Measurement of the Jet Mass Distribution in Boosted Top Quark Decays in CMS

Paolo Gunnellini
on behalf of the CMS Collaboration

QCD@LHC 2019 conference
(Buffalo, USA)
Introduction: top mass measurement

Top mass is a crucial parameter for internal consistency of the Standard Model

- How to measure it?

Resolved Topology:
- Ideogram method
- Template method
- Matrix element method

PRO: extremely precise measurement
CON: difficult to relate it to top-quark pole mass

Boosted Topology:
- Mass of the jet containing top decay products

PRO: independent on top threshold effects and on top-quark mass definition
CON: larger uncertainty than resolved topology

Credits: Dennis Schwarz
Introduction: top jet measurement

Top-jet mass is sensitive to top-quark mass

Top jet mass is an analytically calculable quantity

No ambiguity due to top mass definition

Contributions from perturbative and non-perturbative QCD

Unfolded measurements crucial for testing available theoretical predictions
Introduction: reconstruction strategy

Reconstruction of a “fat jet”, clustering all decay products of the top
- Anti-$k_T$ algorithm ($R = 0.8$)
- Cambridge-Aachen algorithm ($R = 1.2$)
- XCone (novel approach of 13 TeV analysis)

\[ R \sim 2m_{\text{top}} / p_T \]

With a transverse momentum of 400 GeV, top quarks are well included in a jet of $R = 1.0-1.2$

- Study of the jet substructure to discriminate top jets from QCD background
- Selection of top-pair semileptonic events

QCD jets: uniform distribution of particles inside the jets
Top jets: presence of three subjets, one for each decay product

Mass of jet can be measured, e.g., from four-momenta of the subjets or all clustered particles (+soft drop..)
Measurement performed in lepton+jet events

- Top-jet candidate:
  - 1 jet with $p_T > 400$ GeV clustered with the Cambridge-Aachen algorithm ($R = 1.2$)

- Presence of exactly one muon (electron) with $p_T > 35$ (45) GeV

Unfolded absolute and normalized cross sections as a function of the top-jet candidate mass

-> Top-jet mass extraction through templates from simulation
Boosted top jet mass measurement at 8 TeV

Dominant uncertainties:

- Statistical (15% in top-mass peak)
- Jet mass scale (5-10%)
- Jet energy scale (10-15%)

Top mass value extracted from jet mass distribution:

$$m_t = 170.8 \pm 6.0 \text{ (stat)} \pm 2.8 \text{ (syst)} \pm 4.6 \text{ (model)} \pm 4.0 \text{ (theo)} \text{ GeV}$$
$$= 170.8 \pm 9.0 \text{ GeV},$$
Boosted top jet mass measurement at 8 TeV

**IMPROVEMENTS ACHIEVABLE AT 13 TeV:**

- More data!
  - Increase in integrated luminosity
  - $\sigma^{13\text{ TeV}}_{\text{top-pair}} > \sigma^{8\text{ TeV}}_{\text{top-pair}}$

-> Decrease in statistical uncertainties

- Reduced sensitivity for pile-up and UE
  - Possibility to use smaller jets and/or different clustering algorithms

-> Decrease in pile-up and model dependence (jet mass) uncertainty
Boosted top mass measurement at 13 TeV

Single lepton trigger

Selection of exactly one lepton with $p_T > 55$ GeV in $|\eta| < 2.4$

Two X Cone jets, where the top-jet (jet1) candidate needs to have 3 anti-kT (R = 0.4) subjets with $p_T > 30$ GeV in $|\eta| < 2.4$ (one b-tagged)

$M_{\text{jet1}} > M_{\text{jet2+lepton}} \rightarrow$ suppression of unmerged hadronic top jets

$p_T^{\text{miss}} > 50$ GeV

Top mass measurement performed by measuring the mass of the hadronic top jet

Credits: Dennis Schwarz
**X cone jet clustering algorithm**

**Exclusive jet algorithm:**
- It returns exactly $N$ jets as output
- Jet axes found by minimizing the N-jettiness of the event
- Particles are clustered around these axes

**Application to semileptonic top events:**
- Set $N = 2$ and $R = 1.2$
- Calculate $\Delta R$ (lep, jet) for both jets
  - Lowest $\Delta R$ (lep, jet): leptonic jet
  - Highest $\Delta R$ (lep, jet): hadronic jet
- Find subjets: 3 in hadronic jet and 2 in leptonic jet
- Combine subjets in final jets

Figure: Hadronic jet: $p_T = 688$ GeV, $m_{\text{jet}} = 191$ GeV

Ref: JHEP11(2015)072
**X cone jet clustering algorithm**

**Not merged:** top decay products are not associated to one of the three subjets reconstructed in the clustered jet.

**Fully merged:** top decay products are clustered inside the fat jet and are associated to one of the three subjets.

*Use of XCone algorithm maximises the number of selected events and the fully-merged fraction of selected events.*
Boosted top mass measurement

Very clean selection!

Top-jet candidate mass for jets with $p_T > 400$ GeV in $|\eta| < 2.4$

Clear peak observed at the top quark mass in data

Backgrounds from W+jets and single top events ($\leq 3\%$)

Predictions from Monte Carlo generators at NLO:
- POWHEG+PYTHIA8 for top-antitop events,
- POWHEG+PYTHIA8 for single-top events
- MADGRAPH_aMC@NLO+PYTHIA8 for W+jets events
Jet mass: pile-up dependence

Jet mass response extremely stable as a function of number of reconstructed primary vertices

-> Top jet mass: mass of the 3 reconstr. subjets
-> W-jet mass: minimum mass of two subjets in the hadr. top

![Diagram of jet mass response](image-url)
Unfolding procedure

Selection of semileptonic top-pair events
- Exactly one lepton with $p_T > 60$ GeV

Hadronic Top-jet candidate (jet1):
- Jet with $p_T > 400$ GeV and three subsets with $p_T > 30$ GeV

$M_{\text{jet1}} > M_{\text{jet2}} + M_{\text{lepton}}$ -> suppression of unmerged hadronic top jets

**Unfolding takes care of the migration effects from detector to particle level**

Tikhonov regularization used for unfolding results, as implemented in TUnfold package

Jet mass bin widths optimized according to detector resolution and migration effects
Experimental and theoretical uncertainties affecting the absolute top-jet mass cross section

- Statistical uncertainty dominates out the peak of the distribution
- Jet energy scale is the major contribution at masses around the top mass
NLO predictions are able to describe very well the cross sections as a function of the top-jet mass.

Unfolded distributions can be compared to analytical predictions as well.

Visible cross section at particle level:

\[ \sigma = 526.8 \pm 16.7\text{(stat)} \pm 38.1\text{(exp)} \pm 28.7\text{(model,unfold)} \text{ fb} \]
\[ = 526.8 \pm 50.5\text{(tot)} \text{ fb}. \]
Differential cross sections are normalized to the visible measured cross section

**Used for top mass extraction through a template method**

Jet mass distribution is sensitive to the choice of top-quark mass used in simulation

From CDF, for illustrative purposes
Experimental and theoretical uncertainties affecting the normalized top-jet mass cross section

- Statistical uncertainty dominates out the peak of the distribution
- Jet energy scale is the major contribution at masses around the top mass
Top jet mass extraction

From differential cross section as a function of jet mass normalized to the visible cross section

**Extraction through $\chi^2$ minimization**

- Pseudo-data from MC simulation
  - POWHEG + PYTHIA 8
- Mass templates extracted from various Monte Carlo event generator predictions
  - POWHEG + HERWIG++, MADGRAPH_aMC@NLO+ PYTHIA8

Parabolic interpolation of the $\chi^2$-values obtained for each discrete top mass point
Top jet mass extraction

Extraction through $\chi^2$ minimization

- Pseudo-data from MC simulation
  POWHEG + PYTHIA 8

- Mass templates extracted from various Monte Carlo event generator predictions
  POWHEG + HERWIG++, MADGRAPH_aMC@NLO+ PYTHIA8

-> No bias seen on pseudo-data tests

Measured extracted mass follows the value of the input truth mass
Top jet mass extraction

Predictions using a different input top mass exhibit a different position of the jet mass peak

\[ m_t = 172.56 \pm 0.44\text{(stat)} \pm 1.57\text{(exp)} \pm 1.55\text{(model, unfold)} \pm 1.02\text{(theo)} \text{ GeV} \]

\[ m_t = 172.56 \pm 2.47\text{(total)} \text{ GeV} \]

-> factor of 4 improvement with respect to CMS measurement at 8 TeV

-> compatibility with top mass value obtained in the resolved regime

CMS-TOP-19-005
Summary and conclusions

**Top mass is one of the fundamental parameters of the Standard Model**
- Measurement from boosted top jets is independent on mass definition
- Jet masses are analytically calculable and could be compared to unfolded data

**CMS has measured the top mass by using boosted hadronic top jets**

**Novel reconstruction approach using the XConE jet clustering algorithm**
- Improvement in the selection of fully-merged top jets
- Improvement in jet mass resolution

**Unfolded differential absolute cross section measurement as a function of jet mass will be released**
- Crucial for comparisons to various theoretical predictions

**Top mass extraction performed through normalized cross sections as a function of jet mass**
- Improvement of a factor of 4 with respect to CMS measurement at 8 TeV

**Obtained result is compatible with measurement in resolved topologies**

\[
m_t = 172.56 \pm 2.47 \text{(total) GeV} \quad \text{vs} \quad M_t = 172.44 \pm 0.13 \text{ (stat)} \pm 0.47 \text{ (syst) GeV}
\]
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrated cross section [fb]</td>
<td>21</td>
<td>48</td>
<td>260</td>
<td>138</td>
<td>59</td>
</tr>
<tr>
<td>Statistical uncertainty [%]</td>
<td>40.7</td>
<td>23.2</td>
<td>5.6</td>
<td>6.9</td>
<td>12.8</td>
</tr>
<tr>
<td>Experimental uncertainty [%]</td>
<td>11.3</td>
<td>15.9</td>
<td>15.6</td>
<td>34.8</td>
<td>14.9</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>4.7</td>
<td>6.7</td>
<td>11.6</td>
<td>31.4</td>
<td>9.1</td>
</tr>
<tr>
<td>Jet energy resolution</td>
<td>7.2</td>
<td>13.1</td>
<td>8.9</td>
<td>11.4</td>
<td>10.2</td>
</tr>
<tr>
<td>XCone jet correction</td>
<td>4.3</td>
<td>3.5</td>
<td>2.5</td>
<td>8.2</td>
<td>3.4</td>
</tr>
<tr>
<td>b tag</td>
<td>5.6</td>
<td>4.5</td>
<td>4.3</td>
<td>4.8</td>
<td>4.3</td>
</tr>
<tr>
<td>Pileup</td>
<td>1.1</td>
<td>1.2</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>MuID</td>
<td>1.1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>1.1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>MuTrigger</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>1.0</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>ElID</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>ElTrigger</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>ElReco</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>Model uncertainty [%]</td>
<td>33.0</td>
<td>32.1</td>
<td>10.1</td>
<td>22.4</td>
<td>26.7</td>
</tr>
<tr>
<td>FSR</td>
<td>2.8</td>
<td>9.8</td>
<td>3.6</td>
<td>18.3</td>
<td>11.4</td>
</tr>
<tr>
<td>Choice of $m_t$</td>
<td>22.0</td>
<td>15.3</td>
<td>2.1</td>
<td>8.8</td>
<td>10.5</td>
</tr>
<tr>
<td>CR</td>
<td>7.2</td>
<td>10.6</td>
<td>7.7</td>
<td>4.9</td>
<td>1.3</td>
</tr>
<tr>
<td>UE tune</td>
<td>13.6</td>
<td>6.3</td>
<td>4.6</td>
<td>6.0</td>
<td>8.1</td>
</tr>
<tr>
<td>$h_{\text{damp}}$</td>
<td>18.4</td>
<td>22.9</td>
<td>&lt; 1</td>
<td>3.7</td>
<td>19.4</td>
</tr>
<tr>
<td>ISR</td>
<td>4.4</td>
<td>4.7</td>
<td>&lt; 1</td>
<td>2.2</td>
<td>5.0</td>
</tr>
<tr>
<td>Scale</td>
<td>1.1</td>
<td>1.7</td>
<td>1.7</td>
<td>2.7</td>
<td>2.2</td>
</tr>
<tr>
<td>PDF</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>Total Uncertainty [%]</td>
<td>53.8</td>
<td>42.8</td>
<td>19.5</td>
<td>42.0</td>
<td>33.3</td>
</tr>
</tbody>
</table>