First studies towards top-quark pair differential cross section measurement in the dilepton channel at $\sqrt{s} = 13$ TeV with the CMS detector

Mykola Savitskyi, Maria Aldaya, Carmen Diez Pardos et al.

DESY-CMS Top Group, Hamburg
The top quark is special

Heaviest elementary particle known to date: $m_t \approx 173$ GeV

Mass close to scale of electroweak symmetry breaking (EWSB):
$\rightarrow y_t \approx 1$: important role in EWSB?

Decays before hadronizing: $\tau \approx 5 \cdot 10^{-25}$ s $<< 1/\Lambda_{\text{QCD}}$: unique window on "bare" quark

Sensitive to physics phenomena beyond the Standard Model (BSM):
$\rightarrow$ new physics may preferentially couple/decay to top

Major source of background for many Higgs and BSM searches
Why measure differentially?

Precise understanding of top quark distributions is crucial:

- Precision tests of perturbative QCD for top-quark production at different phase space regions
- Tune and test theory predictions and models: → potential to reduce signal modelling systematics
- Enhance sensitivity to BSM physics

LHC is a “top factory”

- Several millions of top-quark pair ($t\bar{t}$) events produced already in Run-I ($\sqrt{s} = 7, 8$ TeV)
- Run-II: expect much larger data sets at $\sqrt{s} = 13(14)$ TeV → important to measure top quark distributions with very high precision
**Top Pair Production and Decay**

**Production cross sections:**

<table>
<thead>
<tr>
<th>Energy, TeV</th>
<th>8</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{\text{incl}}^{\text{NNLO+NNLL}}(t\bar{t}), \text{ pb}$</td>
<td>252.89</td>
<td>831.76</td>
</tr>
</tbody>
</table>

**Weak interaction: Top decay**

$t \rightarrow W+b \sim 100\%$

**Top Pair Branching Fractions**

- \textbf{"alljets"} 46%
- $\tau$+jets 15%
- $e$+jets 15%
- $\mu$+jets 15%
- \textbf{\textit{dileptons}}
- lepton+jets (\textit{\ell}+jets), all jets

**Strong interaction: Top pairs**

- $qq$ annihilation
- $gg$ fusion

<table>
<thead>
<tr>
<th>LHC Energy, TeV</th>
<th>8</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>$gg \rightarrow t\bar{t}$</td>
<td>$\sim$82%</td>
<td>$\sim$90%</td>
</tr>
<tr>
<td>$qq \rightarrow t\bar{t}$</td>
<td>$\sim$18%</td>
<td>$\sim$10%</td>
</tr>
</tbody>
</table>
Selection in Dilepton Channel

1. Trigger conditions

2. **Lepton pair** selection:
   - opposite charge
   - $e$ and $\mu$ isolation criteria
   - $p_T > 20$ GeV, $|\eta| < 2.4$
   - invariant mass: $m_{ll} > 20$ GeV

3. Exclusion of Z-region:
   - in $ee$ and $\mu\mu$: $76$ GeV $< m_{ll} < 106$ GeV

4. Presence of **two jets** ($\text{anti-k}_t$ R=0.4) with $p_T > 30$ GeV and $|\eta| < 2.4$

5. Missing $E_T > 40$ GeV in $ee$ and $\mu\mu$

6. At least one **b-tagged jet**

7. Meaningful solution for kinematic event reconstruction

**Largest backgrounds:** other $t\bar{t}$, Single top, Drell-Yan events
Kinematic Reconstruction

- Measured input: 2 jets, 2 leptons, MET
- Unknowns: $\bar{p}_v, \bar{p}_{\nu} \rightarrow 6$
- Constraints:
  $> m_t, m_{\tilde{t}} \rightarrow 2$
  $> m_{W(+)}, m_{W(-)} \rightarrow 2$
  $> (\bar{p}_v + \bar{p}_{\nu})_T = \text{MET} \rightarrow 2$

Reconstructing each event 100 times and smearing inputs by their resolution:
- top mass fixed to 172.5 GeV
- W mass at RECO level smeared accordingly to W mass distribution
- Jet and lepton energies are corrected for detector effects

Consider weighted average of solutions for all smeared points:

$$p_{x,y,z}^{\text{top}} = \frac{1}{w} \sum_{i=0}^{100} w_i \cdot (p_{x,y,z}^{\text{top}})_i$$
Selection in Dilepton Channel

**Channel**: combined (ee+eμ+μμ)

**Pseudo-data**: poisson-smeared sum of all signal and background simulation samples

**Normalized to**: \( L_{\text{int}} = 5.0 \text{ fb}^{-1} \)

**Observable**: Lepton \( \eta \)

**Signal**: MadGraph+Pythia8

**Selection steps**:
1. Trigger conditions
2. Lepton pair selection:
   - opposite charge
   - \( e \) and \( \mu \) isolation criteria
   - \( p_T > 20 \text{ GeV}, |\eta| < 2.4 \)
   - invariant mass: \( m_{ll} > 20 \text{ GeV} \)
Selection in Dilepton Channel

Channel: combined (ee+eμ+μμ)

Pseudo-data: poisson-smeared sum of all signal and background simulation samples

Normalized to: \(L_{\text{int}} = 5.0\text{ fb}^{-1}\)

Observable: Lepton \(\eta\)

Signal: MadGraph+Pythia8

Selection steps:
1. Trigger conditions
2. Lepton pair selection:
   - opposite charge
   - e and \(\mu\) isolation criteria
   - \(p_T > 20\text{ GeV}, |\eta| < 2.4\)
   - invariant mass: \(m_{ll} > 20\text{ GeV}\)
3. Exclusion of \(Z\) -region:
   - in ee and \(\mu\mu\): \(76\text{ GeV} < m_{ll} < 106\text{ GeV}\)
Channel: combined (ee+eμ+μμ)  
Pseudo-data: poisson-smeared sum of all signal and background simulation samples  
Normalized to: \( L_{int} = 5.0 \text{ fb}^{-1} \)

Observable: Lepton \( \eta \)  
Signal: MadGraph+Pythia8

Selection steps:
1. Trigger conditions
2. Lepton pair selection:  
   - opposite charge  
   - e and \( \mu \) isolation criteria  
   - \( p_T > 20 \text{ GeV}, |\eta| < 2.4 \)  
   - invariant mass: \( m_{ll} > 20 \text{ GeV} \)
3. Exclusion of Z -region:  
   - in ee and \( \mu\mu \): 76 GeV \(< m_{ll} < 106 \text{ GeV} \)
4. Presence of two jets:  
   - with \( p_T > 30 \text{ GeV} \) and within \( |\eta| < 2.4 \)
Selection in Dilepton Channel

Channel: **combined** \((ee+e\mu+\mu\mu)\)

**Pseudo-data**: poisson-smeared sum of all signal and background simulation samples

Normalized to: \(L_{int} = 5.0 \text{ fb}^{-1}\)

**Observable**: Lepton \(\eta\)

**Signal**: MadGraph+Pythia8

**Selection steps:**

1. Trigger conditions
2. Lepton pair selection:
   - opposite charge
   - e and \(\mu\) isolation criteria
   - \(p_T > 20 \text{ GeV}, |\eta| < 2.4\)
   - invariant mass: \(m_{ll} > 20 \text{ GeV}\)
3. Exclusion of Z -region:
   - in ee and \(\mu\mu\): \(76 \text{ GeV} < m_{ll} < 106 \text{ GeV}\)
4. Presence of two jets:
   - with \(p_T > 30 \text{ GeV}\) and within \(|\eta| < 2.4\)
5. Missing \(E_T > 40 \text{ GeV}\) in ee and \(\mu\mu\)
Selection in Dilepton Channel

Channel: combined \((ee+e\mu+\mu\mu)\)

Pseudo-data: poisson-smeared sum of all signal and background simulation samples

Normalized to: \(L_{\text{int}} = 5.0 \text{ fb}^{-1}\)

Observable: Lepton \(\eta\)

Signal: MadGraph+Pythia8

Selection steps:
1. Trigger conditions
2. Lepton pair selection:
   - opposite charge
   - e and \(\mu\) isolation criteria
   - \(p_T > 20 \text{ GeV}, |\eta| < 2.4\)
   - invariant mass: \(m_{ll} > 20 \text{ GeV}\)
3. Exclusion of Z-region:
   - in \(ee\) and \(\mu\mu\): \(76 \text{ GeV} < m_{ll} < 106 \text{ GeV}\)
4. Presence of two jets:
   - with \(p_T > 30 \text{ GeV}\) and within \(|\eta| < 2.4\)
5. Missing \(E_T > 40 \text{ GeV}\) in \(ee\) and \(\mu\mu\)
6. At least one \(b\)-tagged jet
Selection in Dilepton Channel

Channel: **combined** (ee+eμ+μμ)

**Pseudo-data:** poisson-smeared sum of all signal and background simulation samples

Normalized to: \( L_{\text{int}} = 5.0 \text{ fb}^{-1} \)

**Observable:** Lepton \( \eta \)

**Signal:** MadGraph+Pythia8

**Selection steps:**

1. Trigger conditions
2. Lepton pair selection:
   - opposite charge
   - e and μ isolation criteria
   - \( p_T > 20 \text{ GeV}, |\eta| < 2.4 \)
   - invariant mass: \( m_{ll} > 20 \text{ GeV} \)
3. Exclusion of Z -region:
   - in ee and μμ: 76 GeV < \( m_{ll} < 106 \text{ GeV} \)
4. Presence of two jets:
   - with \( p_T > 30 \text{ GeV} \) and within \( |\eta| < 2.4 \)
5. Missing \( E_T > 40 \text{ GeV} \) in ee and μμ
6. At least one b-tagged jet
7. Meaningful solution for kinematic event reconstruction
Selection in Dilepton Channel

Channel: combined (ee+eμ+μμ)

Pseudo-data: poisson-smeared sum of all signal and background simulation samples

Normalized to: \( L_{\text{int}} = 5.0 \text{ fb}^{-1} \)

Observable: Lepton \( \eta \)

Signal: MadGraph+Pythia8

Table given after all cuts:

<table>
<thead>
<tr>
<th>Process</th>
<th>Fraction, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>( tt ) Signal</td>
<td>78.7</td>
</tr>
<tr>
<td>( tt ) Other</td>
<td>14.1</td>
</tr>
<tr>
<td>( tW )</td>
<td>3.3</td>
</tr>
<tr>
<td>W+Jets</td>
<td>0.2</td>
</tr>
<tr>
<td>( DY \to ee/\mu\mu )</td>
<td>2.4</td>
</tr>
<tr>
<td>( DY \to \tau\tau )</td>
<td>0.6</td>
</tr>
<tr>
<td>( tt+Z/W/\gamma )</td>
<td>0.7</td>
</tr>
</tbody>
</table>
Control Distributions

Channel: combined \((ee+e\mu+\mu\mu)\), \(L_{\text{int}} = 5.0\ \text{fb}^{-1}\)

Plots given after full selection

Pseudo-data: poisson-smeared sum of all signal and background simulation samples
Differential Cross Section

- For each observable $X$ the normalized differential cross section in the $i$-th bin is defined as:

$$\left( \frac{1}{\sigma} \frac{d\sigma}{dX} \right)^i = \frac{1}{\sigma} \frac{N_{\text{events}}^i}{\Delta^i_X L}$$

- $N_{\text{events}}^i$ - number of events after background subtraction, efficiency, acceptance and bin-to-bin migration correction
- $\sigma$ - total $t\bar{t}$ cross section in same phase space
- $L$ - integrated luminosity
- $\Delta^i_X$ - bin width

- **Phase space** definition:

  **Top quarks, $t\bar{t}$ system** (obtained via kinematic reconstruction of event) – extrapolated to full phase space after corrections for detector and hadronization effects

  **Leptons or b-jets** – measured in visible phase space ($leptons$: $p_T > 20$ GeV, $|\eta| < 2.4$; $jets$: $p_T > 30$ GeV, $|\eta| < 2.4$) after correction for detector effects

- Bin-to-bin migrations are reduced by binning optimization and corrected by unfolding
Binning and Migrations

- Migration effects studied by:
  
  \[ p_i = \frac{N_{i \text{rec} \& \text{gen}}}{N_{i \text{rec}}} \]  
  - **purity**: sensitive to migrations to \( i \)-th bin

  \[ s_i = \frac{N_{i \text{rec} \& \text{gen}}}{N_{i \text{gen}}} \]  
  - **stability**: sensitive to migrations out of \( i \)-th bin

  \[ \varepsilon_i = \frac{N_{i \text{rec} \& \text{sel}}}{N_{i \text{all \ generated}}} \]  
  - **efficiency** in \( i \)-th bin

- Binning criteria: stability or purity \( \geq ~0.5 \)

- Response matrices are constructed from signal MC
Measured Cross Sections: leptons

- Normalization allows to:
  - reduce systematic uncertainties
  - perform shape comparisons of different theory models to data

- Systematic uncertainty: include JES and JER uncertainties for now
Measured Cross Sections: tops

Normalization allows to:
- reduce systematic uncertainties
- perform shape comparisons of different theory models to data

Systematic uncertainty: include JES and JER uncertainties for now
Measured Cross Sections: $t\bar{t}$-pair

- Normalization allows to:
  - reduce systematic uncertainties
  - perform shape comparisons of different theory models to data

- Systematic uncertainty: include JES and JER uncertainties for now
Conclusions & Outlook

• **Top-quark pair differential cross section measurements:**
  - Essential for constraining the SM
  - Ideal probe for looking for new physics beyond the SM
  - Needed for tuning of PDF sets and modern art Monte-Carlo event generators

• **First studies towards measurements at 13 TeV were presented:**
  - Not latest status & results → currently **Top** secret at CMS and work in progress!
  - Optimize binning with final configuration of the data analysis
  - Evaluate systematic uncertainties
  - Compare to different theory predictions
Conclusions & Outlook

- **Top-quark pair differential cross section measurements:**
  - Essential for constraining the SM
  - Ideal probe for looking for new physics beyond the SM
  - Needed for tuning of PDF sets and modern art Monte-Carlo event generators

- **First studies towards measurements at 13 TeV were presented:**
  - Not latest status & results → currently Top secret at CMS and work in progress!
  - Optimize binning with final configuration of the data analysis
  - Evaluate systematic uncertainties
  - Compare to different theory predictions

Thank you for your attention! :-(
Backup
LHC Collider

**Run 1:** $L_{\text{int}} \approx 25 \text{ fb}^{-1}$ in 2010-2012 at 7/8 TeV

**Run 2:** Phase 0 in 2015-2017
- design energy: $\sqrt{s}=13 \text{ TeV} \sim 14 \text{ TeV}$
- nominal luminosity: $L \sim 1 \cdot 10^{34} \text{cm}^{-2}\text{s}^{-1}$
- bunch spacing: 25 ns
- pile up: ~49
- $L_{\text{int}}$ per year $\sim 45 \text{ fb}^{-1}$

**Planned to collect:** $L_{\text{int}} \sim 75 - 100 \text{ fb}^{-1}$
The CMS Detector

General purpose $4\pi$ detectors:

**Tracker**: Detection and momentum measurement for charged particles

**Calorimeter**: Identification and energy measurement of jets and electrons

**Muon system**: Identification and momentum measurement of muons

---

---

**Phase 0 (2015-2017) upgrade:**

- Complete muon coverage
- Colder tracker
- Photodetectors in HCAL
- New beampipe and infrastructure updates
Visible Phase Space Definition

- **Object definition at generator level:**
  > particles after radiation and hadronization
  > **leptons**: from W-decay
  > **jets**: anti-kT (with cone of $\Delta R=0.5$) algorithm
  > **b-jets**: identified by B-hadrons

- **Directly measured quantities**: leptons and b-jets

Visible particle LVL Phase Space

**Correct for:**
Detector Effects

Measure at visible Phase Space:
- **Leptons**: $p_T > 20$ GeV, $|\eta| < 2.4$
- **Jets**: $p_T > 30$ GeV, $|\eta| < 2.4$
Unfolding

- **Unfolding** techniques correct migrations between bins
- **Response matrix** ($A$): represents bin-by-bin correlations
- Unfolding problem is transformed to $\chi^2$ - minimization problem:
  
  $$\chi^2 = \left( \vec{N} - A \cdot \vec{x} \right)^T \text{COV}_N^{-1} \left( \vec{N} - A \cdot \vec{x} \right) - \tau^2 \cdot K \left( \vec{x} \right)$$

  - $\textbf{N}$: BG corrected data
  - $\textbf{x}$: unfolded result
- Non-physical fluctuations removed by means of the regularization:
  
  $\rightarrow \tau$ – continuous regularization parameter
  
  $\rightarrow$ selected at minimum of average global correlation

- In this measurement regularized unfolding is used