STUDENT JOURNAL OF PHYSICS

Determination of cross section for top quark pair production at LHC and study of jets from top quark decays

Sarvesh Bansal[1]*and **Kajari Mazumdar**[2]
[1]M.Sc Physics 2nd year, Physics Department, IIT Delhi
[2]Department of High Energy Physics, TIFR, Mumbai

Abstract. Currently high energy physicists are trying to find particles beyond the standard model. These new energetic particles will have top quark as their major background. So to understand these new particles we need to understand the kinematics of top quark in a better way. In this project we have estimated the production cross section of top quark pair in LHC proton-proton collision at centre of mass energy of 7 TeV using processed data made available in open platform by CMS collaboration at LHC, CERN. Also in this project we have studied the situation of boosted top quarks when LHC machine will collide protons at an energy of 14 TeV. We have studied the jets from these boosted top quark by generating samples using Pythia Monte Carlo event generator .We used N-subjettiness parameter to find some characteristics of these jets in order to distinguish top signal from mainly QCD multijets events.

Keywords: Top Quark Decay, CMS Collaboration, LHC.

1. INTRODUCTION

In the first part of this project we have estimated the production cross section of top quark [1-4] pair in LHC proton-proton collision at centre-of-mass energy of 7 TeV using processed data made available in open platform by CMS collaboration at LHC, CERN.
(http://opendata.cern.ch/collection/CMS-Derived-Datasets).

This data corresponds to an integrated luminosity of 50 pb^{-1} and each event has at least one muon in the final state as reconstructed by the CMS detector. When a top quark decays (t →bW) it gives 1 or 3 jets depending on semil-leptonic or fully hadronic mode. Thus top quark pair events in the data sample must additionally have jets, with at least two jets due to b-quarks (from top and anti-top) and missing transverse energy due to neutrino. Using Monte Carlo event samples for signal and various background processes we have studied the kinematics of final state objects and applied various selection criteria to enhance the signal-to-background ratio of the final sample. The purity of the selected sample and the efficiency of selection are estimated from Monte Carlo. The cross section is estimated using experimental data.

---

*bansalsarvesh.s@gmail.com

In the second part of the project we have studied the situation of boosted top quarks when LHC machine will collide protons at an energy of 14 TeV. At lower energy collisions the jets from a top quark hadronic decay are distinguishable since they are separated. But at high energy collisions when the top quark is itself boosted, jets from a top decay are not separable always within the detector and may be detected as a fat jet. We have studied these boosted top quark jets by generating samples using Pythia Monte Carlo event generator and used Fastjet method for jet reconstruction. We looked into the jet substructures using N-subjettiness to find some characteristics of these jets in order to distinguish top signal from mainly QCD multijets events.

2. PRODUCTION AND DECAY OF TOP QUARK

Top quark can be produced in high energy colliders by two processes.

- Electroweak interaction
- Strong interaction

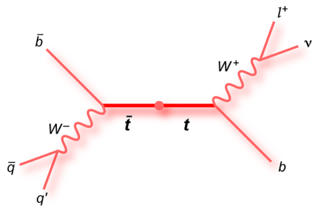
In electroweak interaction only single top quark is produced with other particles and jets, but in strong interaction top quarks are produced in pair and with significantly large rate. Top quark is a heavy and unstable particle so it decays very fast. It decays mostly(branching ratio 99 %)by single mode.

$$\mathrm{t} \rightarrow \mathrm{W}^{+}\mathrm{b}$$

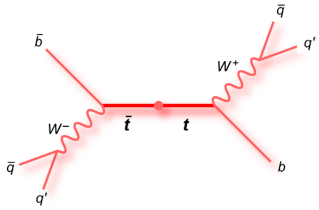
Now W can decay into leptons $\mathrm{W} \rightarrow \ell\nu$ where $\ell = e, \mu, \tau$ or hadrons $\mathrm{W} \rightarrow \mathrm{q}_1\bar{\mathrm{q}}_2$ where $\mathrm{q}_1, \mathrm{q}_2 = \mathrm{u}, \mathrm{d}, \mathrm{c}, \mathrm{s}, \mathrm{b}$. Thus we can have three types of final states when a $\mathrm{t}\bar{\mathrm{t}}$ is produced.Figure 1 displays the cartoons of these (Br is an abbreviation used for branching ratio.).

- Semi-leptonic; $\mathrm{t} \rightarrow \mathrm{b}\ell^{+}\nu$ and $\bar{\mathrm{t}} \rightarrow \bar{\mathrm{b}}\mathrm{q}_1\bar{\mathrm{q}}_2$ and visa versa (Br $\approx 36\%$)
- Fully Hadronic; $\mathrm{t} \rightarrow \mathrm{b}\mathrm{q}_1\bar{\mathrm{q}}_2$ and $\bar{\mathrm{t}} \rightarrow \bar{\mathrm{b}}\mathrm{q}_3\bar{\mathrm{q}}_4$ (Br $\approx 45\%$)
- Dileptonic; $\mathrm{t} \rightarrow \mathrm{b}\ell^{+}\nu$ and $\bar{\mathrm{t}} \rightarrow \bar{\mathrm{b}}\ell^{-}\bar{\nu}$ (Br $\approx 7\%$)

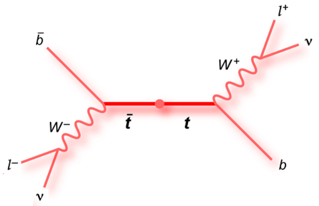
We have done the analysis of semi-leptonic decay. In semi leptonic decay of $t\bar{t}$ event, final state involves a charged lepton, a corresponding neutrino and at least four jets out of which there are two b-tag jets. Semi-leptonic decay can be divided into three modes with three different leptons μ, e, τ.



Semileptonic



Fully hadronic



Dileptonic

Figure 1. Figure shows the final states of top quark pair. Top quark decays via $\mathrm{t} \rightarrow \mathrm{W}^{+}\mathrm{b}$. Further W can decay into leptons $\mathrm{W} \rightarrow \ell\nu$ where $\ell = e, \mu, \tau$ or hadrons $\mathrm{W} \rightarrow \mathrm{q}_1\bar{\mathrm{q}}_2$ where $\mathrm{q}_1, \mathrm{q}_2 = \mathrm{u}, \mathrm{d}, \mathrm{c}, \mathrm{s}, \mathrm{b}$. Antitop quark also decays similarly. So t and $\bar{\mathrm{t}}$ collectively gives three final states a) Semileptonic b) Fully Hadronic c) dileptonic. Details are given in the figure above.

3. ESTIMATION OF CROSS SECTION

3.1 *Background*

There are many other processes which give the same final state as that of $\mathrm{t}\bar{\mathrm{t}}$ event with semileptonic final state. These are considered as backgrounds. These processes are

- **W + jets, Z + jets** : W/Z + jets are produces in hadron-hadron collisions. A quark from one hadron and anti-quark from another hadron get annihilated and produce W/Z. Before annihilation these quarks emit gluons which on hadronisation give us jets.

- **Di-boson**
    1. **WW**: These are produced in following way.
       a) $q\bar{q} \rightarrow Z \rightarrow W^+W^-$ and b) $gg \rightarrow Z \rightarrow W^+W^-$.
    2. **ZZ**: these are also produced in similar way to WW just by replacing WW by ZZ in the final state.
       a) $q\bar{q} \rightarrow Z \rightarrow ZZ$ and b) $gg \rightarrow Z \rightarrow ZZ$.
    3. **WZ**: This can be produced by interaction of quark and antiquark of different kinds which produce a W and W further produce a WZ pair. $q_1\bar{q}_2 \rightarrow W \rightarrow WZ$
- **Single top**: Single top quarks are produced via weak interaction. Single top quarks are produced by three processes.
    1. t-channel: This involves the exchange of a space-like W boson. This process is also called W-gluon fusion, because the b-quark arises from a gluon splitting to bb.
    2. s-channel: This involves the production of a time-like W boson, which then decays to a top and a bottom quark.
    3. tW-channel: involves the production of a real W boson. In this process b-quark absorbs a gluon and then decayes to top quark and real W boson.
- **QCD multijets events**: Quarks and gluons emit many gluons while moving at high energies. These gluons get hadronised and form jets. These jets mimic as they are from the decay of a top quark and hence give a large back ground.

Different process mentioned above have W or Z or jets. W/Z decays to give quarks and leptons(Decay of W boson is discussed in section 2). So in the final states we get leptons and jets which mimic the decay products of top quark. To find the cross section and to improve the signal to background ratio of the selected sample we need to reduce the backgrounds by applying various selection on the data collected in experiments. First we study Monte Carlo samples of various processes and try to analyze the final states of each channel and on the basis of the outcome we apply selection criteria on the data.

Table 1. Total no. of events and event left after the triggering(Cut1) are given in column 2 and column 3 respectively. Column 4 and following columns give the efficiency of different cuts w.r.t to column 3 for various process(0.000 means efficiency of that cut is zero upto three decimals and 0 means no event is left after that cut).

Process	Total events	Cut1	Cut2	Cut3	Cut4	Cut5	Cut6	Cut7
$t\bar{t}$	36941	4515	0.862	0.829	0.784	0.131	0.039	0.036
Z + jets	77729	77729	0.935	0.850	0.412	0.000	0.000	0
Single top	5684	5684	0.896	0.857	0.838	0.020	0.004	0.003
W+Jets	109737	109737	0.868	0.811	0.811	0.000	0	0
WW	4580	4580	0.888	0.846	0.797	0.001	0	0
WZ	3367	3367	0.917	0.861	0.691	0.001	0	0
ZZ	2421	2421	0.951	0.897	0.417	0.000	0	0
QCD	142	142	0.105	0.077	0.077	0.014	0.007	0.007
Data	469384	469384	0.477	0.434	0.405	0.001	0.000	0.000

3.2 *Selection criteria and analysis*

1. Cut1- Only Muon Triggered

   We take only muon triggered events as our data from CMS. Muon triggering means an event will be stored only if the detector detects a muon in the final state. If there is no muon generation then event will be rejected.

2. Cut2- Isolation Criteria < 0.10

   A lepton produced in the decay of W (or Z) is not surrounded by other energetic particles which is opposite to the situation where in QCD process a lepton from the b quark decay (b $\rightarrow c\ell\nu$) will be surrounded by other particles. In the former case the lepton is isolated. For isolation we choose energy deposited by all particles within the 0.4 cone radius around muon with less than 10 % of p_t of muon.

3. Cut3- Leading Muon $>$ 26 Gev

   In case of leptonic decay mode of W, it is a two body decay (lepton and neutrino). Neutrino and muon are massless compared to W, so all the energy of W (rest mass energy and kinetic energy if any) go in as the momentum of muon and neutrino. As the mass of W is around 80 GeV, the maximum momentum of muon is 40 GeV approx. By this selection we reduce the background due to less energetic muons coming from mainly QCD multijet events.

4. Cut4- No second lepton in the event

   In the semileptonic decay of $\mathrm{t\bar{t}}$ there is only one lepton so we need one lepton in the final state. But in Z+jets , ZZ there are more than one lepton. So to reduce background due to these events we choose this criteria.

5. Cut5- Atleast 4 jets with $p_t > 30$ GeV
   In the semileptonic decay channel of $t\bar{t}$ there are two b-tag jets and two quark jets from W, in total 4 jets. But in Z+jets, ZZ there are not many events with 4 jets.

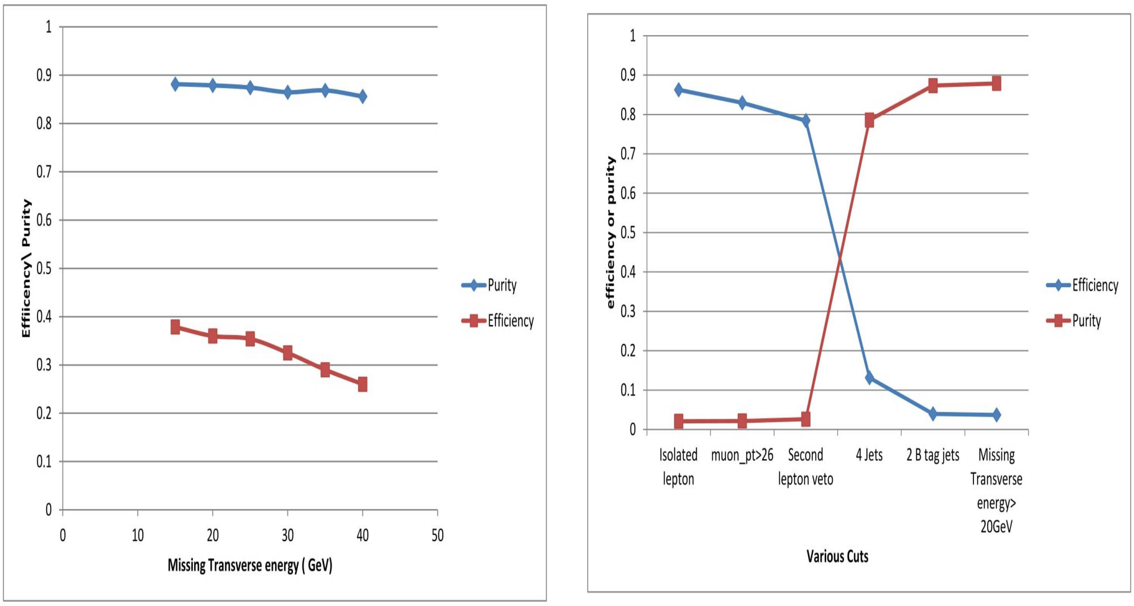


Figure 2. Left plot shows optimization of missing transverse energy and right plot shows how efficiency and purity is varying with various cuts applied. Efficiency is multiplied with the factor of 10 in the left plot.

6. Cut6- Exactly 2 b-tag jets
   We get b and $\bar{b}$ jets from t and $\bar{t}$ decay respectively. Now why exactly 2 b-tag jets? the answer is because we can get b- tag jet in single top event but it will give only one b tag jet. So exactly 2 b-tag jet reduce single top event.

7. Cut7- Missing transverse energy > 20 GeV
   True missing energy is due to the undetected neutrino. But due to detector resolution there may be non zero missing energy in the reconstructed event which do not contain any ν. Since we can't measure the longitudinal component of energy-momentum in the initial state of the hard scatter process, so we deal only with the transverse component. Missing energy is defined as negative of vector sum of all p_t. We choose only 20 GeV here because as missing energy threshold is increased efficiency is decreased but purity remains the same with some small fluctuations. So it is needed to optimize the efficiency and purity. Graph in Figure 2 shows

the variation of efficiency and purity with change in missing energy. Efficiency after various selection criteria is given in table 1.

3.3 *Determination of cross section*

$$\mathrm{crosssection}(\sigma) = \frac{\mathrm{N_{signal}} \times \mathrm{Purity}}{\mathrm{L} \times \epsilon \times \mathrm{Br}}$$

Where the branching ratio (Br) of $\mathrm{t} \rightarrow \mu\nu\mathrm{b}$ = 13.4% , Br of $\bar{\mathrm{t}} \rightarrow \mathrm{q\bar{q}\bar{b}}$ = 66.5%
Br of $\mathrm{t\bar{t}} \rightarrow \mathrm{semileptonicmode} = 2\times$ Br of $\mathrm{t} \rightarrow \mu\nu\mathrm{b}\times$Br of $\bar{\mathrm{t}} \rightarrow \mathrm{q\bar{q}\bar{b}}$ = 17.82 %
$\mathrm{Purity} = \frac{\mathrm{N_{signal}}}{\mathrm{N_{total}}} = 0.878$ Integrated Luminosity (L) = 50 pb^{-1}
$\mathrm{N_{signal}} = \mathrm{N_{data}} - \mathrm{N_{Background}} = 43 - 1.155 = 41.845$

$$\mathrm{Efficiency}(\epsilon) = \frac{\text{total no. of signal events passing all cuts}}{\text{total no. of signal event after trigger}} = 0.036$$

$$\sigma = \frac{41.845 \times 0.878}{50 \times 0.036 \times 0.178} = 114.540\,\mathrm{pb}$$

Estimation of statistical error:

$$(\frac{\Delta\sigma}{\sigma})^2 = (\frac{\Delta \mathrm{N}}{\mathrm{N}})^2 + (\frac{\Delta \mathrm{P}}{\mathrm{P}})^2 + (\frac{\Delta \mathrm{L}}{\mathrm{P}})^2 + (\frac{\Delta\epsilon}{\epsilon})^2 + (\frac{\Delta \mathrm{Br}}{\mathrm{Br}})^2$$

$\Delta\sigma = 27.260$ pb $\sigma = 114.540 \pm 27.260$ pb

4. STUDY OF JETS FROM THE DECAY OF TOP QUARK

4.1 *Hadronic jets in the experiments*

Jets are the collimated bunch of particles (mainly hadrons) coming from the hadroniztion of quarks and gluons. It is mentioned earlier that top decays to W^+b and W^+ further decays to $q\bar{q}$ or $\ell\nu$. If decay mode is leptonic we can detect lepton using tracking detector and calorimeters. But when W^+ decays to quark we can't detect it directly because quarks are unstable and they get hadronised to give many hadrons which we can't resolve in detectors. We get these hadrons as a collimated bunch of particles and this is referred to as jets. From the study of kinematics of jets we can associate a jet from top quark decay.

As there are two quarks in decay of W^+(hadronic mode), we get two jets in the final state and in combination of b quark we get three jets from top quark. As aforementioned, $\mathrm{t\bar{t}}$ event has three decay modes so accordingly we get 2,4 and 6 jets for dileptonic, semileptonic and fully hadronic channels respectively.

In case of fully hadronic decay of $t\bar{t}$ we get 6 jets but this is true when the kinetic energy of top quark is low $i.e$ top quark is not boosted. At lower energy we are able to resolve the three jets on the detector. But when energy of top quark is very high, jets produced from decay of top are also largely boosted. Hence these jets are very close to each other such that we are not able to resolve these jets and three jets appear as one fat jet. So at high energy when jets are close to each other it is needed to study the substructure of jet to associate a jet with the parent top. Figure 3 displays the cartoon of boosted jet.

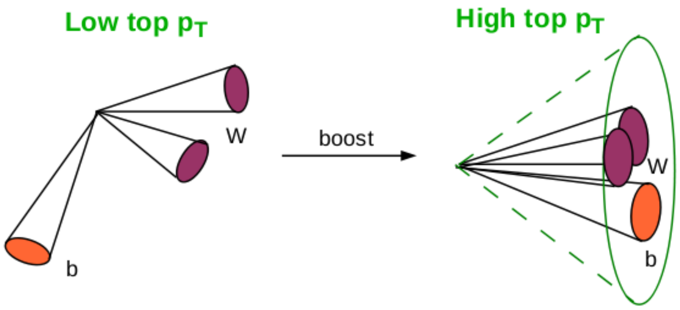


Figure 3. Conversion of three jets into a fatjet at high $\mathrm{p_t}$

4.2 *N-Subjettiness*

N-subjettiness is a variable designed to identify boosted hadronic decaying object like top quark. It is an effective discriminating variable for tagging boosted objects and rejecting the background of QCD jets with large invariant mass. It allow us to break the jets into subjets creating some imaginary axis. N-subjettiness is denoted by τ_{N}. Where 'N' stands for the number of subjets in a jet[1].

$$\tau_{\mathrm{N}} = \frac{1}{\mathrm{d_o}} \times \sum_{\mathrm{k}} \mathrm{p_{T,k}} \times \min[\Delta\mathrm{R_{1,k}}, \Delta\mathrm{R_{2,k}}, \Delta\mathrm{R_{3,k}}....., \Delta\mathrm{R_{N,k}}]$$

Here

$$\mathrm{d_o} = \sum_{\mathrm{k}} \mathrm{p_{T,k}R_o}$$

k runs over the constituent particles in a given jet, p_T,k are their transverse momenta and $R_{J,k}$ is the distance in the $\eta - \phi$ plane between a candidate subjet J and a constituent particle k.

Jet with $\tau_{\mathrm{N}} \approx 0$ have all their particles aligned with the subjet directions and therefore have N subjets. Jet with $\tau_{\mathrm{N}} \gg 0$ have a large fraction of their energy distributed away from the subjet directions and therefore have at least N + 1 subjets. So the value of τ_{N} tells to what degree our estimated number of subjets in a fatjet is right.

4.3 *Analysis for substructure of top quark jet*

For analyzing the conditions at high energy in proton proton collision at LHC we generated top quark pair event using Pythia event generator with centre of mass energy of 14 TeV. When top quark

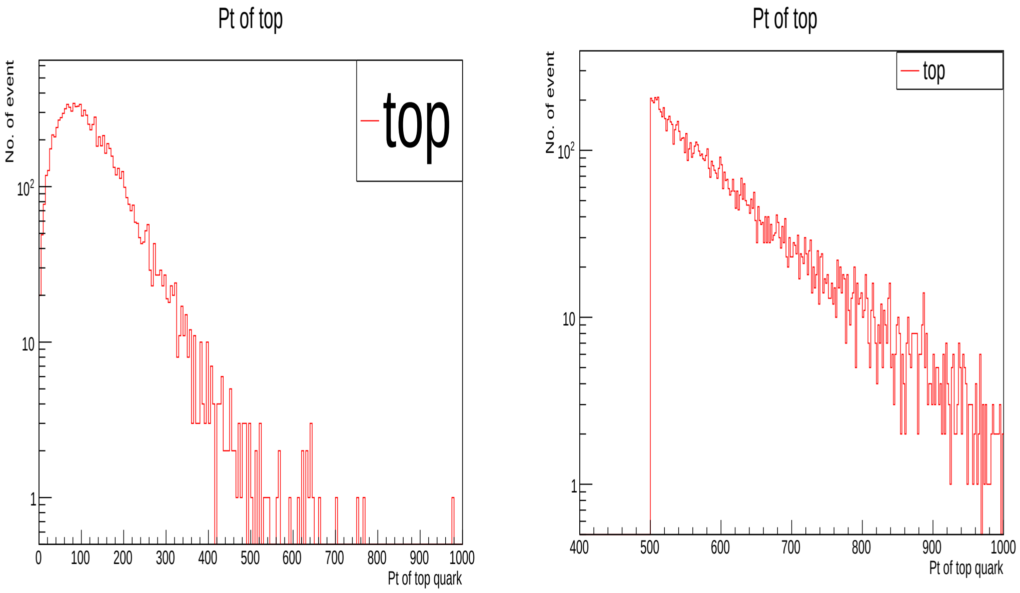


Figure 4. Above graph shows transverse momentum spectra of leading top quark in proton proton collision at 14 TeV. In the left graph it can be seen that number of events are decreasing very fast with increasing $\mathrm{p_t}$. There are much smaller number of events above 500 $\mathrm{p_t}$. Right graph shows the spectra of boosted top quark above 500 $\mathrm{p_t}$.

is generated it shows a distribution of p_t of top quark(Figure 4 left plot). As we are interested in highly boosted top quark we choose only high p_t top quark (eg. above 500 GeV). But as we apply this selection criteria, efficiency of sample become very low $\sim$ 0.29 % .

To get a fair number of events for analysis and to get a good distribution we need to generate a larger no. of events when the p_T of the top quark is above 500 GeV. But to save computing power instead of generating a large sample we have forced top quark to have p_t more than threshold(500 GeV)(Fig 4 right plot). We have also forced the top quark to decay in fully hadronic channel.

At generator level we have full information of final state particles and we reconstructed jets from particles in the final state using Cambridge/Aachen algorithm of Fastjet [5],[6],[7],[8]. For jet clustering we choose radius of cone for fat jet from top quark equal to 1.5. With this we get only 65% jets of the generated top events (Fig 5). After constructing jets we use a standard tool called ”top tagger” to find that if the jets are coming from top or from other particles. Now to find the substructure we use N-subjettiness method. For this we try to find no. of subjets by varing τ_N for $\mathrm{N} = 1, 2, 3$.

QCD multijet processes are the dominant source of boosted fat jets which can mimic hadronic decay of boosted top quark. So we follow the same procedure QCD events and results are shown in Figure 6 and Figure 7.

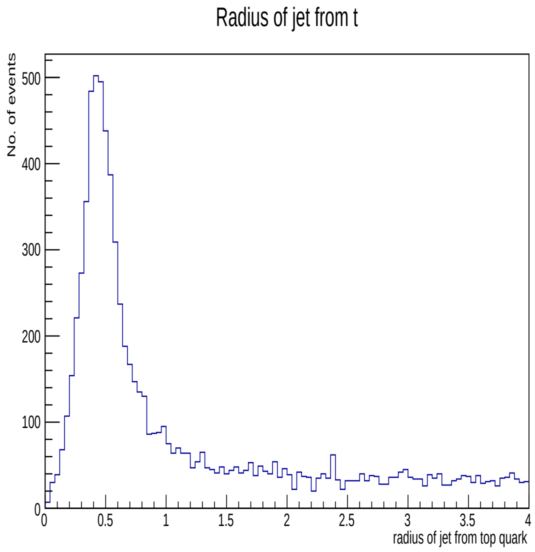


Figure 5. Radius of fat jet from top quark estimated by calculating the distance between b quark and W quark in $\eta - \phi$ space.

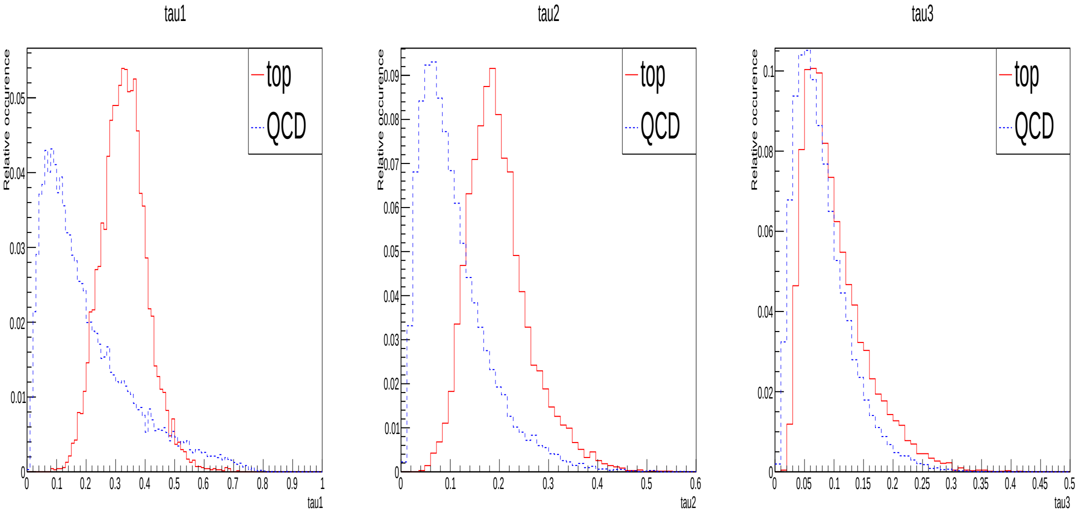


Figure 6: N-subjettiness of jet from top quark and QCD multijet events.

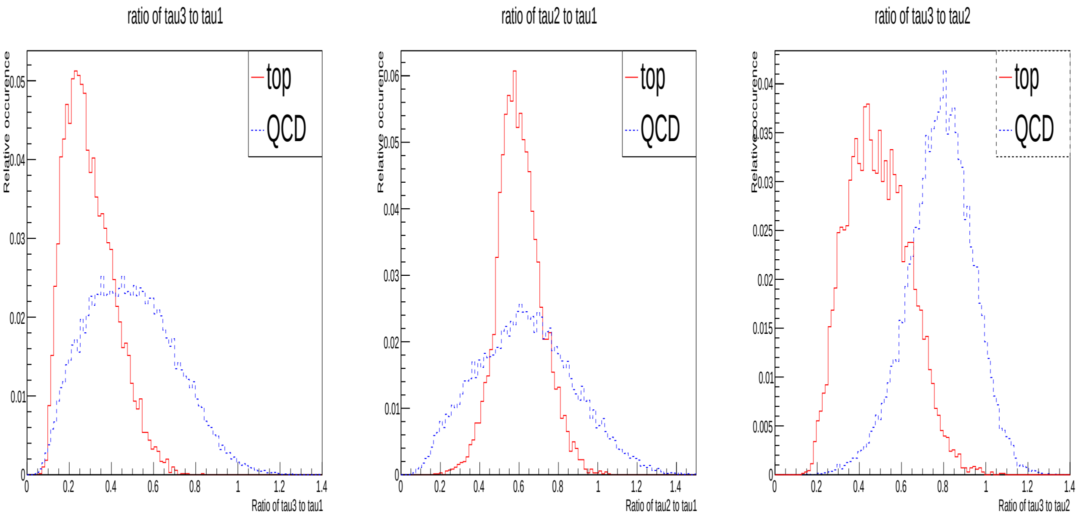


Figure 7: Ratios of N-subjettiness of jet from top quark and QCD multijet events.

4.4 *Observations*

- As the jets from top quark contain three subjets, so it give $\tau_1 \gg 0$(Fig 7 left plot). As we start increasing the number of subjets from one to three, peak of N-subjettiness starts shifting to lower value(Fig 7 middle and right plot). 3-subjettiness for three subjet is approximately equal to zero which confirms that boosted top jet has three subjets.

- In QCD events a boosted jet contains many subjets and these are distributed evenly in the fatjet and not like a top jet. So when we try to find N-subjettiness for QCD multijets it gives a peak near zero for τ_1, τ_2, τ_3. It can be seen from blue line in the graphs.

- When we calculate the ratio of these subjettiness it gives different results for top quark jet and for QCD multijets(Fig 7).

- By using the N-subjettiness variable and its ratios we can distinguish the QCD multijets and top quark events. We can reduce background by applying optimum cut on $\frac{\tau_3}{\tau_2}$.

5. ACKNOWLEDGEMENT

This project would not have been possible without the support of many people. Sarvesh Bansal would also like to thank his colleague Sarvesh Uplap for helping throughout the entire project. He would also like to thank Ph.D students Arvind, Pallabi and Bibhu Prasad for clearing many doubts and helping in project. He would also like to thank TIFR for providing such great opportunity.

6. REFERENCE

1. Jesse Thaler and Ken Van Tilburg,*Identifying boosted objects with N-subjettiness* [arXiv:1011.2268v3 [hep-ph] 25 Jan 2011]
2. T.M. Liss, F.Maltoni, and A. Quadt 2013,*The Top Quark*.
3. C. Clement Stockholm University, Fysikum, S-10691 Stockholm, Sweden, *Properties of the Top Quark*,SLAC Summer Institute on Particle Physics (SSI04), Aug. 2-13, 2004
4. Markus Cristin -ziani, Martijn Mulders,*Top-quark physics at the Large Hadron Collider*,Topical Review,[arXiv:1606.00327v2 [hep-ex] 4 Nov 2016]
5. Jet Algorithms,Vivians Meeting,Philipp Schieferdecker (KIT),April 17th 2009
6. Jet Reconstruction Journal Club 7/28/08
7. Stan Bentvelsen and Irmtraud Meyer,*The Cambridge jet algorithm: features and applications*,CERN-EP/98-043,March 11th, 1998,[arXiv:hep-ph/9803322v2 2 Apr 1998]
8. FastJet user manual, arXiv:1111.6097v1 [hep-ph] 25 Nov 2011