Measurement of the τ-jet performance with first ATLAS Data

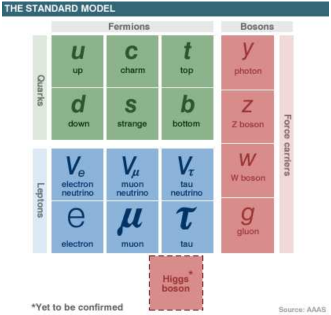
Thies Ehrich

Max-Planck-Institut für Physik
(Werner Heisenberg Institut)
München

Physics at the Large Hadron Collider
3-Dec-2008

Table of Content

The Standard Model of Particle Physics



Source: AAAS

- Theory of three of the four known fundamental interactions
- Gauge Symmetry with the gauge group SU(3)xSU(2)xU(1)
- All experimental tests agreed with its predictions
- Higgs Boson not directly observed (yet)

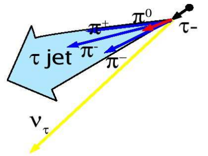
τ-Leptons in the SM

Main Properties of the τ-Lepton

- large mass: m$_\tau$=1.778 GeV
- measureable lifetime: cτ=87.11 μm
- decay modes are well understood from previous experiments

Decay modes

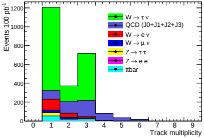
- 35% leptonic:
  - $\tau \rightarrow e\ \nu_\tau \nu_e$
  - $\tau \rightarrow \mu\ \nu_\tau \nu_\mu$
- 65% hadronic:
  - $\tau^+ \rightarrow \pi^+\ \nu_\tau\ n\pi^0$
  - $\tau^+ \rightarrow \pi^+\ \pi^- \pi^+\ \nu_\tau\ n\pi^0$



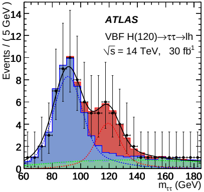
$\Rightarrow$ Focus in this talk: hadronic τ-decays

Why are τs important?

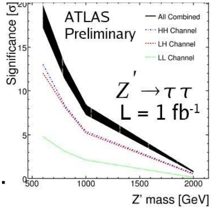
- SM: large number of taus already at very early data taking (e.g. W→ $\tau\nu$)



- Important for the discovery of Higgs Bosons
  - H→ $\tau\tau$ (Standard Model Higgs)
  - A→ $\tau\tau$ (MSSM Higgs)
  - H$^{\pm}$→ $\tau\nu$ (MSSM charged Higgs)



- are used in many new physics searches like Z'→ $\tau\tau$



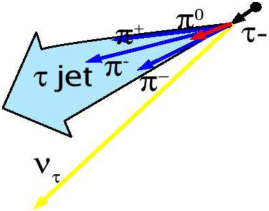
τ-jet Reconstruction with ATLAS

τ-jet in the ATLAS detector

- collimated calorimeter cluster
- low charged tracks multiplicity
- displaced secondary vertex

Main sources of τ-fakes

- jets from light quarks
- electrons
- muons



Two complementary
τ-jet reconstruction algorithms:

- track-based:
  uses tracks as initial
  reconstruction seed
- calorimeter-based:
  uses calorimeter cluster as
  initial seed

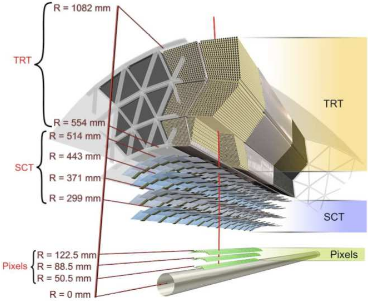
ATLAS Tracking and Calorimetry

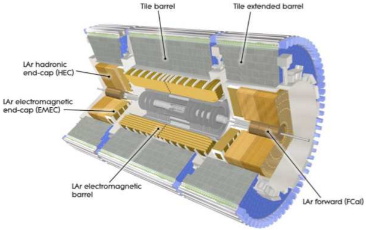
ATLAS Inner Detector

- Pixel: 140M channels, 2.3 m^2
- SCT: 6.2M channel, 61.3 m^2
- TRT: 420k channels

ATLAS Calorimeters

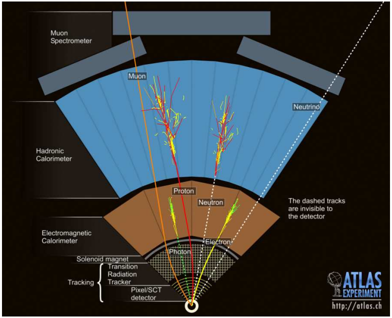
- EM calo (lead/Liquid Argon)
- hadronic barrel calo (iron/plastic,LAr)
- hadronic endcap calo (copper/LAr)
- forward calo (tungsten/LAr)





τ-jet Reconstruction with ATLAS (2)

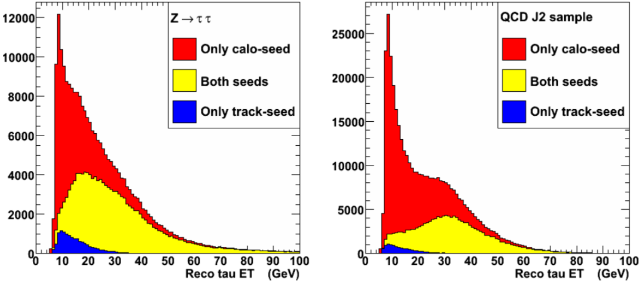
1. for each track seed a calorimeter cluster is searched
2. if a cluster is found calorimeter -based reconstruction is run
3. if no jet found, the candidate is track seeded only
4. use remaining jets for the calorimeter-based reco only



In case of 2:

- the position of the candidate is defined by the track
- energy of the candidate is defined by the calorimeter-based candidate
- the track multiplicity is defined by the track-based candidate

Reconstruction Performance

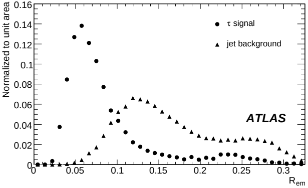


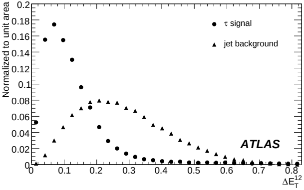
		both seeds	only track-seed	only calo-seed
Signal	Reconstructed	50%	5%	45%
	Matched to Truth	75%	<1%	25%
QCD	Reconstructed	3%	33%	64%

- efficiency loss by selection both seed
- but much lower risk of selection background events

τ-Identification: tauRec (calo seed)

Both algorithms produce a set of identification variables:





EM radius

$$R_{EM} = \frac{\sum_{cells} E_{T_i} \sqrt{(\eta_i - \eta_{cluster})^2 + (\phi_i - \phi_{cluster})^2}}{\sum_{cells} E_{T_i}}$$

EM radius explores the smaller transverse shower porfile of the τ-jet wrt. to jets.

Isolation Fraction

$$\Delta E_T^{12} = \frac{\sum_{cells}^{0.1<\Delta R<0.2} E_{T_i}}{\sum_{cells}^{\Delta R<0.4} E_{T_j}}$$

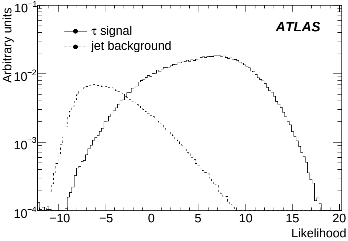
τ-jets are well collimated and have to be isolated

τ-Identification: Likelihood

The discriminating variables are combined to a likelihood function

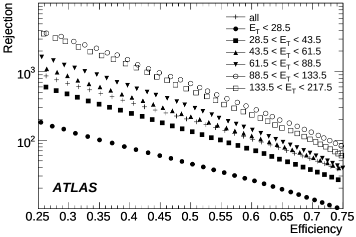
The Likelihood Function

- each of the variables represents a probability density function (pdf) once it is normalized
- the pdf's can be multiplied in order to combine them
- for some reason many people use logarithms though: $L = \Pi f(x;p_T)$, or $\log L = \sum \log f(x;p_T)$



In this study, τ-jets are identified using a likelihood cut llh>4.

Efficiency and Rejection



Definition of Efficiency using MC truth

$$E_{MC} = \frac{\#\text{ identified taus matched to generated taus}}{\#\text{generated taus}}$$

Definition of Rejection using MC truth

$$R_{MC} = \frac{\#\text{generated non-taus}}{\#\text{identified taus matched to generated non-taus}}$$

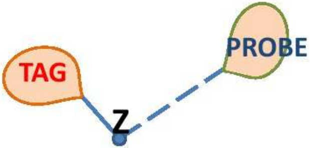
Efficiency Measurement with real Data

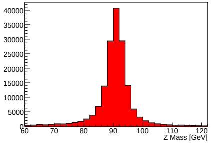
- * Simulated data suffers from large uncertainties.
- * Two of these are the τ reconstruction efficiency/rejection

To measure the efficiency from data, one usually uses the tag-and-probe method:

The tag-and-probe method

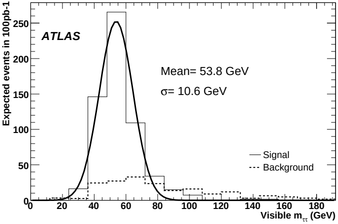
- ○ look for Z$\rightarrow \tau\tau \rightarrow$lep+jet
- ○ select only events with well reconstructed muon
- ○ $\Rightarrow$ this muon is our tag particle
- ○ look for a jet, which fits to the Z mass
- ○ $\Rightarrow$ this jet has to be a τ!
- ○ check, if this jet is identified as a τ (by llh cut)





Z$\longrightarrow \tau_{\mathbf{had}} l$

In principle we can now calculate the efficiency by #taus/#jets



Cuts

- ○ Isolated Lepton
- ○ Missing Energy
- ○ Low Activity
- ○ b-jet Veto
- ○ (τ-Id cuts)

- the measurement is not free of background
- we can estimate the background by selecting particles with same electric charge (charges if main backgrounds QCD and W+Jets are uncorrelated)

Rejection Factors from Z$\rightarrow \mu\mu$,ee+Jets

Idea

- Select Z events with an accompanying jet (Backgrounds: $t\bar{t}$, $b\bar{b}$,...)
- Look at this accompanying jet:
- If backgrounds (including real τs) is small this jet is not a τ-jet
- Check if the jet is identified as a τ-jet (llh>4) ⇒ fake τ-jet

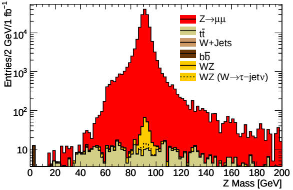
Process	Generator	DS	# Events	Lumi/pb^{-1}
Z$\rightarrow$ ee	Pythia	5144	415k	237
Z$\rightarrow \mu\mu$	Pythia	5145	420k	232
Dijets	Pythia	5802	2.25M	0.012
$t\bar{t}$	Mc@NLO	5200	988k	2189
$b\bar{b}$	Pythia	5701/14	260k	
WZ	Herwig	5987	50k	3055
W+Jets	Alpgen	6108-12	110k	11,30,150,520,475

Cuts for the Z selection: trigger, two (isol.) leptons, exactly one jet, Z mass window

Invariant Mass Z→ $\mu\mu$ and Z→ee

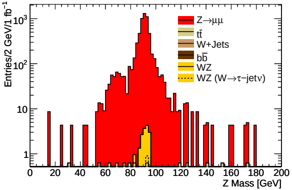
Selection with one jet

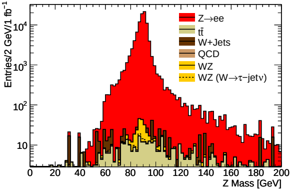
Selection with one jet, llh>4



llh>4

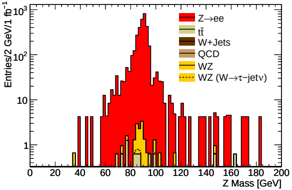
⇒





llh>4

⇒

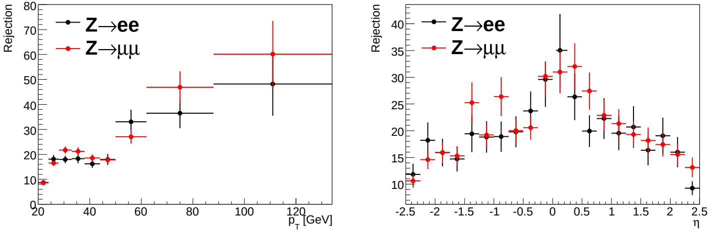


- The background contribution is smaller than 0.2%
- In addition only the backgrounds with real τ contribution introduce an error on the rejection rate measurement (shown for WZ→ $\tau_{had}\nu\ ll$)

Rejection Factors in Z Events

Plot rejection in bins of p_T and η

- p_T=momentum in transverse plane
- η=-ln tan$(\theta/2)$, θ is the polar angle



Overall Results

Process	Data	MC
Z→ ee	19.47±0.72	19.54±0.72
Z→ $\mu\mu$	20.61±0.58	20.73±0.59

- Rejection in Z→ee slightly lower (electron contamination)
- Might be cured by tighter e-ID

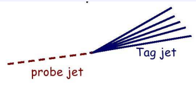
Rejection Factors in QCD events

Reminder:

$$\boxed{\text{Rejection} = \frac{\text{fake } \tau\text{-jets}}{\text{non-}\tau\text{-jets}}}$$

QCD events

- process with overwhelming cross section (10^8 pb) at LHC: pp→jetjet (Di-jets)
- these jets stem mostly from gluons
- this channel is automatically backgound free
- take the jet that triggered the event as a tag jet and probe the other one

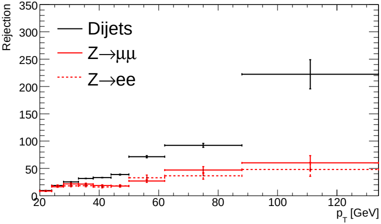


Comparison with Z+Jet Events

Dijets are complementary to Jets in Z events:

⇒ Dijets: mainly gluon\photon-jets

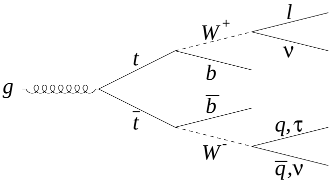
⇒ Z+Jets: mainly quark-jets

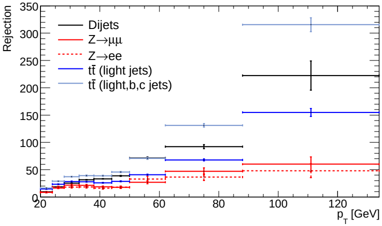


- Rejection much higher for dijets (pdf's trained with dijets)
- deviation gets more pronounced for high p_T

Comparison with $t\bar{t}$ Events

How does the measured rejection from dijets and Z+jets compare to the one in $t\bar{t}$ (based on MC truth in $t\bar{t}$) ?



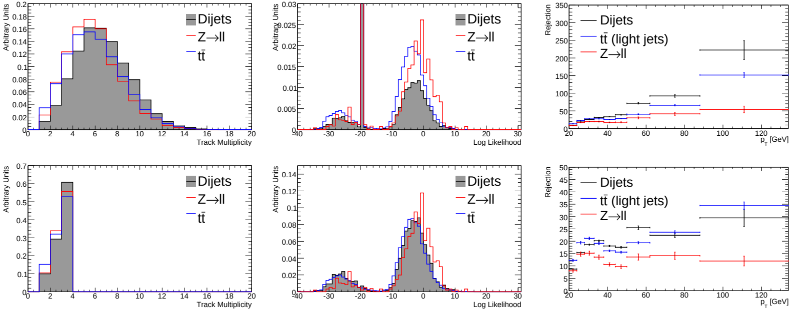


- remove b- and c-jets from $t\bar{t}$ to avoid bias (heavy flavour jets have much higher rejection factors)
- $t\bar{t}$ light jets show higher rejections than Z+jets, but are still below dijets

Track Multiplicity for high p_T

Why is the rejection higher for dijets in high p_T bins?

(plots shown for high p_T bin 88-134 GeV)

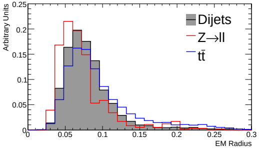


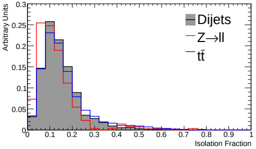
- more tracks in dijets → additional rejection power
- after cut on track multiplicity small shifts are visible in llh:
  - $t\bar{t}$ shifted to smaller values
  - Z jets shifted to higher values

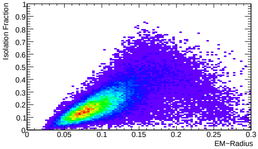
Rejection in p_T, EM-Radius Bins

Which llh variables contribute to the remaining variations ?

- shift of the em radius and isolation fraction distribution in opposite directions
- both variables are correlated, though







EM radius

$$R_{EM} = \frac{\sum\limits_{cells} E_{T_i} \sqrt{(\eta_i - \eta_{cluster})^2 + (\phi_i - \phi_{cluster})^2}}{\sum\limits_{cells} E_{T_i}}$$

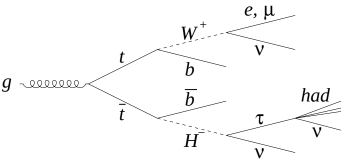
Isolation Fraction

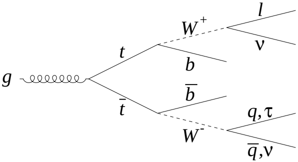
$$\Delta E_T^{12} = \frac{\sum\limits_{cells}^{0.1<\Delta R<0.2} E_{T_i}}{\sum\limits_{cells}^{\Delta R<0.4} E_{T_j}}$$

An Application: $H^{\pm} \rightarrow \tau\nu$

Charged Higgs Bosons

- $H^{\pm}$ appear in models with an extended Higgs Sector
  - SM like with two Higgs Doublets
  - Minimal Supersymmetric extension of the SM (MSSM)
- after the electro weak symmetry breaking five physical Higgs Bosons remain: A,H,h,$H^{\pm}$
- a discovery of $H^{\pm}$ would be a clear sign of physics beyond SM

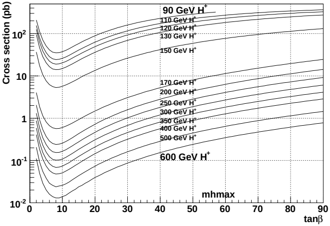


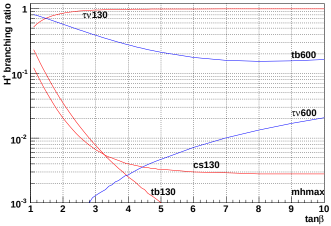


Production and Decay of $H^\pm$

- $H^\pm$ is produced in decays of the top quarks (if light enough)
- high cross sections upto 100 pb (10% of SM $t\bar{t}$)

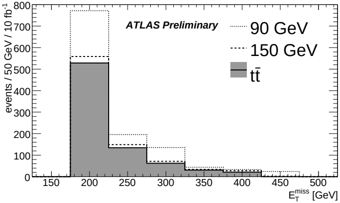
- $H^\pm$ deacys nearly exclusivly into $\tau\nu$ (if light enough)
- competes with SM decay W→ $\tau\nu$ (Br=11%)
- it would be detectable as an excess of τs in $t\bar{t}$ events

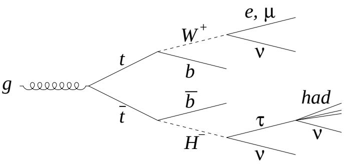




E_T^{miss} after all Cuts

After the selection cuts, we look into the excess of events in the E_T^{miss} distribution:





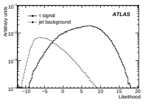
- Here the $t\bar{t}$ contribution was just taken from MC
- But one should rather estimate the SM $t\bar{t}$ contribution from the data itself.
- Lets assume we know the contribution from $t\bar{t} \rightarrow (\tau\nu b)(l\nu b)$

⇒ How to estimate the contribution from fake taus?
e.g. $t\bar{t} \rightarrow (qqb)(l\nu b)$

Towards $t\bar{t}$ Fakerate

Number rec. and id. taus/jets

$$N_{reco} = N^{\tau}_{reco} + N^{jets}_{reco}$$
$$N_{ID} = N^{\tau}_{ID} + N^{jets}_{ID}$$



Define efficiency and rejection

$$e = \frac{N^{\tau}_{ID}}{N^{\tau}_{reco}},\ r = \frac{N^{jets}_{ID}}{N^{jets}_{reco}}$$

With these definitions one calculates the number of fake τ-jets:

$$\boxed{N^{jets}_{ID} = \frac{r}{e-r}(eN_{reco} - N_{ID})}$$

- Already seen: e and r are not constant.

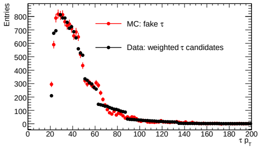
⇒ Parametrize as before and define τ-weights:

$$\omega = \tfrac{r}{e-r}(e-1)\ [\tau \text{ is identified}]$$
$$\omega = \tfrac{r}{e-r}(e-0)\ [\tau \text{ is not identified}]$$

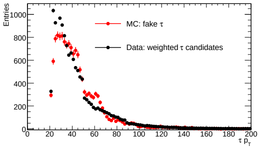
Then the number of misidentified taus is $N^{jets}_{ID} = \sum^{N_{reco}} \omega_i$

Fake τ-p_T from data

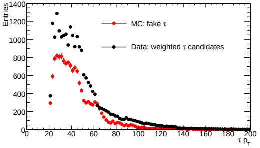
- Efficiency: taken from MC (for now ...)
- Rejection: taken from MC (for $t\bar{t}$), data (for QCD, Z events)



$t\bar{t}$



QCD



Z events

- Estimation of fake τ-jets in $t\bar{t}$ events well under control
- Space for improvement using em radius bins
- This can be done if there is more statistics available

Conclusion and Summary

- τ-jets might be the key to Higgs and new Physics beyond SM
- the identification works with a likelihood function combining several discriminating variables
- the identification efficiency and rejection can be determined directly from the data using Z and QCD events
- this can be used to estimate backgrounds in difficult channels like $t\bar{t}$