Production of Z and W bosons in association with heavy quarks at CMS

Marco Musich[a]
for the CMS Collaboration

(a)INFN Sez. Torino

XX International Workshop on Deep-Inelastic Scattering and Related Subjects
26-30 March 2012, University of Bonn

DIS 2012
Bonn, Germany
26^{th}-30^{th} March 2012

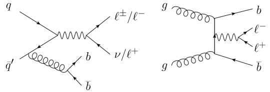
Outline

- The CMS experiment is well suited to look for V+heavy quarks signatures, thanks to its good b tagging performances and robust lepton reconstruction.
- Analysis of V+heavy quarks performed in CMS:
  - $Z + b(b)$, jet based (on 2011 dataset: 2.1 fb^{-1})
  - $Z + bb$, secondary vertex based (on 2011 dataset: 4.6 fb^{-1})
  - $W + c$ (on 2010 dataset: 36 pb^{-1})
- Motivations for these channels:
  - $Z + b(b)$ (main focus of this talk)
    - Measure cross-section with 1 b jet, 2 b jets, 2b in 1 jet $\Rightarrow$ confirm/constrain NLO cross-section and kinematics predictions.
    - Benchmark to $bb + \Phi$ discovery channel in MSSM with large tanβ
    - Background to new physics (e.g. $H \to ZZ \to 4\ell$, $pp \to Z(\ell\ell) + A(bb)$) and SM $Zh, h \to bb$.
  - $W + c \Rightarrow$ put constraints on strange quark *PDF* of the proton in the intermediate x region

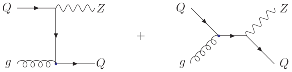
Motivations for detailed $Z + b(\bar{b})$ measurements

- Cross-section calculations: two different schemes should agree at NLO. Preliminary results show difference up to 30% between them. $Z + b$ cross-section measurement could help to resolve it.

- **Fixed flavor scheme** (arXiv:hep-ph/1106.6019)



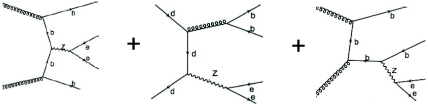
- **Variable flavor scheme** (arXiv:hep-ph/0312024)



- Massive *b*
- Full event description → `aMC@NLO`

- Splitting inside proton PDF
- Massless b → Collinear approach

- Several production mechanisms at the LHC: b quarks fusion, gluon splitting, Z radiation, and Multiple Parton Interactions



- Angular correlations of *bb* pair → comprehension of production mechanism

Cross section definition for $pp \to Z/\gamma^* + b$ (CMS EWK-11-012)

The $pp \to Z/\gamma^* + b, Z/\gamma^* \to \ell\ell$ cross-section is calculated as:

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

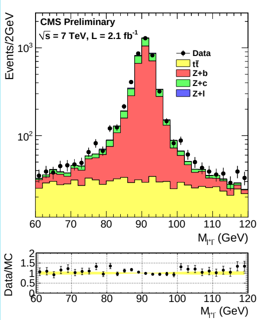
- $N_{\ell\ell+b}$ is the selected number of di-leptons+b events
- $\mathcal{P}$ is the b-jet purity, $f_{t\bar{t}}$ contamination of $t\bar{t}$ events
- ϵ_ℓ and ϵ_b are lepton and b-tagging efficiencies (computed scaling simulation to match efficiencies in data)
- $\mathcal{A}_\ell$ is the lepton acceptance
- $\mathcal{C}_{hadron}$ a correction factor for detector and reconstruction effects
- $\mathcal{L}$ is the luminosity

The cross section is defined with the following requirements:
(i) $\mathbf{p_T^b > 25}$ **GeV and** $|\eta^{\mathbf{b}}| < \mathbf{2.1}$ on any hadron jet containing a b hadron;
(ii) $\mathbf{60 < m_{\ell\ell} < 120}$ **GeV** for the invariant mass of the di-lepton;
(iii) $\mathbf{\Delta R(jet, \ell) > 0.5}$ angular separation between any jet and the Z leptons

Event selection for $Z/\gamma^* + b$ jets

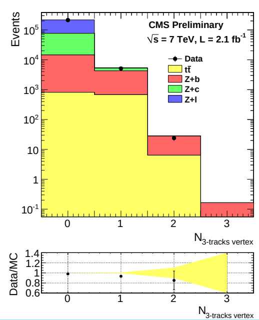
Z from isolated leptons:

- $p_T^{e(\mu)} > 25(20)$ GeV
- $|\eta_{e(\mu)}| > 2.5(2.1)$
- $60 < m_{\ell\ell} < 120$ GeV
- ID criteria + match with HLT objects



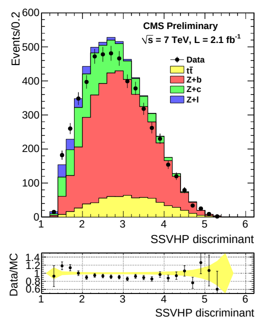
Jet selection:

- anti-$k_T(R = 0.5)$
- $p_T > 25$ GeV
- $|\eta| < 2.1$
- jet lepton separation $\Delta R(\ell, jet) > 0.5$



b-tagging:

- Detached secondary vertex (with at least 3 tracks)
- 1 b-tag eff. $\simeq 35\%$
- $udsg$-mistag $\simeq 0.1\%$



Data-MC comparison for $pp \to Z/\gamma^* + b$

- After rescaling simulation to match data efficiencies, good agreement in overall normalization is found between Data and LO MonteCarlo (`Madgraph`) rescaled to the NLO cross-section. *Some discrepancies in the shapes of kinematic variables*.

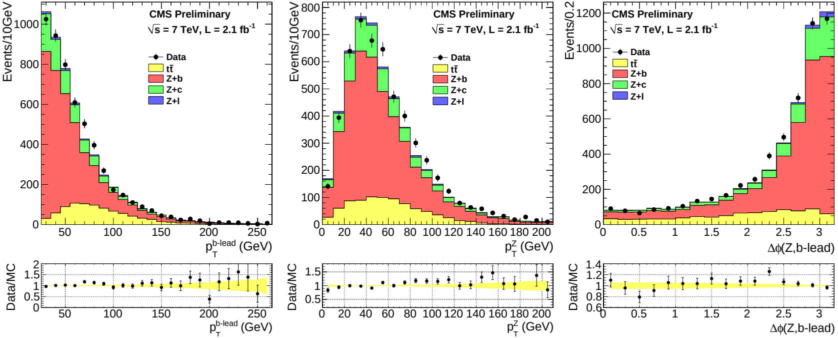


Figure: p_T of the leading-p_T b (left), p_T of the dilepton pairs and (right) $\Delta\phi(Z$,leading-pT b jet) after the high purity b-tagging selection. The yellow bands in the lower plots represent the statistical uncertainty in the MC yield.

Estimation of the backgrounds (Z/γ^*+light or charm and $t\bar{t}$)

- The **$\mathbf{Z} + \mathbf{ucdsg}$** background is estimated assessing the b-purity of the Z + tagged jets sample. The (data driven) extraction of the purity $\mathcal{P}$ is based on a template fit of the mass of the secondary vertex of the tagged jet

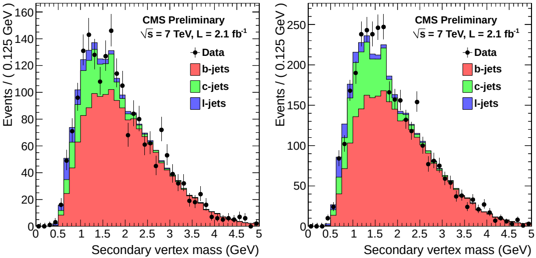


Figure: Secondary vertex mass in Z+tagged jet sample. Left: electron channel, right: muon channel

- The **$\mathbf{t\bar{t}}$** background is extracted from extrapolation of upper sideband of $m_{\ell\ell}$ under the signal region:

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

Extraction of the $Z/\gamma^* + b$ cross-section and comparison with theory

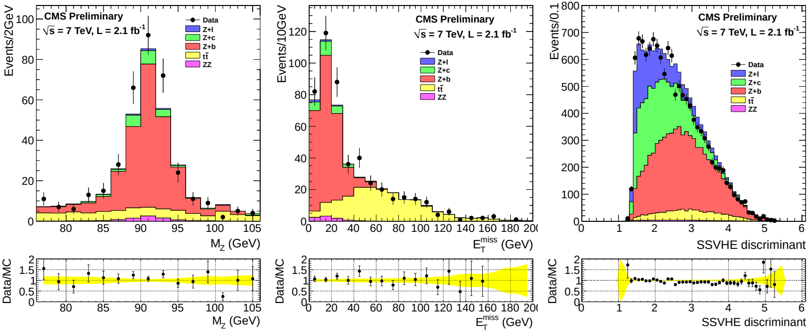
- The detector level yield after correcting for lepton (ϵ_ℓ) and b-tagging (ϵ_b) efficiencies is further corrected with ($\mathcal{A}_\ell \cdot C_{hadron}$) taking into account the lepton acceptance and detector resolution
- The final cross-section is calculated separately in the $Z \to \mu\mu$ and $Z \to ee$ channels, which are found in agreement and then combined:

	cross-section (pb)
$\sigma_{hadron}(Z + b, Z \to \ell\ell)$	6.10 $\pm$ 0.08 (stat) $\pm$ 0.79 (syst) +0.25-0.57 (theory)

- Parton level NLO calculations obtained with the MCFM tool $\Rightarrow \sigma^{parton}_{MCFM} = 4.73 \pm 0.54$ pb
- Parton-to-hadron level correction C_{NP}=0.84$\pm$0.03 is computed using MADGRAPH+PYTHIA and aMC@NLO+HERWIG.
- Final hadron-level-corrected NLO prediction $\sigma^{hadron}_{MCFM} = 3.97 \pm 0.47$ pb
- The measured cross-section in data is found smaller than the hadron-level-corrected NLO prediction

Event selection for $Z/\gamma^* + bb$ (CMS SMP-12-003)

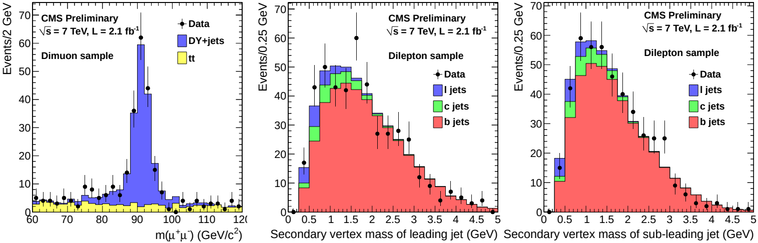
- Same lepton and jet selection as for the inclusive case
- Use high efficiency discriminant (1 b-tag eff. $\simeq 55\%$,*udsg*-mistag $\simeq 1\%$) to keep higher yields $\Rightarrow$ higher $t\bar{t}$ content in the selection
- signal defined in a smaller region in di-lepton mass $76 < m_{\ell\ell} < 106$ GeV and $\not{E}_T < 50$ GeV cut reduces $t\bar{t}$ background



Background estimation for $Z/\gamma^* + bb$

- **$t\bar{t}$**: suppressed with $\not{E}_T$ and $m_{\ell\ell}$ cuts. Estimated with fit to $m_{\ell\ell}$
- **$Z + $ ucdsg**: data driven (b-purity) estimated with two 1D simultaneous template fits to secondary vertex mass distributions. Event purity:

$$\boxed{f_{bb} = 1 - f_{cc} - f_{bl} - f_{lb}}$$



ZZ: from MC normalized to data using CMS cross section σ_{ZZ} measurement

	$\mu\mu + bb$	$ee + bb$
Yields	219	148
bb-purity	$(83 \pm 6)\%$	$(83 \pm 6)\%$
$t\bar{t}$ frac.	$(20 \pm 5)\%$	$(17 \pm 5)\%$
$N_{Z(\ell\ell)Z(bb)}$	5.2 ± 0.2	3.0 ± 0.2

Measurement of the $Z/\gamma^* + bb$ cross-section in multiplicity bin

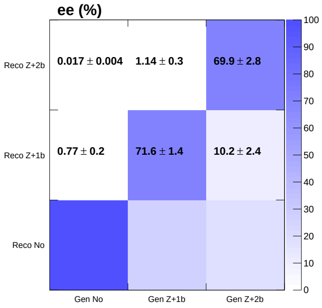
- $Z + 1b$ and $Z + 2b$ yields at hadron level: unfold b-jet multiplicity
- number b-tagged jets $\rightarrow$ reconstructed b-jets $\rightarrow$ hadron-level b-jets
- unfold via the matrix equation:

$$N^{gen}_{Z+ib} = [(\mathcal{A}_\ell^{-1})_{ij}]_{optional} \times (\epsilon_r^{-1})_{jk} \times (\epsilon_\ell^{-1})_{kl} \times (\epsilon_b^{-1})_{lm} \times N^{tag}_{Z_{sel}+mb}$$

- The cross-sections in b jets multiplicity bins are found to be:

Multiplicity bin	$\sigma \pm \delta_\sigma^{stat} \pm \delta_\sigma^{syst} \pm \delta_\sigma^{theo}$ (pb)
$\sigma_{hadr}(Z(\ell\ell) + 1b)$	3.41 $\pm$ 0.05 $\pm$ 0.27 $\pm$ 0.06
$\sigma_{hadr}(Z(\ell\ell) + 2b)$	0.37 $\pm$ 0.02 $\pm$ 0.07 $\pm$ 0.02
$\sigma_{hadr}(Z(\ell\ell) + b)$	3.78 $\pm$ 0.05 $\pm$ 0.31 $\pm$ 0.08

- Cross-sections reported here are the combination of muon and electron channels and are corrected for common minimal lepton acceptance A_ℓ ($p_T^\ell > 20$ GeV and $|\eta|^\ell < 2.5$)



Migration matrix between generator and detector level yields

Data MC comparison of $Z/\gamma^* + bb$

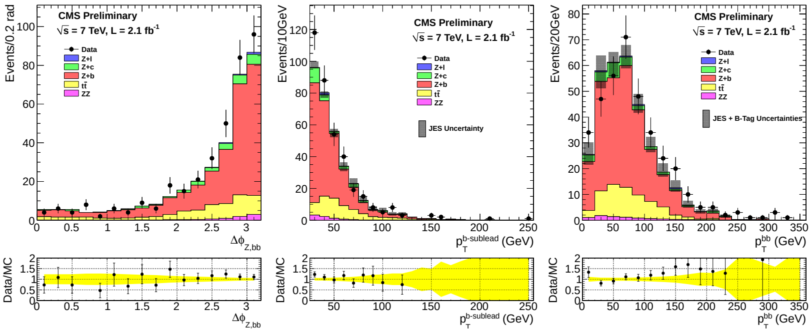


Figure: Left: $\Delta\phi(Z,bb$ pair), center p_T of the sub-leading jet, and p_T distribution of the bb pair after the baseline selection. The black band in upper plots is the systematic uncertainty due to jet energy uncertainty and b-tagging uncertainty. The yellow bands in the lower plots is the statistical uncertainty in the MC yield.

- Some discrepancies in the data/MC comparison of kinematic properties $\Rightarrow$ need further studies, NLO MC simulations.

BB angular correlations with $Z/\gamma^* + bb$ (CMS EWK-11-015)

- The identification of displaced secondary vertices with no use of jets allows to study BB pair production also at small angular separation
- B hadron candidates are reconstructed using the *Inclusive Vertex Finder (IVF)*. Using high impact parameter tracks seeds, additional tracks clustered in L_{3D}/σ_{3D} and angular separation ΔR. Secondary vertices from charm cascade decays are merged to B decays
- Very good angular resolution in B hadron flight direction ($\simeq 0.02$ in ΔR)

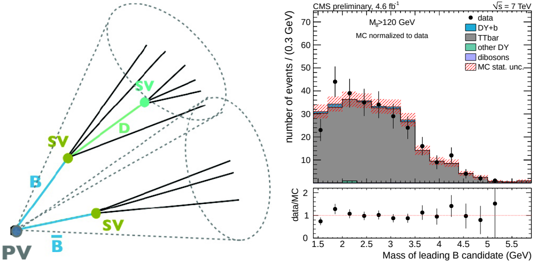


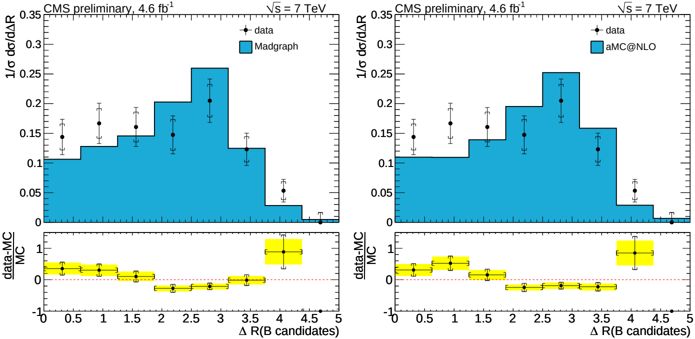
Figure: Left: sketch of the IVF tool, Right: mass of the leading B candidate, in the top dominated region $m(\ell\ell) > 120$ GeV

$\Delta R(BB)$ angular distribution shape in Z/γ^*+bb events

- The $\Delta R(BB)$ shape is calculated via:

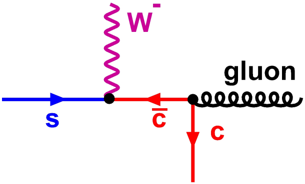
$$\frac{1}{\sigma}\frac{d\sigma}{d\Delta R(BB)} \Rightarrow \frac{1}{\sigma_{visible}}\frac{N_i^{data,fit}\cdot\mathcal{P}_i}{\epsilon_i^{2SV}\cdot\epsilon_i^{\ell}\cdot\mathcal{A}^{\ell}} \qquad i = \Delta R \text{ bin}$$

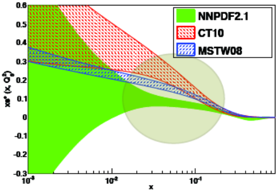
- In each ΔR bin i, $N^{data,fit}$ is extracted from a M.L. fit to $m_{\ell\ell}$
- $N^{data,fit}$ corrected for the IVF purity $\mathcal{P}_i$, the IVF efficiency ϵ_i^{2SV} and the dilepton efficiency and acceptance $(\epsilon_i^{\ell}\cdot\mathcal{A}^{\ell})$. Each bin normalized by $1/\sigma_{visible}$
- Comparison with LO and NLO predictions shows dicrepancies with both



W+charm measurement motivations (CMS EWK-11-013)

- Process $pp \to W + c + X$ sensitive to the proton strange quark content
- At the LHC it is dominated by $\bar{s}g \to W^+\bar{c}$ and $sg \to W^-c$
- Processes like $\bar{d}g \to W + \bar{c}$ and $dg \to W^-c$ are Cabibbo disfavoured
- Processes with b quarks in the final state are even more suppressed $(1 \div 2\%)$
  $\Rightarrow$ more than 10% of the $W + jets$ events at the LHC with $p_T^j > 20$ GeV, contain c jets





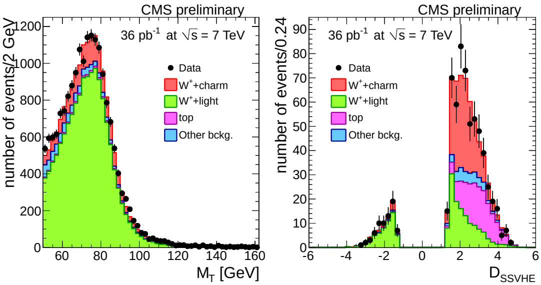
- The cross section ratios:

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

provide important information on the strange and anti-strange quark parton density functions of the proton at the electroweak scale

W+charm cross-section ratios measurements

- The ratios are measured using $W \to \mu\nu$ decays in the kinematic region $p_T^j > 20$ GeV, $|\eta_j| < 2.1$ and with $p_T^\mu > 25$ GeV, $|\eta_\mu| < 2.1$, $M_T(\mu\nu) > 50$ GeV.



Discriminator based on the significance of a 2-track secondary vertex decay length

$$D_{SSVHE} = sign(S) \cdot \log(1 + |S|)$$
$$S = L_{3D}/\sigma_{L_{3D}}(SV)$$

- the charm fraction in the selected $W + jets$ sample extracted from a M.L. fit to the different components of the distribution of the D_{SSVHE} discriminator

$R_{+/-}$	0.92 $\pm$ 0.19 (stat.) $\pm$ 0.04 (syst.)
R_c	0.143 $\pm$ 0.015 (stat.) $\pm$ 0.024 (syst.)

- Results are in agreement with theoretical predictions at next-to-leading order based on available parton distribution functions.

Conclusions

- Using 2010 and 2011 CMS collision data at $\sqrt{s} = 7$ TeV, preliminary measurements of the $Z + b(b)$ cross sections and angular correlations have been presented. The cross-sections have been measured with a total uncertainty of $\simeq$ 1.3%(stat.)$\oplus$12%(syst.) for the inclusive $1b$ case and of $\simeq$ 5%(stat.)$\oplus$19%(syst.) for the inclusive $2b$ case
- Some discrepancies in kinematic properties with LO Monte Carlo prediction have been found
- The study of the BB angular correlations in $Z + bb$ events shows partial disagreement with both LO and NLO predictions
- Finally, the study of the $W + c$ production rate with respect to the W charge and W+light jets rates, measured with a total uncertainty of $\simeq$ 20 %, has also been presented

CMS EWK-11-012 Measurement of the $Z/\gamma^* + b$-jet cross section in pp collisions at $\sqrt{s} = 7$ TeV

CMS SMP-12-003 Measurement of the $Z/\gamma{*}{+}bb$-jets cross section in pp collisions at $\sqrt{s} = 7$ TeV

CMS EWK-11-015 Angular correlation between B hadrons produced in association with a Z boson in pp collisions at $\sqrt{s} = 7$ TeV

CMS EWK-11-013 Measurement of associated charm production in W final states at $\sqrt{s} = 7$ TeV

Backup slides

Calculation of lepton and b tagging efficiencies

Lepton efficiencies with Tag & Probe

- Z mass constraint from a pair of same flavor leptons: **one tag** (high purity), the other **probe** (to measure efficiency given a criterium)
- After di-lepton+jet event topology $\Rightarrow$ **Tag = tight selected lepton**

$$\epsilon_{lepton} = \epsilon_{trk} \times \epsilon_{(reco|trk)} \times \epsilon_{(id|reco)} \times \epsilon_{(iso|id)} \times \epsilon_{(trg|iso)}$$

- Double lepton trigger efficiency: on each leg separately, tag matched to loosest unprescaled single lepton trigger.

$\epsilon_{trig} = (\epsilon_{L1} \times \epsilon_{H2}) + (\epsilon_{L2} \times \epsilon_{H1}) - (\epsilon_{H1} \times \epsilon_{H2})$ L=Low threshold, H=high threshold

b-tag efficiencies

- Estimate per jet efficiency, measured on signal MC sample reweigthed to reproduce per-jet btagging efficiency measured in data
- use data/MC scale factors SF_b,SF_c,SF_l; SF $= \epsilon_{DTA}/\epsilon_{MC}$
- MC b-tag efficiency and c-mistag: from $Z + b$ and $Z + c$ MC sample
- MC event weight calculated taking into accound all possible mistag combinations

Comparison of W+charm results with theoretical predictions

- Constraint of the PDF from the ratios:

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

- CMS measurement has still large experimental uncertainties, but clearly towards discrimination between PDF sets
- Strangeness is where larger differences between PDF sets are found
- Experimental uncertainties to be reduced with the 2011 dataset