Boosted hadronic object identification using jet substructure in ATLAS Run-2

Emma Winkels
on behalf of the ATLAS collaboration

HEPMAD18
Outline

• Jets and jet substructure
• Top and W tagging
• $H \rightarrow bb$ tagging
• Mass-decorrelated taggers
• Summary
Jets
What is a jet?

Jets are objects constructed from the energy deposits left by collimated sprays of particles.

Attempt to group inputs from common sources together.
What is a jet?

Jets are objects constructed from the energy deposits left by collimated sprays of particles.

Attempt to group inputs from common sources together.
What is a jet?

Jets are objects constructed from the energy deposits left by collimated sprays of particles.

Attempt to group inputs from common sources together.
What is a jet?

Jets are objects constructed from the energy deposits left by collimated sprays of particles.

Attempt to group inputs from common sources together.
Run: 311071
Event: 1452867343
2016-10-21 06:34:07 CEST
Inputs to jets

ATLAS uses calorimeter objects and tracks as jet inputs. Calorimeter measures energy of particles. Starts with topological clustering of calorimeter cells.
How to define a jet

- There is no unique way to define a jet
- Different jet algorithms to cluster energy constituents into a jet:
  - Anti-$k_T$: cluster hard (high-$p_T$) and close (small $\Delta R$) energy deposits first
  - $k_T$: cluster soft (low-$p_T$) and close
  - Cambridge/Aachen: cluster close

\[ \Delta R \]

12 constituents
- = large E
- = small E
Boosted jets

Different jet radii ($R$) for different purposes:

- Small jets for quarks and gluons
- Large jets for hadronic decays of $W$, $Z$, $H$, top..
Jet substructure

• Access the inner structure of large jets*

• Jet substructure variables are some function of
  • Number of constituents
  • Energy of the constituents
  • Angular separation of the constituents ($\Delta R$)

• Jet substructure helps us in jet tagging

* We use jet grooming to cut away the soft parts of jets, see 1510.05821
Jet tagging

• Identify the particle that produced the jet.
• Used in broad range of physics analyses.
  • Analyses looking at boosted top/Higgs/W: distinguish these large jets from quark/gluon jets
  • Analyses with $b$-hadrons or $c$-hadrons in the final state: heavy flavour tagging ($b$-, $c$-quark jets)
Top/W tagging

ATLAS-CONF-2017-064
Commissioning of a tagger

ATLAS process from idea to tagger used for physics analyses:

- Use MC to choose tagger
- Check data/MC
- Measure efficiencies in data & MC
- Calculate uncertainties on efficiencies
Two-variable tagging

• Simple cut-based tagging works well:
  • $W$: $m^{\text{comb}} + D_2$
  • $\text{Top}$: $m^{\text{comb}} + \tau_{32}$

• $m^{\text{comb}}$: Combined mass, combines calorimeter clusters with tracks to give more stable mass performance at high-$p_T$.

• $D_2$: Energy correlation ratio, distinguishes between one-prong and two-prong jets.

• $\tau_{32}$: $N$-subjettiness, distinguishes between two-prong and three-prong jets.
Two-variable tagging

• Simple cut-based tagging works well:
  • $W$: $m_{\text{comb}} + D_2$
  • $\text{Top}$: $m_{\text{comb}} + \tau_{32}$

• $m_{\text{comb}}$: Combined mass, combines calorimeter clusters with tracks to give more stable mass performance at high-$p_T$.

• $D_2$: Energy correlation ratio, distinguishes between one-prong and two-prong jets.

• $\tau_{32}$: N-subjettiness, distinguishes between two-prong and three-prong jets.
Machine learning taggers - BDT

- Machine learning techniques allow for the use of multiple variables
- Boosted decision tree (BDT): sequentially adding variables improves classification over two-variable tagging

### W tagging

- ATLAS Simulation Preliminary
- $\sqrt{s} = 13$ TeV, BDT W Tagging
- Trimmed anti-$k_t$ $R = 1.0$ jets
- $\varepsilon_{\text{sig}} = 50\%$
- $p_T^{\text{true}} = [200,2000]$ GeV
- $m_{\text{comb}} > 40$ GeV, $|\eta^{\text{true}}| < 2.0$

### Top tagging

- ATLAS Simulation Preliminary
- $\sqrt{s} = 13$ TeV, BDT Top Tagging
- Trimmed anti-$k_t$ $R = 1.0$ jets
- $\varepsilon_{\text{sig}} = 80\%$
- $m_{\text{comb}} > 40$ GeV, $|\eta^{\text{true}}| < 2.0$
Machine learning taggers - DNN

- ATLAS evaluated TopoDNN* top tagger.
- Uses deep neural network (DNN) with topocluster jet constituents as inputs.
- Performance is better with low-level inputs than standard machine learning taggers.

* More details on TopoDNN: 1704.02124.
Measurements in data

**W enriched sample**

- **ATLAS Preliminary**
  - $\sqrt{s} = 13$ TeV, 36.1 fb$^{-1}$
  - Trimmed anti-$k_t$, $R=1.0$ jets
  - $\Delta R$(large-$R$ jet, b-jet) $> 1.0$
  - $p_T > 200$ GeV

- **Data 2015+2016**

![Graph showing data and prediction for W enriched sample.](image)

**Top enriched sample**

- **ATLAS Preliminary**
  - $\sqrt{s} = 13$ TeV, 36.1 fb$^{-1}$
  - Trimmed anti-$k_t$, $R=1.0$ jets
  - $\Delta R$(large-$R$ jet, b-jet) $< 1.0$
  - $p_T > 350$ GeV

![Graph showing data and prediction for top enriched sample.](image)

---

**Measurements in data**

**W enriched sample**

- **ATLAS Preliminary**
  - $\sqrt{s} = 13$ TeV, 36.1 fb$^{-1}$
  - Trimmed anti-$k_t$, $R=1.0$ jets
  - $\Delta R$(large-$R$ jet, b-jet) $> 1.0$
  - $p_T > 200$ GeV

- **Data 2015+2016**

![Graph showing data and prediction for W enriched sample.](image)

**Top enriched sample**

- **ATLAS Preliminary**
  - $\sqrt{s} = 13$ TeV, 36.1 fb$^{-1}$
  - Trimmed anti-$k_t$, $R=1.0$ jets
  - $\Delta R$(large-$R$ jet, b-jet) $< 1.0$
  - $p_T > 350$ GeV

![Graph showing data and prediction for top enriched sample.](image)

---

**ATLAS-CONF-2017-064**

11/09/2018

Emma Winkels | ewinkels@cern.ch | HEPMAD18
W/top tagging efficiency

Need to measure efficiency in data and get uncertainty on this efficiency*.

• Full ATLAS 2015-2016 dataset of $36.1 - 36.7 \text{ fb}^{-1}$
• Measure top/W tagging efficiency in $t\bar{t}$ lepton+jets samples.
• Measure multijet rejection in dijet and $\gamma$ +jets samples.

* Also done in V+jets: ATLAS-CONF-2018-016
Pile-up is the resulting signal in the detector from other interactions besides the hard scatter we want to look at. Expressed as the mean number of interactions per bunch crossing.
$\epsilon_{MC} = \frac{N_{tagged}^{signal}}{N_{tagged}^{signal} + N_{not\,tagged}^{signal}}$

$\epsilon_{data} = \frac{N_{tagged}^{fitted\,signal}}{N_{tagged}^{fitted\,signal} + N_{not\,tagged}^{fitted\,signal}}$

**W tagger**

**TopoDNN tagger**

---

**ATLAS** Preliminary

- $\sqrt{s} = 13$ TeV, 36.1 fb$^{-1}$
  - lepton+jets selection
  - Trimmed anti-$k_t$, $R=1.0$ jets
  - W tagger ($\epsilon_{sig} = 50\%$): $m_{miss} + D_2$
  - $p_T > 200$ GeV
  - $\text{Total uncert.}$

Data 2015+2016

---

**ATLAS** Preliminary

- $\sqrt{s} = 13$ TeV, 36.1 fb$^{-1}$
  - lepton+jets selection
  - Trimmed anti-$k_t$, $R=1.0$ jets
  - Top tagger ($\epsilon_{sig} = 80\%$): TopoDNN
  - $p_T > 450$ GeV
  - $\text{Total uncert.}$

Data 2015+2016

---

Mean number of interactions per bunch crossing

Data/MC

---

ATLAS-CONF-2017-064

Emma Winkels | ewinkels@cern.ch | HEPMAD18
Multijet rejection vs. pile-up

**W tagger**

- Background rejection ($1/\varepsilon_{W}$):
  - $m^{comb} + D_{2}$
- Data 2015+2016
- $f_{s} = 13$ TeV, 36.7 fb$^{-1}$
- Trimmed anti-$k_{t}$, $R=1.0$ jets
- Multijet Selection
- $W$ tagger ($\varepsilon_{\text{sig}} = 50\%$):

**TopoDNN tagger**

- Background rejection ($1/\varepsilon_{\text{TopoDNN}}$):
- $f_{s} = 13$ TeV, 36.7 fb$^{-1}$
- Trimmed anti-$k_{t}$, $R=1.0$ jets
- Multijet Selection
- Top tagger ($\varepsilon_{\text{sig}} = 80\%$): TopoDNN

**Measure efficiencies in data & MC**
**Calculate uncertainties on efficiencies**

**Use MC to choose tagger**
**Check data/MC**

**ATLAS**

Preliminary

- Mean number of interactions per bunch crossing

---

**ATLAS-CONF-2017-064**

11/09/2018
Emma Winkels | ewinkels@cern.ch | HEPMAD18
H→bb tagging

ATL-PHYS-PUB-2017-010
Overview

Current nominal tagger identifies $b$-jets with multivariate algorithm on anti-$k_T$ $R=0.2$ track-jets. Loses efficiency at high-$p_T$ due to $b$-jets merging.

3 new subjet reconstruction techniques to mitigate this loss.
1 - Variable radius track jets

Subjets with dynamic radius parameter:

$$R_{eff}(p_T) = \frac{\rho}{p_T}$$

with low ($R_{min}$) and high ($R_{max}$) cutoff.

Scans performed over $\rho$, $R_{min}$, $R_{max}$ to find optimal values.
2 – Exclusive-$k_T$

- Recluster trimmed large-R jet calorimeter constituents with $k_T$ algorithm and stop when 2 jets are obtained.
- Splits the large-R jet into two parts, each of which is expected to contain one $b$-hadron.
3 – Centre-of-mass

- Boost jet calorimeter clusters to the centre-of-mass frame of the large-R jet (jet rest frame) and reconstruct subjets.
- Tracks for $b$-tagging are also boosted to the centre-of-mass frame.
Results

New methods show large improvement over nominal tagger for $p_T > 1000$ GeV.
Mass decorrelation
Mass decorrelated taggers

- Jet substructure variables are correlated with jet mass. When you put many of them into a multivariate analysis the correlation gets very strong.
  - **Sculpting of the multijet background** -> resembles the resonance jet mass distribution
  - Depopulates side-band regions
- Aim to decorrelate jet substructure classifiers from jet mass
Designed decorrelated taggers (DDT)

- $\tau_{21}$ variable distinguishes 1-prong from 2-prong jets
- Has a linear relationship to jet scaling variable $\rho^{DDT}$ for masses $> 80$ GeV
Designed decorrelated taggers (DDT)

- $\tau_{21}$ variable distinguishes 1-prong from 2-prong jets
- Has a linear relationship to jet scaling variable $\rho^{\text{DDT}}$ for masses $> 80$ GeV

![Graph showing background rejection and signal efficiency](ATLAS Simulation Preliminary)

- $\sqrt{s} = 13$ TeV
- $W$ jet tagging
- $p_T \in [500, 1000]$ GeV

$\tau_{21}$ after DDT transformation

Standard $\tau_{21}$
Summary
Summary

• Top/W tagging:
  • Machine learning taggers perform better than 2-variable taggers
  • Machine learning tagger with low-level inputs (TopoDNN) performs the best for top tagging
  • Signal efficiency & background rejection in data are well modelled by MC

• H → bb tagging:
  • 3 new subjet reconstruction techniques to overcome loss of efficiency at high $p_T$ due to $b$-jet merging
  • Large improvement over nominal tagger for $p_T > 1000$ GeV

• Mass decorrelated taggers:
  • Various approaches studied with promising results
Fin.
Back-up
The ATLAS coordinate system

\[ \eta = -\ln \tan \frac{\theta}{2} \]

\[ \Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2} \]
Granularity of ATLAS calorimeter

- The hadronic calorimeter is coarser than the EM calorimeter. If we want to use all calorimeter layers we are limited by coarsest layer ($R_{min} \sim 0.2$). But we can also ignore coarse layers and use only very fine layers.
  - EM granularity: $\Delta \eta \times \Delta \phi \approx 0.025 \times \pi / 128$
  - Hadronic granularity: $\Delta \eta \times \Delta \phi \approx 0.1 \times \pi / 32$
Jet trimming

Initial jet  \rightarrow \text{Subjets } k_T \ R = 0.2 \rightarrow \text{Trimmed jet}

\[ \frac{P_T^i}{P_T^{jet}} < f_{cut} = 5\% \]
Combined mass

Track-assisted mass: \( m^{TA} = m^{\text{track}} \times \frac{p_T^{\text{calo}}}{p_T^{\text{track}}} \)

in which \( m^{\text{track}} \) and \( p_T^{\text{track}} \) are the invariant mass and \( p_T \) calculated from tracks associated with the large-R trimmed calorimeter jet and \( p_T^{\text{calo}} \) is the \( p_T \) of the original trimmed large-R jet.

Calorimeter mass: \( m^{\text{calo}} = \sqrt{\left( \sum_i E_i \right)^2 - \left( \sum_i \vec{p}_i \right)^2} \)

\( m^{\text{comb}} = a \times m^{\text{calo}} + b \times m^{TA} \)

with \( a = \frac{\sigma_{\text{calo}}^2}{\sigma_{\text{calo}}^2 + \sigma_{\text{TA}}^2} \) and \( b = \frac{\sigma_{\text{TA}}^2}{\sigma_{\text{calo}}^2 + \sigma_{\text{TA}}^2} \)

where \( \sigma_{\text{calo}} \) and \( \sigma_{\text{TA}} \) are the calorimeter-based jet mass resolution function and the track-assisted mass resolution function respectively.
JSS variables

• Energy correlation ratio: $D_2 = \frac{e_3}{(e_2)^3}$ where $e_2$ and $e_3$ are the 2- and 3-prong energy correlation functions which are sensitive to the 2- and 3-prong structure in a jet.

• N-subjettiness: $\tau_{32} = \tau_3 / \tau_2$ where

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \times \min(\Delta R_{1k}, \Delta R_{2k}, \ldots, \Delta R_{Nk}), \quad d_0 = \sum_k p_{Tk} \times R$$

$R$ is radius parameter of the jet, $p_{Tk}$ is transverse momentum of constituent $k$, and $\delta R_{ik}$ is the distance between the subjet $i$ and the constituent $k$. The N-subjettiness variable $\tau_N$ expresses how well a jet can be described as containing N subjets.

• Jet mass calculated from constituents: $m^2 = (\sum_i E_i)^2 - (\sum_i \vec{p}_i)^2$
**W tagging with ML methods**

**Low pT**

**High pT**

---

**W tagging with ML methods**

**Low pT**

- DNN W
- BDT W
- 2-var optimised tagger
- Trimmed anti-$k_t$, $R = 1.0$ jets
- |$\eta_{true}$| $< 2.0$
- $p_T^{true}$ = [200, 500] GeV
- W tagging

**High pT**

- DNN W
- BDT W
- 2-var optimised tagger
- Trimmed anti-$k_t$, $R = 1.0$ jets
- |$\eta_{true}$| $< 2.0$
- $p_T^{true}$ = [1000, 1500] GeV
- W tagging
TopoDNN

Uses $p_T$, $\eta$, $\varphi$ of 10 leading topoclusters in trimmed large-$R$ jet.
Top/W tagging in data

- Full ATLAS 2015 – 2016 dataset, $L = 36.1 - 36.7$ fb$^{-1}$

- W/top tagging efficiency
  - $t\bar{t}$ decay to single lepton + jets topology
  - $W$: $\Delta R(b$-jet, large-$R$ jet) $> 1.0$
    $p_T(J) > 200$ GeV
  - Top: $\Delta R(b$-jet, large-$R$/HTT jet) $< 1.0/1.5$
    $p_T(J) > 350$ GeV

- Multijet rejection
  - Dijets: $p_T(J) > 450$ GeV
  - $\gamma +$ jets: $p_T(J) > 200$ GeV
W/top tagging efficiency

Need to measure efficiency in data and get uncertainty on this efficiency.

Measure signal-like events in data to fit large-R jet mass for events passing/failing a tagger.
**W/top tagging efficiency vs. jet pT**

### Pre-fit

\[ \epsilon_{\text{MC}} = \frac{N_{\text{tagged signal}}}{N_{\text{tagged signal}} + N_{\text{not tagged signal}}} \]

### Post-fit

\[ \epsilon_{\text{data}} = \frac{N_{\text{tagged signal}}}{N_{\text{tagged fitted signal}} + N_{\text{not tagged fitted signal}}} \]

### ATLAS Preliminary

- \( \sqrt{s} = 13 \) TeV, 36.1 fb\(^{-1}\)
- Lepton+jets selection
- Trimmed anti-\(k_T\) \(R=1.0\) jets
- \( W \) tagger (\( \epsilon_{\text{tag}} = 50\% \)): \( m_{\text{comb}} + D_2 \)
- Total uncert.

### Data 2015+2016

- Data 2015+2016

---

**W tagger**

![Graph showing W tagger efficiency vs. jet pT](image)

**TopoDNN tagger**

![Graph showing TopoDNN tagger efficiency vs. jet pT](image)

---

**11/09/2018**  Emma Winkels | ewinkels@cern.ch | HEPMAD18

47
Multijet rejection vs. jet pT

**W tagger**

- Trimmed anti-$k_T$, $R=1.0$ jets
- Multijet Selection
- $W$ tagger ($\varepsilon_{\text{tag}} = 50\%$): $n^{\text{true}} + D_2$

**TopoDNN tagger**

- Trimmed anti-$k_T$, $R=1.0$ jets
- Multijet Selection
- Top tagger ($\varepsilon_{\text{tag}} = 80\%$): TopoDNN
Variable radius track jets

**ATLAS Simulation Preliminary**

$\rho=30$ GeV, $R_{\text{max}} = 0.4$

- $R=0.2$ Track Jet
- $R_{\text{min}} = 0$
- $R_{\text{min}} = 0.02$
- $R_{\text{min}} = 0.06$
- $R_{\text{min}} = 0.1$

$76$ GeV < $m_{\text{jet}}$ < $146$ GeV

**ATLAS Simulation Preliminary**

$\rho=30$ GeV, $R_{\text{min}} = 0.02$

- $R=0.2$ Track Jet
- $R_{\text{min}} = 0$
- $R_{\text{max}} = 0.2$
- $R_{\text{max}} = 0.3$
- $R_{\text{max}} = 0.4$

$76$ GeV < $m_{\text{jet}}$ < $146$ GeV

Double Subjet B-Labeling Efficiency vs. Higgs Jet $p_T$ [GeV]
Results $H\rightarrow bb$ tagging techniques

**ATLAS Simulation**

Preliminary

$G \rightarrow hh \rightarrow bbbb$

$76 \text{ GeV} < m_{\text{jet}} < 146 \text{ GeV}$

<table>
<thead>
<tr>
<th>Fraction of Subjet Multiplicity</th>
<th>Higgs Jet $p_T$ [GeV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{\text{Subjet}} = 1$</td>
<td><img src="image1.png" alt="Graph" /></td>
</tr>
<tr>
<td>$N_{\text{Subjet}} = 2$</td>
<td></td>
</tr>
<tr>
<td>$N_{\text{Subjet}} \geq 3$</td>
<td></td>
</tr>
</tbody>
</table>

**双质子标签效率**

$76 \text{ GeV} < m_{\text{jet}} < 146 \text{ GeV}$

- $R=0.2$ Track Jet (3)
- VR Track Jet (3)
- ExKt Subjet
- CoM Subjet

ATLAS Simulation Preliminary

$Higgs \text{ Jet } p_T$ [GeV]
H-\(\rightarrow\)bb tagging results

**ATLAS**

Simulation Preliminary

76 GeV < \(m_{\text{jet}}\) < 146 GeV

1500 GeV < \(p_{T,\text{jet}}\) < 2000 GeV

double b-tagging

- R=0.2 Track Jet
- VR Track Jet
- ExKt Subjet
- CoM Subjet

Top Jet Rejection

- R=0.2 Track Jet
- VR Track Jet
- ExKt Subjet
- CoM Subjet
Designed decorrelated taggers

• Linear fit on the relationship between $\tau_{21}$ and rho. Transformation is:

$$\tau_{21}^{DDT} = \tau_{21} - a \times (\rho^{DDT} - 1.5)$$

$a$ is the slope of the fit in the plot.

• DDT transformation removes the linear correlation of $\tau_{21}$ with $\rho$. Since $\rho$ has info on kinematics of the jet (m and $p_T$), the DDT transform yields a JSS discriminant which is decorrelated from the jet mass.
Mass decorrelation

**ATLAS** Simulation Preliminary

$\sqrt{s} = 13$ TeV, W jet tagging
Cuts at $e^{\text{rel}}_{\text{sig}} = 50\%$

Inclusive selection:
- Multijets
- W jets

![Plots showing jet mass distributions](image)