

Bayesian Monte Carlo extraction of sea asymmetry with SeaQuest and STAR data

C. Cocuzza,¹ W. Melnitchouk,² A. Metz,¹ and N. Sato²

¹*Department of Physics, SERC, Temple University, Philadelphia, Pennsylvania 19122, USA*

²*Jefferson Lab, Newport News, Virginia 23606, USA*

Jefferson Lab Angular Momentum (JAM) Collaboration

(Dated: September 1, 2021)

We perform a global QCD analysis of unpolarized parton distributions within a Bayesian Monte Carlo framework, including for the first time W -lepton production data from the STAR Collaboration at RHIC and Drell-Yan di-muon data from the SeaQuest experiment at Fermilab. We assess the impact of these two new measurements on the light antiquark sea in the proton, and the $\bar{d} - \bar{u}$ asymmetry in particular. The SeaQuest data are found to significantly reduce the uncertainty on the $\bar{d}/\bar{u}$ ratio at large parton momentum fractions x , strongly favoring an enhanced $\bar{d}$ sea up to $x \approx 0.4$, in general agreement with nonperturbative calculations based on chiral symmetry breaking in QCD.

I. INTRODUCTION

Over the past five decades, there has been great interest in the structure of the nucleon sea, which encompasses quarks, antiquarks, and gluons that exist beyond the three valence quarks originally proposed as the building blocks of the nucleon [1–3]. In particular, high-energy scattering experiments and global QCD analyses of the data have now conclusively demonstrated a large difference in the momentum distributions of $\bar{u}$ and $\bar{d}$ antiquarks in the proton [4]. The result cannot be explained perturbatively through the splitting of gluons into quark-antiquark pairs [5], and requires nonperturbative mechanisms, such as dynamical chiral symmetry breaking and the pion cloud of the nucleon [6–16], or dynamics related to the Fermi-Dirac statistics of quarks and the Pauli exclusion principle [17–21]. More recently, exploratory studies have been made in extracting information on the isovector sea quark distributions directly from lattice QCD calculations [22–26].

The first experimental indications of a light-quark sea asymmetry came from the CFS group at Fermilab in 1981 [27]. Measuring the Drell-Yan process, where a quark and antiquark from colliding hadrons annihilate into a virtual photon that subsequently decays into a lepton-antilepton pair, they found that the $\bar{d}$ distribution, integrated over parton momentum fraction x , was larger than the integrated $\bar{u}$ distribution. The first high-precision experimental evidence for a sea asymmetry came a decade later from the New Muon Collaboration (NMC) at CERN [28, 29], which used measurements of inclusive deep-inelastic scattering (DIS) on hydrogen and deuterium to test the Gottfried sum rule [30] and determine that the integral $\int_0^1 dx [\bar{d}(x) - \bar{u}(x)]$ must be positive. Further evidence was provided by the NA51 Collaboration at CERN [31] using the Drell-Yan process, and indications for a nonzero asymmetry were also found by the HERMES Collaboration [32] in semi-inclusive DIS of charged pions.

The most conclusive evidence for an excess of $\bar{d}$ over

$\bar{u}$ was provided in 1998 by the Fermilab E866 (NuSea) experiment [33–35], which measured Drell-Yan lepton-pair production cross sections in proton-proton (pp) and proton-deuteron (pD) scattering. At kinematics where the parton momentum fraction of the beam, x_1 , is much greater than that of the target, x_2 , one finds that the ratio of pD to pp cross section, at leading order in the strong coupling, provides direct sensitivity to the $\bar{d}/\bar{u}$ ratio [36],

$$\frac{\sigma_{pD}^{\text{DY}}}{2\sigma_{pp}^{\text{DY}}} \Big|_{x_1 \gg x_2} \approx \frac{1}{2} \left[1 + \frac{\bar{d}(x_2)}{\bar{u}(x_2)} \right]. \quad (1)$$

While the NuSea experiment probed the asymmetry up to momentum fractions $x \approx 0.3$, the subsequent Fermilab E906 (SeaQuest) experiment [37] extended the range up to $x \approx 0.4$, finding some tensions with the NuSea result in the high- x region. The NuSea data had suggested a significant fall in the $\bar{d}/\bar{u}$ ratio for $x \gtrsim 0.3$, albeit with large uncertainties, which is difficult to accommodate in many of the nonperturbative models [6–21]. The SeaQuest experiment was partially motivated to verify this behavior at large x .

An alternative method for extracting the $\bar{d} - \bar{u}$ asymmetry involves W -lepton production in hadronic collisions, whereby a quark and antiquark annihilate into a W boson that decays into a detected lepton and a neutrino. This has been measured in $p\bar{p}$ scattering at the Tevatron and in pp collisions at the Large Hadron Collider (LHC) and the Relativistic Heavy Ion Collider (RHIC). Taking the ratio of the W^+ and W^- cross sections for pp reactions, at leading order in perturbative QCD one has

$$\frac{\sigma_{pp}^{W^+}}{\sigma_{pp}^{W^-}} \approx \frac{u(x_1)\bar{d}(x_2) + u(x_2)\bar{d}(x_1)}{d(x_1)\bar{u}(x_2) + d(x_2)\bar{u}(x_1)}. \quad (2)$$

The most recent high-precision data from the STAR Collaboration at RHIC [38] on the ratios of W -lepton production cross sections have not yet been included in global QCD analyses. Since RHIC has a lower center-of-mass energy of $\sqrt{S} = 0.51$ TeV compared to the Tevatron ($\sqrt{S} = 1.96$ TeV) and the LHC ($\sqrt{S} \geq 7$ TeV),

these data are sensitive to parton distributions at higher values of x , which can potentially provide information on the distributions of light sea quarks in this difficult to measure region.

In this paper, we present the results of a global QCD analysis using the JAM Bayesian Monte Carlo framework, including the new measurements from SeaQuest [37] and STAR [38]. By combining these data with other hadron collider data from Fermilab and the LHC, and with DIS data from fixed-target and collider experiments, we are able to place strong constraints on the light-quark PDFs and assess the impact of the new data on the antiquark asymmetry $\bar{d} - \bar{u}$.

In Sec. II we begin by reviewing the theoretical framework used in this analysis, including a summary of the most relevant factorization formulas. In Sec. III we review the methodology employed for the analysis, as well as the choice of parametrization for the PDFs. The quality of the fit is described in Sec. IV, where we compare the fitted cross sections with the new Drell-Yan and W -lepton data. The results for the PDFs are presented in Sec. V, where we in particular assess the impact of the various datasets on the shape of the $\bar{d} - \bar{u}$ asymmetry and its uncertainties. We also compare our extracted asymmetry with nonperturbative calculations based on chiral symmetry breaking in QCD. Finally, in Sec. VI we summarize our findings and discuss their implications for our understanding of the sea content of the nucleon.

II. THEORETICAL FRAMEWORK

Our theoretical framework is based on fixed order collinear factorization for various high-energy scattering processes involving spin-averaged PDFs, such as inclusive DIS, Drell-Yan lepton-pair production, and weak boson and jet production. Since the new datasets that provide sensitivity to the antiquark distributions in this analysis involve Drell-Yan and weak boson production, we will focus on these observables in this section.

The double differential Drell-Yan lepton-pair production cross section can be written in terms of convolutions of the PDFs in the colliding hadrons with short-distance partonic cross sections $\hat{\sigma}_{ab}^{\text{DY}}$ [39],

$$\frac{d^2\sigma^{\text{DY}}}{dM_{\ell\ell}^2 dY} = \frac{4\pi\alpha^2}{9SM_{\ell\ell}^2} \sum_{ab} \int dx_1 \int dx_2 \quad (3)$$

$$\times f_a(x_1, \mu_F) f_b(x_2, \mu_F) \hat{\sigma}_{ab}^{\text{DY}}(x_1, x_2, S, M_{\ell\ell}, \mu_R, \mu_F),$$

where α is the electromagnetic coupling, S is the invariant mass squared of the reaction, and μ_R and μ_F are the renormalization and factorization scales, respectively. We write the cross section as differential in the invariant mass of the lepton pair, $M_{\ell\ell}^2$, and the rapidity of the lepton pair in the center-of-mass frame, Y . The sum over the quark flavors a, b runs over all partonic channels that can contribute to the Drell-Yan process, for which the

scale is set to $\mu_R = \mu_F = M_{\ell\ell}$. The partonic cross sections $\hat{\sigma}_{ab}^{\text{DY}}$ are computed at next-to-leading order (NLO) in the strong coupling $\alpha_s(\mu_R)$, with the NLO expressions taken from Ref. [39]. For pp scattering, the PDFs $f_{a,b}$ are those in the proton. For proton-deuteron scattering, the x of the parton in the deuteron target is small enough that it is reasonable to approximate the pD cross section by a simple sum of proton and neutron cross sections [40], $\sigma_{pD} \approx \sigma_{pp} + \sigma_{pn}$, with the PDFs in the neutron related to those in the proton through isospin symmetry.

For W -lepton production, the double differential cross section is given by [41]

$$\frac{d^2\sigma^W}{dp_T^\ell d\eta_\ell} = \frac{2}{p_T^\ell} \sum_{ab} \int_{VW}^V dv \int_{VW/v}^1 dw \quad (4)$$

$$\times x_1 f_a(x_1, \mu_F) x_2 f_b(x_2, \mu_F) \hat{\sigma}_{ab}^W(v, w, S, \mu_R, \mu_F),$$

which is differential in the outgoing lepton's pseudorapidity, η_ℓ , and its transverse momentum, p_T^ℓ . Labeling the momenta of the incoming hadrons by P_1 and P_2 , we define the Mandelstam invariants $T \equiv (P_1 - p_\ell)^2$ and $U \equiv (P_2 - p_\ell)^2$, which then allows us to introduce the variables $V \equiv 1 + T/S$ and $W \equiv -U/(S + T)$. Defining v and w as the partonic analogues of V and W , we can write for the parton momentum fractions $x_1 = VW/vw$ and $x_2 = (1 - V)/(1 - v)$. The sum over the quark flavors a, b runs over all partonic channels that can contribute to W -lepton production, for which the renormalization and factorization scales are chosen to be the mass of the W boson, $\mu_R = \mu_F = M_W$. The partonic cross sections $\hat{\sigma}_{ab}^W$ are again computed at NLO in the strong coupling $\alpha_s(\mu_R)$, with the NLO expressions used here taken from Ref. [41].

In practice, the W -lepton production cross sections are usually integrated over a specified range of p_T^ℓ (for example, $p_T^\ell > 25$ GeV for the STAR data), and the cross section is given as differential in the pseudorapidity η_ℓ . Note that the theoretical formalism that leads to Eq. (4) above applies to final states that are fully inclusive in the transverse momentum of the undetected neutrino, p_T^ν . This formalism therefore cannot be used directly to calculate the W -lepton asymmetries from the D0, CDF, and ATLAS collaborations, which place cuts on p_T^ν and would require extending the current theoretical framework to implement such phase space cuts [42].

To avoid this ambiguity, we use the fully inclusive reconstructed W -boson data from D0 and CDF instead of the W -lepton data. Our analysis does not include data from ATLAS, but includes the fully inclusive data from the CMS collaboration, which cover the same kinematic region with similar uncertainties. We can thus retain most of the information provided by these observables, despite the exclusion of the non-fully inclusive W -lepton data from CDF, D0, and ATLAS.

For the DIS theory, we include corrections that are known to be necessary to accurately describe the high- x region. These include target mass corrections, as described by Moffat *et al.* [43], parameterized higher twist

corrections, and nuclear corrections for deuterium data. Further details regarding these corrections will be discussed in a future publication [44].

The scale dependence of the PDFs is determined according to the DGLAP evolution equations [45–47], with the PDFs and α_s evolved according to their renormalization group equation (RGE) at next-to-leading logarithmic accuracy with the boundary condition $\alpha_s(M_Z) = 0.118$. For light and for heavy quarks, the PDFs are evolved using the zero-mass variable-flavor-number scheme. The values of the heavy quark mass thresholds for the evolution of the PDFs and α_s are taken from the PDG values $m_c = 1.28$ GeV and $m_b = 4.18$ GeV in the $\overline{\text{MS}}$ scheme [48].

III. BAYESIAN INFERENCE

Our PDF extraction procedure is based on Bayesian inferences using Monte Carlo techniques developed in previous JAM analyses [49–53]. We parametrize the PDFs at the input scale $\mu_0^2 = m_c^2$ using a generic template function of the form

$$f(x, \mu_0^2) = \frac{N}{\mathcal{M}} x^\alpha (1-x)^\beta (1 + \gamma\sqrt{x} + \eta x), \quad (5)$$

where $\mathbf{a} = \{N, \alpha, \beta, \gamma, \eta\}$ is the set of parameters to be inferred, and $\mathcal{M} = \text{B}[\alpha + 2, \beta + 1] + \gamma \text{B}[\alpha + \frac{5}{2}, \beta + 1] + \eta \text{B}[\alpha + 3, \beta + 1]$ normalizes the function to the second moment to maximally decorrelate the normalization and shape parameters. To characterize the nucleon valence region and discriminate it from the sea components, we parametrize the light-quark and strange PDFs according to

$$\begin{aligned} u &= u_v + \bar{u}, & d &= d_v + \bar{d}, \\ \bar{u} &= S_1 + \bar{u}_0, & \bar{d} &= S_1 + \bar{d}_0, \\ s &= S_2 + s_0, & \bar{s} &= S_2 + \bar{s}_0, \end{aligned} \quad (6)$$

where the dependence on x and the scale μ_0^2 has been suppressed for convenience. The input quark distributions $u_v, d_v, \bar{u}_0, \bar{d}_0, s_0$, and $\bar{s}_0$, as well as the gluon distribution g , are parametrized individually as in Eq. (5). For the sea quark PDFs, the additional functions S_1 and S_2 are also parametrized via Eq. (5), and are designed to allow a more singular small- x behavior compared to the valence distributions by restricting the corresponding α parameter to more negative values. Letting $S_1 \neq S_2$ allows for different small- x behaviors for the light sea quarks and the strange quarks.

The normalization parameter N for the gluon distribution is fixed by the momentum sum rule, while the corresponding normalization parameters for u_v, d_v , and s_0 are fixed by the valence number sum rules. For the $u_v, d_v, g, \bar{u}_0$, and $\bar{d}_0$ distributions, the parameters γ and η in Eq. (5) are included to allow sufficient flexibility. We have verified that including these parameters for S_1, S_2, s_0 , or $\bar{s}_0$ does not lead to significant changes to the final

results, so for these distributions the γ and η parameters are set to zero.

Our Bayesian analysis consists of sampling the posterior distribution given by

$$\mathcal{P}(\mathbf{a}|\text{data}) = \mathcal{L}(\mathbf{a}, \text{data}) \pi(\mathbf{a}), \quad (7)$$

with a likelihood function of Gaussian form,

$$\mathcal{L}(\mathbf{a}, \text{data}) = \exp\left(-\frac{1}{2}\chi^2(\mathbf{a}, \text{data})\right), \quad (8)$$

and a flat prior function $\pi(\mathbf{a})$ that vanishes in regions where the parameters $\mathbf{a}$ give unphysical PDFs. The χ^2 function in (8) is defined as

$$\begin{aligned} \chi^2(\mathbf{a}) &= \sum_{i,e} \left(\frac{d_{i,e} - \sum_k r_e^k \beta_{i,e}^k - T_{i,e}(\mathbf{a})/N_e}{\alpha_{i,e}} \right)^2 \\ &+ \sum_k (r_e^k)^2 + \sum_e \left(\frac{1 - N_e}{\delta N_e} \right)^2, \end{aligned} \quad (9)$$

where $d_{i,e}$ is the experimental data point i from dataset e , and $T_{i,e}$ is the corresponding theoretical value. All uncorrelated uncertainties are added in quadrature and labeled by $\alpha_{i,e}$, while $\beta_{i,e}^k$ represents the k -th source of point-to-point correlated systematic uncertainties for the i -th data point weighted by r_e^k . The latter are optimized per values of the parameters $\mathbf{a}$ via $\partial\chi^2/\partial r_e^k = 0$. We include normalization parameters N_e for each dataset e as part of the posterior distribution per data set, with a Gaussian penalty controlled by the experimentally quoted normalization uncertainties δN_e .

The posterior distribution is sampled via data re-sampling, whereby multiple maximum likelihood optimizations are carried out by adding Gaussian noise with width $\alpha_{i,e}$ to each data point across all data sets. The resulting ensemble of parameter samples $\{\mathbf{a}_k; k = 1, \dots, n\}$ is then used to obtain statistical estimators for a given observable (function of PDFs) $\mathcal{O}(\mathbf{a})$, such as the mean and variance,

$$\text{E}[\mathcal{O}] = \frac{1}{n} \sum_k \mathcal{O}(\mathbf{a}_k), \quad (10a)$$

$$\text{V}[\mathcal{O}] = \frac{1}{n} \sum_k [\mathcal{O}(\mathbf{a}_k) - \text{E}[\mathcal{O}]]^2. \quad (10b)$$

The agreement between data and theory is assessed by defining the “reduced” χ^2 for each dataset e as

$$\chi_{\text{red},e}^2 \equiv \frac{1}{N_{\text{dat}}^e} \sum_i \left(\frac{d_{i,e} - \text{E}[\sum_k r_e^k \beta_{i,e}^k + T_{i,e}/N_e]}{\alpha_{i,e}} \right)^2, \quad (11)$$

with N_{dat}^e the total number of data points for each experiment, and $\text{E}[\dots]$ represents the mean theory as defined in Eq. (10).

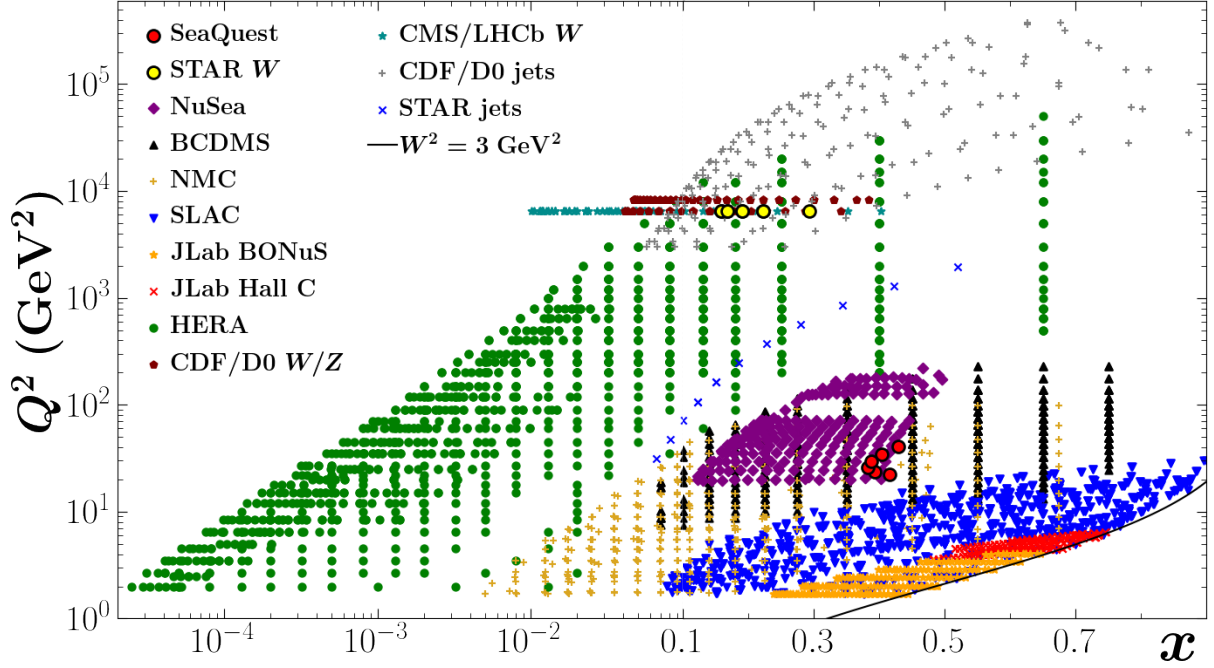


FIG. 1. Kinematic coverage of the datasets included in this analysis. The variable x represents Bjorken- x for DIS and Feynman- x for vector boson and jet production, while the scale Q^2 represents the four-momentum transfer squared for DIS, the mass squared of the intermediate boson for vector boson production, and the transverse momentum squared for jet production. Also indicated is the DIS cut of $W^2 = 3 \text{ GeV}^2$ (solid back line).

IV. QUALITY OF FIT

The datasets used in this analysis comprise inclusive DIS from hydrogen and deuterium, Drell-Yan lepton-pair production in pp and pD scattering, along with weak vector boson and jet production in pp and $p\bar{p}$ collisions. The various datasets are summarized in Table I, and their Born-level kinematics are displayed in Fig. 1. For

DIS, we include F_2 structure function data from fixed target experiments from BCDMS [54], NMC [55, 56], SLAC [57], and Jefferson Lab [58, 59], as well as the reduced neutral and charged current proton cross sections from the combined H1/ZEUS analysis from HERA [60]. The cuts on the four-momentum transfer squared Q^2 and the hadronic final state masses W are $Q^2 > m_c^2$ and $W^2 > 3 \text{ GeV}^2$ for all DIS data.

Beyond DIS, we include Drell-Yan di-muon data from pp and pD collisions from the NuSea [33, 35] and SeaQuest [37] experiments at Fermilab. For weak vector boson mediated processes, we use Z/γ^* cross sections and reconstructed W asymmetries from the CDF and D0 Collaborations at the Tevatron [61–64]; $W^\pm$ -lepton asymmetries from CMS [65–68] and LHCb [69, 70] at the LHC; and the ratio of W^+ -lepton to W^- -lepton cross sections from STAR [38] at RHIC. Finally, we include jet production data from CDF [71], D0 [72], and STAR [73].

The quality of the global fit is summarized in Table I, which shows a global average $\chi_{\text{red}}^2 = 1.12$ for a total of $N_{\text{dat}} = 4421$ data points. The Drell-Yan data on the ratio $\sigma_{pD}^{\text{DY}}/2\sigma_{pp}^{\text{DY}}$ from both the NuSea and SeaQuest experiments are shown in Fig. 2. Note that the data from the two experiments need not overlap, as they were taken at different kinematics with different values of $M_{\ell\ell}$, and for the ranges $0.36 < x_1 < 0.56$ for NuSea [35] and $0.48 < x_1 < 0.69$ for SeaQuest [37]. As suggested in Ref. [37] and seen in Fig. 2 (in combination with Eq. (1)), there is some tension between the new SeaQuest data

TABLE I. Summary of the χ_{red}^2 values and number of data points for individual datasets used in this analysis.

process	ref.	N_{dat}	χ_{red}^2
DIS			
fixed target	[54–59]	2678	1.05
HERA	[60]	1185	1.27
Drell-Yan			
NuSea pp	[33]	184	1.21
NuSea $pD/2pp$	[35]	15	1.30
SeaQuest $pD/2pp$	[37]	6	0.82
W -lepton charge asymmetry			
STAR W^+/W^-	[38]	9	2.02
CMS	[65–68]	45	0.74
LHCb	[69, 70]	16	0.44
W rapidity	[63, 64]	27	1.18
Z rapidity	[61, 62]	56	0.97
inclusive jets	[71–73]	200	1.11
total		4421	1.12

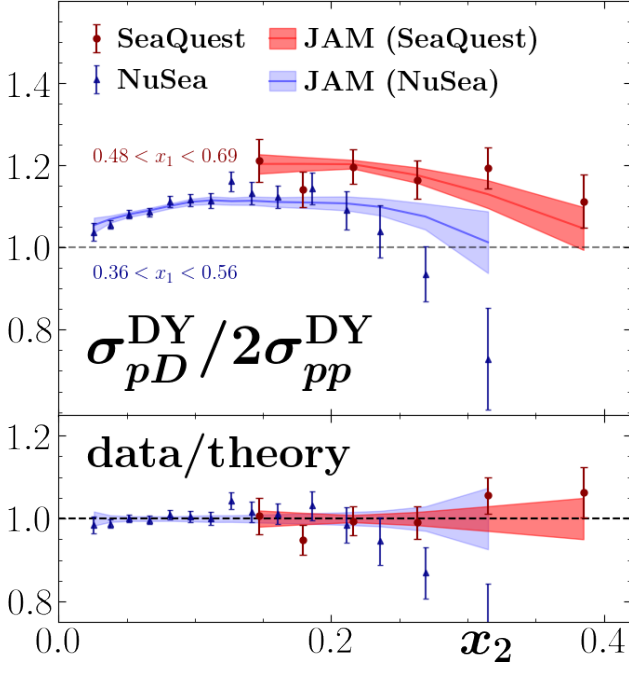


FIG. 2. Drell-Yan cross section ratio $\sigma_{pD}^{\text{DY}}/2\sigma_{pp}^{\text{DY}}$ from SeaQuest [37] (red circles) and NuSea [35] (blue triangles) compared with the JAM results at their respective kinematics (red and blue 1σ uncertainty bands), as a function of the target momentum fraction x_2 , with the corresponding x_1 ranges indicated. The ratio of data to the average theory is illustrated in the lower panel with 1σ theoretical uncertainties at the SeaQuest and NuSea kinematics.

and the earlier NuSea data, particularly above $x = 0.25$, where NuSea suggests $\bar{d}/\bar{u} < 1$ while SeaQuest indicates the opposite. The amount of tension can be illustrated by removing the SeaQuest data, which then leads to an improvement in the χ_{red}^2 for NuSea from 1.30 to 0.57. While the tension exists, we are nonetheless able to attain an acceptable description (within 2σ theoretical uncertainty) of both datasets, as seen in Table I and Fig. 2.

For the new W -lepton data from STAR, shown in Fig. 3, the description suffers slightly at high and at low pseudorapidities, leading to a χ_{red}^2 of 2.02 for these data. From Ref. [38], it is known that this is a common feature of most PDF extractions. This discrepancy is also partially due to some tension with the NuSea data, and the χ_{red}^2 improves to 1.54 upon its removal.

V. EXTRACTED PARTON DENSITIES

Our analysis is based on more than 900 Monte Carlo samples, which we use to ensure the statistical convergence of the PDFs, from which the means and expectation values are then computed using Eqs. (10). The resulting parton densities are displayed in Fig. 4, where we show the valence (u_v and d_v), gluon (g), light anti-

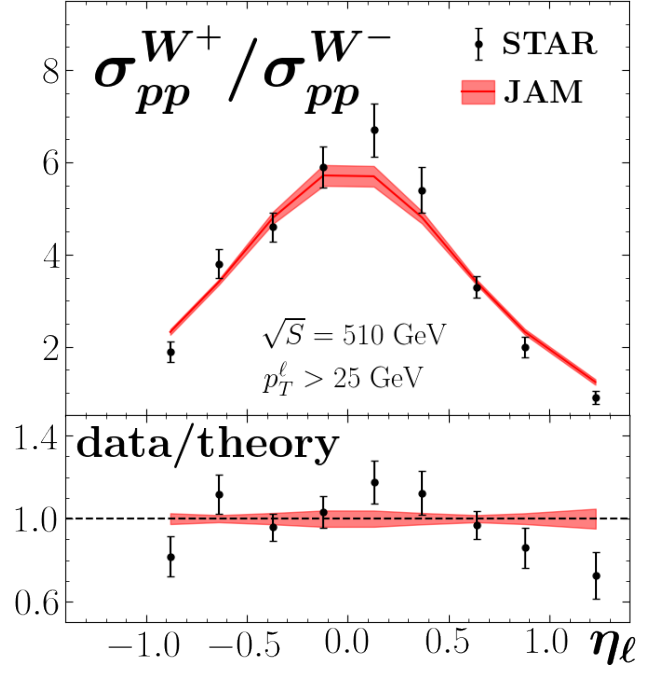


FIG. 3. W -lepton cross section ratio $\sigma_{pp}^{W^+}/\sigma_{pp}^{W^-}$ from STAR [38] (black circles) compared with the JAM fit (red 1σ uncertainty band), as a function of the lepton pseudorapidity, η_ℓ . The ratio of data to average theory is shown in the lower panel together with the 1σ theoretical uncertainty.

quark ($\bar{d} + \bar{u}$), and strange ($s + \bar{s}$) distributions at a scale of $Q^2 = 10 \text{ GeV}^2$.

In the valence sector, our results for both u_v and d_v are slightly larger than those from the other groups near the peaks of the distributions [74–77]. At lower x , the u_v distribution agrees with the other groups, while the d_v PDF agrees best with the CT18 [77] and NNPDF3.1 [74] parametrizations. For the gluon PDF, our results are largely in agreement with other extractions, although at low x there are some differences with the ABMP16 [75] fit. The same is true for the sum of light sea quark distributions, $\bar{d} + \bar{u}$, but the disagreements are with CJ15 [76].

For the strange distribution $s + \bar{s}$, our result is somewhat suppressed at low x relative to the other extractions. This is consistent with previous JAM analyses [52, 53] that included semi-inclusive DIS cross section data, although these data are not included in the present analysis.

The impact on the light-quark sea asymmetry from both the new STAR measurement of W -lepton cross sections and the SeaQuest measurement of Drell-Yan dimuon production is shown in Fig. 5. From the baseline analysis, which excludes these new datasets, the STAR cross section ratios are added first in order to assess their impact. While the STAR data do not lead to significant shifts in the central values of $\bar{d}/\bar{u}$, they do reduce the uncertainties somewhat, by up to $\approx 20\%$ at $x \approx 0.2$. This modest impact can be understood from the fact that the

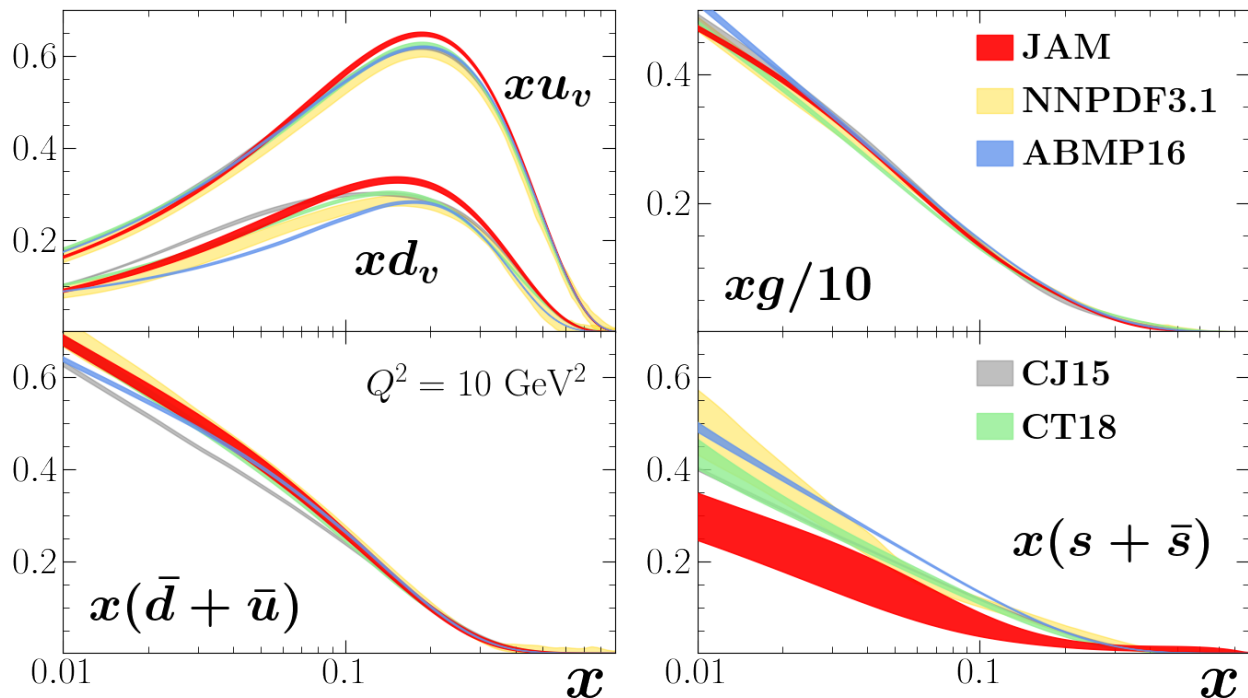


FIG. 4. Comparison of the JAM PDFs from this analysis (red bands) with NLO results from the NNPDF3.1 [74] (gold), ABMP16 [75] (blue), CJ15 [76] (gray), and CT18 [77] (green) parametrizations at the scale $Q^2 = 10 \text{ GeV}^2$, with the bands representing 1σ uncertainty. Note that x multiplied by the PDFs are shown.

NuSea Drell-Yan measurements are more directly sensitive to $\bar{d}/\bar{u}$ (*cf.* Eqs. (1) and (2)), and already provide the bulk of the constraints on the ratio even when compared to high precision W -lepton and reconstructed W data from the Tevatron and LHC. Since the STAR data overlap kinematically with the NuSea measurement, it is therefore difficult to improve on the extraction of $\bar{d}/\bar{u}$ using W -lepton production alone.

After adding the STAR data to the baseline, the SeaQuest Drell-Yan cross section ratios are then included. In this case, the SeaQuest data greatly reduce the uncertainties on the $\bar{d}/\bar{u}$ ratio, by up to $\approx 50\%$ at high x , $x \gtrsim 0.3$. As well as reducing the uncertainty, the addition of the SeaQuest data also increases the $\bar{d}/\bar{u}$ ratio for $x \approx 0.2$, where it remains above unity up to values of $x \approx 0.4$. This is a direct consequence of the extended x range of the data compared with the earlier NuSea results, from $x \approx 0.3$ up to $x \approx 0.4$, with higher precision at the largest x values. This feature is also reflected in the $\bar{d} - \bar{u}$ difference remaining positive across the entire range of x probed.

With the new data from STAR and SeaQuest included, the final $\bar{d}/\bar{u}$ ratio and $x(\bar{d} - \bar{u})$ difference are shown in Fig. 6, compared with the corresponding distributions from several other groups [74–77] at $Q^2 = 10 \text{ GeV}^2$ (see also Refs. [78, 79]). Our results are in agreement, within errors, with those from ABMP16 [75] and CT18 [77] (except with ABMP16 at low $x \lesssim 0.04$), whose ratios also

remain positive. Although there are differences in the shapes of the ratios, our results also largely agree within errors with those from NNPDF3.1 [74] and CJ15 [76]. The disagreement with CJ15 at high x results from their choice of parametrization that forces $\bar{d}/\bar{u} \rightarrow 1$ as $x \rightarrow 1$, and, more importantly, the fact that this fit predates the SeaQuest data, which pull the ratio upwards at large x .

It is instructive to examine the impact of individual datasets on the light-quark sea asymmetry, $x(\bar{d} - \bar{u})$, which we illustrate in the left panel of Fig. 7. Starting with inclusive DIS data only, and excluding data from the NMC experiment, the asymmetry is consistent with zero within large uncertainties. Upon the inclusion of the NMC data [55, 56], the errors are significantly reduced, and the asymmetry gives an indication of deviation from zero in the range $0.01 < x < 0.2$. When W -lepton, reconstructed W and Z boson, and jet production data from RHIC, Tevatron, and the LHC are included (but not the new STAR data [38]), the asymmetry becomes significantly larger, and more distinguishable from zero below $x = 0.3$. The new constraints come primarily from the high precision W and W -lepton asymmetry measurements from the Tevatron and LHC, which are sensitive to $\bar{u}$ and $\bar{d}$ (see Eq. (2)). The further addition of the NuSea [35] Drell-Yan data greatly decreases the uncertainty, showing that these data still provide a strong constraint on the asymmetry even when compared to the Tevatron and LHC W and W -lepton asymme-

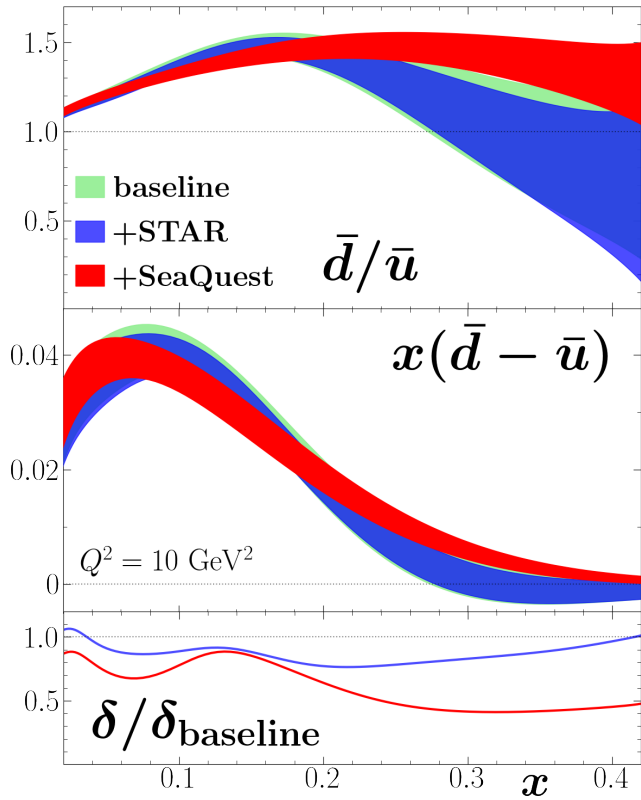


FIG. 5. Impact on the $\bar{d}/\bar{u}$ ratio (top panel) and the asymmetry $x(\bar{d} - \bar{u})$ (middle panel) of the W -lepton data from STAR [38] (blue bands) and the Drell-Yan measurement from SeaQuest [37] (red bands) relative to the “baseline” (green bands) which contains all data except these. The STAR and SeaQuest data are added in succession. The uncertainty on $\bar{d}/\bar{u}$ for these two scenarios normalized to that of the baseline are shown in the bottom panel. All bands represent 1σ uncertainty.

tries. Finally, the inclusion of the new SeaQuest [37] and STAR [38] data reduces the uncertainty on the asymmetry even further, while increasing the magnitude at $x \gtrsim 0.2$, as already seen in Fig. 5 except now displayed on a logarithmic scale.

The impact of the various datasets on the antiquark asymmetry can also be represented in the form of the truncated moment, $\int_{0.01}^1 dx(\bar{d} - \bar{u})$, illustrated in Fig. 7 in the form of the normalized yield of the Monte Carlo replicas. We choose $x = 0.01$ for the lower limit as this is approximately the extent to which existing data provide information on the asymmetry (see Fig. 1). Note that because of large uncertainties associated with the small- x extrapolation, estimates of the total moment are not as meaningful without additional constraints on the $x \rightarrow 0$ behavior. For the same combinations of datasets as described above, one observes that prior to the addition of the NMC data the truncated moments of the replicas can vary widely between -0.2 and $+0.15$. Once the NMC data are added, the moments become almost

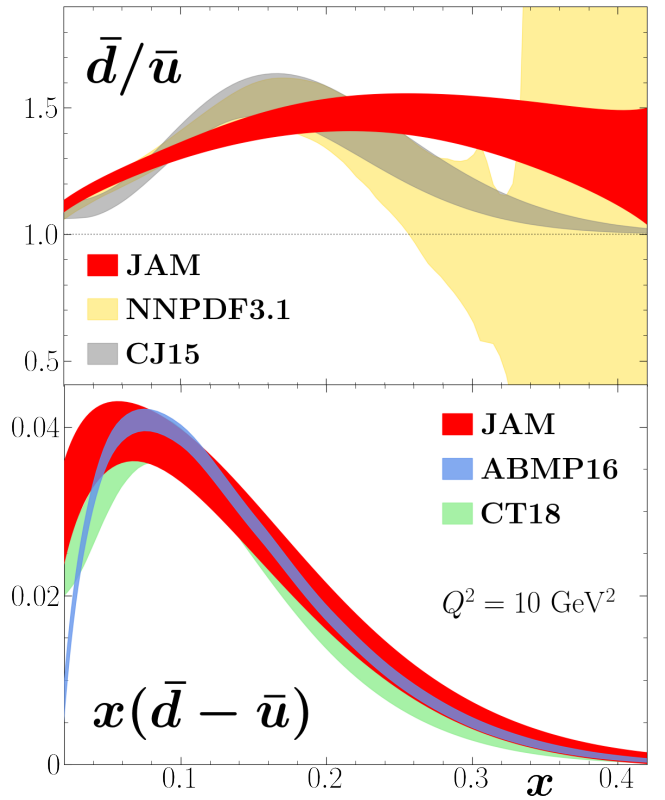


FIG. 6. Comparison of the JAM $\bar{d}/\bar{u}$ and $x(\bar{d} - \bar{u})$ PDFs (red bands) with the NLO parametrizations from NNPDF3.1 [74] (gold), ABMP16 [75] (blue), CJ15 [76] (gray), and CT18 [77] (green) at the scale $Q^2 = 10 \text{ GeV}^2$. All bands represent 1σ uncertainty.

entirely positive, and the yield continues to contract as more data are added. Once all of the data are included, the truncated moments are tightly gathered around 0.1.

The shape and magnitude of the $\bar{d} - \bar{u}$ asymmetry has long been an intriguing puzzle for our understanding of the nonperturbative structure of the nucleon. The most common interpretation of the excess of $\bar{d}$ over $\bar{u}$ in the proton sea has been that associated with chiral symmetry breaking, and the consequent prevalence of the virtual $p \rightarrow n\pi^+$ dissociation [6]. Scattering from the π^+ component of the proton wave function then naturally enhances the $\bar{d}$ distribution, even though some of this will be cancelled by the subdominant $p \rightarrow \Delta^{++}\pi^-$ dissociation, which favors $\bar{u}$ over $\bar{d}$. As an illustration, in Fig. 8 we compare the inferred JAM asymmetry with $\bar{d} - \bar{u}$ at $x > 0$ calculated from a convolution of the $p \rightarrow$ baryon + π splitting functions and the valence PDF of the pion [81–83],

$$(\bar{d} - \bar{u})(x) = [(f_{n\pi^+} + f_{\Delta^0\pi^+} - f_{\Delta^{++}\pi^-}) \otimes \bar{q}_v^\pi](x), \quad (12)$$

where the convolution integral is defined as $f \otimes q = \int_0^1 dy \int_0^1 dz f(y) q(z) \delta(x - yz)$, and y is the light-cone fraction of the proton’s momentum carried by the pion.

For the calculation shown in Fig. 8, the pion PDF

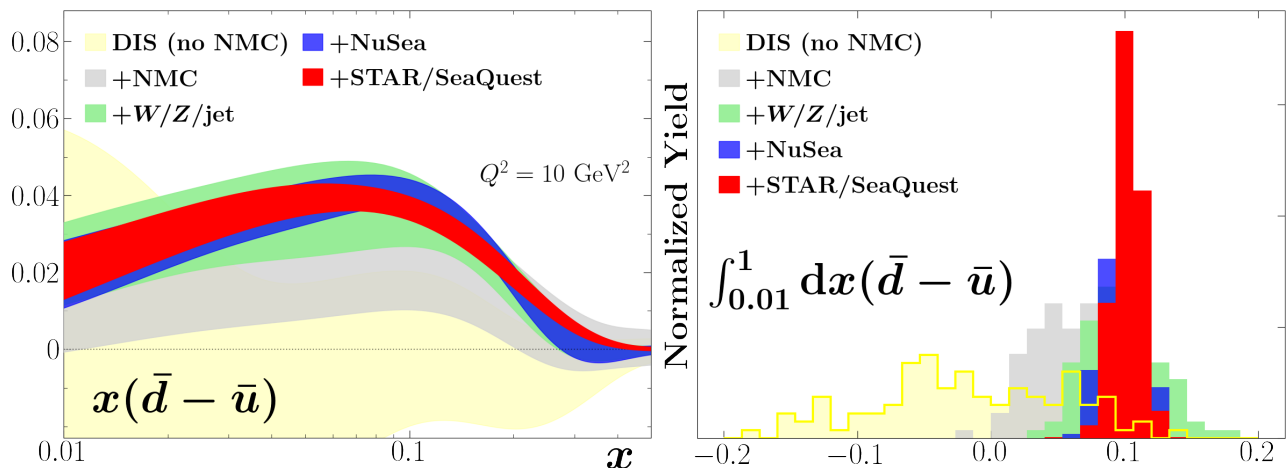


FIG. 7. Comparison of $x(\bar{d} - \bar{u})$ with different combinations of datasets (left panel). First only DIS data [54, 57–60] excluding NMC (gold band). Then data are added successively, with NMC [55, 56] (gray); W -lepton and W asymmetries/ Z boson production/jet production from RHIC [73], Tevatron [61–64, 71, 72], and the LHC [65–70] (green); NuSea [33, 35] (blue); and finally the SeaQuest Drell-Yan [37] and STAR W -lepton ratio [38] (red). For the same combinations of data, the normalized yield of the Monte Carlo replicas for the truncated moment $\int_{0.01}^1 dx(\bar{d} - \bar{u})$ is also shown (right panel).

is taken from the recent NLO analysis of pion-induced Drell-Yan and deep-inelastic leading neutron electroproduction data by Barry *et al.* [80]. The splitting functions are computed at one-pion loop order from chiral effective theory [81–83], using several different models for the ultraviolet regulators [84, 85]. For the regulator mass parameters, we use the values from the global analysis in Ref. [84], for which the integrated splitting function was found to be $\langle n \rangle_{\pi N} = 0.22$. Normalizing the various models of the regulator function, or hadronic form factor, to this value, the resulting band in Fig. 8 can be taken as a representation of the uncertainty on the calculated asymmetry. The uncertainty also includes the errors from the Monte Carlo analysis, although these are small compared to the variations in the models.

The agreement between the calculated asymmetry and the extracted JAM PDFs is generally good within the uncertainties, with the model giving a slightly harder x dependence compared with the global fit. Qualitative agreement with the extracted asymmetry is also found for various other nonperturbative models (chiral loop and other) discussed in Refs. [4, 7–16, 18–21].

VI. SUMMARY AND OUTLOOK

Our global QCD analysis confirms the importance of the recent SeaQuest measurement of pp and pD Drell-Yan cross sections [37]. The inclusion of the SeaQuest data reduces the uncertainty on the $\bar{d}/\bar{u}$ ratio by up to 50% at high values of x , and strongly suggests that the $\bar{d} - \bar{u}$ asymmetry remains positive up to $x \approx 0.4$. The impact from the latest W^+/W^- STAR data [38] on the $\bar{d}/\bar{u}$ ratio, while still important, is less dramatic due to the

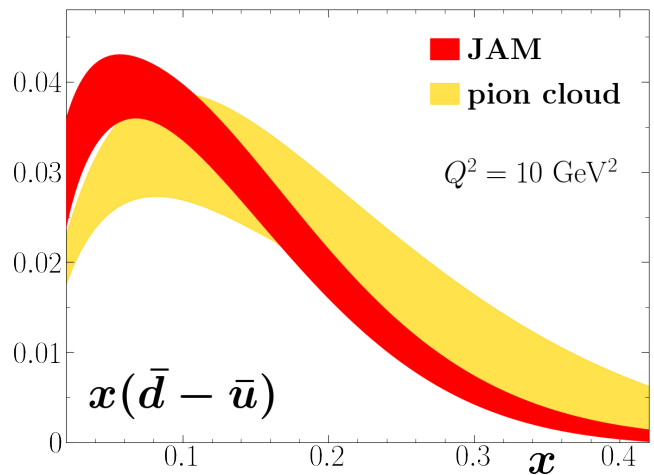


FIG. 8. Comparison of the extracted JAM $x(\bar{d} - \bar{u})$ distribution (red band) at $Q^2 = 10 \text{ GeV}^2$ with results from non-perturbative calculations based on chiral symmetry breaking and the pion cloud [81, 82]. The JAM band represents 1σ uncertainty, while the pion cloud band includes model dependence associated with the shape of the ultraviolet regulator function [80, 84, 85].

lower sensitivity of W -lepton production data compared with the Drell-Yan data.

Although the new SeaQuest data indicate some tension with the earlier NuSea Drell-Yan measurement [34, 35] at large x , a good simultaneous description of both datasets is still possible due to the larger relative uncertainties of the NuSea data at high x . The shape and magnitude of $\bar{d} - \bar{u}$ from the global analysis is consistent with expectations from nonperturbative models, such as those based

on chiral symmetry breaking in QCD, all of which predict a positive asymmetry up to $x \sim 0.4 - 0.5$. In the future, combining this analysis with semi-inclusive DIS data [53] could provide further constraints on the light-quark sea asymmetry in the proton.

ACKNOWLEDGMENTS

We thank F. Ringer and W. Vogelsang for the code used for calculating the W -lepton cross sections. We thank J. J. Ethier for his work on the W -lepton code and

data collection, and P. C. Barry and Y. Zhou for helpful discussions. This work was supported by the U.S. Department of Energy Contract No. DE-AC05-06OR23177, under which Jefferson Science Associates, LLC operates Jefferson Lab, and the National Science Foundation under grant number PHY-2110472. The work of C.C. and A.M. was supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, within the framework of the TMD Topical Collaboration, and by Temple University (C.C.). The work of N.S. was supported by the DOE, Office of Science, Office of Nuclear Physics in the Early Career Program.

-
- [1] M. Gell-Mann, *Phys. Lett.* **8**, 214 (1964).
 - [2] G. Zweig, [preprint CERN-TH-401](#).
 - [3] R. P. Feynman, *Photon-hadron interactions* (Reading, Massachusetts, 1972).
 - [4] D. F. Geesaman and P. E. Reimer, *Rep. Prog. Phys.* **82**, 046301 (2019).
 - [5] D. A. Ross and C. T. Sachrajda, *Nucl. Phys.* **B149**, 497 (1979).
 - [6] A. W. Thomas, *Phys. Lett. B* **126**, 97 (1983).
 - [7] A. W. Schreiber, P. J. Mulders, A. I. Signal, and A. W. Thomas, *Phys. Rev. D* **45**, 3069 (1992).
 - [8] E. M. Henley and G. A. Miller, *Phys. Lett. B* **251**, 453 (1990).
 - [9] S. Kumano and J. T. Londergan, *Phys. Rev. D* **44**, 717 (1991).
 - [10] W. Melnitchouk, A. W. Thomas and A. I. Signal, *Z. Phys. A* **340**, 85 (1991).
 - [11] J. Speth and A. W. Thomas, *Adv. Nucl. Phys.* **24**, 83 (1997).
 - [12] S. Kumano, *Phys. Rep.* **303**, 183 (1998).
 - [13] W. Melnitchouk, J. Speth and A. W. Thomas, *Phys. Rev. D* **59**, 014033 (1998).
 - [14] P. V. Pobylitsa, M. V. Polyakov, K. Goeke, T. Watabe and C. Weiss, *Phys. Rev. D* **59**, 034024 (1999).
 - [15] W. C. Chang and J. C. Peng, *Prog. Part. Nucl. Phys.* **79**, 95 (2014).
 - [16] M. Alberg and G. A. Miller, *Phys. Rev. C* **100**, 035205 (2019).
 - [17] R. D. Field and R. P. Feynman, *Phys. Rev. D* **15**, 2590 (1977).
 - [18] A. W. Schreiber, A. I. Signal and A. W. Thomas, *Phys. Rev. D* **44**, 2653 (1991).
 - [19] C. Bourrely and J. Soffer, *Nucl. Phys.* **B423**, 329 (1994).
 - [20] C. Bourrely and J. Soffer, *Phys. Rev. D* **51**, 2108 (1995).
 - [21] F. M. Steffens and A. W. Thomas, *Phys. Rev. C* **55**, 900 (1997).
 - [22] X. Ji, *Phys. Rev. Lett.* **110**, 262002 (2013).
 - [23] Y. Q. Ma and J. W. Qiu, *Phys. Rev. D* **98**, 074021 (2018).
 - [24] K. Orginos, A. Radyushkin, J. Karpie and S. Zafeiropoulos, *Phys. Rev. D* **96**, 094503 (2017).
 - [25] J. W. Chen, T. Ishikawa, L. Jin, H. W. Lin, Y. B. Yang, J. H. Zhang and Y. Zhao, *Phys. Rev. D* **97**, 014505 (2018).
 - [26] C. Alexandrou, M. Constantinou, K. Hadjiyiannakou, K. Jansen and F. Manigrasso, [arXiv:2106.16065 \[hep-lat\]](#).
 - [27] A. S. Ito *et al.*, *Phys. Rev. D* **23**, 604 (1981).
 - [28] P. Amaudruz *et al.*, *Phys. Rev. Lett.* **66**, 2712 (1991).
 - [29] M. Arneodo *et al.*, *Phys. Rev. D* **50**, R1 (1994).
 - [30] K. Gottfried, *Phys. Rev. Lett.* **18**, 1174 (1967).
 - [31] A. Baldit *et al.*, *Phys. Lett. B* **332**, 244 (1994).
 - [32] K. Ackerstaff *et al.*, *Phys. Rev. Lett.* **81**, 5519 (1998).
 - [33] J. Webb, Ph.D. Thesis, New Mexico State University (2002), [arXiv:hep-ex/0301031](#).
 - [34] E. A. Hawker *et al.*, *Phys. Rev. Lett.* **80**, 3715 (1998);
 - [35] R. S. Towell *et al.*, *Phys. Rev. D* **64**, 052002 (2001).
 - [36] S. D. Ellis and W. J. Stirling, *Phys. Lett. B* **256**, 258 (1991).
 - [37] J. Dove *et al.*, *Nature* **590**, 561 (2021).
 - [38] J. Adam *et al.*, *Phys. Rev. D* **103**, 012001 (2021).
 - [39] T. Becher, M. Neubert and G. Xu, *JHEP* **07**, 030 (2008).
 - [40] P. J. Ehlers, A. Accardi, L. T. Brady and W. Melnitchouk, *Phys. Rev. D* **90**, 014010 (2014).
 - [41] F. Ringer and W. Vogelsang, *Phys. Rev. D* **91**, 094033 (2015).
 - [42] M. Ebert, J. Michel, I. Stewart and F. J. Tackmann, *JHEP* **04**, 102 (2021).
 - [43] E. Moffat, T. C. Rogers, W. Melnitchouk, N. Sato and F. Steffens, *Phys. Rev. D* **99**, no.9, 096008 (2019).
 - [44] C. Cocuzza, W. Melnitchouk, A. Metz and N. Sato, in preparation (2021).
 - [45] V. N. Gribov and L. N. Lipatov, *Sov. J. Nucl. Phys.* **15**, 438 (1972).
 - [46] Y. L. Dokshitzer, *Sov. Phys. JETP* **46**, 641 (1977).
 - [47] G. Altarelli and G. Parisi, *Nucl. Phys.* **B126**, 298 (1977).
 - [48] M. Tanabashi *et al.* (Particle Data Group), *Phys. Rev. D* **98**, 030001 (2018), and 2019 update.
 - [49] N. Sato, W. Melnitchouk, S. E. Kuhn, J. J. Ethier and A. Accardi, *Phys. Rev. D* **93**, 074005 (2016).
 - [50] N. Sato, J. J. Ethier, W. Melnitchouk, M. Hirai, S. Kumano and A. Accardi, *Phys. Rev. D* **94**, 114004 (2016).
 - [51] J. J. Ethier, N. Sato and W. Melnitchouk, *Phys. Rev. Lett.* **119**, 132001 (2017).
 - [52] N. Sato, C. Andres, J. J. Ethier and W. Melnitchouk, *Phys. Rev. D* **101**, 074020 (2020).
 - [53] E. Moffat, W. Melnitchouk, T. Rogers and N. Sato, *Phys. Rev. D* **104**, 016015 (2021).
 - [54] A. C. Benvenuti *et al.*, *Phys. Lett. B* **223**, 485 (1989); *ibid.* **B 237**, 592 (1990).
 - [55] M. Arneodo *et al.*, *Nucl. Phys.* **B483**, 3 (1997).
 - [56] M. Arneodo *et al.*, *Nucl. Phys.* **B487**, 3 (1997).
 - [57] L. W. Whitlow *et al.*, *Phys. Lett. B* **282**, 475 (1992).
 - [58] S. P. Malace *et al.*, *Phys. Rev. C* **80**, 035207 (2009).

- [59] S. Tkachenko *et al.*, [Phys. Rev. C **89**, 045206 \(2014\)](#).
- [60] H. Abramowicz *et al.*, [Eur. Phys. J. C **75**, 580 \(2015\)](#).
- [61] T. Aaltonen *et al.*, [Phys. Lett. B **692**, 232 \(2010\)](#).
- [62] V. M. Abazov *et al.*, [Phys. Rev. D **76**, 012003 \(2007\)](#).
- [63] T. Aaltonen *et al.*, [Phys. Rev. Lett. **102**, 181801 \(2009\)](#).
- [64] V. M. Abazov *et al.*, [Phys. Rev. Lett. **112**, 151803 \(2014\)](#); *ibid.* [114](#), 049901 (2015).
- [65] S. Chatrchyan *et al.*, [JHEP **04**, 050 \(2011\)](#).
- [66] S. Chatrchyan *et al.*, [Phys. Rev. Lett. **109**, 111806 \(2012\)](#).
- [67] S. Chatrchyan *et al.*, [Phys. Rev. D **90**, 032004 \(2014\)](#).
- [68] V. Khachatryan *et al.*, [Eur. Phys. J. C **76**, 469 \(2016\)](#).
- [69] R. Aaij *et al.*, [JHEP **12**, 079 \(2014\)](#).
- [70] R. Aaij *et al.*, [JHEP **01**, 155 \(2016\)](#).
- [71] A. Abulencia *et al.*, [Phys. Rev. D **75**, 092006 \(2007\)](#) [Erratum: [Phys. Rev. D **75**, 119901 \(2007\)](#)].
- [72] V. M. Abazov *et al.*, [Phys. Rev. Lett. **101**, 062001 \(2008\)](#).
- [73] B. I. Abelev *et al.*, [Phys. Rev. Lett. **97**, 252001 \(2006\)](#).
- [74] R. D. Ball *et al.*, [Eur. Phys. J. C **77**, 663 \(2017\)](#).
- [75] S. Alekhin, J. Blümlein, S. Moch and R. Pláčákytė, [Phys. Rev. D **96**, 014011 \(2017\)](#).
- [76] A. Accardi, L.T. Brady, W. Melnitchouk, J. F. Owens and N. Sato, [Phys. Rev. D **93**, 114017 \(2016\)](#).
- [77] T. Hou *et al.*, [Phys. Rev. D **103**, 014013 \(2021\)](#).
- [78] S. Park, A. Accardi, X. Jing and J. F. Owens, [arXiv:hep-ex/2108.057866](#).
- [79] M. Guzzi *et al.*, [arXiv:hep-ex/2108.06596](#).
- [80] P. C. Barry, C.-R. Ji, N. Sato and W. Melnitchouk, [arXiv:2108.05822 \[hep-ph\]](#).
- [81] Y. Salamu, C.-R. Ji, W. Melnitchouk and P. Wang, [Phys. Rev. Lett. **114**, 122001 \(2015\)](#).
- [82] Y. Salamu, C.-R. Ji, W. Melnitchouk, A. W. Thomas, P. Wang and X. G. Wang, [Phys. Rev. D **100**, 094026 \(2019\)](#).
- [83] M. Burkardt, K. S. Hendricks, C.-R. Ji, W. Melnitchouk and A. W. Thomas, [Phys. Rev. D **87**, 056009 \(2013\)](#).
- [84] P. C. Barry, N. Sato, W. Melnitchouk and C.-R. Ji, [Phys. Rev. Lett. **121**, 152001 \(2018\)](#).
- [85] J. R. McKenney, N. Sato, W. Melnitchouk and C.-R. Ji, [Phys. Rev. D **93**, 054011 \(2016\)](#).