UNIVERSITÀ DEGLI STUDI DI MILANO

Data-driven estimation of W+jets background to top quark pair production at LHC

M. I. Besana[1]

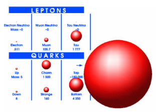
[1]INFN and Università degli Studi di Milano (Italy)

Contact: ilaria.besana@mi.infn.it

Top quark pair production

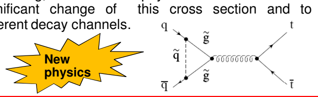
LHC would be a top factory: in the first year of running (with an integrated luminosity of 200 pb^{-1} at 10 TeV) ~ 80.000 top quark pair will be produced. Top quark pair production cross section is indeed enhanced by a factor 100 at 14 TeV (50 at 10TeV) with respect to the one at the now operating colliders. We will be able to perform high-precision measurements.

The top quark is the last particle of the six-member quark family predicted by Standard Model. It was discovered in 1995 at Fermilab.



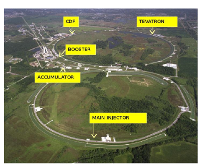
Top quark physics is under investigation at the Tevatron. CDF and D0 collaboration have analyzed more than 3,5 fb^{-1} of data. In particular top quark pair production cross section has been measured with 10% uncertainty.

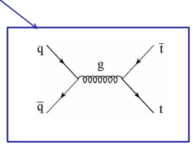
Interesting, because new physics can lead both to a significant change of this cross section and to different decay channels.



Tevatron:

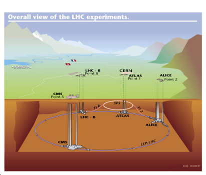
- p-anti p collision
- Centre of mass energy: 1.8 TeV (Run I) and 1.96 TeV (Run II)
- Dominant production process: **quark-antiquark scattering**
- Cross section: 6.97 pb (Run II)
- Important results obtained (top quark mass estimation with 0,7% uncertainty)

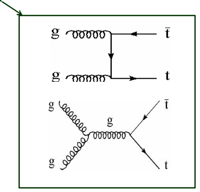




LHC:

- p-p collision
- Design centre of mass energy: 14 TeV (at the beginning 10 TeV)
- Dominant production process: **gluon-gluon fusion**
- Cross section: 833 pb (401 pb at 10 TeV)

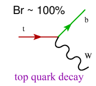




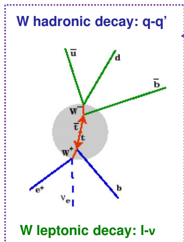
How can we select a top quark pair?

Important features for top quark decay:

- really short decay time (10^{-25} s), so: - it decays before hadronising, thanks to its relatively large width - we can't see top quark, but only its decay products
- decays almost exclusively into Wb, due to the large V_{tb} value of CKM matrix

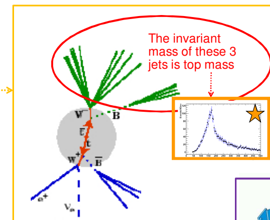


Experimental signature of top quark pair events depends on W decay mode: - ~1/3 into lepton-neutrino, - ~2/3 into quark-antiquark (no b)



In the detector we don't measure partons, but: protons, neutrons, pions, photons, electrons, muons, neutrinos, other particles with laboratory lifetimes >~10ps and the corresponding anti-particles. These particles are clustered in jets.

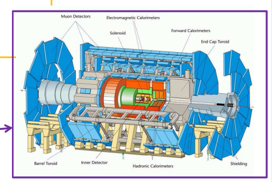
Jets 4-momenta approximate partons 4-momenta



Semileptonic decay is the most studied because of good BR, clear signature (one energetic lepton, at least 4 jets and a neutrino) and the possibility of reconstructing top quark mass.

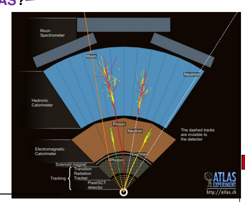
3 decay channels for a top quark pair:

- semileptonic decay: one W decays into ℓ+ν and the other one into q+q̄', BR=4/9
- fully hadronic: both W's decaying into q+q̄', BR=4/9
- fully leptonic: both W's decaying into ℓ+ν, BR=1/9



How particles are detected and identified in ATLAS?

- Electrons:
  - track in the inner detector
  - energy in the electromagnetic calorimeter
  - they stop in the electromagnetic calorimeter
- Muons:
  - track in the inner detector
  - track in muon chambers
- Hadrons:
  - track in the inner detector if they are charged
  - energy both in the electromagnetic and hadronic calorimeter
  - they are usually completely absorbed by hadronic calorimeter
- Neutrinos:
  no interaction in the detector: **ETmiss**



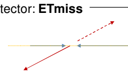
Initial state: Transverse plane: $P_T = 0$; Z axis: $P_z \neq 0$

Constrain on momentum conservation on transverse plane: $\sum P_x = 0$ e $\sum P_y = 0$

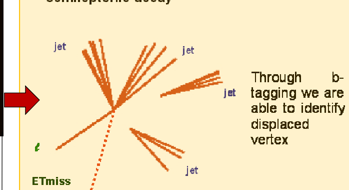
Final state: $P_T = 0$

If a particle is not detected: measured $P_T \neq 0$

$E_T = \sqrt{E_x^2 + E_y^2}$ with $E_{X,Y} = -\sum P_{X,Y}$



Detector view of a top quark pair semileptonic decay



Through b-tagging we are able to identify displaced vertex

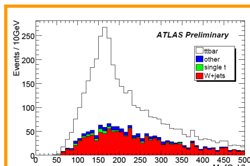
Events are selected asking: one energetic lepton, ETmiss and at least 4 energetic jets.

W+jets background

In order to measure top quark pair production cross section we have to estimate background events:

$$\sigma = \frac{N_{sig}}{\mathcal{L}\times\varepsilon} = \frac{N_{obs}-N_{bkg}}{\mathcal{L}\times\varepsilon}$$

If we don't have a correct estimation of the background processes we will not estimated the true number of signal events: wrong cross section!!

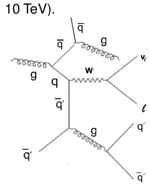


Invariant mass of the 3 jets coming from hadronic top quark decay.

W+jets is the main background.

Uncertainty on σ_{W+jets}	Error on $\sigma_{top\ quark\ pair}$
20%	10%
50%	25%

W can be produced in association with gluons or quarks (initial state or final state radiation). Through hadronisation, they produce a jet. Events with high jet multiplicity are a little fraction of total inclusive cross section, but can't be neglected at LHC energy ($\sigma_{W(->e\nu)+4partons}$ ~ 60 pb at 10 TeV).



Same experimental signature!

W+jets cross section has a big uncertainty : ~100% for W+4jets. No exact calculation for this process, but calculation based on parameters estimated for energy one order of magnitude lower than LHC one.

It' s difficult to measure it from data, because of big top quark contamination.

Data-driven estimation of W+jets background using Z events

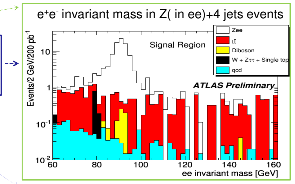
Motivation:

- Z is "easy" to select with high purity (but σ_W~ $10^*\sigma_Z$)
- Ratio W/Z is predicted reasonably well.

Assumption:

$$(W^{SR}/W^{CR})_{data} = (Z^{SR}/Z^{CR})_{data} \cdot C_{MC}, \qquad C_{MC} = \frac{(W^{SR}/W^{CR})_{MC}}{(Z^{SR}/Z^{CR})_{MC}}$$

(log scale) Background contamination: < 4%



where Z_{SR} indicates the number of Z candidate events which pass the same selection criteria as those imposed in the top signal search. W_{CR} and Z_{CR} represent the number of W and Z candidates reconstructed in a control region at low jet multiplicity: 1 jet.

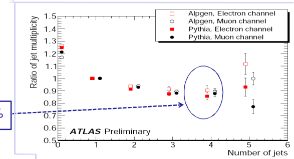
Z events are selected asking:

- two energetic leptons,
- invariant mass of lepton pair in Z-mass window

W events are selected asking:

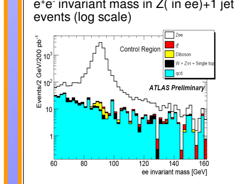
- one energetic lepton,
- ETmiss

Spread ~10%



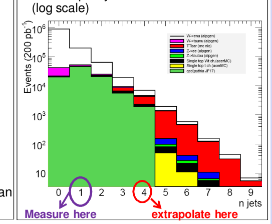
Control region is chosen in order to minimize background contamination and difference between Alpgen and Pythia prediction.

e⁺e⁻ invariant mass in Z(in ee)+1 jet events (log scale)



Background contamination less than 1%.

Jet multiplicity in W in e+ν events (log scale)



Pythia and Alpgen are two different generators. In Pythia the hard scattering is calculated at the lowest order and additional QCD and QED radiation are added in a shower approximation. These predictions are accurate for radiation emitted at small angle, poorest in events with a large number of energetic and widely separated emissions.

In Alpgen hard scattering is calculated using matrix element. Herwig is used to perform the hadronisation and produce initial and final state radiation. There is an algorithm to prevent double counting. Alpgen is target at final state with several, well separated hadronic jets.

Results for electron channel

I'm working on the estimation of W+jets background in the electron channel.

We have obtained :

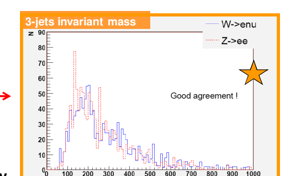
for Alpgen $(W/Z)^{SR}/(W/Z)^{CR}$ = 1 C_{MC}=1

for Pythia $(W/Z)^{SR}/(W/Z)^{CR}$ = 0.8 C_{MC} =0.8

Uncertainty on C_{MC}	
Statistical error for 200 pb^{-1}	11,3%
Purity of the control sample	17,0%
MonteCarlo uncertainty	12.1%
Total error	~ 23%

In the table I've reported a preliminary estimation of the main sources of uncertainty on this method obtained by the Milano group. The statistical contribution depends on the number of expected Z+4jets events, the purity of the control sample is dominated by the uncertainty of QCD contamination in W control region. Monte Carlo uncertainty is the difference between Alpgen and Pythia predictions.

The plot represent the invariant mass of the 3 jets with the highest sum of P_T. There is a good agreement not only in the number of background events, but also in the shape!



Assuming the Standard Model scenario, we expect a measurement of the cross section with a relative uncertainty of ~20% for an integrated luminosity of 200 pb^{-1} at a center of mass energy of 10 TeV.