Estimation of W+jets background to top-quark pair production with first data collected by ATLAS

Maria Ilaria Besana
(Università degli studi di Milano & INFN Milano)

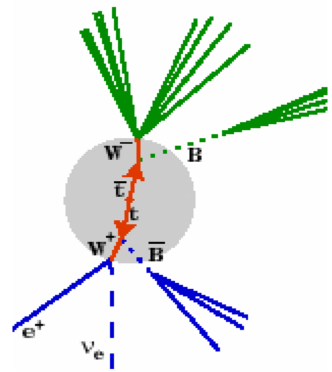
Outline

- Introduction:
  - why top quark physics at the LHC,
  - how we can select top quark pairs with ATLAS detector.
- Data-driven technique developed in order to give an estimation of the most important background: W+jets events.
- Results on first data

Introduction

Top quark pairs: why and how to select them?

- Why top quark pair production cross section measurement?
  - first measurement in a proton-proton collider and at 7 TeV center of mass energy,
  - top quark pair production is a background for new physics
- How we can select top quark pairs?
  - Top quark decays in W+b-jet before hadronising,
  - W can decay:
    - Lepton + neutrino (33%),
    - 2 jets (67%).
  - Top quark pair decay modes:
    - fully hadronic (44%),
    - fully leptonic (11%),
    - semileptonic (45%) .
  - Semileptonic channel is very interesting, because of:
    - good branching ratio,
    - clear experimental signature,
    - invariant mass of 3 jets is equal to top quark mass.



- one energetic lepton,
- missing transverse energy (ETmiss),
- at least 4 energetic jets, 2 of them are b-jets

Top quark pairs events selection

Top quark pair (tT) selection cuts:

- exactly one good lepton,
- ETmiss > 20 GeV,
- at least 4 jets with p_T>25 GeV,
- one of them tagged as a b-jet,
- ETmiss + MT(lepton-ETmiss) > 60 GeV

More details about it in Michele Pinamonti's presentation

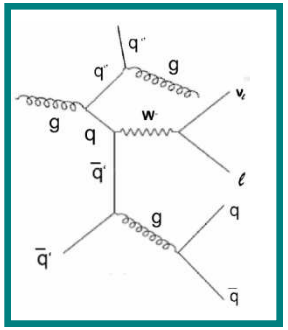
Sample	Expected number of events (2.9 pb^{-1})
tT	30.6
W+jets	3.2
Single top	1.4
QCD	0.9
Other backgrounds	0.2

⟹ **Main background = W+jets !**

W+jets background

- W+jets can have the same experimental signature (cross section for associated production of W and 4 hadronic jets is non negligible at the LHC: ~350 pb).

- W+jets cross section has a big uncertainty : ~60% for W+4jets and ~100% uncertainty on heavy flavour fraction
  - No exact calculation for this process: difficult!
  - Monte Carlo predictions are based on parameters estimated for energy 4 times lower than the LHC one.
  - It's difficult to measure it directly from data because of big top quark contamination.



Need to find a data-driven technique to estimate it from data.

Data driven estimation of W+jets

Data driven technique for W+jets background estimation

- The method is based on 2 main observations:
  - ratio of W+n jets to W+(n+1)jets is expected to be constant as a function of n (Berends scaling),
  - <u>at low jet multiplicity</u> W+jets can be measured from data: lower background contamination with respect to higher jet multiplicity region

- The number of W events with at least 4 jets: can be obtained as:

$$N_{W+\geq 4\,jets} = f_{b-tagged} * \sum_{i\geq 2}\left(\left(\frac{N_{W+2\,jets}}{N_{W+1\,jet}}\right)^{i} * N_{W+2\,jets}\right)$$

where:

N_{W+1jet} = number of selected W+1 jet events

$N_{W+2jets}$ = number of selected W+2jets events

$f_{b\text{-}tagged}$ = fraction of W+≥4jets events that is expected to pass b-tagging requirement

→ from Monte Carlo!

* W candidates events: top quark pair selection cuts applied except the request on jets

Data driven technique for W+jets background estimation

- The method is based on 2 main observations:
  - ratio of W+n jets to W+(n+1)jets is expected to be constant as a function of n (Berends scaling),
  - at low jet multiplicity W+jets can be measured from data: lower background contamination with respect to higher jet multiplicity region

- The number of W events with at least 4 jets: can be obtained as:

$$N_{W+\geq 4\,jets} = f_{b-tagged} * \sum_{i\geq 2}\left(\left(\frac{N_{W+2\,jets}}{N_{W+1\,jet}}\right)^{i} * N_{W+2\,jets}\right)$$

where:

N_{W+1jet} = number of selected W+1 jet events

$N_{W+2jets}$ = number of selected W+2jets events

$f_{b\text{-}tagged}$ = fraction of W+≥4jets events that is expected to pass b-tagging requirement

→ from Monte Carlo!

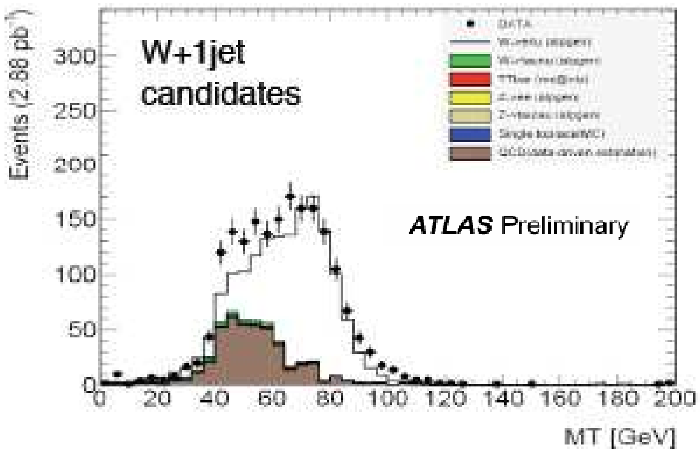
I'm working in particular on W+jets estimation before B tagging in the electron channel.

What we have to do?

- Select W+1jet and W+2jets candidates in data,
- Estimate background contribution: →

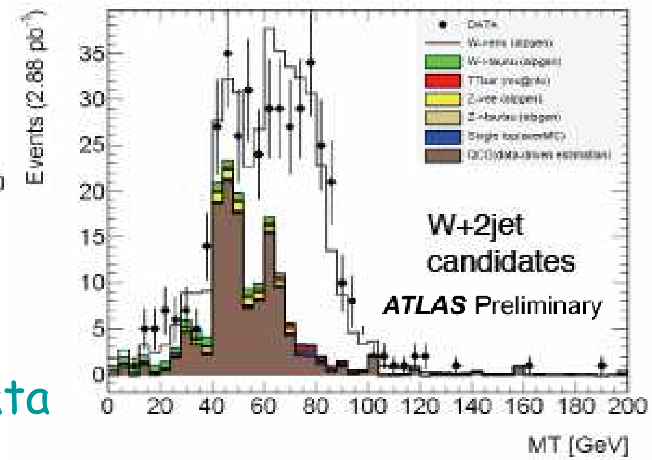
Main backgrounds

W(τν)+jets	Expected rates taken from Monte Carlo, 20% uncertainty
Z+jets	
Single top	Expected rates taken from Monte Carlo, 100% uncertainty
Top quark pair	
QCD	Data-driven technique!



Lepton-ETmiss transverse mass.

Good agreement between data and Monte Carlo: our QCD estimation is reliable



↓ Estimated thanks to the use of an enriched QCD control sample: low ETmiss & looser selection cuts for the lepton.

Maria Ilaria Besana
Università degli Studi & INFN Milano

Results on first data

W+jets estimation

2.88 pb^{-1}

	W+1j	W+2j
Selected events	1789	399
QCD	427	122
Other MC	64	35
Weν measured in data	1298	242
Weν (MC)	1142	260

Background estimation:	
W+2j/W+1j	0.19 ± 0.05
W+2j	242 ± 44
W+4j predicted*	10.9 ± 5.2
W+4j expected	16.4

* Including the W→τν

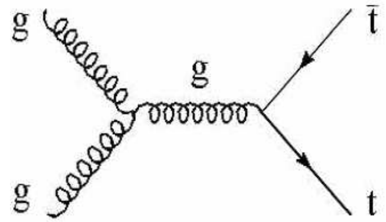
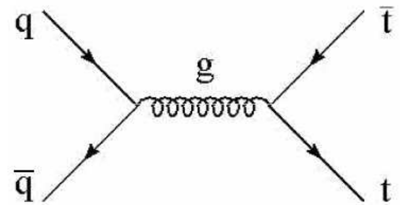
Breakdown of uncertainties	
statistical	28%
Background subtraction	31%
Berends scaling	23%
Total before tagging	48%

Conclusions

- Top physics will be widely studied at the LHC.
- Fundamental for top quark pair cross section measurement: reduce uncertainty on expected number of background events. The biggest background is W+jets.
- I've presented a data-driven technique for W+jets background estimation.
- Thanks to this method we can reduce the uncertainty W+jets background to 50% before b-tagging.

Back-up

Top quark physics: LHC vs Tevatron

	LHC	Tevatron
Center of mass energy	7 TeV	1.96 TeV
Dominant mechanism for top quark pair production	gluon-gluon fusion (82%): 	quark-antiquark annihilation (85%): 
Top quark pair production cross section	~ 160 pb	~6.7 pb
W+4 jets* cross section	~350 pb	~40 pb

* Selection cut applied: p_T > 20 GeV

QCD estiamtion

Matrix method based on this assumption:

ETmiss shape is independent from lepton ID cuts

- Define a control sample with looser selection cuts on the lepton.
- QCD is estimated by solving this 2X2 system:

$$
\begin{aligned}
N^{\text{loose}} &= N^{\text{loose}}_{\text{real}} + N^{\text{loose}}_{\text{fake}}\,, \\
N^{\text{tight}} &= \varepsilon_{\text{real}} N^{\text{loose}}_{\text{real}} + \boxed{\varepsilon_{\text{fake}} N^{\text{loose}}_{\text{fake}}}\,,
\end{aligned}
$$

  - Nloose = number of events with exactly one loose electron (just passin.g RobustMedium)
  - Ntight = number of events with exactly one tight electron (loose + has b-layer and E/p cuts)
  - ε_real measured with Z MC.
  - ε_fake measured for EtMiss < 10 GeV as a function of eta. No correction for W/Z