V+b jets production at CMS

Ernesto Migliore
Università di Torino and INFN

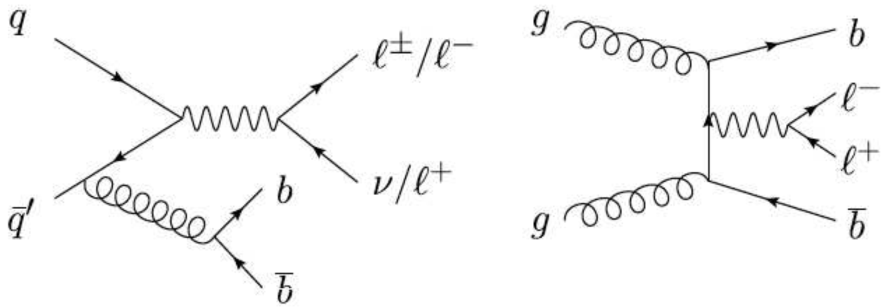
on behalf of the CMS Collaboration

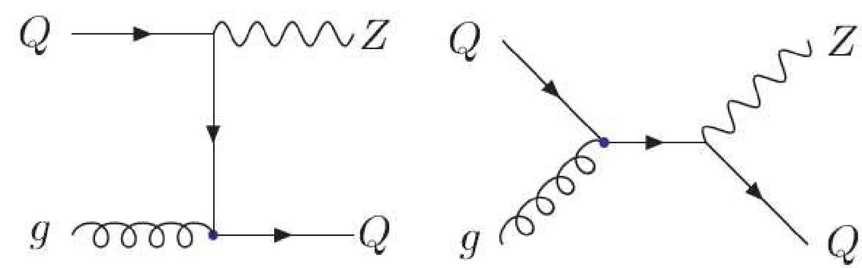
Outline

- Motivations
- Main features of the analyses
- Measurements:
  - Z+jets cross section in bins of b multiplicity
  - angular correlation between B-hadrons in Z + 2 Secondary Vertices events
- Summary and outlook

Motivations for vector boson+b jets analyses

- Interest in the SM process itself (4FS vs. 5FS)

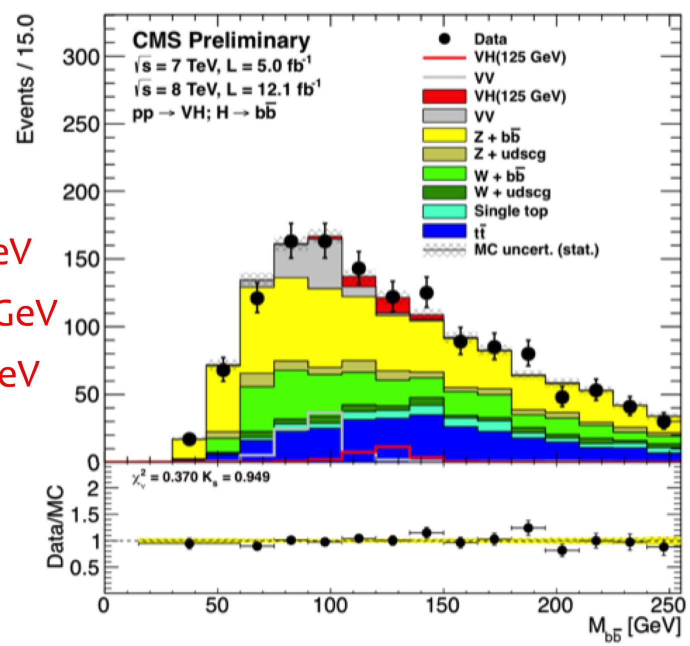




- Interest as one of the largest irreducible backgrounds in several discovery channels e.g. VH→V$b\bar{b}$

CMS PAS HIG-12-044

ℓℓbb $p_T(Z) > 100$ GeV
νℓbb $p_T(W) > 170$ GeV
ννbb $p_T(Z) > 170$ GeV



- MC predictions are the only possibility to estimate this background but large scale factors often used to match MC with data often accompanied by large uncertainties → degradation of the sensitivity of the searches

Features of Z+b analyses

- Exploit
  - robust identification of the leptons (ee or μμ) to identify the Z
  - reconstruction of decay vertex to infer the presence of a decaying B-hadron
- Results so far on the 7 TeV dataset: moderate pile-up (<n_{PV}> from ~6 to ~10)
  - loosest unprescaled triggers are used
    - dimuon p_T thresholds (GeV): 7/7, 8/13, 8/17
    - dielectron p_T thresholds (GeV): 8/17
  - fraction of wrong dilepton–B jet (hadron) vertex association < 1%
- Baseline choice of the MC for the signal:
  inclusive Z/γ* sample from MadGraph+Pythia6 "Z2 tune"+CTEQ6L1 pdfs
  with cross section rescaled to the O(α^2_S) corrections (*k-factor*=1.23)
  ⟶ "signal" = 5FS with massless b quarks in the ME calculation
- MC reweighted to match the trigger/lepton-ID/b-tagging efficiency measured in the data using T&P techniques

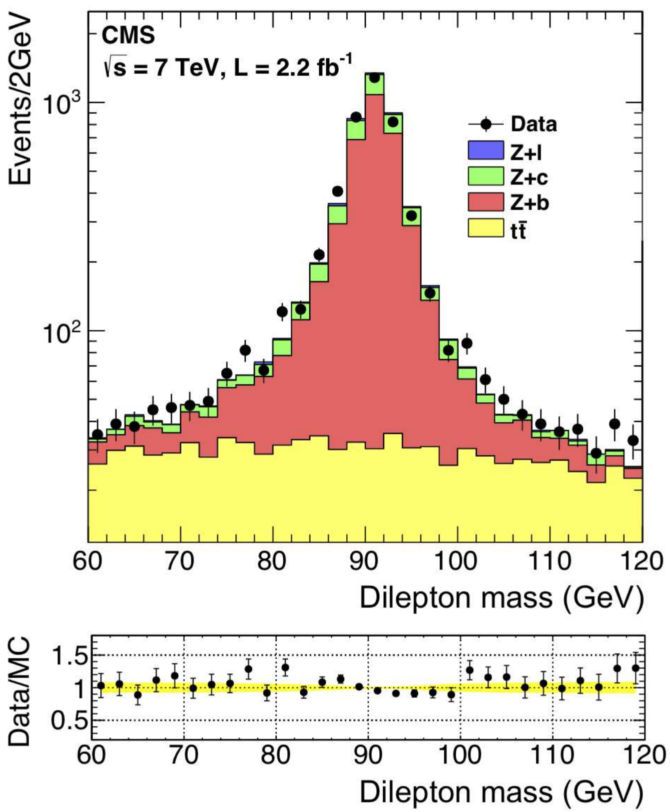
Z+jets in bins of b-jets multiplicity (alias Z+b jets)

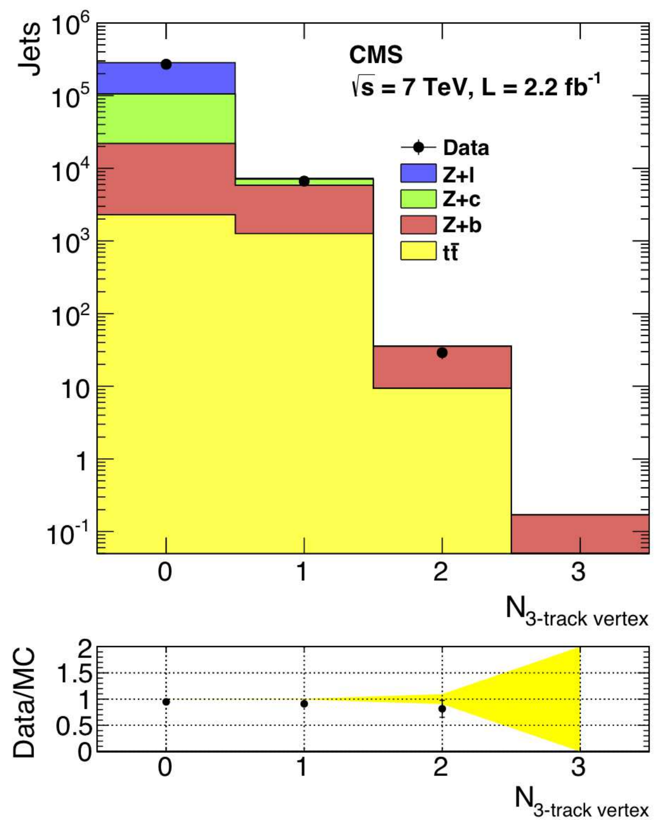
- JHEP 06 (2012) 126: $pp \rightarrow Z/\gamma^{*}+b$ e.g. Z+1b inclusive
- CMS PAS SMP-12-003: $pp \rightarrow Z/\gamma^{*}+1b$ e.g. Z+1b exclusive and
  $pp \rightarrow Z/\gamma^{*}+bb$ e.g. Z+2b inclusive

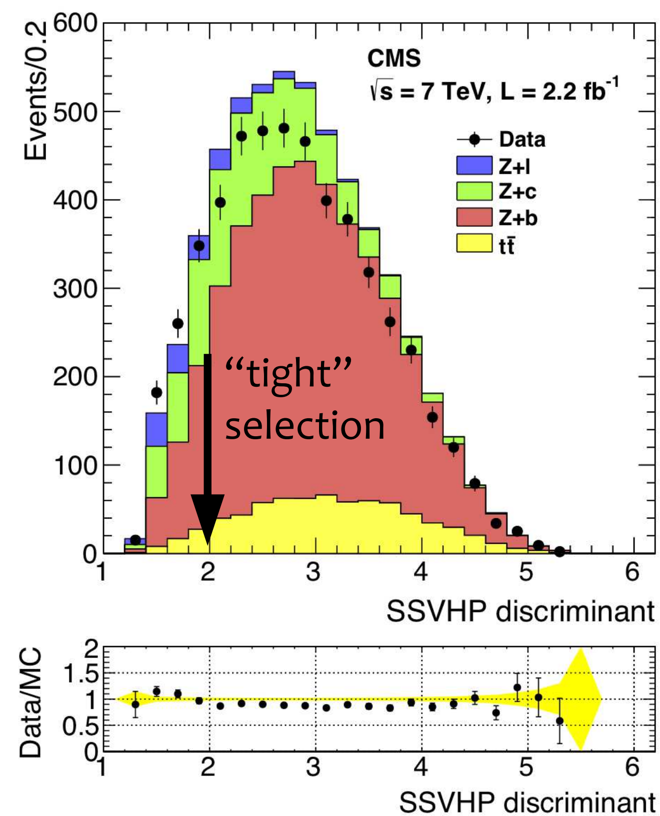
$pp \to Z/\gamma^*+b$: selection of the events

JHEP 06 (2012) 126

- Z/γ^* from isolated leptons
  - $p_T^{\mu(e)}$>20(25) GeV
  - $|\eta^{\mu(e)}|$<2.1(2.5)
  - 60< $m_{\ell\ell}$< 120 GeV
- Presence of additional jets
  - anti k_T (cone R=0.5)
  - p_T^{jet}>25 GeV
  - $|\eta^{jet}|$<2.1
  - $\Delta R(jet,\ell)$>0.5
- b-tagging
  - detached secondary vertex with at least 3 tracks
  - "tight" L_{3D}/σ_{L3D} selection
    - b-jet efficiency ~ 35%
    - udsg-mistag ~ 10^{-3}







JHEP 06 (2012) 126

pp ⟶ Z/γ*+b: extraction of the cross section

- Cross section at the particle level
  - $p_T^b > 25$ GeV and $|\eta^b|<2.1$ hadron level jet (anti k_T R=0.5) containing a B-hadron
  - $60< m_{\ell\ell}< 120$ GeV dilepton (before FSR)
  - $\Delta R(\text{jet},\ell) > 0.5$

$$\sigma_{\text{hadron}}(Z/\gamma^* + \text{b}, Z/\gamma^* \to \ell\ell) = \frac{N(\ell\ell + \text{b}) \times (\mathcal{P} - f_{t\bar{t}})}{\mathcal{A}_\ell \times \mathcal{C}_{\text{hadron}} \times \varepsilon_\ell \times \varepsilon_\text{b} \times \mathcal{L}}$$

b-purity: $\mathcal{P}$

$t\bar{t}$ fraction: $f_{t\bar{t}}$

lepton acceptance: $\mathcal{A}_\ell$

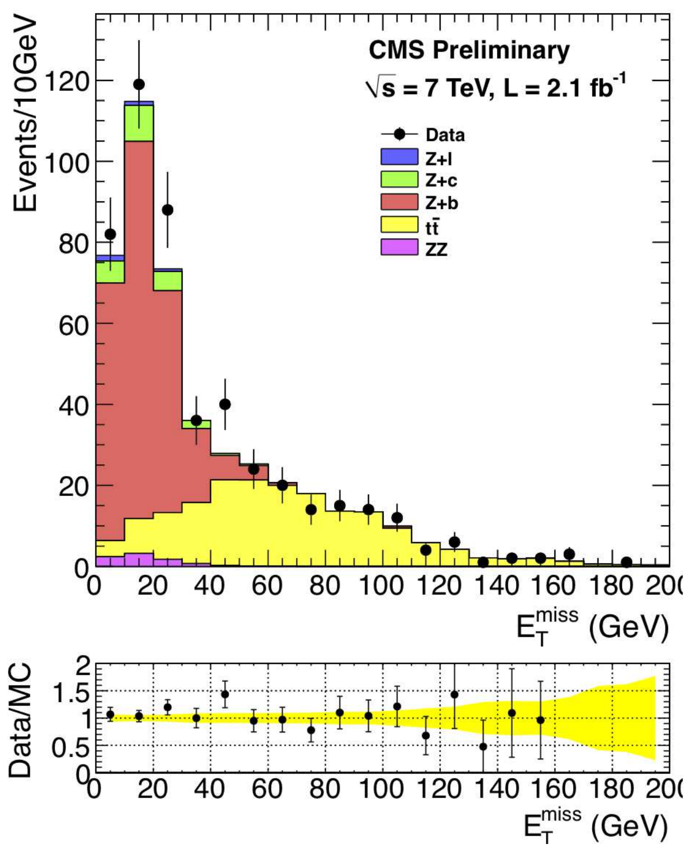
particle-to-detector level correction: $\mathcal{C}_{\text{hadron}}$

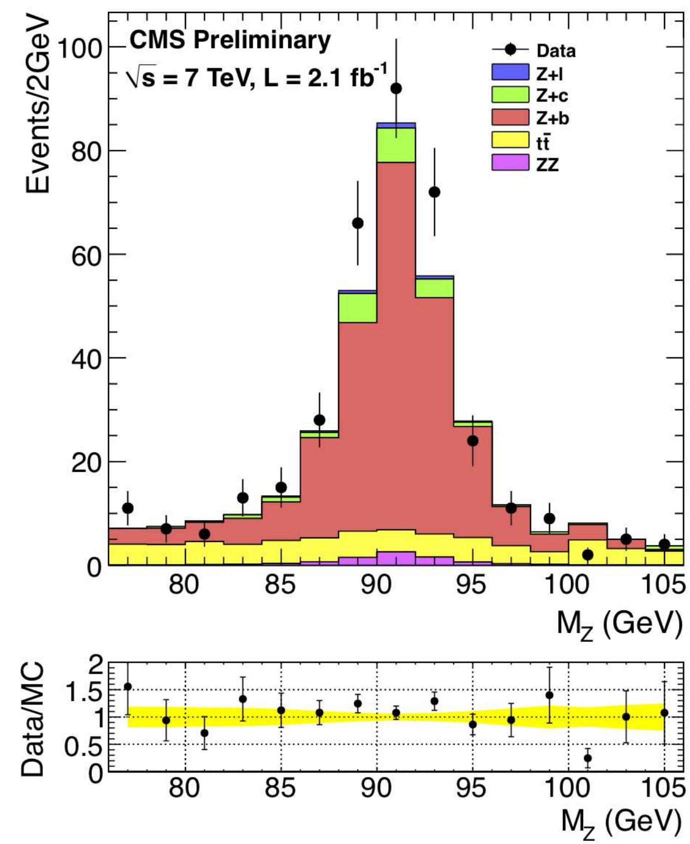
lepton-ID and b-tagging efficiency: $\varepsilon_\ell \times \varepsilon_\text{b}$

- Data-driven estimates of the background subtracted signal yields for μμ (ee)
  - "per -event" b-purity from template fit of the "per-jet" mass of the secondary vertex: P_b~81.5% (83.4%)
  - $t\bar{t}$ fraction from sidebands of the dilepton mass distribution $m_{\ell\ell}$ >120 GeV: f_{tt}~18.4% (18.7%)
- Other correction factors from inclusive Z/γ* simulated events

pp → Z/γ*+bb: selection of events

- Few changes vs. pp → Z/γ*+b to deal with larger fraction of $t\bar{t}$ events wrt Z+jets
  - b-tagging using detached secondary vertex with at least 2 tracks with "medium" L_{3D}/σ_{L3D} selection
    - b-jet efficiency ~ 55%
    - udsg-mistag ~ 10^{-2}
  - E_T^{miss}<50 GeV for Z+bb (eff signal~98%)
  - 76< $m_{\ell\ell}$< 106 GeV





$pp \longrightarrow Z/\gamma^*+bb$: extraction of the cross section

- Cross section at the particle level
  - $p_T^b > 25$ GeV and $|\eta^b|<2.1$ hadron level jet (anti k_T R=0.5) containing a B-hadron
  - "smallest common acceptance" between μμb(b) and eeb(b) $p_T^\ell>20$ GeV, $|\eta^\ell|<2.5$
  - $76< m_{\ell\ell}< 106$ GeV dilepton (before FSR)
  - $\Delta R(\text{jet},\ell) > 0.5$
- Extracting separately Z+1b exclusive and Z+2b inclusive cross sections requires:
  - Determination of background subtracted yields separately for the 1b and 2b cases
    - P_{1b} and P_{2b} from simultaneous 1D+1D template fit of m(SV1) and m(SV2): P_{1b}=56.7% (58.9%) , P_{2b}=83%
    - Fraction of $t\bar{t}$ events from template fit of dilepton mass: $f_{tt\,1b}$=6.4% (7.1%), $f_{tt\,2b}$=20% (17%)
  - Use 2x2 matrices for lepton acceptance, lepton-ID efficiency, b-tagging efficiency and particle-to-detector level corrections factors

$$\mathbb{N}^{sel} = A_l \varepsilon_l \varepsilon_b C_{had} \mathbb{N}^{gen}$$

  - off-diagonal terms introduced by ε_b (e.g. b-tagging inefficiency) and C_{had} (e.g. jet energy resolution)

Z+b jets: cross sections

	Exp [pb]	Th [pb]
Z+1b inclusive JHEP 06 (2012) 126	5.84±0.08(stat)±0.72(syst)+0.25-0.55 (th)	3.97±0.47 [1]
Z+1b exclusive CMS PAS SMP-12-003	3.41±0.05(stat)±0.27(syst)±0.06 (th)	
Z+2b inclusive	0.37±0.02(stat)±0.07(syst)±0.02 (th)	0.33±0.01(stat) [2]
Sum (extrap. to same A_ℓ)	3.78±0.05(stat)±0.31(syst)±0.11 (th) 5.72±0.09(stat) ±0.47(syst)±0.39 (th)	

- Statistics matters: ~5k events for Z+1b inclusive vs. 0.35k events for Z+2b inclusive
- Δσ/σ systematics for μμ (ee):

	Z+1b inclusive	Z+2b inclusive
b-tagging efficiency	3.9% (4.0%)	7.5% (7.4%)
b-jet purity	2.5% (3.5%)	11.0% (10.3%)
Jet Energy Scale	3.8% (3.9%)	6.4% (6.9%)
$t\bar{t}$ fraction	0.5% (0.9%)	9.4% (8.9%)

Z+b jets: cross sections

	Exp [pb]	Th [pb]
Z+1b inclusive (JHEP 06 (2012) 126)	5.84±0.08(stat)±0.72(syst)+0.25-0.55 (th)	3.97±0.47 [1]
Z+1b exclusive (CMS PAS SMP-12-003)	3.41±0.05(stat)±0.27(syst)±0.06 (th)	
Z+2b inclusive	0.37±0.02(stat)±0.07(syst)±0.02 (th)	0.33±0.01(stat) [2]
Sum (extrap. to same A_ℓ)	3.78±0.05(stat)±0.31(syst)±0.11 (th) 5.72±0.09(stat) ±0.47(syst)±0.39 (th)	

- Z+1b inclusive: some tension with [1]
- Z+2b inclusive: good agreement with [2]

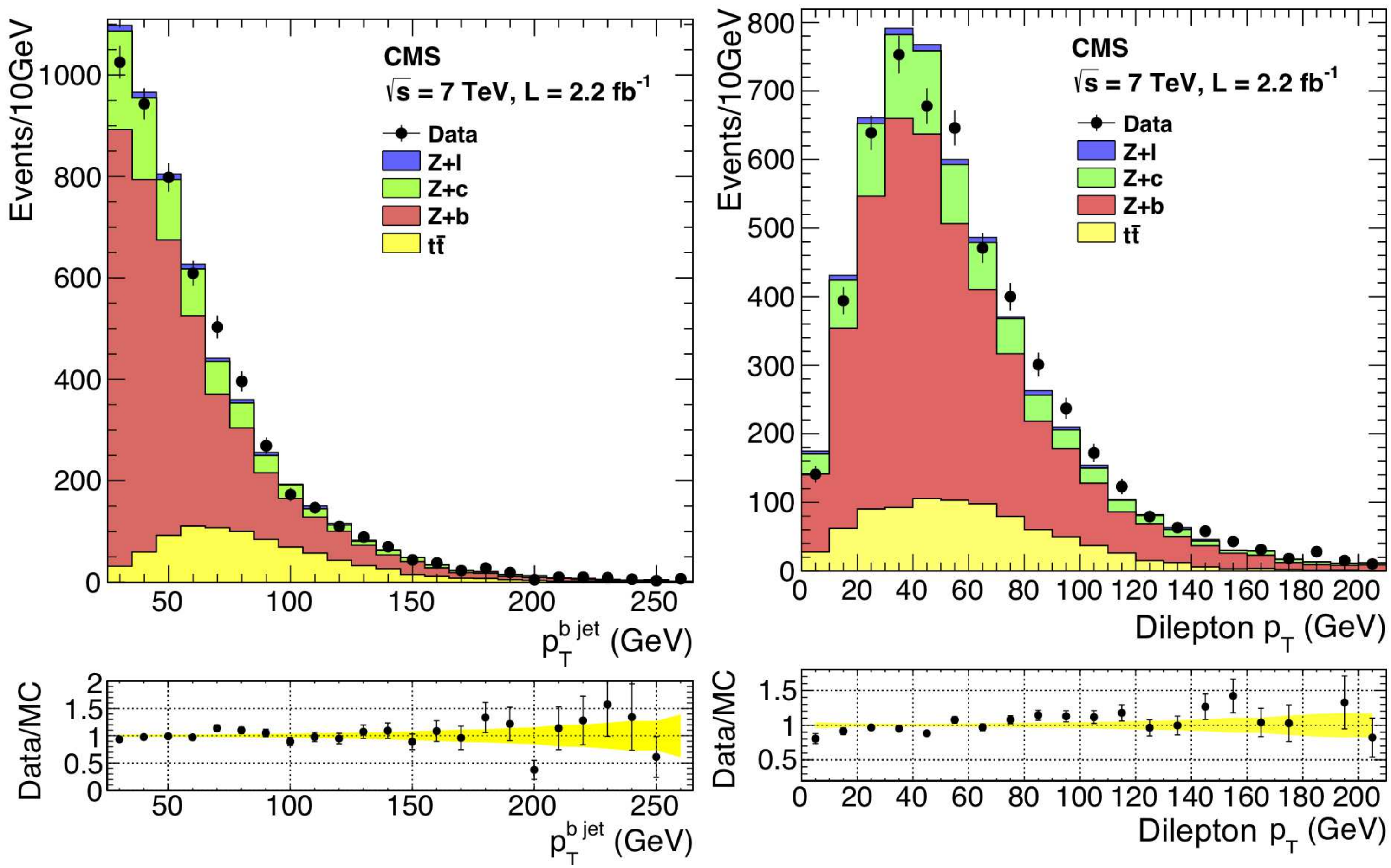
[1] MCFM with parton level to particle level correction derived from MadGraph 5FS+Pythia6

[2] MadGraph 5FS + Pythia6 Drell-Yan scaled to inclusive Z production with $O(\alpha_S^2)$ corrections

JHEP 06 (2012) 126

Z+b jets: kinematic variables at the detector level

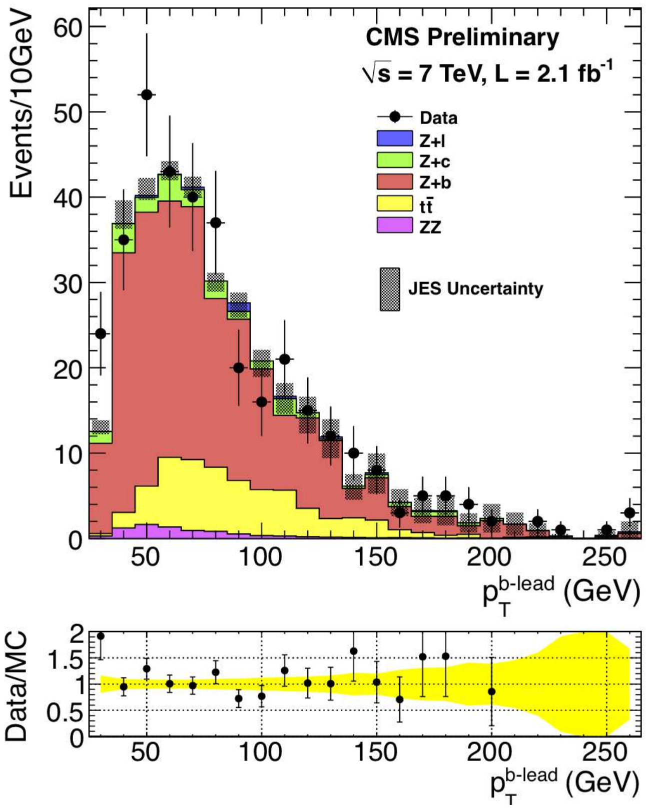
- pp → Z/γ*+b: large statistics (~3.4k signal events) allows to study b-jet kinematics in a b-pure environment

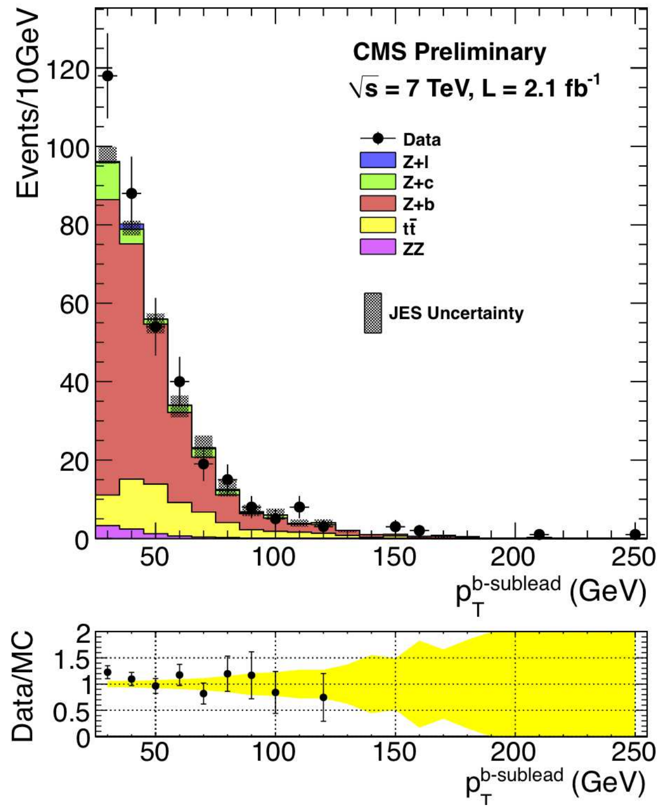


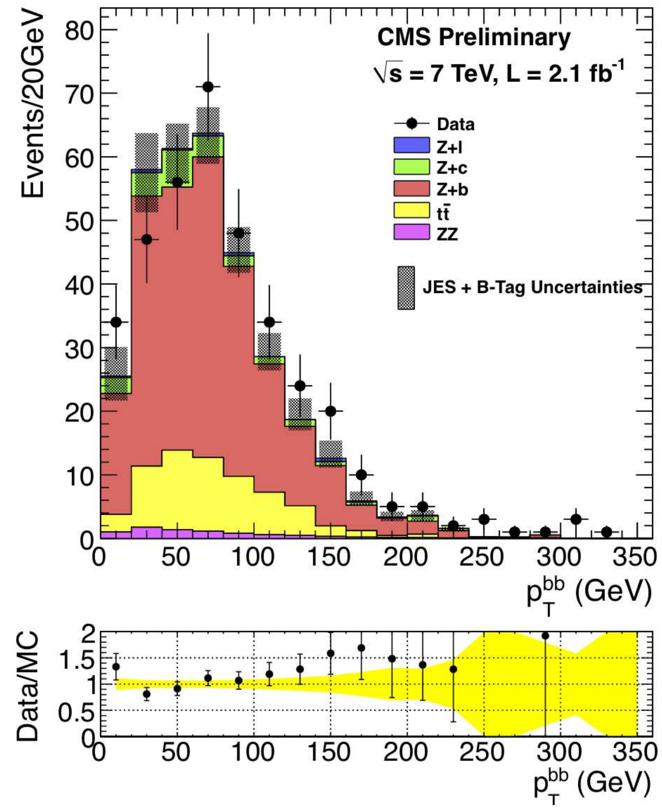
- In general, fair agreement between data and simulation but $p_T^{b\text{-jet}}$ in the data harder than in MG 5Fs simulation (similarly p_T^{Z}) → known limitations of the simulation: massless b in ME, missing higher order terms, decay tables not enough accurate

Z+b jets: kinematic variables at the detector level

- pp → Z/γ*+bb: studies limited by statistics (~0.24k signal events)







- Some excess of events in the data
  - at very low $p_T^{b\text{-jet}}$ both for leading and sub-leading jets
  - at high p_T of the bb pair

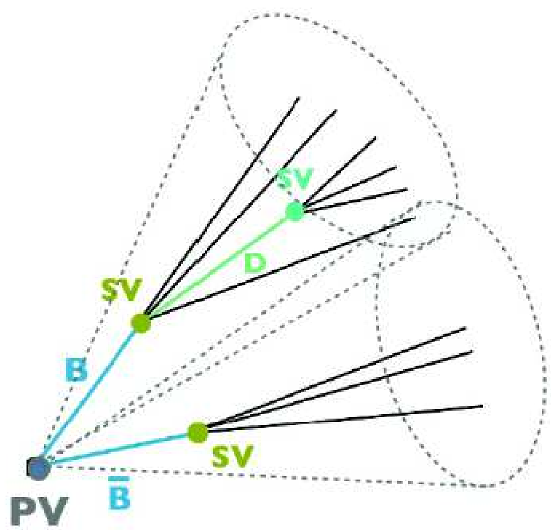
outside the uncertainty of the jet energy corrections (δp_T~3-5%)

Angular correlation between B-hadrons in events with Z+2 SV

- CMS PAS EWK-11-015

Z+2 secondary vertices

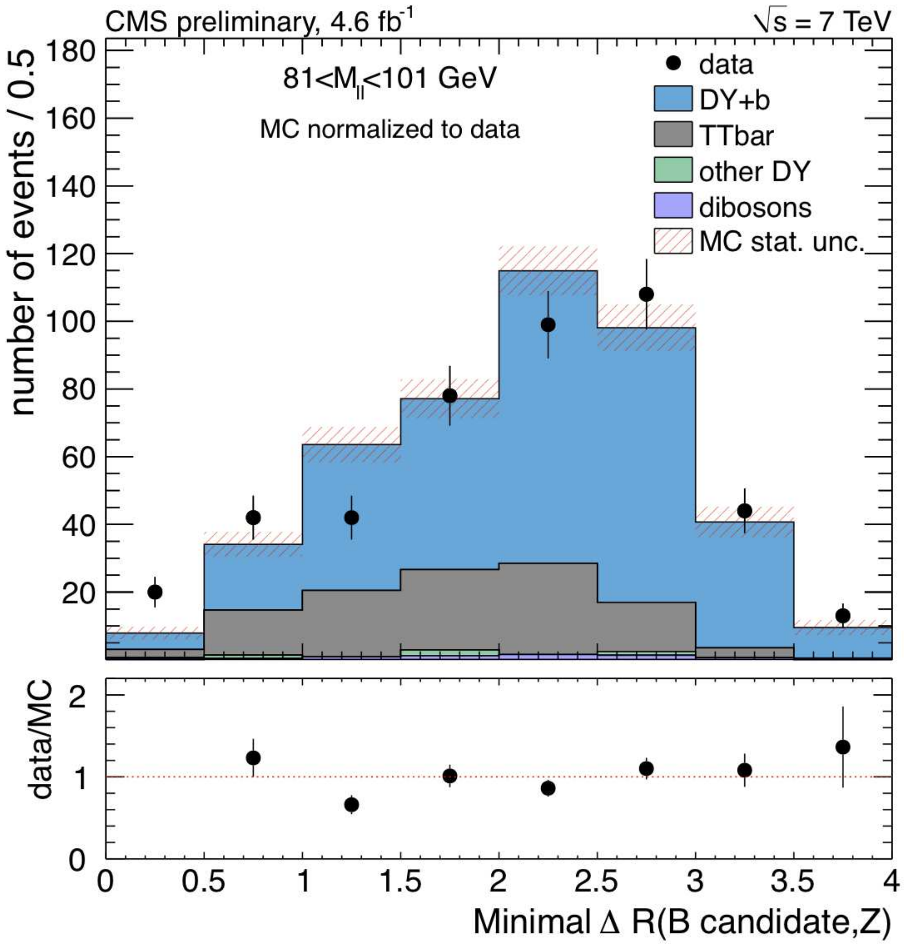
- Z/γ* from isolated leptons
  - $p_T^{\mu(e)}$>20(25) GeV
  - $|\eta^{\mu(e)}|$<2.1(2.5)
  - 60< $m_{\ell\ell}$< 120 GeV

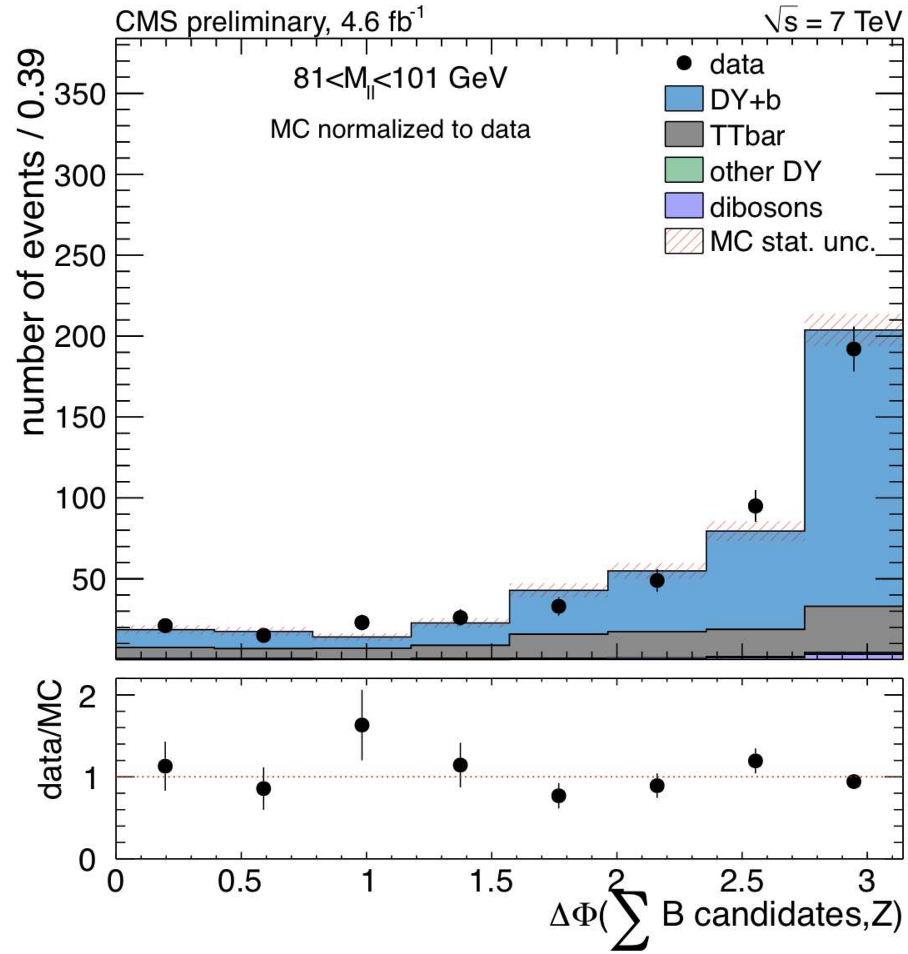


$$\Delta R=\sqrt{(\Delta\phi)^2+(\Delta\eta)^2}$$

- Presence of two secondary vertices found by the **I**nclusive **V**ertex **F**inder:
  - reconstruction of secondary vertices from seed tracks at large impact parameter with respect the primary vertex associated to the dilepton
  - 4-momentum of candidate B-hadron $p_{SV}=\sum p_i$ assuming $\pi^\pm$ tracks
  - candidates from sequential B⟶D+X decays merged on the basis of the angular separation ΔR(SV1,SV2), SV invariant mass, flight direction
- No use of jets
  - not limited by the size of the jet cone ⟶ investigation of the collinear region dominated by subprocess with $q\bar{q}$ initial state
  - measurement of data/MC scale factors for IVF efficiency more difficult ⟶ measurement of absolute cross section not available yet
- $|\eta|$<2, m(SV)>1.4 GeV p_T^{SV} > 8 GeV (≈$p_T^{B\text{-hadron}}$ > 15 GeV)

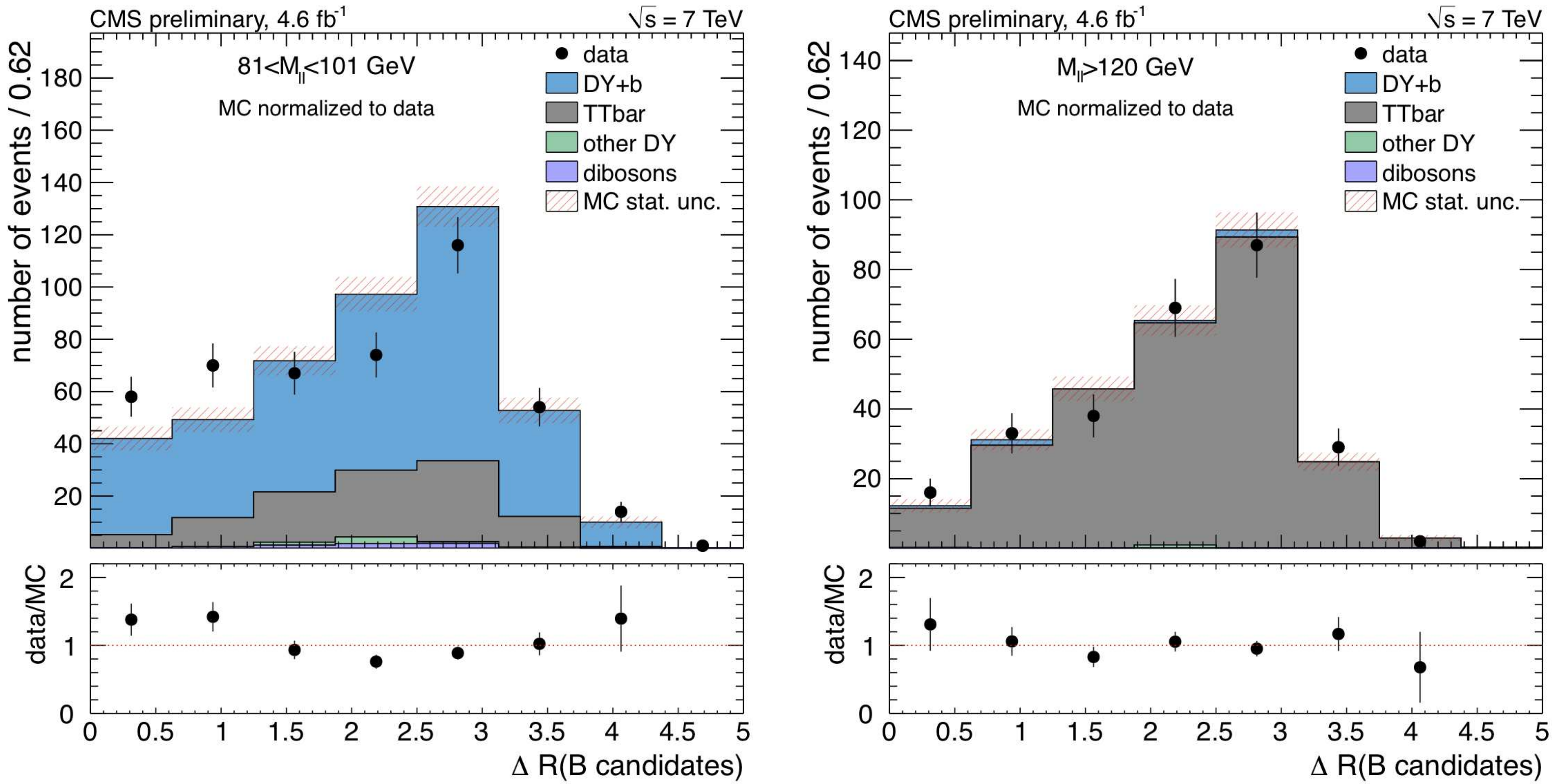
Z+2 SV: angular variables at the detector level





- Good description of extra QCD radiation
- Good control of $b\bar{b}$ events from a second parton-parton scattering

Z+2 SV: ΔR at the detector level



- From simulation σ(ΔR(B,B)-ΔR(SV,SV)) ~ 0.02 , e.g. ~1/30 binning size! → small bin-to-bin migration → no unfolding needed
- Robustness of the method verified in $t\bar{t}$ enriched region

Z+2 SV: ΔR at the particle level

- Measurement in eight bins of ΔR(B,B) correcting to the particle level phase space
  - $p_T^{B\text{-hadron}} > 15$ GeV and $|\eta^{B\text{-hadron}}| < 2.0$
  - $p_T^{\ell} > 20$ GeV, $|\eta^{\ell}| < 2.4$
  - $60 < m_{\ell\ell} < 120$ GeV dilepton (before FSR)

$t\bar{t}$ subtracted

IVF b-purity

$$\frac{1}{\sigma_{visible}} \frac{N_i^{data,fit} \times \mathcal{P}_i}{\varepsilon_i^{2SV} \times \varepsilon_i^{\ell} \times \mathcal{A}_i^{\ell}}$$

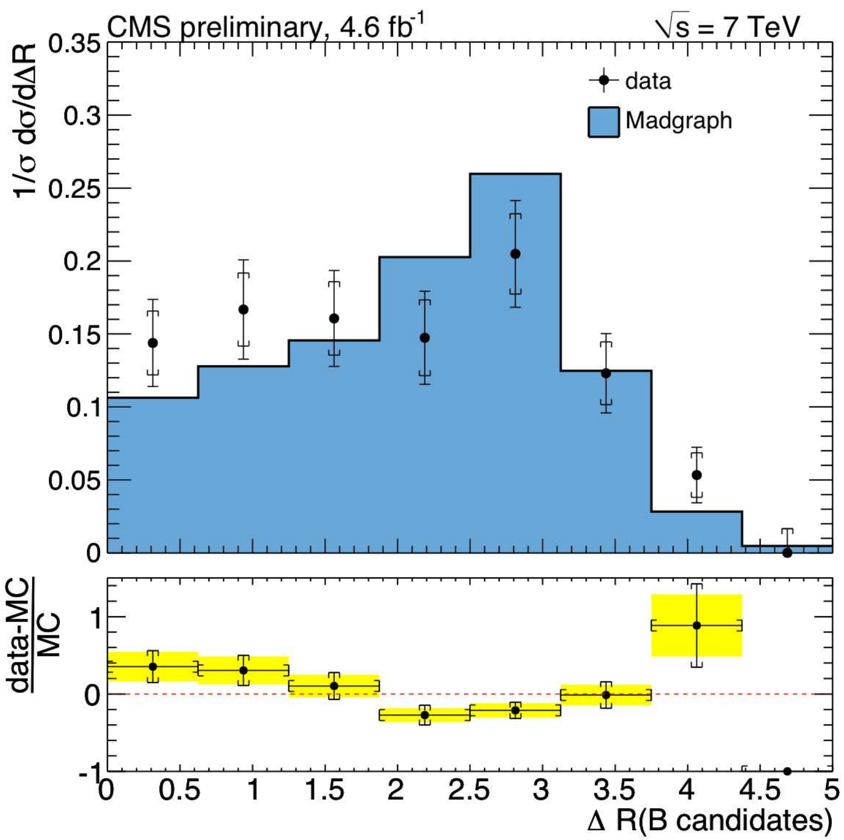
IVF efficiency

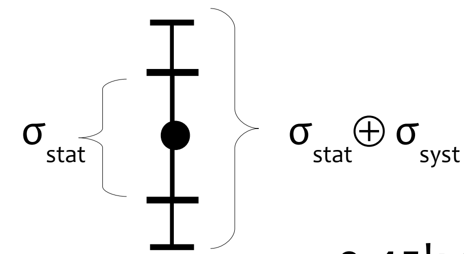
lepton-ID efficiency

lepton acceptance

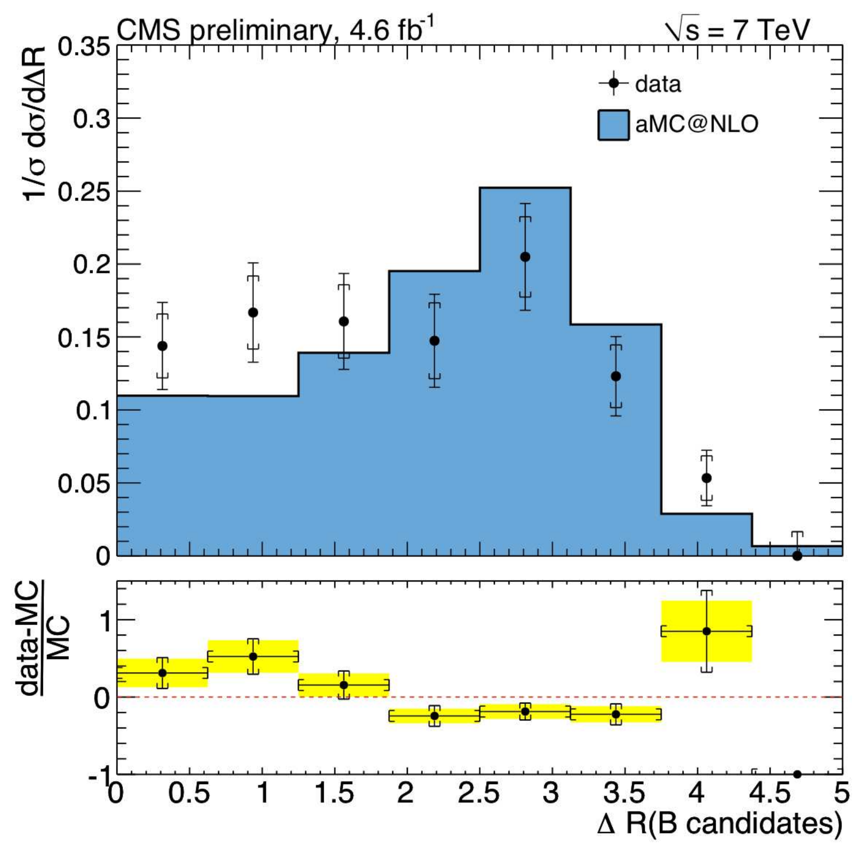
- Data-driven estimates:
  - number of signal events from template fit to dilepton $m_{\ell\ell}$ in ΔR(B,B) bins
- $\sigma_{visible} = \sum \sigma_i \longrightarrow$ normalized differential cross section

Z+2 SV: ΔR at the particle level





~0.45k signal events



- Systematics:
  - DATA-MC difference in $\langle p_T \rangle_{B\ sublead}$ and $\delta p_{T,B}$ from IVF (phase space) 9%
  - IVF purity 4%
- Data showing more flat behavior with respect simulation both with MG (LO) and aMC@NLO

Summary and outlook

- Z+b jets analyses performed so far in CMS have shown
  - Most of the visible events are properly described by a LO massless 5FS MC when an inclusive (e.g. jets of any flavor) LO-to-NLO k-factor is used
  - Some tensions in the description of the event kinematics
    - Z+1b inclusive: p_T^Z harder in data $\longrightarrow$ small benefit in adding additional data
    - Z+2b inclusive: excess at very low p_T^b in data and some discrepancy in p_T^{bb} $\longrightarrow$ need to be re-checked after analyzing the full 7 TeV sample
    - Z+2SV: $1/\sigma\, d\sigma/d\Delta R(B,B)$ more flat in data compared to both LO and NLO MC
      Next steps:
      - measure data/MC scale factors for the IVF efficiency and use $d\sigma/d\Delta R(B,B)$ to understand the origin of the difference in shape
      - extend analysis to additional angular variables and in bins of p_T^Z (e.g. boosted Z) to have better insight into different production mechanisms

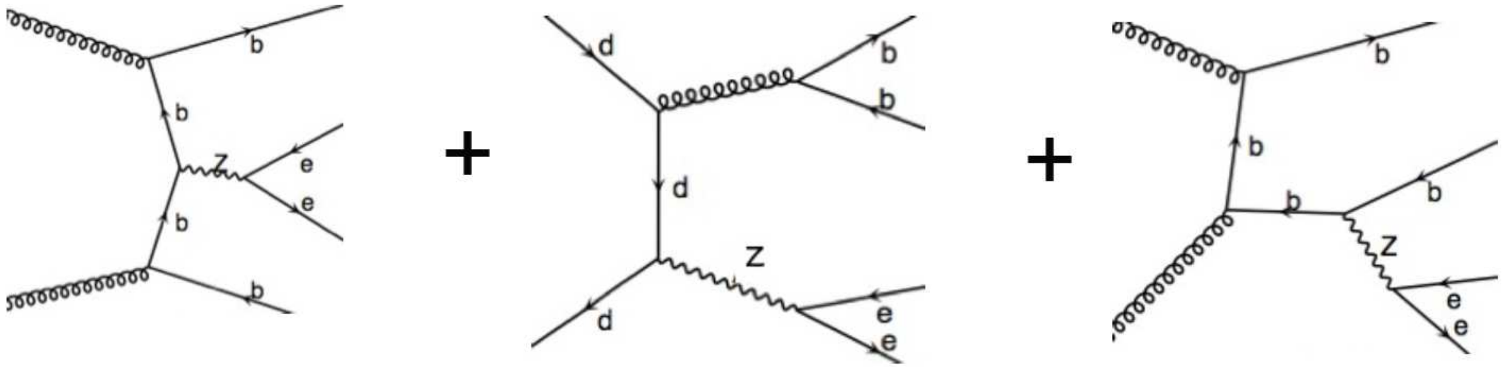
EXTRA-SLIDES

VH⟶Vbb scale factors

Table 6: Data/MC scale factors for each control region in each decay mode for the 2011 7 TeV and the 2012 8 TeV analyses. The errors include the statistical uncertainty from the fit, and a systematic uncertainty accounting for possible data/MC shape differences in the discriminating variables. Electron and muon samples in $Z(\ell\ell)H$ and $W(\ell\nu)H$ are fit simultaneously to determine average scale factors.

Process	$W(\ell\nu)H$	$W(\ell\nu)H$	$Z(\ell\ell)H$	$Z(\ell\ell)H$	$Z(\nu\nu)H$	$Z(\nu\nu)H$
Low p_T	7 TeV	8 TeV	7 TeV	8 TeV	7 TeV	8 TeV
W + udscg	$0.88 \pm 0.01 \pm 0.03$	$1.01 \pm 0.02 \pm 0.01$	–	–	$0.89 \pm 0.01 \pm 0.03$	$0.96 \pm 0.06 \pm 0.03$
$Wb\bar{b}$	$1.91 \pm 0.14 \pm 0.31$	$2.07 \pm 0.15 \pm 0.10$	–	–	$1.36 \pm 0.10 \pm 0.15$	$1.30 \pm 0.17 \pm 0.10$
Z + udscg	–	–	$1.11 \pm 0.03 \pm 0.11$	$1.10 \pm 0.02 \pm 0.06$	$0.87 \pm 0.01 \pm 0.03$	$1.15 \pm 0.07 \pm 0.03$
$Zb\bar{b}$	–	–	$0.98 \pm 0.05 \pm 0.12$	$1.08 \pm 0.04 \pm 0.08$	$0.96 \pm 0.02 \pm 0.03$	$1.12 \pm 0.10 \pm 0.04$
$t\bar{t}$	$0.93 \pm 0.02 \pm 0.05$	$1.07 \pm 0.01 \pm 0.01$	$1.03 \pm 0.04 \pm 0.11$	$1.01 \pm 0.02 \pm 0.06$	$0.97 \pm 0.02 \pm 0.04$	$1.05 \pm 0.07 \pm 0.03$
High p_T	7 TeV	8 TeV	7 TeV	8 TeV	7 TeV	8 TeV
W + udscg	$0.79 \pm 0.01 \pm 0.02$	$0.94 \pm 0.02 \pm 0.01$	–	–	$0.78 \pm 0.02 \pm 0.03$	$0.95 \pm 0.05 \pm 0.02$
$Wb\bar{b}$	$1.49 \pm 0.14 \pm 0.19$	$1.72 \pm 0.16 \pm 0.08$	–	–	$1.48 \pm 0.15 \pm 0.20$	$1.27 \pm 0.18 \pm 0.10$
Z + udscg	–	–	$1.11 \pm 0.03 \pm 0.11$	$1.10 \pm 0.02 \pm 0.06$	$0.97 \pm 0.02 \pm 0.04$	$1.04 \pm 0.07 \pm 0.02$
$Zb\bar{b}$	–	–	$0.98 \pm 0.05 \pm 0.12$	$1.08 \pm 0.04 \pm 0.08$	$1.08 \pm 0.09 \pm 0.06$	$1.15 \pm 0.10 \pm 0.04$
$t\bar{t}$	$0.84 \pm 0.02 \pm 0.03$	$0.99 \pm 0.01 \pm 0.01$	$1.03 \pm 0.04 \pm 0.11$	$1.01 \pm 0.02 \pm 0.06$	$0.97 \pm 0.02 \pm 0.04$	$1.03 \pm 0.07 \pm 0.03$

Z+bb: theory aspects



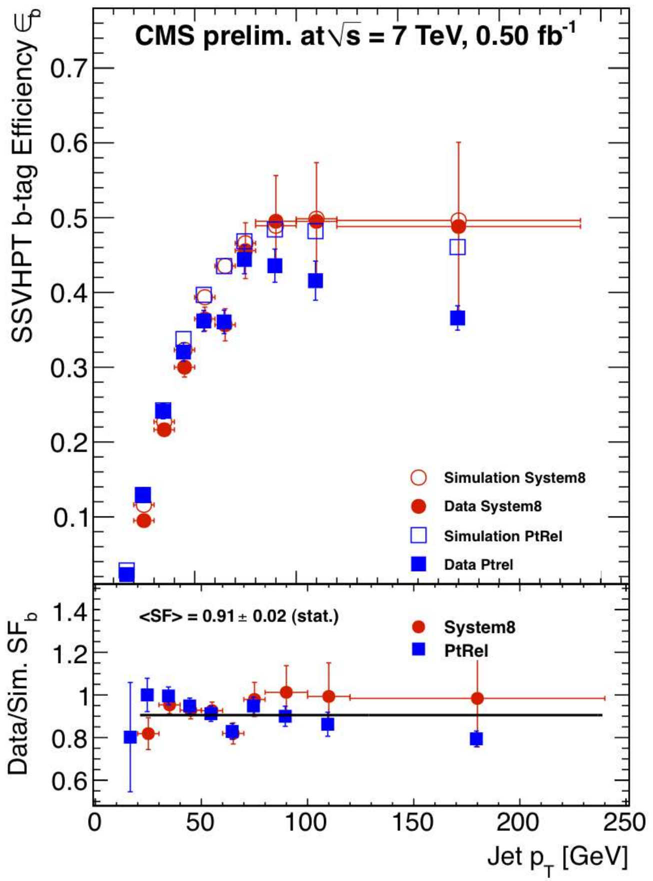
b quark fusion
mostly back-to-back

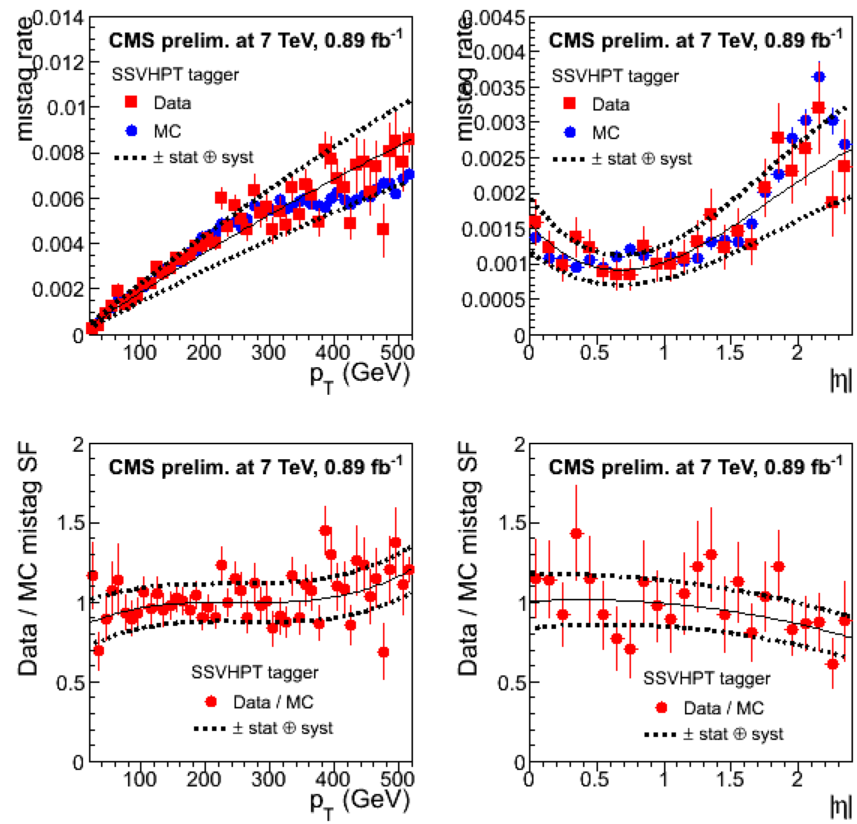
gluon splitting
mostly collinear

Z radiation

- MCFM settings
  - σ(`nproc=261; part=tota`) + σ(`nproc=266; part=real`)
  - `inclusive = .true.`
- MadGraph 5FS + Pythia6 settings
  - k_T MLM matching with matching scale k_T=20 GeV
  - hadron level jet is defined to be a b-jet if there is a B-hadron in ΔR=0.5 cone

b-tagging performance

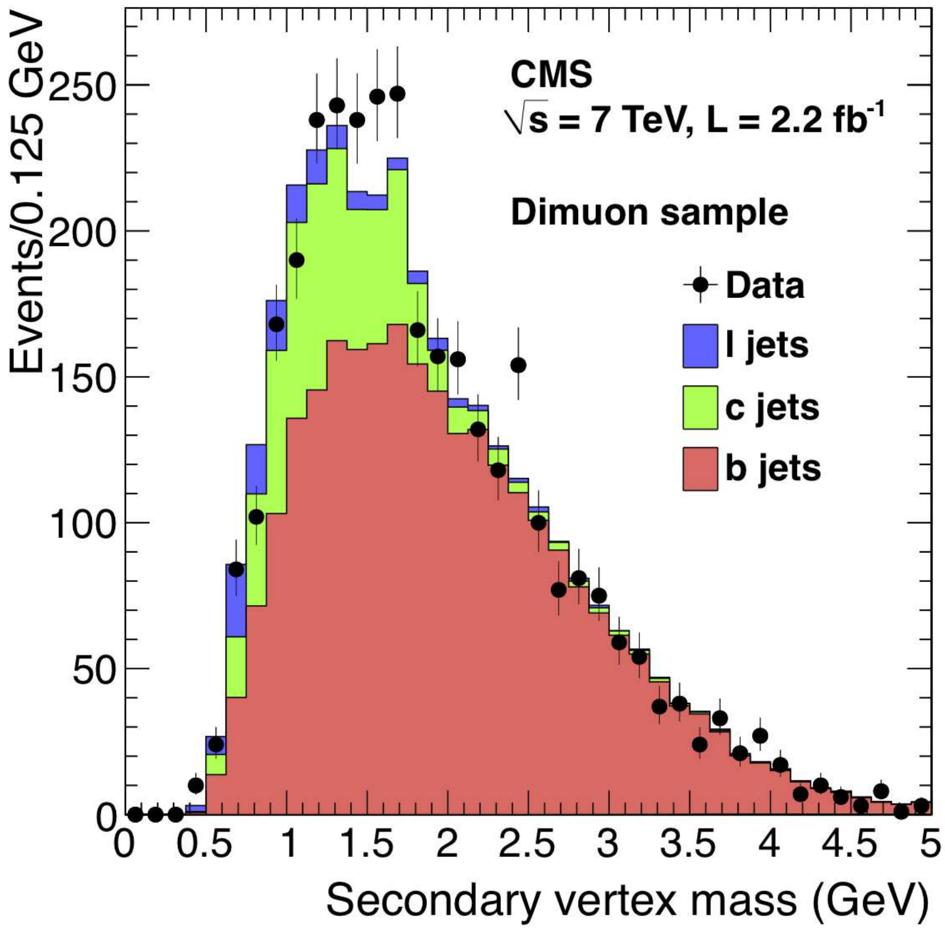


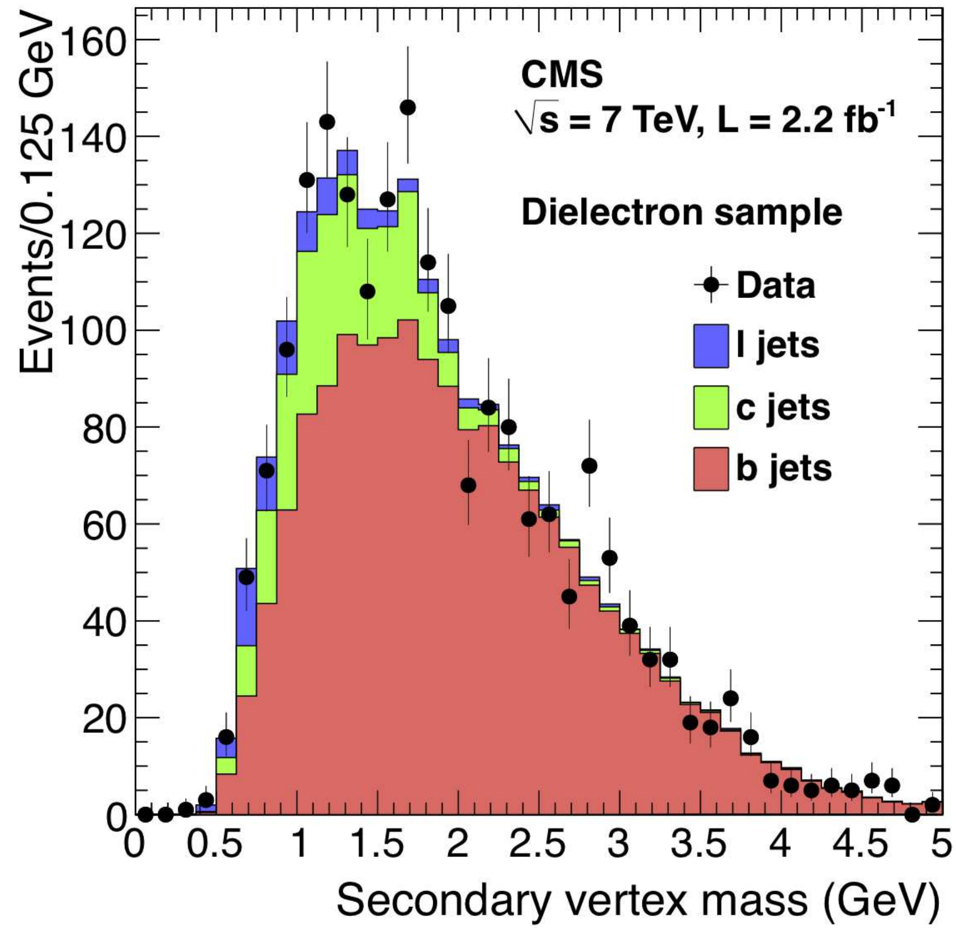


- Dataset: QCD multi-jets p_T>20 GeV
- Uncertainty on SF_b measurements for SSV HEM or HPT discriminants
  - p_T<30 GeV: ~ 12%
  - p_T>30 GeV: evolved from 10% (used in JHEP 06 (2012) 126) to ~3% (used in CMS PAS SMP-12-003)

pp ⟶ Z/γ*+b: data-driven estimates of P_b and f_{tt}

JHEP 06 (2012) 126





$$N^{est}_{t\bar{t}}(in) = \left(\frac{\mathcal{R}^{MC}_{t\bar{t}}}{\mathcal{R}^{MC}_{Z} - \mathcal{R}^{MC}_{t\bar{t}}}\right)\cdot(\mathcal{R}^{MC}_{Z}\cdot N_{obs}(out) - N_{obs}(in)) \qquad \mathcal{R}^{MC}_{t\bar{t}(Z)} = \frac{N^{MC}_{t\bar{t}(Z)}(in)}{N^{MC}_{t\bar{t}(Z)}(out)}$$

pp ⟶ Z/γ*+b: correction factors

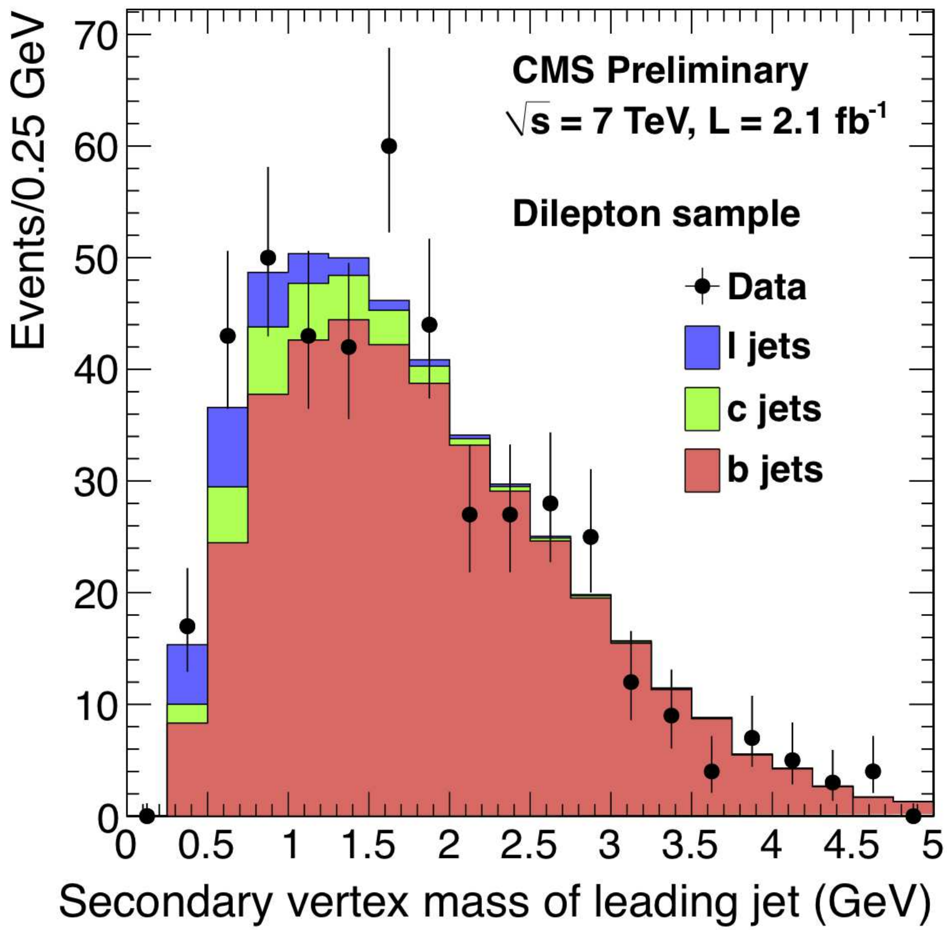
JHEP 06 (2012) 126

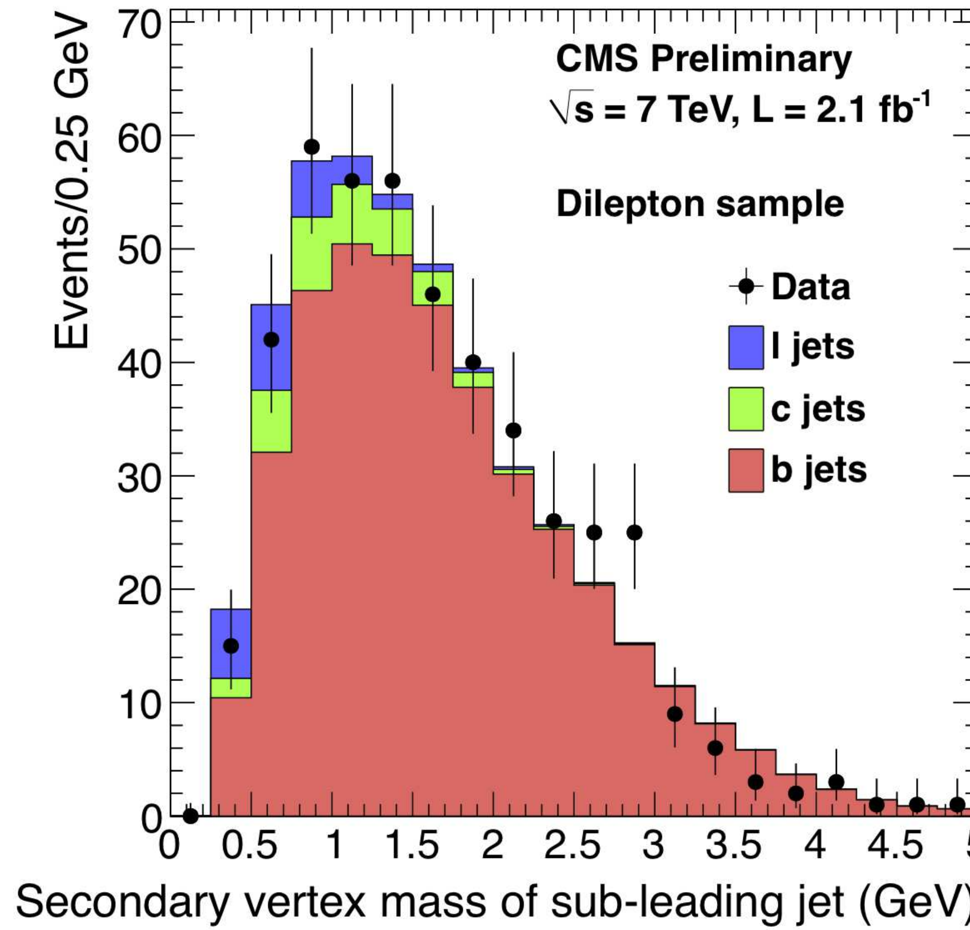
Table 1: Extraction of the cross section $\sigma_{\text{hadron}}(\text{Z}/\gamma^* + \text{b}, \text{Z}/\gamma^* \to \ell\ell)$ for $\ell\ell$ = ee or $\mu\mu$. The uncertainty on each parameter contains all the systematic effects considered in the analysis, summarised in Table 2 and detailed in the main text. The first uncertainty on the cross section results is statistical, while the second is systematic, and the third accounts for limitations of the theory.

Variable	ee+b	$\mu\mu$+b
Selected events	1990	3362
$\mathcal{P}$ (%)	83.4 ± 3.6	81.5 ± 2.9
$f_{\text{t}\bar{\text{t}}}$ (%)	18.7 ± 2.2	18.4 ± 2.3
ε_{b} (%)	35.3 ± 3.5	34.9 ± 3.5
ε_{ℓ} (%)	63.2 ± 2.6	84.4 ± 1.7
$\mathcal{C}_{\text{hadron}}$ (%)	$84.2^{+5.8}_{-0.6}$	$95.0^{+6.6}_{-0.5}$
$\mathcal{A}_{\ell}$ (%)	$55.0^{+3.6}_{-2.1}$	$57.2^{+3.7}_{-2.4}$
$\sigma_{\text{hadron}}(\text{Z}/\gamma^* + \text{b}, \text{Z}/\gamma^* \to \ell\ell)$ (pb)	$5.61 \pm 0.13 \pm 0.73^{+0.24}_{-0.53}$	$5.97 \pm 0.10 \pm 0.73^{+0.25}_{-0.57}$

pp ⟶ Z/γ*+bb

- 1D+1D fit of the masses of the secondary vertices





$P_{2b} = 1 - f_{cc} - f_{lb} - \ - f_{bl}$ with f_{cc}~6% and f_{bl}~5%

Multiple parton interaction

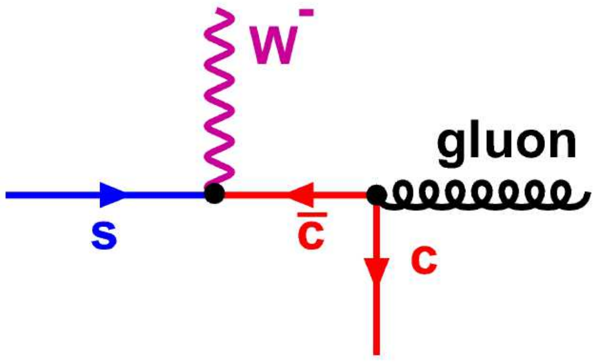
$$\sigma_{MPI} = \frac{m}{2} \times \frac{\sigma_A \times \sigma_B}{\sigma_{eff}}$$

- m=2 for "distinguishable processes"
- σ_{eff}=16±4 mb

	"process A"	σ_A	"process B"	σ_B	σ_{MPI}
Z+b jets	Z/γ* inclusive	3048 nb	$b\bar{b}$ with $p_T^{b\text{-}jet}$>25 GeV	700 nb	0.13 pb
Z+2 SV	Z/γ* inclusive	3048 nb	$b\bar{b}$ with $p_T^{B\text{-}had}$>15 GeV	15 μb	1 pb

W+c jets

- pp $\longrightarrow$ W+c dominated by sg $\longrightarrow W^- c$ and $\bar{s}g \longrightarrow W^+ \bar{c}$



$$R_c = \frac{\sigma(W + c + X)}{\sigma(W + jets + X)}$$

s-quark content of the proton

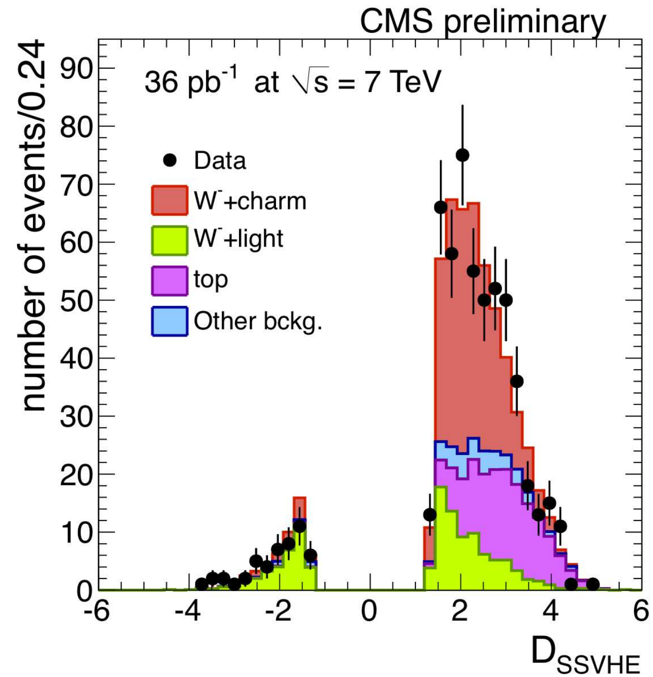
$$R_{+/-} = \frac{\sigma(W^+ + \bar{c} + X)}{\sigma(W^- + c + X)}$$

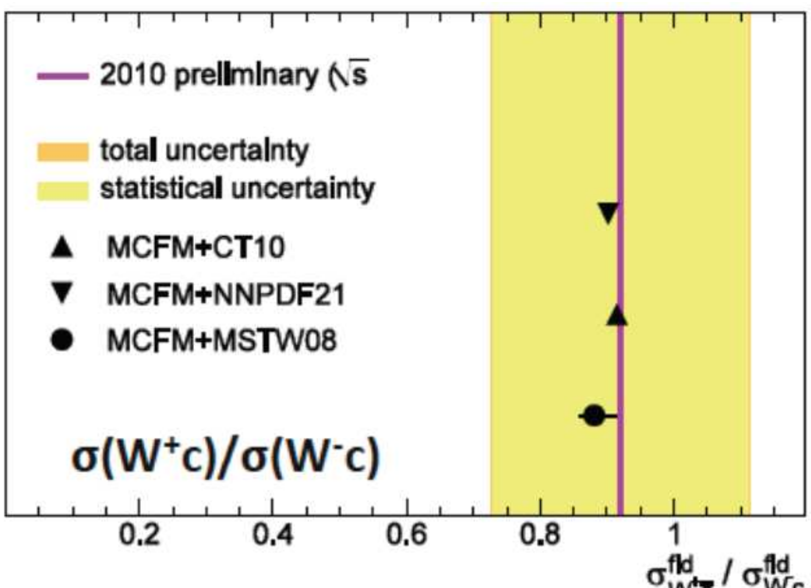
s vs. $\bar{s}$ asymmetry

- Cross section ratios defined at the detector level
  - p_T^{μ}>30 GeV
  - $|\eta^{\mu}|$<2.1
  - $M_T(\mu\nu)$> 50 GeV
  - p_T^{jet}>20 GeV
  - $|\eta^{jet}|$<2.1
  - $\Delta R(jet,\mu) > 0.3$

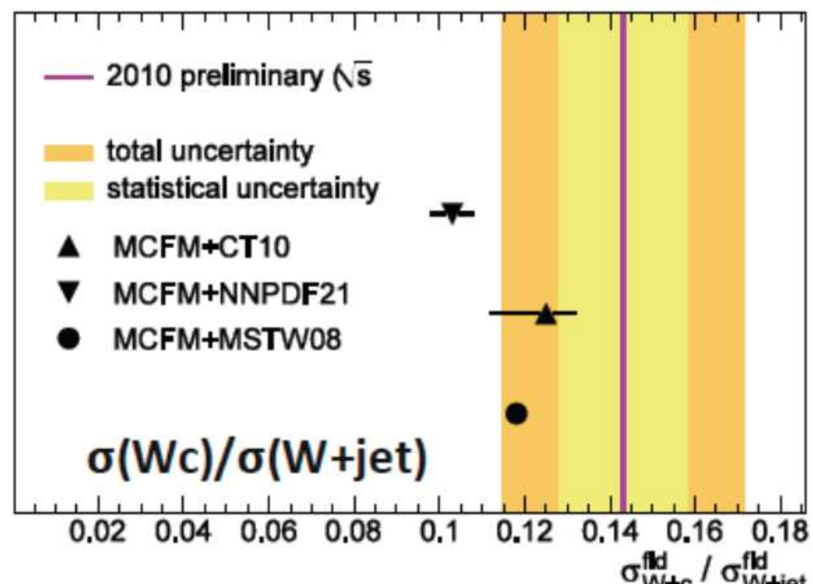
W+c jets

- W+c events (~10% of W+jets) selected by the presence of a detached secondary vertex with at least 2 tracks with
- Background subtracted W+c yield extracted from a Maximum Likelihood template fit to sign(D)*log(1+|D|) where $D=L_{3D}/\sigma_{L3D}$





$R_{\pm} = 0.92 \pm 0.19(stat) \pm 0.04(syst)$



$R_c = 0.143 \pm 0.015(stat) \pm 0.024(syst)$