Complementarity of experimental and lattice QCD data on pion parton distributions

P. C. Barry[1], C. Egerer,[1] J. Karpie[2], W. Melnitchouk[1], C. Monahan[1,3] K. Orginos,[1,3]
Jian-Wei Qiu,[1,3] D. Richards,[1] N. Sato,[1] R. S. Sufian[1,3] and S. Zafeiropoulos[4]

(Jefferson Lab Angular Momentum (JAM) and HadStruc Collaborations)

[1]*Jefferson Lab, Newport News, Virginia 23606, USA*
[2]*Physics Department, Columbia University, New York City, New York 10027, USA*
[3]*Department of Physics, William & Mary, Williamsburg, Virginia 23185, USA*
[4]*Aix Marseille Univ, Université de Toulon, CNRS, CPT, Marseille, France*

We extract pion parton distribution functions (PDFs) in a Monte Carlo global QCD analysis of experimental data together with reduced Ioffe time pseudo-distributions and matrix elements of current-current correlators generated from lattice QCD. By including both experimental and lattice QCD data, our analysis rigorously quantifies both the uncertainties of the pion PDFs and systematic effects intrinsic to the lattice QCD observables. The reduced Ioffe time pseudo-distributions significantly decrease the uncertainties on the PDFs, while the current-current correlators are limited by the systematic effects associated with the lattice. Consistent with recent phenomenological determinations, the behavior of the valence quark distribution of the pion at large momentum fraction is found to be $\sim(1-x)^{\beta_{\text{eff}}}$ with $\beta_{\text{eff}} \approx 1.0$–$1.2$.

I. INTRODUCTION

The pion is one of the most important particles in nature, yet three-quarters of a century after its discovery fundamental questions remain about its properties and behavior. On the one hand, it has long been identified as the pseudo-Goldstone boson associated with chiral symmetry breaking, which governs the low-energy interactions of hadrons. On the other hand, the pion's internal structure in terms of its bound state of quarks and gluons (partons) can be revealed in high-energy scattering reactions. Understanding the structure of the pion provides insight into the nature of the strong force itself.

Details of the internal partonic structure of the pion have remained elusive because it cannot be probed as a fixed target. However, secondary pion beams have been used to scatter from stationary nuclear targets, and inclusive Drell-Yan (DY) lepton-pair production measurements have been made at CERN and Fermilab [1,2]. These, along with prompt photon data, have been used in phenomenological QCD analyses to extract the pion parton distribution functions (PDFs) at moderate and large parton momentum fractions, x [3–9]. Recently, the Jefferson Lab Angular Momentum (JAM) Collaboration utilized the leading neutron (LN) deep-inelastic electroproduction measurements from HERA [10,11], together with the DY data, to capture the small momentum fraction region and constrain the sea quark and gluon distributions in the pion [12]. Further studies in this vein have been performed with the addition of large-p_T differential DY data [13], with its enhanced sensitivity to the gluon distribution, as well as with threshold resummation on the DY hard coefficients [14].

The behavior of the valence quark distribution in the region where the momentum fraction is large has elicited much debate, particularly in the limit where $x \to 1$. Here, the valence PDF is modeled by an asymptotic behavior $\sim(1-x)^{\beta_{\text{eff}}}$, and various nonperturbative and perturbative QCD models predict different values for β_{eff}, ranging from $\beta_{\text{eff}} \approx 1$ to $\beta_{\text{eff}} \approx 2$ [15–35]. Experiments have been proposed to further constrain the large-x behavior using the Sullivan process at Jefferson Lab [36] and the future Electron-Ion Collider (EIC) [37,38]. An earlier analysis using threshold resummation on the DY data found $\beta_{\text{eff}} \approx 2$ [39], while more recently the JAM Collaboration performed a global analysis of pion PDFs using more

advanced threshold resummation technology, particularly the double Mellin method [40], finding $\beta_{\text{eff}} \approx 1.2$ [14].

Experimental cross sections are related to the integrals of PDFs from $x_{\min}$ to 1 with $x_{\min}$ fixed by external kinematics. In fact, no experiment can measure the parton momentum fraction x directly. If the kinematics force $x_{\min}$ to be ~1 or near the exclusive threshold, the role of logarithms in coefficient functions that match the PDFs to cross sections becomes important. Consequently, PDFs extracted from these cross sections become highly sensitive to the perturbative order of the matching coefficients, particularly the threshold effects. In Ref. [14], threshold resummation stemming from soft gluon radiation was applied to the DY hard coefficients up to next-to-leading logarithmic (NLL) accuracy as a potentially necessary theoretical correction to the perturbative expansion. Since the PDFs fall rapidly near threshold, the resummed hard coefficients may contribute substantially to the overall hadronic cross section [39–42]. Threshold resummation was not applied in [14] to the LN cross sections since DIS-like observables have less of an overall impact from large logarithms and are less sensitive to large-x behaviors than for the case of DY [43]. The description of both the DY and LN data was compatible with the inclusion of the double Mellin resummation [14], so both the next-to-leading order (NLO) and NLO + NLL methods of calculating the short distance coefficients remain viable alternatives for extracting PDFs from available experimental data.

On the theoretical side, there has also been considerable recent interest in computing the pion's internal structure through lattice QCD simulations [44–53]. Unlike the experimental cross sections, lattice QCD calculable matrix elements of operators composed of two-quark, two-gluon or two-current correlations with a separation z in position space are matched to PDFs in terms of the short-distance factorization (SDF) formalism through integrals over the entire range of x [54–56]. The extraction of PDFs from these matrix elements is therefore not as sensitive to threshold resummation of the corresponding matching coefficients as it is in the case of DY [57]. The sensitivity to the large-x behavior of the PDFs depends on the relative size of the large-x contribution to the matching coefficient at different values of z. Like the experimental cross sections, these matrix elements are also subject to power corrections from the large-z regime, which limits the lattice data that can be used for extracting PDFs in the SDF approach. On the other hand, in the large momentum effective theory (LaMET) approach, lattice data for the same two-parton correlation matrix elements are taken at all z values and are Fourier transformed into the quasi-PDF in momentum space, which is then matched to the PDF as the pion momentum $P_z \to \infty$. For finite P_z values, the matching between the quasi-PDF and PDF also has power corrections that are proportional to $[x^2(1-x)P_z^2]^{-1}$, which

could hamper the ability to accurately determine the large-x behavior of the PDFs [58].

The LaMET and SDF approaches provide complementary methods for extracting PDFs from lattice calculations, with different systematic uncertainties. The success of these approaches has encouraged significant effort aimed at constraining PDFs through lattice QCD, using a variety of methods and lattice actions [48,59–73]. These methods generally provide more information about PDFs than does the traditional approach of calculating the Mellin moments of PDFs from matrix elements of local operators. That approach has been limited to the lowest few moments by the reduced rotational symmetry of the Euclidean lattice, and generally provides weaker constraints on PDFs than methods that give access to the x dependence directly [74–76].

In the present analysis, we make use of new data from lattice QCD simulations of pion matrix elements with space-separated quark-quark and current-current (CC) correlators, which have well-defined continuum limits, and can be factorized to the standard light-cone PDFs with perturbatively calculable matching coefficients. The lattice data on matrix elements of quark-quark correlators are presented in the form of reduced Ioffe time pseudo-distributions (Rp-ITDs) [55,77]. In this analysis lattice data on both Rp-ITDs and matrix elements of CC correlators are treated on the same footing as experimental data, and are analyzed simultaneously. In the JAM framework, both experimental and lattice data have been used to constrain the unpolarized and polarized nucleon PDFs [78], as well as the nucleon's transversity distribution [79]. The complementarity of the experimental and lattice observables provides an opportunity to learn about both PDFs and the systematic effects associated with lattice calculations of hadron structure.

We proceed with the organization of this paper as follows. In Sec. II we outline the analysis framework, describing both the experimental and lattice observables, and summarizing the QCD analysis methodology employed here. The results of the simultaneous analysis of experimental and lattice QCD data are presented in Sec. III. There, we discuss in detail the impact of the lattice Rp-ITDs and CC correlator data on the pion valence quark PDF, and in particular the quantification of the systematic uncertainties, for both the NLO and NLO + NLL approximations to the hard coefficients. Finally, in Sec. IV we summarize our conclusions and discuss future directions in which this analysis can be taken.

II. ANALYSIS FRAMEWORK

Confinement ensures that quarks and gluons cannot be directly observable. Instead, theoretical and analytic tools are needed to infer the internal partonic structure of hadrons from "good" cross sections, defined as those that can be factorized in QCD into convolutions of soft, universal

nonperturbative parts, such as those involving PDFs, and perturbatively calculable, short-distance hard coefficients. Corrections to these factorized theoretical cross sections are suppressed by powers of the large momentum transfer involved in the scattering process. Predictions follow when cross sections with different hard coefficients, but the same nonperturbative parts, are compared.

While experiments do not detect partons themselves, the detection of hadrons in the final state can reveal the initial-state structure of the hadronic target from the experimental data through Bayesian inference. Such tools allow us to build theoretical observables describing processes involving partons, and to compare them with the available data. Much in the same way as different experimental datasets are used in phenomenological extractions of PDFs, the universality of the PDFs in different lattice observables can also be tested. Here, we treat lattice QCD data on the same footing as experimental data. We use factorization theorems for lattice observables to describe the results in terms of convolutions of hard coefficients with PDFs, and the approach to the extraction of PDFs from lattice data is analogous to the methods used in global QCD analysis of experimental data. In practice, in our analysis we include DY and LN electroproduction observables in conjunction with Rp-ITD and CC correlator lattice QCD data to extract the PDFs in the pion. In the following, we describe the theoretical and analysis framework employed in this analysis.

A. Experimental observables

In the DY process [80] two hadrons collide with an invariant center of mass energy $\sqrt{S}$, producing a detected $\mu^+\mu^-$ lepton pair with invariant mass Q. The cross section can be factorized as a convolution of the PDFs of the incident pion and the target nucleus A, from which the pion scatters, with the perturbatively calculable hard coefficients. Reported differentially in the Feynman variable, $x_F = x_\pi^0 - x_A^0$, and in $\sqrt{\tau} = Q/\sqrt{S}$, the cross section is written as [81]

$$\frac{\mathrm{d}\sigma}{\mathrm{d}x_F \mathrm{d}\sqrt{\tau}} = \frac{4\pi\alpha^2}{9Q^2 S} \sum_{ij} \int_{x_\pi^0}^1 \mathrm{d}x_\pi \int_{x_A^0}^1 \mathrm{d}x_A f_i^\pi(x_\pi, \mu) f_j^A(x_A, \mu)$$
$$\times \mathcal{C}_{ij}^{\mathrm{DY}}(x_\pi, x_\pi^0, x_A, x_A^0, Q, \mu), \quad (1)$$

where $f_{i(j)}^{\pi(A)}$ is a PDF of flavor $i(j)$ in $\pi(A)$. The variable $x_{\pi(A)}^0 = \sqrt{\tau}e^{\pm Y}$ is the minimum momentum fraction to produce the lepton pair, where Y is the rapidity of the lepton pair.

The hard coefficient $\mathcal{C}^{\mathrm{DY}}$ is calculated through perturbative QCD up to a desired order in the strong coupling α_s. We use two forms of the calculation of the short-distance coefficient in DY: (i) fixed order up to $\mathcal{O}(\alpha_s)$, referred to as "NLO," and (ii) fixed order with threshold resummation up

to NLL accuracy, which we refer to as "NLO + NLL." The threshold region corresponds to values of $z = \tau/x_\pi x_A \to 1$, and large logarithms appear in the form $\alpha_s^k \log^{2k-1}(1-z)/(1-z)$ for all orders k. In some regions of x_F and τ, the soft gluon resummation greatly influences $\mathcal{C}^{\mathrm{DY}}$, leading to significant differences between the hard coefficients calculated at NLO and NLO + NLL. These changes are reflected in the resulting PDFs, and the universality of these PDFs must be tested by independent observables. In Ref. [14], several methods of calculating the threshold resummation were studied, based on the "Mellin-Fourier" and "double Mellin" methods. The Mellin-Fourier approach was shown by Lustermans et al. [82] to miss certain terms in the resummation of the same logarithmic orders, whereas the double Mellin method accounted for these corrections [40]. The Mellin-Fourier approaches also showed a worse description of the data at large x_F and τ. In this analysis, we use the preferred double Mellin approach [40] to compute the NLO + NLL result and compare it with that obtained from the pure NLO calculation.

The pion-induced DY data were taken from the Fermilab E615 experiment [1] and the NA10 experiment [2] at CERN. The E615 experiment made use of a pion beam with an energy of 252 GeV, while the NA10 experiment utilized beam energies of 194 and 286 GeV. To avoid the J/ψ and Υ resonance regions and edges of phase space we limit ourselves to the range $4.16 < Q < 7.68$ GeV and $0 < x_F < 0.9$. After these kinematic cuts are made, we fit to 61 data points from the E615 experiment, and 36 and 20 points from the NA10 experiment with the 194 and 286 GeV pion beam energies, respectively. Since the data were collected on a tungsten target, tungsten PDFs need to be used in the calculation, and for these we take the central values from the EPPS16 [83] global nuclear PDF analysis. We find, however, that our results are insensitive to the changes of the tungsten PDFs between the EPPS16 [83] and nCTEQ [84] nuclear PDF parametrizations.

The second method of obtaining information on the structure of the pion is less direct but also relatively well established. In this case, LN electroproduction is used through the Sullivan process [85] to probe the pion structure function in tagged deep-inelastic scattering, $ep \to e'nX$, with the neutron detected in the far forward region. In the one-photon exchange approximation, the electron beam radiates a virtual photon with 4-momentum $q = \ell - \ell'$, where $\ell(\ell')$ is the initial (final) state lepton momentum, and $q^2 \equiv -Q^2$, while the momentum transfer squared between the initial and final nucleons is $t = (p - p')^2$, with $p(p')$ the 4-momentum of the initial (final) state nucleon. The LN cross section is then proportional to the LN structure function,

$$\frac{\mathrm{d}^3\sigma}{\mathrm{d}x\mathrm{d}Q^2\mathrm{d}x_L} = \frac{4\pi\alpha^2}{xQ^4}\left(1 - y_e + \frac{y_e^2}{2}\right)F_2^{\mathrm{LN}}(x, Q^2, x_L), \quad (2)$$

where $x = Q^2/2p \cdot q$ is the Bjorken scaling variable, x_L is the longitudinal momentum fraction the neutron carries relative to the initial proton, $y_e = q \cdot p/q \cdot \ell \approx Q^2/sx$ is the lepton inelasticity, and $s = (\ell + p)^2$ is the total invariant mass squared of the collision. In the limit $|t| \to 0$, which occurs in the region $x_L \to 1$, the virtual photon absorption reaction $\gamma^* p \to nX$ is dominated by the exchange of pions [85–89], which can be described through the $p \to \pi^+ n$ splitting function, $f_{\pi N}$. The LN structure function is represented as a convolution of $f_{\pi N}$ with the structure function of the pion, F_2^π,

$$F_2^{\mathrm{LN}}(x, Q^2, x_L) = 2f_{\pi N}(\bar{x}_L) F_2^\pi(x_\pi, Q^2), \qquad (3)$$

where the splitting function $f_{\pi N}$, evaluated at $\bar{x}_L \equiv 1 - x_L$, is the light-cone momentum distribution of the pion in the nucleon whose functional form can be found in Refs. [85,86,90–95]. A cutoff mass Λ is invoked to regulate the ultraviolet divergence from the $k_\perp^2$ integration in the definition of $f_{\pi N}(\bar{x}_L)$, and is consequently a free parameter in our analysis, see Refs. [12,13,96].

The structure function of the pion, evaluated at $x_\pi = x/\bar{x}_L$, is a convolution of the pion PDFs with the hard coefficients for DIS on a pion,

$$F_2^\pi(x_\pi, Q^2) = \sum_i \int_{x_\pi}^1 \mathrm{d}\xi f_i^\pi(x_\pi/\xi, \mu^2) \mathcal{C}_i^{\mathrm{DIS}}(\xi, \mu^2, Q^2). \qquad (4)$$

Similar to Eq. (1), the range of the integral is from x_π to 1. However, unlike for the DY cross section, here the large logarithms near threshold do not contribute significantly to the perturbative coefficients. For the treatment of the short-distance factorization coefficient, we therefore do not include threshold resummation, and calculate $\mathcal{C}^{\mathrm{DIS}}$ up to NLO in the strong coupling.

The measurements of the LN electroproduction were performed at HERA [10,11], and with focus on the low x_π ($10^{-3} \lesssim x_\pi \lesssim 0.5$) and large Q^2 ($7 < Q^2 < 10^3 \text{ GeV}^2$) regions, provided access to pion PDFs across a wider range of kinematics than with DY experiments alone. While the H1 experiment reported the LN structure function as in Eq. (3), in an effort to reduce systematic uncertainties the ZEUS experiment reported the ratio of the LN structure function to the inclusive structure function,

$$r(x, Q^2, x_L) = \frac{\mathrm{d}^3\sigma/\mathrm{d}x\mathrm{d}Q^2\mathrm{d}x_L}{\mathrm{d}^2\sigma/\mathrm{d}x\mathrm{d}Q^2} \Delta x_L, \qquad (5)$$

where Δx_L is the bin size in x_L measured. Cuts on the data are made for $\bar{x}_L$ to ensure the dominance of the one-pion exchange mechanism. Phenomenologically, in Refs. [12,96] the optimal limit was determined to be $\bar{x}_L < 0.2$, which

leaves 58 and 50 data points for the H1 and ZEUS experiments, respectively.

B. Lattice observables

PDFs are formally defined as Fourier transforms of nonlocal matrix elements of quark and gluon fields evaluated on the light front [97]. The spacetime signature of the Euclidean lattice precludes direct calculation of matrix elements of lightlike separated quark and gluon fields, but does admit computation of the matrix elements of spacelike nonlocal operators. Unlike experimental cross sections, lattice QCD calculated matrix elements of spacelike nonlocal operators are not physical observables, and their values depend on the choice of nonperturbative operator renormalization. Consequently, perturbatively calculable coefficients for factorizing lattice calculated matrix elements will not only depend on the factorization scheme to relate to PDFs, but will also be sensitive to how the lattice calculated matrix elements are renormalized. The matching coefficients take into account only the perturbative dependence of the renormalization scheme, and cannot account for nonperturbative effects in the renormalization of the spacelike nonlocal operators.

In this work, we focus on the lattice calculable matrix elements of spacelike separated quark-quark correlators linked by a straight Wilson line and CC correlators. In a manner similar to experimental cross sections, the renormalized matrix elements of these two types of correlators can be factorized into PDFs with perturbative matching coefficients in the SDF framework [54,56]. The resulted correlation functions provide complementary information on PDFs, however, the results are subjected to different systematic effects. The quark-quark correlator with a Wilson line has a power ultraviolet divergence, which must be removed nonperturbatively. Several nonperturbative renormalization schemes have been proposed in the literature [58,98–103]. In this paper, we use data from lattice QCD calculations with the renormalization defined in the reduced pseudo-PDF approach [55], which removes ultraviolet divergences through construction of suitable ratios of matrix elements.

The nonperturbative renormalization of CC operators is straightforward. In addition, different choices of currents provide additional sensitivities to different combination of PDFs, such as the direct access to the difference of quark and antiquark distributions [51]. CC correlator calculations require lattice computations of four-point functions, instead of the three-point functions in the case of the quark-quark correlator with a Wilson line, making the task more computationally intensive. Comparison of the data obtained from correlation functions of these two complementary correlators may provide a way to estimate systematic uncertainties associated with lattice calculations relevant to hadron structure.

1. Reduced Ioffe time pseudo-distributions

The pion PDFs can be inferred from the Lorentz-invariant Ioffe time pseudo-distribution (pseudo-ITD), $\mathcal{M}$, which is defined as

$$\mathcal{M}(\nu, z^2) = \frac{1}{2p^0} \langle p|\bar{\psi}(0)\gamma^0 \mathcal{W}(z;0)\psi(z)|p\rangle, \quad (6)$$

where $\mathcal{W}(z;0)$ is a straight gauge link of length z in the fundamental representation, $z = (0; 0, 0, z_3)$ is the displacement of the bare quark field, ψ, and $|p\rangle$ is the pion state with momentum p. In the continuum limit, the pseudo-ITD is a function of only Lorentz invariant quantities, namely, the Ioffe time $\nu = p \cdot z$ [104] and the square of the separation.

The data used in this work are presented as a *reduced* pseudo-ITD (Rp-ITD) [49],

$$\mathfrak{M}(\nu, z^2) = \frac{\mathcal{M}(\nu, z^2)}{\mathcal{M}(0, z^2)}, \quad (7)$$

which is the ratio of the function with respect to the $p = 0$ distribution. The ultraviolet divergences that are associated with the gauge link when z is spacelike factorize multiplicatively and contain no dependence on ν [105], and by taking the ratio of the pseudo-ITDs at ν to that at $\nu = 0$ these divergences cancel. This ratio is renormalization group invariant, so the factorization will not have the scheme dependence that other renormalization approaches would have. Additionally, statistical uncertainties and a number of systematic uncertainties cancel [55,102], resulting in smaller errors on the Rp-ITD.

To extract the PDFs of the pion from the Rp-ITD, we expand the Rp-ITD using an operator product expansion [77] in terms of local, nonperturbative, renormalized matrix elements and perturbatively calculable Wilson coefficients [105–107]. The factorization of $\mathfrak{M}(\nu, z^2)$ involves these coefficients and the light-cone Ioffe time distribution, whose Fourier transform with respect to ν gives the parton density $f_q(x)$ for flavor q. The integration range of x in the factorization formula is from $x = -1$ to 1, and from cross symmetry one has $f_q(x) = q(x)$ for positive x, and $f_q(x) = -\bar{q}(-x)$ for negative x, where $|x|$ corresponds to the momentum fraction of the quark and antiquark, respectively. To obtain a meaningful description of the valence quark distribution in the pion, $q_v(x) = q(x) - \bar{q}(x)$, we take the real component of the factorized Rp-ITD $\mathfrak{M}(\nu, z^2)$ [49]. The imaginary component of $\mathfrak{M}(\nu, z^2)$ computed in these datasets cannot be expressed in terms of the physical pion sea quark distribution, because the calculation does not at present include the disconnected contributions.

The Rp-ITD takes the form of a convolution of the PDF with a hard coefficient function, to which correction terms are added to account for systematic uncertainties. The real part of $\mathfrak{M}$ can be expressed as

$$\begin{aligned}
\text{Re}\mathfrak{M}(\nu, z^2) = &\int_0^1 dx\, q_v(x, \mu_{\text{lat}})\mathcal{C}^{\text{Rp-ITD}}(x\nu, z^2, \mu_{\text{lat}}) \\
&+ z^2 B_1(\nu) + \frac{a}{|z|}P_1(\nu) \\
&+ e^{-m_\pi(L-z)}F_1(\nu) + \dots,
\end{aligned} \quad (8)$$

where $q_v \equiv u - \bar{u} = \bar{d} - d$ in π^+, and the $\mathcal{C}^{\text{Rp-ITD}}$ is the cosine transformation of the coefficient function, as discussed in Refs. [77,105–107]. The effects from threshold resummation on these types of matching coefficients for lattice observables exhibiting integrations from $x = 0$ to 1 were shown to be small [57], and in this analysis we therefore keep the short distance coefficient $\mathcal{C}^{\text{Rp-ITD}}$ to NLO. In practice, we apply a truncation on the Taylor expansion of the cosine, expressing Eq. (8) in terms of the moments of the valence quark distribution and moments of the coefficient function $\mathcal{C}^{\text{Rp-ITD}}$, which are calculable analytically [77]. In Eq. (8), μ_{lat} refers to the scale at which the PDF and coefficient function are specified; more details on the choice of this scale appear in Sec. II C.

Beyond the leading twist term, the systematic correction terms in Eq. (8) describe corrections relating to higher twist $[z^2 B_1(\nu)]$, lattice spacing $[(a/|z|)P_1(\nu)]$, and finite volume $[e^{-m_\pi(L-z)}F_1(\nu)]$ effects, where m_π is the pion mass characterizing the lattice ensemble, and L is the spatial extent of the lattice. These can be understood as matrix elements arising from higher twist operators in the OPE, Symanzik improvement of the lattice operators, and continuum finite volume calculations, respectively. While these terms would generally have both ν and z^2 dependence, in practice we neglect the z^2 dependence in these terms since this is further suppressed by α_s or $(z\Lambda_{\text{QCD}})^2$. There also exists a possibility for inclusion of target mass corrections proportional to powers of $m_\pi^2 z^2$. These effects appear from the trace terms in the OPE, just as when analyzing experimental cross sections. In Ref. [108], models have been used to demonstrate that target mass effects will be quite small even when they would naively appear of $O(1)$. The ellipsis in Eq. (8) represents higher order effects and other systematic uncertainties, such as $\mathcal{O}(a\Lambda_{\text{QCD}})$ discretization and pion mass corrections, which cannot be probed given the constraints of the currently available data. Additional systematic effects may also include excited state contamination in the lattice data, which for the pion are less important than in the proton case [109], and perturbative truncation errors.

Each of the functional forms of the accessible systematic correction terms in Eq. (8) is parametrized by a series of ν-dependent functions, as in Ref. [110], given by

$$B_1(\nu) = \sum_n \sigma_{0,n}(\nu) b_n, \qquad (9a)$$

$$P_1(\nu) = \sum_n \sigma_{0,n}(\nu) p_n, \qquad (9b)$$

$$F_1(\nu) = \sum_n \sigma_{0,n}(\nu) f_n, \qquad (9c)$$

where $\sigma_{0,n}(\nu)$ is defined as

$$\sigma_{0,n}(\nu) = \int_0^1 dx \cos(\nu x) x^a (1-x)^b J_n^{(a,b)}(x). \quad (10)$$

Here, $J_n^{(a,b)}(x)$ is a transformed Jacobi polynomial existing in the range $0 \le x \le 1$. Since the set of all $J_n^{(a,b)}(x)$ forms a complete basis of functions for all $a, b > -1$, an infinite series of $\sigma_{0,n}$ could reproduce any function of Ioffe time for any $a, b > -1$. In this analysis, we include a and b in the set of fit parameters. By construction, $\mathfrak{M}(0, z^2) = 1$, and the leading twist contribution satisfies this condition through the valence quark number sum rule. At $\nu = 0$ the systematic corrections must therefore vanish. For this reason, we do not include the $n = 0$ term in the series in Eqs. (9), since these are not guaranteed to vanish at $(n, \nu) = (0, 0)$. Instead, we begin the expansion at $n = 1$ and take up to the $n = 2$ term, since both terms vanish at $\nu = 0$.

We use the available Rp-ITD data from the Jefferson Lab HadStruc group [49], calculated on the ensembles labeled `a127m413` and `a127m413L`, whose parameters are given in Table I. The two datasets each have the same lattice spacing $a \approx 0.127$ fm and pion mass $m_\pi \approx 413$ MeV, but different lattice volumes, with the larger volume dataset `a127m413L` having $L = 32a$ and the smaller volume dataset `a127m413` having $L = 24a$. Each dataset includes the momentum values of $p_i = i \times (2\pi/La)$, where $i \in \{1, 2, 3\}$, and each features discrete separations z up to $6a$, with `a127m413L` including z up to $8a$. The signal-to-noise ratio of the lattice points decays exponentially with momentum, making the calculation of the p_3 points in the

TABLE I. Parameters for the gauge ensembles used in this analysis: lattice spacing (a), pion mass (m_π), inverse gauge coupling (β), spatial (L) and temporal (T) sizes, and the smallest momentum in each ensemble (p_1) [49,51]. Note that while in Ref. [49] the ID names are `a127m415` and `a127m415L` for $m_\pi \approx 415$ MeV, the entries in this table for `a127m413` and `a127m413L` represent the same lattice ensembles that were used in [49].

ID	a (fm)	m_π (MeV)	β	$L^3 \times T$	p_1 (MeV)
`a127m413`	0.127(2)	413(4)	6.1	$24^3 \times 64$	406(6)
`a127m413L`	0.127(2)	413(5)	6.1	$32^3 \times 96$	305(5)
`a94m358`	0.094(1)	358(3)	6.3	$32^3 \times 64$	411(4)
`a94m278`	0.094(1)	278(4)	6.3	$32^3 \times 64$	411(4)

`a127m413` ensemble incompatible with statistical fluctuations. For this reason, we exclude these points. We also cut the $z \le 2a$ data points to avoid discretization errors and contact terms arising from the chosen discretization of the quark action.

2. Current-current correlators

CC correlators provide another class of lattice calculable quantities that can be factorized into PDFs convoluted with perturbative matching kernels [56]. The relevant available data [51] were obtained from a pair of vector and axial-vector currents, which are written in the form

$$\Sigma_{VA}^{\alpha\beta}(z, p) = z^4 Z_V Z_A \langle p | [\bar{\psi}\gamma^\alpha \psi](z)[\bar{\psi}\gamma^\beta \gamma^5 \psi](0) | p \rangle + (V \leftrightarrow A), \qquad (11)$$

where Z_V and Z_A are the renormalization constants of the local lattice vector and axial-vector currents, respectively. Through time-reversal and parity invariance, the CC correlator in Eq. (11) is antisymmetric in the Lorentz indices $\{\alpha, \beta\}$. The CC correlator can then be expressed as two dimensionless pseudo-structure functions, T_1 and T_2, which are functions of the Lorentz invariants ν and z^2 [46]. Choosing the momentum $p = (p^0; 0, 0, p^3)$ with the Lorentz indices $\alpha = 1$, $\beta = 2$, we isolate the function T_1, which can be related to the light-cone valence quark PDF by

$$T_1(\nu, z^2) = \int_0^1 dx q_v(x, \mu_{\text{lat}}) \mathcal{C}^{\text{CC}}(x\nu, z^2, \mu_{\text{lat}}) + z^2 B_1(\nu) + a R_1(\nu) + \dots, \qquad (12)$$

where $\mathcal{C}^{\text{CC}}$ is the matching coefficient, analytically computable through perturbative QCD [46,51,56,103], which is taken up to NLO in this work. As is the case with the Rp-ITD, the threshold resummation for $\mathcal{C}^{\text{CC}}$ is less important over the whole range of $0 \le x \le 1$ than for experimental observables. The leading twist term of Eq. (12) is very similar to the leading twist term in Eq. (8), where $\mathcal{C}^{\text{CC}}$ is an implied integral quantity involving a cosine transformation [51].

The systematic effects added to the leading twist term in Eq. (12) are the power correction $z^2 B_1(\nu)$ and discretization correction $a R_1(\nu)$, with the ellipsis representing other potential systematic correction terms. Attempts were made to fit systematic corrections such as lattice spacing, pion mass, and finite volume, but because the data have such large statistical uncertainties, it was difficult to separate the systematic effects beyond what is included in Eq. (12). We parametrize function $B_1(\nu)$ as in Eq. (9), and the discretization term $R_1(\nu)$ is similarly parametrized as

$$R_1(\nu) = \sum_n \sigma_{0,n}(\nu) r_n, \qquad (13)$$

where $\sigma_{0,n}$ is given in Eq. (10). Unlike in the Rp-ITD case, here we expand Eqs. (9) and (13) in n from 0 to 2 because the quantity $T_1(\nu, z^2)$ does not have a specified normalization given by the leading twist term at $\nu = 0$.

Four datasets were included in our analysis of the HadStruc data [51], with the parameters given in Table I. We refer to the datasets as {a127m413, a127m413L, a94m358, a94m278}, which have lattice spacings of {0.127, 0.127, 0.094, 0.094} fm and pion masses of {413, 413, 358, 278} MeV, respectively. Each ensemble has a total lattice size of $L = 32a$, with the exception of the a127m413 set, which has $L = 24a$. Each dataset spans the range $2a \leq z \leq 4a$, and the a94m358 and a94m278 ensembles further include $z = 5a$ and $6a$. The momentum of the generated lattice data include $p_i = i \times (2\pi/La)$, where $i \in \{1, 2, 3, 4\}$. No kinematic cuts were made on the CC observables.

C. Scale setting

In typical high energy scattering experiments, such as DY lepton-pair production or deep-inelastic LN electro-production, the invariant mass of the virtual photon $Q \equiv \sqrt{Q^2}$ is much greater than any nonperturbative scale, rendering the power correction terms in factorized cross sections small. The renormalization and factorization scales relating the PDFs and hard coefficients to the experimental cross sections are generally set to this probing scale, $\mu = Q$, where asymptotic freedom is exploited and $\log(\mu^2/Q^2)$-type logarithms appearing in the perturbative expansions are suppressed.

The uncertainty in renormalized correlation functions of lattice calculated correlators is independent of how they are factorized into PDFs and the corresponding choice of renormalization and factorization scale, although it impacts the calculation of the perturbative matching coefficients. In the Rp-ITD case, the renormalization constants of lattice calculated ITDs, which have been calculated in Ref. [111], are multiplicative and cancel in the ratio in Eq. (7). For CC observables, the renormalization constants are explicit in Eq. (11), given by Z_V and Z_A. To factorize these renormalized lattice calculated correlators into PDFs convoluted with hard coefficients we need to specify the factorization and renormalization scales, akin to the scales that appear in the treatment of experimental observables. We choose to equate these factorization and renormalization scales, denoted by μ_{lat} in Eqs. (8) and (12). It is convenient to choose this scale to be proportional to the inverse of the separation between fields or currents in order to eliminate large logarithms [55].

It was shown in Ref. [49] that the power corrections of $\mathcal{O}(z^2\Lambda^2_{\text{QCD}})$ are much less than the leading twist component, and the lattice data are well behaved when $1/z < 1$ GeV. We fix μ_{lat} to be a constant across all kinematics, satisfying two requirements: (i) $\alpha_s(\mu_{\text{lat}})$ should
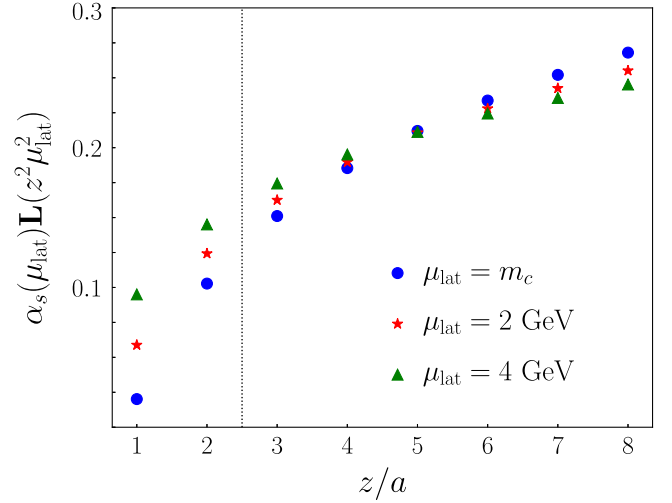

FIG. 1. Typical factor of $\mathcal{O}(\alpha_s)$ in the matching coefficients with explicit dependence on z and μ_{lat}, where $\mathbf{L}(z^2\mu^2_{\text{lat}}) \equiv \log(\frac{1}{4}z^2\mu^2_{\text{lat}}e^{2\gamma_E+1})/2\pi$. Shown are the terms evaluated using 3 fixed scales of $\mu_{\text{lat}} = m_c$ (blue circles), $\mu_{\text{lat}} = 2$ GeV (red stars), and $\mu_{\text{lat}} = 4$ GeV (green triangles) as a function of z in units of lattice spacing $a = 0.127$ fm. The dotted line indicates the lower limit of z/a used in the analysis.

not be too large, and (ii) the product of $\alpha_s(\mu_{\text{lat}})$ and the logarithm appearing in the perturbatively calculable matching coefficients $\sim \log(z^2\mu^2_{\text{lat}})$ is not too large. We investigate these criteria for three scales: $\mu_{\text{lat}} = m_c = 1.27$ GeV, $\mu_{\text{lat}} = 2$ GeV, and $\mu_{\text{lat}} = 4$ GeV. The first requirement is satisfied by all three choices, as the scales provide $\alpha_s \simeq 0.37, 0.30,$ and 0.23, respectively, in the $\overline{\text{MS}}$ renormalization scheme at NLO.

The interplay between $\alpha_s(\mu_{\text{lat}})$ and $\log(z^2\mu^2_{\text{lat}})$ in criterion (ii) is illustrated in Fig. 1, where the term $\alpha_s(\mu_{\text{lat}})\mathbf{L}(z^2\mu^2_{\text{lat}})$ is shown as a function of z in units of $a = 0.127$ fm, with $\mathbf{L}(z^2\mu^2_{\text{lat}}) \equiv \log(\frac{1}{4}z^2\mu^2_{\text{lat}}e^{2\gamma_E+1})/2\pi$. The product of α_s and the logarithm $\mathbf{L}$ is well below unity, indicating that perturbation theory can be trusted at each of the choices of constant μ_{lat} for the kinematics provided. While the results in Fig. 1 are shown at $a = 0.127$ fm, similar types of logarithms that appear in $\mathcal{C}^{\text{CC}}$ for the CC correlators in the form $\mathbf{L}(z^2\mu^2) = \log(\frac{1}{4}z^2\mu^2e^{2\gamma_E})/2\pi$ for all values of a are also well below unity, with the largest value of $\alpha_s(\mu_{\text{lat}})\mathbf{L}(z^2\mu^2_{\text{lat}}) = 0.17$. In the remainder of this paper, we choose the scale $\mu_{\text{lat}} = 2$ GeV, unless otherwise noted.

D. Methodology

We note that each of the expressions in Eqs. (1), (4), (7), and (12) are in the form of a convolution of a PDF (or two PDFs in the DY case) and a hard coefficient function, $\mathcal{C}$. The observables in these equations do not provide a direct method for solving for the PDFs, whose analytic form is not known, on a point-by-point basis, and therefore the

extraction of PDFs from these equations constitutes an inverse problem.

Following the standard approach used in the literature, we parametrize the valence quark distribution q_v, sea quark distribution q_s, and the gluon distribution g at the initial scale μ_0 through the general template function,

$$f(x, \mu_0^2) = \frac{N_f x^{\alpha_f} (1-x)^{\beta_f} (1 + \gamma_f x^2)}{B(\alpha_f + 2, \beta_f + 1) + \gamma_f B(\alpha_f + 4, \beta_f + 1)}, \quad (14)$$

where $f (= q_v, q_s$ or g) labels the parton flavor. The input scale is chosen to be the charm quark mass evaluated at the scale of the charm mass in the $\overline{\text{MS}}$ scheme, $\mu_0 = m_c(m_c) = 1.27$ GeV. In the remainder of the text, we assume the scale and refer to $m_c(m_c)$ simply as m_c. We impose the valence quark number rule and the momentum sum rule, which fix the valence quark and sea quark normalizations, respectively. Even though the lattice data are not sensitive to the sea quark and gluon distributions, they may have an effect through the momentum sum rule and interplay with the effects from the DY and LN experimental data. In a previous analysis [13], it was found that more flexible parametrizations did not change significantly the resulting PDFs and the agreement with the data. Since the parameters γ_s and $\gamma_g = 0$ are not constrainable by the data, we set them both to zero. However, we allow γ_v for the valence quark PDF to be free. In total, there are eight PDF shape parameters to be fitted, along with the Λ cutoff parameter for the LN observables. The scale dependence of the PDFs is determined by solving the DGLAP equations using the zero-mass flavor number scheme evolved up to next-to-leading logarithmic accuracy.

As in previous JAM global QCD analyses of pion PDFs [12–14], we employ Bayesian statistics to sample the posterior distribution according to

$$\mathcal{P}(\boldsymbol{a}|\text{data}) \sim \mathcal{L}(\text{data}|\boldsymbol{a}) \pi(\boldsymbol{a}), \quad (15)$$

where $\pi(\boldsymbol{a})$ is the prior distribution, which sets the boundaries of the fit parameters $\boldsymbol{a}$, $\mathcal{L}$ is the likelihood function,

$$\mathcal{L}(\text{data}|\boldsymbol{a}) = \exp\left(-\frac{1}{2}\chi^2(\boldsymbol{a}, \text{data})\right), \quad (16)$$

which is a Gaussian with the argument χ^2. In practice, maximizing the likelihood function is done by minimizing the χ^2 function, which for a given experimental dataset e is defined by

$$\chi_e^2(\boldsymbol{a}, \text{data}) = \sum_i \left[\frac{d_i^e - \sum_k r_k^e \beta_{k,i}^e - t_i^e(\boldsymbol{a})/n_e}{\alpha_i^e} \right]^2$$
$$+ \left(\frac{1 - n_e}{\delta n_e}\right)^2 + \sum_k (r_k^e)^2, \quad (17)$$

where the sum is over i data points up to N_{dat}^e, with k types of correlated uncertainties. The data points are represented by d_i^e, while the parameter-dependent theory, $t_i^e(\boldsymbol{a})$, is divided by the fitted normalization parameter n_e for each experiment and added with the correlated shift, which includes the nuisance parameters r_k^e and point-to-point correlated uncertainties $\beta_{k,i}^e$. In the second term the δn_e is the overall normalization uncertainty reported by the experiment. The last two terms in Eq. (17) represent penalties such that the fit is disfavored if n_e is far from 1, or if the amount of shift needed on the theory, r_k^e, is large.

Lattice data points are correlated within the same ensemble λ, so the point-to-point statistical uncertainties are not treated in the same way as the uncertainties on the experimental data. Instead, we use the covariance matrix V and minimize

$$\chi_\lambda^2(\boldsymbol{a}, \text{data}) = (\boldsymbol{D}^\lambda - \boldsymbol{T}^\lambda(\boldsymbol{a}))^T V_\lambda^{-1}(\boldsymbol{D}^\lambda - \boldsymbol{T}^\lambda(\boldsymbol{a})), \quad (18)$$

where $\boldsymbol{D}^\lambda$ is the vector of data points of length N_{dat}^λ, and $\boldsymbol{T}^\lambda$ is the set of theory predictions corresponding to the data points. The covariance matrix V follows from the standard definition, $V_{ij} = \text{E}[(\mathcal{O}_i - \text{E}[\mathcal{O}_i])(\mathcal{O}_j - \text{E}[\mathcal{O}_j])]$, where E is the expectation value defined in Eq. (19) below. The total χ^2, as in Eq. (16), that governs the parameter determination in each minimization is $\chi^2 = \sum_e \chi_e^2 + \sum_\lambda \chi_\lambda^2$.

In our Monte Carlo analysis the χ^2 minimizations are executed N_{rep} number of times. Through data resampling for each minimization, Gaussian noise is added to the central value of the data point with 1σ uncertainty width. For χ^2 minimizations using Eq. (18), we diagonalize the covariance matrix and add Gaussian noise in each eigendirection with width given by the magnitude of the eigenvalue. In this way we populate statistics on the posterior distribution, and avoid local minima in parameter space. The resulting parameter sets $\boldsymbol{a}_j$, where j runs from 1 to the number of replicas, N_{rep}, are used to obtain expectation values E and variances V of observables $\mathcal{O}$, defined as

$$\text{E}[\mathcal{O}] = \frac{1}{N_{\text{rep}}} \sum_j \mathcal{O}(\boldsymbol{a}_j), \quad (19a)$$

$$\text{V}[\mathcal{O}] = \frac{1}{N_{\text{rep}}} \sum_j (\mathcal{O}(\boldsymbol{a}_j) - \text{E}[\mathcal{O}])^2. \quad (19b)$$

The agreement between data and theory is quantified using the reduced χ^2, which is defined here as

$$\bar{\chi}_e^2 = \frac{1}{N_{\text{dat}}^e} \sum_i \left[\frac{d_i^e - \text{E}[\sum_k r_k^e \beta_{k,i}^e + t_i^e(\boldsymbol{a})/n_e]}{\alpha_i^e} \right]^2, \quad (20a)$$

$$\bar{\chi}_\lambda^2 = \frac{1}{N_{\text{dat}}^\lambda} (\boldsymbol{D}^\lambda - \text{E}[\boldsymbol{T}^\lambda(\boldsymbol{a})])^T V_\lambda^{-1}(\boldsymbol{D}^\lambda - \text{E}[\boldsymbol{T}^\lambda(\boldsymbol{a})]), \quad (20b)$$

or, generally, $\bar{\chi}^2$. For our analysis, an acceptable $\bar{\chi}^2$ value for each dataset is around 1. The total reported reduced χ^2 is given by $\bar{\chi}^2_{\text{tot}} = (\sum_e \bar{\chi}^2_e N^e_{\text{dat}} + \sum_\lambda \bar{\chi}^2_\lambda N^\lambda_{\text{dat}})/N^{\text{tot}}_{\text{dat}}$, where $N^{\text{tot}}_{\text{dat}}$ is the grand total number of all data points.

III. QCD ANALYSIS RESULTS

While most previous analyses of pion PDFs performed fits to DY data with an NLO expansion for the hard coefficients, several more recent analyses utilized the NLO + NLL approximation with threshold resummation. Both of these approaches give good descriptions of the DY and LN data [14], so that it is difficult to assess the applicability of factorization theorems if only one observable is present, since changes in the hard factors can generally be accommodated by changes in the PDFs. For this reason we present results with NLO calculations for the hard coefficients (referred to as NLO), as well as an NLO + NLL calculation for DY coupled with NLO calculations for the remaining hard coefficients (referred to as "NLO + NLL$_{\text{DY}}$"). Including lattice data in the mix may provide observables that allow us to test the degree of universality for PDFs across experimental and lattice data when using either fixed order or threshold resummation hard coefficients in DY. Lattice data, as we shall see, have sizable systematic effects that currently prevent us reaching definite conclusions about the applicability of threshold correction for DY at large x.

A. Analysis with reduced Ioffe time pseudo-distributions

As baseline fits for our analysis, we consider first as "scenario A" the analysis of DY and LN data determined using the NLO or NLO + NLL$_{\text{DY}}$ methods for the DY cross sections [12,14], without inclusion of lattice data.

A good overall fit to the DY and LN data is obtained, similar to that in the recent JAM analysis [14]. As illustrated in Table II, which summarizes the goodness of fit for each scenario, the number of points in each dataset and the corresponding reduced χ^2, a total reduced $\bar{\chi}^2$ of ≈ 0.8 is found for both the NLO and NLO + NLL$_{\text{DY}}$ baseline fits to the experimental data.

In a first attempt to fit the Rp-ITD data simultaneously with the experimental data, in "scenario B" we include the lattice data in a simplified way, excluding the systematic effects and only fitting to the leading twist terms in Eq. (8). While the DY and LN data were described almost the same as the fits to only experimental data, we found the fit to the smaller lattice volume data, a127m413, was unacceptably large, with a $\bar{\chi}^2$ of 1.98 and 2.63 for the NLO and NLO + NLL$_{\text{DY}}$ methods, respectively. This lack of agreement with the data is an indication that the leading twist convolution term alone is not sufficient to describe the available Rp-ITD data. Consequently, we consider a further "scenario C," which includes all of the available types of systematic terms as shown in Eq. (8).

With the inclusion of the systematic corrections in scenario C, the resulting $\bar{\chi}^2$ are acceptable for all datasets, with a global $\bar{\chi}^2 \approx 0.85$. Importantly, the $\bar{\chi}^2_e$ values do not change much among the scenarios, indicating compatibility of the experimental and lattice data. Interestingly, the agreement of the results with the large lattice volume data, a127m413L, does not change with the inclusion of systematic corrections, and the effect is largely seen in the smaller lattice volume ensemble, a127m413. By performing Monte Carlo inference of all PDF shape parameters and lattice systematic parameters, in the following we investigate the relationship between the PDFs and the systematic uncertainties, including their overall contributions.

TABLE II. Summary of results from global fits to the DY cross sections, LN electroproduction data from HERA, and the RpITD data, including the number of data points N_{dat} and the $\bar{\chi}^2$ values for the NLO and NLO + NLL$_{\text{DY}}$ methods. The Rp-ITD data were fitted at the scale $\mu_{\text{lat}} = 2$ GeV. Scenario A represents the fit to only experimental data, scenario B is fitting both experimental and lattice data with only the leading twist term, and scenario C (in boldface) is the full fit including systematic corrections.

| | | | Scenario A | | Scenario B | | Scenario C | |
| | | | NLO | +NLL$_{\text{DY}}$ | NLO | +NLL$_{\text{DY}}$ | NLO | +NLL$_{\text{DY}}$ |
Process	Experiment	N_{dat}	$\bar{\chi}^2$		$\bar{\chi}^2$		$\bar{\chi}^2$	
DY	E615	61	0.84	0.82	0.83	0.82	**0.84**	**0.82**
	NA10 (194 GeV)	36	0.53	0.53	0.52	0.54	**0.51**	**0.53**
	NA10 (286 GeV)	20	0.80	0.81	0.78	0.79	**0.74**	**0.81**
LN	H1	58	0.36	0.35	0.39	0.39	**0.38**	**0.37**
	ZEUS	50	1.56	1.48	1.62	1.69	**1.59**	**1.62**
Rp-ITD	a127m413L	18	⋯	⋯	1.04	1.06	**1.05**	**1.04**
	a127m413	8	⋯	⋯	1.98	2.63	**1.00**	**1.18**
Total		251	0.82	0.80	0.89	0.92	**0.85**	**0.86**

TABLE III. Summary of Z-sigma levels for the χ^2 distributions for each of the datasets. The absolute σ value by which the peak of the resulting Monte Carlo χ^2 distribution is less than ($^-$) or greater than ($^+$) the expected χ^2 distribution is shown in each entry.

| Process | Experiment | NLO | NLO + NLL$_{DY}$ |
		Z-sigma level	Z-sigma level
DY	E615	0.75^-	0.79^-
	NA10 (194 GeV)	2.32^-	2.26^-
	NA10 (286 GeV)	0.82^-	0.78^-
LN	H1	4.15^-	4.22^-
	ZEUS	2.72^+	2.73^+
Rp-ITD	a127m413L	0.30^+	0.35^+
	a127m413	0.40^+	0.60^+

To assess the significance of the $\bar{\chi}^2$ values in Table II, we perform the Z-sigma level statistical test on each of the datasets analyzed in the global analysis. In this test, the null hypothesis is the expected χ^2 distribution assuming the number of degrees of freedom to be the number of points in the data set, and the resulting χ^2 samples from the Monte Carlo (MC) analyses from scenario C are the alternative hypotheses. We compute the p value using the peak of the resulting MC χ^2 distributions, and the Z-sigma level is the inverse of the normal cumulative distribution function, $Z = \Phi^{-1}(p) \equiv \sqrt{2}\mathrm{erf}^{-1}(2p - 1)$. These values are given in Table III for each of the datasets and represent the number of normal standard deviations from the expected χ^2. The "$-$" or "$+$" superscript on the values indicates their positions below or above the expected χ^2 distribution's mean value, respectively. A large negative value may imply a non-Gaussian χ^2 function is needed to treat the systematic uncertainties. Nevertheless, none of the values are outside of 5σ, indicating a reasonable probability that our results are achieved with the given data. Notably, the NLO and NLO + NLL$_{DY}$ are not in tension for the Rp-ITD data.

The distributions of the free parameters from scenario C are displayed in Fig. 2, which shows the number of Monte Carlo samples with a best fit value in each bin, up to an arbitrary normalization, for both the NLO and NLO + NLL$_{DY}$ methods. The first three rows indicate the PDF shape parameters, as well as the LN cutoff parameter Λ, while the bottom two rows show the distribution of the systematic parameters b_i, p_i and f_i ($i = 1, 2$). Note that the β_v parameter shown in the first row is not the same as the β_{eff} that is used to describe the large-x behavior of the valence quark distribution (see below). In each panel a distinct peak is observed, indicating that the data prefer a certain value and assuring each parameter has settled in a global maximum in the likelihood.

The power correction parameters b_1 and b_2 both have generally positive values, with the b_1 parameter smaller and

peaking slightly above zero, while the b_2 parameter peaks at around 0.1. The lattice spacing error parameters p_1 and p_2 have opposite sign, with the p_1 parameter roughly twice as large as p_2 in magnitude. The finite volume parameters are seen to have the most narrow distributions and have the smallest absolute values. Both f_1 and f_2 are close to zero, with f_1 tending slightly negative and f_2 slightly positive. We discuss the systematic corrections in more detail in Sec. III A 3.

In Fig. 3 we show the resulting predictions of the Rp-ITD from the analysis, along with the leading twist contributions, with the difference between the bands reflecting the contribution from the systematic corrections. The panels illustrate the results using the NLO and NLO + NLL$_{DY}$ methods, as well as for the larger volume ($L = 32a$, a127m413L) and smaller volume ($L = 24a$, a127m413) lattice ensembles. Each separated spectrum represents different values of z, offset for clarity. For small values of ν and z, the difference between the total and leading twist bands is effectively zero, which indicates that the systematic corrections do not play a role there, and that the leading power contribution dominates the total lattice observable. As ν and z increase, the systematic contributions grow in magnitude, particularly for the NLO + NLL$_{DY}$ case, and are negative, as indicated by the separation between the total and leading twist bands. In the case of the NLO method, the bands are somewhat separated, but not as pronounced as for the NLO + NLL$_{DY}$ case, implying different systematic corrections in the two analyses. Because of the difference between the total and leading twist contributions, excluding the systematic corrections would lead to incorrectly extracted PDFs, as the leading twist band with the PDF would have to shift to agree with the data. The systematic correction terms in Eq. (8) must therefore be included in the fit.

The x dependence of the fitted valence quark, sea quark, and gluon distributions, along with the ratios of their uncertainties relative to their mean values, are shown in Fig. 4 at the input scale $\mu = m_c$ from scenarios A, B and C, for both the NLO and NLO + NLL$_{DY}$ cases. When using the NLO hard coefficient in DY, the central values of the valence and sea quark distributions are mostly unaffected by the lattice data. There is a slight change in the gluon distribution, whereby the inclusion of lattice data decreases its magnitude for $x \gtrsim 0.2$. However, the general agreement among the distributions indicates that the lattice and experimental data are compatible.

For the case of the NLO + NLL$_{DY}$ extractions, none of the scenarios are found to match well with each other, suggesting some instability of the PDFs with the inclusion of the lattice data. The experimental data prefer a valence quark distribution with a slightly smaller magnitude at intermediate x. When the lattice data are included, the PDF increases by $\sim 30\%$ in the range $0.2 \lesssim x \lesssim 0.7$. When including the systematic corrections, on the other hand,
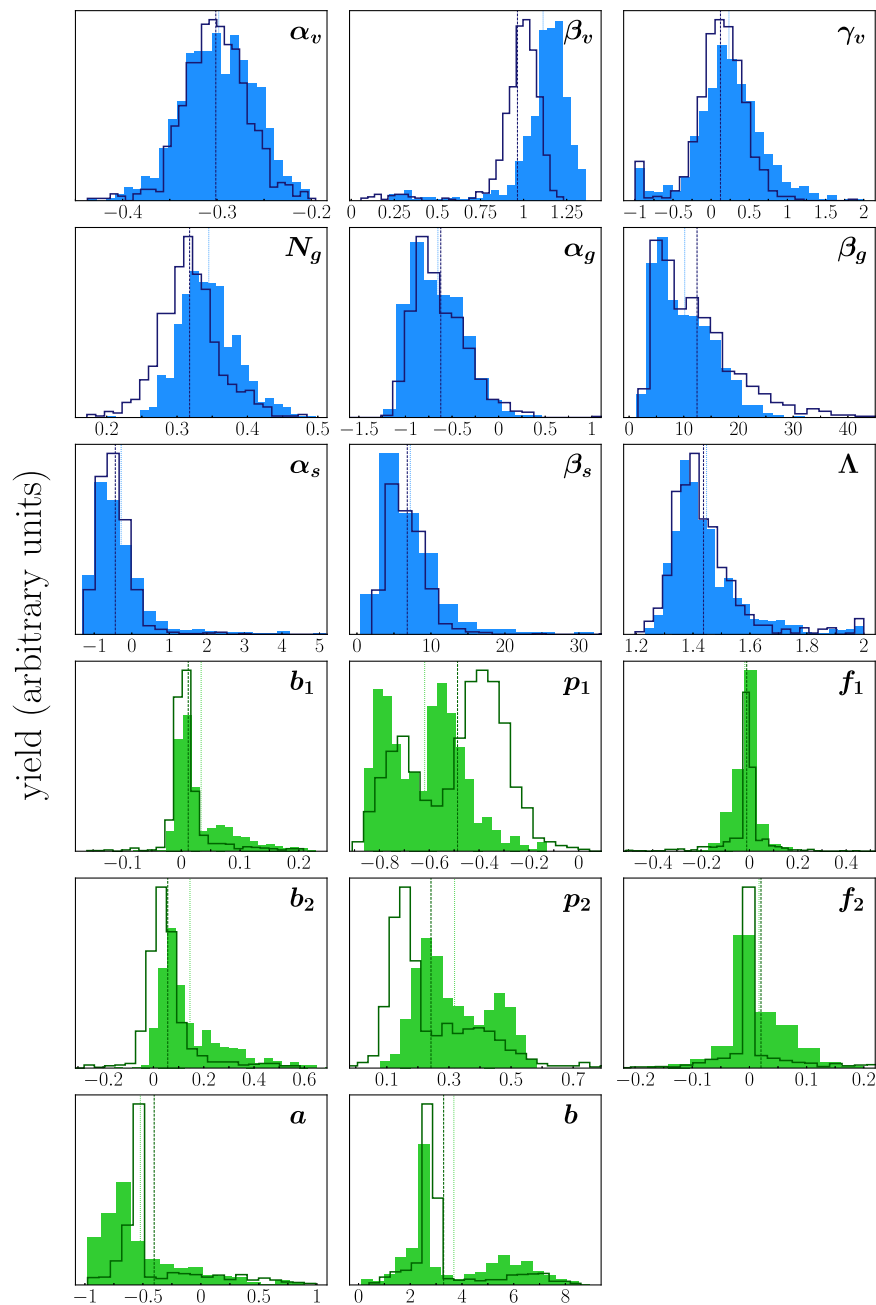
FIG. 2. Distributions of the free parameters describing PDF shapes $\{\alpha_v, \beta_v, \gamma_v; N_g, \alpha_g, \beta_g; \alpha_s, \beta_s\}$ and the ultraviolet cutoff mass Λ in the $N \to \pi N$ splitting function (blue histograms), and the intrinsic lattice systematic parameters $\{b_1, p_1, f_1; b_2, p_2, f_2; a, b\}$ (green histograms) from the Monte Carlo fits, in arbitrary units. Both the NLO (outlined histograms) and NLO + $\mathrm{NLL_{DY}}$ (filled histograms) analyses are shown, with the vertical dotted lines giving the mean values of the parameters.

the PDF shifts downwards, but still mostly does not overlap with the experimental-only results. The large-x sea quark and gluon distributions are suppressed with the inclusion of the lattice data because of indirect constraints from the momentum sum rule. Despite the differences of the PDFs among the scenarios, the description of the experimental data remains unchanged, as indicated in Table II.

The PDFs extracted from only the experimental data carry large uncertainties, especially in the NLO + $\mathrm{NLL_{DY}}$ case, and including the precise lattice data decreases the uncertainty significantly. However, including the systematic corrections again increases the uncertainty of the PDFs, because of the increase in the number of parameters, but nevertheless provides a sizable impact. The behavior of the relative uncertainty in the gluon distribution across the scenarios is opposite to that for the quark distributions, which can be attributed to the redistribution among the parton flavors across the scenarios.
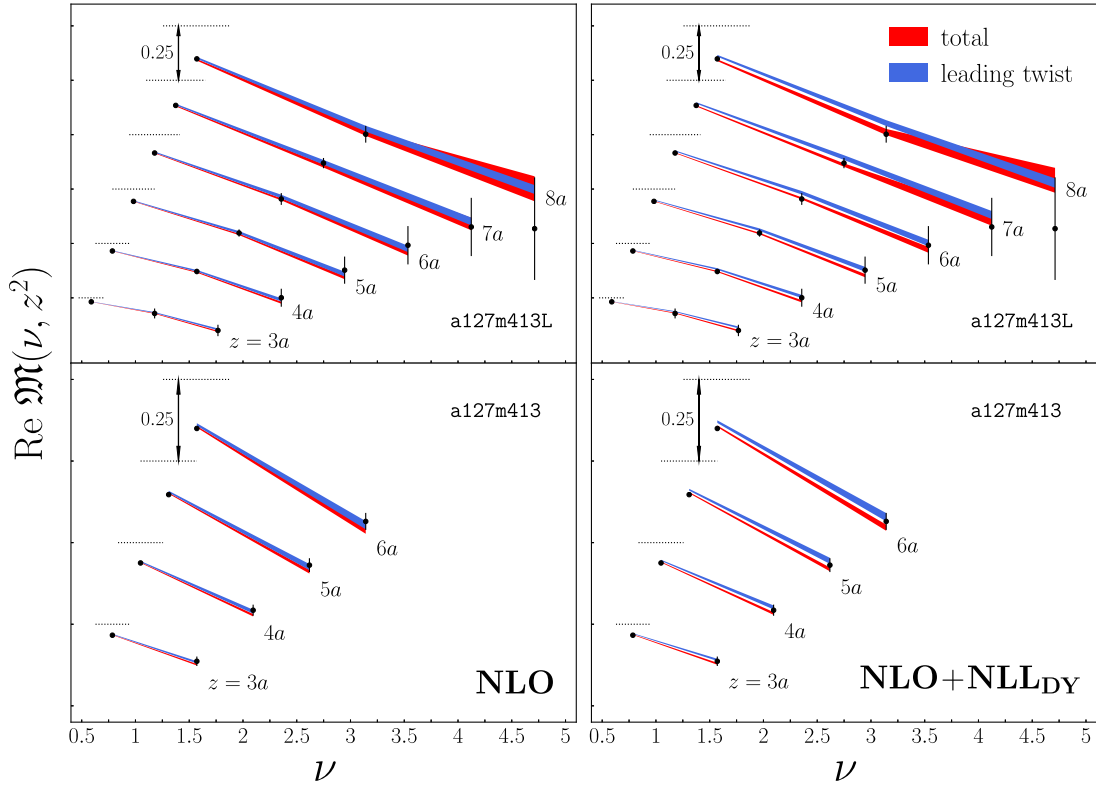
FIG. 3.   Comparison for the fitted reduced Ioffe time pseudo-distributions using NLO (left) and NLO + NLL$_{\text{DY}}$ (right) methods with the larger volume (`a127m413L`, top) and smaller volume (`a127m413`, bottom) lattices (black circles). The total result of the Bayesian inference (1σ red bands) is compared with the leading twist contribution without systematic corrections (1σ blue bands). The horizontal dotted lines for each bin in z indicate 1, the normalization value at $\nu = 0$.

The effective β_v parameter describes the degree of falloff at large x in the valence quark distribution, and operationally we define [112–114]

$$\beta_{\text{eff}}(x, \mu) = \frac{\partial \log |q_v(x, \mu)|}{\partial \log(1 - x)} \qquad (21)$$

at the scale μ. To obtain the PDF when extracting from lattice data alone, precise data over a large range of ν is needed. Joó *et al.* [49] found $\beta_{\text{eff}} \sim 1$, but with a large uncertainty, because of the limited range of ν. The recent analysis of experimental data in Ref. [14] found $\beta_{\text{eff}} \sim 1$ with NLO hard coefficients, and $\beta_{\text{eff}} \sim 1.2$ when using NLO + NLL$_{\text{DY}}$ with double Mellin threshold resummation on the hard coefficients in DY.

In the present analysis, we include the Rp-ITD lattice data and demonstrate in Fig. 5 that the β_{eff} resulting from each method of the short distance DY coefficients agrees, within uncertainties, with the β_{eff} from using experimental data alone [14]. While β_{eff} does not change at all in the NLO case, it is more precisely determined with the inclusion of lattice data for the NLO + NLL$_{\text{DY}}$ calculation, as evident from the shrinking of the uncertainty band relative to the experimental only band. In fact, the β_{eff} of the NLO + NLL$_{\text{DY}}$ analysis is more centered around the lower

portion of the previous band, decreasing slightly its mean value and bringing it closer to $\beta_{\text{eff}} = 1$. The difference in uncertainty reduction in the two cases of the hard coefficients can be attributed to the more dramatic decrease in relative uncertainty of the valence quark distribution in the NLO + NLL$_{\text{DY}}$ method at large x compared with the NLO case, shown in Fig. 4.

1. Which lattice data have maximal impact?

When studying the impact of adding lattice data to global phenomenological extractions of PDFs, it is pertinent to ask which lattice datasets have the most impact. In lattice calculations, all hadrons suffer from an exponential growth of the signal-to-noise ratio as energy increases, but the effect on the pion, as the lightest hadron, is most significant. This signal-to-noise ratio of lattice correlation functions decays exponentially as $\sim \exp[-(E_h - \frac{1}{2}m_\pi n_q)T]$, where E_h is the energy of the hadron, n_q is the number of valence quarks, and T is the Euclidean time separation of the operators [109,115]. Excited state contamination is lessened at large time separations, where the pion has a larger signal-to-noise ratio than the nucleon at a fixed low momentum. In the datasets used here, the relative uncertainty of the lowest momentum state data having $p = 1$
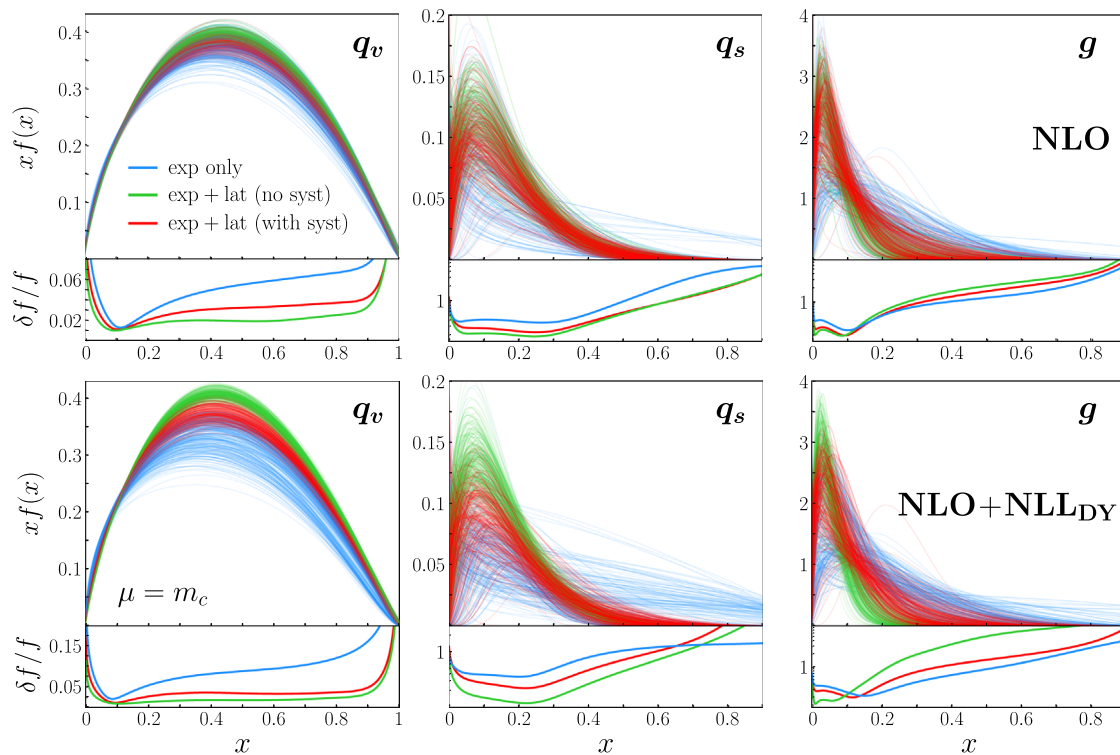
FIG. 4. Valence quark (left), sea quark (middle), and gluon (right) distributions with the 1σ relative uncertainties (underneath each panel) for the NLO (top) and NLO + NLL$_{DY}$ (bottom) methods. All three scenarios are displayed: Extractions from experimental data alone (scenario A, blue curves), from experimental and lattice data without systematic corrections (scenario B, green curves), and from including both experimental and lattice data with systematic corrections (scenario C, red curves). A random subset of 300 of the $\sim$700 total Monte Carlo replicas is shown.

(hereafter referred in units of $2\pi/La$) is an order of magnitude smaller than the $p = 2$ data, even though all momenta were calculated with equal computational cost. As such, we expect the lowest momentum state to drive the impact from the lattice data overall. This feature

also means that utilizing factorization methods that limit power corrections at low momentum is crucial for efficient and precise numerical calculations of parton structure, particularly for the pion, where the effects can be rather dramatic.
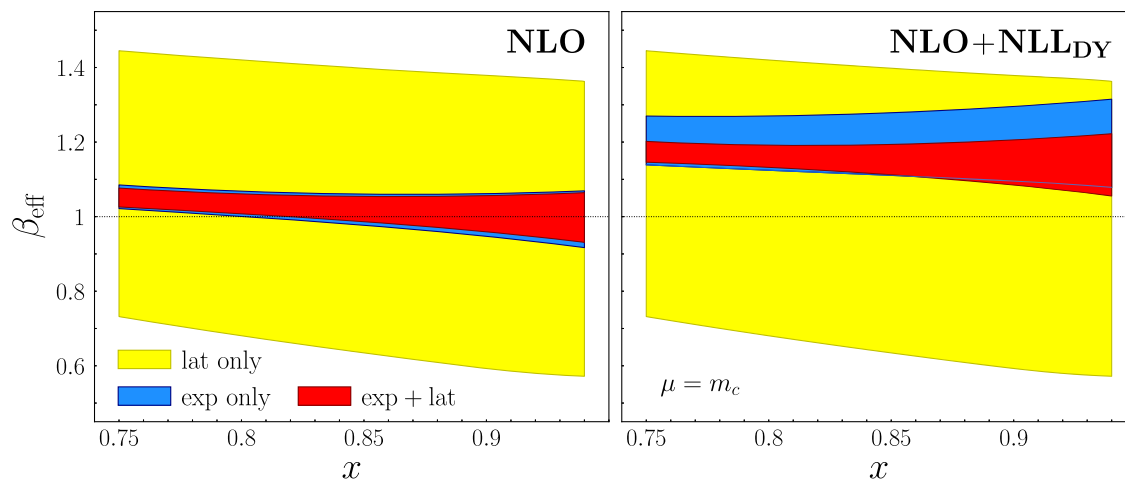
FIG. 5. Effective large-x exponent β_{eff} for the valence quark distribution as a function of x at the input scale $\mu = m_c$ extracted from lattice data alone (yellow bands), experimental data alone (blue bands), and both lattice and experimental data (red bands) from the NLO (left) and NLO + NLL$_{DY}$ (right) methods.

As demonstrated with mock data in Ref. [116], the range of Ioffe time can dramatically change the quality of the resulting PDF, specifically in the low-x region. However, in this combined analysis, the LN data provide constraints at low-x such that an improvement is possible across the whole range of x.

For each value of z, the smallest ν values correspond to the $p = 1$ points. As seen in Fig. 3, the leading twist terms generally agree the best with the total results when ν is small, so that the systematic corrections do not compete with the leading twist contribution. We performed a Monte Carlo analysis of the experimental data combined with the $p = 1$ lattice points with the same systematic corrections included above, again using both NLO and $\mathrm{NLO} + \mathrm{NLL_{DY}}$ coefficients. These analyses included 6 and 4 data points from the `a127m413L` and `a127m413` ensembles, respectively. The resulting $\bar{\chi}^2$ are almost identical to scenario C in Table II, with the NLO analysis producing $\bar{\chi}^2_\lambda = 0.80$ and 1.02 for `a127m413L` and `a127m413`, respectively, and the $\mathrm{NLO} + \mathrm{NLL_{DY}}$ giving $\bar{\chi}^2_\lambda = 0.82$ and 1.05 for the same datasets.

The ratios of the $p = 1$ data to the mean values of the theory are shown in Fig. 6 for each of the Rp-ITD datasets. Since all data points have the same momentum p, the increase in ν is directly proportional to the increase in the z spatial separation. Each of these points is close to unity within the uncertainties of the theory, indicating a good agreement in the analysis. The datasets agree best with the theory at small ν, and deviate in opposite directions at large ν. The theory slightly underpredicts the `a127m413L` data, while marginally overpredicting the `a127m413` data. The difference between NLO and $\mathrm{NLO} + \mathrm{NLL_{DY}}$ is minimal, with a modest preference for the NLO analysis from the `a127m413` dataset, evident at large ν.

In comparison with Fig. 4, which showed the PDFs when taking into account all of the momentum values in `a127m413L` and $p = 1, 2$ in `a127m413`, in Fig. 7 we illustrate the resulting valence quark distribution when including only the $p = 1$ lattice data points. The PDFs inferred from the combined analysis that includes all the lattice data and their systematic corrections are almost indistinguishable from those inferred from the same combined analysis but only including the $p = 1$ lattice data points. As expected, the relative uncertainties decrease when including more lattice data points, but only marginally. The results indicate that the $p = 1$ points drive the analysis from the lattice data and provide the strongest constraints on the PDFs. Despite the limited Ioffe time range of the lattice data, there is clearly merit in performing extractions from the lowest momentum values, as these are the most precise and produce the same qualitative and quantitative results as the analysis with all momenta.

2. Scale variation

Following Sec. II C, we investigate here in detail the effects of varying the scale, μ_{lat}. In addition to the main analysis, which uses $\mu_{\mathrm{lat}} = 2$ GeV, we performed two further combined analyses, varying the scales to $\mu_{\mathrm{lat}} = m_c$ and $\mu_{\mathrm{lat}} = 4$ GeV, using all momenta from the lattice data and the same systematic terms as in Eq. (8). The central values of the valence quark distributions extracted using the input scales $\mu_{\mathrm{lat}} = m_c$ and $\mu_{\mathrm{lat}} = 4$ GeV are shown in Fig. 8 relative to the mean value of the $\mu_{\mathrm{lat}} = 2$ GeV result, for both of the calculational methods for the short-distance DY coefficients.

We find that the $\mu_{\mathrm{lat}} = m_c$ and $\mu_{\mathrm{lat}} = 4$ GeV versions of the analysis slightly deviate from the $\mu_{\mathrm{lat}} = 2$ GeV results, but only by $\lesssim 2\%$ for most of the accessible x range in each case. Importantly, each of the central values lie within the
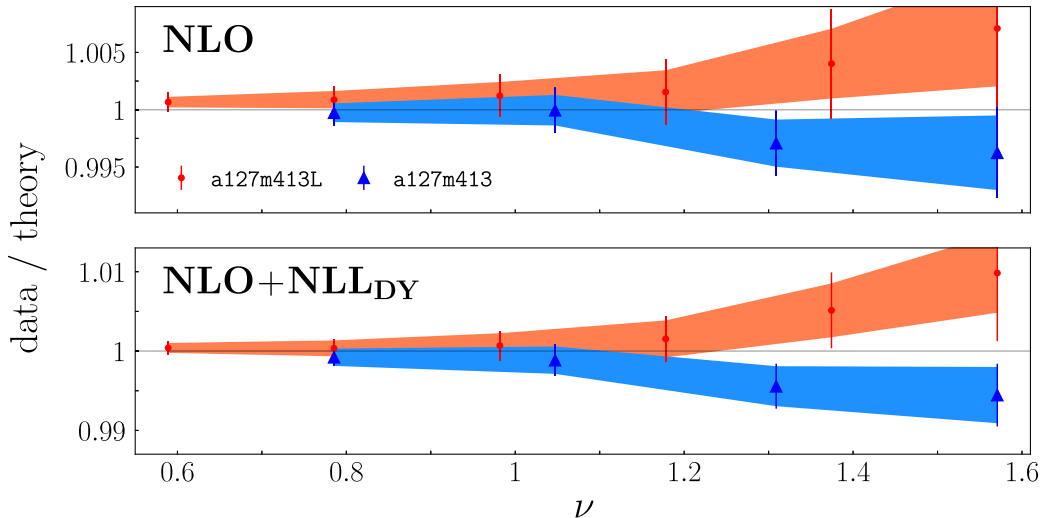

FIG. 6.   Data to theory ratios for the Rp-ITD datasets `a127m413L` (red circles) and `a127m413` (blue triangles) resulting from fits including only the $p = 1$ data points for the NLO (top) and $\mathrm{NLO} + \mathrm{NLL_{DY}}$ (bottom) methods. The bands represent 1σ uncertainties.
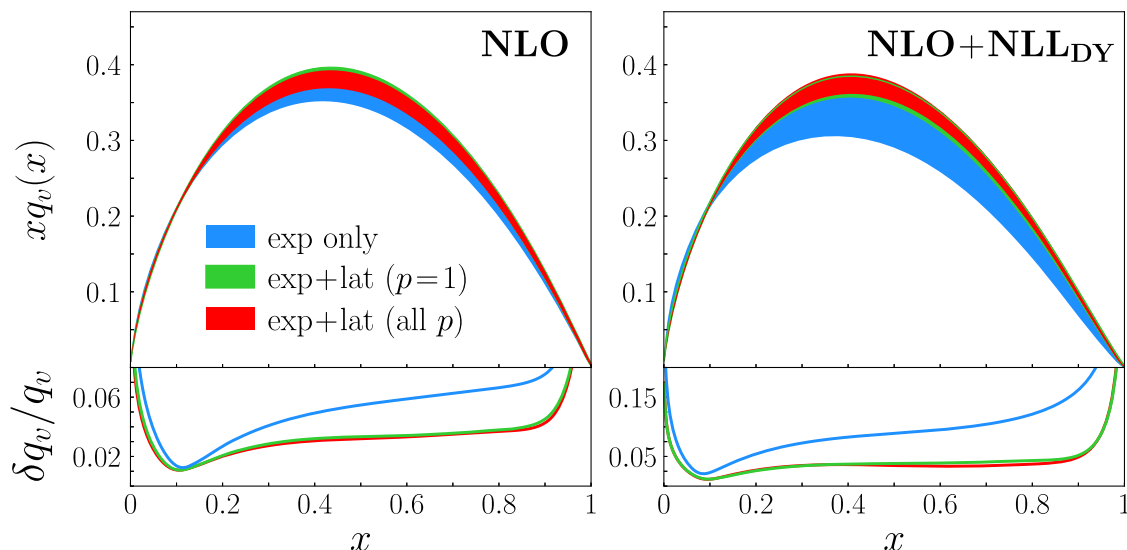
FIG. 7. Valence quark distributions (top) when extracted from experimental data alone (blue), combined with the $p = 1$ lattice data (green), and combined with all the lattice data (red) for the NLO (left) and NLO + NLL$_{DY}$ (right) cases, along with the relative uncertainties (bottom). The bands represent a 1σ uncertainty level.
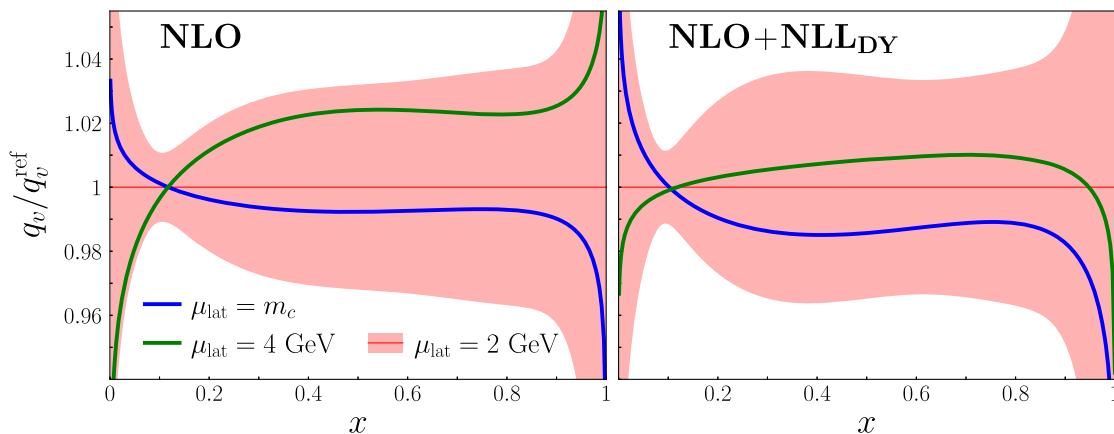

FIG. 8. Central values of the extracted pion valence quark distribution at the input scale from fits setting $\mu_{lat} = m_c$ (blue lines), $\mu_{lat} = 4$ GeV (green lines) relative to q_v^{ref}, which is the central value of the $\mu_{lat} = 2$ GeV main result (red lines and 1σ pink uncertainty bands) for the NLO (left) and NLO + NLL$_{DY}$ (right) cases.

1σ uncertainty band from the $\mu_{lat} = 2$ GeV result. This suggests that the uncertainty band associated with the valence quark PDF has not been underestimated due to scale variation effects, and demonstrates that these effects for the Rp-ITD are largely insignificant.

3. Quantification of Rp-ITD systematics

An important aspect of our analysis is the ability to use the lattice QCD data to provide information on both the pion PDFs and the systematic uncertainties on the lattice calculations. The leading twist contributions include dependence on the PDFs already constrained by the experimental data, whereas the systematic corrections are *a priori* unknown. In Fig. 3, above, we showed that for

small ν, p, and z the leading twist contribution largely equaled the total theory prediction, indicating that the systematic corrections in Eq. (8) are small. The NLO method tended to give smaller deviation between the two curves, while the NLO + NLL$_{DY}$ implied a somewhat more negative systematic correction. In Fig. 9, we show the total systematic corrections associated with the lattice from extractions of both the NLO and NLO + NLL$_{DY}$ analyses. When using the NLO theory for the hard coefficients in the DY calculation, the systematic corrections agree well with those using the NLO + NLL$_{DY}$ theory in both magnitude and uncertainty.

There is considerable overlap between the two sets of bands, and the total systematic corrections are almost indistinguishable. New data are needed to futher discriminate
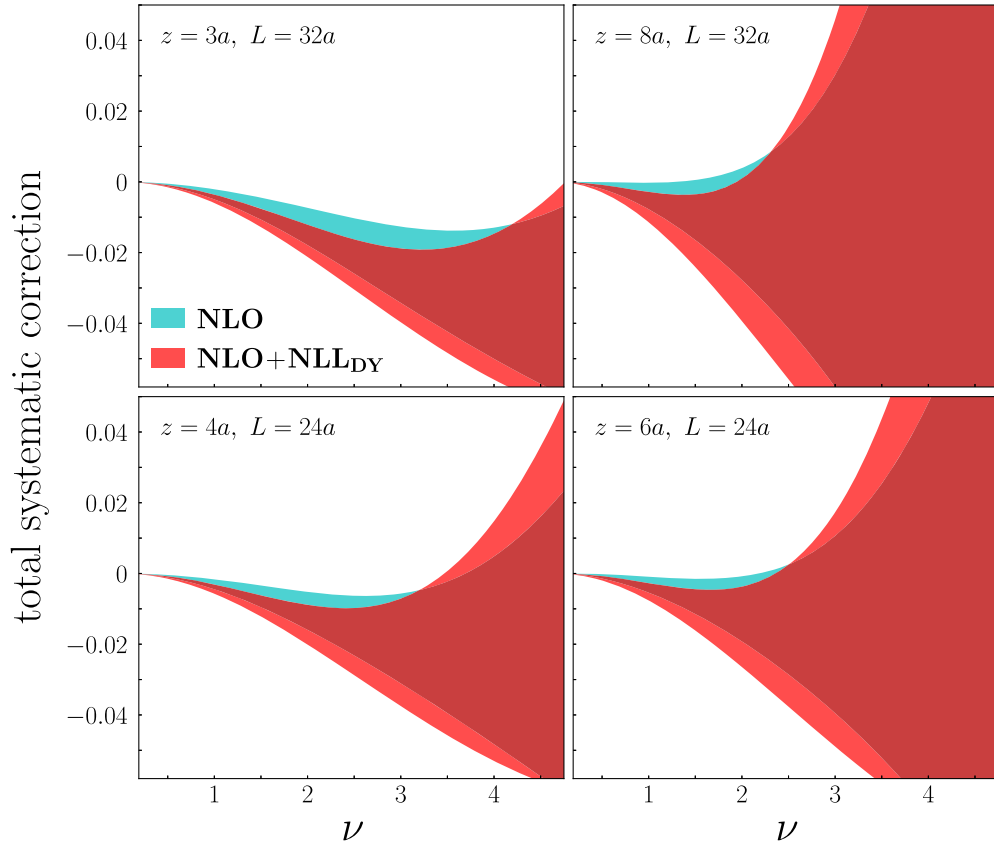
FIG. 9. Total systematic correction versus Ioffe time ν from the NLO (cyan bands) and NLO + NLL$_{DY}$ (red bands) extractions at several representative values of z and L in the available Rp-ITD datasets. The bands represent 1σ uncertainty levels.

between the methods and provide a more trustworthy description of the systematic corrections on these lattices. The general trend between the two methods is common: the systematic corrections are small when $\nu \approx 0$, and increase in magnitude and uncertainty with ν. The full systematic corrections as functions of ν tend to be negative with a downward slope up to a minimum, after which the systematic corrections increase. As z increases, the minimum appears at smaller ν. Additionally, both methods indicate the uncertainty of the systematic corrections increases with z. Avoiding the lattice systematic corrections and performing calculations that focus on regions in which the leading twist contribution dominates can effectively be done at small ν. The systematic uncertainty bands shown in Fig. 9 are either comparable or larger than the statistical uncertainties shown by the lattice data in Fig. 3. To further constrain the PDFs from these data, an improvement should be performed not only on the statistics, but also on the systematics, especially at larger values of ν.

In Fig. 10 we show replicas from the Monte Carlo analysis of the systematic corrections for the power, lattice spacing, and finite volume correction terms in Eq. (8). These corrections are shown for $z = 3a$ and $z = 6a$, for both lattice volumes, and for both calculations of the short distance coefficients in DY. At small z values, the power

correction terms are quite small and the total systematic terms are dominated by the lattice spacing errors. However, at large z, the power corrections terms play a much more important role, while the lattice spacing errors have a lesser impact on the total systematics. The finite volume corrections in both cases are effectively zero and provide no contribution to the overall systematic corrections. Even though we may test the sensitivity to finite volume corrections because of the different lattice volumes, the contribution is negligible compared with the power and spacing corrections, albeit slightly larger for the smaller lattice volume.

B. Current-current correlator analysis

In this section we present results obtained by combining CC correlators as in Eq. (11) with experimental data in our global analysis. The statistical uncertainties on the lattice data for CC correlators [51] are somewhat larger than for the Rp-ITD data, which suggests that the CC correlator data may not constrain the PDFs as strongly as do the Rp-ITD data. However, it is nevertheless important to quantify the extent to which the existing CC correlator data impact the extraction of the pion PDFs, and gauge what kind of uncertainties future lattice simulations should aim for.

For the CC correlator observable, we consider the χ_e^2 as in Eq. (17) for each χ^2 minimization and use $\bar{\chi}_e^2$ as in
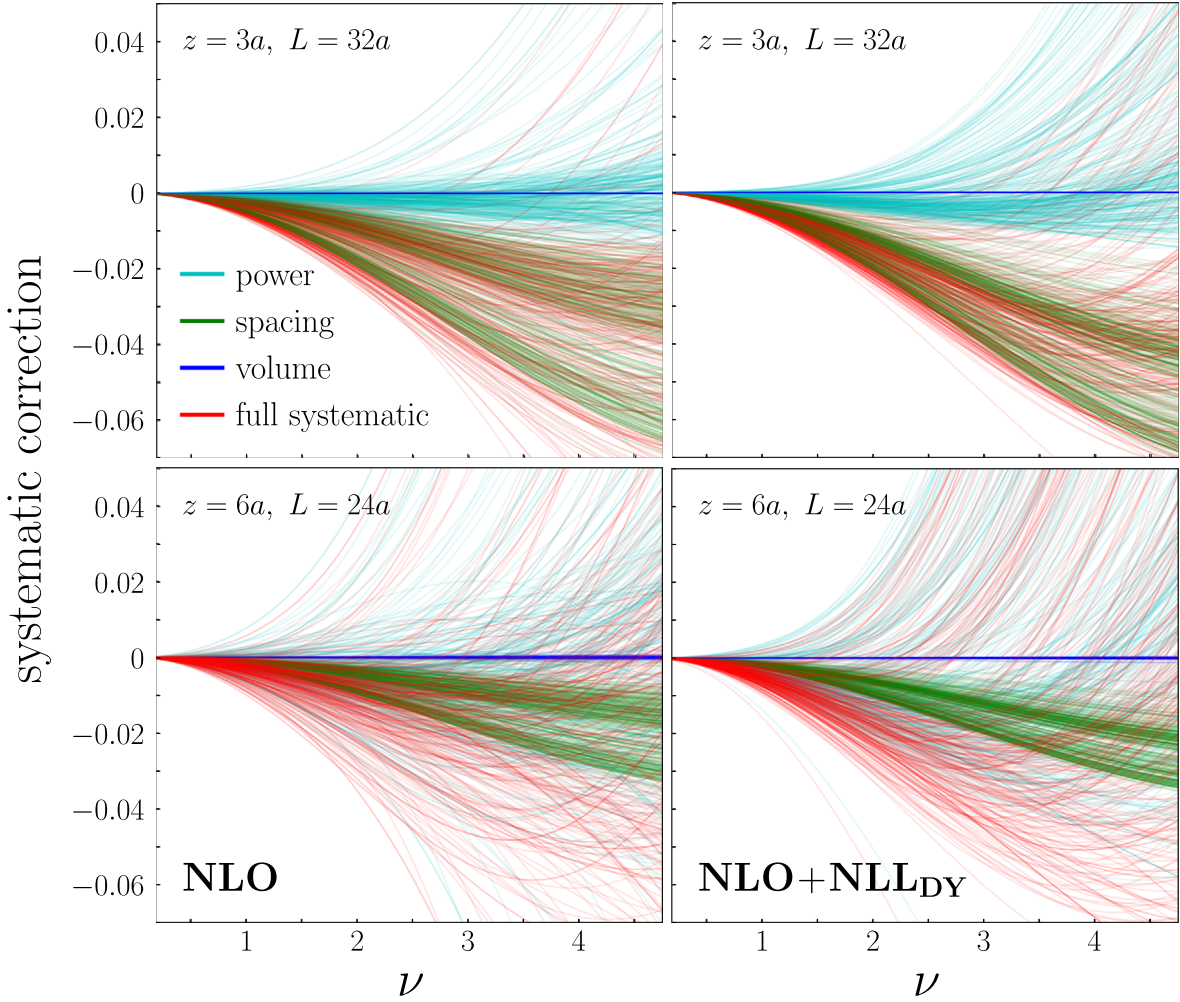
FIG. 10. Contributions of the power (cyan), lattice spacing (green), and finite volume (blue) systematic corrections as in Eq. (8), along with the sum (red), for the (top) `a127m413L` and (bottom) `a127m413` ensembles at various values of the lattice separation z for the NLO (left) and NLO + $\mathrm{NLL}_{\mathrm{DY}}$ (right) methods. Each panel is evaluated at $m_\pi = 413$ MeV and $a = 0.127$ fm, and for clarity a subset of 300 replicas is shown from a total of ≈ 720.

Eq. (20a) for analysis purposes. That is, we do not consider the covariance matrix and the correlation between individual uncertainties. We use the systematic correction terms associated with the power corrections and lattice spacing errors, as shown in Eq. (11). We found that by excluding all systematic effects, there was tension between the experimental and lattice data, evident in the resulting PDFs, which accepted a small set of solutions that were noticeably different from the experimental-only analysis. Additionally, including more systematic corrections increased the uncertainties on the Drell-Yan predictions generated from the experimental-only analysis. These inconsistencies led us to consider various systematic effects, and incorporating power and discretization corrections tamed both the uncertainties on the Drell-Yan predictions and tensions in the PDFs.

In Table IV we show the reduced χ^2 for each of the datasets in the analysis. The agreement of the resulting

TABLE IV. Summary of results for our global analysis of the DY and LN electroproduction data, along with the CC correlator lattice data, including the number of data points fitted N_{dat} and the $\bar{\chi}^2$ values for each of the calculations of the DY hard coefficient. The CC correlator data were fitted with $\mu_{\mathrm{lat}} = 2$ GeV.

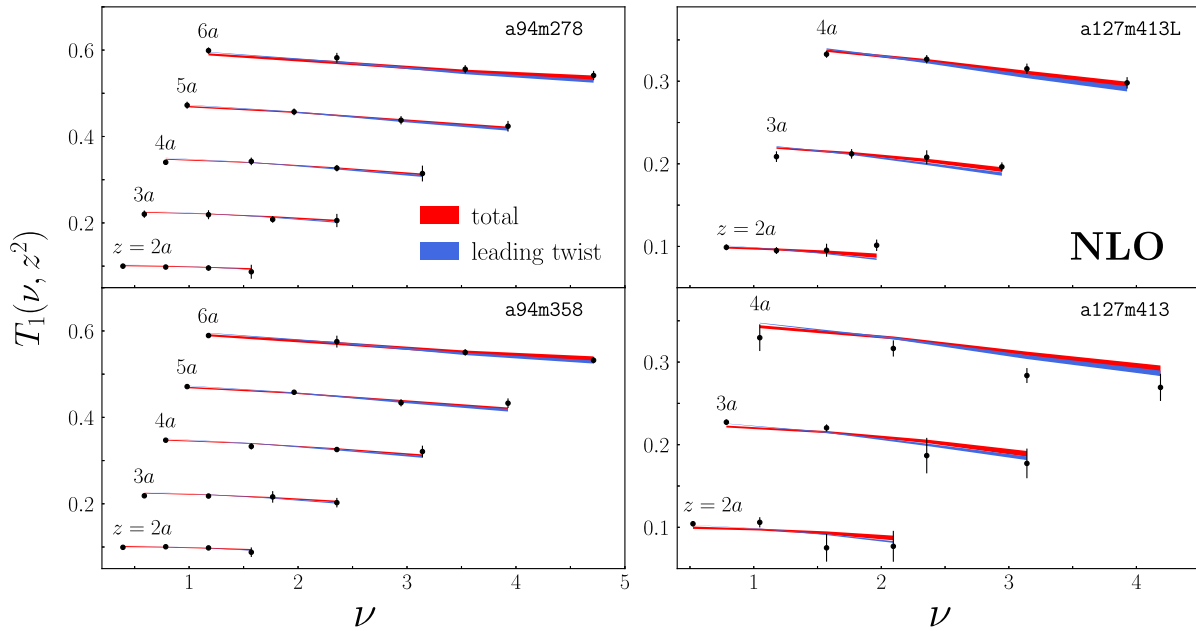|			NLO	NLO + $\mathrm{NLL}_{\mathrm{DY}}$
Process	Experiment	N_{dat}	$\bar{\chi}^2$	$\bar{\chi}^2$
DY	E615 (x_F, Q)	61	0.84	0.81
	NA10 (194 GeV) (x_F, Q)	36	0.53	0.54
	NA10 (286 GeV) (x_F, Q)	20	0.81	0.85
LN	H1	58	0.37	0.35
	ZEUS	50	1.55	1.54
CC	`a94m278`	20	0.33	0.33
	`a94m358`	20	0.47	0.46
	`a127m413L`	12	0.72	0.74
	`a127m413`	12	1.92	1.91
Total		289	0.81	0.80

FIG. 11.　CC correlator data versus Ioffe time ν for the datasets for the full theory as in Eq. (11) (red bands with 1σ uncertainties) and the leading twist contributions (blue bands) for the NLO method. An additive shift of $(z/a - 2)/8$ beyond $z = 2a$ is applied for clarity.

theory with the DY and LN data does not change significantly from the experimental-only results shown as scenario A in Table II. The overall $\bar{\chi}^2 = 0.81$ and 0.80 for the NLO and NLO + NLL$_{DY}$ methods, respectively, indicate a good agreement with the data as a whole, as well as with individual experimental and CC datasets, the only exception being the a127m413 ensemble that has $\bar{\chi}^2 \sim 1.9$ for each of the methods. Noticeably, the $\bar{\chi}^2$ values for each

method are very close, indicating that the CC data do not prefer one method over the other.

In Figs. 11 and 12 we compare the full theory and the leading twist contribution with the lattice data for the NLO and NLO + NLL$_{DY}$ methods, respectively. For this observable, the uncertainties are large in comparison with the Rp-ITD data, especially at small momentum. While the leading twist and total contributions have considerable
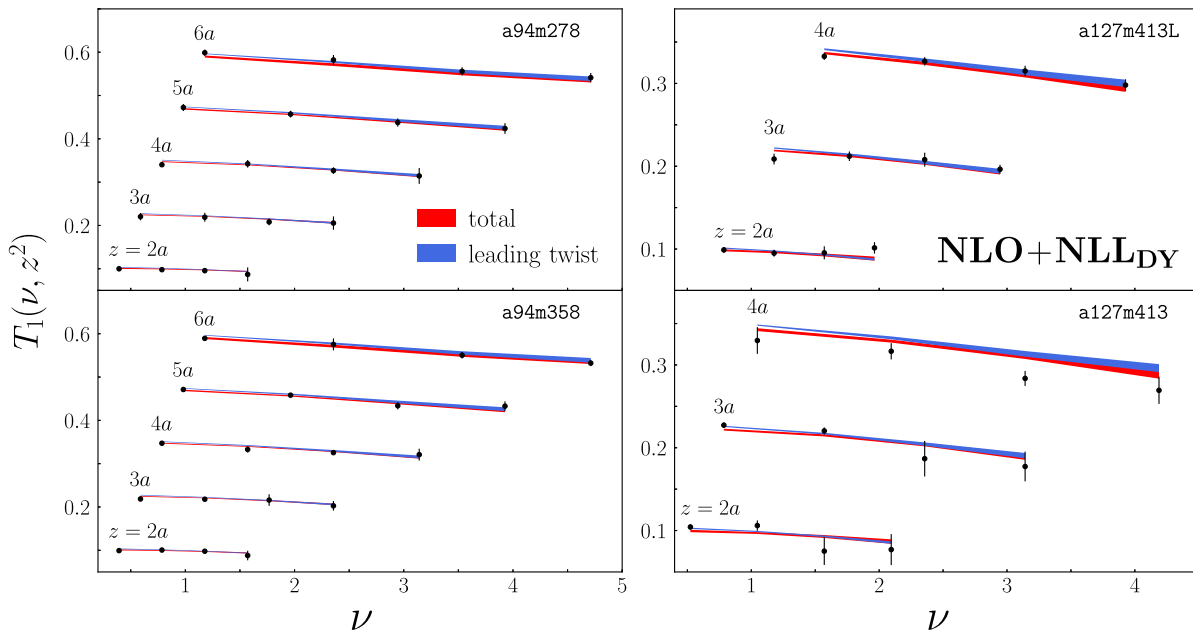
FIG. 12.　Same as Fig. 11, but for the NLO + NLL$_{DY}$ method.
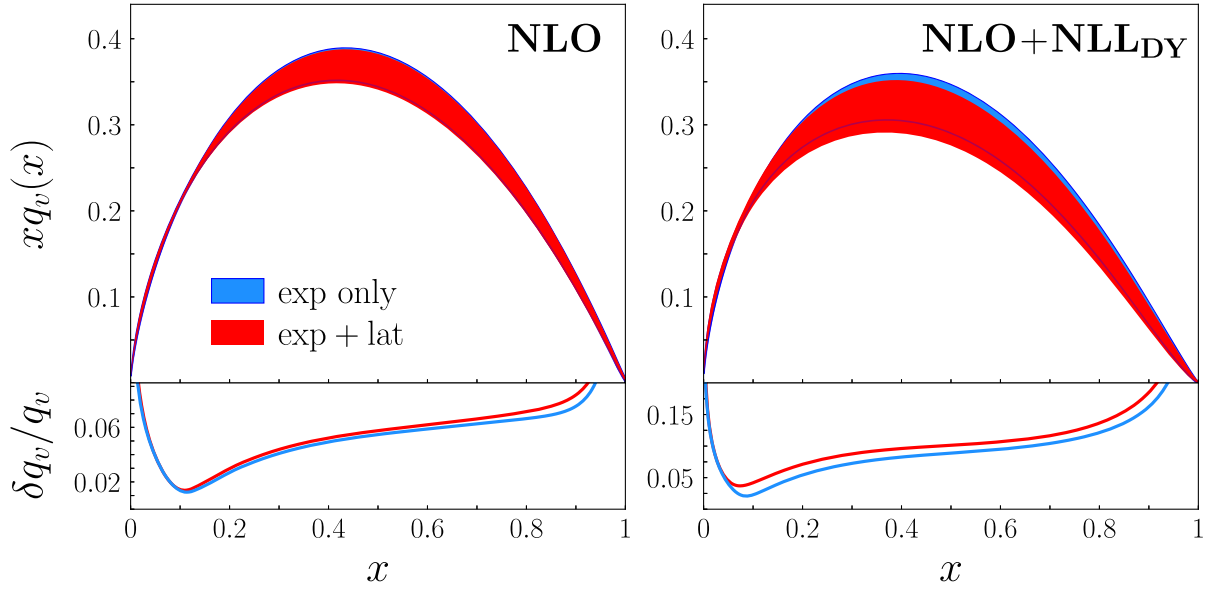
FIG. 13. Valence quark distribution $xq_v(x)$ (top) and the relative uncertainty $\delta q_v/q_v$ (bottom) from the experimental-only analysis (blue 1σ uncertainty bands) and with inclusion of the CC lattice data (red bands) for the NLO (left) and NLO + NLL$_{DY}$ (right) methods.
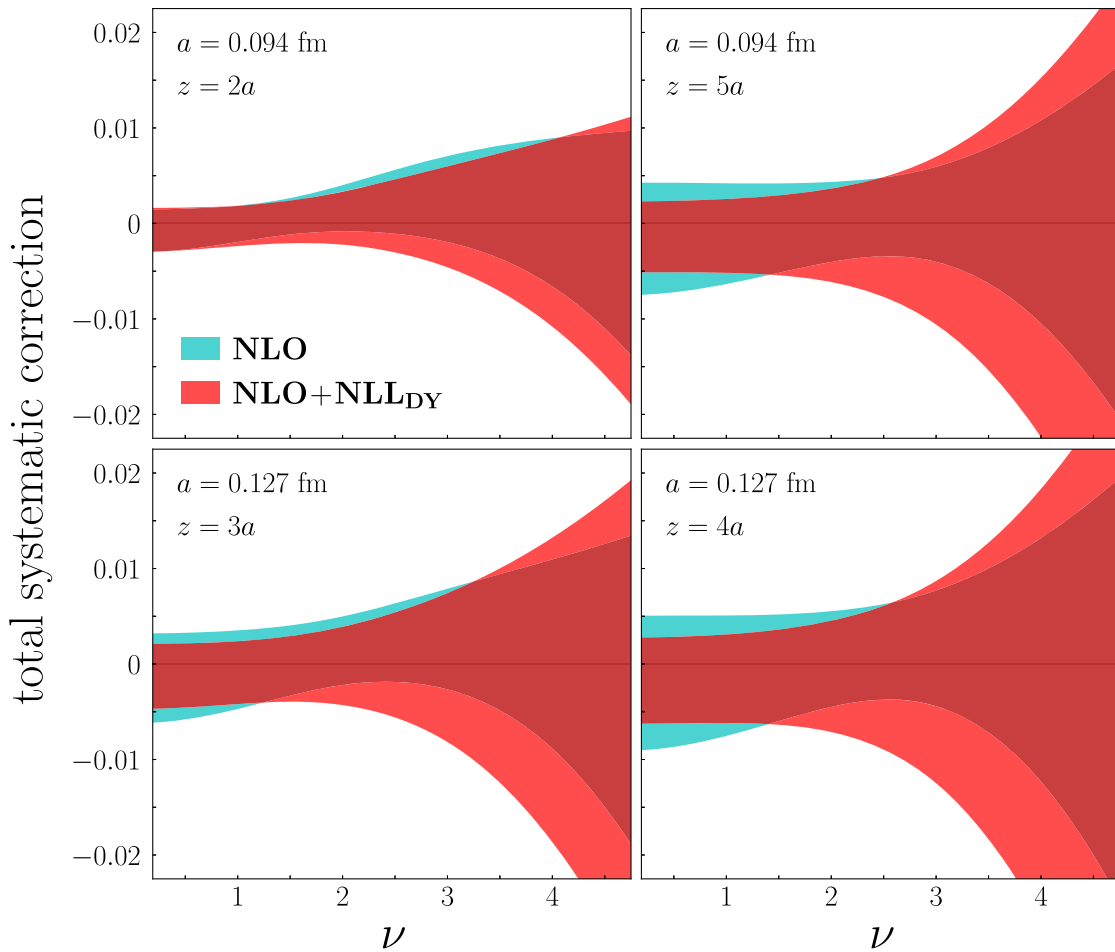
FIG. 14. Total systematic corrections from the CC correlators for $a = 0.094$ fm (top) and $a = 0.127$ fm (bottom) from the NLO (cyan 1σ bands) and NLO + NLL$_{DY}$ (red bands) analyses.

overlap, notably in the lattice data with $a = 0.094$ fm, various features of the data are difficult to capture by using next-to-leading order leading power matching coefficients, and systematic corrections are needed to achieve good agreement with the full set of lattice data.

The valence quark distributions before and after the inclusion of the lattice CC data are shown in Fig. 13 for the NLO and NLO + NLL$_{DY}$ methods, along with their relative uncertainties. The PDFs themselves are rather stable under the inclusion of the lattice data, suggesting compatibility between these lattice and experimental datasets. Unlike the Rp-ITD data, the CC lattice data have essentially no pull on the central values for the NLO + NLL$_{DY}$ analysis, indicating that these data have minor impact on the PDFs. The relative uncertainties are very similar in both cases, with a slight increase when including the CC lattice data compared with the experimental-only analysis. Similar behaviors are seen for the sea quark and gluon distributions (not shown here). Despite the small increase in the relative uncertainties of the PDFs, the relative uncertainties on the predictions for the experimental observables using these

PDFs do not increase, which is a reflection of the nontrivial interplay between different flavors of PDFs.

Figure 14 shows the total systematic corrections from the global analyses for various z and a values that are sensitive to the discretization and power corrections. Again, considerable overlap occurs between the NLO and NLO + NLL$_{DY}$ results, as was similarly observed in the Rp-ITD analysis. Contrary to the Rp-ITD analysis, the systematic corrections at the smallest ν are not necessarily zero by construction, and in fact are slightly negative. In the NLO case, the total correction generally increases as ν increases for all values of z and a, and at large ν trends downwards. While the corrections from using the NLO + NLL$_{DY}$ method follows similar trends as in the NLO analysis, the systematic corrections tend to have larger uncertainties at larger ν.

Even though the existing CC correlator data do not impact the PDFs significantly, the consistency of the extracted PDFs from such a different hadronic matrix element clearly signals the nontrivial success of QCD factorization and universality of PDFs. We also explore
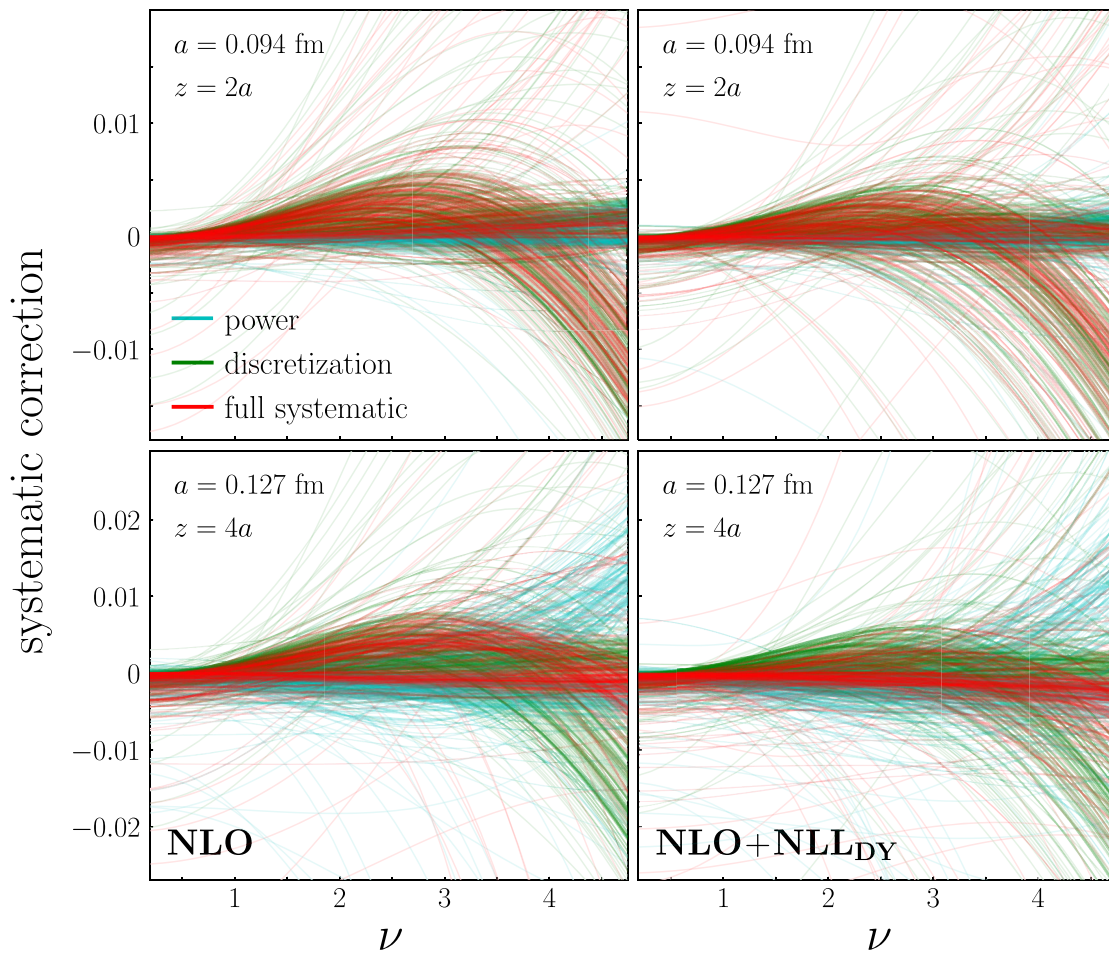

FIG. 15. Systematic corrections associated with the CC correlator lattice data broken down into power (cyan), discretization (green), and total (red) corrections from the NLO (left) and NLO + NLL$_{DY}$ (right) analyses at various a and z values. Shown is a representative subset of 500 replicas.

what may be needed in future in order to achieve an impact. In the current analysis, the total uncertainty associated with the systematic corrections is generally similar to, or larger than, the uncertainty of the total theory shown in Figs. 11 and 12. It is unlikely that reducing statistical uncertainties will significantly improve the uncertainty on the PDF, because of the size of the systematic corrections. We have carried out an impact study in which the statistical uncertainties of the CC lattice data are decreased slowly to match the systematic uncertainty. As expected, the impact on the PDFs was minimal. Reducing the uncertainties on the PDFs from the CC correlator data clearly requires in the first instance a better understanding of the systematic corrections.

Finally, in Fig. 15 we present the breakdown of the systematic corrections in terms of the discretization and power corrections for the NLO and $NLO + NLL_{DY}$ analyses. As expected, the power corrections are quite small, just away from zero when $z = 2a$, but increase in magnitude for $z = 4a$. The discretization corrections are responsible for the general trends of the full systematic uncertainty at small z, but these corrections seem to compete with the power corrections at large z and a. When both z and a are large, there tends to be a tension between the two systematic corrections, resulting in a large uncertainty of the total correction, and the full systematic correction is consistent with zero with a large uncertainty. Further lattice calculations, at different pion masses and lattice spacings and with improved statistical precision, are needed to tame these systematic effects and to test our ansatz for the correction terms.

IV. CONCLUSION

This is the first analysis of its kind, in which experimental data on high-energy pion-nucleus Drell-Yan and leading neutron electroproduction reactions have been supplemented by lattice QCD data on reduced Ioffe time pseudo-distributions and matrix elements of current-current correlators to constrain the PDFs of the pion. One of the main motivations of simultaneously fitting both the experimental and lattice data has been to rigorously quantify the uncertainties on the pion PDFs and identify systematic effects intrinsic to the lattice QCD observables.

The use of the NLO or $NLO + NLL$ short distance coefficients for the DY data is not distinguishable on the basis of goodness of fit to the experimental data. Including the lattice QCD data had *a priori* unknown systematic corrections associated with the lattice ensembles, complicating the ability to distinguish between the two DY hard coefficient calculations. The agreement with the lattice data was similar for the two methods; however, the resulting PDFs were affected in the case of the $NLO + NLL_{DY}$ analysis when including Rp-ITD data. This suggests the need to further investigate possible tensions between the experimental and lattice observables, and the exercise

would benefit from new observables being available. Regardless of this, the lattice data, in conjunction with the experimental measurements, consistently prefer an effective exponent $\beta_{eff} \approx 1.0–1.2$.

The improvement found in the PDF uncertainties with the inclusion of the Rp-ITD lattice data has implications for future global analyses with combined experimental and lattice QCD observables. First, since the $p = 1$ data have such a limited range of Ioffe time, the Fourier transform and the nature of the inverse problem complicate the determination of an x-space quasi-PDF or pseudo-PDF from the lattice data alone, particularly in the low-x region. The precise low-ν data from the lattice can provide a significant constraint on the lowest moments of the PDFs, complemented by the experimental range of data in the low-x region. The $p = 1$ data with $\mathcal{O}(0.1\%)$ statistical precision seem necessary for providing a useful constraint on the PDFs, while adding the $p = 2$ data with $\mathcal{O}(1\%)$ statistical precision did not dramatically improve the result. This fact suggests that future lattice calculations which aim to complement experimental datasets will require subpercent level precision. These types of constraints can be provided by the lattice data through a factorization method which promotes the use of the low momentum data, such as the pseudo-PDF approach.

Through the complementarity of the lattice and experimental data, we were able to quantify the systematic corrections associated with these ensembles and their uncertainties. Decreasing the statistical uncertainties on lattice observables is not sufficient to improve our knowledge of pion PDFs, since statistical and systematic effects are comparable in size. Improved control over systematic effects and reduced statistical uncertainties are both required to further constrain the PDFs. Future lattice calculations performed at smaller lattice spacings will further limit the impact of the power and discretization corrections. Here, the leading twist contributions are dominant, allowing the lattice data to isolate the PDFs more cleanly and provide reliable tests of universality on the PDFs. Finite volume effects were shown to be insignificant for the available Rp-ITD data, so that sampling different lattice volumes may not be essential. On the experimental side, more observables are needed in regions of kinematics that overlap with existing data in order to test the universality of the pion PDFs.

The importance of combining experimental and lattice QCD data was also evident in analyzing the Rp-ITD data with the smallest momentum. The lattice data had a significant impact, despite the limited range of Ioffe time, because of the complementarity of the lattice and experimental data. While this was evident in this analysis with collinear pion PDFs, it should also be noted that this improvement may not be as significant when one considers the nucleon. Because the mass of the nucleon is larger than that of the pion, the statistical noise of low momentum

nucleon lattice QCD correlation functions is generally higher than the pion, though the signal-to-noise ratio of calculations, at fixed computational cost, does not seem to decay as dramatically with momentum as for the pion. Additionally, the collinear isovector nucleon PDFs are already quite well constrained by experimental data, so significantly more precise lattice data may be needed to achieve the relative improvement. Extensions towards nucleon PDFs that are not well constrained by experimental data such as helicity and transversity PDFs could be useful following the methodology presented in this work.

ACKNOWLEDGMENTS

[1] J. S. Conway *et al.*, Phys. Rev. D **39**, 92 (1989).

[2] B. Betev *et al.*, Z. Phys. C **28**, 9 (1985).

[3] J. F. Owens, Phys. Rev. D **30**, 943 (1984).

[4] P. Aurenche, R. Baier, M. Fontannaz, M. N. Kienzle-Focacci, and M. Werlen, Phys. Lett. B **233**, 517 (1989).

[5] P. J. Sutton, A. D. Martin, R. G. Roberts, and W. J. Stirling, Phys. Rev. D **45**, 2349 (1992).

[6] M. Gluck, E. Reya, and A. Vogt, Z. Phys. C **53**, 651 (1992).

[7] M. Gluck, E. Reya, and I. Schienbein, Eur. Phys. J. C **10**, 313 (1999).

[8] K. Wijesooriya, P. E. Reimer, and R. J. Holt, Phys. Rev. C **72**, 065203 (2005).

[9] I. Novikov *et al.*, Phys. Rev. D **102**, 014040 (2020).

[10] F. D. Aaron *et al.*, Eur. Phys. J. C **68**, 381 (2010).

[11] S. Chekanov *et al.*, Nucl. Phys. **B637**, 3 (2002).

[12] P. C. Barry, N. Sato, W. Melnitchouk, and C.-R. Ji, Phys. Rev. Lett. **121**, 152001 (2018).

[13] N. Y. Cao, P. C. Barry, N. Sato, and W. Melnitchouk, Phys. Rev. D **103**, 114014 (2021).

[14] P. C. Barry, C.-R. Ji, N. Sato, and W. Melnitchouk, Phys. Rev. Lett. **127**, 232001 (2021).

[15] Z. F. Ezawa, Nuovo Cimento A **23**, 271 (1974).

[16] P. V. Landshoff and J. C. Polkinghorne, Nucl. Phys. **B53**, 473 (1973).

[17] J. F. Gunion, S. J. Brodsky, and R. Blankenbecler, Phys. Rev. D **8**, 287 (1973).

[18] G. R. Farrar and D. R. Jackson, Phys. Rev. Lett. **43**, 246 (1979).

[19] E. L. Berger and S. J. Brodsky, Phys. Rev. Lett. **42**, 940 (1979).

[20] T. Shigetani, K. Suzuki, and H. Toki, Phys. Lett. B **308**, 383 (1993).

[21] A. Szczepaniak, C.-R. Ji, and S. R. Cotanch, Phys. Rev. D **49**, 3466 (1994).

[22] R. M. Davidson and E. Ruiz Arriola, Phys. Lett. B **348**, 163 (1995).

[23] M. B. Hecht, C. D. Roberts, and S. M. Schmidt, Phys. Rev. C **63**, 025213 (2001).

[24] W. Melnitchouk, Eur. Phys. J. A **17**, 223 (2003).

[25] S. Noguera and S. Scopetta, J. High Energy Phys. 11 (2015) 102.

[26] P. T. P. Hutauruk, I. C. Cloet, and A. W. Thomas, Phys. Rev. C **94**, 035201 (2016).

[27] T. J. Hobbs, Phys. Rev. D **97**, 054028 (2018).

[28] G. F. de Teramond, T. Liu, R. S. Sufian, H. G. Dosch, S. J. Brodsky, and A. Deur, Phys. Rev. Lett. **120**, 182001 (2018).

[29] K. D. Bednar, I. C. Cloët, and P. C. Tandy, Phys. Rev. Lett. **124**, 042002 (2020).

[30] J. Lan, C. Mondal, S. Jia, X. Zhao, and J. P. Vary, Phys. Rev. Lett. **122**, 172001 (2019).

[31] J. Lan, C. Mondal, S. Jia, X. Zhao, and J. P. Vary, Phys. Rev. D **101**, 034024 (2020).

[32] L. Chang, K. Raya, and X. Wang, Chin. Phys. C **44**, 114105 (2020).

[33] Z.-F. Cui, M. Ding, F. Gao, K. Raya, D. Binosi, L. Chang, C. D. Roberts, J. Rodríguez-Quintero, and S. M. Schmidt, Eur. Phys. J. C **80**, 1064 (2020).

[34] A. Kock, Y. Liu, and I. Zahed, Phys. Rev. D **102**, 014039 (2020).

[35] Z. F. Cui, M. Ding, J. M. Morgado, K. Raya, D. Binosi, L. Chang, F. De Soto, C. D. Roberts, J. Rodríguez-Quintero, and S. M. Schmidt, Phys. Rev. D **105**, L091502 (2022).

[36] C. E. Keppel et al., Measurement of tagged deep inelastic scattering (TDIS), Jefferson Lab Experiment PR12-15-006, 2015.

[37] R. Abdul Khalek et al., arXiv:2103.05419.

[38] J. Arrington et al., J. Phys. G **48**, 075106 (2021).

[39] M. Aicher, A. Schafer, and W. Vogelsang, Phys. Rev. Lett. **105**, 252003 (2010).

[40] D. Westmark and J. F. Owens, Phys. Rev. D **95**, 056024 (2017).

[41] H. Shimizu, G. F. Sterman, W. Vogelsang, and H. Yokoya, Phys. Rev. D **71**, 114007 (2005).

[42] S. Catani, M. L. Mangano, P. Nason, and L. Trentadue, Nucl. Phys. **B478**, 273 (1996).

[43] G. F. Sterman and W. Vogelsang, arXiv:hep-ph/0002132.

[44] J.-H. Zhang, J.-W. Chen, L. Jin, H.-W. Lin, A. Schäfer, and Y. Zhao, Phys. Rev. D **100**, 034505 (2019).

[45] Z.-Y. Fan, Y.-B. Yang, A. Anthony, H.-W. Lin, and K.-F. Liu, Phys. Rev. Lett. **121**, 242001 (2018).

[46] R. S. Sufian, J. Karpie, C. Egerer, K. Orginos, J.-W. Qiu, and D. G. Richards, Phys. Rev. D **99**, 074507 (2019).

[47] J.-W. Chen, H.-W. Lin, and J.-H. Zhang, Nucl. Phys. **B952**, 114940 (2020).

[48] T. Izubuchi, L. Jin, C. Kallidonis, N. Karthik, S. Mukherjee, P. Petreczky, C. Shugert, and S. Syritsyn, Phys. Rev. D **100**, 034516 (2019).

[49] B. Joó, J. Karpie, K. Orginos, A. V. Radyushkin, D. G. Richards, R. S. Sufian, and S. Zafeiropoulos, Phys. Rev. D **100**, 114512 (2019).

[50] H.-W. Lin, J.-W. Chen, Z. Fan, J.-H. Zhang, and R. Zhang, Phys. Rev. D **103**, 014516 (2021).

[51] R. S. Sufian, C. Egerer, J. Karpie, R. G. Edwards, B. Joó, Y.-Q. Ma, K. Orginos, J.-W. Qiu, and D. G. Richards, Phys. Rev. D **102**, 054508 (2020).

[52] N. Karthik, Phys. Rev. D **103**, 074512 (2021).

[53] Z. Fan and H.-W. Lin, Phys. Lett. B **823**, 136778 (2021).

[54] Y.-Q. Ma and J.-W. Qiu, Phys. Rev. D **98**, 074021 (2018).

[55] A. V. Radyushkin, Phys. Rev. D **96**, 034025 (2017).

[56] Y.-Q. Ma and J.-W. Qiu, Phys. Rev. Lett. **120**, 022003 (2018).

[57] X. Gao, K. Lee, S. Mukherjee, C. Shugert, and Y. Zhao, Phys. Rev. D **103**, 094504 (2021).

[58] V. M. Braun, A. Vladimirov, and J.-H. Zhang, Phys. Rev. D **99**, 014013 (2019).

[59] V. Braun and D. Müller, Eur. Phys. J. C **55**, 349 (2008).

[60] A. J. Chambers, R. Horsley, Y. Nakamura, H. Perlt, P. E. L. Rakow, G. Schierholz, A. Schiller, K. Somfleth, R. D. Young, and J. M. Zanotti, Phys. Rev. Lett. **118**, 242001 (2017).

[61] C. Alexandrou, K. Cichy, M. Constantinou, K. Jansen, A. Scapellato, and F. Steffens, Phys. Rev. Lett. **121**, 112001 (2018).

[62] H.-W. Lin, J.-W. Chen, X. Ji, L. Jin, R. Li, Y.-S. Liu, Y.-B. Yang, J.-H. Zhang, and Y. Zhao, Phys. Rev. Lett. **121**, 242003 (2018).

[63] M. Bhat, K. Cichy, M. Constantinou, and A. Scapellato, Phys. Rev. D **103**, 034510 (2021).

[64] C. Alexandrou, K. Cichy, M. Constantinou, K. Jansen, A. Scapellato, and F. Steffens, Phys. Rev. D **98**, 091503 (2018).

[65] B. Joó, J. Karpie, K. Orginos, A. Radyushkin, D. Richards, and S. Zafeiropoulos, J. High Energy Phys. 12 (2019) 081.

[66] Z. Fan, X. Gao, R. Li, H.-W. Lin, N. Karthik, S. Mukherjee, P. Petreczky, S. Syritsyn, Y.-B. Yang, and R. Zhang, Phys. Rev. D **102**, 074504 (2020).

[67] B. Joó, J. Karpie, K. Orginos, A. V. Radyushkin, D. G. Richards, and S. Zafeiropoulos, Phys. Rev. Lett. **125**, 232003 (2020).

[68] K. Cichy, L. Del Debbio, and T. Giani, J. High Energy Phys. 10 (2019) 137.

[69] L. Del Debbio, T. Giani, J. Karpie, K. Orginos, A. Radyushkin, and S. Zafeiropoulos, J. High Energy Phys. 02 (2021) 138.

[70] W. Detmold, A. V. Grebe, I. Kanamori, C. J. D. Lin, R. J. Perry, and Y. Zhao, Phys. Rev. D **104**, 074511 (2021).

[71] C. Egerer, R. G. Edwards, C. Kallidonis, K. Orginos, A. V. Radyushkin, D. G. Richards, E. Romero, and S. Zafeiropoulos, J. High Energy Phys. 11 (2021) 148.

[72] T. Khan et al. (HadStruc Collaboration), Phys. Rev. D **104**, 094516 (2021).

[73] C. Egerer et al., Phys. Rev. D **105**, 034507 (2022).

[74] W. Detmold, W. Melnitchouk, and A. W. Thomas, Phys. Rev. D **68**, 034025 (2003).

[75] W. Detmold, W. Melnitchouk, and A. W. Thomas, Mod. Phys. Lett. A **18**, 2681 (2003).

[76] H.-W. Lin et al., Prog. Part. Nucl. Phys. **100**, 107 (2018).

[77] J. Karpie, K. Orginos, and S. Zafeiropoulos, J. High Energy Phys. 11 (2018) 178.

[78] J. Bringewatt, N. Sato, W. Melnitchouk, J.-W. Qiu, F. Steffens, and M. Constantinou, Phys. Rev. D **103**, 016003 (2021).

[79] H.-W. Lin, W. Melnitchouk, A. Prokudin, N. Sato, and H. Shows, Phys. Rev. Lett. **120**, 152502 (2018).

[80] S. D. Drell and T.-M. Yan, Phys. Rev. Lett. **25**, 316 (1970); **25**, 902(E) (1970).

[81] J. C. Collins, D. E. Soper, and G. F. Sterman, Phys. Lett. B **134**, 263 (1984).

[82] G. Lustermans, J. K. L. Michel, and F. J. Tackmann, arXiv:1908.00985.

[83] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, Eur. Phys. J. C **77**, 163 (2017).

[84] K. Kovarik et al., Phys. Rev. D **93**, 085037 (2016).

[85] J. D. Sullivan, Phys. Rev. D **5**, 1732 (1972).

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

[87] W. Melnitchouk and A. W. Thomas, Phys. Rev. D **47**, 3794 (1993).

[88] U. D'Alesio and H. J. Pirner, Eur. Phys. J. A **7**, 109 (2000).

[89] B. Z. Kopeliovich, I. K. Potashnikova, B. Povh, and I. Schmidt, Phys. Rev. D **85**, 114025 (2012).

[90] M. Burkardt, K. S. Hendricks, C.-R. Ji, W. Melnitchouk, and A. W. Thomas, Phys. Rev. D **87**, 056009 (2013).

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

[92] X. G. Wang, C.-R. Ji, W. Melnitchouk, Y. Salamu, A. W. Thomas, and P. Wang, Phys. Rev. D **94**, 094035 (2016).

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

[94] Y. Salamu, C.-R. Ji, W. Melnitchouk, A. W. Thomas, P. Wang, and X. G. Wang, Phys. Rev. D **100**, 094026 (2019).

[95] H. Holtmann, A. Szczurek, and J. Speth, Nucl. Phys. **A596**, 631 (1996).

[96] J. R. McKenney, N. Sato, W. Melnitchouk, and C.-R. Ji, Phys. Rev. D **93**, 054011 (2016).

[97] J. C. Collins, D. E. Soper, and G. F. Sterman, Adv. Ser. Dir. High Energy Phys. **5**, 1 (1989).

[98] M. Constantinou and H. Panagopoulos, Phys. Rev. D **96**, 054506 (2017).

[99] C. Alexandrou, K. Cichy, M. Constantinou, K. Hadjiyiannakou, K. Jansen, H. Panagopoulos, and F. Steffens, Nucl. Phys. **B923**, 394 (2017).

[100] 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).

[101] I. W. Stewart and Y. Zhao, Phys. Rev. D **97**, 054512 (2018).

[102] K. Orginos, A. Radyushkin, J. Karpie, and S. Zafeiropoulos, Phys. Rev. D **96**, 094503 (2017).

[103] Z.-Y. Li, Y.-Q. Ma, and J.-W. Qiu, Phys. Rev. Lett. **126**, 072001 (2021).

[104] B. L. Ioffe, Phys. Lett. B **30**, 123 (1969).

[105] A. Radyushkin, Phys. Rev. D **98**, 014019 (2018).

[106] J.-H. Zhang, J.-W. Chen, and C. Monahan, Phys. Rev. D **97**, 074508 (2018).

[107] T. Izubuchi, X. Ji, L. Jin, I. W. Stewart, and Y. Zhao, Phys. Rev. D **98**, 056004 (2018).

[108] A. Radyushkin, Phys. Lett. B **770**, 514 (2017).

[109] G. P. Lepage, *Theoretical Advanced Study Institute in Elementary Particle Physics* (World Scientific, Singapore, 1990), p. 750.

[110] J. Karpie, K. Orginos, A. Radyushkin, and S. Zafeiropoulos, J. High Energy Phys. 11 (2021) 024.

[111] B. Yoon *et al.*, Phys. Rev. D **95**, 074508 (2017).

[112] E. R. Nocera, Phys. Lett. B **742**, 117 (2015).

[113] A. Courtoy and P. M. Nadolsky, Phys. Rev. D **103**, 054029 (2021).

[114] A. Courtoy and P. M. Nadolsky, in *Proceedings of the 28th International Workshop on Deep Inelastic Scattering and Related Subjects*, arXiv:2108.04122.

[115] G. Parisi, Phys. Rep. **103**, 203 (1984).

[116] J. Karpie, K. Orginos, A. Rothkopf, and S. Zafeiropoulos, J. High Energy Phys. 04 (2019) 057.

[117] J. Towns *et al.*, Comput. Sci. Eng. **16**, 62 (2014).

[118] Jülich Supercomputing Centre, J. Large-Scale Res. Facil. **7**, A182 (2021).