Introduction: Parton distribution functions (PDFs) are important quantities describing the probability densities of quarks and gluons within hadrons. They are not only fundamental properties of quantum chromodynamics (QCD), but also are key inputs to predict cross sections in high-energy scattering experiments [1]. Calculating the $x$-dependence of PDFs from first principles has long been a holy grail in nuclear and high-energy physics. In modern parton physics, the PDFs are defined from the lightcone correlations of quarks and gluons in the hadron, so they involve strong infrared (IR) dynamics and can only be solved by nonperturbative methods such as lattice QCD. However, the direct calculation of PDFs on a Euclidean lattice has been extremely difficult because the real-time dependence of the lightcone makes it infeasible to extract them from lattice simulations with imaginary time. Early studies based on the operator product expansion (OPE) can only access the lower moments of the PDF [2,3,4] from lattice QCD. A similar situation also occurs in lattice calculations of other parton observables, such as the distribution amplitudes (DAs) and generalized parton distributions (GPDs).

In recent years, it was proposed [5,6] that the PDFs and other parton observables can be directly extracted from the lattice by calculating the matrix elements of certain static operators in a boosted hadron state. For the unpolarized quark PDF, the static operator is $O_{\Gamma}(z) = \psi(z)U(z, 0)\bar{\psi}(0)$ where the space-like Wilson line $U(z, 0) = P\exp(-ig \int_0^z dx'A_z(x'))$, and $\Gamma = \gamma_t$ or $\gamma_t$ so that under the infinite Lorentz boost along the $z$-axis $O_{\Gamma}$ approaches the lightcone correlation operator which defines the PDF. The hadron matrix element of $O_{\Gamma}(z)$ can be directly obtained from lattice QCD and its Fourier transformation is known as the quasi-PDF:

$$\tilde{q}_{\Gamma}(x, P_z, \mu) = \int_{-\infty}^{\infty} \frac{dz}{2\pi} e^{ixP_zz} \langle P | O_{\Gamma}(z) | P \rangle . \quad (1)$$

The quasi-PDF is related to the lightcone PDF through a factorization theorem, where the former can be factorized into a perturbative matching coefficient and the latter, up to power corrections suppressed by the nucleon momentum. This factorization theorem is founded in the large-momentum effective theory (LaMET) [6,10], where the matching coefficient can be calculated exactly in perturbation theory.

Since the proposal of LaMET, much progress has been achieved in the theory side, including the matching coefficient connecting the quasi-PDFs to the PDFs at one-loop order [11,12], the nucleon-mass correction [13,14], and the renormalization properties of the quasi-PDF [15,16]. There has also been progress in lattice simulation, progressing from the isovector quark PDF of the nucleon [24,28] to the meson DAs [24,29] and nonperturbative renormalization (NPR) in the regularization-independent momentum subtraction (RI/MOM) scheme [28,30]. Certain technical issues regarding the nonperturbative renormalization were raised and addressed in Refs. [17,18,19,20]. In parallel, there have also been other proposals to calculate the PDFs in lattice QCD [30,31] which are subject...
to their own systematics, but they can be complementary to each other as well as the LaMET approach.

So far, all the lattice calculations of the unpolarized quasi-PDF except Ref. [49] have been done for $O_{\gamma_t}$, which mixes with $O_{\gamma}$ at $O(a^0)$. The operator mixing introduces an additional systematic uncertainty in the nonperturbative renormalization that is not negligible [19, 30, 33, 34], thus limiting the precision of the extracted PDF. Fortunately, according to Refs. [18, 29, 37], the $O_{\gamma_t}$ case is free from operator mixing at $O(a^0)$, although the mixing still exists at $O(a)$ [37]. Therefore, it is highly desirable to start from the quasi-PDF with $O_{\gamma_t}$ to improve the systematic uncertainty from operator mixing in the renormalization procedure.

In this work, we describe a state-of-the-art lattice calculation of the unpolarized isovector quark distribution at physical pion mass. We calculate the quasi-PDF with nucleon momenta as large as 3.0 GeV and use 4 source-sink separations to remove excited-state contamination. The nucleon matrix elements are renormalized using RI/MOM scheme; a new matching calculation is done with the new operator $O_{\gamma_t}$ to connect the RI/MOM quasi-PDF to the $\overline{\text{MS}}$ renormalized lightcone PDF; the nucleon mass correction is included. Our result shows a significant improvement compared to previous lattice studies, in particular in the moderate to large-$x$ region. It also signals a promising trend that the precision of lattice calculations is approaching the precision of phenomenological studies with increasing computing resources.

**Nonperturbative Renormalization and Matching:** A nonperturbative renormalization on the lattice is required to obtain the continuum limit of the quasi-PDF matrix element, which is subject to linear UV divergences. In this work, we follow the RI/MOM scheme elaborated in Refs. [17, 29], and match the result to the $\overline{\text{MS}}$ PDF with the one-loop matching coefficient calculated with the method developed in Ref. [17].

First, the RI/MOM renormalization constant $\tilde{Z}$ is calculated nonperturbatively from the lattice by imposing the following momentum subtraction condition on the matrix element of the quasi-PDF in an off-shell quark state $|p, s, x\rangle$:

$$
\tilde{Z}(z, p^R_z, a^{-1}, \mu_R) = \frac{\text{Tr}[\sum_s \langle p, s | O_{\gamma_t}(z) | p, s \rangle]}{\text{Tr}[\sum_s \langle p, s | O_{\gamma}(z) | p, s \rangle_{tree}]} |_{p_z = -q_z, \mu_z = \mu_z}.
$$

(2)

On the lattice, $\langle p, s | O_{\gamma_t}(z) | p, s \rangle$ is calculated from the amputated Green function of $O_{\gamma_t}$ with Euclidean external momentum.

The renormalization constant $\tilde{Z}(z, p^R_z, a^{-1}, \mu_R)$ depends on the lattice spacing as well as the two scales $p^R_z$ and $\mu_R$. It is used to renormalize the bare nucleon matrix element of the quasi-PDF $\tilde{h}(z, P_z, a^{-1}) = \frac{1}{2p^2} \langle p | O_{\gamma_t}(z) | p \rangle$ in coordinate space,

$$
\tilde{h}_R(z, P_z, p^R_z, \mu_R) = \tilde{Z}^{-1}(z, p^R_z, a^{-1}, \mu_R) \tilde{h}(z, P_z, a^{-1}) |_{a \rightarrow 0}.
$$

(3)

Note that the continuum limit of $\tilde{h}_R(z, P_z, p^R_z, \mu_R)$ is well defined and should be taken before the matching to the PDF. Consequently, the RI/MOM quasi-PDF $\tilde{q}(x, P_z, p^R_z, \mu_R)$ is obtained through the Fourier transform of $\tilde{h}_R(z, P_z, p^R_z, \mu_R)$,

$$
\tilde{q}(x, P_z, p^R_z, \mu_R) = P_z \int \frac{dz}{2\pi} e^{ixp_zz} \tilde{h}_R(z, P_z, p^R_z, \mu_R).
$$

(4)

$\tilde{h}_R(z, P_z, p^R_z, \mu_R)$ and $\tilde{q}(x, P_z, p^R_z, \mu_R)$ are independent of the UV regulator, so the matching between the quasi-PDF and MS PDF can be carried out in the continuum theory with dimensional regularization. In Refs. [18, 36, 49], $\tilde{h}_R(z, P_z, p^R_z, \mu_R)$ was first converted to the MS scheme in coordinate space, and then Fourier transformed into momentum space to obtain the MS quasi-PDF, and finally matched to the MS PDF with the matching coefficient from Refs. [20].

Instead of performing a two-step matching, we choose to directly match $\tilde{q}(x, P_z, p^R_z, \mu_R)$ to the MS PDF, and the matching coefficient for the $O_{\gamma_t}$ case has already been calculated at one-loop order in perturbation theory [17]. Although in principle it is equivalent to the two-step procedure described above, the direct matching can possibly save us from additional systematic uncertainties when we implement them numerically on the lattice data. The efficiency of both procedures will be compared in the end to check consistency.

Following the framework described in Refs. [17, 20], the matching between quasi-PDF $\tilde{q}(x, P_z, p^R_z, \mu_R)$ and the MS PDF $q(y, \mu)$ at scale $\mu$ we obtained is,

$$
\tilde{q}(x, P_z, p^R_z, \mu_R) = \int_{-1}^{1} \frac{dy}{|y|} C \left( \frac{x}{y}, \frac{y P_z}{\mu}, \frac{y P_z}{p^R_z} \right) q(y, \mu) + \mathcal{O} \left( \frac{M^2}{\Lambda^2_{\text{QCD}}}, \frac{A^2_{\text{QCD}}}{p^2_z} \right),
$$

(5)

where $\mu = R^2/p^R_z$, and we have absorbed the anti-quark distribution $q(-y, \mu) = -q(-y, \mu)$ into the region $-1 < y < 0$. The matching coefficient $C$ in this work is calculated at one-loop level and will be presented in a future publication [50]. To predict the PDF using the quasi-PDF obtained by lattice calculation, we invert the matching Eq. [5] by changing the sign of $\alpha_s$ in $C$ to obtain the matching coefficient, which is to be convoluted with $\tilde{q}$ to obtain the prediction $q$.

**Lattice-QCD Calculation Setup:** In this paper, we report the results of a lattice-QCD calculation using clover valence fermions on an ensemble of gauge configurations with lattice spacing $a = 0.09$ fm, box size $L \approx 5.8$ fm and pion mass $M_\pi \approx 135$ MeV with
\[ N_f = 2 + 1 + 1 \text{ (degenerate up/down, strange and charm)} \]

flavors of highly improved staggered quarks (HI8Q) \(^{51}\) generated by MILC Collaboration \(^{52}\). The gauge links are hypercubic (HYP)-smeared \(^{53}\) and then clover parameters are tuned to recover the lowest pion mass of the staggered quarks in the sea \(^{54}\) \(^{-57}\). Only one step of HYP smearing is used to improve the discretization effects; too much smearing may alter the ultraviolet results for the PDF. We use Gaussian momentum smearing \(^{58}\) for the quark field

\[
\psi(x) \rightarrow \psi(x) + \alpha \sum_j U_j(x) e^{i k \cdot \hat{e}_j} \psi(x + \hat{e}_j), \quad (6)
\]

where \(k\) is the desired momentum, \(U_j(x)\) are the gauge links in the \(j\) direction, and \(\alpha\) is a tunable parameter as in traditional Gaussian smearing. Such a momentum source is designed to align the overlap with nucleons of the desired boost momentum, and we are able to reach higher boost momenta for the nucleon states than our previous work at physical pion mass \(^{33}\). In our previous exploratory study, although we varied our Gaussian smearing radius to better overlap with the largest momentum used in the calculation, the smearing of the field is still centered around zero in momentum space. With momentum smearing, the center of the smearing will be shifted to momentum \(k\), which will immediately allow us to reach higher boost momenta with better signal-to-noise ratios in the matrix elements.

To better control the systematics due to contamination by nearby excited states, we vary the Gaussian smearing parameter, \(\alpha\), to better capture the excited state, leaving the signal for the ground-state nucleon cleaner. In addition, we use a simultaneous fit of the nucleon matrix element correlators, using four source-sink nucleon separations, 0.72, 0.81, 0.90, 1.08 fm; the detailed procedure is described in Ref. \(^{57}\) for the nucleon charges. We use multigrid algorithm \(^{59, 60}\) in Chroma software package \(^{61}\) to speed up the physical pion mass clover fermion inversion on the quark propagator. We use multiple values of nucleon boost momenta, \(\vec{P} = \{0, 0, n \pi/2\}\), with \(n \in \{10, 12, 14\}\), which correspond to 2.2, 2.6 and 3.0 GeV nucleon momenta.

We apply RI/MOM scheme nonperturbative renormalization as described in our earlier work \(^{29}\) but switch the \(\Gamma\) to \(\gamma_t\) to avoid operator mixing. We use momentum source for NPR propagators with quark momentum ranging in \([0, 14 \times 2\pi/L]\). We study the momentum dependence using \(\mu_R = 2.3\) and 3.7 GeV, as well as the quark momentum dependence \(p^R_\mu\). Fig. 1 shows the \(p^R_\mu\) dependence at fixed \(\mu_R = 3.7\) GeV where the \(Z(z)\) is the inverse of renormalization constant. We find in the small-\(p^R_\mu\) region there is a notable change in renormalization constants, while at large \(p^R_\mu\), it seems reach a plateau and become stable. Similar behavior is also observed in \(\mu_R = 2.3\) GeV case. Therefore, we pick \(p^R_\mu = 10 \times 2\pi/L\) as our central value for the renormalization constant.

Parton Distribution Function Results and Discussion: We use the “derivative” method proposed in our earlier work \(^{33}\) to improve the truncation error due to the Fourier transformation into \(x\) space; that is, we take the derivative of the renormalized nucleon matrix elements \(\partial_z \tilde{h}_R(z)\) whose Fourier transformation differs from the original in a known way:

\[
\tilde{q}(x) = \int_{-z_{\text{max}}}^{+z_{\text{max}}} dz \frac{e^{ixP_z z}}{iz} \partial_z \tilde{h}_R(z) \quad (7)
\]

\(\partial_z \tilde{h}_R(z)\) is consistent with zero for \(|z| > 15a\) and we took \(z_{\text{max}} = 20a\) in this work. The residual truncation systematics can be quantified by using a known global PDF input by checking how well it reproduces itself at lattice parameters, as outlined in Ref. \(^{33}\).

We also investigate the excited-state contamination in the PDF. Excited-state contamination is notorious for contaminating the well-known nucleon axial charges in many past lattice-QCD calculations. As we increase the nucleon boost momentum, we anticipate that excited-state contamination worsens, since the states are relatively closer to each other; therefore, a careful study of the excited-state contamination is necessary for the LaMET (or quasi-/pseudo-PDF) approach. We use multiple analysis methods to remove excited-state systematics among 4 source-sink separations used in this work. First, we use the “two-sim” analysis described in Ref. \(^{57}\) to obtain the ground-state nucleon matrix elements using all 4 source-sink separations. A second extraction uses this method but only the largest two separations. Then, we use the “two-simRR” analysis (see Ref. \(^{57}\) for details), which includes an additional matrix element related to excited states but with almost doubling the errors comparing with the “two-sim” analysis. Our final matrix element when combining all the three analyses is consistent with “two-simRR” analysis. For the rest of this paper, we will focus on the results using “two-simRR”; it has a larger statistical error due to the inclusion of excited states, but the result will be more reliable.

Using the “two-simRR” renormalized PDF and applying the \(\gamma_t\) matching, we show in Fig. 2 one of the three
nucleon boost momenta, \( P_z = 2.2 \) GeV, before and after applying the matching formula of Eq. [5]. The matching raises the antiquark (i.e., negative-\( x \) region) asymmetry for \( x < -0.05 \), and lowers the positive mid-\( x \) to large \( x \) quark distribution, compared with our exploratory study and heavier-pion PDF. After matching, we study the dependence on the nucleon boost momentum, shown in Fig. [2]. Within the statistical errors, the distribution seems to converge across the three momenta. However, the central values shift noticeably from 2.2 to 3.0 GeV, moving the antiquark distribution toward the asymmetry measured in experiment: \( \bar{d}(x) > \bar{u}(x) \).

3.0 GeV. We use \( O_{2g} \) operator here rather than the \( O_{2g} \) used in Ref. [29]. (Though we showed in Ref. [29] that the final PDF is consistent when including the scalar matrix elements.) We extend our matrix element analysis to include an extra term due to the excited states. We increase at least a factor of 10 in statistics. We also include the complete matching needed from the RI/MOM quasi-PDF to MS lightcone PDF.

While finalizing our analysis, another PDF result at physical pion mass was reported by ETMC [49]. There are several advantages to our work: Larger \( M_\pi L \) (roughly 4 vs 3) to minimize the finite-volume effects, larger nucleon boost momentum to suppress the power corrections, bigger \( zP_z \) range to increase reliable \( x \)-regions, multistate fitting to remove excited-state contamination. It is notable that our results agree with the sea quark asymmetry seen experimentally and agree with zero in unphysical regions \( |x| > 1 \). ETMC’s results have oscillations that continue into the \( |x| > 1 \) regions, which is possibly due to truncation of the Fourier transformation [35].

Other possible sources of systematic uncertainty include: 1) Higher-twist corrections: Taking \( P_z = 3.0 \) GeV, we do not see significant difference in the \( |x| > 0.2 \) regions, indicating that the higher-twist correction is well-controlled when \( x \) is not too small. 2) Truncation effects: At our largest nucleon boost momentum, our largest \( z_{\text{max}}P_z \) is around 27.5. When we reconstruct a known PDF, there is significant difficulty in reproducing the small-\( x \) regions \( |x| < 0.15 \). Note that other lattice-PDF calculations use smaller \( z_{\text{max}}P_z \), yielding smaller reliable regions in \( x \)-space. By extending to large displacement \( z \) or large momentum \( P_z \) on finer lattice spacing in future work, this can be improved straightforwardly. 3) \( O(\alpha_s^2) \) error in perturbation theory: we estimated the \( O(\alpha_s^2) \) error by applying matching to experimental fitted PDF to quasi-PDF using Eq. [5] and then do the inverse matching back to PDF. We find that the error is about the same size as the statistical error.

We have presented new lattice-QCD results for the isovector unpolarized PDF (that is, the up-down quark asymmetry in the proton), which has much potential impact on current PDF estimates in the near future: 1) The isovector PDF at large \( x \) can be used as a constraint in global PDF analysis. The large-\( x \) experimental data are often contaminated by nuclear effects, which are hard to cleanly remove. Many current PDF analyses (see references in Ref. [62]) rely on extrapolation in these regions. There are ongoing LHCb measurements that can potentially improve and constrain the PDF in the large-\( x \) region, but the precision of these data are not yet good enough to make a difference. A recent community whitepaper among lattice and global analysis practitioners [62] predicted that a calculation of the large-\( x \) isovector with 10% final error can improve on the current PDF, especially in the antiquark regions where experimental inputs are even scarcer. Currently, we are able to reproduce the global PDF results; the next step will be

\[
\begin{align*}
\mu \text{matched PDF} = (\mu_R, p^0_R) = (3.7, 2.2) \text{ GeV}.
\end{align*}
\]

FIG. 2. The quasi-PDF and matched PDF with nucleon boost momentum 2.2 GeV. The parameters of the central value of matched PDF is \((\mu_R, p^0_R) = (3.7, 2.2)\) GeV. The matching process lowers the quasi-PDF at large positive \( x \) and enhances the small-\( x \) region’s quark asymmetry. It significantly changes the antiquark asymmetry at this nucleon momentum.

\[
\begin{align*}
P_z = 2.2 \text{ GeV} & \quad P_z = 2.6 \text{ GeV} \quad P_z = 3.0 \text{ GeV}
\end{align*}
\]

FIG. 3. Nucleon boost momentum dependence of the matched unpolarized isovector PDFs. The parameters of the central value of matched PDF is \((\mu_R, p^0_R) = (3.7, 2.2)\) GeV. For quark asymmetry, the shape is consistent throughout most \( x \) regions. However, in the antiquark region, there is a significant change in distribution as momentum increases.

The final result from this paper, shown in Fig. [4] significantly improves on our previous results at physical pion mass [35]. We increase the nucleon momenta used in the calculation from 0.4, 0.8, 1.3 GeV to 2.2, 2.6 and...
to plan improved calculations with total uncertainty less than 10%. 2) With the promising results shown here, we can proceed with similar analyses for the less known polarized PDFs, such as helicity and transversity (the longitudinal and transversely polarized PDF), where the isovector PDFs needed to make impacts for global analysis are less demanding than the unpolarized ones.

Summary and Outlook: In this work, we report the state-of-the-art isovector unpolarized quark distribution using lattice QCD directly at physical pion mass. We use nucleon boosted momenta as large as 3 GeV with high-statistics analysis. We carefully study excited-state systematics whose error is reflected in our final distribution uncertainty. We renormalize our nucleon matrix element using the nonperturbative RI/MOM renormalization, and perform the LaMET one-loop finite-momentum matching and conversion to MS-scheme to connect lattice quasi-distribution to lightcone distribution. We found our final distribution agree well with the global analysis distribution. We carefully examine all possible systematics which will give us better guideline to improve our future calculations and provide better precision distributions. Future direction will be investigating smaller lattice spacing ensembles for reaching even higher boosted momentum such that we can push toward smaller-$x$ region.

ACKNOWLEDGMENTS

We thank the MILC Collaboration for sharing the lattices used to perform this study. The LQCD calculations were performed using the Chroma software suite [61]. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 through ALCC and ERCAP; facilities of the USQCD Collaboration, which are funded by the Office of Science of the U.S. Department of Energy, and supported in part by Michigan State University through computational resources provided by the Institute for Cyber-Enabled Research. JWC is partly supported by the Ministry of Science and Technology, Taiwan, under Grant No. 105-2112-M-002-017-MY3 and the Kenda Foundation. LCJ is supported by the Department of Energy, Laboratory Directed Research and Development (LDRD) funding of BNL, under contract de-sc0012704. YSL is supported by Science and Technology Commission of Shanghai Municipality (Grant No.16DZ2260200) and National Natural Science Foundation of China (Grant No.11655002). HL and YY are supported by the US National Science Foundation under grant PHY 1653405 “CAREER: Constraining Parton Distribution Functions for New-Physics Searches”. JHZ is supported by the SFB/TRR-55 grant “Hadron Physics from Lattice QCD”, and a grant from National Science Foundation of China (No.1405104). YZ is supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, from de-sc0011090 and within the framework of the TMD Topical Collaboration.
