A SEARCH FOR HEAVY NEUTRINOS VIA ELECTROWEAK VECTOR BOSON FUSION

PROCESSES IN PROTON-PROTON COLLISIONS AT THE LARGE HADRON COLLIDER

By

Brandon Joseph Soubasis

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Physics

May 12, 2023

Nashville, Tennessee

Approved:

Will Johns, Ph.D.

Alfredo Gurrola, Ph.D.

Paul Sheldon, Ph.D.

Kelly Holley-Bockelmann, Ph.D.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

The Standard Model of Particle Physics

1.1   Introduction

In high energy physics, the best experimentally consistent knowledge of how particles and matter interact is encompassed in the Standard Model (SM). It successfully describes an extraordinarily large amount of data from various experiments. However, there are a growing number of observations that suggest the SM is incomplete. The underpinning of the SM is quantum field theory, a powerful theory that describes the fundamental principles of quantum mechanics and special relativity. The SM successfully incorporates and describes the electromagnetic interactions - which accounts for atomic, molecular, optical, and condensed matter physics - and the strong (also known as quantum chromodynamics or QCD) and weak interactions responsible for nuclear processes [13]. The gravitational force is the only known fundamental interaction that has not been successfully incorporated into the SM. However, the interaction strength of the gravitational force is small compared to the other three fundamental forces and plays no role in the physics topic described in this thesis.

According to the SM, matter is composed of 12 fermions (spin 1/2 particles), while the fundamental forces of nature between fundamental particles and matter (electromagnetic, weak, and strong) are mediated by bosons (integer spin particles). SM fermions are divided into two subgroups: leptons and quarks. Leptons are involved primarily in electroweak interactions, and thus are characterized by an electric charge which couples to the $\gamma/W^{\pm}/Z^0$ bosons that mediate the interaction. Leptons are divided into first, second, and third generation fermions. The first generation leptons consist of the electron and corresponding electron neutrino, e and ν_e. The second generation leptons consist of the muon, μ, and the corresponding muon neutrino, ν_μ. Similarly, the tau and the tau neutrino, τ and ν_τ , are the third generation leptons. The most significant difference between generations of leptons are the masses of the particles. For example, the mass of the electron is 0.511 MeV/c^2 while the mass of the tau is 1777 MeV/c^2, a difference of almost a factor of 3500. Quarks possess electric charge and interact via electroweak interactions; however, they are also characterized by color charge which couples to the gluons that mediate interactions involving the strong force. Like the leptons, quarks are divided into three generations or groups. The up and down quarks, u and d, are the first generation quarks and combine to form the protons and neutrons within atoms. The charm and strange quarks, c and s, are the second generation quarks, while the top and bottom quarks, t and b, are of the third generation. The particles of the SM, and their properties, as can be seen in Table 1.1. A summary of the four

fundamental forces can be seen in Table 1.2.

The gauge bosons mediate the interactions between elementary particles and matter through the gauge groups of the SM, $SU(3)_C \times SU(2)_L \times U(1)_Y$. The gauge group $SU(3)_C$ governs the underlying symmetries and physics behind the strong force felt by color-charged particles such as quarks. The gauge groups $SU(2)_L$ and $U(1)_Y$ together determine the symmetries and physics of the electroweak force, which is mediated by the $W^\pm/Z^0$ bosons. The $W^\pm$ bosons mediate the weak interactions between particles of different flavors and act only on left-handed particles and right-handed anti-particles. The Z^0 boson interacts with both left-handed particles and right-handed anti-particles. The $SU(2)L$ gauge group governs the physics of left-handed helicity states of fermion pairs such as (u_L, d_L) and $(e_L, \nu e, L)$, while the $U(1)_Y$ gauge group governs the force felt by particles with hypercharge. The Higgs boson, which is a scalar particle that acts as the carrier of the Higgs field, is crucial in explaining the origin of mass in the electroweak force.

Matter is merely a composition of the fundamental particles bound together in some physical state only because bosons are able to exchange information about the way they should interact. For example, matter containing 2 or 3 quarks held together by the strong force (gluons) gives rise to a new set of particles called Hadrons. Protons are just one example of a Hadron composed of 3 quarks (Baryons). Because the strong interaction does not allow quarks to exist in a "free" state (asymptotic freedom), quarks always exist in bound states. In the same way opposite electric charges can attract to form bounded states, the color charge of quarks is stabilized by binding two or three quarks together resulting in objects/particles known as Mesons (quark and anti-quark combinations) and Baryons (three quarks). Mesons and Baryons are collectively known as Hadrons.

1.2 Beyond the Standard Model

The SM of particle physics successfully explores a plethora of experimental observations. However, as experimental observations probe new questions and increasing energies, there is a widespread belief that the SM is an incomplete theory. For example, the SM inability to explain the existence of dark matter (DM) particles. DM is been shown to be a weakly interacting particle which has been observed by astronomers indirectly through observation in stellar rotation curves and gravitational lensing. Another example, in addition to not providing a particle candidate for DM, the SM lacks an explanation for the mass of light neutrinos. More specifically, the SM predicts massless ν_ℓ states, but a non-zero mass of the three generations of neutrinos is implied by the observation of neutrino oscillations. Furthermore, the SM has no explanation for other pressing questions such as neutrino mass hierarchy and matter anti-matter asymmetry in the universe.

Field	$SU(3)_C$	$SU(2)_L$	T^3	$\frac{Y}{2}$	$Q = T^3 + \frac{Y}{2}$	Mass
$Q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$	**3**	**2**	$\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$\frac{1}{6}$	$\begin{pmatrix} \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}$	$\begin{pmatrix} 2.16 MeV \\ 4.67 MeV \end{pmatrix}$
$Q_L = \begin{pmatrix} c_L \\ s_L \end{pmatrix}$	**3**	**2**	$\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$\frac{1}{6}$	$\begin{pmatrix} \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}$	$\begin{pmatrix} 1.27 GeV \\ 93 MeV \end{pmatrix}$
$Q_L = \begin{pmatrix} t_L \\ b_L \end{pmatrix}$	**3**	**2**	$\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$\frac{1}{6}$	$\begin{pmatrix} \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}$	$\begin{pmatrix} 172.76 GeV \\ 4.18 GeV \end{pmatrix}$
u_R	**3**	**1**	0	$\frac{2}{3}$	$\frac{2}{3}$	$2.16 MeV$
d_R	**3**	**1**	0	$-\frac{1}{3}$	$-\frac{1}{3}$	$4.67 MeV$
c_R	**3**	**1**	0	$\frac{2}{3}$	$\frac{2}{3}$	$1.27 GeV$
s_R	**3**	**1**	0	$-\frac{1}{3}$	$-\frac{1}{3}$	$93 MeV$
t_R	**3**	**1**	0	$\frac{2}{3}$	$\frac{2}{3}$	$172.76 GeV$
b_R	**3**	**1**	0	$-\frac{1}{3}$	$-\frac{1}{3}$	$4.18 GeV$
$L_L = \begin{pmatrix} \nu_{e,L} \\ e_L \end{pmatrix}$	**1**	**2**	$\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$-\frac{1}{2}$	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$	$\begin{pmatrix} < 0.8 eV \\ 0.511 MeV \end{pmatrix}$
$L_L = \begin{pmatrix} \nu_{\mu,L} \\ \mu_L \end{pmatrix}$	**1**	**2**	$\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$-\frac{1}{2}$	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$	$\begin{pmatrix} < 0.8 eV \\ 105.66 MeV \end{pmatrix}$
$L_L = \begin{pmatrix} \nu_{\tau,L} \\ \tau_L \end{pmatrix}$	**1**	**2**	$\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$-\frac{1}{2}$	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$	$\begin{pmatrix} < 0.8 eV \\ 1.78 GeV \end{pmatrix}$
e_R	**1**	**1**	0	-1	-1	$0.511 MeV$
μ_R	**1**	**1**	0	-1	-1	$105.66 MeV$
τ_R	**1**	**1**	0	-1	-1	$1.78 GeV$

Table 1.1: Standard Model particle content table including the quantum numbers and masses of fermions, based on the $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetry. The table lists left-handed quark and lepton doublets, as well as right-handed quark and lepton singlets. The quantum numbers of the particles are $SU(3)_C$ color charge, weak isospin T^3, weak hypercharge Y, and electric charge $Q = T^3 + Y/2$. Masses are given in appropriate units. Table adapted from [1].

.

Force	mediator	Acts on	Relative Strength	Range
Strong	gluon	quarks	1	10^{-15} m
Electromagnetism	photon	electrical charges	10^{-2}	∞ $(1/r^2)$
Weak	$W^\pm, Z$	quarks and leptons	10^{-5}	10^{-18} m
Gravity	graviton (hypothetical)	mass and energy	10^{-39}	∞ $(1/r^2)$

Table 1.2: Summary of the four fundamental forces. Table adapted from [2].

Physics Beyond the Standard Model & Heavy Neutrinos

2.1 Introduction

The story of the neutrino is a captivating one in the realm of elementary particles. Proposed in 1930 by Walter Pauli to resolve issues of energy, momentum, and spin conservation in radioactive β-decay, the neutrino was finally confirmed experimentally in the 1950s [14]. Despite its initial discovery, the neutrino continues to elude a full understanding due to its electrical neutrality, incredibly small mass, and weak interaction with matter, making it difficult to study. To this day, the exact masses of neutrinos remain unknown. However, various methods to determine these masses exist, including neutrino oscillation experiments (from sources such as solar, atmospheric, reactor, accelerator, and astrophysical neutrinos), tritium β-decay experiments, neutrinoless double β-decay experiments, and cosmological observations [15]. The simplicity of the Standard Big Bang model also allows for the calculation of relic densities of stable particles, such as neutrinos, if they were once in thermal equilibrium with the thermal radiation bath. [16].

2.2 Neutrino oscillations and mixing

The idea that neutrinos would oscillate between ν and $\bar{\nu}$ states was first proposed by Bruno Pontecorvo in the 1950's [17]. The theory that an electron neutrino could oscillate to a muon neutrino was first considered by Maki, Nakagawa, and Sakata [15]. These two theories worked together to produce the neutrino mass eigenstates $|\nu_j\rangle$, which are described as linear superpositions of the $|\nu_\alpha\rangle$ flavor eigenstates.

$$|\nu_j\rangle = \sum_\alpha U_{\alpha j} |\nu_\alpha\rangle \tag{2.1}$$

where $\alpha = e, \mu, \tau$ and $U_{\alpha j}$ is a 3×3 unitary matrix called the PMNS (Pontecorvo, Maki, Nakagawa, Sakata) matrix. The matrix can be inverted to write the flavour eigenstates as a superposititon of the mass eigenstates:

$$|\nu_\alpha\rangle = \sum_{j=1}^{3} U_{\alpha j}^* |\nu_j\rangle \tag{2.2}$$

Since the mass eigenstates have definite mass (m_j), and energy (E_j), they can be evolved in time as established by the Schrodinger Equation

$$\mathscr{H} |\nu_j(t)\rangle = i\hbar \frac{\partial}{\partial t} |\nu_j(t)\rangle \tag{2.3}$$

where $\mathcal{H}$ is the Hamiltonian, thus yielding

$$|v_j(t)\rangle = e^{-iE_j t}|v_j\rangle \tag{2.4}$$

In the above equation, $|v_j(t=0)\rangle$ is defined as the neutrino mass state at time $t=0$. Quantities are expressed in natural units where $c = \hbar = 1$ and $i^2 = -1$. Combining equation (2.4) with equation (2.2) results in,

$$|v_\alpha(t)\rangle = \sum_{j=0}^{3} U_{\alpha j}^* e^{-iE_j t}|v_j\rangle \tag{2.5}$$

Then combining equation (2.5) with equation (2.1) gives,

$$|v_\alpha(t)\rangle = \sum_{\beta=e,\mu,\tau}\sum_{j=0}^{3} U_{\beta j}U_{\alpha j}^* e^{-iE_j t}|v_\beta\rangle \tag{2.6}$$

Therefore, the probability that a neutrino of flavour state v_α will oscillate into v_β is written as:

$$\begin{aligned}
P(v_\alpha \to v_\beta) &= |\langle v_\beta \rangle v_\alpha|^2 \\
&= |\sum_{j=0}^{N} e^{-iE_j t} U_{\beta j}U_{\alpha j}^*|^2 \\
&= \sum_{j=1}^{3}\sum_{k=1}^{3} U_{\alpha j}^* U_{\alpha k} U_{\beta j} U_{\beta k}^* e^{-i(E_j - E_k)t}.
\end{aligned} \tag{2.7}$$

Assuming two neutrinos that are relativistic, equation (2.7) can be simplified and re-written in a more transparent way as

$$P(v_\alpha \to v_\beta) = sin^2(2\theta)sin^2\left(\frac{\Delta m^2 L}{4E_v}\right), \tag{2.8}$$

where θ is the mixing angle in the PMNS matrix, Δm^2 is the difference in squared mass of the two neutrino states, L the distance travelled by the neutrino as it transitions, and E_v the energy of the neutrinos. From equation (8) one can see that if neutrino oscillations are possible, i.e. $P(v_\alpha \to v_\beta) \neq 0$, then Δm^2 must be non-zero. The $\Delta m^2 \neq 0$ requirement means that neutrinos must have mass and the masses of different neutrino states must be different.

2.2.1 Motivation for Additional Neutrino States

The discovery of neutrino oscillations showed that at least two of the SM neutrinos have mass [18]. This observation provided strong proof of physics beyond the Standard Model. The maximum possible value of neutrino mass has been established from various experiments, including those that study cosmology and

direct measurements, like the decay of tritium. These findings challenge our current understanding of how mass is generated for fundamental particles.

The leading theoretical candidates to explain the neutrino masses involve the so-called "see-saw" mechanism [19, 20, 21]. One example model is provided by the Left-Right symmetric extension to the SM (LRSM), in which the SM group $SU_L(2)$ has a right-handed counterpart, originally introduced to explain the reason for parity non-conservation in weak interactions. The new $SU_R(2)$ group, similar to the $SU_L(2)$, predicts the existence of three new gauge bosons, $W_R^{\pm}$ and Z', and three heavy right-handed neutrino states N_ℓ ($\ell = e, \mu, \tau$), partners of the light neutrinos states ν_ℓ, and can explain the neutrinos' mass hierarchy in the context of the see-saw mechanism [22, 23, 24]. In this context, the smallness of the observed neutrino masses is due to the large mass of a new heavy state N, $m_\nu \sim y_\nu^2 \nu^2 / M_N$, where y_ν is a Yukawa coupling and ν is the Higgs vacuum expectation value in the SM.

A reference process allowed by this model is the resonant production of a W_R from Drell-Yan (DY) processes ($q\bar{q}' \to W_R$), that decays to a N_ℓ and a lepton of the same generation ($W_R \to \ell N_\ell \to \ell + (\ell q\bar{q}')$) and gives two jets and two same flavour leptons in the final state, which is depicted in Figure 6.1.
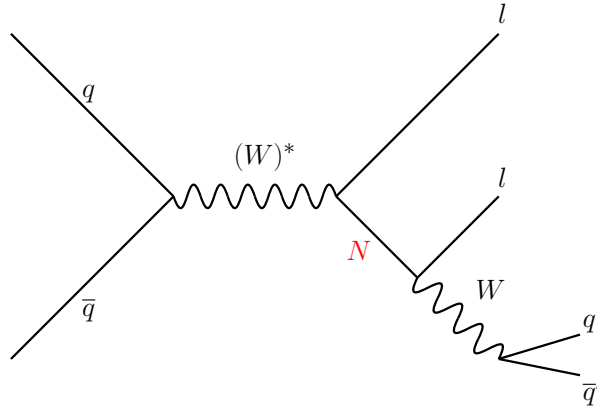
Figure 2.1: A diagram representing the DY production of a Majorana heavy neutrino from a W_R .

The strategy pursued in previous searches targeting the LRSM was to exploit the high mass scale of W_R and the N_ℓ decay to a lepton and two jets by selecting events containing two high transverse momentum (p_T) leptons (opposite-sign or like-sign charge) and two jets that are central in the detector (i.e. pseudorapidity range $|\eta| < 2.4$) [25, 26, 27, 28, 29]. In a collider experiments, transverse momentum, also referred to as transverse energy is a measurement of the particle's momentum perpendicular to the direction of the beam. Pseudorapidity is a measure of the angle between the direction of motion of a particle and the direction of the beam and it is defined as:

$$\eta = -\ln[\tan(\frac{\theta}{2})] \tag{2.9}$$

where θ is the polar angle between the particle direction and the beam direction. The pseudorapidity is useful because it provides a convenient way to describe the direction of particles in a compact range of values, with larger absolute values of pseudorapidity. Furthermore, the invariant mass distribution of the system consisting of two high-p_T leptons and jets produces an excess of signal events, at $\approx m(W_R)$, which can be utilized to discriminate against the smooth and steeply falling SM background distribution. Results of those searches in proton-proton collisions at $\sqrt{s} = 7$, 8, and 13 TeV exclude N_ℓ masses below 1.5 (0.8) TeV, assuming $m(N_\ell)$ is 0.5 (0.4) times the mass of the W_R boson [30, 31, 32]. However, the sensitivity to N_ℓ is dependent on the assumed mass of the W_R boson of the LRSM model. For example, one consequence of the assumption that N_ℓ is lighter than W_R is that no bounds on $m(N_\ell)$ exist for $m(W_R) \sim m(N_\ell)$. Additionally, if the W_R boson is too heavy to provide a large enough cross-section, regardless of $m(N_\ell)$, there is no sensitivity to N_ℓ via the standard DY searches, and thus, another technique must be devised to search for N_ℓ.

It is important to note that the addition of a new heavy neutrino may not only give an answer to the nature of neutrino masses, buy may also help understand several other problems in cosmology and high energy physics. For example, a stable heavy neutrino may be a possible candidate for DM. A heavy neutrino might help explain the matter-antimatter asymmetry of the universe, as a second and third generation of heavy neutrinos would increase the amount of CP-violation. Therefore, experimentally searching for new heavy neutrino states is of fundamental importance.

This dissertation is organized as follows: This section will highlight the status quo of particle physics and our motivation for carrying out this search. Chapter 3 will highlight the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC) apparatus used. Chapter 4 will discuss our physics objects used in the analysis. Chapter 5 will present the dataset. Chapter 6 will give the detailed analysis, results, and some expectations for analyses with data. Chapter 7 will conclude.

CHAPTER 3

The Large Hadron Collider & the Compact Muon Solenoid Detector

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC), as seen in Figure 3.1, is a particle accelerator located near the European Center for Nuclear Research (CERN) [33]. At the LHC, there are four main experiments: CMS (Compact Muon Solenoid), ATLAS (A Toroidal LHC Apparatus), ALICE (A Large Ion Collider Experiment), and LHCb (Large Hadron Collider beauty) [34]. The accelerator provides bunches of protons for six months and heavy ions for one month in a typical year of operation. The proton-proton (pp) collisions are aimed to search for and discover new physics beyond the SM (BSM). Meanwhile, the aim of the heavy ion runs is to study the dynamics of quarks and gluons at high energies. The LHC consists of an underground circular tunnel (100-150 m deep) with a circumference of 27 km, 1232 superconducting dipole magnets to guide the beam through rings, and quadrupole magnets to focus the beams. Using liquid helium, the magnets are cooled down to temperatures of 2.0 K to maintain the superconducting state.
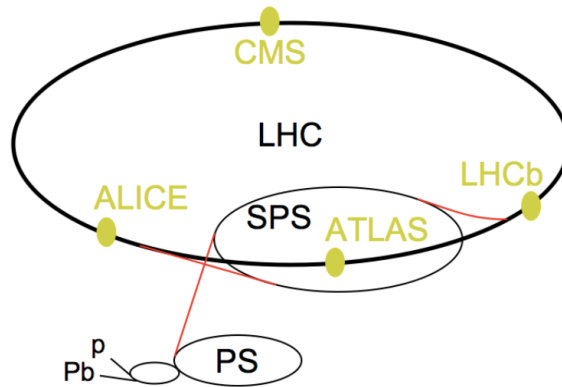
Figure 3.1: Layout of the LHC experiment. The beam transfer lines are shown in red while the four LHC collision points are shown in yellow. [3].

The protons are accelerated by a set of linear and circular accelerators, where their speed increases. The protons are clustered in bunches, spaced in gaps of 25 ns (for Run-II). Initially, in a linear accelerator called LINAC2, the protons are driven until they reach an energy of 50 MeV. Then, the acceleration process continues in the Proton Synchrotron Booster (PSB), where the particles reach an energy of 1.4 GeV. After passing by the PSB, the bunches are injected to the Proton Synchrotron (PS) accelerator, which pushes the

bunches to an energy of 25 GeV. The bunches are then passed to the Super Proton Synchrotron (SPS) where their energy increases to 450 GeV. Finally, the protons are transferred to the main ring of the LHC where theyreached a maximum energy of 6.5 TeV per beam in the Run-II period, before they collide.

When colliding particles, the number of expected events for any process can be shown:

$$N_{\text{process}} = \sigma_{\text{process}} \int L \mathrm{d}t \tag{3.1}$$

where N_{process} is the number of expected events for any specific process, σ_{process} is the associated cross section for the process of interest, and L is the luminosity, which can be seen in Figure 3.2 (left).
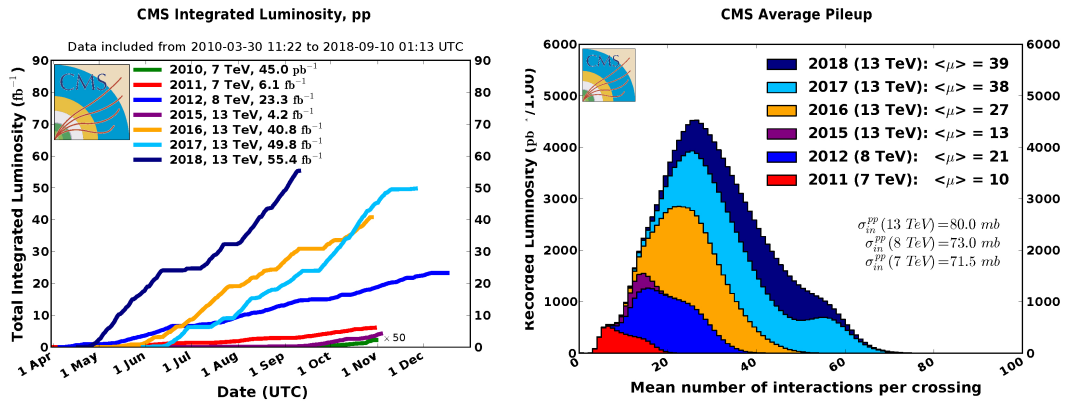
Figure 3.2: Total integrated luminosity of pp collisions by the CMS detector for each year and center-of-mass energy (left). Number of interactions per bunch crossing that is recorded by the CMS detector, also by year (right) [4].

3.2 The Compact Muon Solenoid Detector (CMS)

The central feature of the Compact Muon Solenoid (CMS) detector is a superconducting solenoid with an internal radius of 3 m. The solenoid provides a magnetic field of 3.8 T along the direction of the counterclockwise rotating beam, taken as the z axis of the detector coordinate system, with the center of the detector defined to be at z = 0. The azimuthal angle ϕ is measured in the plane perpendicular to the z axis, while the polar angle θ is measured with respect to this axis. Within the superconducting solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. The electromagnetic calorimeter provides a coverage in pseudorapidity $|\eta| < 1.479$ in the barrel region (EB) and $1.479 < |\eta| < 3.0$ in the two end-cap regions (EE), where pseudorapidity is defined as $\eta = -\ln[\tan(\theta/2)]$. Extensive forward calorimetry complements the coverage provided by the barrel and endcap detectors. Muons are measured in gas ionization detectors embedded in the steel return yoke. The first level of the CMS trigger system, com-

posed of custom hardware processors, uses information from the calorimeters and muon detectors to select up to 100 kHz of the most interesting events. The high-level trigger (HLT) processor farm uses information from all CMS subdetectors to further decrease the event rate to roughly 300 Hz before data storage [34]. A more detailed diagram of the CMS detector, can be found in Figure 3.3.
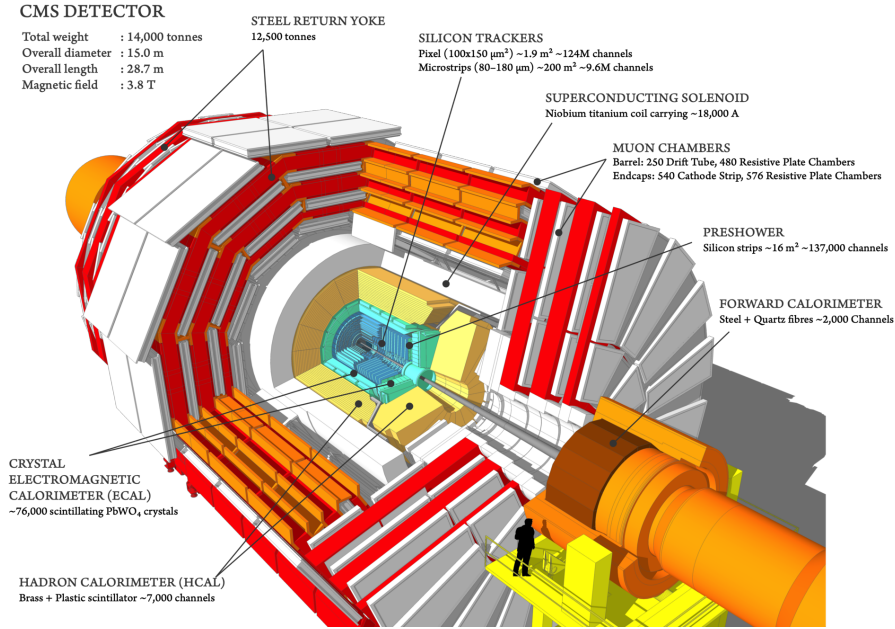
Figure 3.3: Transverse view of the CMS detector [5].

Particles traversing the detector, resulting from the pp collision, leave a different kind of signature depending on their characteristics. Particles that have an electric charge, such as electrons, muons and charged hadrons, leave a signature in the tracker system and bend in the presence of the magnetic field. By measuring the radius of curvature and the direction of the velocity of the particle with respect to the magnetic field, the momentum of the particle can be estimated using the Lorentz Equation (3.2):

$$\frac{d\vec{p}}{dt} = \frac{e}{c}\vec{v} \times \vec{B}, \qquad \frac{dE}{dt} = 0 \tag{3.2}$$

In the equation (3.2), e is the electric charge, c is the velocity of light, $\vec{v}$ is the velocity of the particle, and $\vec{B}$ is the magnetic field. Particles such as electrons and photons deposit a large fraction of their energy in the ECAL. The electrons and photons interact with the calorimeter and generate cascades of particles known as showers. On the other hand, charged pions have dominantly strong (color) interactions and deposit most of their energy in the hadronic calorimeter. Because muons are two hundred times heavier than electrons, they tend to lose a lot less energy in material than electrons do at typical CMS energies. Since muons also do not interact strongly like quarks do with a nucleus, muons are the only charged particles to pass through both

calorimeters and the return yoke of the magnet.

3.2.1 Tracking System

The tracker system is used to measure the momentum of outgoing particles from proton-proton collisions. The tracker reconstructs the paths of charged particles through the magnetic field of the detector using two subcomponents: the pixel detector and the silicon strip detector.

The pixel detector is located closest to the beamline, to handle the high particle density. When a charged particle passes through the pixel detector, it creates electron-hole pairs in the silicon material, which are collected by readout chips (ROCs) for amplification and measurement. The goal is to track the paths of particles from collisions with high precision. The pixel detector covers an area of approximately $|\eta| < 2.5$[6], and the CMS pixel module design is shown in a Figure 3.4.

(a)
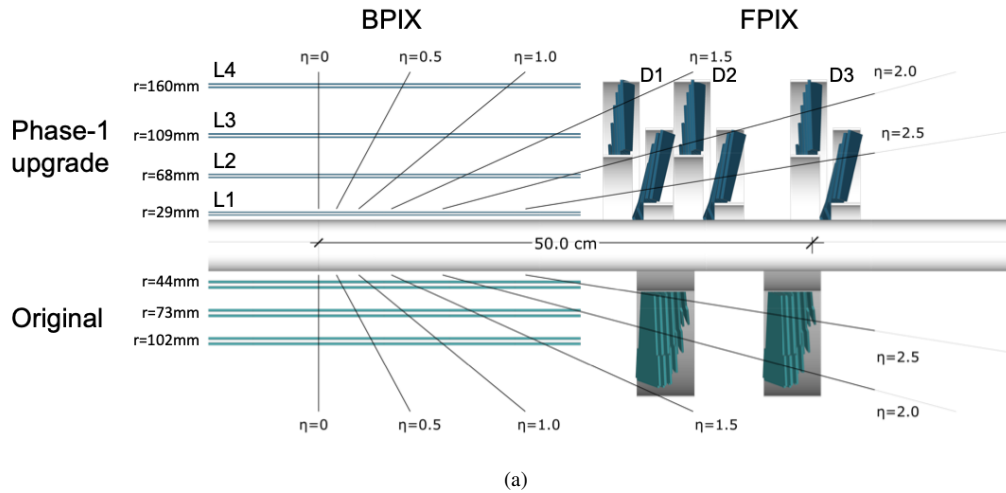
Figure 3.4: Layout of the CMS Phase-1 pixel detector compared to the original detector layout, in longitudinal view[6].

The CMS silicon strip detector is the subsequent layer of detection for the outgoing particles in the CMS detector. It consists of ten layers that extend 130 centimeters from the center of the detector. The layers contain silicon sensors with high response rates and precise spatial resolution. These sensors work similarly to the cells of the pixel detector by detecting the small current produced when a charged particle passes through and ejects electrons from the atoms. The current is collected, amplified, and readout by Analogue Pipeline Voltage (APV) chips[35], allowing for tracking of particles beyond the pixel detector. This is crucial in accurately reconstructing the paths of particles from a collision.

3.2.2 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is used to measure the energy of high-energy photons and electrons. The ECAL consists of a highly segmented array of scintillating crystals, typically lead tungstate ($PbWO_4$), interleaved with thin layers of metal [36]. In addition to $PbWO_4$ crystals, the ECAL for CMS also uses silicon avalanche photodiodes (APDs) and vacuum phototriodes (VPTs) for signal readout. These are used to convert the light produced by particles interacting with the crystals into an electrical signal that can be detected and recorded. As a photon or electron passes through the calorimeter, it ionizes the atoms in the crystal, producing a small amount of light. This light is collected by photomultiplier tubes (PMTs) and converted into a electrical signal. The signal is then amplified and recorded.

With approximately 75,000 individual cells, each equipped with its own PMT, the ECAL can measure the energy deposited by incoming particles with a high degree of precision. The ECAL covers a wide pseudorapidity range, extending out to $|\eta| < 3$ [7]. Figure 3.5 shows a 3D view of the ECAL detector with various sub-detectors (top) as well as coverage in η (below).The energy deposited by a particle in the ECAL can be calculated as follows:

$$E = \int \frac{dE}{dx} dx \tag{3.3}$$

where dE/dx is the energy deposited per unit length and dx is the length of the particle's trajectory through the calorimeter material. In addition to measuring the energy of incoming particles, the ECAL can also be used to identify electrons and photons. This is accomplished by using the shower shape information provided by the calorimeter's segmentation. The shape of the shower produced by an electron is different from that produced by a photon, and this difference can be used to distinguish between the two types of particles. In the barrel section of the ECAL, the detector is made up of a single crystal. In this case, the distinction between electrons and photons is made using the crystal's geometry and the pattern of energy deposition within the crystal. When an electron or photon enters the crystal, it produces a shower of secondary particles through electromagnetic interactions. The shape of the shower depends on the energy and type of the incoming particle, as well as the crystal's geometry. The shower spreads out in a characteristic pattern that can be measured by the photodetector at the end of the crystal. By analyzing the shape and pattern of the energy deposition within the crystal, we can distinguish between electrons and photons with high efficiency. A pre-shower detector (ES) is in front of the two endcaps (EE) to help seperate decays of neutral pions into two close photons. Overall, the ECAL is a critical component of the CMS detector, providing measurements of the energy of incoming photons and electrons and helping to identify the particles produced in proton-proton collisions.
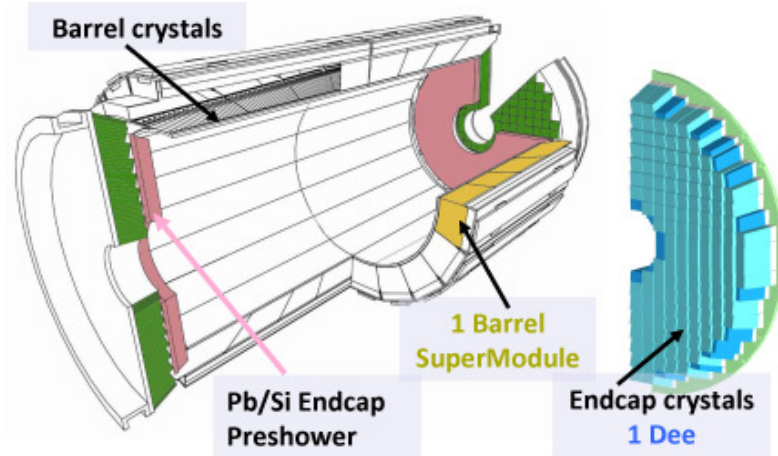
(a)
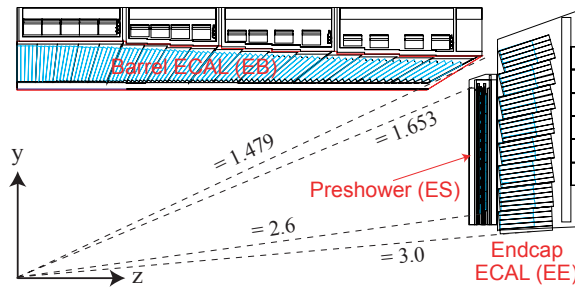
(b)

Figure 3.5: 3D view of the CMS ECAL, showing the location of various sub-detectors (top). Coverage by ECAL in η (below) [7].

3.2.3 Hadronic Calorimeter

Missing Transverse Energy ($\not{E}_{\mathrm{T}}$ or MET) represents the imbalance of energy and momentum in the transverse plane of a particle collision events. In particle collisions, particles are produced in all directions, including those that are not detected. The MET is therefore the sum of the transverse momentum of all the undetected particles, which is inferred from the conservation of momentum in the transverse plane.

The HCAL is essential for measuring the energy and location of showering hadrons as well as the direction of the $\not{E}_{\mathrm{T}}$ by detecting quarks, gluons, and neutrinos. In the "reconstruction" of neutrinos, the HCAL plays an important role in determining the missing transverse energy in an event. The $\not{E}_{\mathrm{T}}$ is a key quantity in the search for neutrino interactions as it represents the transverse momentum imbalance in the event, which is attributed to the undetected neutrinos. By measuring the energies of the hadrons in the HCAL, it is possible to determine the total energy deposited in the detector and subtract it from the known initial energy of the incoming beam particles to calculate the MET. This information, combined with the information from the other sub-detectors

13

in the CMS, can help to reconstruct the neutrino interactions and improve the understanding of the underlying physics processes[37]. Together with the data from the ECAL and the muon system, this system aids in the accurate identification of electrons, photons, and muons.

The hadron barrel (HB), hadron endcap (HE), hadron forward (HF), and hadron outer (HO) calorimeters are the main subdetectors that make up the HCAL. The HCAL's functions include energy measurement and hadron identification. Divided into two sub-detectors: The barrel (HB), which provides coverage to $|\eta| = 1.4$; the endcap (HE), which provides coverage between $1.3 < |\eta| < 3.0$; and the forward calorimeter (HF), which has a coverage from $2.9 < |\eta| < 5.0$. Outside of the magnetic solenoid is where the HO is. It has 12 f-sectors in each of its 5 rings. An f-sector refers to a specific portion or segment of the HCAL that is used to measure the energy of hadrons (particles consisting of quarks and gluons) produced in proton-proton collisions. The energy deposited by these hadrons in the f-sector is used to calculate the total energy deposited in the HCAL and to reconstruct the properties of the colliding particles[38]. A ring in the HCAL refers to a circular grouping of calorimeter cells, which are used to measure the energy of hadrons produced in proton-proton collisions.

Fibers are used to transmit the light from the scintillators in the HB and HE calorimeters to the hybrid photodiodes (HPDs). A charge integrator and encoder then digitize each signal at intervals of 25 ns (QIE). In the HO, silicon photomultipliers (SiPMs) fibers capture the light.
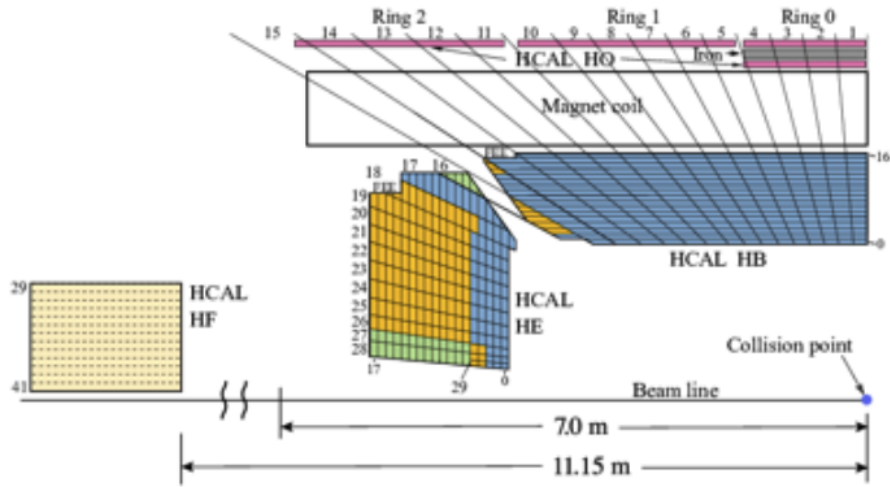
(a)

Figure 3.6: Schematics of a quadrant of the tracking and calorimeters at CMS

3.2.4 Superconducting Solenoid Magnet

The 3.8T superconducting solenoid magnet encircles and encloses the inner tracker, ECAL, and HCAL. Because of the large momentum of charged particles produced in CMS, a strong magnetic field is necessary to bend them for measurement. Figure 3.7 depicts a representation of the magnetic field inside the CMS detector. A conductor made up of three parts—a Rutherford type superconducting cable, a high purity aluminum stabilizer, and an aluminum alloy reinforcement—carries a nominal current of 19.5 kA, which generates the magnetic field [39].
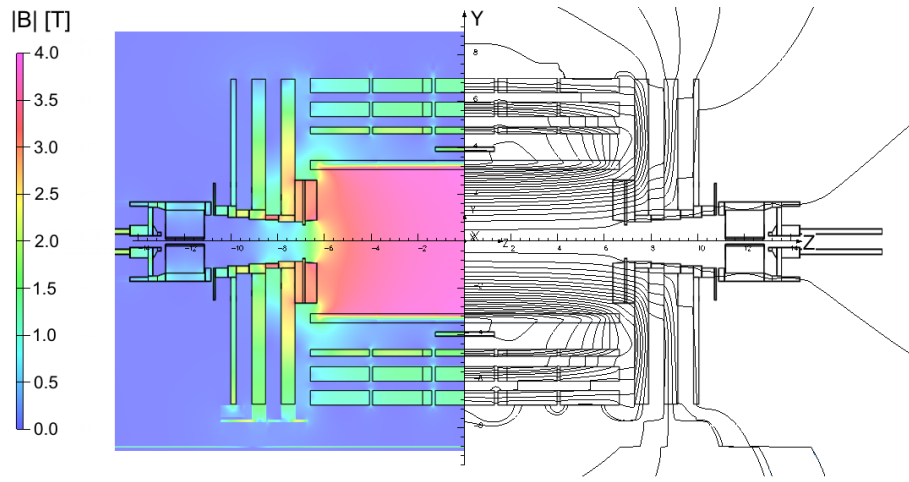
Figure 3.7: Value of $|B|$ (left) and field lines (right) predicted on a longitudinal section of the CMS detector, for the underground model at a central magnetic flux density of 3.8 T. Each field line represents a magnetic flux increment of 6 Wb [8].

3.2.5 The Muon System

The muon system enables accurate identification of muons by matching muon hits and tracks from the tracker. Further, tracker and muon tracks are merged for a better location, momentum, and temporal resolution than either subdetector could achieve separately. The muon system is has two main regions, the barrel region which covers $|\eta| < 1.2$, and the endcap region which covers ($0.9 < |\eta| < 2.4$), as can be seen in Figure 3.8. Three different gas ionization techniques are used by the muon system: drift tubes (DT), cathode strip chambers (CSC), and resistive plate chambers (RPC). The operational capabilities of the muon system are as follows [9]:

- Fast dedicated trigger detectors (RPCs) and detectors with exact spatial resolution can be combined to produce excellent trigger performances on single and multi-muon events and an unmistakable identification of the bunch crossing (DTs, CSCs).

- Three technologies are integrated to create redundancy in the trigger and reconstruction, resulting in the presence of two separate muon systems throughout the whole angular region.

- Operation in a 4 Tesla magnetic field with the ability to measure muons tracks twice—once in the tracker and again in the muon spectrometer—in order to provide excellent results in the measurements of muon momentum and charge across the whole *eta* region, ranging from a few GeV to a TeV.

- High efficiency muon identification rate ($> 95\%$) up to $\eta = 2.4$.
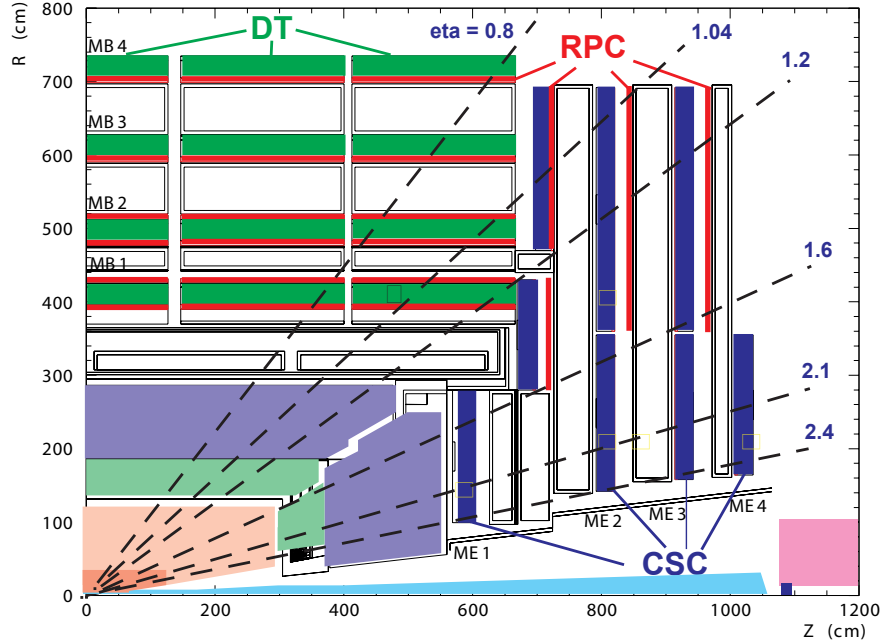
Figure 3.8: Layout of the muon system, highlighting the drift tubes (DT), resistive plate chambers (RPC), and cathode strip chambers (CSC) while also highlighting the barrel region (green) and endcap region (blue) [9].

A cathode wire is located in the core of each DT Chamber, which is made up of several layers of drift tubes arranged in a cylinder form with a readout plane on the exterior. The ionization that the muons leave in the Ar(85%)/C0$_2$(15%) gas mixture, kept at atmospheric pressure as they travel through the drift tubes is what the DTs use to detect the passage of muons. The position of the muon traveling through the chamber is identified using the information that the drift duration of the ionization electrons is related to the radial distance from the cathode wire. To offer an accurate measurement of muons, the DT Chambers are utilized in conjunction with other muon detecting elements, such as the RPCs and CSCs.

The CSCs are gas detectors that use the ionization muons produce as they move through the gas volume of the chamber to calculate their position and momentum. With other hits in the muon system The CSCs are made out of a stack of thin metal strips sandwiched between layers of insulation within a gas volume (40% Ar, 50% C0$_2$, and 10% CF_4). The muons' ionization electrons are drawn to the strips, which form an electrical charge pattern corresponding to the muon's location along the strip. The electronics at either end

of each strip read out the charge, which is then used to calculate the muon's position and momentum. In the endcap regions of the CMS detector, the CSCs measure muons location and direction.

The RPCs are gas detectors that use the ionization muons make as they move through the gas volume of the chamber to determine their position. Two parallel plates make up an RPC, one of which is resistive, acts as the cathode and the other as the anode. The cathode gathers the ionization electrons that muons produce, which results in a current pulse that is proportional to the position of the muon. The electronics read out the current pulse and use that information to locate the muon. The RPCs quickly and effectively measure muons in the CMS detector's barrel areas[40].

3.3 Triggering & Data Acquisition

The massive volumes of data that the detector records must be filtered and events must be chosen by the Trigger and Data Acquisition System (TriDAS) system. The ultimate objective is to decrease the amount of events that are being recorded from millions to roughly 300 occurrences per second for further analysis. Events must fulfill three sets of choices known as trigger levels in order to accomplish this. The detector's event rate is intended to be reduced to 100 kHz with a trigger latency of less than 4 μs at the first level, or Level 1 (L1), using a series of tests based on electronic signals. The L1 trigger algorithms look for distinguishing characteristics such big energy depositions in the ECAL or HCAL, high momentum tracks in the muon system, or a combination of the two. A flow chart of the L1 trigger process can be seen in Figure 3.9.

The following level of triggers, known as Level 2 (L2) and Level 3 (L3) triggers, further reduce the number of occurrences by adding more checks. This level is referred to as the High Level Trigger (HLT). At L2, judgments made during the L1 trigger are validated using more complicated algorithms that benefit from the detector's finer granularity. The output events from L2 finally go through a more thorough reconstruction process at L3. Higher-level objects, also known as trigger objects, are produced during this process. The physics objects employed in data selections are produced by using these refined trigger objects for particle reconstruction and identification.
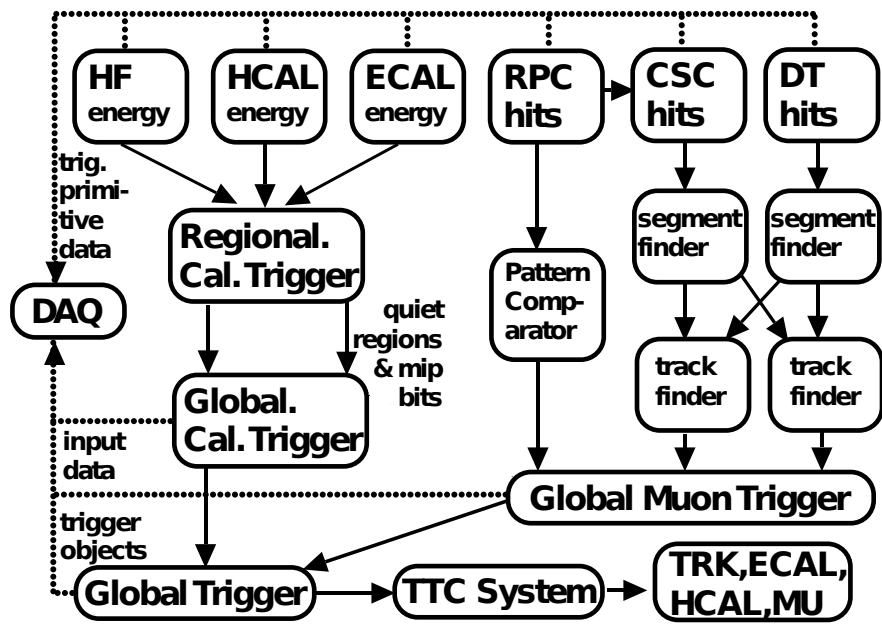
Figure 3.9: The CMS L1 trigger system overview: The ECAL and HCAL data are processed regionally (RCT) and then globally (GCT). The energy depositions (hits) from the resistive-plate chambers (RPC), cathode strip chambers (CSC), and drift tubes (DT) are processed and sent to the global muon trigger (GMT). The GCT and GMT information is combined in the global trigger (GT) which makes the final trigger decision. This decision is conveyed to the tracker (TRK), ECAL, HCAL, or muon systems (MU) via the trigger, timing, and control (TTC) system. The data acquisition system (DAQ) retrieves data from various subsystems for offline storage. [10].

CHAPTER 4

Event Reconstruction & Particle Identification

4.1 Particle Flow Algorithm

The CMS experiment employs the particle flow algorithm to identify and reconstruct the particles that impact the detectors. The algorithm operates through an iterative tracking procedure that ensures the avoidance of duplicated tracks or energy deposits. The following steps are involved in the particle flow algorithm:

- The matching of hits in the muon chamber with tracks in the inner tracker to form "Particle-Flow Muons."

- The reconstruction of electrons by combining tracks and ECAL deposits to form "Particle-Flow Electrons." This is achieved using the Gaussian-Sum Filter method, which is a technique for filtering signals with Gaussian distributions.

- The matching of HCAL hits with tracks to create "Particle-Flow Charged Hadrons." These particles are calibrated assuming they are charged pions.

- The identification of remaining ECAL deposits as "Particle-Flow Photons" and remaining HCAL deposits as "Particle-Flow Neutral Hadrons." The neutral hadrons are given the same energy and mass calibration as charged hadrons.

Overall, the particle flow algorithm takes advantage of the entire CMS detector to reconstruct each particle, making use of the iterative tracking process to prevent double counting of tracks and energy deposits.

The Particle Flow algorithm is a crucial component of the CMS experiment, providing improved performance and results. However, in some rare instances, it can produce undesirable outcomes. For example, events with high-energy hits at the edge of the HCAL were found to result in elevated values for E_T^{miss}. Investigation revealed that these were energetic photons hitting the small portion of the HCAL that is not covered by the ECAL. As there were no corresponding track or ECAL hit, the Particle Flow algorithm mistakenly assumed them to be neutral pions and applied energy calibrations intended for pions rather than photons [41]. This resulted in a disruption of the momentum balance of the system [42]. Figure 4.1 shows a brief overview of how various particles interact with the different detector layers, which will be expanded upon in greater detail in the following sections.
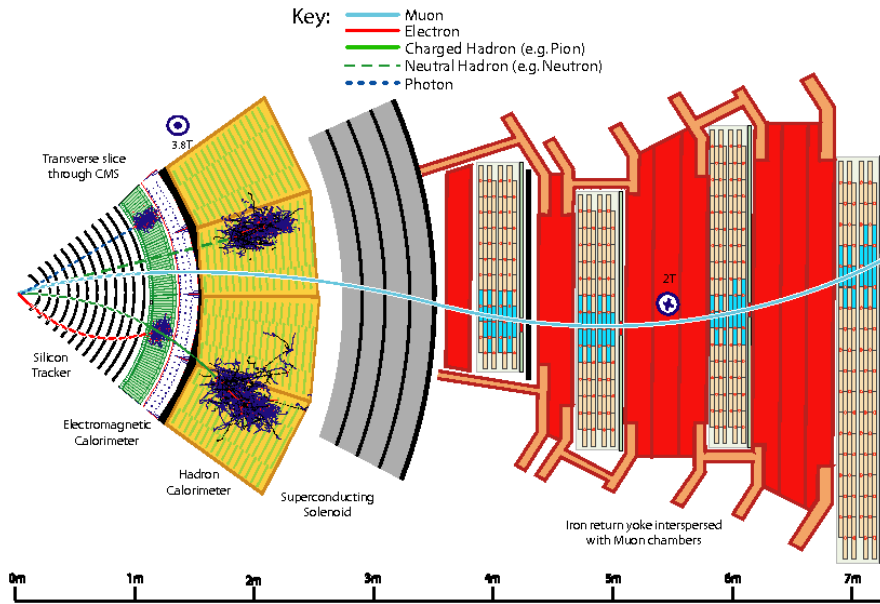
Figure 4.1: Diagrammatic overview of particle interactions with various CMS subdetector layers. Neutrinos notably do no interact with any detector layers, however their production can be inferred when "missing" transverse momentum ($p_{\mathrm{T}}^{\mathrm{miss}}$)is reconstructed in the detector [11].

4.2 Jets

Jets are objects resulting from the decays of quarks or gluons in high-energy particle collisions. Due to confinement in QCD, produced quarks or gluons cannot exist in isolation and quickly undergo hadronization, which produces numerous particles often collimated into a single jet. The composition of a jet can vary and can include hadronic material, photons from π^0 decays, and even muons or electrons from b decays. In the CMS experiment, the particle flow (PF) algorithm is used to identify particles produced in each event, which are then grouped into jets using a jet clustering algorithm. There are many different types of jets, depending on factors like the origin of the jet (b-jet versus pile-up jet), the flavor of the underlying quark that produced the jet (b-jet), whether the jet originated from initial state radiation (ISR) or final state radiation (FSR), and even the algorithm used to group and identify the jet (AK4 jets, AK8 jets). The PF algorithm is used to identify the particles produced in each event, [43] by combining the data from each sub-detector. One commonly used jet clustering algorithm is the anti-k_T algorithm, which prioritizes high momentum particles and results in circular jets in the $\phi - \eta$ plane. Different jet clustering algorithms can be used to achieve different results and jets can be categorized based on their origin, quark flavor, radiation source, and clustering algorithm.

20

4.3 $\not{E}_T$

In high energy particle collisions, some of the energy produced in the event may be undetected, leading to an imbalance in transverse energy. This imbalance is referred to as "missing transverse energy" ($\not{E}_T$) or MET and is calculated as the magnitude of the vector sum of transverse momentum of all reconstructed particles in the event, excluding muons. $\not{E}_T$ is an important observable used to search for new physics phenomena, such as the production of dark matter particles or the observation of new particles that escape detection.

The calculation of $\not{E}_T$ is done by reconstructing the energy and momentum of all particles in the event, including jets, leptons, and photons:

$$\not{E}_T = -\sum_{Objects} \vec{p_T^i} \tag{4.1}$$

These quantities are measured by the various sub-detectors of the CMS experiment, and are then combined into a single particle-level measurement using the particle flow (PF) algorithm. The PF algorithm aims to reconstruct individual particles and measure their energy and momentum with high accuracy. The final step in the calculation of $\not{E}_T$ involves taking the negative vector sum of the transverse momentum of all reconstructed particles.

$\not{E}_T$ provides valuable information about the nature of the particles produced in particle collisions, and can help to uncover new physics phenomena that would otherwise go undetected. It should be noted that $\not{E}_T$ can also result from other effects, such as detector mis-calibration or instrumental effects, and care must be taken to accurately estimate and correct for these contributions [44].

4.4 Electrons & Photons

The reconstruction of electrons and photons in the experiment is similar in nature. Both objects are expected to produce significant energy deposits in the ECAL and electrons, being charged, will also leave tracks in the inner tracker. As the particles move through the detector, they can interact with the surrounding material, resulting in the emission of bremsstrahlung photons in the case of electrons and conversion into electron-positron pairs for photons. The end result is the reconstruction of the original electron or photon as a shower of multiple electrons and photons. An algorithm is then applied to combine the individual particles into a single object and estimate the energy of the initial electron or photon. The electron's momentum also changes due to the emission of bremsstrahlung photons, leading to a change in its curvature, which is taken into account while estimating the track parameters of electrons [45].

The full reconstruction process starts by grouping together crystals with energies exceeding a predefined threshold to form clusters. The cluster with the most energy in a specific region is then defined as the seed

cluster, with a minimum transverse energy ($E_\mathrm{T}^\mathrm{seed}$) above 1 GeV. The ECAL clusters within a defined window around the seed cluster are combined into superclusters, which are then linked to compatible pixel detector tracks. All tracks in the event are tested for compatibility with the electron trajectory hypothesis and ECAL clusters, SCs, and tracks are linked together into particle blocks. These blocks are then resolved into electron and photon objects based on loose selection criteria and a tighter selection is applied to identify electrons and isolated photons [45]. If the object fails both criteria, it is further considered as neutral hadrons, charged hadrons, or nonisolated photons in the PF framework.

The electron reconstruction efficiency is higher than 95% for $E_\mathrm{T}(e) >$20 GeV, determined using a tag-and-probe method with $Z \rightarrow ee$ events. However, the rate of other objects being misidentified as electrons, or fake electrons, can impact the results when studying rare processes with very small cross sections, where the fake electron rate is higher than the production rate of such events.

4.5 Muons

Muons are reconstructed using a combination of information from several subdetectors, such as, from the inner tracker and the muon system. The first step in muon reconstruction is to identify muon candidates in the muon system, which consists of several layers of detectors that are designed to measure the trajectories of muons passing through the the DTs, CSCs, and RPCs detectors. Once a muon candidate has been identified, the information from these subdetectors is combined to determine the muon's momentum and direction. This information is used to form a track, which represents the path of the muon through the detector. In addition to tracking information, muon identification also use information from the ECAL, which measures the energy deposited by the particles passing through the detector. The information is used to identify the type of particle, such as a muon or a pion, and to determine the total energy deposited by the particle. Overall, a combination of information from several subdetectors is used to accurately reconstruct muons and other particles [12]. A number of identification (ID) types for muons are defined for use in CMS analyses. These include:

- "Loose Muon ID" designed to identify muons that are either prompt or result from light or heavy flavor decays. The aim is to keep a low rate of false identification of charged hadrons as muons. A "Loose Muon" is one selected by the PF algorithm and is either a tracker or a global muon.

- "Medium Muon ID" for prompt muons and those from heavy flavor decays. It is a "Loose Muon" with a tracker track that uses hits from more than 80% of the inner tracker layers it traverses. The constraints on the segment compatibility, where the track of a muon passing through the muon system matches with the signals recorded by the system's detector segments, were tuned after the application of the other constraints to target an overall efficiency of 99.5% for muons from simulated W and Z

events.

- "Tight Muon ID" designed to minimize the number of muons from decay in flight or hadronic punch-through. A "Tight Muon" is a "Loose Muon" with a tracker track using hits from at least six layers of the inner tracker, including at least one pixel hit. The muon must be reconstructed as both a tracker and global muon, with the tracker muon having segment matching in at least two muon stations and the global muon fit having a χ^2/dof of less than 10 with at least one hit from the muon system. Additionally, the muon must be compatible with the primary vertex, having a transverse impact parameter $|dXY|$ of less than 0.2 cm and a longitudinal impact parameter $|dz|$ of less than 0.5 cm.

The resulting efficiencies of these IDs can be seen in Figure 4.2 (LooseID, left; TightID, right) for $p_T(\mu) > 20$ GeV, as derived using a $Z \to \mu\mu$ tag and probe method.

(a)   (b)
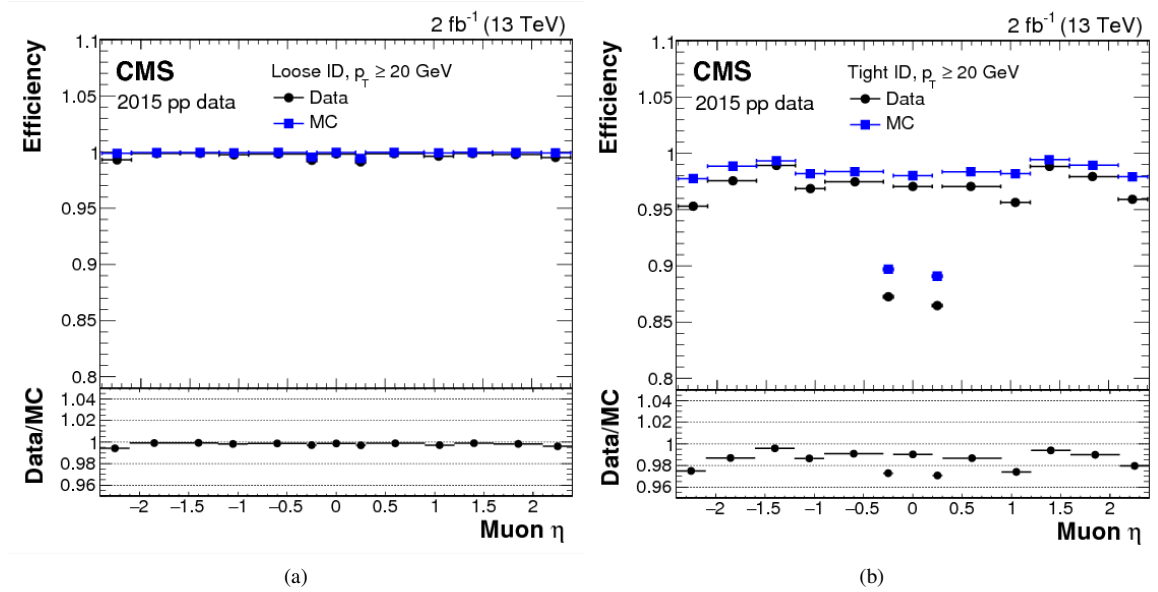
Figure 4.2: Muon identification efficiencies for LooseID (left) & TightID (right) for $p_T(\mu) > 20$ GeV. Data to simulation (MC) agreement is shown below each respective plot. Results were derived using a $Z \to \mu\mu$ tag and probe method [12].

4.6 Hadronic Taus

Tau leptons differ from electrons and muons in that they have enough mass to decay into hadrons approximately 60% of the time. The remaining 40% of the time, they decay into lighter leptons. Detecting hadronic taus (τ_h) in experiments can be challenging as their hadronic jets can resemble those produced by much more common QCD processes but can be identified since there isn't fragmentation and the jets tend to be more collimated. Unlike leptonic taus, hadronic taus need to be reconstructed since leptonic taus appear in the final state as electrons and muons, along with missing momentum from neutrinos [46].

The most widely used method for identifying hadronic taus in the Hadrons Plus Strips (HPS) algorithm. Due to conservation of tau charge, the decay products must include an odd number of charged hadrons, primarily pions, while the neutral components, primarily π^0's, quickly decay into photons. The photons can generate electron-positron pairs which bend in different directions in ϕ, resulting in an azimuthal strip of energy in the ECAL. Hence, the name "Hadrons Plus Strips" [47]refers to the charged and neutral components of the tau decay. HPS algorithm uses as its input anti-k_T jets with $p_\mathrm{T} > 14$ GeV and $|\eta| <2.5$ and performs the following Reconstruction combinations:

- The PF algorithm used to construct combinations of charged and neutral particles consistent with specific τ_h decays. The four-momentum (consisting of p_T, η, ϕ, and mass) of τ_h candidates is then determined.

- Discriminators are used to filter τ_h decays from quark and gluon jets, as well as from electrons and muons. This reduces the occurrence of misidentification in τ_h decays from jets, electrons, and muons.

In this study, all referred hadronic taus are required to be "single prong." This means that the focus is on tau decays that result in a single charged pion, in order to minimize contamination from QCD background.

CHAPTER 5

Analysis Strategy & Monte Carlo Simulation

5.1 Vector Boson Fusion

Vector Boson Fusion (VBF) refers to the fusion of two electroweak vector bosons, typically W or Z bosons, to produce a final state with a high-mass object. This process occurs via the exchange of two intermediate vector bosons in the s-channel, resulting in two highly energetic jets in the forward and backward regions of the detector. The presence of two widely separated forward jets in VBF events provides a distinctive experimental signature, which can be used to efficiently separate the VBF signal from other Higgs production mechanisms, such as gluon fusion or associated production with a vector boson.

Furthermore, VBF processes have favorable kinematic properties, such as large dijet invariant mass (m_{jj}), which is calculated by taking the invariant mass between the two leading jets within an event and can be expressed mathematically as a function of the individual jet p_T and $\Delta\eta_{jj}$ as:

$$m_{jj}^2 \propto 2p_T(j_1)p_T(j_2)\cosh(\Delta\eta_{jj}) \tag{5.1}$$

and large pseudorapidity separation between the two jets ($\Delta\eta_{jj}$), that enable the reconstruction of the Higgs boson mass peak with high precision. These kinematic properties are particularly useful in precision studies of the Higgs sector, where they can provide information on the Higgs couplings to vector bosons, top quarks, and other SM particles [48, 49, 50]. Hence, VBF events are a crucial tool for exploring the mechanism of electroweak symmetry breaking and the nature of the Higgs boson. The clean experimental signature of VBF events and the possibility of high-precision measurements also make them a valuable source of information in searches for new physics beyond the Standard Model.

The focus of this thesis is a search for a new un-observed heavy neutrino state by targeting N_ℓ production via VBF processes [51]. A search for N_ℓ using the VBF topology has not been performed before at a collider, but may present an important avenue for discovery [52]. Figure 5.1 shows an example Feynman diagram for the VBF N_ℓ production mechanism, in particular $W\gamma$ fusion in t-channel diagram containing an off-mass shell W boson. The N_ℓ subsequently decays to a lepton and two jets, resulting in a final state of two forward VBF jets, two additional central jets, and two leptons. Of particular interest for this thesis is the VBF production of the second-generation heavy Majorana Neutrino N_μ in the di-muon plus multijet VBF final state [53].

The production of N_μ via VBF can result in both opposite-sign and same-sign charged dilepton pairs (each giving you half the total contribution). We consider same-sign charge since this is a cleaner signature due to

(a)
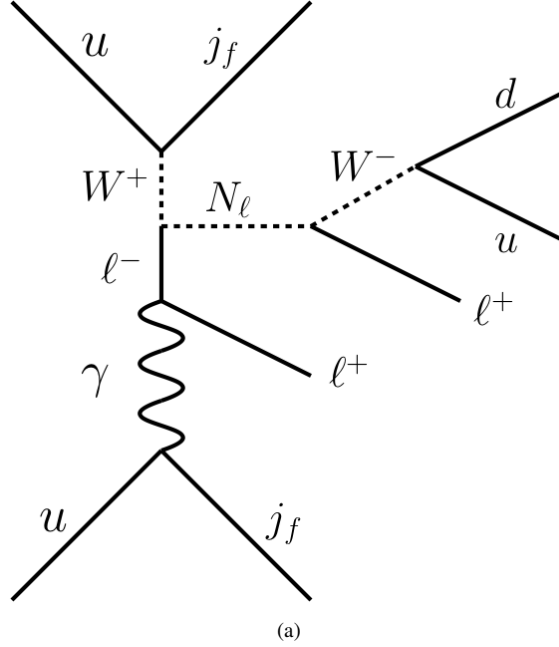
Figure 5.1: A diagram representing the VBF production of a Majorana (fermion that is its own antiparticle) heavy neutrino.

the reduction of SM contributions from $Z(\rightarrow \ell\ell)$ and top quark pair ($t\bar{t}$) backgrounds where opposite-charged pairs contribute. In sections to follow, the same-sign and opposite-sign search sensitivities are compared.

In addition to the two oppositely charged muon candidates defining the final state, the search will require the presence of multiple jets. When the heavy neutrino mass becomes larger, the two jets from the W decay (i.e., $N_\mu \rightarrow W\mu \rightarrow jj\mu$) merge into a single jet since the W is significantly boosted from the N_τ decay. To account for this, we require at least 3 jets, two VBF jets and a third jet with a significantly broader energy flow in the detector. The features of the two VBF jets are that they are forward, in opposite hemispheres of the detector, and have TeV-scale dijet mass. The VBF event selection provides a remarkable handle to help suppress background contributions.

The dominant background, representing $\sim 80\%$ of the total background, is $t\bar{t}$ with associated jets from top decays, two muons from W decays (or one from W and one fake). We will define control samples which are orthogonal to the signal region by modifying a few of the signal selections (i.e. requiring b-tagged jets). These control samples are dominated by $t\bar{t}$ events with $> 90\%$ purity. We will then use these high purity control samples to measure the VBF tagging efficiencies in the data and Monte Carlo (MC) simulated samples and extract correction factors which can be used to correct the $t\bar{t}$ MC prediction in the signal region. The other contributing backgrounds, accounting for $\sim 20\%$ are single top, and diboson. These processes become contributing backgrounds when a dimuon pair is produced from a pair of W decays and multiple jets

resulting from initial state radiation, decays of top quarks, and/or decays of W bosons.

5.2 Monte Carlo Sample Production

The events in this analysis (both signal and background) are generated using the Monte Carlo (MC) method. MC uses random numbers to obtain numerical results. For MC event generators that simulate pp collisions, partons are generated with randomly sampled momenta based on a parton distribution function. The interactions between particles are based on the theory being modeled, and momentum and energy are conserved at each vertex. If interactions are only considered at the tree level, the samples are called leading order (LO). Higher order effects can also be calculated, referred to as next-to-leading order (NLO) or next-to-next-to-leading order (NNLO).

The VBF N_ℓ signal production decay process and the $Z/\gamma^*(\to \ell^+\ell^-)$ + jets, $Z(\to \nu_\ell \bar{\nu}_\ell)$ + jets, $W(\to \ell\nu_\ell)$ + jets backgrounds are simulated at LO precision using MadGraph (MG)[54], where up to four partons in the final state are included in the matrix element calculation. The background processes involving the production of a single vector boson in association with two jets exclusively through pure electroweak interactions are simulated at LO via MG. The QCD multijet background is also simulated at LO using MG. Single top quark and $t\bar{t}$ processes are generated at next-to-leading order (NLO) using the POWHEG generator. The leading order Pythia generator is used to model the diboson processes. The POWHEG and MG generators[55] are interfaced with the Pythia program, which is used to describe the parton shower and the hadronization and fragmentation processes with the CUETP8M1 tune[56]. The LO cross sections are used to normalize simulated signal events, while NLO cross sections are used for simulated backgrounds [57, 58]. For both signal and background simulated events, additional pp interactions (pileup) are generated with Pythia and superimposed on the primary collision process. The simulated events are reweighted to match the pileup distribution observed in data. The simulated samples are processed with a detailed simulation of the CMS apparatus using the GEANT4 package [59]. For all samples, the production of the following process was included: $pp > N_\ell \ell \ell jj$. This was achieved using the following example MG generate command:

```
import model SM_HeavyN_NLO

define mu = mu+ mu-

generate p p > n2 mu j j QCD=0, (n2 > mu j j)
```

The background samples for this analysis are all official CMS MC samples created using the LO event generators MadGraph, POWHEG, and Pythia 8. The list of these samples can be seen in Appendix A.3. The cross sections were calculated using NLO or NNLO diagrams where possible. Samples for DY+jets and W+jets were only created for $H_T > 100$ GeV, so a filter was used to select for $0 < H_T < 100$ GeV, which was applied to the inclusive DY+jets and W+jets samples.

27

5.3 Data Samples

This analysis utilizes proton-proton collision data with a center-of-mass energy of 13 TeV collected by the CMS detector during the 2016, 2017, and 2018 data-taking runs. The data is processed in the NanoAOD format (refer to reference [60] for more information). The total integrated luminosity from the three years of data taking is 137.19 fb^{-1}, including 35.92 fb^{-1}, 41.53 fb^{-1}, and 59.74 fb^{-1} from each year, respectively. Multiple primary datasets will be utilized for different purposes. In the signal search region, the muon PD (found in Table 5.1) will be used for background estimation.

Run II collision data samples: muon primary datasets (NanoAODv6).

Era	Physics sample	Official CMS datasets
2016	Run 2016Bv1	/SingleMuon/Run2016B_ver1-Nano25Oct2019_ver1-v1/NANOAOD
	Run 2016Bv2	/SingleMuon/Run2016B_ver2-Nano25Oct2019_ver2-v1/NANOAOD
	Run 2016C	/SingleMuon/Run2016C-Nano25Oct2019-v1/NANOAOD
	Run 2016D	/SingleMuon/Run2016D-Nano25Oct2019-v1/NANOAOD
	Run 2016E	/SingleMuon/Run2016E-Nano25Oct2019-v1/NANOAOD
	Run 2016F	/SingleMuon/Run2016F-Nano25Oct2019-v1/NANOAOD
	Run 2016G	/SingleMuon/Run2016G-Nano25Oct2019-v1/NANOAOD
	Run 2016H	/SingleMuon/Run2016H-Nano25Oct2019-v1/NANOAOD
2017	Run 2017B	/SingleMuon/Run2017B-Nano25Oct2019-v1/NANOAOD
	Run 2017C	/SingleMuon/Run2017C-Nano25Oct2019-v1/NANOAOD
	Run 2017D	/SingleMuon/Run2017D-Nano25Oct2019-v1/NANOAOD
	Run 2017E	/SingleMuon/Run2017E-Nano25Oct2019-v1/NANOAOD
	Run 2017F	/SingleMuon/Run2017F-Nano25Oct2019-v1/NANOAOD
2018	Run 2018A	/SingleMuon/Run2018A-Nano25Oct2019-v1/NANOAOD
	Run 2018B	/SingleMuon/Run2018B-Nano25Oct2019-v1/NANOAOD
	Run 2018C	/SingleMuon/Run2018C-Nano25Oct2019-v1/NANOAOD
	Run 2018D	/SingleMuon/Run2018D-Nano25Oct2019-v1/NANOAOD

Table 5.1: Run II Muon Primary Datasets

5.4   Data Corrections

Although the CMS detector has been in excellent operation for over a decade, there are some known problems that impact the ability to record and measure certain physical objects. These issues were not taken into account or corrected in MC samples, as they were generated before the problems were identified. Thus, specific adjustments are necessary to achieve consistency between data and MC and to avoid producing biased results. These adjustments descibed below include "L1-PreFiring" for 2016 and 2017 samples, "EE noise veto" for 2017 samples, and "2018 HEM veto" for 2018 samples.

5.4.1   L1 Pre-firing (2016 & 2017)

The L1 Pre-firing problem is caused by a non-negligible probability of the L1 trigger being activated by calorimeter activity before the actual collision event. This leads to a reduction in the recorded events for certain physics processes, especially for high-p_T objects, and can result in a biased selection of events in data compared to simulation. The problem was particularly prevalent in the 2016 and 2017 data-taking periods. To address this issue, specific corrections by the *EGamma* Physics Object Group (POG) at CMS were implemented in the data analysis to account for the effects of the L1 pre-firing and ensure good agreement between data and simulation. The following is the recipe for the probability of an event to pre-fire as a function of the jets and photons present in an event, which can then be applied to MC in the form of an event weight [61]:

$$\omega_{\text{L1PreFiring}} = 1 - P(\text{PreFiring}) = \prod_{i=\text{photons,jets}} (1 - \varepsilon_i^{pref}(\eta, p_{\text{T}})) \tag{5.2}$$

where $\varepsilon_i^{pref}(\eta, p_{\text{T}})$ is the pre-firing probability of a photon/jet measured as a function of p_{T} and η, and ω is the resulting weight derived. In addition, overlap removal between jets and photons is applied.

5.4.2   EE Noise Veto (2017)

In the CMS experiment, an issue with additional noise in the ECAL was discovered in the tails of the $p_{\text{T}^{miss}}$ distributions in 2017. This was due to several factors such as ECAL aging in the high η region, out-of-time pile-up, and selective readout. To resolve the issue, the data and simulation were modified by removing jets and unclustered particles with $p_{\text{T}^{raw}} < 50$ GeV and $2.65 < |\eta| < 3.14$ from the calculation of PF p_{T}^{miss}. Only jets above a threshold of 15 GeV were considered for jet energy resolution (JER) smearing(discussed more in Chapter 8), and these jets were removed from the analysis. However, after the modifications, a disagreement was observed at high-p_{T}^{miss}, which was resolved by rejecting events with ≥ 1 jet with $p_{\text{T}} < 80$ GeV and $2.65 < |\eta| < 3.15$ in the 2017E and 2017F eras. The EE noise veto corrections are applied to relevant data from 2017.

5.4.3 HEM Veto (2018)

The HEM veto in the CMS experiment is a correction applied to the data in 2018 to account for an unpowered section of the hadronic endcap calorimeter (HEM). The HEM covers the region of $-3 \leq \eta \leq -1.65$ and $-1.57 \leq \phi \leq -0.87$ and became unpowered in 2018. This means that any jets that fall within this area in the simulation would not have been detected in the data. To account for this, any jets within this area of the detector in the simulation must be vetoed or excluded from the analysis to ensure accurate comparison with the data. The results of which can be seen in Figure 5.2.
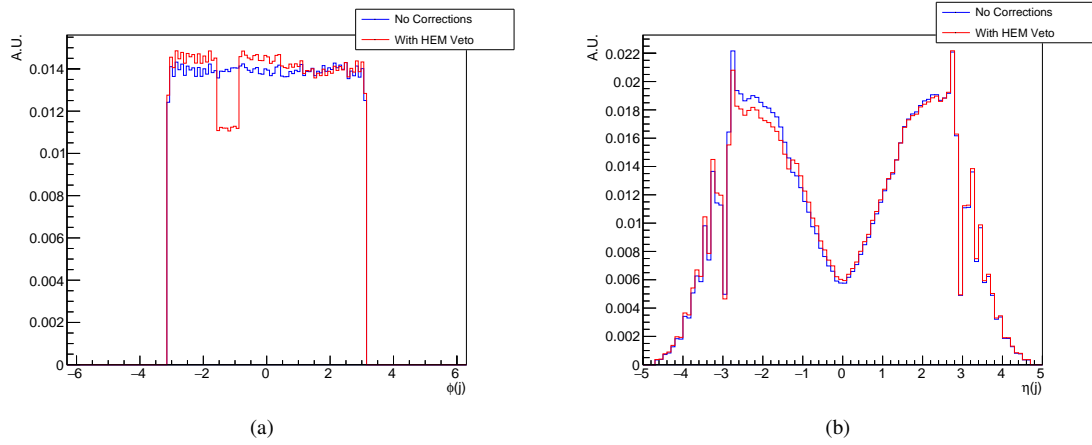
Figure 5.2: Overlaid plots showing before (blue) and after (red) the application of the HEM veto in ϕ (left) and η (right). All distributions have been normalized to unity in order to compare the overall shape.

CHAPTER 6

Event Selection & Signal Optimization

Events used in this search are selected using triggers that require the presence of at least one muon candidate with $p_T > 24$ GeV and $\eta < 2.1$. In addition to the requirements on muon trigger objects, kinematic requirements on p_T and η are imposed on the reconstructed muon candidates used in the signal region (SR) to achieve a trigger efficiency greater than 95%. Events are required to have exactly two muon candidates with $p_T > 10$ GeV, and the highest-p_T muon candidates is required to have $p_T > 30$ GeV. The $\mu\mu$ pair are required to be separated by $\Delta R > 0.4$. Since a primary focus is sensitivity to high-mass N_ℓ, signal events are typically characterized by one high-p_T muon ($p_T(\mu) \approx \frac{1}{3}m_{N_\ell}$). This is a motivating factor for the $p_T > 30$ GeV requirement on one of the muons, which helps to drastically reduce events from Z+jets and W+jets processes where $p_T(\mu) \approx \frac{1}{2}m_{Z,W} = 40, 50$ GeV. The p_T cut on the second muon is motivated by the experimental constraints of reconstruction algorithms at low p_T. Furthermore, in background events, the leptons mainly come from decays of W bosons (e.g. $t\bar{t} \rightarrow bbWW \rightarrow bb\ell\nu_\ell\ell\nu_\ell$) and Z/γ^* bosons (e.g. $Z \rightarrow \ell\ell$), and therefore both muons have similar p_T values. On the other hand, because the first muon in signal events is produced via non-resonant t-channel $W\gamma$ fusion diagrams (making it low-p_T) and the second muon is produced by the decay of a heavy on-mass shell N_ℓ (making it high-p_T), the p_T values of the two muons are largely asymmetric. This motivates a requirement on the absolute value of the scalar difference in p_T between the two muons. Therefore, we require $\Delta p_T > 30$ GeV.

Additionally, we impose a same-sign (SS) electric charge requirement for the two muons, $Q(\mu_1) \times Q(\mu_2) > 0$, in order to reduce the Z+jets background by two orders of magnitude. This requirement is about 50% efficient for signal events. Because the SS requirement is effective at reducing the Z+jets background contribution to the search region, a cut on the invariant mass of the pair is not needed.

In addition to the $\mu\mu$ selection, the final selection is defined by requiring at least two AK4 jets with $p_T > 30$ GeV and $|\eta| < 5.0$. Only jets separated from the μ candidates by $\Delta R > 0.4$ are considered. Depending on the heavy neutrino mass, signal events can have different jet multiplicities and kinematic properties. In the low m_{N_ℓ} mass region (i.e. less than approximately 750 GeV), the W boson has low enough momentum that its decay products are well separated in η-ϕ space. In the high m_{N_ℓ} scenarios, the W boson is significantly boosted such that the 2 quarks from the W decay merge to give rise to a "Fat Jet"(AK8 jet). The kinematics of the low- and high-mass signal regions (SRs) are different, and in both cases are considered separately to obtain the best sensitivity to a broad range of masses. The two SRs used in the analysis are defined as:

- low-mass SR: three or more AK4 jets(j) and exactly 0 AK8 jets(J)

- high-mass SR: two or more AK4 jets(j) and 1 or more AK8 jets(J)

The VBF signal topology is characterized by the presence of two jets in the forward direction, in opposite hemispheres, and with large dijet invariant mass. On the other hand, the jets in background events are mostly central and have small dijet invariant masses. Therefore, the "VBF selection" is imposed by requiring at least two of the aforementioned AK4 jets to have $|\Delta\eta_{jj}| > 4.2$, $\eta_1 \times \eta_2 < 0$, and $m_{jj} > 500$ GeV. This VBF selection is the same for both the high-mass and low-mass SRs. Table 2 lists the full set of signal selection criteria for both the low- and high-mass search regions. In the rare cases ($< 1\%$) where selected events contain more than one dijet candidate satisfying the VBF criteria, the VBF dijet candidate with the largest dijet mass is chosen since studies show it is 97% likely to result in the correct VBF dijet pair for signal events [51].

The signal selection efficiency for VBF μN_μ events depends on the N_μ mass value. The total signal selection efficiency is 4.2% for $m_{N_\mu} = 0.25$ TeV and 10.2% for $m_{N_\mu} = 2.0$ TeV. These efficiencies include the branching fraction of approximately 67% for W decaying to $q\bar{q}'$. Figures 6.1 show the predicted SM background (described in sections to follow) and expected signal, in bins of m_{jj}, for both the low- and high-mass SRs, respectively. The last bin in the mass plots represents the yield for $m_{jj} > 1.5$ TeV (i.e. these bins include the overflow). The predicted background yield is 17.01 ± 3.83 events, with $t\bar{t}$, single-top, and Z+jets composing about $\sim 76\%$, $\sim 13\%$, and $\sim 4\%$ of the total background for the high-mass SR. The respectively simulated distributions corresponding to signal hypotheses with m_{N_μ} values ranging from 0.25-1.5 TeV are also shown for comparison.

Selection	Cut
Trigger	HLT_IsoMu24
Sub-leading μ	$p_T \geq 10$ GeV, $|\eta| < 2.4$, "Tight ID"
Leading μ	$p_T \geq 30$ GeV, $|\eta| < 2.4$, "Tight ID"
$\mu\mu$ combination	$\Delta p_T \geq 30$ GeV, ΔR 0.3, same-sign $\mu\mu$
N(b-jets) = 0	$p_T \geq 20$ GeV, $|\eta| < 2.4$, deep CSV (Cat 1: 'Medium' WP)(Cat 2: 'Tight' WP)
low-mass SR (Cat 1):	
N(AK4 Jets) $\geq$ 3	$p_T \geq 30$ GeV, $|\eta| < 5.0$ Loose ID(2016)
N(AK8 Jets) = 0	$p_T \geq 180$ GeV, $|\eta| < 2.4$, Softdrop [64,105], tau21 < 0.55
high-mass SR (Cat 2):	
N(AK4 Jets) $\geq$ 2	$p_T \geq 30$ GeV, $|\eta| < 5.0$ Loose ID(2016)
N(AK8 Jets) $\geq$ 1	$p_T \geq 180$ GeV, $|\eta| < 2.4$, Softdrop [64,105], tau21 < 0.55
N(VBF jets pair) $\geq$ 1	$\Delta\eta \geq 4.2$, $\Delta R \geq 0.3$, $\eta_1 \times \eta_2 < 0$, $m(jj) > 500$

Table 6.1: Table showing the signal region selections. The selection includes trigger requirements, identification and momentum criteria for the leading and sub-leading muons, and b-tagging requirements for events with no b-tagged jets. The low-mass signal region (Cat 1) requires at least 3 AK4(j) jets with specific p_T and η requirements, and no AK8 jets with soft drop mass in the range of 64-105 GeV and tau21 ratio less than 0.55. The high-mass signal region (Cat 2) requires at least 2 AK4(j) jets and at least 1 AK8(J) jet with the same soft drop mass and tau21 ratio requirements. The deep CSV algorithm is used for b-tagging.

The Soft drop mass is a technique used in high-energy physics to improve the accuracy of jet mass measurements by removing soft radiation from the jet. This technique involves a specific clustering algorithm and iterative declustering process, resulting in a final jet mass known as the soft drop mass, which can be used to distinguish between different physics processes [62].

The Tau21 is defined as the ratio of the energy of particles in the jet that are clustered into two sub-jets (using a specific algorithm) to the total energy of the jet. Specifically, it is the ratio of the energy in the sub-jet with the second-smallest energy to the energy in the sub-jet with the smallest energy [63]. In general, signal jets tend to have a more isotropic energy distribution and are less likely to have a significant fraction of their energy contained in two or more sub-jets (resulting in a smaller value of Tau21), while background jets are more likely to have such a substructure (resulting in a larger value of Tau21). Tau21 is one of several variables used in the identification of jets that originate from the decay of heavy particles, such as W, Z, or Higgs bosons, and in the rejection of jets originating from QCD radiation or other backgrounds.

CSV (Combined Secondary Vertex) is a b-tagging algorithm used in CMS. B-tagging refers to the identification of jets that contain hadrons originating from b-quarks, which are produced in a variety of high-energy
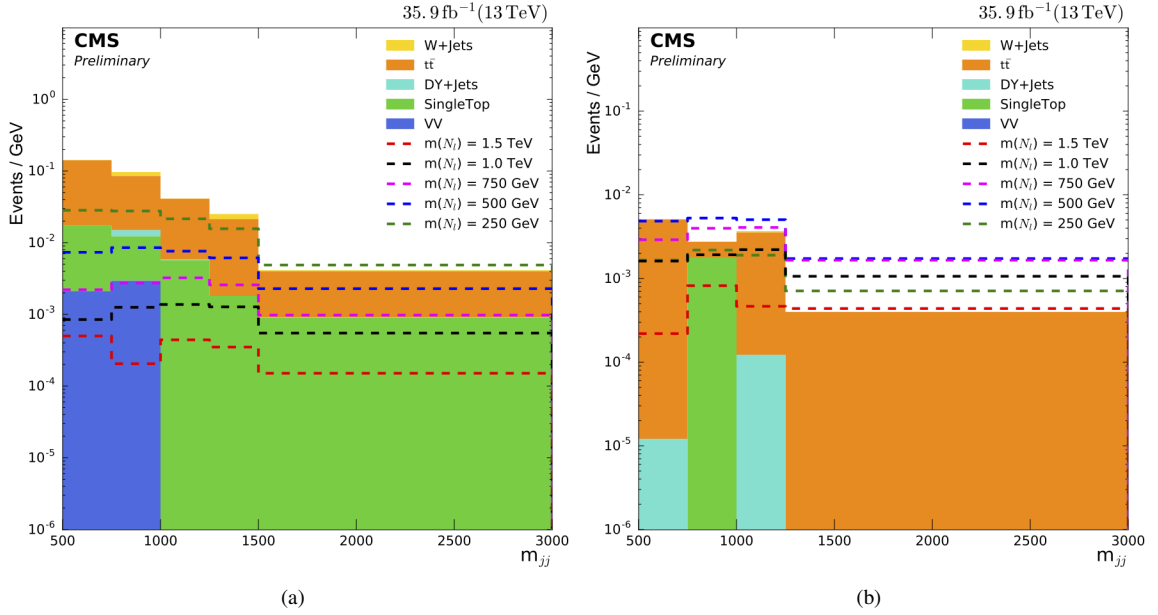
Figure 6.1: The m_{jj} distribution in the low-mass SR. The estimated backgrounds are stacked while the simulated signal distributions are overlaid. (left), and the m_{jj} distribution in the high-mass SR. The estimated backgrounds are stacked while the simulated signal distributions are overlaid. (right).

physics processes. The CSV algorithm utilizes a combination of variables related to the secondary vertices (where b-hadrons decay) and tracks associated with a jet to calculate a discriminant value, which is used to identify jets likely to contain b-hadrons [64].

b-tagging ID & Δp_T Optimization

We have performed a shape based analysis by using the m_{jj} distribution, as shown in Figure 6.1, as our fit variable and searching for a broad excess in the tails of the distribution. The SR selections are optimized to give the best discovery potential, which also gives the best upper 95% confidence level upper limit on the signal cross-section. The calculation of the expected upper limit on the signal cross-section is obtained by using the m_{jj} distribution to construct a combined profile likelihood ratio test statistic in bins of m_{jj} and using the asymptotic CL_s criterion [65]. Systematic uncertainties are taken into account as nuisance parameters, which are removed by profiling, assuming gamma function or log-normal priors for normalization parameters, and Gaussian priors for mass spectrum shape uncertainties. The combination of the low- and high-mass search channels requires simultaneous analysis of the data from the individual channels, accounting for all statistical and systematic uncertainties and their correlations. Correlations among backgrounds, both within a channel and across channels, are taken into consideration in the limit calculation. For example, the uncertainty in the integrated luminosity is treated as fully correlated across channels.

For an example study to illustrate the signal optimization, Figure 6.2 shows the upper limit on signal cross-section as a function of N_μ mass and b-tagging identification criteria (referred to as "working points" or WP). We considered the CSV loose, tight, and medium WPs, which provide different combinations of misidentification rates and bottom quark identification efficiencies. The conclusion shown in Figure 6.2 for the low-mass SR is that the medium b-tagging WP provides better sensitivity (i.e. lower upper limit on signal cross-section). On the other hand, the tight b-tagging WP provides better sensitivity for the high-mass SR. Although we studied the loose, medium, and tight WPs for both SR categories, Figure 6.2 only shows two WPs for each category. These conclusions were validated by comparing the change in the signal and background yields and computing the improvement in the signal significance $\frac{S}{\sqrt{S+B}}$.

A similar optimization process was used to determine how to take advantage of the boosted topology which results from the decay of a heavy N_ℓ (i.e. how to utilize AK8 jets(J) for the high-mass SR). This can be seen by comparing the (≥ 3 AK4, $= 0$ AK8) Medium WP contour with the (≥ 2 AK4, ≥ 1 AK8) Medium WP contour, which shows that the requirement of at least one AK8 jet(J) has a better expected upper limit for TeV scale N_μ masses, while the requirement of zero AK8 jets(J) has better performance for low mass N_μ.
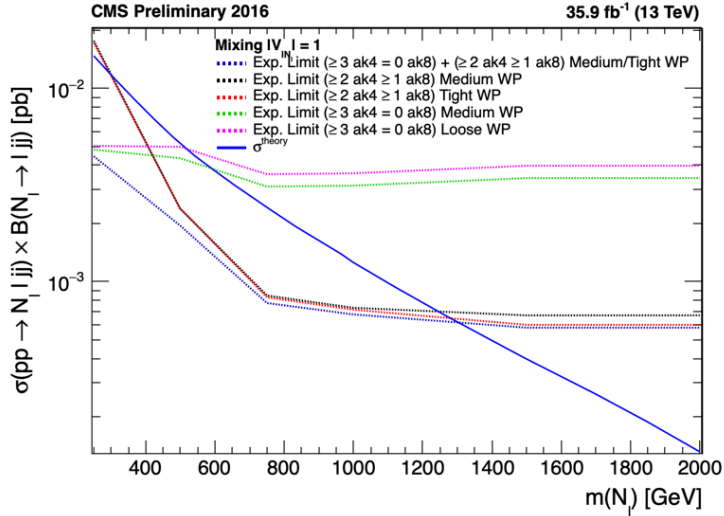
Figure 6.2: Optimization results for b-jet WPs for both signal region.

For some additional example optimization studies, Figures 6.3 show the expected upper limits on signal cross-section as a function of N_μ mass and the Δp_T cut threshold. As noted previously and implied in the Feynman diagram in Figure 6.4, the p_T values of the two muons in signal events are largely asymmetric, while the two muons in background events such as $t\bar{t}$ have Δp_T near zero. We varied the Δp_T cut values to see which gave the best sensitivity. Figures 6.3 shows that a $\Delta p_T > 30$ GeV requirement is optimal for both the low- and high-mass SR. A similar conclusion is obtained by using signal significance instead of the upper

limit on the signal cross-section, as indicated by Figure 6.4.

(a)   (b)
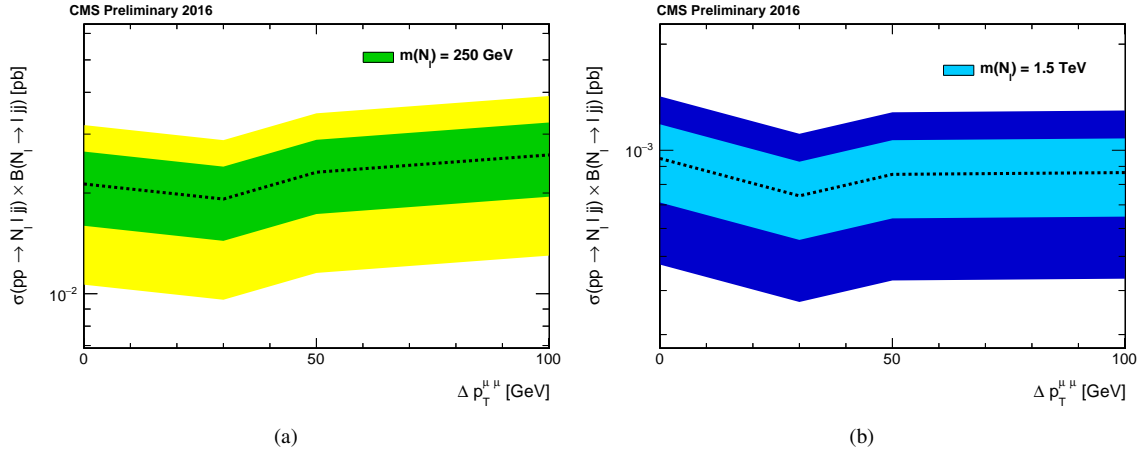
Figure 6.3: The expected upper limits on signal cross-section (and one standard deviation bands) as a function of the Δp_T cut threshold and for N_μ mass of 250 GeV(left) and 1.5 TeV(right).
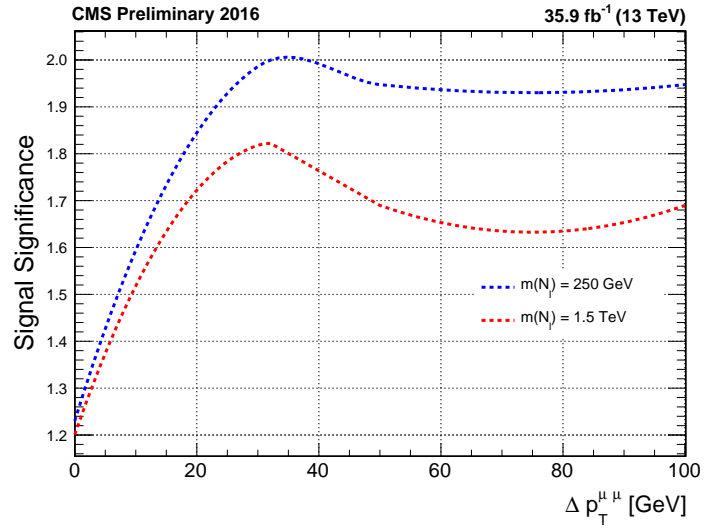
Figure 6.4: The signal significance as a function of the Δp_T cut threshold and for N_μ mass values of 0.25 and 1.5 TeV.

36

CHAPTER 7

Background Estimation

The background refers to any SM process that can produce the same final state particles and mimic their kinematics as the search channel. The goal is to select only signal events and reject all others, but this is not always achievable. Therefore, it is crucial to accurately model the background processes in simulation to ensure that any observed excess is a result of new physics and not poor modeling of known physics. This is accomplished through background estimation studies, which may require scale factors or shape-based corrections to correct for any mismodeling. These studies start by applying central selection cuts to understand the modeling of physics objects in the central region of the detector, followed by VBF selections for the forward region. The dominant background contribution, accounting for 80% of all cases, is the $t\bar{t}$ process for both low- and high-mass SRs. The $t\bar{t}$ background is estimated using a combination of data and simulation.
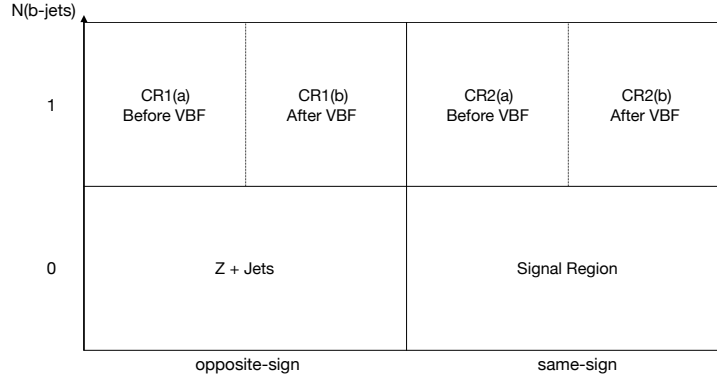
Figure 7.1: ABCD diagram with defined control regions of interest to esimate the background and understand the $t\bar{t}$ background using 2016 dataset. CR1(OS $\mu\mu$) measures a correction factor for the VBF efficiency. CR2(SS $\mu\mu$) measures a correction factor for the central selections.

The estimation of the $t\bar{t}$ background in the signal region is obtained by utilizing 4 control regions (see Figure 7.1) for each low- and high-mass signal region. Since $t\bar{t}$ production is characterized by the production of bottom quarks, the $t\bar{t}$ CRs are obtained with similar selections to the SR, except requiring one jet tagged as a b quark jet. The requirement of a b-jet significantly increases the $t\bar{t}$ purity of the control samples while ensuring that those control samples have negligible signal contamination and similar kinematics and composition of misidentified leptons as the SR. Various control samples obtained with the requirement of exactly two jets are also utilized to validate the performance of the background estimation method. By utilizing 2

b-jets instead of 1 b-jet, the contribution from QCD multijet events is reduced by about a factor of the $j \rightarrow b$ misidentification rate, without affecting the lepton kinematics, the composition of the events, and the m_{jj} shapes.

Initially, it was considered to look only at CR2 (i.e. same-sign $\mu\mu$ events with 1 b-tagged jet) to measure the data-to-simulation correction factor for the VBF efficiency, but CR2 has limited statistics and thus would result in a large statistical uncertainty on the scale factor. To minimize the uncertainty on the scale factor, we take advantage of opposite-sign CRs (CR1 in Figure 7.1) the background is large due to the contribution from $DY(Z \rightarrow \mu\mu)$. The opposite-sign requirement increases the statistics in the control sample by a factor of $f_{j \rightarrow \mu}^{-1}$, where $f_{j \rightarrow \mu}^{-1}$ represents the probability for a jet to be misidentified as a muon. When considering CR1 in estimating the $t\bar{t}$ background, we first check that the VBF jet kinematics in CR2 are the same as in CR1. In other words, the opposite-sign requirement does not bias the VBF selection efficiency. The VBF efficiency in the data ($\varepsilon_{\text{VBF}}^{\text{data}}$) and simulated samples ($\varepsilon_{\text{VBF}}^{\text{MC}}$) is calculated by taking the yield of events that pass the VBF requirements (i.e yield in CR1b minus non-$t\bar{t}$ background yield from MC) and dividing by the yield before the VBF cuts (i.e. yield in CR1a minus non-$t\bar{t}$ background yield from MC). The correction factor for the VBF efficiency is then calculated as

$$SF_{VBF_{eff}} = \varepsilon_{\text{VBF}}^{\text{data}} / \varepsilon_{\text{VBF}}^{\text{MC}} \tag{7.1}$$

We use samples of events obtained with a requirement of exactly 2 b-jets to validate the measured $SF_{VBF_{eff}}$ from CR1. The $t\bar{t}$ background yields in the SR are evaluated using the following equation:

$$N_{SR}^{SS} = N_{SR,SS}^{t\bar{t}} \times SF_{VBF_{eff}} \times SF_{SS,CR2(a)} \tag{7.2}$$

where N_{SR}^{SS} is the predicted $t\bar{t}$ background yield in the SR, $N_{SR,SS}^{t\bar{t}}$ is the $t\bar{t}$ rate predicted by simulation for the SR selection, $SF_{VBF_{eff}}$ the data-over-simulation correction factor for the VBF selection efficiency, and $SF_{SS,CR2(a)}$, which is dominated by background, is the data-over-simulation correction factor for the non-VBF requirements of the SR. The numerator in the calculation of each correction factor is estimated by subtracting from data the contribution from other background events different from that under study, and the statistical uncertainty is propagated to the scale factor uncertainty.

The data and simulated background distributions in CR1 of the low-mass search are shown in Figure 7.2. The selections used to define CR1 of the low-mass search are outlined in Table 7.1. Table 7.2 lists the number of observed events in data as well as the simulated background contributions in low-mass CR1. The uncertainties include only the statistical components. The measured VBF efficiency is 12.9% $\pm$ 2.3% and

Selection	Cut
Trigger	HLT_IsoMu24
Sub-leading μ	$p_T \geq 10$ GeV, $|\eta| < 2.4$, "Tight ID"
Leading μ	$p_T \geq 30$ GeV, $|\eta| < 2.4$, "Tight ID"
$\mu\mu$ combination	**opposite-sign** $\mu\mu$ $\Delta p_T \geq 30$ GeV, $\Delta R \leq 0.3$
N(b-jets) = 1	$p_T \geq 20$ GeV, $|\eta| < 2.4$, deep CSV (Cat 1: 'Medium' WP)(Cat 2: 'Tight' WP)
N(AK4 Jets) ≥ 3	$p_T \geq 30$ GeV, $|\eta| < 5.0$ Loose ID(2016)
N(AK8 Jets) = 0	$p_T \geq 180$ GeV, $|\eta| < 2.4$, Softdrop [64,105], tau21 ≤ 0.55
N(VBF jets pair) ≥ 1	$\Delta\eta \geq 4.2$, $\Delta R \geq 0.3$, $\eta_1 \times \eta_2 < 0$, $m_{jj} > 500$

Table 7.1: Central selection event criteria for CR: OS 0J + $\geq$ 3j (2016 Dataset)

14.0% $\pm$ 2.8% in data and MC, and this results in SF_{VBFeff} of 0.92 $\pm$ 0.03. Figure 7.2 shows good agreement between the data and MC distributions after applying this correction factor.

Process	Yield(before VBF)	Yield(After VBF)
Data	201	26
W + jets	5.98 ± 2.53	0.12 ± 0.06
DY + jets	3.36 ± 2.76	0.01 ± 0.01
Diboson	2.52 ± 0.33	0.37 ± 0.09
$t\bar{t}$	132.91 ± 7.20	18.80 ± 2.60
Single Top	7.27 ± 1.08	2.00 ± 0.48
Total BG	152.05 ± 20.14	21.30 ± 3.24

Efficiency-Data	12.9% $\pm$ 2.4%
Efficiency-MC	14.0% $\pm$ 2.8%
$SF_{VBF_{eff}}$	0.92 $\pm$ 0.05

Table 7.2: Summary of SFs for the CR1 for 2016 dataset, including central selections. Number of observed events and corresponding background predictions is shown. The uncertainties include the statistical component.

The data and simulated background distributions in CR1 of the high-mass search are shown in Figure 7.4. The selections used to define CR1 of the high-mass search are outlined in Table 7.3. Table 7.4 lists the number of observed events in data as well as the simulated background contributions in high-mass CR1. The uncertainties include only the statistical components. The measured VBF efficiency is 15.0% $\pm$ 0.3% and 13.0% $\pm$ 0.3% in data and MC, and respectively this results in SF_{VBFeff} of 1.09 $\pm$ 0.05. Figure 7.4 shows good agreement between the data and MC distributions after applying this correction factor.
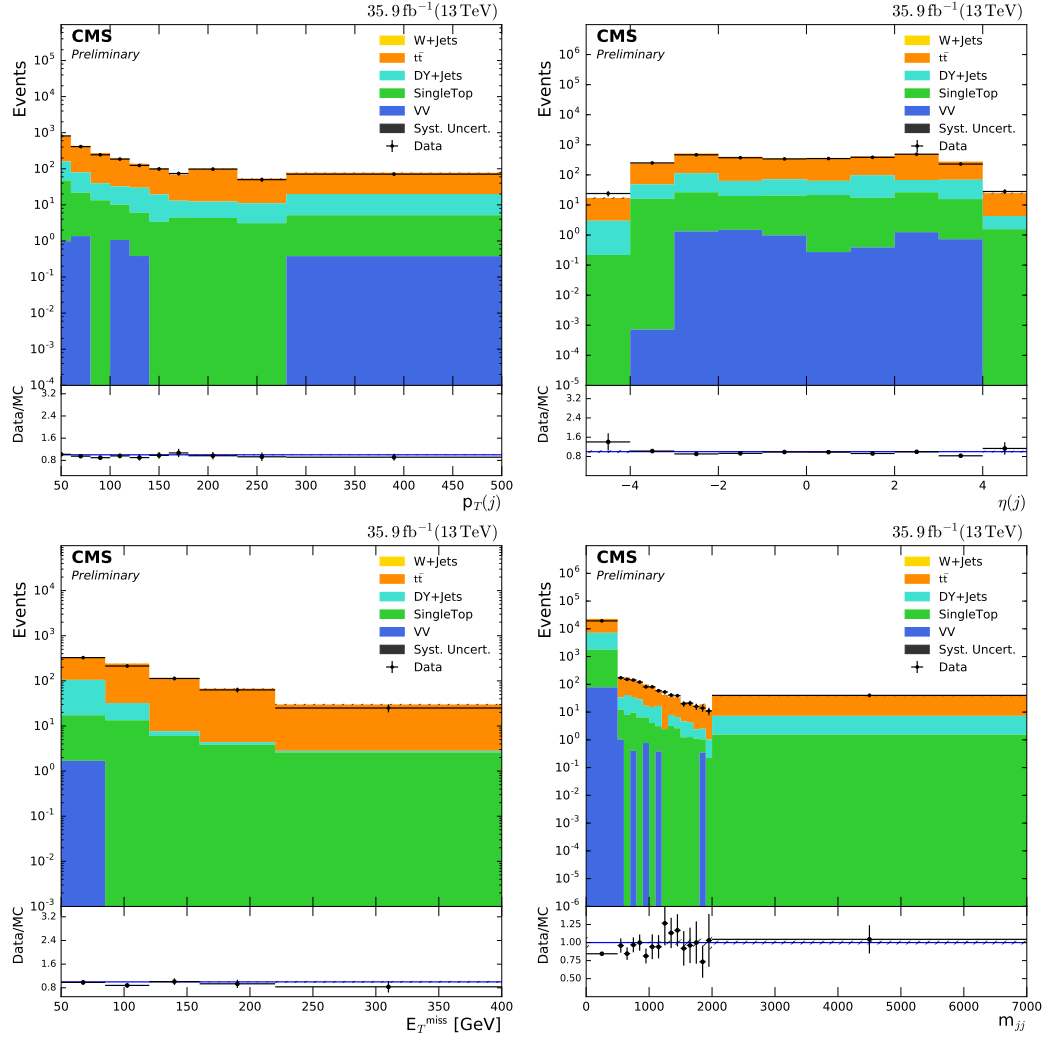
Figure 7.2: This figure shows the distributions of several kinematic variables, including $p_T(j)$, $\eta(j)$, $m(jj)$, and $E^{miss}T$, for the low-mass CR1 dataset. The plots have been scaled by a factor of 0.92. The shapes of these distributions are compared between data and the Monte Carlo (MC) simulation, and show good agreement. Specifically, the shapes of the distributions are similar between the two samples, indicating that the MC simulation is accurately modeling the data. The $pT(j)$ variable represents the transverse momentum of a jet, while $\eta(j)$ represents its pseudorapidity. $m(jj)$ represents the invariant mass of a pair of jets, and E_T^{miss} represents the missing transverse energy in the event.

40
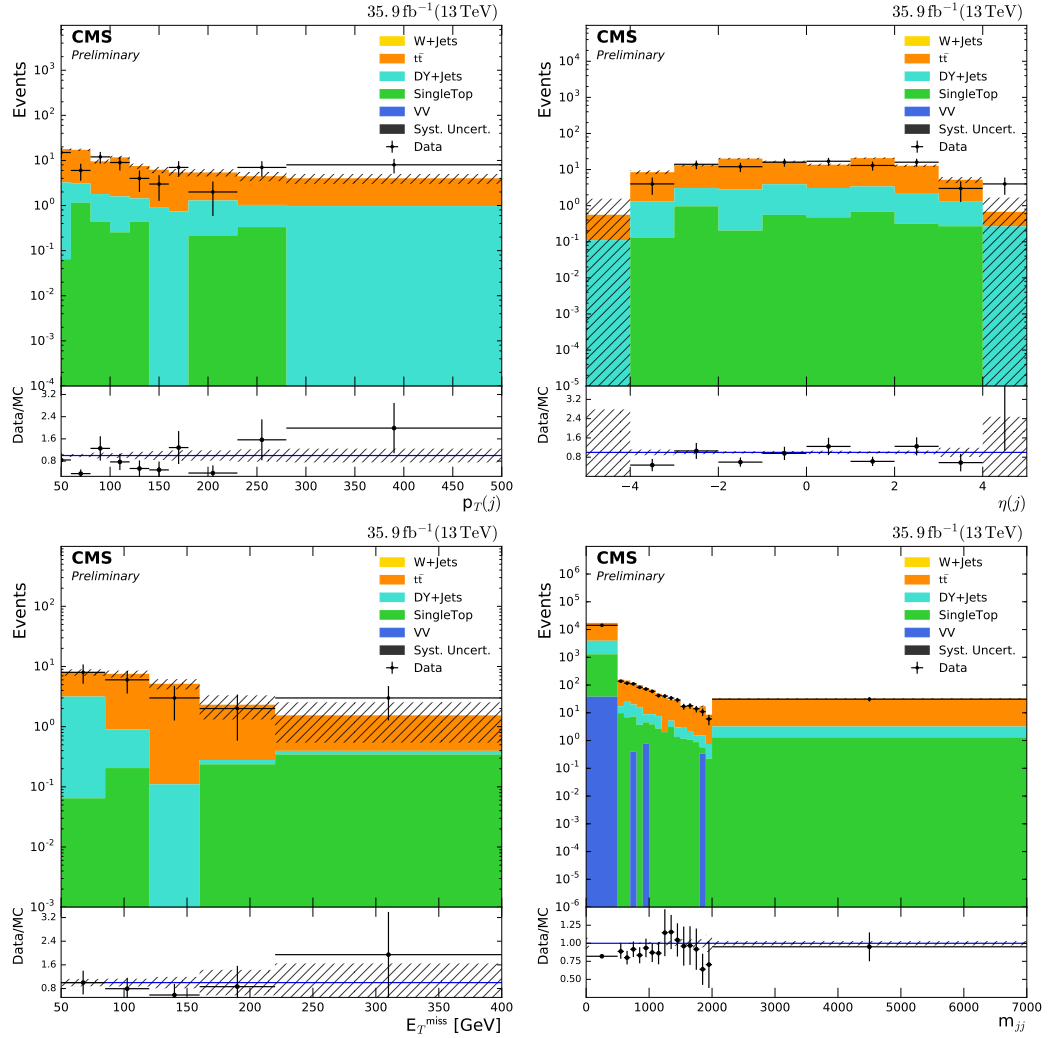
Figure 7.3: $p_T(j)$, $\eta(j)$, $m(jj)$, and E_T^{miss} distributions for high-mass CR1. A scale factor of 1.09 has been applied to these plots. The shapes of the kinematic distributions show good agreement between data and the MC expectation.

Selection	Cut
Trigger	HLT_IsoMu24
Sub-leading μ	$p_T \geq 10$ GeV, $|\eta| < 2.4$, "Tight ID"
Leading μ	$p_T \geq 30$ GeV, $|\eta| < 2.4$, "Tight ID"
$\mu\mu$ combination	**opposite-sign** $\mu\mu$ $\Delta p_T \geq 30$ GeV, $\Delta R \leq 0.3$
N(b-jets) = 1	$p_T \geq 20$ GeV, $|\eta| < 2.4$, deep CSV (Cat 1: 'Medium' WP)(Cat 2: 'Tight' WP)
N(AK4 Jets) ≥ 2	$p_T \geq 30$ GeV, $|\eta| < 5.0$ Loose ID(2016)
N(AK8 Jets) ≥ 1	$p_T \geq 180$ GeV, $|\eta| < 2.4$, Softdrop [64,105], tau21 ≤ 0.55
N(VBF jets pair) ≥ 1	$\Delta\eta \geq 4.2$, $\Delta R \geq 0.3$, $\eta_1 \times \eta_2 < 0$, $m_{jj} > 500$

Table 7.3: Central selection event criteria for CR: OS 1J + $\geq$ 2j (2016 Dataset)

Process	Yield(before VBF)	Yield(After VBF)
Data	15030	22
W + jets	3.01 ± 1.05	0 ± 0
DY + jets	2732.08 ± 72.45	4.00 ± 0.48
Diboson	39.59 ± 3.81	0 ± 0
$t\bar{t}$	14171.7 ± 76.69	19.72 ± 2.89
Single Top	1296.83 ± 16.71	0.852 ± 0.41
Total BG	18243.2 ± 170.71	24.57 ± 7.02

Efficiency-Data	$15.0\% \pm 0.3\%$
Efficiency-MC	$13.0\% \pm 0.3\%$
$SF_{VBF_{eff}}$	0.92 ± 0.006

Table 7.4: Summary of event yields in the CR2(a) and CR2(b) regions for 2016 for before and after applying central and VBF selections. Uncertainties included in this table are only statistical.

The previously described control regions are used to obtain corrections factors for the modeling of the VBF efficiency. In what follows I describe the control regions used to measure correction factors for the modeling of the non-VBF selection (i.e. the same-sign dimuon cuts). This next set of control regions (low- and high-mass regions), are defined by the selections shown in Tables 7.5 and 7.7. The difference here compared to the previous control regions is the requirement of a same-sign dimuon pair. The data and simulated background yields in CR2 are shown in Table 7.6. The SFs are measured by taking data, subtracting non $t\bar{t}$ contributions using simulation, and dividing by the $t\bar{t}$ prediction in simulation. This results in a measured $SF_{SS,CR2a}$ of 1.29 ± 0.10 for low-mass CR2. The data and background distributions for low-mass CR2 are shown in Figure 7.4. The $t\bar{t}$ distributions have been corrected with the measured scale factor. There is good agreement between the data and MC distributions in Figure 7.4.

Figure 7.4: $p_T(j\mu)$, $\eta(\mu)$, $m(\mu\mu)$, and E_T^{miss} distributions for low-mass CR2. A scale factor of 1.29 has been applied to these plots. The shapes of the kinematic distributions show good agreement between data and the MC expectation.

Selection	Cut
Trigger	HLT_IsoMu24
Sub-leading μ	$p_T \geq 10$ GeV, $|\eta| < 2.4$, "Tight ID"
Leading μ	$p_T \geq 30$ GeV, $|\eta| < 2.4$, "Tight ID"
$\mu\mu$ combination	**same-sign** $\mu\mu$ $\Delta p_T \geq 30$ GeV, $\Delta R \leq 0.3$
N(b-jets) = 1	$p_T \geq 20$ GeV, $|\eta| < 2.4$, deep CSV (Cat 1: 'Medium' WP)(Cat 2: 'Tight' WP)
N(AK4 Jets) ≥ 3	$p_T \geq 30$ GeV, $|\eta| < 5.0$ Loose ID(2016)
N(AK8 Jets) = 0	$p_T \geq 180$ GeV, $|\eta| < 2.4$, Softdrop [64,105], tau21 ≤ 0.55
N(VBF jets pair) ≥ 1	$\Delta\eta \geq 4.2$, $\Delta R \geq 0.3$, $\eta_1 \times \eta_2 < 0$, $m_{jj} > 500$

Table 7.5: Central selection event criteria for CR: SS 0J + $\geq$ 3j (2016 Dataset)

Process	Yield
Data	336
W + jets	7.70 ± 2.92
DY + jets	9.02 ± 4.84
Diboson	5.11 ± 0.43
$t\bar{t}$	228.58 ± 9.46
Single Top	20.28 ± 1.71
Total BG	270.70
SF$_{BG}$	1.29 ± 0.10

Table 7.6: Summary of SFs for the CR1 for 2016 dataset, including central selections. Number of observed events and corresponding background predictions for low-mass CR1.

The data and background distributions/yields in high-mass CR2 (see Table 7.7 for the full list of selections) are shown in Figure 7.5. The uncertainties shown are only statistical components. The data yields are in good agreement with the estimated background after correcting the $t\bar{t}$ prediction with the measured scale factor of $SF_{SS,CR2a} = 1.29 \pm 0.10$. The shapes of the kinematic distributions also show good agreement between data and the MC expectation.

Finally, in addition to $t\bar{t}$ there are smaller background contributions from other SM processes such as the production of a single top quark, dibosons, and vector bosons with associated jets (i.e. Z/W+jets). The Z/W+jets backgrounds represent only 6.1% of the total background since they consist of mostly of an opposite-sign dimuon pair (i.e. $Z \to \mu^+\mu^-$) or a single promptly produced muon (i.e. $W \to \mu\nu_\mu$). These two backgrounds are estimated by scaling the yields in the MC simulated samples by the best available cross-sections, and by the integrated luminosity of the data samples. Appropriate known correction factors are also applied as described in previous sections in order to match the performance of various reconstruction

and identification algorithms observed in data. For example, the simulated events are reweighted to match the pileup distribution observed in data. Weights are also applied such that the muon, jet, and b-tagging identification efficiencies also match those observed in data. The process described above is generally the approach taken if the background contribution from a particular process is expected to be small. In certain cases, dedicated validation samples are utilized to confirm the corrected yields from simulation. For example, the modeling of Z+jets background from simulation (after the standard corrections described above) is validated by the agreement in the opposite-sign dimuon mass distribution near the Z-mass peak (90 GeV), and with otherwise similar sections to the SR.

Selection	Cut
Trigger	HLT_IsoMu24
Sub-leading μ	$p_T \geq 10$ GeV, $|\eta| < 2.4$, "Tight ID"
Leading μ	$p_T \geq 30$ GeV, $|\eta| < 2.4$, "Tight ID"
$\mu\mu$ combination	**same-sign** $\mu\mu$ $\Delta p_T \geq 30$ GeV, $\Delta R \leq 0.3$
N(b-jets) = 1	$p_T \geq 20$ GeV, $|\eta| < 2.4$, deep CSV (Cat 1: 'Medium' WP)(Cat 2: 'Tight' WP)
N(AK4 Jets) ≥ 2	$p_T \geq 30$ GeV, $|\eta| < 5.0$ Loose ID(2016)
N(AK8 Jets) ≥ 1	$p_T \geq 180$ GeV, $|\eta| < 2.4$, Softdrop [64,105], tau21 ≤ 0.55
N(VBF jets pair) ≥ 1	$\Delta\eta \geq 4.2$, $\Delta R \geq 0.3$, $\eta_1 \times \eta_2 < 0$, $m_{jj} > 500$

Table 7.7: Central selection event criteria for CR: SS 1J + $\geq$ 2j (2016 Dataset)
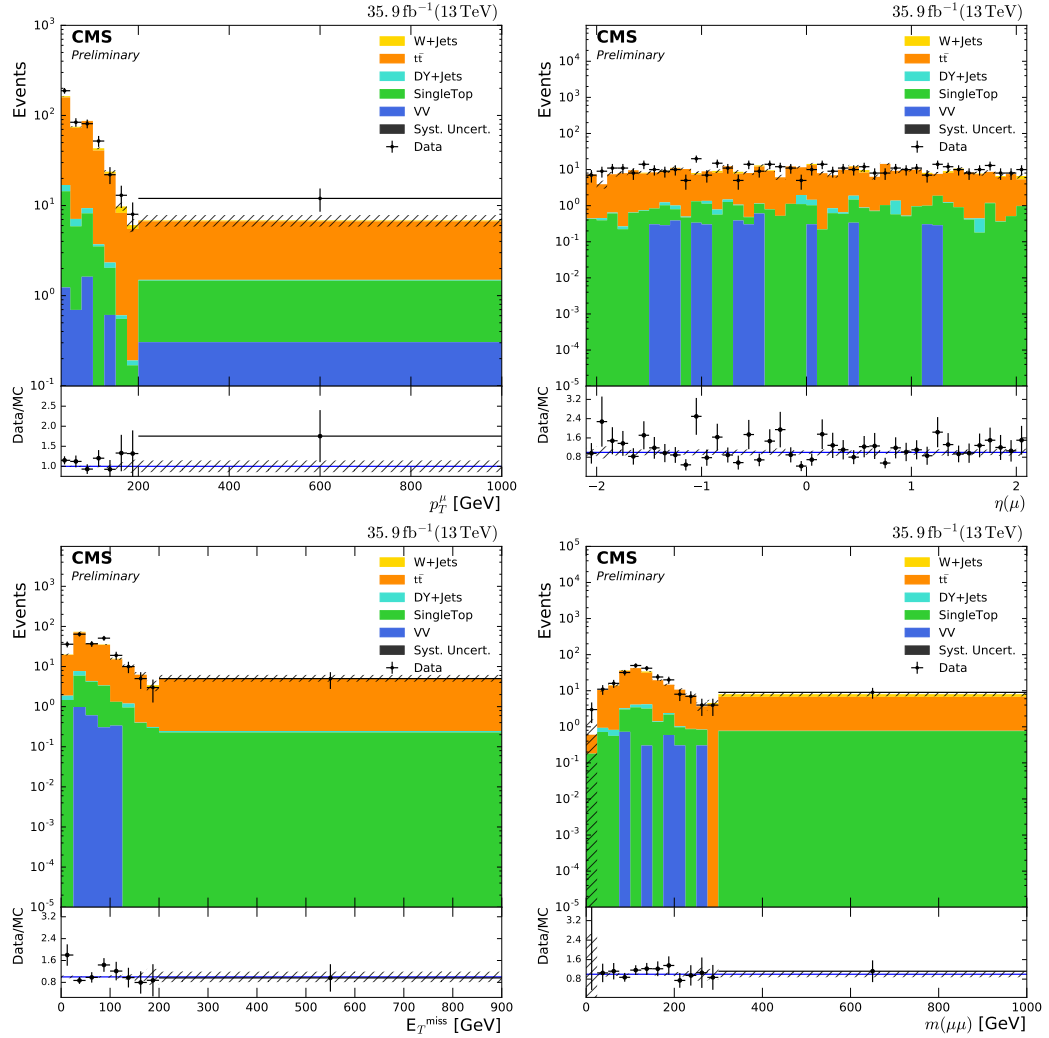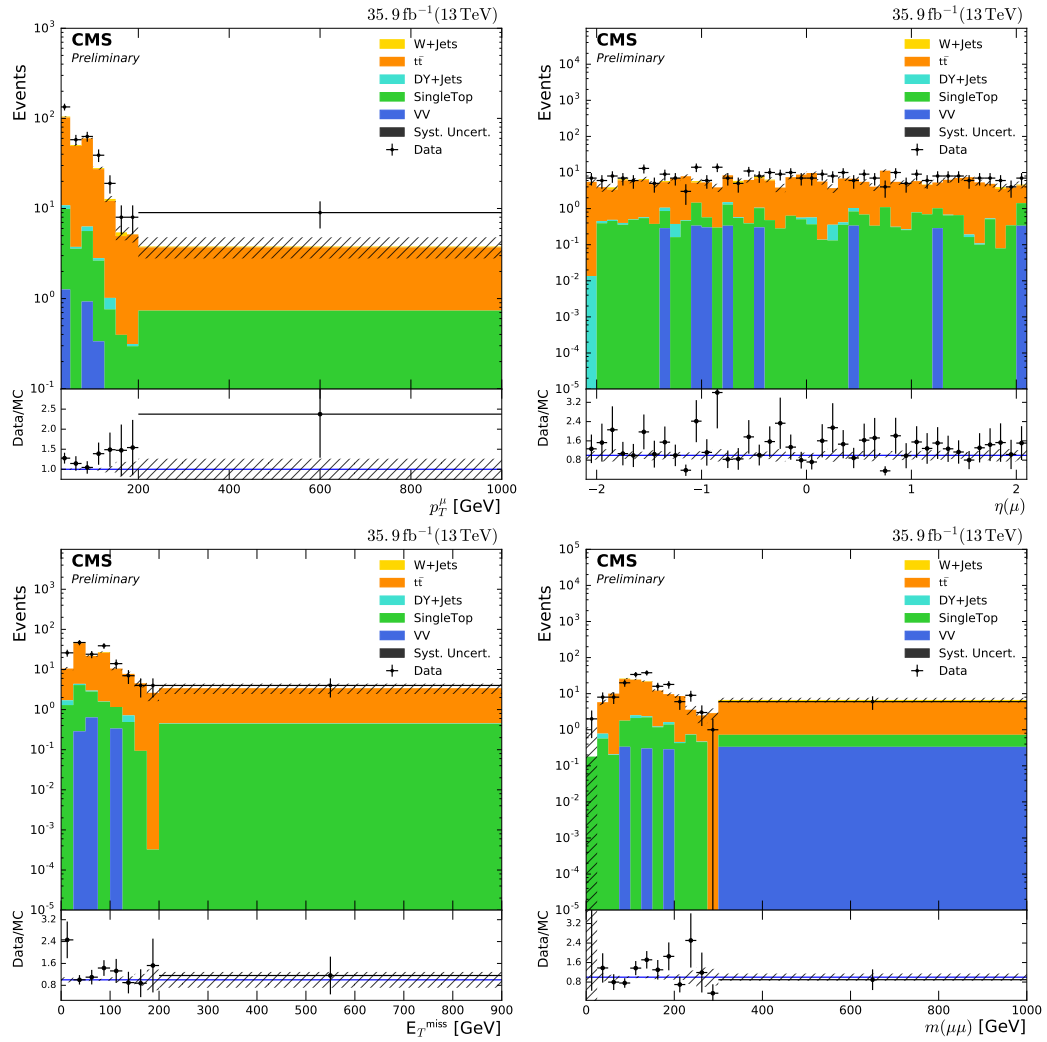
Figure 7.5: $p_T(j\mu)$, $\eta(\mu)$, $m(\mu\mu)$, and E_T^{miss} distributions for high-mass CR2. A scale factor of 1.29 has been applied to these plots.

CHAPTER 8

Systematic Uncertainties

Uncertainty in a data analysis refers to the degree of variation or spread in a measurement around a given value. It can be caused by either natural fluctuations or errors originating from the measuring equipment. There are two main types of uncertainties in data analysis: statistical and systematic.

Statistical uncertainties are those that arise from fluctuations in the data, either in actual recorded data or in simulated Monte Carlo data. They can be characterized by the standard deviation of a Poisson distribution, which is proportional to the square root of the number of events. For example, if an experiment is expected to observe 9 events, the standard deviation would be ± 3 events and the relative statistical uncertainty would be $1/\sqrt{N} = 1/\sqrt{9} = 33.3\%$. If the number of expected events were to increase to 100, the standard deviation would be ± 10 events and the relative statistical uncertainty would be $1/\sqrt{100} = 10\%$. Hence, it can be seen that statistical uncertainties can be reduced by increasing the overall statistics of the experiment through the collection of more data.

Systematic uncertainties, on the other hand, are uncertainties that arise from mismeasurements and errors within the detector. Systematic uncertainties cannot be expected to follow a specific distribution like statistical uncertainties and must be inferred using other methods, such as cross-checks with other detectors, or by varying the simulation parameters and studying the impact on the results. Systematic uncertainties are often larger than statistical uncertainties and can have a significant impact on the results of an experiment. It is important to carefully quantify and understand systematic uncertainties in order to improve the precision and accuracy of the data analysis.

We consider a 1.2% uncertainty for 2016, a 2.3% uncertainty for 2017, and a 2.5% uncertainty for 2018 Luminosity [66, 67, 68]. Uncertainties within years are considered 100% correlated across MC samples and channels within a given year. We also consider the effect on the signal acceptance efficiency by using a 1% momentum scale uncertainty on the muon momentum. The resultant systematic uncertainty on signal and MC based backgrounds is <1%.

Additional studies can be conducted to identify the systematic errors related to other experimental quantities which are not presented here. These studies specifically target high-p_T and far-forward jets in VBF interactions, such as pile-up (PU), jet energy resolution (JER), jet energy scale (JES). These quantities are anticipated to have the greatest impact on the study but have not been thoroughly examined in this analysis.

The total number of PU interactions per bunch crossing can be estimated by calculating the cross section, commonly known as the "minimum-bias" cross section. This measurement aims to minimize the amount of

biasing to ensure the most accurate result possible. The total number of PU interactions per bunch crossing can be calculated using the following formula:

$$\mu = L_{\text{inst}}^{i} \sigma_{\text{inel}} / f_{\text{rev}} \qquad (8.1)$$

The value of μ, representing the total number of interactions between protons, can be determined using the instantaneous luminosity of a single bunch, L_{inst}^{i}, the total cross section for inelastic proton-proton collisions, σ_{inel}, and the LHC orbit frequency, f_{rev}, which is needed to convert the instantaneous luminosity from a per-time measurement to a per-collision measurement. The resulting value of μ from the instantaneous luminosity is an average over a single lumi section, and the distribution of individual events will follow a Poisson distribution centered around this average value. In Run 2 of the CMS experiment, a recommended minimum-bias cross section of 69.2 mb is used, with an associated uncertainty of 4.6%, as reported by the CMS lumi group. To study this further, one can calculate the PU weights using the nominal minimum-bias cross section, as well as the $+1\sigma$ and -1σ variations, and then apply these newly derived PU weights to determine the resulting yields.

The JER and JES are important systematic uncertainties that affect the measurement of jets in high-energy physics analyses. The JER is a measure of the precision with which the energy of a reconstructed jet is measured. The JER is usually expressed as a function of the jet transverse momentum (p_T) and is typically defined as the ratio of the width of the energy response distribution to the mean value, where the energy response is the ratio of the reconstructed jet energy to the true jet energy. The JER can be affected by a variety of factors, such as detector effects, the event pile-up, and the jet reconstruction algorithm. To estimate the JER, the CMS collaboration uses a combination of simulation and data-driven techniques, such as the so-called "smearing method", which adjusts the reconstructed jet energy to match the true jet energy distribution.

JES is a systematic uncertainty that arises from the calibration of the energy measurement of the reconstructed jets. The JES can affect the overall jet energy scale and the shape of the jet energy response as a function of p_T, η, and other variables. One can estimate the JES by the using the combination of simulation and data-driven techniques, such as the "in situ" method, which uses the energy balance of events with two or more well-measured jets to calibrate the jet energy scale. Other techniques used to estimate the JES include the so-called "relative" JES method, which uses the relative differences in jet energy measurements between data and simulation, and the "absolute" JES method, which uses the energy scale of specific physics processes, such as Z boson decays or photon + jet events, to calibrate the jet energy scale.

In summary, the jet energy resolution (JER) and jet energy scale (JES) are important systematic un-

certainties that affect the precision and accuracy of jet measurements and has not yet been perform in this analysis. The CMS JetMET group employs a variety of simulation and data-driven techniques to estimate these uncertainties and to calibrate the jet energy scale [43].

CHAPTER 9

Results

This dissertation presents the analysis of proton-proton collisions, collected by the CMS experiment in 2016 at $\sqrt{s} = 13$ TeV. A search for new particles has been conducted using events from vector boson fusion (VBF) processes and resulting in a final state with two muon particles with same-sign electric charge and at least three jets, two of which satisfy the VBF topology. The results are interpreted in the minimal Type-1 seesaw mechanism, that involves heavy right-handed neutrinos [51]. The VBF topology requires two well-separated jets that appear in opposite hemispheres of the CMS detector, with large dijet invariant mass m_{jj}. The search utilizes events in two different search channels depending on the presence of a boosted pair of merged jets (collectively called a *fatjet*), in order to maximize discovery potential to low- and high-mass N_ℓ. The main result of this thesis is that searching for heavier neutrino states (predicted by extensions of the SM and inferred by the observation of neutrino oscillations) produced through VBF processes, is a key methodology to discovery N_ℓ particles at the LHC. While only presenting the analysis using the collected data in 2016, future aims are to present this analysis using the collected dataset of 2016-2018 (Full run II dataset). Figure 9.1 shows example SRs presented for the collected proton-proton collisions, collected by the CMS experiment in 2016.
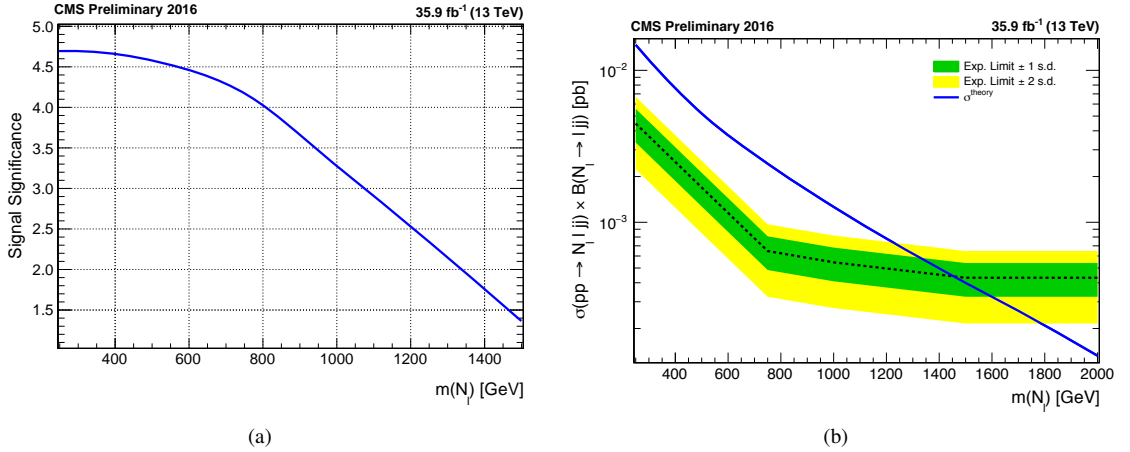
Figure 9.1: (a) Signal significance for the final 2016 dataset (b) Final 2016 combined exclusion of $m(N_\ell) <$ 1.46 TeV, assuming mixing $|V_{lN_\ell}| = 1$. The dashed black curve is the expected upper limit, with one and two standard-deviation bands shown in dark green and light yellow, respectively.

9.1 Conclusions & Future Outlook

The focus of this thesis has been on the analysis of the pp collision data collected in 2016, corresponding to an integrated luminosity of 35.9 fb^{-1}. However, the CMS Collaboration has now re-processed the pp collision data collected in 2017-2018, which increases the total integrated luminosity to 137.1 fb^{-1}. Therefore, the future aim of this analysis is to include the additional 100 fb^{-1} of pp data. This work is already ongoing, and Figure 9.2 shows the low-mass and high-mass SR m_{jj} distributions for the 2017 data (41.5 fb^{-1}). Additionally, the analysis with the full LHC Run II data will be expanded to include the dielectron channels, which target N_e production via VBF. As can be seen from Figure 9.3, the full Run II analysis is expected to achieve a discovery reach with signal significances greater than 5σ (3σ) for N_μ masses up to 1.19 (1.45) TeV.
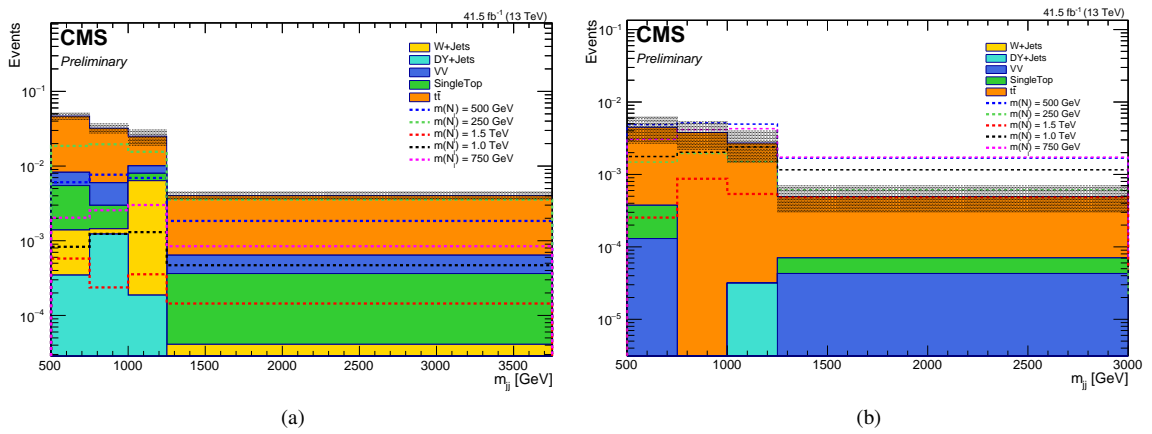
Figure 9.2: A dijet mass distributions for both signal regions for 2017 dataset.
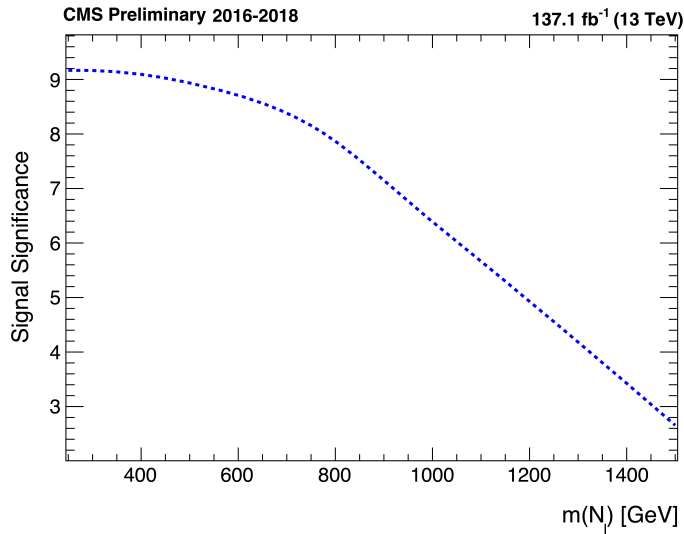
Figure 9.3: Signal significance for final 2016-2018 dataset.

The results of the search for heavy Majorana neutrinos in same-sign dimuon and jet final states in proton-proton collisions at $\sqrt{s} = 13$ TeV, using the dataset corresponding to an integrated luminosity of 35.9 fb^{-1}, are very promising. Although the un-blinded data of events are not shown, the expected standard model background prediction has been compared with the analyzed data. The analysis aims to enhance the sensitivity of previous searches by utilizing events from VBF processes. The results show that the sensitivity of the analysis to heavy neutrino masses extends up to 1500 GeV. The predicted exclusion limit can be extended to a mass of approximately above 2000 GeV by simply scaling the full run II data set.

This analysis stands out due to its significantly greater search sensitivity than previous studies, such as the one mentioned in reference [26]. It has the potential to detect interesting new physics within the search region defined by the expected exclusion, even if no new physics is identified. Furthermore, if no new physics is detected, the analysis can still provide valuable constraints on the left-right symmetric extension of the Standard Model. These constraints can inform and guide future studies in the field, making this analysis particularly unique and valuable.

A.1 Mathematical Treatment of the Lorentz Group and the Poincaré Group

The Lorentz group is defined as the group of transformations which leave the scalar product of Minkowski spacetime invariant [69]. These consist of spacetime rotations and boosts. Let us denote the generator of rotations as J_i and the generator of boosts as K_i. A general Lorentz transformation can then be written as:

$$\Lambda = e^{i\vec{J}\cdot\vec{\theta} + i\vec{K}\cdot\vec{\Phi}} \tag{A.1}$$

The corresponding Lie algebra of these generators is as follows, where $[,]$ denotes the commutator of two objects given by $[x,y] = xy - yx$, and ε_{ijk} is the Levi-Civita symbol:

$$[J_i, J_j] = i\varepsilon_{ijk}J_k$$
$$[J_i, K_j] = i\varepsilon_{ijk}K_k \tag{A.2}$$
$$[K_i, K_j] = -i\varepsilon_{ijk}J_k$$

We see that the rotation generators J_i are closed under commutation, meaning the commutator of two rotation generators returns another rotation generator. The boost generators however are not closed under commutation. If we instead define a new set of generators in the following way:

$$N_i^+ = \frac{1}{2}(J_i + iK_i)$$
$$N_i^- = \frac{1}{2}(J_i - iK_i) \tag{A.3}$$

then we get the following new commutation relations:

$$[N_i^+, N_j^+] = i\varepsilon_{ijk}N_k^+$$
$$[N_i^-, N_j^-] = i\varepsilon_{ijk}N_k^- \tag{A.4}$$
$$[N_i^+, N_j^-] = 0$$

We see that these two new generators close under commutation and that they both obey the Lie algebra of $SU(2)$. Indeed, we have just demonstrated that the Lorentz group can be decomposed into two independent copies of $SU(2)$. This decomposition allows us to further label particles according to their representations

under the Lorentz group, with (0,0) being referred to as the spin-0 or "scalar representation", $(\frac{1}{2},0)$ and $(0,\frac{1}{2})$ both being referred to as the spin-$\frac{1}{2}$ or "spinor representation" (one left handed, the other right handed), and $(\frac{1}{2},\frac{1}{2})$ being referred to as the spin-1 or "vector representation".

Let us now consider the Poincaré group, which consists of the Lorentz group transformations with the addition of spacetime translations. The generator for these spacetime translations will be denoted P_μ. We can then work out the commutation relations of the Poncaré algebra:

$$[J_i, J_j] = i\varepsilon_{ijk}J_k$$
$$[J_i, K_j] = i\varepsilon_{ijk}K_k$$
$$[K_i, K_j] = -i\varepsilon_{ijk}J_k$$
$$[J_i, P_j] = i\varepsilon_{ijk}P_k \tag{A.5}$$
$$[J_i, P_0] = 0$$
$$[K_i, P_j] = i\delta_{ij}P_0$$
$$[K_i, P_0] = -iP_i$$

We can shorten this by defining a new object $M_{\mu\nu}$ such that:

$$J_i = \frac{1}{2}\varepsilon_{ijk}(M_{jk})$$
$$K_i = M_{0i} \tag{A.6}$$

The Poincaré algebra may then be written as:

$$[P_\mu, P_\nu] = 0$$
$$[M_{\mu\nu}, P_\rho] = i(\eta_{\mu\rho}P_\nu - \eta_{\nu\rho}P_\mu) \tag{A.7}$$
$$[M_{\mu\nu}, M_{\rho\sigma}] = i(\eta_{\mu\rho}M_{\nu\sigma} - \eta_{\mu\sigma}M_{\nu\rho} - \eta_{\nu\rho}M_{\mu\sigma} + \eta_{\nu\sigma}M_{\mu\rho})$$

where $\eta_{\mu\nu}$ is the Minkowski metric. From here we can form two Casimir elements which allow us to label the representations of the Poincaré group. The first is:

$$P_\mu P^\mu = m^2 \tag{A.8}$$

which is a continuous variable that we use to label the mass of the particle, and the second is $W_\mu W^\mu$ which is

referred to as the "Pauli-Lubanksi four-vector" and is defined as:

$$W^\mu = \frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}P_\nu M_{\rho\sigma} \tag{A.9}$$

which provides us with a discrete variable that we use to label the spin of the particle. This discussion is especially useful when looking to understand theories beyond the SM such as supersymmetry which attempts to expand the generators of the Poincaré group.

A.2 Mathematical Treatment of Vector Bosons

Mathematically, the vector bosons enter into the SM in a peculiar way. It was previously mentioned that each gauge group of the SM produces a conserved quantity, which is desirable given that conserved charges are readily observed in nature. Consider a "toy" Lagrangian of the form (which is sometimes referred to as "scalar QED"):

$$\mathcal{L} = \partial_\mu \phi^\dagger \partial^\mu \phi - m^2 \phi^\dagger \phi \tag{A.10}$$

where ϕ is a complex scalar field. Now observe when one introduces a gauge transformation (for this example, a $U(1)$ gauge) of the form $e^{iq\alpha}$, where i is the imaginary number, q is a constant, and α is an arbitrary function that is local (i.e. α is a function of spacetime that is allowed to vary from point to point). ϕ now transforms as $\phi(x) \to \phi'(x) = e^{-iq\alpha(x)}\phi(x)$. Let us vary the Lagrangian in this way and observe what happens:

$$\begin{aligned}
\mathcal{L}' &= \partial_\mu \phi'^\dagger \partial^\mu \phi' - m^2 \phi'^\dagger \phi' \\
&= \partial_\mu (e^{-iq\alpha}\phi^\dagger)\partial^\mu(e^{-iq\alpha}\phi) - m^2 e^{-iq\alpha}\phi^\dagger e^{-iq\alpha}\phi \\
&= (iq\partial_\mu\alpha e^{iq\alpha}\phi^\dagger + e^{iq\alpha}\partial_\mu\phi^\dagger)(-iq\partial^\mu\alpha e^{iq\alpha}\phi + e^{iq\alpha}\partial^\mu\phi) - m^2\phi^\dagger\phi \\
&= q^2\partial_\mu\alpha\partial^\mu\alpha\phi^*\phi + iq\partial_\mu\alpha\phi^\dagger\partial^\mu\phi - iq\partial_\mu\phi^\dagger\partial^\mu\alpha + \partial_\mu\phi^\dagger\partial^\mu\phi - m^2\phi^\dagger\phi \\
&\neq \mathcal{L}
\end{aligned} \tag{A.11}$$

It is obvious that the Lagrangian is not invariant under this transformation. Consider however if one is to add a new vector field A_μ to the theory via the covariant derivative and the so-called "minimum coupling" prescription: $\partial_\mu \to D_\mu = \partial_\mu - iqA_\mu$, where A_μ transforms as $A_\mu \to A'_\mu = A_\mu + \partial_\mu\alpha$. Let us first compute

how the covariant derivative acts on the scalar field ϕ as $\phi \rightarrow \phi'$:

$$
\begin{aligned}
D_\mu \phi' &= [\partial_\mu + iqA'_\mu]e^{-iq\alpha}\phi \\
&= [\partial_\mu + iq(A_\mu + \partial_\mu \alpha)]e^{-iq\alpha}\phi \\
&= -iq\partial_\mu \alpha e^{iq\alpha}\phi + e^{-iq\alpha}\partial_\mu \phi + iqA_\mu e^{-iq\alpha}\phi + iq\partial_\mu \alpha e^{-iq\alpha}\phi \\
&= e^{-iq\alpha}\partial_\mu \phi + iqA_\mu e^{iq\alpha}\phi \\
&= e^{-iq\alpha}D_\mu \phi
\end{aligned}
\tag{A.12}
$$

We see that the covariant derivative commutes with the transformation of ϕ. Now let us see how this affects the transformation $\mathscr{L} \rightarrow \mathscr{L}'$:

$$
\begin{aligned}
\mathscr{L} \rightarrow \mathscr{L}' &= (D_\mu \phi')^\dagger (D^\mu \phi') - m^2 \phi'^\dagger \phi' \\
&= (e^{-iq\alpha}D_\mu \phi)^\dagger (e^{-iq\alpha}D^\mu \phi) - m^2 \phi^\dagger \phi \\
&= D_\mu \phi^\dagger D^\mu \phi - m^2 \phi^\dagger \phi \\
&= \mathscr{L}
\end{aligned}
\tag{A.13}
$$

We find that by introducing a new vector field, we were able to ensure the gauge transformation was indeed a good symmetry of the Lagrangian. The power of this method is that by simply starting out with a scalar field (in the case of the SM, we would also need to consider spinor fields) and requiring a local gauge transformation, one is able to derive a conserved charge as well as particle interactions between the two fields. Consider however if we were to introduce a mass term for this vector field, which would be proportional to $\frac{m^2}{2}A_\mu A^\mu$. This additional term would again spoil the symmetry, implying that any new vector field which we introduce must be, similarly to the chiral fermions, massless. This is again at odds with nature however as there are experimentally observed massive bosons. The Higgs mechanism is able to resolve this dilemma, as explained in section.

A.3 Background Samples List

List of background simulation samples for 2016 in the NanoAODv6 data format and its corresponding cross sections in pb, where [*] = RunIISummer16NanoAODv6-PUMoriond17_Nano25Oct2019_102X_mcRun2_asymptotic_v7-v1

Process	Official CMS dataset	Cross section [pb]
$t\bar{t}$	/TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8/[*]/NANOAODSIM	88.29
	/TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8/[*]/NANOAODSIM	377.96
	/TTToHadronic_TuneCP5_PSweights_13TeV-powheg-pythia8/[*]/NANOAODSIM	365.34
Single top	/ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8_TuneCUETP8M1/[*]/NANOAODSIM	3.68
	/ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1/[*]/NANOAODSIM	80.95
	/ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1/[*]/NANOAODSIM	136.02
	/ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/[*]/NANOAODSIM	38.06
	/ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/[*]/NANOAODSIM	38.06
Z+jets H_T-incl.	/DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	18610.0 (NNLO)
	/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	6025.2 (NLO)
Z+jets H_T-binned ($5 < m(\ell\ell) \leq 50$ GeV)	/DYJetsToLL_M-5to50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	224.2
	/DYJetsToLL_M-5to50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	37.2
	/DYJetsToLL_M-5to50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	3.581
	/DYJetsToLL_M-5to50_HT-600toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1.124
Z+jets H_T-binned ($m(\ell\ell) \geq 50$ GeV)	/DYJetsToLL_M-50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	213.4
	/DYJetsToLL_M-50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	65.42
	/DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	7.31
	/DYJetsToLL_M-50_HT-600to800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1.49
	/DYJetsToLL_M-50_HT-800to1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.661
	/DYJetsToLL_M-50_HT-1200to2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.119
	/DYJetsToLL_M-50_HT-2500toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.0028
Diboson	/WWToLNuQQ_13TeV-powheg/[*]/NANOAODSIM	43.53
	/WWTo2L2Nu_13TeV-powheg/[*]/NANOAODSIM	10.48
	/WWTo4Q_13TeV-powheg/[*]/NANOAODSIM	51.723
	/GluGluWWTo2L2Nu_MCFM_13TeV/[*]/NANOAODSIM	0.5906
	/WpWpJJ_QCD_TuneCUETP8M1_13TeV-madgraph-pythia8/[*]/NANOAODSIM	0.02612
	/WWJJToLNuLNu_EWK_QCD_noTop-noHiggs_13TeV-madgraph-pythia8/[*]/NANOAODSIM	2.616
	/WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8/[*]/NANOAODSIM	10.73
	/WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8/[*]/NANOAODSIM	3.054
	/WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/[*]/NANOAODSIM	5.606
	/WZTo3LNu_TuneCUETP8M1_13TeV-powheg-pythia8/[*]/NANOAODSIM	4.43
	/ZZTo2L2Nu_13TeV_powheg_pythia8_ext1/[*]/NANOAODSIM	0.5644
	/ZZTo2L2Q_13TeV_powheg_pythia8/[*]/NANOAODSIM	3.222
	/ZZTo2Q2Nu_13TeV_powheg_pythia8/[*]/NANOAODSIM	4.033
	/ZZTo4L_13TeV_powheg_pythia8/[*]/NANOAODSIM	1.256
	/ZZTo4Q_13TeV_amcatnloFXFX_madspin_pythia8/[*]/NANOAODSIM	6.842
W+jets (H_T-incl.)	/WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	61334.0
W+jets (H_T-binned)	/WJetsToLNu_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1695.0
	/WJetsToLNu_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	532.4
	/WJetsToLNu_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	61.6
	/WJetsToLNu_HT-600To800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	12.4
	/WJetsToLNu_HT-800To1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	5.77
	/WJetsToLNu_HT-1200To2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1.023
	/WJetsToLNu_HT-2500ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.248
QCD (H_T-binned)	/QCD_HT50to100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	246300000.0
	/QCD_HT100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	27990000.0
	/QCD_HT200to300_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1559000.0
	/QCD_HT300to500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	351900.0
	/QCD_HT500to700_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	29070.0
	/QCD_HT700to1000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	5962.0
	/QCD_HT1000to1500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1005.0
	/QCD_HT1500to2000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	101.0
	/QCD_HT2000toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	20.54

Table A.1: List of background simulation samples for 2016 in the NanoAODv6 data format.

List of background simulation samples for 2016 in the NanoAODv6 data format and its corresponding cross sections in pb, where [*] = RunIISummer16NanoAODv6-PUMoriond17_Nano25Oct2019_102X_mcRun2_asymptotic_v7-v1.

Process	Official CMS dataset	Cross section [pb]
QCD (Muon-enriched)	/QCD_Pt-15to20_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/[*]/NANOAODSIM	3819570.0
	/QCD_Pt-20to30_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/[*]/NANOAODSIM	2960198.4
	/QCD_Pt-30to50_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/[*]/NANOAODSIM	1652471.5
	/QCD_Pt-50to80_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/[*]/NANOAODSIM	437504.5
	/QCD_Pt-80to120_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/[*]/NANOAODSIM	106033.7
	/QCD_Pt-120to170_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/[*]/NANOAODSIM	25190.5
	/QCD_Pt-170to300_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/[*]/NANOAODSIM	8654.5
	/QCD_Pt-300to470_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/[*]/NANOAODSIM	797.4
	/QCD_Pt-470to600_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/[*]/NANOAODSIM	79.0
	/QCD_Pt-600to800_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/[*]/NANOAODSIM	25.1
	/QCD_Pt-800to1000_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/[*]/NANOAODSIM	4.7
	/QCD_Pt-1000toInf_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8/[*]/NANOAODSIM	1.6
QCD (EM-enriched)	/QCD_Pt-20to30_EMEnriched_TuneCUETP8M1_13TeV_pythia8/[*]/NANOAODSIM	5352960.0
	/QCD_Pt-30to50_EMEnriched_TuneCUETP8M1_13TeV_pythia8/[*]/NANOAODSIM	9928000.0
	/QCD_Pt-50to80_EMEnriched_TuneCUETP8M1_13TeV_pythia8/[*]/NANOAODSIM	2890800.0
	/QCD_Pt-80to120_EMEnriched_TuneCUETP8M1_13TeV_pythia8/[*]/NANOAODSIM	350000.0
	/QCD_Pt-120to170_EMEnriched_TuneCUETP8M1_13TeV_pythia8/[*]/NANOAODSIM	629664.0
	/QCD_Pt-170to300_EMEnriched_TuneCUETP8M1_13TeV_pythia8/[*]/NANOAODSIM	18810.0
	/QCD_Pt-300toInf_EMEnriched_TuneCUETP8M1_13TeV_pythia8/[*]/NANOAODSIM	1350.0
VBS/VBF diboson	/WpWpJJ_EWK_TuneCUETP8M1_13TeV-madgraph-pythia8/[*]/NANOAODSIM	0.02695
	/WWJJToLNuLNu_EWK_noTop_13TeV-madgraph-pythia8/[*]/NANOAODSIM	0.3439
	/WLLJJ_WToLNu_EWK_TuneCUETP8M1_13TeV_madgraph-madspin-pythia8/[*]/NANOAODSIM	0.01762
	/ZZJJTo4L_EWK_13TeV-madgraph-pythia8/[*]/NANOAODSIM	0.0004454
	/ZZJJ_ZZTo2L2Nu_EWK_13TeV-madgraph-pythia8/[*]/NANOAODSIM	0.002971
WW/ZZ Double Parton Scattering	/WWTo2L2Nu_DoubleScattering_13TeV-pythia8/[*]/NANOAODSIM	0.170300
	/ZZTo4L_DoubleScattering_13TeV-pythia8/[*]/NANOAODSIM	0.929108
VBS/VBF W/Z+Jets	/EWKWPlus2Jets_WToLNu_M-50_13TeV-madgraph-pythia8/[*]/NANOAODSIM	25.81
	/EWKWMinus2Jets_WToLNu_M-50_13TeV-madgraph-pythia8/[*]/NANOAODSIM	20.35
	/EWKZ2Jets_ZToLL_M-50_13TeV-madgraph-pythia8/[*]/NANOAODSIM	3.997
	/EWKZ2Jets_ZToNuNu_13TeV-madgraph-pythia8/[*]/NANOAODSIM	10.04
Higgs	/GluGluHToZZTo4L_M125_13TeV_powheg2_JHUgenV6_pythia8/[*]/NANOAODSIM	0.0129763
	/VBF_HToZZTo4L_M125_13TeV_powheg2_JHUgenV6_pythia8/[*]/NANOAODSIM	0.0010102
	/WPlusH_HToMuMu_M125_13TeV_powheg-pythia8/[*]/NANOAODSIM	0.0001828
	/WMinusH_HToMuMu_M125_13TeV_powheg-pythia8/[*]/NANOAODSIM	0.001159
	/ZH_HToZZ_4LFilter_M125_13TeV_powheg2-minlo-HZJ_JHUgenV6_pythia8/[*]/NANOAODSIM	0.002361
	/ttH_HToZZ_4LFilter_M125_13TeV_powheg2_JHUgenV6_pythia8/[*]/NANOAODSIM	0.0001355
	/VBFHToBB_M-125_13TeV_powheg_pythia8/[*]/NANOAODSIM	2.183
	/GluGluHToBB_M125_13TeV_powheg_pythia8/[*]/NANOAODSIM	25.34
	/WplusH_HToBB_WToQQ_M125_13TeV_powheg_pythia8/[*]/NANOAODSIM	0.339
	/WminusH_HToBB_WToQQ_M125_13TeV_powheg_pythia8/[*]/NANOAODSIM	0.199
	/ZH_HToBB_ZToQQ_M125_13TeV_powheg_pythia8/[*]/NANOAODSIM	0.311
	/ggZH_HToBB_ZToQQ_M125_13TeV_powheg_pythia8/[*]/NANOAODSIM	0.043
	/bbHToBB_M-125_4FS_yb2_13TeV_amcatnlo/[*]/NANOAODSIM	0.310
$t\bar{t}$+X	/TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8/[*]/NANOAODSIM	0.2043
	/TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8/[*]/NANOAODSIM	0.4062
	/TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8/[*]/NANOAODSIM	0.2529
	/TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/[*]/NANOAODSIM	0.5297
	/TTGJets_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8/[*]/NANOAODSIM	3.697
	/TTTT_TuneCUETP8M1_13TeV-amcatnlo-pythia8/[*]/NANOAODSIM	0.009
Vγ+jets	/WGJJToLNu_EWK_QCD_TuneCUETP8M1_13TeV-madgraph-pythia8/[*]/NANOAODSIM	5.664
	/LLAJJ_EWK_MLL-50_MJJ-120_13TeV-madgraph-pythia8/[*]/NANOAODSIM	0.1084
	/LNuAJJ_EWK_MJJ-120_TuneCUETP8M1_13TeV-madgraph-pythia8/[*]/NANOAODSIM	0.776
	/ZGTo2LG_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/[*]/NANOAODSIM	123.8
Triboson	/WWW_4F_TuneCUETP8M1_13TeV-amcatnlo-pythia8/[*]/NANOAODSIM	0.2086
	/WWZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/[*]/NANOAODSIM	0.16510
	/WZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/[*]/NANOAODSIM	0.05565
	/ZZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/[*]/NANOAODSIM	0.1398

Table A.2: List of background simulation samples for 2016 in the NanoAODv6 data format (cont.).

List of background simulation samples for 2017 in the NanoAODv6 data format and its corresponding cross sections in pb, where [*] = RunIIFall17NanoAODv6-PU2017_12Apr2018_Nano25Oct2019_new_pmx_102X_mc2017_realistic_v7-v1, [**] = RunIIFall17NanoAODv6-PU2017_12Apr2018_Nano25Oct2019_102X_mc2017_realistic_v7-v1, [***] = RunIIFall17NanoAODv6-PU2017RECOSIMstep_12Apr2018_Nano25Oct2019_102X_mc2017_realistic_v7-v1, [4*] = RunIIFall17NanoAODv6-PU2017_12Apr2018_Nano25Oct2019_ext_102X_mc2017_realistic_v7-v1, [5*] = RunIIFall17NanoAODv6-PU2017_12Apr2018_Nano25Oct2019_102X_mc2017_realistic_v7-v2 and [6*] = RunIIFall17NanoAODv6-PU2017_12Apr2018_Nano25Oct2019_new_pmx_102X_mc2017_realistic_v7-v2.

Process	Official CMS dataset	Cross section [pb]
$t\bar{t}$	/TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8/[*]/NANOAODSIM	88.29
	/TTToHadronic_TuneCP5_PSweights_13TeV-powheg-pythia8/[*]/NANOAODSIM	377.96
	/TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8/[**]/NANOAODSIM	365.34
Single top	/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8/[**]/NANOAODSIM	3.68
	/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8/[**]/NANOAODSIM	80.95
	/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[**]/NANOAODSIM	136.02
	/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[**]/NANOAODSIM	38.06
	/ST_s-channel_4f_leptonDecays_TuneCP5_PSweights_13TeV-amcatnlo-pythia8/[*]/NANOAODSIM	38.06
Z+Jets H_T-incl.	/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	18610.0 (NNLO)
	/DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	6025.6 (NLO)
Z+jets M-4To50 H_T-binned	/DYJetsToLL_M-4to50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	224.2
	/DYJetsToLL_M-4to50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	37.2
	/DYJetsToLL_M-4to50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	3.581
	/DYJetsToLL_M-4to50_HT-600toInf_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	1.124
Z+jets M-50 H_T-binned	/DYJetsToLL_M-50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	213.4
	/DYJetsToLL_M-50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	65.42
	/DYJetsToLL_M-50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	7.31
	/DYJetsToLL_M-50_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1.49
	/DYJetsToLL_M-50_HT-800to1200_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.661
	/DYJetsToLL_M-50_HT-1200to2500_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	0.119
	/DYJetsToLL_M-50_HT-2500toInf_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.0028
Diboson	/WWTo2L2Nu_NNPDF31_TuneCP5_PSweights_13TeV-powheg-pythia8/[**]/NANOAODSIM	10.48
	/WWTo4Q_NNPDF31_TuneCP5_PSweights_13TeV-powheg-pythia8/[**]/NANOAODSIM	51.723
	/WWToLNuQQ_NNPDF31_TuneCP5_PSweights_13TeV-powheg-pythia8/[**]/NANOAODSIM	43.53
	/GluGluToWWToENEN_13TeV_MCFM701_pythia8/[5*]/NANOAODSIM	0.0457
	/GluGluToWWToENMN_13TeV_MCFM701_pythia8/[**]/NANOAODSIM	0.0457
	/GluGluToWWToENTN_13TeV_MCFM701_pythia8/[**]/NANOAODSIM	0.0457
	/GluGluToWWToMNEN_13TeV_MCFM701_pythia8/[5*]/NANOAODSIM	0.0457
	/GluGluToWWToMNMN_13TeV_MCFM701_pythia8/[**]/NANOAODSIM	0.0457
	/GluGluToWWToMNTN_13TeV_MCFM701_pythia8/[**]/NANOAODSIM	0.0457
	/GluGluToWWToTNEN_13TeV_MCFM701_pythia8/[**]/NANOAODSIM	0.0457
	/GluGluToWWToTNMN_13TeV_MCFM701_pythia8/[**]/NANOAODSIM	0.0457
	/GluGluToWWToTNTN_13TeV_MCFM701_pythia8/[**]/NANOAODSIM	0.0457
	/WpWpJJ_QCD_TuneCP5_13TeV-madgraph-pythia8/[**]/NANOAODSIM	0.02615
	/WWJJToLNuLNu_EWK_QCD_noTop-noHiggs_TuneCP5_13TeV-madgraph-pythia8/[5*]/NANOAODSIM	2.616
	/WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8/[**]/NANOAODSIM	10.73
	/WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8_v2/[**]/NANOAODSIM	3.054
	/WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/[**]/NANOAODSIM	5.606
	/WZTo3LNu_13TeV-powheg-pythia8/[**]/NANOAODSIM	4.43
	/ZZTo2L2Nu_13TeV_powheg-pythia8/[**]/NANOAODSIM	0.5644
	/ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/[**]/NANOAODSIM	3.222
	/ZZTo2Q2Nu_TuneCP5_13TeV_amcatnloFXFX_madspin_pythia8/[**]/NANOAODSIM	4.033
	/ZZTo4L_13TeV_powheg-pythia8/[*]/NANOAODSIM	1.256
W+Jets H_T-incl.	/WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	61334.0 (NLO)
	/WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/[4*]/NANOAODSIM	61334.0 (NLO)
W+Jets H_T-binned	/WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	1695.0
	/WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	532.4
	/WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	61.6
	/WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	12.4
	/WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	5.77
	/WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	1.023
	/WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	0.0248
QCD H_T-binned	/QCD_HT50to100_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	246300000.0
	/QCD_HT100to200_TuneCP5_13TeV-madgraph-pythia8/[**]/NANOAODSIM	27990000.0
	/QCD_HT200to300_TuneCP5_13TeV-madgraph-pythia8/[*]/NANOAODSIM	1559000.0
	/QCD_HT300to500_TuneCP5_13TeV-madgraph-pythia8/[*]/NANOAODSIM	351900.0
	/QCD_HT500to700_TuneCP5_13TeV-madgraph-pythia8/[**]/NANOAODSIM	29070.0
	/QCD_HT700to1000_TuneCP5_13TeV-madgraph-pythia8/[*]/NANOAODSIM	5962.0
	/QCD_HT1000to1500_TuneCP5_13TeV-madgraph-pythia8/[**]/NANOAODSIM	1005.0
	/QCD_HT1500to2000_TuneCP5_13TeV-madgraph-pythia8/[**]/NANOAODSIM	101.8
	/QCD_HT2000toInf_TuneCP5_13TeV-madgraph-pythia8/[**]/NANOAODSIM	20.54

Table A.3: List of background simulation samples for 2017 in the NanoAODv6 data format.

List of background simulation samples for 2017 in the NanoAODv6 data format and its corresponding cross sections in pb, where [*] = RunIIFall17NanoAODv6-PU2017_12Apr2018_Nano25Oct2019_new_pmx_102X_mc2017_realistic_v7-v1, [**] = RunIIFall17NanoAODv6-PU2017_12Apr2018_Nano25Oct2019_102X_mc2017_realistic_v7-v1, [***] = RunIIFall17NanoAODv6-PU2017RECOSIMstep_12Apr2018_Nano25Oct2019_102X_mc2017_realistic_v7-v1, [4*] = RunIIFall17NanoAODv6-PU2017_12Apr2018_Nano25Oct2019_ext_102X_mc2017_realistic_v7-v1, [5*] = RunIIFall17NanoAODv6-PU2017_12Apr2018_Nano25Oct2019_102X_mc2017_realistic_v7-v2 and [6*] = RunIIFall17NanoAODv6-PU2017_12Apr2018_Nano25Oct2019_new_pmx_102X_mc2017_realistic_v7-v2.

Process	Official CMS dataset	Cross section [pb]
QCD Muon enriched	/QCD_Pt-15to20_MuEnrichedPt5_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	3819570.0
	/QCD_Pt-20to30_MuEnrichedPt5_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	2960198.4
	/QCD_Pt-30to50_MuEnrichedPt5_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	1652471.5
	/QCD_Pt-50to80_MuEnrichedPt5_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	437504.5
	/QCD_Pt-80to120_MuEnrichedPt5_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	106033.7
	/QCD_Pt-120to170_MuEnrichedPt5_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	25190.5
	/QCD_Pt-170to300_MuEnrichedPt5_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	8654.5
	/QCD_Pt-300to470_MuEnrichedPt5_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	797.4
	/QCD_Pt-470to600_MuEnrichedPt5_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	79.0
	/QCD_Pt-600to800_MuEnrichedPt5_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	25.1
	/QCD_Pt-800to1000_MuEnrichedPt5_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	4.7
	/QCD_Pt-1000toInf_MuEnrichedPt5_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	1.6
QCD EM Enriched	/QCD_Pt-15to20_EMEnriched_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	1327000.0
	/QCD_Pt-20to30_EMEnriched_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	5352960.0
	/QCD_Pt-30to50_EMEnriched_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	9928000.0
	/QCD_Pt-50to80_EMEnriched_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	2890800.0
	/QCD_Pt-80to120_EMEnriched_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	350000.0
	/QCD_Pt-120to170_EMEnriched_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	62964.0
	/QCD_Pt-170to300_EMEnriched_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	18810.0
	/QCD_Pt-300toInf_EMEnriched_TuneCP5_13TeV_pythia8/[**]/NANOAODSIM	1350.0
VBS/VBF Diboson	/WpWpJJ_EWK_TuneCP5_13TeV-madgraph-pythia8/[**]/NANOAODSIM	0.02696
	/WWJJToLNuLNu_EWK_noTop_TuneCP5_13TeV-madgraph-pythia8/[**]/NANOAODSIM	0.3452
	/WLLJJ_WToLNu_EWK_TuneCP5_13TeV_madgraph-madspin-pythia8/[**]/NANOAODSIM	0.01628
	/ZZJJTo4L_EWK_TuneCP5_13TeV-madgraph-pythia8/[**]/NANOAODSIM	0.00044
WW/ZZ Double Parton Scattering	/WWTo2L2Nu_DoubleScattering_13TeV-pythia8/[**]/NANOAODSIM	0.1703
	/ZZTo4L_TuneCP5_DoubleScattering_13TeV-pythia8/[**]/NANOAODSIM	0.929108
VBS/VBF W/Z+jets	/EWKWMinus2Jets_WToLNu_M-50_TuneCP5_13TeV-madgraph-pythia8/[*]/NANOAODSIM	23.24
	/EWKWPlus2Jets_WToLNu_M-50_TuneCP5_13TeV-madgraph-pythia8/[*]/NANOAODSIM	29.59
	/EWKZ2Jets_ZToLL_M-50_TuneCP5_13TeV-madgraph-pythia8/[*]/NANOAODSIM	4.321
	/EWKZ2Jets_ZToNuNu_TuneCP5_13TeV-madgraph-pythia8/[*]/NANOAODSIM	10.66
Vγ+jets	/LLAJJ_EWK_MLL-50_MJJ-120_TuneCP5_13TeV-madgraph-pythia8/[**]/NANOAODSIM	0.1097
	/LNuAJJ_EWK_MJJ-120_TuneCP5_13TeV-madgraph-pythia8/[**]/NANOAODSIM	0.5345
	/WGJJToLNu_EWK_QCD_TuneCP5_13TeV-madgraph-pythia8/[**]/NANOAODSIM	5.05
	/ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8/[**]/NANOAODSIM	50.43
Higgs	/VBF_HToZZTo4L_M125_13TeV-powheg2_JHUGenV7011-pythia8/[**]/NANOAODSIM	0.0010102
	/VBFHToBB_M-125_13TeV_powheg_pythia8_weightfix/[**]/NANOAODSIM	2.183
	/GluGluHToBB_M125_TuneCP5_13TeV-powheg-pythia8/[**]/NANOAODSIM	25.340
	/GluGluHToZZTo2L2Q_M125_13TeV_powheg2_JHUGenV7011-pythia8/[**]/NANOAODSIM	0.1618
	/GluGluHToZZTo4L_M125_13TeV_powheg2_JHUGenV7011-pythia8/[**]/NANOAODSIM	0.0129763
	/ZH_HToZZ_4LFilter_M125_13TeV_powheg2-minlo-HZJ_JHUGenV7011-pythia8/[**]/NANOAODSIM	0.0002361
	/ZH_HToBB_ZToLL_M125_13TeV_powheg_pythia8/[**]/NANOAODSIM	0.07523
	/ttH_HToZZ_4LFilter_M125_13TeV_powheg2_JHUGenV7011-pythia8/[**]/NANOAODSIM	0.0001355
$t\bar{t}$+X	/TTWJetsToLNu_TuneCP5_PSweights_13TeV-amcatnloFXFX-madspin-pythia8/[6*]/NANOAODSIM	0.2043
	/TTWJetsToQQ_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/[**]/NANOAODSIM	0.4062
	/TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8/[**]/NANOAODSIM	0.2529
	/TTZToQQ_TuneCP5_13TeV-amcatnlo-pythia8/[**]/NANOAODSIM	0.5297
	/TTGJets_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/[*]/NANOAODSIM	3.697
	/TTTT_TuneCP5_13TeV-amcatnlo-pythia8/[**]/NANOAODSIM	0.009
Triboson	/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8/[**]/NANOAODSIM	0.2086
	/WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8/[**]/NANOAODSIM	0.1651
	/WZZ_TuneCP5_13TeV-amcatnlo-pythia8/[4*]/NANOAODSIM	0.05565
	/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8/[4*]/NANOAODSIM	0.0139

Table A.4: List of background simulation samples for 2017 in the NanoAODv6 data format (cont.).

List of background simulation samples for 2018 in the NanoAODv6 data format and its corresponding cross sections in pb, where [*] = RunIIAutumn18NanoAODv6-Nano25Oct2019_102X_upgrade2018_realistic_v20-v1, [**] = RunIIAutumn18NanoAODv6-Nano25Oct2019_102X_upgrade2018_realistic_v20-v2, [***] = RunIIAutumn18NanoAODv6-Nano25Oct2019_102X_upgrade2018_realistic_v20-v3.

Process	Official CMS dataset	Cross section [pb]
$t\bar{t}$	/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	88.29
	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8/[***]/NANOAODSIM	377.96
	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	365.34
Single top	/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-madgraph-pythia8/[*]/NANOAODSIM	3.68
	/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[*]/NANOAODSIM	80.95
	/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[*]/NANOAODSIM	136.02
	/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	38.06
	/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	38.06
Z+jets H_T-incl.	/DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	18610.0
	/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	6025.2
Z+jets M-4to50 H_T-binned	/DYJetsToLL_M-4to50_HT-100to200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	224.2
	/DYJetsToLL_M-4to50_HT-200to400_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	37.2
	/DYJetsToLL_M-4to50_HT-400to600_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	3.581
	/DYJetsToLL_M-4to50_HT-600toInf_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1.124
Z+jets M-50 H_T-binned	/DYJetsToLL_M-50_HT-100to200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	213.4
	/DYJetsToLL_M-50_HT-200to400_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	65.42
	/DYJetsToLL_M-50_HT-400to600_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	7.31
	/DYJetsToLL_M-50_HT-600to800_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1.49
	/DYJetsToLL_M-50_HT-800to1200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.661
	/DYJetsToLL_M-50_HT-1200to2500_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.119
	/DYJetsToLL_M-50_HT-2500toInf_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.0028
Diboson	/WWTo2L2Nu_NNPDF31_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	10.48
	/WWTo4Q_NNPDF31_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	51.723
	/WWToLNuQQ_NNPDF31_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	43.53
	/GluGluToWWToENEN_TuneCP5_13TeV_MCFM701_pythia8/[*]/NANOAODSIM	0.0457
	/GluGluToWWToENMN_TuneCP5_13TeV_MCFM701_pythia8/[*]/NANOAODSIM	0.0457
	/GluGluToWWToENTN_TuneCP5_13TeV_MCFM701_pythia8/[*]/NANOAODSIM	0.0457
	/GluGluToWWToMNEN_TuneCP5_13TeV_MCFM701_pythia8/[*]/NANOAODSIM	0.0457
	/GluGluToWWToMNMN_TuneCP5_13TeV_MCFM701_pythia8/[*]/NANOAODSIM	0.0457
	/GluGluToWWToMNTN_TuneCP5_13TeV_MCFM701_pythia8/[*]/NANOAODSIM	0.0457
	/GluGluToWWToTNEN_TuneCP5_13TeV_MCFM701_pythia8/[*]/NANOAODSIM	0.0457
	/GluGluToWWToTNMN_TuneCP5_13TeV_MCFM701_pythia8/[*]/NANOAODSIM	0.0457
	/GluGluToWWToTNTN_TuneCP5_13TeV_MCFM701_pythia8/[*]/NANOAODSIM	0.0457
	/WpWpJJ_QCD_TuneCP5_13TeV-madgraph-pythia8/[*]/NANOAODSIM	0.02615
	/WWJJToLNuLNu_QCD_noTop_13TeV-madgraph-pythia8/[*]/NANOAODSIM	0.02615
	/WWJJToLNuLNu_EWK_QCD_noTop-noHiggs_13TeV-madgraph-pythia8/[*]/NANOAODSIM	2.616
	/WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8/[*]/NANOAODSIM	10.73
	/WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8/[*]/NANOAODSIM	3.054
	/WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/[*]/NANOAODSIM	5.606
	/WZTo3LNu_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	4.43
	/ZZTo2L2Nu_TuneCP5_13TeV_powheg_pythia8/[*]/NANOAODSIM	0.5644
	/ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/[*]/NANOAODSIM	3.222
	/ZZTo2Q2Nu_TuneCP5_13TeV_amcatnloFXFX_madspin_pythia8/[*]/NANOAODSIM	4.033
	/ZZTo4L_13TeV_powheg_pythia8_TuneCP5/[*]/NANOAODSIM	1.256
	/ZZTo4L_TuneCP5_13TeV_powheg_pythia8/[*]/NANOAODSIM	1.256
W+jets H_T-incl.	/WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	61334.90
W+Jets H_T-binned	/WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1695.0
	/WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	532.4
	/WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	61.6
	/WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	12.4
	/WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	5.77
	/WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1.023
	/WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.0248
QCD H_T-binned	/QCD_HT50to100_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	246300000.0
	/QCD_HT100to200_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	27990000.0
	/QCD_HT200to300_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1559000.0
	/QCD_HT300to500_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	351900.0
	/QCD_HT500to700_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	29070.0
	/QCD_HT700to1000_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	5962.0
	/QCD_HT1000to1500_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1005.0
	/QCD_HT1500to2000_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	101.0
	/QCD_HT2000toInf_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	20.54

Table A.5: List of background simulation samples for 2018 in the NanoAODv6 data format.

List of background simulation samples for 2018 in the NanoAODv6 data format and its corresponding cross sections in pb, where [*] = RunIIAutumn18NanoAODv6-Nano25Oct2019_102X_upgrade2018_realistic_v20-v1, [**] = RunIIAutumn18NanoAODv6-Nano25Oct2019_102X_upgrade2018_realistic_v20-v2, [***] = RunIIAutumn18NanoAODv6-Nano25Oct2019_102X_upgrade2018_realistic_v20-v3.

Process	Official CMS dataset	Cross section [pb]
QCD Muon enriched	/QCD_Pt-15to20_MuEnrichedPt5_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	3819570.0
	/QCD_Pt-20to30_MuEnrichedPt5_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	2960198.4
	/QCD_Pt-30to50_MuEnrichedPt5_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	1652471.5
	/QCD_Pt-50to80_MuEnrichedPt5_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	437504.5
	/QCD_Pt-80to120_MuEnrichedPt5_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	106033.7
	/QCD_Pt-120to170_MuEnrichedPt5_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	25190.5
	/QCD_Pt-170to300_MuEnrichedPt5_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	8654.5
	/QCD_Pt-300to470_MuEnrichedPt5_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	797.4
	/QCD_Pt-470to600_MuEnrichedPt5_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	79.0
	/QCD_Pt-600to800_MuEnrichedPt5_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	25.1
	/QCD_Pt-800to1000_MuEnrichedPt5_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	4.7
	/QCD_Pt-1000toInf_MuEnrichedPt5_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	1.6
QCD EM Enriched	/QCD_Pt-15to20_EMEnriched_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	1327000.0
	/QCD_Pt-20to30_EMEnriched_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	5352960.0
	/QCD_Pt-30to50_EMEnriched_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	99280000.0
	/QCD_Pt-50to80_EMEnriched_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	2890800.0
	/QCD_Pt-80to120_EMEnriched_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	350000.0
	/QCD_Pt-120to170_EMEnriched_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	62964.0
	/QCD_Pt-170to300_EMEnriched_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	18810.0
	/QCD_Pt-300toInf_EMEnriched_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	1350.0
VBS/VBF Diboson	/WpWpJJ_EWK_TuneCP5_13TeV-madgraph-pythia8/[*]/NANOAODSIM	0.02696
	/WWJJToLNuLNu_EWK_noTop_13TeV-madgraph-pythia8/[*]/NANOAODSIM	0.3452
	/WLLJJ_WToLNu_EWK_TuneCP5_13TeV_madgraph-madspin-pythia8/[*]/NANOAODSIM	0.01628
	/ZZJJTo4L_EWK_TuneCP5_13TeV-madgraph-pythia8/[*]/NANOAODSIM	0.00044
WW/ZZ Double Parton Scattering	/WWTo2L2Nu_DoubleScattering_13TeV-pythia8/[*]/NANOAODSIM	0.1703
	/ZZTo4L_TuneCP5_DoubleScattering_13TeV-pythia8/[**]/NANOAODSIM	0.929108
Vγ+jets	/LLAJJ_EWK_MLL-50_MJJ-120_TuneCP5_13TeV-madgraph-pythia8/[*]/NANOAODSIM	0.1097
	/LNuAJJ_EWK_MJJ-120_TuneCP5_13TeV-madgraph-pythia8/[*]/NANOAODSIM	0.5345
	/ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8/[*]/NANOAODSIM	50.43
	/WGJJToLNu_EWK_QCD_TuneCP5_13TeV-madgraph-pythia8/[*]/NANOAODSIM	5.05
VBS/VBF W/Z+jets	/EWKWMinus2Jets_WToLNu_M-50_TuneCP5_13TeV-madgraph-pythia8/[*]/NANOAODSIM	23.24
	/EWKWPlus2Jets_WToLNu_M-50_TuneCP5_13TeV-madgraph-pythia8/[*]/NANOAODSIM	29.59
	/EWKZ2Jets_ZToLL_M-50_TuneCP5_PSweights_13TeV-madgraph-pythia8/[*]/NANOAODSIM	4.321
	/EWKZ2Jets_ZToNuNu_TuneCP5_PSweights_13TeV-madgraph-pythia8/[*]/NANOAODSIM	10.66
Higgs	/GluGluHToZZTo4L_M125_13TeV_powheg2_JHUGenV7011_pythia8/[*]/NANOAODSIM	0.0129763
	/VBF_HToZZTo4L_M125_13TeV_powheg2_JHUGenV7011_pythia8/[*]/NANOAODSIM	0.0010102
	/VBFHToBB_M-125_13TeV_powheg_pythia8_weightfix/[*]/NANOAODSIM	2.183
	/GluGluHToBB_M-125_13TeV_powheg_MINLO_NNLOPS_pythia8/[*]/NANOAODSIM	25.340
	/ZH_HToBB_ZToLL_M125_13TeV_powheg_pythia8/[*]/NANOAODSIM	0.311
	/ZH_HToBB_ZToQQ_M125_13TeV_powheg_pythia8/[*]/NANOAODSIM	0.311
	/ggZH_HToBB_ZToBB_M125_TuneCP5_13TeV_powheg-pythia8/[*]/NANOAODSIM	0.07784
	/ggZH_HToBB_ZToLL_M125_13TeV_powheg_pythia8/[*]/NANOAODSIM	0.006954
	/ggZH_HToBB_ZToQQ_M125_13TeV_powheg_pythia8/[*]/NANOAODSIM	0.04884
	/ttH_HToZZ_4LFilter_M125_13TeV_powheg2_JHUGenV7011_pythia8/[*]/NANOAODSIM	0.0001355
	/GluGluHToZZTo2L2Q_M125_13TeV_powheg2_JHUGenV7011_pythia8/[*]/NANOAODSIM	0.1618
	/ZH_HToZZ_4LFilter_M125_13TeV_powheg2-minlo-HZJ_JHUGenV7011_pythia8/[*]/NANOAODSIM	0.0002361
$t\bar{t}$+X	/TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/[*]/NANOAODSIM	0.2043
	/TTWJetsToQQ_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/[*]/NANOAODSIM	0.4062
	/TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8/[*]/NANOAODSIM	0.2529
	/TTZToQQ_TuneCP5_13TeV-amcatnlo-pythia8/[*]/NANOAODSIM	0.5297
	/TTGJets_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/[*]/NANOAODSIM	3.697
	/TTTT_TuneCP5_13TeV-amcatnlo-pythia8/[*]/NANOAODSIM	0.009
Triboson	/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8/[*]/NANOAODSIM	0.2086
	/WWZ_TuneCP5_13TeV-amcatnlo-pythia8/[*]/NANOAODSIM	0.1651
	/WZZ_TuneCP5_13TeV-amcatnlo-pythia8/[*]/NANOAODSIM	0.05565
	/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8/[*]/NANOAODSIM	0.01398

Table A.6: List of background simulation samples for 2018 in the NanoAODv6 data format (cont.).

References

[1] Geraldine Servant. Concepts in HEP (Fundamental Concepts in Particle Physics).

[2] Matthew Robinson. *Symmetry and the Standard Model*. Springer, 1st edition, 2011.

[3] Sarah Charley. The LHC does a dry run.

[4] Saranya Samik Ghosh and on behalf of the CMS Collaboration. Highlights from the compact muon solenoid (CMS) experiment. *Universe*, 5(1), 2019.

[5] David Barney and Sergio Cittolin. CMS Detector Drawings, Jan 2000.

[6] Lucas Taylor. Silicon pixels, 2011. Accessed: 2020-01-01.

[7] Cristina Biino. The CMS electromagnetic calorimeter: overview, lessons learned during run 1 and future projections. *Journal of Physics: Conference Series*, 587:012001, feb 2015.

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

[9] Pierluigi Paolucci. The CMS Muon system, Apr 2005.

[10] CMS Collaboration. The CMS trigger system. *Journal of Instrumentation*, 12(01):P01020–P01020, jan 2017.

[11] CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. *Journal of Instrumentation*, 12(10):P10003–P10003, oct 2017.

[12] A.M. Sirunyan et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{S}$=13 TeV. *Journal of Instrumentation*, 13(06):P06015–P06015, jun 2018.

[13] G. McCabe. *The Structure and Interpretation of the Standard Model*. Elsevier, 1st edition, 2007.

[14] G Rajasekaran. The Story of the Neutrino. 6 2016.

[15] C. Giunti and C.W. Kim. Fundeamentals of Neutrino Physics and Astrophysics. 2007.

[16] K. Kainulainen and K.A. Olive. Astrophysical and Cosmological Constraints on Neutrino Masses. 2002.

[17] Ubaldo Dore and Lucia Zanello. Bruno Pontecorvo and neutrino physics. 10 2009.

[18] M. Shiozawa. Evidence for neutrino oscillations in atmospheric neutrino observations on behalf of the super-kamiokande and kamiokande collaborations. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 433(1):307–313, 1999.

[19] Tsutomu Yanagida. Horizontal Symmetry and Masses of Neutrinos. *Prog. Theor. Phys.*, 64:1103, 1980.

[20] Rabindra N. Mohapatra and R. E. Marshak. Local B-L Symmetry of Electroweak Interactions, Majorana Neutrinos and Neutron Oscillations. *Phys. Rev. Lett.*, 44:1316–1319, 1980. [Erratum: Phys.Rev.Lett. 44, 1643 (1980)].

[21] T. Ohlsson M. Lindner and G. Seidl. Seesaw mechanisms for Dirac and Majorana neutrino masses. 2002.

[22] P. Minkowski. $\mu \rightarrow e\gamma$ at a rate of one out of 10^9 muon decays? *Phys. Lett. B 67*, 1977.

[23] Rabindra N. Mohapatra and Goran Senjanović. Neutrino mass and spontaneous parity nonconservation. *Phys. Rev. Lett.*, 44:912–915, Apr 1980.

[24] Bhaskar Dutta, Alfredo Gurrola, Will Johns, Teruki Kamon, Paul Sheldon, and Kuver Sinha. Vector Boson Fusion Processes as a Probe of Supersymmetric Electroweak Sectors at the LHC. *Phys. Rev. D*, 87(3):035029, 2013.

[25] Vardan Khachatryan et al. Search for heavy Majorana neutrinos in $\mu^{\pm}\mu^{\pm}+$ jets events in proton-proton collisions at $\sqrt{s} = 8$ TeV. *Phys. Lett. B*, 748:144–166, 2015.

[26] Vardan Khachatryan et al. Search for heavy Majorana neutrinos in ee+ jets and e μ+ jets events in proton-proton collisions at $\sqrt{s} = 8$ TeV. *JHEP*, 04:169, 2016.

[27] Albert M Sirunyan et al. Search for heavy Majorana neutrinos in same-sign dilepton channels in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 01:122, 2019.

[28] Albert M Sirunyan et al. Search for heavy neutrinos and third-generation leptoquarks in hadronic states of two τ leptons and two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 03:170, 2019.

[29] Albert M Sirunyan et al. Search for third-generation scalar leptoquarks and heavy right-handed neutrinos in final states with two tau leptons and two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 07:121, 2017.

[30] Serguei Chatrchyan et al. Search for heavy Majorana Neutrinos in $\mu^{\pm}\mu^{\pm}+$ Jets and $e^{\pm}e^{\pm}+$ Jets Events in pp Collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett. B*, 717:109–128, 2012.

[31] Search for heavy Majorana neutrinos in e+/- e+/- + jets and e+/- mu+/- + jets events in proton-proton collisions at sqrt(s) = 8 TeV. 2015.

[32] Search for heavy Majorana neutrinos in the same-sign dilepton channel in proton-proton collisions at $\sqrt{s} = 13$ TeV. 2018.

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

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

[35] Lucas Taylor. Silicon strips, 2011. Accessed: 2020-01-01.

[36] G. L. Bayatian et al. CMS Physics: Technical Design Report Volume 1: Detector Performance and Software, 2006.

[37] The CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[38] Albert M Sirunyan et al. Calibration of the cms hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13$ tev. *JINST*, 15(05):P05002, May 2020.

[39] CMS Collaboration. *The CMS magnet project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997.

[40] Serguei Chatrchyan et al. The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s}$ = 7 TeV at the LHC. *JINST*, 8:P11002, 2013.

[41] Dordevic, Milos. The cms particle flow algorithm. *EPJ Web Conf.*, 191:02016, 2018.

[42] CMS Collaboration. Performance of the missing transverse energy reconstruction in the 8 tev data. *CMS PAS*, JME-12-002, 2012.

[43] Andrea Malara. Reconstruction of jets and missing transverse momentum at the CMS experiment: Run 2 and perspective for Run 3. *PoS*, ICHEP2020:752. 6 p, Dec 2020. 6 pages, 11 figures, Contribution to 40th International Conference on High Energy physics - ICHEP2020, Accepted for publication on POS https://pos.sissa.it/390/752/pdf.

[44] CMS Collaboration. Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s}$ = 13 TeV using the CMS detector. *Journal of Instrumentation*, 14(07):P07004–P07004, jul 2019.

[45] A.M. Sirunyan et. al. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *Journal of Instrumentation*, 16(05):P05014, may 2021.

[46] Izaak Neutelings. Hadronic tau reconstruction and identification performance in ATLAS and CMS. *PoS*, LHCP2020:045, 2021.

[47] CMS Collaboration. Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS. *Journal of Instrumentation*, 11(01):P01019–P01019, jan 2016.

[48] S. Cahn and J.D. Dawson. Production of very massive higgs bosons. *Physical Review D*, 26(5):996–1005, 1982.

[49] *Physics in Collision VI. Proceedings of the International Conference*. World Scientific, 1997.

[50] J.D. Bjorken. Rapidity gaps and jets as a new-physics signature in very-high-energy hadron-hadron collisions. *Physical Review D*, 47(1):101–113, 1993.

[51] A. Florez et al. Expanding the reach of heavy neutrino searches at the lhc. *Physical Review Letters*, 2017.

[52] Andres G. Delannoy et al. Probing Dark Matter at the LHC using Vector Boson Fusion Processes. *Phys. Rev. Lett.*, 111:061801, 2013.

[53] Andrés Flórez, Kaiwen Gui, Alfredo Gurrola, Carlos Patiño, and Diego Restrepo. Expanding the Reach of Heavy Neutrino Searches at the LHC. *Phys. Lett. B*, 778:94–100, 2018.

[54] Rikkert Frederix, Emanuele Re, and Paolo Torrielli. Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO. *JHEP*, 09:130, 2012.

[55] Michal Czakon and Alexander Mitov. Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders. *Comput. Phys. Commun.*, 185:2930, 2014.

[56] Albert M Sirunyan et al. Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements. *Eur. Phys. J. C*, 80(1):4, 2020.

[57] 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. *JHEP*, 07:079, 2014.

[58] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

[59] S. Agostinelli et al. GEANT4–a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.

[60] Ehatäht, Karl. Nanoaod: a new compact event data format in cms. *EPJ Web Conf.*, 245:06002, 2020.

[61] Laurent Thomas. Reweighting recipe to emulate Level 1 ECAL prefiring.

[62] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler. Soft Drop. *JHEP*, 05:146, 2014.

[63] Jesse Thaler and Ken Van Tilburg. Maximizing Boosted Top Identification by Minimizing N-subjettiness. *JHEP*, 02:093, 2011.

[64] S. Chatrchyan et al. Identification of b-quark jets with the CMS experiment. *JINST*, 8:P04013, 2013.

[65] T. Junk. Interpretation of search results - the cls technique. *European Physical Journal C*, 71:1554, 2011.

[66] CMS Collaboration. Precision luminosity measurement in proton-proton collisions at $\sqrt{s}$= 13 tev in 2015 and 2016 at cms. *European Physical Journal C*, 81, 2021.

[67] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s}$ = 13 TeV. CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2018.

[68] CMS Collaboration. CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s}$ = 13 TeV. CMS Physics Analysis Summary CMS-PAS-LUM-18-002, 2019.

[69] Jakob Schwichtenberg. *Physics from Symmetry*. Springer, 2nd edition, 2018.