Continuum-extrapolated NNLO valence PDF of the pion at the physical point

Xiang Gao,[1,*] Andrew D. Hanlon,[2] Nikhil Karthik,[3,4] Swagato Mukherjee,[2] Peter Petreczky,[2] Philipp Scior,[2] Shuzhe Shi,[5] Sergey Syritsyn,[5,6] Yong Zhao,[1] and Kai Zhou[7]

[1]*Physics Division, Argonne National Laboratory, Lemont, Illinois 60439, USA*
[2]*Physics Department, Brookhaven National Laboratory, Upton, New York 11973, USA*
[3]*Department of Physics, College of William & Mary, Williamsburg, Virginia 23185, USA*
[4]*Thomas Jefferson National Accelerator Facility, Newport News, Virginia 23606, USA*
[5]*Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York 11794, USA*
[6]*RIKEN-BNL Research Center, Brookhaven National Laboratory, Upton, New York 11973, USA*
[7]*Frankfurt Institute for Advanced Studies (FIAS), D-60438 Frankfurt am Main, Germany*

We present lattice QCD calculations of the valence parton distribution function (PDF) of pion employing next-to-next-leading-order (NNLO) perturbative QCD matching. Our calculations are based on three gauge ensembles of $2 + 1$ flavor highly improved staggered quarks and Wilson-Clover valance quarks, corresponding to pion mass $m_\pi = 140$ MeV at a lattice spacing $a = 0.076$ fm and $m_\pi = 300$ MeV at $a = 0.04, 0.06$ fm. This enables us to present, for the first time, continuum-extrapolated lattice QCD results for the NNLO valence PDF of the pion at the physical point. Applying leading-twist expansion for renormalization group invariant (RGI) ratios of bilocal pion matrix elements with NNLO Wilson coefficients, we extract second, fourth, and sixth Mellin moments of the PDF. We reconstruct the Bjorken-x dependence of the NNLO PDF from real-space RGI ratios using a deep neural network as well as from momentum-space matrix elements renormalized using a hybrid scheme. All our results are in broad agreement with the results of global fits to the experimental data carried out by the xFitter and JAM collaborations.

I. INTRODUCTION

Pions are the Nambu-Goldstone bosons of QCD with massless quarks due to spontaneous chiral symmetry breaking. Understanding the structure of the pion, therefore, plays a central role in the study of the strong interaction, in particular in clarifying the relation between hadron mass and hadron structure [1]. The collinear partonic structure of the pion is encoded in the parton distribution functions (PDFs), which describe the collinear-momentum fraction x of a hadron carried by the partons, and can be extracted from deep-inelastic scattering experiments. However, the study of the pion PDFs from experiment is more difficult than the study of the nucleon PDFs due to the sparseness of the experimental data. As a result, the pion PDF is less constrained than the unpolarized quark

nucleon PDF. Therefore, lattice QCD calculations could play an important role in constraining the pion PDF. The PDFs are defined as the Fourier transform of light-cone correlation functions [2] and, therefore, cannot be computed directly on a Euclidean lattice. The Mellin moments of the PDFs can be directly calculated on the lattice, but in practice, the calculations are limited to only the lowest moments due to decreasing signal-to-noise ratios and the power-divergent operator mixing. In the case of the pion, the three lowest moments have been calculated [3].

In recent years, significant progress has been made since large-momentum effective theory (LaMET) [4–6] was proposed. LaMET makes it possible to calculate x-dependent PDFs on the lattice by computing the boosted matrix elements of equal-time extended operators; which, for a quark PDF, is the bilocal quark-bilinear operator,

$$O_\Gamma(z) \equiv \bar{\psi}(z)\Gamma W(z,0)\psi, \qquad (1)$$

where $W(z,0)$ is the Wilson line connecting the quark and antiquark fields, ψ and $\bar{\psi}$, to preserve gauge invariance. For the unpolarized PDF, Γ can be either $\Gamma = \gamma_z$ or $\Gamma = \gamma_0$. The choice $\Gamma = \gamma_0$ has several advantages including being free

of the operator mixing due to the explicit chiral symmetry breaking at finite lattice spacings [7,8] and the absence of additional higher-twist effects proportional to z^μ [9,10]. The Fourier transform of these equal-time matrix elements defines the quasi-PDF (qPDF), which, for hadron states with large momentum, can be perturbatively matched to a light-cone PDF [6,11] up to certain power corrections. Using the same operators, several approaches have also been developed to extract either the Mellin moments of PDFs or x-dependent PDFs, namely, the Ioffe-time pseudodistributions or the pseudo-PDF [9,12]. Alternatively, other approaches have also been proposed such as the short-distance expansion of the current-current correlator [13,14], the operator product expansion (OPE) of a Compton amplitude in the unphysical region [15], the hadronic tensor [16], the heavy-quark operator product expansion [17,18] and so on. Significant progress has been made for the study of nucleon isovector PDFs [19–34] (for reviews, see Refs. [6,35,36]), the flavor decomposition of nucleon PDFs [37,38], gluon PDFs [39–41], and PDFs beyond leading twist [42–45].

Several lattice QCD calculations of the pion PDF have been performed recently [33,46–57] with pion masses heavier than the physical point. There have also been exploratory lattice studies of the structure of the pion radial excitation [58] and of the pion in QCD-like theories [59,60]. In all of these studies, the next-to-leading order (NLO) perturbative matching between the Euclidean time quantities and the light-cone PDF have been used. Recently, next-to-next-to-leading order (NNLO) perturbative matching has been computed [61,62], which is supposed to be more reliable by reducing the perturbation theory uncertainty. Very recently, the NNLO matching has been used to study the pion PDF using the so-called hybrid renormalization scheme in x space [63]. Additionally, a NNLO matching in real space was recently performed in the case of the unpolarized proton PDF in Ref. [19]. The goal of this paper is to study the valence PDF of the pion for physical quark masses with both real-space and x-space matching at NNLO. The lattice calculations are performed at a single lattice spacing, $a = 0.076$ fm. Combining this calculation with the previous ones on finer lattices, but with larger than physical quark mass, we provide estimates of the valence pion PDF and its moments in the continuum limit at the physical point.

The rest of the paper is organized as follows. In Sec. II, we present our lattice setup. In Sec. III, we discuss the lattice calculations of the pion two-point functions, while in Sec. IV, we discuss the extraction of the matrix elements of the quasi-PDF operator. In Sec. V, we briefly review the ratio scheme renormalization used in this work. In Sec. VI, we discuss the determination of even moments of the valence pion PDF using short-distance factorization. In Sec. VII, we discuss the determination of the valence pion PDF through a model-dependent fit of the lattice data. A determination of the valence pion PDF through a deep neural network technique is presented in Sec. VIII. In Sec. IX, we discuss the PDF from the hybrid-scheme renormalization and x-spacing matching. Finally, Sec. X contains our conclusions. Some technical aspects of the calculations including the NNLO matching and the deep neural network (DNN) technique are discussed in the Appendix.

II. LATTICE SETUP

In this paper, the new addition is the dataset at $a = 0.076$ fm with a physical pion mass, and we describe the details for this dataset below. To extract the bare matrix elements of the pion, we computed the pion two-point and three-point functions using the Wilson-Clover action with hypercubic (HYP) smearing gauge fields for the valence quarks on a $2 + 1$ flavor highly improved staggered quark (HISQ) [64] action generated by the HotQCD Collaboration [65] with a pion mass of 140 MeV and lattice spacing $a = 0.076$ fm. The lattice extent $L_t \times L_s^3$ is 64×64^3. We used the tree-level tadpole improved result for the coefficients of the clover term, $c_{sw} = u_0^{-3/4} = 1.0372$, and the quark mass has been tuned so that the valence pion mass is 140 MeV as shown in Table I. Here, u_0 denotes the expectation value of the plaquette on HYP smeared gauge configurations. The gauge link entering the bilocal quark-bilinear operator has been 1-HYP smeared for an improved signal. This setup has recently been used in the calculations of the pion form factor [66] and pion distribution amplitude [67].

The calculations were performed on GPUs, with the QUDA multigrid algorithm [68–71] used for the Wilson-Dirac operator inversions to get the quark propagators. We used the All Mode Averaging [72] technique to increase the statistics with a stopping criterion of 10^{-10} and 10^{-4} for the exact (ex) and sloppy (sl) inversions, respectively. One of the crucial ingredients to access the parton distribution functions of the pion is the large momentum. To obtain an acceptable signal for pions at large momentum, it is necessary to use boosted sources [73]. With the setup above, we are able to achieve momentum as large as

TABLE I. Details of the setup of the $a = 0.076$ fm lattice used in this paper. The number of gauge configuration (#cfgs) and the number of exact and sloppy inversion samples (#ex,#sl) are shown. The momenta are given in lattice units and can be computed by $P_z = 2\pi n_z/(L_s a)$.

Ensembles $a, L_t \times L_s^3$	m_π (GeV)	t_s	z/a	n_z	#cfgs	(#ex,#sl)
$a = 0.076$ fm,	0.14	6,8,	[0,32]	0,1,2,3	350	(5,100)
64×64^3		10		4,5,6,7	350	(10,200)

1.78 GeV for the physical pion mass with a reasonable signal quality. To control the lattice spacing and pion mass dependence, we combine the analysis with two other ensembles with a 300 MeV pion mass, lattice spacing a = 0.04, 0.06 fm and largest momentum up to 2.42 GeV, which has been discussed in detail and analyzed in Refs. [48,50].

A. Pion two-point functions

To construct the boosted pion state, we compute the pion two-point function,

$$C_{2\text{pt}}^{ss'}(t_s; P_z) = \langle \pi_s(\mathbf{x_0}, t_s)\pi_{s'}^\dagger(\mathbf{P}, 0)\rangle, \qquad (2)$$

using the standard pion operator projected to spatial momentum $\mathbf{P}$,

$$\pi_s(\mathbf{x}, t_s) = \bar{d}_s(\mathbf{x}, t_s)\gamma_5 u_s(\mathbf{x}, t_s),$$
$$\pi_s(\mathbf{P}, t_s) = \sum_{\mathbf{x}} \pi_s(\mathbf{x}, t_s)e^{-i\mathbf{P}\cdot\mathbf{x}}. \qquad (3)$$

The subscript s in π_s indicates the choice of quark smearing. In this work, we only work with the choice $\mathbf{P} = (0, 0, P_z)$. Since we use periodic boundary conditions, the hadron momentum is given by

$$P_z = \frac{2\pi n_z}{L_s a} \qquad (4)$$

with n_z listed in Table I. The operator will create infinitely many hadron states with the same quantum numbers as the pion, which are not only the pion ground states but also the excited states. We used momentum-smeared Gaussian-profiled sources in the Coulomb gauge [50] $\pi_s(\mathbf{P}, t)$ ($s = S$) to increase the overlap with the pion ground state and to improve the signal for large momentum. We constructed smeared-smeared (SS) correlators ($s = s' = S$) and smeared-point (SP) correlators ($s = S, s' = P$) to decompose the energy levels of the pion spectrum. We tuned the radius of the Gaussian profile to be 0.59 fm for the $a = 0.076$ fm ensemble. For the boosted smearing, we used the quark momentum $n_z^k = 2$ and 5 for hadron momentum $n_z \in [0, 3]$ and [4, 7], respectively. Since the different pion momentum n_z with the same quark momentum share the forward propagator, this enabled us to save some computation time. The details of the other two ensembles can be found in Ref. [48].

B. Pion three-point functions

We extracted the bare matrix elements from the three-point functions

$$C_{3\text{pt}}(z, \tau, t_s) = \langle \pi_S(\mathbf{x_0}, t_s)O_\Gamma(z, \tau)\pi_S^\dagger(\mathbf{P}, 0)\rangle, \qquad (5)$$

only using the smeared source and sink pion operator separated by a Euclidean time t_s. The isovector operator $O_\Gamma(z, \tau)$ inserted at time slice τ is defined as

$$O_\Gamma(z, \tau) = \sum_{\mathbf{x}} [\bar{u}(x + \mathcal{L})\Gamma W_z(x + \mathcal{L}, x)u(x)$$
$$- \bar{d}(x + \mathcal{L})\Gamma W_z(x + \mathcal{L}, x)d(x)], \qquad (6)$$

with $x = (\mathbf{x}, \tau)$. The quark-antiquark pairs are separated along the momentum direction $\mathcal{L} = (0, 0, 0, z)$, and the Wilson line $W_z(x + \mathcal{L}, x)$ makes sure such measurements are gauge invariant. In this work, we consider the choice $\Gamma = \gamma_0$, since it reduces the higher-twist contribution [9] and is free of mixing due to $\mathcal{O}(a^0)$ chiral symmetry breaking [7,74]. The gauge links that enter the Wilson line were 1-HYP smeared.

III. ANALYSIS OF PION TWO-POINT FUNCTIONS

To extract the pion ground-state matrix elements, one needs to know the energy levels created by the pion operator, which can be determined by analyzing the two-point functions. Below, we discuss the analysis of the two-point function for the $a = 0.076$ fm, physical quark mass ensemble. The details for the spectral analysis of the other two ensembles with a 300 MeV pion mass are given in Ref. [48]. The pion two-point functions constructed from Eq. (2) have the spectral decomposition,

$$C_{2\text{pt}}^{ss'}(t_s) = \sum_{n=0} A_n^s A_n^{s'*}(e^{-E_n t_s} + e^{-E_n(aL_t - t_s)}), \qquad (7)$$

with $E_{n+1} > E_n$ being the energy level and $A_n^s = \langle \Omega | \pi_s | n \rangle$ being the pion overlap amplitude. Here, $|\Omega\rangle$ denotes the vacuum state.

It is expected that the boosted Gaussian smeared sources should have a good overlap with the pion ground state. In Fig. 1, we show the effective mass of the SS (filled symbols) and SP (open symbols) correlators from the $a = 0.076$ fm ensemble. At very small t_s, the effective masses corresponding to the SS correlators are larger than the ones corresponding to SP correlator. This is likely due to the effect that some higher excited states contribute with a negative weight to the SP correlator. At larger t_s, the ordering of SP and SS correlators changes and the effective mass of SS correlators reaches a plateau around $t_s \lesssim 10a$. For the case of $P_z = 0$, we have the pion mass at the physical point around 140 MeV. The lines in the figure are calculated from the dispersion relation $E_0(P_z) = \sqrt{P_z^2 + m_\pi^2}$ with $m_\pi = 140$ MeV, which show nice agreement with the plateaus. The different behavior of the SS and SP correlators can help the extraction of the excited energy states.

We truncated Eq. (7) up to $n = N - 1$ and performed an N-state fit to the two-point functions with time separations in the range $[t_{\min}, 32a]$. We show the fit results for $n_z = 3, 5$
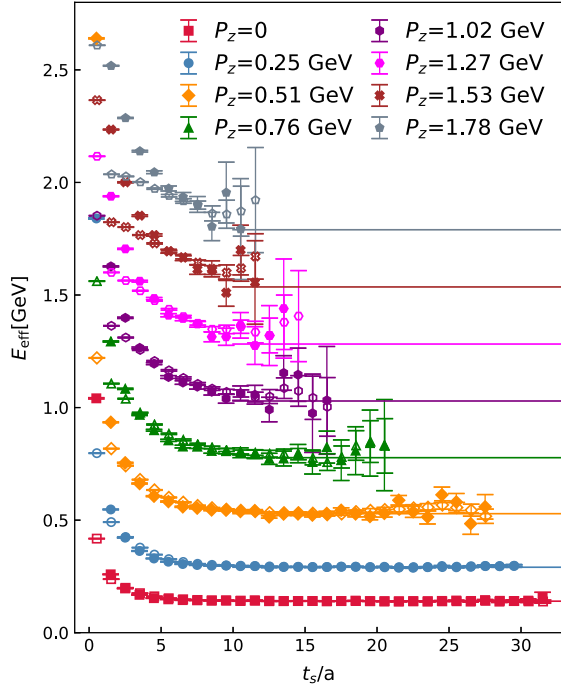
FIG. 1.   Pion effective masses for the $m_\pi = 140$ MeV ensemble are shown. The filled and open symbols correspond to the SS and SP correlators, respectively. The lines are calculated from dispersion relation $E(P_z) = \sqrt{P_z^2 + m_\pi^2}$ with $m_\pi = 140$ MeV.

in Fig. 2 as examples. In the left panels, we show the lowest energy E_0 of the SS correlators extracted from one-state (red points) and two-state (blue points) fits. The values of E_0 from the one-state fits show similar behavior to the effective mass (see Fig. 1) as a function of $t_{\min}$. They reach
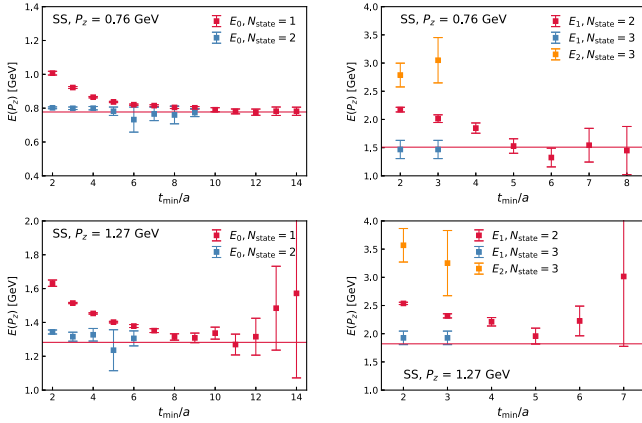

FIG. 2. N-state fit results for $n_z = 3, 5$ on the $m_\pi = 140$ MeV ensemble. The left panels show E_0 from one- and two-state fits to the SS correlators, while the right panels show E_1 from constrained two- and three-state fits to the SS correlators. The lines in left panels are computed from dispersion relation $E_0(P_z) = \sqrt{P_z^2 + m_\pi^2}$ with $m_\pi = 0.14$ GeV, while the ones in the right panels are with $m_\pi' = 1.3$ GeV.

plateaus around $t_s \lesssim 10a$ and agree with the red lines calculated from the dispersion relation. From the two-state fits, we find that the values of E_0 reach plateaus at smaller t_s, namely, $t_{\min} \sim 3a$. Therefore, from one-state and two-state models of the pion two-point function, we can extract the pion ground state using $t_s \geq 10a$ and $t_s \geq 3a$, respectively. To determine the first excited state E_1, we fixed E_0 to be the best estimate from the one-state fit and performed a constrained two-state fit. The values of E_1 from the SS correlators are shown as red points in the right panels of Fig. 2. These values are slowly decreasing with increasing $t_{\min}$ and roughly reach a plateau around $t_{\min} \sim 5a$ within the statistical errors. The plateaus are consistent with the red lines calculated from the dispersion relation $E_1(P_z) = \sqrt{P_z^2 + m_\pi'^2}$ with $m_\pi' \approx 1.3$ GeV, which may suggest a single particle state [58]. To decompose the higher energy states through the three-state fit of the SS correlators, in addition to fixing the E_0, we also added a prior on E_1 from the best estimates from the SP correlators with corresponding errors [48]. The E_1 (blue points) and E_2 (orange points) extracted from the constrained three-state fit of the SS correlators are shown in the right panels of Fig. 2. The three-state fit only works for $t_{\min} \leq 3a$ because of the limited statistics, and the E_2 values do not show $t_{\min}$ dependence within errors. But instead of a single-particle state, the E_2 is more likely to be the combination of the tower of energies beyond E_1, which, however, cannot be decomposed with the limited statistics. From this, we conclude that the spectrum present in the SS correlators can be described by a three-state model when $t_{\min} \gtrsim 2a$ and by a two-state model for $t_{\min} \gtrsim 5a$. These facts will contribute to the analysis of the three-point functions in the next section.

IV. GROUND-STATE BARE MATRIX ELEMENTS FROM THREE-POINT FUNCTIONS

The three-point function defined in Eq. (5) has the spectral decomposition,

$$\langle \pi_S(\mathbf{x}_0, t_s) O_\Gamma(z, \tau) \pi_S^\dagger(\mathbf{P}, 0) \rangle$$
$$= \sum_{m,n} \langle \Omega | \pi_S | m \rangle \langle m | O | n \rangle \langle n | \pi_S^\dagger | \Omega_\Gamma \rangle e^{-\tau E_n} e^{-(t_s - \tau) E_m}, \quad (8)$$

where the overlap amplitudes $A_n^s = \langle \Omega | \pi_S | n \rangle$ and the energy levels E_n are, by definition, the same as for the two-point functions. The quantity $h^B(z, P_z) = \langle 0 | O_\Gamma | 0 \rangle$ is the bare matrix element of the pion ground state. We only computed the three-point function with both the pion source and sink smeared. To take advantage of the high correlation between the three-point and two-point functions, we construct the ratio

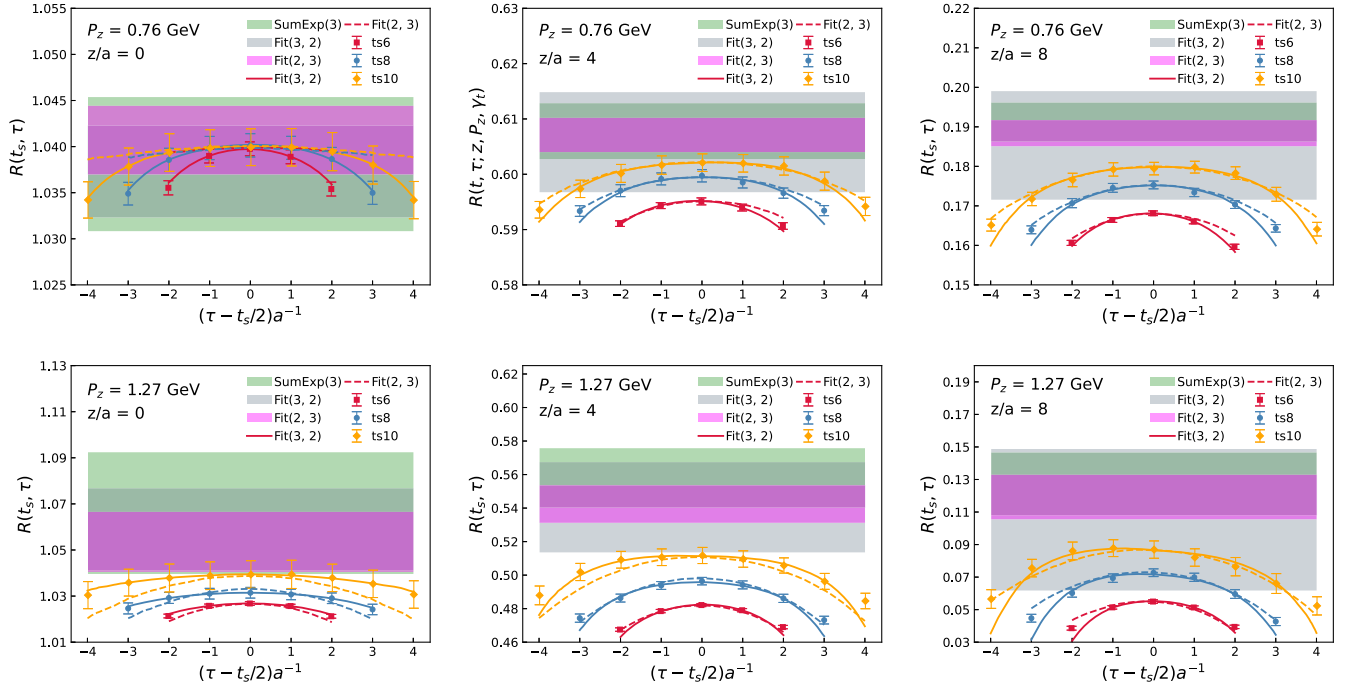$$R(t_s, \tau) = \frac{C_{3pt}(t_s, \tau)}{C_{2pt}^{SS}(t_s)}, \quad (9)$$

FIG. 3.   Ratios $R(t_s, \tau)$ for $n_z = 3$ (upper panels) and 5 (lower panels) of the $m_\pi = 140$ MeV ensemble are shown. The continuous and dashed curves are the central values from Fit(3,2) and Fit(2,3), respectively. The horizontal bands differentiated by their colors are the fit results of SumExp(3), Fit(3,2) and Fit(2,3).

which in the limit $t_s \to \infty$ approaches $R(t_s, \tau) = \langle 0|O_\Gamma|0\rangle$. To take care of the possible wraparound effect, we found it convenient to replace $C_{2pt}^{SS}(t_s)$ with $C_{2pt}^{SS}(t_s) - |A_0|^2 e^{-E_0 t_s}$ using the value of $E_0(P_z)$ from the best estimate of Sec. III. We applied the following two methods to extract the bare matrix elements $h^B(z, P_z)$:

(1) As shown in Eqs. (7) and (8), the spectral decomposition is known for both the three-point and two-point functions. Therefore, we can apply the N-state fits taking the values of $|A_n|^2$ and E_n from the fits of Sec. III to extract the bare matrix elements $h^B(z, P_z)$ of the pion ground state. We will refer to this method by Fit(N, n_{sk}), with N denoting the N-state fit and n_{sk} being the number of τ (time insertions) we skipped on two sides of each time separation.

(2) The sum of the ratios

$$R_{\text{sum}}(t_s) = \sum_{\tau = n_{sk}a}^{t_s - n_{sk}a} R(t_s, \tau) \qquad (10)$$

has fewer excited-state contributions, and its dependence on t_s can be approximated by a linear behavior [75],

$$R_{\text{sum}}(t_s) = (t_s - 2n_{sk}a)h^B(z, P_z)$$
$$+ B_0 + \mathcal{O}(e^{-(E_1 - E_0)t_s}). \qquad (11)$$

However, when t_s is too small, the excited-state contributions cannot be entirely neglected, and one should include the leading-order correction [48] of the form

$$R_{\text{sum}}(t_s) = (t_s - 2n_{sk}a)h^B(z, P_z) + B_0$$
$$+ B_1 e^{-(E_1 - E_0)t_s} + \mathcal{O}(e^{-(E_2 - E_0)t_s}). \qquad (12)$$

We will refer to the above two summation methods by Sum(n_{sk}) and SumExp(n_{sk}) with n_{sk} being the number of τ we skipped.

As discussed in Sec. III, a three-state spectral model can describe two-point functions with $t_s \geq 2a$; thus, Fit(3,2) is justified to do the extrapolation of $R(t_s, \tau)$. In addition, we observed that E_0 from the two-state fit reaches the plateau around $t_{\min} \sim 3a$ and $4a$, where E_1 is consistent within errors as an effective description of the tower of energies larger than E_0. Therefore, we also tried the two-state fit as Fit(2,3) to reduce the number of fit parameters. In Fig. 3, we show the ratios $R(t_s, \tau)$ for $n_z = 3, 5$ with $z/a = 0, 4, 8$ as examples, where the central values of the fit for Fit(3,2) and Fit(2,3) are shown as solid and dashed curves, both describing the $R(t_s, \tau)$ data well within the errors. We tried both summation methods Sum(3) and SumExp(3), as shown in Fig. 4. We show the fit result of the two summation methods and also reconstruct the corresponding bands from the two-state fit Fit(2,3) for comparison. As one can see, SumExp(3) can better describe the data as
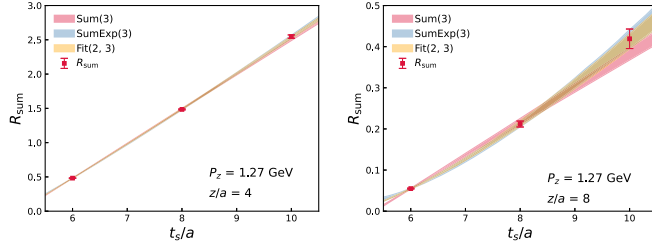
FIG. 4. $R_{\text{sum}}(t_s)$ for $n_z = 5$ with $z/a = 4$ and 8 is shown. The bands are reconstructed from fit strategies Sum(3), SumExp(3), and Fit(2,3).
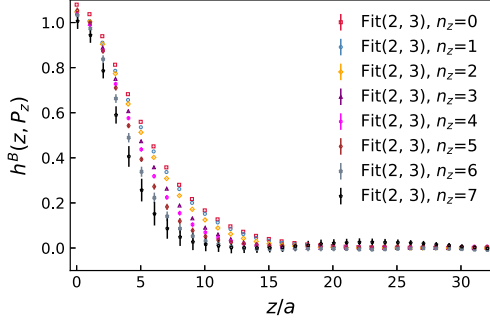

FIG. 5. The bare matrix elements $h^B(z, P_z)$ of the $m_\pi = 140$ MeV ensemble extracted from the Fit(2,3) method are shown for P_z from 0 to 1.78 GeV.

compared to Sum(3), and it is consistent with Fit(2,3), suggesting that the excited-state contamination cannot be totally neglected in the summation methods. Therefore, we only used the corrected summation fit SumExp(n_{sk}) and abandoned Sum(n_{sk}) in our analysis. Finally, we compare the extracted bare matrix elements from the N-state fits and SumExp(3), shown as horizontal bands in Fig. 3, where consistent results can be observed indicating the robustness of our analysis. The results from Fit(3,2) are usually noisier due to having nine fit parameters, whereas Fit(2,3) only has four fit parameters. Since the summation methods are the approximated form of Fit(2,3), we will take the results of Fit(2,3) for our subsequent analysis. To summarize, the bare matrix elements for all momenta using the Fit(2,3) method are shown in Fig. 5. The local bare matrix element $h^B(z = 0, P_z)$, which is the vector current renormalization factor Z_V, has been discussed in Ref. [66].

V. RATIO-SCHEME RENORMALIZATION AND LEADING-TWIST EXPANSION

In Sec. IV, we extracted the bare matrix elements of the pion $h^B(z, P_z, a)$, which then need to be renormalized. It is known that the operator $O_\Gamma(z)$ [cf. Eq. (1)] is multiplicatively renormalizable [76–78],

$$h^B(z, P_z, a) = e^{\delta m|z|} Z(a) h^R(z, P_z, \mu), \quad (13)$$

with $Z(a)$ coming from the fields and vertices renormalization and $e^{\delta m|z|}$ coming from the self-energy divergence of the Wilson line. Nonperturbative renormalization schemes such as RI-MOM [7,8,21,79–83] and the HYBRID scheme [63,84–86] are one possible way to remove the UV divergences. Alternatively, the renormalization factors, including the Wilson line renormalization, cancel out in the ratios of hadron matrix elements evaluated at different momenta, P_z. That is, we construct the renormalization group invariant (RGI) ratios,

$$\mathcal{M}(z, P_z, P_z^0) = \frac{h^B(z, P_z, a)}{h^B(z, P_z^0, a)} = \frac{h^R(z, P_z, \mu)}{h^R(z, P_z^0, \mu)}, \quad (14)$$

where the matrix elements in the rightmost term above are renormalized in the $\overline{\text{MS}}$ scheme due to the ratio being RGI. For $P_z^0 = 0$, the ratio is usually called the reduced Ioffe-time distribution [9,12], and the generalization of such a ratio via the use of nonzero P_z^0 was advocated in Ref. [48]. Because of Lorentz invariance, $\mathcal{M}(z, P_z, P_z^0)$ can be rewritten as $\mathcal{M}(z^2, zP_z, zP_z^0)$, with the term $\lambda = zP_z$ usually referred to as the Ioffe-time in the literature [9]. For small values of z, one can use the leading-twist (twist-2) expansion to relate $\mathcal{M}(z, P_z, P_z^0)$ to the PDF, $q(x, \mu)$, as [9,12,87,88]

$$\mathcal{M}(z, P_z, P_z^0)$$
$$= \frac{\sum_{n=0} c_n(\mu^2 z^2) \frac{(-izP_z)^n}{n!} \langle x^n \rangle(\mu) + \mathcal{O}(\Lambda_{\text{QCD}}^2 z^2)}{\sum_{n=0} c_n(\mu^2 z^2) \frac{(-izP_z^0)^n}{n!} \langle x^n \rangle(\mu) + \mathcal{O}(\Lambda_{\text{QCD}}^2 z^2)}, \quad (15)$$

where $c_n(\mu^2 z^2) = C_n(\mu^2 z^2)/C_0(\mu^2 z^2)$ and where $C_n(\mu^2 z^2)$ are the Wilson coefficients calculated from perturbation theory. The Mellin moments are defined as

$$\langle x^n \rangle(\mu) = \int_{-1}^{1} dx\, x^n q(x, \mu), \quad (16)$$

with $q(x, \mu)$ being the light-cone PDF. The Wilson coefficients have been calculated at NLO [87,89] as well as at NNLO [61,62]. It can be seen that the matrix elements in both numerator and denominator suffer from the higher-twist effects $\mathcal{O}(\Lambda_{\text{QCD}}^2 z^2)$, which are less important when $P_z^0, P_z > \Lambda_{\text{QCD}}$. The ratio-scheme renormalization defined above has the potential to extend the range of z by reducing the higher-twist effects $\mathcal{O}(\Lambda_{\text{QCD}}^2 z^2)$ through the possible cancellation between the numerator and denominator. It was observed in Ref. [48] that the difference in the resulting PDF when using different P_z^0 in the ratio scheme is marginal within the errors. Since the lattice results of $h^B(z, P_z, a)$ for different values of z are strongly correlated in practical calculations, it is convenient to replace the ratio defined by Eq. (14) with the following one:
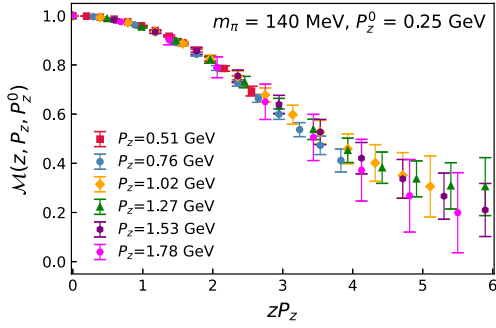
FIG. 6. The ratio scheme renormalized matrix elements $\mathcal{M}(z, P_z, P_z^0)$ with $P_z^0 = 0.25$ GeV are shown for the $m_\pi = 140$ MeV ensemble.

$$\mathcal{M}(z, P_z, P_z^0) = \frac{h^B(z, P_z, a)h^B(0, P_z^0, a)}{h^B(z, P_z^0, a)h^B(0, P_z, a)}. \quad (17)$$

This ratio is equivalent to the one in Eq. (14) in the continuum limit, but it achieves two things. First, it imposes the condition that the value of the $z = 0$ matrix element is momentum independent and therefore removes certain lattice corrections. Second, such a modified ratio has smaller statistical errors because of the correlations between the $z = 0$ and $z \neq 0$ matrix elements. Therefore, we will use Eq. (17) to estimate $\mathcal{M}(z, P_z, P_z^0)$ from our lattice calculations.

The matrix element $h^B(z, P_z)$ could be affected by the finite temporal extent of the lattice, L_t [90–92]. These wraparound effects that are proportional to $e^{-m_\pi L_t}$ could be as large as 3% for $P_z = 0$ and $m_\pi L_t \leq 3.5$. Such wraparound effect was discussed and estimated in Refs. [48,66]. It was found to be important for the zero-momentum case but negligible for the cases of $P_z > 0$. Thus, the use of $P_z^0 \neq 0$ offers yet another practical advantage. For this reason, for the $m_\pi = 140$ MeV ensemble, we consider $P_z^0 = 2\pi/L_s \simeq 0.25$ GeV and omit the $P_z = 0$ lattice data from the analysis in what follows. Finally, in Fig. 6, we show our results for $\mathcal{M}(z, P_z, P_z^0)$ with $P_z^0 = 0.25$ GeV as a function of $\lambda = zP_z$, and different values of P_z.

VI. MODEL-INDEPENDENT DETERMINATION OF EVEN MOMENTS OF THE PION VALENCE QUARK PDF

A. Leading-twist (twist-2) OPE and the fitting method

The leading-twist expansion formula given by Eq. (15) tells us that $\mathcal{M}(z, P_z, P_z^0)$ is sensitive to the moments of the PDF, and these can be extracted by fitting the zP_z and z^2 dependencies of the lattice data for the ratio. As discussed in Sec. II, we are interested in the isovector PDF $q^{u-d}(x)$ of the pion, which is defined on $x \in [-1, 1]$ and obeys $\int_{-1}^{1} dx q^{u-d}(x) = 1$. Under isospin symmetry, the pion isovector PDF q^{u-d} is symmetric with respect to $x = 0$ and is equivalent to the pion valence PDF defined

on $x \in [0, 1]$, namely, $q^{u-d}(|x|) = q^{u-d}(-|x|) = q^v(|x|)$. Because of this, only the terms even in n contribute to Eq. (15). If we assume the positivity of the $\overline{\text{MS}}$ pion valence PDF,[1] we have $\langle x^n \rangle > 0$. Based on positivity, further constraints can also be imposed on the moments as [48]

$$\langle x^{n+2} \rangle - \langle x^n \rangle < 0,$$
$$\langle x^{n+2} \rangle + \langle x^{n-2} \rangle - 2\langle x^n \rangle > 0. \quad (18)$$

To extract the moments of the PDF using Eq. (15), it is necessary to ensure that the leading-twist expansion is a good approximation for the data under consideration. This implies that z cannot be too large, which, given the fact that P_z is less than 3 GeV in present-day lattice calculations, also means that the range in λ is limited. As we see from Fig. 6, our lattice data cover the range up to $\lambda \simeq 6$ with z up to around 0.608 fm. In this range, the sums in Eq. (15) can be truncated to a few terms. As discussed in Appendix A, for realistic pion PDF in this range of λ, the sums can be truncated at $n_{\max} = 8$. Here, we note that there are also higher-twist terms entering the leading-twist formula that are proportional to $(m_\pi^2 z^2)^n$, known as target mass corrections. The target mass corrections can be taken care of by the following replacement [94,95]:

$$\langle x^n \rangle \rightarrow \langle x^n \rangle \sum_{k=0}^{n/2} \frac{(n-k)!}{k!(n-2k)!} \left(\frac{m_\pi^2}{4P_z^2} \right)^k. \quad (19)$$

These corrections are small in the case of the pion. We have to ensure that the higher-twist corrections beyond the target mass corrections are also small. Furthermore, we need to ensure that the perturbative expression for c_n is also reliable. To do this, we fit the λ dependence of $\mathcal{M}(z, P_z, P_z^0)$ at fixed z and check to what extent the obtained moments are independent of z. We perform fits at each z (six data points with different P_z) by truncating the moments up to $n^{\max}$. To stabilize the fits, we impose constraints given by Eq. (18). We tried $n^{\max} = 6$ and 8, where reasonable $\chi^2_{\text{d.o.f}} \lesssim 1$ can always be found when $z \geq 2a$ for $n^{\max} = 6$. As for $n^{\max} = 8$, we found $\langle x^8 \rangle$ is always consistent with 0. We therefore fix $n^{\max} = 6$ for the following discussion in this section. We use leading-order (LO), NLO, and NNLO results for c_n and fix the scale μ to 2 GeV. The coupling constant $\alpha_s(\mu)$ entering $c_n(\mu^2 z^2)$ at NLO and NNLO is evolved from $\alpha_s(\mu = 2 \text{ GeV}) = 0.293$, which is obtained from $\Lambda_{\text{QCD}}^{\overline{\text{MS}}} = 332$ MeV with the five-loop β function and $n_f = 3$, as has been calculated using the same lattice ensembles [96].

In Fig. 7, we show the second moment $\langle x^2 \rangle$ extracted from each fixed z using LO, NLO, and NNLO matching kernels

---

[1] We are aware of Ref. [93], which shows that positivity is not a necessary constraint for $\overline{\text{MS}}$ PDFs (also see references therein). Nevertheless, we still use positivity as a constraint for pion valence quark PDF in our analysis.
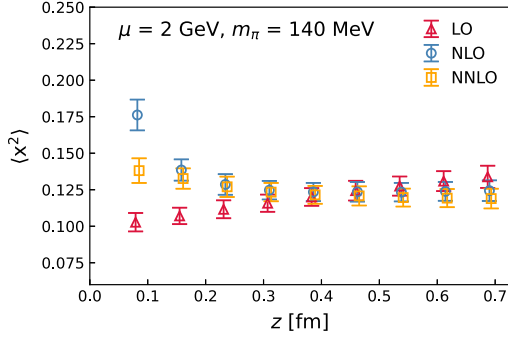
FIG. 7. $\langle x^2 \rangle$ extracted from the ratio-scheme data $\mathcal{M}(z, P_z, P_z^0)$ at each z using leading-order, NLO, and NNLO kernels is shown for the $m_\pi = 140$ MeV ensemble.

for the $a = 0.076$ fm ensemble. Clear z dependence can be observed at LO. Beyond LO, the perturbative kernels are supposed to compensate the z dependence and produce the z-independent plateaus of $\langle x^2 \rangle(\mu)$. We see from the figure that for $z > 0.3$ fm the NLO and NNLO results show no z dependence within errors, suggesting that the leading-twist approximation is reliable in this z range. On the other hand, for smaller z, we see a clear dependence on z, and there is a clear difference between the NLO and NNLO result. This is counterintuitive, as one expects the short-distance factorization to work better for small z. To understand this, we performed additional studies using the results for $\mathcal{M}(z, P_z, P_z^0)$ calculated on the $a = 0.04$ fm and $a = 0.06$ fm ensembles. We varied the renormalization scale roughly by a factor of $\sqrt{2}$, i.e., we used $\mu = 1.4$ GeV and $\mu = 2.8$ GeV, and also performed an analysis where the coupling constant was evolved from the scale $\mu = 2$ GeV to a scale proportional to $1/z$, which corresponds to resumming logs, $\ln(\mu^2 z^2)$. These additional calculations are discussed in detail in Appendix B. There, it is also pointed out that for $z/a \leq 2$ the extractions of the moments are affected by lattice discretization effects. The conclusion of the analysis presented in Appendix B is that the resummation of logs is important for $z < 0.1$ fm and the resummed expressions for $c_n(\mu^2 z^2)$ work well for $z \leq 0.3$ fm. For large z, the resummed results for c_n are not appropriate as the scale in the running coupling constant becomes too low. However, the moments obtained using the resummed result for c_n and $z < 0.3$ fm agree with the ones obtained with the fixed-order result and $z > 0.3$ fm. Therefore, it seems the fixed-order NNLO leading-twist approximation can describe the ratio-scheme data at least up to 0.7 fm, though it was not expected from the beginning. But this allows us to determine higher moments of the pion PDF and constrain the pion PDF in general, as we discuss later.

B. First few moments from a combined fit

To stabilize the fit and extract the higher moments, we perform a combined fit over a range of ratio scheme data in

$[z_{\min}, z_{\max}]$. As has been mentioned, we will use the ratio scheme data $\mathcal{M}(z, P_z, P_z^0)$ constructed from $P_z > P_z^0 = 0.25$ GeV. To take care of possible cutoff effects, we apply the modified form [48],

$$\mathcal{M}(z, P_z, P_z^0) = \frac{\sum_n c_n(\mu^2 z^2) \frac{(-iz P_z)^n}{n!} \langle x^n \rangle + r(a P_z)^2}{\sum_n c_n(\mu^2 z^2) \frac{(-iz P_z^0)^n}{n!} \langle x^n \rangle + r(a P_z^0)^2}. \quad (20)$$

The fit results are shown in Fig. 8 for the 0.076 fm lattice as a function of $z_{\max}$, while the other two finer lattices have been analyzed in Ref. [48], though at NLO level. In the fits, we truncate the moments up to $n^{\max} = 8$ and fix $z_{\min}$ to be $2a$ to avoid the most serious discretization effect at the first lattice grid a. We vary $z_{\max}$ to check the stability of the fit and its dependence on the range of z. With a fixed factorization scale $\mu = 2$ GeV, we find that the second moment $\langle x^2 \rangle$ can be extracted at a very short distance ≈ 0.2 fm and is almost independent of $z_{\max}$. This suggests a good predictive power of the NNLO matching coefficients in the region of z under consideration and that the higher-twist effects or other systematic uncertainty are under control within current statistics. Because of the factorial suppression, the higher moments can only be detected at larger $z P_z$ or $z_{\max}$ as seen from the figure. To estimate the statistical as well as systematic errors from the fit results, we use the same strategy proposed in Ref. [48]: for each bootstrap sample, we evaluate the average and standard deviation of observable A from a range of $z_{\max}$ as $\overline{\text{Mean}(A)}$ and $\overline{\text{SD}(A)}$. Then, we can obtain the statistical errors of $\overline{\text{Mean}(A)}$ and take $\overline{\text{SD}(A)}$ as the systematic error. The estimates using $z_{\max} \in [0.48, 0.72]$ fm are shown as the dark (statistical errors) and light (systematic errors) bands in Fig. 8. We list the moments up to $\langle x^6 \rangle$ extracted from the three ensembles in Table II, while $\langle x^8 \rangle$ is consistent with 0, limited by the statistics, and thus is not shown. As one can see, overall agreement can be observed for different
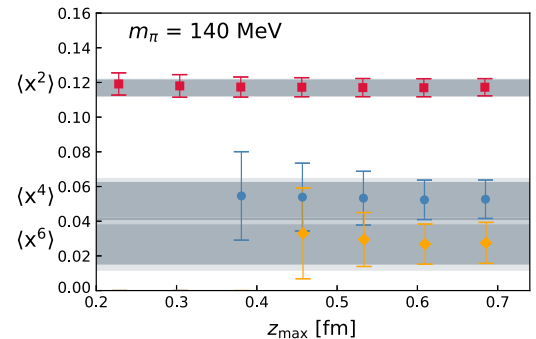

FIG. 8. $\langle x^n \rangle$ with $n = 2, 4, 6$ extracted from the combined fit are shown for the $m_\pi = 140$ MeV lattice. The darker and lighter bands correspond to the statistical and systematic errors. The fits use the ratio-scheme data $\mathcal{M}(z, P_z, P_z^0)$ for $z \in [2a, z_{\max}]$ and $P_z > P_z^0 = 0.25$ GeV.

TABLE II. $\langle x^n \rangle$ with $n = 2, 4, 6$ extracted from the combined fit are shown for the three ensembles at $\mu = 2$ GeV. The statistical errors are in the first brackets, while the systematic errors are in the second brackets, which are estimated by varying $z_{max} \in [0.48, 0.72]$ fm for the fits.

a (fm)	$\langle x^2 \rangle$	$\langle x^4 \rangle$	$\langle x^6 \rangle$	$\chi^2_{\text{d.o.f.}}$
0.076	0.1169(49)(03)	0.0516(110)(19)	0.0267(118)(28)	0.91
0.06	0.1159(35)(09)	0.0385(61)(32)	0.0139(73)(41)	1.4
0.04	0.1108(38)(05)	0.0396(46)(18)	0.0146(43)(16)	0.63
$a^2 \to 0$	0.1104(73)(48)	0.0388(46)(57)	0.0118(48)(48)	1.3

ensembles including the most precise $\langle x^2 \rangle$ with about 5% accuracy. It is worth mentioning that the $a = 0.04$ fm and 0.06 fm ensembles have an unphysical pion mass of 300 MeV, while the $a = 0.076$ fm one is at the physical point, suggesting the mass dependence of the moments is only mild within current statistics.

C. Continuum estimate of the moments making use of the observed weak quark mass dependence

In Fig. 9, we show the $\langle x^2 \rangle$, $\langle x^4 \rangle$, and $\langle x^6 \rangle$ obtained from fits of $\mathcal{M}(z, P_z, P_z^0)$ in $z \in [2a, 0.61]$ fm as a function of a^2. Notice that we have combined the data from the physical pion and the 300 MeV pion together in this plot. As one can see, the mild lattice spacing dependence shows almost no pion mass dependence. It is then reasonable to perform continuum extrapolations assuming a or a^2 dependence under a justified assumption that we can neglect the pion mass dependence. Based on the above observation, we consider two strategies for obtaining the continuum estimate:

(1) Ignoring the pion mass dependence and performing a continuum extrapolation using the following forms:

$$\langle x^n \rangle_a = \langle x^n \rangle_{a \to 0} + d_n a, \tag{21}$$

or

$$\langle x^n \rangle_a = \langle x^n \rangle_{a^2 \to 0} + d_n a^2. \tag{22}$$

We insert the above formulas into Eq. (20) and perform a joint fit of all the data from the three ensembles instead of directly extrapolating the extracted moments.

(2) We only perform the continuum extrapolation of the two ensembles with a 300 MeV pion mass to obtain the value of d_n and then apply the parameter d_n to the physical pion ensemble ($m_\pi = 140$ MeV) to derive the continuum estimate at the physical point. In this procedure, we assume $\langle x^n \rangle_{a^2 \to 0}$ could have pion mass dependence, but the values of d_n have negligible pion mass dependence.

The bands in Fig. 9 are derived from the joint fit of all the three ensembles by applying Eqs. (21) (blue) and (22) (orange) and ignoring the pion mass difference. As one can see, the two bands overlap with each other and pass through the data points with reasonable χ^2/d.o.f. around 1. The pion mass dependence is indeed only mild for the data under consideration as also observed in Ref. [3]. To be conservative, we also apply the second strategy described above. The extrapolated results are shown as the boxes in Fig. 9. Limited by the number of ensembles and the statistics, solving both the lattice spacing and mass dependence makes the extrapolation rather unstable and produces large errors bars covering the estimates from the first strategy. We therefore will simply ignore the pion mass dependence. Considering that the results from Eq. (21) overlap with Eq. (22) with only a slightly larger error and the fact that the Wilson-clover action is $\mathcal{O}(a)$ improved, we will only give the mass-independent continuum estimate using the a^2 correction of Eq. (22) in the following analysis.
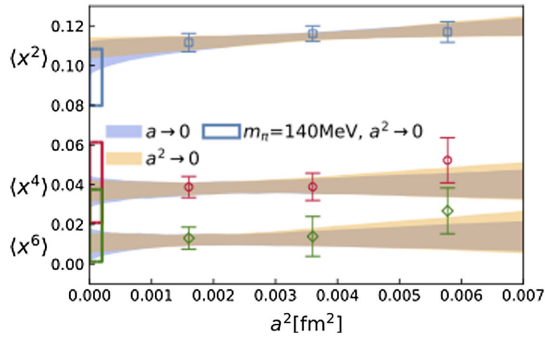

FIG. 9. $\langle x^2 \rangle$, $\langle x^4 \rangle$, and $\langle x^6 \rangle$ are shown as a function of a^2, for which we fitted the matrix elements $\mathcal{M}(z, P_z, P_z^0)$ with $z \in [2a, 0.61]$ fm. The bands are the extrapolation results using Eq. (21) (blue) and Eq. (22) (orange). The empty boxes are the continuum extrapolation of the physical pion mass ensemble (see the text for more details).
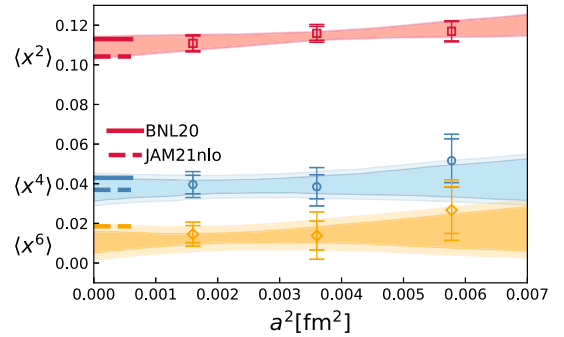

FIG. 10. Our estimates for $\langle x^2 \rangle$, $\langle x^4 \rangle$, and $\langle x^6 \rangle$ as functions of a^2 are shown for the three ensembles. The bands are from the a^2 continuum extrapolation. For comparison, we also show the moments evaluated from the global fit analysis of JAM21nlo [97] and our previous estimates with a 300 MeV pion mass using NLO kernels (BNL20) [48].

In Fig. 10, we show the moments extracted from the three ensembles with both statistical and systematic errors, estimated by varying $z_{max} \in [0.48, 0.72]$ fm. The bands are the continuum estimate using Eq. (22). It can be observed that the systematic errors are small compared to the statistical errors, suggesting the small z_{max} dependence for the data under consideration. Our previous results obtained from NLO kernels with a 300 MeV pion (denoted by BNL20 [48]) are shown for comparison. The good agreement between BNL20 and the new results suggests that the NNLO corrections make only a small difference, and the NLO kernels are mostly sufficient to describe the data evolution with current statistics. The estimated moments are also close to the values obtained from the global fit, JAM21nlo [97].

VII. PION VALENCE PDF FROM MODEL DEPENDENT FITS

As discussed in Sec. VI, our lattice data are only sensitive to the first few moments of the pion valence PDF. Without prior knowledge or constraints on the higher moments, it is impossible to determine the PDFs uniquely. Therefore, as is typical in some global analyses of experiment data, in this section, we try two phenomenology inspired model Ansätze for the PDF to fit the lattice data,

$$q(x; \alpha, \beta) = \mathcal{N} x^\alpha (1 - x)^\beta,$$
$$q(x; \alpha, \beta, s, t) = \mathcal{N}' x^\alpha (1 - x)^\beta (1 + s\sqrt{x} + tx), \quad (23)$$

denoted by Model-$2p$ and Model-$4p$, in which $\mathcal{N}$ and $\mathcal{N}'$ are normalization factors so that $\int_0^1 q^v(x) dx = 1$. The moments of the PDF therefore can be determined from the model parameters, e.g.,

$$\langle x^n \rangle_{(\alpha,...)} = \int_0^1 x^n q^v(x; \alpha, ...) dx. \quad (24)$$

We then reexpress Eq. (20) using the model parameters and minimize

$$\chi^2 = \sum_{P_z > P_z^0}^{P_z^{max}} \sum_{z_{min}}^{z_{max}} \frac{(\mathcal{M}(z, P_z, P_z^0) - \mathcal{M}_{model}(z, P_z, P_z^0; \alpha, ...))^2}{\sigma^2(z, P_z, P_z^0)},$$

$$(25)$$

where we truncate the OPE formula Eq. (20) up to 20th order ($n^{max} = 20$), which is much more than sufficient to describe the data. During the fit, the factorization scale was chosen to be $\mu = 2$ GeV, and the correlation between different P_z and z is taken care of by the bootstrap procedure. To apply the leading-twist approximation, we need to limit the maximum of z, which is chosen to be $z_{max} = 0.72$ fm as was done in Sec. VI.

We first perform the model fit for the $m_\pi = 140$ MeV ensemble using the matrix elements with $z \in [2a, 0.61$ fm]
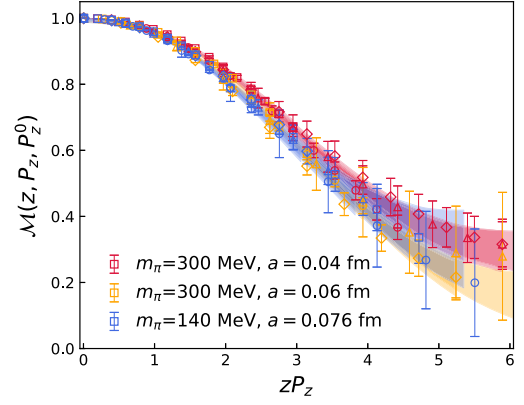

FIG. 11. Ratio scheme renormalized matrix elements $\mathcal{M}(z, P_z, P_z^0)$ of $P_z > P_z^0 = 0.25$ GeV are shown for the three ensembles, with the bands being the fit results of the Model-$4p$ [cf. Eq. (23)] using $z \in [2a, 0.61$ fm].

shown as the black curves in Fig. 12. Then, as with the calculations of the Mellin moments, we perform the mass-independent continuum estimate by a joint fit of the three ensembles with the $\mathcal{O}(a^2)$ correction of Eq. (22) using the ratio-scheme renormalized matrix elements with $P_z > P_z^0 = 2\pi n_z^0/(L_s a)$ and $n_z^0 = 1$ for all three ensembles. The fit result of Model-$4p$ using $z_{max} = 0.61$ fm is shown in Fig. 11. We see that the bands can describe the data well. We vary the $z_{max} \in [0.48, 0.72]$ fm to estimate the systematic errors and show the fit results in Table III and the reconstructed PDFs in Fig. 12 as the red and blue bands. Overall consistency between the two models can be observed in Fig. 12. Furthermore, the results overlap with our previous NLO determination (BNL20) [48] but have smaller errors since we have an additional dataset.
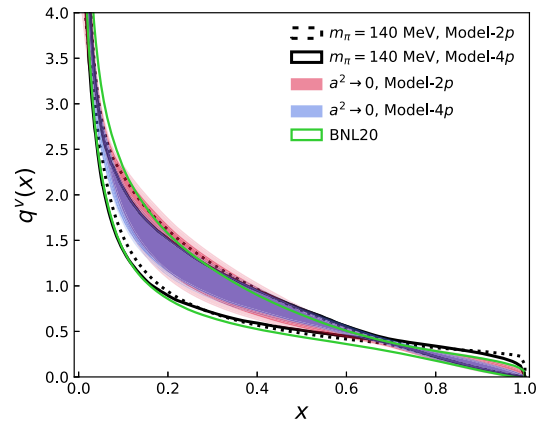

FIG. 12. The two-parameter and four-parameter fit results from the $m_\pi = 140$ MeV ensemble are shown as empty bands. We also show our mass-independent continuum estimate as the filled bands using the data with $z_{max} \in [0.48, 0.72]$ fm to estimate the statistical errors (darker bands) and systematic errors (light bands). For comparison, we also show our previous NLO determination (BNL20) [48].

TABLE III.   The model parameters from the joint fit using the three ensembles with $\mathcal{O}(a^2)$ continuum correction at $\mu = 2$ GeV are shown. The statistical errors are in the first brackets, while the systematic errors are in the second brackets, which are estimated by varying $z_{max} \in [0.48, 0.72]$ fm for the fits.

	α	β	s	t
Model-2p	$-0.45(18)(6)$	$0.79(42)(14)$		
Model-4p	$-0.52(12)(3)$	$1.08(33)(8)$	$-0.34(33)(6)$	$1.82(99)(11)$

The good agreement with the previous results again suggests that the use of NNLO kernels did not change the results significantly.

VIII. DNN REPRESENTATION OF IOFFE-TIME DISTRIBUTION $Q(\lambda,\mu)$

A. DNN representation

The short-distance factorization formula for the renormalized matrix element at leading twist (twist 2) can be written as

$$h^R(z, P_z, \mu) = \int_{-1}^{1} d\alpha \, \mathcal{C}(\alpha, \mu^2 z^2) Q(\alpha\lambda, \mu), \quad (26)$$

where

$$Q(\lambda, \mu) = \int_{-1}^{1} dy \, e^{-iy\lambda} q(y, \mu) \quad (27)$$

and $\mathcal{C}(\alpha, \mu^2 z^2)$ is related to the Wilson coefficients $C_n(\mu^2 z^2)$ [9,12,87]. The explicit form of $\mathcal{C}(\alpha, \mu^2 z^2)$ up to NNLO is given in Appendix D. Therefore, for the ratio scheme matrix element, we could write

$$\mathcal{M}(z, P_z, P_z^0) = \frac{\int_{-1}^{1} d\alpha \bar{\mathcal{C}}(\alpha, \mu^2 z^2) Q(\alpha\lambda, \mu) + r(aP_z)^2}{\int_{-1}^{1} d\alpha \bar{\mathcal{C}}(\alpha, \mu^2 z^2) Q(\alpha\lambda^0, \mu) + r(aP_z^0)^2}, \quad (28)$$

where $\bar{\mathcal{C}}(\alpha, \mu^2 z^2) = \mathcal{C}(\alpha, \mu^2 z^2) / C_0^{\overline{MS}}(\mu^2 z^2)$ so that the integral (without the lattice correction $r(aP_z)^2$) in the numerator and denominator is the standard reduced Ioffe-time distribution [9,12]. As mentioned in previous sections, the range in $(\alpha\lambda)$ in the above equation depends on the largest $\lambda_{max} = z_{max} P_z^{max}$ achieved from the lattice calculation, which therefore limits the information on the PDFs that the lattice data carry. One can either reconstruct the PDFs $q(y, \mu)$ by a certain model as we did in Sec. VII, or one can directly extract the light-cone Ioffe-time distribution $Q(\lambda, \mu)$. In other words, $Q(\lambda, \mu)$ is the first-principles model-independent observable we can derive from the ratio scheme renormalized matrix elements. For this, however, we need to solve the inverse problem in Eq. (28). So far, the DNN technique is probably the most flexible way to achieve this [98,99] and has been used to parametrize the x-dependent PDFs from lattice [100,101]

and also has been proven effective in other physics inverse problems [102,103].

In this work, we express the $Q(\lambda, \mu)$ by the DNN function $f_{DNN}(\boldsymbol{\theta}; \lambda)$,

$$Q_{DNN}(\lambda, \mu) \equiv \frac{f_{DNN}(\boldsymbol{\theta}; \lambda)}{f_{DNN}(\boldsymbol{\theta}; 0)}, \quad (29)$$

where $\boldsymbol{\theta}$ are the DNN parameters. The DNN function $f_{DNN}(\boldsymbol{\theta}; \lambda)$ is a multistep iterative function, constructed layer by layer in composite fashion, to approximate a mapping between two functions in a smooth and unbiased manner. In each layer, the network first performs a linear transformation from the previous layer,

$$z_i^{(l)} = b_i^{(l)} + \sum_j W_{ij}^{(l)} a_j^{(l-1)}, \quad (30)$$

followed by an elementwise nonlinear activation $a_i^{(l)} = \sigma^{(l)}(z_i^{(l)})$ and then propagates to the next layer $(l + 1)$. The first layer represents the input variable λ, and the last layer denotes the corresponding output, $Q_{DNN}(\lambda, \mu)$. Here, $i = 1, \ldots, n^{(l)}$ and $l = 1, \ldots, N$, with $n^{(l)}$ being the width of the lth layer and N being the depth of the DNN. The bias $b_i^{(l)}$ and the weight $W_{ij}^{(l)}$ are the DNN parameters to be optimized (trained), denoted by $\boldsymbol{\theta}$. We then minimize the loss function,

$$J(\boldsymbol{\theta}, r_{mod}) \equiv \frac{\eta}{2}\boldsymbol{\theta} \cdot \boldsymbol{\theta} + \frac{1}{2}\chi^2(\boldsymbol{\theta}, r_{mod}), \quad (31)$$

where the first term is to prevent overfitting and make sure the DNN represented function is well behaved and smooth, while the definition and details of χ^2 can be found inAppendix D. For the training, we vary η from 10^{-1} to 10^{-4} and tried network structures of size $\{1, 16, 16, 1\}$, $\{1, 16, 16, 16, 1\}$, and $\{1, 32, 32, 1\}$ including the input and output layers. We found the results remain unchanged. We therefore chose $\eta = 0.001$ and the DNN structure with four layers, including the input/output layer, to be $\{1, 16, 16, 1\}$. The exponential linear units (elu) were chosen as the action function,

$$\sigma_{elu}(z) = \theta(-z)(e^z - 1) + \theta(z)z. \quad (32)$$

This setup is more than sufficient for the complexity of the data under consideration.

The DNN training process works very much like an interpolation procedure, which is trying to go through the data points as much as possible with many neurons, but smoothly forced by the regularization term $\eta\theta^2$. In other words, the χ^2 of DNN is approaching 0 as much as possible. For our specific task of training the ratio-scheme renormalized matrix elements, part of our data points share the same $\lambda = zP_z$ but different z connected by the perturbative matching kernel $\mathcal{C}(\alpha, \mu^2 z^2)$. The minimum of χ^2 is free of the goodness of model but will depend on how well the kernel can describe the data evolution and usually cannot make it to be zero. In the upper panel of Fig. 13, we show the $\mathcal{M}(z, P_z, P_z^0)$ obtained from the DNN representation of the light-cone Ioffe-time distribution, $Q_{\text{DNN}}(\lambda, \mu)$. In these

training processes, we use the ratio-scheme renormalized matrix elements on $z \in [2a, 0.61 \text{ fm}]$. As one can see, the bands smoothly go through the ratio-scheme data. The results for $Q_{\text{DNN}}(\lambda, \mu)$ are shown in the lower panel of Fig. 13. For comparison, we also reconstruct and show the $\overline{\text{MS}}$ ITD from the moments estimated in Sec. VI (green curves),

$$Q_{\langle x^n \rangle}(\lambda, \mu) = \sum_{n=0} \frac{(-i\lambda)^n}{n!} \langle x^n \rangle, \qquad (33)$$

and the model fits in Sec. VII (blue and yellow curves),

$$Q_{ab}(\lambda, \mu) = \int_0^1 dx \, e^{-ix\lambda} q^v(x; \alpha, \beta),$$

$$Q_{abst}(\lambda, \mu) = \int_0^1 dx \, e^{-ix\lambda} q^v(x; \alpha, \beta, s, t). \qquad (34)$$

The Ioffe-time distributions (ITDs) from all the methods are consistent with each other within the errors. These observations justify the assertion that our data are only sensitive to the first few moments (up to around $\langle x^6 \rangle$ as discussed in Sec. VI), though $Q_{\langle x^n \rangle}$ goes out of control rapidly once λ is beyond $\lambda_{\max} = z_{\max} P_z^{\max}$ due to the truncation of the OPE formula. The $Q_{ab}(\lambda, \mu)$ and $Q_{abst}(\lambda, \mu)$ are more stable by expressing all the Mellin moments with only a few parameters and modeling the large λ behavior beyond $\lambda_{\max} = z_{\max} P_z^{\max}$. The advantage of the DNN is that it does not truncate the matching formula compared to the moments fit, particularly for the case in which the moments are not decaying as fast as a function of order n, and is much more flexible than the model-based fit. $Q_{\text{DNN}}(\lambda, \mu)$ with $\lambda_{\max} = z_{\max} P_z^{\max}$ is therefore the most unbiased first-principles result on the ITD that can be obtained from our lattice data in coordinate space. We also show $Q(\lambda, \mu)$ derived from the JAM global analysis results [97] in Fig. 13, which are in good agreement with our results.

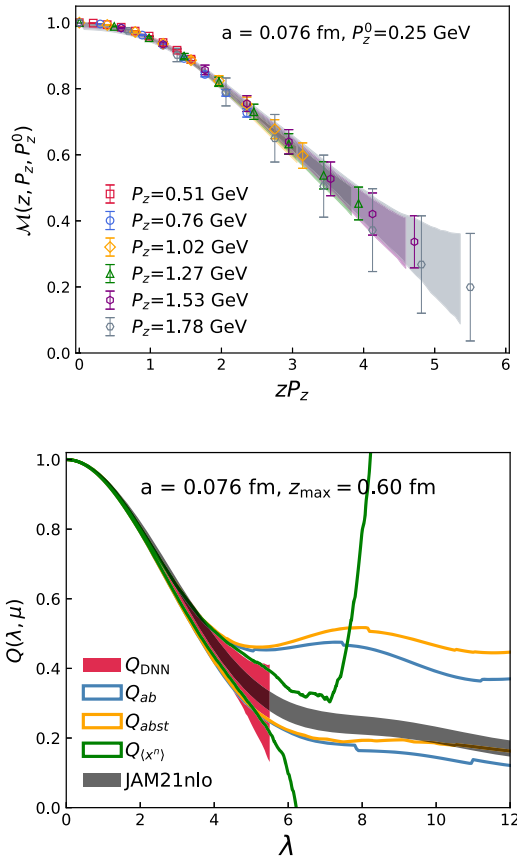
FIG. 13. Upper panel: ratio scheme renormalized matrix elements $\mathcal{M}(z, P_z, P_z^0)$ with $P_z > P_z^0 = 0.25$ GeV for the $m_\pi = 140$ MeV ensemble, where the bands are the fit result from the DNN trained $Q(\lambda, \mu)$. Lower panel: the corresponding DNN represented $Q_{\text{DNN}}(\lambda, \mu)$ is shown, together with the ones constructed from the moments estimated in Sec. VI ($Q_{\langle x^n \rangle}$ as green curves) and the model fits in Sec. VII (Q_{ab} and Q_{abst} as blue and yellow curves). All the training and fits in this figure used data on $z \in [2a, 0.61 \text{ fm}]$. The most recent global analysis result from JAM [97] (black bands) is also shown for comparison.

B. Discussion on the $z_{\max}$ dependence

As has been mentioned, the short-distance factorization scheme suffers from higher-twist corrections proportional to $\mathcal{O}(z^2 \Lambda_{\text{QCD}}^2)$. On the one hand, we want larger z to achieve larger zP_z for a fixed range of P_z values and thereby obtain more information on the PDF from the lattice data. At the same time, we have to ensure that the higher-twist contamination is small. In this section, we discuss how $z_{\max}$ affects the DNN trained results $Q_{\text{DNN}}(\lambda, \mu)$.

In Fig. 14, we show $Q(\lambda, \mu)$ trained from the DNN using data in the range $z \in [2a, z_{\max}]$ with multiple $z_{\max}$. The end of the bands depends on the $z_{\max} P_z^{\max}$ used in the training process. As one can see, increasing $z_{\max}$ has little effect on $Q(\lambda, \mu)$ for small λ but does decrease the errors.
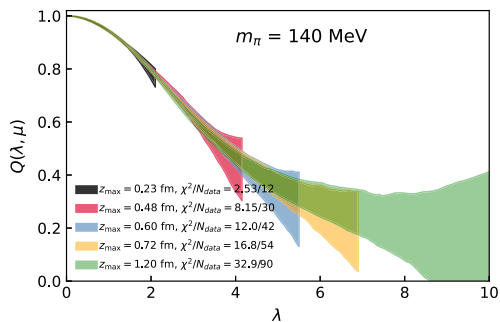
FIG. 14. $Q(\lambda, \mu)$ trained from the DNN using data with $z \in [2a, z_{max}]$ on the $m_\pi = 140$ MeV ensemble. The end of the bands depends on $z_{max} P_z^{max}$.

This observation can be explained from the small λ region of $Q(\lambda, \mu)$, which is dominated by the lower moments that are well constrained by the precise ratio-scheme lattice data with small z. The χ^2/N_{data} are also shown in Fig. 14. As we mentioned, the matching kernel $\bar{C}(\alpha, \mu^2 z^2)$ is supposed to connect the spacelike matrix elements with different z perturbatively. Since the DNN is a flexible parametrization that could pass through all the data points smoothly as much as possible, the nonzero χ^2/N_{data} can only come from the fact that the central values of the lattice data corresponding to the same $z P_z$ but different z are slightly different from the expectation from perturbative evolution. Therefore, the small χ^2/N_{data} indicates that the kernel describes the evolution of the matrix elements as a function of z well within the statistical errors. We also see that the $Q(\lambda, \mu)$ obtained with $z_{max} = 1.2$ fm agrees well with the results obtained with smaller z_{max} within errors. Thus, it is natural to ask how large z_{max} in the DNN analysis could be. To check this, we determine $Q(\lambda, \mu)$ with very small z_{max} but large P_z and then predict the ratio scheme matrix elements for smaller P_z but larger z through the matching. This prediction is then compared to the corresponding lattice result. In Fig. 15, we show the ratio-scheme matrix elements $\mathcal{M}(z, P_z, P_z^0)$ of $P_z = 0.51$ GeV (upper panel) and 0.76 GeV (lower panel) as the black data points, and the bands are predicted from the DNN trained results using P_z up to $P_z^{max} = 1.78$ GeV. We use the data in the range $z \in [2a, z_{max}]$ with short distances $z_{max} = 0.23$ (black) and 0.31 (red) fm. It is interesting to observe that the predicted bands from short distances are consistent with the matrix elements at relatively large distances, suggesting that the matching kernel works well up to $z \sim 0.8$ fm from the prediction of $z_{max} = 0.23$ fm and up to $z \sim 1$ fm from the prediction of $z_{max} = 0.31$ fm within our current statistics. This observation is consistent with what we observed in Sec. VI for the moments extraction and seems to support the argument that the ratio scheme indeed reduces the higher-twist effect $\mathcal{O}(z^2 \Lambda_{QCD}^2)$ by the cancellation between the numerator and denominator, even though it is naively not expected that the leading-twist
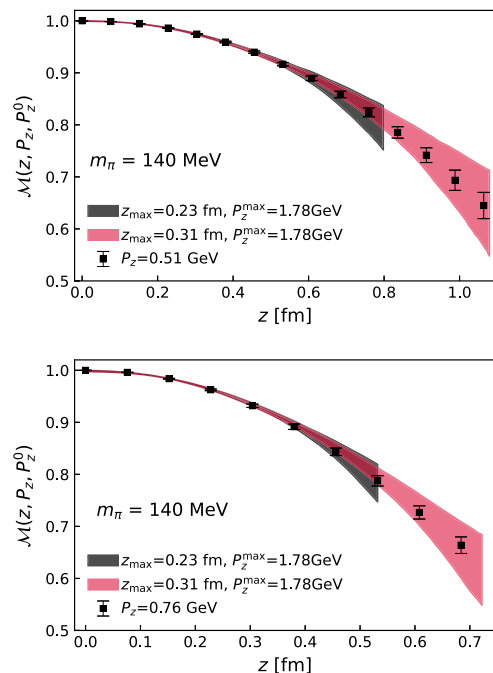

FIG. 15. The ratio-scheme renormalized matrix elements $\mathcal{M}(z, P_z, P_z^0)$ with $P_z = 0.51$ and 0.76 GeV are shown as the data points. The bands are the prediction from the DNN trained results using P_z up to $P_z^{max} = 1.78$ GeV with a short range of z_{max}. The end of the bands in z comes from $z_{max} P_z^{max}/P_z$.

OPE can approximately work up to 1 fm. For the following analysis, we conservatively use z only up to 0.72 fm and vary z_{max} between 0.48 and 0.72 fm to estimate the systematic errors.

C. Discussion on the perturbative order dependence

In this subsection, we explore the perturbative order dependence through the DNN approach, which does not need a truncation of the OPE matching formula. In Fig. 16, the ratio-scheme matrix elements together with the curves reconstructed from the central values of the DNN trained results using LO, NLO, and NNLO matching kernels are shown in the upper panel. It is obvious that the LO matching kernel shows the worst performance among the curves, which only marginally passes within the error bars of the data points. Beyond LO, the curves tend to pass through the central values of the data with much smaller χ^2, though the difference between NNLO and NLO is only mild. This indicates that the leading-twist approximation with fixed-order perturbative kernels works better than LO for the data under consideration, while with current statistics the systematic improvement from NLO to NNLO is small. The distributions for the χ^2 of the bootstrap samples are shown in the lower panel of Fig. 16, in which the χ^2 for the LO fit is significantly larger than the other two cases.
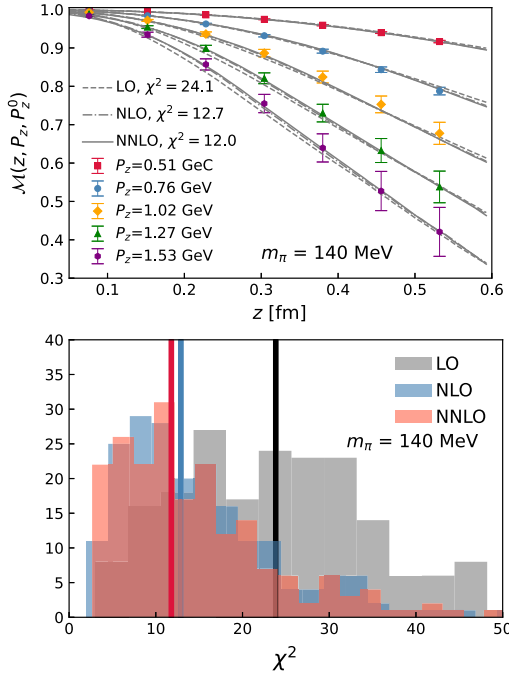
FIG. 16. Upper panel: $\mathcal{M}(z, P_z, P_z^0)$ reconstructed from the DNN trained $Q(\lambda, \mu)$ are shown for the $m_\pi = 140$ MeV ensemble using LO, NLO, and NNLO matching kernels. The central values for the χ^2 of the bootstrap samples are shown in the legends. The fits use $z \in [2a, z_{max}]$ with $z_{max} = 0.61$ fm. Lower panel: the corresponding distribution for the χ^2 of the bootstrap samples with the vertical lines being the median values.

D. Extrapolation of the $Q_{DNN}(\lambda,\mu)$ and FT to the PDFs

The $Q_{DNN}(\lambda, \mu)$ trained from the DNN is the most unbiased and first-principles physical quantity we extracted from the ratio-scheme matrix elements. It is more flexible than the truncated moments extraction (cf. Sec. VI) and the model-based fits (cf. Sec. VII). However, usually one is more interested in the PDF $q(x, \mu)$. The $Q_{DNN}(\lambda, \mu)$ we obtained is limited by the lattice data up to $\lambda_{max} = z_{max} P_z^{max}$, so we need to extrapolate the large λ behavior to infinity to do the Fourier transform (FT).

Inspired by Regge theory, the asymptotic behavior of $Q(\lambda, \mu)$ is dominated by the power law (PL) form,

$$Q_{PL}(\lambda, \mu) = \frac{A}{|\lambda|^{\alpha+1}}. \qquad (35)$$

Limited by the λ_{max} we have, the subleading contribution could also be significant, and therefore this extrapolation has systematic errors from the parametrization. We fit the tail of $Q_{DNN}(\lambda, \mu)$ with $\lambda \in [\lambda_L, \lambda_{max}]$ and vary λ_L to see the λ_L dependence of the extrapolation results. We also impose the constraint $A > 0$ during the extrapolation to ensure the extrapolated ITD is positive and decreasing in λ. Inspired by the asymptotic behavior of the Fourier transform of the two-parameter model $\mathcal{N}x^\alpha(1-x)^\beta$, we also consider the form [63,84],
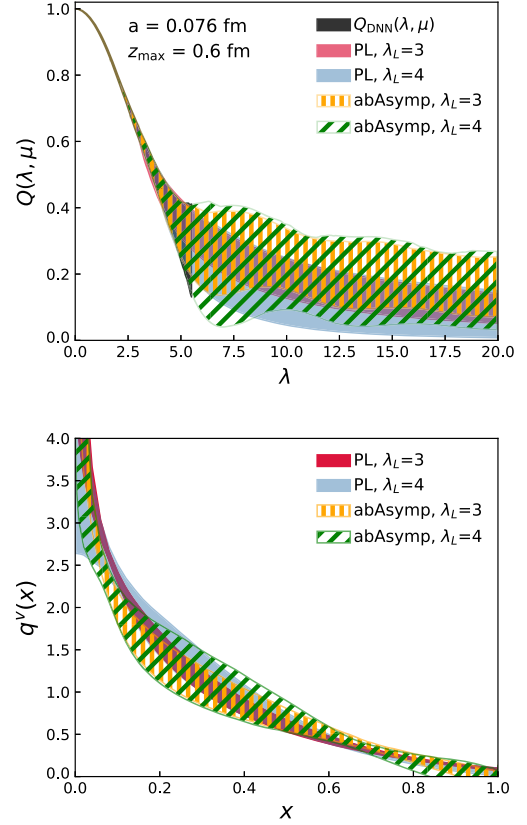
FIG. 17. Upper panel: the DNN trained $Q_{DNN}(\lambda, \mu)$ including the extrapolations using either the power law (denoted by PL) model or the two-parameter asymptotic (denoted by abAsymp) model. Lower panel: the Fourier transform of the extrapolated ITD are shown.

$$Q_{abAsymp}(\lambda, \mu) = \text{Re}\left[A\left(\frac{\Gamma(1+\alpha)}{(-i|\lambda|)^{\alpha+1}} + e^{i\lambda} \frac{\Gamma(1+\beta)}{(i|\lambda|)^{\beta+1}} \right) \right],$$
$$(36)$$

where we have three parameters and we impose the constraint $A > 0$ and $\beta > 0$. We then combine the DNN trained results and the extrapolated results together as

$$Q(\lambda, \mu) = \begin{cases} Q_{DNN}(\lambda, \mu) & \text{for } \lambda \leq \lambda_L \\ Q_{Extra}(\lambda, \mu) & \text{for } \lambda > \lambda_L. \end{cases} \qquad (37)$$

The extrapolated results are shown in the upper panel of Fig. 17, in which $Q_{DNN}(\lambda, \mu)$ is trained using $z \in [2a, 0.61$ fm$]$ for the $m_\pi = 140$ MeV lattice. We fit the extrapolation form using $\lambda \in [\lambda_L, \lambda_{max}]$ with $\lambda_L = 3$ and 4. As one can see, the extrapolation results using different λ_L overlap with each other. The two-parameter model (abAsymp) extrapolated results always show wider error bands compared to the PL model since one more parameter is used in the fit form. We then perform the Fourier transform by combining the discrete Fourier transform and the analytical integral,

$$q(x,\mu) \equiv \sum_{0}^{\lambda_{\max}} 2\frac{\Delta\lambda}{2\pi} Q_{\mathrm{DNN}}(\lambda,\mu)\cos(x\lambda)$$
$$+ 2\int_{\lambda_{\max}}^{\infty} \frac{d\lambda}{2\pi} Q_{\mathrm{Extra}}(\lambda,\mu)\cos(x\lambda), \quad (38)$$

in which $\Delta\lambda$ is the step spacing of λ that was used when we trained from the DNN. The results of the Fourier transform are shown in the lower panels of Fig. 17. The bands from different extrapolation models and different choices of λ_L overlap with each other, and the discrepancy among them mainly exists in the small-x region.

E. Continuum estimate of the DNN represented ITD

We summarize the $Q(\lambda,\mu)$ from the three HISQ ensembles using data with $z \in [2a, 0.61 \text{ fm}]$ in Fig. 18. As expected from the previous analyses of the moments and model fits, the estimates from the three ensembles are close to each other. We again consider the continuum extrapolation only of order $\mathcal{O}(a^2)$ as

$$Q_a(\lambda,\mu) = Q_{a^2 \to 0}(\lambda,\mu) + g(\lambda)a^2, \quad (39)$$

where $g(\lambda)$ is an independent fit parameter for each λ, from which we then perform the extrapolation using the PL and the two-parameter asymptotic (abAsymp) models with λ_L varying from 3 to 4.5. As before, we ignore any weak quark mass dependence in the datasets. To stabilize the two-parameter asymptotic (abAsymp) extrapolation, we impose a prior on the exponent α, which dominates the decay rate as a function of λ, using the corresponding value from the power-law fit sample by sample. The extrapolated results are shown as the red and blue bands in Fig. 18. We compare our continuum estimated ITD to the ones from our model fit Model-4p and the global analysis of JAM21nlo [97]. It can be seen that our results show good agreement with
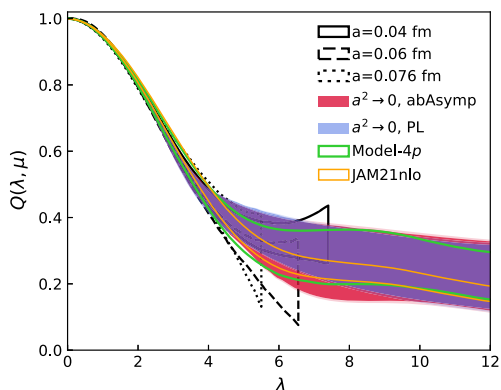

FIG. 18. The $Q(\lambda,\mu)$ reconstructed from the DNN using $z \in [2a, 0.61 \text{ fm}]$ for the three ensembles are shown as the empty bands, while the orange band is the continuum estimate with $\mathcal{O}(a^2)$ corrections. We also show the ITD reconstructed from JAM21nlo [97] for comparison.
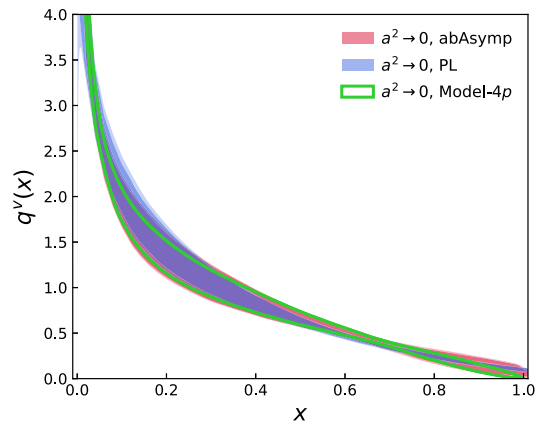

FIG. 19. The results for the pion valence distribution from the Fourier transform of the mass-independent continuum estimate of the DNN-based $Q(\lambda,\mu)$ using either the PL or two-parameter asymptotic (abAsymp) extrapolations are shown. We also show the results from the Model-4p fit for comparison.

JAM21nlo in the small-λ region, while it only covers the JAM2nlo at large λ due to the large errors. In Fig. 19, we show the results for the pion valence distribution from the Fourier transform of the extrapolated ITD, where we vary $z_{\max} \in [0.6, 0.72]$ fm to estimate the statistical and systematic errors. Overall agreement can be observed compared to our model based determination Model-4p, while as expected the DNN-based PDF, which is more flexible, shows larger errors than our model based determination. Since the power-law extrapolation (blue band) does not ensure that the PDFs are only defined in $x \in [0, 1]$, it does not vanish at $x = 1$. This can be considered as a systematic error from the extrapolation model which parametrizes the unknown large-λ behavior. With one more parameter, the results from the two-parameter asymptotic (red band) extrapolation show larger statistical errors but are consistent with 0 at $x = 1$.

IX. PION VALENCE PDF FROM THE QUASI-PDF APPROACH WITH HYBRID RENORMALIZATION

Another way to obtain the valence pion PDF from the matrix elements calculated in this study is to use the quasi-PDF approach of LaMET. Very recently, using the NNLO quasi-PDF matching with the novel hybrid renormalization scheme [84], the valence pion PDF has been calculated with controlled systematic errors in the range $0.03 \lesssim x \lesssim 0.8$ [63]. In that study, two lattice spacings, $a = 0.04$ fm and $a = 0.06$ fm, were used, and the discretization effects turned out to be small. On the other hand, the valence quark masses were larger than the physical ones, corresponding to a pion mass of 300 MeV. In this section we present a calculation of the valence pion PDF within the quasi-PDF approach using NNLO matching and the hybrid renormalization scheme for physical quark masses

$(M_\pi = 140$ MeV$)$ and $a = 0.076$ fm. Our analysis closely follows the strategy outlined in Ref. [63].

The hybrid renormalization scheme is defined by a combination of the ratio scheme at short distances and the explicit subtraction of the self-energy divergence of the Wilson line at large distances:

$$\tilde{R}(z, z_S, P_z) = \begin{cases} \dfrac{h^B(z, P_z, a)}{h^B(z, 0, a)}, & z \le z_S \\ \dfrac{h^B(z, P_z, a)}{h^B(z_S, 0, a)} e^{\delta m(a)|z - z_S|}, & z > z_S. \end{cases} \quad (40)$$

Here, $\delta m(a)$ is the parameter containing the self-energy divergence of the Wilson line. It is determined from the analysis of the free energy of a static quark [104,105]. In this work, we choose $z_S = 0.228$ fm for the $a = 0.076$ fm ensemble. We need to connect h^R in the hybrid scheme to h^R in the $\overline{\text{MS}}$ scheme, and to do this, we need to introduce a parameter, $\bar{m}_0$, that connects the renormalon ambiguity in the renormalization of the Wilson line in the lattice subtraction scheme and $\overline{\text{MS}}$ scheme [63]. As in Ref. [63], to obtain $\bar{m}_0$, we fit our results for $\tilde{R}(z, z_S, P_z = 0)$ to the Ansatz

$$\tilde{R}(z, z_S, P_z = 0) = e^{-\bar{m}_0(\mu)(z - z_S)} \frac{C_0^{\text{NNLO}}(z^2\mu^2) + \Lambda(\mu)z^2}{C_0^{\text{NNLO}}(z_S^2\mu^2) + \Lambda(\mu)z_S^2}, \quad (41)$$

with another parameter $\Lambda(\mu)$ that approximates the higher-twist contributions. We perform the above two-parameter fit with $z \in [z_S + a, z_{\max}]$. At fixed order, $\bar{m}_0(\mu)$ and $\Lambda(\mu)$ are both scale dependent. We choose $\mu = 2$ GeV as the central value of the renormalization scale and vary it by a factor of $\sqrt{2}$ to estimate the perturbative uncertainty. The fit results are shown in Fig. 20. As one can see from the figure, the resulting values of $\bar{m}_0(\mu)$ and $\Lambda(\mu)$ show only mild dependence on $z_{\max}$ for $\mu \ge 2$ GeV. For the lowest value of the renormalization scale, $\mu = 1.4$ GeV, on the other hand, we see significant dependence on $z_{\max}$. This implies that for such low values of μ the perturbative expression for $C_0(\mu^2 z^2)$ is not very reliable because α_s is quite large, and the Ansatz given by Eq. (41) cannot describe the data well. We also see that there is a significant scale dependence for the value of $\bar{m}_0(\mu)$ and $\Lambda(\mu)$, which in turn translates into an uncertainty for the z dependence of $h^R(z, z_S, P_z)$. This uncertainty will get smaller as P_z increases and disappears in the $P_z \to \infty$ limit. For our final estimates of $\bar{m}_0$ and Λ, we chose $z_{\max} = 0.456$ fm.

Following Ref. [63], we calculate the renormalized matrix element in the ratio scheme as

$$\tilde{h}^R(z, z_S, P_z, \mu)$$
$$= \begin{cases} N \dfrac{h^B(z, P_z, a)}{h^B(z, 0, a)} \dfrac{C_0(z^2\mu^2) + \Lambda z^2}{C_0(z^2\mu^2)}, & z \le z_S \\ N \dfrac{h^B(z, P_z, a)}{h^B(z_S, 0, a)} \dfrac{C_0(z_S^2\mu^2) + \Lambda z_S^2}{C_0(z_S^2\mu^2)} e^{\delta m'(z - z_S)}, & z > z_S, \end{cases} \quad (42)$$
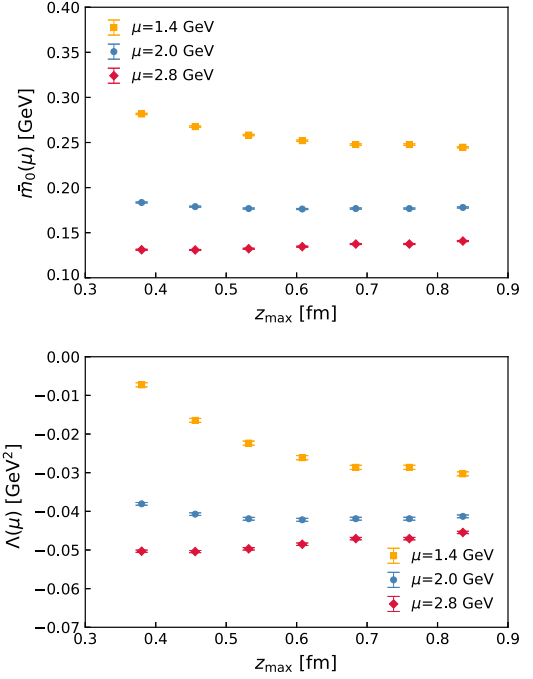

FIG. 20. The results for $\bar{m}_0(\mu)$ and $\Lambda(\mu)$ from fits of $\tilde{R}(z, z_S, P_z = 0)$ with $z \in [z_S + a, z_{\max}]$ to Eq. (41) are shown. We choose the scale $\mu = 2$ GeV and vary it by a factor of $\sqrt{2}$.

where $N = h^B(0, 0, a)/h^B(0, P_z, a)$ and $\delta m' = \delta m + \bar{m}_0$. The above expression for $\tilde{h}^R(z, z_S, P_z)$ is equivalent to Eq. (40) if lattice artifacts and higher-twist contributions can be neglected. The factor N ensures that the renormalized matrix element is equal to 1 at $z = 0$ and removes the lattice artifacts proportional to powers of aP_z. The factor $(C_0(z^2\mu^2 + \Lambda z^2)/C_0(z^2\mu^2))$ was introduced to remove the leading higher-twist contribution that enters through $h^B(z, 0, a)$. Because of this factor and $\bar{m}_0$, the calculated renormalized matrix element in the ratio scheme also depends on μ at any given order of perturbation theory used to evaluate $C_0(\mu^2 z^2)$. At infinite order, this μ dependence will disappear. The renormalized matrix elements at $\mu = 2$ GeV are shown in Fig. 21. One can observe that $\tilde{h}^R$ tends to saturate when $P_z \gtrsim 1$ GeV, though the statistical errors are large.

Since the matrix elements at large z are very noisy and become unreliable, we shall first perform an extrapolation and then do the Fourier transform to obtain the quasi-PDF $\tilde{q}(x)$. For the extrapolation, we use $\tilde{h}^R$ in the range $[z_{\min}, z_{\max}]$, with $z_{\min}$ being the first z where $\tilde{h}^R(z = z_{\min}) < 0.2$ and $z_{\max}$ being either the last z where $\tilde{h}^R(z_{z=\max}) > 0$ or the limit of half the lattice size $(32a)$, i.e., $32a$. Since the spatial correlators of the equal-time matrix elements decay exponentially at large z (with the exception of zero modes), we chose the model combining the exponential and power decay as discussed in detail in Ref. [63],
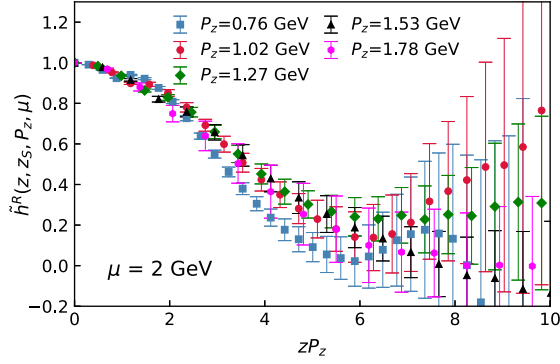
FIG. 21. The hybrid-scheme renormalized matrix elements $\tilde{h}^R(z, z_S, P_z, \mu)$ as a function of zP_z at $\mu = 2$ GeV for the $m_\pi = 140$ MeV ensemble are shown.

$$A \frac{e^{-m_{\mathrm{eff}}|z|}}{|\lambda|^d}, \tag{43}$$

with $\lambda = zP_z$, and the constraints $A, d > 0$ and $m_{\mathrm{eff}} > 0.1$ GeV. A continuity condition is also imposed at the point connecting the data and the extrapolation function. The extrapolated matrix elements for $P_z = 1.53$ and 1.78 GeV are shown in the upper panel of Fig. 22. Then, we perform the FT by combining the discrete FT of the lattice data for $z < z_{\mathrm{min}}$ and the continuous FT of the
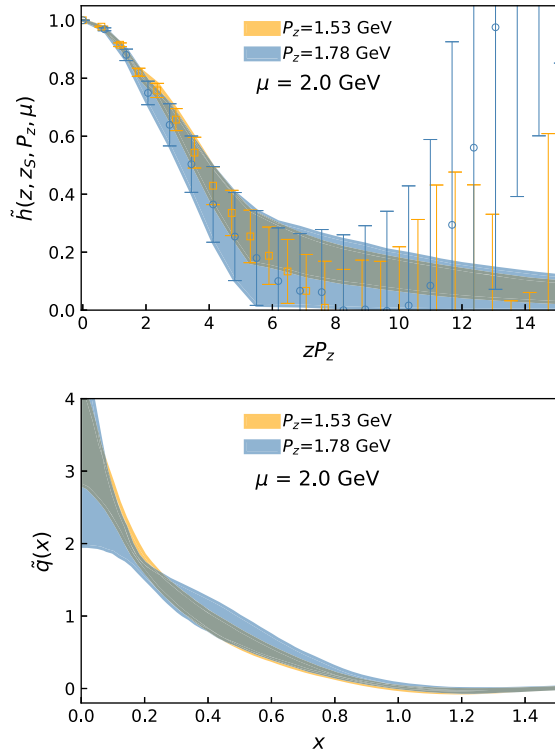
FIG. 22. The extrapolated matrix elements $\tilde{h}^R(z, z_S, P_z, \mu)$ as a function of zP_z at $\mu = 2$ GeV are shown in the upper panel for cases $P_z = 1.53$ GeV and 1.78 GeV. The corresponding quasi-PDFs after a Fourier transform are shown in the lower panel.

extrapolation function for $z \geq z_{\mathrm{min}}$. The corresponding results are shown in the lower panel of Fig. 21. The extrapolation has very weak impact on the moderate-to-large x region but introduces large uncertainty at small x. Nevertheless, after perturbative matching, the latter is beyond the region of x where we have systematic control over [63].

Next, we match the qPDF $\tilde{q}^v(x, \lambda_S, P_z, \mu)$ to the pion valence distribution $q^v(x, \mu)$ in the $\overline{\mathrm{MS}}$ scheme through LaMET [11,84,87,106] using NNLO kernels [61–63],

$$q^v(x, \mu) = \int_{-\infty}^{\infty} \frac{dy}{|y|} C^{-1}\left(\frac{x}{y}, \frac{\mu}{yP_z}, |y|\lambda_S\right) \tilde{q}^v(y, \lambda_S, P_z, \mu)$$
$$+ \mathcal{O}\left(\frac{\Lambda_{\mathrm{QCD}}^2}{(xP_z)^2}, \frac{\Lambda_{\mathrm{QCD}}^2}{((1-x)P_z)^2}\right), \tag{44}$$

where $\lambda_S = z_S P_z$ and $z_S = 0.228$ fm. Then, we will directly derive the PDF with P_z-controlled power
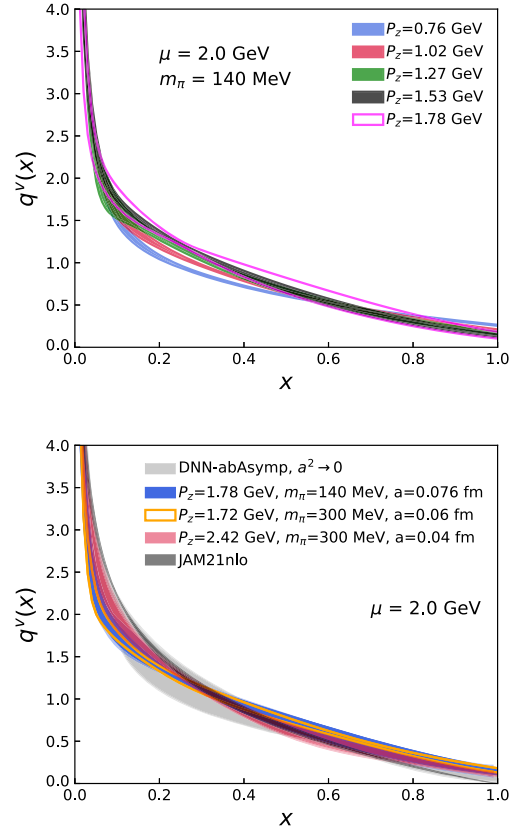

FIG. 23. Upper panel: the pion valence distribution $q^v(x)$ obtained from different P_z for the $m_\pi = 140$ MeV ensemble using NNLO matching is shown. Lower panel: the pion valence distribution $q^v(x)$ obtained from $P_z = 1.78$ GeV (blue) on the $m_\pi = 140$ MeV lattice, $P_z = 1.72$ GeV (orange), and $P_z = 2.42$ GeV (red) on the $m_\pi = 300$ MeV lattices are shown. The darker bands come from the statistical errors, while the weaker bands are the systematic errors from scale variation. We also show our best determination from the DNN for comparison.

corrections for the middle range of x. In the upper panel of Fig. 23, we show the PDF obtained from matching with $P_z \in [0.76, 1.78]$ GeV. We see a significant dependence of the valence pion PDF on P_z. However, as P_z increases, this dependence diminishes, and for the largest three P_z, the resulting valence pion PDFs agree within the estimated errors. The value of q^v for $x \simeq 1$ also decreases with increasing P_z.

Finally, we show $P_z = 1.78$ GeV (blue) for the $m_\pi = 140$ MeV lattice and $P_z = 1.72$ GeV (orange) and 2.42 GeV (red) for the $m_\pi = 300$ MeV lattices from Ref. [63] in Fig. 23, with the darker bands being the statistical errors and the lighter bands being the systematic errors from scale variation. It can be seen that the systematic errors estimated from scale variation are small, even though $\bar{m}_0(\mu)$ shows sizable μ dependence. That is because the renormalon effect, which depends on z, contributes less for larger momentum and is supposed to disappear for infinite momentum. The results from similar momentum, such as 1.78 and 1.72 GeV, basically overlap with each other within the statistical errors, implying both the lattice spacing and mass dependence are small (more details can be found in Appendix C), but differ from the one with larger momentum $P_z = 2.42$ GeV. The the best determination from the DNN in the continuum limit and the JAM21nlo [97] results are also shown for comparison. Though all three LaMET results show some agreement with JAM21nlo in the middle-x region, the one with the highest momentum ($P_z = 2.42$ GeV) overlaps the best.

X. CONCLUSION

We presented lattice calculations of the pion bilocal matrix elements for physical quark masses and a lattice spacing of $a = 0.076$ fm. These results were combined with our previous results for $m_\pi = 300$ MeV, but with very fine lattice spacing of $a = 0.06$ fm and $a = 0.04$ fm. This allowed us to obtain continuum-extrapolated results for pion valance PDF at the physical point. We used the NNLO short-distance factorization of RGI invariant matrix elements to determine up to sixth-order Mellin moments of the pion valence PDF. The inclusion of the NNLO corrections to the perturbative matching did not result in significant changes to the numerical values of the moments, which indicates the convergence of leading-twist approximation with fixed-order perturbation theory. The NNLO correction also played a significant role in our test of the validity of the short-distance factorization as discussed in Appendix B. The pion mass dependence of the moments were found to be very mild. As summarized in Fig. 24, our results for the Mellin moments of the pion valance PDF are in excellent agreement with that from the different phenomenological global fits to experimental data. By fitting model functional forms of the PDF to the zP_z dependence of RGI matrix elements, we inferred its x dependence. Further, we reconstructed the x dependence of the PDF from the
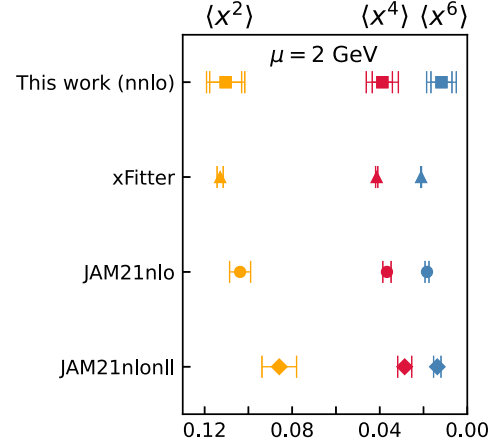

FIG. 24. $\langle x^n \rangle$ with $n = 2$, 4, 6 from the mass-independent continuum estimate using NNLO matching at $\mu = 2$ GeV are shown in the first row. The statistical errors are in the first brackets, while the systematic errors are in the second brackets, which are estimated by varying $z_{max} \in [0.48, 0.72]$ fm for the fits. For comparison, we also show the moments evaluated from global analyses of xFITTER [107], JAM21NLO, and JAM21NLONLL DOUBLE-MELLIN [97].

RGI matrix elements using a DNN. We found that the model fits and the DNN-based reconstruction of the PDF are in very good agreement among themselves, as well as with the phenomenological global fits of experimental data; see Fig. 25. Next, we obtained the x-dependent quasi-PDF from the matrix elements, renormalized in a hybrid scheme. From this quasi-PDF, we determined the x dependence of the PDF using NNLO perturbative matching. As illustrated in Fig. 23, we found the pion mass dependence of the PDF to be small for pion momenta $\gtrsim 1.5$ GeV and that the x
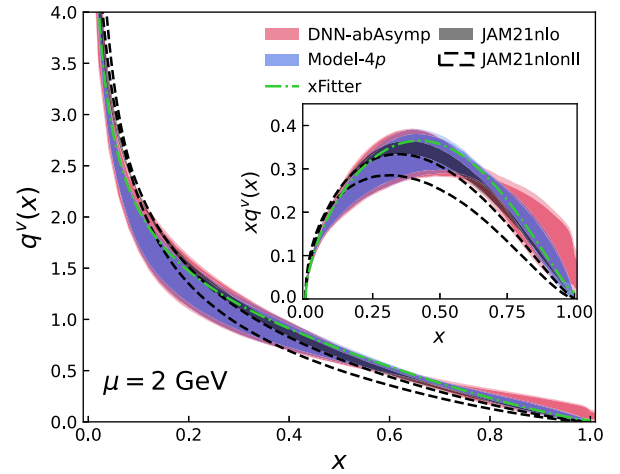

FIG. 25. We show our DNN-based PDF determination in the continuum limit, together with the four-parameter fit results Model-4p, and the global analyses of the experimental data with NLO fixed-order perturbation theory from xFITTER [107] and JAM21NLO [97] as well as the results considering threshold resummation using the double-Mellin method (JAM21NLONLL).

dependence of the PDF are in good agreement with that from the DNN reconstruction and phenomenological global fit results. To conclude, we presented continuum-extrapolated lattice QCD results of the Mellin moments and the x dependence of the valance PDF of pion for physical values of quark masses. The Mellin moments and the x-dependent PDF, determined in multiple ways, agree with each other and with that form phenomenological global fits.

ACKNOWLEDGMENTS

APPENDIX A: TRUNCATION OF THE OPE FORMULA

In this Appendix, we discuss the truncation of the short-distance factorization expression. As mentioned in the main text, in lattice calculations, one probes the reduced Ioffe-time distribution in a limited range of zP_z, namely, up to $(zP_z)_{max} \simeq 6$, and therefore the short-distance factorization
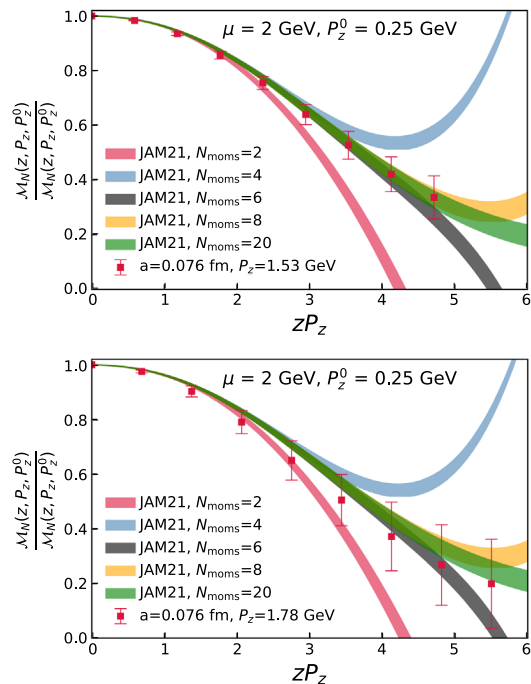

FIG. 26. The reconstructed ratio-scheme matrix elements $\mathcal{M}(z, P_z, P_z^0)$ from the JAM21 global analysis [97] with $P_z^0 = 0.25$ GeV and NNLO matching are shown for $P_z = 1.53$ GeV (top) and $P_z = 1.78$ GeV (bottom). The width of the band corresponds to the uncertainty due to the variation of the renormalization scale.

formula given by Eq. (15) can be truncated. To test how many terms we need to keep in the sums entering Eq. (15), we construct the reduced Ioffe-time distribution using the NNLO global analysis of the pion valence PDF by the JAM Collaboration (JAM21) [97] with NNLO matching, keeping different numbers of terms. We also compare it with our lattice results. This analysis is shown in Fig. 26. Keeping ten terms in the sum, i.e., considering up to the 20th moment, almost reproduces the exact result. Keeping only the first four terms in the sum results in truncation errors that are smaller than the scale uncertainty, and given the errors of the lattice results, it is even possible to truncate the sum at the term proportional to $\langle x^6 \rangle$. We note our lattice results for reduced Ioffe-time distribution agree very well with those obtained from JAM21.

APPENDIX B: THE RELIABILITY OF THE PERTURBATIVE MATCHING AT DIFFERENT z

In this Appendix, we scrutinize the validity of the perturbative matching for different values of z. To do this, we revisit our results for the ratio scheme matrix elements and the extraction of the moments at smaller lattice spacings, $a = 0.04$ and 0.06 fm, and unphysical pion mass, $m_\pi = 300$ MeV. In Fig. 27, we show the second moment extracted for fixed values of z at LO, NLO, and NNLO. We see the same tendencies as for $a = 0.076$ fm
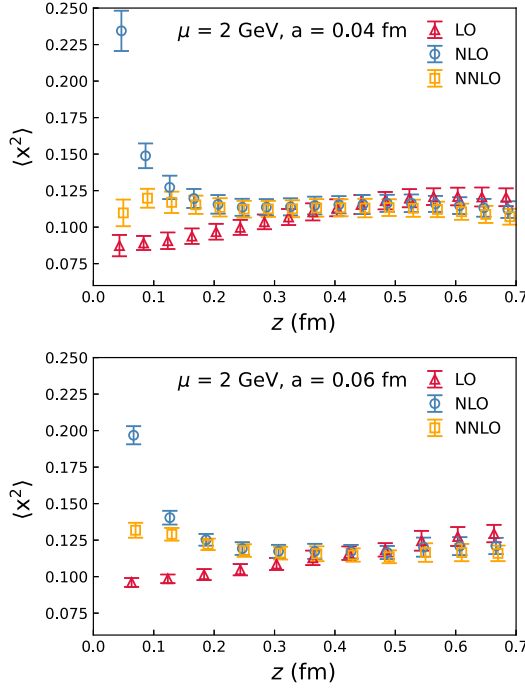
FIG. 27. $\langle x^2 \rangle$ extracted from the ratio-scheme data $\mathcal{M}(z, P_z, P_z^0)$ at each z using LO, NLO, and NNLO kernels is shown for the a $= 0.04$ fm ensemble (top) and 0.06 ensemble (bottom).

shown in Fig. 7. In particular, the NNLO result has the smallest z dependence. At the smallest two values of z, the second moment is too high. However, in physical units, the values of z for which the second moment is too high are shifted to smaller z, confirming that this effect is mostly due to lattice discretization.

The ratio-scheme matrix element defined in Eq. (14) is a RGI quantity, from which we can extract the Mellin moments perturbatively through the leading-twist OPE formula Eq. (15). For the following discussion, we consider the $P_z = 0$ case and write the OPE formula as

$$\mathcal{M}(z, P_z, P_z^0 = 0) = \sum_{n=0}^{\infty} \frac{(-izP_z)^n}{n!} \langle x^n \rangle_{\mathrm{LO}}(z), \quad (B1)$$

where we define

$$\langle x^n \rangle_{\mathrm{LO}}(z) = c_n(z^2\mu^2)\langle x^n \rangle(\mu), \qquad c_n(z^2\mu^2) = \frac{C_n(z^2\mu^2)}{C_0(z^2\mu^2)}. \quad (B2)$$

On the lattice side, $\langle x^n \rangle_{\mathrm{LO}}(z)$ is obtained by fitting the results for the reduced Ioffe-time distribution with the polynomial form, and the results are shown in Figs. 7 and 27 as open triangles. At leading order, $c_n = 1$ and $\langle x^n \rangle_{\mathrm{LO}}(z)$ is just the usual Mellin moment. Beyond leading order, the z dependence of $\langle x^n \rangle_{\mathrm{LO}}(z)$ should match the z dependence obtained in lattice QCD calculations if

perturbation theory is sufficiently accurate. This ensures that the $\langle x^n \rangle(\mu)$ extracted from the lattice are independent of z. We study $\langle x^n \rangle_{\mathrm{LO}}(z)$ for $x < 0.3$ fm where perturbation theory may be reliable.

In this work, we used perturbative kernels up to NNLO level,

$$C_n^{\mathrm{NNLO}}(\mu^2 z^2) = 1 + \frac{\alpha_s(\mu)}{2\pi} C_n^{(1)}(\mu^2 z^2) + \frac{\alpha_s^2(\mu)}{2\pi} C_n^{(2)}(\mu^2 z^2),$$
$$(B3)$$

where $C_n^{(1)}(\mu^2 z^2)$ and $C_n^{(2)}(\mu^2 z^2)$ are the one-loop and two-loop coefficients that contain terms proportional to $\ln(\mu^2 z^2 e^{2\gamma_E}/4)$. For example, at NLO, $C_n^{(1)}(\mu^2 z^2)$ reads [87]

$$C_n^{(1)}(z^2\mu^2) = C_F \left[\left(\frac{3 + 2n}{2 + 3n + n^2} + 2H_n \right) \ln \frac{\mu^2 z^2 e^{2\gamma_E}}{4} \right.$$
$$\left. + \frac{5 + 2n}{2 + 3n + n^2} + 2(1 - H_n)H_n - 2H_n^{(2)} \right],$$
$$(B4)$$

with H_n being the harmonic numbers. If μ is not too different from $2/(ze^{\gamma_E})$, the logarithms are not too large, and the perturbative expansion for the Wilson coefficients is well behaved. However, if z is varied in a large range, then the logarithms could become large and need to be resummed. This can be done with the help of the renormalization group equation

$$\left(\frac{\partial}{\partial \ln \mu^2} + \beta(a_s)\frac{\partial}{\partial a_s} - \gamma_n \right) C_n(z^2\mu^2) = 0, \quad (B5)$$

in which the anomalous dimension reads (up to NNLO)

$$\gamma_n = a_s \left[\frac{3C_F}{2} - \int_0^1 dx\, x^n P(0)_{qq}(x) \right]$$
$$+ a_s^2 \left\{ \int_0^1 dx\, x^n \left[P_{qq}^{V(1)}(x)\theta(x) - P_{q\bar{q}}^{(1)}(x)\theta(x) \right] \right.$$
$$+ \left[C_F^2 \left(-\frac{5}{8} + \frac{2\pi^2}{3} \right) + C_F C_A \left(\frac{49}{24} - \frac{\pi^2}{6} \right) \right.$$
$$\left. \left. + C_F n_f T_F \left(-\frac{5}{6} \right) \right] \right\}$$
$$= a_s \gamma_n^{(1)} + a_s^2 \gamma_n^{(2)}, \quad (B6)$$

and the running coupling $a_s = \alpha_s/(2\pi)$ obeys

$$\beta(a_s) = \frac{da_s}{d\ln\mu^2} = -a_s^2 \beta_0 - a_s^3 \beta_1,$$
$$\beta_0 = \frac{11C_A - 4n_f T_F}{6}, \qquad \beta_1 = \frac{102 - 38n_f/3}{4}. \quad (B7)$$

Then, solving the RG equation with boundary condition $Q_0 = k \frac{2}{z e^{\gamma_E}}$, one obtains the leading-logarithm (LL) resummed NLO kernels (NLOevo),

$$C_n^{\mathrm{NLO_{evo}}}(\mu^2 z^2) = C_n^{\mathrm{NLO}}(Q_0^2 z^2) \left(\frac{a_s(\mu)}{a_s(Q_0)} \right)^{-\gamma_n^{(1)}/\beta_0}, \quad \text{(B8)}$$

or NLL resummed NNLO kernels (NNLOevo),

$$C_n^{\mathrm{NNLO_{evo}}}(\mu^2 z^2) = C_n^{\mathrm{NNLO}}(Q_0^2 z^2)$$

$$\times e^{\frac{\gamma_n^{(1)} \ln \frac{a_s(\mu)}{a_s(Q_0)}}{\beta_0} - \frac{\left(-\beta_1 \gamma_n^{(1)} + \beta_0 \gamma_n^{(2)}\right) \ln \frac{\beta_0 + \beta_1 a_s(\mu)}{\beta_0 + \beta_1 a_s(Q_0)}}{\beta_0 \beta_1}}, \quad \text{(B9)}$$

where $\gamma_n^{(1)}$ and $\gamma_n^{(2)}$ are the anomalous dimension of the nth moments. For a conventional choice $k = 1$ of $Q_0^2 = \frac{4k^2}{z^2 e^{2\gamma_E}}$, terms proportional to logarithms $\ln(\mu^2 z^2 e^{2\gamma_E}/4)$ are all cancelled. For example, the NLO + LL kernels read

$$C_n^{\mathrm{NLO_{evo}}}(\mu^2 z^2)$$

$$= \left\{ 1 + \frac{\alpha_s(Q_0)}{2\pi} C_F \left[\left(\frac{3 + 2n}{2 + 3n + n^2} + 2H_n \right) \ln(k^2) \right. \right.$$

$$\left. \left. + \frac{5 + 2n}{2 + 3n + n^2} + 2(1 - H_n)H_n - 2H_n^{(2)} \right] \right\}$$

$$\times \left(\frac{a_s(\mu)}{a_s(Q_0)} \right)^{-\frac{\gamma_n^{(1)}}{\beta_0}}. \quad \text{(B10)}$$

As discussed above, the coefficients $c_n(\mu^2 z^2)$ are supposed to compensate the z dependence of $\langle x^2 \rangle_{\mathrm{LO}}(z)$ and produce a z-independent $\langle x^2 \rangle(\mu)$. Now, we can contrast the z dependence of $\langle x^2 \rangle_{\mathrm{LO}}(z)$ with the perturbative prediction of $c_n(\mu^2 z^2)$ at different order and different choices of the renormalization scale. We perform such a comparison for the $a = 0.04$ fm ensemble, where we have the largest range in z and the smallest discretization errors. To see the predictive power of the perturbative result, we can fix $\langle x^2 \rangle(\mu)$ at some z_0 and then predict the $\langle x^2 \rangle_{\mathrm{LO}}(z)$ at different values of z. Here, we choose $z_0 = 0.16$ fm, which is in the middle of the range $z \in [0, 0.3 \text{ fm}]$. In the upper panels of Fig. 28, we show the fixed-order predictions, i.e., NLO (left) and NNLO (right), for $\langle x^2 \rangle_{\mathrm{LO}}(z)$. We choose the factorization scale $\mu = 2$ GeV as the central value and vary it by a factor of $\sqrt{2}$. As one can see, the NLO result with $\mu \gtrsim 2$ GeV can describe most of the data points. For $z > 0.1$ fm, the NLO result can describe the z dependence of $\langle x^2 \rangle_{\mathrm{LO}}(z)$ reasonably well for $\mu = 2$ GeV and $\mu = 2.8$ GeV. However, the scale variation of the NLO result is quite large, and for lower values of the renormalization scale, e.g., $\mu = 1.4$ GeV, the NLO result fails to describe the lattice data. On the contrary, the NNLO result shows a very small scale dependence for $z > 0.1$ fm and describes the lattice data very well. Thus, the use of NNLO largely improves the quality of the perturbative matching. For small separations, $z < 0.1$ fm, the NLO result fails to describe the lattice calculations and has large scale
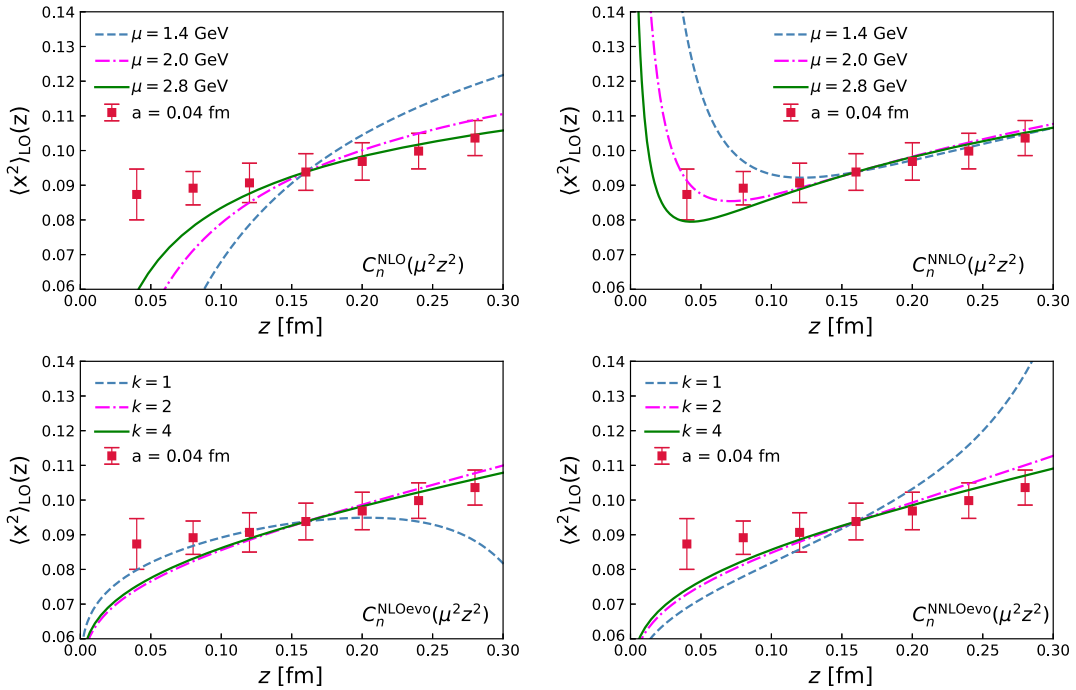

FIG. 28. $\langle x^2 \rangle_{\mathrm{LO}}(z)$ extracted from the ratio scheme matrix elements for the $a = 0.04$ fm ensemble compared to the perturbative predictions at NLO and NNLO (top panels) and NLO$_{\mathrm{evo}}$ and NNLO$_{\mathrm{evo}}$ (bottom panels) for different renormalization scales. The perturbative results have been normalized to the lattice result at $z = 0.16$ fm.

dependence. The NNLO result for these distances does a better job in describing the lattice data but has large scale dependence. Furthermore, the NLO and NNLO results have very different shapes for $z < 0.1$ fm, explaining the discrepancy between the NLO and NNLO results seen in Figs. 7 and 27 at the smallest z. From this, we conclude that there are large logs in the perturbative result which need to be resummed. The resummed results given by Eqs. (B8) and (B9) are shown in the lower panels of Fig. 28. Here, we use $\mu = 2$ GeV and set $k = 1, 2, 4$ for $Q_0 = k\frac{2}{ze^{\gamma_E}}$. As one can see, the naive choice $Q_0 = \frac{2}{ze^{\gamma_E}}$ ($k = 1$) that eliminates all the logarithms $\ln(\mu^2 z^2 e^{2\gamma_E}/4)$ does not give the best description of the lattice data points. When $k \gtrsim 2$, the data under consideration can be well described except the first two points which suffer from significant discretization effects. This suggests $Q_0 \simeq \frac{4}{ze^{\gamma_E}}$ is the better choice to set the scale in α_s for the renormalization group improved perturbative result. We also note that NLO_{evo} and NNLO_{evo} results are very similar and describe the lattice data for $z < 0.3$ fm, except when the scale Q_0 becomes too low and α_s is very large.

We summarize the above discussion with the following. the fixed-order NLO and NNLO results describe the z dependence of the lattice results well for $z > 0.1$ fm, provided $\mu \geq 2$ GeV, but are not reliable for $z < 0.1$ fm. For NNLO, smaller values of the scale μ are possible since the scale variation is tiny. The NNLO result also works at larger z, which shows a controlled scale dependence. The RG-improved results, NLO_{evo} and NNLO_{evo}, describe the lattice data well for $z < 0.3$ fm, if the coupling is fixed at $Q_0 \simeq \frac{4}{ze^{\gamma_E}}$, suggesting that this scale sets the running of α_s in the perturbative expression.

It may appear surprising that the NNLO result seems to capture the z dependence of the ratio scheme matrix elements at $z \simeq 0.3$ fm or even larger. In Ref. [48], it is suggested that this may be due to the fact that higher-twist contributions, while non-negligible for $z > 0.3$ fm, cancel

out in the ratio scheme. In Ref. [63], it is shown that the NNLO and NNNLO corrections to $C_0(\mu^2 z^2)$ are large for $z > 0.3$ fm (cf. Fig. 6 in that paper). Our analysis suggest that some of these large corrections, which are also present in $C_n(\mu^2 z^2), n > 0$, cancel out in the ratio $c_n = C_n/C_0$, rendering the perturbative expansion reliable even at relatively large values of z. The analysis in this Appendix shows that moments of the pion PDF can be reliably extracted using only $z < 0.3$ fm if RG-improved matching is used. In this region of z, the perturbative matching is reliable. We also see, however, that using the fixed-order NNLO result allows the moments to be obtained reliably at large z, where the applicability of the perturbative matching may seem to be doubtful. But the two determinations agree. This is because some higher-order perturbative corrections to the Wilson coefficients C_n cancel out in C_n/C_0 and there are also cancellations of the higher-twist contributions. This opens up the possibility to reliably determine the PDF using the ratio scheme in current lattice calculations, which are typically performed on lattices with $a > 0.04$ fm.

APPENDIX C: MASS DEPENDENCE OF PDF FROM HYBRID RENORMALIZATION AND *x*-SPACE MATCHING

To understand the pion mass and lattice spacing dependence of the results, we compare the pion valence PDF obtained in this work with the earlier calculation performed at two lattice spacings, $a = 0.04$ fm and $a = 0.06$ fm, with a pion mass $m_\pi = 300$ MeV [63] in Fig. 29. It can be observed that the $q^v(x)$ from different calculations show some discrepancy at $P_z \sim 1$ GeV, while they start to overlap when $P_z \gtrsim 1.5$ GeV. On the one hand, the lattice spacing dependence is mild as expected from previous sections, and on the other hand, it also suggests the pion mass may play a role for low momenta but decay rapidly for large P_z.
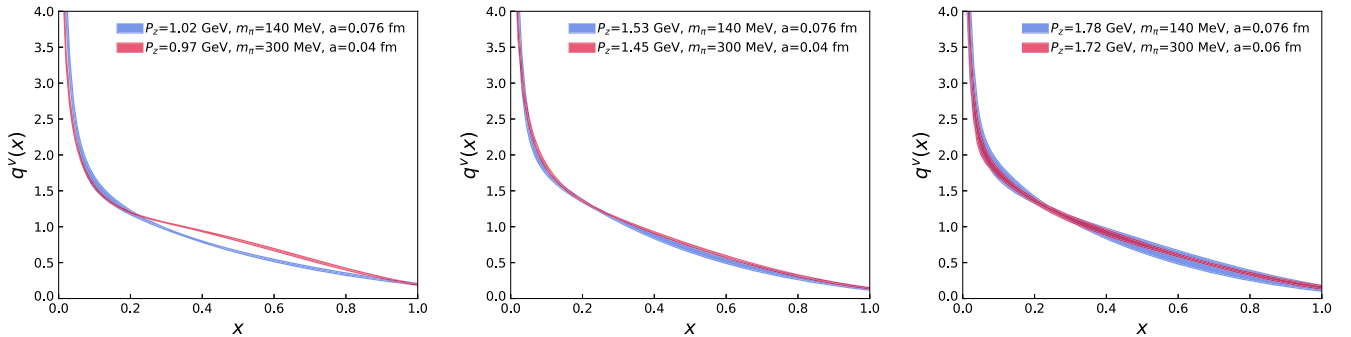

FIG. 29.   The pion valence distribution $q^v(x)$ obtained from an NNLO LaMET matching is shown. We show the results from the $m_\pi = 140$ MeV case (blue) and the $m_\pi = 300$ MeV cases [63] (red) for comparison.

APPENDIX D: THE MATCHING STRATEGY AND THE DNN REPRESENTATION OF THE ITD $Q(\lambda, \mu)$

In Sec. VIII, we apply Eq. (28) to solve for the light-cone ITD $Q(\lambda, \mu)$ from the ratio-scheme matrix elements. Starting from the standard reduced Ioffe-time distribution, one has the factorization formula

$$\tilde{Q}(z, P_z) = \int_{-1}^{1} d\alpha \frac{\mathcal{C}(\alpha, \mu^2 z^2)}{C_0^{\overline{MS}}(\mu^2 z^2)} Q(\alpha\lambda, \mu)$$

$$= Q(\lambda, \mu) + \int_{-1}^{1} d\alpha [Q(\alpha\lambda, \mu) - Q(\lambda, \mu)]$$

$$\times \frac{n_{NNLO}(\alpha, \mu^2 z^2)}{d_{NNLO}(\mu^2 z^2)}, \tag{D1}$$

where we rewrite the perturbative kernels $\bar{\mathcal{C}}(\alpha, \mu^2 z^2)$ by

$$n_{NNLO}(\alpha, \mu^2 z^2) = \sum_{i=1}^{2} \sum_{j=0}^{i} \left(\frac{\alpha_s}{2\pi}\right)^i n_{i,j}(\alpha)[L(\mu^2 z^2)]^j,$$

$$d_{NNLO}(\mu^2 z^2) = 1 + \sum_{i=1}^{2} \sum_{j=0}^{i} \left(\frac{\alpha_s}{2\pi}\right)^i d_{i,j}[L(\mu^2 z^2)]^j, \tag{D2}$$

and

$$L(\mu^2 z^2) \equiv 2\gamma_E + \ln\frac{\mu^2 z^2}{4}, \tag{D3}$$

$$d_{1,0} = \frac{5C_F}{2}, \tag{D4}$$

$$d_{1,1} = \frac{3C_F}{2}, \tag{D5}$$

$$d_{2,0} = C_F\left[\frac{469\beta_0}{48} + \frac{223C_F - 94C_A}{96} + \frac{\pi^2(8C_F - 15C_A)}{36} + 2(C_A - 4C_F)\zeta(3)\right], \tag{D6}$$

$$d_{2,1} = \left[C_F^2\left(-\frac{5}{8} + \frac{2\pi^2}{3}\right) + C_F C_A\left(\frac{49}{24} - \frac{\pi^2}{6}\right) - \frac{5}{6}C_F n_f T_F\right] + \frac{5C_F}{2}\left(\frac{3}{2}C_F + \beta_0\right), \tag{D7}$$

$$d_{2,2} = \frac{3C_F}{4}\left(\frac{3}{2}C_F + \beta_0\right), \tag{D8}$$

$$n_{1,0}(\alpha) = C_F\left[2(1-\alpha) - \frac{1+\alpha^2}{1-\alpha} - \frac{4\ln(1-\alpha)}{1-\alpha}\right], \tag{D9}$$

$$n_{1,1}(\alpha) = -C_F\frac{1+\alpha^2}{1-\alpha}, \tag{D10}$$

as well as $n_{2,0}(\alpha)$, $n_{2,1}(\alpha)$, and $n_{2,2}(\alpha)$ in more complicated forms. The constants in the formulas are $C_F = 4/3$,

$T_F = 1/2$, $C_A = 3$, $n_f = 3$ (3 flavor in this work), and $\beta_0 = (11C_A - 4n_f T_F)/6$. With this setup, we then numerically saved $n_{NNLO}(\alpha, \mu^2 z^2)$ [mainly for $n_{2,j}(\alpha)$] for frequent calls.

As discussed in the main text of Sec. VIII, we express the light-cone ITD $Q(\lambda, \mu)$ by the DNN,

$$Q_{DNN}(\lambda, \mu) \equiv \frac{f_{DNN}(\boldsymbol{\theta}; \lambda)}{f_{DNN}(\boldsymbol{\theta}; 0)}, \tag{D11}$$

and minimize the loss function,

$$J(\boldsymbol{\theta}, r_{mod}) \equiv \frac{\eta}{2}\boldsymbol{\theta} \cdot \boldsymbol{\theta} + \frac{1}{2}\chi^2(\boldsymbol{\theta}, r_{mod}). \tag{D12}$$

The $\frac{\eta}{2}\boldsymbol{\theta} \cdot \boldsymbol{\theta}$ term is to make sure the DNN represented function has good shape and is smooth. The χ^2 is defined as

$$\chi^2(\boldsymbol{\theta}, r_{mod})$$

$$= \sum_{P_z > P_z^0}^{P_z^{max}} \sum_{z_{min}}^{z_{max}} \frac{(\mathcal{M}(z, P_z, P_z^0) - \mathcal{M}_{DNN}(z, P_z, P_z^0; \boldsymbol{\theta}, r_{mod}))^2}{\sigma^2(z, P_z, P_z^0)}, \tag{D13}$$

with $\sigma(z, P_z, P_z^0)$ being the statistical errors of the ratio-scheme matrix elements $\mathcal{M}(z, P_z, P_z^0)$. We analytically solve the gradients of χ^2,

$$\frac{\partial\chi^2}{2\partial r_{mod}}$$

$$= \sum_{P_z > P_z^0}^{P_z^{max}} \sum_{z_{min}}^{z_{max}} \frac{\mathcal{M}(z, P_z, P_z^0) - \mathcal{M}_{DNN}(z, P_z, P_z^0)}{\sigma^2(z, P_z, P_z^0)[\tilde{Q}_{DNN}(z, P_z^0) + r_{mod}(aP_z^0)^2]}$$

$$\times [(aP_z)^2 - \mathcal{M}_{DNN}(z, P_z, P_z^0)(aP_z^0)^2], \tag{D14}$$

$$\frac{\nabla_{\boldsymbol{\theta}}\chi^2}{2}$$

$$= \sum_{P_z > P_z^0}^{P_z^{max}} \sum_{z_{min}}^{z_{max}} \frac{\mathcal{M}(z, P_z, P_z^0) - \mathcal{M}_{DNN}(z, P_z, P_z^0)}{\sigma^2(z, P_z, P_z^0)[\tilde{Q}_{DNN}(z, P_z^0) + r_{mod}(aP_z^0)^2]}$$

$$\times \frac{1}{f_{DNN}(0)}\left\{\nabla_{\boldsymbol{\theta}} f_{DNN}(zP_z)\right.$$

$$+ \int_{-1}^{1}[\nabla_{\boldsymbol{\theta}} f_{DNN}(\alpha zP_z) - \nabla_{\boldsymbol{\theta}} f_{DNN}(zP_z)]\frac{n_{NNLO}(\alpha, \mu^2 z^2)}{d_{NNLO}(\mu^2 z^2)}d\alpha$$

$$- \mathcal{M}_{DNN}(z, P_z, P_z^0)\left[\nabla_{\boldsymbol{\theta}} f_{DNN}(P_z^0 z)\right.$$

$$+ \int_{-1}^{1}[\nabla_{\boldsymbol{\theta}} f_{DNN}(\alpha P_z^0 z) - \nabla_{\boldsymbol{\theta}} f_{DNN}(P_z^0 z)]\frac{n_{NNLO}(\alpha, \mu^2 z^2)}{d_{NNLO}(\mu^2 z^2)}d\alpha\right]$$

$$+ r_{mod}[(aP_z)^2 - \mathcal{M}_{DNN}(z, P_z, P_z^0)(aP_z^0)^2]\nabla_{\boldsymbol{\theta}} f_{DNN}(0)\left.\right\}. \tag{D15}$$

To optimize the parameters, we apply the Adam optimization method [108] for the gradient descent in this work.

[1] A. C. Aguilar *et al.*, Eur. Phys. J. A **55**, 190 (2019).

[2] J. Collins, *Foundations of Perturbative QCD* (Cambridge University Press, Cambridge, England, 2013), Vol. 32.

[3] M. Oehm, C. Alexandrou, M. Constantinou, K. Jansen, G. Koutsou, B. Kostrzewa, F. Steffens, C. Urbach, and S. Zafeiropoulos, Phys. Rev. D **99**, 014508 (2019).

[4] X. Ji, Phys. Rev. Lett. **110**, 262002 (2013).

[5] X. Ji, Sci. China Phys. Mech. Astron. **57**, 1407 (2014).

[6] X. Ji, Y.-S. Liu, Y. Liu, J.-H. Zhang, and Y. Zhao, Rev. Mod. Phys. **93**, 035005 (2021).

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

[8] 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).

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

[10] S. Bhattacharya, K. Cichy, M. Constantinou, J. Dodson, X. Gao, A. Metz, S. Mukherjee, A. Scapellato, F. Steffens, and Y. Zhao, arXiv:2209.05373.

[11] X. Xiong, X. Ji, J.-H. Zhang, and Y. Zhao, Phys. Rev. D **90**, 014051 (2014).

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

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

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

[15] 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).

[16] K.-F. Liu and S.-J. Dong, Phys. Rev. Lett. **72**, 1790 (1994).

[17] W. Detmold and C. J. D. Lin, Phys. Rev. D **73**, 014501 (2006).

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

[19] M. Bhat, W. Chomicki, K. Cichy, M. Constantinou, J. R. Green, and A. Scapellato, Phys. Rev. D **106**, 054504 (2022).

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

[21] 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).

[22] C. Alexandrou, K. Cichy, M. Constantinou, J. R. Green, K. Hadjiyiannakou, K. Jansen, F. Manigrasso, A. Scapellato, and F. Steffens, Phys. Rev. D **103**, 094512 (2021).

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

[24] C. Alexandrou, K. Cichy, M. Constantinou, K. Hadjiyiannakou, K. Jansen, A. Scapellato, and F. Steffens, Phys. Rev. D **99**, 114504 (2019).

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

[26] Y.-S. Liu *et al.* (Lattice Parton Collaboration), Phys. Rev. D **101**, 034020 (2020).

[27] Y.-S. Liu, J.-W. Chen, L. Jin, R. Li, H.-W. Lin, Y.-B. Yang, J.-H. Zhang, and Y. Zhao, arXiv:1810.05043.

[28] 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).

[29] J.-W. Chen, L. Jin, H.-W. Lin, Y.-S. Liu, Y.-B. Yang, J.-H. Zhang, and Y. Zhao, arXiv:1803.04393.

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

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

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

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

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

[35] M. Constantinou *et al.*, Prog. Part. Nucl. Phys. **121**, 103908 (2021).

[36] M. Constantinou *et al.*, arXiv:2202.07193.

[37] C. Alexandrou, M. Constantinou, K. Hadjiyiannakou, K. Jansen, and F. Manigrasso, Phys. Rev. D **104**, 054503 (2021).

[38] C. Alexandrou, M. Constantinou, K. Hadjiyiannakou, K. Jansen, and F. Manigrasso, Phys. Rev. Lett. **126**, 102003 (2021).

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

[40] Z. Fan, R. Zhang, and H.-W. Lin, Int. J. Mod. Phys. A **36**, 2150080 (2021).

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

[42] S. Bhattacharya, K. Cichy, M. Constantinou, A. Metz, A. Scapellato, and F. Steffens, Phys. Rev. D **104**, 114510 (2021).

[43] S. Bhattacharya, K. Cichy, M. Constantinou, A. Metz, A. Scapellato, and F. Steffens, Phys. Rev. D **102**, 114025 (2020).

[44] S. Bhattacharya, K. Cichy, M. Constantinou, A. Metz, A. Scapellato, and F. Steffens, Phys. Rev. D **102**, 034005 (2020).

[45] S. Bhattacharya, K. Cichy, M. Constantinou, A. Metz, A. Scapellato, and F. Steffens, Phys. Rev. D **102**, 111501 (2020).

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

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

[48] X. Gao, L. Jin, C. Kallidonis, N. Karthik, S. Mukherjee, P. Petreczky, C. Shugert, S. Syritsyn, and Y. Zhao, Phys. Rev. D **102**, 094513 (2020).

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

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

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

[53] P. C. Barry *et al.* (Jefferson Lab Angular Momentum (JAM), HadStruc Collaborations), Phys. Rev. D **105,** 114051 (2022).

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

[55] C. Shugert, X. Gao, T. Izubichi, L. Jin, C. Kallidonis, N. Karthik, S. Mukherjee, P. Petreczky, S. Syritsyn, and Y. Zhao, in *Proceedings of the 37th International Symposium on Lattice Field Theory* (2020), arXiv:2001.11650.

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

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

[58] X. Gao, N. Karthik, S. Mukherjee, P. Petreczky, S. Syritsyn, and Y. Zhao, Phys. Rev. D **103,** 094510 (2021).

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

[60] N. Karthik and R. Narayanan, Phys. Rev. D **106,** 014503 (2022).

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

[62] L.-B. Chen, W. Wang, and R. Zhu, Phys. Rev. Lett. **126,** 072002 (2021).

[63] X. Gao, A. D. Hanlon, S. Mukherjee, P. Petreczky, P. Scior, S. Syritsyn, and Y. Zhao, Phys. Rev. Lett. **128,** 142003 (2022).

[64] E. Follana, Q. Mason, C. Davies, K. Hornbostel, G. P. Lepage, J. Shigemitsu, H. Trottier, and K. Wong (HPQCD, UKQCD Collaborations), Phys. Rev. D **75,** 054502 (2007).

[65] A. Bazavov *et al.*, Phys. Rev. D **100,** 094510 (2019).

[66] X. Gao, N. Karthik, S. Mukherjee, P. Petreczky, S. Syritsyn, and Y. Zhao, Phys. Rev. D **104,** 114515 (2021).

[67] X. Gao, A. D. Hanlon, N. Karthik, S. Mukherjee, P. Petreczky, P. Scior, S. Syritsyn, and Y. Zhao, Phys. Rev. D **106,** 074505 (2022).

[68] J. Brannick, R. C. Brower, M. A. Clark, J. C. Osborn, and C. Rebbi, Phys. Rev. Lett. **100,** 041601 (2008).

[69] M. A. Clark, R. Babich, K. Barros, R. C. Brower, and C. Rebbi, Comput. Phys. Commun. **181,** 1517 (2010).

[70] R. Babich, M. A. Clark, B. Joo, G. Shi, R. C. Brower, and S. Gottlieb, in *Proceedings of the SC11 International Conference for High Performance Computing, Networking, Storage and Analysis* (2011), arXiv:1109.2935.

[71] M. A. Clark, B. Joó, A. Strelchenko, M. Cheng, A. Gambhir, and R. Brower, arXiv:1612.07873.

[72] E. Shintani, R. Arthur, T. Blum, T. Izubuchi, C. Jung, and C. Lehner, Phys. Rev. D **91,** 114511 (2015).

[73] G. S. Bali, B. Lang, B. U. Musch, and A. Schaefer, Phys. Rev. D **93,** 094515 (2016).

[74] J.-W. Chen, T. Ishikawa, L. Jin, H.-W. Lin, Y.-B. Yang, J.-H. Zhang, and Y. Zhao, Chin. Phys. C **43,** 103101 (2019).

[75] L. Maiani, G. Martinelli, M. Paciello, and B. Taglienti, Nucl. Phys. **B293,** 420 (1987).

[76] X. Ji, J.-H. Zhang, and Y. Zhao, Phys. Rev. Lett. **120,** 112001 (2018).

[77] J. Green, K. Jansen, and F. Steffens, Phys. Rev. Lett. **121,** 022004 (2018).

[78] T. Ishikawa, Y.-Q. Ma, J.-W. Qiu, and S. Yoshida, Phys. Rev. D **96,** 094019 (2017).

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

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

[81] H.-W. Lin, Phys. Lett. B **824,** 136821 (2022).

[82] H.-W. Lin, J.-W. Chen, and R. Zhang, arXiv:2011.14971.

[83] H.-W. Lin, Phys. Rev. Lett. **127,** 182001 (2021).

[84] X. Ji, Y. Liu, A. Schäfer, W. Wang, Y.-B. Yang, J.-H. Zhang, and Y. Zhao, Nucl. Phys. **B964,** 115311 (2021).

[85] Y.-K. Huo *et al.* (Lattice Parton Collaboration (LPC)), Nucl. Phys. **B969,** 115443 (2021).

[86] J. Hua, M.-H. Chu, P. Sun, W. Wang, J. Xu, Y.-B. Yang, J.-H. Zhang, and Q.-A. Zhang (Lattice Parton Collaboration), Phys. Rev. Lett. **127,** 062002 (2021).

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

[88] I. Balitsky and V. Braun, Nucl. Phys. **B311,** 541 (1989).

[89] A. V. Radyushkin, Phys. Lett. B **781,** 433 (2018).

[90] R. A. Briceño, J. V. Guerrero, M. T. Hansen, and C. J. Monahan, Phys. Rev. D **98,** 014511 (2018).

[91] H.-W. Lin and R. Zhang, Phys. Rev. D **100,** 074502 (2019).

[92] W.-Y. Liu and J.-W. Chen, Phys. Rev. D **104,** 054508 (2021).

[93] J. Collins, T. C. Rogers, and N. Sato, Phys. Rev. D **105,** 076010 (2022).

[94] J.-W. Chen, S. D. Cohen, X. Ji, H.-W. Lin, and J.-H. Zhang, Nucl. Phys. **B911,** 246 (2016).

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

[96] P. Petreczky and J. H. Weber, Eur. Phys. J. C **82,** 64 (2022).

[97] P. C. Barry, C.-R. Ji, N. Sato, and W. Melnitchouk (Jefferson Lab Angular Momentum (JAM) Collaboration), Phys. Rev. Lett. **127,** 232001 (2021).

[98] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, Neural Netw. **6,** 861 (1993).

[99] A. Kratsios, Ann. Math. Artif. Intell. **89,** 435 (2021).

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

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

[102] S. Shi, K. Zhou, J. Zhao, S. Mukherjee, and P. Zhuang, Phys. Rev. D **105,** 014017 (2022).

[103] L. Wang, S. Shi, and K. Zhou, Phys. Rev. D **106,** L051502 (2022).

[104] A. Bazavov, N. Brambilla, P. Petreczky, A. Vairo, and J. H. Weber (TUMQCD Collaboration), Phys. Rev. D **98,** 054511 (2018).

[105] P. Petreczky, S. Steinbeißer, and J. H. Weber, Proc. Sci. LATTICE2021 (**2022**) 471 [arXiv:2112.00788].

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

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

[108] D. P. Kingma and J. Ba, arXiv:1412.6980.