The impact of the 2019 CMS muon endcap upgrade with GEMs on the search for the H$\rightarrow \tau^+\tau^- \rightarrow \mu^+\mu^-$ final state

by
Ankit Durgasis Mohapatra

Bachelor of Science
Physics and Space Sciences
Florida Institute of Technology
2015

A thesis
submitted to the Graduate School of
Florida Institute of Technology
in partial fulfillment of the requirements
for the degree of

Master of Science
in
Physics

Melbourne, Florida
December 2015

We the undersigned committee hereby recommends that the attached document be accepted as fulfilling in part the requirements for the degree of Master of Science in Physics.

"The impact of the 2019 CMS muon endcap upgrade with GEMs on the search for the H$\rightarrow \tau^+\tau^- \rightarrow \mu^+\mu^-$ final state"
a thesis by Ankit Durgasis Mohapatra

Marcus Hohlmann, Ph.D.
Professor, Physics and Space Science
Major Advisor

Francisco Yumiceva, Ph.D.
Assistant Professor, Physics and Space Sciences

Aaron Welters, Ph.D.
Assistant Professor, Mathematical Sciences

Daniel Batcheldor, Ph.D.
Associate Professor and Department head, Physics and Space Sciences

ABSTRACT

The impact of the 2019 CMS muon endcap upgrade with GEMs on the search for the H$\rightarrow \tau^+\tau^- \rightarrow \mu^+\mu^-$ final state

by

Ankit Durgasis Mohapatra

Thesis Advisor: Marcus Hohlmann, Ph.D.

This thesis outlines the study of the H$\rightarrow \tau^+\tau^- \rightarrow \mu^+\mu^-$ channel with the CMS detector. The analysis used Monte-Carlo samples with the GEM geometry and other upgrades to be made in 2019. Studies were done to investigate the effect of lowering the dimuon transverse momentum, p_T, threshold for the $\mu\mu$ channel. It was found that lowering the p_T trigger threshold for the leading and the sub-leading muon from (20 GeV, 10 GeV) to (12 GeV, 5 GeV), respectively doubles the acceptance of events in this channel. Two boosted decision trees (BDTs) were utilized to distinguish between the three event classes: H$\rightarrow \tau^+\tau^-$, Z$\rightarrow \tau^+\tau^-$ and Z$\rightarrow \mu^+\mu^-$. A significance of 26.6 was seen with a cut of -0.13 on the BDT output for the event classes: Z$\rightarrow \tau^+\tau^-$ (signal) and Z$\rightarrow \mu^+\mu^-$ (background), while a significance of 22.4 was seen with a cut value of -0.16 for the event classes: H$\rightarrow \tau^+\tau^-$ (signal) and Z$\rightarrow \tau^+\tau^-$ (background).

Acknowledgements

I would like to express my gratitude to Dr. Marcus Hohlmann, my research and thesis advisor. Without his support and valuable guidance, this research would not be possible in the first place.

I would like to thank the members of my thesis committee, Dr. Francisco Yumiceva and Dr. Aaron Welters for agreeing to be in my committee. I would also like to thank everyone in the CMS GEM Collaboration, especially Dr. Kerstin Hoepfner and Dr. Paolo Giacomelli for all the useful upgrade physics discussions. I would also like to thank Dr. Francisco Yumiceva for clearing my CMSSW doubts promptly. I would like to thank Dr. Rosamaria Venditti for sharing the Bari analysis framework with me. I would like to thank Vallary Bhopatkar, currently a doctoral candidate in our physics group, for all the useful discussions on the multivariate analysis towards the end of my thesis.

I would like to thank Dr. Tania Moulik for motivating me to pursue research in particle physics in India. I would also like to thank Dr. Rob Veenhof for sharing his vast knowledge of gaseous detector simulation with me during my undergraduate research project at CERN.

I would like to thank my room mate, Jiri Kapralek for being very supportive, and for cheering me up when I was down. Last but not the least, I would like to thank my parents for showing keen interest in my studies, and supporting my career decisions.

Dedication

This thesis is dedicated to my parents, Dhirendra Kumar Mohapatra and Anjali Mohapatra.

Contents

1 Standard Model and the Higgs boson **1**

1.1 Standard Model 1

1.2 Higgs boson 3

1.3 H$\rightarrow \tau\tau \rightarrow \mu\mu$ final state 8

2 LHC and CMS Detector **11**

2.1 Large Hadron Collider 11

2.2 The CMS detector 13

2.2.1 Detector description 13

2.2.2 CMS geometry 15

2.2.3 Silicon tracker 16

2.2.4 Muon system 16

2.2.5 Trigger 18

3 Boosted decision trees **20**

3.1 Decision trees 20

3.2 Boosting decision trees 22

4 Physics Analysis **24**

4.1 Simulation datasets 24

4.2 Kinematic studies 25

4.2.1 Acceptance studies 31
4.3 Selection of the H→ $\tau\tau \rightarrow \mu\mu$ events 38
4.4 Multivariate analysis 38
4.4.1 BDT separation of Z→ $\tau\tau$ from Z→ $\mu\mu$ 50
4.4.2 BDT separation of H→ $\tau\tau$ from Z→ $\tau\tau$ 59

5 Summary and Conclusion **68**

A MC datasets **70**

B Comparison of the different MVA methods **71**

C BDT implementation in TMVA **74**

D Optimization of BDTs **77**
D.1 BDTs for H→ $\tau\tau$ (signal) and Z→ $\tau\tau$ (background) 77
D.2 BDT for Z→ $\tau\tau \rightarrow \mu\mu$ (signal) and Z→ $\mu\mu$ 79

E Computing framework **80**

List of Figures

1.1 The leptons, quarks and force carriers in the standard model [1]. 2

1.2 Higgs production processes: (a) gluon fusion (gg→H), (b) vector boson fusion (qqH or q$\bar{q}$H) (c) vector boson ($W^{\pm}$H or ZH) and (d) Higgs boson produced in association with a top-quark pair (t$\bar{t}$H) [2]. 4

1.3 Standard Model Higgs boson production cross sections at E_{CM} = 8 TeV [3]. 5

1.4 Standard Model Higgs boson total production cross sections for E_{CM} = 7, 8 and 14 TeV [3]. 6

1.5 Standard Model Higgs boson decay branching ratios [3]. 7

1.6 Tau decay modes [4]. 9

2.1 Overall view of the LHC experiments [5]. 12

2.2 A cut-away view of the CMS detector. It occupies a volume of cylindrical shape of about 22 m in length and 15 m in diameter, and has a total weight of 14000 t [6]. 14

2.3 Radial section of the CMS detector illustrating particle paths and interactions [6]. 15

2.4 Location of the proposed GE1/1 detector (highlighted in red) within the CMS muon system. A quadrant of the R-z cross-section of the CMS detector is shown[7]. 18

3.1 Schematic view of a decision tree. Starting from the root node (the topmost box), a sequence, of binary splits using the discriminating variables (MET, dimuonratiopt, etc.) is applied to the dataset. The variable that gives the best separation between signal and background is applied at each node. The same variable may thus be used at several nodes (for instance MET), while others might not be used at all. The leaf nodes at the bottom end of the tree are colored differently for signal and background depending on the majority of events that end up in them. 21

4.1 Level 1 muon trigger rates before and after the GE1/1 upgrade at a luminosity of 2×10^{34} $cm^{-2}s^{-1}$ for constant efficiency of 94%. MS1/1 denotes the first endcap muon station Level 1 trigger in both cases, i.e. with CSC-only or with the combination CSC and GEM trigger information. With the addition of GE1/1, the bending angle between the two stations can be used and the trigger rate is greatly reduced [7]. 26

4.2 P_T distribution for the leading and the sub-leading muons in the H$\rightarrow \tau\tau \rightarrow \mu\mu$ sample with GEMs for case-1. 27

4.3 P_T distribution for the leading and the sub-leading muons in the H$\rightarrow \tau\tau \rightarrow \mu\mu$ sample with GEMs for case-2. 28

4.4 P_T distribution for the leading and the sub-leading muons in the H$\rightarrow \tau\tau \rightarrow \mu\mu$ sample with GEMs for case-3. 29

4.5 Correlation between the reconstructed muon p_T and $|\eta|$ for the H$\rightarrow \tau\tau \rightarrow \mu\mu$ sample with GEMs. The plot includes all reconstructed muons which have $|\eta|<2.4$. 30

4.6 Reconstructed muon $|\eta|$ distribution for the H$\rightarrow \tau\tau \rightarrow \mu\mu$ sample with GEMs as a function of the p_T thresholds. The three vertical lines are drawn to differentiate between the different η regions: $|\eta|<1.5$ (barrel), $1.5<|\eta|<2.2$ (endcap with CSC+GEMs) and $2.2<|\eta|<2.4$ (endcap with CSCs only). 31
4.7 Acceptance plot for the H$\rightarrow \tau\tau \rightarrow \mu\mu$ sample with GEMs for case-1. The gradient of the acceptance through the offline point indicates the optimal direction for increasing the acceptance for this channel by potentially lowering p_T thresholds for the dimuon trigger. 34
4.8 Acceptance plot for the H$\rightarrow \tau\tau \rightarrow \mu\mu$ sample with GEMs for case-2. The gradient of the acceptance through the offline point indicates the optimal direction for increasing the acceptance for this channel by potentially lowering p_T thresholds for the dimuon trigger. 35
4.9 Acceptance plot for the H$\rightarrow \tau\tau \rightarrow \mu\mu$ sample with GEMs for case-3. The gradient of the acceptance through the offline point indicates the optimal direction for increasing the acceptance for this channel by potentially lowering p_T thresholds for the dimuon trigger. 36
4.10 Acceptance plot for the H$\rightarrow \tau\tau \rightarrow \mu\mu$ sample with GEMs for $|\eta|<2.4$. The gradient of the acceptance through the offline point indicates the optimal direction for increasing the acceptance for this channel by lowering the p_T thresholds for the dimuon trigger. 37
4.11 MET in GeV(linear scale). 40
4.12 Azimuthal angle between the positive muon momentum and the missing transverse energy in radians (linear scale). 41

4.13 Common logarithm of the muon DCA significance of d_{xy} for the leading muons (linear scale). 42
4.14 Common logarithm of the muon DCA significance of d_{xy} for the sub-leading muons (linear scale). 43
4.15 Common logarithm of the muon DCA significance of d_z for the leading muons (linear scale). 44
4.16 Common logarithm of the muon DCA significance of d_z for sub-leading muons (linear scale). 45
4.17 The ratio of the transverse momentum of the dimuon system to the scalar sum of muon transverse momenta (linear scale). . . . 46
4.18 Pseudo-rapidity of the dimuon system. 48
4.19 The angle ω^* between three-momentum of the positively charged muon and production plane of the dimuon system (linear scale). 49
4.20 Correlation matrix for the $Z\rightarrow \tau\tau \rightarrow \mu\mu$ signal sample. The linear correlation coefficients are given in %. The DCASig(μ) of the d_z for the leading and the sub-leading muons are highly correlated (94 %). The short-hand notations used in this figure are described in table 4.2. 51
4.21 Correlation matrix for the $Z\rightarrow \mu\mu$ background sample. The linear correlation coefficients are given in %. The DCASig(μ) of the d_z for the leading and the sub-leading muons are highly correlated (95 %). The short-hand notations used in this figure are described in table 4.2. 52
4.22 ROC curve for $Z\rightarrow \tau\tau \rightarrow \mu\mu$ (signal) and $Z\rightarrow \mu\mu$ (background). 54
4.23 BDT discriminant distribution for the signal ($Z\rightarrow \tau\tau$) and the background ($Z\rightarrow \mu\mu$). 55

4.24 Signal Z($\rightarrow \tau\tau$) and background (Z$\rightarrow \mu\mu$) efficiency, signal purity, signal efficiency times purity and resulting significance when cutting on BDT output value. 56
4.25 Dimuon invariant mass on a log scale for signal (Z$\rightarrow \tau\tau$) and background (Z$\rightarrow \mu\mu$) after pre-selection cuts. No BDT cuts have been applied. The plots are shown for illustrating the effect of the BDT selection. The samples do not reflect the actual cross-section $\times$ branching ratio of the samples. 58
4.26 Dimuon invariant mass on a log scale for signal (Z$\rightarrow \tau\tau$) and background (Z$\rightarrow \mu\mu$) after the BDT cut (-0.13). The plots are shown for illustrating the effect of the BDT selection. The samples do not reflect the actual cross-section $\times$ branching ratio of the samples. 59
4.27 Correlation matrix for the H$\rightarrow \tau\tau$ signal sample. The linear correlation coefficients are given in %. The DCASig(μ) of the d_z for the leading and the sub-leading muons are highly correlated (96 %). The short-hand notations used in this figure are described in table 4.2. 61
4.28 Correlation matrix for the Z$\rightarrow \tau\tau$ background sample. The linear correlation coefficients are given in %. The DCASig(μ) of the d_z for the leading and the sub-leading muons are highly correlated (93 %). The short-hand notations used in this figure are described in table 4.2. 62
4.29 ROC Curve for H$\rightarrow \tau\tau$ (signal) and Z$\rightarrow \tau\tau$ (background). . . . 63
4.30 BDT discriminant distribution for the signal (H$\rightarrow \tau\tau$) and the background (Z$\rightarrow \tau\tau$). 64

4.31 Signal (H→ $\tau\tau$) and background (Z→ $\tau\tau$) efficiency, signal purity, signal efficiency times purity and resulting significance when cutting on BDT output value. 65
4.32 Dimuon invariant mass on a log scale for signal (H→ $\tau\tau$) and background (Z→ $\tau\tau$) after pre-selection cuts. No BDT cuts have been applied. The plots are shown for illustration of the effect of the BDT selection. The samples do not reflect the actual cross-section × branching ratio of the samples. 66
4.33 Dimuon invariant mass on a log scale for signal (H→ $\tau\tau$) and background (Z→ $\tau\tau$) after the BDT cut (-0.16). The plots are shown primarily for illustration of the effect of the BDT selection. The samples do not reflect the actual cross-section × branching ratio of the samples. 67

B.1 ROC curve for the different MVAs for signal (H→ $\tau\tau$) and background (Z→ $\tau\tau$). 72
B.2 ROC curve values for the different MVAs. 73

D.1 ROC curve values for the BDTs with different parameters. . . . 78
D.2 ROC curve values for BDTs with different parameters. 79

List of Tables

1.1 Fundamental forces. 2

1.2 Topological branching fraction (in %) of $\tau\tau$ to the six final states in decreasing order. 10

4.1 Kinematic acceptance (in %) for the H$\rightarrow \tau\tau \rightarrow \mu\mu$ sample with GEMs for different eta regions. The overall detector acceptance includes all muons which have $|\eta|<2.4$. 33

4.2 Method-specific ranking of discriminant variables for Z$\rightarrow \tau\tau$ (signal) and Z$\rightarrow \mu\mu$ (background). The short-hand notation for some of the discriminants are given in brackets. 50

4.3 ROC integral values. 53

4.4 BDT cut value for the maximum significance. 57

4.5 Method-specific ranking of the discriminant variables for H$\rightarrow \tau\tau$ (signal) and Z$\rightarrow \tau\tau$ (background). 60

4.6 ROC integral values. 63

4.7 BDT cut value for the maximum significance. 65

A.1 Monte Carlo datasets. 70

Chapter 1

Standard Model and the Higgs boson

1.1 Standard Model

The standard model of particle physics is an extremely successful physics model. It describes three fundamental interactions, each of which involves an exchange of force carrying bosons (integer spin 1 particles) between fermions (half-integer spin particles). The elementary particles which make up all known matter in the universe fall into one of the two categories: leptons or quarks. The sets of leptons and quarks are each organized into three generations. The quarks, leptons and the force carrying bosons are shown in fig. 1.1.

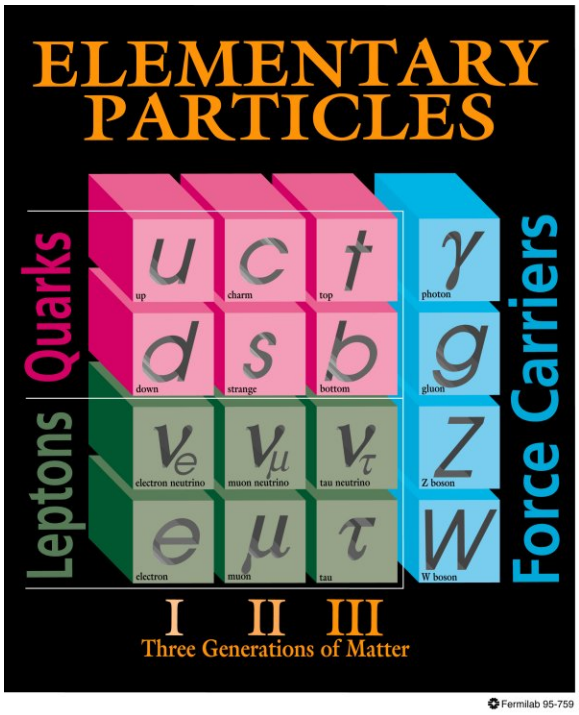


Figure 1.1: The leptons, quarks and force carriers in the standard model [1].

There are four fundamental forces in nature. Table 1.1 lists the name, strength, range and mediating particle for these forces.

Table 1.1: Fundamental forces.

	Relative strength	Range (in meters)	Mediating particle
Strong	1	10^{-15}	gluons
Electromagnetic	10^{-1}	∞	photons (γ)
Weak	10^{-4}	10^{-18}	Vector bosons: $W^{\pm}$ and Z^0
Gravity	10^{-43}	∞	probably gravitons

1.2 Higgs boson

The Standard Model describes the interactions of the elementary particles. However, according to the Standard Model, the particles are all supposed to be massless. The photon and the gluon are massless, but not the rest. In the Standard Model, the mass problem is solved by the spontaneous breaking of the gauge symmetry through what is known as the Higgs mechanism [8]. It is worthwhile to mention that the mass of the Higgs boson is not predicted by the standard model, rather it is a free parameter of the theory.

The major processes by which the Standard Model Higgs boson may be produced are: gluon fusion ($gg \rightarrow H$), vector boson fusion (qqH or $q\bar{q}H$), Higgs boson produced in association with W/Z and Higgs boson produced in association with a top-quark pair ($t\bar{t}H$). Fig. 1.2. shows the Feynman diagrams for the different Higgs production processes.

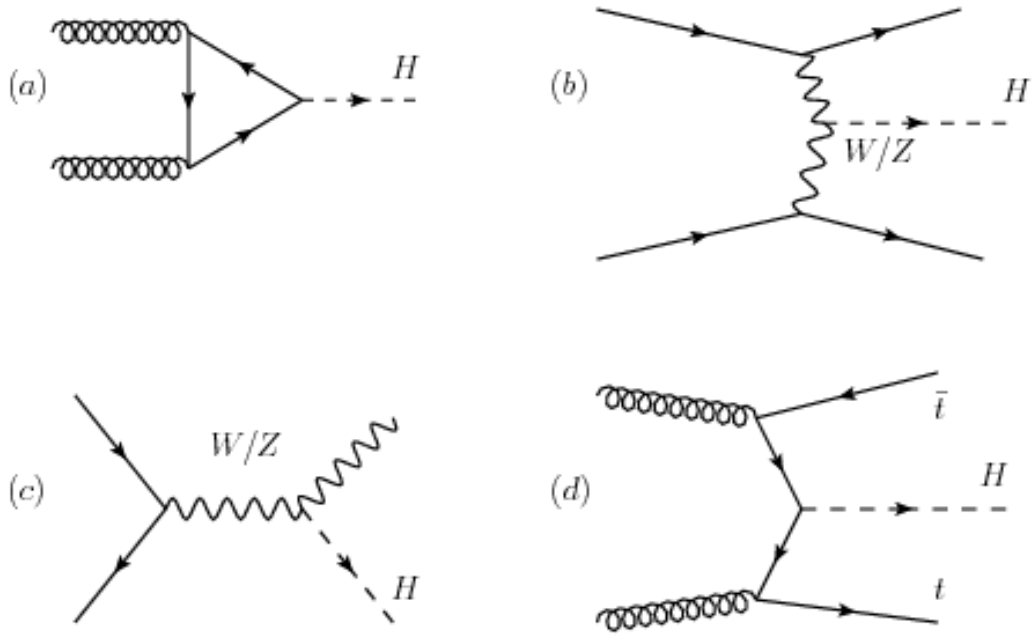


Figure 1.2: Higgs production processes: (a) gluon fusion (gg→H), (b) vector boson fusion (qqH or q$\bar{q}$H) (c) vector boson ($W^{\pm}$H or ZH) and (d) Higgs boson produced in association with a top-quark pair (t$\bar{t}$H) [2].

In fig. 1.3 the production cross section is shown as a function of the standard model Higgs mass. The dominating Higgs production mechanism is the gluon fusion process for all the possible Higgs masses less than 1000 GeV/c^2. At the highest masses, a significant part of the cross section is from the vector boson fusion. The total production cross-sections for E_{CM} = 7, 8 and 14 TeV are compared in fig. 1.4

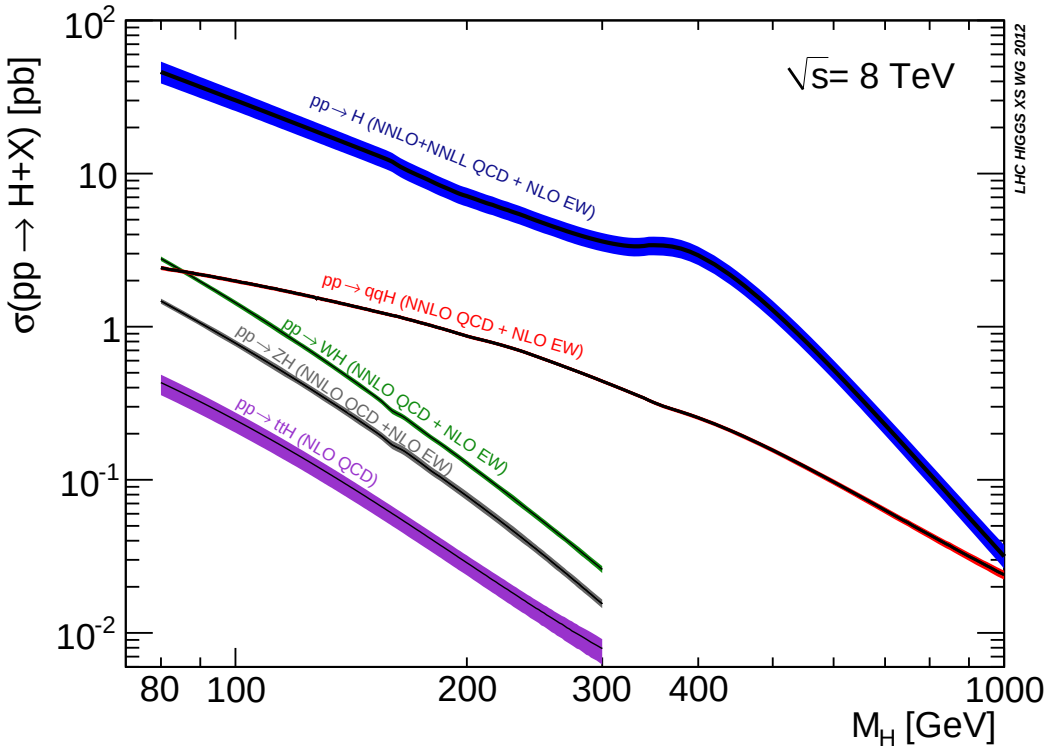


Figure 1.3: Standard Model Higgs boson production cross sections at $E_{CM} = 8$ TeV [3].

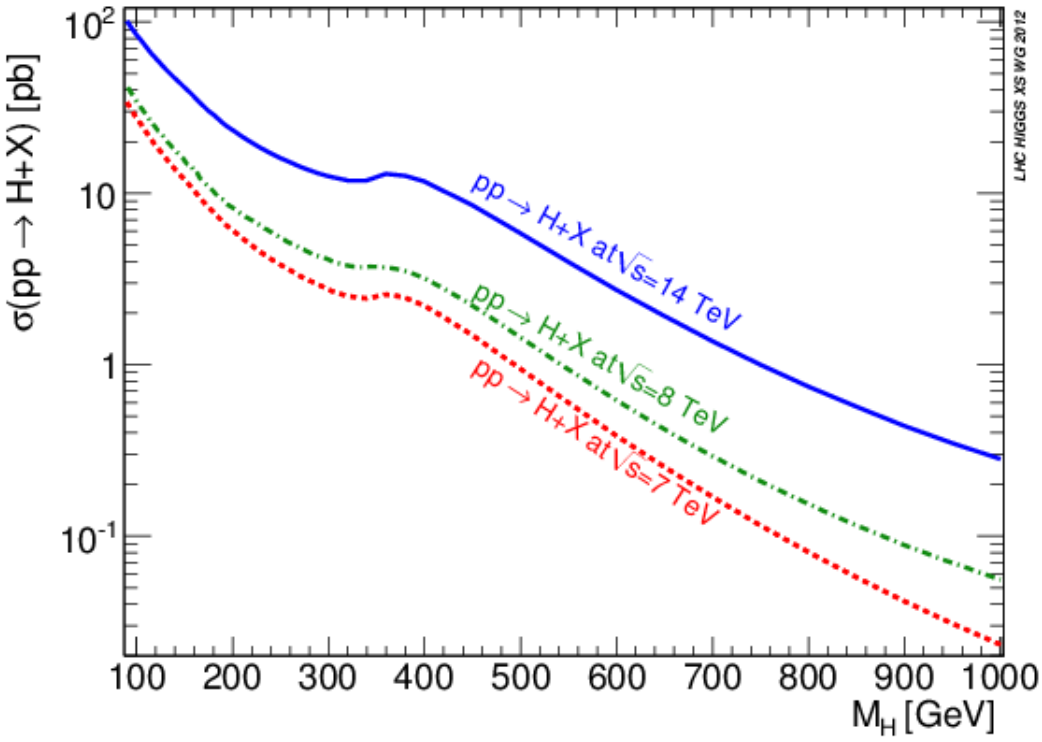


Figure 1.4: Standard Model Higgs boson total production cross sections for E_{CM} = 7, 8 and 14 TeV [3].

The Higgs boson decays on a very fast time-scale (190×10^{-15} seconds[9]), so the only thing we measure is what it decays into. The Higgs can decay to two fermions $f\bar{f}$, two photons $\gamma\gamma$, two gluons gg, and to the weak mediating bosons $W^+ W^-$ and ZZ. Fig. 1.5 shows the branching ratio for each decay channel. The two fermion decay modes: $H \to b\bar{b}$ and $H \to \tau\tau$ are studied to understand the nature of the Higgs boson interaction with the fermions. For a Standard Model Higgs boson with M_H=125 GeV, the $H \to b\bar{b}$ channel despite having the highest branching ratio, suffers from large backgrounds and poor mass resolution of the b quark system. The Higgs boson has a relatively large branching ratio into τ leptons, making the $H \to \tau\tau$ decay channel very promising. This

is a significant motivation for CMS to have strong tau identification capabilities.

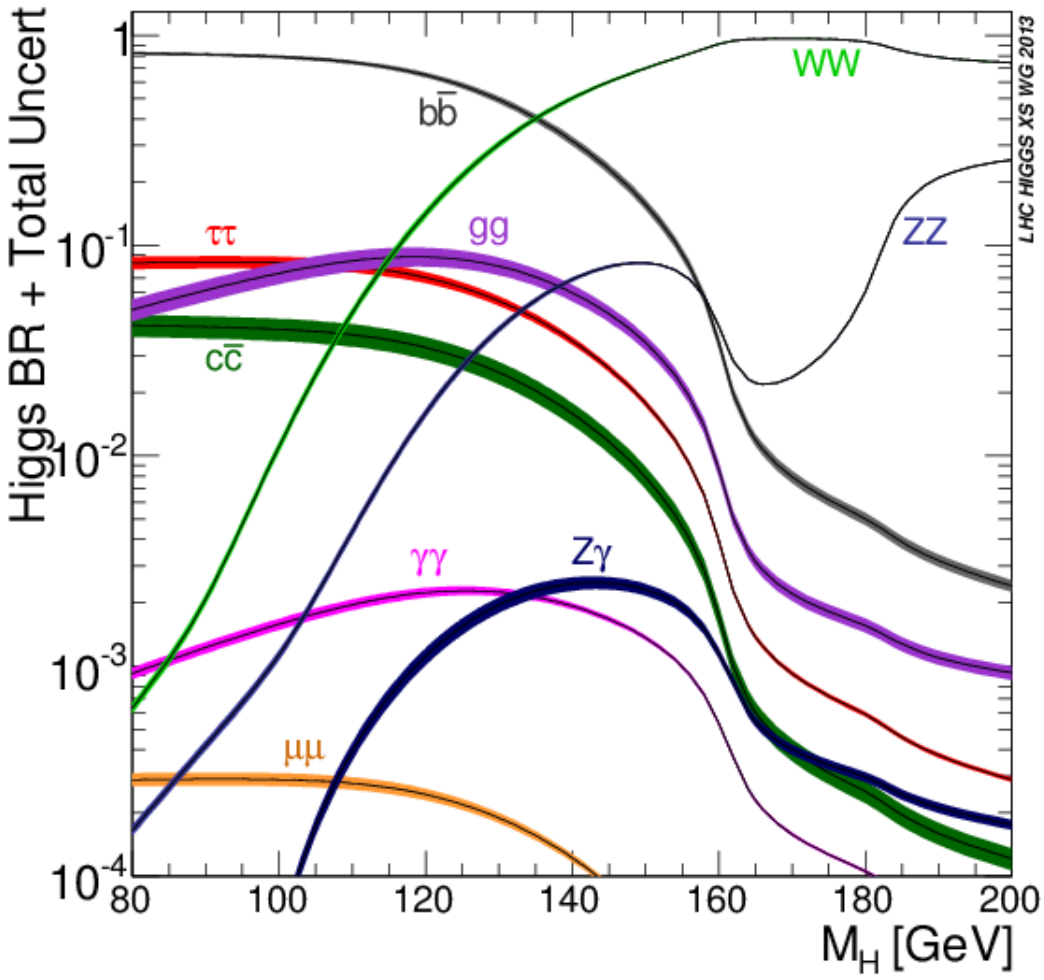


Figure 1.5: Standard Model Higgs boson decay branching ratios [3].

1.3 H$\rightarrow \tau\tau \rightarrow \mu\mu$ final state

The final state, which will be referred to as H$\rightarrow \tau\tau \rightarrow \mu\mu$ in this thesis, in full detail is shown below:

$$H \rightarrow \tau^+\tau^- \rightarrow \mu^+\nu_\mu\bar{\nu_\tau}\mu^-\bar{\nu_\mu}\nu_\tau$$

The tau lepton can decay into hadrons and leptons. Fig. 1.6 summarizes the major tau decays. The tau lepton decays hadronically approximately 65% of the time; however, it decays only 17% of the time to a muon. This information is used to calculate the topological branching fraction for the $\tau\tau$ to the different final states in table 1.2.

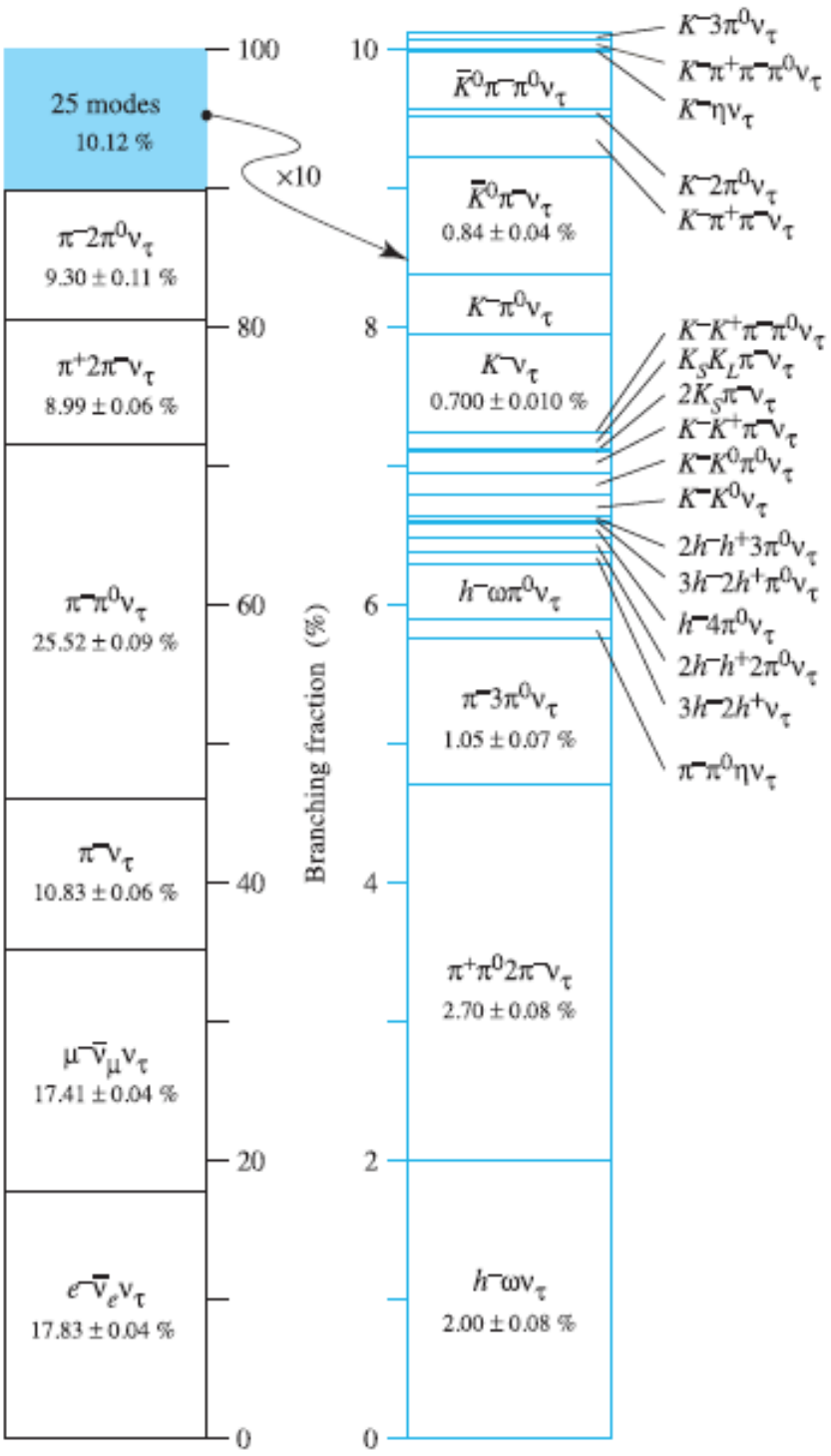


Figure 1.6: Tau decay modes [4].

Table 1.2: Topological branching fraction (in %) of $\tau\tau$ to the six final states in decreasing order.

Decay channel	Topological branching fraction(in %)
$\tau_{had}\tau_{had}$	42
eτ_{had}	23
$\mu\tau_{had}$	23
ee	6
eμ	3
$\mu\mu$	3

It is seen from the table 1.2 that $\tau\tau \rightarrow \mu\mu$ has a small branching fraction. The dominant backgrounds for this channel are: Z$\rightarrow \tau\tau$ and Z$\rightarrow \mu\mu$. The high background coupled with the low topological branching ratio are the major challenges associated with this channel; this warrants the need to increase the event acceptance. It is worth exploring because it gave an improvement of the order of 10% to the run-1 Higgs results when combined with the other decay channels [10], [11].

Chapter 2

LHC and CMS Detector

2.1 Large Hadron Collider

The Large Hadron Collider (LHC) [12] is a hadron-hadron collider hosted inside the former LEP tunnel. The collider contains high frequency accelerating cavities, focusing quadrupole magnets, and superconducting dipole magnets for the bending of the protons (or heavy ions) in the plane of the 27 km long accelerator ring. The layout is shown in figure 2.1.

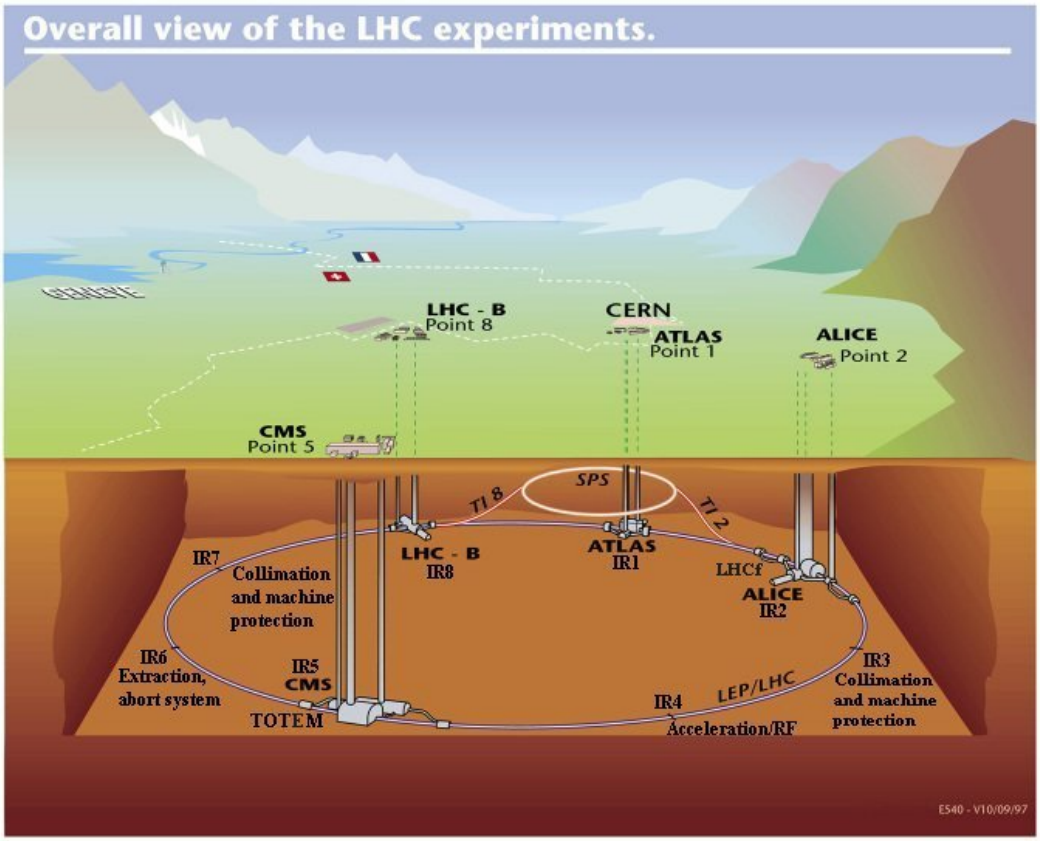


Figure 2.1: Overall view of the LHC experiments [5].

In total, there are 4 collision points (major experiments) on the LHC ring:

- A Toroidal LHC ApparatuS (ATLAS): A general purpose experiment.
- Compact Muon Solenoid (CMS): The second general purpose experiment.
- A Large Ion Collider Experiment (ALICE): Experiment designed to study the quark-gluon plasma.
- LHCb: Experiment designed to use b-quarks to study matter-antimatter asymmetry.

2.2 The CMS detector

The Compact Muon Solenoid (CMS) [13] is a general purpose detector operating at the LHC collider at CERN. It is a particle detector that is designed to study a wide range of particles and phenomena produced in high-energy collisions in the LHC.

2.2.1 Detector description

Fig. 2.2 gives a cut-away view of the CMS detector. The particle detection at CMS is shown in fig. 2.3. Particles emerging from the collisions first encounter a silicon tracker that charts their positions as they move through the detector, which allows the measurement of their momentum. Next are the calorimeters that measure the energy of the particles. The lead tungstate Electromagnetic Calorimeter (ECAL) measures the energy of photons and electrons (particles which interact electro-magnetically, hence the name), whereas the brass Hadron Calorimeter (HCAL) detects any particle made up of quarks.

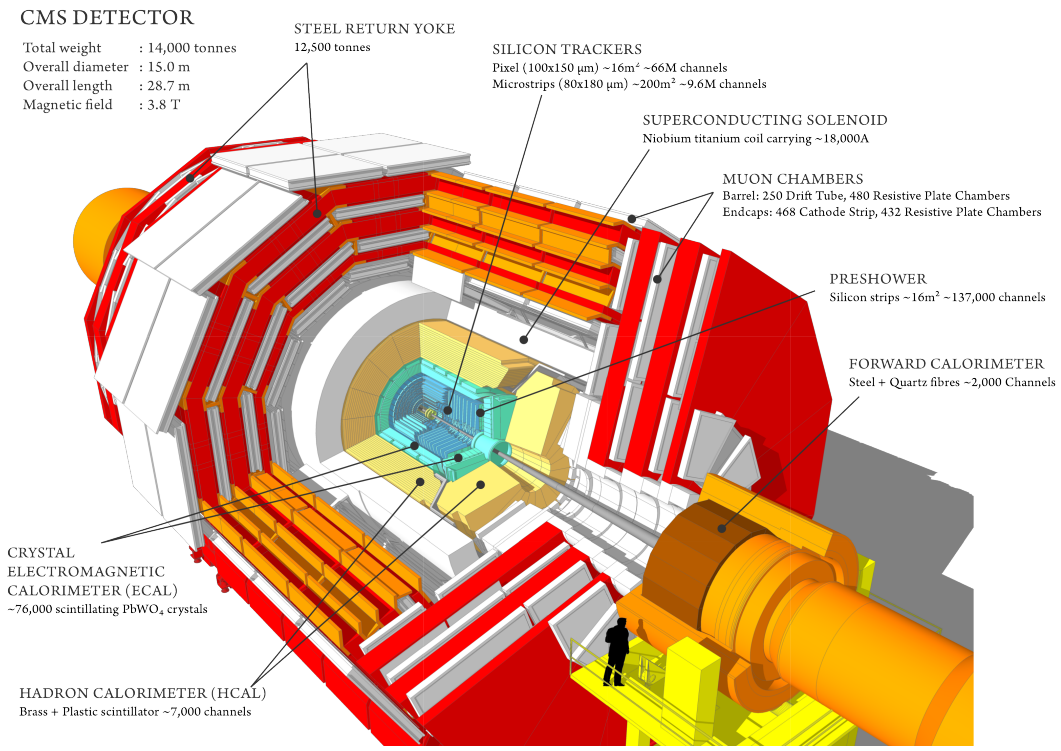


Figure 2.2: A cut-away view of the CMS detector. It occupies a volume of cylindrical shape of about 22 m in length and 15 m in diameter, and has a total weight of 14000 t [6].

All particles but muons and neutrinos are stopped by the calorimeters and the iron magnet. The muon tracks are measured by four layers of muon detectors that are interleaved with the iron yoke. The neutrinos are indirectly inferred from the missing transverse energy in the event.

Bunches of particles collide up to 40 million times per second within the LHC, so a trigger system that saves only potentially interesting events is essential. This reduces the number of interactions recorded from one billion to O(100) per second.

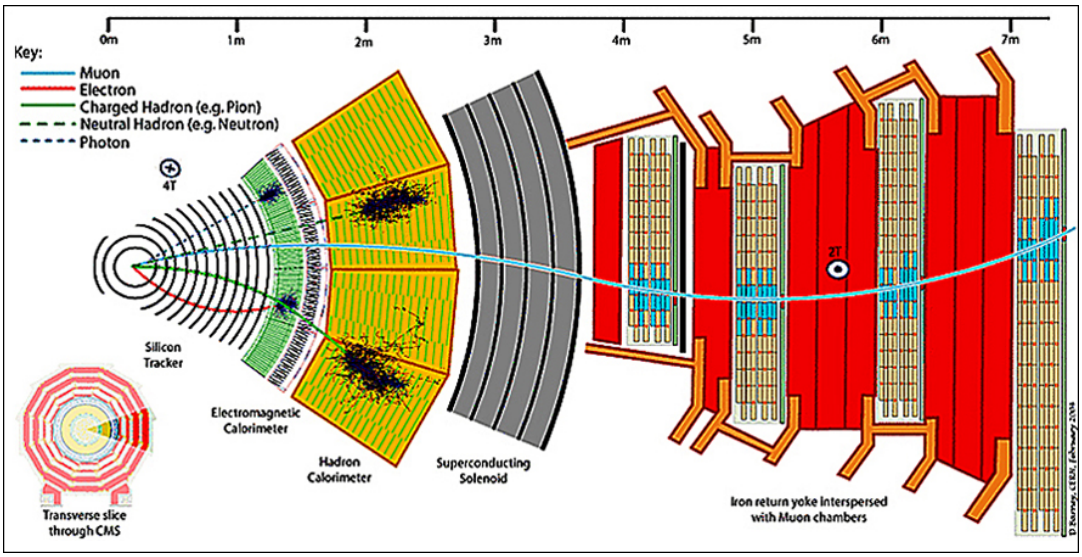


Figure 2.3: Radial section of the CMS detector illustrating particle paths and interactions [6].

2.2.2 CMS geometry

The interaction point is defined as the origin of the CMS Cartesian coordinates [13]. The z-axis is along the beam direction, the positive x-axis points into the center of the LHC ring, and the positive y-axis points upwards to form a right handed coordinate system. The azimuthal angle is measured from the x-axis in the x-y plane.

The polar angle θ is the angle from the beam axis. However, a quantity called the rapidity is favored to express angle with respect to the beam line because differences in rapidity are invariant under Lorentz boosts. A massless approximation to rapidity is the pseudo-rapidity, defined as:

$$\eta = \text{-ln}(\tan(\theta/2)) \tag{2.1}$$

An alternative definition of η [1] is:

$$\frac{1}{2} \times ln(\frac{|p| + p_z}{|p| - p_z}) \tag{2.2}$$

|p| is the magnitude of the three-momentum and p_z is the z-component of the momentum of the particle. Angular separation between points in the $\eta - \phi$ plane is defined as:

$$\Delta R = \sqrt{(\phi' - \phi)^2 + (\eta' - \eta)^2} \tag{2.3}$$

Sub-detectors relevant to the $\mu\mu$ channel are briefly discussed in the following subsections.

2.2.3 Silicon tracker

The CMS silicon tracker [14] records the paths taken by charged particles by finding their positions at a number of key points; this is used to calculate the momentum of the particles (the more curved the path in the solenoidal magnetic field, the less momentum the particle had). The tracker consists of the silicon pixels and microstrip detectors. As particles travel through the tracker, the pixels and the microstrips produce tiny electric signals that are amplified and detected. The tracker employs sensors with 75 million separate electronic read-out channels.

2.2.4 Muon system

Muon detection is one of CMS's most important tasks. Muons can penetrate several meters of iron and are not stopped by any calorimeter. Chambers to detect muons are placed at the very edge of the experiment where they are the only particles likely to register a signal. The momentum of a muon is measured by fitting a curve to hits among the four muon stations; this combined with the

[1] This definition is later used in the analysis to calculate one of the BDT input variables

tracker measurements precisely traces the muon's path.

The CMS muon sub-detector consists of Drift Tubes (DTs), Cathode Strip Chambers (CSCs) and Resistive Plate Chambers (RPCs). DTs extend up to $|\eta| < 1.2$ and CSCs cover $1.0<|\eta|< 2.4$. RPCs provide a redundant and fast trigger and coarse position measurement for $|\eta|<1.6$.

Muon trigger studies for the CMS Phase 1 Upgrade [15] indicate substantial efficiency losses for muons with $p_T < 25$ GeV in the endcap region. One of the planned improvements [16], [7] is the installation of an additional set of muon detectors, GE1/1, that use gas electron multipliers (GEM) [17] in the first endcap muon station to improve the muon triggering and reconstruction in the region $1.6<|\eta|<2.2$ for Run-3 which will start in 2019. The location of the proposed GE1/1 is shown in the fig. 2.4

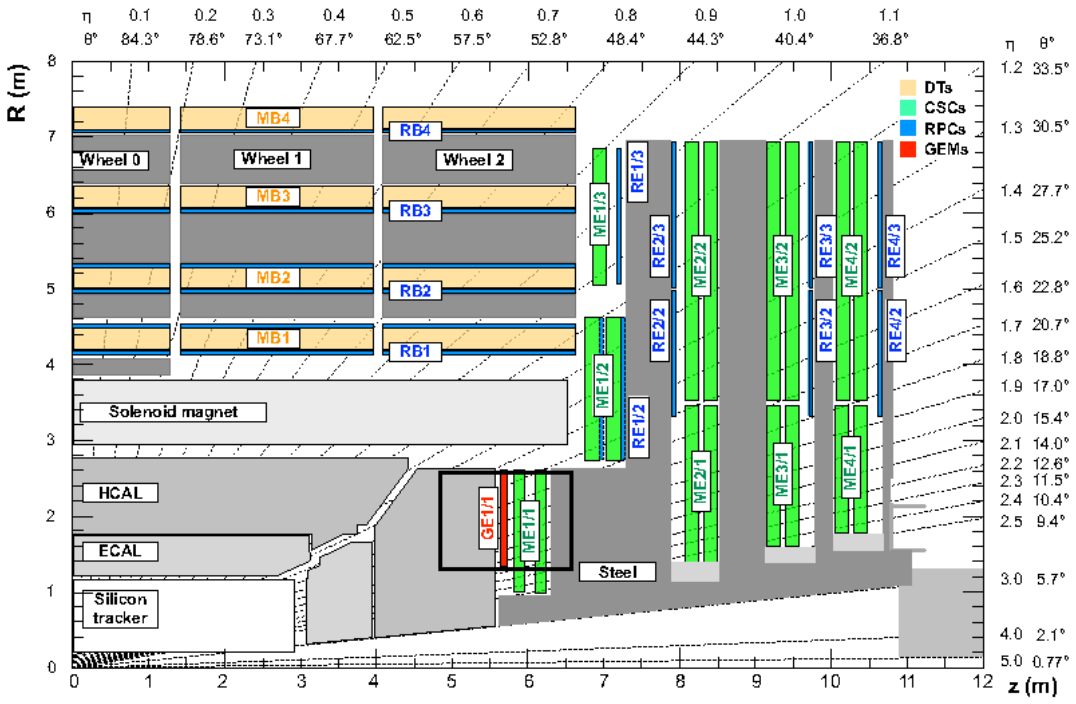


Figure 2.4: Location of the proposed GE1/1 detector (highlighted in red) within the CMS muon system. A quadrant of the R-z cross-section of the CMS detector is shown[7].

2.2.5 Trigger

When CMS is performing at its peak, about one billion proton-proton interactions will take place every second inside the detector. It is impossible to read and store such massive amount of data coming from these events. We therefore need a trigger [18] that can select the potentially interesting events and achieve a reduction of O(10^{-5}), which can be read out and stored for subsequent analysis. However, with time between bunch crossing as low as 25 (50) nanoseconds, which corresponds to a crossing frequency of 40 (20) MHz, the required rejection is too large to be achieved in a single step, and hence the selection task is split into steps: Level 1 (hardware based) and High Level Trigger (software based).

The Level 1 (L1) trigger is an extremely fast and automatic process that looks for simple signs of interesting physics. The L1 trigger makes decisions based on the the information from the calorimeter and the muon systems. It also uses global sums of E_T and missing E_T, tested against several E_T and p_T thresholds. It is expected that the inclusion of GE1/1 would help in maintaining the Run-2 L1 trigger thresholds for Run-3 and beyond.[7]

The output from the L1 trigger is of the order of 100 kHz; the HLT is required to reduce this to a few hundred Hz. High Level Triggers make further combinations and other topological calculations on the digital list of objects transmitted from L1. The HLT utilizes the full event data for the decision to keep an event. The last stage of HLT processing does reconstruction and event filtering.

Chapter 3

Boosted decision trees

In order to separate the background from the signal, multivariate analysis was used. Different multivariate techniques were studied, out of which boosted decision trees (BDTs) performed better than most other methods.This is shown in Appendix B.

3.1 Decision trees

A decision tree is a machine-learning technique which combines several weak classifiers to create a more powerful multivariate discriminant. It is a structure of cuts organized into nodes. A node is the point in the tree in which a variable and a cut value are provided and an event is determined to either pass (signal) or fail (background) it. As seen in figure 3.1, a tree begins at a primary (root) node and splits into two secondary nodes depending on the result of the root node cut (A yes will put it in the left node, and a no will move it to the right node). A new cut is then applied at the secondary nodes. Each of these nodes carry a cut criteria and a tested candidate would again advance from this node either to the left or the right daughter node. Each cut path eventually stops at

some terminal node or "leaf" (once the stopping criteria is fulfilled), where the event will be classified as a signal or a background event. TMVA v4.02 package [19] provided by root v6.04.02 [20] was used for the multivariate analysis.

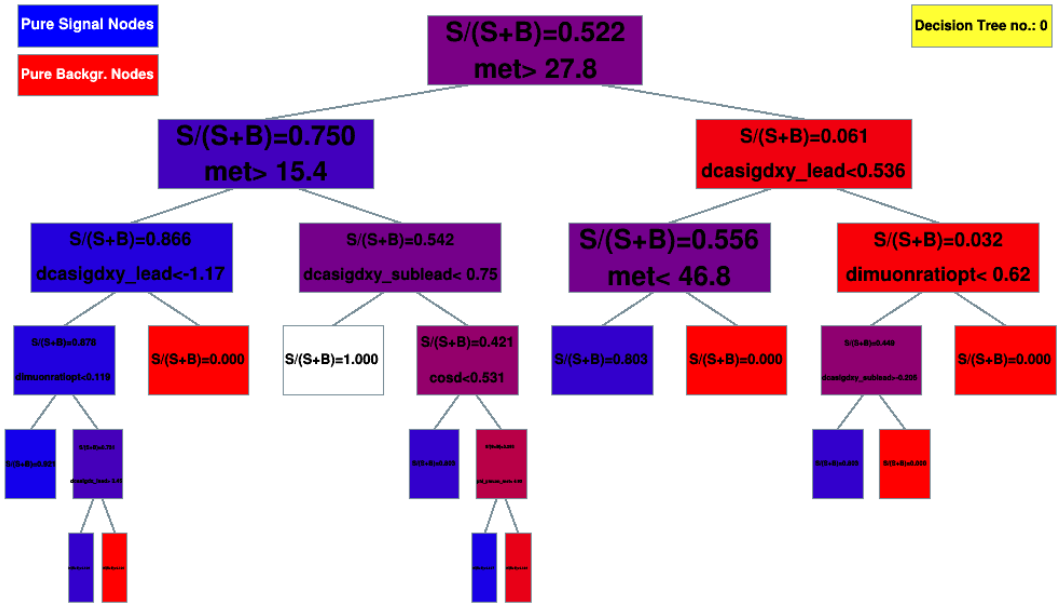


Figure 3.1: Schematic view of a decision tree. Starting from the root node (the topmost box), a sequence, of binary splits using the discriminating variables (MET, dimuonratiopt, etc.) is applied to the dataset. The variable that gives the best separation between signal and background is applied at each node. The same variable may thus be used at several nodes (for instance MET), while others might not be used at all. The leaf nodes at the bottom end of the tree are colored differently for signal and background depending on the majority of events that end up in them.

3.2 Boosting decision trees

Boosting refers to an effective method of producing a stronger classifier by combining weak classifiers (algorithms or variables) which are rough and moderately inaccurate. This is achieved by repeatedly applying the weak classifier to different versions (re-weighted) of the training data. TMVA supports five boosting options for decision trees: Adaptive, Gradient, Bagging, Decorrelation + Adapative and Fisher discriminant. Adaptive boost (Adaboost) was extensively used for the analysis. The different steps involved in a boosted decision tree classification problem along with their implementation in TMVA are outlined below:

- All the training events or a subset of them, in case one is using bagging, are given to the first tree. Bagging refers to the process of selecting a fixed number of random events from the parent sample (specified by a number between 0 and 1, where 1 refers to the entire dataset). Bagging is observed to improve the performance of the decision tree D. The first node from where the classification starts is called a root node. If the UseBaggedSample flag is set to kFALSE, then all the events are given to each tree, else a percentage of the events (specified by the flag BaggedSampleFraction) is chosen randomly (Poisson distribution) from the events and given to the decision trees.

- The minimum and maximum values of the discriminant variables for the event sample are determined; the separation gain is calculated at the grid points (the number of grid points is specified by nCuts), and the cut which maximizes the separation gain (called the optimal gain) is chosen for each variable. Then, the variable with the cut which maximizes the separation gain is chosen. The separation gain is calculated as follows:

Separation index = Parent index - left index - right index

The indices are calculated using gini index which is given as p(1-p) where p is the number of events in a particular node.

- The error value (err) for each tree is calculated by counting the number of misclassified events (signal events classified as background and vice-versa). The misclassified events are then multiplied by the boostweight, $(\frac{1-err}{err})^{\beta}$. After this step, the events are weighted inside a loop, and the next decision tree sees events with new modified weights.

- The next tree is then grown in a similar way, until a stop criterion is specified.

- In the end, an event is classified as a signal or a background by taking a weighted majority vote; if a particular event is classified by 200 trees, it is classified as a signal if it ends up in the signal leaf node in the majority of the trees.

Chapter 4

Physics Analysis

4.1 Simulation datasets

The official Run-1 analysis [11] of this channel used three different signal samples: gluon fusion (gg→H), vector boson fusion (qqH or q$\bar{\text{q}}$H) and Higgs boson produced in association with top-quark pair (tt$\bar{H}$). However, this study just focuses on one of the signal samples, namely, gluon fusion. In the beginning, official CMS Monte Carlo (MC) samples were used, but it turned out that the number of events which had at least two muons were about 2000, which is not enough statistics for our study. Instead private samples were generated in the FIT computing cluster (App. E). One of the limitations of the official samples was that the tau particles decayed to all the channels; this was circumvented by forcing the taus to decay to two muons using the tauola interface in Pythia 6 in the private samples [21],[22],[23],[24]. The private MC samples were generated with no pile-up. Pileup refers to the overlapping secondary proton-proton collisions on top of the primary interaction; it will be a major challenge for the high luminosity LHC runs (Run-2 and beyond).

The two prominent backgrounds for the dimuon final state were studied: Drell-

Yan to Tau Tau (Z$\rightarrow \tau\tau$) and Drell-Yan to Mu Mu (Z$\rightarrow \mu\mu$). In order to get around the problem of low statistics, private samples were generated for both the event classes. For the $Z \rightarrow \tau\tau$ irreducible background, taus were forced to decay into muons. An irreducible background has the same kind of particles as the signal event. The final state topology for H$\rightarrow \tau\tau \rightarrow \mu\mu$ (signal) and Z$\rightarrow \tau\tau \rightarrow \mu\mu$ (background) event classes are identical and characterized by two muons and four neutrinos. A summary of all the simulated event samples is given in table A.1 in the appendix A.

4.2 Kinematic studies

It is important to understand the various kinematic distributions for the muons in the signal sample, which would help us to optimize the p_T cut for the trigger studies. Fig. 4.1 shows the single muon trigger rate curves before and after the GE1/1 upgrade for the region $1.6<|\eta|<2.2$ [7]. The upgrade will help in keeping the L1 trigger thresholds at low p_T values, which is important for the physics studies ranging from new physics searches to the measurements in the Higgs sector. The trigger rate for the Run2 (Run1) L1 trigger threshold of 25 (14) GeV is (was) 16 (7) kHz [25]. From fig. 4.1, it is seen that the L1 trigger threshold for Run3 with GE1/1 can be lowered to about 9 GeV (10 GeV), while maintaining the same Run2 (Run1) L1 trigger rate.

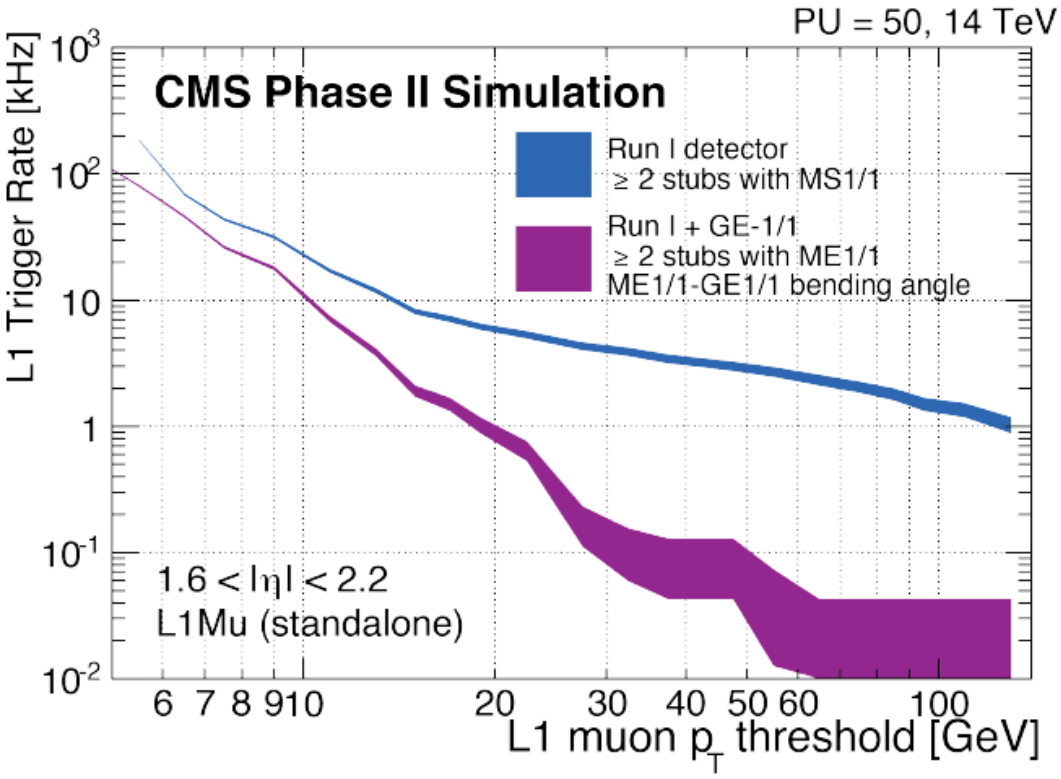


Figure 4.1: Level 1 muon trigger rates before and after the GE1/1 upgrade at a luminosity of 2×10^{34} cm^{-2}s^{-1} for constant efficiency of 94%. MS1/1 denotes the first endcap muon station Level 1 trigger in both cases, i.e. with CSC-only or with the combination CSC and GEM trigger information. With the addition of GE1/1, the bending angle between the two stations can be used and the trigger rate is greatly reduced [7].

In order to isolate and understand the impact of the GEMs in the muon reconstruction, the p_T distribution for the $H \rightarrow \tau\tau \rightarrow \mu\mu$ in the Run-3 configuration of CMS, which includes GEMs, is shown for different η regions in figures 4.2, 4.3 and 4.4. The different η regions, which will be designated by case numbers from now on, are listed below:

- Case 1: Leading and sub-leading muon in $1.5 < |\eta| < 2.2$, i.e. both in the

endcap region covered by GE1/1

- Case 2: Leading muon in 1.5<$|\eta|$<2.2 (endcap) and sub-leading muon in $|\eta|$<1.5 (barrel)
- Case 3: Leading muon in $|\eta|$<1.5 (barrel) and sub-leading muon in 1.5<$|\eta|$<2.2 (endcap)

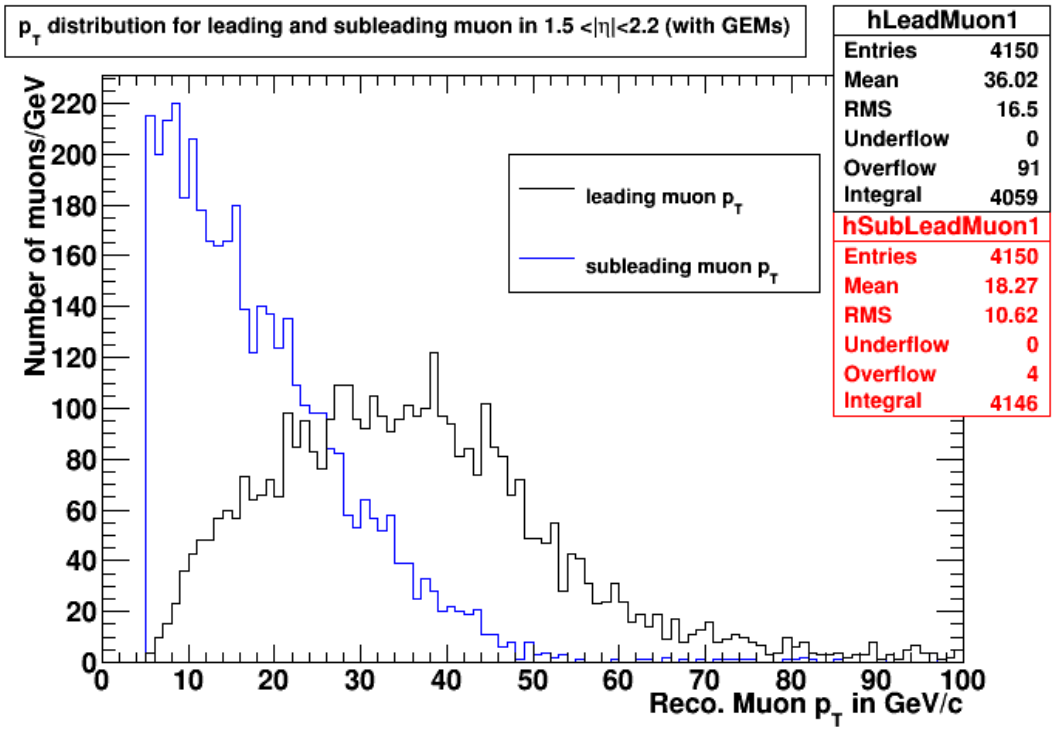


Figure 4.2: P_T distribution for the leading and the sub-leading muons in the H$\rightarrow \tau\tau \rightarrow \mu\mu$ sample with GEMs for case-1.

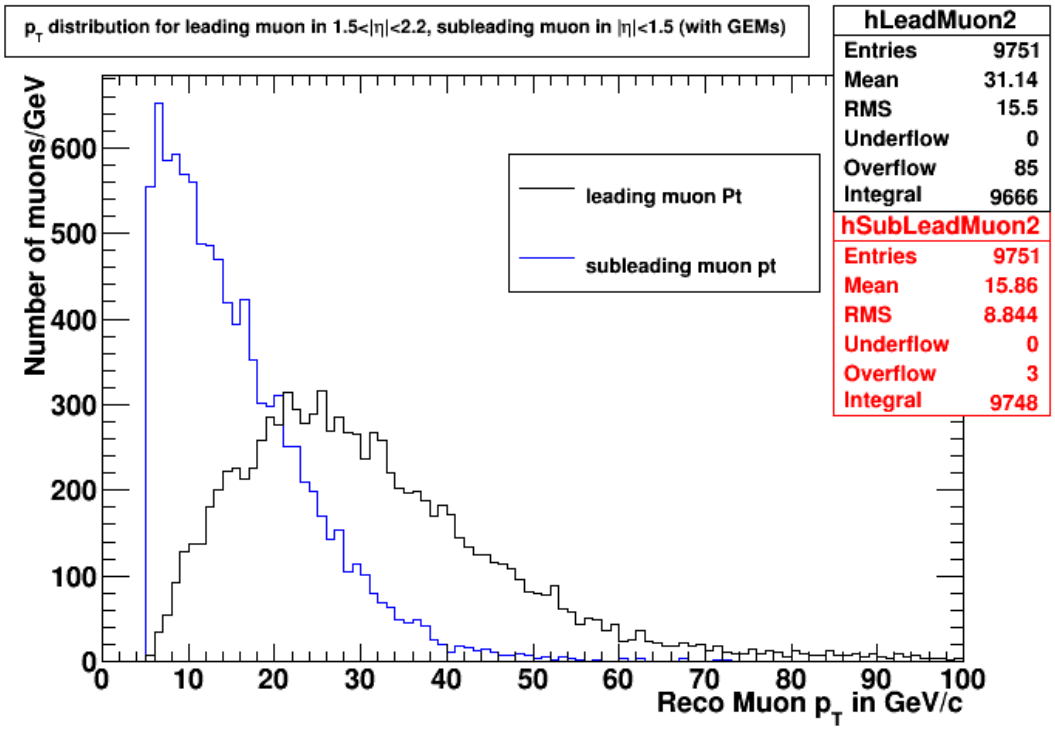


Figure 4.3: P_T distribution for the leading and the sub-leading muons in the H$\rightarrow \tau\tau \rightarrow \mu\mu$ sample with GEMs for case-2.

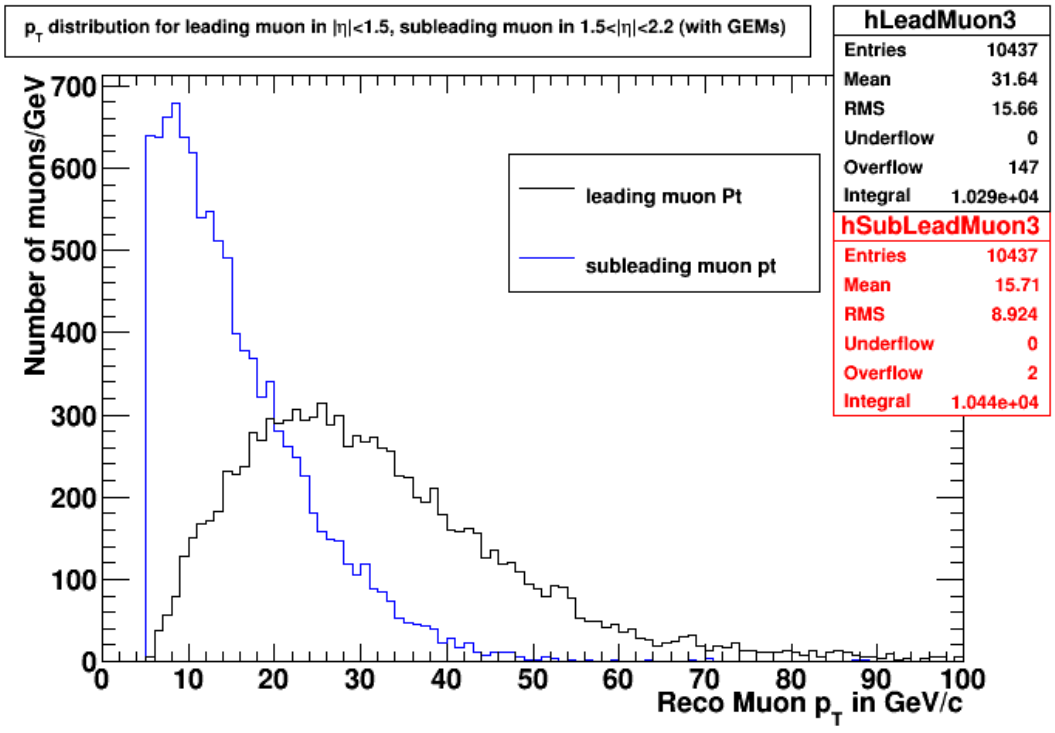


Figure 4.4: P_T distribution for the leading and the sub-leading muons in the H$\rightarrow \tau\tau \rightarrow \mu\mu$ sample with GEMs for case-3.

The kinematic phase space in fig. 4.5 shows that the muons in this channel have a soft p_T spectrum in the detector acceptance ($|\eta|$<2.4); most of the muons are concentrated in the central region and have $p_T < 20$ GeV. This necessitates the need to lower the muon p_T thresholds in order to keep a high trigger and reconstruction efficiency.

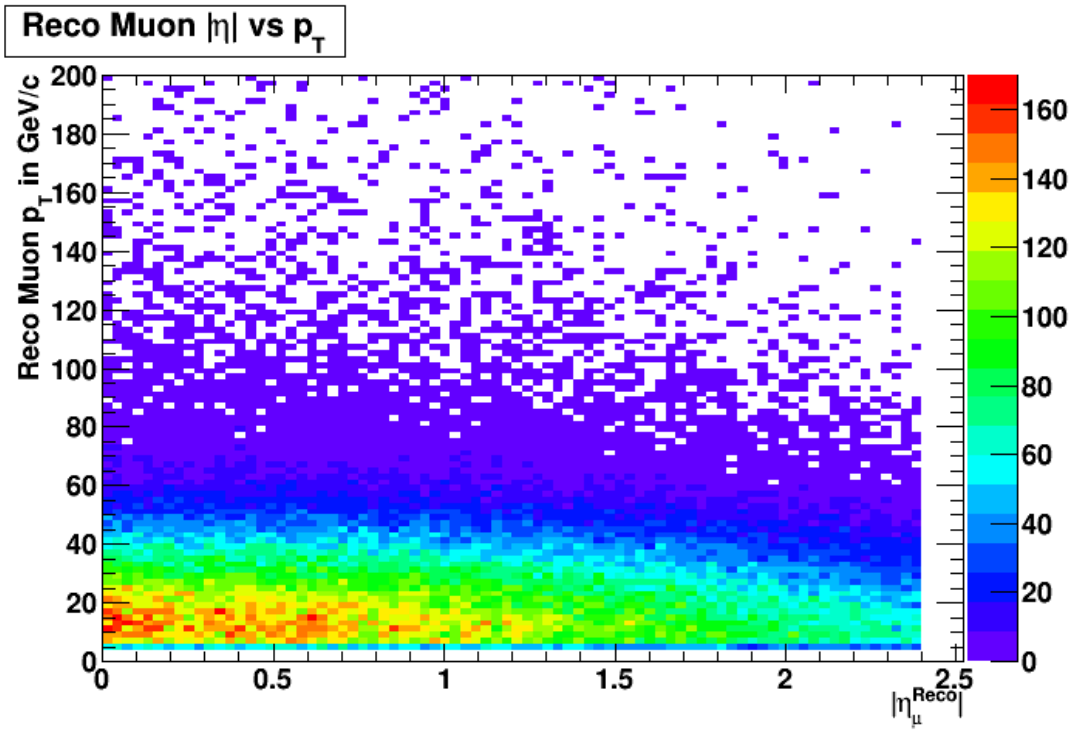


Figure 4.5: Correlation between the reconstructed muon p_T and $|\eta|$ for the H$\rightarrow \tau\tau \rightarrow \mu\mu$ sample with GEMs. The plot includes all reconstructed muons which have $|\eta|<2.4$.

Fig. 4.6 shows the $|\eta|$ for all the reconstructed muons as a function of the p_T thresholds. As can be seen in the figure, 21% of the muons lie in the forward region ($1.5<|\eta|<2.2$); this serves as a motivation to study the H$\rightarrow \tau\tau \rightarrow \mu\mu$ channel in order to probe the possible improvements from the inclusion of GEMs in the muon reconstruction. The importance of keeping a low trigger threshold for muons is underlined by the fact that decreasing the p_T threshold from 35 GeV to 15 GeV leads to a 50% increase in the event acceptance.

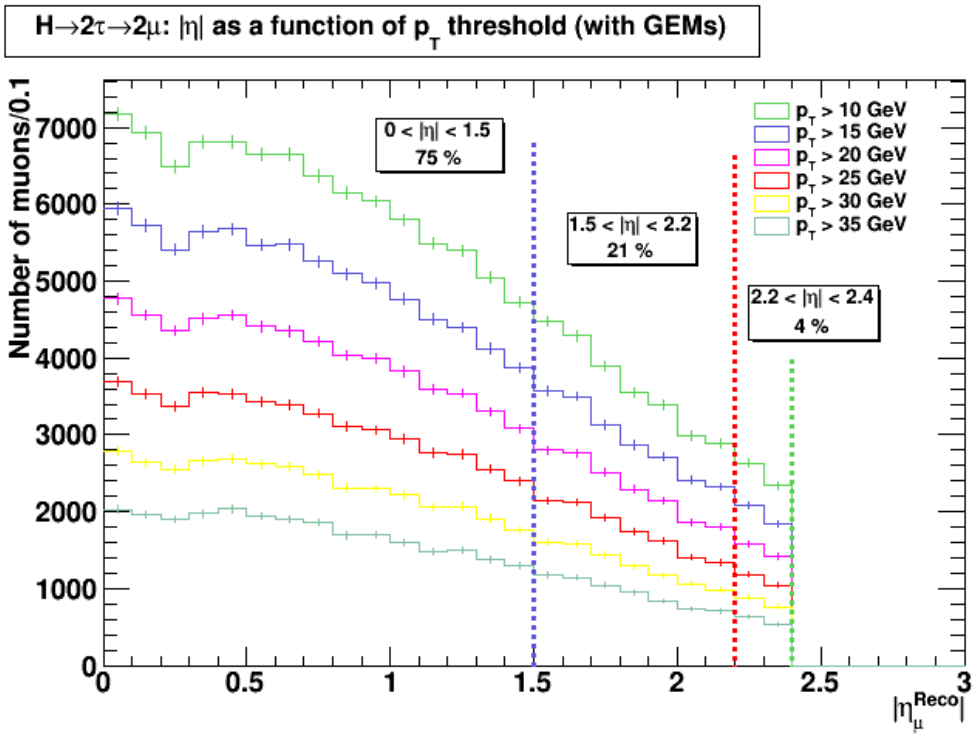


Figure 4.6: Reconstructed muon $|\eta|$ distribution for the H$\rightarrow \tau\tau \rightarrow \mu\mu$ sample with GEMs as a function of the p_T thresholds. The three vertical lines are drawn to differentiate between the different η regions: $|\eta|<1.5$ (barrel), $1.5<|\eta|<2.2$ (endcap with CSC+GEMs) and $2.2<|\eta|<2.4$ (endcap with CSCs only).

4.2.1 Acceptance studies

The search for the SM Higgs boson decaying to τ leptons, is performed in five different final states [11]. Table 1.2 shows the branching fraction of $\tau\tau$ to these final states.

The main challenges in the $\mu\mu$ final state are the large Drell-Yan background and the small topological branching ratio; it is necessary to increase the acceptance of the events in the $\mu\mu$ decay channel. In order to achieve increased acceptance, we can lower the trigger thresholds for the leading and the sub-leading muon.

Two dimensional acceptance plots are useful in quantifying the (increased) kinematic acceptance by lowering the p_T trigger thresholds for the leading and the sub-leading muon for the 2019 CMS run with the GEMs. The acceptance (in %) is calculated as follows:

$$\text{Acceptance (in \%)} = \frac{n_1}{n_2} \times 100 \tag{4.1}$$

where n_1 is the the number of events which have leading and sub-leading muon p_T greater than the pair of dimuon p_T trigger values, and n_2 is the total number of accepted events. For instance, the acceptance for (18 GeV, 5 GeV) would be proportional to the number of events which have p_T values greater than 18 GeV and 5 GeV for the leading and the sub-leading muon, respectively. Figs. 4.7, 4.8, 4.9 and 4.10 show the percentage of events which are accepted for a given cut on the p_T values of the leading and the sub-leading muon. Table 4.1 summarizes the acceptance values for three pairs of cut (trigger) values. These values are listed below:

- (12 GeV, 5 GeV): L1 trigger value for double muons (Run-2)
- (17 GeV, 8 GeV): HLT trigger path for double muons (Run-1)
- (20 GeV, 10 GeV): Offline selection (Run-1)

Table 4.1: Kinematic acceptance (in %) for the H$\rightarrow \tau\tau \rightarrow \mu\mu$ sample with GEMs for different eta regions. The overall detector acceptance includes all muons which have $|\eta|<2.4$.

	(12 GeV, 5 GeV)	(17 GeV, 8 GeV)	(20 GeV, 10 GeV)
Leading muon in $1.5<|\eta|<2.2$ Sub-leading muon in $1.5<|\eta|<2.2$	96	66	48
Leading muon in $1.5<|\eta|<2.2$ sub-leading muon in $|\eta|<1.5$	96	62	42
Leading muon in $|\eta|<1.5$ Sub-leading muon in $1.5<|\eta|<2.2$	95	62	41
Overall detector acceptance	95	63	43

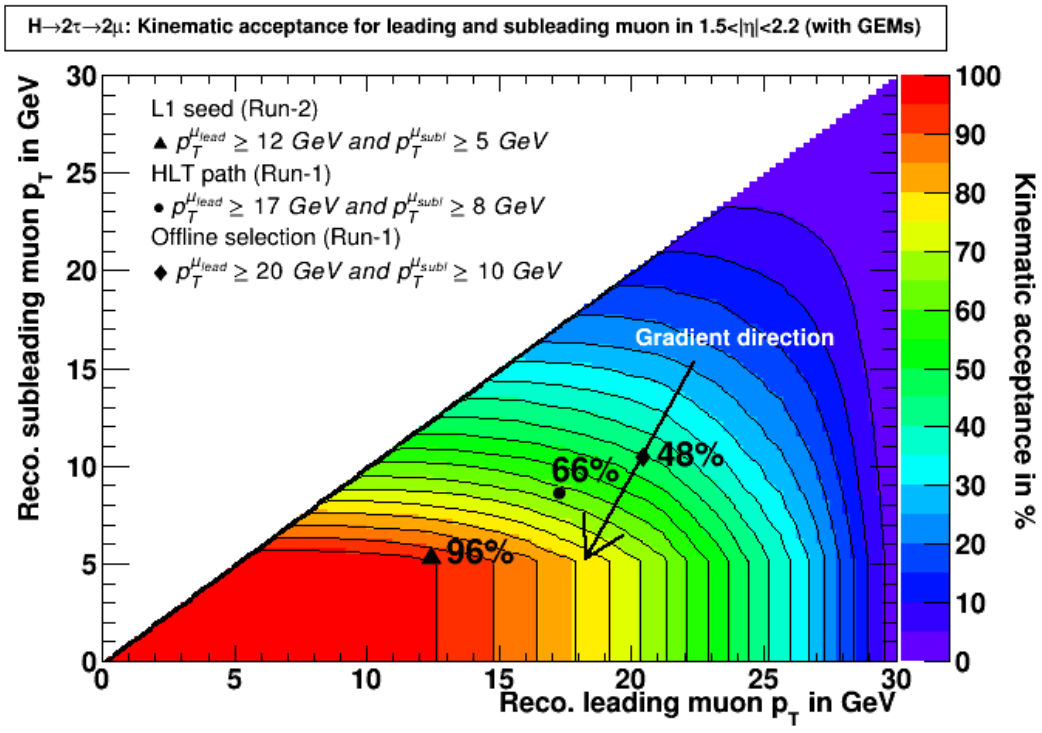


Figure 4.7: Acceptance plot for the H$\rightarrow \tau\tau \rightarrow \mu\mu$ sample with GEMs for case-1. The gradient of the acceptance through the offline point indicates the optimal direction for increasing the acceptance for this channel by potentially lowering p_T thresholds for the dimuon trigger.

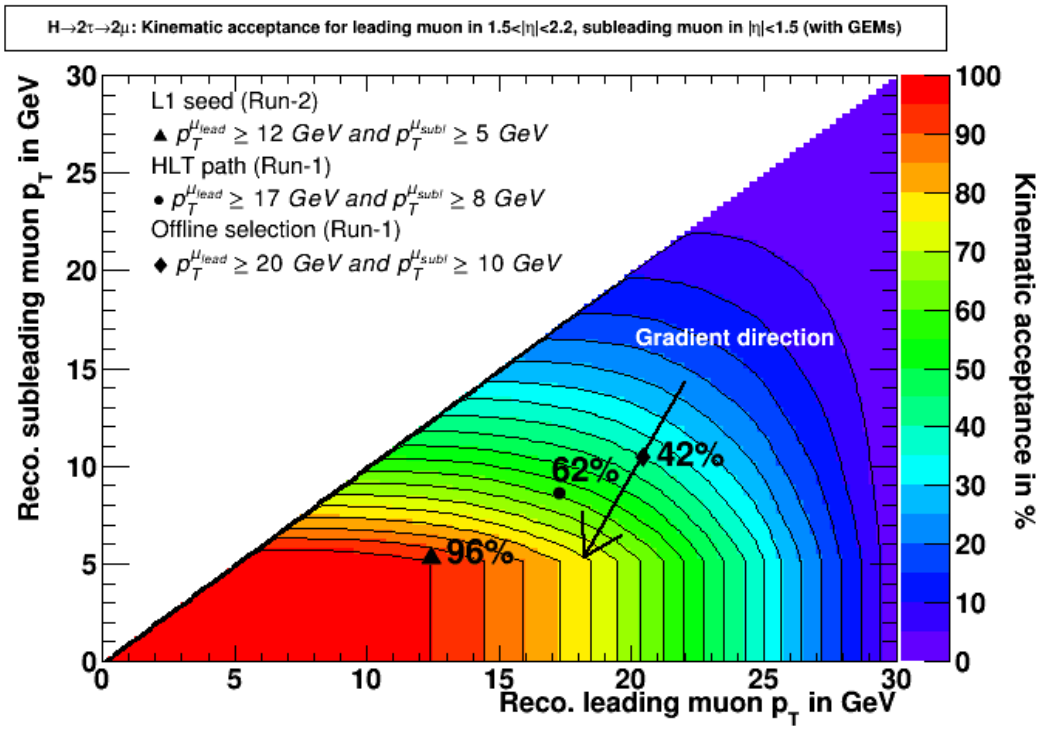


Figure 4.8: Acceptance plot for the H$\rightarrow \tau\tau \rightarrow \mu\mu$ sample with GEMs for case-2. The gradient of the acceptance through the offline point indicates the optimal direction for increasing the acceptance for this channel by potentially lowering p_T thresholds for the dimuon trigger.

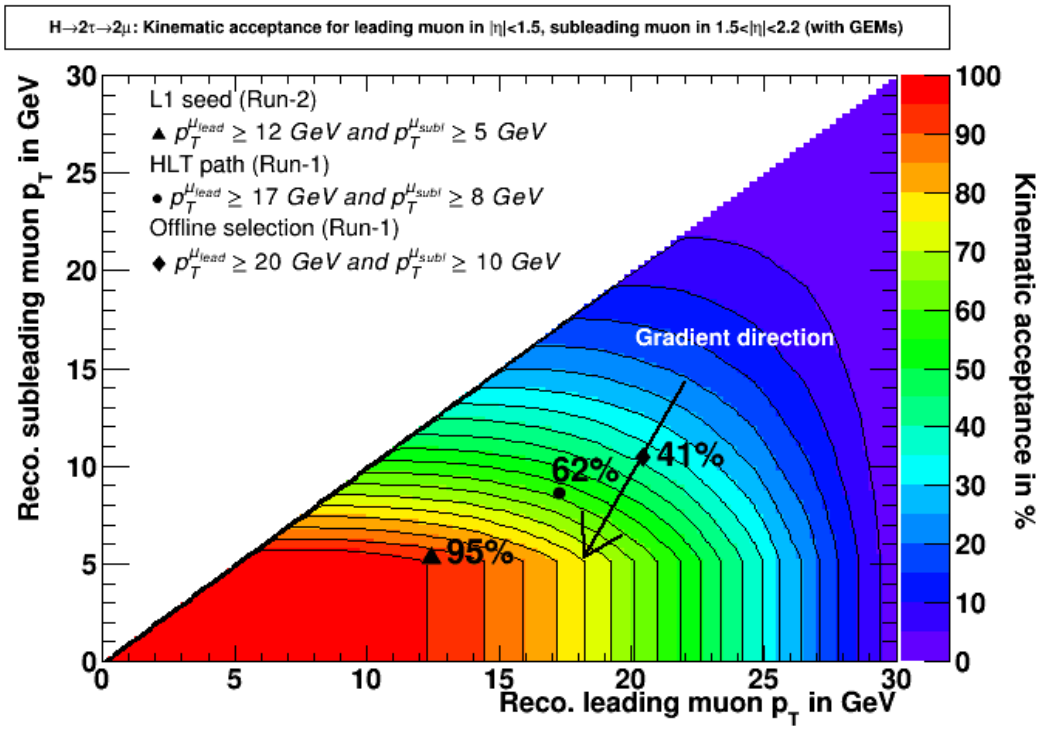


Figure 4.9: Acceptance plot for the H$\rightarrow \tau\tau \rightarrow \mu\mu$ sample with GEMs for case-3. The gradient of the acceptance through the offline point indicates the optimal direction for increasing the acceptance for this channel by potentially lowering p_T thresholds for the dimuon trigger.

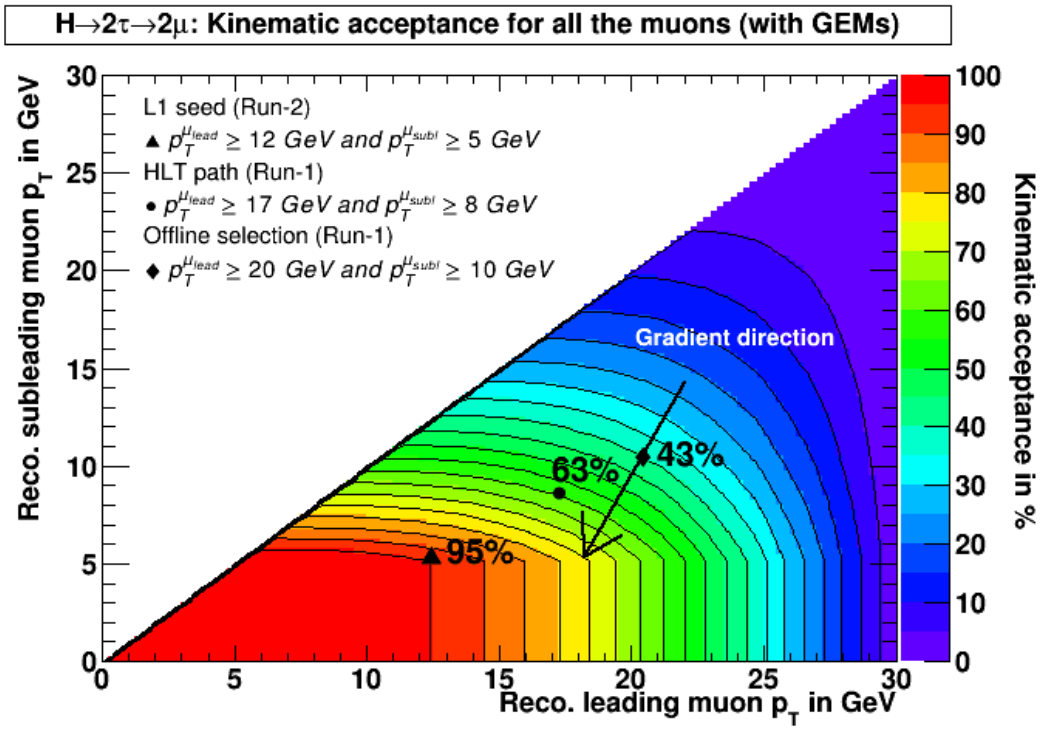


Figure 4.10: Acceptance plot for the H→ $\tau\tau$ → $\mu\mu$ sample with GEMs for $|\eta|$<2.4. The gradient of the acceptance through the offline point indicates the optimal direction for increasing the acceptance for this channel by lowering the p_T thresholds for the dimuon trigger.

The acceptance plots show that lowering the trigger thresholds from (20 GeV, 10 GeV) to (12 GeV, 5 GeV) increases the acceptance of events (in all the η regions) in the $\mu\mu$ decay channel by a factor of more than two (∼45 % to ∼90%). The contour plots (and the gradient direction arrow) show that it is more effective to lower the cut on the sub-leading muons to increase the acceptance values.

4.3 Selection of the H$\rightarrow \tau\tau \rightarrow \mu\mu$ events

In this section, we describe the different pre-selection cuts which are applied to select muons and not some other particles which fake a muon. Events with at least two muons are selected in the beginning of the analysis. Subsequently, the muons are selected if they have opposite charges, $|\eta|$ less than 2.4, are reconstructed within the inner tracker and the muon stations, and reconstructed as so-called particle flow muons [26]. The selected muons must fulfill the following criteria:

- The global muon must have at least one good hit in the muon stations.
- The muon track must have more than 5 hits in the inner tracker and at least one pixel hit.
- $\chi^2/\text{ndof} < 10$ for the global muon track fit.
- Transverse impact parameter w.r.t. the primary vertex $|\text{d}_{xy}| < 0.04$ cm.
- Longitudinal impact parameter w.r.t. the primary vertex $|\text{d}_z| < 0.1$ cm.

4.4 Multivariate analysis

MC samples with no pileup (PU) are used for the multivariate analysis; however, the distribution of the input variables which are dependent on MET (variables involving MET), are shown with both PU and no PU to visualize the impact of PU on that variable. The input variables are chosen to exploit two key facts to separate signal from background: difference in spin of the Z and Higgs boson, and decay topology of the tau leptons (presence of neutrinos). The distributions of the input variables are normalized to unit area in order to compare the shapes

for the the three different event classes. The following input variables are used for the the multivariate analysis[10]:

- **Missing E_T.** The missing transverse energy distribution is shown in fig. 4.11. Missing E_T is calculated as:

$$\text{MET} = \sqrt{(MET_{p_x})^2 + (MET_{p_y})^2} \tag{4.2}$$

  For this analysis, Type-1 and muon corrected particle flow MET is used. The Type-I correction is a propagation of the jet energy corrections (JEC) to MET. The JEC is done by removing non-linear response of the calorimeter to jet energy at various η and E_T. Type-I correction replaces the vector sum of transverse momenta of particles which can be clustered as jets with the vector sum of the transverse momenta of the jets to which JEC is applied. Muon corrections to MET are done by adding the muon energy deposits in the calorimeter to MET (raw), and then subtracting the muon transverse momentum (which can be measured precisely in the muon chambers and the tracker) [27]. The distribution for the H$\rightarrow \tau\tau \rightarrow \mu\mu$ peaks at the highest value among the PU0 samples.This can be explained by the its production mechanism [28]. In gluon fusion, Higgs is produced alongside a high p_T jet, hence it has a finite transverse momentum. The $\tau^+\tau^-$ are ultra-relativistic, hence their decay products are almost collinear. This large momentum transfer produces an observable missing E_T in the transverse plane. In case of Z$\rightarrow \mu\mu$, most of the transverse momentum is carried by the two muons, hence the MET value for this channel is low. The MET contribution from the PU events explain the right shift in the Z$\rightarrow \mu\mu$ MET distribution with PU50.

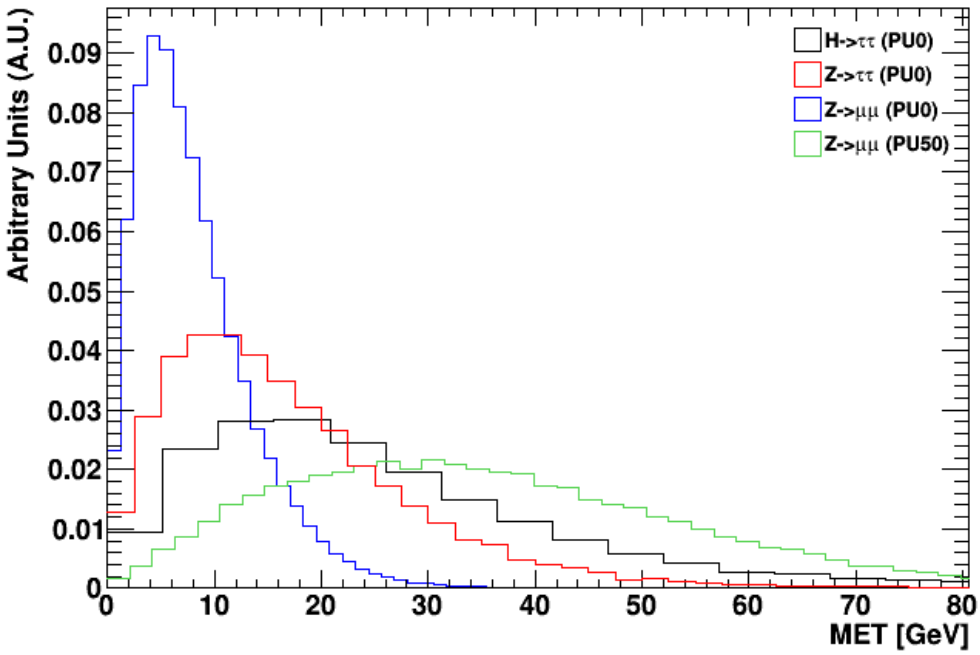


Figure 4.11: MET in GeV(linear scale).

- **The azimuth angle between the direction of the three-momentum of the positive muon and the missing transverse energy, $\Delta\Phi(\mu^+, MET)$.**

  Fig. 4.12 shows the difference in the azimuth angle between the direction of the MET and the direction of the three-momentum of the positive muon. The distribution for the Z$\rightarrow \mu\mu$ PU50 sample is shown along with the PU0 sample; from the distribution it seems that pileup has no influence on this variable. The shape of the distribution is explained by the decay topology of the taus. The decay products of the tau leptons (as a consequence of their ultra-relativistic nature) are almost collinear and not back to back, which explains why the peak at 3 radians (which roughly corresponds to 180 degrees) is much smaller compared to the peak near

zero. The distribution for Z→ $\mu\mu$ is almost flat. This discriminant is one of the better classifiers for the background studies 4.4.1

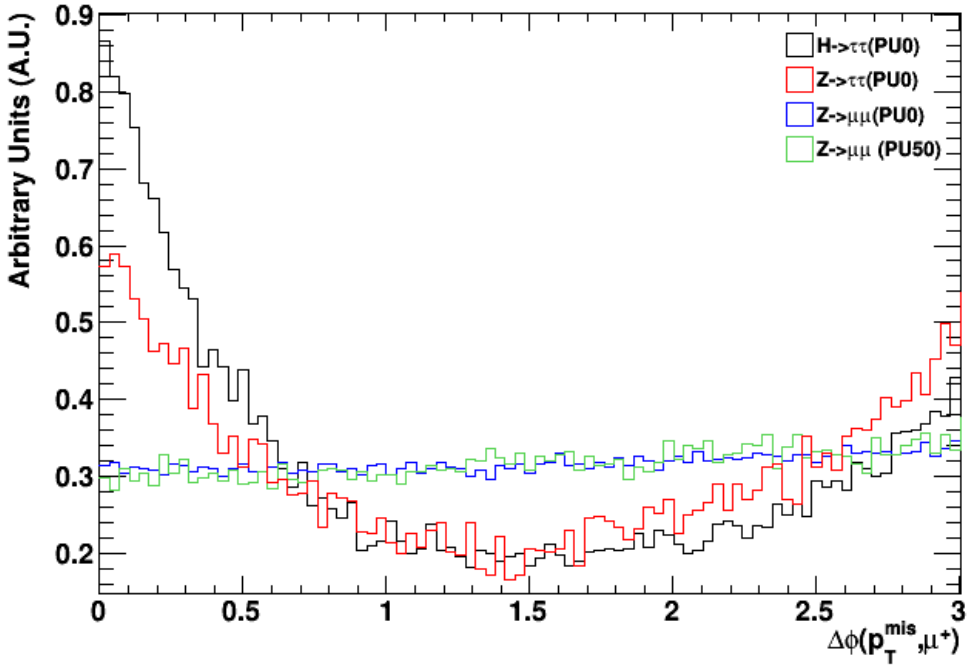


Figure 4.12: Azimuthal angle between the positive muon momentum and the missing transverse energy in radians (linear scale).

- **The muon distance of closest approach (DCA) significance, DCASig(μ) of d_{xy} and d_z for both the leading and the sub-leading muon.**

  DCA significance of the transverse (longitudinal) impact parameter d_{xy} (d_z) can be calculated as:

$$\text{DCASig}(\mu) \text{ of } d_{xy} = \frac{d_{xy}}{\sigma_{d_{xy}}} \tag{4.3}$$

  where d_{xy} is the transverse impact parameter of the muon and $\sigma_{d_{xy}}$ is the error in the impact parameter. The DCA significance of the longitudi-

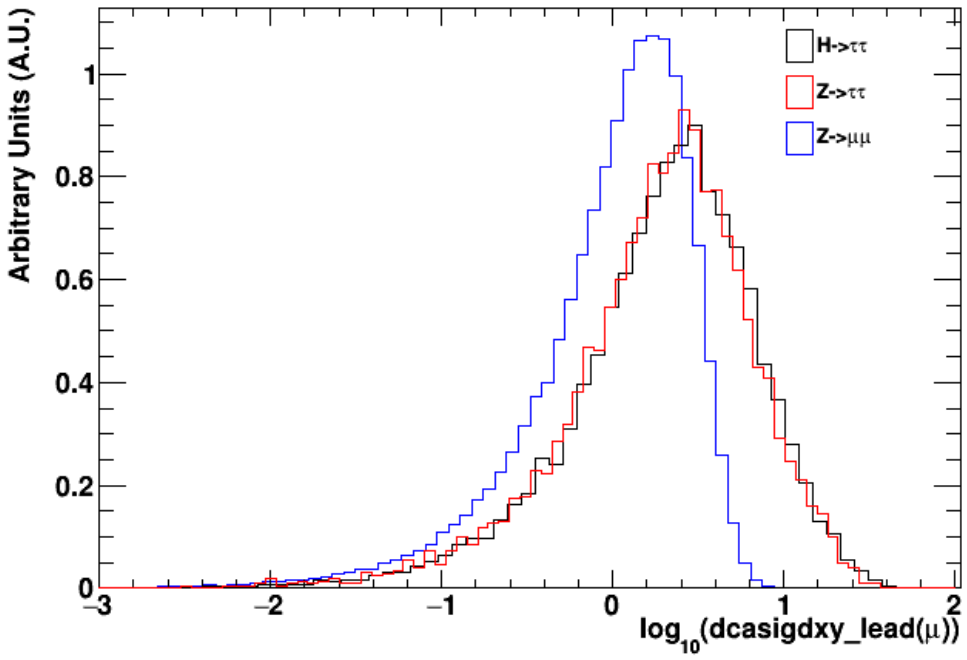


Figure 4.13: Common logarithm of the muon DCA significance of d_{xy} for the leading muons (linear scale).

nal impact parameter can be calculated by replacing d_{xy} and $\sigma_{d_{x,y}}$ with d_z and σ_{d_z}, respectively. Figs. 4.13, 4.14, 4.15 and 4.16 show the common logarithm of the DCASig for the transverse and longitudinal impact parameter for the MC samples. The longitudinal impact parameter for the decay channels involving an intermediate tau lepton are shifted to the right because of the tau lepton lifetime (293.2×10^{-15} seconds) [29].

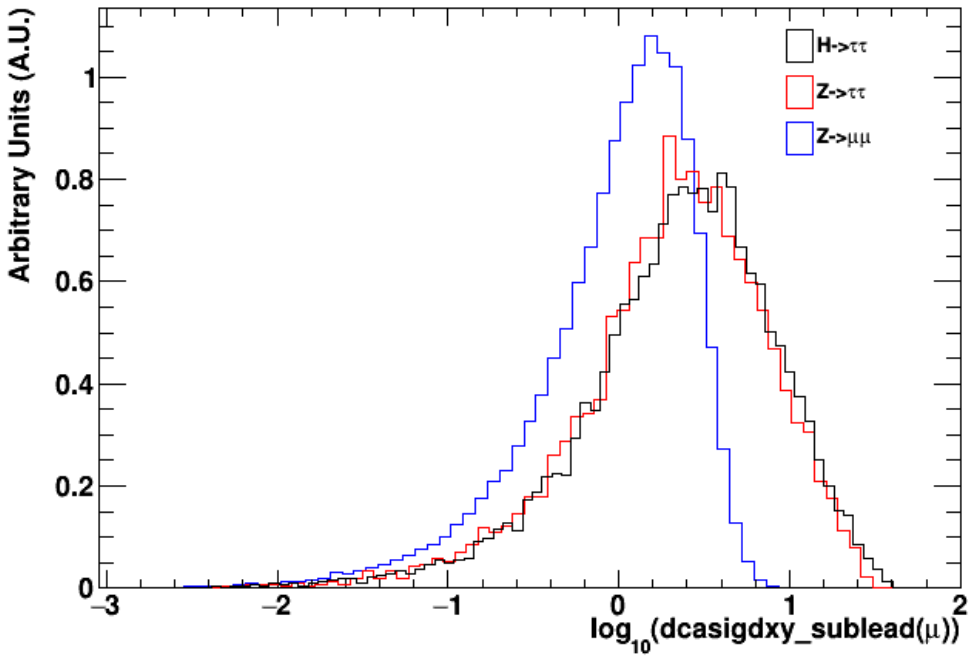


Figure 4.14: Common logarithm of the muon DCA significance of d_{xy} for the sub-leading muons (linear scale).

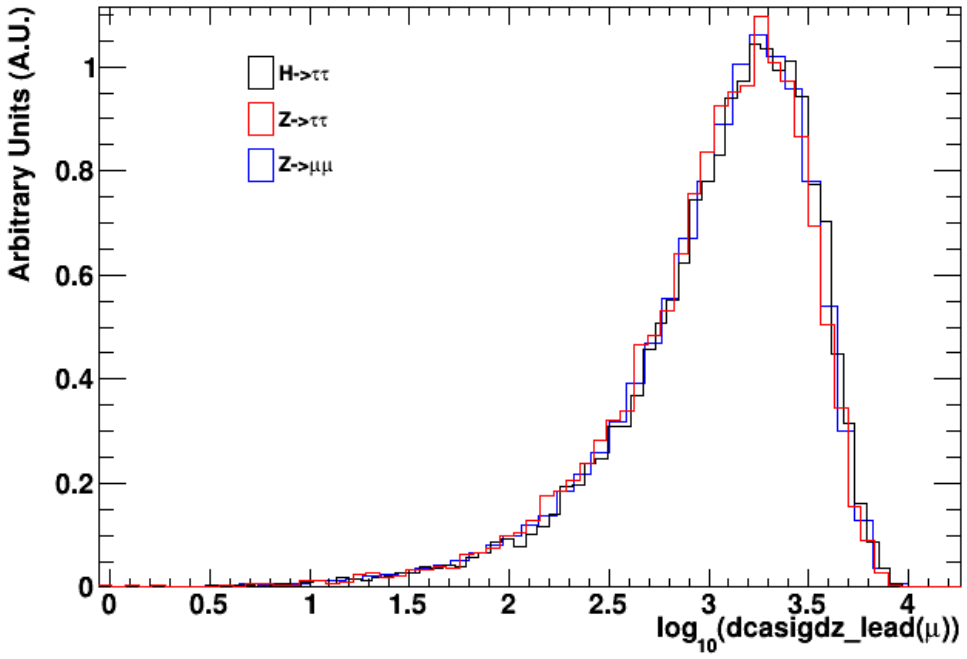


Figure 4.15: Common logarithm of the muon DCA significance of d_z for the leading muons (linear scale).

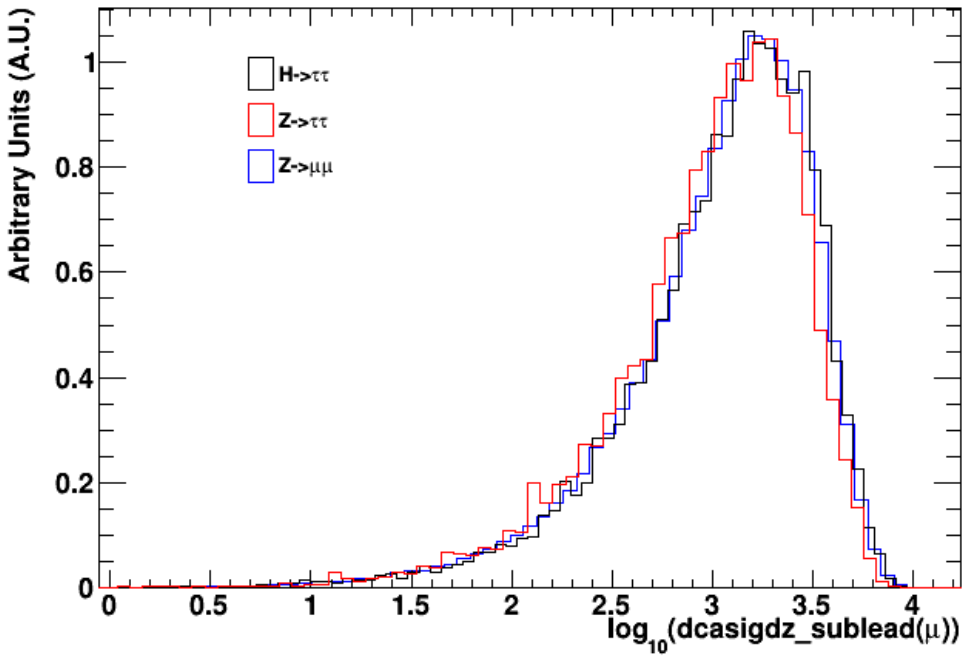


Figure 4.16: Common logarithm of the muon DCA significance of d_z for subleading muons (linear scale).

- **The ratio of the transverse momentum of the dimuon system to the scalar sum of the positive and negative muon momenta, $p_T(2\mu)/\sum p_T(\mu^i)$.**

  The transverse momentum of the dimuon system is defined as:

$$p_T(2\mu) = \sqrt{(p_x^+ + p_x^-)^2 + (p_y^+ + p_y^-)^2} \tag{4.4}$$

  where p_x^+ (p_x^-) and p_y^+ (p_y^-) are the x and y components of the positive (negative) muon momentum. Then, the ratio (r) can be calculated as:

$$\mathrm{r} = \frac{p_T(2\mu)}{p_T(\mu^+) + p_T(\mu^-)} \tag{4.5}$$

  where $p_T(\mu^+)$ ($p_T(\mu^-)$) is the transverse momentum of the positive (negative) muon. Figure 4.17 shows the distribution of this discriminant for the

pile-up zero signal (H$\rightarrow \tau\tau \rightarrow \mu\mu$) and background samples (Z$\rightarrow \tau\tau \rightarrow \mu\mu$ and Z$\rightarrow \mu\mu$). The Z$\rightarrow \mu\mu$ has the highest value for the scalar sum of the individual muons (roughly 40 GeV), while in the case of H$\rightarrow \tau\tau$ and Z$\rightarrow \tau\tau$, some amount of transverse momentum is taken away by the neutrinos, so the value of the scalar sum decreases (smaller denominator). This explains the peak at a smaller value for Z$\rightarrow \mu\mu$. H$\rightarrow \tau\tau$ has the largest value of MET, hence its peak is shifted to the right.

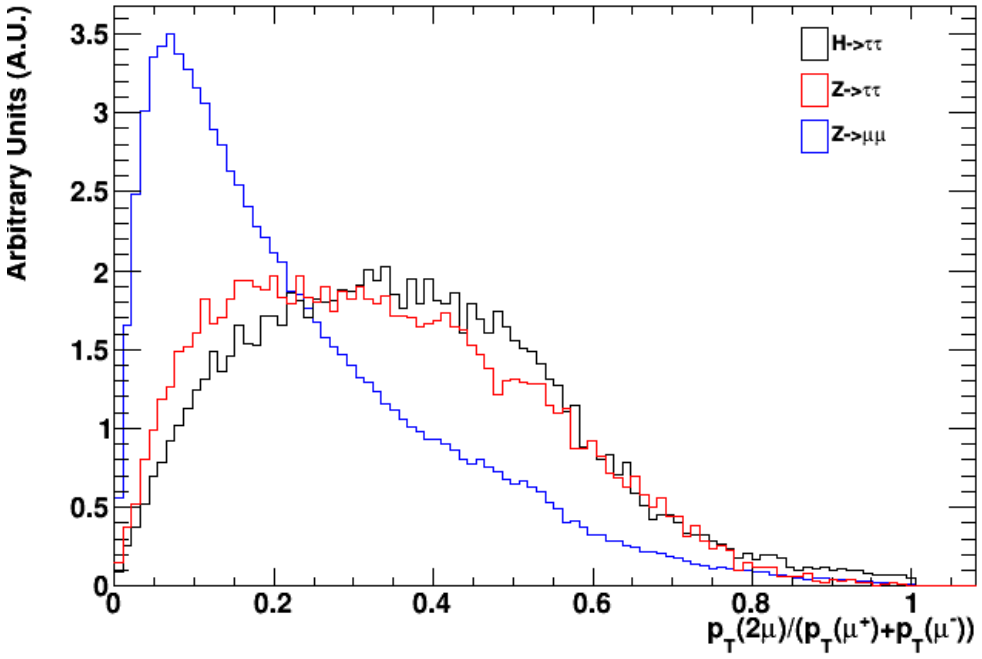


Figure 4.17: The ratio of the transverse momentum of the dimuon system to the scalar sum of muon transverse momenta (linear scale).

- **The pseudo-rapidity of the dimuon system, $\eta(2\mu)$.**

  Pseudo-rapidity of the dimuon system is shown in the figure 4.18, and is

calculated as:

$$\eta(2\mu) = \frac{1}{2} \times ln(\frac{|p(2\mu)| + p_z(2\mu)}{|p(2\mu)| - p_z(2\mu)}) \tag{4.6}$$

$|\mathrm{p}(2\mu)|$ is the magnitude of the three-momentum of the dimuon system and is calculated as:

$$|p(2\mu)| = \sqrt{(p_x^+ + p_x^-)^2 + (p_y^+ + p_y^-)^2 + (p_z^+ + p_z^-)^2} \tag{4.7}$$

p_z is the component along the beam-axis and can be calculated as:

$$p_z(2\mu) = p_z^+ + p_z^- \tag{4.8}$$

The dimuon η distribution for all the MC samples is roughly symmetrical about the origin; however the events in the Z$\rightarrow \mu\mu$ decay channel are more spread out compared to the other two channels. In case of H$\rightarrow \tau\tau$, and Z$\rightarrow \tau\tau$, some amount of momentum is taken by the neutrinos. which is not the case with the Z$\rightarrow \mu\mu$, hence the width of the distribution is inversely proportional to the MET value for the channel; this explains why Z$\rightarrow \mu\mu$ has the widest distribution and H$\rightarrow \tau\tau$ has the narrowest.

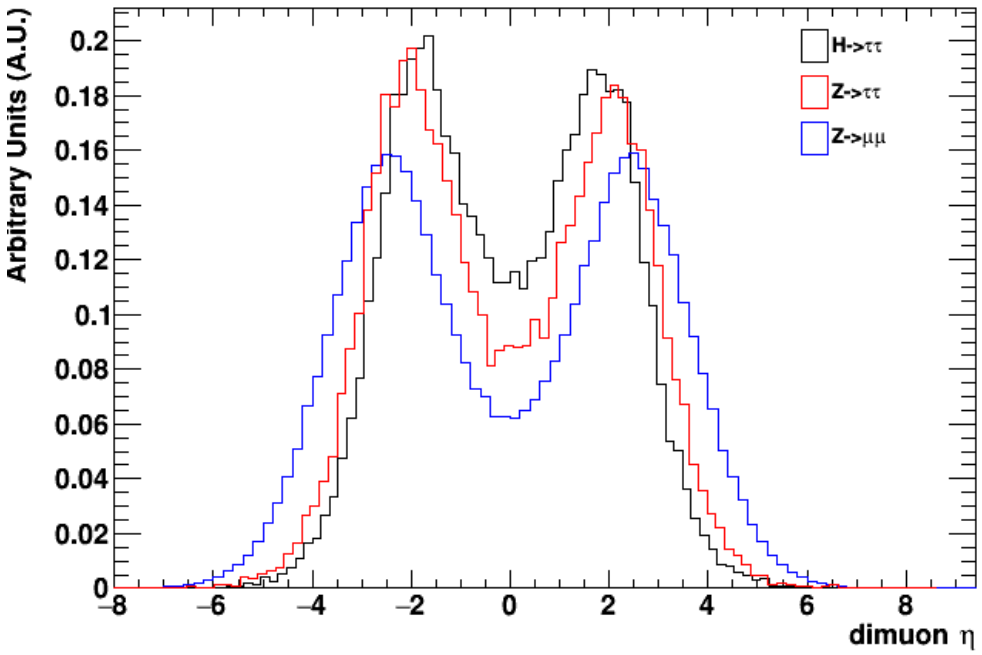


Figure 4.18: Pseudo-rapidity of the dimuon system.

- **The angle ω^* between three-momentum of the positively charged muon and production plane of the dimuon system.**

  The cosine of the angle ω^* is shown in the fig. 4.19 and is calculated as follows:

  - The individual muon and dimuon parameters in the center-of-mass (cms) frame are determined.
  - The muon four-momentum is then Lorentz boosted into the dimuon rest-frame.
  - The cosine of the angle is then determined by taking the dot product of the unit vector along the three-momentum of the positive muon and the unit vector along the direction of the dimuon rest-frame.

The Z→ $\tau\tau$ and the H→ $\tau\tau$ samples are fairly flat on an average; although the distribution does show fluctuations. The Z→ $\mu\mu$ PU0 and PU50 samples exhibit a drop at the boundary points, but the PU50 exhibits a more pronounced peak around the origin.

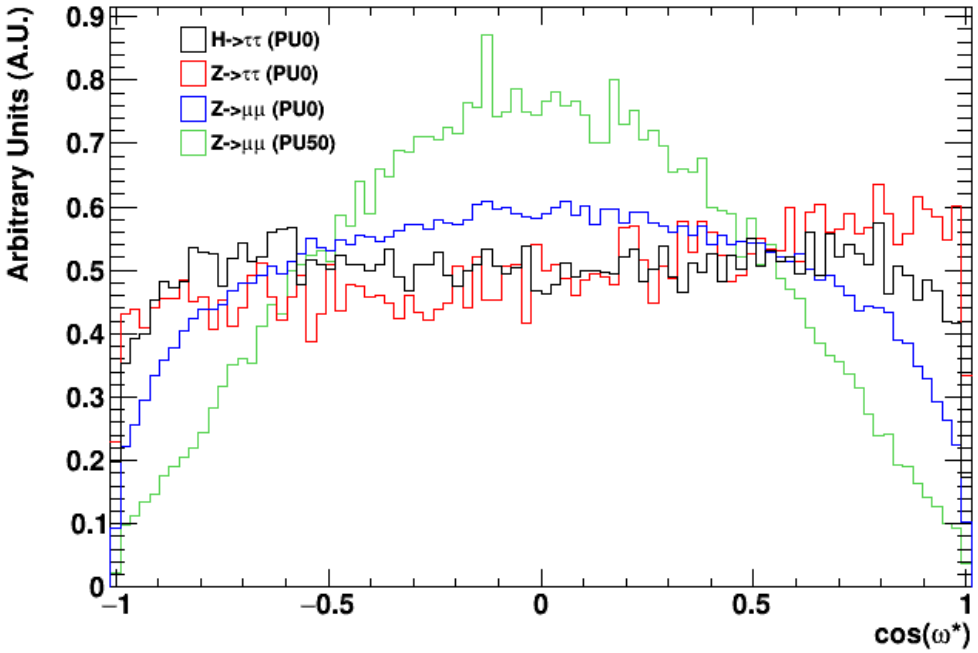


Figure 4.19: The angle ω^* between three-momentum of the positively charged muon and production plane of the dimuon system (linear scale).

These discriminating variables are given as input to the BDT, which is then used to distinguish between the event classes: H→ $\tau\tau \rightarrow \mu\mu$, Z→ $\tau\tau \rightarrow \mu\mu$ and Z→ $\mu\mu$.

4.4.1 BDT separation of Z$\rightarrow \tau\tau$ from Z$\rightarrow \mu\mu$

A BDT is constructed to distinguish between the event classes, Z$\rightarrow \tau\tau \rightarrow \mu\mu$ (signal) and Z$\rightarrow \mu\mu$ (background). Only the samples with zero pile-up are used here. Table 4.2 shows the ranking of the input variables. Figs. 4.20 and 4.21 show the correlation matrices for the variables. The DCA significance of the transverse impact parameter (d_{xy}) for the leading and sub-leading muon are the best discriminants for separating Z$\rightarrow \tau\tau$ from Z$\rightarrow \mu\mu$. The worst performing discriminants are the DCA Significance of longitudinal impact parameter (d_z) for the leading and the sub-leading muons, and they seem to be highly correlated as well; zero to negligible improvement in the separation power of the BDT is achieved by including both of them, hence only DCA Significance of the leading muon is used for the subsequent analysis.

Table 4.2: Method-specific ranking of discriminant variables for Z$\rightarrow \tau\tau$ (signal) and Z$\rightarrow \mu\mu$ (background). The short-hand notation for some of the discriminants are given in brackets.

Rank	Variable	Variable Importance
1	dcasig(μ) of d_{xy} for sub-leading muon (dxysl)	1.500×10^{-1}
2	dcasig(μ) of d_{xy} for leading muon (dxyl)	1.284×10^{-1}
3	MET	1.216×10^{-1}
4	$p_T(2\mu)/p_T(\mu^+)+p_T(\mu^-)$ (pT)	1.203×10^{-1}
5	cos(ω^*) (cosd)	1.129×10^{-1}
6	$\Delta\phi(p_T^{\mathrm{mis}}, \mu^+)$ (phi)	1.084×10^{-1}
7	$\eta(2\mu)$ (eta)	9.184×10^{-2}
8	DCAsig(μ) of d_z for leading muon (dzl)	8.470×10^{-2}
9	DCAsig(μ) of d_z for sub-leading muon (dzsl)	8.190×10^{-2}

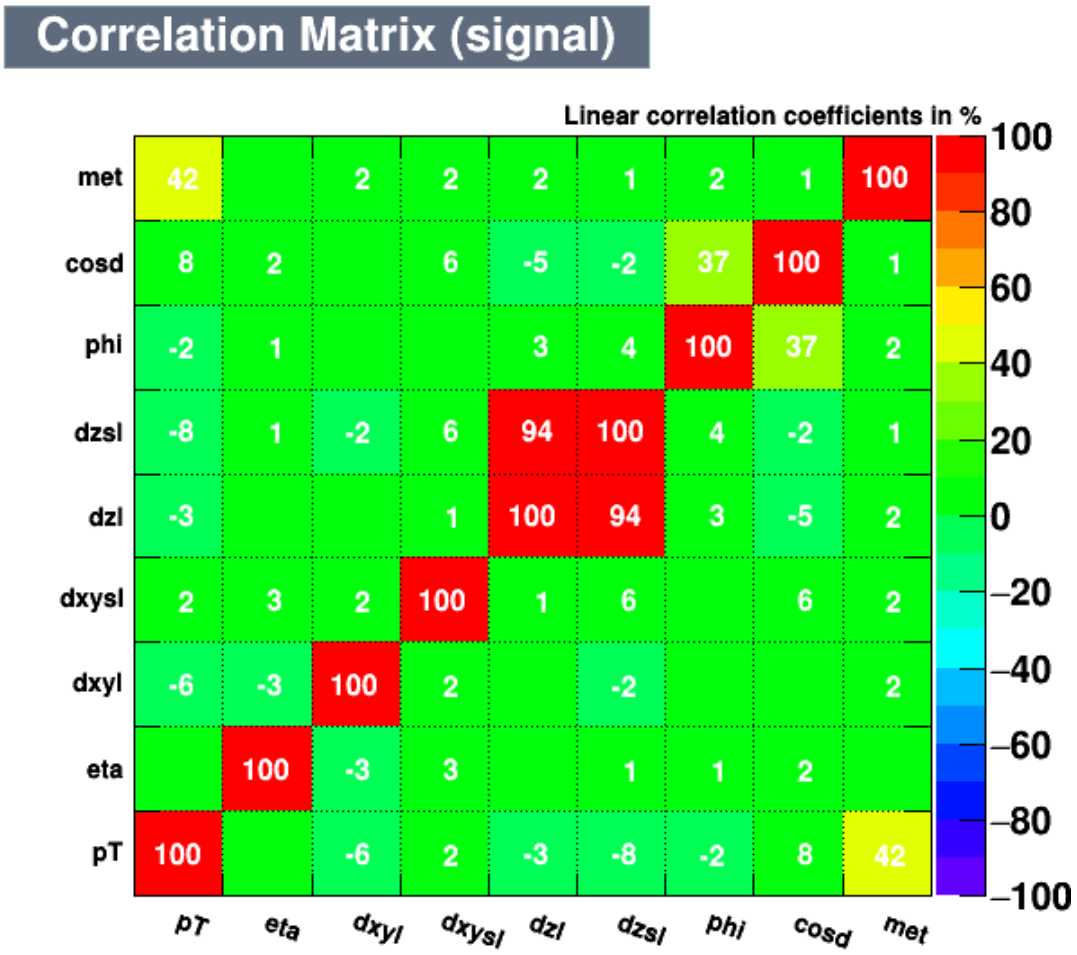


Figure 4.20: Correlation matrix for the Z$\rightarrow$ $\tau\tau$ $\rightarrow$ $\mu\mu$ signal sample. The linear correlation coefficients are given in %. The DCASig(μ) of the d_z for the leading and the sub-leading muons are highly correlated (94 %). The short-hand notations used in this figure are described in table 4.2.

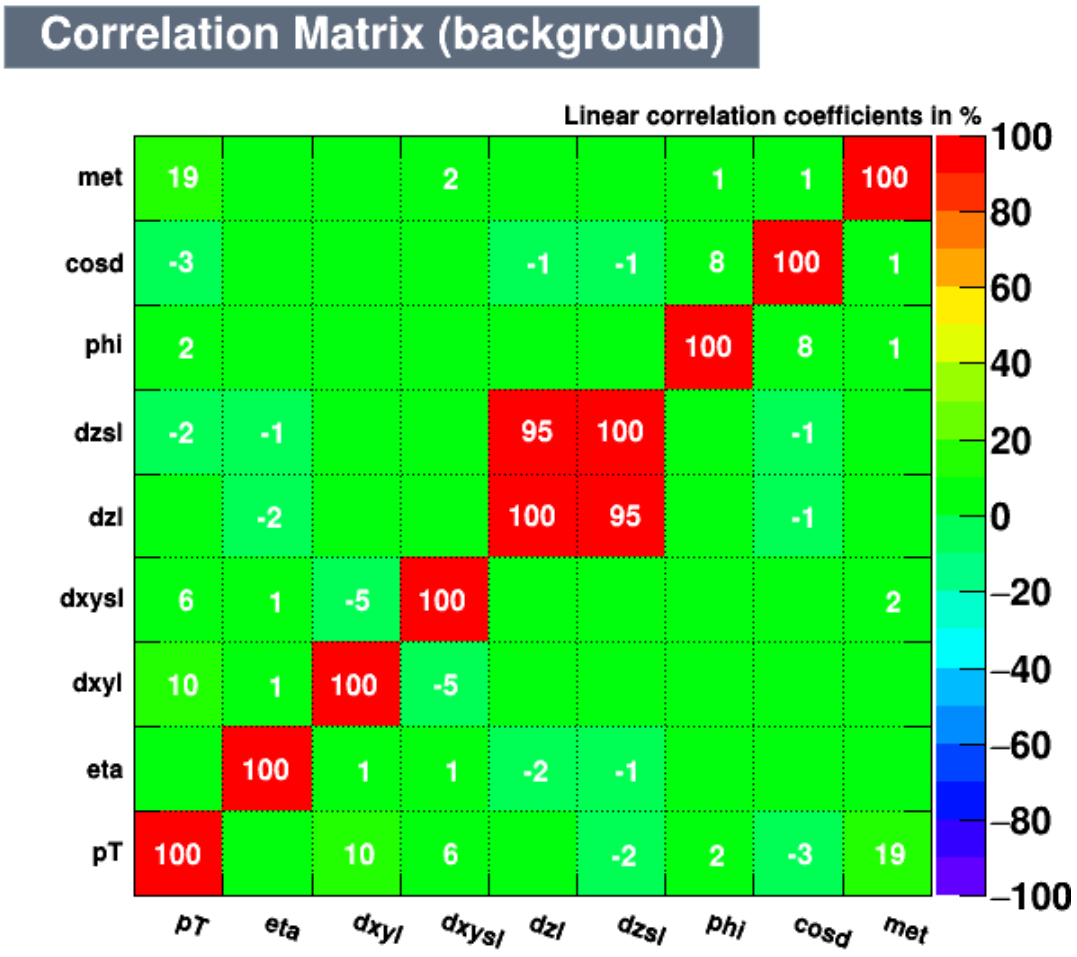


Figure 4.21: Correlation matrix for the Z→ $\mu\mu$ background sample. The linear correlation coefficients are given in %. The DCASig(μ) of the d_z for the leading and the sub-leading muons are highly correlated (95 %). The short-hand notations used in this figure are described in table 4.2.

The performance of the BDT is gauged by looking at the Receiver Operating Characteristics (ROC) curve. A ROC curve is a graphical plot which illustrates the performance of a binary classifier as its discrimination threshold is varied. It is created by plotting the background rejection (1-false positive rate) against

signal efficiency (true positive rate) at various threshold settings. The value of the ROC integral (measure of the area under the ROC curve) gives a good quantitative idea about the performance of the decision tree. A value closer to 1 (classifier has a zero misclassification rate; all signals are classified as signals and all backgrounds are classified as backgrounds) implies good performance, while a value closer to 0.5 (half of the signal events are misclassified as background events and vice-versa) implies random guessing (The probability of the classifier classifying an event correctly is 0.5, which is the probability of getting a head when tossing a coin).

The discriminatory power of a BDT depends on the shape of the input variable distributions. The value of the ROC integral (area under the ROC curve) for discriminating Z$\rightarrow \tau\tau$ from Z$\rightarrow \mu\mu$ is summarized in the table 4.3. Figure 4.22 shows the ROC curve.

Table 4.3: ROC integral values.

Signal	Background	ROC Integral value
Z$\rightarrow \tau\tau$	Z$\rightarrow \mu\mu$	0.913

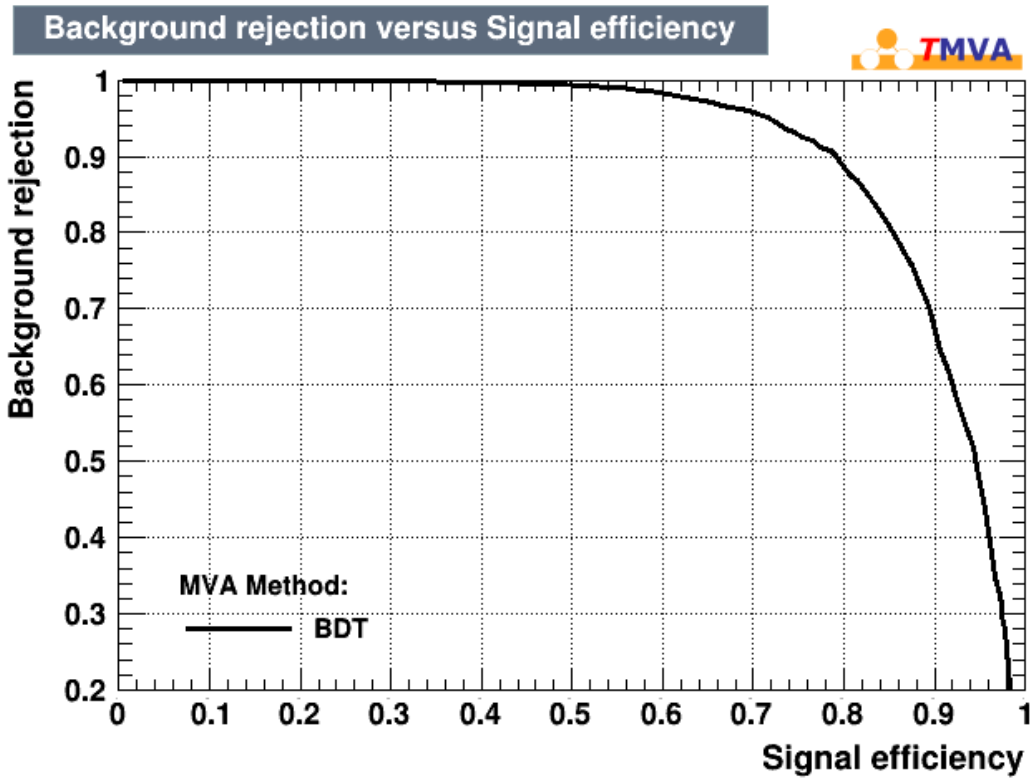


Figure 4.22: ROC curve for Z$\rightarrow \tau\tau \rightarrow \mu\mu$ (signal) and Z$\rightarrow \mu\mu$ (background).

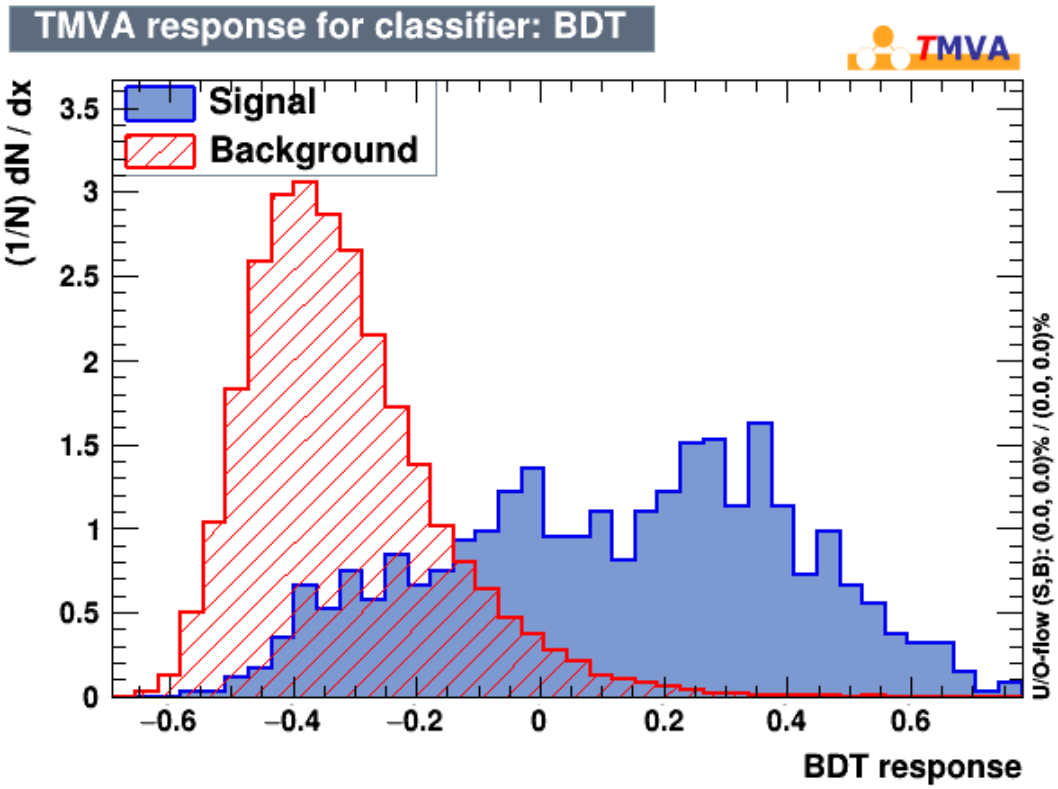


Figure 4.23: BDT discriminant distribution for the signal (Z→ $\tau\tau$) and the background (Z→ $\mu\mu$).

The MVA classifier response is compared for the signal and the background in fig. 4.23. The MVA response is calculated as [19]:

$$\text{MVA classifier response} = \frac{1}{N_{collection}} \times \sum_{i}^{N_{collection}} ln(\alpha_i) \cdot h_i(x) \qquad (4.9)$$

x is the n-tuple of the input variables (say dimuon eta, MET, etc.). $N_{collection}$ is the total number of classifiers (in our case total number of BDTs) in the collection. α is the error rate (misclassification rate). The value of h(x) is +1 for signal and -1 for background. Small (large) values for MVA classifier response indicate a background-like (signal-like) event. This output (response) can be used to put a cut to increase the signal purity in a sample with unknown

signal and background composition. Fig 4.24 shows the signal and background efficiencies along with the purities, versus the cut value. Signal (background) efficiency is defined as:

$$\text{Signal (background) efficiency} = \frac{\text{Number of selected signal (background) events}}{\text{Number of true signal (background) events}} \tag{4.10}$$

and signal purity is defined as:

$$\text{Signal purity} = \frac{\text{Number of true signal events}}{\text{Total number of events}} \tag{4.11}$$

Also shown in the fig. 4.24 are the significance and the signal efficiency times signal purity. The optimal cut is shown in the table 4.4.

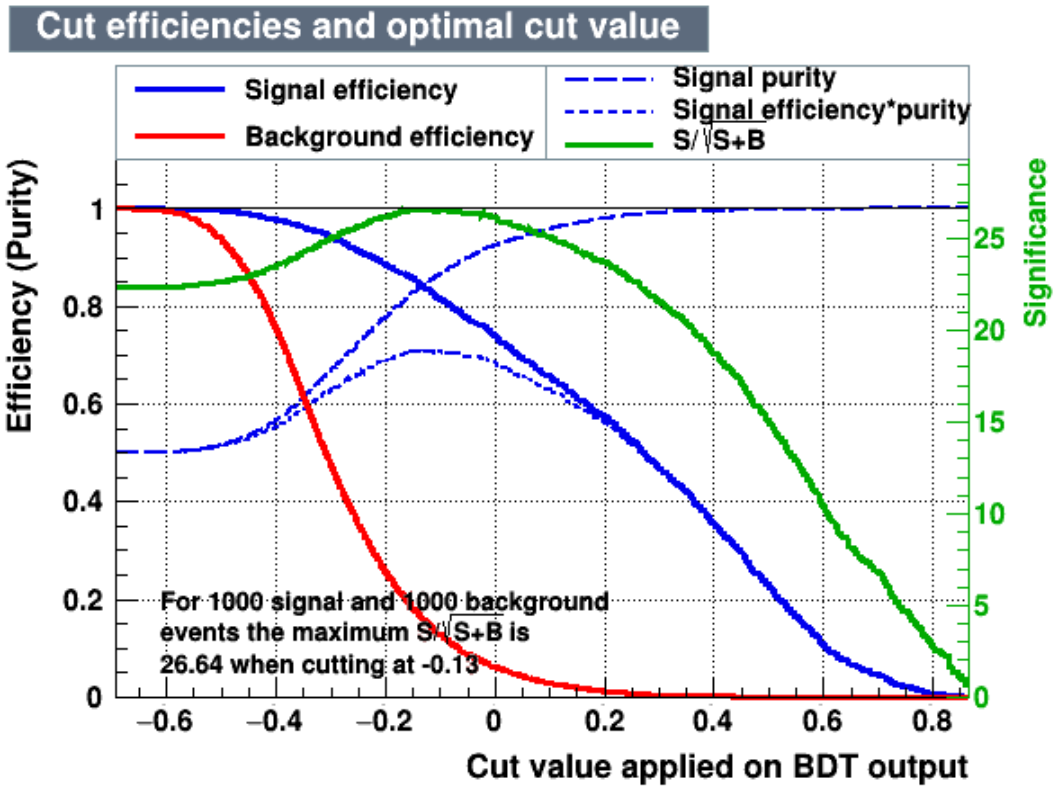


Figure 4.24: Signal Z($\rightarrow \tau\tau$) and background (Z$\rightarrow \mu\mu$) efficiency, signal purity, signal efficiency times purity and resulting significance when cutting on BDT output value.

Table 4.4: BDT cut value for the maximum significance.

Cut value	Significance
-0.13	26.64

Fig. 4.25 shows the invariant dimuon mass on a log scale for Z$\rightarrow \tau\tau$ (signal) and Z$\rightarrow \mu\mu$ (background) after pre-selection. We can use the optimal BDT cut value to increase the signal purity of the data sample. Fig. 4.26 shows the invariant dimuon mass on a log scale for Z$\rightarrow \tau\tau$ (signal) and Z$\rightarrow \mu\mu$ (background) for events that have a BDT response value greater than -0.13. It is important to note that these plots are shown here only to illustrate the effect of the BDT cuts on the signal purity of the data samples. The data samples do not reflect the actual cross-section $\times$ branching ratio (BR).

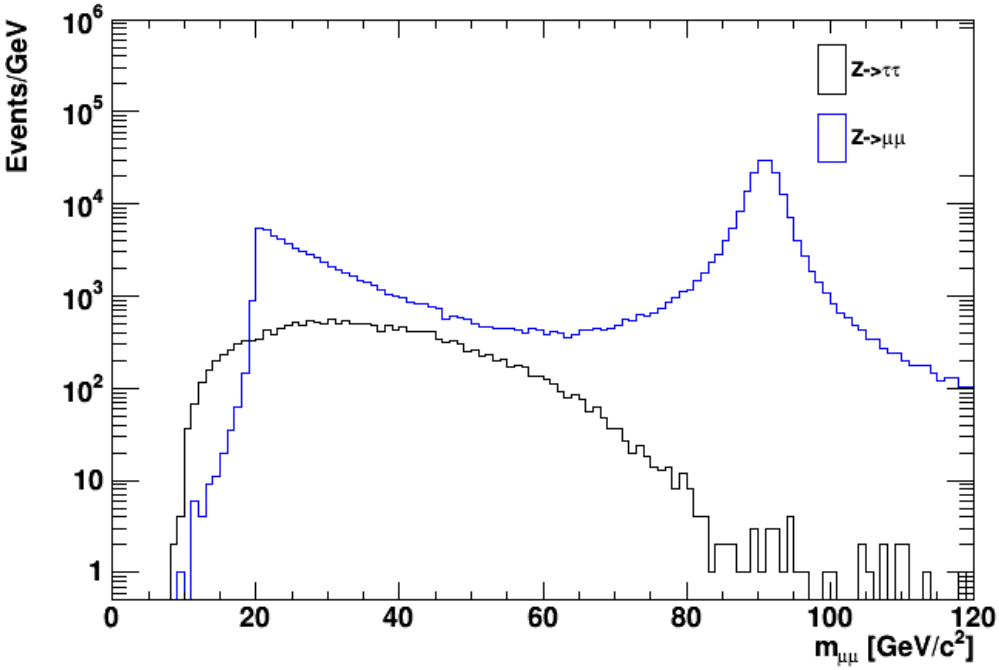


Figure 4.25: Dimuon invariant mass on a log scale for signal (Z$\rightarrow \tau\tau$) and background (Z$\rightarrow \mu\mu$) after pre-selection cuts. No BDT cuts have been applied. The plots are shown for illustrating the effect of the BDT selection. The samples do not reflect the actual cross-section $\times$ branching ratio of the samples.

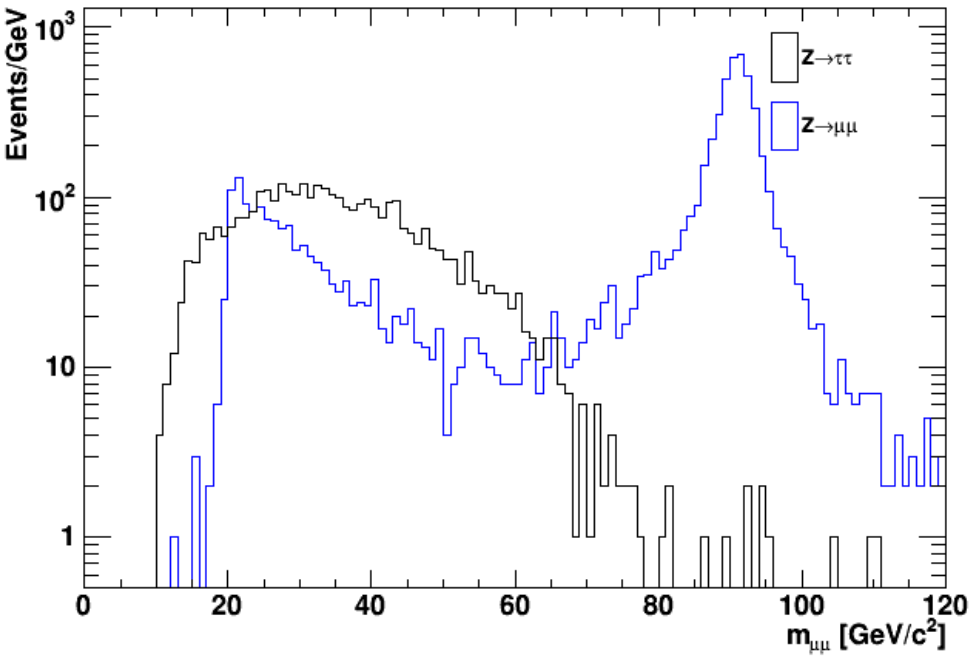


Figure 4.26: Dimuon invariant mass on a log scale for signal (Z→ $\tau\tau$) and background (Z→ $\mu\mu$) after the BDT cut (-0.13). The plots are shown for illustrating the effect of the BDT selection. The samples do not reflect the actual cross-section × branching ratio of the samples.

In fig. 4.25, it is seen that for the mass region greater than 20 GeV, the dimuon events from Z→ $\mu\mu$ form a smooth background over the Z→ $\tau\tau$ events. However, if only dimuon events that pass the BDT cut are selected, then we can clearly see a bump in the mass region between 20 GeV and 60 GeV, as shown in fig. 4.26.

4.4.2 BDT separation of H→ $\tau\tau$ from Z→ $\tau\tau$

The Z→ $\tau\tau$ is a major irreducible background. The final state topology for H→ $\tau\tau$ → $\mu\mu$ (signal) and Z→ $\tau\tau$ → $\mu\mu$ (background) event classes are both characterized by two muons and four neutrinos. Only the samples with zero

pile-up are used here. A BDT is constructed to distinguish between them. Table 4.5 shows the ranking of the input variables. Figs. 4.27 and 4.28 show the correlation matrices for the variables. Missing transverse energy (MET) is the best discriminant for separating H$\rightarrow \tau\tau$ (signal) from Z$\rightarrow \tau\tau$ (background). The worst performing discriminants are the DCA Significance of longitudinal impact parameter (d_z) for the leading and the sub-leading muons, and they seem to be highly correlated; zero to negligible improvement in the separation power of the BDT is achieved by including both of them, hence only the DCA Significance of the leading muon is used for the subsequent analysis.

Table 4.5: Method-specific ranking of the discriminant variables for H$\rightarrow \tau\tau$ (signal) and Z$\rightarrow \tau\tau$ (background).

Rank	Variable	Variable Importance
1	MET	1.431×10^{-1}
2	$\cos(\omega^*)$	1.235×10^{-1}
3	$\eta(2\mu)$	1.177×10^{-1}
4	$\Delta\phi(p_T^{\text{mis}}, \mu^+)$	1.138×10^{-1}
5	$p_T(2\mu)/p_T(\mu^+)+p_T(\mu^-)$	1.090×10^{-1}
6	dcasig(μ) of d_{xy} for leading muon	1.045×10^{-1}
7	dcasig(μ) of d_{xy} for sub-leading muon	1.043×10^{-1}
8	dcasig(μ) of d_z for leading muon	9.360×10^{-2}
9	dcasig(μ) of d_z for sub-leading muon	9.045×10^{-2}

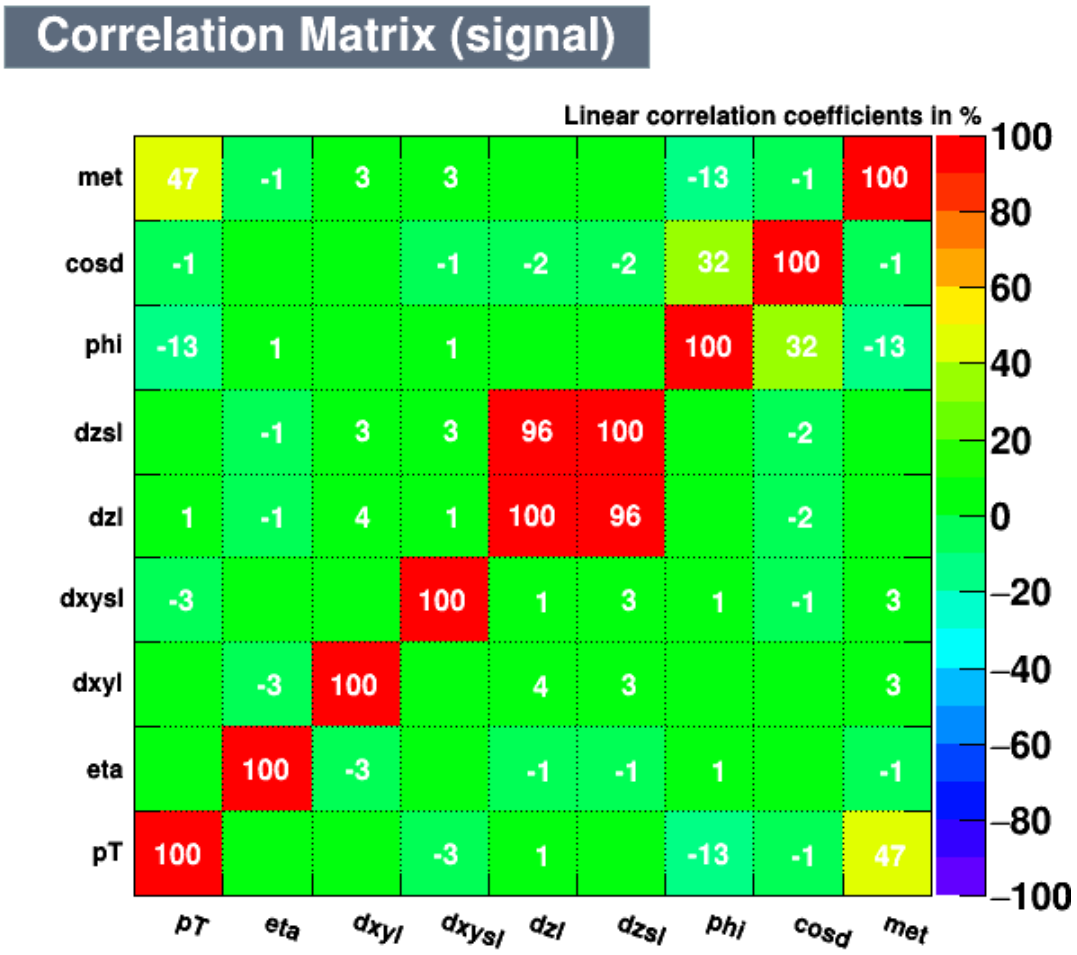


Figure 4.27: Correlation matrix for the H$\rightarrow \tau\tau$ signal sample. The linear correlation coefficients are given in %. The DCASig(μ) of the d_z for the leading and the sub-leading muons are highly correlated (96 %). The short-hand notations used in this figure are described in table 4.2.

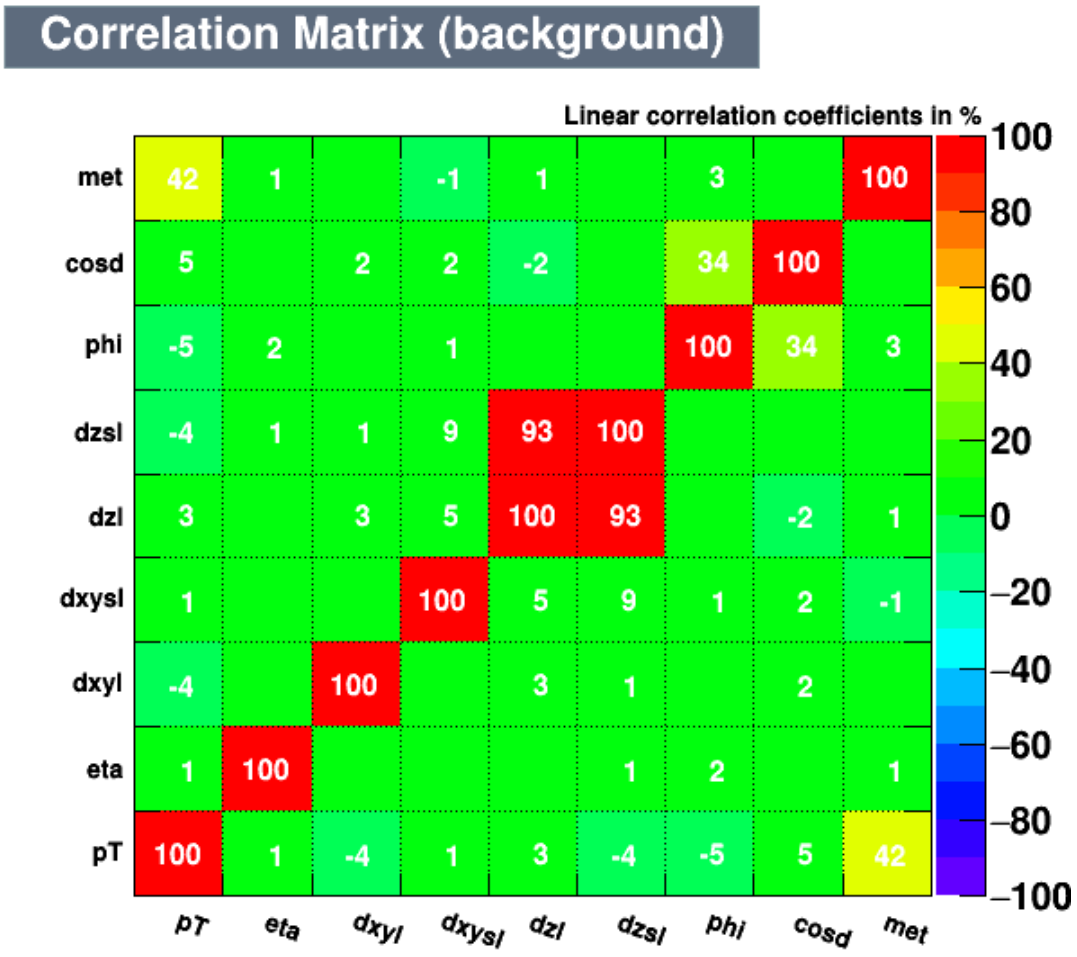


Figure 4.28: Correlation matrix for the Z→ $\tau\tau$ background sample. The linear correlation coefficients are given in %. The DCASig(μ) of the d_z for the leading and the sub-leading muons are highly correlated (93 %). The short-hand notations used in this figure are described in table 4.2.

The performance is gauged by looking at the ROC curve. The BDT is much more efficient at separating events between Z→ $\tau\tau \rightarrow \mu\mu$ and Z→ $\mu\mu$ compared to H→ $\tau\tau$ and Z→ $\tau\tau$. The value of the ROC integral (area under the ROC curve) is summarized in table 4.6. Figure [4.29 shows the ROC curve. The MVA

classifier response is compared for the signal and the background in fig. 4.30.

Table 4.6: ROC integral values.

Signal	Background	ROC Integral value
H$\rightarrow \tau\tau$	Z$\rightarrow \tau\tau$	0.671

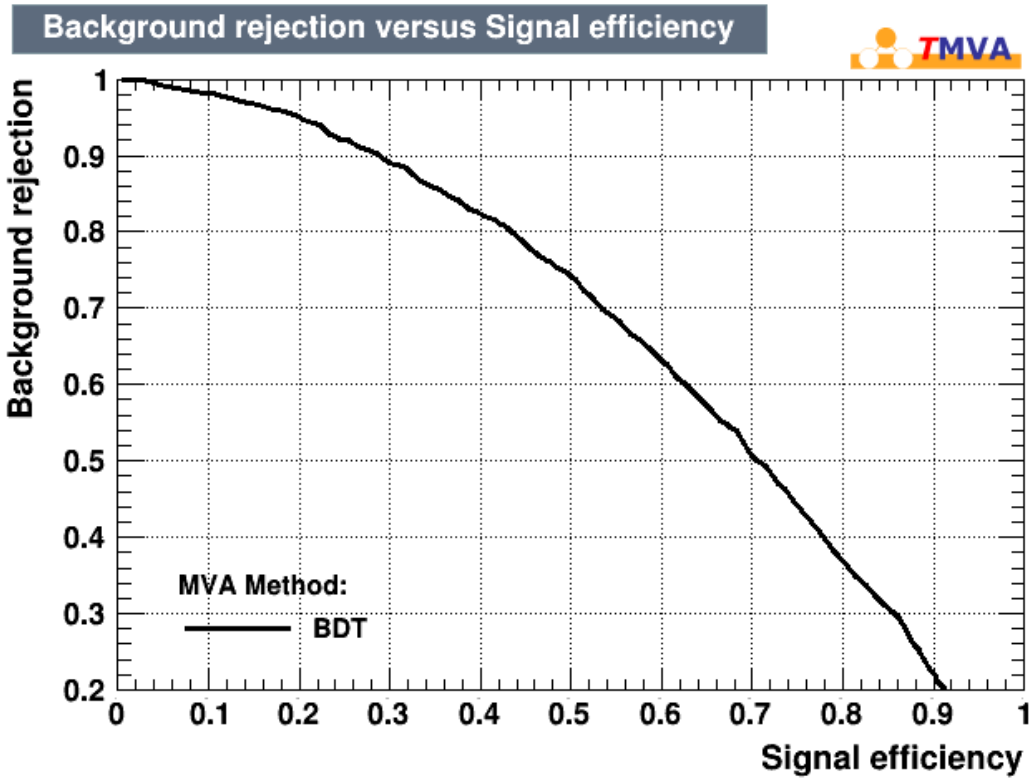


Figure 4.29: ROC Curve for H$\rightarrow \tau\tau$ (signal) and Z$\rightarrow \tau\tau$ (background).

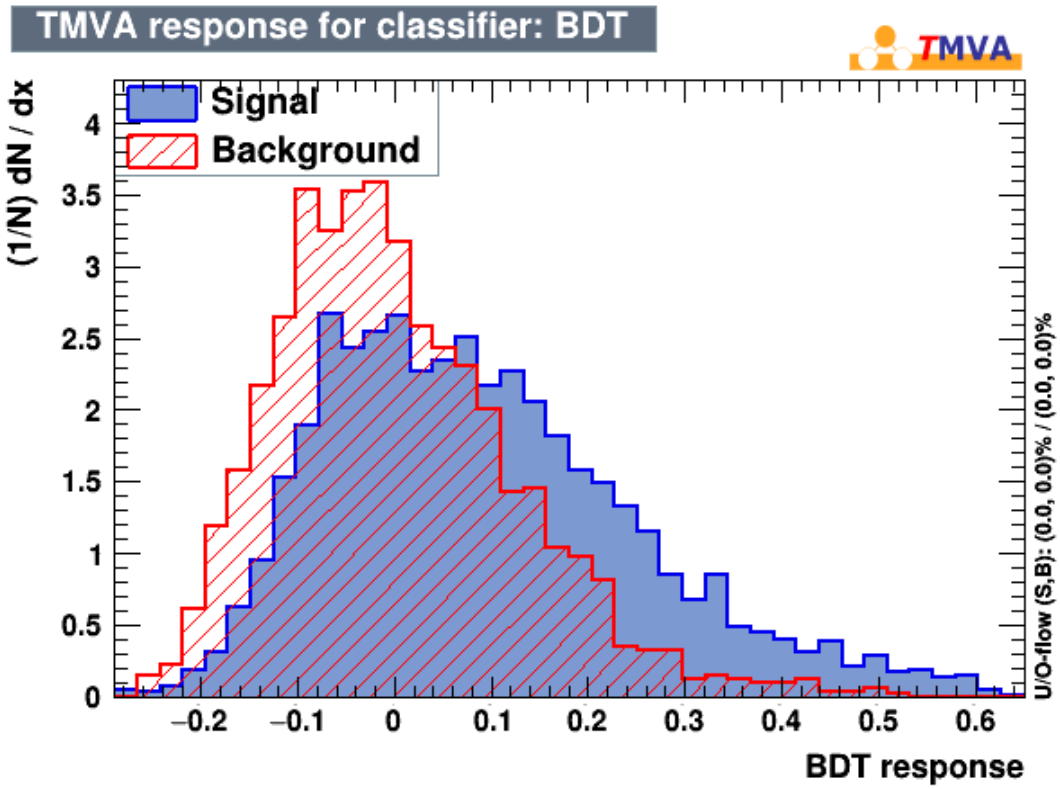


Figure 4.30: BDT discriminant distribution for the signal (H$\rightarrow \tau\tau$) and the background (Z$\rightarrow \tau\tau$).

This output can be used to put a cut to increase the signal purity in a sample with unknown signal and background composition. Fig 4.31 shows the signal and the background efficiencies along with the purities, the significance and the signal efficiency times signal purity versus the cuts. The optimal cut is shown in the table 4.4.

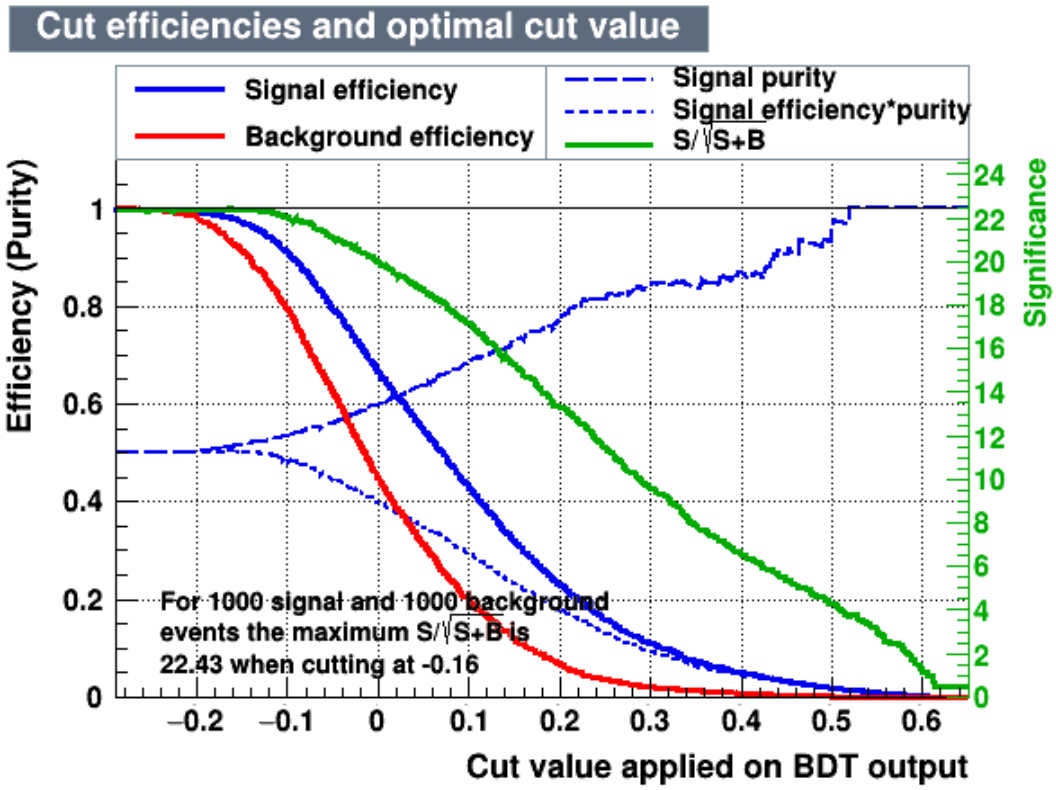


Figure 4.31: Signal (H→ $\tau\tau$) and background (Z→ $\tau\tau$) efficiency, signal purity, signal efficiency times purity and resulting significance when cutting on BDT output value.

Table 4.7: BDT cut value for the maximum significance.

Cut value	Significance
-0.16	22.43

Fig. 4.32 shows the invariant dimuon mass on a log scale for H→ $\tau\tau$ (signal) and Z→ $\tau\tau$ (background) after pre-selection. We can use the optimal BDT cut value to select only those events that have a BDT value greater than the optimal cut and plot the invariant dimuon mass for those events. Fig. 4.33

shows the invariant dimuon mass on a log scale for H$\rightarrow \tau\tau$ (signal) and Z$\rightarrow \tau\tau$ (background) for events that have a BDT response value greater than -0.16. It is important to note that these plots are shown here mainly to illustrate the effect of the BDT cuts on the signal purity of the data samples. The data samples do not reflect the actual cross-section $\times$ branching ratio (BR).

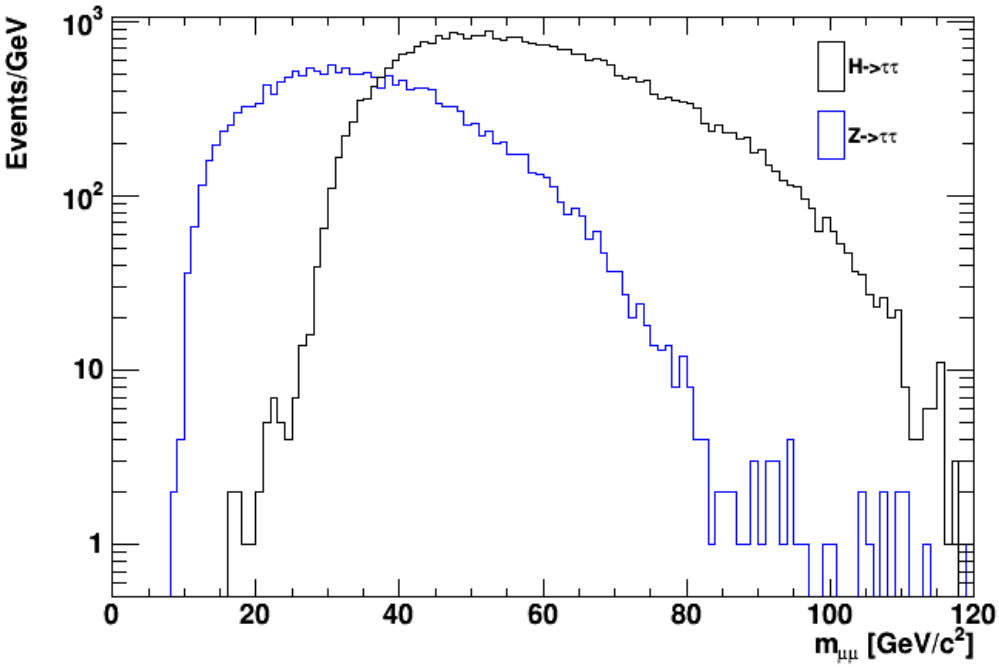


Figure 4.32: Dimuon invariant mass on a log scale for signal (H$\rightarrow \tau\tau$) and background (Z$\rightarrow \tau\tau$) after pre-selection cuts. No BDT cuts have been applied. The plots are shown for illustration of the effect of the BDT selection. The samples do not reflect the actual cross-section $\times$ branching ratio of the samples.

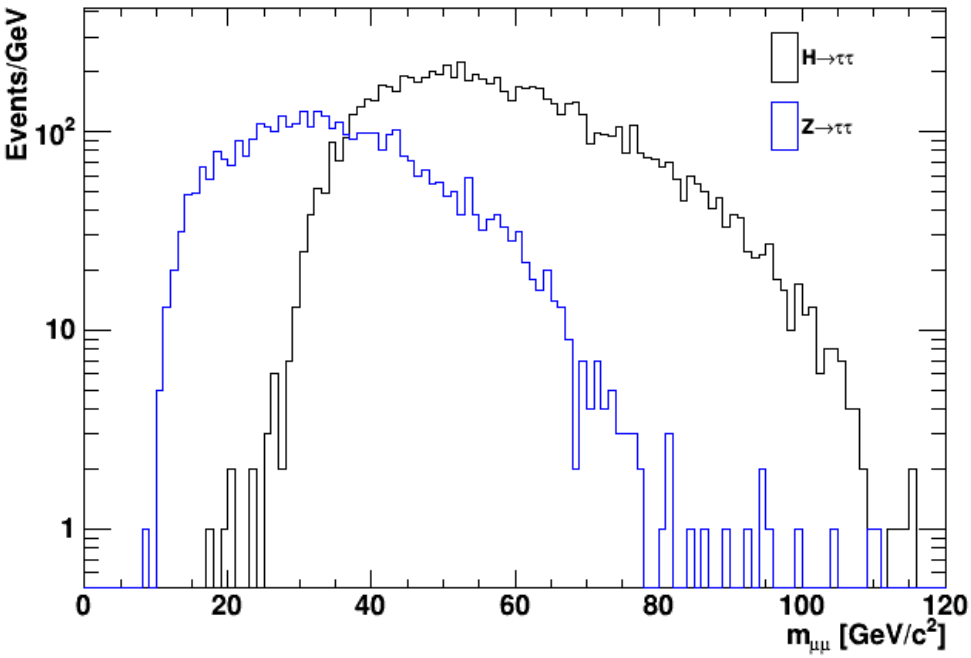


Figure 4.33: Dimuon invariant mass on a log scale for signal (H→ $\tau\tau$) and background (Z→ $\tau\tau$) after the BDT cut (-0.16). The plots are shown primarily for illustration of the effect of the BDT selection. The samples do not reflect the actual cross-section × branching ratio of the samples.

Chapter 5

Summary and Conclusion

This master's thesis describes an analysis of the H$\rightarrow \tau^+\tau^- \rightarrow \mu^+\mu^-$ channel for the CMS experiment at the LHC. The high DY background and the small branching fraction of the $\tau\tau \rightarrow \mu\mu$ (3%) make the $\mu\mu$ channel very challenging; hence the need for increased event acceptance and multivariate analysis in this channel.

The effect of lowering the dimuon p_T threshold on the event acceptance was studied for the leading and the sub-leading muon in different eta regions. In all the cases, lowering the dimuon p_T trigger threshold from (20 GeV, 10 GeV) to (12 GeV, 5 GeV), respectively doubles the acceptance of events. The GE1/1 upgrade should allow us to do that while keeping the low L1 p_T trigger rate in check. It was found that the event acceptance was more sensitive to changes in the sub-leading muon p_T thresholds than to changes in the leading muon p_T.

Two boosted decision trees (BDTs) were trained to distinguish between the three event classes: H$\rightarrow \tau^+\tau^-$, Z$\rightarrow \tau^+\tau^-$ and Z$\rightarrow \mu^+\mu^-$. The decision tree that was used to separate the Z$\rightarrow \tau^+\tau^-$ (signal) and Z$\rightarrow \mu^+\mu^-$ (background) events, had a ROC curve integral value of 0.907, while the second decision tree which was trained to separate H$\rightarrow \tau^+\tau^-$ (signal) and Z$\rightarrow \tau^+\tau^-$ (background) had a

ROC curve integral value of 0.671. This result can be interpreted as follows: If random pairs of signal and background events are given to the first decision tree, then it would classify the pairs correctly 90% of the time, whereas the second decision tree would correctly classify 67% of the random pair of events. The significance was calculated as a function of the BDT variable for the large-size Monte Carlo samples used.

- A significance ($S/\sqrt{S+B}$) of 26.6 was found with a cut of -0.13 on the BDT output for the event class: $Z \rightarrow \tau^+\tau^-$ (signal) and $Z \rightarrow \mu^+\mu^-$ (background),

- A significance of 22.4 was found with a cut value of -0.16 for the event class: $H \rightarrow \tau^+\tau^-$ (signal) and $Z \rightarrow \tau^+\tau^-$ (background).

Cuts tighter than the quoted values should be typically used on the Run-3 data because of the small expected signal size. Finally, the effect of the BDT selection on signal purity was illustrated by comparing the dimuon invariant mass before and after applying the optimal BDT for the two different analyses.

Appendix A

MC datasets

Physics process	COM energy (TeV)	Number of events generated	Simulated pileup	GEMs used	Data Format	Official?
Gluon fusion to Higgs, Higgs decaying to Tau	13	462681	Yes (PU40)	No	AODSIM	Yes
Gluon fusion to Higgs Higgs decaying to Tau Tau decaying to muons	14	200000	No	No	RECOSIM	No
Drell Yan to Tau	14	999772	Yes (PU50)	Yes	AODSIM	Yes
Drell Yan to Mu	14	999772	Yes (PU50)	Yes	AODSIM	Yes
Drell Yan to Mu	14	993597	No	Yes	GEN-SIM	Yes
					-RECO	
Drell Yan to Tau, Tau decaying to muons	14	199700	No	Yes	RECOSIM	No

Table A.1: Monte Carlo datasets.

Appendix B

Comparison of the different MVA methods

Various MVA methods are provided by the TMVA package. In order to achieve the best separation between signal and background, different MVAs were trained, tested and evaluated. The MVAs used for the comparison studies are:

- LD
- BDT
- PDERS
- MLPBNN
- Likelihood
- RuleFit
- SVM
- KNN
- FDA_GA

- Cuts
- LikelihoodPCA
- CutsD

Figs. B.1 and B.1 show the ROC curves for the different MVA techniques. It is seen that the BDTs perform better than most MVA methods except LD.

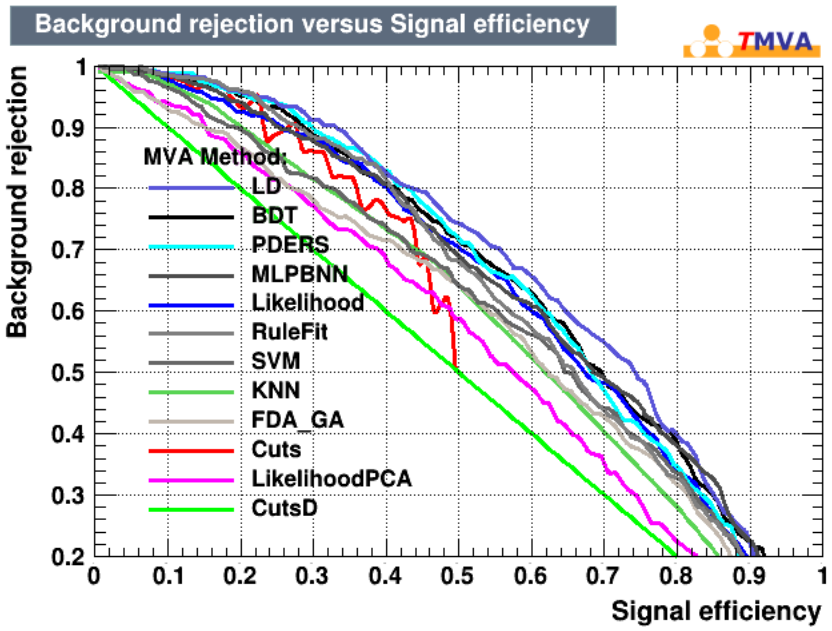
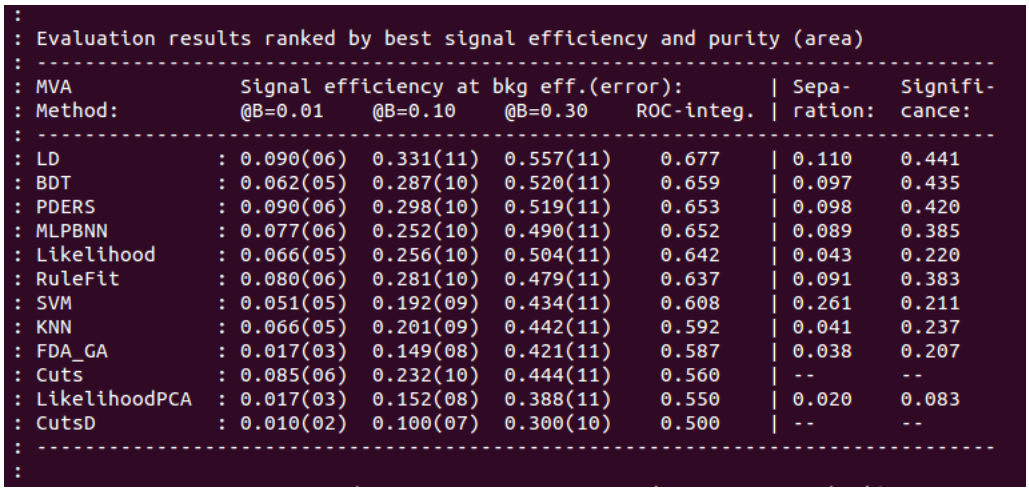


Figure B.1: ROC curve for the different MVAs for signal (H$\rightarrow \tau\tau$) and background (Z$\rightarrow \tau\tau$).



```
:
: Evaluation results ranked by best signal efficiency and purity (area)
: ------------------------------------------------------------------------------
: MVA              Signal efficiency at bkg eff.(error):       | Sepa-    Signifi-
: Method:          @B=0.01    @B=0.10    @B=0.30    ROC-integ. | ration:  cance:
: ------------------------------------------------------------------------------
: LD             : 0.090(06)  0.331(11)  0.557(11)    0.677    | 0.110    0.441
: BDT            : 0.062(05)  0.287(10)  0.520(11)    0.659    | 0.097    0.435
: PDERS          : 0.090(06)  0.298(10)  0.519(11)    0.653    | 0.098    0.420
: MLPBNN         : 0.077(06)  0.252(10)  0.490(11)    0.652    | 0.089    0.385
: Likelihood     : 0.066(05)  0.256(10)  0.504(11)    0.642    | 0.043    0.220
: RuleFit        : 0.080(06)  0.281(10)  0.479(11)    0.637    | 0.091    0.383
: SVM            : 0.051(05)  0.192(09)  0.434(11)    0.608    | 0.261    0.211
: KNN            : 0.066(05)  0.201(09)  0.442(11)    0.592    | 0.041    0.237
: FDA_GA         : 0.017(03)  0.149(08)  0.421(11)    0.587    | 0.038    0.207
: Cuts           : 0.085(06)  0.232(10)  0.444(11)    0.560    | --       --
: LikelihoodPCA  : 0.017(03)  0.152(08)  0.388(11)    0.550    | 0.020    0.083
: CutsD          : 0.010(02)  0.100(07)  0.300(10)    0.500    | --       --
: ------------------------------------------------------------------------------
:
```

Figure B.2: ROC curve values for the different MVAs.

Appendix C

BDT implementation in TMVA

Functions used in Adaboost: (boostweight gets multiplied to the event weight)

- Loops over the events, and gets the weight for each (remember getweight gives weight*boostweight)
- Then uses Checkevent to see if the event is misclassified (uses checkevent to get the node type, and then matches that with the node type from the data) **CheckEvent()**:
  - Gets the pointer to the root node
  - If the node is an intermediate root (0), then it checks if the event goes to the right or left, and then gets the pointer to the right/left daughter
  - In the end it gets the node type, converts it to double and returns it
- After looping, calculates err() function
- Calculates the boost weight as Log((1-err)/err)*adaboostbeta, and boostfactor as exp(boostweight)

- If it is a positive weight event, then it sets the boostweight as the current boostweight times boostfactor
- Iterates and calculates the new weights for all the events
- Again iterates, and if it is a signal, then it scales the boostweight by globalnormweight*sigbkgfraction (which is actually 1), for bkg, only the first term
- In the end just returns the boostweight

BDT training:MethodBDT::Train (L1085-L1339) This function does the training for the trees

- Calls InitEventSample()
- Books important training and testing histograms (if enabled boost monitoring)
- Adds a new decision tree recursively using a while loop (stop condition is number of trees and a flag continueBoost), and sets it variables using SetNvars
- Calls BuildTree (returns number of nodes) function to count the number of nodes before and after pruning: (L271-L506:DecisionTree.cxx)
  - Calling buildtree w/o the second argument (node) starts a tree with a root node
  - If the number of events >= 2*min. node size, then only it proceeds with building a tree
  - Uses TrainNodeFast (L909-1335:DecisionTree.cxx) and TrainNodefull (which returns separationgaintotal) to find the optimal cut (i)Separation

Gain is implemented in (L68-104:Separationbase.cxx) (ii)Calculated as parentindex-leftindex-rightindex (each index is got by the using getseparationindex)

- Proceeds to call boost function, if the boost weight is greater than zero, then that gets added to boostweight, else the boosting stops (boostweight == 0), and the continueboost is set to false

Appendix D

Optimization of BDTs

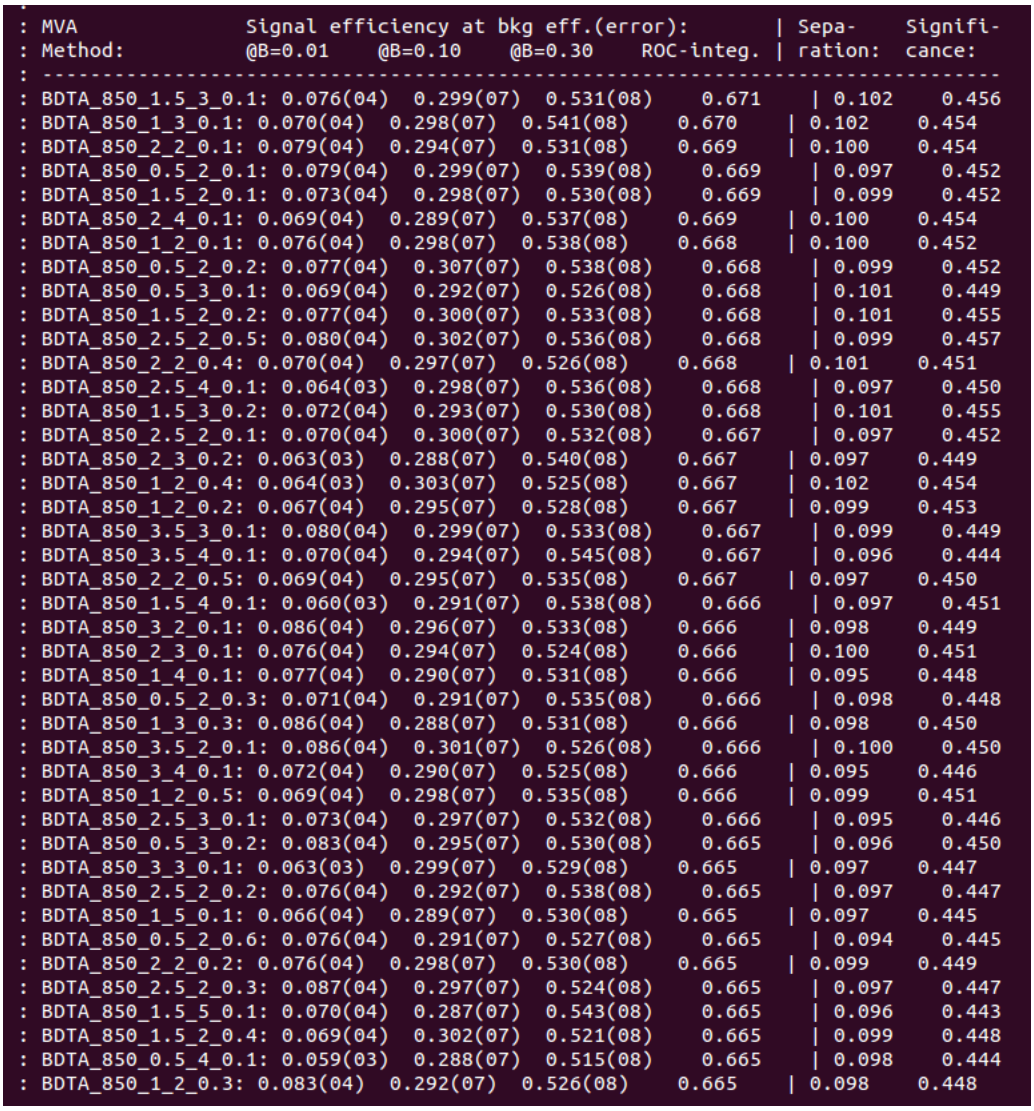
D.1 BDTs for H$\rightarrow \tau\tau$ (signal) and Z$\rightarrow \tau\tau$ (background)

BDT from the TMVA package, has many parameters which can be optimized. The guideline outlined in the TMVA manual [19] include the following points (default values are enclosed in brackets):

- Small tree depths (3)
- Large number of trees (850)
- Slow learning rate (0.5)
- Minimum number of events in the leaf-node (2.5

The number of trees are varied from 800 to 1200 in steps of 50, tree depth is varied from 2 to 5, the minimum percentage of the events in the leaf node is varied from 2% to 5% in steps of 0.5, and the learning rate is varied from 0.1 to 0.6 in steps of 0.1. It is found that no significant improvement is achieved by changing the number of trees, hence the tree number is kept at the default value

of 850. The convention for naming the BDT tree is: MVANameBoosttype-Tree number-% of events in the leafnode-tree depth-learning rate. A screenshot of the ROC curve integral value for the best performing BDTs is shown in fig. D.1



MVA Method:	Signal efficiency at bkg eff.(error): @B=0.01	@B=0.10	@B=0.30	ROC-integ.	Separation:	Significance:
BDTA_850_1.5_3_0.1:	0.076(04)	0.299(07)	0.531(08)	0.671	0.102	0.456
BDTA_850_1_3_0.1:	0.070(04)	0.298(07)	0.541(08)	0.670	0.102	0.454
BDTA_850_2_2_0.1:	0.079(04)	0.294(07)	0.531(08)	0.669	0.100	0.454
BDTA_850_0.5_2_0.1:	0.079(04)	0.299(07)	0.539(08)	0.669	0.097	0.452
BDTA_850_1.5_2_0.1:	0.073(04)	0.298(07)	0.530(08)	0.669	0.099	0.452
BDTA_850_2_4_0.1:	0.069(04)	0.289(07)	0.537(08)	0.669	0.100	0.454
BDTA_850_1_2_0.1:	0.076(04)	0.298(07)	0.538(08)	0.668	0.100	0.452
BDTA_850_0.5_2_0.2:	0.077(04)	0.307(07)	0.538(08)	0.668	0.099	0.452
BDTA_850_0.5_3_0.1:	0.069(04)	0.292(07)	0.526(08)	0.668	0.101	0.449
BDTA_850_1.5_2_0.2:	0.077(04)	0.300(07)	0.533(08)	0.668	0.101	0.455
BDTA_850_2.5_2_0.5:	0.080(04)	0.302(07)	0.536(08)	0.668	0.099	0.457
BDTA_850_2_2_0.4:	0.070(04)	0.297(07)	0.526(08)	0.668	0.101	0.451
BDTA_850_2.5_4_0.1:	0.064(03)	0.298(07)	0.536(08)	0.668	0.097	0.450
BDTA_850_1.5_3_0.2:	0.072(04)	0.293(07)	0.530(08)	0.668	0.101	0.455
BDTA_850_2.5_2_0.1:	0.070(04)	0.300(07)	0.532(08)	0.667	0.097	0.452
BDTA_850_2_3_0.2:	0.063(03)	0.288(07)	0.540(08)	0.667	0.097	0.449
BDTA_850_1_2_0.4:	0.064(03)	0.303(07)	0.525(08)	0.667	0.102	0.454
BDTA_850_1_2_0.2:	0.067(04)	0.295(07)	0.528(08)	0.667	0.099	0.453
BDTA_850_3.5_3_0.1:	0.080(04)	0.299(07)	0.533(08)	0.667	0.099	0.449
BDTA_850_3.5_4_0.1:	0.070(04)	0.294(07)	0.545(08)	0.667	0.096	0.444
BDTA_850_2_2_0.5:	0.069(04)	0.295(07)	0.535(08)	0.667	0.097	0.450
BDTA_850_1.5_4_0.1:	0.060(03)	0.291(07)	0.538(08)	0.666	0.097	0.451
BDTA_850_3_2_0.1:	0.086(04)	0.296(07)	0.533(08)	0.666	0.098	0.449
BDTA_850_2_3_0.1:	0.076(04)	0.294(07)	0.524(08)	0.666	0.100	0.451
BDTA_850_1_4_0.1:	0.077(04)	0.290(07)	0.531(08)	0.666	0.095	0.448
BDTA_850_0.5_2_0.3:	0.071(04)	0.291(07)	0.535(08)	0.666	0.098	0.448
BDTA_850_1_3_0.3:	0.086(04)	0.288(07)	0.531(08)	0.666	0.098	0.450
BDTA_850_3.5_2_0.1:	0.086(04)	0.301(07)	0.526(08)	0.666	0.100	0.450
BDTA_850_3_4_0.1:	0.072(04)	0.290(07)	0.525(08)	0.666	0.095	0.446
BDTA_850_1_2_0.5:	0.069(04)	0.298(07)	0.535(08)	0.666	0.099	0.451
BDTA_850_2.5_3_0.1:	0.073(04)	0.297(07)	0.532(08)	0.666	0.095	0.446
BDTA_850_0.5_3_0.2:	0.083(04)	0.295(07)	0.530(08)	0.665	0.096	0.450
BDTA_850_3_3_0.1:	0.063(03)	0.299(07)	0.529(08)	0.665	0.097	0.447
BDTA_850_2.5_2_0.2:	0.076(04)	0.292(07)	0.538(08)	0.665	0.097	0.447
BDTA_850_1_5_0.1:	0.066(04)	0.289(07)	0.530(08)	0.665	0.097	0.445
BDTA_850_0.5_2_0.6:	0.076(04)	0.291(07)	0.527(08)	0.665	0.094	0.445
BDTA_850_2_2_0.2:	0.076(04)	0.298(07)	0.530(08)	0.665	0.099	0.449
BDTA_850_2.5_2_0.3:	0.087(04)	0.297(07)	0.524(08)	0.665	0.097	0.447
BDTA_850_1.5_5_0.1:	0.070(04)	0.287(07)	0.543(08)	0.665	0.096	0.443
BDTA_850_1.5_2_0.4:	0.069(04)	0.302(07)	0.521(08)	0.665	0.099	0.448
BDTA_850_0.5_4_0.1:	0.059(03)	0.288(07)	0.515(08)	0.665	0.098	0.444
BDTA_850_1_2_0.3:	0.083(04)	0.292(07)	0.526(08)	0.665	0.098	0.448

Figure D.1: ROC curve values for the BDTs with different parameters.

The ROC curves for all the different parameters are shown in fig. D.2

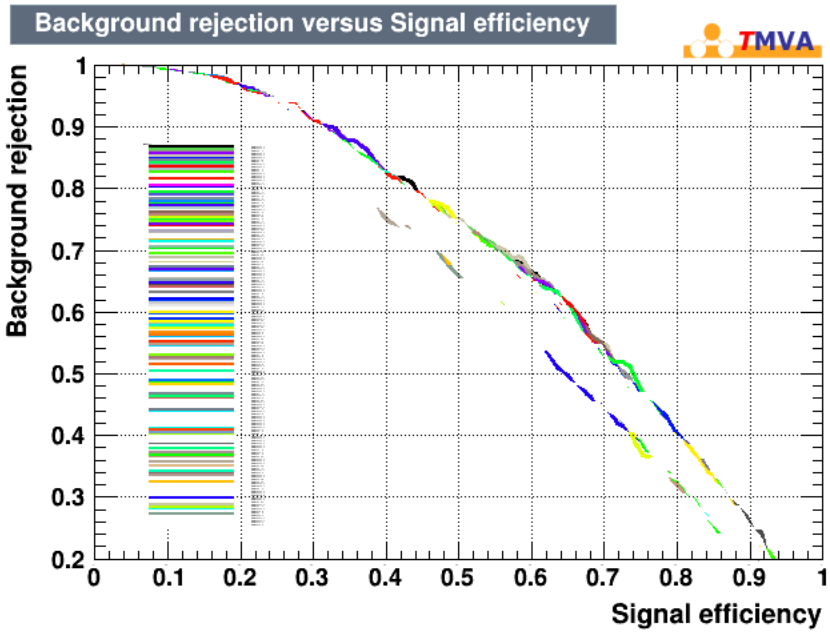


Figure D.2: ROC curve values for BDTs with different parameters.

A few more interesting parameters which can also be optimized are listed below:

- BaggedSampleFraction(0.5)
- nCuts(20)

Preliminary fine-tuning showed that the default values for this variables are the most optimal.

D.2 BDT for Z$\rightarrow \tau\tau \rightarrow \mu\mu$ (signal) and Z$\rightarrow \mu\mu$

The ROC curve value for the BDT is already close to 0.9, no fine-tuning is necessary.

Appendix E

Computing framework

All of the analysis was done in the Open Science Grid (OSG) and LHCC grid Tier-3 cluster at Florida Tech which runs CENTOS 6 on the Compute Element (CE), Storage Element (SE) and the 20 worker nodes. The cluster used HT-Condor as the job manager, and runs jobs from CMS and OSG. 5 nodes are reserved for local users. The batch job submission capabilities offered by the cluster helped immensely in calculating the BDT input variables.

Bibliography

[1] `http://nicadd.niu.edu/~hedin/lq3.html`.

[2] http://inspirehep.net/record/1204788/files/h-prod.png.

[3] https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG.

[4] K. A. Olive et al. Review of Particle Physics. *Chin. Phys.*, C38:090001, 2014.

[5] http://www.atlas.ch/photos/lhc.html.

[6] http://cms.web.cern.ch/news/cms-detector-design.

[7] CMS Collaboration. CMS TECHNICAL DESIGN REPORT FOR THE MUON ENDCAP GEM UPGRADE. Technical Report CERN-LHCC-2015-012. CMS-TDR-013, CERN, Geneva, Jun 2015.

[8] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[9] Vardan Khachatryan et al. Limits on the Higgs boson lifetime and width from its decay to four charged leptons. *Phys. Rev.*, D92(7):072010, 2015.

[10] Agni Bethani. *Neutral Higgs boson searches in the $H\rightarrow \tau\tau \rightarrow \mu\mu$.* Ph.D. thesis, Karlsruhe Institute of Technology, July 2013.

[11] CMS Collaboration. Search for neutral higgs bosons in di-tau final states with the full 2011 and 2012 dataset (Moriond top-up). CMS Physics Analysis Summary CMS AN-2013/011, 2013.

[12] Lyndon Evans and Philip Bryant. LHC machine. *Journal of Instrumentation*, 3(08):S08001, 2008.

[13] The CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004, 2008.

[14] V Karimäki, M Mannelli, P Siegrist, H Breuker, A Caner, R Castaldi, K Freudenreich, G Hall, R Horisberger, M Huhtinen, and A Cattai. *The CMS tracker system project: Technical Design Report.* Technical Design Report CMS. CERN, Geneva, 1997.

[15] A Tapper and Darin Acosta. CMS Technical Design Report for the Level-1 Trigger Upgrade. Technical Report CERN-LHCC-2013-011. CMS-TDR-12, CERN, Geneva, Jun 2013.

[16] CMS GEM collaboration. `https://twiki.cern.ch/twiki/bin/viewauth/MPGD/CmsGEMCollaboration`.

[17] F. Sauli. GEM: A new concept for electron amplification in gas detectors. *Nuclear Instruments and Methods in Physics Research A*, 386:531–534, 1997.

[18] CMS Collaboration. *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems.* Technical Design Report CMS.

[19] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, and Helge Voss. TMVA: Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007.

[20] Rene Brun and Fons Rademakers. Root - an object oriented data analysis framework. In *AIHENP'96 Workshop, Lausane*, volume 389, pages 81–86, 1996.

[21] Tauola and Tauola++ Interfaces to CMSSW. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideTauolaInterface`.

[22] Pythia6 Interface to CMSSW. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuidePythia6Interface`.

[23] SWGuideCMSDriver. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideCmsDriver`.

[24] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.

[25] J. Brooke. Performance of the CMS Level-1 Trigger. *PoS*, ICHEP2012:508, 2013.

[26] Florian Beaudette. The CMS Particle Flow Algorithm. In *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013)*, pages 295–304, 2013.

[27] http://physics.ucsd.edu/students/courses/fall2009/physics214/MET.pdf.

[28] Alexander Belyaev, Renato Guedes, Stefano Moretti, and Rui Santos. Higgs boson phenomenology in $\tau^+\tau^-$ final states at the lhc. *Journal of High Energy Physics*, 2010(7), 2010.

[29] M. Acciarri et al. Measurement of the lifetime of the τ lepton. *Phys. Lett.*, B479:67–78, 2000.