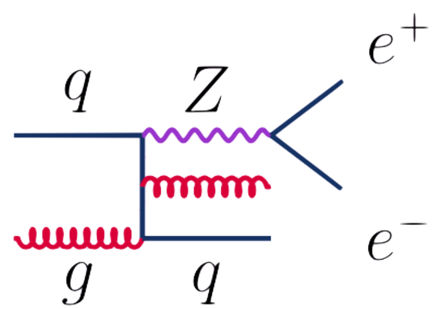
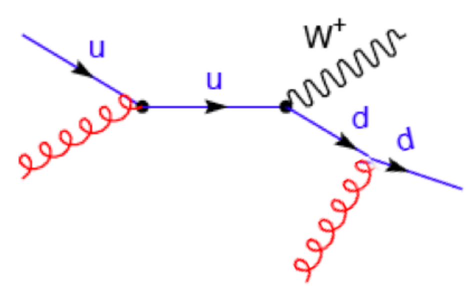
Rates of Jets Produced in Association with Vector Bosons at CMS

Kira Grogg
University of Wisconsin
August 9, 2011

Motivation and Goals

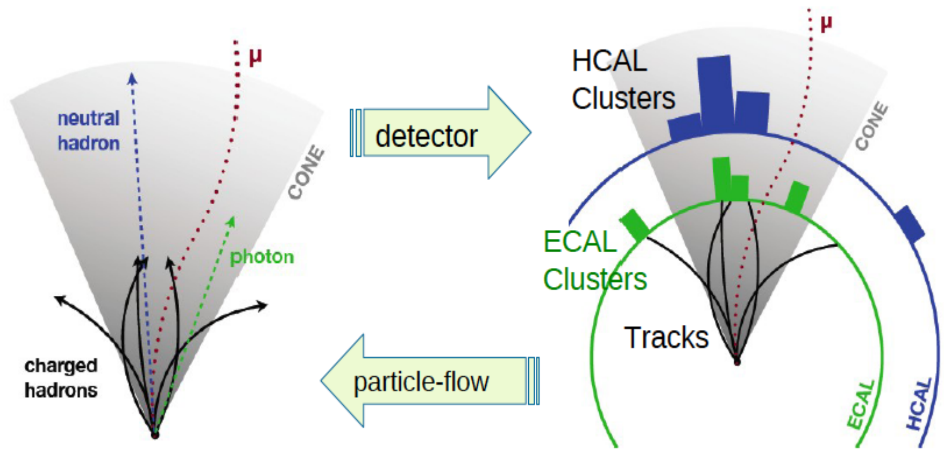
- V+jets characterized by jets, high energy leptons and significant missing E_T in the final state
  - *Major background for new physics*
- Test of perturbative QCD calculations
  - Verification of theoretical cross-section and parton distribution functions (PDFs)
- Goal: measurement of the rate of events with a vector boson produced with the presence of jets
  - Jets are considered when above a given E_T threshold
  - Inclusive rate of n jets (i.e., ≥ n jet) is given and events are not corrected for acceptance, for comparison with multiple theoretical models
- Start with ratio measurements where systematics from jet energy scale, luminosity, and lepton selection partially cancel
- Comparisons with ME+PS Monte Carlo





Object Reconstruction

- **Muons** reconstructed using silicon tracker and muon chambers
  - Identification based on compatibility of tracker, calorimeters and muon chambers measurements
- p_T resolution for EWK ~1-2%
- **Electrons** reconstructed using silicon tracker and $PbWO_4$ crystal calorimeter
  - Identification based on shower shape, Had/Em, track matching
- E_T resolution for EWK ~1%
- **Jets** and **Missing E_T** reconstructed using particle flow technique
  - All constituent particles—electrons, muons, photons, neutral hadrons, and charged hadrons—are reconstructed from information in all sub-detectors
  - Jets reconstructed from particles using anti-k_T algorithm with cone radius of 0.5



Analysis Flow

2010 LHC data of 36.1 pb^{-1} CMS-PAS-EWK-10-015

Select V candidates
$N_{selected}$

Select and count jets
- $E_T > 30$ GeV
- *n jets*

Extract signal
- Exclusively by jet multiplicity
- $\boldsymbol{N_V}$

Correct yields for efficiency
- ε_{tot}

Correct jet multiplicity for detector effects
- ***Unfolding***

$$\sigma_{acc}(n\,jets) = \frac{N_V \cdot U}{\varepsilon_{tot} \cdot L}$$

Plot ratios

$$\sigma(V + njets) / \sigma(V_{tot})$$

$$\sigma(V + njets) / \sigma(V + (n-1)\,jets)$$

$$\sigma(V + njets) / \sigma(V + (n-1)\,jets) = \alpha + \beta \cdot n$$

Cross section given as ratios to reduce systematics
Final jet counting is "inclusive" (i.e., ≥ n jets)

W→enu and Z→ee Selection

- Electron selection
  - p_T > 20 GeV
  - $|\eta|$ < 2.5, excluding 1.4442 < $|\eta|$ < 1.566
  - Identification, isolation, conversion rejection (see backup)

  (Acceptance: p_T and $|\eta|$ requirements)

- Check for 2nd electron with
  - Identification, isolation, conversion rejection
    - looser to increase statistics
  - p_T > 10 GeV
  - $|\eta|$ < 2.5, exclu. 1.4442 < $|\eta|$ < 1.566
  - 60 < M_{ee} < 120 GeV

  If exists, event is Z(ee)
  If does not exist, event is W(enu)

- No muons with p_T > 15
- Transverse impact parameter δ_{xy} < 0.035
- HLT object match
- For W(enu) only: M_T > 20 GeV
  - From electron and Particle Flow Missing E_T
  - Necessary for data-driven fitting

$$m_T = \sqrt{2 p_T^{(e)} p_T^{(\nu)} (1 - \cos \Delta\phi)}$$

W→munu and Z→mumu Selection

- Global and Tracker muon with
  - $p_T > 20$ GeV
  - $|\eta| < 2.1$
  - Tracker muon: $n_{trk} > 10$, $n_{pixhits} > 1$, $d_{xy} < 2$mm
  - Global muon: one valid hit, $c^2/ndf < 10$, 2 segments match track muon
  - Combined relative isolation < 0.15

  (first two items: Acceptance)

- Check for 2nd muon with
  - Global
  - $p_T > 10$ GeV
  - $|\eta| < 2.4$
  - $60 < M_{mumu} < 120$ GeV
  - Looser to increase statistics

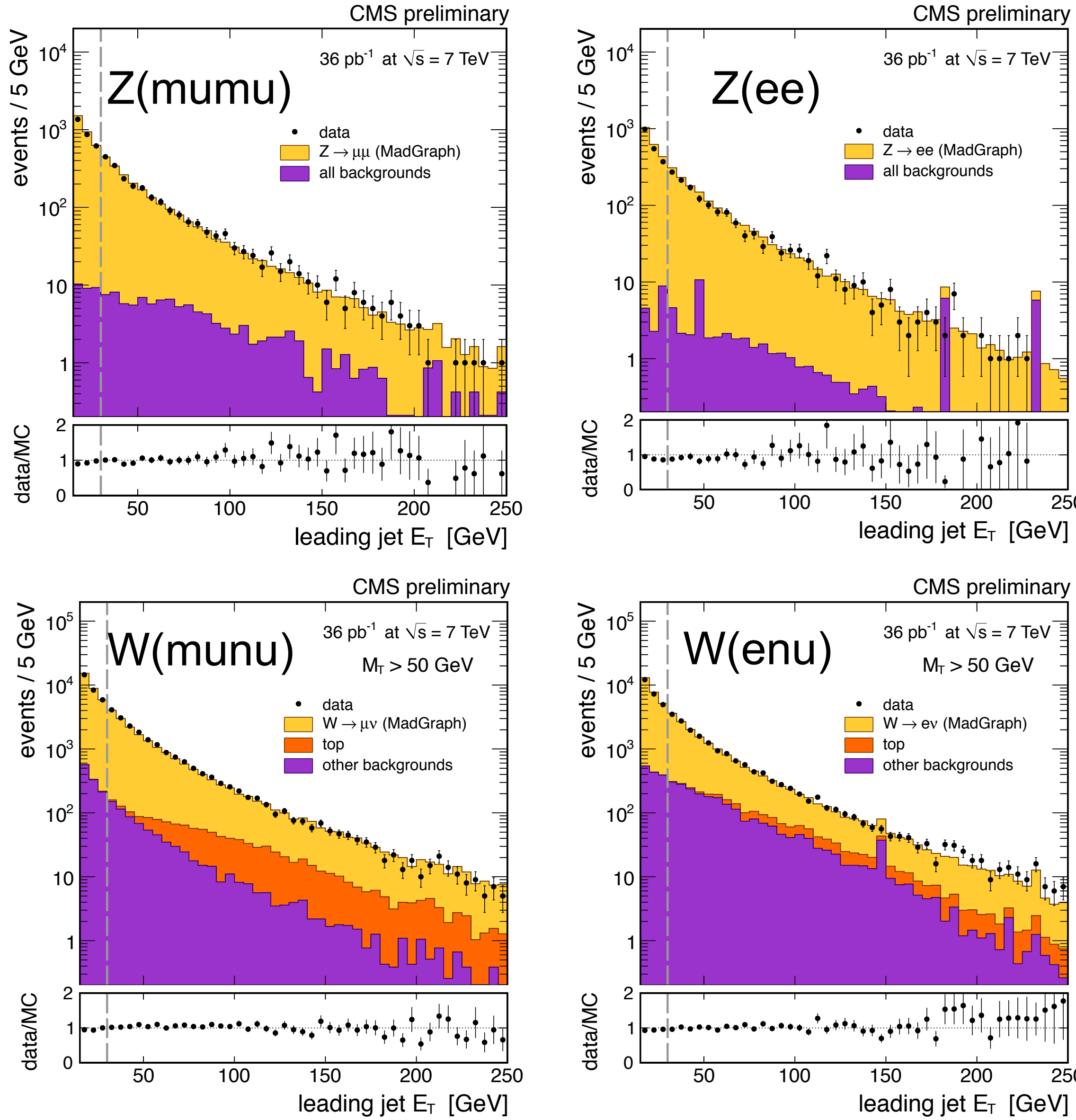
  If exists, event is Z(mumu)
  If does not exist, event is W(munu)

- For W(munu) only: $M_T > 20$ GeV
  - From muon and Particle Flow Missing E_T
  - Necessary for data-driven fitting

$$m_T = \sqrt{2 p_T^{(e)} p_T^{(v)} (1 - \cos\Delta\phi)}$$

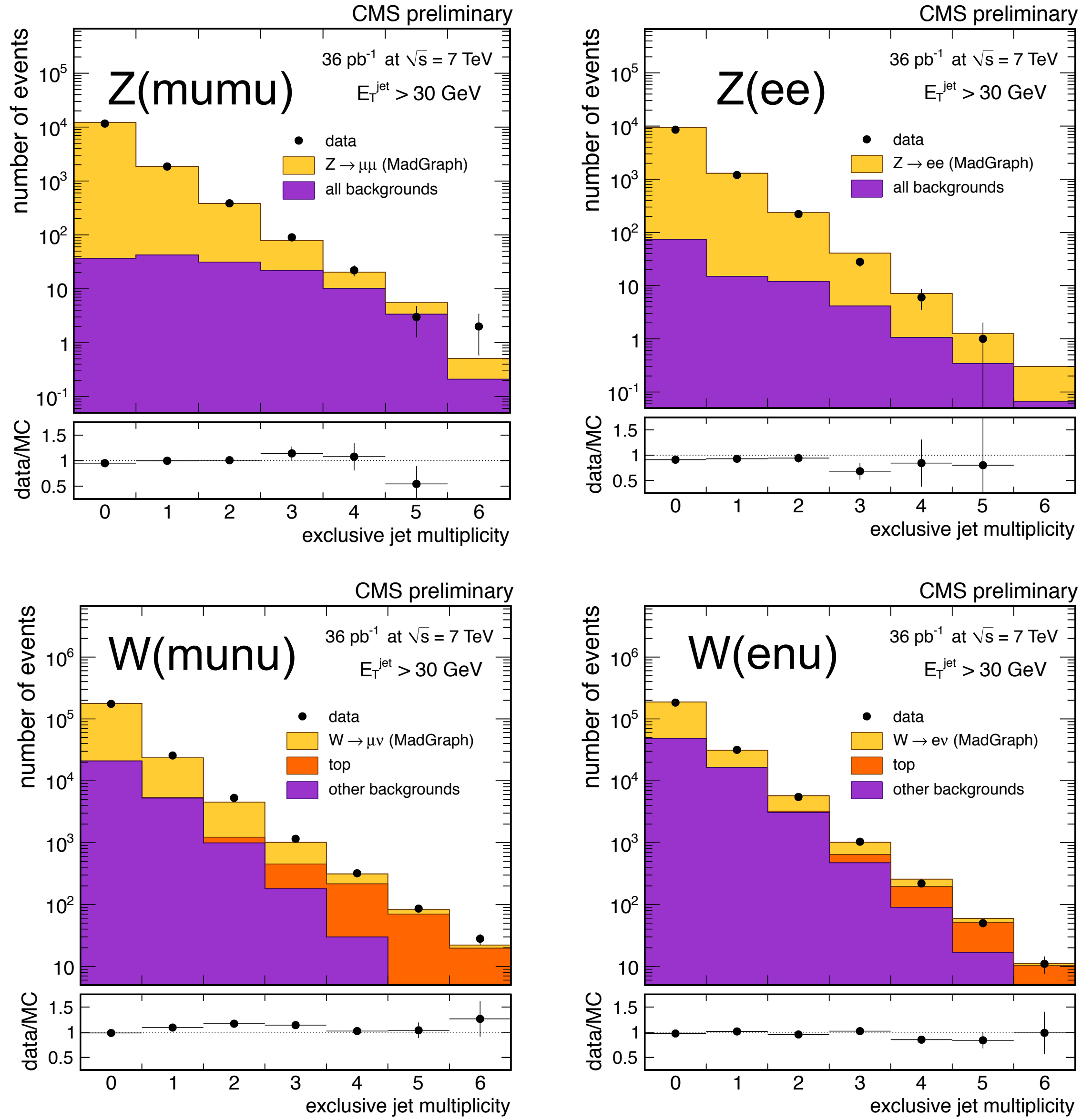
Leading Jet Transverse Energy

- Leading jet E_T after full selection applied
  - Dashed line indicates jet threshold of > 30 GeV
  - $|\eta| < 2.4$
- Madgraph MC (signal) normalized to NNLO cross sections, backgrounds to (N)LO
- W M_T > 50 GeV to enhance signal
- Agreement with MC is very good



Exclusive Jet Multiplicity

- Comparison of data to MC raw jet rates
- Jet $E_T > 30$ GeV
  - $|\eta| < 2.4$
- MadGraph MC (signal) normalized to NNLO cross sections, backgrounds to (N)LO
- Data in agreement with MadGraph+Pythia MC



Full Selection Efficiencies

- Selection efficiency combines tag and probe and MC results
  - Tag and probe on Z+jets data and MC samples
    - Fits to invariant mass for probes passing and failing selection cuts
    - W: $\varepsilon_{reconstruction}$ x $\varepsilon_{selection}$ x $\varepsilon_{trigger}$
    - Z: $\varepsilon_{reconstruction}$ x $\varepsilon_{selection}$ x $\varepsilon_{trigger}$ x $\varepsilon'_{reconstruction}$ x $\varepsilon'_{looser\ selection}$
  - MC efficiency: full selection / gen leptons in acceptance
    - Acceptance: generator lepton p_T > 20 GeV, eta < 2.5 (2.1 for muons)
  - Final efficiency used to correct yields (after signal extraction):
    - MC * T&P data / T&P MC
- Muon efficiency from tag and probe only, and events are corrected before signal extraction

	0 jets	1 jets	2 jets	3 jets	≥ 4 jets
ε (Muons)	0.952	0.925	0.915	0.916	0.843
ε (Wenu)	0.718	0.659	0.599	0.557	0.471
ε (Zee)	0.666	0.620	0.582	0.578	0.477

Signal Extraction of Z→ll

- Extended unbinned maximum likelihood fit to the di-lepton invariant mass
- Signal modeled with Crujiff function (modified gaussian with left and right tails independent)

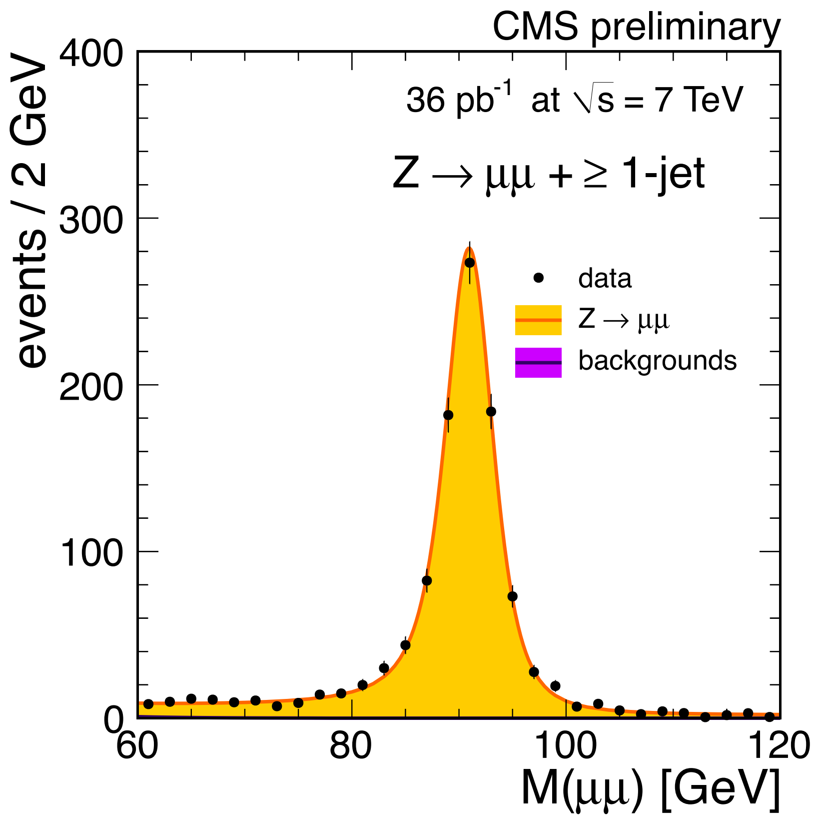
$$F_S(M_{ll};\alpha_L,\alpha_R,\sigma_L,\sigma_R,M_0,N_S) = N_S\, e^{-\frac{(M_{ll}-M_0)^2}{2\sigma^2+\alpha^2(M_{ll}-M_0)^2}}$$

$\sigma = \sigma_L(\sigma_R)$ for $M_{ll} < M_0(M_{ll} > M_0)$ and $\alpha = \alpha_L(\alpha_R)$ for $M_{ll} < M_0(M_{ll} > M_0)$.

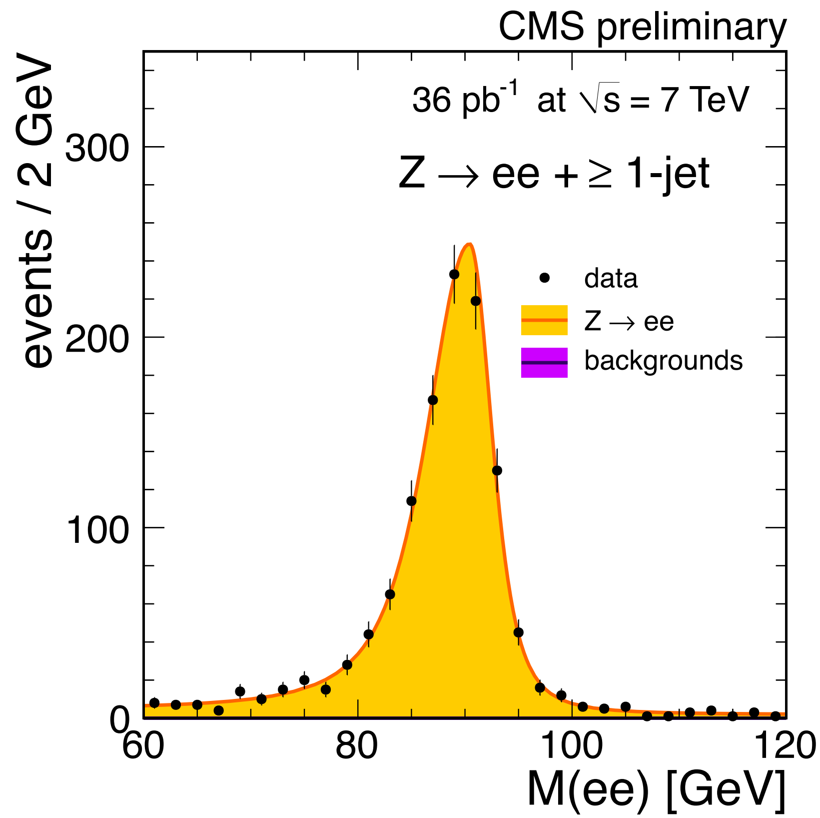
  - α_L determined from high purity data sample and fixed
  - All other parameters floated, but constrained to be the same for n ≥ 1 samples
- All backgrounds modeled with an exponential, floated for all bins
- For muons, events are weighted by efficiency as a function of n-jets, p_T, and eta before fitting

Z→ll Fit Results

Z → mumu fit for 1 jets



Z → ee fit for 1 jets



Background too low to be visible

Signal Extraction of W→lnu

- Use W M_T to distinguish signal from the majority of backgrounds
- Use number of b-tagged jets to distinguish signal from top
  - Top decays to W, so it also peaks in M_T
  - Data-driven method, not relying on MC cross sections
  - See next slide for PDF
- Perform 2D fits of M_T x $n_{btagged}$
- Species:
  - Signal (W) : cruijff or double cruijff (0-2 jets)
    - Mean and resolutions of signal are floated (for 0, 1 & 2 jets)
    - Mean for signal (3 & 4 jets) is floated
  - Top (ttbar, single top): cruijff
    - Parameters fixed to MC values
    - Divided in to three subspecies based on number of b-jets (0, 1, ≥ 2)
  - Others (QCD, Z, W→τν, photons): cruijff
    - Initially fit to ID-inverted data sample
    - All parameters floated

Parameterization of N-btagged Jets

- Probability distribution function for n b-tagged jets:

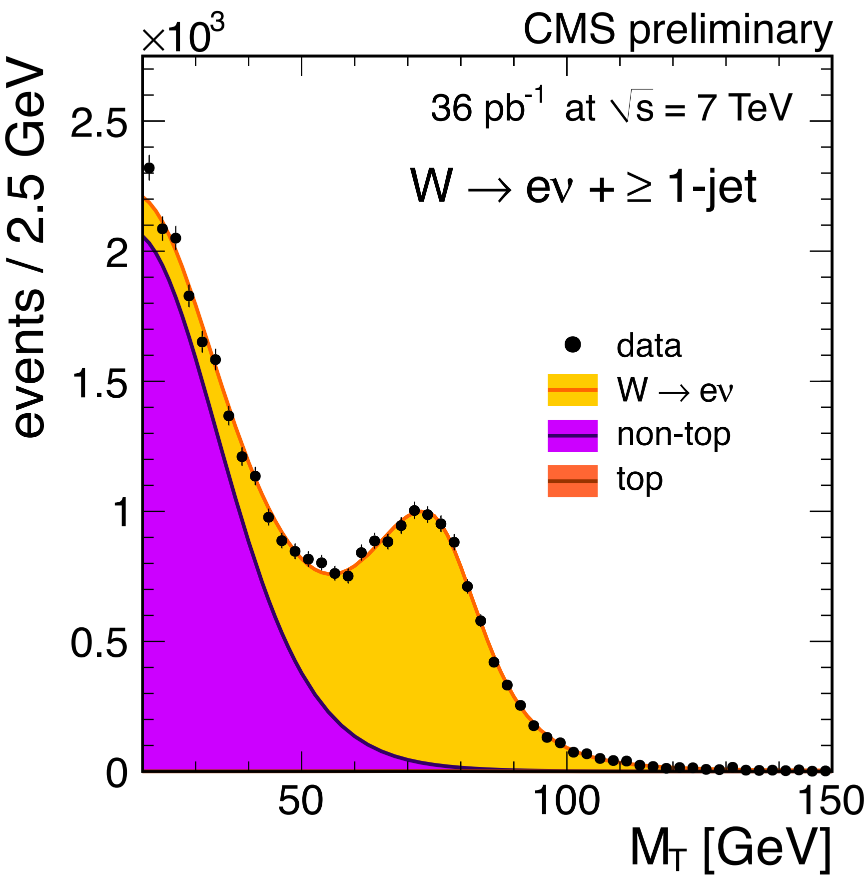
$$P(n_j^{tagged}|n_j, n_{bj}, \epsilon_{nob}, \epsilon_b) =$$

$$\begin{cases} (1-\epsilon_{nob})^{n_j-n_{bj}} \cdot (1-\epsilon_b)^{n_{bj}} & n_j^{tagged} = 0 \\ (1-\epsilon_{nob})^{n_j-n_{bj}-1} \cdot \epsilon_{nob} \cdot (n_j - n_{bj}) \cdot (1-\epsilon_b)^{n_{bj}} + & n_j^{tagged} = 1 \\ \qquad (1-\epsilon_{nob})^{n_j-n_{bj}} \cdot (1-\epsilon_b)^{n_{bj}-1} \cdot (\epsilon_b) \cdot n_{bj} & \\ 1 - P(0) - P(1) & n_j^{tagged} \geq 2 \end{cases}$$

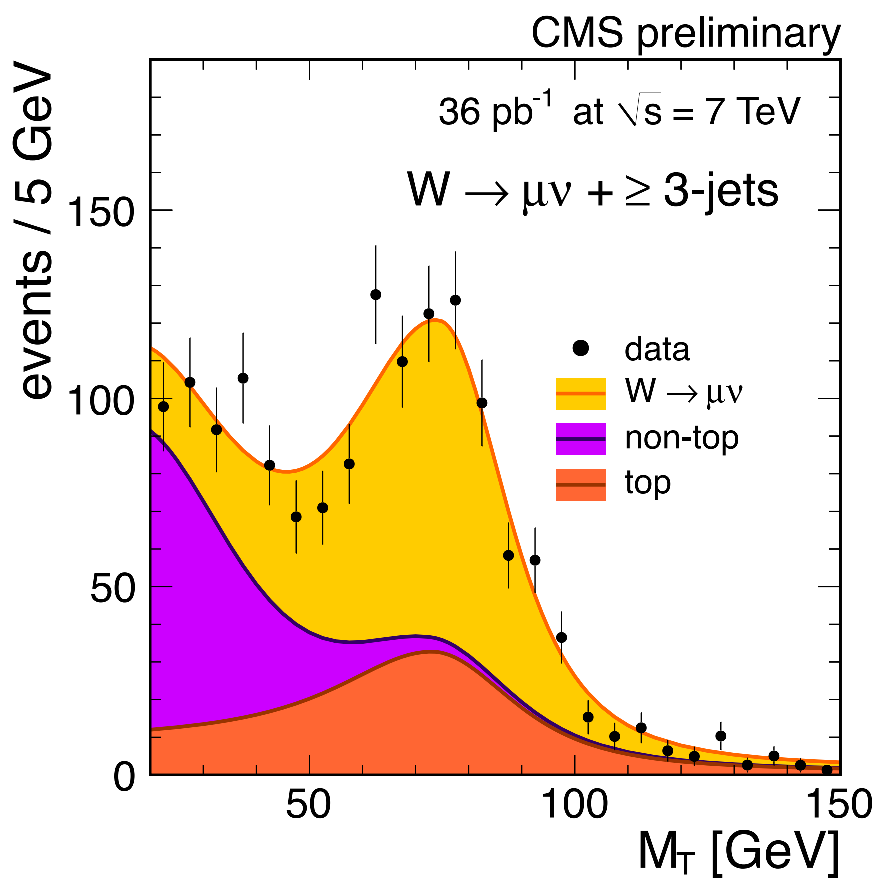
  - n_b = number of b-tagged jets
  - n_{bj} = number of jets in acceptance that are b-flavored (true)
  - ε_{nob} = mistag rate
    - 2.42 ± 0.03 (stat) ± 0.5 (syst)% from MC and validated on data
  - ε_b = tag rate
    - 63 ± 6.3% from MC and validated on data

W→enu and W→munu M_T Fit Results

1 jet events, electron

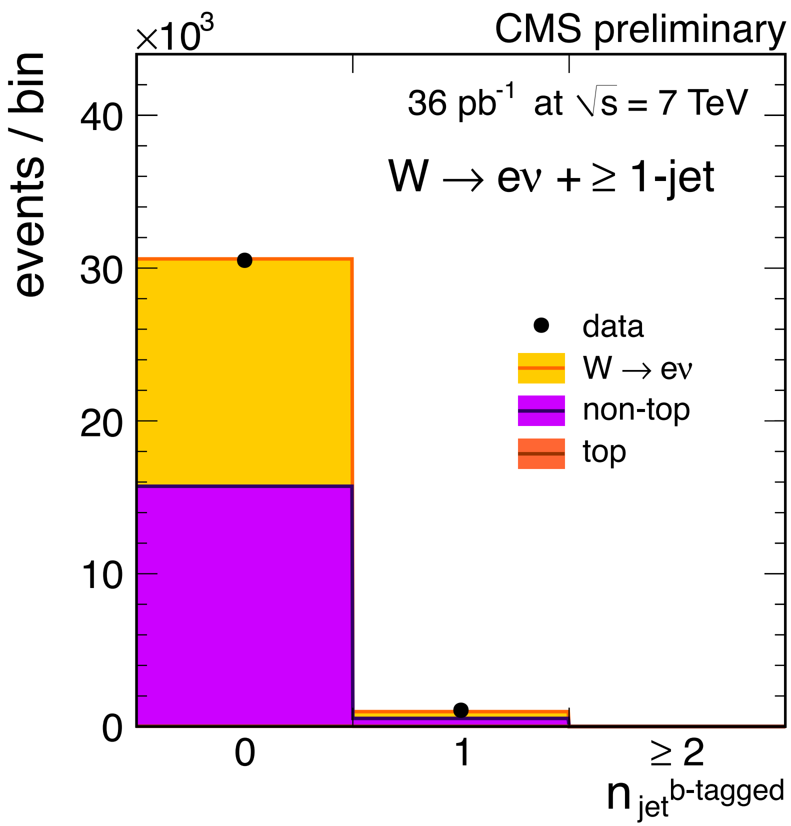


3 jet events, muon

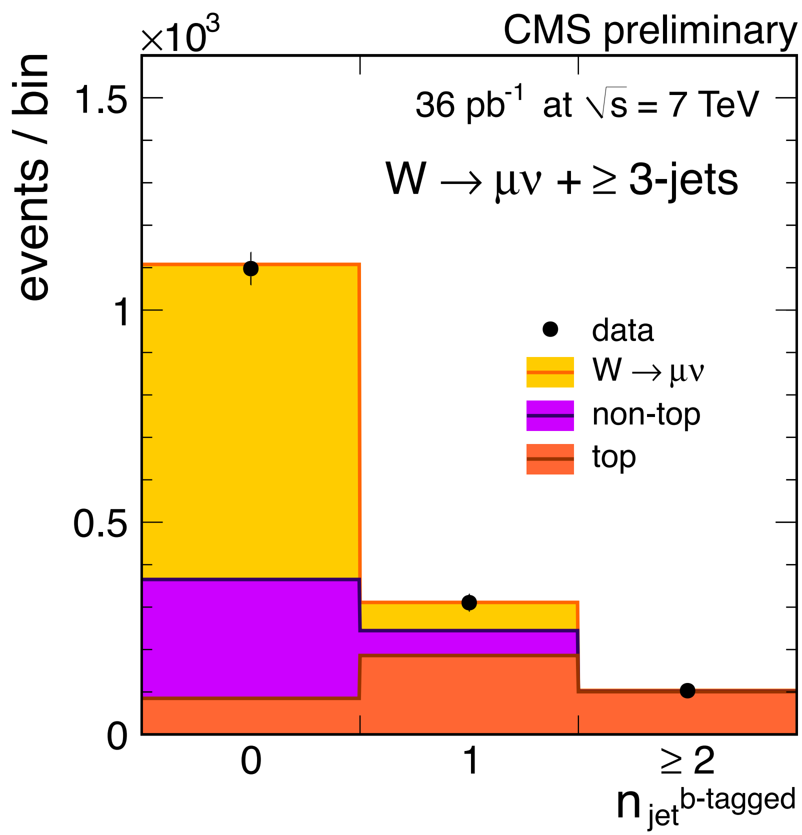


W→enu and W→munu n_{bjets} Fit Results

1 jet events, electron

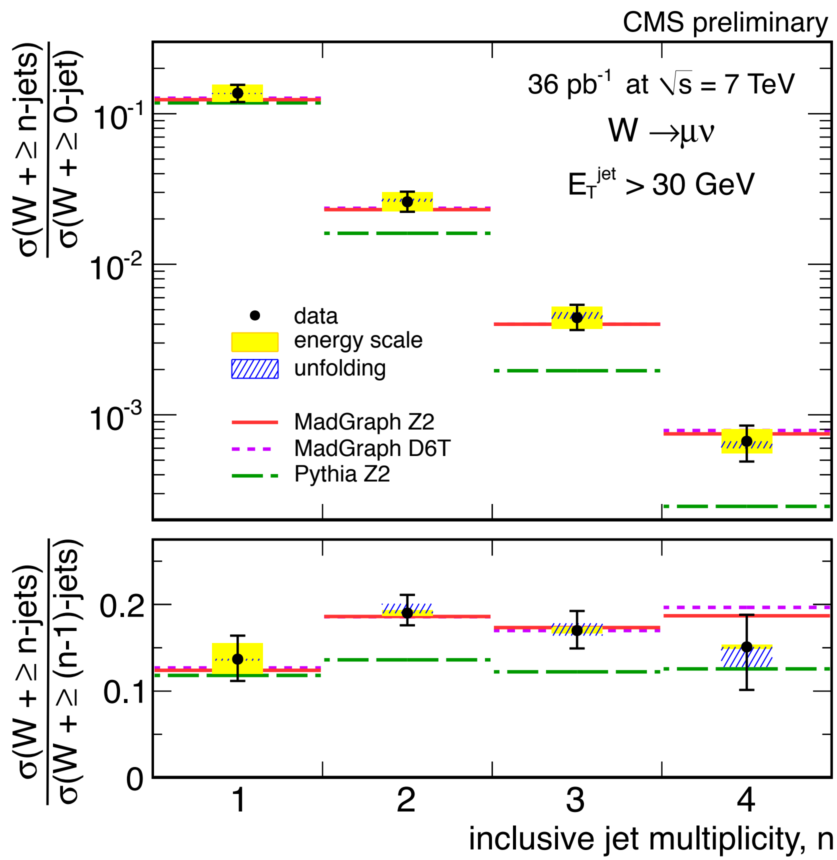


3 jet events, muon

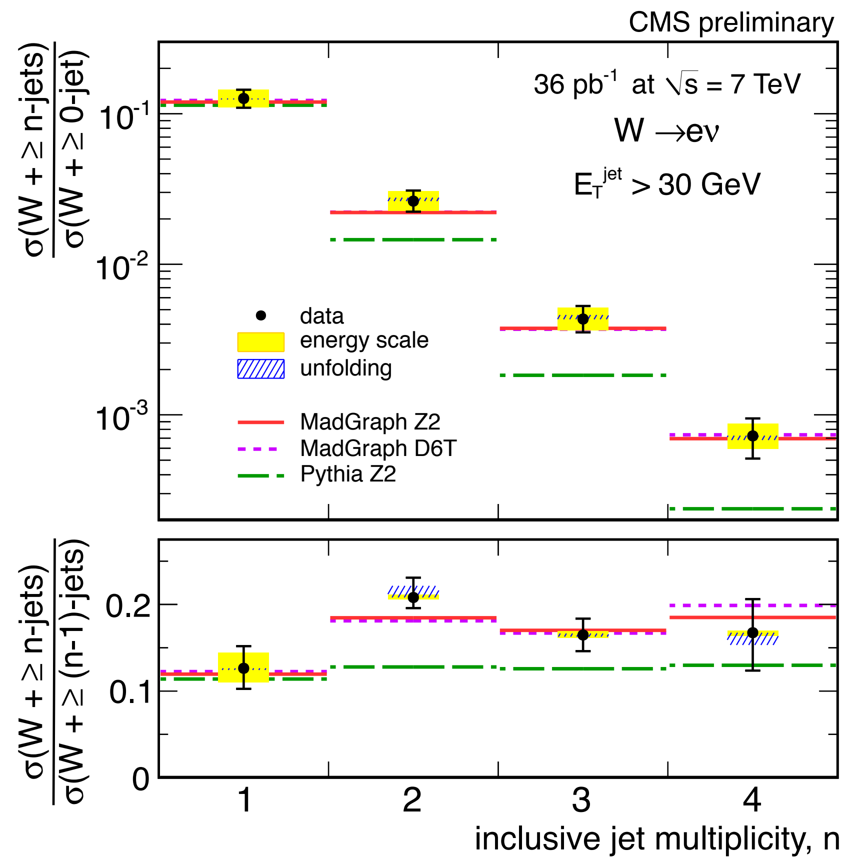


W+Jets Cross Section Ratios

W→munu + jets



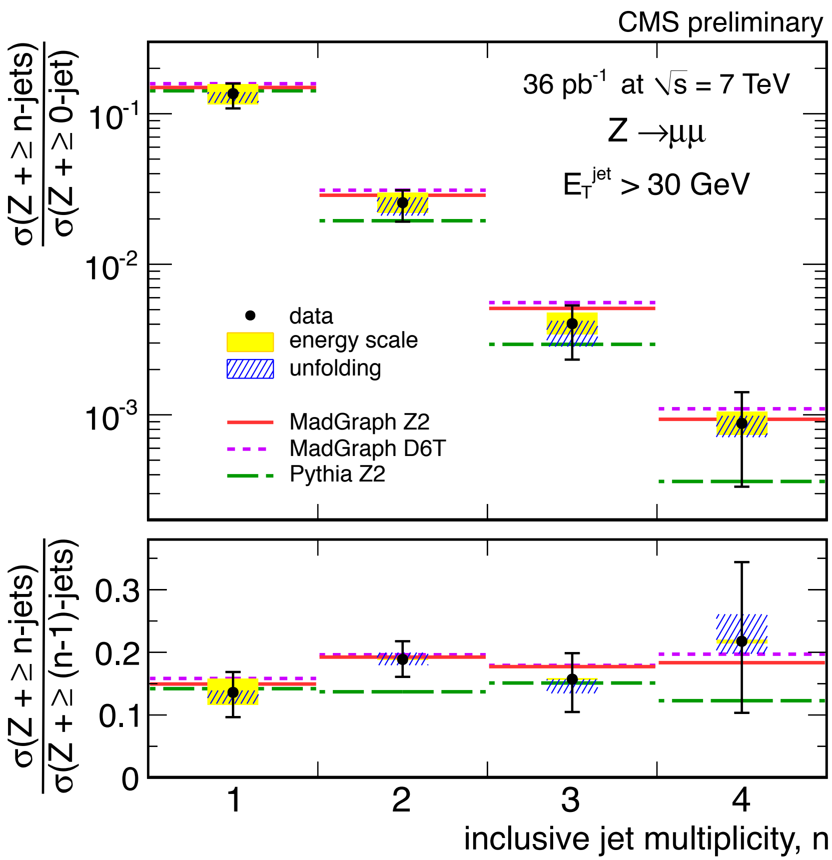
W→enu + jets



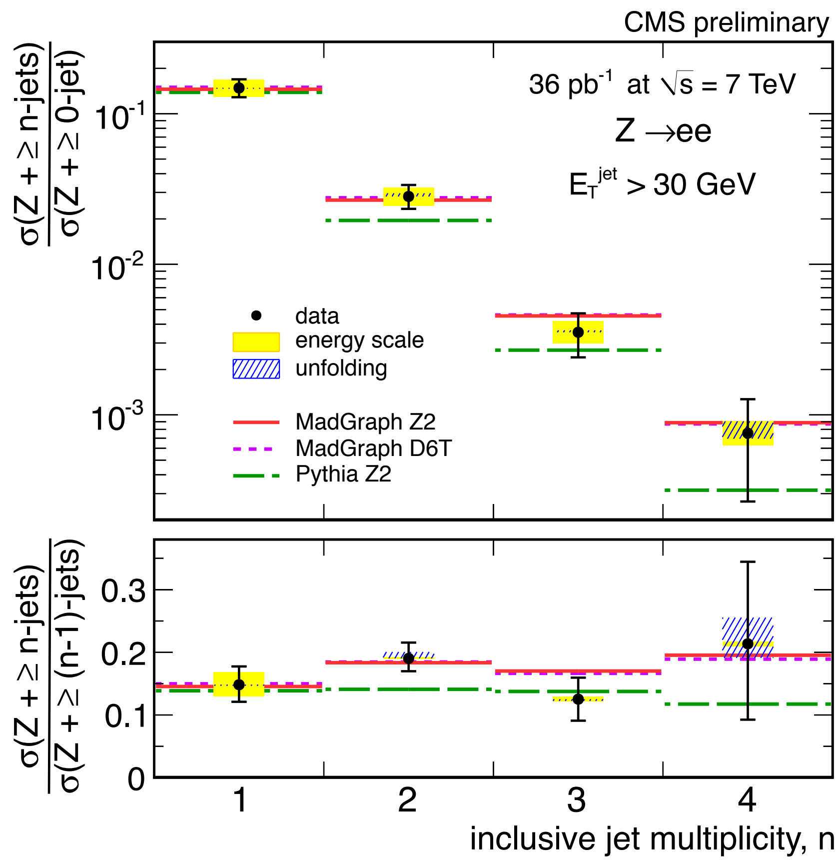
Jet multiplicity unfolded using singular value decomposition to correct for migrations
Results agree with ME matched MCs but not with LO Pythia (as expected)

Z+Jets Cross Section Ratios

Z→mumu + jets



Z→ee + jets



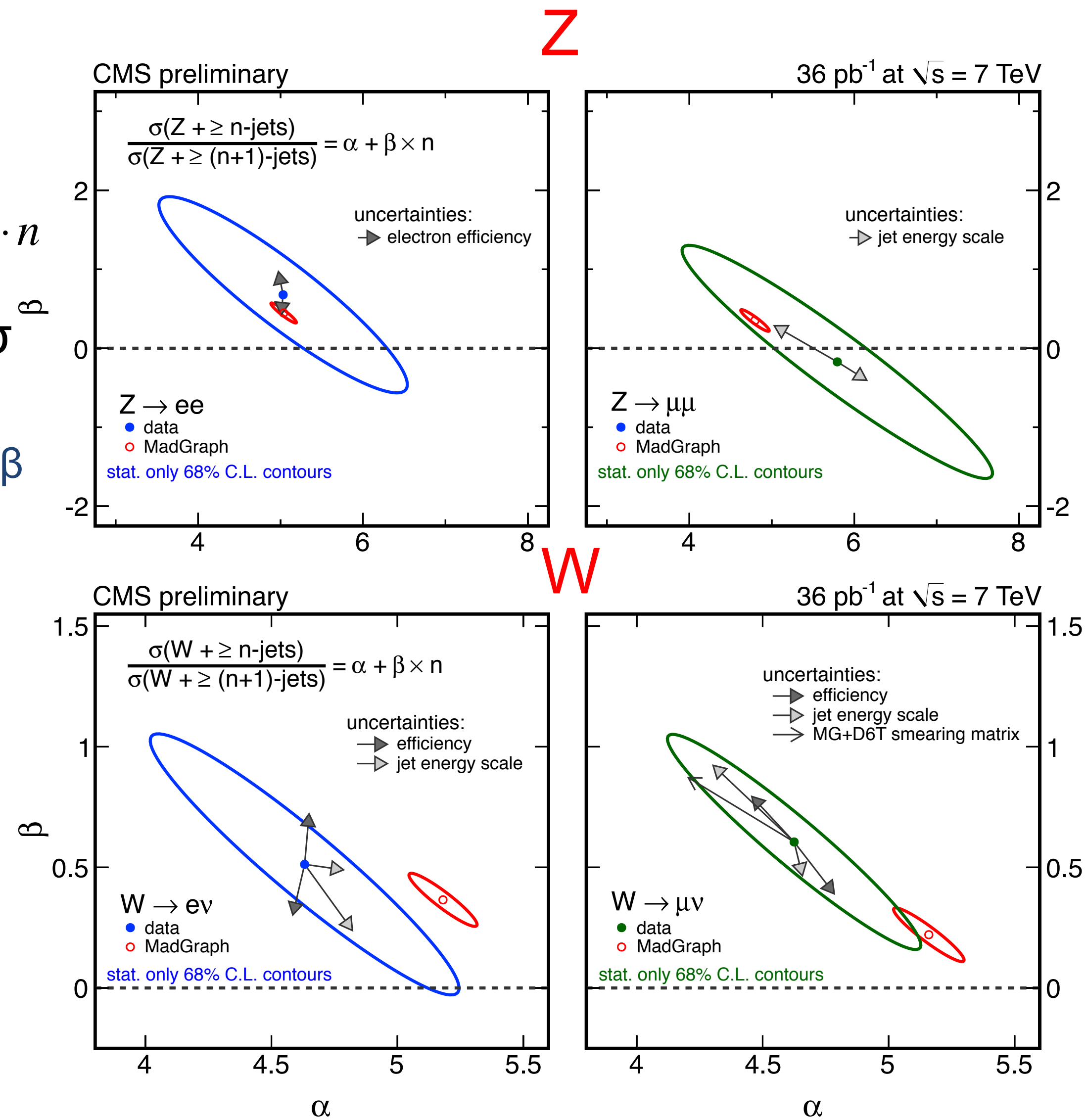
Jet multiplicity unfolded using singular value decomposition to correct for migrations
Results agree with ME matched MCs but not with LO Pythia (as expected)

Berends-Giele Scaling

- A further constraint is placed requiring

$$\sigma(V+njets)/\sigma(V+(n+1)jets)=\alpha+\beta\cdot n$$

- Naïve LO expectation of σ ratio ~ α ~ α_s^{-1}
  - Include additional deviation β
- The B-G scaling fit is similar to the previously described signal extraction
- Yield is fit for 1 jet bin, α and β fit for all channels
- Agreement between data and MC within 1 or 2 stand. dev.



Conclusions

- Presented results for V + jets using **36 pb^{-1}**
  - Jet E_T threshold of 30 GeV
- The analysis makes extensive use of data-driven methods for efficiency and background subtraction
- The results are in agreement with Madgraph Monte Carlo predictions (ME+PS)
  - Poor agreement with Pythia, as expected
- First direct measurement of Berends-Giele scaling
- **Future plans** – already collected 1.25 fb^{-1}
  - Absolute cross sections and unfolding of the jet energy spectra
  - Dijet masses
  - Comparisons with NLO MCs

Backup

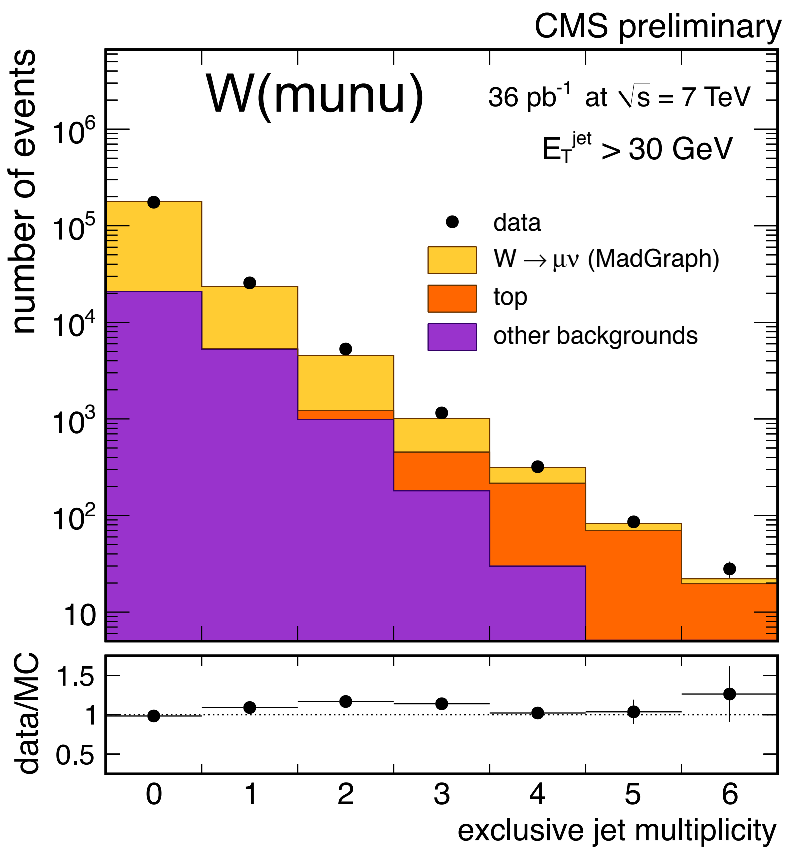
Event Selection

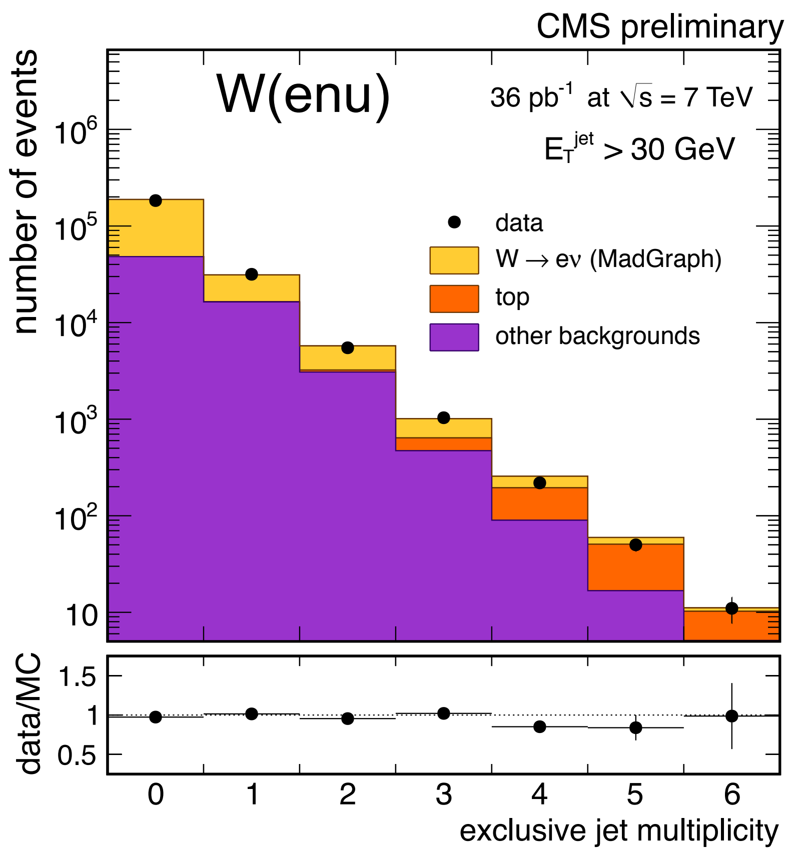
- Gsf electrons
  - $p_T > 20$ GeV
  - $|\eta| < 2.5$, exclu. $1.4442 < |\eta| < 1.566$
  - WP80 (see table)
    - ID
    - Conversion rejection
    - Isolation
      - relative to p_T, ΔR cone of 0.3
- No 2nd electron forming Z mass with 1st
  - $!(60 < m_Z < 120)$
- No muons with $p_T > 15$
- Transverse impact parameter $\delta_{xy} < 0.035$
- HLT object match
- $M_T > 20$ GeV
  - From gsfElectron and PF MET
  - Necessary for data-driven fitting

WP 80	Barrel	Endcap
Identification		
$\sigma_{i\eta i\eta}$	0.01	0.03
$\Delta\phi_{in}$	0.03	0.02
$\Delta\eta_{in}$	0.004	0.005
H/E	0.025	0.025
Conversion rejection		
Missing hits	0 OR	
Dist	(0.02 AND	
$\Delta\cot(\theta)$	0.02)	
Isolation		
Track iso	0.09	0.04
Ecal iso	0.07	0.05
Hcal iso	0.10	0.025

Exclusive jet multiplicity for W

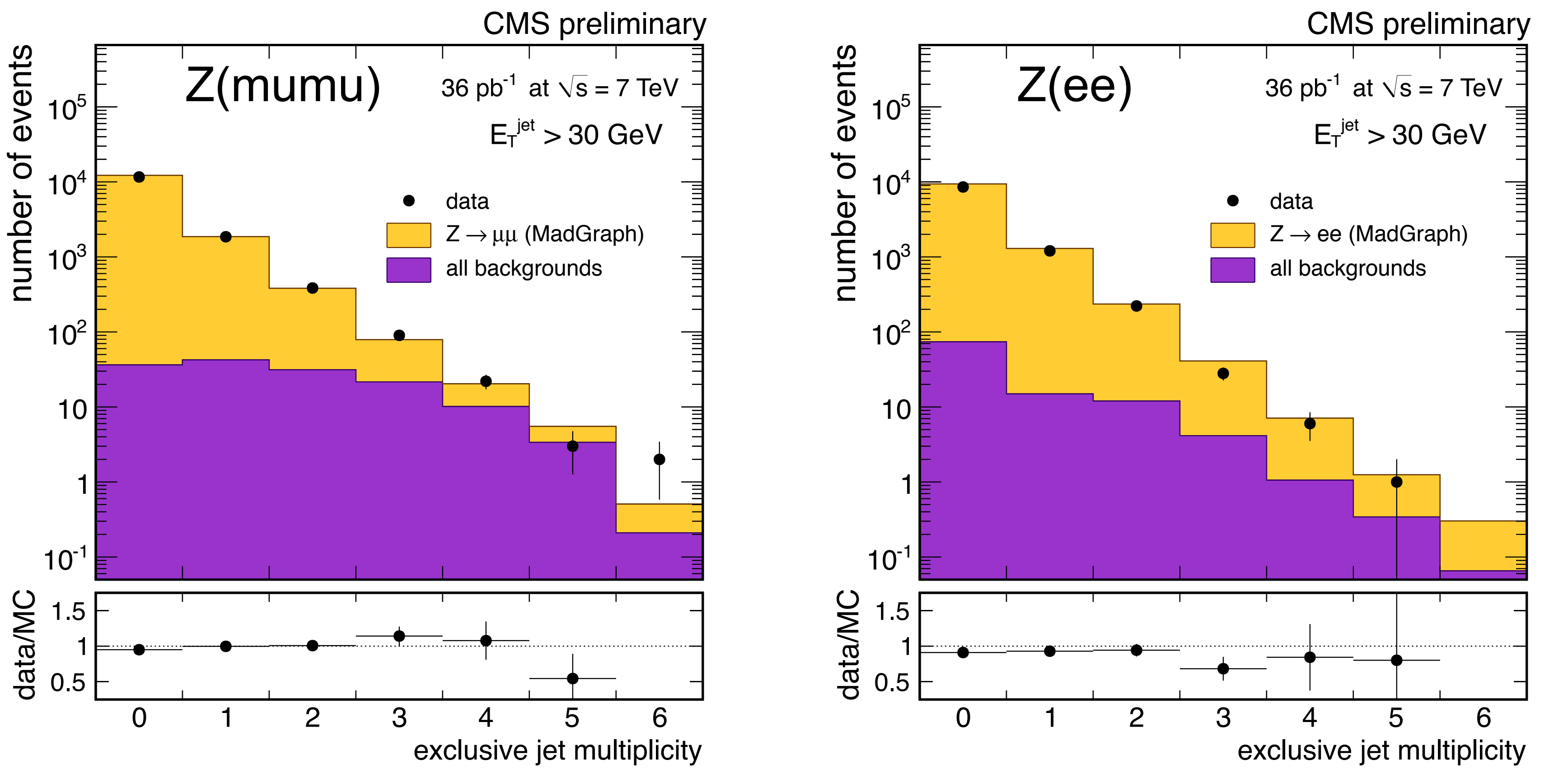
Data – MC comparisons of raw jet rates for the W signal region, M_T > 50 GeV, showing good agreement with MADGRAPH + PYTHIA





Exclusive jet multiplicity for Z

Data – MC comparisons of raw jet rates for the Z signal region showing good agreement with MADGRAPH + PYTHIA



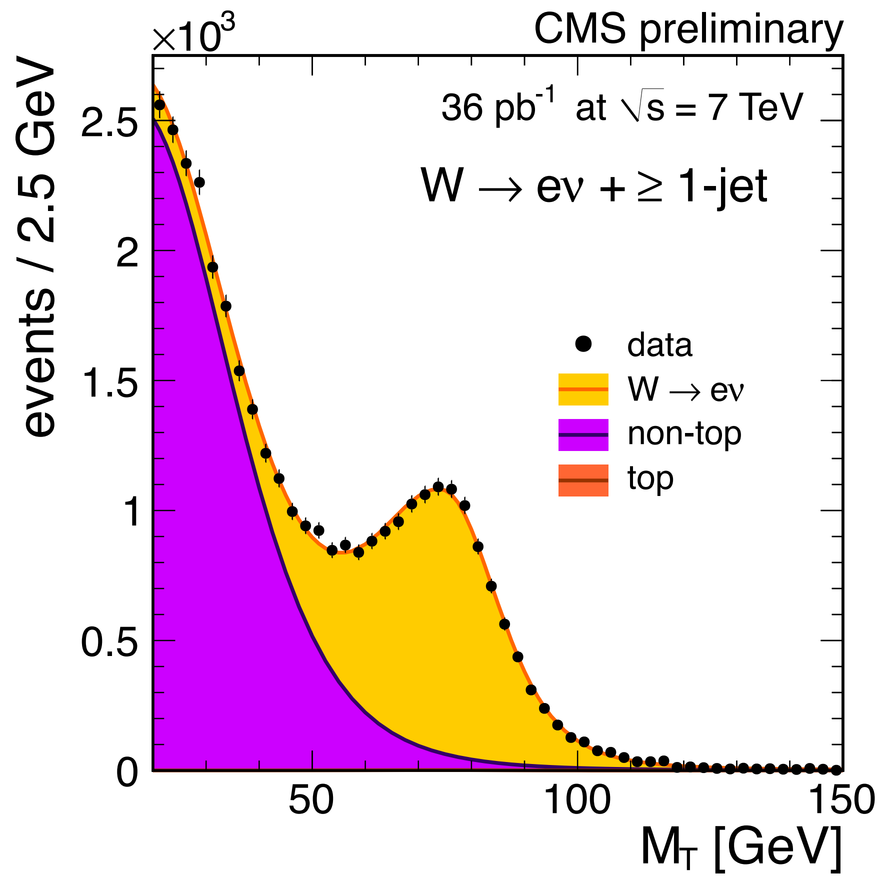
Efficiency of Lepton Selection

- Measured using tag and probe on Z+jets sample
  - Invariant mass for pass and fail samples are fit for signal and background
  - Efficiency found as a function of jet bin
- W: $\varepsilon_{reconstruction} \times \varepsilon_{selection} \times \varepsilon_{trigger}$
- Z: $\varepsilon_{reconstruction} \times \varepsilon_{selection} \times \varepsilon_{trigger} \times \varepsilon'_{reconstruction} \times \varepsilon'_{looser\ selection}$
  - Where ε is for probe $p_T > 20$ GeV, and ε' is for probe $p_T > 10$ GeV

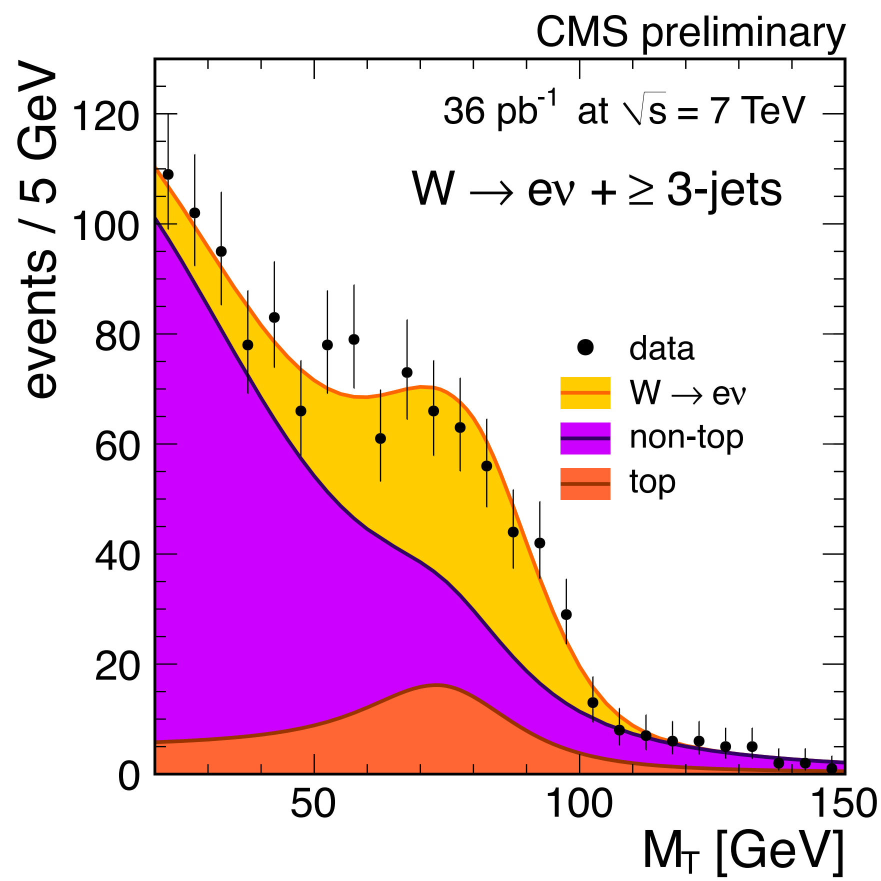
- Electrons: Uncertainty from choice of fitting line shape
  - BW+CrystalBall vs Double Crujiff (both with exp bkgd)
  - Averaged both fits for the central value
  - Using jet $p_T > 15$ GeV for jet counting for adequate statistics
- Muons: determined as a function of jet bin, p_T, and eta
  - Measured by p_T and eta for 0 and 1 bins, extrapolated to n > 1

W→enu M_T Fit Results

1 jet events

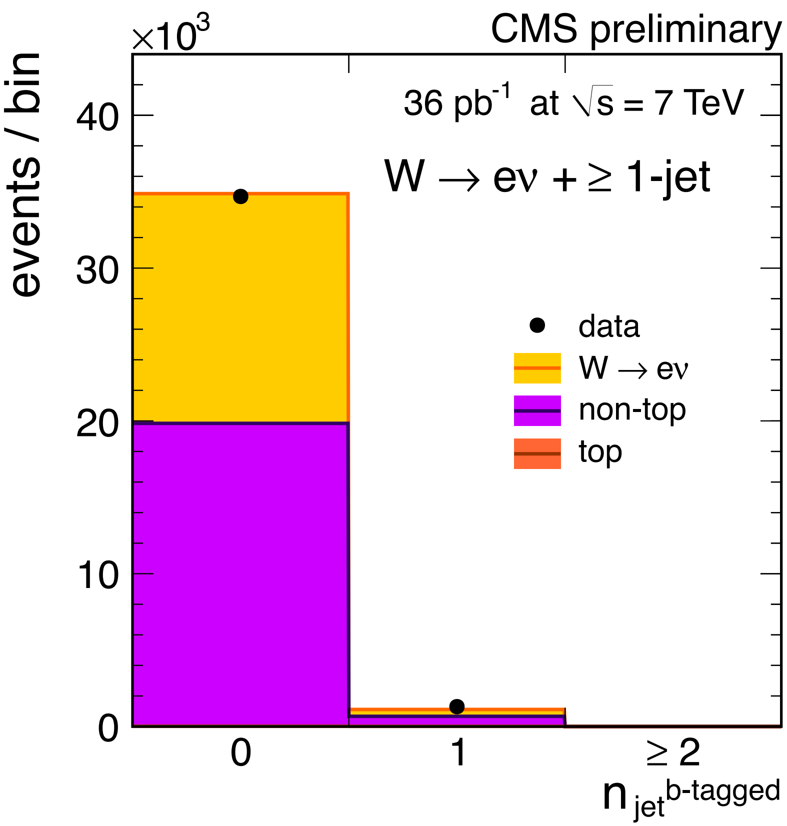


3 jet events

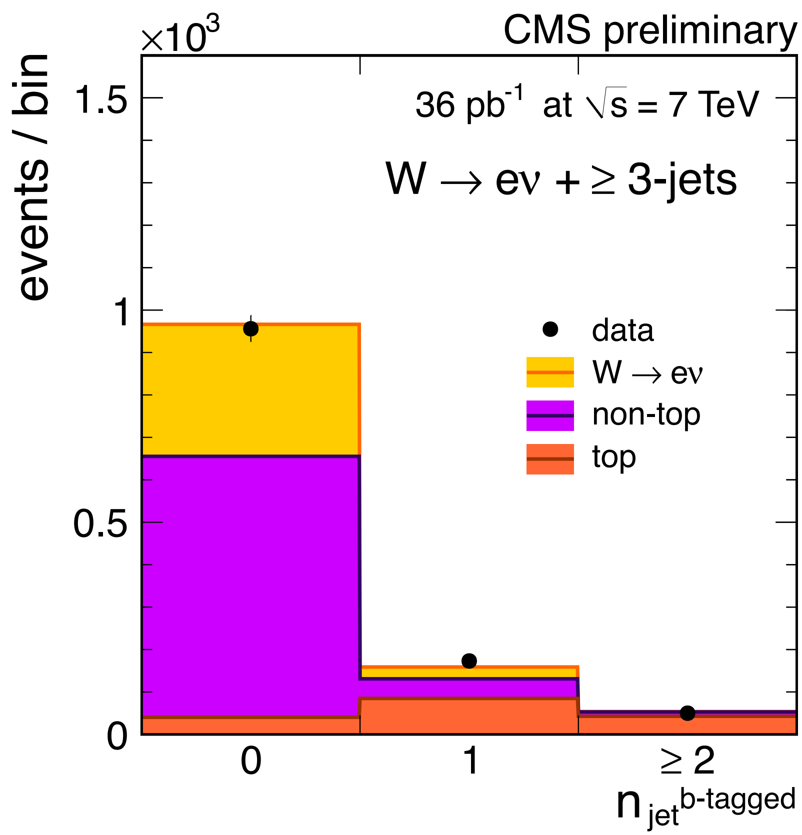


W→enu n_{bjets} Fit Results

1 jet events

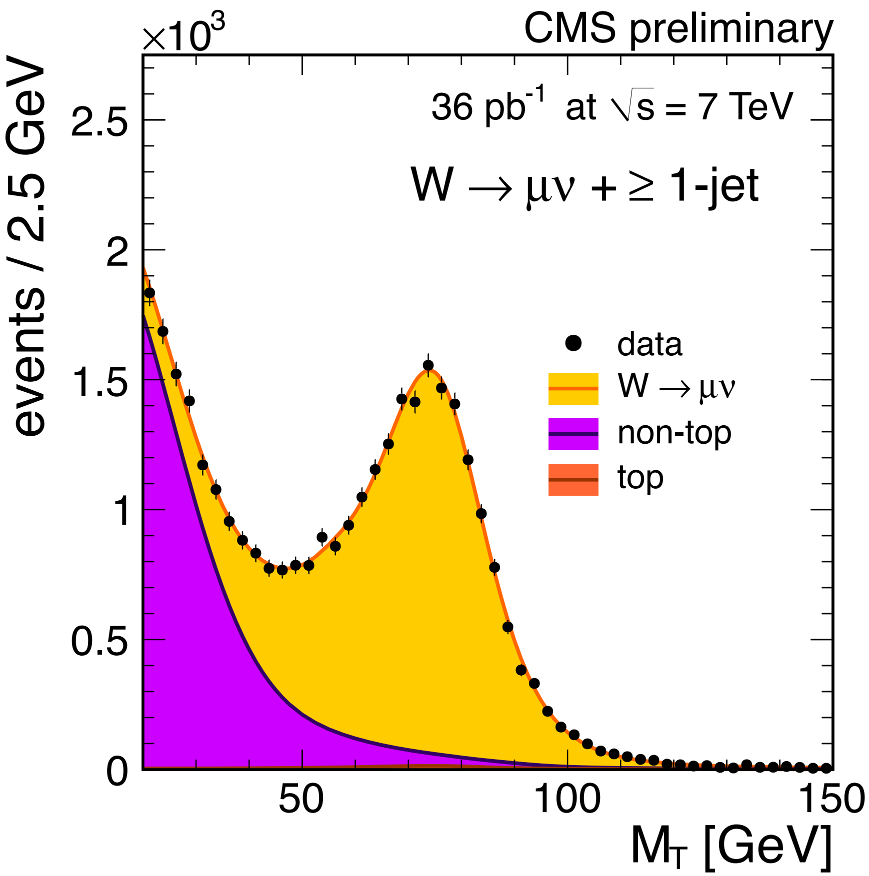


3 jet events

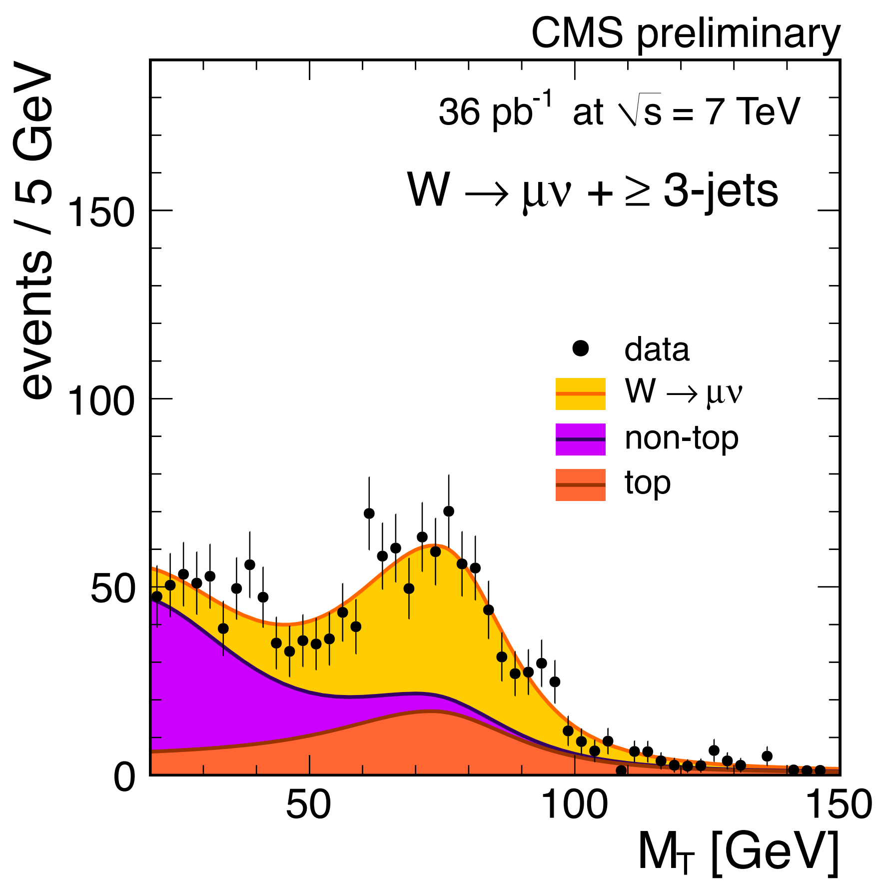


W→munu M_T Fit Results

1 jet events

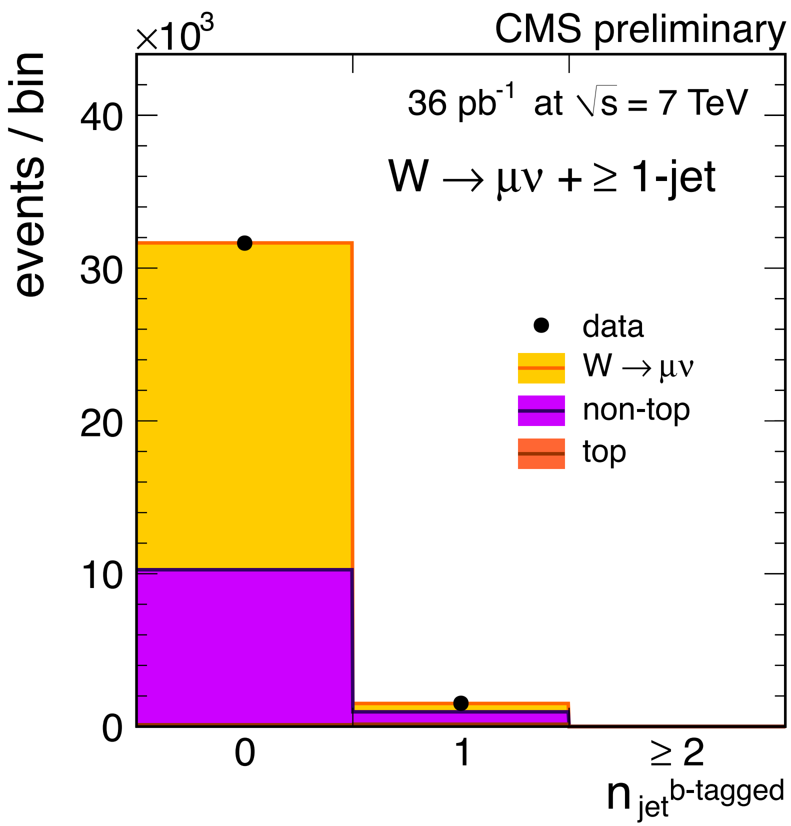


3 jet events

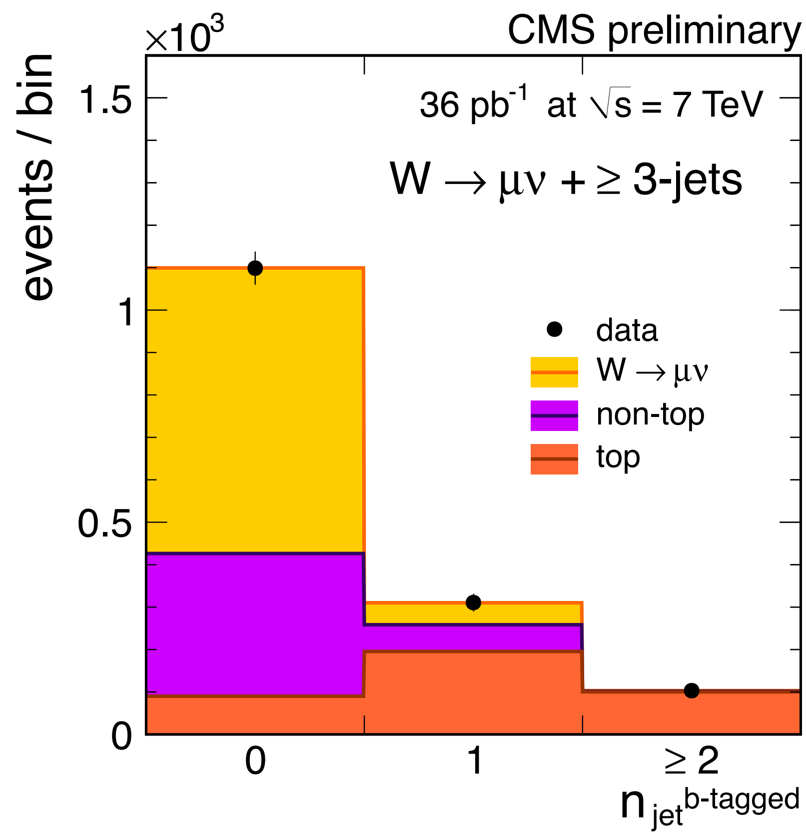


W→munu n_{bjets} Fit Results

1 jet events



3 jet events



Total systematic uncertainties for W→lnu

Official JetMET recommendations followed using W(mnu)+Jets data and MC samples

- JES uncertainties considered (correlated between jet bins)
  - Jet Energy Scale
  - Jet Energy Resolution
  - MET Resolution
  - Flavor dependence
  - PU residual 500MeV offset
  - CMSSW Release

- Lepton efficiency and signal extraction uncertainty (from varying the constrained parameters) are uncorrelated

Uncertainties on jet rate in $W \rightarrow e\nu$ events [%]					
Jet multiplicity	0	1	2	3	≥ 4
Jet counting	∓ 5	± 8	$^{+11}_{-10}$	$^{+14}_{-12}$	$^{+16}_{-15}$
Lepton efficiency	± 3	$^{+6}_{-5}$	$^{+7}_{-6}$	± 10	$^{+24}_{-12}$
Signal extraction		± 0.1	± 0.4	± 2.9	± 8.5
Total systematics	± 6	± 10	$^{+13}_{-12}$	$^{+18}_{-16}$	$^{+30}_{-21}$
Statistical uncertainty	± 0.3	± 1.0	± 2.4	± 7.5	± 22
Uncertainties on jet rate in $W \rightarrow \mu\nu$ events [%]					
Jet multiplicity	0	1	2	3	≥ 4
Jet counting	∓ 5	± 8	$^{+11}_{-10}$	$^{+14}_{-12}$	$^{+16}_{-15}$
Lepton efficiency	± 3	± 6	± 4	± 10	± 17
Signal extraction		± 0.1	± 0.4	± 2.9	± 8.5
Total systematics	± 6	± 10	$^{+13}_{-12}$	$^{+19}_{-17}$	± 26
Statistical uncertainty	± 0.2	± 0.8	± 2.3	± 6.5	± 27

Total systematic uncertainties for Z→ll

- Same JES uncertainties as for W
- Similar efficiency
- No fit systematics
  - Parameters floated

Uncertainties on jet rate in $Z \to e^+e^-$ events [%]					
Jet multiplicity	0	1	2	3	≥ 4
Jet counting	∓ 5	± 8	$^{+11}_{-10}$	$^{+14}_{-12}$	$^{+16}_{-15}$
Efficiency	± 3	$^{+6}_{-5}$	$^{+7}_{-6}$	± 10	$^{+24}_{-12}$
Total systematics	± 6	± 10	$^{+13}_{-12}$	$^{+18}_{-16}$	$^{+30}_{-21}$
Statistical uncertainty	± 1.0	± 3.0	± 8.0	± 20	± 47
Uncertainties on jet rate in $Z \to \mu^+\mu^-$ events [%]					
Jet multiplicity	0	1	2	3	≥ 4
Jet counting	∓ 5	± 8	$^{+11}_{-10}$	$^{+14}_{-12}$	$^{+16}_{-15}$
Efficiency	± 3	$^{+6}_{-5}$	$^{+7}_{-6}$	± 10	$^{+24}_{-12}$
Total systematics	± 6	± 10	$^{+13}_{-12}$	$^{+18}_{-16}$	$^{+30}_{-21}$
Statistical uncertainty	± 1.1	± 2.7	± 5.2	± 18	± 35

Propagation of uncertainties

- Unfolding is performed on exclusive n-jet bins (i.e., n=0, n=1, n=2, n=3, n≥4)
  - After unfolding the inclusive rates are calculated
- Unfolding is done multiple times for uncertainty calculations:
  - with statistical errors only
  - with statistical + uncorrelated systematics
    - Lepton efficiency, fit
  - central values shifted by correlated systematics
    - Jet counting
  - changing the unfolding method
    - Different tune (Z2 vs D6T), generator (Madgraph vs Pythia), or algorithm (SVD vs Bayes)

Jet Selection

- AntiK$_T$5 Particle Flow Jets
  - L1FastJet, L2+L3+L2L3Residual corrections
  - p_T > 30 GeV
    - Sensitive to the matrix element
    - Smaller pile-up correction needed
  - |eta| < 2.4
  - Remove if selected electron is within $\Delta R < 0.3$
  - Muons are excluded from jet list before clustering

PFlow Jet ID variable	Selection
chargedEmEnergyFraction	< 0.99
neutralHadronEnergyFraction	< 0.99
neutralEmEnergyFraction	< 0.99
chargedHadronEnergyFraction	> 0
chargedMultiplicity	> 0

- Jet energy uncertainty:
  - Add in quadrature: JEC + PU + Flavor
    - JEC dependent on eta and p_T (~3%)
    - PU dependent on jet p_T (~1.2 % for 30 GeV jet)
    - Flavor set to 2-3%

Selection Efficiency: Tag & Probe

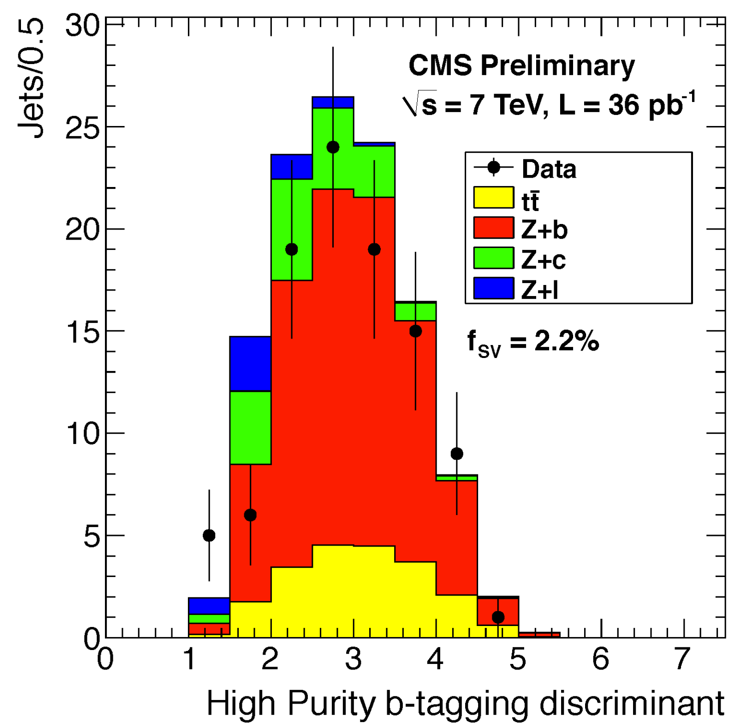
- Use data-driven "Tag-and-probe" method as part of the efficiency calculation
  - Start from Z/γ^* + jets data sample (very little background)
    - Two electrons forming an invariant mass, $60 < m_{ee} < 120$ GeV
  - One electron, the "tag", passes full selection (reduces background)
  - Second "probe" electron is divided into two samples
    - Passing the desired requirement
      - i.e., reconstruction, WP80, or HLT
    - Failing the same requirement
  - Fits are performed on the passing and failing samples to extract the number of Z electrons from the remaining background
  - Efficiency is the number of probes passing the current requirement relative to the total number of probes, e.g., $\varepsilon_{trigger} = N_{trig} / N_{WP80}$
    - $\varepsilon_{T\&P} = \varepsilon_{reconstruction} \times \varepsilon_{selection} \times \varepsilon_{trigger}$

See T&P fits

Observation of Z+b with Z→ll

CMS-PAS-EWK-10-015

- Z+b benchmark for high tanβ MSSM Higgs searches
- H+b NLO prediction has large uncertainties
  - 30% scheme dependence (variable vs fixed flavor schemes)
  - Z+b data should help to clarify
- Select Z+≥1 jet events
  - Jet ET > 25 GeV; separated from lepton by ΔR >0.5
  - Require secondary vertex
  - M_T < 40 GeV to reject top
  - 29 dieletron and 36 dimuon events after selection
- B-tagging descriminant variable built from flight distance between PV and SV
  - SSVHE: high efficiency selection with ≥2 tracks attached to SV
  - SSVHP: high purity selection with ≥3 tracks attached to SV

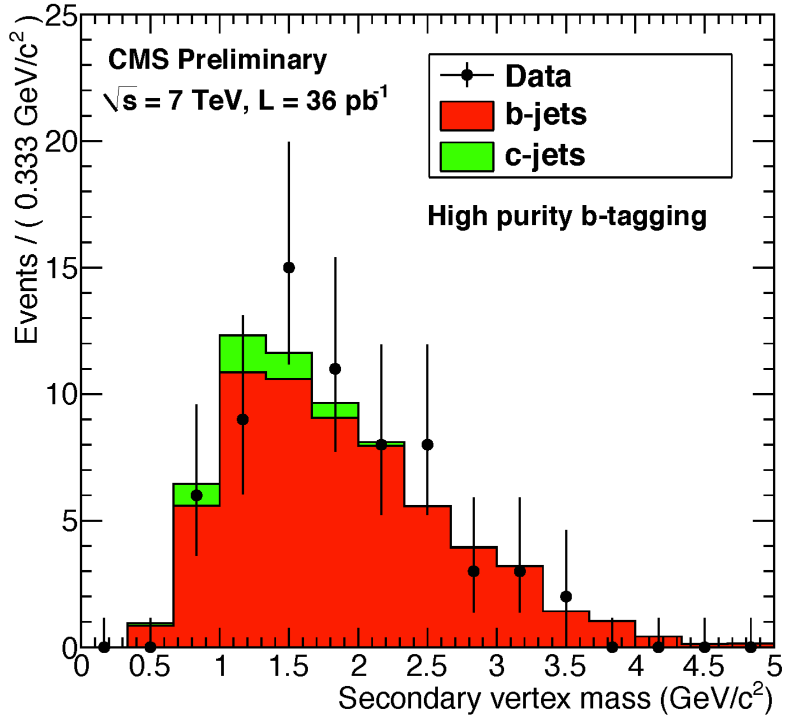


Z+b/Z+jets Ratio

CMS-PAS-EWK-10-015

- Determine Z+b purity in selected sample from binned ML fit:
  - of SV mass or B-tag discriminant shape
  - MC templates for b, c, ligh-jet components

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



→ Results compatible with Madgraph (scaled to NLO) & MCFM

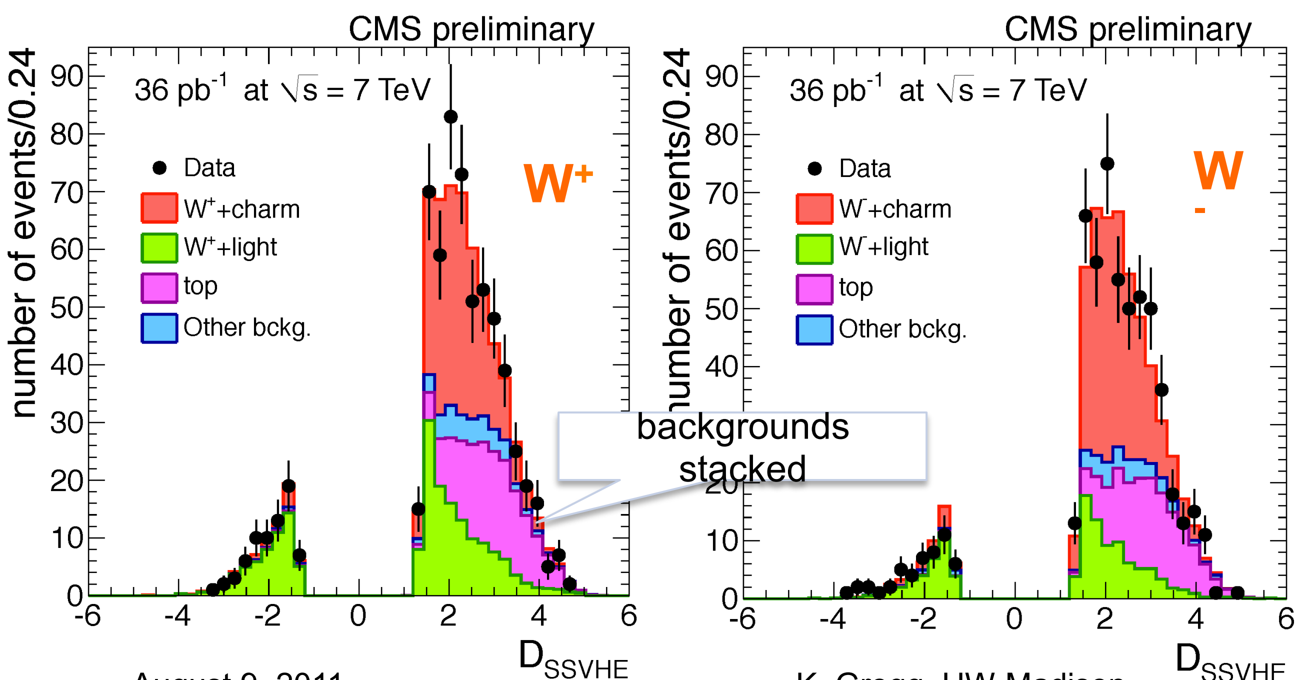
→ Limited statistics: scheme dependence cannot be resolved yet

Sample	$\frac{pp\to ee+b+X}{pp\to ee+j+X}$ (%) $p_T^e > 25$ GeV, $\lvert\eta^e\rvert < 2.5$	$\frac{pp\to\mu\mu+b+X}{pp\to\mu\mu+j+X}$ (%) $p_T^\mu > 20$ GeV, $\lvert\eta^\mu\rvert < 2.1$
Data SSVHE	$4.3 \pm 0.6(stat) \pm 1.1(syst)$	$5.1 \pm 0.6(stat) \pm 1.3(syst)$
Data SSVHP	$5.4 \pm 1.0(stat) \pm 1.2(syst)$	$4.6 \pm 0.8(stat) \pm 1.1(syst)$
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	$4.3 \pm 0.5(theory)$	$4.7 \pm 0.5(theory)$

Study of W+c with W→μν

CMS-PAS-EWK-11-013

- Process dominated by sbar g → W^+ cbar and sg→ W^- c
- Probes s and sbar content of proton
- Select W+≥1 jet events in muon channel
  - M_T > 50 GeV to reject QCD background
  - Jet E_T > 20 GeV
  - Require SV with ≥2 associated tracks and significantly displaced from PV
- B-tagging descriminant variable D_{SSVHE} built from flight distance between PV and SV



ML fit of signal, top, W +light quarks, DY components to observed D_{SSVHE}

Negative values of D_{SSVHE} due to detector resolution effects and well suited to constrain light quark component

$(W^{+}+c)/(W^{-}+c)$ and $(W+c)/(W+jets)$ ratios

CMS-PAS-EWK-11-013

- For leading jet with E_T>20GeV and $|\eta|$<2.1:

$$R_c^{\pm} \equiv \sigma(W^{+}\bar{c})/\sigma(W^{-}c)$$

$$R_c^{\pm} = 0.92 \pm 0.19\,(stat.) \pm 0.04\,(syst.)$$

  - Leading source of sys error: PDf uncertainties, pile-up effect and background templates

$$R_c \equiv \sigma(W+c)/\sigma(W+\ jets)$$

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

  - Leading source of sys error: Tracking resolution

➜ Results in agreement with NLO predictions