Parton physics of the large-N_c mesons

Nikhil Karthik[1,2,*] and Rajamani Narayanan[3,†]

[1]*Department of Physics, William and Mary, Williamsburg, Virginia 23185, USA*
[2]*Thomas Jefferson National Accelerator Facility, Newport News, Virginia 23606, USA*
[3]*Department of Physics, Florida International University, Miami, Florida 33199, USA*

(Received 26 May 2022; accepted 22 June 2022; published 7 July 2022)

We initiate the studies on the structural physics of the tower of stable large-N_c mesons through a first computation of the collinear quark structure of a large-N_c pion using lattice Monte Carlo methods. We adapt the large-N_c continuum reduction for the determination of meson correlation functions involving the spatially extended quasiparton distribution function operators as a perfect strategy to concentrate only on the short perturbative length scales. We find the internal structures of a pion in the large-N_c and $N_c = 3$ theories to be quite similar. Interestingly, we find hints that even the observed differences could arise to a large extent via the different perturbative QCD evolution in the two theories from similar initial conditions at low-factorization scales.

DOI: 10.1103/PhysRevD.106.014503

I. INTRODUCTION

Quantum chromodynamics (QCD) in the limit of large number of colors, N_c, at a fixed 't Hooft coupling [1,2] $\lambda = N_c\alpha_s$, is greatly simplified by being a planar model in which quarks are naturally quenched, and it is well known to be a realistic QCD-like theory that approximately reproduces many features in the real world, such as the ratios of low-lying meson masses [3–8]. The next frontier in QCD physics is to understand the structural aspects of hadrons in more detail, so as to relate the emergent properties of hadrons, such as their masses and spins, to those of the short-distance quark-gluon (parton) degrees of freedom and their interactions (e.g., see [9–11]). In this respect, the large-N_c limit motivates and crystallizes concepts in parton phenomenology, such as the linear Regge trajectories (proven in two dimensions [2]), the dipole approach to Balitsky-Fadin-Kuraev-Lipatov formalism [12], and the concept of quark-hadron duality [13,14] to name a few.

The large-N_c baryons [15] are $\mathcal{O}(N_c)$ heavier degrees of freedom that can be described as a chiral soliton [16–19]. Such an identification has lead to mean-field theory studies of the parton distributions inside a nucleon (for initial works, see [20–23]). In contrast, the large-N_c mesons are the leading lighter degrees of freedom, and nonperturbative methods (e.g., lattice simulations) are the only way to study them. With the access to an infinite tower of completely stable large-N_c mesons of different J^{PC}, it is an ideal realization of QCD that is conducive to investigate the partonic origin of hadron physics. Development of such realistic models of mesons as hard scatterers is especially important due to the reinvigorated experimental [11,24,25] and theoretical efforts towards the meson structures, especially of the pion, the Goldstone mode of chiral symmetry breaking (refer to [26] for a review, and [27–36] for recent numerical works). Quite surprisingly, despite the continued effort to understand large-N_c QCD over the years, the partonic nature of the large-N_c mesons is to a large extent unknown. To our knowledge, the study in Ref. [37] of the distribution amplitude of a pion within a large-N_c Regge model is a singular work toward this direction.

Through the present work, we bridge this persisting gap in our understanding of the canonical toy model of QCD through a first computation of quark distribution function of the large-N_c pion, and thereby, lay the framework for comparative studies of internal structures of different stable species of mesons. As an important feature of the large-N_c theory, we present the large-N_c continuum reduction as a novel tailor-made approach for the operator product expansion (OPE) based strategies [38–41] to perform parton physics on the lattice. We display the schematic of the central idea of the calculation in Fig. 1(a), and we elaborate on it in the following discussion. As an initial work in this direction, we keep the discussion simple by summarizing the main techniques and results in the main text, and by referring the reader to various appendixes for the elaborate details.

*nkarthik.work@gmail.com
†rajamani.narayanan@fiu.edu

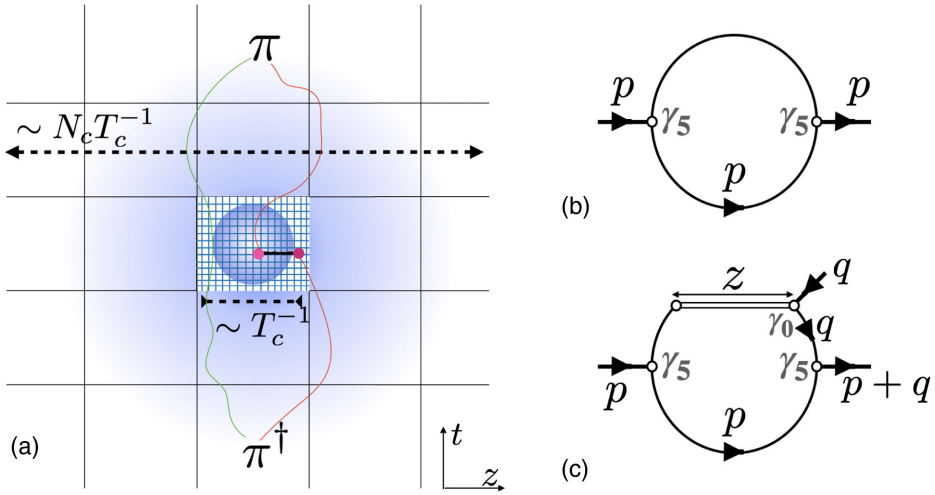

FIG. 1. (a) Schematic of large-N_c continuum reduction for quasi-PDF operator evaluated within a pion. The gauge fields on the $\approx N_c \ell$ sized box are obtained as replicas of gauge fields within a ℓ sized box, with $\ell \approx T_c^{-1}$ the deconfinement temperature. The quarks hopping on such crystalline configuration are labeled by their positions in a periodic ℓ^4 box and their Bloch momenta. The correlation functions in the larger box can be obtained using lattice implementation of momentum space Feynman diagrams that use quark propagators in the ℓ^4 box. (b),(c) The momentum space Feynman diagrams implemented directly on the lattice. The lines are quark propagators. The arrows show the off-shell 4-momentum injected at the vertices. The two-point function of a pion is shown in (b). The three-point function of the quasi-PDF operator (double line) with pion creation and annihilation operators is shown in (c).

II. BASICS OF CONTINUUM REDUCTION

Owing to the absence of a center symmetry breaking deconfinement phase transition in two Euclidean space-time dimensions, the large-N_c QCD_2 is well known to be reducible to a single-site matrix model [42,43]. We can extend the Eguchi-Kawai reduction [43] to dimensions $d > 2$ [44–48], provided we preserve the $U^d(1)$ center symmetry by reducing the theory not to a point, but instead to a small box of volume ℓ^d, with $\ell \geq T_c^{-1}$ the inverse of deconfinement temperature. The powerful aspect of the large-N_c reduction is that we can exactly find the expectation values of gauge-invariant quantities (such as a $w_1 \times w_2$ Wilson loop) in $\mathbb{R}^d$ just from the expectation values of the same quantity on the reduced ℓ^d periodic torus (even if $w_1, w_2 > \ell$) through folding. We can regulate the reduced continuum theory on an L^d periodic lattice using a lattice coupling $b = (g^2 N_c)^{-1}$ in the limit $N_c \to \infty$ at fixed b, using L greater than a critical $L_c(b)$. The asymptotic scaling of $L_c(b)$ defined the critical size, $\ell_c = T_c^{-1}$ [48]. As a corollary, we can unfold the torus by tessellating $\mathbb{R}^d$ with the gauge configuration in ℓ^d box, resulting in a path integral over crystalline configurations. Consequently, the quarks are labeled by position $x \in \ell^d$ and the Bloch momentum q. Thereupon, we can write functions F of a lattice Dirac operator $\not{D}$ in $\mathbb{R}^d$, such as its propagator G, in terms of functions F^L of Dirac operator $\not{D}^L$ in L^d periodic lattice as

$$F_{x,y}(U_\mu) = \int \frac{d^d q}{(2\pi)^d} e^{i\frac{q\cdot(x-y)}{L}} F^L_{x,y}(U_\mu e^{i\frac{q_\mu}{L}}). \tag{1}$$

We discuss the details behind such a construction in Appendix A 1. By using such a relation, along with the global $U^d(1)$ center symmetry, we can reduce all n-point functions of quark bilinears in $\mathbb{R}^d$ to computations of n-point functions on L^d periodic lattice. For example, as derived in Appendix A 2, we can write the two-point function, $\tilde{C}_{2\text{pt}}(p) = \langle \pi(p)\pi^\dagger(p)\rangle$, for a pion ($\pi = \bar{d}\gamma_5 u$) in momentum space as

$$\tilde{C}_{2\text{pt}}(p) = \langle \text{Tr}[\gamma_5 G^L(U_\mu e^{-ip_\mu})\gamma_5 G^L(U_\mu)]\rangle, \tag{2}$$

where $p = (p_0, \mathbf{p})$ is the continuous-valued Euclidean four momentum of the pion, the trace is over spin, color, and the entire L^d lattice, and the ensemble average $\langle \ldots \rangle$ is with respect to the pure gauge action. We show the Feynman diagram for Eq. (2) in Fig. 1(b). From the spectral decomposition, $\tilde{C}_{2\text{pt}}(p_0, \mathbf{p}) = \sum_{i=0} 2A_i E_i(\mathbf{p})(p_0^2 + E_i^2(\mathbf{p}))^{-1}$, we can obtain the long-distance energy spectrum, $E_i(\mathbf{p})$, and amplitudes, A_i. Alternatively, we can access the spectrum from the multi-exponential, $A_i e^{-E_i(\mathbf{p})t_s}$, decay of $C_{2\text{pt}}(t_s; \mathbf{p}) = \int \frac{dp_0}{2\pi} \tilde{C}_{2\text{pt}}(p_0, \mathbf{p}) e^{ip_0 t_s}$, in the Euclidean time t_s. We see that the long-distance hadronic spectral physics is trivialized by the ability to capture $|\mathbf{x}|, t_s \gtrsim \Lambda_{\text{QCD}}^{-1}$ using only simulation of a box of size $\ell \approx \Lambda_{\text{QCD}}^{-1}$.

III. ZOOMING IN ON PARTON SCALES WITH CONTINUUM REDUCTION

This large-N_c continuum reduction leads to a key simplification in lattice QCD computations of parton

distributions. Many recent developments [38–41] in the *ab initio* computations of the Bjorken x-dependent parton distribution functions (PDFs), $f(x,\mu)$ at a $\overline{\text{MS}}$ factorization scale μ, and related quantities, rely on the leading-twist expansion of certain equal-time renormalized invariant amplitudes, $\mathcal{M}(\nu,z^2)$ with $\nu=-z\cdot P$, involving an operator pair [38,41] or a bilocal extended operator [39,40] with a spatial separation $z_\mu=z_3\delta_{\mu,3}$ that is evaluated within a state $|P\rangle$ of an on-shell hadron moving with momentum $P=(E(\mathbf{p}),\mathbf{p})$. Through lattice Monte Carlo determination of $\mathcal{M}$, we can relate it to $f(x,\mu)$ through an OPE truncated at leading-twist terms (see [49]),

$$\mathcal{M}(\nu,z^2)=\sum_{n=0}\frac{(i\nu)^n}{n!}C_n(\mu^2z^2)\int_{-1}^{1}x^nf(x,\mu)dx. \quad (3)$$

In the absence of higher-twist corrections, the Wilson coefficients C_n capture $\ln(-z^2\mu^2)$-type QCD contributions to $\mathcal{M}$ using perturbation theory and leads to $f(x,\mu)$ at a chosen scale μ. Thus, along with the necessity of nonzero P_3, the short-distance $|z|$ is crucial for the validities of the OPE, the perturbation theory, and for ignoring higher-twist terms. On the other hand, the leading-twist expansion is performed within hadronic in and out states, and therefore, having control of the long-distance aspects of QCD is equally important. Applying the above formalism to the large-N_c theory is much simpler—we can capture the long-distance hadronic states easily by the virtue of continuum reduction, leaving only the relevant partonic scales for z below the inverse deconfinement transition temperature, T_c^{-1}, to be captured by Monte Carlo sampling of gauge fields within T_c^{-1} extent.

With this realization, we extend the continuum reduction approach to n-point functions involving an extended operator, such as the u-quark quasi-PDF operator $\mathcal{O}(z;q)\equiv\sum_x e^{iq\cdot x}\bar{u}_x\gamma_0W_{x,x+z}u_{x+z}$, for purely spatial $z=(0,0,0,z_3)$, and $W_{x,x+z}$ is a straight Wilson line connecting x to $x+z$. The spatial part is $\mathbf{q}=0$ for the PDF we want to study. Following our discussion of the twp-point function and the method of folding Wilson loops of any size on an L^4 lattice, we can similarly write the three-point function $\tilde{C}_{3\text{pt}}(z,p,q)\equiv\langle\pi(p+q)\mathcal{O}(z;q)\pi^\dagger(p)\rangle$, as

$$\tilde{C}_{3\text{pt}}=\sum_x\langle\text{tr}([\gamma_0W^L]_{x,x+z}[G^L\gamma_5G^L\gamma_5G^L]_{x+z,x})\rangle, \quad (4)$$

where the gauge links U_μ entering the propagators from left to right are multiplied by phases 1, e^{-ip_μ}, and e^{iq_μ} respectively. The trace is over color and spin, and W^L is the folded Wilson line obtained by wrapping around the periodic lattice if $|z_3|\geq L$. In Appendix A 3, we present a detailed derivation of the above equation. We show the Feynman diagram for Eq. (4) in Fig. 1(c). Note that the quark line disconnected piece in $\tilde{C}_{3\text{pt}}$ is N_c^{-1} suppressed, and therefore as another large-N_c advantage, we have ignored it in the above equation. As in the two-point function, we can obtain the required bare quasi-PDF matrix element,

$$2P_0h^B(z,P)\equiv\langle\pi;P|\mathcal{O}(z)|\pi;P\rangle, \quad (5)$$

through the spectral analysis of Eq. (4) either in momentum space or in the real-space t_s after Fourier transforming $\tilde{C}$ with respect to p_0 to form $C_{3\text{pt}}(z,t_s,\mathbf{p},q)$. A convenient choice $q=(0,\mathbf{0})$ gives the so-called summation method [50], wherein, $C_{3\text{pt}}(z,t_s,\mathbf{p},q=0)/C_{2\text{pt}}(t_s,\mathbf{p})=t_sh^B(z,\mathbf{p})+\text{constant}$, up to $O(e^{-(E_1-E_0)t_s})$ excited-state corrections.

IV. COMPUTATIONAL DETAILS

We implemented the continuum reduction approach to determine the u-quark PDF of the large-N_c pion in $d=4$. As a first exploratory study, we performed our computation at a fixed simulation point at a large but finite value of $N_c=17$ on an $L=8$ lattice using a coupling $b=0.355$, which is in the confined phase [48]. The lattice spacing in units of string tension [51,52] is $\sqrt{\sigma}a=0.254(2)$. Due to the finite large N_c, the $U^d(1)$ center symmetry reduces to $Z_{N_c}^d$ discrete symmetry, and therefore, we quantized the lattice momenta in units of $2\pi/(LN_c)$ and multiples thereof, to leave the above results intact. In this way, we effectively enlarged the 8^4 lattice into a 68×136^3 lattice. We used a Wilson-Dirac operator coupled to smeared gauge links for D^L and tuned the quark mass to produce a pion of mass $m_\pi=0.86\sqrt{\sigma}$. We stochastically computed $\tilde{C}_{2\text{pt}}(p)$ and $\tilde{C}_{3\text{pt}}(z,p,q=0)$ at all values of p_0 at each given $\mathbf{p}=(0,0,P_3)$, using 15000-32000 configurations, and Fourier transformed them into functions of t_s. We studied nine different spatial momentum $P_3/\sqrt{\sigma}\in[0,5.82]$. We elaborate further on the lattice setup in Appendix B. We determined h^B using summation type fits to $C_{3\text{pt}}/C_{2\text{pt}}$ ratio. For further details on the spectral analysis of two-point and three-point functions, the reader can refer to Appendixes C and D, respectively. Since $h^B(z,P)$ is multiplicatively renormalizable [53,54], we took the renormalization group invariant ratio [55] of quasi-PDF matrix elements at $P_3\neq0$ with respect to $P_3=0$ to form the pseudo Ioffe-time distribution (pseudo-ITD), $\mathcal{M}(\nu,z^2)$.

V. COLLINEAR QUARK STRUCTURE OF THE LARGE-N_c PION

We show the real and imaginary parts of the u-quark pseudo-ITD $\mathcal{M}(\nu,z^2)$ as a function of ν in Fig. 2. As seen from Eq. (3), the two are governed by $u-\bar{u}$ and $u+\bar{u}$ PDFs respectively. In $SU(3)$ theory, $u+\bar{u}$ PDF mixes with gluon PDF; however, this mixing is N_c^{-1} suppressed and hence ignored here. The data points are the result of our

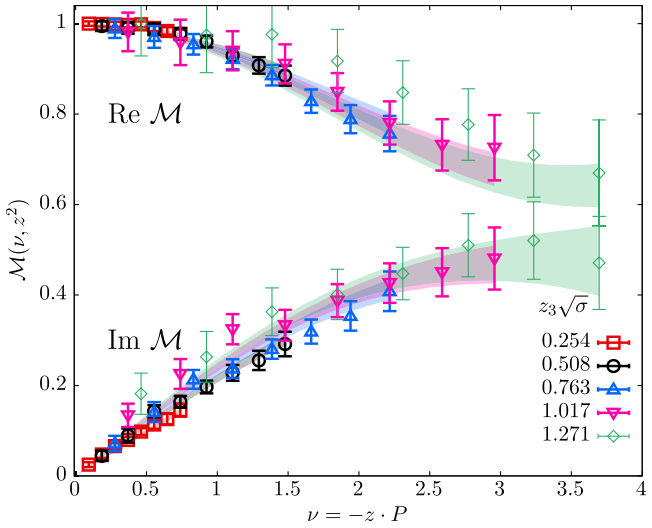

FIG. 2. The real and imaginary parts of the pseudo-ITD of pion $\mathcal{M}(\nu, z^2)$ in the large-N_c limit are shown. The lattice data from different values of quark-antiquark separation z_3 are shown using different colored symbols. The bands are fits to the leading-twist OPE with the large-N_c NLO Wilson coefficients at $\overline{\text{MS}}$ scale $\mu = 4.55\sqrt{\sigma}$. The fit parameters are the Mellin moments.

lattice computation from different (z_3, P_3) put together. The near-continuous set of momenta we were able to use helped us pack the range of ν with data points. In the large-N_c limit, the string tension $\sqrt{\sigma}$ sets a fiducial scale that distinguishes perturbative and nonperturbative length scales; therefore, we restricted the data for $\mathcal{M}$ to only those up to the borderline $\sqrt{\sigma} z_3 \leq 1.27$. At the same time, a cautious use of $|P_3| < a^{-1}$ only let us scan a range of $\nu < 3.5$. The near universality of the data with respect to the scaling variable ν points to the viability of perturbative OPE methods in the large-N_c theory. By fitting the lattice data using the leading-twist OPE in Eq. (3), we extracted the Mellin moments $\langle x^n \rangle_{u\pm\bar{u}} \equiv \int_0^1 x^n f_{u\pm\bar{u}}(x,\mu) dx$ at a scale $\mu = 4.55\sqrt{\sigma}$ using the one-loop result [49] for the Wilson coefficients $C_n(\mu^2 z^2)$ in the large-N_c limit; for this we used the leading-order value $\lim_{N_c\to\infty} C_F(N_c)\alpha_s(\mu) = 0.39$ using $\Lambda_{\overline{\text{MS}}}/\sqrt{\sigma} = 0.503$ [56,57]. We chose a scale $\mu \approx a^{-1}$ so that it is characteristic of the typical small z_3 used in this work. We gather the technical details for the OPE fits and on the perturbative factors in Appendix E and Appendix F 1, respectively. We find for the first few moments

$$[\langle x \rangle_{u+\bar{u}}, \langle x^3 \rangle_{u+\bar{u}}] = [0.25(1), 0.10(2)] \quad (\text{via Im}\,\mathcal{M}),$$
$$[\langle x^2 \rangle_{u-\bar{u}}, \langle x^4 \rangle_{u-\bar{u}}] = [0.13(2), 0.10(2)] \quad (\text{via Re}\,\mathcal{M}), \quad (6)$$

with correlated $\chi^2/\text{df} \sim 39/26$ in the two cases. As a cross-check that a perturbative OPE framework is working for the chosen range in z_3, we used the fixed-z^2 moments analysis [58] as a diagnostic tool [30,59] to detect any corrections—as discussed in Appendix G, we found the method to work well within statistical errors. In addition to the above Mellin moments analysis, we also performed fits to the valence $u - \bar{u}$ data assuming a phenomenologically motivated functional form [60], $f_{u-\bar{u}}(x;\alpha,\beta,s) = \mathcal{N} x^\alpha (1-x)^\beta (1 + s x^2)$. Since our access to the range of ν is limited in this work, and the small-x region is believed to be harder to access on the lattice, we imposed a prior that $\alpha \in [-0.6, -0.4]$ motivated by the Regge phenomenology. With the caveat of using an Ansatz, we found the data to be best described by a large-x exponent $\beta = 0.7(3)$, similar to what is seen in recent lattice $SU(3)$ QCD results [27–30] as well as by global fits [60,61]. From an indirect estimation of the valence momentum fraction, $\langle x \rangle_{u-\bar{u}} = 0.23(2)$ from the PDF Ansatz fit, we find it to be the same as $\langle x \rangle_{u+\bar{u}}$ within errors; thus, there might only be a negligible amount of anti-u in the large-N_c pion wave function.

VI. PHENOMENOLOGY

In order to use large-N_c theory as a model system to compare and contrast the $SU(3)$ QCD with, we first set the GeV scale in the $SU(\infty)$ world through a choice $\sqrt{\sigma} = 0.44$ GeV that is known [3,7,62] to result in a low-energy meson spectrum that is numerically similar to the real world; this choice implies, $[a^{-1}, \mu, m_\pi] = [1.73, 2.00, 0.38]$ GeV in our computation. We use the $\overline{\text{MS}}$ Ioffe time distribution (ITD), $\mathcal{M}^{\overline{\text{MS}}}_{u\pm\bar{u}}(\nu,\mu)$, which are the cosine (for $u - \bar{u}$) and sine (for $u + \bar{u}$) Fourier transforms of PDFs from x to ν space, to justifiably perform this comparison within the range of ν spanned by our lattice data. In Fig. 3, we compare $\mathcal{M}^{\overline{\text{MS}}}_{u\pm\bar{u}}$ for the large-N_c pion (red band), as inferred from the model-independent fits to Mellin moments, with the JAM20 global fit result [60] (gray band) for the real-world pion at $\mu = 2$ GeV. We find a good agreement between the two theories in the case of the valence $u - \bar{u}$ ITD. We suspect that the observed tendency for $SU(\infty)$ data to peel off at $\nu \approx 3$ could be a systematic effect due to the absence of constraint from data beyond $\nu = 3.5$, and in fact, such a feature is absent in the ITD reconstructed from $f_{u-\bar{u}}(x;\alpha,\beta,s)$ (purple band). In the inset of Fig. 3, we also see a nearly similar x dependence of our Ansatz-based reconstruction of $f_{u-\bar{u}}(x)$ (purple band) and the JAM20 result. Thus, the valence structure of pion is likely to be weakly dependent on N_c. It appears that features like the valence quarks that carry ≈50% of the pion momentum at few GeV resolutions could be typical in $SU(N_c)$ theories.

In the bottom panel of Fig. 3, we show a similar comparison between $u + \bar{u}$ ITDs at $\mu = 2$ GeV. Here, we see a visible difference between the two theories. Based on a better agreement seen in the valence sector, we ask if the difference seen in the singlet $u + \bar{u}$ distribution could originate from the perturbative radiative processes in the large-N_c and $SU(3)$ QCD; as a main difference, the $g \to q\bar{q}$ splitting is absent when $N_c \to \infty$. In Appendix F 2, we

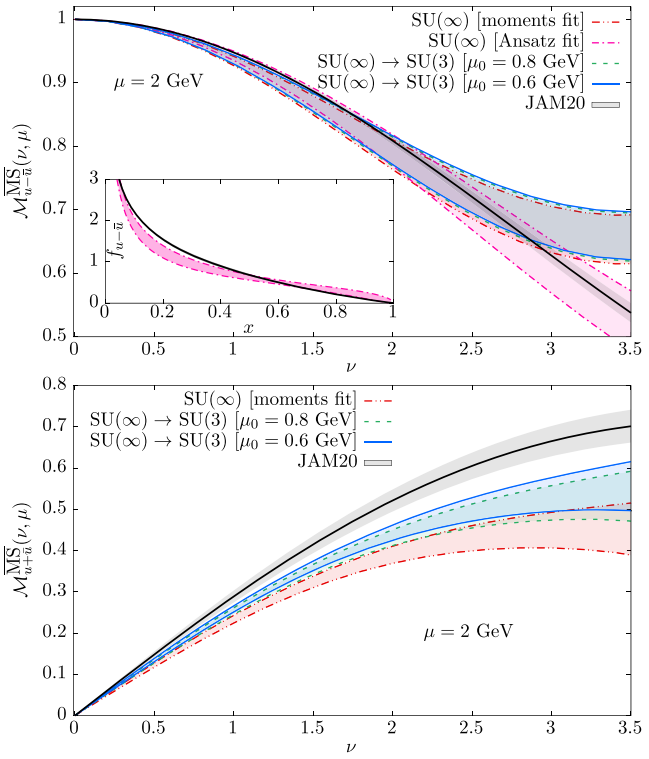

FIG. 3. The comparison of $\overline{\rm MS}$ ITDs in large-N_c QCD [SU(∞)] with the global fit results (JAM20) for the ITDs in SU(3) QCD at scale $\mu = 2$ GeV. The top and bottom panels are $u - \bar{u}$ and $u + \bar{u}$ ITDs, respectively. The red bands (moments fit) are results from OPE analysis by fitting Mellin moments. The green and blue bands are expectations for ITDs in SU(3) QCD based on the assumption of a nearly similar large-N_c PDF at a lower factorization scale $\mu_0 = 0.8$ and 0.6 GeV, respectively. The results for $u - \bar{u}$ ITD and PDF assuming an ansatz $f_{u-\bar{u}}(x, \alpha, \beta, s)$ are shown as purple bands (Ansatz fit) in the top panel and its inset, respectively.

discuss the perturbative evolution aspects in the large-N_c limit. Working under a premise that the large-N_c and SU(3) theories have similar $u + \bar{u}$ PDFs at a low-factorization scale μ_0, we first evolved the pairs, $[2\langle x^n\rangle_{u+\bar{u}}, \langle x^n\rangle_g]$ at $\mu = 2$ GeV in the large-N_c theory to a scale μ_0 (= 0.8 to 0.6) GeV, and evolved that result back to $\mu = 2$ GeV using 3 flavor SU(3) QCD DGLAP evolution. Since we have not explicitly calculated the gluon moments for large-N_c theory, we used the sum rule $\langle x\rangle_g = 1\text{–}2\langle x\rangle_{u+\bar{u}}$ and simply set the other higher moments of the small-x dominant gluon to be negligible. As we are looking only for qualitative tendencies, we performed the evolution at leading-logarithmic order using the same $\Lambda_{\overline{\rm MS}}$ in both theories. We show the resulting ITDs based on evolutions from $\mu_0 = 0.8$ and 0.6 GeV as the green and the blue bands in Fig. 3 lower panel. Remarkably, the QCD evolution pulls the large-N_c result closer to the JAM20 result when successively smaller μ_0 are used. Thus, large-N_c QCD presents itself as an interesting model system for singlet parton physics where $g \to q\bar{q}$ splitting is switched off, with all other splitting remaining intact. Such a procedure only leads to a negligible effect in valence $u - \bar{u}$ ITD as seen in Fig. 3 top panel.

VII. DISCUSSION

We presented the large-N_c mesons as an interesting uncharted model system for understanding partonic physics, using the continuum reduction. Our first lattice computation of large-N_c pion structure shows indeed that the structural properties in the large-N_c theory are likely to be similar to our real world, as has been seen in the meson spectrum; their differences seem to be even more interesting as it gives us a version of QCD where the sea is not radiatively proliferated with quark-antiquark pairs, and hence could help us understand the role of sea quarks in real-world QCD. Not to be mistaken, the method needs to be improved by going to finer coupling, larger N_c, and also cross-checked with a complementary twisted Eguchi-Kawai reduction [63]. An easy generalization of the method to QCD_2 might help in pruning the Monte Carlo methods by direct comparisons with analytical results [2,64–67]. A large-N_c an advantage could be the exponential suppression [68–70], of small instantons in the large-N_c limit that might suppress instanton-induced power corrections [71–73] to the OPE at typical short distances reached in contemporary lattice calculations. It would be interesting to extend this work to probe the differences in gluon structures of the radial and angular stable excitations [31,74] of the ground-state mesons, and perform x-dependent spin physics of stable higher-spin large-N_c mesons, such as the ρ.

ACKNOWLEDGMENTS

The authors thank R. Edwards, K. Orginos, and J. Qiu for the valuable discussions. N. K. thanks P. Barry for helping with the JAM20 data. The authors thank and acknowledge William & Mary Research Computing for providing computational resources and technical support that have contributed to the results reported within this paper [75]. The work was performed on the *Femto* and *Meltemi* computing clusters at William & Mary. R. N. acknowledges partial support by the NSF under Grant No. PHY-1913010. N. K. is supported by Jefferson Science Associates, LLC under U.S. DOE Contract No. DE-AC05-06OR23177 and in part by U.S. DOE Grant No. DE-FG02-04ER41302.

APPENDIX A: DETAILS ON CONTINUUM REDUCTION

We consider a L^d periodic lattice. The gauge action in terms of plaquettes U_p is

$$S_g = bN\sum_p \mathrm{Tr}(U_p + U_p^\dagger), \tag{A1}$$

and gauge fields on an infinite d-dimensional lattice obey the periodic condition,

$$U_\mu(x) = U_\mu(x + L\hat{\nu}); \quad \forall\, \mu, \nu. \tag{A2}$$

As long as the lattice coupling $b < b_1(L)$ the theory is in the confined phase and infinite volume results can be computed exactly with finite lattice spacing effects. Below, we derive expressions for certain n-point functions of quark bilinears (mesons) using the continuum reduction framework. We specify the n-point functions as $G^{(n)}$. We derive results using continuum reduction for a general dimension d, but we finally used only $d = 4$ in this work.

1. Quark propagator in infinite lattice from finite periodic lattice: Bloch wave functions and $U^d(1)$ global symmetry

Consider an operator $F(U)$ on an infinite lattice obtained by copying the gauge fields from the L^d lattice using periodicity. As per Bloch's theorem, the eigenvalue problem takes the form

$$\sum_y^\infty F_{x,y}(U) q_y^i = \lambda_i(p) q_x^i(p); \quad q_{x+nL}^i = e^{ip\cdot n} q_x^i(p), \tag{A3}$$

where n is a tuple of integers. Under a gauge transformation g,

$$U_\mu^g(x) = g_x U_\mu(x) g_{x+\hat{\mu}}^\dagger, \quad F_{x,y}(U^g) = g_x F_{x,y}(U) g_y^\dagger. \tag{A4}$$

Of particular interest to us will be Abelian gauge transformations of the form $g_x = e^{-i\frac{p\cdot x}{L}}$ on the infinite lattice. Under these gauge transformations,

$$U_\mu^g(x) = U_\mu(x) e^{i\frac{p_\mu}{L}} \quad \text{(written in short as } Ue^{i\frac{p}{L}}\text{)}, \tag{A5}$$

and we can rewrite the eigenvalue problem as

$$\sum_y^\infty F_{x,y}(Ue^{i\frac{p}{L}}) q_y^{ig}(p) = \lambda_i(p) q_x^{ig}(p); \quad q_x^{ig}(p) = e^{-i\frac{p\cdot x}{L}} q_x^i(p);$$
$$q_{x+nL}^{ig}(p) = q_x^{ig}(p). \tag{A6}$$

One can use the periodicity of $q_x^{ig}(p)$ to further rewrite the eigenvalue problem as

$$\sum_y^L F_{x,y}^L(Ue^{i\frac{p}{L}}) q_y^{ig}(p) = \lambda_i(p) q_x^{ig}(p);$$
$$F_{x,y}^L(U) \equiv \left[\sum_{n=-\infty}^\infty F_{x,y+nL}(U)\right], \tag{A7}$$

and the induced operator F^L on the finite periodic lattice satisfies

$$F_{x,y}^L(U) = F_{x,y+nL}^L(U) = F_{x+nL,y}^L(U) \tag{A8}$$

for any vector n with integer entries. The above eigenvalue equation is for a finite size matrix on a finite periodic lattice. We can write the operator and its inverse on the infinite lattice using their finite volume counterparts as

$$F_{x,y}(U) = \int \frac{d^dp}{(2\pi)^d} \sum_i \lambda_i(p) q_x^i(p) [q_y^i(p)]^\dagger$$
$$= \int \frac{d^dp}{(2\pi)^d} e^{i\frac{p\cdot(x-y)}{L}} F_{x,y}^L(Ue^{i\frac{p}{L}});$$
$$F_{x,y}^{-1}(U) = \int \frac{d^dp}{(2\pi)^d} e^{i\frac{p\cdot(x-y)}{L}} [F^L]_{x,y}^{-1}(Ue^{i\frac{p}{L}}). \tag{A9}$$

One can extend the above relation to a product of operators:

$$\sum_z^\infty A_{x,z}(Ue^{i\phi}) B_{z,y}(Ue^{i\chi}) = \sum_z^\infty \int \frac{d^dp}{(2\pi)^d}\frac{d^dq}{(2\pi)^d} e^{i\frac{p\cdot(x-z)}{L}} e^{i\frac{q\cdot(z-y)}{L}} A_{x,z}^L(Ue^{i\phi}e^{i\frac{p}{L}}) B_{z,y}^L(Ue^{i\chi}e^{i\frac{q}{L}})$$
$$= \sum_z^L \sum_{k=-\infty}^\infty \int \frac{d^dp}{(2\pi)^d}\frac{d^dq}{(2\pi)^d} e^{i\frac{p\cdot(x-z-kL)}{L}} e^{i\frac{q\cdot(z+kL-y)}{L}} A_{x,z}^L(Ue^{i\phi}e^{i\frac{p}{L}}) B_{z,y}^L(Ue^{i\chi}e^{i\frac{q}{L}})$$
$$= \int \frac{d^dp}{(2\pi)^d} e^{i\frac{p_j\cdot(x-y)}{L}} \left[\sum_z^L A_{x,z}^L(Ue^{i\phi}e^{i\frac{p}{L}}) B_{z,y}^L(Ue^{i\chi}e^{i\frac{p}{L}})\right]. \tag{A10}$$

One application of the above reduction we will use involves the quark propagator, $G_{x,y}(U)$, in a fixed gauge field background and a smearing operator, $S_{x,y}^\phi(U)$, in a fixed gauge field background where ϕ labels the type of smearing. Specific to this work, $S(U_\mu)$ is the Wuppertal smearing kernel, and $S_{x,y}^\phi(U_\mu) = S(e^{i\phi_\mu} U_\mu)$ for a phase $\phi = (0, \phi_1, \phi_2, \phi_3)$ with nonzero spatial components in general, which we also set to 0. Hence, what follows is for a more general case than actually used in the present computation. The smearing operator is typically diagonal

in spinor space. We will assume $S^{\phi}(U)$ is Hermitian as is true in most cases that are typically functions of the covariant Laplacian operator. A form of reduction we will need is

$$\sum_{x',x''}^{\infty} S^{\phi_1}_{x,x'}(U) G_{x',x''}(U) S^{\phi_2}_{x'',y}(U) = \int \frac{d^d p}{(2\pi)^d} e^{i\frac{p\cdot(x-y)}{L}} G^{L;\phi_1\phi_2}_{x,y}(U e^{i\frac{p}{L}}), \tag{A11}$$

where we have used different smearing operators on either side of the unsmeared quark propagator and we define the smeared propagator on the finite periodic lattice by

$$G^{L;\phi_1\phi_2}_{x,y}(U) = \sum_{x',x''}^{L} S^{L\phi_1}_{x,x'}(U) G^{L}_{x',x''}(U) S^{L\phi_2}_{x'',y}(U) \tag{A12}$$

which satisfies

$$G^{L;\phi_1\phi_2}_{x,y}(U) = G^{L;\phi_1\phi_2}_{x,y+nL}(U) = G^{L;\phi_1\phi_2}_{x+nL,y}(U) \tag{A13}$$

for any vector n with integer entries as expected of a propagator on a periodic lattice.

2. Two-point function of mesons

Let

$$M^{\Gamma}_{ij}(x) = \sum_{x',x''} \bar{q}^{(i)}_{x'} S^{\phi_i}_{x',x}(U) \Gamma S^{\phi_j}_{x,x''}(U_\mu) q^{(j)}_{x''} \tag{A14}$$

be a gauge invariant meson operator located at x. Since $S_{x,y} \propto \delta_{x_0,y_0}$, the sums above are actually restricted to s time slice containing x by construction, but written as a sum over the entire space-time. The indices i, j provide the quark flavor indices and Γ specifies the type of fermion in a spinor space. For the case of pion, considered in the paper,

$$\pi(x) = M^{\gamma_5}_{du}(x) = \sum_{x',x''} \bar{d}_{x'} S_{x',x}(U e^{i\phi}) \gamma_5 S_{x,x''}(U e^{-i\phi}) u_{x''}. \tag{A15}$$

Using Eq. (A11), the two-point function of a meson of a type Γ_1 with type Γ_2 in the infinite lattice is

$$G^{(2)}(x,y) = \langle M^{\Gamma_1}_{ij}(x) [M^{\Gamma_2}_{ij}]^{\dagger}(y) \rangle, \tag{A16}$$

which after Wick contraction yields

$$G^{(2)}(x,y;U) = \int \frac{d^d q}{(2\pi)^d} \frac{d^d q'}{(2\pi)^d} e^{i\frac{(q-q')\cdot(x-y)}{L}} \times \mathrm{tr}\,[\Gamma_1 G^{L;\phi_2\phi_2}_{x,y}(U e^{i\frac{q}{L}}) \Gamma_2 G^{L;\phi_1\phi_1}_{y,x}(U e^{i\frac{q'}{L}})]. \tag{A17}$$

If we write down the two-point function in momentum space using

$$\tilde{G}^{(2)}(p',p;U) = \sum_{x,y}^{\infty} e^{i(p'\cdot x + p\cdot y)} G^{(2)}(x,y;U) \tag{A18}$$

and split the infinite sum over x and y into blocks of finite sums over finite periodic lattice (i.e., $x \to x + n \bmod L$, and replace sum to be over the periodic x and n) and invoke the periodicity property in Eq. (A13), we will arrive at condition $Lp + q' - q = 0$ along with momentum conservation $p + p' = 0$. Using these, we can write the two-point function as

$$\tilde{G}^{(2)}(p',p;U) = \delta(p'+p) \int \frac{d^d q}{(2\pi)^d} \mathrm{Tr}[\Gamma_1 G^{L;\phi_2\phi_2}(U e^{i\frac{q}{L}}) \times \Gamma_2 G^{L;\phi_1\phi_1}(U e^{i\frac{q}{L}+p'})], \tag{A19}$$

where Tr denotes the trace over the entire lattice and spin. Invoking the $U^d(1)$ global symmetry present in the confined phase, we can shift $U e^{i\frac{q}{L}} \to U$, and we can write the propagator in momentum space at a fixed gauge field background as

$$\tilde{G}^{(2)}(p',p;U_\mu) = \delta(p'+p)\,\mathrm{Tr}[\Gamma_1 G^{L\phi_2\phi_2}(U) \Gamma_2 G^{L\phi_1\phi_1}(U e^{-ip})]. \tag{A20}$$

For the pion, for which $\phi_1 = -\phi_2 = \phi$ so as to preserve isospin symmetry during quark smearing, we defined the two-point function in the main text as

$$\tilde{C}_{2\mathrm{pt}}(p) = \mathrm{Tr}[\gamma_5 G^{L;-\phi,-\phi}(U) \gamma_5 G^{L;\phi,\phi}(U e^{-ip})], \tag{A21}$$

with explicit smearing factors included in the detailed expression.

3. Quasi-PDF-pion-pion three-point function

The fermion bilinear connected by a spatial Wilson line from w to $w+z$ for $z = (0,0,0,z_3)$, i.e., the quasi-PDF operator, is given by

$$\mathcal{O}(w;z) = \bar{u}_w \gamma_0 W_{w;w+z} u_w. \tag{A22}$$

Our focus will be on the three-point function

$$G^{(3)}(x,y,w;z) \equiv \langle \pi(x)\mathcal{O}(w;z)\pi^\dagger(y)\rangle. \quad \text{(A23)}$$

Strictly in the large-N_c limit, we can ignore the quark-line disconnected diagrams and write

$$\begin{aligned} G^{(3)}(x,y,w;z) = \sum_{x',x'',y',y''} & \mathrm{tr}(\gamma_5 S_{x,x'}(Ue^{-i\phi})G_{x',w}(U) \\ & \times \gamma_0 W_{w,w+z}(U) G_{w+z,x''}(U) \\ & \times S_{x'',y}(Ue^{-i\phi})\gamma_5 S_{y,y'}(Ue^{i\phi}) \\ & \times G_{y',y''}(U) S_{y'',x}(Ue^{i\phi})). \end{aligned} \quad \text{(A24)}$$

Replacing infinite lattice propagators by propagators on periodic lattice as we have done before,

$$\begin{aligned} G^{(3)}(x,y,w;z) = \int \frac{d^4q}{(2\pi)^4}\frac{d^4q'}{(2\pi)^4}\frac{d^4q''}{(2\pi)^4} & e^{-iq\cdot(w-x)/L} \\ & \times e^{-iq'\cdot(y-w-z)/L} e^{-iq''\cdot(x-y)/L} \\ & \times \mathrm{tr}(\gamma_5 G^{L;-\phi,\varnothing}_{x,w}(Ue^{iq/L})\gamma_0 W_{w,w+z}(U) \\ & \times G^{L;\varnothing;-\phi}_{w+z,y}(Ue^{iq'/L})\gamma_5 G^{L;\phi,\phi}_{y,x}(Ue^{iq''/L})), \end{aligned} \quad \text{(A25)}$$

where we have used the following notation:

$$\begin{aligned} G^{L;-\phi,\varnothing}_{x,y}(U) &= \sum_{x'} S^L_{x,x'}(Ue^{-i\phi})G^L_{x',y}(U), \\ G^{L;\varnothing,-\phi}_{x,y}(U) &= \sum_{x'} G^L_{x,x'}(U)S^L_{x',y}(Ue^{-i\phi}), \\ G^{L;\phi,\phi}_{x,y}(U) &= \sum_{x',x''} S^L_{x,x'}(Ue^{i\phi})G^L_{x',x''}(U)S^L_{x'',y}(Ue^{i\phi}). \end{aligned} \quad \text{(A26)}$$

Fourier transforming over $(x,w,y) \to (p',Q,p)$ on the infinite lattice keeping z fixed,

$$\begin{aligned} \tilde{G}^{(3)}(p',Q,p;z,U) &= \sum_{x,w,y}^{\infty} \int \frac{d^4q}{(2\pi)^4}\frac{d^4q'}{(2\pi)^4}\frac{d^4q''}{(2\pi)^4} e^{-iq\cdot(w-x)/L} e^{-iq'\cdot(y-w-z)/L} \\ & \times e^{-iq''\cdot(x-y)/L} e^{i(p'\cdot x + Q\cdot w + p\cdot y)} \\ & \times \mathrm{tr}(\gamma_5 G^{L;-\phi,\varnothing}_{x,w}(Ue^{iq/L})\gamma_0 W_{w,w+z}(U) \\ & \times G^{L;\varnothing,-\phi}_{w+z,y}(Ue^{iq'/L})\gamma_5 G^{L;\phi,\phi}_{y,x}(Ue^{iq''/L})). \end{aligned} \quad \text{(A27)}$$

We can split the infinite sum over x, y, w into blocks of sums over finite periodic lattice, invoke the periodicity property in Eq. (A13), and use the folded property of Wilson lines. This will result in

$$\begin{aligned} & q - q'' + p'L = 0; \qquad q' - q + QL = 0; \\ & q'' - q' + pL = 0, \end{aligned} \quad \text{(A28)}$$

which includes the momentum conservation, $p' + Q + p = 0$. We arrive at

$$\begin{aligned} \tilde{G}^{(3)}(p',Q,p;z,U) &= \delta(p'+Q+p)\int \frac{d^4q'}{(2\pi)^4}\sum_{x,w,y}^{L} \mathrm{tr}(\gamma_5 G^{L;-\phi,\varnothing}_{x,w}(Ue^{i(\frac{q'}{L}+Q)}) \\ & \times \gamma_0 W_{w,w+z}(Ue^{i\frac{q'}{L}}) G^{L;\varnothing,-\phi}_{w+z,y}(Ue^{i\frac{q'}{L}})\gamma_5 G^{L;\phi,\phi}_{y,x}(Ue^{i(\frac{q'}{L}-p)})). \end{aligned} \quad \text{(A29)}$$

Using $U^d(1)$ symmetry,

$$\begin{aligned} \tilde{G}^{(3)}(p',Q,p;z,U) &= \delta(p'+Q+p)\sum_{x,w,y}^{L} \mathrm{tr}(\gamma_0 W_{w,w+z}(U) G^{L;\varnothing,-\phi}_{w+z,y}(U) \\ & \times \gamma_5 G^{L;\phi,\phi}_{y,x}(Ue^{-ip})\gamma_5 G^{L;-\phi,\varnothing}_{x,w}(Ue^{iQ})). \end{aligned} \quad \text{(A30)}$$

We referred to the above equation in the main text, now with an explicit specification of quark smearing factors, as

$$\begin{aligned} \tilde{C}_{3\mathrm{pt}}(z,p,Q) = \sum_{x,w,y}^{L} & \mathrm{tr}(\gamma_0 W_{w,w+z}(U) G^{L;\varnothing,-\phi}_{w+z,y}(U) \\ & \times \gamma_5 G^{L;\phi,\phi}_{y,x}(Ue^{-ip})\gamma_5 G^{L;-\phi,\varnothing}_{x,w}(Ue^{iQ})). \end{aligned} \quad \text{(A31)}$$

We used $Q = 0$ in this work.

APPENDIX B: DETAILS OF THE LATTICE CALCULATION

In the present work, we used a fixed large value of $N_c = 17$, since it is the smallest value of N_c beyond which the $1/N_c$ corrections are typically found to be small in previous works. We used L^4 lattices in this paper with $L = 8$. We used the standard single plaquette Wilson gauge action and set the lattice coupling $b = 0.355$, such that it is close to being the largest b possible on $L = 8$ and keep the lattice gauge theory in the confined phase (phase 0c). We used the critical $L_c(b) = 6.6$ for the value of b. Since fermion loops are $1/N_c$ suppressed, the quenched lattice computation of fermionic quantities is exact in the large-N_c limit. Each update of the gauge fields on the entire lattice was made up of $\frac{N_c(N_c-1)}{2}$ SU(2) heat-bath updates on every link followed by one SU(N) over-relaxation update on every link [48]. We performed 100 such updates between measurements to avoid autocorrelation. To make sure the configurations thermalized to the 0c phase, we successively decreased the value of b from a higher value of $b = 0.365$. By monitoring the gap in the Polyakov loop eigenvalues [48] in all four directions, we ensured that the configurations were in the correct phase. We computed the pion-pion

two-point and pion-quasi-PDF-pion three-point functions [Eqs. (1) and (3) in the main text] on every configuration at 9 different values of momentum,

$$P_3 a = \frac{2\pi n_3}{N_c L} = \frac{2\pi n_3}{136}, \tag{B1}$$

for $n_3 = 0, 2, 4, \ldots, 16$. We used gradually more numbers of configuration $N_{\rm cfg}$ (∼32 K) at the higher momenta compared to the lower ones ($N_{\rm cfg} \sim 12$ K). We have collected the details of the statistics in Table I.

We evaluated the two-point and three-point functions [Eqs. (1) and (3) in the main text] stochastically. Namely, for the two-point function, the stochastic estimator using noise vectors ξ,

$$\tilde{C}_{\rm 2pt}(p) = \overline{\xi^\dagger \gamma_5 G^L(U)\gamma_5 G^L(Ue^{ip})\xi} = \overline{\chi^\dagger(0)\phi(p)}, \tag{B2}$$

with $\phi(p) \equiv G^L(Ue^{iq})\xi$ and $\chi(0) \equiv \gamma_5[G^L(U)]^\dagger\gamma_5\xi$. The combined noise and ensemble average is

$$\overline{\xi^\dagger A \xi} = \frac{1}{N_{\rm vec} N_{\rm cfg}} \sum_{i=1}^{N_{\rm vec}\times N_{\rm cfg}} \xi_i^\dagger A \xi_i. \tag{B3}$$

We used $N_{\rm vec} = 3$ number of $\mathbb{Z}_2$ noise vectors for ξ; that is $\xi^{\alpha,a,x} = \frac{1\pm i}{\sqrt{2}}$ for spin, color, and position indices α, a, x, respectively. We further diluted the noise vectors over even-odd lattice sites and over the two chiral projections. For the three-point function, we used the stochastic estimator as

$$\tilde{C}_{\rm 3pt}(z,p,q) = \sum_x \overline{\xi_x^\dagger \gamma_0 W_{x,x+z}\phi_{x+z}(p)};$$
$$\phi(p) = G^L(U)\gamma_5 G^L(Ue^{-ip})\gamma_5 G^L(Ue^{iq})\xi. \tag{B4}$$

In this work, we only used $q = 0$ above. We used two steps of Stout smearing for the gauge links that are used to construct the Wilson line W.

We used Wilson-Dirac operator $\not{D}^L(m_w)$ to compute the propagators $G^L = [\not{D}^L]^{-1}$. We improved the Dirac operator by using gauge links that are smeared by two steps of the large-N_c version of the Stout smearing [76]. With smearing, we expect the zero quark mass to be in the region of the Wilson mass $m_w = [-0.38, -0.39]$. We tuned to $m_w = -0.36$ to realize a pion mass that was feasible given the computational resource available to us. We implemented the Wilson-Dirac inversion using the BiCG-Stab algorithm [77].

We used smeared quark sources in the construction of two-point and three-point functions using a smearing kernel $S(U; N_{\rm wup}, \delta)$ for the Wuppertal smearing [78]. We implemented Wuppertal smearing using $(N_{\rm wup}, \delta) = (40, 0.6)$, which we chose to be optimal through a set of initial tuning runs. We kept the radius of Wuppertal smearing fixed at all pion momenta. A puzzling experience during the tuning process at non-zero momenta was the negligible effect of phased momentum smearing [79], $S(e^{i\phi}U)$, which typically improves the signal-to-noise ratio at higher momenta in the SU(3) QCD at some value of ϕ; we did not find any such improvement within statistical errors during the tuning phase in which we used only about ∼1000 configurations. Therefore, we simply used unphased ($\phi = 0$) Wuppertal kernel for quark smearing at all momenta.

TABLE I. The table lists the momenta $P_3 = (\frac{2\pi}{L_s N_c})n_3$, and the amount of statistics at each momentum. The statistics comes from two sources; namely, independent number of gauge field configurations (second column) and the number of Z_2 stochastic vectors in each configuration. We used 3 Z_2 random vectors which are diluted in chirality and in even-odd lattice sites, which comes out to 12 sets of inversions over the components of the noise vectors. We fixed this for all momenta. To convert P_3 to GeV, $P_z = 0.081 n_3$ GeV. All the momenta used in this work are below the lattice a^{-1} scale.

n_3	$P_3 a$	$P_3/\sqrt{\sigma}$	Configurations
0	0	0	15353
2	0.092	0.363	16320
4	0.185	0.727	11520
6	0.277	1.090	19200
8	0.370	1.454	30720
10	0.462	1.817	27552
12	0.554	2.181	30720
14	0.647	2.544	30720
16	0.739	2.910	30504

APPENDIX C: SPECTRAL CONTENT OF PION TWO-POINT FUNCTIONS

1. Construction

We determined the two-point function $\tilde{C}_{\rm 2pt}(p)$ with $p = (p_0, 0, 0, P_3)$ for $p_0 = 2\pi n_0/L_0^{\rm eff}$ using $n_0 \in [0, \frac{L_0^{\rm eff}}{2}]$ and the effective temporal extent of $L_0^{\rm eff} = 68 = N_c L/2$. We could have used an effective temporal extent of up to 136, but instead we used a smaller one to make the computation easier, and an effective temporal extent of 68 is quite comparable to what is being used in present structural computations in SU(3) QCD. One possibility to investigate the spectral content in the two-point function is to fit the data at different fixed spatial P_3 to

$$\tilde{C}_{\rm 2pt}(p_0, P_3) = B + \sum_{i=0}^{N_{\rm st}-1} \frac{|A_i|^2 \sinh(aE_i(P_3))}{\cosh(aE_i(P_3) - \cos(ap_0))};$$
$$A_i = \langle E_i|\pi^\dagger|0\rangle, \tag{C1}$$

where one can truncate the momentum space spectral decomposition at $N_{\rm st}$ number of states. It is to be remembered

that, even if higher excited states might not contribute to the p_0 dependence of the correlator in the range of smaller p_0, they can still contribute a momentum-independent constant to the above correlator. Therefore, we corrected such a truncated series by a constant term, B, to account for such effects of all other higher excited states.

In this paper, we used the computed momentum space two-point functions to Fourier transform them into real space, so that we could perform a rather traditional lattice QCD analysis via effective masses and multiexponential fits. That is, the real-space correlator is

$$C_{2\mathrm{pt}}(t_s, P_3) = \sum_{n_0=0}^{L_t^{\mathrm{eff}}-1} \tilde{C}_{2\mathrm{pt}}(p_0, P_3) e^{ip_0 \frac{t_s}{a}};$$
$$\tilde{C}_{2\mathrm{pt}}(L_0^{\mathrm{eff}} - n_0, P_3) = \tilde{C}_{2\mathrm{pt}}(n_0, P_3). \quad \text{(C2)}$$

Note that we have used the momenta p and their integer quanta n interchangeably as arguments above, and we will do so in the rest of the text without any obvious confusion.

After the above Fourier transformation, we performed the usual N_{st}-state fits to study their spectral content,

$$C_{2\mathrm{pt}}(t_s, P_3) = \sum_{i=0}^{N_{\mathrm{st}}-1} |A_i|^2 (e^{-E_i \frac{t_s}{a}} + e^{-E_i (L_0^{\mathrm{eff}} - \frac{t_s}{a})}), \quad \text{(C3)}$$

and obtained their effective masses by solving $\frac{\cosh(E_i(L_0^{\mathrm{eff}}/2 - (t_s+a)/a))}{\cosh(E_i(L_0^{\mathrm{eff}}/2 - t_s/a))} = \frac{C_{2\mathrm{pt}}(t_s+a, P_3)}{C_{2\mathrm{pt}}(t_s, P_3)}$.

2. An issue with long-tailed distributions in the zero-momentum case

First, we discuss the case of zero spatial momentum $P_3 = 0$ which we found to be challenging within the stochastic approach of constructing trace along with the Fourier transform to real space. In the top panels of Fig. 4, we show the momentum space correlator $\tilde{C}(p_0, P_3)$ as a function of temporal momentum p_0; the top-left and top-right panels show the results at $P_3 = 0$ and

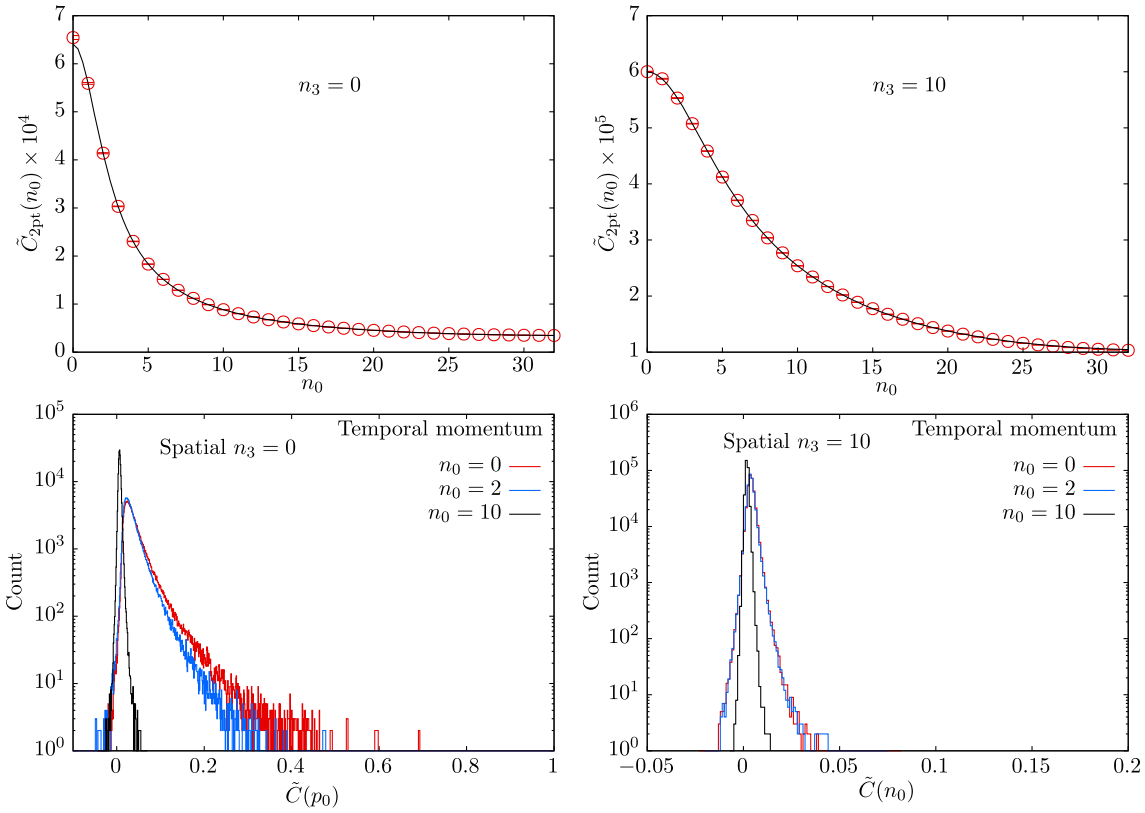

FIG. 4. Top: sample pion-pion two-point functions $\tilde{C}_{2\mathrm{pt}}(n_0; n_3)$ at spatial momenta $n_3 = 0$ (left) and $n_3 = 10$ (right) as a function of temporal momentum n_0. The black curves are fits to the momentum space correlator to Eq. (C1) with $N_{\mathrm{st}} = 3$ to data from $n_0 \in [1, 20]$. For $n_3 = 10$, the fit automatically passes through the $n_0 = 0$ data point. For $n_3 = 0$, the $n_0 = 0$ data point is slightly above the expectation. This causes a problem for $n_3 = 0$ real-space correlator construction and shows up as a pathological state with near zero mass. The possible issue is purely numerical and has to do with $n_3 = 0$, $n_0 = 0$ being the pion susceptibility that is difficult to evaluate stochastically due to the long-tailed nature of its Monte Carlo histogram. Bottom: Such Monte Carlo histograms for $\tilde{C}_{2\mathrm{pt}}(n_0, n_3)$ are shown in the bottom two panels for $n_3 = 0$ and $n_3 = 10$. Indeed the $n_3 = 0$, $n_0 = 0$ case is long tailed, and it might require even larger statistics to evaluate it robustly. The histograms immediately get narrower at nonzero n_3 and n_0 and thereby, do not cause any issues.

$P_3/(2\pi/136) = n_3 = 10$, respectively. The black curves are the best fit curves using Eq. (C1) truncated at $N_{\rm st} = 3$ and fitted over a range of $n_0 \in [1, 20]$. The fits work well with $\chi^2/{\rm dof} \approx 1$. However, we note that the curve for $P_3 = 0$ when extrapolated to $n_0 = 0$ is slightly, but in a statistically significant manner, below the actual stochastically evaluated data point for $\tilde{C}(p_0 = 0, P_3 = 0)$. Such a problem existed only in the $P_3 = 0$ case, and at other nonzero P_3 (such as $n_3 = 10$ case on the top right) the fitted curve automatically passed through $n_0 = 0$ data point as well. While the problem is easy to fix by avoiding the $n_0 = 0$ data point while performing fits, it causes a problem when reconstructing real-space correlators via Fourier transform; namely, if $n_0 = 0$ is not evaluated very accurately, then its effect is to add a spurious low-mass state into the real-space correlator. The origin of the problem is easy to understand. In the bottom left and bottom right panels of Fig. 4, we show the Monte Carlo histogram of the stochastic estimator in Eq. (B2) for the two spatial momenta. For each of them, we have shown the histograms at three values of n_0. We see that for $n_0 = 0, n_3 = 0$, which is nothing but pion susceptibility, the distribution is very long tailed, and hence, it is likely that the difficulty we are finding is due to the inability to robustly estimate the mean and the statistical error of such a long-tailed distribution. At nonzero n_0, the distribution gets narrower. Also, the distribution at $n_0 = 0$ gets narrower at nonzero n_3, and hence we were able to reconstruct real-space correlators well. Having understood the problem, we found the following procedure to correct the $n_0 = 0$ data points to solve the issue; we took the lowest three nonzero $n_0 = 1, 2, 3$ data points that are dominated by the ground-state E_0, and solved a system of equations,

$$\frac{A}{\cos(ap_0) - \cosh(aE_0)} + B = \tilde{C}_{\rm 2pt}(p_0, P_3), \quad \text{(C4)}$$

to find the unknown parameters A, B, E_0. Using them, we corrected the $n_0 = 0$ point with the estimated value $\frac{A}{1-\cosh(aE_0)} + B$. Using this corrected $n_0 = 0$ data point, we used Eq. (C2) to perform the Fourier transform and obtained the correlator as a function of t_s. With this procedure, the spurious low mode disappeared at $P_3 = 0$. At nonzero P_3, such a procedure did not have any significant effect at all.

3. Analysis of two-point functions in real space

In the left panel of Fig. 5, we show our "raw" data for $\tilde{C}_{\rm 2pt}(p_0, P_3)$ that we directly computed on the lattice as a function of p_0 for the nine different spatial P_3. As such, we find our determination of $\tilde{C}_{\rm 2pt}(p_0, P_3)$ to be smoothly varying in both p_0 and P_3, up to an issue noted above for $P_3 = 0$. The well-determined nature of $\tilde{C}$ in Fourier space is deceiving, as the long-distance exponential falloff in the Fourier transformed $C_{\rm 2pt}(t_s)$ comes from delicate cancellations between different $\tilde{C}_{\rm 2pt}(p_0)$, resulting in noise at larger t_s. This can be seen in the right panel of Fig. 5, where we show such an inferred real-space two-point function $C_{\rm 2pt}(t_s, P_3)$ using Eq. (C2). We have displayed the results at various spatial momenta as a function of t_s/a.

In the determination of the quasi-PDF matrix element, the spectral data of a two-point function does not enter in the summation type analysis we performed (as we discuss in the next section). We present our analysis of the spectral content of the two-point function now for the sake of completion. In Fig. 6, we show the effective mass $E_{\rm eff}(t_s)$

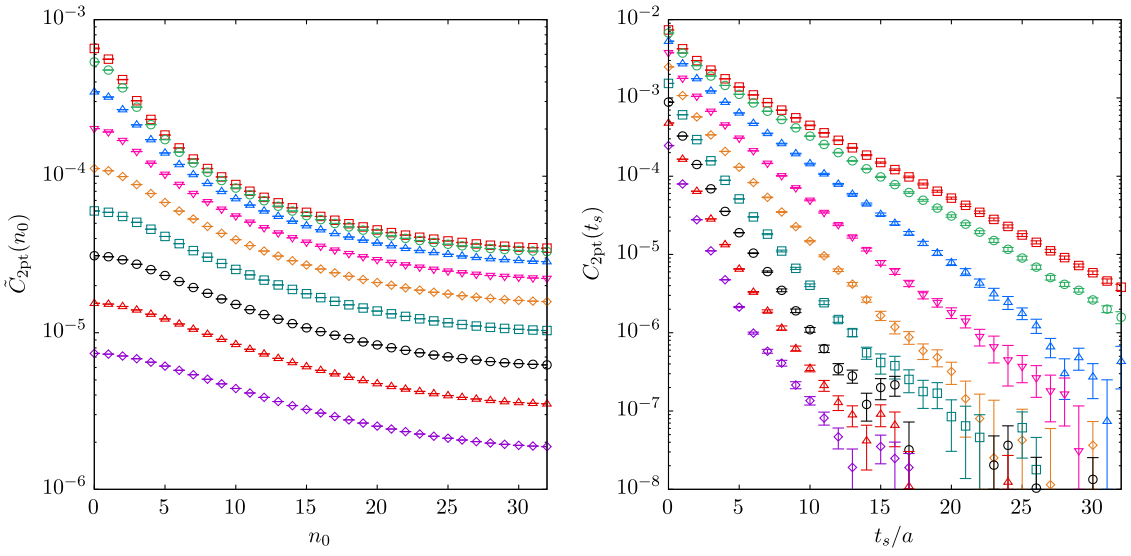

FIG. 5. Pion two-point function correlator as a function of temporal Euclidean momentum n_0 (left) and as a function of temporal separation t_s/a (right) as constructed from the momentum space correlations via Fourier transformation. The different symbols from top to bottom in the two panels are the data points at different spatial momentum along the z direction, n_3, from $n_3 = 0$ to 16 in steps of two.

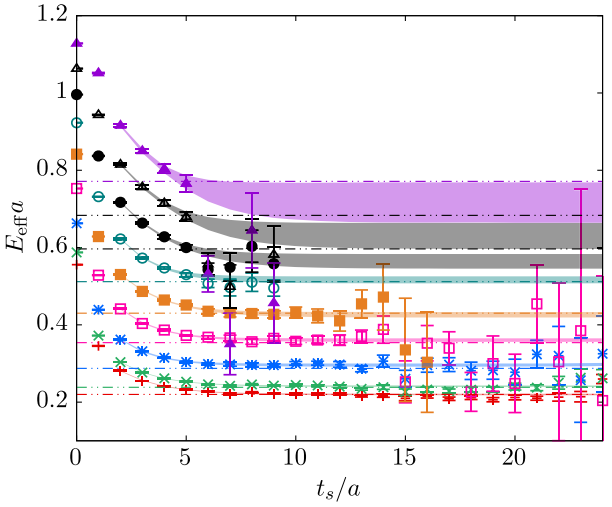

FIG. 6. The effective mass as determined from the real-space two-point function $C_{\rm 2pt}(t_s, P_3)$. The different symbols are the results at different spatial momenta $P_3 = 2\pi n_3/136$ from $n_3 = 0, 2, \ldots, 16$, from bottom to top. The bands are two-state fits to correlators. The horizontal lines are expectations from continuum single-particle dispersion.

determined from $C_{\rm 2pt}(t_s, P_3)$. We were able to perform stable two-state fits [$N_{\rm st} = 2$ in Eq. (C3)] to $C_{\rm 2pt}(t_s)$. For this, we used a fit range $t_s \in [2a, 15a]$ for $n_3 = 0, 2, 4, 6, 8$, $t_s \in [2a, 10a]$ for $n_3 = 10$, 12, and $t_s \in [2a, 6a]$ for $n_3 = 14$, 16. We used a smaller minimum of $2a$ so as to be sensitive to excited states, and at the same time make the fits stable. We changed the maximum range of t_s so as to avoid the noisier data points, as well as those that are not well determined after all the intricate cancellations in the Fourier transform from $\tilde{C}_{\rm 2pt}(p_0)$ resulting in orders of magnitude smaller values for $C_{\rm 2pt}(t_s)$ as seen in Fig. 5 (for example, the $t_s > 6a$ data points in Fig. 6 for $n_3 = 14$, 16 that are suddenly pulled to smaller values than expected, and it is clear that they are not well determined numerically and might need more precise data for $\tilde{C}$). In this way, we found the pion mass in our calculation to be $m_\pi a = 0.219(2)$ in lattice units. We show the resulting effective mass curves from the two-state fits at different momentum n_3 as the bands in Fig. 6. For comparison, the expected values for $E(P_3)$ from one-particle dispersion relation based on the value of $m_\pi a = 0.219$ are shown as the dot-dashed horizontal lines in the figure. We see that the resulting ground-state energies agree with the continuum dispersion within errors. As a curious observation that is unrelated to the results in the paper, we found the first excited state energy at $n_3 = 0$ to be $aE_1(P_3 = 0) = 0.79(4)$ from the two-state fits. Using a string-tension value of $\sqrt{\sigma} = 440$ MeV, we find this value to be about 1.3 GeV which seems to agree quite nicely with the pion radial excitation pole mass in SU(3) QCD. Thus, the usage of string tension to set the large-N_c GeV scale has its advantage as noted in the main text. However, we only found a poor agreement of the momentum dependence of $E_1(P_3)$ with a single particle dispersion curve and hence we cannot rule out the possibility of the agreement with pion (1300) at $P_3 = 0$ to be a numerical coincidence in our calculation, and the E_1 could simply be effectively capturing the tower of excited states.

APPENDIX D: DETERMINATION OF BARE PION QUASI-PDF MATRIX ELEMENTS

1. Construction of t_s-dependent three-point function

We used the stochastic estimator in Eq. (B4) to determine the three-point function in momentum space, $\tilde{C}_{\rm 3pt}(z, p, q)$. We found it computationally simpler and cheaper to fix the momentum insertion q and scan the entire set of p_0 for each choice of pion's spatial momentum P_3. We chose $q_0 = 0$ as this choice is the summation method as noted in the main text. Further, we set the spatial part of q also to zero, as we want the forward matrix element in this work for the case of PDF; thus $q = 0$. We reconstructed the t_s dependence of the three-point function as

$$C_{\rm 3pt}(z, t_s, P_3) = \sum_{n_0=0}^{L_0^{\rm eff}-1} \tilde{C}_{\rm 3pt}(z, p_0, P_3, q=0) e^{ip_0 \frac{t_s}{a}};$$
$$\tilde{C}_{\rm 3pt}(z, L_0^{\rm eff} - n_0, n_3, q=0) = -\tilde{C}_{\rm 3pt}(z, n_0, n_3, q=0). \quad \text{(D1)}$$

The second identity is simply due to the usage of γ_0 in the definition of a quasi-PDF operator, which makes the three-point function to be proportional to p_0, and hence, antisymmetric with respect to n_0 and $L_0^{\rm eff} - n_0$. As is usual, we used $z = (0, 0, 0, z_3)$ along the z axis. By using a folded Wilson line, we scanned $z_3/a \in [-16, 16]$. In the end, we only used $|z_3|/a \le 6$ so as to ensure the applicability of perturbation theory.

Let us make the connection of a three-point function with $q_0 = 0$ to the summation method obvious. We suppress the arguments for z, p in the two- and three-point functions for the sake of brevity, and both of them should be understood to be at the same p below. For the sake of argument, let $C_{\rm 3pt}(t_s, \tau)$ be the three-point function by Fourier transforming with respect to both p_0 and q_0; in that case, τ is the insertion time of the quasi-PDF operator τ, and it could be both within and outside the pion source and sink locations. For the case $\tau \le t_s$, we can do a spectral decomposition to get

$$C_{\rm 3pt}(t_s, \tau) = \sum_{i,j=0} A_i^* A_j \langle i|\mathcal{O}|j\rangle e^{-E_i(t_s-\tau)-E_j\tau}; \quad A_i = \langle 0|\pi^\dagger|E_i\rangle \quad \text{(D2)}$$

for $\tau \le t_s$. The sum within this region gives $\sum_{\tau=0}^{t_s} C_{\rm 3pt}(t_s, \tau) \sim |A_0|^2 \langle 0|\mathcal{O}|0\rangle t_s e^{-E_i t_s} + \text{const} \times e^{-E_0 t_s}$, up

to $O(e^{-E_1 t_s})$ excited state corrections. When the operator is "outside" the source and sink, $\tau > t_s$, then the terms which are not exponentially suppressed with the effective temporal extent, L_0^{eff}, are of the form

$$C_{3\text{pt}}(t_s,\tau) = \sum_{i,j}\langle 0|\mathcal{O}|i\rangle e^{-E_i\tau}e^{-E_j t_s}\langle i|\pi|j\rangle\langle j|\pi^\dagger|0\rangle, \quad \text{(D3)}$$

for $\tau > t_s$. Since the state j cannot be the ground-state pion, and $i \neq j$, the sum over $\tau > t_s$ cannot have the linear piece and will contribute simply as yet another exponentially suppressed excited state contribution. Therefore, we can sum over τ for all values from 0 to L_0^{eff}, and the linear piece in t_s gives the information on the ground-state matrix element. The sum is $\sum_{\tau=0}^{L_0^{\text{eff}}} C_{3\text{pt}}(t_s,\tau) = C_{3\text{pt}}(t_s, q_0 = 0)$. To cancel off the amplitudes A_i, we form the ratio

$$R(t_s) \equiv \frac{C_{3\text{pt}}(t_s, q_0 = 0)}{C_{2\text{pt}}(t_s)}. \quad \text{(D4)}$$

From the arguments above, we see that

$$R(t_s) = \langle E_0|\mathcal{O}|E_0\rangle t_s + C + O(e^{-(E_1-E_0)t_s}). \quad \text{(D5)}$$

By fitting the linear t_s dependence of $R(t_s)$, we obtained the bare quasi-PDF matrix element from the slope.

2. Extraction of ground-state bare quasi-PDF matrix elements

We fit the functional form

$$R(t_s; P_3, z_3) = t_s h^B(z_3, P_3) + C, \quad \text{(D6)}$$

to the summed ratio $R(t_s)$ at different z_3 and P_3, using $h^B(z_3, P_3)$ and C as fit parameters over ranges $t_s \in [t_s^{\min}, t_s^{\max}]$. At the precision allowed by our data, we restricted the value of $t_s^{\min} = 2a, 3a, 4a$, and finally used $t_s^{\min} = 3a$, which in physical units, $t_s^{\min}\sqrt{\sigma} = 0.76$, is on the verge of the typical nonperturbative mass-gapped scales.

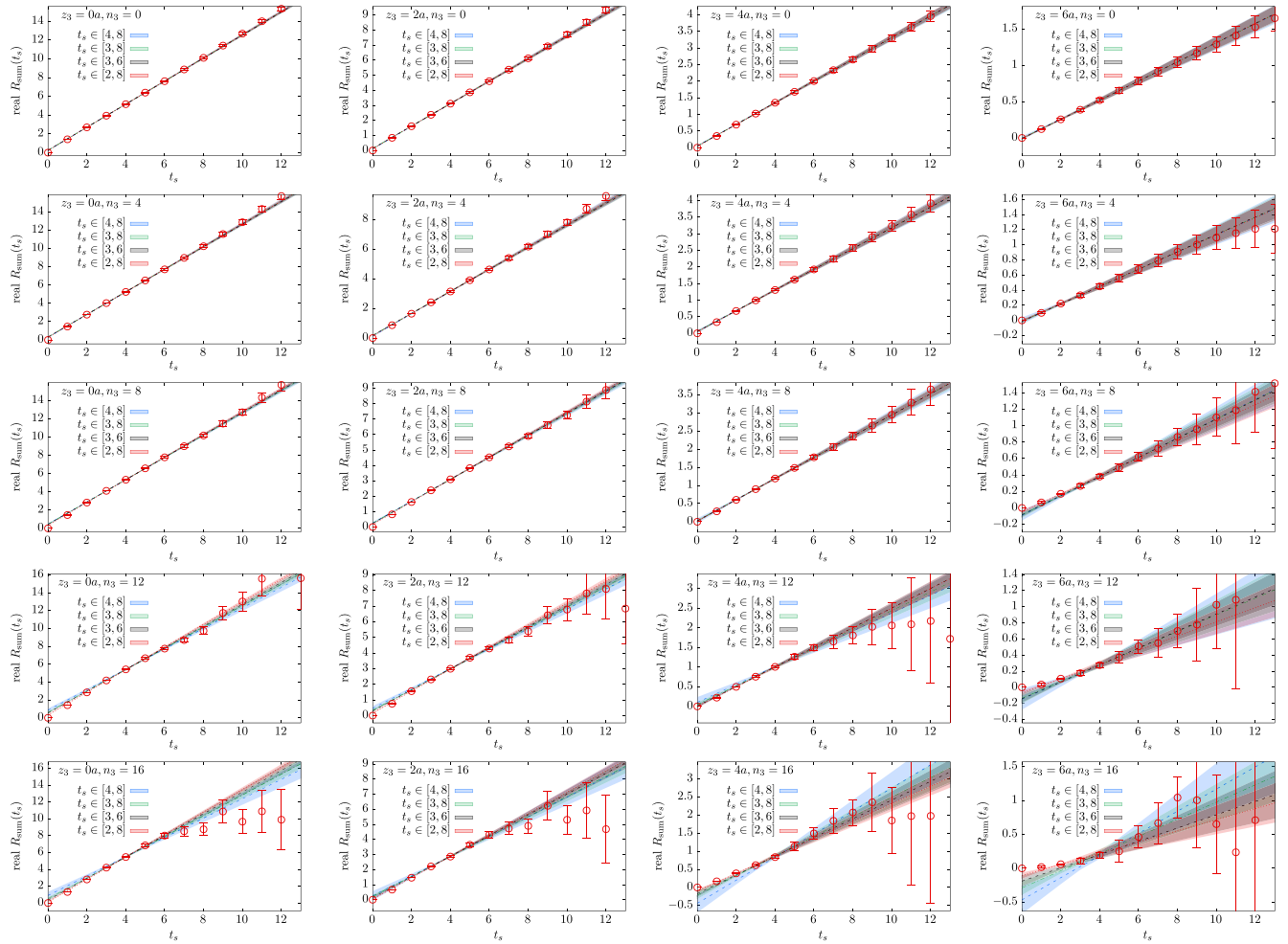

FIG. 7. Extraction of the real part of the ground-state bare quasi-PDF matrix element, $\text{Re}h^B(z_3, n_3)$, via summation method. The bands are straight line fits $h^B t_s + C$ to the data over different ranges of t_s that is specified in the legend. The different panels show the data and the fits from different n_3 (rows) and different z_3/a (columns).

By changing $t_s^{\min}$ to $2a$ and $4a$, we checked for the level of consistency as a means to test for the residual presence of excited state contributions. We used $t_s^{\max}$ between $6a$ and $8a$, and avoided points beyond $8a$ to not use the noisy as well as stochastically not-so-well-determined points at even larger $t_s^{\max}$. In the various panels of Figs. 7 and 8, we show the data points and fitted straight lines to determine $\mathrm{Re}h^B$ and $\mathrm{Im}h^B$, respectively. Each column in these figures shows the fits at $z_3 = 0, 2a, 4a, 6a$ at fixed momentum n_3. The different rows show them at $n_3 = 0$, 4, 8, 12, 16. In each panel, the points are our lattice determination of $R(t_s)$. We have shown the fits to Eq. (D6) for different fit ranges as the bands; the slopes of these lines are the needed values of h^B. Within the statistical errors, it is clear that the data nicely agrees with a linear t_s dependence in the ranges of t_s specified above. At the smaller n_3, where the data quality is better, we see that the fitted bands from the various ranges agree quite well. At larger $n_3 = 12$ to 16, the data quality for $t_s > 6a$ is quite poor, and the fits that start from $t_s = 4a$ and include $t_s > 6a$ data points behave quite differently. Therefore, for $n_3 \geq 12$, we restricted $t_s^{\max} = 6a$.

In Figs. 9 and 10, we show the resulting z_3 dependence of the quasi-PDF matrix element $h^B(z_3, P_3)$ from the summation-type fits over the different t_s ranges. We have slightly displaced the different estimations for clarity. We see that the estimations using $t_s \in [3a, 8a]$ are quite consistent with those using $[4a, 8a]$ for the momenta $n_3 \leq 10$. Therefore, we used the estimated values of h^B from $t_s \in [3a, 8a]$ in the main text. For the higher momenta, as we noted above, we see that estimations using $t_s^{\max} > 6a$ are not reliable. Within the larger statistical errors at the higher momenta $n_3 \geq 12$, we find the $t_s \in [3a, 6a]$ estimates are consistent with the shorter $t_s \in [2a, 6a]$ estimates, and also within the larger errors of the $t_s \in [4a, 8a]$ estimates which are biased with the poorly determined data beyond $t_s > 6a$. Therefore, we chose the range containing $t_s \in [3a, 6a]$ for the set of momenta $n_3 \geq 12$.

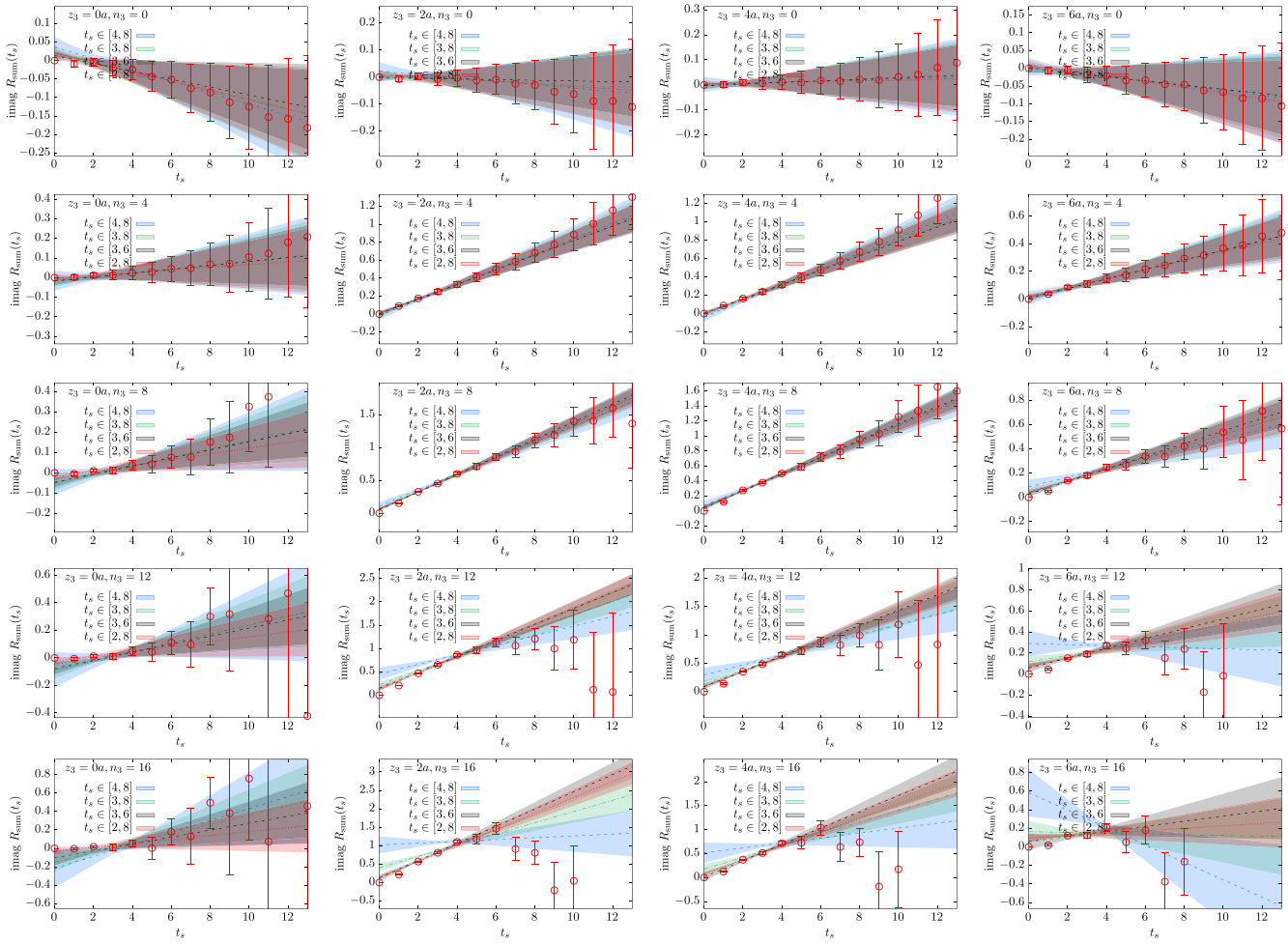

FIG. 8. Extraction of the imaginary part of the ground-state bare quasi-PDF matrix element, $\mathrm{Im}h^B(z_3, n_3)$, via summation method. The bands are straight line fits $h^B t_s + B$ to the data over different ranges of t_s. The different panels show the data and the fits from different n_3 (rows) and different z_3/a (columns).

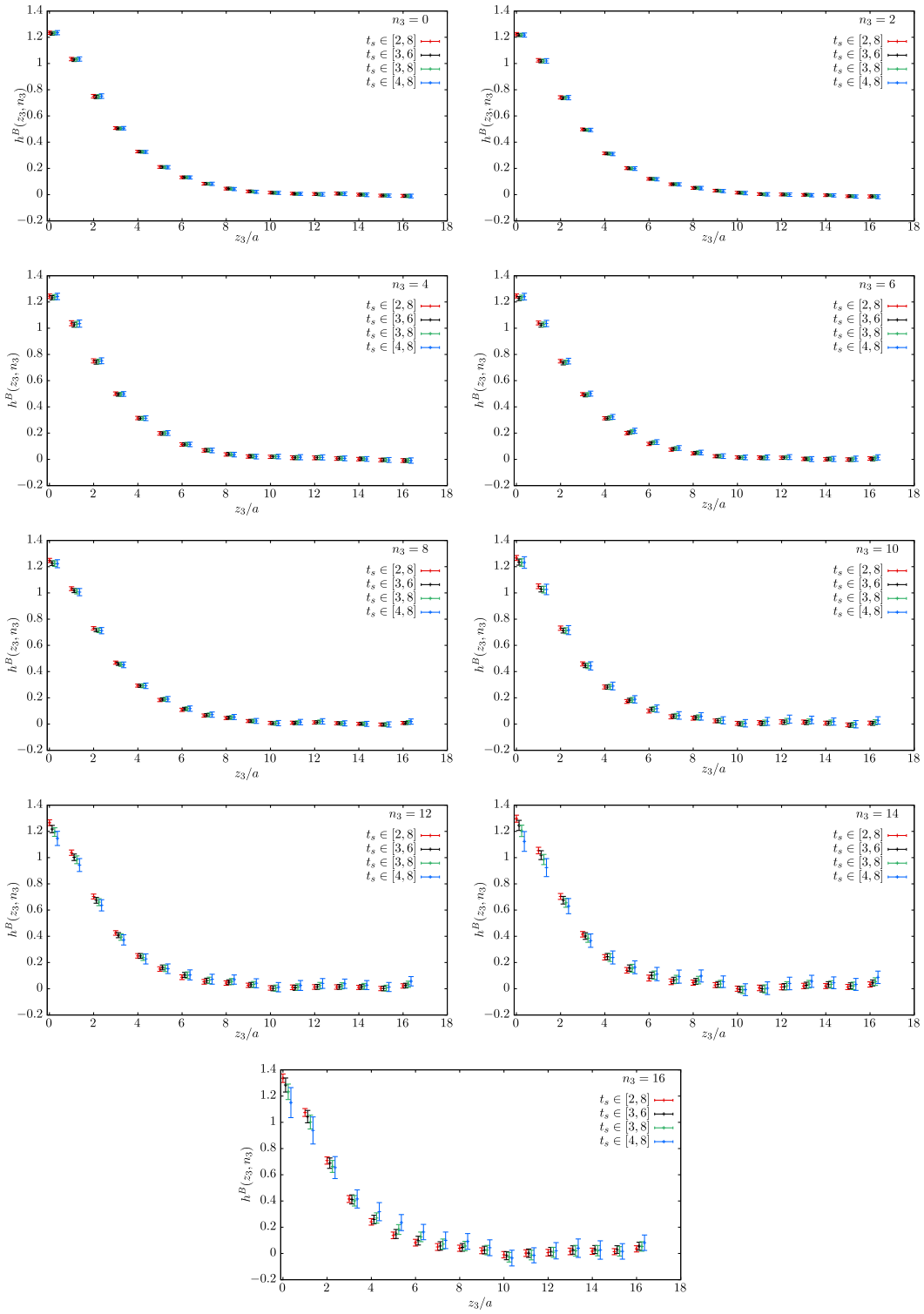

FIG. 9. The real part of the bare quasi-PDF matrix element $h^B(z_3, n_3)$ is shown as a function of z_3/a at different spatial momentum $\propto n_3$ used in this work as separate panels. The extrapolated results from summation methods over different fit ranges in t_s are shown together in the plots. We used extrapolations from $t_s \in [3a, 8a]$ for $n_3 \in [0, 10]$ and $t_s \in [3a, 6a]$ for $n_3 \in [12, 16]$ to avoid badly determined points beyond $t_s \geq 6a$. Using ranges with even larger minimum t_s was not feasible and forms a limitation of this work.

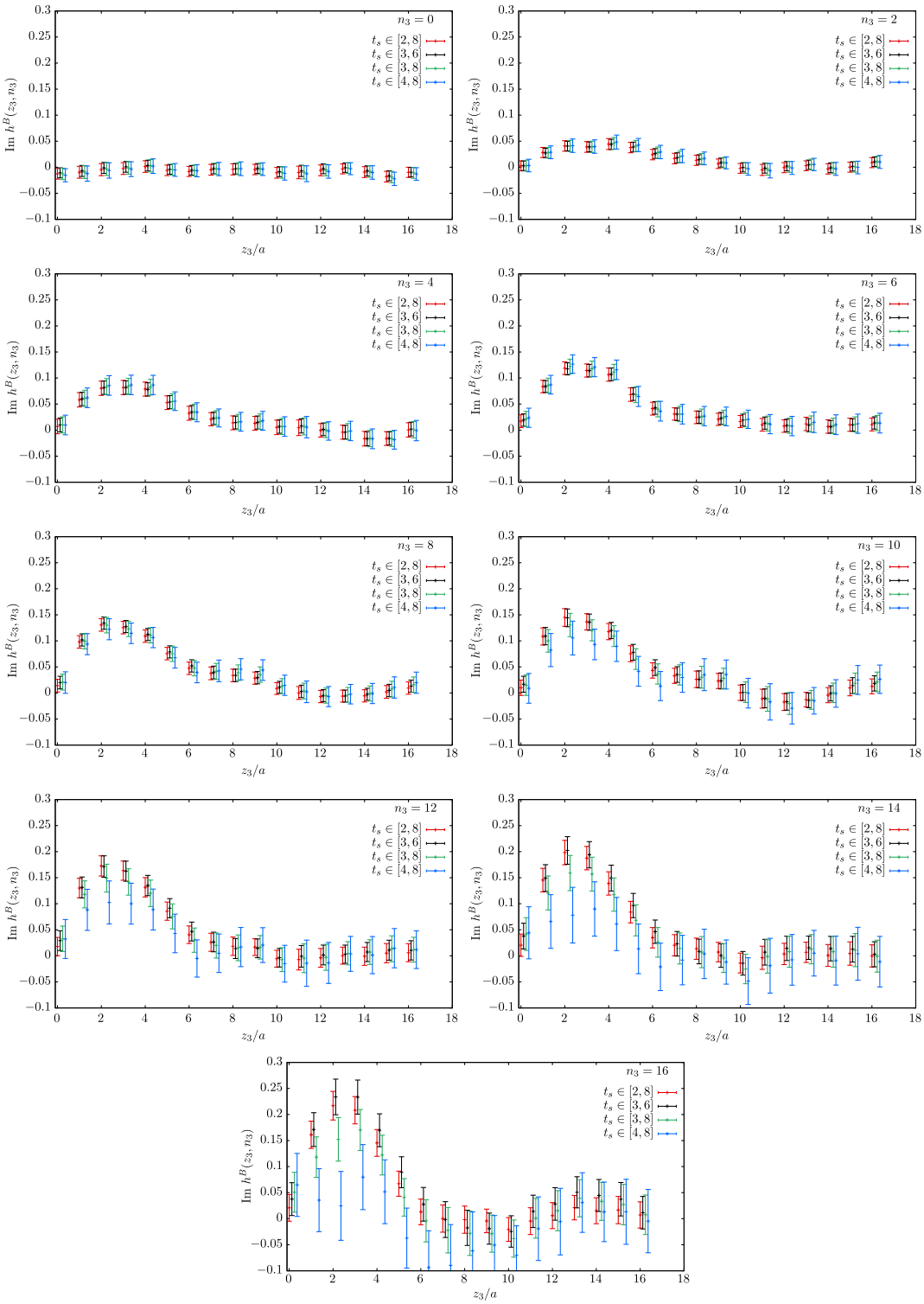

FIG. 10. The imaginary part of the bare quasi-PDF matrix element $h^B(z_3, n_3)$ is shown as a function of z_3/a at different n_3 used in this work as separate panels. The description is similar to Fig. 9.

As a cross-check, we present the values of $h^B(z_3 = 0, P_3)$ as a function of P_3 in Fig. 11. The $z_3 = 0$ matrix element is nothing but the pion matrix element of the local vector current operator, and hence measures the inverse of the vector current renormalization factor, Z_V. If the extraction of matrix elements is done correctly and there is no momentum-dependent lattice corrections, we should not find any P_3 dependence in Z_V. Indeed, we find that to be the case in Fig. 11 up to statistical errors.

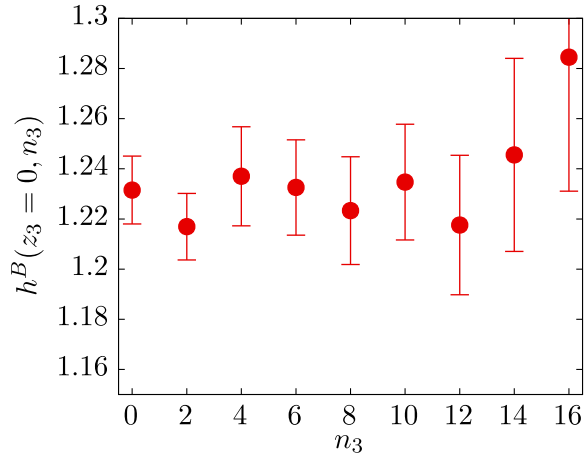

FIG. 11. A cross-check on the near-constant behavior the extracted bare matrix element of local current operator $h^B(z_3 = 0, n_3)$ as a function of momentum n_3 in lattice units. The bare matrix element is an estimate of the inverse of the vector current renormalization constant Z_V.

APPENDIX E: IMPLEMENTATION OF LEADING-TWIST OPE AND CONSTRUCTION OF $\overline{\text{MS}}$ ITD

We implemented the leading-twist OPE using the truncated form of Eq. (3) written explicitly as

$$\begin{aligned}\text{Re}\mathcal{M}(\nu, z_3^2) &= 1 + \sum_{n=1}^{N_{\max}} \frac{(-1)^n \nu^{2n}}{(2n)!} C_{2n}(z_3^2\mu^2)\langle x^{2n}\rangle_{u-\bar{u}},\\ \text{Im}\mathcal{M}(\nu, z_3^2) &= \sum_{n=1}^{N_{\max}} \frac{-(-1)^n \nu^{2n-1}}{(2n-1)!} C_{2n-1}(z_3^2\mu^2)\langle x^{2n-1}\rangle_{u+\bar{u}}.\end{aligned} \tag{E1}$$

We used the truncation as $N_{\max} = 4$ to fit the data up to $\nu = 3.5$, and we checked that the results do not change well within errors when $N_{\max}$ is changed from 3 to 4. The Wilson coefficients C_n are the isovector quark coefficients, usually written explicitly as C_n^{qq}. For the imaginary part, which is not a isovector quantity, one would have to include the corresponding C_n^{qq} and C_n^{qg} which will cause mixing with quark and gluon PDFs [80]; here, in the large-N_c limit, the C_n^{qg} which are proportional to $\alpha_s T_F(N_c)$ are $1/N_c$ suppressed and hence, we have considered only the C_n^{qq} Wilson coefficients above and in the main text.

Using the above OPE, we performed combined fits to the z_3 and P_3 dependencies of the lattice data. We performed two types of fits in the main text:

(1) Moments fit: here, we used the Mellin moments $\langle x^n\rangle$ entering the OPE as the free fit parameters. Since, we are assuming no functional form for the x dependence of the PDF, we referred to these types of fits as the model-independent fit analysis. Assuming the positivity of the underlying $u + \bar{u}$ and $u - \bar{u}$ PDFs help impose additional constraints [30] on their Mellin moments. We implemented such inequalities using a change of variable from moments to λ_i,

$$\langle x^{2n}\rangle_{u-\bar{u}} = \sum_{i=n}^{N_{\max}} \sum_{j=i}^{N_{\max}} e^{-\lambda_j}, \tag{E2}$$

and similarly for odd moments $\langle x^{2n-1}\rangle_{u+\bar{u}}$.

(2) PDF ansatz fit: Here, we assumed a global fit analysis inspired ansatz for the x dependence of the valence PDF, $f_{u-\bar{u}}(x) = \mathcal{N} x^\alpha (1-x)^\beta (1 + sx^2)$, with $\int_0^1 f_{u-\bar{u}}(x)dx = 1$. In practice, it results in Mellin moments $\langle x^{2n}\rangle_{u-\bar{u}}(\alpha, \beta, s)$, that in turn enter Eq. (E1). We fit the parameters α, β, and s in this manner. We imposed a prior $\alpha \in [-0.4, -0.6]$ based on Regge intercept expectation for a valence PDF. We did not perform an equivalent analysis for $u + \bar{u}$, as it was not clear if we should assume it to be a combination of ansatz for valence PDF and sea-quark PDF, and what prior to impose on small-x behavior of sea-quarks in large-N_c theory. Therefore, we avoided such issues here by performing only moments fit to the $u + \bar{u}$ case.

In the main text, we constructed the $\overline{\text{MS}}$ ITD at $\mu = 2$ GeV based on the analysis of pseudo-ITD lattice data above. These ITDs are defined as

$$\begin{aligned}\mathcal{M}^{\overline{\text{MS}}}_{u-\bar{u}}(\nu, \mu) &\equiv \int_0^1 f_{u-\bar{u}}(x, \mu)\cos(x\nu)dx;\\ \mathcal{M}^{\overline{\text{MS}}}_{u+\bar{u}}(\nu, \mu) &\equiv \int_0^1 f_{u+\bar{u}}(x, \mu)\sin(x\nu)dx.\end{aligned} \tag{E3}$$

In practice, the construction of $\overline{\text{MS}}$ ITD using the limited range of ν is simplified into a truncated series in ν as

$$\begin{aligned}\mathcal{M}^{\overline{\text{MS}}}_{u-\bar{u}}(\nu, \mu) &= 1 + \sum_{n=1}^{N_{\max}} \frac{(-1)^n \nu^{2n}}{(2n)!}\langle x^{2n}\rangle_{u-\bar{u}}(\mu),\\ \mathcal{M}^{\overline{\text{MS}}}_{u+\bar{u}}(\nu, \mu) &= \sum_{n=1}^{N_{\max}} \frac{-(-1)^n \nu^{2n-1}}{(2n-1)!}\langle x^{2n-1}\rangle_{u+\bar{u}}(\mu),\end{aligned} \tag{E4}$$

using the best fit estimates of the Mellin moments from the leading-twist analysis.

APPENDIX F: DETAILS REGARDING THE PERTURBATIVE ASPECTS

1. Coupling constant and Wilson coefficients

We borrowed various existing perturbative results computed for general N_c, and we simply used the large-N_c values of the color factors, $C_F(N_c) \to N_c/2$, $C_A(N_c) \to N_c$ and $T_F(N_c) \to 1/2$ in those expressions. In the absence of a nonperturbative running of the large-N_c $\overline{\text{MS}}$ coupling,

we simply used the LO 't Hooft coupling in the large-N_c limit,

$$\lambda(\mu) \equiv \lim_{N_c\to\infty} \alpha_s(\mu) N_c = \frac{1}{\frac{11}{12\pi}\ln(\frac{\mu^2}{\Lambda^2_{\overline{\rm MS}}})}. \tag{F1}$$

With $\sqrt{\sigma} = 0.44$ GeV to set the scale, we used $\Lambda_{\overline{\rm MS}} = 0.22$. At $\mu = 2$ GeV, we get $\lambda(2\text{ GeV}) = 0.778$. For the Wilson coefficients that enter the leading-twist OPE in Eq. (3), we used the one-loop expressions in Ref [49] with the replacement $C_F(N_c)\alpha_s(\mu) \to \lambda(\mu)/2 = 0.389$. The Wilson coefficients for the $u+\bar{u}$ PDF would differ in the SU(3) QCD due to it being a flavor singlet quantity. In the large-N_c limit, such differences due to the mixing terms [$\propto T_F(N_c)$] are subleading in $1/N_c$, and hence, we simply used the nonsinglet Wilson coefficients C_n for odd values of n.

2. Large-N_c LO DGLAP evolution

In the main text, we checked whether a universal initial condition at a low-factorization scale μ_0 could explain the observed differences between SU(3) QCD and in large-N_c theory. For this, we performed the DGLAP evolution of large-N_c PDF (or equivalently its ITD) from scale μ to a lower scale μ_0, that is then used as an initial condition for 3 flavor SU(3) QCD evolution back to scale μ using corresponding DGLAP evolution in Mellin space. That is, taking

$$X_n^{(N_c)}(\mu) = [2\langle x^n\rangle_{u-\bar{u}}(\mu), 2\langle x^n\rangle_{u+\bar{u}}(\mu), \langle x^n\rangle_g(\mu)] \tag{F2}$$

as the array of quark and gluon moments in $SU(N_c)$ QCD, we evolved them as

$$X_n^{(N_c)}(\mu') = \begin{pmatrix} P_{\rm NS}^{qq,N_c}(n,\mu,\mu') & 0 & 0 \\ 0 & P_{\rm S}^{qq,N_c}(n,\mu,\mu') & P_{\rm S}^{gq,N_c}(n,\mu,\mu') \\ 0 & P_{\rm S}^{qg,N_c}(n,\mu,\mu') & P_{\rm S}^{gg,N_c}(n,\mu,\mu') \end{pmatrix} \cdot X_n^{(N_c)}(\mu), \tag{F3}$$

where $P^{ij,N_c}(n,\mu,\mu')$ are the $SU(N_c)$ theory DGLAP factors from parton species i to species j in Mellin space (e.g., textbook such as [81]) that evolve the moments from scale μ to μ'. The subscript S and NS specify singlet and nonsinglet respectively. In this paper, we used a LO DGLAP evolution, at which order $P^{qq}_{NS} = P^{qq}_S$. At LO, the evolution depends on μ only via the logarithms $\frac{\ln(\mu/\Lambda_{\overline{\rm MS}})}{\ln(\mu_0/\Lambda_{\overline{\rm MS}})}$. Since we were only interested in capturing the qualitative behavior of $u+\bar{u}$ ITD in the large-N_c theory and $SU(3)$ theory, we simply used $\Lambda_{\overline{\rm MS}} = 0.22$ GeV in the DGLAP factors of both the theories. One should note that the ratios of twist-2 operator anomalous dimensions to β-function coefficient, $\gamma_n^{(1)}/\beta_0$ have a finite limit when $N_c \to \infty$. In the $N_c \to \infty$ limit, the cross term $P_{\rm S}^{gq,N_c} \to 0$, and hence $u+\bar{u}$ evolves without mixing with the gluon. On the other hand, the term $P_{\rm S}^{qg,\infty}$ is nonzero as gluon radiation from a quark line is still a leading process in N_c counting.

In this work, we only computed the quark moments $\langle x^n\rangle_{u+\bar{u}}$ in the large-N_c theory, and we did not explicitly compute the gluon PDF in the large-N_c pion. Therefore, we deduced the leading moment $\langle x\rangle_g = 1\text{–}2\langle x\rangle_{u+\bar{u}}$ from the momentum sum rule. Since we expect the gluon PDF to be contributing dominantly in the small-x region, we assumed that the next moment $\langle x^3\rangle_g$ (and all other higher odd moments) can be neglected. With these inputs from the large-N_c theory, we followed the chain of evolution,

$$X_n^{(\infty)}(\mu) \to X_n^{(\infty)}(\mu_0) \to X_n^{(3)}(\mu). \tag{F4}$$

Using such an expectation $X_n^{(3)}(\mu)$ for SU(3) QCD moments at $\mu = 2$ GeV based on the above evolution, we constructed the corresponding $\overline{\rm MS}$ ITD by using Eq. (E4).

APPENDIX G: EFFICIENCY OF ONE-LOOP LARGE-N_c LEADING-TWIST OPE

The leading-twist OPE can be applied to the lattice data at fixed values of z_3 [58], so as to capture the ν dependence coming only via variation in the momentum P_3. Such an application has been found [30,59] to be a nice diagnostic of the effectiveness of perturbative as well as leading-twist frameworks in a region of z_3, and as way to detect corrections to the framework. In the left panel of Fig. 12, we show such a z_3-dependent leading nontrivial moment $\langle x\rangle$, from the analysis of ${\rm Im}\mathcal{M}(\nu, z_3^2)$. In the right panel, we show a similar z_3 dependence of $\langle x^2\rangle$ from ${\rm Re}\mathcal{M}(\nu, z_3^2)$. We used $\mu = 2$ GeV in the scale set by $\sqrt{\sigma} = 0.44$ GeV as explained above. The red filled circles are the results using one-loop Wilson coefficients in the OPE. If one loop is sufficient, and if there are no higher-twist corrections to the OPE and z_3-dependent lattice spacing corrections to the continuum OPE, then one should observe a plateau in the moments as a function of z_3. In the range $z_3 \in [2a, 5a]$ that we used, we see an approximate plateau in the one-loop results in the two panels. Our determinations of the two moments via combined fits to the entire data in the range of $z_3 \in [2a, 5a]$ is shown as the gray bands, which are consistent with the plateau in the data. We skipped the $z_3 = a$ point to be cautious of avoiding any lattice corrections at those separations. Given the quality of our data, we did not add any lattice spacing and higher-twist corrections by hand to the leading-twist continuum OPE. To see the effect of one-loop evolution in z_3 effected by the Wilson coefficients, we also plot the results using tree-level (i.e., set $\alpha_s = 0$) in the two panels in Fig. 12. The effect of one loop is rather small in comparison with typical statistical errors, but it is quite pronounced at shorter $z_3 = 1a - 3a$.

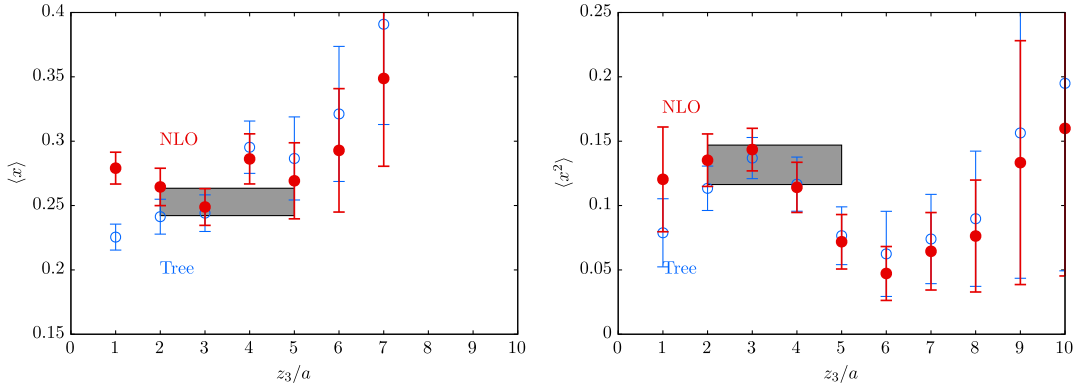

FIG. 12. Fixed-z^2 analysis of pseudo-ITD $\mathcal{M}(\nu, z^2)$ by fitting the leading-twist OPE to the $P_3 z_3$ dependence at different values of z_3 using the first few Mellin moments as fit parameters. The left and right panels show the resulting z_3 dependent $\langle x \rangle_{u+\bar{u}}$ and $\langle x^2 \rangle_{u-\bar{u}}$ moments, respectively. The filled red circles are the result of performing such an analysis using one-loop large-N_c Wilson coefficients. The open circles are obtained by setting $C_n = 1$, that is, to their tree-level values. Our estimates of those moments based on a combined fit to both z^2 and ν dependencies of $\mathcal{M}$ data in the range $z_3 \in [2a, 5a]$ are shown as gray bands.

[1] G. 't Hooft, Nucl. Phys. **B72**, 461 (1974).
[2] G. 't Hooft, Nucl. Phys. **B75**, 461 (1974).
[3] M. G. Pérez, A. González-Arroyo, and M. Okawa, J. High Energy Phys. 04 (2021) 230.
[4] T. DeGrand and Y. Liu, Phys. Rev. D **94**, 034506 (2016); **95**, 019902(E) (2017).
[5] P. Hernández, C. Pena, and F. Romero-López, Eur. Phys. J. C **79**, 865 (2019).
[6] G. S. Bali, F. Bursa, L. Castagnini, S. Collins, L. Del Debbio, B. Lucini, and M. Panero, J. High Energy Phys. 06 (2013) 071.
[7] G. S. Bali and F. Bursa, J. High Energy Phys. 09 (2008) 110.
[8] L. Del Debbio, B. Lucini, A. Patella, and C. Pica, J. High Energy Phys. 03 (2008) 062.
[9] R. Abdul Khalek *et al.*, arXiv:2203.13199.
[10] A. Accardi *et al.*, Eur. Phys. J. A **52**, 268 (2016).
[11] J. Dudek *et al.*, Eur. Phys. J. A **48**, 187 (2012).
[12] A. H. Mueller, Nucl. Phys. **B415**, 373 (1994).
[13] E. C. Poggio, H. R. Quinn, and S. Weinberg, Phys. Rev. D **13**, 1958 (1976).
[14] M. A. Shifman, in *Proceedings of the 8th ISHFP* (World Scientific, Singapore, 2000), Vol. 3, pp. 1447–1494, 10.1142/9789812810458_0032.
[15] E. Witten, Nucl. Phys. **B160**, 57 (1979).
[16] T. H. R. Skyrme, Nucl. Phys. **31**, 556 (1962).
[17] E. Witten, Nucl. Phys. **B223**, 433 (1983).
[18] G. S. Adkins, C. R. Nappi, and E. Witten, Nucl. Phys. **B228**, 552 (1983).
[19] D. Diakonov, V. Y. Petrov, and P. V. Pobylitsa, Nucl. Phys. **B306**, 809 (1988).
[20] D. Diakonov, V. Petrov, P. Pobylitsa, M. V. Polyakov, and C. Weiss, Nucl. Phys. **B480**, 341 (1996).
[21] D. Diakonov, V. Y. Petrov, P. V. Pobylitsa, M. V. Polyakov, and C. Weiss, Phys. Rev. D **56**, 4069 (1997).
[22] H. Weigel, L. P. Gamberg, and H. Reinhardt, Mod. Phys. Lett. A **11**, 3021 (1996).
[23] L. P. Gamberg, H. Reinhardt, and H. Weigel, Phys. Rev. D **58**, 054014 (1998).
[24] A. C. Aguilar *et al.*, Eur. Phys. J. A **55**, 190 (2019).
[25] B. Adams *et al.*, arXiv:1808.00848.
[26] C. D. Roberts, D. G. Richards, T. Horn, and L. Chang, Prog. Part. Nucl. Phys. **120**, 103883 (2021).
[27] R. S. Sufian, J. Karpie, C. Egerer, K. Orginos, J.-W. Qiu, and D. G. Richards, Phys. Rev. D **99**, 074507 (2019).
[28] 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).
[29] T. Izubuchi, L. Jin, C. Kallidonis, N. Karthik, S. Mukherjee, P. Petreczky, C. Shugert, and S. Syritsyn, Phys. Rev. D **100**, 034516 (2019).
[30] 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).
[31] X. Gao, N. Karthik, S. Mukherjee, P. Petreczky, S. Syritsyn, and Y. Zhao, Phys. Rev. D **103**, 094510 (2021).
[32] N. Karthik, Phys. Rev. D **103**, 074512 (2021).
[33] H.-W. Lin, J.-W. Chen, Z. Fan, J.-H. Zhang, and R. Zhang, Phys. Rev. D **103**, 014516 (2021).
[34] X. Gao, A. D. Hanlon, S. Mukherjee, P. Petreczky, P. Scior, S. Syritsyn, and Y. Zhao, Phys. Rev. Lett. **128**, 142003 (2022).
[35] W. Detmold, A. V. Grebe, I. Kanamori, C. J. D. Lin, S. Mondal, R. J. Perry, and Y. Zhao (HOPE Collaboration), Phys. Rev. D **105**, 034506 (2022).
[36] J. Hua *et al.*, arXiv:2201.09173.
[37] E. Ruiz Arriola and W. Broniowski, Phys. Rev. D **74**, 034008 (2006).
[38] V. Braun and D. Müller, Eur. Phys. J. C **55**, 349 (2008).
[39] X. Ji, Phys. Rev. Lett. **110**, 262002 (2013).
[40] A. V. Radyushkin, Phys. Rev. D **96**, 034025 (2017).
[41] Y.-Q. Ma and J.-W. Qiu, Phys. Rev. D **98**, 074021 (2018).
[42] D. J. Gross and E. Witten, Phys. Rev. D **21**, 446 (1980).
[43] T. Eguchi and H. Kawai, Phys. Rev. Lett. **48**, 1063 (1982).

[44] R. Narayanan, Acta Phys. Pol. B **40**, 3231 (2009), arXiv:0910.3711.
[45] R. Narayanan and H. Neuberger, Proc. Sci. LATTICE2007 (2007) 020 [arXiv:0710.0098].
[46] R. Narayanan, H. Neuberger, and F. Reynoso, Phys. Lett. B **651**, 246 (2007).
[47] R. Narayanan and H. Neuberger, Phys. Rev. Lett. **91**, 081601 (2003).
[48] J. Kiskis, R. Narayanan, and H. Neuberger, Phys. Lett. B **574**, 65 (2003).
[49] T. Izubuchi, X. Ji, L. Jin, I. W. Stewart, and Y. Zhao, Phys. Rev. D **98**, 056004 (2018).
[50] L. Maiani, G. Martinelli, M. L. Paciello, and B. Taglienti, Nucl. Phys. **B293**, 420 (1987).
[51] J. Kiskis and R. Narayanan, Phys. Lett. B **681**, 372 (2009).
[52] B. Lucini, M. Teper, and U. Wenger, J. High Energy Phys. 01 (2004) 061.
[53] T. Ishikawa, Y.-Q. Ma, J.-W. Qiu, and S. Yoshida, Phys. Rev. D **96**, 094019 (2017).
[54] X. Ji, J.-H. Zhang, and Y. Zhao, Phys. Rev. Lett. **120**, 112001 (2018).
[55] K. Orginos, A. Radyushkin, J. Karpie, and S. Zafeiropoulos, Phys. Rev. D **96**, 094503 (2017).
[56] C. Allton, M. Teper, and A. Trivini, J. High Energy Phys. 07 (2008) 021.
[57] S. Datta and S. Gupta, Proc. Sci. LAT2009 (**2009**) 178 [arXiv:0910.2889].
[58] J. Karpie, K. Orginos, and S. Zafeiropoulos, J. High Energy Phys. 11 (2018) 178.
[59] C. Egerer *et al.* (HadStruc Collaboration), Phys. Rev. D **105**, 034507 (2022).
[60] P. C. Barry, C.-R. Ji, N. Sato, and W. Melnitchouk (Jefferson Lab Angular Momentum (JAM) Collaboration), Phys. Rev. Lett. **127**, 232001 (2021).
[61] P. C. Barry, N. Sato, W. Melnitchouk, and C.-R. Ji, Phys. Rev. Lett. **121**, 152001 (2018).
[62] M. J. Teper, arXiv:hep-lat/9711011.
[63] A. González-Arroyo and M. Okawa, Phys. Lett. B **755**, 132 (2016).
[64] I. Bars and M. B. Green, Phys. Rev. D **17**, 537 (1978).
[65] Y. Jia, S. Liang, L. Li, and X. Xiong, J. High Energy Phys. 11 (2017) 151.
[66] Y. Jia, S. Liang, X. Xiong, and R. Yu, Phys. Rev. D **98**, 054011 (2018).
[67] M. Burkardt, Phys. Rev. D **62**, 094003 (2000).
[68] E. Witten, Nucl. Phys. **B149**, 285 (1979).
[69] M. J. Teper, Z. Phys. C **5**, 233 (1980).
[70] B. Lucini and M. Teper, J. High Energy Phys. 06 (2001) 050.
[71] P. Nason and M. Palassini, Nucl. Phys. **B444**, 310 (1995).
[72] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, Nucl. Phys. **B147**, 385 (1979).
[73] N. Andrei and D. J. Gross, Phys. Rev. D **18**, 468 (1978).
[74] A. Holl, A. Krassnigg, and C. D. Roberts, Phys. Rev. C **70**, 042203 (2004).
[75] https://www.wm.edu/it/rc.
[76] C. Morningstar and M. J. Peardon, Phys. Rev. D **69**, 054501 (2004).
[77] A. Frommer, V. Hannemann, B. Nockel, T. Lippert, and K. Schilling, Int. J. Mod. Phys. C **05**, 1073 (1994).
[78] S. Gusken, U. Low, K. H. Mutter, R. Sommer, A. Patel, and K. Schilling, Phys. Lett. B **227**, 266 (1989).
[79] G. S. Bali, B. Lang, B. U. Musch, and A. Schäfer, Phys. Rev. D **93**, 094515 (2016).
[80] W. Wang, S. Zhao, and R. Zhu, Eur. Phys. J. C **78**, 147 (2018).
[81] R. G. Roberts, *The Structure of the Proton: Deep Inelastic Scattering*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, England, 1994).