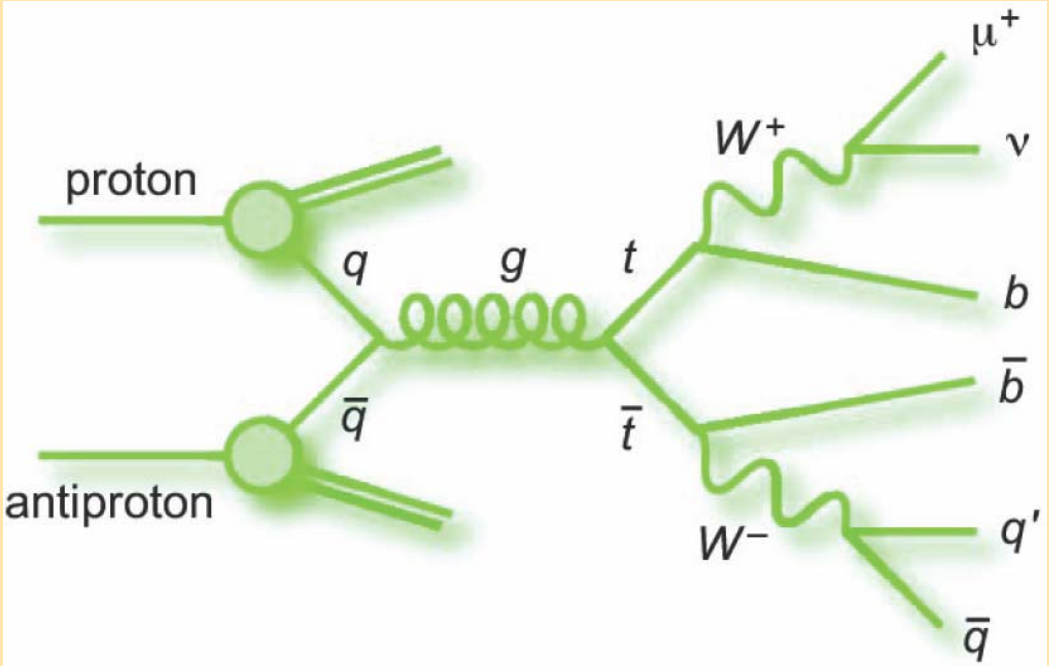
Top pair production cross section in the lepton+jets channel : Methods used in DØ

Sabine Crépé-Renaudin,
LPSC Grenoble



Warning

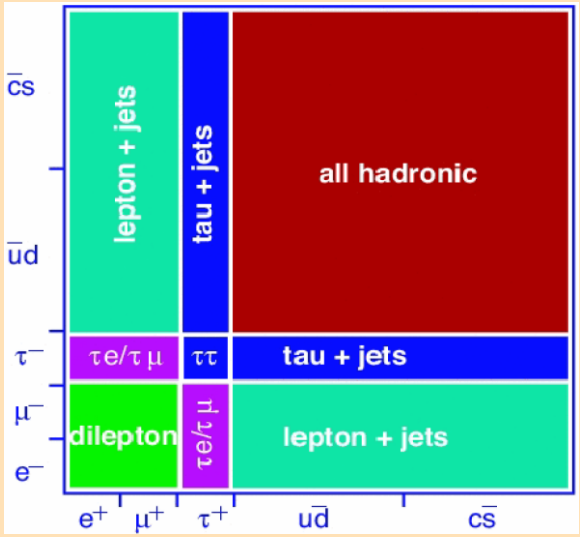
Some of the results shown here are in the approval process by the DØ collaboration and are thus not official.

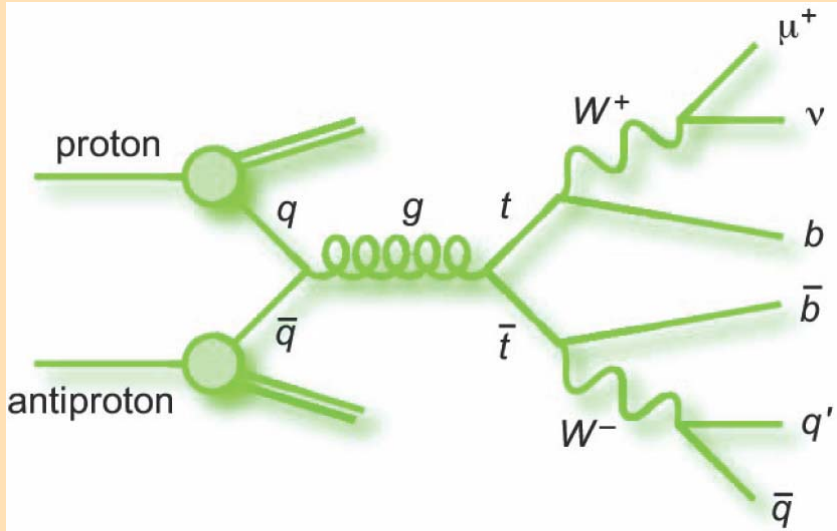
They are used to illustrate the panel of methods used in DØ to measure the ttbar cross section in the lepton + jets channel.

They have been obtained on ~360 pb^{-1} of Run II data.

Top pair production : the l+jets decay channel

- Top pair decays : t t → W b Wb with W → lυ and W → qq





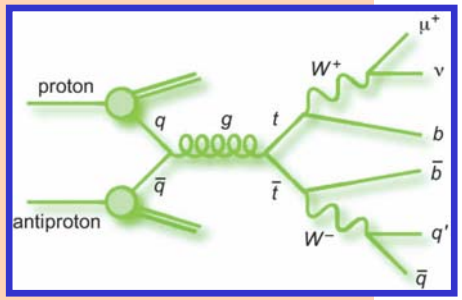
⇨ good compromise between statistics and background level

Two sets of analysis :

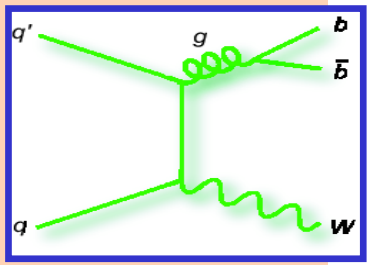
- l = e, μ or τ → e, μ : 35 % of the tt decays ⇒ several strategies detailled
- l = tau → hadrons : 10 % of the tt decays ⇒ will be summarized

The e, μ + jets decay channel

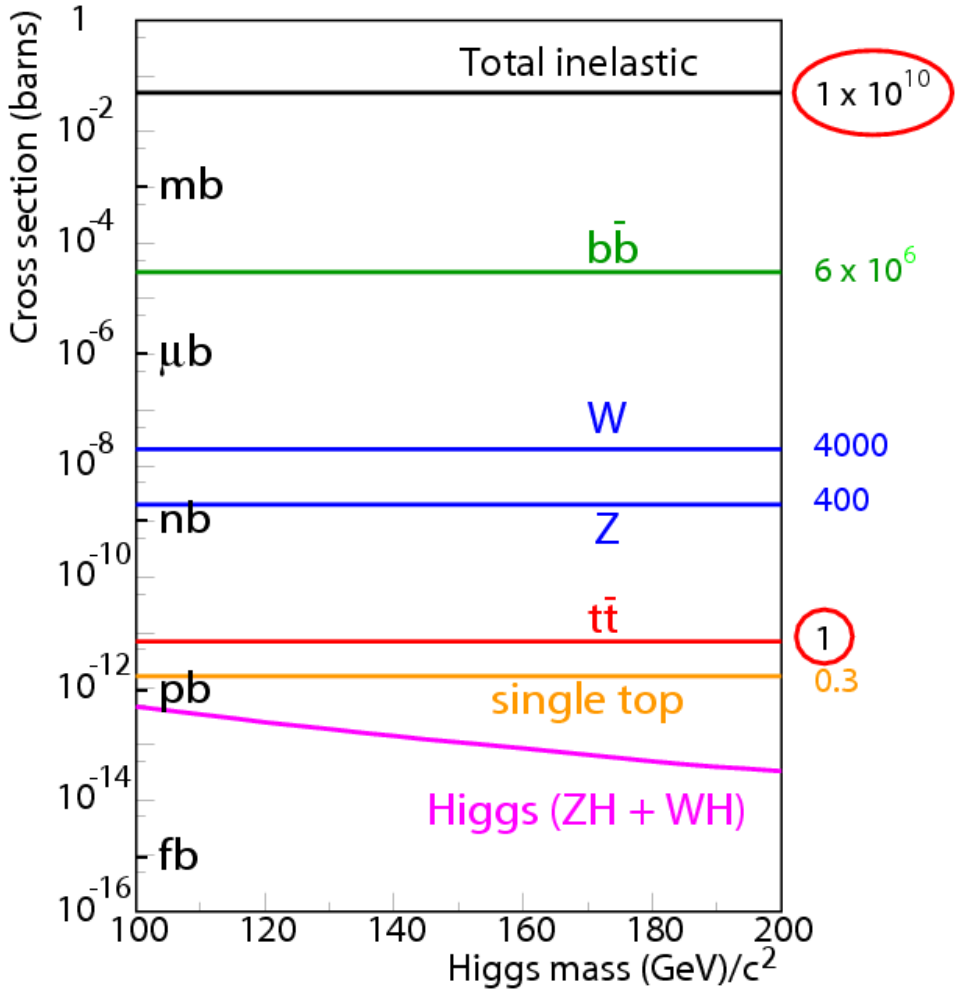
- Signal signature :
  - 1 high p_T isolated lepton
  - Missing E_T from υ
  - 4 jets



- Backgrounds :
  - Physics background
    - W+jets



  - Instrumental backgrounds
    - Multijets (QCD)
    - Z+jets
    - Dibosons
    - Other top channels including single top



Analysis scheme

1. W and jets selection : cuts and efficiency (common to all analysis)
   - Trigger
   - Selection cuts
   - Cuts against backgrounds
   - ⇨ On data and Monte Carlo background samples
2. Background evaluation (common to all analysis)
   - Multi-jets : based on data → "matrix method"
   - Others (not W+jets) : based mainly on MC
3. W+jets and top signal separation : different tools
   - Use event topology
   - Use b-tagging : displaced vertices or soft lepton tag
4. Cross section extraction : $\sigma(tt) = \frac{N_{data} - N_{bkg}}{L * BR * \varepsilon^{trig} * \varepsilon^{tt}}$
5. Systematics evaluation

Selection of a W and jets sample

- Data quality : checked at run to event level
- Trigger requirements :
  - ≥1 e or μ with p_T > 15-20 GeV
  - ≥1 jet with E_T > 20-25 GeV
- Selection :
  - Good vertex
  - ≥3 jets : E_T(leading) > 40 GeV E_T(others) > 20 GeV, |η| < 2.5
  - 1 isolated lepton :
    - Loose selection : p_T > 20 GeV, |η(e)| < 1.1 |η(μ)| < 2.0

    from primary vertex, matched with trigger and track from the central tracking system
    - Tight selection : tight cut on e → likelihood, μ → isolation/pT
  - Missing Transverse Energy (MET) > 20 GeV
- Anti-background cuts
  - Against multi-jet : Δ(φ,MET)
  - Against Z+jets : dilepton mass not in Z window
- Orthogonality cuts wrt dilepton analysis
  - No other isolated lepton

- QCD events with a misreconstructed jet as an e or μ leads to mismeasured p_T and thus MET in the direction or back to back to the lepton; same behavior for a semileptonic decay of a jet with a muon
- Cut is optimized with respect to tt (efficiency x purity)

e+jets channel

μ+jets channel

Efficiency measurement

- For trigger
  - Trigger efficiency measured on data for each trigger level and per object (lepton, jet …) and then combined (see Benoit)
- For selection
  - Use MC for top signal and small or top related backgrounds
  - Use data for multijets background
  - Use a mix of information from data and MC for W+jets and Z+jets

⇨ Top pair (≥ 3 jets) preselection efficiency : ~ 21 % (12, 9 % bin 3, ≥ 4)

Note 1 :

Suppose that object reconstruction is optimized and Monte Carlo faithful to reality; if not : correct MC →

" scale factor " (Cf Benoit)

Note 2 :

Monte Carlo samples : mainly Alpgen for the hard scatter and Pythia for underlying event and hadronisation. PDF : CTEQ5-6L.

Multijets event evaluation

- Determined from data using the matrix method
  - i.e. take advantage of the fact that, for a fake lepton and a true lepton passing the loose selection, the probability to pass a tight selection is different.
  - Tight selection : for e+jets → likelihood, for μ+jets → isolation

- Matrix method :
  - $N^{loose} = N_{QCD}^{loose} + N_{sig}^{loose}$ (1)
  - $N^{tight} = N_{QCD}^{tight} + N_{sig}^{tight}$

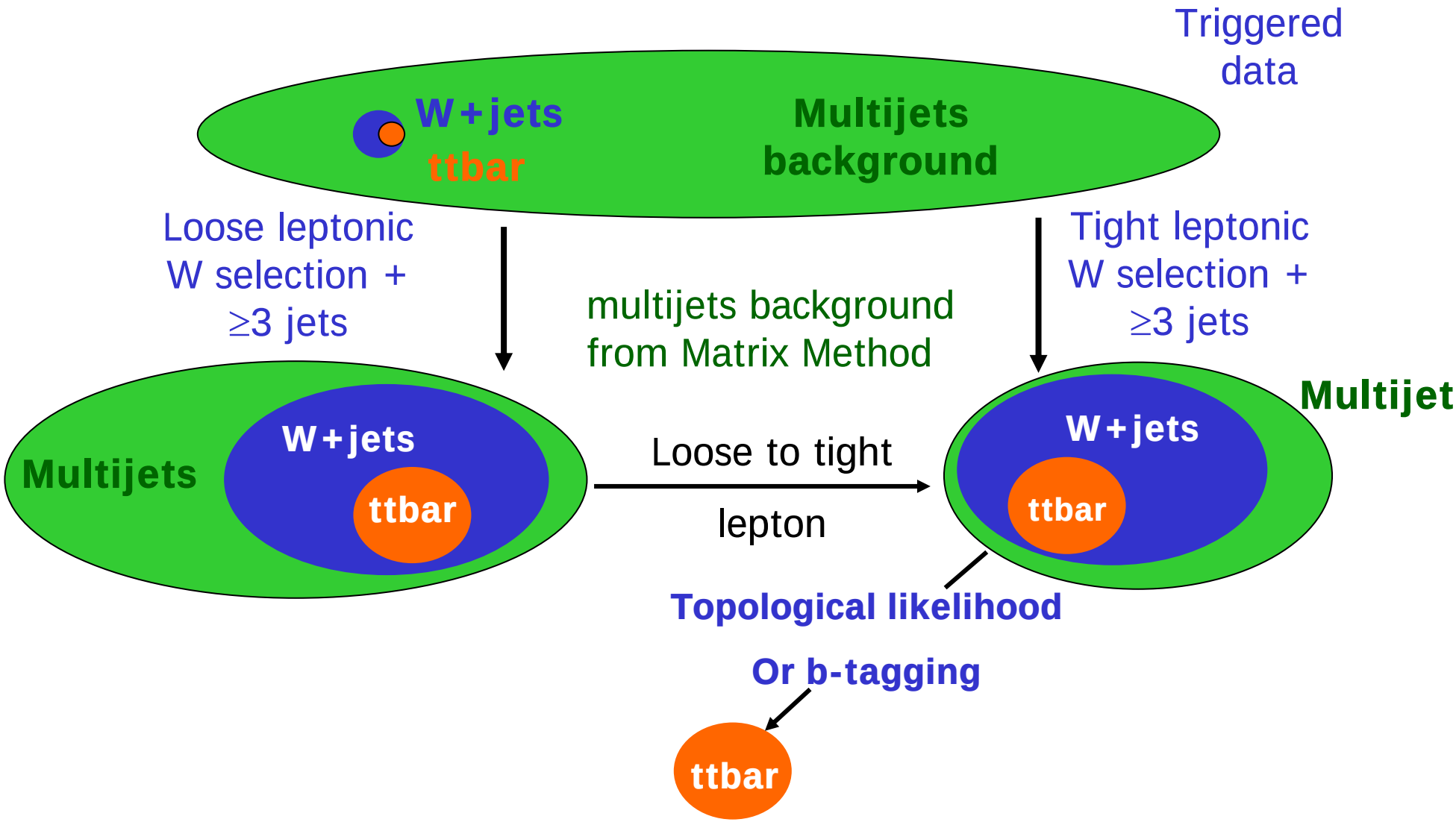
  $\Rightarrow N^{tight} = \varepsilon_{QCD}\, N_{QCD}^{loose} + \varepsilon_{sig}\, N_{sig}^{loose}$ (2)

  where ε_{QCD} and ε_{sig} are probability to pass the tight cut
  they are measured independently on data (low MET region) and W+jets MC resp.

  - Solve (1) + (2) ⇒ get N_{QCD}^{tight} and N_{sig}^{tight}

QCD ⇔ fake lepton
sig ⇔ true lepton
= W + Z + top + …

Sample composition



Topological Analysis

- The aim is to use the specific topology of top events
  - Necessity to combine several variables in a likelihood discriminant with signal = ttbar and background = W+jets (in the 4th jet bin) :

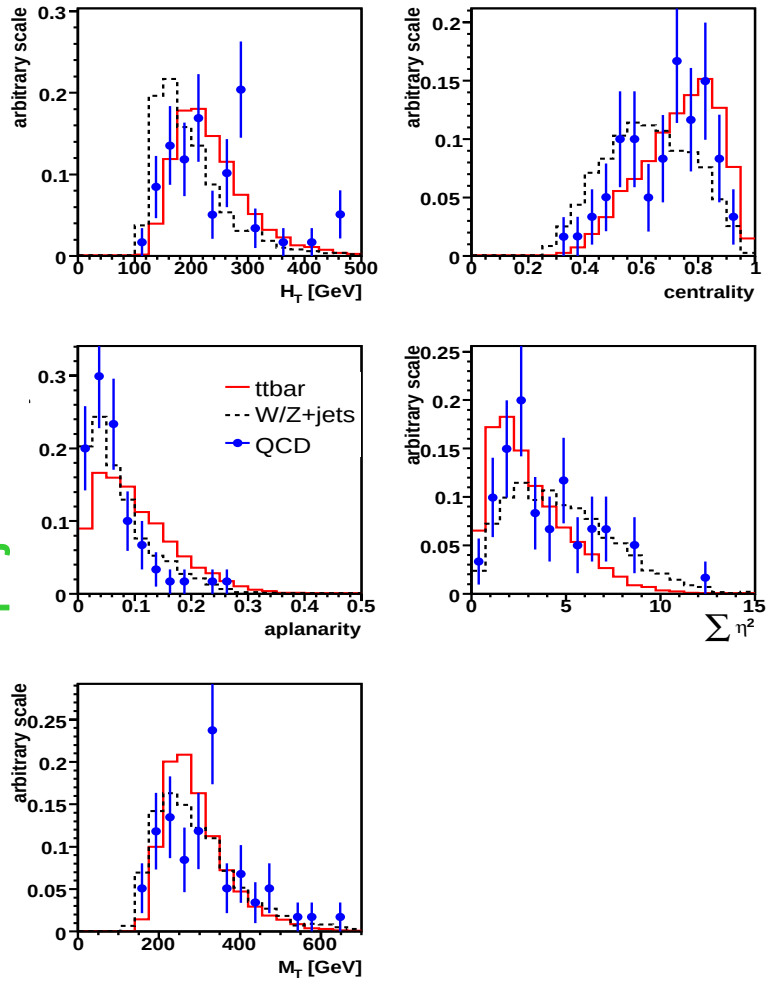
$$D = \frac{S(x_1, x_2, \ldots)}{S(x_1, x_2, \ldots) + B(x_1, x_2, \ldots)} \approx \frac{\exp(\sum_i (\ln \frac{s_i}{b_i}))}{\exp(\sum_i (\ln \frac{s_i}{b_i})) + 1}$$

  - The variables are chosen between the ones - if used alone - minimize the uncertainty on the cross section (err stat and main syst - jets -) (and whose MC distribution reproduces well reality).
  - The likelihood function is optimized to minimize the uncertainty on cross section using all systematic sources that influence the likelihood shape.
  - Templates are built for signal and all backgrounds.

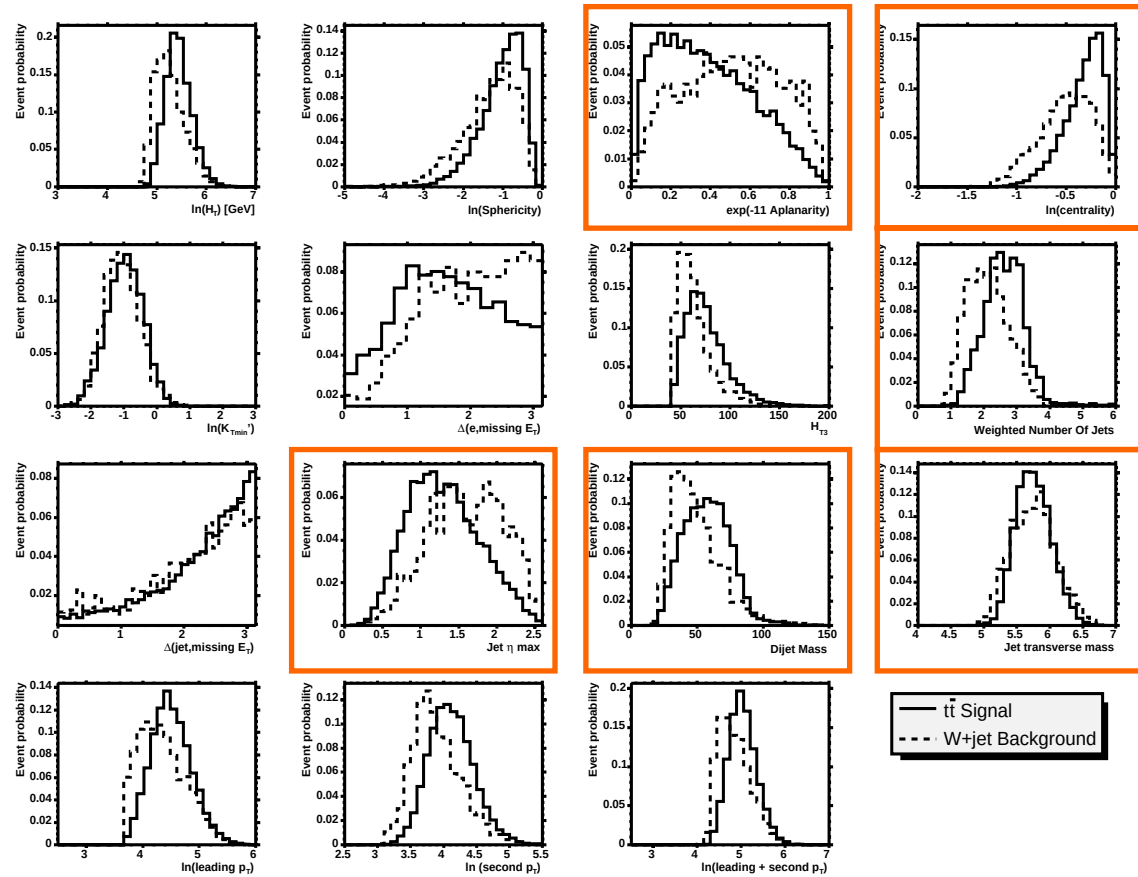
- Number of ttbar events is extracted by performing a maximum likelihood fit to the likelihood discriminant distributions (the number of multijets event is constrained to the output of the matrix method) .

Topological analysis : discriminant variables

μ+jets channel



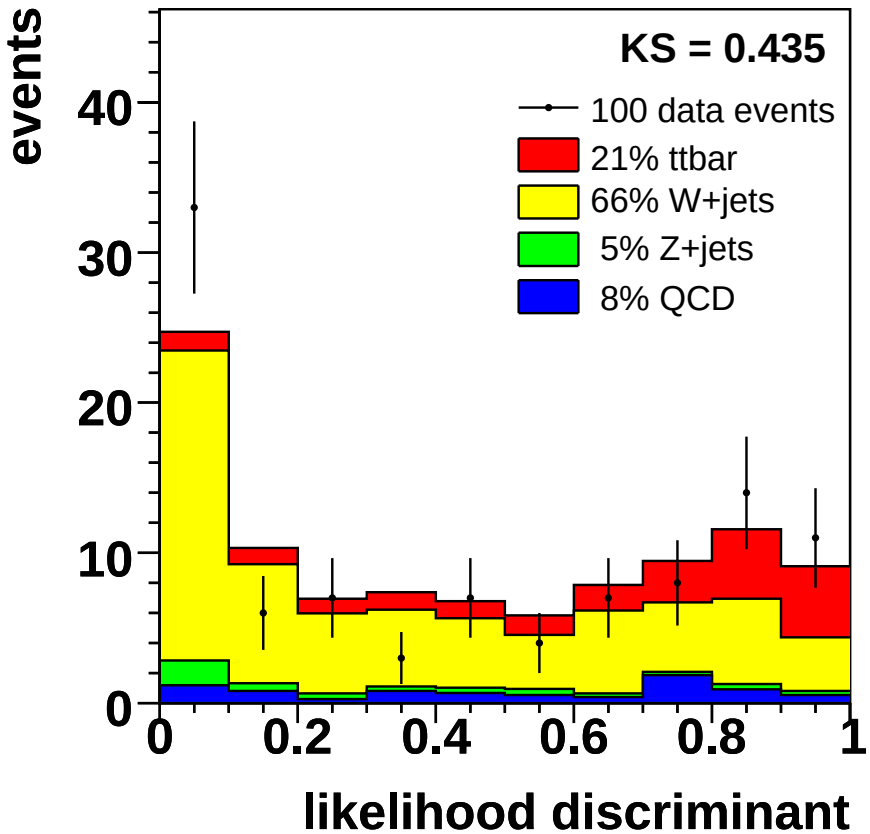
e+jets channel

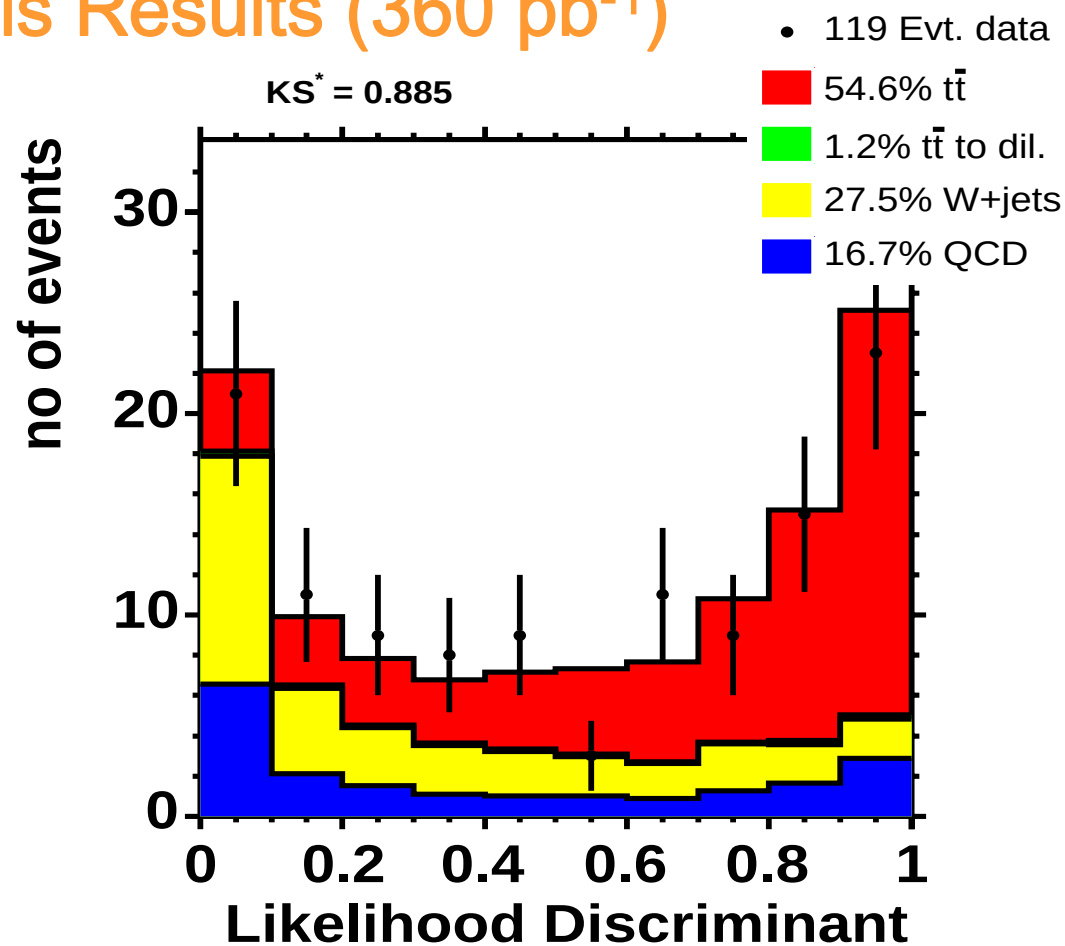


μ+jets variables : H_T, centrality=H_T/H, aplanarity (momentum tensor), $\Sigma\eta^2$ (jet) and m_TW

e+jets variables : centrality, aplanarity, weighted number of jets, jet η max,
minimum dijet mass, jet m_T

Topological Analysis Results (360 pb^{-1})





σ (μ+jets) = 3.6 + 1.8 - 1.7 (stat) + 0.5 - 0.5 (syst) + 0.25 - 0.22 (lumi)

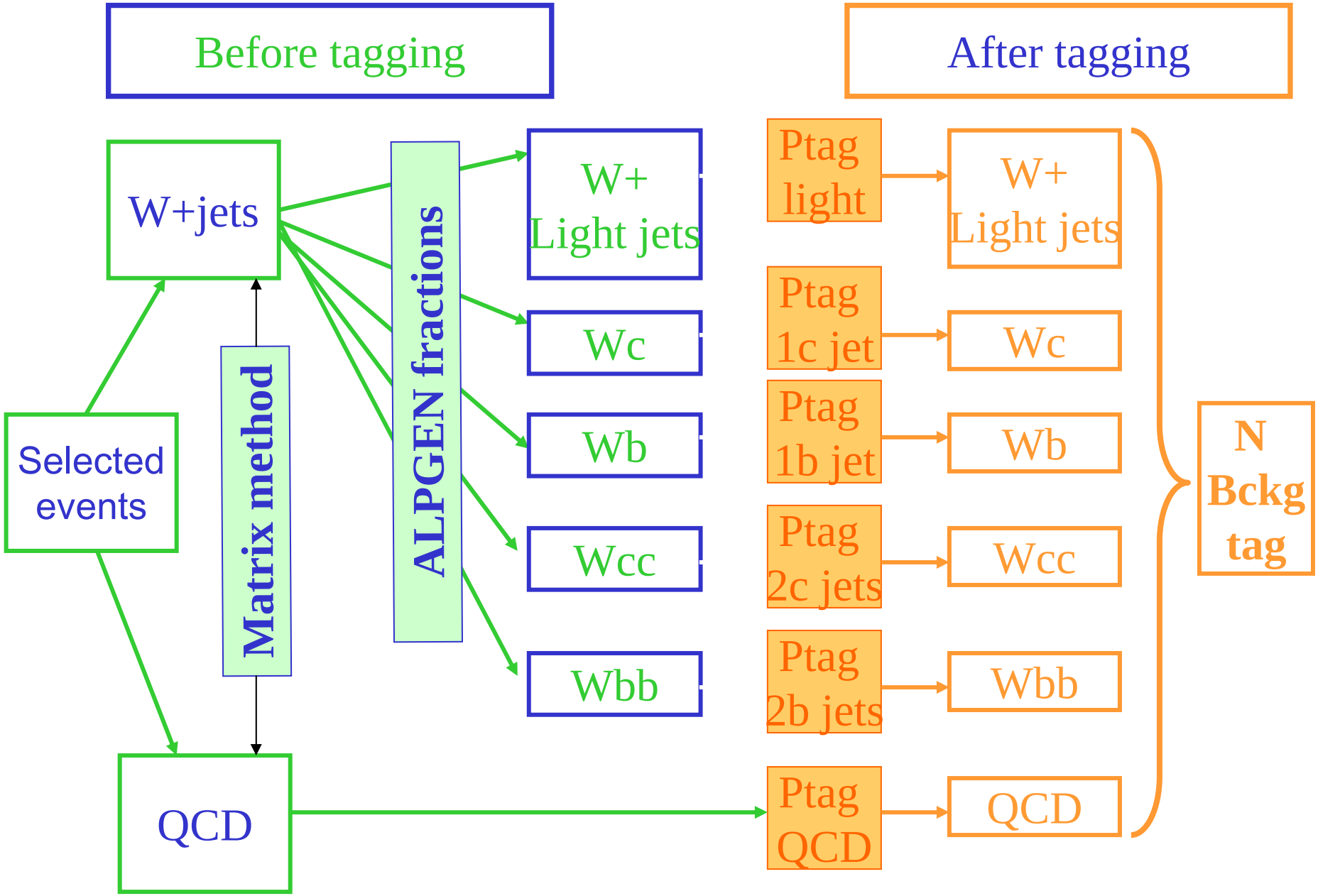
σ (e+jets) = 11.4 + 2.4 -2.3 (stat) + 1.1 -1.2 (syst) + 0.8 -0.7 (lumi)

B-tagging analysis

- The presence of two b jets in top pair events allows to discriminate top from background events and enhance greatly signal to background ratio
- Two b-tagging techniques have been used to measure the ttbar cross section : lifetime tagging and Soft Muon Tagging
- Both b-tagging analysis follow similar scheme :

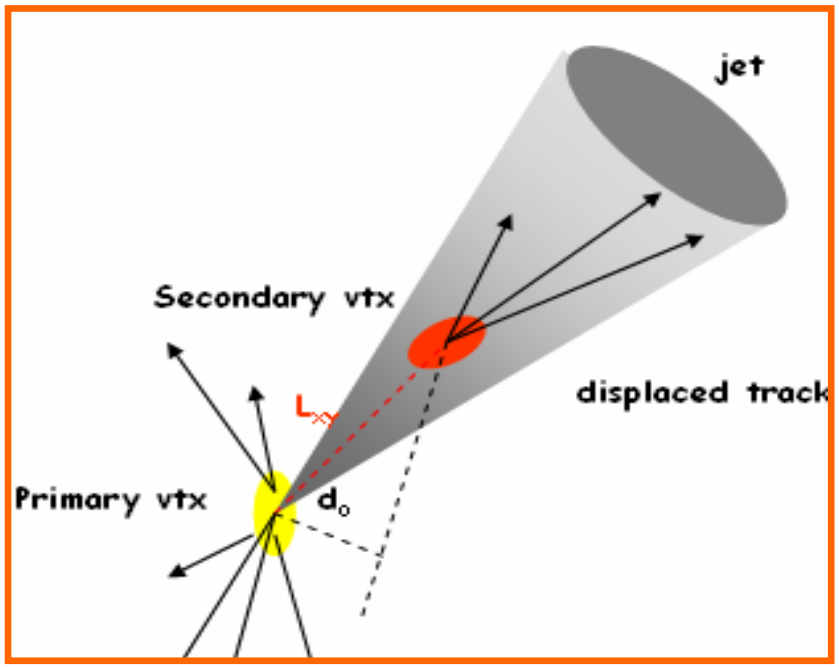
1. After W+jets selection, data sample are tagged
2. Number of QCD and W+jets events are deduced from matrix method **
3. The tagging efficiency is measured for all backgrounds either on data or MC samples : tricky case → the W+jets background
4. Cross section is deduced using a maximum likelihood method

** the output of the matrix method is the number of events containing a true lepton : number of ttbar events have to be deduced from them to get the number of W+jets event. This is taken into account in the likelihood method.


Before tagging
After tagging
W+jets
Selected events
Matrix method
QCD
ALPGEN fractions
W+ Light jets
Wc
Wb
Wcc
Wbb
Ptag light
Ptag 1c jet
Ptag 1b jet
Ptag 2c jets
Ptag 2b jets
Ptag QCD
W+ Light jets
Wc
Wb
Wcc
Wbb
QCD
N Bckg tag

Lifetime tagging

⇨ b jets have large track multiplicity and contain B-hadrons which travel Lxy ~ 3mm before decaying.



⇨ ttbar have ~ 60 % tagging probability

SVT (Secondary Vertex Tagger) algorithm starts from track jets and

- Reconstructs secondary vertex
  - $\geq$ 2 tracks with $p_T \geq$ 1GeV,
  - $\geq$ 1 SMT hit
  - impact parameter significance >3.5
- Remove tracks associated with K^0_S, Λ^0 and photon conversions ($\gamma \rightarrow e^+e^-$)
- Positive tag:
  - Secondary vertex within a jet with a decay length significance $L_{xy}/\sigma_{Lxy}>7$
- Negative tag:
  - Secondary vertex within a jet with a decay length significance $L_{xy}/\sigma_{Lxy}<-7$ (due to resolution effects)

Soft Lepton Tagging

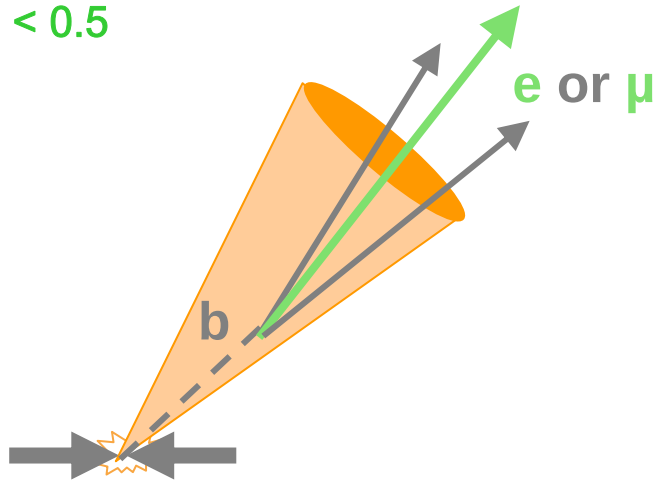
- In top pair production with lepton + jets decay
  - 2 b jets
    - $b \rightarrow l + X \sim 10.7$ %
    - $b \rightarrow c \rightarrow l + X \sim 9.6$ %
  - 1 c jets in 1/3 W decays
    - $c \rightarrow l + X \sim 9,6$ %

⇨ $b \rightarrow l$ in more than 20 % of the cases for both $l = e$ and $l = \mu$

⇨ ~ 40 % of events have a jet $\rightarrow l$ with $l = e$ or μ

- Soft muon tag :
  - 1 medium μ with $p_T > 4$ GeV, $|\eta| < 2$, not a cosmic, $\Delta R(\text{jet}) < 0.5$
  - associated with a track

- Efficiencies
  - 46 % efficiency to tag a muon in a jet
  - ~ 0.4 muon in jet per ttbar event and 16 % tagging efficiency in MC for ttbar
  - ~ 10 and 19 % efficiency to tag a b and bb event
  - ~ 5 and 10% efficiency to tag a c and cc event
  - ~ 0.2 % probability to tag a light jet



B-tagging analysis : Cross-section

- Analysis done for each jet multiplicity bin :
  - 1 and 2 jets : dominated by background ⇒ allow to show that background is correctly evaluated
  - 3 and ≥4 jets : used for cross section measurement
- Cross section :

$$\sigma(tt) = \frac{N_{data} - N_{bkg}}{L * BR * \varepsilon^{trig} * \varepsilon^{tt}}$$

  - The observed σtt depends on the number of selected events in data (loose, tight, tagged or not) : N^{obs}
  - We want the true σtt which is related to N^{true}:

$$N^{true}(\sigma tt) = Ntt^{true}(\sigma tt) + Nbgd^{true}$$

  - The cross section is the one which maximize the likelihood giving the Poisson probability that N^{obs} is observed when N^{true} is expected : N^{true} is let floating but its value is constrained to N^{obs}

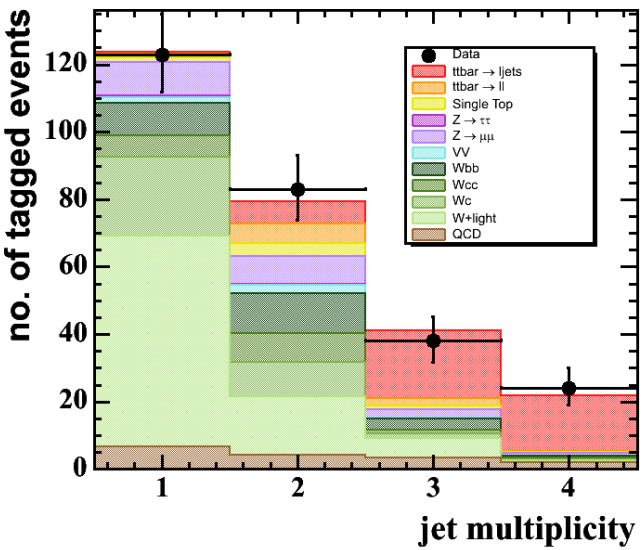
$$\mathscr{L}(N^{true}, \sigma tt) = P(N^{obs}, N^{true}(\sigma tt))$$

  - To take advantage of the different behaviour of signal and background, the likelihood used is the product of Poisson probability for each channel and event class : electron and muon, 3rd and 4th jet bins. Lifetime tagging uses also 1 and 2 tags :

$$\mathscr{L}(N^{true}_i, \sigma tt) = \prod P(N^{obs}_i, N^{true}_i(\sigma tt))$$

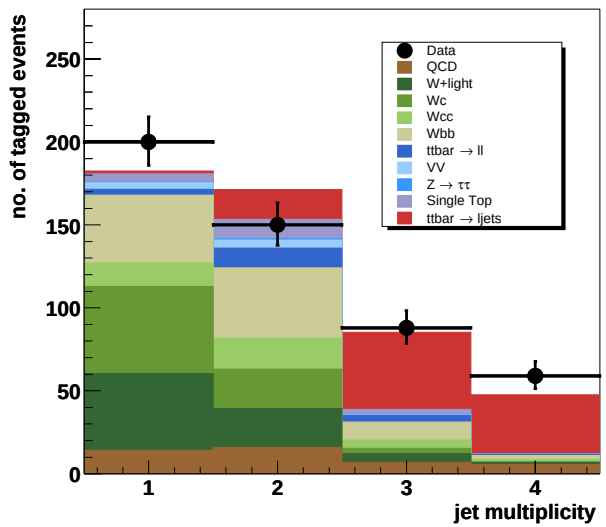
B-tagging methods : Results (360 pb^{-1})

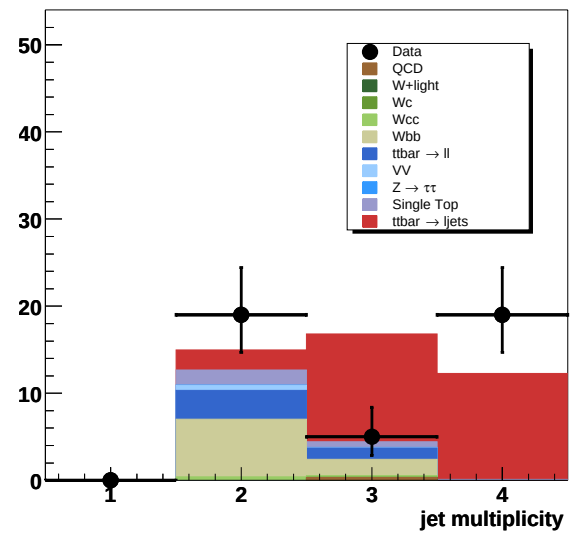
Soft Muon Tagging



σ = 8.42 +2.26 -2.05 (stat + syst)
± 0.55 (lumi)

Lifetime Tagging





σ = 7.7 + 1.1 -1.0 (stat + syst)
± 0.5 (lumi)

Systematics

- Topological analysis
  - 10000 pseudo experiments are generated
  - Likelihood templates are recomputed varying the source of systematic uncertainties
  - ttbar nber of events is fitted using unmodified and modified templates
  - Systematics is the gaussian mean of the relative difference

  - Dominant systematics : template statistic (W+jets), JES, nber of loose-tight event for QCD evaluation

- B-tagging analysis : Nuisance parameter likelihood
  - All systematics are let free in the likelihood according to a gaussian distribution with width equals to their uncertainty
    ➔ Offsets in σ(tt)
    ➔ Takes correlations naturally into account

  - Dominant systematics : tag rate (semi leptonic b-tag, fake rate), W fractions, stat in matrix method

τ+jets channel

- Only hadronic τ decays are considered (3 types : π, ρ, 3 prong)
  - identified as narrow jets using a NN (profile, isolation, track isolation …)
- Analysis :
  - Selection :
    - Trigger : 4 jets E_T > 10 GeV |η|<3.6
    - 4 jets, MET, 1 τ and 2 b-jets SVT tagged
  - Backgrounds :
    - W+jets evaluated with Alpgen sample
    - QCD : τ fake rate is measured on a b-veto sample vs η and p_T
  - Topological NN
    - Topological variables : Aplanarity, sphericity, centrality, HT, √s
    - Top and W mass likelihood, b jets pT and decay significance

- Results (combination of 1 and multi prong):

$$\sigma = 5.05 + 4.31 - 3.46 \text{ (stat)} + 0.68\text{-}0.67 \text{ (syst)} \pm 0.55 \text{ (lumi)}$$

- Dominant systematics : JES, MC stat, b-tagging

Conclusion

- Top events in l+jets channel have to be extracted from important QCD and W+jets backgrounds
  - Different strategies have been developed :
    - Topological likelihood
    - B-tagging
  - After top event selection, amount of main background (W+jets and QCD) are mostly evaluated using data

- Total uncertainty ~ 15% is approaching theoretical one. Statistical uncertainties reach systematics level …