

# The intrinsic charm quark valence distribution of the proton

Richard D. Ball,<sup>1</sup> Alessandro Candido,<sup>2,3</sup> Juan Cruz-Martinez,<sup>3</sup> Stefano Forte,<sup>2</sup> Tommaso Giani,<sup>4,5</sup>  
Felix Hekhorn,<sup>2,6,7</sup> Giacomo Magni,<sup>4,5</sup> Emanuele R. Nocera,<sup>8</sup> Juan Rojo,<sup>4,5</sup> and Roy Stegeman<sup>1</sup>

(The NNPDF Collaboration)

<sup>1</sup>*The Higgs Centre for Theoretical Physics, University of Edinburgh,  
JCMB, KB, Mayfield Rd, Edinburgh EH9 3JZ, Scotland*

<sup>2</sup>*Tif Lab, Dipartimento di Fisica, Università di Milano and INFN,  
Sezione di Milano, Via Celoria 16, I-20133 Milano, Italy*

<sup>3</sup>*CERN, Theoretical Physics Department, CH-1211 Geneva 23, Switzerland*

<sup>4</sup>*Department of Physics and Astronomy, Vrije Universiteit, NL-1081 HV Amsterdam*

<sup>5</sup>*Nikhef Theory Group, Science Park 105, 1098 XG Amsterdam, The Netherlands*

<sup>6</sup>*University of Jyväskylä, Department of Physics,*

*P.O. Box 35, FI-40014 University of Jyväskylä, Finland*

<sup>7</sup>*Helsinki Institute of Physics, P.O. Box 64, FI-00014 University of Helsinki, Finland*

<sup>8</sup>*Dipartimento di Fisica, Università degli Studi di Torino and INFN,  
Sezione di Torino, Via Pietro Giuria 1, I-10125 Torino, Italy*

(Dated: March 18, 2024)

We provide a first quantitative indication that the wave function of the proton contains unequal distributions of charm quarks and antiquarks, i.e. a nonvanishing intrinsic valence charm distribution. A significant nonvanishing valence component cannot be perturbatively generated, hence our results reinforce previous evidence that the proton contains an intrinsic (i.e., not radiatively generated) charm quark component. We establish our result through a determination of the parton distribution functions (PDFs) of charm quarks and antiquarks in the proton. We propose two novel experimental probes of this intrinsic charm valence component:  $D$ -meson asymmetries in  $Z+c$ -jet production at the LHCb experiment, and flavor-tagged structure functions at the Electron-Ion Collider.

**Introduction.** The possible existence of charm quarks as intrinsic constituents of the proton, on the same footing as the much lighter up, down, and strange quarks, has fascinated physicists for more than four decades [1, 2]. Charm quarks and antiquarks are heavier ( $m_c \sim 1.5$  GeV) than the proton itself ( $m_p \sim 1$  GeV). They are copiously pair-produced through the perturbative QCD radiation of gluons and light quarks that generates their scale dependence. An intrinsic charm (IC) component is the scale-independent result that is left after subtracting this radiative contribution.

A plethora of experimental and theoretical studies have tried to either identify or reject the presence of IC in the proton [3–6]. We have recently presented a determination of intrinsic charm in the proton from a global analysis of parton distribution functions (PDFs) [7–9]. This study found evidence for IC at the  $3\sigma$  level, and was supported by independent constraints from forward  $Z$  production with charm jets at the LHCb experiment [10].

In Ref. [9] we determined the distributions of charm quarks and antiquarks assuming equality of the intrinsic (scale-independent) charm and anticharm PDFs, i.e. the vanishing of the charm valence PDF

$$c^-(x, Q^2) = c(x, Q^2) - \bar{c}(x, Q^2). \quad (1)$$

The valence charm PDF  $c^-(x, Q^2)$  must have vanishing integral over  $x$  at all scales  $Q^2$ , because the proton does not carry the charm quantum number, but the PDF itself may well be nonzero, as it happens for the strange valence

PDF  $s^- = s - \bar{s}$ . Indeed, a nonvanishing charm valence component is always generated, like for any other quark flavor, by perturbative QCD evolution [11]. However, any perturbatively generated valence charm component is tiny in comparison to all other PDFs, including those of heavy quarks. Hence, any evidence of a sizable valence charm PDF is a definite sign of its intrinsic nature. Model calculations [2, 12], while in broad agreement on the shape of total IC PDF, widely differ in predictions for the shape and magnitude of the intrinsic valence charm component. Model calculations of IC complemented with input from lattice QCD [13] also predict a non-vanishing valence component.

Here we investigate this issue by performing a data-driven determination of the intrinsic valence charm PDF of the proton, based on the same methodology as in [9]. We generalize the NNPDF4.0 PDF determination by introducing an independent parametrization of the charm and anticharm PDFs, determine them from a global QCD analysis, and subtract the perturbatively generated contributions by transforming all PDFs to the three-flavor-number scheme (3FNS) in which perturbative charm vanishes so any residual charm PDF is intrinsic.

We find a non-zero charm valence PDF, with a positive valence peak for  $x \sim 0.3$ , whose local significance is close to two sigma. We demonstrate the stability of this result with respect to theoretical, dataset, and methodological variations. We then propose two novel experimental probes to further scrutinize this asymmetry between

charm and anticharm PDFs:  $D$ -meson asymmetries in  $Z+c$ -jet production at LHCb [10, 14], and flavor-tagged structure functions at the upcoming Electron-Ion Collider (EIC) [15, 16].

**Methodology.** As in Ref. [9], we follow the NNPDF4.0 methodology, theory settings and dataset [8], the only modifications being related to the independent parametrization of the charm valence PDF. Firstly, the neural network architecture is extended with an additional neuron in the output layer in order to independently parametrize  $c^-(x, Q_0)$ , Eq. (1), at the PDF parametrization scale  $Q_0 = 1.65$  GeV. In the default PDF basis (“evolution basis”, see App. B) this extra neuron is taken to parametrize the valence non-singlet combination  $V_{15} = (u^- + d^- + s^- - 3c^-)$ , with  $q^- \equiv q - \bar{q}$ . In an alternative basis (“flavor basis”) it instead parametrizes  $\bar{c}$ : so in both cases the valence component is obtained by taking linear combinations of the neural network outputs. In our previous analysis [8], the assumption of vanishing intrinsic valence was enforced by setting  $V_{15} = V = \sum_i q_i^-$  in the evolution basis or  $\bar{c} = c$  in the flavor basis at the scale  $Q_0$ .

In addition to experimental constraints, a non-zero charm valence must, as mentioned, satisfy the sum rule

$$Q_{15} \equiv \int_0^1 dx V_{15}(x, Q_0) = 3, \quad (2)$$

$$Q_c \equiv \int_0^1 dx (c - \bar{c})(x, Q_0) = 0, \quad (3)$$

in the evolution or flavor basis respectively. This sum rule is enforced in the same manner as that of the strange valence sum rule [8]. Finally, to ensure cross-section positivity (at  $Q^2 = 5$  GeV<sup>2</sup>) separately for charm- and anticharm-initiated processes, we replace the neutral current  $F_2^c$  positivity observable (sensitive only to  $c^+$ ) with its charged current-counterparts  $F_2^{c,W^-}$  and  $F_2^{\bar{c},W^+}$ . The charm PDFs  $xc$  and  $x\bar{c}$  themselves are not required to be positive-definite [17–19]. Integrability and preprocessing are imposed as in NNPDF4.0. We have verified that results are stable upon repeating the hyperoptimization of all parameters defining the fitting algorithm, and thus we keep the same settings as in [8].

**The valence charm PDF.** As explained in Ref. [9], intrinsic charm is the charm PDF in the 3FNS, where charm is treated as a massive particle that does not contribute to the running of the strong coupling or the evolution of PDFs. In the absence of intrinsic charm (“perturbative charm”, henceforth), the charm and anticharm PDFs in the 3FNS vanish identically. In the four-flavor-number scheme (4FNS), in which charm is treated as a massless parton, these PDFs are determined by perturbative matching conditions between the 3FNS and the 4FNS [20]. In our data-driven approach, the charm and anticharm PDFs, instead of being fixed by perturbative

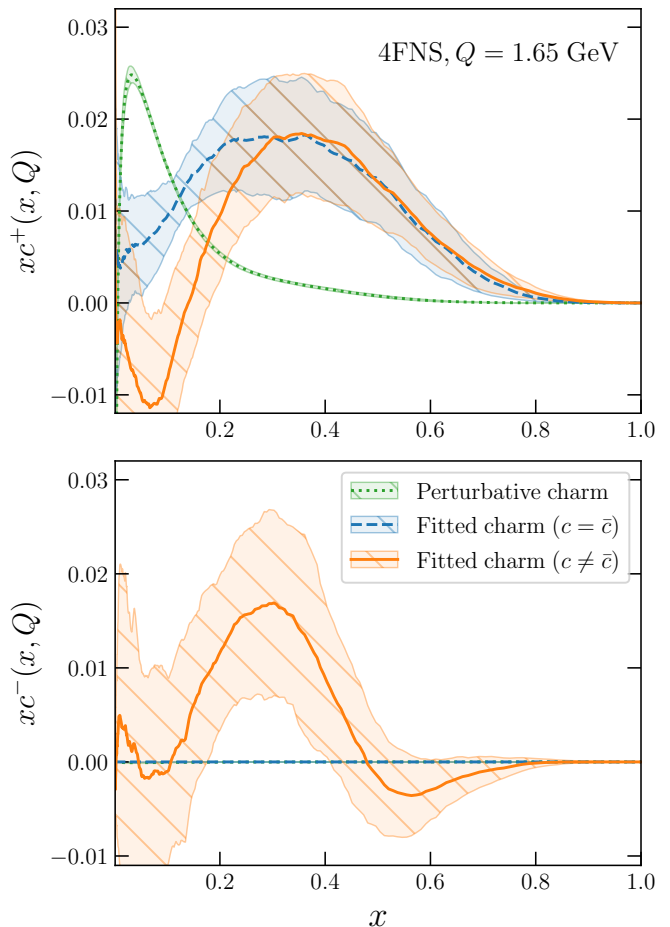


FIG. 1: The charm total  $xc^+$  (top) and valence  $xc^-$  (bottom) PDFs in the 4FNS at  $Q = 1.65$  GeV. The perturbative and data-driven results are compared, in the latter case either assuming  $c^- = 0$  (as in [9]) or  $c^-$  determined from data.

matching conditions, are determined from data on the same footing as the light quark PDFs. The deviation of data-driven charm from perturbative charm, i.e., in the 3FNS the deviation of the charm and anticharm PDFs from zero, is identified with the intrinsic component. In practice, we parametrize PDFs at  $Q_0 = 1.65$  GeV in the 4FNS, and then invert the matching conditions to determine the intrinsic component in the 3FNS.

In Fig. 1 we show  $xc^+$  and  $xc^-$  in the 4FNS at  $Q = 1.65$  GeV, i.e. just above the charm mass that we take to be  $m_c = 1.51$  GeV, determined using next-to-next-to leading order (NNLO) QCD theory. The bands are 68% confidence level (CL) PDF uncertainties. We show both the purely perturbative and data-driven results, in the latter case both for  $c = \bar{c}$  (same as in [9]) and  $c \neq \bar{c}$ . Note that the purely perturbative valence PDF vanishes at  $Q = m_c$  at NNLO, and only develops a tiny component at one extra perturbative order (N<sup>3</sup>LO), or at higher scales. Hence, a nonvanishing valence component in the 4FNS provides already evidence for intrinsic charm.

Upon allowing for a vanishing valence  $xc^-$  component,

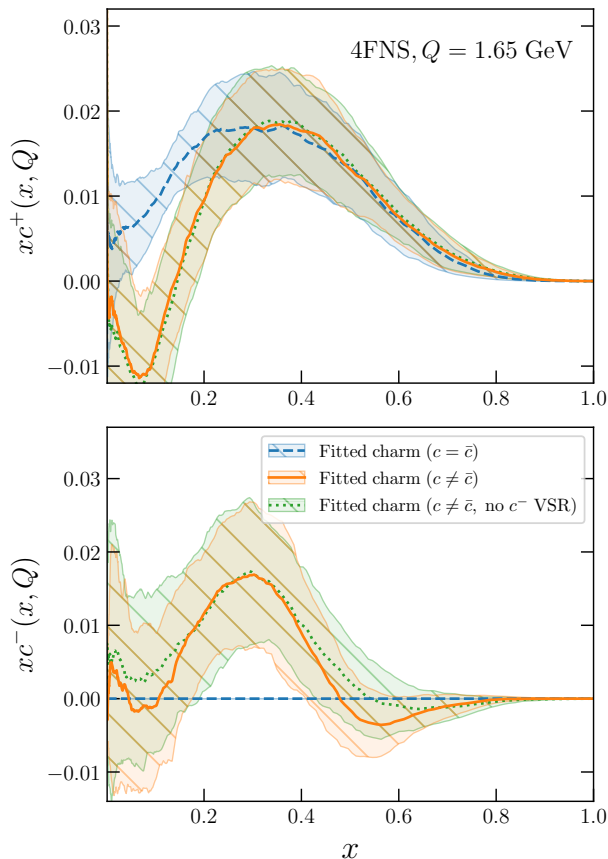


FIG. 2: Same as Fig. 2, now without imposing the charm valence sum rule Eq. (2) when  $c \neq \bar{c}$ .

the total charm  $xc^+$  is quite stable, especially around the peak at  $x \sim 0.4$ . This total charm PDF is also somewhat suppressed for smaller  $x \lesssim 0.2$  as compared to the baseline result. In terms of fit quality, the  $\chi^2$  per data point for the global dataset decreases from 1.162 to 1.151, corresponding to an improvement by about 50 units in absolute  $\chi^2$ . The main contributions to this decrease comes from neutral current deep-inelastic scattering and LHC gauge boson production data (see App. A).

The valence component is nonzero and positive at more than one sigma level in the  $x \in [0.2, 0.4]$  region, and consistent with zero within the large PDF uncertainties elsewhere. The size and shape of the valence charm PDF seen in Fig. 1 are stable upon variations of PDF parametrization basis (App. B), the value of  $m_c$ , (App. C), the input dataset (App. D), and the kinematic cuts in  $W^2$  and  $Q^2$  (App. E). All other PDFs are mostly left unaffected by having allowed for a nonvanishing valence charm.

Whereas in our default determination we have imposed the charm valence sum rule Eq. (2), we have also repeated our determination without imposing this theoretical constraint. We then obtain  $Q_c = 0.07 \pm 0.14$  and the resulting charm PDFs are shown in Fig. 2. This

result demonstrates that the valence sum rule is actually enforced by the data, and our result is data-driven.

**Intrinsic valence charm.** The intrinsic valence charm PDF is now determined by transforming back to the 3FNS scheme, and is displayed in Fig. 3 (upper panel), together with its 4FNS counterpart already shown in Fig. 1. An estimate of the missing higher order uncertainties (MHOU) related to the truncation of the perturbative expansion is also included. This, as in [9], is estimated as the change in the 3FNS PDF when the transformation from the 4FNS to the 3FNS is performed to one higher perturbative order, i.e. N<sup>3</sup>LO [21–29], as this is estimated to be the dominant missing higher order correction.

The 3FNS and 4FNS valence PDFs turn out to be quite close, implying that for the valence PDF, unlike for the total charm PDF, the theory uncertainty is smaller than the PDF uncertainty. We thus find that the intrinsic (3FNS) charm valence is nonzero and positive roughly in the same  $x$  region as its 4FNS counterpart.

The statistical significance of the nonvanishing valence is quantified by the pull, defined as the median PDF in units of the total uncertainty, shown in Fig. 3 (bottom). The local significance of the intrinsic valence is slightly below two sigma, peaking at  $x \sim 0.3$ . The significance of the total intrinsic component is similar to that found in Ref. [9], namely about three sigma for  $x \sim 0.5$ . As in Ref. [9], we also show the results found in fit variants including the EMC  $F_2^c$  [30] and LHCb  $Z + c$  data [10], both of which increase the local significance.

The results of Figs. 1-3 suggest that the intrinsic valence component may be nonzero, but their significance falls below the three sigma evidence level. We thus propose two novel experimental observables engineered to probe this valence charm component.

**Charm asymmetries in  $Z + c$  at LHCb.** The LHCb LHC Run 2 data, which, as shown in Ref. [9], reinforce the evidence for an intrinsic total charm component, correspond to measurements of forward  $Z$  production in association with charm-tagged jets [10]. They are presented as a measurement of  $\mathcal{R}_j^c(y_Z)$ , the ratio between  $c$ -tagged and untagged jets in bins of the  $Z$ -boson rapidity  $y_Z$ , and they are obtained from tagging  $D$ -mesons from displaced vertices. The higher statistics available first at Runs 3 and 4 and later at the HL-LHC will enable the reconstruction of the exclusive decays of  $D$ -mesons, and thus the separation of charm and anticharm-tagged final states. We thus define the asymmetry

$$\mathcal{A}_c(y_Z) \equiv \frac{N_j^c(y_Z) - N_j^{\bar{c}}(y_Z)}{N_j^c(y_Z) + N_j^{\bar{c}}(y_Z)}, \quad (4)$$

where  $N_j^c$  ( $N_j^{\bar{c}}$ ) is defined in the same manner as  $\mathcal{R}_j^c$  [10], but now restricted to events with  $D$ -mesons containing a charm quark (antiquark). This asymmetry is directly sensitive to a possible difference between the charm and anticharm PDFs in the initial state.

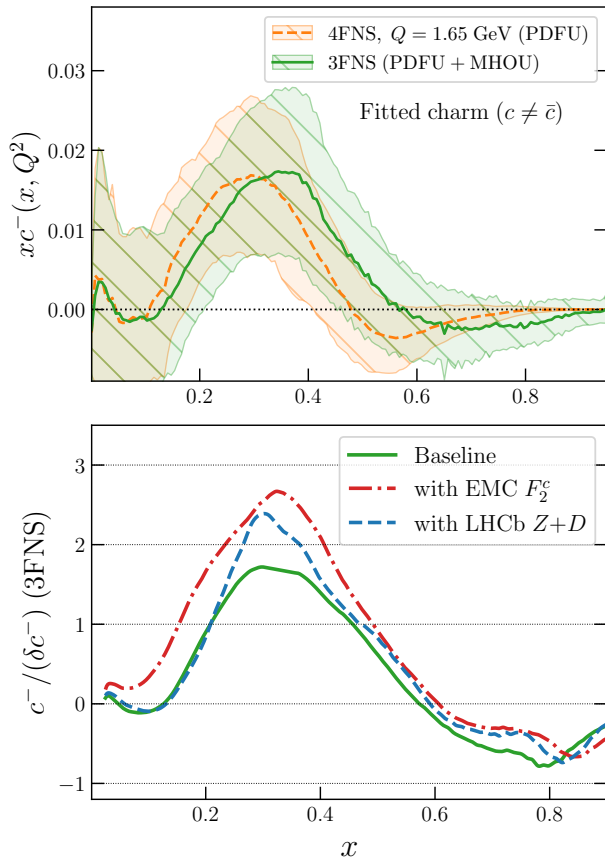


FIG. 3: Top: the 3FNS (intrinsic) valence charm PDF  $xc^-$ , compared to the 4FNS result (same as Fig. 1 bottom). The 3FNS also includes MHO due to the inversion from the 4FNS to the 3FNS. Bottom: the pull for valence  $xc^-$  charm PDF in the 3FNS. Results are shown both for the default fit and also when including the EMC  $F_2^c$  and LHCb  $Z+c$  data

In Fig. 4 we display the asymmetry  $\mathcal{A}_c(y_Z)$ , Eq. (4), computed for  $\sqrt{s} = 13$  TeV using the PDFs determined here, that allow for a nonvanishing valence component, as well as the default NNPDF4.0 with  $c = \bar{c}$ . Results are computed using MG5\_AMC@NLO [31] at leading order (LO) matched to PYTHIA8 [32, 33], with the same  $D$ -meson tagging and jet-reconstruction algorithm as in [10, 14]. The leading order parton-level result is also shown.

It is apparent from Fig. 4 that, even though the forward-backward asymmetry of the  $Z$  decay generates a small asymmetry  $\mathcal{A}_c \neq 0$  even when  $c = \bar{c}$  [34, 35], the LO effect due to an asymmetry between  $c$  and  $\bar{c}$  PDFs is much larger, and stable upon showering and hadronization corrections. Indeed, higher-order QCD corrections largely cancel in the ratio  $\mathcal{A}_c(y_Z)$ .

In Fig. 4 we also display projected uncertainties for the LHCb measurement of this asymmetry at Run 3 and at the HL-LHC (see App. F for details), showing that a valence component of the same size as our central

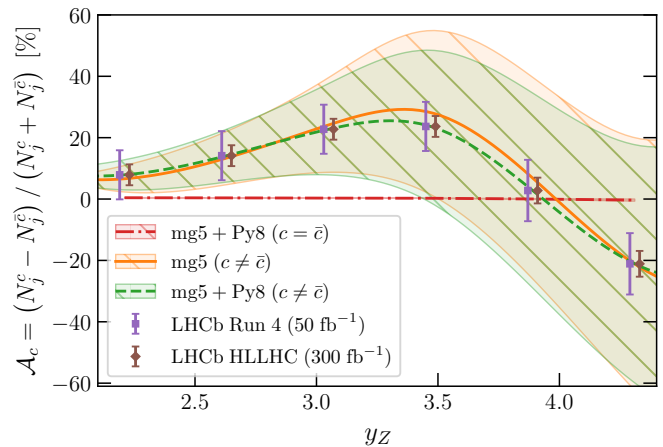


FIG. 4: The charm asymmetry  $\mathcal{A}_c(y_Z)$ , Eq. (4), in  $Z+c$ -jet production at LHCb ( $\sqrt{s} = 13$  TeV) evaluated at LO matched to parton showers with the nonvanishing valence PDF determined here. The pure LO result and the result with vanishing charm valence are also shown for comparison. The bands correspond to one-sigma PDF uncertainties. Projected statistical uncertainties for LHCb measurements at Run 4 ( $\mathcal{L} = 50 \text{ fb}^{-1}$ ) and the HL-LHC ( $\mathcal{L} = 300 \text{ fb}^{-1}$ ) are also shown.

prediction could be detected respectively at about a two sigma or four sigma level.

**Charm-tagged DIS at the EIC.** A standard probe of the charm component of the proton is the deep-inelastic charm structure function  $F_2^c$  [30, 36–38] and the associate deep-inelastic reduced charm production cross-section  $\sigma_{\text{red}}^{c\bar{c}}$ . Correspondingly, the charm valence can be determined from the reduced cross-section asymmetry

$$\mathcal{A}_{\sigma^{c\bar{c}}}(x, Q^2) \equiv \frac{\sigma_{\text{red}}^c(x, Q^2) - \sigma_{\text{red}}^{\bar{c}}(x, Q^2)}{\sigma_{\text{red}}^{c\bar{c}}(x, Q^2)}. \quad (5)$$

A measurement of this observable requires reconstructing final-state  $D$ -mesons by identifying their decay products. At the future EIC this will be possible with good precision using the proposed ePIC detector [15, 39, 40].

The predicted asymmetry  $\mathcal{A}_{\sigma^{c\bar{c}}}$  at  $Q^2 = 20 \text{ GeV}^2$  is shown in Fig. 5; results are shown at the reduced charm (parton) cross-section level, evaluated with YADISM [41] at NNLO accuracy. As in Fig. 4, we show results obtained both using the PDFs determined here, that allow for a nonvanishing valence component, as well as the default NNPDF4.0 with  $c = \bar{c}$ . We also display the projected statistical uncertainties [15] at the EIC running at  $\sqrt{s} = 63 \text{ GeV}$  for  $\mathcal{L} = 10 \text{ fb}^{-1}$  (see App. G). It is clear that a nonvanishing charm valence component can be measured at the EIC to very high significance even for a moderate amount of integrated luminosity.

In addition to the charm-tagged structure function  $F_2^{c\bar{c}}$ , at the EIC complementary sensitivity to the charm valence content of the proton would be provided by the charm-tagged parity-violating structure function  $x F_3^{c\bar{c}}(x, Q^2)$ . This observable has the advantage that

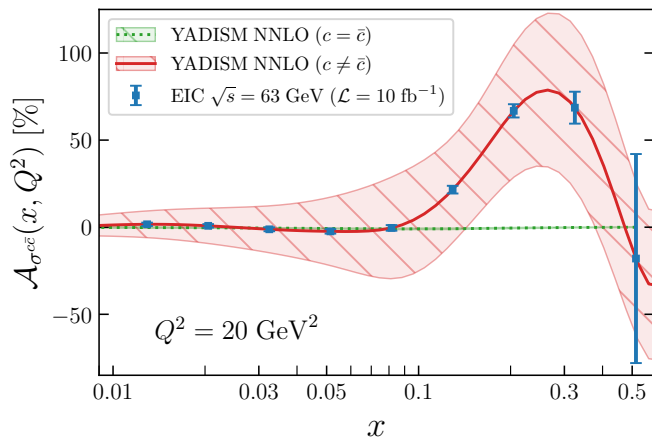


FIG. 5: The reduced charm-tagged cross-section asymmetry  $\mathcal{A}_{\sigma^{c\bar{c}}}$ , Eq. (5), at  $Q^2 = 20 \text{ GeV}^2$  computed at NNLO QCD using the nonvanishing valence PDF determined here. The result with vanishing charm valence is also shown for comparison. The bands correspond to one-sigma PDF uncertainties. The projected statistical uncertainties at the EIC [15] (running at  $\sqrt{s} = 63 \text{ GeV}$  for  $\mathcal{L} = 10 \text{ fb}^{-1}$ ) are also shown.

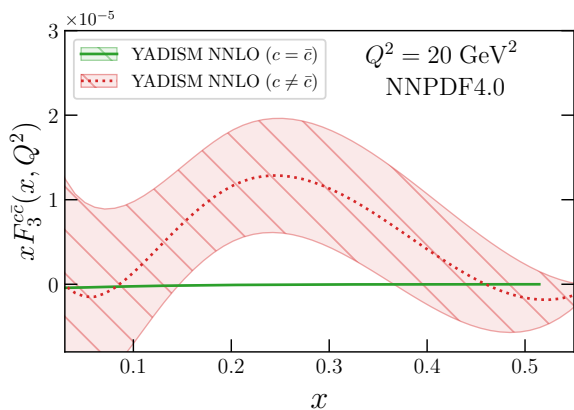


FIG. 6: Same as Fig. 5 for the charm-tagged parity-violating structure function  $xF_3^{c\bar{c}}(x, Q^2)$  at the EIC (no projection for the statistical accuracy of the EIC measurement is available).

at LO is already proportional to  $xc^-$ , hence provides a direct constraint on valence charm. Predictions for this observable, are presented in Fig. 6. Even in the absence of detailed predictions for prospective EIC measurements of this observable, it is clear that its measurement would significantly constrain the charm valence PDF.

**Outlook.** Our main conclusion is that current experimental data provide support for the hypothesis that the valence charm PDF may be nonzero, even though with the NNPDF4.0 dataset it is not possible to reach three-sigma evidence. Whereas the situation may improve somewhat with future PDF determinations based on the full LHC Run-3 dataset, dedicated observables

such as the LHCb charm asymmetry Eq. (4) as well as charm production at the EIC Eq. (5) will be needed in order to achieve firm evidence or discovery. Other experimental probes that could be explored in this context include open charm production and asymmetries at the LHC, in particular for forward (LHCb [42, 43]) and far-forward (FASER $\nu$  [44], SND@LHC [45], and the Forward Physics Facility [46, 47]) detectors. Progress in lattice computations might well also provide further constraints.

From the theory point of view, ongoing efforts towards a NNPDF determination based on N<sup>3</sup>LO calculations should reduce some of the theory uncertainties affecting the current determination. On a more speculative vein, it might also be interesting to investigate an intrinsic bottom quark component and its eventual asymmetry.

*Acknowledgments.* We are grateful to Rhorry Gauld for extensive discussions concerning  $Z$ +charm production, and to Reynier Cruz-Torres and Barak Schmookler by providing the EIC projections for  $F_2^c$ . We thank Thomas Boettcher, Dan Craik, Philip Ilten, Patrick Koppenburg, Niels Tuning, and Michael Williams, for information about the LHCb  $Z$ +charm measurements and the associated projections. R. D. B. and R. S. are supported by the U.K. Science and Technology Facility Council (STFC) grant ST/T000600/1. T. G. is supported by NWO (Dutch Research Council) via an ENW-KLEIN-2 project. F. H. is supported by the Academy of Finland project 358090 and is funded as a part of the Center of Excellence in Quark Matter of the Academy of Finland, project 346326. E.R. N. is supported by the Italian Ministry of University and Research (MUR) through the “Rita Levi-Montalcini” Program. J. R. and G. M. are partially supported by NWO (Dutch Research Council).

- 
- [1] S. J. Brodsky, P. Hoyer, C. Peterson, and N. Sakai, *The Intrinsic Charm of the Proton*, *Phys. Lett.* **B93** (1980) 451–455.
  - [2] S. J. Brodsky, A. Kusina, F. Lyonnet, I. Schienbein, H. Spiesberger, and R. Vogt, *A review of the intrinsic heavy quark content of the nucleon*, *Adv. High Energy Phys.* **2015** (2015) 231547, [[arXiv:1504.06287](#)].
  - [3] P. Jimenez-Delgado, T. Hobbs, J. Londergan, and W. Melnitchouk, *New limits on intrinsic charm in the nucleon from global analysis of parton distributions*, *Phys.Rev.Lett.* **114** (2015), no. 8 082002, [[arXiv:1408.1708](#)].
  - [4] **NNPDF** Collaboration, R. D. Ball, V. Bertone, M. Bonvini, S. Carrazza, S. Forte, A. Guffanti, N. P. Hartland, J. Rojo, and L. Rottoli, *A Determination of the Charm Content of the Proton*, *Eur. Phys. J.* **C76** (2016), no. 11 647, [[arXiv:1605.06515](#)].
  - [5] M. Guzzi, T. J. Hobbs, K. Xie, J. Huston, P. Nadolsky, and C. P. Yuan, *The persistent nonperturbative charm enigma*, *Phys. Lett. B* **843** (2023) 137975, [[arXiv:2211.01387](#)].
  - [6] T.-J. Hou, S. Dulat, J. Gao, M. Guzzi, J. Huston,

- P. Nadolsky, C. Schmidt, J. Winter, K. Xie, and C. P. Yuan, *CT14 Intrinsic Charm Parton Distribution Functions from CTEQ-TEA Global Analysis*, *JHEP* **02** (2018) 059, [[arXiv:1707.00657](#)].
- [7] **NNPDF** Collaboration, R. D. Ball et al., *An open-source machine learning framework for global analyses of parton distributions*, *Eur. Phys. J. C* **81** (2021), no. 10 958, [[arXiv:2109.02671](#)].
- [8] **NNPDF** Collaboration, R. D. Ball et al., *The path to proton structure at 1% accuracy*, *Eur. Phys. J. C* **82** (2022), no. 5 428, [[arXiv:2109.02653](#)].
- [9] **NNPDF** Collaboration, R. D. Ball, A. Candido, J. Cruz-Martinez, S. Forte, T. Giani, F. Hekhorn, K. Kudashkin, G. Magni, and J. Rojo, *Evidence for intrinsic charm quarks in the proton*, *Nature* **608** (2022), no. 7923 483–487, [[arXiv:2208.08372](#)].
- [10] **LHCb** Collaboration, R. Aaij et al., *Study of Z Bosons Produced in Association with Charm in the Forward Region*, *Phys. Rev. Lett.* **128** (2022), no. 8 082001, [[arXiv:2109.08084](#)].
- [11] S. Catani, D. de Florian, G. Rodrigo, and W. Vogelsang, *Perturbative generation of a strange-quark asymmetry in the nucleon*, *Phys. Rev. Lett.* **93** (2004) 152003, [[hep-ph/0404240](#)].
- [12] T. J. Hobbs, J. T. Londergan, and W. Melnitchouk, *Phenomenology of nonperturbative charm in the nucleon*, *Phys. Rev. D* **89** (2014), no. 7 074008, [[arXiv:1311.1578](#)].
- [13] R. S. Sufian, T. Liu, A. Alexandru, S. J. Brodsky, G. F. de Téramond, H. G. Dosch, T. Draper, K.-F. Liu, and Y.-B. Yang, *Constraints on charm-anticharm asymmetry in the nucleon from lattice QCD*, *Phys. Lett. B* **808** (2020) 135633, [[arXiv:2003.01078](#)].
- [14] T. Boettcher, P. Ilten, and M. Williams, *Direct probe of the intrinsic charm content of the proton*, *Phys. Rev. D* **93** (2016), no. 7 074008, [[arXiv:1512.06666](#)].
- [15] M. Kelsey, R. Cruz-Torres, X. Dong, Y. Ji, S. Radhakrishnan, and E. Sichtermann, *Constraints on gluon distribution functions in the nucleon and nucleus from open charm hadron production at the Electron-Ion Collider*, *Phys. Rev. D* **104** (2021), no. 5 054002, [[arXiv:2107.05632](#)].
- [16] R. Abdul Khalek et al., *Science Requirements and Detector Concepts for the Electron-Ion Collider: EIC Yellow Report*, *Nucl. Phys. A* **1026** (2022) 122447, [[arXiv:2103.05419](#)].
- [17] A. Candido, S. Forte, and F. Hekhorn, *Can  $\overline{MS}$  parton distributions be negative?*, *JHEP* **11** (2020) 129, [[arXiv:2006.07377](#)].
- [18] J. Collins, T. C. Rogers, and N. Sato, *Positivity and renormalization of parton densities*, *Phys. Rev. D* **105** (2022), no. 7 076010, [[arXiv:2111.01170](#)].
- [19] A. Candido, S. Forte, T. Giani, and F. Hekhorn, *On the positivity of  $M\overline{S}$  parton distributions*, [arXiv:2308.00025](#).
- [20] M. Buza and W. L. van Neerven,  *$O(\alpha_s^{**2})$  contributions to charm production in charged-current deep-inelastic lepton hadron scattering*, *Nucl. Phys. B* **500** (1997) 301–324, [[hep-ph/9702242](#)].
- [21] I. Bierenbaum, J. Blümlein, and S. Klein, *The Gluonic Operator Matrix Elements at  $O(\alpha_s^2)$  for DIS Heavy Flavor Production*, *Phys. Lett. B* **672** (2009) 401–406, [[arXiv:0901.0669](#)].
- [22] I. Bierenbaum, J. Blümlein, and S. Klein, *Mellin Moments of the  $O(\alpha_s^3)$  Heavy Flavor Contributions to unpolarized Deep-Inelastic Scattering at  $Q^2 \gg m^2$  and Anomalous Dimensions*, *Nucl. Phys. B* **820** (2009) 417–482, [[arXiv:0904.3563](#)].
- [23] J. Ablinger, J. Blümlein, S. Klein, C. Schneider, and F. Wissbrock, *The  $O(\alpha_s^3)$  Massive Operator Matrix Elements of  $O(n_f)$  for the Structure Function  $F_2(x, Q^2)$  and Transversity*, *Nucl. Phys. B* **844** (2011) 26–54, [[arXiv:1008.3347](#)].
- [24] J. Ablinger, A. Behring, J. Blümlein, A. De Freitas, A. Hasselhuhn, A. von Manteuffel, M. Round, C. Schneider, and F. Wißbrock, *The 3-Loop Non-Singlet Heavy Flavor Contributions and Anomalous Dimensions for the Structure Function  $F_2(x, Q^2)$  and Transversity*, *Nucl. Phys. B* **886** (2014) 733–823, [[arXiv:1406.4654](#)].
- [25] J. Ablinger, J. Blümlein, A. De Freitas, A. Hasselhuhn, A. von Manteuffel, M. Round, and C. Schneider, *The  $O(\alpha_s^3 T_F^2)$  Contributions to the Gluonic Operator Matrix Element*, *Nucl. Phys. B* **885** (2014) 280–317, [[arXiv:1405.4259](#)].
- [26] A. Behring, I. Bierenbaum, J. Blümlein, A. De Freitas, S. Klein, and F. Wißbrock, *The logarithmic contributions to the  $O(\alpha_s^3)$  asymptotic massive Wilson coefficients and operator matrix elements in deeply inelastic scattering*, *Eur. Phys. J. C* **74** (2014), no. 9 3033, [[arXiv:1403.6356](#)].
- [27] J. Ablinger, J. Blümlein, A. De Freitas, A. Hasselhuhn, A. von Manteuffel, M. Round, C. Schneider, and F. Wißbrock, *The transition matrix element  $a_{gq}(n)$  of the variable flavor number scheme at  $o(\alpha_s^3)$* , *Nuclear Physics B* **882** (May, 2014) 263–288.
- [28] J. Ablinger, A. Behring, J. Blümlein, A. De Freitas, A. von Manteuffel, et al., *The 3-Loop Pure Singlet Heavy Flavor Contributions to the Structure Function  $F_2(x, Q^2)$  and the Anomalous Dimension*, [arXiv:1409.1135](#).
- [29] J. Blümlein, J. Ablinger, A. Behring, A. De Freitas, A. von Manteuffel, C. Schneider, and C. Schneider, *Heavy Flavor Wilson Coefficients in Deep-Inelastic Scattering: Recent Results*, *PoS QCDEV2017* (2017) 031, [[arXiv:1711.07957](#)].
- [30] **European Muon** Collaboration, J. J. Aubert et al., *Production of charmed particles in 250-GeV  $\mu^+$  - iron interactions*, *Nucl. Phys. B* **213** (1983) 31–64.
- [31] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **1407** (2014) 079, [[arXiv:1405.0301](#)].
- [32] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852–867, [[arXiv:0710.3820](#)].
- [33] P. Skands, S. Carrazza, and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 Tune*, *European Physical Journal* **74** (2014) 3024, [[arXiv:1404.5630](#)].
- [34] R. Gauld, U. Haisch, B. D. Pecjak, and E. Re, *Beauty-quark and charm-quark pair production asymmetries at LHCb*, *Phys. Rev. D* **92** (2015) 034007, [[arXiv:1505.02429](#)].
- [35] R. Gauld, U. Haisch, and B. D. Pecjak, *Asymmetric heavy-quark hadroproduction at LHCb: Predictions and applications*, *JHEP* **03** (2019) 166, [[arXiv:1901.07573](#)].
- [36] S. Forte, E. Laenen, P. Nason, and J. Rojo, *Heavy*

- quarks in deep-inelastic scattering, *Nucl. Phys.* **B834** (2010) 116–162, [arXiv:1001.2312].
- [37] R. D. Ball, M. Bonvini, and L. Rottoli, *Charm in Deep-Inelastic Scattering*, *JHEP* **11** (2015) 122, [arXiv:1510.02491].
- [38] **H1, ZEUS** Collaboration, H. Abramowicz et al., *Combination and QCD analysis of charm and beauty production cross-section measurements in deep inelastic ep scattering at HERA*, *Eur. Phys. J.* **C78** (2018), no. 6 473, [arXiv:1804.01019].
- [39] R. A. Khalek, J. J. Ethier, E. R. Nocera, and J. Rojo, *Self-consistent determination of proton and nuclear PDFs at the Electron Ion Collider*, *Phys. Rev. D* **103** (2021), no. 9 096005, [arXiv:2102.00018].
- [40] N. Armesto, T. Cridge, F. Giuli, L. Harland-Lang, P. Newman, B. Schmookler, R. Thorne, and K. Wichmann, *Impact of Inclusive Electron Ion Collider Data on Collinear Parton Distributions*, arXiv:2309.11269.
- [41] A. Candido, A. Garcia, G. Magni, T. Rabemananjara, J. Rojo, and R. Stegeman, *Neutrino Structure Functions from GeV to EeV Energies*, *JHEP* **05** (2023) 149, [arXiv:2302.08527].
- [42] **LHCb** Collaboration, R. Aaij et al., *Open charm production and asymmetry in pNe collisions at  $\sqrt{s_{NN}} = 68.5$  GeV*, *Eur. Phys. J. C* **83** (2023), no. 6 541, [arXiv:2211.11633].
- [43] **LHCb** Collaboration, R. Aaij et al., *First Measurement of Charm Production in its Fixed-Target Configuration at the LHC*, *Phys. Rev. Lett.* **122** (2019), no. 13 132002, [arXiv:1810.07907].
- [44] **FASER** Collaboration, H. Abreu et al., *First Direct Observation of Collider Neutrinos with FASER at the LHC*, *Phys. Rev. Lett.* **131** (2023), no. 3 031801, [arXiv:2303.14185].
- [45] **SND@LHC** Collaboration, R. Albanese et al., *Observation of Collider Muon Neutrinos with the SND@LHC Experiment*, *Phys. Rev. Lett.* **131** (2023), no. 3 031802, [arXiv:2305.09383].
- [46] J. L. Feng et al., *The Forward Physics Facility at the High-Luminosity LHC*, *J. Phys. G* **50** (2023), no. 3 030501, [arXiv:2203.05090].
- [47] J. M. Cruz-Martinez, M. Fieg, T. Giani, P. Krack, T. Mäkelä, T. Rabemananjara, and J. Rojo, *The LHC as a Neutrino-Ion Collider*, arXiv:2309.09581.

## Appendix

We collect here a number of supplementary results, none of which is required for the understanding of the main text, which is self-contained. The purpose of this Appendix is (a) to ensure reproducibility of our results and (b) to document the set of tests of stability and robustness that we have carried out. In particular: we document the fit quality for our default determination; we test for stability upon changes of parametrization basis, variations of the quark mass, choice of dataset and kinematic cuts; and we provide details of our projections for measurements at the HL-LHC and the EIC.

### Appendix A: Fit quality and data impact

We compare the fit quality for the PDF determination presented here with  $c \neq \bar{c}$  to the published NNPDF4.0 determination with  $c = \bar{c}$ , by showing in Table I the experimental  $\chi^2$  per data point for different groups of processes and for the total dataset. We refer to [8] for the definition of the dataset and of the  $\chi^2$  and the process categories (see in particular Table 5.1 of Ref. [8]).

The largest reduction in absolute  $\chi^2$  upon allowing for a nonvanishing charm valence component is in collider DIS (i.e. HERA), and the largest reduction in  $\chi^2$  in charged-current collider DIS. The largest impact is seen in the large  $Q^2$ , large  $x$   $e^+p$  CC bins such as shown in Fig. 7, consistent with the observation that the intrinsic charm PDFs are localized at large  $x$ . Note that HERA data for the  $F_2^c$  charm structure function, that are included in the fit, have no impact on intrinsic charm because they are in the medium-to-low  $x$  region where the charm PDF is dominated by the perturbative component.

### Appendix B: Parametrization basis dependence

The NNPDF4.0 set is determined by choosing by default as a parametrization basis for PDFs the eigenstates of QCD evolution; as a consistency check results in Ref. [8] are presented by alternatively choosing flavor and antiflavor eigenstates. Because here we are determining a difference between PDFs, it is especially important to check stability upon choice of basis. In the evolution basis, the charm PDFs are parametrized through the two combinations

$$T_{15}(x, Q^2) = (u^+ + d^+ + s^+ - 3c^+)(x, Q), \quad V_{15}(x, Q^2) = (u^- + d^- + s^- - 3c^-)(x, Q), \quad (\text{B1})$$

Dataset	$n_{\text{dat}}$	$\chi^2/n_{\text{dat}} (c \neq \bar{c})$	$\chi^2/n_{\text{dat}} (c = \bar{c})$
DIS NC (fixed-target)	973	1.24	1.26
DIS CC (fixed-target)	908	0.86	0.86
DIS NC (collider)	1127	1.18	1.19
DIS CC (collider)	81	1.23	1.28
Drell-Yan (fixed-target)	195	1.02	1.00
Tevatron $W, Z$ production (inclusive)	65	1.06	1.09
LHC $W, Z$ production (inclusive)	463	1.35	1.37
LHC $W, Z$ production ( $p_T$ and jets)	150	0.99	0.98
LHC top-quark pair production	64	1.28	1.21
LHC jet production	171	1.25	1.26
LHC isolated $\gamma$ production	53	0.76	0.77
LHC single $t$ production	17	0.36	0.36
<b>Total</b>	<b>4616</b>	<b>1.151</b>	<b>1.162</b>

TABLE I: The values of the experimental  $\chi^2$  per data point for the different groups of processes entering the NNPDF4.0 determination as well as for the total dataset. We compare the results of the baseline NNPDF4.0 fit ( $c = \bar{c}$ ) with the results of this work ( $c \neq \bar{c}$ ).

at a scale  $Q = Q_0 = 1.65$  GeV. In the flavor basis, they are parametrized as  $c(x, Q)$  and  $\bar{c}(x, Q)$ . Note that in neither case the total and valence combinations  $c^\pm$  are elements of the basis.

In Fig. 8 the  $xc^\pm$  PDFs found using either basis are compared. Agreement at the one sigma level is found for all  $x$ . The main qualitative features are independent of the basis choice, specifically the presence of a positive valence peak around  $x \sim 0.3$  for  $xc^-$ . Results in the flavor basis display larger PDF uncertainties, possibly because the flavor basis has not undergone the same extensive hyperoptimization as the fits in the evolution basis.

### Appendix C: Dependence on the value of the charm mass

We verify the independence of our results on the value of the charm quark mass, by repeating our determination as the (pole) charm mass is varied from the default  $m_c = 1.51$  GeV to  $m_c = 1.38$  GeV and 1.64 GeV. Note that we always choose the scale  $\mu_c = m_c$  as matching scale between the 4FNS and 3FNS, hence this is also varied alongside  $m_c$ . The total and valence charm PDFs  $xc^\pm$  at  $Q = 1.65$  GeV in the 4FNS are displayed in Fig. 9. The result is found to be essentially independent of the charm mass value, in agreement with the corresponding result of Ref. [8] (see Fig 8.6)

### Appendix D: Dataset dependence

We study the effect of removing some datasets from our determination, with the dual goal of checking the stability of our results, and investigating which data mostly determine the valence charm PDF. Specifically, we remove the LHCb  $W, Z$  data, which was found in [9] to dominate the constraints on the total charm PDF from all collider measurements, and we determine the PDFs only using DIS structure function data. The valence charm PDF found in either case is compared to the default in Fig. 10. Removing the LHCb electroweak data leaves  $xc^-$  mostly unchanged, hence the valence PDF appears to be less sensitive to this data than the total charm. When only including DIS data a nonzero



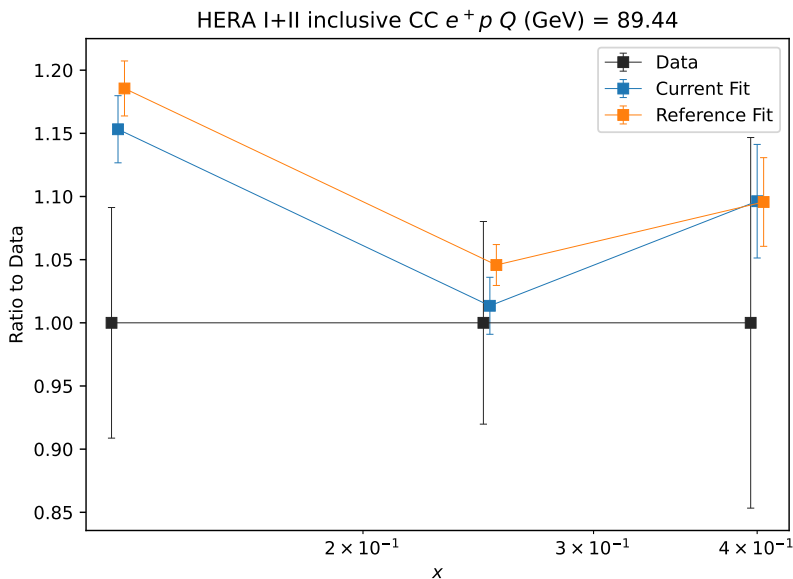


FIG. 7: Comparison between the data (black) and theory prediction obtained using the  $c \neq \bar{c}$  (blue) and  $c = \bar{c}$  (orange) PDF sets of Table I, for HERA  $e^+p$  charged-current DIS in one of the largest  $Q^2$  bins,  $Q = 89.44$  GeV, shown as a ratio to the experimental data.

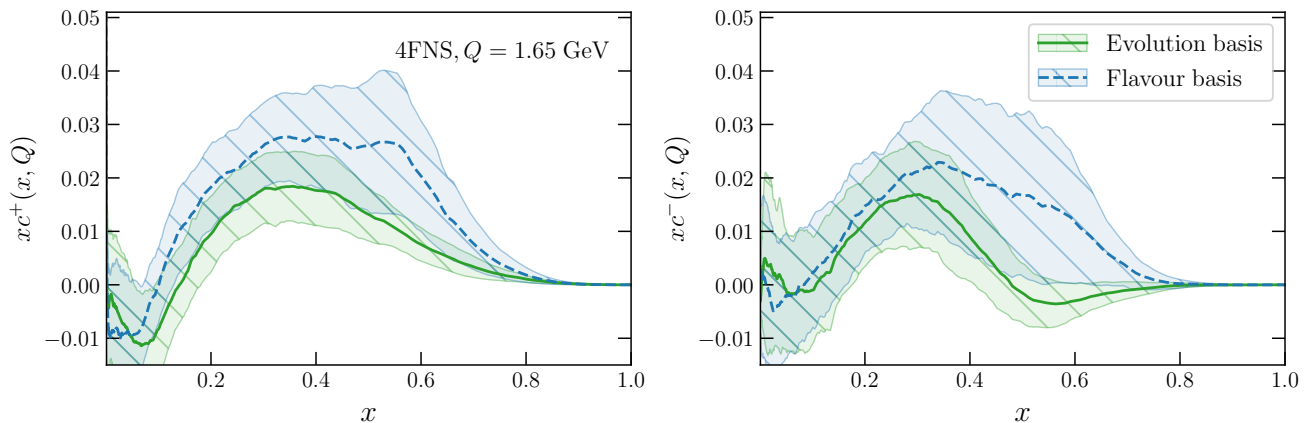


FIG. 8: Comparison between the total  $xc^+$  (left) and valence  $xc^-$  (right) charm PDFs in the 4FNS at  $Q = 1.65$  GeV, obtained parametrizing PDFs in the evolution basis (default) and in the flavor basis.

valence component is still found but now with a reduced significance: the result is consistent with zero at the one sigma level.

### Appendix E: Kinematic cuts

The NNPDF4.0 dataset only includes data with  $Q^2 \geq 3.5$  GeV<sup>2</sup> and  $W^2 \geq 12.5$  GeV<sup>2</sup>, in order to ensure the reliability of the leading-twist, fixed-order perturbative approximation. It is important to verify that results for intrinsic charm are stable upon variation of these cuts, as this checks that intrinsic charm is not contaminated by possible nonperturbative corrections not accounted for in the global PDF fitting framework. To this purpose, we have raised the  $W^2$  cut in steps of 2.5 GeV<sup>2</sup> up to 20 GeV<sup>2</sup>, and the  $Q^2$  cut up to 5 GeV<sup>2</sup>. Results are displayed in Fig. 11, and prove satisfactory stability: upon variation of the  $W^2$  cut nothing changes, and upon variation of the  $Q^2$  cut (which removes a sizable amount of data) the central value is stable and the uncertainty only marginally increased.

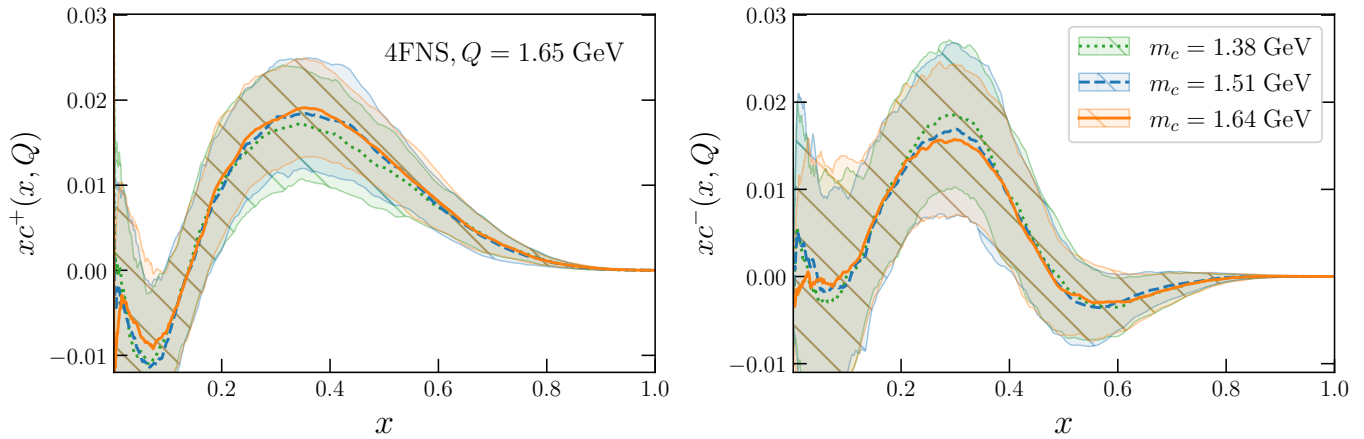


FIG. 9: Comparison of the total  $xc^+$  (left) and valence charm  $xc^-$  (right) PDFs in the 4FNS at  $Q = 1.65$  GeV as the charm pole mass is varied about the default central value  $m_c = 1.51$  GeV by  $\pm 0.13$  GeV.

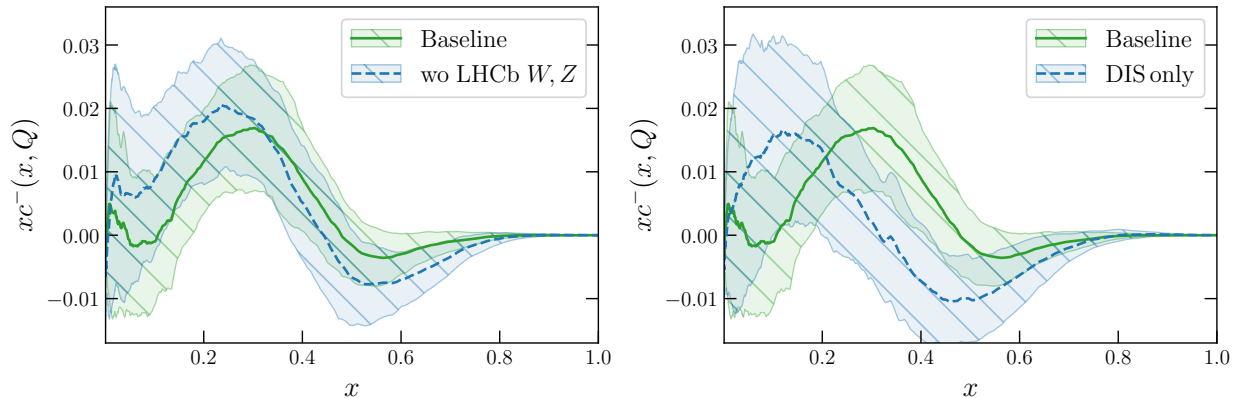


FIG. 10: The charm valence PDF in the default determination compared to a determination in which LHCb  $W, Z$  inclusive production data are excluded (left) and a determination based on DIS structure functions only (right).

### Appendix F: The $Z$ +charm asymmetry at LHCb

The projected statistical uncertainties for the future LHCb measurement of  $\mathcal{A}_c$  shown in Fig. 4 are obtained extrapolating from those of the Run 2 data by correcting both for the higher luminosity and for the acceptance associated to the different charm-tagging procedure required in this case. The uncertainties obtained in the Run 2 measurement [10] and based on an integrated luminosity of  $\mathcal{L} = 6 \text{ fb}^{-1}$  are rescaled to the expected luminosity to be accumulated by LHCb by the end of Run 4,  $\mathcal{L} \sim 50 \text{ fb}^{-1}$ , and at the HL-LHC,  $\mathcal{L} \sim 300 \text{ fb}^{-1}$ . Furthermore, the Run 2 measurement was based on charm-meson tagging with displaced vertices, with a charm-tagging efficiency of  $\epsilon_c \sim 25\%$ . The asymmetry  $\mathcal{A}_c$  requires separating charm from anticharm in the final states, which in turn demands reconstructing the  $D$ -meson decay products. The associated efficiency is estimated by weighting the  $D$ -meson branching fractions to the occurrence of each species in the LHCb  $Z$ +charm sample, resulting in an efficiency of  $\epsilon_c \sim 3\%$ . The uncertainty on the asymmetry is then determined by using error propagation with  $N_j^c = N_j^{\bar{c}}$ , neglecting the dependence of the uncertainty on the value of the asymmetry itself.

### Appendix G: Charm structure functions at the Electron-Ion Collider

The projected statistical uncertainties for the future EIC measurement of the charm-tagged asymmetry  $\mathcal{A}_{\sigma_{c\bar{c}}}$  shown in Fig. 5 are obtained as follows. We adopt the projections from [15] for the kinematic coverage in the  $(x, Q^2)$  plane and the expected statistical precision, based on running at a center-of-mass energy of  $\sqrt{s} = 63$  GeV for  $\mathcal{L} = 10 \text{ fb}^{-1}$ . These projections entail that measurements of charm production at the EIC will cover the region  $1.3 \text{ GeV}^2 \lesssim Q^2 \lesssim 120 \text{ GeV}^2$  and  $5 \times 10^{-4} \lesssim x \lesssim 0.5$ . Charm production is tagged from the reconstruction of  $D^0$  and  $\bar{D}^0$  exclusive decays,

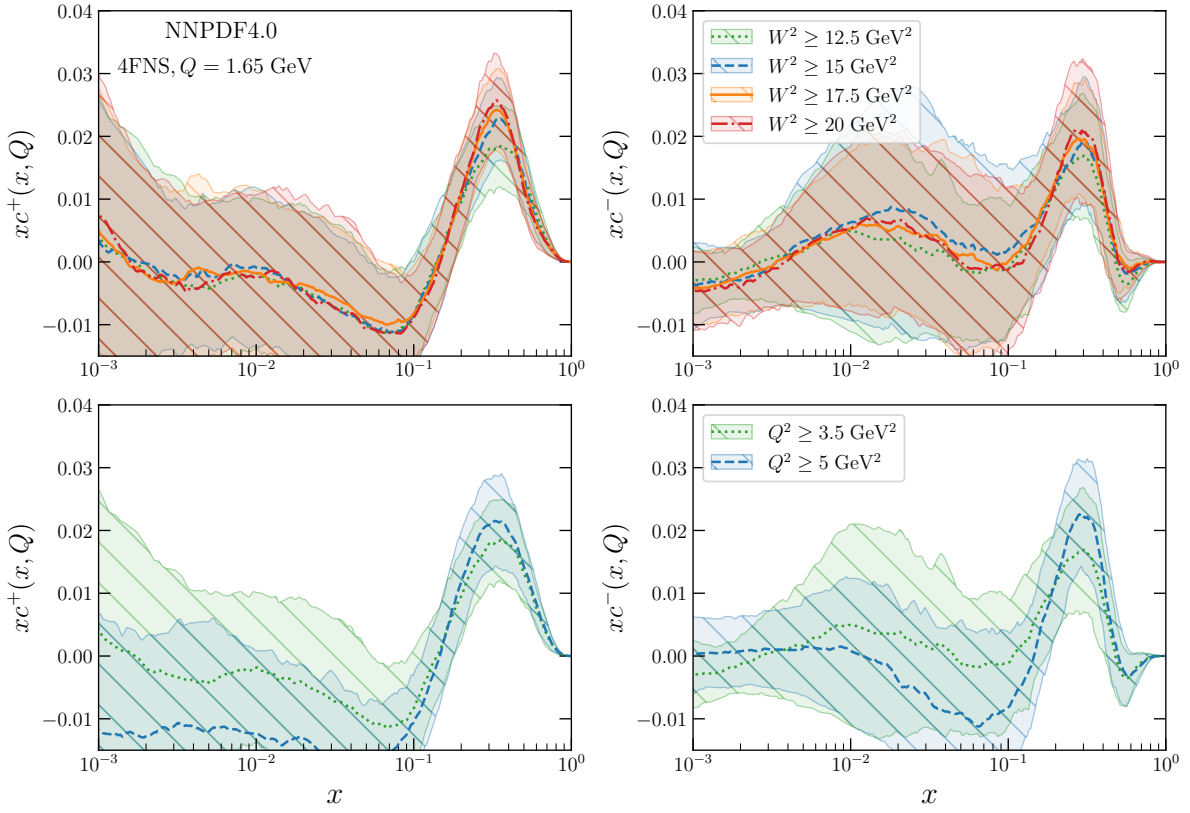


FIG. 11: The variation in the 4FNS total (left) and valence (right) charm PDFs at  $Q = 1.65$  GeV as the  $W^2$  cut is raised to  $20$  GeV $^2$  in steps of  $2.5$  GeV $^2$  (top) and the  $Q^2$  cut is raised to  $5$  GeV $^2$  (bottom). The kinematic cuts in the baseline fit are  $Q^2 \geq 3.5$  GeV $^2$  and  $W^2 \geq 12.5$  GeV $^2$ .

and a detailed estimate of experimental uncertainties would require a full detector simulation. Here, however, we limit ourselves to estimating the statistical accuracy on the asymmetry Eq. (5), which is expressed in terms of reduced cross-sections, defined as in Ref. [38] in terms of charm structure functions. For this, we take the statistical uncertainties provided in [15] and increase them by a factor  $\sqrt{2}$  since the measured sample has to be separated into  $D$ - and  $\bar{D}$ -tagged subsamples.