τ-ID from D0 to ATLAS

Volker Büscher,
Michael Heldmann,
Ingo Torchiani

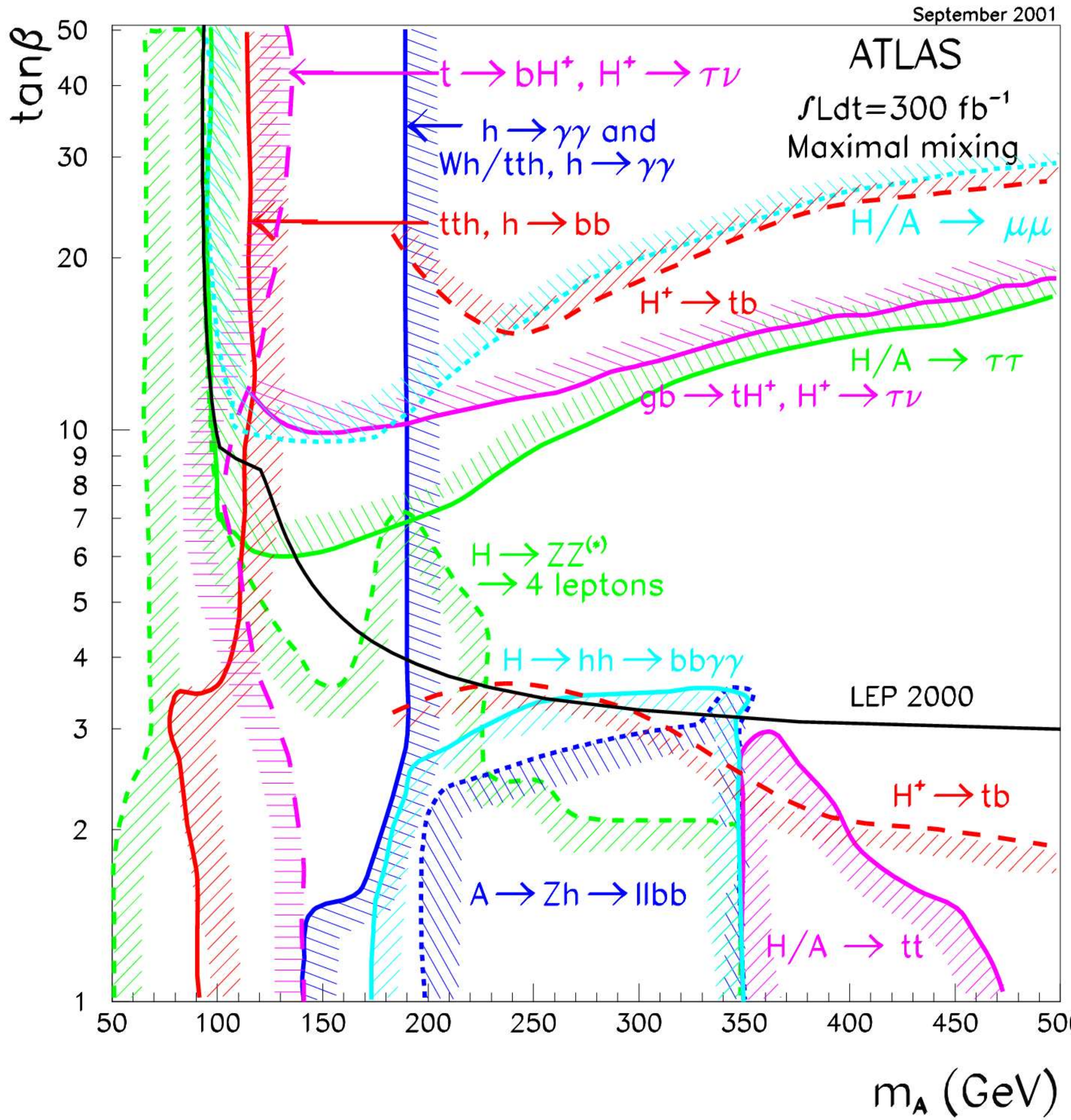
University of Freiburg

- Motivation
- D0 ATLAS
- preliminary comparisons
- preliminary conclusions
- summary

A big part of the work (all D0 related work) presented was actually done by Ingo

Motivation

- **At ATLAS we expect a big number of final states involving taus**
- **Channels using taus**
  - $A^0/H^0 \to \tau\ \tau$
  - $H^+ \to \tau\ \nu$
  - SUSY with production of $\tilde{\tau} \to \tau + \chi^0{}_1$
  - Standardmodell Higgs (VBF qq H → qq $\tau\ \tau$)
  - $Z \to \tau\ \tau$ (for comissioning)
  - τ are perhaps the only way to access the chiral structure of SUSY
- →τ's are an important signature



Motivation

- Since this is the TeV4LHC workshop, the questions are:
  - What can ATLAS learn from D0 about tau reconstruction and identification ?
  - How can we transfer this knowledge to ATLAS ?
- The steps we would like to follow are:
  - compare the D0 algorithm to what we use at ATLAS
  - look for input on how we can improve our algorithm
- many ATLAS Analysis rely on the understanding of τ identification
  - will we reach the performance we see on the MC at the moment ?
  - → learn from the D0 comparison between MC and data
- check if the description of MC-Generators of the low multiplicity jets is correct with D0 data
- get input on how to measure the performance using data

Tau Identificaton

- **How can one identify τ-leptons ?**
- most important decay modes
  - Leptonical decay modes
    - $\tau \rightarrow \nu_\tau + \nu_e + e$ (17.4%)
    - $\tau \rightarrow \nu_\tau + \nu_\mu + \mu$ (17.8%)

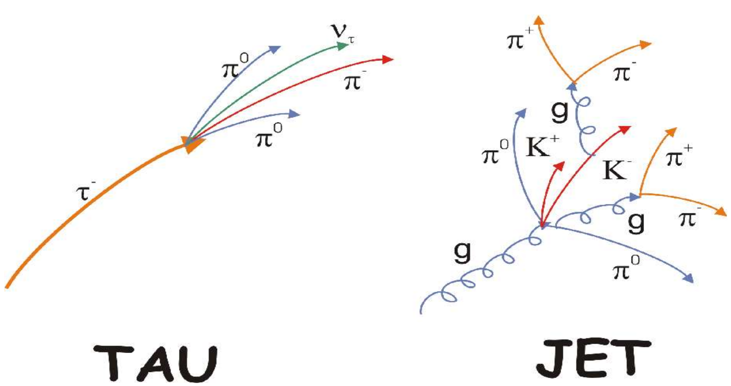
1 track
only difference
from prompt leptons:
impact parameter

  - Hadronical decay modes
    - 1 prong
      - $\tau \rightarrow \nu_\tau + \pi^c$ (11.0%)
      - $\tau \rightarrow \nu_\tau + \pi^c + \times \pi^0$ (36.2%)
    - 3 prong
      - $\tau \rightarrow \nu_\tau + 3 \bullet \pi^c + \times \pi^0$ (15.2%)

1 track, impact parameter
shower shape, energy sharing
find the photon cluster

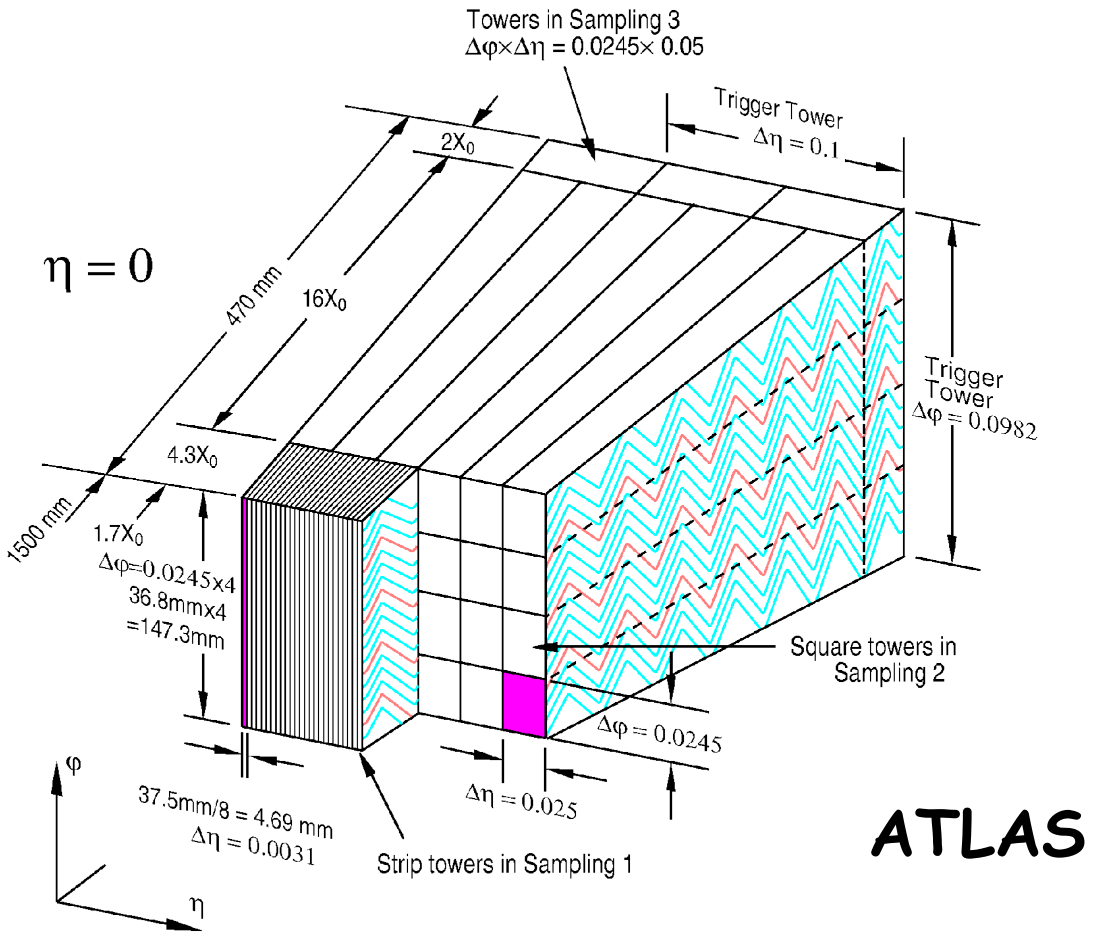
3 track, impact parameters,
secondary vertex

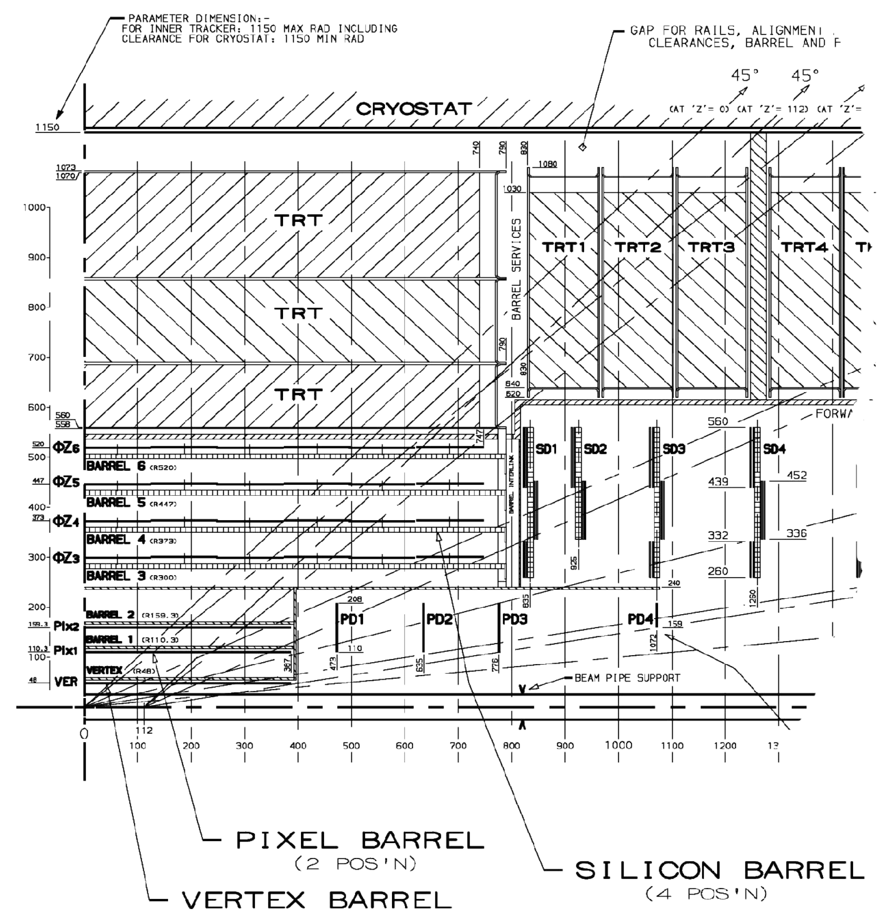
- **$\rightarrow$ τs are colimated calorimeter objects with one or three associated tracks**



ATLAS Calorimeter and ID

- The τ identification makes use of tracks and calorimeter objects
- Atlas has a presampler (0.025x0.1), η-strip-layer (0.003x0.1), middle (0.025x0.025) , back layer (0.05x0.025) and three hadronical layers with 0.1x0.1 and 0.2x0.1
- The ID has three pixel layers, four stereo microstrip layers and a straw tube tracker

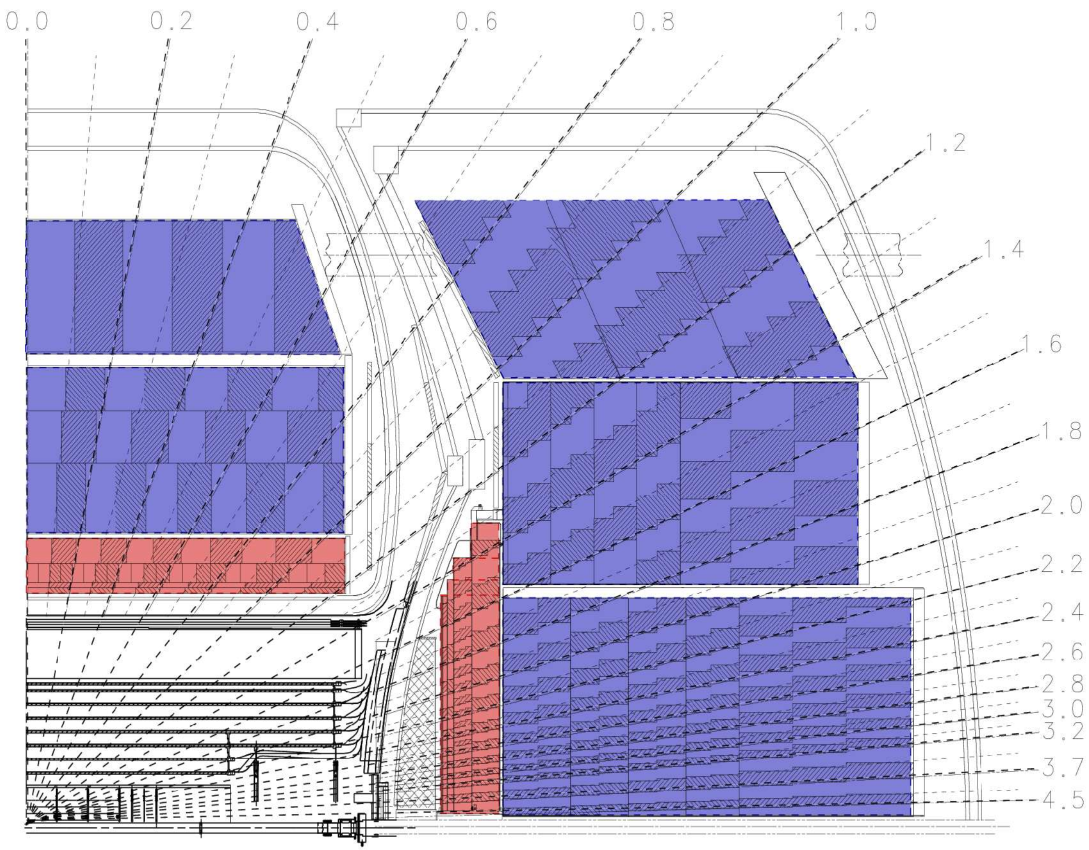


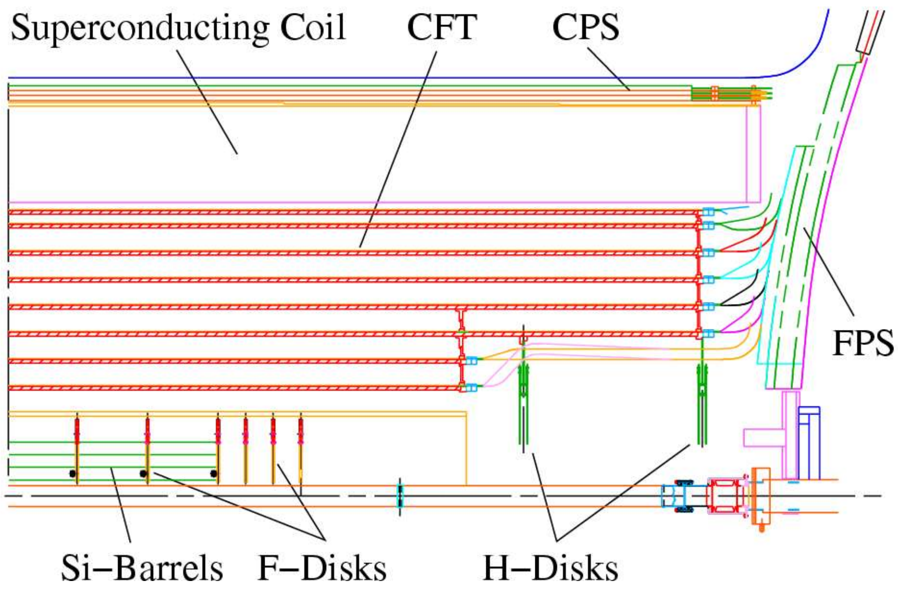


D0 Calorimeter and ID

- D0 has a an EM with four layers of 0.1x0.1, 0.1x0.1, 0.05x0.05, 0.1x0.1 and four hadronical layers 0.1x0.1
- The ID consists of a silicon tracker with four stereo layers and eight stereo layers for the fiber tracker
- the ID of D0 covers $|\eta| < 3.0$!

D0





Strategy for TeV4LHC (1)

- The goal is to understand what difference we can expect from the results in MC to performance with real data
  - We think we can establish a chain of understanding
  - ATLAS Algorithm on ATLAS MC → D0 Algorithm on ATLAS MC → D0 Algorithm on D0 MC → D0 Algorithm on D0 data
- **Steps on the way to establish the chain**
  - select signal sample in D0 data
  - select background sample in D0 data
    - study preselection (called reconstruction step in ATLAS)
    - study TauID in D0 with few key variables
    - study full TauID in D0 using the ANN

If all steps show agreement between data and MC what does that mean for ATLAS ?

Strategy for TeV4LHC (2)

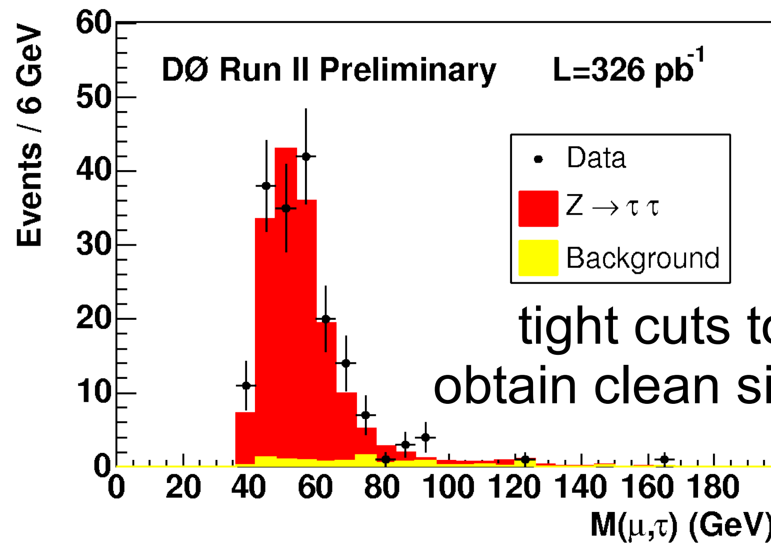
- if all steps show agreement between data and MC what does that mean for ATLAS ?
- **steps to show that results are transferable to ATLAS**
  - implemented all variables D0 uses at ATLAS
  - produce a comparable sample for signal and background
  - compare few key input variables
  - establish a TauID using these few D0 input variables
  - establish a TauID using the counterpart ATLAS variables
  - show that both TauID select similar areas in phasespace
  - show that these variables take into account the gross performance

Strategy for TeV4LHC (3)

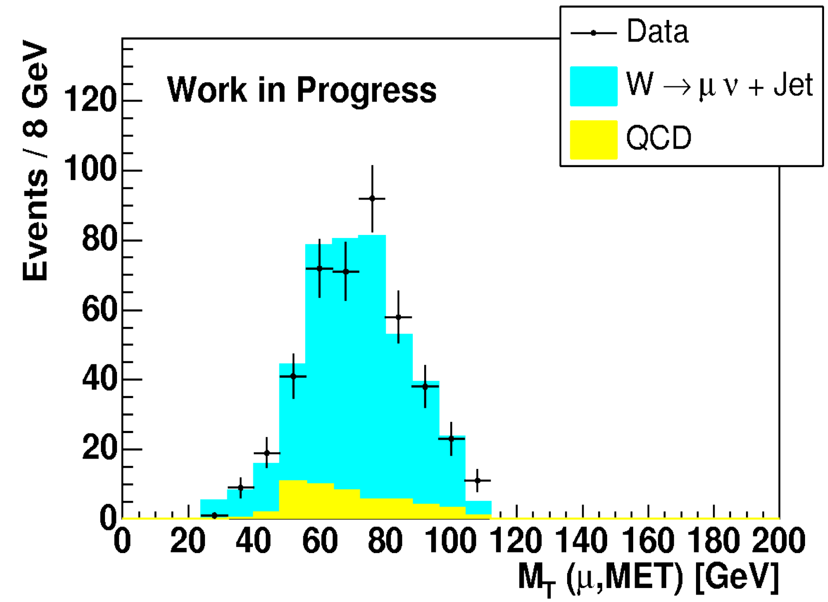
- Of course because of differences in the detector design the "translation" of variables is not uniquely defined
- our convention:
  - D0 EM3 (finely granulated layer in the EM) → ATLAS EM2
  - D0 EM1, EM2 → ATLAS η-strip layer
  - tower granularity in both cases 0.1x0.1
- energy thresholds have been adjusted to match different cell sizes, noise levels and so on ...

$W \rightarrow \mu\ \nu$ selection in D0 data

- **D0: signal sample is $Z \rightarrow \tau\ \tau$, background is $W \rightarrow \mu\ \nu$**
  - for the background only 100pb^{-1} used up to now, x4 available
  - $W \rightarrow \mu\ \nu$ is used because
    - it is a relevant background for many studies
    - it allows to obtain unbiased jets (using a single μ trigger) down to rather low p_T
  - **preselection**
    - $p_T(\mu)$ > 25 GeV, |eta(μ)| < 1.5
    - p_T(jet) > 15 GeV, |eta(jet)| < 1.5
    - MET > 20 GeV, m_T > 30 GeV
    - m(μ, track) < 60 GeV
    - $\Delta\varphi$(MET, jet) > 0.4



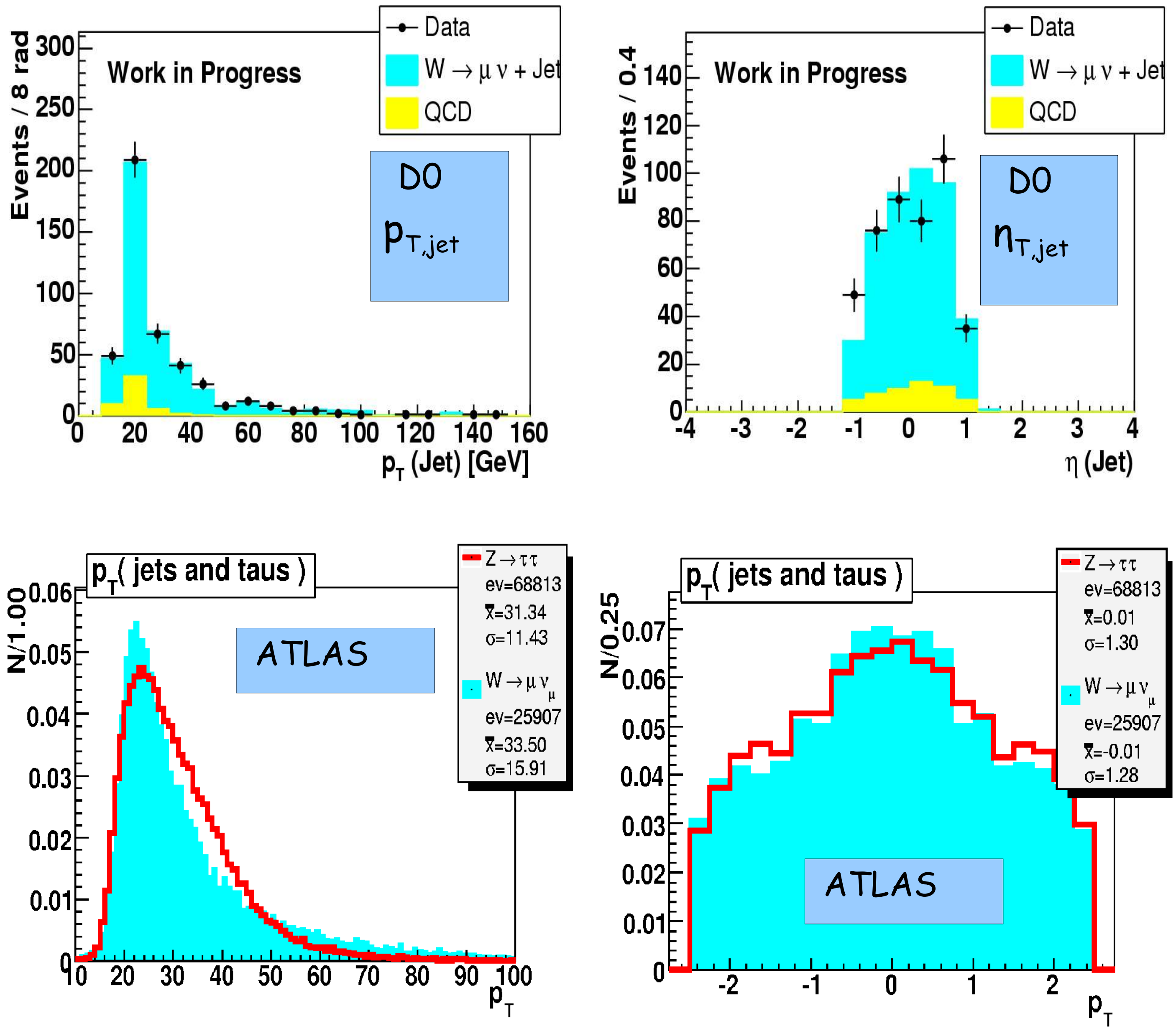
tight cuts to obtain clean signal



- ATLAS: signal sample is $Z \rightarrow \tau\ \tau$, background is $W \rightarrow \mu\ \nu$
  - same preselection, sample was produced specifically for TeV4LHC, using PYTHIA inclusive W like D0 (filtered in MC truth level for ATLAS to reduce CPU-time)

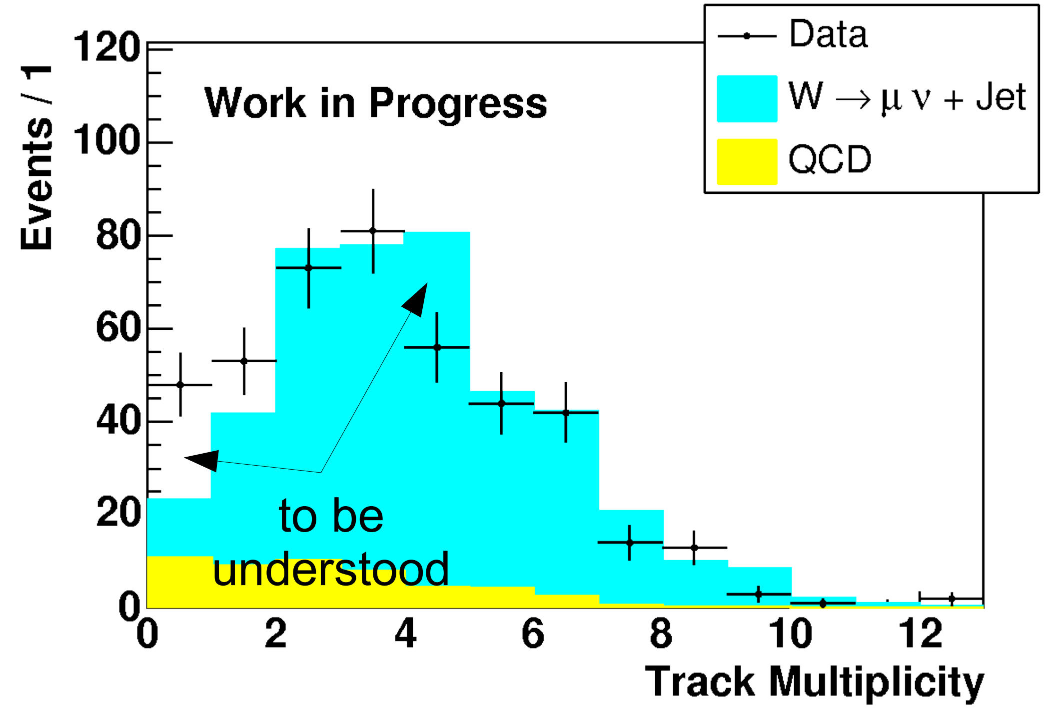
$W \rightarrow \mu\ \nu$ jets

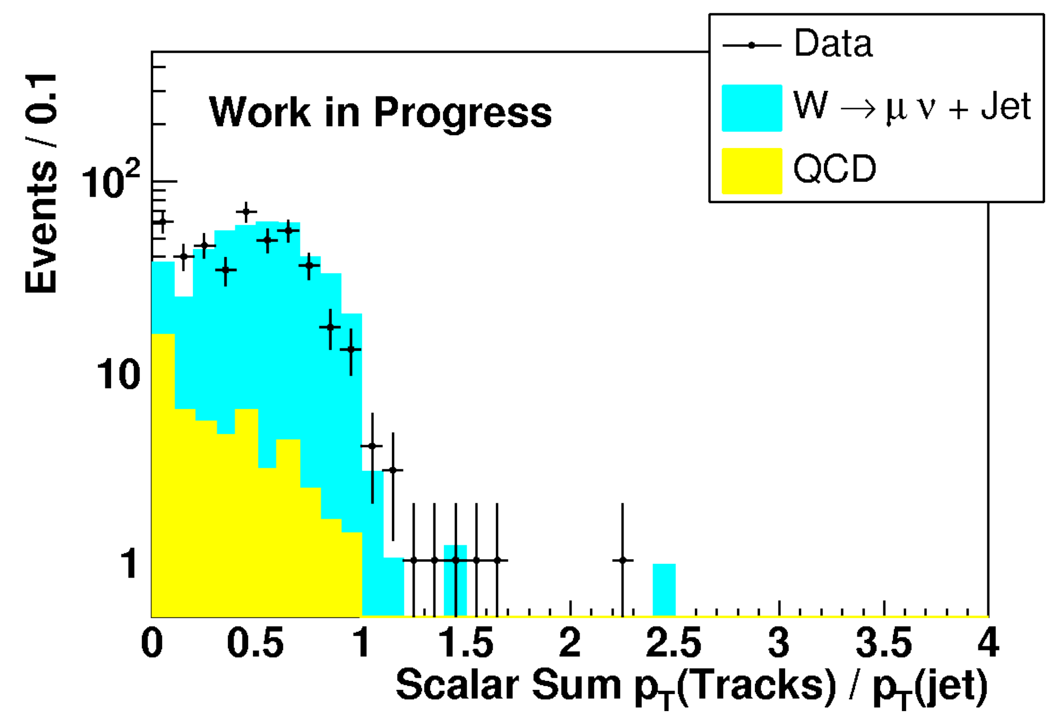
- **for this study we are interested in the jets**
- **the p_T spectrum is very important since all shape properties depend very much on p_T**
- **resonable overlap between D0 and ATLAS can observed**
- **➡ we are not facing very different kinematics here**
- **➡ jets should possess similar shape properties**



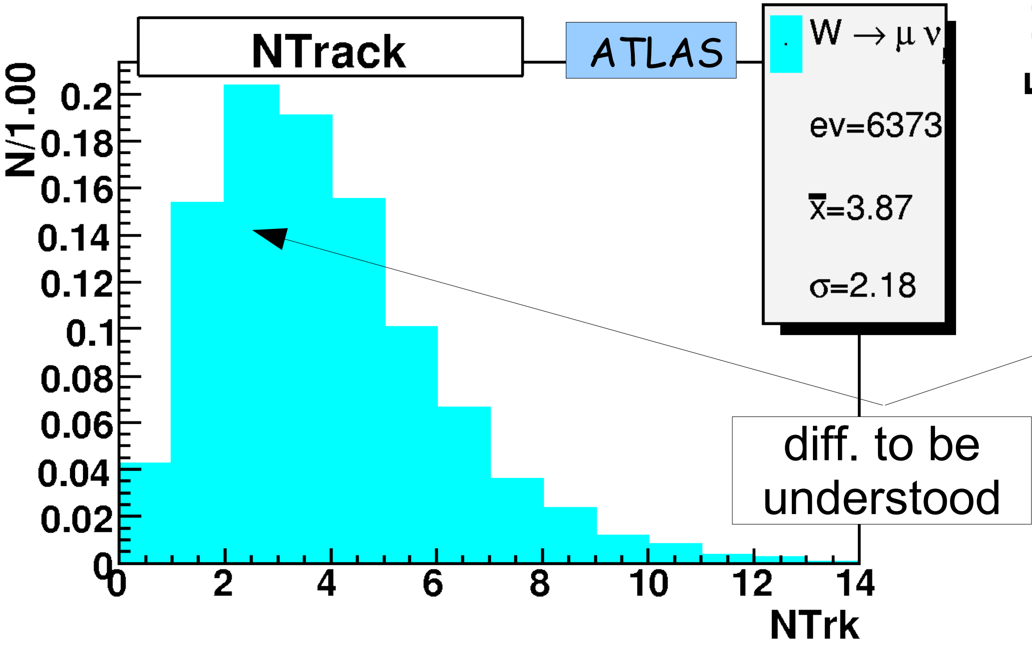
Preselection

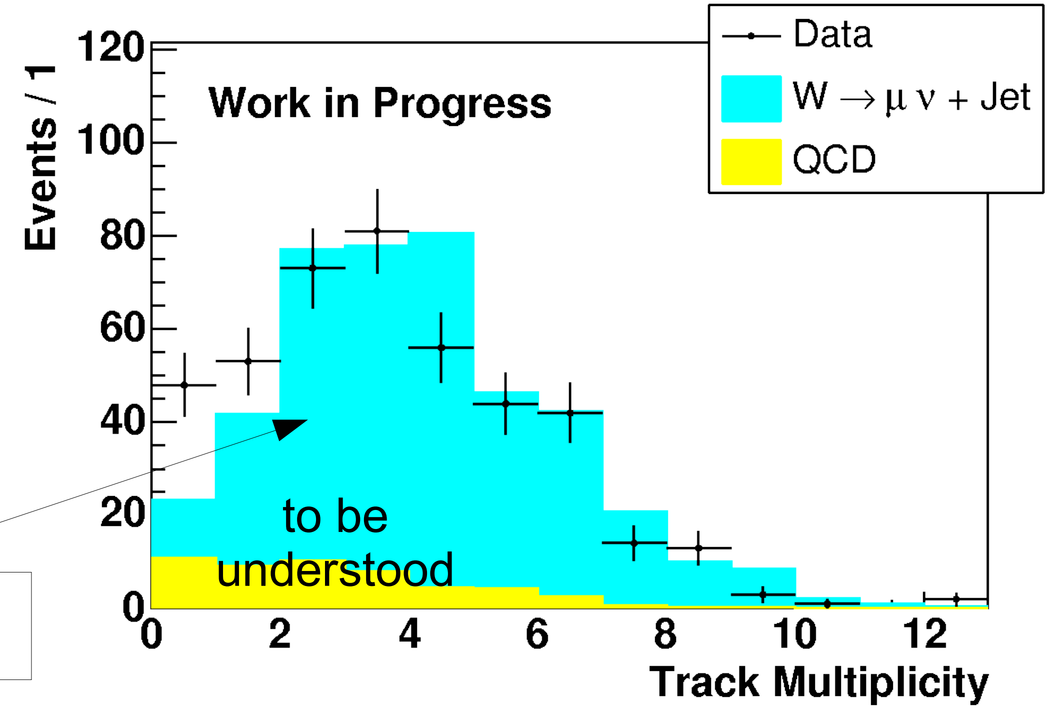
- Preselection means the step from jet ➞ τ-candidate
- D0 and ATLAS use both a seperate cluster-finder for the τs
- at least 1 track is required
- **question 1: are low multiplicity jets well described in the MC ?**
  - question 1a: is the 0-3-track fraction correctly described in the MC ?
  - question 1b: is the charged track multiplicity correctly described ?
- selecting only clusters with > 0 track gives the biggest contribution to the preselection efficiency
- low multiplicity region is not perfectly modellled ➞ correction factors in D0 tau analyses

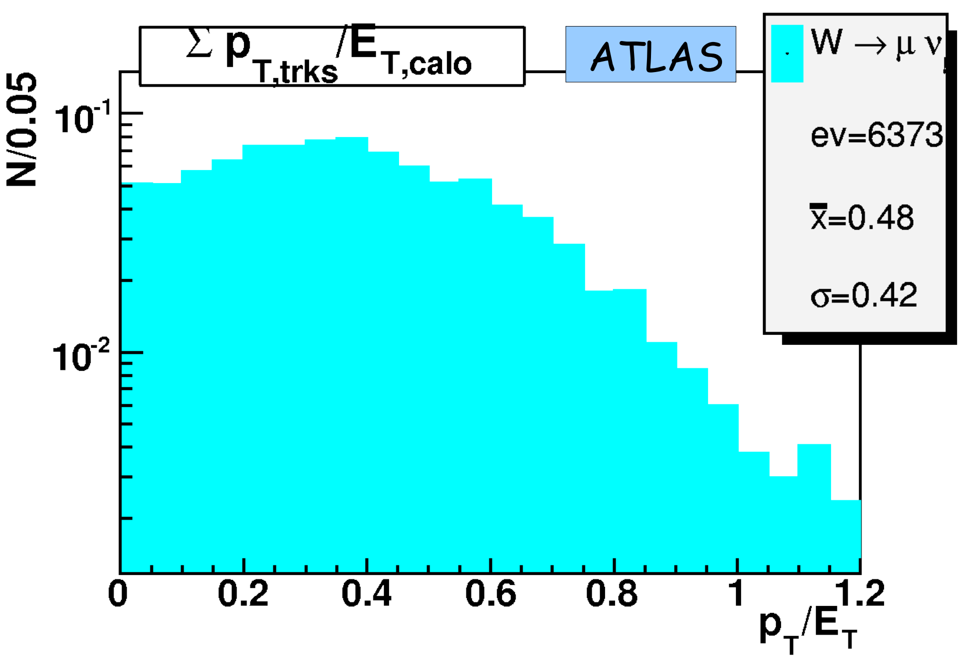


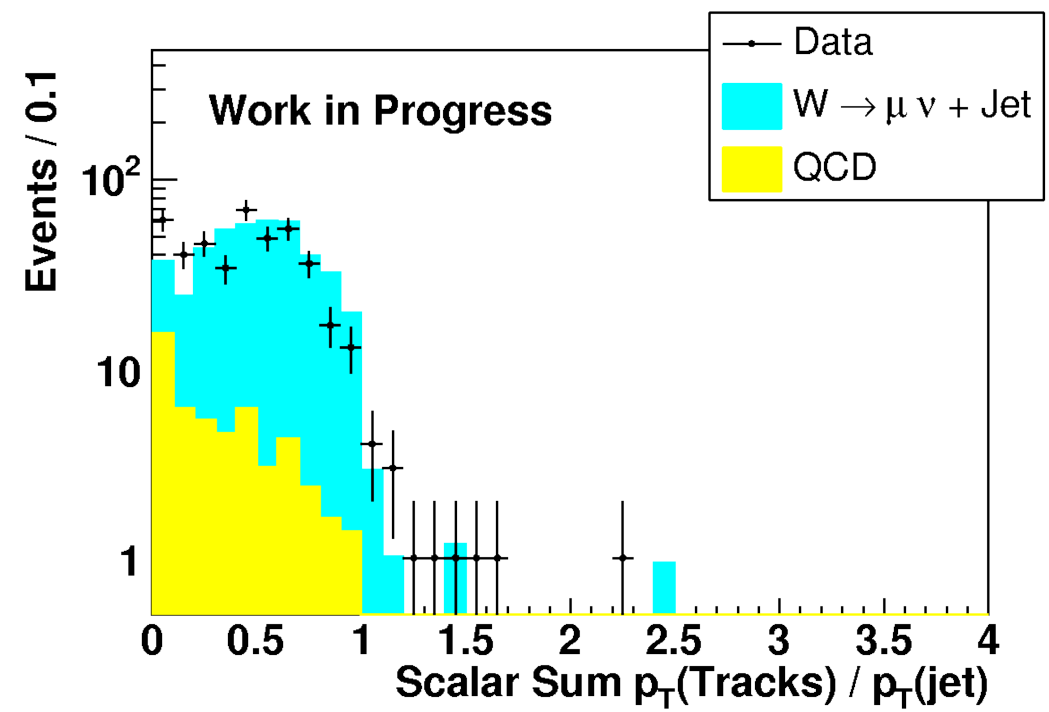


Preselection

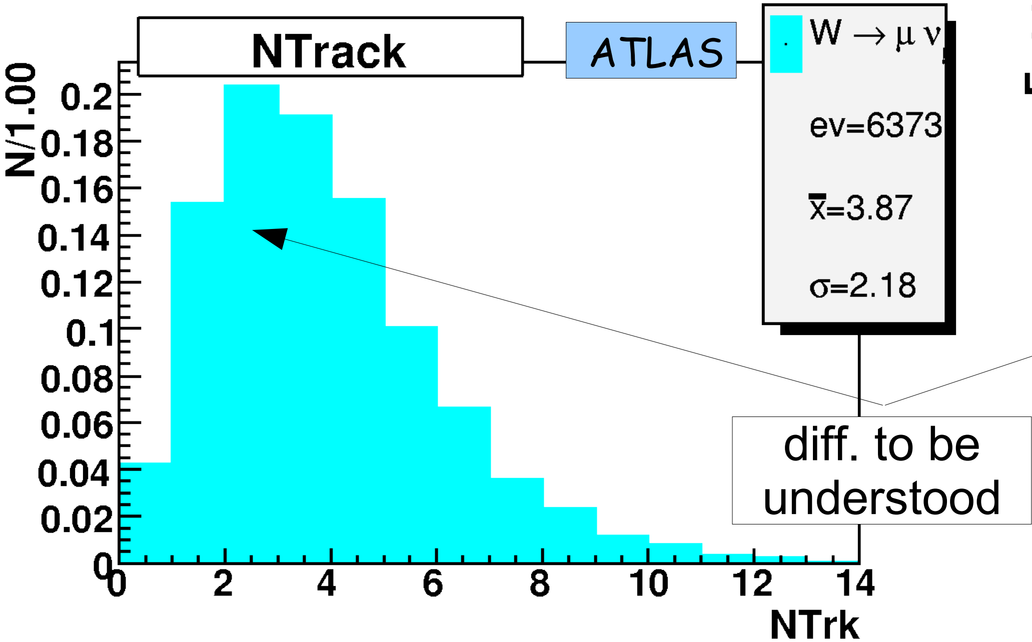


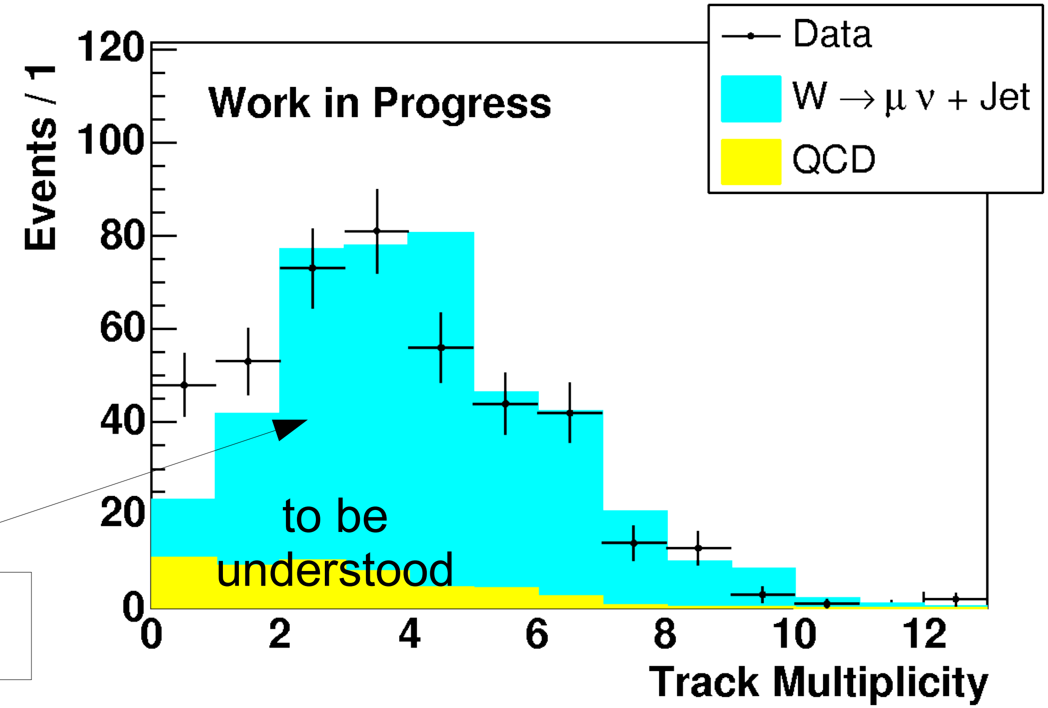






Preselection



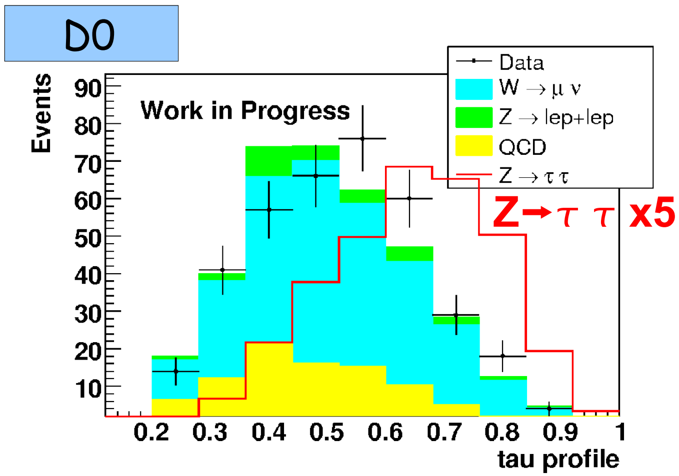


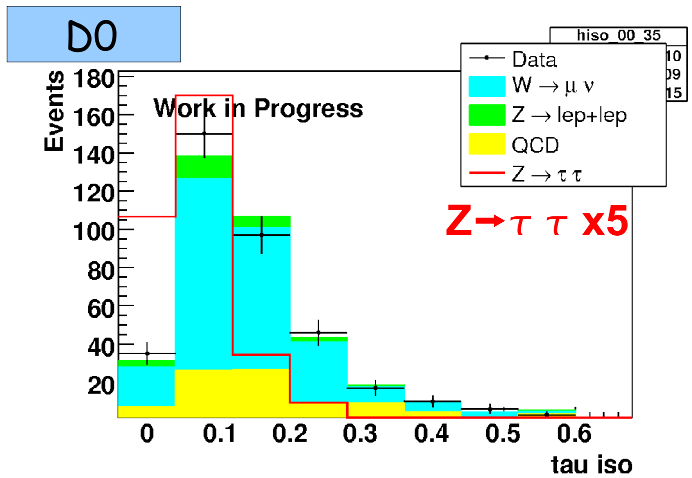
- low multiplicity jets show some discrepancies between data and MC
- in some bins the difference amounts to a factor of 2
- overall uncertainty should be lower that x2
- → to be investigated in more detail
- ATLAS spectra show higher 1 track fraction than D0
- → to be investigated in more detail
- conclusion: there are differences but the enviroment seems comparable

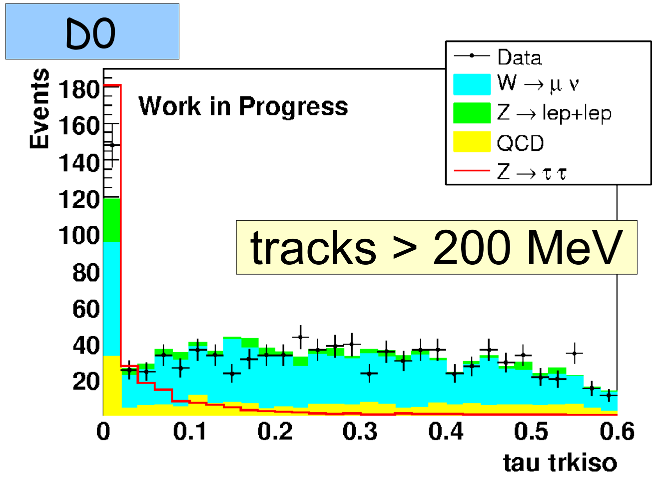
Input variables for a "SlimID"

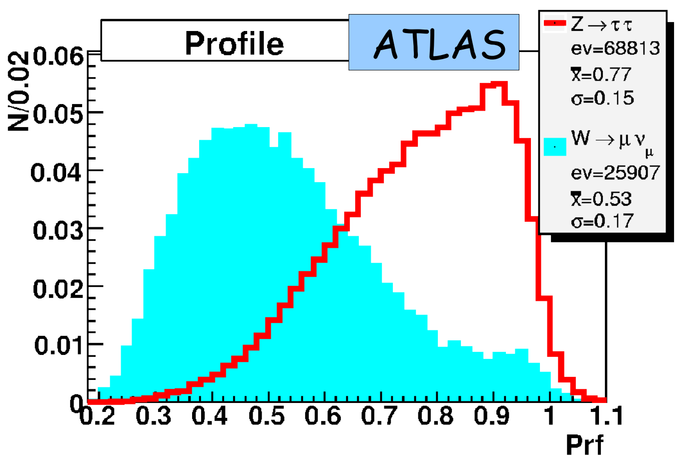
- **Strategy is to build a TauID ("SlimID") based on three main D0 input variables**
- **Profile** = (E_T(Tower1) + E_T(Tower2))/E_T(0.5), Towers defined on $\Delta\eta$ x $\Delta\varphi$=0.1x0.1 granularity
- **Isolation** = (E_T(0.5) + E_T(0.3)) / E_T(0.3)
- **TrackIsolation** = p_T(τ tracks) / p_T(all tracks)
  - τ-Tracks are defined in the following way
    - only tracks with $\Delta R < 0.5$, $p_T > 1.5$ GeV are considered, and sorted in p_T
    - the first track is always a τ track, the second/third tracks are τ tracks if their invariant mass together < 1.1 / 1.7 GeV
- all D0 variables were implemented in ATLAS (not only the once shown here ➡ see last TeV4LHC talk)

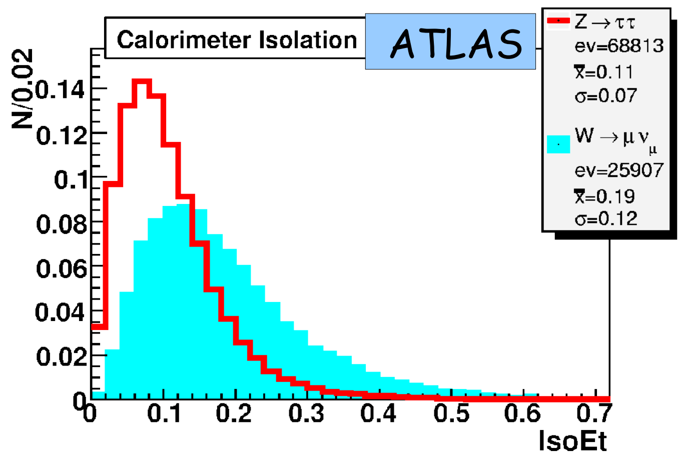
Comparison of input variables

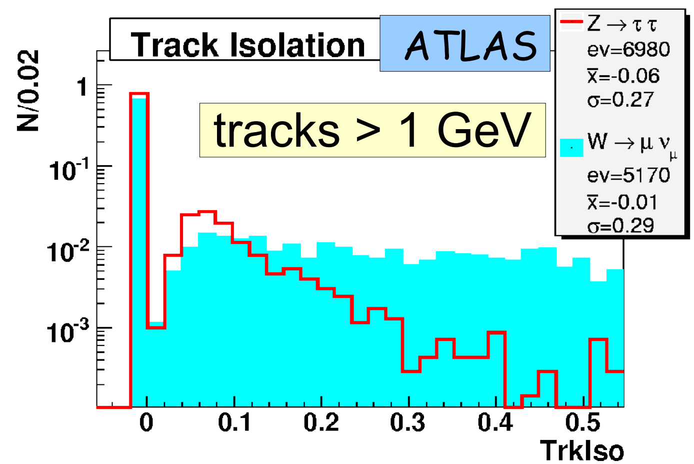








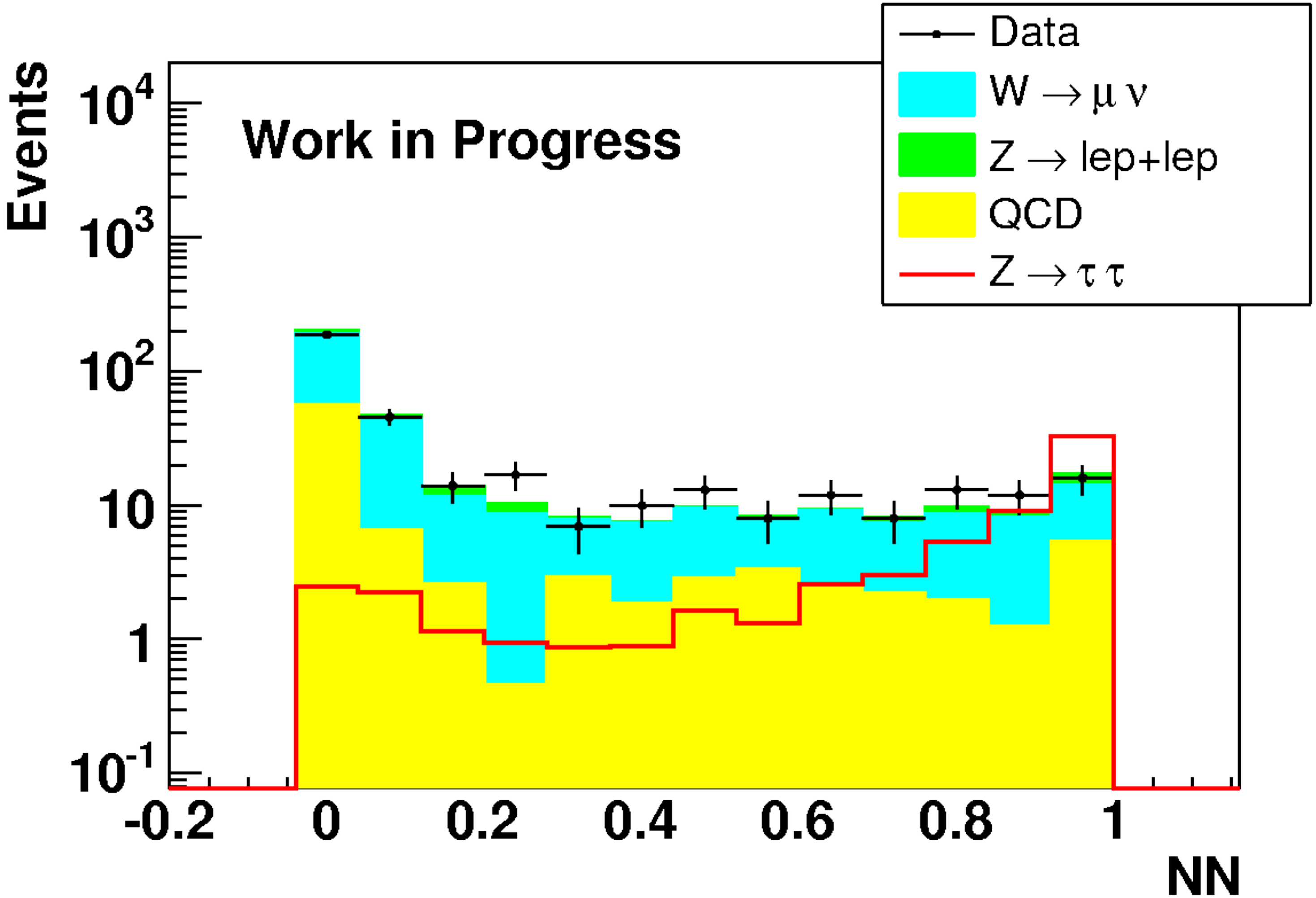




- **D0 shows nice agreement for the input variables**
- **TrackIsolation shows biggest difference ➡ comes from the tecnical drawback that at ATLAS there are only tracks > 1 GeV readily available**
- **variables certainly seem comparable ➡ we don't probe a very different phase space region in D0 and ATLAS of the jet shape**

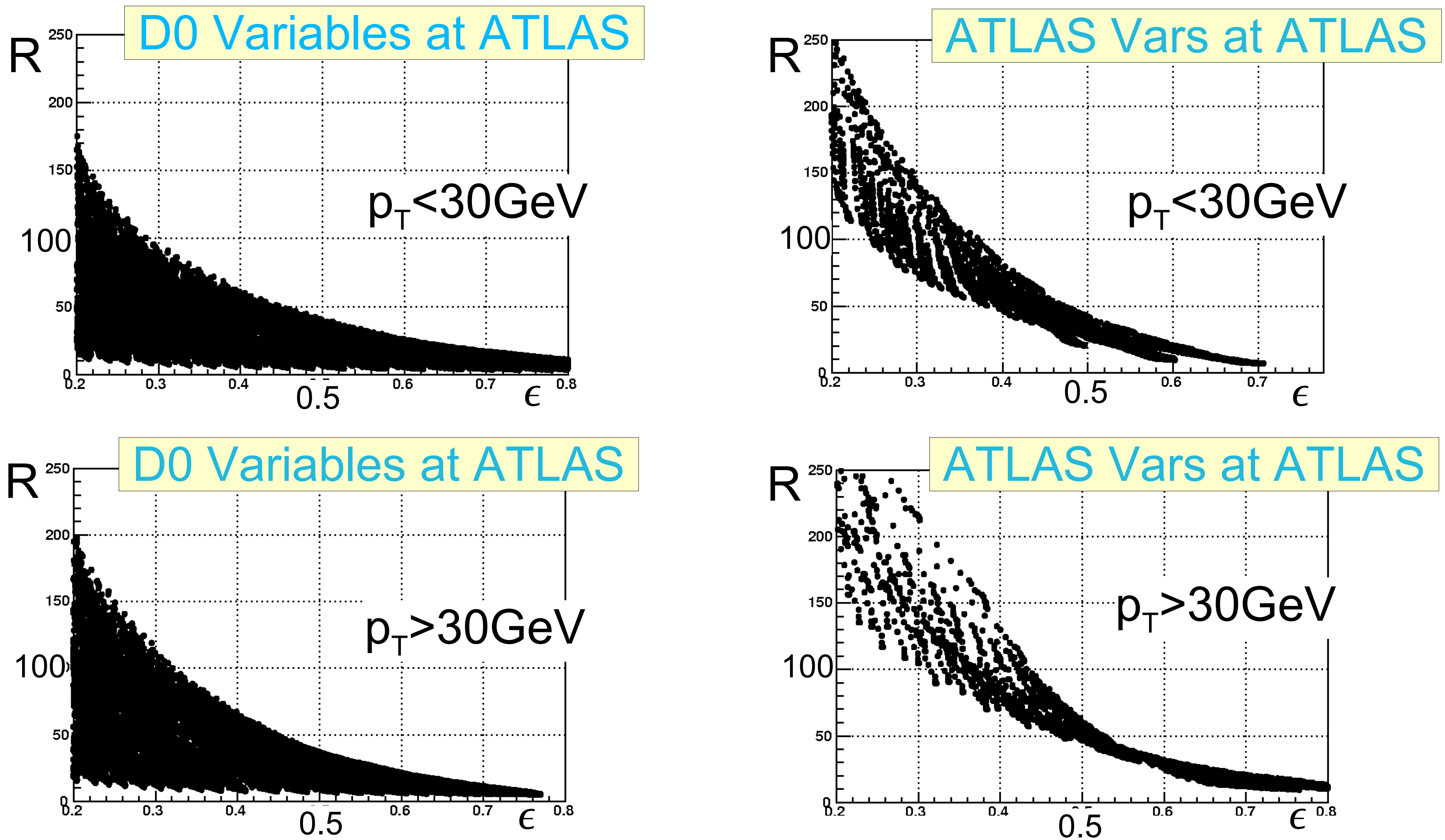
D0 SlimID and FullID

- also in D0 a scan in Profile, Isolation and TrackIsolation was performed
- both this simple cut and the cut on the NN show nice agreement between DATA and MC
- →ANN plot shows that also the full ID performance is well described in the MC even in the very tau like region of e.g ANN > 0.9
- now again: how does tthis transfer to ATLAS ? Do we really get the gross effect looking only at three variables ?



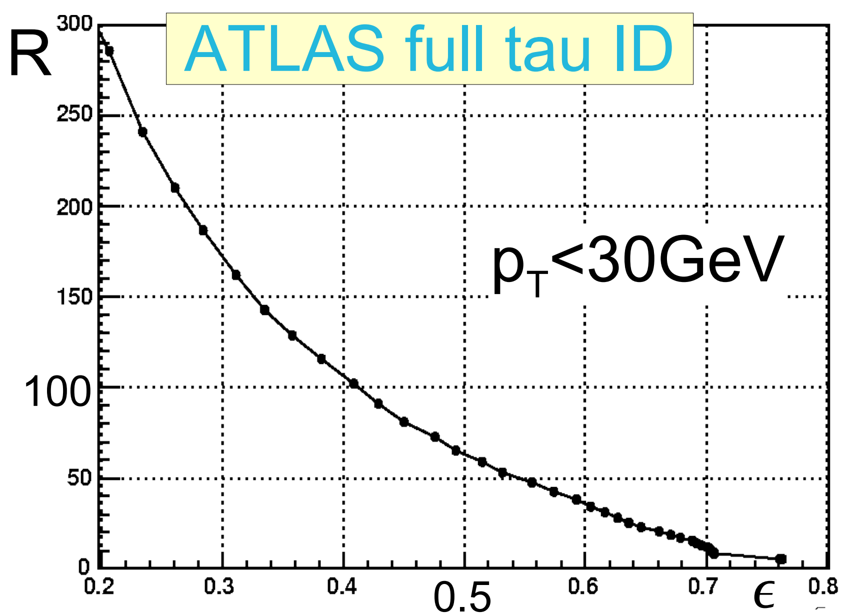
SlimID

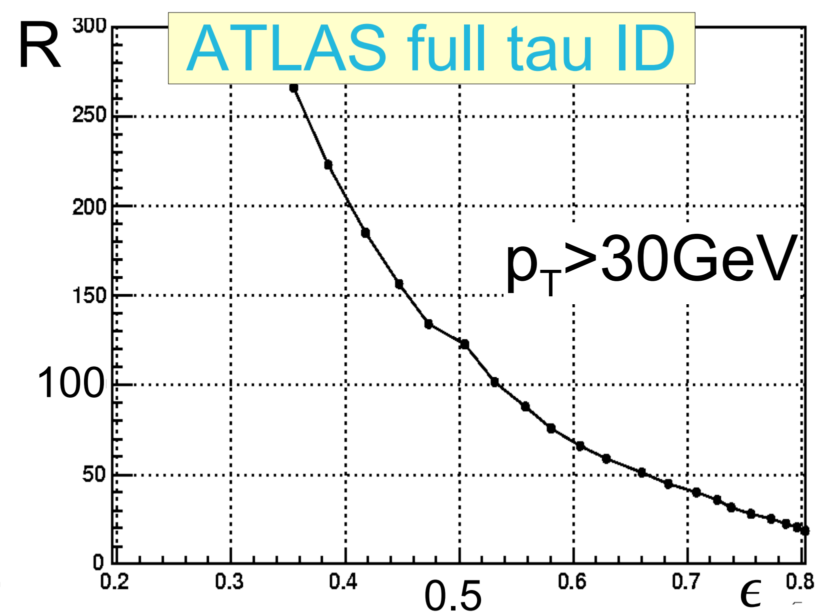
- Idea: implement a comparable "TauID" based on only three basic D0 variables: Prf, Iso, TrkIso



SlimID

- As expected both D0 and ATLAS variables show lower performance for lower p_T
- As expected ATLAS variables show a little bit better performance than D0 variables (mainly because Prf does not make use of the fine granularity calo)
- but performances in both cases are reasonably comparable
- ➞ another check passed that D0 results are transferable to ATLAS
- one caveat: a test showed that for signal both sets of variables selected 86% same candidates, and for fakes upto ~50% (varying with choosen efficiency)
- numbers for the full tau ID show that it improves upon the simple cuts by ~50%
- ➞we are loosing something here, but a big chunk is taken into account





Summary

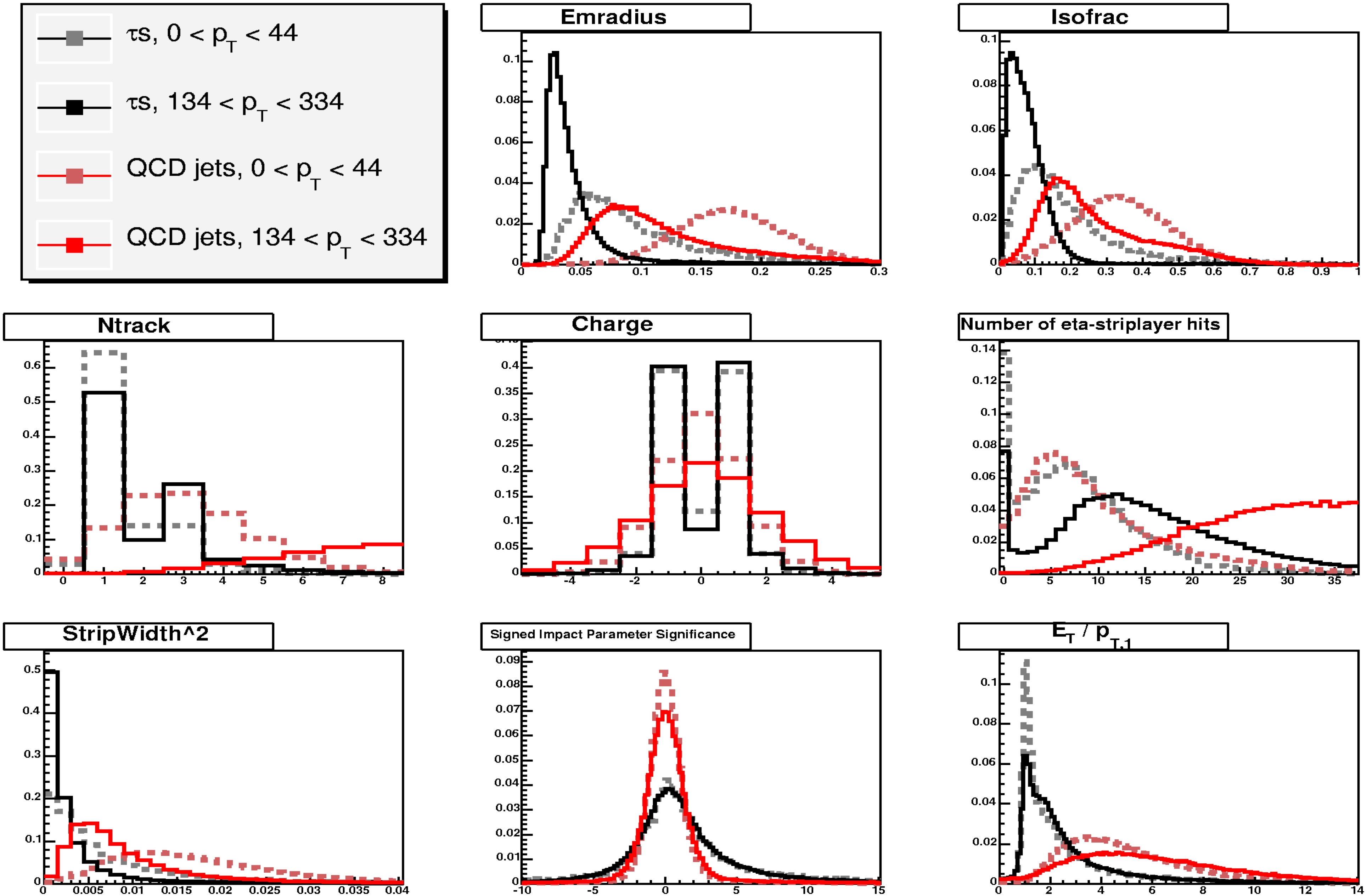
- D0 shows that the preselection performance (so the track multiplicity inside clusters) showed some but small descrepancy between MC and DATA
- it was also shown that the NN performance was correctly described showing no descrepancy within the statistical error between MC and DATA
- through implementing the D0 variables at ATLAS a comparison shows that results from D0 are applicable to ATLAS
- not to 100% but to a big part
- it also shows that the gross effect can be understood studying the 3 main variables (Prf, Iso, TrkIso)
- shown results are preliminary and have to be rechecked
- if everything holds, prediction for the potential of ATLAS involving tau final states should not suffer from a systematic uncertainty of > 100 % (with some safety margin) per τ
- this conclusion can only be drawn for the low (for ATLAS) p_T zone covered here

Backup Slides

TauID at ATLAS

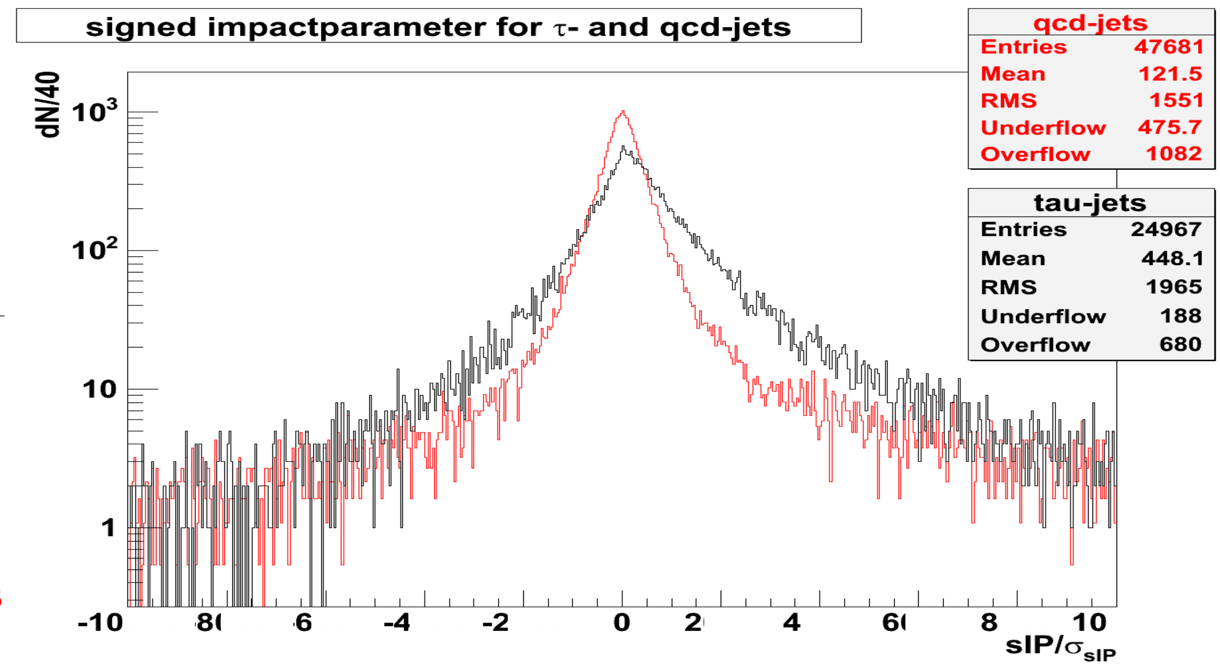
- Half a year ago a τ group has formed for ATLAS ➞ there is very much work in progress
- I will show only the “standard” way at the moment because we are more interested on D0 side anyway but there exist other algorithms as well
- Our τ-reconstruction package tauRec starts from clusters found by a sliding window algorithm
- We use the following quantities to discriminate τs against jets
  - R_{em}= Radius of the cluster in the em-calorimeter ΔR=0.4
  - ΔE_T^{12} = Fraction of the transverse Energy between ΔR=0.2 and ΔR=0.1 around the center of the cluster
  - N_{tr} = Number of Tracks within 0.3, p_T>2GeV
  - N_{em} / N_{strip}= Number of Hits in EM calo/η-Strip, E_T>200MeV
  - $E_{T,width,strip}$ = Width in the η-Strip
  - $E_{T,em}/E_T$, Charge, $E_{T,had}/\Delta p_{T(tracks)}$
  - Lifetime Signed Impact Parameter
  - for 3 prong decays: secondary vertex

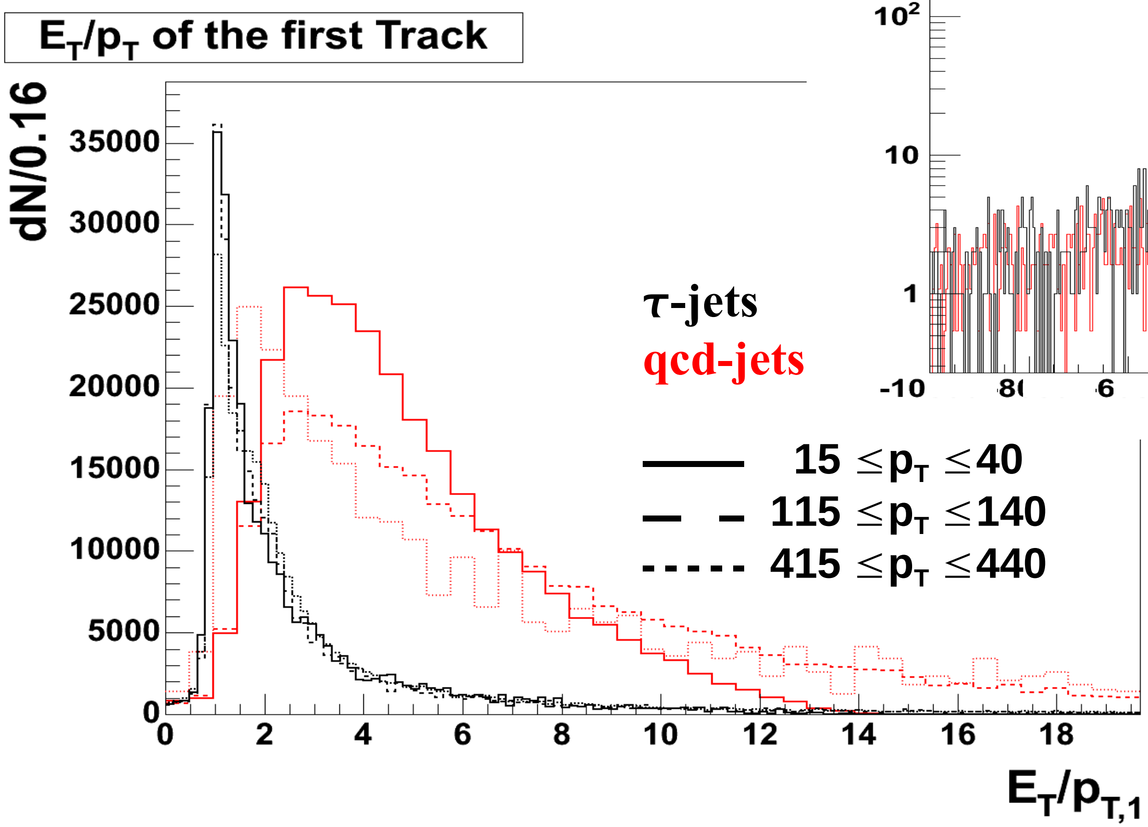
TauID at ATLAS



TauID at ATLAS

- Impactparametersignificance A_0/σ_{ip} : only 2d information, no reconstructed primary vertex (soon to come)
- Sign is defnied as $sign(\sin(\varphi_{track} - \varphi_{cluster}))$

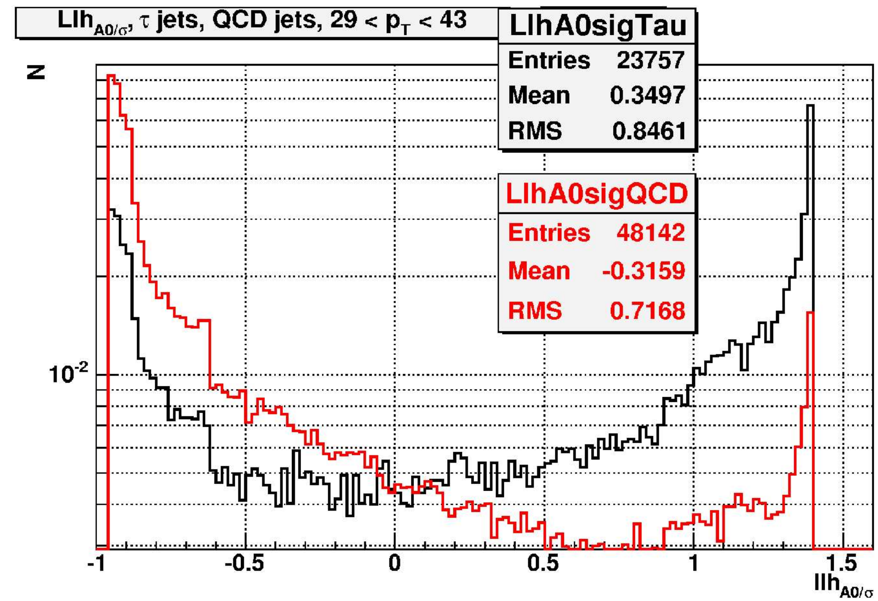


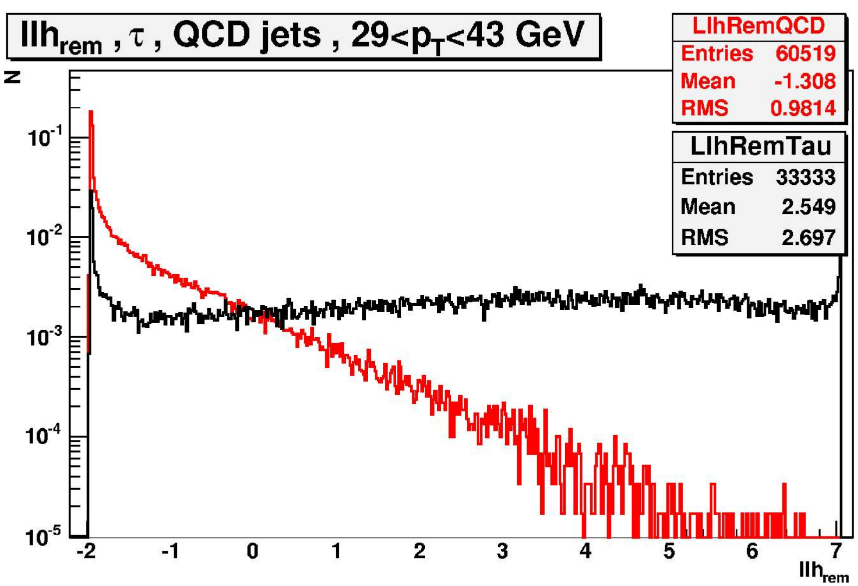


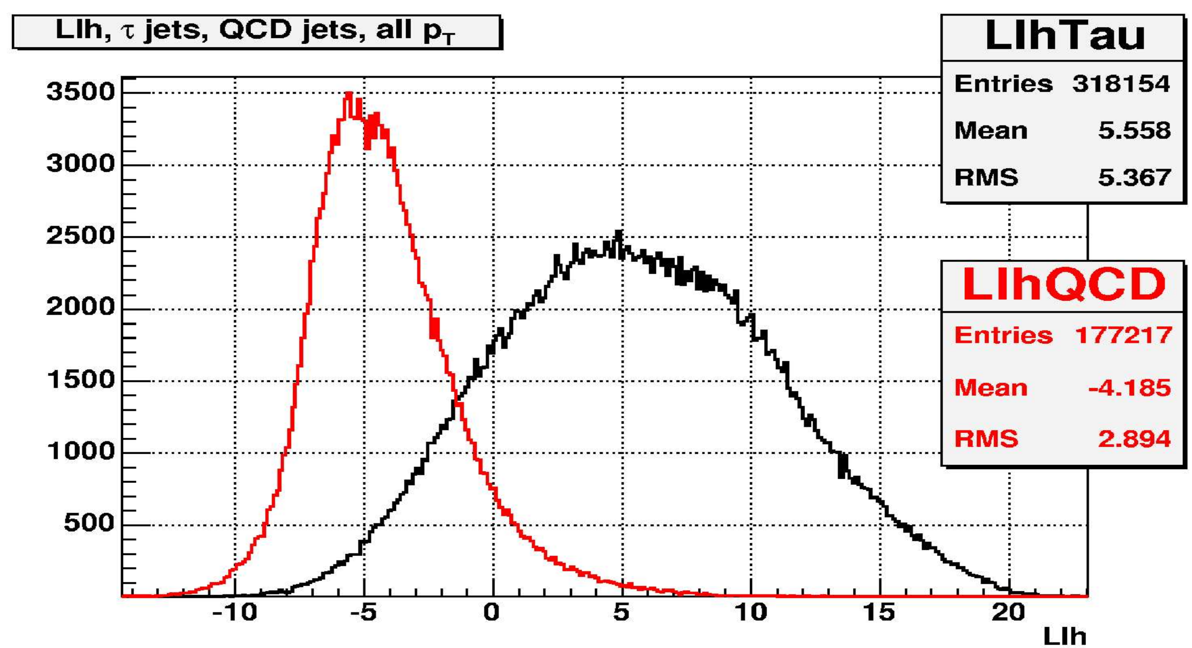
- E_T/p_T : Ratio of total to charged transverse Energy of the Jet
- Shows E_T dependance for QCD-Jets but none for τ-Jets

TauID at ATLAS

- All variables are then combined into a LikelihoodRatio
- preselection cut before the Llh: $1 \leq N_{Tr} \leq 3$
- 3 discreet variables, N_{tr}, N_{strip}, Charge, Llh directly from histograms
- 4 continous variables, R_{em}, ΔE_T^{12}, $E_{T,width,strip}$, $A0/\sigma$, fitted with arbitrary functions (normaly gaus*polynom)
- 10 p_T bins from 15 to 600 GeV, for Noise and NoNoise







TauID at D0

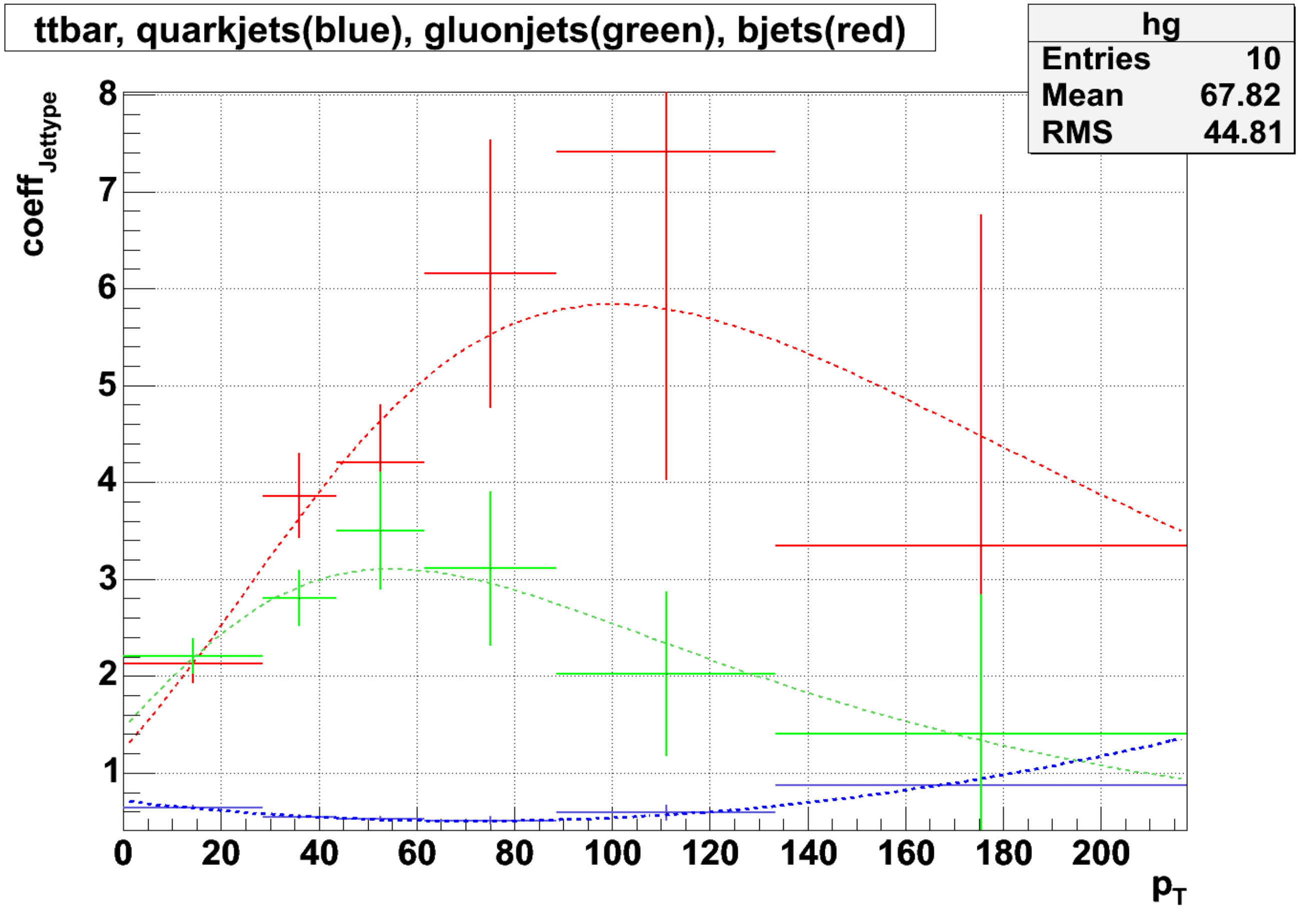
- D0 makes use of similar variables but all are defined in a slightly different way
- D0 defines a τ – **Type** as follows
  - Type 1 : 1-prong – no em subcluster ($\tau \rightarrow \pi + \nu$)
  - Type 2 : 1-prong – with em subcluster ($\tau \rightarrow \pi + \nu + x\ \pi^0$)
  - Type 3 : 3-prongs (more than one τ track)
- The em subcluster is found by the following algorithm
  - Find the leading cells in the em layer with finest granularity
  - collect all neighbour cells
  - around the leading neighbour cell, in turn collect all its neighbour cells
  - collect all em cells (from other layers) which have an overlap with any of the so far collected cells
  - if their energy > 800 MeV they are called the em subcluster
- τ-Tracks are defined in the following way
  - only tracks with $\Delta R < 0.5$, $p_T > 1.5$ GeV are considered, and sorted in p_T
  - the first track is always a τ track, the second/third tracks are τ tracks if their invariant mass together < 1.1 / 1.7 GeV

TauID at D0

- all variables act within a cone of 0.5 around the calo center
  - definition variablename(x) means variable calculated using objects within dR < x around the calo center
- Profile = $(E_T(\text{Tower1}) + E_T(\text{Tower2}))/E_T(0.5)$, Towers defined on $\Delta\eta \times \Delta\varphi = 0.1 \times 0.1$ granularity
- Isolation = $(E_T(0.5) + E_T(0.3)) / E_T(0.3)$
- M(Track1, em subcluster)
- $p_{T1} / E_T = p_T$ of the leading tracks divided by the calorimeter energy
- EM12Frac = $(E_T(\text{EM1}) + E_T(\text{EM2})) / E_T$, where ET(EM1) means trans.energy in the first em layer
- trkiso = $p_T(\tau \text{ tracks}) / p_T(\text{all tracks})$
- $e1e2/E_T$ = sqrt(sum(τ tracks p_T) * E_T(EM)) / $E_T(0.3)$
- em3iso = $E_T(\text{em subcluster}) / E_T(\text{EM3})$
- ntr1030 = number of tracks within 10° – 30° around the calo center

Plans

- this project has just begun
- a lot of details have still to be understood and differences made as small as possible (algorithm and samples)
- of course we need to finish the implementation of all variables
- the identification of tau types does not work at ATLAS → need to do something reasonably similar
- the samples have to be choosen carefully and enough statistics has to be available



- at ATLAS we see that the rejection varies very much with the type of jet you are rejecting

Plans

- → the samples we use for comparisons (also for the backgrounds) should be as similar as possible
- they should match in p_T, η, and jettype

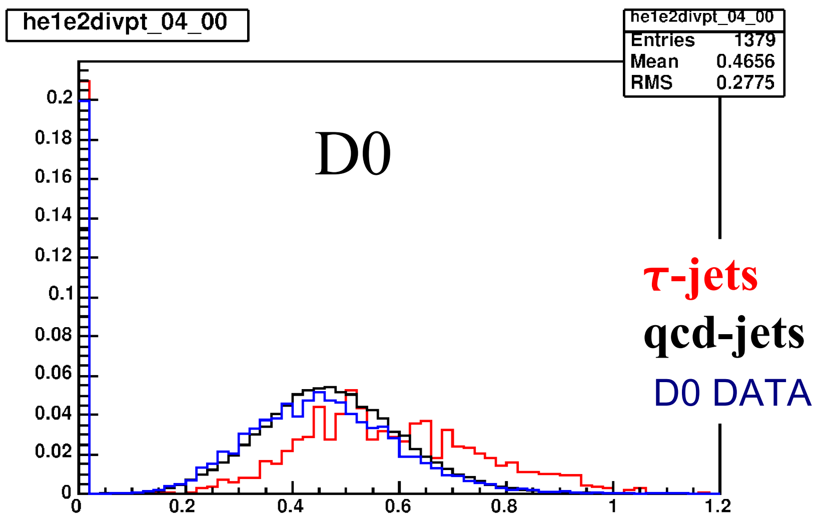
- → will probably use W+jet in the future for jets, sticking to $Z \rightarrow \tau\ \tau$ for τs
- the preselection of D0 data should be imitated selecting the samples for ATLAS
- the p_T and η distribution of taus and jets should be as similar as possible
- the influence of the reconstruction has to be understood (perhaps normalizing to jets, but also the jets may show differences)
- → make a nice comparison between D0 MC, D0 data and ATLAS MC

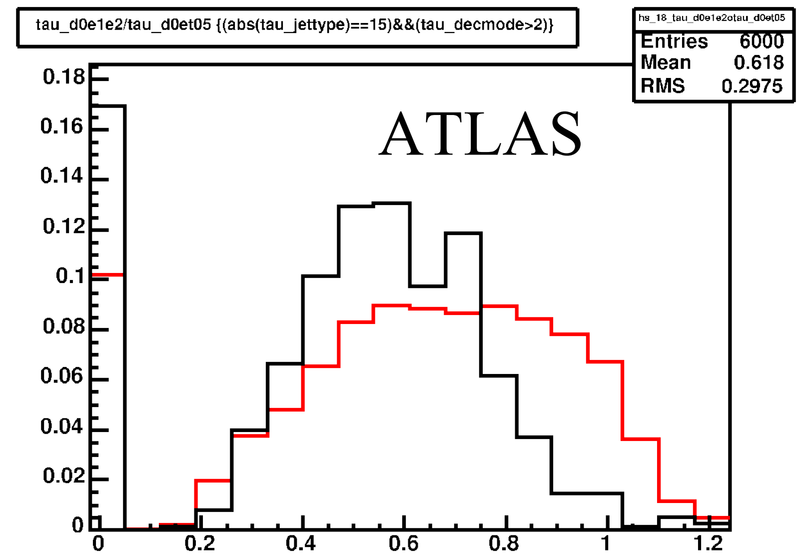
TeV for LHC(3)

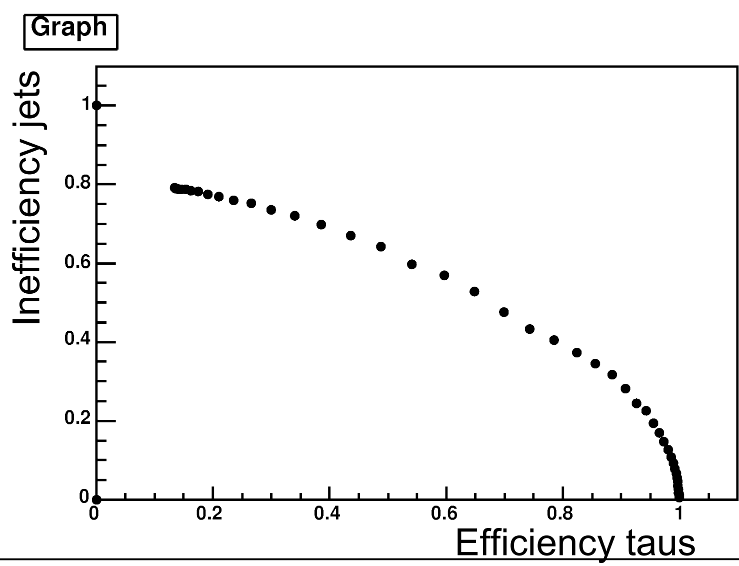
Comparison e1e2/ET(0.3)

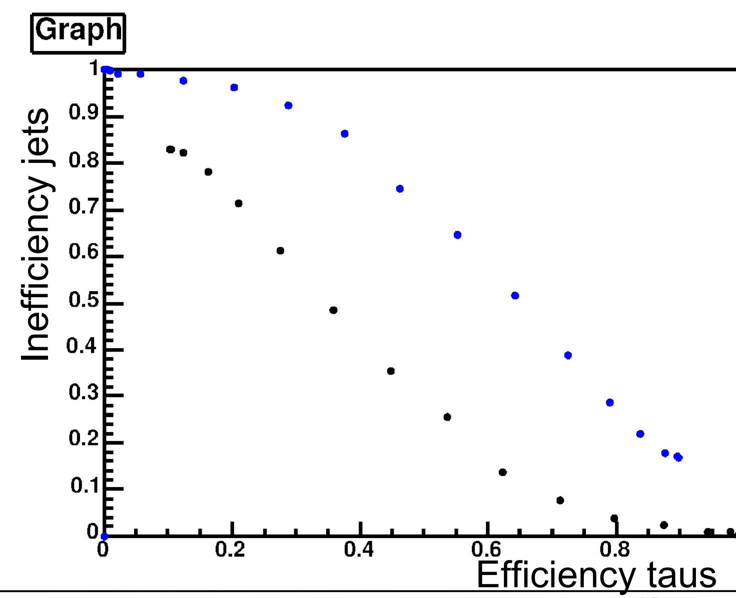
- e1e2/ET is a measure of the difference in energy compared to the em calorimeter

- e1e2/ETshows weaker discrimination for both
- distributions due not match perfectly, ATLAS show cases with very few EM energy
- the performance show different dependency on the efficiency
- but overall trend seems ok





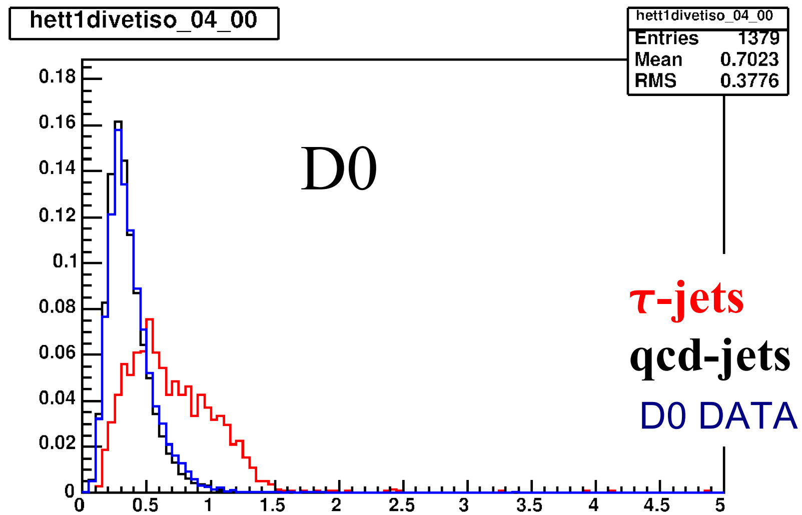


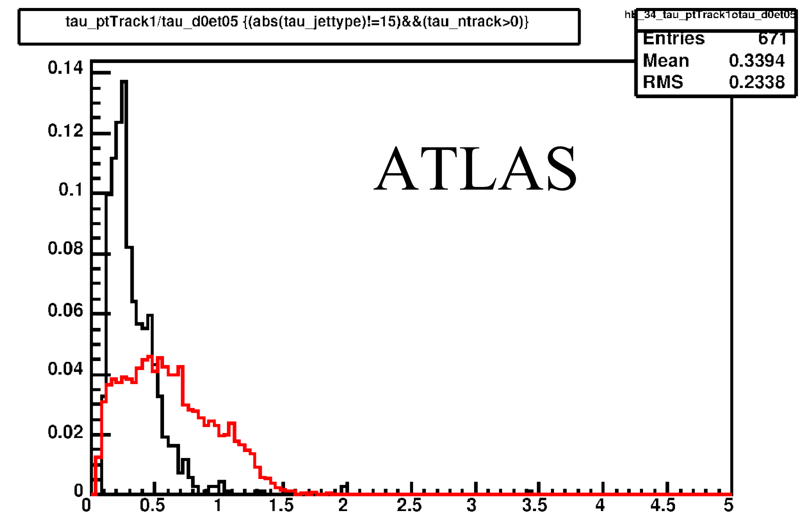


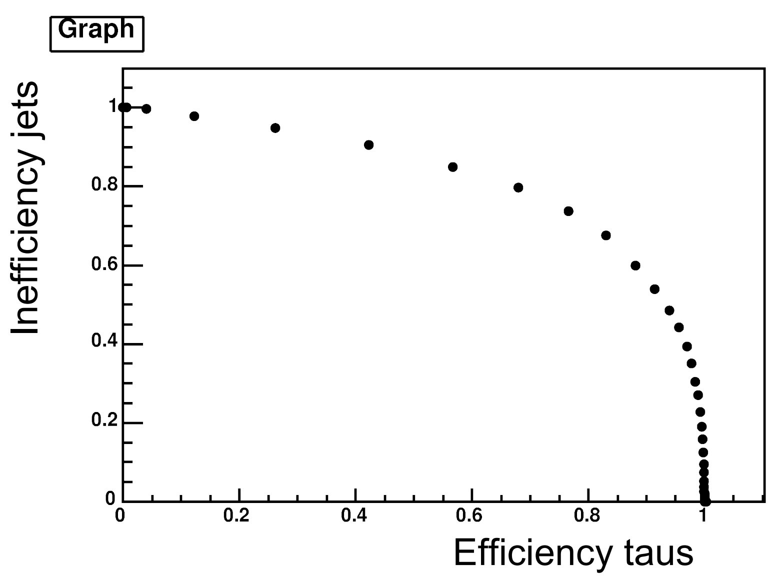
Comparison pT(Tr1)/ET(0.5)

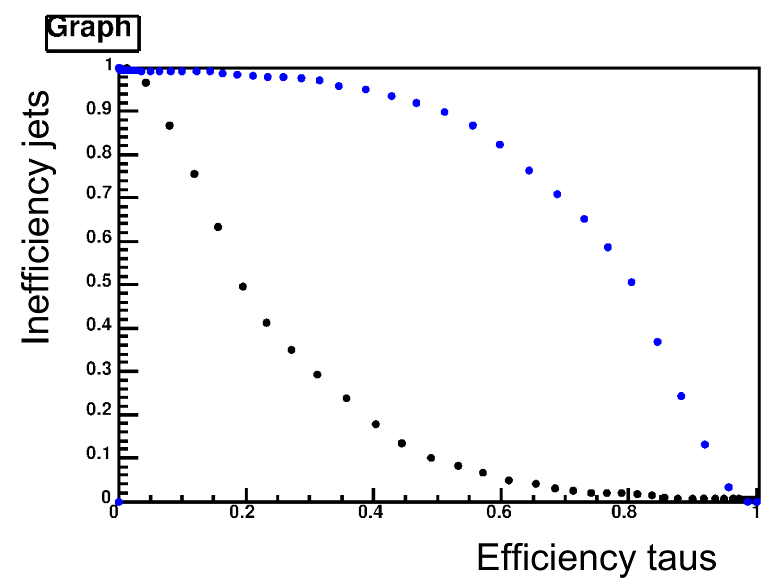
- pT of the leading tracks devided by the total calorimeter energy

- shows a good performance
- distributions are comparable
- the performance shows different behaviour for high efficiency





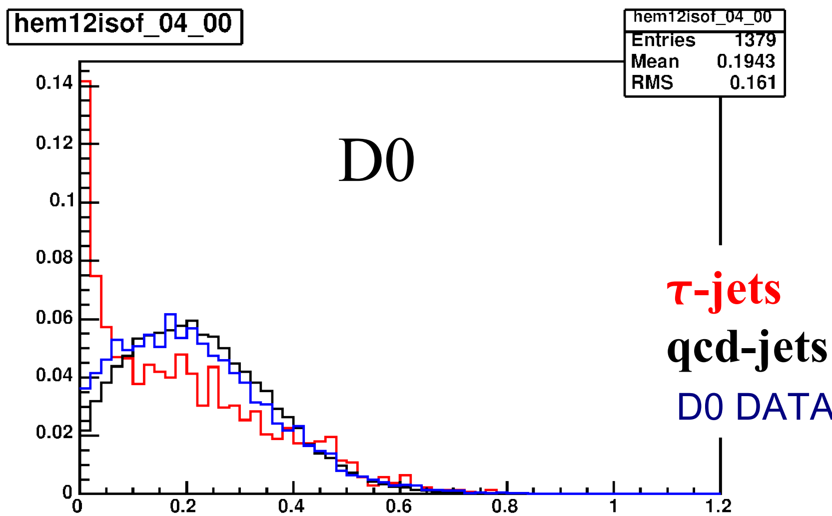


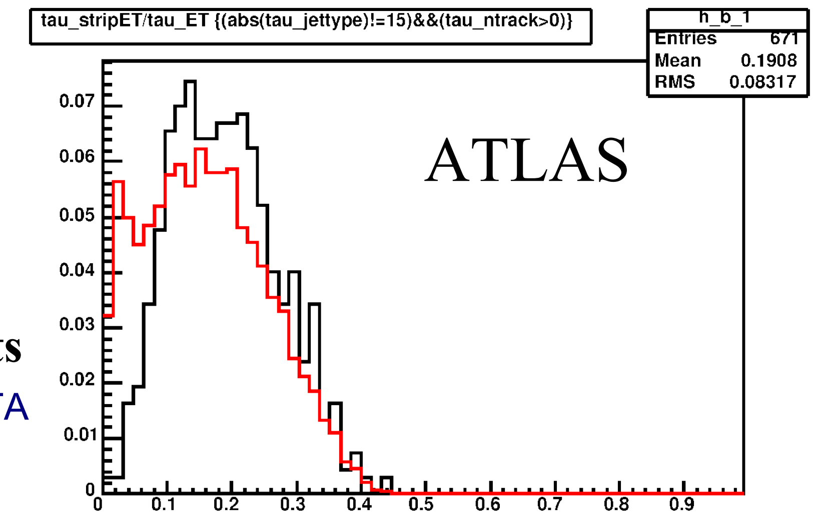


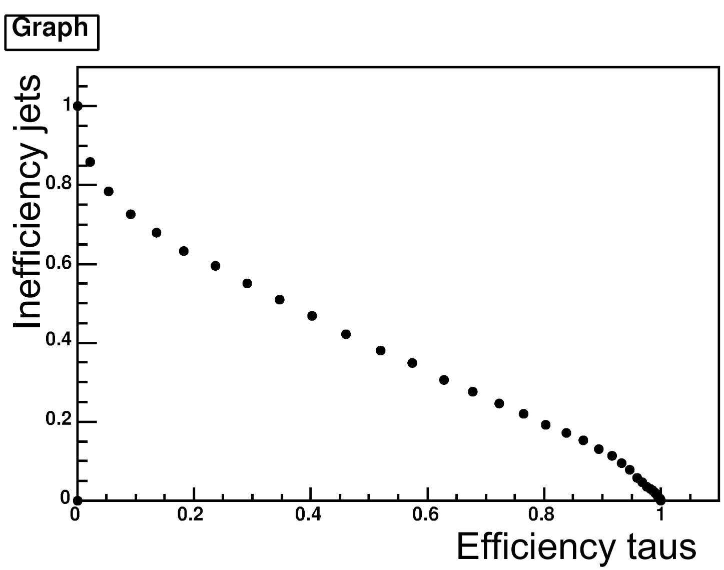
Comparison EM12IsoF

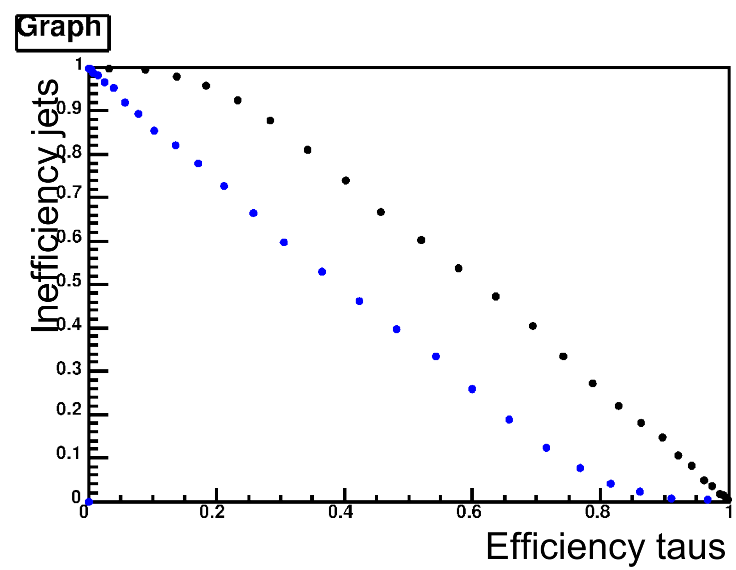
- energy in the first two em layers devided by the total calorimeter energy

- for ATLAS I used only the eta-strip-layer, but this already has ~4.3X0 → fewer cases were there is no energy
- the resulting distributions are therefore quite different
- the performance shows a different behaviour









Comparison Profile

- sorry, don't have the profile for ATLAS yet, the equivalent quantity is EMRadius

- for both ATLAS and D0 the “profile” is an important variable showing good discrimination
- only for the interest: the distributions are mirrowed but show some similarity
- we will soon have the D0 style profile for ATLAS