

FLORIDA STATE UNIVERSITY
COLLEGE OF ARTS AND SCIENCES

SEARCH FOR HIGGS BOSON DECAYS TO LONG-LIVED SCALAR PARTICLES WITH
REGIONS OF INTEREST AND MACHINE LEARNING IN CMS

By

SUHO KIM

A Dissertation submitted to the
Department of Physics
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2022

Suho Kim defended this dissertation on November 7, 2022.

The members of the supervisory committee were:

Ted Kolberg
Professor Directing Dissertation

Elizabeth Hammock
University Representative

Todd Adams
Committee Member

Kohsaku Tobioka
Committee Member

Peter Hoeflich
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with university requirements.

To my parents, who always suspected I'd end up here

ACKNOWLEDGMENTS

First, I thank my supervisor Ted Kolberg for instructing me through the CMS experiment. Thanks to him, I have a more thorough understanding of collaboration as a physicist and a member of society. It was an honor to be a member of your group. I am grateful for Dr. Adam's kind welcome to the FSU CMS community as the head of the FSU-CMS group. Before entering FSU-CMS, I was worried and suspicious of my capability to work with such talented physics students in the department. Thanks to his generous welcome, I could immerse myself in FSU-CMS community and develop strong personal relationships with colleagues in FSU-CMS group. Dr. Yohay was quitessential for my better understanding of the reconstruction algorithms of tau leptons, which are infamously difficult and complicated in high energy physics. Dr. Askew's incisive questions helped me not to panic in public speeches. I am grateful for Dr. Prosper's QCD research, diversifying our EXO-dominated FSUCMS community. I still remember my Lepto-quark pre-exercises and feel thankful for it.

I can not ignore our great theoretical high energy physics community as well. Dr. Tobioka's connection to superlative high energy theorists made our fifth floor community so rich. I was lucky to be able to join seminars of world-class theorists thanks to his connection to those people. I want to thank Dr. Reina for giving such great lectures on the most complicated subject of physics study.

I want to express my gratitude to Dr. Hoeflich for his nuclear-Astrophysics class. I sincerely enjoyed the class, which were outside of the high energy physics. I also want to thank Dr. Liz Hammock, for introducing me to her neuroscience project on the early developmental stage of primates. It is great to become an FSU graduate student when you can get introduced to those superb research on a flight back home just because you are a Nole.

I want to express my appreciation to Andrew Evan Hart for implementing the ROI concept in CMSSW codes. We will communicate more in the future. I also need to thank Karla Pena for her continuous communication with the B-parking community, which was critical for my research's development. Pedro Vieira de Castro Ferreira da Silva is undoubtedly the kindest person in CMS collaboration. I was sincerely happy and felt worthy of working for and along with him for the HGCAL simulation work. Too bad I got to know him only in the middle of 2021, but too lucky that I got to know him before my departure.

This journey would have been a lot less memorable if I were not with all of my FSU colleagues. Caleb Benetti, thank you for being my true American friend. I hope we keep that friendship after this journey. Thank you Karem Esther Penalo Castillo. I believe we were quite a pair. I want to congratulate Ryan Kim for his excellent work in FSU-CMS. I also want to express my admiration for all of my FSU-CMS colleagues: Mutia, Nick, Ali, Bobby, Rafeey, Purbita, Pampa, Mac, and Redwan. You people are truly talented and inspirational. I want to thank my 2016 colleagues as well: Bal, Angelica, Yuting, Jing, Lawrence, Jason. No doubt, 2016 was the year. Last, I want to thank Dr. Daniel Diaz, my 2016 friend and FSU-CMS EXO collaborator. It was an honor to be a friend of yours and to have worked with you.

Last but not least, thank you, mom, dad, and my sister.

TABLE OF CONTENTS

List of Tables	viii
List of Figures	ix
List of Symbols	xiii
Abstract	xiv
1 Introduction	1
2 The Standard Model and Its Future	6
2.1 The Standard Model	6
2.1.1 Fermion sector	6
2.1.2 Gauge Sector	8
2.1.3 Scalar Sector	8
2.2 Motivation for Neutral Naturalness	9
3 CMS Detector	12
3.1 The LHC and CMS	12
3.1.1 Calorimetry	14
3.1.2 The superconducting Magnet of CMS	18
3.1.3 The Muon Chamber of CMS	19
3.2 Tracker of CMS	20
3.3 Trigger of CMS	21
4 B Parking Trigger Strategy	23
4.1 Trigger Paths	24
4.2 Integrated Luminosity and pileup weight for the HLT path	27
5 Physics Object Definitions	34
5.1 Muons	34
5.2 Jets	36
5.3 Taus	36
5.4 Region of Interest	36
5.4.1 Tracks	39

5.4.2	Vertex Fitter	40
5.4.3	ROI formation	40
6	Machine Learning	44
6.1	Machine Learning Input Variables	44
6.2	Fine-Tuning of ML environemnt	45
6.3	ML scores of the ROIs	45
7	Event Selection, Signal and Control Regions	54
7.1	Isolation criteria for muons	54
7.2	Delta Phi(lead ROI,sublead ROI)	55
7.3	DeltaR(ROI, jet)	57
7.4	Leading muon's transverse impact parameter significance to PV	58
8	Background Estimation	61
8.1	ABCD method	62
8.2	Validation of ABCD method	67
9	Systematic Uncertainties	72
10	Results	73
11	Conclusions	75
Appendix		
A	Data Samples	76
A.1	BrilCalc command for B-Parking HLT path	76
A.2	Monte Carlo Samples	77
A.2.1	Signal Model and Simulation	77
A.2.2	Background Monte Carlo	78
B	Machine Learning Training Variable	81
References	83	
Biographical Sketch	89	

LIST OF TABLES

4.1	HLT trigger paths used in the analysis	26
4.2	Data and MC Global tags used for 2018 datasets	26
6.1	ROI (trackCluster) variables by category	45
6.2	ROI (Annulus) variables by category	46
6.3	Parameters for ML network	47
7.1	Ptbinned Isocut	55
8.1	Region A’s cutflow histogram by bin entry (weighted)	65
8.2	Region A’s cutflow histogram by bin event number	66
8.3	c τ ,MS vs Signal Contamination in A’ (Bkg :108k)	67
9.1	Systematic Uncertainty Table	72
10.1	Final Data entry in each region	73
10.2	Median value of 95% CL exclusion limit on the branching ratio $\mathbf{B}(H \rightarrow SS \rightarrow \tau^+\tau^-\tau^+\tau^-)$	74
A.1	Datasets used in the analysis	76
A.2	Signal Monte Carlo samples CAMPAIGN:RunIIAutumn18MiniAOD-rp_102X_upgrade2018_realistic_v15-v1	78
A.3	QCD MuEnriched Pt5 background Monte Carlo samples	79
A.4	W,Z,H boson background Monte Carlo samples	79
A.5	Top background Monte Carlo samples	79
A.6	Monte Carlo sample summary	80
B.1	Output scores for different Epoch training	81
B.2	Output scores for different Signal and Background MC combination	82

LIST OF FIGURES

1.1	Feynman diagram of ZH signal process studied in the most recent CMS analysis paper (2021) [45]	3
1.2	The analysis' signal process feynman diagram. The Higgs is created in gg production mode. The LLP scalar decays into τ lepton pair.	4
1.3	τ lepton's decay mode pie chart	4
1.4	τ lepton's detection mode pie chart. Prong means a charged track reconstructed in the tracker volume of CMS. It corresponds to π^\pm	5
2.1	A cartoon display of Higgs Portal process	10
2.2	A diagram display of Higgs Portal process	11
3.1	Picture of the CERN complex [54]	13
3.2	Picture of CMS viewed from the beam direction [55]	15
3.3	Cartoon of CMS with its subpart annotated [56]	16
3.4	Cross-section of CMS detector as a particle traverses through the apparatus [56] . . .	16
3.5	Schematic view of the ECAL. [57]	17
3.6	Cross-section of the HCAL in CMS detector. It shows the Barrel, endcap, front, and outside portions. [58]	17
3.7	Cross-section of the Muon Chamber in CMS detector. It shows the 4 layers of the drift tube (DT) cross-section viewed from the z-axis. [55]	19
3.8	Cross-section of the trackers in CMS detector. It shows the pixels in the inner tracker for more precise vertexing and the silicon strips on the outer trackers. Silicon strips are tilted with respect to previous layers of strips. [59]	20
3.9	Cartoon diagram displaying the Trigger and Data Acquisition System process, along with parking of datasets/	21
4.1	The figures show two different loop diagrams for $B \rightarrow K^* \ell^+ \ell^-$ processes. The FCNC is forbidden in the SM and there is no tree level process for $B \rightarrow K^* \ell^+ \ell^-$. Thus, these two processes are the leading contributors for $B \rightarrow K^* \ell^+ \ell^-$, which are highly suppressed.[60]	24
4.2	If the Lepton universality is not satisfied ($R(K^*, D^*) \neq 1$), it implies there is new physics hidden in the diagram. It could be interpreted in terms of Lepto-quark or Electroweak Z boson which couples to right-hand chiral leptons.[60]	25

4.3	pt of the scalar products	26
4.4	DeltaR of the scalar products	27
4.5	lifetime of the scalar products in the lab frame	28
4.6	The plots show the trigger efficiency for each HLT path with respect to LLP's lifetime. Each line denotes mass scale of each LLP. Please note the efficiency is set to 0 for MS=7 GeV $c\tau = 1\text{mm}$ due to absence of Monte Carlo (MC).	29
4.7	The plots show the trigger efficiency for each HLT path with respect to LLP's lifetime. Each line denotes mass scale of each LLP. The analysis uses Mu9_IP6 for Era A,B of data and Mu12_IP6 for Era C,D of data. Please note the efficiency is set to 0 for MS=7 GeV $c\tau = 1\text{mm}$ due to absence of Monte Carlo (MC).	30
4.8	Bunch crossing of dataset /ParkingBPH1/Run2018A-UL2018_MiniAODv2-v1/MINIAOD	31
4.9	Bunch crossing of dataset /ParkingBPH1/Run2018B-UL2018_MiniAODv2-v1/MINIAOD	31
4.10	Bunch crossing of dataset /ParkingBPH1/Run2018C-UL2018_MiniAODv2-v1/MINIAOD. Please note that the HLT path for EraC has higher muon object's p_T threshold with 12 GeV (compared to 9 GeV in EraA).	32
4.11	Bunch crossing of dataset /ParkingBPH1/Run2018D-UL2018_MiniAODv2-v1/MINIAOD. Please note that the HLT path for EraD has higher muon object's p_T threshold with 12 GeV (compared to 9 GeV in EraA).	32
4.12	Bunch crossing of Monte Carlo Simulation for physics process of Drell-Yan. The dataset is /DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 /RunIIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v2/ MINIAODSIM	33
4.13	PUweight calculated for Era A of B-parking dataset	33
5.1	Data/MC of muon objects p_T , η , energy, nMu	35
5.2	Data/MC of jet objects for p_T , η , energy, nJet	37
5.3	Data/MC of τ objects for p_T , η , energy, nJet	38
5.4	Data/MC of MET objects for p_T , energy	39
5.5	The figures display cartoons of the step-by-step formation of the Region of Interest (ROI). After LLP travels from the Primary Vertex (PV) to Secondary Vertex (SV), it leaves no track, and decays into a tau lepton pair. Tau lepton pair's tracks are recorded in the tracker, as denoted in green color lines. The first figure shows the vertex (red star) fitted by KalmanVertexFitter in CMSSW from two charged tracks that decayed from a neutral LLP. The second figure shows fitting procedure of another vertex. With one of tau leptons decaying into a muon, the muon track crosses the partner tau lepton track and forms a new vertex near the original vertex. The third figure shows the	

clustering of the vertex, with the second vertex formed by the extended tracks being clustered into the original ROI if the vertices are within a 1cm limit.	41
5.6 The figures display cartoons of the step-by-step formation of the shell structure. The first figure shows the axis of the cone when the axis goes through the center of ROI. The second figure shows the construction of the isolation plane, with $\Delta = 0.3$ parameter. The third figure shows the construction of the isolation shell, with the isolation plane becoming a sphere.	42
5.7 The figures display cartoons of the step-by-step inclusion of isolation track information. The first figure shows an electron or charged hadron track (orange color) originating from a Pile-up interaction. One can see the pileup track goes through the isolation shell in the cartoon. The second figure shows that the track is outside of the ROI and inside of the isolation shell. Thus, the track becomes an isolation track.	43
6.1 Simple cartoon showing ML network for vertex and annulus shell.	46
6.2 Signal versus Background for loglead, where the ROI score is the highest ROI of the event. Left plot is for MS-15_ctauS-10mm point, whereas the right plot is for MS-15_cauS-100mm point	48
6.3 Signal versus Background for logexclead (sublead), where the ROI score is the second highest ROI (outside of $d\Phi=0.4$ from leading ROI) of the event. Left plot is for MS-15_ctauS-10mm point, whereas the right plot is for MS-15_cauS-100mm point	49
6.4 Data/MC agreement for loglead and logexclead scores	50
6.5 Data/MC agreement for leading and subleading ROI distributions	51
6.6 Data/MC agreement for leading and subleading ROI distributions	52
6.7 Data/MC agreement for leading and subleading ROI distributions	53
7.1 ΔR value of tau lepton pair from LLP decay is plotted. $\tau s'$ mass are low enough to be resolved when decayed from heavier LLP.	56
7.2 Signal versus Background for Delta Phi(leadROI, subleadROI). The plot is for MS-15_ctauS-100mm point	56
7.3 Delta R(Jet, leadingROI). Left plot is for MS-15_ctauS-10mm point, whereas the right plot is for MS-15_cauS-100mm point	57
7.4 leading muon's transverse impact parameter value to the primary vertex. Left plot is for MS-15_ctauS-10mm point, whereas the right plot is for MS-15_cauS-100mm point	58
7.5 Amplified histograms of Log10muIPSig	59
7.6 Data/MC agreement for distributions	60
8.1 Data/MC agreement for distributions. High ROI scores and high Log10(muIPSig) values lack MC statistics	61

8.2	Cutflow histogram of Signal versus background. The signal is MS15 GeV-ct10mm point. Last bin of the cutflow histogram is ABCD region selection. Top left plot is for region A, whereas the bottom right plot is for region D	63
8.3	Final bin of cutflow	64
8.4	Cartoon description of ABCD method validation. The first figure shows a simpler definition of validation region. The second figure shows a more sophisticated definition of validation regions.	68
8.5	Profile plot of VR. Correlation factor is 0.105. ProfileX is rather flat. X-axis is log10(1-leading ROI score). Y-axis is Log10(muIPSig).	69
8.6	Profile plot of VR1. Correlation factor is 0.124. ProfileX is rather flat. X-axis is log10(1-leading ROI score). Y-axis is Log10(muIPSig).	69
8.7	Profile plot of VR2. Correlation factor is 0.1229. ProfileX is rather flat. X-axis is log10(1-leading ROI score). Y-axis is Log10(muIPSig).	70
8.8	Profile plot of VR3. Correlation factor is 0.1203. ProfileX is rather flat. X-axis is log10(1-leading ROI score). Y-axis is Log10(muIPSig).	70
10.1	Current Preliminary limit plots for lower mass LLP	74
10.2	Current Preliminary limit plots for higher mass LLP	74
A.1	Leading Feynman diagrams for ggH production mode	78

LIST OF SYMBOLS

The following short list of symbols are used throughout the document. The symbols represent quantities that I tried to use consistently. The symbols follow CMS Technical Design Report (TDR) style.

D0	D0 experiment
GeV	1 Giga electron-Volt
TeV	1 Tera electron-Volt
\mathcal{L}_{int}	Integrated luminosity
p_T	Transverse momentum
η	Pseudorapidity
E_T	Transverse energy
H_T	Scalar sum of hadronic jet's transverse energy
E_T^{miss}	Missing transverse energy
Pp	Proton-to-Proton collision
H	Higgs
π^0	Pion (pi-zero)
K_S^0	K short
J/ψ	J-Psi meson
B_s^0	strange B meson
Υ	Upsilon meson
g	Gluon
b	b quark
t	t quark
e	Electron lepton
μ	Muon lepton
τ	Tau lepton
j	Jet
$t\bar{t}$	t-tbar
M_{Pl}	Planck mass
MS	Mass scale of particle
$c\tau$	Lifetime of particle in its rest frame
ℓ	Electrically charged leptons
ν_l	Generic neutrino

ABSTRACT

We present a search for long-lived particles (LLPs) produced in gluon fusion Higgs production mode (ggH), using a novel strategy of Regions of Interest (ROIs). Regions of Interest are collections of pair-wise track vertices fitted by the vertex fitter in CMSSW. The analysis focuses on LLPs with lifetimes of mm-m ranges, and result in decay of collections of pair-wise track vertices. Variables of the constructed ROIs become inputs for our Deep Neural Network (DNN) Machine Learning (ML) algorithms, as the main discriminator between the signal and the background. We look for LLPs that decay into two tau leptons. This final state is particularly interesting, given tau lepton final state exclusion limits are mostly omitted in preceding analyses, due to the difficulty of reconstructing tau leptons. To trigger on this final state, we exploit the B-parking trigger. The B-parking trigger is a novel High-Level Trigger path in CMS detector, targetting soft displaced muons. The trigger was installed in 2018, totaling an integrated Luminosity value of $\mathcal{L}_{int} = 44\text{fb}^{-1}$. No excess of events over the standard model expectation is observed. The results are interpreted in the context of exotic Higgs decays to a pair of long-lived scalars (S). We set limits on the branching ratio of the Higgs to LLPs, $\mathbf{B}(H \rightarrow SS \rightarrow \tau^+\tau^-\tau^+\tau^-)$, as a function of the proper lifetime. The analysis has the strongest discriminant power for the LLP's $c\tau$ 10 mm and 15 GeV mass. The analysis' exclusion limit on the branching ratio for τ lepton final state ($\mathbf{B}(H \rightarrow SS \rightarrow \tau^+\tau^-\tau^+\tau^-)$) is one of the most stringent results from the LHC detectors. We place limits on the Higgs branching ratio to this final state down to 9% at 10 mm for 7 GeV, 5% at 10 mm for 15 GeV, 5% at 100 mm for 40 GeV, and 18% at 100 mm for 55 GeV.

CHAPTER 1

INTRODUCTION

The discovery of particles at the electroweak scale, such as the top quark at Fermilab’s CDF and D0 experiment [1, 2] and the Higgs boson at the Large Hadron Collider (LHC) at CERN [3, 4], led to the discovery of all constituents in the Standard Model (SM). The SM describes the nature of fundamental particles and their interactions with precision. Despite its success, the SM suffers from a few obstacles: the evidence of neutrino masses and mixing [5], the observations of bullet clusters confirming the presence of dark matter (DM) [6, 7, 8, 9, 10], and baryon-antibaryon asymmetry [11] all remain unexplained in the framework of SM. In addition, the SM suffers from the naturalness problem. One needs to look for physics Beyond the Standard Model (BSM) to solve such issues.

Problems with the Standard Model have led to both theoretical and experimental attempts to explore BSM physics. To search for the BSM, the high energy physics (HEP) community has completed many types of research, both on the theoretical and experimental sides. The theoretical high energy physics community approached the issue with two main approaches. The first approach tackles the issue of precision. Although the SM is a well-constrained model, further precision of particles in the SM, especially those in the electroweak scale, gives new insight into BSM physics. For instance, the CDF collaboration recently discovered a 7σ deviation of the W boson mass from the SM prediction [12]. The W boson mass deviation has been interpreted for new physics using the framework of the Standard Model Effective Field Theory (SMEFT) [13]. In SMEFT, the SM operators from the SM Lagrangian are used to build 5,6,8-dimensional new terms for the Lagrangian. The SMEFT scalar, fermion, or vector extension hints at new insight into the BSM and its phenomenology [13]. The other approach aims to build an entirely new Lagrangian term, which can be appended to the current SM Lagrangian term. The BSM Lagrangian introduces new particle fields and should address unsolved issues in the current SM framework. The new particles’ mass scale can range from the collider level to the astrophysical level.

To find the new BSM particles, the experimental high energy physics community has conducted many different kinds of research based on collider and astrophysical data. Experimental physicists invested strenuous effort in data from collider physics since the lightest particles from popular theory, such as supersymmetry (SUSY), were within reach of high energy colliders. Experimentalists

searching for the BSM particles can also divide their main approaches into two, the first with prompt decay of the BSM particle and the second with a long-lived lifetime signature. CMS analysis targeting the BSM particles with prompt lifetime has been studied and resulted in no deviation from the SM prediction [14]. However, the second approach, which has not been fully investigated, is when particles decay with a Long-Lived signature, in other words, Long-Lived particles (LLP). This signature is uniquely exciting and challenging. It requires a different analysis strategy depending on the BSM particle's mass scale (MS) and lifetime ($c\tau$). Thus, HEP experimental scientists perceive this frontier as the blue ocean for HEP experimentalists, to the extent that they are planning a new detector to target LLPs solely [15].

In this dissertation, we focus on the LLPs originating from the LHC collisions, review preceding analyses, and propose a novel strategy. Searches for LLPs decaying into final states containing jets were investigated at the Tevatron ($\sqrt{s} = 1.96$ TeV) by both CDF [16] and D0 [17] Collaborations, at the LHC by "A Toroidal LHC Apparatus" (ATLAS) and LHCb Collaborations at $\sqrt{s} = 7$ TeV [18, 19], by CMS, ATLAS and LHCb Collaborations at $\sqrt{s} = 8$ TeV [20, 21, 22, 23, 24, 25, 26]. More recently, CMS [27, 28, 29, 30, 31, 32] and ATLAS Collaborations [33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44] released publications with data obtained at $\sqrt{s} = 13$ TeV.

CMS Collaboration released a new result in 2021. In the new paper, the Higgs decaying into LLPs is created in association with a Z vector boson [45]. This analysis' Feynman diagram is displayed in figure ???. The new analysis sheds light on LLPs with lighter masses thanks to the clean dilepton trigger from the Z vector boson. The preceding analyses successfully put the exclusion limit on the branching ratio, the Higgs to the LLPs to b and d-quark, below unitarity. However, the exclusion limit for τ final state has been omitted or presented with values above 1.

However, the Leptophilic model for Twin Higgs and other Higgs models are also highly motivated [46]. Continuous neglect of τ final state limit is not only poor practice but overlooks an important unexplored phase space. This analysis searches for LLPs originating from the Higgs Portal model with the Higgs' Leptophilic nature. Since the coupling strength of the Higgs' to SM fermions are quantified by the Yukawa couplings, focus on Leptophilic Higgs translates into a focus on τ final state decaying back from the LLPs via the Higgs Portal. The mass limit on the scalar mass is 55GeV; we investigate only on-shell neutral scalar particles from the Higgs. A minimum 7 GeV mass for scalar particles is required to create on-shell τ -lepton pairs. The Feynman diagram for scalar particle production mechanism is depicted in Figure ??.

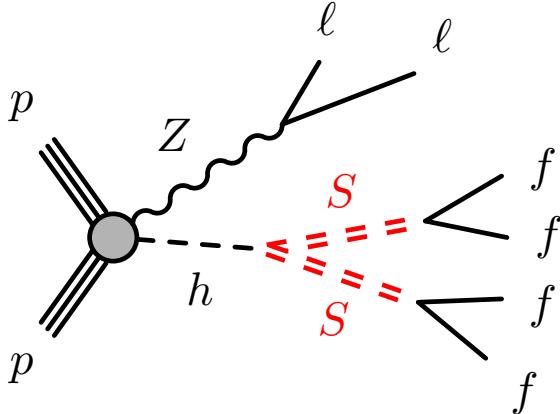


Figure 1.1: Feynman diagram of ZH signal process studied in the most recent CMS analysis paper (2021) [45]

The main challenge and reason for omitting a τ lepton analysis are the different decay modes of τ leptons. τ leptons decay hadronically and leptonically, with several different sub-decay modes. Its decay mode pie chart and CMS detection category pie chart are displayed in figures ?? and ?? respectively. Developing analysis strategies to optimize the search for each sub-decay mode is exceptionally complicated—the main reason for the omission or no reasonable exclusion limit in preceding LHC results. A displaced vertex search can be more efficient than displaced objects (jet, muon, electrons) search to be inclusive of all τ leptons’ decay modes. We exploit the newly developed Regions of Interest mechanism in the tracker volume. The mechanism fits pair-wise tracks of Lost-tracks and PackedPFCandidates classes in CMS data into a vertex, and cluster the vertices to form Regions of Interest. ROIs save all relevant track and fitted vertex qualities along with isolation information. These variables are input for Machine Learning (ML) algorithms for a data-scientific search method.

Another challenge is that CMS searches are not optimal for detecting Higgs boson decays due to its decay products’ relatively low momenta. Higgs produced in association with Z vector boson analysis [45] overcame this barrier with the help of dilepton trigger. Although the ggH production mode gives the largest Higgs cross-section, it further complicates the trigger strategy. This analysis exploits the τ lepton’s leptonic decay, where the τ lepton decays into a soft muon. We use a trigger of the B Parking High-Level Trigger (HLT) Path implemented in CMS for the 2018 portion of Run 2 to trigger on the soft muon.

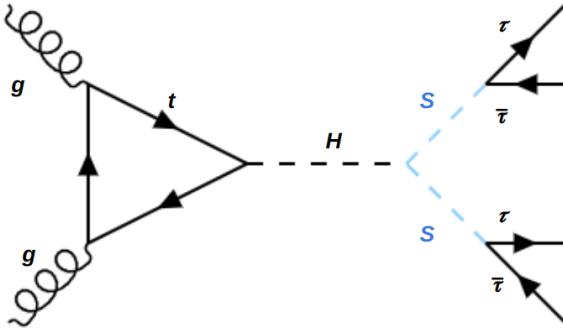


Figure 1.2: The analysis' signal process feynman diagram. The Higgs is created in gg production mode. The LLP scalar decays into τ lepton pair.

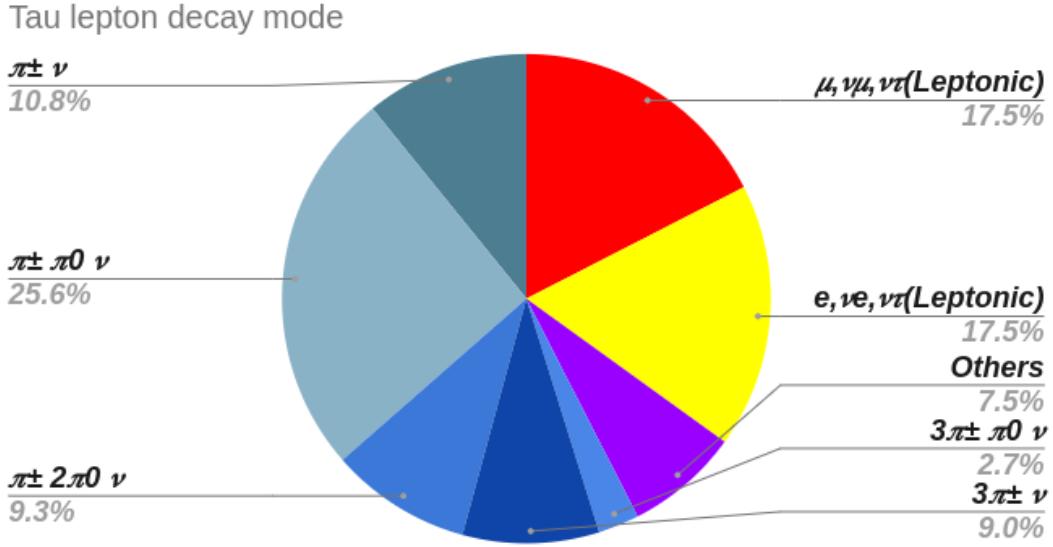


Figure 1.3: τ lepton's decay mode pie chart

The rest of the dissertation is organized as follows. In Section 2, we discuss the theoretical background of the BSM and LLPs in more detail. In Section 3, CMS detector is thoroughly discussed with emphasis on the tracker and the trigger system, which are relevant detector parts for the analysis. We discuss how the analysis exploited the b-parking trigger in Section 4, with a description of its original motive for the trigger's installation. The physics objects and formation of Regions of Interest are clarified in Section 5. The machine learning algorithms are further explained in Section 6. The event selections are presented in Section 7. Section 8 describes the data driven background estimation method and the method's validation process. Section 9 describes the systematic uncertainties. Finally, Section 10 presents the search results. We conclude with

Tau lepton detection mode

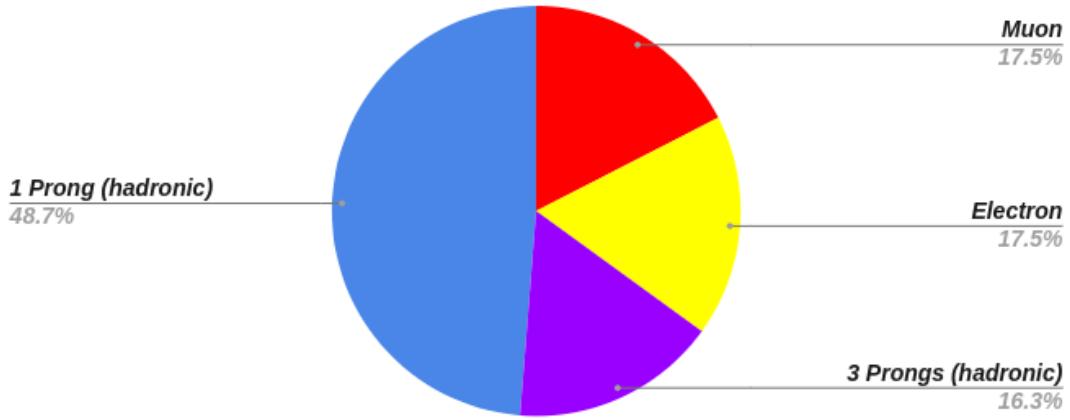


Figure 1.4: τ lepton's detection mode pie chart. Prong means a charged track reconstructed in the tracker volume of CMS. It corresponds to π^\pm

Section 11.

CHAPTER 2

THE STANDARD MODEL AND ITS FUTURE

In Section 1, we introduced the few obstacles facing the SM: The existence of dark matter, baryon-antibaryon asymmetry, and the evidence of neutrino masses and mixing. The SM Lagrangian, written as in formula 2.1, does not have particle fields that can explain those phenomena.

$$\mathcal{L}_{SM} = \mathcal{L}_{Higgs} + \mathcal{L}_{Gauge} + \mathcal{L}_{Kinetic} + \mathcal{L}_{Yukawa} \quad (2.1)$$

Since it can not be explained by any of the particles' fields in the SM, it requires the addition of new particles' fields or new terms in the current SM Lagrangian expression. New terms in the SM Lagrangian entail new vertices in the Feynman diagram, which open the door for a new understanding of high energy physics. However, if those observations did not exist, the SM is complete within its framework except for the naturalness problem. The naturalness problem originates from the SM Higgs being a scalar particle. The formula 2.1 has three fields: boson, fermion, and scalar.

2.1 The Standard Model

2.1.1 Fermion sector

Quantum Field Theory (QFT), humanity's mathematical framework used for the SM, explains matter as an excited state of fermion fields derived from the canonical quantization of the SM Lagrangian's fermion fields. The fermions' fields have baryon symmetry, and slightly broken isospin, and chiral symmetries in the SM Lagrangian. Baryon symmetry is evident in $\mathcal{L}_{Kinetic}$ and \mathcal{L}_{Yukawa} of formula 2.1. Formula 2.2 and 2.3 shows $\mathcal{L}_{Kinetic}$ and \mathcal{L}_{Yukawa} Lagrangian terms' expanded forms for different chirality and up-type and down-type fermions.

$$\mathcal{L}_{Kinetic} = \bar{Q}_L \cdot i\gamma^\mu D_\mu Q_L + \bar{d}_R \cdot i\gamma^\mu D_\mu d_R + \bar{u}_R \cdot i\gamma^\mu D_\mu u_R + h.c \text{ (Hermitian Conjugate).} \quad (2.2)$$

$$\mathcal{L}_{Yukawa} = \bar{Q}_L \cdot Y_{ij}^d H \cdot d_R + \bar{Q}_L^* \cdot Y_{ij}^u H^* \cdot u_R + h.c \text{ (Hermitian Conjugate).} \quad (2.3)$$

In formular 2.2, Dirac matrices are summed up over Lorentz indices for Lorentz invariance. In formula 2.2 and 2.3, d_R , u_R are right-hand fermions for down-type (ℓ for leptons) and up-type (ν_l

for leptons) fermions respectively. Q_L is a left-hand fermion doublet as expressed in formula 2.4.

$$Q_L = \begin{bmatrix} u_L \\ d_L \end{bmatrix} \quad (2.4)$$

Since fermions with left-hand chirality behave differently from fermions with right-hand chirality, the covariant derivatives are also different for left-hand kinetic term and right-hand kinetic terms, being defined in the formula 2.5 and in 2.6.

$$D_\mu = i\partial_\mu - \frac{1}{2}g\mathcal{T} \cdot W_\mu - \frac{1}{2}g'YB_\mu \quad (2.5)$$

$$D_\mu = i\partial_\mu - \frac{1}{2}g'YB_\mu \quad (2.6)$$

\mathcal{T} is isospin for SU(2) weak force, and g , W are couplings and fields for the weak theory. g' and B are respective variables for quantum electrodynamics (QED). As a fermion transforms as in formula 2.7, we can see all the terms in the formula 2.2 and 2.3 are invariant.

$$q \rightarrow q \cdot e^{i\alpha_B}. \quad (2.7)$$

Thus, fermion sector of the SM Lagrangian satisfies the baryon symmetry.

The SM's weak force satisfies the Special unitary group, SU(2) symmetry. Isospin is the quantum number for the SU(2) symmetry of the SM. Up-type fermions and down-type fermions have $(\frac{1}{2}, \frac{-1}{2})$ isospin values in the SM. If the isospin symmetry were perfect in the SM, it could have many implications for baryons and others in the universe. Baryons of protons and neutrons consist of 3 quark fermions: (u,u,d) and (u,d,d) respectively. As SU(2) group generator transits the isospin quantum number from up-type to down-type, proton and neutron would have the same mass if up quark and down quark had same masses. However, up quark and down quark have different masses. Thus, the isospin symmetry is slightly broken in the SM and result in slight difference in mass of protons and neutrons.

The SM fermions also have chiral symmetry breaking. Fermions can transform under chiral rotation as given in formular 2.8.

$$\begin{bmatrix} u'_L \\ d'_L \end{bmatrix} = e^{i\alpha_5} \cdot \begin{bmatrix} u'_L \\ d'_L \end{bmatrix}, \quad u'_R = e^{-i\alpha_5} \cdot u_R, \quad d'_R = e^{-i\alpha_5} \cdot d_R, \quad (2.8)$$

Transformation phase is in opposite direction for left-hand and right-hand fermions. Under the transformation formular of 2.8, the mass of fermions are not always invariant under chiral rotation. Formular 2.9 shows the change of fermion's mass under chiral rotation.

$$m_q \cdot (\bar{u}'_R u'_L + \bar{d}'_R d'_L) = m_q \cdot e^{2i\alpha} \cdot (\bar{u}_R u_L + \bar{d}_R d_L) \quad (2.9)$$

The mass is only invariant when the $m_q = 0$. However, fermions' masses are not zero and the chiral symmetry breaking is proportional to the masses of the fermions. It results in “Technical naturalness” of the SM fermions’, which makes quantum correction to be also proportional to the same chiral symmetry breaking. It provides a significant advantage in QFT’s renormalization. QFT’s renormalization understands the observed mass of particles in terms of its bare mass and correction as in 2.10.

$$m^2 = m_0^2 + \delta m^2. \quad (2.10)$$

For fermions, the mass correction term (δm_f^2) is multiplicative to the chiral symmetry breaking as in formula 2.11.

$$\delta m_f^2 = m_f \cdot \frac{g^2}{16\pi^2} \cdot \ln \frac{\Lambda}{m_f}, \quad \Lambda = \text{New Physics Mass Scale} \quad (2.11)$$

The mass correction term at a new BSM physics scale is multiplied by the chiral symmetry breaking. Its multiplicative nature protects fermion fields’ from extreme radiative correction at BSM scale in the renormalization process.

2.1.2 Gauge Sector

Likewise, the bosons’ fields satisfy special QFT framework symmetries. The boson satisfies the U(1), SU(2), or SU(3) gauge symmetry and is expressed in gauge terms as in formula 2.12.

$$\mathcal{L}_{Gauge} = -\frac{1}{4}B_{\mu\nu} \cdot B^{\mu\nu} - \frac{1}{4}W_{\mu\nu}^T \cdot W_T^{\mu\nu} - \frac{1}{4}G_{\mu\nu}^\alpha \cdot G_\alpha^{\mu\nu}. \quad (2.12)$$

, where B, W, and G are QED, weak, quantum chromodynamics (QCD) fields, T, α are SU(2), SU(3) group indices, respectively.

Gauge symmetry, like technical naturalness for fermions, also provides one significant advantage for renormalization. In case of gauge symmetry, the correction term (δm_f^2) of formula 2.10 becomes zero and protects boson fields from the divergence of radiative correction in the renormalization process.

2.1.3 Scalar Sector

Unlike fermions or gauge bosons, the scalar field of the Higgs boson is not protected by any symmetry. It is subject to large radiative corrections, especially from the top quark loop. Thus, the necessary radiative corrections are enormous for the SM to be valid up to the Planck or Grand Unification Theory (GUT) scales of 10^{16} GeV. One needs exorbitant fine-tuning to fit the Higgs

mass at the observed value of 125 GeV. A simple mathematical expression is shown in formula 2.13

$$125 \text{ GeV} = m_0^2 - \frac{g^2}{16\pi^2} \cdot \Lambda^2, \quad \Lambda = O(10^{16}) \quad (2.13)$$

One of the most popular solutions to this problem is Supersymmetry (SUSY), which assigns chirality to the Higgs particle. SUSY solves the fine-tuning problem, neutrino masses, and provides a candidate for DM.

Unfortunately, the LHC has found no significant excess over the SM background in their search for SUSY[14]. Although the non-observation of supersymmetric partner particles does not invalidate SUSY, it makes it less attractive among the particle physics community. Non-observation of superpartners, particularly the stop (scalar partner of the top quark), has pushed its mass beyond 1 TeV. It generates a “little hierarchy” problem, but an alternative “neutral naturalness” solution remains.

2.2 Motivation for Neutral Naturalness

The top partners are not charged under the SM color group in the framework of neutral naturalness. Because of being colorless, their production cross section is much smaller, and the present limits on the top SUSY partner particles are well below 1 TeV. Examples of neutral naturalness models are the Twin Higgs [47], Folded SUSY [48], and the Quirky Little Higgs [49] models. Twin Higgs realizes the Higgs as a pseudo-Goldstone boson that protect the weak scale from radiative corrections up to scales of order 5 - 10 TeV [47]. Naturalness constrains the masses of most of the twin/mirror partners not to exceed a few hundred GeV [47]. Twin baryons can constitute some or all of the dark matter in the universe, depending on the baryon asymmetry in the mirror sector. [47]. Thus, Twin Higgs model can solve both the naturalness problem and dark matter issue of the SM. Folded SUSY postulates that one loop quadratic divergences of the Standard Model Higgs field are cancelled by “folded partners”, and the gauge quantum numbers for the folded partners are a bit different from conventional SUSY [48]. The familiar squarks and gauginos need not be present [48]. The only folded partners with masses less than 50 GeV are the gluons of mirror color [48]. These folded gluons will confine into folded glueballs at a scale of order a few GeV [48].

In the Quirky Little Higgs scenario, there are some new fermions that couple to a new non-Abelian gauge group referred to as “infracolor” [49]. The fermions charged under this new gauge group are called quirks in analogy to the traditional quarks [49]. The dark top quirk and top quark

are related by an $SU(6)$ bulk gauge symmetry in which their respective confining gauge groups are embedded [49]. Quirky Little Higgs theorizes that top loop contribution to the Higgs mass is cancelled by an uncolored heavy “top quirk” charged under a different $SU(3)$ gauge group [49]. The Higgs in this model is a pseudo-Nambu-Goldstone boson and its mass parameter is protected by an $SU(3)_W$ symmetry [49]. Depending on the “quirk” color scale, glueball phenomenology can arise or not [49]. If the infrared scale is larger than MeV scale, the quirk-anti-quirk pair annihilates mainly into hidden sector glueballs, but also into two photons or two leptons [49]. The displaced leptons may provide the easiest search strategy [49].

Theoretical models provide the possibility of neutral Long-Lived Particles, which may be produced in the proton-proton collisions of the LHC and decay back to SM particles far from the interaction point (IP).[50] If the mirror QCD gluons form scalar glueballs, the SM Higgs boson can become a “Higgs Portal” between the SM and BSM mirror QCD scalar glueballs. In the Mirror SM and Twin SM models, only the SM Higgs boson can interact with both SM QCD and mirror QCD particles as in figure ???. BSM mirror QCD scalar glueballs can only decay back to SM particles

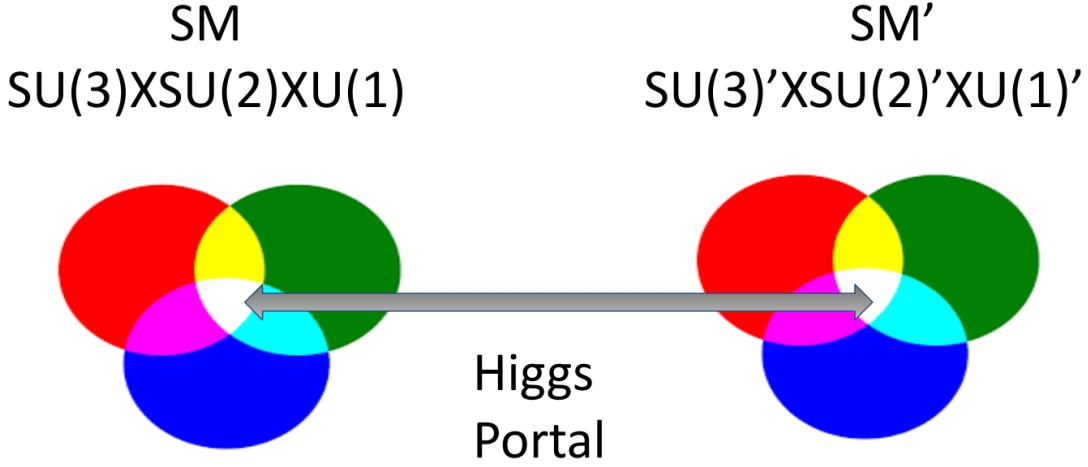


Figure 2.1: A cartoon display of Higgs Portal process

via Higgs boson decay.

Because of its decay as an off-shell Higgs boson, its cross section is highly suppressed. The decay branching ratio to the highest mass fermions will be highest following the Yukawa couplings. The decay ratio into b quarks or τ leptons is highest depending on the mirror scalar’s mass. However, if the Higgs has leptophilic behavior, the decay ratio into τ leptons is always the highest. The displaced decays of the scalars would lead to exotic signatures in CMS, such as distant innermost

tracker hits, displaced vertices, and displaced jets. The phenomenology of long-lived decay increased interest in a neutral naturalness among the particle physics community [51, 52]. The long-lived scalar model process is shown in Figure ??.

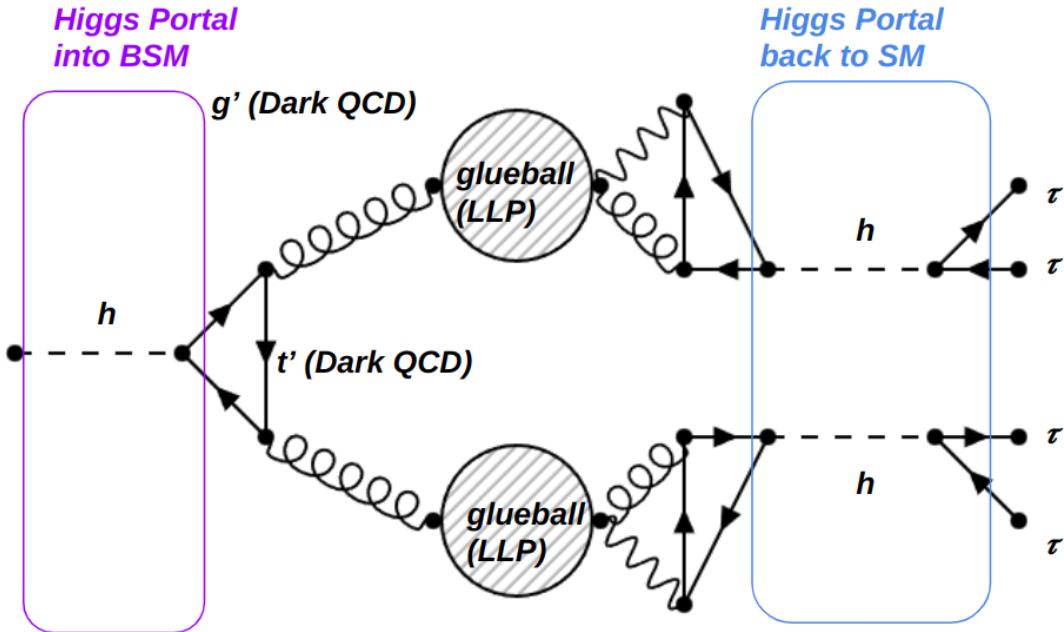


Figure 2.2: A diagram display of Higgs Portal process

CHAPTER 3

CMS DETECTOR

Since the discovery of the neutron in the 30s, humanity’s aspiration to understand the most fundamental constituents and laws of physics has only accelerated. The Discovery of pions, Kaons, and other hadrons from the cosmic rays in the 40s and 50s led to an understanding of the substructure of particles, which culminated in Gell-Mann’s Eightfold way. Deep inelastic scattering performed in the 60s by the Stanford Linear Accelerator Center (SLAC) confirmed the existence of “partons” or “quarks” inside those particles. Observance of J/ψ and Υ mesons in SLAC and MIT put the third-generation fermions in the collection of particles. Gargamelle bubble chamber in CERN succeeded in detecting muon neutrinos, postulated by Pauli in the 30s from the beta decay experiment. CERN’s UA1 and UA2 in the 80s confirmed the existence of the first particles in the electroweak scale, namely the W and Z bosonsFermilab’s CDF and D0 in the 90s confirmed the existence of the top quark, the heaviest particle in the SM. Fermilab’s CDF and D0 in the 90s confirmed the existence of t , the heaviest particle in the SM. As much as we appreciate the predecessor physicists for their phenomenal analytic and statistical works, we also need to appreciate the experimental apparatus’s evolution. Although the simple design of a bubble chamber or detection of cosmic rays is still a helpful insight for particle physics research, physicists wanted to create the cosmic and very early phase of the universe on our terra. Progress from the linear accelerator to the TeV scale circular high energy accelerator was achieved by physicists, engineers, and others. Its pinnacle came with the construction of the LHC, situated in CERN, the home of UA(1-5).

3.1 The LHC and CMS

The LHC is the world’s largest and highest-energy particle collider. The LHC was built for a decade, from 1998 to 2008. The construction was completed in collaboration with 100 countries and 10000 scientists around the globe, demonstrating the ethos of global cooperation of the physics community and its majestic scale. The LHC construction, which cost 5 billion US Dollars for its construction, costs 5.5 billion dollars per year for its electric and computing power consumption [53]. The LHC is built in a tunnel 328 feet underground at CERN, on the Franco-Swiss border

near Annecy, France, and Geneva, Switzerland. The LHC shoots bunches of protons and lead(Pb) ions near the speed of light, enabled by the 27 km ring of superconducting magnets with several accelerating apparatuses. One can infer the LHC name's origin, given that it is 27km long, shoots hadrons of protons, and collides with each other at 0.9997 fractions of the speed of light.

CERN is an accelerator complex that includes a succession of increasingly higher-energy machines. Each machine accelerates a beam of particles to a threshold of desired energy and injects

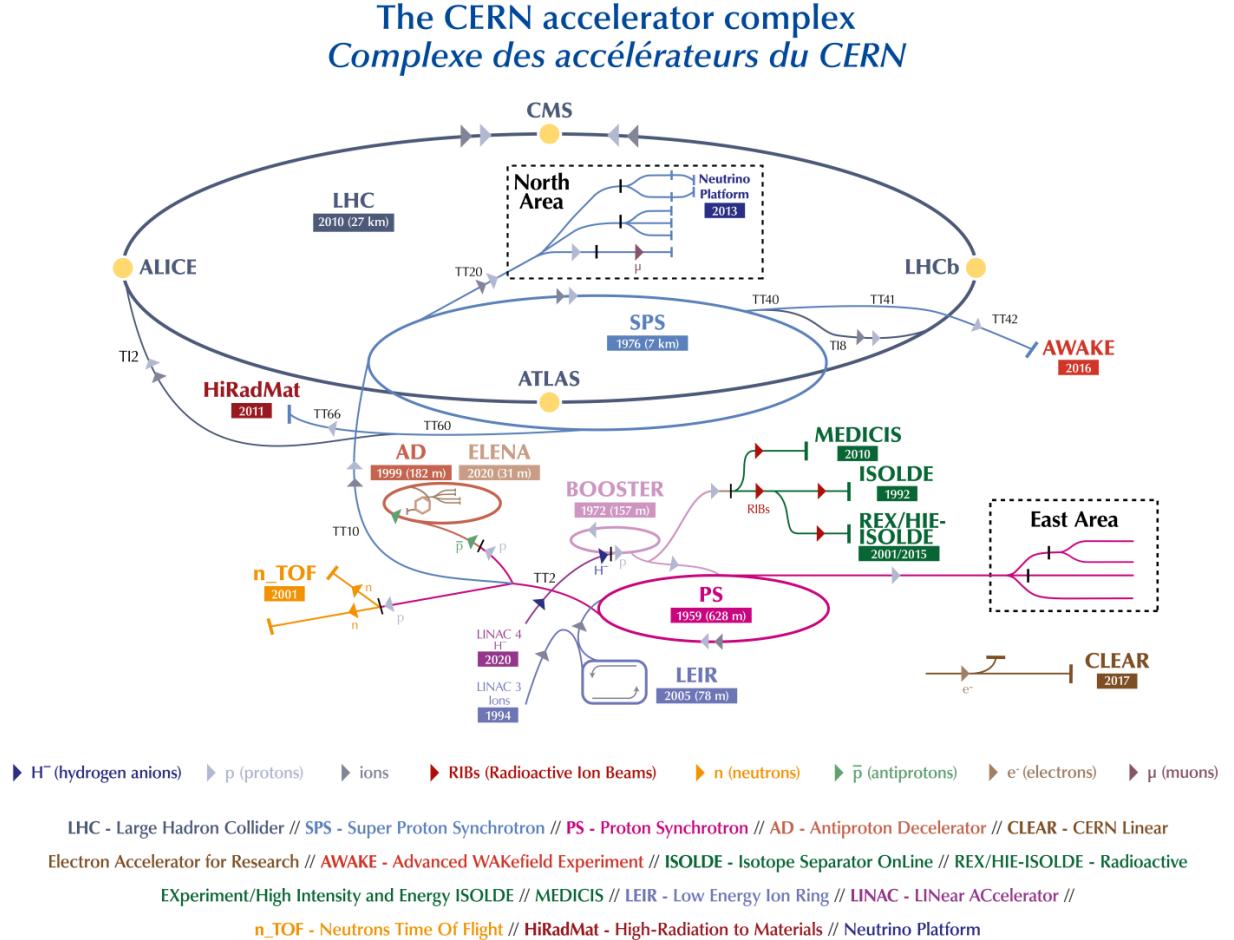


Figure 3.1: Picture of the CERN complex [54]

the beam into the next machine in the chain. The next machine brings the beam to even higher energy and repeats the cycle until it enters the LHC. The LHC is the last element of this chain, in which the beams reach their highest energies. The LHC is cooled to 1.9K and maintained at ultrahigh vacuum status.

The first goal of LHC was to discover the last remaining piece of the SM, the Higgs boson. The LHC accomplished its primary goal with the discovery of the Higgs boson by CMS and ATLAS on July 4th, 2012. However, the LHC’s goal does not stop there. As mentioned in chapter 2, several unanswered questions in high energy physics still exist. As the highest energy collider currently in operation, LHC remains the best option to address these questions. It wants to discover the SUSY particles, dark matter, and other exotic particles to help us better understand the most fundamental nature of the universe.

In order to tackle these issues more efficiently, there are four main detectors in the LHC: Compact Muon Solenoid (CMS), A Toroidal LHC Apparatus (ATLAS), A Large Ion Collider Experiment (ALICE), and LHCb. CMS and ATLAS are general-purpose colliders. They were used for the discovery of the Higgs Boson in 2012. They are used for an entire range of high energy physics, investigating the SUSY particles to dark matter to precision QCD to Lepton Universality. ALICE is a lead-lead collider, targeting to study a phase of matter called the Quark-Gluon Plasma (QGP). A study of QGP helps us better understand quarks and gluons’ behavior when they escape the confinement of the QCD. Similar research is done in Brookhaven National Laboratory’s Pioneering High Energy Nuclear Interaction eXperiment (BNL’s PHENIX). LHCb experiment uses LHC as a b-factory. It wants to test or challenge “Lepton Universality,” claiming that interaction between leptons and a gauge boson measures the same for each lepton. Similar research has been done in BaBar of SLAC and is ongoing in BelleII of KoEnerugi Kasokuki kenkyu kiko (KEK).

The analysis of this dissertation entirely derives from data obtained in CMS. The following subsections detail the parts and functions of each part of CMS. CMS consists of 5 main parts: the tracker, the electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL), the superconducting magnet, and the Muon chamber. We will review each part’s hardware information and its role in the entire CMS. The review order is identical to a particle’s trajectory from the beamspot as in figure ??, except that the tracker is placed at the end for its connection to my analysis’ trigger strategy.

3.1.1 Calorimetry

ECAL of CMS. Electron and photon energies are measured in the ECAL with high precision. CMS has a compact scintillating crystal calorimeter with excellent performance for energy resolution. The ECAL consists of 75,000 lead-tungstate (PbWO_4) crystals with coverage in pseudorapidity up to 3.0. As photons and electrons shower through the crystals, their radiation light is

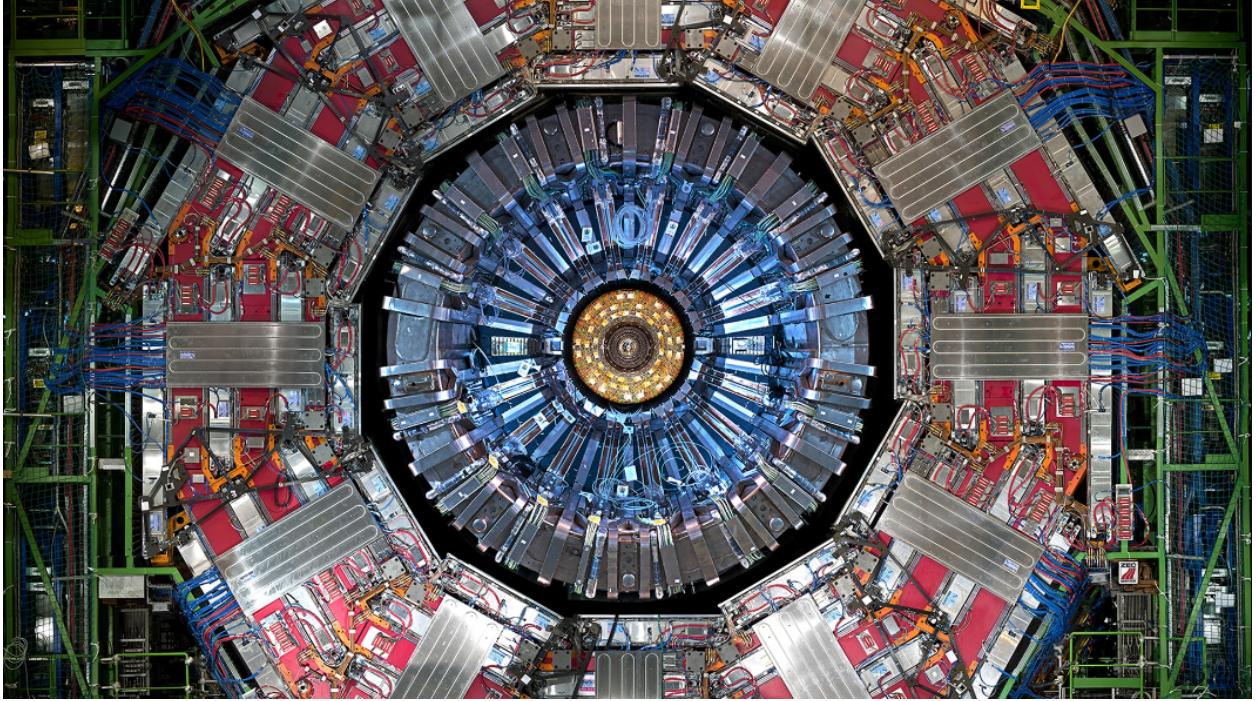


Figure 3.2: Picture of CMS viewed from the beam direction [55]

detected by silicon avalanche photodiodes (APDs) in the barrel and vacuum phototriodes (VPTs) in the endcap. ECAL distinguishes photons from electrons based on the energy dispersion on the η - $i\phi$ map as the photon's energy is deposited into a wider area. ECAL can also distinguish π^0 , which decay into two photons, from photons originating from the hard scattering. ECAL's pre shower system installed in the front of the endcap is used for π^0 rejection and the detection of photons.

HCAL of CMS. The ECAL is surrounded by brass/scintillator sampling Hadron Calorimeter (HCAL) with pseudorapidity (η) up to 3.0. Strongly interacting SM particles, such as quarks, and gluons, deposit their energy into the HCAL. Due to the color confinement of QCD, these particles are in the form of hadrons, and hadrons' shower is messier than electromagnetically interacting particles. We identify these strongly interacting particles as a jet, a narrow cone of hadrons, and other particles produced by the hadronization of a quark or gluon. HCAL plays an essential role in identifying and measuring neutrinos by measuring jets' energy and direction. Since the initial momentum in the transverse plane (xy-plane), which is a plane perpendicular (\perp) to the beam (\hat{z}) direction, is 0, the final transverse momentum is also equal to 0. Suppose the final transverse momentum is not equal to 0, or there is missing transverse energy (MET), E_T^{miss} . In that case,

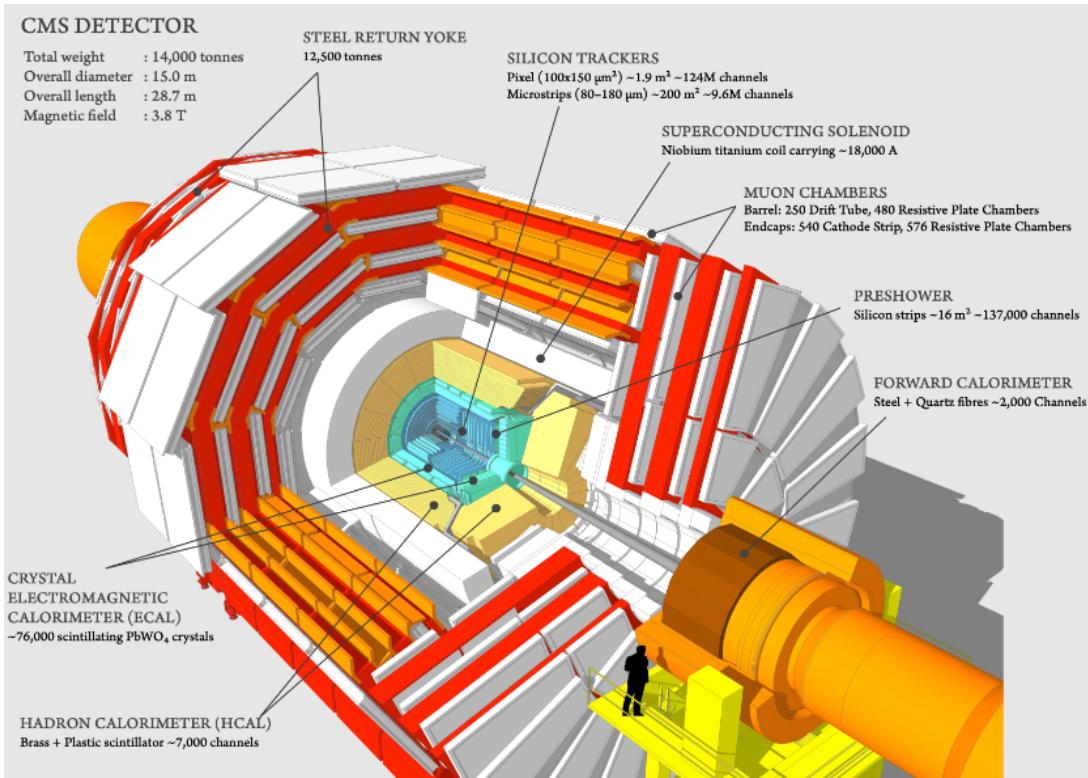


Figure 3.3: Cartoon of CMS with its subpart annotated [56]

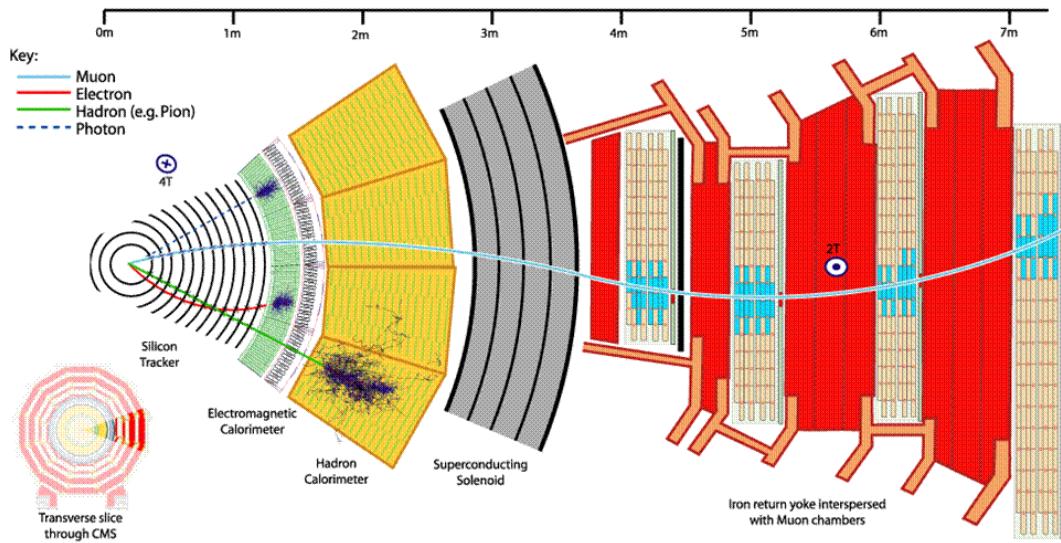


Figure 3.4: Cross-section of CMS detector as a particle traverses through the apparatus [56]

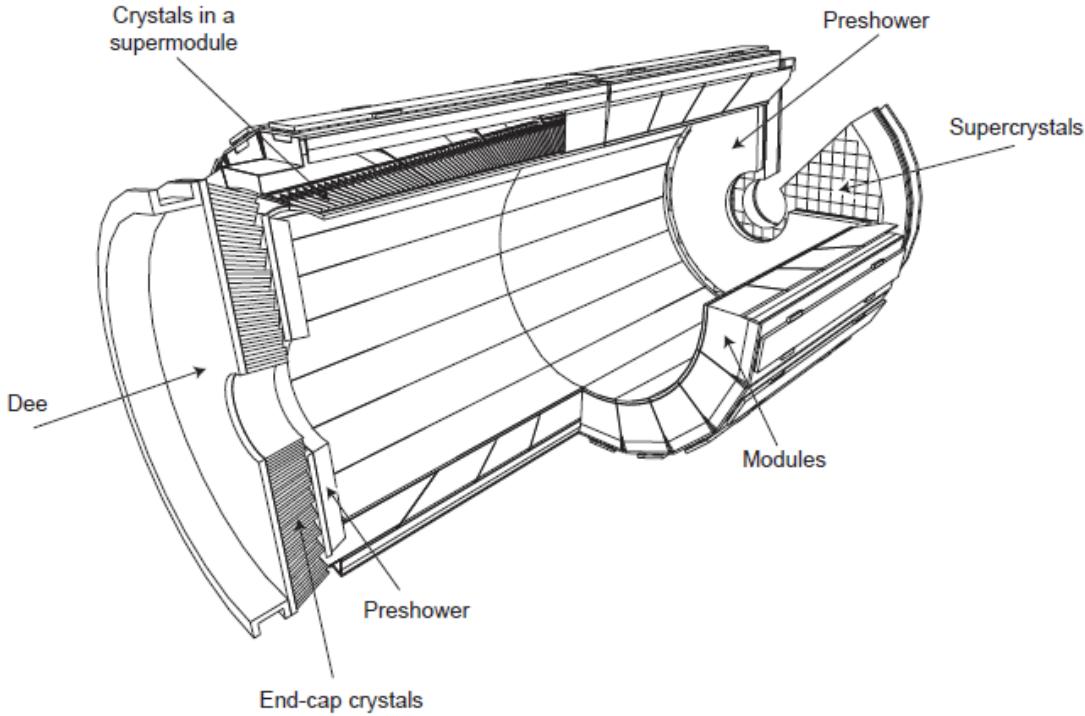


Figure 3.5: Schematic view of the ECAL. [57]

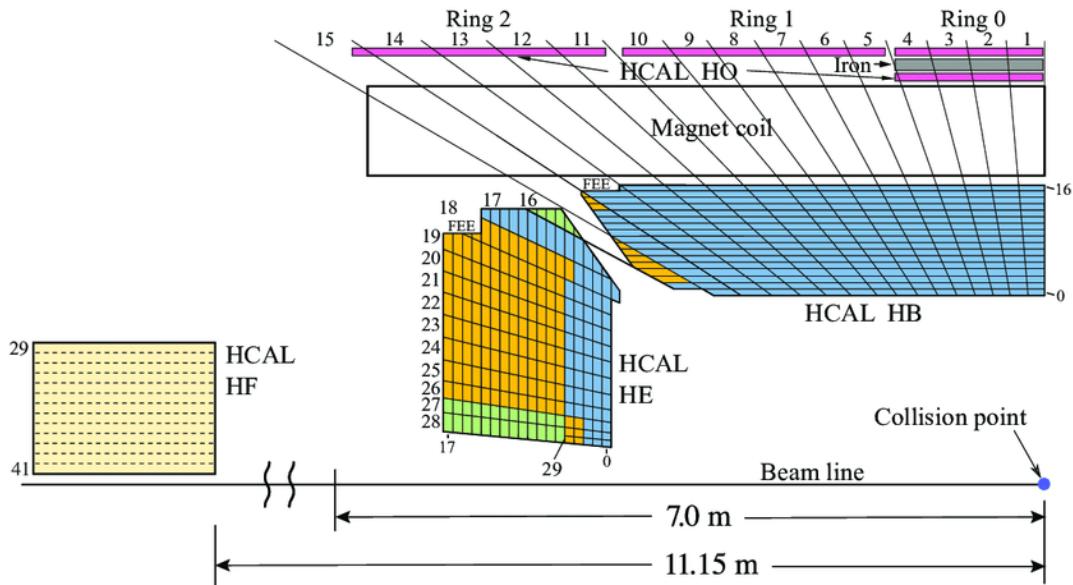


Figure 3.6: Cross-section of the HCAL in CMS detector. It shows the Barrel, endcap, front, and outside portions. [58]

it means there is an undetected particle leaving the collider in the opposite direction of all other detected particles' momentum vector summation, signaling the identification of neutrinos. However,

it could also be interpreted as a sign of a BSM particle based on other information. Thus, E_T^{miss} plays a crucial role in finding new BSM particles, like the supersymmetric partners of quarks and gluons. In addition, precise measurement of high energy jets is vital for searches for high mass Standard Model and SMEFT study.

Since HCAL plays such a crucial role in detecting new physics, the structure of HCAL is also highly complex. Wavelength-shifting (WLS) fibers transport the scintillation light produced by the scintillator tiles to photodetectors. This light is detected by hybrid photodiodes (HPDs) that can operate in high axial magnetic fields. While most of the HCAL’s Barrel (HB) and endcap (HE) are positioned inside CMS magnet, HCAL outside and front (HO,HF) are located outside the magnet to detect particles from high-energy showers. The HB and HE sample calorimeters are with 50 mm thick copper absorber plates interleaved with 4 mm thick scintillator sheets. The HCAL’s Barrel is complemented by a “tail-catcher” to ensure that hadronic showers are sampled with nearly 11 hadronic interaction lengths to contain high-energy jets. HF is installed at each end of CMS detector, which provides coverage up to a pseudorapidity of 5.0, with steel absorber plates used for the harsher radiation environment of the forward systems. The HF ensures full geometric coverage for measuring the transverse energy in the event. HCAL’s comprehensive geometrical coverage and precise energy measurements are crucial for BSM searches in Vector-Boson Fusion (VBF) Higgs production mode, where high energetic jets decay back-to-back at high pseudorapidity. Muons and tau leptons deposit only a tiny fraction of their energy in the calorimeters and are identified with tracking and muon detector subsystems’ information at the reconstruction level.

3.1.2 The superconducting Magnet of CMS

The superconducting magnet encompasses the inner tracker and the calorimetry, while outside of the superconducting magnet is the flux return system and muon detector. The superconducting magnet is 13m-long, 5.9 m inner diameter, 12 kilo-ton, 4 T, the core part of CMS. CMS magnet system consists of a superconducting coil, the magnet yoke (barrel and endcap), a vacuum tank, ancillaries such as cryogenics, power supplies, and process controls. The magnetic flux is returned via a 1.5 m thick saturated iron yoke interleaved in four stations of Muon Chamber.

The magnetic field provided by the superconducting magnet is essential for the momentum measurement of charged particles. Electrically charged particles’ trajectories are bent inside the magnetic field due to the Lorentz force. The particles leave their trajectories in the tracker. Furthermore, the trajectory is used to calculate the momentum of each particle. The 4T magnetic

field also enables the detection of isolated electrons produced by the decay of b , W , and Z particles. CMS ECAL uses these electrons for calibration accuracy to a fraction of a percent.

3.1.3 The Muon Chamber of CMS

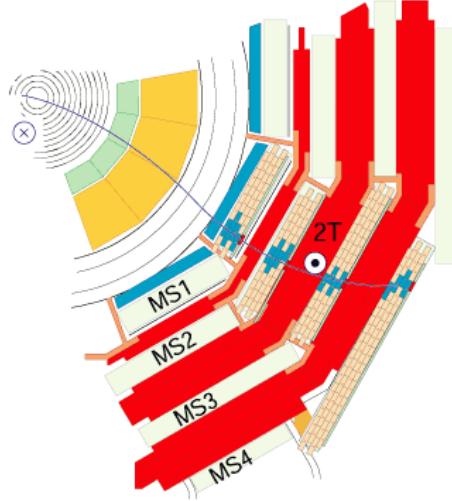


Figure 3.7: Cross-section of the Muon Chamer in CMS detector. It shows the 4 layers of the drift tube (DT) cross-section viewed from the z-axis. [55]

Muons, stable in CMS perspective due to their lifetime, provide clean signatures, unlike hadrons. Muon system's reconstruction efficiency is better than 98% over the full pseudorapidity range. Each muon station has several layers of aluminum drift tubes (DT) for the barrel region and cathode strip chambers (CSCs) for the endcap region. The DT and CSC detectors are used to obtain a precise measurement of the four vectors of the muons. The DT and CSC are complemented by resistive plate chambers (RPCs), fast gaseous muon detectors that provide a muon trigger system. The considerable thickness of the absorber material (iron) ensures muons do not escape the detector, thereby increasing its identification efficiency.

Muons detected and triggered by the Muon Chamber feed into muon reconstruction algorithms. In reconstruction algorithms, position, direction vectors, and an estimate of the transverse muon momentum are used as seeds for the track fits using the Kalman filter technique. The result is a collection of reco::track objects, reconstructed as “standalone muons.” For each standalone muon track, a search for tracks matching it among those reconstructed in the inner tracking system (“inner tracks” or “silicon tracks”) is performed. Based on the Kalman filter technique, the best-matching tracker track and “standalone muon” pair gives a collection of reco::Track objects referred

to as “global muons.” A complementary approach considers all tracker tracks to be potential muon candidates. It provides a collection of reco::Track objects referred to as “tracker muons.” with compatible signatures in the calorimeters and the muon system.

3.2 Tracker of CMS

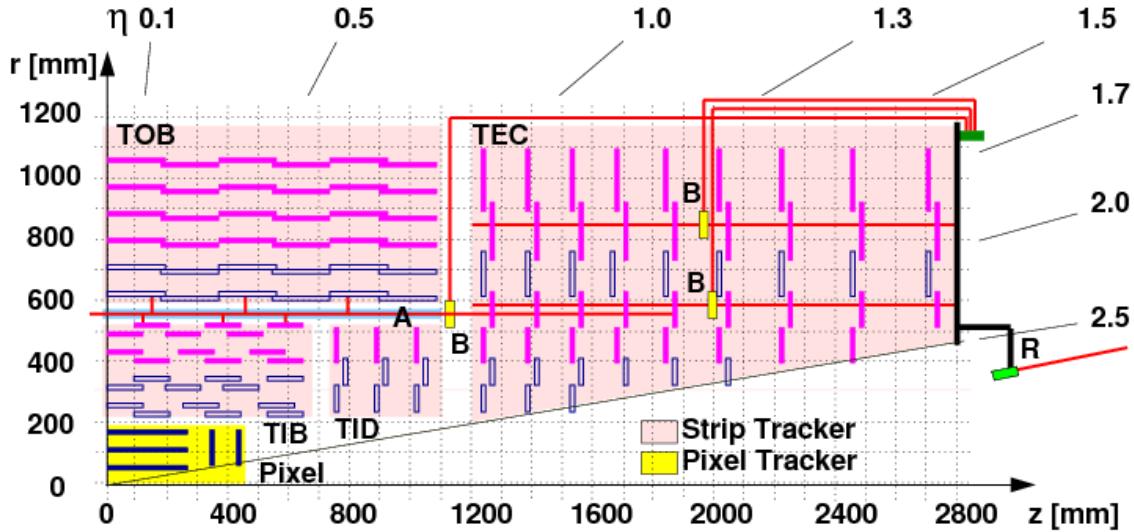


Figure 3.8: Cross-section of the trackers in CMS detector. It shows the pixels in the inner tracker for more precise vertexing and the silicon strips on the outer trackers. Silicon strips are tilted with respect to previous layers of strips. [59]

The inner tracker is the first detector material sitting around the LHC beampipe. It consists of the pixel cells in the innermost part and silicon strip sensors in the outer part of the tracker. The innermost tracking material consists of 3 layers of silicon pixel detectors. In the year of 2017 of Run 2, an additional pixel layer was added for even better performance. The square pixel detectors with high granularity are extremely radiation resistant and provide the most accurate position information. The outer layers of the tracker consist of strip sensors, which are more financially affordable. They are 5.8 m in length, 2.6m in diameter, in ten layers, and consist of 25,000 silicon strip sensors in a cylindrical shape. The system provides analog data from about 10 million channels. Each layer is oriented to a slight off-angle to the previous layer for 2D position measurement.

CMS tracking system, with the magnetic field coming from the superconducting magnet, reconstruct muons, electrons, and charged hadrons’ tracks with high momentum resolution with an

efficiency better than 98% in pseudorapidity up to 2.5. They provide the required granularity and precision to deal with high-track multiplicities. For muons, with information from the measurements of the muon chamber, the tracking system contributes to even better resolution.

The tracker is quintessential for many physics analyses. Its closest placement to the interaction point provides precise vertex reconstruction and measurement of the tracks' impact parameter (IP). The trackers' reconstruction for secondary vertices is critical for b-tagging, a method of detecting jets arising from b decays. It is a frequently used tool in many CMS groups. For instance, physicists studying Lepton Universality use b-tagging, enabled by the tracker's good vertex reconstruction efficiency, to identify the physics process involving B mesons to Kaons. In addition, physicists use b-tagging to identify top quarks. Top quarks predominantly decay into b-jets and are good portals for BSM physics along with the Higgs boson. The tracker's impact parameter values also play a significant role in discovering BSM physics targetting for LLPs. LLPs IP values are starkly different from SM particles' IP values.

3.3 Trigger of CMS

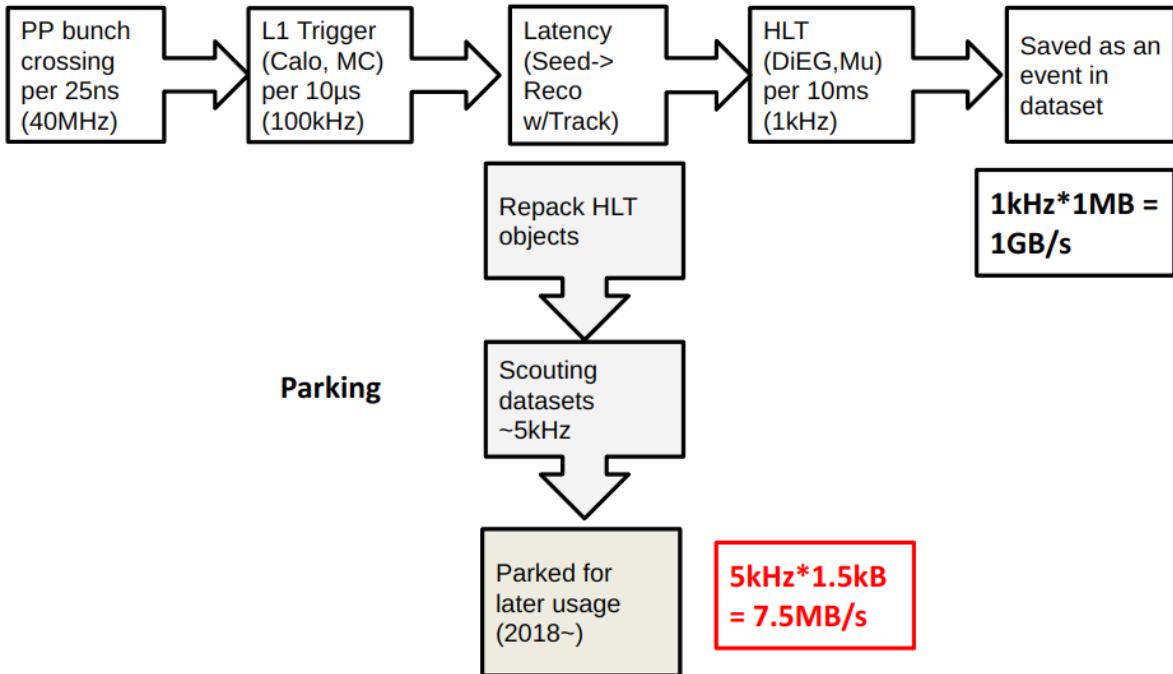


Figure 3.9: Cartoon diagram displaying the Trigger and Data Acquisition System process, along with parking of datasets/

The Discovery of new physics requires high energy and extensive statistics. New physics, as it has not been discovered so far, has a small cross-section. Extensive statistics are necessary to confirm the existence of the small cross-section signal over standard model backgrounds. In order to achieve the high statistics, CMS beam crossing happens at a very high rate with bunches of protons shoted together. CMS beam crossings occur at a rate of 40 million per second (40MHz) with a spacing of 25ns, while 25-50 collisions per bunch crossing are usual figures for Run 2 data of CMS. It totals about 1 billion events occurring in CMS detector per second.

Proton-proton interactions are messy and produce many tracks, given their QCD nature. However, most of the events are “not interesting” to even be considered candidates for potentially interesting events. These events occur too quickly to be recorded and would take up vast amounts of disk space to store, which would be a waste of electronic resources. In order to filter and save only “interesting” events in the extreme environment, CMS employs the Trigger and Data Acquisition System. It selects the most interesting hundred or so events per second for storage with fast electronics and resolution. CMS Data Acquisition System and Triggering are described below.

The first level of triggering, Level 1 (L1), is a hardware trigger. It uses hardware processors to rapidly select or reject events based on information from the ECAL and Muon Chamber. The Level 1 trigger reduces the event rate from 40MHz to 100kHz. An event passing the L1 trigger is transmitted to the Data Acquisition System. In addition, latency is invoked in CMS, so data-taking or additional L1 trigger does not happen until the full Trigger and Data Acquisition System is finalized. The passed event is reconstructed by CMS Software (CMSSW) with information from the 16M channels in CMS subdetector systems. It also prepares the passed event to be scouted, then parked into a dataset. This strategy is implemented for the physics process where the physics event happens frequently. The CPU cannot store all the information with full reconstruction, but it is worth revisiting due to its characteristics.

Events that pass the L1 trigger are subject to the High Level Trigger. The High Level Trigger (HLT) is software-based. The HLT decision reduces the event rate to about 1kHz for storage, with a processing time of about 40ms per event. It also uses tracker information and requires a more stringent pt, η cut on triggering objects. An event that passed a specific HLT is saved into the HLT’s dataset and publicly available to CMS collaborators for their analysis. CMS, running at designed luminosity, records 12 Petabytes of data with the storing scheme.

A rough schematic of the trigger and data acquisition process is depicted in figure ??.

CHAPTER 4

B PARKING TRIGGER STRATEGY

The analysis signal process contains displaced SM τ leptons in their final state. To exploit the leptonic decay of τ lepton with significant IP, specifically with the final muon state for a clean signal, the B-Parking triggers are used. CMS implemented the B-Parking trigger in 2018 of Run 2 to research lepton universalities. As described in chapter 3, lepton universality tests claim that interaction between leptons and a gauge boson measures the same for each lepton. In mathematical expression, $R(K^*, D^*)$ defined in formulae 4.1 and 4.2, are tested.

$$R(K^*) = \frac{B \rightarrow K^* \mu^+ \mu^-}{B \rightarrow K^* e^+ e^-} \quad (4.1)$$

$$R(D^*) = \frac{B \rightarrow D^* \tau^\pm \nu}{B \rightarrow D^* \mu^\pm \nu} \quad (4.2)$$

The first ratio specifically attracts the investigation of many physicists. $R(K^*)$'s physics process is highly suppressed because Flavor-Changing Neutral Current (FCNC) is not allowed in the SM. For a b to change its flavor to a strange quark, it has to go through a loop process with two additional vertices, suppressing the cross-section. Di-muon and di-electron signature of the highly suppressed process is very clean, making it an optimal channel to test lepton universality. Its Feynman diagrams are shown in figures of ?? and ?? Because of this physics reason, the muon final state of B mesons are desired for research of $R(K^*, D^*)$. Consequently, the B-parking trigger requires a soft muon with modest displacement (measured using impact parameter) from the primary vertex, as in b-tagging. It requires a muon with transverse momentum (p_T) of 7-12 GeV with impact parameter (IP) significance 3-6.

PP collisions in LHC produce enormous events, which could trigger the B parking trigger paths. As discussed in chapter 3, the Current CPU capacity of CMS is limited and not capable of reconstructing the entire event at such a high trigger rate at the HLT level. Thus, CMS scouts events, meaning it writes events that passed L1 trigger to a temporary dataset. Later, full HLT and RECO steps are implemented and serve as a B-Parking dataset. The prescale factor for BPH triggers is 5-6.

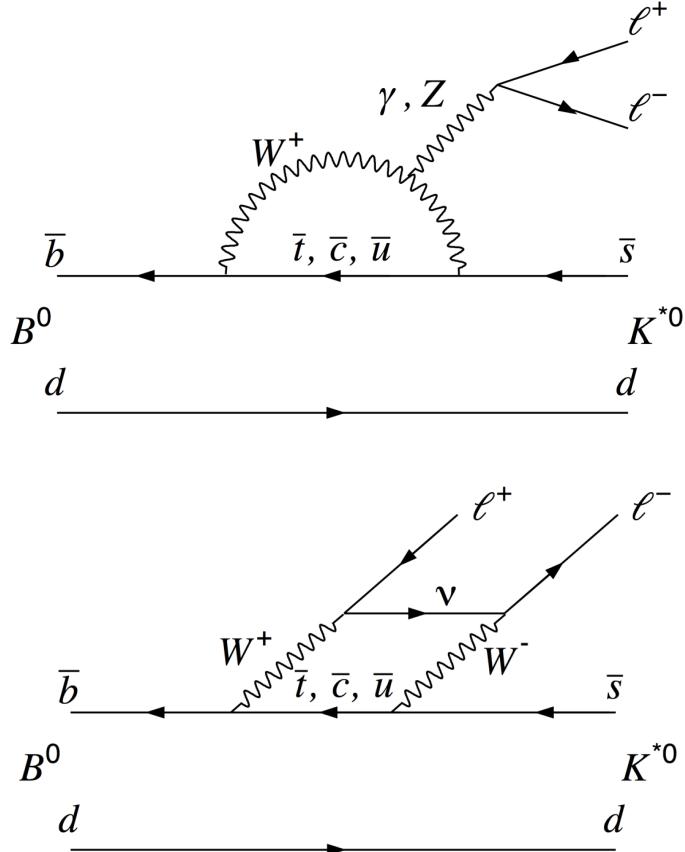


Figure 4.1: The figures show two different loop diagrams for $B \rightarrow K^* \ell^+ \ell^-$ processes. The FCNC is forbidden in the SM and there is no tree level process for $B \rightarrow K^* \ell^+ \ell^-$. Thus, these two processes are the leading contributors for $B \rightarrow K^* \ell^+ \ell^-$, which are highly suppressed.[60]

4.1 Trigger Paths

We use data collected by the B-Parking triggers for the year of 2018. The exact names of paths for B-Parking triggers are listed in Table 4.1. To compare with these data obtained from CMS detector, we get the Monte Carlo (MC) datasets for simulation of signal and background events. Monte Carlo methods are computational algorithms to obtain physics results with statistical randomness. CMS group uses this statistical instrument to publish various signal and background physics process MC datasets. Using the generated signal MC, We plotted distributions of the generator level LLP's physics variables. One can gauge triggering muons' IP significance values, pt , and isolation information.

Figure ?? and ?? are the trigger efficiency of various BPH trigger paths for different mass scale and lifetime points of the signal ($H \rightarrow SS \rightarrow \tau^+ \tau^- \tau^+ \tau^-$) sample. The signal process shows

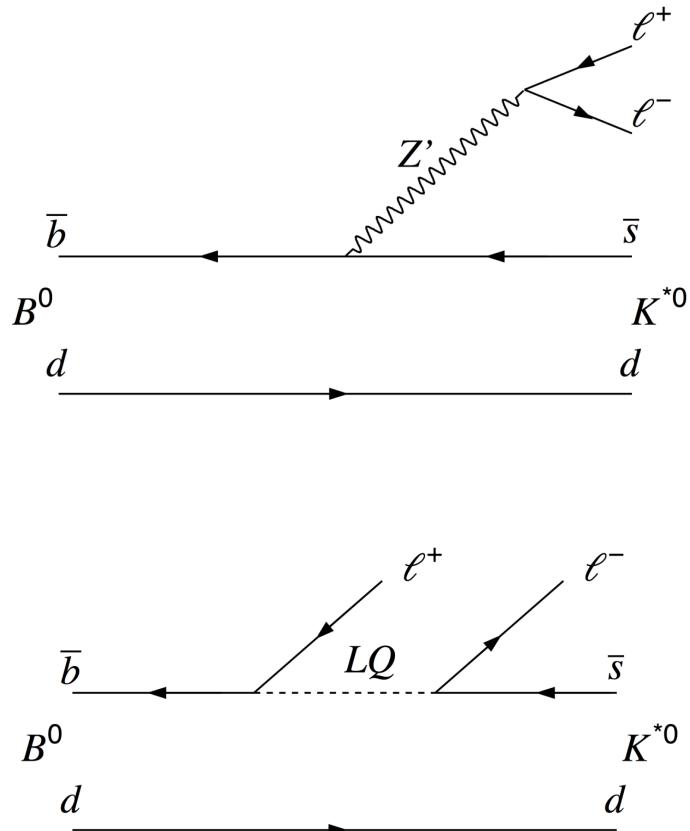


Figure 4.2: If the Lepton universality is not satisfied ($R(K^*, D^*) \neq 1$), it implies there is new physics hidden in the diagram. It could be interpreted in terms of Lepto-quark or Electroweak Z boson which couples to right-hand chiral leptons.[60]

an overall good efficiency. Signal points with LLP's $c\tau = 10\text{mm}$ and $c\tau = 100\text{mm}$ show the best performance. Signal points with LLP's $c\tau = 1000\text{mm}$ likely decay outside of the tracker region, leaving no track's impact parameter, and fails to pass the trigger. On the other hand signal point with LLP's $c\tau = 1\text{mm}$ may not reach the first-pixel detector, which is at 2.7cm from the beam spot, and fails to pass the trigger. We can confirm this explanation by observing that lighter LLP has better trigger efficiency for a shorter lifetime thanks to a more significant boost and vice versa for heavier LLP.

In contrast to the signal, the background processes have poor trigger efficiency. $t\bar{t}$ Jets, Single Tops, and QCD pass the trigger at low but non-negligible efficiency. All these background processes have heavy flavor particles for their final state (b or t quark). From the trigger efficiency and enormous cross-section of the QCD process, we can infer that the QCD process will become a

Table 4.1: HLT trigger paths used in the analysis

Data sample	Trigger
ParkingBPH*-Run2018A	HLT_Mu9_IP6_part*
ParkingBPH*-Run2018B	HLT_Mu9_IP6_part*
ParkingBPH*-Run2018C	HLT_Mu12_IP6_part*
ParkingBPH*-Run2018D	HLT_Mu12_IP6_part*

Table 4.2: Data and MC Global tags used for 2018 datasets

Data 2018	106X_dataRun2_v29
MC 2018	106X_upgrade2018_realistic_v11_L1v1

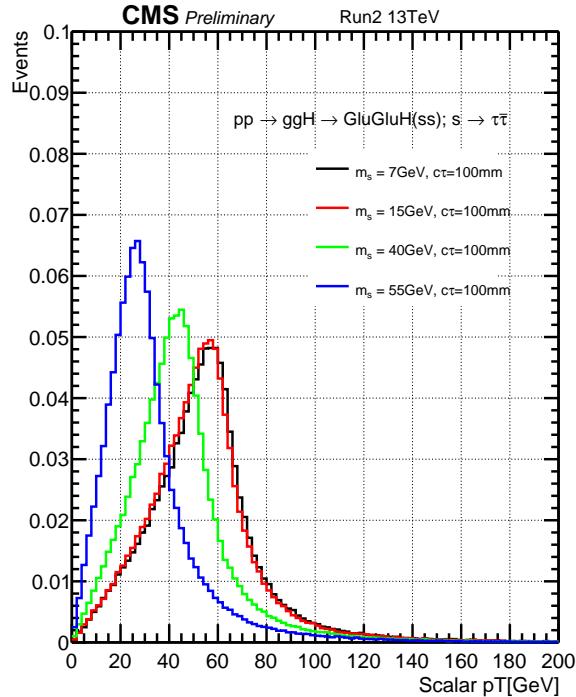


Figure 4.3: p_T of the scalar products

significant contribution to our background. Many physics analyses groups across CMS use the QCD MC datasets. Thanks to their popularity and need for heavy statistics, CMS group generates

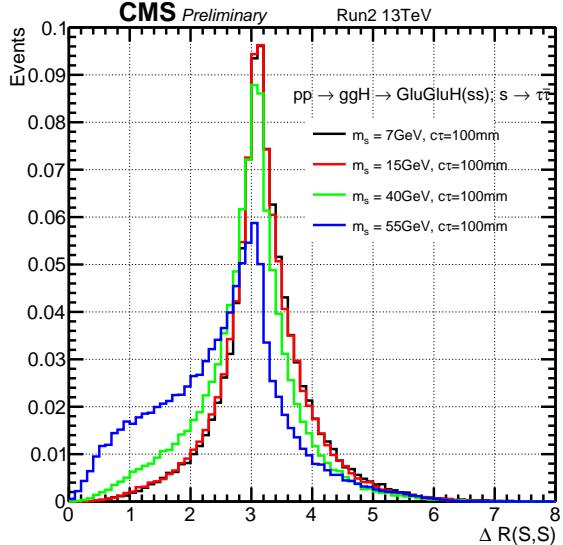


Figure 4.4: DeltaR of the scalar products

QCD MC datasets filtered to different groups' focuses. We use QCD's Muon Enriched datasets to improve background events' statistics, which would pass the B-Parking trigger. CMS group names the dataset as "QCD-MuEnrichedPt5", with QCD's muon decay product passing a p_T threshold of 5 GeV. QCD-MuEnriched has better efficiency for higher p_T bin samples since higher p_T bin samples tend to have more b-jets for their final state. Drell-Yan and W-Jet processes have a very poor trigger efficiency due to the absence of heavy flavor particles in their final state.

4.2 Integrated Luminosity and pileup weight for the HLT path

The integrated luminosity for each era has been summarized in table A.1 in Appendix A. The information was obtained with commands in section A.1 of Appendix A. The integrated Luminosity totals at $44 fb^{-1}$ lower than $58.7 fb^{-1}$ for the year of 2018. The bunch-crossing for the B-parking HLT path is also very different from other HLT paths. As expected, b-parking data are recorded during lower bunch-crossing runs due to its extreme rate in CMS collider.

It is vital to adjust this bunch-crossing variable for MC simulation to model the data correctly. To achieve this purpose, we apply pileup weight to the MC simulation. Pileup weight is simply a number of collisions per bunch-crossing of data divided by the number of collisions per bunch-crossing of MC for a specific era. Pileup (PU) weight values are calculated for each era of data-taking

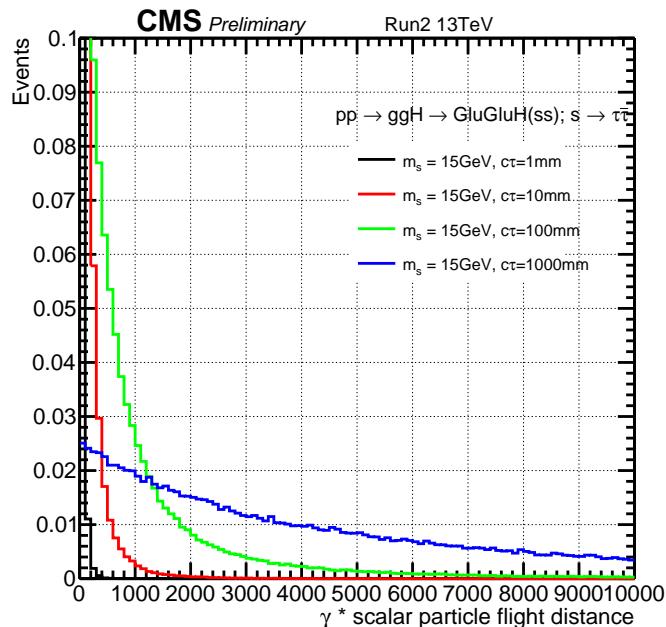
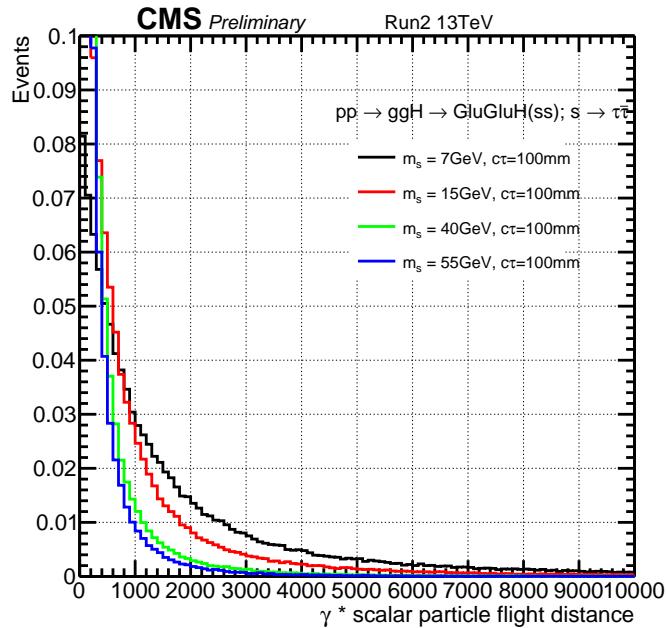


Figure 4.5: lifetime of the scalar products in the lab frame

(A, B, C, D). Figure ?? shows the BPH1-Era A's HLT_Mu9_IP6 HLT path's Data PU distribution.

Figure ?? shows the BPH1-Era B's HLT_Mu9_IP6 HLT path's Data PU distribution.

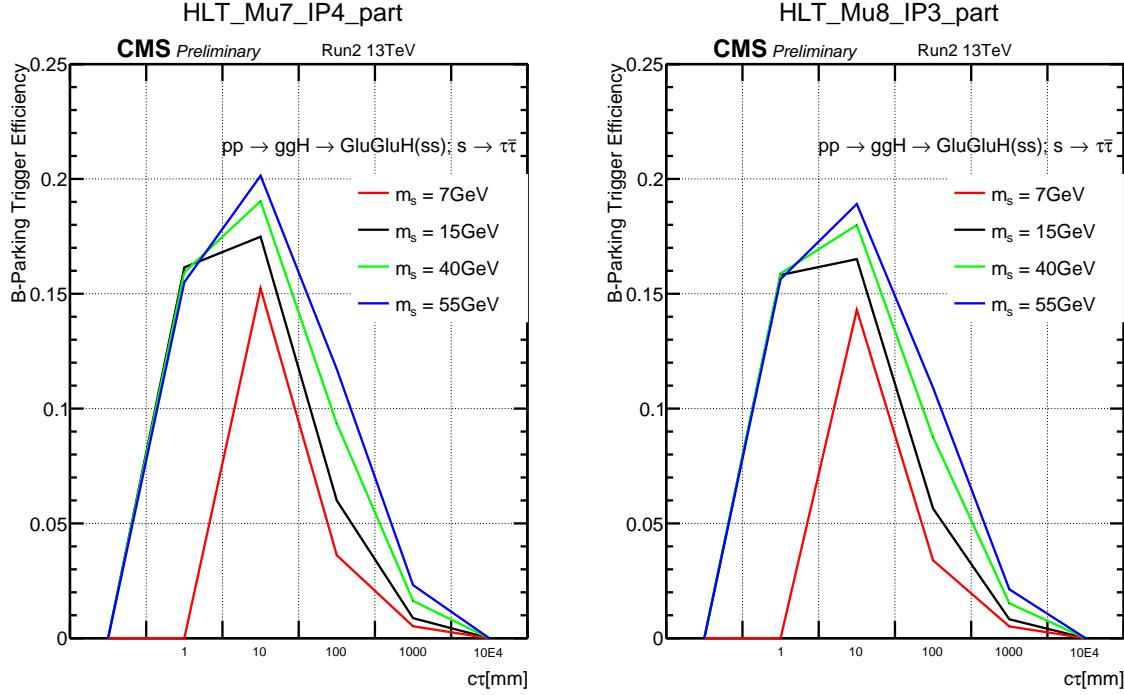


Figure 4.6: The plots show the trigger efficiency for each HLT path with respect to LLP's lifetime. Each line denotes mass scale of each LLP. Please note the efficiency is set to 0 for MS=7 GeV $c\tau = 1\text{mm}$ due to absence of Monte Carlo (MC).

Figure ?? shows the BPH1-Era C's HLT_Mu12_IP6 HLT path's Data PU distribution.

Figure ?? shows the BPH1-Era D's HLT_Mu12_IP6 HLT path's Data PU distribution.

Figure ?? shows DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 MC PU distribution.

Figure ?? shows resultant such PUWeight from BPH1_A HLT_Mu9_IP6 and DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8.

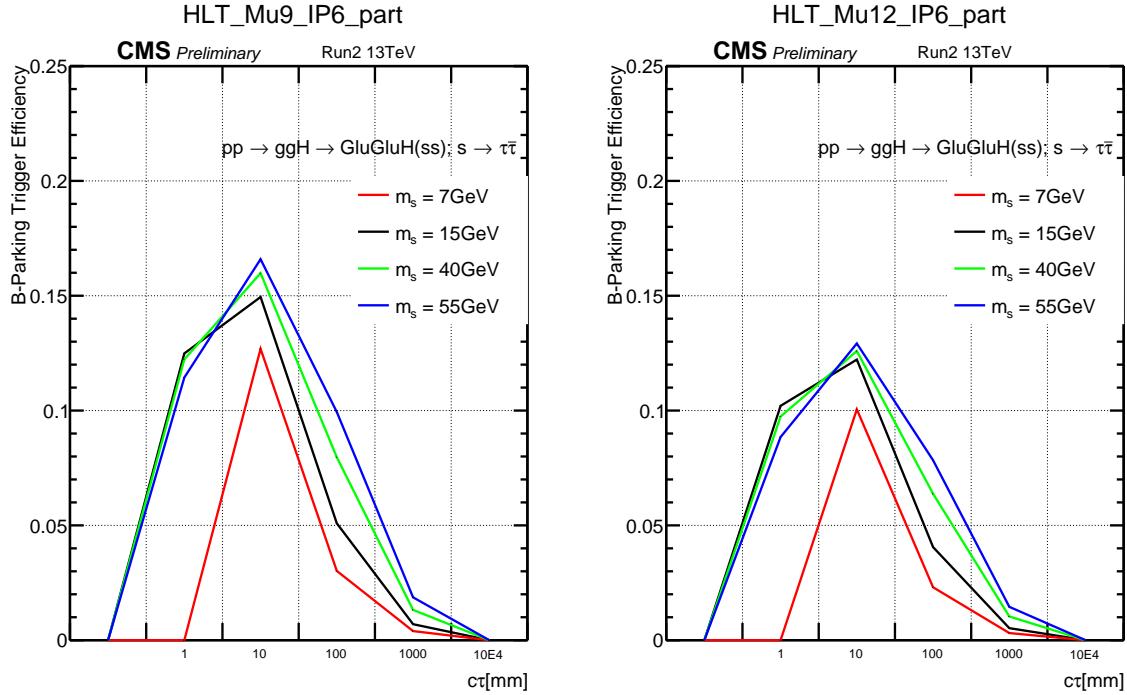


Figure 4.7: The plots show the trigger efficiency for each HLT path with respect to LLP's lifetime. Each line denotes mass scale of each LLP. The analysis uses Mu9_IP6 for Era A,B of data and Mu12_IP6 for Era C,D of data. Please note the efficiency is set to 0 for $MS=7\text{ GeV}$ $c\tau = 1\text{mm}$ due to absence of Monte Carlo (MC).

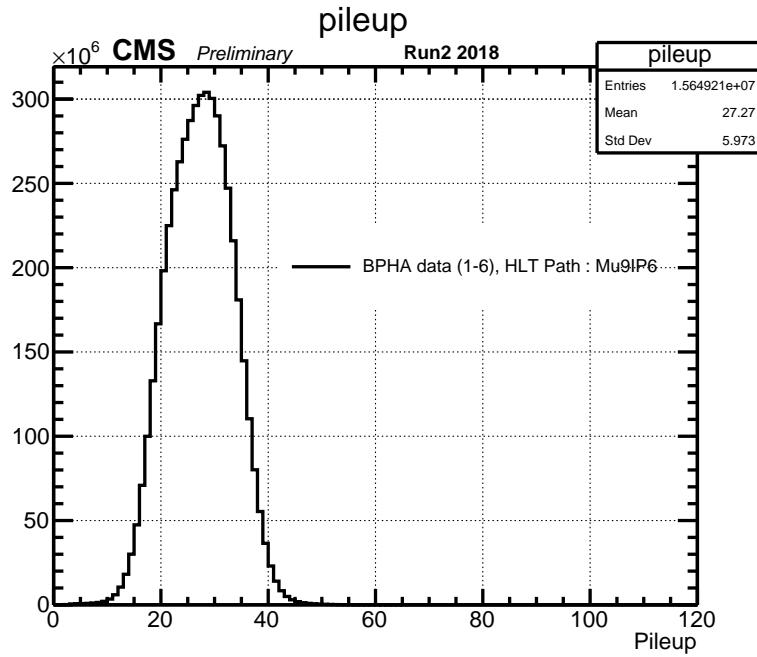


Figure 4.8: Bunch crossing of dataset /ParkingBPH1/Run2018A-UL2018_MiniAODv2-v1/MINIAOD

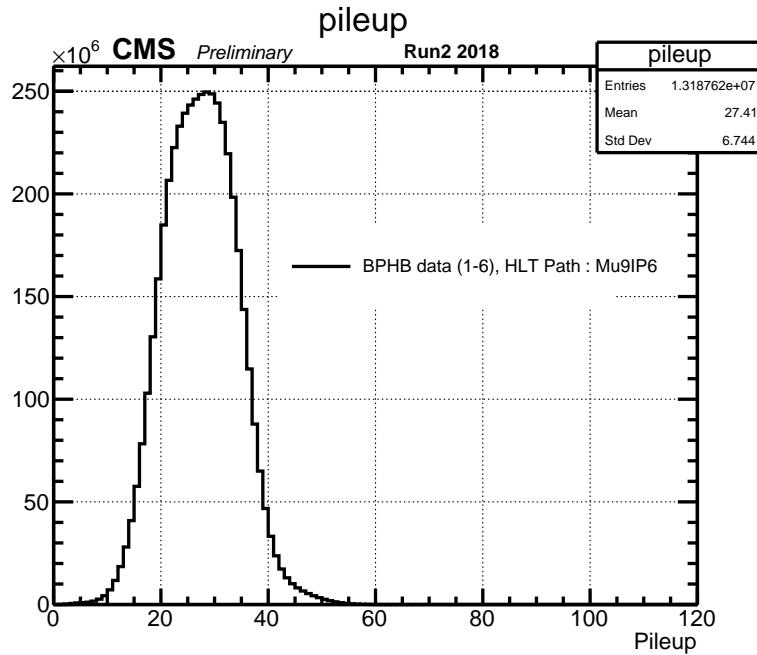


Figure 4.9: Bunch crossing of dataset /ParkingBPH1/Run2018B-UL2018_MiniAODv2-v1/MINIAOD

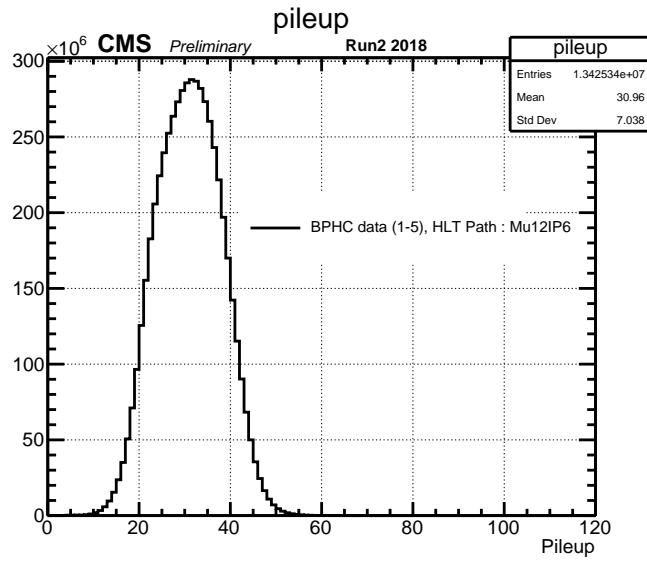


Figure 4.10: Bunch crossing of dataset /ParkingBPH1/Run2018C-UL2018_MiniAODv2-v1/MINIAOD. Please note that the HLT path for EraC has higher muon object's p_T threshold with 12 GeV (compared to 9 GeV in EraA).

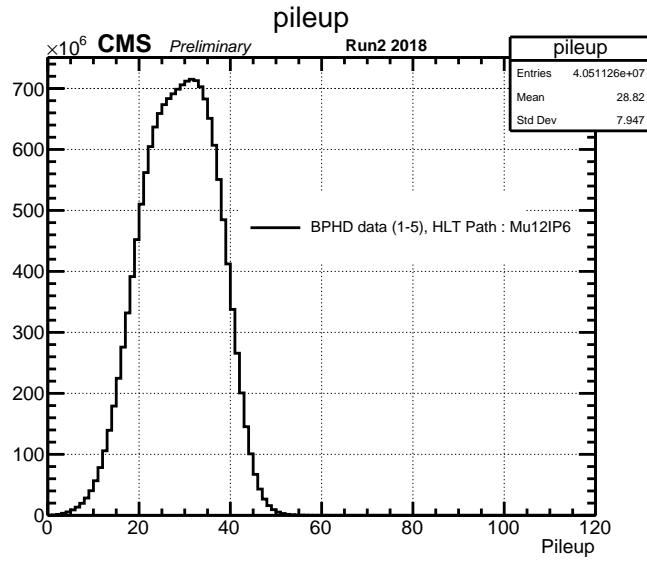


Figure 4.11: Bunch crossing of dataset /ParkingBPH1/Run2018D-UL2018_MiniAODv2-v1/MINIAOD. Please note that the HLT path for EraD has higher muon object's p_T threshold with 12 GeV (compared to 9 GeV in EraA).

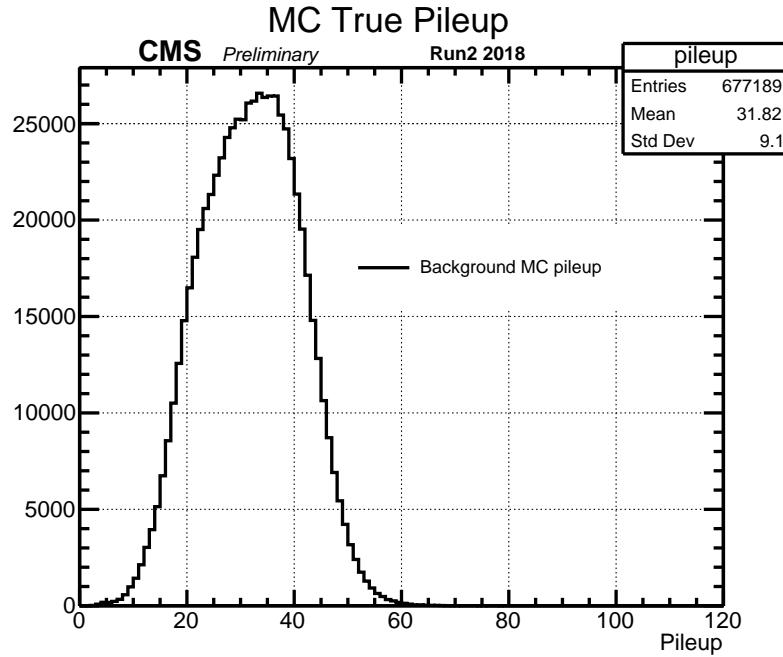


Figure 4.12: Bunch crossing of Monte Carlo Simulation for physics process of Drell-Yan. The dataset is /DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 /RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v2/ MINIAODSIM

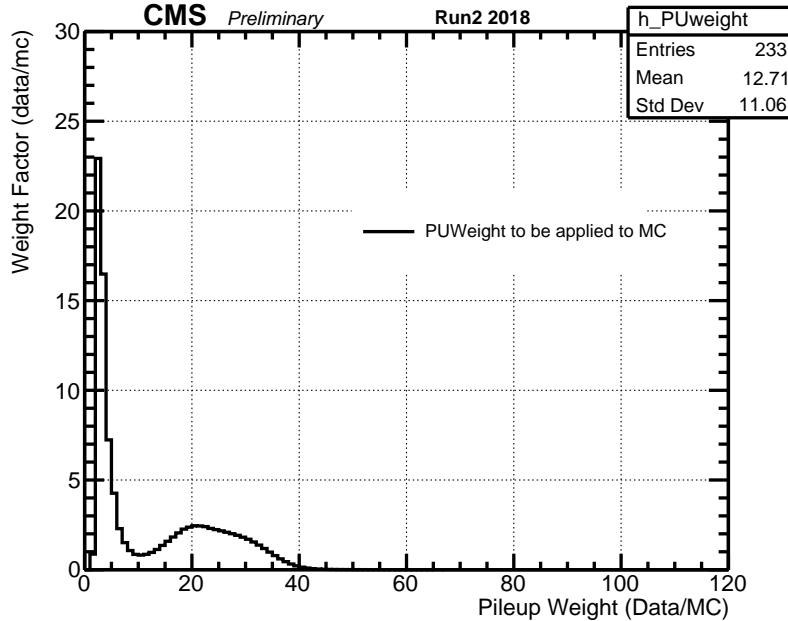


Figure 4.13: PUweight calculated for Era A of B-parking dataset

CHAPTER 5

PHYSICS OBJECT DEFINITIONS

This section provides the definitions of physics objects used in the analysis. The analysis uses two main CMS reconstructed objects: muons and jets. Photons do not show anywhere in our signal process. τ 's of our signal process can decay into electrons with the same probability as into muons. However, the analysis uses b-parking and only targets the muon decay mode of the τ . Electrons are not an interest of our analysis either.

Although the signal's final decay involves τ lepton, we do not use CMS reconstructed τ . Given that τ decay results in many different objects (muon, electron, and different combinations of hadrons) as shown in figure ??, CMS uses a special reconstruction algorithm for τ object. Particle-Flow jets (PFjets), with all particle identification information within the jet, are the starting point for the τ reconstruction. With π^0 reconstructed in ECAL as a strip, the Hadronic-Plus-Strip (HPS) algorithm reconstructs τ (at 60% efficiency in Run 1) when a charged hadron and a strip are reconstructed in the PFJet [61]. Therefore, CMS reconstructed τ collection only includes hadronically decayed τ lepton. One of our signal process tau lepton decays into muon, in a leptonic mode, for the trigger. Although other tau lepton could still decay hadronically, usage of CMS reconstructed τ leptons in the analysis would be accompanied by a division of several different sub-channels (muon, electron, had), which result in 27 (3^3) different combinatorics. In order to circumvent this difficulty, we approach a tracker-oriented method with regions of interest to analyze other τ leptons in the signal process.

We use muons, jets, and only tracks (ROIs) for this analysis.

5.1 Muons

The analysis sources the “SlimmedMuons” collection from the MINIAOD (A simpler format of Analysis Object Data files produced by CMS) MC datasets to produce `selectedPatMuons`. Muon objects are required to have

- $p_T \geq 12 \text{ GeV}$ to reach B-Parking trigger plateau
- $|\eta| \leq 1.5$ for L1 seed $|\eta|$ cut in B-Parking trigger

- reco::Muon object matched to the B-Parking trigger object
- Pass the Loose ID criterion (isLooseMuon) as described in the Muon POG [62].

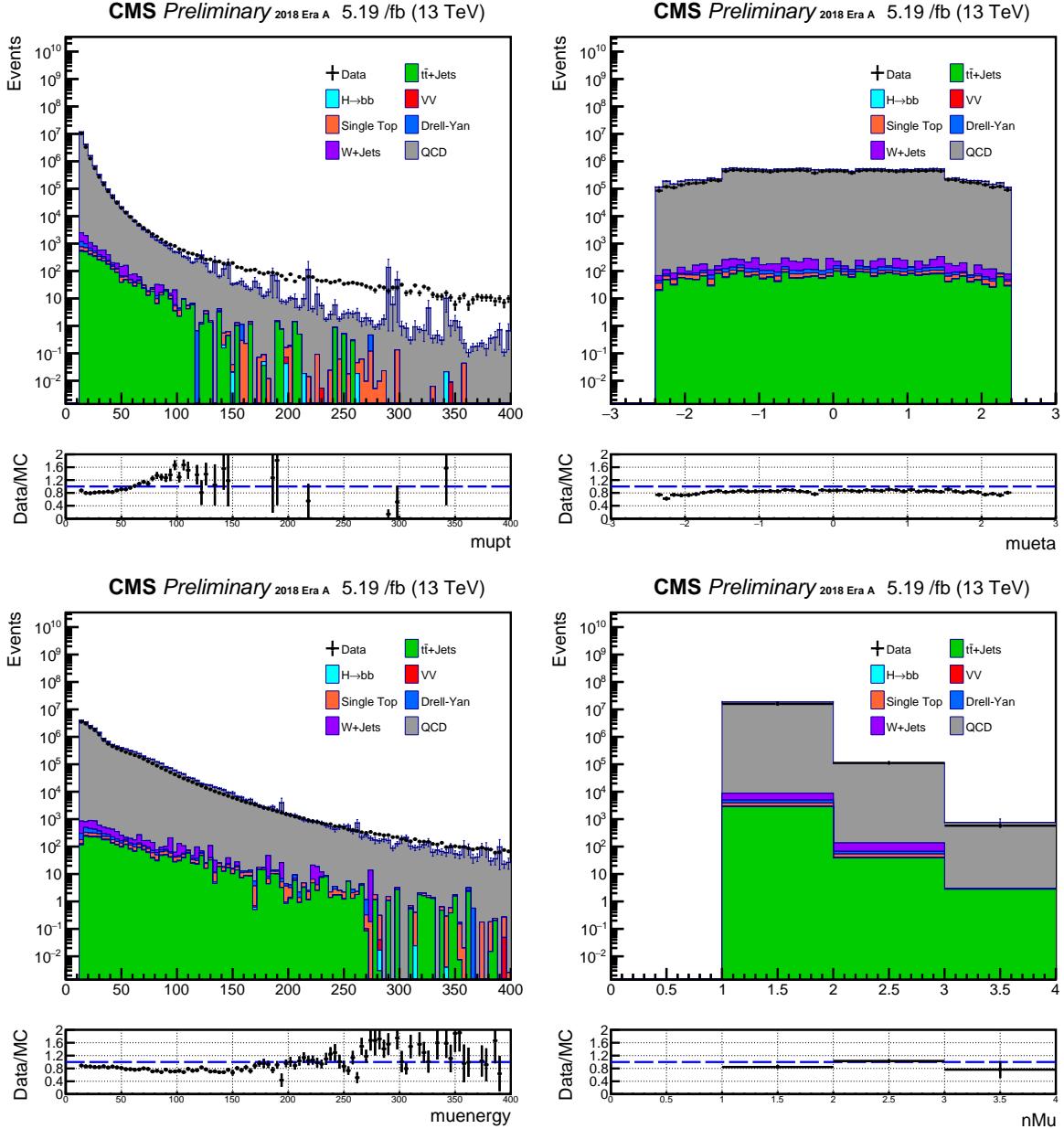


Figure 5.1: Data/MC of muon objects p_{T} , η , energy, n_{Mu}

The motivations for isolation requirements on muons are further discussed in Section 7.

5.2 Jets

The analysis sources “SlimmedJets” collection from MINIAOD dataset to produce `selectedJets`. CMS reconstructs jets from calorimeter energy deposits using the anti- k_T clustering algorithm with a distance parameter of $R = 0.4$ [63]. Then, the calojets are input into the Particle-Flow (PF) algorithms to produce the PFJets collection. Then, variables in PFJets class are slimmed to be saved into MINIAOD files. Jet objects require

- $p_T \geq 20 \text{ GeV}$
- $|\eta| \leq 2.4$
- $0 \leq \text{emEnergyFraction} \leq 0.9$
- $0 \leq \text{energyFractionHadronic} \leq 0.9$
- No selected electron or muon within $\Delta R = 0.4$

The energy fraction cuts above are taken from the recommended Run2 Tight jet ID cuts for particle flow jets [64].

5.3 Taus

Although we do not use τ leptons for event selection or background estimation, we still plotted essential variables of τ leptons for review. The analysis sources “PAT::slimmedTaus” from MINIAOD for MC and “RECO::slimmedTaus” for Data to produce `selectedTaus`. τ ’s hadronic decay can be reconstructed with PFJets’ charged hadrons in HCAL, and 2 γ s from π^0 decays in ECAL. τ objects require

- $p_T \geq 20 \text{ GeV}$
- $|\eta| \leq 2.4$

5.4 Region of Interest

Tracks contain many essential qualities, such as the impact parameter significance and the four-vectors. LLPs decay in the tracker, so these track qualities should be good discriminant against the background. However, we can not save all tracks in the event with all track information, as much as we have to filter out uninteresting events with the trigger and Data Acquisition system in CMS.

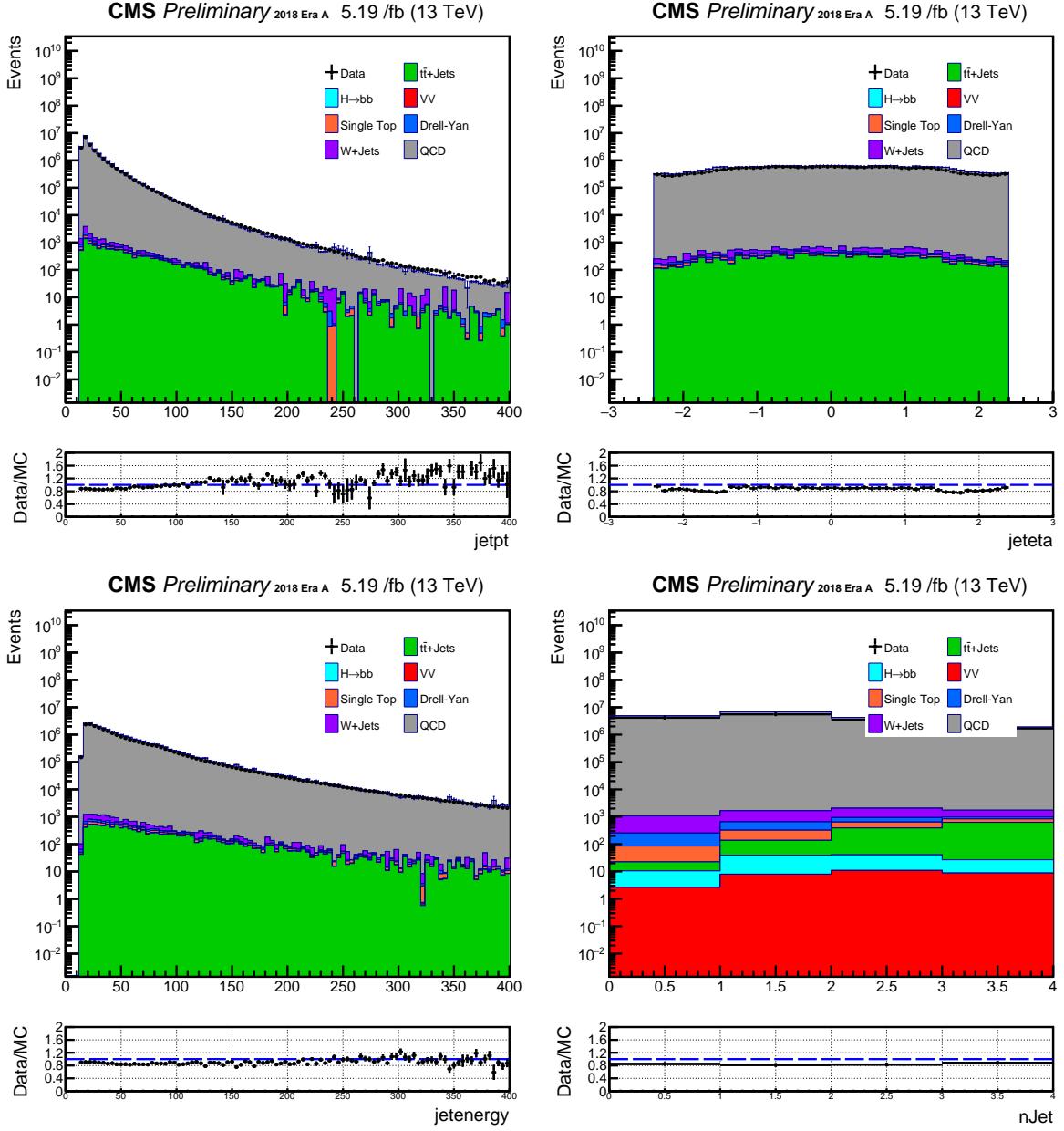


Figure 5.2: Data/MC of jet objects for p_T , η , energy, nJet

In our signal process, a geometrical concept can play a vital role in sorting out only relevant tracks for our analysis. LLP has no electric or color charge as described in section 2. LLP leaves no tracks in the tracker and decays into 2 charged τ leptons, which decay into at least one charged track (muon, electron, at least one charged hadron). Like in the pair production, two charged tracks often travel opposite directions. Thus, a geometric convergence of 2 charged tracks should point to the decay vertex of the neutral LLP. The tracker algorithm we use to construct this “Interesting

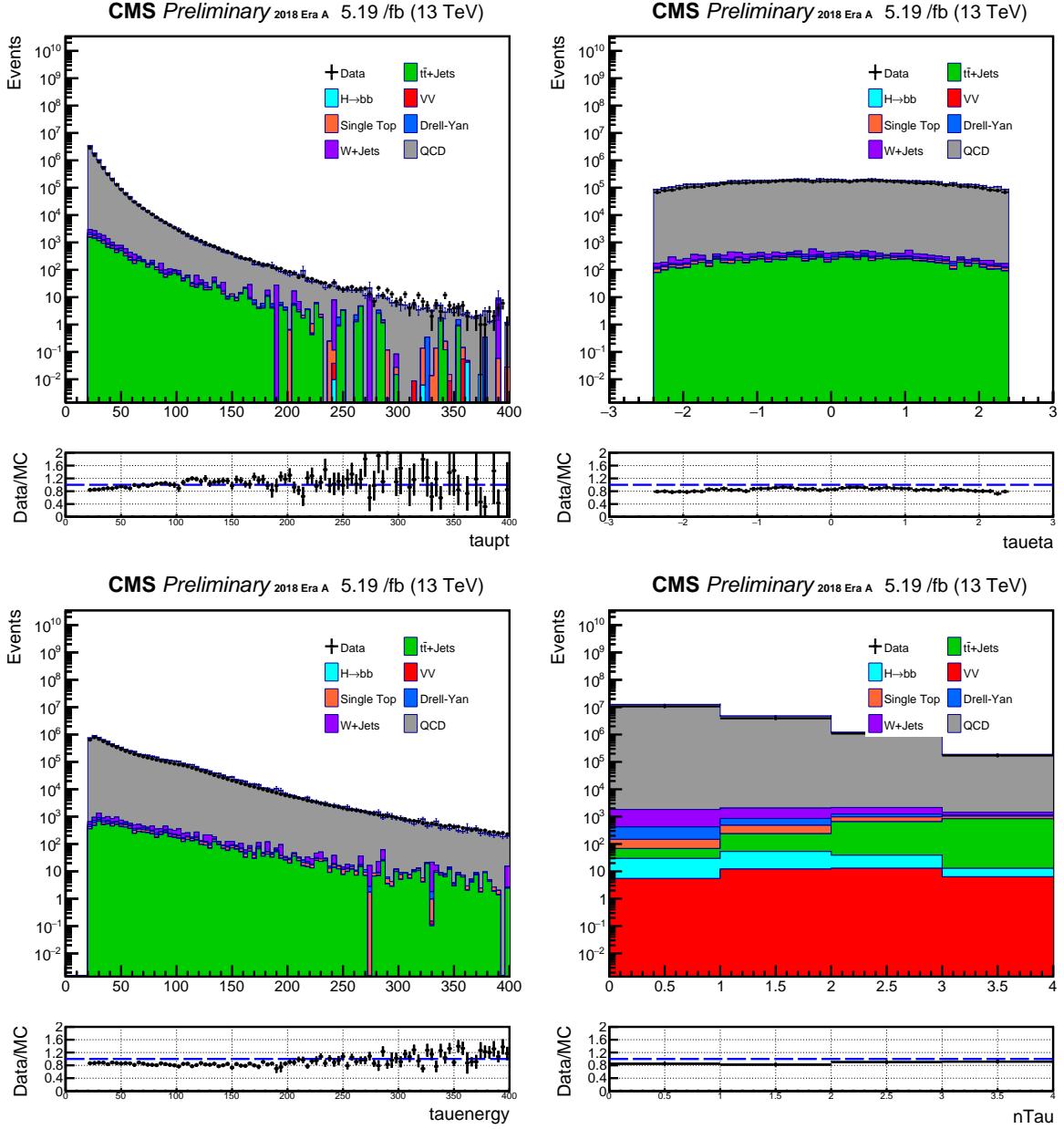


Figure 5.3: Data/MC of τ objects for p_T , η , energy, nJet

“Geometric Region” provides us a converged vertex. The area around this vertex is referred to as the “Region of Interest” (ROI). The complete reconstruction procedure of the Regions of Interest is detailed in the following subsections. An ROI requires

- Good quality track selection
- Vertex Fitted from pair-wise tracks by V0Fitter in CMSSW

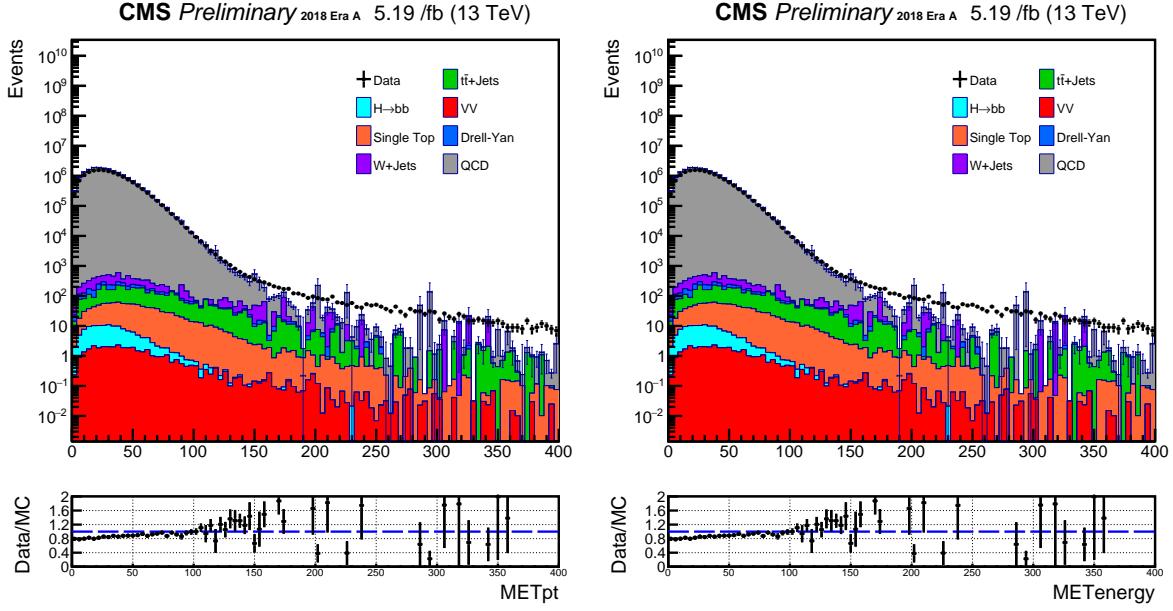


Figure 5.4: Data/MC of MET objects for p_T , energy

- Clustering the fitted vertices to form a Region of Interest (ROI)
- Looking for tracks around $\Delta R = 0.3$ around ROI to save its isolation information

5.4.1 Tracks

The analysis sources “packedPFCandidates” and “lostTracks” from MINIAOD. Track parameters and covariance values will be propagated along the ROI production process and no value should be non-physical

- !isinf(tracks.parameter) and !isnan(tracks.parameter)
- !isinf(tracks.covariance) and !isnan(tracks.covariance)
- Number of valid hits > 3
- $p_T \geq 0.35$
- Track $IPSig_{XY} \geq 2.$
- Track $IPSig_Z \geq -1.$
- Track normalized $\chi^2 \leq 10.$

5.4.2 Vertex Fitter

The analysis sources “offlineBeamspot” from MINIAOD for beamspot reference. Vertex fitter is “KalmanVertexFitter” with vertex cuts as below.

- Vertex $\chi^2 \leq 6.63$
- Transverse Decay distance significance ≥ 15 .
- V0mass ≥ 13000 GeV
- $\cos(\theta_{XY})$ between x and p of V0 candidate ≥ 0
- $\cos(\theta_{XYZ})$ between x and p of V0 candidate ≥ -2

5.4.3 ROI formation

Fitted vertices are clustered to form a Region of Interest (ROI). The clustering steps can be further detailed as below.

- A fitted vertex is merged with another vertex if they are within 1cm.
- A new ROI is formed where the position is the center of mass of 2 merged ROIs.
- Clustering is repeated until there is no other vertex within 1cm limit.

Although vertex clustering can be done with the step above, we can be creative and acquire more information. Isolation, containing information about the ROI’s environment, could also be helpful information. We can obtain information about isolation by defining an isolation shell. The shell’s construction step is described below.

- A cone of $\Delta R \leq 0.3$ around the center of the ROI is defined, where ΔR is calculated with respect to the PV.
- A plane that is perpendicular to the axis of the cone and that contains the center of the ROI is the isolation plane. It becomes a circle.
- Make the circle into a 3D sphere.
- Any tracks that pass through that sphere but not through the ROI are the annulus tracks.

With these essential tracks’ information, we are ready to discriminate signals from the background.

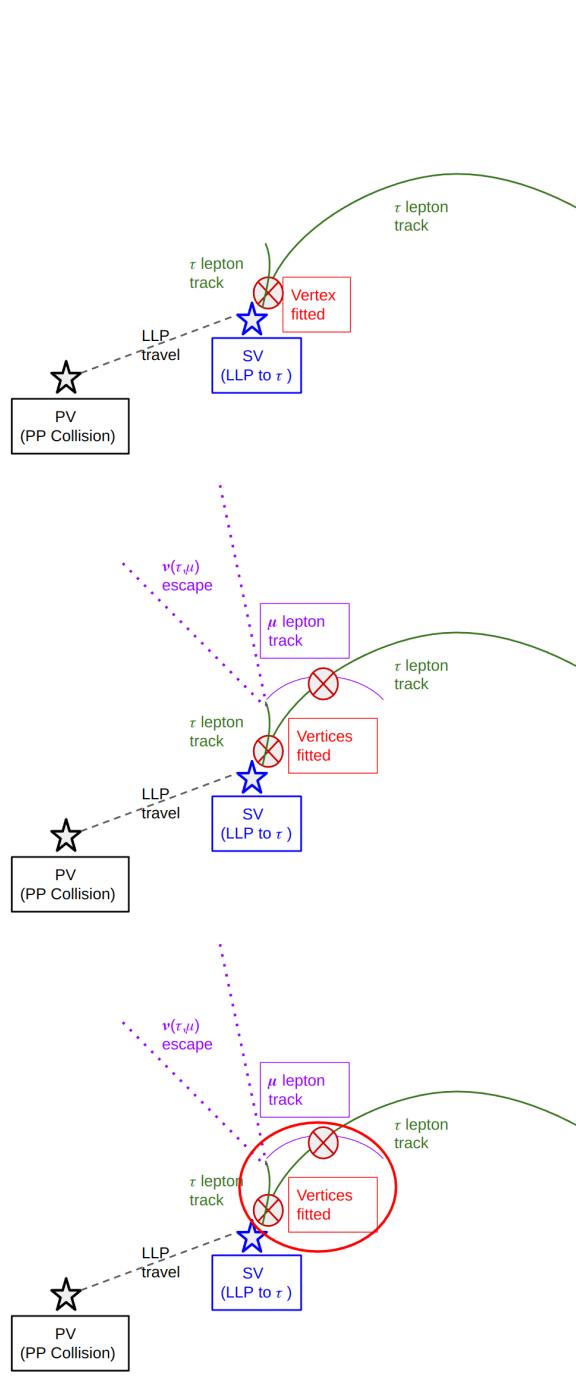


Figure 5.5: The figures display cartoons of the step-by-step formation of the Region of Interest (ROI). After LLP travels from the Primary Vertex (PV) to Secondary Vertex (SV), it leaves no track, and decays into a tau lepton pair. Tau lepton pair's tracks are recorded in the tracker, as denoted in green color lines. The first figure shows the vertex (red star) fitted by KalmanVertexFitter in CMSSW from two charged tracks that decayed from a neutral LLP. The second figure shows fitting procedure of another vertex. With one of tau leptons decaying into a muon, the muon track crosses the partner tau lepton track and forms a new vertex near the original vertex. The third figure shows the clustering of the vertex, with the second vertex formed by the extended tracks being clustered into the original ROI if the vertices are within a 1cm limit.

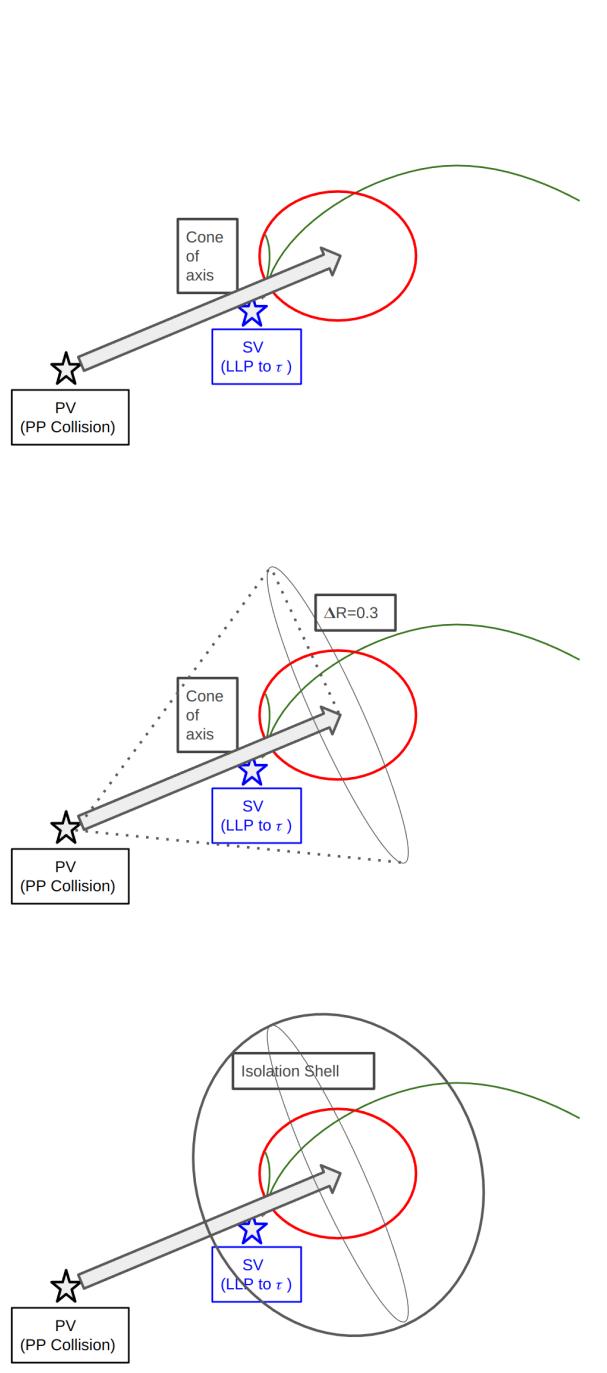


Figure 5.6: The figures display cartoons of the step-by-step formation of the shell structure. The first figure shows the axis of the cone when the axis goes through the center of ROI. The second figure shows the construction of the isolation plane, with $\Delta = 0.3$ parameter. The third figure shows the construction of the isolation shell, with the isolation plane becoming a sphere.

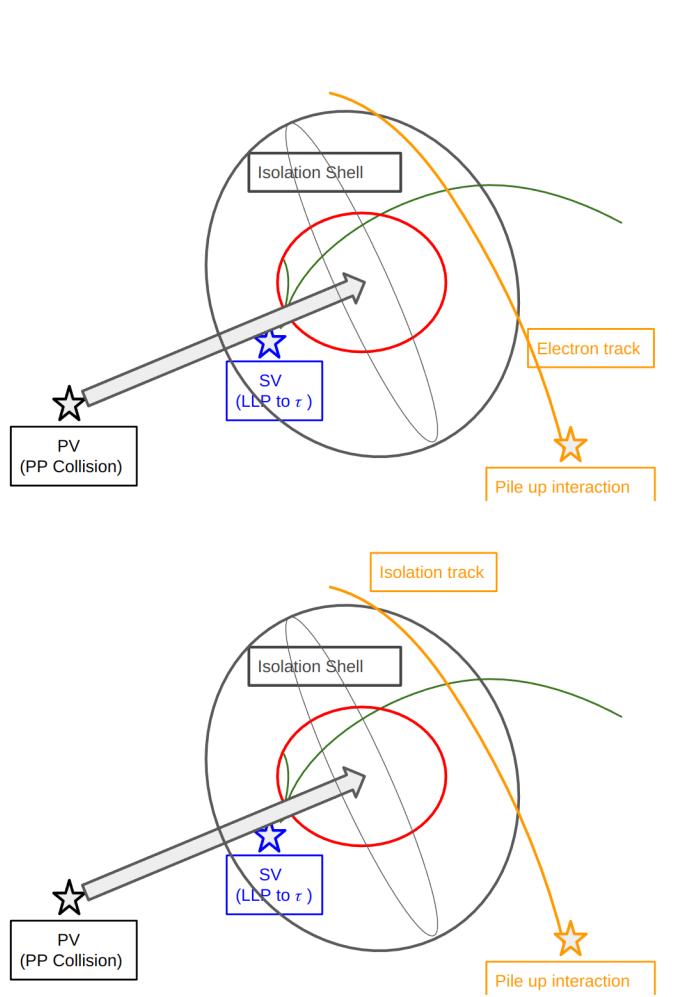


Figure 5.7: The figures display cartoons of the step-by-step inclusion of isolation track information. The first figure shows an electron or charged hadron track (orange color) originating from a Pile-up interaction. One can see the pileup track goes through the isolation shell in the cartoon. The second figure shows that the track is outside of the ROI and inside of the isolation shell. Thus, the track becomes an isolation track.

CHAPTER 6

MACHINE LEARNING

ROIs are artificial regions created by the CMSSW mechanism, displaced vertex candidates. ROIs contain thorough information about fitted tracks, vertices, and isolation, following the formation procedure described in the previous section. The exhaustive variables saved in each ROI contain information that directly and indirectly tells whether the ROI is from the signal or background process. Given the extensive amount of variables within ROIs, it is inappropriate to use ROIs' single or a few variables as our tagging variables, like in ZH analysis [45]. It is also inefficient to implement a cut-based approach due to having so many variables from ROI. The optimization process for all these vertex, isolation, and track variables would be extremely time-consuming, inefficient, and error-prone. Therefore, the analysis exploits machine learning (ML) for this multivariate analysis. Boosted decision trees will also face a similar problem as cut-based analysis. Deep Neural Networks are the most adaptable ML algorithm for this analysis method, in which the algorithm inputs an extensive list of ROI variables into a neural network and receives a final score to discriminate ROIs that arise in signal process versus background processes. The list of variables inputted into the ML algorithms is provided in subsection 6.1.

The analysis uses Keras-Tensorflow for its ML platform. CMSSW includes Keras-Tensorflow, which enables simple and easy usage of Keras-Tensorflow with simple commands in various CMS remote clusters. For CMSSW_10_6_12 (to have access to the B Parking trigger path), the Keras-Tensorflow version is 2.3.1. It runs with CPUs and GPUs.

6.1 Machine Learning Input Variables

As discussed in chapter 5, the ROI has vertex and shell (isolation) information. Vertex information is trained with 27 different variables, which are listed and categorized in Tables 6.1.

Shell information is trained with 9 different variables, which are listed in Tables 6.2.

We input vertex and shell information into separate ML networks. Thus, we get two final products from the ML networks, one for the vertex information and the other for the shell information.

Table 6.1: ROI (trackCluster) variables by category

TrackCluster	Position	TrackClusters.vx() - primaryVertex.X()
	Position	TrackClusters.vy() - primaryVertex.Y()
	Position	TrackClusters.vz() - primaryVertex.Z()
	Covariance	TrackClusters.vertexCovariance()(0,0)
	Covariance	TrackClusters.vertexCovariance()(0,1)
	Covariance	TrackClusters.vertexCovariance()(0,2)
	Covariance	TrackClusters.vertexCovariance()(1,0)
	Covariance	TrackClusters.vertexCovariance()(1,1)
	Covariance	TrackClusters.vertexCovariance()(1,2)
	Track0,1	Track0,1.pt
	Track0,1	Track0,1.eta
	Track0,1	Track0,1.phi
	Track0,1	Track0,1.dxy
	Track0,1	Track0,1.dz
	Track0,1	Track0,1.dxyError
	Track0,1	Track0,1.dzError
	Track0,1	Track0,1.normalizedChi2
	Track0,1	Track0,1.HighPurityInt

6.2 Fine-Tuning of ML environemnt

CMSSW includes an image container for Keras-Tensorflow, and researchers can also submit remote batch jobs for its ML training. The analysis tested multiple variables (such as epoch numbers, batch sizes, phi sizes, and f sizes) of our DNN layers thanks to the submission of remote batch jobs with CMSSW’s Keras-Tensorflow image container. The analysis tested different signal vs. background dataset combinations for maximum discriminant power. The list of mass scale and lifetime tested for signal points are listed in the appendix B. Different SM physics process (and their compositions) tested for background process are also listed in the appendix B.

The final Tensorflow product used for the analysis was trained with parameters in table 6.3. Keras-Tensorflow DNN variable information was fine-tuned with these variables for the maximum Area Under the Curve (AUC) value. Its information is listed in the table 6.3.

6.3 ML scores of the ROIs

More than 10 ROIs per event often exist from our ROI reconstruction mechanism. However, we do not need all ROIs per event, as many of them are low-scoring ROIs that come from track

Table 6.2: ROI (Annulus) variables by category

Annulus	pfCandidate/LostTracks pfCandidate/LostTracks pfCandidate/LostTracks pfCandidate/LostTracks pfCandidate/LostTracks pfCandidate/LostTracks pfCandidate/LostTracks pfCandidate/LostTracks pfCandidate/LostTracks pfCandidate/LostTracks	pfCandidate/LostTracks.pt pfCandidate/LostTracks.eta pfCandidate/LostTracks.phi pfCandidate/LostTracks.dxy pfCandidate/LostTracks.dz pfCandidate/LostTracks.dxyError pfCandidate/LostTracks.dzError pfCandidate/LostTracks.normalizedChi2 pfCandidate/LostTracks.HighPurityInt
---------	--	--

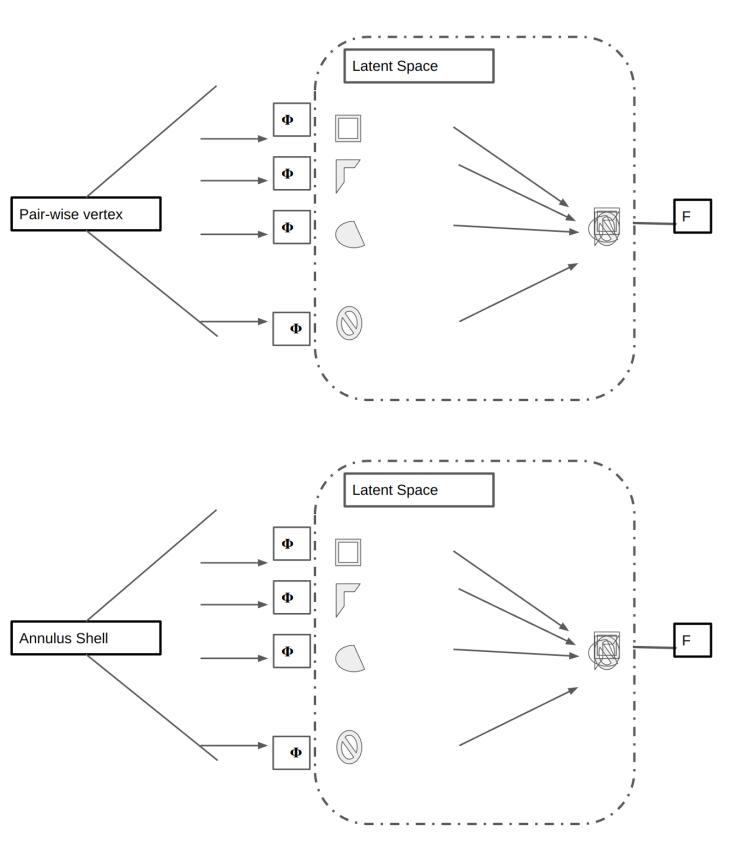


Figure 6.1: Simple cartoon showing ML network for vertex and annulus shell.

misconstruction and others. The signal process has two scalar decays as in Figure A.1, and it is reasonable to use only the two highest-scoring ROIs for our analysis. Thus, we define the leading ROI (leadROI) as an ROI with the highest vertex score in the event. Likewise, it would be easy to define the subleading ROI (subleadROI) as an ROI with the second highest vertex score in the

Table 6.3: Parameters for ML network

Epoch	300
batch size	250
Phi sizes	(64,128,256) for vertex ,(32,64,128) for shell
f sizes	(256,128,32)
Signal	ggHSSTo4Tau-MS15GeV-c τ 100mm
Background	QCD_Pt120-170_MuEneriched and TTJets

event. However, the subleadROI is not simply the second highest-scoring ROI of the event due to the non-negligible lifetime of τ leptons in the detector. After obtaining an ROI with the highest ML vertex score, we search for the next highest scored ROI, $\Delta\Phi > 0.4$, from the leading ROI to avoid counting ROI from the same LLP decay. The ordering is done by vertex discriminant since it is more potent than the shell discriminant. For better numerical representation, we use log base 10 value of ROI vertex score subtracted from 1. Mathematical definition of loglead and logexclead are in formulae 6.1 and 6.2.

$$\text{loglead} = \text{Log10}(1 - \text{ROI}'s \text{ leading Vertex score}) \quad (6.1)$$

$$\text{logexclead} = \text{Log10}(1 - \text{ROI}'s \text{ subleading Vertex score}), \text{ with } \Delta\Phi(\text{lead}, \text{sublead}) \geq 0.4. \quad (6.2)$$

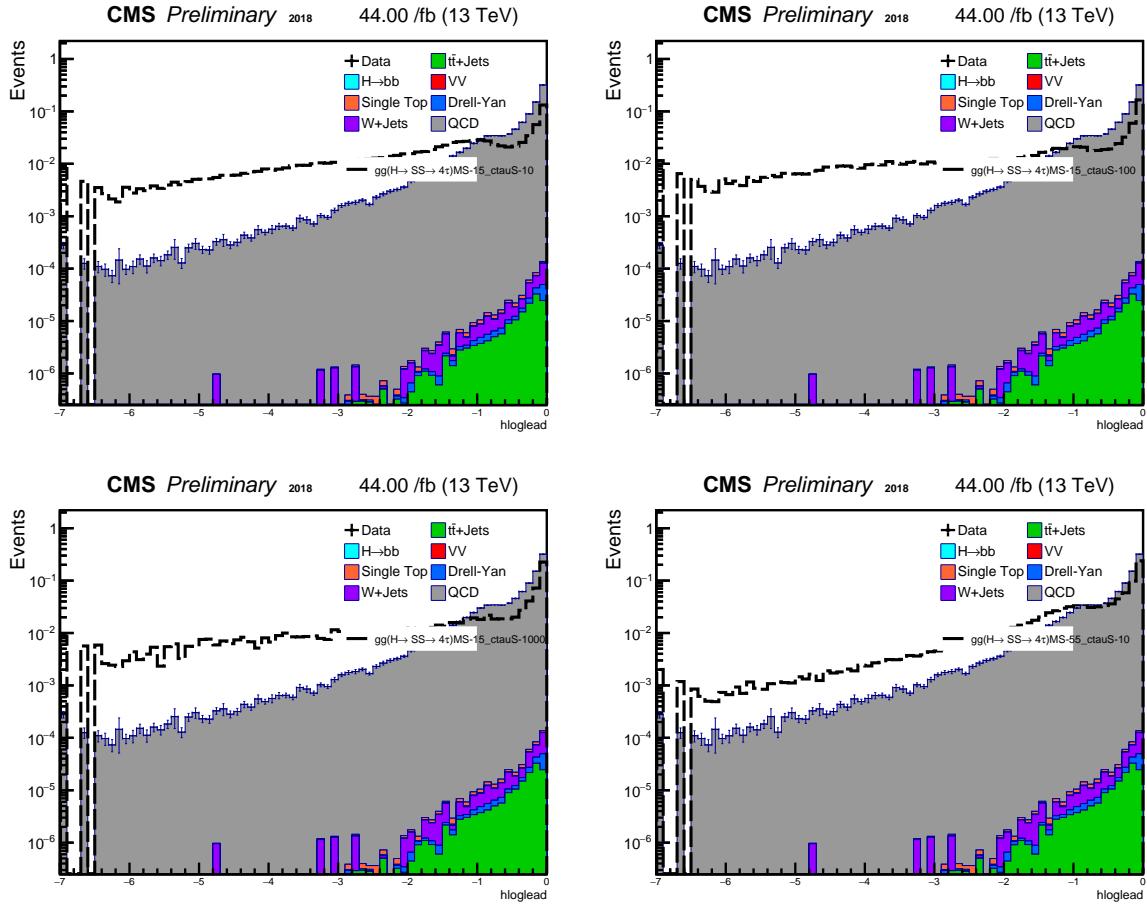


Figure 6.2: Signal versus Background for loglead, where the ROI score is the highest ROI of the event. Left plot is for MS-15_ctauS-10mm point, whereas the right plot is for MS-15_ctauS-100mm point

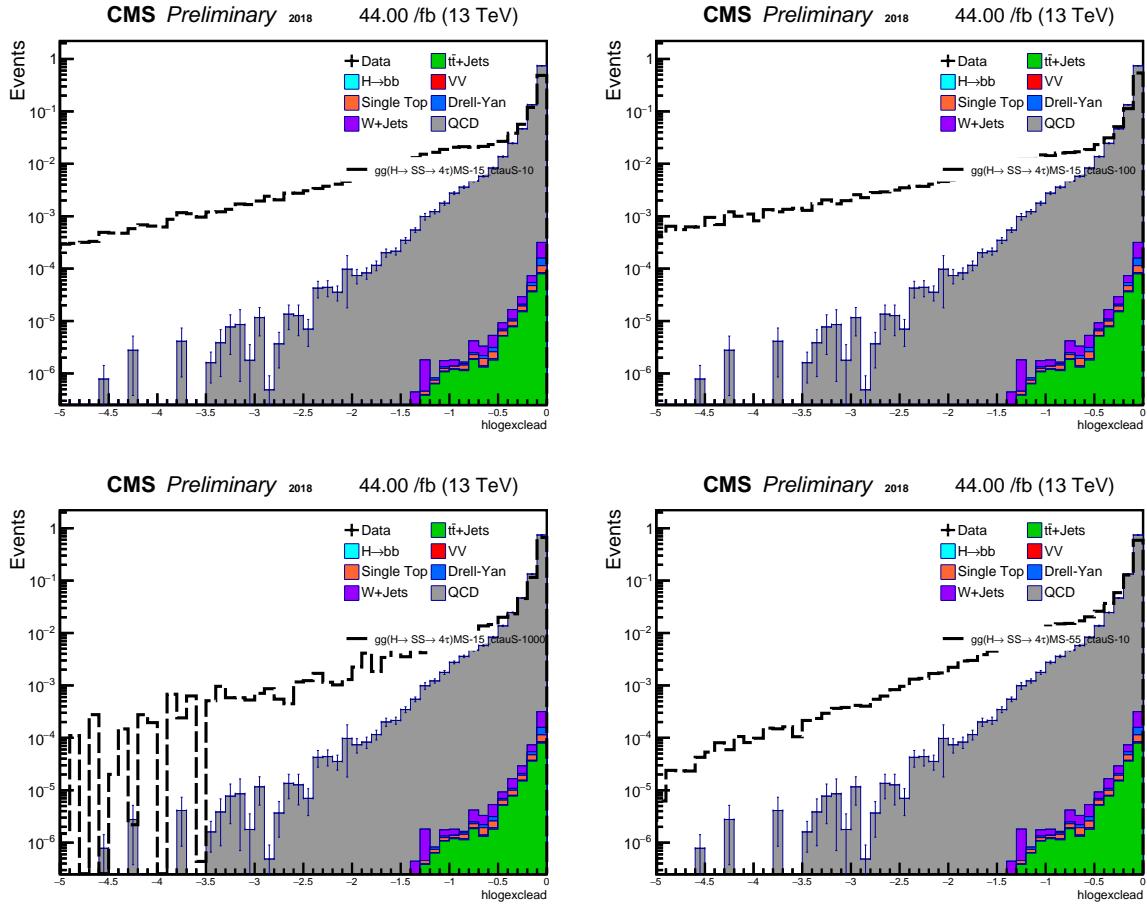
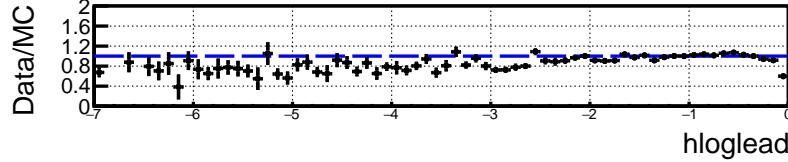
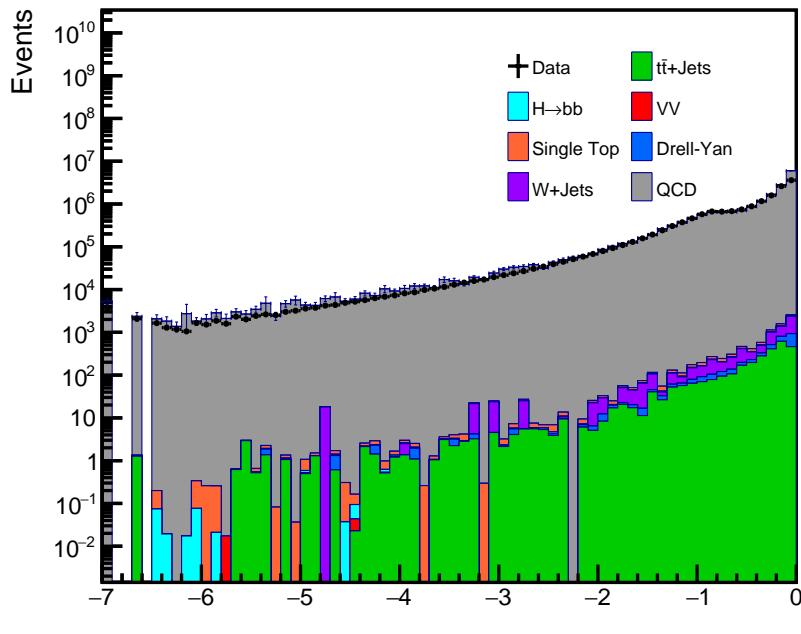


Figure 6.3: Signal versus Background for logexlead (sublead), where the ROI score is the second highest ROI (outside of $d\Phi=0.4$ from leading ROI) of the event. Left plot is for MS-15_ctauS-10mm point, whereas the right plot is for MS-15_ctauS-100mm point

CMS Preliminary 2018 Era A 5.19 /fb (13 TeV)



CMS Preliminary 2018 Era A 5.19 /fb (13 TeV)

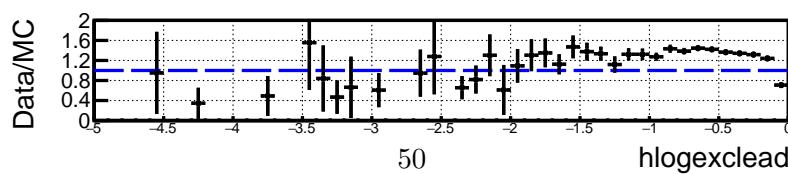
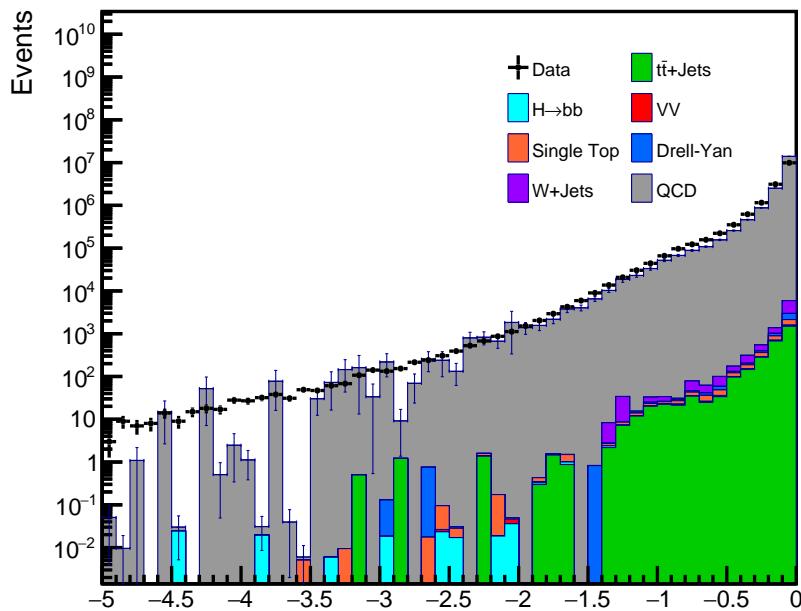


Figure 6.4: Data/MC agreement for loglead and logexlead scores

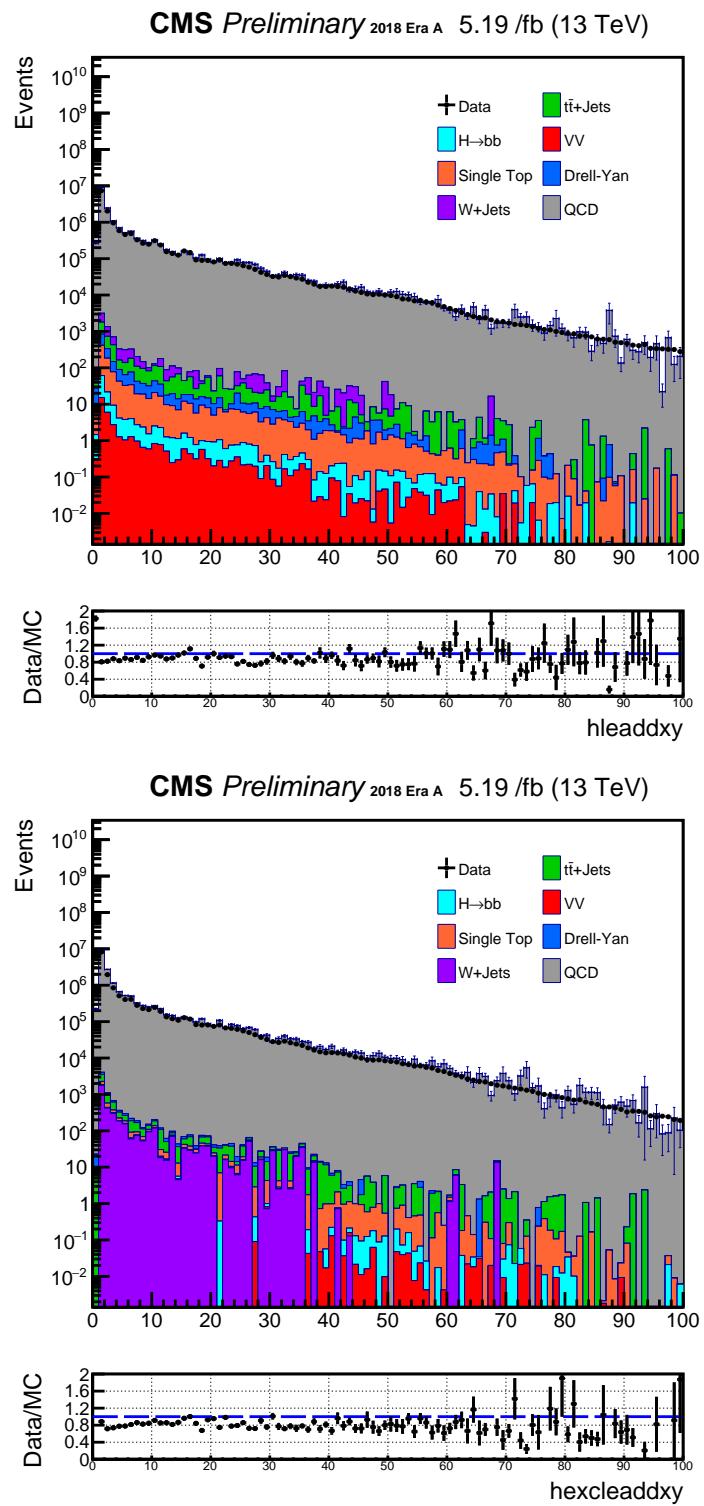


Figure 6.5: Data/MC agreement for leading and subleading ROI distributions

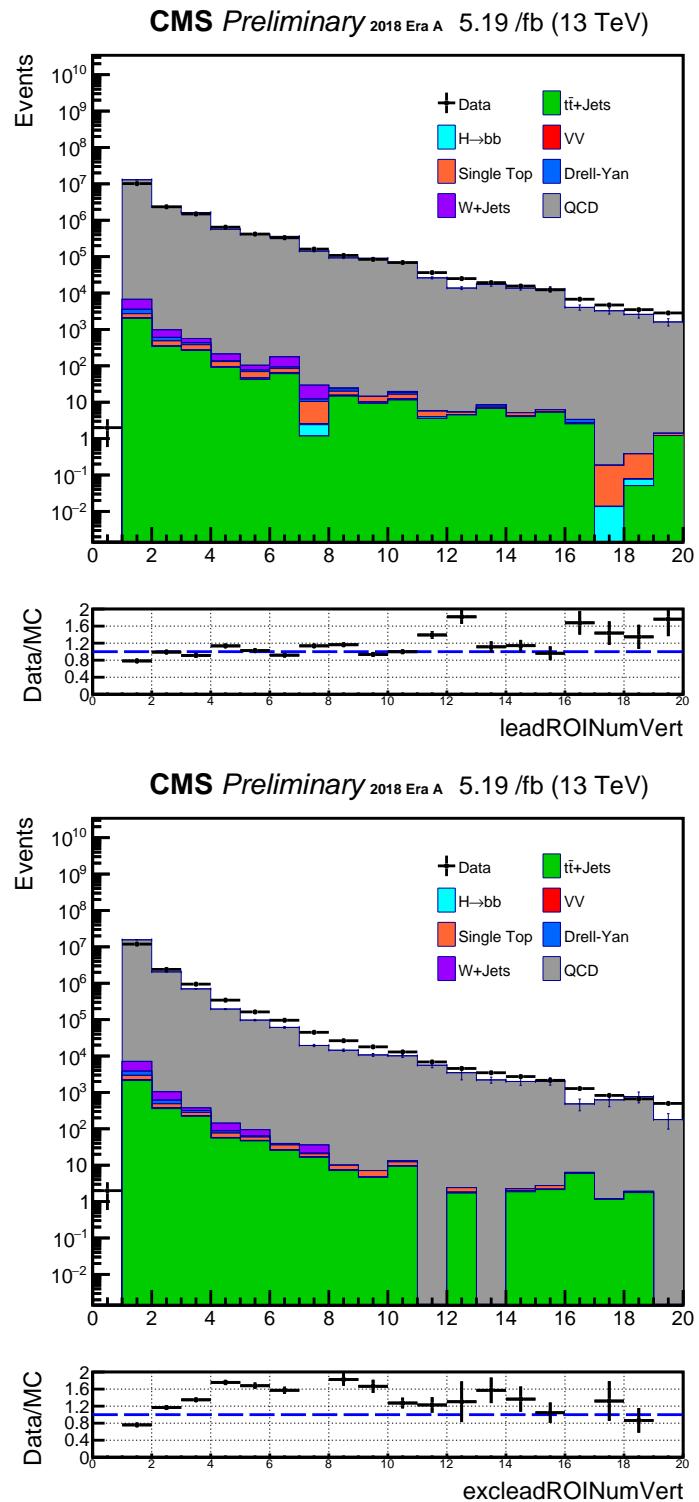


Figure 6.6: Data/MC agreement for leading and subleading ROI distributions

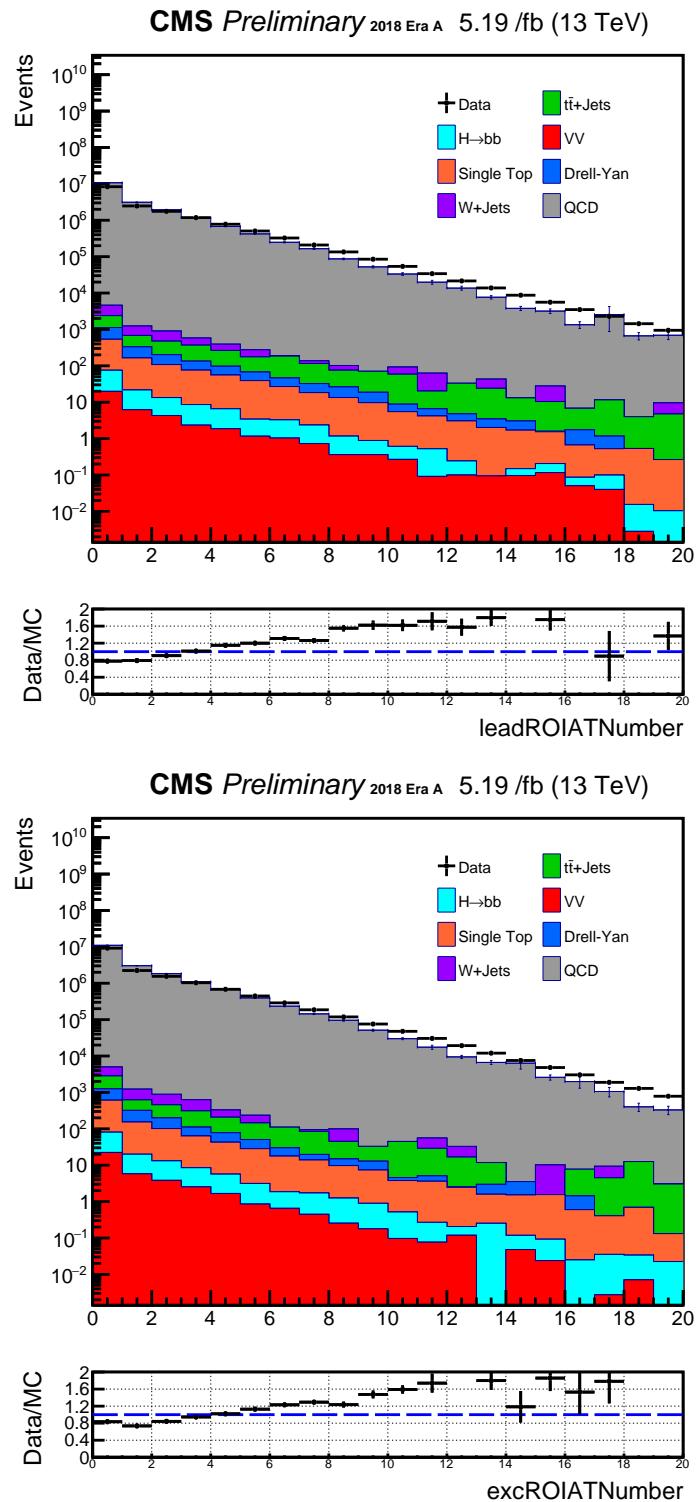


Figure 6.7: Data/MC agreement for leading and subleading ROI distributions

CHAPTER 7

EVENT SELECTION, SIGNAL AND CONTROL REGIONS

We have introduced the loglead and logexclead variables, which were both highly discriminant ROI scores(Score ordering is based upon the vertex score). Although we might have successful event selection with just triggers and ROI vertex scores, we could strengthen the discriminant power further if we consider objects other than the trigger and ROIs(tracks). These variables could be

- Non-relevant to ROIs (muon, jet object)
- Missing input variable in the ROI training

These are not ML training inputs; something ML cannot learn. So they are not discriminated against with scoring of the ROIs. The analysis implements other event selection categories itemized below for this phenomenological situation.

- $\Delta\Phi(\text{leadROI}, \text{subleadROI})$
- $\Delta R(\text{leadROI}, \text{Jet}) < 0.6$
- 1 Isolated μ
- Leading μ 's transverse impact parameter significance to PV

Each of the cuts for the items above is motivated and explained in following subsections.

7.1 Isolation criteria for muons

As discussed in chapter 4, the main background for our analysis comes from QCD process, especially from the B-meson decay of QCD events with high ROI scores. Leptonic decay of B-meson generates muons, which trigger the B-parking trigger of the analysis. Muons of B-meson decay have inferior isolation quality. They come with other hadron tracks near themselves in B-meson decay process. In contrast, muons of τ lepton decay have better isolation quality. As shown in the figure ??, τ lepton has no hadron decay associated with muon when it decays into a muon. In order to eliminate the dominant B-meson background from the QCD process, the analysis applies a PFISOLoose in selecting muon objects. The precise definition of PFISOLoose is defined below.

Table 7.1: Pt binned Isocut

MC sample	events with BPH+1 μ (noISO)	events with BPH+1 μ (ISO)
15to20	1	0.545
20to30	1	0.407
30to50	1	0.295
50to80	1	0.144
80to120	1	0.078
120to170	1	0.049
170to300	1	0.032
ggH4 τ -MS7	1	0.619
ggH4 τ -MS15	1	0.855
ggH4 τ -MS40	1	0.910

- $(\Sigma pT(ch.hadfromPV) + \max(0, \Sigma ET(neut.had) \Sigma ET(phot) - 0.5 * \Sigma pT(ch.hadfromPU))) / P_T(\mu) < 0.25$

Due to its poorer isolation quality from the boost, some muons decayed from τ lepton with a 7 GeV mass fail isolation cut. However, applying the PFISOLoose cut on muons is still beneficial since it removes more background events than signal events. The table 7.1 demonstrates event yield drop before and after requiring PFISOLoose cut on muons, classified by its signal and background process.

One can also gauge the isolation of muon object based on the generator level of ΔR of tau lepton pair as in figure ??.

7.2 Delta Phi(lead ROI,sublead ROI)

This analysis looks for displaced vertices in the tracker region, coming from the decays of exotic LLPs from Higgs produced in gluon fusion mode, leaving the SM Higgs boson without boost. The largest mass of the exotic LLPs is 55 GeV, ranging down to 7 GeV. Thus, exotic LLPs decayed from the SM Higgs become boosted, with their momentum vectors pointing back-to-back in the SM Higgs rest frame. Exotic LLPs with lighter mass are more boosted than heavier LLPs since less LLP mass means more leftover energy into kinetic energy. Given that ROIs corresponding to an exotic LLP's decay should have the highest ROI score, one should expect that the leading ROI and subleading ROI in a single event would be back-to-back. Thus, in signal events, $\Delta\Phi(\text{lead}$

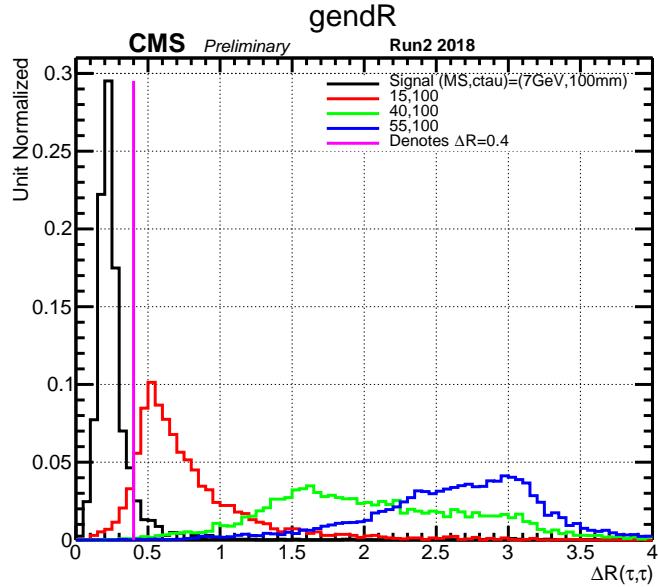


Figure 7.1: ΔR value of tau lepton pair from LLP decay is plotted. $\tau s'$ mass are low enough to be resolved when decayed from heavier LLP.

ROI,sublead ROI) tends to have high values, while the background processes tend to have a more random distribution. This analysis applies a cut above 2.2 to reduce background contribution.

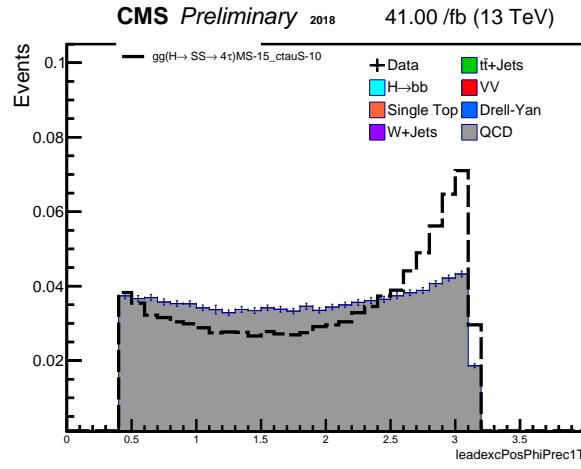


Figure 7.2: Signal versus Background for Delta Phi(leadROI, subleadROI). The plot is for MS-15_ctauS-100mm point

7.3 DeltaR(ROI, jet)

τ leptons of the signal process can also decay hadronically, while only one of the τ leptons decay into a muon to trigger the B parking trigger. When τ leptons decay, its decay shower can get clustered in the calorimeter and reconstructed as a jet. Given that the τ lepton's on-shell mass is a fixed value, τ lepton and its hadronic decay products (to be clustered into a jet) have a specific kinematic phase space. Thus, the $\Delta R(\text{ROI}, \text{jet})$ has a distribution with a peak at a value around 0.3-0.6. Meanwhile, the QCD background has a different distribution shape. Given the hadronic nature of the process, jet multiplicity is high. Higher jet multiplicity makes the $\Delta R(\text{ROI}, \text{jet})$ value somewhat randomized, resulting in a flat distribution.

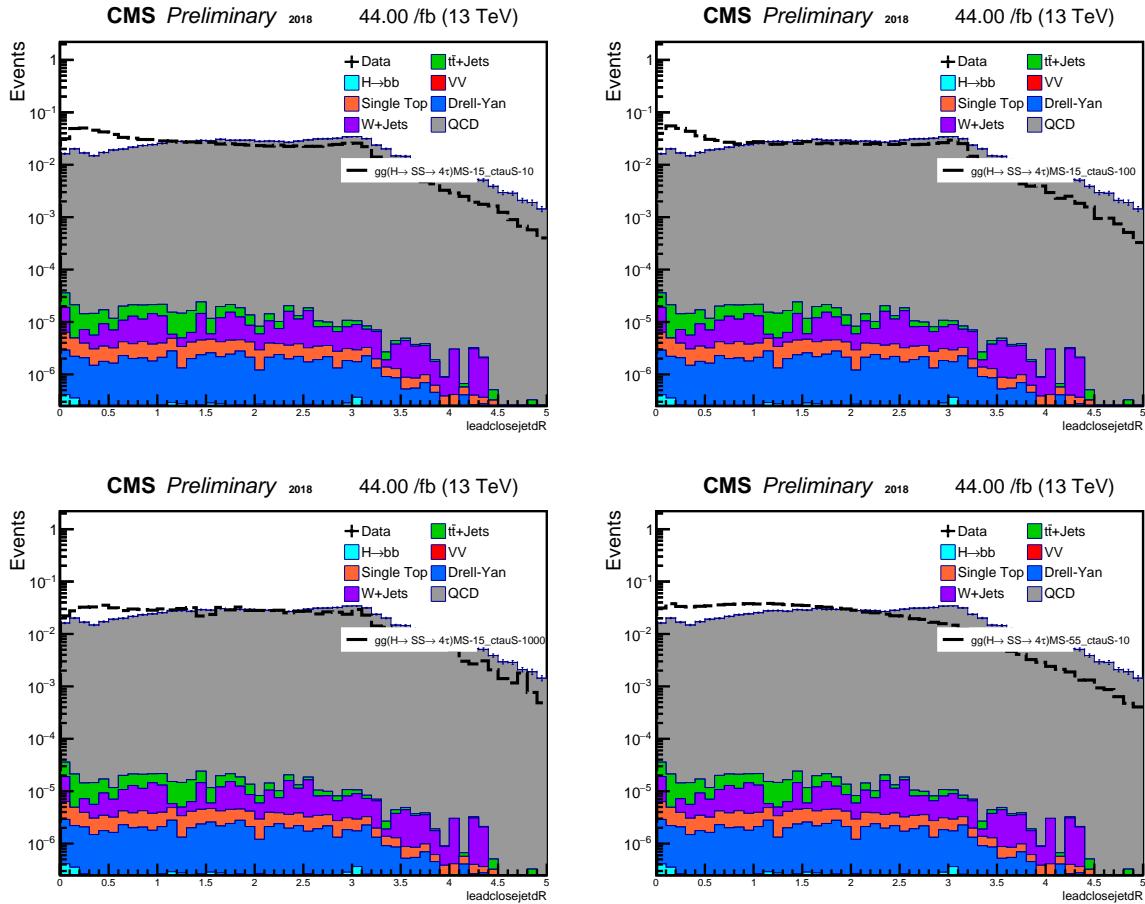


Figure 7.3: Delta R(Jet, leadingROI). Left plot is for MS-15_ctauS-10mm point, whereas the right plot is for MS-15_ctauS-100mm point

7.4 Leading muon's transverse impact parameter significance to PV

With the B-Parking trigger, triggering muons have significant transverse displacement (impact parameter) in both background and signal processes. However, displacement in the signal process is more significant than in the background process. The signal process has a minimum of $c\tau = 1\text{mm}$, which is longer than the B-meson lifetime. Thus, triggering muon object's transverse impact parameter to PV (muIPSig) is larger in the signal process than in the background process. In order to account for the error values associated with the parameter, we use impact parameter significance, defined as the ratio of the impact parameter divided by its error. In order to adjust the orders of the magnitude observed for the impact parameter significance (muIPSig) variable, we exploit the log value of the IPSig. The analysis implements a cut on $\text{Log10}(\text{muIPSig})$.

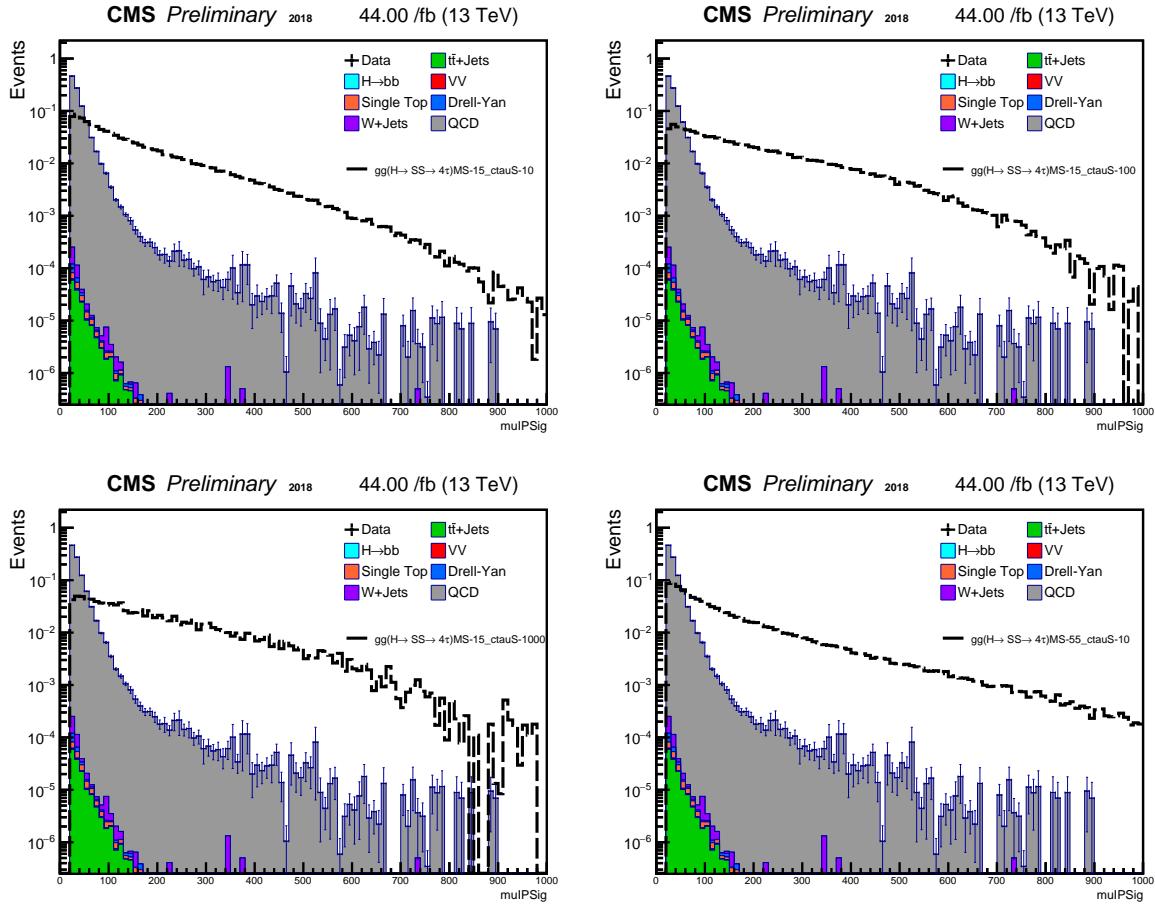


Figure 7.4: leading muon's transverse impact parameter value to the primary vertex. Left plot is for MS-15_ctauS-10mm point, whereas the right plot is for MS-15_ctauS-100mm point

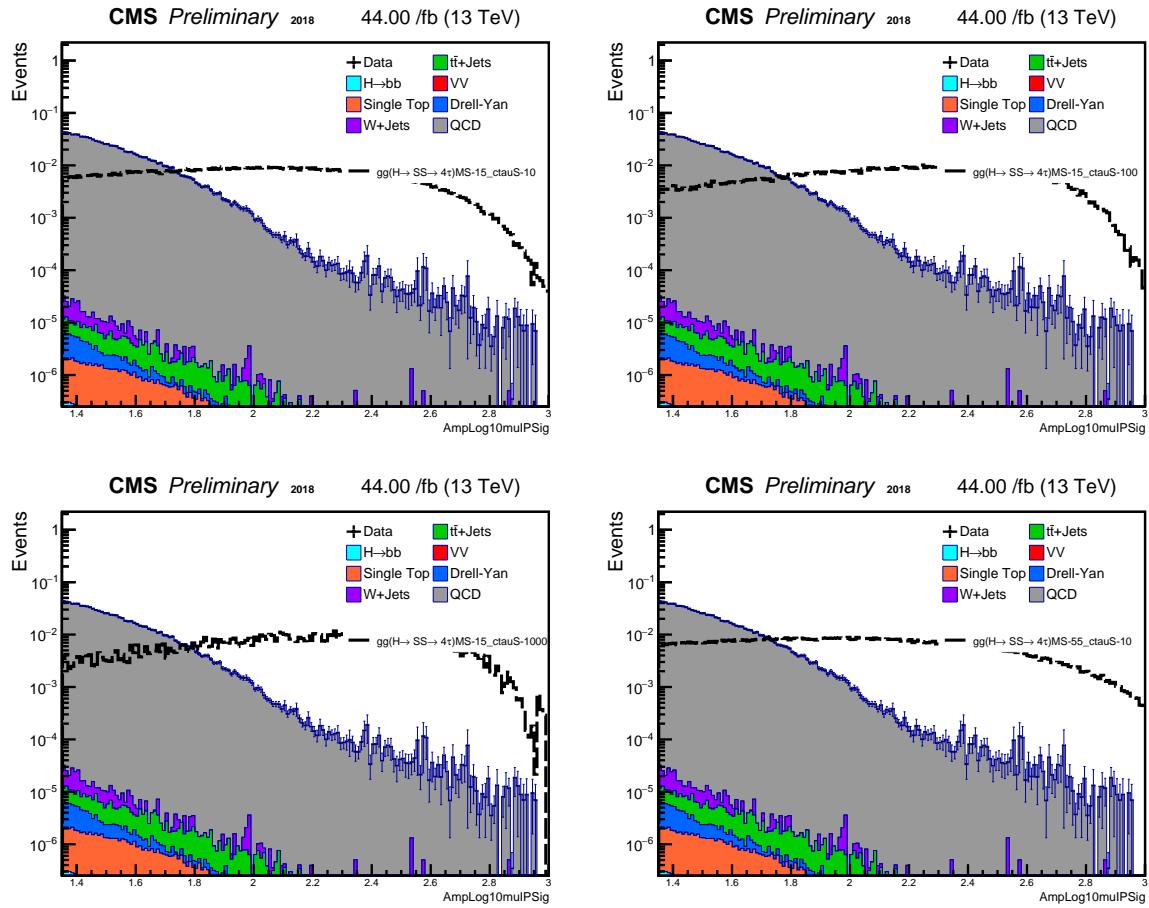


Figure 7.5: Amplified histograms of Log10muIPSig

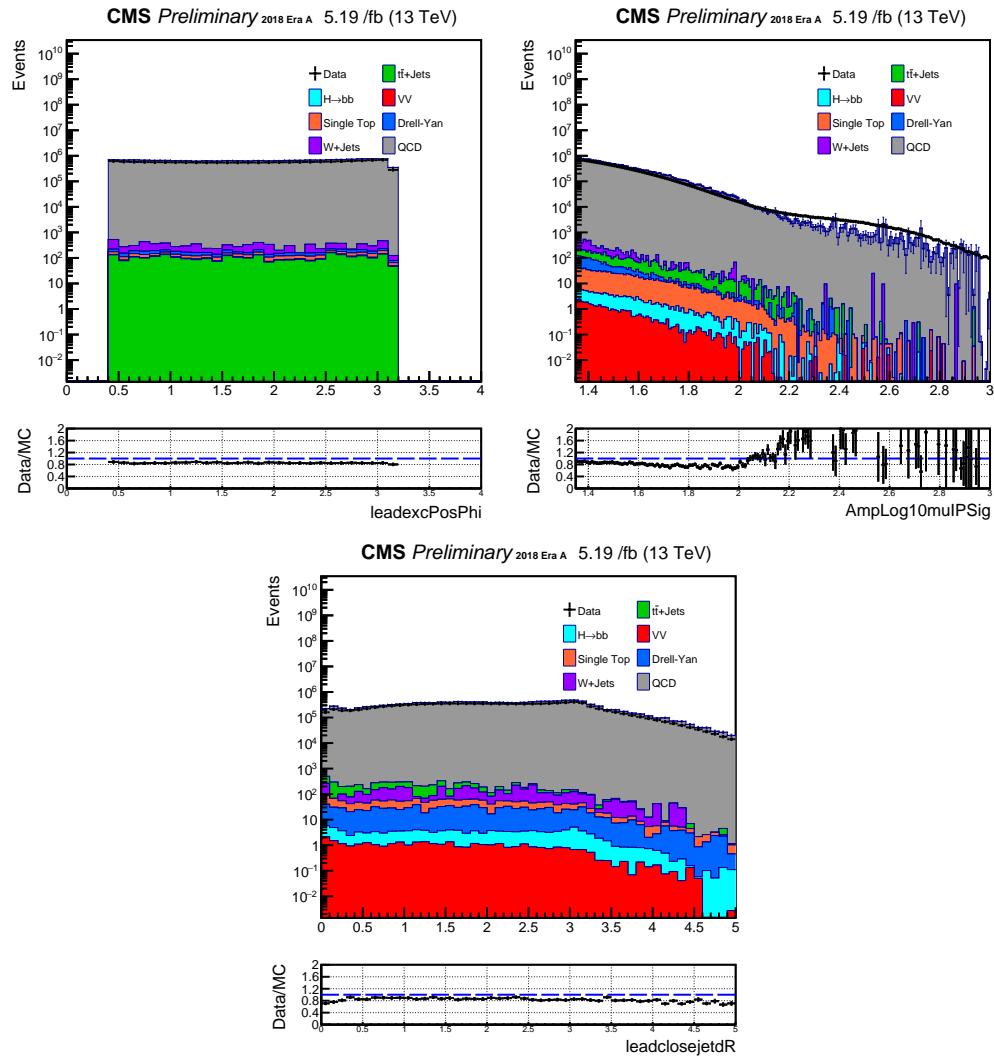


Figure 7.6: Data/MC agreement for distributions

CHAPTER 8

BACKGROUND ESTIMATION

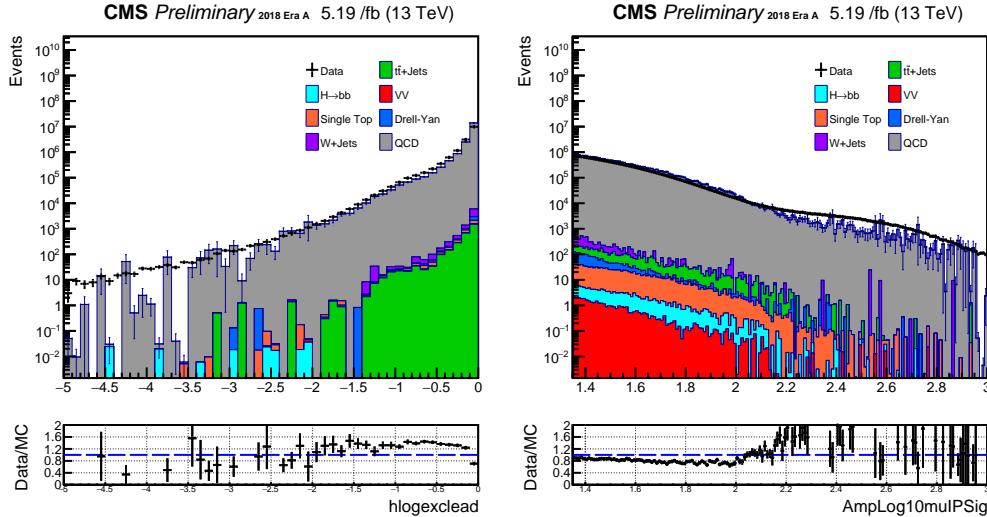


Figure 8.1: Data/MC agreement for distributions. High ROI scores and high $\text{Log10}(\mu\text{IPSig})$ values lack MC statistics

Section 7 listed several different discriminant variables for the analysis. Leading ROI vertex score, subleading ROI vertex score, and leading muon's IP value have the most discriminatory power. Although the background MC samples describe the data's distribution reasonably well, one can also see discrepancies at higher vertex scores and IPSig values. The discrepancy originates from 2 different reasons. First, the lack of statistics is a fundamental deficiency for the MC samples. Unlike 11.9 billion b-parking data events, MC sample events are simulated with extremely intensive computing resources. Thus, its event number is below 100 million at most, failing to describe distributions for extreme values. Second, the QCD MC's data description accuracy is lacking compared to other physics processes. QCD background, the main background for the analysis, is a quantum-chromodynamics process. QCD involves much uncertainty and relies on probabilistic description at low energy. Therefore, QCD's description accuracy is lacking compared to other processes, such as WJets and the Drell-Yan process. Since MC does not adequately describe the data, the analysis uses data-driven background estimation method.

8.1 ABCD method

ABCD method is the most preferred option for any background estimation method thanks to its simplicity. ABCD method is the simplest form of transfer factor. A ratio of background events at two control regions (CR) is applied to another CR's background event to infer background events in signal regions (SR) without unblinding. The simplest mathematical for is described in formula 8.1.

$$SR : CR1 = CR2 : CR3, SR = CR1 * \frac{CR2}{CR3} \quad (8.1)$$

ABCD method requires two variables used for its estimation to be independent of each other for the background process.

Unfortunately, leading ROI and subleading ROI vertex scores are correlated in the QCD background process. ROIs from B-mesons score higher in our TensorFlow. In QCD processes, the B-mesons are likely to be pair produced. Therefore, when the leading ROI vertex score is high due to a B-meson, the subleading ROI vertex score is also high because the other meson is produced at the same time. Thus, due to their correlation, leading and subleading scores cannot be our ABCD discriminant variable candidates.

The analysis selects leading ROI and leading muon's IP value as its ABCD discriminant variables. After implementing all other cuts (sublead ROI, $\Delta\Phi(lead, sublead)$, Annulus score, 1 Isolated μ , $\Delta R(leadROI, Jet)$), we tested the correlation factor between the leading ROI, and leading mu's IPSig values for each background process. The values are pretty minimal except for 2-3 processes where there were too few entries to derive a solid conclusion due to statistical limitations.

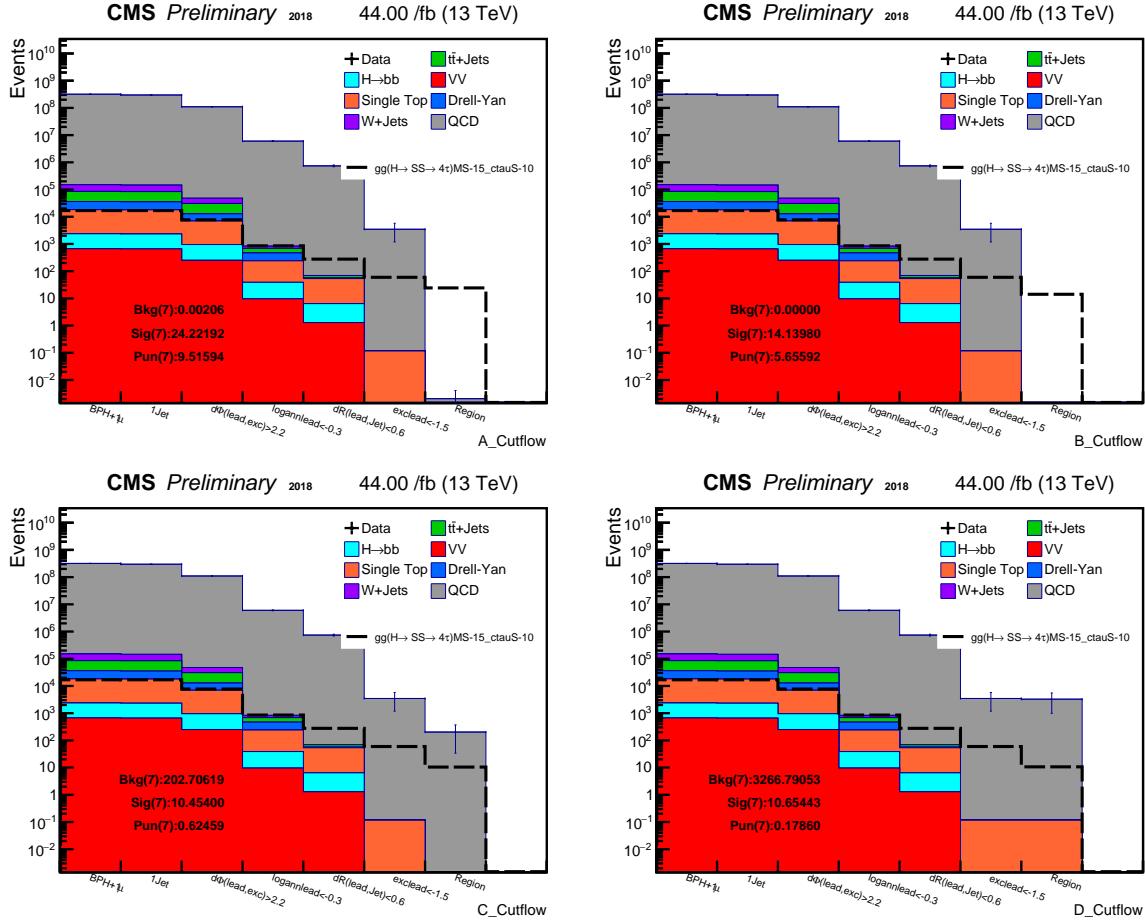


Figure 8.2: Cutflow histogram of Signal versus background. The signal is MS15 GeV-ct10mm point. Last bin of the cutflow histogram is ABCD region selection. Top left plot is for region A, whereas the bottom right plot is for region D

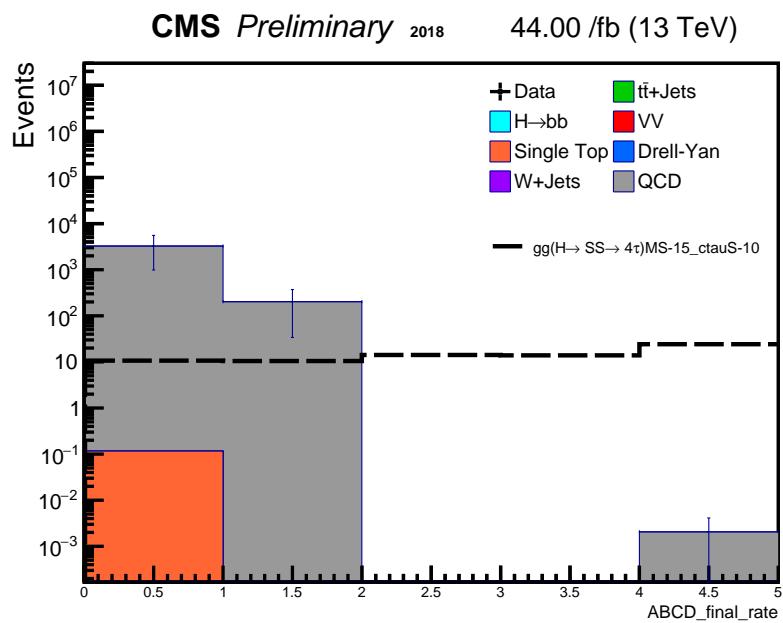


Figure 8.3: Final bin of cutflow

Table 8.1: Region A's cutflow histogram by bin entry (weighted)

	BPH+1mu	1Jet	dPhi(lead,sub)	logannlead	dR(lead,Jet)	sublead	Region
MS15 GeV ct 10mm	1.6854e+04	1.6841e+04	7.7175e+03	8.5419e+02	2.7683e+02	5.9470e+01	2.4222e+01
TotBkgHist	3.1907e+08	2.9827e+08	1.1018e+08	6.0251e+06	7.4162e+05	3.4695e+03	2.0616e-03
QCD-Pt-15to20	4.5618e+07	4.0658e+07	1.3980e+07	1.1258e+06	6.0823e+04	0.0000e+00	0.0000e+00
QCD-Pt-20to30	1.1366e+08	1.0529e+08	3.8168e+07	2.1684e+06	2.2104e+05	0.0000e+00	0.0000e+00
QCD-Pt-30to50	1.1632e+08	1.1025e+08	4.1278e+07	2.0473e+06	3.3234e+05	2.0504e+03	0.0000e+00
QCD-Pt-50to80	3.4997e+07	3.3757e+07	1.3312e+07	5.3557e+05	9.7238e+04	1.1086e+03	0.0000e+00
QCD-Pt-80to120	6.9466e+06	6.8067e+06	2.8163e+06	1.1089e+05	2.1125e+04	2.8541e+02	0.0000e+00
QCD-Pt-120to170	9.8312e+05	9.7707e+05	4.1147e+05	2.9160e+04	7.0735e+03	6.3799e+00	0.0000e+00
QCD-Pt-170to300	3.6131e+05	3.5851e+05	1.5924e+05	6.6890e+03	1.7730e+03	1.6482e+01	0.0000e+00
QCD-Pt-300to470	2.2089e+04	2.2052e+04	1.0082e+04	3.9742e+02	1.2324e+02	1.8960e+00	0.0000e+00
QCD-Pt-470to600	1.5958e+03	1.5927e+03	7.4614e+02	2.7121e+01	9.9211e+00	1.2208e-01	0.0000e+00
QCD-Pt-600to800	5.8035e+02	5.7973e+02	2.7517e+02	8.3416e+00	3.7266e+00	6.3688e-03	0.0000e+00
QCD-Pt-800to1000	9.3006e+01	9.2944e+01	4.4374e+01	1.5118e+00	7.9019e-01	2.3820e-03	2.0616e-03
QCD-Pt-1000toInf	2.5854e+01	2.5830e+01	1.2238e+01	3.5332e-01	1.7023e-01	7.8399e-03	0.0000e+00
WJetsToLNu	6.5196e+04	6.0905e+04	1.6914e+04	1.4396e+02	0.0000e+00	0.0000e+00	0.0000e+00
DYJetsToLL-M-50	1.8101e+04	1.7297e+04	5.8069e+03	2.2737e+02	3.0498e+00	0.0000e+00	0.0000e+00
TTJets	4.8794e+04	4.8736e+04	1.8003e+04	2.1718e+02	1.2788e+01	0.0000e+00	0.0000e+00
ST-t-channel-top-5f	4.3424e+03	4.2675e+03	1.6261e+03	6.4960e+01	2.2447e+01	0.0000e+00	0.0000e+00
ST-s-channel-4f-leptonDecays	1.0063e+04	9.8535e+03	3.8657e+03	1.1469e+02	1.8001e+01	0.0000e+00	0.0000e+00
ST-tW-top	1.4277e+03	1.4234e+03	5.3647e+02	2.1466e+01	6.6034e+00	1.1637e-01	0.0000e+00
WW	2.5191e+02	2.4836e+02	9.0591e+01	5.6034e+00	8.4678e-01	0.0000e+00	0.0000e+00
WZ	2.4202e+02	2.3858e+02	9.2235e+01	2.4949e+00	6.0096e-02	0.0000e+00	0.0000e+00
ZZ	1.7331e+02	1.6886e+02	6.7381e+01	1.4956e+00	3.8633e-01	0.0000e+00	0.0000e+00
GluGluHToBB	1.6967e+03	1.6652e+03	7.0197e+02	2.9198e+01	5.1113e+00	0.0000e+00	0.0000e+00
VVhist	6.6723e+02	6.5581e+02	2.5021e+02	9.5939e+00	1.2932e+00	0.0000e+00	0.0000e+00
ggHbbhist	1.6967e+03	1.6652e+03	7.0197e+02	2.9198e+01	5.1113e+00	0.0000e+00	0.0000e+00
SThist	1.5833e+04	1.5544e+04	6.0283e+03	2.0111e+02	4.7051e+01	1.1637e-01	0.0000e+00
DYhist	1.8101e+04	1.7297e+04	5.8069e+03	2.2737e+02	3.0498e+00	0.0000e+00	0.0000e+00
TThist	4.8794e+04	4.8736e+04	1.8003e+04	2.1718e+02	1.2788e+01	0.0000e+00	0.0000e+00
WJhist	6.5196e+04	6.0905e+04	1.6914e+04	1.4396e+02	0.0000e+00	0.0000e+00	0.0000e+00
QCДhist	3.1892e+08	2.9812e+08	1.1014e+08	6.0242e+06	7.4155e+05	3.4694e+03	2.0616e-03

Table 8.2: Region A's cutflow histogram by bin event number

	BPH+1mu	1Jet	dPhi(lead,sub)	logannlead	dR(lead,Jet)	sublead	Region
SigHist	2.2694e+05	2.2681e+05	1.0352e+05	1.2648e+04	4.1410e+03	8.5500e+02	3.4200e+02
TotHist	5.3177e+05	5.2132e+05	2.1245e+05	1.0073e+04	2.3160e+03	2.3000e+01	1.0000e+00
QCD-Pt-15to20	2.7520e+03	2.5260e+03	8.6400e+02	7.0000e+01	4.0000e+00	0.0000e+00	0.0000e+00
QCD-Pt-20to30	4.8608e+04	4.5901e+04	1.6665e+04	1.0160e+03	1.0700e+02	0.0000e+00	0.0000e+00
QCD-Pt-30to50	9.0101e+04	8.6625e+04	3.2269e+04	1.7380e+03	2.7400e+02	1.0000e+00	0.0000e+00
QCD-Pt-50to80	6.8298e+04	6.6504e+04	2.5810e+04	1.1720e+03	2.1300e+02	3.0000e+00	0.0000e+00
QCD-Pt-80to120	6.5713e+04	6.4742e+04	2.6313e+04	1.1520e+03	2.1400e+02	3.0000e+00	0.0000e+00
QCD-Pt-120to170	2.3205e+04	2.3121e+04	9.5490e+03	8.6400e+02	2.1800e+02	1.0000e+00	0.0000e+00
QCD-Pt-170to300	6.8059e+04	6.7665e+04	2.9518e+04	1.2610e+03	3.3200e+02	2.0000e+00	0.0000e+00
QCD-Pt-300to6470	3.8598e+04	3.8550e+04	1.7291e+04	7.0500e+02	2.1900e+02	4.0000e+00	0.0000e+00
QCD-Pt-470to600	1.4636e+04	1.4614e+04	6.6820e+03	2.6400e+02	1.0000e+02	1.0000e+00	0.0000e+00
QCD-Pt-600to800	1.5791e+04	1.5777e+04	7.3730e+03	2.4500e+02	1.0200e+02	1.0000e+00	0.0000e+00
QCD-Pt-800to1000	2.8626e+04	2.8612e+04	1.3489e+04	4.8500e+02	2.4900e+02	4.0000e+00	0.0000e+00
QCD-Pt-1000toInf	8.4270e+03	8.4220e+03	3.9280e+03	1.1800e+02	5.7000e+01	2.0000e+00	0.0000e+00
WJetsToLNu	6.9900e+02	6.7400e+02	2.0000e+02	1.0000e+01	0.0000e+00	0.0000e+00	0.0000e+00
DYJetsToLL-M-50	2.3600e+03	2.2800e+03	7.8300e+02	3.7000e+01	3.0000e+00	0.0000e+00	0.0000e+00
TTJets	1.0453e+04	1.0447e+04	3.9980e+03	1.7900e+02	6.3000e+01	0.0000e+00	0.0000e+00
ST-t-channel-top-5f	5.7510e+03	5.6790e+03	2.1330e+03	9.6000e+01	2.9000e+01	0.0000e+00	0.0000e+00
ST-s-channel-4f-leptonDecays	1.6739e+04	1.6465e+04	6.5460e+03	2.7700e+02	5.4000e+01	0.0000e+00	0.0000e+00
ST-tW-top	5.8810e+03	5.8680e+03	2.1960e+03	8.9000e+01	2.9000e+01	1.0000e+00	0.0000e+00
WW	8.9600e+02	8.8600e+02	3.1900e+02	1.7000e+01	3.0000e+00	0.0000e+00	0.0000e+00
WZ	1.1810e+03	1.1670e+03	4.4900e+02	1.7000e+01	2.0000e+00	0.0000e+00	0.0000e+00
ZZ	9.2200e+02	9.0500e+02	3.5600e+02	1.0000e+01	2.0000e+00	0.0000e+00	0.0000e+00
GluGluHToBB	1.4070e+04	1.3888e+04	5.7160e+03	2.5100e+02	4.2000e+01	0.0000e+00	0.0000e+00
WJhist	6.9900e+02	6.7400e+02	2.0000e+02	1.0000e+01	0.0000e+00	0.0000e+00	0.0000e+00
DYhist	2.3600e+03	2.2800e+03	7.8300e+02	3.7000e+01	3.0000e+00	0.0000e+00	0.0000e+00
VVhist	2.9990e+03	2.9580e+03	1.1240e+03	4.4000e+01	7.0000e+00	0.0000e+00	0.0000e+00
TThist	1.0453e+04	1.0447e+04	3.9980e+03	1.7900e+02	6.3000e+01	0.0000e+00	0.0000e+00
ggHbbhist	1.4070e+04	1.3888e+04	5.7160e+03	2.5100e+02	4.2000e+01	0.0000e+00	0.0000e+00
SThist	2.8371e+04	2.8012e+04	1.0875e+04	4.6200e+02	1.1200e+02	1.0000e+00	0.0000e+00
QCDDhist	4.7281e+05	4.6306e+05	1.8975e+05	9.0900e+03	2.0890e+03	2.2000e+01	1.0000e+00

8.2 Validation of ABCD method

Although background MC's correlation factor(\hat{r}) values are small, we need to verify that the data's correlation factor is negligible since data will be used for our background estimation. However, direct computation of $\hat{r}(\text{leadROI}, \text{Log10(muIPSig)})$ for data is problematic because CMS collaboration stipulates “blinding.” In general, a high signal ratio is expected in the signal region. In order to avoid bias for discovery and to change their analysis strategy accordingly, CMS collaboration recommends physicists blind themselves from looking into data entry in the signal region until community acknowledgment is granted. This concept is called “blinding.” Since one can not access data entry in the signal region, the data's correlation factor(\hat{r}) can not be computed. Therefore, one needs to find an alternative way to estimate data's $\hat{r}(\text{leadROI}, \text{Log10(muIPSig)})$ without unblinding data in the signal region.

Lifetime	MS7	MS15	MS40	MS55
1mm	N/A	7.43	1.68	.947
10mm	343.2	833.6	767.2	427.8
100mm	141.1	538.4	1011	1155
1000mm	13.4	68.5	129.7	163.1

Table 8.3: $c\tau, MS$ vs Signal Contamination in A' (Bkg :108k)

For this purpose, we define a new ABCD region, referred to as A'B'C'D'. A'-D' region is an orthogonal plane to the original ABCD, where the $\Delta\Phi(\text{lead}, \text{exlead})$ category of the event selection is flipped while all other event selection cuts are removed. Region A', a correspondent Signal region of the orthogonal plane, should be much lower in the signal ratio (background dominated): Only so, looking into data entry in A' is not problematic. In table ??, we confirm that the background dominates the A' region while the signal is low regardless of the signal's MS and ctau.

In figure ??, we can see the correlation factor is small at about 10%, still a negligible value. With $\hat{r}(\text{leadROI}, \text{Log10(muIPSig)})$ value in A'-D' region, It seems sufficient to claim that there should be a negligible correlation between leadROI and Log10(muIPSig) for the A-D region as well. However, We could have 1 more diagnostic question before unblinding. What if $\hat{r}(\text{leadROI}, \text{Log10(muIPSig)})$ value changes with respect to $\Delta\Phi(\text{lead}, \text{exlead})$? If the formula in 8.2 is not true, $\hat{r}(\text{leadROI}, \text{Log10(muIPSig)})$ obtained from A'-D' is not a good estimate for $\hat{r}(\text{leadROI}, \text{Log10(muIPSig)})$ for A-D.

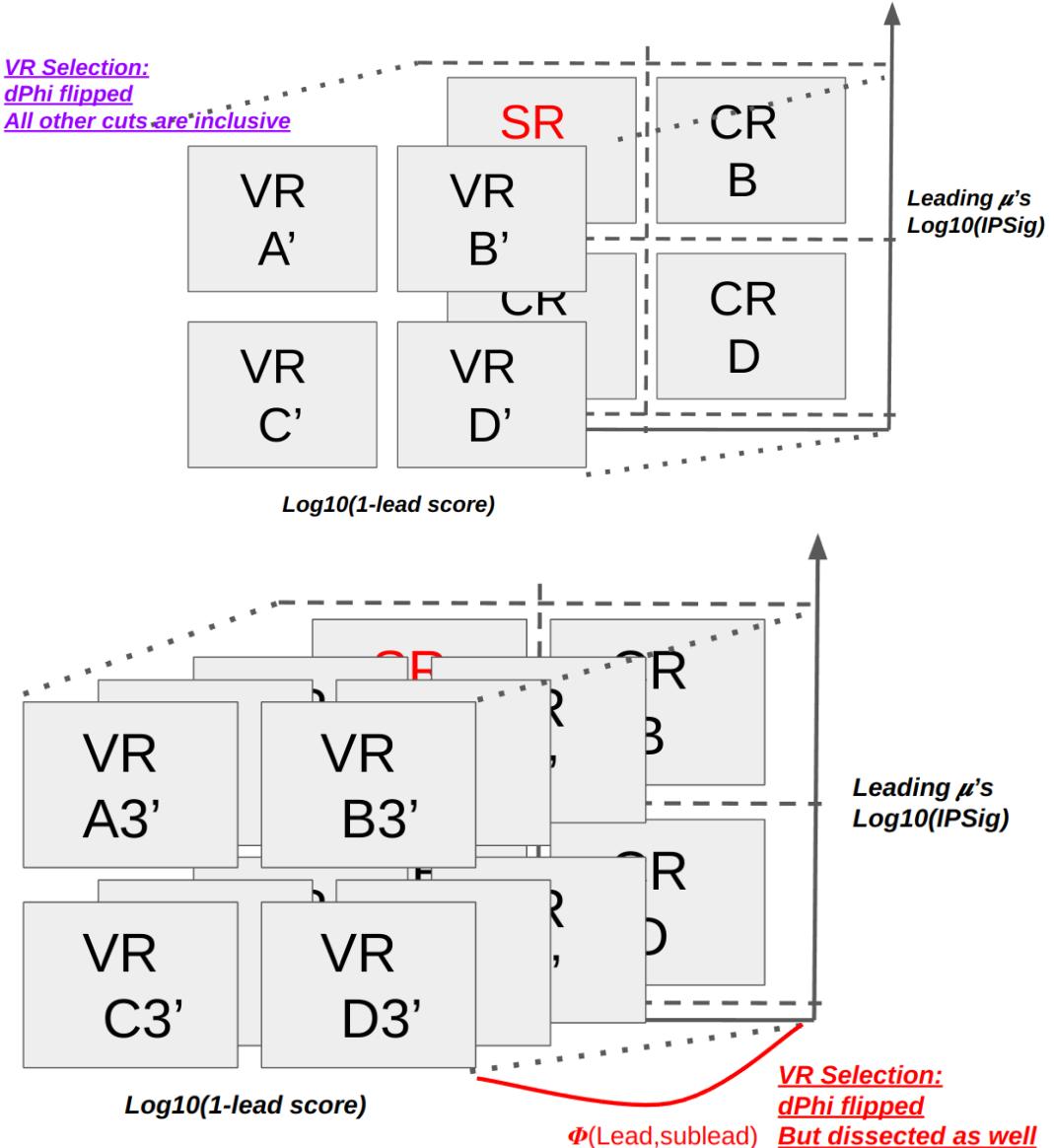


Figure 8.4: Cartoon description of ABCD method validation. The first figure shows a simpler definition of validation region. The second figure shows a more sophisticated definition of validation regions.

$$\frac{d\hat{r}(\text{leadROI}, \text{Log10(muIPSig)})}{d\Delta\Phi(\text{lead, exc})} = 0 \quad (8.2)$$

We subdivide A'-D' to check that formula 8.2 is true. A1'-D1' are orthogonal planes where $\Delta\Phi(\text{lead, exclead})$ is closest to A-D. Increasing numeric value indicates how far the $\Delta\Phi(\text{lead, exclead})$

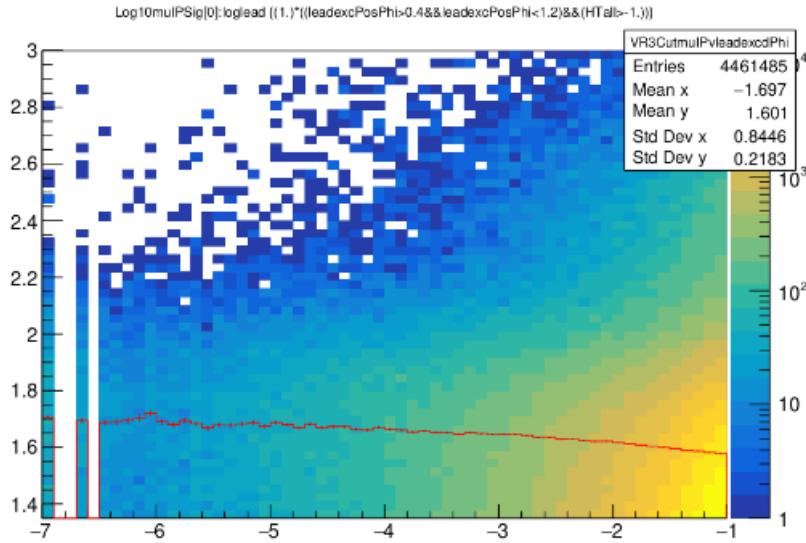


Figure 8.5: Profile plot of VR. Correlation factor is 0.105. ProfileX is rather flat. X-axis is $\log_{10}(1\text{-leading ROI score})$. Y-axis is $\log_{10}(\text{muIPSSig})$.

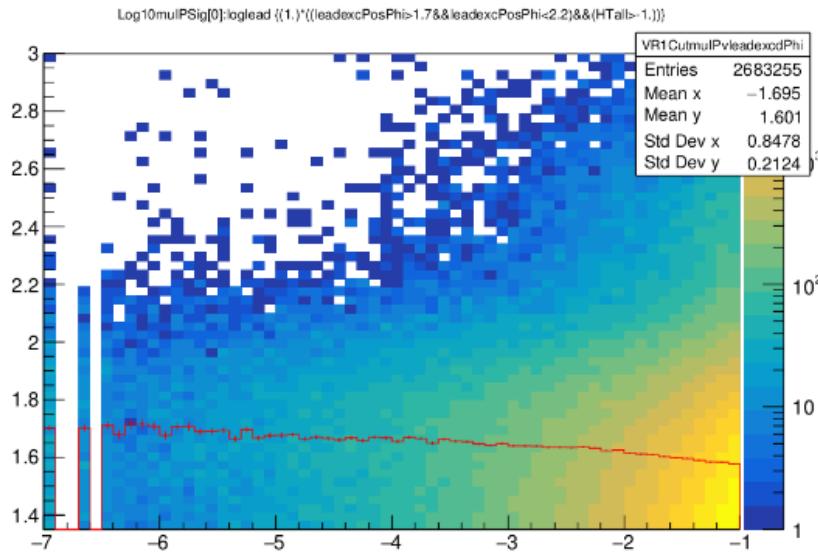


Figure 8.6: Profile plot of VR1. Correlation factor is 0.124. ProfileX is rather flat. X-axis is $\log_{10}(1\text{-leading ROI score})$. Y-axis is $\log_{10}(\text{muIPSSig})$.

is from the A-D. $\hat{r}(\text{leadROI}, \log_{10}(\text{muIPSSig}))$ value for each of A'-D' region is plotted and summarized in figures ??-?? .

As we scan $\hat{r}(\text{leadROI}, \log_{10}(\text{muIPSSig}))$ for each $\Delta\Phi(\text{lead}, \text{exclead})$ section, very small trend with respect to $\Delta\Phi(\text{lead}, \text{exclead})$ is observed. We can claim the formula 8.2 is true and

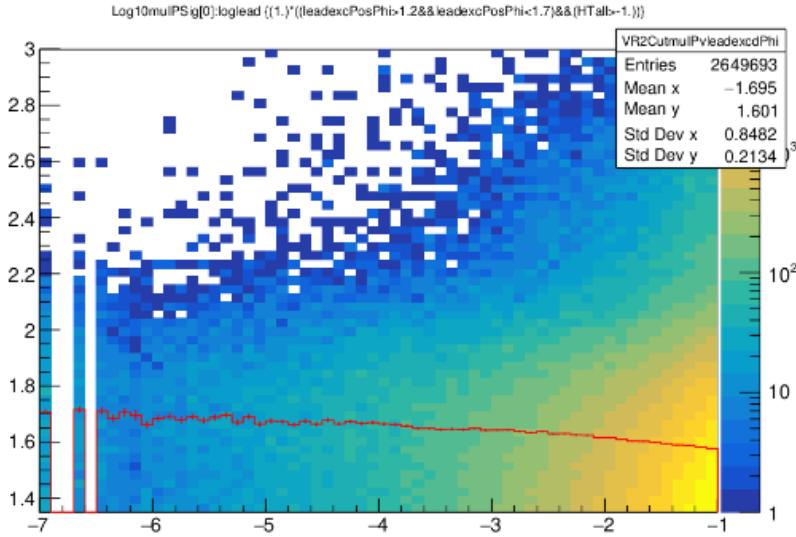


Figure 8.7: Profile plot of VR2. Correlation factor is 0.1229. ProfileX is rather flat. X-axis is $\log_{10}(1\text{-leading ROI score})$. Y-axis is $\log_{10}(\text{muIPSSig})$.

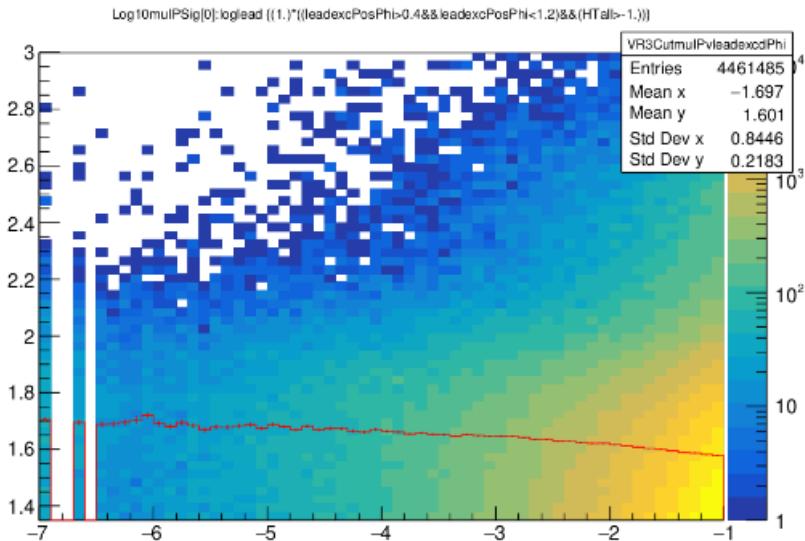


Figure 8.8: Profile plot of VR3. Correlation factor is 0.1203. ProfileX is rather flat. X-axis is $\log_{10}(1\text{-leading ROI score})$. Y-axis is $\log_{10}(\text{muIPSSig})$.

$\hat{r}(\text{leadROI}, \log_{10}(\text{muIPSSig}))$ from A'-D' can be translated into A-D region as well. Since there is little correlation between the two variables for the ABCD region, we validated the background estimation method. With the background rate in the signal region estimated by the ABCD method with data, and with signal MC, we can claim whether there exists a signal over the SM expectation

or not. If it does not, we can set an exclusion limit on the Higgs branching ratio to our final state at a specific value.

Before that final step, systematic uncertainty is described in the following section for exclusion limit calculation.

CHAPTER 9

SYSTEMATIC UNCERTAINTIES

Table 9.1 shows a summary of the systematic uncertainties included for the exclusion limit calculation,

Table 9.1: Systematic Uncertainty Table

Uncertainty	Signal (%)	Background (%)
Luminosity	2	2
Muon ID,ISO Scale Factor	2	2
ROI Score	10	-
Background Estimation Method	20	-
Statistical error	-	20

CMS records the delivered luminosity, which is accompanied by its uncertainty. This uncertainty is universal and applied to every analysis. For data collected in 2018, the luminosity uncertainty is found to be 1.8%, which is applied to both signal and background [65]. The analysis uses muon objects applied with Loose ID and ISO criteria. Lepton ID, ISO scale factors, provided by the muon POG [62], also carry uncertainty, albeit the magnitude is minimal. The ROI score uncertainty is obtained by adjusting the leading ROI's vertex score. As discussed in chapter 8, we validated the ABCD method using a validation region of A'-D'. The correlation factor of 2 variables in A'-D' was a small value of 11%. It tells us there could be an 11% or slightly larger correlation in the A-D region. A positive correlation implies more background concentrated in the SR(A) than its estimated rate(B^*C/D), which would be about 11%. To be conservative with the estimation method, we set uncertainty sourced by the background estimation method at 20%. Last, our signal MC in SR(A) has enough statistics. However, it is not infinite. Therefore, we need to consider the statistical error of the signal MC in the signal region. We have 300 events in signal region A, which results in 6% of statistical error from $\frac{\sqrt{N}}{N}$, $\frac{17}{300}$.

CHAPTER 10

RESULTS

CMS analyses use the “Higgs Analysis” group’s “Combined Limit” to set limit values on various BSM processes. The tool comes with various statistical options developed by statisticians and physicists. We use the Combined Tool’s “AsymptoticLimits” Approach to set limit on the branching ratio of $\mathbf{B}(H \rightarrow SS \rightarrow \tau^+\tau^-\tau^+\tau^-)$. Assuming data events are in agreement with the SM expectation, we calculated the expected values on the exclusion limit for the branching ratio of $\mathbf{B}(H \rightarrow SS \rightarrow \tau^+\tau^-\tau^+\tau^-)$. We input the expected background yield (B^*C/D) for actual background yield to calculate the expected limit values. After calculating the expected limit, we unblind ourselves and look at the actual data event in the SR(A). There are nine events in SR(A), which is within the statistical uncertainty of the SM expectation. Thus, we did not find the BSM LLPs in our data from the analysis strategy.

Table 10.1: Final Data entry in each region

Region	Entry	Uncertainty
CR(D)	2151	± 46.37
CR(C)	311	± 17.63
CR(B)	45	± 6.708
SR(B^*C/D)	6.506	± 2.551
SR(A)	9	± 3

The limits are calculated for LLP’s all MS and $c\tau$ combinations. The best performance is observed for 15 GeV and 10mm point. The table shows each combination’s exclusion limit for 95% confidence level (CL). The results show improvement from previous tracker analysis’ τ lepton final state limit [45]. The values are one of the most stringent results from the LHC results.

Lifetime	MS7	MS15	MS40	MS55
1mm	N/A	184.4	65.9	184.4
10mm	0.0934	0.0402	0.5984	2.006
100mm	0.3359	0.0514	0.0484	0.1742
1000mm	63.50	35.50	1.650	1.468

Table 10.2: Median value of 95% CL exclusion limit on the branching ratio $\mathbf{B}(H \rightarrow SS \rightarrow \tau^+ \tau^- \tau^+ \tau^-)$

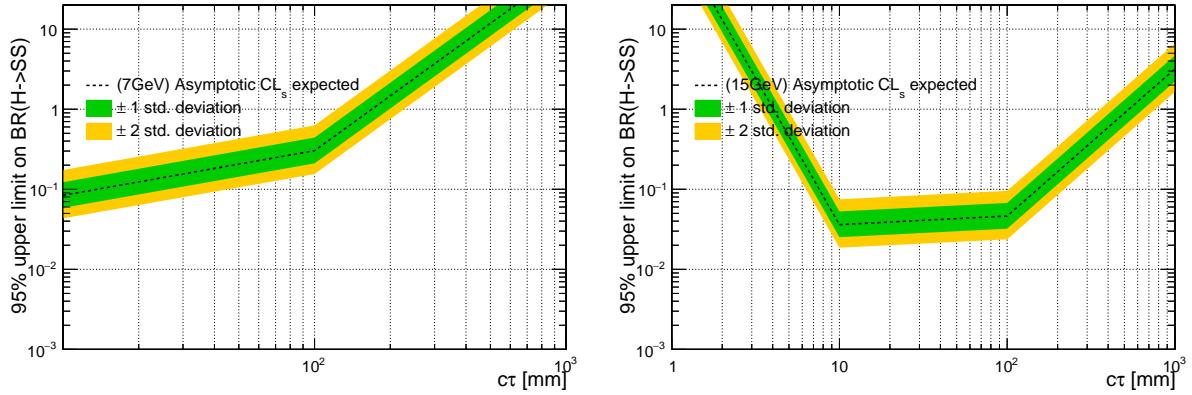


Figure 10.1: Current Preliminary limit plots for lower mass LLP

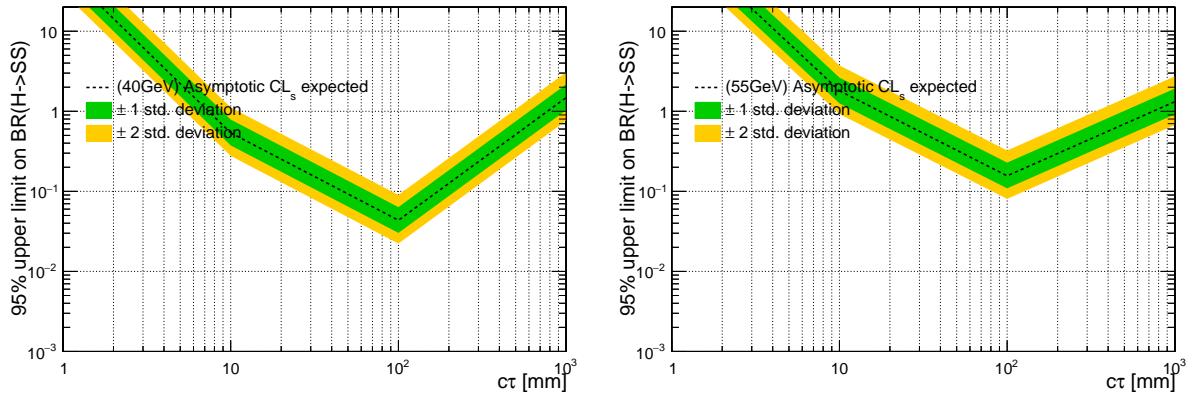


Figure 10.2: Current Preliminary limit plots for higher mass LLP

CHAPTER 11

CONCLUSIONS

The SM still needs extra particles to solve the naturalness problem. The theory of neutral naturalness, which includes Twin Higgs' models and others, postulates long-lived neutral scalar particles. The SM Higgs boson can become a portal to the BSM LLP, and the SM Higgs boson can be created in CMS of the LHC. We used CMS data to observe if we could detect LLPs via the Higgs Portal. Preceding analyses already tested LLPs' decays into b and d quarks, concluded CMS data could not detect LLPs via b and d quark channels, and set limits on the branching ratio of the Higgs to LLPs. However, preceding analyses were unsuccessful in LLP's τ channel decay due to the difficulty in trigger and reconstruction strategy. The analysis exploited a novel B-Parking HLT path to trigger the LLP's τ channel decay while maximizing the signal process's cross-section by targeting the ggH production mode. We reconstructed Regions of Interest, a purely tracker-based object, to avoid subdividing τ lepton's multiple decay modes. Signal ROIs' vertices and track parameters are very different from background ROIs' because of LLP's distant decay in the tracker volume. We used ML for the multivariate analysis tool. We used only the leading and the subleading scores of an event due to two LLP presence in a signal event. Muon's IP significance value was also used for event selection. We used the ABCD method to predict the background rate in the signal region. We validated the method without unblinding by looking at the correlation factor at the $\Delta\phi$ inverted regions. Systematic uncertainty originates from \mathcal{L}_{int} recording, lepton scale factors, background estimation, and statistical uncertainties. With these, we used the HiggsAnalysis' CombineTools to set limits on the branching ratio of the Higgs to LLPs for τ lepton channel. We observed no data over the SM expectation, set the limits on the branching ratio of the Higgs to LLPs, $\mathbf{B}(H \rightarrow SS \rightarrow \tau^+\tau^-\tau^+\tau^-)$, to a stringent value. The exclusion limit is one of the most competitive results from CMS analyses, with the branching ratio set to a single digit for several combinations of MS and $c\tau$.

APPENDIX A

DATA SAMPLES

A.1 BrilCalc command for B-Parking HLT path

The analysis uses B Parking datasets. Data was collected during the 2018 portion of Run 2 and corresponds to an integrated luminosity of 50 fb^{-1} .

Table A.1: Datasets used in the analysis

Data sample	Integrated Luminosity (pb^{-1})
/ParkingBPH1/Run2018A-UL2018_MiniAODv2-v1/MINIAODD	867.13
/ParkingBPH2/Run2018A-UL2018_MiniAODv2-v1/MINIAODD	859.38
/ParkingBPH3/Run2018A-UL2018_MiniAODv2-v1/MINIAODD	866.07
/ParkingBPH4/Run2018A-UL2018_MiniAODv2-v1/MINIAODD	867.38
/ParkingBPH5/Run2018A-UL2018_MiniAODv2-v1/MINIAODD	867.12
/ParkingBPH6/Run2018A-UL2018_MiniAODv2-v1/MINIAODD	866.98
Total	5194.1
/ParkingBPH1/Run2018B-UL2018_MiniAODv2-v1/MINIAODD	882.57
/ParkingBPH2/Run2018B-UL2018_MiniAODv2-v1/MINIAODD	881.37
/ParkingBPH3/Run2018B-UL2018_MiniAODv2-v1/MINIAODD	882.31
/ParkingBPH4/Run2018B-UL2018_MiniAODv2-v1/MINIAODD	882.32
/ParkingBPH5/Run2018B-UL2018_MiniAODv2-v1/MINIAODD	882.30
/ParkingBPH6/Run2018B-UL2018_MiniAODv2-v1/MINIAODD	399.63
Total	4810.5
/ParkingBPH1/Run2018C-UL2018_MiniAODv2-v1/MINIAODD	1171.8
/ParkingBPH2/Run2018C-UL2018_MiniAODv2-v1/MINIAODD	1171.8
/ParkingBPH3/Run2018C-UL2018_MiniAODv2-v1/MINIAODD	1168.2
/ParkingBPH4/Run2018C-UL2018_MiniAODv2-v1/MINIAODD	1171.8
/ParkingBPH5/Run2018C-UL2018_MiniAODv2-v1/MINIAODD	1165.3
Total	5848.9
/ParkingBPH1/Run2018D-UL2018_MiniAODv2-v1/MINIAODD	5640.4
/ParkingBPH2/Run2018D-UL2018_MiniAODv2-v1/MINIAODD	5619.6
/ParkingBPH3/Run2018D-UL2018_MiniAODv2-v1/MINIAODD	5634.3
/ParkingBPH4/Run2018D-UL2018_MiniAODv2-v1/MINIAODD	5650.0
/ParkingBPH5/Run2018D-UL2018_MiniAODv2-v1/MINIAODD	5622.1
Total	28166.4
ParkingBPH Total	44019.9

```

brilcalc lumi --byls --normtag
/cvmfs/cms-bril.cern.ch/cms-lumi-pog/Normtags/normtag_PHYSICS.json
-i processedLumis.json --hltpath HLT_Mu9_IP6_part0_v* -o output.csv

```

Please note that we get luminosity for the specific BPH HLT path and save the output for pileupReCalc_HLTpaths.py

```

pileupReCalc_HLTpaths.py -i output.csv
--inputLumiJSON pileup_latest.txt -o My_HLT_corrected_PileupJSON.txt
--runperiod Run2

```

where, pileup_latest.txt is from

```

/afs/cern.ch/cms/CAF/CMSCOMM/COMM_DQM/certification/Collisions18/
13TeV/PileUp/pileup_latest.txt

```

Then, we substitute the usual pileup_latest.txt to My_HLT_corrected_PileupJSON.txt to obtain the PU distribution of data.

```

pileupCalc.py -i processedLumis.json
--inputLumiJSON My\_HLT\_\_corrected\_PileupJSON.txt
--calcMode true --minBiasXsec 69200 --maxPileupBin 120 --numPileupBins 120
MyDataPileupHistogram.root

```

With histograms above, one can create a PUWeight histogram distribution for an era's specific HLT path with commands below.

```

PUdata->Scale(1./PUdata->Integral(0,-1));
PUmc->Scale(1./PUmc->Integral(0,-1));
PUdata->Divide(PUmc);

```

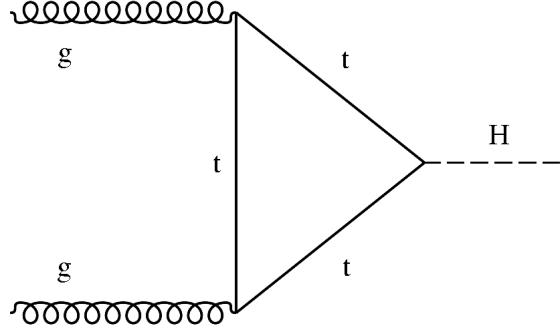
A.2 Monte Carlo Samples

A.2.1 Signal Model and Simulation

The ggH production process (see Figure A.1) is generated at next-to-next-to-leading order (NNLO) and next-to-next-to-leading-log (NNLL) QCD and next-to-leading order (NLO) EW accuracies [66]. The Higgs boson mass is set to 125GeV for all signal samples. The cross sections, computed at NNLO+NNLL QCD and NLO EW accuracies and obtained from CERN Yellow Report 3, are 4.414 pb. The CMS detector response is modeled with GEANT4 [67].

Table ?? lists the signal Monte Carlo samples. Due to lack of Ultra-Legacy campaign for the signal Monte Carlo, the analysis used the pre-Ultra-Legacy campaign signal Monte Carlo.

Figure A.1: Leading Feynman diagrams for ggH production mode



Sample
/ggH_HToSSTo4Tau_MH-125_TuneCP5_13TeV-powheg-pythia8/CAMPAIGN/MINIAODSIM
/ggH_HToSSTo4Tau_MH-125_MS-55_ctauS-1_TuneCP5_13TeV-powheg-pythia8/CAMPAIGN/MINIAODSIM
/ggH_HToSSTo4Tau_MH-125_MS-55_ctauS-10_TuneCP5_13TeV-powheg-pythia8/CAMPAIGN/MINIAODSIM
/ggH_HToSSTo4Tau_MH-125_MS-55_ctauS-100_TuneCP5_13TeV-powheg-pythia8/CAMPAIGN/MINIAODSIM
/ggH_HToSSTo4Tau_MH-125_MS-55_ctauS-1000_TuneCP5_13TeV-powheg-pythia8/CAMPAIGN/MINIAODSIM
/ggH_HToSSTo4Tau_MH-125_MS-40_ctauS-1_TuneCP5_13TeV-powheg-pythia8/CAMPAIGN/MINIAODSIM
/ggH_HToSSTo4Tau_MH-125_MS-40_ctauS-10_TuneCP5_13TeV-powheg-pythia8/CAMPAIGN/MINIAODSIM
/ggH_HToSSTo4Tau_MH-125_MS-40_ctauS-100_TuneCP5_13TeV-powheg-pythia8/CAMPAIGN/MINIAODSIM
/ggH_HToSSTo4Tau_MH-125_MS-40_ctauS-1000_TuneCP5_13TeV-powheg-pythia8/CAMPAIGN/MINIAODSIM
/ggH_HToSSTo4Tau_MH-125_MS-15_ctauS-1_TuneCP5_13TeV-powheg-pythia8/CAMPAIGN/MINIAODSIM
/ggH_HToSSTo4Tau_MH-125_MS-15_ctauS-10_TuneCP5_13TeV-powheg-pythia8/CAMPAIGN/MINIAODSIM
/ggH_HToSSTo4Tau_MH-125_MS-15_ctauS-100_TuneCP5_13TeV-powheg-pythia8/CAMPAIGN/MINIAODSIM
/ggH_HToSSTo4Tau_MH-125_MS-15_ctauS-1000_TuneCP5_13TeV-powheg-pythia8/CAMPAIGN/MINIAODSIM
/ggH_HToSSTo4Tau_MH-125_MS-7_ctauS-1_TuneCP5_13TeV-powheg-pythia8/CAMPAIGN/MINIAODSIM
/ggH_HToSSTo4Tau_MH-125_MS-7_ctauS-10_TuneCP5_13TeV-powheg-pythia8/CAMPAIGN/MINIAODSIM
/ggH_HToSSTo4Tau_MH-125_MS-7_ctauS-100_TuneCP5_13TeV-powheg-pythia8/CAMPAIGN/MINIAODSIM
/ggH_HToSSTo4Tau_MH-125_MS-7_ctauS-1000_TuneCP5_13TeV-powheg-pythia8/CAMPAIGN/MINIAODSIM

Table A.2: Signal Monte Carlo samples

CAMPAIGN:RunIIAutumn18MiniAOD-rp_102X_upgrade2018_realistic_v15-v1

A.2.2 Background Monte Carlo

All samples were processed as recommended in the PPD Run2 Analysis Guideline [68]. All background Monte Carlo samples have been completely produced in the Ultra-Legacy campaign as well. Thus, the analysis used the Ultra-Legacy campaign background Monte Carlo. The asterisk mark denotes the campaign abbreviation, RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1. Tables A.3-A.6 summarizes the background Monte Carlo used in this analysis.

a

Table A.3: QCD MuEnriched Pt5 background Monte Carlo samples

Sample
/QCD_Pt-15to20_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*-v3/MINIAODSIM
/QCD_Pt-20to30_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*-v4/MINIAODSIM
/QCD_Pt-30to50_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*-v3/MINIAODSIM
/QCD_Pt-50to80_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*-v3/MINIAODSIM
/QCD_Pt-80to120_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*_ext1-v2/MINIAODSIM
/QCD_Pt-120to170_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*_ext1-v2/MINIAODSIM
/QCD_Pt-170to300_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*-v3/MINIAODSIM
/QCD_Pt-300to470_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*_ext3-v1/MINIAODSIM
/QCD_Pt-470to600_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*_ext1-v2/MINIAODSIM
/QCD_Pt-600to800_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*-v1/MINIAODSIM
/QCD_Pt-800to1000_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*_ext3-v2/MINIAODSIM
/QCD_Pt-1000toInf_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*-v1/MINIAODSIM

Table A.4: W,Z,H boson background Monte Carlo samples

Sample
/DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/MINIAODSIM
/WJetsToLNu_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/MINIAODSIM
/WW_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/MINIAODSIM
/WZ_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/*-v3/MINIAODSIM
/ZZ_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/MINIAODSIM
/GluGluHToBB_M125_13TeV_amcatnloFXFX_pythia8/*-v1/MINIAODSIM

Table A.5: Top background Monte Carlo samples

Sample
/TTJets_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/MINIAODSIM
/ST_s-channel_4f_hadronicDecays_TuneCP5_13TeV-madgraph-pythia8/*_ext1-v1/MINIAODSIM
/ST_t-channel_top_5f_TuneCP5_13TeV-powheg-pythia8/*-v1/MINIAODSIM
/ST_t-channel_antitop_5f_TuneCP5_13TeV-powheg-pythia8/*-v1/MINIAODSIM
/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/*_ext1-v1/MINIAODSIM
/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/*_ext1-v1/MINIAODSIM

Table A.6: Monte Carlo sample summary

Sample
/QCD_Pt-15to20_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*-v3/MINIAODSIM
/QCD_Pt-20to30_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*-v4/MINIAODSIM
/QCD_Pt-30to50_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*-v3/MINIAODSIM
/QCD_Pt-50to80_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*-v3/MINIAODSIM
/QCD_Pt-80to120_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*_ext1-v2/MINIAODSIM
/QCD_Pt-120to170_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*_ext1-v2/MINIAODSIM
/QCD_Pt-170to300_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*-v3/MINIAODSIM
/QCD_Pt-300to470_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*_ext3-v1/MINIAODSIM
/QCD_Pt-470to600_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*_ext1-v2/MINIAODSIM
/QCD_Pt-600to800_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*-v1/MINIAODSIM
/QCD_Pt-800to1000_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*_ext3-v2/MINIAODSIM
/QCD_Pt-1000toInf_MuEnrichedPt5_TuneCP5_13TeV_pythia8/*-v1/MINIAODSIM
/DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/MINIAODSIM
/WJetsToLNu_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/MINIAODSIM
/WW_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/MINIAODSIM
/WZ_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/*-v3/MINIAODSIM
/ZZ_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/*-v2/MINIAODSIM
/GluGluHToBB_M125_13TeV_amcatnloFXFX_pythia8/*-v1/MINIAODSIM
/TTJets_TuneCP5_13TeV-madgraphMLM-pythia8/*-v1/MINIAODSIM
/ST_s-channel_4f_hadronicDecays_TuneCP5_13TeV-madgraph-pythia8/*_ext1-v1/MINIAODSIM
/ST_t-channel_top_5f_TuneCP5_13TeV-powheg-pythia8/*-v1/MINIAODSIM
/ST_t-channel_antitop_5f_TuneCP5_13TeV-powheg-pythia8/*-v1/MINIAODSIM
/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/*_ext1-v1/MINIAODSIM
/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/*_ext1-v1/MINIAODSIM
/ggH_HToSSTo4Tau_MH-125_TuneCP5_13TeV-powheg-pythia8/*-v1/MINIAODSIM
/ggH_HToSSTo4Tau_MH-125_TuneCP5_13TeV-powheg-pythia8/*-v1/MINIAODSIM

APPENDIX B

MACHINE LEARNING TRAINING VARIABLE

Loss, Accuracy, validation loss, validation accuracy, and AUC values are recorded for the number of epochs run.

Table B.1: Output scores for different Epoch training

Epoch	loss	acc	val_loss	val_acc	AUC
100	0.2161	0.9074	0.2469	0.8942	0.9387
150	0.2039	0.9120	0.2393	0.8983	0.9414
200	0.1934	0.9151	0.2523	0.8953	0.9408
250	0.1977	0.9144	0.2459	0.8982	0.9399
300	0.1738	0.9272	0.2573	0.8977	0.9387
350	0.1607	0.9332	0.2693	0.8934	0.9403
400	0.1459	0.9387	0.2823	0.8970	0.9394

Loss, Accuracy, and AUC values are recorded for the combination of signal point and the background process.

Table B.2: Output scores for different Signal and Background MC combination

	Signal	Background	Loss	Accuracy	AUC
ggH_HToSSSTo4Tau_MH-125_MS-7_ctauS-10	QCDMuEnrichedPt5_Pt20-30	0.1852	0.9244	0.9696	
	QCDMuEnrichedPt5_Pt470-600	0.1597	0.9361	0.9727	
	TTJets	0.1681	0.9320	0.9680	
ggH_HToSSSTo4Tau_MH-125_MS-15_ctauS-10	QCDMuEnrichedPt5_Pt20-30	0.2133	0.9093	0.9598	
	QCDMuEnrichedPt5_Pt470-600	0.1687	0.9306	0.9679	
	TTJets	0.1838	0.9242	0.9610	
ggH_HToSSSTo4Tau_MH-125_MS-15_ctauS-100	QCDMuEnrichedPt5_Pt20-30	0.0747	0.9755	0.9791	
	QCDMuEnrichedPt5_Pt470-600	0.1596	0.9361	0.9727	
	TTJets	0.1681	0.9320	0.9680	
ggH_HToSSSTo4Tau_MH-125_MS-40_ctauS-100	QCDMuEnrichedPt5_Pt20-30	0.1898	0.9209	0.9695	
	QCDMuEnrichedPt5_Pt470-600	0.1765	0.9275	0.9698	
	TTJets	0.1576	0.9379	0.9705	
ggH_HToSSSTo4Tau_MH-125_MS-55_ctauS-100	QCDMuEnrichedPt5_Pt20-30	0.1898	0.9227	0.9635	
	QCDMuEnrichedPt5_Pt470-600	0.1480	0.9399	0.9674	
	TTJets	0.1433	0.9450	0.9706	

REFERENCES

- [1] S.Abach et al. “Search for High Mass Top Quark Production in pp Collisions at s = 1.8 TeV.” In: *Phys. Lett.* 74 (1995), p. 12. DOI: 10.1103/PhysRevLett.74.2422. arXiv: 9411001 [hep-ex].
- [2] F.Abe et al. “Observation of Top Quark Production in pp Collisions with the Collider Detector at Fermilab.” In: *Phys. Lett.* 74 (1995), p. 21. DOI: 10.1103/PhysRevLett.74.2626. arXiv: 9503002 [hep-ex].
- [3] Serguei Chatrchyan et al. “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC.” In: *Phys. Lett.* B716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].
- [4] Georges Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC.” In: *Phys. Lett.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].
- [5] Y.Fukuda et al. “Evidence for Oscillation of Atmospheric Neutrinos.” In: *Phys. Lett.* 81 (1998), p. 9. DOI: 10.1103/PhysRevLett.81.1562. arXiv: 9807003 [hep-ex].
- [6] Matthew Baumgart et al. “Non-Abelian Dark Sectors and Their Collider Signatures.” In: *JHEP* 04 (2009), p. 014. DOI: 10.1088/1126-6708/2009/04/014. arXiv: 0901.0283 [hep-ph].
- [7] David E. Kaplan, Markus A. Luty, and Kathryn M. Zurek. “Asymmetric Dark Matter.” In: *Phys. Rev.* D79 (2009), p. 115016. DOI: 10.1103/PhysRevD.79.115016. arXiv: 0901.4117 [hep-ph].
- [8] Yuk Fung Chan et al. “LHC Signatures of a Minimal Supersymmetric Hidden Valley.” In: *JHEP* 05 (2012), p. 155. DOI: 10.1007/JHEP05(2012)155. arXiv: 1112.2705 [hep-ph].
- [9] Keith R. Dienes and Brooks Thomas. “Dynamical Dark Matter: I. Theoretical Overview.” In: *Phys. Rev.* D85 (2012), p. 083523. DOI: 10.1103/PhysRevD.85.083523. arXiv: 1106.4546 [hep-ph].
- [10] Keith R. Dienes, Shufang Su, and Brooks Thomas. “Distinguishing Dynamical Dark Matter at the LHC.” In: *Phys. Rev.* D86 (2012), p. 054008. DOI: 10.1103/PhysRevD.86.054008. arXiv: 1204.4183 [hep-ph].
- [11] Yanou Cui and Brian Shuve. “Probing Baryogenesis with Displaced Vertices at the LHC.” In: *JHEP* 02 (2015), p. 049. DOI: 10.1007/JHEP02(2015)049. arXiv: 1409.6729 [hep-ph].
- [12] T Aaltonen et al. “High-precision measurement of the W boson mass with the CDF II detector.” In: *Science* 376 (2022) 6589, 170-176 (2022). DOI: 10.1126/science.abk1781.

- [13] Satoshi Mishima and Motoi Endo. “New physics interpretation of W-boson mass anomaly.” In: (2022), p. 21. DOI: doi.org/10.48550/arXiv.2204.05965. arXiv: 2204.05965 [hep-ph].
- [14] *CMS Supersymmetry Physics Results*. Tech. rep. Geneva: CERN, 2022. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS>.
- [15] Jared Barron and David Curtin. “On the Origin of Long-Lived Particles.” In: *Journal of High Energy Physics* 61 (2020), p. 39. DOI: doi.org/10.48550/arXiv.2007.05538. arXiv: 2007.05538 [hep-ph].
- [16] T. Aaltonen et al. “Search for heavy metastable particles decaying to jet pairs in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV.” In: *Phys. Rev.* D85 (2012), p. 012007. DOI: [10.1103/PhysRevD.85.012007](https://doi.org/10.1103/PhysRevD.85.012007). arXiv: 1109.3136 [hep-ex].
- [17] V. M. Abazov et al. “Search for Resonant Pair Production of long-lived particles decaying to b anti-b in p anti-p collisions at $s^{**}(1/2) = 1.96$ -TeV.” In: *Phys. Rev. Lett.* 103 (2009), p. 071801. DOI: [10.1103/PhysRevLett.103.071801](https://doi.org/10.1103/PhysRevLett.103.071801). arXiv: 0906.1787 [hep-ex].
- [18] Georges Aad et al. “Search for a light Higgs boson decaying to long-lived weakly-interacting particles in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector.” In: *Phys. Rev. Lett.* 108 (2012), p. 251801. DOI: [10.1103/PhysRevLett.108.251801](https://doi.org/10.1103/PhysRevLett.108.251801). arXiv: 1203.1303 [hep-ex].
- [19] Roel Aaij et al. “Search for long-lived particles decaying to jet pairs.” In: *Eur. Phys. J.* C75.4 (2015), p. 152. DOI: [10.1140/epjc/s10052-015-3344-6](https://doi.org/10.1140/epjc/s10052-015-3344-6). arXiv: 1412.3021 [hep-ex].
- [20] Georges Aad et al. “Search for long-lived, weakly interacting particles that decay to displaced hadronic jets in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector.” In: *Phys. Rev.* D92.1 (2015), p. 012010. DOI: [10.1103/PhysRevD.92.012010](https://doi.org/10.1103/PhysRevD.92.012010). arXiv: 1504.03634 [hep-ex].
- [21] Georges Aad et al. “Search for massive, long-lived particles using multitrack displaced vertices or displaced lepton pairs in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector.” In: *Phys. Rev.* D92.7 (2015), p. 072004. DOI: [10.1103/PhysRevD.92.072004](https://doi.org/10.1103/PhysRevD.92.072004). arXiv: 1504.05162 [hep-ex].
- [22] Khachatryan et al. “Search for long-lived neutral particles decaying to quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 8$ TeV.” In: *Phys. Rev. D* 91 (1 Jan. 2015), p. 012007. DOI: [10.1103/PhysRevD.91.012007](https://doi.org/10.1103/PhysRevD.91.012007). URL: <https://link.aps.org/doi/10.1103/PhysRevD.91.012007>.
- [23] Georges Aad et al. “Search for pair-produced long-lived neutral particles decaying in the ATLAS hadronic calorimeter in pp collisions at $\sqrt{s} = 8$ TeV.” In: *Phys. Lett.* B743 (2015), pp. 15–34. DOI: [10.1016/j.physletb.2015.02.015](https://doi.org/10.1016/j.physletb.2015.02.015). arXiv: 1501.04020 [hep-ex].

- [24] R. Aaij et al. “Updated search for long-lived particles decaying to jet pairs.” In: *Eur. Phys. J. C* 77.12 (2017), p. 812. DOI: 10.1140/epjc/s10052-017-5178-x. arXiv: 1705.07332 [hep-ex].
- [25] Roel Aaij et al. “Search for massive long-lived particles decaying semileptonically in the LHCb detector.” In: *Eur. Phys. J. C* 77.4 (2017), p. 224. DOI: 10.1140/epjc/s10052-017-4744-6. arXiv: 1612.00945 [hep-ex].
- [26] Roel Aaij et al. “Search for long-lived heavy charged particles using a ring imaging Cherenkov technique at LHCb.” In: *Eur. Phys. J. C* 75.12 (2015), p. 595. DOI: 10.1140/epjc/s10052-015-3809-7. arXiv: 1506.09173 [hep-ex].
- [27] Albert M Sirunyan et al. “Search for new long-lived particles at $\sqrt{s} = 13$ TeV.” In: *Phys. Lett. B* 780 (2018), pp. 432–454. DOI: 10.1016/j.physletb.2018.03.019. arXiv: 1711.09120 [hep-ex].
- [28] Albert M Sirunyan et al. “Search for long-lived particles with displaced vertices in multijet events in proton-proton collisions at $\sqrt{s} = 13$ TeV.” In: *Phys. Rev. D* 98.9 (2018), p. 092011. DOI: 10.1103/PhysRevD.98.092011. arXiv: 1808.03078 [hep-ex].
- [29] Albert M Sirunyan et al. “Search for long-lived particles decaying into displaced jets in proton-proton collisions at $\sqrt{s} = 13$ TeV.” In: *Phys. Rev. D* 99.3 (2019), p. 032011. DOI: 10.1103/PhysRevD.99.032011. arXiv: 1811.07991 [hep-ex].
- [30] Albert M Sirunyan et al. “Search for long-lived particles using nonprompt jets and missing transverse momentum with proton-proton collisions at $\sqrt{s} = 13$ TeV.” In: *Phys. Lett. B* 797 (2019), p. 134876. DOI: 10.1016/j.physletb.2019.134876. arXiv: 1906.06441 [hep-ex].
- [31] Albert M Sirunyan et al. “Search for new particles decaying to a jet and an emerging jet.” In: *JHEP* 02 (2019), p. 179. DOI: 10.1007/JHEP02(2019)179. arXiv: 1810.10069 [hep-ex].
- [32] *Search for long-lived particles decaying into displaced jets.* Tech. rep. CMS-PAS-EXO-19-021. Geneva: CERN, 2020. URL: <https://cds.cern.ch/record/2717071>.
- [33] M. Aaboud et al. “Search for the Higgs boson produced in association with a vector boson and decaying into two spin-zero particles in the $H \rightarrow aa \rightarrow 4b$ channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.” In: *JHEP* 10 (2018), p. 031. DOI: 10.1007/JHEP10(2018)031. arXiv: 1806.07355 [hep-ex].
- [34] Morad Aaboud et al. “Search for long-lived particles in final states with displaced dimuon vertices in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.” In: *Phys. Rev. D* 99.1 (2019), p. 012001. DOI: 10.1103/PhysRevD.99.012001. arXiv: 1808.03057 [hep-ex].
- [35] Morad Aaboud et al. “Search for the Production of a Long-Lived Neutral Particle Decaying within the ATLAS Hadronic Calorimeter in Association with a Z Boson from pp Collisions at $\sqrt{s} = 13$ TeV.” In: *Phys. Rev. Lett.* 122.15 (2019), p. 151801. DOI: 10.1103/PhysRevLett.122.151801. arXiv: 1811.02542 [hep-ex].

- [36] Morad Aaboud et al. “Search for long-lived particles produced in pp collisions at $\sqrt{s} = 13$ TeV that decay into displaced hadronic jets in the ATLAS muon spectrometer.” In: *Phys. Rev. D* 99.5 (2019), p. 052005. DOI: 10.1103/PhysRevD.99.052005. arXiv: 1811.07370 [hep-ex].
- [37] Morad Aaboud et al. “Search for heavy long-lived multicharged particles in proton-proton collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector.” In: *Phys. Rev. D* 99.5 (2019), p. 052003. DOI: 10.1103/PhysRevD.99.052003. arXiv: 1812.03673 [hep-ex].
- [38] Morad Aaboud et al. “Search for heavy charged long-lived particles in the ATLAS detector in 36.1 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV.” In: *Phys. Rev. D* 99.9 (2019), p. 092007. DOI: 10.1103/PhysRevD.99.092007. arXiv: 1902.01636 [hep-ex].
- [39] Morad Aaboud et al. “Search for long-lived neutral particles in pp collisions at $\sqrt{s} = 13$ TeV that decay into displaced hadronic jets in the ATLAS calorimeter.” In: *Eur. Phys. J. C* 79.6 (2019), p. 481. DOI: 10.1140/epjc/s10052-019-6962-6. arXiv: 1902.03094 [hep-ex].
- [40] Georges Aad et al. “Search for heavy neutral leptons in decays of W bosons produced in 13 TeV pp collisions using prompt and displaced signatures with the ATLAS detector.” In: *JHEP* 10 (2019), p. 265. DOI: 10.1007/JHEP10(2019)265. arXiv: 1905.09787 [hep-ex].
- [41] Georges Aad et al. “Search for Magnetic Monopoles and Stable High-Electric-Charge Objects in 13 Tev Proton-Proton Collisions with the ATLAS Detector.” In: *Phys. Rev. Lett.* 124.3 (2020), p. 031802. DOI: 10.1103/PhysRevLett.124.031802. arXiv: 1905.10130 [hep-ex].
- [42] Georges Aad et al. “Search for displaced vertices of oppositely charged leptons from decays of long-lived particles in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.” In: *Phys. Lett. B* 801 (2020), p. 135114. DOI: 10.1016/j.physletb.2019.135114. arXiv: 1907.10037 [hep-ex].
- [43] Georges Aad et al. “Search for long-lived neutral particles produced in pp collisions at $\sqrt{s} = 13$ TeV decaying into displaced hadronic jets in the ATLAS inner detector and muon spectrometer.” In: *Phys. Rev. D* 101.5 (2020), p. 052013. DOI: 10.1103/PhysRevD.101.052013. arXiv: 1911.12575 [hep-ex].
- [44] Georges Aad et al. “Search for light long-lived neutral particles produced in pp collisions at $\sqrt{s} = 13$ TeV and decaying into collimated leptons or light hadrons with the ATLAS detector.” In: *Eur. Phys. J. C* 80.5 (2020), p. 450. DOI: 10.1140/epjc/s10052-020-7997-4. arXiv: 1909.01246 [hep-ex].
- [45] A.Tumeysan et al. “Search for long-lived particles produced in association with a Z boson in proton-proton collisions at $s = 13$ TeV.” In: *JHEP*. (2021), p. 36. DOI: 10.1103/PhysRevLett.74.2422. arXiv: 2110.13218 [hep-ex].
- [46] Brooks Thomas Shufang Su. “The LHC Discovery Potential of a Leptophilic Higgs.” In: *Phys. Rev. D* 79:095014, 2009 (2009), p. 25. DOI: 10.1103/PhysRevD.79.095014. arXiv: 0903.0667 [hep-ph].

- [47] Z. Chacko, Hock-Seng Goh, and Roni Harnik. “The Twin Higgs: Natural electroweak breaking from mirror symmetry.” In: *Phys. Rev. Lett.* 96 (2006), p. 231802. doi: 10.1103/PhysRevLett.96.231802. arXiv: hep-ph/0506256 [hep-ph].
- [48] Gustavo Burdman et al. “Folded supersymmetry and the LEP paradox.” In: *JHEP* 02 (2007), p. 009. doi: 10.1088/1126-6708/2007/02/009. arXiv: hep-ph/0609152 [hep-ph].
- [49] Haiying Cai, Hsin-Chia Cheng, and John Terning. “A Quirky Little Higgs Model.” In: *JHEP* 05 (2009), p. 045. doi: 10.1088/1126-6708/2009/05/045. arXiv: 0812.0843 [hep-ph].
- [50] Nathaniel Craig et al. “Naturalness in the Dark at the LHC.” In: *JHEP* 07 (2015), p. 105. doi: 10.1007/JHEP07(2015)105. arXiv: 1501.05310 [hep-ph].
- [51] David Curtin and Christopher B. Verhaaren. “Discovering Uncolored Naturalness in Exotic Higgs Decays.” In: *JHEP* 12 (2015), p. 072. doi: 10.1007/JHEP12(2015)072. arXiv: 1506.06141 [hep-ph].
- [52] Csaba Csaki et al. “Searching for displaced Higgs boson decays.” In: *Phys. Rev.* D92.7 (2015), p. 073008. doi: 10.1103/PhysRevD.92.073008. arXiv: 1508.01522 [hep-ph].
- [53] *Facts and Figures about the LHC*. 2022. URL: <https://home.cern/resources/faqs/facts-and-figures-about-lhc>.
- [54] *Facts and figures about the LHC*. URL: <https://home.cern/resources/faqs/facts-and-figures-about-lhc>.
- [55] *Detector*. URL: <https://cms.cern/detector>.
- [56] Scarlet Norberg et al. *Overview of CMS Physics Goals and Detector*. 2015. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSExperiment>.
- [57] Alessandro Bartoloni et al. “The CMS ECAL barrel HV system.” In: *Journal of Instrumentation* (2013). doi: 10.1088/1748-0221/8/02/C02039.
- [58] CMS Collaboration. “Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13$ TeV.” In: *JINST* 15 (2019). doi: doi.org/10.1088/1748-0221/15/05/P05002.
- [59] CMS Collaboration. “Alignment of the CMS silicon tracker during commissioning with cosmic rays.” In: *JINST* 5 (2010). doi: 10.1088/1748-0221/5/03/T03009. arXiv:0910.2505: 1501.04020 (physics.ins-det).
- [60] R.Aaij et al. “Test of lepton universality with $B^0 \rightarrow K^{*0} l^+ l^-$ decays.” In: *JHEP* 08 (2017). doi: [doi:10.48550/arXiv.1705.05802](https://doi.org/10.48550/arXiv.1705.05802). eprint: 1705.05802 (hep-ex).
- [61] CMS collaboration. “Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2.” In: (2016).

- [62] Khachatryan et al. *Muon POG – Muon Identification Run2*. 2020. URL: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideMuonIdRun2>.
- [63] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The anti- k_t jet clustering algorithm.” In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ex].
- [64] Khachatryan et al. *Jet Identification Run2 – 2018*. 2020. URL: <https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2018>.
- [65] *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV*. Tech. rep. CMS-PAS-LUM-18-002. Geneva: CERN, 2019. URL: <https://cds.cern.ch/record/2676164>.
- [66] C. Mariotti Heinemeyer et al. “Handbook of LHC Higgs Cross Sections: 3. Higgs Properties.” In: (2013), p. 404. DOI: 10.5170/CERN-2013-004. arXiv: 1307.1347 [hep-ph].
- [67] S. Agostinelli et al. “GEANT4: A Simulation toolkit.” In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.
- [68] Khachatryan et al. *Physics Performance Datasets – PPD Run2 Analysis Recipe Summary*. 2020. URL: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/PdmVAnalysisSummaryTable>.

BIOGRAPHICAL SKETCH

The author was born in Ulsan, Republic of Korea. The author was originally educated in Republic of Korea and traveled to the United States of America for his higher education. After finishing high school, the author completed a Bachelor of Science degree in physics at Washington University in St.Louis. He received his degree in Spring of 2016. While his time in Washington University in St.Louis, he studied physics, mathematics, computer science, and many other humanity courses such as African-American theater, German, and Economics. Following a four years of work in physics, the author decided to pursue the Doctoral degree in physics. The author went to Florida State University to pursue graduate work. He joined the CMS group in 2018 and continued his work in the CMS collaboration for 4 years. From 2018 to 2020, he worked as an assistant researcher for the [45] analysis. Starting 2021, he joined the ROI exotica group, specializing focus on the tau lepton final state.