background estimation for semi-leptonic top pair production

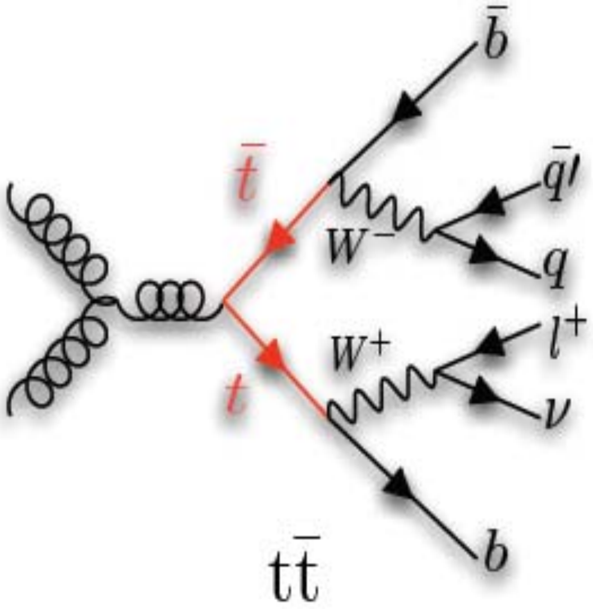
towards a top pair production cross-section measurement

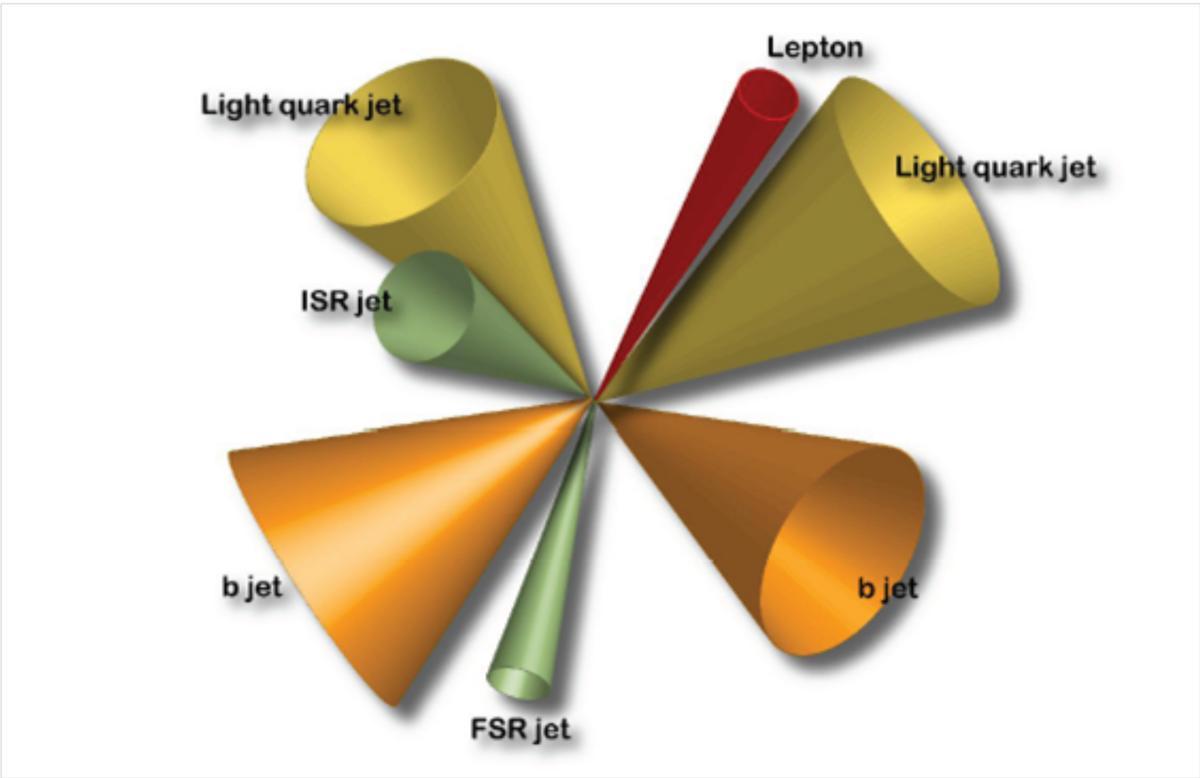
Clemens Lange (DESY)
DESY top & friends meeting
DESY
15th November, 2010

HELMHOLTZ
ASSOCIATION

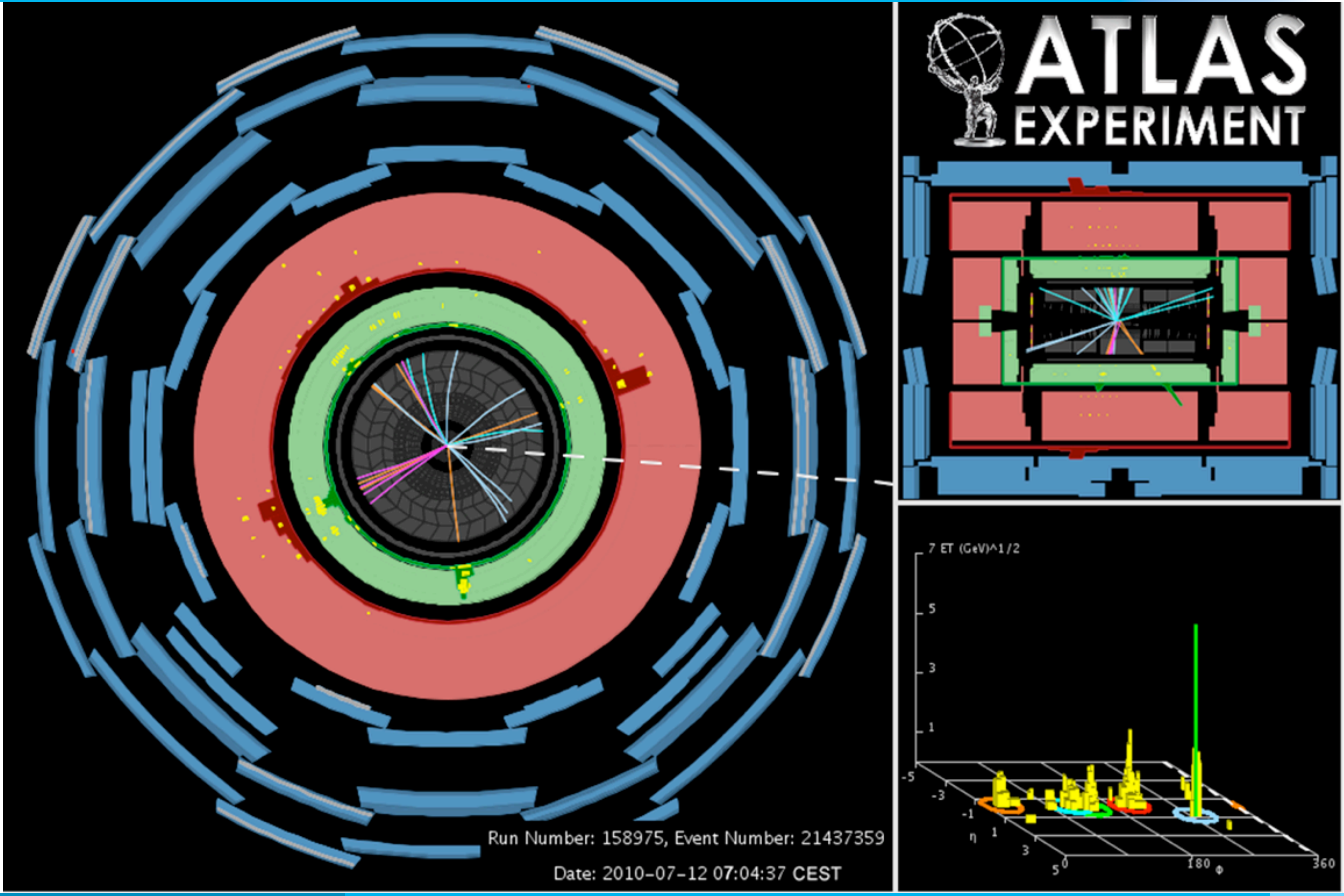
top quark signal selection

- event selection follows signal topology
- require most efficient trigger for signal
- exactly one isolated high-quality lepton (electron or muon) that matches with trigger (electron), $p_T > 20$ GeV
- require ≥ 4 jets, $p_T > 25$ GeV
- one of them b-tagged
- MET > 20 GeV (neutrino)
- topological cut on MET to further reduce QCD



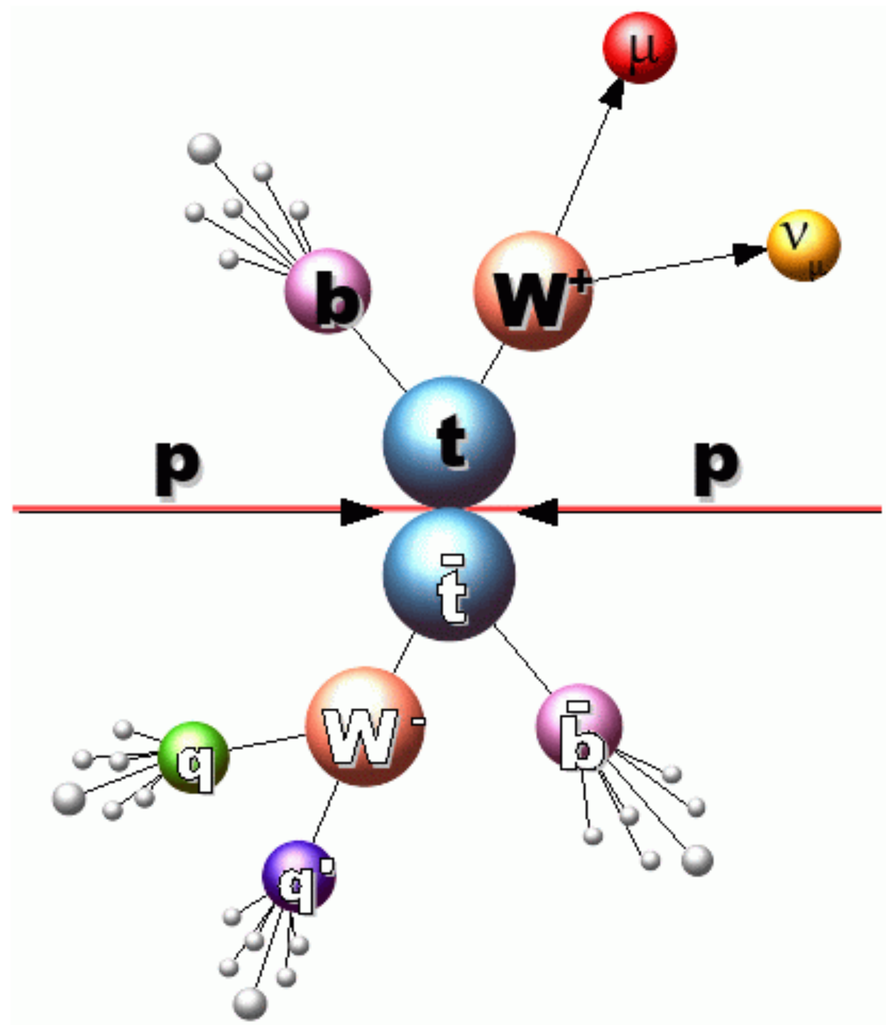


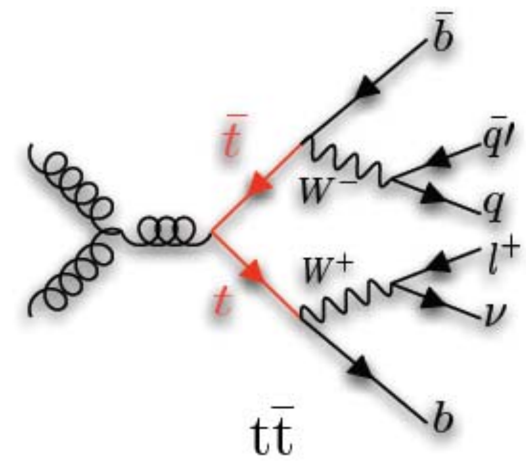
top quark candidate at ATLAS

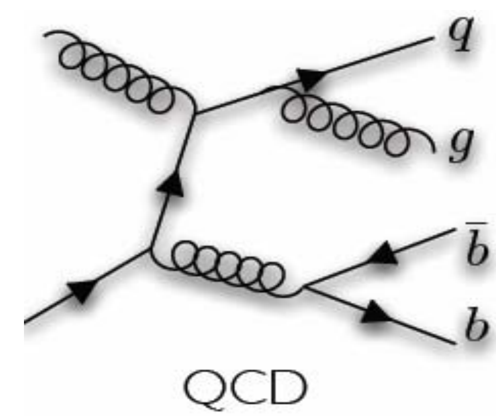


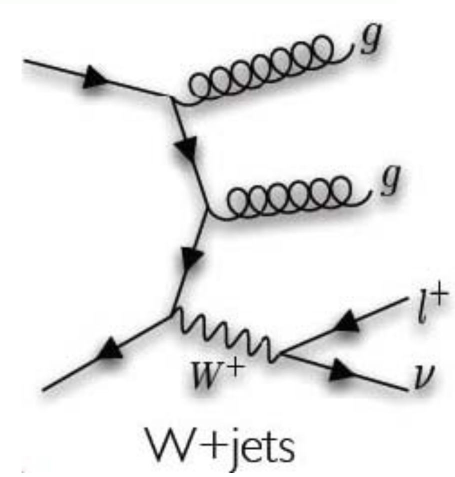
- need to consider final states
- high jet multiplicity
- isolated leptons
- b-tagged jets

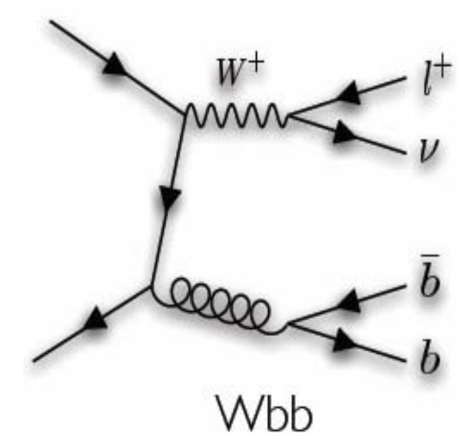
- major backgrounds:
  - W+Jets & W + heavy flavour
  - QCD multijets











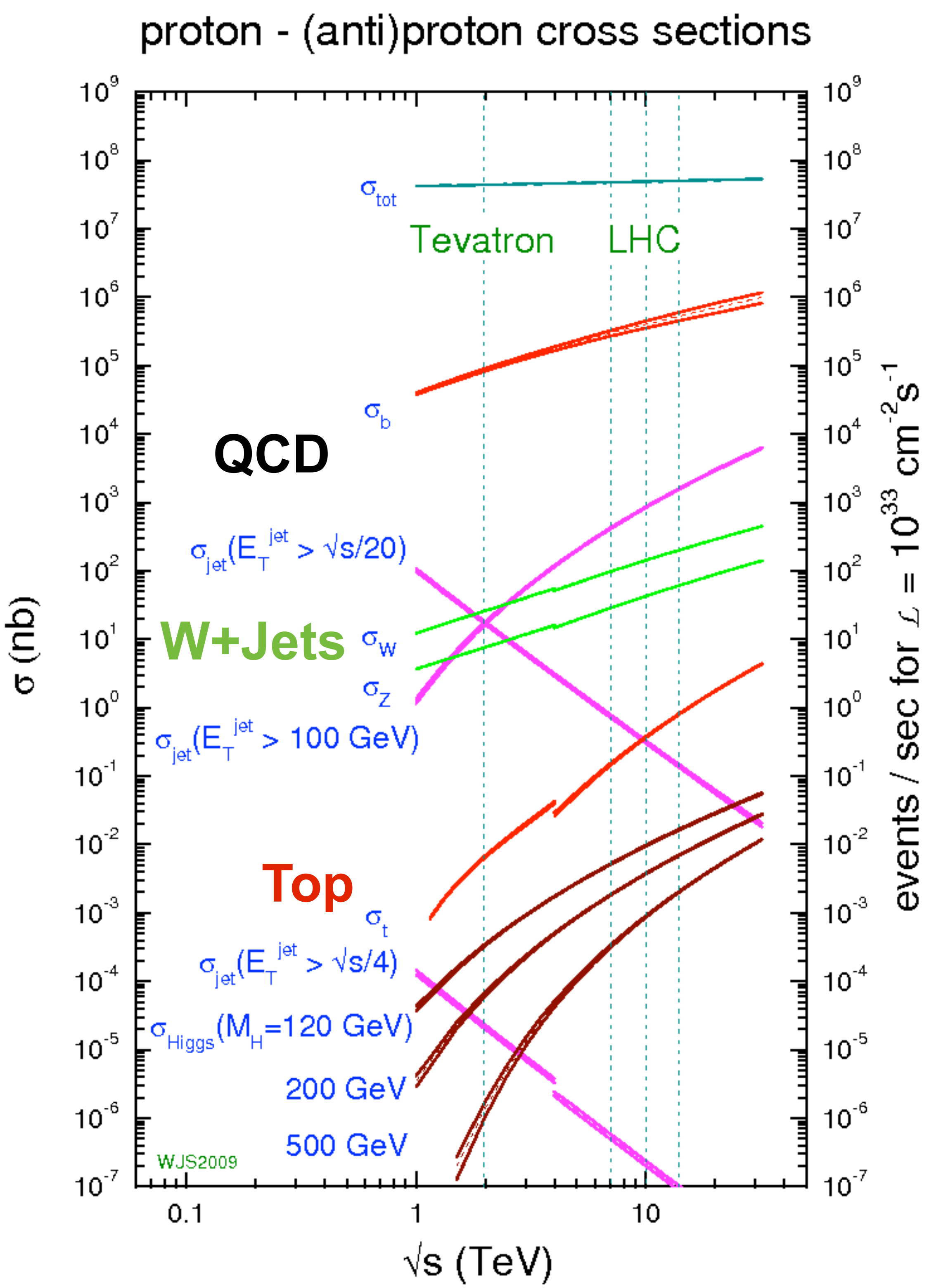
top and background cross-sections

- signal and background cross-sections calculated in theory
- top pair production at 7 TeV and m_t = 172.5 GeV (Moch, Uwer):

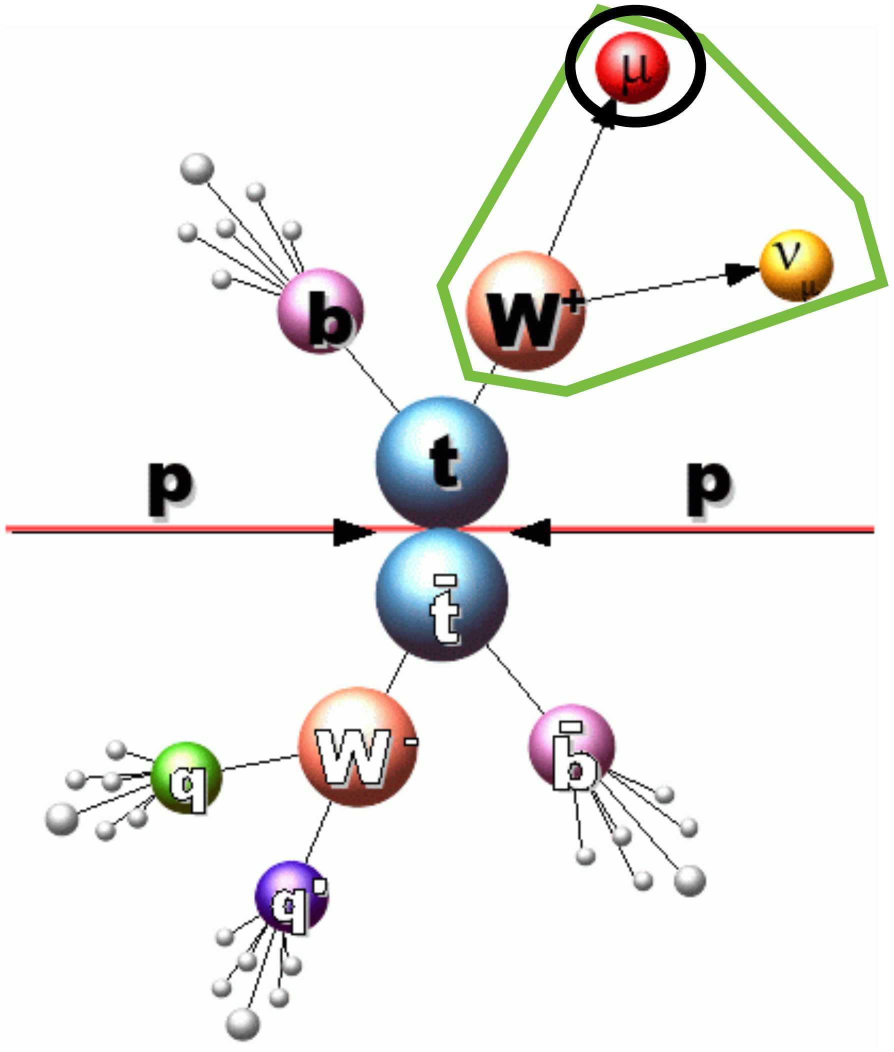
$$\sigma_{t\bar{t}} = 164.57^{+8.34}_{-11.33}\ \text{pb}$$

- W+jets production: σ ~ 30 nb
- QCD production: σ ~ μb/mb
- to calculate cross-section in cut and count experiment, need to evaluate background:

$$\sigma = \frac{N_{\text{tot}} - N_{\text{bg}}}{\epsilon \times \mathcal{L}}$$

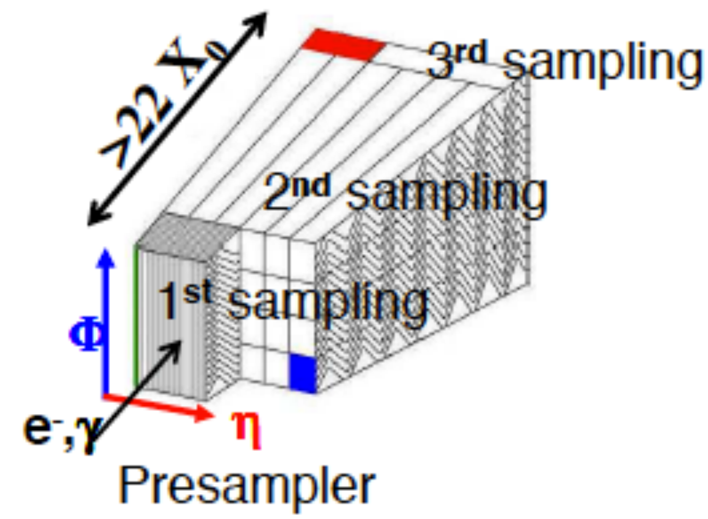


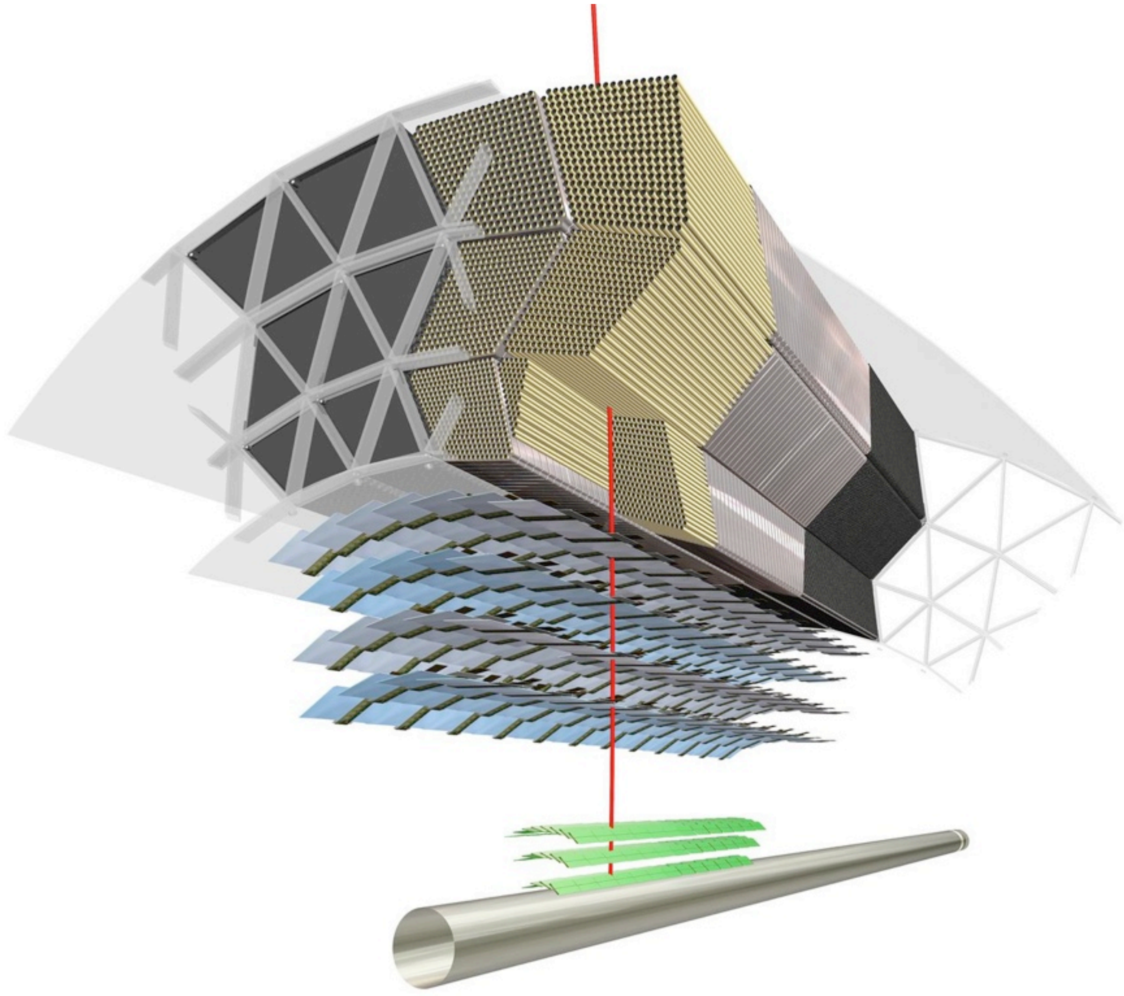
- model all processes in MC
- apply tight selection cuts to reject backgrounds
- cannot reject W+jets background completely due to similar topology ➞ **estimate using data-driven approach**
- QCD contributes only due to mis-identified leptons (electron fake rate of order 10^{-3}-10^{-4})
- QCD MC production limited by computing power
- mostly instrumental background, difficult to model ➞ **take from data**
- my work: **QCD estimation in electron channel**



intermezzo: electron identification

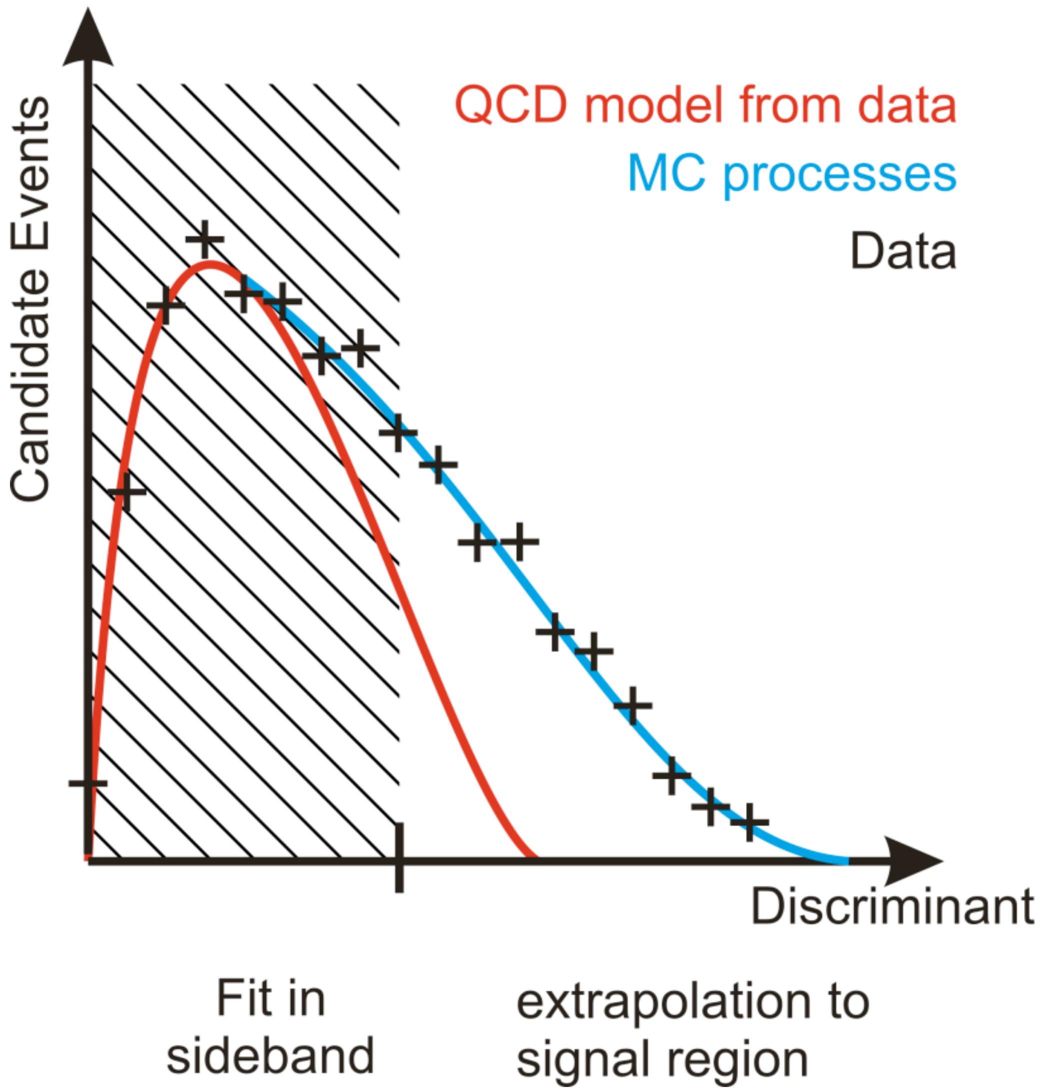
- electron produced at interaction point travel through inner detectors
- track can be reconstructed from hits
- electron is then stopped in electro-magnetic calorimeter
- to obtain high-quality electron candidates we can cut on (particle identification (PID) variables):
  - inner detector track quality
  - track match with calorimeter cluster
  - shower shape
  - leakage into hadronic calorimeter
- also require isolation





modelling QCD from data

- need to find region where QCD dominates while staying as close as possible to signal selection
- select sample orthogonal to standard top selection by inverting cut on electron PID selection ➞ **anti-electron sample**
- find distribution that is sensitive to lepton fakes ➞ **missing transverse energy** (QCD here mostly instrumental background)
- shape of QCD background taken from data, but model provides no cross-section ➞ **determine amount of QCD background from fit**



details on the fitting procedure

- binned Likelihood fit using TMinuit
- allow background contributions to vary from Standard Model expectation within expected uncertainties

Likelihood function

$$-\log\mathcal{L} = \sum_i^{\text{bins}} (\mu_i - N_i^{data}\log\mu_i) + \mathcal{G}$$

$$\mu_i = \sum_k^{\text{processes}} N_i^{\text{MC}} \cdot \beta_k$$

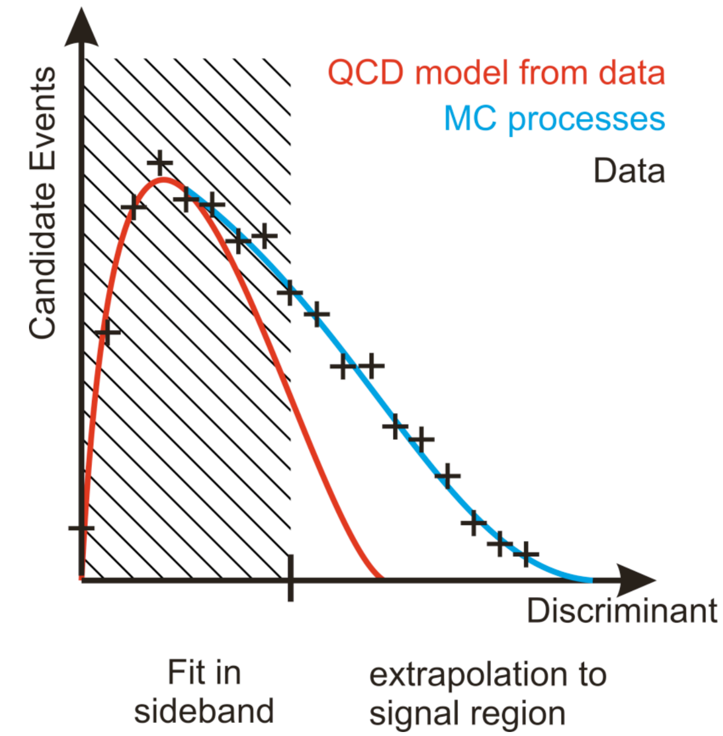
Gaussian constraint

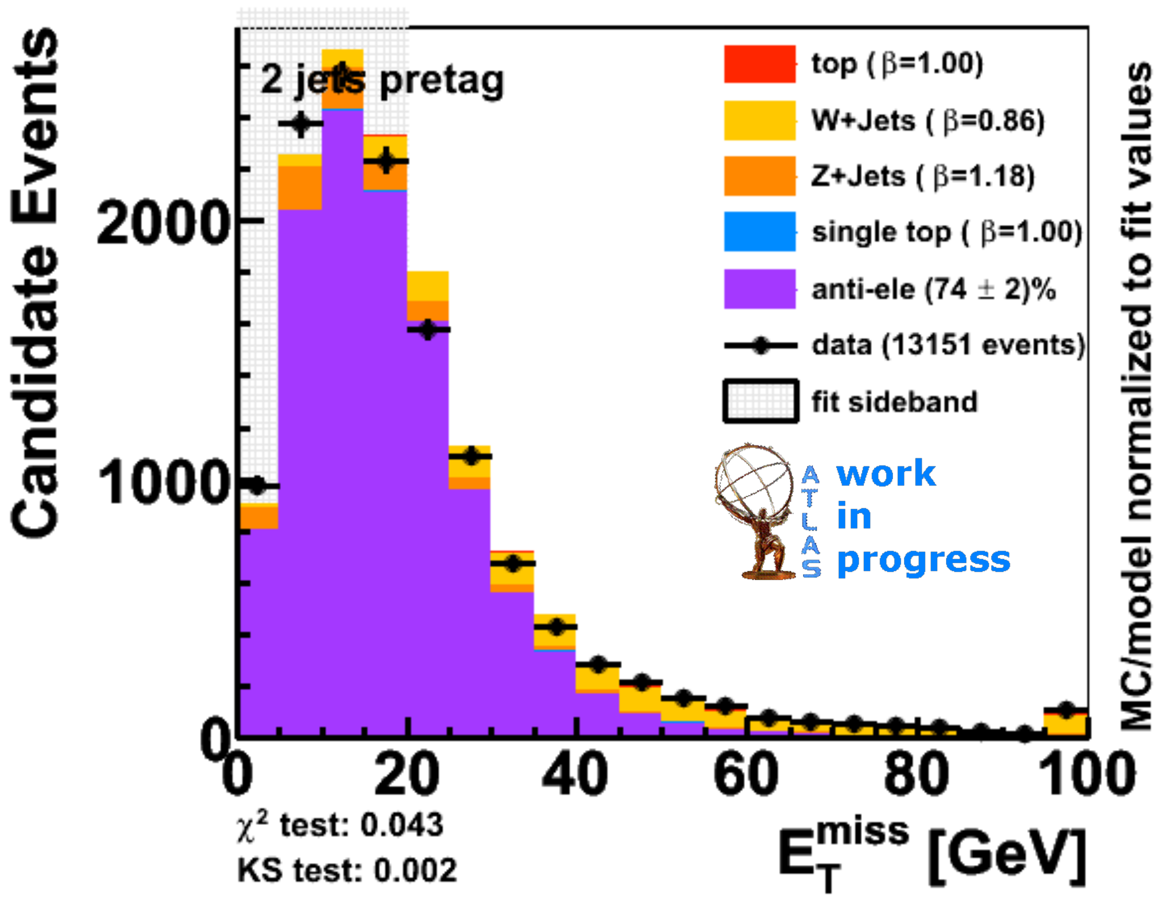
$$\mathcal{G} = -\frac{1}{2}\sum_k^{\text{processes}} \left(\frac{\beta_k - 1}{\sigma_k}\right)^2$$

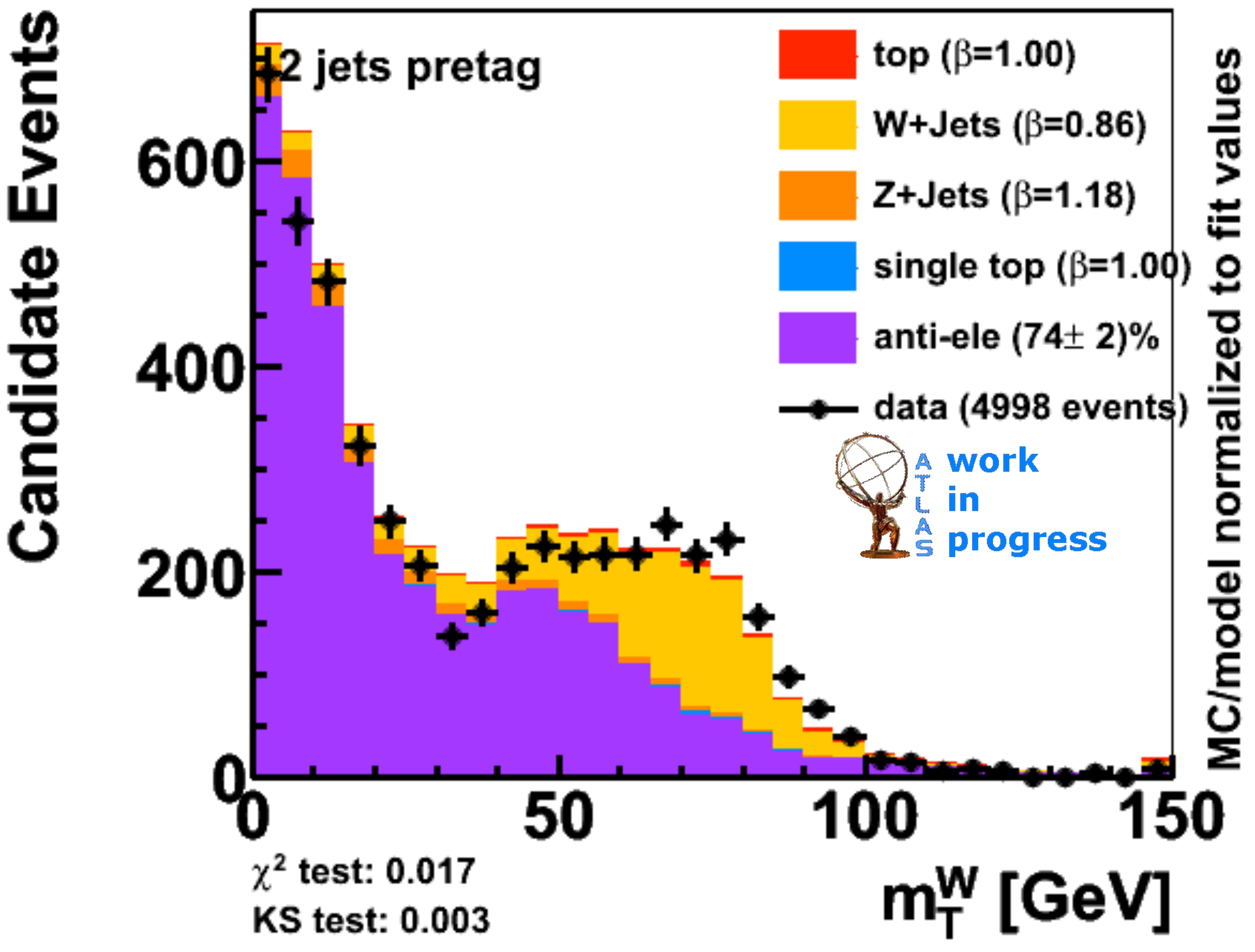
$$\beta_k = \frac{N^{\text{fit}}}{N^{\text{expected}}}$$

anti-electron model

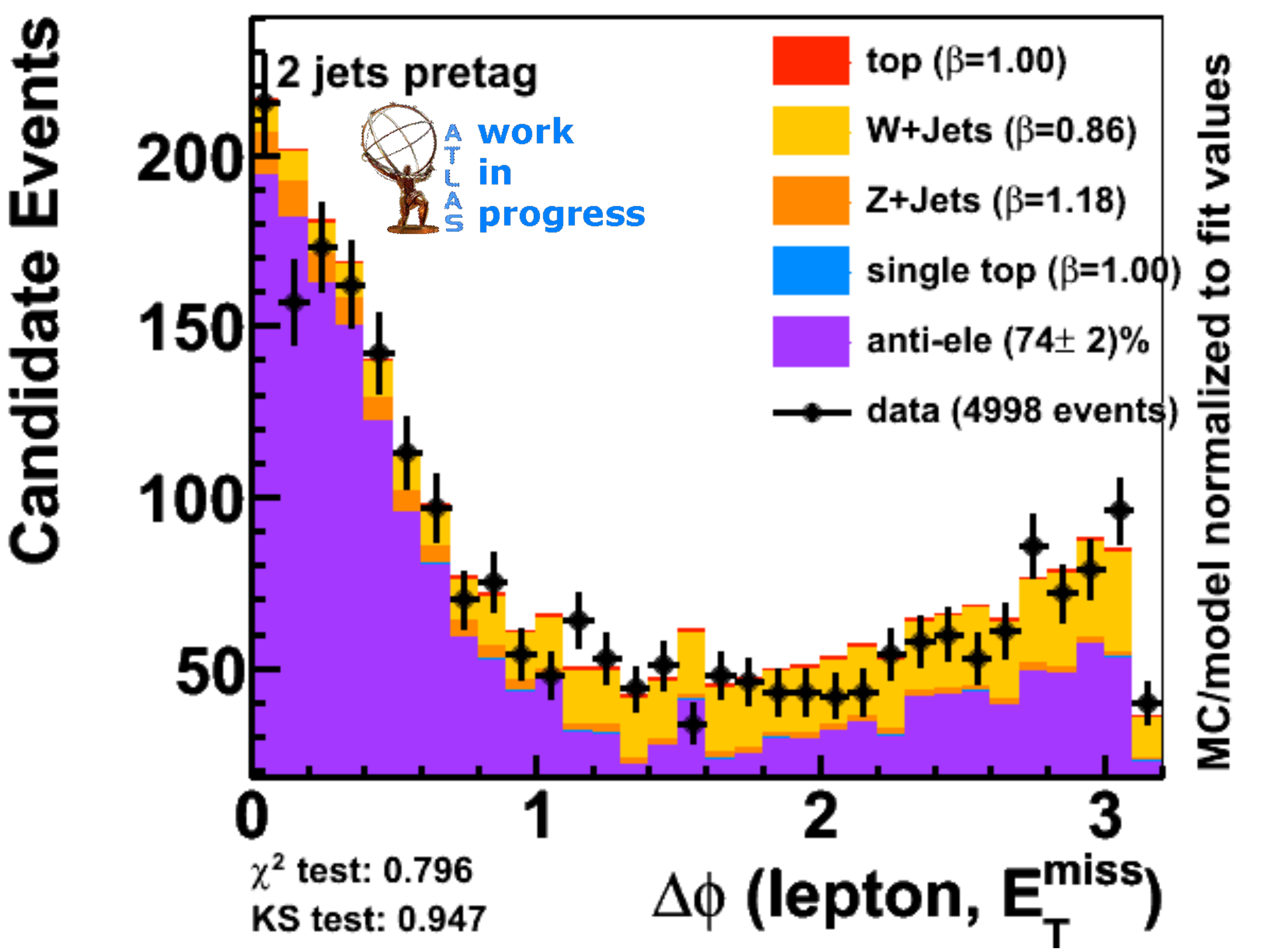
- if we can fit in sideband we can perform another fit (e.g. W+jets background) in the signal region later
- top selection applies MET cut of 20 GeV ➞ **use 0-20 GeV sideband for fit and extrapolate**
- all processes taken from MC and scaled to SM expectation
- **challenge**: tune background model to agree with signal-selected data
- having a full background model should also allow us to look into other kinematic distributions





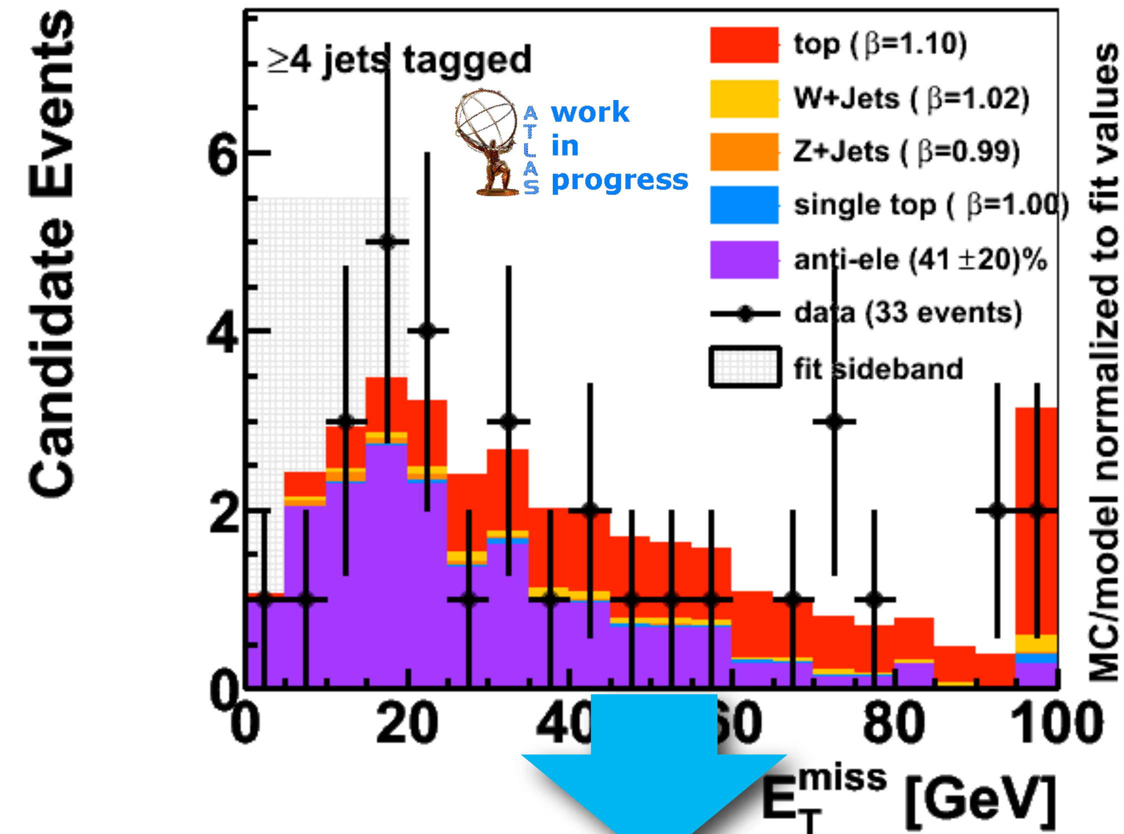


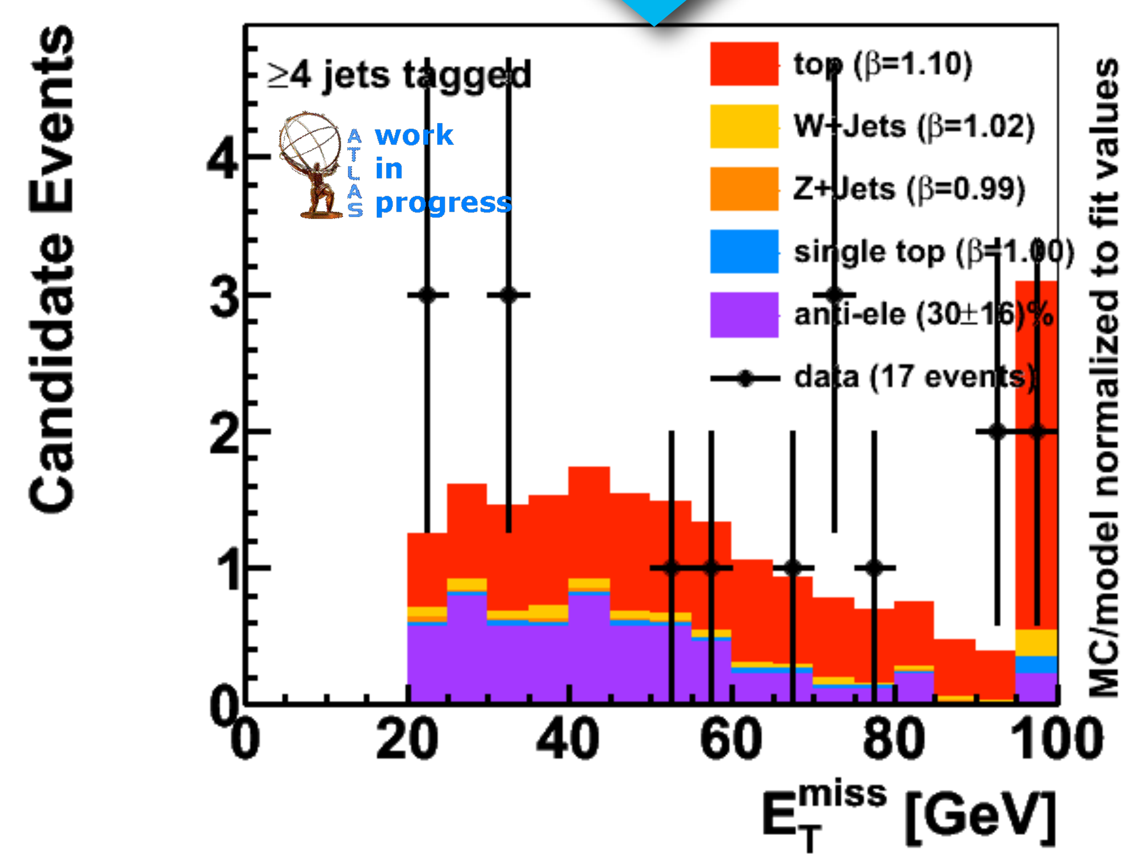
good agreement with data!



top signal region

- estimated QCD fraction very high in previous plots
- need to reduce QCD contribution to „see“ signal
- introduced additional QCD-killer cut
- problem: kills sideband region
- solution: fit before cut is applied, extrapolate and then apply cut (for each process)
- QCD fraction reduced substantially (though very limited statistics here)
- n.b.: there is a similar method using jets for the QCD modell





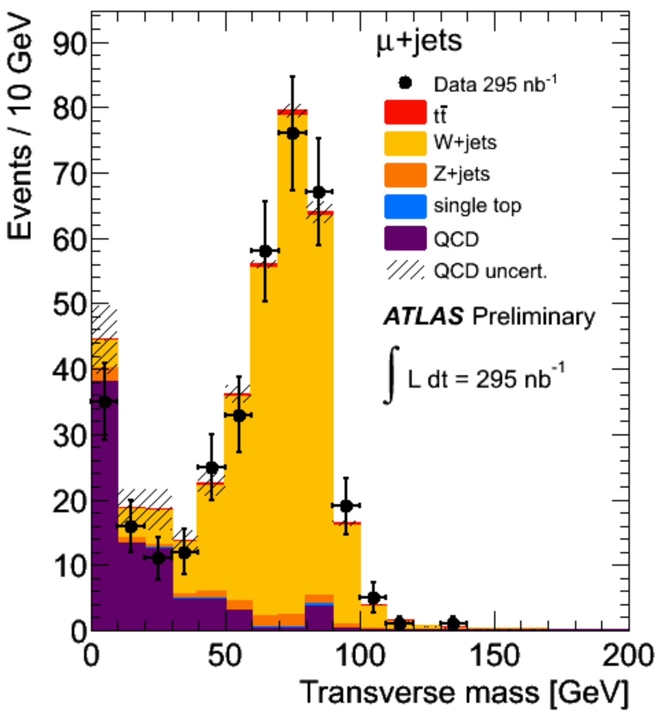
QCD estimation in muon channel

- multijet background predominantly from heavy quark decays (non-prompt leptons)
- top group using matrix method (c.f. ATLAS-CONF-2010-087)
- define „loose“ and „tight“ sample, where „tight“ ∈ „loose“
- measure ε_{fake} in multijet dominated region, ε_{real} from Zμμ events
- two approaches for ε_{fake} estimation:
  - use low MET
  - high muon impact parameter

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

$$
\varepsilon_{\mathrm{real}} = \frac{N^{\mathrm{tight}}_{\mathrm{real}}}{N^{\mathrm{loose}}_{\mathrm{real}}} \qquad \varepsilon_{\mathrm{fake}} = \frac{N^{\mathrm{tight}}_{\mathrm{fake}}}{N^{\mathrm{loose}}_{\mathrm{fake}}}
$$

$$
N^{\mathrm{tight}}_{fake} = \frac{\varepsilon_{\mathrm{fake}}}{\varepsilon_{\mathrm{real}} - \varepsilon_{\mathrm{fake}}} \left(N^{\mathrm{loose}} \varepsilon_{\mathrm{real}} - N^{\mathrm{tight}} \right)
$$

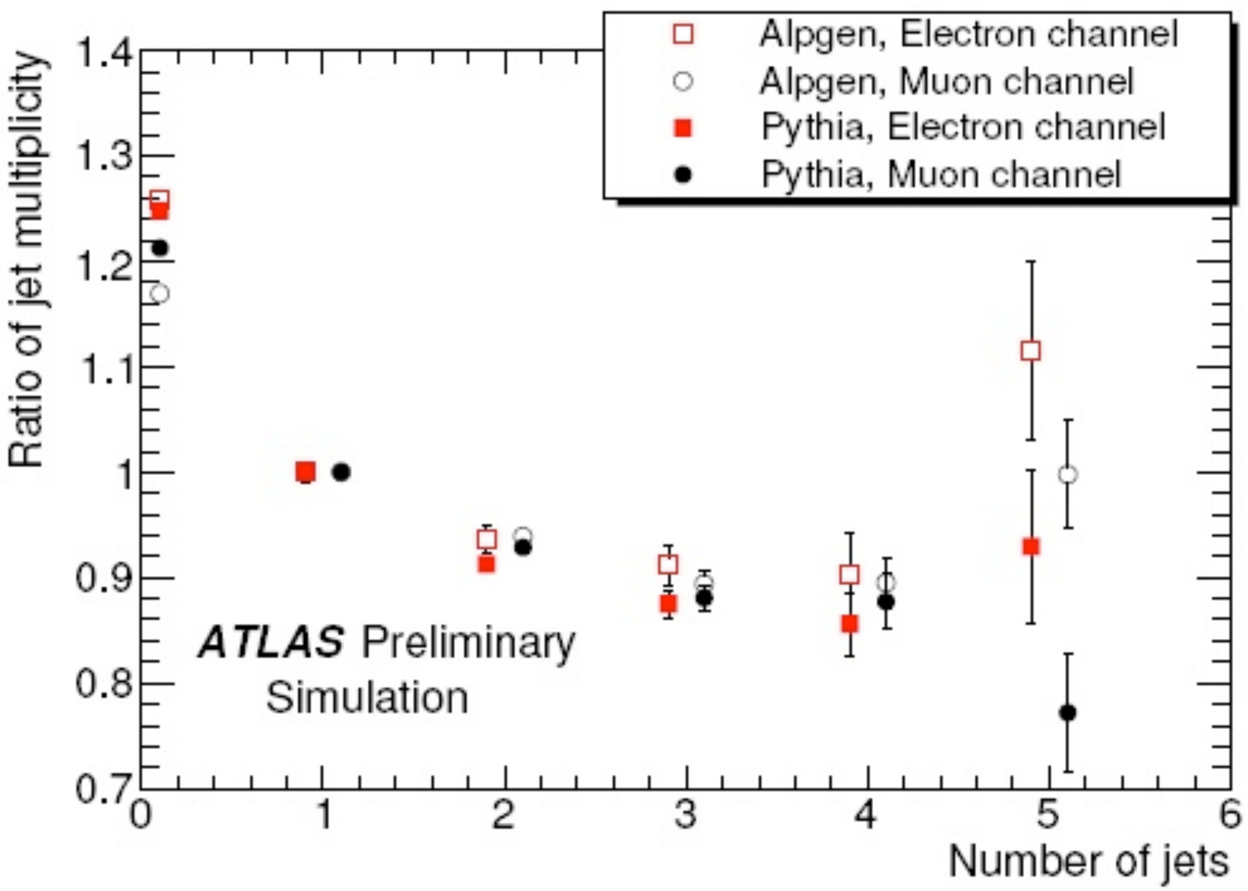


W+jets background estimation (ATL-PHYS-PUB-2009-087)

- W to Z ratio predicted with smaller uncertainty than W+jets MC predictions
- extrapolate from low jet multiplicity control region (CR) to top signal region (SR)
- caveat: QCD contribution in W+jets CR significant - needs to be subtracted
- different approach now used

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

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



summary

- top pair cross-section measurement challenging
- anti-electron model allows to model QCD background from data
- matrix method used for muon channel
- W+jets background also obtained from data

- all ingredients available for an ATLAS top pair cross-section measurement