$Z \rightarrow \tau\tau$ visible mass analysis in the first data

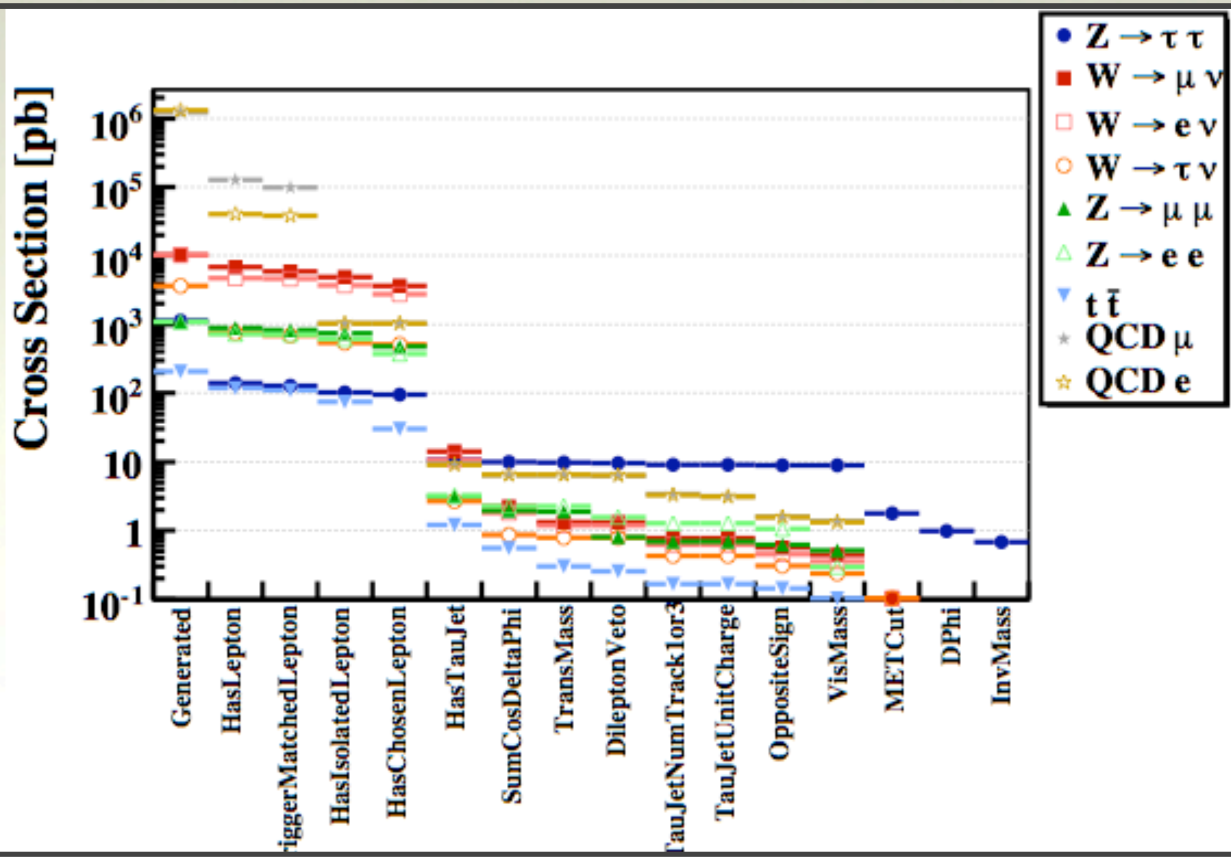
Outline

- The $Z \rightarrow \tau\tau$ visible mass analysis at 10 TeV:
  - Tau-jet scale determination,
  - The problem of the background and the present strategy to subtract it
- An alternative method for the background evaluation using track multiplicities
  - Control regions for the background: QCD and $W \rightarrow l\nu$
  - First try on estimation…and difficulties
- Technicalities
- Conclusions and prospects for the 7 TeV analysis in MC and data

$Z \rightarrow \tau\tau \rightarrow l\tau_{had}$ visible mass analysis

Select a high purity sample → very strong cuts
- Control samples for channels containing τ
- Tau-jet scale determination
- Measurement of the cross section pp → Z x BR($Z \rightarrow \tau\tau \rightarrow l\tau_{had}$)

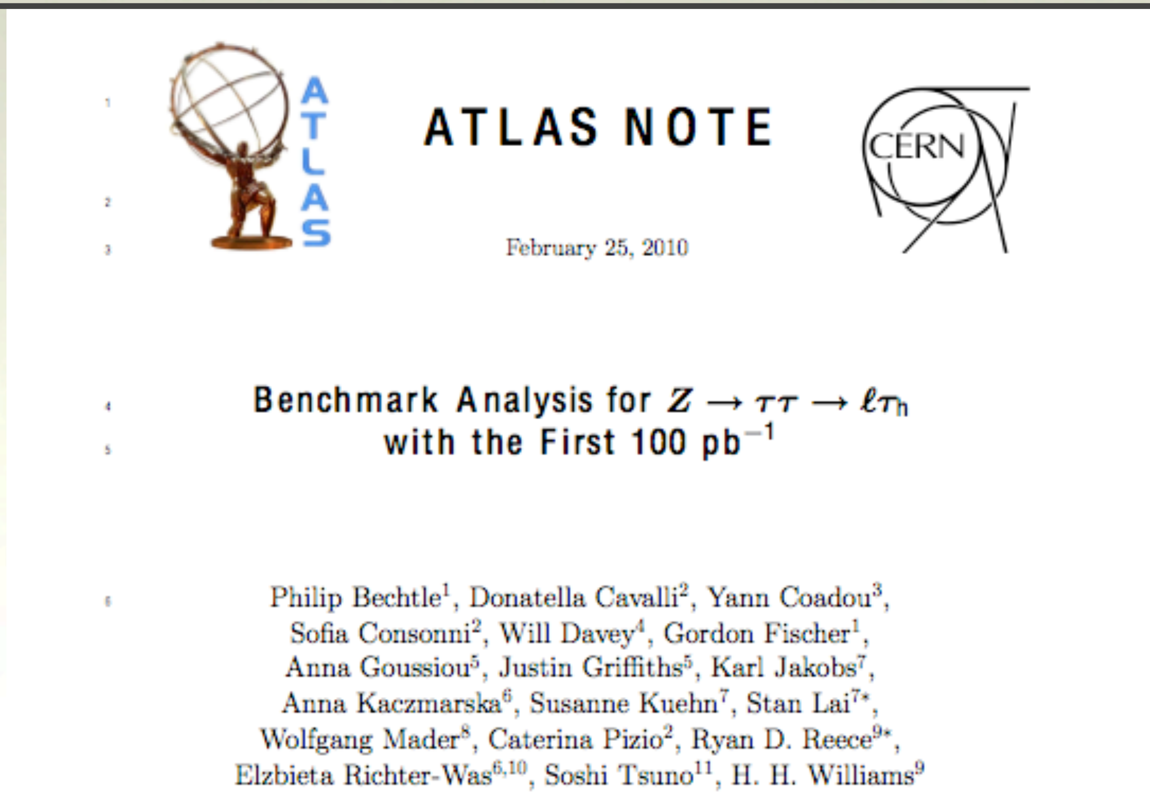
Note submitted last week (10 TeV): Benchmark Analysis for $Z \rightarrow \tau\tau \rightarrow l\tau_{had}$ with the First 100 pb^{-1}



$Z \rightarrow \tau\tau \rightarrow l\tau_{had}$ visible mass analysis

Select a high purity sample → very strong cuts
- Control samples for channels containing τ
- Tau-jet scale determination
- Measurement of the cross section pp → Z x BR($Z \rightarrow \tau\tau \rightarrow l\tau_{had}$)

Note submitted last week (10 TeV): Benchmark Analysis for $Z \rightarrow \tau\tau \rightarrow l\tau_{had}$ with the First 100 pb^{-1}



ATLAS NOTE

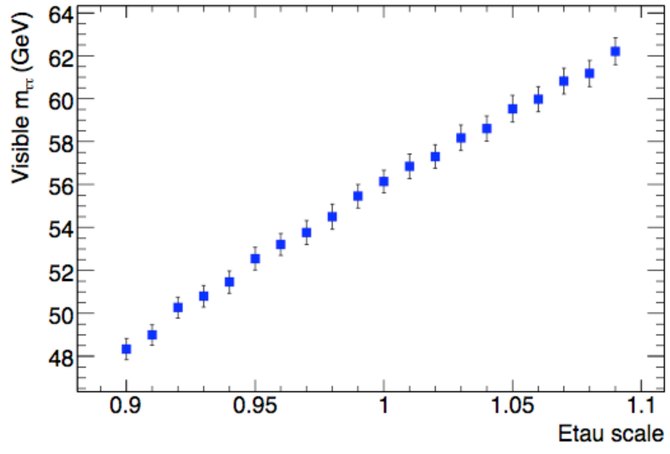
February 25, 2010

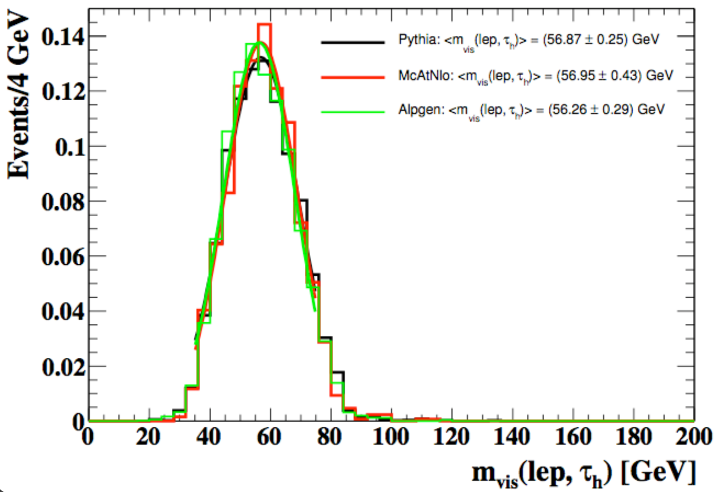
Benchmark Analysis for $Z \rightarrow \tau\tau \rightarrow \ell\tau_h$ with the First 100 pb^{-1}

Philip Bechtle[1], Donatella Cavalli[2], Yann Coadou[3], Sofia Consonni[2], Will Davey[4], Gordon Fischer[1], Anna Goussiou[5], Justin Griffiths[5], Karl Jakobs[7], Anna Kaczmarska[6], Susanne Kuehn[7], Stan Lai[7*], Wolfgang Mader[8], Caterina Pizio[2], Ryan D. Reece[9*], Elzbieta Richter-Was[6,10], Soshi Tsuno[11], H. H. Williams[9]

Tau-jet scale determination

Strategy: reconstruct the invariant mass of the visible decay products, proportionality to the tau-jet scale

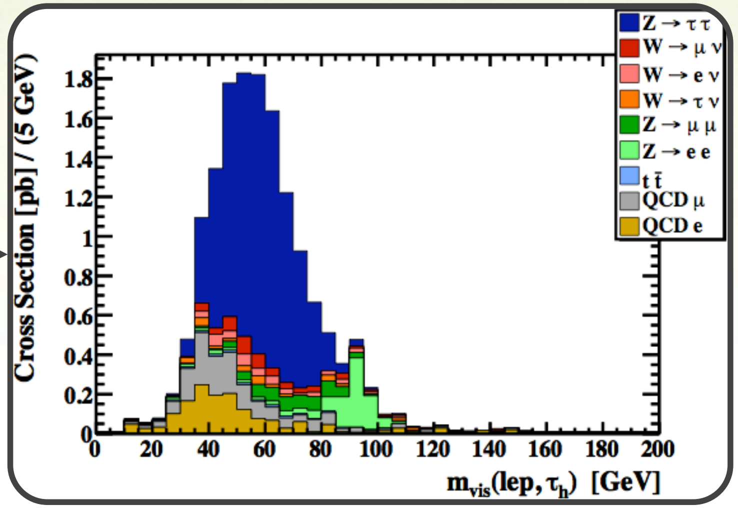




Reference from Montecarlo Truth: 58.86 GeV, 1% systematics (different MCs)

Keep control over the residual backgrounds

- Evaluate bg, especially QCD
- Subtraction procedure: OS-SS events (delicate for W+jets)

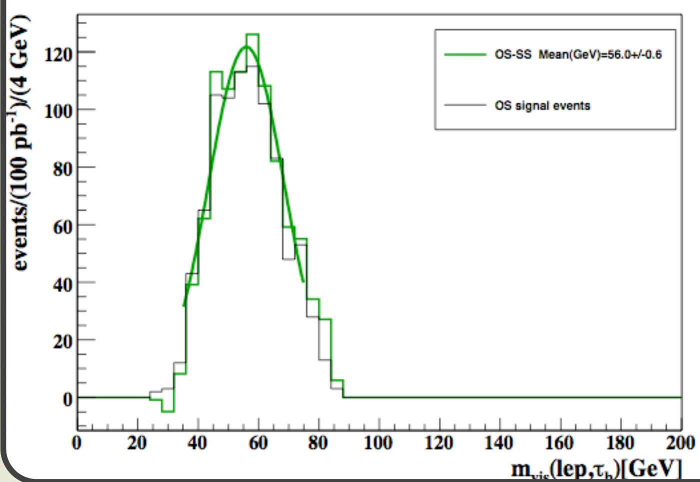


Background subtraction: OS-SS

$Z \to \tau\tau$ SameSign events: negligible (0.12 events in 100 pb^{-1}, VisibleMass in [0, 200] GeV)

QCD contributes in a similar way to OS and SS
- OS: 3.2 ± 0.4 events
- SS: 3.4 ± 0.4 events (1 pb^{-1})
- Not possible to check the visible mass shape is the same (low statistics)

$W \to l\nu$ + jets is not OS-SS symmetric
- can correct for the number (using a control region)
  - W SS correction: 4% discrepancy wrt real value
  - SS correction (taking into account all the other non qcd backgrounds): 27% discrepancy wrt real value
- use a flat shape for the correction



Fitting in the [35, 75] GeV VisibleMass range, peak position:

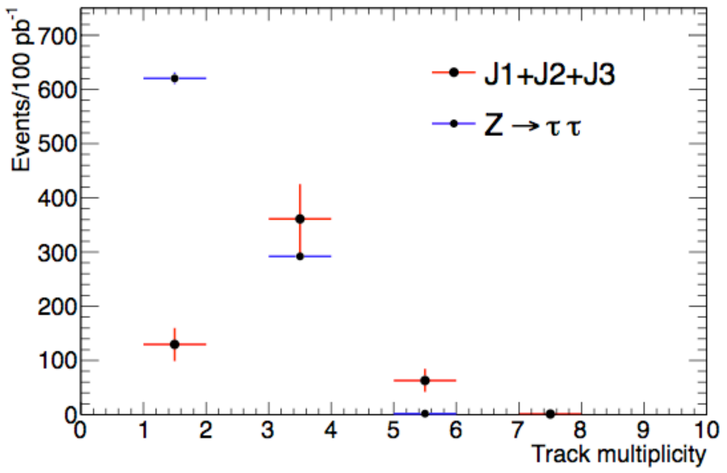
OS signal: 56.68 ± 0.61 GeV
OS-SS signal + background (no QCD): 56.0 ± 0.6 GeV

1.5 % shift in the peak value

In the following another complementary method to evaluate the backgrounds, based on track multiplicities, is shown. This method may be used even in the $W \to \tau_{had}\,\nu$

The method: track multiplicity in signal and QCD

Real taus → low track multiplicity
(signal region: 1 or 3 tracks)
Fakes from QCD → high track multiplicity



Strategy:

- count events with high track multiplicity (mainly QCD background)
- use them to evaluate the 1 or 3 track region background contamination (QCD contribution)

$$QCD = High|_{signal\,region} \cdot \frac{1or3}{High}\bigg|_{control\,region}$$

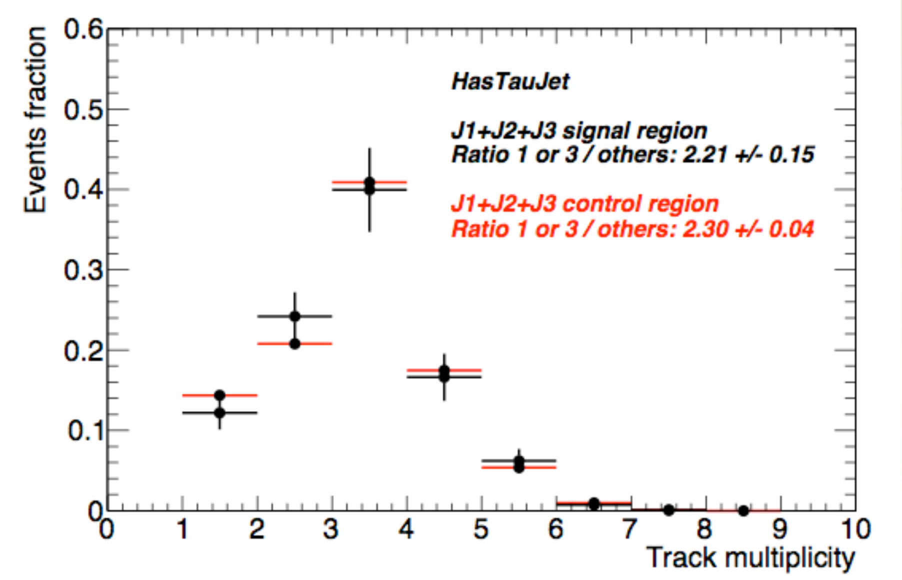
- How to measure the low/high track multiplicity ratio for QCD events?
- How is the high track multiplicity region contaminated by signal and other backgrounds?

A QCD control region

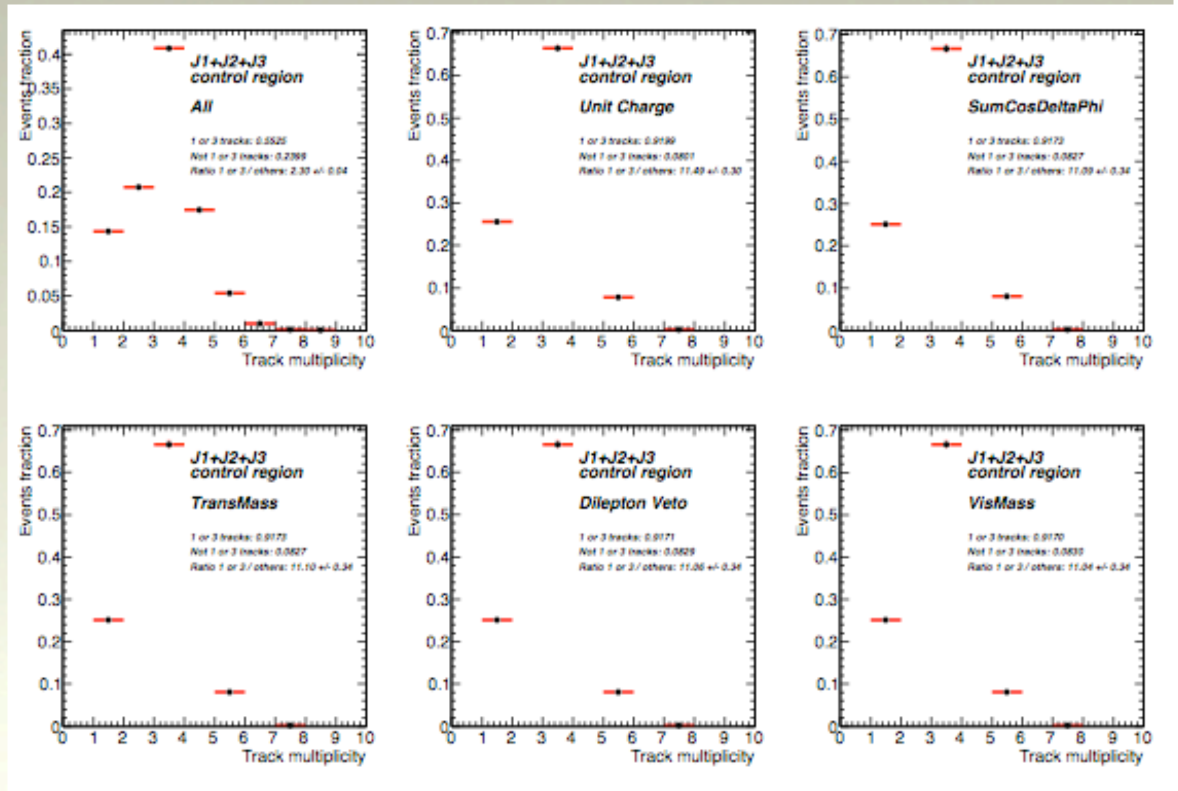
Strategy:
Ask the tag lepton to fail isolation
• signal and all other backgrounds have isolated leptons → almost pure QCD sample
• necessary to work at the beginning of the cut flow to save statistics
• but: very low statistics anyway in available MC samples

Improving the control region:
• Factorize Tau Identification → crucial: verify this does not change Low/High
• Move Unit Charge cut immediatly after HasTauJet and use the shape at this level
→ crucial: verify following cuts do not change the shape further

	Muon	Electron
Isolated (signal)	$etcone40/p_T < 0.1$ $nucone40 \leq 0$	$etcone30/p_T < 0.12$ $nucone40 \leq 0$
Non Isolated (control)	$etcone40/p_T \geq 0.1$ $nucone40 > 0$	$etcone30/p_T \geq 0.12$ $nucone40 > 0$



A QCD control region



Stability of the ratio
after UnitCharge!
ratio = 11.49 ± 0.30

Adding up signal and all the other backgrounds to the control region:

ratio = 11.41 ± 0.29 → 7 ‰ difference ! Very pure control region

Problems with ID-factorization

• Tau-ID does not contain an explicit dependence on track multiplicity, but nevertheless favours low track number candidates (variables built from first three leading tracks)
• Factorization does not take into account this fact properly

Asking only for a lepton and a Tau Jet

$\text{Ratio}_{\text{ID factorized}} = 2.30 \pm 0.03$
$\text{Ratio}_{\text{ID not factorized}} = 4.29 \pm 0.73$

Try to work without factorizing ID: try using JF17 (high statistics, $\sim 10^7$ events)

Too low statistics anyway:
• 6 events surviving for the signal region
• 0 events in the signal region, high track multiplicity
• 99 events surviving in the control region
• 5 events in the control region, high track multiplicity

A W control region

We have not only QCD in the signal region:
- About 40 % of the background is non-QCD, of which about 50 % is due to $W \rightarrow l\nu$
- W background is mainly due to additional jets faking taus
- Can we evaluate even the W background using this method? Is the ratio the same?

Strategy for a W control region:
- Ask for events with high transverse mass: TransMass in [60, 90] GeV instead of < 50 GeV
- Reverse SumCosDeltaPhi cut: SumCosDeltaPhi > -0.15
- Pure control region (4 % contamination due to signal and other backgrounds)

$\text{Ratio}_{\text{signal region}} = 31.5 \pm 11.7$
$\text{Ratio}_{\text{control region}} = 28.0 \pm 4.3$

Compatibility between signal and control region

No compatibility between QCD (ratio = 11.49 ± 0.30) and W
Hypothesis: different kinds of jets
J filtered samples → rich in b's → high track multiplicity
W samples: rich in light flavour jets → lower track multiplicity

Difficulties in estimating the background

Try to evaluate QCD background in the signal region:
• Use ID-factorization anyway (ratio underestimation)
• no OppositeSign requirement
• VisibleMass range [0, 200] GeV

Real value = 490 ± 71
Estimate = 739 ± 243
Discrepancy = (51± 54) %

Erros are dominated by the low statistics in the signal region, high multiplicity

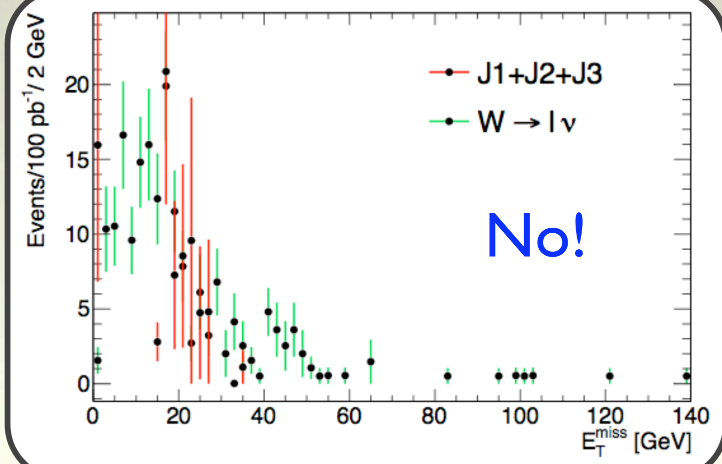
Try to evaluate W background in the signal region:
• no OppositeSign requirement
• VisibleMass range [0, 200] GeV

Real value = 181 ± 11
Estimate = 166 ± 66
Discrepancy = (- 8 ± 37) %

Difficulties in estimating the background

Try to evaluate
- QCD, $Z \rightarrow ee$, $Z \rightarrow \mu\mu$ using the QCD ratio
- $W \rightarrow l\nu$ using the W ratio

Real value = 869 ± 72
Estimate = 929 ± 252
Discrepancy = (- 6.2 ± 30) %



Problem: how is it possible to distinuish W from QCD events in the signal region, high track multiplicity? Use $E_{T\ miss}$?

As a last possibility: evaluate all the jet-like backgrounds using the QCD ratio (dominant background).

Real value = 869 ± 72
Estimate = 828 ± 245
Discrepancy = (- 4.7 ± 29) %

Discrepanices compensate by chance? If evaluated separately QCD is overstimated, W is underestimated

Technicalities

Benchmark analysis for $Z \rightarrow \tau\tau$ (in svn): very slow to rerun

Ntuple
- selected objects
- flags to reproduce the cut flow

Python scripts:
- can reproduce the analysis using the flags
- difficulties running on many files
- not trivial to change cuts that involve object choices

C ++ analysis (with Susanne Kuhen, Freiburg):
- fully reproduce the benchmark cut flow
- possibility to play with object choices
- still limited by what goes in the ntuple (no possibility to change preselection, overlap removal, truth matching…)

D3PD (TauD3PD maker in svn)

C ++ analysis adapted (with Susanne):
- fully reproduce the benchmark cut flow
- possibility to play with object choices
- possibility to play with all preselections and matchings

work in progress…

Summarising…

A benchmark analysis for visible mass $Z \rightarrow \tau\tau$ is ready: results at 10 TeV

- Background subtraction using SS events
- Complementary background evaluation using track multiplicity
    - It is possible to define a pure control region for QCD, in which tau-jet fakes have the same 1or3 tracks / >3 tracks ratio
    - It is possible to define a pure control region for $W \rightarrow l\nu$, in which the tau-jet fakes have the same 1or3 tracks / >3 tracks ratio
    - Ratios for $W \rightarrow l\nu$ and QCD are different
    - If it were possible to distinguish W from QCD the jet-like backgrounds could be estimated within statistical errors. These errors, though, are very big (~ 30 %)
    - MC statistics for QCD samples does not allow to get to better conclusions

Plans for the future

- Re-make the analysis at 7 TeV MC full visible using D3PDs
- Look at data using D3PDs and the same analysis program
- Study further issues related to the tau-jet scale determination (generator level studies with Powegh)
- Use the OS-SS method to subtract the background and check the numbers with the complementary method

Backup slides

$Z \to \tau\tau \to l\tau_{had}$ visible mass analysis

	$Z \to \tau\tau$	$W \to l\nu$	$Z \to ll$	$t\bar{t}$	QCD Full μ	QCD Full electron	QCD AF-II
generated	1128(1)	24370(22)	2196(2)	205.5(4)	1212003(2790)	1302219(2800)	926090000(211000)
has lepton	141.1(5)	12300(14)	1923(2)	119.1(3)	127463(860)	40775(420)	181120(2480)
lepton passed trigger	129.0(5)	11264(13)	1550(2)	109.6(3)	97394(750)	38393(420)	26733(713)
lepton passed isolation	102.4(4)	9256(11)	1359(2)	74.4(2)	1039(25)	1043(25)	2213(74)
$p_{\mathrm{T}}(\ell) < 40$ GeV	95.0(4)	6976(11)	851(13)	30.1(2)	1039(25)	1043(25)	2213(74)
has tau-jet	10.6(1)	28 (7)	6.4(1)	1.20(3)	9.3(3)	9.1(4)	21(1)
$\Sigma \cos \Delta\phi > -0.15$	9.9(1)	4.9(3)	4.2(1)	0.55(2)	6.4(3)	6.5(3)	15.5(9)
$m_{\mathrm{T}}(\ell, E_{\mathrm{T}}^{\mathrm{miss}}) < 50$ GeV	9.7(1)	3.3(3)	4.1(1)	0.30(2)	6.4(3)	6.5(3)	15.5(9)
dilepton veto	9.7(1)	3.3(3)	2.38(7)	0.26(1)	6.4(3)	6.4(3)	15.5(9)
$N_{\mathrm{tracks}}(\tau_{\mathrm{h}}) = 1$ or 3	9.1(1)	1.8(2)	1.98(7)	0.17(1)	3.4(3)	3.3(3)	8.5(7)
$\lvert\mathrm{Charge}(\tau_{\mathrm{h}})\rvert = 1$	9.1(1)	1.8(2)	1.98(7)	0.17(1)	3.1(2)	3.1(3)	7.8(7)
opposite sign	9.0(1)	1.3(2)	1.68(6)	0.14(1)	1.6(1)	1.5(2)	3.7(5)
$m_{\mathrm{vis}}(\ell, \tau_{\mathrm{h}}) = 25 - 85$ GeV	8.9(1)	1.0(2)	0.81(4)	0.104(9)	1.4(1)	1.3(1)	3.2(5)

Background subtraction: OS-SS

	$Z \to \tau\tau$	$W \to l\nu$	$Z \to ll$	$t\bar{t}$	Signal+bkg	All bkg
OS SR	8.9	1.04	0.81	0.1	10.85	1.95
SS SR	0.12	0.4	0.17	0.01	0.7	0.57
OS CR	0.01	3.67	0.04	0.1	3.81	3.81
SS CR	0	1.22	0.03	0.03	1.28	1.28

$$\text{Estimated } W\ OS_{SR} = SS_{CR} * (OS/SS)_{CR} * \frac{OS^W_{SR}}{OS^W_{CR}}$$

1.08 pb

$$\text{Number of events to be subtracted} = SS_{SR} + SS_{CR} * \left(\frac{OS^W_{SR}}{OS^W_{CR}} * ((OS/SS)_{CR} - 1) \right)$$

1.42 pb

Tau CutSafe ID variables

Calo-track variables

$$\frac{E_{\mathrm{T}}^{\mathrm{total}}}{p_{\mathrm{T1}}} = \frac{\sum_i E_{\mathrm{T},i}^{\mathrm{EM}} + \sum_i E_{\mathrm{T},i}^{\mathrm{Had}}}{p_{\mathrm{T1}}}$$

$$\frac{E_{\mathrm{T}}^{\mathrm{EM}}}{p_{\mathrm{T}}^{\mathrm{total}}} = \frac{\sum_i E_{\mathrm{T},i}^{\mathrm{EM}}}{\sum_{j=1}^{N} p_{\mathrm{T},j}^{track}}$$

$$\frac{E_{\mathrm{T}}^{\mathrm{Had}}}{p_{\mathrm{T}}^{\mathrm{total}}} = \frac{\sum_i E_{\mathrm{T},i}^{\mathrm{Had}}}{\sum_{j=1}^{N} p_{\mathrm{T},j}^{\mathrm{track}}}$$

$$\frac{p_{\mathrm{T}}^{\mathrm{total}}}{E_{\mathrm{T}}^{\mathrm{total}}} = \frac{\sum_{k=1}^{N} p_{\mathrm{T},k}^{\mathrm{track}}}{\sum_i E_{\mathrm{T},i}^{\mathrm{EM}} + \sum_j E_{\mathrm{T},j}^{\mathrm{Had}}}$$

$$W_{\mathrm{track}}^{\tau} = \frac{\sum \left(\Delta R^{\mathrm{track}}\right)^2 \cdot p_{\mathrm{T}}^{\mathrm{track}}}{\sum p_{\mathrm{T}}^{\mathrm{track}}} - \frac{\left(\sum \Delta R^{\mathrm{track}} \cdot p_{\mathrm{T}}^{\mathrm{track}}\right)^2}{\left(\sum p_{\mathrm{T}}^{\mathrm{track}}\right)^2}$$

Calo variables

$$R_{\mathrm{EM}} = \frac{\sum_{i=1}^{\Delta R<0.4} E_{\mathrm{T},i}^{\mathrm{EM}} \sqrt{\left(\eta_i^{\mathrm{EM}} - \eta_{\mathrm{caloseed}}\right)^2 + \left(\phi_i^{\mathrm{EM}} - \phi_{\mathrm{caloseed}}\right)^2}}{\sum_{i=1}^{\Delta R<0.4} E_{\mathrm{T},i}^{\mathrm{EM}}}$$

$$W_{\mathrm{strip}} = \sqrt{\frac{\sum_{i}^{\Delta R<0.4} E_{\mathrm{T},i}^{\mathrm{strip}} \left(\eta_i - \eta_{\mathrm{caloseed}}\right)^2}{\sum_{i}^{\Delta R<0.4} E_{\mathrm{T},i}^{\mathrm{strip}}}}$$

$$f_{\mathrm{iso}} = \frac{\sum_{i}^{0.1<\Delta R<0.2} E_{\mathrm{T},i}^{\mathrm{EM}}}{\sum_{j}^{\Delta R<0.4} E_{\mathrm{T},j}^{\mathrm{EM}}}$$

$$\frac{E_{\mathrm{T}}^{\mathrm{EM}}}{E_{\mathrm{T}}^{\mathrm{total}}} = \frac{\sum_{i}^{\Delta R<0.4} E_{\mathrm{T},i}^{\mathrm{EM}}}{\sum_{i}^{\Delta R<0.4} E_{\mathrm{T,i}}^{\mathrm{EM}} + \sum_{j}^{\Delta R<0.4} E_{\mathrm{T},j}^{\mathrm{Had}}}$$