A study of the impact of double-differential top distributions from CMS on parton distribution functions

Michał Czakon\textsuperscript{(a,1)}, Sayipjamal Dulat\textsuperscript{(b,2)}, Tie-Jiun Hou\textsuperscript{(c,3)}, Joey Huston\textsuperscript{(d,4)}, Alexander Mitov\textsuperscript{(e,5)}, Andrew S. Papanastasiou\textsuperscript{(e,f,6)}, Ibrahim Sitiwaldi\textsuperscript{(a,7)}, Zhite Yu\textsuperscript{(d,8)}, and C.–P. Yuan\textsuperscript{(d,9)}

\textsuperscript{(a)}Institut für Theoretische Teilchenphysik und Kosmologie, RWTH Aachen University, D-52056 Aachen, Germany

\textsuperscript{(b)}School of Physics Science and Technology Xinjiang University, Urumqi, Xinjiang 830046 China.

\textsuperscript{(c)}Department of Physics, College of Sciences, Northeastern University, Shenyang 110819, China

\textsuperscript{(d)}Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824 U.S.A.

\textsuperscript{(e)}Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK

\textsuperscript{(f)}MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Crewe Road, Edinburgh EH4 2XU, UK

(Dated: December 19, 2019)
Abstract

LHC $t\bar{t}$ data have the potential to provide constraints on the gluon distribution, especially at high $x$, with both ATLAS and CMS performing differential measurements. Recently, CMS has measured double-differential $t\bar{t}$ distributions at 8 TeV. In this paper, we examine the impact of this data set on the gluon distribution, using the recently calculated double-differential NNLO predictions for that data. No significant impact is found when the CMS data is added to the CT14HERA2 global PDF fit, due to the larger impact of the inclusive jet data from both the Tevatron and the LHC. If the jet data are removed from the fit, then an impact is observed. If the CMS data is scaled by a larger weight, representing the greater statistical power of the jet data, a roughly equal impact on the gluon distribution is observed for the $t\bar{t}$ as for the inclusive jet data. For data samples with higher integrated luminosity at 13 TeV, a more significant impact of the double-differential $t\bar{t}$ data may be observed.

PACS numbers: 12.38.-t,12.38.Bx,12.38.Aw
I. INTRODUCTION

One of the limitations in searches for potential new physics at the LHC is the theoretical uncertainty in predictions for the standard model backgrounds to the new physics. In general, new physics is expected to occur at high masses, and thus requires the colliding partons to have relatively large fractions ($x$) of the parent protons’ momenta. These theoretical uncertainties include those related to the parton distribution functions (PDFs) at high $x$, especially those of the gluon distribution, the most poorly known PDF in this kinematic region. Until recently, the only data included in global PDF fits sensitive to the value of the high $x$ gluon were those from inclusive jet production. Older parton distribution functions have used only jet data from the Tevatron; with newer generations of PDFs, jet data from the LHC has been added and generally has a significance equal to, or greater than, the Tevatron jet data, due to the wider kinematic range and the smaller systematic errors.

For high transverse momentum jet production, however, the gluon distribution is sub-dominant, with $qq$ scattering being the dominant sub-process, followed by $gq$ scattering. Top pair production, on the other hand, is dominated by the $gg$ initial state, and thus provides a direct handle on the gluon distribution. For top pair production at high mass, rapidity or
transverse momentum, the sensitivity continues to high momentum fraction $x$ values.

Both ATLAS and CMS have measured top pair production with variables such as the $t\bar{t}$ mass, rapidity ($y$), either of the individual top quark (anti-quark), or of the pair, and the transverse momentum ($p_T$), again either of the individual top quark (anti-quark), or of the combination $[1, 2]$. This data has been included in previous PDF fits (see, for example, Ref. [3]) and will be included in CT18. Each distribution, or combination, has a sensitivity to the gluon distribution. Recently, CMS has measured double-differential top pair distributions, using combinations of the above variables $[4]$, which have the potential to provide a greater sensitivity to the initial state gluon distribution, when combined with the recent NNLO calculation of such double-differential distributions. The NNLO calculation of top pair production of course also depends greatly on the value of $\alpha_s(m_Z)$, which itself is anti-correlated with the high $x$ gluon distribution. In this paper, we explore the relative sensitivity and importance of the double-differential top-pair distributions, to the high $x$ gluon distribution, with current data, and with extrapolations to what might be expected from future data.

This paper is organized as follows: Sec.II describes the double-differential measurements of CMS. Sec.III then discusses the theoretical framework for the calculation of the double-differential theoretical predictions, and their inclusion in fastNNLO. Sec.IV then explores the correlation between the measured distributions and the parton $x$ values of the gluon distribution. The correlations indicate the kinematic range over which the data may have some influence on the gluon distribution in the global PDF fits. Correlation, however, is not sufficient by itself to describe the impact. In Sec.V ePump $[5]$, is used to update the PDFs in the CTEQ-TEA fitting framework. We discuss the impact of adding the CMS double-differential top data to the CT14HERA2 global fit $^1$, with and without jet data, from the Tevatron and the LHC, included in the original CT14HERA2 data set.

Finally, Sec.VI concludes, and offers a projection of the impact of additional data at the LHC.

---

1 CT14HERA2 is the latest published CT set at the time of writing of this paper, with the CT18 paper in progress. The gluon distribution for CT14HERA2 is similar to that obtained in CT14, except at very high $x$ where CT14 has a harder gluon.
II. EXPERIMENTAL OVERVIEW

In this work, we consider the double-differential top-quark data from CMS [3], which consists of the following normalized $t\bar{t}$ distributions: the transverse momentum of the top quark ($p_T^t$) as a function of the top rapidity ($y_t$), the top-antitop system transverse momentum ($p_T^{t\bar{t}}$) as a function of the $t\bar{t}$ mass ($m_{t\bar{t}}$), the pseudo-rapidity separation of the top pair ($\Delta \eta_{t\bar{t}}$) as a function of the $t\bar{t}$ mass ($m_{t\bar{t}}$), the rapidity of the top quark ($y_t$) as a function of the $t\bar{t}$ mass ($m_{t\bar{t}}$), the $t\bar{t}$ rapidity ($y_{t\bar{t}}$) as a function of the $t\bar{t}$ mass ($m_{t\bar{t}}$), the azimuthal separation ($\Delta \phi_{t\bar{t}}$) of the top and anti-top as a function of the $t\bar{t}$ mass ($m_{t\bar{t}}$). The last distribution is particularly sensitive to the effects of soft gluon radiation, so will not be used in the following comparison to fixed-order predictions. The data sets, the number of data points in each set, and the internal CTEQ-TEA reference number are given in Table 1 below. The data were taken at 8 TeV with an integrated luminosity of 19.7 fb$^{-1}$. The statistical and systematic errors are typically of similar size, with the largest systematic error being due to the jet energy scale. In the original CMS paper [3], the data were compared to NLO fixed-order predictions, to NLO+parton shower predictions and to fixed-order approximate NNLO predictions (for several observables). For this paper, comparisons are made to full NNLO predictions.

TABLE I: The double-differential $t\bar{t}$ data sets used in this study. The ID number refers to the internal references inside the CTEQ-TEA fitting code.

| ID | data | no. of data points |
|----|------|-------------------|
| 573 | $d^2\sigma/dy_tdp_T^t$ | 16 |
| 574 | $d^2\sigma/dm_{t\bar{t}}dp_T^{t\bar{t}}$ | 16 |
| 575 | $d^2\sigma/dm_{t\bar{t}}d\Delta \eta_{t\bar{t}}$ | 12 |
| 576 | $d^2\sigma/dm_{t\bar{t}}dy_{t\bar{t}}$ | 16 |
| 577 | $d^2\sigma/dm_{t\bar{t}}dy_{t\bar{t}}$ | 16 |
III. NNLO CALCULATION OF DIFFERENTIAL TOP-PAIR DISTRIBUTIONS

In this work we calculate the NNLO QCD corrections to one- and two-dimensional top quark-pair differential distributions at the LHC. The distributions are defined in terms of the following top quark kinematic variables: $p_T^t$, $m_{tt}$, $p_T^{\bar{t}}$, $y_t$, $y_{t\bar{t}}$ and $\Delta \eta_{t\bar{t}}$. The $p_T^t$ and $y_t$ distributions are averaged over the corresponding top and antitop distributions. Our binning follows the CMS collaboration’s 8 TeV measurement [4].

We use $m_t = 173.3$ GeV and utilize the dynamic scales derived in Ref. [6] (see also Ref. [7]):

\begin{align*}
\mu_0 &= \frac{m_T}{2} = \frac{1}{2} \sqrt{p_T^{2,t} + m_t^2}, \quad (1) \\
\mu_0 &= \frac{H_T}{4} = \frac{1}{4} \left( \sqrt{p_T^{2,t} + m_t^2} + \sqrt{p_T^{2,\bar{t}} + m_t^2} \right). \quad (2)
\end{align*}

Specifically, we compute the one-dimensional average top $p_T^t$ distribution with the scale Eq. (1) while all other one-dimensional distributions and all two-dimensional ones are computed with the help of the scale Eq. (2). Scale variation is derived with the help of the usual 7-point variation of the factorization and renormalization scales around the central scale $\mu_0$.

The calculations performed in this work are used to produce tables in the fastNLO format [8, 9]. More details about our fastNLO tables can be found in Ref. [10]. These tables have the advantage that predictions can be recalculated very fast with any PDF set or for any value of $\alpha_S$. As a cross-check, we have also provided in electronic format two binned predictions based on the NNPDF30 nnlo as 0118 [11] and CT14nnlo as 0111 [12] PDF sets. All predictions can be downloaded from the following webpage (http://www.precision.hep.phy.cam.ac.uk/results/ttbar-fastnlo/).

In this work we follow the STRIPPER approach [13–15] previously applied to top-pair production in Refs. [6, 16–19]. We have implemented it in a flexible, fully-differential partonic Monte Carlo program which, in principle, is able to calculate any infrared safe partonic observable. Further technical details can be found in Ref. [20]. Two-dimensional distributions in NNLO QCD have recently also been calculated in Ref. [21] for a different LHC setup.
IV. DOUBLE-DIFFERENTIAL TOP DATA AND THE CORRELATION TO THE GLUON DISTRIBUTION

The double-differential $t\bar{t}$ data are expected to have the strongest correlation with the gluon PDF, as the dominant $t\bar{t}$ production mechanism at the LHC is through $gg$ fusion. This argument can be demonstrated quantitatively by examining the correlation cosines for the $t\bar{t}$ data as a function of the gluon momentum fraction $x$. The quantity $\cos \phi$, the correlation cosine, characterizes whether the PDF degrees of freedom of quantities $X$ and $Y$ are correlated ($\cos \phi \approx 1$), anti-correlated ($\cos \phi \approx -1$), or uncorrelated ($\cos \phi \approx 0$)\cite{22}. In this case, $X$ and $Y$ are the gluon distribution and the double-differential $t\bar{t}$ data.

As an example, the correlation cosines between $t\bar{t}$ data and the gluon and up quark distributions are shown in Fig. 1 for the CMS data set labelled 573, $(d^2\sigma/dy_{t}\,dp_{Tt})^2$. There are 16 data points in this data set, so the plot contains 16 curves. It is apparent that gluon PDF has a stronger correlation with the $t\bar{t}$ data, and in particular that it is mainly the gluon PDF for $x > 0.01$ range that has strong correlations. Note that approximately half of the data points have a strong correlation with the gluon distribution at an $x$ value above 0.1 and a strong anti-correlation with the gluon near $x$ of 0.01, while (approximately) the other half of the data points have the opposite behavior. There is not much correlation in the $x$

\footnote{The correlations with the other PDF flavors are similar to that observed for the up quark.}
FIG. 2: The correlations between the $t\bar{t}$ observables in data sets 574 to 577 ($m_{t\bar{t}}{p_T}$, $m_{t\bar{t}}-\Delta\eta_{t\bar{t}}$, $m_{t\bar{t}}$-$y_t$ and $m_{t\bar{t}}$-$y_{t\bar{t}}$ double differential distributions) and the CT14HERA2 gluon PDFs, as a function of $x$. Each curve corresponds to a separate data point.

region around 0.1, and as will be seen later, the constraints on the gluon distribution by the top data sets tend to be weaker here.

The high $x$ gluon in particular still has a great deal of uncertainty in global PDF fits. The range around 0.01 is also of interest as it plays a role in Higgs boson production through $gg$ fusion. The correlations between the other CMS $t\bar{t}$ observables and the gluon PDF are shown in Fig. 2, where the same conclusion holds. Although this demonstrates that the CMS 2D $t\bar{t}$ observables depend highly on gluon PDFs in these $x$ ranges, it does not necessarily mean that the $t\bar{t}$ data are going to have a strong impact on the determination of the gluon PDF in a global fit.

The impact of a data set on a global PDF fit has been discussed in Ref. [23], as involving not only a correlation between the data and specific PDFs, in a given $x$ range, but also

---

3 It is interesting that data sets 576 and 577 have strong correlations more uniformly spread in $x$ than do the other distributions.
a sensitivity of the data to those PDFs. The sensitivity is determined by the number of 
data points, the kinematic range they cover, and the magnitudes of the statistical and 
systematic errors (and the correlations of the latter). It can be shown that [3, 23] one of 
the strongest sensitivities (per data point) for the gluon distribution is given by the CMS 
double-differential top data (not included in CT14 or CT14HERA2, but will be in CT18). 
Although, the CMS double-differential top data has one of the highest sensitivities per data 
point, the largest data set sensitivities belong to the HERA I+II data set and the CMS 7 
TeV inclusive jet data set. The former has a relatively low average sensitivity per data point 
but has 1120 data points. The latter has a moderate sensitivity to the gluon distribution per 
data point, but has 185 data points, most with reasonably small statistical and systematic 
errors. In the next section, we will examine the actual impact of the top and jet data sets 
on a global PDF fit using the program ePump. As shown in Ref. [3 5], ePump can quickly 
provide quantitative information on the impact of a given data set to updating PDFs and 
their error bands, including the information on the relevant parton flavor and $x$ range.

V. RESULTS OF PDF FITTING

As introduced in Ref. [5], ePump is a convenient software tool that allows an examination 
of the impact of a new data set, without the need to perform a complete global PDF 
fit. The $\chi^2$ and dof for each of the CMS double-differential top data sets, compared to 
NNLO predictions using CT14HERA2, are shown in Table II first without including the 
data set in the fit, and then including the data set via ePump. The data provided by the 
experiment group [4] are normalized distributions, with correlated systematic uncertainties 
and correlation matrices of statistical uncertainties. Due to the loss of one degree of freedom 
when constructing normalized distributions, the correlation matrices are singular, so when 
we use the data to update PDFs (or to calculate the $\chi^2$), we discard the bin with the largest 
values of kinematic variables and the corresponding correlation coefficients for each data set, 
as instructed by the experiment paper [4].

Several aspects can be immediately noticed. First, the $\chi^2$/dof for each of the data sets is 
on the order of 1.5-2, except for data set 574 which has a $\chi^2$/dof of over 5. Second, there is 
minimal improvement in the $\chi^2$/dof when each data set is included in the global fit.

When each double-differential $t\bar{t}$ data set is individually added to CT14HERA2, no strong
TABLE II: The $\chi^2$ for each 2D $t\bar{t}$ data set, calculated with the original global-fit CT14HERA2 PDFs and ePump updated CT14HERA2 PDFs.

| ID  | data                                | dof | $\chi^2$ before updating | $\chi^2$ after updating |
|-----|-------------------------------------|-----|--------------------------|--------------------------|
| 573 | $d^2\sigma/dy_t dp_T$               | 15  | 35.5                     | 34.9                     |
| 574 | $d^2\sigma/dm_{t\bar{t}} dp_T$     | 15  | 82.3                     | 80.6                     |
| 575 | $d^2\sigma/dm_{t\bar{t}} d\Delta\eta_{t\bar{t}}$ | 11  | 22.1                     | 22.0                     |
| 576 | $d^2\sigma/dm_{t\bar{t}} dy_{t\bar{t}}$ | 15  | 20.2                     | 20.1                     |
| 577 | $d^2\sigma/dm_{t\bar{t}} dy_{t\bar{t}}$ | 15  | 23.8                     | 23.5                     |

FIG. 3: The updated gluon PDF when the data set 577 (CMS 8TeV $m_{t\bar{t}}$-$y_{t\bar{t}}$ double differential data) is added to CT14HERA2 using ePump with weight=1. The suffix “.54” is to stress that there are 54 eigen-sets used in CT14HERA2 global fit, without the two gluon extreme sets. The letter “e” is to note that the PDF was obtained by ePump. Left: PDF central values. Right: Error bands.

Impacts are observed on the gluon distribution. A similar result was noted for the influence of the single-differential top measurements [3]. Among the double-differential $t\bar{t}$ data sets, ID 577 (differential in $m_{t\bar{t}}$ and $y_{t\bar{t}}$) has the most noticeable impact on the gluon distribution. Fig. 3 shows the updated gluon PDF and its error band when data set 577 is added to CT14HERA2 using ePump.

The central value of the gluon PDF changes only at large $x$ values, in a region basically unconstrained by present data, and there is no notable change in the gluon PDF uncertainty. This is the result of the $t\bar{t}$ data being dominated by the effects of the other data sets,
FIG. 4: The updated gluon PDF and its error band when data set 577 (CMS 8TeV $m_{t\bar{t}}-y_{t\bar{t}}$ double differential data) is added to CT14HERA2mJ using ePump with weight=1. The suffix “.54” is to stress that there are 54 eigen-sets used in CT14HERA2 global fit, without the two gluon extreme sets. The letter “e” is to note that the PDF was obtained by ePump. Left: PDF central values. Right: Error bands.

including the jet data, included in the CT14HERA2 fit. If the jet data from the Tevatron and LHC are removed from this fit (which yields a new set of global fit PDFs, named PDF CT14HERA2mJ), then the impact of the $t\bar{t}$ double-differential data is noticeably larger. The ePump-updated gluon PDF of CT14HERA2mJ, using data set 577, is shown in Fig. 4 (Note that CT14HERA2mJ now serves as the reference set.) Some clear trends are observed. The gluon distribution at high $x$ (above 0.15) is larger than that preferred by CT14HERA2 (but still somewhat smaller than that preferred by CT14HERA2mJ). The PDF uncertainty is still larger than that of CT14HERA2 for $x \geq 10^{-4}$. A comparison of Fig. 3 and 4 indicates that the double-differential $t\bar{t}$ data has an impact on the best-fit gluon-PDF in the large $x$ region, but in the presence of the jet data, the impact on gluon-PDF error band is diminished. Hence, we conclude that the sensitivity of each $t\bar{t}$ data point is about the same as the jet data, but the overall sensitivity of the $t\bar{t}$ data set is far less than the jet data due to the much smaller number of total data points in the $t\bar{t}$ data set.

The level of agreement of the $t\bar{t}$ data with the NNLO predictions, using CT14HERA2 or CT14HERA2mJ, can be observed by comparing the theory predictions and the data for each data point. Fig. 5 and 6 show the comparisons for the CMS 8TeV $m_{t\bar{t}}-y_{t\bar{t}}$ data set, for CT14HERA2 and CT14HERA2mJ, respectively. In the comparison of the data to the NNLO prediction using CT14HERA2, the data points are shifted according to the optimal
FIG. 5: The comparison between the NNLO theory prediction of CT14HERA2 and experiment data for each data point for the $m_{t\bar{t}}$-$y_{t\bar{t}}$ double differential data set. The blue data points indicate the level of agreement with the NNLO predictions using CT14HERA2, without allowing for any systematic error shifts in the fit for each data point. The blue error bars indicate the size of the systematic errors. The black diamonds indicate the impact of the systematic error shifts that lead to the best fit with CT14HERA2, and the green triangles show the impact of the data shifts in the global fit after the data set is included in the global fit. In doing this comparison, only the uncorrelated uncertainties and correlated systematic uncertainties are included, without the correlation matrices of statistical uncertainties. Thus we can have clear definitions of shifted data here and show 16 data points at the same time.

Systematic error shifts leading to best agreement with the theoretical prediction. The shifted data points are closer to the theory prediction, as expected. Similar results are obtained for the other double-differential observables.

In total, there are 305 jet data points included in the CT14HERA2 PDF fit, including data from CDF and D0 at the Tevatron, and ATLAS and CMS at the LHC.\textsuperscript{4} The statistical errors vary from less than 1% at low transverse momentum to tens of percent at high $p_T$. In contrast, as shown in Table II, there are 16 data points for all but one of the CMS double-

\textsuperscript{4} Specifically, the data include 1.13 fb\textsuperscript{-1} from CDF and 0.70 fb\textsuperscript{-1} from D0, at a center-of-mass energy of 1.96 TeV and 4.5 fb\textsuperscript{-1} from ATLAS and 5 fb\textsuperscript{-1} from CMS, at a center-of-mass energy of 7 TeV.
FIG. 6: The comparison between the NNLO theory prediction of CT14HERA2mJ and experiment data for each data point for the $m_{t\bar{t}}$-$y_{t\bar{t}}$ double differential data set. The blue data points indicate the level of agreement with the NNLO predictions using CT14HERA2mJ, without allowing for any systematic error shifts in the fit for each data point. The blue error bars indicate the size of the systematic errors. The black diamonds indicate the impact of the systematic error shifts that lead to the best fit with CT14HERA2mJ, and the green triangles show the impact of the data shifts in the global fit after the data set is included in the global fit. In doing this comparison, only the uncorrelated uncertainties and correlated systematic uncertainties are included, without the correlation matrices of statistical uncertainties. Thus we can have clear definitions of shifted data here and show 16 data points at the same time.

The differential top data with statistical errors that vary from 2% to 17%, and systematic errors on the order of 3-17%. Thus, there is a factor of 19 times more jet data points than double-differential $t\bar{t}$ data points.

An interesting exercise is to increase the weight for the CMS $t\bar{t}$ data in the PDF updating (using ePump), using either CT14HERA2mJ or CT14HERA2 global-fit PDFs as the base, to a level that corresponds either to the statistical power of the full jet data, or to that of the most important single jet data in the PDF fit, the CMS 7 TeV data set [3]. A weight of 19 (=305/16) would correspond to having a similar number of data points for the top data as for the entire jet data set, and can also be considered as corresponding to an effective decrease
in the statistical and systematic errors of the top data. As an additional comparison, the
impact of increasing the weight of the CMS \(t\bar{t}\) data to that of the largest impact jet data
set in CT14HERA2 (7 TeV CMS jet data with 133 data points) has also been considered,
using a weight of 8 (=133/16). To provide intermediate results, weights of 3 and 5 are also
considered. It should be stressed that increasing the weight is not exactly equivalent to an
increased luminosity, since there is no change in the central values of the data, i.e. the jitter
from the existing data due to limited statistics is preserved in the re-weighted data, reducing
somewhat its impact on the PDF fit. The results are shown in Tables III and IV for the
cases of the CT14HERA2 and CT14HERA2mJ PDFs, respectively.

TABLE III: The \(\chi^2\) for each 2D \(t\bar{t}\) data set, calculated with the original global-fit CT14HERA2
PDFs (i.e., \(w = 0\)) and ePump updated CT14HERA2 PDFs with different weights.

| ID | data | dof | \(\chi^2\) |
|----|------|-----|---------|
|    |      |     | \(w = 0\) | \(w = 1\) | \(w = 3\) | \(w = 5\) | \(w = 8\) | \(w = 19\) |
| 573 | \(d^2\sigma / dy_t dp_T^t\) | 15  | 35.5    | 34.9    | 33.9    | 33.1    | 32.2    | 30.0    |
| 574 | \(d^2\sigma / dm_{t\bar{t}} dp_T^{t\bar{t}}\) | 15  | 82.3    | 80.6    | 77.8    | 75.6    | 73.1    | 68.1    |
| 575 | \(d^2\sigma / dm_{t\bar{t}} d\Delta \eta_{t\bar{t}}\) | 11  | 22.1    | 22.0    | 21.8    | 21.6    | 21.3    | 20.3    |
| 576 | \(d^2\sigma / dm_{t\bar{t}} dy_t\) | 15  | 20.2    | 20.1    | 20.0    | 19.9    | 19.8    | 19.5    |
| 577 | \(d^2\sigma / dm_{t\bar{t}} dy_{t\bar{t}}\) | 15  | 23.8    | 23.5    | 23.1    | 22.9    | 22.6    | 22.1    |

First, note that the starting \(\chi^2\) values are larger for CT14HERA2mJ than for
CT14HERA2, especially for data sets 573 and 577. As will be shown later, this is because
the gluon distribution the double-differential top data prefer is closer to that of CT14HERA2
than CT14HERA2mJ. For the CT14HERA2 fit, there is a slow decrease in \(\chi^2\) as the addi-
tional weight increases, due to the constraining influence of the jet data. There is a faster
decrease for the CT14HERA2mJ fit. However, the exact magnitude of the decrease in \(\chi^2\)
should not be taken too seriously, since increasing the weight does not change the central
values of the data, i.e. the jitter from the existing data due to limited statistics is preserved
in the re-weighted data. Nevertheless, Tables III and IV tell us that the data sets 573 and
577 have better agreement with the theory calculations and other data sets employed in the
CT14HERA2mJ and CTEQ14HERA2 global analyses. What we would like to focus on here
is the impact to the gluon-PDF error bands when the weight of the \(t\bar{t}\) data set is increased in
TABLE IV: The $\chi^2$ for each 2D $t\bar{t}$ data set, calculated with the original global-fit CT14HERA2mJ PDFs (i.e., $w = 0$) and ePump updated CT14HERA2mJ PDFs with weight 1 or corresponding higher weights. Note that weight 8 and 19 are also applied to data set 575, although it contains 12, not 16, data points. This is because these 12 data points were also constructed out of the same data as the other data sets and thus they are scaled by the same factor.

| ID | data | dof | $\chi^2$ |
|----|------|-----|----------|
|    |      |     | $w = 0$ | $w = 1$ | $w = 3$ | $w = 5$ | $w = 8$ | $w = 19$ |
| 573 | $d^2\sigma/dy_t dp_T^t$ | 15  | 50.9    | 41.1    | 33.5    | 30.1    | 27.3    | 22.3    |
| 574 | $d^2\sigma/dm_{t\bar{t}} dp_T^{t\bar{t}}$ | 15  | 70.9    | 69.0    | 66.7    | 65.5    | 64.4    | 63.1    |
| 575 | $d^2\sigma/dm_{t\bar{t}} d\Delta_{t\bar{t}}$ | 11  | 26.4    | 25.9    | 25.0    | 24.2    | 23.2    | 20.4    |
| 576 | $d^2\sigma/dm_{t\bar{t}} dy_t$ | 15  | 27.6    | 25.5    | 23.1    | 21.9    | 20.9    | 19.7    |
| 577 | $d^2\sigma/dm_{t\bar{t}} d\Delta_{t\bar{t}}$ | 15  | 45.3    | 34.7    | 27.5    | 25.1    | 23.6    | 22.3    |

the PDF updating. We observe that the inclusion of data sets 573 and 577 show a noticeable improvement in $\chi^2$ when included in the CT14HERA2mJ fit, but not in the CT14HERA2 fit. They show further improvement for CT14HERA2mJ on the use of higher weights.

We now consider the impact on the gluon distribution, first considering the weight of 19, again weighting an individual double-differential $t\bar{t}$ data set to have the equivalence of the total jet data in CT14HERA2. The ePump updated gluon PDFs with this weight are shown in Figs. 7-11, where it can be seen that the $t\bar{t}$ data has a similar constraint on the gluon PDFs as does the jet data (included in the CT14HERA2 fit), both for the value of the central PDF and the size of the error band. The central gluon distribution that is thus obtained does not always agree with that obtained using the jet data (CT14HERA2), but the error bands are all of similar size. Note in particular that data set 573 prefers a somewhat stronger gluon at moderate $x$, and data set 574 prefers a stronger gluon at high $x$ than do the other data sets. As a result the gluon in the region sensitive to gluon-gluon Higgs boson production ($x=0.01$) is larger in the case of 573 and smaller in the case of 574. Furthermore, we would like to note that due to the different composition of hard scattering processes contributing to the production of $t\bar{t}$ and jet productions at the LHC, the weighted

\footnote{In fact, data set 574 is the only $t\bar{t}$ data set that prefers a significantly stronger gluon at high $x$}
$t\bar{t}$ data provide a slightly narrower gluon-PDF error band for $x$ around 0.3, as compared to the jet data, cf. Figs. 7 and 8 when using CT14HERA2mJ global fit as the base for PDF updating.

**FIG. 7:** The updated gluon PDF and its error band when adding 573 (which is differential in $y_t$ and $p_T$), using ePump with weight=19, compared to CT14HERA2mJ and CT14HERA2e. Hereafter, CT14HERA2e is obtained by adding jet data back to CT14HERA2mJ using ePump, which is very similar to the CT14HERA2 PDF set [3]. Left: PDF central values. Right: Error bands.

**FIG. 8:** The updated gluon PDF and its error band when adding 574 (which is differential in $m_{t\bar{t}}$ and $p_T$), using ePump with weight 19, compared to CT14HERA2mJ and CT14HERA2e. Left: PDF central values. Right: Error bands.

We have also examined the impact of a smaller weight, 8, which corresponds to having a similar number of data points for the $t\bar{t}$ data as for the single strongest jet data set included in the CT14HERA2, that of the CMS 7 TeV inclusive jet cross section. In Figs. 12-16 we show the results of ePump updated gluon PDFs when each $t\bar{t}$ data set is added to CT14HERA2mJ
FIG. 9: The updated gluon PDF and its error band when adding 575 (which is differential in $m_{t\bar{t}}$ and $\Delta \eta_{t\bar{t}}$), using ePump with weight 19, to CT14HERA2mJ and CT14HERA2e. Left: PDF central values. Right: Error bands.

FIG. 10: The updated gluon PDF and its error band when adding 576 (which is differential in $m_{t\bar{t}}$ and $y_t$), using ePump with weight 19, compared to CT14HERA2mJ and CT14HERA2e. Left: PDF central values. Right: Error bands.

with weight 8. We find that the $t\bar{t}$ data have similar size effects with weight 8, as observed with weight 19. This is true especially for data set 577 ($m_{t\bar{t}} - y_t$), where we find almost the same impact on the gluon PDF as for jets, except that jet data lead to a smaller error band in the $x$ range between 0.1 to 0.2, where the correlations were observed to be weaker. There may be a type of saturation of impact that takes place as the weight is increased. If so, then the relative impact of the $t\bar{t}$ data may increase faster than expected by the use of a simple ratio of data points as has been used here.

It is then useful to examine the impact of even smaller weights, 3 and 5, in Figs. 17 and 18.
FIG. 11: The updated gluon PDF and its error band when adding data set 577 (which is differential in $m_{t\bar{t}}$ and $y_{t\bar{t}}$), using ePump with weight 19, compared to CT14HERA2mJ and CT14HERA2e. Left: PDF central values. Right: Error bands.

FIG. 12: The updated gluon PDF and its error band when adding 573 (which is differential in $y_{t}$ and $p_{T}^{t}$), using ePump with weight 8, compared to CT14HERA2mJ and CT14HERA2e. Left: PDF central values. Right: Error bands.

for data set 577. Here we see, as expected, intermediate results, which nonetheless indicate that even moderate increases in integrated luminosity samples could potentially lead to a noticeable impact of the double-differential top data.

An alternative way of displaying the impact of a new data set on the resulting PDF distributions is to examine the length of the shift vector $d^0$, of the best-fit position in PDF parameter space, from the original set of parameters for CT14HERA2 to those preferred by the fit with the inclusion of the new data set. The vector $d^0$ is 27-dimensional, corresponding to the number of free parameters in the CT14HERA2 global PDF fit. A value of $d^0$ of the
FIG. 13: The updated gluon PDF and its error band when adding 574 (which is differential in \(m_{t\bar{t}}\) and \(p_T^{\ell}\)), using ePump with weight 8, compared to CT14HERA2mJ and CT14HERA2e. Left: PDF central values. Right: Error bands.

FIG. 14: The updated gluon PDF and its error band when adding 575 (which is differential in \(m_{t\bar{t}}\) and \(\Delta \eta_{t\bar{t}}\)), using ePump with weight=8, compared to CT14HERA2mJ and CT14HERA2e. We still use weight 8 instead of \(133/12=11\) because these 12 data points were also constructed out of the same \(t\bar{t}\) data. Left: PDF central values. Right: Error bands.

order of 1, indicates that the new best-fit vector touches the 90% CL boundary, i.e. there is a very large impact (change) from this new data set, while a value of \(d^0\) smaller than 0.1 would imply no large change of the PDFs. In Table VI, the values of \(d^0\) are shown for the results of including data sets 573-577 using CT14HERA2 with weights of 1, 3, 5, 8 or 19. In Table V, the values of \(d^0\) are shown for the results of including data sets 573-577 using CT14HERA2mJ with weights of 1, 3, 5, 8 or 19. The values of \(d^0\) increase with weight, as expected. The impact is greater with CT14HERA2mJ than with CT14HERA2. The largest
FIG. 15: The updated gluon PDF and its error band when adding 576 (which is differential in $m_{t\bar{t}}$ and $y_t$), using ePump with weight 8, compared to CT14HERA2mJ and CT14HERA2e. Left: PDF central values. Right: Error bands.

FIG. 16: The updated gluon PDF and its error band when adding data set 577 (which is differential in $m_{t\bar{t}}$ and $y_{t\bar{t}}$), using ePump with weight 8, compared to CT14HERA2mJ and CT14HERA2e. Left: PDF central values. Right: Error bands.

The values of $d^0$ results from data sets 573 and 574 for CT14HERA2 and data sets 573 and 577 for CT14HERA2mJ. Note that a large value of $d^0$ results from the pull of that data set away from the gluon PDF obtained in CT14HERA2mJ. The smallest values of $d^0$ are from those data sets that lead to the smallest apparent differences between either CT14HERA2 or CT14HERA2mJ. Note that a small value of $d^0$ could also result from a data set that provided strong constraints, but already agreed with the predictions using that PDF.
FIG. 17: The updated gluon PDF and its error band when adding data set 577 (which is differential in $m_{t\bar{t}}$ and $y_{t\bar{t}}$), using ePump with weight=3, compared to CT14HERA2mJ and CT14HERA2e. Left: PDF central values. Right: Error bands.

FIG. 18: The updated gluon PDF and its error band when adding data set 577 (which is differential in $m_{t\bar{t}}$ and $y_{t\bar{t}}$), using ePump with weight=5, compared to CT14HERA2mJ and CT14HERA2e. Left: PDF central values. Right: Error bands.

VI. DISCUSSION AND CONCLUSIONS

The LHC can be correctly characterized as a top factory. Precise measurements of the $t\bar{t}$ final state allows for a better understanding of the production mechanisms and in particular, can allow for a determination of the gluon distribution, especially at high $x$, where it is currently relatively unconstrained. The determination of the gluon distribution, and indeed of all of the PDFs, needs to take place in the context of a global PDF fit, which includes a wide variety of data, including top production. Up to now, only singly differential top measurements have been included in global PDF fits. Double-differential measurements
TABLE V: $t\bar{t}$ data list. Shift lengths of best-fit point when added to CT14HERA2 and CT14HERA2mJ, using ePump with various weights.

| ID | data                      | $d^0$ CT14HERA2 | $d^0$ CT14HERA2mJ |
|----|---------------------------|----------------|-------------------|
|    |                           | $w = 1$        | $w = 1$           | $w = 3$ | $w = 5$ | $w = 8$ | $w = 19$ |
| 573| $d^2\sigma/dy_tdp_T^t$    | 0.06           | 0.22              | 0.45    | 0.60    | 0.75    | 1.1      |
| 574| $d^2\sigma/dm_{tt}dp_T^{t\bar{t}}$ | 0.10          | 0.10              | 0.24    | 0.34    | 0.43    | 0.61     |
| 575| $d^2\sigma/dm_{tt}d\Delta\eta_{tt}$ | 0.03          | 0.05              | 0.15    | 0.24    | 0.36    | 0.73     |
| 576| $d^2\sigma/dm_{tt}dy_T$   | 0.02           | 0.10              | 0.25    | 0.34    | 0.43    | 0.59     |
| 577| $d^2\sigma/dm_{tt}dy_{T\bar{t}}$ | 0.04          | 0.22              | 0.43    | 0.54    | 0.62    | 0.75     |

TABLE VI: $t\bar{t}$ data list. Shift lengths of best-fit point when added to CT14HERA2 using ePump with various weights.

| ID | data                      | $d^0$ CT14HERA2 |
|----|---------------------------|----------------|
|    |                           | $w = 1$        | $w = 3$ | $w = 5$ | $w = 8$ | $w = 19$ |
| 573| $d^2\sigma/dy_tdp_T^t$    | 0.06           | 0.16    | 0.25    | 0.37    | 0.73    |
| 574| $d^2\sigma/dm_{tt}dp_T^{t\bar{t}}$ | 0.10          | 0.28    | 0.43    | 0.61    | 1.06    |
| 575| $d^2\sigma/dm_{tt}d\Delta\eta_{tt}$ | 0.03          | 0.08    | 0.13    | 0.20    | 0.44    |
| 576| $d^2\sigma/dm_{tt}dy_T$   | 0.02           | 0.06    | 0.09    | 0.13    | 0.26    |
| 577| $d^2\sigma/dm_{tt}dy_{T\bar{t}}$ | 0.04          | 0.10    | 0.15    | 0.21    | 0.35    |

have the potential of providing more detailed information on the gluon distribution. With the double-differential measurements taken by CMS, and the recent calculation of these observables to NNLO, it is now possible to use the double-differential data in a global PDF fit at NNLO, the order needed for precision determinations.

Including the CMS double-differential top data with the nominal weight of one does not greatly impact the gluon distribution due to the greater influence of the inclusive jet data. A more sizeable impact is observed in the fit when the jet data is removed. We have seen that applying a weight factor of 19 for the CMS double-differential $t\bar{t}$ data leads to a similar constraining power on the gluon distribution function as the jet data included in the CT14HERA2 global PDF fit. However, an almost equivalent constraining power can be
reached using a lower weight value of 8. Such a sample is effectively present in the current 13 TeV data taken in Run 2 (especially allowing for the impact of the increased center-of-mass energy). However, the LHC jet data will also increase proportionately. Even now, the 8 TeV CMS jet data set is more constraining than the 7 TeV data set, as will be shown in the CT18 paper. It is not clear in such an enlarged set of data what the relative influences of the top and inclusive jet data would be, but a greater integrated luminosity may have a larger impact on the top data as compared to the jet data, both in terms of the relative statistical and the relative systematic errors. Furthermore, due to the different composition of hard scattering processes contributing to the production of $t\bar{t}$ and jet productions at the LHC, precision $t\bar{t}$ data may constrain gluon-PDF error band in somewhat different (large) $x$ regions as compared to jet data, cf. Figs. 7 and 8. In addition, it may be possible to combine more than one double-differential set of observables, if the statistical correlations are taken into account, further strengthening the impact of the data.

Acknowledgments

A.M. thanks the Department of Physics at Princeton University for hospitality during the completion of this work. The work of M.C. was supported by the Deutsche Forschungsgemeinschaft under grant 396021762 - TRR 257 and by a grant from BMBF. The work of A.M. and A.P. has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 683211); it was also supported by the UK STFC grants ST/L002760/1 and ST/K004883/1. A.P. is a cross-disciplinary post-doctoral fellow supported by funding from the University of Edinburgh and Medical Research Council (core grant to the MRC Institute of Genetics and Molecular Medicine). The work of S.D. and I.S. was supported by the National Natural Science Foundation of China under the Grant No. 11965020 and No. 11847160. We thank our CTEQ-TEA colleagues for support and discussions. This work was supported by the U.S. National Science Foundation under Grants No. PHY-1719914 and PHY-1707812. C.-P. Yuan is also grateful for the support from the Wu-Ki Tung endowed chair in particle physics.
[1] G. Aad et al. [ATLAS Collaboration], Eur. Phys. J. C 76, no. 10, 538 (2016) [arXiv:1511.04716 [hep-ex]].

[2] V. Khachatryan et al. [CMS Collaboration], Eur. Phys. J. C 75, no. 11, 542 (2015) [arXiv:1505.04480 [hep-ex]].

[3] T. J. Hou, Z. Yu, S. Dulat, C. Schmidt and C.-P. Yuan, Phys. Rev. D 100, 114024 (2019) [arXiv:1907.12177 [hep-ph]].

[4] A. M. Sirunyan et al. [CMS Collaboration], Eur. Phys. J. C 77, no. 7, 459 (2017) [arXiv:1703.01630 [hep-ex]].

[5] C. Schmidt, J. Pumplin, C. P. Yuan and P. Yuan, Phys. Rev. D 98, no. 9, 094005 (2018) [arXiv:1806.07950 [hep-ph]].

[6] M. Czakon, D. Heymes and A. Mitov, JHEP 1704, 071 (2017) [arXiv:1606.03350 [hep-ph]].

[7] M. Czakon, A. Ferrogli, D. Heymes, A. Mitov, B. D. Pecjak, D. J. Scott, X. Wang and L. L. Yang, JHEP 1805, 149 (2018) [arXiv:1803.07623 [hep-ph]].

[8] T. Kluge, K. Rabbertz and M. Wobisch, [hep-ph/0609285].

[9] D. Britzger et al. [fastNLO Collaboration], [arXiv:1208.3641 [hep-ph]].

[10] M. Czakon, D. Heymes and A. Mitov, arXiv:1704.08551 [hep-ph].

[11] R. D. Ball et al. [NNPDF Collaboration], JHEP 1504, 040 (2015) [arXiv:1410.8849 [hep-ph]].

[12] S. Dulat et al., Phys. Rev. D 93, no. 3, 033006 (2016) [arXiv:1506.07443 [hep-ph]].

[13] M. Czakon, Phys. Lett. B 693, 259 (2010) [arXiv:1005.0274 [hep-ph]].

[14] M. Czakon, Nucl. Phys. B 849, 250 (2011) [arXiv:1101.0642 [hep-ph]].

[15] M. Czakon and D. Heymes, Nucl. Phys. B 890, 152 (2014) [arXiv:1408.2500 [hep-ph]].

[16] M. Czakon, P. Fiedler and A. Mitov, Phys. Rev. Lett. 115, no. 5, 052001 (2015) [arXiv:1411.3007 [hep-ph]].

[17] M. Czakon, D. Heymes and A. Mitov, Phys. Rev. Lett. 116, no. 8, 082003 (2016) [arXiv:1511.00549 [hep-ph]].

[18] M. Czakon, P. Fiedler, D. Heymes and A. Mitov, JHEP 1605, 034 (2016) [arXiv:1601.05375 [hep-ph]].

[19] A. Behring, M. Czakon, A. Mitov, A. S. Papanastasiou and R. Poncelet, Phys. Rev. Lett. 123, no. 8, 082001 (2019) [arXiv:1901.05407 [hep-ph]].
[20] M. Czakon, A. van Hameren, A. Mitov and R. Poncelet, JHEP 1910, 262 (2019) [arXiv:1907.12911 [hep-ph]].

[21] S. Catani, S. Devoto, M. Grazzini, S. Kallweit and J. Mazzitelli, JHEP 1907, 100 (2019) [arXiv:1906.06535 [hep-ph]].

[22] J. Campbell, J. Huston and F. Krauss, “The Black Book of Quantum Chromodynamics : A Primer for the LHC Era”. OXFORD UNIVERSITY PRESS, ISBN 978-0-19-965274-7, First edition published in 2018.

[23] B. T. Wang, T. J. Hobbs, S. Doyle, J. Gao, T. J. Hou, P. M. Nadolsky and F. I. Olness, Phys. Rev. D 98, no. 9, 094030 (2018) [arXiv:1803.02777 [hep-ph]].