Exercise/Lesson #2

Scientific Data Analysis Lab course

Alexis Pompili - UniBA

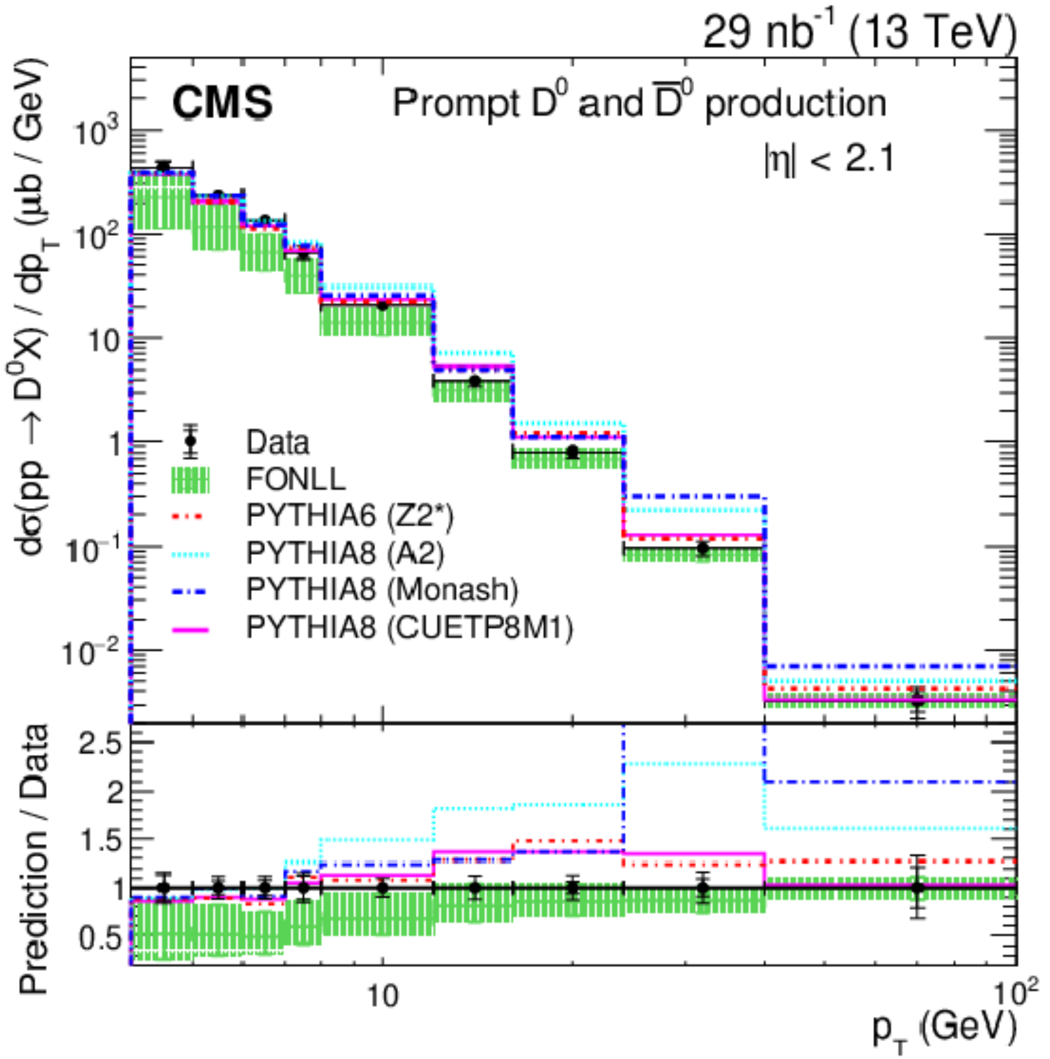
PART 1 : Differential production cross section for prompt D^0 (and $\bar{D}^0$) mesons with CMS data

PART 2 : Comparison of ALICE and CMS data for this production cross section: study of a possible discrepancy

PART 1 : Differential production cross section for prompt D^0 (and $\overline{D}^0$) mesons with CMS data

In this first part of the exercise we want to reproduce this typical differential production cross section plot as example to learn the use of **`TGraphErrors`** and **`TGraphAsymErrors`**.

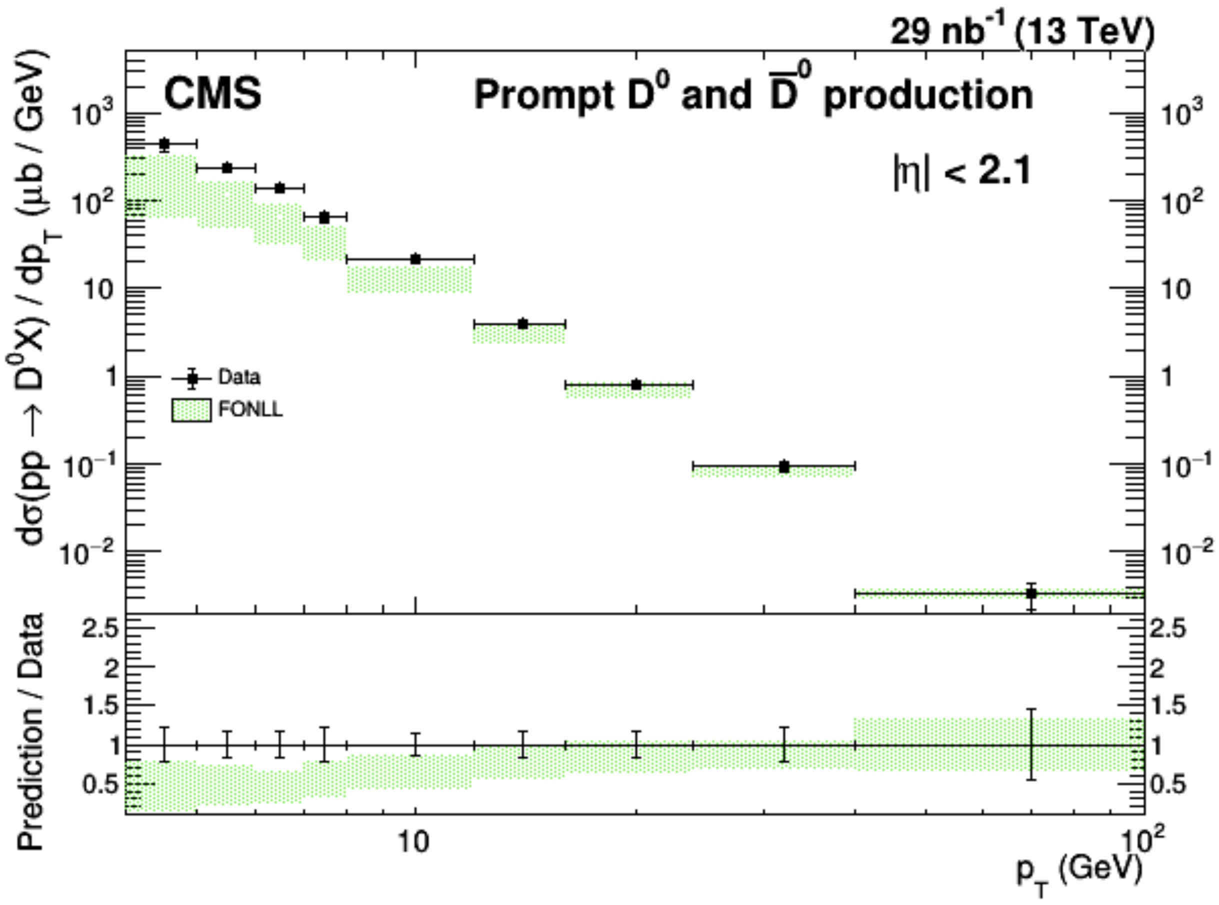
Specifically this plot is in upper Fig.5 of https://arxiv.org/abs/2107.01476 published in JHEP 11 (2021) 225.



We will reproduce the plot doing a comparison only with FONLL theoretical predictions (not PYTHIA)
Also we will sum in quadrature statistical and systematical uncertainties in each bin.

```
[[pompili@pompilic7 Esercitazione-2]$ root -l
[root [0] .x cms_D0_xsec.C
 Info in <TCanvas::Print>: file /home/pompili/SDAL-2022/Esercitazione-2/CMS_D0_xsec.png has been created
 root [1]
```

We obtain this plot….



… with the following code (file **cms_D0_xsec.C**)

```
#include <TH1.h>
#include <TF1.h>
#include <TF2.h>
#include <TStyle.h>
#include <TCanvas.h>
#include <TGraph.h>
#include <TGraphErrors.h>
#include <TString.h>
#include <TLine.h>
#include "TPad.h"
//
//
void cms_D0_xsec() {
  //
  gROOT->Reset();
  gROOT->Clear();
  gStyle->SetOptLogx();
  //
  ////////////////////////////////////////////////////////////////////
  //
  // -- 9 bins therefore 10 limiting numbers (inf and sup extremes)
  //
  Float_t vpt[10] = {4.,5.,6.,7.,8.,12.,16.,24.,40.,100.};
  Float_t vptm[9];
  Float_t vpte[9];
  //
  // We need to identify the average positions and half-widths of the bins
  //
  for(Int_t i=0;i<9;i++){ vptm[i]=(vpt[i+1]+vpt[i])/2; }
  //
  // Vanno individuate le semilarghezze dei bin
  for(Int_t j=0;j<9;j++){ vpte[j]=(vpt[j+1]-vpt[j])/2; }
  //
  // From Tab.5 - column D0+d0bar (in the CMS paper: https://arxiv.org/pdf/2107.01476.pdf)
  // CMS - Intervalli 4-5, 5-6, 6-7, 7-8, 8-12, 12-16, 16-24, 24-40, 40-100
  //
  // CMS - Incertezza totale (stat+syst) - simmetrica
  Float_t vD0[9] = {430.24,230.12,135.84,65.71,20.97,3.93,0.812,0.097,0.0033};
  Float_t vD0_err[9] = {69.3987,26.403,16.1875,10.1786,2.0494,0.4767,0.09982,0.015399,0.001091};
  //
  TGraphErrors *gr_data = new TGraphErrors(9, vptm, vD0, vpte, vD0_err);
  //
  // FONLL - ha incertezza asimmetrica
  Float_t FONLL[9] = {220.0,117.4,66.0,38.96,14.125,3.175,0.6925,0.08375,0.00326667};
  Float_t FONLL_err_min[9] = {157.2,71.2,35.0,18.38,5.52,0.915,0.155,0.01375,0.0004};
  Float_t FONLL_err_plus[9] = {107.0,47.4,22.6,11.42,3.545,0.655,0.1225,0.0125,0.00037667};
  //
  TGraphAsymmErrors *gr_theo = new TGraphAsymmErrors(9, vptm, FONLL, vpte, vpte, FONLL_err_min, FONLL_err_plus);
  //
  //-- note the syntax : TGraphAsymmErrors(n,x,y,exl,exh,eyl,eyh); l=low=min & h=high=plus
  //
  ////////////////////////////////////////////////////////////////////
```

```
//////////////////////////////////////////////////////////////////////////
//
//======================== CMS vs Theory (FONLL) : make the ratio
//
Float_t ratio_theo_over_data[9];
for (Int_t k=0; k<9; k++)
   {
    ratio_theo_over_data[k] = FONLL[k] / vD0[k];
   }
//
//--theoretical predictions and these data are not correlated: if ratio is R=a/b then the formula for the error propagation is:
//                                                             (e_R)^2 = R^2 * [ (e_a)^2/a^2 + (e_b)^2/b^2 ] , namely:
//                                                             e_R = R * sqrt( (e_a)^2/a^2 + (e_b)^2/b^2 )
//
Float_t ratio_theo_over_data_err_min[9];
for (Int_t h=0; h<9; h++)
   {
    ratio_theo_over_data_err_min[h] = ratio_theo_over_data[h] * sqrt( ((FONLL_err_min[h]*FONLL_err_min[h]) / (FONLL[h] * FONLL[h])) + ((vD0_err[h] * vD0_err[h]) / (vD0[h] * vD0[h])) );
   }
//
Float_t ratio_theo_over_data_err_plus[9];
for (Int_t l=0; l<9; l++)
   {
    ratio_theo_over_data_err_plus[l] = ratio_theo_over_data[l] * sqrt( ((FONLL_err_plus[l]*FONLL_err_plus[l]) / (FONLL[l] * FONLL[l])) + ((vD0_err[l] * vD0_err[l]) / (vD0[l] * vD0[l])) );
   }
//
TGraphAsymmErrors *gr_ratio_theo_over_data = new TGraphAsymmErrors(9, vptm, ratio_theo_over_data, vpte, vpte, ratio_theo_over_data_err_min, ratio_theo_over_data_err_plus); // asymmetry on y-err only
//
/////////////////////////////////////////////////////////////////////////////
//
Float_t ratio_data_over_data[9];
for (Int_t y=0; y<9; y++)
   {
    ratio_data_over_data[y] = vD0[y]/vD0[y];
   }
//
Float_t ratio_data_over_data_err[9];
//
for (Int_t z=0; z<9; z++)
   {
    ratio_data_over_data_err[z] = ratio_data_over_data[z] * sqrt( ((vD0_err[z]*vD0_err[z]) / (vD0[z]*vD0[z])) + ((vD0_err[z]*vD0_err[z]) / (vD0[z]*vD0[z])) );
   }
//
TGraphErrors *gr_selfratio = new TGraphErrors(9, vptm, ratio_data_over_data, vpte, ratio_data_over_data_err);
//
//////////////////////////////////////////////////////////////////////////////
```

```
////////////////////////////////////////////////////////////////////////////////
/////////////////////////////////-- PLOTTING --////////////////////////////////
//
TMultiGraph *mg = new TMultiGraph(); // https://root.cern.ch/doc/master/classTMultiGraph.html
//
mg->Add(gr_theo,"2p");
mg->Add(gr_data,"EP");
//
gr_data->SetMarkerColor(1);
gr_data->SetMarkerStyle(20);
gr_data->SetMarkerSize(0.75);
//
gr_theo->SetFillColor(kGreen);
gr_theo->SetFillStyle(3002);
gr_theo->SetMarkerColor(0);
//
gr_ratio_theo_over_data->SetFillColor(kGreen);
gr_ratio_theo_over_data->SetFillStyle(3002);
gr_ratio_theo_over_data->SetMarkerColor(0);
gr_ratio_theo_over_data->SetTitle("");
gr_ratio_theo_over_data->GetXaxis()->SetTitle("p_{T} (GeV)");
//
gr_ratio_theo_over_data->GetYaxis()->SetTitle("Prediction / Data");
gr_ratio_theo_over_data->GetXaxis()->SetTitleSize(0.12);
gr_ratio_theo_over_data->GetYaxis()->SetTitleSize(0.12);
gr_ratio_theo_over_data->GetXaxis()->SetLabelSize(0.11);
gr_ratio_theo_over_data->GetYaxis()->SetLabelSize(0.1);
gr_ratio_theo_over_data->GetXaxis()->SetTitleOffset(1.1);
gr_ratio_theo_over_data->GetYaxis()->SetTitleOffset(0.39);
//
gr_ratio_theo_over_data->SetMinimum(0.1);
gr_ratio_theo_over_data->SetMaximum(2.7);
gr_ratio_theo_over_data->GetXaxis()->SetLimits(4, 100);
//
mg->SetTitle("");
mg->GetXaxis()->SetLabelSize(0); // this eliminates the x-labels for mg i.e. p1
mg->GetXaxis()->SetTitle("");
mg->GetYaxis()->SetTitleSize(0.062);
mg->GetYaxis()->SetTitleOffset(0.72);
mg->GetYaxis()->SetLabelSize(0.05);
mg->GetYaxis()->SetTitle("d#sigma(pp #rightarrow D^{0}X) / dp_{T} (#mub / GeV)");
mg->SetMinimum(0.002);
mg->SetMaximum(5000.);
mg->GetXaxis()->SetLimits(4, 100);
//
//
TCanvas *MyC = new TCanvas("MyC","CMS");
//
```

```
    TCanvas *MyC = new TCanvas("MyC","CMS");
    //
    auto p1 = new TPad{"p1", "", 0.05, 0.3, 0.95, 0.95};
    p1->SetBottomMargin(0); // --> eliminate the white margin arund the pad
    p1->SetTickx(2);
    p1->SetTicky(2);  // --> this makes labels and scale values on both sides
    auto p2 = new TPad{"p2", "Ratio", 0.05, 0, 0.95, 0.3};
    p2->SetTopMargin(0);
    p2->SetBottomMargin(0.3);
    //p2->SetLeftMargin(0.1);
    p2->SetTicky(2); // --> this makes labels and scale values on both sides
    p1->Draw();
    p2->Draw();
    //
    p1->cd();
    p1->SetLogy();
    //------------------ Drawing the Tgraph previously added at Multigraph
    mg->Draw("a");
    //
    auto leg = new TLegend{0.13, 0.3, 0.3, 0.4}; // xmin, ymin, xmax, ymax with respect to pad's corners
      leg->AddEntry(gr_data, "Data", "LPE");
      //
      leg->SetBorderSize(0); // to get te rectangular area without border
      leg->AddEntry(gr_theo, "FONLL", "F"); // option F makes the rectangular area
      leg->Draw();

      // Scritta sezione d'urto
      TLatex latex{};
      latex.SetTextSize(0.063);
      latex.DrawLatex(45, 5800,"29 nb^{-1} (13 TeV)");
      latex.SetTextSize(0.075);
      latex.DrawLatex(4.5, 1060, "CMS");
      latex.SetTextSize(0.07);
      latex.DrawLatex(12, 1050, "Prompt D^{0} and #bar{D}^{0} production");
      latex.DrawLatex(45, 150, "|#eta| < 2.1"); // report the fiducial region in eta
    //
    p2->cd();
    //------------------ plot the ratios
    gr_ratio_theo_over_data->Draw("a2p");
    gr_selfratio->Draw("psame");
    //
    //
    //
    /////////////
    //
    MyC->SaveAs("/home/pompili/SDAL-2022/Esercitazione-2/CMS_D0_xsec.png");
    //
    delete gr_data;
    delete gr_theo;
    delete gr_ratio_theo_over_data;
    delete mg;
    delete gr_selfratio;
    //
    delete leg;
    delete p1;
    delete p2;
    delete MyC;
    //
}
```

Part 1.2 Comparison with ALICE results

In the CMS paper JHEP 11 (2021) 225 there is a comparison with an older ALICE result at 7TeV [EPJ C 77 (2017) 550]

Figure 8 shows the comparison with the ALICE results [8, 9] for the D^{*+}, D^0, and D^+ cross sections at $\sqrt{s} = 7$ TeV in the range $1 < p_T < 24$ GeV ($0 < p_T < 36$ GeV for the D^0) and for the rapidity region $|y| < 0.5$. Between the two ALICE measurements, the more recent one [9] is chosen for the comparison. It should be noted that the cross section definition by ALICE includes a factor of 1/2 that accounts for the fact that the measured yields include particles and antiparticles, while the cross sections are given for particles only. The same is true for the corresponding FONLL predictions, as well. To provide a relevant comparison, the CMS measurements are given for $p_T < 24$ GeV ($p_T < 40$ GeV for the D^0), for a better comparison with the ALICE points. Both sets of results are consistent with the respective FONLL predictions and close to their upper edge, as shown in the lower two panels.

(Fig.8, upper/right/b)

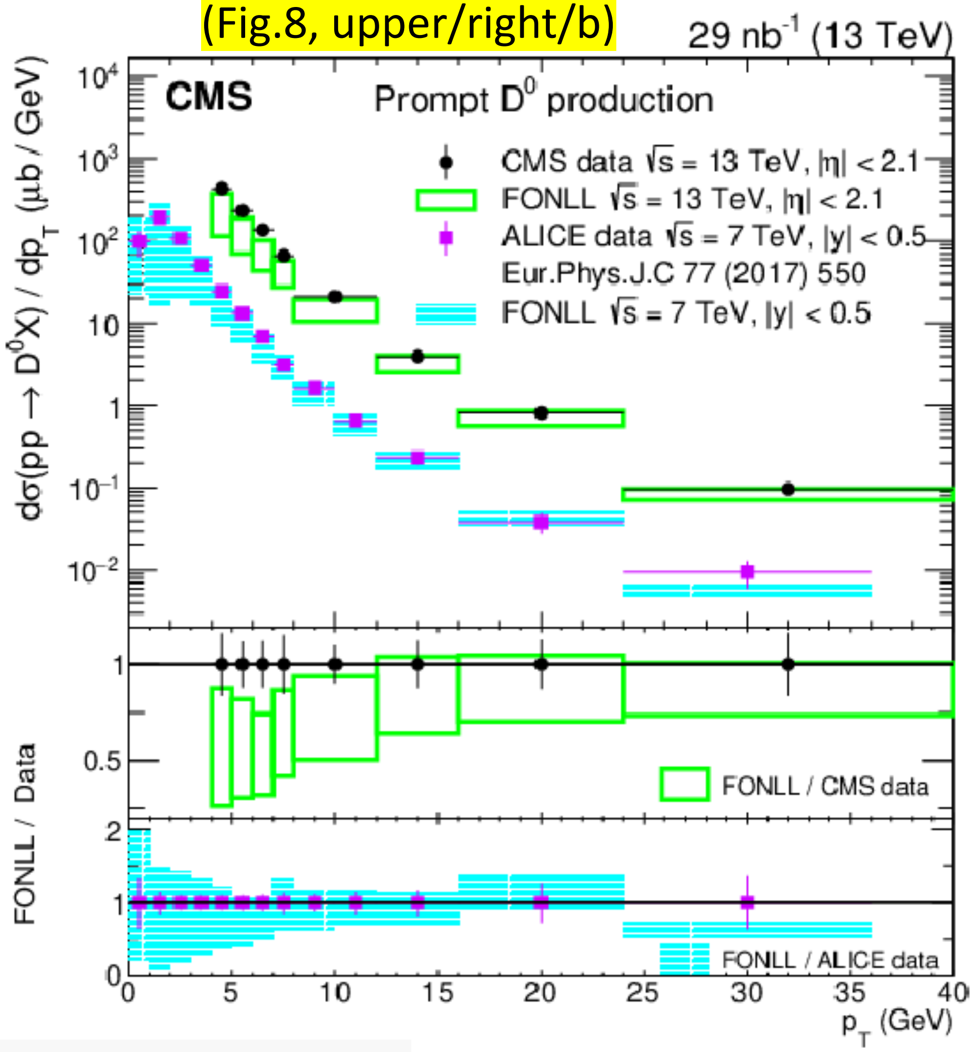
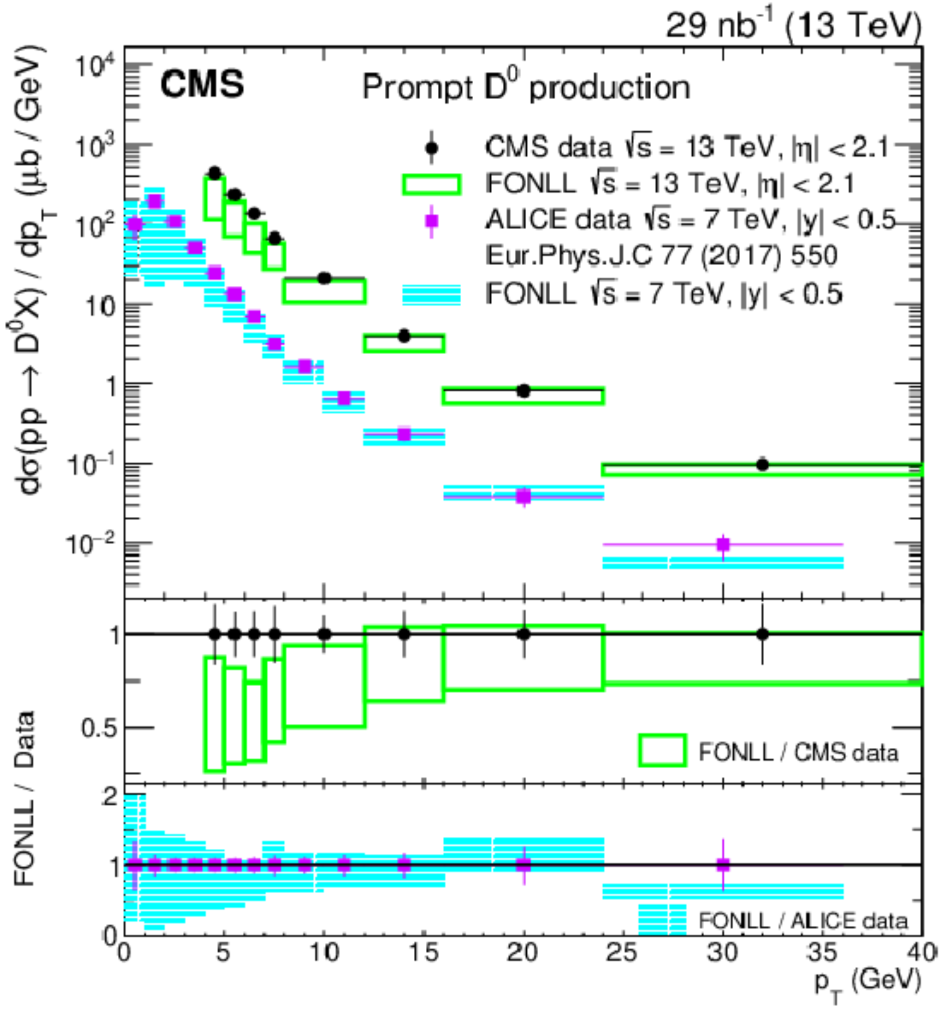


Figure 8-b: Differential cross section $d\sigma/dp_T$ for prompt $D^0 + \bar{D}^0$ meson production with $p_T < 24$ GeV from CMS (black circles, this paper) at $\sqrt{s} = 13$ TeV and ALICE [7] (magenta squares) at $\sqrt{s} = 7$ TeV and $|y| < 0.5$. The corresponding predictions from FONLL are shown by the unfilled and filled boxes, respectively. The cross section definition by ALICE includes a factor of 1/2 that accounts for the fact that the measured yields include particles and antiparticles while the cross sections are given for particles only. The same is true for the corresponding FONLL predictions, as well. The vertical lines on the points give the total uncertainties in the data, and the horizontal lines show the bin widths. The two lower panels give the ratios of the FONLL predictions to the CMS and ALICE data, shown by circles and squares, respectively.



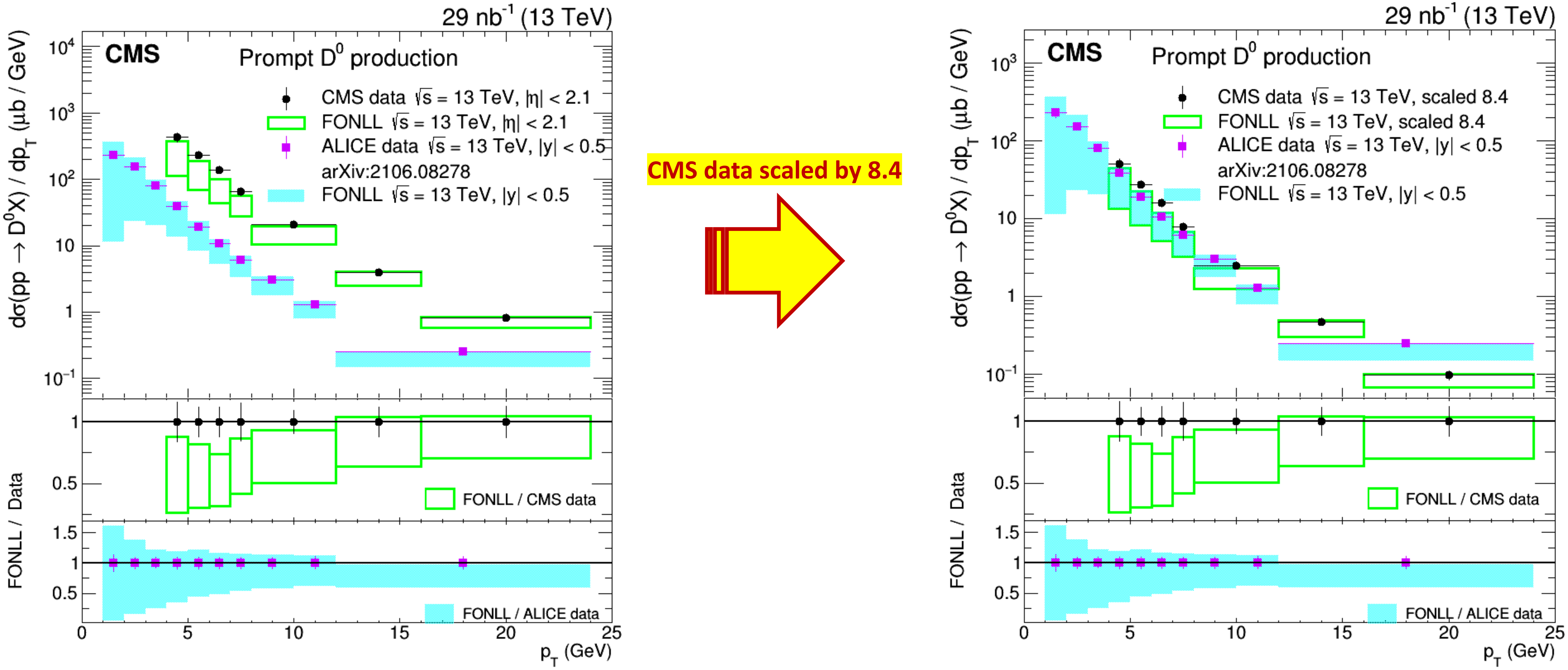
For an effective comparison the results have to be scaled (one to the other; for instance the CMS one to the ALICE one which is more central: the approx. scaling factor to apply is 8.4, to both data and FONLL predictions, in order to take into account the different (pseudo-)rapidity regions).

Further difference comes from the different center-of-mass energy (13TeV vs 7TeV)!

While CMS draft paper (arXiv:2107.01476) was being submitted also ALICE had very recently submitted a draft paper (arXiv:2106.08278) - later published as Phys. Rev. Lett. 128 (2022) 012001 - with the same measurement for D^0 mesons but at 13TeV.
At this point the two collaborations setup a comparison to check if results were compatible or not. This is discussed in the next slides.

PART 2 : Comparison of ALICE and CMS data for this production cross section: study of a possible discrepancy

For a visual comparison a flat scaling (factor 8.4) is applied at the beginning:



However to make a direct comparison we need to be rigorous as discussed in the following steps.
As an exercise we try to reproduce here with a good approximation the direct comparison that was done in 2021.

Step-0) Identify the bins where **both** experiments have results (with enough statistics per bin):
4-5, 5-6, 6-7, 7-8 , 8-12, 12-24GeV

Step-1) Flat scaling between $|\eta|$ < 2.1 and |y|< 0.5 is not fully correct since there is a (small) p_T dependence (derived from FONLL)!

The data can be ordered in this table (for simplicity asymmetric uncertainties are "symmetrized"):

p_T - interval	bin center and width		#CMS	scale factor	#ALICE
4-5	**4.5**	**0.5**	430.24 ± 69.40	7.81 ± 0.38	$38.64^{+3.39}_{-3.67}$
5-6	**5.5**	**0.5**	230.12 ± 26.40	7.83 ± 0.35	$18.91^{+1.72}_{-1.91}$
6-7	**6.5**	**0.5**	135.85 ± 16.19	7.82 ± 0.325	$10.67^{+1.00}_{-1.10}$
7-8	**7.5**	**0.5**	65.71 ± 10.18	7.81 ± 0.31	$6.13^{+0.58}_{-0.64}$
8-12	**10.0**	**2.0**	20.97 ± 2.05	7.76 ± 0.295	$2.155^{+0.20}_{-0.22}$
12-24	**18.0**	**6.0**	1.85 ± 0.23	7.60 ± 0.37	$0.25^{+0.03}_{-0.03}$

Step-2) Scale CMS data according to the formulas:

$$\#\text{CMS-scaled} = \frac{\#CMS}{SF} \qquad \sigma_{CMS-scaled} = \frac{\#CMS}{SF} \cdot \sqrt{\left(\frac{\sigma_{CMS}}{\#CMS}\right)^2 + \left(\frac{\sigma_{SF}}{SF}\right)^2}$$

Step-3) For simplicity simmetrize the ALICE error by simple arithmetic mean: $\sigma = \frac{\sigma_+ + \sigma_-}{2}$

Step-4) Two approaches are possible : the ***multiplicative*** and the ***additive***

The **multiplicative is not suitable** as discussed in the next slide**:**

considering the ALICE/CMS or CMS/ALICE ratios brings to a different results [*] for each bin (see example)

and the choice of one or of the other would be arbitrary!

Thus we prefer the additive approach for the comparison of the two experimental results! (see next-to-next slide)

[*] The fact that error propagation gives different results according to the ratio order is not surprising since the uncertainties are not really multiplicative !

To be explicit, by multiplicative approach we mean:

CMS : $C \pm \varepsilon_C$ includes the uncertainties : stat., syst., scaling error

ALICE : $A \pm \varepsilon_A$ includes the uncertianties : stat., syst., lumi (a posteriori).

Ratio ALICE/CMS : $R = \frac{A}{C}$ $\varepsilon_R = \sqrt{\left(\frac{\partial R}{\partial A}\right)^2 \varepsilon_A^2 + \left(\frac{\partial R}{\partial C}\right)^2 \varepsilon_C^2} = \cdots = R \cdot \sqrt{\frac{\varepsilon_A^2}{A^2} + \frac{\varepsilon_C^2}{C^2}}$

Ratio CMS/ALICE : $R' = \frac{C}{A}$ $\varepsilon_{R'} = \sqrt{\left(\frac{\partial R'}{\partial A}\right)^2 \varepsilon_A^2 + \left(\frac{\partial R'}{\partial C}\right)^2 \varepsilon_C^2} = \cdots = R' \cdot \sqrt{\frac{\varepsilon_A^2}{A^2} + \frac{\varepsilon_C^2}{C^2}}$

Sigma ALICE/CMS : $\sigma = \frac{|1-R|}{\varepsilon_R} = \frac{|1-R|}{R} \cdot \left(\frac{\varepsilon_A^2}{A^2} + \frac{\varepsilon_C^2}{C^2}\right)^{-\frac{1}{2}} = \left|\frac{1}{R} - 1\right| \cdot (\dots)^{-\frac{1}{2}}$

Sigma CMS/ALICE : $\sigma' = \frac{|1-R'|}{\varepsilon_{R'}} = \frac{|1-R'|}{R'} \cdot \left(\frac{\varepsilon_A^2}{A^2} + \frac{\varepsilon_C^2}{C^2}\right)^{-1/2} = \left|\frac{1}{R'} - 1\right| \cdot (\dots)^{-\frac{1}{2}} = |R - 1| \cdot (\dots)^{-\frac{1}{2}}$

Example: bin 6-7GeV :

$C = 17.40 \pm 2.20$

$A = 10.67 \pm 1.05$

$R = \frac{A}{C} \cong 0.61$

$\varepsilon_R \cong 0.61 \cdot 0.16 \cong 0.10$ (TOT err)

$R' = \frac{C}{A} \cong 1.63$

$\varepsilon_{R'} \cong 1.63 \cdot 0.16 \cong 0.26$ (TOT err ')

$\sigma = \frac{|1-0.61|}{0.10} \cong 3.9$

$\sigma' = \frac{|1-1.63|}{0.26} \cong 2.4$

Step-5) *Additive approach:*

Consider the difference **|#CMS_scaled-#ALICE|** to be compared with 1;
its uncertainty is simply:

$$\sigma_{diff} = \sqrt{\sigma^2_{\#CMS\_scaled} + \sigma^2_{\#ALICE}}$$

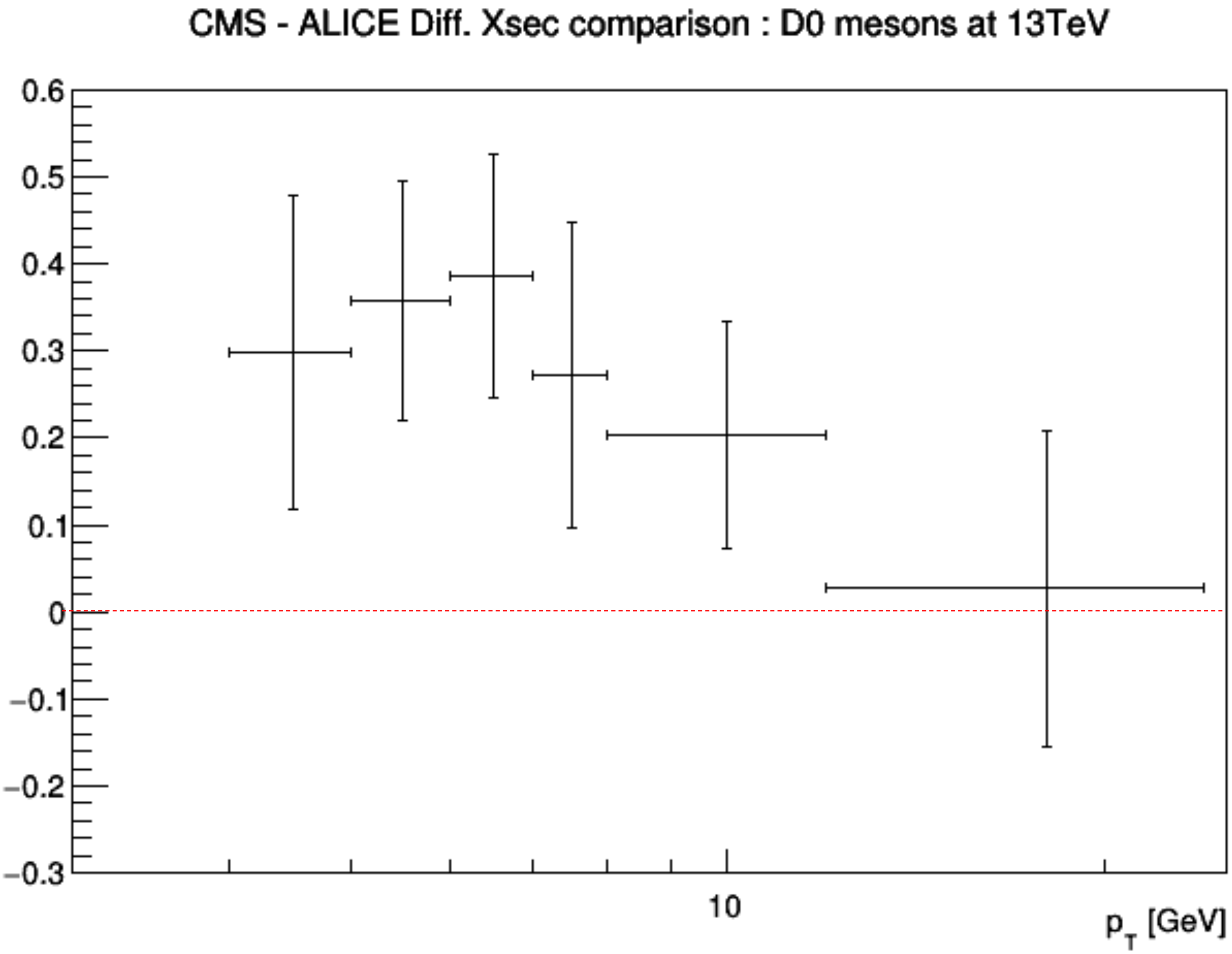
Deviation from 1 is obtaned by dividing the difference by the uncertainty and the descrepancy would be given in units of sigmas.

Alternatively one can consider the #CMS as ***normalization*** and compare the normalized difference with 0,
… considering the uncertainty as $\frac{\sigma_{diff}}{\#CMS\_scaled}$

$$\frac{|\#CMS\_scaled - \#ALICE|}{\#CMS\_scaled}$$ (thus also uncertanty is normalized)

Using this approach we obtain reasonable significance values of discrepancy!

It can be checked (as exercise) that these values lie in between the values obtained by the two ratios in each bin (considering |1-R| and comparing it to 0 with uncertainty σ_R for the two different ratios, ALICE-CMS & CMS-ALICE).


CMS - ALICE Diff. Xsec comparison : D0 mesons at 13TeV
0.6
0.5
0.4
0.3
0.2
0.1
0
−0.1
−0.2
−0.3
10
p_T [GeV]

Backup

Cinematica ad un collider adronico

Slide by C.Gemme

- Momento trasverso p_T (stima l'hardness del processo)
  - le particelle che sfuggono alla rivelazione hanno angolo piccolo
  - il momento si conserva "trasverso" $\Sigma p_T^i \sim 0$
- Momento longitudinale p_Z ed energia E
  - le particelle che sfuggono hanno p_z grande
  - P_z visibile (rivelato) non e' conservato
- Angolo polare θ non e' invariante, ma lo e' la rapidita' *y* e la pseudorapidita' η

$$y = \frac{1}{2}\ln\frac{E+p_Z}{E-p_Z} \approx \eta = -\ln\left(\tan\frac{\vartheta}{2}\right)$$

Se m~0

- ϕ come variabile azimutale