Data analysis in high-energy physics

Search for top—anti-top quark resonances in proton-proton collisions with ATLAS data

CSU NUPAC Tutorials 2019

Gabriel Palacino

Thursday, April 04, 2019
We are going to implement some event selection criteria based on the final state of our signal.

We will learn how to apply quality requirements to the physics analysis objects.

We will reconstruct new kinematic variables.
What do we expect to see in a $Z'\rightarrow tt$ event?

- We are focusing on the one-lepton+jets channel:
  - 1 lepton only: electron or muon
  - The event must be selected by an electron or muon trigger
What do we expect to see in a $Z'\rightarrow tt$ event?

- We are focusing on the one-lepton+jets channel:
  - 1 lepton only: electron or muon
  - The event must be selected by a electron or muon trigger

- $W$ boson decays leptonically to a lepton+neutrino
  - The neutrino leaves ATLAS undetected
  - This is observed as Missing $E_T$
What do we expect to see in a $Z'\rightarrow tt$ event?

- We are focusing on the one-lepton+jets channel:
  - 1 lepton only: electron or muon
  - The event must be selected by a electron or muon trigger

- W boson decays leptonically to a lepton+neutrino
  - The neutrino leaves ATLAS undetected
  - This is observed as Missing $E_T$

- Two b-quarks and two quarks from the top quarks and W boson, respectively
  - Four jets are expected from decay of $Z'\rightarrow tt$
What do we expect to see in a $Z'\rightarrow tt$ event?

- We are focusing on the one-lepton+jets channel:
  - 1 lepton only: electron or muon
  - The event must be selected by an electron or muon trigger

- $W$ boson decays leptonically to a lepton+neutrino
  - The neutrino leaves ATLAS undetected
  - This is observed as Missing $E_T$

- Two $b$-quarks and two quarks from the top quarks and $W$ boson, respectively
  - Four jets are expected from decay of $Z'\rightarrow tt$

- Jets from $b$-quarks can be tagged as such
  - $b$-jet tagging efficiency is less than 100%
  - To prevent significant loss of signal events, we will require only one of the four jets to be $b$-tagged.
What do we expect to see in a $Z'$—$\rightarrow tt$ event?

- We are focusing on the one-lepton+jets channel:
  - 1 lepton only: electron or muon
  - The event must be selected by an electron or muon trigger

- $W$ boson decays leptonically to a lepton+neutrino
  - The neutrino leaves ATLAS undetected
  - This is observed as Missing $E_T$

- Two $b$-quarks and two quarks from the top quarks and $W$ boson, respectively
  - At least four jets are expected from decay of $Z'$—$\rightarrow tt$

- Jets from $b$-quarks can be tagged as such
  - $b$-jet tagging efficiency is less than 100%
  - To prevent significant loss of signal events, we will require only one of the jets to be $b$-tagged.

- Let's not forget that we also need a primary vertex in the event.
Selecting leptons

How do we make sure that the leptons we are selecting are good for analysis?

• Leptons can be produced in jets.
• Leptons in jets are surrounded by other particles that leave tracks in the ID and deposit energy in the calorimeter.
• We are interested in so-called isolated leptons
• There are two types of isolation: track and calorimeter isolation
  ‧ Track isolation: it is measured in terms of the total $p_T$ of the tracks within a cone of radius $R=\sqrt{\Delta \eta^2 + \Delta \phi^2}$ centered about the lepton (for us $R=0.3$)
  ‧ Calorimeter isolation: it is measured in terms of the total $E_T$ in the calorimeter within a cone of radius $R=\sqrt{\Delta \eta^2 + \Delta \phi^2}$ centered about the lepton (for us $R=0.2$)
Selecting leptons

How do we make sure that the leptons we are selecting are good for analysis?

- Lepton isolation can be relative to the lepton $p_T$ (we will use this type of isolation)
- There are several levels identification criteria that depend on likely we require the object to be an actual lepton
- These combine information from several subdetectors to classify the lepton as Loose, Medium or Tight
- We will require **Tight lepton** identification requirements.

![Identification Efficiency Graph]

**Figure 1**: The efficiency to identify electrons from $Z \rightarrow ee$ decays (left) and the efficiency to identify hadrons as electrons (background rejection, right) estimated using simulated dijet samples. The efficiencies are obtained using Monte Carlo simulations, and are measured with respect to reconstructed electrons. The candidates are matched to true electron candidates for $Z \rightarrow ee$ events. For background rejection studies the electrons matched to true electron candidates are not included in the analysis. Note that the last bin used for the optimisation of the ID is 45-50 GeV, which is why the signal efficiency increases slightly more in the 50 GeV bin than in others, and the background efficiency increases in this bin as well.

The electron identification performance may be influenced by the parasitic collisions taking place in the same beam crossing (in-time pileup) or a consecutive bunch crossing (out-of-time pileup) as the hard $pp$ collision producing the electron candidate. The number of reconstructed primary vertices is indicative of the level of pileup in each event, with the average number of primary vertices (eight per event) corresponding to an average pileup of 13.7. Since some shower shape distributions depend on the number of pileup collisions per bunch crossing, the cut on the LH discriminant value is loosened as a function of the number of primary vertices. This is done to ensure that the LH identification remains efficient at high pileup, without drastically increasing the amount of background accepted by the LH selection. The optimisation included simulations with a number of pileup collisions of up to 40, covering the range of the pileup observed in 2015.

At high $E_T$, some of the calorimeter variable distributions are different from the typical distributions obtained with $Z \rightarrow ee$ and used to construct the LH PDFs. Higher energy electrons tend to deposit relatively smaller fractions of their energy in the early layers of the EM calorimeter, and more in the later layers of the EM calorimeter or even in the hadronic calorimeter.

Loose and Medium were deemed to be loose enough to be robust against these $E_T$-dependent changes. However, the tighter requirement used in Tight would lead to inefficiencies at high $E_T$, if not handled properly. Thus, for electron candidates with $E_T$ above 125 GeV, Tight uses the same discriminant selection as Medium but adds rectangular cuts on $w_{\text{stot}}$ and $E/p$, which were found to be particularly effective at discriminating signal from background at high $E_T$.

In addition to the multivariate approach used in the LH method described so far, a cut-based method using a set of rectangular cuts on the electron ID discriminating variables was used in Run-1. This method encompasses a similar set of operating points. The cut-based Loose operating point relies primarily on information from the hadronic calorimeter and the first two layers of the EM calorimeter for distinguishing signal from background. The cut-based Medium operating point adds information from the TRT, the transverse impact parameter, and the third layer of the EM calorimeter, in addition to tighter cuts on the

How do we make sure that the leptons we are selecting are good for analysis?

- Lepton isolation can be relative to the lepton $p_T$ (we will use this type of isolation)
- There are several levels identification criteria that depend on likely we require the object to be an actual lepton
- These combine information from several subdetectors to classify the lepton as Loose, Medium or Tight
- We will require **Tight lepton** identification requirements.
Defining good leptons

- Tight ID: `let_type & 520 = true` (& is a bitwise AND in C++)
- Lepton $p_T > 25$ GeV
- Relative track isolation: $\text{lep}_{ptcone30}/\text{lep}_{pt} < 0.15$
- Relative calorimeter isolation: $\text{lep}_{etcone20}/\text{lep}_{pt} < 0.15$
- Leptons in the direction of the arrow are considered as good
Selecting Jets

• The jets we use have been reconstructed with the Anti-\(k_T\) algorithm with \(R=0.4\)
• Jets must have a \(p_T > 25\) GeV
• Jets must also be within \(|\eta| < 2.5\) (tracking fiducial region)
• In a future session we will learn about how the energy of the jet is actually calibrated (which is done before we get to the ROOT ntuple stage we started from)
Selecting Jets

- In the low-jet $p_T$ region ($p_T < 50$ GeV), jets may come from pileup.
- The Jet Vertex Fraction measures the fraction of the tracks associated to the jet that comes from the primary vertex.
- For jets with $p_T < 50$ GeV and $|\eta|<2.4$ we require that
  - $\text{JVF (jet}_j\text{vf)} > 0.5$

![Diagram showing jet vertex fraction and event topology](image)

Table 1: Track selection criteria for all tracks considered for jet-association.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>z_0</td>
</tr>
<tr>
<td>$</td>
<td>\eta</td>
</tr>
<tr>
<td>$\chi^2/\text{ndf} \leq 3.0$</td>
<td></td>
</tr>
<tr>
<td>Si hits ($\text{Pix} + \text{SCT}$) $\geq 7$</td>
<td></td>
</tr>
<tr>
<td>Pixel hits $\leq 5$</td>
<td></td>
</tr>
<tr>
<td>SCT hits $\leq 20$</td>
<td></td>
</tr>
<tr>
<td>TRT hits $\leq 40$</td>
<td></td>
</tr>
</tbody>
</table>

$Z'(2500 \text{ GeV}) \rightarrow tt$
b-jet Tagging

- Jets are considered as coming from a b-quark or b-tagged based on the output of sophisticated algorithm
- In our case we use the output of the MV1 algorithm
- We will use it at a working point at which the b-jet tagging efficiency is 70%
  - MV1 output (jet_MV1) >= 0.7892

Jet

Primary Vertex

Secondary Vertex

Displaced Tracks

Jet

L_{xy}

d_0

\nu

\nu

l^+

v

W^+

\bar{q}

q'

Z'(2500 GeV)\rightarrow tt
• The $Z'$ is a heavy resonance and the top quarks will have a large transverse momentum
• The neutrino from the $W$ boson will have a large transverse momentum too
• We set a requirement on:
  ‣ $\text{MET} > 30 \text{ GeV}$
Constructing new kinematic variables

- In this final state there is one thing we know: there are two W bosons
- The hadronically decaying W boson is hard to reconstruct
- We don’t have the full kinematics of the neutrino from the leptonically decaying W
- What do we do then to exploit the presence of the W bosons?
  - Transverse mass of the W: $M_T^2 = m_T^2 + p_T^2 + 2(E_T,1 E_T,2 - \bar{p}_T,1 \cdot \bar{p}_T,2)$
  
  \[
  E_T^2 = m^2 + (\bar{p}_T)^2
  \]

This can be easily obtained using TLorentzVectors!
Summary of the Event Selection

- The event has one good vertex (hasGoodVertex)
- Exactly one good lepton
  - Tight lepton ID
  - $p_T > 25$ GeV
  - Track isolation: $\text{lep}_\text{ptcone30}/\text{lep}_\text{pt} < 0.15$
  - Calorimeter isolation: $\text{lep}_\text{etcone20}/\text{lep}_\text{pt} < 0.15$
- At least four good jets
  - $p_T > 25$ GeV
  - $|\eta| < 2.5$
  - if $p_T < 50$ GeV and $|\eta|<2.4$, $\text{JVF} > 0.5$
- At least one b-jets
  - $\text{MV1 output} \geq 0.7892$
- MET > 30 GeV
- Transverse mass of the leptonic W: $m_T(\text{lep}, \text{MET}) > 30$ GeV
Thanks!