Tau Identification and Search for SM $H \to \tau\tau$ with ATLAS

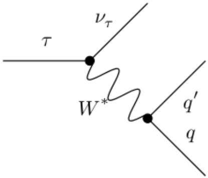
Théo MEGY
Albert-Ludwigs-Universität Freiburg
12.09.2017

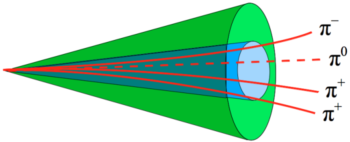
Outline

1. τ_{had} identification
   - Reconstruction
   - Identification
   - Measurement
2. SM H$\rightarrow \tau\tau$ Analysis
   - Motivation
   - Run 2 analysis
   - Impact of τ_{had} identification
3. Conclusion

τ_{had} reconstruction

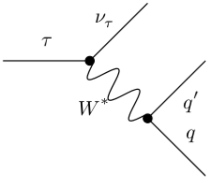
- Hadronic decay in 65% of cases
- Forms a **narrow jet**
    - 1 or 3 charged pions (prong)

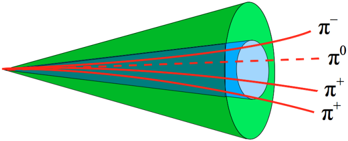




τ_{had} reconstruction

- Hadronic decay in 65% of cases
- Forms a **narrow jet**
  - 1 or 3 charged pions (prong)



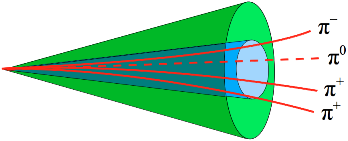


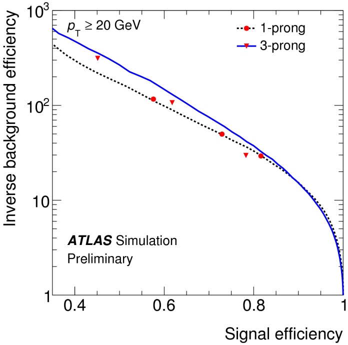
- $\tau_{\mathbf{had}}$ **candidate reconstruction**:
  - Seed: jet formed using anti-k_t algorithm
  - Track matched to τ_{had} if ΔR<0.2 and p_T>1 GeV
  - Other leptons or jets can be reconstructed as τ_{had} $\Rightarrow$ Fakes

τ_{had} identification

- **Identification**:
  - Discriminate against jet fakes
  - Boosted Decision Tree using info from internal structure of τ_{had} candidate
  - True τ_{had}: low track multiplicity, narrow calorimeter activity, etc

- Examples of input variable:
  - Fraction of tracks p_T in isolation region





- Separate BDT for 1 and 3 prong

τ_{had} identification efficiency measurement

- **Measure the identification efficiency from ATLAS data**

- Signal: $Z \rightarrow \tau\tau$
- *Tag-and-probe* approach:
  - Select events triggered by a muon (tag) and containing a τ_{had} candidate (probe) before Id
- Backgrounds: W+jet, $Z \rightarrow ll$, top (t$\bar{t}$), Multijet

- Extract efficiency:
  - Compare number of reconstructed τ_{had} before and after identification

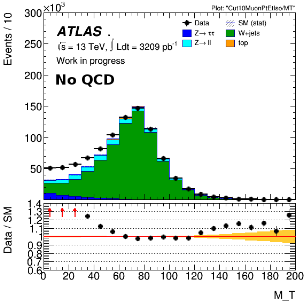
Event Selection

- Goal: Select $Z \rightarrow \tau(\rightarrow \mu\nu_\mu\nu_\tau)\ \tau_{had}$ events
- Pre-selection applied:
  - 1 single trigger matched muon with $p_T \geq 22$ GeV
  - Electron veto
  - At least 1 τ_{had} candidate with |q|=1, 1 or 3 tracks and $p_T \geq 20$ GeV
    $\rightarrow$ Pick candidate with highest p_T

Event Selection

- Goal: Select $Z \rightarrow \tau(\rightarrow \mu\nu_\mu\nu_\tau)\ \tau_{had}$ events
- Pre-selection applied:
  - 1 single trigger matched muon with $p_T \geq 22$ GeV
  - Electron veto
  - At least 1 τ_{had} candidate with |q|=1, 1 or 3 tracks and $p_T \geq 20$ GeV
    $\rightarrow$ Pick candidate with highest p_T

- Signal Region:
  $M_{\mathrm{T}} \leq 50$ GeV
  SumCosDPhi $\geq$ -0.1
  $42 \leq$ visible mass $(\mu, \tau) \leq 82$ GeV



$$M_{\mathrm{T}} = \sqrt{2p_T(\mu).E_{\mathrm{T}}^{\mathrm{miss}}.(1 - cos\Delta\Phi(\mu, E_{\mathrm{T}}^{\mathrm{miss}}))}$$

SumCosDphi$= cos\Delta\Phi(\mu, E_{\mathrm{T}}^{\mathrm{miss}}) + cos\Delta\Phi(\tau_{had}, E_{\mathrm{T}}^{\mathrm{miss}})$

Event Selection

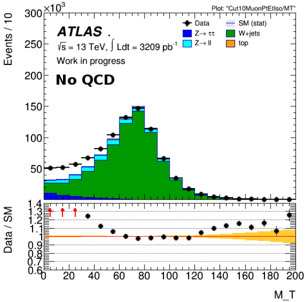
- Goal: Select $Z \rightarrow \tau(\rightarrow \mu\nu_\mu\nu_\tau)\ \tau_{had}$ events
- Pre-selection applied:
  - 1 single trigger matched muon with $p_T \geq 22$ GeV
  - Electron veto
  - At least 1 τ_{had} candidate with |q|=1, 1 or 3 tracks and $p_T \geq 20$ GeV
    $\rightarrow$ Pick candidate with highest p_T

- Signal Region:

  $M_{\mathrm{T}} \leq 50$ GeV
  SumCosDPhi $\geq$ -0.1
  $42 \leq$ visible mass $(\mu, \tau) \leq 82$ GeV

- W+jet Control Region:

  $M_{\mathrm{T}} \geq 60$ GeV
  $E_{\mathrm{T}}^{\mathrm{miss}} \geq 30$ GeV
  SumCosDPhi ≤ 0



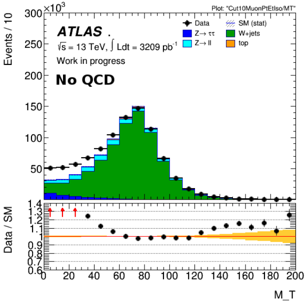
$$M_{\mathrm{T}} = \sqrt{2p_T(\mu).E_{\mathrm{T}}^{\mathrm{miss}}.(1 - cos\Delta\Phi(\mu, E_{\mathrm{T}}^{\mathrm{miss}}))}$$

SumCosDphi=$cos\Delta\Phi(\mu, E_{\mathrm{T}}^{\mathrm{miss}}) + cos\Delta\Phi(\tau_{had}, E_{\mathrm{T}}^{\mathrm{miss}})$

Event Selection

- Goal: Select $Z \rightarrow \tau(\rightarrow \mu\nu_\mu\nu_\tau)\ \tau_{had}$ events
- Pre-selection applied:
  - 1 single trigger matched muon with $p_T \geq 22$ GeV
  - Electron veto
  - At least 1 τ_{had} candidate with |q|=1, 1 or 3 tracks and $p_T \geq 20$ GeV
    $\rightarrow$ Pick candidate with highest p_T

- Signal Region:
  $M_{\mathrm{T}} \leq 50$ GeV
  SumCosDPhi $\geq$ -0.1
  $42 \leq$ visible mass $(\mu, \tau) \leq 82$ GeV
- W+jet Control Region:
  $M_{\mathrm{T}} \geq 60$ GeV
  $E_{\mathrm{T}}^{\mathrm{miss}} \geq 30$ GeV
  SumCosDPhi ≤ 0
- Multijet Control Region
  SR with inverted muon isolation



$$M_{\mathrm{T}} = \sqrt{2p_T(\mu).E_{\mathrm{T}}^{\mathrm{miss}}.(1 - cos\Delta\Phi(\mu, E_{\mathrm{T}}^{\mathrm{miss}}))}$$

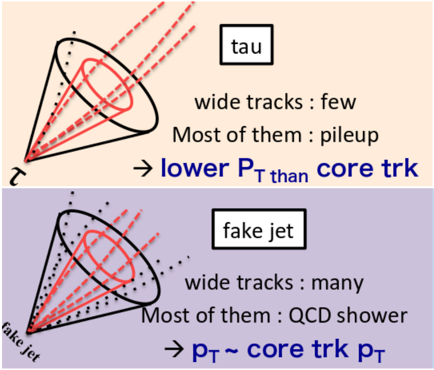
SumCosDphi=$cos\Delta\Phi(\mu, E_{\mathrm{T}}^{\mathrm{miss}}) + cos\Delta\Phi(\tau_{had}, E_{\mathrm{T}}^{\mathrm{miss}})$

Track multiplicity fit method

Track multiplicity $=$

Core tracks $+$
Outer tracks satisfying
$\min(\frac{p_T^{core}}{p_T^{outer}} \times \Delta R(core, outer)) \leq 4$

Core: $0 \leq \Delta R \leq 0.2$
Outer: $0.2 \leq \Delta R \leq 0.6$

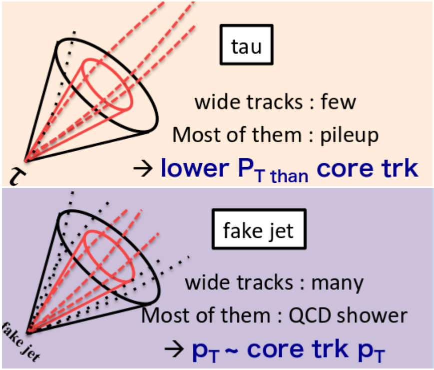


Track multiplicity fit method

Track multiplicity =

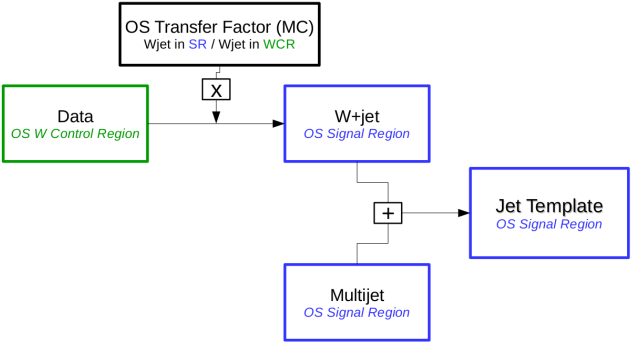
Core tracks +
Outer tracks satisfying
$\min(\frac{p_T^{core}}{p_T^{outer}} \times \Delta R(core, outer)) \leq 4$

Core: $0 \leq \Delta R \leq 0.2$
Outer: $0.2 \leq \Delta R \leq 0.6$

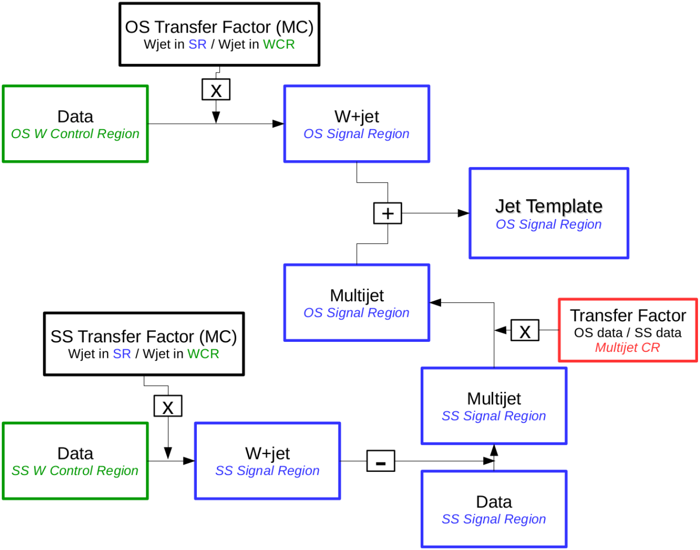


- Definition of 3 templates:
  - **Tau template** = MC $Z \rightarrow \tau\tau$ and top($t\bar{t}$) with truth matched τ's
  - **Lepton template** = MC $Z \rightarrow ll$, top and $Z \rightarrow \tau\tau$ with lepton fake τ's
  - **Jet template** (W+jet and Multijet) $\Rightarrow$ Data driven

Construction of the Jet template



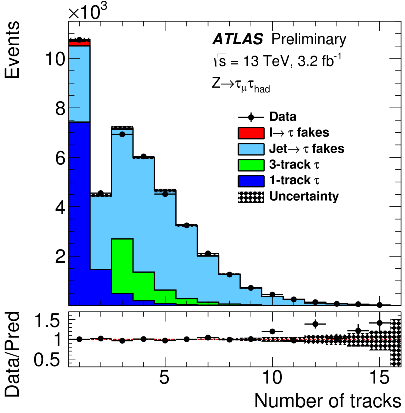
Construction of the Jet template



Pre-ID fit - Results

- **Fit in pre-ID region**
  - Float Tau templates with 1 common parameter
  - Float Jet template
  - Constrain Lepton template to MC prediction

- **Fit in pre-ID region**
  - Float Tau templates with 1 common parameter
  - Float Jet template
  - Constrain Lepton template to MC prediction



$$\chi^2/ndf = 1.278$$

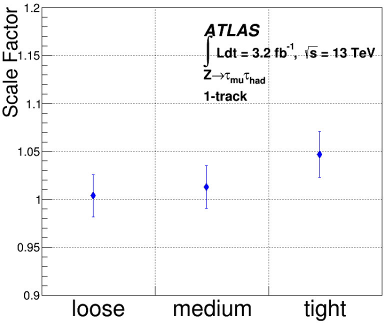
Tau normalisation factor $= 0.960 \pm 0.015$

Jet normalisation factor $= 0.970 \pm 0.011$

τ_{had} identification efficiency

- To get Tau ID efficiency:
  - Build templates in passed ID region (1 and 3 prong)
  - Apply jet normalisation factor extracted from pre-ID fit
  - Tau = data - jet - lepton

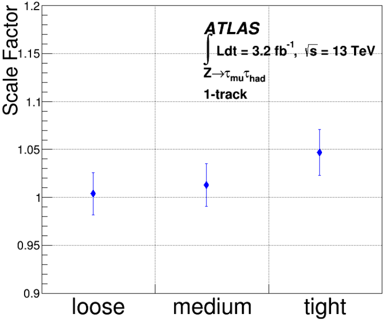
- Tau ID efficiency $= \dfrac{Tau\ \ Passed\ \ ID}{Tau\ \ Before\ \ ID}$
- Scale Factor $= \dfrac{data\ \ efficiency}{MC\ \ efficiency}$



τ_{had} identification efficiency

- To get Tau ID efficiency:
  - Build templates in passed ID region (1 and 3 prong)
  - Apply jet normalisation factor extracted from pre-ID fit
  - Tau = data - jet - lepton

- Tau ID efficiency $= \frac{Tau\ \ Passed\ \ ID}{Tau\ \ Before\ \ ID}$
- Scale Factor $= \frac{data\ \ efficiency}{MC\ \ efficiency}$



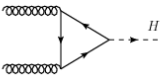
- **Currently**:
  - New method being developed
  - Simultaneous fit in all ID regions
  - SF binned in p_T^τ

Outline

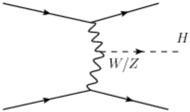
1. τ_{had} identification
   - Reconstruction
   - Identification
   - Measurement
2. **SM H$\rightarrow \tau\tau$ Analysis**
   - Motivation
   - Run 2 analysis
   - Impact of τ_{had} identification
3. Conclusion

SM H$\rightarrow \tau\tau$ Analysis - Motivation

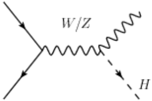
- **Why searching for H$\rightarrow \tau\tau$?**
  - Want to measure Higgs boson coupling to fermions
  - Clear signature $\Rightarrow$ Exploit **gluon Fusion** (ggF) + **Vector Boson Fusion** (VBF)



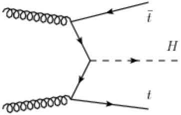
GGF (88%)



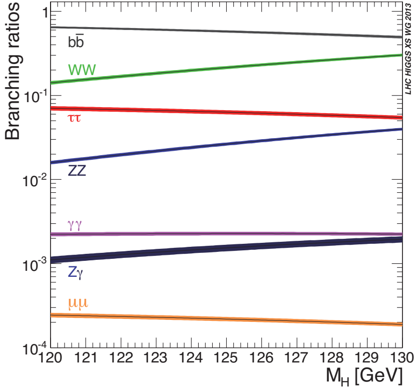
VBF (6.6%)



VH (5%)

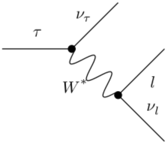


ttH (0.4%)

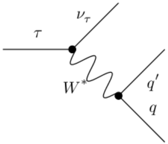


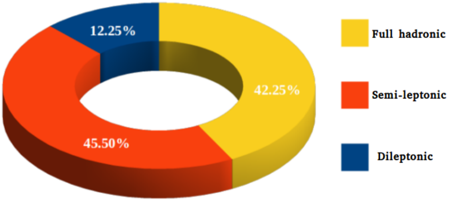
SM H$\to \tau\tau$ Analysis - Tau decay

Leptonic decay



Hadronic decay

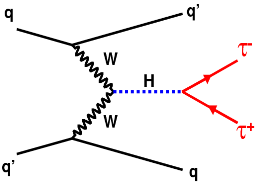


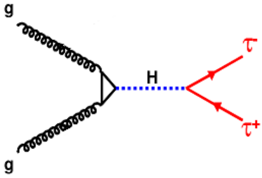


3 channels to investigate and combine

- ATLAS Run 1 analysis $\Rightarrow$ 4.5 σ significance ; ATLAS + CMS $\Rightarrow$ 5.5 σ
- Run 2: Rediscover H$\to \tau\tau$ with ATLAS at $\sqrt{s}$ = 13 TeV
  $\Rightarrow$ Cut-based and multivariate analysis pursued
- CMS Run 2 $\Rightarrow$ 4.9 σ ; Run 1 + Run 2 $\Rightarrow$ 5.9 σ

SM H$\rightarrow \tau\tau$ Analysis - Analysis categories





- **VBF category**

  Particular topology: 2 forward high p_T jets
  - $N_{jets} \geq 2$, $p_T^{jet0} > 40$ GeV and $p_T^{jet1} > 30$ GeV
  - $\Delta\eta_{jj} \geq 3$, $m_{jj} \geq 400$ GeV
  - Taus between jets' pseudo-rapidity gap

- **Boosted category**

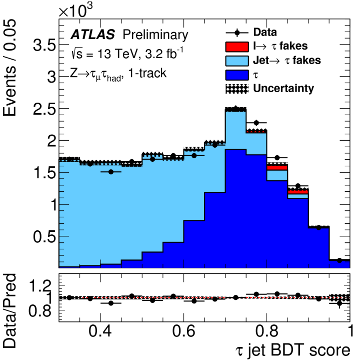
  Enriched in ggF events
  - Failing VBF selection
  - $p_T^H > 100$ GeV

Outline

Impact of τ_{had} identification

- Can we improve the sensitivity/significance by changing τ_{had} ID conditions ?
- Study performed in **semi-leptonic channel** with 2015+2016 dataset

1 Prong

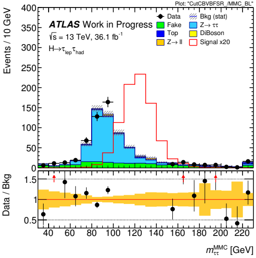


- Currently used: **Medium** WP
  (cut at τ jet BDT score $\approx$ 0.6)
- Suppress fakes with tighter cut
- Try to apply **Continuous tau ID**
  Split SR in **Tight** ID and **Medium not Tight** ID

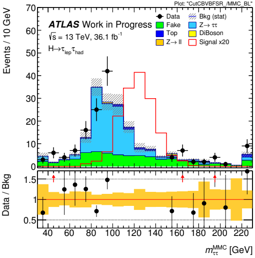
Continuous Tau ID

- Inclusive VBF signal region:

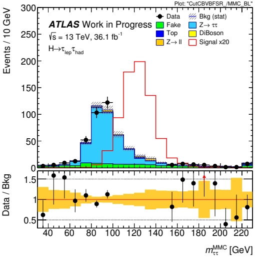
Medium



Medium not Tight



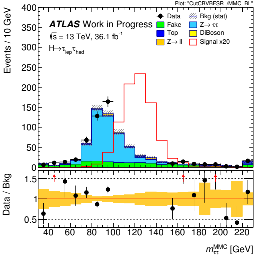
Tight



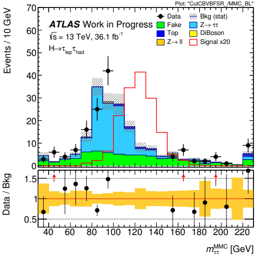
Continuous Tau ID

- Inclusive VBF signal region:

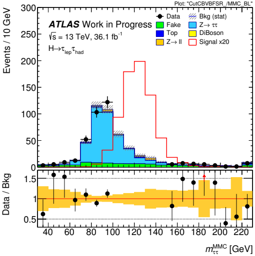
Medium



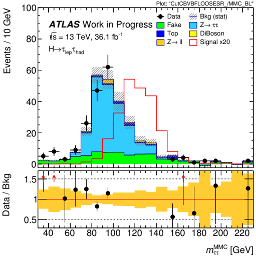
Medium not Tight

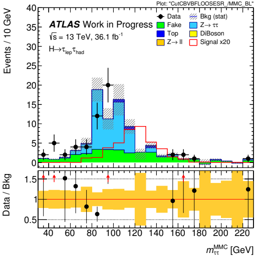


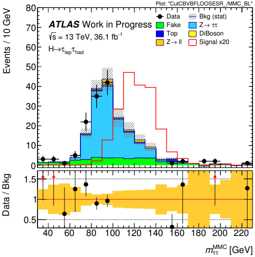
Tight



- Low VBF signal region:







Continuous Tau ID

- Perform fit in the semi-leptonic channel
  - Tau ID syst. for Medium and Tight treated independently
- Expected significance:
  - Increase in VBF from Medium to Continuous identification
  - Stable in Boosted
  - **Few percent increase** in combined fit

Continuous Tau ID

- Perform fit in the semi-leptonic channel
  - Tau ID syst. for Medium and Tight treated independently
- Expected significance:
  - Increase in VBF from Medium to Continuous identification
  - Stable in Boosted
  - **Few percent increase** in combined fit
- **But**: low statistics in Medium not Tight region

  $\Rightarrow$ Especially dramatic in Low VBF SR
- **Impact on the Fit**:
  - Automatic procedure for Nuisance Parameter treatment: prune, use shape and/or normalisation information
  - Compare treatment in Medium not Tight/Tight with Medium ($Z \rightarrow \tau\tau$)
  - **Medium not Tight**: several NP using only norm., instead of norm+shape
  - **Tight**: only few NP not using shape anymore

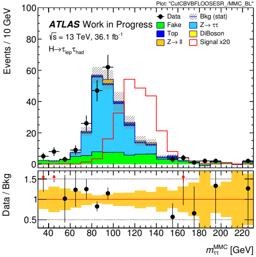
Continuous Tau ID - Very Tight ID

- **Define tighter Tau ID cut**
  - Split Medium region evenly
    + suppress fakes further
  - Scanned a few values
  - Optimal split obtained for **BDT score > 0.73**
  - Simplified case: specific 1/3 prong cuts normally needed
- No corresponding SF
  - Use Tight SF for Very Tight
  - Use Medium SF for Medium not Very Tight

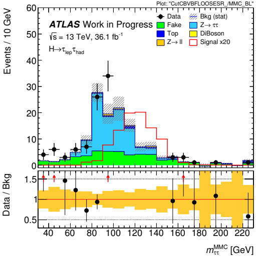
Continuous Tau ID - Very Tight ID

- Low VBF signal region:

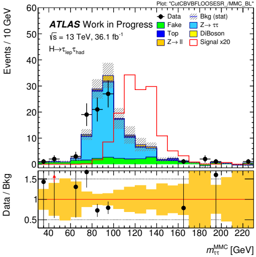
Medium



Medium not Very Tight



Very Tight (bdt > 0.73)



- Better fakes rejection in Very Tight region
- Improved statistics in Medium not Very Tight
- **Treatment of Nuisance Parameters** ($Z\rightarrow\tau\tau$):
  - Still lots of NP losing shape information (Low VBF SR)

Continuous Tau ID - Very Tight ID

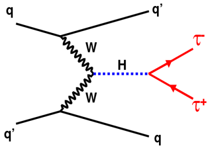
- Expected significance:
    - Increase in both VBF and Boosted
    - **6% increase** in the combined significance
- Caveats:
    - Non-optimised WP
    - Need dedicated tau ID SF
    - NP treatment $\rightarrow$ still suffer from low statistics

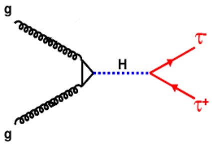
Conclusion

- **Tau identification in ATLAS**:
  - Select true τ_{had} and discriminate against jet fakes
  - Using multivariate algorithm
  - Need of dedicated measurement in ATLAS data
- **H$\rightarrow \tau\tau$ and tau identification**:
  - Expected significance raised by using continuous tau identification
  - Attempt to define tighter ID cut to fight low stats
  - Potential improvement of significance by 6%
- **Next steps**:
  - Define of fully optimised WPs (1 and 3 prong)
  - Derive corresponding SFs
  - Use continuous tau identification in MVA analysis

Backup

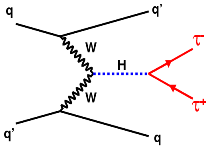
SM H$\rightarrow \tau\tau$ Analysis - Analysis categories

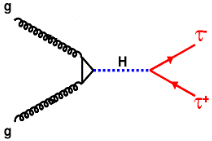




	VBF	**Boost**
	At least two jets with $p_T^{j1} > 40\,\text{GeV}$ and $p_T^{j2} > 30\,\text{GeV}$	Failing VBF selection
	$\Delta\eta_{jj} > 3$, $m_{jj} > 300\text{GeV}$	$p_T^H > 100\,\text{GeV}$
	leptons,taus between jets' pseudo-rapidity gap	

SM H$\rightarrow \tau\tau$ Analysis - Analysis categories

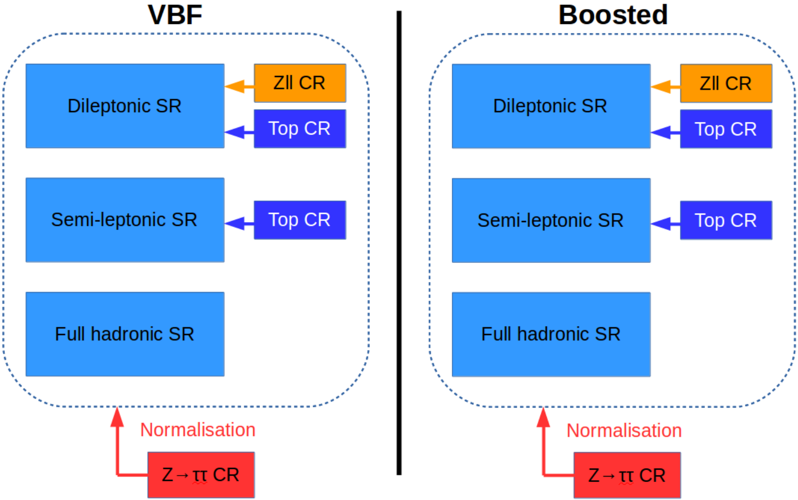




<table>
<tr><td></td><td colspan="3">VBF
At least two jets with $p_T^{j1} > 40$ GeV and $p_T^{j2} > 30$ GeV
$\Delta\eta_{jj} > 3$, $m_{jj} > 300$GeV
leptons,taus between jets' pseudo-rapidity gap</td><td colspan="2">Boost
Failing VBF selection
$p_T^H > 100$ GeV</td></tr>
<tr><td>ll</td><td>VBF High-pt
$p_T^H > 100$ GeV
and $m_{jj} > 400$GeV</td><td>VBF Low-pt
$p_T^H < 100$ GeV
and $m_{jj} > 400$GeV</td><td></td><td>Boost High-pt
$p_T^H > 140$ GeV</td><td>Boost Low-pt
$p_T^H < 140$ GeV</td></tr>
<tr><td>lh</td><td>VBF Tight
$m_{jj} > 500$GeV
and $p_T^H > 100$ GeV</td><td>VBF Loose
$m_{jj} < 500$GeV
or $p_T^H < 100$ GeV</td><td></td><td>Boost High-pt
$p_T^H > 140$ GeV and
$\Delta R_{l\tau} < 1.5$</td><td>Boost Low-pt
$p_T^H < 140$ GeV or
$\Delta R_{l\tau} > 1.5$</td></tr>
<tr><td>hh
$p_T^{j1} > 70$ GeV</td><td>VBF High-pt
$p_T^H > 140$ GeV and
$\Delta R_{\tau\tau} < 1.5$</td><td>VBF Low-pt Tight
$p_T^H < 140$ GeV or $\Delta R_{\tau\tau} < 1.5$ and
$m_{jj} > (-250\Delta\eta_{jj} + 1550)$ GeV</td><td>VBF Low-pt Loose

$m_{jj} < (-250\Delta\eta_{jj} + 1550)$ GeV</td><td>Boost High-pt
$p_T^H > 140$ GeV and
$\Delta R_{\tau\tau} < 1.5$</td><td>Boost Low-pt
$p_T^H < 140$ GeV or
$\Delta R_{\tau\tau} > 1.5$</td></tr>
</table>

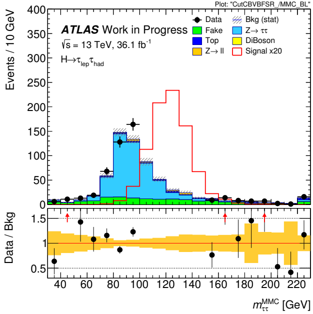
SM H$\to\tau\tau$ Analysis - Fit Model



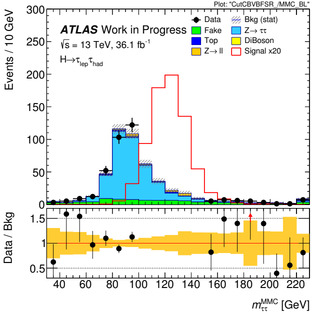
Impact of τ_{had} identification

- Apply **tight tau ID**
- Inclusive VBF signal region:

Medium



Tight



- Statistics reduced, but fakes also

Impact of τ_{had} identification

- Statistical power increased with tight ID working point
- Does it propagate to final significance ?
- Perform **fit in the semi-leptonic channel**

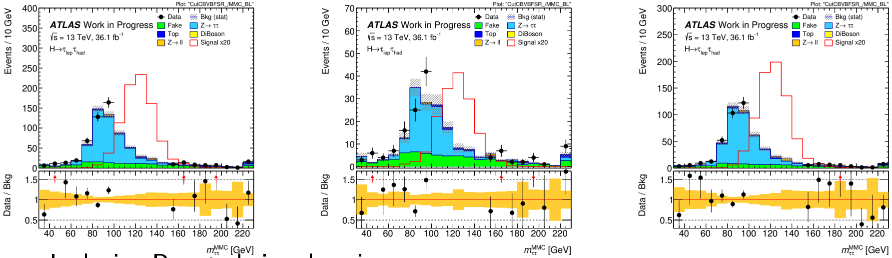
Impact of τ_{had} identification

- Statistical power increased with tight ID working point
- Does it propagate to final significance ?
- Perform **fit in the semi-leptonic channel**
- Loss in significance
- **Next step**: Also use information from **medium not tight** ID
  $\Rightarrow$ Continuous tau ID

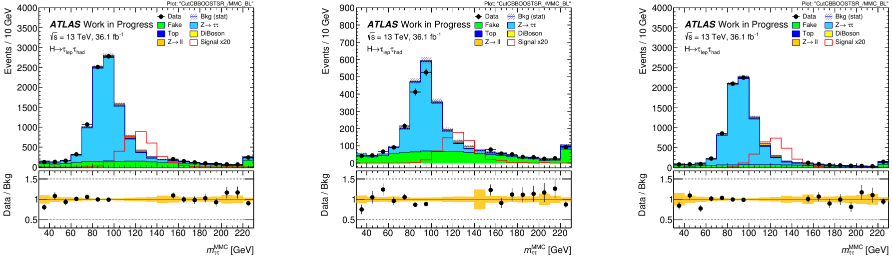
Fitting studies - Split Tau ID

- Inclusive VBF signal region:

Medium	Medium not Tight	Tight



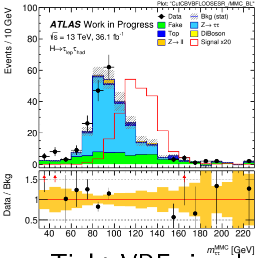
- Inclusive Boosted signal region:



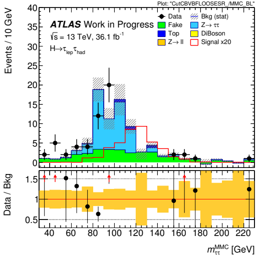
Fitting studies - Split Tau ID

- Loose VBF signal region:

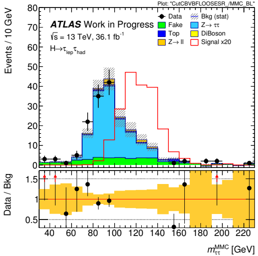
Medium



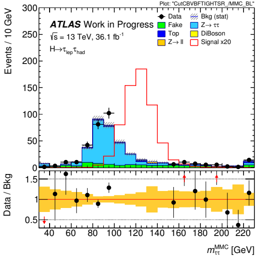
Medium not Tight

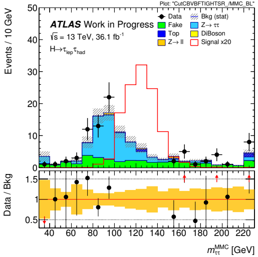


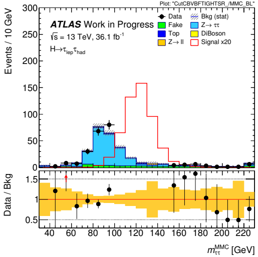
Tight



- Tight VBF signal region:



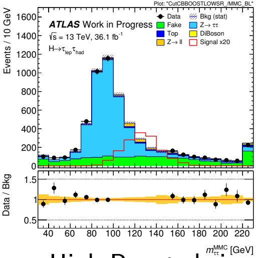




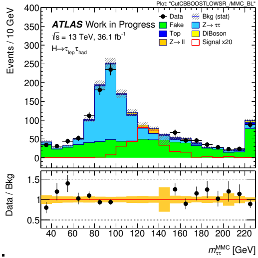
Fitting studies - Split Tau ID

- Low Boosted signal region:

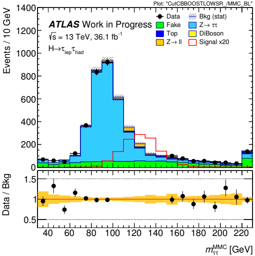
Medium



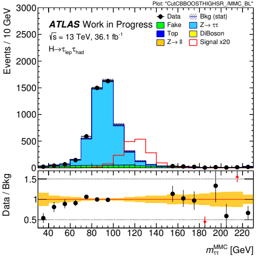
Medium not Tight

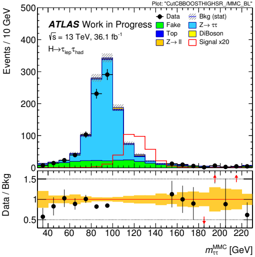


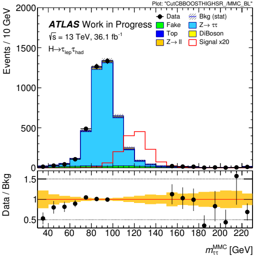
Tight



- High Boosted signal region:



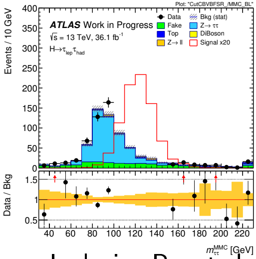


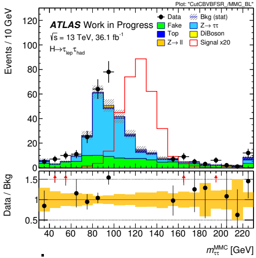


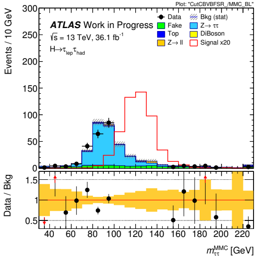
Fitting studies - Very Tight Tau ID

- Inclusive VBF signal region:

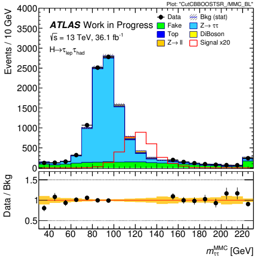
Medium | Medium Not Very Tight | Very Tight (bdt > 0.73)

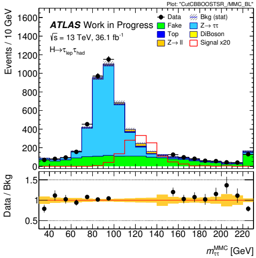


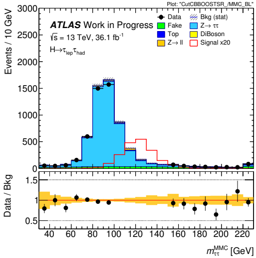




- Inclusive Boosted signal region:



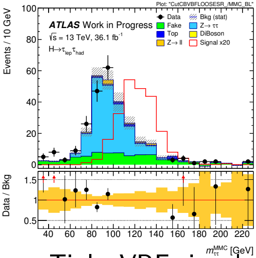




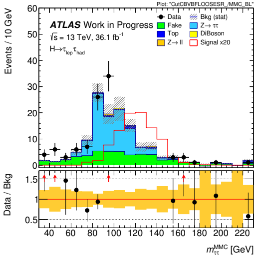
Fitting studies - Very Tight Tau ID

- Loose VBF signal region:

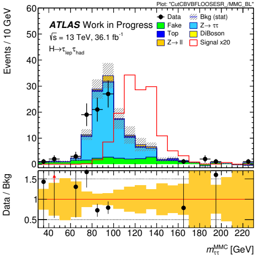
Medium



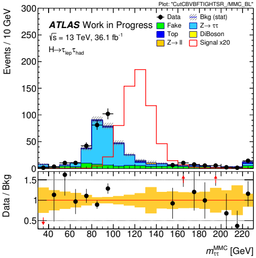
Medium Not Very Tight

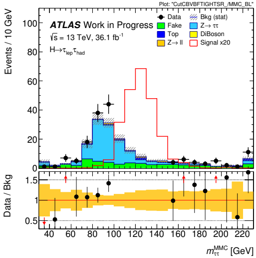


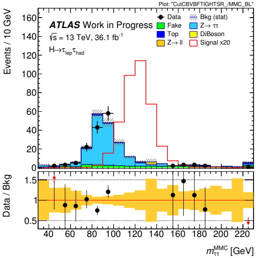
Very Tight (bdt > 0.73)



- Tight VBF signal region:



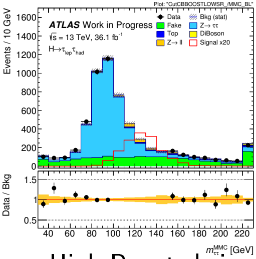


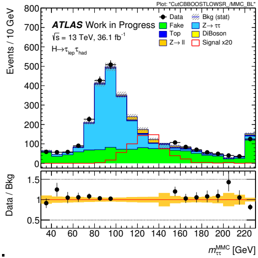


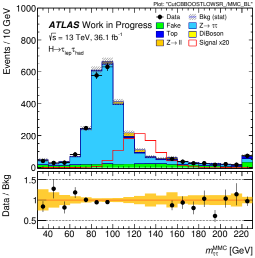
Fitting studies - Very Tight Tau ID

- Low Boosted signal region:

Medium | Medium Not Very Tight | Very Tight (bdt > 0.73)







- High Boosted signal region: