Top quark pair production with additional jets (heavy flavor) from CMS

Tae Jeong Kim (Hanyang University)

For the first MadAnalysis 5 workshop on LHC recasting in Korea 22/August/2017, at High 1
• Mass 14000 tonnes
• Height = 15 m
• Length = 28.7 m
• measure the direction and energy of the particles
• Select 100 events per second among billion events
Integrated luminosity = $\int_{0}^{t} L_{int} dt$ collected at 13 TeV
Introduction to top quark

• Predicted as isospin partner of bottom quark (bottom quark was discovered in 1977)

• The top quark is the most massive of the quarks \( \sim 173 \text{ GeV} \) in the standard model: \( m(t) > m(H) > m(Z) > m(W) \)

• Because of its mass, it was not observed directly at the LEP

• Discovered in 1995 by CDF and D0 experiment

• Lifetime is very short: \( \tau = \frac{1}{\Gamma} = 4 \times 10^{-25} \text{ s}, \Gamma = 1.5 \text{ GeV} \)

• It decays before it can form strongly interacting bound states

• The top quark properties can be studied without non-perturbative effects
Top quark discovery

- Discovered in 1995 by CDF and D0 with only 17 events (D0) and 19 events (CDF)
- Produced 1000s events at Tevatron
top quark pair production at the LHC

Quark annihilation ~10%  Gluon-gluon fusion ~90%

Top mass | Central (pb) | Scale uncert. | PDF+$\alpha_S$ uncert. | Mass uncert.
---|---|---|---|---
172.5 GeV | 831.76 | +19.77-29.20 | +35.06-35.06 | +23.18-22.45
173.2 GeV | 815.96 | +19.37-28.61 | +34.38-34.38 | +22.67-21.95

• With Run 2 data in 2016 only corresponding to around 36 $f b^{-1}$, LHC already produced ~million events after the final selection.
Top quark decay

• ~100% top decays to $bW$

• High $p_T$ isolated leptons
• missing transverse energy ($\nu$)
• Jets, b-jets

Top Pair Decay Channels

<table>
<thead>
<tr>
<th>Decay Channel</th>
<th>all-hadronic</th>
<th>tau+jets</th>
<th>muon+jets</th>
<th>electron+jets</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{c}s$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\bar{u}d$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\bar{t}\tau$</td>
<td>$\tau\tau$</td>
<td>$\tau\tau$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\bar{\mu}\mu$</td>
<td>$\mu\mu$</td>
<td>$\mu\mu$</td>
<td>$\mu\mu$</td>
<td></td>
</tr>
<tr>
<td>$e^-$</td>
<td>$e^-$</td>
<td>$e^-$</td>
<td>$e^-$</td>
<td>$e^-$</td>
</tr>
<tr>
<td>$W$ decay</td>
<td>$e^+$</td>
<td>$\mu^+$</td>
<td>$\tau^+$</td>
<td>$ud$</td>
</tr>
</tbody>
</table>

Dilepton ($e/\mu$) ~ 6.7%
Lepton+jets ~ 35%
All hadronic ~ 45%
Top quark candidate
Motivation of $t\bar{t} +$ additional jets

• $t\bar{t}b\bar{b}$ is the main irreducible background for $t\bar{t}H(b\bar{b})$ searches

• $t\bar{t}jj(c\bar{c})$ is the reducible background faking jets

• It is crucial to understand precisely the $t\bar{t}jj$ and $t\bar{t}b\bar{b}$ processes as these are also the main background for most of new physics searches such as four top search, FCNC, SUSY, ...

• In particular, differential distributions will allow us to check the validity of the QCD calculation involving top quark pair plus additional quarks or gluons.
• For Run 1, the LO MadGraph with up to three additional partons interfaced with PYTHIA 6 is reference $t\bar{t}$ sample.
  • It described most of kinematic distributions except top quark $p_T$ spectrum.
• For Run 2,

<table>
<thead>
<tr>
<th>Event generator</th>
<th>Parton shower</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>POWHEG</td>
<td>Pythia 8</td>
<td>Default for Run 2</td>
</tr>
<tr>
<td>MadGraph5_aMC@NLO</td>
<td>Pythia 8 / Herwig++</td>
<td></td>
</tr>
<tr>
<td>MadGraph5_aMC@NLO</td>
<td>Pythia 8 with FxFx /MLM merging</td>
<td></td>
</tr>
<tr>
<td>POWHEG</td>
<td>Herwig 7/ Herwig++</td>
<td></td>
</tr>
</tbody>
</table>

Pileup reweighting done : MC samples are reweighted to match the distribution of the average number of interactions in data
Particle-level objects

- **Particle-level objects (based on the stable particle)**
  - **Electrons and muons**: $p_T > 25$ GeV and $|\eta|<2.5$
    - not originate from a hadron, Adding the four-momentum of all photons within $\Delta R = 0.1$
  - **Jets**: $p_T > 25$ GeV and $|\eta|<2.5$
    - clustering all stable particle except the selected $e, \mu$ and radiated photons as well as neutrinos using the anti-$k_t$ algorithm with $R=0.4$.
    - Neutrinos from hadron decay are included
  - **b jets**: ghost matching technique - b hadron momentum is scaled down to a negligible value and included in the jet clustering.
Particle level top quarks

• In the lepton + jets, take the jet permutation by minimizing following quantity

\[
K^2 = [M(p_\nu + p_\ell + p_{b_\ell}) - m_t]^2 \\
+ [M(p_{j_1} + p_{j_2}) - m_W]^2 \\
+ [M(p_{j_1} + p_{j_2} + p_{b_h}) - m_t]^2
\]

• b-tagged jets are assigned first with higher priority
lepton + jets at 13 TeV (CMS)

- Only one electron or muon
  \( p_T > 30 \text{ GeV}, |\eta| < 2.1 \)
- No additional lepton
  \( p_T > 15 \text{ GeV}, |\eta| < 2.4 \)
- At least four jets
  \( p_T > 30 \text{ GeV}, |\eta| < 2.4 \)
- At least two b-tagged jets

Number of permutation increases drastically → more likely select a wrong permutation.
lepton + jets at 13 TeV (CMS)

Recoiling objects depend on the transverse momentum of $t\bar{t}$

- Data has softer $t\bar{t}$ system $p_T$ spectrum
- Theoretical uncertainties are reduced in the particle-level measurements as these have less dependency on theory-based extrapolation
lepton + jets at 13 TeV (CMS)

Most of MC predictions overshoot data

PRD 95, 092001 (2017)
lepton + jets at 13 TeV (CMS)

NNLO has a better agreement with data

PRD 95, 092001 (2017)
Tuning with $\alpha_s^{ISR}$ and $h_{damp}$

POWHEG: $h_{damp}$ = controls of the $p_T$ of the first additional emission beyond the Born configuration (default is top quark mass 172.5 GeV) → regulate the high-$p_T$ emission against top quark pair system recoils

PYTHIA 8: $\alpha_s^{ISR}$ is the value of the strong coupling at $m_Z$ (default is 0.1365)

damping real emission generated by POWHEG with a factor $\frac{h_{damp}^2}{(p_T^2+h_{damp}^2)}$

Tune using the 8 TeV data

$h_{damp} = 1.581^{+0.658}_{-0.585} \times m_t, \quad \alpha_s^{ISR} = 0.1108^{+0.0145}_{-0.0142}$
$t\bar{t}$ $p_T$ spectrum and jet multiplicity

CMS Preliminary 2.3 fb$^{-1}$ (13 TeV)

NLO generators agree with data within uncertainty
LO order of MG5_aMC@NLO (MLM configuration) and aMC@NLO do not agree with data

CUETP8M2T4 is new tune
CUETP8M1 is old tune
Comparison with different $h_{damp}$ with $h_{damp} = 1.581m_{top}$ being the central value

Data disfavors vanishing $h_{damp}$
$t\bar{t}$ $p_T$ spectrum in 2D (CMS)

The 0-additional jet case has the worst agreement data and theory predictions.
Additional jet distribution from CMS

$1/\sigma_{\text{full}} \frac{d\sigma}{d\Delta R_{jj}}$

$1/\sigma_{\text{full}} \frac{d\sigma}{d m_{jj}}$
Jet multiplicity in lepton + jets with 36 $fb^{-1}$

- POWHEG+Pythia8 prediction of the jet multiplicity is consistent with data
- The jet multiplicity from previous 8 TeV measurements was used for CUETP8M2T4 tune
- The tune accurately described the jet multiplicity on a larger dataset with a higher $\sqrt{s}$
Cut flow for $t\bar{t}b\bar{b}$ in dilepton decay mode

**Figure:**

- **Events:**
  - $\text{CMS} \quad 2.3 \text{ fb}^{-1} (13 \text{ TeV})$
  - Legend: $\text{t\bar{t}b\bar{b}}$, $\text{t\bar{t}LF}$, $\text{Single t}$, $\text{t\bar{t}bj}$, $\text{t\bar{t}c\bar{c}}$, $\text{t\bar{t}V}$, $\text{Data}$

**Table:**

<table>
<thead>
<tr>
<th>Process</th>
<th>$e^+e^-$</th>
<th>$\mu^+\mu^-$</th>
<th>$e^{\pm}\mu^{\mp}$</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t\bar{t}bb$</td>
<td>6.3 ± 0.4</td>
<td>8.6 ± 0.4</td>
<td>24 ± 1</td>
<td>39 ± 1</td>
</tr>
<tr>
<td>$t\bar{t}bj$</td>
<td>16 ± 1</td>
<td>21 ± 1</td>
<td>57 ± 2</td>
<td>95 ± 2</td>
</tr>
<tr>
<td>$t\bar{t}c\bar{c}$</td>
<td>7.7 ± 0.4</td>
<td>11 ± 1</td>
<td>27 ± 1</td>
<td>46 ± 1</td>
</tr>
<tr>
<td>$t\bar{t}LF$</td>
<td>157 ± 2</td>
<td>220 ± 2</td>
<td>596 ± 1</td>
<td>972 ± 4</td>
</tr>
<tr>
<td>$t\bar{t}$ others</td>
<td>18 ± 1</td>
<td>19 ± 1</td>
<td>61 ± 1</td>
<td>99 ± 1</td>
</tr>
<tr>
<td>$t\bar{t}$ V</td>
<td>2.5 ± 0.1</td>
<td>3.2 ± 0.2</td>
<td>7.3 ± 0.2</td>
<td>14 ± 1</td>
</tr>
<tr>
<td>Single $t$</td>
<td>6.6 ± 0.8</td>
<td>8.4 ± 0.8</td>
<td>23 ± 2</td>
<td>39 ± 2</td>
</tr>
<tr>
<td>$Z+$jets</td>
<td>0.8$^{+1.0}_{-0.8}$</td>
<td>5.4 ± 1.5</td>
<td>0.6 ± 0.5</td>
<td>6.8 ± 1.9</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>215 ± 2</td>
<td>297 ± 3</td>
<td>796 ± 4</td>
<td>1311 ± 6</td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td>186</td>
<td>288</td>
<td>682</td>
<td>1156</td>
</tr>
</tbody>
</table>

**Obs/Exp:**

- $0$, $1$, $2$, $3$, $\geq 4$

**Dataset:**

- **CMS:**
  - (13 TeV)
  - $2.3 \text{ fb}^{-1}$

**ArXiv:**

- arXiv:1705.10141
**tt\bar{t} + Heavy flavor from CMS**

- Rearrange jets in b-tagging algorithm discriminator
- Using b-tagging algorithm discriminator of third and fourth jets

**Simulation CMS (13 TeV)**

![Graph of ttbb and ttLF](image)

[arXiv:1705.10141](https://arxiv.org/abs/1705.10141)
### $t\bar{t}b\bar{b}(jj)$ cross sections

**Full phase space vs Visible phase space**

<table>
<thead>
<tr>
<th>Phase Space (PS)</th>
<th>Parton level</th>
<th>Particle level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visible PS</td>
<td>–</td>
<td>4 (b) jets and 2 leptons (e, $\mu$)</td>
</tr>
<tr>
<td>Full PS</td>
<td>$t$, $\bar{t}$ and 2 (b) jets (not from $t$ or $\bar{t}$)</td>
<td>–</td>
</tr>
</tbody>
</table>

**leptons :** $p_T > 20$ GeV, $|\eta| < 2.4$ / **Jets :** $p_T > 20$ GeV, $|\eta| < 2.5$

### Results

<table>
<thead>
<tr>
<th>Phase space</th>
<th>$\sigma_{t\bar{t}b\bar{b}}$ [pb]</th>
<th>$\sigma_{t\bar{t}jj}$ [pb]</th>
<th>$\sigma_{t\bar{t}b\bar{b}} / \sigma_{t\bar{t}jj}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visible</td>
<td>Measurement</td>
<td>0.088 ± 0.012 ± 0.029</td>
<td>3.7 ± 0.1 ± 0.7</td>
</tr>
<tr>
<td></td>
<td>SM (POWHEG)</td>
<td>0.070 ± 0.009</td>
<td>5.1 ± 0.5</td>
</tr>
<tr>
<td>Full</td>
<td>Measurement</td>
<td>4.0 ± 0.6 ± 1.3</td>
<td>184 ± 6 ± 33</td>
</tr>
<tr>
<td></td>
<td>SM (POWHEG)</td>
<td>3.2 ± 0.4</td>
<td>257 ± 26</td>
</tr>
</tbody>
</table>

- Theoretical ratios are lower than the measured values
- But consistent within two standard deviations
Comparisons with various MC predictions

CMS Unpublished
$\sqrt{s} = 13\text{TeV}, 2.3 \text{ fb}^{-1}$

Full phase space

CMS Unpublished
$\sqrt{s} = 13\text{TeV}, 2.3 \text{ fb}^{-1}$

Full phase space

- Measurement
- Stat  Total
- POWHEG v2 P8M1
- MG5_aMC@NLO [FxFx] P8M1
- MG5_aMC@NLO [MLM] P8M1
- POWHEG v2 H++ EE5C
- MG5_aMC@NLO H++ EE5C

arXiv:1705.10141
Comparisons with various MC predictions

Visible phase space

\[ \sqrt{s} = 13\text{TeV}, 2.3 \text{ fb}^{-1} \]

**CMS Unpublished**

Visible phase space

- Measurement
  - Stat
  - Total

- POWHEG v2 P8M1
- MG5_aMC@NLO [FxFx] P8M1
- MG5_aMC@NLO [MLM] P8M1

`arXiv:1705.10141`
Differential $t\bar{t}b\bar{b}$ cross sections

$\Delta R = \sqrt{\Delta \phi(b, b)^2 + \Delta \eta(b, b)^2}$
Differential $t\bar{t}b\bar{b}$ cross sections

CMS

19.7 fb$^{-1}$ (8 TeV)

Events / 80 GeV

CMS

$\sigma (t\bar{t}b\bar{b})$ [pb/GeV]

Data

Theory

Stat.

Syst.

Stat. $\oplus$ Syst.

MC

Madgraph+Pythia6

MC@NLO+Herwig6

Powheg+Pythia6

Powheg+Herwig6

19.7 fb$^{-1}$ (8 TeV)

Conclusion

• We are towards very precision measurement in top quark measurement, in particular, using differential distributions.

• $t\bar{t}+X$ is the main background of many new physics. It is crucial to control experimental and systematic uncertainties.

• Differential $t\bar{t}b\bar{b}$ cross section measurement with more data of $36 \text{ fb}^{-1}$ is in progress.

• More interaction between theory and experiments is required.