Study of physics algorithms to setup an efficient trigger for the selection of boosted top quark production where the top decays semi-leptonically, in the CMS experiment

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Murli Kartik Maurya

Indian Institute of Science Education and Research, Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.
April, 2018

Supervisor: Prof. Sudeshna Banerjee
DHEP, Tata Institute of Fundamental Research, Mumbai-400005 India

Certificate

This is to certify that this thesis entitled Study of physics algorithms to setup an efficient trigger for the selection of boosted top quark production where the top decays semi-leptonically, in the CMS experimenttowards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Murli Kartik Maurya, at Tata Institute of Fundamental Research, Mumbai under the supervision of Prof. Sudeshna Banerjee DHEP, Tata Institute of Fundamental Research, Mumbai-400005 India, during the academic year 2017-2018.

Murli Kartik Maurya

Supervisor

Prof. Sudeshna Banerjee

This thesis is dedicated to my family and my maternal grandfather.

Declaration

I hereby declare that the matter embodied in the report entitled Study of physics algorithms to setup an efficient trigger for the selection of boosted top quark production where the top decays semi-leptonically, in the CMS experiment are the results of the work carried out by me at the Department of high energy physics, Tata Institute of Fundamental Research, under the supervision of Prof. Sudeshna Banerjee DHEP, Tata Institute of Fundamental Research, Mumbai-400005 India and the same has not been submitted elsewhere for any other degree.

Murli Kartik Maurya

Sudeshna Banerjee

Supervisor

Prof. Sudeshna Banerjee

Acknowledgments

First, I would like to thank Prof. Sudeshna Banerjee for supervising this work. At every moment she has been very supportive and has given me the best advice. I want to thank her for careful review of my work and always encourages me to learn more about the fabrics of high energy physics.This thesis would not have been the same without her.

I also want to thank Dr. Bhavana Gomber, Dr. Zhenbin Wu, and Bajrang Sutar. They had taught me so many things in this field and always tried to explain me different concepts and entertain all my queries during my work period like teammates.

I am also grateful to faculty members in TIFR and from another institute Prof. Kajari Mazumdar, Prof. Sridhara Dasu and many more for their contribution to this work.

Because of this thesis, I got the chance to meet beautiful people in TIFR like Apoorva, Shivangi, Jhansi, Siddhesh and many more who keeps me motivated and confident all the time. I would not forget the boisterous laugh of Bajrang sir and Jhansi while watching any random video indeed fills the room with joy.

The last word goes to my parents, family members, friends and my all sisters for their unwavering support and care, which I cannot thank enough.

Abstract

This thesis presents a study of two topics; the first one is the study of jets of radius 0.8, reconstructed from anti-kt jet clustering algorithm for the semi-leptonic decay of $t\bar{t}$. The performance of the parameters of reconstructed jets which are coming from the similar to the data of the CMS has been presented here. The results are based on $t\bar{t}$ and QCD/Multi jets simulated events Monte Carlo (MC) data samples, these data samples have been generated at the different luminosities of the proton-proton collisions at LHC. The LHC machine, as well as the detector for the various experiments, will be upgraded in the next few years, with his upgrade machine luminosity of proton-proton collision will be five times greater than the present. This creates very much complex environment because the number of particles after the collision will increase, this will make the study of desired processes harder also due to the higher center of mass energy of the collision, many of the particles (e.g., top quark) will be highly boosted. Therefore, the study of shapes top quark jet (using NJettiness variables) will play an important role in identifying the top quark jet and its substructure.

The second topic of this thesis presents the results on the study of vivado High-Level Synthesis (HLS) software and its application for implementation of sophisticated algorithms e.g., algorithm for detecting electron and photon (γ) in Electromagnetic Calorimeter (ECAL) in CMS and passing this information to the Field Programmable Gate Arrays (FPGAs) hardware and to set up an efficient trigger for the CMS experiment.

Contents

List of Figures

List of Tables

Introduction

Since 1960, the theory of standard model (SM) of particle physics arose out of attempts by theorists (Glashow, Weinberg, Salam) to describe the fundamental forces (Electromagnetic, Weak and Strong interactions) in the universe and classifying all known elementary particles.

With the experimental discovery of $W^{\pm}$ and Z bosons in 1983 and SM has repeatedly examined and experimentally tested the various properties of $W^{\pm}$ and Z with great accuracy, which made electroweak theory widely accepted. Proposed model for quarks by physicists Murray Gell-Mann and George Zweig in 1964 were provided with evidence from experiments at Stanford linear accelerator centre (SLAC) in 1968 and with the discovery of top quark in Tevatron (Fermilab) in 1995. After the discovery of Higgs boson in LHC compatible with the scalar boson of the standard model (SM) in July 2012 and which has made physicists to study a rich variety of its decay products. LHC in its run I from 2011-2012, Collider had reached peak luminosity of $7.7 \times 10^{33} cm^{-2} s^{-1}$ more than 75 percent of its design luminosity and delivered an integrated luminosity of 25 fm^{-1} to Compact Muon Solenoid (CMS) and A Toroidal LHC ApparatuS (ATLAS) detectors, which result an confirmation of 125 GeV carrier standard model Higgs boson and serves as a efficacious evidence for (Higgs) mechanism by which spontaneous symmetry breaking in SM occurs, giving rise to masses of fundamental fermions and bosons. No evidence for the existence of an extended scalar sector has been observed, but some intriguing excesses ($H \to \mu^{\pm}\tau^{\pm}$ flavor violating decays, t$\bar{t}$ production, ...) need more data to be confirmed.In 2015,LHC for run II has allowed to collect 3 fm^{-1} data at 13TeV CM energy.

The Phase II upgrade[1] of LHC is aimed to answering plethora of fundamental questions in physics like nature of dark matter we observe in the universe? Properties of higgs boson? Behaviour of QCD under harsh environment? The basic goal of Phase II upgrade is to maintain the excellent performance of CMS detector in terms of efficiency, resolution and background rejection for all the physics objects used in the analysis of the data. The major concerns of high luminosity[1] upgrade in LHC (HL-LHC) is radiation damage to CMS detector from high integrated luminosity[2] and very high pile up[3] from high

[1]Luminosity gives a measure of how many collisions are happening in a particle accelerator per unit area per unit time.

[2]The integral of the delivered luminosity over time is called integrated luminosity.

[3]Proton beam at the LHC is bunched rather than continuous so every time bunch crossing happens,

instantaneous luminosity[4]. For HL-LHC Phase II the upgraded luminosity is of $2 \times 10^{35} cm^{-2} s^{-1}$ and centre of mass energy 14 TeV. Which increases the interaction rate and collision pileup beyond the capabilities of existing envisioned detector and trigger technologies. The nominal scenario is to operator at leveled luminosity of $5 \times 10^{34} cm^{-2} s^{-1}$, corresponding to mean pileup of 140 interaction per beam crossing.

This thesis is devoted to the study of physics algorithm for setup an efficient trigger in high luminosity environment and study of boosted top quarks decay analysis. In this project we had studied two topics, one topic deals with study of physics process by its data analysis and another is about firmware development studies in phase II trigger upgrades and more specifically electron/gamma algorithm implementation in synchronous way with hardware (FPGAs) for trigger upgrade using software HLS Vivado.

multiple protons interact (collides) with other which results few 'Hard scattering' interactions (Head-on collision) for interesting physics processes but there are other soft interactions also present which results the undesired background processes which results "pile up" event and time difference between two subsequent bunch crossings is 25 ns so there two types of pile up's in-time (within the same bunch crossing) and out-time (between the two subsequent bunch crossings). Our aim is always to increase the number of interesting events means more numbers of particle per bunch crossing means more pile up too

[4]The instantaneous luminosity is a measure of how many particles pass through a surface of unit area in unit time.

Chapter 1

Preliminary studies on high energy physics

In order to explore sub-atomic structure, physicists uses particle as a probe by increasing their momentum and making them go up to nearly the speed of light. The initial part of this chapter covers the theoretical background of particle physics and design and later part describes construction of CMS detector in LHC and the basic studies of high energy physics processes in CMS experiment which will be use full for our analysis.

1.1 The standard model of particle physics

The standard model (SM)[2] of particle physics is the theory of strong, weak and electromagnetic interactions of elementary particles. The reason of condescending faith on this theory is because it had passed stringent experimental tests. It is a gauge theory of $SU(3) \times SU(2) \times U(1)$ symmetry group.

SM does not include gravity, as SM beautifully describes near the electroweak symmetry breaking scale (246 GeV). But SM is not valid at energies above Planck scale (1.22×10^{19} GeV), where gravity can no longer be ignored.

Standard model of particle physics consists of fermions (half integer spin) and bosons

(integer spin). Gauge bosons mediates the forces: the photon mediates electromagnetism, gluon mediates the strong force and $W^{\pm}$, Z mediates the weak forces and there is the scalar (zero spin) higgs boson. Fermions (leptons and quark) are constituents of matter. Table 1.1 and 1.2 summarises some of the properties of quarks and leptons respectively. Standard model only describes visible matter in the universe. But unfortunately SM does not include the properties of dark matter particles. which also shows standard model theory is incomplete.

Particle	charge(e)	mass(MeV)
u(up)	$\frac{2}{3}$e	2.2
d(down)	-$\frac{1}{3}$e	4.7
s(strange)	-$\frac{1}{3}$e	96
c(charm)	$\frac{2}{3}$e	1.28×10^3
b(bottom)	-$\frac{1}{3}$e	4.18×10^3
t(top)	$\frac{2}{3}$e	173.1×10^3

Table 1.1: Properties of the quarks. e represents Coulomb charge[3].

Particle	charge(e)	mass(MeV)
electron(e)	-e	0.51099
electron neutrino(ν_e)	0	$\simeq 0$
muon(μ)	-e	105.65
muon neutrino(ν_μ)	0	$\simeq 0$
tau(τ)	-e	1776.86
tau neutrino(ν_τ)	0	$\simeq 0$

Table 1.2: Properties of the leptons. e represents Coulomb charge. Neutrino's mass not zero but very small compared to other SM particles[3].

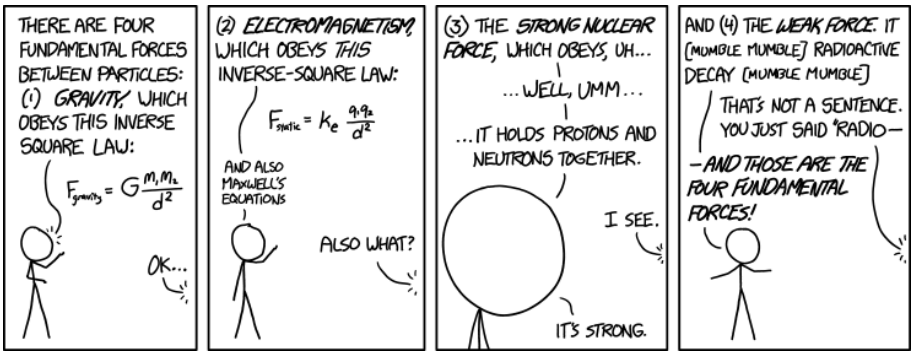


Figure 1.1: Fundamental forces[4]

1.2 The LHC and the CMS detector: A brief overview

The Large Hadron Collider (LHC)[6] is a ring shaped ($\sim$ 27Km circumference), most power full accelerator on this planet situated underground near Geneva. Where proton-proton beams (which travels very close to the speed of light in opposite directions) are getting collided with each other at collision point. LHC also has the facility for lead-lead or lead-proton collisions. But our focus is only on proton-proton collision.

In LHC, each proton beam has 2808 bunches with each bunch contains 10^{11} protons.To increase the probability of collision, 1.2×10^{11} (at start) numbers of protons in a bunch are squeezed with help of thousands of different magnets up to the width (human hair is of the width of $50\mu m$) $20\mu m$[6]. One way to increase the luminosity is to increase the number of bunches. At full luminosity the LHC uses a bunch spacing of 25 ns (or 7.5 m). Which corresponds to frequency of 40 MHz or bunches should pass each of collision points in LHC 40 million times per second. The average bunch crossing frequency in LHC is 30 MHz.

So, Bunch crossing on average is 30 million times per second and LHC produces 40 collisions between 200 billion particles. LHC generates 1 billion particle collision per second.The LHC started with 7 TeV centre of mass energy in 2010 and delivered 6.14 fb^{-1}, 23.30 fb^{-1} at 8 TeV in 2012 and, after long shutdown (LS1), of 4.22 fb^{-1} at 13 TeV in 2015 and is currently running at 13 TeV.

The interaction rate, $\frac{dN}{dt}$ depends on the process cross section σ and on luminosity $\mathcal{L}$ as:

$$\frac{dN}{dt} = \mathcal{L}\sigma \tag{1.1}$$

and

$$\mathcal{L} = \frac{\gamma f k_B N_P^2}{4\pi\epsilon_n\beta^*} \tag{1.2}$$

where γ is the Lorentz factor, f is the revolution frequency, k_B is the number of bunches, N_P is the number of protons/bunch, ϵ_n is the normalized transverse emittance (with a design value of 3.75 μm), β^* is the betatron function at IP (interaction point), and F is the reduction factor due to the crossing angle.

The LHC luminosity status (2017) at 13 TeV collisions are given in Fig. 1.2[7].

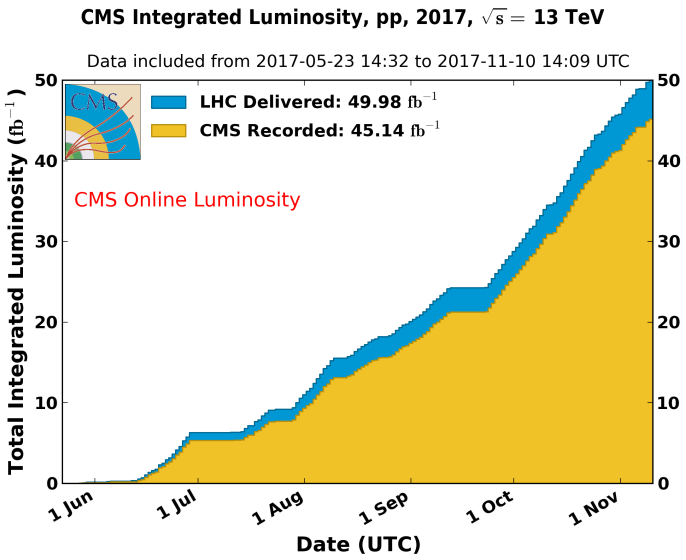


Figure 1.2: Online total integrated luminosity

1.2.1 LHC's detectors and the CMS detector

There are seven experiments installed at the LHC: A Large Ion Collider Experiment (ALICE), A Toroidal LHC ApparatuS (ATLAS), the Compact Muon Solenoid (CMS), the Large Hadron Collider beauty (LHCb) experiment, the Large Hadron Collider forward (LHCf) experiment, the TOTal Elastic and diffractive cross section Measurement (TOTEM) experiment, and Monopole and Exotics Detector at the LHC (MoEDAL)[6].

The Compact Muon Solenoid (CMS)[5] detector of LHC is a cylindrical shaped layered detector, which detects the stable particles at different layers and it has most powerful solenoid magnet of 4 Tesla (T). The collective information of particle from different layers of CMS is used to reconstruct the whole track of the particle. When the two protons collide with each other at the collision point in the middle of CMS, the quarks present inside the highly energetic protons (uud) interacts with each other and with the process of hadronization make up the hadrons which as a stable particle are getting detected in CMS and other leptons are also getting detected.

The CMS follows right-handed coordinate system at collision point, where y-axis points

vertically and x-axis points radially at the centre of the LHC. The z-axis is orthogonal to x-y plane (along the beam direction). The θ longitudinal angle/ polar angle/ centre of mass scattering angle is measured from z-axis in a plane perpendicular to x-y plane and the azimuthal/ transverse angle ϕ measured in x-y plane from x-axis. Transverse component of momentum of a particle lie in x-y plane.

The θ used to define angle of particle relative to beam axis which is called pseudorapidity η:

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{1.3}$$

Pseudorapidity is a spacial case of rapidity ζ for high energy particles, where ζ is the hyperbolic angle of rotations in lorentz transformation between two inertial frames.

$$\zeta = \tanh^{-1}(\beta) \tag{1.4}$$

where $\beta = \frac{v}{c}$, $-1 < \beta < 1$ and $-\infty < \zeta < \infty$.

And, rapidity can also be written in terms of energy and momentum:

$$\zeta = \frac{1}{2}\ln\left(\frac{E + |P|c}{E - |P|c}\right) \tag{1.5}$$

and ζ relative to beam axis is denoted by y:

$$y = \frac{1}{2}\ln\left(\frac{E + P_z c}{E - P_z c}\right) \tag{1.6}$$

This can be interpreted as rapidity of boost along the beam axis which takes observer from lab frame to frame in which particle moves perpendicular to the beam.In condition $E >> m$, y boils down to η and differences between y or η intervals of two particle is lorentz invariant. The distance between two physics objects is governed in $\eta - \phi$ space because it is frame invariant.

CMS performs many functions in LHC as a detector like Bending particles, Identifying tracks, Measuring energy, Detecting muons. It consists different layers of sub-detectors like Tracker, Electromagnetic calorimeter (ECAL), Hadron calorimeter (HCAL), Muon detector.

1. Tracker
   The tracker[8] sub detector is used to calculate the momentum of a particle by tracking its path through a magnetic field, the curvature of track of the particle is inversely proportional to the momentum of the particle. The curvature of the tracks of only charged particles is constructed with help magnetic field which deflects the opposite charged particles in opposite directions with particular curvature. Tracker is fully made up of silicon divided into two parts layers silicon pixel and silicon micro strips, particle while travelling through both of these layers produce electric signals, which later amplified and getting detected. Silicon pixel layer itself composed of 3 layers at a distance of 4cm, 7 cm and 11cm. Silicon strips contains 10 layes.

2. Electromagnetic calorimeter (ECAL)
   The ECAL measures the energy of the electron and photons, and covers $\eta = |3|$. Electromagnetic shower coming from electron or photon while entering the ECAL lead tungstun crystals.

3. Hadron calorimeter (HCAL)
   The HCAL, surrounds the ECAL which measures energy deposits of hadrons in jets and measure missing transverse energy (MET). HCAL covers up to $|\eta|= 3$ and $|\eta|= 5$ by iron/quartz fibre calorimeter or hadron forward (HF) ensuring nearly full geometric coverage for measurement of transverse energy in the event.

4. Magnet
   A 3.8 Tesla superconducting solenoid magnet of 12.9 m length used to curve the tracks of the charged particles.

5. Muon detector
   It is used to detect muons and to measure their momentum. It detects muons using gaseous detectors made up of 40 % Ar, 50 % CO_2 and 10 % CF_4, which gets ionised by muons and those ions later acts as a signal.

The CMS detector is illustrated in Fig. 1.3[5].

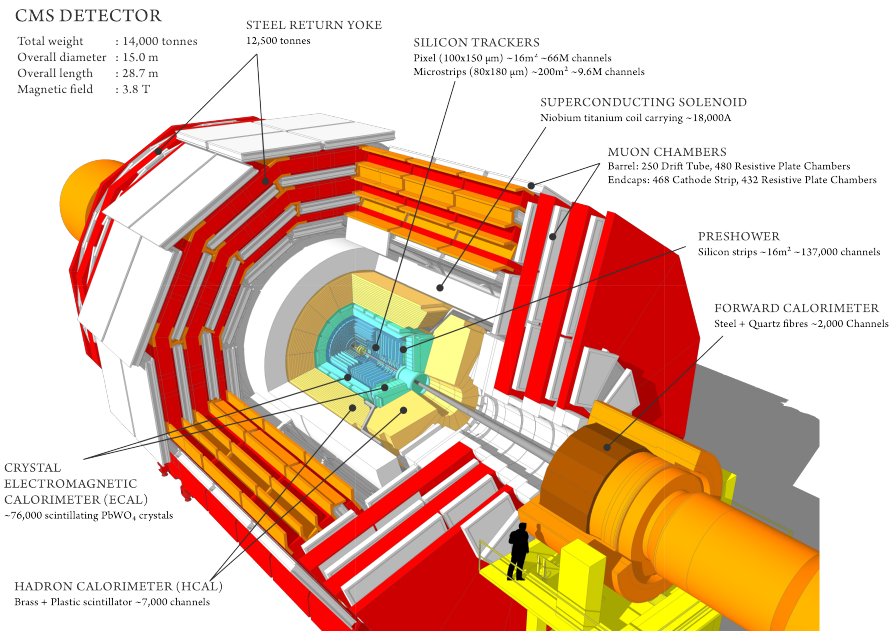


Figure 1.3: The Compact Muon Solenoid (CMS) detector.

1.3 A brief overview of trigger upgrade

In order to manage with high collision rate, 2-level Trigger system[9] uses limited information from each event[1] to decide whether or not to record the event. This greatly reduces the amount of data storage and more efficiently store the information of interesting physics object and so CMS uses a series of triggers to identify potentially interesting events and record them. The first level, the Level-1 (L1) trigger, uses custom-built electronics (like Field Programmable Gate Array (FPGA)) in order to reduce the output rate from 40 MHz to 100 kHz.

Events which satisfy some relatively loose set of criteria are passed to the second level, the high-level trigger (HLT), which further reduces the rate 400 Hz, where more sophisticated algorithms, much closer to those used in the offline reconstruction, are used to decide whether or not to store an event. A certain fraction of events, called minimum Bias

[1]Particles collide at high energies inside CERN's detectors, creating new particles that decay in complex ways as they move through layers of sub detectors. The sub detectors register each particle's passage and microprocessors convert the particles' paths and energies into electrical signals, combining the information to create a digital summary of the "collision event". The raw data per event is around one million bytes (1 Mb), produced at a rate of about 600 million events per second.

events are collected that have passed through the whole trigger system without putting any constraints/selection criteria whatsoever. These kind of events are used for calibration.

With the high luminosity (HL), CMS is also going to be upgraded, the performance issues associated with high pile-up (PU) needs to be mitigate which heavily depends on particle-flow (PF)[2] event reconstruction algorithms. To study interesting physics processes requires trigger electronics (L1) to be upgraded and trigger rate should be decreased by improving transverse momentum p_T resolution to obtain lower rates without loss of efficiency and by mitigating the effect of the combinatorial backgrounds[3] arising from pileup.

In phase II tracker by introducing tracking information at L1 improving the precise momentum measurement which will ensures increased background rejection in early stage. To overcome the radiation damage the calorimeters end caps will also be replaced by high granularity calorimeter with improved segmentation.

1.4 Event generation: using PYTHIA

From first principle we cannot find/calculate the structure of LHC events because it involves large amount of cascade of particle decays and can involves tons physics processes or its an evolution of few body hard scattering to complex multi particle final state as given in Fig. 1.4[10]. So, its impossible to calculate scattering amplitudes of an event using feynman diagrams. They involves several competing mechanisms which contribute perturbativly and nonperturbativly. We can use event generator like PYTHIA 8226[11] for simulating various events kinds using standard model physics we know, with random numbers providing quantum variability. PYTHIA 8 is very use full tool for predicting event rates and topologies, simulating possible background, study detector requirements.

PYTHIA works for either hadron-hadron or lepton-lepton collisions with CM energies goes up to 100TeV (but its not advisable to go such high energy). The outgoing particles are produced in vacuum and the simulation of the interaction of the produced particles

[2]The particle-flow event reconstruction aims at reconstructing and identifying all stable particles in the event, i.e., electrons, muons, photons, charged hadrons and neutral hadrons, with a thorough combination of all CMS sub-detectors towards an optimal determination of their direction, energy and type.

[3]Fake physics objects arising from inevitable different processes (background).

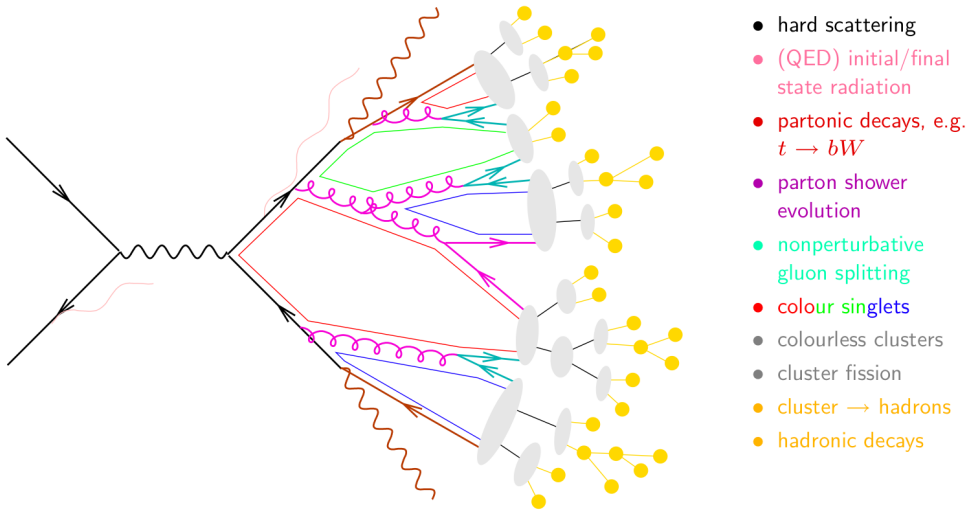


Figure 1.4: Generation of an $e^-e^+ \to t\bar{t} \to b\bar{b}W^-W^+$ event

with detector material is not included in PYTHIA. In PYTHIA hard processes $2 \to 1$, $2 \to 2$ with some $2 \to 3$ are available like QCD Processes, Electroweak(EW) Processes, Top Production, Higgs processes, SUSY Processes, New gauge boson processes, Compositeness processes etc. It contains currently, sixteen parton distribution function (PDF)[4] sets for the proton come built in.

1.4.1 Brief aspects of coding in PYTHIA

After success full installation in the system, in the one of its sub directory named **examples** contains paired files(mainNN.h and mainNN.cc) and to execute the test program run **make mainNN** later run executable **./mainNN** after success full compilation and we can get output in separate file with **./mainNN > output.txt**. Inside the main code the first step is to create event generator with **"Pythia pythia;"** it initializes all the default values for the Settings and the Particle Data databases. PYTHIAs settings and particle data can be changed by the two methods

pythia.readString(string);

[4]To investigate the behaviour and dynamics of fundamental particles quarks and gluons in nucleons, we need very high resolution (Perturbative QCD regime) and at that scale we can only talk in terms of probability(QM) distribution of quarks and gluons like a snapshot of internal environment of proton which shape the probability of finding quarks and gluons within a proton which is called "parton distribution function".

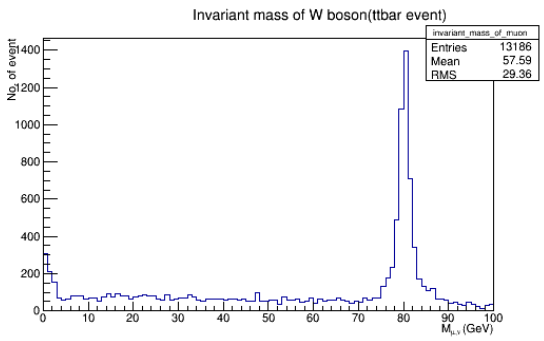


Figure 1.5: The invariant mass of W boson in $t\bar{t}$ event using pythia.

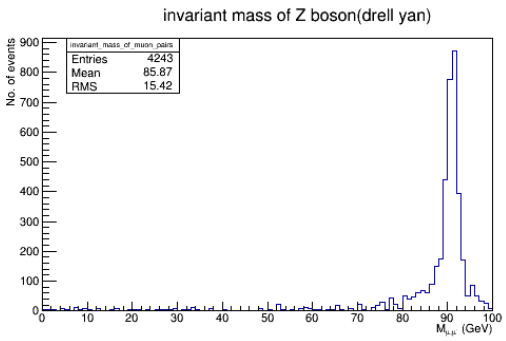


Figure 1.6: The invariant mass of Z boson in Drell Yan event

for changing a single variable, and

pythia.readFile(fileName);

for changing a set of variables, one per line in the input file.

pythia.next();

is required to generate the next event. This method would be located inside an event loop, where a required number of events are to be generated. The key output of the **pythia.next()** command is the event record found in **pythia.event**. A process-level summary of the event is stored in **pythia.process**. As an output we are getting lots of information about an event like the identity according to the PDG particle codes, the status (production reason, decayed or not), two mother and two daughter indices (can represent ranges), a colour and an anti colour tag, the four-momentum and mass,a production scale, a four-vector representing the production vertex, a pointer to the relevant particle data table entry and a pointer back to the event the particle belongs to.

Other information like incoming beams, the event type, kinematics of the hard process, values of parton distributions and couplings, event weights and cross section statistics. We can interface root software with PYTHIA by making changes in code file and run with modified command[5]. We had studied two processes Drell Yan and $t\bar{t}$ production in PYTHIA using root software for plotting. In $t\bar{t}$ production ($gg \to t\bar{t}$) we had plotted for invariant mass of W boson which is decaying into muon and nutrino and it matched approximately with with actual mass 80.385±0.015 $\frac{GeV}{c^2}$ (Fig. 1.5) and similarly in Drell Yan process we had plotted invariant mass for Z boson ($M_z \sim 91.18$GeV) which is decaying

[5]g++ main12.cc -o main12 -I../include -O2 -std=c++98 -pedantic -W -Wall -Wshadow -fPIC -L../lib -Wl,-rpath,../lib -lpythia8 -ldl 'root-config –cflags' 'root-config –glibs'

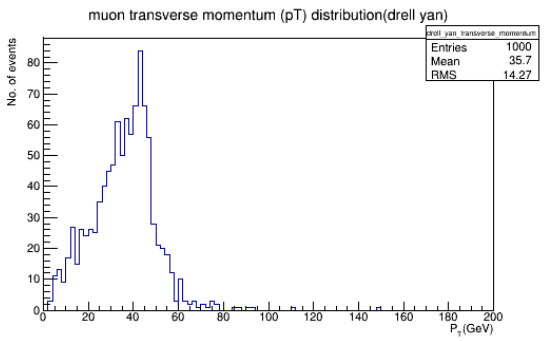


Figure 1.7: The p_T distribution of μ in Drell Yan event

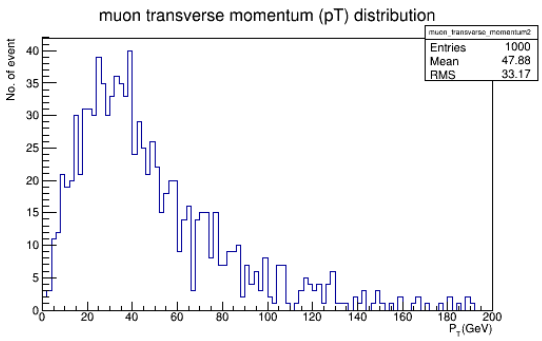


Figure 1.8: The p_T distribution of μ in $t\bar{t}$ event using pythia.

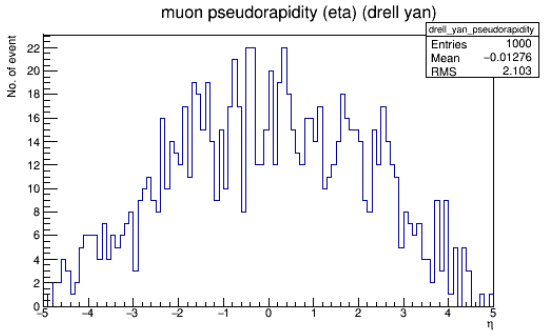


Figure 1.9: The η distribution of μ in Drell Yan event

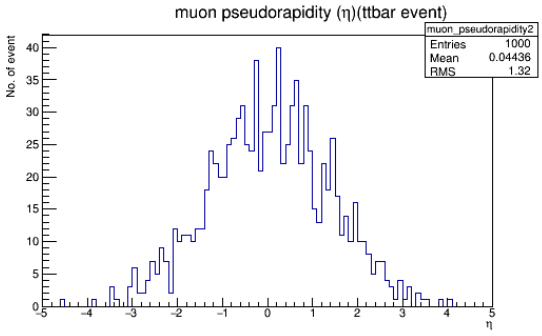


Figure 1.10: The η distribution of μ in $t\bar{t}$ event

into muon pair (Fig. 1.6) and In Fig. 1.7 and 1.8, we have plotted p_T distribution of μ from $t\bar{t}$ and Drell Yan event, as we can see there are more high p_T μ's in $t\bar{t}$ event as compared to in Drell Yan event. In Fig. 1.9 and 1.10, is for μ's pseudorapidity(η) in $t\bar{t}$ and Drell Yan[6] event.

So, We can use PYTHIA to study different types of particle production in proton-proton collisions at various centre of mass energies.

[6]Drell Yan is a high energy hadronhadron scattering between quark and antiquark of opposite hadrons, which creates a virtual photon or Z boson which then decays into a pair of oppositely-charged leptons.

1.5 Jet reconstruction algorithms

Jet is a collimated spray of particles(electron, photon, charged and neutral hadrons) resulting from hadronization of partons. Jet algorithms reduces the complicated multiparticle event into simple final state of few jets, it acts as a mapping between parton level to hadron level final state. The study of jets is important to test perturbative QCD, to probe proton structure and looking for new physics. Jet algorithms gives a set rules for grouping particles into jet. Jets has to be infrared and collinear (IRC) safe and it is the property that if one modifies an event by a collinear splitting or the addition of a soft emission, the set of hard jets that are found in the event should remain unchanged. In the IRC unsafe algorithm, a soft emission leads to a different set of final state jets and thus to the lack of cancellation of soft and collinear divergences (KLN theorem).

There are two broad classes of jet algorithms[12,13,14,15,16], the Cone algorithm and Sequential recombination algorithm (SRA). In cone algorithm, we put together particles in some specific conical angular region (in $\eta - \phi$ space) around some seed particle so that 4-momentum sum of particles in given region coincides with the cone axis, then we can call it a stable cone which is quite a brute force method. Until 2007, all the cone jet algorithms used (JetClu, MidPoint, ...) were IRC unsafe, But later using SISCone algorithm we can make IRC safe cone algorithms. In SRA, we identify the pair of particles that are closest in some symmetric distance measure d_{ij} depending upon their transverse momentum, then recombines them and put them separately and repeat it again till any halting condition comes up. Depending upon the distance measure there are 3 different kinds of algorithms:

$$\begin{aligned} d_{ij} &= min(P_{T_i}^n, P_{T_j}^n)\left(\frac{\triangle R_{ij}}{R_0}\right)^2 \\ d_{iB} &= P_{T_i}^n \end{aligned} \tag{1.7}$$

Where d_{ij} is the distance measure between all pair of particles i and j, d_{iB} is the distance measure between every particle i and beam, $\triangle R_{ij} = \sqrt{(y_i^2 - y_j^2)^2 + (\phi_i^2 - \phi_j^2)^2}$ and R_0 is the jet radius. If the $n = 2$ then d_{ij} and d_{iB} will be for **kt Algorithm**, if $n = -2$ then d_{ij} and d_{iB} will be for **anti-kt Algorithm** and if $d_{ij} = \left(\frac{\triangle R}{R_0}\right)^2$ and $d_{iB} = 1$ it makes **c/A Algorithm**.

If $R_0 = 1$, then we look for smallest between d_{ij} and d_{iB} if it is d_{ij}, then we replace i and j with single new object whose momentum $p_{T_i} + p_{T_j}$ called **pseudojet** and if it is d_{iB}, then we remove i and call it **beamjet**. After repeating the process until smallest d_{ij} or d_{iB} is above the threshold d_{cut}, then all the particles/pseudojets left are then that event's **jet**. Current prevailing scenario dictates anti-kt algorithm is working more efficiently and producing good results.In LHC analysis, SRA algorithms are mostly used.

1.6 Event Reconstruction in CMS

To understand the standard model processes, it can be modeled with Monte Carlo (MC) event simulation based on CMS detector response using event generators and softwares like PYTHIA, MADGRAPH etc. The information from simulation (generator level) has to be matched with information coming from reconstruction of event.

1.6.1 Particle flow algorithm in CMS

A particle Flow (PF) algorithm[17, 18, 19] event reconstruction algorithm has been aimed to identify and reconstruct individually each physics objects and retrace the whole event picture of p-p collision using information coming from all sub detector parts of CMS. The individual PF particles - e^-, μ^-, γ, $h^\pm$, h^0 by getting the information at every possible level of the CMS detector and storing their information like p_T, energy, type etc. has made possible the reconstruction of more complicated objects like Jets, Missing Transverse Energy (MET), reconstruct and identify τ decay products etc. PF Jet reconstruction example is illustrated in Fig. 1.12[20].

PF algorithm consists of:

1. Fundamental ingredients
    - calorimeter clustering
    - tracking and extrapolation to the calorimeters
    - muon identification

- electron pre-identification

2. Linking topologically connected elements

3. Particle identification and reconstruction

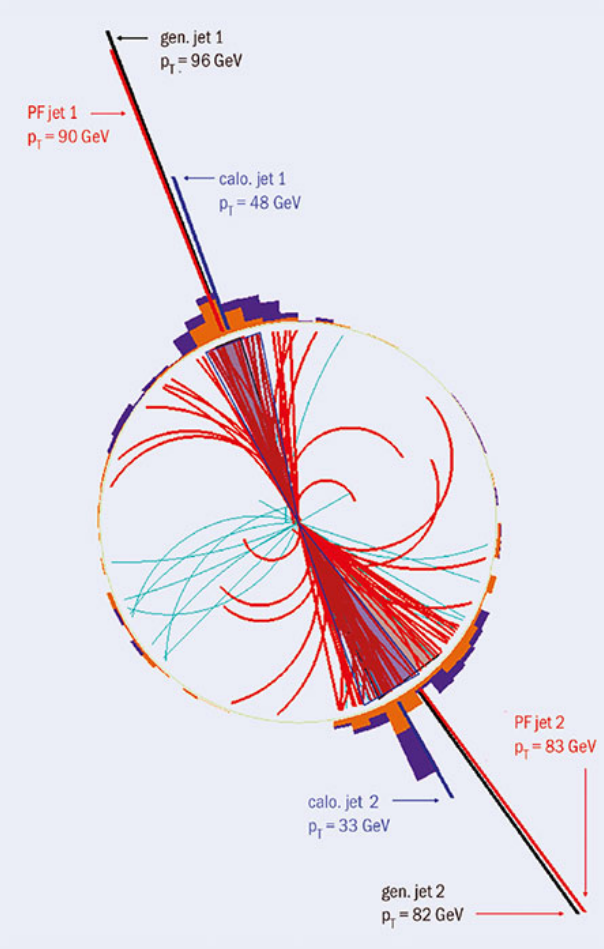


Figure 1.11: Jet reconstruction in a simulated dijet event. The reconstructed particles clustered in the two jets are displayed with thicker lines. For clarity, unclustered particles with $pT < 1$ GeV are not shown. The particle-flow jet transverse momentum, indicated as a radial line, is compared to the momenta of the corresponding generated and calorimeter jets.

An event originate from primary vertex[7] p, from p interacting particles starts interacting with detector (CMS) like in tracker only charged particles leaves their tracks (because they are in electromagnetic field) and from their radius of curvature we can find the particle's

[7] $p - p$ collision point in space

momentum. Since photons does not leave their tracks in tracker but they store their energy in ECAL so in ECAL γ and e^- are detected as they deposits their energy in ECAL clusters. But sometimes ECAL deposits does not belong to any silicon tracks (reducing the possibility of charged particle candidate) + any photon candidate which shows that this new candidate can't be charged and neither photon also which leads to possibility of neutral hadron(h^0) candidate so we need to go into HCAL, as photon candidate does not deposit any energy in HCAL so this new candidate is h^0.

So getting information from ECAL+HCAL, we can construct a jet by jet clustering as shown in the Fig. 1.11, where dark purple rectangular energy deposits are from Calorimeters. In particle flow, link algorithm actually links all the information from all the sub detectors and reconstruct particles candidate.

1.7 Chapter summary

The higgs boson particle discovered in LHC, matches with the scalar boson of SM and which further strengthens the theory of standard model. Simulation software gives a very good insight of an p-p collisions and with the help event reconstruction algorithms like PF we can reconstruct particles and further reconstruct other complicated objects like jets, MET etc.

Chapter 2

Top quark decay event analysis

In this chapter focus is on semi-leptonic decay of $t\bar{t}$ at different luminosities of the proton-proton collisions at LHC. The discussion on generation of response and resolution plots of reconstructed Anti-kT jets of radius 0.8 (AK8) in $t\bar{t}$ decay using Monte Carlo (MC) data and compare it with QCD events in 0 and 140 pileup. In the end, the Study of jet substructure has been describe for AK8 jets.

2.1 Motivation and the basics

After the discovery of 125 GeV scalar boson, LHC is looking for new physics - beyond standard model (BSM) which predicts for giving solution of hierarchy problem. In which LHC will produce new heavy particles involving involving decay channels with top quark, higgs boson and W/Z bosons.

Therefore, study of top quarks becomes important in different kinematic conditions. There are many other unexplored question needs to be answer like prevailing theory dictates that particles gain mass through interactions with the Higgs field, so why do top quarks interact so much more with the Higgs than do any other known particles? which can be studied by making direct measurement of top-higgs interaction side by side, but this has not happened yet as it requires high production rates. At tevatron collider ($t\bar{t}$) production [bib20], was dominated by $q\bar{q}$ around 81% and $g\bar{g}$ around 19% and decay modes

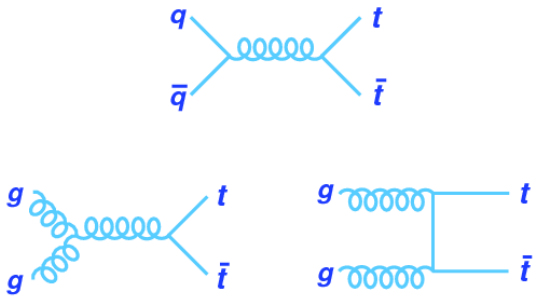


Figure 2.1: Tree level Feynman diagrams of $t\bar{t}$ production

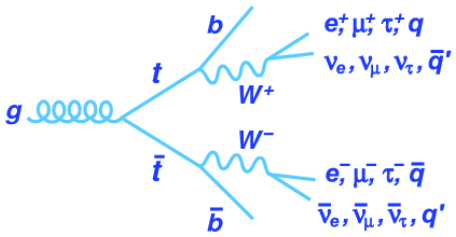


Figure 2.2: Tree level Feynman diagrams of $t\bar{t}$ decay.

are shown in Fig. 2.1[21].

Mostly top quark never hadronizes as the mean lifetime of top quark $\sim 10^{-25}$s and typical mean lifetime for hadronization is $\sim 10^{-24}$s. Top quark decays into W and b quark Fig. 2.2 and for our analysis we are studying semi-leptonic decay of $t\bar{t}$.

2.2 Top quark decay jets analysis and results

The study of efficiency of jets reconstruction of $t\bar{t}$ decay products are carried out by MC data samples used through out this document, together with different jet reconstruction algorithms. The MC samples used in this analysis are stored in fnal area [1]. The MC samples used in this documents are:

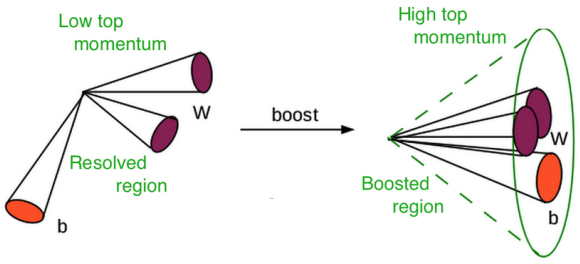
QCD PU140	$t\bar{t}$ PU140
QCD PU0	$t\bar{t}$ PU0

Table 2.1: Monte Carlo (MC) samples

In $t\bar{t}$ decay, there are 4 jets, 2 coming from W boson hadronic decay and 2 are b quark jets. Foe high pT top quark decay in boosted regime where in lab frame the decay products with momentum at high energies is in the same direction as the momentum of top. Jets can be imagined as the cone made up of closely packed spray of hadrons coming from parent

[1] /eos/uscms/store/user/benwu/Phase2L1/FatJet/Oct02withGenv2 ntuples

parton. In high energy scenario these cones started to approach each other, which makes the identification of jets even more difficult as illustrated in Fig. 2.3[22]. MC samples have reconstructed jets (recojets) information using anti-kt jet clustering algorithm with cone size of 0.8 (AK8) from other sub detector parts and generated particle level jet (genjet).



/all

Figure 2.3: Resolved and Boosted region in top quark decay.

The different reconstructed AK8 jet information in MC samples are -

1. Calo : corrected calorimeter tower
2. RawCalo : uncorrected calorimeter tower
3. TK : Tracks
4. TkVtx : Tracks from primary vertex
5. PF : Particle flow object reconstructed from level I particle flow algorithm
6. PUPPI : puppi algorithm on particle flow

PUPPI is a new method of pileup mitigation by implementing "Pileup per particle identification" (PUPPI)[23], which removes the pileup selectively for each particle. These tree stores information like jetmass distribution, eta distribution, p_T distribution etc. The Fig. 2.4 and 2.5 shows the reconstructed (reco) AK8 jet p_T distribution of $t\bar{t}$ for pileup 0 and 140 and other reco p_T distribution plots for QCD is in [Appendix A].

The Jet mass gives the invariant mass of the interaction that produced them, that is boosted top quark in our case. Which is the combine energy of 3 jets coming from

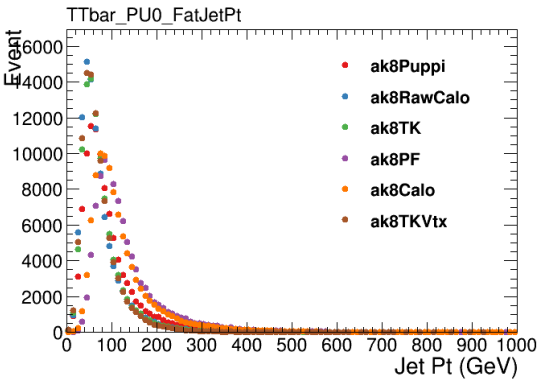


Figure 2.4: The reco p_T distribution for $t\bar{t}$PU0

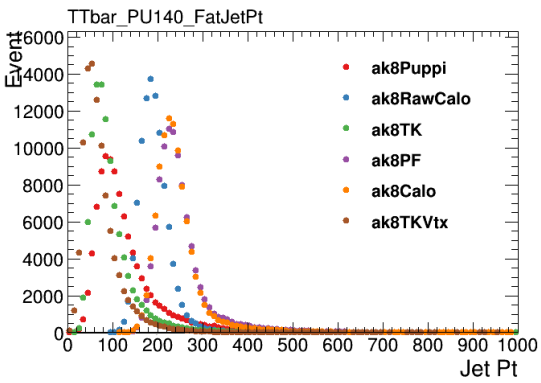


Figure 2.5: The reco p_T distribution for $t\bar{t}$PU140

calorimeter energy clusters.

$$\left(m^{Jet}\right)^2 = \left(\sum_i E_i\right)^2 - \left(\sum_i P_i\right)^2 \tag{2.1}$$

where E_i and P_i are energy and 3- momentum of i^{th} jet constituent.

Fig. 2.6 and Fig. 2.7 are showing the jet mass distribution for $t\bar{t}$ at pileup (PU) 0 and PU140 and we can see that in the $t\bar{t}$ PU140, the peak position is at far away from actual top mass because these reco AK8 jets are not matched GenTop jet. So, we are getting lots PU contribution and even for mass matched with GenTop, it is never going to be $\sim$ 173 GeV exactly.

In $t\bar{t}$ PU0, peak more on left side because there is no pile up in this case. Similar kind of behaviour we obtained for QCD in PU140 and PU0[Appendix A].

2.2.1 Study of response and resolution of AK8 jets

To check for efficiency of reco AK8 jets in $\eta = |2.5|$, the study of response and resolution of reco AK8 jets is needed. The simulated jet response is the ratio of arithmetic means of matched reconstructed and generator level (genjet) jets transverse momenta is plotted against genjet p_T for jet radius R=0.8 from different sub detector area in specific $\eta = |2.5|$ region. So, we had generated response plots which are between the average value of the

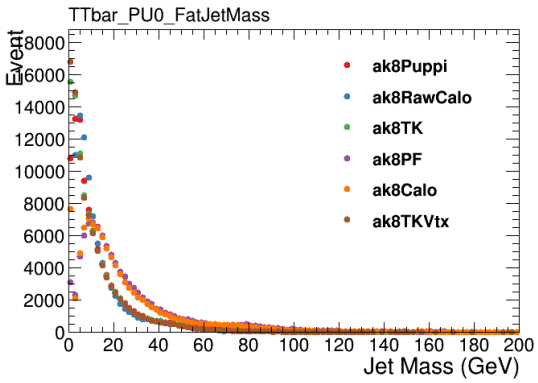


Figure 2.6: The reco Jet mass distribution for $t\bar{t}$PU0

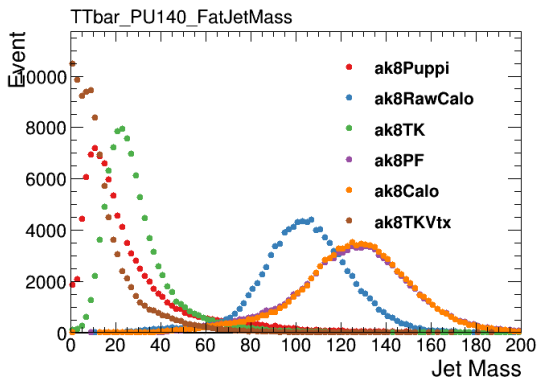


Figure 2.7: The reco Jet mass distribution for $t\bar{t}$PU140

ratio of matched reco AK8 jet p_T and genjet (<recojet p_T/genjet p_T >) versus genjet p_T The matching condition for reco AK8 jet with genjet is if the distance between the two in $\eta - \phi$ space is less than 0.4, then we call it matched pair.

These plots are generated with non uniform binning for genjet p_T on horizontal axis. The values of (<recojet p_T/genjet p_T >) are coming from the series of projection plots fetched from parent scatter plot between recojet p_T/genjet p_T and genjet p_T given in Fig. 2.8 and fitted with Gaussian curve for those uniform binning. The scatter plots between reco AK8 jet p_T and genjet p_T has been illustrated in Fig. 2.8 and Fig. 2.10 for $t\bar{t}$ PU140 of AK8Puppi and with added feature of events number density in Fig. 2.9 and Fig. 2.11.

The simulated response plots for AK8 jets p_T are shown in Fig. 2.12 to Fig. 2.15 with different processes.

We can see in high pileup PU140 of $t\bar{t}$ and QCD given in Fig. 2.13 and Fig. 2.15, PUPPI has shown the best response of matched AK8 jets for high p_T regime tending to 1 and for PU0 case in $t\bar{t}$ and QCD, the PF is working better as expected because there is no pileup, PUPPI acts on PF to mitigate pileup but there is no pileup in this case.

The jet p_T resolution[24] plots are made by fetching the ratio of σ and average value of Gaussian fitted projections of scatter plots of recojet p_T/genjet p_T and genjet p_T with non uniform binning. The resolution plots for AK8 jets p_T are shown in Fig. 2.16 to Fig. 2.19 with different processes.

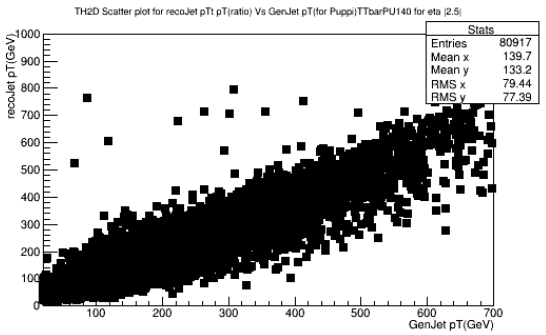


Figure 2.8: The scatter plot of $t\bar{t}$ PU140 for AK8 Puppi between recojet $p_T$and genjet p_T.

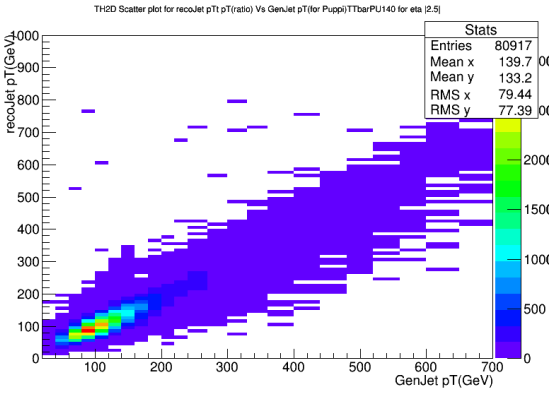


Figure 2.9: The scatter plot of $t\bar{t}$PU140 for AK8Puppi between recojet $p_T$and genjet p_T with number density.

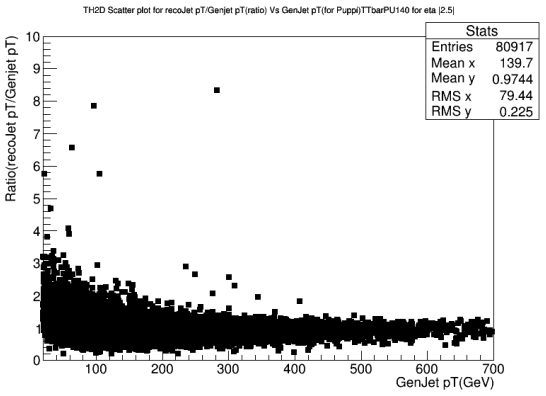


Figure 2.10: The scatter plot of $t\bar{t}$ PU140 for AK8 Puppi between recojet p_T/genjet p_T and genjet p_T.

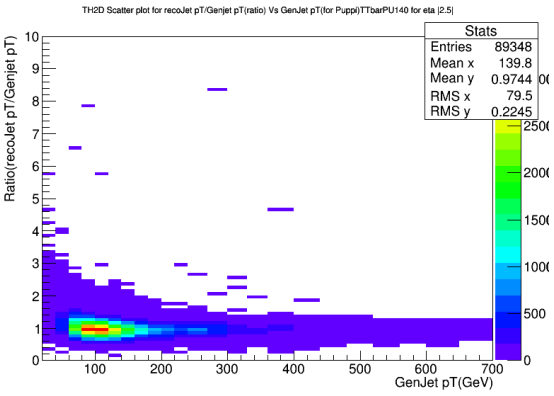


Figure 2.11: The scatter plot of $t\bar{t}$ PU140 for AK8Puppi between recojet p_T/genjet p_T and genjet p_T with number density.

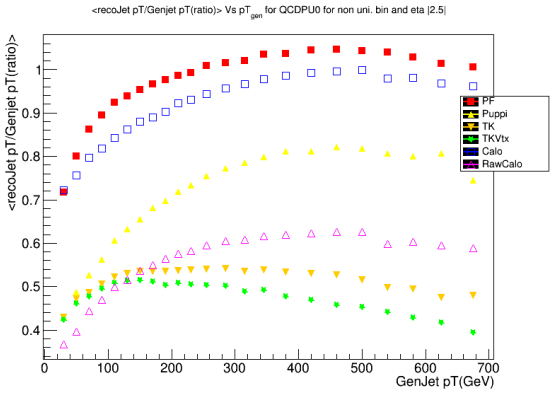


Figure 2.12: The simulated response plot for QCD PU0.

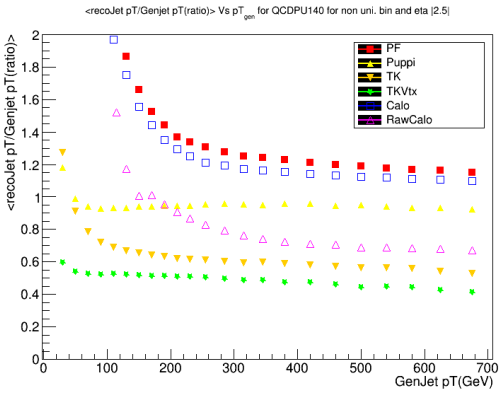


Figure 2.13: The simulated response plot for QCD PU140.

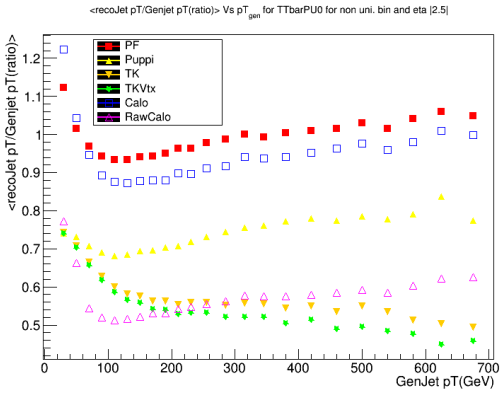


Figure 2.14: The simulated response plot for $t\bar{t}$ PU0.

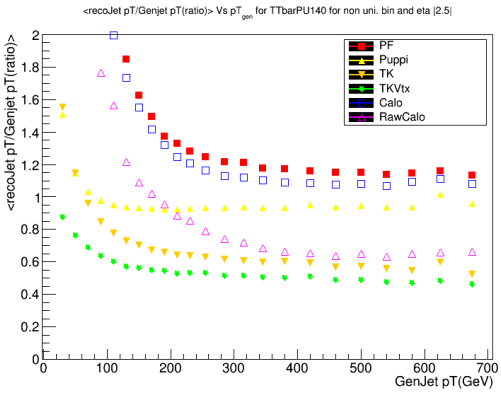


Figure 2.15: Simulated response plot for $t\bar{t}$ PU140.

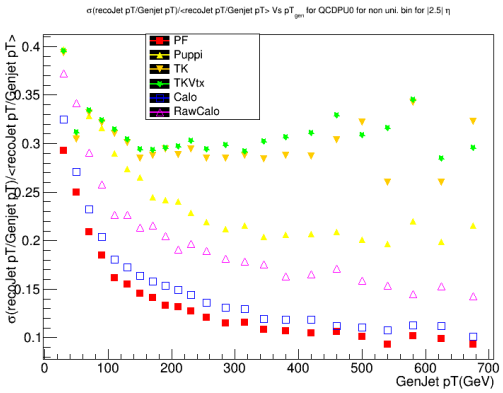


Figure 2.16: The resolution plot of AK8jets for QCD PU0.

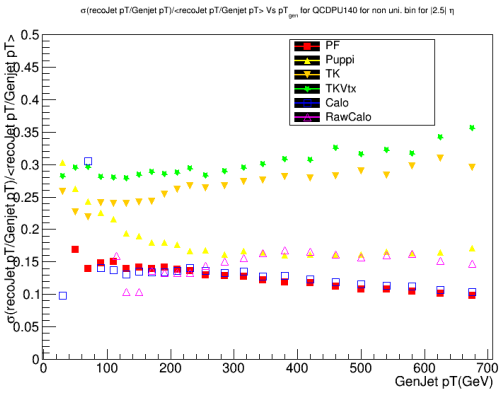


Figure 2.17: Resolution plot of AK8 jets for QCD PU140.

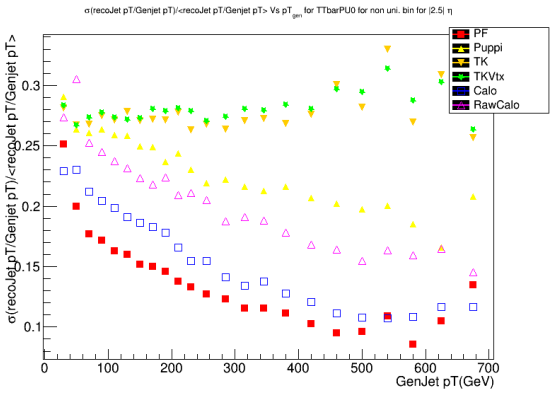


Figure 2.18: The resolution plot of AK8 jets for $t\bar{t}$ PU0.

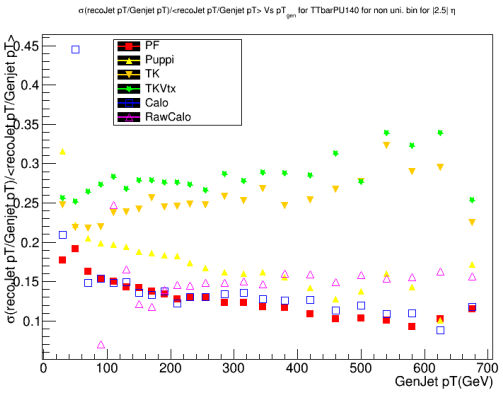


Figure 2.19: The resolution plot of AK8jets for $t\bar{t}$ PU140.

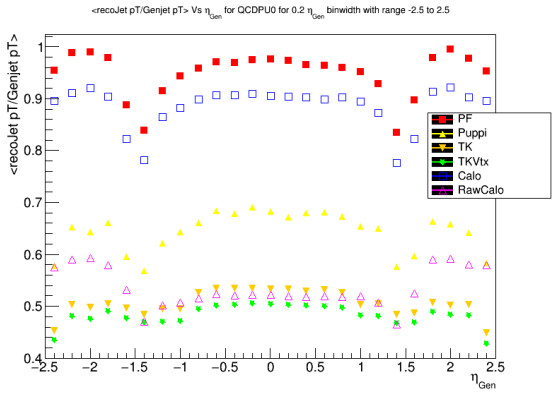


Figure 2.20: Eta simulated response plot for QCDPU0.

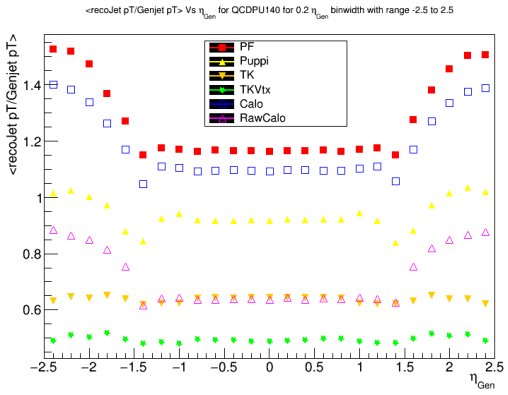


Figure 2.21: Eta simulated response plot for QCDPU140.

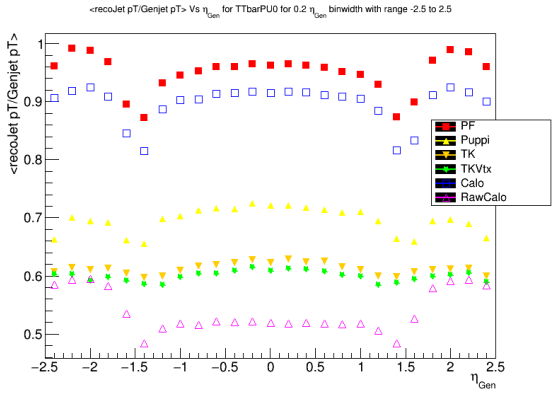


Figure 2.22: The eta simulated response plot for $t\bar{t}$ PU0.

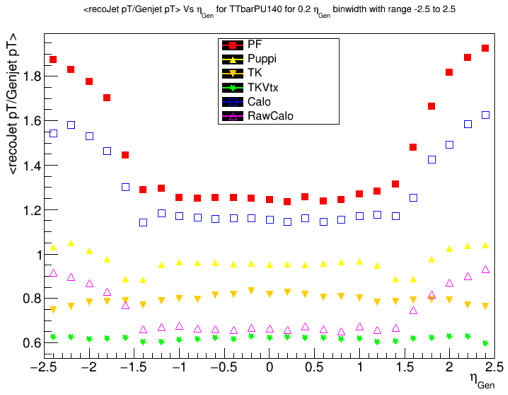


Figure 2.23: The eta simulated response plot for $t\bar{t}$ PU140.

We can see the AK8 PUPPI jets performance increases for $t\bar{t}$ PU140 in Fig. 2.19 as compared to $t\bar{t}$ PU0 and similar kind of behaviour has been observed for QCD PU140.

We had similar response and resolution plots for matched AK8 jets p_T for genjet eta $\eta = |2.5|$ cut. The eta response plots are shown in Fig. 2.20 to Fig. 2.23 , again it shows that AK8 PUPPI jet has better response for $t\bar{t}$ PU140 and QCD PU140 in barrel region $\eta = |1|$ as compared to AK8 PF jets, as the average value for the ratio of matched reco AK8 jet p_T with genjet p_T is very close to 1. Whereas for PU0 of QCD and $t\bar{t}$ its AK8 PF jet which perform better.

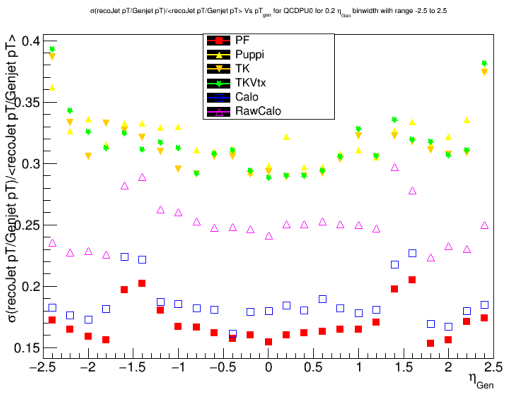


Figure 2.24: The eta resolution plot of AK8 jets for QCD PU140.

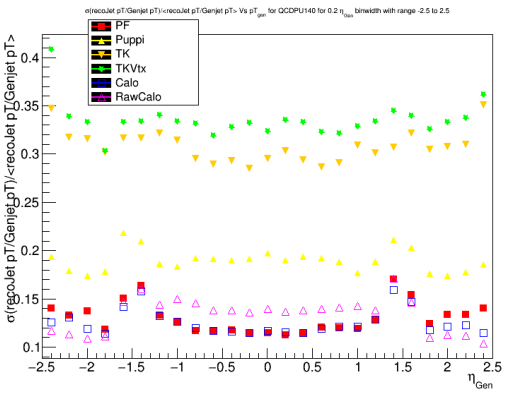


Figure 2.25: The eta resolution plot of AK8 jets for QCD PU0.

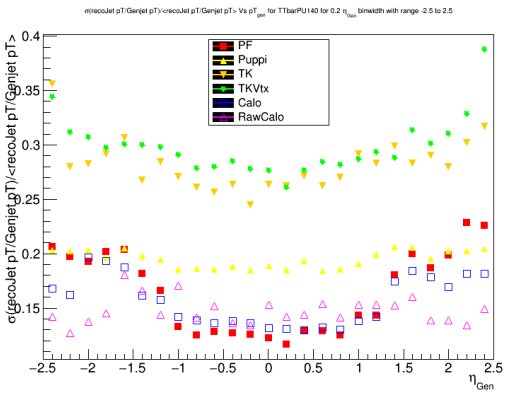


Figure 2.26: The eta resolution plot for AK8 jets of $t\bar{t}$ PU140.

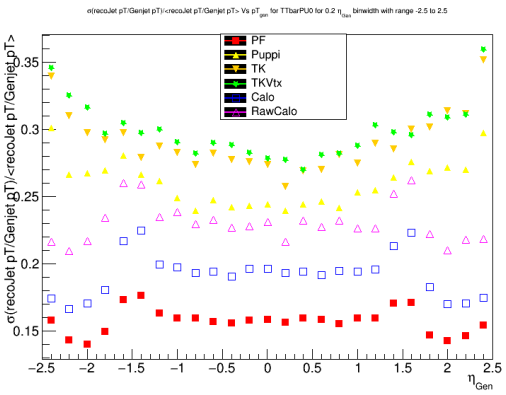


Figure 2.27: The eta resolution plot for AK8 jets of $t\bar{t}$ PU0.

The eta resolution plots are shown in Fig. 2.24 to Fig. 2.27, the $t\bar{t}$ and QCD in PU140 AK8 PUPPI got improved as compared to in PU0.

In conclusion, the reconstructed PUPPI AK8 jets and particle flow jets had shown the best response and resolution performance in large pileup and in resolution the AK8PF jets had shown better results than AK8PUPPI jets which is kind of unique behaviour till now and is going to be analyse more in comparison with AK4 jets[Appendix B]. The response and resolution plots are help full in study of AK8 jets in high pileup environment.

2.3 Anti-kT jets substructure study

The boosted top quark can decay 2 ways- one is leptonically, but the decay products has smaller angular separation between lepton and b quark jet. So, its very tricky to isolate lepton but could be achieve by shrinking the isolation cone around the lepton in function of its p_T. The second way is top decay hadronically, the decay products might be too close together for reconstruct separately three jets. That is why are using anti-kt jets of radius 0.8 . To study AK8 fatjet, we are required to apply algorithm on it to recover the (sub structure)particle level jets of top quark decay fatjet, for this we had used the NJettiness variable τ_N[25,26].

$$\tau_N = \frac{2}{Q^2} \sum_k min\Big(q_a.p_k, q_b.p_k, q_1.p_k, q_2.p_k, \cdots, q_N.p_k\Big) \tag{2.2}$$

Where $q_a, q_b, q_1, q_2 \cdots, q_N$ are a fixed set of mass less reference momenta for two beams and Nsignal jets and p_k is the momenta of all measured particle in final state.

τ_N quantifies the distance of particle from beam and jets or it gives us the inclusive event shape as how the N energetic jets event looks. The $\tau_N \sim 0$ means event contain infinitely N narrow jets. So, we we are trying to find subjets inside each top jet and impose kinematical constraints.

The MC samples for AK8 jets had default Njettiness information from CMSSW software for $t\bar{t}$ and QCD. Njettiness variable provide for discrimination between jets coming from boosted top quark decays and other processes. For boosted top quarks the variable ratio $\tau_{32} = \tau_3/\tau_2$ and $\tau_{21} = \tau_2/\tau_1$ are important. In high pile up scenario, there are high branching fraction of pairs of gauge boson and top quark. So, the study of new physics has to be done by effectively identifying the final state of gauge boson and top quark.

In semi leptonic decay of top quark, the way to study internal structure of jets and to make distinction between jets originating from boosted electroweak bosons and top quarks ('W jets' and 'Top jets'). So, we need to tag the boosted object using NJettiness variable.

The variable τ_{21} discriminates two prong objects like boosted W, Z, higgs boson and τ_{32} is for 3 prong objects like boosted top quark. In this document we had focused on top quark decay into b jet and W boson. If W boson decay hadronically into 2 quarks, top jet

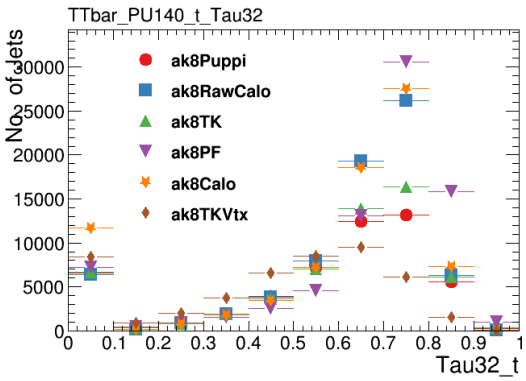


Figure 2.28: The τ_{32} matched with GenTop distribution for $t\bar{t}$ PU140 without any p_T cut.

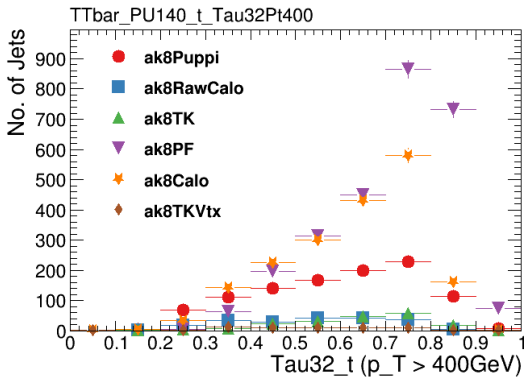


Figure 2.29: The τ_{32} matched with GenTop distribution for $t\bar{t}$ PU140 with $p_T >$ 400GeV

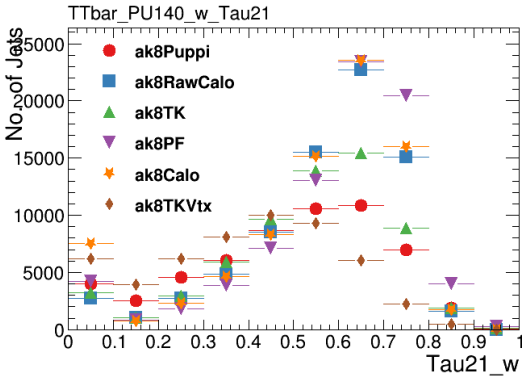


Figure 2.30: The τ_{21} matched with GenW distribution for $t\bar{t}$ PU140 without any p_T cut.

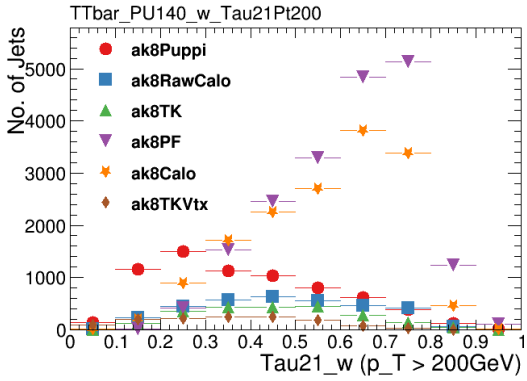


Figure 2.31: The τ_{21} matched with GenW distribution for $t\bar{t}$ PU140 with $p_T >$ 200GeV

will have 3 lobes of energy. So, τ_{32} has to be effective discriminating variable of top jet.

The reproduced distribution plots of τ_{21} and τ_{32} for boosted top quark are shown in Fig. 2.28 to Fig. 2.31 for $t\bar{t}$ PU140 and in Fig. 2.30 to Fig. 2.33 for $t\bar{t}$ PU0. We had matched reco AK8 jets with matching condition as dR < 0.4 and ignore if it has lepton with dR $<$ 0.4. The AK8 PUPPI jet is looking promising with τ_{21} matched with Gen W for higher p_T cut ($p_T > 200$ GeV) in $t\bar{t}$ PU140, as there are more number of jets with 2 sub jets identified inside fatjet in Fig. 2.31 as compared to no p_T cut in PU140 itself in Fig. 2.30. Whereas in τ_{32}, AK8 PUPPI jet does not look good for τ_{32} matched to GenTop in PU140.

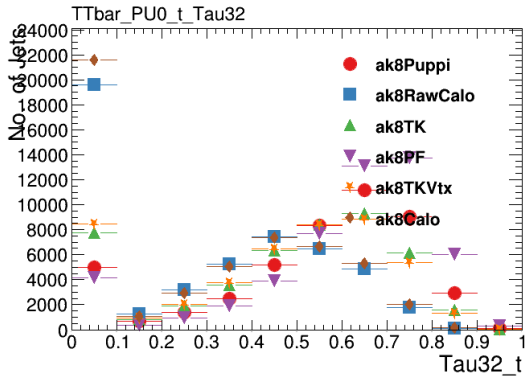


Figure 2.32: The τ_{32} matched with GenTop distribution for $t\bar{t}$ PU0 without any p_T cut.

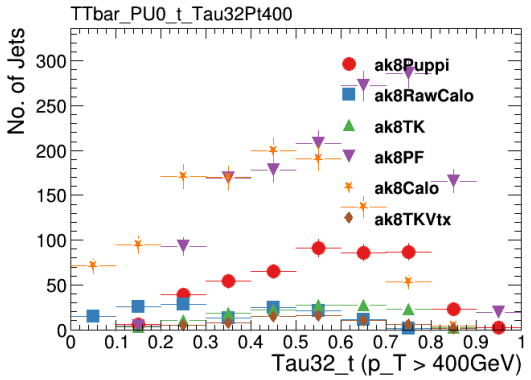


Figure 2.33: The τ_{32} matched with GenTop distribution for $t\bar{t}$ PU0 with $p_T >$ 400GeV.

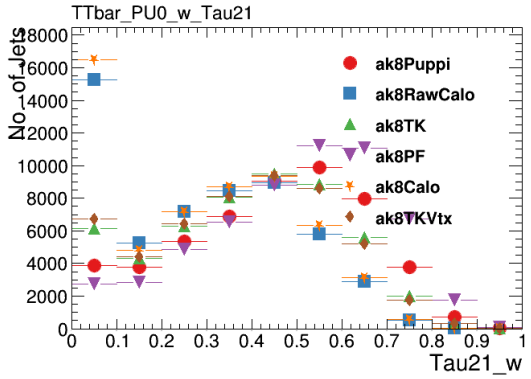


Figure 2.34: The τ_{21} matched with GenW distribution for $t\bar{t}$ PU0 without any p_T cut.

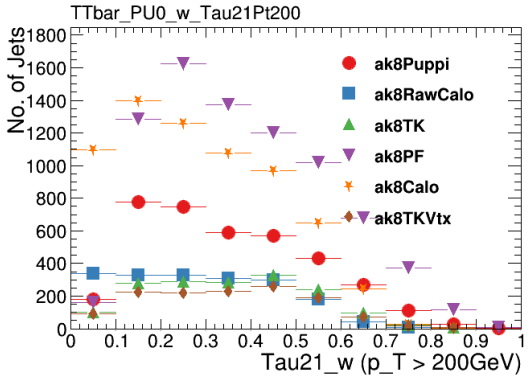


Figure 2.35: The τ_{21} matched with GenW distribution for $t\bar{t}$ PU0 with $p_T >$ 200GeV.

So, in conclusion NJettiness variable gives the good insight of fatjet and it is pileup sensitive. The τ_{21} variable working well for AK8PUPPI is good in performing well for W but not for top quark jet $t\bar{t}$ PU140.

2.4 Chapter summary

The response and resolution plots for p_T of reconstructed AK8 jets based on PUPPI and PF algorithms have been shown versus p_T as well as η of the jet. In principle, the PUPPI algorithm should improve the resolution of the jet in comparison to the PF algorithm in the high pileup scenario. But as it has been shown in this study there is no improvement in the resolution of PUPPI and which could be a unique characteristic of AK8 jets. This is being investigated further. The plots for the distribution of Njettiness variable have also been shown with these algorithms. It is has been found that NJettiness variable is an efficient tool for discriminating boosted object. Specifically, the τ_{21} variable is suitable for the AK8 jets built with PUPPI algorithm for identifying jets coming from W bososn in high p_T and high pileup scenario. The τ_{32} variable did not show a good performance for identifying boosted top jet and this is under further investigation.

Chapter 3

The e/γ Algorithm using Vivado high level synthesis (HLS) software

In experimental high energy physics the increasing popularity of Vivado HLS for trigger applications on FPGAs in CMS experment at LHC to check the performance and efficiency of hardware actually bridges the gap of keeping the consistency between trigger firmware and its corresponding c++ model. This chapter is devoted to investigation of HLS tools and testing of e/γ algorithm on vivado HLS for CMS experiment.

3.1 Basics of HLS

The phase II of HL-LHC will going to make very harsh environment in LHC, as proton-proton collisions per bunch crossing (pileup or PU) will be between 140 and 200 which means lots of data would be there but we only need to store interesting data(trigger comes into the picture for selection) and LHC collides proton bunches in every 25ns. So, our algorithms has to be very fast and efficient. The trigger system reduces the 40 MHz collision rate to a manageable data storage rate of 400 Hz and electronics of L1trigger do that. In HL-LHC, the trigger rates for physics objects will exceed the current trigger threshold. To make electronics more efficient CMS is using xilinx

FPGAs(field-programmable gate array)[1].

Xilinx HLS software- Vivado HLS transforms C-specification into a register transfer level (RTL) implementation that synthesizes into Xilinx FPGAs. RTL is a design abstraction which models a synchronous digital circuit in terms of how digital signal(data) between hardware registers and logical operations performed on those signals. The synchronous digital circuit consits two kinds of elements:registers and combinational logic. RTL abstraction used in hardware description language (HDLs) like vertilog and VHDL to create HL representation of a circuit. The Vivado High-Level Synthesis compiler enables C, C++ and SystemC programs to be directly targeted into Xilinx devices without the need to manually create RTL. Vivado HLS provides pragma directive which are used to improve the performance of the code, reduce latency[2], reduce the resources utilization for RTL code.

INITIATION INTERVAL(II):The function Initiation Interval (II) is the number of clock cycles before the function can accept new input data. If II=11, This means it takes 11 clock cycles before the function can initiate a new set of input reads and start to process the next set of input data. The time to perform one complete execution of a function is referred to as one transaction.

Here are list of some important HLS directives:

1. **Pipeline directive:**pragma reduces the initiation interval for a function or loop by allowing the concurrent execution of operations.

   - **syntax:** # pragma HLS pipeline II=$< int >$ enable_flush rewind

   - $II =< int >$**:**Specifies the desired initiation interval for the pipeline. Vivado HLS tries to meet this request. Based on data dependencies, the actual result might have a larger initiation interval. The default II is 1.

   - **enable_flush:**An optional keyword which implements a pipeline that will flush and empty if the data valid at the input of the pipeline goes inactive.

[1]A field-programmable gate array (FPGA) is an integrated circuit designed to be configured by a customer or a designer after manufacturing hence 'field-programmable'. It contain an array of programmable logic blocks, and a hierarchy of re configurable interconnects that allow the blocks to be 'wired together', like many logic gates that can be inter-wired in different configurations. FPGAs are used in hardware of Trigger L1

[2]The latency of the function is the number of clock cycles required for the function to compute all output values.

- **rewind:**An optional keyword that enables rewinding, or continuous loop pipelining with no pause between one loop iteration ending and the next iteration starting

2. **Unroll directive:**Unroll loops to create multiple independent operations rather than a single collection of operations. The UNROLL pragma transforms loops by creating multiples copies of the loop body in the RTL design, which allows some or all loop iterations to occur in parallel.

   □ **syntax:** #pragma HLS unroll factor=< N > region skip_exit_check

   - **factor=**< N >**:** Specifies a non-zero integer indicating that partial unrolling is requested. The loop body is repeated the specified number of times, and the iteration information is adjusted accordingly. If factor= is not specified, the loop is fully unrolled.
   - **region:**An optional keyword that unrolls all loops within the body (region) of the specified loop, without unrolling the enclosing loop itself.
   - **skip_exit_check:**An optional keyword that applies only if partial unrolling is specified with factor=.

3. **array_partition directive:**Partitions an array into smaller arrays or individual elements. This partitioning:

   (a) Results in RTL with multiple small memories or multiple registers instead of one large memory.

   (b) Effectively increases the amount of read and write ports for the storage.

   (c) Potentially improves the throughput of the design.

   (d) Requires more memory instances or registers.

   □ **syntax:** # pragma HLS arraypartition variable=<name> <type> factor=<int> dim=<int>

   - **variable=**< *int* >**:** A required argument that specifies the array variable to be partitioned.
   - < *type* >**:** Optionally specifies the partition type. The default type is complete. Other types are Block and Cyclic.

- **factor=**$< int >$**:** Specifies the number of smaller arrays that are to be created.
- **dim=**$< int >$**:** Specifies which dimension of a multi-dimensional array to partition. Specified as an integer from 0 to N, for an array with N dimensions.

Other pragma directives are given in [27].

3.2 From HLS to CTP7 card

The logic for a particular problem undergoes several stages before it is finally realized on an FPGA. This can be labelled as follows:

1. Coding for the problem in High Level Synthesis:Vivado HLS
2. Testing if the logic implemented is correct: Test Bench in Vivado HLS
3. Virtual synthesis of the IP to get an estimate of the time and resource utilization of the IP: Synthesis in Vivado HLS
4. To generate an IP: Export IP in Vivado HLS
5. To be implemented on the FPGA card: Generate bit map file in Vivado Tool
6. To implement physically on Virtex-7 FPGA sitting in a test crate at UWM: Use eagle25 login to manipulate the input file and download the output file.

The twiki page at [28] explains in detail how the entire process is carried out.

3.2.1 Vivado HLS workshop

Prof.Sridhara Dasu(UWM) conducted a workshop on Vivado HLS in august 2017, at TIFR explaining about how to generate a bit file of a HT code written by Sridhara[29]. Where we had to modify the HT code to implement a LUT (look up table) to see the 10-bit ET values and add them if they were above a certain set threshold for both, rgnET and hfET. After successfully installing hls vivado into our system and following the steps given in

twiki till synthesis design after editing ctp7_top.vhd file. This file contains all the port and component mapping for the inputs and outputs of your IP. For example, for MakeHT, all the 252 rgnET and 144 hfET input ports along with 1 HT output port and saving all the changes in ctp7_top.vhd file.Next, to generate bitstream, three important steps:

1. Run synthesis
2. Run Implementation
3. Genrate Bitstream

After successfully generating the bitsteam, a ctp7_top.bit file is generated in the impl1 folder. This file needs to be loaded on the FPGA and tested. To do this, we have been facilitated with eagle25, a Super User login which enables us to use the ZYNQ processor which has a linux platform running on it, which in turn communicates with the Virtex-7 FPGA.

Login command for superuser: su ctp7hls

Password: hls

To run our bit file on this FPGA, follow the steps given on Twiki page mentioned above. Do NOT rename the bit file as then it will give a Permission Denied error while copying it to the /tem/. folder on eagle25.

This workshop turns out as a essential preliminary study about Vivado HLS and FPGAs.

3.3 e/γ algorithm

The precise measurement electron energy is quite a difficult task in CMS experiment as electron looses most of its energy in tracker itself before reaching into ECAL due bremsstrahlung. So, its becomes essentially important to measure the energy of radiated photons in order to measure accurate energy of electron. The e/γ objects are detected by ECAL in CMS and in (η, ϕ) space. ECAL is made up of crystals and $17\eta * 4\phi$ towers = $17 * 4 * 5 * 5$ crystal. After, we had gone through the ClusterFinder code [30] and understanding algorithm which basically works like this:

1. Identify for every 5x5 crystal ('trigger tower'), the peak crystal position in eta-phi space.
    - Use energy-weighted position algorithm
2. Calculate the 3x3 cluster energy around the peak crystal.
    - For each tower in a calo-layer1 card $17\eta * 4\phi$ towers,results in:
        - Peak position in eta and phi
        - 5x5 tower sum
        - 3x3 cluster energy around the seed
3. Merge the 3x3 clusters with peak energy at the tower boundaries

The output of clusterfinder algorithm are peak position in $\eta x\phi$, 3x3 cluster energy and 5x5 tower energy sum. Cluster Finder Algorithm for CaloLayer1 card and synthesized using HLS Vivado and the HLS Vivado performance for Target device:Xilinx Virtex-7 690T and Target clock time:4.16 sec. And it results in with Latency of 53 clock cycle and estimated time 3.64 sec. Now, for e/γ algorithm we have to make cluster size 3x5 instead of 3x3, the extended region in ϕ direction is because to recover energy lost due to Bremsstrahlung.

So, making 3x5 cluster size, we further modify the clusterfinder code and calculating 2x5(Right) and 2x5(Left) cluster inside 3x5 cluster and get cluster energy of 2x5 by picking bigger energy from left or right. After implementing all the changes in the code we ran it on Vivado HLS with same configurations and the results are shown in Fig. 3.1 and Fig. 3.2. The latency of 50 clock cycle and estimated time 3.64 ns with resource utilisation of 36 percent.So we managed to bring latency down to 50cc by making a judicious use of pragma directives in the code.

After implementing the 3x5 cluster energy part, the next task was to calculate 9ηx15ϕ isolation region around the peak position and after implementing this in code, the output results was bit of unexpected as the resource utilisation shoots high as 121 percent and latency is also high 61cc as illustrated in Fig. 3.3 and Fig. 3.4.

So, in order to reduce the resource utilisation and latency we figured out the other smarter way to calculate isolation region this time 10ηx15ϕ around the 5x5 tower, by

Performance Estimates

Timing (ns)

Summary

Clock	Target	Estimated	Uncertainty
ap_clk	4.16	3.64	0.52

Figure 3.1: Timing of clusterfinder code after making 3x5 cluster

Latency (clock cycles)

Summary

Latency		Interval		
min	max	min	max	Type
50	50	6	6	function

Detail

Instance

Loop

Utilization Estimates

Summary

Name	BRAM_18K	DSP48E	FF	LUT
DSP	-	-	-	-
Expression	-	-	0	27491
FIFO	-	-	-	-
Instance	-	-	78880	84276
Memory	-	-	-	-
Multiplexer	-	-	-	32359
Register	-	-	66692	12000
Total	0	0	145572	156126
Available	2940	3600	866400	433200
Utilization (%)	0	0	16	36

Figure 3.2: Latency and resource utilization of clusterfinder code after making 3x5 cluster

```
================================================================
== Utilization Estimates
================================================================
* Summary:
+-----------------+---------+-------+--------+--------+
|       Name      | BRAM_18K| DSP48E|   FF   |   LUT  |
+-----------------+---------+-------+--------+--------+
|DSP              |        -|      -|       -|       -|
|Expression       |        -|      -|       0|   27492|
|FIFO             |        -|      -|       -|       -|
|Instance         |        -|      -|  319192|  431076|
|Memory           |        -|      -|       -|       -|
|Multiplexer      |        -|      -|       -|   32727|
|Register         |        -|      -|  104426|   36592|
+-----------------+---------+-------+--------+--------+
|Total            |        0|      0|  423618|  527887|
+-----------------+---------+-------+--------+--------+
|Available        |     2940|   3600|  866400|  433200|
+-----------------+---------+-------+--------+--------+
|Utilization (%)  |        0|      0|      48|     121|
+-----------------+---------+-------+--------+--------+
```

Figure 3.3: The resource utilization of clusterfinder code after making 9ηx15ϕ isolation region

```
== Vivado HLS Report for 'getClustersInCard'
================================================================
* Date:           Mon Feb 19 06:31:21 2018

* Version:        2016.4 (Build 1756540 on Mon Jan 23 19:31:01 MS
* Project:        clusterfinder2
* Solution:       solution1
* Product family: virtex7
* Target device:  xc7vx690tffg1927-2


================================================================
== Performance Estimates
================================================================
+ Timing (ns):
    * Summary:
    +--------+-------+----------+------------+
    |  Clock | Target| Estimated| Uncertainty|
    +--------+-------+----------+------------+
    |ap_clk  |   4.16|      3.64|        0.52|
    +--------+-------+----------+------------+

+ Latency (clock cycles):
    * Summary:
    +-----+-----+-----+-----+----------+
    |  Latency  |  Interval | Pipeline |
    | min | max | min | max |   Type   |
    +-----+-----+-----+-----+----------+
    |   61|   61|    6|    6| function |
    +-----+-----+-----+-----+----------+
```

Figure 3.4: Latency and estimated timing of clusterfinder code after making 9ηx15ϕ isolation region

extending the tower in $\eta - \phi$ space, the implementation of 10ηx15ϕ isolation region in the code is still in progress.

In conclusion, in e/γ algorithm, our next task is to calculate the 10ηx15ϕ isolation region and from isolation region we have to remove the 3x5 cluster energy around a peak position so that we can find isolated e/γ objects and simultaneously implementing it on Vivado HLS.

3.4 Chapter summary

The Vivado HLS firmware is a powerful tool for implementing complex algorithms in (FPGA)hardware. The implementation of e/γ algorithm has been help full in understanding the firmware and hardware performance in CMS.

Chapter 4

Conclusion and Outlook

The high Luminosity Large hadron Collider (HL-LHC) is a major upgrade which increases the chances for probing the sensitivity of beyond standard model physics with integrated luminosity of 3000 fb^{-1}. The study of production and decay of heavy particles like scalar Higgs boson and Top quark will lead to potential searches for new physics.

In this thesis, the study of response and resolution of fatjets of radius 0.8 (reconstructed from Anti-kt jet clustering algorithm-AK8) for semi-leptonic decay of top quark event in boosted regime has been described. The reconstructed AK8 PUUPI jets provide the best response in high pileup scenario. In terms of resolution also AK8 PUPPI jets performance got improved for 140 pileup as compared to 0 pileup case but PUPPI algorithm does not improve the resolution of the jet in high pileup case. This could be a feature of AK8 jets but has to be investigated further.The variation of the response and resolution has been studied with p_T and η of the jet.

The study of jet substructure using NJettiness variable shows that τ_{21} variable is efficient for jet coming from W boson. But the variable τ_{32} does not give good performance in identifying boosted top jet.

In HL-LHC regime, the study of $t\bar{t}$ production and deacy is very important in high pileup scenarios for search of new physics and for studying properties of Higgs boson also.

The second part of this thesis focuses on understanding Vivado HLS software tools and

e/γ algorithm implementation using this software in the FPGAs. The study has shown that by tuning the algorithm using the hardware resources can be optimized in FPGAs. The vivado HLS software helps in implementation of the algorithm in a user friendly way on hardware systems.

Bibliography

[1] Contardo, D and Klute, M and Mans, J and Silvestris, L and Butler, J, Technical Proposal for the Phase-II Upgrade of the CMS Detector, Geneva 2015.

[2] Griffiths, David J, *Introduction to elementary particles; 2nd rev. version*, Wiley, Physics textbook, New York, NY, 2008.

[3] Patrignani, C. and others, Review of Particle Physics,Particle Data Group, Chin. Phys., C40, 2016.

[4] xkcd, xkcd website, https://xkcd.com/1489/.

[5] CMS. CMS website. https://cms.cern/detector/.

[6] Lefevre, Christiane, LHC: the guide (English version), 2009.

[7] CMS Collaboration. CMS-Public twiki. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults

[8] Karimki, V and Mannelli, M and Siegrist, P and Breuker, H and Caner, A and Castaldi, R and Freudenreich, K and Hall, G and Horisberger, R and Huhtinen, M and Cattai, A., Technical Design Report CMS, CMS Collaboration, Geneva, 1997.

[9] Collaboration, CMS., The Phase-2 Upgrade of the CMS L1 Trigger Interim Technical Design Report, CERN-LHCC-2017-013. CMS-TDR-017, CERN-LHCC-2017-013, Geneva, Sep 2017.

[10] https://www.gk-eichtheorien.physik.uni-mainz.de/Dateien/Zeppenfeld-3.pdf

[11] Sjöstrand, Torbjörn and Mrenna, Stephen and Skands, Peter., A brief introduction to PYTHIA 8.1, 178, 11, Computer Physics Communications, 2008.

[12] Cacciari, Matteo and Salam, Gavin P and Soyez, Gregory., The anti-kt jet clustering algorithm, 2008, 04, Journal of High Energy Physics, IOP Publishing, 2008.

[13] Salam, Gavin P ., Towards jetography, The European Physical Journal C, Springer, 2010.

[14] Atkin, Ryan., *Review of jet reconstruction algorithms*, Journal of Physics: Conference Series, IOP Publishing, 2015.

[15] Coco, Victor and Soyez, Gregory and Rojo-Chacon, Juan and Sander, Christian and Delsart, Pierre-Antoine., Jets and jet algorithms, 2009.

[16] Rangel, Murilo., Experimental aspects of jet physics at LHC, arXiv preprint arXiv:1608.00057, 2016.

[17] Beaudette, Florian., The CMS Particle Flow Algorithm, arXiv preprint arXiv:1401.8155, 2014.

[18] CMS collaboration and CMS Collaboration and others, Particle-flow event reconstruction in CMS and performance for jets, taus and MET, CMS-PAS-PFT-09-001, 2009.

[19] CMS Collaboration. CMS-Public twiki. https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideParticleFlow

[20] CMS Collaboration and others., Particle flow and global event description in CMS, preparation.

[21] Aaltonen, T and Abazov, VM and Abbott, B and Acharya, BS and Adams, M and Adams, T and Alexeev, GD and Alkhazov, G and Alton, A and González, B Alvarez and others., *Combination of the top-quark mass measurements from the Tevatron collider*, Physical Review D, 86, APS, 2012.

[22] de fin de Grado, Trabajo., Top tagging at the LHC experiments with proton-proton collisions at s= 13TeV, 2015.

[23] Bertolini, Daniele and Harris, Philip and Low, Matthew and Tran, Nhan., ileup per particle identification, Journal of High Energy Physics, 2014, Springer, 2014.

[24] ryan, Vardan and Erbacher, Robin and Carrillo Montoya, Camilo Andres and Carvalho, Wagner and Górski, Maciej and Kotlinski, Danek and Anderson, Jacob and Jez, Pavel and Ujvari, Balazs and Ozturk, Sertac and others., Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV, JINST, 12, CERN-PH-EP-2015-305, P02014, CMS-JME-13-004-003, 2016.

[25] Stewart, Iain W and Tackmann, Frank J and Waalewijn, Wouter J., N jettiness: an inclusive event shape to veto jets, 105, 9, Physical review letters, APS, 2010.

[26] Usai, Emanuele and Collaborations, CMS and others., Boosted top: experimental tools overview, arXiv preprint arXiv:1501.0090, 2015.

[27] xilinx xilinx website. https://www.xilinx.com/html_docs/xilinx2017_1/sdsoc_doc/topics/pragmas/concept-Intro_to_HLS_pragmas.html

[28] CMS Collaboration. CMS-Public twiki.
https://twiki.cern.ch/twiki/bin/viewauth/CMS/L1TriggerPhase2HLSProjects

[29] github github website.
https://github.com/SridharaDasu/VivadoHLSProjects/tree/master/HT

[30] github github website.
https://github.com/SridharaDasu/VivadoHLSProjects/tree/master/ClusterFinder

Appendices

Appendix A

The reco Jet mass and transverse momentum distribution

The reco jet mass distribution for QCD PU0 and PU140 in Fig. A.1 and Fig. A.2. The reco jet p_T distribution for QCD PU0 and PU140 in Fig. A.3 and Fig. A.4.

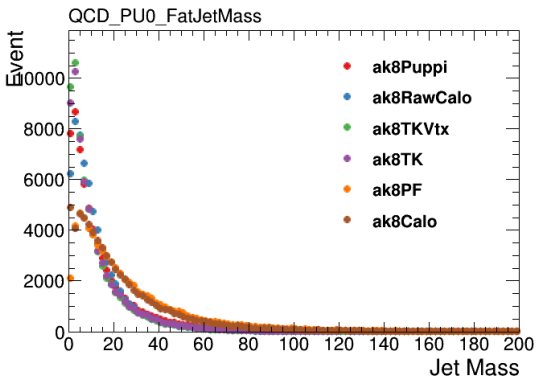


Figure A.1: The reco Jet mass distribution for QCD PU0

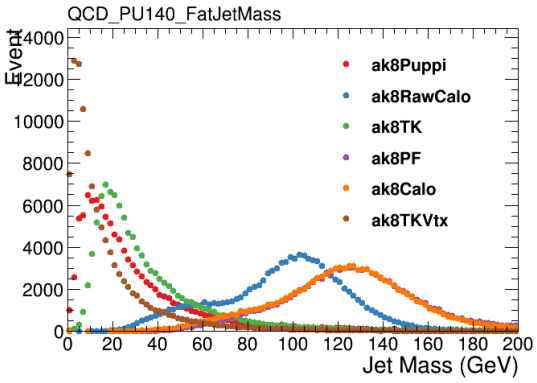


Figure A.2: The reco Jet mass distribution for QCD PU140

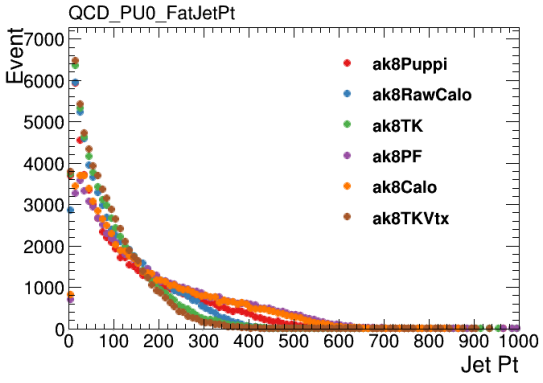


Figure A.3: The reco Jet p_T distribution for QCD PU0

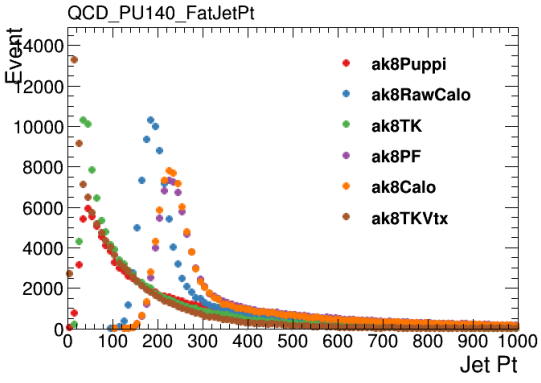


Figure A.4: The reco Jet p_T distribution for QCD PU140

Appendix B

The response and resolution plots for AK4 jets

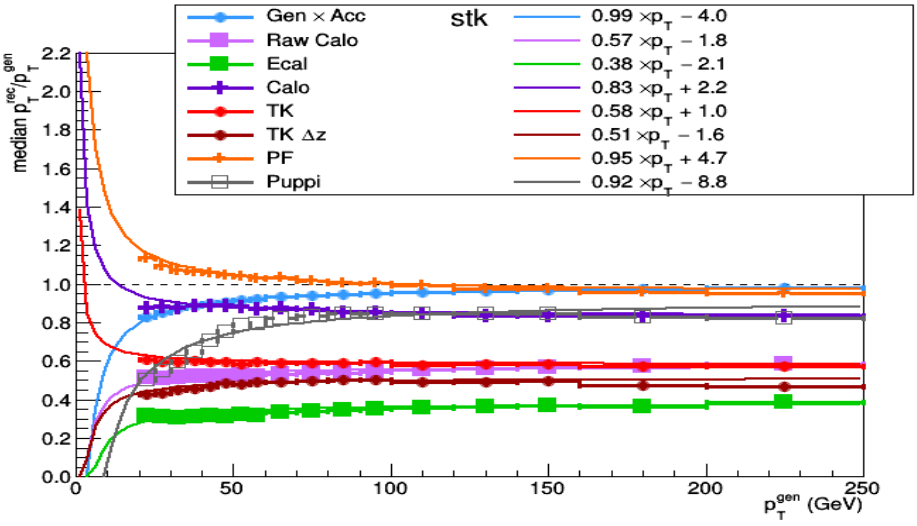


Figure B.1: Simulated jet p_T response versus generated particle level jet p_T for jet radius R= 0.4

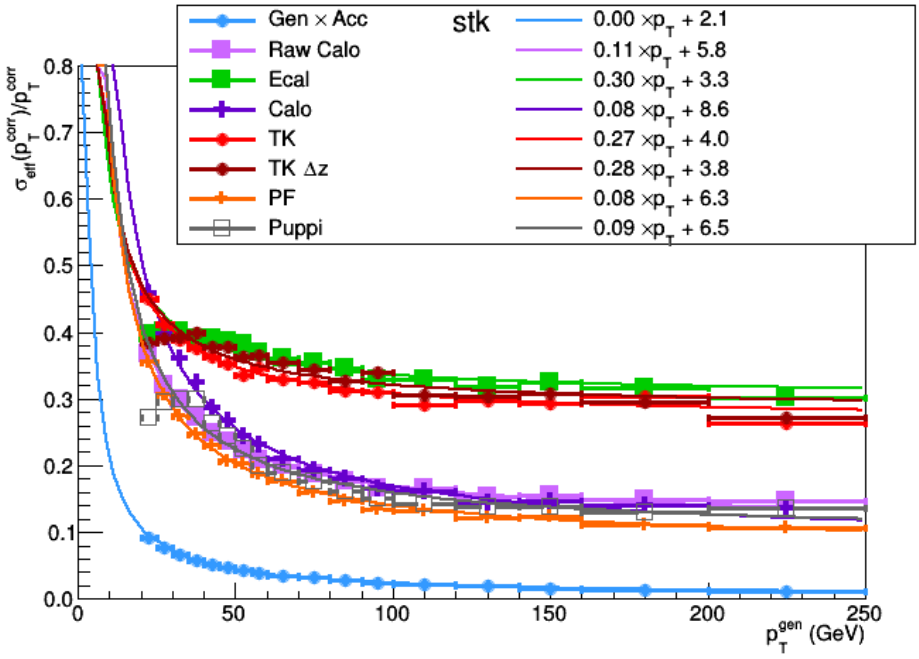


Figure B.2: Jet p_T resolution versus generated particle level jet p_T for jet radius R= 0.4