Measurement of boosted top quark in CMS

Debarati Roy
On behalf of the CMS collaboration

DAE Symposium, 2020
15.12.2020
Boosted top quark (t):

- Measuring properties of t at high momenta ($p_T \sim 400$ GeV) enables detailed studies of kinematic regime that has not been accessible before.

  - Sensitive to new physics
  - Crucial test for SM validation at such high energy regime.

- The more energy the more likely to have higher boost $\Rightarrow$ more boosted t.

- Boosted t decay products are such collimated that it can be reconstructed as a large radius jet with $R \sim 2m_t/p_T$ jet.

Such a large radius jet can be identified by various methods as proposed in jet substructure study!
Boosted top quark latest Measurements in CMS:


Measurement of Jet Mass Distribution in Boosted Top Quark Decays:

- First measurement of the jet mass distribution in hadronic decays of boosted t in ℓ+jets final state at 13 TeV.

- Products of the hadronically decaying boosted t is reconstructed as a large radius jet with $p_T > 400$ GeV.

- Measured differential $t\bar{t}$ production cross section as a function of the jet mass ($m_{\text{jet}} = (\sum p_i)^2$) with integrated luminosity of 35.9 fb$^{-1}$.

- A new jet reconstruction technique (XCon) used for first time at LHC improves the precision by $\sim 3$ relative to earlier measurement ([EPJ C 77 (2017) 467](#)) where CA jet algorithm used ($R = 1.2$) at 8 TeV.

- Peak position of the $m_{\text{jet}}$ distribution is sensitive to the top quark mass ($m_t$) and the data are used to extract a value of $m_t$ to assess the measurement’s sensitivity.
XCon algorithm (a brief overview):

• N-jettiness measures the degree to which the hadrons in the final state are aligned along N jet axes or the beam direction.

• There exists varieties of N-jettiness measure.

• Any choice of measure along with specific algorithm to minimize N-jettiness defines an exclusive jet algorithm.

• XCon : an exclusive cone jet algorithm with requirement of boundaries to approximate circles in the $\eta$-$\phi$ plane, can be achieved by an appropriate choice of jet and beam measures.

• XCon algorithm returns a fixed number of jets, relevant for physics applications where no. of jets known in advance.
Base level event selection:

<table>
<thead>
<tr>
<th>Object</th>
<th>e channel</th>
<th>μ channel</th>
</tr>
</thead>
<tbody>
<tr>
<td>trigger requirement</td>
<td>1 isol. e with $p_T &gt; 27$ GeV OR $p_T &gt; 115$ GeV with no isol. cond. OR 1 γ with $p_T &gt; 175$ GeV, $</td>
<td>\eta</td>
</tr>
<tr>
<td>lepton $p_T$</td>
<td>$&gt; 55$ GeV</td>
<td>$&gt; 55$ GeV</td>
</tr>
<tr>
<td>lepton $</td>
<td>\eta</td>
<td>$</td>
</tr>
<tr>
<td>lepton isolation</td>
<td>2D$^1$</td>
<td>Cone isol. when $p_T &lt; 120$ GeV, 2D$^1$ when $&gt; 120$ GeV</td>
</tr>
<tr>
<td>XCone fat ($R_{jet} = 1.2$) jet with $N_{jet} = 2$ (1$^{st}$ step)</td>
<td>$p_T &gt; 200$ GeV</td>
<td>$p_T &gt; 200$ GeV</td>
</tr>
<tr>
<td>XCone sub ($R_{sub} = 0.4$) jet with $N_{sub} = 3$ (2$^{nd}$ step)</td>
<td>$</td>
<td>\eta</td>
</tr>
</tbody>
</table>

XCone selection results in exactly 2 large radius XCones jets with 3 XCones subjets each!

$^1\Delta R$(lepton, next AK4 jet) > 0.4 or $p_{T}^{rel}$(lepton, next AK4 jet) > 40 GeV
Few important strategies to note:

- $p_4$ of lepton candidate subtracted from $p_4$ of jet if $\Delta R(\ell,j) < 0.4$.
- JEC from AK4 jet applied to XCon jetsub.
- Additional correction applied to XCon jetsub momenta: estimated from hadronic $t\bar{t}$ events ($\sim 2\%$).

- $p_4$ of 3 XCon jetsubs combined to form final XCon jet!
- XCon jet with max. $\Delta R$ from $\ell$ used to perform measurement.
Full event selection:

- XCon e jet1 subjets $p_T > 30$ GeV, $|\eta| < 2.5$
- XCon e jet2 subjets $|\eta| < 2.5$
- XCon e jet1 $p_T > 400$ GeV
- $m_{jet1} > m_{jet2+lepton}$

with MET $> 50$ GeV, $\geq 1$ b-tagged jet

Nice, narrow peak close to $m_t$!

- Much sharper peak with XCon e for a fully contained large jet!
- XCon e results in a large improvement of resolution in $m_{jet}$ ($\sim 6\%$) as compared to CA ($\sim 14\%$).
• Regularised unfolding based on a least-squares fit used.
• Prior to unfolding background contributions subtracted from data.
• Sideband regions considered to constrain migrations into & out of measurement phase space.
• Among uncertainties JES (max. 31%) & FSR (max. 18%) dominate.
• $112 < m_{\text{jet}} < 232$ GeV : $\sigma = 527 \pm 15(\text{stat}) \pm 39(\text{exp}) \pm 29(\text{model})$ fb.

$m_t = 172.6 \pm 0.4(\text{stat}) \pm 1.6(\text{exp})$
$\pm 1.5(\text{model}) \pm 1.0(\text{theo})$ GeV.

Extracted $m_t$ from fit to normalized differential cross section!
Measurement of $t\bar{t}$ differential cross section using top quarks at large transverse momenta in $pp$ collisions at $\sqrt{s} = 13$ TeV:

- Measured in all-jet & $\ell$+jets $t\bar{t}$ final states where at least one $t$ is boosted & its decay products reconstructed as a large radius jet with $p_T > 400$ GeV.

- 2$^{nd}$ $t$ required to decay either in a similar way (all-jet scenario), or leptonically ($\ell$+jets scenario), reconstructed as $e$ or $\mu$, a b jet & MET due to undetected neutrino.

- Cross section extracted as a function of kinematic variables of individual $t$ or of the $t\bar{t}$ system with 35.9 fb$^{-1}$ integrated luminosity.
all-jet channel:

Baseline event selection criteria:

<table>
<thead>
<tr>
<th>Object</th>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>trigger</td>
<td>2 AK8 jets ($p_{T1} &gt; 280$ GeV, $p_{T2} &gt; 200$ GeV), trimmed mass &gt; 30 GeV, at least one of these is b-tagged (only for SR)</td>
</tr>
<tr>
<td>AK8 jet $p_T$</td>
<td>$&gt; 400$ GeV</td>
</tr>
<tr>
<td>AK8 jet $</td>
<td>\eta</td>
</tr>
<tr>
<td>AK8 jet mass</td>
<td>$50 &lt; m_{SD} &lt; 300$ GeV</td>
</tr>
<tr>
<td>lepton veto</td>
<td>$p_T &gt; 20$ GeV, $</td>
</tr>
</tbody>
</table>

Further event categories:

<table>
<thead>
<tr>
<th>Region</th>
<th>Trigger</th>
<th>Offline Requirements</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR</td>
<td>signal</td>
<td>Base+$NN &gt; 0.8+2btags+m_{jet1,2}^{SD} \in (120,220)$ GeV</td>
<td>signal region</td>
</tr>
<tr>
<td>$SR_A$</td>
<td>signal</td>
<td>Base+$NN &gt; 0.8+2btags$</td>
<td>QCD fit region</td>
</tr>
<tr>
<td>$SR_B$</td>
<td>signal</td>
<td>Base+$2btags+m_{SD}^{jet1,2} \in (120,220)$ GeV</td>
<td>signal systematics region</td>
</tr>
<tr>
<td>CR</td>
<td>control</td>
<td>Base+$NN &gt; 0.8+0btag+m_{jet1,2}^{SD} \in (120,220)$ GeV</td>
<td>QCD control region</td>
</tr>
</tbody>
</table>

NN cut 0.8 used to compromise between enough $t\bar{t}$ events & signal purity
Background estimation in all-jet channel:

- Multijet contributes as most dominating background even after significant reduction due to event selection criteria (NN cut, b-tagging).
- Background estimated with data: shape with CR & $N_{\text{multijet}}$ with SR$_{A}$ using a binned maximum likelihood fit of $m_{SD}$ of the AK8 jet with highest $p_{T}$.

Signal extraction:

$$S(x) = D(x) - R_{\text{yield}} N_{\text{multijet}} Q(x) - B(x)$$

- $\bar{t}t$ signal extracted from data by subtracting contribution from background.

$$R_{\text{yield}} \equiv \frac{N^{\text{SR}}_{\text{multijet}}}{N^{\text{SR}_{A}}_{\text{multijet}}} = \frac{N^{\text{CR}}_{\text{multijet}}}{N^{\text{CR}_{A}}_{\text{multijet}}}$$

$\bar{t}t$ & multijet production normalized according to fitted values of respective yields.
Final Result in all-jet channel:

• Unfolding: simple response matrix inversion without regularization in parton & particle phase space carried out.

\[
\frac{d\sigma_i^{\text{unf}}}{dx} = \frac{1}{\mathcal{L}\Delta x_i} \frac{1}{f_{2,i}} \sum_j \left( R_{ij}^{-1} f_{1,j} S_j \right)
\]

\( f_1 \) & \( f_2 \): fractions to quantify overlap between different phase spaces

• Signal extraction, unfolding with different response matrices, extrapolation to particle or parton phase space repeated for every source of uncertainty estimation!

• Dominating uncertainties: JES, b-tagging.

• Good agreement between data & theory in terms of normalized cross section (theory predicts higher than measurement in abs. cross section!).

<table>
<thead>
<tr>
<th>Observable</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_T^{t1} )</td>
<td>&gt; 400 GeV</td>
</tr>
<tr>
<td>(</td>
<td>\eta^{t1}</td>
</tr>
<tr>
<td>( m_{t\bar{t}} )</td>
<td>&gt; 800 GeV</td>
</tr>
</tbody>
</table>

Particle level

<table>
<thead>
<tr>
<th>Observable</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N_{jets} )</td>
<td>&gt; 1</td>
</tr>
<tr>
<td>( p_T^{jet1,2} )</td>
<td>&gt; 400 GeV</td>
</tr>
<tr>
<td>(</td>
<td>\eta^{jet1,2}</td>
</tr>
<tr>
<td>( m_{SD} )</td>
<td>(120, 220) GeV</td>
</tr>
<tr>
<td>( m_{jj} )</td>
<td>&gt; 800 GeV</td>
</tr>
</tbody>
</table>

CMS

35.9 fb\(^{-1}\) (13 TeV)
### Event selection criteria:

<table>
<thead>
<tr>
<th>Object</th>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>trigger</strong></td>
<td>Single lepton with no isol. cond. (1 e with $p_T &gt; 45$ GeV, $</td>
</tr>
<tr>
<td><strong>lepton</strong></td>
<td>Exactly 1 $e(\mu)$ with Tight(Medium) ID, $p_T &gt; 50$ GeV, $</td>
</tr>
<tr>
<td><strong>lepton veto</strong></td>
<td>Medium ID &amp; no $\text{minIso}$ requirement</td>
</tr>
<tr>
<td><strong>AK4 jet (b jet candidate)</strong></td>
<td>At least 1 with $p_T &gt; 50$ GeV, $</td>
</tr>
<tr>
<td><strong>AK8 jet (t jet candidate)</strong></td>
<td>At least 1 with $p_T &gt; 400$ GeV, $</td>
</tr>
<tr>
<td>$p_T^{miss}$</td>
<td>$&gt; 50(35)$ GeV depending on $e$ or $\mu$</td>
</tr>
</tbody>
</table>

*Summed track $p_T$ divided by lepton $p_T$, for tracks in a cone around the lepton whose radius scales inversely with lepton $p_T$.\n
**Summed track $p_T$ divided by lepton $p_T$, for tracks in a cone around the lepton whose radius scales inversely with lepton $p_T$.**

| $|\Delta \phi(\hat{p}_T^X, \hat{p}_T^{miss})| < 1.5 p_T^{miss} / 110$ GeV |

---

**Event categories:**

- **0t:** t jet candidate fails $t$ tag (most background)
- **110b:** t jet candidate passes $t$ tag, b jet candidate fails $b$ tag
- **111b:** t and b jet candidates both pass tag (most signal)

More multijet reduction
Background estimation & Simultaneous Likelihood Fit in $\ell$+jets channel:

- Background estimated from simulation except multijet.
- Multijet estimated with data sideband region by inverting lepton isol. \((0.1 < \text{miniIso} < 0.2)\) & relaxing lepton ID criteria.
- Predicted contributions from signal & other background subtracted from data in sideband region to obtain multijet events.

- Signal strength, tagging efficiency, background normalization extracted through a simultaneous binned maximum-likelihood fit to data across different analysis categories (0t, 1t0b, 1t1b).
- Systematic uncertainties enter as nuisance parameters.

Posterior kinematic distribution in maximum-likelihood fit
Final Result in $\ell$+jets channel:

- Unfolding: matrix inversion without regularization in parton & particle phase space:
  - Parton level
    - Selection: semileptonic $t\bar{t}$, top quark $p_T > 400$ GeV
      - $\mu/e$ (isPromptFinalState, or isDirectPromptTauDecayProductFinalState from isPromptDecayed $\tau$)
    - Top quark: isLastCopy top quark (after FSR) with sign opposite the lepton
  - Particle level
    - Selection
      - $\geq 1$ particle-level AK8 jet with $p_T > 400$ GeV, $|\eta| < 2.4$, and $105 < m_{\text{jet}} < 220$ GeV
      - $\geq 1$ particle-level AK4 jet with $p_T > 50$ GeV, $|\eta| < 2.4$ originating from a b quark
      - $\mu/e$ with $p_T > 50$ GeV, $|\eta| < 2.1$
    - Top jet: leading particle-level AK8 jet

- Before Unfolding background subtracted in 1t1b signal region.
- Data in $e$ & $\mu$ channels combined before unfolding
- Differential cross section extracted from 1t1b category.

Shape describes well overall, theory overpredicts data (consistent with previous studies).
Summary & Outlook:

• Exploiting full Run-2 dataset allows CMS to extend measurement to higher boosts & establish boosted regime for a number of precision measurements in the top quark sector in Run 3 & at the high-luminosity LHC.

• Jet Mass measurement in boosted top quark decay studies the potential of a new technique that results in significant improvement & can be further explored in great detail in near future with boosted kinematics.

• Measurement of differential $t\bar{t}$ cross section provides useful inputs to gain deeper understanding in studying predictions from different MC event generators as well as to perform various measurements in boosted regime with different topologies (e.g. extending phase space to reconstruct large radius jet with semi-leptonically decaying top quark).

• Ongoing studies in CMS puts significant effort to develop methods to identify boosted leptonic top quark (to establish leptonic top tagger).
• Extra Material
Introduction:

• LHC an ideal laboratory to study properties of top quark (t) with unprecedented precision.

• Being heaviest elementary particle, t does not form bound states & decays almost immediately => unique opportunity to study production & decay of a ‘bare’ quark, at energy scales much larger than those typically involved for other quarks.

• Large mass ($m_t$) used to verify self-consistency of SM. Also plays a crucial role in electroweak sector, because among all SM particles it is the largest contributor in terms of radiative corrections to the mass & self-coupling of the Higgs boson.

• Provides important inputs to detector calibration, its performance, as well as to reconstruction algorithms.
More energy $\rightarrow$ boost

As LHC energy increases, objects are produced with more $p_T$.

- In high $p_T$ regime difference between daughter & mother particles is sufficient to provide a boost to the daughter particle: boosted daughter decay then forms a “fatjet” or “large-R jet” e.g. $\text{top}(\rightarrow qq^\prime b)$, $\text{Higgs}(\rightarrow bb^\prime)$, $\text{W/Z}(\rightarrow qq^\prime)$: Leads to subjets inside a fatjet!

![Diagram showing normal and high-$p_T$ scenarios for jet analysis]

more luminosity $\rightarrow$ pile-up

We are inundated with additional pp interactions on top of our process of interest -> fatjet reduces combinatorial nightmare!
More energetic $X$ ($X \to ab$), more collimated decay products ($\text{top} \to q\bar{q}'b$, $W \to q\bar{q}$)

- Boosted characteristics in LHC

- We exploit the “substructure” of the fatjet to identify original particles.

- Fatjet not only includes particles coming from interesting decay, but also from pile-up, UE => needs to be Groomed!