Background estimation of Mono-Higgs in $b\bar{b}$ final state using 2018 data of the CMS detector at the LHC, CERN

A project submitted to the School of Physics, University of Hyderabad in partial fulfillment for the award of the degree of

Master of Science in Physics

By

Prayag Yadav

Reg. No. - 19IPMP03

School Of Physics
University Of Hyderabad
Hyderabad 500046
May-2024

DECLARATION

I hereby declare that this project entitled **'Background estimation of Mono-Higgs in $b\bar{b}$ final state using 2018 data of the CMS detector at the LHC, CERN'** is carried out by me at the School of Physics, University of Hyderabad, under the supervision of Dr. Bhawna Gomber.

No part of this project has been previously submitted for a degree or diploma or any other qualification at this University or any other.

Prayag Yadav
19IPMP03

CERTIFICATE

This is to certify that this project entitled **'Background estimation of Mono-Higgs in $b\bar{b}$ final state using 2018 data of the CMS detector at the LHC, CERN'** is carried out by Prayag Yadav (19IPMP03) under my supervision at the School of Physics, University of Hyderabad, in partial fulfillment of the requirements for the award of the degree of Master of Science in Physics. No part of this project has been previously submitted for a degree or diploma, or any other qualification at this university or any other.

Dr. Bhawna Gomber
Project Supervisor

Dean School of Physics

ACKNOWLEDGEMENTS

I extend my deepest gratitude to my supervisor, Dr. Bhawna Gomber for supporting me at every step of the work. I thank Bisnupriya Sahu, PhD student, to help me with technical assistance whenever needed. I would like to thank Shivani Lomte, PhD student at University of Wisconsin, for fruitful discussions about the project.

I would like to thank School of Physics, University of Hyderabad and University of Wisconsin to provide access to the necessary infrastructure to complete this project.

I am immensely fortunate to have the support of my loving family and friends. Their support made a tangible effect on the completion of this project.

Dedicated to my mother

Abstract

Dark matter is an elusive state of matter whose presence has verifiable astronomical and cosmological implications. Super-symmetric models attempt to include dark matter and provide processes involving interaction between dark matter and standard model particles. One of the final states of such an interaction is the production of a Higgs boson along with large missing transverse momentum. The Higgs boson in turn decays into two bottom quarks. This analysis involves the identification of the signal (large $p_T^{miss} + H \rightarrow b\bar{b}$) from various standard model backgrounds which are estimated by defining control regions. The two major backgrounds for this analysis are top pairs ($t\bar{t}$) and $Z \rightarrow \nu\bar{\nu}$. These two backgrounds are estimated by four different control regions: single muon, single electron, double muon, and double electron. This thesis is a study of the single electron control region which involves top quarks and electrons .

Contents

List of Figures

List of Tables

1 Introduction

1.1 Dark matter

Darkness there was at first, by darkness hidden; the universe was undistinguished like water. That which, becoming, by the void was covered; that One by the force of heat came into being.
- Rigveda

The universe, as we know today, is composed of 5% ordinary matter, 27% dark matter and 68% dark energy[6]. Ordinary matter consists of all the visible matter - the galaxies, stars, planets, gases and radiation. Dark matter is the invisible form of matter whose effects can only be seen via gravity. Dark energy is another unknown form of energy required to account for the accelerated expansion of the universe.

Dark matter is an elusive state of matter necessary to account for the large scale structure of the present cosmos. Various astronomical evidences and cosmological evidences support the existence of it. Astronomical evidences include rotational curves of galaxies[7] and gravitational lensing data [8]. Cosmological evidences include the evidence from anisotropy of cosmic microwave background[9].

1.2 Ways to search for dark matter

There are mainly three modes of efforts to search for dark matter:

1. **Direct search:** Direct search assumes direct interaction of standard model particles with dark matter. Experiments like XENON [10], IceCube [11] and DAMA/LIBRA [12] detect scintillation light from nuclei recoiling from dark matter collisions.

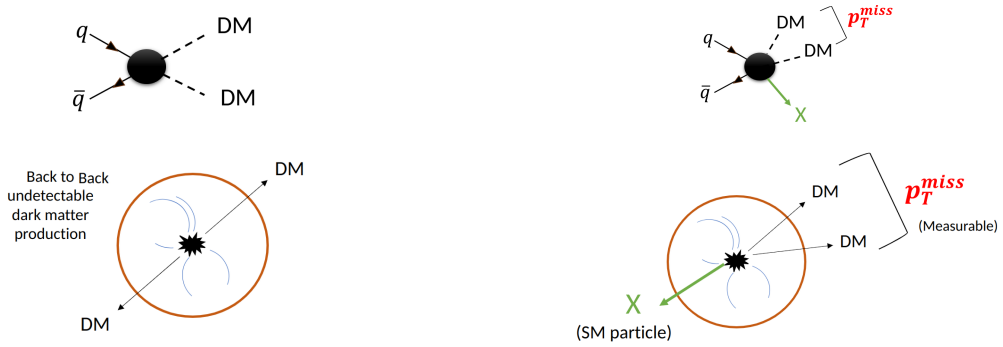
2. **Indirect search:**[13] Annihilation of dark matter particle pairs (denoted as $\chi\bar{\chi}$) may lead to an anomalous excess in the cosmic ray radiation from galaxy halos. Searches for such signatures above the known astronomical backgrounds come under indirect detection of dark matter.

3. **Collider search:**[14] If Dark matter particles can be produced in particle colliders, they are produced as a result of interaction between Standard Model particles and Dark Matter particles. Searches probing for such signatures are called collider searches.

1.3 Mono-X searches

Particle colliders like the Large Hadron Collider(LHC)[15] can collide protons to a center of mass energy of $\sqrt{s} = 13.6\ TeV$. The final state particles produced as result of such collision, tend to travel in the transverse plane of the detectors. Interesting topologies arise depending upon the number of particles in the final state.

Consider a process in which two quarks produce a DM particle pair (fig. 1a). To conserve the momentum in the center of mass frame, the two dark matter particles are produced back to back. This type of topology is of little use because it cannot be detected in a search.



(a) Back-to-back topology

(b) Mono-X topology

Figure 1: Advantage of the Mono-X topology for DM searches

Now, consider a process in which two quarks produce a DM particle pair and a standard model particle 'X' (fig. 1b). In such a topology, the SM particle X recoils from the dark matter pair which goes undetected in the detectors. Measuring the recoil of X gives an idea about the DM particles. This is done by calculating the "missing transverse momentum p_T^{miss}" of the event which is found by applying the conservation of momentum in the transverse plane.

Among various Mono-X searches, Mono-Higgs searches have unique advantages:

1. It involves no initial state radiation and hence a cleaner final state for analysis
2. Since the Higgs boson is directly involved with the mechanism to provide mass to particles, it is more closely connected to DM production
3. The final state has a high missing transverse energy tail which helps to separate the signal from background.

The highest branching ratio of Higgs decay is through two bottom quarks[16]:

$$H \to b\bar{b} \quad (58\%) \tag{1}$$

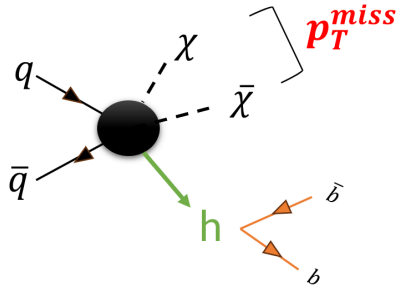


Figure 2: The Mono-Higgs topology

In this analysis, we are interested in the large p_T^{miss} + $H \to b\bar{b}$ final state.

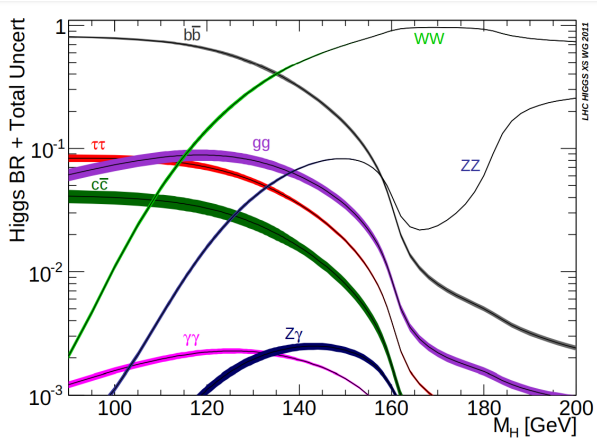


Figure 3: The branching ratio of Higgs decay for various low mass ranges[1]

2 The experimental setup

2.1 Large Hadron Collider

The Large Hadron Collider (LHC) is the world's largest particle accelerator located near Geneva at the French-Swiss border. It is a 27 km long ring of superconducting magnets which accelerate H^+ ions (protons) close to the speed of light. The journey of a proton starts from a small cylinder containing H_2 gas. Hydrogen is ionized and protons thus produced are accelerated through a series of smaller rings before reaching the LHC. At LHC, the protons travel in bunches within two rings, running opposite to each other. Each bunch is only 25 nanosecond apart in time. They reach an energy of 6.8 TeV each.

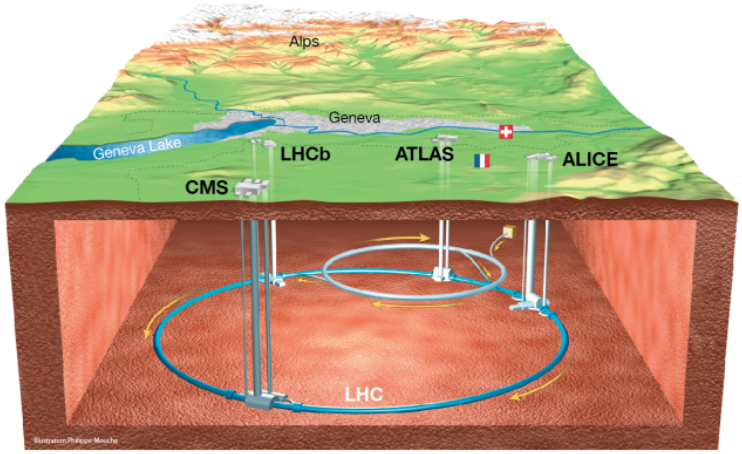


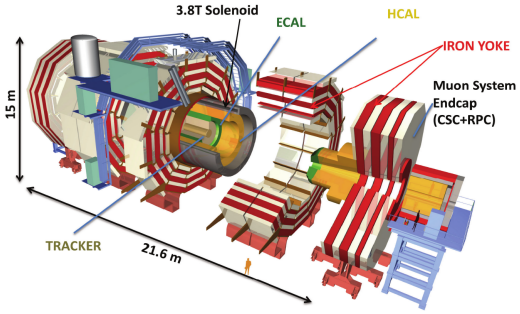
Figure 4: A schematic depiction of LHC site and the associated experiments [17]

The two LHC rings meet at 4 different locations in the LHC ring where detectors are placed to detect the products of the proton-proton collisions (fig 4). The experiments at the four collision points are CMS, LHCb, ATLAS and ALICE. CMS and ATLAS are general

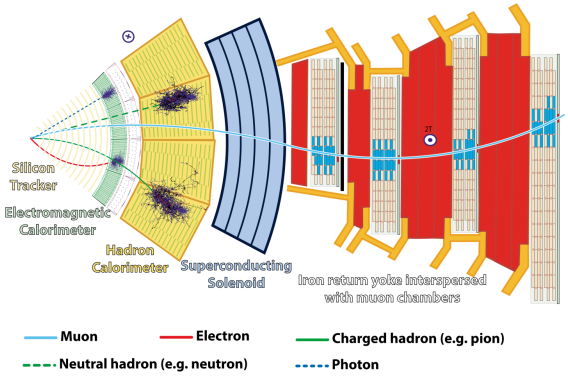
purpose detectors while LHCb is involved in B-physics and ALICE is involved in heavy-ion collisions.

2.2 Compact Muon Solenoid

The Compact Muon Solenoid (CMS) is a general purpose detector located 100m underground at Cessy, near Geneva (also called point 5). It uses a powerful solenoid electromagnet at 3.8 Tesla to measure the momentum and charge of particles. Such a strong electromagnet allowed for a compact design of the whole detector (hence the name). The detector is built like an onion with layers upon layers of different particle detectors.



(a) A schematic representation of the CMS detector[18]



(b) The various layers in CMS help identify different particles

Figure 5: The CMS detector

The CMS detector is designed to produce good:

1. muon identification and momentum resolution
2. charged particle momentum resolution and reconstruction efficiency
3. electromagnetic energy resolution, good di-photon and di-electron mass resolution
4. missing-transverse-energy and dijet-mass resolution.

All this is accomplished by the harmonious working of the various sub-detectors in CMS. The main subdetector systems at CMS are:

2.2.1 Inner Tracker System

The tracker is the innermost sub-detector of the CMS. It is designed to detect precise and efficient trajectories of all charged particles. The inner tracking system has two parts - Pixel detector and Silicon Strip detectors. Pixel detectors are made up of 124 million silicon pixels responsible for good primary vertex resolution (upto a few tens of μm). Surrounding the pixel detectors are multiple layers of silicon strips which track the trajectory of charged particles.

2.2.2 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is a hermetic, homogeneous detector made up of 61,200 Lead Tungstate ($PbWO_4$) crystals mounted in a central barrel part accompanied by 7324 crystals in each of the two end caps. A pre-shower detector is installed at the face of the two end caps. The ECAL measure the energy of electrons and photons. The choice of the scintillating crystals ($PbWO_4$) has been made considering various factors like short scintillation duration, high photoelectric absorption coefficient and short radiation length.[19]

2.2.3 Hadronic Calorimeter

The hadronic calorimeter (HCAL) quantifies the energies of hadronic particles and jets of particles arising from quarks or gluons. It consist of hadron calorimeter barrel and hadron end caps which are located outside the ECAL. HCAL is made up of alternating layers of absorbing and scintillating material which helps in producing a hadronic shower.

2.2.4 Muon Systems

Muons are almost 200 times heavier than electrons and can pass through several meters of iron without interacting. So aptly, muon detectors are placed as the outermost layer of the CMS detector. Muon chambers are made up of drift tubes, cathode strip chambers and resistive plate chambers. They are interleaved between return iron yoke of the solenoid and experience different intensities of magnetic fields in different parts of the iron yoke. This, along with tracker information helps to determine Muon's momentum accurately.

2.3 Data Acquisition

CMS has a high rate of data production and it would be impractical to save all of it. Therefore choosing useful events from the incoming data at real time is of paramount importance. This task is done by the Trigger System. The trigger system is designed to choose the best physics relevant events in real time. The trigger system consists of two subunits: Level 1 Trigger and High Level Trigger.

2.3.1 L1 Trigger

The Level 1 trigger is a hardware based processing system which employs the use of programmable electronics like FPGAs extensively. The L1 trigger reduces the input crossing rate of 40MHz to an output rate of 100KHz[20]. It takes into account, various physics criteria to decide keep an event or not.

2.3.2 HLT

High Level Trigger (HLT) is a software based processing system which employs commercial microprocessor-based compute farms for trigger. The trigger farm is composed of about 2000 PCs. HLT takes information from L1 at 100 kHz and makes selections to produce a final rate of about 800Hz. HLT is much slower than L1 and has to account for possible failures

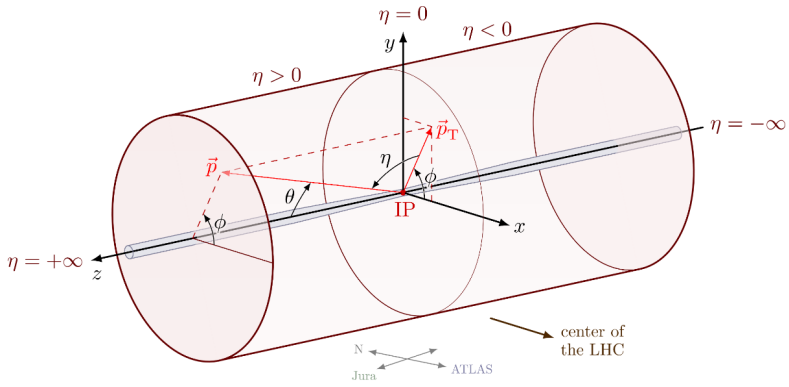


Figure 6: Schematic representation of the various coordinate systems at CMS[21]

of computing nodes. HLT builds one full event from the event fragments it gets as input. It is responsible for jet reconstruction, object identification, b-tagging etc.

2.4 The coordinate system

The CMS experiment is mostly cylindrically symmetric. We can construct a coordinate system with the beam direction as z-axis, the azimuthal angle ϕ traversing the cross section of the detector and the polar angle θ. Note that the knowledge of the radius ρ is already known from the construction plan of the detector. Therefore any 3-momentum vector can be described as

$$(p_T, \theta, \phi) \tag{2}$$

where p_T is the transverse momentum of the particle.

Generally, it is advantageous to use a Lorentz-invariant quantity instead of θ. Such a quantity is the pseudo-rapidity η :

$$\eta = -ln\left(tan\left(\frac{\theta}{2}\right)\right) \tag{3}$$

$$\theta = 0 \;\Rightarrow\; \eta = \infty \tag{4}$$

$$\theta = \frac{\pi}{2} \;\Rightarrow\; \eta = 0 \tag{5}$$

$$\theta = \pi \;\Rightarrow\; \eta = -\infty \tag{6}$$

Therefore any 3-momentum vector can be described as

$$(p_T, \eta, \phi) \tag{7}$$

Figure 6 represents the various coordinate systems at CMS. The interaction point(IP) of the pp-collision is the center of the coordinate systems. The y-axis is vertical to ground level and x-axis is horizontal to ground level pointing towards the center of LHC ring. The z-axis points in the beam direction.

3 Analysis strategy

The final state signature is large missing transverse energy along with a Higgs boson decaying to a bottom quark pair.

$$Final\ State:\ Large\ p_T^{miss} + H(\rightarrow b\bar{b}) \tag{8}$$

3.1 Backgrounds

A certain number of SM backgrounds mimic the final state of this analysis. The major backgrounds for this final state are:

1. $t\bar{t}$ decay
2. $Z(\rightarrow \nu\bar{\nu})$ + jets
3. Single top quarks
4. $W(\rightarrow l\nu)$ + jets
5. Drell-Yan(DY) + jets
6. The production of the single top quark in association with W boson (tW)
7. Diboson (WW, WZ, ZZ)
8. The associated production of a Higgs Boson with vector bosons (WH and ZH)

Among these, the most significant backgrounds are the $t\bar{t}$ and $Z(\rightarrow \nu\bar{\nu})\ +\ jets$. They are estimated by data driven methods by constructing control regions (see 3.2). Other backgrounds are estimated by Monte-Carlo simulations.

It is commonplace to divide the search for a specific signature into non-intersecting analysis regions. These regions are called signal region and control regions.

3.2 Signal region and control region

Signal Region (SR) is the region with event selections that allow the final state to pass and removes any other scenarios. For example, the final state has no leptons; so leptons are vetoed (events with at least one lepton are removed) from the signal region.

Control Region (CR) is the region with event selections that allow only those events which are characteristics of a SM background process. The final state is devoid of any lepton. There are SM processes which mimic the final state. For example, consider the process in which a top-antitop pair decays semi-leptonically to produce a pair of b-quarks, two additional quarks, a lepton(electron or muon) and a neutrino. The CR accepts two b-quarks, additional jets and exactly one lepton.

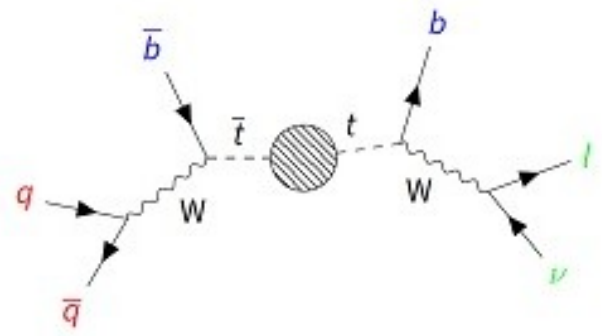


Figure 7: Semi-leptonic decay of $t\bar{t}$[22]

3.3 Control Region classification

There are two types of CRs in the analysis: the top CR and the Z+jets CR. The top CR is characterized by one lepton in the final state. The Z+jets CR is characterized by two leptons(lepton and anti-lepton of the same kind) in the final state. Depending upon the type of lepton - electron or muon , we have $2 \times 2 = 4$ CRs : Top(e) , Top(μ), Z(ee)+jets and Z($\mu\mu$)+jets .

The jets produced from the b-quarks have an angular separation which depends upon the boost of the initial Higgs boson. Based on the angular separation we can further divide the various CRs into boosted and resolved subcategories (see 4.1.2). That means we have 4 subdivisions for Top CR (fig. 8)and 4 subdivisions for Z+jets CR.

In this thesis, the **Boosted Top(e) CR** is discussed and results associated with it are presented.

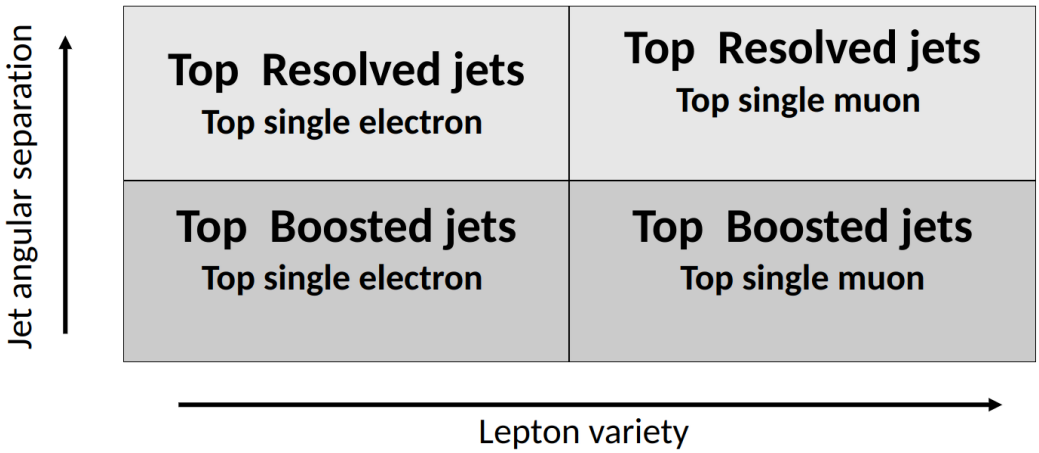


Figure 8: The subdivisions of the Top Control Region

3.4 The COFFEA Analysis Framework

With the advent of HEP-relevant packages like Uproot[23] and Awkward Array[24], Python[25] has gained much traction in high energy physics in recent years. The bulk of the analysis is done using Python as the programming language. Utilizing the native 'jagged' structure of the data stored in root[26] files, frameworks have been created which enable fast and reliable performance comparable to C++ analyses.

3.4.1 NanoAOD

The data stored at CERN is saved in root files designed specially for HEP. The raw data is raw detector readout data store as is and never deleted. AOD (Analysis Object Data) files contain all the information needed for an analysis. It contains data about the various analysis objects like electrons, photons, jets etc. MiniAOD is a more efficient format of data storage which stores a subset of AOD data relevant for 80% of physics analyses and in the process reduces disk and i/o resource requirements significantly[27].

NanoAOD is yet another scheme of root files that achieves 20 times data reduction from MiniAOD by retaining only high level information about physics objects and reducing the number of significant digits stored in variables[28]. The data format used in this analysis is NanoAOD because it integrates well with a Python-based analysis.

3.4.2 Columnar analysis

Traditional analysis with C++ processes data in an event by event basis. The is in contrast with the way data is saved in a root file. Data is saved within branches in a columnar fashion in a root file. Event by event processing first unpacks the columnar structure and runs computation in events loops.

Modern CPUs and GPUs are designed to perform highly vectorized calculations faster. A columnar way of analysis utilizes this hardware advantage in a way that is native to its syntax. A columnar analysis also opens up ways to optimize scalability of the analysis and enables parallel processing.

3.4.3 COFFEA and parallel processing

Columnar Object Framework For Effective Analysis (COFFEA) [29] is a Python framework that combines the various HEP related packages under one framework. It provides an array-based syntax to handle event data which is not so alien to NumPy syntax. By using modern big-data tools like Dask[30], the scalable codes written using COFFEA achieve great parallel processing capabilities.

COFFEA is used as the primary analysis tool in this thesis. NanoAOD root files are processed by creating a 'COFFEA processor' and submitted in large compute clusters at the University of Wisconsin, Madison[31]. The processed data is store in '.coffea' files. Plots and insights are generated using mplhep[32] and Matplotlib[33] using the output files.

4 Analysis objects

4.1 Jets and b-tagging

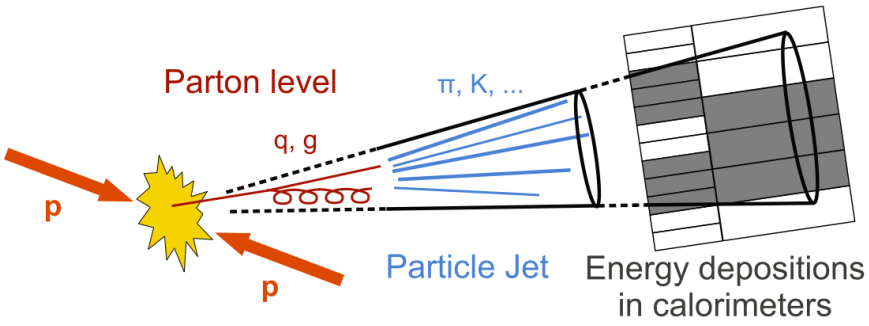


Figure 9: A cartoon of the formation of jets at CMS [34]

Due to color-confinement isolated quarks or gluon can't be observed experimentally. They undergo a process of hadronization, producing color-neutral hadrons in a collimated spray of particles. This spray of particles is called a particle jet. Unlike the signature of electrons or photons, jets deposit their energy over a larger area on the calorimeters. By measuring the area of energy deposition, total energy of the deposition, information about vertex position from tracker and several other parameters, jets are reconstructed and their flavor is determined. This is done by machine learning algorithms at the L1 Trigger level and further at the HLT level.

Jets reconstructed through anti-KT algorithm (see 4.1.1) and particle flow criteria are chosen for this analysis. Depending upon the angular width ($\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$) two types of jets are obtained from the anti-k_T algorithm - ak4 jets ($\Delta R = 0.4$) and ak8 jets ($\Delta R = 0.8$).

4.1.1 k_T and anti-k_T algorithms

Consider a set of particles which may or maybe not be fragments from a jet. Denoting the transverse momentum of each of these particles as k_T, defining the desired cone size R, the k_T algorithm follows these steps:

1. For each particle 'i' calculate d_{ij} and d_{iB} for each ij pair :

$$d_{ij} = min(k_{Ti}^2, k_{Tj}^2)\frac{\Delta R_{ij}^2}{R^2} \tag{9}$$

$$d_{iB} = k_{Ti}^2 \tag{10}$$

2. Find the minimum of d_{ij} and d_{iB}.

3. If d_{ij} pair is minimum then recombine the pair (add the Lorentz 4-vectors).

4. If d_{iB} is minimum then call 'i' a jet and remove it from the list of particles.

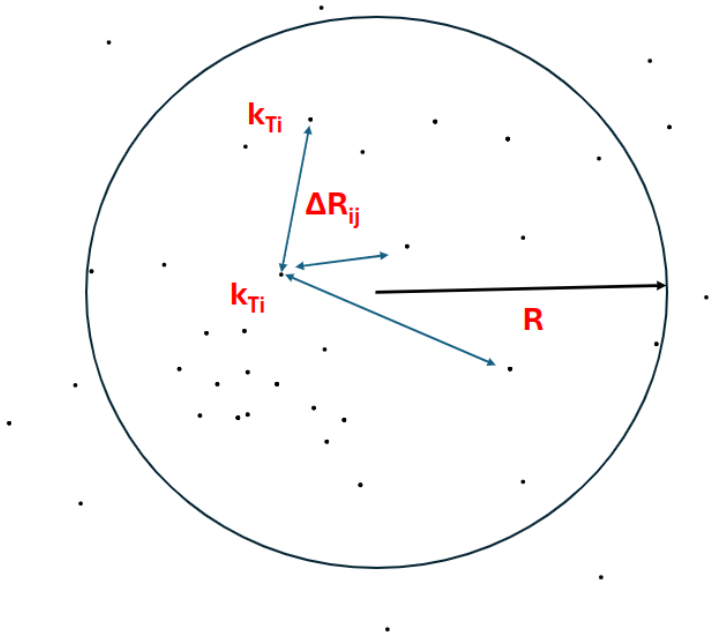


Figure 10: A schematic representation of the hits on a calorimeter matrix. Each black dot denotes a particle hit.

5. Repeat from step 1 until no particles are left.

The anti-k_T clustering algorithm is similar to the k_T algorithm, with a different d_{ij}:

$$d_{ij} = max(\frac{1}{k_{Ti}^2}, \frac{1}{k_{Tj}^2})\frac{\Delta R_{ij}^2}{R^2} \tag{11}$$

4.1.2 Resolved and Boosted scenario

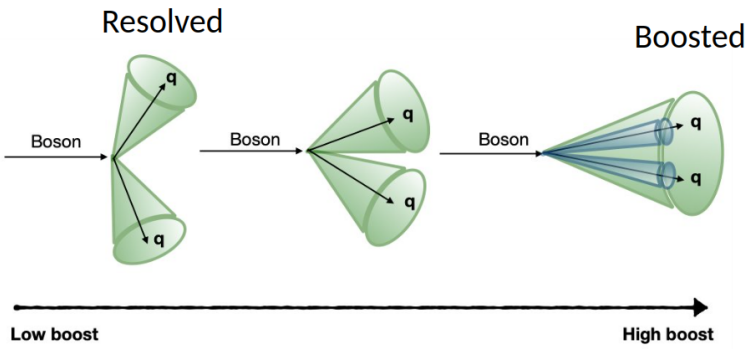


Figure 11: Cartoon depicting the angular separtion of jets vs their boost

Depending upon the boost of the initial particle, the subsequent ak4 jets produced may either be well separated (resolved) or starting to overlap (boosted). To account for both the

scenarios two different sizes of jets are used - the usual ak4 jets for resolved case and a larger ak8 jet (or FatJet) for the boosted case (fig. 11).

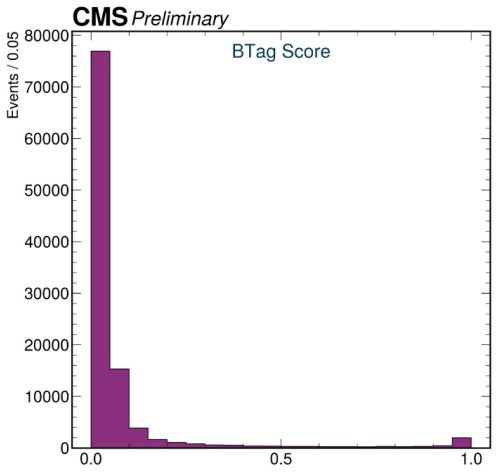
4.1.3 b-Tagging

The flavor of jets are determined by tagging algorithms. b-Tagging algorithms use machine learning and deep neural networks to assign a score between 0 to 1 to each jet. 1 corresponds to 100% confidence of a jet being a b-jet and 0 corresponds to 100% confidence of a jet not being a b-jet. Figure 12a shows the distribution of the b-Tag score in a dataset of 1 million events in the MET dataset from 2018UL.

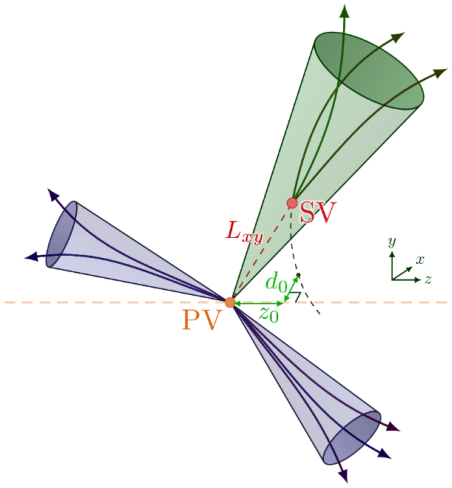
b-Jets have unique characteristics which help in it's identification:

1. The hadrons created from b-quarks have enough lifetime to travel a certain distance before decaying. This manifests as the b-jet having a displaced secondary vertex from the primary vertex at which the b-quark was created.

2. b-quark is heavy and therefore it's jets have high transverse momenta, have wider cones and have higher particle multiplicity than other jets.



(a) b-Tag score of ak4 jets for 1 million events from 2018UL MET dataset.



(b) The b-jet (green) has a primary vertex and a secondary vertex separated apart.[35]

Figure 12: b-Tagging of jets

The b-Tagging algorithm used in this analysis is called `bTagDeepFlavB`. The signal efficiency can be tuned by requiring the value of the b-Tag score to be greater than a pre-defined 'weight-parameter'. Accordingly, the b-jets can be categorized as tight, medium and loose.

4.2 Missing transverse energy

The missing transverse energy(MET or p_T^{miss}) is defined as the negative vector sum of tranverse momentum of all the particles in an event:

$$p_T^{miss} = -\sum_i (p_T)_i \tag{12}$$

This holds from the conservation of momentum in the transverse plane.

p_T^{miss} contains information about all the invisible particles in the detector: Neutrinos, Anti-Neutrinos and proposed hypothetical dark matter particles.

4.3 Recoil

In the control regions, the role of p_T^{miss} is performed by Recoil. Recoil is defined as the 2D vector sum of p_T^{miss} and the p_T of the lepton used as a tag for the CR.

$$\vec{U} = \vec{p_T}^{miss} + \vec{p_T}^{e \text{ or } \mu} \tag{13}$$

4.4 Photons, electrons, Muons, and Taus

Photons and electrons are absorbed in the electromagnetic calorimeter and have unique energy deposition patterns (for example, the Molière radius of the photon in the $PbWO_4$ crystal is narrower than that of the electron). Distinction between electron and photon candidates are made with a cut-based selections and other algorithmic selections. Signal efficiency is also tuned with respect to various parameters and tight,medium and loose electron/photons are defined.

Muons and Taus are also identified and isolated from other objects using ΔR criteria.

5 Event selections and object selections

5.1 Datasets

Dataset Name	Events selected for
`/EGamma/NANOAOD/Run2018[A-D]/UL2018_MiniAODv2_NanoAODv9-v1/*`	Single-Electron CR 2018

Table 1: The datasets used in the analysis

Table 1 gives the dataset used in the analysis. Apart from this, a long list of Monte-Carlo simulated samples were used. The number of events in each of these datasets is in the order of few hundreds of million.

5.2 Event selection criteria

Various cuts or selections were applied to select the best events for analysis. These selections are shown in table 2.

Selection order	**Event Selection**	**Description**
0	$No\ cuts$	Raw
1	$MET - Filters$	To remove MET noise
2	$Electron\ Trigger$	Choose Single-Electron events
3	$N(FatJet) = 1$	One ak8 Jet only
4	$N(IsoAddJet) \leq 2$	Two additional Jets
5	$N(IsoLoosebTagJet) = 1$	One loose b-Tagged Jet
6	$N(e) = 1$ & $N(\mu) = 0$	One Electron and no Muon
7	$p_T^{miss} > 50\ GeV$	To remove QCD backgrounds
8	$Recoil > 250\ GeV$	Final state requirement for boosted category
9	$N(\tau) = 0$	Tau veto
10	$N(\gamma) = 0$	Photon veto
11	$HEMveto$	Removing HEM affected events (detector issue in 2018)

Table 2: Event selections for the boosted Top electron control region

5.2.1 Trigger

Trigger Name	**Purpose**
`HLT_Ele32_WPTight_Gsf`	To select single electrons with $p_T > 32\ GeV$ in 2018 dataset

Table 3: The triggers used in the analysis

An HLT Trigger is used to select single electron events in the initial cuts itself. This has the added advantage of reducing the computing time required for the analysis.

5.2.2 MET filters

MET filters reduce detector noise and error. The list of MET filters used in the analysis are given in table 4.

Filter Name	**Filter Name**
`goodVertices`	`BadPFMuonFilter`
`globalSuperTightHalo2016Filter`	`BadPFMuonDzFilter`
`HBHENoiseFilter`	`eeBadScFilter`
`HBHENoiseIsoFilter`	`ecalBadCalibFilter`
`EcalDeadCellTriggerPrimitiveFilter`	

Table 4: The MET-filters used in the analysis

5.2.3 Corrections and Scale Factors

To account for different efficiencies of algorithms in Data events and Monte-Carlo simulated events, appropriate scale factors are applied on histograms produced from MC events. Some of these scale factors(SFs) and corrections are Electron-Trigger SF, b-Tagging SF, Pile-Up reweight SF, Jet Energy Calibration and Jet Energy Resolution, L1 prefiring and Top reweighting [36]. Each of these also carry an uncertainty with them.

In 2018, power supply of two HCAL modules in range η : -3.0 to -1.3 and ϕ : -1.57 to -0.87 was compromised. Therefore, in this analysis we remove any event which has at least one jet or fatjet lying in this range.This is called HEM veto.

5.3 Object selections

5.3.1 Fatjet selection

Fatjets (ak8 jets) were selected using a series of selection shown in table 5. Since fatjets contain two overlapping ak4 jets, their mass should be close to the Higgs mass. This is asserted by choosing the mass range of fatjets between 70 GeV to 150 GeV.

Selection	Description
$p_T > 30\ GeV$	To remove low pt jet noise
$\|\eta\| < 2.5$	Extent of calorimeters
$jetID \geq 2$	Tight ak8jets
$70\ GeV < M_{softdrop} < 150\ GeV$	Mass around Higgs mass
$\Delta R(Fatjet, e) > 0.4$ & $\Delta R(Fatjet, \mu) > 0.4$	Jets isolated from Electrons and Muons

Table 5: Fatjet(ak8 jet) selections in the analysis

5.3.2 Jet selection

Selections for additional jets is given in table 6. Unlike fatjets, there is no selection of the mass of jets.

Selection	Description
$p_T > 30\ GeV$	To remove low pt jet noise
$\|\eta\| < 2.5$	Extent of calorimeters
$jetID \geq 2$	Tight ak4jets
$\Delta R(jet, e) > 0.4$ & $\Delta R(jet, \mu) > 0.4$	Jets isolated from Electrons and Muons

Table 6: Jet(ak4 jet) selections in the analysis

5.3.3 Lepton selections

Selection	Description
$p_T > 40\ GeV$	To be about the trigger's turn on (32 GeV)
$\|\eta\| < 2.5$	Extent of calorimeters
$cutBased \geq 4$	Tight Electrons
gapcuts	Remove Electrons from ECAL gap

Table 7: Electron selections in the analysis[2]

Selection	Description
$p_T > 15\ GeV$	Remove low p_T noise
$\|\eta\| < 2.4$	Extent of Muon chambers
`isPFcand`	Is a particle flow candidate
`isTracker` & `isGlobal`	Has signature in tracker and globally
`looseId`	Loose Muons
`pfRelIso04_all < 0.25`	Relative isolation from other objects

Table 8: Muon selections in the analysis[3]

Selection	Description
$p_T > 20\ GeV$	Remove low p_T noise
$\|\eta\| < 2.3$	Extent of calorimeters
`idDecayModeOldDMs` & `decayMode` $\neq$ `5 or 6`	Veto "experimental 2-prong"
`idDeepTau2017v2p1VSe` $\geq$ `8` & `idDeepTau2017v2p1VSmu` $\geq$ `2` & `idDeepTau2017v2p1VSjet` $\geq$ `8`	Parameters for tau tagging algorithms

Table 9: Tau selections in the analysis[4]

Electrons are selected keeping in mind the trigger threshold of Electron trigger. The trigger threshold, also as trigger turn-on, is 32 GeV here. So the p_t cut is 40 GeV. A summary of all the electron triggers are given in table 7. Similarly, Muon and Tau selections are shown in table 8 and 9 respectively.

5.3.4 Photon selections

Photons have contribution neither to the signal region nor the control regions. Loose photons are selected to be removed from the analysis using the selection mentioned in table 10.

Selection	Description
$p_T > 20\ GeV$	Remove low p_T noise
$\lvert\eta\rvert < 2.5$	Extent of calorimeters
`cutBased` $\geq$ `1`	Loose Photons

Table 10: Photon selections in the analysis[5]

6 Results

After applying all the event selections, various analysis objects were selected following the object selections. Relevant kinematic variables like p_T. η, ϕ and mass were plotted for Electrons, Fatjets, additional jets, MET and Recoil. The various backgrounds were also scaled according to luminosity and the cross-sections of the relevant MC process. A summary of the results are shown in the subsequent subsections.

6.1 Kinematic Plots

6.1.1 Electron kinematics

Electrons are an import part of the Top(e) control region because they tag the process. The plots in 13 show the expected shapes in the three kinematic variables. The Y-axis is the number of events which pass the selection in log scale. X-axis denotes the kinematic variables (p_T, η or ϕ).

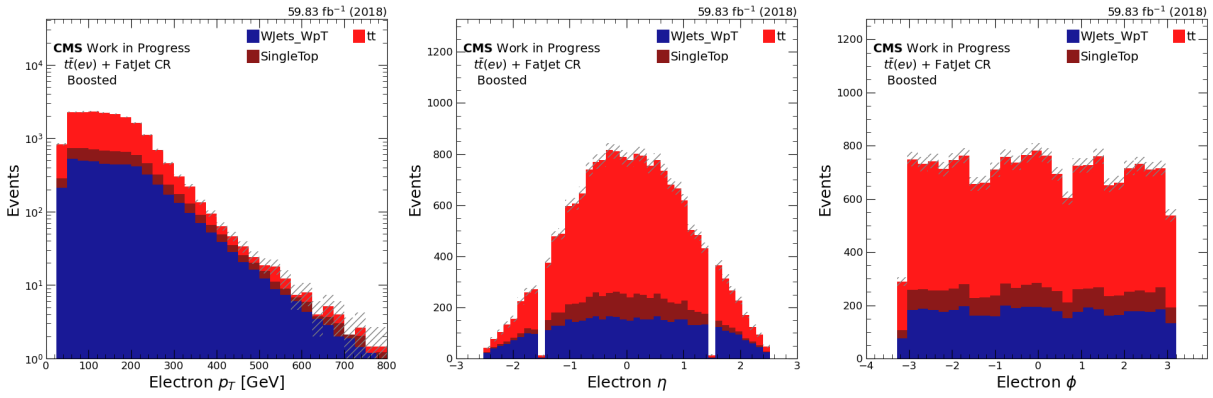


Figure 13: Electron Kinematics

6.1.2 MET kinematics

Another important parameter in the analysis is Missing Transverse Energy. As aspected, MET has smooth variation over the the momentum ranges and is almost uniform across the ϕ direction (see 14).

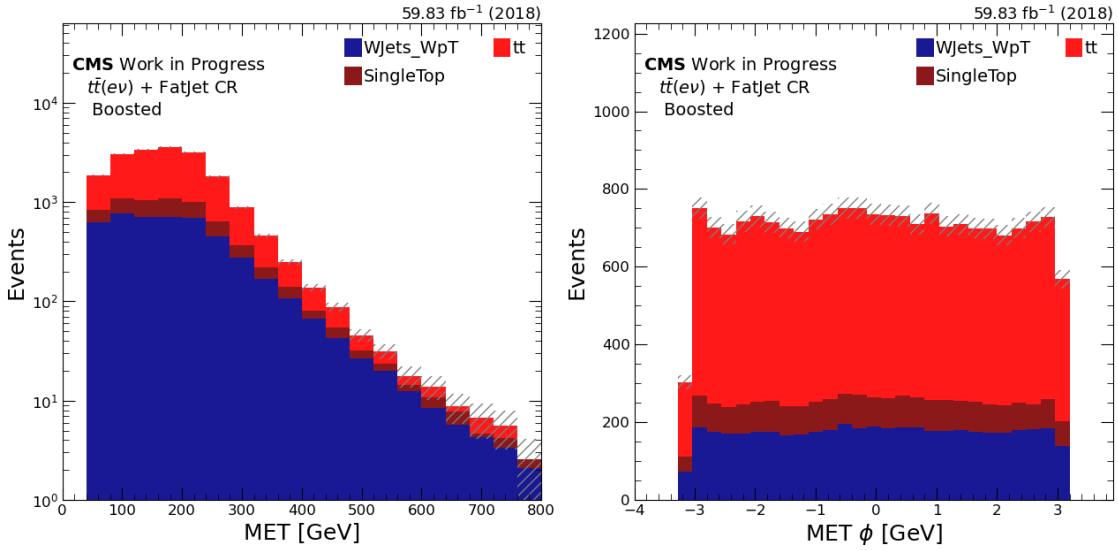


Figure 14: MET Kinematics

6.1.3 Recoil kinematics

Similar to MET, Recoil too shows a smooth variation over the the momentum ranges and is almost uniform across the ϕ direction. The Recoil ¿ 250 GeV cut is clearly visible in the Recoil vs Events plot in 15.

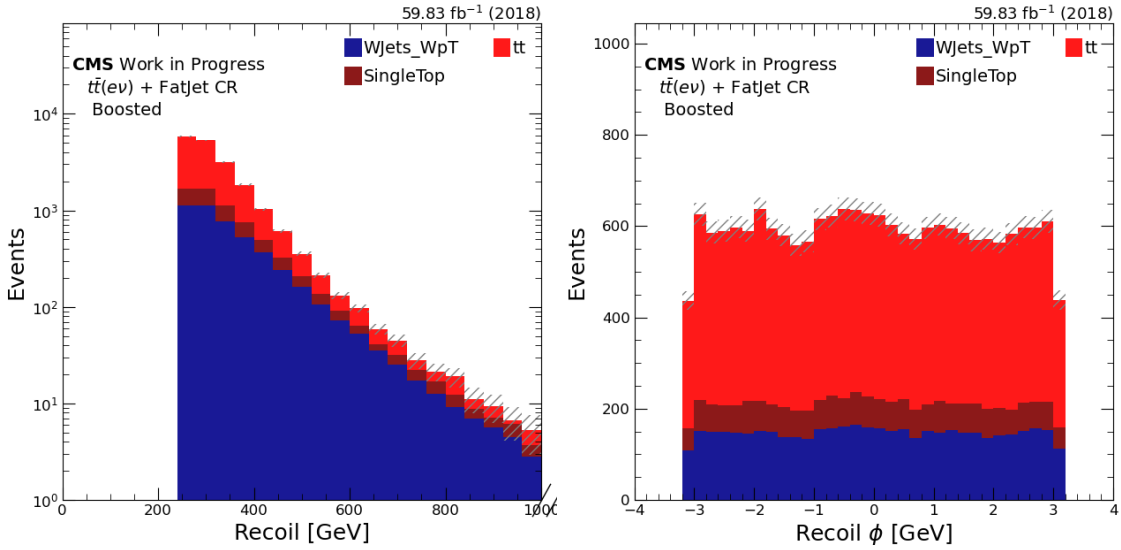


Figure 15: Recoil Kinematics

6.1.4 Fatjet kinematics

Fatjet is a probe of the Higgs boson in the signal final state. The plots in 16 and 17 show a mass peak below 125 GeV (mass of Higgs boson). So it is safe to say that the majority of contribution from the Top background, stays below the Higgs mass. The mass shown in the plots is the softdrop mass which is the mass of the jet obtained from a soft-drop clustering algorithm.

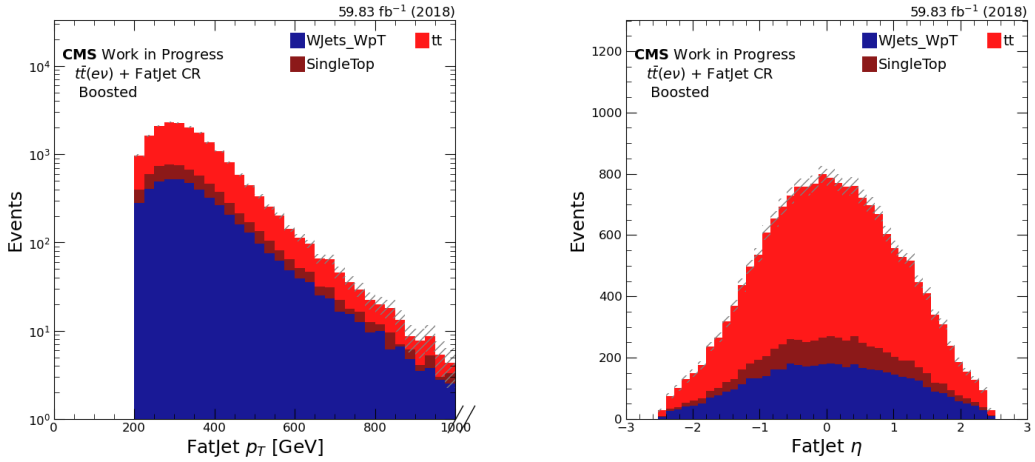


Figure 16: FatJet Kinematics (a)

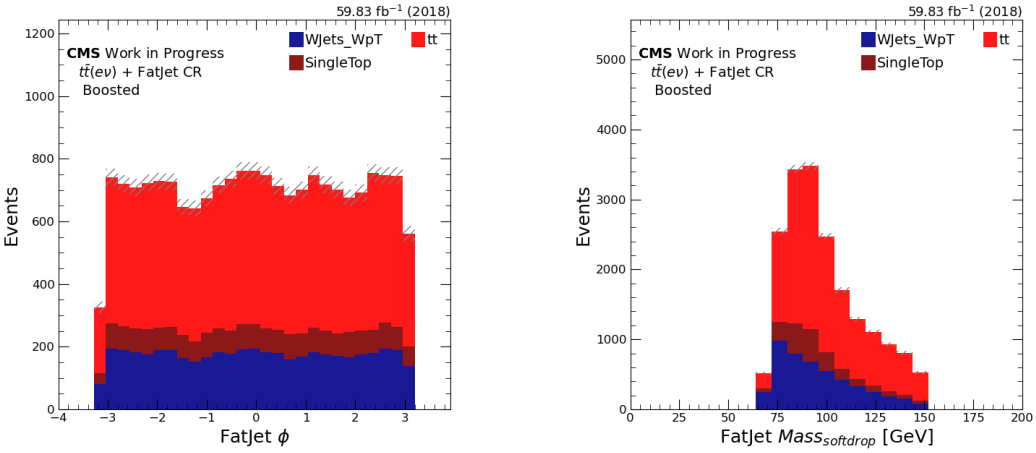


Figure 17: FatJet Kinematics (b)

7 Conclusion

Top(e) - boosted control region was studied successfully. The plots in 6.1 shows the contribution of various processes to the final state p_T^{miss} + $H \rightarrow b\bar{b}$. As expected, $t\bar{t}$, Single top

and W+Jets are the major contributors to the background, in that order.

Estimating one control region is a small step forward to a full analysis which is beyond the scope of this thesis. Nevertheless, study of this single CR gave a great insight into the nature of particle detection and the statistical estimation of particles in a full analysis.

References

[1] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira, "Standard model higgs-boson branching ratios with uncertainties," *The European Physical Journal C*, vol. 71, Sept. 2011.

[2] JanLukasSpah, "Cutbasedelectronidentificationrun2." https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2#Offline_selection_criteria_for_V, 2023. [Accessed 19-05-2024].

[3] WonJun, "Swguidemuonselection." https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonSelection, 2024. [Accessed 19-05-2024].

[4] DanielWinterbottom, "Tauidrecommendationforrun2." https://twiki.cern.ch/twiki/bin/view/CMS/TauIDRecommendationForRun2, 2023. [Accessed 19-05-2024].

[5] RezaGoldouzian, "Cutbasedphotonidentificationrun2." https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedPhotonIdentificationRun2, 2022. [Accessed 19-05-2024].

[6] OpenStax, "Astronomy — Simple Book Production — courses.lumenlearning.com." https://courses.lumenlearning.com/suny-geneseo-astronomy/. [Accessed 16-05-2024].

[7] S. C. Venkataramani and A. C. Newell, "Pattern dark matter and galaxy scaling relations: Is dark matter the self-organized behavior and manifestation of things we already knew?," *The European Physical Journal Special Topics*, vol. 230, p. 2139–2165, June 2021.

[8] J. M. Kubo, J. Annis, F. M. Hardin, D. Kubik, K. Lawhorn, H. Lin, L. Nicklaus, D. Nelson, R. R. R. Reis, H.-J. Seo, M. Soares-Santos, A. Stebbins, and T. Yunker, "The sloan nearby cluster weak lensing survey," *The Astrophysical Journal*, vol. 702, p. L110, aug 2009.

[9] "Planck reveals an almost perfect Universe — esa.int." https://www.esa.int/Science_Exploration/Space_Science/Planck/Planck_reveals_an_almost_perfect_Universe. [Accessed 16-05-2024].

[10] D. Akerib, X. Bai, S. Bedikian, E. Bernard, A. Bernstein, A. Bolozdynya, A. Bradley, D. Byram, S. Cahn, C. Camp, M. Carmona-Benitez, D. Carr, J. Chapman, A. Chiller, C. Chiller, K. Clark, T. Classen, T. Coffey, A. Curioni, E. Dahl, S. Dazeley, L. de Viveiros, A. Dobi, E. Dragowsky, E. Druszkiewicz, B. Edwards, C. Faham, S. Fiorucci, R. Gaitskell, K. Gibson, M. Gilchriese, C. Hall, M. Hanhardt, B. Holbrook, M. Ihm, R. Jacobsen, L. Kastens, K. Kazkaz, R. Knoche, S. Kyre, J. Kwong, R. Lander, N. Larsen, C. Lee, D. Leonard, K. Lesko, A. Lindote, M. Lopes, A. Lyashenko, D. Malling, R. Mannino, Z. Marquez, D. McKinsey, D.-M. Mei, J. Mock, M. Moongweluwan, M. Morii, H. Nelson, F. Neves, J. Nikkel, M. Pangilinan, P. Parker, E. Pease,

K. Pech, P. Phelps, A. Rodionov, P. Roberts, A. Shei, T. Shutt, C. Silva, W. Skulski, V. Solovov, C. Sofka, P. Sorensen, J. Spaans, T. Stiegler, D. Stolp, R. Svoboda, M. Sweany, M. Szydagis, D. Taylor, J. Thomson, M. Tripathi, S. Uvarov, J. Verbus, N. Walsh, R. Webb, D. White, J. White, T. Whitis, M. Wlasenko, F. Wolfs, M. Woods, and C. Zhang, "The large underground xenon (lux) experiment," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 704, p. 111–126, Mar. 2013.

[11] R. Abbasi *et al.*, "Searching for Dark Matter from the Sun with the IceCube Detector," *PoS*, vol. ICRC2021, p. 020, 2022.

[12] R. Bernabei, P. Belli, A. Bussolotti, F. Cappella, R. Cerulli, C. Dai, A. d'Angelo, H. He, A. Incicchitti, H. Kuang, J. Ma, A. Mattei, F. Montecchia, F. Nozzoli, D. Prosperi, X. Sheng, and Z. Ye, "The dama/libra apparatus," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 592, p. 297–315, July 2008.

[13] F. Donato, "Indirect searches for dark matter," *Physics of the Dark Universe*, vol. 4, pp. 41–43, 2014. DARK TAUP2013.

[14] A. Boveia and C. Doglioni, "Dark matter searches at colliders," *Annual Review of Nuclear and Particle Science*, vol. 68, p. 429–459, Oct. 2018.

[15] CERN, "Large hadron collider."

[16] "CERNYellowReportPageBR - LHC Physics - TWiki — twiki.cern.ch." https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR. [Accessed 17-05-2024].

[17] "Vue d'ensemble du LHC — cds.cern.ch." https://cds.cern.ch/record/1708847/export/hx?ln=en. [Accessed 17-05-2024].

[18] "Introduction to CMS Detector Pre-Exercise — cms-opendata-workshop.github.io." https://cms-opendata-workshop.github.io/workshop2023-lesson-cms-detector/01-introduction/index.html, 2013. [Accessed 17-05-2024].

[19] P. A. al., "Energy resolution of the barrel of the cms electromagnetic calorimeter," *Journal of Instrumentation*, vol. 2, p. P04004, apr 2007.

[20] "Real time analysis with the CMS Level-1 Trigger — CMS Experiment — cms.cern." https://cms.cern/news/real-time-analysis-cms-level-1-trigger. [Accessed 18-05-2024].

[21] "CMS coordinate system — tikz.net." https://tikz.net/axis3d_cms/. [Accessed 18-05-2024].

[22] "Curtis McLennan — phys.ksu.edu." https://www.phys.ksu.edu/reu/archive/2022/mclennan.html. [Accessed 18-05-2024].

[23] J. Pivarski, P. Das, C. Burr, D. Smirnov, M. Feickert, T. Gal, L. Kreczko, N. Smith, N. Biederbeck, O. Shadura, M. Proffitt, benkrikler, H. Dembinski, H. Schreiner, J. Rembser, M. R., C. Gu, J. Rübenach, M. Peresano, and R. Turra, "scikit-hep/uproot: 3.12.0," July 2020.

[24] J. Pivarski, C. Escott, N. Smith, M. Hedges, M. Proffitt, C. Escott, J. Nandi, J. Rembser, bfis, benkrikler, L. Gray, D. Davis, H. Schreiner, Nollde, P. Fackeldey, and P. Das, "scikit-hep/awkward-array: 0.13.0," July 2020.

[25] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009.

[26] R. Brun, F. Rademakers, P. Canal, A. Naumann, O. Couet, L. Moneta, V. Vassilev, S. Linev, D. Piparo, G. GANIS, B. Bellenot, E. Guiraud, G. Amadio, wverkerke, P. Mato, TimurP, M. Tadel, wlav, E. Tejedor, J. Blomer, A. Gheata, S. Hageboeck, S. Roiser, marsupial, S. Wunsch, O. Shadura, A. Bose, CristinaCristescu, X. Valls, and R. Isemann, "root-project/root: v6.18/02," June 2020.

[27] G. Petrucciani, A. Rizzi, C. Vuosalo, and on behalf of the CMS Collaboration, "Mini-aod: A new analysis data format for cms," *Journal of Physics: Conference Series*, vol. 664, p. 072052, dec 2015.

[28] M. Peruzzi, G. Petrucciani, A. Rizzi, and for the CMS Collaboration, "The nanoaod event data format in cms," *Journal of Physics: Conference Series*, vol. 1525, p. 012038, apr 2020.

[29] L. Gray, N. Smith, A. Novak, P. Fackeldey, B. Tovar, Y.-M. Chen, G. Watts, and I. Krommydas, "coffea," May 2024.

[30] M. Rocklin, "Dask: Parallel computation with blocked algorithms and task scheduling," in *Proceedings of the 14th python in science conference*, no. 130-136, Citeseer, 2015.

[31] "CMS-T2 Resources — University of Wisconsin - 2013;Madison — hep.wisc.edu." https://www.hep.wisc.edu/cms/comp/resource.html. [Accessed 18-05-2024].

[32] A. Novak, H. Schreiner, and M. Feickert, "mplhep," Sept. 2023.

[33] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[34] "Jets at CMS and the determination of their energy scale — CMS Experiment — cms.cern." https://cms.cern/news/jets-cms-and-determination-their-energy-scale. [Accessed 18-05-2024].

[35] I. Neutelings, "B tagging jets — tikz.net." https://tikz.net/jet_btag/. [Accessed 19-05-2024].

[36] A. Askew, R. Khurana, J. R. Komaragiri, D. Kumar, M. Mittal, P. C. Tiwari, and S.-S. Yu, "Analysis note(ver. 6): Search for dark matter produced in association with a higgs boson decaying to a pair of bottom quarks."