

# Probing parton distribution functions at large $x$ via Drell-Yan forward-backward asymmetry

Yao Fu<sup>1,\*</sup>, Raymond Brock<sup>2,†</sup>, Daniel Hayden<sup>2,‡</sup> and C.-P. Yuan<sup>2,§</sup>

<sup>1</sup>*Department of Modern Physics, University of Science and Technology of China, Jinzhai Road 96, Hefei, Anhui 230026, China*

<sup>2</sup>*Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, USA*



(Received 2 August 2023; accepted 7 February 2024; published 4 March 2024)

The forward-backward asymmetry of the Drell-Yan process in dilepton decays at high invariant masses can be used to probe the parton distribution functions at large  $x$ . The behavior of three modern parton distribution function sets (CT18NNLO, MSHT20, and NNPDF4.0) are compared, and updated under various scenarios via ePump using proton-proton collision pseudodata generated at  $\sqrt{s} = 13$  TeV with  $3000 \text{ fb}^{-1}$  of integrated luminosity.

DOI: [10.1103/PhysRevD.109.054006](https://doi.org/10.1103/PhysRevD.109.054006)

## I. INTRODUCTION

Many of the future, high luminosity, precision measurements at the Large Hadron Collider (LHC) will be limited by modeling and theoretical uncertainties. Prominent among these are uncertainties related to knowledge of parton distribution functions (PDFs). Kinematics of the wide range of measurements and searches at the LHC span parton  $x$  ranges from very low where gluon and antiquark densities are relevant to high- $x$  regions dominated by valence quark densities. An intriguing approach to further limiting PDF uncertainties might be to, rather than rely on global, all-purpose PDF sets, create targeted “boutique” PDF sets designed for specific purposes. Judicious kinematic selection of LHC data as strategic, supplementary inputs to PDF fitting might lead to legitimate reductions from such special-purpose PDF sets. Neutral-current Drell-Yan (NCDY) backgrounds constitute a trial of this idea as high mass resonant and nonresonant searches are standard targets for new physics searches but they are now compromised by PDF uncertainties, and will be more-so in the future. Standard model (SM) backgrounds that must be modeled very well are experimentally uncomplicated and

amenable to high experimental precision. So NCDY studies seem a good way to explore this idea of specialized PDFs.

We tested that approach in a previous study [1] and demonstrated how judiciously-chosen NCDY data below  $m_{\ell\ell} = 1$  TeV, can result in new PDFs that bring significant improvement to high-mass PDF uncertainties. That paper studied the triple differential cross section in selected regions of the variables  $m_{\ell\ell}$ ,  $\cos(\theta^*)$ , and  $\eta_{\ell\ell}$  and found choices that enhanced well-understood PDFs by modifying a standard set into a set designed for specific new physics searches; hence, “boutique”. We then simulated the sensitivity of those modified PDFs for a future LHC search for high-mass  $Z' \rightarrow \ell\ell$  final states and found significant improvement in the theoretical uncertainties of such a search. This idea can be extended and multiple calculations are underway to study additional variables from other measurables and reactions beyond NCDY.

This paper outlines the next calculation in this series, specifically looking at the behavior of various PDF sets in the description of the forward-backward asymmetry ( $A_{\text{FB}}$ ) at high invariant masses. We do not simulate a specific new physics search here, but are interested in what we can learn about PDFs in high-mass regions. The ePump Package [2] is then used to understand the effect of using these updated PDF nominal values and uncertainties if nature behaved according to the other PDFs considered. We note that others have begun exploring ways to improve NCDY PDF uncertainties at high invariant mass as well [3–9]. This study differs from other studies through its use of ePump, and careful consideration of the tolerance used, as well as the focus on  $A_{\text{FB}}$  specifically in the high-mass region. An important aim of this study is to show that the approach of using quantitatively new data to update the central value of a PDF and its associated uncertainties is not just a novel

\*yfu@mail.ustc.edu.cn

†brockr@msu.edu

‡haydend5@msu.edu

§yuanch@msu.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.

feature of CTEQ-TEA, but can also be used for other modern PDF sets.

The following equation describes the leading-order relationship between initial state parton momentum fractions and NCDY kinematic variables in hadron collisions,

$$x_1, x_2 = \frac{Q}{\sqrt{s}} e^{\pm y}, \quad (1)$$

where  $x_1, x_2$  are the momentum fractions of partons in the incoming beam,  $Q$  is the energy scale,  $\sqrt{s}$  is the collision energy, and  $y$  is the rapidity of the DY pair. Note that this equation only holds exactly at leading order, but is appropriate for this study in which low- $p_T$  data dominates. According to Eq. (1), high mass events directly correspond to the high- $x$  region in PDFs. In modern PDFs, the valence quarks are well-known for  $x \sim 0.3$  while the sea-quark PDFs, especially the strangeness PDFs, are hardly known for  $x > 0.3$ . However, for  $x > 0.7$  even the valence PDFs are not well-determined. Kinematic distributions for the high-mass Drell-Yan process, especially the forward-backward asymmetry ( $A_{\text{FB}}$ ) directly provide valence parton information in that large- $x$  region. Therefore,  $A_{\text{FB}}$  measurements in the high-mass region at the LHC can provide complementary information on large- $x$  PDFs to the science program at the Electron-Ion Collider (EIC) [10]. This leads to the motivation to use the dilepton invariant mass range of 0.5 TeV to 5.0 TeV, to focus on the high-mass region away from the  $Z$ -boson peak. No other kinematic criteria are applied to the events studied, though most events fall in the range of dilepton rapidity,  $|y| < 3$ , and dilepton  $p_T$  between 2 GeV and 40 GeV.

In this study, three modern PDF sets are compared; CT18NNLO [11], MSHT20 [12], and NNPDF4.0 [13] by generating proton-proton ( $pp$ ) collision pseudodata corresponding to  $\sqrt{s} = 13$  TeV and an integrated luminosity of  $3000 \text{ fb}^{-1}$ . These pseudodata are used as input to the ePump Package to update either CT18NNLO or MSHT20 as the underlying theory input. By updating the PDFs, both the central value of the PDF, and the 68% confidence level (C.L.) variation uncertainty (i.e., for CT18NNLO this uncertainty describes the variation of the underlying eigenvectors) will change. This study also carefully considers the role of the tolerance used in ePump when updating

PDF sets with new data, and discusses how to appropriately set these parameters to obtain a realistic result. Note that using ePump to update the NNPDF Hessian sets is not possible as it was found that many eigenvector sets have  $\chi^2$  less than the NNPDF central set, a feature which has been extensively discussed [14].

## II. FORWARD-BACKWARD ASYMMETRY AND DILUTION FACTOR

In this work, we follow Ref. [15] and define the forward-backward asymmetry ( $A_{\text{FB}}$ ) at hadron colliders in the Collins-Soper (CS) frame [16], which is a special rest frame of the DY pair, as

$$A_{\text{FB}} = \frac{N_F - N_B}{N_F + N_B}, \quad (2)$$

where  $N_F$  and  $N_B$  are the number of forward and backward events, respectively. In the NCDY process, the scattering angle  $\theta$  is defined by the momentum direction of the outgoing fermion  $f_j$  relative to the momentum direction of the incoming fermion  $f_i$ . The sign of  $\cos \theta$  is used to define forward ( $\cos \theta > 0$ ) and backward ( $\cos \theta < 0$ ) events. At hadron colliders, due to the unknown momentum direction of the incoming quarks, forward and backward events are defined in the CS rest frame [16], where the polar and azimuthal angles are defined relative to the two hadron-beam directions. The  $z$ -axis is defined in the rest frame of the DY pair, bisecting the angle between the incoming hadron momentum and the negative of the other hadron momentum. The cosine of the polar angle  $\theta^*$  between the momentum direction of the outgoing lepton  $l^-$  and the  $\hat{z}$ -axis in the CS frame is defined as the scattering angle of the DY pair at hadron colliders, which can be calculated directly from the laboratory frame lepton quantities by

$$\cos \theta_{\text{CS}}^* = c \frac{2(p_1^+ p_2^- - p_1^- p_2^+)}{m_{ll} \sqrt{m_{ll}^2 + p_{T, ll}^2}}, \quad (3)$$

where the scalar factor  $c$  (either 1 or  $-1$ ) is defined for the Tevatron and the LHC, respectively, as

$$c = \begin{cases} 1, & \text{for the Tevatron (a proton-antiproton collider)} \\ \vec{p}_{Z, ll} / |\vec{p}_{Z, ll}|, & \text{for the LHC (a proton-proton collider)}. \end{cases} \quad (4)$$

And thus, the sign of the  $z$ -axis is defined as the proton beam direction for the Tevatron, and as the sign of the boost direction of the lepton pair with respect to the  $z$ -axis in the laboratory frame on an event-by-event basis for the LHC. The variables  $p_{Z, ll}$ ,  $m_{ll}$ , and  $p_{T, ll}$  denote the longitudinal momentum, invariant mass, and transverse momentum of the dilepton system, respectively, and,

$$p_i^\pm = \frac{1}{\sqrt{2}}(E_i \pm p_{z,i}), \quad (5)$$

where the lepton (antilepton) energy and longitudinal momentum are  $E_1$  and  $p_{z,1}$  ( $E_2$  and  $p_{z,2}$ ), respectively. DY events are therefore defined as forward ( $\cos \theta_{\text{CS}}^* > 0$ ) or backward ( $\cos \theta_{\text{CS}}^* < 0$ ) according to the direction of the outgoing lepton in this frame of reference. In the case of the LHC, one can define another frame such that the  $z$ -axis is oriented to the quark direction of motion. We use  $\cos \theta_q^*$  to denote this case, and define a coefficient  $c$  as

$$c = \vec{p}_q / |\vec{p}_q|. \quad (6)$$

We then use  $\cos \theta_h^*$  to denote the case using the lepton-pair momentum to define the  $z$ -axis. One can easily obtain the relationship between  $\cos \theta_h^*$  and  $\cos \theta_q^*$ ,

$$\begin{aligned} \cos \theta_h^* &= \cos \theta_q^*, & \text{for } E(q) > E(\bar{q}) \\ \cos \theta_h^* &= -\cos \theta_q^*, & \text{for } E(q) < E(\bar{q}). \end{aligned} \quad (7)$$

As introduced in previous studies [15,17,18], the dilution factor ( $D$ ) quantifies the probability that the energy of the antiquark is larger than the energy of quark. When the quark carries higher energy, the number of forward and backward events in the two different frames will be the same, which has a probability of  $(1 - D)$ . When the antiquark carries higher energy, the number of forward and backward events in the two different frames will have a different sign, which has a probability of  $D$ . Finally, the number of forward and backward events  $N_F^h$  and  $N_B^h$  defined by the  $\cos \theta_h^*$  can be written as

$$\begin{aligned} N_F^h &= (1 - D)N_F^q + DN_B^q, \\ N_B^h &= (1 - D)N_B^q + DN_F^q, \end{aligned} \quad (8)$$

where the  $N_F^q$  and  $N_B^q$  represent the number of forward and backward events defined by the  $\cos \theta_q^*$ . As a result, the relationship between the  $A_{\text{FB}}$  defined by the  $\cos \theta_h^*$ , and the  $A_{\text{FB}}$  defined by the  $\cos \theta_q^*$ , can be roughly written as

$$A_{\text{FB}}^h \approx (1 - 2D)A_{\text{FB}}^q, \quad (9)$$

where  $A_{\text{FB}}^h$  and  $A_{\text{FB}}^q$  are the asymmetries defined by  $\cos \theta_h^*$  and  $\cos \theta_q^*$ , respectively. Writing this for different flavors, Eq. (9) can be written precisely as

$$A_{\text{FB}}^h = \sum_f \frac{N_f}{N} (1 - 2D_f) A_{\text{FB}}^f, \quad (10)$$

where  $f$  represents the index of flavors which is directly coupled with  $Z$  boson. Here the  $f$  contains five flavors:  $u\bar{u}$ ,  $d\bar{d}$ ,  $s\bar{s}$ ,  $c\bar{c}$ , and  $b\bar{b}$ . The  $N$  and  $N_f$  represent the total number of events, and the number of events with a specific flavor, respectively.  $D_f$  represents the dilution factor in the process with a certain flavor, and  $A_{\text{FB}}^f$  is defined by the  $\cos \theta_q^*$  with flavor dependence. The up-type and down-type flavor coupled with  $Z$  boson have different  $A_{\text{FB}}$  values, where a more detailed discussion can be found in the  $A_{\text{FB}}$  factorization study [15]. The dilution factor  $D$  defined in Eq. (9) can be roughly treated as an average of all the flavor combinations. Figure 1 shows the  $A_{\text{FB}}$  and dilution factor as a function of dilepton mass for all flavor combinations,  $D$ , only  $u\bar{u}$  contributions,  $D_u$ , and only  $d\bar{d}$  contributions,  $D_d$ , for the CT18 PDF. The calculation is done by using the MCFM program [19] at next-to-leading order (NLO), interfaced to APPLgrid [20]. The  $s$ ,  $c$ , and  $b$  quark contribution are about the same as their antiquarks in CT18 PDFs, so the dilution factor  $D_s$ ,  $D_c$ , and  $D_b$  will be 0.5, and the  $A_{\text{FB}}$  defined by  $\cos \theta_h^*$  for those processes will be zero. For CT18NNLO,  $\bar{s}(x, Q_0)$  is assumed to be identical to

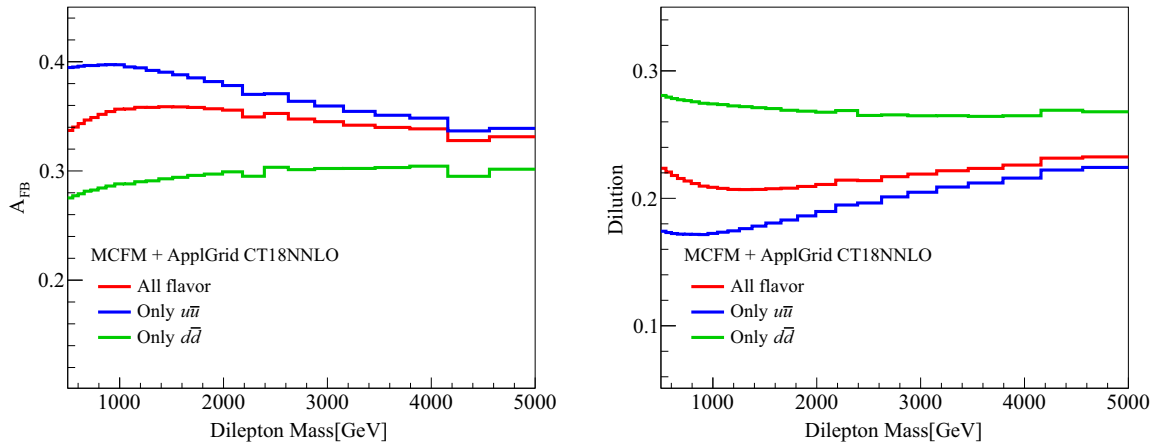


FIG. 1. The  $A_{\text{FB}}$  (left) and dilution (right) spectrum as a function of dilepton mass for MCFM at NLO accuracy in the high mass region ( $500 \text{ GeV} < M_{ll} < 5000 \text{ GeV}$ ), using the CT18NNLO PDF set. The red, blue, and green curves represent the all-flavor combination, only  $u\bar{u}$  contribution, and only  $d\bar{d}$  contribution, respectively.

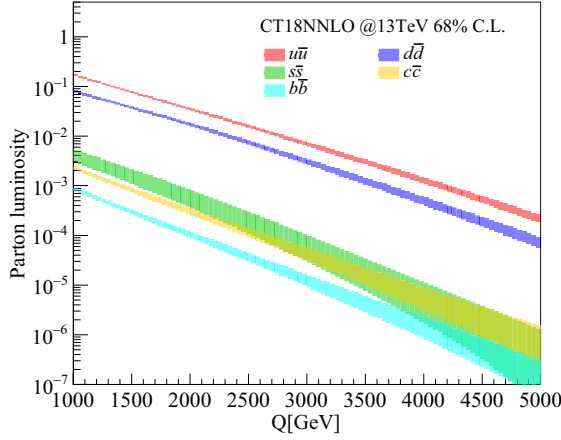


FIG. 2. Parton luminosities of CT18NNLO at the high-invariant mass region for the initial state of  $u\bar{u}$ ,  $d\bar{d}$ ,  $s\bar{s}$ ,  $c\bar{c}$ , and  $b\bar{b}$ . The width of the curves represents the PDF uncertainty for 68% C.L.

$s(x, Q_0)$  at  $Q_0 = 1.3$  GeV. When the PDFs are evolved to higher energy scale  $Q$ ,  $\bar{s}(x, Q_0)$  and  $s(x, Q_0)$  can differ slightly at next-to-next-to-leading order (NNLO) and beyond. However, even in the case of allowing  $\bar{s}(x, Q_0)$  not equal to  $s(x, Q_0)$ , such as in CT18As [21], the current constraint on the strangeness asymmetry is such that the dilution factor originating from the  $s\bar{s}$  production channel will also be small. As shown in Fig. 1 (right), the dilution factor of the  $u\bar{u}$  process is smaller than that of the  $d\bar{d}$  process, while the  $A_{\text{FB}}^h$  of the  $u\bar{u}$  process is larger than the  $d\bar{d}$  process.

Figure 2 shows the parton luminosities for  $u\bar{u}$ ,  $d\bar{d}$ ,  $s\bar{s}$ ,  $c\bar{c}$ , and  $b\bar{b}$  processes. This shows that the  $u\bar{u}$  process is dominant in the high-invariant mass region. As a result, the  $A_{\text{FB}}^h$  value of the all flavor combination is closer to the  $A_{\text{FB}}^h$  in the  $u\bar{u}$  process.

To specify what kind of flavor and  $x$  region can be constrained, we write the dilution factor in parton language. According to the definition of the dilution factor, the equivalent in parton language can be written as

$$D_q = \frac{f_{\bar{q}}(x_1)f_q(x_2)}{f_q(x_1)f_{\bar{q}}(x_2) + f_{\bar{q}}(x_1)f_q(x_2)}, \quad \text{with } x_1 > x_2. \quad (11)$$

From this one can easily see that the dilution factor is sensitive to the relative difference between quarks and antiquarks. Figure 3 shows the comparison of  $\bar{u}/u$  and  $\bar{d}/d$  quark PDFs for the CT18 PDF. Above  $x > 0.6$  the relative difference between  $\bar{u}$  and  $u$  is larger than the difference between  $\bar{d}$  and  $d$ , which explains the difference of  $D_u$  and  $D_d$  shown in Fig. 1.

This separation between  $\bar{u}$  and  $u$  and  $\bar{d}$  and  $d$  suggest that  $A_{\text{FB}}$  in the high-invariant mass region can provide sensitive information about the relative difference between quarks and antiquarks, namely, the valence information in the

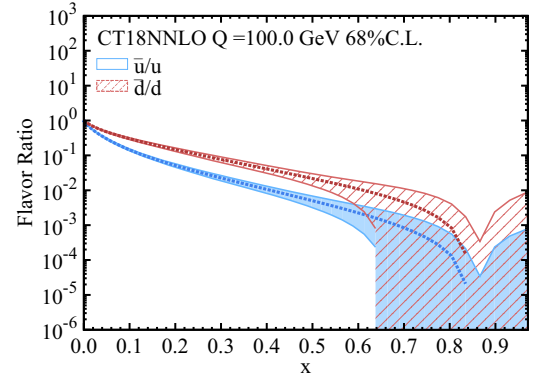


FIG. 3. Central value and uncertainty of CT18 of  $\bar{u}/u$  and  $\bar{d}/d$  PDFs.

large- $x$  region, which is not currently constrained by existing measurements at the LHC.

### III. PDF UPDATING USING $A_{\text{FB}}$ IN THE HIGH-MASS REGION

In this section, we quantitatively show the impact of using  $A_{\text{FB}}$  in the high-invariant mass region, to update the PDFs under study and their associated uncertainties, using the ePump package. ePump is meant to be used as a tool to approximate an update to a given PDF set and its respective PDF uncertainties in response to new kinds of input data, in this study,  $A_{\text{FB}}$ . Should an ePump exercise suggest that these new kinds of data might inform the central PDF and/or reduce PDF uncertainties, then these new data types should be considered as additions to complete, global fitting. In this study pseudodata were generated for  $A_{\text{FB}}$  in the high-invariant mass region using ResBos [22,23] at  $\text{N}^3\text{LL} + \text{NNLO}$  in QCD at  $\sqrt{s} = 13$  TeV. As this study is concerned with high-mass DY, no kinematic cuts were applied, such as lepton  $p_T$ , dilepton  $p_T$ , or rapidity. To confirm the robustness of the ResBos cross section calculation the estimates were also compared to MCFM [19] at NLO, and found to be in good agreement. Each PDF under study was used to generate the theory template, required by ePump to perform the update. Electroweak parameters can affect  $A_{\text{FB}}$  distributions, but a previous study [17] shows that this mainly affects the Z-peak region. So we leave the electroweak parameters set to be the same between pseudodata and theory templates. In order to study the impact of various possible datasets, CT18, MSHT20, and NNPDF4.0 were each used to generate pseudodata, which are then subsequently used to update the nominal PDF under study.

For the pseudodata, samples were generated corresponding to an integrated luminosity of  $3000 \text{ fb}^{-1}$  in order to study the impact of data in future at the high-luminosity LHC. No kinematic cuts are imposed in this study.  $A_{\text{FB}}$  in the invariant mass region from 500–5000 GeV were used to perform the PDF updating (using 25 mass bins of varying size to preserve good statistical precision).



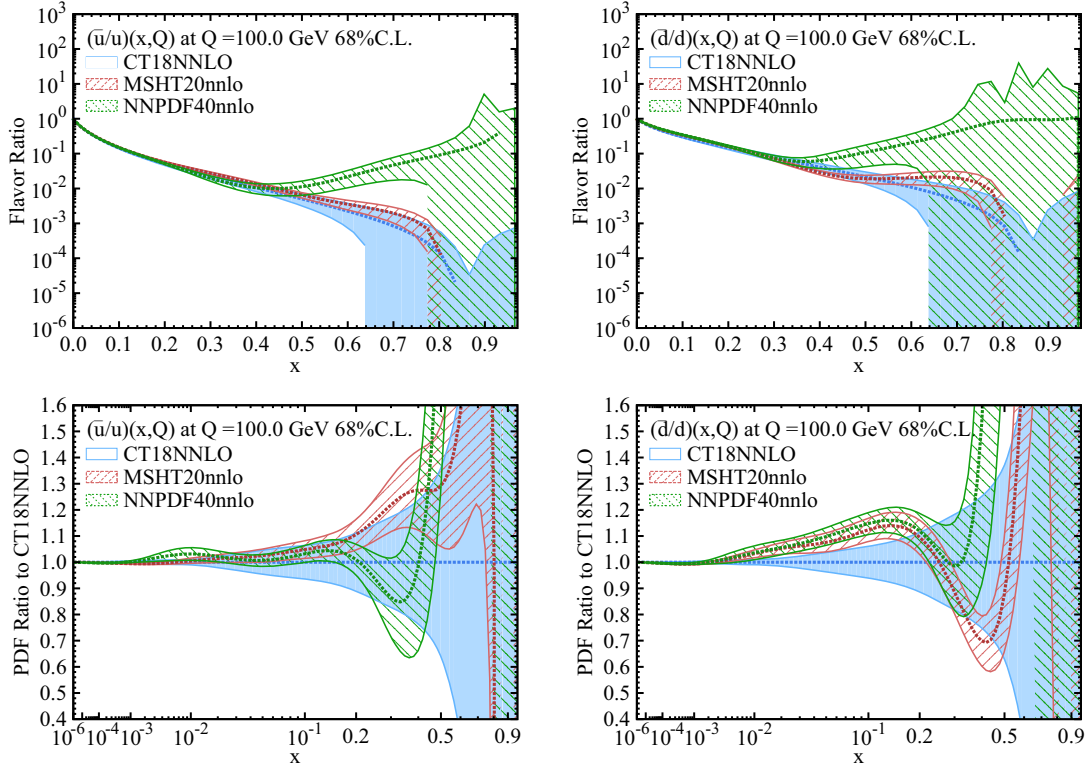


FIG. 4. Central value and uncertainty of CT18, MSHT20, and NNPDF40 (top), and ratio (bottom) of the central value and uncertainty to the CT18NNLO central value of the  $\bar{u}/u$  (left), and  $\bar{d}/d$  (right) PDFs.

As a starting point, Fig. 4 shows the central value and uncertainty of CT18, MSHT20, and NNPDF40, as well as their ratio compared to CT18, for  $\bar{u}/u$ , and  $\bar{d}/d$ .

#### IV. PDF COMPARISONS

Figure 5 shows the MCFM calculation at NLO for the  $A_{FB}$  spectrum and the dilution factor as a function of dilepton

invariant mass in the high-invariant mass region with CT18, MSHT20, and NNPDF4.0 PDFs. Both Figs. 4 and 5 suggest different behavior for NNPDF4.0 as compared with that of CT18 and MSHT20, and the dilution factor can explain the drop-off in  $A_{FB}$  for NNPDF4.0 at high-invariant masses compared to CT18 and MSHT20. Also, Fig. 5 (right) shows a growing feature for the dilution

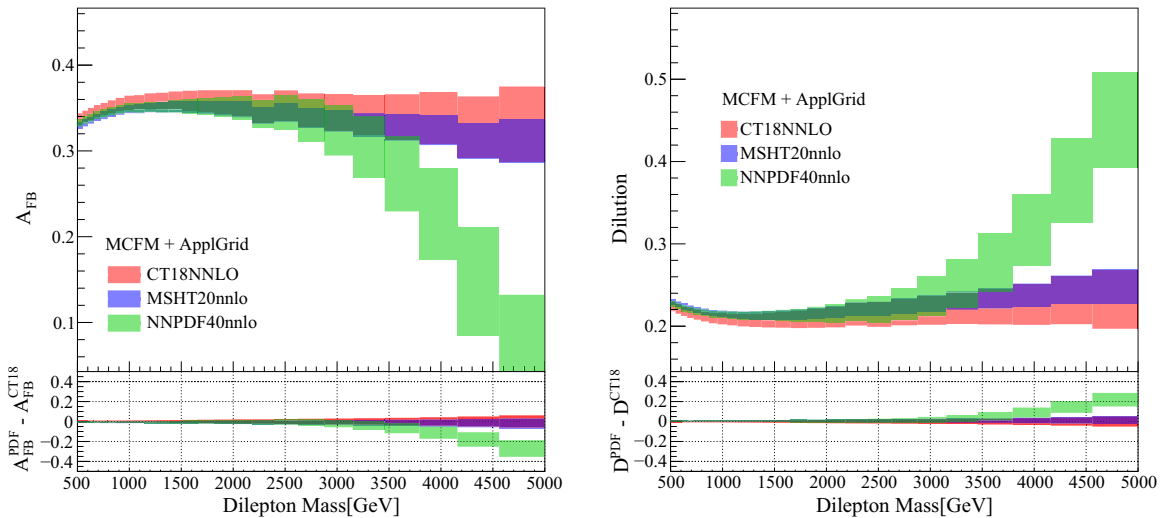


FIG. 5. Comparison of the  $A_{FB}$  (left) and dilution (right) with different PDF inputs for MCFM at NLO accuracy in the high-mass region ( $500 \text{ GeV} < M_{ll} < 5000 \text{ GeV}$ ). The band represents the PDF uncertainty.

factor as a function of invariant mass for NNPDF4.0. According to Eq. (9), when the dilution factor  $D$  is larger than 0.5, a negative  $A_{\text{FB}}^h$  will be observed and that appears to be the case for NNPDF4.0 in the very high-invariant mass region ( $> 5000$  GeV). In order to explain the negative  $A_{\text{FB}}$  in very high-invariant mass regions, a naive toy model has been discussed [24].

Comparing  $A_{\text{FB}}$  and the dilution factor described by CT18 and MSHT20 also reveals some differences in the methodology of the PDF global analysis, with the adoption of different datasets, higher-order theoretical calculations, and different choices of the nonperturbative parametrization forms of various parton PDFs.

### A. PDF-updating for CT18

Firstly, the CT18 PDF was updated using pseudodata generated by CT18 itself, based on the flavor combination of  $\bar{u}/u$ , and  $\bar{d}/d$ . This is because according to the definition of the dilution factor in Eq. (11), the most sensitive flavor combination of  $A_{\text{FB}}$  data is  $\bar{u}/u$ , and  $\bar{d}/d$ . Figure 6 shows the ratio of CT18 and its uncertainties before and after updating. Since CT18 was used for both the pseudodata and the theory templates, the central values of the PDF set

does not change, as expected. Similarly, the error band after the update shows only a slight reduction of a few percent compared to the original. This is due to the relatively small number of events in the high-invariant mass region. A similar study [5] which used xFitter instead of ePump to update CT18 using  $A_{\text{FB}}$  from the Z-peak region, found a relative improvement in the PDF uncertainty at  $x = 10^{-4}$  of  $\sim 63\%$ . We also calculated the result of limiting our inputs to the Z-peak region instead of the high-mass region, and found an improvement in the PDF uncertainty of  $\sim 53\%$  for the same  $x$  range. However, xFitter and ePump use different tolerances to weight the importance of new data in the PDF update, which is an important topic handled in more detail in the Appendix of this paper. Figure 7 then shows the results of updating the CT18 PDF using the pseudodata generated by MSHT20 and NNPDF4.0, respectively. The  $\chi^2$  of the MSHT pseudodata changes when performing the PDF updating from 23.3 to 12.6, while the  $\chi^2$  of the NNPDF pseudodata changes when the PDF is updated by NNPDF pseudodata from 40.6 to 21.8. Due to the different  $A_{\text{FB}}$  distributions predicted by MSHT20 and NNPDF4.0, the effects of new  $A_{\text{FB}}$  data in the high-mass region are absorbed into the  $\bar{u}/u$ , and  $\bar{d}/d$  PDFs in the large- $x$  region.

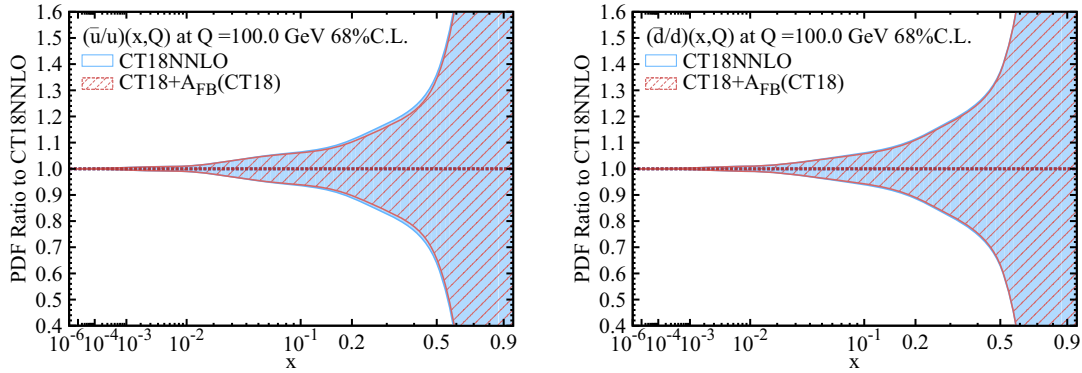


FIG. 6. PDF update of CT18 for  $\bar{u}/u$  (left), and  $\bar{d}/d$  (right) using  $A_{\text{FB}}$  pseudodata generated using CT18. The central value and uncertainty are compared to the CT18 central value.

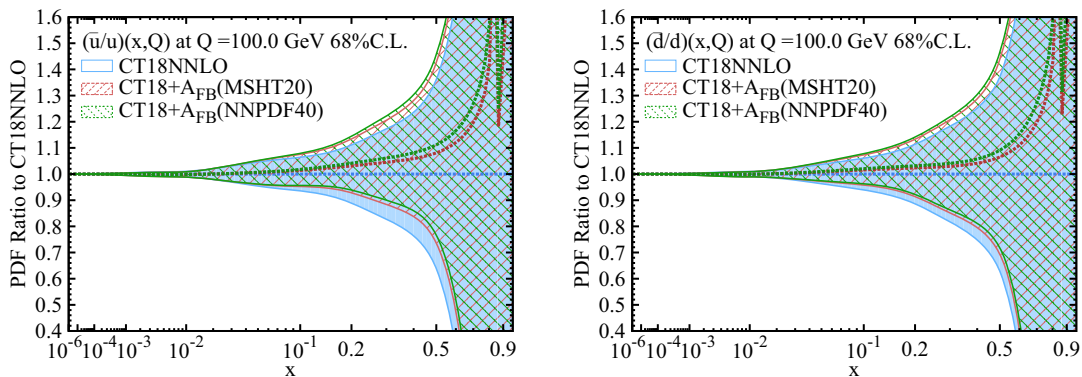


FIG. 7. PDF update of CT18 for  $\bar{u}/u$  (left), and  $\bar{d}/d$  (right) using  $A_{\text{FB}}$  pseudodata generated using MSHT20 and NNPDF4.0. The central value and uncertainty are compared to the CT18 central value.

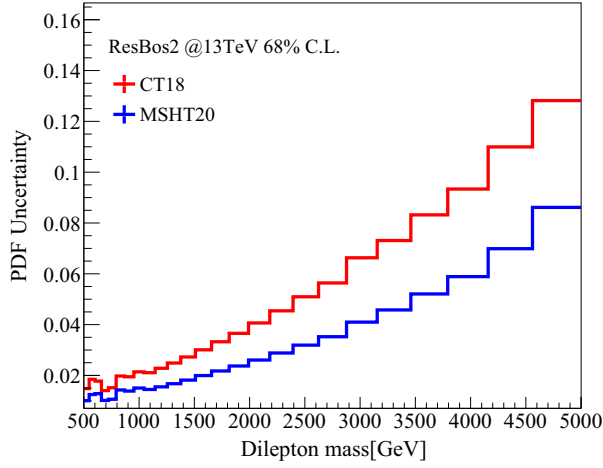


FIG. 8. PDF uncertainty of  $A_{FB}$  in high-invariance mass region corresponding to 68% C.L. for CT18 and MSHT20.

The updated  $\bar{u}/u$ , and  $\bar{d}/d$  PDFs show obvious deviations compared to the CT18 PDF.

### B. PDF updating for MSHT20

In this section, another set of PDF updating results are presented based on using MSHT20 as the nominal PDF set. The pseudodata are kept the same as the previous part of this study. The difference in this section is that the theory templates are generated using the MSHT20 PDF instead of CT18. Firstly, in MSHT20 the nonperturbative shape parameters are different with respect to CT18, which leads to different behavior for  $\bar{u}/u$  and  $\bar{d}/d$ , and means it is interesting to repeat the study using MSHT20 as the nominal PDF. Secondly, since the tolerance used in MSHT20 is about 10, which is smaller than CT18 by a factor of 3 (The average tolerance for CT18 is about 30 for 68% C.L.), the impact of new data on MSHT20 could be much stronger. A more in-depth discussion on the effect of using different tolerances is described in the Appendix. However, as shown in Fig. 8, the PDF uncertainty of  $A_{FB}$  for MSHT20 is much smaller than for CT18. The updating

procedure also conveys such information, which when coupled with the smaller tolerance being used, finally ends up with a similar result in terms of relative change to the CT18 case. When using CT18 pseudodata to update MSHT, the  $\chi^2$  of CT18 pseudodata is changed from 23.4 to 9.0. When using NNPDF pseudodata to update MSHT, the  $\chi^2$  of NNPDF pseudodata is changed from 4.8 to 3.4.

Figure 9 shows the PDF updating results when using the pseudodata generated by CT18, MSHT20, and NNPDF40 to update MSHT20.

## V. CONCLUSION

This study presents the result of using the kinematic information from forward-backward asymmetry in the Drell-Yan process at high-dilepton invariant mass (0.5 TeV to 5.0 TeV) to update various PDF sets (CT18NNLO, MSHT20, and NNPDF4.0). Pseudodata were generated for  $pp$  collisions at  $\sqrt{s} = 13$  TeV, for an integrated luminosity of  $3000 \text{ fb}^{-1}$ , using the separate PDFs, and input into ePump to update the PDF central value and 68% C.L. variation uncertainties for CT18NNLO and MSHT20. NNPDF4.0 was found to have a different behavior in  $A_{FB}$  versus dilepton invariant mass compared to the other PDFs considered. This is related to the dilution factor, and as such the most sensitive parts of the PDF to these effects are the ratio of  $\bar{u}/u$  and  $\bar{d}/d$ . This is particularly important in the context of “boutique” PDFs as it shows that  $A_{FB}$  at high-dilepton invariant mass is a useful observable for constraining the large- $x$  region in processes which are sensitive to  $\bar{u}/u$  and/or  $\bar{d}/d$ . Using a dataset of  $3000 \text{ fb}^{-1}$  appears to be sufficient to differentiate among the PDFs, such that if CT18NNLO or MSHT20 were assumed as the underlying theory, but nature behaved closer to NNPDF4.0, the former PDF sets would be updated by the new input data accordingly. While this new information from  $A_{FB}$  at high invariant mass is effective at shifting the nominal prediction of PDFs, it was also shown that the effect on the PDF variation uncertainty is fairly modest. Given the broad

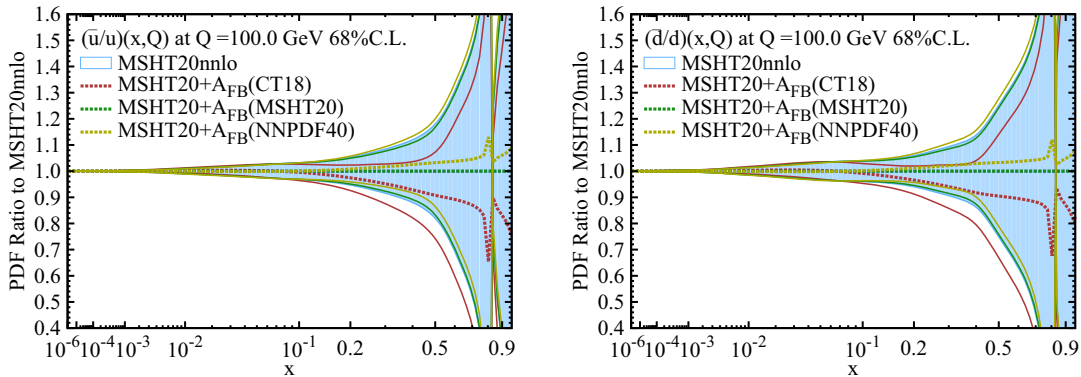


FIG. 9. PDF update for  $\bar{u}/u$  (left), and  $\bar{d}/d$  (right) using  $A_{FB}$  pseudodata generated using CT18, MSHT20, and NNPDF40. The central value and uncertainty are compared to the MSHT20 central value.

range of dilepton invariant mass studied, we investigated splitting up the range into smaller chunks (0.5–1.0 TeV, 1.0–3.0 TeV, and 3.0–5.0 TeV). Using each of these ranges separately to perform the PDF updating it was found that the 0.5–1.0 TeV region contributes the overwhelming sensitivity, with around 20% of the sensitivity coming from the range 1.0–3.0 TeV, and a small contribution from the 3.0–5.0 TeV range. This was not completely unexpected given the number of events expected in each region for this integrated luminosity.

One could also make the argument against using high mass data in global PDF fits for the fear of somehow absorbing new physics into Standard Model (SM) PDFs. However, even low-mass data could hide new physics in the tails of some distributions given enough data, and at some point SM-only PDFs would not be able to describe such a departure from the SM. Furthermore, given the different PDF behaviors observed in this study, the community should be aware that some PDFs (such as NNPDF4.0) might predict distributions of  $A_{\text{FB}}$  versus dilepton invariant mass that are similar to many nonresonant new physics models (predicting high-mass drop-off in  $A_{\text{FB}}$ ). It should be noted that the choice of tolerance when using `ePump` is important as it causes the new input data to have a stronger (for lower tolerances) or weaker (for higher tolerances) effect on the resulting PDF update. Choosing a tolerance that accurately reflects the original PDF fit is vital to ensure a realistic resulting update.

Further studies are planned for specifically high-mass Drell-Yan searches to explore the notion that “boutique” PDFs created for specific uses might be developed using specified measurables to help constrain SM backgrounds in regions where, without more precise PDF modeling, new physics might lay hidden. Finally, we note that when real data become available, a full global fit has to be carried out to explore the need of new nonperturbative functional forms of the PDFs at the initial  $Q_0$  scale to better describe the data, especially when the data provides new constraints on PDFs at very large- or small- $x$  regions.

## ACKNOWLEDGMENTS

We thank Liang Han and Siqi Yang for discussions. This work is partially supported by the U.S. National Science Foundation under Grants No. PHY-2013791, No. PHY-2111226, and No. PHY-2310291. C. P. Yuan is also grateful for the support from the Wu-Ki Tung endowed chair in particle physics.

## APPENDIX: TOLERANCE DEPENDENCE IN PDF UPDATING

In this section the impact of tolerance choices in the PDF updating method is discussed. In global QCD analysis,

PDFs are obtained by minimizing the  $\chi^2$  function. The uncertainties are defined in the relevant neighborhood of the global minimum as

$$\Delta\chi^2 \leq T^2, \quad (\text{A1})$$

where  $T$  is the tolerance parameter. In the Hessian method, the PDFs are parametrized by  $N$  parameters  $\{z_i; i = 1, N\}$ . The  $\chi^2$  function can be approximated by a quadratic expansion by the parameters  $z$ , so that the  $\chi^2$  function can be written as

$$\Delta\chi^2 = T^2 \sum_{i=1}^N z_i^2. \quad (\text{A2})$$

In the `ePump` package, theory templates need to be provided as part of the PDF updating procedure. However, `ePump` does not know what the tolerance parameter corresponds to for a given theory template. Users must provide an input of a fixed tolerance parameter  $T$ , or a set of dynamical tolerances  $\{T_i\}$ . The tolerance parameter is used to tell the `ePump` package how large the deviation from the global minimum is for a given theory template, which is vital to ensure that the PDF update due to the new dataset is realistic. As shown in Sec. IV B, the original PDF uncertainty as a function of  $A_{\text{FB}}$  for the MSHT20 PDF is relatively small; however, this does not lead to a larger impact when we perform PDF updating with the same pseudodata statistical precision. As long as we input a corresponding tolerance associated with the original PDF uncertainty, the final updating result will make sense. If the tolerance parameter is smaller than the real one, it means that the PDF uncertainty of the given theory template is treated as if it deviated less from the global minimum than in reality. In other words, a smaller tolerance means that `ePump` gives a larger weight to the new dataset, leading to a more significant impact. Figure 10 shows the PDF updating result for CT18 using pseudo-data generated by NNPDF4.0, under different tolerance choices. The result of using  $T = 1$ , as done in the default `xFitter` profiling [25], gives a much stronger impact in the update compared to  $T = 10$ , using the same pseudodata input. Much more detailed discussions can be found in the Appendix F of [11].

For CT18 the dynamical tolerances are used in the global QCD analysis. The effective tolerance parameter is about 100 for 90% C.L., or equivalently 37 for 68% C.L. If the tolerance parameter is set to 1 for the 68% C.L., it effectively means that a weight of 37 is given to the new dataset. This will dramatically overestimate the impact of new dataset.



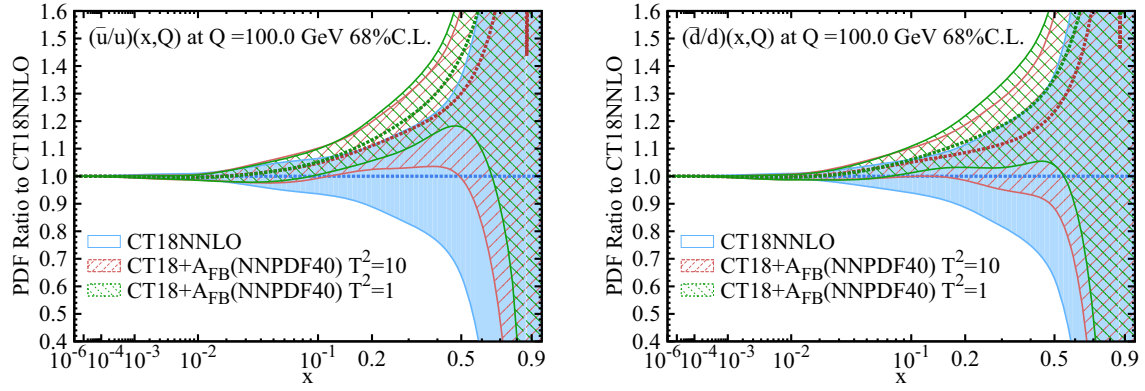


FIG. 10. PDF update for  $\bar{u}/u$  (left), and  $\bar{d}/d$  (right) using  $A_{\text{FB}}$  pseudodata generated using NNPDF4.0. The central value and uncertainty are compared to the CT18 central value. The tolerance is set to be 10 and 1, respectively.

- 
- [1] C. Willis, R. Brock, D. Hayden, T.-J. Hou, J. Isaacson, C. Schmidt, and C.-P. Yuan, *Phys. Rev. D* **99**, 054004 (2019).
  - [2] C. Schmidt, J. Pumplin, and C. P. Yuan, *Phys. Rev. D* **98**, 094005 (2018).
  - [3] J. Fiaschi, F. Giuliani, F. Hautmann, S. Moch, and S. Moretti, *Phys. Lett. B* **841**, 137915 (2023).
  - [4] J. Fiaschi, F. Giuliani, F. Hautmann, and S. Moretti, *J. High Energy Phys.* **02** (2022) 179.
  - [5] J. Fiaschi, F. Giuliani, F. Hautmann, and S. Moretti, *Nucl. Phys.* **B968**, 115444 (2021).
  - [6] S. Amoroso, J. Fiaschi, F. Giuliani, A. Glazov, F. Hautmann, and O. Zenaiev, *Phys. Lett. B* **821**, 136613 (2021).
  - [7] E. Accomando *et al.*, *J. High Energy Phys.* **10** (2019) 176.
  - [8] E. Accomando, J. Fiaschi, F. Hautmann, and S. Moretti, *Eur. Phys. J. C* **78**, 663 (2018); **79**, 453(E) (2019).
  - [9] E. Accomando, J. Fiaschi, F. Hautmann, and S. Moretti, *Phys. Rev. D* **98**, 013003 (2018); **99**, 079902(E) (2019).
  - [10] R. Abdul Khalek *et al.*, arXiv:2203.13199.
  - [11] T.-J. Hou *et al.*, *Phys. Rev. D* **103**, 014013 (2021).
  - [12] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, *Eur. Phys. J. C* **81**, 341 (2021).
  - [13] R. D. Ball *et al.* (NNPDF Collaboration), *Eur. Phys. J. C* **82**, 428 (2022).
  - [14] X. Jing *et al.*, *Phys. Rev. D* **108**, 034029 (2023).
  - [15] S. Yang, Y. Fu, M. Liu, L. Han, T.-J. Hou, and C.-P. Yuan, *Phys. Rev. D* **106**, 033001 (2022).
  - [16] J. C. Collins and D. E. Soper, *Phys. Rev. D* **16**, 2219 (1977).
  - [17] Y. Fu, S. Yang, M. Liu, L. Han, T.-J. Hou, C. Schmidt, C. Wang, and C. P. Yuan, *Chin. Phys. C* **45**, 053001 (2021).
  - [18] S. Yang, Y. Fu, M. Liu, R. Zhang, T.-J. Hou, C. Wang, H. Yin, L. Han, and C. P. Yuan, *Eur. Phys. J. C* **82**, 368 (2022).
  - [19] J. Campbell and T. Neumann, *J. High Energy Phys.* **12** (2019) 034.
  - [20] T. Carli, D. Clements, A. Cooper-Sarkar, C. Gwenlan, G. P. Salam, F. Siegert, P. Starovoitov, and M. Sutton, *Eur. Phys. J. C* **66**, 503 (2010).
  - [21] T.-J. Hou, H.-W. Lin, M. Yan, and C. P. Yuan, *Phys. Rev. D* **107**, 076018 (2023).
  - [22] C. Balazs and C. P. Yuan, *Phys. Rev. D* **56**, 5558 (1997).
  - [23] J. Isaacson, ResBos2: Precision resummation for the LHC era, Ph.D. thesis, Michigan State University, 2017.
  - [24] R. D. Ball, A. Candido, S. Forte, F. Hekhorn, E. R. Nocera, J. Rojo, and C. Schwan, *Eur. Phys. J. C* **82**, 1160 (2022).
  - [25] S. Alekhin *et al.*, *Eur. Phys. J. C* **75**, 304 (2015).