V+heavy flavor jets and constraints to PDFs in CMS

Juan Pablo Fernández Ramos (CIEMAT)
on behalf of the CMS Collaboration

EPS-HEP2018 @ Ghent
July 10-17, 2018

CMS-SMP-15-009

CMS-SMP-17-014

CMS-PAS-18-013

Introduction

- Measurements of σ(pp→Z/W+c) provide tests QCD predictions. **Results sensitive to hard scattering process & associated soft QCD radiation**

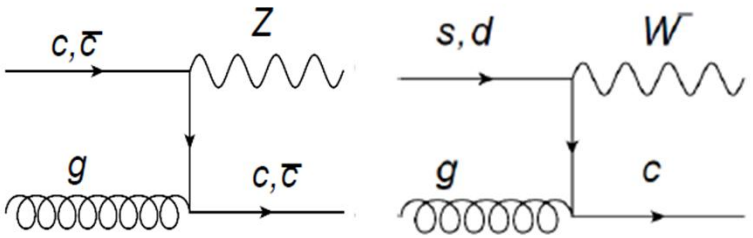
- Allows better understanding of proton structure. Z/W+c jets tests PDF for c,s respectively.

- Intrinsic Charm component inside the proton enhances σ(Z+c) @ ↑pt(Z).

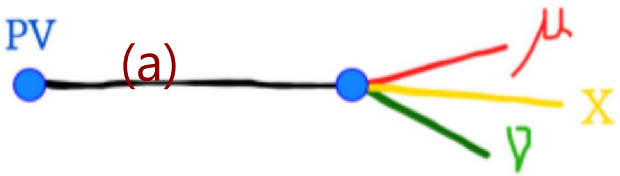
- Background to some SM processes and in searches

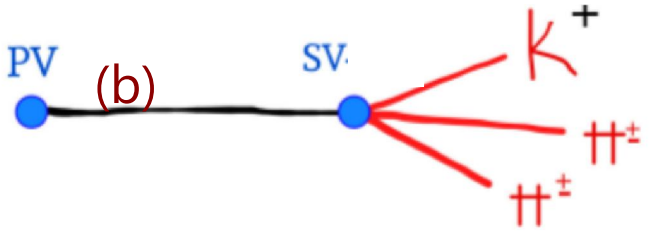
Analysis strategy:

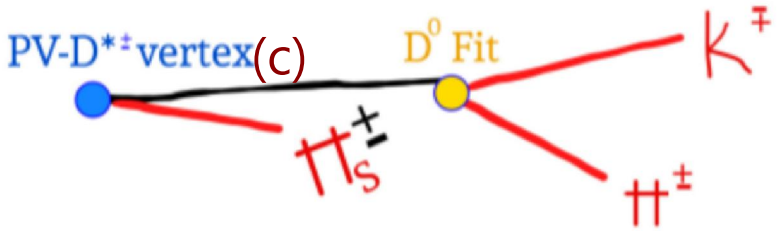
- Standard $Z \rightarrow l^+l^-$, $W^+ \rightarrow l^+\nu$ selection
- Isolated leptons with $p_T(l)>20(26)$ GeV and $|\eta(l)|<2.4$ from Z(W)
- anti-k_T jet: $p_T^{jet} >25$ GeV & $|\eta^{jet}|<2.5$
- Z:Dilepton invariant mass:[71,111]GeV



Z/W+heavy flavor selection:







Identify jets from heavy flavor quarks through the decay of charm hadrons in several final states:

- (a) Semileptonic decay of c/b hadrons: muon in a jet
- (b) D hadron inclusive decays: SV in a jet
- (c) $D^{*\pm}$, $D^{\pm}$ exclusive decays

Samples:

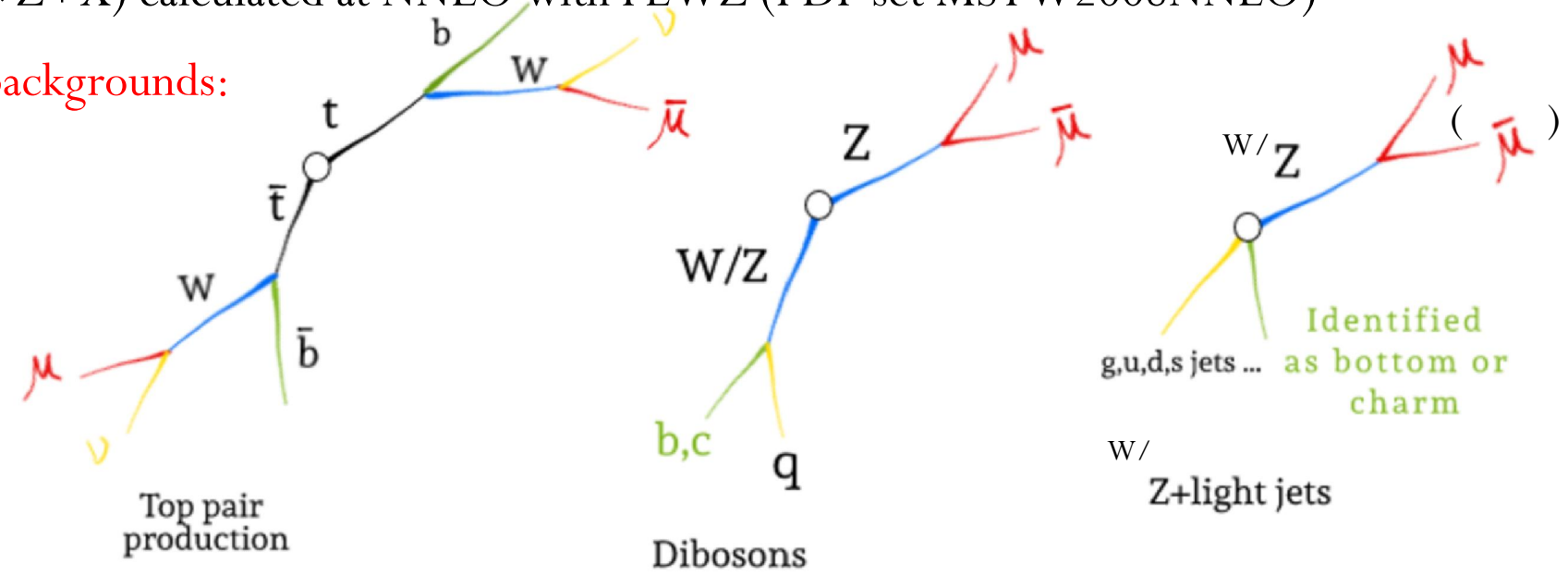
DATA: 2012 8 TeV (19.7 fb^{-1} ± 0.5) for Z+c & W+c , 2016 13 TeV (35.7 fb^{-1}) for W+c

Signal MC: DY&W+jets generated w. MADGRAPH5@LO+PYTHIA6(PDF set CTEQ6L)

W+jets generated w. MADGRAPH5@NLO+PYTHIA8(PDF set NNPDF3.0)

σ(pp→Z+X) calculated at NNLO with FEWZ (PDF set MSTW2008NNLO)

Main backgrounds:



Contributions from $t\bar{t}$, diboson, W/Z+light processes (from simulations except $t\bar{t}$ from data for Z analysis).

Z: Missing transverse energy < 40 GeV (to reduce $t\bar{t}$ background).

W: $M_T(l,\nu) > 50$ GeV

Data-MC differences in lepton trigger, identification and isolation efficiencies corrected (tag & probe method). Pileup events included in the MC.

Semileptonic selection for Z and W analysis (SL channel)

- μ inside a jet (taking part of a secondary vertex for Z). This reduces the light contribution more than standard b-tagging algorithms.
- $p_T^{\mu}<25$ GeV, with $p_T^{\mu}/p_T^{jet}<0.6$, $|\eta^{\mu}|<2.5$
- non-isolated, $I_{comb}/p_T^{\mu}>0.2$

Semileptonic candidates:

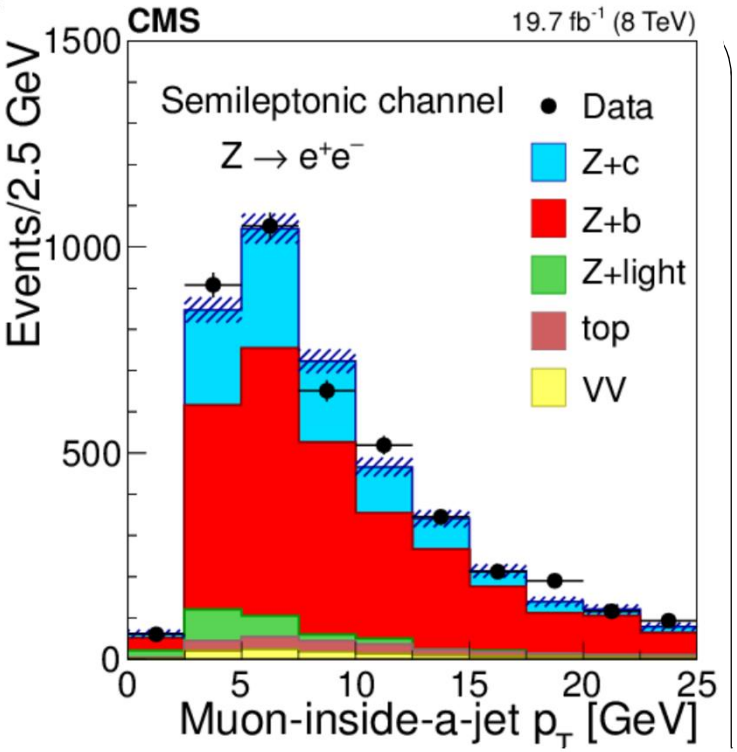
4145 $Z \rightarrow e^+e^-$ 5258 $Z \rightarrow \mu^+\mu^-$

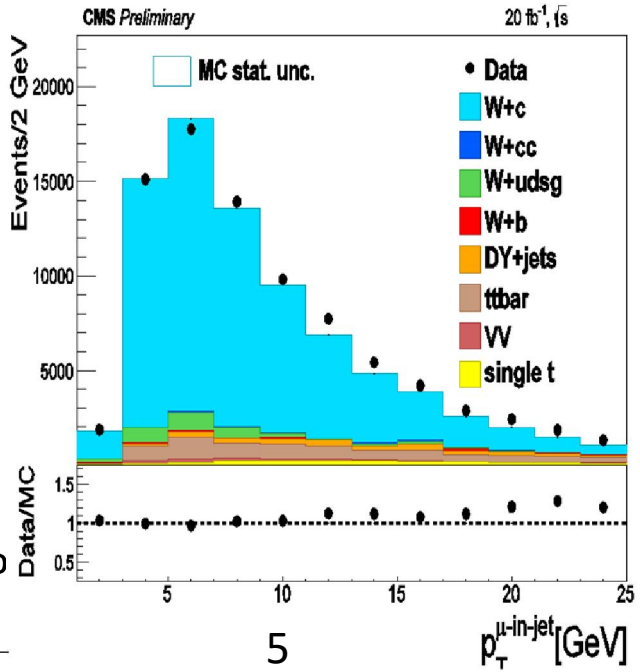
52K $W \rightarrow e\nu$ 32K $W \rightarrow \mu\nu$

Relative contributions:

Z+c:25% Z+b:65% Z+light:5% Others:5%

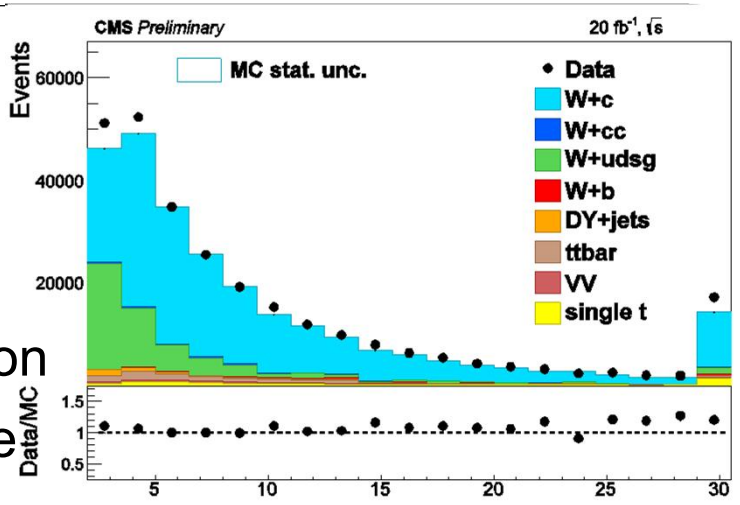
W+c:80% W+b:0.5% W+light:4% top:10% Others:6%





Inclusive D selection for W analysis (SV channel)

- Reconstructed secondary vertex in jet
- Vertex_mass>0.55 GeV and SV flight distance significance>3.5 to reduce light jet contamination
- In case of several jets with SV in the event, take the **highest p_T jet**



SV flight distance significance

131K W → μν 117K W → eν

W+c:75% W+b:0.5% W+light:15% top:6% Others:3.5%

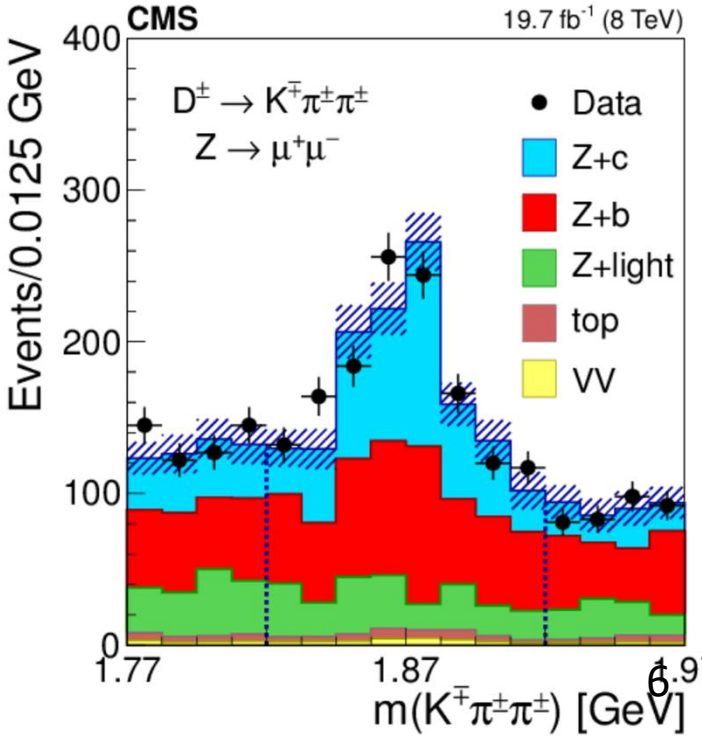
Exclusive $D^{\pm}$ Selection for Z analysis ($D^{\pm}$)

Use jets with a 3 tracks secondary vertex & search for $D^{\pm} \rightarrow K^{\mp}\pi^{\pm}\pi^{\pm}$ resonant peak.

Non resonant bkg. in the signal region sub-tracted from the neighboring sidebands

490±48 $D^{\pm}$ ($Z \rightarrow \mu^{+}\mu^{-}$) 375±44 $D^{\pm}$ ($Z \rightarrow e^{+}e^{-}$)

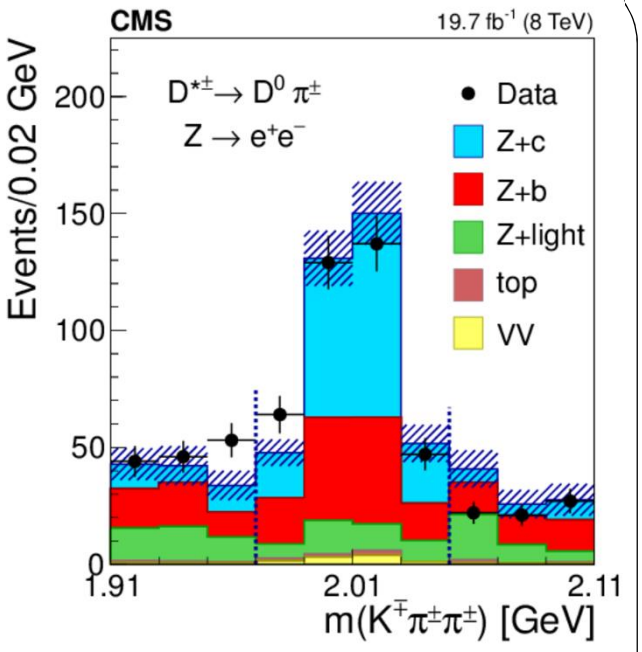
Z+c: ~60% Z+b: ~35% Others:<5%



Exclusive D(2010)$^{*\pm}$ Selection for Z (W) analysis (D$^{*\pm}$ channel)

- $D^{*\pm} \rightarrow D^0 \pi_s^{\pm}$ [$D^0 \rightarrow K^-\pi^+$(+c.c.)] decay chain.
- Kaon: track with sign opposite to π_s

- D^0 vertex with $L_{xy}/\sigma(L_{xy})>3(0)$, D^0 vertex prob.>0.05
- $p_T(D^*)>0(4)$ && $p_T(D^*)/\Sigma p_{T(\text{cone } 0.4)} > 0(0.2)$
- $p_T(K)>1.75(1)$, $p_T(\pi)>0.75(1)$, $p_T(\pi_s)>0.5(0.35)$ GeV
- $|\Delta R(D^*, \text{jet})| < 0.5$, $|\Delta R(D^0, \pi_s)| < 0.1(0.15)$.
- $|m(D^0)-1.865|<100(35)$ MeV, $|\Delta m-145|< 5$ (1) MeV
- Signal region : $1.97 < m(D^*) < 2.05$ GeV
- Sidebands : $0.06 < |m(D^{*\pm})-2.01| < 0.12$ GeV



Non resonant background in the signal region subtracted from the neighboring sidebands

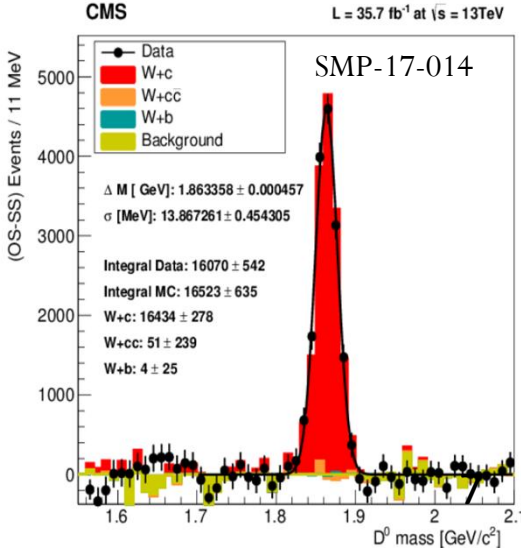
(wrong charge $D^0 \rightarrow K^-\pi^-$(+c.c.) in W+c analysis)

After sideband subtraction:

309 ± 22 $D^{*\pm}(Z \rightarrow \mu^+\mu^-)$ 234 ± 22 $D^{*\pm}(Z \rightarrow e^+e^-)$

19.2 ± 0.6 x 10^3 $D^{*\pm}$ $(W \rightarrow \mu\nu)$

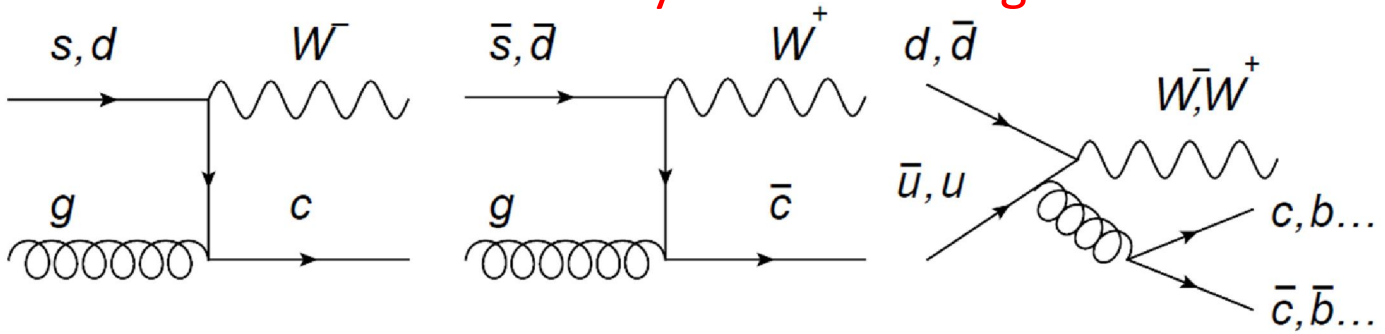
Z+c:~65% Z+b:~30% Z+light:<1% Others($t\bar{t}$+VV):<4%
W+c:>98%



The simulation is reweighted to match the experimental values of
c→ $D^{*\pm} \rightarrow D^0 \pi_s^{\pm}$ [$D^0 \rightarrow K^-\pi^+$] (PDG+L.Gladilin,Eur.Phys.J.C75(2015)19)

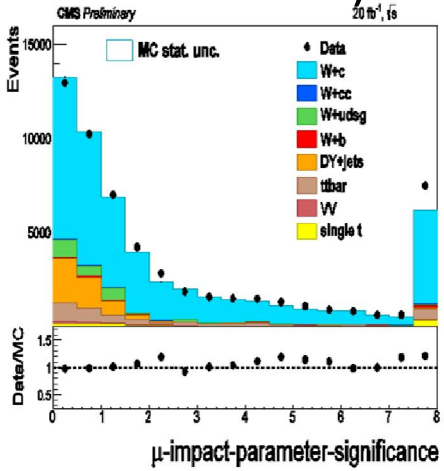
Selection of W+c samples

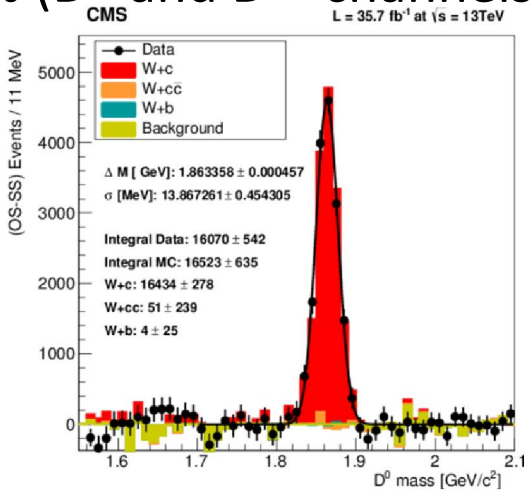
- W→ eν, μν plus jets with similar selection to Z+HF
- Identification of heavy flavor jet: **μ or SV in jet from D-hadron inclusive decays or $D^{*\pm}$**
- OS–SS subtraction to remove symmetric backgrounds

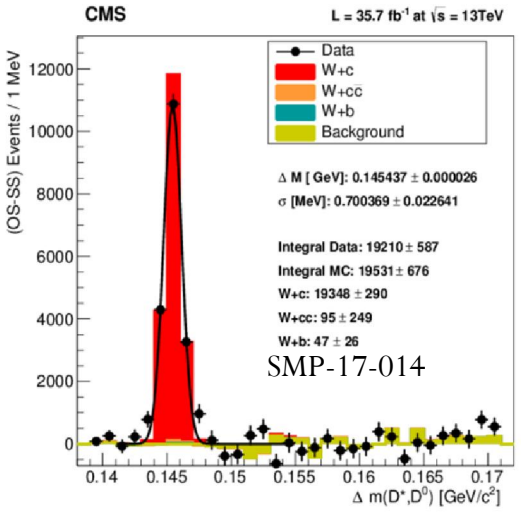


OS event *OS event* *Bckg: 50% OS, 50% SS*

After OS-SS subtraction the purity in W+c of the resulting sample is > 80% (SL and SV channels) and >98% ($D^{\pm}$ and $D^{*\pm}$ channels)





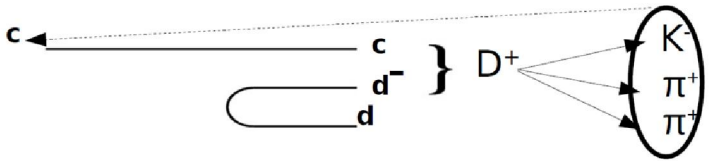


SMP-17-014

Charm charge determination of W+c samples

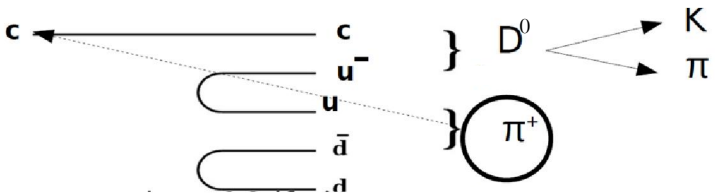
•In the SL(D^*) channel the charm charge is that of the μ(D^*)

• In the SV channel, the charge of the SV vertex : $\sum_{\text{Vertex-tracks}} q_{\text{tracks}}$



•If vertex-charge == 0 use charge of closest PV-track (p_T>0.3 && ΔR(track, SV vertex)<0.1).

The sign of that track, the closest track to the $\text{Charm}_{\text{hadron}}$ in the process of fragmentation, tells you whether we have a c or a cbar :



•Charge definition: OS if charges of the SV and the lepton from the W decay are opposite.

•OS : $Q_{W\to e(\mu)} \neq Q_{SV}$ SS : $Q_{W\to e(\mu)} = Q_{SV}$

Signal extraction (Z+c)

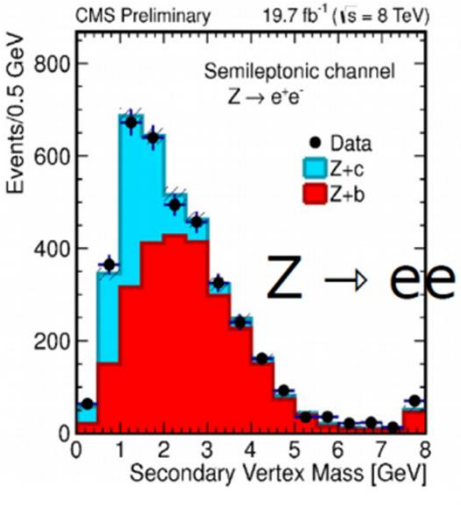
Z: Total # of observed Z+c/Z+b extracted from a χ^2 minimization fit of the Z+c/Z+b templates to the experimental distributions of vertex mass and JP discriminants (likelihood estimate of prob. of jet tracks to come from primary vertex)

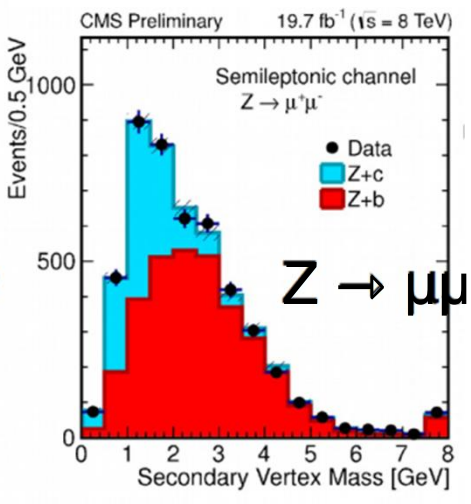
$$\chi^2 = \sum_i \frac{(n_i^{data} - \mu_{\mathrm{Z+c}} N_i^{\mathrm{Z+c}} - \mu_{\mathrm{Z+b}} N_i^{\mathrm{Z+b}})}{(\sigma_i^d)^2 + (\sigma_i^{\mathrm{MC}})^2}$$

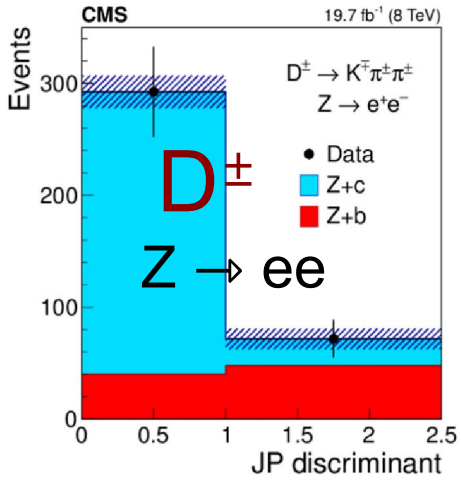
n_i = Number of events in data (after subtraction of remaining background from Z+light, tt and VV)

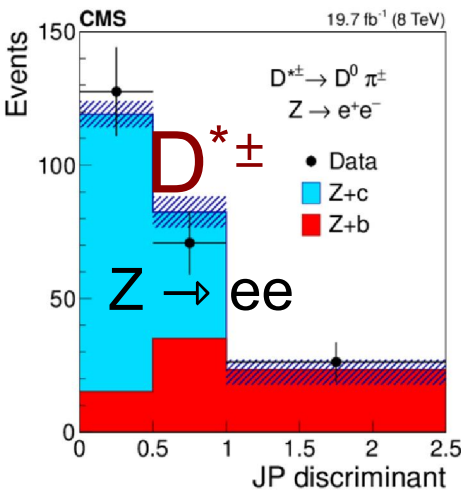
N_i^{Z+c}, N_i^{Z+b} = Number of Z+c, Z+b

Parameters to fit: μ_{z+c} & μ_{z+b}









c/b separation clearer in the D* mode (the soft pion comes from the PV for c→ D* and not for b→ B→D*)

μ_{z+c} & μ_{z+b} in the 0.9-1.1 range

Results: σ(W+c) @ 13 TeV

Charm from D^* channel

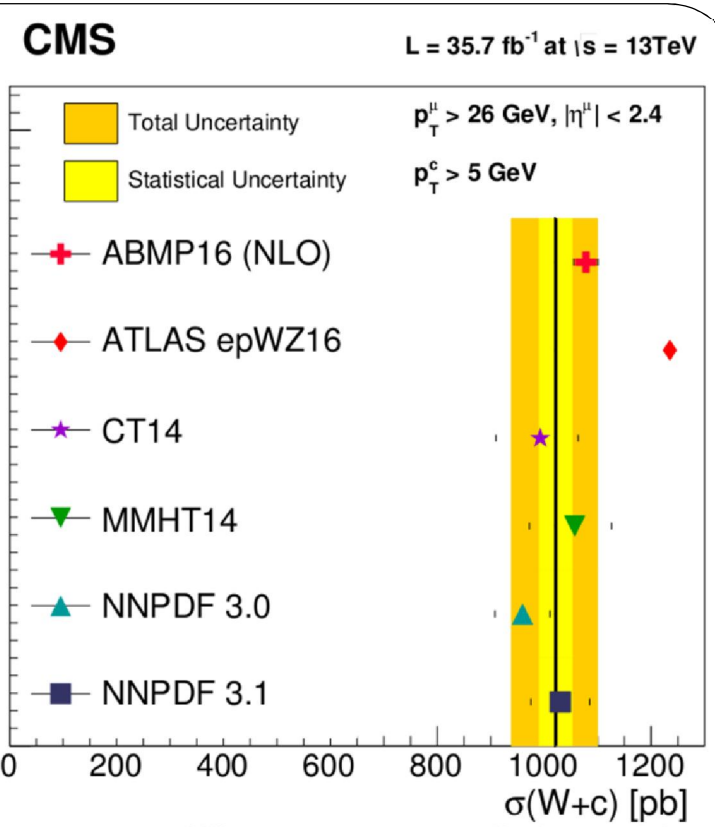
$$\sigma = N_{(W+D^*)data} / L\, B\, A_c\, \varepsilon_c$$

$$A_c\, \varepsilon_c = N_{W+D^*(reco)} / N_{W+c(gen)}$$

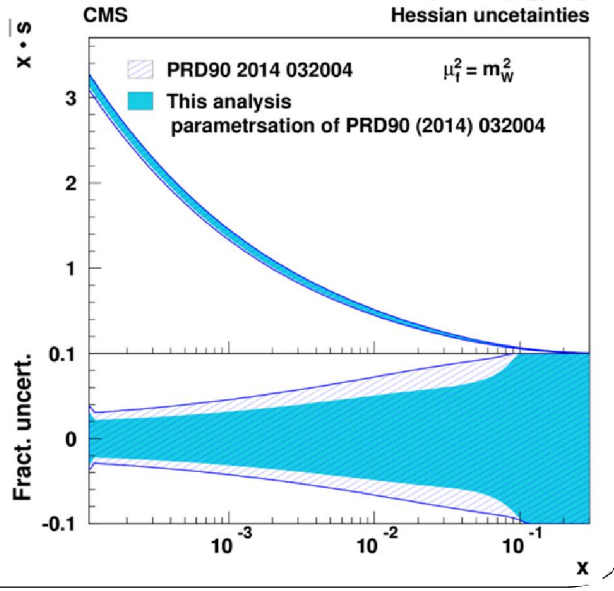
$\sigma(W+ c) = 1026 \pm 31(stat)^{+76}_{-72}(syst)$ pb

$\sigma(W^+ + c)/\sigma(W^- + c) = 0.968 \pm 0.055(stat)^{+0.015}_{-0.028}(syst)$

In good agreement with the theoretical predictions at NLO using different PDF sets (except f



From a QCD analysis at NLO together with inclusive DIS measurements and earlier results from CMS on W+c production and the lepton charge asymmetry in W-production : The strange quark distribution and strangeness suppression factor agree with results from neutrino-scattering experiments.



Results: σ(W+c) @ 8 TeV

Charm from SL and SV channels

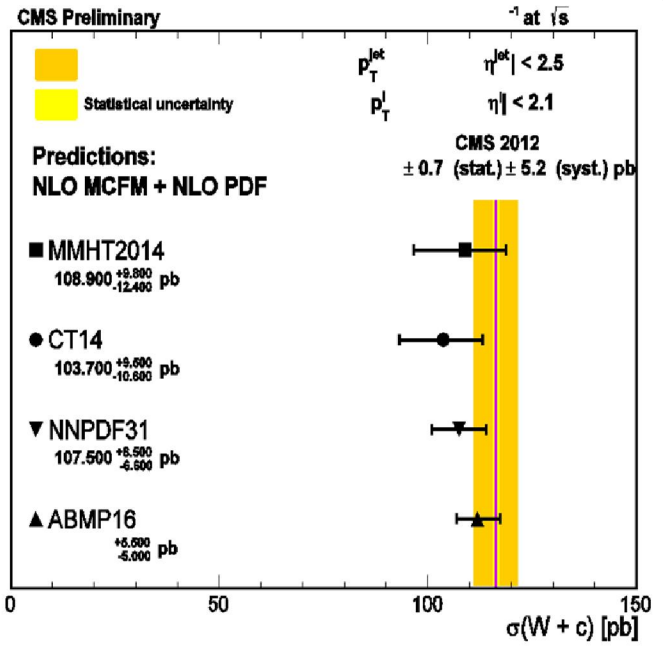
$$\sigma = N_{(W+c)data} / L\, B\, A_c\, \varepsilon_c$$

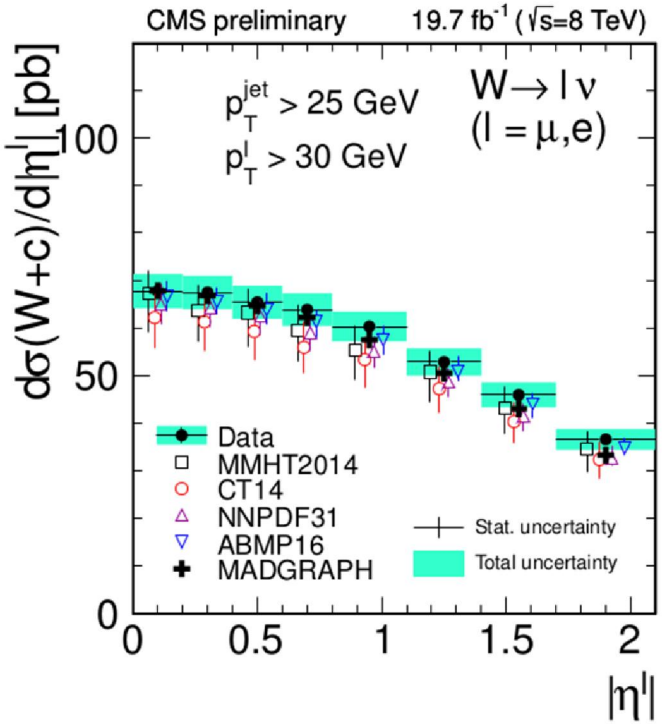
$$A_c\, \varepsilon_c = N_{W+c(reco)} / N_{W+c(gen)}$$

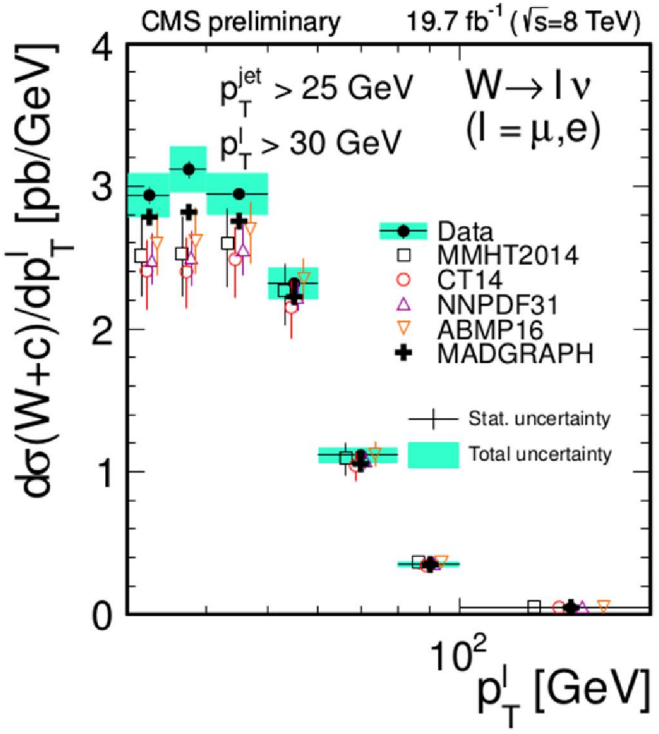
$\sigma(W+c) = 116.3 \pm 0.7(stat) \pm 5.2(syst)$ pb

$\sigma(W^+ + c)/\sigma(W^- + c) = 0.986 \pm 0.011(stat) \pm 0.013(syst)$

In good agreement with the theoretical predictions at NLO using different PDF sets

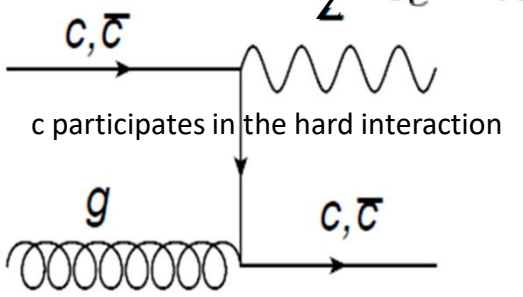




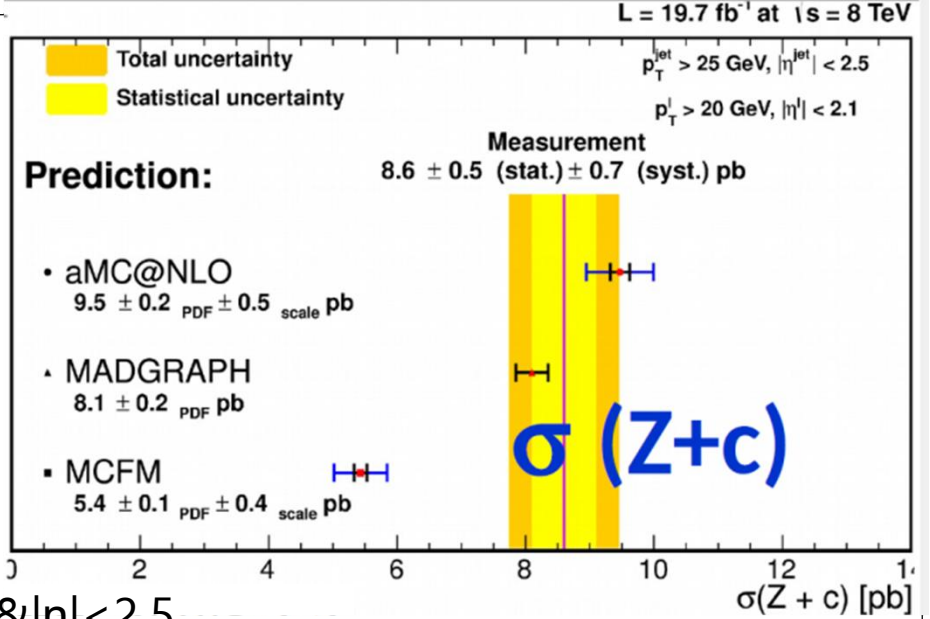


Results: σ(Z+c) @ 8 TeV

$$\sigma(\mathrm{Z}+\mathrm{c}) = \frac{N_{\mathrm{Z+c}}^{fitted}}{\epsilon_c^{\mathrm{Z+c}} \mathcal{L}}$$

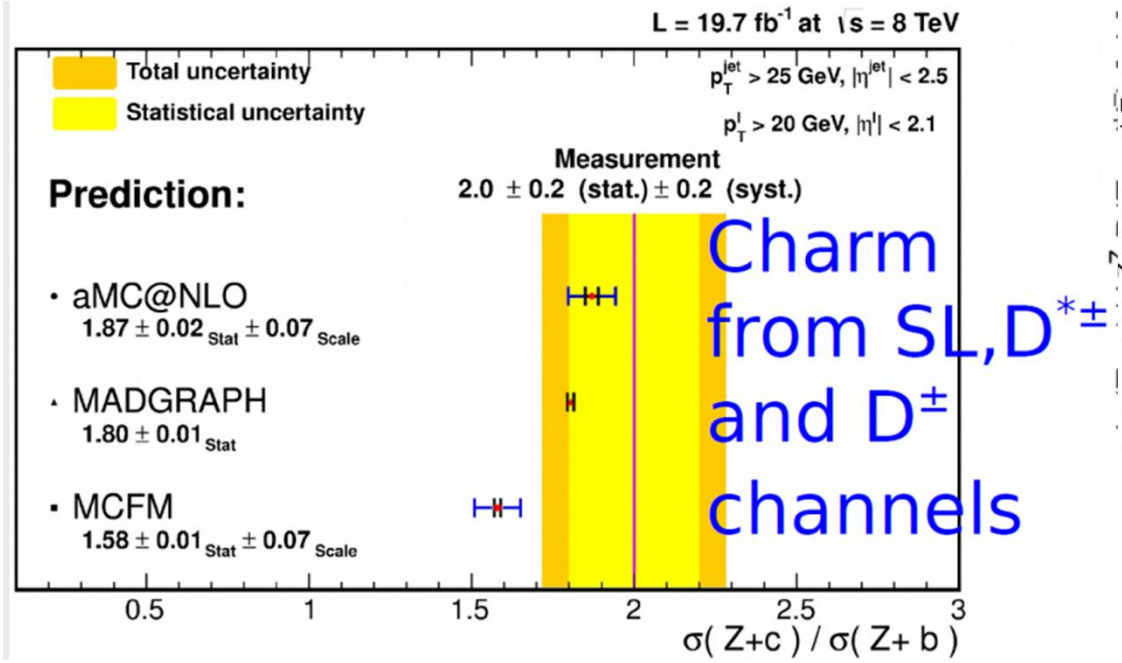


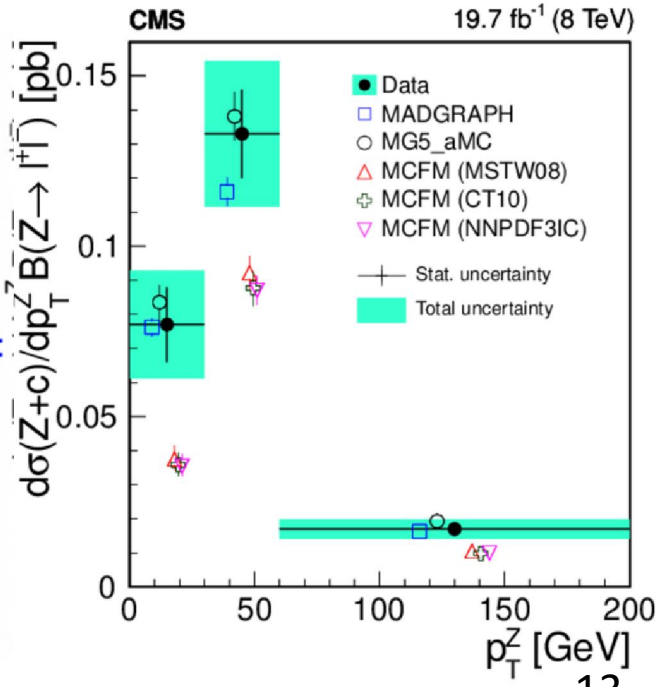
c participates in the hard interaction



μ,e(pT)>25 GeV & |η|<2.1, **jet(pT)>25 GeV** &|η|<2.5

MCFM do not include contributions from GS





Conclusions

Evaluated W/Z + c associated production, inclusive and differential

Z in agreement with predictions from MadGraph5 amc@nlo (and Madgraph renormalized to a FEWZ calculation for Z+c). W in agreement with the theoretical predictions at NLO using different PDF sets

There has been a lot of improvement in the last decades and there is more to come from both , theoretical and experimental results

Back up

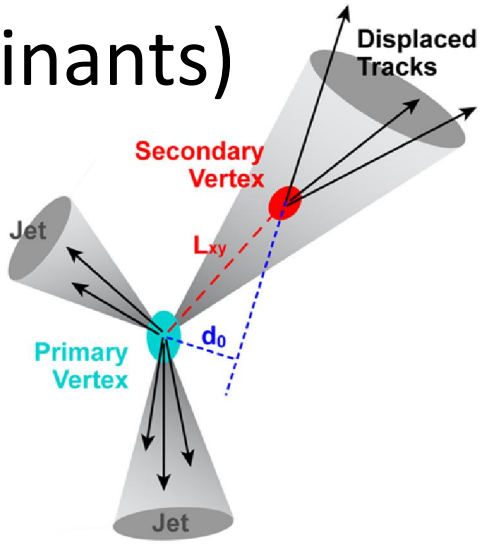
Z+c Selection: b/c separation (discriminants)

- Vertex mass (for semileptonic mode)

$$M_{\text{vertex}}^{\text{corr}} = \sqrt{M_{\text{vertex}}^2 + p_{\text{vertex}}^2 \sin^2\theta} + p_{\text{vertex}} \sin\theta,$$

Correction included to account for unidentified neutral decay products

- JP (for D hadron modes): likelihood estimate of prob. of jet tracks to come from primary vertex



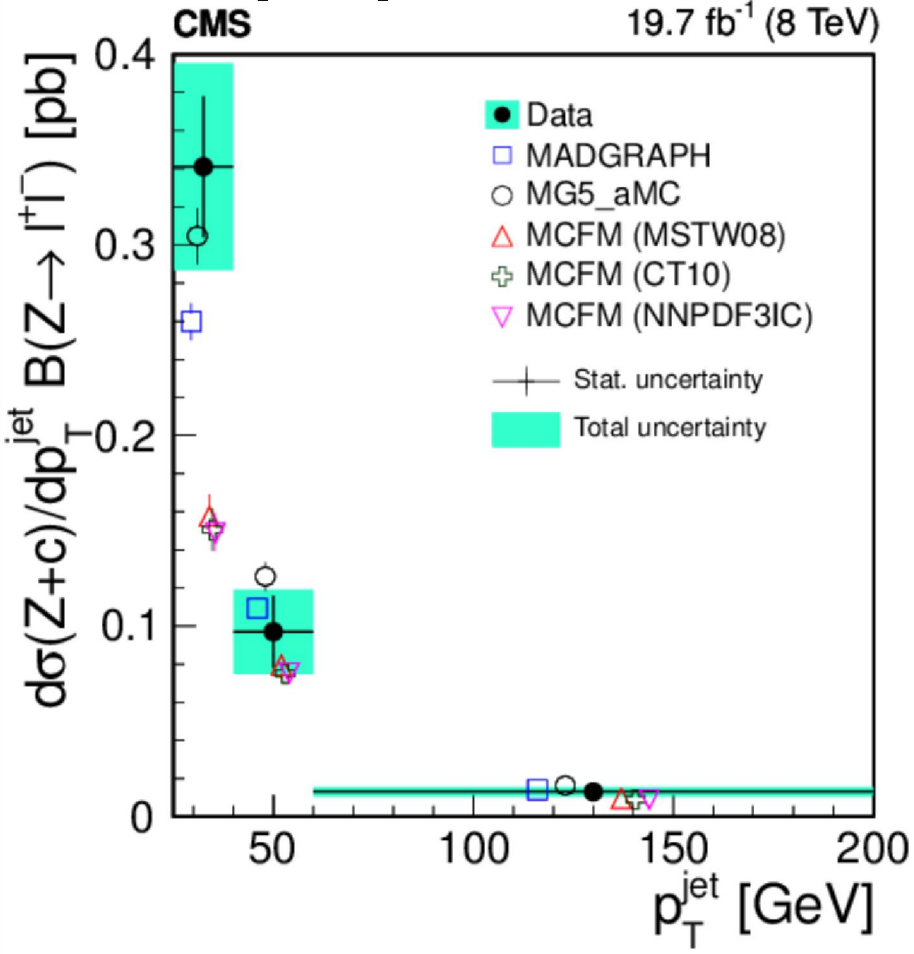
The larger the IP of a track the more inconsistent w.r.t. PV

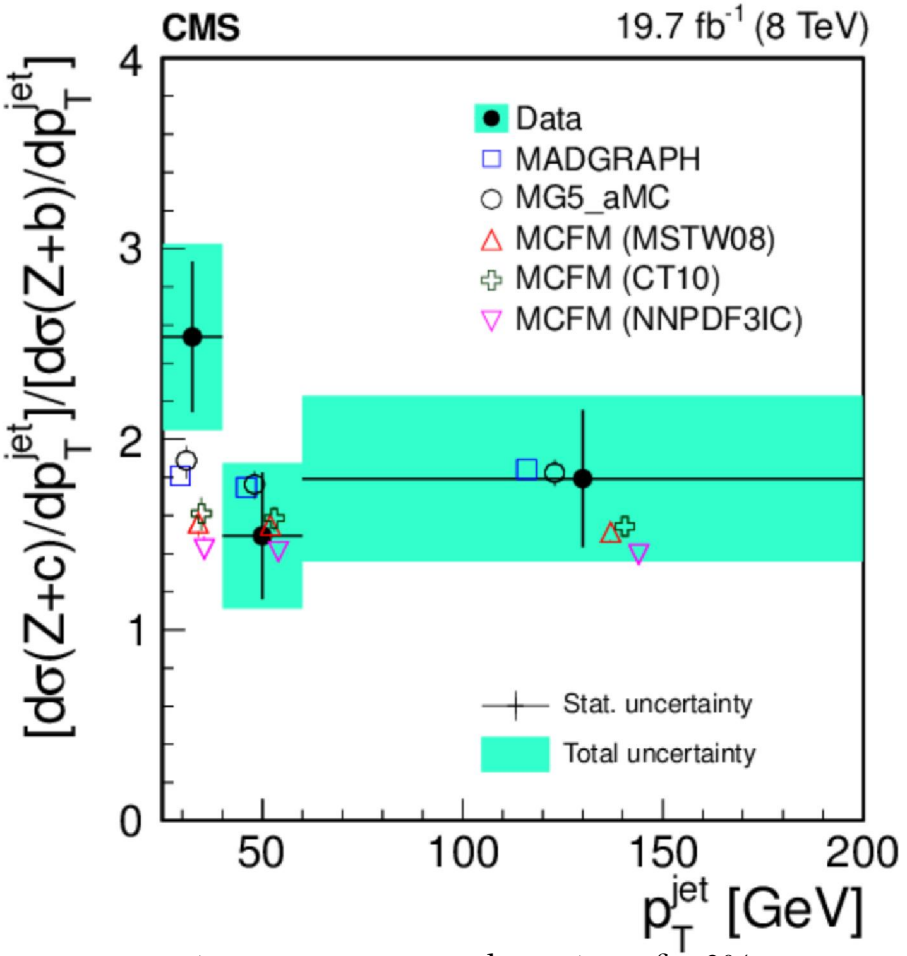
Modeling strategy

- Z+c :
  - Shape : data driven (W+charm) [1st time]
  - Normalization taken from MC after applying vertex-efficiency corrections
- Z+b :
  - Shape : from MC but corrected with data (ttbar)
  - Normalization from MC after vertex-efficiency corrections
- Z+light and Dibosons: shape and normalization from MC
- ttbar: Data driven

Differential cross sections as a function of p_T^{jet}

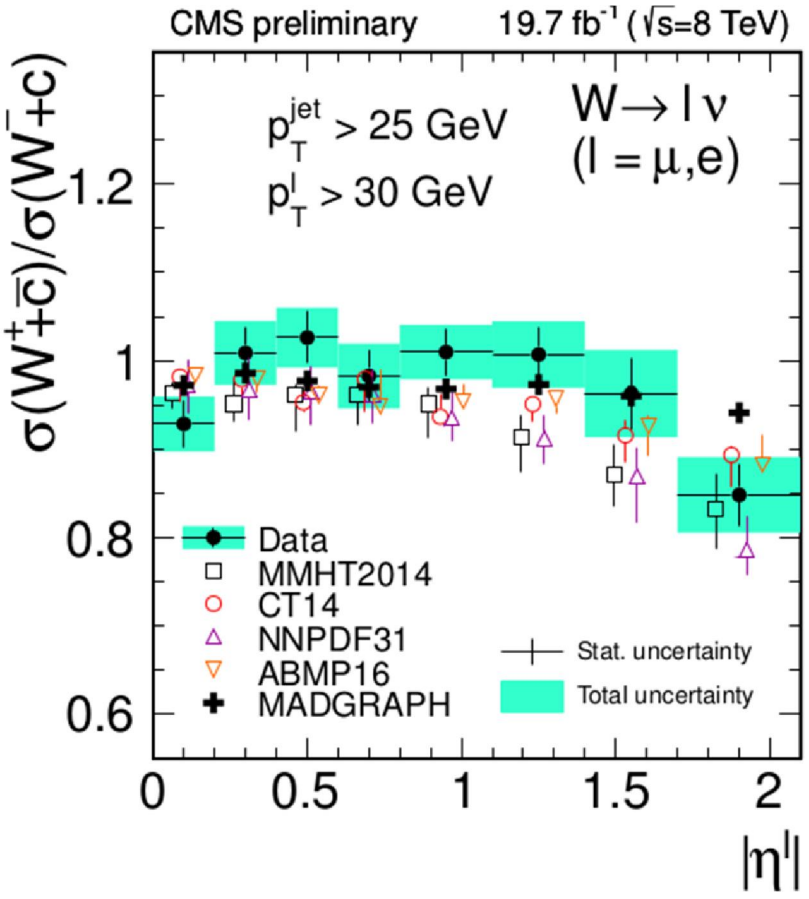
Bins [GeV] : 25-40, 40-60 and 60-200

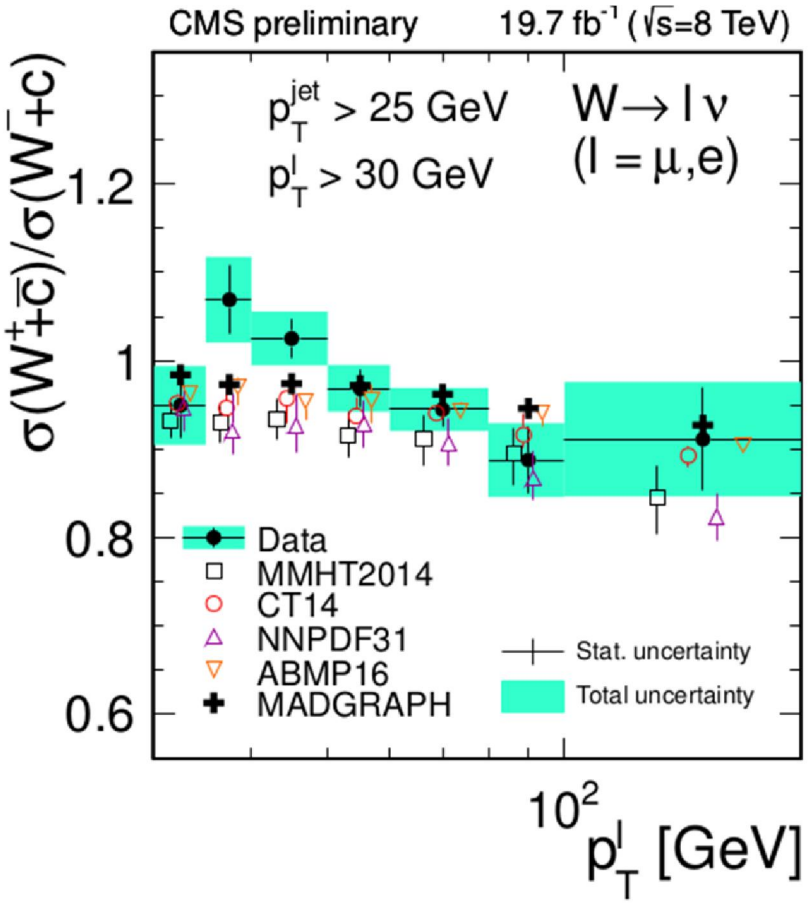




If the proton momentum fraction taken by the charm quark component (intrinsic + per-turbative) is of ~2%,an increase in

- Comparison with MCFM (differential cross section ratio)





ss section determinatic

Semileptonic mode			
Channel	N_{Z+c}^{signal}	$\mathcal{C}_{Z+c}$ (%)	$\sigma(Z+c)$ [pb]
$Z \to e^+e^-$	1066 ± 95	0.63 ± 0.03	$8.6 \pm 0.8 \pm 1.0$
$Z \to \mu^+\mu^-$	1449 ± 144	0.81 ± 0.03	$9.1 \pm 0.9 \pm 1.0$
$Z \to \ell^+\ell^-$	$\sigma(Z+c) = 8.8 \pm 0.6\,(stat) \pm 1.0\,(syst)$ pb		

Channel	N_{Z+b}^{signal}	$\mathcal{C}_{Z+b}$ (%)	$\sigma(Z+c)/\sigma(Z+b)$
$Z \to e^+e^-$	2606 ± 114	2.90 ± 0.08	$1.9 \pm 0.2 \pm 0.2$
$Z \to \mu^+\mu^-$	3240 ± 147	3.93 ± 0.10	$2.2 \pm 0.3 \pm 0.2$
$Z \to \ell^+\ell^-$	$\sigma(Z+c)/\sigma(Z+b) = 2.0 \pm 0.2\,(stat) \pm 0.2\,(syst)$		

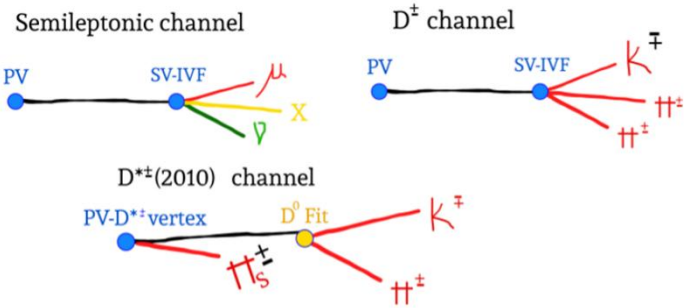
$D^{\pm}$ mode			
Channel	N_{Z+c}^{signal}	$\mathcal{C}_{Z+c}$ (%)	$\sigma(Z+c)$ [pb]
$Z \to e^+e^-$	276 ± 55	0.13 ± 0.02	$10.9 \pm 2.2 \pm 0.9$
$Z \to \mu^+\mu^-$	316 ± 75	0.18 ± 0.02	$8.8 \pm 2.0 \pm 0.8$
$Z \to \ell^+\ell^-$	$\sigma(Z+c) = 9.7 \pm 1.5\,(stat) \pm 0.8\,(syst)$ pb		

$D^{*\pm}(2010)$ mode			
Channel	N_{Z+c}^{signal}	$\mathcal{C}_{Z+c}$ (%)	$\sigma(Z+c)$ [pb]
$Z \to e^+e^-$	151 ± 31	0.11 ± 0.01	$7.3 \pm 1.5 \pm 0.5$
$Z \to \mu^+\mu^-$	247 ± 28	0.14 ± 0.01	$9.3 \pm 1.1 \pm 0.7$
$Z \to \ell^+\ell^-$	$\sigma(Z+c) = 8.5 \pm 0.9\,(stat) \pm 0.6\,(syst)$ pb		

Combination	
$Z \to \ell^+\ell^-$	$\sigma(Z+c) = 8.8 \pm 0.5\,(stat) \pm 0.6\,(syst)$ pb

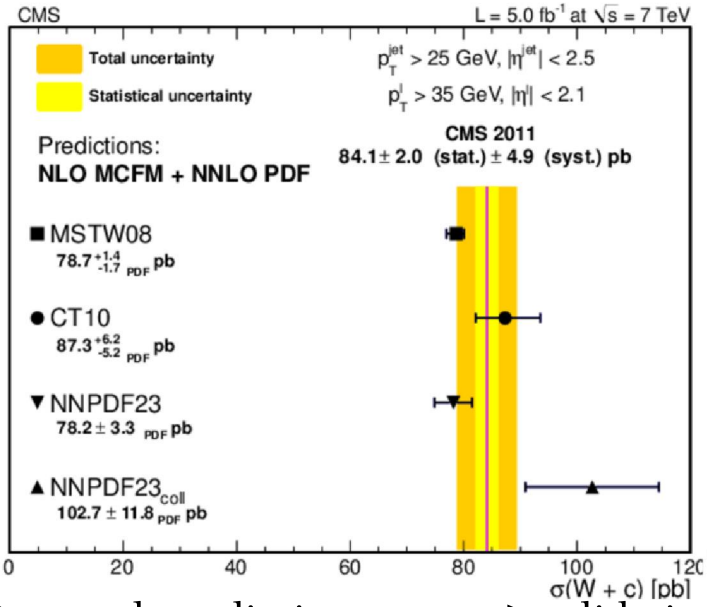
W+c

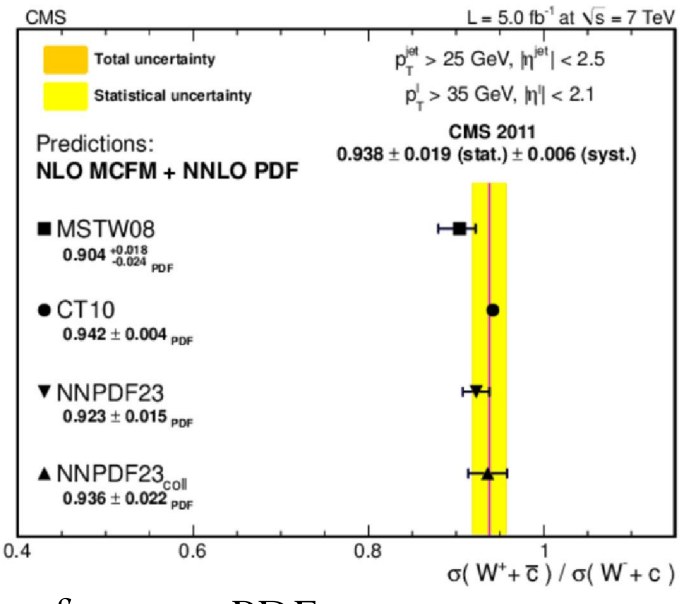
- W→ eν or μν

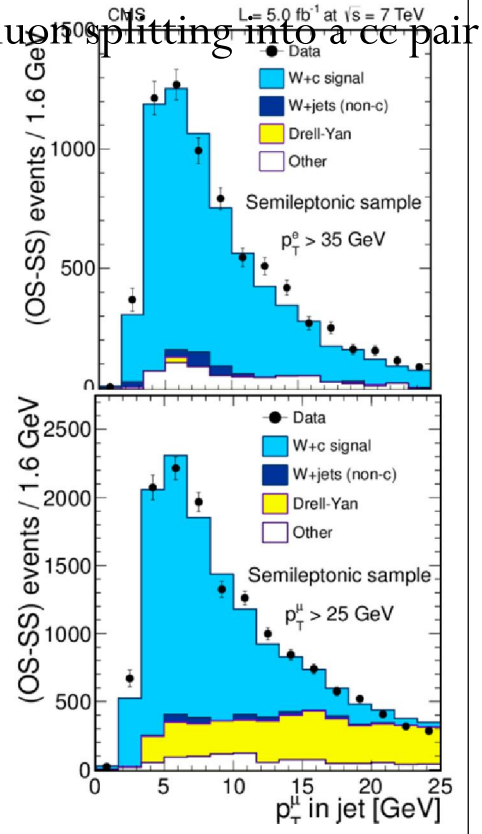


Identification of heavy flavor: **μ in jet or D-hadron**

The MCFM predictions for this process do not include contributions from gluon splitting into a cc pair







Data and predictions agree->validation of strange PDFs

First evidence for an asymmetry in the W^++c and W^-+c production.

Modeling strategy

Now that we have chosen the variables to separate the different contributions we need a way to model properly each of them

This is called template modeling and has two parts:

- Modeling properly the shape
- Accurate determination of tagging efficiency

- Z+c :

  Shape : data driven (W+charm) [1st time]

  Normalization taken from MC after applying vertex-efficiency corrections

- Z+b :

  Shape : from MC but corrected with data (ttbar)

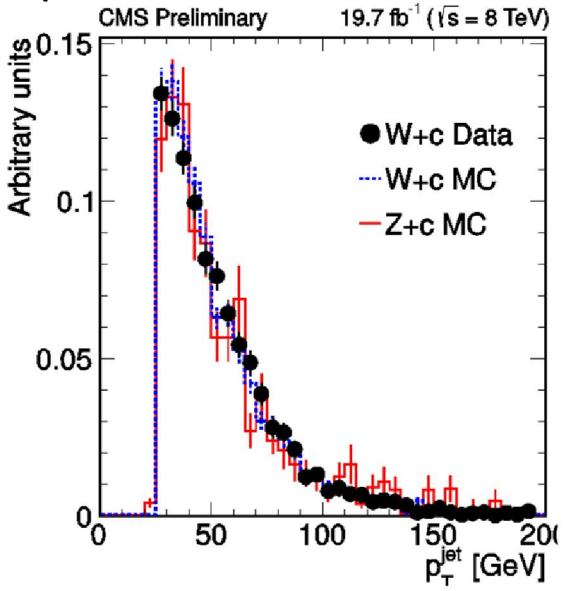
  Normalization from MC after vertex-efficiency corrections

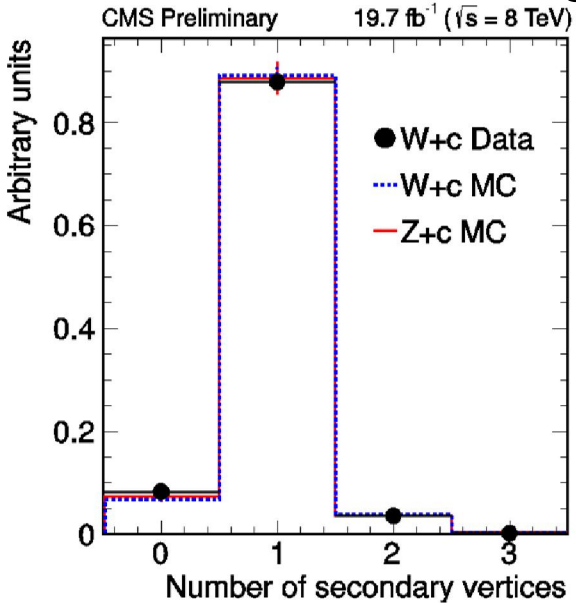
- Z+light and Dibosons: shape and normalization from MC
- ttbar: Data driven

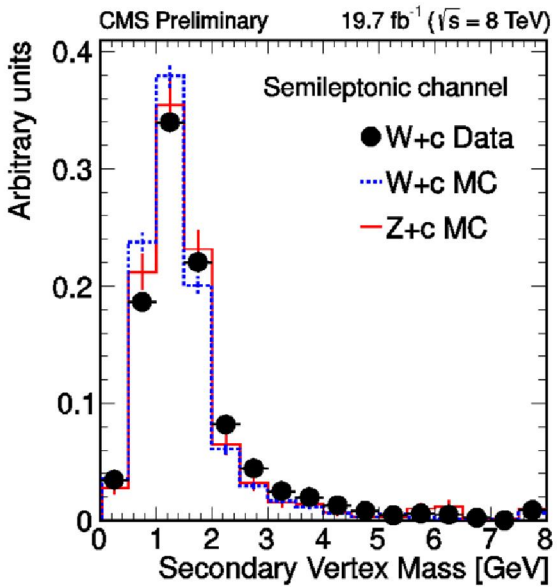
Template (shape) modeling for Z+c

Comparison of c-jets from Z+c and W+c processes

(data from W+c : after subtraction of remaining (little) background)







- *Agreement in general distributions (p_T^{jet} , N_{SV})*
- Discriminant distributions (SV-mass and JP) **W+c MC** and **Z+c MC** agree
- JP prob **W+c MC** and **W+c data** agree and validates the **Z+c MC** description
- SV-mass **W+c MC** and **W+c data** do not agree

The shape is not well modeled by the **W+c MC**. We take the shape of SV-mass f