New parton distribution functions from a global analysis of quantum chromodynamics

Sayipjamal Dulat,[1,2,*] Tie-Jiun Hou,[3,†] Jun Gao,[4,‡] Marco Guzzi,[5,§] Joey Huston,[2,∥] Pavel Nadolsky,[3,¶] Jon Pumplin,[2,**] Carl Schmidt,[2,††] Daniel Stump,[2,‡‡] and C.-P. Yuan[2,§§]

[1]*School of Physics Science and Technology, Xinjiang University, Urumqi, Xinjiang 830046, China*
[2]*Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, USA*
[3]*Department of Physics, Southern Methodist University, Dallas, Texas 75275-0181, USA*
[4]*High Energy Physics Division, Argonne National Laboratory, Argonne, Illinois 60439, USA*
[5]*School of Physics and Astronomy, University of Manchester, Manchester M13 9PL, United Kingdom*

We present new parton distribution functions (PDFs) at next-to-next-to-leading order (NNLO) from the CTEQ-TEA global analysis of quantum chromodynamics. These differ from previous CT PDFs in several respects, including the use of data from LHC experiments and the new D0 charged-lepton rapidity asymmetry data, as well as the use of a more flexible parametrization of PDFs that, in particular, allows a better fit to different combinations of quark flavors. Predictions for important LHC processes, especially Higgs boson production at 13 TeV, are presented. These CT14 PDFs include a central set and error sets in the Hessian representation. For completeness, we also present the CT14 PDFs determined at the LO and the NLO in QCD. Besides these general-purpose PDF sets, we provide a series of (N)NLO sets with various α_s values and additional sets in general-mass variable flavor number schemes, to deal with heavy partons, with up to three, four, and six active flavors.

I. INTRODUCTION

Run 1 at the Large Hadron Collider (LHC) was a great success, culminating in the discovery of the Higgs boson [1,2]. No physics beyond the Standard Model was discovered in this run; however, run 2, with a larger center-of-mass energy and integrated luminosity, will allow for an increased discovery potential for new physics. Precision measurements of the Higgs boson and of various electroweak observables will be performed with extraordinary accuracy in new kinematic regimes in run 2. Run-1 achievements, such as the combined ATLAS-CMS measurement of the Higgs boson mass with 0.2% accuracy [3], will soon be superseded. For both precision measurements and for discovery of possible new physics, it is important to have the proper tools for the calculation of the relevant cross sections. These tools include both matrix element determinations at higher orders in perturbative QCD and electroweak theory and precision parton distribution

functions (PDFs). The need for precision PDFs was driven home by the recent calculation of the inclusive cross section for gluon-gluon fusion to a Higgs boson at next-to-next-to-next-to-leading order (NNNLO) [4]. As this tour-de-force calculation has significantly reduced the scale dependence of the Higgs cross section, the PDF and α_s uncertainties become the dominant remaining theoretical uncertainty (as of the last PDF4LHC recommendation).

The CT10 parton distribution functions were published at NLO in 2010 [5], followed by the CT10 NNLO parton distribution functions in 2013 [6]. These PDF ensembles were determined using diverse experimental data from fixed-target experiments, HERA and the Tevatron collider, but without data from the LHC. In this paper, we present a next generation of PDFs, designated as CT14. The CT14 PDFs include data from the LHC for the first time, as well as updated data from the Tevatron and from HERA experiments. Various CT14 PDF sets have been produced at the LO, NLO and NNLO and are available from LHAPDF [7].

The CTEQ-TEA philosophy has always been to determine PDFs from data on inclusive, high-momentum transfer processes, for which perturbative QCD is expected to be reliable. For example, in the case of deep-inelastic lepton scattering, we only use data with $Q > 2$ GeV and $W > 3.5$ GeV. Data in this region are expected to be relatively free of nonperturbative effects, such as higher twists or nuclear corrections. Thus, there is no need to introduce phenomenological models for nonperturbative corrections beyond the leading-twist perturbative contributions.

*sdulat@msu.edu
†tiejiunh@mail.smu.edu
‡jgao@anl.gov
§marco.guzzi@manchester.ac.uk
∥huston@pa.msu.edu
¶nadolsky@physics.smu.edu
**pumplin@pa.msu.edu
††schmidt@pa.msu.edu
‡‡stump@pa.msu.edu
§§yuan@pa.msu.edu

033006-1

For the majority of processes in the CT14 global analysis, theoretical predictions are now included at the NNLO level of accuracy. In particular, a NNLO treatment [8] of heavy-quark mass effects in neutral-current deep-inelastic scattering (DIS) is realized in the Aivazis-Collins-Olness-Tung scheme with kinematic rescaling (ACOT-χ) [9–11] and is essential for obtaining correct predictions for LHC electroweak cross sections [12,13]. We make two exceptions to this rule, by including measurements for charged-current DIS and inclusive jet production at NLO only. In both cases, the complete NNLO contributions are not yet available, but it can be argued based on our studies that the expected effect of missing NNLO effects is small relative to current experimental errors (cf. Sec. II). For both types of processes, the NLO predictions have undergone various benchmarking tests. A numerical error was discovered and corrected in the implementation of the Simplified ACOT-χ scheme (SACOT-χ) for charged-current DIS, resulting in relatively small changes from CT10 (within the PDF uncertainties).

As in the CT10 global analysis, we use a charm pole mass of 1.3 GeV, which was shown to be consistent with the CT10 data in Ref. [6]. The PDFs for u, d, s (anti)quarks and the gluon are parametrized at an initial scale of 1.295 GeV, and the charm quark PDF is turned on with zero intrinsic charm as the scale Q reaches the charm pole mass.

The new LHC measurements of W/Z cross sections directly probe the flavor separation of u and d (anti)quarks in an x range around 0.01 that was not directly assessed by the previously available experiments. We also include an updated measurement of electron charge asymmetry from the D0 Collaboration [14], which probes the d quark PDF at $x > 0.1$. To better estimate variations in relevant PDF combinations, such as $d(x, Q)/u(x, Q)$ and $\bar{d}(x, Q)/\bar{u}(x, Q)$, we increased the number of free PDF parameters to 28, compared to 25 in CT10 NNLO. As another important modification, CT14 employs a novel flexible parametrization for the PDFs, based on the use of Bernstein polynomials (reviewed in the Appendix). The shape of the Bernstein polynomials is such that a single polynomial is dominant in each given x range, reducing undesirable correlations among the PDF parameters that sometimes occurred in CT10. In the asymptotic limits of $x \rightarrow 0$ or $x \rightarrow 1$, the new parametrization forms allow for the possibility of arbitrary constant ratios of d/u or $\bar{d}/\bar{u}$, in contrast to the more constrained behavior assumed in CT10.

The PDF error sets of the CT14 ensemble are obtained using two techniques, the Hessian method [15] and Monte Carlo sampling [16]. Lagrange multiplier studies [17] have also been used to verify the Hessian uncertainties, especially in regions not well constrained by data. This applies at NNLO and NLO; no error sets are provided at LO due to the difficulty in defining meaningful uncertainties at that order.

A central value of $\alpha_s(M_Z)$ of 0.118 has been assumed in the global fits at NLO and NNLO, but PDF sets at alternative

values of $\alpha_s(m_Z)$ are also provided. CT14 prefers $\alpha_s(M_Z) = 0.115^{+0.006}_{-0.004}$ at NNLO (0.117 ± 0.005 at NLO) at 90% confidence level (C.L.). These uncertainties from the global QCD fits are larger than those of the data from LEP and other experiments included into the world average [18]. Thus, the central PDF sets are obtained using the value of 0.118, which is consistent with the world average value and was recommended by the PDF4LHC group [19]. For the CT14 LO PDFs, we follow the precedent begun in CTEQ6 [20] by supplying two versions, one with a one-loop $\alpha_s(M_Z)$ value of 0.130 and the other with a two-loop $\alpha_s(M_Z)$ value of 0.118.

The flavor composition of CT14 PDFs has changed somewhat compared to CT10 due to the inclusion of new LHC and Tevatron data sets, to the use of modified parametrization forms, and to the numerical modifications discussed above. The new PDFs are largely compatible with CT10 within the estimated PDF uncertainty. The CT14 NNLO PDFs have a softer strange quark distribution at low x and a somewhat softer gluon at high x, compared to CT10 NNLO. The d/u ratio has decreased at high x in comparison to CT10, as a consequence of replacing the 2008 D0 electron charge asymmetry (0.75 fb^{-1} [21]) measurement by the new 9.7 fb^{-1} data set [14]. The d/u ratio approaches a constant value in the $x \rightarrow 1$ limit due to the input physics assumption that both d_{val} and u_{val} behave as $(1 - x)^{a_2}$ at $x \rightarrow 1$ with the same value of a_2 (reflecting expectations from spectator counting rules), but allowing for independent normalizations. The $\bar{d}/\bar{u}$ ratio has also changed as a consequence of the new data and the new parametrization form.

The organization of the paper is as follows. In Sec. II, we list the data sets used in the CT14 fit and discuss further aspects of the global fits for the central CT14 PDFs and for the error sets. In Sec. III, we show various aspects of the resultant CT14 PDFs and make comparisons to CT10 PDFs. In Sec. IV, we show comparisons of NNLO predictions using the CT14 PDFs to some of the data sets used in the global fits. Specifically, we compare to experimental measurements of jet, W and Z, $W + c$ cross sections. In Sec. V, we discuss NNLO predictions using the CT14 PDFs for Higgs boson production via the gluon-gluon fusion channel and for top-quark and antiquark pair production. Our conclusion is given in Sec. VI.

II. SETUP OF THE ANALYSIS

A. Overview of the global fit

The goal of the CT14 global analysis is to provide a new generation of PDFs intended for widespread use in high-energy experiments. As we generate new PDF sets, we include newly available experimental data sets and theoretical calculations and redesign the functional forms of PDFs if new data or new theoretical calculations favor it. All changes—data, theory, and parametrization—contribute to the differences between the old and new generations of PDFs in ways that are correlated and frequently

cannot be separated. The most important, but not the only, criterion for the selection of PDFs is the minimization of the log-likelihood χ^2 that quantifies agreement of theory and data. In addition, we make some "prior assumptions" about the forms of the PDFs. A PDF set that violates them may be rejected even if it lowers χ^2. For example, we assume that the PDFs are smoothly varying functions of x, without abrupt variations or short-wavelength oscillations. This is consistent with the experimental data and sufficient for making new predictions. No PDF can be negative at the input scale Q_0, to preclude negative cross sections in the predictions. Flavor-dependent ratios or cross section asymmetries must also take physical values, which limits the range of allowed parametrizations in extreme kinematical regions with poor experimental constraints. For example, in the CT14 parametrization we restricted the functional forms of the u and d PDFs so that $d(x, Q_0)/u(x, Q_0)$ would remain finite and nonzero at $x \rightarrow 1$; cf. the Appendix. We now review every input of the CT14 PDF analysis in turn, starting with the selection of the new experiments.

B. Selection of experiments

The experimental data sets that are included in the CT14 global analysis are listed in Tables I (lepton scattering) and II (production of inclusive lepton pairs and jets). There are a total of 2947 data points included from 33 experiments, producing a χ^2 value of 3252 for the best fit (with $\chi^2/N_{pt} = 1.10$). It can be seen from the values of χ^2 in Tables I and II that the data and theory are in reasonable agreement for most experiments. The variable S_n in the last column is an "effective Gaussian variable," first introduced in Sec. V of Ref. [5] and defined for the current analysis in Refs. [6,22]. The effective Gaussian variable quantifies compatibility of any given data set with a particular PDF fit

in a way that is independent of the number of points $N_{pt,n}$ in the data set. It maps the χ^2_n values of individual experiments, whose probability distributions depend on $N_{pt,n}$ in each experiment (and thus, are not identical), onto S_n values that obey a cumulative probability distribution shared by all experiments, independently of $N_{pt,n}$. Values of S_n between -1 and $+1$ correspond to a good fit to the nth experiment (at the 68% C.L.). Large positive values ($\gtrsim 2$) correspond to a poor fit, while large negative values ($\lesssim -2$) are fit unusually well.

The goodness of fit for CT14 NNLO is comparable to that of our earlier PDFs, but the more flexible parametrizations did result in improved agreement with some data sets. For example, by adding additional parameters to the $\{u, \overline{u}\}$ and $\{d, \overline{d}\}$ parton distributions, somewhat better agreement was obtained for the BCDMS and NMC data at low values of Q. The quality of the fit can be also evaluated based on the distribution of S_n values, which follows a standard normal distribution (of width 1) in an ideal fit. As in the previous fits, the actual S_n distribution (cf. the solid curve in Fig. 1) is somewhat wider than the standard normal one (the dashed curve), indicating the presence of disagreements, or tensions, between some of the included experiments. The tensions have been examined before [5,51–53] and originate largely from experimental issues, almost independent of the perturbative QCD order or PDF parametrization form. A more detailed discussion of the level of agreement between data and theory will be provided in Sec. IV.

1. Experimental data from the LHC

Much of these data have also been used in previous CT analyses, such as the one that produced the CT10 NNLO PDFs. As mentioned, no LHC data were used in the CT10

TABLE I. Experimental data sets employed in the CT14 analysis. These are the lepton deep-inelastic scattering experiments. $N_{pt,n}, \chi^2_n$ are the number of points and the value of χ^2 for the nth experiment at the global minimum. S_n is the effective Gaussian parameter [5,6,22] quantifying agreement with each experiment.

ID No.	Experimental data set		$N_{pt,n}$	χ^2_n	$\chi^2_n/N_{pt,n}$	S_n
101	BCDMS F_2^p	[23]	337	384	1.14	1.74
102	BCDMS F_2^d	[24]	250	294	1.18	1.89
104	NMC F_2^d/F_2^p	[25]	123	133	1.08	0.68
106	NMC σ_{red}^p	[25]	201	372	1.85	6.89
108	CDHSW F_2^p	[26]	85	72	0.85	−0.99
109	CDHSW F_3^p	[26]	96	80	0.83	−1.18
110	CCFR F_2^p	[27]	69	70	1.02	0.15
111	CCFR xF_3^p	[28]	86	31	0.36	−5.73
124	NuTeV $\nu\mu\mu$ semi-inclusive DIS	[29]	38	24	0.62	−1.83
125	NuTeV $\bar{\nu}\mu\mu$ semi-inclusive DIS	[29]	33	39	1.18	0.78
126	CCFR $\nu\mu\mu$ semi-inclusive DIS	[30]	40	29	0.72	−1.32
127	CCFR $\bar{\nu}\mu\mu$ semi-inclusive DIS	[30]	38	20	0.53	−2.46
145	H1 σ_r^b	[31]	10	6.8	0.68	−0.67
147	Combined HERA charm production	[32]	47	59	1.26	1.22
159	HERA1 combined DIS	[33]	579	591	1.02	0.37
169	H1 F_L	[34]	9	17	1.92	1.7

TABLE II.   The same as Table I, showing experimental data sets on Drell-Yan processes and inclusive jet production.

ID No.	Experimental data set		$N_{pt,n}$	χ^2_n	$\chi^2_n/N_{pt,n}$	S_n
201	E605 Drell-Yan process	[35]	119	116	0.98	−0.15
203	E866 Drell-Yan process, $\sigma_{pd}/(2\sigma_{pp})$	[36]	15	13	0.87	−0.25
204	E866 Drell-Yan process, $Q^3 d^2\sigma_{pp}/(dQdx_F)$	[37]	184	252	1.37	3.19
225	CDF run-1 electron A_{ch}, $p_{T\ell} > 25$ GeV	[38]	11	8.9	0.81	−0.32
227	CDF run-2 electron A_{ch}, $p_{T\ell} > 25$ GeV	[39]	11	14	1.24	0.67
234	D0 run-2 muon A_{ch}, $p_{T\ell} > 20$ GeV	[40]	9	8.3	0.92	−0.02
240	LHCb 7 TeV 35 pb^{-1} W/Z $d\sigma/dy_\ell$	[41]	14	9.9	0.71	−0.73
241	LHCb 7 TeV 35 pb^{-1} A_{ch}, $p_{T\ell} > 20$ GeV	[41]	5	5.3	1.06	0.30
260	D0 run-2 Z rapidity	[42]	28	17	0.59	−1.71
261	CDF run-2 Z rapidity	[43]	29	48	1.64	2.13
266	CMS 7 TeV 4.7 fb^{-1}, muon A_{ch}, $p_{T\ell} > 35$ GeV	[44]	11	12.1	1.10	0.37
267	CMS 7 TeV 840 pb^{-1}, electron A_{ch}, $p_{T\ell} > 35$ GeV	[45]	11	10.1	0.92	−0.06
268	ATLAS 7 TeV 35 pb^{-1} W/Z cross sec., A_{ch}	[46]	41	51	1.25	1.11
281	D0 run-2 9.7 fb^{-1} electron A_{ch}, $p_{T\ell} > 25$ GeV	[14]	13	35	2.67	3.11
504	CDF run-2 inclusive jet production	[47]	72	105	1.45	2.45
514	D0 run-2 inclusive jet production	[48]	110	120	1.09	0.67
535	ATLAS 7 TeV 35 pb^{-1} incl. jet production	[49]	90	50	0.55	−3.59
538	CMS 7 TeV 5 fb^{-1} incl. jet production	[50]	133	177	1.33	2.51

fits. Nonetheless, the CT10 PDFs have been in good agreement with LHC measurements so far.

As the quantity of the LHC data has increased, the time has come to include the most germane LHC measurements into CT fits. The LHC has measured a variety of Standard Model cross sections, yet not all of them are suitable for determination of PDFs according to the CT method. For that, we need to select measurements that are experimentally and theoretically clean and are compatible with the global set of non-LHC hadronic experiments.

In the CT14 study, we select a few such LHC data sets at $\sqrt{s} = 7$ TeV, focusing on the measurements that provide novel information to complement the non-LHC data. From vector boson production processes, we selected W/Z cross sections and the charged-lepton asymmetry measurement from ATLAS [46], the charged-lepton asymmetry in the electron [45] and muon decay channels [44] from CMS,

FIG. 1.   Best-fit S_n values of 33 experiments in the CT14 analysis.

and the W/Z lepton rapidity distributions and charged-lepton asymmetry from LHCb [41]. The ATLAS and CMS measurements primarily impose constraints on the light quark and antiquark PDFs at $x \gtrsim 0.01$. The LHCb data sets, while statistically limited, impose minor constraints on $\overline{u}$ and d PDFs at $x = 0.05$–0.1.

Upon including these measurements, we can relax the parametric constraints on the sea (anti)quark PDFs of u, $\overline{u}$, d, and $\overline{d}$. In the absence of relevant experimental constraints in the pre-CT14 fits, the PDF parametrizations were chosen so as to enforce $\overline{u}/\overline{d} \to 1$, $u/d \to 1$ at $x \to 0$ in order to obtain convergent fits. As reviewed in the Appendix, the CT14 parametrization form is more flexible, in the sense that only the asymptotic power x^{a_1} is required to be the same in all light-quark PDFs in the $x \to 0$ limit. This choice produces wider uncertainty bands on u_v, d_v, and $\overline{u}/\overline{d}$ at $x \to 0$, with the spread constrained by the newly included LHC data.

From the other LHC measurements, we now include single-inclusive jet production at ATLAS [49] and CMS [50]. These data sets provide complementary information to Tevatron inclusive jet production cross sections from CDF run 2 [47] and D0 run 2 [48] that are also included. The purpose of jet production cross sections is primarily to constrain the gluon PDF $g(x, Q)$. While the uncertainties from the LHC jet cross sections are still quite large, they probe the gluon PDF across a much wider range of x than the Tevatron jet cross sections.

One way to gauge the sensitivity of a specific data point to some PDF $f(x, Q)$ at a given x and Q is to compute a correlation cosine between the theoretical prediction for this point and $f(x, Q)$ [13,15,54]. In the case of CT10 NNLO, the sensitivity of the LHC charge asymmetry data

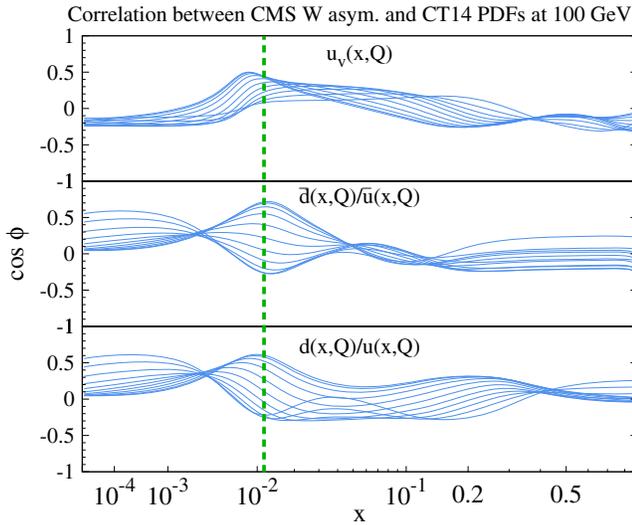Correlation between CMS W asym. and CT14 PDFs at 100 GeV
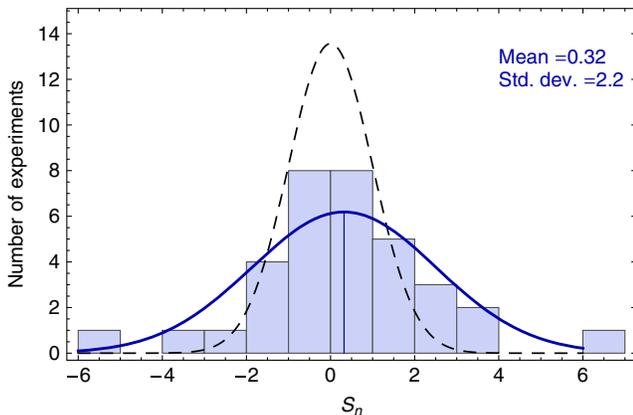
FIG. 2. The correlation cosine $\cos\phi$ [13] between the PDF $f(x, Q = 100~\text{GeV})$ at the specified x value on the horizontal axis and NNLO predictions for muon CMS charge asymmetry [44].

sets to the valence PDF combinations at $x = 0.01$–0.1 was established by this method in Sec. VII C of [6]. However, the somewhat large strength of correlations at small x that had been observed suggested the possibility that CT10 light-quark parametrizations were not sufficiently flexible in the x region probed by the LHC charge asymmetry.

Since CT14 has adopted more flexible parametrizations for the affected quark flavors, the above correlations with u_v, d_v, and d/u at small x are now somewhat relaxed, as illustrated by the newly computed correlations between CT14 NNLO and CMS A_{ch} data in Fig. 2. Each line shows $\cos\phi$ between $f(x, Q)$ and the NNLO prediction for one of the bins of the data. When the PDF uncertainty receives a large contribution from $f(x, Q)$, $\cos\phi$ comes out to be close to ± 1, say, $|\cos\phi| > 0.7$. With the new parametrization form, the CMS charge asymmetry is reasonably, but not exceptionally, correlated with both $\overline{d}/\overline{u}$ and d/u at $x \sim 0.01$ corresponding to central-rapidity production of weak bosons at $\sqrt{s} = 7$ TeV (indicated by a vertical dashed line in the figure). The correlation with u_v and d_v is smaller than in CT10.
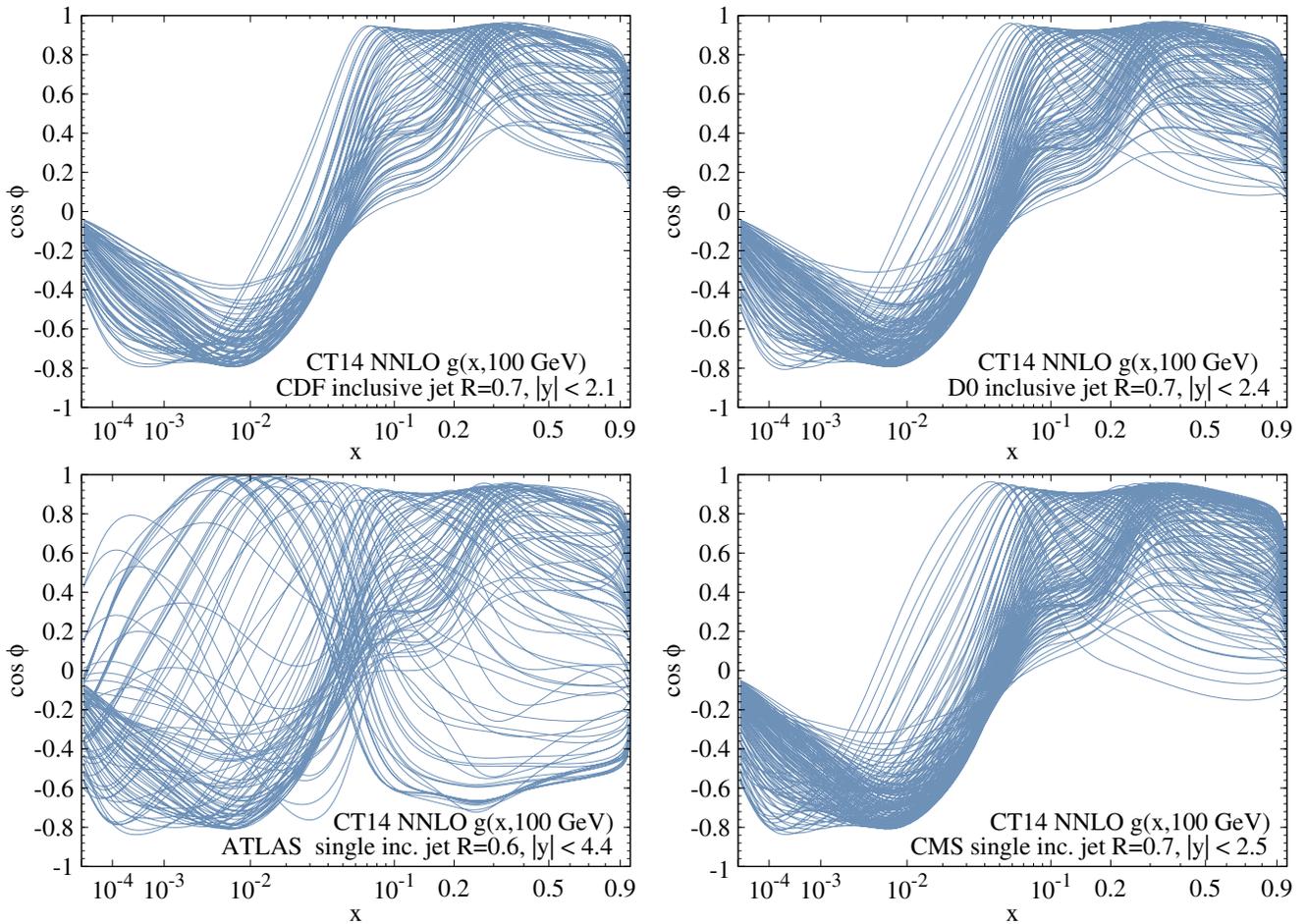

FIG. 3. The correlation cosine $\cos\phi$ [13] between the g-PDF at the specified x value on the horizontal axis and NLO predictions for the CDF [47] (upper left panel), D0 [48] (upper right panel), ATLAS [49] (lower left panel) and CMS [50] (lower right panel) inclusive jet cross sections at $Q = 100$ GeV.

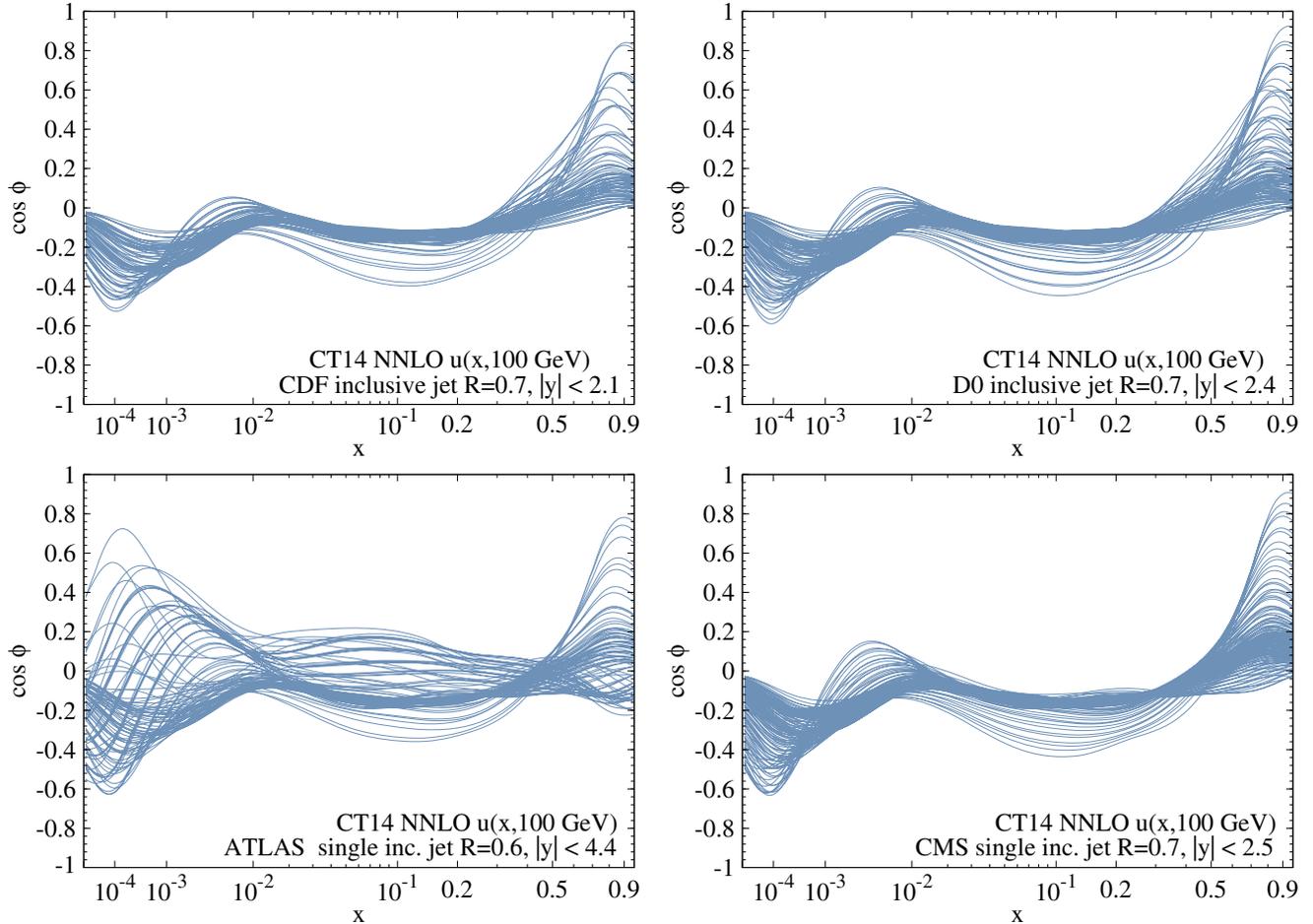FIG. 4. The correlation cosine $\cos \phi$ [13] between the u-PDF at the specified x value on the horizontal axis and NLO predictions for the CDF [47] (upper left panel), D0 [48] (upper right panel), ATLAS [49] (lower left panel) and CMS [50] (lower right panel) inclusive jet cross sections at $Q = 100$ GeV.

For the ATLAS [49], CMS [50], CDF [47] and D0 [48] inclusive jet data sets, the correlation cosine, $\cos \phi$, for gluon PDF is plotted in Fig. 3 using NLO QCD theory to evaluate the theoretical cross section. Again, the lines correspond to individual p_{Tj} bins of the data. We observe that the CDF and D0 jet cross sections are highly correlated with the gluon PDF $g(x, Q)$ at $x \gtrsim 0.05$ and anticorrelated at small x as a consequence of the momentum sum rule. The ATLAS and CMS jet cross sections are highly correlated with $g(x, Q)$ in a much wider range, $x > 0.005$. In contrast, the PDF-induced correlation of the jet cross sections with the quark PDFs, such as $u(x, Q)$ in Fig. 4, is at most moderate. The ATLAS and CMS jet data therefore have the potential to reduce the gluon uncertainty, but significant reduction will require the data from run 2.

2. High-luminosity lepton charge asymmetry from the Tevatron

Forward-backward asymmetry (A_{ch}) distributions of charged leptons from inclusive weak boson production at the Tevatron are uniquely sensitive to the average slope of

the ratio $d(x, Q)/u(x, Q)$ at large x, of order 0.1 and above. In the CT14 analysis, we include several data sets of A_{ch} measured at $\sqrt{s} = 1.8$ and 1.96 TeV by the CDF and D0 Collaborations. The CDF run-1 data set on A_{ch} [38,55], which was instrumental in resolving conflicting information on the large-x behavior of $u(x, Q)$ and $d(x, Q)$ from contemporary fixed-target DIS experiments [56–59], is supplemented by the CDF run-2 data set at 170 pb^{-1} [39]. A_{ch} data at $\sqrt{s} = 1.96$ TeV from D0 in the electron [14] and muon [40] decay channels, for 9.7 and 0.3 fb^{-1}, are also included. In all A_{ch} data sets, we include subsamples with the cuts on the transverse momentum $p_{T\ell}$ of the final-state lepton specified in Table II.

The electron data set (9.7 fb^{-1}) from D0 that we now include replaces the 0.75 fb^{-1} counterpart set [21], first included in CT10. This replacement has an important impact on the determination of the large-x quark PDFs; thus, these new A_{ch} data sets are perhaps the most challenging and valuable among all that were added in CT14.

The D0 A_{ch} data have small experimental errors and hence push the limits of the available theoretical

calculations. Relatively small differences in the average slope (with respect to x) of the d/u ratio in the probed region can produce large variations in χ^2_n for the Tevatron charge asymmetry [56–58]. By varying the minimal selection cuts on $p_{T\ell}$ of the lepton, it is possible to probe subtle features of the large-x PDFs. For that, understanding of the transverse momentum dependence in both experiment and theory is necessary, which demands evaluation of transverse momentum resummation effects.

When the first Tevatron run-2 A_{ch} data sets were implemented in CT fits, significant tensions were discovered between the electron and muon channels, and even between different $p_{T\ell}$ bins within one decay channel. The tensions prompted a detailed study in the CT10 analysis [5]. The study found that various $p_{T\ell}$ bins of the electron and muon asymmetries from D0 disagree with DIS experiments and among themselves.

In light of these unresolved tensions, we published a CT10 PDF ensemble at NLO, which did not include the D0 run-2 A_{ch} data and yielded a d/u ratio that was close to that ratio in CTEQ6.6 NLO. An alternative CT10W NLO ensemble was also constructed. It included four $p_{T\ell}$ bins of that data and predicted a harder d/u behavior at $x \to 1$. When constructing the counterpart CT10 NNLO PDFs in [6], we took an in-between path and included only the two most inclusive $p_{T\ell}$ bins, one from the electron [21] and one from the muon [40] samples. This choice still resulted in a larger d/u asymptotic value in CT10 NNLO than in CTEQ6.6.

The new A_{ch} data for 9.7 fb^{-1} in the electron channel are more compatible with the other global fit in the data that we included. Therefore, CT14 includes the D0 A_{ch} measurement in the muon channel with $p_{T\ell} > 20$ GeV [40] and in the electron channel with $p_{T\ell} > 25$ GeV [14]. The replacement does not affect the general behavior of the PDFs, except that the CT14 d/u ratio at high x follows the trends of CTEQ6.6 NLO and CT10 NLO, rather than of CT10W NLO and CT10 NNLO.

3. New HERA data

CT14 includes a combined HERA-1 data set of reduced cross sections for semi-inclusive DIS production of open charm [32] and measurements of the longitudinal structure function $F_L(x, Q)$ in neutral-current DIS [34]. The former replaces independent data sets of charm structure functions and reduced cross sections from H1 and ZEUS [60–63]. Using the combined HERA charm data set, we obtain a slightly smaller uncertainty on the gluon at $x < 0.01$ and better constraints on charm mass than with independent sets [64]. The latter HERA data set, on F_L, is not independent from the combined HERA set on inclusive DIS [33] but has only nine data points and does not significantly change the global χ^2. Its utility is primarily to prevent unphysical solutions for the gluon PDF at small x at the stage of the PDF error analysis.

4. Other LHC results

One class of LHC data that could potentially play a large role [13] in the determination of the gluon distribution, especially at high x, is the differential distributions of $t\bar{t}$ production, now available from ATLAS [65] and CMS [66,67]. However, these data are not included into our fit, as the differential NNLO $t\bar{t}$ cross section predictions for the LHC are not yet complete and the total cross section measurements lack statistical power. [68]. In addition, constraints on the PDFs from $t\bar{t}$ cross sections are mutually correlated with the values of QCD coupling and top-quark mass. NLO electroweak corrections, playing an important role [69,70] for these data, are still unavailable for some $t\bar{t}$ kinematic distributions. Once these calculations are completed, they will be incorporated in future versions of CT PDFs. For now, we simply show predictions from CT14 for the $t\bar{t}$ distributions using the approximate NNLO calculations in Sec. V.

C. Summary of theoretical calculations

1. QCD cross sections

The CT14 global analysis prioritizes the selection of published data for which NNLO predictions are available, and theoretical uncertainties of various kinds are well understood. Theoretical calculations for neutral-current DIS are based on the NNLO implementation [8] of the SACOT-χ factorization scheme [9–11] with massive quarks. For inclusive distributions in the low-mass Drell-Yan process, NNLO predictions are obtained with the program VRAP [71,72]. Predictions for W/Z production and weak boson charge asymmetries with $p_{T\ell}$ cuts are obtained with the NNLL-(approximately NNLO) program ResBos [73–76], as in the previous analyses.

As already mentioned in the introduction, two exceptions from this general rule concern charged-current DIS and collider jet production. Both have unique sensitivities to crucial PDF combinations but are still known only to NLO. The CCFR and NuTeV data on inclusive and semi-inclusive charge-current DIS are indispensable for constraining the strangeness PDF; single-inclusive jet production at the Tevatron and now at the LHC are essential for constraining the gluon distribution. Yet, in both categories, the experimental uncertainties are fairly large and arguably diminish the impact of missing NNLO effects. Given the importance of these measurements, our approach is then to include these data in our NNLO global PDF fits but evaluate their matrix elements at NLO.

According to this choice, we do not rely on the use of threshold resummation techniques [77,78] to approximate the NNLO corrections in jet production. Nor do we remove the LHC jet data due to the kinematic limitations of such resummation techniques [79]. A large effort was invested in the CT10 and CT14 analyses to estimate the possibility of biases in the NNLO PDFs due to using NLO cross sections

for jet production [80,81]. The sensitivity of the central PDFs and their uncertainty to plausible NNLO corrections was estimated with a variation of the Cacciari-Houdeau method [82], by introducing additional correlated systematic errors in jet production associated with the residual dependence on QCD scales and a potential missing contribution of a typical magnitude expected from a NNLO correction. These exercises produced two conclusions. First, the scale variation in the NLO jet cross section is reduced if the central renormalization and factorization scales are set equal to the transverse momentum p_T of the individual jet in the data bin. This choice is adopted both for the LHC and Tevatron jet cross sections. In the recently completed partial NNLO calculation for jets produced via gg scattering [83,84], this scale choice leads to an NNLO/NLO K factor that both is smaller than for the alternative scale equal to the leading jet's p_T and is relatively constant over the range of the LHC jet measurements [85]. Second, the plausible effect of the residual QCD scale dependence at NLO can be estimated as a correlated uncertainty in the CT10 NNLO fit. Currently it has marginal effect on the central PDF fits and the PDF uncertainty.

The CT14 analysis computes NLO cross sections for inclusive jet production with the help of FASTNLO [86] and APPLGRID [87] interfaces to NLOJET++ [88,89]. A series of benchmarking exercises that we had completed [80,81] verified that the fast interfaces are in good agreement among themselves and with an independent NLO calculation in the program MEKS [80]. Both ATLAS and CMS have measured the inclusive jet cross sections for two jet sizes. We use the larger of the two sizes (0.6 for ATLAS and 0.7 for CMS) to further reduce the importance of NNLO corrections.

2. Figure-of-merit function

In accord with the general procedure summarized in Ref. [6], the most probable solutions for CT14 PDFs are found by a minimization of the function

$$\chi^2_{\text{global}} = \sum_{n=1}^{N_{\text{exp}}} \chi^2_n + \chi^2_{th}. \tag{1}$$

This function sums contributions χ^2_n from N_{exp} fitted experiments and includes a contribution χ^2_{th} specifying theoretical conditions ("Lagrange multiplier constraints") imposed on some PDF parameters. In turn, the χ^2_n are constructed as in Eq. (14) of Ref. [6] and account for both uncorrelated and correlated experimental errors. Section III of that paper includes a detailed review of the statistical procedure that we continue to follow. Instead of repeating that review, we shall briefly remind the reader about the usage of the tolerance and quasi-Gaussian S variables when constructing the error PDFs.

The minimum of the χ^2_{global} function is found iteratively by the method of steepest descent using the program MINUIT. The boundaries of the 90% C.L. region around the minimum of χ^2_{global}, and the eigenvector PDF sets quantifying the associated uncertainty, are found by iterative diagonalization of the Hessian matrix [15,17]. The 90% C.L. boundary in CT14 and CT10 analyses is determined according to two tiers of criteria, based on the increase in the global χ^2_{global} summed over all experiments, and on the agreement with individual experimental data sets [5,6,22]. The first type of condition demands that the global χ^2 does not increase above the best-fit value by more than $\Delta\chi^2 = T^2$, where the 90% C.L. region corresponds to $T \approx 10$. The second condition introduces a penalty term P, called tier-2 penalty, in χ^2 when establishing the confidence region, which quickly grows when the fit ceases to agree with any specific experiment within the 90% C.L. for that experiment. The effective function $\chi^2_{\text{eff}} = \chi^2_{\text{global}} + P$ is scanned along each eigenvector direction until χ^2_{eff} increases above the tolerance bound, or rapid χ^2_{eff} growth due to the penalty P is triggered.

The penalty term is constructed as

$$P = \sum_{n=1}^{N_{\text{exp}}} (S_n)^k \theta(S_n) \tag{2}$$

from the equivalent Gaussian variables S_n that obey an approximate standard normal distribution independently of the number of data points $N_{pt,n}$ in the experiment. Every S_n is a monotonically increasing function of the respective χ^2_n given in [22,90]. The power $k = 16$ is chosen so that $(S_n)^k$ sharply increases from zero when S_n approaches 1.3, the value corresponding to the 90% C.L. cutoff. The implementation of S_n is fully documented in the Appendix of Ref. [22].

3. Correlated systematic errors

In many of the data sets included in the CT14 analysis, the reported correlated systematic errors from experimental sources dominate over the statistical errors. Care must therefore be taken in the treatment of these systematic errors to avoid artificial biases in the best-fit outcomes, such as the bias described by D'Agostini in [91,92].

Our procedure for handling the systematic errors is reviewed in Secs. III C and VI D of [6]; see also a related discussion in the Appendixes of [20,81]. The correlated errors for a given experiment, and effective shifts in the theory or data that they cause, are estimated in a linearized approximation by including a contribution in the figure-of-merit function χ^2 proportional to the correlation matrix. A practical implementation of this approach runs into a dilemma of distinguishing between the additive and multiplicative correlated errors, which are often not separated in the experimental publications, but must follow different prescriptions to prevent the bias. It is the matrix $\beta_{i,\alpha}$ of *relative* correlated errors that is typically published; the *absolute* correlated errors must be reconstructed from $\beta_{i,\alpha}$
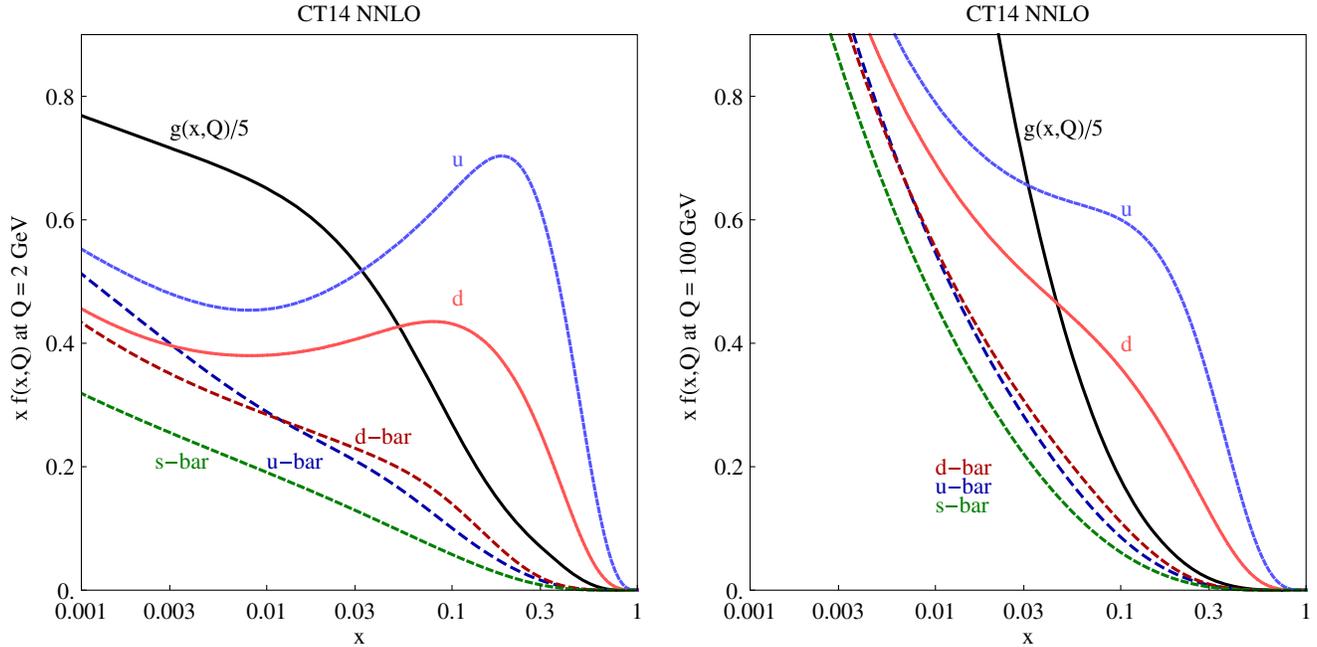
FIG. 5.   The CT14 parton distribution functions at $Q = 2$ GeV and $Q = 100$ GeV for $u, \bar{u}, d, \bar{d}, s = \bar{s}$, and g.

by following the prescription for either the additive or multiplicative type.

In inclusive jet production, the choice between the additive and multiplicative treatments modifies the large-x behavior of the gluon PDF. This has been studied in the CT10 NNLO analysis; cf. Sec. VI D of [6]. In general, the dominant sources of systematic error, especially at the Tevatron and LHC, should be treated as multiplicative rather than additive, that is, by assuming that the relative systematic error corresponds to a fixed fraction of the theoretical value and not of the central data value. The final CT14 PDFs were derived under this assumption, by treating the systematic errors as multiplicative in all experiments.[1] Of course, this is just one option on the table: alternative candidate fits of the CT14 family were also performed, by treating some correlated errors as additive. They produced the PDFs that generally lie within the quoted uncertainty ranges, as in the previous exercise documented in [6].

III. OVERVIEW OF CT14 PDFS AS FUNCTIONS OF x AND Q

Figure 5 shows an overview of the CT14 parton distribution functions, for $Q = 2$ and 100 GeV. The function $xf(x, Q)$ is plotted versus x, for flavors $u, \overline{u}, d, \overline{d}, s = \bar{s}$, and g. We assume $s(x, Q_0) = \bar{s}(x, Q_0)$, since their difference is consistent with zero and has large

---

[1]According to terminology adopted in Refs. [6,81], CT14 implements the correlated errors according to the "extended T" prescription for all experiments, i.e., by normalizing the relative correlated errors by the current theoretical value in each iteration of the fit.

uncertainty [93]. The plots show the *central fit* to the *global data* listed in Tables I and II, corresponding to the lowest total χ^2 for our choice of PDF parametrizations.

The relative changes between the CT10 NNLO and CT14 NNLO ensembles are best visualized by comparing their PDF uncertainties. Figure 6 compares the PDF error bands at 90% confidence level for the key flavors, with each band normalized to the respective best-fit CT14 NNLO PDF. The solid blue and dashed red error bands are obtained for CT14 and CT10 NNLO PDFs at $Q = 100$ GeV, respectively.

Focusing first on the u and d flavors in the upper four subfigures, we observe that the u and $\overline{u}$ PDFs have mildly increased in CT14 at $x < 10^{-2}$, while the d and $\overline{d}$ PDFs have become slightly smaller. These changes can be attributed to a more flexible parametrization form adopted in CT14, which modifies the $SU(2)$ flavor composition of the first-generation PDFs at the smallest x values in the fit.

The CT14 d-quark PDF has increased by 5% at $x \approx 0.05$, after the ATLAS and CMS W/Z production data sets at 7 TeV were included. At $x \gtrsim 0.1$, the update of the D0 charge asymmetry data set in the electron channel, reviewed in Sec. II B 2, has reduced the magnitude of the d quark PDFs by a large amount and has moderately increased the $u(x, Q)$ distribution.

The $\overline{u}(x, Q)$ and $\overline{d}(x, Q)$ distributions are both slightly larger at $x = 0.01$–0.1 because of several factors. At $x = 0.2$–0.5, where there are only very weak constraints on the sea quark PDFs, the new parametrization form of CT14 results in smaller values of $\overline{u}(x, Q)$ and larger values $\overline{d}(x, Q)$, as compared to CT10, although for the most part within the combined PDF uncertainties of the two ensembles.
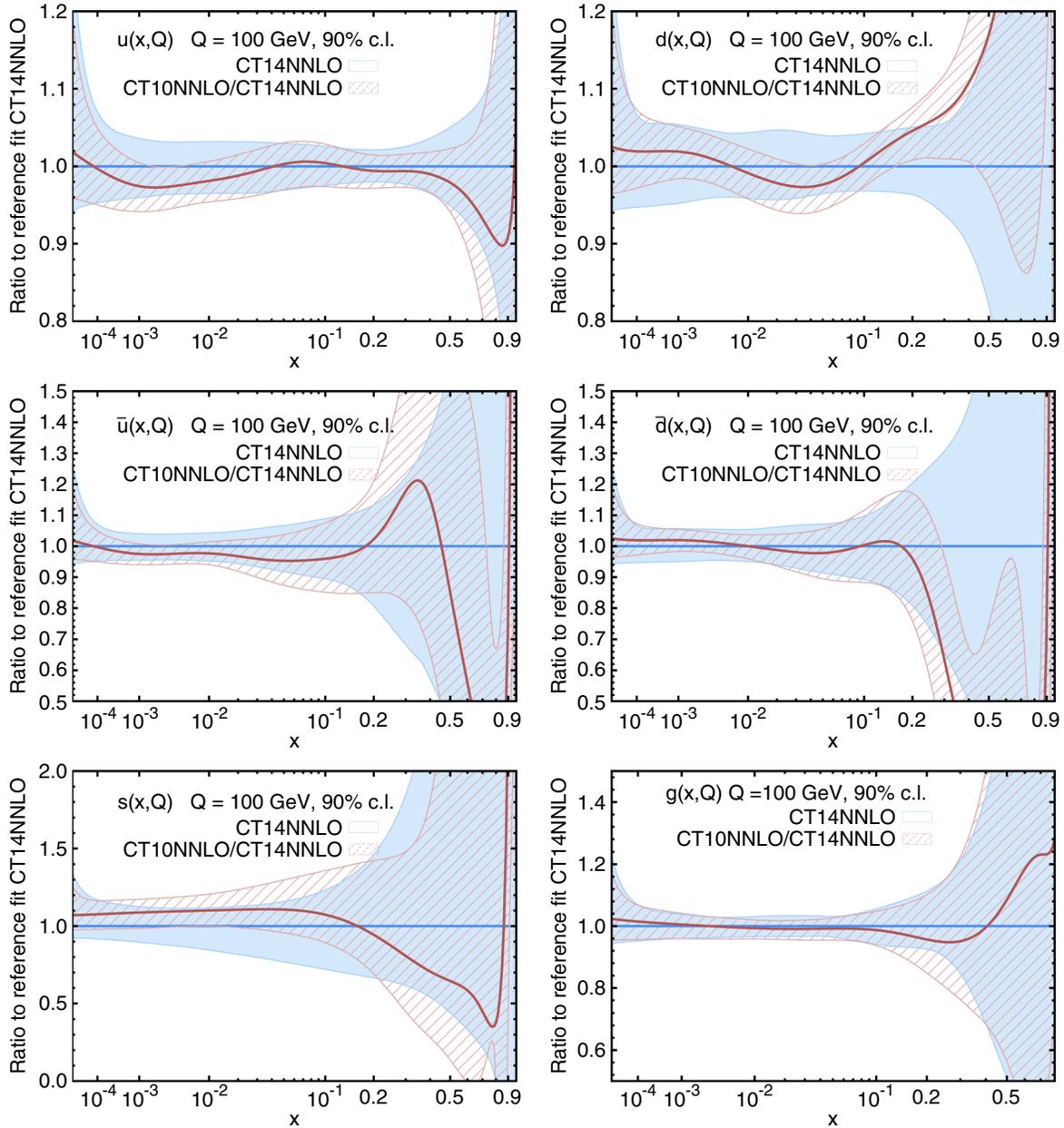
FIG. 6. A comparison of 90% C.L. PDF uncertainties from CT14 NNLO (solid blue curve) and CT10 NNLO (dashed red curve) error sets. Both error bands are normalized to the respective central CT14 NNLO PDFs.

The central strangeness PDF $s(x, Q)$ in the third row of Fig. 6 has decreased for $0.01 < x < 0.15$, but within the limits of the CT10 uncertainty, as a consequence of the more flexible parametrization, the corrected calculation for massive quarks in charged-current DIS, and the inclusion of the LHC data. The extrapolation of $s(x, Q)$ below $x = 0.01$, where no data directly constrain it, also lies somewhat lower than before; its uncertainty remains large and compatible with that in CT10. At large x, above about 0.2, the strange quark PDF is essentially unconstrained in CT14, just as in CT10.

The central gluon PDF (last frame of Fig. 6) has increased in CT14 by 1%–2% at $x \approx 0.05$ and has been

somewhat modified at $x > 0.1$ by the inclusion of the LHC jet production, by the multiplicative treatment of correlated errors, and by the other factors discussed above. For x between 0.1 and 0.5, the gluon PDF has increased in CT14 as compared to CT10.

Let us now review the ratios of various PDFs, starting with the ratio d/u shown in Fig. 7. The changes in d/u in CT14 NNLO, as compared to CT10 NNLO, can be summarized as a reduction of the central ratio at $x > 0.1$, caused by the 9.7 fb^{-1} D0 charge asymmetry data, and an increased uncertainty at $x < 0.05$ allowed by the new parametrization form. At $x > 0.2$, the central CT14 NNLO ratio is lower than that of CT10 NNLO, while their
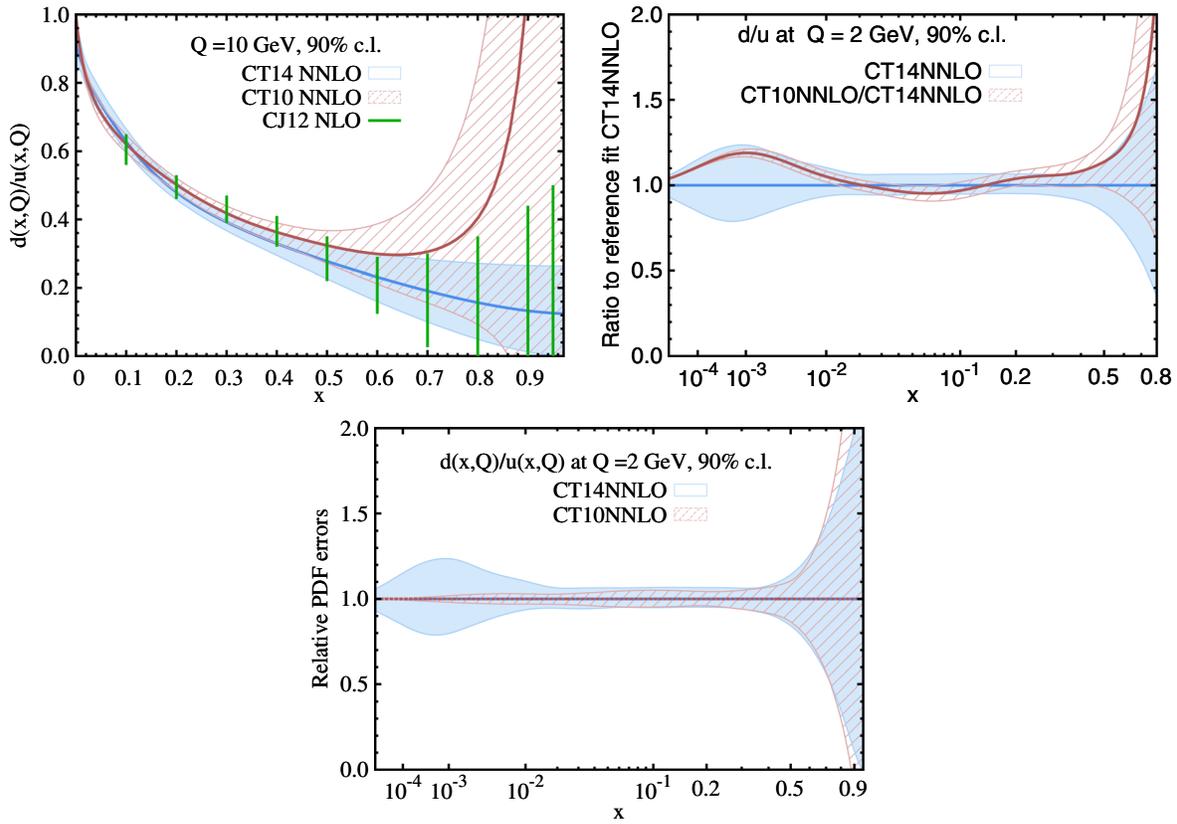
FIG. 7. A comparison of 90% C.L. uncertainties on the ratio $d(x, Q)/u(x, Q)$ for CT14 NNLO (solid blue) and CT10 NNLO (dashed red), and CJ12 NLO (green lines) error ensembles.

relative PDF uncertainties remain about the same. This can be better seen from a direct comparison of the relative PDF uncertainties (normalized to their respective central PDFs) in the third inset. The collider charge asymmetry data constrain d/u at x up to about 0.4. At even higher x, outside of the experimental reach, the behavior of the CT14 PDFs reflects the parametrization form, which now allows d/u to approach any constant value at $x \to 1$.

At such high x, the CTEQ-JLab analysis (CJ12) [94] has independently determined the ratio d/u at NLO, by including the fixed-target DIS data at lower W and higher x that are excluded by a selection cut $W > 3.5$ GeV in CT14 and by considering higher-twist and nuclear effects that can be neglected in the kinematic range of CT14 data. The CT14 uncertainty band on d/u at NNLO lies for the most part between the CJmin and CJmax predictions at NLO that demarcate the CJ12 uncertainty; cf. the first inset of Fig. 7. We see that the CT14 predictions on d/u at $x > 0.1$, which were derived from high-energy measurements that are not affected by nuclear effects, fall within the CJ12 uncertainty range obtained from low-energy DIS with an estimate of various effects beyond leading-twist perturbative QCD. The ratio should be stable to inclusion of NNLO effects; thus, the two ensembles predict a similar trend for collider observables sensitive to d/u.

Turning now to the ratios of sea quark PDFs in Fig. 8, we observe that the uncertainty on $\overline{d}(x, Q)/\overline{u}(x, Q)$ in the left inset has also increased at small x in CT14 NNLO. At $x > 0.1$, we assume that both $\overline{u}(x, Q_0)$ and $\overline{d}(x, Q_0)$ are proportional to $(1 - x)^{a_2}$ with the same power a_2; the ratio $\overline{d}(x, Q_0)/\overline{u}(x, Q_0)$ can thus approach a constant value that comes out to be close to 1 in the central fit, while the parametrization in CT10 forced it to vanish. The uncertainty on $\overline{d}/\overline{u}$ has also increased across most of the x range.

The overall reduction in the strangeness PDF at $x > 0.01$ leads to a smaller ratio of the strange-to-nonstrange sea quark PDFs, $(s(x, Q) + \overline{s}(x, Q))/(\overline{u}(x, Q) + \overline{d}(x, Q))$, presented in the right inset of Fig. 8. At $x < 0.01$, this ratio is determined entirely by parametrization form and was found in CT10 to be consistent with the exact $SU(3)$ symmetry of PDF flavors, $(s(x, Q) + \overline{s}(x, Q))/(\overline{u}(x, Q) + \overline{d}(x, Q)) \to 1$ at $x \to 0$, albeit with a large uncertainty. The $SU(3)$-symmetric asymptotic solution at $x \to 0$ is still allowed in CT14 as a possibility, even though the asymptotic limit of the central CT14 NNLO has been reduced and is now at about 0.6 at $x = 10^{-5}$. The uncertainty of strangeness has increased at such small x and now allows $(s(x, Q) + \overline{s}(x, Q))/(\overline{u}(x, Q) + \overline{d}(x, Q))$ between 0.35 and 2.5 at $x = 10^{-5}$.
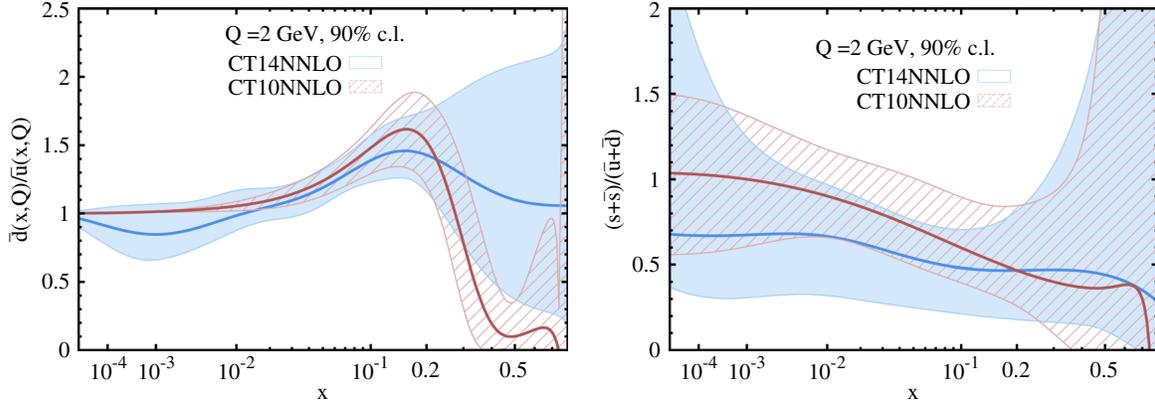
FIG. 8.　A comparison of 90% C.L. uncertainties on the ratios $\bar{d}(x, Q)/\bar{u}(x, Q)$ and $(s(x, Q) + \bar{s}(x, Q))/(\bar{u}(x, Q) + \bar{d}(x, Q))$, for CT14 NNLO (solid blue curve) and CT10 NNLO (dashed red curve) error ensembles.

IV. COMPARISONS WITH HADRONIC EXPERIMENTS

A. Electroweak total cross sections at the LHC

Measurements of total cross sections for production of massive electroweak particles at hadron colliders provide cornerstone benchmark tests of the Standard Model. These relatively simple observables can be both measured with high precision and predicted in NNLO QCD theory with small uncertainties. In this subsection, we collect NNLO theory predictions based on CT14 and CT10 NNLO PDFs for inclusive W and Z boson production, top-quark pair production, Higgs-boson production (through gluon-gluon fusion), at the LHC with center-of-mass energies of 8 and 13 TeV. These theoretical predictions can be compared to the corresponding experimental measurements. We also examine correlations between PDF uncertainties of the total cross sections in the context of the Hessian formalism, following the approach summarized in Ref. [13]. PDF-driven correlations reveal relations between PDF uncertainties of QCD observables through their shared PDF parameters.

The masses of the top quark and Higgs boson are set to $m_t^{\text{pole}} = 173.3$ GeV and $m_H = 125$ GeV, respectively, in this work. The W and Z inclusive cross sections (multiplied by branching ratios for the decay into one charged-lepton flavor), are calculated by using the VRAP v0.9 program [71,72] at NNLO in QCD, with the renormalization and factorization (μ_R and μ_F) scales set equal to the invariant mass of the vector boson. The total inclusive top-quark pair cross sections are calculated with the help of the program TOP++ v2.0 [95,96] at NNLO + NNLL accuracy, with QCD scales set to the mass of the top quark. The Higgs boson cross sections via gluon-gluon fusion are calculated at NNLO in QCD by using the IHIXS v1.3 program [97], in the heavy-quark effective theory (HQET) with finite top quark mass correction, and with the QCD scales set equal to the invariant mass of the Higgs boson.

Figures 9–12 show central predictions and 90% C.L. regions for (W^+, W^-), $(Z, W^\pm)$, $(t\bar{t}, Z)$ and $(t\bar{t}, ggH)$ pairs of
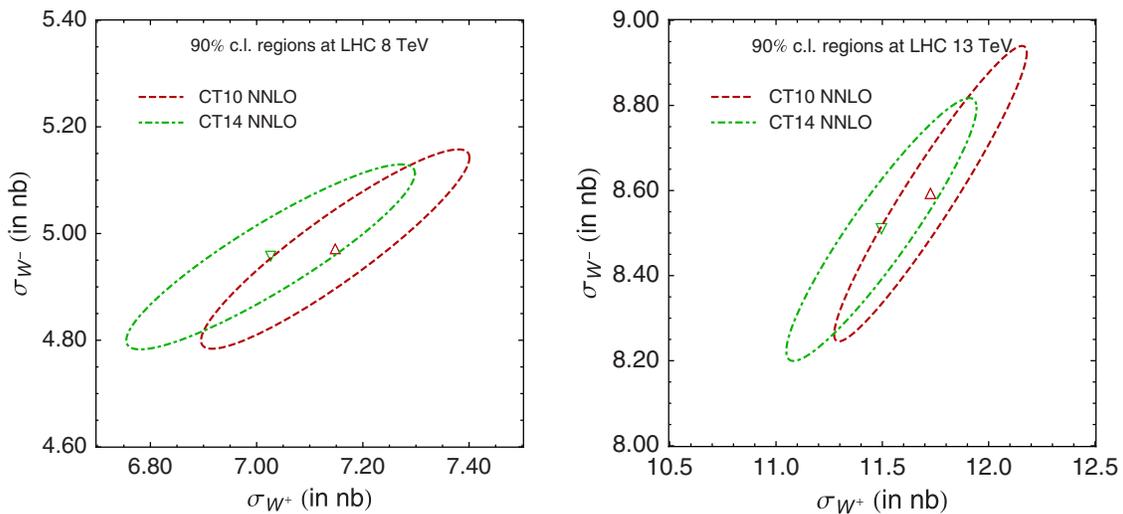

FIG. 9.　The CT14 and CT10 NNLO 90% C.L. error ellipses for the W^- and W^+ cross sections, at the LHC 8 and 13 TeV.
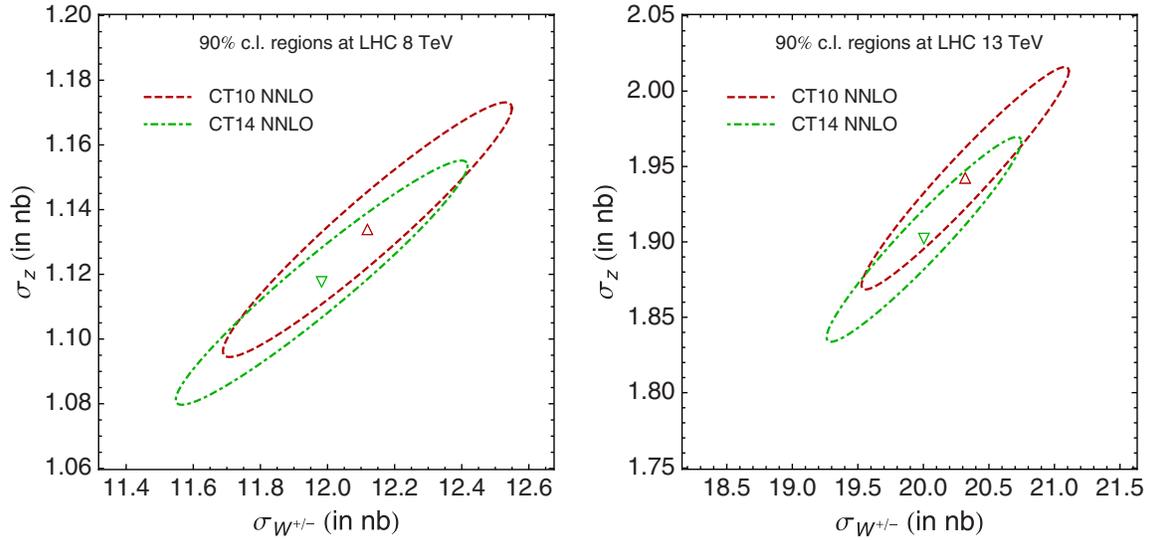
FIG. 10.   CT14 and CT10 NNLO 90% C.L. error ellipses for Z and $W^\pm$ cross sections, at the LHC 8 and 13 TeV.
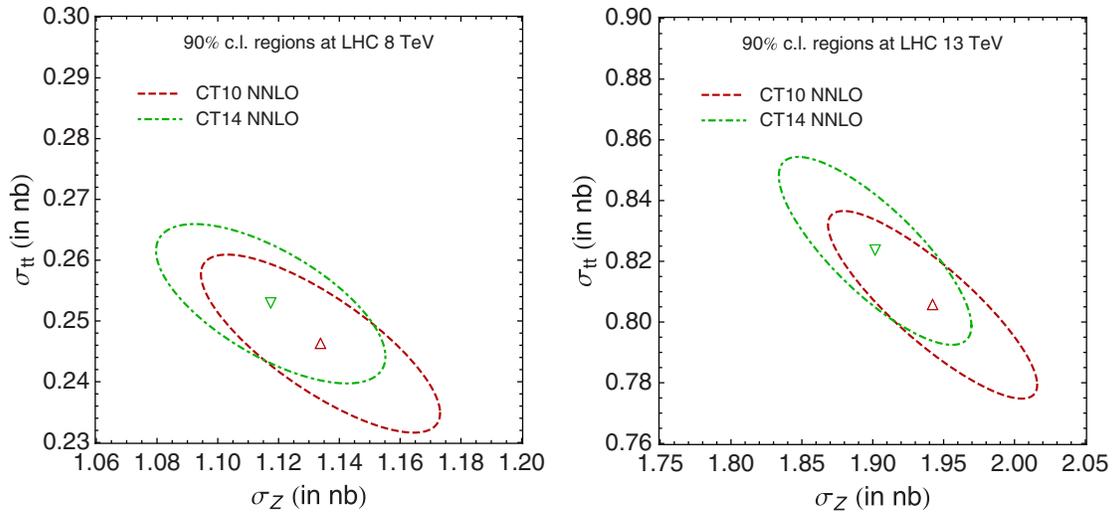

FIG. 11.   CT14 and CT10 NNLO 90% C.L. error ellipses for $t\bar{t}$ and Z cross sections, at the LHC 8 and 13 TeV.
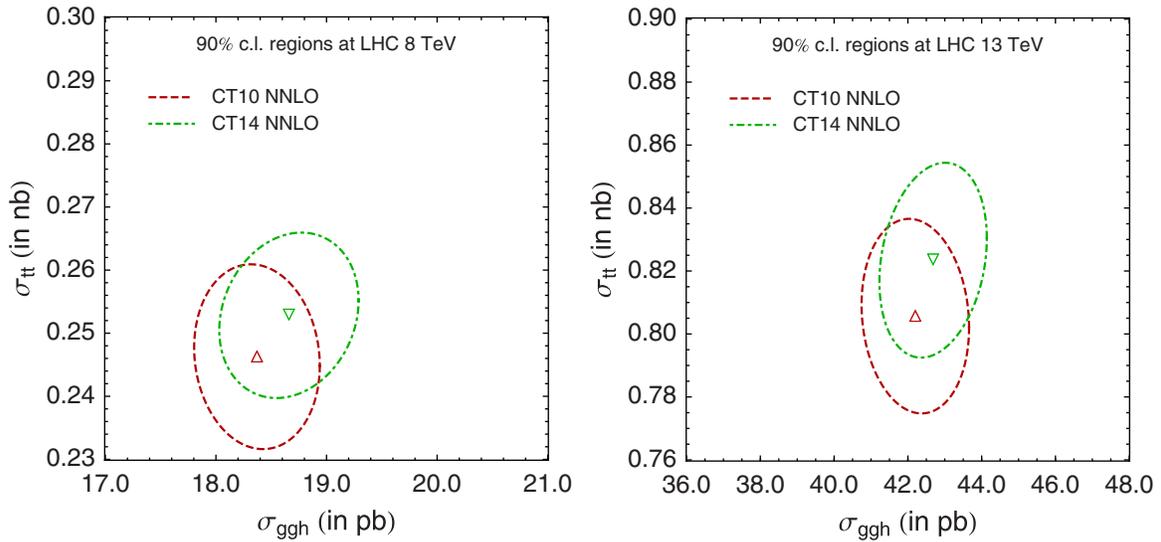

FIG. 12.   CT14 and CT10 NNLO 90% C.L. error ellipses for $t\bar{t}$ and ggH cross sections, at the LHC 8 and 13 TeV.

inclusive cross sections at the LHC 8 and 13 TeV. In each figure, two elliptical confidence regions are shown, obtained with either CT14 or CT10 NNLO PDFs. These can be used to read off PDF uncertainties and correlations for each pair of cross sections. For example, Figs. 9 and 10 indicate that the PDF-induced uncertainties, at the 90% C.L., are about 3.9%, 3.7%, and 3.7% for W^+, W^-, and Z boson production at the LHC 13 TeV, respectively, with CT14 NNLO PDFs. As compared to the results using CT10 NNLO PDFs, the ratio of the total inclusive cross sections of W^+ to W^- productions at the LHC 13 TeV is smaller by about 1% when using CT14 NNLO PDFs, which also provide a slightly larger error (by about a half percent) in that ratio. Specifically, the CT14 NNLO predictions of that ratio at the 68% C.L. are $1.42^{+1.2\%}_{-1\%}$ at LHC 8 TeV and $1.35^{+1\%}_{-1\%}$ at LHC 13 TeV, respectively. The central predictions at 8 TeV are in agreement with the recent CMS measurements [98]. They also show that the electroweak gauge boson cross sections are highly correlated with each other; in fact, much of the uncertainty is driven in this case by the small-x gluon [13].

In Fig. 11, we observe a moderate anticorrelation between the top-quark pair and the Z boson production cross sections. This is a consequence of the proton momentum sum rule mediated by the gluon PDF [13]. In Fig. 12, the Higgs boson cross section through gluon-gluon fusion does not have a pronounced correlation or anticorrelation with the top-quark cross section, because they are dominated by the gluon PDF in different x regions. The Higgs boson and $t\bar{t}$ cross section predictions are further examined in Sec. V. As a result of the changes in PDFs from CT10 to CT14, both the calculated Higgs boson and top-quark pair production cross sections have increased slightly, while the electroweak gauge boson cross sections have decreased. However, the changes of the central predictions are within the error ellipses of either CT14 or CT10.

B. LHC and Tevatron inclusive jet cross sections

We now turn to the comparisons of CT14 PDFs with new LHC cross sections on inclusive jet production. We argued in Sec. II that PDF uncertainty of inclusive jet production at the LHC is strongly correlated with the gluon PDF in a wider range of x than in the counterpart measurements at the Tevatron. The true potential of LHC jets for constraining the gluon PDF also depends on experimental uncertainties, which we can now explore for the first time using the CMS and ATLAS data on inclusive jet cross sections at 7 TeV.

We first note that, in the context of our analysis, the single-inclusive jet measurements at the LHC are found to be in reasonable consistency with the other global data, including Tevatron run-2 single-inclusive jet cross sections measured by the CDF and D0 Collaborations. The values of χ^2 for the four jet experiments (ID = 504, 514, 535, and
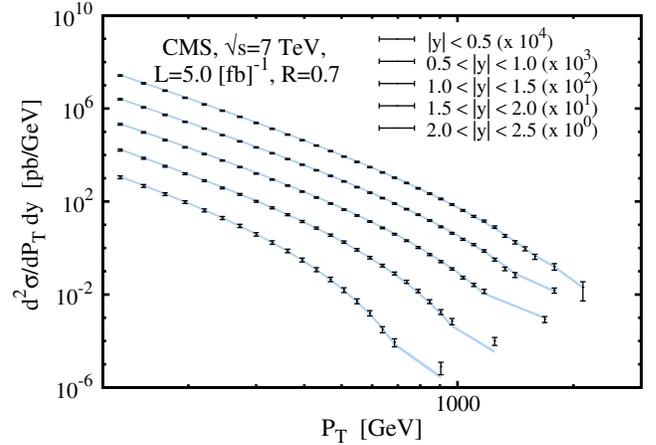

FIG. 13. Comparison of data and theory for the CMS 7 TeV inclusive jet production, for CT14 NNLO PDFs. Measurements of $d^2\sigma/dp_T dy$ for five rapidity bins are plotted as functions of jet p_T. The points are data with total experimental errors, obtained by adding the statistical and systematic errors in quadrature. The bands are theoretical calculations with 68% C.L. PDF uncertainties.

538) are listed at the end of Table II. We obtain very good fits ($\chi^2/N_{pt} = 1.09$ and 0.55) to the D0 and ATLAS jet data sets and moderately worse fits ($\chi^2/N_{pt} = 1.45$ and 1.33) to the CDF and CMS data sets. The description of the Tevatron jet data sets has been examined as a part of the CT10 NNLO study [6], where it was pointed out that the χ^2 for the CDF run-2 measurement tends to be increased by random, rather than systematic, fluctuations of the data. In regards to describing the Tevatron jet data, the CT14 NNLO PDFs follow similar trends as CT10 NNLO.

1. CMS single-inclusive jet cross sections

Figure 13 shows a comparison between the measurements for the CMS inclusive jet data at 7 TeV [50] and NLO theory prediction [86,99,100] utilizing CT14 NNLO PDFs. We discussed earlier in the paper that the missing NNLO contributions to the hard-scattering cross section can be anticipated to be small under our QCD scale choices, compared to the experimental uncertainty.

The CMS data, with 5 fb^{-1} of integrated luminosity, employ the anti-k_T jet algorithm [101] with jet radius $R = 0.7$. The measurements are divided into five bins of rapidity and presented as a function of the p_T of the jet, with a total of 133 data points. The theoretical prediction based on the CT14 NNLO PDFs reproduces the behavior of experimental cross sections across 13 orders of magnitude.

Figure 14 provides a more detailed look at these distributions, by plotting the shifted central data values divided by the theory. The data are shifted by optimal amounts based on the treatment of the systematic errors as nuisance parameters; cf. Ref. [6]. The error bars for the shifted data include only uncorrelated errors, i.e., statistical and uncorrelated systematic errors added in quadrature.
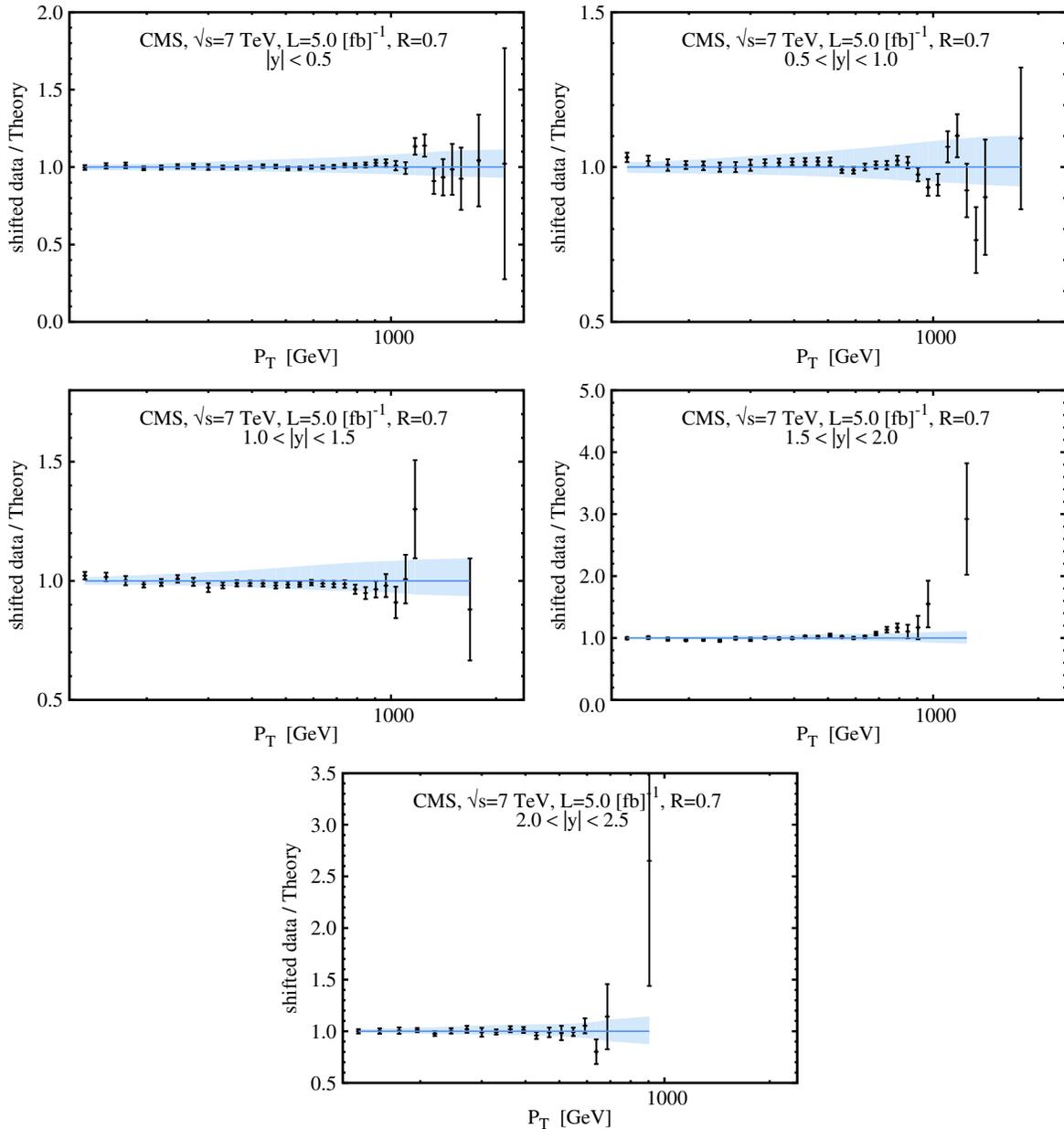
FIG. 14.   The same as Fig. 13, shown as the ratio of shifted data for CMS 7 TeV divided by theory. The error bars correspond to total uncorrelated errors. The shaded region shows the 68% C.L. PDF uncertainties.

Here we notice moderate differences (up to a few tens of percent of the central prediction) between theory and shifted data, which elevate χ^2 for this data set by about 2.5 standard deviations for the central CT14 PDF set or less for the error PDF sets.

Although they are not statistically significant, the origin of these mild discrepancies can be further explored by studying the correlated shifts allowed by the systematic uncertainties. In our implementation of systematic errors [6], each correlated uncertainty α is associated with a normally distributed random nuisance parameter λ_α. When $\lambda_\alpha \neq 0$, it may effectively shift the central value of the data point i in the fit by

$$\beta_{i,\alpha}\lambda_\alpha = \sigma_{i,\alpha}X_i\lambda_\alpha,$$

where $\sigma_{i,\alpha}$ is the published fractional 1-σ uncertainty of data point i due to systematic error α. X_i is the cross section value that normalizes the fractional systematic uncertainty [6], set equal to the theoretical value T_i in the procedure of the current analysis.

Each λ_α is adjusted to optimize the agreement between theory and data. Figure 15 shows a histogram of the best-fit λ_α for the 19 sources of the systematic errors published by CMS [50]. In an ideal situation, the optimized $\{\lambda_\alpha : \alpha = 1...19\}$ would be normally distributed with a mean value of 0 and standard deviation of 1. The actual distribution of the

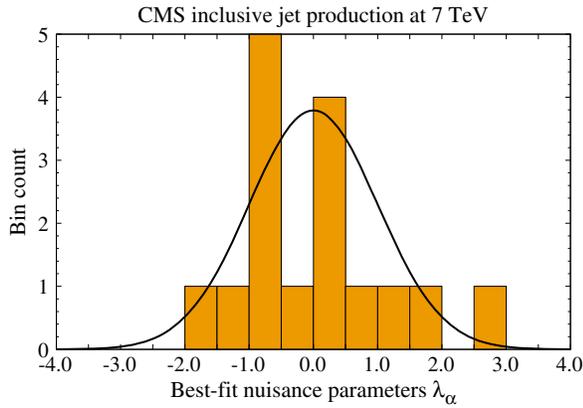FIG. 15.   Histogram of optimized nuisance parameters λ_α for the sources of correlated systematic errors of the CMS 7 TeV inclusive jet production. The curve is the standard normal distribution expected in the ideal case.

λ_α values in Fig. 15 appears to be somewhat narrower than the standard normal one. This and relatively high $\chi^2/N_{pt} = 1.33$ may indicate that either uncorrelated systematic uncertainties are underestimated, or higher-order theoretical calculations are needed to describe the data.

2. ATLAS single-inclusive jet cross sections

Equivalent comparisons for the ATLAS 7 TeV inclusive jet production with 37 pb^{-1} of integrated luminosity [49] are presented in Figs. 16–18. In this case, we compare to data in seven bins of rapidity for the anti-k_T jet algorithm [101] with jet radius $R = 0.6$. The agreement is excellent in all figures, not the least because both statistical and systematic errors are still large in this early data set. Among 119 sources of experimental errors that were identified, many have little impact on the best fit. The resulting distribution of the nuisance parameters in Fig. 18
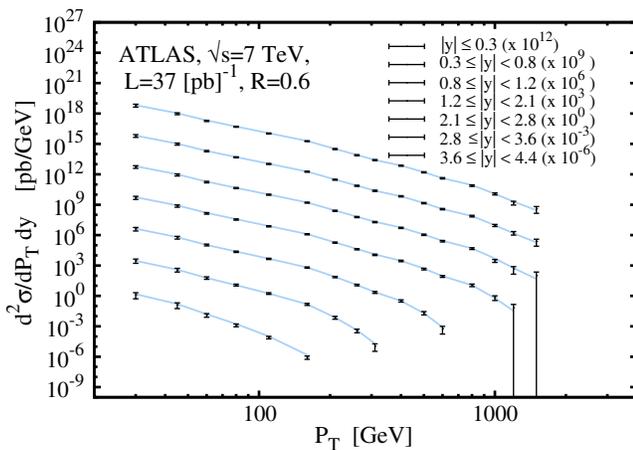
FIG. 16.   Comparison of data on $d^2\sigma/dp_T dy$ and NLO theory for the ATLAS 7 TeV inclusive jet production, using CT14 NNLO PDFs.

at the best fit is much narrower than the ideal Gaussian distribution, indicating that most of the correlated sources need not deviate from their nominal values when the PDFs are fitted.

To summarize, Figs. 13–18 demonstrate that CT14 PDFs agree with both sets of CMS and ATLAS single-inclusive jet cross sections. The ATLAS Collaboration also measured inclusive jet production at center of mass energy $\sqrt{s} = 2.76$ TeV and published ratios between the 2.76 and 7 TeV measurements in Ref. [102]. These two measurements are well described by the theory prediction using CT14, with a $\chi^2/N_{pt} \approx 1$. Furthermore, the ATLAS Collaboration published the inclusive jet measurements using another choice of jet radius of 0.4 [49]. Both ATLAS and CMS Collaborations measured cross sections for dijet production [49,50] based on the same data sample of the single-inclusive jet measurements. These measurements are not included in the CT14 global analysis because of the correlations between the two (dijet and single-inclusive jet) data sets. However, it has been verified that the CT14 analysis gives a good description for all these data sets as well.

C. Differential cross sections for lepton pair production at the LHC

1. Charged-lepton pseudorapidity distributions in W/Z boson production

Differential cross sections for production of massive vector bosons set important constraints on the flavor composition of the proton, notably on the u and d quarks, antiquarks and their ratios. Figure 19 compares CT14 NNLO theoretical predictions with pseudorapidity ($|\eta_\ell|$) distributions of charged leptons from inclusive $W^\pm$ and Z^0 production and decay in the 2010 ATLAS 7 TeV data sample with 33–36 pb^{-1} of integrated luminosity [46]. Theoretical predictions are computed using the program RESBOS. The black data points represent the unshifted central values of the data. The error bars indicate the total (statistical + systematic) experimental error. The blue band is the CT14 PDF uncertainty evaluated at the 68% C.L. These three measurements share correlated systematic errors. From the figures, we see that the data are described well by theory over the entire rapidity range, even in the absence of correlated systematic shifts. The PDF uncertainties are similar in their size to those of the experimental measurements and, overall, the theory predictions are within one standard deviation of the data.

2. Influence of W boson charge asymmetry measurements at the LHC

Another handy observable for determining the parton distribution functions is the *charge asymmetry* for W^+ and W^- bosons produced in pp or $p\bar{p}$ collisions. This process has been measured both at the Tevatron and at the LHC. As
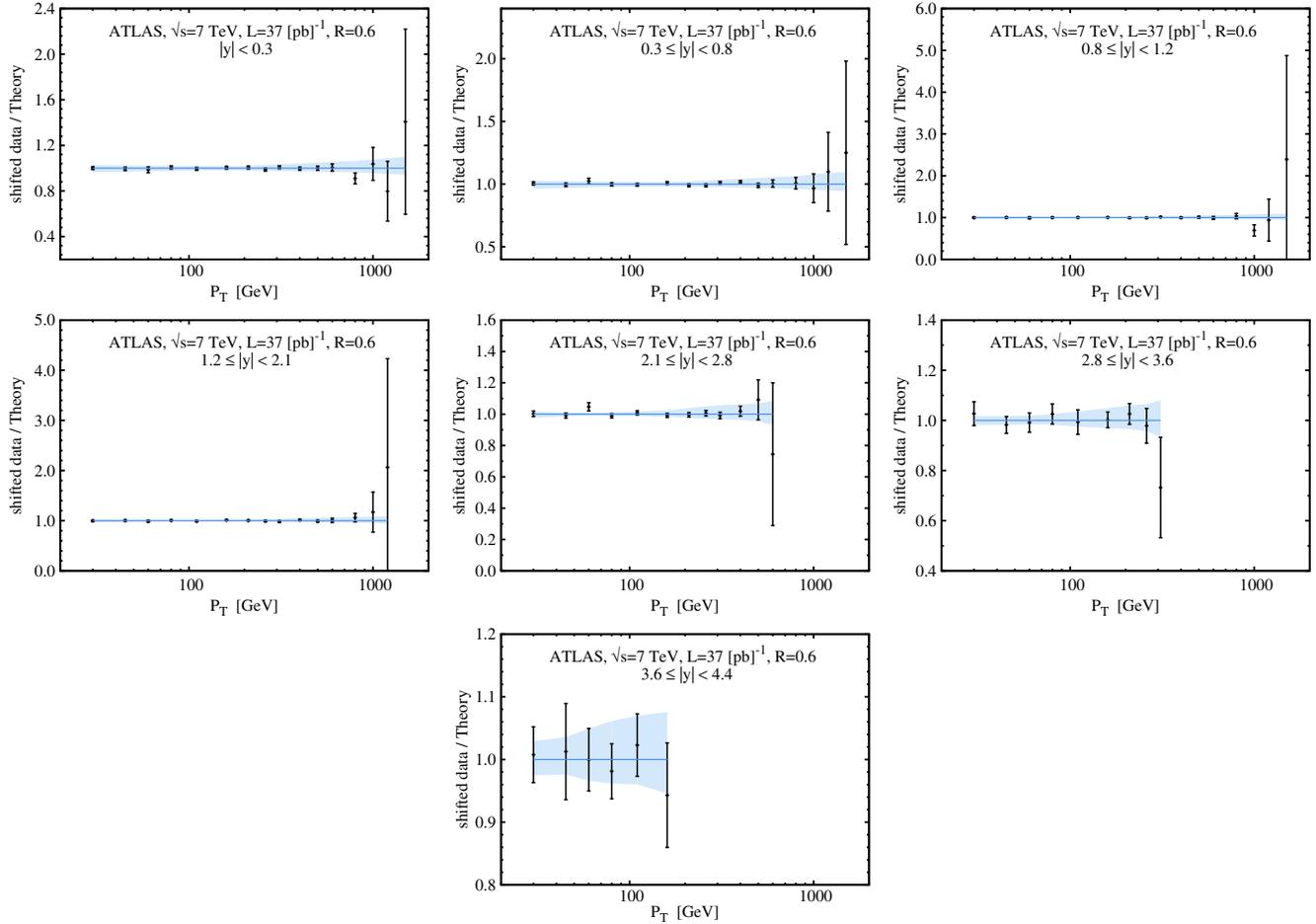
FIG. 17.   The same as Fig. 16, shown as the ratio of shifted data for ATLAS 7 TeV divided by the NLO theory. The error bars correspond to total uncorrelated errors. The shaded region shows the 68% C.L. PDF uncertainties.

the asymmetry involves a ratio of the cross sections, many experimental systematic errors cancel, leading to very precise results. Without these collider data, the main information about the difference between the light flavors,
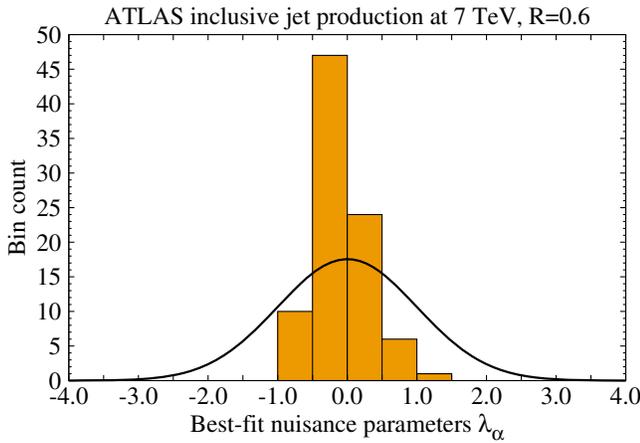

FIG. 18.   Histogram of optimized nuisance parameters λ_α for the sources of correlated systematic errors of the ATLAS 7 TeV inclusive jet production.

$d, \bar{d}$ and $u, \bar{u}$, would come from the BCDMS and NMC experiments, which are measurements of muon deep-inelastic scattering on proton and deuteron targets. Under the assumption of charge symmetry between the nucleons, the difference of the proton and deuteron cross sections distinguishes between the u and d PDFs in a nucleon. However, the deuteron measurements are subject to nuclear binding corrections, which have been estimated by introducing nuclear models [94,103,104], but are not calculable from first principles. In contrast, the $W^\pm$ charge asymmetry data from the Tevatron and LHC colliders directly provide information about the difference between d and u flavors, without the need for nuclear corrections. By including the ATLAS and CMS charge asymmetry data, we are able to obtain, for the first time, direct experimental constraints on the differences of the quark and antiquark PDFs for u and d flavors at $x \approx 0.02$ typical for the 7 TeV kinematics.

Figure 20 shows a comparison of data and theory, for the lepton charge asymmetry of inclusive $W^\pm$ production, from the ATLAS experiment at the LHC 7 TeV [46]. These asymmetry data are correlated with the W/Z rapidity
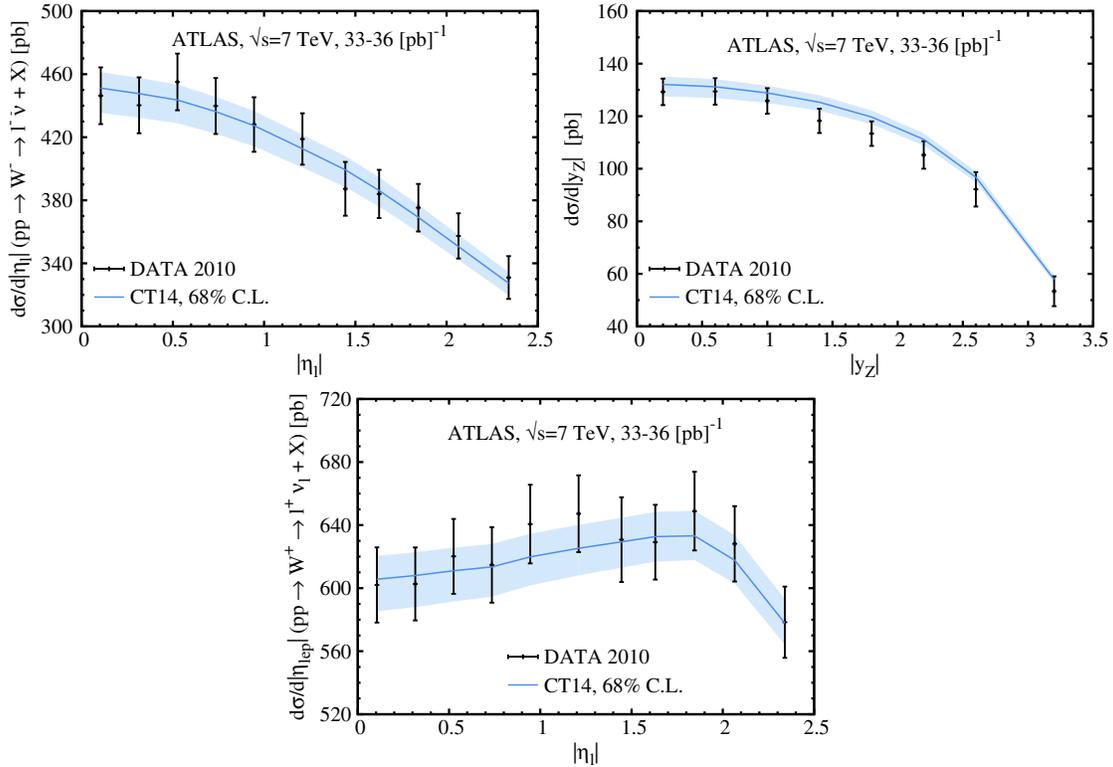
FIG. 19. Comparison between the 2010 ATLAS measurements [46] of the $W^{\pm}$ charged-lepton pseudorapidity and Z boson rapidity distributions at $\sqrt{s}=7$ TeV and the RESBOS theory using CT14 NNLO PDFs.

measurements discussed in the previous subsection; all four W/Z data sets are included in the CT14 global analysis using a *shared* correlation matrix from the ATLAS publication [46]. The measurement was carried out with several kinematic cuts. The lepton transverse momentum was required to be greater than 20 GeV, the missing transverse energy to be greater than 25 GeV, and the lepton-neutrino transverse mass to be greater than 40 GeV. The shaded region is the PDF uncertainty of CT14 NNLO at 68% C.L.
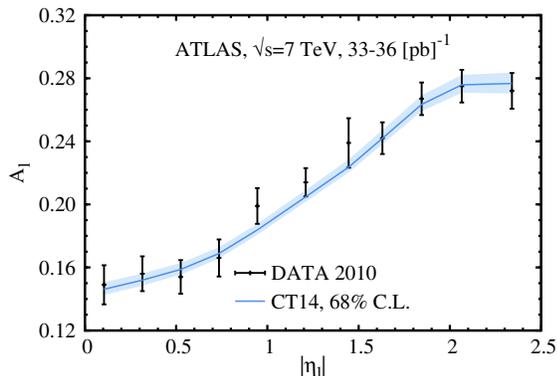
FIG. 20. $W^{\pm}$ charge asymmetry as a function of lepton pseudorapidity measured by the ATLAS Collaboration, compared to the 68% C.L. CT14 NNLO uncertainty band. The kinematic requirements are $p_{T\ell} > 20$ GeV, $p_{T\nu_{\ell}} > 25$ GeV and $M_T^{\ell\nu_{\ell}} > 40$ GeV.

Again the points with error bars represent the unshifted data with the experimental errors added in quadrature. The data fluctuate around the CT14 predictions and are described well by the CT14 error band.

Figure 21 presents a similar comparison of the unshifted data and CT14 NNLO theory for the charge asymmetry of decay muons [44] and electrons [45] from inclusive $W^{\pm}$ production from the CMS experiment at the LHC 7 TeV. The asymmetry for muons is measured with 4.7 fb^{-1} of integrated luminosity, with $p_{T\ell} > 25$ and 35 GeV; the asymmetry for electrons is measured with 840 pb^{-1} and $p_{T\ell} > 35$ GeV. Here we note that the CMS measurement does not apply a missing E_T cut to A_{ch}, contrary to the counterpart ATLAS A_{ch} measurement. Theory predictions are the same for both the muon and electron channels with the same cuts. The muon and electron data are consistent with one another, but the muon data have smaller statistical and systematic uncertainties, as is apparent in Fig. 21. All three subsets of CMS A_{ch} agree with predictions using CT14; their χ^2 is further improved by optimizing the correlated shifts. The electron data and the muon data with the $p_{T\ell}$ cut of 35 GeV are included in the CT14 global analysis. The muon data with a $p_{T\ell}$ cut of 25 GeV are not included in the CT14 analysis but nevertheless are well described.

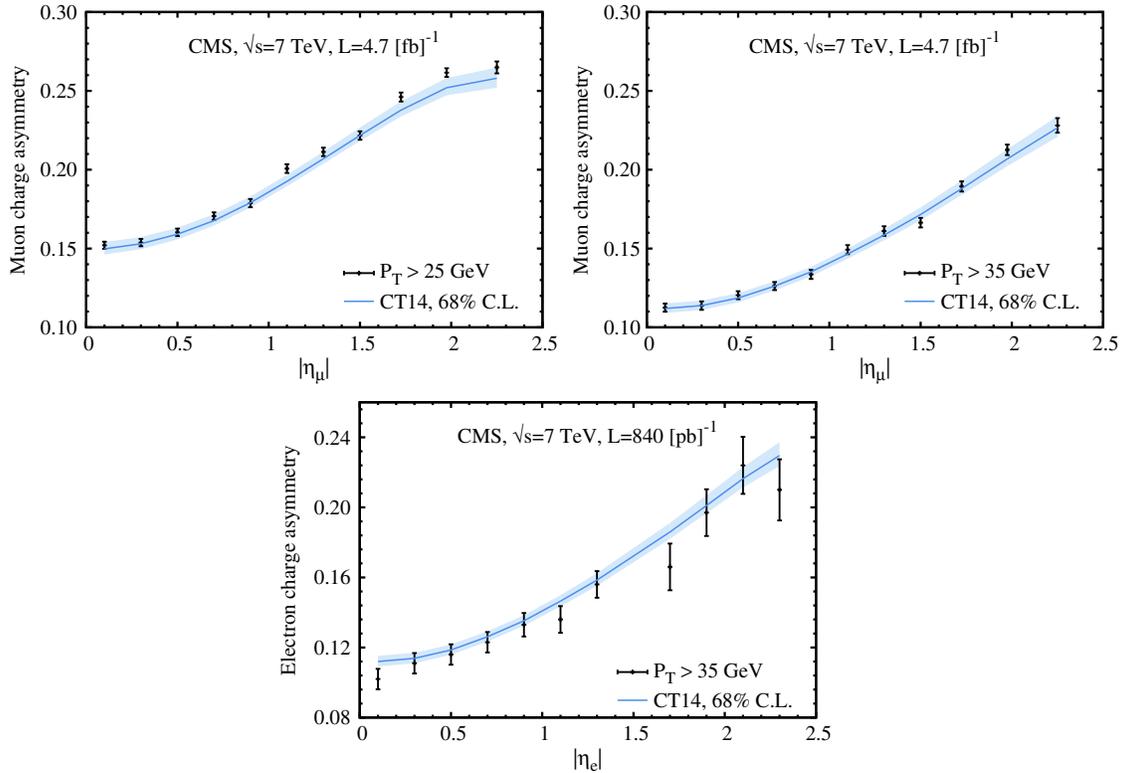In the LHCb measurement of the charged-lepton asymmetry at 7 TeV [41], the muons are required to have a

FIG. 21. Charge asymmetry of decay muons and electrons from $W^{\pm}$ production measured by the CMS experiment. The data values have $p_{T\ell} > 25$ or 35 GeV for the muon data and $p_{T\ell} > 35$ GeV for the electron data. The vertical error bars on the data points indicate total (statistical and systematic) uncertainties. The curve shows the CT14 theoretical calculation; the shaded region is the PDF uncertainty at 68% C.L.

transverse momentum greater than 20 GeV. The corresponding comparison of the CT14 NNLO predictions to the LHCb A_{ch} data is shown in Fig. 22. The LHCb case is especially interesting, as the LHCb acceptance for charged leptons extends beyond the rapidity range measured by ATLAS and CMS. Thus, the LHCb results are sensitive to the u and d quark PDFs at larger x values than at the
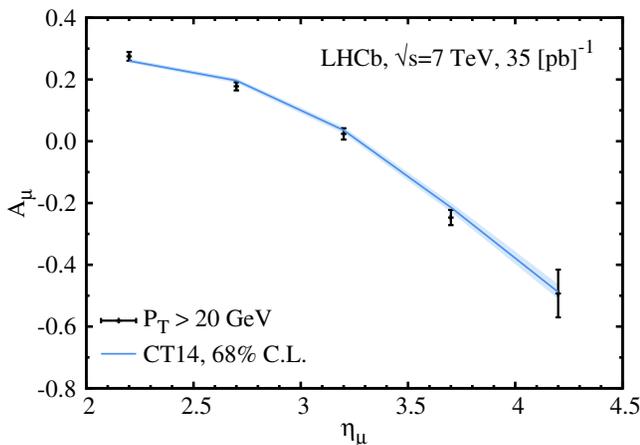

FIG. 22. Charge asymmetry of decay muons from $W^{\pm}$ production measured by the LHCb experiment.

ATLAS or CMS. Good agreement between data and theory is again observed.

3. Production of Drell-Yan pairs at ATLAS

In Figs. 23 and 24, we compare CT14 NNLO predictions to ATLAS 7 TeV measurements of differential cross sections for production of high-mass [105] and low-mass [106] Drell-Yan pairs, plotted as a function of dilepton invariant mass $m_{\ell\ell}$. The experimental cross sections correspond to the "electroweak Born level," unfolded from the raw data by correcting for electroweak final-state radiation. The high-mass data sample corresponds to $116 < m_{\ell\ell} < 1000$ GeV. At low dilepton masses, we compare to the combined electron + muon sample at $26 < m_{\ell\ell} < 66$ GeV for $L = 1.6$ fb^{-1} in the upper row, as well as to the muon sample at $12 < m_{\ell\ell} < 66$ GeV for $L = 35$ pb^{-1} in the lower row. Fiducial acceptance cuts on the decay leptons are specified inside the figures. Correlated experimental uncertainties are included in the comparison.

On the theory side, the cross sections are calculated at NNLO in QCD with APPLGRID interface [87] to FEWZ [107–110], and including photon-scattering contributions. Experimental uncertainties in these cross sections tend to be larger than the PDF uncertainties, as illustrated by the figures; hence, we only compare these data to the CT14
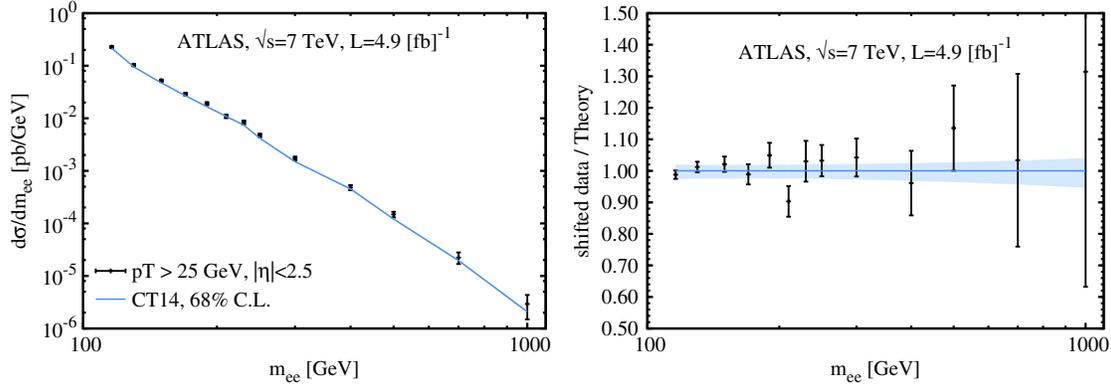
FIG. 23.   Invariant mass distributions of Drell-Yan pairs in the high-mass region by ATLAS 7 TeV [105], with superimposed NNLO predictions based on CT14 NNLO PDFs. The left subfigure shows the differential cross sections as a function of the dilepton mass m_{ee}. The right subfigure shows the ratio of ATLAS shifted data to CT14 theory predictions.

predictions *a posteriori*, without actually including them in the CT14 fit.

It can be observed in the figures that CT14 NNLO PDFs agree well with the high-mass and low-mass data samples both in terms of the cross sections (in the left subfigures) and ratios of the shifted data to theoretical predictions (right subfigures). The PDF uncertainty bands, indicated by light blue color, approximate the average behavior of the experimental data without systematic discrepancies.

D. $W^{\pm}$ charge asymmetry from the D0 experiment at the Tevatron

We reviewed above that, historically, measurements of $W^{\pm}$ charge asymmetry at the Tevatron have been important in the CTEQ-TEA global analysis. For example, the CTEQ6 PDFs (circa 2002) and CT10 PDFs (circa 2010–2012) included the $W^{\pm}$ asymmetry data from the CDF and D0 experiments to supplement the constraints on u and d quark PDFs at $x > 0.1$ from fixed-target DIS experiments.
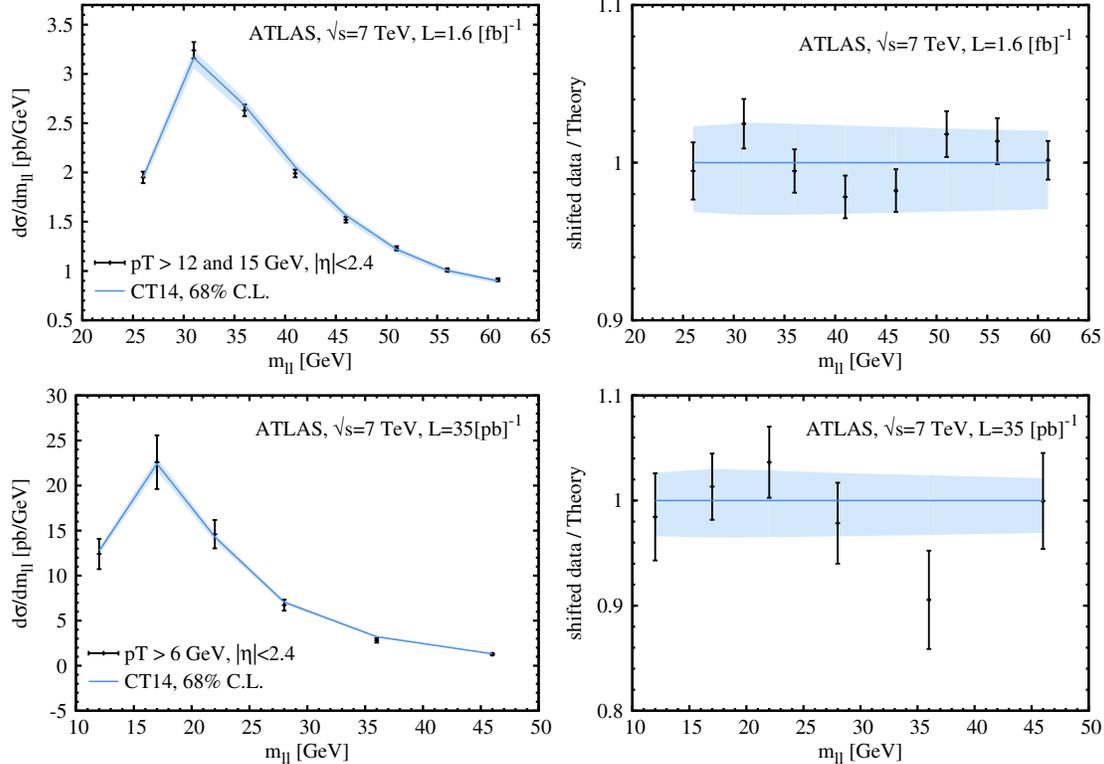

FIG. 24.   The same as in Fig. 23, for ATLAS 7 TeV differential distributions of Drell-Yan pairs in the low-mass and extended low-mass regions [106].
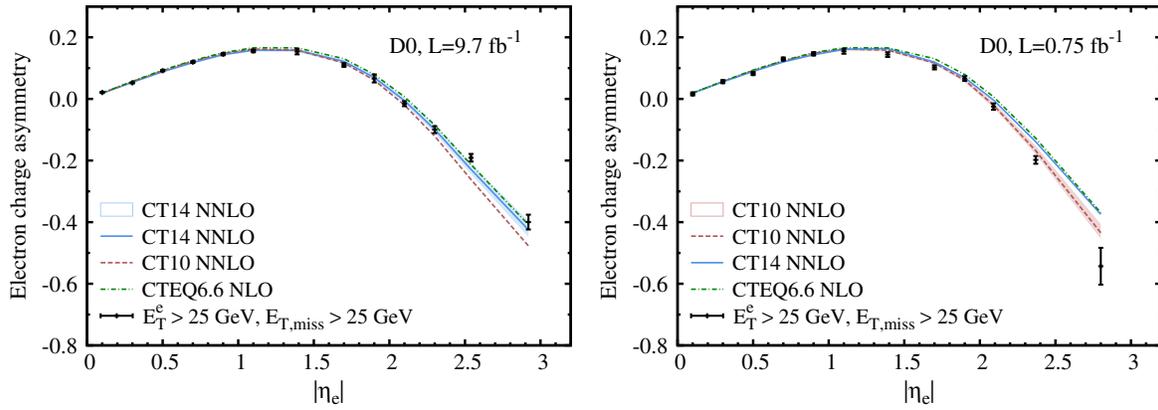
FIG. 25.   Charge asymmetry of decay electrons from $W^\pm$ production measured by the DØ experiment in run 2 at the Tevatron with high (left) and low (right) luminosities, compared to several generations of CTEQ-TEA PDFs.

The charge asymmetry at the Tevatron probes the differences of the *slope in x* of the PDFs for u and d flavors.

A new $W^\pm$ charge asymmetry measurement from the D0 experiment at the Tevatron has recently been published, using the full integrated luminosity (9.7 fb^{-1}) from run 2 [14]. The experimental uncertainties, both statistical and systematic, are smaller than in the previous A_{ch} measurement [21]. Figure 25 compares the D0 run-2 data and various theoretical predictions at NNLO for both the latest (left) and the previous D0 data set (right). We show the unshifted data with the total experimental errors as error bars and the 68% C.L. PDF uncertainties as the shaded regions. As an alternative representation, Fig. 26 shows the differences between theory and shifted data, where the error bars represent the uncorrelated experimental errors. From the two figures, we conclude that it is difficult to fit both data sets well, given the smallness of the systematic shifts associated with A_{ch}. While the 9.7 fb^{-1} electron data set is in better agreement with the global data, including the D0 muon [40] and CDF [39] A_{ch} measurements, the best-fit χ^2/N_{pt} for the 9.7 fb^{-1} sample remains relatively high

(about 2) and is sensitive to detailed implementation of NNLO corrections. In-depth studies on the D0 asymmetry data will be presented in a forthcoming paper. When the high-luminosity D0 A_{ch} measurement was substituted for the low-luminosity one, we observed reduction in the d/u ratio at $x > 0.1$ compared to CT10W NLO and CT10 NNLO sets.

In total, constraints from the LHC and Tevatron W/Z differential cross sections and asymmetries lead to important changes in the quark sector PDFs, as documented in Sec. III. At $x \lesssim 0.02$, we obtain more realistic error bands for the u, $\bar{u}$, d, $\bar{d}$ PDFs upon including the ATLAS and CMS data sets. At $x > 0.1$, the high-luminosity D0 charge asymmetry and other compatible experiments predict a softer behavior of $d(x, Q)/u(x, Q)$ than in CT10W.

E. Constraints on strangeness PDF from CCFR, NuTeV, and LHC experiments

Let us now turn to the strangeness PDF $s(x, Q)$, which has become smaller at $x > 0.05$ in CT14 compared to our previous analyses, CT10 and CTEQ6.6. Although the CT14 central $s(x, Q)$ lies within the error bands of either
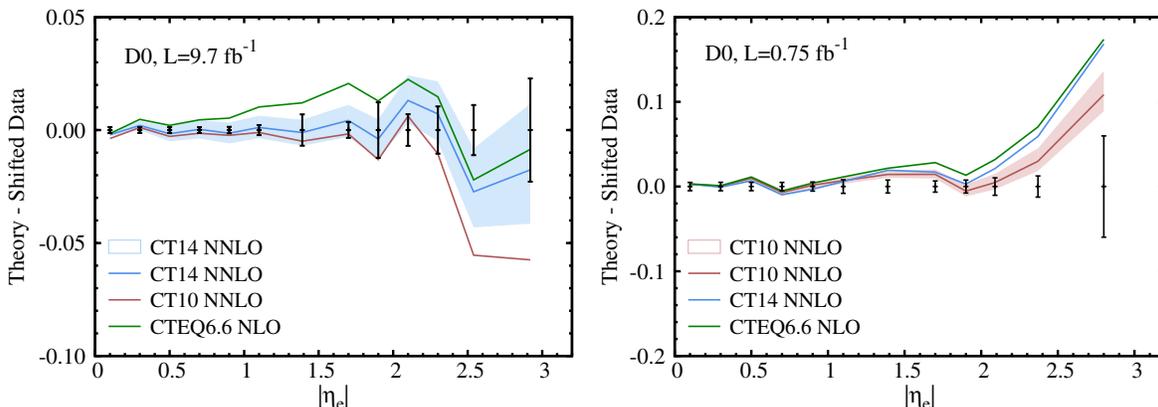

FIG. 26.   The same as Fig. 25, plotted as the difference between theory and shifted data for A_{ch} from D0 run 2 (9.7 fb^{-1}).

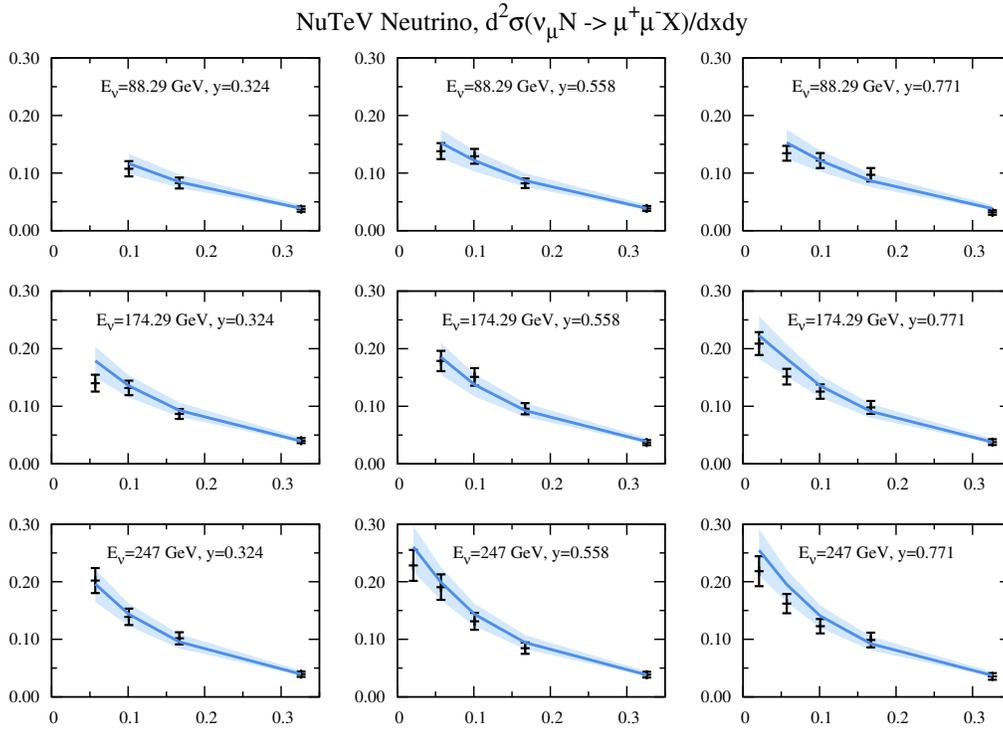NuTeV Neutrino, $d^2\sigma(\nu_\mu N \to \mu^+\mu^- X)/dxdy$

FIG. 27. Comparison of data and theory for the NuTeV measurements of dimuon production in neutrino-iron collisions. The data are expressed in the form of $d^2\sigma/dxdy$ and shown as a function of x for a certain y and neutrino energy.

earlier PDF set, it is important to verify that it is consistent with the four fixed-target measurements that are known to be sensitive to $s(x, Q)$: namely, measurements of dimuon production in neutrino and antineutrino collisions with iron targets, from the CCFR [30] and NuTeV [29] Collaborations (ID = 124–127).

Predictions using previous CTEQ PDFs were in agreement with these four experiments. In Table I for CT14, the four corresponding χ^2 values are also good. Supporting evidence comes from the point-by-point comparisons in Figs. 27 and 28, between the theoretical cross sections for CT14 NNLO PDFs and the dimuon data from the NuTeV experiment in neutrino and antineutrino scattering. The analogous comparisons for the CCFR experiment are in Figs. 29 and 30. Given the size of the measurement errors and of the PDF uncertainty, it is clear that CT14 central predictions provide a good description of the dimuon cross sections. Also, our estimate for the *uncertainty* of the strange PDF looks reasonable: it is comparable to the measurement errors for these cross sections, which are known to be sensitive mostly to the strange quark PDF.

Nevertheless, the CT14 central strangeness PDF lies on the lower side of the CT10 PDF uncertainty in some kinematic ranges. As mentioned in the introduction, the reduction is in part attributable to elimination of a computational error (wrong sign of a term) in the treatment of heavy-quark mass effects in charged-current DIS in post-CTEQ6.5 analyses and in part from other sources,

especially introduction of the LHC W/Z data, and more flexible parameterizations for all PDF flavors.

The ATLAS and CMS experimental collaborations have recently published studies on the strangeness content of the proton and have come to somewhat discrepant conclusions. On the ATLAS side, two papers were published, one in 2012 [111] and one in 2014 [112]. In the 2012 study, the inclusive DIS and inclusive $W^\pm$ and Z boson production measurements [46] were employed to determine the strangeness fraction of the proton for one value of (x, Q). In the 2014 study, the ATLAS 7 TeV $W + c$-jet, $W + D^{(\star)}$ [112], and inclusive $W^\pm/Z$ cross sections were used. These two analyses determined the ratio (r^s) of strange to down-sea quark PDF,

$$r^s \equiv 0.5 \frac{(s + \bar{s})}{\bar{d}} \quad \text{at } x = 0.023, \qquad Q = 1.4 \text{ GeV}. \quad (3)$$

They find

$$\begin{aligned} r^s &= 1.00^{+0.25}_{-0.28} \quad \text{ATLAS(2012)}, \\ r^s &= 0.96^{+0.26}_{-0.30} \quad \text{ATLAS(2014)}, \end{aligned} \quad (4)$$

which imply a rather large strangeness density.

In 2014, the CMS Collaboration [44] determined the ratios

NuTeV Anti-Neutrino, $d^2\sigma(\bar{\nu}_\mu N \rightarrow \mu^+\mu^- X)/dxdy$
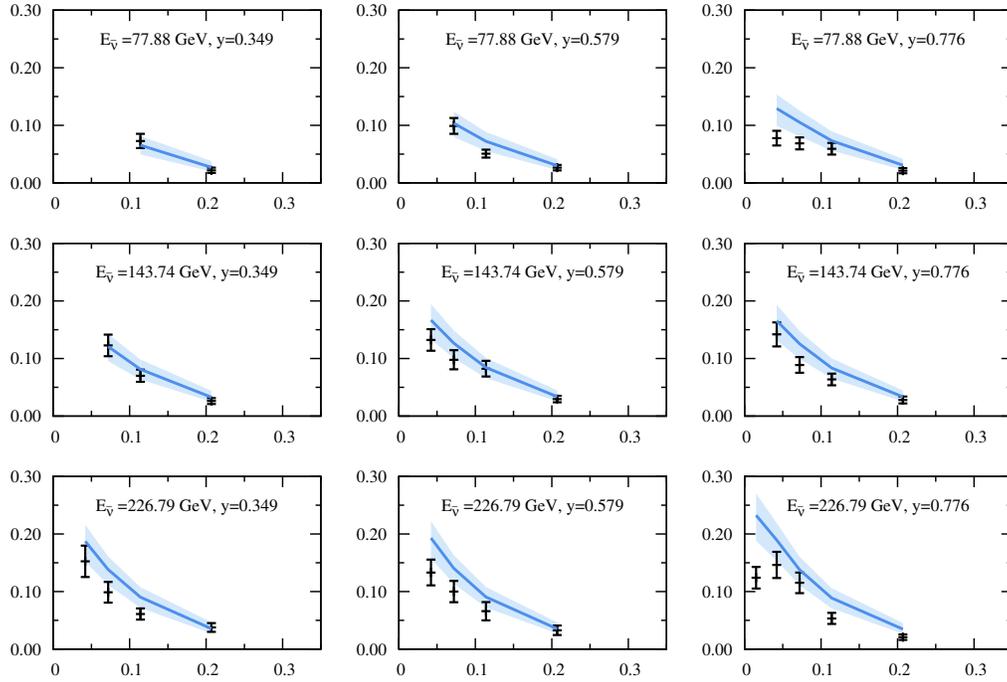

FIG. 28.   The same as Fig. 27, for the NuTeV measurements of dimuon production in antineutrino-iron collisions.

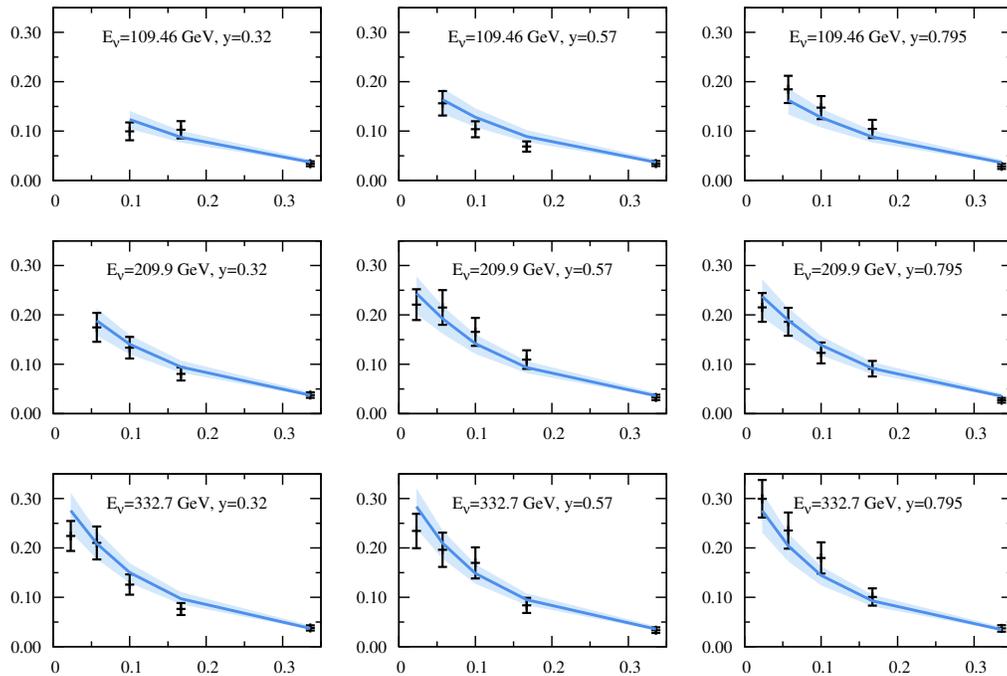CCFR Neutrino, $d^2\sigma(\nu_\mu N \rightarrow \mu^+\mu^- X)/dxdy$

FIG. 29.   The same as Fig. 27, for the CCFR measurements of dimuon production in neutrino-iron collisions.

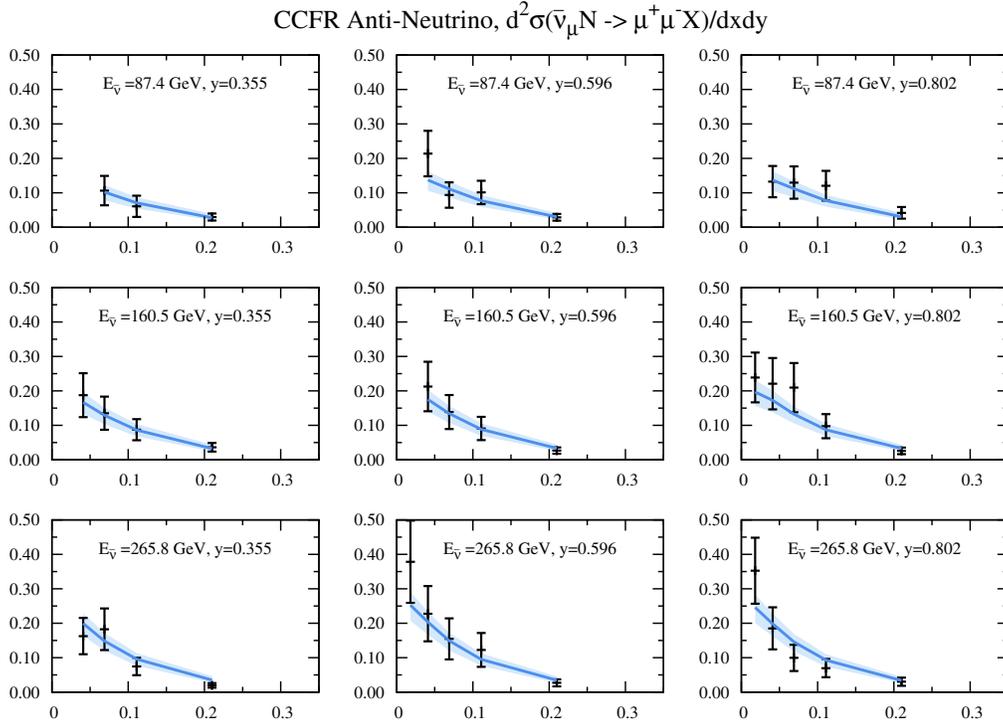CCFR Anti-Neutrino, $d^2\sigma(\bar{\nu}_\mu N \to \mu^+\mu^- X)/dxdy$

FIG. 30.   The same as Fig. 27, for the CCFR measurements of dimuon production in antineutrino-iron collisions.

$$R^s \equiv \frac{(s + \bar{s})}{\bar{u} + \bar{d}} \quad \text{at } x = 0.023, \qquad Q = 1.4 \text{ GeV} \quad (5)$$

and

$$\kappa^s(Q^2) = \frac{\int_0^1 x[s(x, Q^2) + \bar{s}(x, Q^2)]dx}{\int_0^1 x[\bar{u}(x, Q^2) + \bar{d}(x, Q^2)]dx}, \qquad (6)$$

by using inclusive DIS, the charge asymmetry of decay muons from $W^\pm$ production [44], and W + charm production differential cross sections [113] at 7 TeV. They obtain

$$R^s = 0.65^{+0.19}_{-0.17},$$
$$\kappa^s(Q^2 = 20 \text{ GeV}^2) = 0.52^{+0.18}_{-0.15} \quad \text{CMS(2014).} \quad (7)$$

Notice that ATLAS and CMS use two different definitions, r^s and R^s, for the strangeness fraction, which are supposed to coincide at the initial scale $Q_0 = 1.4$ GeV, if $\bar{u}(x, Q_0) = \bar{d}(x, Q_0)$.[2]

For comparison, at the factorization scale $Q = 1.4$ GeV and $x = 0.0234$, the CT14 and CT10 predictions are

---

[2]Both ATLAS and CMS studies are performed in the HERAFitter framework [114] and assume SU(2)-symmetric sea quark PDF parametrizations at the initial scale $Q_0 = 1.4$ GeV.

$$r^s_{\text{CT14 NNLO}} = \frac{\bar{s}(x, Q)}{\bar{d}(x, Q)} = 0.53 \pm 0.20,$$
$$r^s_{\text{CT10 NNLO}} = \frac{\bar{s}(x, Q)}{\bar{d}(x, Q)} = 0.76 \pm 0.17. \qquad (8)$$

Both CT14 and CT10 indicate a smaller strangeness than the ATLAS result and are compatible with CMS; the r^s ratio is smaller for CT14 than for CT10.

The NOMAD Collaboration has also completed a study of the strange quark PDF, relying on $\nu + Fe \to \mu^+ + \mu^- + X$ measurements [115] at lower energies than NuTeV and CCFR. They find that the strangeness suppression factor is

$$\kappa^s(20 \text{ GeV}^2) = 0.591 \pm 0.019, \qquad (9)$$

also yielding a smaller strangeness density than the ATLAS result. In another recent study by Alekhin and collaborators [116], the strange quark distribution and the ratios r^s and κ^s were determined in a QCD analysis including the NuTeV, CCFR, NOMAD and CHORUS measurements. The study uses the fixed-flavor-number scheme for the heavy-flavor treatment. Their main result is $\kappa^s(20 \text{ GeV}^2) = 0.654 \pm 0.030$. The CT14 and CT10 predictions for this quantity are

$$\kappa^s_{\text{CT14 NNLO}} = 0.62 \pm 0.14,$$
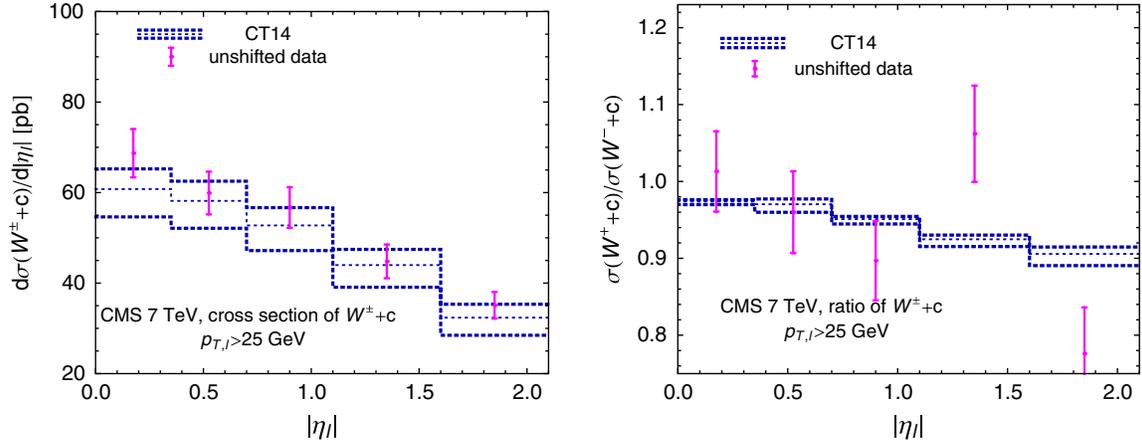$$\kappa^s_{\text{CT10 NNLO}} = 0.73 \pm 0.11. \qquad (10)$$

FIG. 31. Comparison of the CT14 prediction to $W^{\pm} + c$ differential cross sections (left) and to the ratio of $W^{+} + \bar{c}$ to $W^{-} + c$ cross sections (right) from the CMS measurement at 7 TeV.

The CT14 calculation is consistent with the NOMAD central value. However, the CT14 PDF uncertainty is considerably larger than the uncertainty quoted in the NOMAD paper, partly because of a different convention for the PDF uncertainty.

F. The CMS $W + c$ production measurement

Another experimental measurement that has direct access to the strange quark distribution is the associated production of W boson and charm quark at the LHC. Such a measurement was reported by the CMS Collaboration for $\sqrt{s} = 7$ TeV and a total integrated luminosity of 5 fb^{-1} [113]. Cross sections and ratios of cross sections with the observed W^{+} and W^{-} bosons were measured differentially with respect to the absolute value of the pseudorapidity of

the charged lepton from the W boson decay. As the theoretical cross section is not yet known at NNLO, the data were not directly included in the global fit but are compared here to NLO calculations based on MCFM 6.0 [117], assuming a nonzero charm quark mass and excluding contributions from gluon splitting into a $c\bar{c}$ pair. The renormalization and factorization scales are set to the virtuality of the W boson. The transverse momentum of the charged lepton is required to be at least 25 GeV. The theoretical calculation applies the same kinematical cuts as in the experimental analysis, but at the parton level.

The left panel of Fig. 31 shows the pseudorapidity distribution of the decay charged lepton from W boson decay in $W^{\pm} + c$ production at 7 TeV. The format of the figures is the same as in the previous comparisons. The total
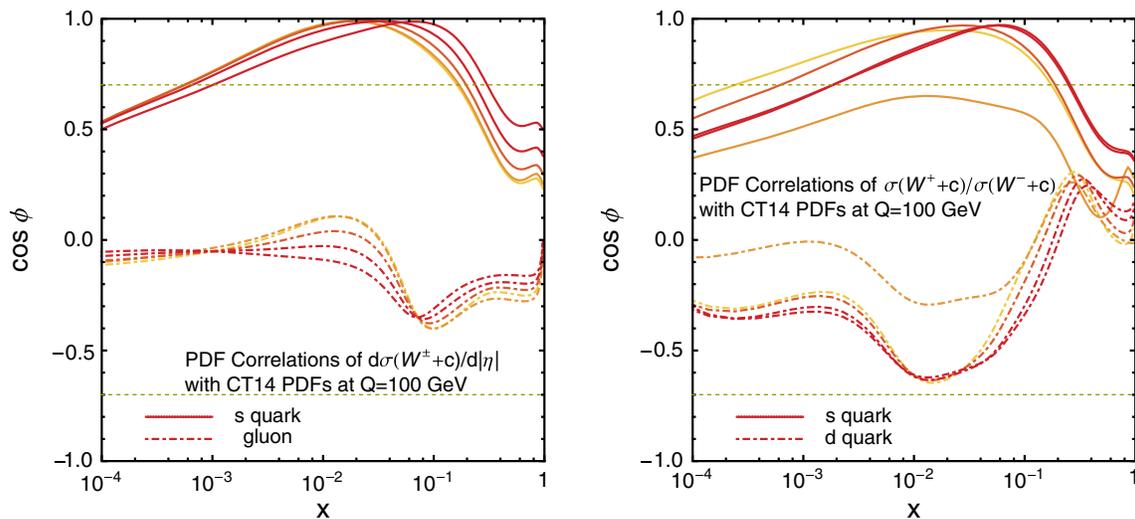

FIG. 32. Correlation cosines between the PDFs of select flavors, the $W^{\pm} + c$ cross section, and the W^{+}/W^{-} cross section ratio, as a function of x in the PDF.
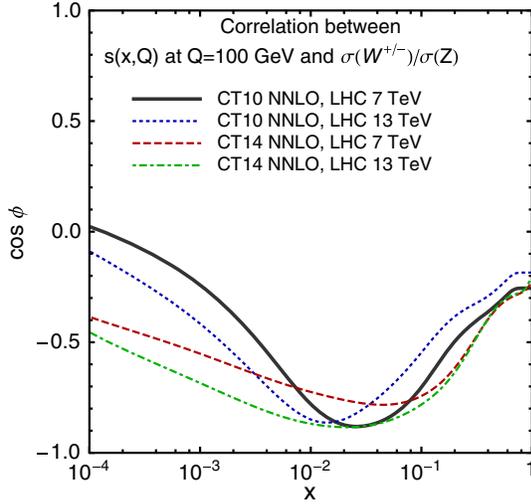
FIG. 33.   Correlation cosines between the ratio of total cross sections for $W^\pm$ production and Z boson production, and s quark PDF from CT14 and CT10 NNLO sets, for LHC 7 and 13 TeV.

experimental errors in the figures are reasonably close to the 68% C.L. PDF uncertainties. With further experimental and theoretical improvements, the process may contribute to the reduction of the PDF uncertainty.

The right panel shows the *ratio* of charged-lepton rapidity distributions in $W^+ + \bar{c}$ and $W^- + c$ production, which provides a handle on the strangeness asymmetry $s - \bar{s}$. The CT14 parametrization allows for no intrinsic s asymmetry at the initial scale Q_0. [At higher scales, a tiny asymmetry is generated by three-loop Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution.] Our prediction reproduces the average trend of the data; however, the experimental errors are larger than the PDF uncertainties.

The specific x ranges that are probed by CMS $W^\pm + c$ cross sections can be identified by plotting correlation cosines [13] between the PDFs of various flavors, the $W^\pm + c$ cross section, or cross section ratios. Figure 32 shows such correlation cosines for the s quark, gluon, and d quark PDFs at the factorization scale of 100 GeV. Lines in darker colors correspond to bins with larger rapidities. In the case of the differential cross section, the PDF correlations are most significant for the strange quark distributions at $x = 0.01–0.1$, as indicated by their strong correlations with $\cos\phi \sim 1$. The gluon does not play a significant role, due to its relatively smaller uncertainty in the same x region. In the case of the cross section ratio, the correlation with the strangeness is still dominant. But also, at large rapidity, the d quark contribution to the W^- cross sections is mildly anticorrelated, indicating that the ratio has marginal sensitivity to $d(x, Q)$ at x around 0.01.

Another well-known probe of the strangeness content of the proton is provided by the ratio of total cross sections for LHC $W^\pm$ and Z boson production [13]. The correlation cosine between $\sigma(W^\pm)/\sigma(Z)$ and $s(x, Q)$ can be viewed in

Fig. 33. As expected, we observe strong anticorrelation in a certain x range at all LHC center-of-mass energies. Compared to CT10, the x region of the strongest sensitivity shifts to higher x, and the x dependence gets flatter in CT14.

V. IMPACT ON HIGGS BOSON AND $t\bar{t}$ CROSS SECTIONS AT THE LHC

Gluon fusion provides the largest cross section for production of a Higgs boson. It was the most important process for the discovery of the Higgs boson in 2012, and it continues to be essential for detailed studies of Higgs boson properties. A great deal of benchmarking of hard cross sections and PDFs for the gg initial state was carried out both before and after the discovery [19,80,81,118–120]. This was motivated in part by the fact that the PDF uncertainty for the gg initial state was comparable to the renormalization and factorization scale uncertainties in the theoretical cross section at NNLO for producing a Higgs boson through gluon fusion. The recent calculation of the gluon fusion process at NNNLO [4] has reduced the scale uncertainty in the hard cross section still further, making the PDF uncertainty even more critical.

Similarly, production of a $t\bar{t}$ final state is crucial to many analyses at the LHC, as both a Standard Model signal and as a background to new physics. By far the dominant subprocess for $t\bar{t}$ production at the LHC is $gg \to t\bar{t}$, making $t\bar{t}$ production an important benchmark for understanding the gg PDF luminosity [13], especially with the current calculation of the $t\bar{t}$ total inclusive cross section now available at NNLO [95,96].

Using CT10 PDFs, we have recently performed detailed analyses of the predictions for $gg \to H$ and $t\bar{t}$ cross sections, as well as their uncertainties from both the PDFs and the strong coupling α_s [121,122]. In this section we update these studies and review CT14 predictions for $gg \to H$ and $t\bar{t}$ total and differential cross sections.

A. Higgs boson from gluon fusion at the LHC

We begin with an analysis of the PDF and α_s uncertainties for $gg \to H^0$. For this, we have utilized the NNLO code iHix s 1.3 [97], choosing the Higgs boson mass to be $M_H = 125$ GeV, and with both the renormalization and factorization scales fixed at $\mu = M_H$. Here, we have included the finite top quark mass correction (about 7%) to the fixed-order NNLO result obtained using the HQET (with infinite top-quark mass approximation).

To calculate the 90% C.L. PDF and α_s uncertainties of an arbitrary cross section X according to the most conventional (Hessian) method [15], we provide error PDFs (56 in the case of CT14) to probe independent combinations of the PDF parameters for the central $\alpha_s(M_Z) = 0.118$, plus two additional PDFs obtained from the best fits with $\alpha_s(M_Z) = 0.116$ and 0.120. Using the error sets, the combined PDF + α_s uncertainty on X is estimated by
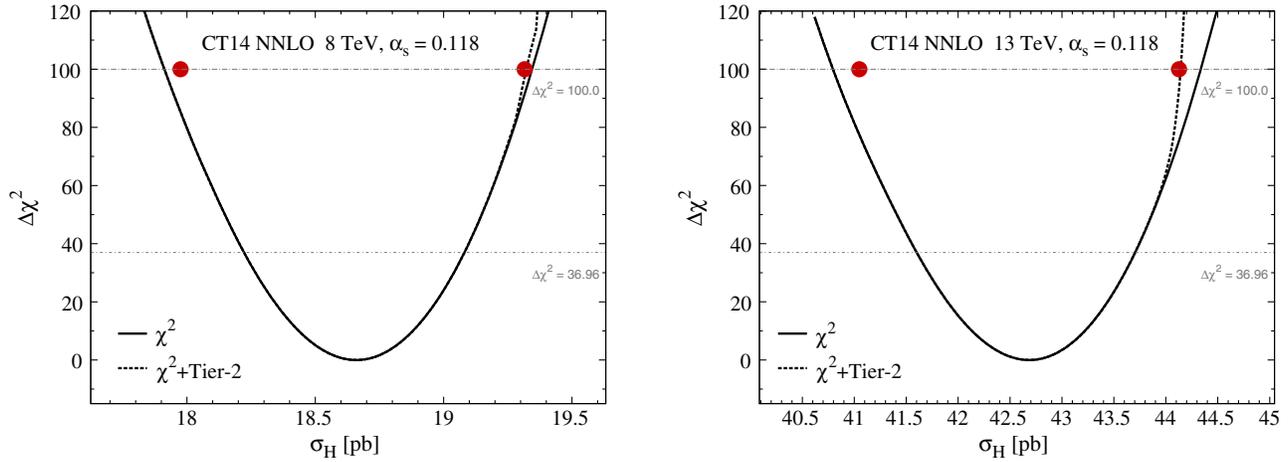
FIG. 34.   Dependence of the increase in χ^2 in the constrained CT14 fit on the expected cross section σ_H at the LHC 8 and 13 TeV, for $\alpha_s(M_Z) = 0.118$. The solid and dashed curves are for the constrained fits without and with the tier-2 penalties, respectively. The red dots correspond to the upper and lower 90% C.L. limits calculated by the Hessian method.

adding the PDF and α_s uncertainties in quadrature [123]. The quadrature-based combination is exact if χ^2 has a quadratic dependence, and X has a linear dependence on the PDF fitting parameters in the vicinity of the best fit. To account for some mild nonlinearities, asymmetric errors are allowed in the positive and negative directions of each eigenvector in the fitting parameter space.

Another method for estimating the PDF and α_s uncertainties on X introduces Lagrange multipliers (LMs) [17]. It does not rely on any assumptions about the functional dependence of X on the PDF parameters. Instead, the PDFs are refitted a number of times, while fixing X to take some user-selected value in each fit. Then the uncertainty in X can be estimated by looking at how χ^2 in the series of fits varies depending on the input value of X. The downside of the LM method is that it requires to repeat the PDF fit many times in order to calculate

the uncertainty of each given observable. It is clearly impractical for general-purpose experimental analyses; however, it can be straightforwardly performed for a few selected observables. As a side benefit, the LM method also provides an easy way to see which experimental data sets in the PDF global analysis have the most impact on the PDF dependence of X. Thus, in this section we will perform both the LM and Hessian analyses of the uncertainties for the Higgs boson and $t\bar{t}$ cross sections at the LHC.

We first do these calculations while keeping the strong coupling fixed at its central value of $\alpha_s(M_Z) = 0.118$ recommended by the PDF4LHC group. The uncertainties obtained this way are purely due to the PDFs. The results of the LM analysis are illustrated by Fig. 34, where we plot the change $\Delta\chi^2$ in χ^2 as a function of the tentative cross section σ_H for Higgs boson production via gluon fusion in pp
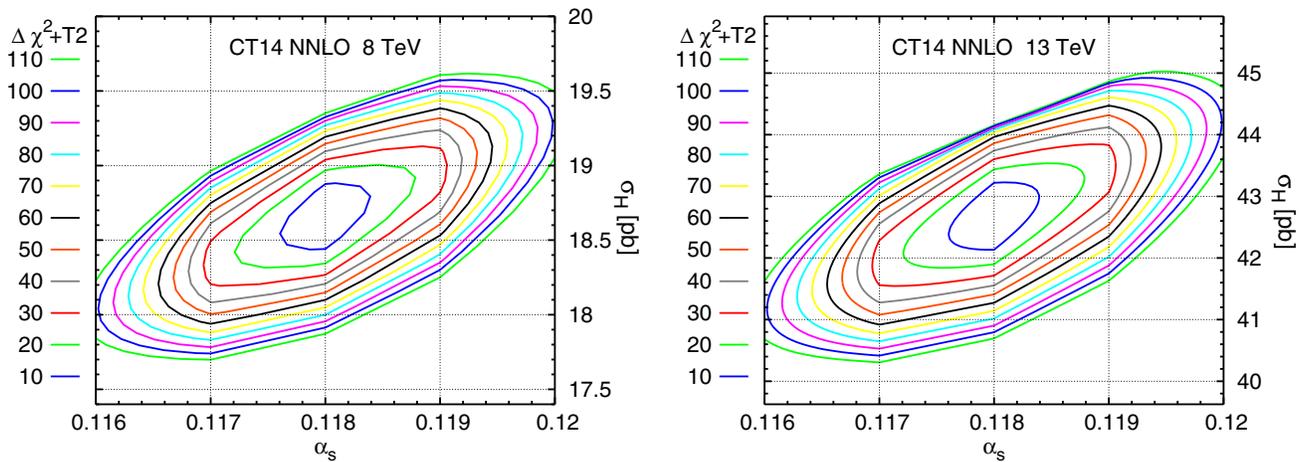

FIG. 35.   Contour plots of $\Delta\chi^2(\alpha_s(M_Z), \sigma_H)$ plus tier-2 penalty in the $(\alpha_s(M_Z), \sigma_H)$ plane, for σ_H at the LHC 8 and 13 TeV.

TABLE III. Uncertainties of $\sigma_H(gg \to H)$ computed by the Hessian and LM methods, with tier-2 penalty included. The 68% C.L. errors are given as percentages of the central values. The PDF-only uncertainties are for $\alpha_s(M_Z) = 0.118$.

$gg \to H$ (pb), PDF unc., $\alpha_s = 0.118$	8 TeV	13 TeV
68% C.L. (Hessian)	$18.7 + 2.1\% - 2.3\%$	$42.7 + 2.0\% - 2.4\%$
68% C.L. (LM)	$+2.3\% - 2.3\%$	$+2.4\% - 2.5\%$
$gg \to H$ (pb), PDF $+ \alpha_s$ unc.	8 TeV	13 TeV
68% C.L. (Hessian)	$18.7 + 2.9\% - 3.0\%$	$42.7 + 3.0\% - 3.2\%$
68% C.L. (LM)	$+3.0\% - 2.9\%$	$+3.2\% - 3.1\%$

collisions at energies $\sqrt{s} = 8$ and 13 TeV. $\Delta\chi^2 = 0$ corresponds to the best-fit PDFs to the CT14 experimental data set, so that the minimum of the approximately

parabolic curves is at our best-fit prediction for σ_H. Nonzero $\Delta\chi^2$ are obtained with an extra constraint that enforces σ_H to take the values on the horizontal axis that deviate from the best-fit ones. We have plotted the changes of both the simple χ^2 (solid line) and the $\chi^2 +$ tier-2 penalty (dashed line), in order to see the effects of requiring that no particular data set is too badly fit in the global analysis. (As defined in the Appendix of Ref. [22] and in [5], the tier-2 penalty makes use of the variable S_n, which gives a measure of the goodness of fit for each individual data set. A large S_n means that the experiment is not consistent with the theory.) We see that the two curves are almost identical over much of the range plotted, only beginning to diverge when σ_H is far from the best-fit value, and one or more experimental data sets can no longer be satisfactorily fit.

We can estimate asymmetric errors $(\delta\sigma_H)_\pm$ at the 90% C.L. by allowing a tolerance $\Delta\chi^2 = T^2$, with T of
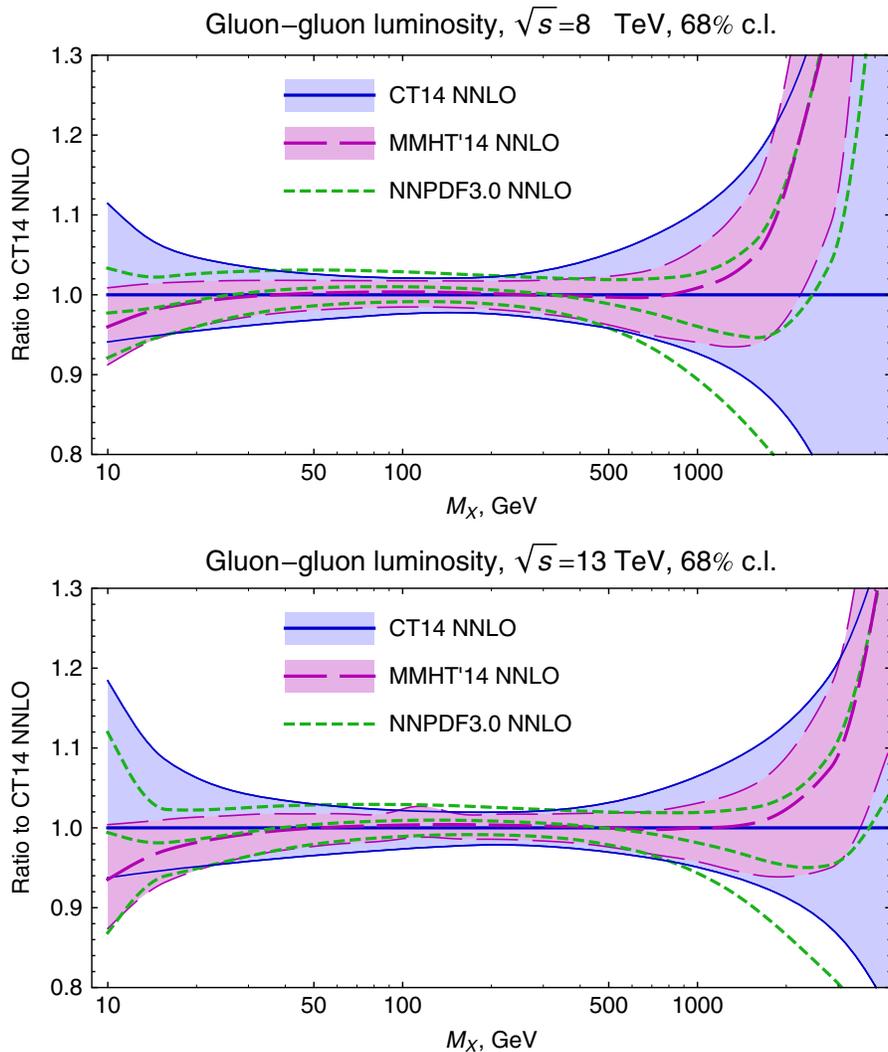

FIG. 36. The gg PDF luminosities for CT14, MMHT2014 and NNPDF3.0 PDFs at the LHC with $\sqrt{s} = 8$ and 13 TeV, with $\alpha_s(M_Z) = 0.118$.

TABLE IV. Higgs boson production cross sections (in pico-barns) for the gluon fusion channel at the LHC, at 8 and 13 TeV center-of-mass energies, obtained using the CT14, MMHT2014, NNPDF3.0, and CT10 PDFs, with a common value of $\alpha_s(M_Z) = 0.118$. The errors given are due to the PDFs at the 68% C.L.

	CT14	MMHT2014	NNPDF3.0	CT10
8 TeV	$18.66^{+2.1\%}_{-2.3\%}$	$18.65^{+1.4\%}_{-1.9\%}$	$18.77^{+1.8\%}_{-1.8\%}$	$18.37^{+1.7\%}_{-2.1\%}$
13 TeV	$42.68^{+2.0\%}_{-2.4\%}$	$42.70^{+1.3\%}_{-1.8\%}$	$42.97^{+1.9\%}_{-1.9\%}$	$42.20^{+1.9\%}_{-2.5\%}$

about 10. Given the nearly parabolic nature of these plots, we see that the 68% C.L. errors can be consistently defined using a range corresponding to $\Delta\chi^2 = (T/1.645)^2$. The 90% C.L. and 68% C.L. tolerance values are indicated by the upper and lower horizontal lines, respectively, in each of the plots. Finally, the red dots are the upper and lower 90% C.L. limits from the Hessian method analysis. They agree quite well with the LM analysis using the χ^2 + tier-2 penalty at both 8 and 13 TeV. The effect of the tier-2 penalty is modest; the deviations from the parabolic behavior are small.

We next perform a LM scan by allowing both the σ_H cross section and $\alpha_s(M_Z)$ to vary as "fitting parameters" and by including the world-average constraints on $\alpha_s(M_Z)$ directly into the χ^2 function. (Details can be obtained in Ref. [122].) We examine χ^2 as a function of $(\alpha_s(M_Z), \sigma_H)$ and trace out contours of constant χ^2 + tier-2 penalty in the (α_s, σ_H) plane in Fig. 35, for $\sqrt{s} = 8$ and 13 TeV.

A contour here is the locus of points in the (α_s, σ_H) plane along which the constrained value of χ^2 + tier-2 is constant. We see from Fig. 35 that the values of σ_H and $\alpha_s(M_Z)$ are strongly correlated, as expected, since the gg fusion cross section is proportional to $\alpha_s(M_Z)^2$. Larger values of $\alpha_s(M_Z)$ correspond to larger values of σ_H for the same goodness of fit to the global data, even though there is a partially compensating decrease of the gg luminosity. The effect of the tier-2 penalty is very small, being most noticeable for values of α_s

around its global average of 0.118, which results in a squeezing of the ellipses in that region.

Table III recapitulates the results from Figs. 34 and 35 by listing the central values of σ_H, the PDF uncertainties, and combined PDF + α_s uncertainties as obtained by the Hessian and LM methods. Here, the PDF + α_s uncertainty at 68% C.L. is obtained from the result at 90% C.L. by a scaling factor of 1/1.645.

The gg PDF luminosities for CT14, MMHT2014 [124] and NNPDF3.0 [79] PDFs at 13 TeV are shown in Fig. 36. The parton luminosity is defined as in Ref. [125]. All central values and uncertainty bands agree very well among the three global PDFs, in the x range sensitive to Higgs production. In Table IV, we compare the predictions for σ_H from CT14 with those from MMHT2014, NNPDF3.0, and CT10. Compared to CT10, predicted σ_H values for CT14 NNLO have increased by 1%–1.5%. Along with the changes also present in the updated PDFs from the two other PDF groups, the modest increase in the CT14 gluon brings σ_H from the global PDF groups into a remarkably good agreement. The projected spread due to the latest NNLO PDFs in the total cross section σ_H at 13 TeV will be about the same in magnitude as the scale uncertainty in its NNNLO prediction.

Besides providing an estimate of the PDF uncertainty, the LM analysis allows us to identify the experimental data sets that are most sensitive to variations of σ_H. In the LM scan of σ_H, we monitor the changes of the equivalent Gaussian variable S_n for each included experimental data set. In the plots of S_n values vs σ_H, of the type presented in Fig. 37, we select the experiments whose S_n (closely related to χ^2_n) depends strongly on σ_H. Such experiments typically impose the tightest constraints on σ_H, when their S_n quickly grows with σ_H.

We see that, although the CMS 7 TeV inclusive jet data (538) is relatively poorly fit by CT14 NNLO, it is also not very sensitive to the expected Higgs cross section. The data sets most relevant to the Higgs cross section are the HERA
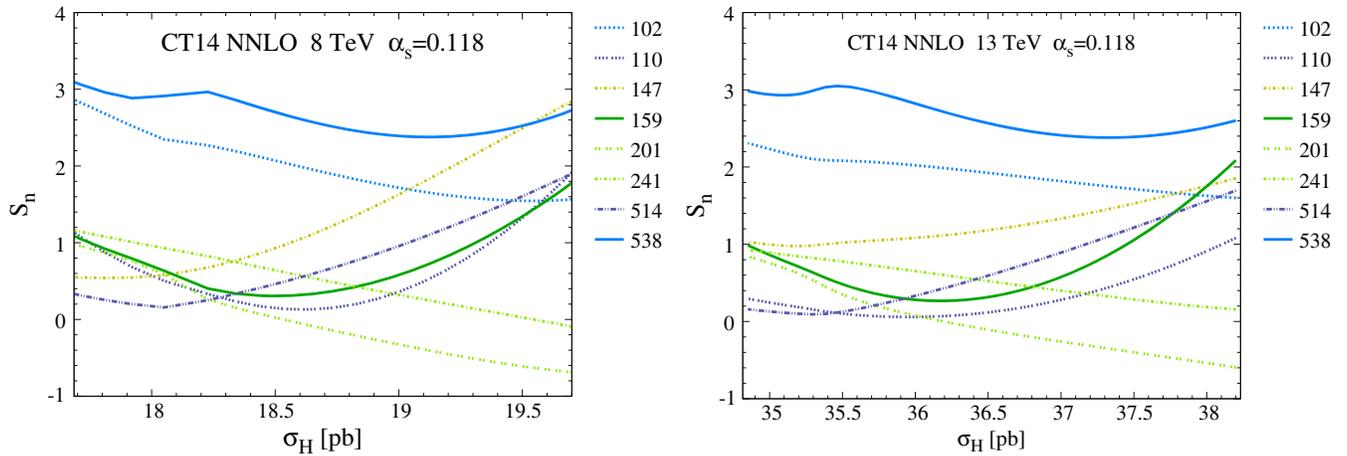

FIG. 37. The equivalent Gaussian variable S_n versus σ_H at the LHC with $\sqrt{s} = 8$ and 13 TeV.

TABLE V.   CT14 NNLO total inclusive cross sections for top-quark pair production at LHC center-of-mass energies of 7, 8, and 13 TeV.

$pp \to t\bar{t}$ (pb), PDF unc., $\alpha_s = 0.118$	7 TeV	8 TeV	13 TeV
68% C.L. (Hessian)	$177 + 4.4\% - 3.7\%$	$253 + 3.9\% - 3.5\%$	$823 + 2.6\% - 2.7\%$
68% C.L. (LM)		$+4.8\% - 4.6\%$	$+2.9\% - 2.9\%$
$pp \to t\bar{t}$ (pb), PDF + α_s	7 TeV	8 TeV	13 TeV
68% C.L. (Hessian)	$+5.5\% - 4.6\%$	$+5.2\% - 4.4\%$	$+3.6\% - 3.5\%$
68% C.L. (LM)		$+5.1\% - 4.7\%$	$+3.6\% - 3.5\%$

TABLE VI.   Measurements of total inclusive cross sections for top-quark pair production at LHC center-of-mass energies of 7, 8, and 13 TeV, for an assumed top-quark mass of 172.5 GeV.

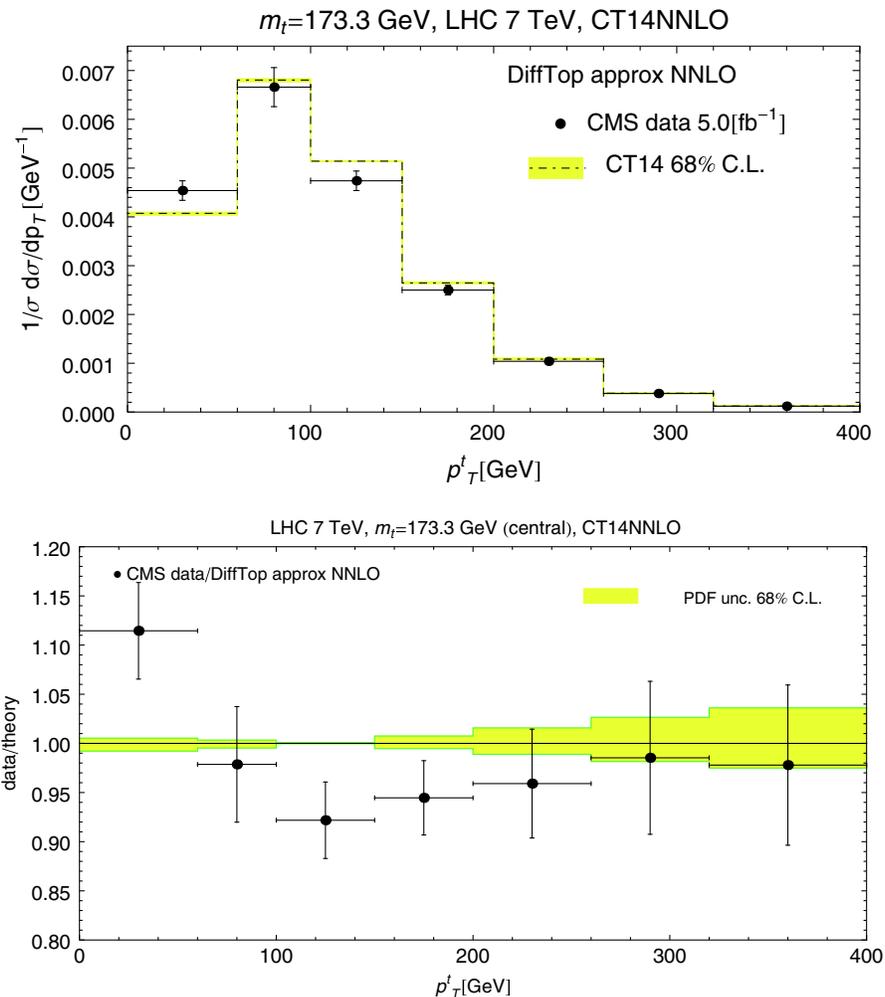| $\sigma_{t\bar{t}}^{\mathrm{exp}}$ (pb)	7 TeV (dilepton channel)	8 TeV (lep + jets)
ATLAS [126,127]	$177 \pm 20^{(\mathrm{stat})} \pm 14^{(\mathrm{syst})} \pm 7^{(\mathrm{lumi})}$	$260 \pm 1^{(\mathrm{stat})}{}_{-23}^{+22}{}^{(\mathrm{syst})} \pm 8^{(\mathrm{lumi})} \pm 4^{(\mathrm{beam})}$
CMS [128,129]	$161.9 \pm 2.5^{(\mathrm{stat})}{}_{-5.0}^{+5.1}{}^{(\mathrm{syst})} \pm 3.6^{(\mathrm{lumi})}$	$228.4 \pm 9.0^{(\mathrm{stat})}{}_{-26}^{+29}{}^{(\mathrm{syst})} \pm 10^{(\mathrm{lumi})}$
	7 TeV (lepton + jets, dilepton, all-jets)	8 TeV (dilepton channel)
ATLAS and CMS Combined [130,131]	$173.3 \pm 2.3^{(\mathrm{stat})} \pm 7.6^{(\mathrm{syst})} \pm 6.3^{(\mathrm{lumi})}$	$241.5 \pm 1.4^{(\mathrm{stat})} + 5.7^{(\mathrm{syst})} \pm 6.2^{(\mathrm{lumi})}$

FIG. 38.   Normalized final-state top-quark p_T differential distribution at CMS 7 TeV.
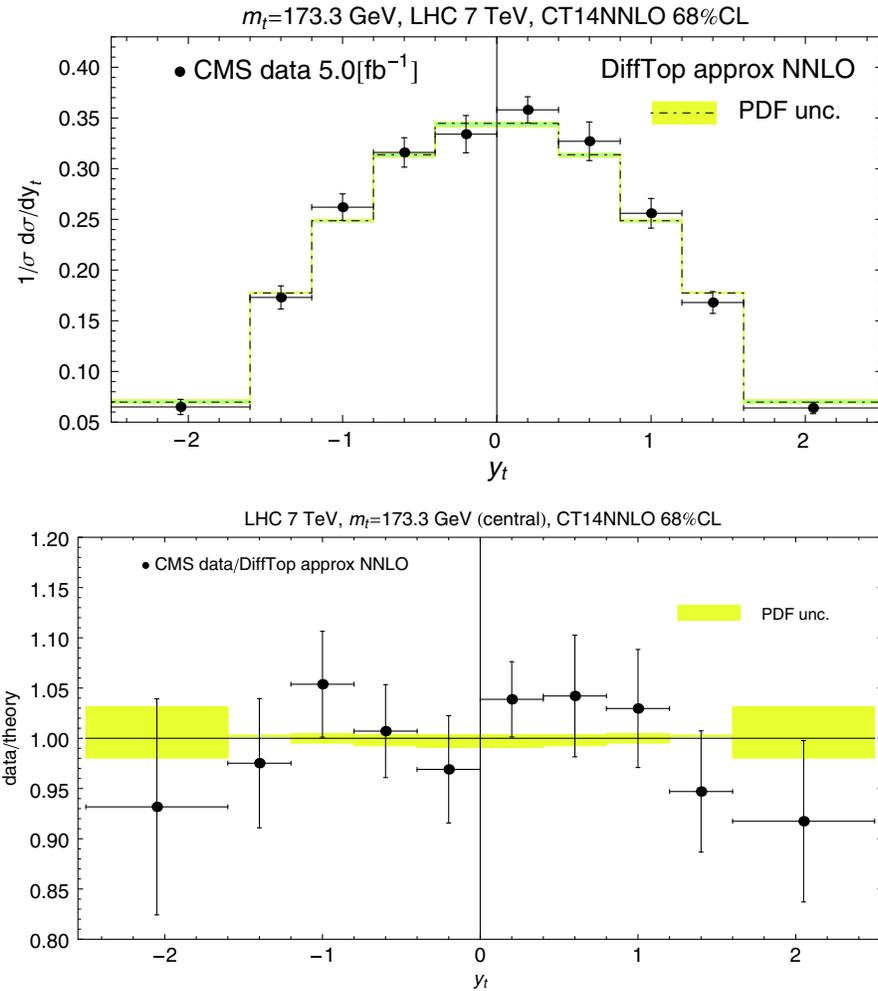
FIG. 39. Normalized final-state top-quark rapidity distribution at CMS 7 TeV.

inclusive data set (159) at both larger and smaller values of σ_H, as well as combined charm production cross sections from HERA (147); D0 run-2 inclusive jet (514); and CCFR F_2^p (110) at larger σ_H. At small σ_H, the most sensitive data set is BCDMS F_2^d (102), with some sensitivity also from E605 Drell-Yan (201) and LHCb 7 TeV charge asymmetry (241). Sensitivity of σ_H to CCFR dimuon data observed with CT10 [122] is no longer present.

B. $t\bar{t}$ production cross section at the LHC

Next, we consider theoretical predictions and their uncertainties for the total inclusive cross section for $t\bar{t}$ production at the LHC, and also present some differential cross sections.

In the $t\bar{t}$ case, the comparison between the Hessian and Lagrange multiplier methods for finding uncertainties is very similar to that found for the Higgs cross section. Therefore, we just present our final estimates for the total inclusive cross section from the Top++ code [96], given in Table V. Recent experimental measurements of the total

inclusive cross section for top-quark pair production at the LHC are given in Table VI, together with ATLAS and CMS combined determinations at $\sqrt{s} = 7$ and 8 TeV.

For comparison, predictions and PDF-only errors using CT10NNLO PDFs give $\sigma_{t\bar{t}} = 246^{+4.1\%}_{-3.4\%}$ pb at 8 TeV and $\sigma_{t\bar{t}} = 806^{+2.5\%}_{-2.2\%}$ pb at 13 TeV at 68% C.L. Here we find that the Hessian and the LM methods are in very good agreement in CT14 at $\sqrt{s} = 13$ TeV and agree slightly worse at $\sqrt{s} = 8$ TeV. Measurements of $t\bar{t}$ pair production can potentially constrain the gluon PDF at large x, if correlations between the gluon, $\alpha_s(M_Z)$ and the top-quark mass are accounted for. Given the current experimental precision of $t\bar{t}$ measurements, the impact of such data in a global PDF fit is expected to be moderate; related exploratory studies can be found in Refs. [132,133].

In Figs. 38–40, the normalized top-quark transverse momentum p_T and rapidity y distributions at approximate NNLO $(\mathcal{O}(\alpha_s^4))$ are compared to the CMS [67] and ATLAS [65] measurements, at a center-of-mass energy $\sqrt{s} = 7$ TeV. The yellow bands represent the CT14 PDF uncertainty evaluated at the 68% C.L. with the program
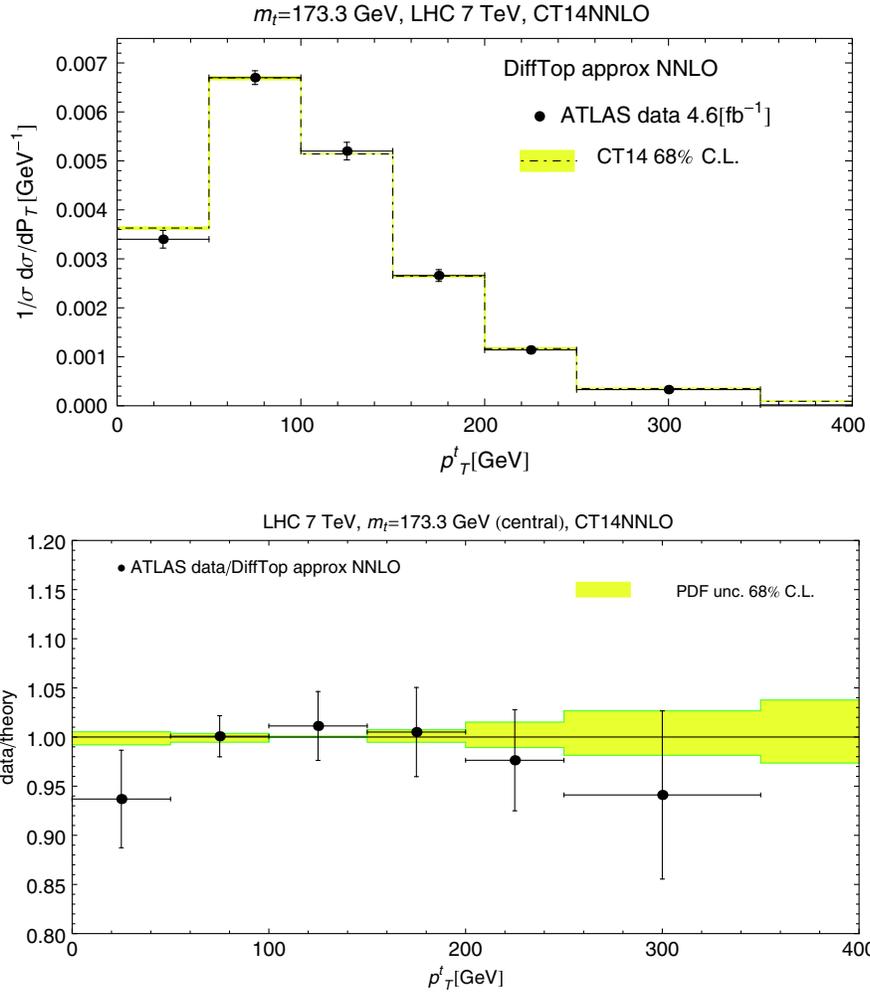
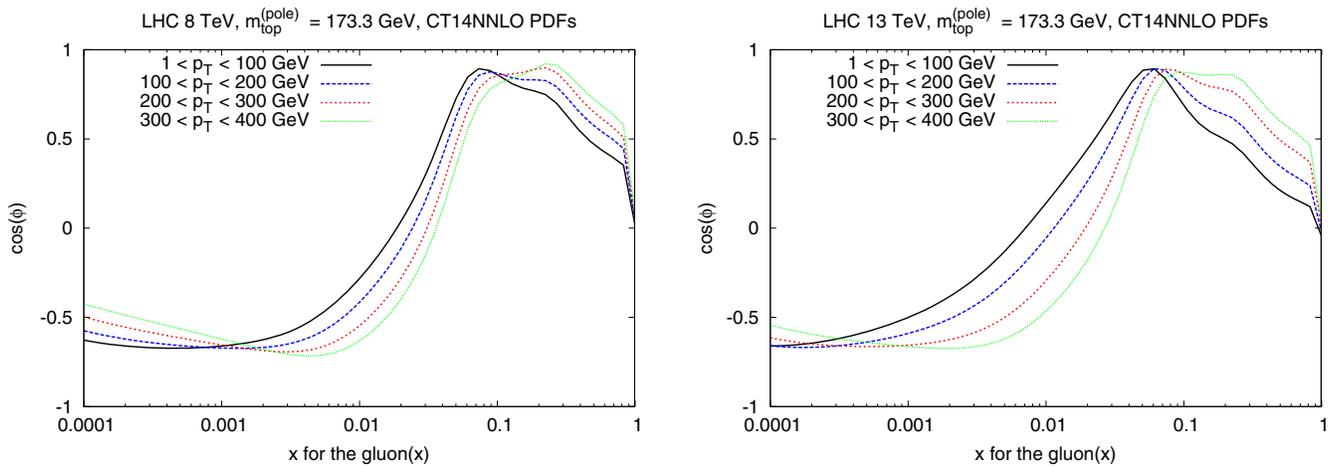FIG. 40.   Normalized final-state top-quark p_T differential distribution at ATLAS 7 TeV.

FIG. 41.   The correlation cosine as a function of x gluon for the top-quark p_T distribution in $t\bar{t}$ production at the LHC at $\sqrt{s} = 8$ and 13 TeV.
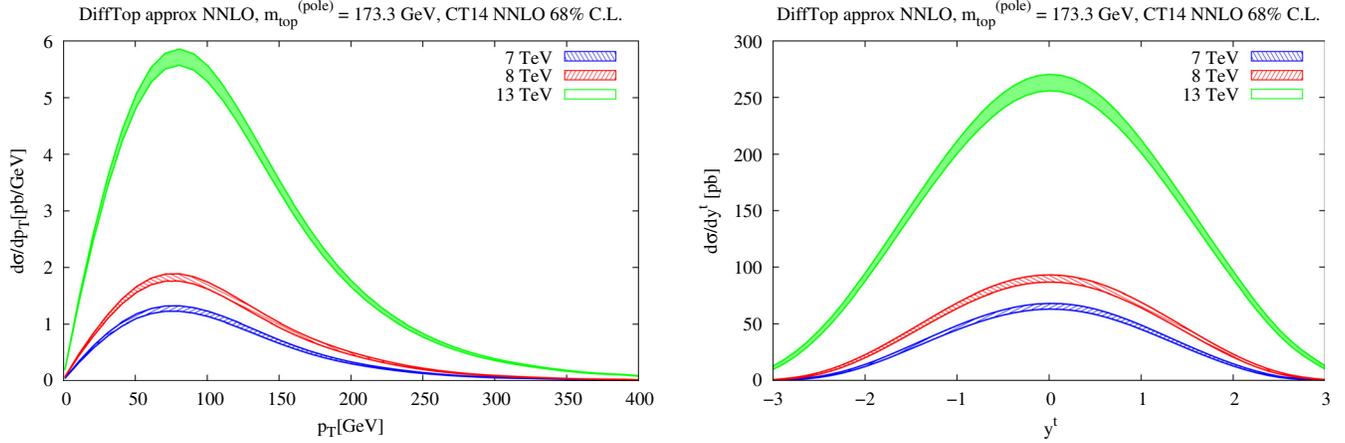
FIG. 42.   Absolute differential p_T and y distributions for the final-state top quark in $t\bar{t}$ production at the LHC at $\sqrt{s} = 7$, 8, and 13 TeV.

DIFFTOP [133] based on QCD threshold expansions beyond the leading logarithmic approximation, for one-particle inclusive kinematics. The value of the top-quark mass here is $m_t = 173.3$ GeV in the "pole mass" definition. In Fig. 41 the correlation cosine between the differential top-quark p_T distribution and the momentum fraction x carried by the gluon is shown, in four different p_T bins at the LHC with $\sqrt{s} = 8$ and 13 TeV. The cosine correlation at $\sqrt{s} = 7$ TeV exhibits identical features to that of $\sqrt{s} = 8$ TeV. It is therefore omitted. A strong correlation between the p_T distribution and large-x gluon ($x \approx 0.1$) is observed for both LHC energies, although the cosines exhibit different patterns of x dependence. Finally, in Fig. 42 we present the absolute, rather than normalized, differential p_T and y distributions for top-quark production, together with the relative PDF uncertainties, at the LHC $\sqrt{s} = 7$, 8 and 13 TeV.

VI. DISCUSSION AND CONCLUSION

In this paper, we have presented CT14, the next generation of NNLO (as well as LO and NLO) parton distributions from a global analysis by the CTEQ-TEA group. With rapid improvements in LHC measurements, the focus of the global analysis has shifted toward providing accurate predictions in the wide range of x and Q covered by the LHC data. This development requires a long-term multiprong effort in theoretical, experimental, and statistical areas.

In the current study, we have added enhancements that open the door for long-term developments in CT14 methodology geared toward the goals of LHC physics. This is the first CT analysis that includes measurements of inclusive production of vector bosons [41,44–46] and jets [49,50] from the LHC at 7 and 8 TeV as input for the fits. We also include new data on charm production from DIS at HERA [32] and precise measurements of the electron charge asymmetry from D0 at 9.7 fb^{-1} [14]. These measurements allow us to probe new combinations of quark flavors that were not resolved by the previous data sets. As most of these measurements contain substantial correlated systematic uncertainties, we have implemented these correlated errors and have examined their impact on the PDFs.

On the theory side, we have introduced a more flexible parametrization to better capture variations in the PDF dependence. A series of benchmark tests of NNLO cross sections, carried out in the runup for the CT14 fit for all key fitted processes, has resulted in better agreement with most experiments and brought accuracy of most predictions to the truly NNLO level. We examined the PDF errors for the important LHC processes and have tested the consistency of the Hessian and Lagrange multiplier approaches. Compared to CT10, the new inputs and theoretical advancements resulted in a softer d/u ratio at large x, a lower strangeness PDF at $x > 0.01$, a slight increase in the large-x gluon (of order 1%), and wider uncertainty bands on d/u, $\bar{d}/\bar{u}$, and $q - \bar{q}$ combinations at x of order 0.001 (probed by LHC W/Z production). Despite these changes in central predictions, the CT14 NNLO PDFs remain consistent with CT10 NNLO within the respective error bands.

Some implications of CT14 predictions for phenomenological observables were reviewed in Secs. IV and V. Compared to calculations with CT10 NNLO, the $gg \rightarrow H$ total cross section has increased slightly in CT14: by 1.6% at the LHC 8 TeV and by 1.1% at 13 TeV. The $t\bar{t}$ production cross sections have also increased in CT14 by 2.7% at 8 TeV and by 1.4% at 13 TeV. The W and Z cross sections, while still consistent with CT10, have slightly changed as a result of reduced strangeness. Common ratios of strangeness and nonstrangeness PDFs for CT14 NNLO, shown in Eqs. (8) and (10), are consistent with the independent ATLAS, CMS, and NOMAD determinations within the PDF uncertainties.

The final CT14 PDFs are presented in the form of one central and 56 Hessian eigenvector sets at NLO and NNLO. The 90% C.L. PDF uncertainties for physical observables can be estimated from these sets using the symmetric [20] or asymmetric [5,54] master formulas by adding contributions from each pair of sets in quadrature. These PDFs are determined for the central QCD coupling of $\alpha_s(M_Z) = 0.118$, consistent with the world-average α_s value. For estimation of the combined PDF $+ \alpha_s$ uncertainty, we provide two additional best-fit sets for $\alpha_s(M_Z) = 0.116$ and 0.120. The 90% C.L. variation due to $\alpha_s(M_Z)$ can be estimated as half of the difference in predictions from the two α_s sets. The PDF $+ \alpha_s$ uncertainty, at 90% C.L., and including correlations, can also be determined by adding the PDF uncertainty and α_s uncertainty in quadrature [123].

At leading order, we provide two PDF sets, obtained assuming one-loop evolution of α_s and $\alpha_s(M_Z) = 0.130$; and two-loop evolution of α_s and $\alpha_s(M_Z) = 0.118$. Besides these general-purpose PDF sets, we provide a series of (N)NLO sets for $\alpha_s(M_Z) = 0.111$–0.123 and additional sets in heavy-quark schemes with up to three, four, and six active flavors. Phenomenological applications of the CT14 series and the special CT14 PDFs (such as allowing for nonperturbative intrinsic charm contribution) will be discussed in a followup study [134].

Parametrizations for the CT14 PDF sets are distributed in a standalone form via the CTEQ-TEA Web site [135] or as a part of the LHAPDF6 library [7]. For backward compatibility with version 5.9.X of LHAPDF, our Web site also provides CT14 grids in the LHAPDF5 format, as well as an update for the CTEQ-TEA module of the LHAPDF5 library, which must be included during compilation to support calls of all eigenvector sets included with CT14 [136].

APPENDIX: PARAMETRIZATIONS IN CT14

Parton distribution functions are measured by parameterizing their x dependence at a low scale Q_0. For each choice of parameters, the PDFs are computed at higher scales by DGLAP evolution; and the parameters at Q_0 are adjusted to optimize the fit to a wide variety of experimental data. Traditional parametrizations for each flavor are of the form

$$xf_a(x, Q_0) = x^{a_1}(1-x)^{a_2} P_a(x), \qquad (A1)$$

where the x^{a_1} behavior at $x \to 0$ is guided by Regge theory, and the $(1-x)^{a_2}$ behavior at $x \to 1$ is guided by spectator counting rules. The remaining factor $P_a(x)$ is assumed to be slowly varying, because there is no reason to expect fine structure in it even at scales below Q_0, and evolution from those scales up to Q_0 provides additional smoothing.

In the previous CTEQ analyses, $P_a(x)$ in Eq. (A1) for each flavor was chosen as an exponential of a polynomial in x or $\sqrt{x}$; e.g.,

$$P(x) = \exp(a_0 + a_3\sqrt{x} + a_4 x + a_5 x^2) \qquad (A2)$$

for $u_v(x)$ or $d_v(x)$ in CT10 [6]. The exponential form conveniently enforces the desired positive-definite behavior for the PDFs, and it suppresses nonleading behavior in the limit $x \to 0$ by a factor $\sqrt{x}$, which is similar to what would be expected from a secondary Regge trajectory. However, this parametrization has two undesirable features. First, because the exponential function can vary rapidly, the power laws x^{a_1} and $(1-x)^{a_2}$, which formally control the $x \to 0$ and $x \to 1$ limits, need not actually dominate in practical regions of small x (say $x \lesssim 0.001$) or large x (say $x \gtrsim 0.6$). Second, the qualitative similarity of $\exp(a_3\sqrt{x})$, $\exp(a_4 x)$, and $\exp(a_5 x^2)$ to each other causes the parameters a_3, a_4, a_5 to be strongly correlated with each other in the fit. This correlation may destabilize the χ^2 minimization and compromise the Hessian approach to uncertainty analysis, since that approach is based on a quadratic dependence of χ^2 on the fitting parameters, which is only guaranteed close to the minimum.

We introduce a better style of parametrization in CT14. We begin by replacing $P_a(x)$ by a polynomial in $\sqrt{x}$, which avoids the rapid variations invited by an exponential form. Low-order polynomials have been used previously by many other groups; however, polynomials with higher powers were less widespread. We add them now to provide more flexibility in the parametrization. In particular, for the best-constrained flavor combination $u_v(x) \equiv u(x) - \bar{u}(x)$ we use a fourth-order polynomial

$$P_{u_v} = c_0 + c_1 y + c_2 y^2 + c_3 y^3 + c_4 y^4, \qquad (A3)$$

where $y = \sqrt{x}$. But rather than using the coefficients c_i directly as fitting parameters, we reexpress the polynomial as a linear combination of *Bernstein polynomials*:

$$\begin{aligned} P_{u_v} = \; & d_0 p_0(y) + d_1 p_1(y) + d_2 p_2(y) \\ & + d_3 p_3(y) + d_4 p_4(y), \end{aligned} \qquad (A4)$$

where

$$p_0(y) = (1 - y)^4,$$
$$p_1(y) = 4y(1 - y)^3,$$
$$p_2(y) = 6y^2(1 - y)^2,$$
$$p_3(y) = 4y^3(1 - y),$$
$$p_4(y) = y^4. \tag{A5}$$

This reexpression does not change the functional form of P_{u_v}: it is still a completely general fourth-order polynomial in $y = \sqrt{x}$. But the new coefficients d_i are less correlated with each other than the old c_i, because each Bernstein polynomial is strongly peaked at a different value of y. [The flexibility of the parametrization can be increased by using higher-order polynomials; the generalization of Eq. (A5) to higher orders is obvious—the numerical factors are just binomial coefficients.]

In practice, we refine this procedure as follows. First, as a matter of convenience, we set $d_4 = 1$ and supply in its place an overall constant factor, which is determined by the number sum rule $\int_0^1 u_v(x)dx = 2$. We then set $d_3 = 1 + a_1/2$ to suppress deviations from the $(1 - x)^{a_2}$ behavior of $u_v(x)$ at large x by canceling the first subleading power of $(1 - x)$ in P_{u_v}:

$$xu_v(x) \rightarrow \text{const} \times (1 - x)^{a_2} \times [1 + \mathcal{O}((1 - x)^2)]$$
$$\text{for } x \rightarrow 1. \tag{A6}$$

We use the same parametrization for $d_v(x) \equiv d(x) - \overline{d}(x)$, with the same parameter values a_1 and a_2; but, of course, independent parameters for the coefficients of the Bernstein polynomials and the normalization, which is set by $\int_0^1 d_v(x)dx = 1$.

Tying the valence a_1 parameters together is motivated by Regge theory and supported by the observation that the value of a_1 obtained in the fit is not far from the value expected from Regge theory. (The a_1 values for $u, d, \overline{u}$, and $\overline{d}$ are expected to be close to 0 from the Pomeron trajectory; but that leading behavior is expected to cancel in $u_v = u - \overline{u}$ and $d_v = d - \overline{d}$, revealing the subleading vector meson Regge trajectory at $a_1 \simeq 0.5$.) Not counting the two normalization parameters that are constrained by quark-number sum rules, we are left with a total of eight fitting parameters for the valence quarks. This is the same number of parameters as were used in CT10 NNLO. As a consistency check, we find that allowing the a_1 and a_2 parameters for d_v to be independent of those for u_v would reduce $\chi^2 \approx 3380$ by less than one unit. Allowing those parameters to be free would also not substantially increase the uncertainty range given by the Hessian procedure, except at very large x, where the fractional uncertainty is already very large. The additional fractional uncertainty at small x generated by allowing different a_1 powers is also not important, because that uncertainty only appears in the

valence quantities $u(x) - \overline{u}(x)$ and $d(x) - \overline{d}(x)$; while most processes of interest are governed by the much larger $u(x)$, $d(x)$, $\overline{u}(x)$, $\overline{d}(x)$ themselves.

In addition to theoretical arguments that the power laws a_2 should be the same for u_v and d_v [137], χ^2 tends to be insensitive to the differences. A large portion of the data included in the global fit are from electron and muon DIS on protons, which is more sensitive to u and $\overline{u}$ than to d and $\overline{d}$ because of the squares of their electric charges. Hence, when similar parametrizations are used for P_{u_v} and P_{d_v}, the uncertainties of $a_1(d_v)$ and $a_2(d_v)$ are relatively large.

Our assumption $a_2(u_v) = a_2(d_v)$ forces $u_v(x)/d_v(x)$ to approach a constant in the limit $x \rightarrow 1$. It allows our phenomenological findings to be relevant for the extensive discussions of what that constant might be [94,104]. However, the experimental constraints at large x are fairly weak: we can find excellent fits over the range $-0.5 < a_2(d_v) - a_2(u_v) < 1.2$ at an increase of only five units in χ^2. Hence both $u_v(x)/d_v(x) \rightarrow 0$ and $u_v(x)/d_v(x) \rightarrow \infty$ at $x \rightarrow 1$ remain fully consistent with the data. However, our assumption $a_2(u_v) = a_2(d_v)$ does not restrict the calculated uncertainty range materially in regions where it is not already very large.

By way of comparison, if we use the CT10 NNLO [6] form (A2) for u_v and d_v, we obtain a slightly better fit (χ^2 lower by 8) with an unreasonable $a_2 \approx 0.1$. Similar behavior led us to fix $a_2 = 0.2$ in CT10 NNLO.

In a different comparison, the MSTW2008 fit [16,138] uses a parametrization for u_v and d_v that is equivalent to Eq. (A3) with $c_3 = c_4 = 0$, with the power-law parameters a_1 and a_2 allowed to differ between u_v and d_v. If we use this MSTW parametrization for the valence quarks at our $Q_0 = 1.3$ GeV, in place of the form we have chosen, the best-fit χ^2 increases by 64, even though the total number of fitting parameters is the same. This decline in the fit quality comes about because the freedom to have $a_2(u_v) \neq a_2(d_v)$ and $a_2(u_v) \neq a_2(d_v)$ is not actually very helpful, as noted above; so setting $c_3 = c_4 = 0$ does not leave an adequate number of free parameters.

The more recent MMHT2014 [124] PDF fit uses full fourth-order polynomials for u_v and d_v. In our fit, however, we find that no significant improvement in χ^2 would result from treating $d_3(u_v)$ and $d_3(d_v)$ as free parameters, rather than choosing them to cancel the first subleading behavior at $x \rightarrow 1$, as we have done.

Meanwhile the HERA PDF fits [33,139,140] use much more restricted forms, equivalent to $c_1 = c_2 = c_3 = 0$ for u_v and $c_1 = c_2 = c_3 = c_4 = 0$ for d_v. Those forms are far too simple to describe our data set: using them in place of our choice increases χ^2 by more than 200.

We made a case in previous work [141] to repackage polynomial parametrizations like (A3) as linear combinations of Chebyshev polynomials of argument $1 - 2\sqrt{x}$. This method has been adopted in the recent MMHT2014 fit [124]. However, we now contend that repackaging based

on a linear combination of Bernstein polynomials, as we do in CT14, is much better. The full functional forms available in the fit are, of course, the same either way. But, because each of the Bernstein polynomials has a single peak, and the peaks occur at different values of x, the coefficients that multiply those polynomials mainly control distinct physical regions and are therefore somewhat independent of each other.

In contrast, every Chebyshev polynomial of argument $1 - 2\sqrt{x}$ has a maximum value ± 1 at both $x = 0$ and $x = 1$, along with an equal maximum magnitude at some interior points. All Chebyshev polynomials are important over the entire range of x, so their coefficients are strongly correlated in the fit. This causes minor difficulties in finding the best fit and major difficulties in using the Hessian method to estimate uncertainties based on orthogonal eigenvectors. Furthermore, using Bernstein polynomials makes it easy to enforce the desired positivity of the PDFs in the $x \to 0$ and $x \to 1$ limits, because each of those limits is controlled by a single polynomial.

We use a similar parametrization for the gluon, but with a polynomial of a lower order, because the data provide fewer constraints on the gluon distribution:

$$P_g(y) = g_0[e_0 q_0(y) + e_1 q_1(y) + q_2(y)], \quad \text{(A7)}$$

where

$$q_0(y) = (1 - y)^2,$$
$$q_1(y) = 2y(1 - y),$$
$$q_2(y) = y^2. \quad \text{(A8)}$$

However, in place of $y = \sqrt{x}$, we use the mapping

$$y = 1 - (1 - \sqrt{x})^2 = 2\sqrt{x} - x. \quad \text{(A9)}$$

This mapping makes $y = 1 - (1 - x)^2/4 + \mathcal{O}((1 - x)^3)$ and hence

$$P_g(y) \to \text{const} + \mathcal{O}((1 - x)^2) \quad \text{(A10)}$$

in the limit $x \to 1$. This is an alternative way to suppress the first subleading power of $(1 - x)$ at $x \to 1$. We have five free parameters to describe the gluon distribution, including g_0 which governs the fraction of momentum carried by the gluons. The best fit has $a_2 = 3.8$, with the range $2.6 < a_2 < 5.0$ allowed by an increase of only 5 in χ^2.

In contrast, CT10 NNLO [6] again used the form (A2) for the gluon distribution, where a_2 was frozen at an arbitrary value of 10 because χ^2 was rather insensitive to it. That left the same number of free parameters as are used here but did not allow anything to be learned about the behavior at very large x.

If we use (A2) for the gluon in our present fit, the resulting χ^2 is nearly as good, but again this choice yields almost no information about the sixth parameter a_2: a range of $\Delta\chi^2 = 1$ includes $-0.4 < a_2 < 12$. The negative a_2 part of that range corresponds to an integrably singular gluon probability density at $x \to 1$, which is not actually forbidden theoretically, but would be totally unexpected. This older parametrization would bring in unmotivated complexity in the large-x region that is not indicated by any present data. To test that our parametrization has adequate flexibility, we made similar fits using somewhat higher-order Bernstein polynomials, including up to a total of ten more free parameters. We calculated the uncertainty for the $gg \to H$ cross section at 8 TeV using the Lagrange multiplier method and found very little variation in the range of the prediction. We also calculated the range of uncertainty in $\alpha_s(m_Z)$ obtained from our fits at 90% confidence (including our tier-2 penalty). The extra freedom in parametrization increased the uncertainty range only slightly: 0.111–0.121 using the CT14 parametrization; 0.111–0.123 using the more flexible one.

The sea quark distributions $\bar{d}$ and $\bar{u}$ were parametrized using fourth-order polynomials in y with the same mapping $y = 2\sqrt{x} - x$ that was used for the gluon. We assumed $\bar{u}(x)/\bar{d}(x) \to 1$ at $x \to 0$, which implies $a_1(\bar{u}) = a_1(\bar{d})$. As the strangeness content is constrained rather poorly, we used a minimal parametrization $P_{s+\bar{s}} = \text{const}$, with a_1 tied to the common a_1 of $\bar{u}$ and $\bar{d}$. Even fewer experimental constraints apply to the strangeness asymmetry, so we have assumed $s(x) = \bar{s}(x)$ in this analysis. Thus, we have just two parameters for strangeness in our Hessian method: a_2 and normalization. In view of more upcoming data on measuring the asymmetry in the production cross sections of $W + \bar{c}$ and $W + c$ from the LHC, we plan to include $s(x) \neq \bar{s}(x)$ in our next round of fits.

In all, we have eight parameters associated with the valence quarks, five parameters associated with the gluon, and 13 parameters associated with sea quarks, for a total of 26 fitting parameters. Hence there are 52 eigenvector sets generated by the Hessian method that captures most of the PDF uncertainty.

The Hessian method tends to underestimate the uncertainty for PDF variations that are poorly constrained, because the method is based on the assumption that χ^2 is a quadratic function of the fitting parameters; and that assumption tends to break down when the parameters can move a long way because of a lack of experimental constraints. This can be seen, for example, for the case of the small-x gluon uncertainty, by a Lagrange multiplier scan in which a series of fits are made with different values of the independent variable $g(x, Q)$ at $x = 0.001$, $Q = Q_0$.

In order to include the wide variation of the gluon distribution that is allowed at small x, we therefore supplement the Hessian sets with an additional pair of sets that

were obtained using the Lagrange multiplier method: one with enhanced gluon and one with suppressed gluon at small x, as was already done in CT10. In CT14, we also include an additional pair of sets with enhanced or suppressed strangeness at small x; although it is possible that treating $a_1(s)$ as a fitting parameter independent from $a_1(\overline{u}) = a_1(\overline{d})$ would have worked equally well.

In summary, we have a total of 56 error sets: 2×26 from the Hessian method, supplemented by two extremes of small-x gluon, and two extremes of small-x strangeness. Uncertainties from all pairs of error sets are to be summed in quadrature using the master formulas [5,20,54]. In comparison, CT10 NNLO had 50 error sets. The increased flexibility in the CT14 parametrization is warranted by better experimental constraints and its improved fit to the data. Indeed, fitting the CT14 data set using the old CT10 parametrizations yields a best fit that is worse by 60 units in χ^2.

[1] ATLAS Collaboration, Phys. Lett. B **716**, 1 (2012).
[2] CMS Collaboration, Phys. Lett. B **716**, 30 (2012).
[3] ATLAS Collaboration and CMS Collaboration, Phys. Rev. Lett. **114**, 191803 (2015).
[4] C. Anastasiou, C. Duhr, F. Dulat, F. Herzog, and B. Mistlberger, Phys. Rev. Lett. **114**, 212001 (2015).
[5] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, and C.-P. Yuan, Phys. Rev. D **82**, 074024 (2010).
[6] J. Gao, M. Guzzi, J. Huston, H.-L. Lai, Z. Li, P. Nadolsky, J. Pumplin, and D. Stump, and C.-P. Yuan, Phys. Rev. D **89**, 033009 (2014).
[7] https://lhapdf.hepforge.org/.
[8] M. Guzzi, P. M. Nadolsky, H.-L. Lai, and C.-P. Yuan, Phys. Rev. D **86**, 053005 (2012).
[9] M. A. G. Aivazis, J. C. Collins, F. I. Olness, and W.-K. Tung, Phys. Rev. D **50**, 3102 (1994).
[10] J. C. Collins, Phys. Rev. D **58**, 094002 (1998).
[11] W.-K. Tung, S. Kretzer, and C. Schmidt, J. Phys. G **28**, 983 (2002).
[12] W.-K. Tung, H.-L. Lai, A. Belyaev, J. Pumplin, D. Stump, and C.-P. Yuan, J. High Energy Phys. 02 (2007) 053.
[13] P. M. Nadolsky, H.-L. Lai, Q. H. Cao, J. Huston, J. Pumplin, D. Stump, W.-K. Tung, and C.-P. Yuan, Phys. Rev. D **78**, 013004 (2008).
[14] V. M. Abazov et al. (D0 Collaboration), Phys. Rev. D **91**, 032007 (2015); **91**, 079901(E) (2015).
[15] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H.-L. Lai, and W.-K. Tung, Phys. Rev. D **65**, 014013 (2001).
[16] G. Watt and R. S. Thorne, J. High Energy Phys. 08 (2012) 052.
[17] D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H.-L. Lai, and W.-K. Tung, Phys. Rev. D **65**, 014012 (2001).
[18] K. A. Olive et al. (Particle Data Group Collaboration), Chin. Phys. C **38**, 090001 (2014).
[19] S. Alekhin et al., arXiv:1101.0536.
[20] J. Pumplin, D. R. Stump, J. Huston, H.-L. Lai, P. M. Nadolsky, and W.-K. Tung, J. High Energy Phys. 07 (2002) 012.
[21] V. Abazov et al. (D0 Collaboration), Phys. Rev. Lett. **101**, 211801 (2008).
[22] S. Dulat, T.-J. Hou, J. Gao, J. Huston, J. Pumplin, C. Schmidt, D. Stump, and C.-P. Yuan, Phys. Rev. D **89**, 073004 (2014).
[23] A. C. Benvenuti et al. (BCDMS Collaboration), Phys. Lett. B **223**, 485 (1989).
[24] A. C. Benvenuti et al. (BCDMS Collaboration), Phys. Lett. B **237**, 592 (1990).
[25] M. Arneodo et al. (New Muon Collaboration), Nucl. Phys. **B483**, 3 (1997).
[26] J. P. Berge et al., Z. Phys. C **49**, 187 (1991).
[27] U. K. Yang et al. (CCFR/NuTeV Collaboration), Phys. Rev. Lett. **86**, 2742 (2001).
[28] W. G. Seligman et al., Phys. Rev. Lett. **79**, 1213 (1997).
[29] D. A. Mason, Ph.D. thesis, University of Oregon, Report No. FERMILAB-THESIS-2006-01, UMI-32-11223.
[30] M. Goncharov et al. (NuTeV Collaboration), Phys. Rev. D **64**, 112006 (2001).
[31] A. Aktas et al. (H1 Collaboration), Eur. Phys. J. C **40**, 349 (2005).
[32] H. Abramowicz et al. (H1 and ZEUS Collaborations), Eur. Phys. J. C **73**, 2311 (2013).
[33] F. Aaron et al. (H1 and ZEUS Collaborations), J. High Energy Phys. 01 (2010) 109.
[34] F. Aaron et al. (H1 Collaboration), Eur. Phys. J. C **71**, 1579 (2011).
[35] G. Moreno et al., Phys. Rev. D **43**, 2815 (1991).
[36] R. Towell et al. (NuSea Collaboration), Phys. Rev. D **64**, 052002 (2001).
[37] J. Webb et al. (NuSea Collaboration), arXiv:hep-ex/0302019.
[38] F. Abe et al. (CDF Collaboration), Phys. Rev. Lett. **77**, 2616 (1996).
[39] D. Acosta et al. (CDF Collaboration), Phys. Rev. D **71**, 051104 (2005).
[40] V. Abazov et al. (D0 Collaboration), Phys. Rev. D **77**, 011106 (2008).
[41] R. Aaij et al. (LHCb Collaboration), J. High Energy Phys. 06 (2012) 058.
[42] V. Abazov et al. (D0 Collaboration), Phys. Lett. B **658**, 112 (2008).
[43] T. A. Aaltonen et al. (CDF Collaboration), Phys. Lett. B **692**, 232 (2010).
[44] CMS Collaboration, Phys. Rev. D **90**, 032004 (2014).

[45] CMS Collaboration, Phys. Rev. Lett. **109**, 111806 (2012).

[46] ATLAS Collaboration, Phys. Rev. D **85**, 072004 (2012).

[47] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. D **78**, 052006 (2008).

[48] V. Abazov *et al.* (D0 Collaboration), Phys. Rev. Lett. **101**, 062001 (2008).

[49] ATLAS Collaboration, Phys. Rev. D **86**, 014022 (2012).

[50] CMS Collaboration, Phys. Rev. D **87**, 112002 (2013).

[51] J. C. Collins and J. Pumplin, arXiv:hep-ph/0105207.

[52] J. Pumplin, Phys. Rev. D **81**, 074010 (2010).

[53] J. Pumplin, Phys. Rev. D **80**, 034002 (2009).

[54] P. M. Nadolsky and Z. Sullivan, *Proceedings of Snowmass 2001*, eConf C010630, P510 (2001).

[55] F. Abe *et al.* (CDF Collaboration), Phys. Rev. Lett. **74**, 850 (1995).

[56] E. L. Berger, F. Halzen, C. S. Kim, and S. Willenbrock, Phys. Rev. D **40**, 83 (1989); **40**, 3789 (1989).

[57] A. D. Martin, R. G. Roberts, and W. J. Stirling, Mod. Phys. Lett. A **04**, 1135 (1989).

[58] H.-L. Lai, J. Botts, J. Huston, J. G. Morfin, J. F. Owens, J. Qiu, W.-K. Tung, and H. Weerts, Phys. Rev. D **51**, 4763 (1995).

[59] A. D. Martin, W. J. Stirling, and R. G. Roberts, Phys. Rev. D **50**, 6734 (1994).

[60] F. D. Aaron *et al.* (H1 Collaboration), Eur. Phys. J. C **71**, 1769 (2011); **72**, 2252(E) (2012).

[61] A. Aktas *et al.* (H1 Collaboration), Eur. Phys. J. C **45**, 23 (2006).

[62] S. Chekanov *et al.* (ZEUS Collaboration), Phys. Rev. D **69**, 012004 (2004).

[63] J. Breitweg *et al.* (ZEUS Collaboration), Eur. Phys. J. C **12**, 35 (2000).

[64] J. Gao, M. Guzzi, and P. M. Nadolsky, Eur. Phys. J. C **73**, 2541 (2013).

[65] ATLAS Collaboration, Phys. Rev. D **90**, 072004 (2014).

[66] CMS Collaboration, Eur. Phys. J. C **75**, 542 (2015).

[67] CMS Collaboration, Eur. Phys. J. C **73**, 2339 (2013).

[68] A. Mitov, in Proceedings of the 23rd Workshop on Deep Inelastic Scattering and Related Subjects, DIS 2015 (unpublished), https://indico.cern.ch/event/341292/session/3/contribution/257.

[69] J. H. Kühn, A. Scharf, and P. Uwer, Phys. Rev. D **91**, 014020 (2015).

[70] J. H. Kühn, A. Scharf, and P. Uwer, Eur. Phys. J. C **51**, 37 (2007).

[71] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, Phys. Rev. Lett. **91**, 182002 (2003).

[72] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, Phys. Rev. D **69**, 094008 (2004).

[73] C. Balazs, J. Qiu, and C.-P. Yuan, Phys. Lett. B **355**, 548 (1995).

[74] C. Balazs and C.-P. Yuan, Phys. Rev. D **56**, 5558 (1997).

[75] F. Landry, R. Brock, P. M. Nadolsky, and C.-P. Yuan, Phys. Rev. D **67**, 073016 (2003).

[76] M. Guzzi, P. M. Nadolsky, and B. Wang, Phys. Rev. D **90**, 014030 (2014).

[77] M. C. Kumar and S. O. Moch, Phys. Lett. B **730**, 122 (2014).

[78] N. Kidonakis and J. F. Owens, Phys. Rev. D **63**, 054019 (2001).

[79] R. D. Ball *et al.* (NNPDF Collaboration), J. High Energy Phys. 04 (2015) 040.

[80] J. Gao, Z. Liang, D. E. Soper, H.-L. Lai, P. M. Nadolsky, and C.-P. Yuan, Comput. Phys. Commun. **184**, 1626 (2013).

[81] R. D. Ball *et al.*, J. High Energy Phys. 04 (2013) 125.

[82] M. Cacciari and N. Houdeau, J. High Energy Phys. 09 (2011) 039.

[83] A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, and J. Pires, Phys. Rev. Lett. **110**, 162003 (2013).

[84] J. Currie, A. Gehrmann-De Ridder, E. W. N. Glover, and J. Pires, J. High Energy Phys. 01 (2014) 110.

[85] S. Carrazza and J. Pires, J. High Energy Phys. 10 (2014) 145.

[86] M. Wobisch *et al.* (fastNLO Collaboration), arXiv:1109.1310.

[87] 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).

[88] Z. Nagy, Phys. Rev. Lett. **88**, 122003 (2002).

[89] Z. Nagy, Phys. Rev. D **68**, 094002 (2003).

[90] T. Lewis, Australian Journal of statistics **30A**, 160 (1988).

[91] G. D'Agostini, Nucl. Instrum. Methods Phys. Res., Sect. A **346**, 306 (1994).

[92] G. D'Agostini, Report Nos. CERN-99-03 and CERN-YELLOW-99-03.

[93] H.-L. Lai, P. M. Nadolsky, J. Pumplin, D. Stump, W.-K. Tung, and C.-P. Yuan, J. High Energy Phys. 04 (2007) 089.

[94] J. F. Owens, A. Accardi, and W. Melnitchouk, Phys. Rev. D **87**, 094012 (2013).

[95] M. Czakon, P. Fiedler, and A. Mitov, Phys. Rev. Lett. **110**, 252004 (2013).

[96] M. Czackon and A. Mitov, Comput. Phys. Commun. **185**, 2930 (2014).

[97] C. Anastasiou, S. Buehler, F. Herzog, and A. Lazopoulos, J. High Energy Phys. 12 (2011) 058.

[98] CMS Collaboration, Phys. Rev. Lett. **112**, 191802 (2014).

[99] S. D. Ellis, Z. Kunszt, and D. E. Soper, Phys. Rev. Lett. **69**, 1496 (1992).

[100] Z. Kunszt and D. E. Soper, Phys. Rev. D **46**, 192 (1992).

[101] M. Cacciari, G. P. Salam, and G. Soyez, Eur. Phys. J. C **72**, 1896 (2012).

[102] ATLAS Collaboration, Eur. Phys. J. C **73**, 2509 (2013).

[103] A. D. Martin, A. J. T. M. Mathijssen, W. J. Stirling, R. S. Thorne, B. J. A. Watt, and G. Watt, Eur. Phys. J. C **73**, 2318 (2013).

[104] A. Accardi, W. Melnitchouk, J. F. Owens, M. E. Christy, C. E. Keppel, L. Zhu, and J. G. Morfin, Phys. Rev. D **84**, 014008 (2011).

[105] ATLAS Collaboration, Phys. Lett. B **725**, 223 (2013).

[106] ATLAS Collaboration, J. High Energy Phys. 06 (2014) 112.

[107] K. Melnikov and F. Petriello, Phys. Rev. D **74**, 114017 (2006).

[108] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, Comput. Phys. Commun. **182**, 2388 (2011).

[109] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, Comput. Phys. Commun. **184**, 208 (2013).

[110] Y. Li and F. Petriello, Phys. Rev. D **86**, 094034 (2012).
[111] ATLAS Collaboration, Phys. Rev. Lett. **109**, 012001 (2012).
[112] ATLAS Collaboration, J. High Energy Phys. 05 (2014) 068.
[113] CMS Collaboration, J. High Energy Phys. 02 (2014) 013.
[114] S. Alekhin *et al.*, Eur. Phys. J. C **75**, 304 (2015).
[115] O. Samoylov *et al.* (NOMAD Collaboration), Nucl. Phys. **B876**, 339 (2013).
[116] S. Alekhin, J. Bluemlein, L. Caminadac, K. Lipka, K. Lohwasser, S. Moch, R. Petti, and R. Placakyte, Phys. Rev. D **91**, 094002 (2015).
[117] J. M. Campbell and R. K. Ellis, Nucl. Phys. B, Proc. Suppl. **205–206**, 10 (2010).
[118] S. Dittmaier *et al.* (LHC Higgs Cross Section Working Group Collaboration), arXiv:1101.0593.
[119] S. Dittmaier *et al.*, arXiv:1201.3084.
[120] S. Heinemeyer *et al.* (LHC Higgs Cross Section Working Group Collaboration), arXiv:1307.1347.
[121] C. Schmidt, S. Dulat, J. Gao, M. Guzzi, T.-J. Hou, J. W. Huston, P. Nadolsky, J. Pumplin, D. Stump, and C.-P. Yuan, *Proc. Sci.*, DIS2014 (2014) 146.
[122] S. Dulat, T.-J. Hou, J. Gao, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, and D. Stump, and C.-P. Yuan, Phys. Rev. D **89**, 113002 (2014).
[123] H.-L. Lai, J. Huston, Z. Li, P. Nadolsky, J. Pumplin, D. Stump, and C.-P. Yuan, Phys. Rev. D **82**, 054021 (2010).

[124] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, Eur. Phys. J. C **75**, 204 (2015).
[125] J. M. Campbell, J. W. Huston, and W. J. Stirling, Rep. Prog. Phys. **70**, 89 (2007).
[126] ATLAS Collaboration, Phys. Lett. B **707**, 459 (2012).
[127] ATLAS Collaboration, Phys. Rev. D **91**, 112013 (2015).
[128] CMS Collaboration, J. High Energy Phys. 11 (2012) 067.
[129] CMS Collaboration, Report No. CMS-PAS-TOP-12-006.
[130] CMS Collaboration, Report No. CMS-PAS-TOP-14-016.
[131] ATLAS Collaboration, Reports Nos. ATLAS-CONF-2012-134 and ATLAS-COM-CONF-2012-166.
[132] M. Czakon, M. L. Mangano, A. Mitov, and J. Rojo, J. High Energy Phys. 07 (2013) 167.
[133] M. Guzzi, K. Lipka, and S. O. Moch, J. High Energy Phys. 01 (2015) 082.
[134] CTEQ-TEA group (to be published).
[135] http://hep.pa.msu.edu/cteq/public/ct14.html.
[136] http://hep.pa.msu.edu/cteq/public/ct14/lhapdf/v5/.
[137] S. Brodsky (private communication).
[138] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, Eur. Phys. J. C **63**, 189 (2009).
[139] H. Abramowicz *et al.* (H1 and ZEUS Collaborations), Eur. Phys. J. C **75**, 580 (2015).
[140] O. Zenaiev *et al.*, Eur. Phys. J. C **75**, 396 (2015).
[141] J. Pumplin, Phys. Rev. D **82**, 114020 (2010).