W+jets as a top background at the LHC

Ayana T Holloway Arce
Berkeley Workshop on V+jets

Introduction

- What are LHC experiments planning to measure with top quarks?
  - strong and electroweak production rates, mechanisms
  - top mass to ~ 1 GeV precision
  - weak interaction vertex: W polarization, CKM constants…
  - rare decays
  - calorimeter energy scale calibrations
  - efficiency to identify b jets
- For many precision measurements: ~insensitive to V+jets
  - I'll show the techniques, review the existing studies
- **Nevertheless:** we wish to understand the content/kinematics of top-like samples
  - study: how would we deal with W+jets in an inclusive 'lepton + jets' cross section measurement?

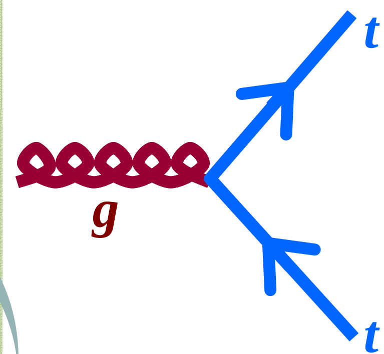
Top quark prospects

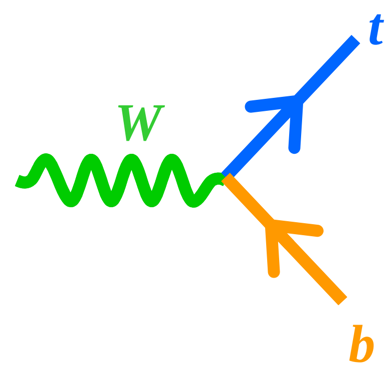
Pair production

- Obviously very important:
  - dominant channel: ~800 pb
  - 'laboratory' for t properties (except V_{tb})
  - now biggest background to single top
- Signatures: B($t \rightarrow W+b$) ~100%
  - 2 W bosons, 2 (bottom) jets

Electroweak production

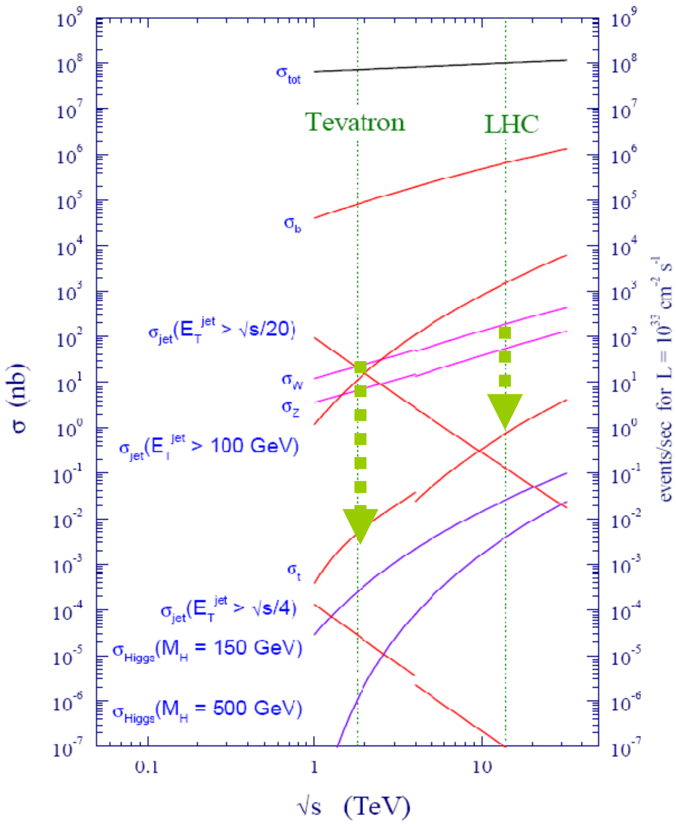
- Early improvement over Tevatron:
  - tq: 247 pb at LHC
  - Wt: 62 pb
  - tb: 11 pb
- Signatures: $t \rightarrow W+b$ and q, W, or b
  - apply jet vetoes, kinematic cuts

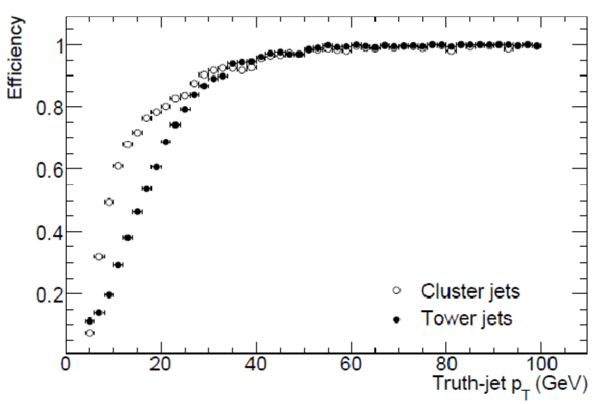




Rate of ~1 evt/s and selection efficiency ~ a few %: statistics are no obstacle!

a new regime





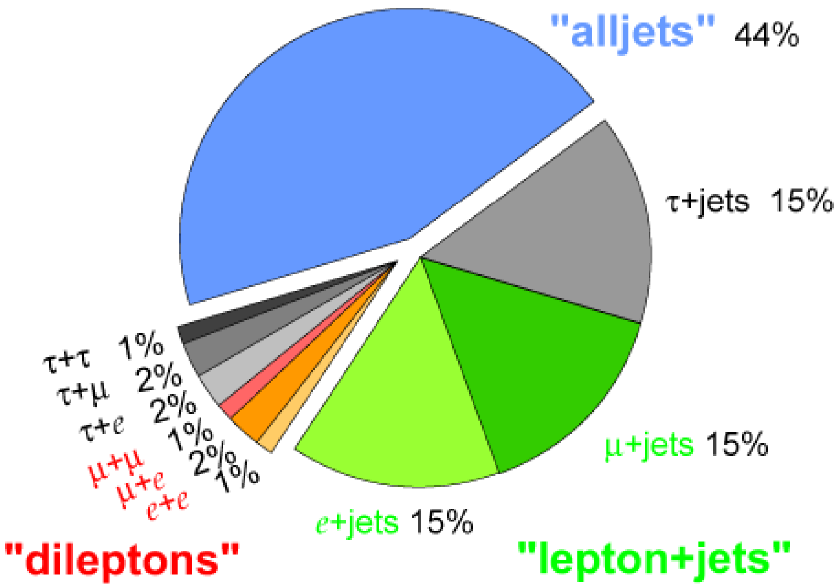
Tevatron

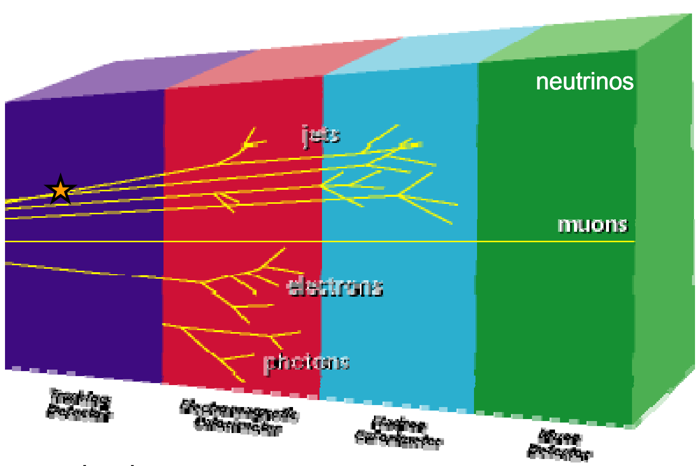
- n(tag) + n(lepton) >= 2
- "jets" ~ 15 GeV E_T
- biggest σ uncertainties:
  - Jet energy scale
  - b-tag efficiency (~5%)
  - W+HF fraction uncertainty

LHC

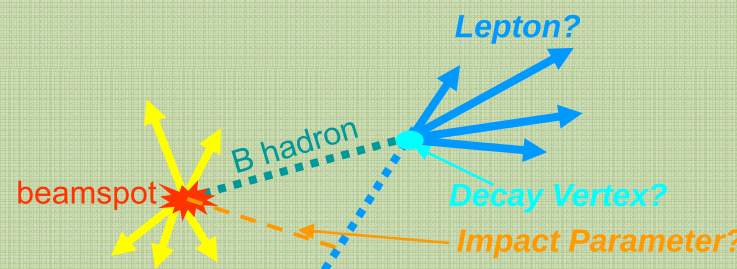
- n(tag) + n(lepton) >= 1
- "jets" ~ 25 GeV E_T
- new environment, production mechanism
  - Jet energy scale, pileup, ISR/FSR, …

$t\bar{t}$ measurements: prerequisites





- Event reconstruction:
  - good e/μ identification
  - calibrated jet/missing E_T
  - heavy flavor tagging



- Note: 'lepton+jets' eventually loses luster!

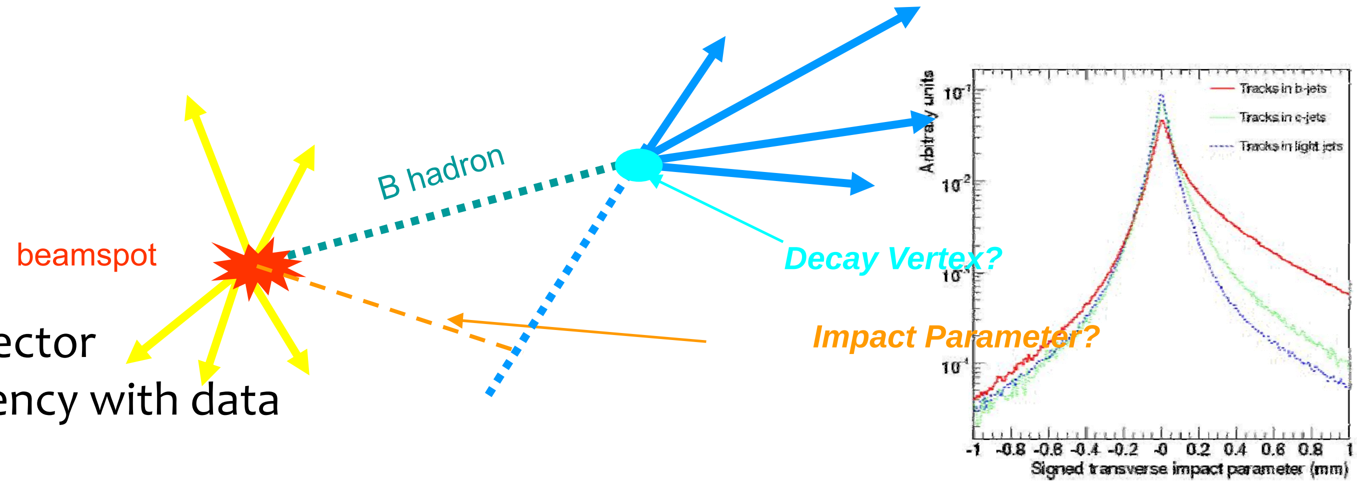
	Lepton+jets	Dilepton
σ(tt): 1 fb^{-1}	1.2⊕10%	0.9⊕11%
t mass: 10 fb^{-1}	0.4⊕1.3%	0.5⊕1.1%

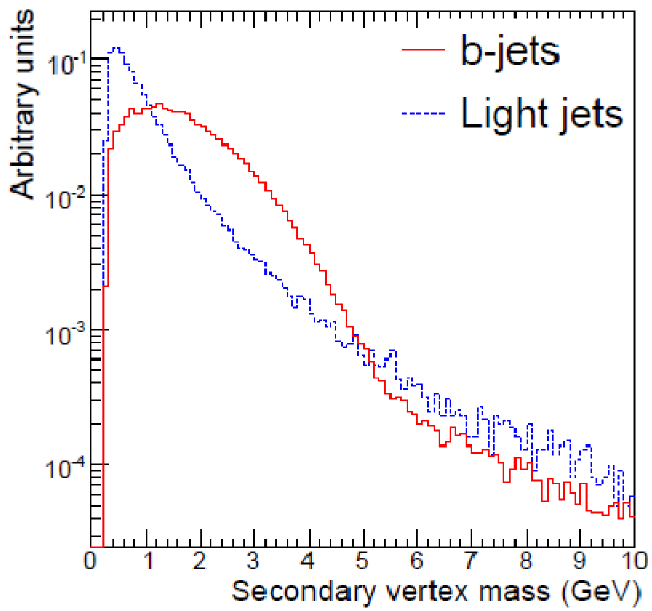
CMS: J. Phys. **G34** (2007)

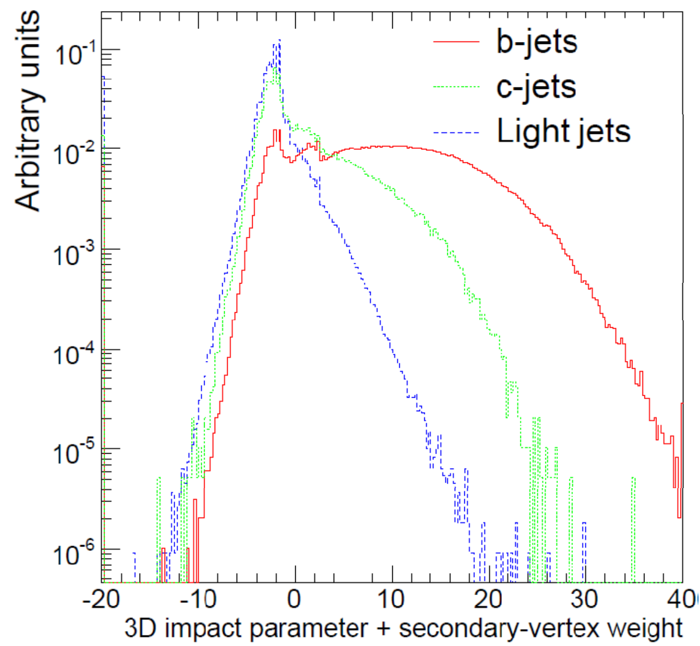
aside: ATLAS b-tagging

two tasks:

1. align inner detector
2. calibrate efficiency with data

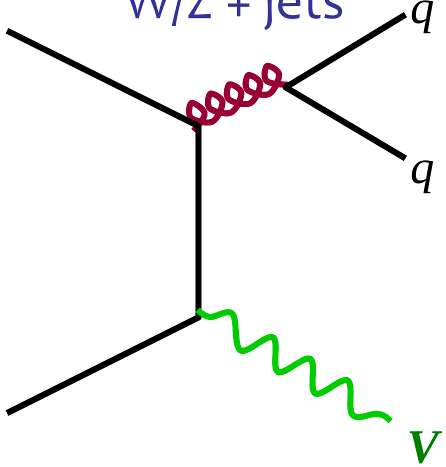




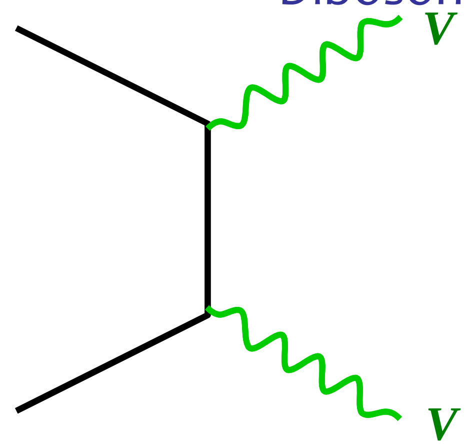


$t\bar{t}$ measurements: backgrounds

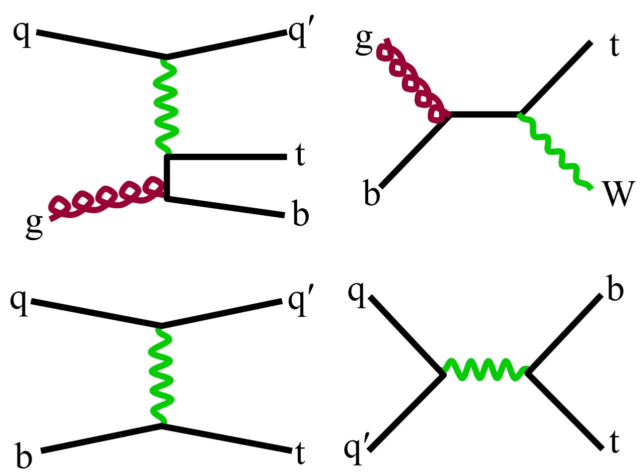
W/Z + jets



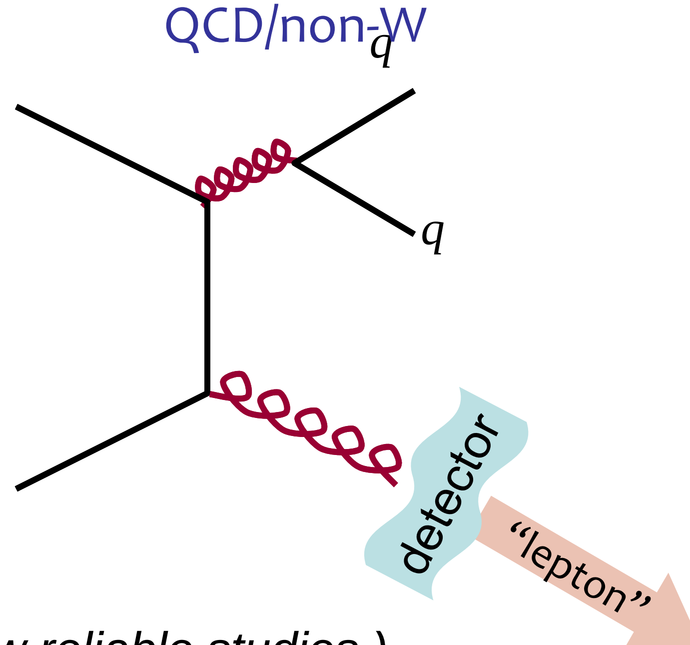
Diboson



Single top

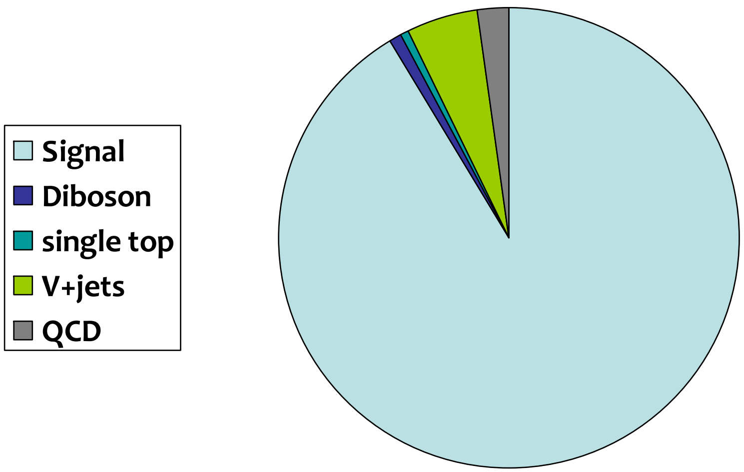


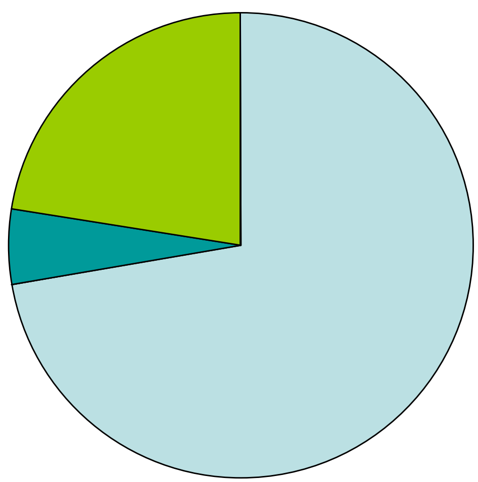
QCD/non-W



(few reliable studies)

tt measurements: backgrounds





Tevatron l+jets w/ 1 b-tag

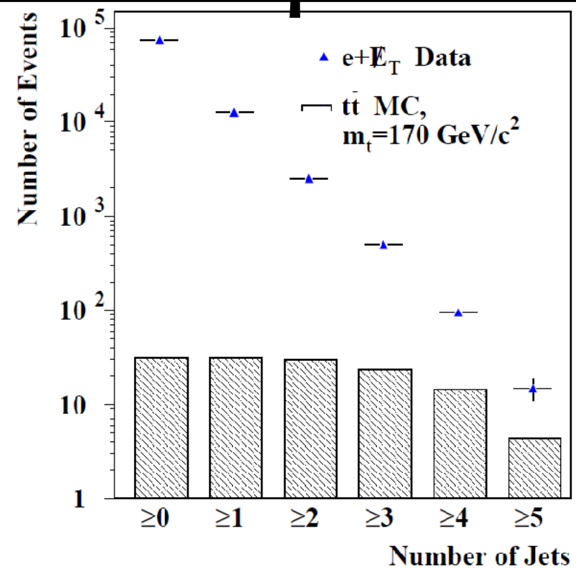
- e/μ + 3 jets above 30 GeV, 1 tag, H_T cut
- Initial purity ~0.4
- tagging 'tight': 60 ± 3%

LHC l+jets inclusive

- e + 3 jets above 40 GeV + 1 above 25
- ~75% purity before tagging

Chronology: W in *tt* →lepton+jets

- Perennial techniques:
  - Dibosons: fixed
  - Single top: fixed
  - QCD: loose W selection
- Run-I:
  - Measure σ for W+≥3jets
    - correct for top signal
  - Measure (***b+c+q***) tag rate of dijet sample
  - apply to W+jets sample
    - assumes F_{HF} isn't sample dependent
    - Systematic: 10% variation among samples (DØ)

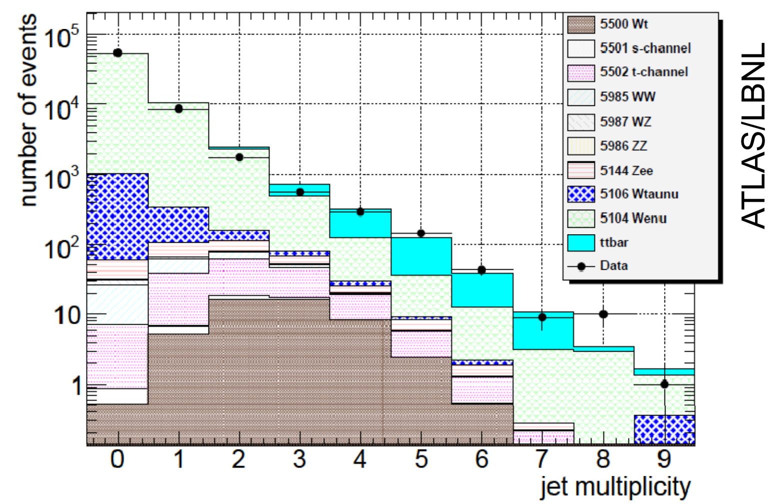


DØ PRD67 (2003)

- (mostly) Run-II:
  - basically, separate "mistags:"
    - in tagger efficiency
    - in W+jets sample
  - use (calibrated) predictions for F_{HF}(W)

Chronology: W in *tt* →lepton+jets

- Perennial techniques:
  - Dibosons: fixed
  - Single top: fixed
  - QCD: loose W selection
- Early LHC results:
  - often, assume no b-tagging, so separating top/W+jets requires other tools
    - top mass distributions, or
    - longer extrapolation in N(jets)

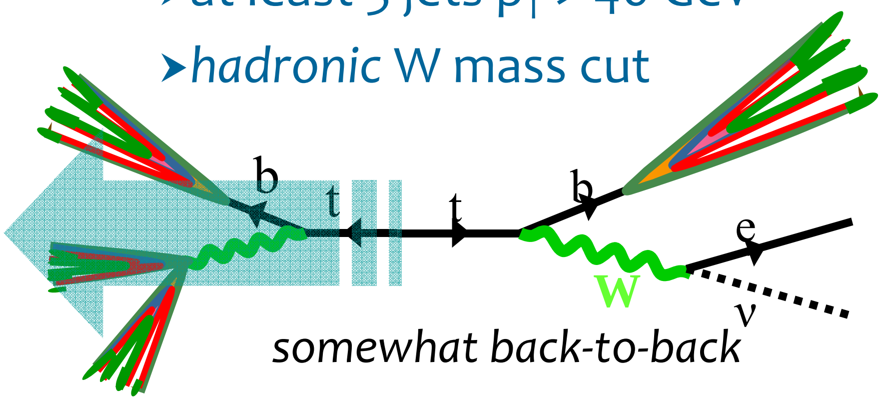


- later LHC results:
  - b-tag (efficient against W+jets)
  - top sample will calibrate this efficiency
    - reintroducing sensitivity to HF fractions in W+jets?

ATLAS *tt* l+jet event selection

fitting 3-jet mass

- Event selection:
  - e/μ: |η| < 2.4, p_T > 20
  - missing ET > 20
  - at least 4 jets p_T > 25 GeV, |η| < 2.5
  - at least 3 jets p_T > 40 GeV
  - *hadronic* W mass cut



- Estimating σ_{tt}
  - Fit for signal contribution in hadronic top mass

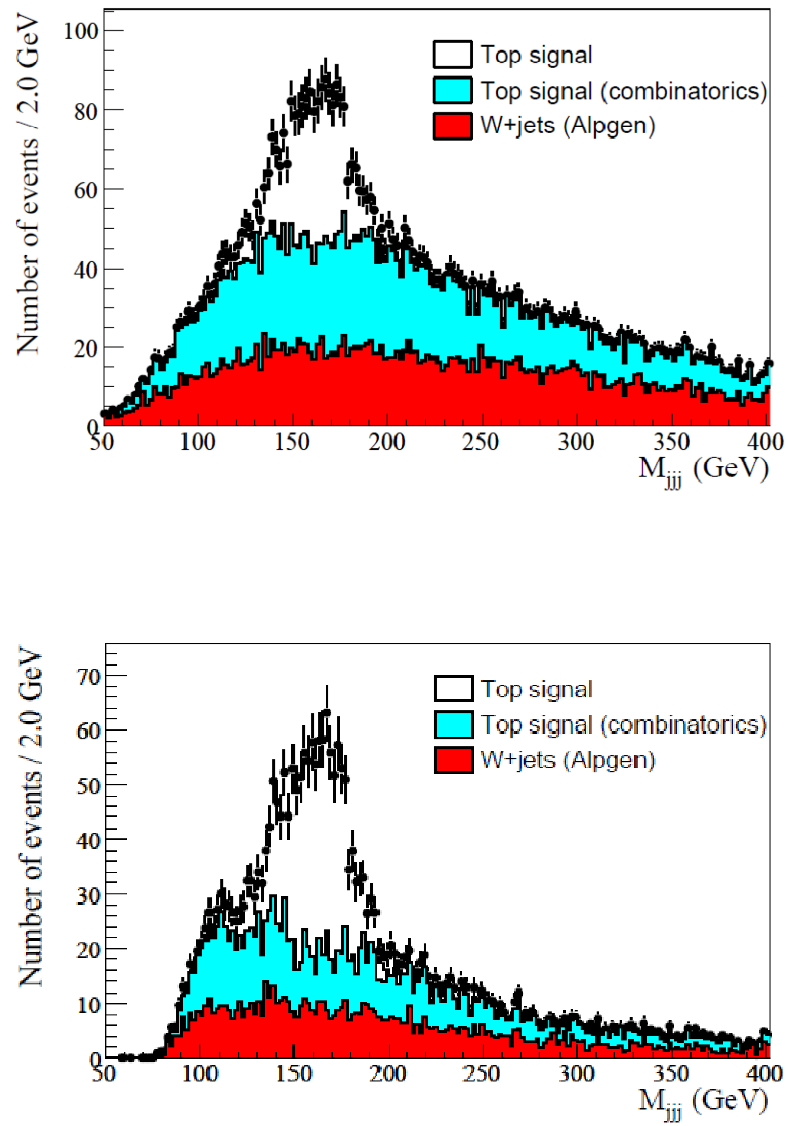
fitting N_{jet} distribution

- Event selection:
  - e channel: |η| < 2.4, p_T > 25
  - missing E_T > 25 GeV
  - at least 4 jets p_T > 25 GeV, |η| < 2.5
  - $m_t(W)$ > 40 GeV
- Estimating σ_{tt}
  - 'fit' N_{jet} distribution

Background sensitivity in mass fit

- Hadronic W mass cut
  - Select the three jets with the highest sum p_T
  - Make all 3 “W” masses, and
  - remove event if **no** $|m_W - m_{jj}| < 20$ GeV

100 pb^{-1}



c.f. ATLAS preliminary ATL-PUB 2005-024

Background sensitivity in mass fit

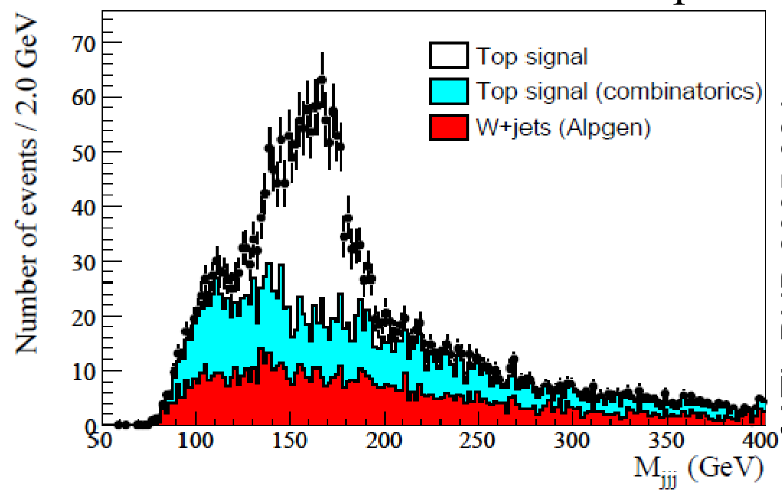
Details:

- Signal model
  - MC@NLO + HERWIG
- Background model
  - ALPGEN: W+4 jets (only!) + PYTHIA
  - blue is misreconstructed *tt*
- Cross section is derived from fit to gaussian signal + background model
  - $\sigma \sim N_{gauss} * F_{MC}$ where
  - $F_{MC} = <N_{gauss}/N_{gauss}>$ total
  - reduce sensitivity to **peaking** backgrounds with fixed width of signal
  - argument: what's in the continuum doesn't matter
- optimism: use very weak constraints on BG shape (4-7 degree polynomial)
  - there could be **new** BG components…

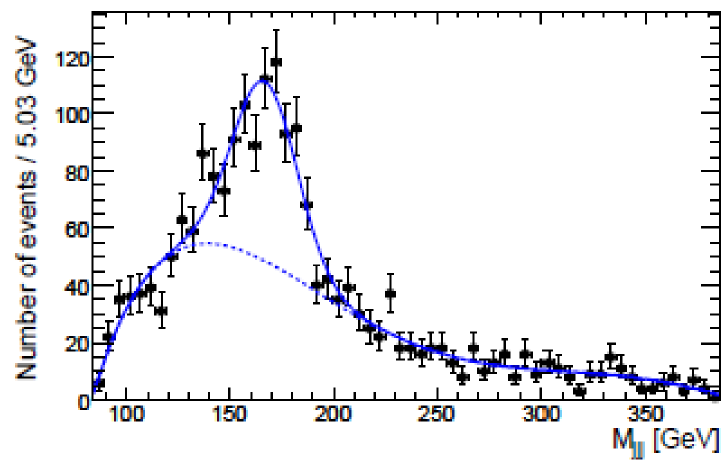
Recent work:

  - improved BG model (matched ALPGEN 2.06 + HERWIG samples)

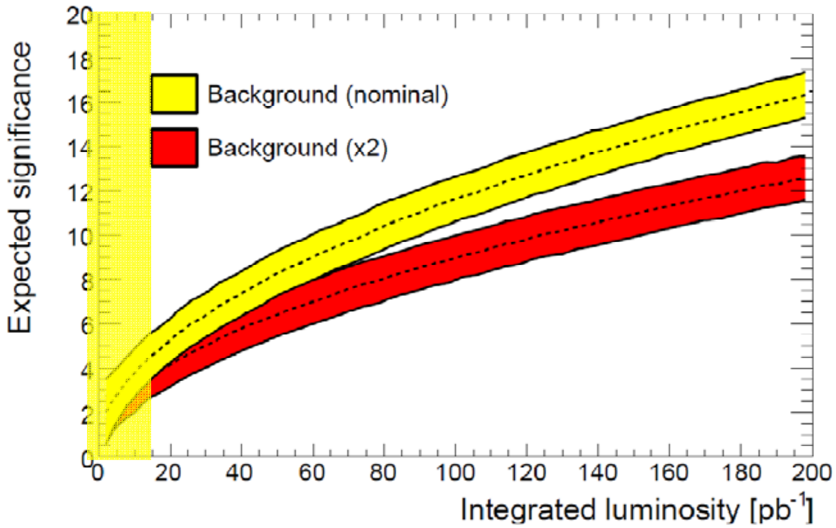
100 pb^{-1}



cf ATL-PUB 2005-024



The bigger picture: an N_{jet} fit



- We need more info to understand W+j backgrounds → look at N_{jets}
  - especially in first data: mass fit can be unlucky, so study is designed for ~16 pb^{-1}
- **Fit** simultaneously to top and W+j shapes
  - others are 'fixed'

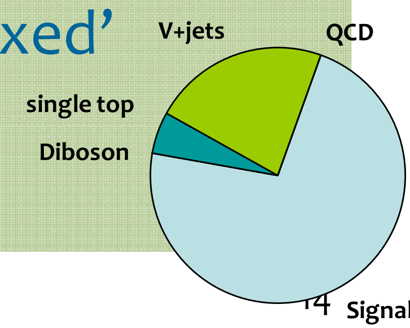
$$\sigma_{tt} = \frac{N_{t\bar{t}}}{\mathscr{L}\,\varepsilon_{\text{trigger}}\left(A\varepsilon_{\text{reco}}\right)}$$

$N_{obs} - N_W - N_{other}\big|_{\geq 4\text{ jets}}$

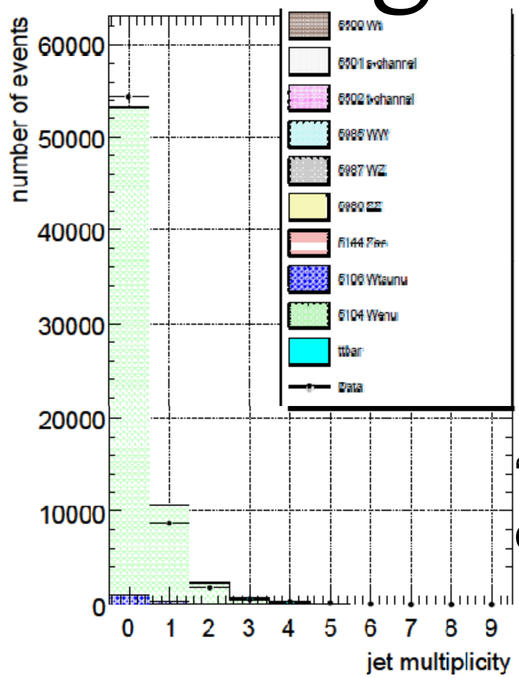
Normalized MC shape

Fit over 5 jet bins

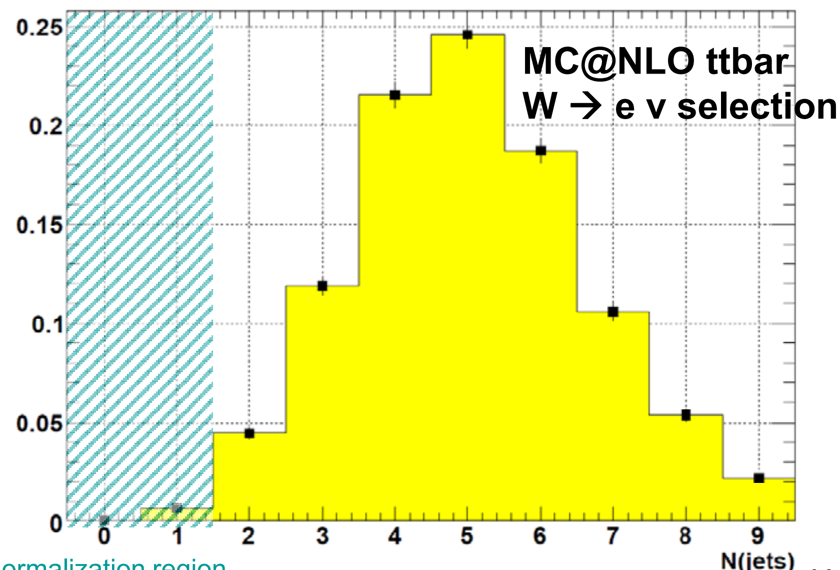
Normalized, data-corrected MC shape



Background treatment in N_{jet} fit

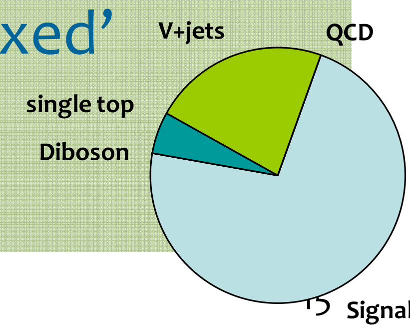


"Data:" black box of mixed MC



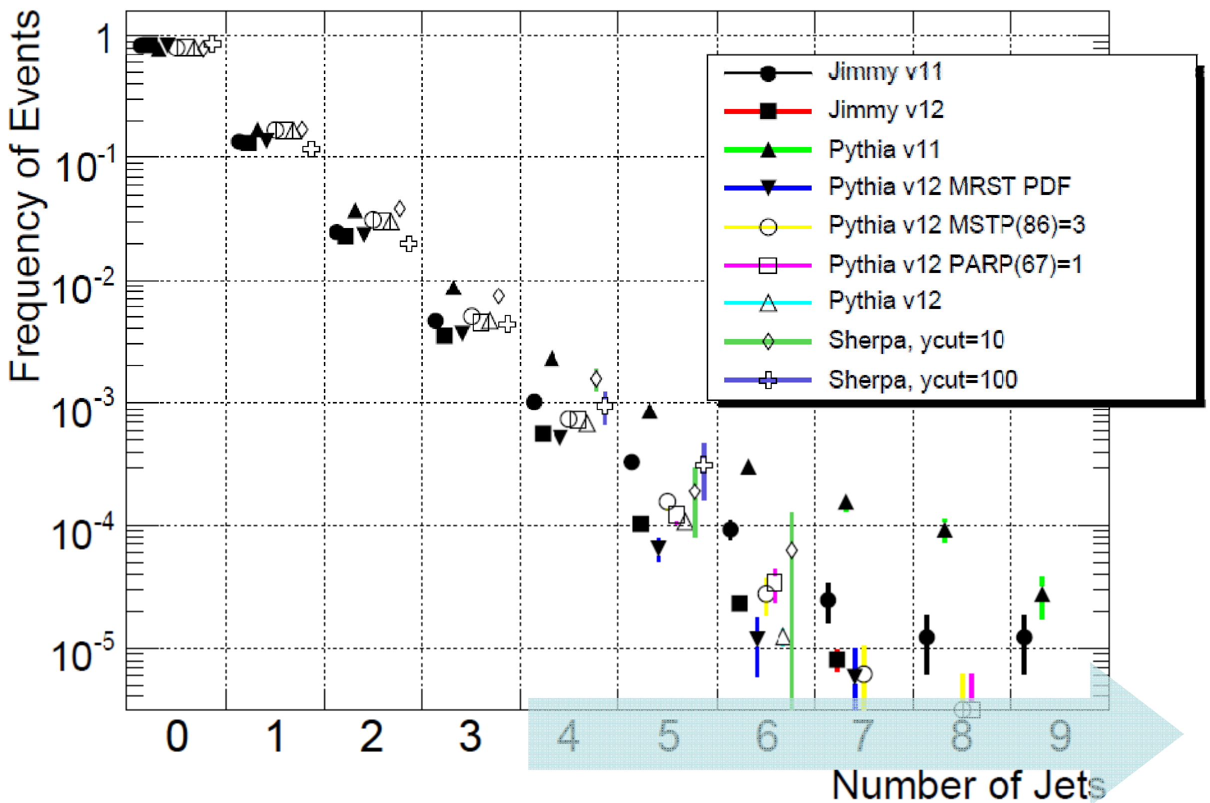
W normalization region

- We need more info to understand W+j backgrounds → look at N_{jets}
  - ➤ especially in first data: mass fit can be unlucky, so study is designed for ~16 pb^{-1}
- **Fit** simultaneously to top and W+j shapes
  - ➤ others are 'fixed'



Background uncertainty in N_{jet} fit

- apply baseline analysis cuts (lepton and missing E_T) to various generators (*Ian's principle: available, validated…*)
- compare particle-level N(jet) shapes:
  - Resulting BG uncertainty: 50 - 100% in signal region from 1-2 jet normalization method
  - cross section systematic: → 25-50% systematic from background

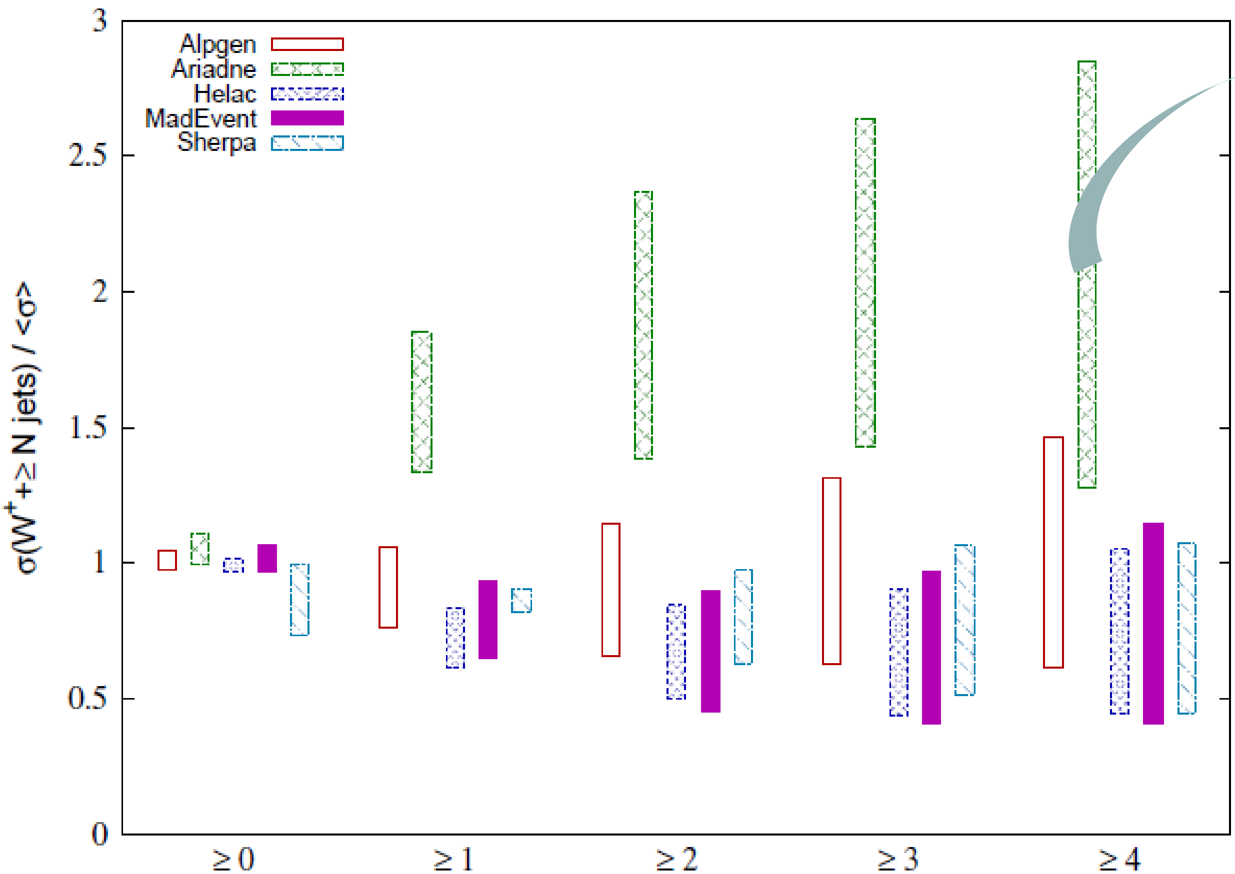


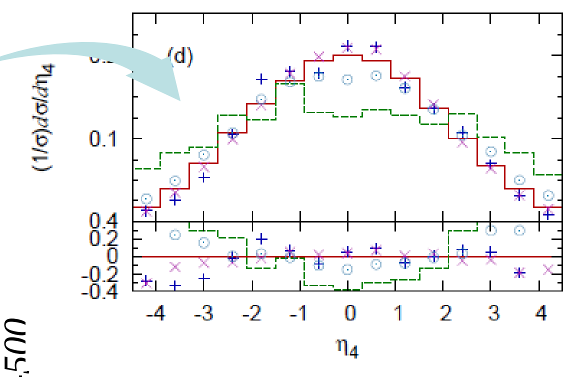
Generators:
- Herwig 6.510
- Pythia v11= 6.323, v12 = 6.403
- Sherpa 1.09

Default PDF: CTEQ6L (compare MRST2004)

Background uncertainty in N_{jet} fit

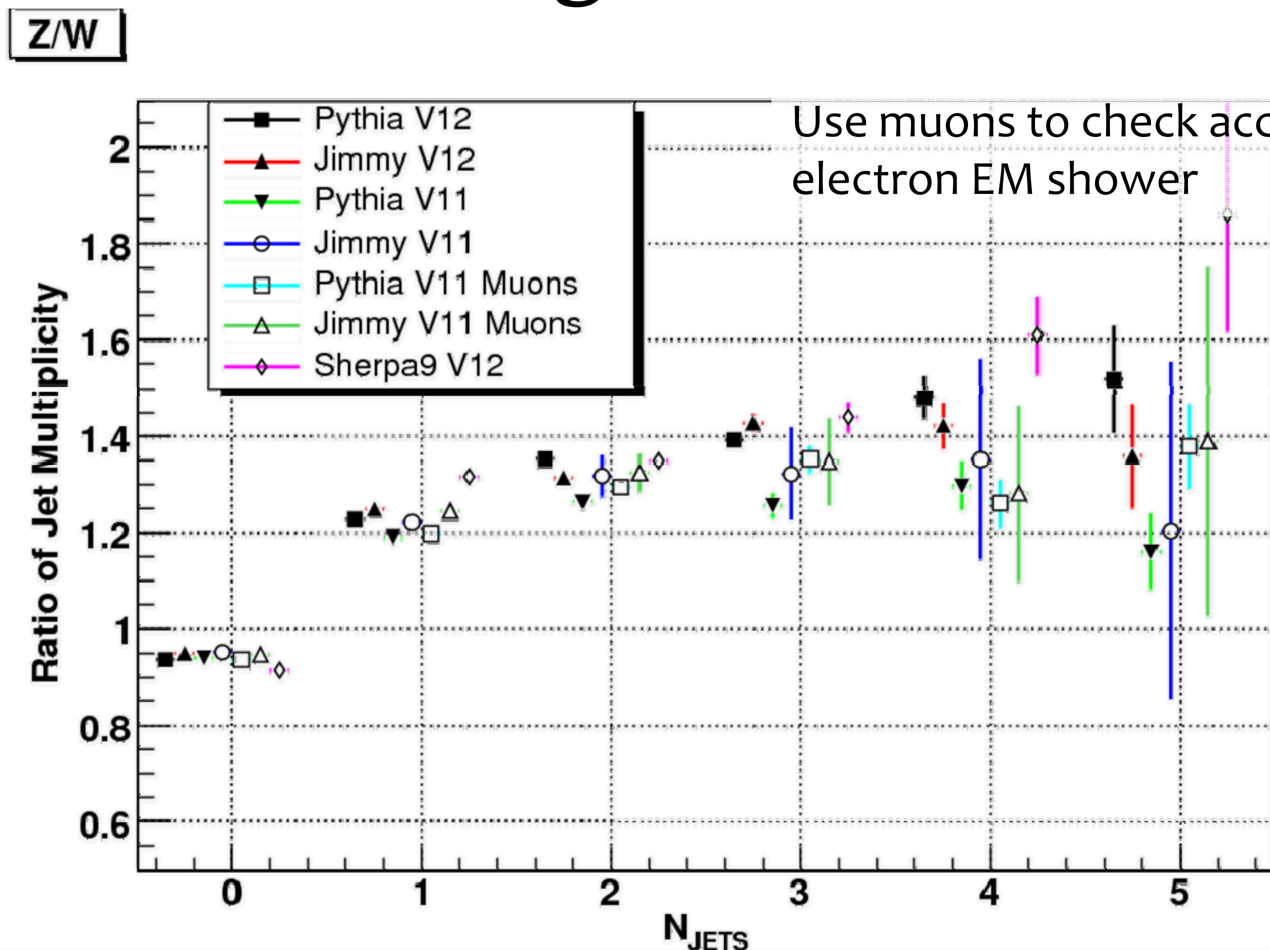
- Is agreement better with **matched** generators?
  - look ~~at~~ up inclusive W+jets cross sections: each of 5 generators compared to mean (σ) for jet E_T> 20 GeV
  - still large uncertainties →look for a shape calibration





Eur. Phys. J. C53 (2008) 473-500

Jet counting calibration with Z+jets



Use muons to check acceptance bias from electron EM shower

- Ratio of (normalized) Z and W distributions
  - use same *trigger path*, lepton requirement, *jet definition*
  - require a second lepton (same cuts) and $70 < m_Z < 110$

Jet counting calibration with Z+jets

- Predictions for W/Z ratio are more uniform: the residual uncertainty is ~15%

$$\left.\frac{W(\geq 4)}{W(\leq 1)}\right|_{\rm data} = \frac{[Z(\geq 4)/Z(\leq 1)]_{\rm data}}{[Z(\geq 4)/Z(\leq 1)]_{\rm MC}} \left.\frac{W(\geq 4)}{W(\leq 1)}\right|_{\rm MC}$$

  - Also use Z+4 jets to calculate 'missed' Z +jets: 2 ± 0.6% of signal region

- But… this method adds a 25% error (in 16 pb^{-1}) from the number of reconstructed Z bosons

Jet counting homework

- emphasis was exploring small dataset, trying in-situ calibrations on an intentionally miscalibrated sample:
  - trigger and lepton ID efficiencies, energy scale calibrations…
- Now: start developing a clear picture of lepton+jets channel
  - optimising cuts for (somewhat) higher luminosity
  - Comparing matched ME generators for background
    - kinematics, including hadronic top mass
  - Improve W+jets calibration procedure
    - check for possible instrumental effects varying across N(jet)
    - understand the limiting theoretical systematic

b-tag efficiency measurement

- ATLAS ongoing studies: “global efficiency” likelihood fit to N(tagged)

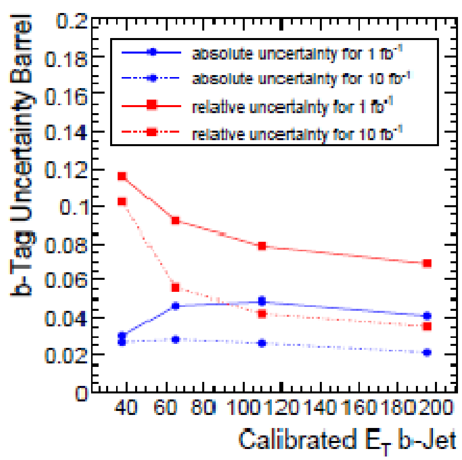
$$< N_n >= (L \cdot \sigma_{t\bar{t}} \cdot A_{pre-tag}) \cdot \sum_{i,j,k} F_{ijk} \sum_{i'j'k'} A_i^{i'} \cdot \varepsilon_b^{i'} \cdot (1-\varepsilon_b)^{i-i'} \cdot A_j^{j'} \cdot \varepsilon_c^{j'} \cdot (1-\varepsilon_c)^{j-j'} \cdot A_k^{k'} \cdot \varepsilon_l^{k'} \cdot (1-\varepsilon_l)^{k-k'}$$

  - data points: N(0,1,2,3…) tags → measure charm and top cross section
    - almost as many bbc (+X) as bbq (+X) events
  - but, no background measurement
- b-tag efficiency is insensitive to W+jets:
  - ~1% from all BG, (~2% ISR : combinatorics again)
    - top cross section picks up ~5% uncertainty

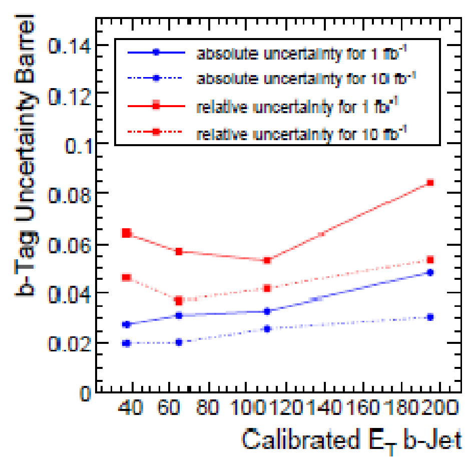
b-tag efficiency measurement

- CMS: "differential efficiency:" kinematic fit (requiring m_t,m_W) ➔ ~5% error in 1fb^{-1}
  - again: neglect W+jet: after tagging one jet, ~3% remaining W contamination
    - combinatoric errors dominate
  - dilepton measurement (pretag sample, no m(Z) cut): assumed 20% uncertainty on Z+jets:

L+jets

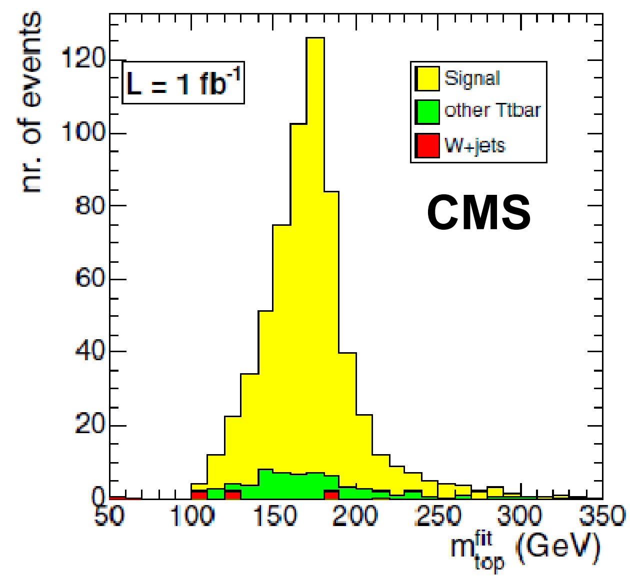


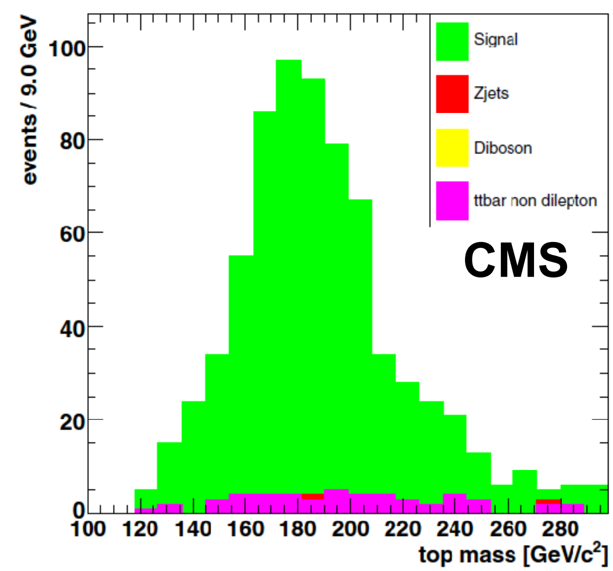
Dilepton



Mass measurement

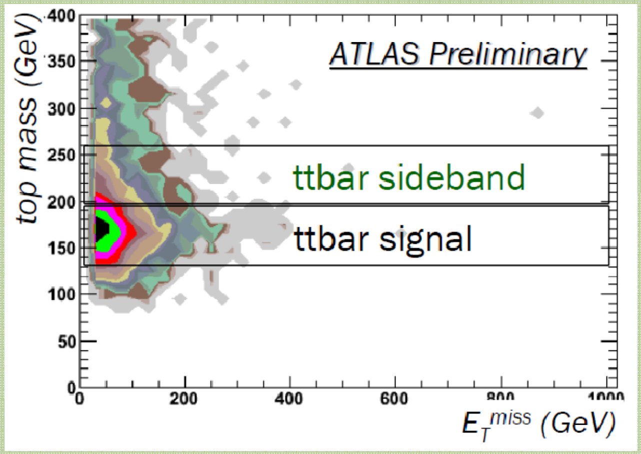
- upper: lepton +jets
- lower: dilepton
  - kinematic fits
  - tight tags *and* tag vetoes





Summary and outlook

- V+jets is no longer our biggest worry for cross sections or top properties, but…
- … historically, understanding top and W+jets together allows new measurements
- Looking at a loose top sample:
  - using reconstructed Z+jets will help separate W+jets contribution
  - mass and other kinematics can then pick out top
    - essential for understanding SUSY
- Priorities: explore systematics of any kind in W/Z calibration
- due diligence investigating W+HF in some "clean" channels



next talk →