive}} = 1/3$ [38].

In contrast to many expectations, the value reported by the NMC from an analysis of deep-inelastic muon-hydrogen and muon-deuterium scattering data, $S_G^{\text{NMC}} = 0.235 \pm 0.026$ [97,98] at $Q^2 = 4$ GeV^2, was significantly lower than the naive flavor symmetric expectation, suggesting a strong violation of the sum rule and the first compelling evidence for $\bar{d} \neq \bar{u}$. This finding prompted tremendous excitement in the nuclear and particle physics community and spurred considerable work on both the theoretical [99,100] and experimental [101–104] fronts to better understand this violation. In particular, since higher-order corrections to the Gottfried sum in Eq. (20) from perturbative gluon radiation were found [105] to be very small numerically, it suggested that significant nonperturbative effects were likely to be responsible for the asymmetry. Such a nonperturbative effect was, in fact, predicted [106] from chiral symmetry breaking and the associated pion cloud of the nucleon and has become a standard explanation for the $\bar{d}$ excess of $\bar{u}$ in the proton [107,108]. A recent analysis also examined the impact of extrapolation methods and higher-twist effects on S_G [37].

The uncertainty on the NMC value for S_G arises from statistical and systematic errors, including from extrapolations of $F_2^p - F_2^n$ into the unmeasured regions at $x \to 0$ and $x \to 1$. The NMC analysis assumed that effects from nuclear corrections to F_2^n extracted from inclusive deuterium data were negligible, defining $(F_2^n/F_2^p)_{\text{NMC}} \equiv 2F_2^d/F_2^p - 1$, which assumes that $R_{d/N} = 1$. As discussed in the previous sections, however, nuclear binding and smearing corrections at large x [6,12,15] and nuclear shadowing at low x [109–112] give rise to clear deviations of $R_{d/N}$ from unity. In the present analysis, considerable effort has been made to account for the nuclear corrections in the extraction of the neutron F_2, as discussed in Sec. III, and in the following we examine the impact of those corrections on the Gottfried integral.

The analysis in Sec. III combined proton and deuteron F_2 structure function measurements from SLAC [50,85], BCDMS [88,89], NMC [84,87,95], and JLab [10] at a set of fixed Q^2 values. An illustration of the isovector nucleon structure function data, shifted to a common value of $Q^2 = 4$ GeV^2, is shown in Fig. 7 for the measured x range between $x = 0.009$ and $x = 0.603$. The plot also estimates the effects of neglecting deuteron corrections in the extraction of the neutron structure function by comparing the isovector $F_2^p - F_2^n$ structure function to the $2F_2^p - F_2^d$ combination of proton and deuteron structure functions. As one can see, deuteron corrections are generally comparable to the experimental uncertainties in the mid-x region and tend to suppress the isovector structure function or, equivalently, enhance the neutron structure function.
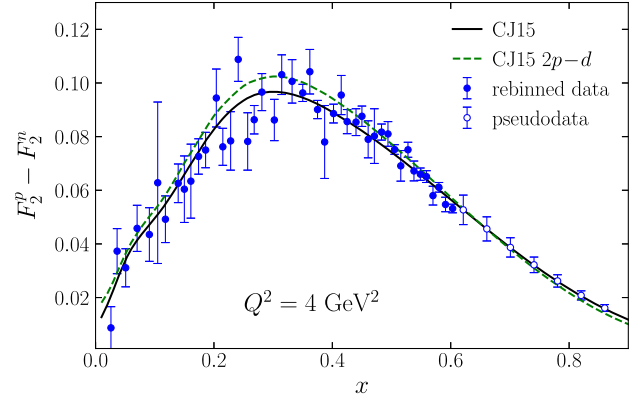

FIG. 7. Isovector nucleon structure function $F_2^p - F_2^n$ versus x from the combined reanalysis at $Q^2 = 4$ GeV^2 (filled circles), compared with the CJ15 global fit (solid black line) and the CJ15 $2F_2^p - F_2^d$ structure function combination (dashed green line), which would coincide with the isovector combination in the absence of nuclear effects. The pseudodata points in the extrapolated region at $x > 0.603$ (open circles) are generated from the CJ15 calculation with 5% nominal model uncertainties on F_2^p and F_2^n.

Computing the integral over all x values, as needed for the Gottfried sum [Eq. (19)], requires extrapolating the structure functions beyond the measured region to $x = 0$ and $x = 1$. Although our aim in the combined analysis of the world F_2^n data is to provide the best possible constraints on the structure functions and their moments, with minimal theoretical bias, such extrapolations will inevitably introduce model dependence into the procedure, especially for the $x \to 0$ behavior. We will discuss the uncertainties introduced into the extracted functions and moments from both the $x \to 0$ and $x \to 1$ extrapolations in the following.

To proceed, we consider the contributions to the total Gottfried integral (19) from individual x regions at a fixed value of Q^2,

$$S_G(Q^2) = I_G(0, x_{\text{low}}; Q^2) + I_G(x_{\text{low}}, x_{\text{high}}; Q^2) \\ + I_G(x_{\text{high}}, 1; Q^2), \tag{22}$$

where x_{low} and x_{high} are the lower and upper bounds of the experimental data, both of which vary with Q^2 (see Table III). The smallest value of x_{low} for any of the bins from the NMC data [84,87,95] is $x = 0.004$. The individual contributions to $S_G(Q^2)$, along with the total extrapolated moments, are displayed in Fig. 8 for several Q^2 values between $Q^2 = 2$ and 30 GeV^2. From kinematics, one finds that the values of both x_{low} and x_{high} increase with increasing Q^2. This has the effect of increasing the relative contribution from the extrapolated region at low x and decreasing the relative contribution from the large-x extrapolation.

TABLE III. Contributions $I_G(x_{min}, x_{max}; Q^2)$ to the Gottfried integral from different intervals of x at fixed Q^2 values, along with the total integral $S_G(Q^2)$. The experimentally measured region corresponds to (x_{low}, x_{high}). The range $(0.004, 1)$ is provided for comparison with the NMC experiment [97,98].

Q^2 (GeV^2)	x_{low}	x_{high}	$(0, x_{low})$	(x_{low}, x_{high})	$(x_{high}, 1)$	$(0.004, 1)$	S_G
				I_G			
2	0.005	0.422	0.011(6)	0.175(13)	0.039(8)	0.218(16)	**0.228(16)**
4	0.009	0.603	0.018(7)	0.188(19)	0.011(2)	0.207(19)	**0.218(20)**
6	0.014	0.690	0.025(8)	0.177(21)	0.004(0)	0.195(22)	**0.207(23)**
8	0.024	0.747	0.035(7)	0.199(22)	0.002(0)	0.224(22)	**0.236(23)**
10	0.028	0.781	0.038(6)	0.175(31)	0.001(0)	0.204(33)	**0.214(31)**
12	0.035	0.819	0.045(7)	0.210(14)	0.000(0)	0.245(15)	**0.256(16)**
15	0.037	0.851	0.048(7)	0.189(14)	0.000(0)	0.225(14)	**0.237(16)**
20	0.053	0.877	0.062(9)	0.189(10)	0.000(0)	0.240(11)	**0.252(13)**
30	0.072	0.896	0.077(10)	0.166(13)	0.000(0)	0.232(14)	**0.243(16)**

For the measured region $I_G(x_{low}, x_{high}; Q^2)$, the uncertainties were estimated using $\gtrsim 10^4$ Monte Carlo simulations based on the total systematic and statistical errors reported by each experiment, as discussed in Sec. III. At the ith iteration, each data point D_j was shifted by $\lambda_i \delta_j^{sys}$, where λ_i is a random number generated by the standard normal distribution, and δ_j^{sys} is the total systematic uncertainty of the jth data point. All points with their statistical uncertainties were then binned in x with a bin size $\Delta x = 10^{-4}$. A trapezoid integration method with adaptive step size was used with the rebinned data points to obtain an integral I_i ($i = 1, ..., 10^4$). The mean and standard deviation of I_i were recorded as the central value and total error of $I_G(x_{low}, x_{high}; Q^2)$, respectively. In practice, the step size Δx was varied between 10^{-5} and 10^{-1} to check the integral stability, with $\Delta x = 10^{-4}$ found to be stable and in good agreement with corresponding analytical results.
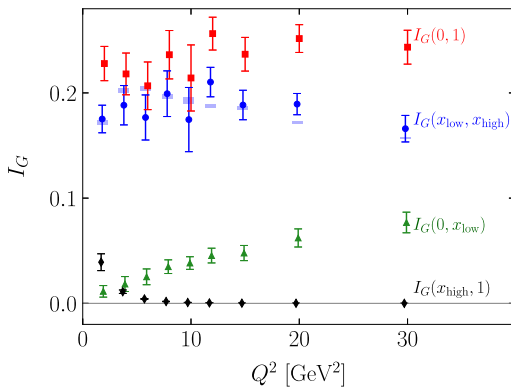

FIG. 8. Variation with Q^2 of the contributions to the Gottfried integral $I_G(x_{min}, x_{max}; Q^2)$ from different x intervals, including from the measured (blue circles), $x \to 0$ extrapolated (green triangles), and $x \to 1$ extrapolated (black diamonds) regions, along with the total integral (red squares). The blue shaded bands show the integral in the measured region calculated from the CJ15 PDFs.

For the unmeasured low-x region contribution $I_G(0, x_{low}; Q^2)$, we followed previous analyses [42,84,87,95] by assuming a parametrization for the nonsinglet structure function inspired by Regge theory, with functional form $F_2^p - F_2^n = Ax^\alpha$. The parameters A and α were fitted to data at $0.01 < x < 0.1$ for each value of Q^2. We found good fits to the data with $\alpha = 0.6$, which also described well the nonsinglet structure function calculated perturbatively with the CJ15 PDFs in the unmeasured $x < x_{min}$ region. The corresponding contributions to GSR integrals were then calculated analytically.

At the lowest Q^2 value, $Q^2 = 2$ GeV^2, the unmeasured region contributes $\approx 5\%$ of the total integral. This fraction becomes larger with increasing Q^2 due to the more restricted range of low-x data accessible at the higher Q^2 values (see Fig. 8), rising up to $\approx 13\%$ at $Q^2 = 10$ GeV^2. Since the $x \to 0$ extrapolated contribution to the Gottfried integral is not negligible, it is important to estimate the systematic uncertainty arising from this component. For this purpose, the value of α was varied between 0.5 and 0.7, and the resulting differences in the truncated GSR integral are taken as a measure of the systematic uncertainty. A small contribution from the uncertainty in the normalization parameter A was also folded into the total error. For all Q^2 points considered, the magnitude of the uncertainty was found to be $\approx 3\% - 4\%$ of the total integral.

The importance of the low-x contribution to the Gottfried integral is also evident in Fig. 9, which shows the dependence of $I_G(x_{min}, 1; Q^2)$ on the lower limit x_{min} of the x integration at a fixed value of $Q^2 = 4$ GeV^2. The results are compared with the original extraction from the NMC data analysis [87,95], as well as from the more recent analysis of Abbate and Forte (AF) [42], and illustrate the saturation of the integral as $x_{min} \to 0$. While the results from our global reanalysis agree with the previous results at higher x_{min} values $x_{min} \gtrsim 0.2$, at lower x_{min} our extracted integrals are slightly below both the earlier NMC and AF results. In our analysis, we also
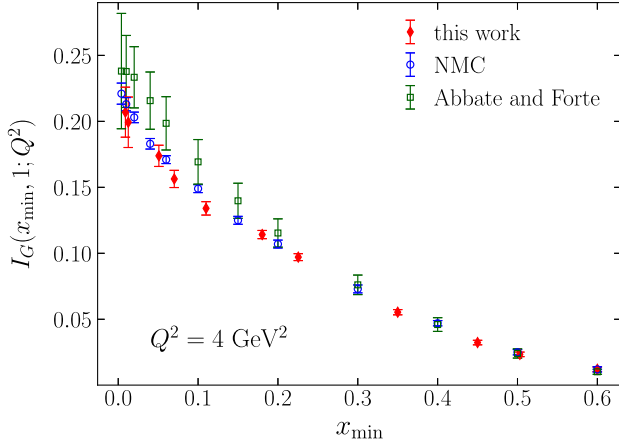
FIG. 9. Dependence of the Gottfried integral $I_G(x_{min}, 1; Q^2)$ on the lower limit of the x integration x_{min} at $Q^2 = 4$ GeV^2 in our analysis (red diamonds), compared with the NMC extraction [87,95] (open blue circles) and the analysis by Abbate and Forte [42] (open green squares).

removed shadowing corrections from the deuterium data, resulting in a 1%–2% increase of F_2^d at $x \lesssim 0.1$ and hence a comparable increase in the extracted F_2^n and a decrease of the isovector $F_2^p - F_2^n$ and the corresponding integrated value in the same region of x.

For the unmeasured large-x region, $x_{high} < x < 1$, the inelastic contribution to the integral was evaluated from the CJ15 PDFs [7], as well as an empirical fit that describes structure function data in both the DIS and resonance regions [113,114]. The latter naturally cuts off at x values larger than the pion threshold $x_\pi = Q^2/(W_\pi^2 - M^2 + Q^2)$, where $W_\pi = M + m_\pi$ is the minimum mass of the inelastic final state. The former suffers from the "threshold problem" affecting TMC calculations in momentum space [115–118], and the integrals are limited to $x < x_\pi$. In either case, the elastic contribution to the Gottfried integral was not included. The average of the two calculated values gives the central value of the high-x contribution, and their difference provides an estimate of the systematic extrapolation uncertainty.

The final results for the contributions to the Gottfried integral from the various regions of x are summarized in Table III. At a reference scale of $Q^2 = 4$ GeV^2, we find the contribution from the region $x > 0.004$, corresponding to the range given in the NMC analysis [98], to be $I_G(0.004, 1) = 0.207(19)$, where for notational convenience we suppress the Q^2 dependence in I_G. This is slightly smaller than the value found in the original NMC data analysis [84], $I_G^{NMC}(0.004, 0.8) = 0.221(21)$, but is consistent within the uncertainties. It is also compatible with the more recent AF determination [42], $I_G^{AF}(0.004, 0.8) = 0.228(44)$, which has a somewhat larger uncertainty. Including the contributions from the extrapolation into the unmeasured regions $x \to 0$ and $x \to 1$, we find the total Gottfried integral at $Q^2 = 4$ GeV^2 to be $S_G = 0.218(20)$, which again is slightly smaller than the values extracted from
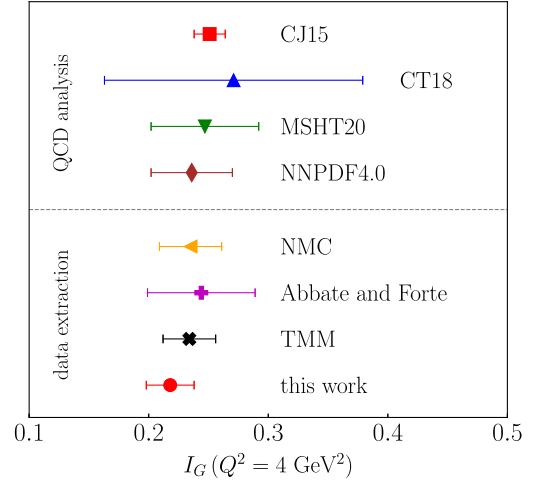

FIG. 10. Gottfried integral $I_G(0, 1; Q^2 = 4$ GeV$^2)$ from data extractions (bottom) and from theoretical calculations using PDF obtained in recent global QCD analyses (top). The smaller CJ15 uncertainty compared to the other QCD calculations is due to the collaboration's choice of a nominal $t = 1.646$ tolerance factor, which is smaller than in the dynamical tolerance criteria adopted in the other analyses. The large CT18 uncertainty is due to bigger PDF uncertainties at $x \lesssim 0.001$ compared to NNPDF4.0 and MSHT20.

the NMC data alone, $S_G^{NMC} = 0.235(26)$ [84], or in the AF analysis, $S_G^{AF} = 0.242(48)$ [42], or from the recent reanalysis of NMC data with the truncated methods, $S_G^{TMM} = 0.234(22)$ [37]. A comparison of these data extractions is shown in Fig. 10, along with calculations using global QCD analyses from CJ15 [7], CT18 [119], MSHT20 [73], and NNPDF4.0 [120].[1] Taken collectively, these latter values appear to slightly overestimate the data extractions.

To estimate the higher-twist contributions to the Gottfried integral, we compare the results using the CJ15 calculation, including and neglecting the subleading $1/Q^2$ corrections to the structure functions. We find effects for S_G of the order 0.1%, so that to a good approximation we can use Eq. (20) to convert our extracted S_G value to an integrated LO flavor asymmetry in the proton of $\Delta = 0.173(30)$ at $Q^2 = 4$ GeV^2, which is slightly larger than the NMC result $\Delta^{NMC} = 0.148(39)$ [84], but consistent within the quoted uncertainties. It is also somewhat larger than the integrated asymmetry extracted from the E866 Drell-Yan data [104] (which were used to constrain the $\bar{d}/\bar{u}$ ratio) when combined with the parametrization of $\bar{d} + \bar{u}$ from the CTEQ5 global QCD analysis [125], $\Delta^{E866} = 0.118(12)$. Our reanalysis of the neutron structure function data therefore suggests a stronger violation of SU(2) flavor asymmetry in the proton sea than reported in the previous studies.

Within the chiral effective field theory framework of Refs. [107,126], in which the nonperturbative sea in the

---

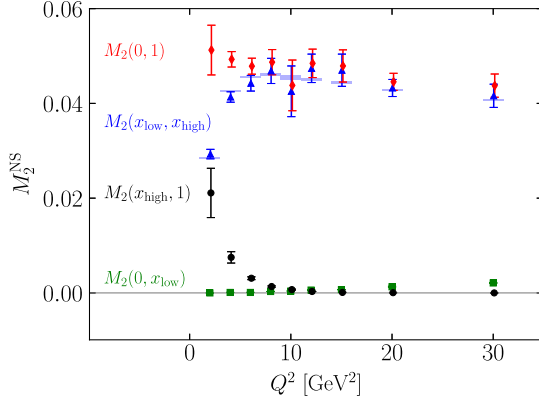[1]The calculations were done with open-source code APFEL++ [121,122] using PDF sets from the LHAPDF library [123,124].

FIG. 11. Nonsinglet Nachtmann moment M_2^{NS} extracted from world DIS data at various values of Q^2. The total moments (red diamonds) are broken down into contributions from different x intervals: the measured region $[x_{low}, x_{high}]$ (blue triangles), the $x > x_{high}$ extrapolation region (black circles), and the $x < x_{low}$ extrapolation region (green squares). The blue shaded bands represent the integral in the measured region calculated using CJ15 structure functions instead of data points.

proton is generated from pseudoscalar pion loops, an integrated asymmetry of $\Delta \approx 0.18$ would correspond to an average π^+ multiplicity in the quantum fluctuation of a proton to a π^+ and neutron state of $\langle n \rangle_{\pi^+ n} \approx 0.25$. In comparison, a smaller value of $\Delta \approx 0.12$ as obtained from the E866 analysis [104] would correspond to a π^+ multiplicity of $\langle n \rangle_{\pi^+ n} \approx 0.15$. The larger deviation from the Gottfried sum rule observed in our analysis implies, therefore, a $\gtrsim 50\%$ larger pion cloud than that suggested by the previous studies.

B. Nonsinglet moments

In recent years, developments in lattice QCD have enabled precision calculations of moments of PDFs from first principles [32,33]. Comparison of the calculated

nonsinglet quark distribution moments can be made with moments extracted from experimental data, providing a valuable test of the lattice methodology and various high-order corrections, as well as of QCD itself. The nonsinglet moment of the u- and d-quark PDFs accessible to lattice QCD is given by

$$\langle x \rangle_{u^+ - d^+} = \int dx\, x[u(x) + \bar{u}(x) - d(x) - \bar{d}(x)], \quad (23)$$

which corresponds to the difference between the momentum carried by u and d quarks in the proton, computed at some resolution scale usually set by the lattice spacing. For a direct comparison between the calculated PDF moments and those extracted from experiment, we consider the Nachtmann moment of the F_2^N structure function of the nucleon N, which accounts for kinematical target mass effects associated with higher-spin operators,

$$M_2^{p-n}(Q^2) = \int_0^1 dx\, \frac{\xi^3}{x^3} \left[\frac{3 + 9r + 8r^2}{20} \right] F_2^{p-n}(x, Q^2), \quad (24)$$

where $\xi = 2x/(1+r)$ is the Nachtmann scaling variable [127,128], with $r = \sqrt{1 + 4M^2 x^2/Q^2}$. We neglect the small difference between the proton and neutron masses, by setting the nucleon mass to $M = 0.939$ GeV.

Using the same rebinned F_2^p and F_2^n datasets used in the GSR analysis and deploying the integration technique described in the previous section to the integral in Eq. (18), we then extract the nonsinglet moments,

$$M_2^{NS}(Q^2) = M_2^p(Q^2) - M_2^n(Q^2), \quad (25)$$

from the world DIS data. To gauge the relative contribution of the measured and extrapolation regions to the full moment, we also compute the truncated moments $M_2^{NS}(x_{min}, x_{max}; Q^2)$ by restricting the x integration in

TABLE IV. Contributions to the nonsinglet Nachtmann moment $M_2^{NS}(x_{min}, x_{max}; Q^2)$ from various intervals of x at fixed Q^2 values from 2 to 10 GeV^2, along with the total integral. The experimentally measured region corresponds to (x_{low}, x_{high}). Analysis methods are the same as described in Sec. IV A. The column labeled "HT" gives the higher-twist contribution to the total integral $M_2^{NS}(0, 1; Q^2)$ as a percentage.

			M_2				
Q^2 (GeV^2)	x_{low}	x_{high}	$(0, x_{low})$	(x_{low}, x_{high})	$(x_{high}, 1)$	$(0, 1)$	HT (%)
2	0.005	0.408	0.0000(0)	0.0293(10)	0.0211(52)	**0.0513(53)**	3.7
4	0.009	0.603	0.0001(0)	0.0414(11)	0.0075(12)	**0.0493(16)**	1.1
6	0.014	0.690	0.0001(0)	0.0442(16)	0.0031(3)	**0.0479(17)**	0.5
8	0.024	0.747	0.0003(0)	0.0467(26)	0.0014(2)	**0.0487(27)**	0.3
10	0.028	0.791	0.0004(0)	0.0425(52)	0.0007(1)	**0.0438(54)**	0.2
12	0.035	0.819	0.0006(0)	0.0474(30)	0.0003(0)	**0.0485(30)**	0.2
15	0.037	0.851	0.0007(0)	0.0470(34)	0.0001(0)	**0.0479(34)**	0.2
20	0.053	0.877	0.0013(1)	0.0433(18)	0.0001(1)	**0.0446(18)**	0.1
30	0.072	0.896	0.0021(1)	0.0416(25)	0.0000(0)	**0.0438(25)**	0.1

TABLE V. Nonsinglet $\langle x \rangle_{u^+ - d^+}$ PDF moments at $Q^2 = 4$ GeV^2 obtained from DIS data extractions, lattice QCD calculations, and PDF global analyses. The lattice values are envelopes of the calculations shown in Fig. 12 for each quark flavor scheme, and the global QCD analyses results are described in the main text.

Data extractions		Lattice QCD envelope		
This work	E06-009	$N_f = 2$	$N_f = 2 + 1$	$N_f = 2 + 1 + 1$
0.143(5)	0.138(14)	0.189(23)	0.160(49)	0.173(21)

Global PDF analyses							
MMHT2014	CJ15	NNPDF3.1	CT18	JAM21	ABMP2016	HERAPDF2.0	Combined
0.151(4)	0.152(2)	0.152(3)	0.156(7)	0.157(2)	0.167(4)	0.188(3)	$0.156^{(+20)}_{(-6)}$

Eq. (24) to the $[x_{\min}, x_{\max}]$ interval. The results, including the contributions from each of the integration regions, are shown in Fig. 11 as a function of Q^2 and listed in Table IV. The low-x extrapolation region gives a negligible contribution to the moments, while the contribution from the large-x region increases with decreasing Q^2; indeed, the unmeasured large-x interval widens as Q^2 becomes smaller. As with the GSR, the extracted moments are only weakly dependent on Q^2, suggesting a possibly large cancellation between the proton and neutron HT components.

Within the QCD factorization approach, the nonsinglet Nachtmann moment $M_2^{\rm NS}$ can be related to the nonsinglet moment $\langle x \rangle_{u^+ - d^+}$ of the quark PDFs by dividing out the perturbative Wilson coefficient and subtracting possible HT contributions,

$$\frac{3}{C_2} M_2^{\rm NS} = \langle x \rangle_{u^+ - d^+} + {\rm HT}, \qquad (26)$$

where to $\mathcal{O}(\alpha_s)$ the Wilson coefficient $C_2 = 1 + 1.0104 \alpha_s(Q^2)/4\pi$ [129]. We can estimate the size of the HT contribution by calculating the CJ15-calculated $M_2^{\rm NS}$ with and without HT corrections, reporting the relative effect in Table IV. We find that the HT contribution is typically smaller than the extracted data uncertainty and becomes comparable to this only at the lowest Q^2 value.

In particular, at $Q^2 = 4$ GeV^2, the HT contribution is negligible compared to the $M_2^{\rm NS}$ extraction uncertainties, and inverting Eq. (26) we find $\langle x \rangle_{u^+ - d^+} = 0.143(5)$ at this scale. This result is shown in Fig. 12 compared with a recent experimental extraction from precision proton and deuteron F_2 structure function data from the E06-009 experiment at JLab [36] and with various lattice QCD calculations as reviewed in Refs. [32,33]. In addition, our extracted moment is also compared with an average of PDF nonsinglet moments using PDF parametrizations from recent global QCD analyses, using Eq. (23) directly. The corresponding numerical values of all the moments are given in Table V.

In the analysis of the E06-009 data [36], nuclear corrections were removed from the deuteron data by

adopting the same nuclear convolution model as that utilized here, but using as input the phenomenological structure functions that were fitted directly to inclusive DIS data [113]. The E06-009 analysis also considered the elastic contribution to the nonsinglet moment and resulted in a value of $\langle x \rangle_{u^+ - d^+}^{\rm E06\text{-}009} = 0.138(14)$. With the elastic contribution removed, this would be reduced by 3% to $\langle x \rangle_{u^+ - d^+}^{\rm E06\text{-}009} = 0.133(14)$, which is well within the quoted uncertainties. As can be appreciated from Fig. 12, the E06-009 moment is consistent with that found in our analysis, albeit with a larger uncertainty.

The nonsinglet PDF moments extracted from experimental data appear systematically below those from lattice QCD calculations. This is clear from the comparisons in Fig. 12, where recent lattice QCD moments [130–135] have been extrapolated to the physical pion mass, as reviewed in Ref. [33]. The small tension with the data extraction may be indicative of residual unaccounted for systematic effects in the lattice extraction of the PDF moments.
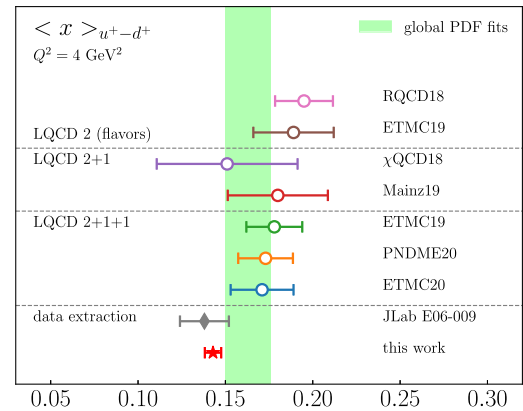

FIG. 12. A comparison of the nonsinglet PDF moment $\langle x \rangle_{u^+ - d^+}$ extracted from DIS structure functions data (solid symbols) and from various lattice QCD (LQCD) calculations (open circles) at $Q^2 = 4$ GeV^2. The green vertical band shows the unweighted combination of moments from recent global PDF analyses discussed in the text.

The nonsinglet moments $\langle x \rangle_{u^+ - d^+}$ obtained from an average of recent global QCD analyses, namely, the ABMP16 [136], CJ15 [7], CT18 [119], HERAPDF2.0 [137], JAM19 [138], MMHT2014 [72], and NNPDF3.1 [139] PDF parametrizations, are compared in Fig. 12 with some recent lattice QCD simulations [32,33] and extractions from data. How to average observables calculated from different global PDF fits is an open question [24,120], and in this analysis we quote the median of the individual central values as the central value of the combined moment and asymmetric standard deviations from the median as systematic errors. The statistical uncertainties on the individual calculations do not vary substantially, and a simple average represents these well. The averaged result is found to be $\langle x \rangle_{u^+ - d^+}^{\text{PDF}} = 0.156(4)_{-(6)}^{+(20)}$, with the error band in Fig. 12 representing the sum in quadrature of the statistical and systematic uncertainties.

The phenomenological results thus obtained from the global QCD analyses lie between the DIS data extraction and the lattice QCD calculations. This may suggest that DIS data are in mild tension with the Drell-Yan and jet data from proton-proton collisions that have been included in the global analyses to complement the DIS data. On the other hand, the tension between lattice simulations and phenomenological results may be milder than that indicated by comparison with the data extractions alone.

V. CONCLUSIONS

We have performed a detailed reanalysis of the world's inclusive DIS data, simultaneously obtained on protons and deuterons, to extract the structure function F_2^n of the free neutron. To account for the nuclear effects in the deuteron, we have consistently applied the nuclear correction calculation from the CJ15 global QCD analysis [7], which includes calculated nuclear smearing and, in order to minimize the theoretical uncertainties, fitted nucleon off-shell corrections. Special attention has been devoted to the normalization of the proton and deuteron experimental datasets and to the treatment of correlated systematic errors, as well as the quantification of procedural and theoretical uncertainties. The data themselves have been carefully cross-normalized and shifted point by point, as allowed by the correlated systematic uncertainties, which turns out to be essential for minimizing the uncertainties in the extracted F_2^n.

As applications of the extracted neutron structure function, we considered a reevaluation of the Gottfried sum rule, including for the first time its Q^2 dependence, and extracted the nonsinglet $F_2^p - F_2^n$ moment that provides an experimental benchmark for precision lattice QCD simulations. In both cases, we have carefully evaluated the statistical and systematic uncertainties. The Q^2 dependence was found to be rather flat, within the uncertainties, indicating strong cancellations of higher-twist power corrections for the proton and neutron.

A comprehensive database including the world data on inclusive DIS from proton and deuteron targets, as well as the extracted neutron structure function and neutron-to-proton ratio, has been made publicly available (see Appendix A). To facilitate replication of our study and a comparison with other nuclear correction models, as well as for general applications, we also provide precomputed (x, Q^2) grids in LHAPDF format for calculating the modified CJ15 PDFs used in our study, named CJ15nlo_mod, and the corresponding DIS structure functions (see Appendix B). As a demonstration of the use of the DIS structure function grids, we illustrate neutron excess correction ratios for neutral and charged current DIS on nuclear targets (see Appendix C). We expect these resources will be useful in future phenomenological applications requiring data on the neutron F_2^n structure function.

The supporting datasets for this paper are openly available from the GitHub repository [140]. The modified CJ15nlo_mod PDF set, as well as the corresponding proton, neutron, and deuteron CJ15nlo_mod_SF structure functions are publicly available on the CJ website [74].

APPENDIX A: THE CTEQ-JLab DIS DATABASE

Starting from the existing experimental datasets used in the CJ15 global QCD analysis [7], a comprehensive database was constructed of unpolarized DIS measurements with proton and deuteron targets, along with the neutron target structure functions extracted in this analysis. The experimental observables included in the database are

(i) the F_2 structure functions, and their d/p and n/d ratios,

(ii) the longitudinal to transverse cross section ratio $R = \sigma_L / \sigma_T$, and

(iii) the reduced cross section $\sigma_{\text{red}}(x, Q^2) = F_2(x, Q^2) - (y^2 / Y^+) F_L(x, Q^2)$,

where $y = \nu / E$ is the lepton inelasticity, and $Y^+ = 1 + (1 - y)^2 + 2M^2 x^2 y^2 / Q^2$, with E the incident lepton energy and ν the energy transfer in the target rest
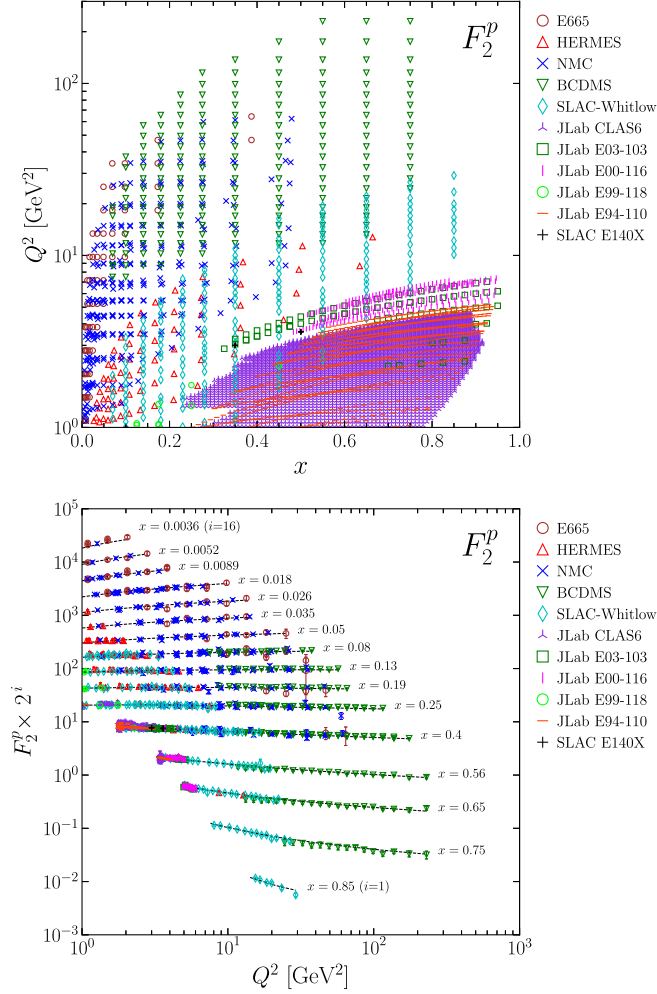
FIG. 13. Kinematic range (upper) and proton F_2^p data included in this database, binned in x (lower). The proton F_2^p values are cut on $W^2 = 3.5$ GeV^2. Note that the available $\sigma_{\rm red}$ data will have a different (larger) kinematic coverage.

FIG. 14. As in Fig. 13, but for the deuteron F_2^d data.

frame. The reduced cross section datasets were extracted from experimental cross sections according to

$$\sigma_{\rm red}(x, Q^2) = \frac{xQ^4}{2\pi\alpha^2 Y^+} \frac{{\rm d}^2\sigma}{{\rm d}x{\rm d}Q^2}.$$

For the first time, we have also included the full DIS datasets from the JLab 6 GeV program, which expanded the kinematic coverage in the high-x region (see Figs. 13 and 14 below). A list of included experiments and observables is shown in Table VI. The F_2^n and F_2^n/F_2^p data points were extracted, respectively, from the matched proton and deuteron data and from the measured d/p and n/p ratios (see Sec. III A). Extensive efforts were made to collect details of normalization and correlated and uncorrelated systematic uncertainties on all the datasets.

All datasets are maintained in a public GitHub repository [140] in both Excel and plain text formats (the latter
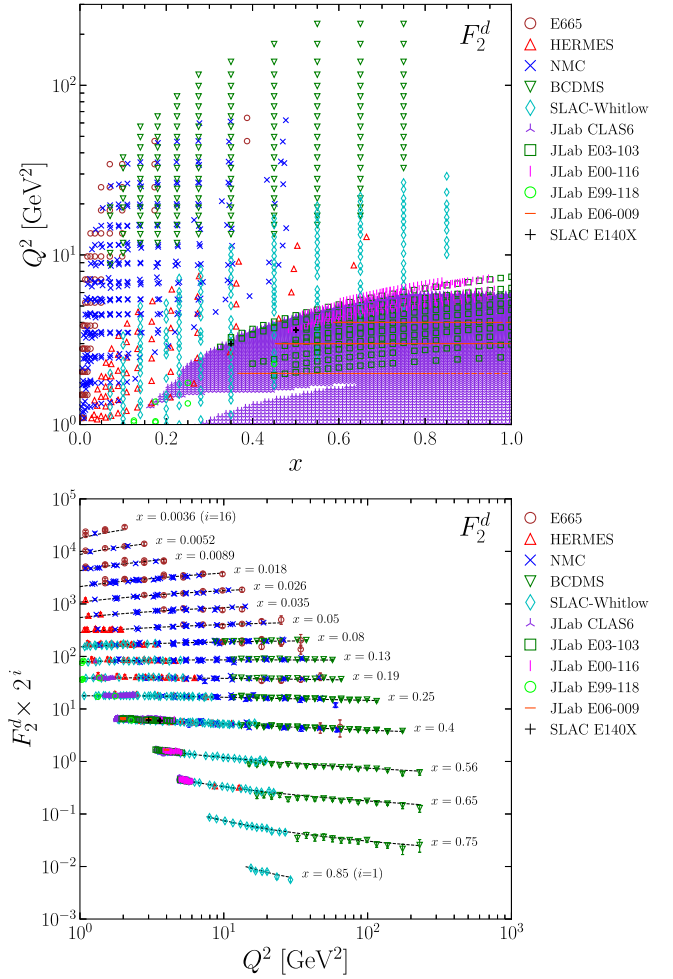
compatible with the CJ15 fitting package). A five-digit identifier is assigned to each dataset, and the content, references, source of uncertainties, and related calculations are documented in each README file.

APPENDIX B: THE CJ15nlo_mod PDF AND STRUCTURE FUNCTION GRIDS

The careful evaluation of theoretical uncertainties in the extraction of the neutron structure functions in the main text necessitated the repeated evaluation of NLO F_2 structure functions using a modified version of the 49-member strong CJ15nlo PDF set, which represents uncertainties stemming from variations of 19 PDF parameters, two off-shell parameters, and three HT parameters [7] with a nominal 90% confidence level [7]. As a result of this effort, we present here the CJ15nlo_mod set of PDFs, as well as a corresponding set of calculated DIS structure functions.

The modification of the published CJ15 PDF error set was needed to take care of deviations from the assumed Gaussian behavior of PDFs and observables around the

TABLE VI.   List of experiments and observables included in the DIS database. Datasets marked with "$*$" were also included in the CJ15 global QCD analysis. The extracted F_2^n and $R_{n/p}$ are also provided.

Experiment	$\sigma_{\rm red}$	F_2	σ_L/σ_T	F_2^n extracted	$R_{n/p}$ extracted
SLAC-Whitlow [85]	$p, d, d/p$	$p*, d*$	p, d	✓	✓
SLAC-E140 [141]			d		
SLAC-E140x [86]	p, d	p, d	p, d	✓	
NMC [87,95]	$p, d, d/p*$	$p*, d$		✓	✓
BCDMS [88,89]	p, d	$p*, d*$	p, d	✓	
JLab E06-009 [142]	d	d			
JLab E94-110 [143,144]	p	p			
JLab E03-103 [92,93]	p, d	p, d		✓	
JLab E99-118 [91,145,146]	$p, d, d/p$	$p, d, d/p$		✓	
JLab JLCEE96 [147]	p				
JLab E00-116 [94,148]	p, d	$p*, d*$		✓	
JLab CLAS6 [149–152]	p, d	p, d			
JLab BONuS [10,80]		$n, n/d*$			✓
HERA I + II [137]	$p*$				
HERMES [90]	$p, d, d/p$	$p*, d*$		✓	✓
E665 [153]		p, d			

fitted parameter values in the least constrained regions of parameter space, such as for observables sensitive to the d quark at large-x values. We have, for example, observed non-negligible deviations from the assumed quadratic behavior of the χ^2 function in the parameter subspace spanned by the power correction parameters, which, in turn, is correlated with the d-quark parameters governing its large momentum behavior. As a result, the curvature of the χ^2 function in the vicinity of the minimum is typically underestimated by the eigenvalues of the numerically evaluated Hessian matrix, which then does not faithfully capture the global fit uncertainties. One consequence of relevance for the present paper is that the uncertainties of the $R_{d/N}$ ratio calculated with the master formula (10) and the published error PDF set are typically smaller than those displayed in the figures of Ref. [7] that, for simplicity, used the Hessian approximation for the structure function ratios.

In order to compensate for the observed deviations from the expected Gaussian behavior, we have then followed the procedure discussed in Refs. [27,154]. Namely, we have rescaled the CJ15nlo PDF error sets along each eigendirection to ensure that they produce a $\Delta\chi^2 = 1.646$ in the fitted datasets compared to the best-fit CJ15nlo PDF set. The rescaling factors are close to 1 in the majority of the cases, except, indeed, for the least constrained eigendirections, but produce non-negligible effects, for example, for the evaluation of the uncertainties of the $R_{d/N}$ ratio. Further discussion of Gaussian deviations and possible remedies can be found in the aforementioned references.

Finally, to facilitate the reproduction of our results, as well as for general use, we have made publicly available on the CJ website [74] the modified 49-member CJ15nlo_mod PDF set, as well as the corresponding proton, neutron, and deuteron CJ15nlo_mod_SF structure functions, as precomputed x, Q^2 grids on the CJ Collaboration's web page. All grids are provided in LHAPDF format [155] for easy Python and C access and will soon be submitted for inclusion in the official LHAPDF website [156]. The F_2 neutral current structure functions are presented both with and without TMC and HT corrections. The F_L neutral current structure function and the charged current structure functions are only available at leading twist, since the CJ15 fit lacked the data to constrain their power corrections. The indexing of structure functions within each LHAPDF grid follows the conventions laid out for the inclusive DIS studies presented in Ref. [157] and is explicitly discussed in Ref. [158].

APPENDIX C: ISOSCALAR CORRECTIONS WITH CJ15_mod STRUCTURE FUNCTIONS

When comparing the experimentally measured lepton-nucleus cross sections for different nuclear targets, a correction to account for the proton and neutron number imbalance (or target nonisoscalarity) is often necessary to isolate dynamical nuclear effects that go beyond the trivial difference between the proton and neutron scattering cross section contributions. This process has been widely discussed in the literature, and we briefly rederive it here before discussing the use of the CJ15nlo_mod proton and neutron structure functions, described in Appendix B, for calculating the needed correction factor.

For a nucleus A with Z protons and N neutrons, the nuclear effects in the lepton-nucleus cross section can be quantified by considering the ratio

$$R_A^{\rm nuc} = \frac{\sigma_A}{Z\sigma_p + N\sigma_n} \tag{C1}$$

of the nuclear cross section σ_A to the sum of Z free proton and N free neutron cross sections, denoted by σ_p and σ_n, respectively. (The subscript A labels the nucleus under consideration, rather than the specific atomic number.) It is then straightforward to obtain

$$R_A^{\text{nuc}} = \frac{\sigma_A f_A^{\text{iso}}}{\frac{1}{2} A(\sigma_p + \sigma_n)}, \qquad (\text{C2})$$

where the so-called "isospin correction factor,"

$$f_A^{\text{iso}} = \left(\frac{A}{2}\right) \frac{1 + \sigma_n/\sigma_p}{Z + N\sigma_n/\sigma_p}, \qquad (\text{C3})$$

only depends on the ratio of free neutron to free proton cross sections. Note that the denominator of the right-hand side of Eq. (C2) is just a scaled, isospin-symmetric free nucleon target cross section. By analogy, one can then interpret

$$\sigma_A^{\text{iso}} \equiv \sigma_A f_A^{\text{iso}} \qquad (\text{C4})$$

as an isospin-symmetrized nuclear cross section, with f_A^{iso} providing the conversion from the measured σ_A to the isoscalar cross section σ_A^{iso}. Having thus removed the trivial nuclear effects due to the number imbalance of protons and neutrons in a given nucleus, one can compare and contrast nuclear effects in different nuclei on the same footing, such as in the double ratios $(\frac{1}{A}\sigma_A^{\text{iso}})/(\frac{1}{B}\sigma_B^{\text{iso}})$ of the symmetrized cross sections for nuclei A and B.

In practice, the cross section ratio data are often converted to structure function ratios $\sigma_A/\sigma_d \approx F_2^A/F_2^d$, before applying the isoscalarity corrections, under the assumption that the longitudinal to transverse cross section ratio σ_A^L/σ_A^T is independent of the nuclear target. Repeating the derivation above, one can see that the isoscalar correction factor f_A^{iso} would analogously apply to the nuclear structure function. However, one should keep in mind that, while a relatively small effect, the A independence assumption for the longitudinal to transverse ratio may not be accurate at large-x and low-Q^2 values [145,147,159].

Experimentally, in the absence of free neutron targets, the nuclear modification ratio (C1) is often approximated by the "nuclear EMC ratio,"

$$R_A^{\text{EMC}} = \left(\frac{2}{A}\right) \frac{\sigma_A^{\text{iso}}}{\sigma_d}, \qquad (\text{C5})$$

of the per-nucleon nuclear and deuteron cross sections, with the isoscalar correction applied to the numerator (although sometimes this correction is not applied). A nonzero nuclear modification in the deuteron $R_d^{\text{nuc}} \neq 1$, was recently observed by the BONuS Collaboration [80], with further precision measurements planned [55], and shown to

be non-negligible, especially at large values of x (see also Fig. 2). The neutron data extracted in Ref. [80] obviate the need for using deuteron targets as proxies for the combination of free proton and neutron targets, enabling one to directly measure R_A^{nuc} and fully expose the nuclear effects in the nuclear target.

The CJ15nlo_mod proton and neutron structure functions obtained in this work (see Appendix B) can be directly used to calculate the isoscalar correction factor f_A^{iso} at any x and Q^2. Indeed, for events with not too large lepton energy loss, one can approximate

$$\frac{\sigma_n}{\sigma_p} \approx \frac{F_2^n}{F_2^p} \qquad (\text{C6})$$

and obtain

$$f_A^{\text{iso}} \approx \left(\frac{A}{2}\right) \frac{1 + F_2^n/F_2^p}{Z + NF_2^n/F_2^p}, \qquad (\text{C7})$$

which depends only on the F_2^n/F_2^p ratio. (Note that the CJ15 PDF fit from which the CJ15nlo_mod is derived assumes isoscalar, multiplicative higher-twist corrections. For determination of f_A^{iso} one can therefore use the leading-twist structure functions that were perturbatively calculated at NLO using the CJ15nlo_mod PDF sets. This is important when dealing with charged current DIS data, which were not included in the CJ15 analysis.)

In Fig. 15, the CJ15_mod F_2^n/F_2^p ratio is compared to several widely used empirical fits. F2ALLM [160] fitted F_2 data prior to 1997 above the resonance region and is provided with and without Q^2 dependence. Arrington 2011 [3] used a similar deuteron smearing and off-shell corrections as CJ15 and results in a similar F_2^n/F_2^p ratio. Segarra [161] assumed a universal modification of nucleons in short-range correlated (mostly neutron-proton) pairs and included recent light nuclear target data, as well as data from heavy nuclei, in their fits. In general, the F_2^n/F_2^p ratio from all these models are in good agreement at small x, but show large systematic deviations at large x stemming from the adopted neutron extraction procedure. Among the considered fits and parametrizations, CJ15nlo_mod is the only Q^2-dependent model that also provides a rigorous uncertainty estimation. The isoscalar correction factor for ^{56}Fe in Fig. 15 is calculated using Eq. (C7) with different F_2^n/F_2^p ratio models and reveals a model dependence of the order of 10% at large values of x.

The neutron-to-proton ratio may also be utilized in accounting for neutron excess effects in neutrino-nucleus scattering. In contrast to electron scattering, here the neutrino-neutron cross section is actually larger than the neutrino-proton cross section. The charged current neutron-to-proton ratio $R_{n/p}^{\text{CC}} = F_2^{n,\text{CC}}/F_2^{p,\text{CC}}$, calculated using CJ15
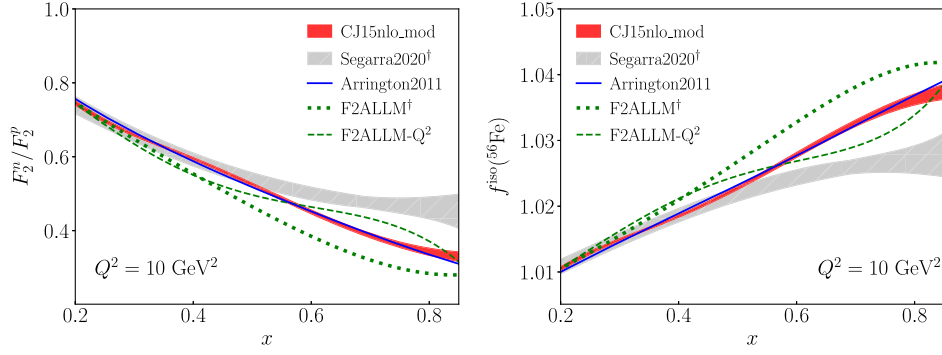
FIG. 15.   Neutron-to-proton F_2 structure function ratio (left) and isoscalar corrections for ^{56}Fe (right) at $Q^2 = 10$ GeV^2 from this work and several empirical fits including the F2ALLM model with and without Q^2 dependence [160] and more recent fits by Arrington *et al.* [3] and Segarra *et al.* [161]. Models with Q^2 dependence are labeled by †.
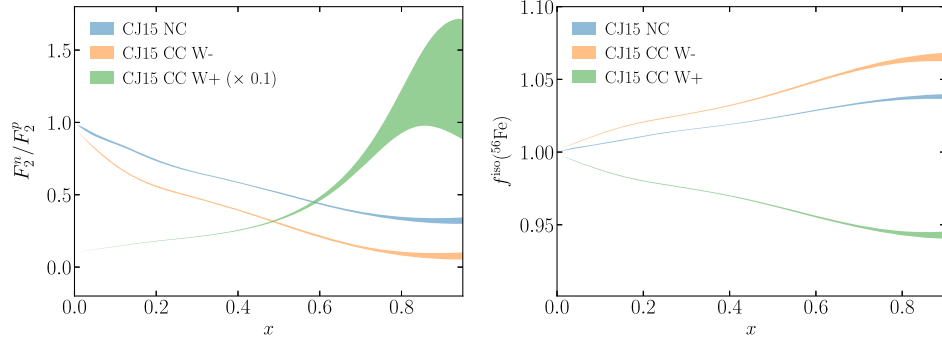

FIG. 16.   Left: neutron-to-proton F_2 ratio calculated at $Q^2 = 20$ GeV^2 using CJ15 PDFs. The positively charged CC W^+ ratio, that at LO is equal to the reciprocal of the negatively charged current (CC) W^- ratio and grows as $x \to 1$, is compared to the neutral current (NC) ratio and the CC W^- ratio on a linear scale. Right: the isoscalar correction factor for ^{56}Fe calculated from corresponding neutron-to-proton ratios.

PDFs, is shown in Fig. 16 for W^- exchange ($e^- p \to \nu X$ and $\bar{\nu} p \to e^+ X$) and W^+ exchange ($e^+ p \to \bar{\nu} X$ and $\nu p \to e^- X$). At leading order, those two ratios are related using isospin symmetry by

$$R_{n/p}^{W^+} \approx \frac{1}{R_{n/p}^{W^-}}. \tag{C8}$$

In the $x \to 1$ limit, one finds

$$R_{n/p}^{W^-} \to \frac{d}{u}, \tag{C9}$$

which can be compared with the corresponding neutral current ratio, $R_{n/p}^{\mathrm{NC}} \to \frac{1}{4}(1 + \frac{15}{4} d/u)$. The isoscalar

corrections calculated with F_2^{CC} are of the same order as those from the NC case, as shown on the right panel of Fig. 16.

At large x the W^- ratio $R_{n/p}^{W^-}$ is essentially a shifted down version of the NC ratio and decreases toward zero following the behavior of the fitted d/u ratio. Conversely, the ratio $R_{n/p}^{W^+}$ for W^+ exchange grows rapidly as $x \to 1$; this growth is tamed if the d/u ratio tends to a finite limit, as in the CJ15 fit. The $R_{n/p}^{W^+}$ structure function ratio therefore seems to be a particularly sensitive probe of the large-x behavior of the d and u-quark PDFs and of the nucleon's nonperturbative structure. It could be measured in $e^+ + p(d) \to \bar{\nu} + X$ reactions with a positron beam at JLab12 [162] and at the Electron-Ion Collider [61] or in $\nu + p(d) \to e^- + X$ processes at high-energy neutrino facilities [163].

[1] A. Bodek and J. L. Ritchie, Phys. Rev. D **23**, 1070 (1981).
[2] J. Arrington, F. Coester, R. J. Holt, and T.-S. H. Lee, J. Phys. G **36**, 025005 (2009).
[3] J. Arrington, J. G. Rubin, and W. Melnitchouk, Phys. Rev. Lett. **108**, 252001 (2012).
[4] A. Accardi, M. E. Christy, C. E. Keppel, W. Melnitchouk, P. Monaghan, J. G. Morfín, and J. F. Owens, Phys. Rev. D **81**, 034016 (2010).
[5] A. Accardi, W. Melnitchouk, J. F. Owens, M. E. Christy, C. E. Keppel, L. Zhu, and J. G. Morfin, Phys. Rev. D **84**, 014008 (2011).
[6] J. F. Owens, A. Accardi, and W. Melnitchouk, Phys. Rev. D **87**, 094012 (2013).
[7] A. Accardi, L. T. Brady, W. Melnitchouk, J. F. Owens, and N. Sato, Phys. Rev. D **93**, 114017 (2016).
[8] T. Aaltonen *et al.*, Phys. Rev. Lett. **102**, 181801 (2009).
[9] V. M. Abazov *et al.*, Phys. Rev. Lett. **112**, 151803 (2014); **114**, 049901(E) (2015).
[10] S. Tkachenko *et al.*, Phys. Rev. C **89**, 045206 (2014); **90**, 059901(A) (2014).
[11] A. Accardi, Proc. Sci. DIS2015 (**2015**) 001.
[12] S. I. Alekhin, S. A. Kulagin, and R. Petti, Phys. Rev. D **96**, 054005 (2017).
[13] S. I. Alekhin, S. A. Kulagin, and R. Petti, Phys. Rev. D **105**, 114037 (2022).
[14] C. Cocuzza, C. E. Keppel, H. Liu, W. Melnitchouk, A. Metz, N. Sato, and A. W. Thomas, Phys. Rev. Lett. **127**, 242001 (2021).
[15] W. Melnitchouk and A. W. Thomas, Phys. Lett. B **377**, 11 (1996).
[16] N. Isgur, Phys. Rev. D **59**, 034013 (1999).
[17] S. J. Brodsky, M. Burkardt, and I. Schmidt, Nucl. Phys. **B441**, 197 (1995).
[18] R. J. Holt and C. D. Roberts, Rev. Mod. Phys. **82**, 2991 (2010).
[19] C. D. Roberts, R. J. Holt, and S. M. Schmidt, Phys. Lett. B **727**, 249 (2013).
[20] A. Courtoy and P. Nadolsky, Phys. Rev. D **103**, 054029 (2021).
[21] R. D. Ball, E. R. Nocera, and J. Rojo, Eur. Phys. J. C **76**, 383 (2016).
[22] E. Hammou, Z. Kassabov, M. Madigan, M. L. Mangano, L. Mantani, J. Moore, M. M. Alvarado, and M. Ubiali, J. High Energy Phys. 11 (2023) 090.
[23] L. T. Brady, A. Accardi, W. Melnitchouk, and J. F. Owens, J. High Energy Phys. 12 (2012) 19.
[24] A. Accardi *et al.*, Eur. Phys. J. C **76**, 471 (2016).
[25] X. Cid Vidal *et al.*, CERN Yellow Rep. Monogr. **7**, 585 (2019).
[26] R. D. Ball, A. Candido, S. Forte, F. Hekhorn, E. R. Nocera, J. Rojo, and C. Schwan, Eur. Phys. J. C **82**, 1160 (2022).
[27] A. Accardi, T. J. Hobbs, X. Jing, and P. M. Nadolsky, Eur. Phys. J. C **81**, 603 (2021).
[28] M. Sajjad Athar and J. G. Morfín, J. Phys. G **48**, 034001 (2021).
[29] E. Catano-Mur, arXiv:2206.03542.
[30] A. Abed Abud *et al.*, arXiv:2203.06100.
[31] D. Ruterbories *et al.*, Phys. Rev. Lett. **129**, 021803 (2022).
[32] H.-W. Lin *et al.*, Prog. Part. Nucl. Phys. **100**, 107 (2018).

[33] M. Constantinou *et al.*, Prog. Part. Nucl. Phys. **121**, 103908 (2021).
[34] K. U. Can *et al.*, Phys. Rev. D **102**, 114505 (2020).
[35] V. Guzey *et al.*, Phys. Rev. C **86**, 045201 (2012).
[36] I. Albayrak *et al.*, Phys. Rev. Lett. **123**, 022501 (2019).
[37] A. Kotlorz, D. Kotlorz, and O. V. Teryaev, Int. J. Mod. Phys. A **36**, 1 (2021).
[38] K. Gottfried, Phys. Rev. Lett. **18**, 1174 (1967).
[39] A. L. Kataev, arXiv:hep-ph/0311091.
[40] J. Dove *et al.*, Nature (London) **590**, 561 (2021); **604**, E26 (2022).
[41] M. Arneodo *et al.*, Phys. Rev. D **50**, R1 (1994).
[42] R. Abbate and S. Forte, Phys. Rev. D **72**, 117503 (2005).
[43] A. Bodek *et al.*, Phys. Rev. D **20**, 1471 (1979).
[44] G. B. West, Phys. Lett. **37B**, 509 (1971).
[45] W. B. Atwood and G. B. West, Phys. Rev. D **7**, 773 (1973).
[46] W. Melnitchouk, A. W. Schreiber, and A. W. Thomas, Phys. Lett. B **335**, 11 (1994).
[47] W. Melnitchouk, A. W. Schreiber, and A. W. Thomas, Phys. Rev. D **49**, 1183 (1994).
[48] A. Y. Umnikov, F. C. Khanna, and L. P. Kaptar, Z. Phys. A **348**, 211 (1994).
[49] L. L. Frankfurt and M. I. Strikman, Phys. Rep. **160**, 235 (1988).
[50] L. Whitlow, E. Riordan, S. Dasu, S. Rock, and A. Bodek, Phys. Lett. B **282**, 475 (1992).
[51] U.-K. Yang and A. Bodek, Phys. Rev. Lett. **82**, 2467 (1999).
[52] W. Melnitchouk, I. R. Afnan, F. R. P. Bissey, and A. W. Thomas, Phys. Rev. Lett. **84**, 5455 (2000).
[53] A. D. Martin, A. J. T. M. Mathijssen, W. J. Stirling, R. S. Thorne, B. J. A. Watt, and G. Watt, Eur. Phys. J. C **73**, 2318 (2013).
[54] R. D. Ball, E. R. Nocera, and R. L. Pearson, Eur. Phys. J. C **81**, 37 (2021).
[55] BONuS12 Collaboration, The structure of the free neutron at large x-Bjorken, Jefferson Lab Experiment Proposal E12-06-113 (2006).
[56] N. Baillie *et al.*, Phys. Rev. Lett. **108**, 142001 (2012); **108**, 199902(E) (2012).
[57] D. Abrams *et al.*, Phys. Rev. Lett. **128**, 132003 (2022).
[58] S. I. Alekhin, S. A. Kulagin, and R. Petti, Phys. Rev. D **107**, L051506 (2023).
[59] S. Malace *et al.*, Precision measurements of the F2 structure function at large x in the resonance region and beyond, Jefferson Lab Experiment Proposal E12-10-002 (2010).
[60] The SoLID Collaboration, SoLID Preliminary Conceptual Design Report (2014) (to be published).
[61] R. Abdul Khalek *et al.*, Nucl. Phys. **A1026**, 122447 (2022).
[62] J. Adam *et al.*, Phys. Rev. D **103**, 012001 (2021).
[63] S. Chatrchyan *et al.*, Phys. Rev. D **90**, 032004 (2014).
[64] S. Chatrchyan *et al.*, Phys. Rev. Lett. **109**, 111806 (2012).
[65] G. Aad *et al.*, Phys. Rev. D **85**, 072004 (2012).
[66] R. Aaij *et al.*, J. High Energy Phys. 08 (2015) 039.
[67] V. Khachatryan *et al.*, Eur. Phys. J. C **76**, 469 (2016).
[68] P. Jimenez-Delgado, W. Melnitchouk, and J. F. Owens, J. Phys. G **40**, 093102 (2013).
[69] J. Gao, L. Harland-Lang, and J. Rojo, Phys. Rep. **742**, 1 (2018).

[70] K. Kovařík, P. M. Nadolsky, and D. E. Soper, Rev. Mod. Phys. **92**, 045003 (2020).

[71] J. J. Ethier and E. R. Nocera, Annu. Rev. Nucl. Part. Sci. **70**, 43 (2020).

[72] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, Eur. Phys. J. C **75**, 204 (2015).

[73] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, Eur. Phys. J. C **81**, 341 (2021).

[74] CTEQ-JLab Collaboration, CJ Collaboration homepage, https://www.jlab.org/theory/cj (accessed: April 2024).

[75] D. Acosta et al., Phys. Rev. D **71**, 051104 (2005).

[76] V. M. Abazov et al., Phys. Rev. D **88**, 091102 (2013).

[77] V. M. Abazov et al., Phys. Rev. D **91**, 032007 (2015); **91**, 079901(E) (2015).

[78] S. A. Kulagin, G. Piller, and W. Weise, Phys. Rev. C **50**, 1154 (1994).

[79] S. A. Kulagin, W. Melnitchouk, G. Piller, and W. Weise, Phys. Rev. C **52**, 932 (1995).

[80] K. A. Griffioen et al., Phys. Rev. C **92**, 015211 (2015).

[81] W. Melnitchouk, R. Ent, and C. Keppel, Phys. Rep. **406**, 127 (2005).

[82] S. Li, A. Accardi, and I. Fernando, in Exploring QCD with Tagged Processes, Paris, 2021 (2021), https://indico.ijclab .in2p3.fr/event/7455/.

[83] T. J. Hobbs, B.-T. Wang, P. M. Nadolsky, and F. I. Olness, Phys. Rev. D **100**, 094040 (2019).

[84] M. Arneodo et al., Phys. Lett. B **364**, 107 (1995).

[85] L. W. Whitlow, Ph.D. thesis, Stanford University, 1990.

[86] L. Tao, Ph.D. thesis, The American University, 1994.

[87] M. Arneodo et al., Nucl. Phys. **B483**, 3 (1997).

[88] A. C. Benvenuti et al., Phys. Lett. B **223**, 485 (1989).

[89] A. Benvenuti et al., Phys. Lett. B **237**, 592 (1990).

[90] A. P. Balachandran and A. R. de Queiroz, J. High Energy Phys. 11 (2011) 126.

[91] V. Tvaskis, Ph.D. thesis, Amsterdam University, 2004.

[92] J. Seely et al., Phys. Rev. Lett. **103**, 202301 (2009).

[93] A. Daniel, Ph.D. thesis, University of Houston, 2007.

[94] S. Malace, Ph.D. thesis, Hampton University, 2006.

[95] M. Arneodo et al., Nucl. Phys. **B487**, 3 (1997).

[96] J. Pumplin, D. R. Stump, J. Huston, H.-L. Lai, P. Nadolsky, and W.-K. Tung, J. High Energy Phys. 07 (2002) 12.

[97] P. Amaudruz et al., Phys. Rev. Lett. **66**, 2712 (1991).

[98] M. Arneodo et al., Phys. Rev. D **50**, R1 (1994).

[99] S. Kumano, Phys. Rep. **303**, 183 (1998).

[100] D. F. Geesaman and P. E. Reimer, Rep. Prog. Phys. **82**, 046301 (2019).

[101] A. Baldit et al., Phys. Lett. B **332**, 244 (1994).

[102] K. Ackerstaff et al., Phys. Rev. Lett. **81**, 5519 (1998).

[103] E. A. Hawker et al., Phys. Rev. Lett. **80**, 3715 (1998).

[104] R. S. Towell et al., Phys. Rev. D **64**, 052002 (2001).

[105] D. A. Ross and C. T. Sachrajda, Nucl. Phys. **B149**, 497 (1979).

[106] A. W. Thomas, Phys. Lett. **126B**, 97 (1983).

[107] Y. Salamu, C.-R. Ji, W. Melnitchouk, and P. Wang, Phys. Rev. Lett. **114**, 122001 (2015).

[108] A. Accardi, C. E. Keppel, S. Li, W. Melnitchouk, G. Niculescu, I. Niculescu, and J. F. Owens, Phys. Lett. B **801**, 135143 (2020).

[109] B. M. Badelek and J. Kwiecinski, Nucl. Phys. **B370**, 278 (1992).

[110] V. R. Zoller, Phys. Lett. B **279**, 145 (1992).

[111] W. Melnitchouk and A. W. Thomas, Phys. Rev. D **47**, 3783 (1993).

[112] G. Piller, W. Ratzka, and W. Weise, Z. Phys. A **352**, 427 (1995).

[113] M. E. Christy and P. E. Bosted, Phys. Rev. C **81** (2010).

[114] P. E. Bosted and M. E. Christy, Phys. Rev. C **77** (2008).

[115] I. Schienbein et al., J. Phys. G **35**, 053101 (2008).

[116] A. Accardi and J.-W. Qiu, J. High Energy Phys. 07 (2008) 090.

[117] F. M. Steffens, M. D. Brown, W. Melnitchouk, and S. Sanches, Phys. Rev. C **86**, 065208 (2012).

[118] J. Blumlein, Prog. Part. Nucl. Phys. **69**, 28 (2013).

[119] T.-J. Hou et al., Phys. Rev. D **103**, 014013 (2021).

[120] J. Butterworth et al., J. Phys. G **43**, 023001 (2016).

[121] V. Bertone, S. Carrazza, and J. Rojo (APFEL Collaboration), Comput. Phys. Commun. **185**, 1647 (2014).

[122] V. Bertone, Proc. Sci. DIS2017 (**2018**) 201.

[123] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, Eur. Phys. J. C **75**, 132 (2015).

[124] J. R. Andersen et al., arXiv:1405.1067.

[125] H. L. Lai, J. Huston, S. Kuhlmann, J. Morfin, F. I. Olness, J. F. Owens, J. Pumplin, and W. K. Tung, Eur. Phys. J. C **12**, 375 (2000).

[126] Y. Salamu, C.-R. Ji, W. Melnitchouk, A. W. Thomas, and P. Wang, Phys. Rev. D **99**, 014041 (2019).

[127] O. Nachtmann, Nucl. Phys. **B63**, 237 (1973).

[128] O. W. Greenberg and D. Bhaumik, Phys. Rev. D **4**, 2048 (1971).

[129] T. Weigl and W. Melnitchouk, Nucl. Phys. **B465**, 267 (1996).

[130] G. Bali, S. Collins, M. Göckeler, R. Rödl, A. Schäfer, and A. Sternbeck, Phys. Rev. D **100**, 014507 (2019).

[131] Y.-B. Yang, J. Liang, Y.-J. Bi, Y. Chen, T. Draper, K.-F. Liu, and Z. Liu, Phys. Rev. Lett. **121**, 212001 (2018).

[132] T. Harris, G. von Hippel, P. Junnarkar, H. B. Meyer, K. Ottnad, J. Wilhelm, H. Wittig, and L. Wrang, Phys. Rev. D **100**, 034513 (2019).

[133] C. Alexandrou et al., Phys. Rev. D **101**, 034519 (2020).

[134] C. Alexandrou, S. Bacchio, M. Constantinou, J. Finkenrath, K. Hadjiyiannakou, K. Jansen, G. Koutsou, H. Panagopoulos, and G. Spanoudes, Phys. Rev. D **101**, 094513 (2020).

[135] S. Mondal, R. Gupta, S. Park, B. Yoon, T. Bhattacharya, and H.-W. Lin, Phys. Rev. D **102**, 054512 (2020).

[136] S. I. Alekhin, J. Blümlein, S. Moch, and R. Placakyte, Phys. Rev. D **96**, 014011 (2017).

[137] H. Abramowicz et al., Eur. Phys. J. C **75**, 580 (2015).

[138] N. Sato, C. Andres, J. J. Ethier, and W. Melnitchouk, Phys. Rev. D **101**, 074020 (2020).

[139] R. D. Ball et al., Eur. Phys. J. C **77**, 663 (2017).

[140] CTEQ-JLab Collaboration, CJ database, https://github .com/JeffersonLab/CJ-database (accessed: April 2024).

[141] S. Dasu et al., Phys. Rev. D **49**, 5641 (1994).

[142] I. H. Albayrak, Ph.D. thesis, Hampton University, 2011.

[143] Y. Liang et al., Phys. Rev. C **105**, 065205 (2022).

[144] M. E. Christy et al., Phys. Rev. C **70**, 015206 (2004).

[145] V. Tvaskis et al., Phys. Rev. C **81**, 055207 (2010).

[146] V. Tvaskis *et al.*, Phys. Rev. Lett. **98**, 142301 (2007).
[147] M. I. Niculescu, Ph.D. thesis, Hampton University, 1999.
[148] S. P. Malace *et al.*, Phys. Rev. C **80**, 035207 (2009).
[149] M. Osipenko *et al.*, Phys. Rev. D **67**, 092001 (2003).
[150] M. Osipenko *et al.*, Phys. Rev. C **73**, 045205 (2006).
[151] M. Osipenko *et al.*, arXiv:hep-ex/0309052.
[152] M. Osipenko *et al.*, arXiv:hep-ex/0507098.
[153] A. V. Kotwal, Ph.D. thesis, Harvard University, 1995.
[154] N. T. Hunt-Smith, A. Accardi, W. Melnitchouk, N. Sato, A. W. Thomas, and M. J. White, Phys. Rev. D **106**, 036003 (2022).
[155] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, Eur. Phys. J. C **75**, 132 (2015).
[156] https://lhapdf.hepforge.org/ (accessed: September 2023).
[157] R. Abdul Khalek *et al.*, Nucl. Phys. **A1026**, 122447 (2022).
[158] https://jeffersonlab.github.io/txgrids (accessed: September 2023).
[159] V. Guzey, L. Zhu, C. E. Keppel, M. E. Christy, D. Gaskell, P. Solvignon, and A. Accardi, Phys. Rev. C **86**, 045201 (2012).
[160] H. Abramowicz and A. Levy, arXiv:hep-ph/9712415.
[161] E. P. Segarra, A. Schmidt, T. Kutz, D. W. Higinbotham, E. Piasetzky, M. Strikman, L. B. Weinstein, and O. Hen, Phys. Rev. Lett. **124**, 092002 (2020).
[162] A. Accardi *et al.*, Eur. Phys. J. A **57**, 261 (2021).
[163] R. Petti, Phys. Lett. B **834**, 137469 (2022).

Correction: Errors in the denominator in Eqs. (8) and (9) have been fixed. The inline equation in the third paragraph in Sec. IV A contained an error and has been fixed.