THE CATHOLIC UNIVERSITY OF AMERICA

Search for Dark Matter produced in association with Monotop in the fully leptonic channel in proton-proton collisions at 13 TeV with the Compact Muon Soleniod experiment

A DISSERTATION

Submitted to the Faculty of
Department of Physics
School of Arts and Sciences
Of The Catholic University of America
In Partial Fulfillment of the Requirements
For the Degree
Doctor of Philosophy

By

Rishabh Uniyal

Washington, D.C.

2023

Search for Dark Matter produced in association with Monotop in the fully leptonic channel in proton-proton collisions at 13 TeV with the Compact Muon Soleniod experiment

Rishabh Uniyal, Ph.D.

Director: Dr. Aaron Dominguez, Ph.D.

Co-Director: Dr. Rachel Bartek, Ph.D.

A search for dark matter has been performed using the monotop model, with the dark matter signature in the form of large missing energy, presence of a b-tagged jet and an isolated lepton (electron and muon). The data analyzed in this analysis was collected from 2016-2018 amounting to an integrated luminosity of 137 fb^{-1} from the proton-proton collisions at a center-of-mass energy ($\sqrt{s}$) of 13 TeV and upper exclusion limits on the mass of the dark matter mediator and the dark matter particle have been set at 95% confidence level.

This dissertation by Rishabh Uniyal fulfills the dissertation requirement for the doctoral degree in Physics approved by Dr. Aaron Dominguez, Ph.D., as Director, by Dr. Rachel Bartek, Ph.D. as Co-Director, and by Dr. Grzegorz Kalicy, Ph.D., as Reader.

Dr. Aaron Dominguez, Ph.D., Director

Dr. Rachel Bartek, Ph.D., Co-Director

Dr. Grzegorz Kalicy, Ph.D., Reader

To Maa, Papa, and Taitun

ACKNOWLEDGEMENTS

Graduate school was one heck of a journey! I have no idea how I made it through, but what I have a very good idea of is, the gratitude towards all the people who helped me in one way or the other to get through.

Firstly, Maa, Papa and Taitun for all the support, love, encouragement and happiness which motivated to keep going.

Secondly, my thesis advisers Aaron and Rachel for not just teaching physics but teaching us to be kind, empathetic, understanding, patient and doing things not just for the sake of doing them. The Fermilab Monotop group Doug and Matteo for allowing me to join the group and Michael, Sebastian, Ji-eun and Sunil for all the support.

Thirdly, my best friends from my undergrad, Raavala, Baba, Sam, Parashar and Jaini for all the fun times, it's been 10 years since we graduated undergrsad. My grad school friends, my housemates during grad school Andreza, Rafael and their cat garfs.

Lastly, everyone who helped me in one way or the other.

It is Done! thank you to all.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

The Standard Model (SM) of Particle Physics is one of the most complete and versatile theories which describes the fundamental particles and the forces with which they interact. The SM has been successful at describing many features of the nature that we have observed in our experiments. From the precise measurement of the electron's magnetic dipole moment to the discovery of the Higgs Boson. The predictions of the SM have just been really precise. While the SM may has been really successful, there are some phenomenon like absence of the description of Dark Matter (DM) in the SM which prevents it from taking the title of the most complete and unified theory.

Dark Matter is a form of material that neither emits, absorbs or reflects any Electromagnetic radiation. It does seem to have mass given its Gravitational effects. Dark Matter contributes to around 24% of the total mass energy of the universe (or 5 times the mass of ordinary matter) but does not interact directly with light, so we have not observed it yet. Astrophysical observations like Gravitational Lensing [2] (bending of light coming from distant galaxies by massive galaxies or galaxy clusters) and motion of galaxies at speeds so high for ordinary matter to sustain, Galaxy Cluster Collisions [3] and Dark Matter seeded Galaxy formations [4] have provided evidence for the presence of Dark Matter.

There are different ways in which Dark Matter could be detected– direct detection, indirect detection and making Dark Matter on the Earth and then detecting it. The Large Hadron Collider (LHC) collides proton beams at the highest energy in the world. The proton beams collide at four main interaction points at the LHC. At each interaction point is a detector. In this thesis we will show the search for Dark Matter, based on the proton-proton collision data recorded by the Compact Muon Solenoid (CMS) experiment at the LHC thus performing a data exploration journey for the quest of dark matter.

This thesis is consists of 7 chapters. **Chapter 1** gives an introduction about the Standard Model of particle physics followed by explaining the Electromagnetic force, the Strong force and the weak interactions. We will also talk about motivations we want to look for beyond standard model of particle physics searches. In **Chapter 2** we discuss the motivations for searching Dark Matter including discussions on the Galactic Rotation curves, Gravitational lensing and Dark Matter seeded galaxy formations. In **Chapter 3** we will talk about the experimental setup (The Compact Muon Solenoid Experiment) used for performing the analysis. We will discuss the technical functioning of different sub-detectors and how the detector records data in general. In **Chapter 4**, we will discuss the High-Luminosity upgrade of the LHC, why did we need the upgrade and what are the key changes in the CMS detector to mitigate the effects of increase luminosity. We also discuss in details the upgrade related to the pixel detector and the research and development efforts at the Catholic University by presenting some initial studies .In **Chapter 5** we will discuss the theoretical foundations of the monotop model which predicts the dark matter and the statistical analysis procedure used to extract results. In **Chapter 6** we describe the analysis strategy including event selection, datasets used, corrections and scale factors applied, types of objects (Electrons, muons, etc.) used, control regions selections to study the backgrounds, systematic uncertainties and statistical tests for validating the results. And Finally, in **Chapter 7**, we discuss the results and future work.

1.1 The Standard Model of Particle Physics

Talking about the limitations has always been the first thing that gets discussed in the context of standard model (SM), but the SM has been by-far the most successful theory in describing the fundamental particles and the forces driving their interactions but of course as physicists, we have to keep questioning a theory until it gives answers to our experimental observations. In this section, we will first discuss the success of the SM with a brief discussion and followed by its shortcomings and the search for physics beyond the SM.

The Standard Model of particle physics is a theoretical framework that describes the fundamental building blocks of matter and the forces that govern their interactions. It is based on the idea that all matter is made up of tiny, indivisible particles called elementary particles, which are governed by three out of four fundamental forces: the strong nuclear force, the weak nuclear force, and the electromagnetic force. The SM's success can be gauged by the Nobel prize wining discovery of the Higgs boson in 2012, which completed the last missing piece of the SM.

The SM theory is expressed using quantum field theory, which is a mathematical framework combining classical mechanics and quantum mechanics. According to the SM the particles are seen as quantized excitation called field. The SM is relies on two key invariance rules, the gauge invariance and frame invariance. In the former, the SM is expressed as $SU(3) \times SU(2) \times U(3)$ gauge invariant, where the groups correspond to the three fundamental interaction forces. In the later, the SM is invariant under rotation, translation and boosts.

The SM elementary particles that make up all matter are divided into two categories: quarks and leptons. Quarks are the building blocks of protons and neutrons. There are six types of quarks: up, down, charm, strange, top, and bottom, and there are six types of leptons: electrons, muons, taus, and their corresponding neutrinos. The fundamental forces in SM are described by the exchange of particles called force carriers. The strong nuclear force, which holds the quarks together inside protons and neutrons, is mediated by particles called gluons. The weak nuclear force, which is responsible for certain types of radioactive decay, is mediated by particles called $W^{\pm}$ and Z^0 bosons. The electromagnetic force, which is responsible for the interaction between electrically charged particles, is mediated by particles called photons. See fig. 1.1

1.1.1 The Electromagnetic Interactions

The electromagnetic force is one of fundamental forces which governs the interaction between electromagnetically charged particles, this interaction is mediated by the exchange of photons. The mathematical theory which describes this interaction is called quantum electro-

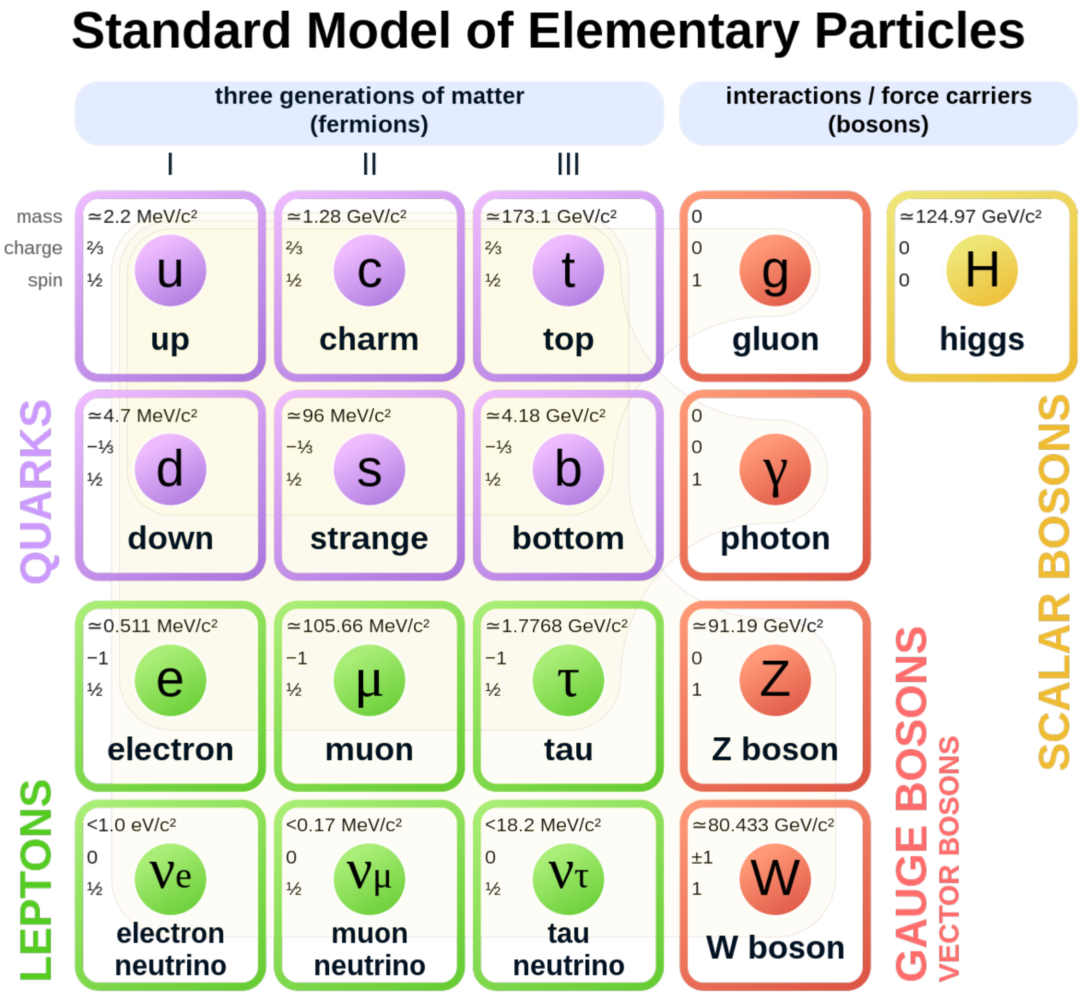


Figure 1.1: This figure adapted from [1] shows the division of particles and bosons according to the SM. Here, the fermions are divided into three generations and boson. The fermions consist of quarks and leptons and the bosons consist of gauge bosons which could be scalar or vector.

dynamics (QED). The Lagrangian density of the Dirac field ψ describing the propagation of fermions is given by eq. 1.1

$$\mathcal{L} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi \tag{1.1}$$

where γ_{μ} are Dirac matrices.

The U(1) group represents the group of rotations around a fixed axis, and the above Lagrangian must be invariant under a U(1) transformation. To do this, an additional term (e)is added to the Lagrangian as given by eq 1.2 and A_{μ} is the gauge transformation given by eq 1.3

$$\mathcal{L} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi - (q\bar{\psi}\gamma_{\mu}\psi)A_{\mu} \tag{1.2}$$

$$A_{\mu} \rightarrow A_{\mu} + \partial_{\mu}\lambda \tag{1.3}$$

Now, we introduce a kinetic term for the photon vector field which must be invariant under 1.3, thus giving the complete Lagrangian as:

$$\mathcal{L} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi - (q\bar{\psi}\gamma_{\mu}\psi)A_{\mu} - \frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} \tag{1.4}$$

So, the above Lagrangian describes the field which interacts with a particle of charge q, by the exchange of a photon (m=0, s=1). The electromagnetic coupling constant is defined as:

$$\alpha_{em}(\mu) = \frac{q(\mu)^2}{4\pi} \tag{1.5}$$

where q(μ) describes the dependence of the charge q on the energy scale of the interaction (μ).

1.1.2 The Strong Interactions

As the name suggests that this is a strong force and is actually the strongest of all the four fundamental forces. The SM theory that describes is force is called quantum chromodynamics (QCD). Analogous to QED, in QCD we see that the strong force exists between particle possessing a property called 'color' charge, this 'color' has nothing to do with the colors in the visible light spectrum but is used as a representation. There are three 'color' charges, the red, green and blue. The strong force interaction takes place with the exchange of a mediator or gauge boson called gluon. In order to quantify the ever large zoo particles physicists have used the concept of group theory, and used the local gauge invariance of the QCD lagrangian given by eq

$$\mathcal{L}_{QCD} = -\frac{1}{4}F^{a\mu\nu}F^{a}_{\mu\nu} + \sum_{f=flavor} \bar{q}_f(i\not{D} - m_f)q_f \tag{1.6}$$

where: q_f is the color charge triplet

$$q_f = \begin{pmatrix} q_f^{red} \\ q_f^{green} \\ q_f^{blue} \end{pmatrix} \tag{1.7}$$

$$D_\mu = \partial_\mu + igA^a_\mu T^a \tag{1.8}$$

where, T^a are generators of $SU_c(3)$ for a = 1,2....8 the field strength tensor $F^a_{\mu\nu}$ is defined as:

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - gf^{abc}A^b_\mu A^c_\nu \tag{1.9}$$

where g is a function of the energy scale of interaction (μ), and the qcd coupling constant $\alpha_s(\mu)$ is defined as

$$\alpha_s(\mu) = \frac{g_s(\mu)^2}{4\pi} \tag{1.10}$$

One important thing to note about the QCD interaction is the gluon-gluon self interaction indicated by the first term in the QCD lagrangian in eq 1.6, this interaction is due to the 'color' charge

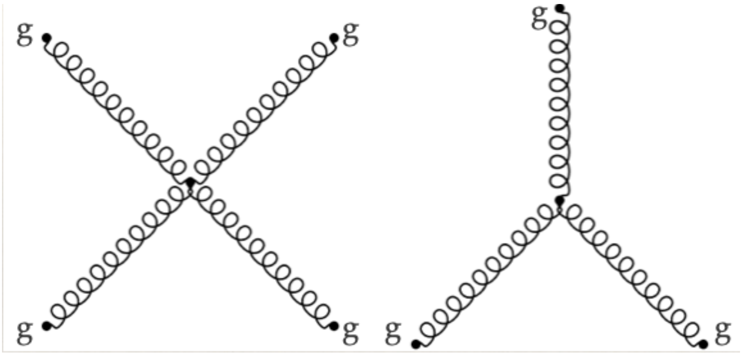


Figure 1.2: This figure shows the gluon-gluon self interaction in a three and four gluon setup in a QCD interaction. This occurs because of the non-abelian nature of the SU(3) interaction account for the color charge carried by the gluons.

carried by the gluons and allows them to interact with other gluons. Fig 1.2 shows the 4-gluon and 3-gluon self interaction.

1.1.3 The Electro-Weak Interactions

In the 1970's the unification of Electromagnetic and the weak force was proposed by the Wienberg-Salam-Galshow model, the model was explained by using the SU(2) X U(1) gauge symmetry. The $SU(2)_L$ gauge symmetry corresponds to the left handed weak isospin triplets and the $U(2)_Y$ corresponds to the singlet coupling to a weak hypercharge.

The electroweak lagrangian is defined as

$$\mathcal{L}_{EWK} = \mathcal{L}_g + \mathcal{L}_f + \mathcal{L}_h + \mathcal{L}_Y \tag{1.11}$$

The $\mathcal{L}_g$ describe the gauge interaction between the W vector boson triplet and the B vector boson singlet defined by 1.12, where, $W_{a\mu\nu}$ (for a =1,2,3) and $B_{\mu\nu}$ are the field tensors.

$$\mathcal{L}_g = -\frac{1}{4}W_a^{\mu\nu}W_{\mu\nu}^a - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} \tag{1.12}$$

The $\mathcal{L}_f$ describe the fermionic interaction between the W vector boson triplet and the B vector boson singlet defined by 1.13, where, j represents the three generations of fermions, $\not{D}$ is contraction of the 4-gradient with the Dirac matrices as defined above.

$$\mathcal{L}_f = \bar{Q}_j i\not{D} Q_j + \bar{u}_j i\not{D} u_j + \bar{d}_j i\not{D} d_j + \bar{L}_j i\not{D} L_j + \bar{e}_j i\not{D} e_j \tag{1.13}$$

The $\mathcal{L}_h$ describes the Higgs field lagrangian and the interaction of the higgs boson field h with the gauge bosons defined by equation 1.14

$$\mathcal{L}_h = |D_\mu h|^2 - \lambda \left(|h|^2 - \frac{v^2}{2} \right)^2 \tag{1.14}$$

The above equations have been adapted from [2]

1.2 Limitations of the Standard Model

Even though the SM has been successful in describing the fundamental particles and their interactions, there are definitely some areas where it has not been as successful in explaining the new physics. In this section we will talk about the limitations of the SM and what lies beyond the SM.

The gravitational force is not included in the SM. In particle physics we deal with particles in the sub-atomic domain, like electrons, protons (quarks), etc. The gravitational force between these fundamental particle is negligible compared to the gravitational force between two rocks, the SM has not been able to explain the reason for why the magnetic force of a tiny magnet is significantly larger that the gravitational force of the earth on a small object.

This above problem thus leads us to the SM has also not being able to explain the hierarchy problem [3], According to the hierarchy problem, we have not been able to find a reason why the fundamental forces are not comparable in terms of the strength of the forces, an example of which is the fact that the weak force is 10^{34} times stronger that the gravitational force.

Next, the matter-antimatter asymmetry [4]. At the time of the Big Bang [5], both matter and antimatter were present in equal amounts but now as we look around, everything is matter with not so much antimatter. So, there had to be some sort of external force which created an imbalance, leading to our surrounding to become matter dominant. The cause of this effect is still unknown and the SM doesn't seem to have an explanation for this.

The presence if dark matter has been confirmed by the astrophysical observations such as galactic rotation curves, gravitational lensing and dark matter seeded galaxy formations. However, the SM doesn't seem to include particles which could have potential dark matter properties. See sec 2.1 for detailed discussion on these observed effects.

The above unexplained scenarios thus lead to look into a beyond the SM theory. So, in this thesis we'll be exploring the monotop model which gives us a process for the presence of dark matter.

CHAPTER 2

Motivations for Dark Matter Search

This chapter provides an overview of the currently known evidences of the presence of Dark Matter. There have been observations which indicate the presence of Dark Matter but most of these observations have been indirect in nature. These observation are the result of gravitational interaction of the Dark Matter and the astrophysical constructs. So, the motivation behind this research analysis is to find direct evidence for the presence of Dark Matter, it could either be in the form of one or more fundamental particles (which is the hypothesis researched in this thesis) or a completely new phenomenon unexplained by the Standard Model.

This chapter will explain the topics of Galactic rotation curves, Gravitational lensing and Dark Matter seeded galaxy formations.

2.1 Why are we searching for Dark Matter?

2.1.1 Galactic Rotation Curves

Galactic rotation curves are made by calculating the rotational velocity of the stars along the length of the galaxy and measuring the corresponding distance of the star from the center. Figure 2.1 shows that the observed rotation curve for the galaxy Messier 33, which seems to deviate significantly from the theoretical prediction based on the amount of observed luminous matter. One might also find an interesting observation in the tail of rotation curve in Figure 2.1. The flat tail distribution is apparently very common in rotation curve of most galaxies regardless of their diverse shapes, sizes and locations as shown in Figure 2.2. These observations point to the fact that the density profile for a galaxy may be different from the ones in which all mass is concentrated at the center. Modeling of this density profile by using N-body numerical simulation

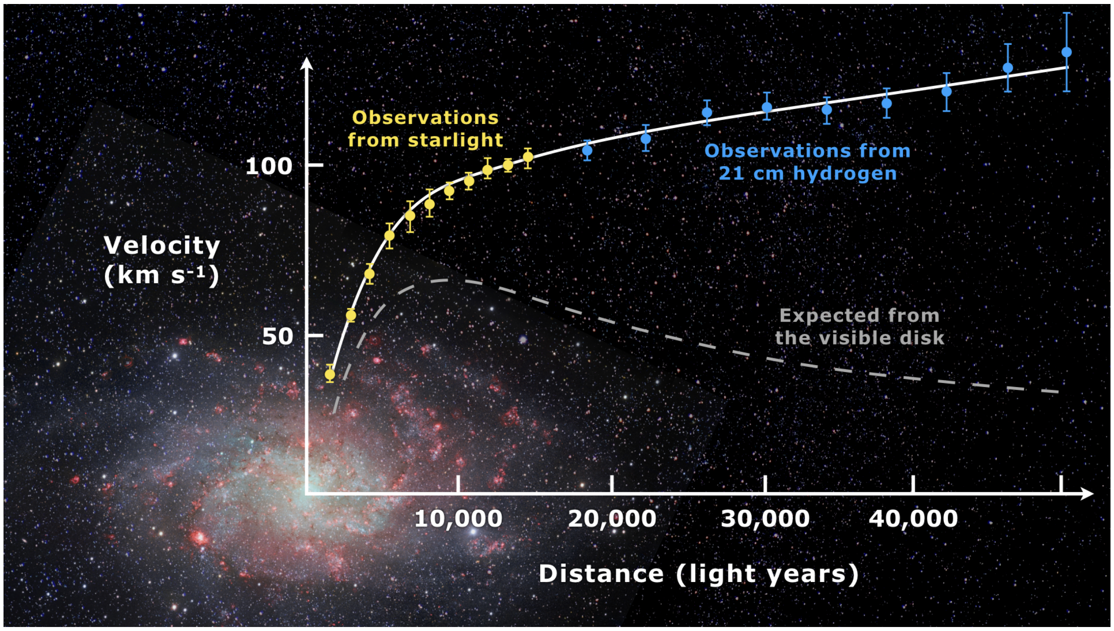


Figure 2.1: This figure depicts the rotational curve of spiral galaxy Messier 33 (yellow and blue points with error bars), and a predicted one from distribution of the visible matter (gray line). The discrepancy between the two curves can be accounted for by adding a dark matter halo surrounding the galaxy.

can help us profile shape of dark matter distributions around the galaxies. One such profile is the Novarro-Frenk-White profile (NFW) [6], given by the following expression 2.1.

$$\frac{\rho(r)}{\rho_{crit}} = \frac{\delta_c}{\frac{r}{r_s}\left(1+\frac{r}{r_s}\right)^2} \tag{2.1}$$

where r is the distance from the center of the galaxy, δ_c is the dimensionless characteristic density, r_s is the radius scale and $\rho_{crit} = 3H^2/8\pi G$ is the critical density for closure.

From Figure 2.3 we can see that observations clearly indicate approximately constant dark matter density in the inner parts of the galaxies, while the cosmological numerical simulations indicate a steep power-law like distribution. This deviation in the theoretical predictions is referred to as the "$core-cusp$" problem [8]. However, the dark matter scientific community seems to have opinions on the reasons for the "$core - cusp$" problem arising either from incomplete understanding of galaxy core dynamics or experimental limitations in the measurements.

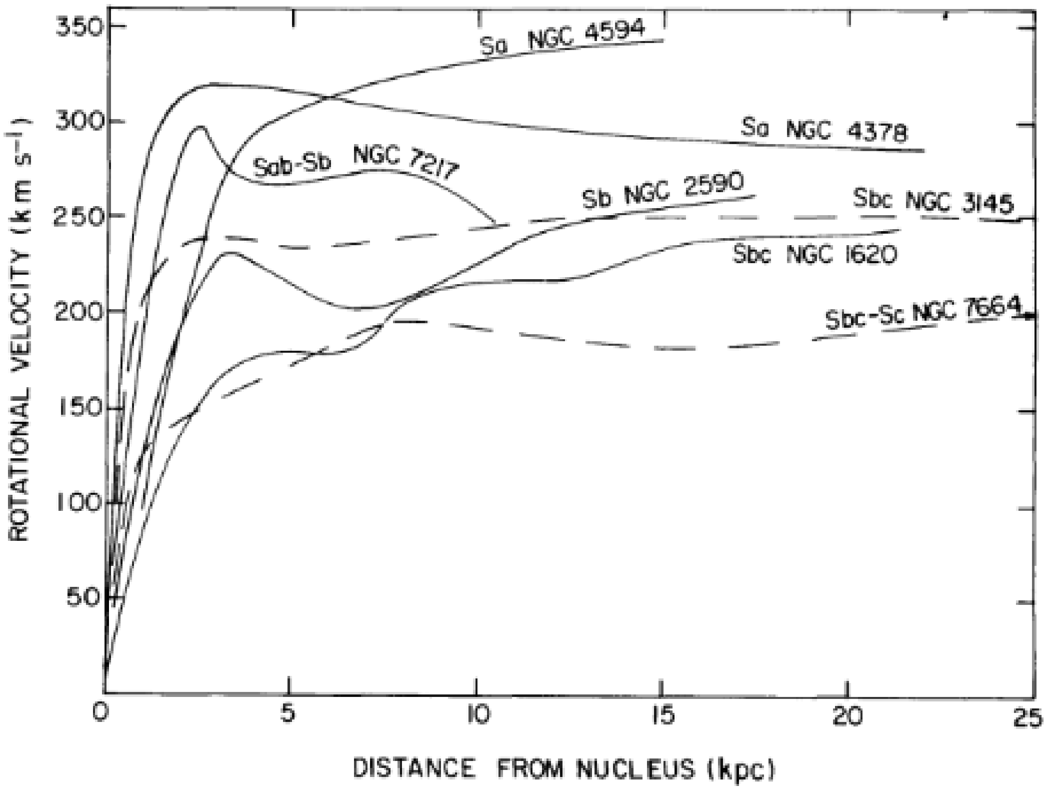


Figure 2.2: This figure depicts the rotational curves of many galaxies as observed by [7]. The flat tail characteristic at large distance is very much evident

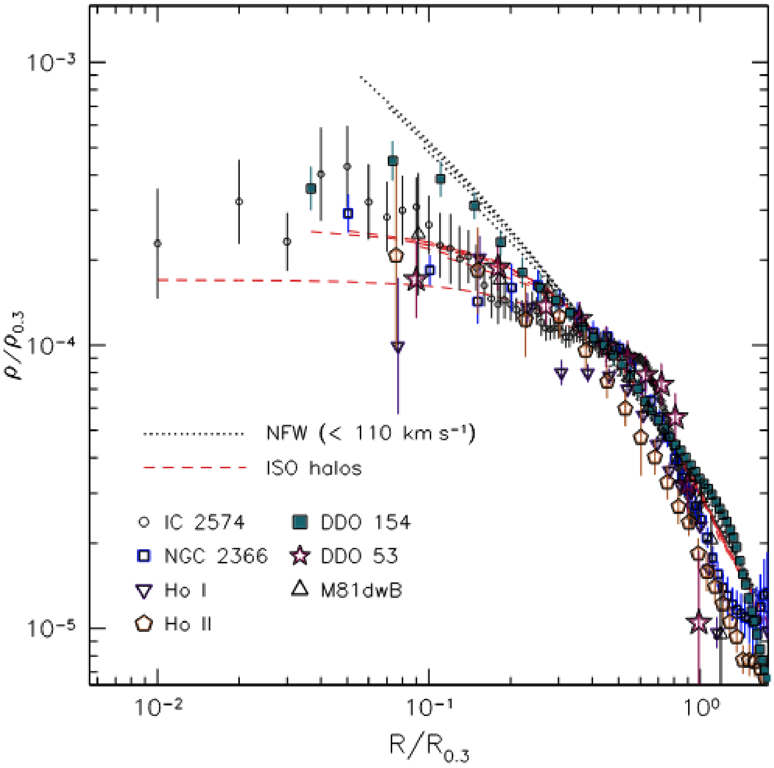


Figure 2.3: This figure depicts the dark matter density profiles of seven dwarf galaxies indicated by the data points and the theoretical prediction model indicated by the black dotted lines. The deviation of the observations from the theoretical predictions as we move to the center of the galaxy is referred as the ”core-cusp” problem.

Be that as it may, we can say that dark matter is necessary for the explanation of the velocities in the galactic rotation curves.

2.1.2 Gravitational Lensing

The phenomenon of gravitational lensing is the bending of light when it travels near massive objects like the sun in analogy with the bending of light through a lens. It is a consequence of Einstein’s General theory of Relativity [9] which states that the path of the light is bent by gravity when it travels close to a massive object like the sun. This phenomenon was experimentally observed by Arthur Eddington and Andrew Claude Crommelin and Charles Davidson separately at the island of Principe (West Africa) and Sobral (Brazil). They measured the position of a bright

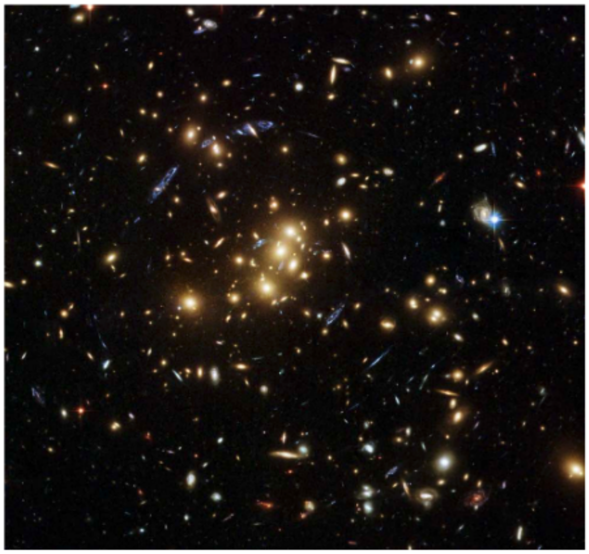


Figure 2.4: This figure shows the effect of gravitational lensing around the galaxy cluster CL0024+17. The elongated blue objects are physically unassociated galaxies lying behind the cluster and appear as tangential arcs. They appear to be distorted because of the gravitational lensing effects. (Figure credit:NASA/ESA/M.J. Jee (John Hopkins University))

group of stars called Haydes, which at the time of the eclipse were behind the sun and were easier to observe because of their brightness. After analysing the photographs it was observed that the stars behind the sun were indeed visible confirming the Einstein's theory of the curvature of the space-time and the phenomenon of gravitational lensing.

Gravitational lensing has been the most successful technique to investigate dark matter. The deflection of the rays of light by this phenomena creates effects such as shifting, distorting and magnifying the images of background galaxies, as shown in figure 2.4. Probing such effects has provided constraints on the properties of dark matter such as mass, mean density relative to matter and size [10].

The assembly of galaxy clusters involves complex physical processes, these processes affect the baryonic mass and the dark matter differently, thus distorting the distributions of baryonic mass and the dark matter.Galaxy clusters contain three basic ingredients: galaxies, intra-cluster gas and dark matter. The ”bullet cluster”, a two cluster setup that collided about 150 years ago

Figure 2.5: This figure illustrates the X-ray emission coming from the hot, intra-cluster gas in the galaxy cluster merger known as the Bullet Cluster [12]. The data suggests that there is a significant difference between the distribution of dark matter and the observed plasma in this merger, which provides strong evidence for the presence of additional, hidden dark matter. The figure includes images from X-ray (left) and lensing (right) observations by NASA, the European Space Agency, and several other institutions.

[11, 12] has provided the most direct evidence for the effect of these processes as shown in the left image of Figure 2.5 indicating that the ingredients have separated. Similar to particle production at a particle collider, the trajectories of the cluster debris are also governed by the properties of its constituents.

Since individual galaxies inside the cluster are well spaced meaning that, they have very low collisional cross-section, implying their continued movement during the collision far from the point of impact. While the intra-cluster gas is uniformly spread and has a large collisional cross-section, thus slowing it down. One may think that all the mass in the cluster can be defined by the above two ingredients, but a significant of amount of mass is observed to be located near the galaxies. For the mass to have travelled this distance it must have some sort of self-interaction collisional cross-section. And gravitational lensing observations require the third ingredient of the bullet cluster which is the dark matter to explain the above effect.

2.1.3 Dark Matter Seeded Galaxy formations

One of the very interesting questions that is posed with the existence Dark Matter is its role in the formation of galaxies. In the research domain of Dark Matter, there have been many theories that have been proposed which explain the formation and evolution of galaxies.

One such theory is the ”bottom-up” scenario [13], in this case tiny clumps of dark matter combine to form large structures such as the galaxies. A strong argument in support of this theory is the cosmic microwave background, which basically tells us that in the early stages of the expansion of the universe it was uniform.

in which small clumps of dark matter form first and eventually merge to create larger structures such as galaxies. This theory is supported by observations of the cosmic microwave background, which show that the universe was much more uniform in the early stages of its expansion.

Another theory is the ”top-down” scenario [14], in this case larger structure like the galaxies decompose into smaller dark matter particles supported by the currently observed large scale structure of the universe showing the universe is not uniform but made from smaller clumps to build large structures like galaxies

Another theory, the Cold Dark Matter theory [15] hypothesizes that dark matter particles are slow moving and can form clumps on small scales.

2.2 Dark Matter search at the LHC

As discussed in section 2.1, astrophysical observations like galactic rotation curve 2.1.1, gravitational lensing 2.1.2 have provided evidence for the presence of dark matter. At the LHC, assuming that dark matter particles and SM particles interact, if the collisions of protons happens in such a way that a dark matter candidate is produced while following the laws of conservation of energy and momenta. At the same time if the Center of mass energy of the collision is large enough to produce a dark matter candidate, we can expect to signatures pointing to the presence

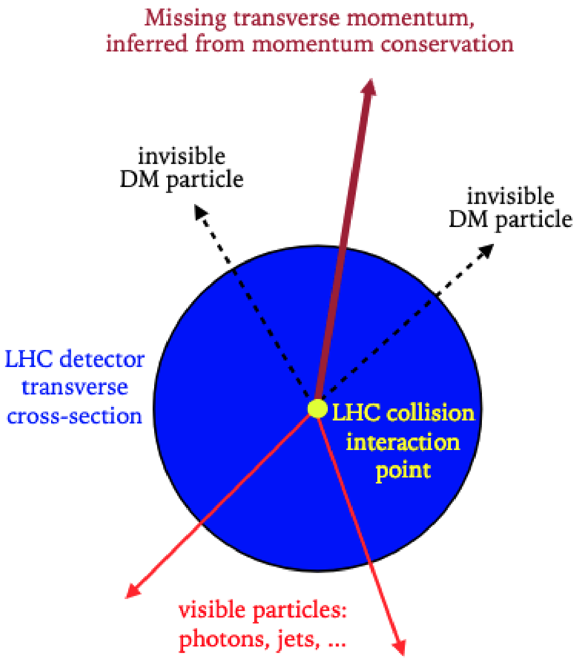


Figure 2.6: This image shows the missing transverse momentum signature of the dark matter at the LHC. adapted from [16]

of dark matter. As we have known that the dark matter is weakly interacting the possible signature at the LHC would be shown as a large missing energy as shown in fig, 2.6

CHAPTER 3

The CMS Experiment at the LHC

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [17] is the world's largest and most powerful particle accelerator. Inside the accelerator, two highly energetic proton beams collide up to a center-of-mass energy of 13 TeV. The LHC is constructed inside a 27 km tunnel, 100 m under the ground, located partly in Switzerland and France.

At the LHC two energetic proton beams travel in opposite directions in separate beam pipes. The proton beams travel in bunches separated by 25 ns in time. The beams travel at relativistic speeds and collide at four major interaction points at the LHC. The LHC has four main detector experiments at each interaction point: ALICE (A Large Ion Collider Experiment), ATLAS (A Toroidal LHC Apparatus), CMS (Compact Muon Solenoid), and LHCb (Large Hadron Collider beauty). ATLAS and CMS are general purpose-detectors with broad physics objectives ranging from studying the Standard Model to searching Beyond Standard Model (BSM) physics such as Dark Matter. While ALICE and LHCb are more physics dedicated detectors, ALICE and LHCb are dedicated to heavy-ion physics and studying b-quark physics respectively.

3.1.1 LHC Computing Grid

LHC produces data at a tremendous rate such that even after filtering 99% of the data, it stores around 88 Petabytes (PB) of data. This data must be stored and made easily accessible to physicists located all over the world. The Worldwide LHC computing grid (WLCG) [18] is responsible for providing resources for storage, distribution for easy data retrieval and analysis of the data collected by the LHC. It combines over 900,000 PB of computing cores from over 170 locations in 42 countries making it the world's largest distributed computing grid. It provides infrastructure which serves the need of around 12,000 physicists by providing real time data access and processing power. During Run II, LHC recorded 0-8 Gb/s, with a global data transfer rate of 60 Gb/s.

There are four main technical components which describe the structure of the WLCG. 1.) Excellent connectivity and networking infrastructure allowing the WLCG to initiate data distribution to hundreds of collaborating institutes. 2.) Each grid center manages a large collection of computer hardware and storage systems. 3. Middle-ware which act as a mediator between the operating systems of the computers and physics application software. 4.) To analyze the data from the LHC and to be able to extract relevant physics information, custom physics analysis based software are needed.

Physically WLCG consists of four layers based on the functionality and computing resources see Figure 3.1. Tier-0 is based at CERN in Geneva, Switzerland. It is responsible for the safe-keeping of the raw data (first copy), first pass reconstruction of the raw data into meaningful information. It distributes the raw data and reconstruction output to the Tier-1s, and reprocessing of data when the LHC is not running.Tier-1 consists of 13 large computer centers responsible for the safe-keeping of a proportional share of raw and reconstructed data, large-scale reprocessing and safe-keeping of corresponding output, distribution of data to Tier 2s and safe-keeping of a share of simulated data produced at these Tier 2s. Tier-2 usually provide facility to store sufficient data computing resources depending on specific analyses needs. Tier-3 are local small computing and storage facilities forming local clusters for accessing WLCG Tiers-0,1,2.

3.2 The Compact Muon Solenoid Experiment

3.2.1 Introduction

The Compact Muon Solenoid (CMS) is a general purpose detector located in eastern France at the Large Hadron Collider (LHC) tunnel 100 m underground. It is a general purpose

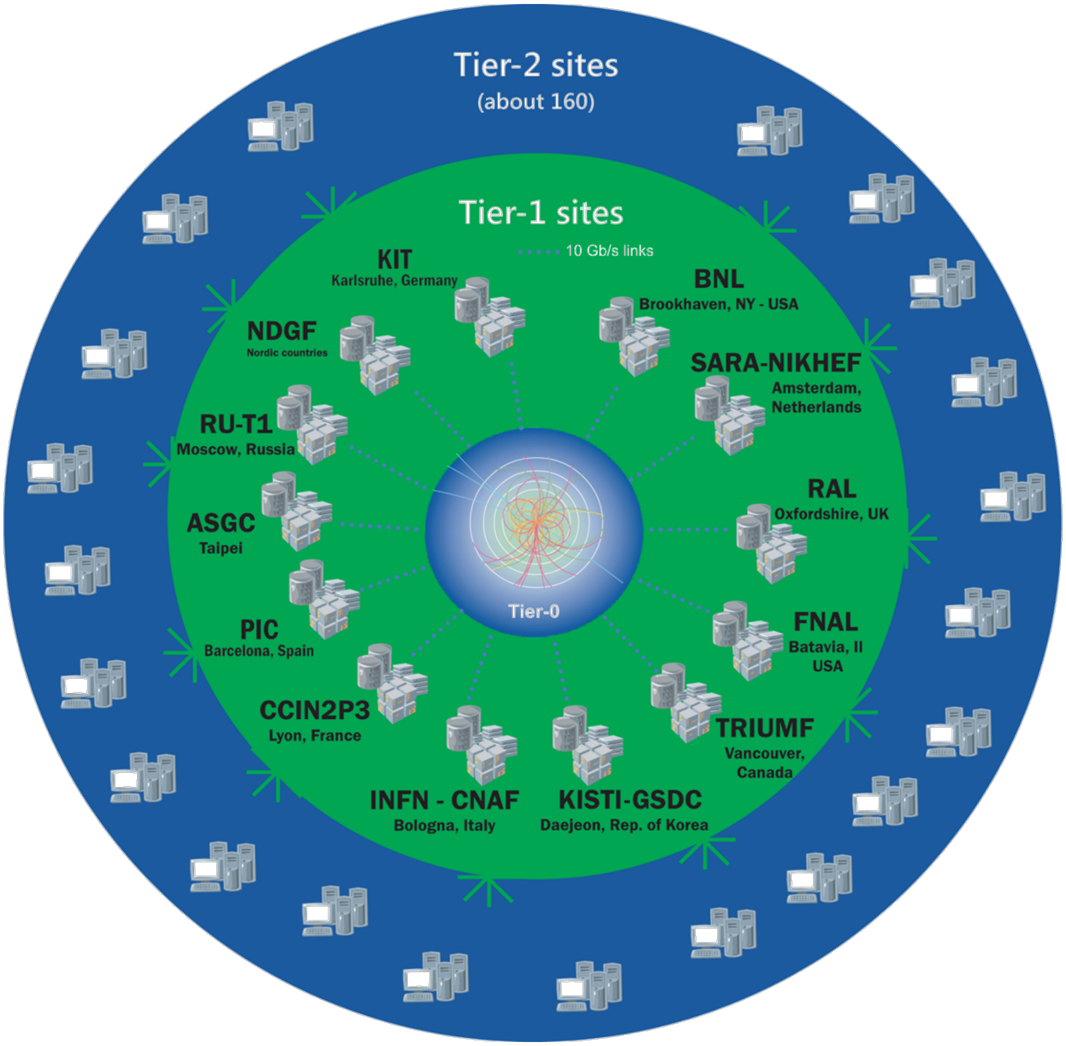


Figure 3.1: This figure adapted from [19] illustrates the 4 layers which compose the WLCG (Tier-0,1,2,3). Tier-0 is the CERN Data Center shown at the center. The green area shows the Tier-1 and outer blue region shows the 160 Tier-2 sites.

detector because of its broad physics searches ranging from studying the Standard Model (SM) to searching for particles that could potentially make Dark Matter (DM). CMS is a cylindrical shaped detector with 21 m in length and 15 m in cross-sectional diameter. The three main characteristics of the CMS experiment are its compactness, accurate detection of Muon particles in the muon system and its magnetic solenoid. The superconducting magnetic solenoid at the core of the CMS detector produces a continuous magnetic field of 3.8 T which is of the order of O(5) of Earth's magnetic field (0.25 - 0.65 gaus). Such large magnetic field is required for a precise measurement of the momentum of high-energy charged particles. All CMS detector subsystems are enclosed by the magnetic solenoid, the muon system, on the other hand, is on the outside. Muons are the particles that are directly detected by the CMS, with a special property that they neither stop nor decay within the boundaries of the detector. They undergo very little energy loss and thus act as a powerful medium to study high-energy process in the presence of high background. The CMS detector is cylindrical shaped with an onion like structure, having several concentric layers of subsystems. These subsystems help prepare ”photographic images” of each collision event by determining the properties of the particles in the collision. CMS has the following basic subsystems ordered from innermost to outermost:

1. Silicon Tracking Detector
2. Electromagnetic Calorimeter
3. Hadronic Calorimeter
4. Superconducting Solenoid Magnet
5. Muon System

The CMS detector can detect 5 general categories of particles as shown in the figure below:

1. Electrons
2. Photons
3. Charged Hadrons
4. Neutral Hadrons
5. Muons

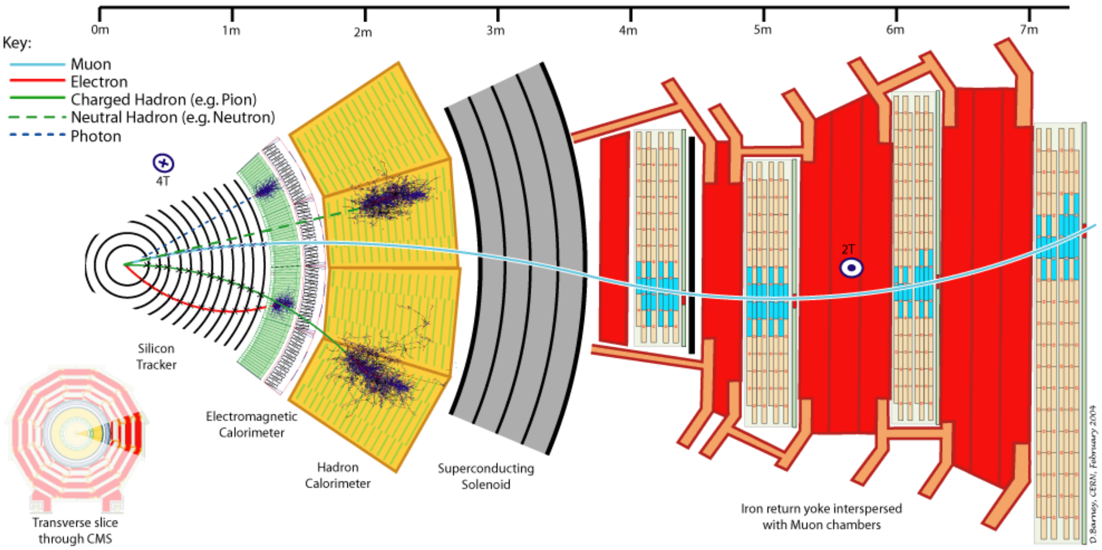


Figure 3.2: This Figure shows the cross-sectional slice of the CMS detector. Here we have shown the different sub-systems of CMS, trajectories and energy deposits of the 5 general categories of particles detected by CMS by one or more of its sub-systems

3.2.2 Coordinate System

The definition of a coordinate system is very crucial in quantifying the position of a particle inside CMS with respect to the origin. CMS uses a right-handed coordinate system defined as follows and illustrated in Figure 3.3 – the origin is defined as the nominal pp collision point, the x-axis points inwards towards the center of the LHC ring, the y-axis points upwards perpendicular to the x-axis and the z-axis points in the counter-clockwise beam direction. However, in hadron collider physics a more traditional coordinate, pseudo-rapidity η is used, defined as

$$\eta = -\ln\tan\left(\frac{\theta}{2}\right) \tag{3.1}$$

For a particle of three-momentum $\mathbf{p}$ with z-component p_z, pseudorapidity can be written as

$$\eta = \frac{1}{2}\ln\left(\frac{|\mathbf{p}|+p_z}{|\mathbf{p}|-p_z}\right) \tag{3.2}$$

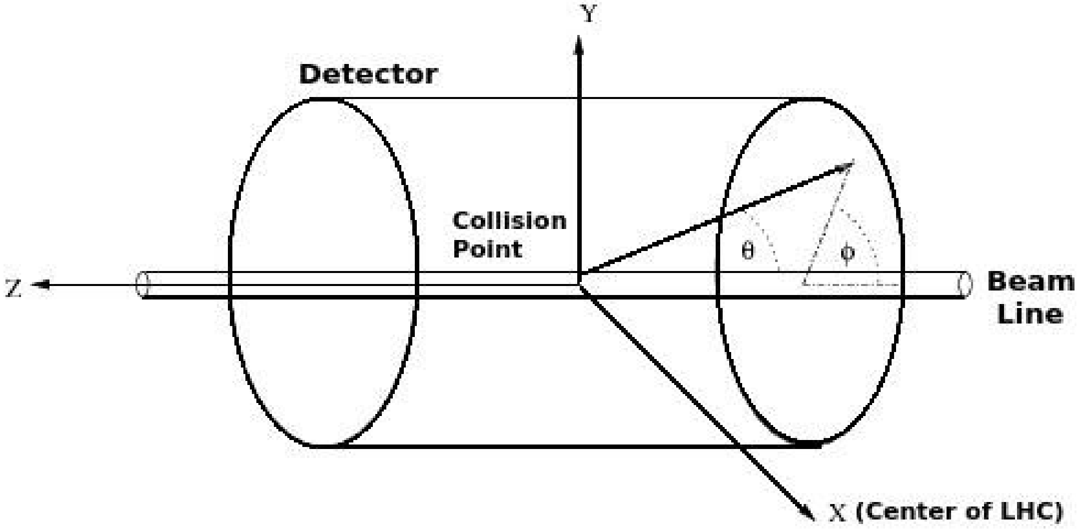


Figure 3.3: This Figure shows the coordinate system used by the CMS detector adapted from [20], illustrating the x-axis, y-axis, z-axis, radial angle θ (measured wrt to the z-axis and the azimuthal angle ϕ (measured in x-y plane from the x-axis).

In the limit of relativistic particles (high velocity and low mass limit), the pseudo-rapidity approximates the rapidity y given by

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right), \tag{3.3}$$

The rapidity is invariant under Lorentz boost transformations and is only dependent on the polar angle θ and not on the energy of the particle. From Figure 3.4, we see that higher values of η correspond to low values of θ, this region is also referred as ”forward” or high η regions. In Figure 3.4, we observe two regions which cover approximately $|\eta|< 1.2$ and $1.2 < |\eta|< 3$, these regions are known as the barrel and the endcap respectively.

3.2.3 Silicon Tracking Detector

Tracking the paths of the particles produced after proton-proton collision inside CMS is of crucial importance. The tracking information from the silicon tracking detector is used to accurately reconstruct the momenta of particles such as electrons, muons or particle jets (colli-

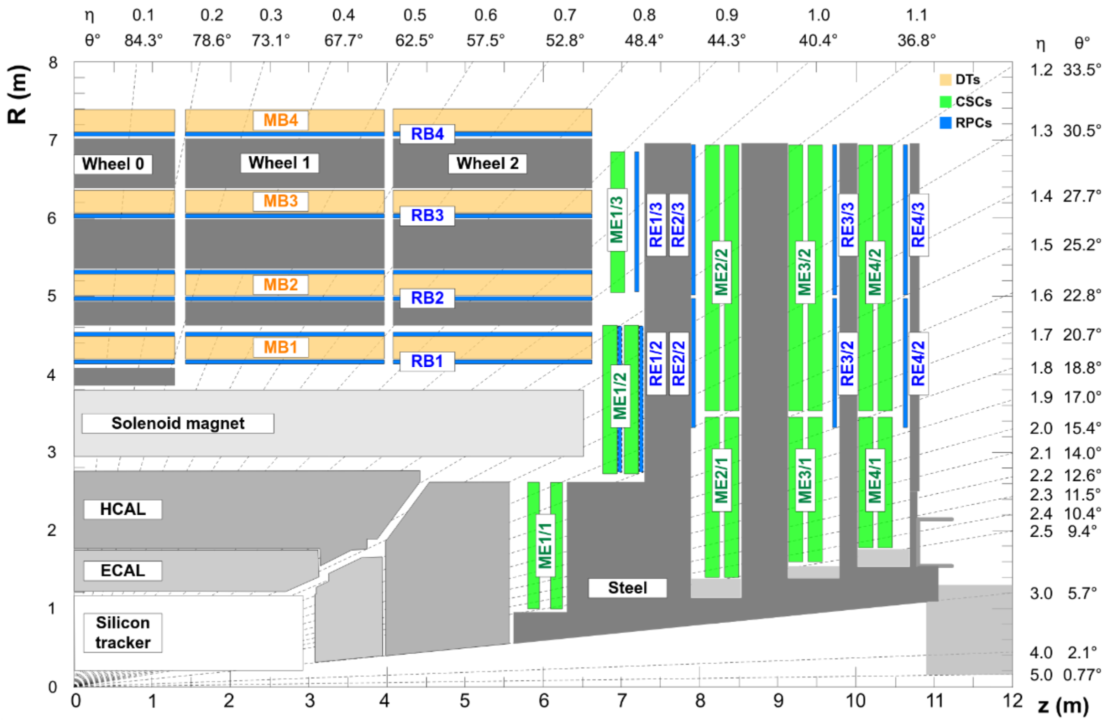


Figure 3.4: This Figure shows one quadrant of CMS in the r-z plane adapted from [21], illustrating the positions of subsystems in barrel and endcap regions

mated sprays of particles initiated by quarks and gluons), mitigate the effect of many overlapping proton-proton collisions (“pileup”), or correctly measure missing momentum. As the charged particles traverse the silicon sensor they create electron-hole pairs, which in the presence of an applied voltage is measured as current.

The magnetic field is responsible for bending the charged particles, the higher the bending smaller is the momenta associated with the particles. The CMS Tracker is located inside the magnetic solenoid and is composed of two sub-systems namely, the pixel and the strip, illustrated in Figure 3.5 and elaborated in the following sections.

3.2.3.1 Pixel

The Pixel sub-system is located in the high particle density region ($r < 20cm$). Being very close to the nominal proton-proton interaction point the pixel detector is exposed to most

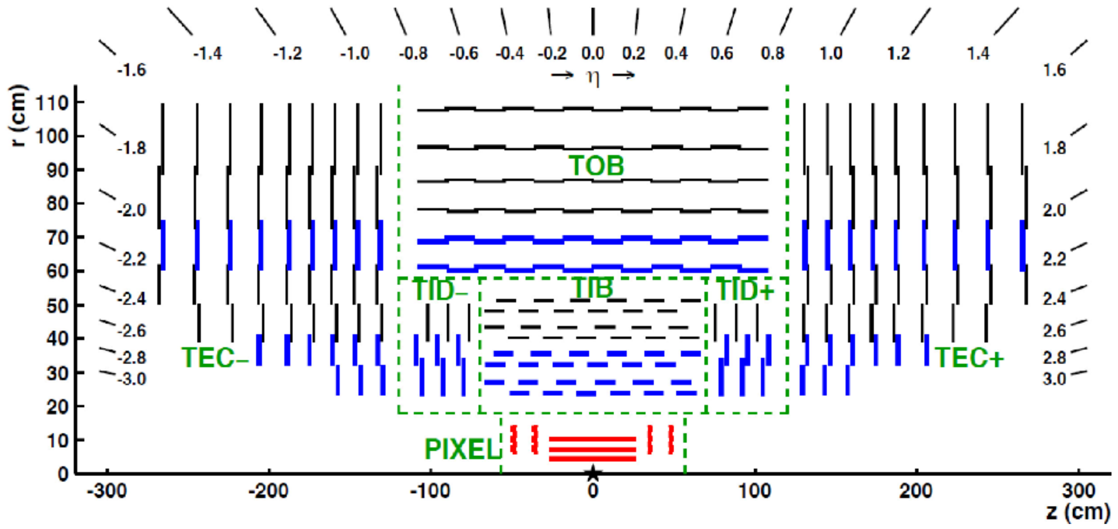


Figure 3.5: This Figure shows schematic of the upper half of the CMS Tracker in the *r*-*z* plane adapted from [22], illustrating the positions of the pixel (red) and strip (blue).

radiation. Dynamic inefficiencies like decreasing hit efficiency with increasing instantaneous luminosity, increased fake rates, reduced resolution in the inner layers as Pile up (PU) increased in the LHC, were some of the reasons which lead to the Phase I upgrade of the pixel detector. In this thesis we perform analysis of the data from all Run II years (2016, 2017, 2018) so, here we describe both the 2016 (Pixel Phase 0) and 2017/2018 (Pixel Phase I) detectors. Figure 3.6 shows the layout of the 2016 (Phase 0) and 2017/2018 (Phase I) pixel detector and the geometry of the upgraded barrel layers (BPIX) and forward/backward disks (FPIX) detectors.

The 2016 (Phase 0) pixel [24] consisted of three barrel layers (BPIX) at radii of 4.4 cm, 7.3 cm and 10.2 cm, and two forward/backward disks (FPIX) at longitudinal positions of ±34.5 cm and ±46.5 cm and extending in radius from about 6 cm to 15 cm. The BPIX consists of 48 million pixels covering a total area of 0.78 m^2 and the FPIX has 18 million pixels covering an area of 0.28 m^2. These pixels produce a minimum of 3 independent hits along the paths of the charged particles with single hit resolutions between $10 - 20\,\mu\mathrm{m}$.

The 2017 (Phase I) pixel [25] added an additional layer to the barrel layer and additional 4 layers to the endcap region. The layers in barrel region are arranged at radii of 2.9 cm, 6.8 cm,

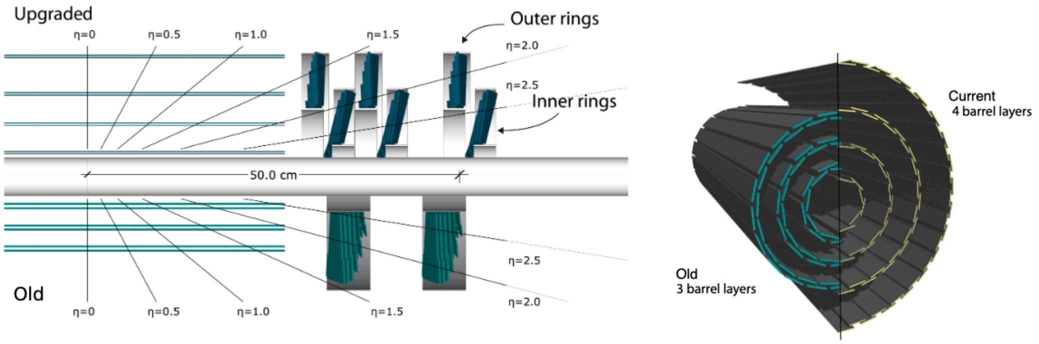


Figure 3.6: [Left]This Figure shows the comparison of the CMS pixel detector in Phase 0 (below center) with 3 barrel layers, 2 endcap disks and Phase 1 with 4 barrel layers, 6 endcap disks, adapted from [23]. The image on the right shows the cross-sectional view

10.9 cm and 16 cm. The innermost layer is made closer to the interaction point by reducing the radius of the beam pipe. The endcap region consisting of 6 layers is organized into three disks with two layers each at a distance of ±29.1 cm, ±39.6 cm and ±51.6 cm from the origin along the z-axis. The number of pixels in BPIX increased from 48 million to 79 million and in the FPIX region, from 18 million to 45 million covering a total area of 2 m^2. The updated tracker allows the CMS to record a minimum of 4 independent hits per track, increasing the efficiency of track reconstruction, reduced fake rates, enhanced resolution of the momentum and impact parameter at high PU.

As shown in Figure 3.7 the silicon sensor is composed of many layers stacked together. Each pixel unit on the sensor is bump bonded to the read-out channel on the readout-chip (ROC). These bump bonds create an electrical connection between the pixel and the readout channel on the chip. The ROC is connected via wire-bonds to the high density interconnect (HDI) for carrying the signal to the front-end electronics. The pixel sensors in the barrel are $n - on - n$ semi-conductors of size 100 µm × 150 µm, giving a spatial resolution of 15 µm-20 µm. There are 4,160 pixels per sensor and 16 sensors per module.

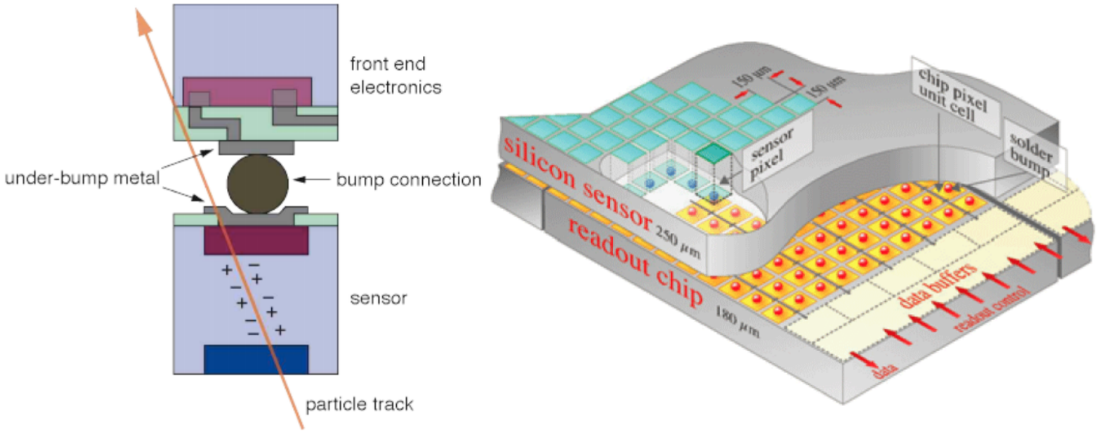


Figure 3.7: [Left] signal read-out mechanism in a typical CMS silicon sensor. [Right] Anatomy of a silicon sensor showing distinct components like silicon bulk, ROC and bump bonds with dimensions.

3.2.3.2 Strip Tracker

The Silicon Strip Tracker (SST) surrounds the Pixel detector [26], the SST consists of a total of 10 layers distributed in two barrels, Tracker Inner Barrel (TIB) and Tracker Outer Barrel (TOB). The barrels are enclosed by wheels grouped into 4 parts, two Tracker Inner Disks (TIDs) with three wheels per side, two Tracker End Caps (TECs) with nine wheels per side. The SST has a diameter of 2.4 m and a length of 5 m. with an active area of 198 m^2 distributed across 24,244 sensors it is the biggest silicon detector to be built. Figure 3.5 illustrates the layout of the strips.

3.2.4 Electromagnetic Calorimeter

The electromagnetic Calorimeter (ECAL) [27] is a homogeneous and hermetic detector which measures the energy of the electromagnetically interacting particles like photons, electrons and positrons. It consists of two main components as shown in Figure 3.8- the barrel region composed of 61,200 lead tungstate ($PbWO_4$) crystals covering the region in $|\eta| < 1.479$, and the two endcap regions composed of 14,648 crystals covering the region $1.479 < |\eta| < 3$. Avalanche photodiodes (APDs) are used in the barrel region and vacuum phototriodes (VPTs) are used in

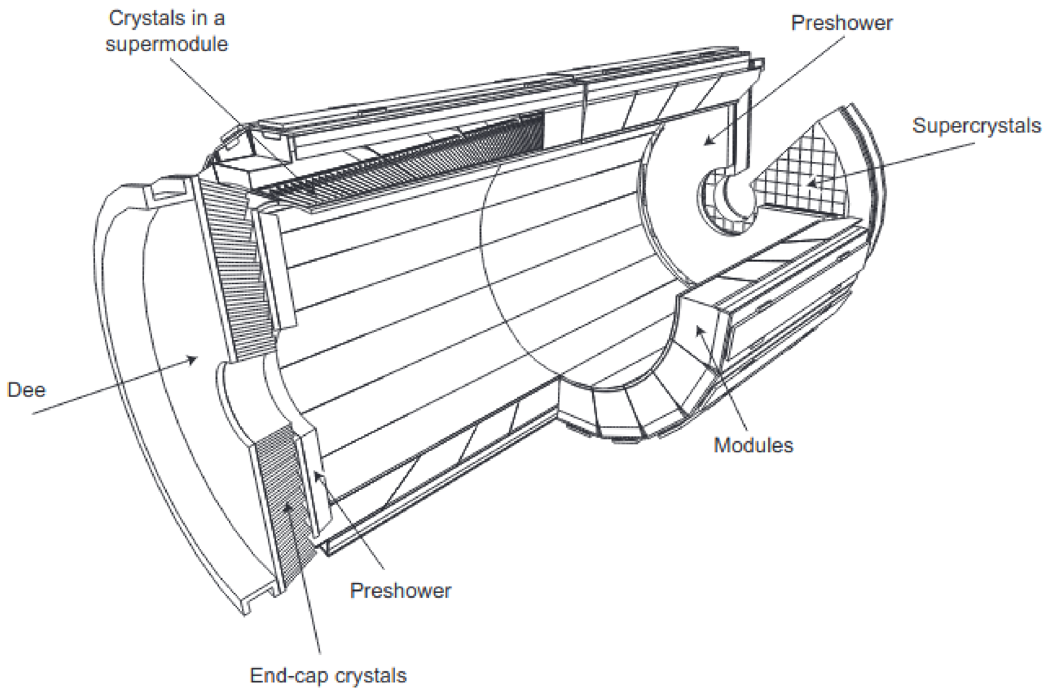


Figure 3.8: Schematic represntation of the CMS ECAL components adapted from [27]

the endcap region. One of the main motivations in the design of the ECAL was its capability to detect two photons coming from the Higgs boson decay. This capability is enhanced by the good energy resolution of the ECAL.

The choice of $PbWO_4$ crystals is motivated by the following reasons. Its short radiation length (0.89 cm), small *Moliére* (2.19 cm)and high density (8.27 gcm^{-3}) results in high resolution and compactness of ECAL. It's a fast scintillator (meaning 80% of the scintillation light is emitted within 25 *ns* of the particle's incidence on the crystal). The $PbWO_4$ crystals in the endcap have a truncated pyramidal shape which makes the light collection non-uniform along the crystal length. This non-uniformity is corrected by depolishing one of the lateral face of the crystal as shown in figure 3.9.

When an electron/photon traverses the ECAL crystals, it results in a cascade of electromagnetic interactions, producing a shower of particles such that the sum of there energy is

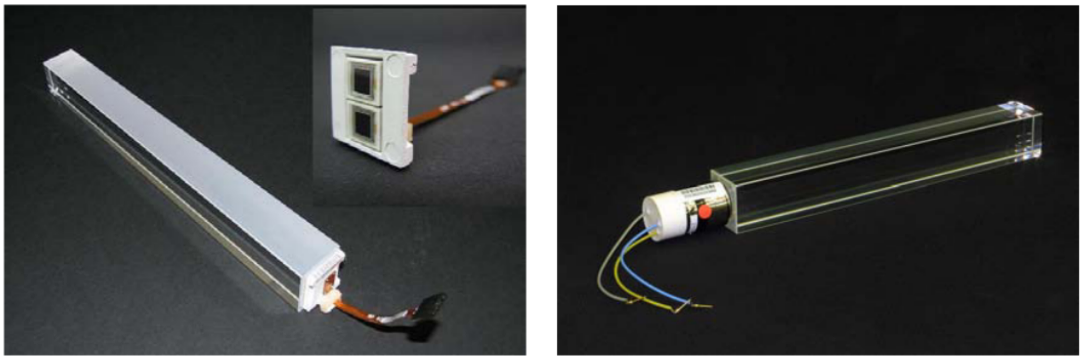


Figure 3.9: $PbWO_4$ crystals with photodetectors attached. Left panel: A barrel crystal with the upper face depolished and the APD capsule. In the insert, a capsule with the two APDs. Right panel: An endcap crystal and VPT

proportional to the energy of the incident particle. Inside the ECAL these particles release photons which are detected by the photo-detectors at the end of the crystals. APDs (with an active area of $5\,\mathrm{mm} \times 5\,\mathrm{mm}$) in the barrel (two APDs per crystal) and VPTs (with an active area of $280\,\mathrm{mm}^2$) in the endcap (one VPT per crystal) glues at the back of the crystal as shown in Figure 3.9.

The energy resolution for particles with energy below 500 GeV, the shower leakage from the rear of the calorimeter becomes significant. It is parametrized in 3.4

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \tag{3.4}$$

where S is the stochastic term, N the noise term and C the constant term. The contribution to the stochastic term is mainly due to event-to-event fluctuations and pre-shower energy deposit fluctuations. The contribution to the noise term is due to electronics and pileup noise. Finally, the constant term is due to the non-uniform longitudinal light collection. The experimental values of these parameters were found to be 2.8%, 0.12% and 0.30% for the first, second and third terms respectively. See 3.5

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12\%}{E}\right)^2 + (0.30\%)^2\,, \tag{3.5}$$

3.2.5 Hadronic Calorimeter

The hadronic calorimter (HCAL) [28] is located outside the ECAL, it is responsible for the measurement of the hadronic jets, and neutrinos or exotic particles resulting in apparent missing transverse energy. The hadronic jets are initiated by the charged and neutral hadron particles like kaons, pions and baryons.

HCAL is sampling calorimeter. It finds a particle's position, energy and arrival time using alternating layers of brass absorbers and tiles of plastic scintillator. When a hadronic particle hits the brass absorber layer it results in the production of many secondary particles. These secondary particles further interact with the absorber layer to create shower of particles. As this shower develops, the particles pass through the alternating layers of active scintillation material causing them to emit blue-violet light. Unlike ECAL, the particles are not absorbed in the scintillator hence the energy of the particles is sampled; therefore the total energy of the hadrons is not measured but estimated from energy clusters resulting in low resolution of the detector.

The HCAL is organized into 4 sections in a longitudinal view as shown in Figure 3.10. It consists of hadron barrel (HB), hadron outer (HO), hadron endcap (HE) and hadron forward (HF). HB covers $|\eta|< 1.4$ region. HE covers a substantial portions in the rapidity range $1.3 < |\eta|< 3$, this region receives around 34% of the particles produced in the final state. In the central pseudorapidity region, the combined stopping power of EB plus HB does not provide sufficient containment for hadron showers. So, HO is placed outside the solenoid to cover the $|\eta|< 1.3$ region and measure the energy tails of the particles passing the HB and the magnet. The HF, made of quartz fiber scintillator and steel as absorption material, covers the forward region $3.0 < |\eta|< 5.2$.

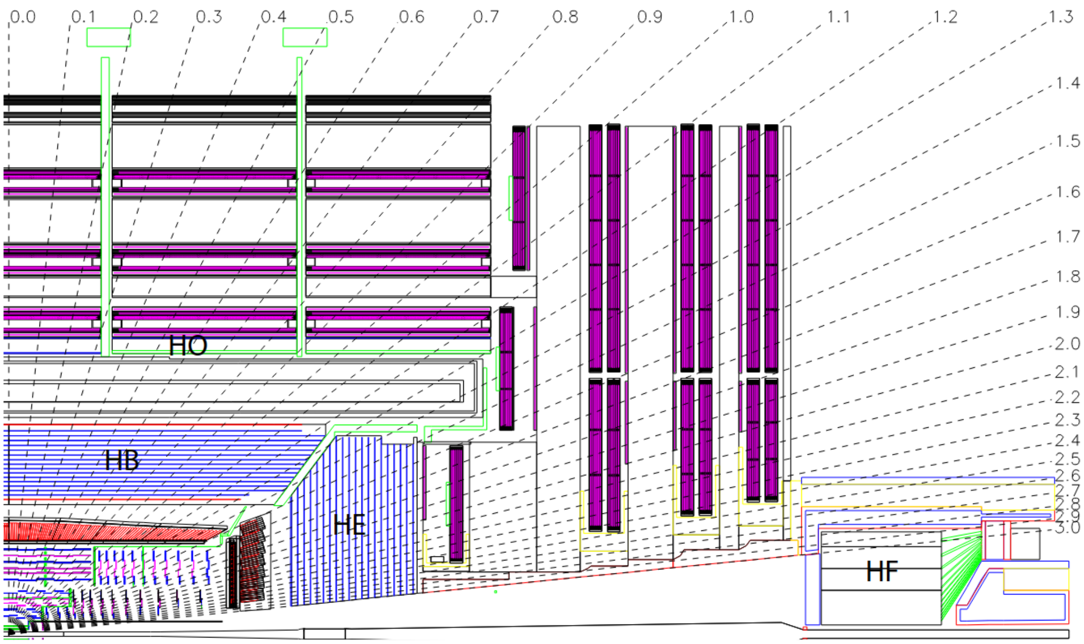


Figure 3.10: Layout of CMS's hadronic calorimeter. All components are shown in this r-z projection. The coverage of the hadronic calorimeter is roughly $|\eta| < 4$. The dashed lines are at fixed η values

3.2.6 Solenoid Magnet

The solenoid magnet at the CMS experiment completely surrounds the calorimeters (ECAL and HCAL) and the trackers (Silicon and Pixel). It has a length of 12.5 m and a diameter of 6.3 m composed of 4 layers of NbTi coils reinforced with aluminum conductors. The magnet is cooled down to a temperature of 4.5K using liquid helium to achieve superconductivity and produce a magnetic field of 3.8 T along the direction of the beam. This magnetic field is $O(5)$ greater than Earth's magnetic field (10^{-5}T). The solenoid has a total mass of 220 tonnes and draws a current of 19 kA and stores an energy of 2.6 GJ making it the world's strongest magnet system.

The majority (67 %) of the magnetic flux of the solenoid is returned to a steel return yoke of mass 10,000 tons. While the rest of the magnetic flux creates a stray magnetic field. The yoke layers are scattered within the muon chambers, this allows to stop particles other than the muons to get through to the muon chamber, thus ensuring high purity muon samples.

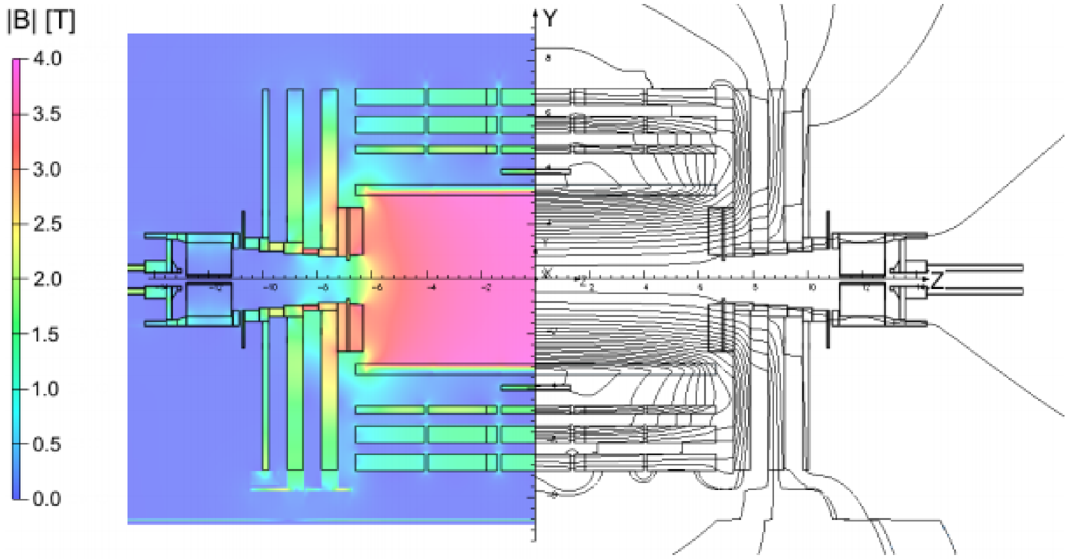


Figure 3.11: Layout of magnetic field produced by the solenoid in the CMS [29]. The center region of the detector shows a uniform magnetic field of 3.8 T, the solenoid was originally designed to reach a peak magetic field of 4T but it was slightly reduced to 3.8T to extend the life of the solenoid

3.2.7 Muon System

The Muon system consists of the outer-most layer of the CMS detector. Its purpose is to detect muons which are charged particles like electrons and positrons with the only difference being the mass of muon is 200 times. The primary reason for detecting muons is that, we expect the them to be produced as decay products of new particles such as the Higgs decaying into four muons. One advantage in the detection of muons in the muon system is that the most of the particles other than muons are stopped in the inner layers of the CMS detector, this results in lower background for identifying muons.

The Muon system consists of three sub-detectors- 250 drift tubes (DTs), 540 cathode strip chambers (CSCs) and 610 resistive plate chambers (RPCs). DTs and RPCs are arranged in concentric cylinders around the beamline while the RPCs and CSCs makeup the endcaps, which cover the barrel part as shown in figure 3.12.

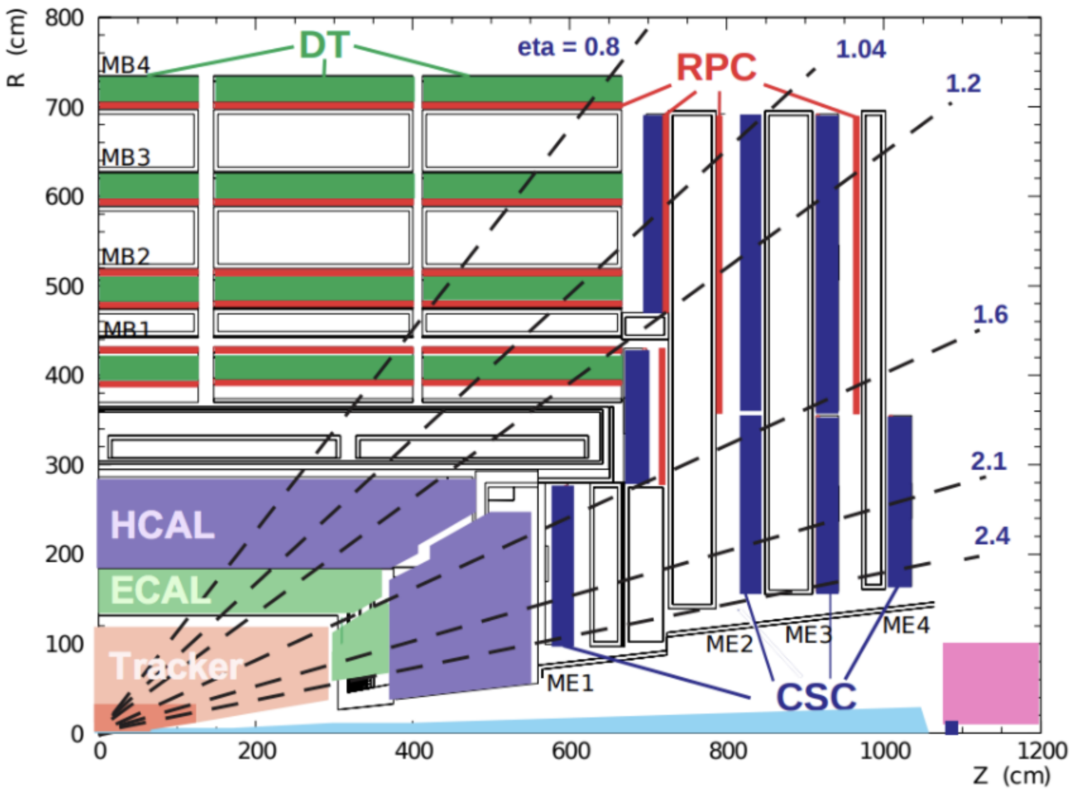


Figure 3.12: Layout of CMS's Muon showing the DTs, CSCs and RPCs fitted between the layers of the return yolk. All components are shown in this r-z projection. The coverage of the muon system is roughly $|\eta| < 2.4$. The dashed lines are at fixed η values

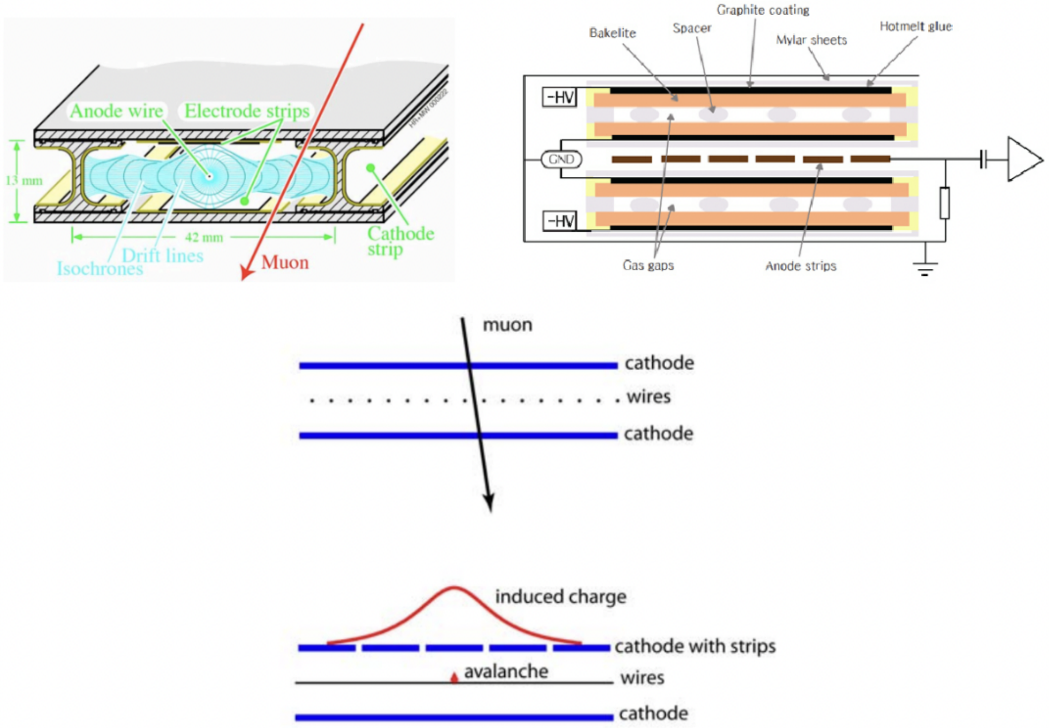


Figure 3.13: Top left diagram shows the mechanism of passing muons through DT ionizes the gas atoms producing electrons. These electrons move in the presence of electric field thus producing a signal for the presence of muons.

The DTs [27] measure the muon position in the barrel part of the detector in the $|\eta| < 1.2$ region. They consist of 4 barrel layers parallel to the beam axis placed alternatively between the return yolk layers of the magnet. Each DT layer is formed by a 4 cm wide Aluminum tube containing a stretched wire in a gas volume. The wire is maintained at a potential difference, this potential difference creates an electric field which attracts the electrons from the ionized gas atoms. When a muon passes the DT it ionizes the gas atoms producing electrons, these electrons drift in the presence of electric field towards the wires and the signal produced by these electrons acts as a proxy for the muons. This mechanism is illustrated in Figure 3.13

The cathode strip chambers are used in end-caps covering a region of $0.9 < |\eta| < 2.4$ which corresponds to the region of uneven magnetic field and high particle rates. CSCs provide precise space and time information about the presence of muons. CSCs consist of arrays of positively charged wires crossed with negatively charged copper cathode strips within a gas volume oriented perpendicular to each other. When muons pass through, they ionize the gas atoms and produce electrons, which travel to the anode wires creating an avalanche of electrons. Positive ions move away from the wire and towards the copper cathode, also inducing a charge pulse in the strips, at right angles to the wire direction. Since the strips and the wires are perpendicular to each other, we get two position coordinates for each passing muon.This mechanism is illustrated in bottom plot of Figure 3.13

The RPC chambers are fitted between alternating layers of DT and CSC layers, they prove a fast measurement of the muon momentum used to trigger event selection. The RPCs also use the same mechanism as the CSC and DT and combine a good spatial resolution with the time resolution of one nanosecond. The mechanism is illustrated in right diagram Figure 3.13

3.2.8 Trigger and Data Acquisition System

On an average there are 40 Million collisions per second or to say 40 Million events every second. While the size of 1 Event is around 1 MB. Doing some quick math tells us that we'll be producing 40 TB of data every second (40 TB/s). That is an insane amount of data bandwidths. Given the detector is running for several days without stops would produce an unprecedented amount of data. While not all data are interesting and limitations imposed by storage and data transfer bandwidths, we need a trigger system which selects events which may contain some new physics information and discard other non-interesting events. CMS has a very advanced trigger system to select and store these events from the pp collisions at the LHC. This system has two layers, a hardware-based Level-1 trigger and a software-based High-level trigger [30].

3.2.9 Level-1 trigger

The Level-1 (L1) trigger is a fast hardware trigger that makes a decision to accept or reject an event in less than 4 μs. This system uses coarse information from a subset of CMS detectors:

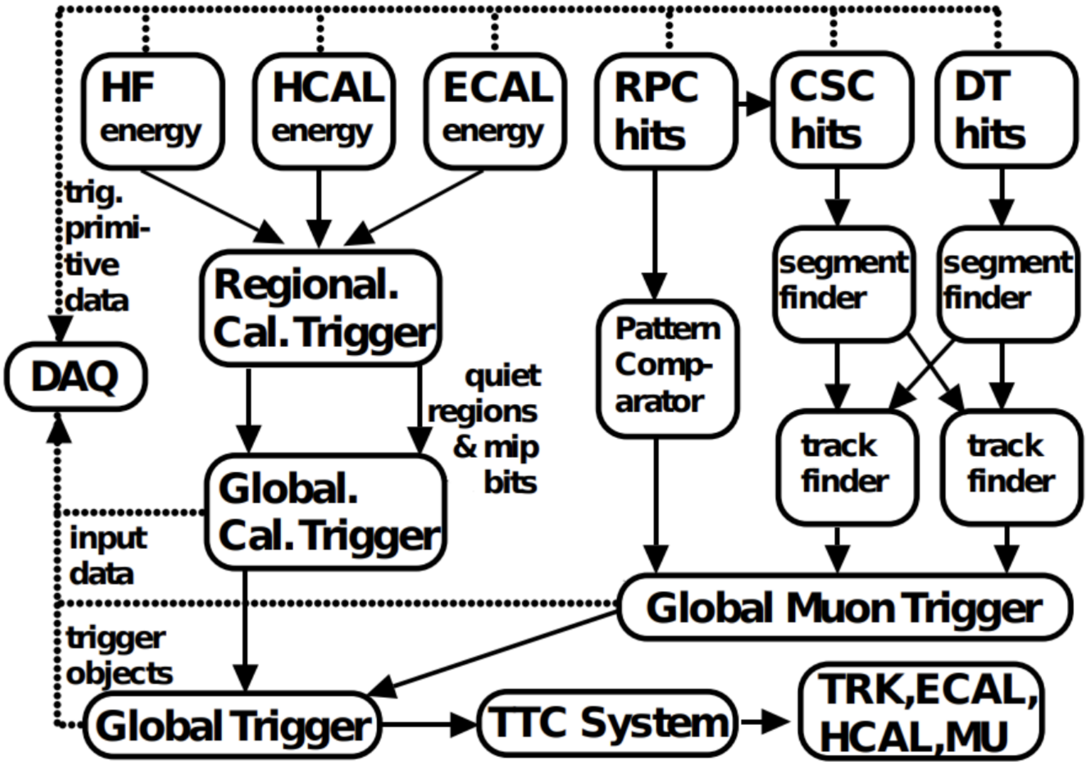


Figure 3.14: CMS's L1 trigger system [30]. The main data processed by the lower-level trigger are calorimeter and muon information only since it must be very fast.

ECAL, HCAL, and the muon chambers. The result of the L1 trigger is distributed to all CMS detectors and an accept decision triggers the readout of all 55 million channels. The L1 trigger system reduces the event rate from40 MHz to a more acceptable bandwidth of 100 kHz.

Figure 3.14 shows how organization of the L1 trigger system. The energy information from the calorimeters and the hit information from the muon chambers is processed in first. Here the regional information from the calorimeters is used to construct photon/electron candidates using the energy deposit measurements. These objects are then passed to the global calorimeter trigger where clusters of energy are grouped into jets and identification of electrons and photons. These objects are then passed to the global trigger where the final decision is taken by combining the information from the muon system.

Similary, the muon chamber hits are reconstructed as muon paths using several processing algorithms. The hit segments are combined to form tracks using the track-finding algorithm. Results from all the muon detectors are combined into the global muon trigger. Finally, the global trigger combines the information from the global muon and calorimeter triggers to make a global accept/reject decision on an event.

3.2.10 High-level trigger

It is a software based trigger which uses very sophisticated algorithms to produce an output rate of 100 Hz from an input of 100 KHz, amounting to a reduction factor of 1000. It uses the entire information available from the detector from tracking to reconstruction to identification and apply a selection criteria on the events. HLT then communicates its decision to the Event Filter Farm (13,000 processors) to tell them which event passed the HLT criteria. The HLT has a menu of trigger paths that is configurable at runtime. A trigger path is a set of filters on several physics objects that the event must satisfy. If an event passes at least one full trigger path, it is stored permanently for offline reconstruction.

CHAPTER 4

The Upgrade of the CMS detector for High-Luminosity LHC

4.1 Why are we upgrading the LHC and the CMS detector?

The High Luminosity Large Hadron Collider (HL-LHC) [31] is an upgrade to the Large Hadron Collider (LHC) to start operation in 2027. It will increase the design luminosity of the LHC from 1.5-1.9 × $10^{34}\ cm^{-2}\ s^{-1}$ to 5 × $10^{34}\ cm^{-2}\ s^{-1}$ also reaching a peak luminosity of 7.5 × $10^{34}\ cm^{-2}\ s^{-1}$, this will allow us to collide more proton bunches and thus generating more data for analysis. The size of data collected by the HL-LHC in a year would be larger than all the data that the LHC has generated since 2010, the HL-LHC is expected to deliver an integrated luminosity of 300 fb^{-1} in a year and a total luminosity of 3000 fb^{-1} over ten years of its expected life . This new large dataset will allow researchers to explore new energy domains and explore new physics phenomena which were previously not accessible, study the accessible ones with higher precision and sensitivity. However there are many challenges in achieving the upgrade.

- The upgrade depends on many state-of-the-art technologies that need to be applied to the HL-LHC. Some of them include the high-field superconducting radiation tolerant magnets, compact and high precision RF cavities, vacuum and cryogenics.
- With the increase in the p-p collision rate by a factor of 10, we expect increase in trigger rate 6.3, pileup 6.2.1 and data transfer and storage bandwidths.

The main goals of the HL-LHC are:

- High precision measurements of the properties of the higgs boson and the longitudinal components of the massive vector bosons. The main aim being able to precisely understand if there are any deviations from the standard model predictions.

- At higher energy domains see if the higgs boson is accompanied by particles which could be potential dark matter candidates.

4.2 CMS Tracker Upgrade

The above challenges are not just central to the LHC but the CMS detector too. With the increase in the number of collisions, the CMS detector will see increased pileup rates (25 → 200), increase number of particle hits in the detector sensors upto 3.2 GHz cm^{-2}, increase in the number of electronic read-out of the signal, increased trigger rates (100 → 750 kHz), increased radiation (300 → 3000 fb^{-1}), higher operational temperatures and increased power load. Such conditions pose a serious challenge for consistent and effective operation of the existing detectors thus necessitating the need for the upgrade the CMS detector to exploit the increase in luminosity provided by the HL-LHC [32], although there are many sub-detectors 3.2.3 within CMS, we are going to discuss the upgrades related to the tracker 3.2.3 and in particular the pixel sub-detector (also called the Inner Tracker) only, as most of the research and development done at the Catholic University's High energy physics detector lab was on the Pixel detector.

4.2.1 Limitations of the current tracker

From Fig 4.1 shows a quarter section of the CMS phase I tracker in the r-z co-ordinate frame. It consists of two main sections, the pixel and the strip, the pixel is below the 200 mm in the radial (y-axis) direction marked in green color and below 500 mm in the z-direction while beyond 200 mm in the radial direction is the strip made from single sided and double sided modules marked in red and blue respectively. The pixel detector has already been updated once in 2016/2107 due to a DC-DC converter issue [33] thus naming it as 'Phase I' pixel. Both the pixel and strip parts of the tracker are sensitive to operational conditions of the LHC so with increase in the luminosity the detectors would suffer from significant damage and performance degradation [32]. Radiation can cause damage to the pixel sensors, leading to a decrease in their

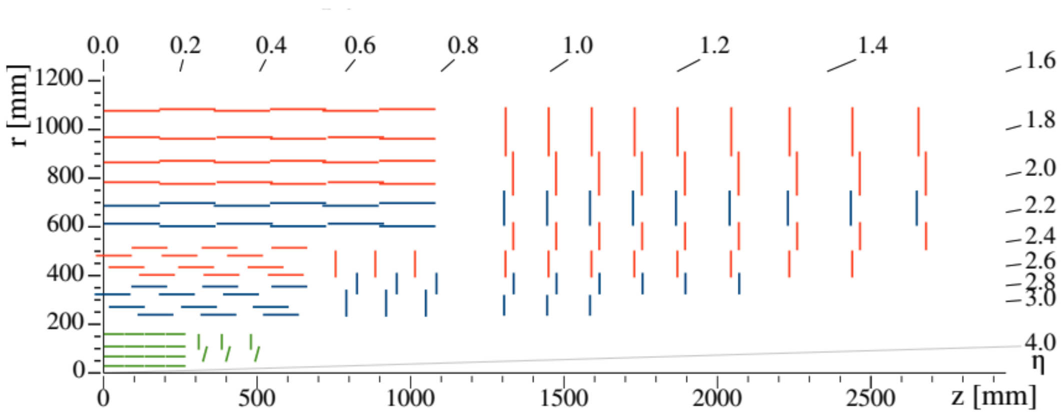


Figure 4.1: This Figure shows the 1/4th section of the CMS Phase I tracker in the r-z co-ordinate frame. The green makings show the pixel detector and the red and blue show the single-sided and double-sided modules in the strip tracker.

ability to collect charge and also the Lorentz angle. This can result in a decrease in the charge sharing ability of the pixels, causing the spatial resolution to deteriorate, leading to a reduction in measuring the hit efficiency of the pixel sensors. The strip sensors are affected by increase in the sensor depletion voltage and of the leakage current, the increase in leakage current can be controlled by operating a lower temperatures but the sensor depletion voltage issue would still exist. As particle tracking and reconstruction from the tracker are central to the physics analysis, the deterioration in the performance of the tracker would result in poor quality data, making the data not very reliable for physics explorations.

4.2.2 Why is the upgrade required?

The upgrade of the CMS tracker would result in the performance because of the following features:

- Radiation tolerance: Radiation tolerance of the detector components would be key for excellent performance of CMS. With the increase in luminosity the number of particles incident on unit area (Fluence) of the sensor is increases significantly causing a very dosage of radiation. According to simulation results the different regions of the

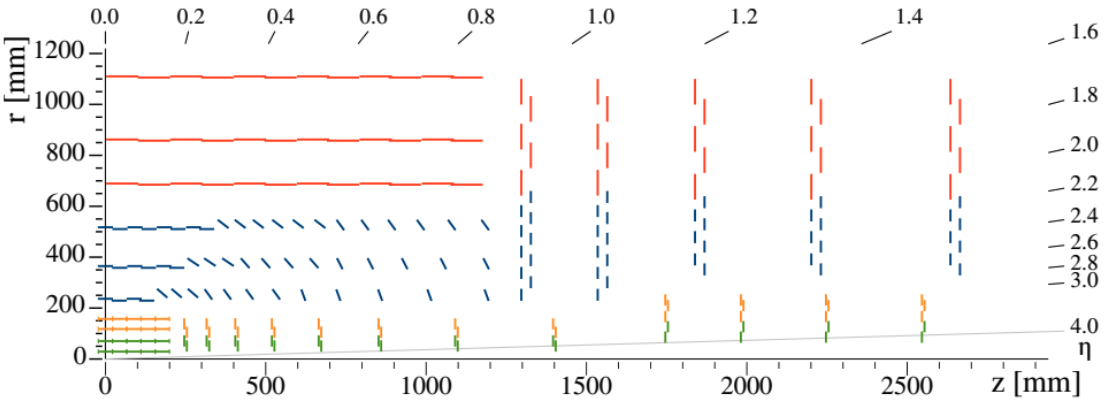


Figure 4.2: This Figure adapted from [32] shows the 1/4th section of the CMS Phase II tracker in the r-z co-ordinate frame. The green makings show the pixel detector and the red and blue show the single-sided and double-sided modules in the strip tracker.

detector are expected to see an increase of one order in magnitude during the HL-LHC nominal run. The current detector components would be incapable of handling such fluence levels of the particles so the CMS detector must be upgraded.

- Increased Granularity: High pileup upto 200 is expected in the HL-LHC run, so, to ensure that the tracking performance of the detector is high we need high channel density for the detector to keep up with larger number of particle that would be intercepted.
- Improved two-track separation: In the current tracker, due to hit merging the track finding performance for highly energetic jet is limited. At high luminosity this performance needs to be improved.
- Reduced Material budget in the Tracker: The presence of the amount of material in the detector called material budget affects the performance of the detector. The material radiates energetic radiations which gets recorded as energy deposits causing poor performance of the calorimeters and the CMS detector overall.
- Extended Tracking coverage: With an extended coverage of $\eta = 4$ would increase the acceptance in the forward region of the tracker and the calorimeters.

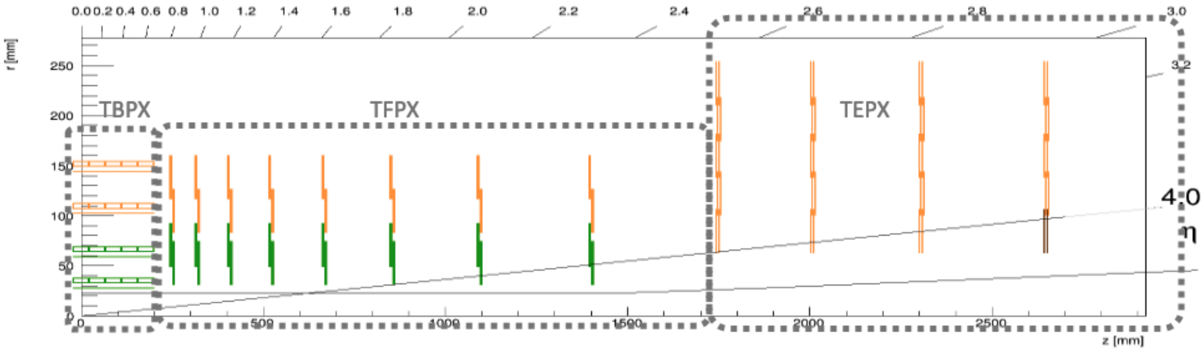
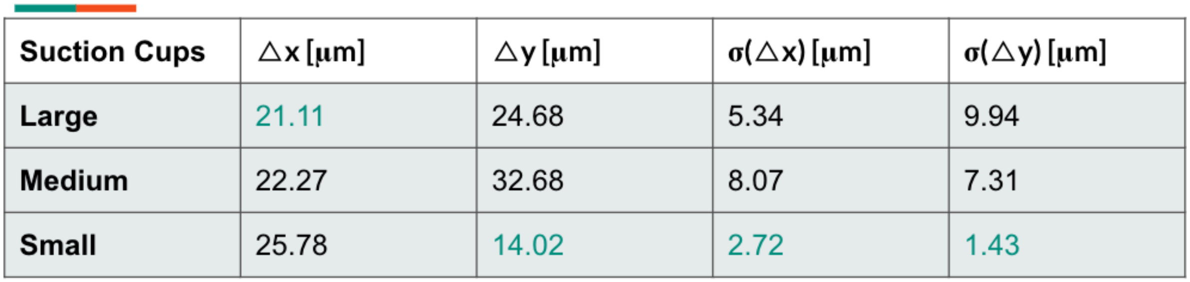


Figure 4.3: This Figure adapted from [32]shows the quadrant of the CMS Phase II Inner tracker in the r-z co-ordinate frame. The green and orange markings show the pixel detector. The green markings are the pixel modules with two read-out chips and the orange ones are the one with four read-out chips. TBPX is the Tracker Barrel Pixel, TFPX is the Tracker Forward Pixel and TEPX is the Tracker Endcap Pixel.

4.2.3 Upgraded Inner Tracker

The inner tracker is the first line of defense to intercept the shower of particles from the p-p collisions. The upgraded inner tracker has to cope up with the effects of increased luminosity. Some of the notable effects include radiation doasage of 1.2 Grad, a hadron fluence of upto 2.3 $\times 10^{16}$ neq/cm^2, a pileup of 140-200 collisions per bunch crossing, the design also includes the ease of replacement of the degraded components.

Fig 4.3 shows the Upgraded inner tracker in the r-z plane with a η acceptance of 4. It consists of three components: TBPX, TFPX and TEPX. TBPX stands for Tracker Barrel Pixel, It consists of four layers, the inner two layers have are modules with two read-out chips (ROC) and the outer two layers are made of pixel modules with four read-out chips. TFPX stands for Tracker Forward Barrel Pixel, it consists of eight layers, with a similar two-ROC and four-ROC combinations. Then we have the TEPX which stands for Tracker Endcap Pixel consisting of four layers with a similar ROC setup. Fig 4.8 shows the inner structure of the arrangement of modules in different layers.



Suction Cups	△x [μm]	△y [μm]	σ(△x) [μm]	σ(△y) [μm]
Large	21.11	24.68	5.34	9.94
Medium	22.27	32.68	8.07	7.31
Small	25.78	14.02	2.72	1.43

Figure 4.4: This table shows comparison of the placement deviations from the expected and observed position of the x and y coordinates when placing and object from one position to a target position.

4.2.4 Pixel RD53A module assembly at Catholic University

The Catholic University has a mechanical gantry setup for assembling the modules with micro-meter precision. The gantry is `AGS15000 Series Gantry` see fig 4.6 custom fabricated by Aerotech [34], it offers a complete 3-dimensional movement across the dimensions of the gantry and a 360 degree rotational motion the head of the gantry.

As mentioned before, because of increase in luminosity, elevated hit occupancy of 3.5 GHz $cm^{-2}$is expected in the pixel sensor. This demands the development of a new read-out chip (ROC) with higher granularity of 50 × 50 μm^2, increased power demand while also keeping the less amount of material inside the detector, and a trigger latency of 12.4 μs. Fig 4.5 shows the 3D rendering of the RD53A module [35], the assembly of which is shown in the Fig 4.7.

Prior to assembling the module we performed studies for finding the movement precision of the gantry using different size of suction cups. Fig 4.4 shows that the small sized suction cups have the smallest standard deviation in the measurements, so we used the small sized suction cups for assembling the module.

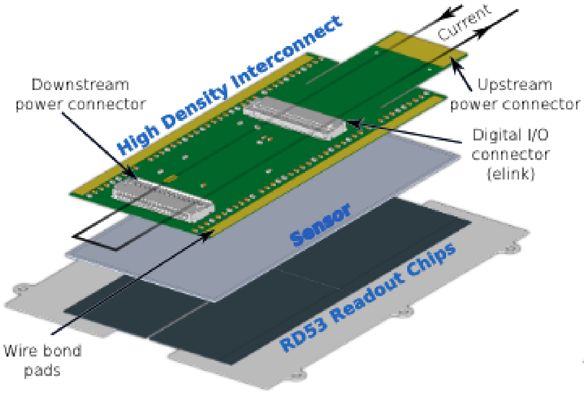


Figure 4.5: This figure shows a 3-dimensional representation of the RD53A module adapted from [32]

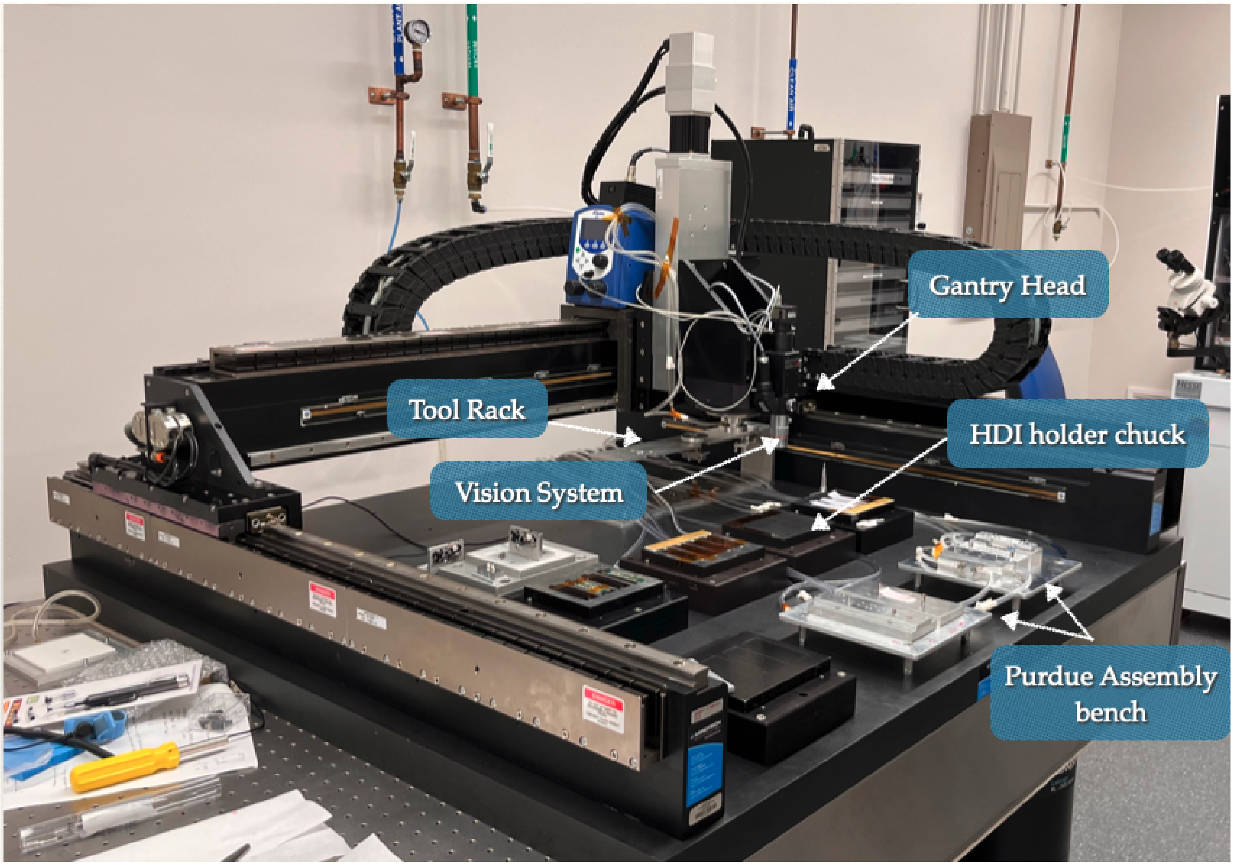
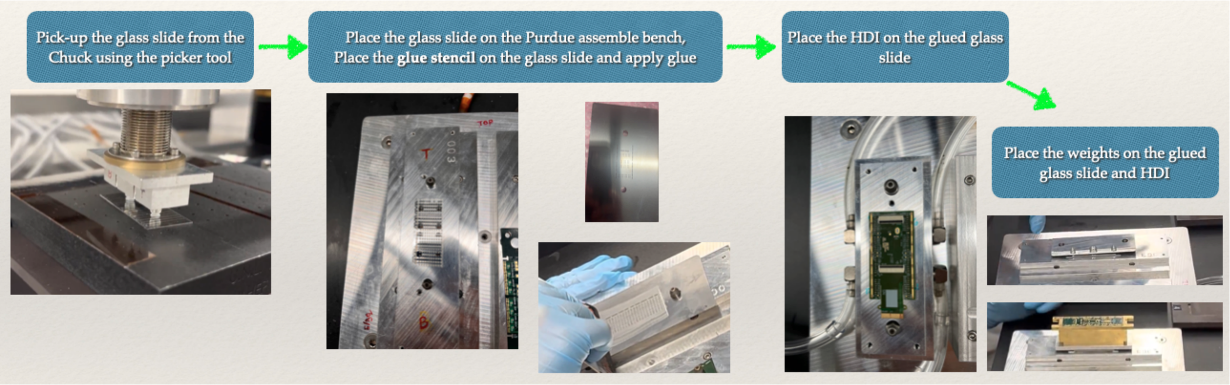


Figure 4.6: This Figure shows the Aerotech Gantry setup at the Catholic university, The RD53A module is assembled with the help of the components marked above.



Pick-up the glass slide from the Chuck using the picker tool

Place the glass slide on the Purdue assemble bench, Place the **glue stencil** on the glass slide and apply glue

Place the HDI on the glued glass slide

Place the weights on the glued glass slide and HDI

Figure 4.7: This Figure shows the steps for assembling the RD53A module, First step is moving the glass slide from the chuck to the purdue assembly bench and then the glue stencil is used to apply the glue then we place the HDI on top of the glued glass slide and leave it for curing overnight

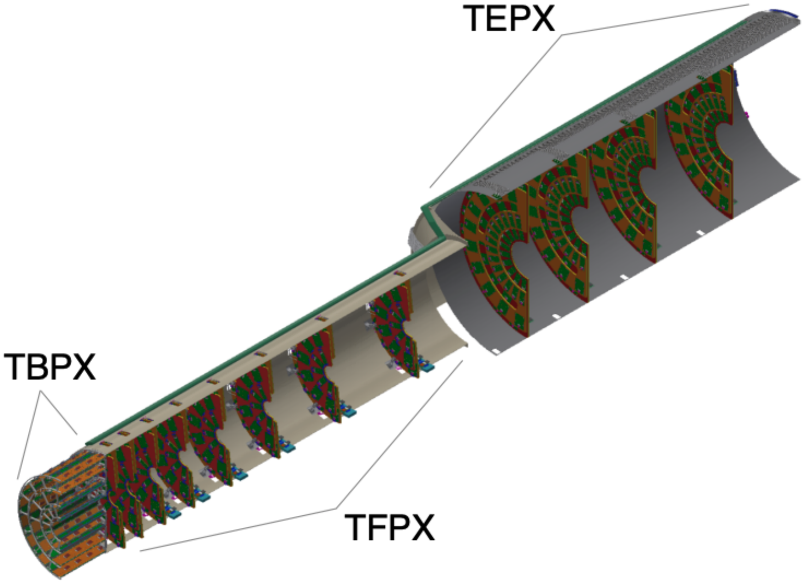


Figure 4.8: This figure shows the view of one quarter of the Inner Tracker, showing the TBPX ladders and TFPX and TEPX dees inside the supporting structures. The pixel modules are shown as orange elements in TBPX and as green elements in TFPX and TBPX. The dees are depicted as red and orange surfaces, adapted from [32]

CHAPTER 5

Theoretical Foundations of Monotop Analysis

This Chapter will focus on the explaining the theory behind the monotop analysis and the importance of the top quark.

5.1 The Top quark

The top quark was discovered at the Fermilab Tevatron almost 25 years ago. It is the weak isospin partner of the bottom quark. Its discovery resulted in completion of the three generation structure of the standard model (SM). The top quark mass was measured to be $m_t = 176 \pm 13$ GeV, a charge of 2/3 times the charge on the electron and a lifetime of $5 \times 10^{-25} s$ making it the heaviest of all the fermions known till date in the standard model. Because of it's large mass and correspondingly short lifetime it behaves differently than other quarks.

The top quark decays before it hadronizes, passing its information to the decay products. Therefore making it possible to infer its properties from the decay products in the detector. The top quark decays into a a bottom quark and a W boson, and since the W boson is an unstable particle it decays into a quark anti-quark pair of different flavors (hadronic decay channel) or to a charged lepton and a neutrino (leptonic decay channel) with the following branching ratios.

Table 5.1. W boson decay modes

Decay channel	Branching ratio
$W^{\pm} \rightarrow q\bar{q}'$	68.32%
$W^{\pm} \rightarrow e^{\pm}\bar{\nu}_e$	10.46%
$W^{\pm} \rightarrow \mu^{\pm}\bar{\nu}_{\mu}$	10.50%
$W^{\pm} \rightarrow \tau^{\pm}\bar{\nu}_{\tau}$	10.75%

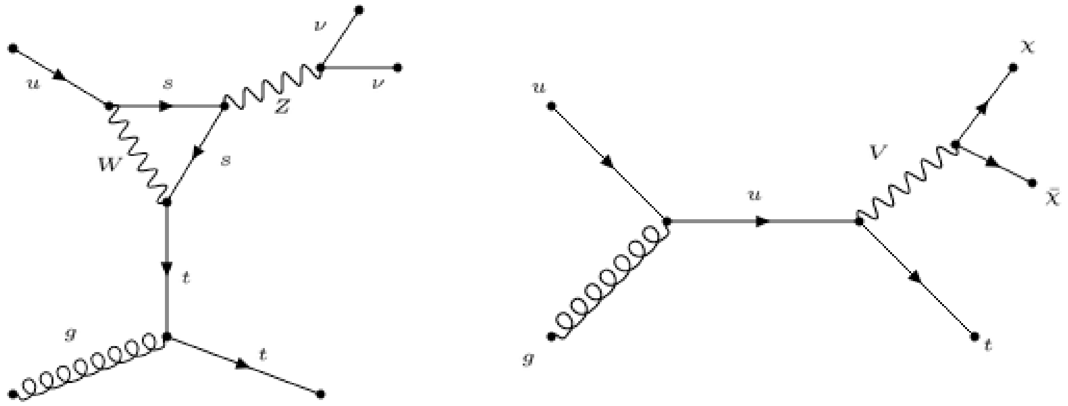


Figure 5.1: [Left]: Shows the tree-level Feynman diagram for Standard model monotop production. [Right] Shows the Leading-order Feynman diagram for the production of monotop signatures in the Dark Matter model [36]. It is flavor changing neutral interaction mediated by a hypothetical vector boson V decaying into two Dark Matter candidates χ.

As the top quark decays into a W boson and a bottom quark, it's evident that the hadronic and leptonic decay probabilities of the top quark will be proportional to the hadronic and leptonic decay probabilities of the W boson. So, it can be implied from the above table that, around 70% of times the top quark decays into three quarks (two quarks from W boson and one bottom quark) which further hadronize to produce jets. While around 30% of the times the top quark would decay leptonically to a lepton, its corresponding lepton neutrino and a bottom quark. In this thesis we will be exploring the leptonic channel for the search for Dark Matter (DM), so our final state signature would a lepton, its corresponding neutrino and a jet coming from a bottom quark.

In accordance with the SM, at the LHC, top quarks are predominantly produced in pairs ($t\bar{t}$) through strong interactions and as single top via the electroweak interaction. In the following section we will be looking at the production of a single top quark (Monotop) in association with missing transverse energy due to the two DM candidates, as an extension to the SM.

5.2 The Monotop Model

This section details the theoretical motivation behind the monotop model [36]. The monotop model is based on an alternative approach where a final state signature is proposed that no standard model process can reach at tree-level. The final state signature in our case is a top quark produced in association with missing transverse energy. In SM this production mode is suppressed both by a loop factor and by two powers of non-diagonal Cabibbo-Kobayashi-Maskawa matrix elements [37]. In Figure 5.1 we see that the left image shows the Standard model production of the monotop signature. The loop consisting of the W boson and two strange quarks allows for the production of Z boson and the top quark and if the Z boson decays into two neutrinos we get the monotop signature in accordance with rules of the Standard model. While the right image in Figure 5.1 shows the tree-level production of the monotop signature. To allow for this tree-level production the monotop model suggests two main mechanisms [36]. Either the top quark is produced (resonantly or not) in association with missing energy of a fermionic nature or through a flavor-changing interaction with an invisible bosonic state, as illustrated on the left and right panels of Fig. 5.2, respectively. This physics analysis utilizes the non-resonant monotop model so, we will discuss in detail about the flavor changing neutral interaction model (Non-Resonant model).

5.2.1 Non-Resonant Monotop Model

As shown in the right panel of figure 5.2, we see that a top quark is produced in association with an invisible bosonic field V (Spin 1). The invisible bosonic field couples with the top and light u/c quarks in a flavor changing way. There are two bosonic states associated with this non-resonant model, a bosonic vector field V with spin 1 and a bosonic scalar field with spin 0. These bosonic states are unstable because of their coupling to quarks. Since we are looking for a missing energy signature in the detector there are two ways in which that can be achieved in the monotop model. First, either the bosonic fields (Spin 0/Spin 1) are long lived such that,

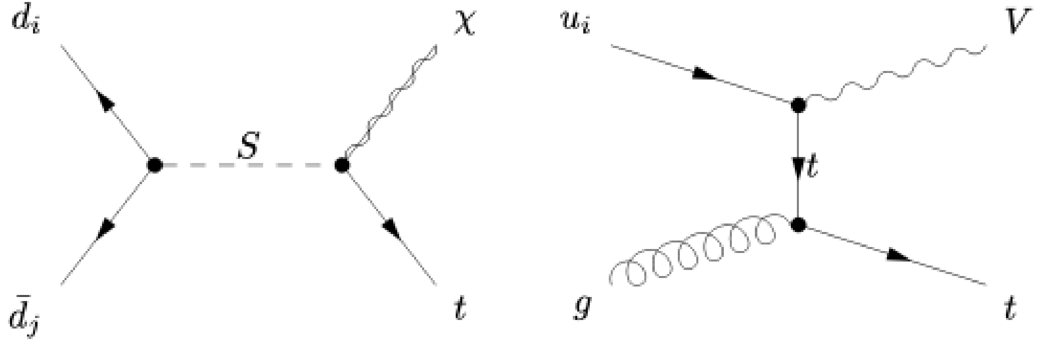


Figure 5.2: [Left] Representative Feynman diagrams leading to monotop signatures, through the resonant exchange of a colored scalar field S and [Right] via a flavor-changing interaction with a vector field V (right). In these two examples, the missing energy is carried by the V and χ particles. The images have been adapted from [36]

they decay outside the detector or, they decay into a pair of neutral stable particles. The monotop model discussed in this thesis would focus on the bosonic state decaying into two stable dark matter candidates. To ensure that the new invisible boson leads to missing energy signature in the detector, leads us to an assumption that bosonic field (V/Φ) decays into dark matter candidates (χ,$\bar{\chi}$). The Lagrangian [38] for this (Flavor changing type interaction) model is

$$\mathcal{L} = \mathcal{L}_{SM} + \mathcal{L}_{kin} + \Phi\bar{u}[a^0_{FC} + b^0_{FC}\gamma_5]u + V_\mu\bar{u}\gamma^\mu[a^1_{FC} + b^1_{FC}\gamma_5]u + h.c. \tag{5.1}$$

$\mathcal{L}_{SM}$ is the Standard Model Lagrangian, which is supplemented by kinetic and gauge interaction terms for all new states included in the Lagrangian $\mathcal{L}_{kin}$. The third and the fourth terms describe the flavor changing associated production of the top quark and the invisible scalar Φ or a vector V, respectively. u represents the up-type quarks (u,c,t) and the strength of the interactions of the two bosonic states (V/Φ) and a pair of up-type quarks is modeled via two 3 × 3 matrices in flavor space $a^{\{0,1\}}_{FC}$ and $b^{\{0,1\}}_{FC}$. The monotop model is also chosen such that flavor changing neutral interaction only relates the up-type quark of the first generation with third generation of quarks. Which implies that all other coupling other than those between the first and third generation of quarks would tend to 0. i.e. $a^{0,1}_{1,3(FC)}$, $a^{0,1}_{3,1(FC)} \neq 0$ and $b^{0,1}_{1,3(FC)}$, $b^{0,1}_{3,1(FC)} \neq 0$, while all other couplings will have zero strength.

So, the main signatures associated with the monotop production can be classified according to the top quark decays.

$pp \rightarrow t + \not{E}_T \rightarrow bW + \not{E}_T \rightarrow bl\bar{\nu}_l + \not{E}_T$ (Leptonic decay mode)

$pp \rightarrow t + \not{E}_T \rightarrow bW + \not{E}_T \rightarrow bjj + \not{E}_T$ (Hadronic decay mode)

In this thesis we will be using the leptonic channel for the search of dark matter candidates. Next section will include a detailed discussion about the vector and scalar invisible bosons.

5.2.2 b-tagging and Deep Jet

Identifying jets produced by b quarks (b jets) is crucial for many physics analyses. In the past, multivariate analysis techniques were commonly used for flavor tagging, but more recently, deep learning algorithms have gained popularity because they can handle complex multi-classification tasks well. In this analysis we have used the DeepJet [39] classification algorithm to select b-tagged jets with high efficiency. In this section we will talk about the technical details of the algorithm and see its efficiency.

deepJet is a machine learning based algorithm in the paradigm of deep learning [40]. It is deep neural network that uses convolutional and long short-term memory layers to analyze properties of up to charged and neutral particle-flow jets and up to 4 secondary vertices associated with each jet. The particle-flow jets and vertices are processed separately using 1x1 convolutional layers, with different numbers of hidden layers used for each type of input. The output of the convolutional layers is then fed into LSTMs and combined with global jet properties, which are first processed through a dense layer before being passed through several hidden layers see fig 5.3. The goal of the algorithm is to analyze the properties of the particle-flow jets and vertices in order to make a prediction or classification.

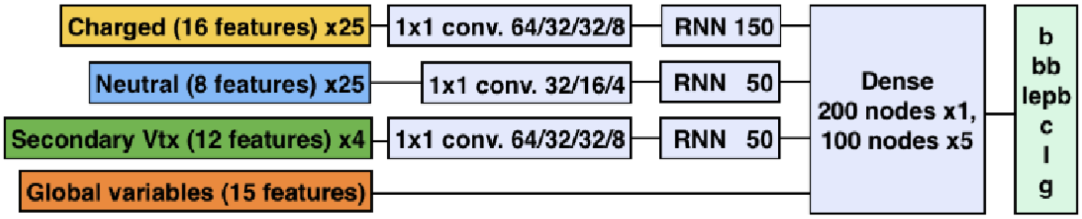


Figure 5.3: this figure shows the scale factors as a function of jet pT obtained from the Kin and the LTSV methods, separately. Adapted from [40]

The DeepJet algorithm is evaluated at three different operating points, referred to as loose, medium, and tight. These points are defined by the values of the discriminator distribution at which the rate of wrongly identifying a light jet as a b jet is 10%, 1%, and 0.1%, respectively. The performance of the DeepJet algorithm on jets with $|\eta|< 2.5$ and $p_T > 30$ GeV from QCD multijet MC is used to determine the operating points. The performance evaluation of the algorithm on the working points is done using two methods: the Kin and the Lifetime tagging secondary vertex (LSTV) methods. The former is used for measuring the b-tagging efficiency in the dilepton top pair events, the selected events contain a oppositely charged lepton pair with $p_T > 25$ GeV and $\geq$ two jets with $p_T > 30$ GeV. The efficiencies and correction factors are calculated by a shape-based fit of a boosted decision tree discriminant that utilises kinematic variables to distinguish b jets from ISR/FSR jets. In the latter method (LSTV) measures the b-tagging efficiency in the multi-jet events, the selected events have a $p_T > 20$ GeV and a muon with $p_T > 5$ GeV within $\delta R < 0.4$ of the jet. Unlike the Kin method, a jet probability tagger is used as a discriminant. Since, we have used the Medium working point for b-tagging ref fig 5.4

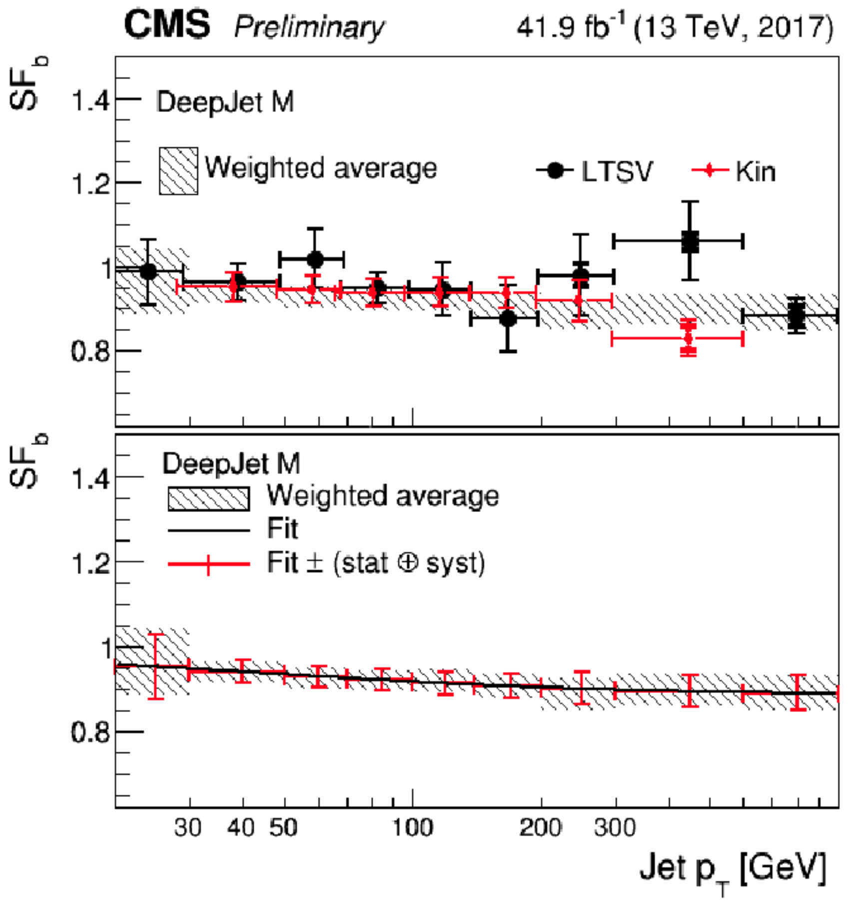


Figure 5.4: This figure shows the scale factors for efficiency of DeepJet algorithm with Medium working point.

5.3 Statistical Analysis Procedure

5.3.1 Fundamental Concepts

Statistics plays a key role in particle physics. The CMS experiment collects data and this data is used either to make an inference about the probabilistic model (searching for a new phenomena) or to estimate the parameters of the model (precise measurements of a parameter of the SM, ex: the mass of the top quark). Such a statistical inference uses two main methods, first, the frequentist method which estimates the probability of an outcome by measuring the frequency of the outcome in a repeatable experiment. However one thing to note is that the frequentist method does not give the probability for a hypothesis or for the value of the parameter. The probability P(X) of an outcome X is defined as

$$P(X) = \frac{\# \, of \; favorable \; outcomes}{\# \, of \; total \; possible \; outcomes} \tag{5.2}$$

Second, the bayesian inference, it expresses one's degree of belief, for instance given a probability distribution function for a parameter, $f(x;\theta)$ we can express our degree of belief about the possible value that parameter can have. Bayesian inference is more like a subjective probability, it requires prior knowledge about the values that the parameter possesses before the experiment. With the usage of Bayes' theorem one can estimate the posterior probability of the parameter having a particular value.

$$P(Y|X) = \frac{P(X|Y)\; P(Y)}{P(X)} \tag{5.3}$$

In Eq 5.3,$P(Y|X)$ represents the probability of occurrence of Y under the assumption that X has happened and vice versa for $P(X|Y)$, $P(X)$ and $P(Y)$ are called priors and represent the general probabilities for the outcome X and Y.

5.3.2 Limit setting procedure and statistical concepts

In order to quantify the fact that the number of observed events is in accordance with the SM predictions or to quantify the presence of monotop signal in the observed data, a statistical procedure is developed which is used by CMS collaboration [41].

In order to quantify our results we use the concept of setting upper limits on the observed cross section of monotop signal (accessed via a proxy called signal strength modifier ($\mu = \sigma_{obs}/\sigma_{SM}$), where σ is the cross section) to exclude the region of the phase space not compatible with the observed data. The reason for the choice of μ for limit setting lies in the fact that it gives us a handle on changing the signal strength by the same factor for all production modes. since, $\mu = \sigma_{obs}/\sigma_{SM}$) and when the expected SM cross section (σ_{SM}) is greater than the observed cross section (σ_{obs}), the model (monotop model in this case) is excluded at confidence level (95%) corresponding to a given significance ($\alpha = 5\%$). A significance of 5% means that that there is a 5% chance of incorrectly rejecting the Null hypothesis (Type I error).

5.3.3 Hypothesis testing and test statistic

To understand the concept of hypothesis testing, we need to first understand what is a hypothesis, why do we use hypothesis testing and how to perform a hypothesis test. A hypothesis is nothing but a statement based on knowledge or a phenomenon that has been observed. For instance, in the context of this analysis we can say we have two hypothesis: a background-only hypothesis that is, our data is composed of only background like events or a signal+background hypothesis, meaning that our data contains a potential signal as described by the monotop model, see sec 5.2. In order to quantify the results of the experiment we use hypothesis testing to either reject or fail to reject the Null hypothesis. We define a Null ($H_0(b)$) and alternate ($H_a(s+b)$) hypothesis and then use a test statistic to evaluate the p-values associated with the test and for a given significance if the p-value $< \alpha$ (0.05) we either reject or fail to reject (p-value$> \alpha$) the Null hypothesis at a confidence level defined by 1-α.

A very important property that a test statistic must possess is the discriminating power. According to Neyman-Pearson [42] lemma, a test statistic that has the maximum discriminating power for a hypothesis at a given significance is given by the likelihood ratio of the hypothesis. The likelihood function is given by eq. 5.4 and eq. 5.5, represent the signal+background and background-only hypothesis, respectively.

$$\mathcal{L}(x|\mu,\theta) = P(x|\mu \cdot s(\theta) + b(\theta))p(\tilde{\theta}|\theta) \tag{5.4}$$

$$\mathcal{L}(x|\mu,\theta) = P(x|b(\theta))p(\tilde{\theta}|\theta) \tag{5.5}$$

Here $P(x|\mu \cdot s(\theta) + b(\theta))$ is defined as the product of probabilities of all the bins in the transverse mass (m_T^W) distribution, which is our discriminating variable. $p(\tilde{\theta}|\theta)$ is the p.d.f. for the nuisance parameters taken into consideration given by eq 5.6, this equation represents the probability distribution of observing n_i number of events in the i_{th} bin product over all bins, where $\mu \cdot s_i + b_i$ is the expected number of events in the i^{th} bin.

$$P(x|\mu \cdot s(\theta) + b(\theta)) = \prod_i \frac{(\mu_i \cdot s_i)^{n_i}}{n_i!} exp(-\mu \cdot s_i + b_i) \tag{5.6}$$

a test statistic q_μ is defined to discriminate between the null and the alternate hypothesis given by eq. 5.7

$$q_\mu = -2\ln\left(\frac{\mathcal{L}(x|\mu \cdot s(\hat{\theta}_\mu) + b(\hat{\theta}_\mu))}{\mathcal{L}(x|\hat{\mu} \cdot s(\hat{\theta}) + b(\hat{\theta})})\right) \tag{5.7}$$

eq. 5.7 is the profile likelihood ratio where x represents the data , μ is the signal strength modifier, the numerator is called the conditional likelihood where $\hat{\theta}_\mu$ is the value of θ for a given μ which

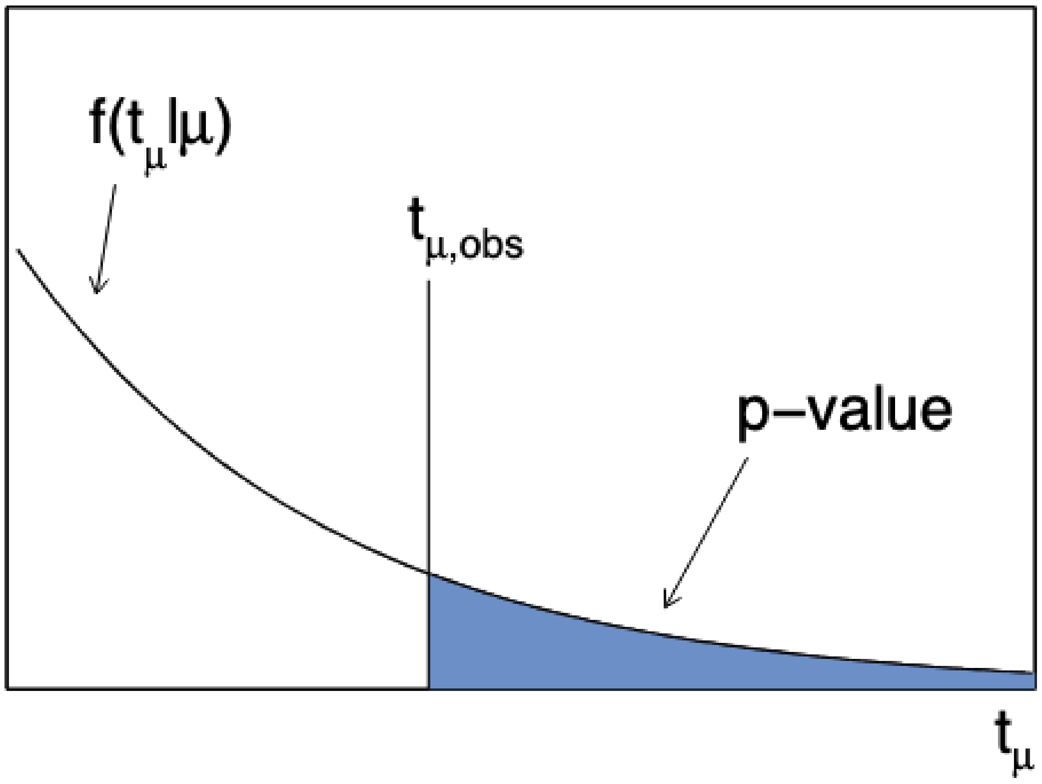


Figure 5.5: This figure adapted from [43] shows the p-value calculation from observed data.

maximizes the likelihood. The denominator is called the global likelihood where $\hat{\theta}_\mu$ and $\hat{\mu}$ are the values of θ and μ which gives give the global maxima.

Once the test statistic is defined, to quantify the level of disagreement between the background-only and signal+background hypothesis, we compute a p-value defined as:

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu|\mu)dq_\mu \tag{5.8}$$

Where $q_{\mu,obs}$ is the value of the test statistic q_μ obtained from data and $f(q_\mu|\mu)$ is the p.d.f. of the test statistic [43] for a given μ.

we can define the p-value calculation equations for the background-only and signal+background hypotheses as:

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu|\mu, \hat{\theta}_{\mu,obs}) dq_\mu \tag{5.9}$$

$$1 - p_b = \int_{q_{\mu,obs}}^{\infty} f(q_\mu|0, \hat{\theta}_{0,obs}) dq_\mu \tag{5.10}$$

The CL_s criterion is then defined as:

$$CL_s(\mu) = \frac{p_\mu}{1 - p_b} = \frac{CL_{s+b}}{CL_b} \tag{5.11}$$

So, for a specified significance level of α, if the CL_s value is less than α ($CL_s \leq \alpha$) at unity signal strength modifier value ($\mu = 1$), can be used to reject the monotop signal+background hypothesis with a (1 - α) × 100% confidence level.

CHAPTER 6

Statistical Framework of Monotop Analysis

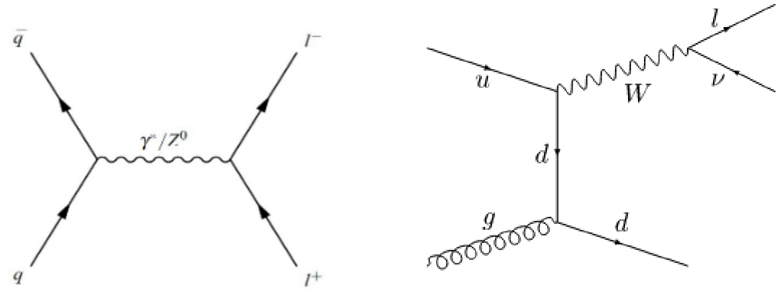
This chapter will explain the analysis strategy for the monotop analysis in the leptonic channel

6.1 Analysis Strategy

This section will detail the analysis strategy for the search of leptonically decaying mono-top signature. As we have seen in section 5.2, the final state signature of the mono-top events includes missing transverse energy and leptonic or hadronic decay of the top quark. In this analysis since we focus on the leptonic decay of mono-top as the final state signature along side the missing transverse energy. The theoretical mono-top models being tested in this analysis expects a large missing transverse energy due to the recoil of the top quark against the higher mass of the dark matter mediator, which further decays into two invisible dark matter candidates.

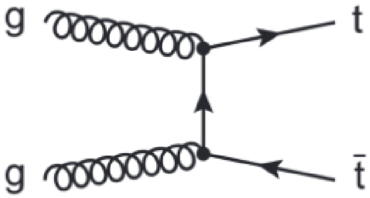
The final state signature we are targeting is the presence of a large missing energy (from the proposed DM candidates and lepton neutrino from the single top quark decay) and the presence of a b-tagged jet alongside a prompt lepton, called leptonic decay. However there are several SM processes that produce the same final state signature. We will briefly detail them below.

The Drell-Yan process occurs when a quark and an anti-quark from separate hadrons annihilate creating a Z boson (Z^0) or a virtual photon (γ^*) with other jets, which further decay into a pair of oppositely charged leptons. If there is a miss-reconstruction of one of the lepton may result in signal like signature in the final state with a lepton, a B-jet and missing energy. A Leading-order Feynman diagram is shown in Figure 6.1.a. The total production cross section of Drell-Yan process (see Table 6.2) is significantly higher that the expected cross section of the

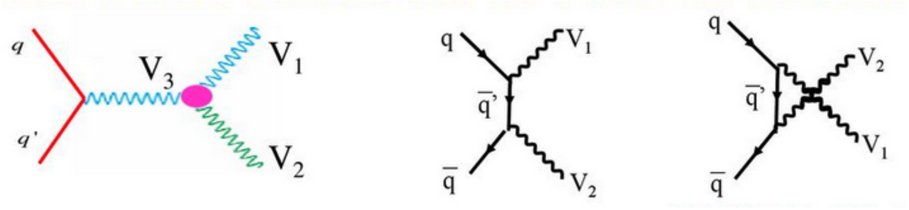


(a) Drell-Yan Jets

(b) W(l ν) Jets



(c) $t\bar{t}$ Jets



(d) Di-Boson

Figure 6.1: Leading-order Feynman diagram for a.) Drell-Yan process where a neutral Z boson decays to a lepton and miss-reconstructed lepton may result in signal like signature. b.) W boson production decaying into charged lepton and neutrino in association with jets. c.) $t\bar{t}$ production decaying into a W boson and b-jet. d.) WW, ZZ, WZ production with bosons decaying with signal like signatures.

signal samples (see Table 6.1) in the mono-top model. So, the Drell-Yan process is an important background to be estimated in our analysis.

The other background of relevance is the production of W boson decaying into a lepton and a neutrino in association with other jets. This final state could possibly mimic our signal final states if one of the additional jets is mis-reconstructed as a b-jet. Fig 6.1.b shows the Leading-order Feynman diagram for this process. In addition, the total cross section of the W boson production is significantly higher than the expected cross section of the signal in the mono-top model thus giving us an indication that this background needs highly precise measurements.

the $t\bar{t}$ background is another background which may produce signal like signature. Fig 6.1.c shows the Leading-order Feynman diagram. From table 6.2 we see that the total $t\bar{t}$ cross section is also significantly higher than the expected cross section of the signal in the mono-top model. The $t\bar{t}$ background can decay via a hadronic, semi-leptonic and di-leptonic modes. As we know that in this analysis we are looking for a final state comprising of a lepton, a b-tagged AK4 jet and missing energy, the semi-leptonic decay of $t\bar{t}$ would be the major background. In semi-leptonic decay if one the top decays hadronically producing a AK4 jet mis-tagged as a b-jet and the other decay leptonically producing a lepton, we could see a signal like signature from the $t\bar{t}$ background. The other decay mode would be a di-leptonic decay which is suppressed because of the small branching ratio. and finally, in all hadronic mode the missing energy contribution is not enough to be accounted as a signal-like signature. Figure 6.2 shows the decay mode branching ratios for the $t\bar{t}$ process.

The discriminating variable used in the statistical analysis is the transverse mass of the W boson. The transverse mass is calculated by combining the information from the lepton, p_T^l and the MET, $\not{E}_T$ according to the following formula:

$$m_T^W = \sqrt{2p_T^l \not{E}_T(1 - cos(\Delta\phi(l - \not{E}_T)))} \tag{6.1}$$

Where p_T^l is the transverse momentum of the leading lepton, $\not{E}_T$ is the missing transverse energy and $\Delta\phi(l - \not{E}_T))$ is the difference of the ϕ between the $\not{E}_T$ and lepton.

The transverse mass has a distinct signature for W boson processes in which the W boson decays into a lepton l and a neutrino ν. For such cases in which the W boson is produced on-shell the distribution of transverse mass peaks at and around the W boson mass of 80 GeV and any larger contributions of W → lν may come from off-shell W boson decays or due to incorrect reconstruction of the lepton and neutrino.

While for leptonic mono-top events as the missing energy contribution primarily comes from the DM candidates thus resulting in larger contributions to the transverse mass greater than the W boson mass. So, that's the reason we chose m_T^W as the discriminating variable between

	decay mode	Br
di-lepton	$t\bar{t} \rightarrow \hat{l}\nu_{\hat{l}} b\hat{l}'\nu_{\hat{l}'} b$	4.5%
	$t\bar{t} \rightarrow \hat{l}\nu_{\hat{l}} b\tau\nu_{\tau} b$	4.8%
	$t\bar{t} \rightarrow \tau\nu_{\tau} b\tau'\nu_{\tau'} b$	1.3%
all-hadronic	$t\bar{t} \rightarrow qq'bq''q'''b$	45.7%
lepton+jets	$t\bar{t} \rightarrow \hat{l}\nu_{\hat{l}} bqq'b$	28.8%
	$t\bar{t} \rightarrow \tau\nu_{\tau} bqq'b$	15.2%

Figure 6.2: : Shows decay modes for a $t\bar{t}$ process

signal and backgrounds as we expect the signal to have a large transverse mass compared to the dominating W+jets and $t\bar{t}$ backgrounds which have transverse mass closer W-boson mass. Since the W+Jets and $t\bar{t}$ + Jets are the major backgrounds in the leptonic channel, we create signal and control regions to determine the signal contribution and normalization of the W+Jets and $t\bar{t}$ backgrounds. These regions are distinguished based on the number of b-tagged AK4 jets in the event.

6.2 Datasets

Between 2015 and 2018, a total of 150.26 fb^{-1} of data at center- of-mass energy of 13 TeV was recorded, of which 137.64 fb^{-1} is used in the analysis. This loss in data occurs because of the malfunctioning of sub-detectors at different times of data collection. In order to maintain high quality data set such events from problematic detector parts are discarded.

6.2.1 Pileup re-weighting

Whenever $proton - proton$ collisions happens at the LHC, the hard scattered protons are accompanied by soft scattered $p - p$ collisions. During the reconstruction of the jets at CMS [27] the actual p_T of the jets is found to be different from the final-state particles that comprise

Table 6.1. Non-resonant mono-top signal samples split according to the mass of the mediator M_V and the mass of the DM candidates M_χ

M_V [GeV]	M_χ [GeV]	Cross Section [pb]
200	50	59.59
195	100	49.31
200	150	0.2533
300	100	18.69
295	150	2.373
300	300	0.0189
500	150	4.315
495	250	0.704
500	500	0.002262
1000	150	0.3727
995	500	0.07102
1000	1000	5.344E-05
2000	500	0.01297
1995	1000	0.002655
2000	150	0.01414
1750	150	0.02879
1750	700	0.02263
1700	800	0.0189
2500	750	0.003237
2495	1250	0.0006786
3000	1000	0.0008935
2995	1500	0.0001901

the jet. This additional contribution comes from particles of the soft-scattered $p - p$ collisions near the primary vertex (PV) of the hard $p - p$ collision [44] and is called *pileup*. As these soft collisions do not contain any interesting physics, their effect must be mitigated because of their contributions to actual calorimeter deposits and tracks. The Monte Carlo (MC) [45] simulation used for modeling the backgrounds donot account for the presence of *pileup*, so we have to derive event-wise weights from the data and apply a re-weighting procedure to include the *pileup* effects in the MC. We do this by dividing the simulated *pileup* with the observed *pileup* and using this ratio to re-weight the MC events. Figure 6.3 shows weights associated with the additional particle contribution from the soft collisions to the jets originating from a hard collision PV for the entire Run II data.

Table 6.2. Background samples used in the analysis for all years

Sample	Cross Section [pb]
ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8	3.3024
ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85
ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	80.95
ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	136.02
TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	91.4936
TTToHadronic_TuneCP5_13TeV-powheg-pythia8	374.292
TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	365.9744
DYJetsToLL_0J_TuneCP5_13TeV-amcatnloFXFX-pythia8	5125.0
DYJetsToLL_1J_TuneCP5_13TeV-amcatnloFXFX-pythia8	951.4
DYJetsToLL_2J_TuneCP5_13TeV-amcatnloFXFX-pythia8	358.3
DYJetsToLL_Pt-100To250_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	94.39
DYJetsToLL_Pt-250To400_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	3.656
DYJetsToLL_Pt-400To650_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	0.4969
DYJetsToLL_Pt-50To100_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	395.1
DYJetsToLL_Pt-650ToInf_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	0.04689
WJetsToLNu_0J_TuneCP5_13TeV-amcatnloFXFX-pythia8	53300.0
WJetsToLNu_1J_TuneCP5_13TeV-amcatnloFXFX-pythia8	8947.0
WJetsToLNu_2J_TuneCP5_13TeV-amcatnloFXFX-pythia8	3335.0
WJetsToLNu_Pt-100To250_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	757.7
WJetsToLNu_Pt-250To400_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	27.53
WJetsToLNu_Pt-400To600_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	3.511
WJetsToLNu_Pt-600ToInf_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	0.5426
WW_TuneCP5_13TeV-pythia8	118.7
WZ_TuneCP5_13TeV-pythia8	65.5443
ZZ_TuneCP5_13TeV-pythia8	15.8274

The *pileup* can not be measured directly as it comes from soft pp interaction it is directly correlated to the number of primary vertices (N_{PV}). Larger the N_{PV} larger is the pileup. Figure 6.4 shows the mean number of interactions per bunch crossing for the year 2016, 2017 and 2018.

6.2.2 Prefire weights

The data collected during the 2016-2017 era suffered from an effect in which the gradual timing shift of ECAL was not properly propagated to L1 trigger primitives (ref sec 3.2.9), this resulted in falsely associating the high η trigger primitives with previos bunch crossing. Since the L1 trigger 3.2.9 doesn't allow two consecutive bunch crossings to fire, it resulted in rejection of events if a significant amount of ECAL energy is found in the region of $2. < |\eta| < 3$. The effect is strongly eta and pt dependent and prefiring rates well above 50% can be reached for jets with p_T of several hundreds of GeV and $2.75 < |\eta| < 3$ [46]. To resolve this effect a weight is associated with each object can cause a prefire, the final event weight is obtained as the product of the non

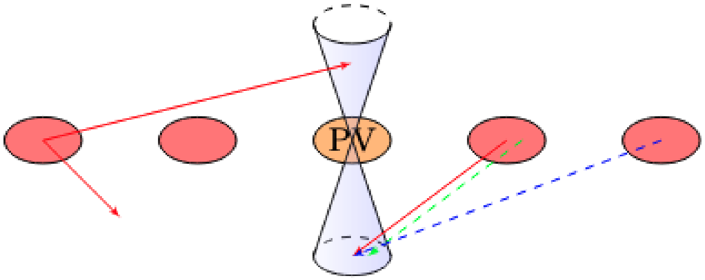
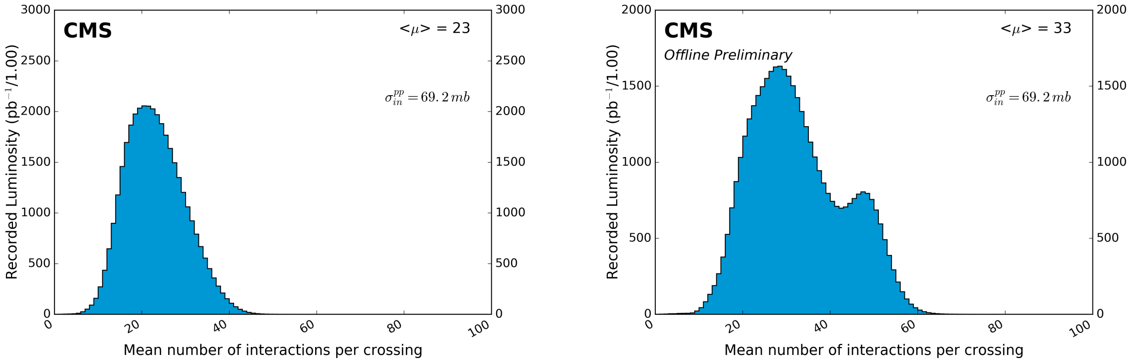
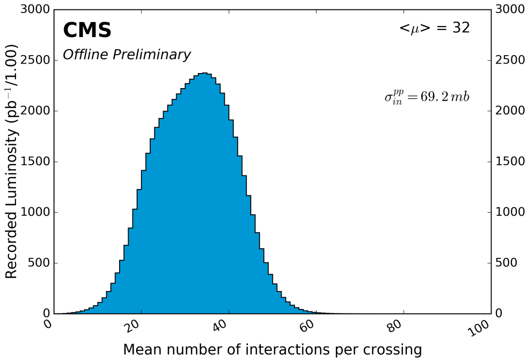


Figure 6.3: This figure taken from [44] shows two jets coming from a primary vertex and additional soft pp interactions as red circles. The arrows show the particles from the other collisions contribute to the jets from the primary vertex, where the colored line represent different particles.



(a) 2016

(b) 2017



(c) 2018

Figure 6.4: This figure shows the true pileup distribution of the data observed by the CMS [27] for 2016 (a), 2017 (b), and 2018 (c).

prefiring probability of all objects (measured using unprefirable events) given by eq: 6.2. This issue is only observed in data and so the corrections are only applied to data only.

$$w = 1 - P(prefiring) = \prod_{i=photons,jets,muons} \left(1 - \epsilon_i^{pref}\left(\eta, p_T^{EM,\mu}\right)\right) \tag{6.2}$$

6.2.3 Primary Datasets

The stream of data recorded by CMS [27] is organized into **Primary datasets** based on the relevant high level trigger selections as discussed in sections. 3.2.9, 3.2.10. The main ideology behind this organization is to group together events which contain similar physics information. These events are non-exclusive in nature meaning that an event can occur in multiple *primary datasets* based on the HLT selection. In this analysis the datasets relevant to us for single electron final state are EGamma for 2018 and Single Electron *primary datasets* for 2017, 2016 (Pre and Post VFP), and for single muon final state are Single Muon *primary datasets* for 2016 (Pre and Post VFP), 2017 and 2018. Table 6.3 shows the data samples used in for analysis in this thesis.

The signal and background sample processes were generated using MC simulation. V+Jets background samples were centrally produced at using MADGRAPH5 aMC@NLO [47] generator interfaced with PYTHIA [48] using the CP5 tunes for fragmentation and hadronization. Top samples were produced centrally with PYTHIA POWHEG using the CP5 tunes for fragmentation and hadronization and Di-Boson (WW, WZ, ZZ) is simulated with PYTHIA8 [48]. The last step is passing these events through the CMS detector using Geant4 [27] simulation to apply the effects of the detector to the events. Table 6.2 shows the list of background samples used in the analysis.

Signal samples were privately produced using MADGRAPH5 aMC@NLO [47] generator interfaced with PYTHIA [48] using the CP5 tunes for fragmentation and hadronization. They are generated across a mass grid of the dark matter (M_χ) and mediator mass (M_V), such that $M_\chi < \frac{M_V}{2}$.

Table 6.3. datasets used in the analysis for all years

Sample	Cross Section [pb]
2018	
/EGamma/Run2018A-UL2018_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
/EGamma/Run2018B-UL2018_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
/EGamma/Run2018C-UL2018_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
/EGamma/Run2018D-UL2018_MiniAODv2_NanoAODv2-v1/NANOAOD	1.0
/SingleMuon/Run2018A-UL2018_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
/SingleMuon/Run2018B-UL2018_MiniAODv2_NanoAODv9-v2/NANOAOD	1.0
/SingleMuon/Run2018C-UL2018_MiniAODv2_NanoAODv9-v2/NANOAOD	1.0
/SingleMuon/Run2018D-UL2018_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
2017	
/SingleElectron/Run2017B-UL2017_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
/SingleElectron/Run2017C-UL2017_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
/SingleElectron/Run2017D-UL2017_MiniAODv2_NanoAODv9-v2/NANOAOD	1.0
/SingleElectron/Run2017E-UL2017_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
/SingleElectron/Run2017F-UL2017_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
/SingleMuon/Run2017B-UL2017_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
/SingleMuon/Run2017C-UL2017_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
/SingleMuon/Run2017D-UL2017_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
/SingleMuon/Run2017E-UL2017_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
/SingleMuon/Run2017F-UL2017_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
/SingleMuon/Run2017G-UL2017_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
/SingleMuon/Run2017H-UL2017_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
2016 PreVFP	
/SingleElectron/Run2016B-ver1_HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD	1.0
/SingleElectron/Run2016C-HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD	1.0
/SingleElectron/Run2016D-HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD	1.0
/SingleElectron/Run2016E-HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD	1.0
/SingleElectron/Run2016F-HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD	1.0
/SingleMuon/Run2016B-ver1_HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD	1.0
/SingleMuon/Run2016C-HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD	1.0
/SingleMuon/Run2016D-HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD	1.0
/SingleMuon/Run2016E-HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD	1.0
/SingleMuon/Run2016F-HIPM_UL2016_MiniAODv2_NanoAODv9-v2/NANOAOD	1.0
2016 PostVFP	
/SingleElectron/Run2016F-UL2016_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
/SingleElectron/Run2016G-UL2016_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
/SingleElectron/Run2016H-UL2016_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
/SingleMuon/Run2016F-UL2016_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
/SingleMuon/Run2016G-UL2016_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0
/SingleMuon/Run2016H-UL2016_MiniAODv2_NanoAODv9-v1/NANOAOD	1.0

The couplings between the dark matter mediator (M_V) and quark (q) and dark matter mediation (M_V) and dark matter particle (M_χ) represented by $g_{V,q}$ and $g_{V,\chi}$ are set to 1 and 0.25 respectively, as in the theory paper [49]. Each signal sample has around 250k to 300k events and table 6.1 shows the signal mass grid with their respective cross sections.

	Electron	Muon
2018	Ele32_WPTight_Gsf **OR** Photon200	IsoMu24
2017	Ele35_WPTight_Gsf **OR** Photon200	IsoMu27
2016pre	Ele27_WPTight_Gsf **OR** Photon175	IsoMu24 **OR** IsoTkMu24
2016 post	Ele27_WPTight_Gsf **OR** Photon175	IsoMu24 **OR** IsoTkMu24

Figure 6.5: This table show the combination of triggers used in the analysis

6.3 Triggers

This section will present the motivations for selecting a particular HLT path and the specifications of the properties of the selections used in the analysis. The HLT trigger path is a selection mechanism which makes sure that only events which pass the relevant triggers are stored in primary datasets and discarded otherwise. Figure 6.5 shows the combination of HLTs used in this analysis.

The trigger paths described in Figure 6.5 are relevant based on the final state signature of the signal sample, which is the presence of an isolated lepton and large missing energy. These paths are dependent on the year in which the data were recorded. Table 6.4 shows the description of the triggers across different eras.

In the leptonic analysis, we require the presence of an isolated electron and that is the reason we have used the triggers mentioned in table 6.4. In the channel with electron in the final state we require the triggers `HLT_EleYY_WPTight_Gsf_vX` and `HLT_PhotonYY_vX`. The former asks for a trigger level selection of only those events which have an isolated electron in the final state with p_T threshold (denoted as `YY`) at 32, 35 and 27 GeV for 2018, 2017 and 2016 (pre & post VFP) respectively. In latter case, we ask for non-isolated high p_T electron with a p_T threshold of 200 GeV for 2018 & 2017 and 175 GeV for 2016 (pre & post VFP) eras. The reason for choosing the combination of isolated and non-isolated triggers is as follows. Isolation in this context implies the absence of other particles near the electrons, these particles may come from some hadronic activity after collisions. As we move towards selecting high p_T electrons the efficiency of the isolated electron trigger drops, this drop is partially accounted by including the non-isolated high p_T trigger path. Similarly, we require the presence of an isolated muon in the final state and Table 6.4 lists the muon channel triggers as well. For collecting such events with final state muons, we require the events to pass `HLT_IsoMuYY_vX` trigger with p_T threshold (denoted as YY) of the muon to be 24 GeV for 2016 (pre & post VFP) & 2018 and 27 GeV for 2017. We also require the selection of isolated muons with hits in the tracker `HLT_IsoTkMu24_vX` of the CMS detector with a p_T threshold of 24 GeV for 2016 (pre & post VFP) era.

The quantity which allows us to quantify the trigger selections is the efficiency of the HLT path. The efficiency of a trigger path is defined as the ratio of the number of events passing the HLT path and the total number of events that passes the reference trigger selection. The trigger efficiencies are derived for both data sample and the corresponding MC sample for a particular HLT path and since the two efficiencies are different, this effect is propagated to the MC samples through the trigger scale factors. These scale factor is just a ratio of the efficiency of the HLT path in data (ϵ_{data}) with the efficiency of the HLT path in MC (ϵ_{MC}). Figures 6.6 and 6.7 shows the efficiencies for the electron channel HLT paths in table 6.4 for electrons passing a tight CutBased Id selection and muons respectively.

6.4 Object Reconstruction

6.4.1 Electrons

It is of utmost importance to select electrons which can be reconstructed to gain access to events which provide crucial inference about the interesting physics at the LHC [17]. The electrons are charged particles which interact electromagnetically inside the detector, they leave a signature in the form of an energy deposit inside the ECAL, and these energy deposits are matched with the tracks created by hits in silicon tracker to find the kinematical information (p_T, η, ϕ) about the electrons [52].

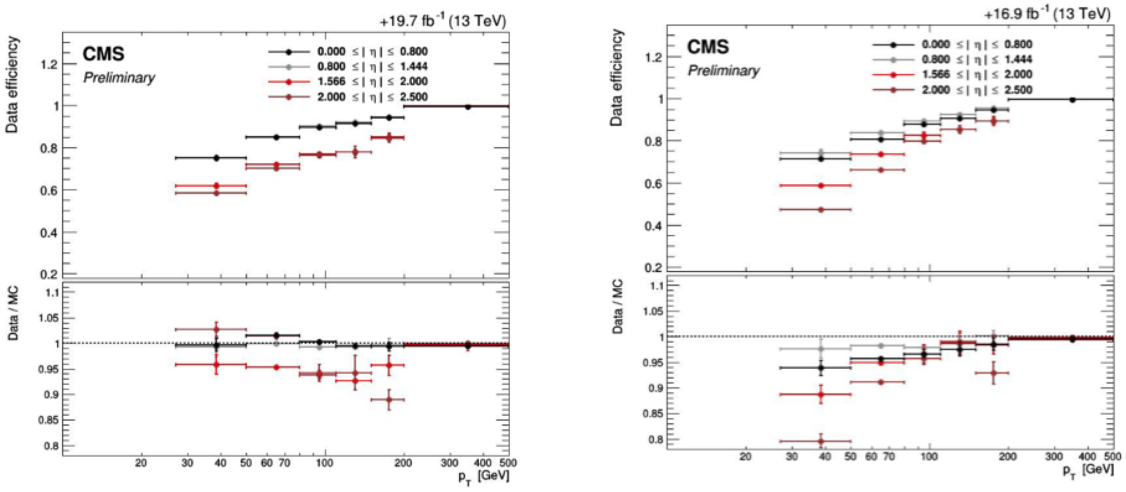
Table 6.4. High Level trigger objects used in the analysis

Signature	Trigger Name
Electron Channel	
2018	
isolated electron	HLT_Ele32_WPTight_Gsf_vX
high p_T electron	HLT_Photon200_vX
2017	
isolated electron	HLT_Ele35_WPTight_Gsf_vX
high p_T electron	HLT_Photon200_vX
2016 PreVFP	
isolated electron	HLT_Ele27_WPTight_Gsf_vX
high p_T electron	HLT_Photon175_vX
2016 PostVFP	
isolated electron	HLT_Ele27_WPTight_Gsf_vX
high p_T electron	HLT_Photon175_vX
Muon Channel	
2018	
isolated Muon	HLT_IsoMu24_vX
2017	
isolated Muon	HLT_IsoMu27_vX
2016 PreVFP	
isolated Muon	HLT_IsoMu24_vX
isolated Muon	HLT_IsoTkMu24_vX
2016 PostVFP	
isolated Muon	HLT_IsoMu24_vX
isolated Muon	HLT_IsoTkMu24_vX

It is important to note that not all electrons produced and detected are of relevance. In order to select relevant electrons we require the electron objects to satisfy certain criteria. So, in this section we will describe the two collections of electrons used in this analysis namely, loose electrons and tight electrons.

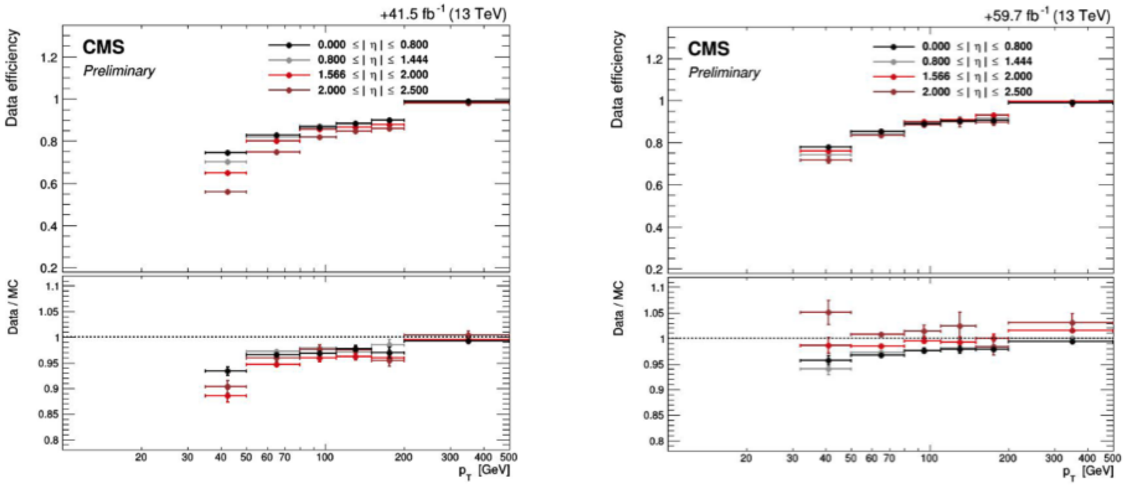
Loose Electrons:

From Table 6.5 we can see the selection criteria for selection loose electrons, this criteria basically enhances the efficiency of electron selection (95% [53]) with a not-so-great identification rate. The p_T threshold is set at 15 GeV with a veto `CutBasedID` [53]. Alongside these selections, the loose electrons should also satisfy an upper threshold selection of 0.05 in the ECAL barrel (EB) and 0.1 in the ECAL endcap (EC), for the transverse (relative to the primary vertex)



(a) 2016 Pre VFP

(b) 2016 Post VFP



(c) 2017

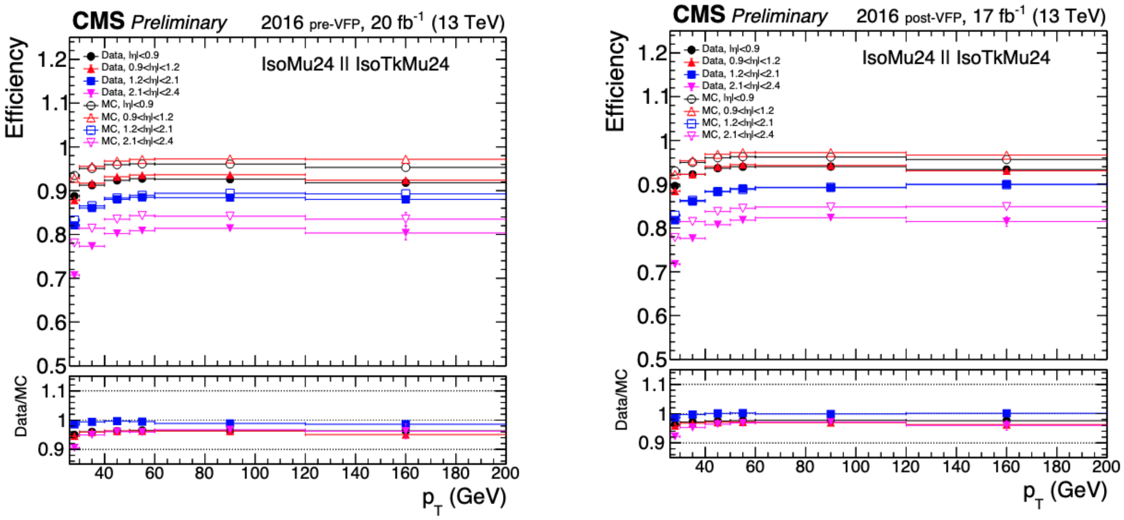
(d) 2018

Figure 6.6: This figure taken from [50] shows the trigger efficiencies as a function of p_T for the electron channel HLT paths in Table 6.4 for electrons passing tightCutbased id for 2016, 2017and 2018 eras.

impact parameter d_{xy}. A threshold of 0.1 in the ECAL barrel (EB) and 0.2 in the ECAL endcap (EC) on the longitudinal (relative to the primary vertex) impact parameter d_z is also required.

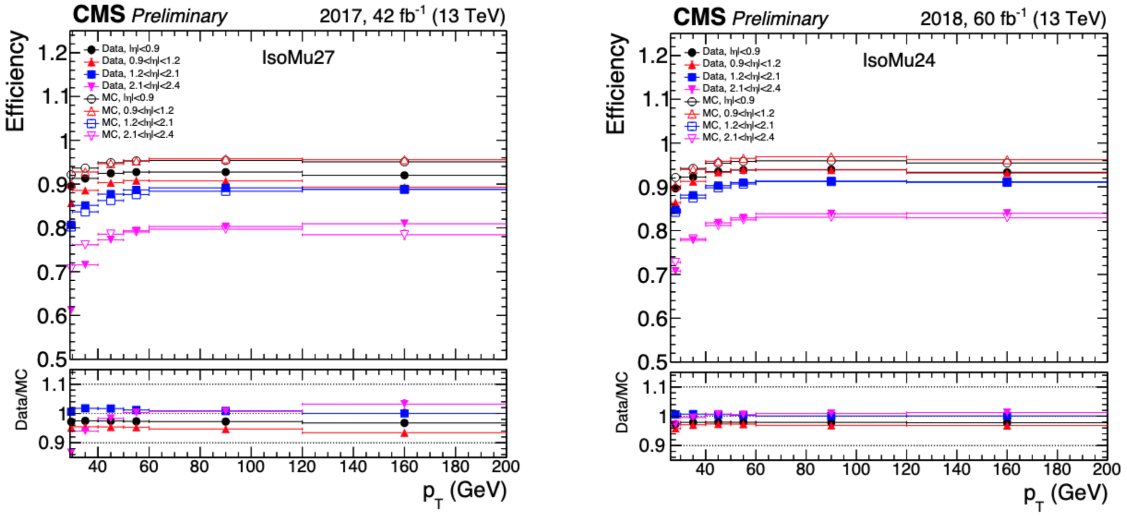
Tight Electrons:

From Table 6.5 we can see the selection criteria for selection tight electrons, this criteria basically enhances the selection of electrons with a low mis-identification rate and a rather lower efficiency of electron selection (70% [53]). The p_T threshold is set at 40 GeV with a tight `CutBasedID` [53]. Alongside these selections, the tight electrons should also satisfy an upper



(a) 2016 Pre VFP

(b) 2016 Post VFP



(c) 2017

(d) 2018

Figure 6.7: This figure taken from [51] shows the trigger efficiencies as a function of p_T for the muon channel HLT paths in Table 6.4 for 2016, 2017and 2018 eras.

threshold selection of 0.05 in the ECAL barrel (EB) and 0.1 in the ECAL endcap (EC), for the transverse (relative to the primary vertex) impact parameter d_{xy}. A threshold of 0.1 in the ECAL barrel (EB) and 0.2 in the ECAL endcap (EC) on the longitudinal (relative to the primary vertex) impact parameter d_z is also required.

ID and Reconstruction Corrections:

Since in reality the reconstruction and identification of electron objects [52] is not perfect inside the CMS detector, corrections for the observed detector effects in recorded data need to be propagated to the MC simulation of the samples. These corrections incorporated through scale factors ($\epsilon_{data}/\epsilon_{MC}$) allow the MC to have similar efficiencies ($\epsilon_{reconstruction}$ & ϵ_{id}) as the observed data. The scale factors depend on the p_T the electron and η of the ECAL supercluster and are made available centrally by the respective electron object group for reconstruction and ids for `2018, 2017, 2016 (pre & post VFP))` eras [54]-[55].

Table 6.5. Loose and Tight electron selection criteria used

Type	p_T[GeV]	$\lvert\eta\rvert$	Electron ID	IP_{xy} [cm]	IP_z [cm]
loose	>15	<2.5	veto ID	0.05(EB), 0.1(EC)	0.1(EB), 0.2(EC)
tight	>40	<2.5	tight ID	0.05(EB), 0.1(EC)	0.1(EB), 0.2(EC)

6.4.2 Muons

As discussed before, in the case of a leptonic decay mode of monotop, we require the presence of an electron or a muon in the final state alongside large transverse missing energy. Thus, making muon a critical object for probing the leptonic monotop signature. Analogous to electrons, in this analysis we have used two collections of muons namely, loose muons and tight muons. From table 6.6, we see the the p_T threshold to be set at 20 GeV for loose muon passing a loose Id and loose isolation selection criteria inside an $|\eta|< 2.4$ of the CMS detector. Similarly for tight muons, a p_T threshold of 40 set for a tight muon passing a tight Id and tight isolation requirement inside an $|\eta|< 2.4$ of the CMS detector.

As observed in the efficiencies in the loose and tight ids for electrons, for loose muon Id the selection efficiency is higher than the tight muon Id but the tight muon Id has a lower misidentification rate. From [56], we see that for loose muon Id the efficiency is less than 99% with a misidentification rate of less 0.5%. while for tight muon Id, the efficiency is less than the loose muon Id efficiency at 96% but a better misidentification rate at 0.3%.

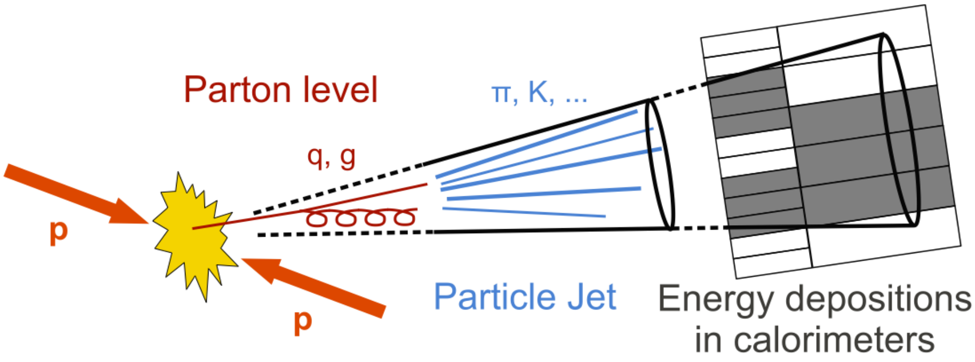


Figure 6.8: This figure shows the pp collisions resulting in a quark-gluon production which hadronize to produce jets

In order to account for the detector effects in the MC simulation, we need to apply some corrections to the efficiencies of isolation and identification (ID) measurements of the muons in the MC simulation. These corrections are propagated by means of scale factors, they depend on the η and p_T of the muons. These scale factors are made available centrally through the corresponding muon object group. The id and isolation scale factors are available for `2018, 2017, 2016 (pre & postVFP))` [57].

Table 6.6. Loose and Tight muon selection criteria used

Type	p_T[GeV]	$\|\eta\|$	muon ID	muon isolation
loose	>20	<2.4	loose ID	loose
tight	>40	<2.4	tight ID	tight

6.4.3 Jets

When protons collide at the LHC, quarks and gluons are produced. These particles are color-charged and cannot be isolated due to the color confinement property of strong force. So, the quarks and gluons undergo hadronization and further creating more quark-gluon pair which hadronize futher resulting in a hadronic shower and recorded as a jet signature in the detector. See Figure 6.8. The collections of jets used in the analysis are the particle-flow charged hadron subtraction jets (AK4 PF Jets) with a cone radius of 0.4 cm in the η-ϕ plane of the CMS detector. We will also look at the AK4 Jets tagged as coming from the hadronization of a b-quark.

6.4.3.1 AK4 PF CHS Jets

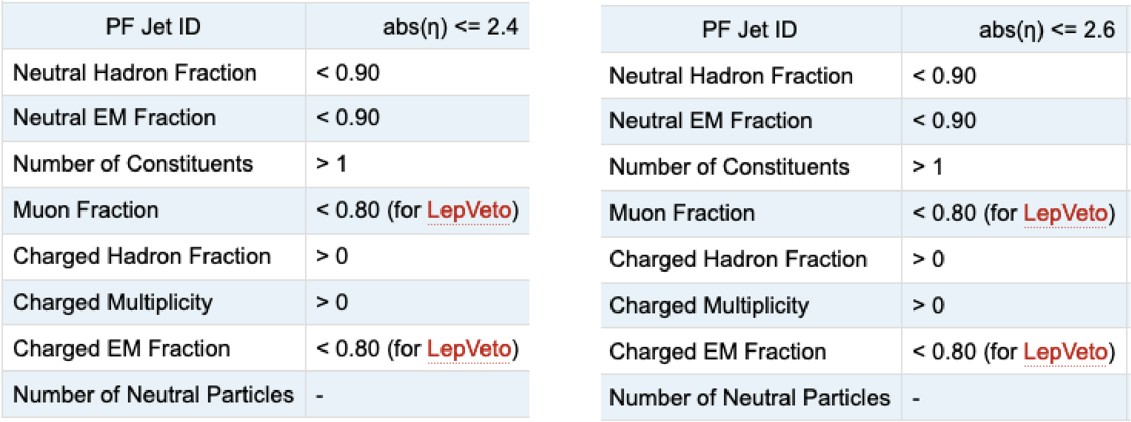
First, to understand the naming of these jets as `AK4 PF CHS` Jets. Jets are reconstructed by clustering the energy deposits in the ECAL with the tracks in the silicon tracker (also know as PF candidates as they are constructed by the Particle Flow algorithm [58]), CMS uses the algorithm named anti-k_T (AK) algorithm [39] to cluster jets with a jet radius of 0.4 cm. The effects of pileup are handled by the Charged Hadronic Subtraction (CHS) algorithm [59].

In the anti-k_T clustering algorithm the distance between object is measured by the equation 6.3. Where $\Delta_{ij}^{2p} = \left(y_i - y_j\right)^{2p} + \left(\phi_i - \phi_j\right)^{2p}$, k_{ti}, y_i, and ϕ_i are respectively the transverse momentum, rapidity and azimuth of particle i. d_{ij} refers to the distance between the objects i and j which d_{iB} refers to the distance between the object i and the Beam. The reason behind the naming of the algorithm anti-k_T (p<0) is to differentiate from the already existing k_T algorithm (p>0).

$$d_{ij} = min\left(\frac{1}{d_{ti}^{2p}}, \frac{1}{d_{tj}^{2p}}\right)\frac{\Delta_{ij}^{2p}}{R^2}, \;\; d_{iB} = \frac{1}{k_{ti}^2} \tag{6.3}$$

The performance of reconstructed jets is different in data and in simulated samples. Because of the presence of detector effects the reconstruction of true physical jets is inaccurate in both data and simulated samples. To account for these effects a correction is derived using a multijet event sample [60]. There are two type of jet corrections that are applied (as a function of jet η and p_T) in this analysis, the jet energy scale (JES) and the jet energy resolutions (JER). The JES correct the p_T of the jets, while the JERs correct the resolution of the simulated samples to match with data [60], the JER and JES are combinedly called as jet energy corrections or JEC for short. More details in section 6.4.3.2 and 6.4.3.3

Prior to applying the JEC (JER/JES), the jets have to pass a selection criteria in order to ensure that the reconstructed jets correspond to the real physics jets [61]. The identification criteria is shown in figure 6.9 for 2016, 2017 & 2018 eras:



PF Jet ID	abs(η) <= 2.4
Neutral Hadron Fraction	< 0.90
Neutral EM Fraction	< 0.90
Number of Constituents	> 1
Muon Fraction	< 0.80 (for LepVeto)
Charged Hadron Fraction	> 0
Charged Multiplicity	> 0
Charged EM Fraction	< 0.80 (for LepVeto)
Number of Neutral Particles	-

(a) 2016

PF Jet ID	abs(η) <= 2.6
Neutral Hadron Fraction	< 0.90
Neutral EM Fraction	< 0.90
Number of Constituents	> 1
Muon Fraction	< 0.80 (for LepVeto)
Charged Hadron Fraction	> 0
Charged Multiplicity	> 0
Charged EM Fraction	< 0.80 (for LepVeto)
Number of Neutral Particles	-

(b) 2017/2018

Figure 6.9: This figure taken from [62] shows the identification criteria for jet selection

After this selection criteria the analysis also employs a $p_T > 30$ GeV selection on the jets. In order to account to effect of particles inside the jet due to pileup, a pile tight JetId is used as discriminator on jets with $p_T < 50$ GeV. It is BDT discriminator which is used to distinguish the prompt jets from the pileup jets [63].

6.4.3.2 JER

The JER corrections help in adjusting the resolution in simulated samples to match with the resolution data. There can be two scenarios for updating the resolution (4 momentum) of the jet, one where a particle level jet is found around the reconstructed jet such that the distance between the two jets $\Delta R < R/2$, and the other, where the particle-level jet can not be found within $\Delta R < R/2$. The correction formulae are shown in eq 6.4 and 6.5 respectively for the two cases. where s_{JER} is the resolution scale factor, p_T is the transverse momenta of the reconstructed jet and $p_T^{particle}$ is the the transverse momenta of the particle-level jet, $\mathcal{N}(0, \sigma_{JER})$ is Gaussian normal distribution with mean 0 and standard deviation σ_{JER}.

These correction factors are applied as a function of the η and p_T of the reconstructed jet and are provided centrally by the CMS collaboration [64]

$$c_{JER} = 1 + (s_{JER} - 1)\,\frac{p_T - p_T^{particle}}{p_T} \tag{6.4}$$

$$c_{JER} = 1 + \mathcal{N}(0, \sigma_{JER})\,\sqrt{max\left(s_{JER}^2 - 1, 0\right)} \tag{6.5}$$

6.4.3.3 JES

The jet energy scale corrections are applied via a factorized method meaning that it is a step by step process to counter different effects. At each factorized step a correction factor is applied to the jets as a function of η & p_T. The first step involves accounting for the contribution from soft collisions to the jets, also known as pileup. Following which the detector effects are taken into consideration, the differences are approximated as a function of η & p_T. These correction are provided centrally by the CMS collaboration [64].

6.4.3.4 AK4 Heavy flavor tagged PF CHS Jets

As discussed in section 6.1, that the final state signature in mono-top requires the presence of a lepton, a b-tagged jet and large missing energy. A b-tagged jet is jet coming from hadronization of b-quark, CMS uses the DeepJet algorithm [65] which a neural network based model for classification of the jet based on the working points in table 6.7. In this analysis we have used the medium working point and the corresponding corrections are applied and uncertainties propagated.

As we have understood from previous sections that in order for the simulated samples to behave similar to data we need to update the efficiencies of the simulated samples to match with data. a scale factor is derived as the ratio of efficiency of data with the MC sample. These scale factors depend on jet p_T, η and hadron flavor of the jet at generator level.

As you will see in section 6.5 we have

Table 6.7. b-tag working points used in the analysis

Working Point	discriminant value ≥			Mistag Probability
	2016	2017	2018	
loose	0.0614	0.0521	0.0494	10%
medium	0.3093	0.3033	0.2770	1%
tight	0.7221	0.7489	0.7264	0.1%

6.4.4 Missing Transverse Energy

The CMS detector has many cylindrical layers which provides a hermetic coverage to technically detect all the particles produced during p-p collisions. It is because of this property of the CMS detector which makes it possible to measure the kinematic properties of all the particles produced. Since the collisions take place in the rest frame of the beam, the net transverse momenta of the beam is zero i.e. the total transverse momenta vector $\mathbf{p_T^{miss}}$ of the outgoing particles should be zero by the law of conservation of momentum. However, some particles like the neutrinos or the hypothesized dark matter particles are not detected by the CMS detector which results in a non-zero net transverse momenta $\mathbf{p_T^{miss}}$. So, $\mathbf{p_T^{miss}}$ is used as a probe for such weak-interacting particles.

The reconstruction of the missing energy $\mathbf{p_T^{miss}}$ is done using the PF candidates analogous to the the jet reconstruction. As is known that the reconstruction efficiency is affected by the detector and the mis-reconstruction effects, a correction factor is introduced which propagates the corrections applied to jets (ref 6.4.3.2 and 6.4.3.3) to the missing energy as in eq 6.6, these corrections are also know as Type I corrections.

$$p_T^{miss,corr} = p_T^{miss} - \sum_{jets} \left(p_{T,jet}^{corr} - p_{T,jet} \right) \tag{6.6}$$

On top of the Type I corrections, met x-y shift corrections [66] are applied to the missing energy. These corrections help in correcting the azimuthal (ϕ) angle dependence on the reconstructed met to combat the effects of an-isotropic detector response, detector mis-alignment and displacement of the beam spot.

Large missing energy may not always point to interesting physics phenomena but may also act as a proxy for uninteresting effects such as detector noise, cosmic rays and beam-halo particles. Such met is called `fake met , false met or anamolous met`. So, we have used some algorithms which use timing, pulse-shape and signal topology to identify and remove false met. The filters [67] in tab 6.8 have been recommended to remove fake met and are applied to both data and MC.

Table 6.8. Met Filters used in the analysis

Filter name	2016 (pre & postVFP)	2017	2018
primary vertex filter: removes events without reconstructed primary vertices	✓	✓	✓
beam halo filter: removes beam halo backgrounds	✓	✓	✓
HBHE noise filter: removes events with poor HCAL performance in general	✓	✓	✓
HBHEiso noise filter: removes events with poor isolated HCAL channels	✓	✓	✓
EcalDeadCellTriggerPrimitiveFilter: removes events with dead ECAL channels that fired HLT triggers	✓	✓	✓
Bad PF Muon Filter removes events with poor PF muon reconstruction	✓	✓	✓
Bad PF Muon Dz Filter:	✓	✓	✓
ee badSC noise filter	✓	✓	✓
ECAL bad calibration filter update	✓	✗	✗

6.5 Event Selection

In this section, we will describe in detail the selection criteria on the events which maximize the contribution of signal like events in the signal region and quantify major backgrounds (as discussed in section 6.1 in the control region regions dominated by W+Jets and $t\bar{t}$ backgrounds, respectively [68].

The distinguishing variable used between signal and background like events is the transverse mass of W boson 6.7. This variable allows for compare events dominated by signal-like events and identify the signal like signature in the signal region with a characteristic shape different from the background like events. As we require the presence of a single b-tagged jet, a lepton and large met in the signal region, on top of the basic event selection (pre-selection) we use the number of b-tagged jets selection criteria to differentiate between the signal and background control regions.

This section will first describe the basic selection criteria for event selection, followed by the background control region and signal region definitions.

$$m_T^W = \sqrt{2p_T^l \not{E}_T(1 - cos(\Delta\phi(l - \not{E}_T)))} \tag{6.7}$$

6.5.1 Pre-Selection

As discussed in the section 6.1, the signal region is dominated by the presence of large missing energy from the DM candidate decay an isolated lepton and a b-tagged jet which come from the decay of the single top quark. The top quark recoils against the DM candidate before decaying into a b-tagged jet and a W boson. The W boson further decays into a lepton and its corresponding anti-neutrino. The mono-top signal signature of consist of a single high p_T jet which is tagged as a originating from the hadronization of a B hadron. See sec 5.2.2 on how to classify a jet as a b-jet.

1. The magnitude of missing transverse momentum of the W boson m_T^W >150 GeV. As the phase space for < 150 GeV would be covered by the hadronic monotop channel [69]. In addition, we also require the presence of isolated leptons which pass the tight identification criteria as defined in section 6.4 and an isolated $high - p_T$ jets ($p_T > 70$ GeV) with radius of 0.4 cm recoiling against the DM candidates. These AK4 jets would later be tagged as b-jets and used as a discriminant between the signal and background control regions.
2. The angle between missing energy and leading AK4 Jet in transverse plane of the detector i.e. $\Delta\phi(\not{E}_T, Jet) > 1.5$. This selection removes the mis-modelled signal region in the detector. The mis-modelling is the result of QCD events for which the momenta are incorrectly measured.
3. The missing transverse momentum $\not{E}_T > 100$ GeV to only consider region of the phase space which are well modelled in the simulationand counter the effects of QCD multi-jet events.
4. In the 2018 data taking period, due to power failure of a 40° section of one end of the Hadronic Calorimeter (HEM) so the events in data collected from this particular region of the detector have to be removed. The events with missing energy ($\not{E}_T > 470$ GeV) pointing to the region -0.62 < $\phi(\not{E}_T) < -1.62$ are veto'd. Also, jets with $p_T > 30$ in the previously mention HEM region and -3 < η <-1.3 are also not considered in the analysis.

6.5.2 Background Control Regions

As mentioned in sec 6.1, W+jets and $t\bar{t}$ are the major backgrounds which could possibly mimic the monotop signature. It is very essential to quantify the estimation of these backgrounds to get a precise estimate of them in the signal region. The number of b-tagged jets is used a differentiating object between the signal and background control region. In the signal region we require the presence of exactly one b-tagged jet which for W+jets and $t\bar{t}$ we require zero and at least two b-jets, respectively.

6.5.2.1 W Control Region

There is motivation behind the creation of the W+jets control region. As we know from section 6.1 that in the final state signature we are looking for a lepton (e/μ), a b-tagged jet and large missing transverse momentum primarily coming from the dark matter candidate and lepton neutrino. We can also expect a similar signature coming from the W+jets background, the W boson decays into lepton and its corresponding lepton neutrino (W→lν) and the additional jet could come from a light flavor quark or gluon and mis-tagged as a jet coming from a b-quark, thus giving a leptonic monotop signal like signature. Although at CMS the mistaging probability for such jets is only 1% with the `DeepJet`, see section 5.2.2 algorithm (Table 6.7), the large cross section of the W+jets background results in significant number of such events mimicking a potential monotop signal.

We have divided the W+jet background control region in two, a region with electron in the final state and another region with muon in the final state. In the former, we require exactly one electron which passes the tight identification criteria, sec 6.4.1 and zero muons passing the loose identification, sec 6.4.2 selection criteria, while in the later case, we require exactly one muon which passes the tight identification criteria, sec 6.4.2 and zero electrons passing the loose identification, sec 6.4.2 selection criteria. This criteria ensures that we select a very high percentage of good prompt leptons and thus allowing us to make a good estimation on the W+jets background prediction. Figures 6.10, 6.10 shows the distributions of p_T, η and ϕ for electrons and muons, respectively for the year 2018. These distribution are dominated by the W($l\nu$) with additional jets type events. As we can see that the simulation is fairly well modeled in the W+jets control region.

In addition to the above criteria on leptons, we require the presence on AK4 jets tagged as coming from a b-quark. The AK4 jets are clean jets meaning, events containing jets having a loose lepton or photon within a cone radius of 0.4 cm of the jet are discarded. Fig 6.10.d shows the p_T distribution of leading jet for the year 2018, showing good agreement between simulation and data in the W+jets control region. For figures of the W control region in era 2017, 2016postVFP and 2016preVFP ref to Appendix 1 fig A.1, A.3 and A.2. We also require the events to pass at least one of the single electron (muon) triggers to select events with prompt electrons (muons) as in tab 6.4.

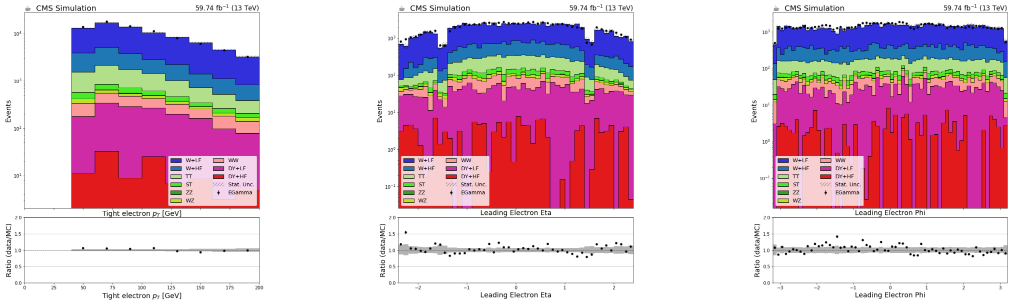
From Fig 6.10.e. we see that the transverse mass of the W boson (m_T^W) is modeled nicely as can be seen from the data-simulation ratio plot. For more figures relevant to era 2017, 2016postVFP and 2016preVFP ref to Appendix 1 fig A.1, A.3 and A.2.

Similarly we can see good agreement between data and simulation in the W control region with muon in the final state. See fig 6.11 and Appendix 1 fig A.4, A.6 and A.5 for control region plots for the other eras.

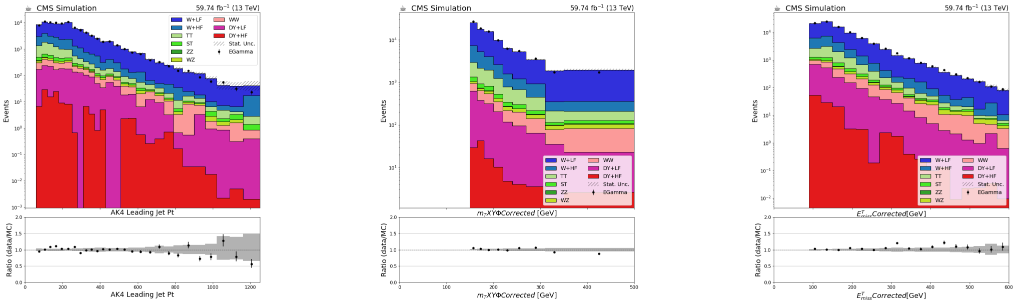
6.5.2.2 Top ($t\bar{t}$) Control Region

As mentioned in the section 6.1 that alongside the W→l ν + other jets background we also have the top quark-top anti-quark pair ($t\bar{t}$)with additional jets. It is a major background whose final state could result in a mono-top signal like event. In the $t\bar{t}$ + jets background because of the presence of a b-tagged jet and a charged lepton could result in final states which mimic the mono-top signal. So, We create the $t\bar{t}$+jets control region to get a good estimate of this background. Below we talk about the selection criteria imposed on objects to enrich the control region with events coming from $t\bar{t}$+jets background.

The only distinction in the selection criteria between the W boson control region, see sec 6.5.2 and the top control region is the requirement on the number of AK4 b-tagged jets in the event. In the top control region we require at least two AK4 b-tagged jets which pass the medium working point see sec 6.7. This selection enriches the top control region with events coming from $t\bar{t}$, thus also suppressing events coming from W→lν with additional jets background.



(a) Leading electron p_T (b) Leading electron η (c) Leading electron ϕ

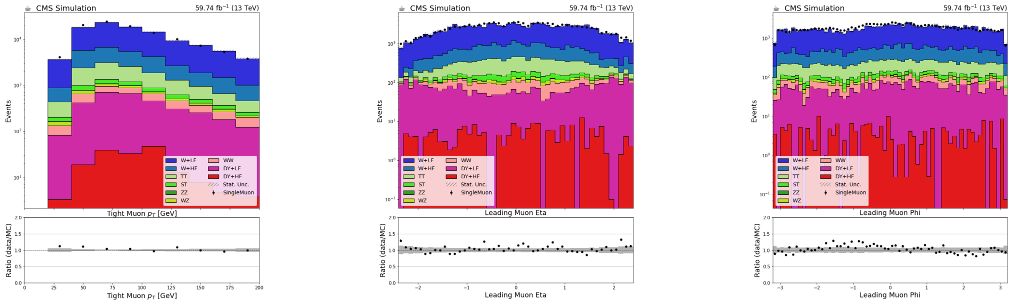


(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

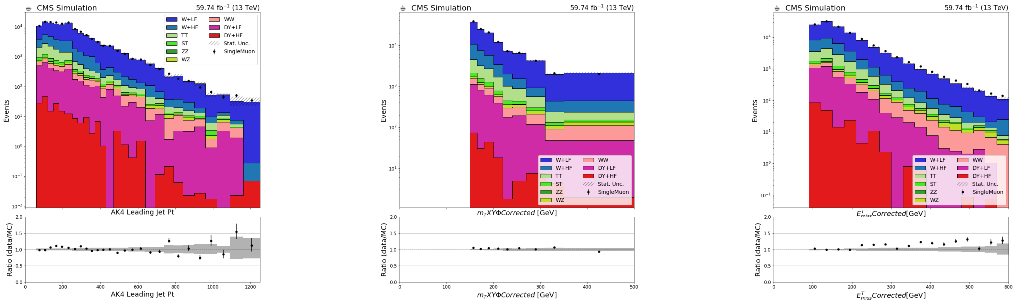
Figure 6.10: This figure a.) shows the p_T distribution of the leading electron. b.) shows the η distribution of the leading electron. c.) shows the ϕ distribution of the leading electron. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the W Control region with electron in the final state for 2018 era. Gray bands indicate statistical uncertainties on simulated background samples

Figure 6.12.e. and 6.13.e show the transverse mass distribution of the W boson for the 2018 era in the $t\bar{t}$ control region with electron and muon in the final state respectively. We see that the data and simulation are well modeled in the region of interest described by the preselection criteria see sec 6.5.1. For m_T^W distribution in the era 2016-2017 for both electron and muon final states in the top control region ref Appendix 1 Fig A.7-A.12

The distribution of the kinematic variables (p_T, η, ϕ)of the electrons (muons) shown in Fig 6.12 (Fig 6.12 for muons). We observe that there is a normalization factor of around 10% which will be taken care of by the maximum likelihood fit but other than that simulation agrees very well with the observed data implying that we can get a very good estimate of the background processes. refer Appendix 1 Fig A.7-A.12 for distribution relevant to the eras 2016-2017.



(a) Leading muon p_T (b) Leading muon η (c) Leading muon ϕ

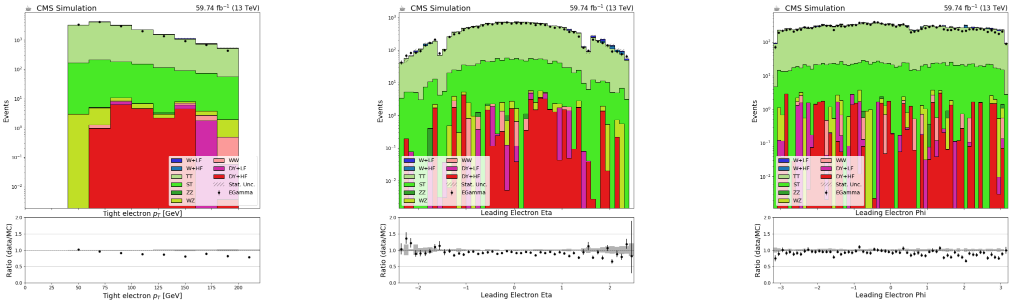


(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

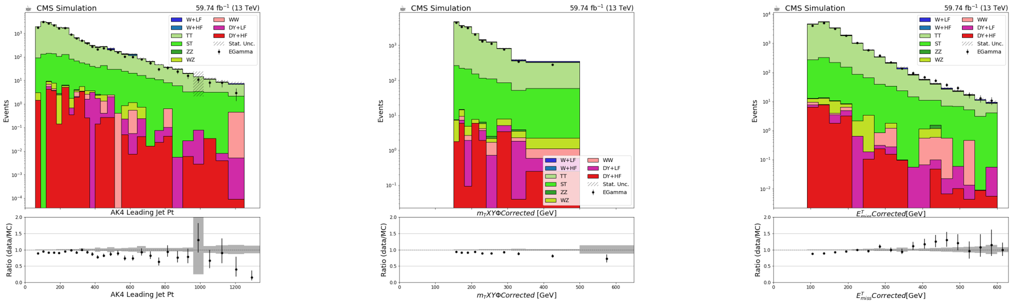
Figure 6.11: This figure a.) shows the p_T distribution of the leading electron. b.) shows the η distribution of the leading muon. c.) shows the ϕ distribution of the leading muon. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the W Control region with muon in the final state for 2018 era. Gray bands indicate statistical uncertainties on simulated background samples

Since, in the top control region we require the presence of at least two AK4 b-tagged jets passing the medium working point criteria. Fig 6.12.d. shows the distribution of the AK4 jet p_T, which seems fairly well modeled.

For distributions of relevant variable please refer to Appendix 1 Fig A.7-A.12 showing that the top control region is indeed dominated by the $t\bar{t}$+jet events with a good data and simulation modeling.



(a) Leading electron p_T (b) Leading electron η (c) Leading electron ϕ

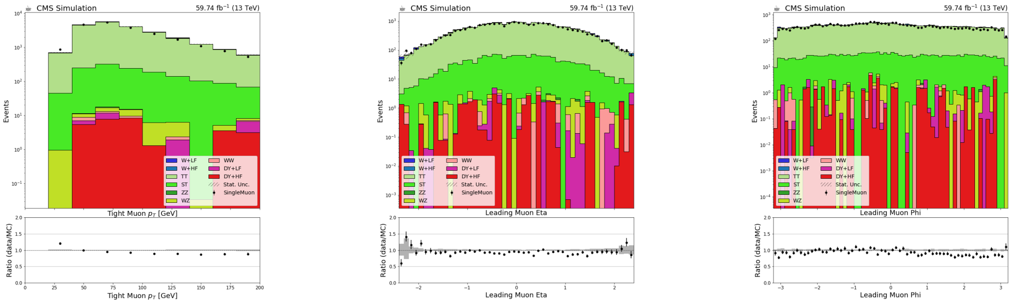


(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

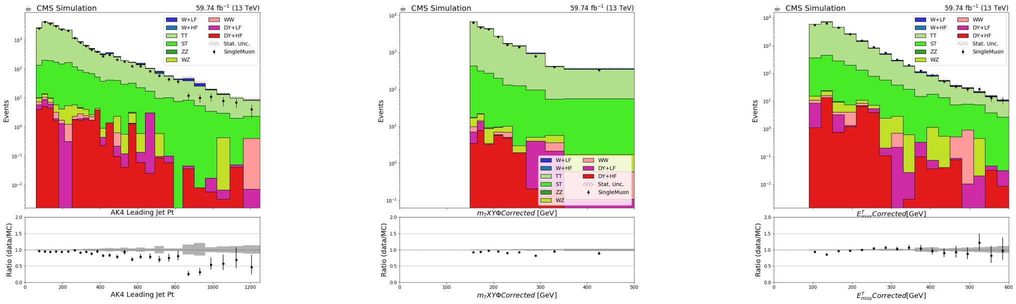
Figure 6.12: This figure a.) shows the p_T distribution of the leading electron. b.) shows the η distribution of the leading electron. c.) shows the ϕ distribution of the leading electron. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the Top Control region with electron in the final state for 2018 era. Gray bands indicate statistical uncertainties on simulated background samples

6.5.3 Signal Region

As the name suggests, signal region refers to the set of selection criteria which tend to maximize the selection of events which are monotop signal-like. In case of leptonic channel we want to maximize the selection of events which have a single AK4 b-tagges jet passing the medium deepJet [58] working point (see sec 6.7, a lepton (electron /muon) which passes the tight lepton selection as defined in section 6.4.1 and 6.4.2 and large missing energy. The one criteria which differentiates the signal region from the other background control regions is the number of AK4 b-tagged jets, in the signal region we require exactly on AK4 medium b-tagged jet.



(a) Leading muon p_T (b) Leading muon η (c) Leading muon ϕ

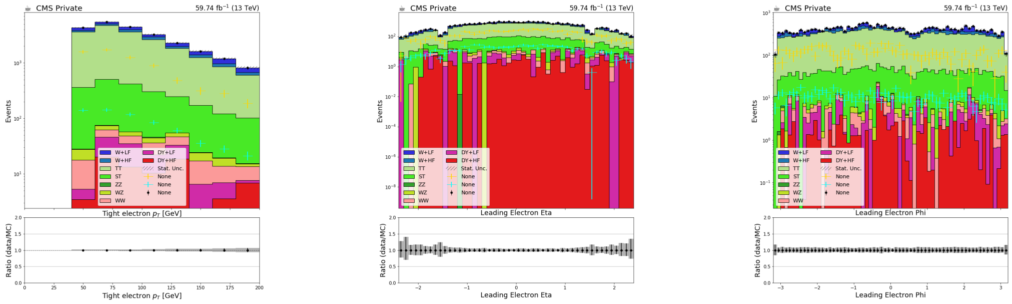


(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

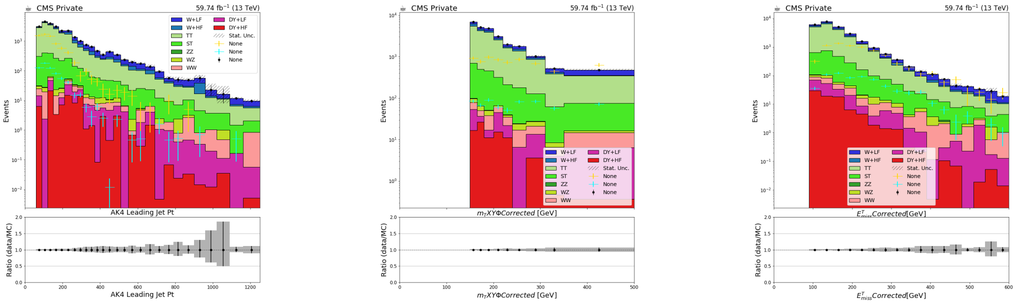
Figure 6.13: This figure a.) shows the p_T distribution of the leading electron. b.) shows the η distribution of the leading muon. c.) shows the ϕ distribution of the leading muon. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the Top Control region with muon in the final state for 2018 era. Gray bands indicate statistical uncertainties on simulated background samples

Figure 6.14 (a,b,c) and 6.15 (a,b,c) shows the distribution of kinematic variable (p_T, η, ϕ) for electrons and muons, respectively for the era 2018. In these plots you can see that we have plotted pseudo data and not the real observed data basically to not induce any bias in our observations. We have also plotted two signal samples labeled as gold and cyan color as representation of the signal distribution. The transverse mass of W boson m_T^W is shown if figure 6.14 (e) and 6.15 (e) in the electron and muon final states for the year 2018.

Similarly, we can also see the p_T of the AK4 jet distribution and the missing transverse momentum in fig 6.14 (d,f) and 6.15 (d,f) in the electron and muon channels respectively for the era 2018. Similar plots for the eras 2016-2017 are available in Appendix 1 Fig A.7-A.12.



(a) Leading electron p_T (b) Leading electron η (c) Leading electron ϕ

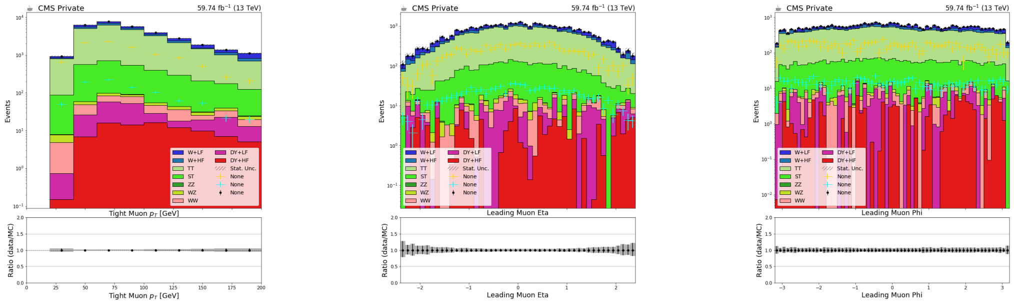


(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

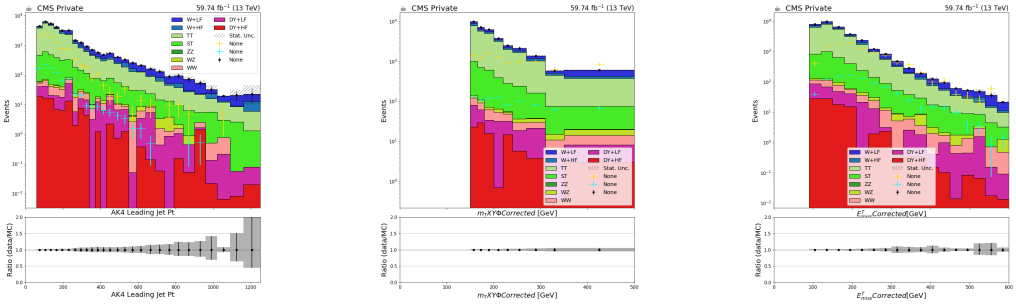
Figure 6.14: This figure a.) shows the p_T distribution of the leading electron. b.) shows the η distribution of the leading electron. c.) shows the ϕ distribution of the leading electron. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the Signal region with electron in the final state for 2018 era. Gray bands indicate statistical uncertainties on simulated background samples. The black dots are the pseudo-data for representation (and not real data). The gold and the cyan colored crosses are the signal samples M_χ=100, M_V=195 and M_χ=50, M_V=200 respectively

6.6 Systematic Uncertainties

Just like most things in the world, imperfection is an inherent feature possessed by most things and the CMS detector is not exclusive to this. The CMS detector is composed of many subcomponents which help us in measuring the properties of particles produced in a proton-proton collision. There are instances when the resolution of these measurements is not as great because of obvious limitations of the components or because of problems in data recording arising from the detector effects. These effects are taken into account by the application of scale factors, but these scale factors have errors associated with the effects they are trying to correct for. Such



(a) Leading muon p_T (b) Leading muon η (c) Leading muon ϕ



(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

Figure 6.15: This figure a.) shows the p_T distribution of the leading electron. b.) shows the η distribution of the leading muon. c.) shows the ϕ distribution of the leading muon. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the Signal region with muon in the final state for 2018 era. Gray bands indicate statistical uncertainties on simulated background samples. The black dots are the pseudo-data for representation (and not real data). The gold and the cyan colored crosses are the signal samples M_χ=100, M_V=195 and M_χ=50, M_V=200 respectively

errors are called systematic uncertainties and in the following section we will describe them. See Tab 6.10

6.6.1 Luminosity

The integrated luminosity collected by the CMS experiment has an uncertainty associated with its measurement. These uncertainty measurements are calculated centrally and are shown in tab 6.9. These uncertainties are uncorrelated across years and are treated for both signal and background simulation samples.

Table 6.9. Luminosity uncertainties used in the analysis

year	uncertainty
2016 preVFP	1.2%
2016 PostVFP	1.2%
2017	2.3%
2018	2.5%

6.6.2 Jet Energy Corrections

In sec 6.4.3.2 & 6.4.3.3 we mentioned the jet energy resolution (JER) and jet energy scale (JES) are applied to the jets. The JES correct the pT of the jets, while the JERs correct the resolution of the simulated samples to match with data, the JER and JES are together called as jet energy corrections or JEC. Since the transverse mass of W boson (m_T^W) depends on the met which have the effects of JEC propagated we need to take into account the uncertainties associated with these corrections. The signal region and background control regions also depend on AK4 jets so it becomes crucial to include the JEC systematic in the analysis. These uncertainties are un-correlated across the 2016-2018 era and therefore are parameterized as independent in the analysis.

6.6.3 Electron Id and reconstruction

As mentioned in sec 6.4.1, we have applied correction factors for electron identification and reconstruction. These correction factors have a certain uncertainty associated with the measurements. The uncertainties are propagated to the analysis which vary the shape of the distributions. The uncertainties in the statistical model are independent across all the eras and are un-correlated.

6.6.4 Muon Id and isolation

As mentioned in sec 6.4.2, we have applied correction factors for electron identification and isolation. These correction factors have a certain uncertainty associated with the measurements. The uncertainties are propagated to the analysis which vary the shape of the distributions. The uncertainties in the statistical model are independent across all the eras and are un-correlated.

6.6.5 Trigger

As described in section 6.3, we have used a combination of correction factors for selecting events with electron and muons see tab. 3.2.10. These trigger scale factor measurements have uncertainties associated with them, these uncertainties are up and down shape variations with respect to the nominal scale factors. The uncertainties affect the shape and rate of the distribution and are independent across data collected across different years and are un-correlated.

6.6.6 B-Tagging

The signal region and the top control region in the leptonic channel require the presence of b-tagged AK4 jet(s). CMS uses the deepJet [62] algorithm for identification of jet which come from hadronization of a b-quark, this b-tagging mechanism has scale factors which are applied to the simulation to propagate real detector b-tagging efficiency to the simulation. The scaling factors have uncertainties in their measurements and need to be parameterized in the statistical model as a variation. The uncertainties are shape up/down shape variations and considered as un-correlated across the all the eras in the analysis.

6.6.7 Pileup

As described in sec 6.2.1 that in order to account for effects of particles produced by the soft p-p collisions causing excess deposits in the ECAL and tracking system we apply correction to events. These correction weights have uncertainties in their measurements. So, we use distribution based shape uncertainties which are un-correlated across all era in the analysis and allows a shape variation over the nominal pileup weights.

6.6.8 L1 Prefire

As described in sec 6.2.2, we described the effect on L1 trigger primitives from the gradual timing shift of the ECAL, this results in self rejection of events due to this mismatch. We apply correction weights to events in data for 2016 and 2017 years, however there is a certain uncertainty in the measurements of these weights so, we take this variation into account, these uncertainties are correlated across years and are parameterized as dependent in the analysis.

6.6.9 Parton distribution Function

Signal and background simulation samples produced at the LHC use Monte Carlo technique but in order to calculate their cross section and precise encoding the momentum structure of the partons inside the hadron needs to done. This ensure that the simulation data are comparable to the experimental data. To ensure that these calculations and encoding of the momentum structure is propagated to the simulation, we use parton density functions (PDF)[70]. At leading order calculations, this information propagation is non-tedious but at higher orders this information propagation through PDF seems to break down. So, to take this effect into account we have theoretical PDF uncertainties associated with PDF used for generation of simulation samples and are correlated across years.

6.7 Statistical Analysis

The objective of any experimental analysis is to have an initial theoretical model and compare the predictions of this theoretical model with the observed data. This comparison is usually done using statistical hypothesis testing [71] by setting a Null hypothesis and an Alternate hypothesis. This comparison is quantified by calculating a p-value and assigning a certain

Table 6.10. Luminosity uncertainties used in the analysis

Uncertainty	2016preVFP	2016PostVFP	2017	2018	correlation
Luminosity	✓	✓	✓	✓	✗
JEC	✓	✓	✓	✓	✗
Electron Id & Reconstruction	✓	✓	✓	✓	✗
Muon Id & Isolation	✓	✓	✓	✓	✗
Trigger	✓	✓	✓	✓	✗
B-tagging	✓	✓	✓	✓	✗
Pileup	✓	✓	✓	✓	✗
L1 Prefire	✓	✓	✓	✓	✓
PDF	✓	✓	✓	✓	✓

level of confidence in the results based on a significance level (α). In this analysis we have used a statistical software called `combine` [72] for performing the necessary statistical test and extracting the final limits. This software uses the input as in the form text files called data cards 6.16 and work-spaces which are in `ROOT` [73]format. The general workflow used in this analysis is as follows:

- Perform the background only asimov fit on the combined Signal region and Background control region
  1. Create individual text-file data cards and work-spaces for all control and signal regions
  2. Combine all individual data cards based for each era you are considering using the combine commands. for ex: combine all datacards for 2018

     CombineCards.py CR1.txt CR2.txt Signal.txt > combined.text
  3. Now convert this combined file to a combine workspace (assuming you have `ROOT` workspaces for each region like (CR1.root, CR2.root and Signal.root)

     text2workspace combined.txt -o workspace.root

  4. now use the combine `FitDiaganostics` command to either do a background only fit or a signal+background only asimov (`-t -1` option in the command below) fit.

     Background-only fit: (expectSignal 0)

     ```
     combine -M FitDiagnostics workspace.root –expectSignal 0 –minos=all –cminDefaultMinimizerStrategy 0 –forceRecreateNLL -t -1
     ```

     Signal+Background-only fit:(expectSignal 1)

     ```
     combine -M FitDiagnostics workspace.root –expectSignal 1 –minos=all –cminDefaultMinimizerStrategy 0 –forceRecreateNLL -t -1
     ```

  5. Now get the nuisance parameters using the following command

     ```
     python diffNuisances.py fitDiagnostics.root –all
     ```

  6. Now plots the good looking pull and impacts plots with the built-in `diffNuisances.py` script.
  7. Once satisfied with the impacts and pulls, next step is do the goodness of fit tests, start with individual control regions and then do it for the all the regions combined still on the asimov toy datasets.

     ```
     python diffNuisances.py fitDiagnostics.root –sortBy=impact -g Plot_2018.root –all
     ```

  8. Now create plots for prefit and postfit distributions in the background-only fit
  9. Perform goodness-of-fit tests for the postfit distributions in the individual control region and then combined control regions
  10. When everything looks good, calculate expected exclusion limits for each signal sample

- Perform the background only fit on the combined Signal region and Background control region. This step is used for calculating the best fit values of the parameters using maximum likelihood estimation method. The general idea is to perform a simultaneous fit in the combination of signal and background control regions, then the best

```
imax 1
jmax 1
kmax *
--------------------------------------------------------------------------------
shapes * * simple-shapes-df_input.csv $CHANNEL:$PROCESS:nominal,sum_w:sum_ww $CHANNEL:$PROCESS:$SYSTEMATIC,sum_w:sum_ww
--------------------------------------------------------------------------------
bin         bin1
observation 85
--------------------------------------------------------------------------------
bin             bin1   bin1
process         signal background
process         0      1
rate            10     100
--------------------------------------------------------------------------------
lumi    lnN     1.10   1.0
bgnorm  lnN     1.00   1.3
alpha   shapeN2 -      1   uncertainty on background shape and normalization
sigma   shapeN2 0.5    -   uncertainty on signal resolution. Assume the histogram is a 2 sigma shift,
#                          so divide the unit gaussian by 2 before doing the interpolationkk
```

Figure 6.16: a sample combine datacard

estimate values of the parameters in the dominating backgrounds from the control region are propagated to the signal region.

In the following section we will take you through a step-by-step process for validating our statistical model. First, we will present the Maximum likelihood fits with Asimov toys [74], then we shall analyze the pulls and impacts of the nuisance parameters and find the best-fit value of the nuisance parameters, we then follow this with the simultaneous goodness-of-fit of for combined signal and background control regions. Next, we present the expected limits followed by the observed limits and stating an exclusion region for this search.

6.7.1 Tests on Asimov Toy datasets

In this step we want to understand the nuisance parameters in our statistical model, we perform a maximum likelihood fit in all the control and signal regions at a signal strength value r=0. By asimov[1] datasets we mean that the model isn't exposed to real data but data is sampled from the existing simulation distribution in each control region. Fig 6.17-6.19, shows the constraint of the pulls in the combined signal and control regions for 2018, 2017 and 2016 eras,

---

[1] The name of the Asimov data set is inspired by the short story Franchise, by Isaac Asimov . In the book, the election voting is done by the most representative person.

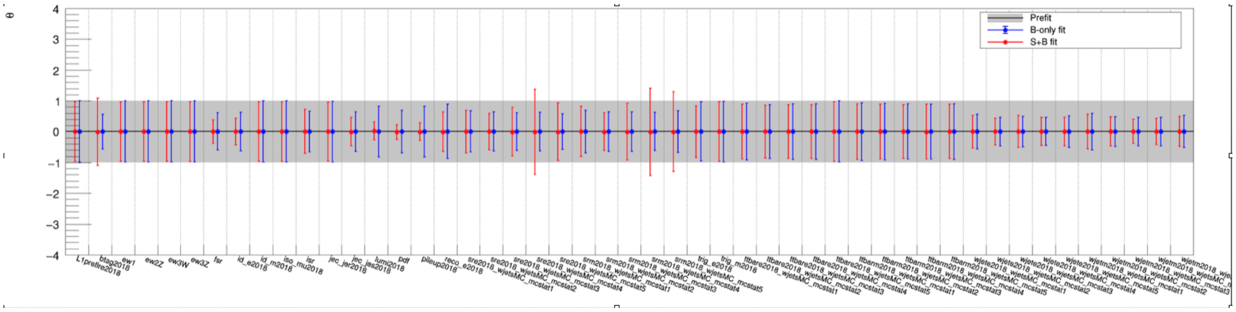


Figure 6.17: Pulls for 2018 era using Asimov toy

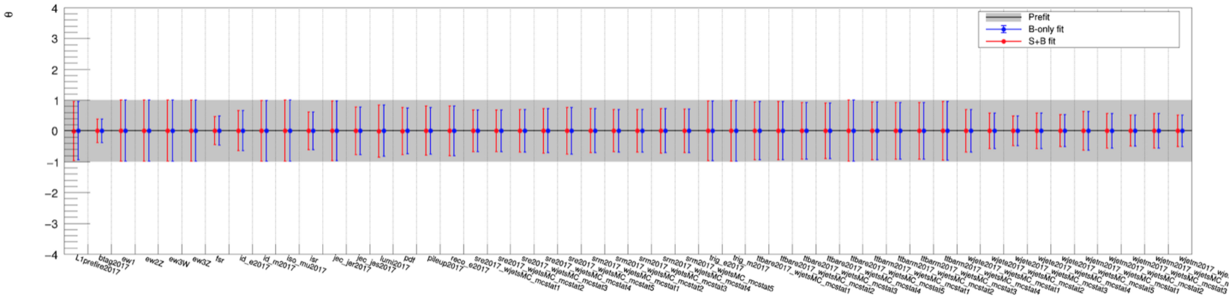


Figure 6.18: Pulls for 2017 era using Asimov toy

respectively. Here we observe that the all the systematic uncertainties are behaving well. Ref Fig 6.21-6.22 for the systematic with highest impact respective eras.

6.7.2 Goodness of Fit tests

Goodness-of-fit test are performed in two cases, first we want to understand how well is our background modeled in the control region only. W($l\nu$) → jets and T$\bar{T}$ are the dominant backgrounds in our region of interest (signal region). We will using the estimates from these control region to be used in the signal region so, it's highly critical to have the high p-value for the goodness-of-fit test statistic. We have use here the Kolmogorov-Smirnoff (KS) [75] test for the goodness-of-fit. Tab 6.11 shows the p-values when using the KS test, we observe that the gof test fails for top CR (μ) in 2018, W CR (μ) in 2016 postVFP and W CR (e) for 2016 preVFP. This

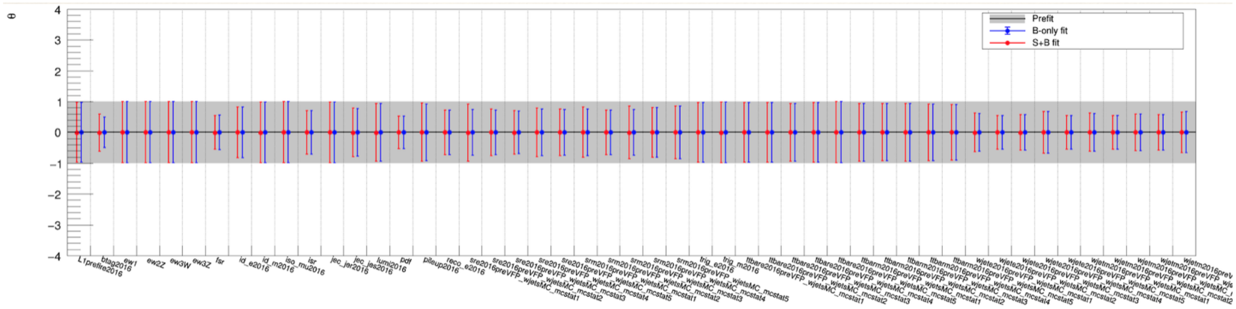


Figure 6.19: Pulls for 2016PreVFP era using Asimov toy

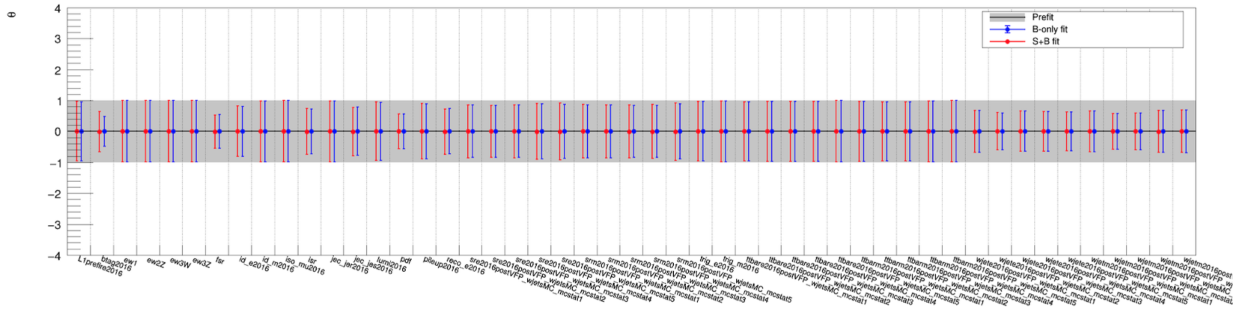


Figure 6.20: Pulls for 2016PostVFP era using Asimov toy

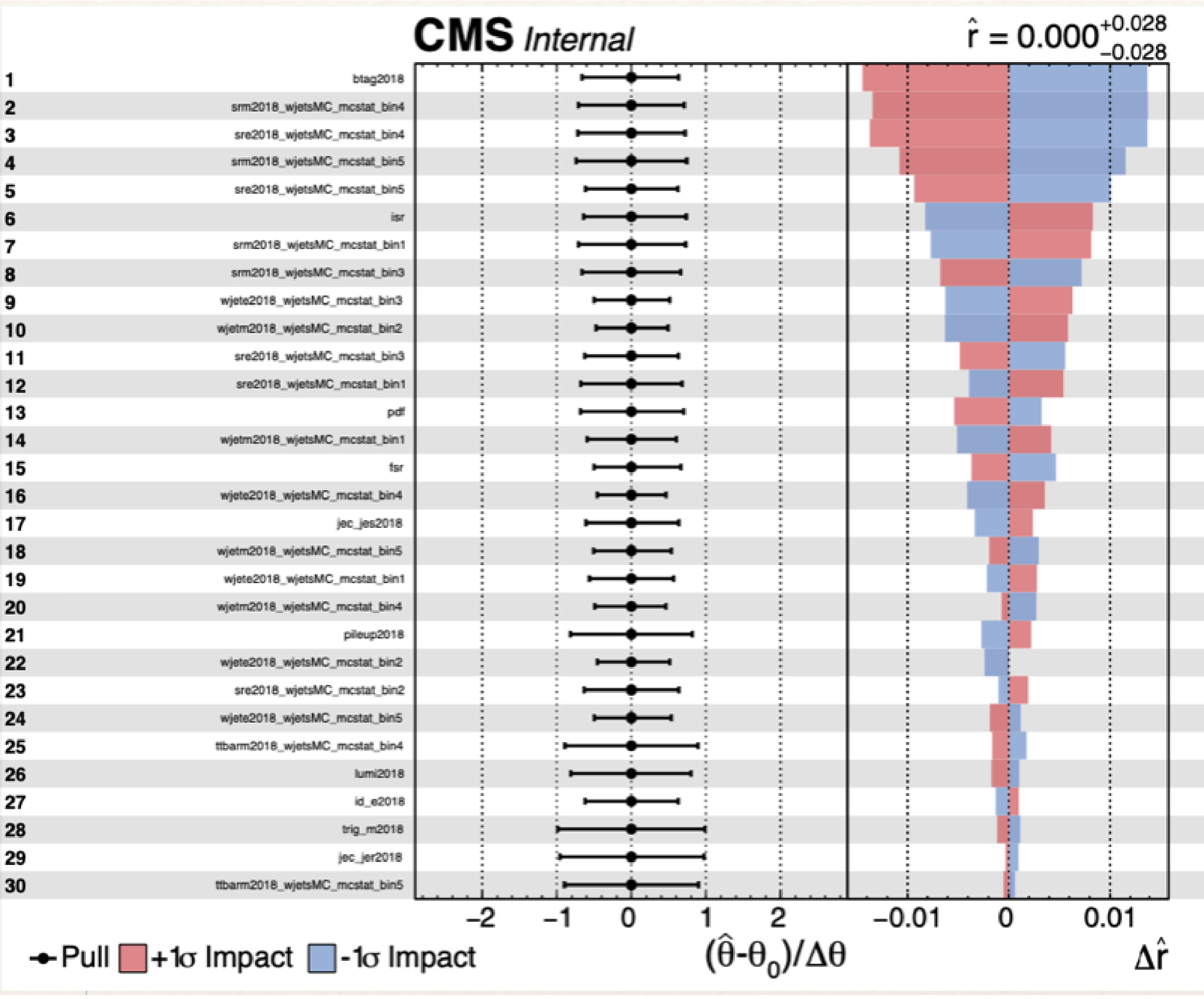


Figure 6.21: impacts for 2017 era using Asimov toy

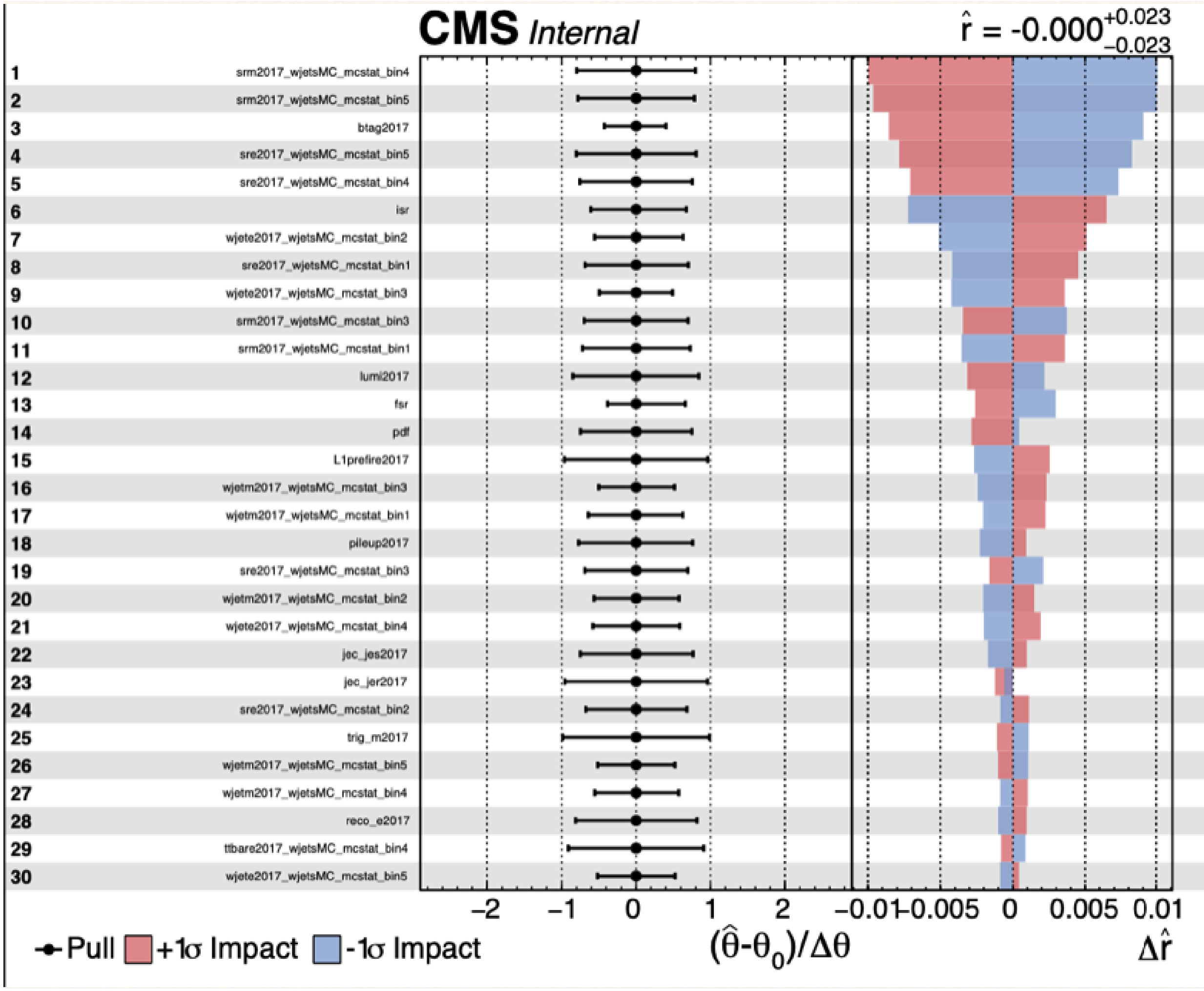


Figure 6.22: impacts for 2017 era using Asimov toy

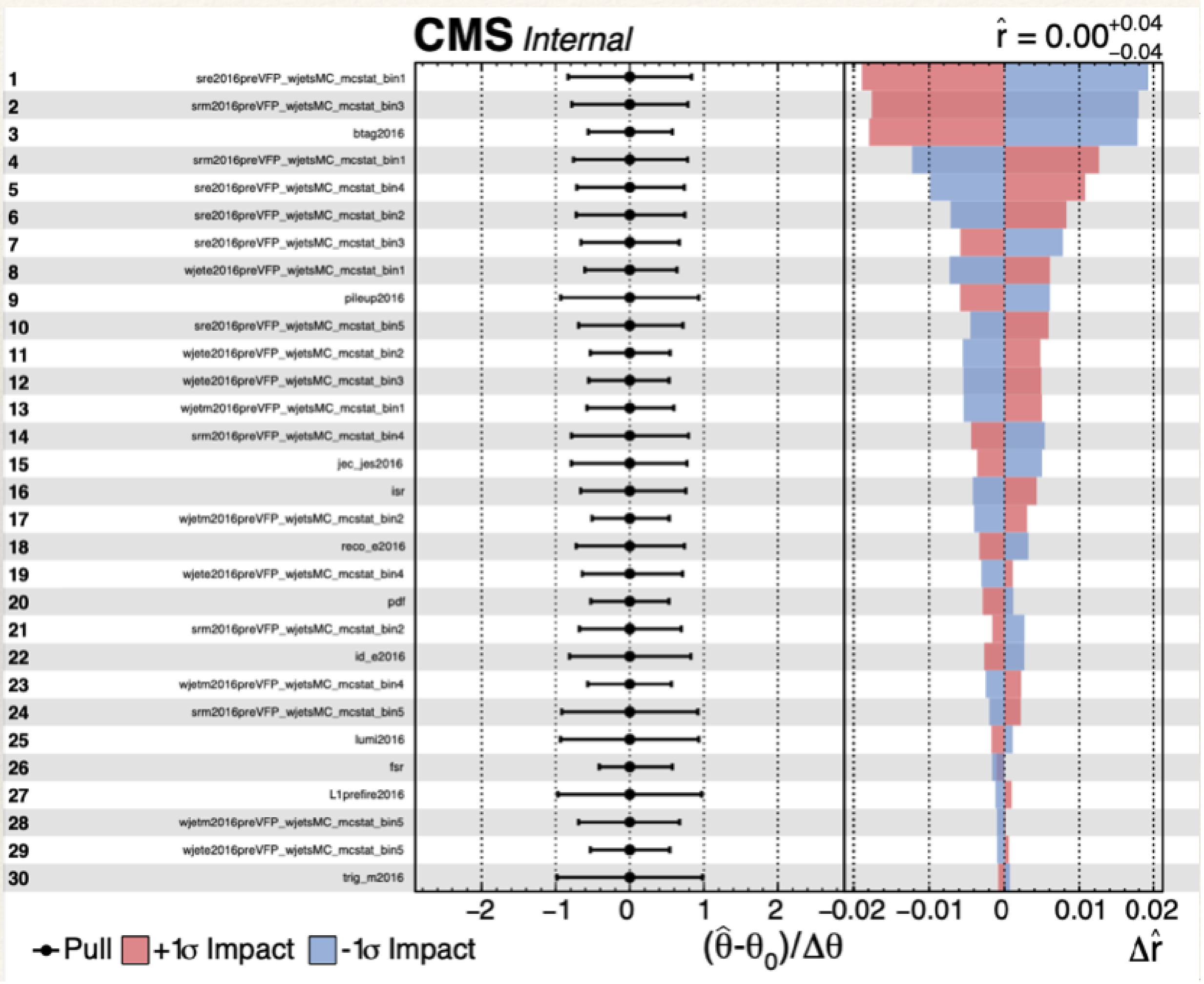


Figure 6.23: impacts for 2016PreVFP era using Asimov toy

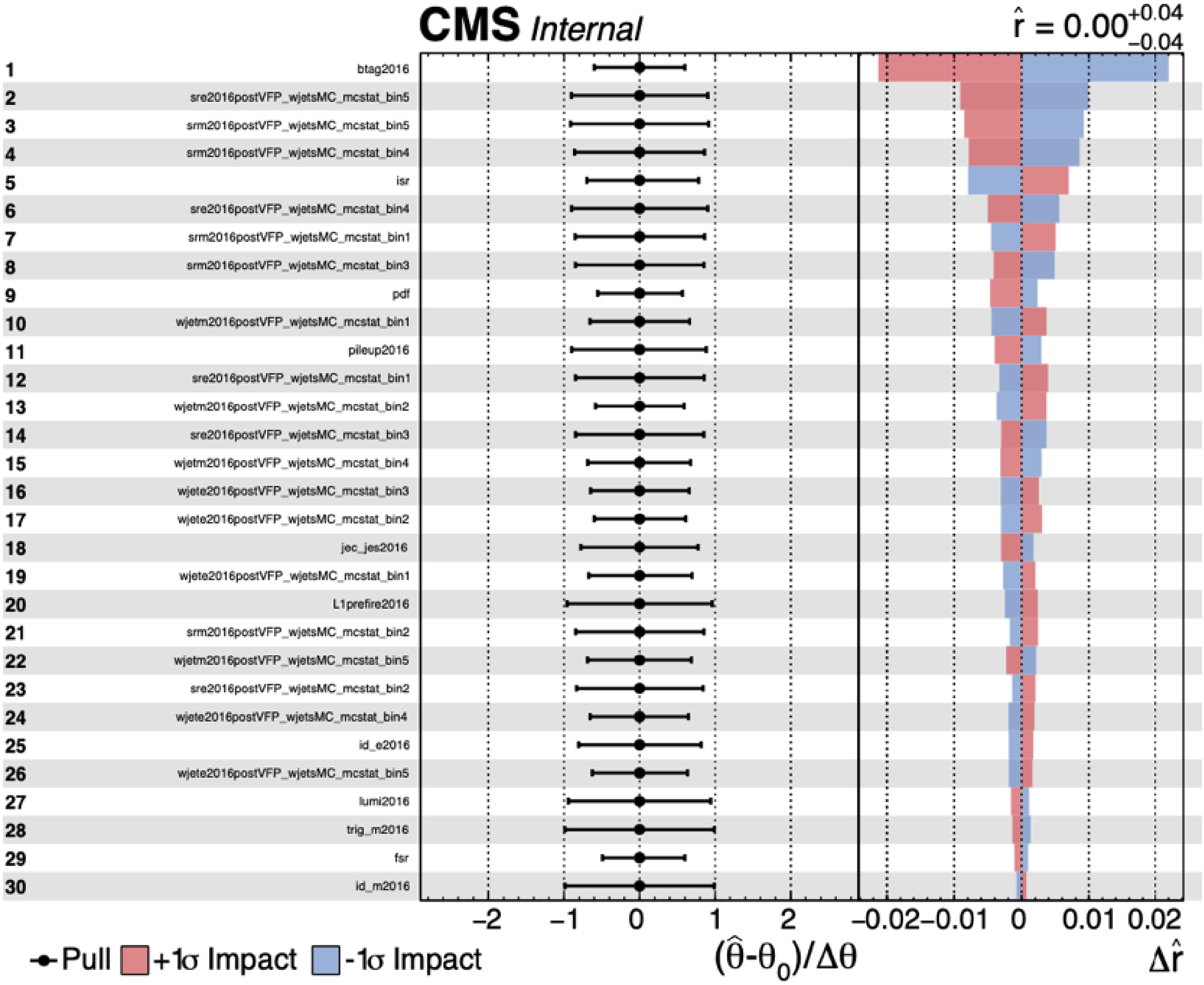


Figure 6.24: impacts for 2016PostVFP era using Asimov toy

would need further investigation in the future. Next, we perform the goodness-of-fit test in the

Table 6.11. Kolmogorov-Smirnoff Goodness of fit test control regions

Year	WCR(e)	WCR(μ)	TCR(e)	TCR(μ)
2018	0.295	0.012	0.155	0.312
2017	0.509	0.635	0.082	0.457
2016PreVFP	0.321	0.335	0.177	0.009
2016PostVFP	0.119	0.549	0.035	0.280

simultaneous region including control and signal regions. Tab 6.12, shows the gof p-values for the signal and control regions for all eras. We observe here that the gof test p-value for the SR(μ), WCR(e) and WCR(μ) fail for 2016 preVFP era and WCR(e) and WCR(μ) for 2016 postVFP era. To accommodate this issue we would also present the upper exclusion limits for 2018 and 2017 combined era. Overall, the gof test values look good implying good modeling in the signal region.

Table 6.12. Kolmogorov-Smirnoff Goodness of fit test including signal and control regions

Year	SR(e)	SR(μ)	WCR(e)	WCR(μ)	TCR(e)	TCR(μ)
2018	0.228	0.194	0.914	0.892	0.122	0.156
2017	0.932	0.700	0.516	0.316	0.208	0.836
2016PreVFP	0.602	0.002	0.003	0.002	0.353	0.651
2016PostVFP	0.052	0.246	0.054	0.032	0.527	0.598

6.7.3 Best fit impacts

Figures 6.25-6.28 show the impacts after the best-fit of the nuisance parameters.

6.7.4 Prefit and PostFit distributions

Figures 6.29-6.32 show the postfit distribution with bestfit values of the nuisance parameters.

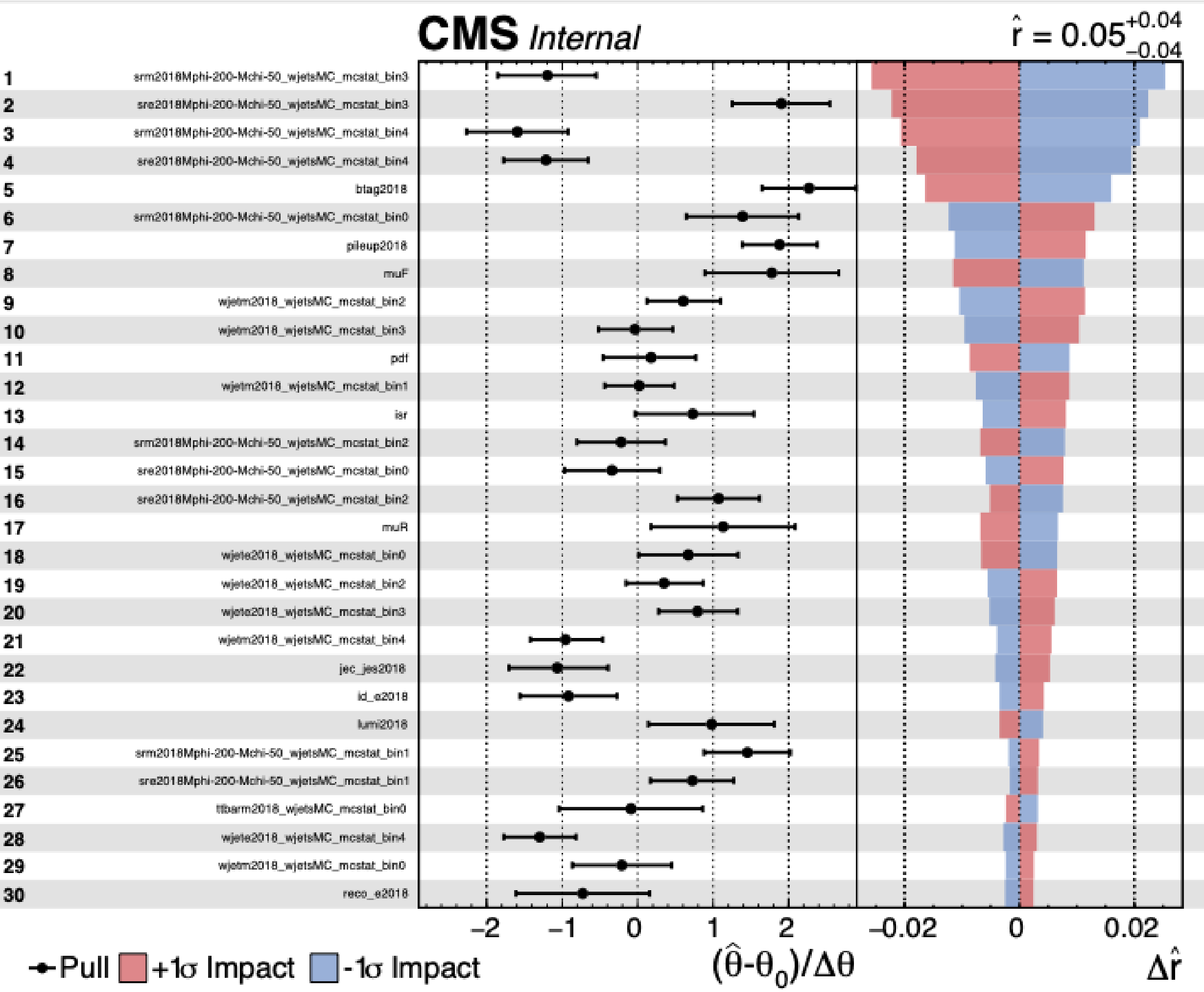


Figure 6.25: impacts for 2018 era showing the best-fit values of the systematic uncertainties

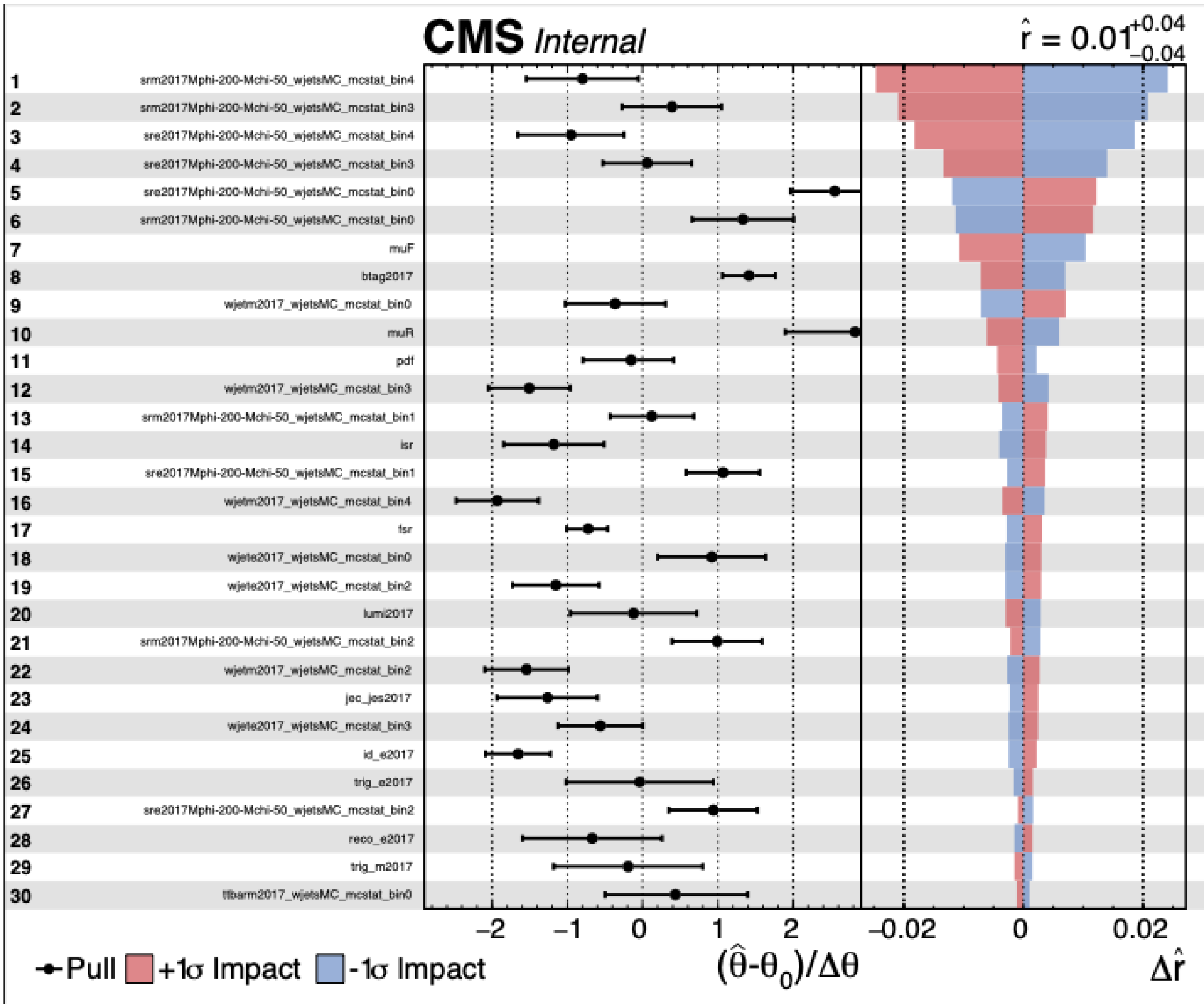


Figure 6.26: impacts for 2017 era showing the best-fit values of the systematic uncertainties

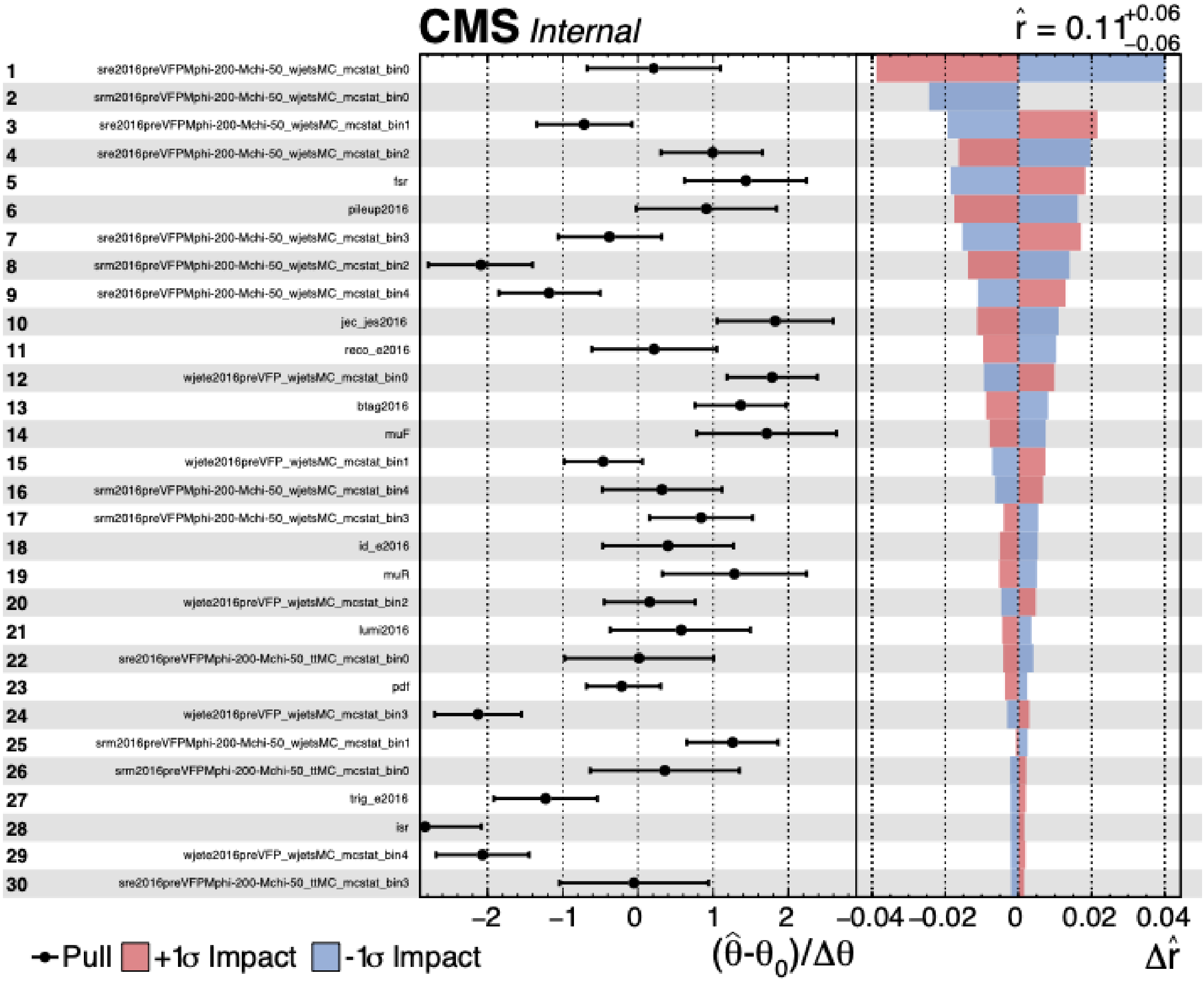


Figure 6.27: impacts for 2016PreVFP era showing the best-fit values of the systematic uncertainties

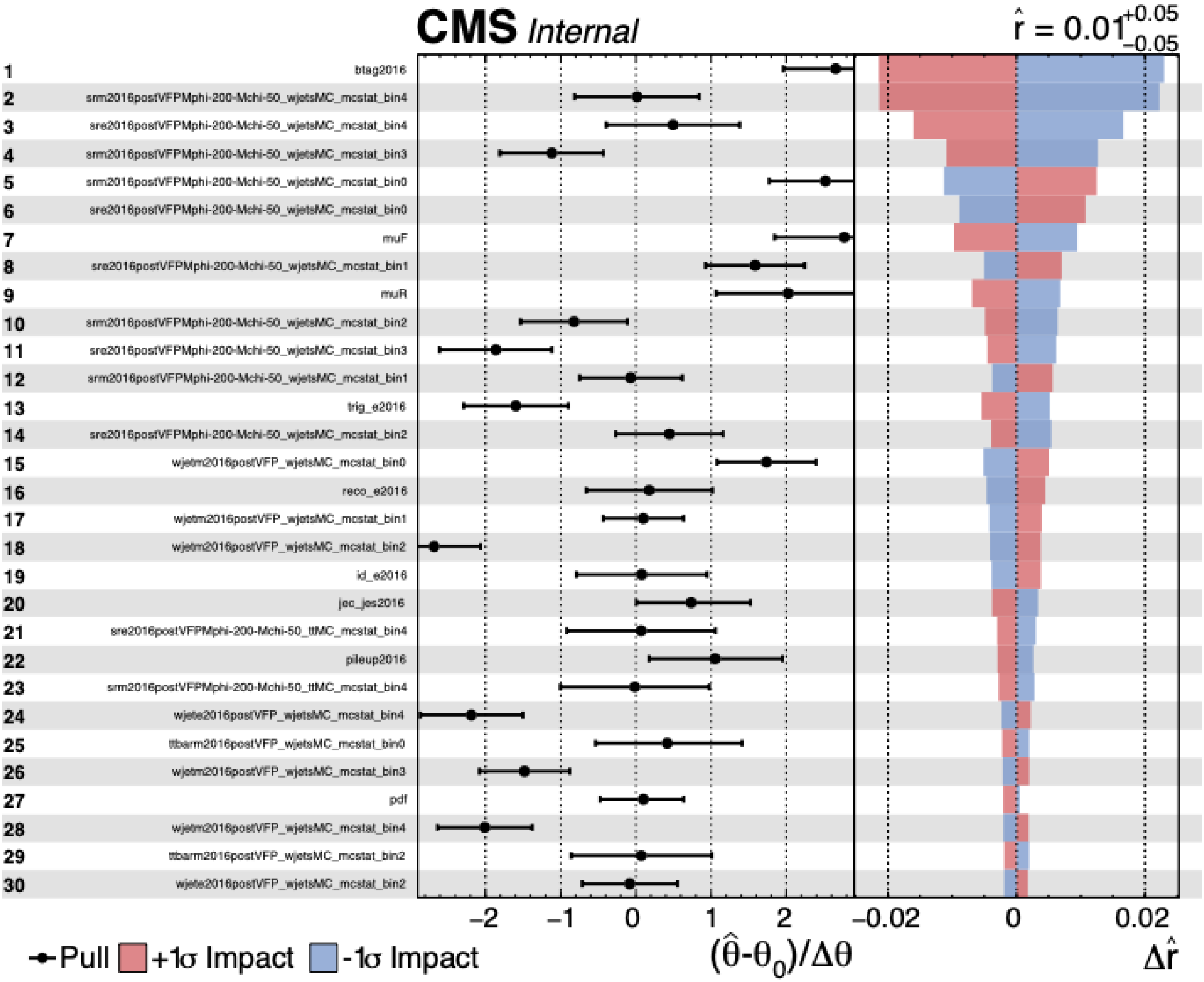


Figure 6.28: impacts for 2016PostVFP era showing the best-fit values of the systematic uncertainties

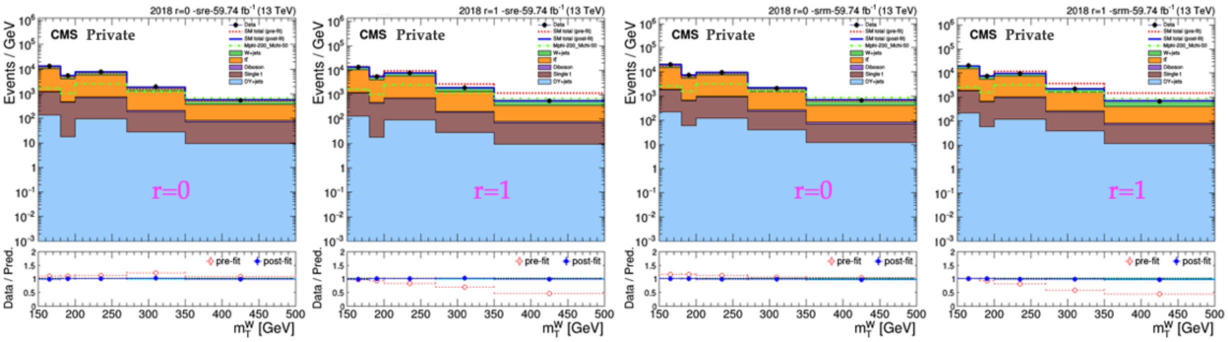


Figure 6.29: PostFit distributions for 2018: This Figure shows the distribution Data/MC before and after the fit in signal region. The first and second figure is the Signal region with electron in final state with signal strength r=0 and r=1, respectively. The third and fourth figure is the Signal region with muon in final state with signal strength r=0 and r=1, respectively. The red dotted line is the prefit and the blue dotted line is postfit. The cyan colored band is the postfit uncertainty, the neon green colored line is the signal $M_V = 200\,GeV\,and\,M_\phi = 50\,GeV$ and other backgrounds are stacked on top of each other.

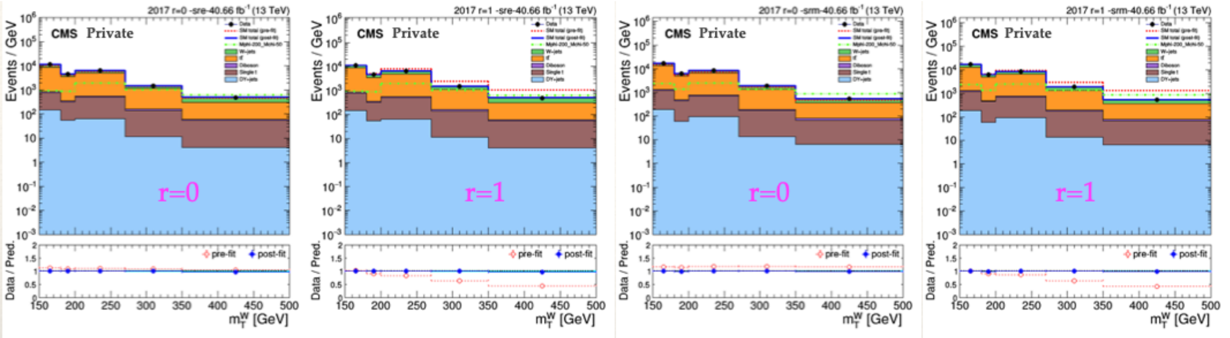


Figure 6.30: PostFit distributions for 2017: This Figure shows the distribution Data/MC before and after the fit in signal region. The first and second figure is the Signal region with electron in final state with signal strength r=0 and r=1, respectively. The third and fourth figure is the Signal region with muon in final state with signal strength r=0 and r=1, respectively. The red dotted line is the prefit and the blue dotted line is postfit. The cyan colored band is the postfit uncertainty, the neon green colored line is the signal $M_V = 200\,GeV\,and\,M_\phi = 50\,GeV$ and other backgrounds are stacked on top of each other.

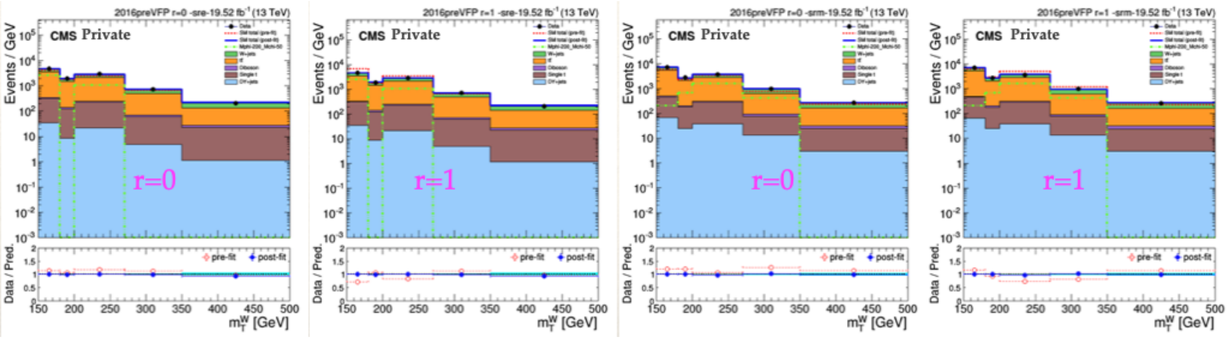


Figure 6.31: PostFit distributions for 2016preVFP: This Figure shows the distribution Data/MC before and after the fit in signal region. The first and second figure is the Signal region with electron in final state with signal strength r=0 and r=1, respectively. The third and fourth figure is the Signal region with muon in final state with signal strength r=0 and r=1, respectively. The cyan colored band is the postfit uncertainty, the neon green colored line is the signal $M_V = 200\, GeV\ and\ M_\phi = 50\, GeV$ and other backgrounds are stacked on top of each other.

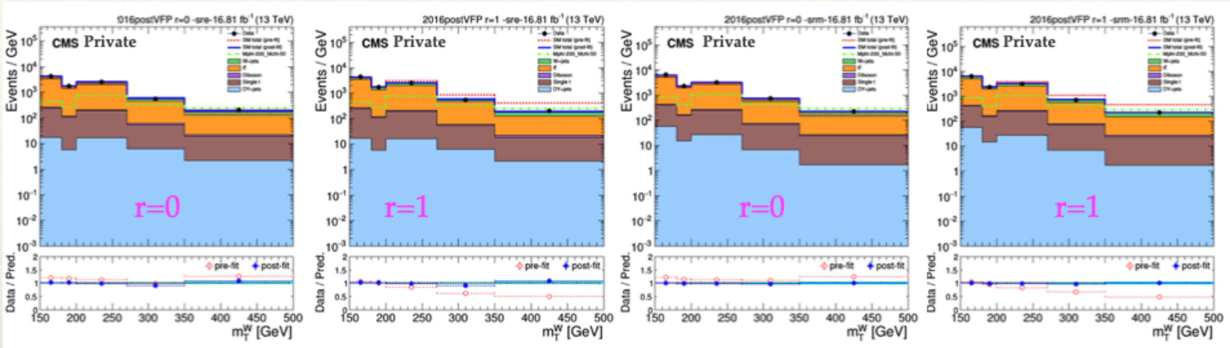


Figure 6.32: PostFit distributions for 2016postVFP: This Figure shows the distribution Data/MC before and after the fit in signal region. The first and second figure is the Signal region with electron in final state with signal strength r=0 and r=1, respectively. The third and fourth figure is the Signal region with muon in final state with signal strength r=0 and r=1, respectively. The cyan colored band is the postfit uncertainty, the neon green colored line is the signal $M_V = 200\, GeV\ and\ M_\phi = 50\, GeV$ and other backgrounds are stacked on top of each other.

6.8 Results

6.8.1 Upper Exclusion limits

Exclusion upper limits are set at 95 % confidence level using the modified frequentist CL_s method as discussed in sec 5.3.2 and [41]-[43], these limit are calculated using the LHC collaboration wide used statistical tool called `combine` [72].

In order to gauge the exclusion power of the model prior to calculating the expected limits, the upper exclusion limits are first calculated for the background only model also called the expected limits. Fig 6.33 shows the expected asymptotic limits obtained by using the `combine` tool, This figure shows a 2D grid representing the values of the median expected upper limits for different mass points see Tab 6.1. The solid black line is the profile showing the region where the median expected upper limit on the signal strength modifier (μ) at 95 % confidence level is equal to 1 and the dashed black lines represent the profile the median expected upper limit on the signal strength modifier (μ) at 68 % confidence level is equal to 1. While we only generated a few signal mass points, we performed interpolation based on the following equations to get a continuous mass grid. The calculations are based on [76] and [77].

$$width_{dm} = \frac{M_V}{24\pi} \times \sqrt{1 - \frac{4M_\chi^2}{M_V^2}} \times g_{dm}^2 \times \left(2\left(1 - \frac{M_\chi^2}{M_V^2}\right) + 6\frac{M_\chi^2}{M_V^2}\right) \tag{6.8}$$

$$width_{quarks} = \frac{M_V}{4\pi} \times g_{quark}^2 \times \left(1 - \frac{172.5^2}{M_V^2}\right) \times \left(1 - \frac{172.5^2}{2M_V^2} - \frac{172.5^4}{2M_V^4}\right) + \frac{M_V}{4\pi} \times g_{quark}^2 \tag{6.9}$$

$$BR(M_V, M_\chi) = \frac{width_{dm}}{width_{quarks} + width_{dm}} \tag{6.10}$$

where g_{dm} =1 and g_{quarks} =0.25, M_V is the mediator mass and M_χ is dark matter mass from Fig 6.33, at M_V = 200 GeV, we have generated MC signal sample M_χ = [50, 150]. We want to find the expected limit values for M_χ = [0, 25, 50, 75] based on the value of the limit at (200, 50). We restrict our interpolations to $M_\chi < \frac{M_V}{2}$.

for point M_V =200, M_χ=0:

$$Expected\ Limit(M_V = 200, M_\chi = 0) = \frac{Expected\ Limit(200, 50)}{BR(200, 50) \times BR(200, 0)} \tag{6.11}$$

Similarly for point M_V =200, M_χ=25:

$$Expected\ Limit(M_V = 200, M_\chi = 25) = \frac{Expected\ Limit(200, 50)}{BR(200, 50) \times BR(200, 25)} \tag{6.12}$$

Finally, fig 6.34 shows the observed exclusion limit for the monotop signal across the combination of 2016-2018 eras. Mediator mass of below 900 GeV and DM candidate masses below 400 GeV are excluded at 95% confidence level. The observed limits are within 68% variation band of the expected limits. As the gof test 6.12 is for 2016 needs further investigation, we also show the limits for only 2017-2018 era where the gof test look good. See fig 6.35.

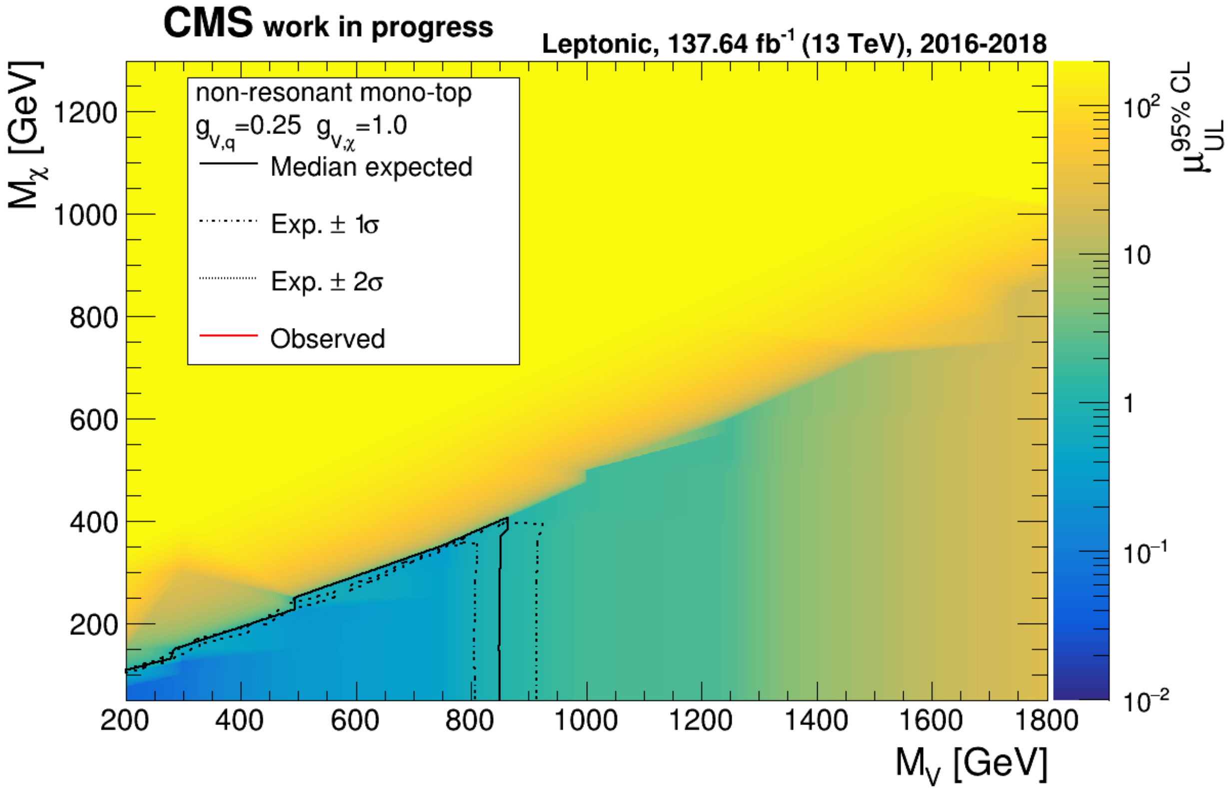


Figure 6.33: This plot is a 2-d plot showing the expected upper exclusion limits on the signal strength modifier $\mu_{UL}^{95\%CL}$ at 95 % confidence level. The x-axis is the mediator mass (M_V) and the y-axis is the dark matter candidate mass M_χ. The color gradient is the median expected upper limits. The solid black line represents the profile for which the median expected upper limit is one, the dashed black line is the profile for which the expected upper limit is one with 68 % confidence level. The area under the black dashed line is the exclusion region at 95% confidence level.

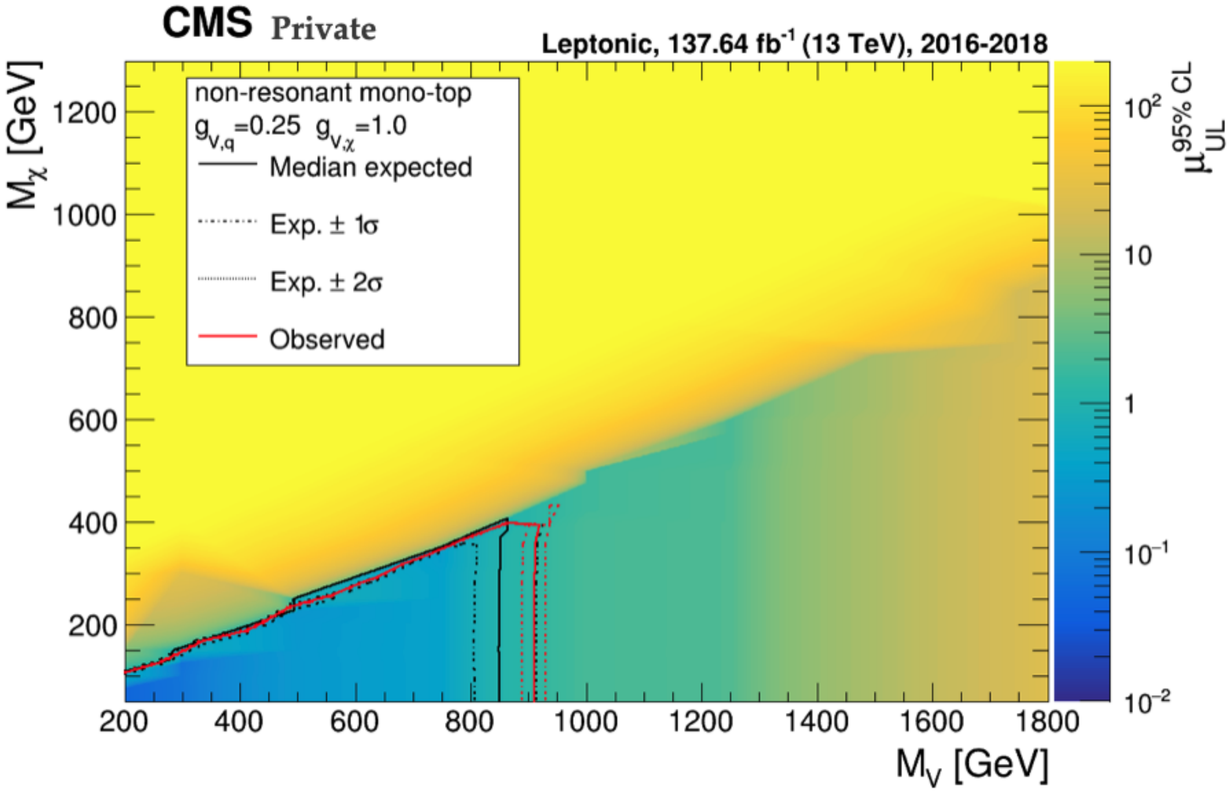


Figure 6.34: This plot is a 2-d plot showing the expected upper exclusion limits on the signal strength modifier $\mu_{UL}^{95\%CL}$ at 95 % confidence level. The x-axis is the mediator mass (M_V) and the y-axis is the dark matter candidate mass M_χ. The color gradient is the median expected upper limits. The solid black line represents the profile for which the median expected upper limit is one, the dashed black line is the profile for which the expected upper limit is one with 68 % confidence level. The area under the black dashed line is the expected exclusion region at 95% confidence level and the area under the red dashed line is the observed exclusion region at 95% confidence level

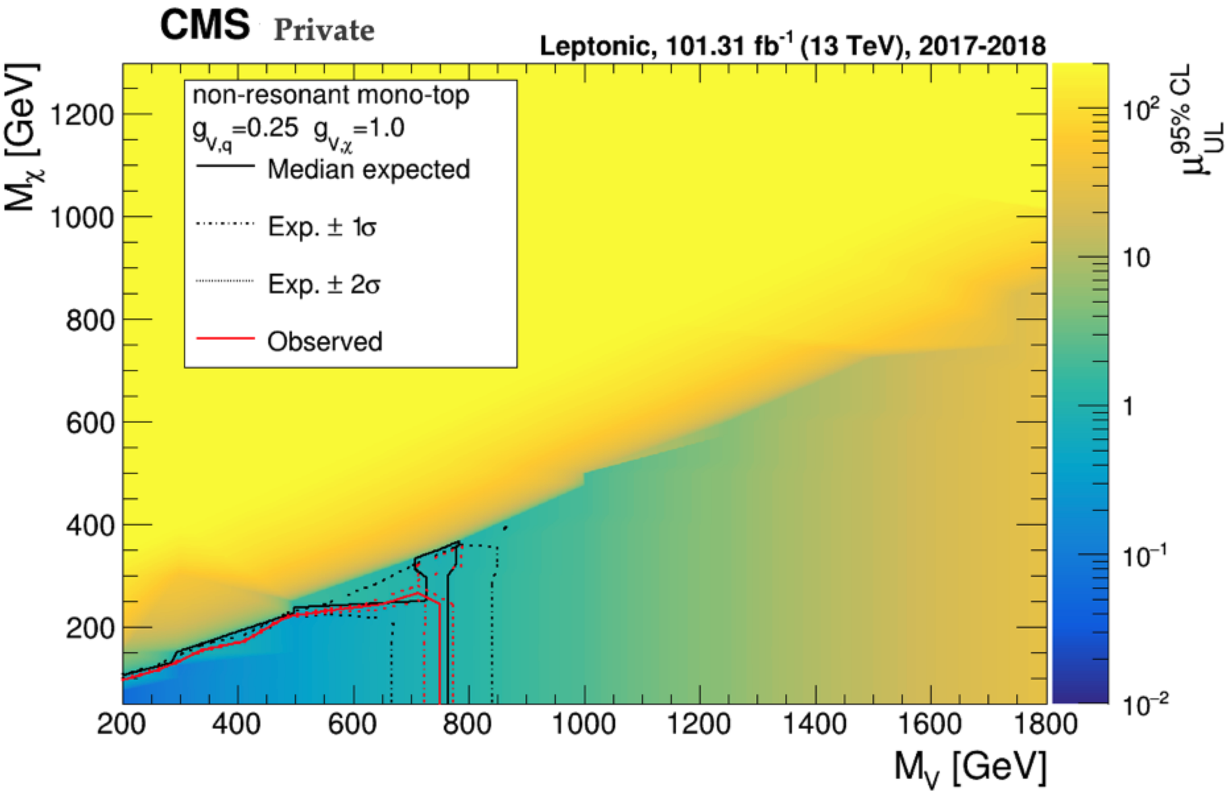


Figure 6.35: This plot is a 2-d plot showing the expected upper exclusion limits for the 2017-2018 era, on the signal strength modifier $\mu_{UL}^{95\%CL}$ at 95 % confidence level. The x-axis is the mediator mass (M_V) and the y-axis is the dark matter candidate mass M_χ. The color gradient is the median expected upper limits. The solid black line represents the profile for which the median expected upper limit is one, the dashed black line is the profile for which the expected upper limit is one with 68 % confidence level. The area under the black dashed line is the exclusion region at 95% confidence level.

CHAPTER 7

Conclusion

In this dissertation we have presented a search for dark matter using the monotop model with electron and muon in the final states by analysing the data collected in the years 2016-2018 amounting to around 137 fb^{-1}. The analysis has been performed by utilizing a python-based data analysis framework called `coffea` [78]. Previous analysis had been performed for only the muon final states thus making this analysis to cover the full leptonic channel.

After the statistical analysis we presented the upper exclusion limits for the entire leptonic channel. Next steps for this analysis would be first understand the discrepancies in the 2016 data analysis, as was observed in the goodness of fit test 6.12 and then combining with the hadronic channel. As discussed in chapter 4 that the CMS is expected to undergo an upgrade in preparations of the HL-LHC thus allowing for 10 times more data to be analyzed and thus help in exploring the higher energy scales for the search for new physics beyond the SM.

REFFERENCES

[1] Wikipedia Commons. Sm. `https://en.wikipedia.org/wiki/Standard_Model`. [Online; accessed 21-Nov-2022].

[2] Wikipedia Commons. Sm. `https://en.wikipedia.org/wiki/Electroweak_interaction`. [Online; accessed 21-Nov-2022].

[3] Fred Jegerlehner. The hierarchy problem of the electroweak standard model revisited, 2013.

[4] Michael Dine and Alexander Kusenko. Origin of the matter-antimatter asymmetry. *Reviews of Modern Physics*, 76(1):1–30, dec 2003.

[5] Particle Data Group, W. M. Yao, C. Amsler, D. Asner, R. M. Bamett, J. Beringer, P. R. Burchat, Cd Carone, C. Caso, O. Dahl, G. D'ambrosio, A. Dcgouvca, M. Doscr, S. Eidelman, J. L. Feng, T. Ghcrghctta, M. Goodman, C. Grab, D. E. Groom, A. Gurtu, K. Hagiwara, K. G. Hayes, J. J. Hernandez-Rey, K. Hikasa, H. Jawahcry, C. Kolda, Y. Kwon, M. L. Mangano, A. V. Manohar, A. Masoni, R. Miqucl, K. Monig, H. Murayama, K. Nakamura, S. Navas, K. A. Olive, L. Rape, C. Patrignani, A. Picpkc, G. Punzi, G. Raffclt, J. G. Smith, M. Tanabashi, J. Tcrning, N. A. Tornqvist, T. G. Trippc, P. Vogcl, T. Watari, C. G. Wohl, R. L. Workman, and P. A. Zyla. Review of particle physics. *Journal of Physics G: Nuclear and Particle Physics*, 33(1), July 2006. Publisher Copyright: © 2006, Institute of Physics Publishing. All rights reserved.

[6] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. A universal density profile from hierarchical clustering. *The Astrophysical Journal*, 490(2):493–508, dec 1997.

[7] V. C. Rubin, Jr. Ford, W. K., and N. Thonnard. Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa -¿ Sc. *ApJ*, 225:L107–L111, November 1978.

[8] W. J. G. de Blok. The core-cusp problem. *Advances in Astronomy*, 2010:1–14, 2010.

[9] Alan A Coley and David L Wiltshire. What is general relativity? *Physica Scripta*, 92(5):053001, 2017.

[10] Richard Massey, Thomas Kitching, and Johan Richard. The dark matter of gravitational lensing. *Reports on Progress in Physics*, 73(8):086901, jul 2010.

[11] Douglas Clowe, Maruš a Bradač, Anthony H. Gonzalez, Maxim Markevitch, Scott W. Randall, Christine Jones, and Dennis Zaritsky. A direct empirical proof of the existence of dark matter. *The Astrophysical Journal*, 648(2):L109–L113, aug 2006.

[12] Richard Massey, Thomas Kitching, and Johan Richard. The dark matter of gravitational lensing. *Reports on Progress in Physics*, 73(8):086901, jul 2010.

[13] A. Yu. Ignatiev and R. R. Volkas. Mirror dark matter and large scale structure. *Physical Review D*, 68(2), jul 2003.

[14] Joel R. Primack and Michael A. K. Gross. Hot dark matter in cosmology, 2000.

[15] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. The structure of cold dark matter halos. *The Astrophysical Journal*, 462:563, may 1996.

[16] Oliver Buchmueller, Caterina Doglioni, and Lian-Tao Wang. Search for dark matter at colliders. *Nature Physics*, 13(3):217–223, mar 2017.

[17] Salvatore Rappoccio. The experimental status of direct searches for exotic physics beyond the standard model at the large hadron collider. *Reviews in Physics*, 4:100027, nov 2019.

[18] J. Flix Molina, A. Forti, M. Girone, and Andrea Sciaba. Operating the Worldwide LHC Computing Grid: current and future challenges. *J. Phys. Conf. Ser.*, 513:062044, 2014.

[19] Flavia Donno and Maarten Litmaath. Data management in WLCG and EGEE. Worldwide LHC Computing Grid. Technical report, CERN, Geneva, Feb 2008.

[20] Genessis Perez. Unitarization models for vector boson scattering at the lhc, 01 2018.

[21] Kwan Lee, S.W. Cho, S.Y. Choi, Beixue Hong, Yunsun Go, M.H. Kang, J.H. Lim, Sangho Park, Anna Cimmino, Shannon Crucy, Alexis Fagot, Mazlum Gul, Alberto Rios, Michael Tytgat, Nikolaos Zaganidis, Soher Aly, Yasser Assran, Amr Radi, A. Sayed, and Juozas Vaitkus. Radiation tests of real-sized prototype rpcs for the phase-2 upgrade of the cms muon system. *Journal of Instrumentation*, 11:C08008–C08008, 08 2016.

[22] B. Vormwald. The cms inner tracker – transition from lhc run i to run ii and first experience of run ii, 2016.

[23] Tai Sakuma and Thomas McCauley. Detector and event visualization with SketchUp at the CMS experiment. *Journal of Physics: Conference Series*, 513(2):022032, jun 2014.

[24] CMS collaboration. CMS Technical Design Report for the Pixel Detector Upgrade, 9 2012.

[25] A. Kumar. Near future upgrades of the cms pixel detector. *Journal of Instrumentation*, 10, 2015.

[26] J.-L. Agram. Cms silicon strip tracker performance. *Physics Procedia*, 37:844–850, 2012. Proceedings of the 2nd International Conference on Technology and Instrumentation in Particle Physics (TIPP 2011).

[27] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[28] J. Freeman. Innovations for the CMS HCAL. *Int. J. Mod. Phys. A*, 25:2421–2436, 2010.

[29] S Chatrchyan et al. Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays. *JINST*, 5:T03021, 2010.

[30] S. Chatrchyan et al. The CMS trigger system. *Journal of Instrumentation*, 12(01):P01020–P01020, jan 2017.

[31] CMS Collaboration. met-filters. `https://indico.cern.ch/event/512424/attachments/1246970/1855908/HL-ILO_General_Interest.pdf`. [Online; accessed 21-Nov-2022].

[32] CMS collaboration. Technical Proposal for the Phase-II Upgrade of the CMS Detector, 6 2015.

[33] L Feld, W Karpinski, K Klein, J Merz, J Sammet, and M Wlochal. A dc-dc converter based powering scheme for the upgrade of the cms pixel detector. *Journal of Instrumentation*, 6(11):C11031, nov 2011.

[34] Aerotch. Aerotech series 1500 Gantry.

[35] Branislav Ristic. Prototype Module Construction for the High Luminosity Upgrade of the CMS Pixel Detector. *PoS*, Vertex2019:058, 2020.

[36] J. Andrea, B. Fuks, and F. Maltoni. Monotops at the lhc. *Physical Review D*, 84(7), Oct 2011.

[37] T. N. Pham. Ckm matrix elements, 2011.

[38] Jean-Laurent Agram, Jeremy Andrea, Michael Buttignol, Eric Conte, and Benjamin Fuks. Monotop phenomenology at the large hadron collider. *Phys. Rev. D*, 89:014028, Jan 2014.

[39] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-k˙t jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, apr 2008.

[40] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, jan 2015.

[41] CMS Collaboration. Combined results of searches for the standard model higgs boson in pp collisions at 7 tev”. *Physics Letters B*, 710(1):26–48, mar 2012.

[42] Marc Hallin. *Neyman-Pearson Lemma*. John Wiley & Sons, Ltd, 2014.

[43] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2), feb 2011.

[44] A. Perloff. Pileup measurement and mitigation techniques in CMS. *J. Phys. Conf. Ser.*, 404:012045, 2012.

[45] Rahmat Rahmat, Rob Kroeger, and Andrea Giammanco. The fast simulation of the CMS experiment. *J. Phys. Conf. Ser.*, 396:062016, 2012.

[46] CMS Collaboration. prefire-reweighting. `https://twiki.cern.ch/twiki/bin/viewauth/CMS/L1PrefiringWeightRecipe`. [Online; accessed 21-Nov-2022].

[47] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7), jul 2014.

[48] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Computer Physics Communications*, 191:159–177, jun 2015.

[49] Jean-Laurent Agram, Jeremy Andrea, Michael Buttignol, Eric Conte, and Benjamin Fuks. Monotop phenomenology at the large hadron collider. *Physical Review D*, 89(1), jan 2014.

[50] CMS Collaboration. Electron trigger efficiencies. `https://indico.cern.ch/event/1146225/contributions/4835158/attachments/2429997/4160813/HLTSFsWprime%20.pdf`. [Online; accessed 21-Nov-2022].

[51] CMS Collaboration. Muon trigger efficiencies. `https://indico.cern.ch/event/1005603/contributions/4239813/attachments/2194259/3709453/20210222_ULTriggerSF_moh_v1.pdf`. [Online; accessed 21-Nov-2022].

[52] CMS Collaboration. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *Journal of Instrumentation*, 16(05):P05014, may 2021.

[53] CMS Collaboration. Electron cut-based identification efficiencies. `https://twiki.cern.ch/twiki/bin/view/CMS/EgammaUL2016To2018#General_note_about_ID_SFs`. [Online; accessed 20-Nov-2022].

[54] CMS Collaboration. Electron reco-id efficiencies. `https://twiki.cern.ch/twiki/bin/view/CMS/EgammaUL2016To2018#SFs_for_Electrons_UL_2018`. [Online; accessed 21-Nov-2022].

[55] CMS Collaboration. Electron reco-id efficiencies. `https://twiki.cern.ch/twiki/bin/view/CMS/EgammaUL2016To2018#SFs_for_Photons_UL_2016_postVFP`. [Online; accessed 21-Nov-2022].

[56] CMS Collaboration. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at 13 TeV. *Journal of Instrumentation*, 13(06):P06015–P06015, jun 2018.

[57] CMS Collaboration. Muon iso-id efficiencies. `https://twiki.cern.ch/twiki/bin/view/CMS/MuonUL2018`. [Online; accessed 21-Nov-2022].

[58] Florian Beaudette. The cms particle flow algorithm, 2014.

[59] CMS Collaboration. Pileup Removal Algorithms. Technical report, CERN, Geneva, 2014.

[60] CMS Collaboration. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *Journal of Instrumentation*, 12(02):P02014–P02014, feb 2017.

[61] CMS Collaboration. Jet algorithms performance in 13 TeV data. Technical report, CERN, Geneva, 2017.

[62] CMS Collaboration. Jet identification. `https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVUL`. [Online; accessed 21-Nov-2022].

[63] CMS Collaboration. Pu identification. `https://twiki.cern.ch/twiki/bin/view/CMS/PileupJetIDUL`. [Online; accessed 21-Nov-2022].

[64] CMS Collaboration. Pu identification. `https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution#JER_Scaling_factors_and_Uncertai`. [Online; accessed 21-Nov-2022].

[65] E. Bols, J. Kieseler, M. Verzetti, M. Stoye, and A. Stakia. Jet flavour classification using DeepJet. *Journal of Instrumentation*, 15(12):P12012–P12012, dec 2020.

[66] CMS Collaboration. met-xy-corrections. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMetAnalysis#xy_Shift_Correction`. [Online; accessed 21-Nov-2022].

[67] CMS Collaboration. met-filters. `https://twiki.cern.ch/twiki/bin/view/CMS/MissingETOptionalFiltersRun2#UL_data`. [Online; accessed 21-Nov-2022].

[68] Rishabh Uniyal and CMS Collaboration. Search for monotop in the leptonic channel in proton-proton collisions at $\sqrt{s} = 13$ Tev in the CMS detector. In *APS April Meeting Abstracts*, volume 2021 of *APS Meeting Abstracts*, page B19.001, January 2021.

[69] CMS Collaboration. Search for dark matter in events with energetic, hadronically decaying top quarks and missing transverse momentum at $$ \sqrt{s}=13 $$ TeV. *Journal of High Energy Physics*, 2018(6), jun 2018.

[70] Andy Buckley, James Ferrando, Stephen Lloyd, Karl Nordström, Ben Page, Martin Rüfenacht, Marek Schönherr, and Graeme Watt. LHAPDF6: parton density access in the LHC precision era. *The European Physical Journal C*, 75(3), mar 2015.

[71] Roger B. Davis and Kenneth J. Mukamal. Hypothesis testing. *Circulation*, 114(10):1078–1082, 2006.

[72] CMS Collaboration. Higgs combine tool. `https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/`. [Online; accessed 21-Nov-2022].

[73] Rene Brun, Fons Rademakers, Philippe Canal, Axel Naumann, Olivier Couet, Lorenzo Moneta, Vassil Vassilev, Sergey Linev, Danilo Piparo, Gerardo GANIS, Bertrand Bellenot, Enrico Guiraud, Guilherme Amadio, wverkerke, Pere Mato, TimurP, Matevž Tadel, wlav, Enric Tejedor, Jakob Blomer, Andrei Gheata, Stephan Hageboeck, Stefan Roiser, marsupial, Stefan Wunsch, Oksana Shadura, Anirudha Bose, CristinaCristescu, Xavier Valls, and Raphael Isemann. root-project/root: v6.18/02, August 2019.

[74] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2), feb 2011.

[75] Muhammad Aslam. Introducing kolmogorov–smirnov tests under uncertainty: An application to radioactive data. *ACS Omega*, 5(1):914–917, 2020. PMID: 31956845.

[76] Andreas Albert, Mihailo Backovic, Antonio Boveia, Oliver Buchmueller, Giorgio Busoni, Albert De Roeck, Caterina Doglioni, Tristan DuPree, Malcolm Fairbairn, Marie-Helene Genest, Stefania Gori, Giuliano Gustavino, Kristian Hahn, Ulrich Haisch, Philip C. Harris, Dan Hayden, Valerio Ippolito, Isabelle John, Felix Kahlhoefer, Suchita Kulkarni, Greg Landsberg, Steven Lowette, Kentarou Mawatari, Antonio Riotto, William Shepherd, Tim M. P. Tait, Emma Tolley, Patrick Tunney, Bryan Zaldivar, and Markus Zinser. Recommendations of the lhc dark matter working group: Comparing lhc searches for heavy mediators of dark matter production in visible and invisible decay channels, 2017.

[77] Idir Boucheneb, Giacomo Cacciapaglia, Aldo Deandrea, and Benjamin Fuks. Revisiting monotop production at the LHC. *Journal of High Energy Physics*, 2015(1), jan 2015.

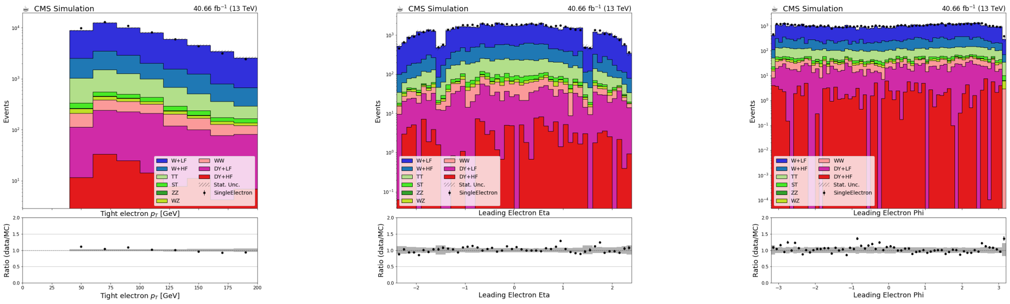
[78] Lindsey Gray, Nicholas Smith, Benjamin Tovar, Andrzej Novak, Jayjeet Chakraborty, Peter Fackeldey, Nikolai Hartmann, Gordon Watts, Douglas Thain, Giordon Stark, BenGalewsky, Jonas Rübenach, Benjamin Fischer, Devin Taylor, MoAly98, Dmitry Kondratyev, Paul Gessinger, Yi-Mu ”Enoch” Chen, Joosep Pata, Anna Woodard, Andreas Albert, slehti,

Zoe Surma, Alexx Perloff, Kevin Pedro, dnoonan08, Andrew Hennessee, Karol Krizka, kmohrman, and Lukas. Coffeateam/coffea: Release v0.7.20, November 2022.

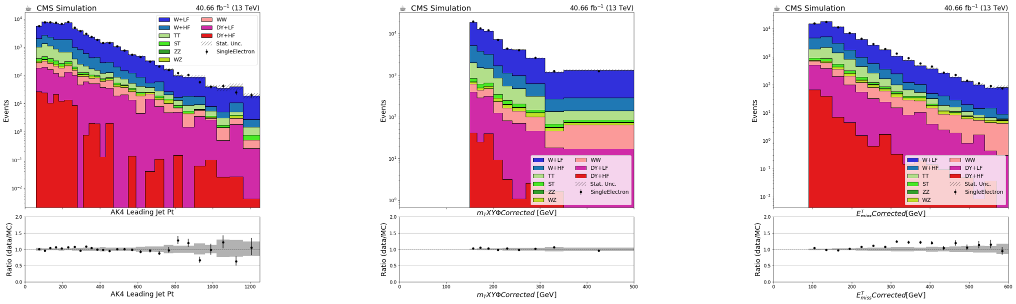
APPENDIX A

Control Region plots

The collision data collection for Run II at the CMS experiment was performed over a period of three years, from 2016-2018. In this appendix we will present the rest of the control region plots corresponding to years 2016-2017. Kindly, refer to the caption of the figures for more details

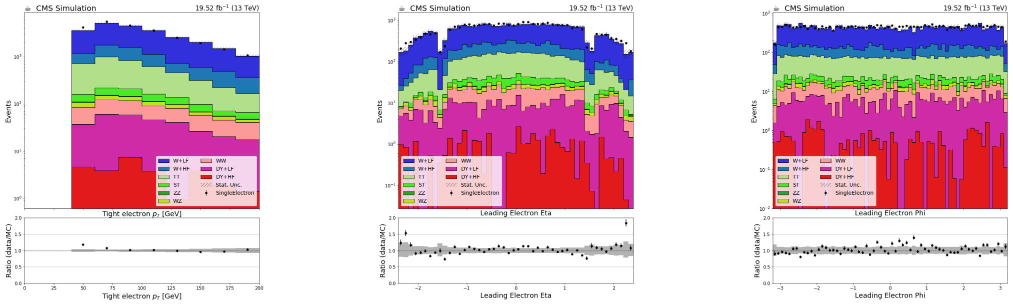


(a) Leading electron p_T (b) Leading electron η (c) Leading electron ϕ

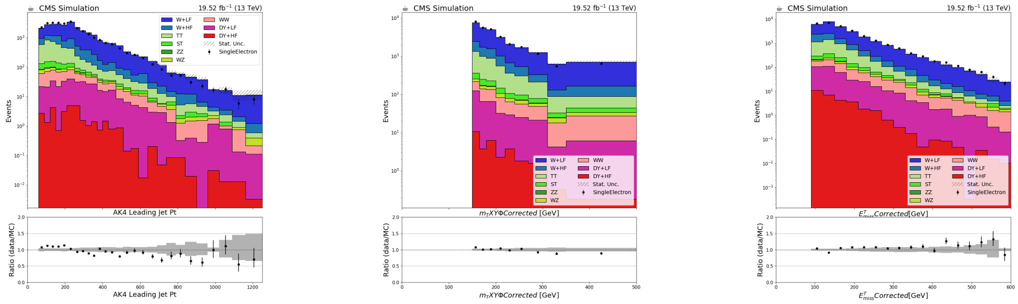


(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

Figure A.1: This figure a.) shows the p_T distribution of the leading electron. b.) shows the η distribution of the leading electron. c.) shows the ϕ distribution of the leading electron. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the W Control region with electron in the final state for 2017 era.

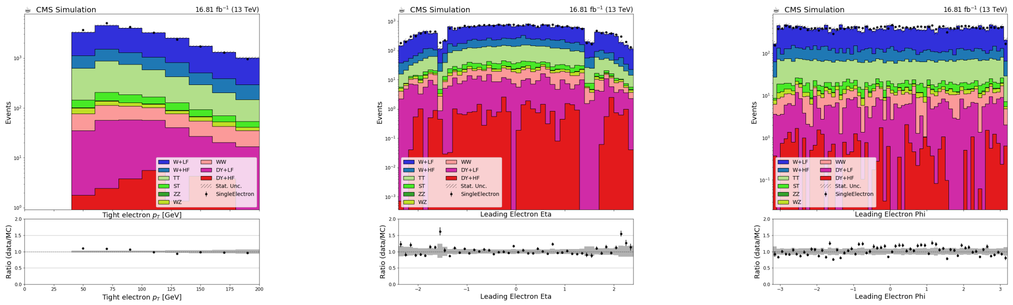


(a) Leading electron p_T (b) Leading electron η (c) Leading electron ϕ

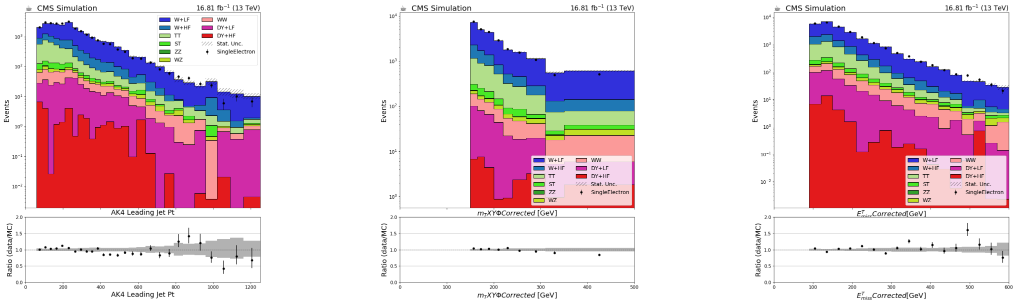


(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

Figure A.2: This figure a.) shows the p_T distribution of the leading electron. b.) shows the η distribution of the leading electron. c.) shows the ϕ distribution of the leading electron. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the W Control region with electron in the final state for 2016preVFP era.

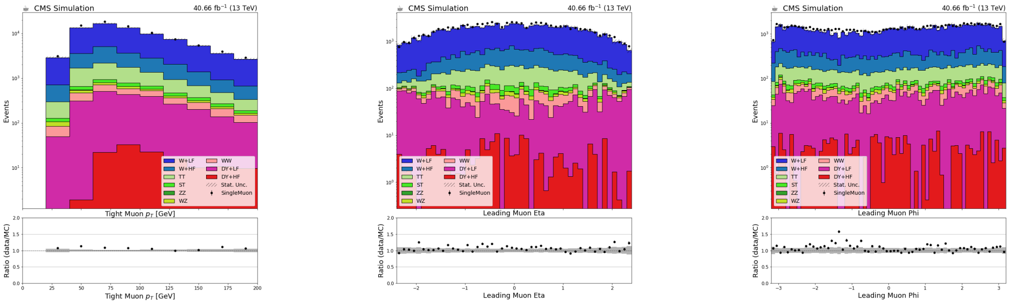


(a) Leading electron p_T (b) Leading electron η (c) Leading electron ϕ

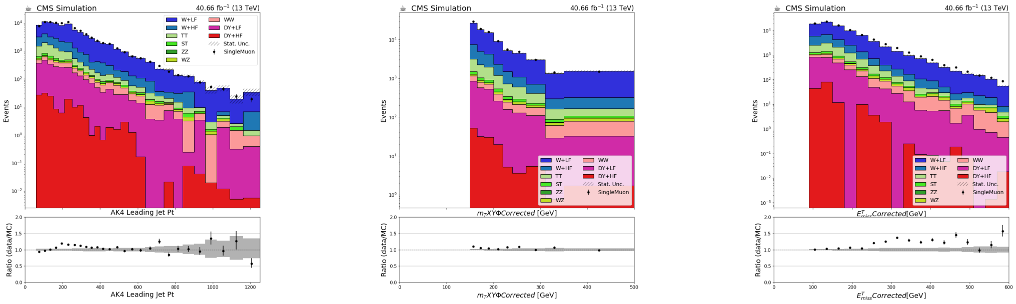


(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

Figure A.3: This figure a.) shows the p_T distribution of the leading electron. b.) shows the η distribution of the leading electron. c.) shows the ϕ distribution of the leading electron. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the W Control region with electron in the final state for 2016postVFP era.

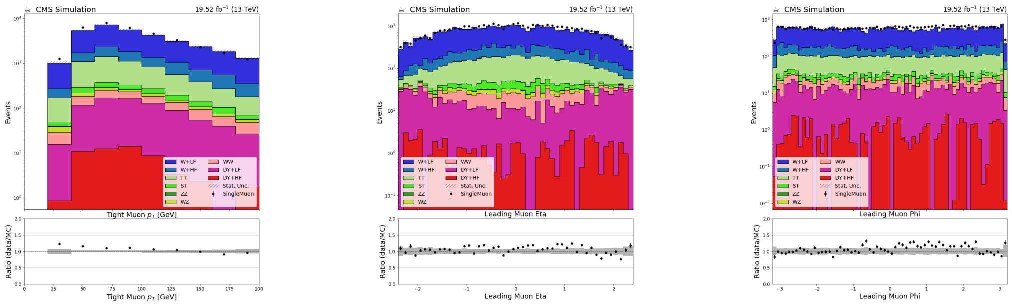


(a) Leading muon p_T (b) Leading muon η (c) Leading muon ϕ

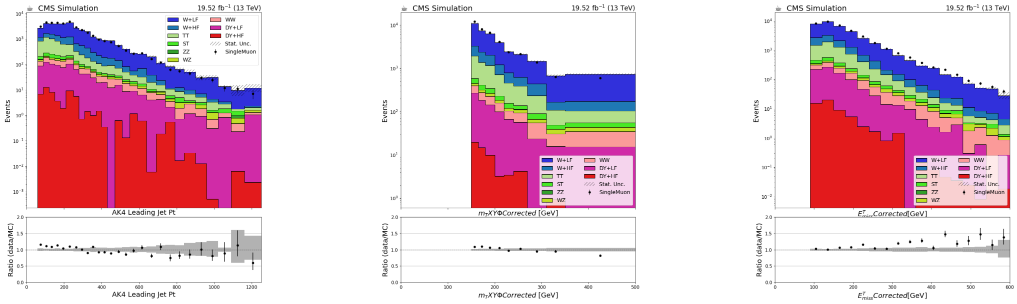


(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

Figure A.4: This figure a.) shows the p_T distribution of the leading muon. b.) shows the η distribution of the leading muon. c.) shows the ϕ distribution of the leading muon. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the W Control region with electron in the final state for 2017 era.

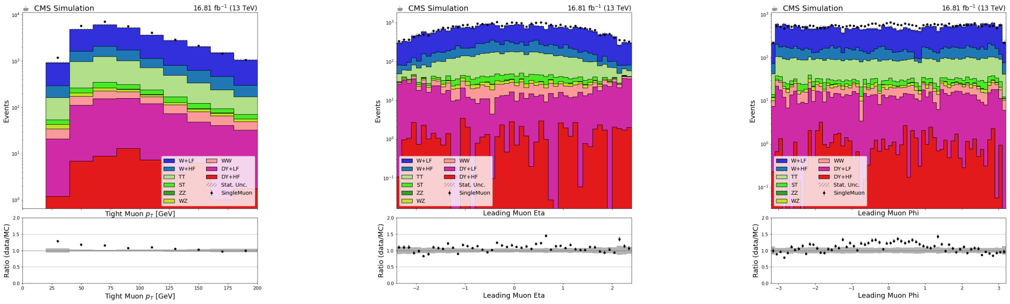


(a) Leading muon p_T (b) Leading muon η (c) Leading muon ϕ

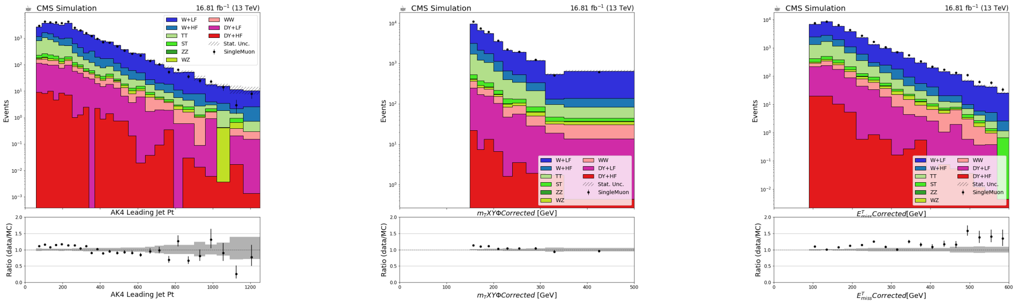


(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

Figure A.5: This figure a.) shows the p_T distribution of the leading muon. b.) shows the η distribution of the leading muon. c.) shows the ϕ distribution of the leading muon. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the W Control region with electron in the final state for 2016preVFP era.

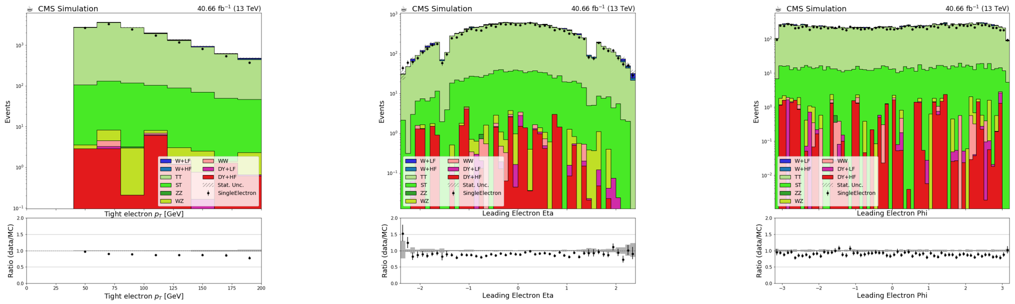


(a) Leading muon p_T (b) Leading muon η (c) Leading muon ϕ

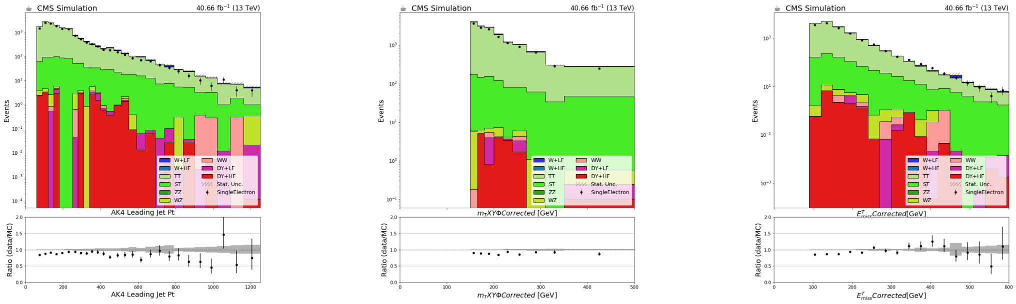


(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

Figure A.6: This figure a.) shows the p_T distribution of the leading muon. b.) shows the η distribution of the leading muon. c.) shows the ϕ distribution of the leading muon. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the W Control region with electron in the final state for 2016postVFP era.

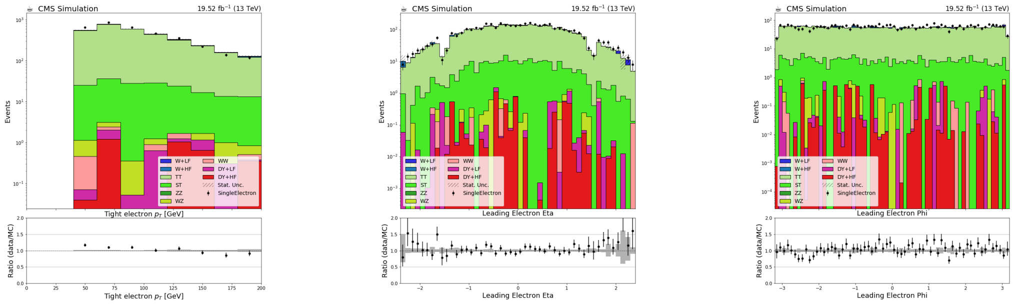


(a) Leading electron p_T (b) Leading electron η (c) Leading electron ϕ



(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

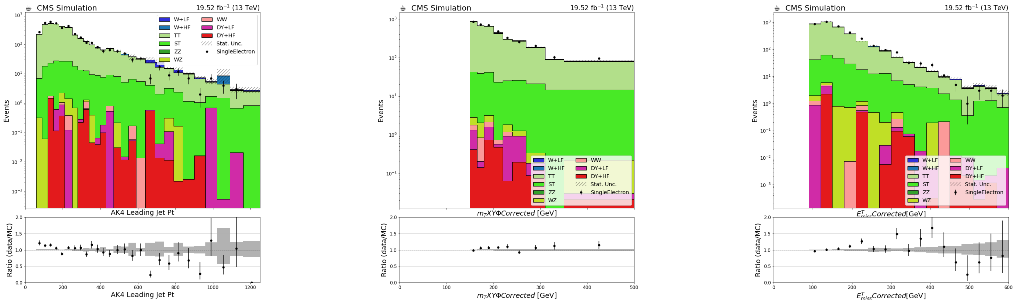
Figure A.7: This figure a.) shows the p_T distribution of the leading electron. b.) shows the η distribution of the leading electron. c.) shows the ϕ distribution of the leading electron. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the top Control region with electron in the final state for 2017 era.



(a) Leading electron p_T

(b) Leading electron η

(c) Leading electron ϕ

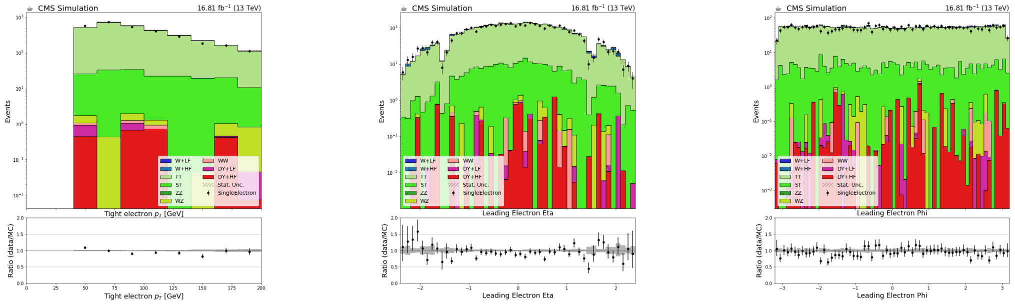


(d) Leading AK4 jet p_T

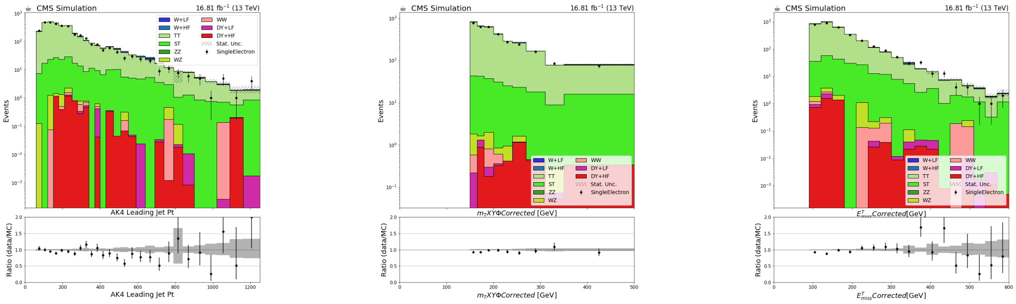
(e) m_T^W

(f) $\not{E}_T$

Figure A.8: This figure a.) shows the p_T distribution of the leading electron. b.) shows the η distribution of the leading electron. c.) shows the ϕ distribution of the leading electron. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the W Control region with electron in the final state for 2016preVFP era.

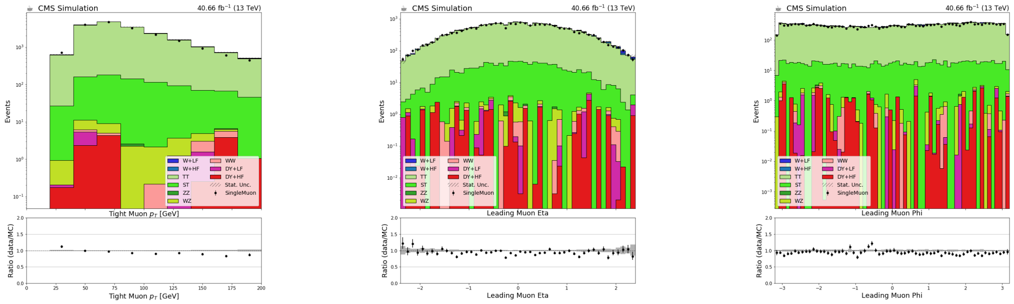


(a) Leading electron p_T (b) Leading electron η (c) Leading electron ϕ

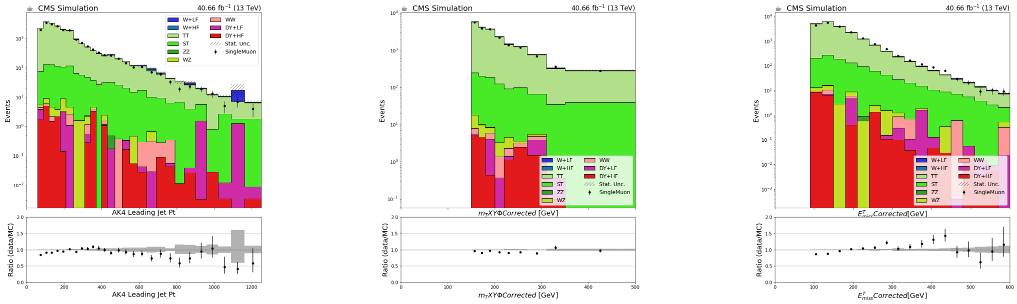


(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

Figure A.9: This figure a.) shows the p_T distribution of the leading electron. b.) shows the η distribution of the leading electron. c.) shows the ϕ distribution of the leading electron. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the top Control region with electron in the final state for 2016postVFP era.

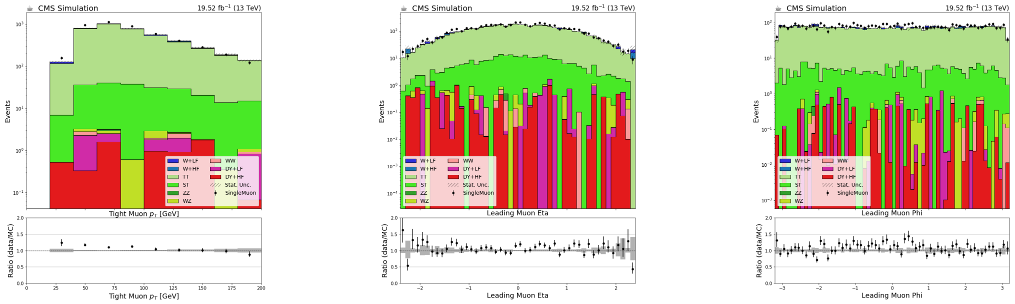


(a) Leading muon p_T (b) Leading muon η (c) Leading muon ϕ

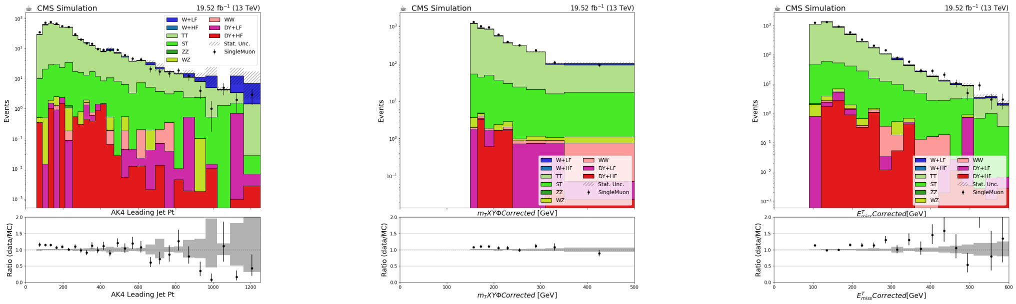


(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

Figure A.10: This figure a.) shows the p_T distribution of the leading muon. b.) shows the η distribution of the leading muon. c.) shows the ϕ distribution of the leading muon. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the top Control region with electron in the final state for 2017 era.

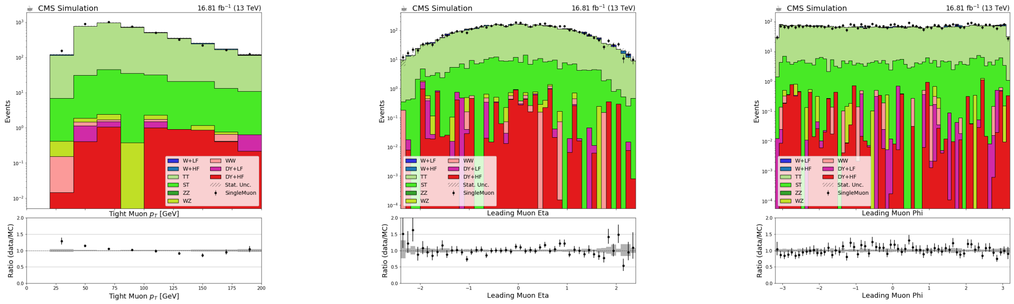


(a) Leading muon p_T (b) Leading muon η (c) Leading muon ϕ

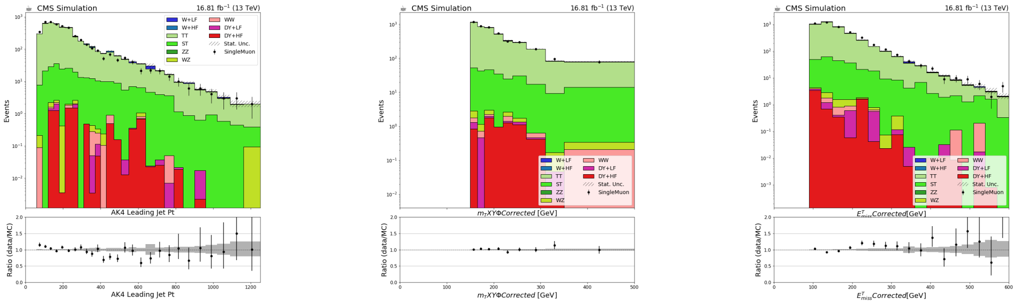


(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

Figure A.11: This figure a.) shows the p_T distribution of the leading muon. b.) shows the η distribution of the leading muon. c.) shows the ϕ distribution of the leading muon. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the top Control region with electron in the final state for 2016preVFP era.

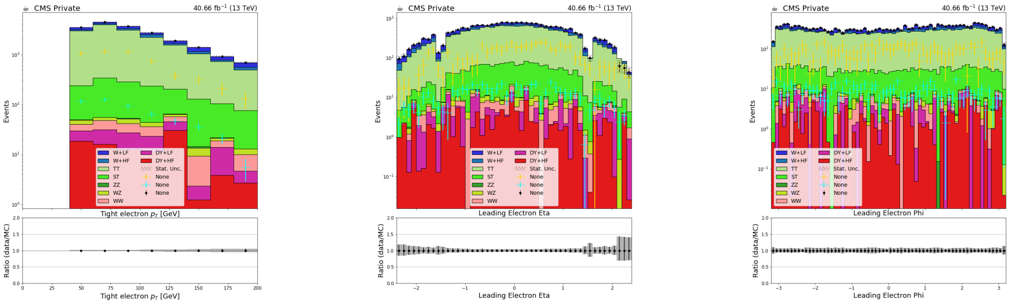


(a) Leading muon p_T (b) Leading muon η (c) Leading muon ϕ

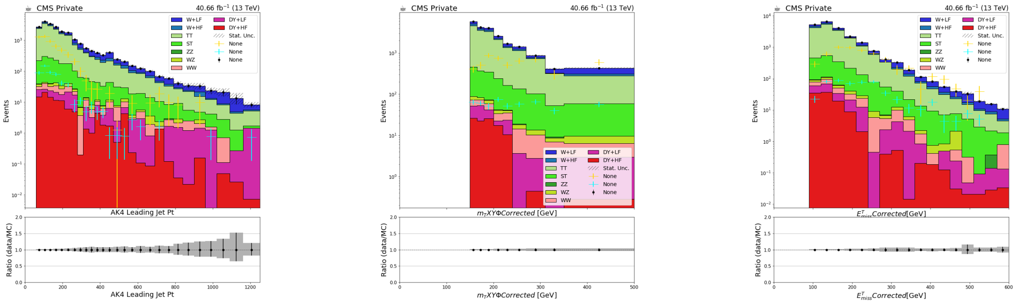


(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

Figure A.12: This figure a.) shows the p_T distribution of the leading muon. b.) shows the η distribution of the leading muon. c.) shows the ϕ distribution of the leading muon. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the top Control region with electron in the final state for 2016postVFP era.

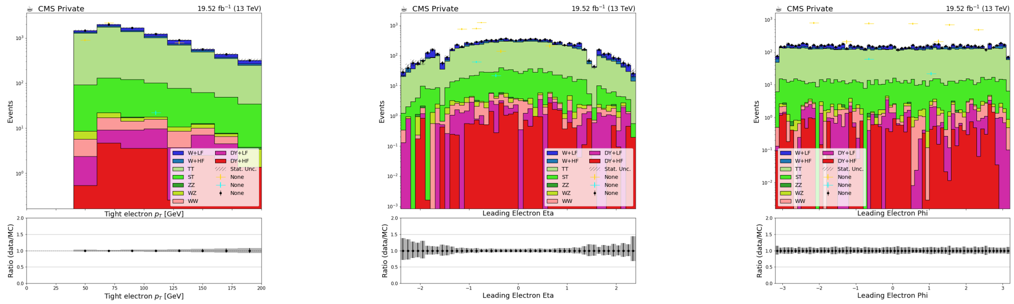


(a) Leading electron p_T (b) Leading electron η (c) Leading electron ϕ

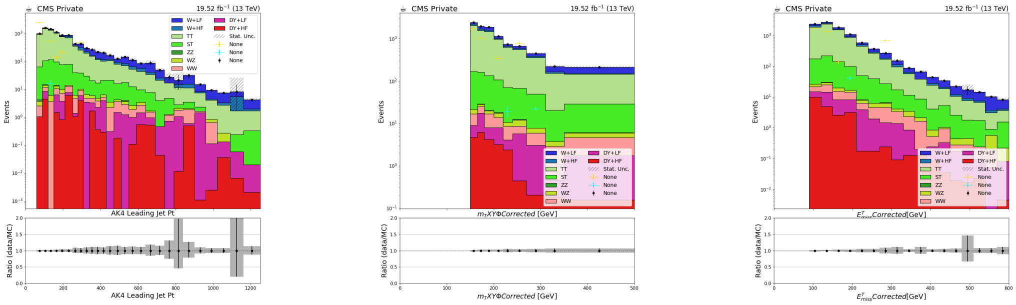


(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

Figure A.13: This figure a.) shows the p_T distribution of the leading electron. b.) shows the η distribution of the leading electron. c.) shows the ϕ distribution of the leading electron. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the Signal region with electron in the final state for 2017 era.

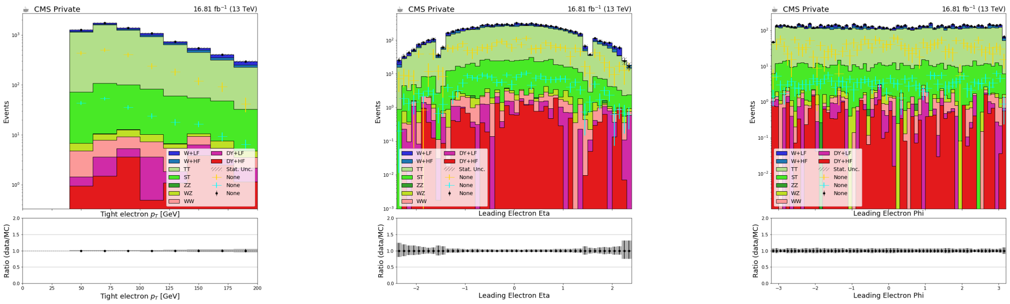


(a) Leading electron p_T (b) Leading electron η (c) Leading electron ϕ

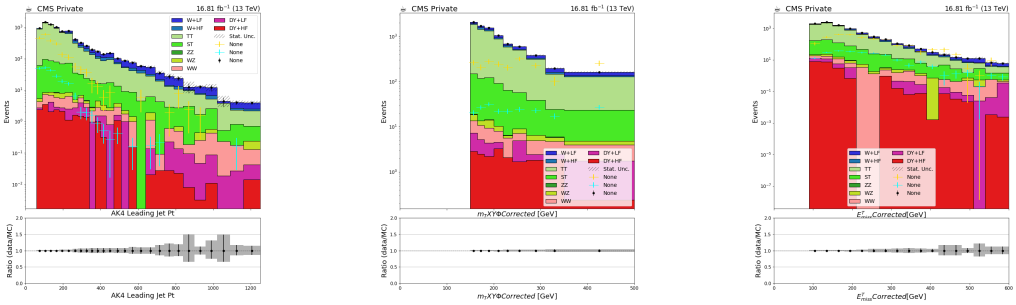


(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

Figure A.14: This figure a.) shows the p_T distribution of the leading electron. b.) shows the η distribution of the leading electron. c.) shows the ϕ distribution of the leading electron. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the Signal region with electron in the final state for 2016preVFP era.

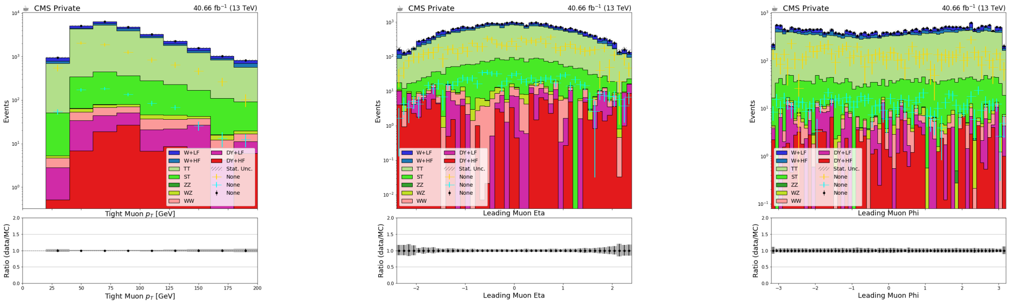


(a) Leading electron p_T (b) Leading electron η (c) Leading electron ϕ

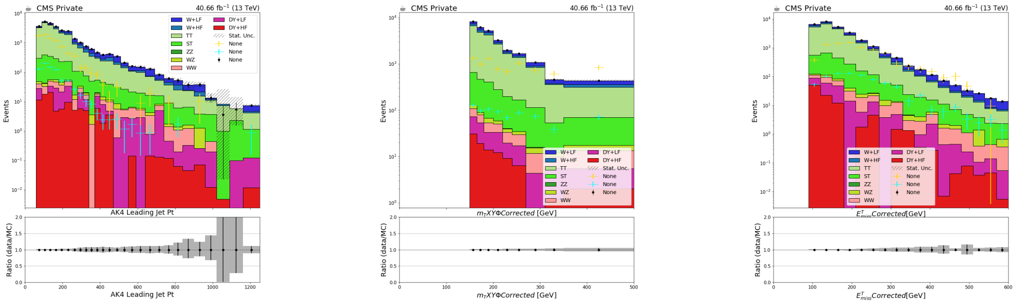


(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

Figure A.15: This figure a.) shows the p_T distribution of the leading electron. b.) shows the η distribution of the leading electron. c.) shows the ϕ distribution of the leading electron. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the Signal region with electron in the final state for 2016postVFP era. Gray bands indicate statistical uncertainties on simulated background samples. The black dots are the pseudo-data for representation (and not real data). The gold and the cyan colored crosses are the signal samples M_χ=100, M_V=195 and M_χ=50, M_V=200 respectively

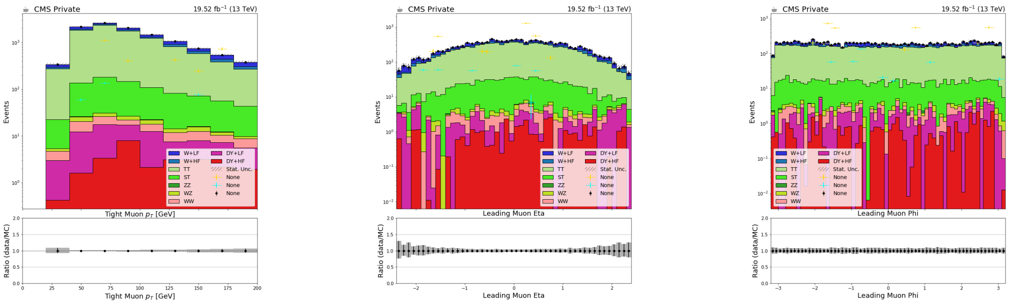


(a) Leading muon p_T (b) Leading muon η (c) Leading muon ϕ



(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

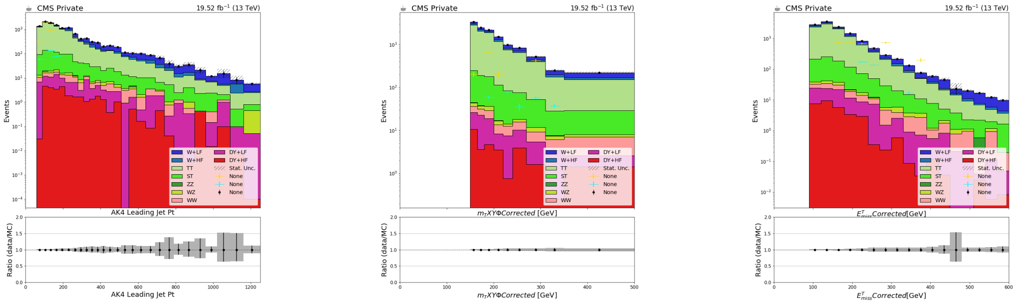
Figure A.16: This figure a.) shows the p_T distribution of the leading muon. b.) shows the η distribution of the leading muon. c.) shows the ϕ distribution of the leading muon. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the Signal region with muon in the final state for 2017 era. Gray bands indicate statistical uncertainties on simulated background samples. The black dots are the pseudo-data for representation (and not real data). The gold and the cyan colored crosses are the signal samples M_χ=100, M_V=195 and M_χ=50, M_V=200 respectively



(a) Leading muon p_T

(b) Leading muon η

(c) Leading muon ϕ

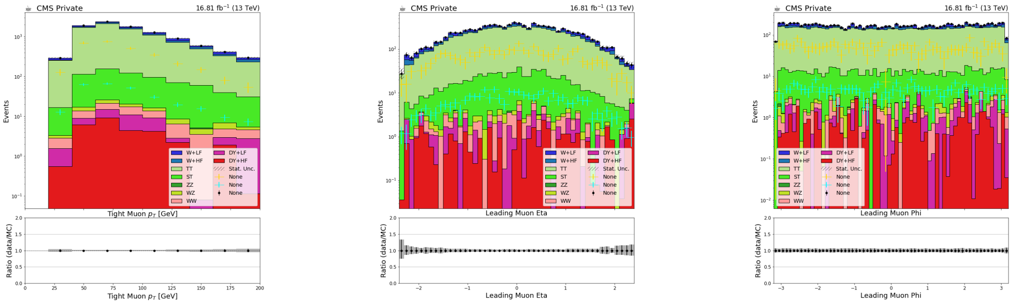


(d) Leading AK4 jet p_T

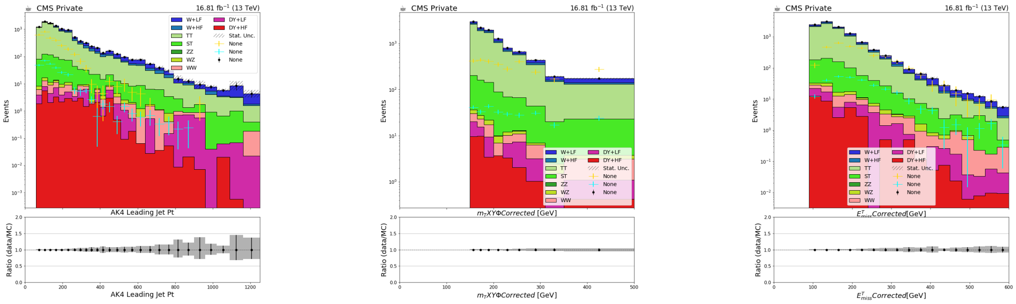
(e) m_T^W

(f) $\not{E}_T$

Figure A.17: This figure a.) shows the p_T distribution of the leading muon. b.) shows the η distribution of the leading muon. c.) shows the ϕ distribution of the leading muon. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the Signal region with muon in the final state for 2016preVFP era. Gray bands indicate statistical uncertainties on simulated background samples. The black dots are the pseudo-data for representation (and not real data). The gold and the cyan colored crosses are the signal samples M_χ=100, M_V=195 and M_χ=50, M_V=200 respectively



(a) Leading muon p_T (b) Leading muon η (c) Leading muon ϕ



(d) Leading AK4 jet p_T (e) m_T^W (f) $\not{E}_T$

Figure A.18: This figure a.) shows the p_T distribution of the leading muon. b.) shows the η distribution of the leading muon. c.) shows the ϕ distribution of the leading muon. d.) shows the p_T distribution of the leading AK4 jet. e.) shows the m_T^W, transverse mass of the W boson. f.) shows the missing p_T distribution in the Signal region with muon in the final state for 2016postVFP era. Gray bands indicate statistical uncertainties on simulated background samples. The black dots are the pseudo-data for representation (and not real data). The gold and the cyan colored crosses are the signal samples M_χ=100, M_V=195 and M_χ=50, M_V=200 respectively

APPENDIX B

Pulls and Impacts

The collision data collection for Run II at the CMS experiment was performed over a period of three years, from 2016-2018. In this appendix we will present the rest of the impact and pull plots corresponding to years 2016-2017. Kindly, refer to the caption of the figures for more details

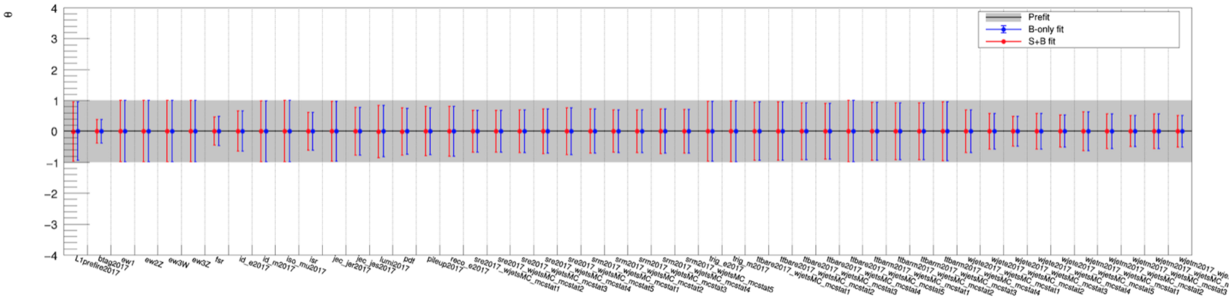


Figure A.1: Pulls for 2018 era using Asimov toy

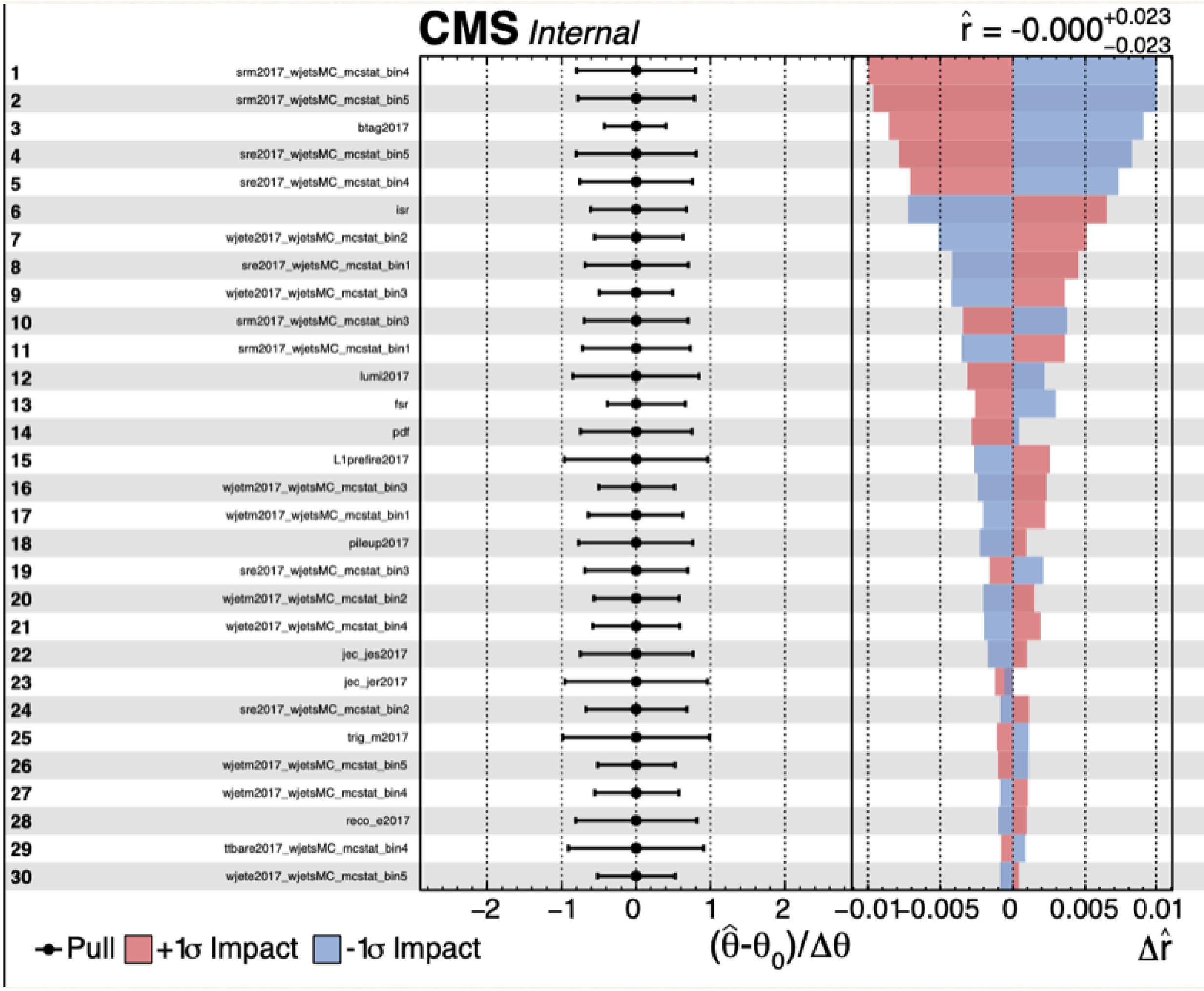


Figure A.2: impacts for 2018 era using Asimov toy

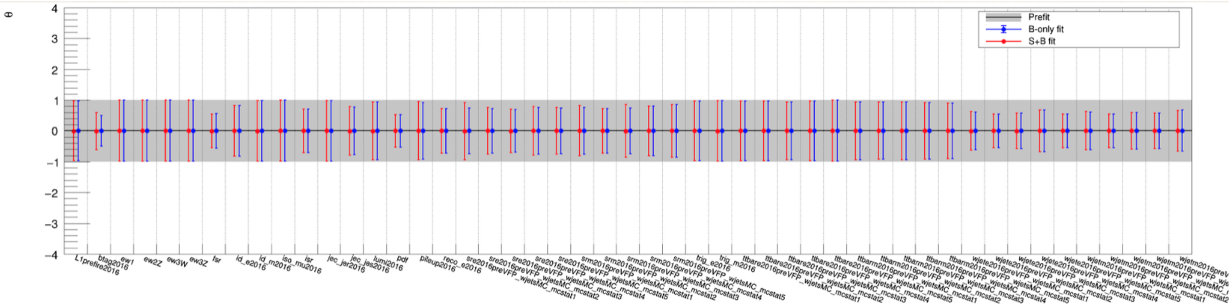


Figure A.3: Pulls for 2016preVFP era using Asimov toy

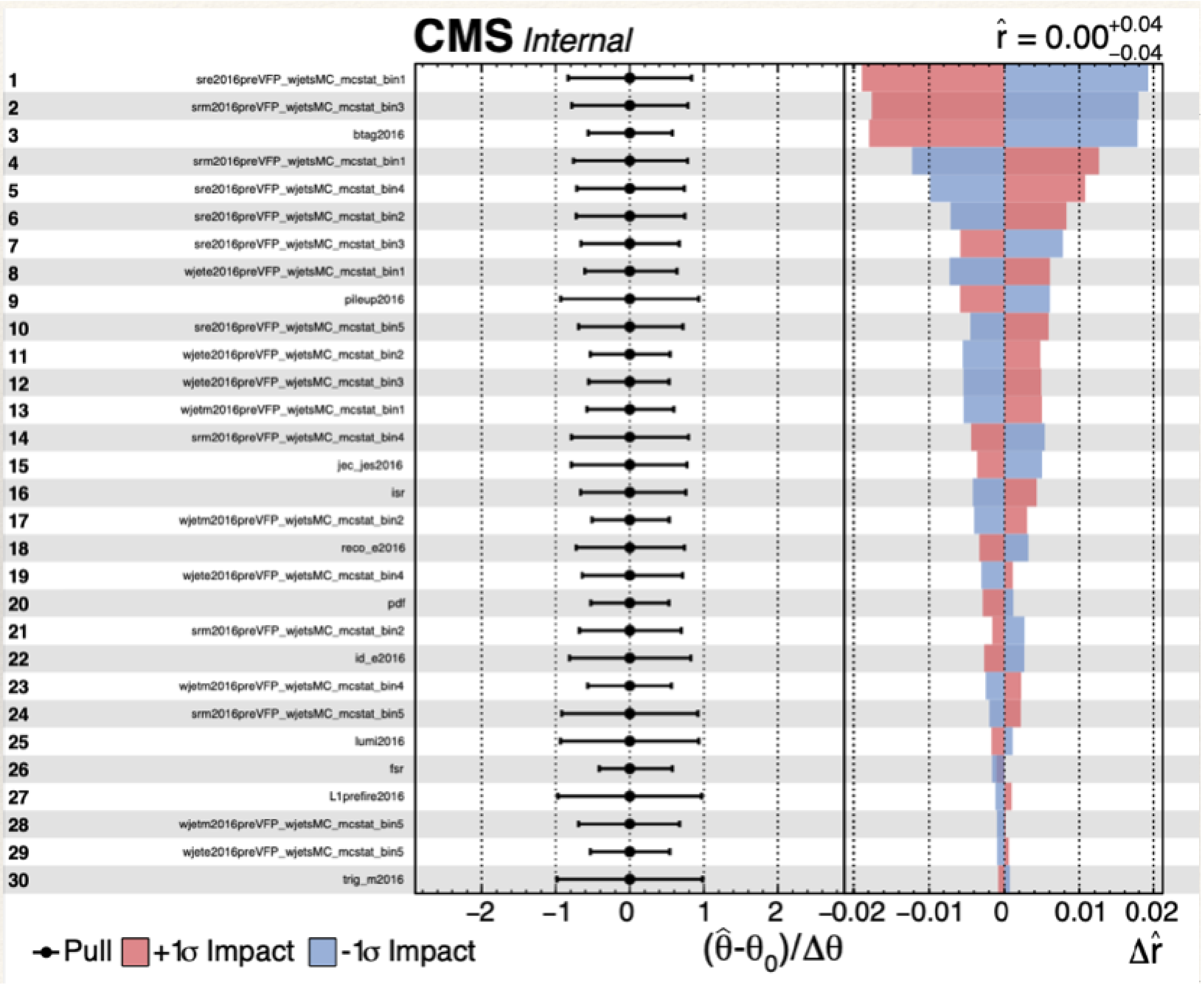


Figure A.4: impacts for 2018 era using Asimov toy

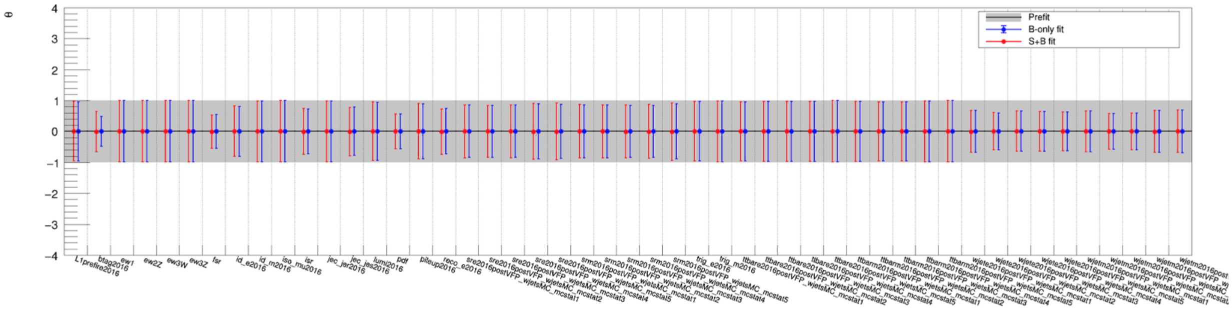


Figure A.5: Pulls for 2016postVFP era using Asimov toy

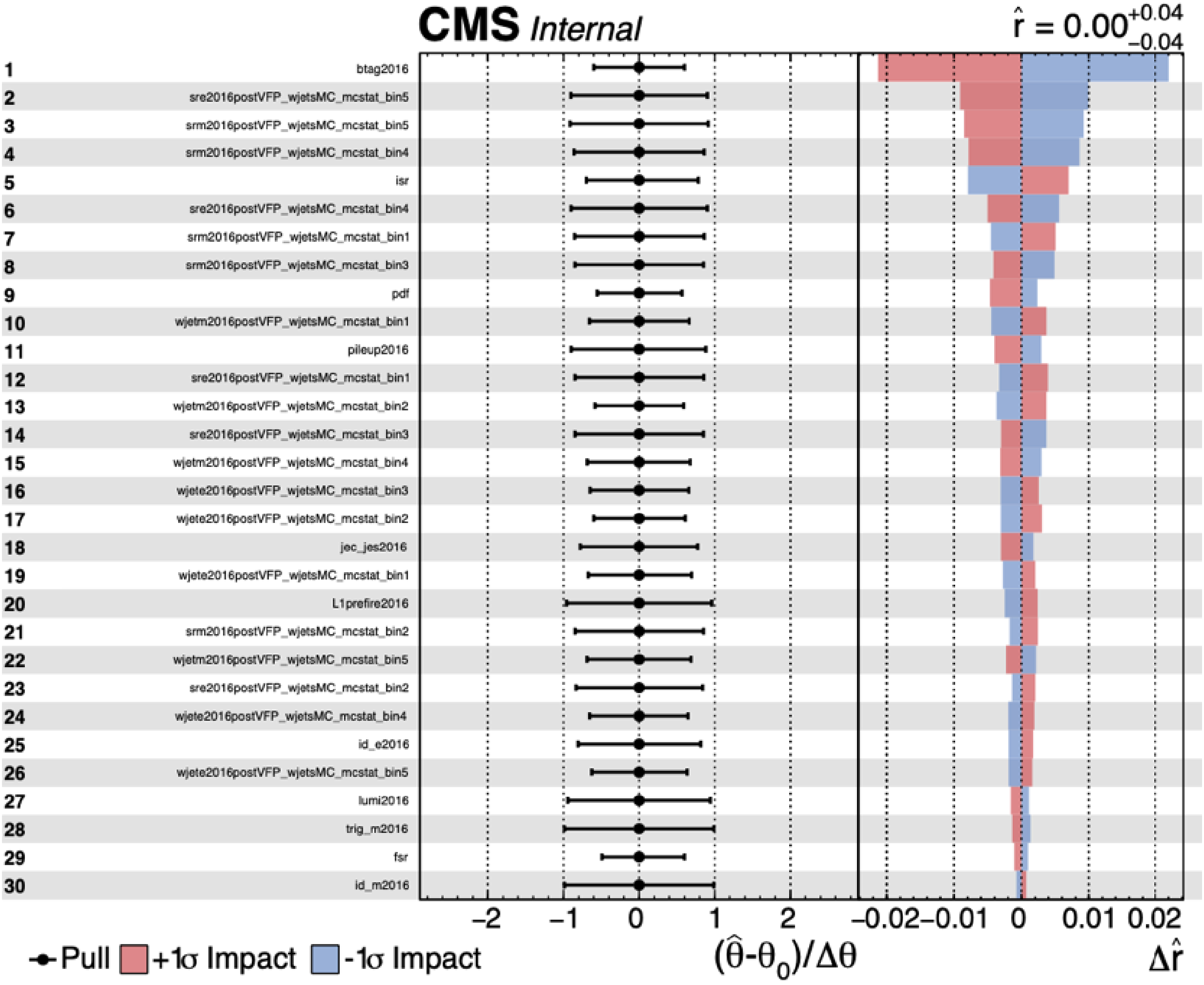


Figure A.6: impacts for 2018 era using Asimov toy