Tau identification using multivariate techniques in ATLAS

Dugan O’Neil

On Behalf of the ATLAS Collaboration
Simon Fraser University

ACAT 2011
Motivation

Introduction to Taus

Features of Tau Jets

MV Techniques Used

Performance

Recent Examples:

- $W \rightarrow \tau \nu$ (2010 data)
- $t \bar{t} \rightarrow \tau + \mu$ (2011 data)
Motivation

- τ leptons are present in important signatures of new physics at the LHC.
- SM Higgs favours decays to τ over other leptons.
- SUSY Higgs (charged and neutral) have substantial branching ratios to τ.
- Other signatures with preference for third generation decays, SUSY stau production, etc. predict an excess of τ.
- Important SM sources of τ (at LHC) include W, Z, t̄t.
General $\tau$ Characteristics

- $m_\tau = 1.8 \text{GeV}$
- Lifetime: $c\tau = 87 \mu m$
- Hadronic decays are a well-collimated collection of charged and neutral pions.
- Most have 1 or 3 charged tracks (prongs)
- Leading pion direction reproduces visible $\tau$ direction well.

In this presentation, only hadronic $\tau$ decays will be discussed. Leptonic modes are generally considered as part of prompt-electron or prompt-muon channels.
The jet cross section is many orders-of-magnitude higher than interesting $\tau$ signatures.

Hadronic $\tau$ decays have few basic features to distinguish them from gluon or quark initiated jets:
- lower particle multiplicity
- narrower cone
- different average composition (eg. EM energy fraction)

These features are described by (correlated) distinguishing variables built in each subdetector in ATLAS.
Reconstruction and Identification

- Finding $\tau$s at ATLAS is divided into two steps: Reconstruction and Identification (ID).
- Reconstruction involves building a $\tau$ candidate. Very little separation between signal and background:
  - Find a reconstructed jet in the calorimeter above 10GeV. Associate tracks to this jet.
  - Find a 6GeV track in the tracker. Associate calorimeter clusters to this track.
  - Calculate the distinguishing variables needed for ID.
- Identification uses the distinguishing variables to separate $\tau$s from jets. Three options in ATLAS:
  - Cuts.
  - Projective Likelihood.
  - Boosted Decision Tree.
Example Distinguishing Variables (2010 data)
Example Distinguishing Variables (2010 data)

ATLAS Preliminary

\[ m_{\text{clusters}} \text{ [GeV]} \]

\[ m_{\text{track}} \text{ [GeV]} \]

\[ S_{\text{flight}} \]

\[ W \rightarrow \tau \nu + Z \rightarrow \tau \tau \]

\[ \text{dijet Monte Carlo} \]

1 prong 15 GeV<p_T<60 GeV

\[ 2010 \text{ dijet data} \int dt L = 23 \text{ pb}^{-1} \]

\[ \text{ATLAS Preliminary} \]

\[ W \rightarrow \tau \nu + Z \rightarrow \tau \tau \]

\[ \text{dijet Monte Carlo} \]

3 prongs 15 GeV<p_T<60 GeV

\[ 2010 \text{ dijet data} \int dt L = 23 \text{ pb}^{-1} \]
Complications and Constraints

- Width-related variables change with $E$. The rate of change is not the same for signal and background.
- The distinguishing variables can depend on the instantaneous luminosity (pileup dependence).
- 1 and 3 prong taus have different backgrounds and very different background levels.
- Light-quark, heavy-quark and gluon initiated jets can have different fake rates.
- Sometimes the fake is not a jet at all. Electrons can look more like our definition of a $\tau$, than a real $\tau$ does.
- Different strategies are employed to deal with these complications.
We started with a relatively short-list of variables we expected to be well-modeled when data-taking started.

Defined simple cuts, LLH and BDT based on this conservative list. Verified variables in data.

2010 $W \rightarrow \tau \nu$, $Z \rightarrow \tau \tau$ observations made with simple cuts.

2011 retraining of all techniques (with more variables) led to significant improvements.

2011 physics analyses and $W \rightarrow \tau \nu$ cross section paper (based on 2010 data) then moved to MV techniques. Cuts used as reference, for performance studies and trigger.

I will only describe LLH and BDT:

- superior performance
- easier to optimize with more variables
- continuous output score
MV Discriminants (2010 data)

ATLAS Preliminary

1 prong 15 GeV < p_T < 60 GeV

2010 dijet data ∫ dt L = 23 pb⁻¹

ATLAS Preliminary

1 prong 15 GeV < p_T < 60 GeV

2010 dijet data ∫ dt L = 23 pb⁻¹
In 2011, the two MV discriminants were optimized in a similar way:
- same dijet data sample used to extract jet background.
- same triggers, selection cuts.
- MC signal (e.g. $Z \rightarrow \tau\tau$).
- multiple “bins” in luminosity.
- 2 “bins” in n-prong.
- same variable definitions (though slightly different lists).
Projective Likelihood Identification (2011)

- Separate variable list for single-prong and multi-prong $\tau$s.
- The likelihood function is a product of the distributions of ID variables:

$$L_{S(B)} = \prod_{i=1}^{N} p_{i}^{S(B)}(x_i)$$

(neglects correlations between variables)

- The discriminant is then

$$d = \ln \left( \frac{L_S}{L_B} \right) = \sum_{i=1}^{N} \ln \left( \frac{p_i^{S}(x_i)}{p_i^{B}(x_i)} \right)$$

- Use several separate likelihoods:
  - 3 $p_T$ bins
  - 3 luminosity bins ($N_{vtx}$)
  - 2 prong bins
Decision Trees in a Nutshell

- Turns a simple cut-based approach into a MV technique.
- Keep working on any events which fail cuts.
- Adding well-modeled variables should not hurt performance.
- Score assignment (D) depending leaf-node they land on.
Decision Trees in a Nutshell

- Turns a simple cut-based approach into a MV technique.
- Keep working on any events which fail cuts.
- Adding well-modeled variables should not hurt performance.
- Score assignment (D) depending leaf-node they land on.
Decision Trees in a Nutshell

- Turns a simple cut-based approach into a MV technique.
- Keep working on any events which fail cuts.
- Adding well-modeled variables should not hurt performance.
- Score assignment (D) depending leaf-node they land on.
Decision Trees in a Nutshell

- Turns a simple cut-based approach into a MV technique.
- Keep working on any events which fail cuts.
- Adding well-modeled variables should not hurt performance.
- Score assignment (D) depending leaf-node they land on.
Decision Trees in a Nutshell

- Turns a simple cut-based approach into a MV technique.
- Keep working on any events which fail cuts.
- Adding well-modeled variables should not hurt performance.
- Score assignment (D) depending leaf-node they land on.
Decision Trees in a Nutshell

- Turns a simple cut-based approach into a MV technique.
- Keep working on any events which fail cuts.
- Adding well-modeled variables should not hurt performance.
- Score assignment (D) depending leaf-node they land on.
Decision Trees in a Nutshell

- Turns a simple cut-based approach into a MV technique.
- Keep working on any events which fail cuts.
- Adding well-modeled variables should not hurt performance.
- Score assignment $(D)$ depending leaf-node they land on.
Boosting in a Nutshell

- We are using adaptive boosting (AdaBoost)
- Basic principal on DT, create a forest of trees:
  - train a tree $T_k$
  - $T_{k+1} = \text{modify}(T_k)$
- Check which events are misclassified in $T_k$, increase their weight in $T_{k+1}$ (boost).
- Obtain score from weighted average of all trees, where misclassification rate determines tree weight.
- Dilutes piecewise nature of DT, improves performance.
Results: Separating $\tau$ from Jets

ATLAS Preliminary
2011 data, $\int dt L = 160$ pb$^{-1}$

1-prong, $p_\tau > 20$ GeV

ATLAS Preliminary
2011 data, $\int dt L = 160$ pb$^{-1}$

3-prong, $p_\tau > 20$ GeV
Physics Applications: $W \rightarrow \tau \nu$

- Search for $W \rightarrow \tau \nu$ is very challenging. All you have is the $\tau$ and some MET.
- Analysis performed on 2010 dataset using 2010 version of BDT TauID. Cut on BDT output score.
- First paper submitted (PLB) from LHC with $W \rightarrow \tau \nu$ cross section (CERN-PH-EP-2011-122).

$$\sigma \times BR(W \rightarrow \tau \nu) = 11.1 \pm 0.3^{(stat)} \pm 1.7^{(syst)} \pm 0.4^{(lumi)}nb$$
Physics Applications: $t\bar{t} \rightarrow \mu + \tau$

- Goal to measure the $t\bar{t}$ cross section in the $\mu + \tau$ channel.
- After pre-selection cuts, dominant background is $t\bar{t} \rightarrow \mu + jet$, where the jet fakes a $\tau$.
- The only remaining feature is Tau ID.
- Jets from $t\bar{t}$ are quark-dominated. Standard fakes rates produced for $Z$, $W$ analyses are gluon-dominated.
- Rather than cutting on BDT output, use the shape.
- Quark-rich template from data, $\tau$ signal from MC. Fit the BDT output shape in data to get the signal normalization.
Physics Applications: \( t\bar{t} \rightarrow \mu + \tau \)

\[ \sigma_{t\bar{t}} = 142 \pm 21_{\text{stat.}} \pm 20_{16} \pm 5_{\text{lumi.}} \text{ pb} \]
Identifying hadronic $\tau$ decays at LHC is important for several interesting physics channels.

MV techniques bring several advantages:
- Superior performance.
- Continuous output distribution (shape info).
- Ease of adding new variables, re-optimizing.

ATLAS has now seen the SM $\tau$ sources $W, Z, \text{top}$ and has public exclusion results for both neutral and charged Higgs using tau signatures.
EXTRA SLIDES
**Train/Grow/Learn**

- First node holds all events
- For each variable, find the best cut
- Select best variable + cut and produce Failed and Passed branches
- Repeat recursively on each node
- Invoke a stopping condition. Terminal node = leaf.
Details - Splitting a node

**Impurity** $i(t)$

- maximum for equal mix of signal and background
- symmetric in $p_{signal}$ and $p_{background}$
- minimal for signal only or background only
- strictly concave $\Rightarrow$ reward purer nodes

- Decrease of impurity for split $s$ of node $t$ into children $t_L$ and $t_R$:
  \[ \Delta i(s, t) = i(t) - p_L \cdot i(t_L) - p_R \cdot i(t_R) \]
- Aim: find split $s^*$ such that:
  \[ \Delta i(s^*, t) = \max_{s \in \{\text{splits}\}} \Delta i(s, t) \]
- Maximizing $\Delta i(s, t) \equiv$ minimizing overall tree impurity

**Examples**

- **Gini**
  \[ Gini = 1 - \sum_{i=s,b} p_i^2 = \frac{2sb}{(s+b)^2} \]
- **Entropy**
  \[ \text{entropy} = - \sum_{i=s,b} p_i \log p_i \]
Decision Trees

Measure and Apply

- Take trained tree and run on independent simulated sample, determine purities.
- Apply to Data
- Should see enhanced separation (signal right, background left)
- Could cut on output and measure, or use whole distribution to measure.
Boosting - Use the Forest
Decision Trees - Boosting

Boosting

- Recent technique to improve performance of a weak classifier
- Recently used on DTs by GLAST and MiniBooNE
- Basic principal on DT:
  - train a tree $T_k$
  - $T_{k+1} = \text{modify}(T_k)$

AdaBoost algorithm

- Adaptive boosting
- Check which events are misclassified by $T_k$
- Derive tree weight $\alpha_k$
- Increase weight of misclassified events
- Train again to build $T_{k+1}$
- Boosted result of event $i$:
  \[ T(i) = \sum_{n=1}^{N_{\text{tree}}} \alpha_k T_k(i) \]

- Averaging dilutes piecewise nature of DT
- Usually improves performance

<table>
<thead>
<tr>
<th>Source of Uncertainty</th>
<th>( \frac{\delta C_W}{C_W} )</th>
<th>( \frac{\delta N_{EW}}{N_{EW}} )</th>
<th>( \frac{\delta N_{QCD}}{N_{QCD}} )</th>
<th>( \frac{\delta \sigma_{W \rightarrow \tau_h \nu_\tau}}{\sigma_{W \rightarrow \tau_h \nu_\tau}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trigger efficiency</td>
<td>6.1%</td>
<td>6.1%</td>
<td>-</td>
<td>7.0%</td>
</tr>
<tr>
<td>Energy scale</td>
<td>6.7%</td>
<td>8.7%</td>
<td>-</td>
<td>8.0%</td>
</tr>
<tr>
<td>(\tau_h) ID efficiency</td>
<td>9.6%</td>
<td>4.1%</td>
<td>-</td>
<td>10.3%</td>
</tr>
<tr>
<td>Jet (\tau_h) misidentification</td>
<td>-</td>
<td>7.2%</td>
<td>-</td>
<td>1.1%</td>
</tr>
<tr>
<td>Electron (\tau_h) misidentification</td>
<td>-</td>
<td>4.5%</td>
<td>-</td>
<td>0.7%</td>
</tr>
<tr>
<td>Pile-up reweighting</td>
<td>1.4%</td>
<td>1.2%</td>
<td>-</td>
<td>1.6%</td>
</tr>
<tr>
<td>Electron reconstruction/identification</td>
<td>-</td>
<td>1.2%</td>
<td>-</td>
<td>0.2%</td>
</tr>
<tr>
<td>Muon reconstruction</td>
<td>-</td>
<td>0.3%</td>
<td>-</td>
<td>0.04%</td>
</tr>
<tr>
<td>Underlying event modeling</td>
<td>1.3%</td>
<td>1.1%</td>
<td>-</td>
<td>1.5%</td>
</tr>
<tr>
<td>Cross section</td>
<td>-</td>
<td>4.5%</td>
<td>-</td>
<td>0.7%</td>
</tr>
<tr>
<td>QCD estimation: Stability/correlation</td>
<td>-</td>
<td>-</td>
<td>2.7%</td>
<td>0.2%</td>
</tr>
<tr>
<td>QCD estimation: Sig./EW contamination</td>
<td>-</td>
<td>-</td>
<td>2.1%</td>
<td>0.1%</td>
</tr>
<tr>
<td>Monte Carlo statistics</td>
<td>1.4%</td>
<td>2.4%</td>
<td>6.0%</td>
<td>1.5%</td>
</tr>
<tr>
<td>Total systematic uncertainty</td>
<td>13.4%</td>
<td>15.2%</td>
<td>6.9%</td>
<td>15.1%</td>
</tr>
</tbody>
</table>
Physics Applications: $W \rightarrow \tau \nu$

![Graph showing ATLAS W → τν, W → eν, W → μν](image)
Physics Applications: $t\bar{t} \rightarrow \mu + \tau$ - Cutflow (1p)

<table>
<thead>
<tr>
<th>Cut</th>
<th>$\bar{t}(\mu, \tau)$</th>
<th>$\bar{t}(\mu + \text{jets})$</th>
<th>$\bar{t}(\mu\ell)$</th>
<th>W+jets</th>
<th>Z+jets</th>
<th>Single top</th>
<th>Diboson</th>
<th>Total</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trigger</td>
<td>2693 ± 10</td>
<td>20880 ± 40</td>
<td>4679 ± 10</td>
<td>5680000 ± 6800</td>
<td>867000 ± 630</td>
<td>7692 ± 40</td>
<td>8212 ± 40</td>
<td>6591500 ± 6900</td>
<td>8872361</td>
</tr>
<tr>
<td>Isolated $\mu$</td>
<td>2243 ± 10</td>
<td>13700 ± 20</td>
<td>2064 ± 10</td>
<td>5419300 ± 7000</td>
<td>416800 ± 460</td>
<td>6316 ± 40</td>
<td>6453 ± 40</td>
<td>5866900 ± 7000</td>
<td>8113657</td>
</tr>
<tr>
<td>$\geq 1 \tau$ candidate</td>
<td>497 ± 5</td>
<td>2042 ± 10</td>
<td>131 ± 2</td>
<td>74340 ± 660</td>
<td>13740 ± 80</td>
<td>470 ± 10</td>
<td>680 ± 10</td>
<td>91900 ± 670</td>
<td>154513</td>
</tr>
<tr>
<td>$N_{\text{jet}} \geq 2$</td>
<td>401 ± 4</td>
<td>1941 ± 10</td>
<td>103 ± 2</td>
<td>7330 ± 90</td>
<td>1516 ± 30</td>
<td>227 ± 10</td>
<td>169 ± 10</td>
<td>11690 ± 100</td>
<td>16385</td>
</tr>
<tr>
<td>$E_T^{\text{miss}} &gt; 30$ GeV</td>
<td>351 ± 4</td>
<td>1562 ± 10</td>
<td>93 ± 2</td>
<td>5237 ± 80</td>
<td>654 ± 20</td>
<td>178 ± 10</td>
<td>117 ± 10</td>
<td>8191 ± 80</td>
<td>8930</td>
</tr>
<tr>
<td>$H_T &gt; 200$ GeV</td>
<td>346 ± 4</td>
<td>1553 ± 10</td>
<td>92 ± 2</td>
<td>4332 ± 60</td>
<td>514 ± 20</td>
<td>170 ± 5</td>
<td>102 ± 5</td>
<td>7109 ± 60</td>
<td>7304</td>
</tr>
<tr>
<td>$\geq 1 b$-jet</td>
<td>235 ± 3</td>
<td>979 ± 10</td>
<td>56 ± 1</td>
<td>163 ± 10</td>
<td>17 ± 3</td>
<td>87 ± 3</td>
<td>6 ± 1</td>
<td>1543 ± 10</td>
<td>1593</td>
</tr>
</tbody>
</table>
Physics Applications: $t\bar{t} \rightarrow \mu + \tau$ - Cutflow (1p)

<table>
<thead>
<tr>
<th>Cut</th>
<th>$t\bar{t}(\mu, \tau)$</th>
<th>$t\bar{t}(\mu + \text{jets})$</th>
<th>$t\bar{t}(\mu\ell)$</th>
<th>$W + \text{jets}$</th>
<th>$Z + \text{jets}$</th>
<th>Single top</th>
<th>Diboson</th>
<th>Total</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trigger</td>
<td>2693 ± 10</td>
<td>20880 ± 40</td>
<td>4679 ± 10</td>
<td>5680000 ± 6800</td>
<td>867000 ± 630</td>
<td>7692 ± 40</td>
<td>8212 ± 40</td>
<td>6591500 ± 6900</td>
<td>8872361</td>
</tr>
<tr>
<td>Isolated $\mu$</td>
<td>2243 ± 10</td>
<td>13700 ± 20</td>
<td>2064 ± 10</td>
<td>5419000 ± 7000</td>
<td>416831 ± 460</td>
<td>6316 ± 40</td>
<td>6453 ± 40</td>
<td>5866900 ± 7000</td>
<td>8113657</td>
</tr>
<tr>
<td>≥ 1 $\tau$ candidate</td>
<td>500 ± 5</td>
<td>5100 ± 10</td>
<td>437 ± 4</td>
<td>220300 ± 1100</td>
<td>22350 ± 110</td>
<td>1310 ± 20</td>
<td>1412 ± 20</td>
<td>251400 ± 1200</td>
<td>503283</td>
</tr>
<tr>
<td>N$_{\text{jet}} \geq$ 2</td>
<td>371 ± 4</td>
<td>4794 ± 10</td>
<td>331 ± 4</td>
<td>19900 ± 160</td>
<td>2896 ± 40</td>
<td>583 ± 10</td>
<td>336 ± 10</td>
<td>29210 ± 160</td>
<td>44759</td>
</tr>
<tr>
<td>$E_{\text{miss}} &gt;$ 30 GeV</td>
<td>326 ± 4</td>
<td>3850 ± 10</td>
<td>297 ± 3</td>
<td>14400 ± 130</td>
<td>1134 ± 20</td>
<td>461 ± 10</td>
<td>230 ± 10</td>
<td>20700 ± 130</td>
<td>23932</td>
</tr>
<tr>
<td>$H_{\text{T}} &gt;$ 200 GeV</td>
<td>321 ± 4</td>
<td>3823 ± 10</td>
<td>293 ± 3</td>
<td>11860 ± 100</td>
<td>922 ± 20</td>
<td>441 ± 10</td>
<td>200 ± 10</td>
<td>17900 ± 110</td>
<td>19522</td>
</tr>
<tr>
<td>≥ 1 $b$-jet</td>
<td>206 ± 3</td>
<td>2465 ± 10</td>
<td>182 ± 3</td>
<td>535 ± 30</td>
<td>33 ± 4</td>
<td>226 ± 10</td>
<td>13 ± 2</td>
<td>3660 ± 30</td>
<td>4086</td>
</tr>
</tbody>
</table>