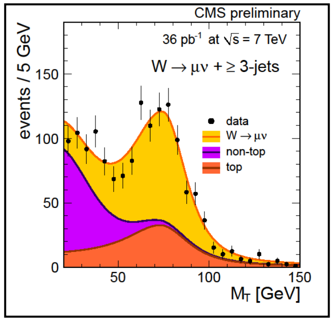
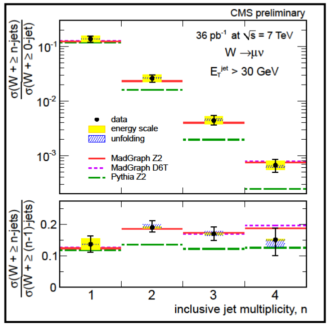
W/Z+jets with the CMS detector

Massimo Nespolo
on behalf of
CMS Collaboration

Jet reconstruction and spectroscopy at hadron collider (Pisa)
April 18-19, 2011

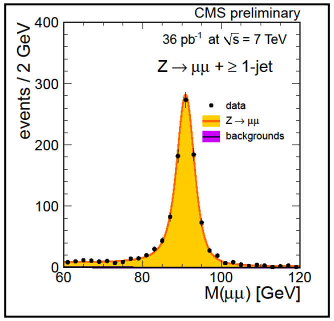
W/Z+jets with the CMS detector

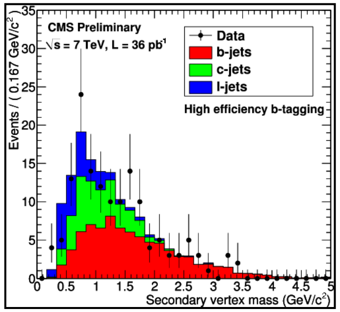




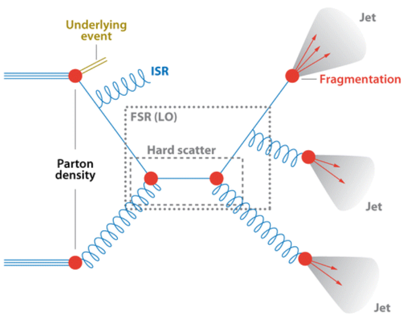
Outline:

- Why study *W/Z*+jets:
  - Check description of *QCD*.
  - Background for Higgs and BSM.
- Method:
  - Selections, fits, systematics…
- Results (with 36 pb^{-1}, all 2010 data):
  - Ratios of rates.
  - Berends-Giele scaling.
  - *Z* +*b* production.



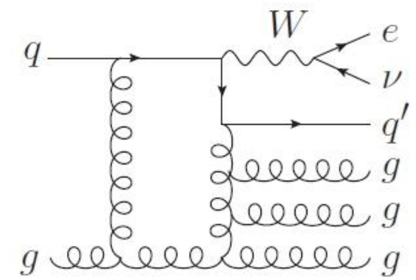


Motivation and goal



ME+PS (tree):
Normalization?
Scale?

NLO: recently done up to 4(3) jets for *W*(*Z*)



Idea: use events tagged by *W/Z* to ***probe QCD*** in a clean environment

Provide ***bkg normalization*** for lepton+MET+jets final state:

1. Top, Higgs.
2. BSM (e.g. SUSY).

Analysis ***potentially sensitive*** to new physics by itself:

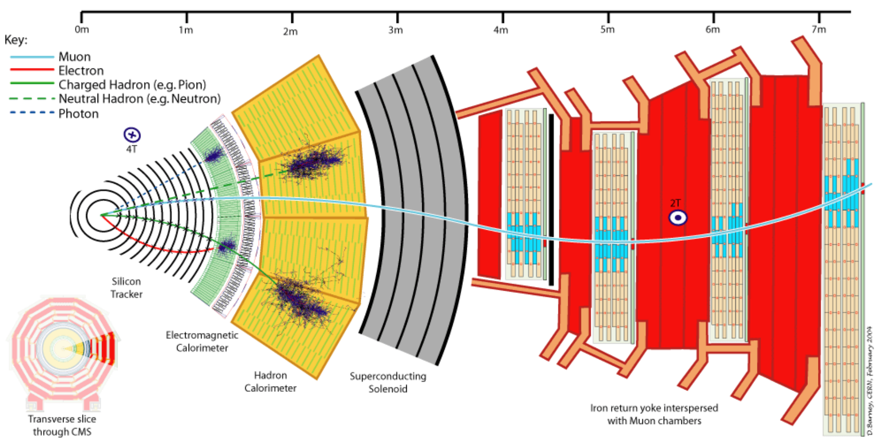
1. Excess of jets at high-p_T.
2. Deviation from scaling.

Signal selection (muon channel)

Trigger: 1 μ with p_T > (9-15) GeV, depending on luminosity

Requires one muon:

1. Well fitted from hits in tracker and muon chambers, passing ID to reject punch-through and decay in-flight.
2. $p_T > 20$ GeV, $|\eta| < 2.1$.
3. Isolated: $(\Sigma p_T(tk) + \Sigma E_T(had+em))/p_{T\mu} < 15\%$.



Is there a second μ?

- $p_T > 10$ GeV, $|\eta| < 2.4$
- $60 < M_{\mu\mu}$ (GeV) < 120

Yes: *Z* sample

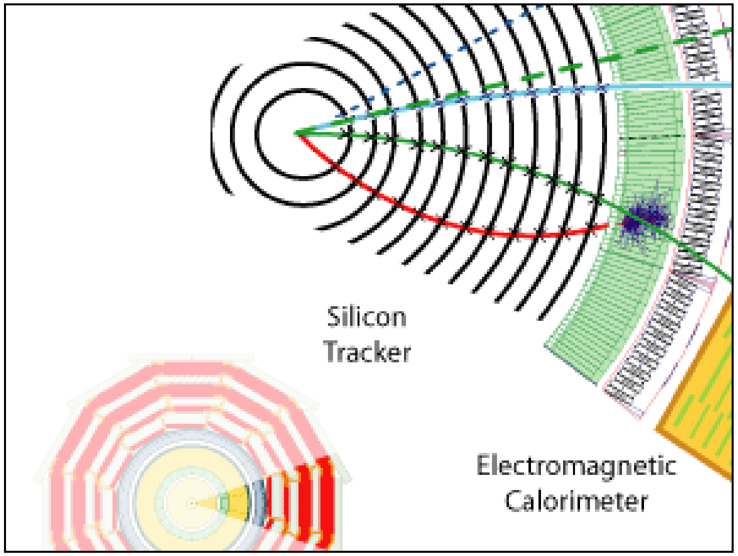
No (M_T>20 GeV): *W* sample

Signal selection (electron channel)

Trigger: 1 e with p_T > (10-17) GeV, according to increasing luminosity

Requires one electron:

1. With ***tight*** (80% efficient) isolation, electron identification, and conversion rejection criteria satisfied.
2. $p_T > 20$ GeV, $|\eta| < 2.5$ ($1.44 < |\eta| < 1.57$ ***excluded***).
3. Matched to the trigger primitive.



Is there a second e?

- $p_T > 10$ GeV, $|\eta| < 2.5$
- $60 < M_{ee}$ (GeV) < 120

Yes (***loose*** ID): *Z* sample

No (M_T>20 GeV): *W* sample

Jets reconstruction

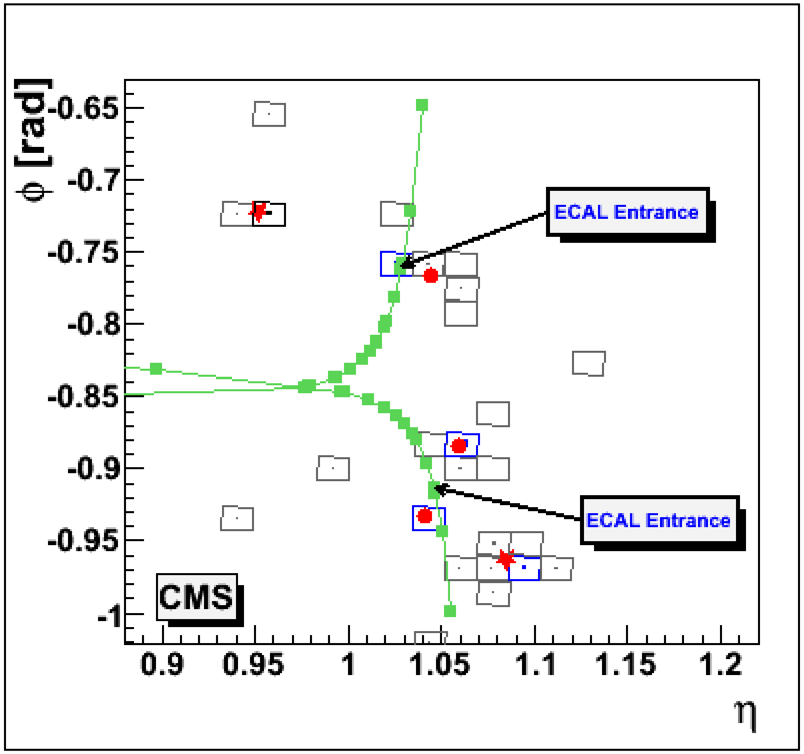
- Algorithm: ***anti-KT with ΔR = 0.5*** (default in CMS).
- Input: list of "particles", identified by Particle Flow.

- E_T > 30 GeV, $|\eta| < 2.4$ (tracker acceptance).
- Loose ID to ***remove noise*** of calorimeters.
- Data-driven ***energy calibration*** applied.
- ***Pile-up*** jet-energy offset removed with *FastJet*.

Leptons from VB-decay have ***not*** be counted as jets:

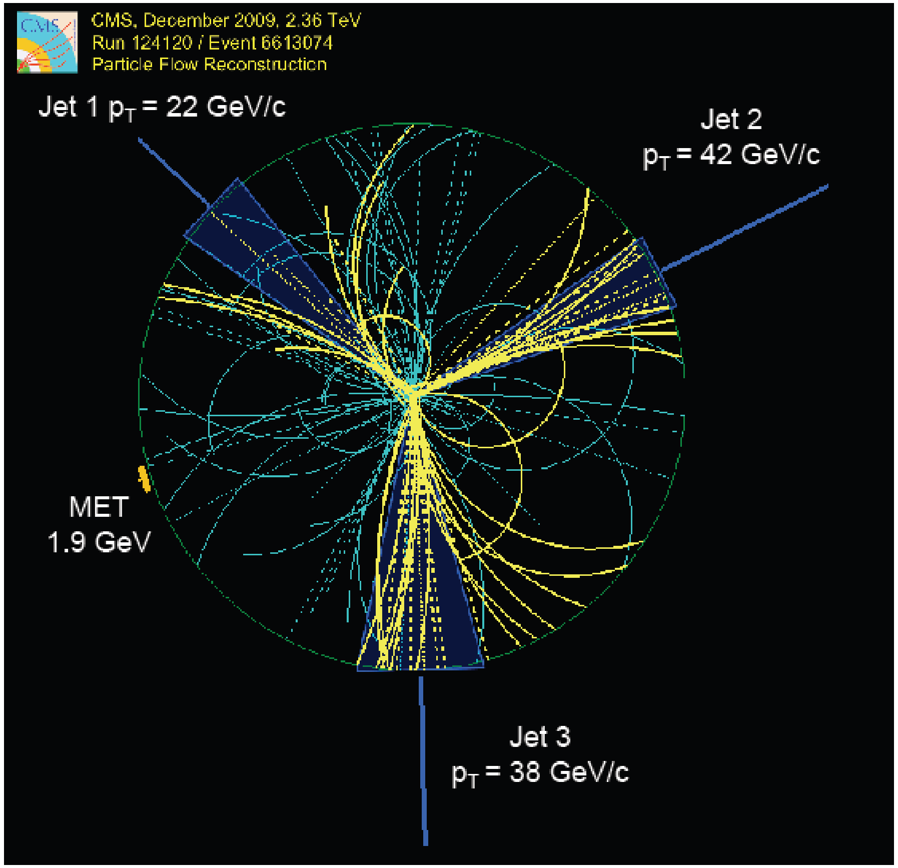
1. Muons: removed from particles before PF jets are clustered
2. Electrons: only consider jets in $\Delta R > 0.3$ away from electrons

Particle Flow (PF): the idea

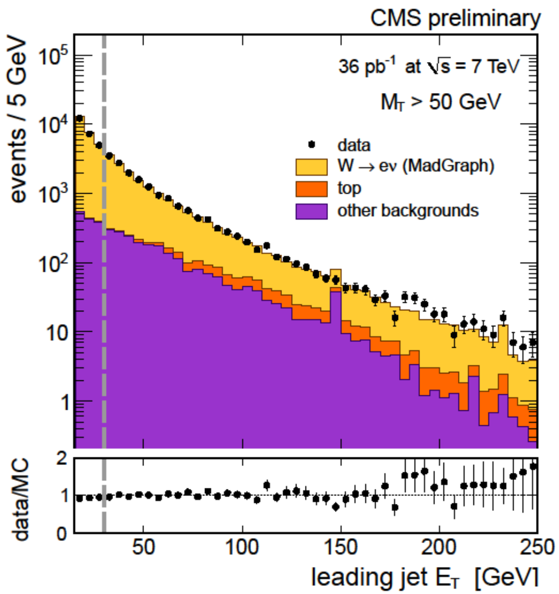


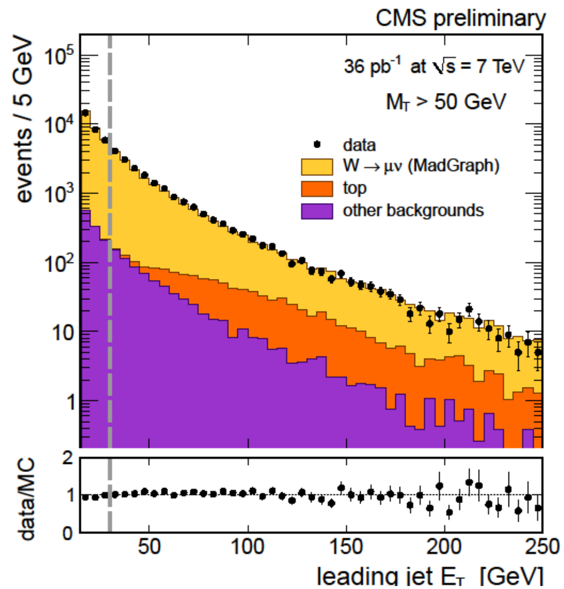
Combine all the information provided by CMS subdetectors, by linking tracks, clusters, calo towers... together

Output: a list of "gen-like" particles with no overlap nor double counting



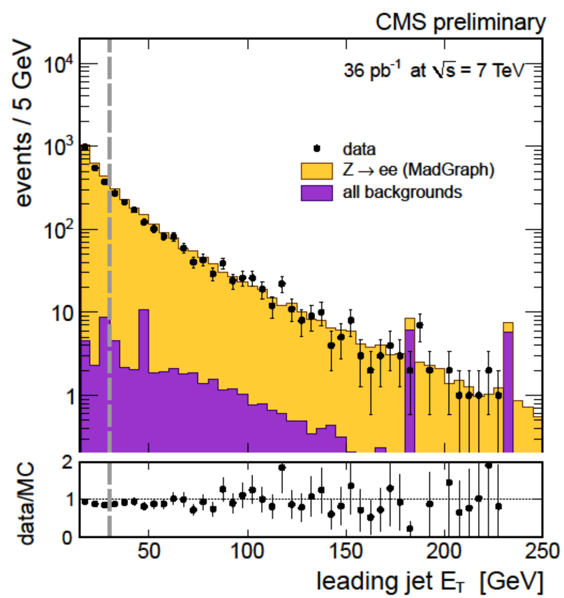
Data vs MC: jet p_T (1-jet events)

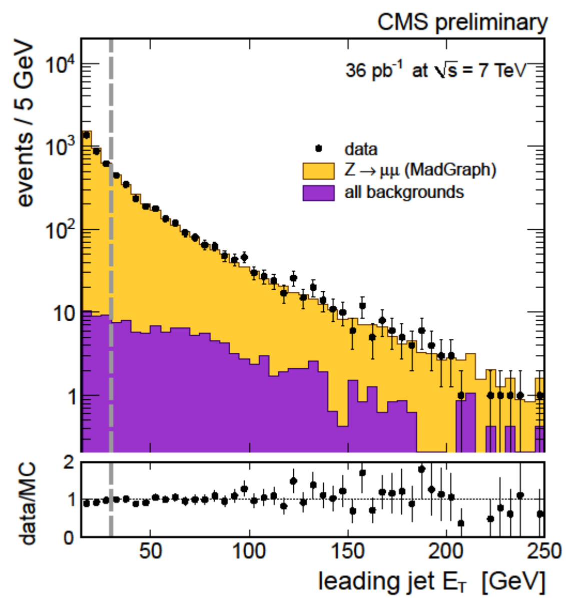




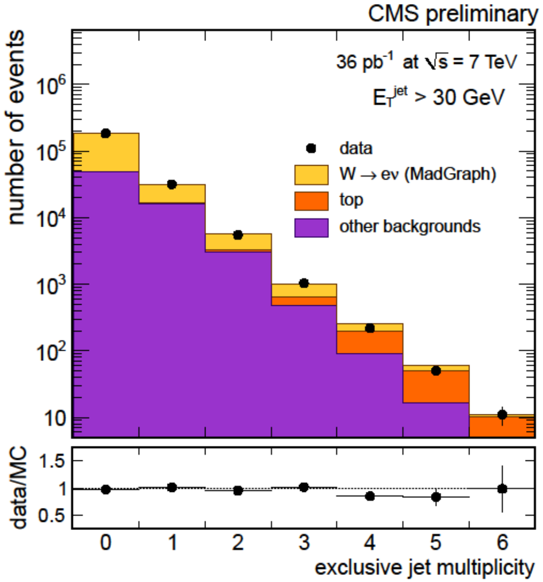
W boson
(M_T > 50 GeV)

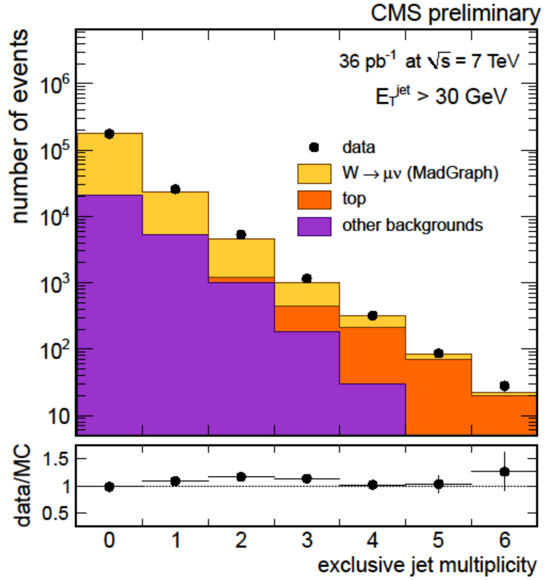
Z boson
(very clean)





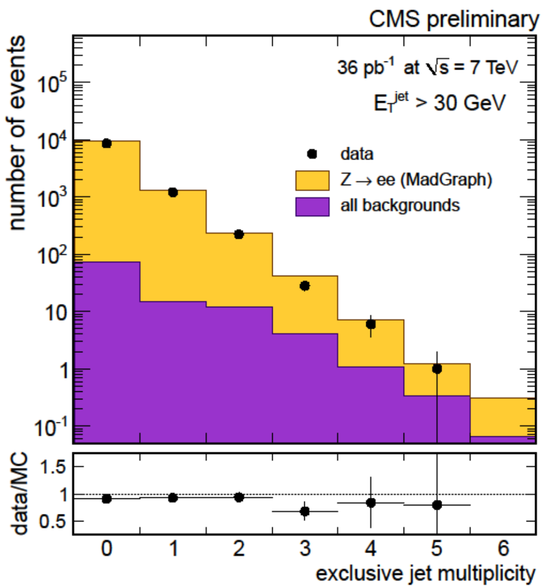
Data vs MC: exclusive jet multiplicity

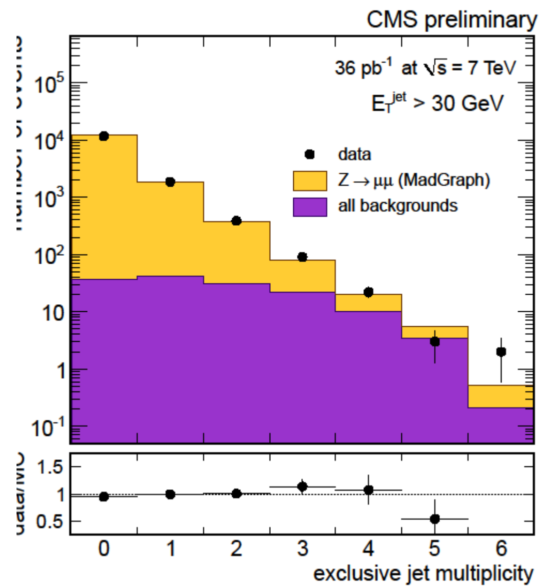




W boson
(M_T > 50 GeV)

Z boson
(very clean)



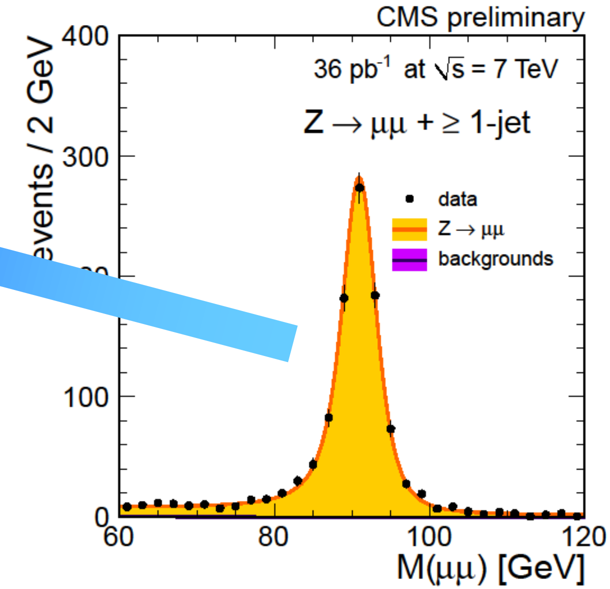


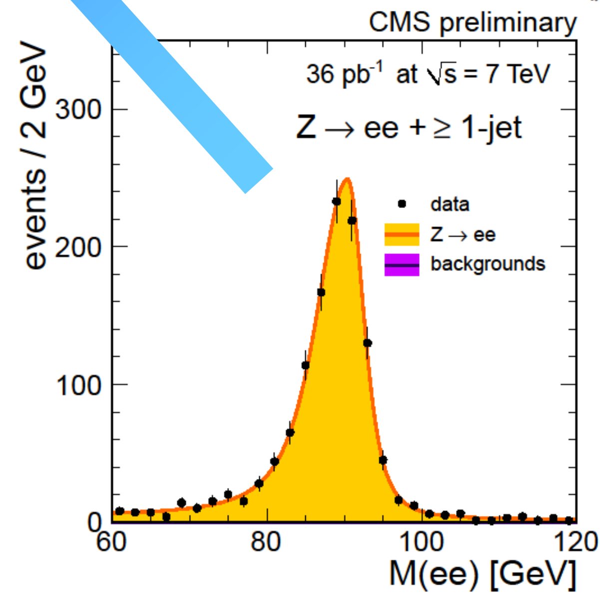
Signal extraction (Z boson)

Very clean resonant signal
with expected low background

Unbinned Maximun-Likehood fit
of di-lepton invariant mass spectra

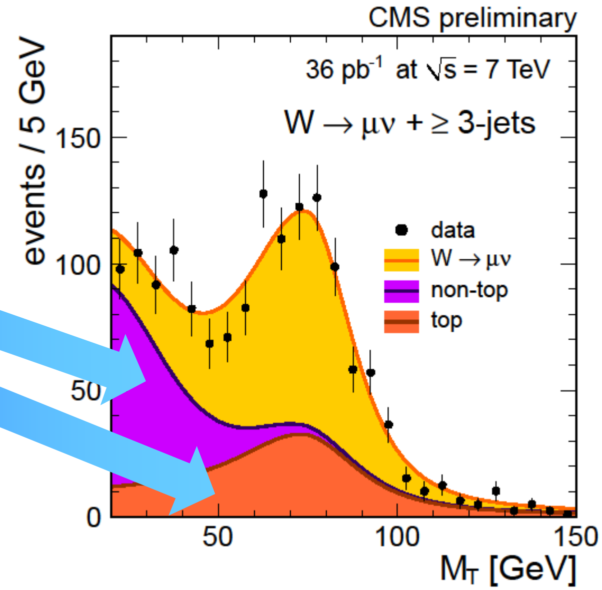
- Functional forms.
- Cruijff for signal, exponential accounting for all backgrounds.
- Floating background.
- Signal parameters floating, but kept equal for all jet multiplicity.



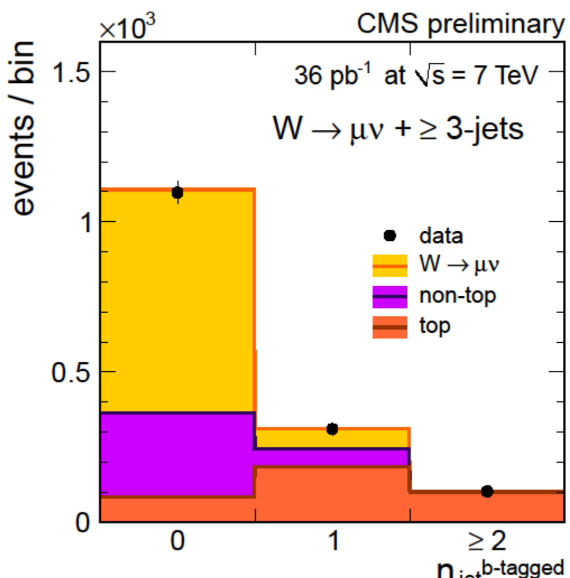


Signal extraction (*W* boson)

Two backgrounds in M_T spectra:
- Not peaking (***QCD+EWK***).
- Peaking (***top***, for W+3,4 jets).



UML fit in 2 dimensions:
M_T vs $n_{jets}^{b\text{-tagged}}$



- Functional forms.
- Fit in range $20 < M_T < 150$ GeV.
- More or less parameters floating according to the statistics.
- b-tagging efficiency and mis-tag prob. from data (di-leptonic top).

Efficiencies and corrections

Lepton efficiencies, measured with Tag & Probe on *Z* events

Factorized as:

1. *Reconstruction*: (ECAL supercluster → ele / track → mu)
2. Selection (differs for first/second lepton in Z events)
3. Trigger (only first leg)

For *W* boson, fit range $M_T > 20$ GeV

1. Correct from MC
2. Verified on data in Z events

Unfolding jet multiplicity spectrum

1. Extract migration matrix $R(n_{RECO}, n_{GEN})$ from MC
2. Singular value decomposition to "unsmear" n_{jet} distribution

What do we measure?

We measured 2 types of ***ratios***:
production of *n* jets over total cross-section, and over (*n*-1) jets

$$\frac{\sigma(\mathrm{V}+\geq n\text{–jets})}{\sigma(\mathrm{V}+\geq 0\text{–jets})} \qquad \frac{\sigma(\mathrm{V}+\geq n\text{–jets})}{\sigma(\mathrm{V}+\geq (n-1)\text{–jets})}$$

Ratio reduces systematics

- full cancellation of luminosity (big uncertainty, 10%).
- partial cancellation of jet energy scale.
- only lepton efficiency vs the number of jets matters.

Systematic uncertainties

Main source:
change in the jet multiplicity due to ***uncertainty on jet energy***

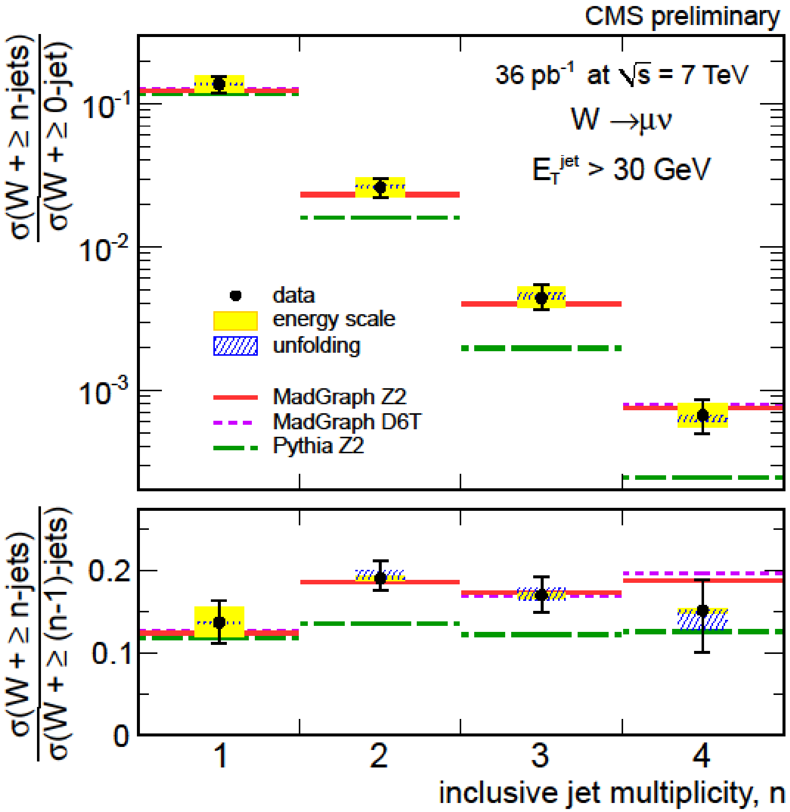
- Uncertainties on data-driven JEC.
- JEC flavour dependence (from MC).
- Pile-up removal (500 MeV each jet in MC).

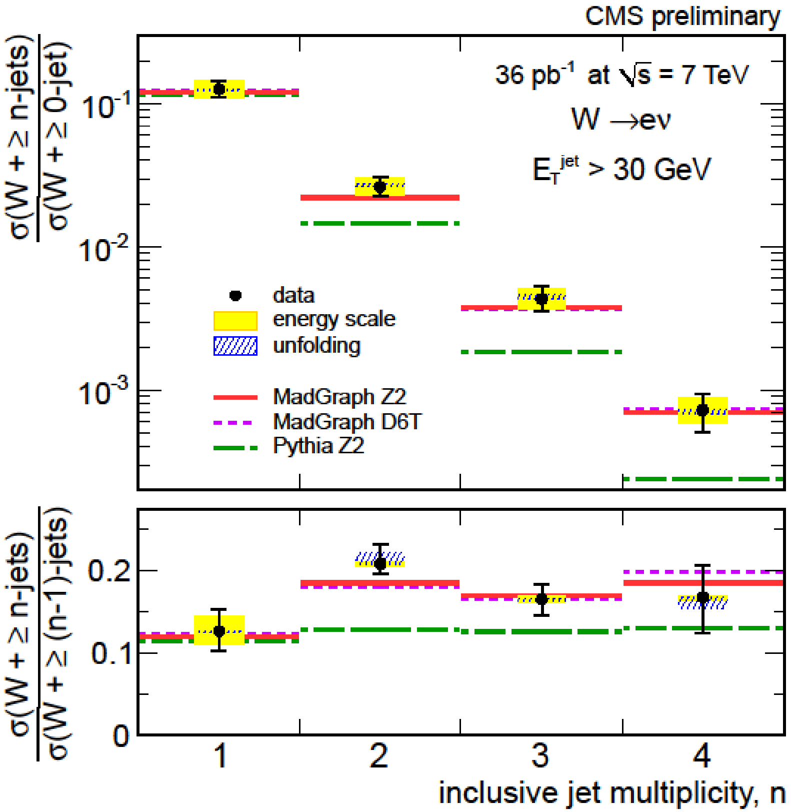
- Jet energy resolution.
- Pile-up residual effect after FastJet subtraction.
- For the *W*, effect on MET (from fit to M_T on data).

Other sources on systematic uncertainties:

1. Reconstruction/selection ***efficiency*** (close to the previous one).
2. ***Signal extraction*** (important only at high jet multiplicity)

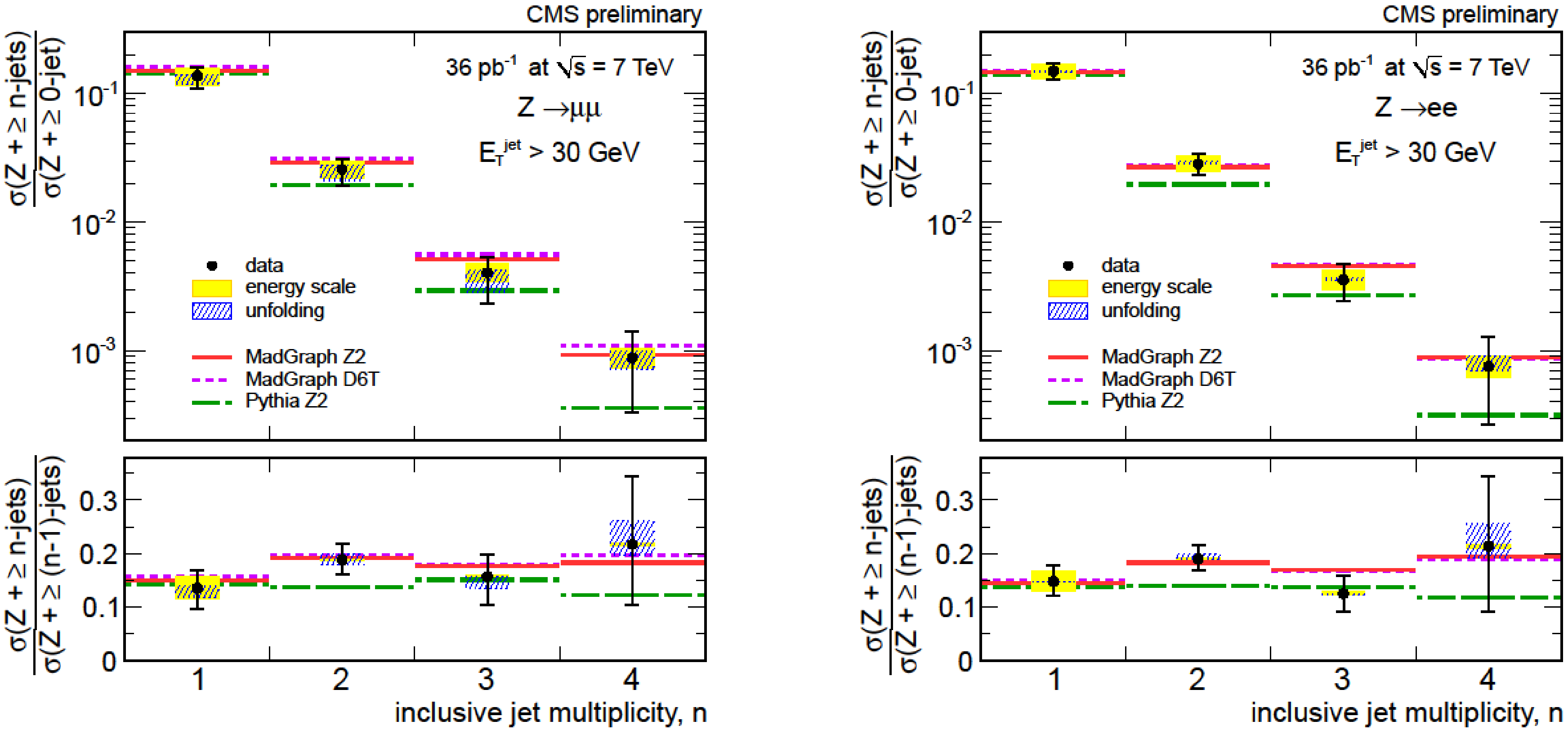
Inclusive jet rates (*W*): $E_T > 30$ GeV





1. Excellent agreement with expectation from ME+PS (MadGraph).
2. PS alone (PYTHIA) starts to fail for $n_{\text{jets}} \geq 2$.

Inclusive jet rates (Z): E_T > 30 GeV



1. Agreement data-MC is good again for ME+PS (MadGraph).
2. PYTHIA is also compatible with data (bigger errors).

Berends-Giele scaling

LO calculation would predict a ***constant value*** for the ratio:

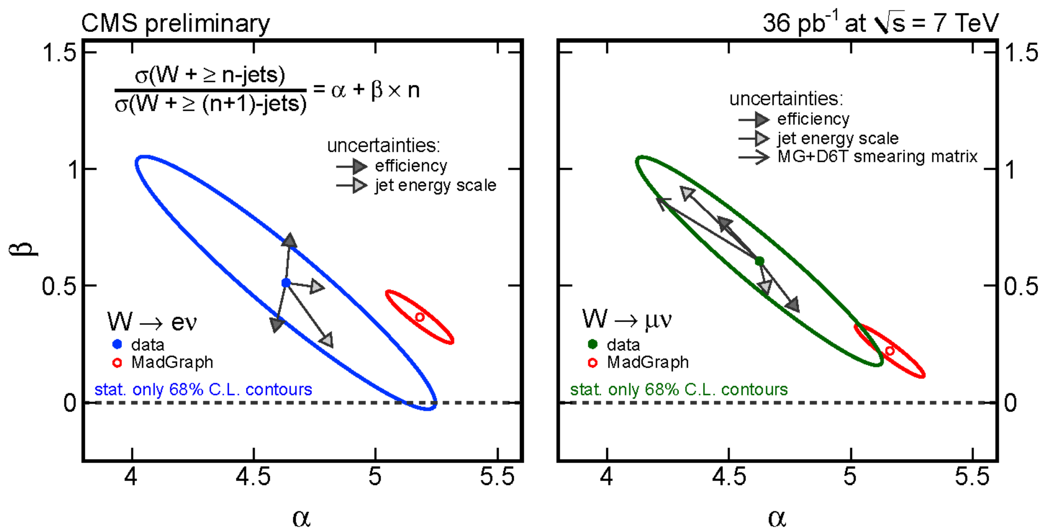
$$C_n = \frac{\sigma_n}{\sigma_{n+1}}$$

NLO corrections and/or phase space effects could violate this proportionality, so we tested the scaling with a second fit

$$C_n = \alpha + \beta n$$

- Fit on ***exclusive*** jet multiplicity bin (uncorrelated).
- Events with ***no jet*** recoiling against VB were ***excluded***.
- ***Bin-to-bin migration*** (det. effects) is taken into account.

Berends-Giele scaling: results

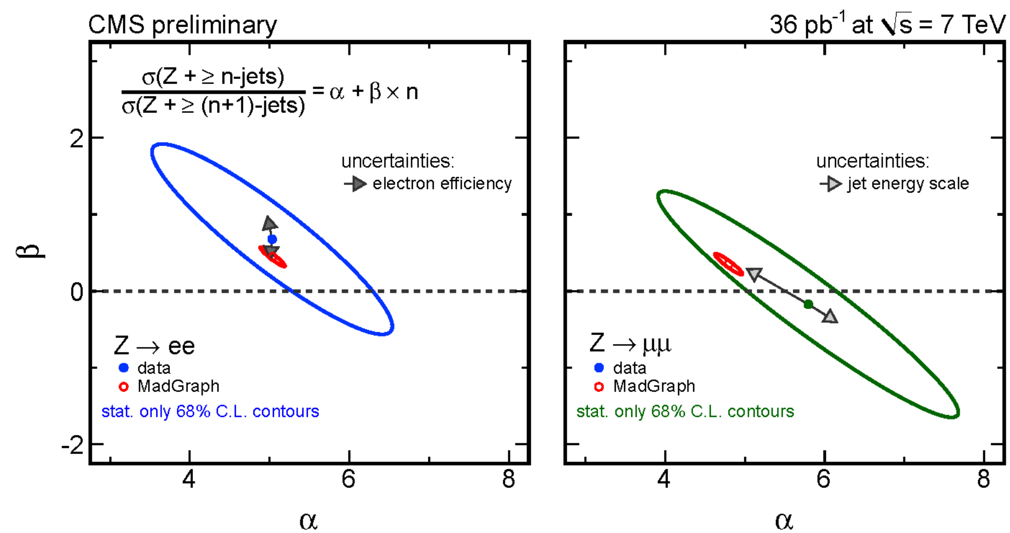


W boson

1. β is 1.0 σ far from 0.0.
2. Differences due to the ΔR cut (for e only)

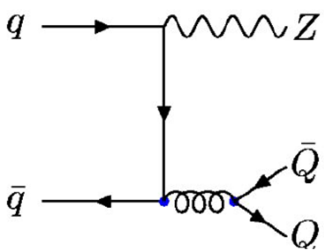
Z boson

1. β is 0.5 σ far from 0.0.
2. Very good agreement with ME+PS MC

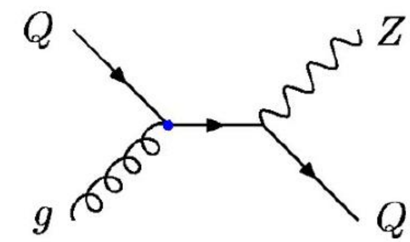


Z + *b*-jets

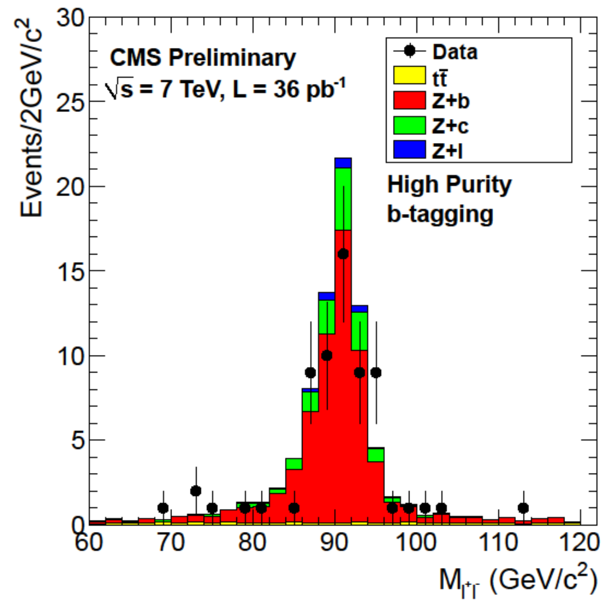
Benchmark channel for H(MSSM), but up to 30% difference between:



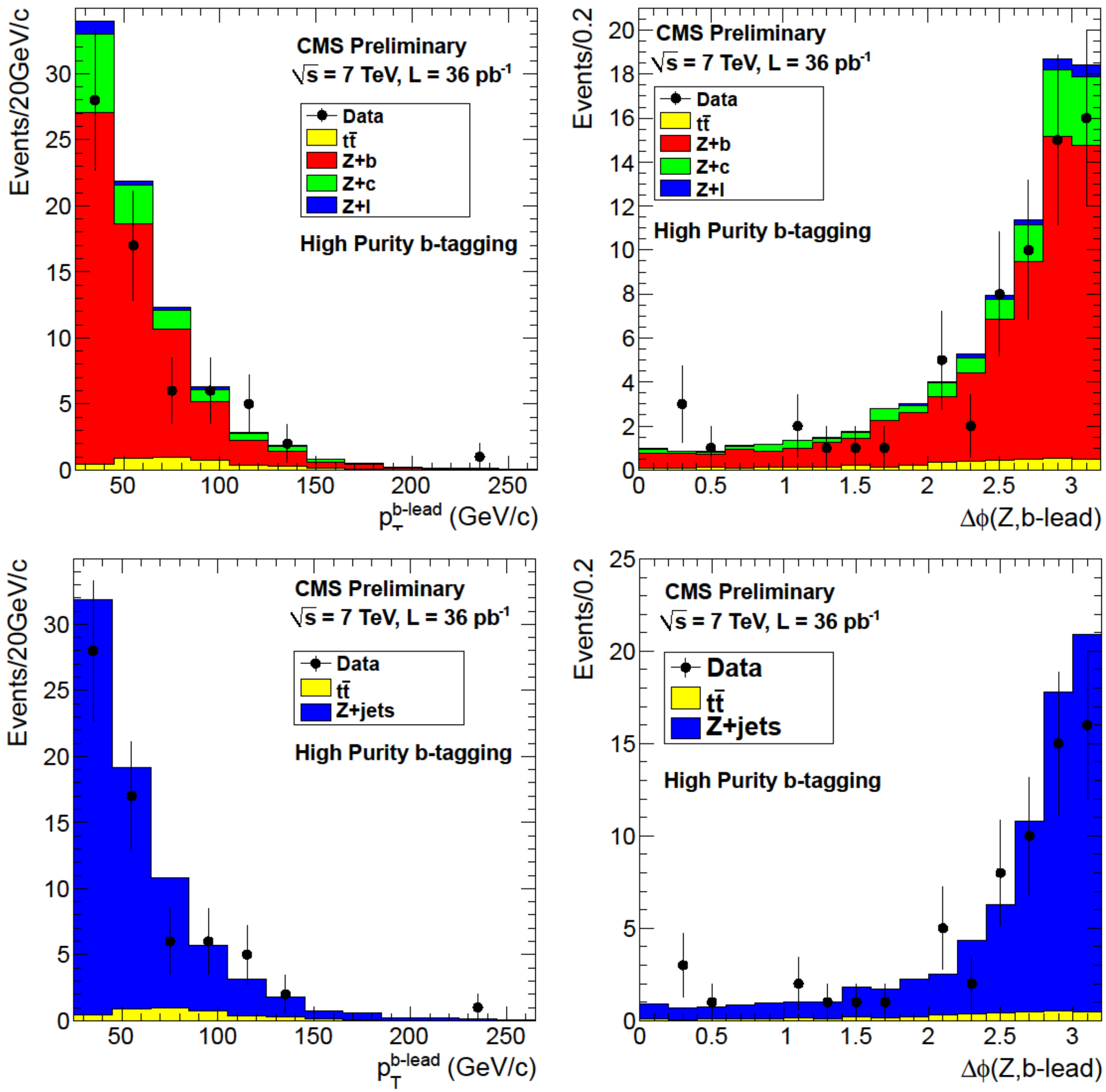
Fixed flavour ⟷ Variable flavour



1. At least 1 Z (into ee and μμ).
2. At least 1 PF jet, with $E_T > 25$ GeV, $|\eta| < 2.1$.
3. ΔR(jet, lepton) > 0.5.
4. ≥1 secondary vertex in jet.
5. 2 versions of b-tagging: High Purity and High Efficiency.
6. MET < 40 GeV, 60 < M_{ll} < 120



Fixed vs variable flavour



MadGrah
Fixed-flavour

No distinction between schemes (limited statistics in the tails)

MadGrah
Variable-flavour

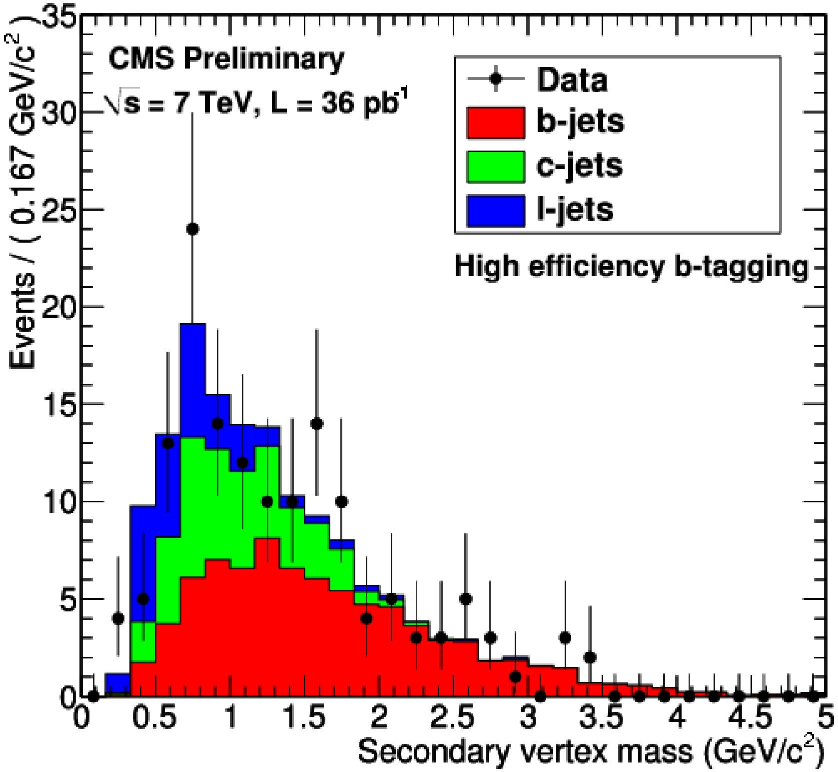
(Z+b)/(Z+jets) ratio

- *Z+b* purity is extracted from fit to the secondary vertex mass

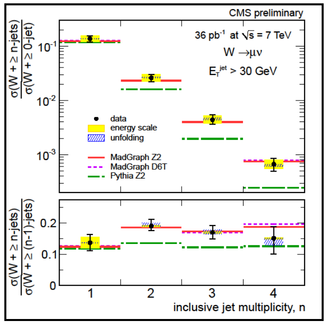
Purity (%)	SSVHE	SSVHP
data	55±9	88±11
MC	57±3	82±4

- Results are compatible with MadGraph(*) and MCFM NLO calculations

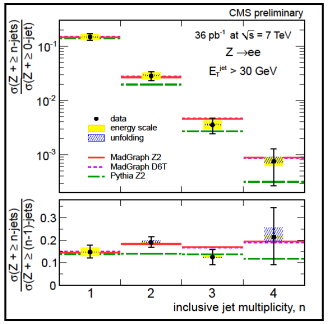
(*) *Z+b* and *Z+c* with $p_{T,jet}$ >15 GeV scaled to MCFM x-sec



Sample	$\mathcal{R}(Z \to ee)$ (%), $p_T^e > 25$ GeV, $\lvert\eta^e\rvert < 2.5$	$\mathcal{R}(Z \to \mu\mu)$ (%), $p_T^\mu > 20$ GeV, $\lvert\eta^\mu\rvert < 2.1$
Data HE	$4.3 \pm 0.6(stat) \pm 1.1(syst)$	$5.1 \pm 0.6(stat) \pm 1.3(syst)$
Data HP	$5.4 \pm 1.0(stat) \pm 1.2(syst)$	$4.6 \pm 0.8(stat) \pm 1.1(syst)$
MADGRAPH	$5.1 \pm 0.2(stat) \pm 0.2(syst) \pm 0.6(theory)$	$5.3 \pm 0.1(stat) \pm 0.2(syst) \pm 0.6(theory)$
MCFM	$4.3 \pm 0.5(theory)$	$4.7 \pm 0.5(theory)$



In short…



- We studied the production of jets with *W/Z*:
  - PF jets were used, for the best response.
  - Good agreement data-MC for jet p_T and jet multiplicity.
- We measured:
  - Ratios (V+*n* jets)/(V) and (V+*n* jets)/(V+(*n-1*) jets) .
  - The Berends-Giele scaling.
- Production of *Z+b*:
  - Measured ratio *(Z+b)/(Z+jets)*.
  - Good agreement data-MC (also for *b* fraction).