Measurement of the differential cross section for top-quark pair production in the dilepton channel at $\sqrt{s} = 13$ TeV with the CMS detector

Maria Aldaya, Till Arndt, Carmen Diez Pardos, Alexander Grohsjean, Ali Harb, Johannes Hauk, Eleni Ntomari, Mykola Savitskyi
(DESY)
Precise understanding of top quark distributions is crucial:

- Precision tests of perturbative QCD for top quark production at different phase space regions
- Theory predictions and models need to be tuned and tested with measurements: potential to reduce signal modelling systematics
- Extract/use for PDF fits
- Enhance sensitivity to New Physics
- Background for Higgs, rare processes and many BSM searches

From Run-I and early Run-II: need better understanding of \( p_T(\text{top}) \), \( p_T(\ttbar) \), \( m(\ttbar) \)

- \( p_T(\text{top}) \) spectrum softer in data observed at 7 & 8 TeV
- Impact on many precision measurements and searches
General analysis strategy

- **Goal**: measure $\sigma(\bar{t}t)$ as a function of top quark or $\bar{t}t$ system observables
  - look into slope of $p_T^{(\text{top})}$ at 13 TeV
  - consistency check with theory calculations

- **Main analysis ingredients**
  - Event selection
  - $\bar{t}t$ kinematic reconstruction
  - Bin-wise cross section measurement
  - **Unfolding**: correct for detector effects & acceptance to parton or particle level after background subtraction

- **Differential $\bar{t}t$ cross sections**
  - Normalize to in-situ measured $\sigma(\bar{t}t)$: mostly shape uncertainties contribute

Following the same analysis strategy as public results by CMS:
- at 8 TeV → Eur. Phys. J. C 75 (2015) 542 (dilepton & l+jets)
- at 13 TeV → PAS TOP-15-010 (dilepton)
Event selection in dilepton channel

- Dilepton trigger selection
- ≥ 2 high-$p_T$ leptons (ee, $\mu\mu$, $\mu e$)
  - $p_T > 20$ GeV, $|\eta| < 2.4$
  - opposite charge
  - isolation criteria
- QCD veto: $m_{ll} > 20$ GeV
- ≥ 2 jets: anti-$k_T$ $R = 0.4$
  - $p_T > 30$ GeV, $|\eta| < 2.4$
  - Jet cleaning: $\Delta R(l, \text{jet}) > 0.4$ against selected leptons
- ≥ 1 b-tagged jets
- ee, $\mu\mu$ channels: $E_T^{\text{miss}} > 40$ GeV
  - Z veto: $|m_Z - m_{ll}| > 15$ GeV

In addition: kinematic reconstruction of $t\bar{t}$ system → event excluded, if no solution found

Largest backgrounds:
- other $t\bar{t}$
- single top (tW)
- Drell-Yan process
Kinematic distributions: particle objects

- After b-tagging requirement
- Combined channel shown
- Dominant backgrounds: $t\bar{t}$ other, single top, $Z+$jets
- $t\bar{t}$ other includes all non-dilepton decays
- Reference $t\bar{t}$ prediction: Powheg+Pythia8
- MC corrected for efficiencies in data

Dominated by statistical uncertainty

**Left:** first $L = 42 \text{ pb}^{-1}$ of 13 TeV data (50ns)
Kinematic distributions: particle objects

- After b-tagging requirement
- Combined channel shown
- Dominant backgrounds: $t\bar{t}$ other, single top, $Z+jets$
- $t\bar{t}$ other includes all non-dilepton decays
- Reference $t\bar{t}$ prediction: Powheg+Pythia8
- MC corrected for efficiencies in data

In general, good data-to-MC agreement

Left: first $L = 42 \text{ pb}^{-1}$ of 13 TeV data (50ns)

Right: $L = 2.3 \text{ fb}^{-1}$ (13 TeV, 25ns)

CMS PAS-TOP-15-010

Run II: 2015 25ns data

Leptons / 15 GeV

$P_T(\text{lep})$

$P_T(b\text{-jets})$
Kinematic reconstruction of $t\bar{t}$ system

- Measured input: 2 jets, 2 leptons, MET
- Unknowns: $p_\nu^+, p_\nu^-$ → 6
- Constraints:
  - $m_t, m_\bar{t} → 2$
  - $m_{W^+}, m_{W^-} → 2$
  - $(p_\nu^+ + p_\nu^-)_T = \text{MET} → 2$
- 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:
  $$p_{x,y,z}^{\text{top}} = \frac{1}{W} \sum_{i=0}^{100} w_i \cdot (p_{x,y,z}^{\text{top}})_i$$
Data: L = 2.3 fb⁻¹  
(13 TeV, 25ns)

- Very pure $\bar{t}t$ signal after full event selection & kinematic reconstruction (~80%)
- Combined channel shown
- Reference $\bar{t}t$ prediction: Powheg+Pythia8

In general, good data-to-MC agreement
Kinematic distributions: $t\bar{t}$ system

- Very pure $t\bar{t}$ signal after full event selection & kinematic reconstruction (~80%)

- Combined channel shown

- Reference $t\bar{t}$ prediction: Powheg+Pythia8

Data: $L = 2.3 \text{ fb}^{-1}$ (13 TeV, 25ns)

In general, good data-to-MC agreement
Number of events

Response matrix $A_{ij}$

Binning

Chosen to limit migration effects in and out of bins:

- purity ($p_i$) & stability ($s_i$): $\geq 40$-$50\%$
- $\approx$ flat in all bins

$$p_i = \frac{N_i^{\text{rec & gen}}}{N_i^{\text{rec}}} \quad s_i = \frac{N_i^{\text{rec & gen}}}{N_i^{\text{gen}}}$$

Phase space

- Correct back to parton level in full phase space
- Top quark definition: before decay and after QCD radiation

Regularized unfolding

- Basic unfolding - simple inversion of response matrix $A_{ij}$:
  $$N_{i,\text{unf}} = A_{ij}^{-1} N_{j,\text{measured}}$$
- Regularization used to remove large statistical fluctuations (SVD)
Results: \( p_T^{(\text{top})}, y^{(\text{top})} \)

- **Left**: \( L = 42 \text{ pb}^{-1} \)

- Full phase space, parton level

- Reference \( \bar{t}t \) prediction used for unfolding: \textit{PowhegV2+Pythia8}

- **Left**: dominated by stat. uncertainties → hard to make any conclusions on observed trends
Results: $p_T(top), |y(top)|$

Left: $L = 42 \text{ pb}^{-1}$

Right: $L = 2.3 \text{ fb}^{-1}$

- Full phase space, parton level
- Reference $\bar{t}t$ prediction used for unfolding: PowhegV2+Pythia8

- Left: dominated by stat. uncertainties → hard to make any conclusions on observed trends
- Right: softer data in $p_T(top)$ → consistent with Run-I
- Reasonable data-vs-MC agreement for $|y(top)|$
Results: $p_T(\bar{t}t), |y(\bar{t}t)|, m(\bar{t}t)$

- Full phase space, parton level
- Reference $t\bar{t}$ prediction used for unfolding: PowhegV2+Pythia8
- Softer data in $p_T$ (top) → consistent with Run-I
- Reasonable agreement between data and predictions for $|y(\bar{t}t)|$
- Harder $p_T(\bar{t}t)$ in MadGraph and aMC@NLO predictions

Data: $L = 2.3 \text{ fb}^{-1}$
**Summary**

---

### Top quark pair differential cross section measurements:

- Essential for constraining the SM
- Ideal probe for looking for new physics beyond the SM

---

### Results using $L = 2.3$ fb$^{-1}$ of 13 TeV data recorded by CMS in 2015:

- Measurement dominated by systematical uncertainty: 5-20% precision
  $\rightarrow$ already a precision test of pQCD
- In general, data described reasonably well by all MC predictions
- $p_T^{\text{top}}$: looks like slope observed at 8 TeV is similar to one appeared at 13 TeV
### Summary

**Top quark pair differential cross section measurements:**

- Essential for constraining the SM
- Ideal probe for looking for new physics beyond the SM

**Results using $L = 2.3\; fb^{-1}$ of 13 TeV data recorded by CMS in 2015:**

- Measurement dominated by systematical uncertainty: 5-20% precision
  → already a precision test of pQCD
- In general, data described reasonably well by all MC predictions
- $p_T\,(\text{top})$: looks like slope observed at 8 TeV is similar to one appeared at 13 TeV

---

**Thank you for your attention!**
Backup
Binneing and migrations

- **Migration effects** studied by:

  \[ \pi_i = \frac{N_i^{\text{rec \& gen}}}{N_i^{\text{rec}}} \]  - **purity**: sensitive to migrations to \( i \)-th bin

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

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

- **Binning criteria:**
  - stability or purity \( \geq \sim 40\text{-}50\% \)
  - \( \approx \) flat in all bins
  - diagonal response matrix

- Example for measurement in bins of \( p_T^{\text{top}} \)
Unfolding techniques correct migrations between bins

Response matrix (A): represents bin-by-bin correlations

Unfolding problem is transformed to $\chi^2$ - minimization problem:

\[ \chi^2 = \left( \vec{N} - A \cdot \vec{x} \right)^T \text{COV}_{\vec{N}}^{-1} \left( \vec{N} - A \cdot \vec{x} \right) - \tau^2 \cdot K \left( \vec{x} \right) \]

- $\boldsymbol{N}$: BG corrected data
- $\boldsymbol{x}$: unfolded result

Non-physical fluctuations removed by means of the regularization:

> $\tau$ – continuous regularization parameter

> selected at minimum of average global correlation

- Signal $\bar{t}t$ reference sample used for unfolding: PowhegV2+Pythia8
Systematic uncertainties

Normalization: mostly shape uncertainties contribute

Each uncertainty propagated through analysis chain individually
- For each source, the corresponding efficiency, resolution or scale is changed by its uncertainty or similar
- Systematic unc. per bin: difference of the changed result wrt nominal value

Experimental uncertainties
- Trigger efficiency, Lepton ID/Iso, JES, JER, b-tagging, Pile-Up reweighting, Background Cross Sections (30% variations for all samples)

Signal model uncertainties
- $Q^2$ scale, Top mass, Hadronization model, Generator model, PDF
- Measurement dominated by statistical uncertainty in all bins of each observable

- **Hadronization**: PowhegV2+Pythia8 vs PowhegV2+Herwig++

- **Generator**: PowhegV2+Pythia8 vs aMC@NLO(FxFx)+Pythia8

- **Typical dominant uncertainties**: medians of the distribution of uncertainties over all bins for rapidity (all other) observables

| Source          | Uncertainty (%) |
|-----------------|-----------------|
| Generator       | 3.4 (1.6)       |
| Hadronization   | 2.3 (2.9)       |
| PDF             | 1.5 (0.5)       |
| JES             | 1.2 (1.2)       |
| JER             | 0.7 (0.8)       |
| b-tagging       | 0.6 (0.9)       |
Preliminary results: overview of uncertainties

Measurements mostly dominated by signal model uncertainties

for $L = 2.26 \text{ fb}^{-1}$
(13 TeV, 25ns)