Unpolarized proton PDF at NNLO from lattice QCD with physical quark masses

Xiang Gao,[1,*] Andrew D. Hanlon[2,†] Jack Holligan[3,4,‡] Nikhil Karthik,[5,6] Swagato Mukherjee,[2]
Peter Petreczky,[2] Sergey Syritsyn,[7,8] and Yong Zhao[1]

[1]*Physics Division, Argonne National Laboratory, Lemont, Illinois 60439, USA*
[2]*Physics Department, Brookhaven National Laboratory, Bldg. 510A, Upton, New York 11973, USA*
[3]*Department of Physics, University of Maryland, College Park, Maryland 20742, USA*
[4]*Center for Frontier Nuclear Science, Stony Brook University, Stony Brook, New York 11794, USA*
[5]*American Physical Society, Hauppauge, New York 11788, USA*
[6]*Department of Physics, Florida International University, Miami, Florida 33199, USA*
[7]*RIKEN-BNL Research Center, Brookhaven National Laboratory, Upton, New York 11973, USA*
[8]*Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York 11790, USA*

We present a lattice QCD calculation of the unpolarized isovector quark parton distribution function (PDF) of the proton utilizing a perturbative matching at next-to-next-to-leading-order (NNLO). The calculations are carried out using a single ensemble of gauge configurations generated with $N_f = 2 + 1$ highly-improved staggered quarks with physical masses and a lattice spacing of $a = 0.076$ fm. We use one iteration of hypercubic smearing on these gauge configurations, and the resulting smeared configurations are then used for all aspects of the subsequent calculation. For the valence quarks, we use the Wilson-clover action with physical quark masses. We consider several methods for extracting information on the PDF. We first extract the lowest four Mellin moments using the leading-twist operator product expansion approximation. Then, we determine the x dependence of the PDF through a deep neural network within the pseudo-PDF approach and additionally through the framework of large-momentum effective theory utilizing a hybrid renormalization scheme. This is the first application of the NNLO matching coefficients for the nucleon directly at the physical point.

I. INTRODUCTION

Parton distribution functions (PDFs) are universal quantities that can be used to compute cross sections for various hard-scattering processes. Their universality makes them particularly useful, in that they can be determined from some subset of processes and then used in predictions of other processes. Additionally, their determination gives insights into the structure of hadrons. The partonic internal-structure of the nucleon was established long ago from deep inelastic scattering experiments at SLAC. This prompted significant experimental and theoretical efforts to reduce the uncertainty in these quantities. Using the data provided from several

experiments (e.g. the Tevatron, HERA, the LHC, etc.), the collinear structure for unpolarized and polarized nucleons has been determined to a few-percent accuracy from various recent global analyses [1–4]. The success of these programs has been significant, yet there is still a need to further reduce the uncertainties on PDFs, and much remains unknown about the full internal structure of the nucleon. For example, there is little experimental data for the distribution of a transversely polarized nucleon. Further, even less is known about the generalized parton distributions (GPDs) and the transverse momentum dependent (TMD) PDFs. These quantities will be better understood through data collected from the JLab 12 GeV upgrade [5] and the electron-ion collider [6,7].

Calculations of these quantities from first principles would be very useful. There has been significant progress in the necessary theoretical developments for the determination of PDFs from lattice QCD. This will allow for supplementing the existing experimental data in order to further reduce the global analysis uncertainties, as well as fill the gaps where little experimental data exists. For recent reviews on the methods and progress on computing light-cone PDFs from lattice QCD see e.g. Refs. [8–14].

[*]gaox@anl.gov
[†]ahanlon@bnl.gov
[‡]holligan@umd.edu

Although there are several methods that have been considered for the extraction of PDFs from the lattice, two methods in particular have been extensively used in recent years with tremendous success. These are the quasidistribution approach based on large-momentum effective theory (LaMET) [15,16], and the pseudodistribution approach [17,18] based on a short-distance factorization (SDF) that has also been applied in the current-current correlator approach [19,20]. The two methods become equivalent at infinite momentum, but differ in their systematics at finite momentum, and there has been much work regarding their differing advantages [21]. SDF relies on the validity of the leading-twist operator product expansion (OPE), which can also be used to directly extract the first few Mellin moments. While LaMET relies on large momentum to control the power corrections in the matching which become large for values of $x \sim 0$ and $x \sim 1$.

There have been several applications of these methods to the unpolarized isovector quark PDF of the nucleon, e.g. from the quasi-PDF approach [22–28] and the pseudo-PDF approach [18,29–34]. In this paper, we extract this PDF using both methods, which allows us to better understand the different systematics between each method. This is the first such calculation utilizing a matching kernel at next-to-next-to-leading order (NNLO) directly with physical quark masses.

The rest of the paper is organized as follows. First, in Sec. II we describe the lattice setup. Then, in Sec. III we extract the bare matrix elements needed for the PDF determination. Next, in Sec. IV we extract the first few Mellin moments via the leading-twist OPE approximation. In Sec. VI we use a deep neural network (DNN) to overcome the inverse problem that arises within the pseudeo-PDF method. Our final PDF extraction method is given in Sec. VII and uses a matching in x-space within the LaMET framework from hybrid renormalized matrix elements. Finally, our conclusions are given in Sec. VII.

II. LATTICE DETAILS

The setup and calculations used here are very similar to our previous work on the pion valence PDF [35,36]. For convenience, we repeat the more salient details.

We use a single ensemble of $N_f = 2 + 1$ highly improved staggered quarks (HISQ) [37] with physical masses, a size of $64^3 \times 64$, and a lattice spacing of $a = 0.076$ fm generated by the HotQCD collaboration [38]. In the valence sector, we use the tree-level tadpole-improved Wilson-clover action with $c_{sw} = u_0^{-3/4} = 1.0372$ (where u_0 is the plaquette expectation value on the hypercubic (HYP) smeared gauge configurations), physical quark masses, and one step of HYP smearing [39]. This is the same ensemble used in our previous work [36].

Our calculations are carried out on GPUs with the QLua software suite [40], which utilizes the multigrid solver in QUDA [41,42] for the quark propagators. In order to

TABLE I. Some details on the ensemble and the statistics gathered for our calculation. The n_z give the momentum projection at the sink for the three-point functions in integer units of the lattice. The k_z are the boost momentum for the smeared quark fields. The various sink-source separations are given by t_{sep}. And, the number of sources used for exact and sloppy solves is given by #ex and #sl, respectively.

Ensembles $a, L_t \times L_s^3$	m_π (GeV)	N_{cfg}	n_z	k_z	t_{sep}/a	(#ex,#sl)
$a = 0.076$ fm	0.14	350	0	0	6	(1, 16)
64×64^3			0	0	8,10	(1, 32)
			0	0	12	(2, 64)
			1	0	6,8,10,12	(1, 32)
			4	2	6	(1, 32)
			4	2	8,10,12	(4, 128)
			6	3	6	(1, 20)
			6	3	8	(4, 100)
			6	3	10,12	(5, 140)

reduce the computational cost, we employ all-mode averaging (AMA) [43], which involves performing several low-precision (sloppy) solves in addition to a small number of high-precision (exact) solves to correct for the bias introduced from the sloppy solves alone. The stopping criterion for our solver is 10^{-10} and 10^{-4} for exact and sloppy solves, respectively.

The nucleon interpolating operators used in this work are given by

$$N_\alpha^{(s)}(x,t) = \varepsilon_{abc} u_{a\alpha}^{(s)}(x,t)(u_b^{(s)}(x,t)^T C \gamma_5 d_c^{(s)}(x,t)), \quad (1)$$

where $C = \gamma_t \gamma_y$ is the charge-conjugation matrix, and the superscript on the quark fields indicates whether the quarks are smeared ($s = S$) or not ($s = P$). Additionally, we use momentum smearing [44] in order to improve the overlaps with highly boosted hadron states. Our choice for the boosted quarks is $\frac{2\pi k_z}{L}\hat{z}$, where k_z depends on the choice of the final momentum projection in our three-point functions $\vec{p}_f \equiv \frac{2\pi n_z}{L}\hat{z}$. Note that one might naively choose $k_z = n_z/3$, as there are three valence quarks in a baryon. However, it was shown in Ref. [44] that a value of $k_z \approx 0.45 n_z$ was the optimal choice for the nucleon at their pion mass $m_\pi \approx 295$ MeV, and we make similar choices for k_z in this work.

Some of the more important details of our setup can be found in Table I.

A. Two-point correlation functions

The two-point correlation functions we compute are given by

$$
\begin{aligned}
&C^{2pt}(\vec{p}, t_{sep}; \vec{x}, t_0) \\
&= \sum_{\vec{y}} e^{-i\vec{p}\cdot(\vec{y}-\vec{x})} \mathcal{P}_{\alpha\beta}^{2pt} \langle N_\alpha^{(s)}(\vec{y}, t_{sep} + t_0) \bar{N}_\beta^{(s')}(\vec{x}, t_0) \rangle, \quad (2)
\end{aligned}
$$

where we project to the positive parity states with $\mathcal{P}^{2\mathrm{pt}} = \frac{1}{2}(1 + \gamma_t)$. In order to increase statistics, we average these correlators with the negative time correlators while changing the projector to $\mathcal{P}^{2\mathrm{pt}} = \frac{1}{2}(1 - \gamma_t)$. The change in parity projector is needed because the backward propagating antibaryon is not the antiparticle of the forward propagating baryon but of its parity partner. We always use a smeared source ($s' = S$) but consider both smeared ($s = S$) and unsmeared ($s = P$) sinks which leads to SS and SP correlators respectively. The SS and SP correlators share the same energy eigenstates, but differ in their overlaps onto these states allowing for a more reliable determination of the spectrum by looking for agreement in the spectrum determined from each.

For each value of k_z, we computed two-point correlators with momentum projections at the sink of $\vec{p} = \frac{2\pi n_z}{L}\hat{z}$ for all $n_z \leq 10$, which is larger than the optimal n_z even for our largest k_z. All of these momenta are not strictly needed, as we only included a much smaller number of sink momentum projections in the three point functions. However, the larger range of momentum for the two-point functions comes at a minimal extra cost and gives us more confidence in our analysis to clearly see a consistent momentum dependence for the spectrum.

B. Three-point correlation functions

The three-point correlation functions we computed are of the form

$$C_\Gamma^{3\mathrm{pt}}(\vec{p}_f, \vec{q}, t_{\mathrm{sep}}, t_{\mathrm{ins}}, z; \vec{x}, t_0)$$
$$= \sum_{\vec{y}, \vec{z}_0} e^{-i\vec{p}_f \cdot (\vec{y} - \vec{x})} e^{-i\vec{q} \cdot (\vec{x} - \vec{z}_0)} \mathcal{P}_{\alpha\beta}^{3\mathrm{pt}}$$
$$\times \langle N_\alpha(\vec{y}, t_{\mathrm{sep}} + t_0) \mathcal{O}^\Gamma(\vec{z}_0 + z\hat{z}, t_{\mathrm{ins}} + t_0) \bar{N}_\beta(\vec{x}, t_0) \rangle, \quad (3)$$

where $\vec{q} = \vec{p}_f - \vec{p}_i$, and the inserted operator is given by

$$\mathcal{O}^\Gamma(\vec{z}_0 + z\hat{z}, t_{\mathrm{ins}} + t_0)$$
$$= \bar{q}(\vec{z}_0, t_{\mathrm{ins}} + t_0)\Gamma\tau_3$$
$$\times W(\vec{z}_0, t_{\mathrm{ins}} + t_0; \vec{z}_0 + z\hat{z}, t_{\mathrm{ins}} + t_0)q(\vec{z}_0 + z\hat{z}, t_{\mathrm{ins}} + t_0), \quad (4)$$

where $q(\vec{z}, t)$ is an isospin doublet consisting of the light quark fields u and d, W is a Wilson line of length z that connects the positions of the quark fields via a straight spatial path along the z-axis, and τ_3 results in the isovector combination (i.e. $\bar{u}\Gamma Wu - \bar{d}\Gamma Wd$) which leads to the cancellation of the disconnected diagrams. The gauge links entering the Wilson line are the 1-step HYP-smeared gauge links. The sink momenta considered here are all in the z-direction $\vec{p}_f = \frac{2\pi}{L}n_z\hat{z}$ with $n_z \in \{0, 1, 4, 6\}$ leading to momentum in physical units of $P_z = \{0, 0.25, 1.02, 1.53\}$ GeV, where $P_z \equiv \frac{2\pi}{L}n_z$.

In order to access the unpolarized PDF, we use $\mathcal{P}^{3\mathrm{pt}} = \frac{1}{2}(1 + \gamma_t)$. Additionally, we have the choice of $\Gamma = \gamma_t, \gamma_z$. But, to avoid mixing under renormalization, we use $\Gamma = \gamma_t$ [45,46].

III. ANALYSIS OF CORRELATION FUNCTIONS

Our goal is to obtain the bare ground-state matrix elements from the three-point functions. However, there can be significant unwanted contributions to these correlators coming from excited states. These contributions are especially important for the present work at the physical pion mass, as the contamination from higher states becomes more severe as the pion mass is lowered. Thus, in general we must include the effects of excited states in our analysis. In practice, it is very difficult to reliably separate these effects from the desired ground state matrix elements directly from the three-point functions. Fortunately, we can extract the lowest few energies from the two-point functions and use this information in our fits to the three-point functions.

A. Two-point function analysis

Our aim in this subsection is twofold. First, as already pointed out, we would like to determine the energies that contribute to the three-point functions. And, second, we need to understand at what t_{sep} and t_{ins} various energies are no longer relevant. In Fig. 1, we show the effective energies for all two-point correlators we have computed, along with predictions for the ground-state energies for all $n_z \leq 10$ using the ground-state energy for $n_z = 0$ with the continuum dispersion relation.
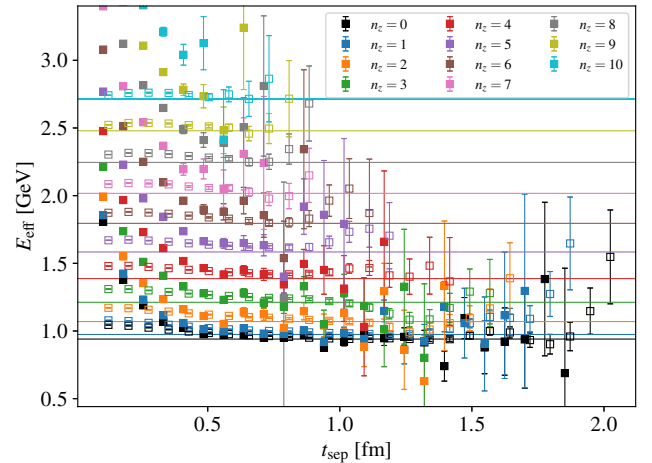

FIG. 1.   The effective energies from SS (filled) and SP (open) two-point correlators for all $n_z \leq 10$. The horizontal lines correspond to predictions of the ground-state energies for each n_z using the continuum dispersion relation with the estimate for the ground-state energy extracted from a one-state fit to the $n_z = 0$ SS two-point correlator.
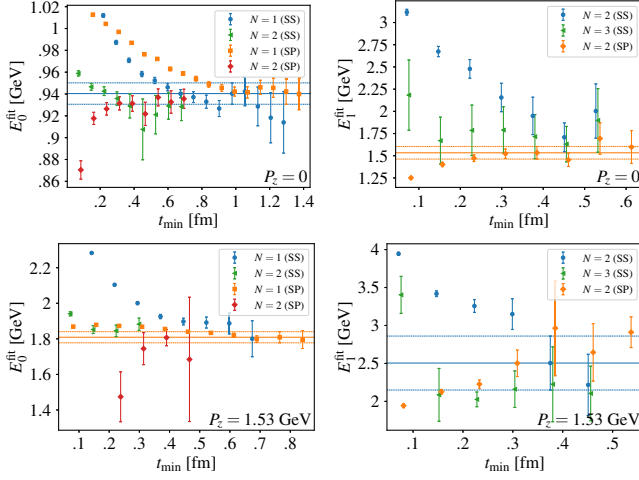
FIG. 2. The $t_{\min}$ dependence of the (left) ground state and (right) first excited state for (upper) $P_z = 0$ and (lower) $P_z = 1.53$ GeV. The fits on the left-hand side are done without any constraints, while the fits on the right-hand side are done with the ground-state energy constrained from the best one-state fit to the correlators with the same momentum. The legend indicates the number of states N included in the fit function and the correlator being fit to (i.e. either SS or SP).

The fit form we use comes from the spectral decomposition for the two-point functions

$$C^{2\mathrm{pt}}(\vec{p}, t_{\mathrm{sep}}) = \sum_n A_\alpha^{(n)}(\vec{p}) A_\alpha^{(n)}(\vec{p})^* e^{-E_n(\vec{p}) t_{\mathrm{sep}}}, \quad (5)$$

where $A_\alpha^{(n)} \equiv \langle \Omega | N_\beta \mathcal{P}_{\beta\alpha}^{2\mathrm{pt}} | n, \vec{p} \rangle$ ($|\Omega\rangle$ denotes the vacuum state), thermal effects are ignored, and t_0 is shifted to zero. We then truncate the number of contributing states to N and rearrange to find

$$C_N^{2\mathrm{pt}}(\vec{p}, t_{\mathrm{sep}}) = C_0 e^{-E_0 t_{\mathrm{sep}}} \left[1 + \sum_{i=1}^{N-1} R_i \prod_{j=1}^{i} e^{-\Delta_{j,j-1} t_{\mathrm{sep}}} \right], \quad (6)$$

where $\Delta_{j,j-1} \equiv E_j(\vec{p}) - E_{j-1}(\vec{p})$, $R_i = C_i/C_0$, and $C_n = A_\alpha^{(n)}(\vec{p}) A_\alpha^{(n)}(\vec{p})^*$. This particular form is useful in that it guarantees a proper ordering of the states.

As a first step, we perform unconstrained one-state fits to all the two-point correlators, and then pick the best fit for each momentum. For each momentum, the different fits involve two choices: (i) either SS or SP correlators, and (ii) the minimum time separation $t_{\min}$ included in the fit. The largest time separation $t_{\max}$ included in the fit is chosen to be the largest time separation such that $|C(t)| \geq 3\delta C(t)$ for all $t \leq t_{\max}$, where $\delta C(t)$ is the error of $C(t)$. The best fit is determined based on observing stability in the results from small changes in $t_{\min}$ in combination with a p-value $\gtrsim 0.1$. If multiple fits satisfy these criteria, then fits to SS correlators are preferred and $t_{\min}$ is chosen to lie roughly in the middle of the region of stability. In the left-hand side of Fig. 2, we show the dependence of the ground-state energy on the choice of $t_{\min}$ for the smallest and largest values of
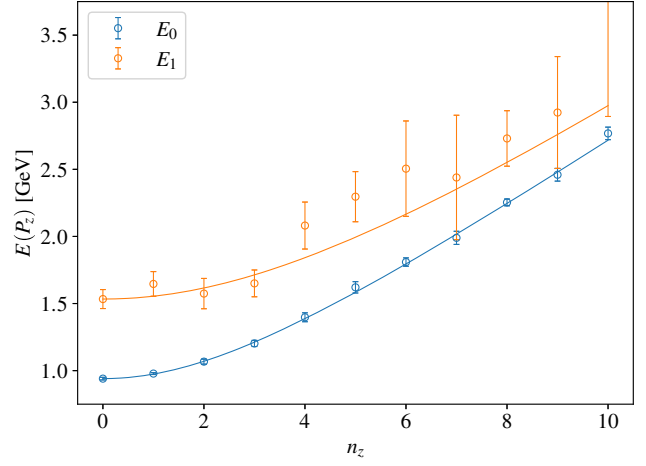
FIG. 3. The ground (blue) and first excited (orange) states from one- and two-state fits, respectively, for each momentum with $n_z \leq 10$. The curves come from predictions for the energies with $n_z \neq 0$ using the continuum dispersion relation with the mean value of the $n_z = 0$ energies.

momentum used in the three-point functions. For each momentum, we then perform two-state fits with a prior on the ground state energy coming from the best unconstrained one-state fit for the same momentum. The right-hand side of Fig. 2 shows the dependence of the first-excited state energy on the choice of $t_{\min}$ for the same momenta as in the ground state figures. Our best fit estimates for the first two energies with $n_z = 0$ are

$$E_0 = 0.9345(57) \text{ GeV}, \qquad E_1 = 1.533(71) \text{ GeV}. \quad (7)$$

In Fig. 3, the ground state and first excited state from one- and two-state fits, respectively, are plotted as a function of momentum. The mean value of the energies for $n_z = 0$ are used to predict the energies for $n_z \neq 0$ using the continuum dispersion relation. These predictions are shown as lines in Fig. 3. As expected, the ground state energies follow the continuum dispersion relation quite well, with a mass consistent with the proton, giving confidence in our control over excited states, even at large momentum. More surprisingly, the first excited state appears to also follow a continuum dispersion relation with a mass that is a little larger than the Roper resonance, which is unexpected for calculations using the Wilson-clover action [47]. The small disagreement with the Roper mass could also be from uncontrolled finite-volume effects, which are generally very important for resonances. Further, we know that several multi-hadron states should lie below our estimate of the first-excited state, but our results suggest our operators must have very poor overlap onto those states.

B. Three-point function analysis

The standard approach to extract the bare matrix elements is to form an appropriate ratio of the
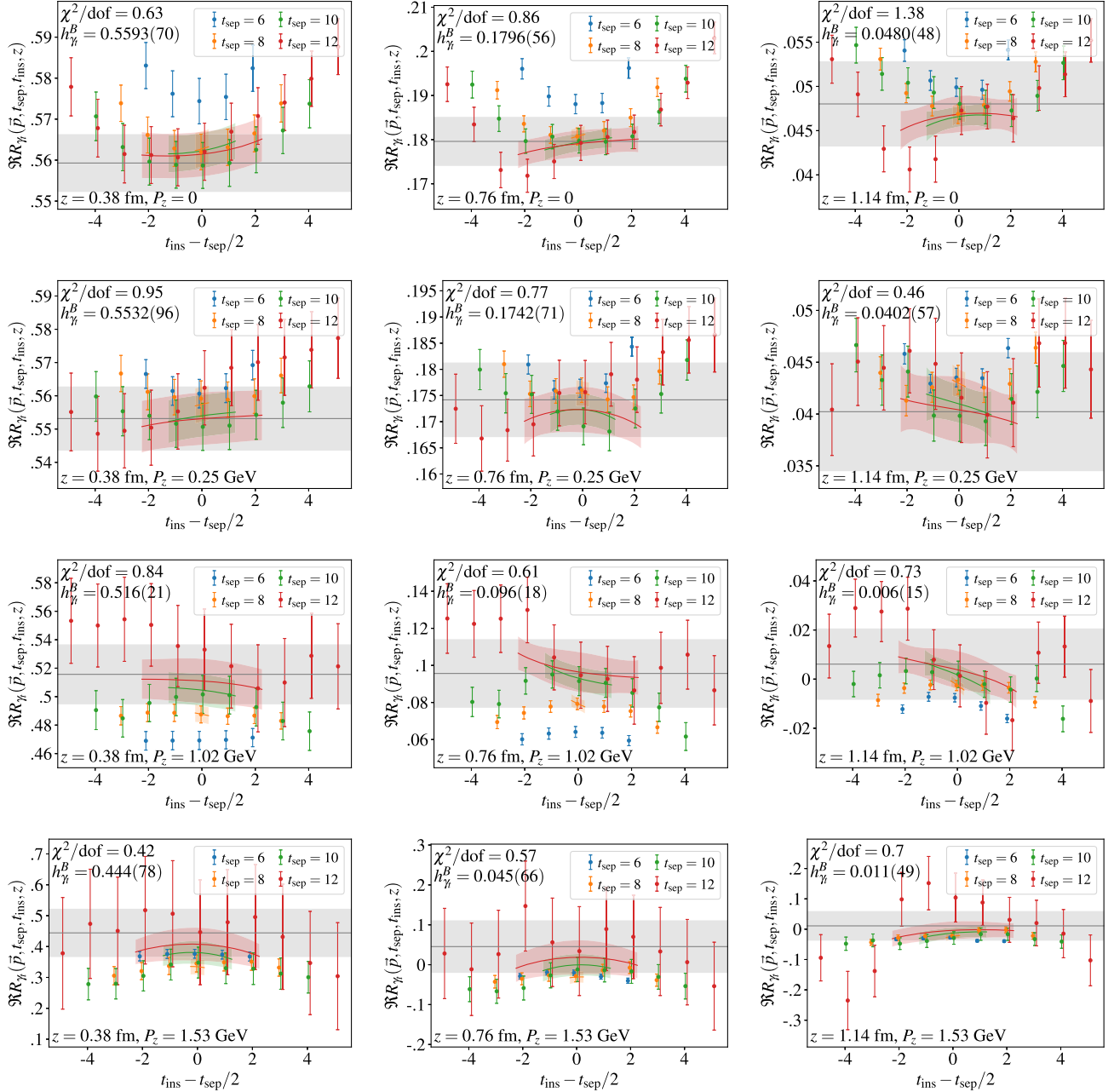
FIG. 4. The real parts of the ratio of three-point to two-point functions for all values of momentum (one for each row) and a few representative values of Wilson-line length z (one for each column). The χ^2/dof reported, estimate for the ground-state bare matrix element (also represented by a gray band), and t_{sep} fit bands come from the preferred fit strategy, i.e. the two-state fit to the ratio R_{γ_t} with $n_{\mathrm{exc}} = 3$, where n_{exc} is the number of data points nearest both the source and sink that are not included in the fit. The range in t_{ins} of the t_{sep} bands covers the included data points in the fit.

three-point functions to the two-point functions. This is convenient in that in the limit as t_{sep} and t_{ins} become large, the ratio approaches the desired ground-state bare matrix element. As an added benefit, the ratio leads to a cancellation of correlations, resulting in reduced statistical errors.

To determine the needed ratio, it is beneficial to give the spectral decomposition of the three-point function

$$C_\Gamma^{3\mathrm{pt}}(\vec{p}_f, \vec{p}_i, t_{\mathrm{sep}}, t_{\mathrm{ins}}, z)$$
$$= \sum_{m,n} A_\alpha^{(m)}(\vec{p}_f) A_\alpha^{(n)}(\vec{p}_i)^* h_\Gamma^B(\vec{p}_f, m; \vec{p}_i, n; z)$$
$$\times e^{-E_m(\vec{p}_f)(t_{\mathrm{sep}} - t_{\mathrm{ins}})} e^{-E_n(\vec{p}_i) t_{\mathrm{ins}}}, \quad (8)$$

where $A_\alpha^{(n)} \equiv \langle \Omega | N_\beta \mathcal{P}_{\beta\alpha}^{3\mathrm{pt}} | n, \vec{p} \rangle$, $h_\Gamma^B(\vec{p}_f, m; \vec{p}_i, n; z)$ are the bare matrix elements, thermal effects are ignored, and it is
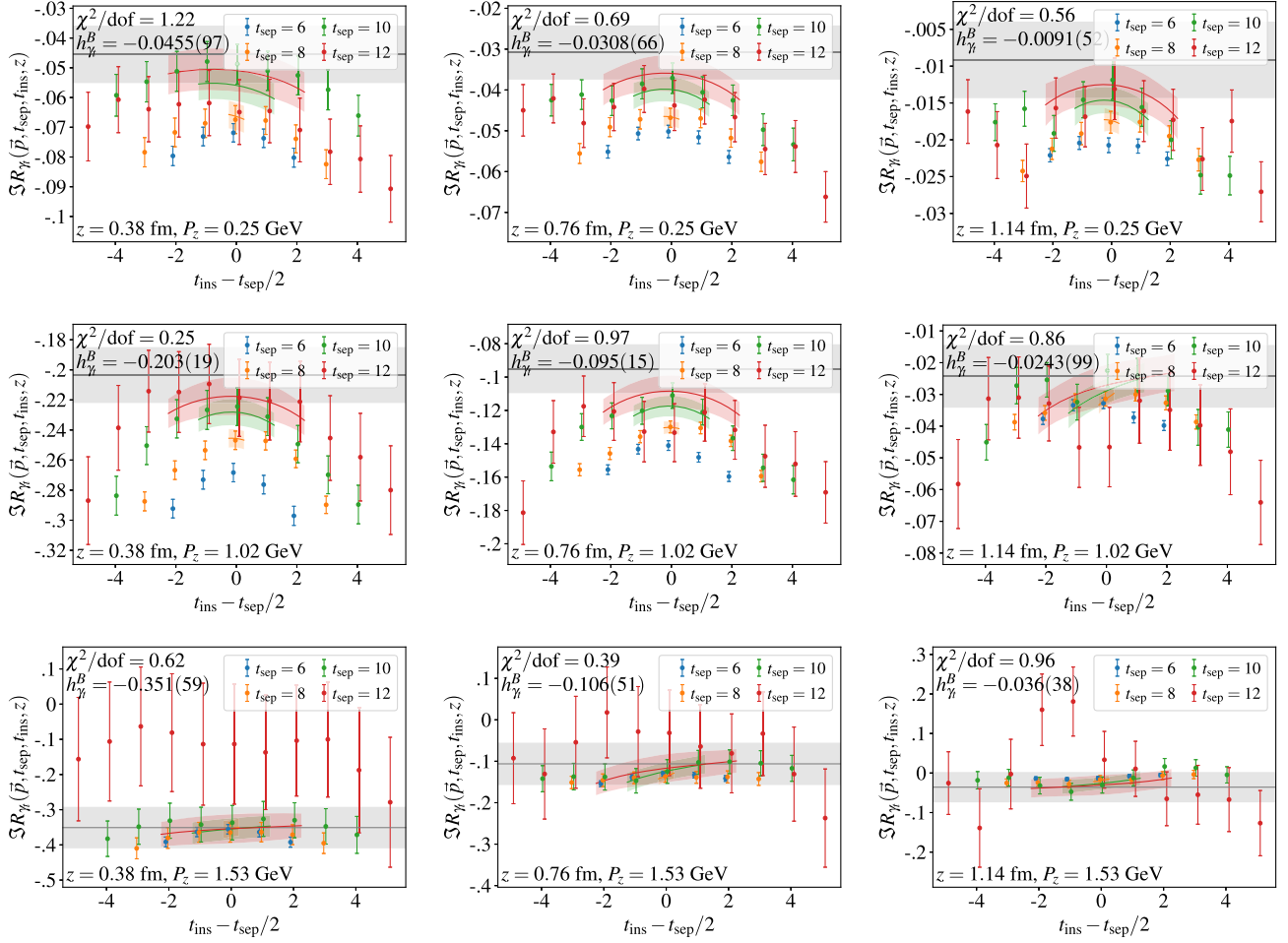
FIG. 5.   The same as Fig. 4, but for the imaginary parts. Additionally, the zero-momentum matrix elements are not shown, as these are all consistent with zero (as they are expected to be).

assumed that t_0 has been shifted to zero. The energies entering the spectral decomposition are the same for both the three-point and two-point functions. However, the overlap factors are only the same if $\mathcal{P}^{\mathrm{2pt}}$ and $\mathcal{P}^{\mathrm{3pt}}$ are chosen appropriately. In our case, this is trivially true since we consider $\mathcal{P}^{\mathrm{2pt}} = \mathcal{P}^{\mathrm{3pt}} = \frac{1}{2}(1 + \gamma_t)$. Therefore, in the case where $\vec{p} \equiv \vec{p}_f = \vec{p}_i$, used in this work, we form

$$R_{\gamma_t}(\vec{p}, t_{\mathrm{sep}}, t_{\mathrm{ins}}, z) \equiv \frac{C_{\gamma_t}^{\mathrm{3pt}}(\vec{p}, \vec{p}, t_{\mathrm{sep}}, t_{\mathrm{ins}}, z)}{C^{\mathrm{2pt}}(\vec{p}, t_{\mathrm{sep}})}, \quad (9)$$

where we now indicate the use of $\Gamma = \gamma_t$ (and continue to do so throughout the remainder of the paper). This ratio can easily be seen to obey

$$\lim_{t_{\mathrm{ins}}, t_{\mathrm{sep}} \to \infty} R_{\gamma_t}(\vec{p}, t_{\mathrm{sep}}, t_{\mathrm{ins}}, z) = h_{\gamma_t}^B(\vec{p}, 0; \vec{p}, 0; z). \quad (10)$$

The fit functions we consider are based on truncating the spectral decomposition in both the numerator and denominator of $R_{\gamma_t}(\vec{p}, t_{\mathrm{sep}}, t_{\mathrm{ins}}, z)$ to the same number of states N

$$R_{\gamma_t}^N(\vec{p}, t_{\mathrm{sep}}, t_{\mathrm{ins}}, z)$$
$$= \frac{\sum_{m,n=0}^{N-1} h_{\gamma_t}^{\prime B} \prod_{l,k,r=1}^{m} e^{-\Delta_{l,l-1}t_{\mathrm{sep}}} e^{(\Delta_{k,k-1} - \Delta_{r,r-1})t_{\mathrm{ins}}}}{1 + \sum_{i=1}^{N-1} R_i \prod_{j=1}^{i} e^{-\Delta_{j,j-1}t_{\mathrm{sep}}}}, \quad (11)$$

where $\Delta_{i,i-1}$, R_i, and $h_{\gamma_t}^{\prime B}$ are the fit parameters, and we have suppressed the function arguments. For all of our fits, we prior the $\Delta_{i,i-1}$ and R_i from the corresponding two-point function fit. Note that the $h_{\gamma_t}^{\prime B}(\vec{p}, z; m, n)$ can be written in terms of the original matrix elements found in Eq. (8) as

$$h_{\gamma_t}^{\prime B}(\vec{p}, z; m, n) \equiv \frac{A_\alpha^{(m)}(\vec{p}) A_\alpha^{(n)}(\vec{p})^* h_{\gamma_t}^B(\vec{p}, m; \vec{p}, n; z)}{A_\alpha^{(0)}(\vec{p}) A_\alpha^{(0)}(\vec{p})^*}. \quad (12)$$

For convenience, in what follows, we use $h_{\gamma_t}^B(z, P_z, a)$ to denote the ground-state bare matrix element, where a is the lattice spacing, as these are the only matrix elements used in the subsequent analysis.

Further, we also consider the summation method as an alternative strategy, which sums $R_{\gamma_t}(\vec{p}, t_{\mathrm{sep}}, t_{\mathrm{ins}}, z)$ over a subset of $t_{\mathrm{ins}} \in [n_{\mathrm{exc}} + 1, t_{\mathrm{sep}} - n_{\mathrm{exc}} - 1]$
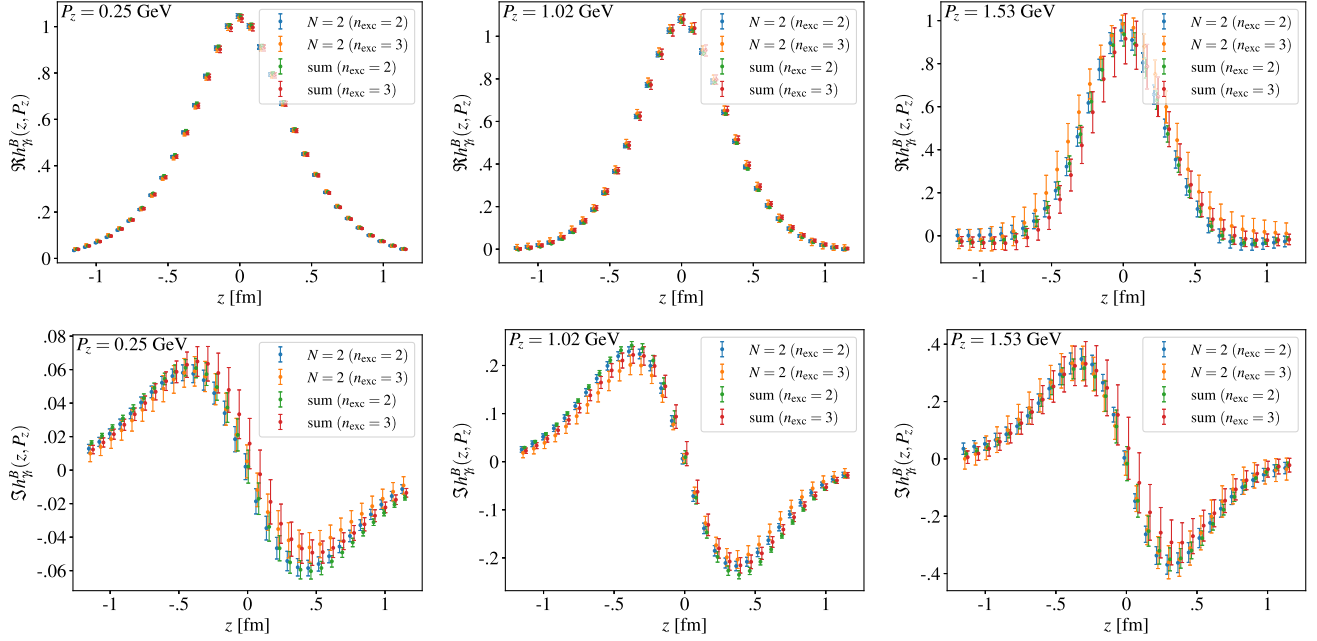
FIG. 6.   The Wilson-line length dependence of the (upper) real and (lower) imaginary parts of the ground-state bare matrix elements from two-state and summation fits with $n_{\text{exc}} = 2, 3$ for the three nonzero values of momentum considered (one for each column).

$$S_{\gamma_t}^{n_{\text{exc}}}(\vec{p}, t_{\text{sep}}, z) \equiv \sum_{t_{\text{ins}} = n_{\text{exc}} + 1}^{t_{\text{sep}} - n_{\text{exc}} - 1} R_{\gamma_t}(\vec{p}, t_{\text{sep}}, t_{\text{ins}}, z), \quad (13)$$

which in turn reduces the leading excited-state contamination to be $\mathcal{O}(e^{-\Delta_{1,0} t_{\text{sep}}})$, as opposed to $\mathcal{O}(e^{-\Delta_{1,0} t_{\text{sep}}/2})$ present in the ratio of Eq. (9). The summed ratio can then be fit to a linear function

$$S_{\gamma_t}^{n_{\text{exc}}}(\vec{p}, t_{\text{sep}}, z) = B_0 + t_{\text{sep}} h_{\gamma_t}^B(z, P_z, a). \quad (14)$$

We will use the summation fits as a consistency check for our multi-state fits directly to the ratio of three-point to two-point functions.

For our limited number of sink-source separations and statistics, we find the three-state fits to be unreliable, therefore we only consider fits that include up to two states in the ratio of Eq. (9). Further, to reduce the effects of excited states as much as possible, we remove some of the insertion times near the source/sink symmetrically. The number of excluded times n_{exc} is given such that the insertion times included in the fit are $t_{\text{ins}} \in [n_{\text{exc}} + 1, t_{\text{sep}} - n_{\text{exc}} - 1]$. In order to retain enough data for our fits, we only consider $n_{\text{exc}} \leq 3$ meaning that all fits will include insertion times in which more than two states are contributing, as evidenced from significant t_{min} dependence still present in the two-state fits shown in Fig. 2 for $t_{\text{sep}} \leq n_{\text{exc}} + 1$. We must therefore use an effective value for the gap $\Delta_{1,0}$ that mocks up the effects of all higher excited states present for the smallest/largest insertion time included in the fits. To this

end, we simply use the extracted value of $\Delta_{1,0}$ coming from a two-state fit to the appropriate SS two-point function with $t_{\text{min}} = n_{\text{exc}} + 1$.

In order to reduce the effects of excited states as much as possible, our preferred fit is the two-state fit to the ratio Eq. (9) with $n_{\text{exc}} = 3$, which completely excludes $t_{\text{sep}} = 6$ from the fit. In Figs 4 and 5, we show the results using this fit strategy for all computed values of momentum and a few representative values of Wilson-line length z. Then, in Fig. 6, we compare the extracted ground-state bare matrix elements as a function of the Wilson-line length z from two-state fits and the summation method with $n_{\text{exc}} = 2, 3$. There is reasonable consistency among the various fits, but there is still some tension in a handful of cases, which further motivates our use of the more conservative fits with $n_{\text{exc}} = 3$.

IV. MELLIN MOMENTS FROM THE LEADING-TWIST OPE

The bare matrix elements $h_{\gamma_t}^B(z, P_z, a)$ are multiplicatively renormalizable, and therefore we can cancel the renormalization factors, which only depend on the lattice spacing a and the Wilson-line length z, by forming the ratio [26]

$$\mathcal{M}_{\gamma_t}(\lambda, z^2; P_z^0) = \frac{h_{\gamma_t}^B(z, P_z, a)}{h_{\gamma_t}^B(z, P_z^0, a)} \Big/ \frac{h_{\gamma_t}^B(0, P_z, a)}{h_{\gamma_t}^B(0, P_z^0, a)}, \quad (15)$$

where $\lambda \equiv z P_z$ is referred to as the Ioffe time and the ratio itself is the Ioffe time pseudo-distribution (pseudo-ITD). This quantity is thus a renormalization-group invariant quantity. The additional $z = 0$ matrix elements appearing

in the ratio are not strictly required but they enforce an exact normalization and further reduce correlations and systematics. The lattice spacing dependence of the pseduo-ITD is suppressed for convenience. The ratio for the specific case of $P_z^0 = 0$, used in what follows, is referred to as reduced pseudo-ITD [18,29–34], and will be denoted by $\mathcal{M}_{\gamma_t}^0(\lambda, z^2)$.

Using the OPE for the γ_t matrix elements, we can extract the first few Mellin moments by fitting the pseudo-ITD to

$$
\mathcal{M}_{\gamma_t}(\lambda, z^2; P_z^0)
$$
$$
= \frac{\sum_{n=0} C_n(\mu^2 z^2) \frac{(-i\lambda)^n}{n!} \langle x^n \rangle(\mu) + \mathcal{O}(\Lambda_{\mathrm{QCD}}^2 z^2)}{\sum_{n=0} C_n(\mu^2 z^2) \frac{(-i\lambda^0)^n}{n!} \langle x^n \rangle(\mu) + \mathcal{O}(\Lambda_{\mathrm{QCD}}^2 z^2)}, \quad (16)
$$

where $C_n(\mu^2 z^2)$ are Wilson coefficients which have been computed up to next-to-next-to-leading order (NNLO) [48,49], $\langle x^n \rangle(\mu)$ are the Mellin moments at a factorization scale μ defined by

$$
\langle x^n \rangle(\mu) = \int_{-1}^{1} \mathrm{d}x \, x^n q^{u-d}(x, \mu), \quad (17)
$$

and

$$
q^{u-d}(x, \mu) \equiv \begin{cases} q^u(x, \mu) - q^d(x, \mu), & x \geq 0 \\ q^{\bar{u}}(-x, \mu) - q^{\bar{d}}(-x, \mu), & x < 0, \end{cases} \quad (18)
$$

where $q^f(x, \mu)$ and $q^{\bar{f}}(x, \mu)$ are the PDFs for the quark and antiquark of flavor f defined for $x \in [0, 1]$. Additionally, although the effects are small, we include the target mass corrections by making the following substitution:

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

The Wilson coefficients depend on the strong coupling constant $\alpha_s(\mu)$, and we use the same estimates as in Ref. [36], which gives $\alpha_s(\mu = 2 \text{ GeV}) = 0.2930$.

There are a few things to note regarding our fits to the ratio of OPEs. First, we consider the leading-twist approximation where the higher-twist corrections that come in as $\mathcal{O}(\Lambda_{\mathrm{QCD}}^2 z^2)$ are ignored. Therefore, we must empirically determine at what value of z^2 this approximation breaks down, which can be done by looking for a strong dependence of our results on z^2. Second, since we have chosen $P_z^0 = 0$, the denominator in Eq. (16) becomes unity. With these simplifications, it can easily be seen that the real and imaginary parts of the ratio of OPEs in Eq. (16) correspond to even and odd moments, respectively. Therefore, we separately fit the real and imaginary parts of the reduced pseudo-ITD to
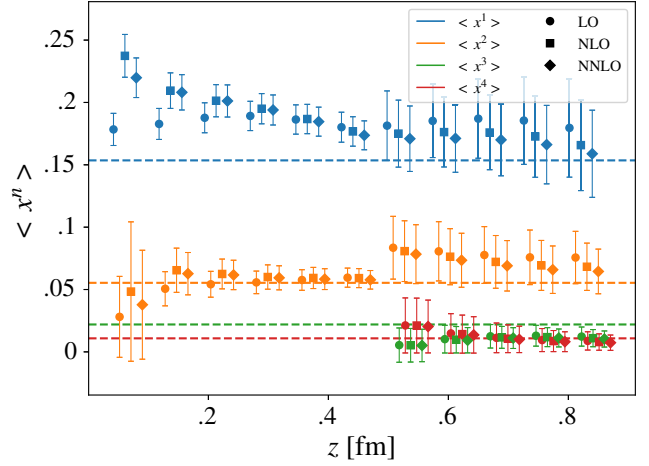

FIG. 7. Results for the lowest four Mellin moments as a function of z from fits of the reduced pseudo-ITD at fixed z with $n_z \in [1, 4, 6]$ to the leading-twist OPE using LO, NLO, and NNLO Wilson coefficients evaluated at $\mu = 2$ GeV. Only the first two moments are extracted for $z \leq 6a$. The horizontal dashed lines correspond to the central values of the moments extracted from the global analysis of NNPDF4.0 [4] defined at a scale $\mu = 2$ GeV.

$$
\mathrm{Re}\mathcal{M}_{\gamma_t}^0(\lambda, z^2) = \sum_{n=0}^{\lfloor N_{\max}/2 \rfloor} C_{2n}(\mu^2 z^2) \frac{(-i\lambda)^{2n}}{(2n)!} \langle x^{2n} \rangle,
$$
$$
\mathrm{Im}\mathcal{M}_{\gamma_t}^0(\lambda, z^2) = \sum_{n=1}^{\lceil N_{\max}/2 \rceil} C_{2n-1}(\mu^2 z^2) \frac{(-i\lambda)^{2n-1}}{(2n-1)!} \langle x^{2n-1} \rangle,
$$
$$
(20)
$$

respectively, where the moments $\langle x^n \rangle$ are the fit parameters (except $\langle x^0 \rangle$, which is fixed to one) and $N_{\max}$ is the largest moment considered.

As a first test of the validity of the leading-twist approximation, we perform fits including data at a fixed value of z^2 only. As the higher-twist effects enter as $\mathcal{O}(\Lambda_{\mathrm{QCD}}^2 z^2)$, observation of a dependence in the extracted moments from a fixed-z^2 analysis on z would likely indicate non-negligible higher-twist effects, invalidating the leading-twist approximation in the region of z where this dependence is observed. However, it should be noted that additional systematics from discretization effects and large logs (see Appendix B in Ref. [36]) can lead to dependence of the moments on z in the small-z region (i.e. $z \lesssim 0.2$ fm) as well. Additionally, the fixed-z^2 analysis also gives us an opportunity to determine the dependence on the perturbative order of the Wilson coefficients. The results are shown in Fig. 7, with a comparison to the moments determined from the global analysis of NNPDF4.0 [4] shown as dashed lines. The fits use a value of $\mu = 2$ GeV when evaluating the Wilson coefficients, and the NNPDF4.0 results are also defined at the scale $\mu = 2$ GeV. We found that for values of $z \leq 6a \sim 0.456$ fm, the data is only
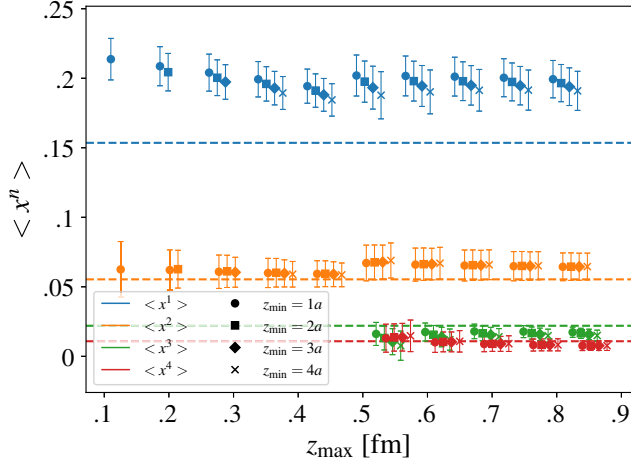
FIG. 8. Results for the lowest four Mellin moments from uncorrelated fits of the reduced pseudo-ITD to the leading-twist OPE as a function of z_{max}, with $z \in [z_{min}, z_{max}]$ and $n_z \in [1, 4, 6]$. The results use the NNLO Wilson coefficients evaluated at $\mu = 2$ GeV. Only the first two moments are considered for $z_{max} \leq 6a$. The horizontal dashed lines are the same as in Fig. 7.

sensitive to the lowest Mellin moment. Therefore, we only include a higher moment in the fits when $z > 6a$. This can be understood by evaluating the reduced pseudo-ITD using Eq. (16) with the moments extracted from the NNPDF4.0 global analysis [4]. From this, we can determine the dependence on the number of included moments in the leading-twist OPE. We find that for $z \lesssim 0.4$ fm, there are no significant effects for any $N_{max} \geq 2$. Therefore, a choice of $N_{max} = 2$ is sufficient in this region. But, beyond this, including a higher moment begins to make a difference.

There are a few interesting observations from these fits. First, the only dependence on z or the perturbative order is for $\langle x \rangle$. Further, the perturbative order of the Wilson coefficients only seems to matter at very small values of

z, where there is some mild z-dependence for $\langle x \rangle$ beyond leading order. Following this, the results are rather independent of z for $z \gtrsim 0.4$ fm. These effects are likely some combination of discretization effects and the need for resummation of large logs which was done in [36,50,51].

Next, we consider the inclusion of a range of z for our fits, where we still include all three values of nonzero momentum. Up to this point, all fits have been fully correlated, however we found difficulty in obtaining reliable correlated fits when including multiple values of z due to a high condition number for the covariance matrix. For the cases in which a correlated fit was possible, the results are in agreement with a fully uncorrelated fit. Therefore, we continue to use uncorrelated fits in what follows. The results are shown in Fig. 8.

As expected from the fixed-z^2 analysis, larger values of z_{min} tend to bring the value for $\langle x \rangle$ down closer to the global fit results. But, there is generally good agreement among all fits. As our final fit result, we choose $z_{min} = 4a$ to avoid discretization effects and large logs, and $z_{max} = 10a$ to avoid higher-twist effects. The fit results using these choices for the real and imaginary parts of the reduced pseudo-ITD are shown in Fig. 9. Our final results for the lowest four Mellin moments at $\mu = 2$ GeV are

$$\langle x \rangle = 0.191(14), \quad \langle x^3 \rangle = 0.0149(47), \quad \chi^2_{dof} = 0.97,$$
$$\langle x^2 \rangle = 0.065(10), \quad \langle x^4 \rangle = 0.0083(43), \quad \chi^2_{dof} = 0.26,$$
$$(21)$$

which can be compared to the results from NNPDF4.0 also defined at $\mu = 2$ GeV

$$\langle x \rangle = 0.15355(52), \quad \langle x^3 \rangle = 0.02197(12),$$
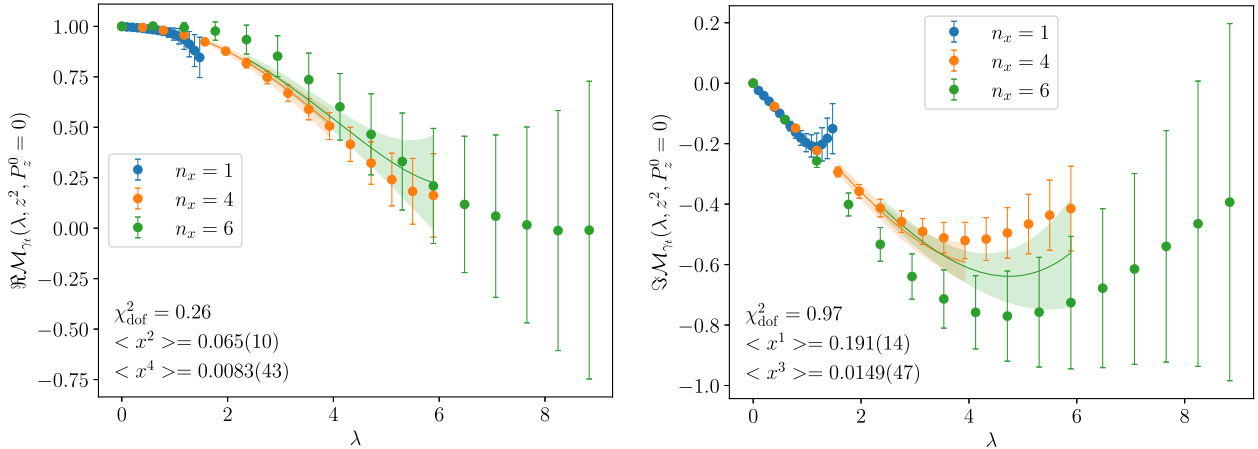$$\langle x^2 \rangle = 0.05536(22), \quad \langle x^4 \rangle = 0.010846(78). \quad (22)$$

FIG. 9. The (left) real and (right) imaginary parts of the reduced pseudo-ITD for the three momentum used in this work. The data come from our preferred fit strategy described in Sec. III B. The shaded bands correspond to the fits using the leading-twist OPE with NNLO Wilson coefficients evaluated at $\mu = 2$ GeV and including two moments in both the real and imaginary parts.

Some discussion is in order regarding our extracted value of $\langle x \rangle$, which is $2 - 3\sigma$ larger than the value obtained from phenomenological global fits. There is a long history of lattice calculations consistently obtaining larger values as well. First, there are several calculations that also utilized nonlocal operators to obtain the first few moments which see similar disagreements with the estimates from the global fits [29–31,34,52]. Further, these moments can also be extracted from local twist-2 operators, of which there are several lattice calculations which also see a value larger than expected from experiment [53–56]. However, there was an older study consistent with the global fits that argued excited-state effects could cause larger extractions for $\langle x \rangle$ [57]. Finally, a recent study also obtained results consistent with the global phenomenological fits by improving their analysis strategy, focusing on excited state contamination, which also suggested that discretization effects could produce shifts of $\langle x \rangle$ upwards [58]. More work is needed to fully resolve this issue.

V. PDF FROM LEADING-TWIST OPE: DNN RECONSTRUCTION

We now turn our attention to the extraction of the PDF itself. In this section, we use the pseudo-distribution approach [17,18], which requires the solution of an ill-posed inverse problem that we solve via the use of a deep neural network (DNN). We used this method in our previous work [36], and we repeat the pertinent details here.

Additionally, once we obtain the PDF, the moments can be extracted from it, allowing for a comparison to our extractions in the previous section.

A. Method

The leading-twist factorization formula for the renormalized matrix element can be written as

$$h^R(z, P_z, \mu) = \int_{-1}^{1} d\alpha \mathcal{C}(\alpha, \mu^2 z^2) \int_{-1}^{1} dy e^{-iy\alpha\lambda} q(y, \mu), \quad (23)$$

where $q(y, \mu)$ is the light-cone PDF, and the matching kernel $\mathcal{C}(\alpha, \mu^2 z^2)$ can be inferred from the Wilson coefficients $C_n(\mu^2 z^2)$ [17,18,59], which connect the position-space matrix elements to the x-dependent PDFs. However, it has been shown that the ratio-scheme data $\mathcal{M}_{\gamma_t}(\lambda, z^2; P_z^0)$ only contains information on the first few moments of the PDFs [36,60,61]. To determine the x-dependence with limited information, priors have to be used. Conventionally one may choose certain models inspired by the end-point behavior of the PDFs like,

$$q(x) = A x^\alpha (1 - x)^\beta (1 + \text{subleading terms}), \quad (24)$$

where the subleading terms can be modeled [60] which however may lead to a bias. To reduce the model bias, more general functional forms like the Jacobi polynomial

basis [32,33,62] are proposed to make a more flexible PDF parametrization. In addition, the deep neural network technique, which in principle can approximate any function with enough complexity in a smooth and unbiased manner, is probably the most flexible method. The DNN has been proposed to parametrize the full PDF function [63,64] and the Ioffe-time distribution [36]. In this work, instead we apply the DNN to only represent the subleading terms of the PDFs, which limits its contribution, in order to keep the end-point behavior and avoid any serious overfitting.

In our case, the real and imaginary parts of the ratio-scheme matrix elements $\mathcal{M}_{\gamma_t}(\lambda, z^2; P_z^0)$ with $P_z^0 = 0$ are related to $q^-(x)$ and $q^+(x)$, respectively, (e.g. see Ref. [34])

$$q^-(x) \equiv q^u(x) - q^d(x) - (q^{\bar{u}}(x) - q^{\bar{d}}(x)),$$
$$q^+(x) \equiv q^u(x) - q^d(x) + (q^{\bar{u}}(x) - q^{\bar{d}}(x)), \quad (25)$$

which are defined for $x \in [0, 1]$ (see Eq. (18) and the surrounding text). Using the DNN we parametrize these functions as

$$q^-(x; \alpha^-, \beta^-, \boldsymbol{\theta}^-) \equiv A^- x^{\alpha^-} (1 - x)^{\beta^-}$$
$$\times [1 + \delta \sin(f^-_{\text{DNN}}(x, \boldsymbol{\theta}^-))]$$
$$q^+(x; \alpha^+, \beta^+, A^+, \boldsymbol{\theta}^+) \equiv A^+ x^{\alpha^+} (1 - x)^{\beta^+}$$
$$\times [1 + \delta \sin(f^+_{\text{DNN}}(x, \boldsymbol{\theta}^+))], \quad (26)$$

where A^- is fixed by the normalization condition for q^-, i.e. $\int_0^1 dx q^-(x; \alpha^-, \beta^-, \boldsymbol{\theta}^-) = 1$, and A^+ is a free parameter. The $f^-_{\text{DNN}}(x, \boldsymbol{\theta}^-)$ and $f^+_{\text{DNN}}(x, \boldsymbol{\theta}^+)$ are DNN functions whose contributions are limited by $|\delta \sin(f^\pm_{\text{DNN}})| \lesssim \delta$. With this setup, we can make sure the DNNs are only subleading contributions. In the case of $\delta = 0$, contributions from the DNN are disabled and the forms in Eq. (26) simply reduce to the two $(q^-(x; \alpha^-, \beta^-))$ and three $(q^+(x; \alpha^+, \beta^+, A^+))$ parameter model fit. Although in this work we fix δ to be a constant, it could also be a function of x if one assumes the contribution from subleading terms varies for different local x.

The DNN functions $f^\pm_{\text{DNN}}(x, \boldsymbol{\theta}^\pm)$ are composite multi-step iterative functions, constructed layer by layer. The first layer is made up of a single node (i.e. a_1^1) and corresponds to the value of the input variable x. Then the hidden layers first perform a linear transformation,

$$z_i^{(l)} = b_i^{(l)} + \sum_j W_{ij}^{(l)} a_j^{(l-1)}, \quad (27)$$

followed by the nonlinear activation $\sigma^{(l)}(z_i^{(l)})$ whose output gives the input to the next layer $a_i^{(l)}$. The particular activation function we used is the so-called exponential linear unit $\sigma_{\text{elu}}(z) = \theta(-z)(e^z - 1) + \theta(z)z$. Finally, the
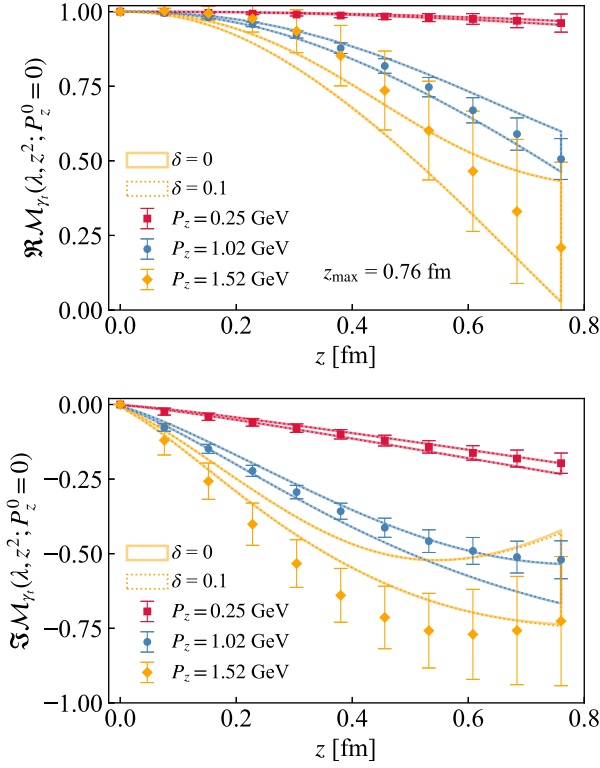
FIG. 10. The DNN training results using the ratio-scheme renormalized matrix elements in the range $z \in [2a, 10a]$ (27 data points) for the real part (upper panel) and imaginary part (lower panel) are shown. The results using $\delta = 0$ and $\delta = 0.1$ are shown as the solid and dotted curves, respectively.
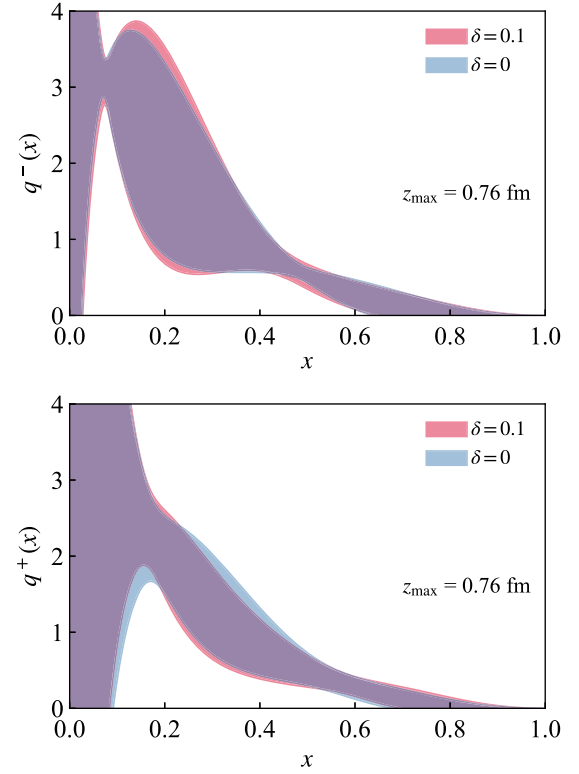

FIG. 11. The DNN represented PDFs using the ratio-scheme renormalized matrix elements in the range $z \in [2a, 10a]$ (27 data points) for the real part (upper panel) and imaginary part (lower panel) are shown. The results with $\delta = 0$ and $\delta = 0.1$ are shown as the blue and red bands, respectively.

last layer produces the output $f_{\mathrm{DNN}}^{\pm}(x, \boldsymbol{\theta}^{\pm})$, which is then used to evaluate $q^{\pm}(x; A^{\pm}, \alpha^{\pm}, \beta^{\pm}, \boldsymbol{\theta}^{\pm})$. The lower indices $i = 1, \ldots, n^{(l)}$ label the particular node within the lth layer, where $n^{(l)}$ is the number of nodes in the lth layer. The upper indices, in parenthesis, $l = 1, \ldots, N$ label the individual layers, where N is the number of layers (i.e the depth of the DNN). The biases $b_i^{(l)}$ and weights $W_{ij}^{(l)}$, denoted by $\boldsymbol{\theta}^{\pm}$, are the DNN parameters to be optimized (trained) by minimizing the loss function,

$$ J(\boldsymbol{\theta}^{\pm}) \equiv \frac{\eta}{2} \boldsymbol{\theta}^{\pm} \cdot \boldsymbol{\theta}^{\pm} + \frac{1}{2} \chi^2(\boldsymbol{\theta}^{\pm}, \alpha^{\pm}, \beta^{\pm}, \ldots), \quad (28) $$

where the first term is to prevent overfitting and makes sure the DNN-represented function is well behaved and smooth. The definition and details of the χ^2 function can be found in the Appendix. Given our low statistics, a simpler network structure like $\{1, 16, 16, 1\}$ (where each entry gives the number of nodes in each layer) is good enough to approximate the $f_{\mathrm{DNN}}^{\pm}(x, \boldsymbol{\theta}^{\pm})$ smoothly. Practically, we vary η from 10^0 to 10^{-2}, and tried network structures of size $\{1, 16, 16, 1\}$, $\{1, 16, 16, 16, 1\}$ and $\{1, 32, 32, 1\}$. We found the results remain unchanged. We therefore chose $\eta = 0.1$ and the DNN structure with four layers, including the input/output layer, to be $\{1, 16, 16, 1\}$.

B. DNN represented PDF

We train $q^-(x)$ and $q^+(x)$ using the ratio-scheme data $\mathcal{M}_{\gamma_t}(\lambda, z^2; P_z^0)$ with $z \in [2a, z_{\max}]$, where we skip the first point $z = a$ to avoid the most serious discretization effects. In the upper panel of Fig. 10, we show the results using $z \in [2a, 10a]$ (27 data points in total) for the real part with $\delta = 0$ (solid curves) and $\delta = 0.1$ (dotted curves). As one can see, the curves from the DNN go through the data points well, especially for the smaller momenta which are more precise. We vary δ to control the contributions from the DNN, where larger δ allows more flexibility of the PDF parametrization. However from $\delta = 0$ to $\delta = 0.1$, the fit results barely change, which is also reflected in the χ^2 (total number of 27 data points) which only evolves from 14.078 to 14.077 between the different fits. The corresponding PDFs are shown in the upper panel of Fig. 11, where the error bands increase for larger δ. These observation suggest the DNN provides a very flexible parametrization but the data is very limited so that the simple two-parameter model $q^-(x; A^-, \alpha^-, \beta^-)$ can already describe the data well. It is expected that for the imaginary part $q^+(x)$, the three-parameter (A^+, α^+, β^+) fit will be extremely unstable. With the idea that antiquarks are supposed to have little contribution at large x
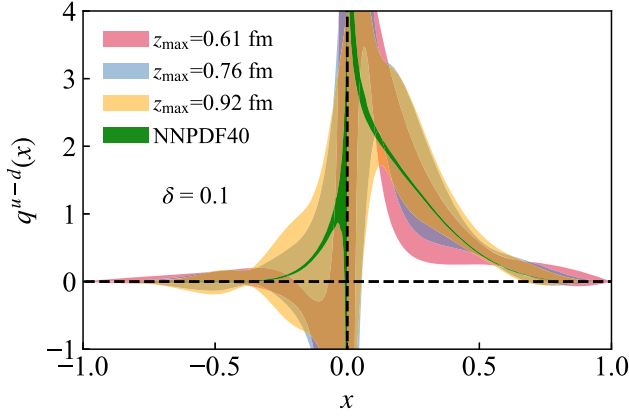
FIG. 12. The isovector PDF of the nucleon using data in the range $z \in [2a, z_{\max}]$ and $\delta = 0.1$ are shown. For comparison, we show the global analysis results from NNPDF4.0 [4].

for the nucleon, where $q^+(x)$ and $q^-(x)$ are both dominated by $q^u(x) - q^d(x)$, we choose to prior β^+ in $q^+(x)$ using the result for β^- in $q^-(x)$. The results for the imaginary part of $\mathcal{M}_{\gamma_t}(\lambda, z^2; P_z^0)$ and for $q^+(x)$ are shown in the lower panels of Fig. 10 and Fig. 11, respectively. Similarly, with a larger δ, the fit results do not change much, with χ^2 evolving from 21.28 to 21.09 and the error bands of the PDFs becoming slightly larger. In what follows, we set $\delta = 0.1$, but larger values of δ will play an important role for more precise data in the future.

As higher-twist contamination can enter at large values of z, we train the PDFs using different $z_{\max}$ to study its dependence. The results are shown in Fig. 12, where we combine $q^-(x)$ and $q^+(x)$ to obtain the isovector PDF $q^{u-d}(x)$, defined for $x \in [-1, 1]$, using the relations in Eq. (25). It is observed that the $z_{\max}$ dependence is small relative to the large errors, suggesting that higher-twist effects are less important compared to the statistics for our data. Compared to the global analysis from NNPDF4.0 [4], our results have significantly larger errors, leading to agreement with the global analysis results in most regions of x. We summarize the fit results and the χ^2 in Table II. As one can see, the errors on the fit parameters are quite large due to the limited statistics we were able to obtain. The first

few moments inferred from the model fit results are also shown, which are consistent with the model-independent extraction in Sec. IV.

VI. x-SPACE MATCHING

This section demonstrates the calculation of the unpolarized proton PDF using the method of LaMET. This offers a consistency check with the previous method of the leading-twist OPE combined with the deep neural network. In addition, it offers a more direct approach to the x-dependent PDF due to the ability to extrapolate to infinite distance and, thus, Fourier transform to momentum space.

A. Method

The data is analyzed by the same method as that laid out in [35]. The process involves renormalizing the matrix elements in the hybrid-scheme; extrapolation of the renormalized matrix elements to infinite distance; Fourier transforming the matrix elements to momentum space and, finally, matching our data to the light-cone PDF.

1. Renormalization

The first step is to renormalize the bare matrix elements which are multiplicatively renormalizable [65–67] as

$$h_{\gamma_t}^{B}(z, P_z, a) = e^{-\delta m(a)|z|} Z_{\gamma_t}(a) h_{\gamma_t}^{R}(z, P_z), \quad (29)$$

where $Z_{\gamma_t}(a)$ contains the logarithmic ultraviolet (UV) divergences and is independent of z, and $\delta m(a)$ includes the linear UV divergences coming from the Wilson-line self-energy. Although the ratio scheme can be used at small values of z where the OPE is valid, we need a method that does not introduce nonperturbative effects in the IR region at large values of z. Thus, we use the recently developed hybrid scheme [68] which explicitly includes the factor involving $\delta m(a)$ at large z.

An estimate for $\delta m(a)$ can be determined from the static quark-antiquark potential, leading to a value of $a\delta m(a) = 0.1597(16)$ taken from Refs. [69–73]. However, the quantity $\delta m(a)$ has a scheme dependence, and in order to match to $\overline{\text{MS}}$ we must determine the necessary shift

TABLE II. The parameters from the DNN training with $\delta = 0.1$ using the data with $z \in [2a, z_{\max}]$ are shown, where $N_{\text{data}} = 21, 27, 33$ for $z_{\max} = 0.61, 0.76, 0.92$ fm. The moments inferred from the model fit results are also shown.

	$z_{\max}$ [fm]	α^-	β^-	A^-	χ^2/N_{data}	$\langle x^2 \rangle$	$\langle x^4 \rangle$
$q^-(x)$	0.61	$-0.47(0.75)$	1.38(2.38)		0.64	0.071(14)	0.022(18)
	0.76	0.01(1.12)	2.72(3.45)		0.52	0.070(12)	0.015(12)
	0.92	0.16(1.18)	3.49(3.76)		0.44	0.069(12)	0.014(9)

	$z_{\max}$ [fm]	α^+	β^+	A^+	χ^2/N_{data}	$\langle x \rangle$	$\langle x^3 \rangle$
$q^+(x)$	0.61	$-0.99(1.15)$	1.38(2.38)	0.25(0.73)	0.74	0.202(19)	0.027(16)
	0.76	$-0.09(1.84)$	2.72(3.45)	1.61(6.94)	0.78	0.201(18)	0.030(12)
	0.92	0.47(2.12)	3.49(3.76)	4.6(22.1)	0.80	0.200(17)	0.031(10)

$\bar{m}_0 \equiv \delta m' - \delta m$ where $\delta m'$ subtracts the linear divergence in the Wilson line and matches the result to the $\overline{\text{MS}}$ scheme which includes an $\mathcal{O}(\Lambda_{\text{QCD}})$ renormalon ambiguity [74]. Following Ref. [35], $\bar{m}_0$ can be estimated from fitting a ratio of the bare matrix elements, including factors of $e^{\delta m(a)z}$, to a form motivated from the OPE

$$e^{\delta m(a)(z-z_0)} \frac{h^B_{\gamma_t}(z, 0, a)}{h^B_{\gamma_t}(z_0, 0, a)} = e^{-\bar{m}_0(z-z_0)} \frac{C_0(\mu^2 z^2) - \Lambda z^2}{C_0(\mu^2 z_0^2) - \Lambda z_0^2}, \quad (30)$$

where $z_0 = 3a \sim 0.228$ fm, $C_0(\mu^2 z^2)$ is computed at NNLO and is the only leading-twist Wilson coefficient contributing to the OPE at zero momentum, and the terms $(\bar{m}_0, \Lambda)$ are fit parameters. The value of z_0 is chosen large enough to neglect discretization effects and small enough to neglect higher-twist effects which become important for $z \gtrsim 0.2$ fm. The Λ term allows for the inclusion of larger values of z in order to capture some of the higher-twist effects.

In fixed order perturbation theory, the two parameters $(\bar{m}_0, \Lambda)$ depend on the renormalization scale μ. As such, we must use a different set of fitting parameters for our calculation of the full PDF at different energy scales. The two aforementioned parameters are independent of the external state. We fit the parameters $\bar{m}_0$ and Λ from both the nucleon and pion matrix elements, with the latter calculated in Ref. [36], and find any tension would lead to differences on the order of 3% for the expected linear power corrections [68] at the largest value of $P_z \sim 1.53$ GeV, which is much less than the statistical uncertainty. Therefore, we continue with the use of the parameters fitted from the proton matrix elements. The $\bar{m}_0$ and Λ values computed at different energy scales and up to different values of z_{max} are shown in Fig. 13.

Finally, we can form the renormalized matrix elements in the hybrid scheme as

$$\begin{aligned} &h^R_{\gamma_t}(\lambda, \lambda_s, P_z, \mu) \\ &= N \frac{h^B_{\gamma_t}(z, P_z, a)}{h^B_{\gamma_t}(z, 0, a)} \frac{C_0(\mu^2 z^2) - \Lambda z^2}{C_0(\mu^2 z^2)} \theta(z_s - z) \\ &+ N e^{\delta m'(z-z_s)} \frac{h^B_{\gamma_t}(z, P_z, a)}{h^B_{\gamma_t}(z_s, 0, a)} \frac{C_0(\mu^2 z_s^2) - \Lambda z_s^2}{C_0(\mu^2 z_s^2)} \theta(z - z_s), \quad (31) \end{aligned}$$

where $N = h^B_{\gamma_t}(0, 0, a)/h^B_{\gamma_t}(0, P_z, a)$ is a normalization, the correction Λz^2 to $h^B_{\gamma_t}(z, 0, a)$ is included for small z, we choose $z_s = 3a \sim 0.228$ fm, we have explicitly included $\delta m' = \delta m + \bar{m}_0$ at large z, and the lattice spacing dependence for the renormalized matrix elements have been suppressed for convenience. The form of the hybrid scheme comes from using the ratio scheme for $z < z_s$; for $z > z_s$, we use the matrix element with the linear divergence removed by the exponential involving $\delta m'$. The additional factors on the side of $z > z_s$ are used to enforce continuity at $z = z_s$.
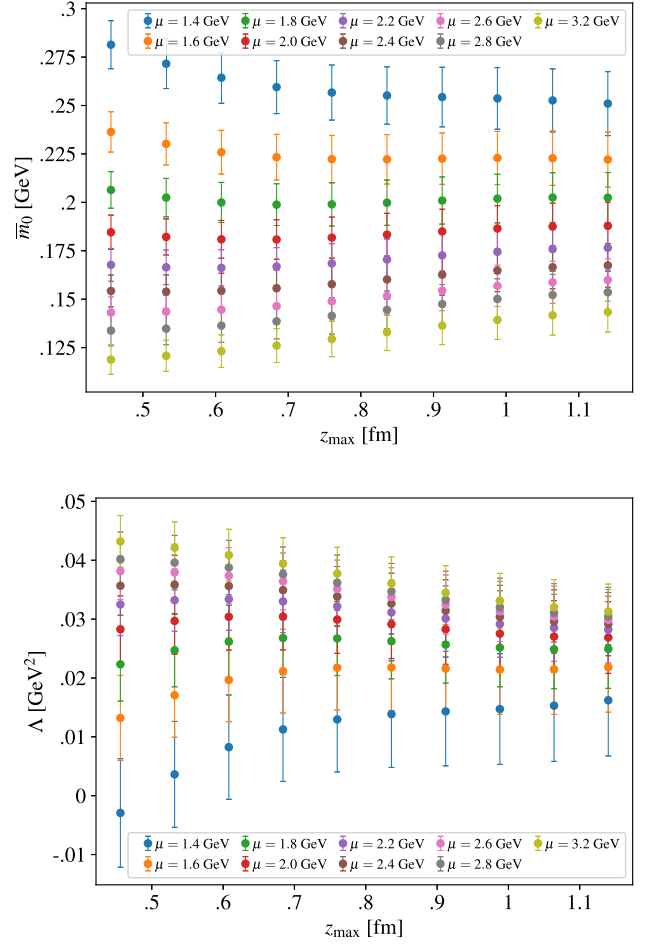

FIG. 13. (Upper) $\bar{m}_0$ and (lower) Λ values computed from fits using Eq. (30) at different energy scales μ, $z_0 = 3a \sim 0.228$ fm, and up to different maximum z-values, z_{max}. The Wilson coefficient $C_0(\mu^2 z^2)$ is computed at NNLO.

2. Large-λ extrapolation

In order to avoid unphysical oscillations in x-space, it is important that the Fourier transform not be truncated, which requires the matrix elements up to infinite λ. Of course, the lattice calculation can only produce values up to $\lambda_{\text{max}} = P_z z_{\text{max}}$. Additionally, the signal can quickly deteriorate at large values of λ. Therefore, we extrapolate to infinite λ and perform the Fourier transform by a discrete sum over the data up to some maximum $\lambda_L = P_z z_L$ beyond which the extrapolation function is integrated to infinity. With sufficiently large λ_{max}, the matrix elements fall close to zero at λ_{max}, so the extrapolation will mainly affect the small x region which is outside our prediction with LaMET.

We use the exponential decay model

$$\frac{A e^{-m_{\text{eff}} \lambda / P_z}}{|\lambda|^d}, \quad (32)$$

where (A, m_{eff}, d) are fit parameters with the constraints $m_{\text{eff}} > 0.1$ GeV, $A > 0$, and $d > 0$, as was done in Ref. [35]. The constraint $m_{\text{eff}} > 0.1$ GeV helps to ensure
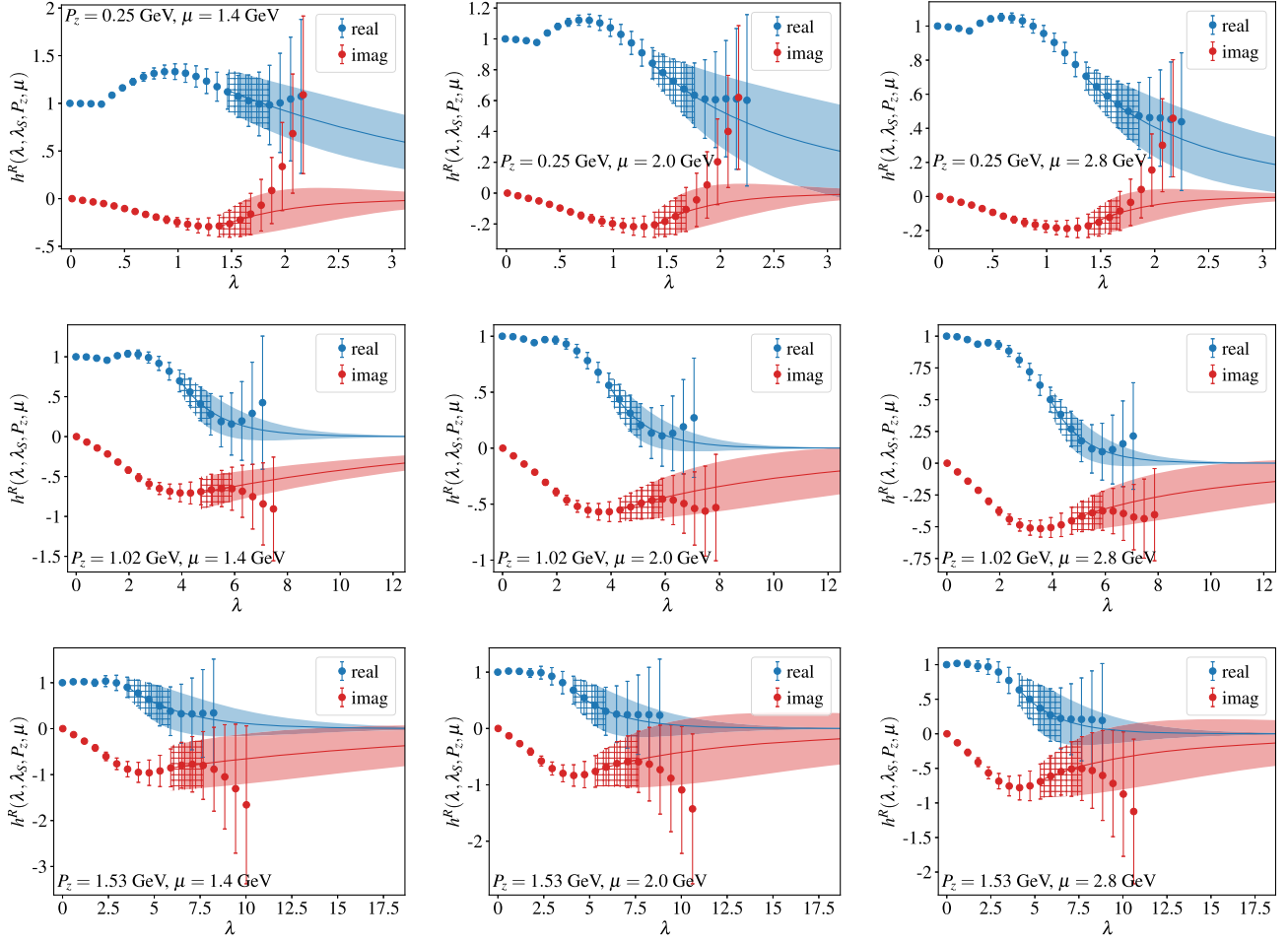
FIG. 14. The real (blue) and imaginary (red) parts of the hybrid renormalized matrix elements along with their large-λ extrapolations using Eq. (32). The three rows from top to bottom correspond to momenta $P_z = 0.26$, 1.02, and 1.53 GeV; the three columns from left to right correspond to energy scales $\mu = 1.4$, 2.0, and 2.8 GeV. The data points themselves come from the lattice calculation, while the bands correspond to the extrapolation fit result in the range where we use the extrapolation rather than the data. The hatches show the range of λ used for the extrapolation fit.

suppression at large values of λ, and does not noticeably affect the Fourier transform in the moderate-to-large x region. This large-distance behavior can be derived under the auxiliary field formulation of the Wilson line [65,67], where the nonlocal quark bilinear operator can be expressed as the product of two heavy-to-light currents. At large spacelike separation, this current product drops off exponentially with the decay constant given by the binding energy of the heavy-light meson. From this behavior, the extrapolation model of Eq. (32) is derived. A more detailed derivation of the large-λ model is provided in App. B of [35].

The real and imaginary parts of $h_{\gamma_t}^R(\lambda, \lambda_s, P_z, \mu)$ are fitted separately to Eq. (32). Care must be taken in choosing the data points used in the fit to the extrapolation model. Data points at too large a value of λ will have a poor signal-to-noise ratio and give large uncertainties in the extrapolation parameters. Data points at too small a value of λ will not capture the exponential decay expected at finite momentum. Our general guide is to select points for which

$h_{\gamma_t}^R(\lambda, \lambda_s, P_z, \mu)$ becomes compatible with zero (if such points exist) or just before the point in which the matrix element begins to grow, contrary to the exponential decay expected from Eq. (32).

Results of these fits for various P_z and μ are shown in Fig. 14, where the hatches indicate the range of λ used for the fit while the shaded bands start from λ_L. The value of λ corresponding to a data point that most closely resembles the extrapolation at that value of λ becomes the chosen value of λ_L. Note that our criteria for choosing λ_L can lead to different values chosen for the real and imaginary parts. The resulting choices for z_L are shown in Table III.

3. Fourier transform

Next, we obtain the quasi-PDF from the Fourier transform of the renormalized matrix elements

$$\tilde{q}(y, z_s, P_z, \mu) = \int \frac{dz P_z}{2\pi} e^{iy P_z z} h_{\gamma_t}^R(z, z_s, P_z, \mu). \quad (33)$$

TABLE III. The values of z_L chosen for the real (z_L^{re}) and imaginary (z_L^{im}) parts of the hybrid renormalized matrix elements for each $P_z = \frac{2\pi}{L} n_z$ and μ considered in this work.

n_z	μ (GeV)	z_L^{re}/a	z_L^{im}/a
1	1.4	16	16
	2.0	17	16
	2.8	17	16
4	1.4	12	13
	2.0	12	14
	2.8	12	14
6	1.4	8	11
	2.0	9	12
	2.8	9	12

To perform this integral, we exploit the symmetry property $h_{\gamma_t}^R(-z, z_s, P_z, \mu) = h_{\gamma_t}^R(z, z_s, P_z, \mu)^*$. Expanding the integrand of Eq. (33), we find that

$$
\begin{aligned}
\mathrm{Re}\, e^{iyP_z z} h_{\gamma_t}^R(z, z_s, P_z, \mu) &= \cos(yP_z z)\mathrm{Re}h_{\gamma_t}^R(z, z_s, P_z, \mu) \\
&\quad - \sin(yP_z z)\mathrm{Im}h_{\gamma_t}^R(z, z_s, P_z, \mu), \\
\mathrm{Im}\, e^{iyP_z z} h_{\gamma_t}^R(z, z_s, P_z, \mu) &= \sin(yP_z z)\mathrm{Re}h_{\gamma_t}^R(z, z_s, P_z, \mu) \\
&\quad + \cos(yP_z z)\mathrm{Im}h_{\gamma_t}^R(z, z_s, P_z, \mu).
\end{aligned}
\tag{34}
$$

Thus, the real and imaginary parts of the integrand in Eq. (33) are symmetric and antisymmetric, respectively. Therefore, the imaginary part vanishes under integration. After simplifying the integral, we divide it into a part that sums over the data up to some maximum z_L and a part that performs the integral analytically from z_L to infinity using the extrapolated fit function

$$
\begin{aligned}
&\tilde{q}(y, z_s, P_z, \mu) \\
&= \left[\sum_{z=0}^{z_L^{\mathrm{re}}/a} \frac{z_L^{\mathrm{re}} P_z}{\pi N_{z_L}^{\mathrm{re}}} + \int_{z_L^{\mathrm{re}}}^{\infty} \frac{dz P_z}{\pi} \right] \mathrm{Re}h_{\gamma_t}^R(z, z_s, P_z, \mu) \cos(zP_z y) \\
&\quad - \left[\sum_{z=0}^{z_L^{\mathrm{im}}/a} \frac{z_L^{\mathrm{im}} P_z}{\pi N_{z_L}^{\mathrm{im}}} + \int_{z_L^{\mathrm{im}}}^{\infty} \frac{dz P_z}{\pi} \right] \mathrm{Im}h_{\gamma_t}^R(z, z_s, P_z, \mu) \sin(zP_z y),
\end{aligned}
\tag{35}
$$

where z_L^{re} and z_L^{im} correspond to the values of z_L for the real and imaginary parts, respectively, and $N^{\mathrm{re/im}} \equiv z_L^{\mathrm{re/im}}/a + 1$.

4. Light-cone matching

The final step of the analysis is to match our momentum space quasi-PDF to the light cone. The light-cone PDF $q(x, \mu)$ is related to the quasi-PDF $\tilde{q}(y, z_s, P_z, \mu)$ via

$$
\begin{aligned}
q(x, \mu) &= \int_{-\infty}^{\infty} \frac{dy}{|y|} C^{-1}\left(\frac{x}{y}, \frac{\mu}{yP_z}, |y|\lambda_s\right) \tilde{q}(y, z_s, P_z, \mu) \\
&\quad + \mathcal{O}\left(\frac{\Lambda_{\mathrm{QCD}}^2}{x^2 P_z^2}, \frac{\Lambda_{\mathrm{QCD}}^2}{(1-x)^2 P_z^2}\right) \\
&\equiv C^{-1}\left(\frac{x}{y}, \frac{\mu}{yP_z}, |y|\lambda_s\right) \otimes \tilde{q}(y, z_s, P_z, \mu) \\
&\quad + \mathcal{O}\left(\frac{\Lambda_{\mathrm{QCD}}^2}{x^2 P_z^2}, \frac{\Lambda_{\mathrm{QCD}}^2}{(1-x)^2 P_z^2}\right),
\end{aligned}
\tag{36}
$$

where $C^{-1}\left(\frac{x}{y}, \frac{\mu}{yP_z}, |y|\lambda_s\right)$ is the inverse matching kernel and we use the notation "$\otimes$" as a short-hand for the integral. We can write the full matching kernel as a series expansion in the strong coupling α_s

$$
C\left(\frac{x}{y}, \frac{\mu}{yP_z}, |y|\lambda_s\right) = \delta\left(\frac{x}{y} - 1\right) + \sum_{n=1}^{\infty} \alpha_s^n C^{(n)}\left(\frac{x}{y}, \frac{\mu}{yP_z}, |y|\lambda_s\right)
\tag{37}
$$

with its inverse being defined as

$$
C^{-1}\left(\frac{x}{z}, \frac{\mu}{zP_z}, |z|\lambda_s\right) \otimes C\left(\frac{z}{y}, \frac{\mu}{yP_z}, |y|\lambda_s\right) = \delta\left(\frac{x}{y} - 1\right),
\tag{38}
$$

where $\delta(\frac{x}{y} - 1)$ is the Dirac delta function. We can obtain a series solution for C^{-1} by combining Eqs. (37) and (38)

$$
\begin{aligned}
C^{-1}\left(\frac{x}{y}, \frac{\mu}{yP_z}, |y|\lambda_s\right) &= \delta\left(\frac{x}{y} - 1\right) - \alpha_s C^{(1)}\left(\frac{x}{y}, \frac{\mu}{yP_z}, |y|\lambda_s\right) \\
&\quad + \alpha_s^2 C^{(1)}\left(\frac{x}{z}, \frac{\mu}{zP_z}, |z|\lambda_s\right) \\
&\quad \otimes C^{(1)}\left(\frac{z}{y}, \frac{\mu}{yP_z}, |y|\lambda_s\right) \\
&\quad - \alpha_s^2 C^{(2)}\left(\frac{x}{y}, \frac{\mu}{yP_z}, |y|\lambda_s\right) + \mathcal{O}(\alpha_s^3).
\end{aligned}
\tag{39}
$$

As demonstrated in Appendix C. 1 of [35], the replacement of the full integral in Eq. (36) by matrix multiplication reduces the computational cost of light-cone matching with negligible loss of accuracy. We thus form a matching matrix C_{xy} with the x and y indices corresponding to those in Eq. (37).

Treating the quasi-PDF as the LO approximation in LaMET, we construct two matrices, C_{xy}^{NLO} and C_{xy}^{NNLO}, to achieve NLO and NNLO results, respectively, for the light-cone PDF

$$
q_x^{\mathrm{LO}} = \tilde{q}_x,
\tag{40}
$$

$$
q_x^{\mathrm{NLO}} = \tilde{q}_x - \delta y \sum_y C_{xy}^{\mathrm{NLO}} \tilde{q}_y,
\tag{41}
$$

$$
q_x^{\mathrm{NNLO}} = \tilde{q}_x - \delta y \sum_y C_{xy}^{\mathrm{NLO}} \tilde{q}_y - \delta y \sum_y C_{xy}^{\mathrm{NNLO}} \tilde{q}_y,
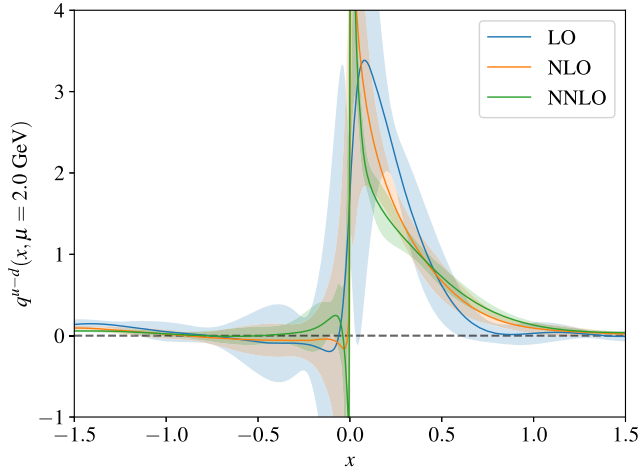\tag{42}
$$

FIG. 15. Dependence of the x-space matched isovector-quark PDF on the perturbative order used in the matching kernel. The results shown use the largest value of momentum computed in this work (i.e. $P_z = 1.53$ GeV).
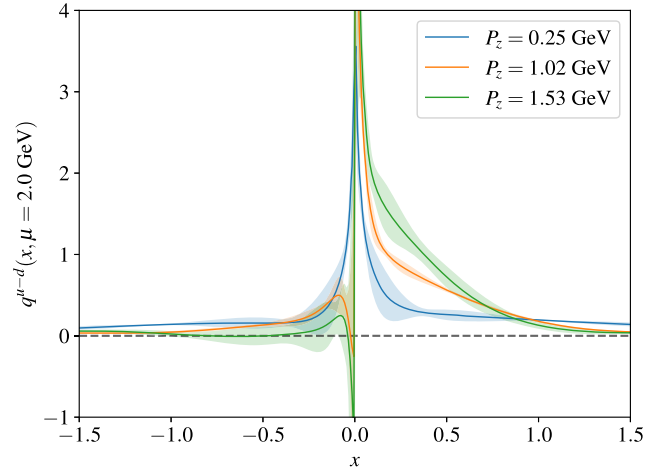

FIG. 16. Dependence of the x-space matched isovector-quark PDF on the momentum of the nucleon. The results shown use the matching kernel computed at NNLO.

where $\delta y = 0.001$, corresponding to a discretization of the integral in Eq. (36).

The corrections to the matching in Eq. (36) mean that our LaMET calculation breaks down as $x \to 0$ and $x \to 1$. Our range of validity is explained in Sec. VI B 1.

B. Results

As a first attempt to understand some of the systematics involved with the LaMET approach, we start by studying the perturbative matching order dependence with our largest value of $P_z = 1.53$ GeV, which is shown in Fig. 15. There does appear to be convergence going from NLO to NNLO, at least in the middle-x region where LaMET is expected to hold. Additionally, we also show the dependence on momentum using the NNLO matching in Fig. 16. There is a significant dependence on the momentum for most values of x, which indicates that the momenta are still not large enough to sufficiently suppress the power corrections. Therefore, we present the PDF obtained at $P_z = 1.53$ GeV as our final result.

1. Valid range of x

The LaMET approach is only valid in the middle ranges of $|x| \in [x_{\min}, x_{\max}]$, which arises from the power corrections in Eq. (36). The power counting in Eq. (36) is based on the argument that the active and spectator partons must carry hard momentum, and that power corrections in the $\overline{\text{MS}}$ scheme usually begin at quadratic order. In QCD, they are closely related to the renormalon ambiguities in the leading-twist coefficient functions [74–76]. In Ref. [74], the power corrections in the quasi-PDF were estimated by a renormalon analysis, and it was concluded that they behave as $\mathcal{O}(\Lambda_{\text{QCD}}^2/(x^2(1-x)P_z^2))$, whereas an independent analysis led to $\mathcal{O}(\Lambda_{\text{QCD}}^2/(x^2 P_z^2))$ [75]. On the other

hand, the large infared logarithms in the end-point regions, i.e., the DGLAP logarithms $\ln\frac{\mu^2}{4x^2 P_z^2}$ at small x and threshold logarithms $\ln\frac{\mu^2}{4(1-x)P_z^2}$ at large x [50,51], also indicate that nonperturbative effects become important when $2xP^z$ and $2\sqrt{1-x}P^z$ become close to Λ_{QCD}.

As our statistical precision is not adequate to properly assess all systematics rigorously, we determine the values of $x_{\min}$ and $x_{\max}$ by requiring that $x_{\min}P^z \sim \Lambda_{\text{QCD}}$ and $\sqrt{1-x_{\max}}P^z \sim \Lambda_{\text{QCD}}$. In our matching coefficient, the strong coupling α_s is defined in the $\overline{\text{MS}}$ scheme with $\Lambda_{\text{QCD}}^{\overline{\text{MS}}} = 332$ MeV [77]. Therefore, using our largest value of $P_z \approx 1.53$ GeV, we find $x_{\min} \approx 0.217$ and $x_{\max} \approx 0.953$.

In Fig. 17, we show a comparison of our final estimate at $\mu = 2.0$ GeV from x-space matching with $P_z = 1.53$ GeV,
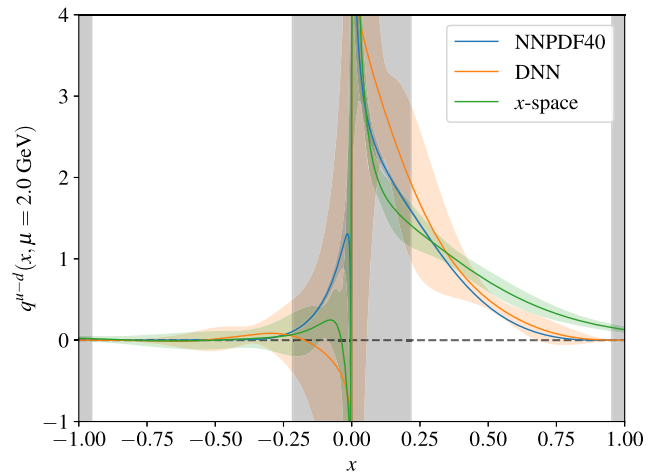

FIG. 17. Comparison of the x-dependence of the isovector-quark PDF from the global analysis of NNPDF4.0, the DNN, and x-space matching. The gray bands correspond to the regions of x where we do not rigorously trust the results of LaMET.

the DNN with $z_{max} = 0.76$ fm and $\delta = 0.2$, and the global analysis from NNPDF4.0 [4]. Our x-space matching results in the region $0.5 < x < 0.953$ still have noticeable differences from the NNPDF4.0 analysis, as we also observe in Fig. 15 that the NLO and NNLO corrections are significant compared to the LO result. This indicates that the systematic uncertainties, most likely the unresummed large logarithms at $x \sim 1$ [50], the P^z-dependent power corrections, and the lattice discretization effects are not well under control. Achieving higher momentum and taking the continuum limit will be necessary to improve the calculation.

As one final check on the systematics, we would like to understand any relationship between the discrepancy observed for $\langle x \rangle$ and that observed for the PDF. Therefore, we use the moments determined from the NNPDF4.0 global analysis to reconstruct the reduced pseudo-ITD using the ratio of OPEs in Eq. (16). Next, when forming the hybrid-renormalized matrix elements in Eq. (31), we use this reconstructed reduced pseudo-ITD for $z \leq z_S$, rather than using the lattice data itself. Finally, we proceed in the same manner to extract the PDF with this newly constructed hybrid-renormalized matrix elements. If the origin of the discrepancy we see in $\langle x \rangle$ was the same as that for the PDF itself, this new PDF should agree better with the PDF from the NNPDF4.0 analysis. However, while we do see a shift in the extracted PDF towards the results from NNPDF4.0 for large x, the difference is less than 1σ and, therefore, cannot fully explain the discrepancy we see in the PDFs.

VII. CONCLUSIONS

In this paper, we have extracted information on the unpolarized isovector quark PDF of the proton from lattice QCD using various methods. This is the first such work to utilize an NNLO matching coefficient in a calculation performed with physical quark masses directly. The significance of the NNLO term is demonstrated in Figs. 7 and 15 for the Mellin moments and the PDF, respectively. The difference between matching at NLO and NNLO (within the valid x-range, in the case of the PDF) is small but non-negligible. This demonstrates good convergence of the matching process. With a physical pion mass, excited state contamination in the three-point functions is significant, and we therefore had to utilize several fitting methods to ensure that effects from these unwanted states are under control. From the bare matrix elements, we then used the ratio scheme along with the leading-twist OPE approximation to extract the lowest four moments of the PDF. We found the lowest moment $\langle x \rangle$ to lie above the result from the global phenomenological fits performed by NNPDF4.0 [4], which is quite typical in several other lattice calculations. The effect is likely due to unaccounted-for systematics (e.g. discretization effects, higher-twist contributions, etc.).

Next, we determined the x-dependent PDFs utilizing a DNN to solve the inverse problem arising in the pseudo-PDF approach. Although we found agreement with the results from NNPDF4.0, the errors were significantly larger, which suggests that our statistical errors dominated any potential systematic error.

Our final method involved performing the matching directly in momentum space using the hybrid renormalization scheme with LaMET. The results show some tension with both our DNN results and the global results from NNPDF4.0, even in the region of x in which we expect power corrections to be under control. This is not entirely unexpected, as we did not observe a convergence in the PDF results as P_z was increased, suggesting a need for larger P_z. Further, since these calculations were performed with a single ensemble, the size of discretization effects has not been determined. Despite these limitations, the tension we see in our results is not substantial enough to question the validity of the methods and is a promising first step towards a more thorough investigation.

Our future work will include more statistics, more ensembles, and larger momenta, which will allow for control over the remaining unaccounted-for systematics in our current calculations.

a high-performance computing cluster operated by the Laboratory Computing Resource Center at Argonne National Laboratory.

The computation of the correlators was carried out with the QLua software suite [40], which utilized the multigrid solver in QUDA [41,42]. The analysis of the correlation functions was done with the use of lsqfit [78] and gvar [79]. Several of the plots were created with Matplotlib [80].

APPENDIX: χ^2 AND THE DETAILS OF THE DNN TRAINING

In this work, our ratio scheme renormalized matrix elements $\mathcal{M}_{\gamma_t}(\lambda, z^2; P_z^0)$ used $P_z^0 = 0$ which is the standard reduced pseudo-ITD [18,29–34], denoted here as $\tilde{Q}(\lambda, z^2) = \mathcal{M}_{\gamma_t}(\lambda, z^2; P_z^0 = 0)$, which can be related to the PDF by

$$
\begin{aligned}
\tilde{Q}(\lambda, z^2) &= \int_{-1}^{1} d\alpha \int_{-1}^{1} dy \frac{C(\alpha, \mu^2 z^2)}{C_0(\mu^2 z^2)} e^{-iy\alpha\lambda} q(y, \mu) \\
&= \int_{-1}^{1} dy q(y) \frac{\int_{-1}^{1} d\alpha e^{-iy\alpha\lambda} C(\alpha, \mu^2 z^2)}{C_0(\mu^2 z^2)} \\
&= \int_{-1}^{1} dy q(y) \bar{C}(y\lambda, \mu^2 z^2),
\end{aligned} \tag{A1}
$$

where $C_0(\mu^2 z^2)$ is the zeroth-order Wilson coefficient, and

$$
\bar{C}(\lambda, \mu^2 z^2) \equiv \frac{\int_{-1}^{1} d\alpha e^{-i\lambda\alpha} C(\alpha, \mu^2 z^2)}{C_0(\mu^2 z^2)}. \tag{A2}
$$

One then can study the real and imaginary part of $\tilde{Q}(\lambda, z^2)$ separately. Here we only discuss the real part while the imaginary part can be derived by changing $\cos(\lambda\alpha)$ to $\sin(\lambda\alpha)$. Up to NNLO, we express the convolution kernel as

$$
\begin{aligned}
C_{\text{NNLO}}(\lambda, \mu^2 z^2) &\equiv \int_{-1}^{1} d\alpha \cos(\lambda\alpha) C(\alpha, \mu^2 z^2) \\
&= \cos(\lambda) d_{\text{NNLO}}(\mu^2 z^2) \\
&\quad + \int_{0}^{1} d\alpha [\cos(\lambda\alpha) - \cos(\lambda)] \\
&\quad \times n_{\text{NNLO}}(\alpha, \mu^2 z^2),
\end{aligned} \tag{A3}
$$

with

$$
C_{\text{NNLO},0}(\mu^2 z^2) \equiv C_{\text{NNLO}}(\lambda = 0, \mu^2 z^2) = d_{\text{NNLO}}(\mu^2 z^2), \tag{A4}
$$

where

$$
d_{\text{NNLO}}(\mu^2 z^2) = 1 + \sum_{i=1}^{2} \sum_{j=0}^{i} \left(\frac{\alpha_s}{2\pi}\right)^i d_{i,j} [L(\mu^2 z^2)]^j, \tag{A5}
$$

$$
n_{\text{NNLO}}(\alpha, \mu^2 z^2) = \sum_{i=1}^{2} \sum_{j=0}^{i} \left(\frac{\alpha_s}{2\pi}\right)^i n_{i,j}(\alpha) [L(\mu^2 z^2)]^j, \tag{A6}
$$

where

$$
L(\mu^2 z^2) \equiv 2\gamma_E + \ln\frac{\mu^2 z^2}{4}, \tag{A7}
$$

$$
d_{1,0} = \frac{5 C_F}{2}, \tag{A8}
$$

$$
d_{1,1} = \frac{3 C_F}{2}, \tag{A9}
$$

$$
d_{2,0} = C_F \left[\frac{469\beta_0}{48} + \frac{223 C_F - 94 C_A}{96} \right. \\
\left. + \frac{\pi^2(8 C_F - 15 C_A)}{36} + 2(C_A - 4 C_F)\zeta(3) \right], \tag{A10}
$$

$$
d_{2,1} = \left[C_F^2 \left(-\frac{5}{8} + \frac{2\pi^2}{3} \right) + C_F C_A \left(\frac{49}{24} - \frac{\pi^2}{6} \right) - \frac{5}{6} C_F n_f T_F \right] \\
+ \frac{5 C_F}{2} \left(\frac{3}{2} C_F + \beta_0 \right), \tag{A11}
$$

$$
d_{2,2} = \frac{3 C_F}{4} \left(\frac{3}{2} C_F + \beta_0 \right), \tag{A12}
$$

$$
n_{1,0}(\alpha) = C_F \left[2(1-\alpha) - \frac{1+\alpha^2}{1-\alpha} - \frac{4\ln(1-\alpha)}{1-\alpha} \right], \tag{A13}
$$

$$
n_{1,1}(\alpha) = -C_F \frac{1+\alpha^2}{1-\alpha}, \tag{A14}
$$

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 $\beta_0 = (11 C_A - 4 n_f T_F)/6$. Then we denote, and numerically compute, that

$$
c_{i,j}(\lambda) \equiv \int_{0}^{1} d\alpha [\cos(\lambda\alpha) - \cos(\lambda)] n_{i,j}(\alpha), \tag{A15}
$$

which gives,

$$
\begin{aligned}
&\bar{C}_{\text{NNLO}}(\lambda, \mu^2 z^2) \\
&= \cos(\lambda) + \frac{\int_{0}^{1} d\alpha [\cos(\lambda\alpha) - \cos(\lambda)] n_{\text{NNLO}}(\alpha, \mu^2 z^2)}{d_{\text{NNLO}}(\mu^2 z^2)} \\
&= \cos(\lambda) + \frac{\sum_{i=1}^{2} \sum_{j=0}^{i} (\alpha_s/2\pi)^i c_{i,j}(\lambda) L^j(\mu^2 z^2)}{1 + \sum_{i=1}^{2} \sum_{j=0}^{i} (\alpha_s/2\pi)^i d_{i,j} L^j(\mu^2 z^2)}. \tag{A16}
\end{aligned}
$$

In this work, we express the PDF $q^-(x)$ by DNN together with extra factors,

$$
q_{\text{DNN}}^-(y) \equiv N y^{\alpha} (1-y)^{\beta} [1 + \delta \sin(f_{\text{DNN}}^-(y, \boldsymbol{\theta}^-))], \tag{A17}
$$

with δ being some number that limits the contribution of subleading terms. The normalization condition yields

$$N(\alpha^-, \beta^-, \boldsymbol{\theta}^-)$$
$$= \left\{ 2 \int_0^1 dy y^{\alpha^-} (1-y)^{\beta^-} [1 + \delta \sin(f^-_{\text{DNN}}(y, \boldsymbol{\theta}^-))] \right\}^{-1}. \quad (A18)$$

The loss function is defined as

$$J(\alpha^-, \beta^-, \boldsymbol{\theta}^-) \equiv \frac{\eta}{2} \boldsymbol{\theta}^- \cdot \boldsymbol{\theta}^- + \frac{1}{2} \chi^2(\alpha^-, \beta^-, \boldsymbol{\theta}^-), \quad (A19)$$

where

$$\chi^2(\alpha^-, \beta^-, \boldsymbol{\theta}^-) \equiv \sum_{i,j} [\text{Cov}^{-1}]_{ij} [\tilde{Q}_{\text{DNN}}(P_i, z_i) - \tilde{Q}(P_i, z_i)]$$
$$\times [\tilde{Q}_{\text{DNN}}(P_j, z_j) - \tilde{Q}(P_j, z_j)]. \quad (A20)$$

[1] S. Alekhin, J. Blümlein, S. Moch, and R. Placakyte, Phys. Rev. D **96**, 014011 (2017).
[2] T.-J. Hou *et al.*, Phys. Rev. D **103**, 014013 (2021).
[3] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, Eur. Phys. J. C **81**, 341 (2021).
[4] R. D. Ball *et al.* (NNPDF Collaboration), Eur. Phys. J. C **82**, 428 (2022).
[5] J. Dudek *et al.*, Eur. Phys. J. A **48**, 187 (2012).
[6] A. Accardi *et al.*, Eur. Phys. J. A **52**, 268 (2016).
[7] R. Abdul Khalek *et al.*, Nucl. Phys. **A1026**, 122447 (2022).
[8] K. Cichy and M. Constantinou, Adv. High Energy Phys. **2019**, 3036904 (2019).
[9] Y. Zhao, Int. J. Mod. Phys. A **33**, 1830033 (2019).
[10] A. V. Radyushkin, Int. J. Mod. Phys. A **35**, 2030002 (2020).
[11] X. Ji, Y.-S. Liu, Y. Liu, J.-H. Zhang, and Y. Zhao, Rev. Mod. Phys. **93**, 035005 (2021).
[12] M. Constantinou, Eur. Phys. J. A **57**, 77 (2021).
[13] M. Constantinou *et al.*, Prog. Part. Nucl. Phys. **121**, 103908 (2021).
[14] K. Cichy, Proc. Sci., LATTICE2021 (**2022**) 017.
[15] X. Ji, Phys. Rev. Lett. **110**, 262002 (2013).
[16] X. Ji, Sci. China Phys. Mech. Astron. **57**, 1407 (2014).
[17] A. V. Radyushkin, Phys. Rev. D **96**, 034025 (2017).
[18] K. Orginos, A. Radyushkin, J. Karpie, and S. Zafeiropoulos, Phys. Rev. D **96**, 094503 (2017).
[19] V. Braun and D. Müller, Eur. Phys. J. C **55**, 349 (2008).
[20] Y.-Q. Ma and J.-W. Qiu, Phys. Rev. Lett. **120**, 022003 (2018).
[21] X. Ji, arXiv:2209.09332.
[22] C. Alexandrou, K. Cichy, M. Constantinou, K. Jansen, A. Scapellato, and F. Steffens, Phys. Rev. Lett. **121**, 112001 (2018).
[23] C. Alexandrou, K. Cichy, M. Constantinou, K. Hadjiyiannakou, K. Jansen, A. Scapellato, and F. Steffens, Phys. Rev. D **99**, 114504 (2019).
[24] J.-W. Chen, L. Jin, H.-W. Lin, Y.-S. Liu, Y.-B. Yang, J.-H. Zhang, and Y. Zhao, arXiv:1803.04393.
[25] Y.-S. Liu *et al.* (Lattice Parton Collaboration), Phys. Rev. D **101**, 034020 (2020).
[26] 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).
[27] 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).
[28] H.-W. Lin, J.-W. Chen, and R. Zhang, arXiv:2011.14971.
[29] B. Joó, J. Karpie, K. Orginos, A. Radyushkin, D. Richards, and S. Zafeiropoulos, J. High Energy Phys. 12 (2019) 081.
[30] B. Joó, J. Karpie, K. Orginos, A. V. Radyushkin, D. G. Richards, and S. Zafeiropoulos, Phys. Rev. Lett. **125**, 232003 (2020).
[31] M. Bhat, K. Cichy, M. Constantinou, and A. Scapellato, Phys. Rev. D **103**, 034510 (2021).
[32] J. Karpie, K. Orginos, A. Radyushkin, and S. Zafeiropoulos (HadStruc Collaboration), J. High Energy Phys. 11 (2021) 024.
[33] 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.
[34] M. Bhat, W. Chomicki, K. Cichy, M. Constantinou, J. R. Green, and A. Scapellato, Phys. Rev. D **106**, 054504 (2022).
[35] X. Gao, A. D. Hanlon, S. Mukherjee, P. Petreczky, P. Scior, S. Syritsyn, and Y. Zhao, Phys. Rev. Lett. **128**, 142003 (2022).
[36] X. Gao, A. D. Hanlon, N. Karthik, S. Mukherjee, P. Petreczky, P. Scior, S. Shi, S. Syritsyn, Y. Zhao, and K. Zhou, Phys. Rev. D **106**, 114510 (2022).
[37] 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).
[38] A. Bazavov *et al.*, Phys. Rev. D **100**, 094510 (2019).
[39] A. Hasenfratz and F. Knechtli, Phys. Rev. D **64**, 034504 (2001).
[40] A. Pochinsky, QLua lattice software suite, https://usqcd.lns.mit.edu/qlua (2008–present).
[41] M. A. Clark, R. Babich, K. Barros, R. C. Brower, and C. Rebbi, Comput. Phys. Commun. **181**, 1517 (2010).
[42] R. Babich, M. A. Clark, B. Joo, G. Shi, R. C. Brower, and S. Gottlieb, in *SC11 International Conference for High Performance Computing, Networking, Storage and Analysis* (ACM, 2011), arXiv:1109.2935.

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

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

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

[46] J.-W. Chen, T. Ishikawa, L. Jin, H.-W. Lin, J.-H. Zhang, and Y. Zhao (LP3 Collaboration), Chin. Phys. C **43**, 103101 (2019).

[47] M. Sun *et al.* (xQCD Collaboration), Phys. Rev. D **101**, 054511 (2020).

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

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

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

[51] Y. Su, J. Holligan, X. Ji, F. Yao, J.-H. Zhang, and R. Zhang, arXiv:2209.01236.

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

[53] G. S. Bali, S. Collins, B. Gläßle, M. Göckeler, J. Najjar, R. H. Rödl, A. Schäfer, R. W. Schiel, A. Sternbeck, and W. Söldner, Phys. Rev. D **90**, 074510 (2014).

[54] A. Abdel-Rehim *et al.*, Phys. Rev. D **92**, 114513 (2015); **93**, 039904(E) (2016).

[55] C. Alexandrou, M. Constantinou, K. Hadjiyiannakou, K. Jansen, C. Kallidonis, G. Koutsou, A. Vaquero Avilés-Casco, and C. Wiese, Phys. Rev. Lett. **119**, 142002 (2017).

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

[57] J. R. Green, M. Engelhardt, S. Krieg, J. W. Negele, A. V. Pochinsky, and S. N. Syritsyn, Phys. Lett. B **734**, 290 (2014).

[58] K. Ottnad, D. Djukanovic, T. Harris, H. B. Meyer, G. von Hippel, and H. Wittig, Proc. Sci. LATTICE2021 (**2022**) 343.

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

[60] 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).

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

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

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

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

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

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

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

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

[69] A. Bazavov *et al.* (HotQCD Collaboration), Phys. Rev. D **90**, 094503 (2014).

[70] A. Bazavov, N. Brambilla, H. T. Ding, P. Petreczky, H. P. Schadler, A. Vairo, and J. H. Weber, Phys. Rev. D **93**, 114502 (2016).

[71] A. Bazavov, P. Petreczky, and J. H. Weber, Phys. Rev. D **97**, 014510 (2018).

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

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

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

[75] W.-Y. Liu and J.-W. Chen, Phys. Rev. D **104**, 094501 (2021).

[76] R. Zhang, J. Holligan, X. Ji, and Yushan (to be published).

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

[78] P. Lepage, gplepage/lsqfit: lsqfit version 12.0.3 (2022), https://github.com/gplepage/lsqfit.

[79] P. Lepage, gplepage/gvar: gvar version 11.9.5 (2022), https://github.com/gplepage/gvar.

[80] J. D. Hunter, Comput. Sci. Eng. **9**, 90 (2007).