Optimization of top quark pair kinematic reconstruction using CMS Open Data

Spitsyn Mykyta

Taras Shevchenko National University of Kyiv

Supervisors: Oleksandr Zenaiev

Introduction

- This project is about improving experimental measurements of top quark production
- It is not directly related to xFitter, but hopefully in the future improved measurements, obtained with methods described in this project, will be used in xFitter

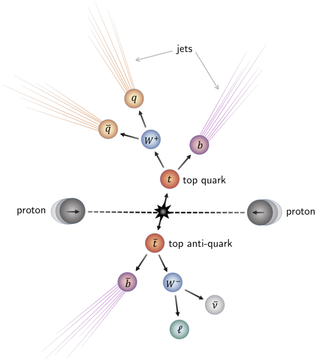
Tasks

- Explore $t\bar{t}$ reconstruction methods in the dileptonic decay channel using CMS Open Data
- Evaluate the efficiency, resolution, and parameter bias for each method and compare results with CMS
- Determine the differential and total $t\bar{t}$ production cross sections using each reconstruction method, compare results and assess their accuracy

Decay channels of the $t\bar{t}$ system

$t\bar{t}$ decays via three channels:

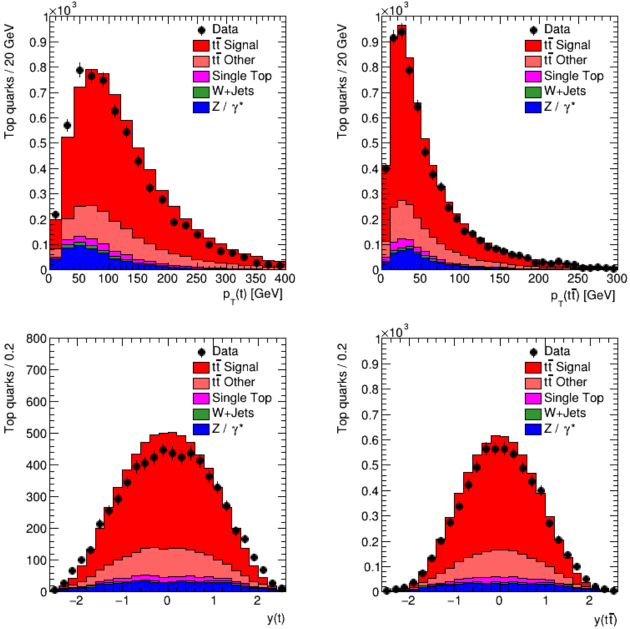
- Dileptonic channel (10.5%): $t\bar{t} \to W^+bW^-\bar{b} \to \ell^+\nu b\ell^-\bar{\nu}\bar{b}$
- Lepton+jet channel (43.8%): $t\bar{t} \to W^+bW^-\bar{b} \to q\bar{q}'b\ell^-\bar{\nu}\bar{b}$ or $\ell^+\nu bq\bar{q}'\bar{b}$
- All-jets channel (45.7%): $t\bar{t} \to W^+bW^-\bar{b} \to q\bar{q}'bq\bar{q}\bar{b}$



Data used

This study uses CMS Open Data. The datasets include both Monte Carlo generated samples and real CMS detector data.

1. Experimental data:
   - Dileptonic decays
     - $\mu^{\pm}\mu^{\mp}$
     - $e^{\pm}e^{\mp}$
     - $e^{\pm}\mu^{\mp}$
2. Monte Carlo data
   - Signal (MadGraph + Pythia6)
   - Background:
     - $t\bar{t}$ (decay via τ leptons) (MadGraph + Pythia6)
     - t (POWHEG + Pythia6)
     - Drell–Yan process ($Z \to \mu^{+}\mu^{-}, Z \to e^{+}e^{-}$) (MadGraph + Pythia6)
     - W^{+} bosons and jets


Top quarks / 20 GeV
$\times 10^3$
Data
$t\bar{t}$ Signal
$t\bar{t}$ Other
Single Top
W+Jets
Z / γ^*
$p_T(t)$ [GeV]
$p_T(t\bar{t})$ [GeV]
Top quarks / 0.2
y(t)
$y(t\bar{t})$

Analytical reconstruction methods

- FKR – Full Kinematic Reconstruction. Reconstructs the four-momenta of t and $\bar{t}$ using energy–momentum conservation and known intermediate particle masses. The system of equations is solved 100 times with detector resolution smearing.
- SKR – Simple Kinematic Reconstruction. A simplified variant of FKR without detector resolution smearing; the kinematic equations are solved only once.
- LKR – Loose Kinematic Reconstruction. Does not reconstruct ν and $\bar{\nu}$ separately but treats them as a $\nu\bar{\nu}$ system, increasing efficiency but preventing reconstruction of t and $\bar{t}$ separately, only the $t\bar{t}$ system.

Full kinematic reconstruction

Goal: obtain $\vec{p}_t$, $\vec{p}_{\bar{t}}$

- Measured input: 2 leptons, 2 jets, $p_{\mathrm{T}}^{\mathrm{miss}}$
- Unknowns: $\overline{p}_\nu$, $\overline{p}_{\overline{\nu}}$,
- Constraints:
  - $m_t, m_{\bar{t}}$
  - m_{W^+}, m_{W^-}
  - $p_{\mathrm{T}}^{\mathrm{miss}} = p_{\nu,T} + p_{\overline{\nu},T}$
- For each pair of jets equations of the form $m_t^2 = E^2 - p^2$ are used
- If there are several combinations, b-tagged jets and solutions with lowest $m_{t\bar{t}}$ are prefered and the 2nd order equations become 4th order

Loose kinematic reconstruction

The kinematic variables of the $\nu\overline{\nu}$ system are derived as follows:

1. The transverse momentum $\vec{p}_T$ of the $\nu\overline{\nu}$ system is set equal to $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$
2. The $\nu\overline{\nu}$ longitudianl momentum $p_{z,\nu\overline{\nu}}$ is set to that of the lepton pair $p_{z,\nu\overline{\nu}} = p_{z,\ell\overline{\ell}}$
3. The energy $\nu\overline{\nu}$ of the $E_{\nu\overline{\nu}}$ system is defined as $E_{\nu\overline{\nu}} = E_{\ell\overline{\ell}}$ for $p_{\nu\overline{\nu}} < E_{\ell\overline{\ell}}$ and $E_{\nu\overline{\nu}} = p_{\ell\overline{\ell}}$ for $p_{\nu\overline{\nu}} > E_{\ell\overline{\ell}}$ (to ensure that $m_{\nu\overline{\nu}} \geq 0$)
4. The four momentum $\ell\overline{\ell}\nu\overline{\nu}$ is calculated
5. For $m_{\ell\overline{\ell}\nu\overline{\nu}} < 2m_W = 2 \times 80.4$ GeV, the mass component of the four-momentum of $\ell\overline{\ell}\nu\overline{\nu}$ is set to $2m_W$, ensuring that $m_{W^+W^-} \geq 2m_W$
6. The four momentum of the $t\overline{t}$ system is calculated by using the four momentum of the $\ell\overline{\ell}\nu\overline{\nu} + b\overline{b}$

Evaluation of reconstruction methods

One ot the most important characteristics of the kinematic reconstruction is efficiency – the ratio of reconstructed to Monte Carlo generated events. Additionally, resolution and parameter bias are evaluated as functions of invariant mass, transverse momentum, rapidity, and azimuthal angle of the $t\bar{t}$ system. Average efficiencies:

Reconstruction method	Efficiency
FKR	82%
SKR	67%
LKR	96%

Efficiency, resolution, and bias

Reconstruction efficiency E and its uncertainty ΔE are defined as:

$$E = \frac{N_{\text{reco}}}{N_{\text{MCgen}}} \qquad \Delta E = \sqrt{\frac{E(1-E)}{N_{\text{MCgen}}}}$$

N_{reco}, N_{MCgen} – number of reconstructed and generated events

Bias and its uncertainty:

$$\bar{r} = \frac{1}{n}\sum_{i}^{n} r_i \qquad \Delta r = \frac{\sigma}{\sqrt{n}}$$

where σ – resolution,
$r_i = x_{\text{rec, i}} - x_{\text{gen, i}}$, i – event index, x – parameter value (y, p_T, M, φ)
x_{gen} – true (MC generated) value, x_{rec} – reconstructed value,
n – number of reconstructed events in a bin.

Reconstruction resolution:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i}^{n}\left(r_i - \overline{r}\right)^2}$$

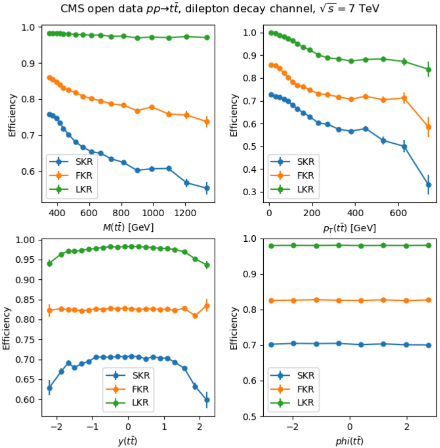
Resolution uncertainty:

$$\Delta\sigma = \sqrt{\frac{\mu_4 - \sigma^4}{n}}$$

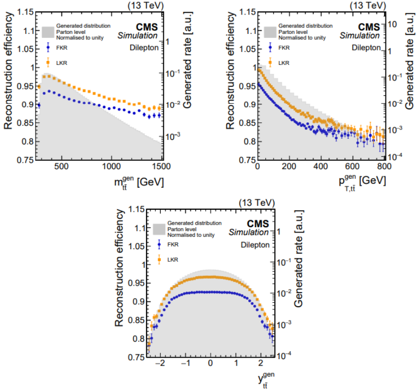
μ_4 – fourth central moment:

$$\mu_4 = \frac{1}{n}\sum_{i=1}^{n}\left(r_i - \overline{r}\right)^4$$

Efficiency

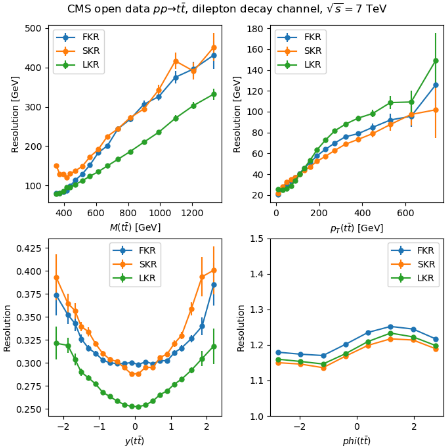


CMS open data pp→tt̄, dilepton decay channel, √s = 7 TeV
Efficiency
SKR
FKR
LKR
M(tt̄) [GeV]
pT(tt̄) [GeV]
y(tt̄)
phi(tt̄)

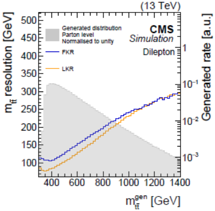


(13 TeV)
CMS
Simulation
Dilepton
Generated distribution
Parton level
Normalised to unity
FKR
LKR
Reconstruction efficiency
Generated rate [a.u.]
m tt̄ gen [GeV]
p T,tt̄ gen [GeV]
y tt̄ gen

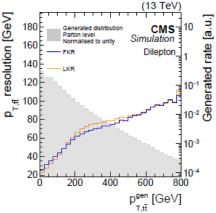
Resolution



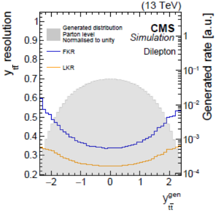
CMS open data $pp \to t\bar{t}$, dilepton decay channel, $\sqrt{s} = 7$ TeV
FKR
SKR
LKR
Resolution [GeV]
$M(t\bar{t})$ [GeV]
$p_T(t\bar{t})$ [GeV]
Resolution
$y(t\bar{t})$
$phi(t\bar{t})$



(13 TeV)
CMS
Simulation
Dilepton
Generated distribution
Parton level
Normalised to unity
FKR
LKR
$m_{t\bar{t}}$ resolution [GeV]
Generated rate [a.u.]
$m_{t\bar{t}}^{gen}$ [GeV]

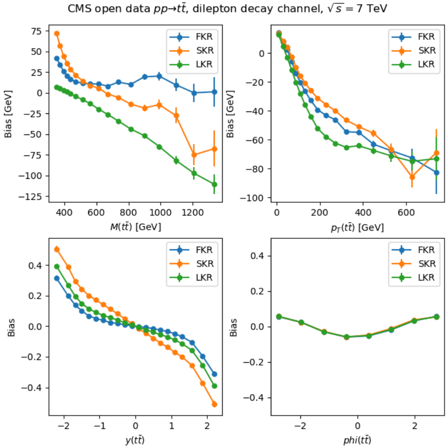


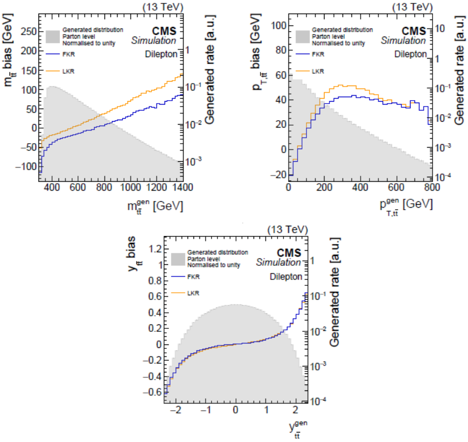
(13 TeV)
CMS
Simulation
Dilepton
Generated distribution
Parton level
Normalised to unity
FKR
LKR
$p_{T,t\bar{t}}$ resolution [GeV]
Generated rate [a.u.]
$p_{T,t\bar{t}}^{gen}$ [GeV]



(13 TeV)
CMS
Simulation
Dilepton
Generated distribution
Parton level
Normalised to unity
FKR
LKR
$y_{t\bar{t}}$ resolution
Generated rate [a.u.]
$y_{t\bar{t}}^{gen}$

Bias


CMS open data $pp \to t\bar{t}$, dilepton decay channel, $\sqrt{s} = 7$ TeV
FKR
SKR
LKR
Bias [GeV]
$M(t\bar{t})$ [GeV]
$p_T(t\bar{t})$ [GeV]
Bias
$y(t\bar{t})$
$phi(t\bar{t})$


(13 TeV)
CMS
Simulation
Dilepton
Generated distribution
Parton level
Normalised to unity
FKR
LKR
Generated rate [a.u.]
$m_{t\bar{t}}$ bias [GeV]
$m^{gen}_{t\bar{t}}$ [GeV]
$p_{T,t\bar{t}}$ bias [GeV]
$p^{gen}_{T,t\bar{t}}$ [GeV]
$y_{t\bar{t}}$ bias
$y^{gen}_{t\bar{t}}$

Total and differential $t\bar{t}$ production cross sections

Differential cross section:

$$\frac{\mathrm{d}\sigma}{dy} = \frac{N_{\mathrm{sig}}}{ELB\Delta Y} \quad \text{where } N_{\mathrm{sig}} = N_{\mathrm{DATA}} - N_{\mathrm{MC\ background}},$$

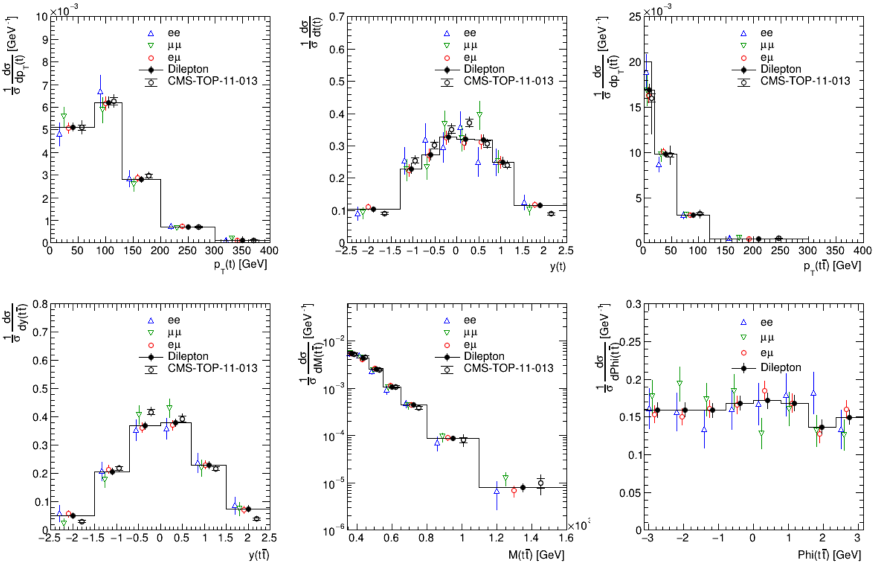
E = $\frac{N_{\mathrm{MCreco}}}{N_{\mathrm{MCgen}}}$ – efficiency

L = 2.5 fb^{-1} – luminosity, B = 4.6% – branching ratio

Total cross section:

$$\sigma = \int \frac{\mathrm{d}\sigma}{dy}$$

Differential cross sections



Total cross sections

CMS result at $\sqrt{s} = 7$ TeV:

$\sigma_{t\bar{t}} = 173.6 \pm 2.1(\text{stat})^{+4.5}_{-4.0}(\text{syst}) \pm 3.8(\text{lumi})$

FKR

Parameter	Total cross section $\sigma \pm \Delta\sigma$, pb	$\varepsilon, \%$
$\vec{p}_T(t\bar{t})$	176.2 ± 4.1	1.5
$y(t\bar{t})$	182.4 ± 5.1	5.0
$M(t\bar{t})$	169.8 ± 4.0	2.2
$\varphi(t\bar{t})$	160.5 ± 4.2	8.2

LKR

Parameter	Total cross section $\sigma \pm \Delta\sigma$, pb	$\varepsilon, \%$
$\vec{p}_T(t\bar{t})$	175.2 ± 4.1	1
$y(t\bar{t})$	184.5 ± 5.1	6
$M(t\bar{t})$	173.1 ± 4.0	0.3
$\varphi(t\bar{t})$	160.5 ± 4.2	7.5

SKR

Parameter	Total cross section $\sigma \pm \Delta\sigma$, pb	$\varepsilon, \%$
$\vec{p}_T(t\bar{t})$	171.6 ± 4.0	1.5
$y(t\bar{t})$	166.8 ± 5.2	1.1
$M(t\bar{t})$	158.7 ± 4.4	8.5
$\varphi(t\bar{t})$	158.7 ± 4.2	8.5

Conclusion

- Three analytical reconstruction methods of the $t\bar{t}$ system were studied using CMS Open Data
- Efficiency, resolution, and parameter bias were determined for each method; results agree with CMS
- For the first time, these characteristics were studied as a function of azimuthal angle, showing detector resolution asymmetry
- LKR is optimal for reconstructing the $t\bar{t}$ kinematics, while FKR is optimal for reconstructing t and $\bar{t}$ separately
- Total and differential $t\bar{t}$ production cross sections were obtained relative to various kinematic variables
- Results are consistent with CMS measurements within statistical uncertainties
- Demonstrated that CMS Open Data allow performing full-scale physics research.

Next steps

- Improvement of Loose kinematic reconstruction: better constraints and improved analysis
- Addition of new variable $\Delta\varphi_{t\bar{t}}$
- Development of new reconstruction methods

Thank you for your attention!