W-bosons and b's

Pierluigi Catastini

Harvard University

Queen Mary PPRC Seminar

15-Mar-2013

Motivation

Aim:

- Verify rather sophisticated W+b-jets theoretical predictions.
- Understand W+b-jets with the same precision as V+jets.
- Understand discrepancies previously seen between measurements and theoretical predictions.

W+b production is irreducible background in many key analyses

- Higgs associated production (VH) with H-->bb
- Single top
- New heavy particle searches (ex: light stop)

Normalization estimated using ad-hoc methods

Kinematic shapes based on MC simulations, where predictions differ

Comparing LHC data with most recent NLO calculations

4-flavor scheme (*u, d, c, s*) calculation paired to parton shower (Powheg, MC@NLO)

Combined 4- and 5-flavor (*u, d, c, s, b*) scheme calculation (MCFM)

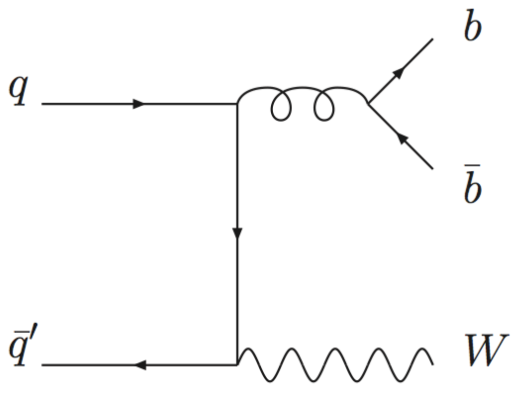
Production Mechanism(s)

Measurements of W+b-jets include three production categories

1. *b* in final state:

matrix element (LO, NLO),

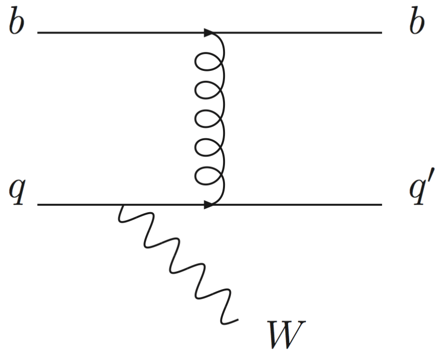
gluon splitting in parton shower.

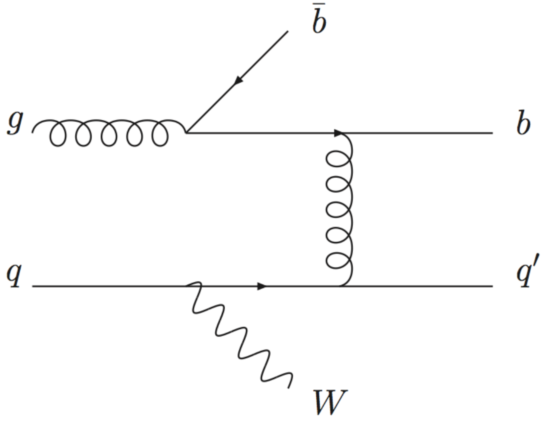


2. *b* in initial state:

5-flavor scheme (LO, NLO),

gluon splitting in initial state (NLO).



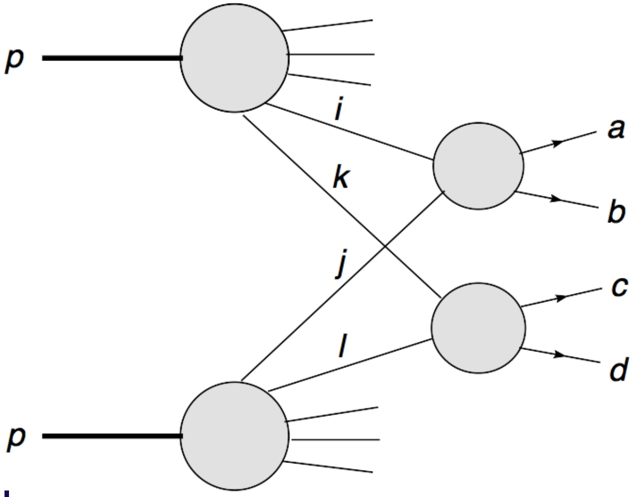


Production Mechanism(s)

3. Double-parton-interactions (DPI):

V and *b* produced by different parton-parton interactions in the same pp collision

- Negligible at the Tevatron
- Very small (3%) in Z+b-jets analysis at LHC
- Large (30%) in W+b-jets analysis (1-jet, low p_T) at LHC



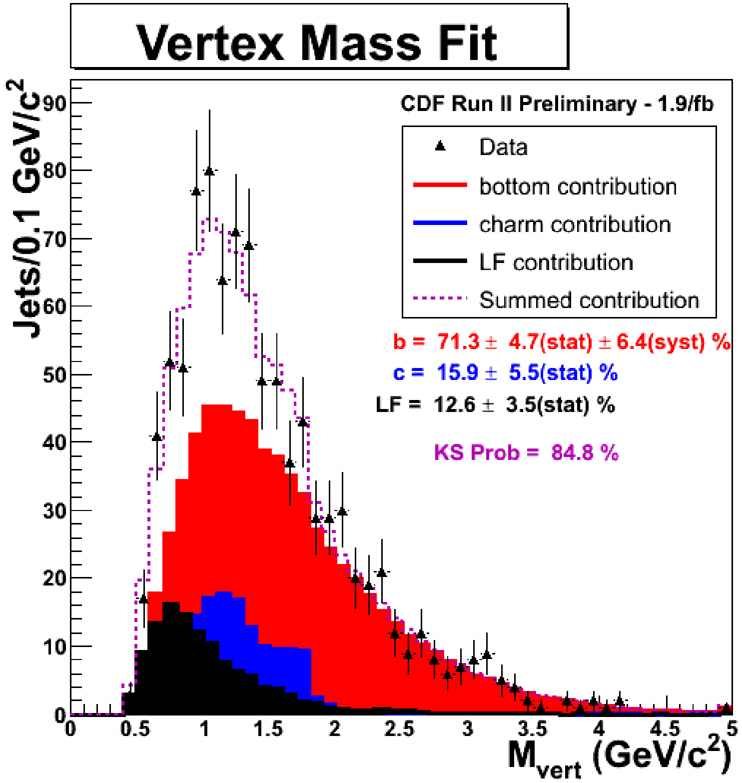
Large theoretical uncertainty:

- Z+b-jets analysis: +/- 100%
- W+b-jets analysis: $^{+39\%}_{-28\%}$ from ATLAS measurement of σ_{eff} in W+2-jets events
- In ATLAS W+b-jet analysis, DPI uncertainty is comparable to scale uncertainty

A little bit of History

First Measurement from CDF

- p-pbar collisions at sqrt(s)=1.96 TeV
- One or two jets, reconstructed with a cone algorithm with R=0.4
- jet ET > 20 GeV and |η|< 2.0
- Events with at least one b-tagged (ultra- tight secondary vertex requirements)
- Use vertex mass to discriminate between b, c and light jets.



PRL 104, 131801 (2010)

$$\sigma_{W+b} \times Br(W \to l\nu)$$

$$2.74 \pm 0.27 \pm 0.42\ pb$$

$$ALPGEN = 0.78\ pb$$

$$NLO\ pQCD = 1.22 \pm 0.14\ pb$$ (MCFM)

Measurement significantly higher than NLO predition: 2.8σ

Then ATLAS with 35 pb-1

- p-p collisions sqrt(s) = 7 TeV.

- The SV0 b-tagging algorithm is based on requiring a displaced secondary vertex reconstructed within a jet with a decay length significance > 5.85.

- SV0 mass used to separate b-jets from c- and light-jets on a statistical basis.

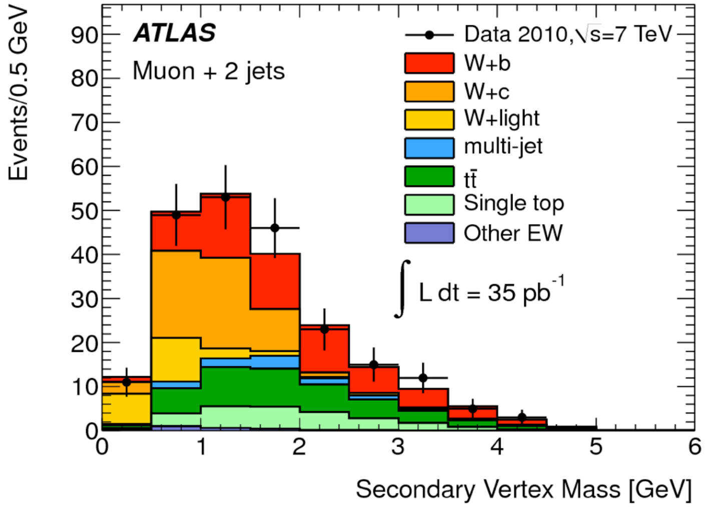
  - 1 b-tagged jet
  - 1 or 2 jet
  - Fit each jet bin separately for e and μ

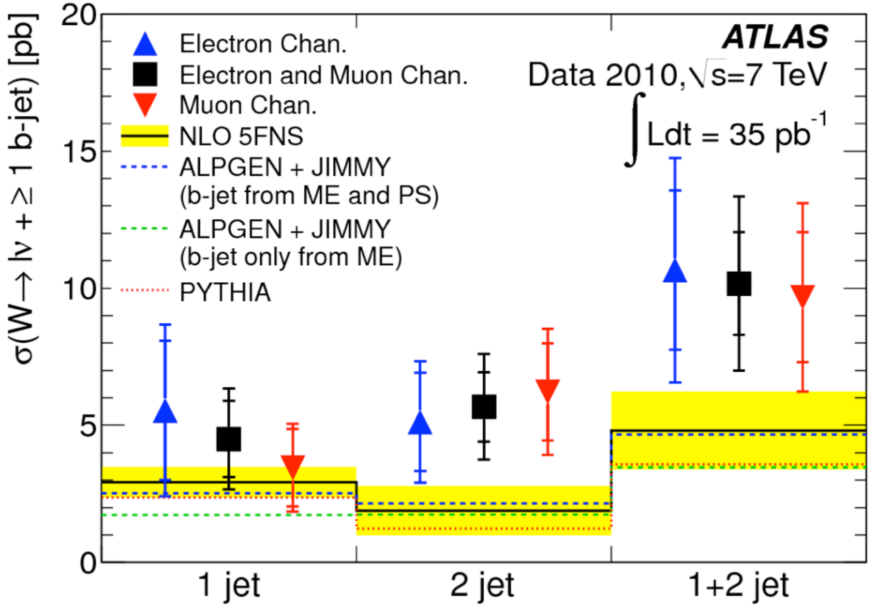
NLO prediction obtained in the 5 flavor number scheme [F. Caola *et al.* arXiv:1107.3714]

NLO agrees within 1.5 sigma with the measurements

Still some tension with predictions.

Phys.Lett. B707 (2012) 418-437





More recently D0

- p-pbar collisions at sqrt(s)=1.96 TeV
- At least 1 b-jet
- Jet PT > 20 GeV and |η|< 2.0
- Use vertex mass to discriminate between b, c and light jets.

Measured (combining the muon and electron W decay channels)

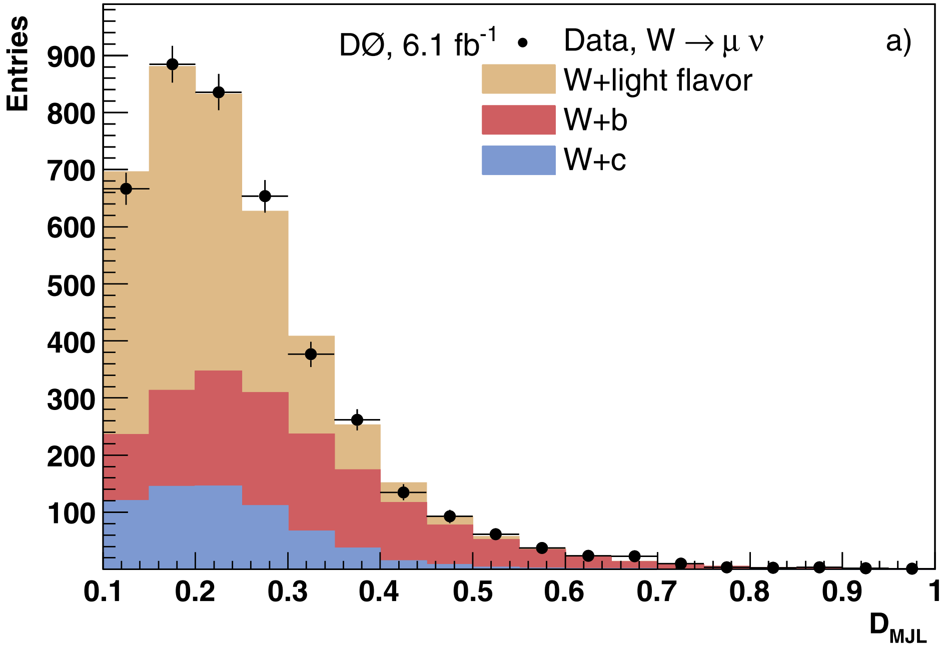
$$\sigma(W+b)\cdot\mathcal{B}(W\to\ell\nu)$$

$$= 1.05 \pm 0.03\,(\text{stat.}) \pm 0.12\,(\text{syst.})\ \text{pb.}$$

Predicted

$$1.34^{+0.40}_{-0.33}\ (\text{scale}) \pm 0.06\ (\text{PDF})^{+0.09}_{-0.05}\ (m_b)\ \text{pb.}$$

Phys.Lett. B718 (2013) 1314

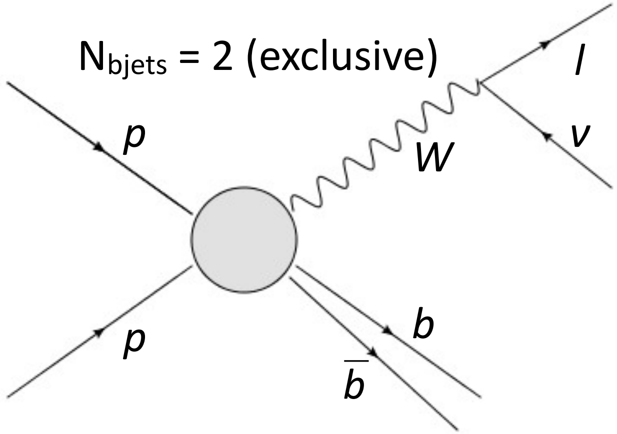


Central value low, but perfectly compatible with prediction within uncertainties.

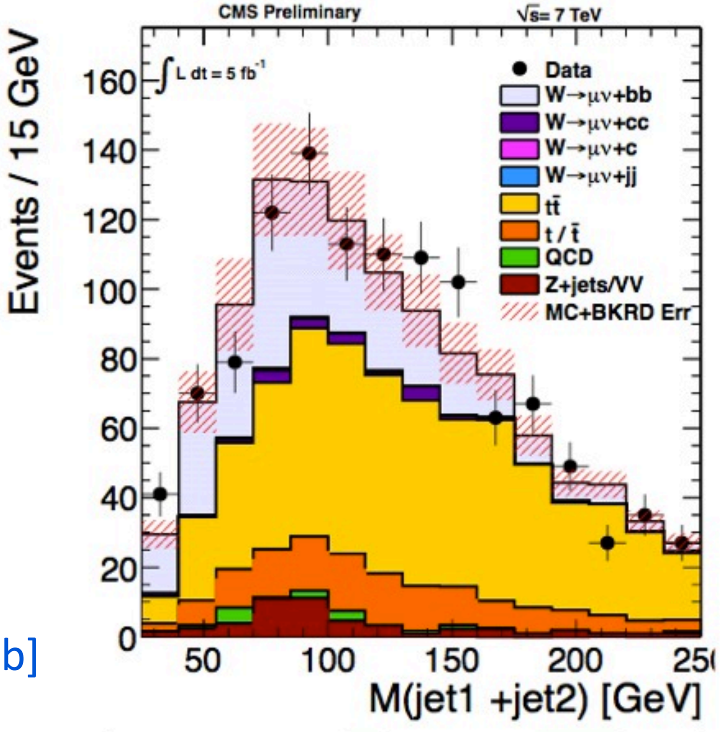
Out of the press: CMS Wbb

- Slide “stolen” from Moriond QCD talk by Konstantinos Theofilatos.
- No additional information available for now.
- Double tagged W+bb cross section measurement

  The two b-jets are well separated in DR and both are b-tagged.



$N_{bjets} = 2$ (exclusive)



	σ(W→bb)
Data (CMS)	0.53 ± 0.05 stat ± 0.1 sys [pb]
MCFM (MSTW08NNLO)	0.52 ± 0.03 [pb]

Complementary phase space w.r.t. the W + 1 b-jet measurement, very good Data/Theory agreement; Important for WH with H→bb

Few comments

- Four results from four experiments and two colliders.

  - CDF significantly higher than prediction. Not confirmed by D0 (same collider, similar fiducial region).
  - Some tension at ATLAS in specific fiducial regions.
  - Perfect agreement at CMS.

- Are these results sensitive to the same W+b production processes ? NO

  - Tevatron is p-pbar at sqrt(s) = 1.96 TeV
    - Quark initiated processes more important than gluon initiate ones.
    - DPI contribution negligible

  - LHC is p-p at sqrt(s) = 8 TeV
    - Gluon initiated processes very important.
    - ATLAS measurement quite exclusive and very sensitive to single b and collinear final state b-quarks.
    - CMS measurement is only sensitive to well separated b-jets
    - Depending on the selection, DPI is very important.

New W+b-jets results from ATLAS with 4.6 fb^{-1}

arXiv:1302.2929, Submitted to JHEP

Analysis Strategy

✦ What we measure:

- W+b-jets cross section, inclusive and differential
  - Exactly 1 b-tagged jet (unfold to 1 or 2 b-jets)
  - Differential: Number of additional jets (0 or 1)
  - Differential: p_T of b-jet
  - DPI contribution is NOT subtracted.

✦ How we measure it:

- 1. Select W+jets, require 1 b-tagged jet
- 2. Estimate backgrounds
  - W+c, W+light : use b-tagging distributions
  - QCD, top, single top : use kinematic distributions
- 3. Unfold to fiducial region

W+b-jet Selection

W boson

Isolated lepton (e, μ), $p_T > 25$ GeV

$E_T^{miss} > 25$ GeV

$m_T(W) > 60$ GeV

reduce multijet background

b-tag

reduce top background

Exactly 1 b-tag

“CombNN” neural network combines two techniques:

1. *b*- and *c*-vertex reconstruction along common line of flight,
2. simple likelihood based on impact parameter significance of all tracks.

High purity working point:

- 40-60% for b-jets
- ~10% for c-jets, ~0.1% for light-jets

Jets (Anti-k_T, R = 0.4)

$p_T > 25$ GeV

$|y| < 2.1$

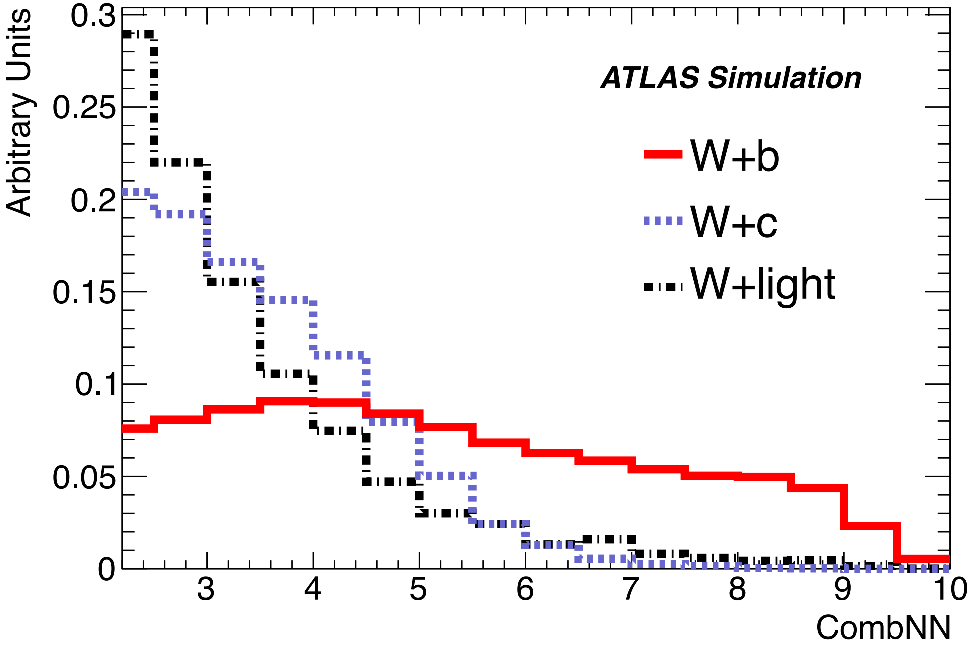
full jet cone within tracking region

NJets < 3

reduce top background

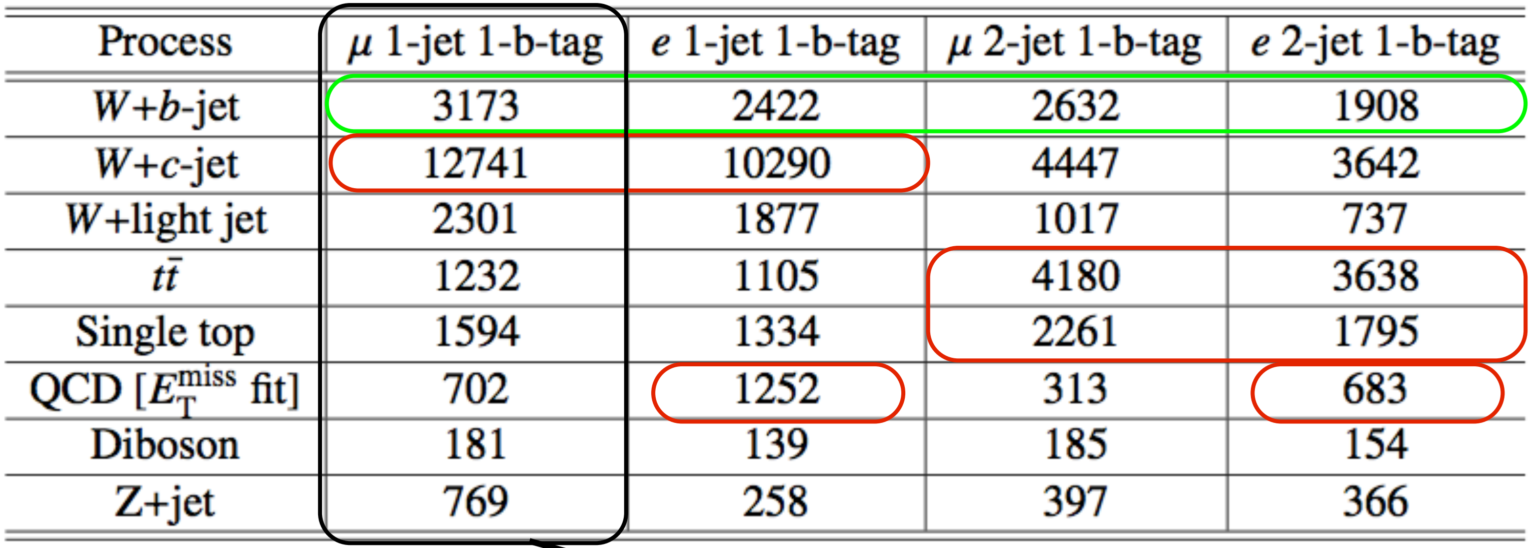
> 75% of tracks from leading collision

reduce pile-up



Sample Composition

4 analysis regions: signal and background after all fits



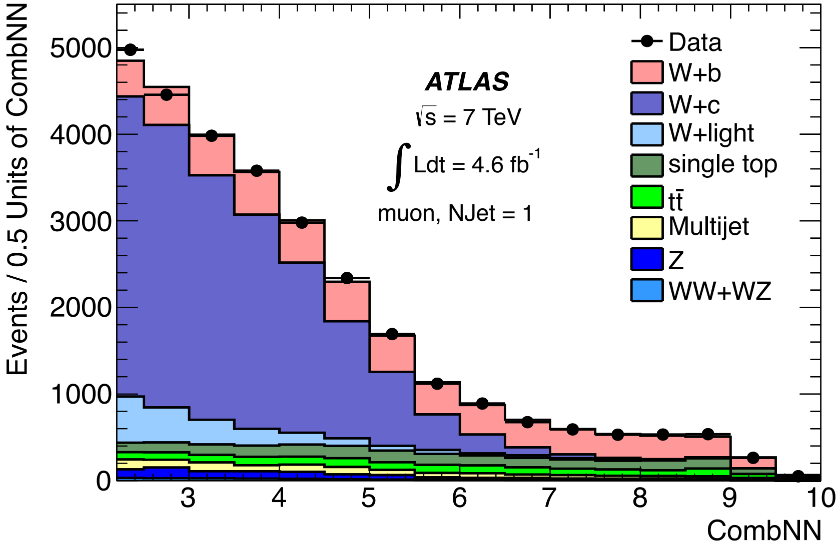
Process	μ 1-jet 1-b-tag	e 1-jet 1-b-tag	μ 2-jet 1-b-tag	e 2-jet 1-b-tag
W+b-jet	3173	2422	2632	1908
W+c-jet	12741	10290	4447	3642
W+light jet	2301	1877	1017	737
$t\bar{t}$	1232	1105	4180	3638
Single top	1594	1334	2261	1795
QCD [E_T^{miss} fit]	702	1252	313	683
Diboson	181	139	185	154
Z+jet	769	258	397	366

Signal is ~20% of sample !

Background composition very different in each analysis region.

Striking:

- W+c-jet in 1-jet region
- Single-top and tt in 2-jet region
- Multijet in electron channel



Estimate Backgrounds

Completely independent estimates in electron and muon channels

W+c, W+light: use b-tagging distributions

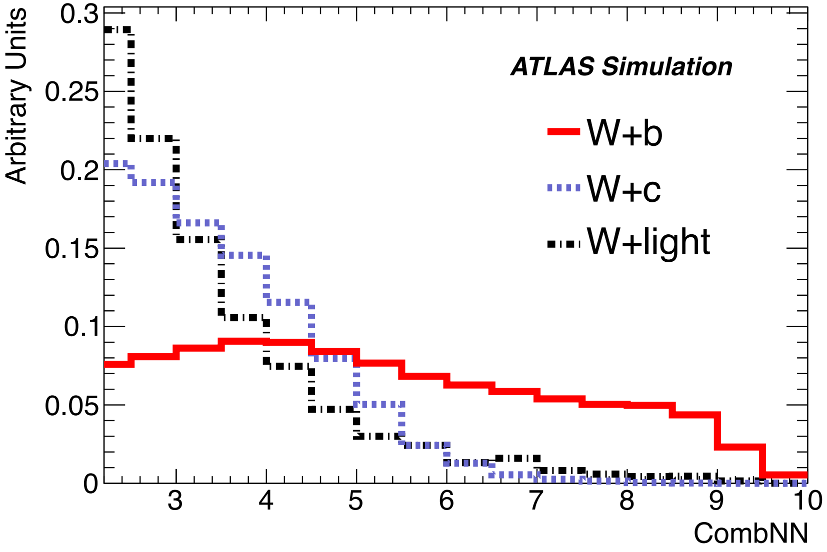
CombNN tagger value is best discriminator btw. b-jets, c-jets and light-jets

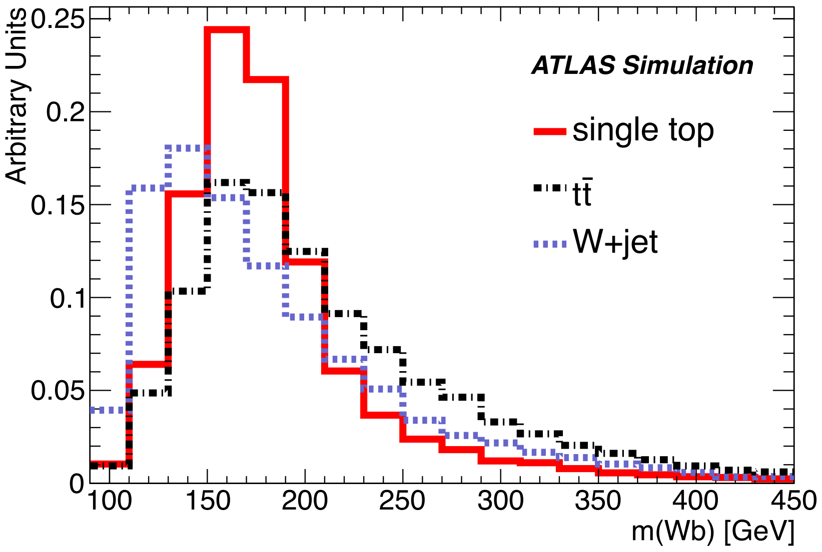
Backgrounds with b-jets: use kinematic properties

Multijet: Missing Transverse Energy [E_T^{miss}]

tt: Number of Jets

Single top: Mass of W-b pair [m_{Wb}]





Estimate Multijet

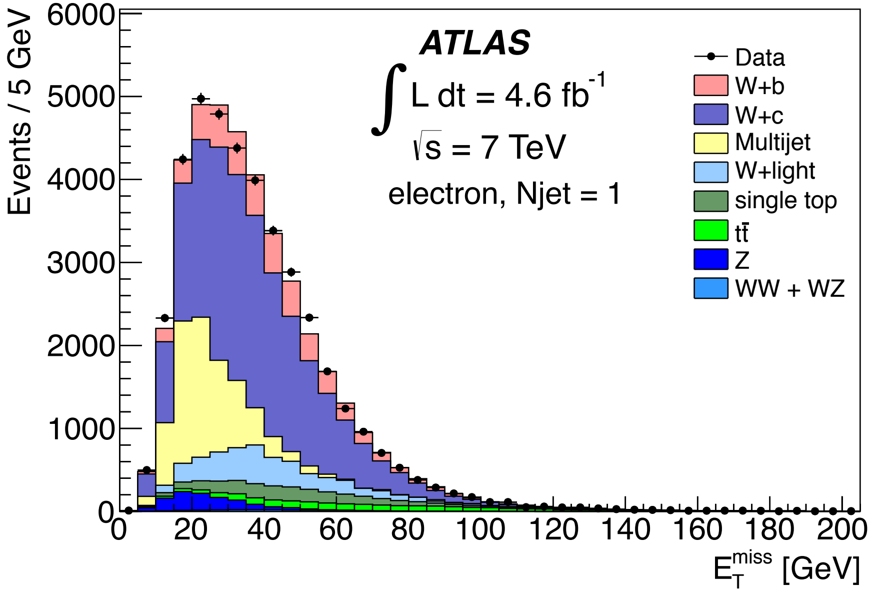
Multijet template shape obtained from data

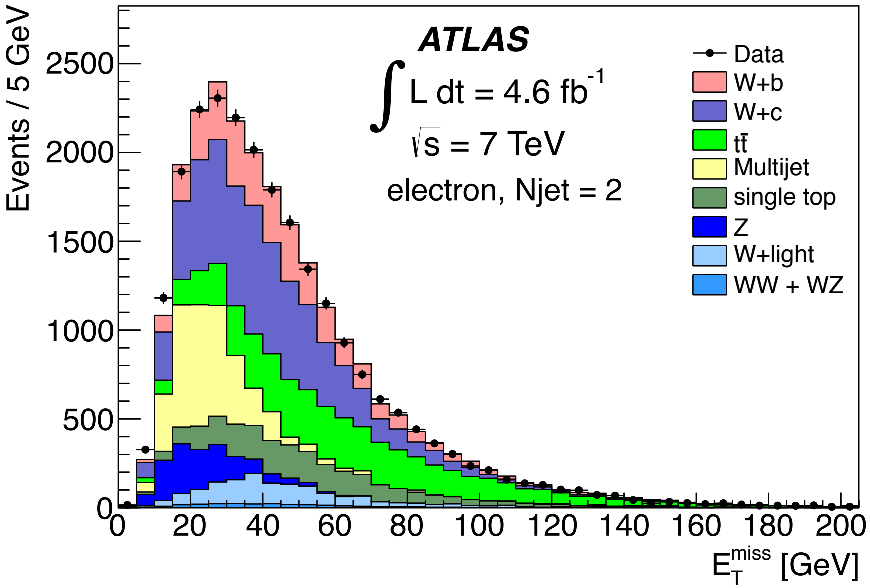
Modified leptonic selection, based on studies in heavy-flavor dijet samples

Multijet normalization obtained from fit to E_T^{miss} distribution

Normalization uncertainty from cross-check fits with $m_T(W)$ and p_T^{lepton} distributions

- Good agreement, leading to 50% uncertainty



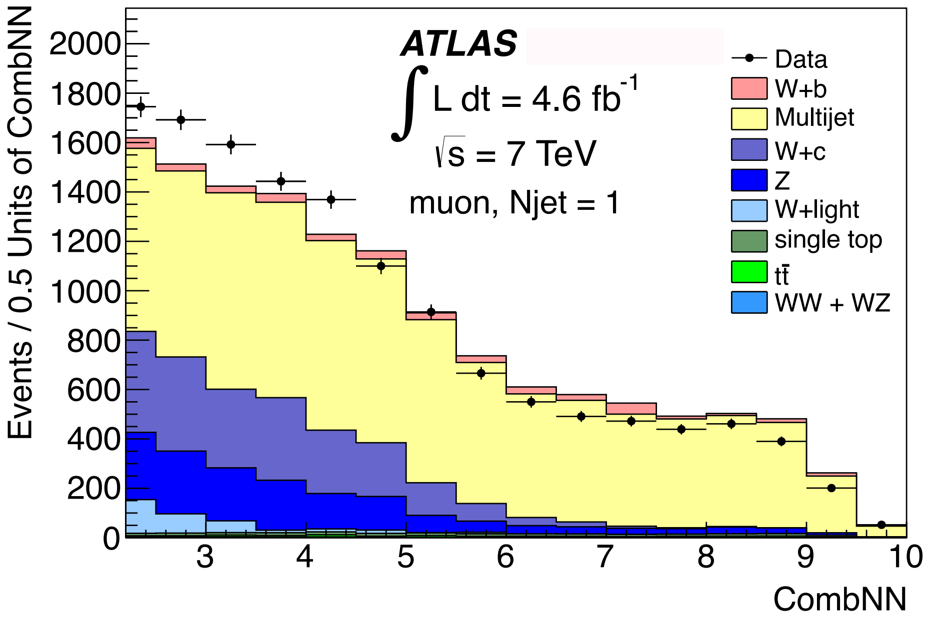


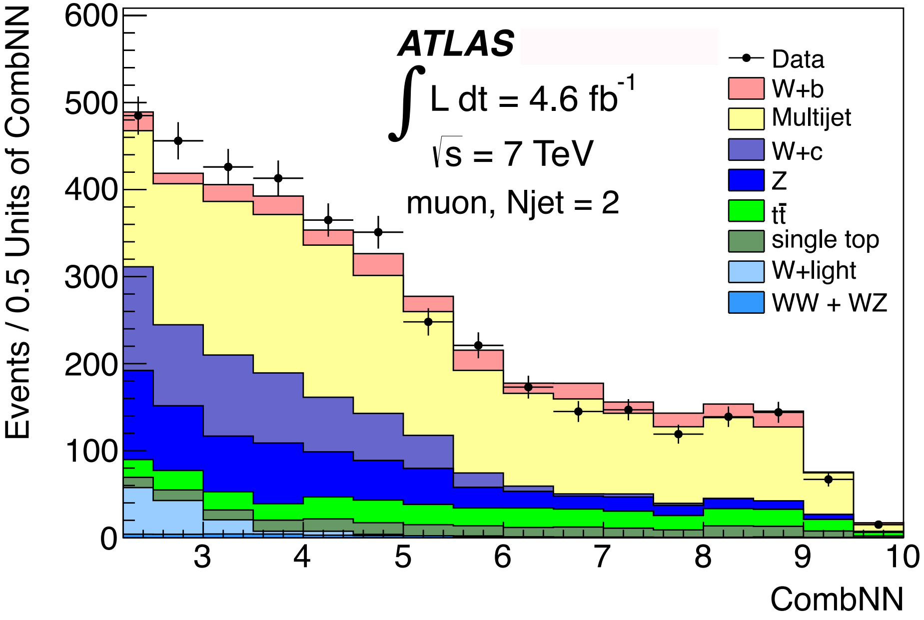
Multijet: shape uncertainty

Multijet template shape obtained from data

Shape uncertainty from control region: low E_T^{miss}, low $m_T(W)$

Correction as a function of the CombNN weight defined by the data/ Multijet model ratio in the control region.





Estimate tt

tt normalization obtained in control region: “At least 4 jets, exactly 1 b-tag”

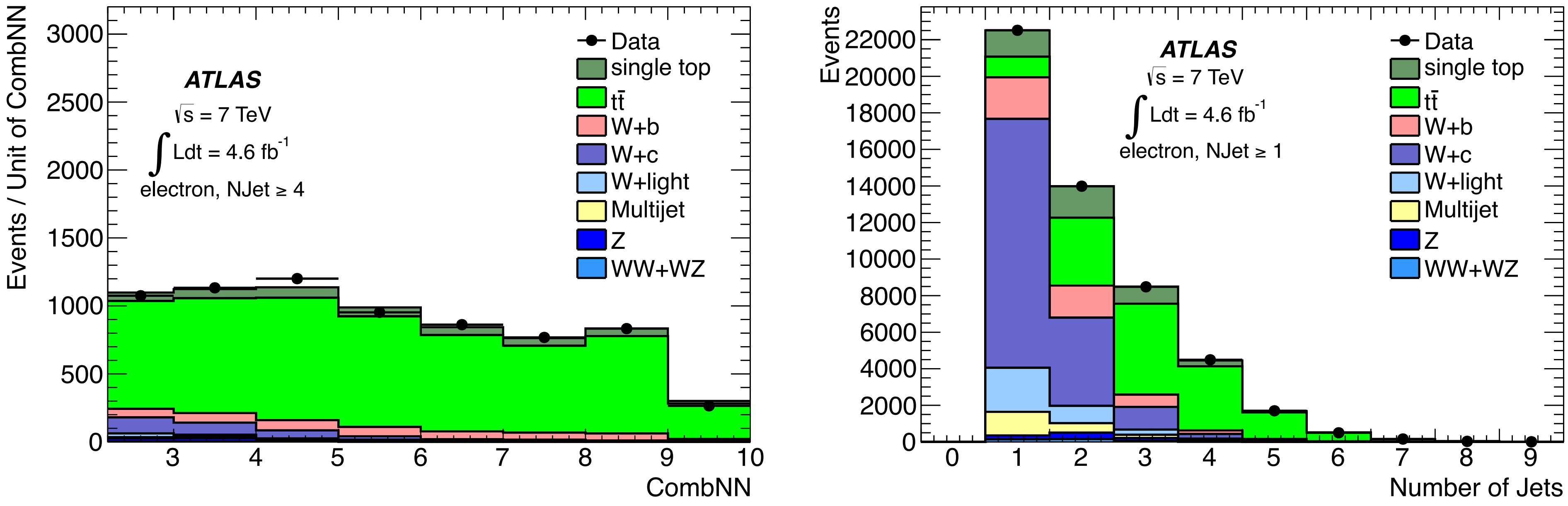
Fitted correction factors ~ 10%

Cross-checks: 3-jets control region, number-of-jets distribution

- Normalization uncertainty: 10%, based on fit results

Apply correction factor to to signal regions (1-jet 1 b-tag, 2-jet 1 b-tag)

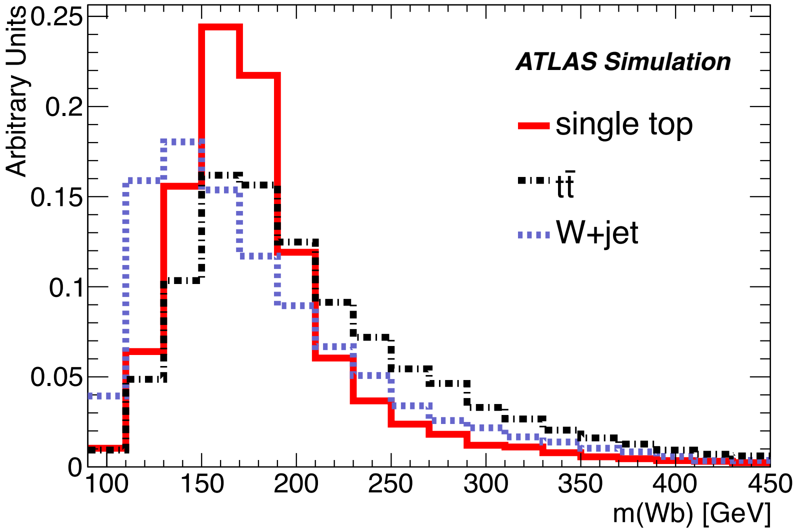
Systematic uncertainties evaluated by rerunning fit with modified Monte Carlo samples



Estimate single-top

No control region: use signal region kinematics

- Invariant mass of (W, b-tagged jet): m_{Wb}
- Cross-check: $H_T = p_T^{lepton} + p_T^{jets} + E_T^{miss}$

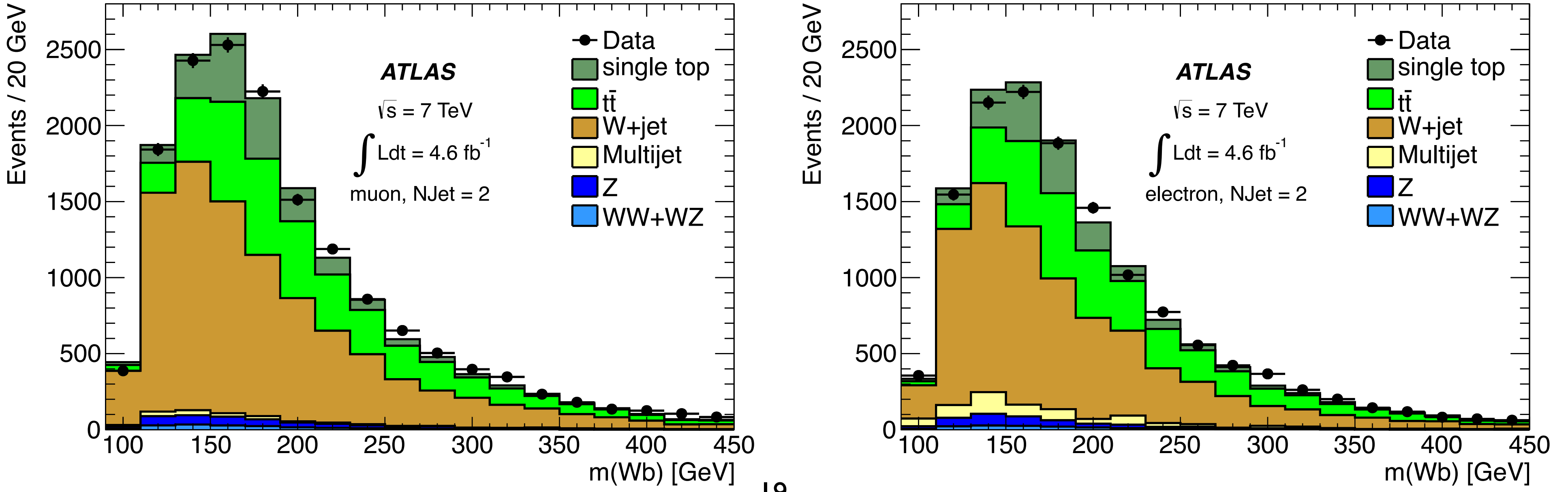


1-jet region: inconsistent picture, no measurement

- H_T vs. m_{Wb} , AcerMC vs MC@NLO
- Use MC normalization with 50% uncertainty

2-jet bin: consistent picture

- Fitted correction factors ~ 15%
- Normalization uncertainty: 20%, based on fit results



Separate b from c/light

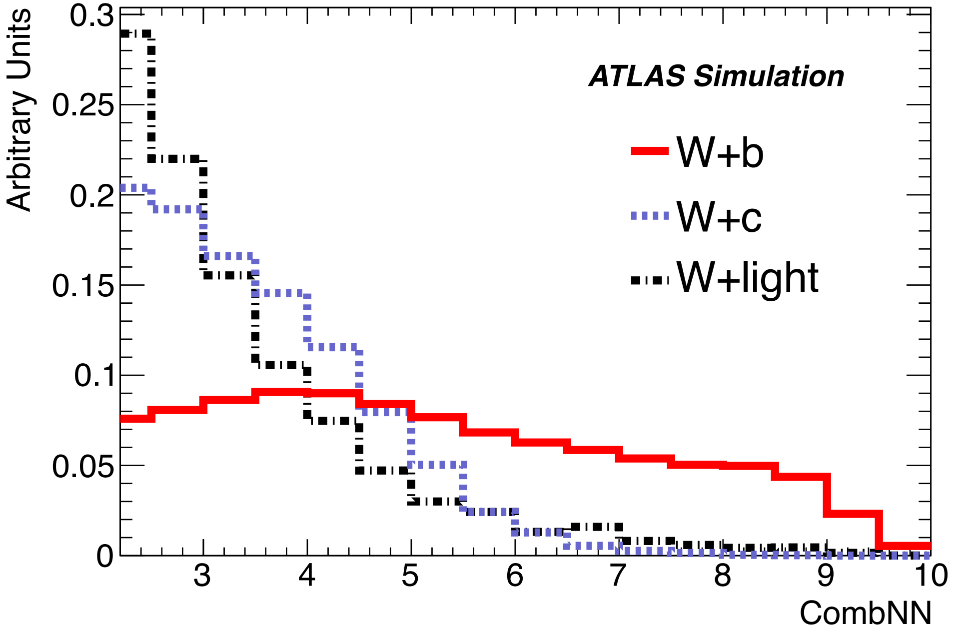
Use b-tagger distribution (CombNN) to separate b, c, light jets

Maximum likelihood fit

W+b, W+c, W+light: normalization completely free

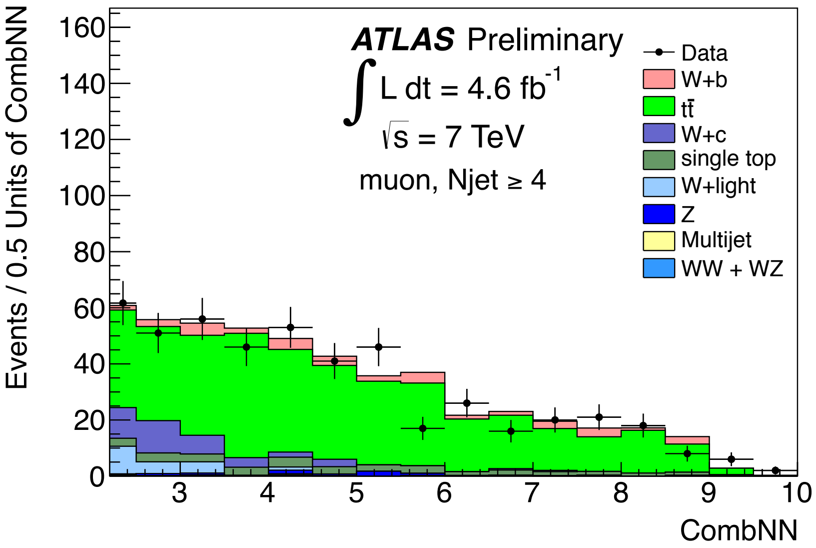
Other bkgs: Gaussian nuisance parameters (based on previous estimates)

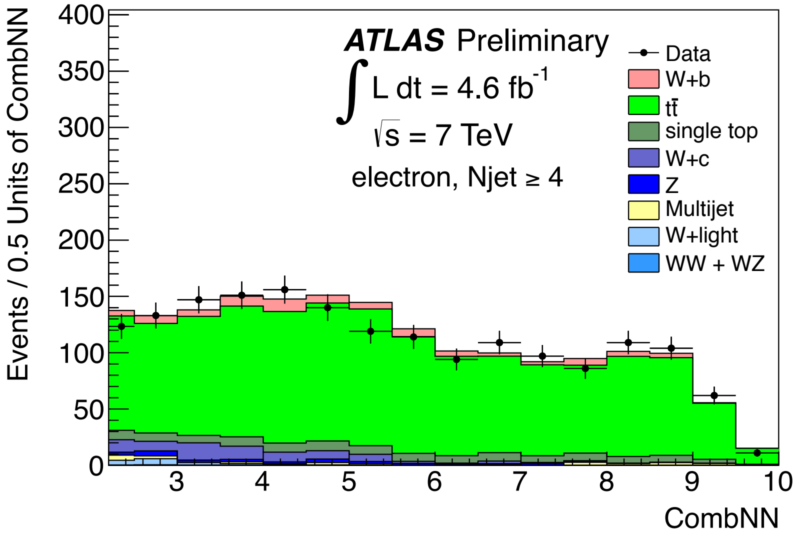
- Multijet (50%), tt (10%), single-top (50% for 1-jet, 20% for 2-jet)



CombNN shapes

- Clean samples of b-jets can be obtained by selecting events with at least 4 jets and 1 b-tag. We checked data/MC shapes agreement considering different jet -pT bins.
- Almost pure samples of b-jet are selected by requiring at least 4 jets and 2 b-tags. The differences in shape between data and MC are used to estimate the b-jets CombNN shape systematics.





- CombNN shape for c-jets: checked using enriched W+c-jet samples (OS soft muon tags). Systematics estimated by producing new shapes after varying the number of tracks in decay vertices.
- CombNN shape systematics for light-jets estimated by comparing templates extracted from Pythia and Herwig programmes.

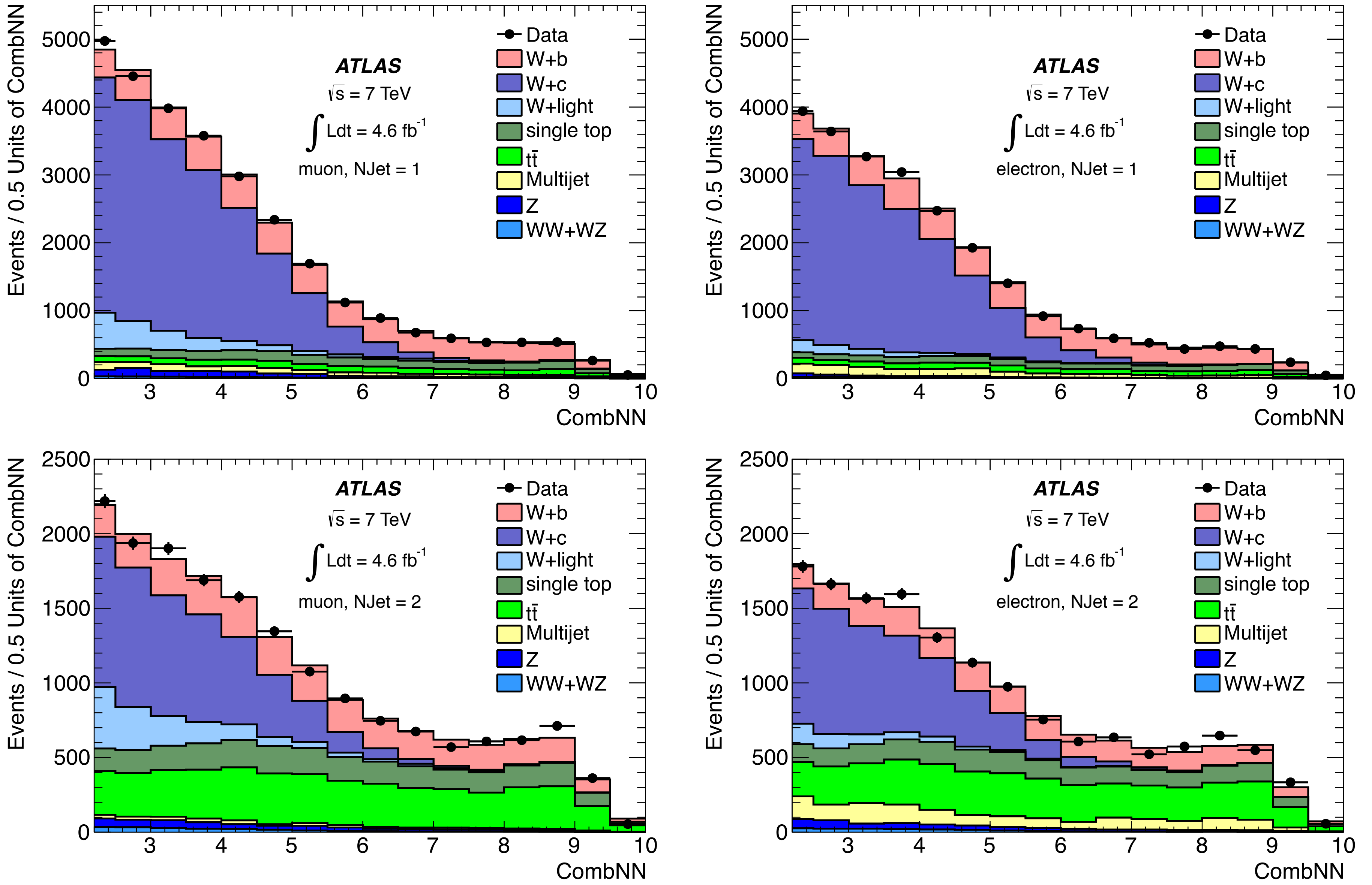
Flavor Content Fit Results

- Correction factors to be applied to the expectation, estimated by the binned ML fit for each process in the four analysis regions.

- The W+b expectation is Alpgen scaled to the W NNLO inclusive cross-section

Process	μ 1-jet	e 1-jet	μ 2-jet	e 2-jet
W+b-jets	1.68 ± 0.14	1.98 ± 0.16	1.14 ± 0.10	1.16 ± 0.13
W+c-jets	1.22 ± 0.04	1.30 ± 0.05	1.04 ± 0.09	1.10 ± 0.10
W+light-jets	0.70 ± 0.22	0.28 ± 0.25	1.15 ± 0.33	0.67 ± 0.44
$t\bar{t}$	1.00 ± 0.10	1.00 ± 0.10	1.02 ± 0.10	1.01 ± 0.10
Single-top	1.07 ± 0.34	1.02 ± 0.36	1.08 ± 0.19	1.01 ± 0.19
Diboson	1.00 ± 0.10	1.00 ± 0.10	1.00 ± 0.10	1.00 ± 0.10
Z+jets	1.00 ± 0.10	1.00 ± 0.10	1.00 ± 0.10	1.00 ± 0.10
Multijet	1.12 ± 0.47	0.80 ± 0.40	0.67 ± 0.49	1.79 ± 0.42

Flavor Content Fit Results



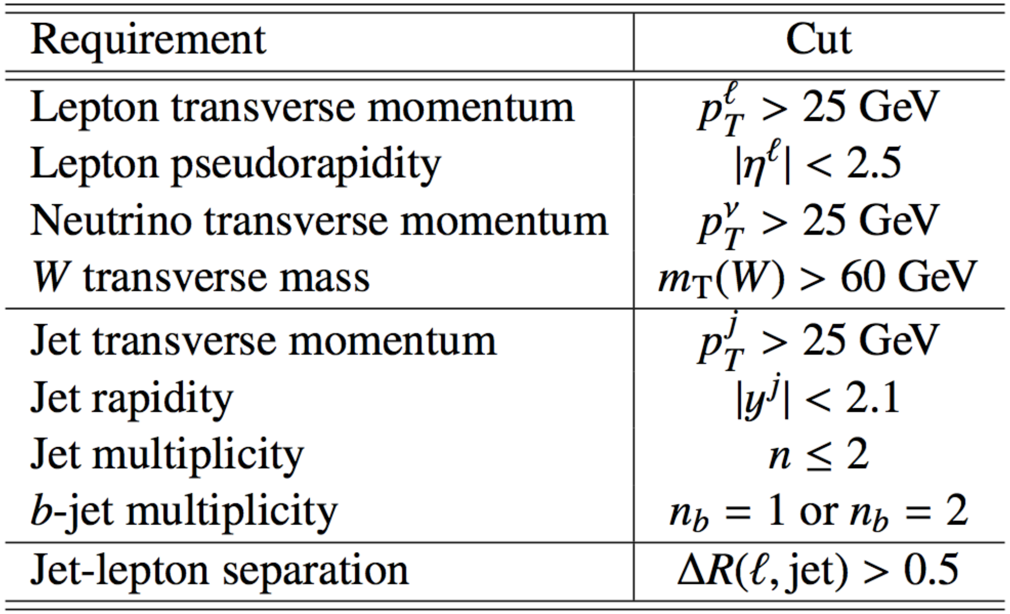
Unfolding to fiducial region

Measurement presented in fiducial region

Kinematic phase space well covered by detector acceptance

Defined at particle level

- Hadron-level jets built from particles with τ > 10ps
- “b-jet” = matched (ΔR < 0.3) with weakly decaying B-hadron with p_T > 5 GeV



Requirement	Cut
Lepton transverse momentum	$p_T^\ell > 25$ GeV
Lepton pseudorapidity	$\lvert\eta^\ell\rvert < 2.5$
Neutrino transverse momentum	$p_T^\nu > 25$ GeV
W transverse mass	$m_\mathrm{T}(W) > 60$ GeV
Jet transverse momentum	$p_T^j > 25$ GeV
Jet rapidity	$\lvert y^j\rvert < 2.1$
Jet multiplicity	$n \leq 2$
b-jet multiplicity	$n_b = 1$ or $n_b = 2$
Jet-lepton separation	$\Delta R(\ell, \mathrm{jet}) > 0.5$

Unfolding performed using Alpgen/Herwig/Jimmy MC

Uncertainties

1. Affecting fit statistical uncertainty

Gaussian nuisance parameters for backrounds

Statistical uncertainty on the W+b-jets yield is increased by a factor of two with respect to all non-W bkgs fixed

2. Affecting fit results (shapes)

All the “template” systematics (b-, c-, light-jet, multijet)

Other systematics also affect shape:

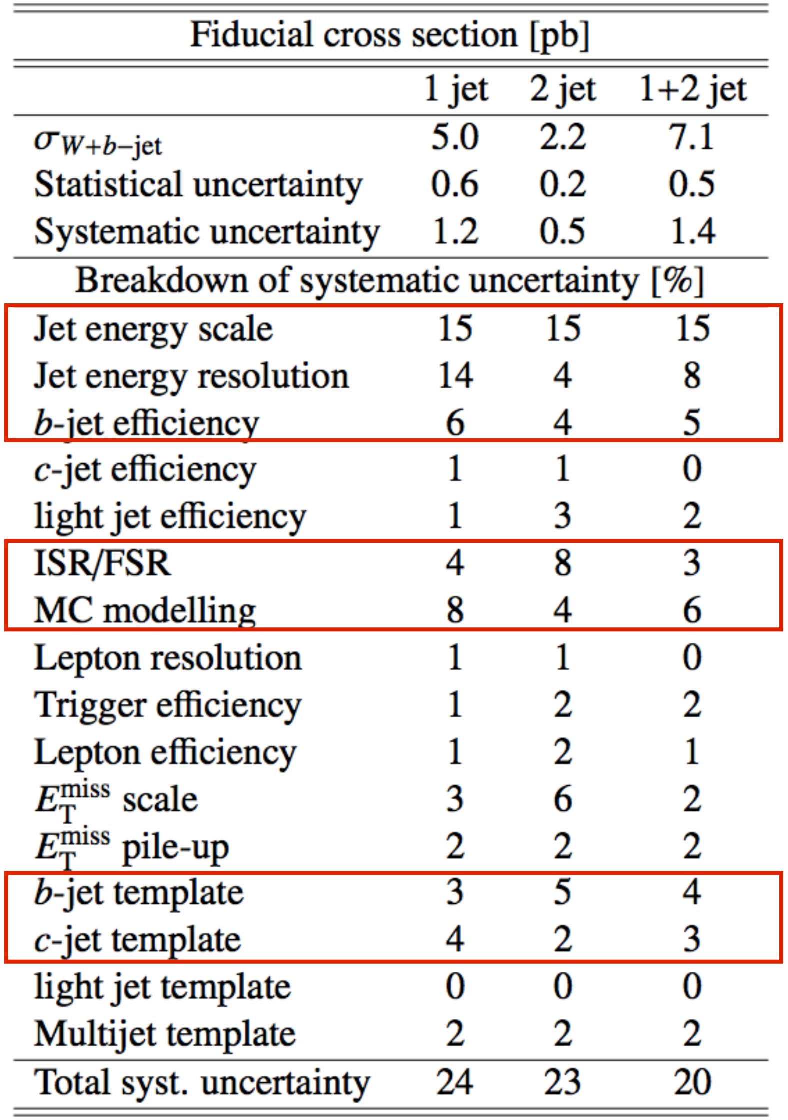
- Jet p_T influenced by scale, resolution, ISR/FSR, MC modeling
- Jet p_T influences b-tagging (CombNN) shapes
- Jet p_T influences number-of-jets distribution for tt normalization

3. Affecting unfolding results (acceptance)

Acceptance mostly affected by jet energy and b-tagging uncertainties

2 and # 3 evaluated by rerunning analysis chain

Latter two are evaluated by rerunning full analysis chain using pseudo-experiments with systematically varied MC



Fiducial cross section [pb]			
	1 jet	2 jet	1+2 jet
$\sigma_{W+b\text{-jet}}$	5.0	2.2	7.1
Statistical uncertainty	0.6	0.2	0.5
Systematic uncertainty	1.2	0.5	1.4
Breakdown of systematic uncertainty [%]			
Jet energy scale	15	15	15
Jet energy resolution	14	4	8
b-jet efficiency	6	4	5
c-jet efficiency	1	1	0
light jet efficiency	1	3	2
ISR/FSR	4	8	3
MC modelling	8	4	6
Lepton resolution	1	1	0
Trigger efficiency	1	2	2
Lepton efficiency	1	2	1
E_T^{miss} scale	3	6	2
E_T^{miss} pile-up	2	2	2
b-jet template	3	5	4
c-jet template	4	2	3
light jet template	0	0	0
Multijet template	2	2	2
Total syst. uncertainty	24	23	20

Comparison with Theory

Comparisons at particle level with 3 sets of predictions

Monte Carlo	**Process**	**Proton model**	**Precision**	**Parton shower, underlying event**	**Double parton interactions (DPI)**
Alpgen/Herwig/Jimmy	W+b/c/light-jets	4 flavors	LO with NNLO k-factor = 1.2	Yes	Yes
Powheg/Pythia	W+bb	4 flavors	NLO	Yes	No
MCFM	W+b+X	4 and 5 f.	NLO	No	No

Corrections: non-perturbative (parton shower, underlying event) **and DPI**

Correction	1 jet	2 jets	
Non-perturbative	0.92 ± 0.02 (had.) ± 0.03 (UE)	0.96 ± 0.05 (had.) ± 0.03 (UE)	← Multiplicative
DPI [pb]	1.02 ± 0.05 (stat) ${}^{+0.40}_{-0.29}$ (syst)	0.32 ± 0.02 (stat) ${}^{+0.12}_{-0.09}$ (syst)	← Additive

Theoretical uncertainties calculated for MCFM

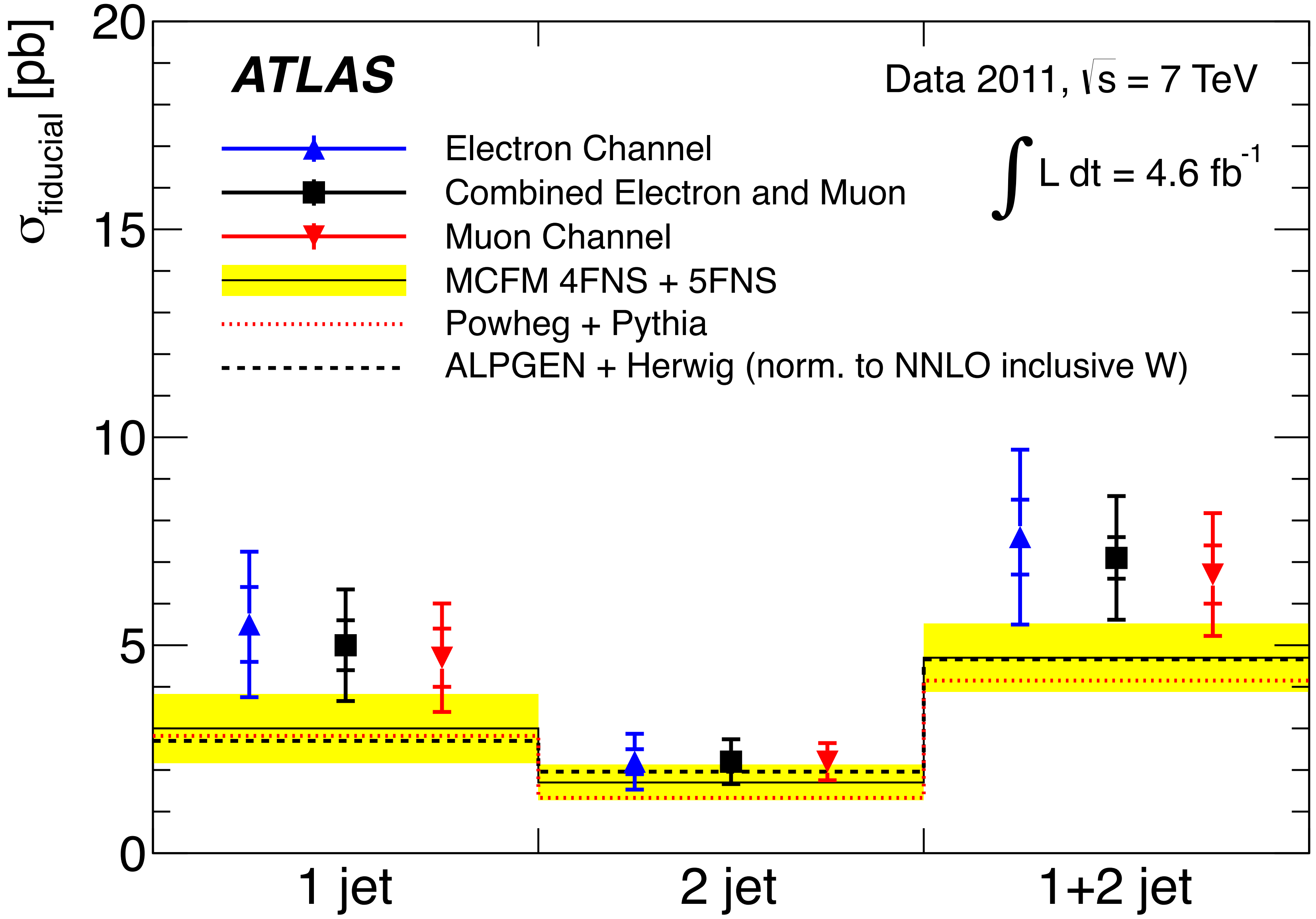
Scale (15-20%)

- Calculated by varying scale from μ/4 to 4μ as done in [arXiv:1107.3714]
- Stewart/Tackmann procedure used to estimate jet-veto uncertainty [arXiv:1107.2117]

DPI (15-20%)

- ATLAS measurement of σ_{eff} in W+2-jets used to assign a DPI uncertainty.

Results: W+b-jet



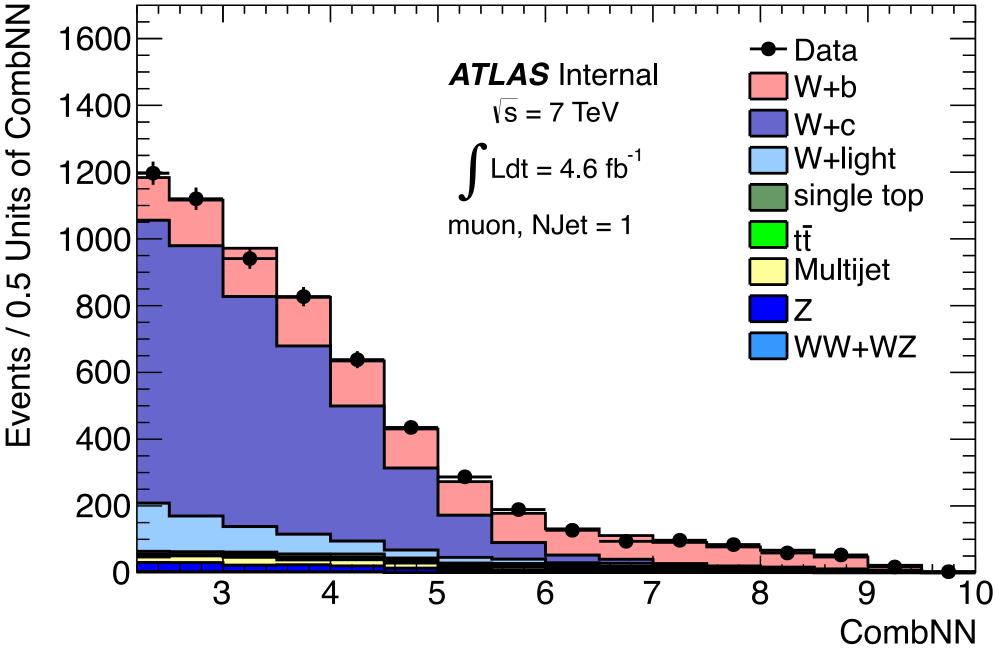
Differential Measurement

Differential: p_T of b-jet (up to 140 GeV)

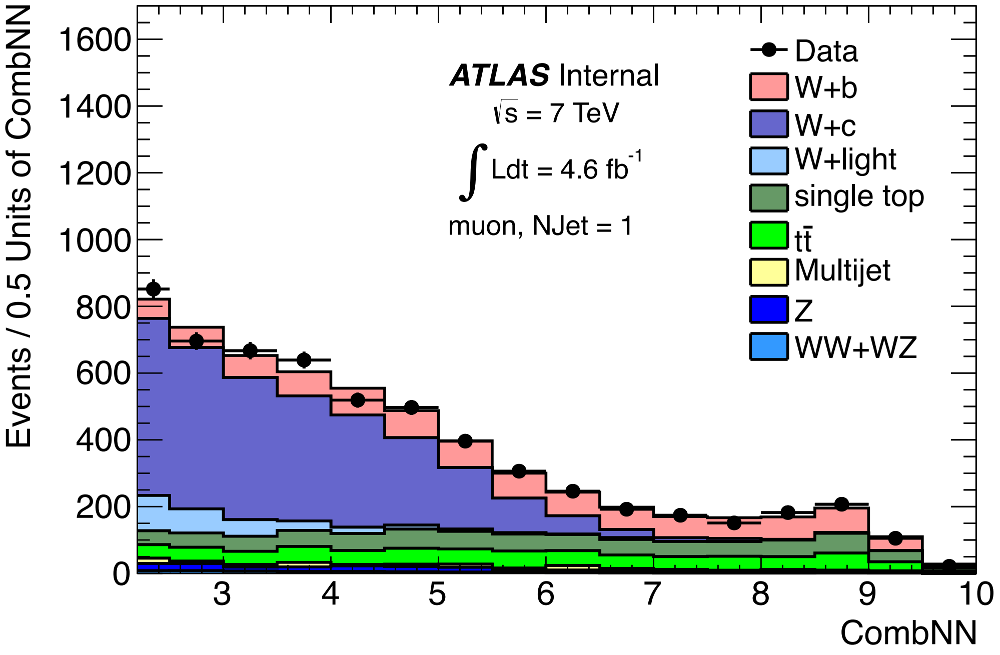
Background composition varies greatly as a function of p_T (st, tt concentrated at high p_T)

Also W+b-jets compositions varies (DPI concentrated at low p_T)

1-Jet; pT in [25, 30] GeV



1-Jet; pT in [60, 140] GeV



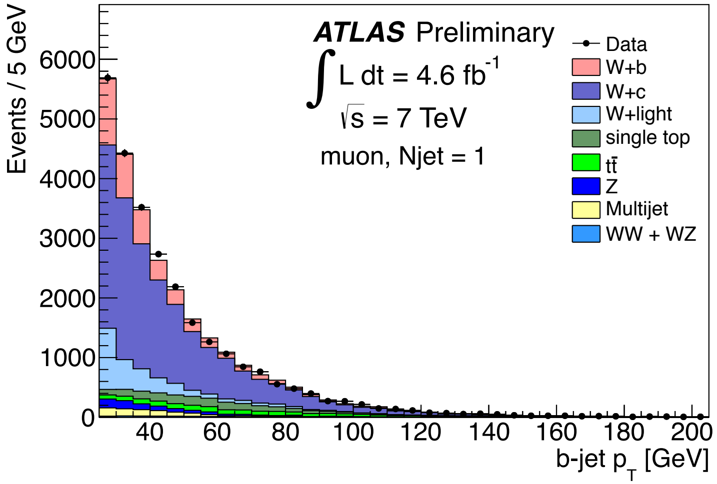
Differential Measurement

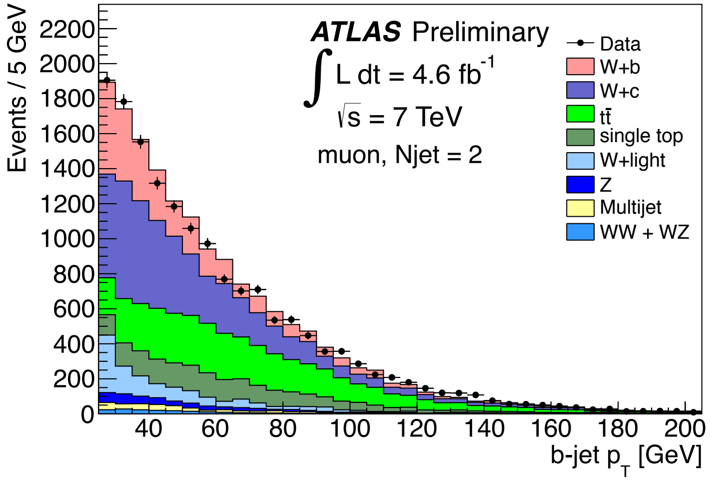
Repeat CombNN fit in each bin

Use correction factors and uncertainties from inclusive measurement.

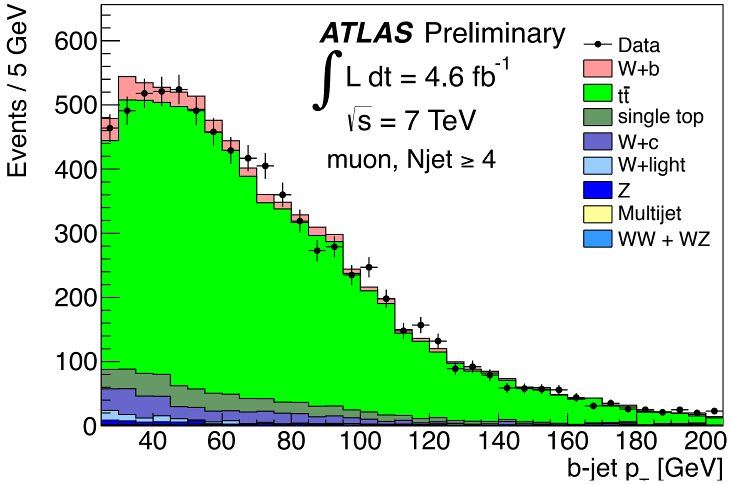
Background model reproduces b-jet pT distribution fairly well.

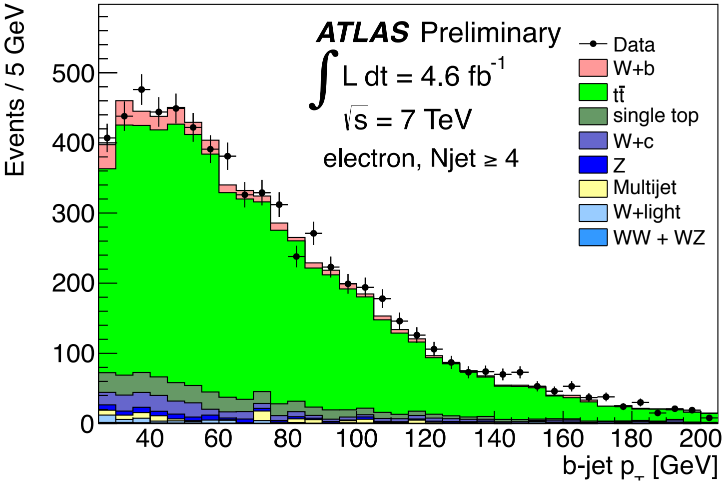
Analysis regions





tt control regions



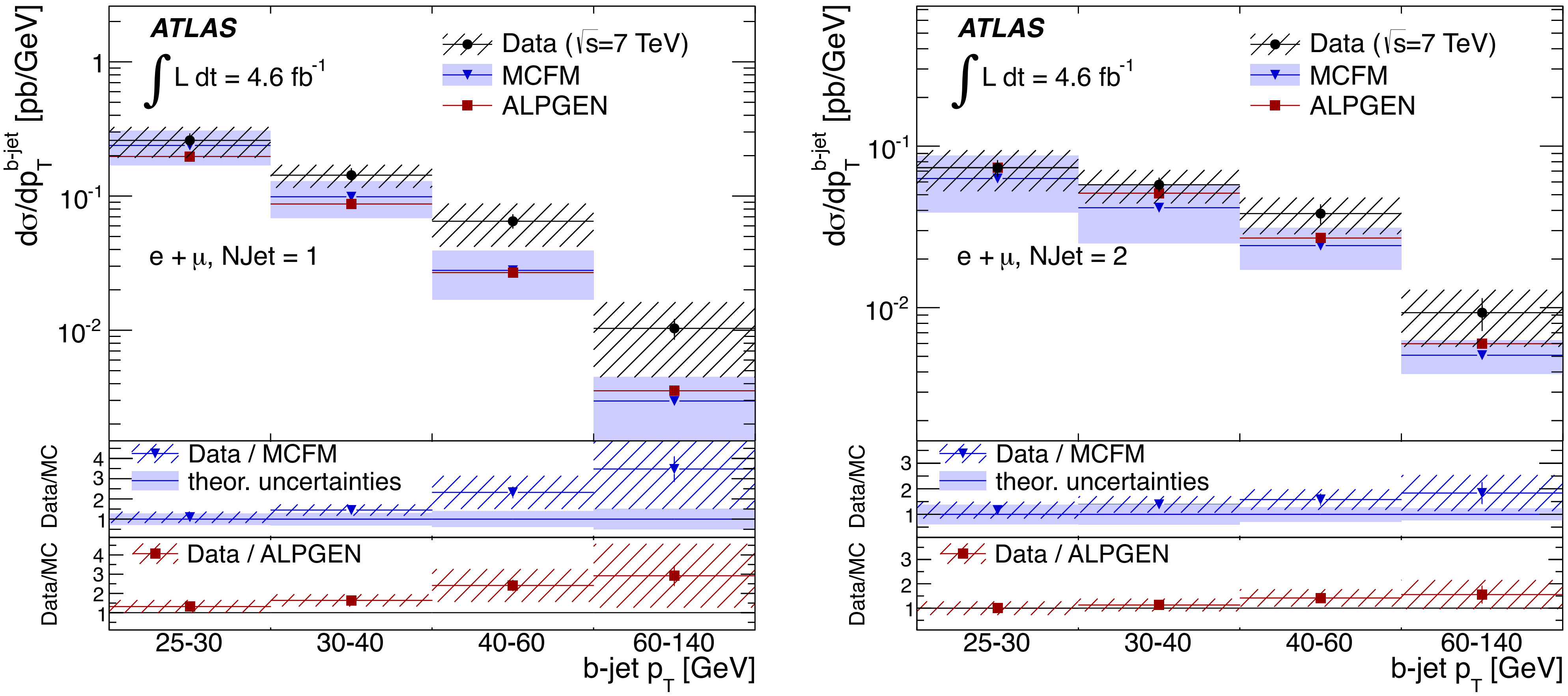


Differential Measurement

Repeat CombNN fit in each bin

Use correction factors and uncertainties from inclusive measurement.

Background model reproduces b-jet pT distribution fairly well.



W+b-jet + single-top

We also quote the Wb cross-section WITHOUT subtracting the single top, both inclusive and differential:

- s-top is a delicate background: interesting check
- Significantly reduced total uncertainties.

The practical implementation:

- Add together single top and W+b template, treated as a single component of the fit.
- Fit the data to estimate the Wb+single top number of events.
- Unfold the fit results into a fiducial cross-section, using the same region of the W+b.
- The latter step uses a response matrix built by combining Wb and single top, assuming the relative normalization (5% additional systematic associated to this assumption)
- Systematics computed as for W+b cross-section.

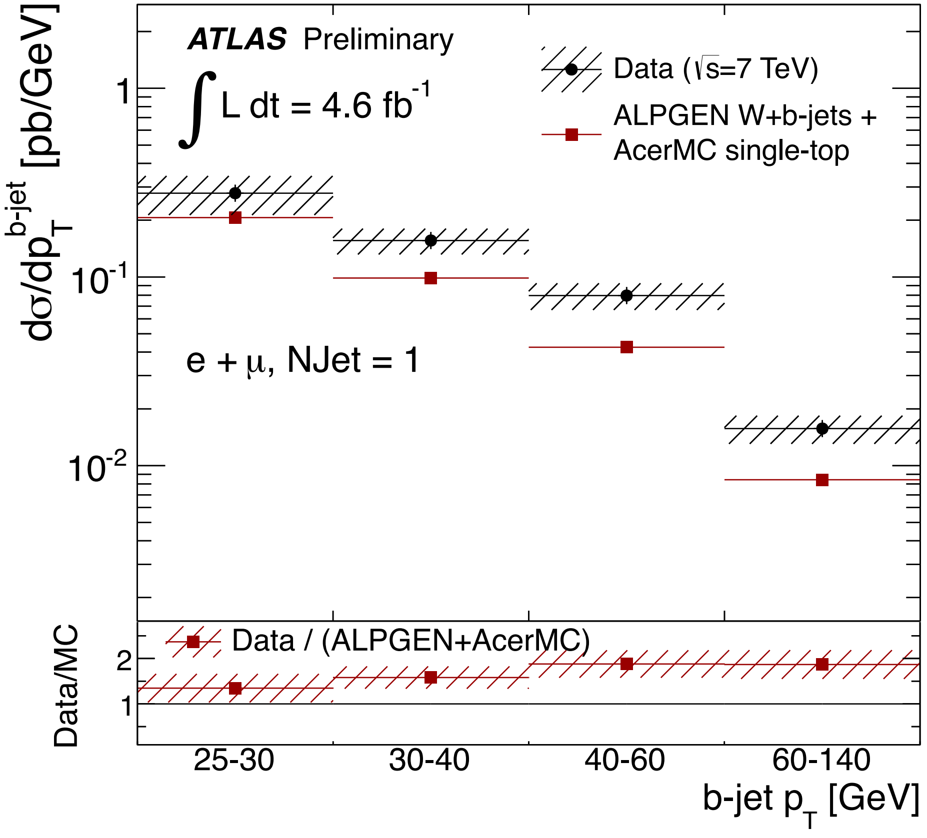
W+b-jet + single-top

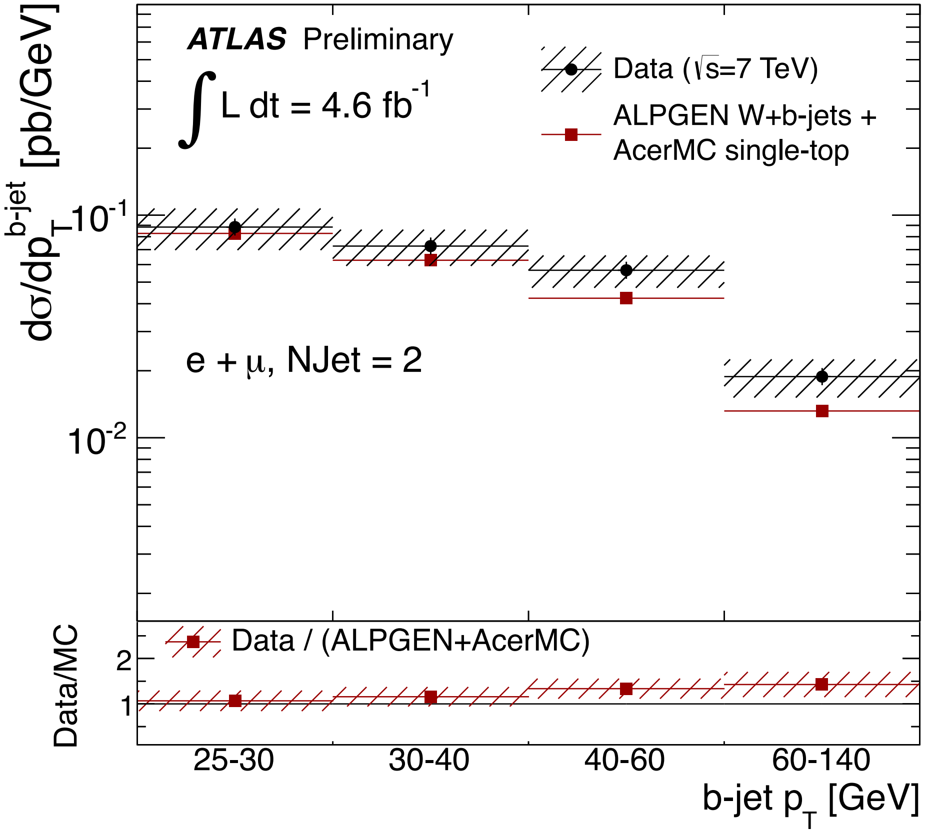
Inclusive:

$$\sigma_{\rm fid}\ (1\ \text{jet}) = 5.9 \pm 0.2\ (\text{stat}) \pm 1.3\ (\text{syst})\ \text{pb},$$
$$\sigma_{\rm fid}\ (2\ \text{jet}) = 3.7 \pm 0.1\ (\text{stat}) \pm 0.8\ (\text{syst})\ \text{pb},$$
$$\sigma_{\rm fid}\ (1{+}2\ \text{jet}) = 9.6 \pm 0.2\ (\text{stat}) \pm 1.7\ (\text{syst})\ \text{pb}.$$

Differential:

u





Conclusions

W boson + b-jets measurements have gained momentum

Precision era is starting, differential measurements can lead the way

Techniques used in W+b-jets measurements are feeding into other analyses, where W+b-jets are backgrounds

Next steps

Differential measurements and ratios of W+b-jets

- Compare 8 TeV with 7 TeV results
- Understand V+c-jets (largest background)

Understand DPI at both theoretical and experimental level <-- Necessary!

Deep connection to challenging performance issues

- Low momentum jets (in pile-up conditions)
- High purity b-tagging (vs. c-jets)

Backup

Dataset, Simulation

Dataset

4.6 fb^{-1} of 7 TeV *pp* collisions collected in 2011

Trigger: Single lepton (electron or muon)

Monte Carlo Simulation

W+jets, Z+jets: ALPGEN + HERWIG parton shower + JIMMY underlying event (including DPI). Normalized to NNLO inclusive W

Top pair production (tt): POWHEG + PYTHIA (norm. to NLO)

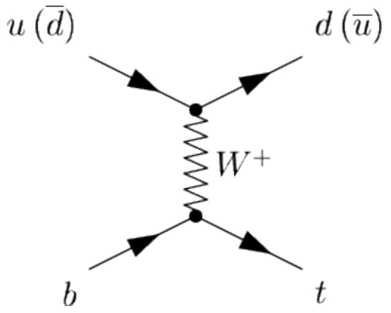
Single top production (st): ACERMC + PYTHIA (norm. to NLO)

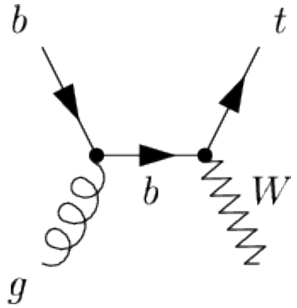
Diboson (WW, WZ): HERWIG (norm. to NLO)

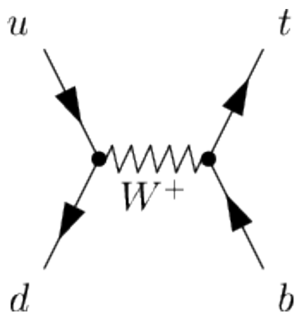
Pile-up (additional *pp* interactions): PYTHIA

ATLAS detector simulation: GEANT

Single-top production

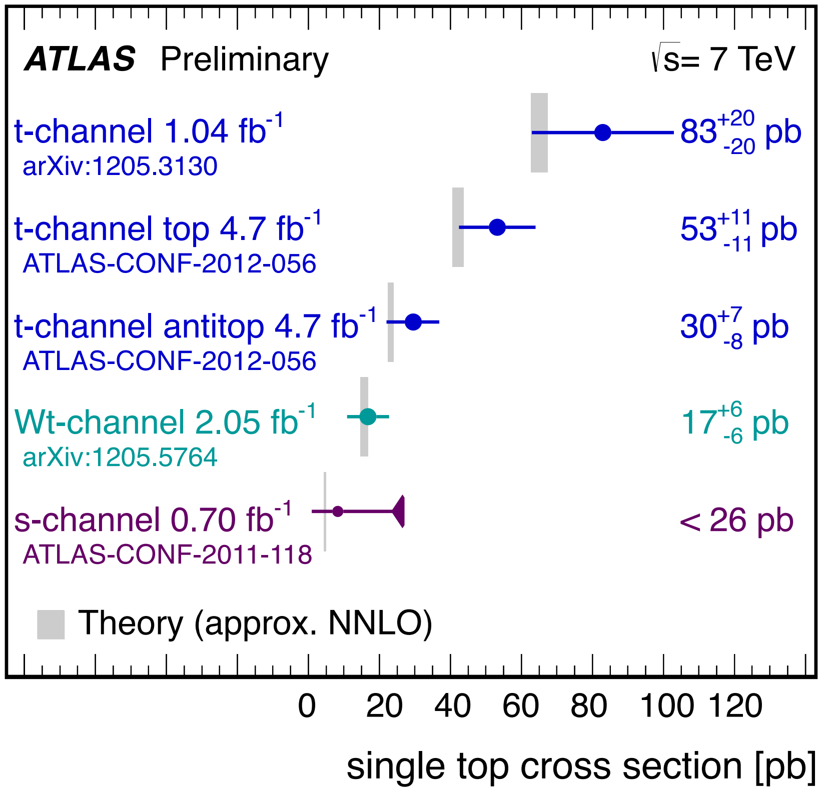






Single-top production mechanisms:

- t-channel
  - Measured in very similar analysis regions: “2-jet 1 b-tag” and “3-jet 1 b-tag”
- Wt
  - Measured using 2 leptons
- s-channel:
  - Limits set using “2-jet 2-b-tag” region



Non-perturbative corrections

Hadronization and UE correction (MCFM):

Estimated in each jet multiplicity and pT bin using particle jets in Powheg with UE switched off.

Systematics on hadronization evaluated using Powheg interfaced to Herwig/Jimmy

Systematics on UE evaluated changing AUET2B into Perugia2011 tune.

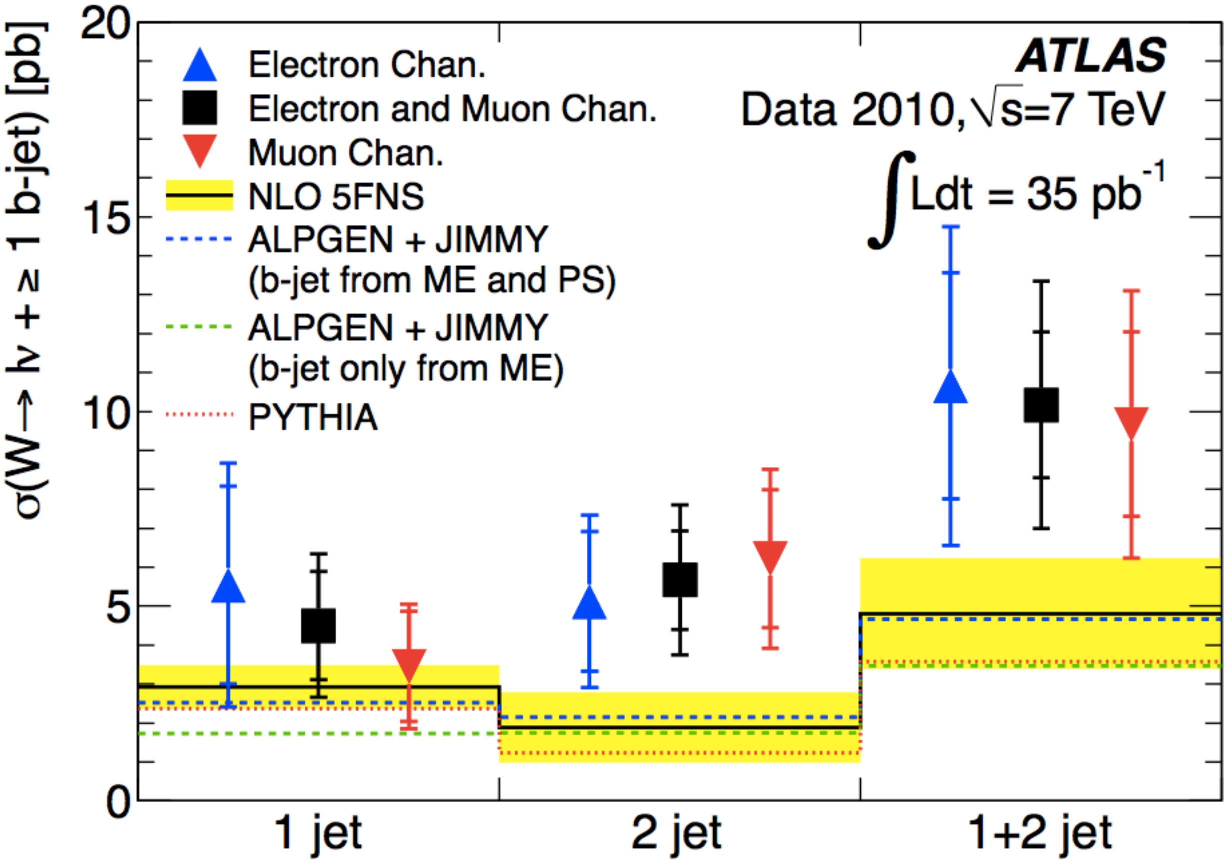
Previous W+b-jet measurement

Differences

2010 dataset: 35 pb^{-1}

Different fiducial region (WmT > 40 vs 60)

Different treatment of single top background (used MC expectation)



Some Questions

How do you estimate the experimental MPI uncertainties?

They affect unfolding (MPI has lower pT --> lower acceptance)

- We use many variations of alpgen (Scale, PDF, MLM matching pT), and ME vs PS, which all affect the MC pT spectrum of the whole W+b component by large amounts.

Why not use forward jets for single-top?

Larger experimental uncertainties

Larger theoretical uncertainties on st (FSR/ISR) and on Wb (unknown)

Why is the W+light fit so jumpy?

Consistency is reasonable.

No QCD shape uncertainties here (which are different for e and mu).

Irrelevant for W+b-jet result: if we fix W+light, the result is the same.

Why not differential in DeltaPhi?

Understanding the backgrounds in bins of deltaPhi is more difficult

In current analysis, bkgs are treated independently in each pT bin

- makes sense since composition varies a lot from bin to bin
- very little assumptions made on shapes of backgrounds

In DeltaPhi, composition is more stable, so would need to make strong assumptions about shape of backgrounds.

Some Questions

Why do you allow for multiple b-jets in the fiducial region?

Extrapolation is smaller than if we excluded them.

In fact most events with two b-jets fall in the 1-b-tag analysis region, so it is natural to include them instead of subtracting them.

What is that ΔR(jet, lepton) cut?

Used in all V+jets analysis. Don't want the lepton from the V to be included in the jet energy.