Data Driven Background Estimation for H^+ in ATLAS

Yoram Rozen

Technion - Israel Inst. Of Technology

On behalf of the ATLAS collaboration

Direct search via H->$\tau\nu$, H->$c\bar{s}$

- Fake τ
- True τ (Embedding)
- Multi-jet background
- Matrix method

Technion Israel Institute of Technology

$H^+ \to \tau^+\nu_\tau$

- Three channels:
  - τ->hadrons:
    - $t\bar{t} \to WbH\bar{b} \to q\bar{q}bH\bar{b}$
    - $t\bar{t} \to WbH\bar{b} \to l\nu bH\bar{b}$

    Patrick talk earlier today
  - τ->lν:
    - $t\bar{t} \to WbH\bar{b} \to q\bar{q}bH\bar{b}$ Catrin talk morning session
- Many common backgrounds
- Heavy use of data to estimate the background
- ATLAS publications with 37 pb^{-1} and with 4.7fb^{-1}
  - JHEP 1206 (2012) 39; arXiv:1204.2760; CERN-PH-EP-2012-083 (4.7 fb^{-1})
  - ATLAS-CONF-2011-051 (37 pb^{-1})
  - ATLAS-CONF-2012-011 (4.7 fb^{-1})

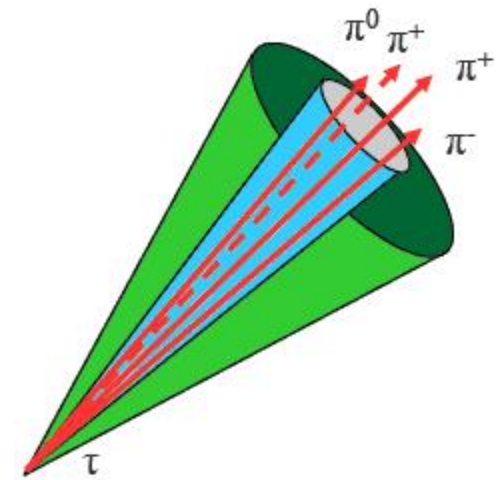
$H^+ \to c\bar{s}$

- $t\bar{t} \to WbH\bar{b} \to l\nu b\bar{b}c\bar{s}$ Catrin talk
- ATLAS-CONF-2011-094 (35 pb^{-1})

Fake rate

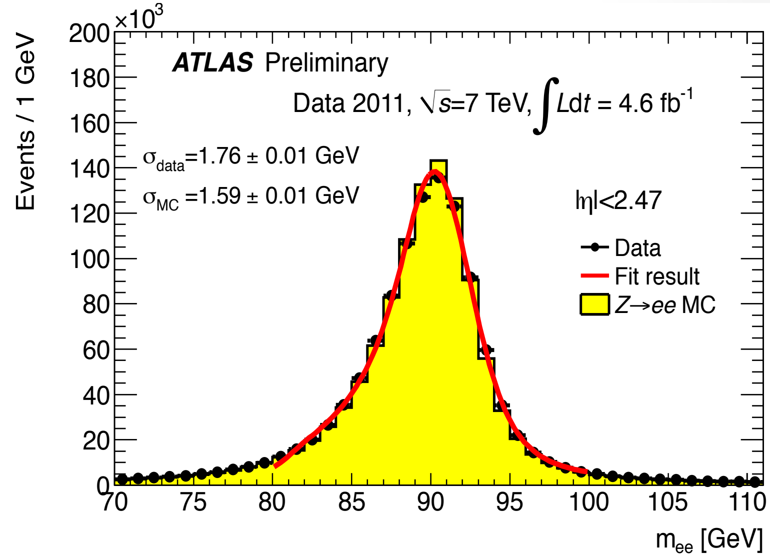
Deals with: e, jet ->τ

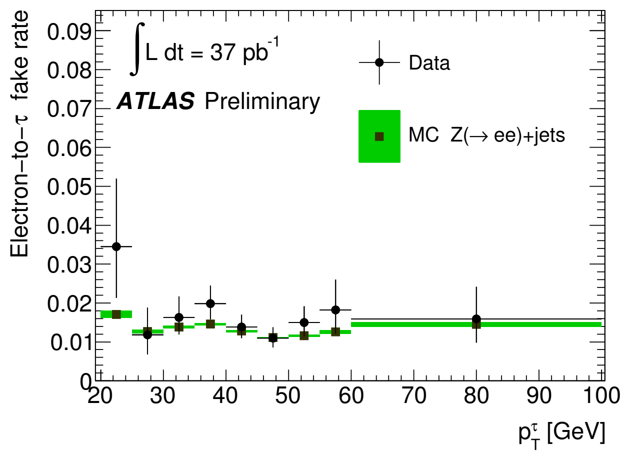
- An electron or a jet identified as a τ they are dubbed "fake"
- Method:
  - Find the fake rate defined as (#of fakes)/(total # τ candidates)
  - Sum the # of objects (w/ the appropriate selection) multiplied by the above fake rate.
- Different application for each object.
- Done in bins of p_T (and η).

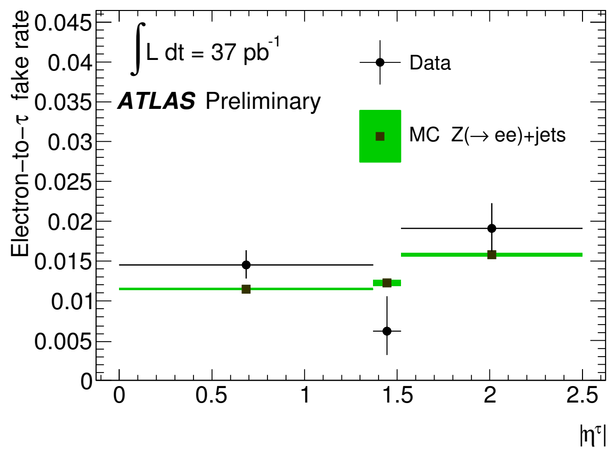


e → τ fakes

- "Tag and probe" method
  - Use clean Z->e⁺e⁻ signal
  - One tight electron to "tag"
  - Other electron to probe the probability to be identified as a τ-jet



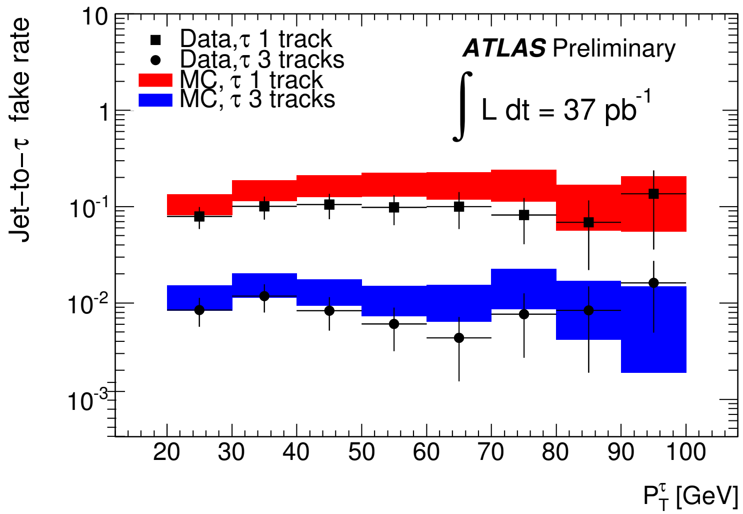




- Similar method for μ but with 2-3 orders smaller -> negligible

Jets → τ fakes (1) (37pb^{-1} publication only)

- γ+jets event are used. Identified by the γ trigger.
- Binned by the number of tracks in the jet and p_T.



- Systematics include:
  - Contamination (real τ) from processes like QCD and Z,W
  - Control sample uncertainty and correlation to other methods.

Jets→τ fakes (2) – Current method

- Non-τ jets are used by selecting a W+jets sample:
  - b-jet veto to reject tt events.
  - Leptonic W.
- Miss-identification probability is measured in bins of p_T and η.

  Typical values: 7% (1 prong) and 2% (3 prong)
- Applying the probabilities to the number of jets in the final sample (after removing b-tagged jets).

Things to consider (systematics):

- Object related
- q/g ratio difference between the target tt and the W+jets.
- True taus in the sample.
- Control sample size.

Embedding:

Deals with: true τ background (hadronic)

- $t\bar{t}$ -> Xμ is similar to $t\bar{t}$->Xτ in everything but μ .vs. τ
- If the μ is replaced by a τ we have a guaranteed background environment with everything but the lepton having data characteristics. (also used in H->ττ)

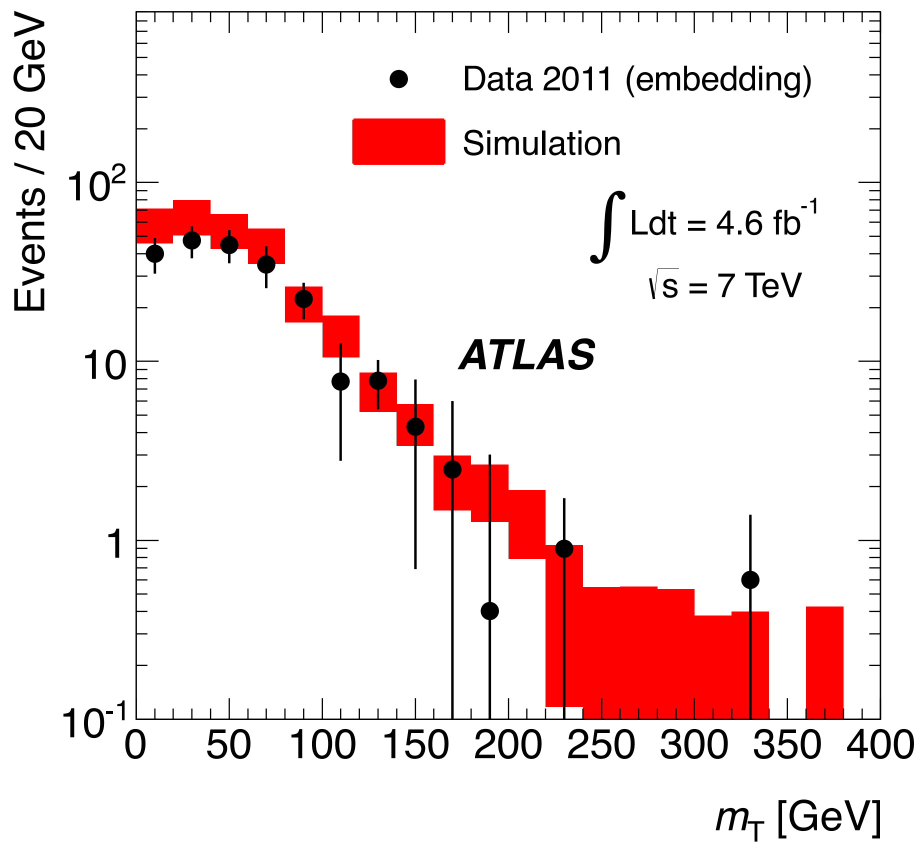
Procedure:

- Selected $t\bar{t}$ evens containing a μ (from the decaying W)
- Scale the μ momentum to compensate for the mass difference
- Simulate a τ with the scaled μ 4-momenta
- Embed the simulated τ back in the data event
- **Run the H^+ analysis to get the normalization of the tt background:**

$$N^{tt-bkg} = N_{sel}^{EMB}(1 - f_{W\to\tau\to\mu})\frac{\varepsilon_{trig}^{\tau}}{\varepsilon_{sel}^{\mu}}B(\tau \to had)$$

- Since embedding uses $t\bar{t}$ events and universality predicts exactly the number of τs, an absolute bkg prediction can be made.
- m_T shape is the alternative used in the first (37pb^{-1}) pub. to normalize the bkg in a $t\bar{t}$ dominated region

For the τ+jets channel



Embedding (cont.)

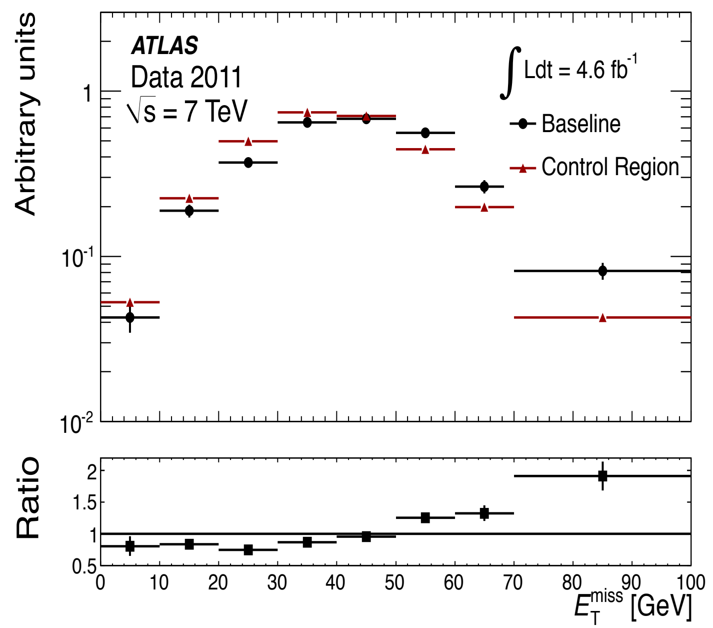
Systematics:

- Control sample uncertainty and statistics
- τ simulation (identification, energy scale)
- μ isolation
- Embedding parameters
- m_T shape due to τ energy scale

Control region (inverted selection)

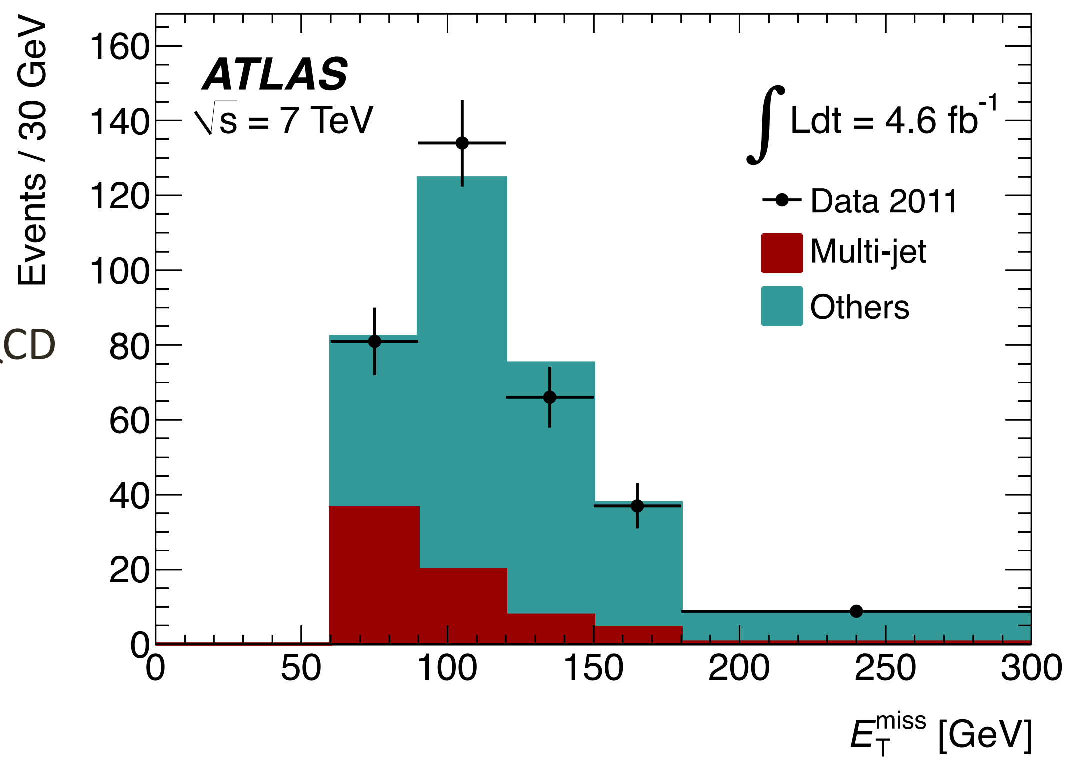
Deals with: multi-jet background

- Template method fitting the E_T^{miss} shape
- Starting with loose τ selection but rejecting the tight selection and a b-veto-> a sample of non-selected events with similar characteristics is obtained.
- E_T^{miss} shape must be similar to the baseline shape (checked at an early selection stage)



9-10-2012

- E_T^{miss} shape of other SM and the QCD is fitted to the data.
- Overall normalization and QCD fraction are the only fitted parameters.



Systematics:

- Fit procedure (range, binning)
- $t\bar{t}$ and W+jets shape and their relative norm.
- Sample size

Matrix method

Deals with:non-prompt muons (only H->c$\bar{s}$)

- Published result with $35pb^{-1}$
- Multi-jet background in the μ channel is estimated with "matrix method"

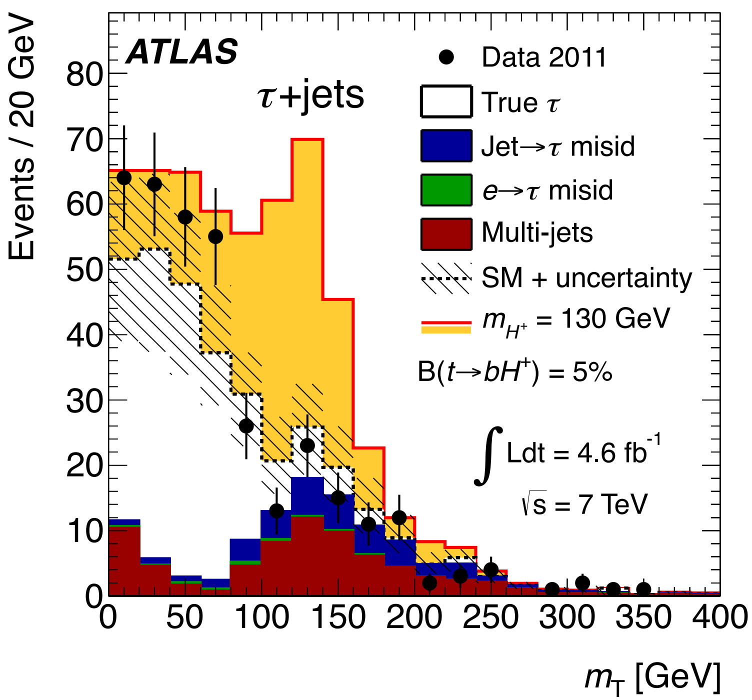
$$N_{loose} = N_{loose}^{real} + N_{loose}^{fake}$$

$$N_{std} = rN_{loose}^{real} + fN_{loose}^{fake}$$

  - r is estimated from Z->μμ
  - f is estimated from 2 control regions design to avoid prompt muons
    - Low E_T^{miss} for QCD source
    - High E_T^{miss} but high impact parameter muons
- For the electron channel a likelihood template fit of E_T^{miss} is used

Background Summary

τ+jets channel



Sample	Event yield (τ+jets)
True τ (embedding method)	210 ± 10 ± 44
Misidentified jet→ τ	36 ± 6 ± 10
Misidentified $e \to \tau$	3 ± 1 ± 1
Multi-jet processes	74 ± 3 ± 47
∑ SM	330 ± 12 ± 65
Data	355
$t \to bH^+$ (130 GeV)	220 ± 6 ± 56
Signal+background	540 ± 13 ± 85

9-10-2012

Summary

- Background is no longer divided into physics sources which depend on XS for estimation
- Smaller dependence on simulation

  -> Reduction in systematic uncertainties

- Background sources are divided into the analysis objects
- All background sources can be estimated in a data driven way.
- Some of the methods are applicable for other searches as well.