Search for rare W boson decays to three charged pions using machine learning techniques in proton-proton collisions at the Large Hadron Collider

by

James V. Natoli

B.S., University of Maryland, College Park, 2019

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2025

Abstract

A search is presented for the rare decay of a W boson into the fully hadronic final state of three charged pions. This search uses a combination of proton-proton collision data collected by the CMS experiment at the CERN LHC at center-of-mass energies 13 TeV and 13.6 TeV corresponding to integrated luminosities of 137.6 fb^{-1} and 61.9 fb^{-1}, respectively. A ditau trigger is used to select events. This takes advantage of algorithms designed to reconstruct and identify pions from the hadronic decay of tau leptons. As a result of the very low transverse momentum distribution of the third pion, we select isolated tracks which satisfy an overall charge requirement on the three objects of plus or minus one. This search uses machine learning tools to improve the sensitivity towards signal. In addition, this thesis presents two projects completed in service of the CMS collaboration. The first quantifies radiation damage to the physical components of the detector and derives necessary corrections. The second encompasses the development and maintenance of a framework which synchronizes critical data used for alignment and calibrations of the detector.

Search for rare W boson decays to three charged pions using machine learning techniques in proton-proton collisions at the Large Hadron Collider

by

James V. Natoli

B.S., University of Maryland, College Park, 2019

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2025

Approved by:

Major Professor
Yurii Maravin

Copyright

Abstract

A search is presented for the rare decay of a W boson into the fully hadronic final state of three charged pions. This search uses a combination of proton-proton collision data collected by the CMS experiment at the CERN LHC at center-of-mass energies 13 TeV and 13.6 TeV corresponding to integrated luminosities of $137.6\,\mathrm{fb}^{-1}$ and $61.9\,\mathrm{fb}^{-1}$, respectively. A ditau trigger is used to select events. This takes advantage of algorithms designed to reconstruct and identify pions from the hadronic decay of tau leptons. As a result of the very low transverse momentum distribution of the third pion, we select isolated tracks which satisfy an overall charge requirement on the three objects of plus or minus one. This search uses machine learning tools to improve the sensitivity towards signal. In addition, this thesis presents two projects completed in service of the CMS collaboration. The first quantifies radiation damage to the physical components of the detector and derives necessary corrections. The second encompasses the development and maintenance of a framework which synchronizes critical data used for alignment and calibrations of the detector.

Table of Contents

List of Figures

List of Tables

Acknowledgments

My experience in graduate school has been influenced by countless people. I can not possibly acknowledge everyone who has helped me on this journey but I would nonetheless like to try.

First, I'd like to thank Yurii Maravin, my advisor. His relentless support was a necessary constant during my years of research. We approached and overcame several serious hurdles in the analysis together and each time I felt encouraged and excited. As I took on larger roles in the CMS collaboration I always felt confident knowing that I had his endorsement. I'd also like to thank Keti Kaadze for her guidance and geniality. Together, she and Yurii curate a positive and comfortable environment for our research group. They are a wonderful team and it was my pleasure to work under their tutelage and get to know their lovely family.

Next, I want to express my gratitude towards the rest of the K-State High Energy Physics group. Professors Tim Bolton and Glenn Horton-Smith provided valuable perspective from their expertise in neutrino physics. I am also thankful towards Andrew Ivanov and Victor Turchin for agreeing to serve on my committee. I want to acknowledge Professors Loren Greenman, Daniel Rolles, Bharat Ratra, and J.T. Laverty for their guidance, both in and out of the classroom.

I'd also like to express my appreciation for Dennis Roy. He was willing to assist me with all matters ranging from debugging python code to ordering desserts in French. His model of work ethic and selfless cooperation served as a guide for me as I progressed in my graduate career. I also owe many thanks to our other postdoctoral researcher, Giulia Sorrentino, for being an excellent role model of confidence and success. I am grateful for all of the fellow graduate students that I had the pleasure of working closely in research with: Guru, Athar, Naila, and Braden. Especially to Braden, who has spent far more time with me than anyone should: thank you. From Manhattan to Geneva and back, we navigated planes, trains, busses, bikes, and boats together. Your friendship, trust, and kindness made my life as a graduate student unambiguously better.

I'd also like to thank all of the K-State physics graduate students for their friendship. Specifically to Lydia for welcoming me to K-State, to Jo Lynn for being a constant beam of light even in dark places, and to Sajed for being staunchly dependable: thank you. I made many friends both in Manhattan and Geneva, all of whom I am grateful towards and very happy to have known. Thank you to Tyler, Rahasya, and Bill for countless ridiculous evenings and for keeping graduate school as light-hearted as possible. I owe specific appreciation to Iris and Jesse, the first friends I made through research. They showed me how fun physics could be with friends by your side and inspired me to continue making connections.

I want to thank my friends from Bel Air: Rachel, Natalie, Shannon, Jordan, Eric, Kevin, Nate, and Bernie. It is an honor have grown up together and I'm continually grateful for your friendship. To Bernie in particular who embarked on a graduate journey at the same time as me, I want to thank you for your advice and friendship.

I'd like to thank the members of the HCAL group in the CMS collaboration. It was my pleasure to work closely with one piece of the detector for this long, importantly because it allowed me to meet so many great people. Specifically I want to thank Salavat Abdoulline, Pasha Parygin, and Alberto Belloni and Sarah Eno. Sarah and Alberto fostered my initial interest in particle physics and laid the foundation for me to have a successful graduate career. There are too many people at CERN to thank, but I want to specifically acknowledge Ana, the barista at the cafe in building 14. Her welcoming smile and a kind words (in both French and English) went a long way to making me feel welcome in a new and strange environment. I want to thank Kim Coy for her help from the very beginning when I needed a reminder to accept the admission to K-State to the very end when she helped me turn in all of my materials to graduate on time.

Finally, thank you to my family: my dad, sister, brother, and mother. Without your strength, love, and guidance none of this would have been possible.

Dedication

For my family

Chapter 1

Introduction

Particle physics emerged as a modern scientific field over the last century to describe the properties and interactions of the most fundamental objects in nature. The standard model of particle physics (SM) is the hallmark achievement of this effort. It can be used to predict many physical quantities, all of which stand up to strict experimental verification. Despite the accuracy of the SM there are still gaps in our understanding of the universe. As a result, modern particle physics experiments continue to make increasingly precise measurements of physical quantities. Any observed difference from the SM prediction of a quantity could provide a hint to what is missing. At the same time, experiments are searching for new particles which might exist. Since the discovery of the Higgs boson in 2012, all particles predicted by the SM have been observed. Any additional particles would require theoretical extensions that could also explain some of the unresolved questions.

In this thesis a search for a rare decay of the W boson to three charged pions is presented. The SM predicts that the W boson can decay to hadronic particles such as pions. None have yet been experimentally observed due to their low decay rates and experimental complexities. It is therefore a necessary aspect of the SM to probe as any deviation could be an indication of new physics.

The remainder of Section 1 of this thesis will describe the scientific theory underpinning the SM. Following this will be a technical description of the Large Hadron Collider (LHC)

Generation	Particle	Mass [GeV]	Charge [e]
1	e	0.511×10^{-3}	-1
2	μ	0.106	-1
3	τ	1.776	-1
1	ν_{e}	$< 0.8 \times 10^{-6}$	0
2	ν_{μ}	$< 1.9 \times 10^{-4}$	0
3	ν_{τ}	$< 18.2 \times 10^{-3}$	0

Table 1.1: *Summary of the properties of the six fundamental leptons*[2].

and Compact Muon Solenoid (CMS) detector in Section 2. In Section 3, the details are specified for the search for a rare decay of the W boson to three charged pions. Section 4 will discuss efforts performed to support the continued successful operation of the CMS detector, followed by results and conclusions in Section 5.

1.1 Elementary particles

The particles of the SM are organized into categories based on certain physical properties. The broadest of these categories are fermions and bosons. Fermions are particles with half-integer spin values that obey Fermi-Dirac statistics while bosons have integer spin and obey Bose-Einstein statistics.

The fermions consist of leptons (ℓ^-) and quarks (q), each of which have three generations of particles. Each generation of leptons consists of one electrically charged and one electrically neutral particle. In increasing order of mass, the charged leptons are the electron (e), muon (μ), and tau (τ)[2]. Each of these has the exact same charge, $-1e$, given in units of the electric charge.

The neutral leptons are collectively known as neutrinos. Each charged lepton has a neutrino counterpart, aptly named the electron neutrino (ν_{e}), muon neutrino (ν_{μ}), and tau neutrino (ν_{τ}). They are nearly massless and are known to oscillate between the three flavors[2]. However, this behavior is only relevant at distances much larger than the length scales of collider physics. The properties of the leptons are summarized in Table 1.1.

The quarks are also electrically charged, though they do not have integer values of electric

Generation	Particle	Mass [GeV]	Charge [e]
1	u	0.216×10^{-3}	$+\frac{2}{3}$
1	d	0.4170×10^{-3}	$-\frac{1}{3}$
2	c	1.273	$+\frac{2}{3}$
2	s	0.094	$-\frac{1}{3}$
3	t	172.5	$+\frac{2}{3}$
3	b	4.183	$-\frac{1}{3}$

Table 1.2: *Summary of the properties of the six fundamental quarks*[2].

charge. Instead, they are broken into two types: up-type quarks, with charge $+\frac{2}{3}$, and down-type quarks, with charge $-\frac{1}{3}$. The up-type quarks include the up (u), charm (c), and top (t) quarks while the down-type quarks include the down (d), strange (s), and bottom (b) quarks. Like the leptons, the flavors of quarks are ordered in increasing mass. The quarks differ furthermore from the leptons in that they carry color charge which allows them to form bound states known as hadrons. The names, symbols, and properties of the quarks are summarized in Table 1.2.

Each fermion has an associated antiparticle, denoted as ℓ^+, $\bar{\mathrm{q}}$, and $\bar{\nu}$, respectively, for charged leptons, quarks, and neutrinos. The properties of antiparticles are identical to their corresponding particles, with the exception of electric charge which is inverted.

The bosons are not organized into generations like the fermions. The photon (γ) and the gluon (g) are electrically neutral and massless. The Z and Higgs (H) bosons are electrically neutral but massive. Lastly, the W boson is electrically charged ($\pm 1e$) and has a mass. The Higgs boson has spin-0 while the other bosons all have spin-1. The gluon is the only boson which carries color charge, like the quarks. The names, symbols, and properties of the bosons are summarized in Table 1.3.

Particle	Symbol	Mass [GeV]	EM Charge [e]	Spin
Photon	γ	0	0	1
Gluon	g	0	0	1
W	$W^{\pm}$	80.4	±1	1
Z	Z	91.2	0	1
Higgs	H	125.2	0	0

Table 1.3: *Summary of the properties of the SM bosons*[2].

1.2 Fundamental forces

The physical properties of the elementary particles described in Section 1.1 determine how each type of particle interacts with itself and others. Each of the spin-1 bosons is a gauge boson for an associated physical force and each force has an associated quantity. This section describes each of these forces qualitatively while Section 1.3 provides a more formal description.

1.2.1 Electromagnetic force

All electrically charged particles interact with each other by exchanging photons. As a result, they experience an attraction or repulsion known as the electromagnetic (EM) force. This allows for atoms to form, as negatively charged electrons bind to the positively charged nucleus. Molecules are also the result of ions experiencing attractive forces due to their opposite charges. At an even larger scale, the action of a compass needle pointing towards the magnetic north pole is the result of this force as well.

1.2.2 Weak force

The weak force results from the exchange of the $W^{\pm}$ and Z bosons. These particles have a very large mass, 80.4 GeV and 91.1 GeV, respectively[2]. Each fermion participates in these interactions, however due to the large mass of the mediating particles the effective range of the force is very small. At low energies, with respect to the EM force, the weak force is 10^{-5} as powerful[3]. Radioactive β-decay and nuclear fission are the result of the weak nuclear force.

1.2.3 Strong force

Particles carrying color charge interact via the exchange of gluons. The resulting attraction is 1000 times more powerful than the EM force and is responsible for binding quarks together into hadrons such as protons and neutrons[3]. Color charge differs from electric charge in that it comes in three varieties: red, green, and blue. Quarks carry one of these color charges while antiquarks carry one of the anticolor charges: antired, antigreen, and antiblue. The gluon carries one color and one anticolor which yields eight different varieties. As a consequence of this, gluons are also able to interact with themselves.

The strong force does not decrease with distance like the EM force. Instead, the attraction between two color charged particles increases as they travel apart. Eventually enough energy is stored in the bond that a quark-antiquark pair is created. This phenomenon is known as "color confinement" and makes it impossible to isolate a single quark. Only color neutral bound states of multiple quarks can exist.

1.2.4 Gravitational force

In the macroscopic world the effects of gravity are easily observable. Planets, stars, and galaxies are all held together through the mutual attraction of mass. However on the scale of individual particles with very small masses the force of gravity is a factor 10^{-34} as strong as the EM force[3]. As a result the effects of gravity in high energy physics experiments are negligible.

1.3 Mathematical formalism

The SM can be expressed as a Lagrangian density, $\mathcal{L}$, of the form[4]

$$\mathcal{L} = \mathcal{L}\left(\phi_r(x), \partial_\alpha \phi_r(x)\right), \tag{1.1}$$

where $\phi_r(x)$ are vector fields and $\partial_\alpha \phi_r(x)$ are the 4-dimensional first derivatives of these fields, $\frac{\partial \phi(x)}{\partial x^\alpha}$. In order to derive the equations of motion for these fields the action integral over an arbitrary region of space, $S(\Omega)$, is required to be stationary[4], as shown in Eq. 1.2 and Eq. 1.3:

$$S(\Omega) = \int_\Omega \mathcal{L}\left(\phi_r(x), \partial_\alpha \phi_r(x)\right) d^4x, \tag{1.2}$$

$$\delta S(\Omega) = 0, \tag{1.3}$$

where δ represents a variation in the action, $S(\Omega)$. If variations of these fields can be described as

$$\phi_r(x) \rightarrow \phi_r(x) + \delta\phi_r(x), \tag{1.4}$$

then Eq. 1.3 reduces to

$$\frac{\partial \mathcal{L}}{\partial \phi_r(x)} - \frac{\partial}{\partial x^\alpha}\left(\frac{\partial \mathcal{L}}{\partial\left(\partial_\alpha \phi_r(x)\right)}\right) = 0, \tag{1.5}$$

which is the familiar Euler-Lagrange equation. By defining momenta conjugate to the fields and enforcing the canonical commutation relations this theory is quantized. When the resulting equation of motion of a field describes a particle with specific properties, that particle can be interpreted physically as an excitation, or quanta, of the original field. As an example, consider the Lagrangian density,

$$\mathcal{L} = \frac{1}{2}\left(\partial_\alpha \phi^\dagger(x)\partial^\alpha \phi(x) - \mu^2 \phi^\dagger(x)\phi(x)\right), \tag{1.6}$$

where μ is some constant with units of inverse length. The resulting equations of motion are found to be:

$$\begin{aligned} \left(\partial_\alpha \partial^\alpha + \mu^2\right)\phi(x) = 0, \\ \left(\partial_\alpha \partial^\alpha + \mu^2\right)\phi^\dagger(x) = 0. \end{aligned} \tag{1.7}$$

These describe a spin-0, electromagnetically charged particle-antiparticle pair with mass μ. The charged pi-mesons ($\pi^\pm$), known as pions, are an example of this type of particle. Accordingly, charged pions can be interpreted as the quanta of the complex scalar field, $\phi(x)$, with mass 139.57 GeV.

1.3.1 Gauge invariance

Noether's theorem states that, if a given Lagrangian density is invariant under a set of transformations, then there is an associated conserved quantity. Equation 1.4 can be rewritten as a global phase transformation in the context of the complex, scalar fields $\phi(x)$ and $\phi^\dagger(x)$:

$$\begin{cases} \phi_r(x) \to \phi_r'(x) = \phi_r(x)e^{i\epsilon} \\ \phi_r^\dagger(x) \to \phi_r^{\dagger\prime}(x) = \phi_r^\dagger(x)e^{-i\epsilon} \end{cases}, \tag{1.8}$$

where ϵ is some real parameter. The Lagrangian density described in Eq. 1.6 is invariant under this global phase transformation described in Eq. 1.8. The resulting conserved quantity, $J^\mu(x)$, is shown in Eq 1.9 and can be interpreted as electromagnetic charge.

$$J^\alpha(x) = \phi^\dagger(x)\partial^\alpha\phi(x) - \phi(x)\partial^\alpha\phi^\dagger(x) \tag{1.9}$$

$$\partial_\alpha J^\mu(x) = 0 \tag{1.10}$$

For other theories a different form of phase invariance must be imposed. For example, the equation of motion for free spin-$\frac{1}{2}$ charged particles such as electrons and quarks can be derived from the following Dirac Lagrangian density:

$$\mathcal{L}_0 = \bar{\psi}(x)\left[i\gamma^\mu \frac{\partial}{\partial x^\mu} - m\right]\psi(x), \tag{1.11}$$

where γ^μ are the 4×4 Dirac matrices and m is the rest mass of the particles. These particles are known to interact with the electromagnetic field. Therefore, an interaction term, $\mathcal{L}_\text{I}$, is added to the Lagrangian, yielding

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_0 + \mathcal{L}_\text{I} \\ &= \bar{\psi}(x)\left[i\gamma^\mu \frac{\partial}{\partial x^\mu} - m\right]\psi(x) + e\bar{\psi}(x)\gamma^\mu\psi(x)A_\mu(x), \end{aligned} \tag{1.12}$$

where $A_\mu(x)$ is a vector potential. However, the addition of this interaction term breaks the global phase invariance. To resolve this, the transformation of the vector potential,

$$A_\mu(x) \to A'_\mu(x) = A_\mu(x) + \partial_\mu f(x), \tag{1.13}$$

is coupled with a local phase transformation of the fields,

$$\begin{cases} \phi_r(x) \to \phi'_r(x) = \phi_r(x)e^{if(x)} \\ \phi_r^\dagger(x) \to \phi_r^{\dagger\prime}(x) = \phi_r^\dagger(x)e^{-if(x)} \end{cases} . \tag{1.14}$$

The local phase transformation described by Eq. 1.14 is a generalization of the global phase transformation described by Eq. 1.8. The phase now includes an arbitrary function of the space-time coordinate, $f(x)$. This is also known as a gauge transformation and theories such as the Lagrangian in Eq. 1.12 which are invariant under these are known as gauge theories. This specific gauge transformation corresponds to a $U(1)$ symmetry and the quanta associated with this vector potential is the massless photon, which can be considered a $U(1)$ gauge boson.

1.3.2 Electroweak theory

Despite the successful formulation of quantum electrodynamics (QED) as a fully renormalized $U(1)$ gauge theory, there still existed many phenomenon which could not be explained by this model. Specifically, the weak interactions of radioactive decay appeared to be mediated by massive, charged bosons. Efforts began to associate these interactions with a symmetry similar to $U(1)$ and QED. In 1953, Chen Ning Yang and Robert Mills proposed a theory based on $SU(2)$ symmetry. This yields three massless vector bosons with associated charges $+1, -1,$ and 0. However, this is insufficient because the short range of the weak interactions necessitates a large mass for the associated mediating particle.

The solution, proposed simultaneously by Robert Brout, François Englert, and Peter Higgs in 1964, is found by introducing a complex isodoublet scalar field with four independent real components. The Lagrangian for this field has certain symmetries and is characterized by a parameter, μ^2. When $\mu^2 > 0$ there is a single, unique ground state of the system and all of the quanta of the field have identical, positive masses. However, when $\mu^2 < 0$, the ground state is degenerate and the system settles into one out of the infinite configurations. This is known as spontaneous symmetry breaking: the ground state of a system does not possess a symmetry that the Lagrangian does. In this context, three of the components of the original field appear as massless scalar quanta known as Goldstone bosons. While there is no physical evidence for these particles, a remedy is found by modifying the Lagrangian for the original isodoublet scalar such that it is invariant under the combined $SU(2) \otimes U(1)$ transformation[5].

Now there are four additional vector quanta, three from the $SU(2)$ and one from $U(1)$. Remarkably, under spontaneous symmetry breaking the three Goldstone bosons do not appear as particles but instead impart an additional degree of freedom in the form of linear polarization to three of the vector quanta. This linear polarization of a vector field manifests as mass. This procedure of imparting a mass on the $SU(2)$ vector bosons via the addition of a scalar field is known as the Brout-Englert-Higgs mechanism and the remaining massive boson associated with scalar field is called the Higgs boson. This procedure yields the following physical bosons: one massive neutral scalar, two massive charged vectors, one massive neutral vector, and one massless neutral vector. These are the Higgs, $W^{\pm}$, Z, and photon, respectively[5].

1.3.3 Quantum chromodynamics

The theory of strong interactions can be described by a Lagrangian density similar to QED, but with the notable difference that the quark fields possess an additional degree of freedom known as color charge. Accordingly, the relativistic quantum field theory describing the interactions of particles with color is called quantum chromodynamics (QCD).

As mentioned in Sec. 1.2.3, there are three varieties of color charge: red, green, and blue[4]. The wavefunction for a given quark can be written as a product of spin and color parts, where the color part is usually denoted $\chi^c = r, g, b$ with

$$r = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad g = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{1.15}$$

Color theory explains the absence of free quarks in nature. Only color neutral bound states can exist as free particles. This requirement allows only color-anticolor (e.g. $r\bar{r}$) bound states known as mesons, or three color (rgb or $\bar{r}\bar{g}\bar{b}$) bound states, known as baryons. Color neutral combinations of four or five quarks can be experimentally produced, but do not occur naturally[6].

The gluon field and the resulting interactions between it and the quark fields are a consequence of requiring the free quark Lagrangian density to be invariant under $SU(3)$ transformations. The QCD Lagrangian density is therefore

$$\mathcal{L}_{QCD} = \bar{\Psi}^f(x) \left[i\gamma^\alpha \partial^\alpha + i g_s \lambda_j \frac{A_j^\alpha}{2} - m_f \right] \Psi^f(x) - \mathcal{L}_G, \tag{1.16}$$

where

- $\Psi^f(x)$ is the quark field with $f = u, d, s, c, b, t$
- g_s is a real coupling constant
- λ_j are eight matrices ($j = 1, 2, \ldots, 8$) representative of the $SU(3)$ group
- A_j^α are eight gauge fields ($j = 1, 2, \ldots, 8$) called gluon fields
- m_f is the mass of the different quarks
- $\mathcal{L}_G$ are the gluon field interaction terms

The $\mathcal{L}_G$ interaction terms reveal some specific behaviors of the gluon field which distinguish it from the photon:

$$
\begin{aligned}
\mathcal{L}_G = -\frac{1}{4}F_{i\mu\nu}(x)F_i^{\mu\nu}(x) + g_s f_{ifk} A_{i\mu}(x)A_{j\nu}(x)\partial^\mu A_k^\mu(x) \\
-\frac{1}{4}g_s^2 f_{ifk} f_{ilm} A_j^\mu(x)A_k^\nu(x)A_{l\mu}(x)A_{m\nu}(x)
\end{aligned}
\tag{1.17}
$$

The first term represents eight massless spin-1 gluon fields. The second and third terms represent gluon self-interactions. These arise because the gluon field itself is color charged. This is also why the free gluon is not observed, as only color neutral states exist as free particles.

To summarize, the SM is the combination of the quantum field theories discussed in this section. It combines electroweak theory and quantum chromodynamics to provide a coherent description of the elementary particles and their interactions. It is described by a Lagrangian density which is invariant under $SU(3) \otimes SU(2) \otimes U(1)$ symmetry transformations and the quanta of the fields represent all known particles.

1.4 Limitations of the Standard Model

Since the formulation of the SM as a gauge invariant quantum field theory, it has been used to make accurate predictions of physical quantities. Despite these achievements, there are still several observed phenomena which can not be explained with the SM. A few are enumerated below:

- neutrinos are massless in the SM, however they are known to be have some non-zero mass due to the observation of oscillations between flavors,
- there is no explanation for the observed asymmetry between matter and antimatter in the universe. Only very few SM processes produce matter and anti-matter at different rates and the deviation is not significant enough to account for the near total absence of antimatter,
- ample astrophysical evidence exists for dark matter and dark energy but there is no

description in the SM for these phenomena,

- the SM lacks a mechanism to explain why the gravitational force is many orders of magnitude weaker than the other three forces.

Modern particle physics experiments seek to make precise measurements of physical processes and compare them to the SM predictions. A deviation, even very small, could be evidence of new physics. Section 2 will describe the technical aspects involved in making these precise measurements.

Chapter 2

The Experiment

Huge amounts of data must be produced in order to study physics processes which are predicted to occur only very rarely. The Large Hadron Collider (LHC) was designed to collide high energy particles at rates large enough for such rare processes to be observed. Since the beginning of data collection in 2010, the LHC has progressively increased both the energy and rate of these collisions to record-setting values. During Run-2 (2015-2018) protons were collided at a center of mass energy of 13 TeV. The protons now collide at a center of mass energy of 13.6 TeV for the ongoing Run-3 (2022-2026) with plans to increase again to 14 TeV towards the end of the decade. The LHC produces collisions at four different locations for experiments to collect data. The CMS (Compact Muon Solenoid) and ATLAS (A Toriodal LHC ApparatuS) experiments are both general-purpose detectors. Two special-purpose detectors known as ALICE (A Large Ion Collider Experiment) and LHCb (Large Hadron Collider beauty) occupy the remaining locations. Over the years, five additional experiments have been built to supplement the LHC physics program. These are much smaller than those previously mentioned and are variously positioned near the four main detectors, working in conjunction with them. This chapter will first describe the LHC in Section 2.1, followed by the CMS detector in Section 2.2

2.1 The Large Hadron Collider

The LHC is the largest component of the accelerator complex operated by the European Organization for Nuclear Physics (CERN, from the French, *Organisation européenne pour la recherche nucléaire*), shown in Fig. 2.1. It is a ring 27 km in circumference located 100 m underground in the countryside near Geneva, Switzerland. The ring consists of two high vacuum pipes inside of which beams of protons circulate in opposite directions close to the speed of light. Superconducting magnets are placed at regular intervals to steer the beams around the curves. In addition to the four locations where the beams can be made to collide there are also structures present along the ring to collimate, accelerate, and eventually remove the beams.

The protons which constitute these beams begin as the nuclei of negative hydrogen atoms. They are first accelerated by the Linear Acelerator 4 (LINAC4) to an energy of 160 MeV and then injected into the Proton Synchrotron Booster (PSB). At injection, the ions are passed through a thin carbon foil which strips the electrons away leaving just the protons. This process is known as "charge-exchange injection" and allows for minimal beam losses. While in the PSB, the protons are accelerated again to an energy of 2 GeV.

The protons then pass through the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS), where they are accelerated to 26 GeV and 450 GeV, respectively. The beams are subsequently transferred to the LHC where they are accelerated to their final collision energy. Currently, this is 6.8 TeV per beam for proton-proton collisions. Once the beams reach the desired energy they are made to collide at the four different points around the ring, providing data for the experiments.

The LHC uses 1232 dipole magnets to steer the beams around the curves and 392 quadrupole magnets to focus the beams. These magnets are superconducting which requires them to be kept at very low temperatures with liquid Helium. Additional magnets are used to further focus the beams just prior to collisions.

In addition to the collision energy, the LHC is designed to maximize luminosity, which is a measure of the collision rate. The instantaneous luminosity is defined for head-on collisions

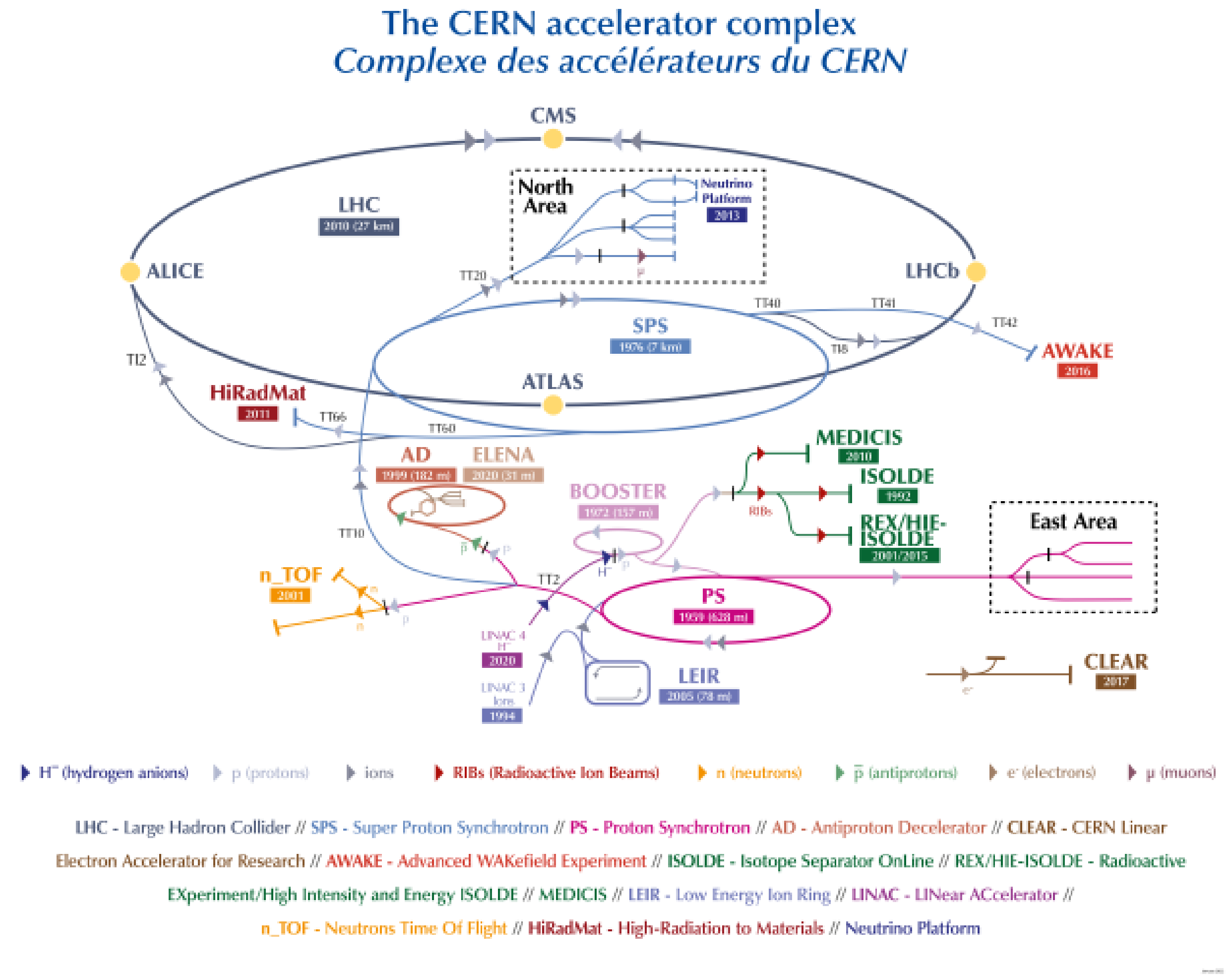

Figure 2.1: *Full diagram of the accelerator complex at CERN. The LHC is shown in dark blue, with the four main experiments located at various locations. In actuality, ALICE and LHCb are much closer to the ATLAS detector than shown here. The CMS detector's position is accurately shown directly opposite ATLAS*[7].

of two beams with Gaussian profiles as[3]:

$$\mathcal{L} = f \frac{n_1 n_2}{4\pi\sigma_x\sigma_y}, \tag{2.1}$$

where f is the frequency of collisions measured in Hz, n_1 and n_2 are the number of colliding particles, and σ_x and σ_y, respectively, are the root-mean-square horizontal and vertical beam sizes. The two proton beams are segmented into bunches which collide every 25 ns at each of the four interaction points during normal operations. This corresponds to a frequency of

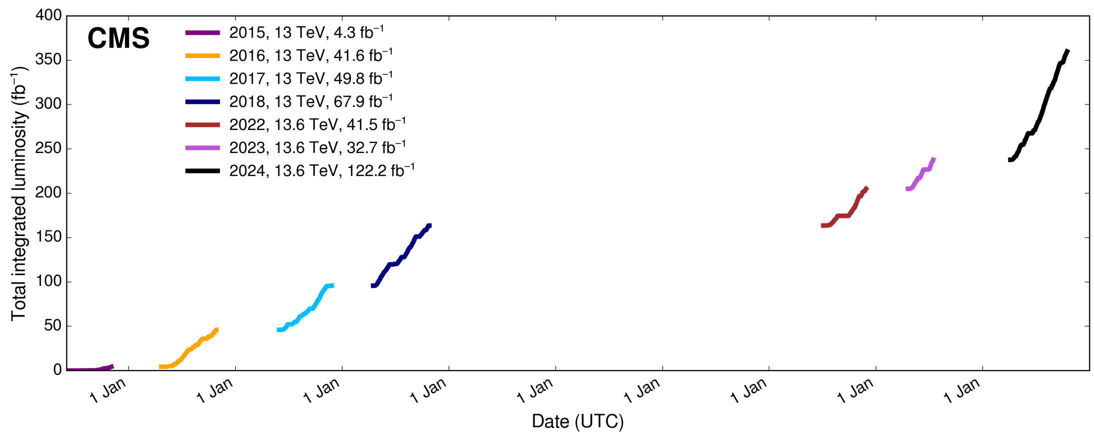

Figure 2.2: *Integrated luminosity of proton-proton collisions delivered to the CMS detector versus date. Seven years of data taking are shown in the different colors. Mechanical issues in 2022 and 2023 limited the amount of luminosity that could be delivered.*[9]

40 MHz. The nominal number of protons per bunch is 1.15×10^{11} and the LHC was designed to accommodate 2808 bunches per beam, however there are practical limitations to this[8]. The peak design luminosity of the LHC is $1.0 \times 10^{34}\,\mathrm{s}^{-1}\,\mathrm{cm}^{-2}$, however this has since been exceeded[8]. A large instantaneous luminosity increases the likelihood of a hard-scattering collision between protons. Over the first three years of Run-3, there were an average of 49 hard-scattering interactions during a given bunch crossing at the CMS detector, an increase from an average of 34 over Run-2.

The beams circulate for as long as the beam intensity remains above a certain threshold. When the time comes to discard the beams, they are steered out of the ring by magnets and into a designated location. This process is known as a "beam dump" and can also be initiated automatically by one of several safety systems. It is useful to measure the accumulated luminosity by integrating over a period of time. The integrated luminosity shown in Fig. 2.2 for the past seven years of data-taking is given in units of inverse femtobarns (fb^{-1}), where $1\,\mathrm{fb}^{-1} = 10^{-39}\,\mathrm{cm}^2$. In 2024, the average time for beams to circulate was 7.72 hours, which includes premature beam dumps. If no technical issues arose, the beams were kept in the machine until $850\,\mathrm{pb}^{-1}$ was delivered — typically about 13 hours.

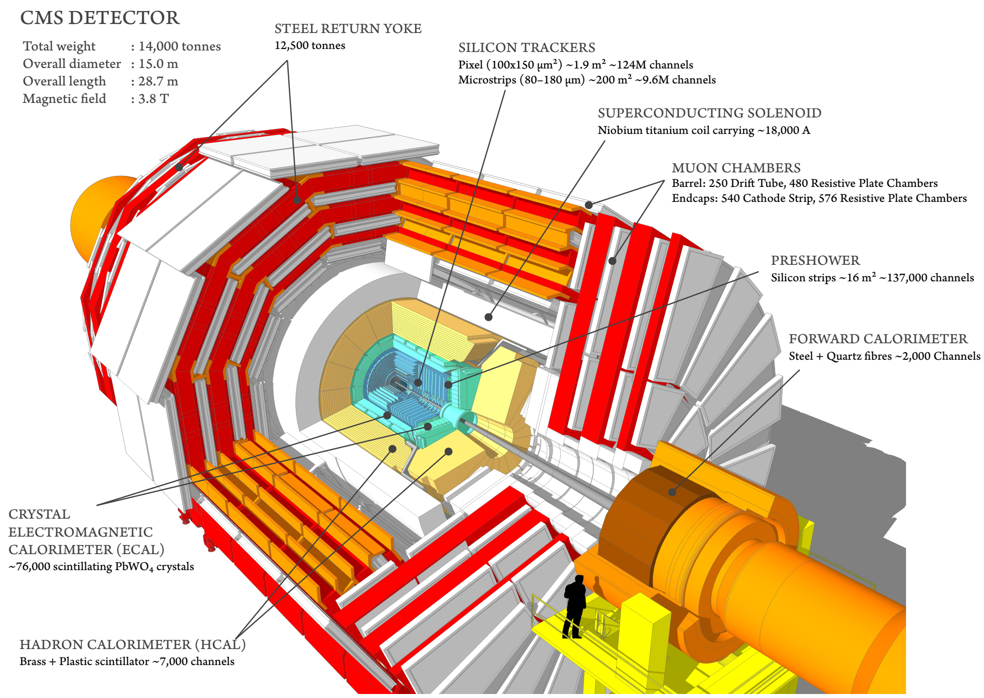

Figure 2.3: *A labeled diagram of the CMS detector with a portion cut away such that the inner layers are visible. A human is shown for scale in the bottom right.*[10]

2.2 The CMS experiment

The CMS detector is designed to accurately and efficiently observe and record the interactions produced in its volume by the proton-proton collisions at the LHC. It is 14.6 m tall, 21.6 m long, and weighs about 12,500 t. The detector is structured in concentric cylindrical layers of sub-detectors. Each one is sensitive to specific types of particles. Together they achieve hermetic coverage over most of the angular space surrounding the interaction point.

The various sub-detectors can be seen in Fig 2.3. These will be described in the following sections, along with the superconducting solenoid magnet (2.2.5) and trigger system (2.2.7). Lastly, section 2.3 describes how information from these different sub-detectors is used to reconstruct and identify particles produced in the collisions.

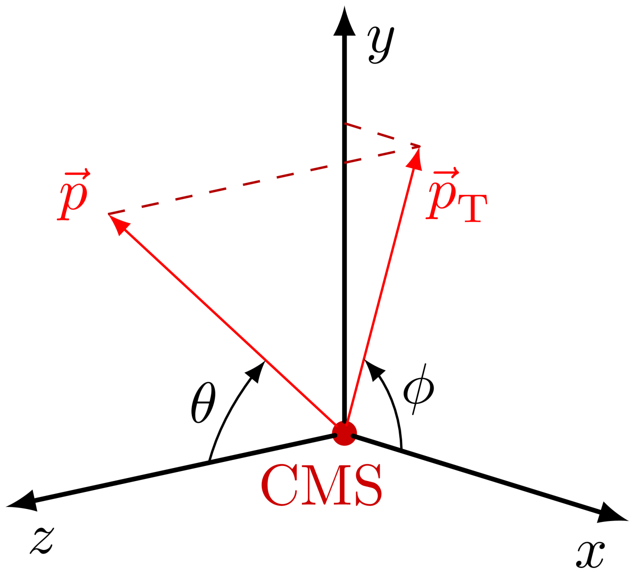

Figure 2.4: *The coordinate system used by the CMS experiment. Momentum ($\vec{p}$) and its component in the x-y plane ($\vec{p_T}$) are also shown as the two labeled red arrows*[11]

2.2.1 Detector coordinate system

The CMS experiment uses a spherical coordinate system with the x axis pointing in towards the center the LHC ring, the y axis pointing up in the air, and the z axis pointing along the beam pipe in a counterclockwise direction if viewed from above. The x-y plane is referred to as the transverse plane. These axis are shown in Fig 2.4, with the polar (θ) and azimuthal (ϕ) angles also shown.

It is useful to define the rapidity, y, as[3]:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \tag{2.2}$$

where E is the energy of a particle and p_z is the component of its momentum along the z axis. In the massless limit, when $E \gg m$, where m is the mass of a particle, this reduces to the more commonly used variable, pseudorapidity[3]:

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right). \tag{2.3}$$

Differences in the pseudorapidities of two particles, measured as $\Delta\eta$, are Lorentz invari-

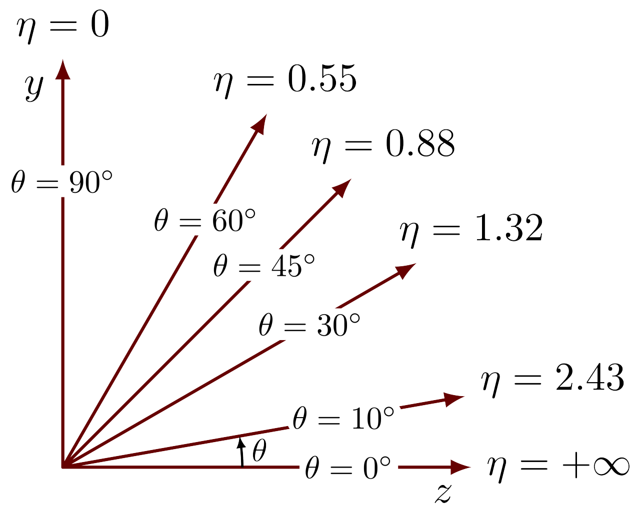

Figure 2.5: *Values of pseudorapidity for common polar angles*[11]

ant. This value can therefore be used to define the separation of two particles in the y-z plane independent of the rest frame. Additionally, the number of jets produced at the LHC is constant as a function of η. Some values of η are shown in Fig. 2.5, as an example of how it compares to common polar angles.

The variable ΔR is commonly used to define distances in $\eta - \phi$ space. For two particles i and j it is defined as

$$\Delta R_{i,j} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}. \tag{2.4}$$

2.2.2 Tracker

The tracker is the innermost layer of the CMS detector. It is cylindrical in shape and surrounds the interaction point, extending 5.8 m along the beam pipe. The beam pipe extends to a radius of 26 mm and the tracker instrumentation begins at 29 mm. The tracker receives extremely high radiation doses due to its proximity to the interactions. The particle density is also very high, especially when inside a radius of 200 mm. In order to maintain an occupancy of 1% or less, a silicon pixel detector is used in this region, with a standard pixel size of $100 \times 150\,\mu\text{m}^2$. The pixel detector consists of four barrel layers and three endcap layers[12]. As charged particles traverse these layers they deposit energy. Deposits in adjacent layers that follow a consistent trajectory are grouped together to form a track. The tracker also reconstructs secondary vertices corresponding to long-lived particles, such as τ leptons, which are important to many analyses.

Outside of 200 mm, a silicon strip tracker is used. The strips are organized into ten layers in the barrel region and twelve in the endcap region. Between a radius of 200 mm and 550 mm, the strip size is $10\,\mathrm{cm} \times 80\,\mu\mathrm{m}$. Further out, from 550 mm to 1100 mm, a larger strip size of $25\,\mathrm{cm} \times 180\,\mu\mathrm{m}$ is used. The strip sizes are adjusted to maintain low occupancy levels[13].

The pixel and silicon strip detector combine to provide efficient and precise measurements of the particle trajectories up to $|\eta| = 2.5$. This information is essential to the reconstruction algorithms described in section 2.3.

2.2.3 Electromagnetic Calorimeter (ECAL)

The ECAL is next to the Tracker, moving outwards radially from the interaction point. It is designed to measure the energy of electrons and photons using scintillating lead tungstate ($PbWO_4$) crystals. The crystals emit blue-green light as electrons and photons interact electromagnetically with the nuclei[13]. Two different types of photodetectors are used to collect and amplify this light signal: avalanche photodiodes (APD) in the barrel region ($|\eta| < 1.479$) and vacuum phototriodes (VPT) in the endcap region ($1.479 < |\eta| < 3.0$).

Both the crystals and the photodiodes perform worse as a function of increasing temperature. ECAL therefore features a robust, water-based cooling system capable of maintaining a temperature of $(18.00 \pm 0.05)\,^{\circ}\mathrm{C}$[13].

Electrons with energy above 10 MeV primarily radiate photons as they interact with the electric field of a nucleus in a process known as *bremsstrahlung*. Above similar energies, photons primarily convert to an electron and a positron in the presence of atomic nucleus in a process known as pair production. When these processes occur sequentially and repetitively they form a cascade of energy known as electromagnetic (EM) shower.

Several properties of $PbWO_4$ make it an ideal material for the ECAL. First, it is radiation-

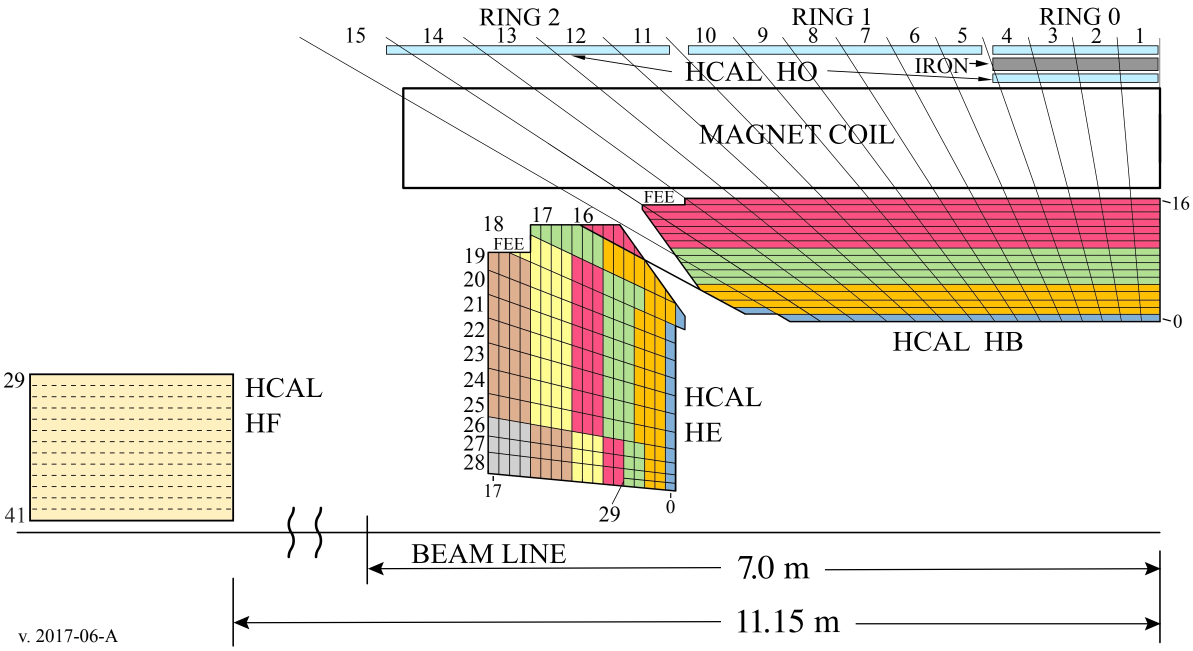

Figure 2.6: *The relative sizes and locations of the HB, HE, HO, and HF detectors*[14]

hard, allowing it to operate without replacement for many years. Additionally, the scintillation time is of the same order of magnitude as the time between bunch crossings. The Molière radius defines an area in which 90% of the EM shower energy is contained. This is 2.2 cm for $PbWO_4$, which decreases the likelihood that showers from different particles will overlap[13]. Lead tungstate also features a short radiation length if 0.89 cm[13], defined as average distance over which the energy of an electron is reduced by a factor of $1/e$. Given the thickness of ECAL, it contains more than 25 radiation lengths[13], which is enough to prevent energy from leaking in to other parts of the detector.

2.2.4 Hadronic Calorimeter (HCAL)

Continuing radially outwards from the ECAL, the next subdetector is the HCAL. It is essential to the physics program of CMS given the large numbers of jets expected from proton-proton collisions. This sub-detector, shown in Figure 2.6, consists of four distinct elements: barrel (HB), endcap (HE), outer (HO), and forward (HF).

The HB and HE sections operate as sampling calorimeters and consist of alternating layers of brass and plastic scintillator material. Particles traversing the volume interact with the dense brass and deposit energy in the scintillating tiles. The resulting light is then carried to silicon photomultipiers (SiPMs) by a combination of wavelength shifting and clear fibers. Both HB and HE consists of a 9 mm thick initial layer of Bicron BC408, followed by 16 layers of Kuraray SCSN81. These are 3.7 mm thick, except for the furthest out layer in HB, which is 9 mm. HB occupies a region between a radius of 1.77 m and 2.95 m and extends to $|\eta| < 1.3$. HE extends the coverage further out to $|\eta| < 3.0$[13].

The HF consists of two identical detectors located 11.15 m on either side of the interaction point. These receive a significantly larger particle flux than the other HCAL elements, as they are positioned 12.5 cm from the beam pipe and extend roughly from $3.0 < |\eta| < 5.0$. There are no other CMS detector components between HF and the interaction point because this is outside of the tracker acceptance region.

In order to operate in this environment through the LHC data taking periods, HF consists of radiation-hard quartz fibers within a steel absorber. These fibers capture Cherenkov light produced by shower of charged particles and transmit it via air-core light guides to photomultiplier tubes (PMTs). Two different lengths of fibers are used in HF. One half of the fibers are known as Long (L) and extend the full 165 cm longitudinal length of the detector. The other half of the fibers are accordingly known as Short (S) and are positioned 22 cm back from the front of the detector[13]. The L fibers are sensitive to both EM and hadronic particles, whereas the S fibers will only receive between about 20% and 40% of the EM shower, increasing with the particle energy[15]. Partitioning the fibers in this way allows for some level of differentiation between the types of particles depositing energy in the detector.

HO is a small layer situated outside of the magnet (Section 2.2.5) in the barrel region which uses hybrid photodiodes (HPDs) to collect and amplify the light signals. It is necessary to provide adequate stopping power for all hadronic showers[13].

2.2.5 Magnet

A principal feature of the CMS detector is the superconducting solenoid magnet. Surrounding the tracker, ECAL, and HCAL, it has an inner diameter of 5.9 m and is 12.5 m long[13]. The solenoid itself consists of four layers of niobium-titanium (Nb-Ti) wire and is contained within a vacuum cryostat which maintains the low temperatures necessary for superconductivity. When the magnet is powered during physics data-taking, a homogeneous 3.8 T field is generated inside the solenoid volume along the direction of the beam pipe.

Charged particles moving in this magnetic field will follow a curved trajectory due to the Lorentz force. The magnitude of the component of this force perpendicular to the field, F_T, is

$$F_T = q \cdot v_T \cdot B, \tag{2.5}$$

where q is charge of the particle, v_T is the component of velocity in the transverse plane, and B is the magnitude of the magnetic field. This force is a centripetal force and so Eq. 2.5 can be rewritten as

$$m\frac{v_T^2}{r} = q \cdot v_T \cdot B, \tag{2.6}$$

where m is the mass of the particle and r is the radius of the curve of the particles path. Equation 2.6 can then be rearranged as

$$m \cdot v_T = q \cdot B \cdot r, \tag{2.7}$$

where $m \cdot v_T = p_{\mathrm{T}}$. Equation 2.7 shows how the strong magnetic field allows for the momentum of charged particles to be measured from the curvature of the particles trajectory.

Supporting the detector and also serving as a return path for the magnetic field flux is an iron yoke, weighing 10,000 t. Figure 2.7 shows the relative positioning of the solenoid and the yoke with the magnetic field strength and field lines overlaid. With the exception of muons and neutrinos, most particles that were not absorbed by the calorimeters will be

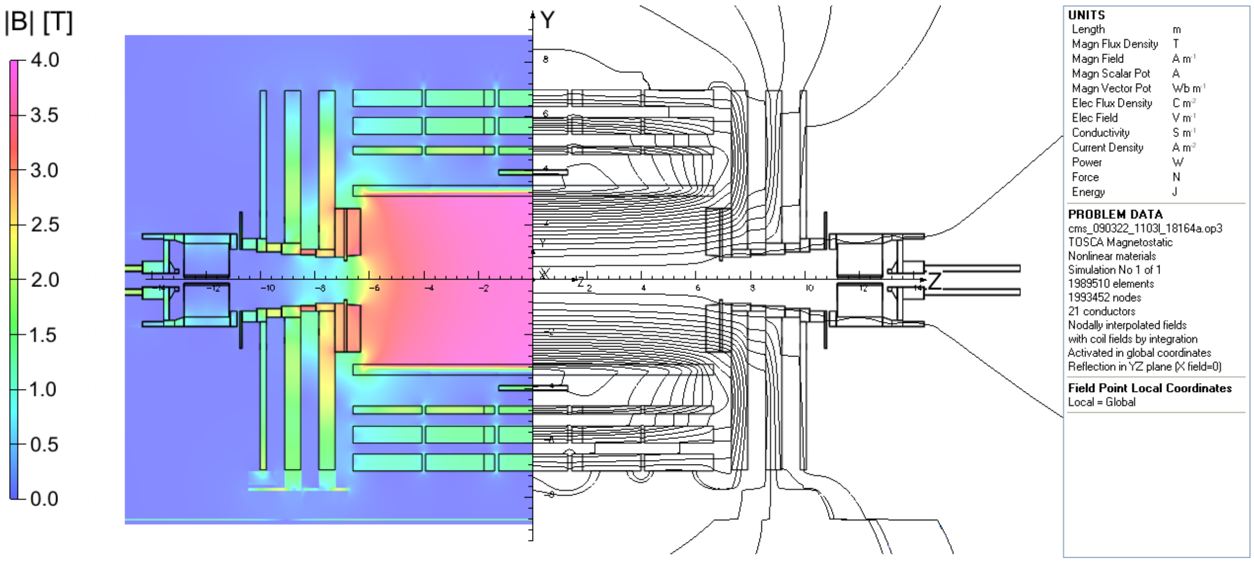

Figure 2.7: *Diagram of the CMS magnet system, showing the magnitude, $|B|$, of the magnetic field (left) and the field lines (right)*[16]*. The region of space colored red is the interior of the solenoid. The light green rectangles radially outside of the solenoid are the pieces of the iron yoke.*

stopped by the iron yoke.

2.2.6 Muon System

The muon system is the outermost section of the CMS detector and uses different types of gas ionization detectors to measure the position and momentum of muons. Unlike electrons, muons do not suffer significant radiative losses in the ECAL and therefore provide excellent mass resolution as final state particles. An efficient and accurate muon system is therefore necessary for high precision analyses.

The barrel region, $|\eta| < 1.2$, and endcap region, $0.9 < |\eta| < 2.4$, are instrumented, respectively, with drift tubes (DTs) and cathode strip chambers (CSCs). Both of these technologies provide excellent spatial resolution for energy deposits. The region $|\eta| < 1.6$ is instrumented with resistive plate chambers (RPCs). These do not have as good a spatial resolution as DTs or CSCs but provide a sharp p_{T} threshold for the trigger systems[13]. A gas electron multiplier (GEM) detector was added to the endcap region of the muon system prior to the start of data taking in 2022. This detector operates on the principle of gas

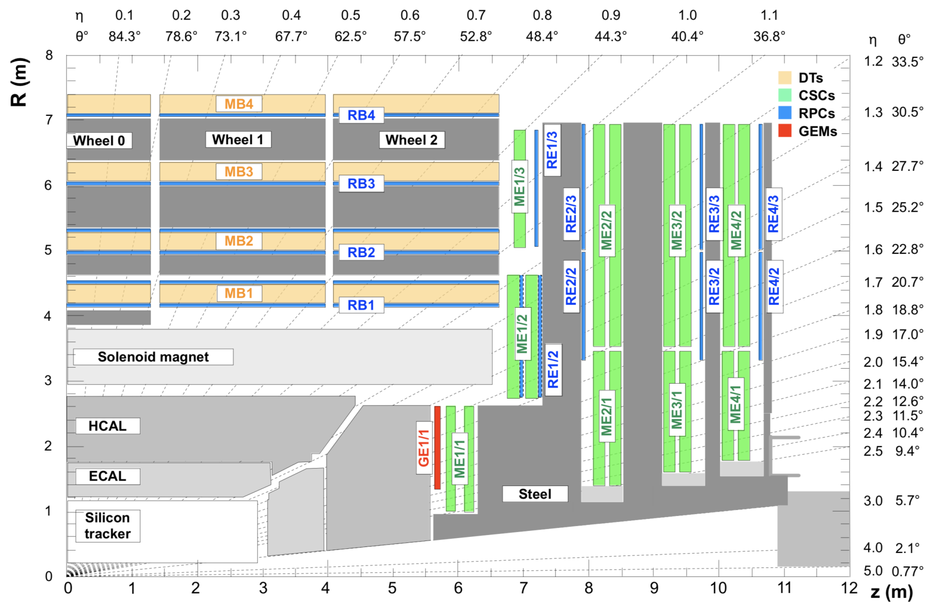

Figure 2.8: *A labeled diagram of CMS muon system. The newly added GEM detector can be seen colored red in the endcap section. Also labeled are the other detector components, colored different shades of grey*[17].

ionization, similar to the other three technologies mentioned. It provides coverage in the region $1.6 < |\eta| < 2.2$ where there are no RPCs and is designed to bolster the performance of the trigger system and muon reconstruction[17].

Figure 2.8 shows the relative positioning of these various muon detection systems amid the magnet yoke. The other detector components mentioned in the previous sections can be seen as well.

2.2.7 Trigger System

The LHC produces collisions at a rate of 40 MHz, however due to practical limitations of memory storage it is not possible to record event data from all of these collisions. Instead, CMS uses a two-tiered system to efficiently and accurately identify only those events which are of high interest to physicists.

Level-1 Trigger

The first tier in the CMS trigger system is implemented entirely in hardware and is known as the Level-1 (L1) Trigger. The L1 Trigger uses only information from the muon system and the calorimeters (ECAL and HCAL) to reduce the event rate to 100 kHz. The decision to keep or discard the event is made in about 4 μs. This latency is much larger than the 25 ns between bunch crossings. As a result, the trigger logic is pipelined, which allows for multiple events to be processed concurrently[18].

This pipeline is structured hierarchically to organize information from increasingly large regions of the detector. This architecture is shown in Fig. 2.9. Local triggers within the calorimeters and muon systems construct particle candidates, known as trigger primitives. Next, a set of regional triggers covering the full calorimeter space combine and sort these objects, before passing them on to either a Global Calorimeter or Global Muon Trigger. These triggers determine, respectively, the highest-rank calorimeter or muon object and pass this information on to the Global Trigger, which makes the final decision, known as Level-1 Accept (L1A)[13].

High-Level Trigger

Events which pass the L1 Trigger are further evaluated by the second tier known as the High-Level Trigger (HLT). The HLT is implemented entirely in software and uses the full readout from the detector to further reduce the event rate from 100 kHz to 1 kHz. A computing farm runs reconstruction algorithms which combine information from all sub-detectors to reconstruct more complex objects. On average, the HLT takes 300 ms to make a decision which determines if the raw data will be saved for complete offline reconstruction[20].

A single HLT trigger is a series of specific requirements based on the available event information. These requirements have a physical motivation and exploit known characteristics of physics processes. The triggers are grouped together based on the general type of objects they target. Several hundred HLT triggers are active during data taking allowing for many different types of events to be studied.

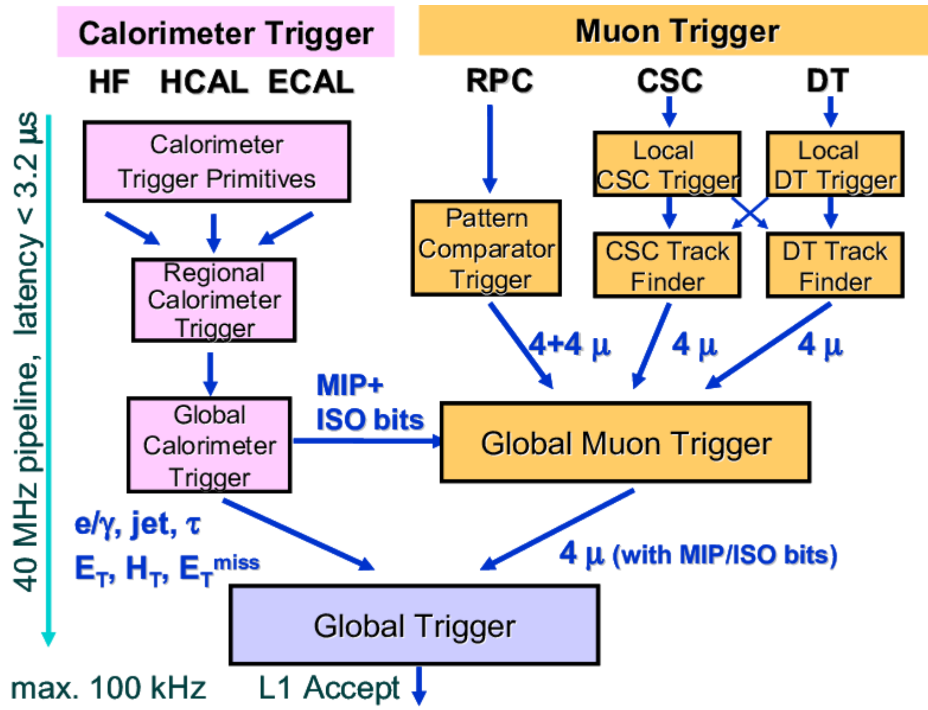

Figure 2.9: *The architecture and data flow of the synchronous and pipelined L1 Trigger system*[19]

2.3 Object reconstruction

Only events which the trigger system accepts undergo the global event reconstruction. The following sections describe the process by which energy deposits in the detector are reconstructed into more complex objects for use in analyses.

2.3.1 Track reconstruction

Charged particles traversing the CMS tracker leave energy deposits which are used to reconstruct particle trajectories. This is done by iteratively applying a method known as the Combinatorial Track Finder (CTF)[21]. On each repetition, the following four steps are performed:

- **Track seeding:** starting with the inner layers of the tracker, a "seed" is generated from two or three hits in successive layers. More than 85% of seeds are obtained from a series of three pixel hits known as a pixel triplet. Seeds can also be formed from pairs of pixel hits along with a vertex or a strip hit. These methods help to identify forward or displaced particles. Lastly, a seed can be made from a pair of hits in the strip detector to account for particles which might not leave any hits in the pixel detector. Seeds define a rough first estimate of the track parameters and associated uncertainties.

- **Track finding:** track candidates are built from seeds using an approach based on the Kalman filter (KF) method[22]. An extrapolation of the trajectory based on the seed parameters is used to search for additional compatible hits. These hits are added to the track candidate and the parameters are updated accordingly. The extrapolation takes into account the effects of the detector material on the trajectory and its uncertainty.

- **Track fitting:** once all of the hits are associated with a track, it is refitted and smoothed with a KF. Hits deemed to be outliers are removed and the track is refit until only hits consistent with the track parameters remain. The fitting procedure ensures that the track can be accurately extrapolated both inwards to the interaction region and outwards to the calorimeters and muon system[22].

- **Track selection:** a significant number of the track candidates found at this point in the procedure are not actually associated with a charged particle that existed. Therefore, the final step is to apply a variety of quality requirements. In general, the requirements are more strict on track candidates which pass through fewer layers. These requirements can also be adjusted to accommodate the evolution of data-taking conditions.

To reduce the combinatorial complexity, hits associated with high-quality tracks are removed after each iteration[22]. These methods are able to reconstruct tracks with $|\eta| < 2.5$, $p_{\mathrm{T}} > 0.1\,\mathrm{GeV}$, and a displacement from the beam line of $< 60\,\mathrm{cm}$. The efficiencies in the barrel ($|\eta| < 1.4$) and endcap ($1.4 < |\eta| < 2.5$) regions were measured to be 94% and 85%, respectively[22].

2.3.2 Vertex positioning

The position of all interaction vertices in a given event is determined from the trajectories of reconstructed tracks. Tracks are selected for this calculation only if they originate close to the center of the interaction point. This region, known as the beam spot, is the location in the CMS detector where the LHC beams collide and is measured from an average over many events[22]. Next, tracks are clustered based on the z-coordinate of their closest approach to the center of the beam spot. These clusters become the candidate vertices. To reject false vertices, only those which have at least two tracks that are incompatible with other vertices are kept. Lastly, an adaptive vertex fitter is used to estimate the parameters for each vertex[23]. This takes into account each constituent track's likelihood to be associated with a given vertex.

The reconstructed vertex with the largest sum of p_{T} is denoted as the primary vertex (PV) and all others are considered pileup. This vertex has the highest likelihood of being associated with a noteworthy physics process.

2.3.3 Particle flow

Tracking and vertex information is combined with clusters of energy in the calorimeters via an algorithm known as particle-flow (PF)[24]. This yields a comprehensive list of reconstructed final-state particles. The PF algorithm proceeds by identifying objects in the following order: muons, electrons, photons, and lastly charged hadrons. The information associated with a particle that has already been reconstructed is ignored for future steps. The particle candidates reconstructed by the PF algorithm serve as inputs for additional algorithms which reconstruct higher-level objects, such as jets and tau leptons.

2.3.4 Muons

First, muons are reconstructed using information from the tracker and muon systems. As minimally ionizing particles (MIPs), they do not deposit significant amounts of energy in the calorimeters. The hits left behind through ionization of the traversed material are connected to form tracks in both the tracker and the muon systems. Then, these separate tracks can be extrapolated and matched together to form muon candidates.

Energy from hadronic showers can occasionally reach the muon system and be mistakenly reconstructed as a muon candidate. To prevent these candidates from being misidentified as muons, an isolation requirement is applied on additional tracks and energy deposits within a distance $\Delta R < 0.3$ to the muon trajectory. The combined p_{T} of the nearby tracks and E_{T} (transverse energy) of the energy deposits is required to be no more than 10% of the muon p_{T}.

2.3.5 Electrons and photons

Similar identification algorithms are used for electrons and photons because they are almost always produced in conjunction with one another through either the radiation of bremsstrahlung photons or production of electron-positron pairs. To account for the abrupt changes in track trajectories that occur when a substantial amount of energy is lost, the tracks used to seed electron candidates are fit with a Gaussian-sum filter (GSF)[25]. These secondary particles will also deposit energy in ECAL, which must be properly associated with the original particle in order to make accurate energy measurements. Therefore, neighboring clusters of energy in ECAL are grouped together to form"superclusters".

Electron candidates are formed from a GSF track matched with a supercluster that has fewer than three additional matched tracks. Photon candidates are formed from superclusters with $E_{\mathrm{T}} > 10\,\mathrm{GeV}$ that are not matched to any GSF track. To prevent the misidentifcation of hadrons as electrons or photons, the HCAL energy within $\Delta R < 0.15$ of the supercluster is required to be no more than 10% of the supercluster energy.

2.3.6 Hadrons

Finally, the PF algorithm uses the remaining information to reconstruct charged and neutral hadrons. Energy clusters in HCAL not matched to any track are considered neutral hadron candidates. One neutral hadron, the neutral pion (π^0), decays almost immediately to a pair of non-isolated photons which will not be reconstructed according to the methods described in Section 2.3.5. As a result, ECAL energy clusters of more than 400 MeV with no matching track are grouped into pairs of photons. If the total energy of the photon pair is greater than 1.5 GeV, then these are kept as photon candidates and assumed to be the result of π^0 decay. The remaining energy deposits and tracks are linked to form charged hadron candidates. The PF algorithm does not attempt to identify the specific type of charged or neutral hadron.

2.3.7 Jets

Large amounts of quarks and gluons are produced in the LHC collisions. These particles do not leave a discrete, single signature in the detector. Instead, by nature of their interactions, they quickly fragment and hardronize into a multitude of charged and neutral particles. As these ensuing particles interact with the detector and decay, the energy of the initial quark or gluon is dispersed across a local region of the detector. By collecting and associating this collimated spray of energetic hadrons, it is possible to reconstruct an object known as a jet[26].

The CMS experiment uses an algorithm known as anti-k_T[27] to cluster together nearby particles in $\eta - \phi$ space. The following two distances are defined for this algorithm

$$d_{ij} = \min\left(\frac{1}{p_{\mathrm{T},i}^2}, \frac{1}{p_{\mathrm{T},j}^2}\right) \cdot \frac{\Delta R_{ij}^2}{R^2}, \tag{2.8}$$

$$d_{iB} = \frac{1}{p_{\mathrm{T},i}^2}, \tag{2.9}$$

where R is a distance parameter and $p_{\mathrm{T},i}^2$ and $p_{\mathrm{T},j}^2$ are the transverse momenta of two different particles. The algorithm executes the following steps in succession[26]:

1. The distances d_{ij} and d_{iB} are calculated considering all PF particles
2. The minimum between d_{ij} and d_{iB} is identified
3. If this minimum is d_{ij}, combine the particles i and j as a 4-vector sum and return to

the first step

4. Otherwise, denote particle i as a jet, remove it from further consideration, and return to the first step

This process continues until all particles have been clustered into jets. The result of this algorithm is a number of roughly conical jets centered around high-p_{T} particles.

A standard R value of 0.4 is used to reconstruct quark (u, d, s, c, b) and gluon jets, whereas 0.8 is used for boosted massive bosons (W, Z, H) and top quark jets[28]. Jets reconstructed with these values are known as AK4 and AK8 jets, respectively. Two competing factors have influenced these values. First, the distance parameter needs to be large enough to include all of the energy associated with the jet. A value of R that is too small will fail to capture some particles, thereby miscalculating the energy of the original particle that created the jet. Second, the distance parameter needs to be small enough to limit the contributions from background radiation. The value of 0.4 and 0.8 balance these considerations for their respective jets and are used by both the CMS and ATLAS collaborations in jet reconstruction.

Some of the particles reconstructed by the PF algorithm will necessarily be from pileup regardless of the choice of distance parameter. Therefore, two techniques are employed to mitigate the effects of pileup on jet reconstruction: charged-hadron subtraction (CHS) and pileup per particle identification (PUPPI).

The CHS technique identifies charged hadrons which are not associated with the primary vertex of the event. These are then removed from the list of PF particles prior to jet reconstruction[24]. Since CHS uses charged tracks, it can only be applied within the acceptance of the tracker, $|\eta| < 2.5$. Additionally, it does not remove any contributions from neutral pileup particles, which account for one third of the jet energy[24].

The PUPPI algorithm assigns each particle a weight from 0 to 1, which represents its' likelihood to originate from the primary vertex of the event. In an ideal scenario, all pileup particles would receive a weight of 0 and those coming from the hard scattering would receive a weight of 1. This weight is calculated by comparing the particles under consideration to those known to come from pileup[29]. The weight is then used to rescale the 4-momenta of particles, which can now be used in the jet clustering algorithms. This was found to be more effective than applying a simple cut on the weight because the rescaled values correct jet shapes and kinematics[30].

2.3.8 Hadronic tau leptons

Tau leptons are unstable and decay very quickly. They are therefore reconstructed and identified using the signatures of their visible decay products. When the visible decay product is either an electron or a muon, the tau decay is deemed leptonic and the standard reconstructions algorithms for electrons and muons, described in Sections 2.3.4 and 2.3.5, will reconstruct the corresponding lepton. Otherwise, the visible decay products consist of hadrons which are identified as described in Section 2.3.6. The branching fraction ($\mathcal{B}$) of these hadronic tau (τ_{h}) decays is 64.8%, compared to 17.8% and 17.4% for electrons and muons, respectively. About 95% of τ_{h} decays result in either one or three charged pions in addition to zero, one, or two neutral pions. The various dominant hadronic decay modes are listed in Table 2.1.

Decay mode	Branching Fraction [%]
$\tau^- \to \mathrm{h}^-\nu_\tau$	11.5
$\tau^- \to \mathrm{h}^-\pi^0\nu_\tau$	25.9
$\tau^- \to \mathrm{h}^-\pi^0\pi^0\nu_\tau$	9.5
$\tau^- \to \mathrm{h}^-\mathrm{h}^+\mathrm{h}^-\nu_\tau$	9.8
$\tau^- \to \mathrm{h}^-\mathrm{h}^+\mathrm{h}^-\pi^0\nu_\tau$	4.8
Other	3.3
Total	64.8

Table 2.1: *Branching fractions for dominant hadronic decay modes of the τ lepton. The symbols "h^+" and "h^-" denote a generic charged hadron, usually a pion or a Kaon.*

A dedicated algorithm known as hadrons-plus-strips (HPS)[31;32] was developed to reconstruct tau candidates. This algorithm takes into account the presence of an isolated charged track and the specific energy distribution that results from a $\pi^0 \to \gamma\gamma$ decay. The HPS algorithm is seeded with jets, reconstructed as described in Section 2.3.7, with $p_{\mathrm{T}} > 14\,\mathrm{GeV}$

$|\eta| < 2.5$.

Using the PF electron and photon constituents of the jet, the HPS algorithm iteratively constructs "strips", which will represent possible π^0 candidates. First, the highest p_T electron or photon not already associated with a strip is selected as the starting point. The initial η and ϕ values for the strip are set using this object. Next, additional electrons and photons with $p_\mathrm{T} > 0.5\,\mathrm{GeV}$ located within an $\eta - \phi$ isolation region are added to the strip. Only electrons or photons which have not already been added to a strip are considered. After each particle is added the η and ϕ values of the strip are recalculated using a p_T weighted sum of the constituents. As soon as there are no more eligible electrons or photons in the vicinity of the strip, the process repeats starting with the next highest p_T electron or photon not already associated with a strip. Only strips with $p_\mathrm{T} > 2.5\,\mathrm{GeV}$ are kept as π^0 candidates.

The isolation region for the strip is defined in the η and ϕ directions by the following functions:[33]

$$f(p_\mathrm{T}) = 0.20 \cdot p_\mathrm{T}^{-0.66}, \tag{2.10}$$

$$\Delta\eta = f\left(p_\mathrm{T}^{e/\gamma}\right) + f\left(p_\mathrm{T}^{\mathrm{strip}}\right), \tag{2.11}$$

$$g(p_\mathrm{T}) = 0.35 \cdot p_\mathrm{T}^{-0.71}, \tag{2.12}$$

$$\Delta\phi = g\left(p_\mathrm{T}^{e/\gamma}\right) + g\left(p_\mathrm{T}^{\mathrm{strip}}\right), \tag{2.13}$$

where $p_\mathrm{T}^{e/\gamma}$ is the p_T of the electron or photon jet constituent that is being added to the strip, and $p_\mathrm{T}^{\mathrm{strip}}$ is the p_T of the strip itself.

These functions define a rectangular region in $\eta-\phi$ space which is larger in the ϕ direction to account for the bending of the trajectories of charged particles in the field of the CMS magnet[33]. In the context of tau decays, the behavior of these functions reflects how the distribution of energy from π^0 decays varies depending on the p_T of the parent τ_h candidate. At high p_T the electrons and photons will be dispersed across a smaller area, therefore the size of the window is reduced. This helps to avoid including objects not associated with the decay of the π^0. Conversely, at low p_T the electrons and photons will be more spread out and so the window size is increased to make sure all of the decay products are included.

Once strips have been formed, the HPS algorithm considers the PF charged hadron constituents of the jet which are not associated with a strip. The tracks corresponding to these charged hadrons must have $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$ and $d_{xy} < 0.1\,\mathrm{cm}$. This ensures, respectively, that they are of good quality and are compatible with the primary vertex of the event. Hadronic tau candidates are formed by assembling different combinations of these charged hadrons and π^0 candidates.

The HPS algorithm reconstructs τ_{h} candidates in one of seven decay modes (DM). Five of these correspond to the physical decays listed in Table 2.1: $\mathrm{h}^{\pm}$, $\mathrm{h}^{\pm}\pi^0$, $\mathrm{h}^{\pm}\pi^0\pi^0$, $\mathrm{h}^{\pm}\mathrm{h}^{\mp}\mathrm{h}^{\pm}$, and $\mathrm{h}^{\pm}\mathrm{h}^{\mp}\mathrm{h}^{\pm}\pi^0$. The other two are non-physical due to charge conservation and are included to account for track inefficiencies: $\mathrm{h}^{\pm}\mathrm{h}^{\pm/\mp}$ and $\mathrm{h}^{\pm}\mathrm{h}^{\pm/\mp}\pi^0$. For convenience, the DMs are numbered with integers according to the following formula

$$\mathrm{DM} = 5 \times \big(\mathrm{N}(\mathrm{h}^{\pm}) - 1\big) + \mathrm{N}(\pi^0), \tag{2.14}$$

where $\mathrm{N}(\mathrm{h}^{\pm})$ is the number of charged hadrons and $\mathrm{N}(\pi^0)$ is the number of π^0's present in the decay. The hadronic decays can also be referred to by the number of charged pions, or "prongs", that were reconstructed. For example, the $\mathrm{h}^{\pm}$, $\mathrm{h}^{\pm}\pi^0$, and $\mathrm{h}^{\pm}\pi^0\pi^0$ decays are all called "one-prong" or "single-prong" decays.

Four of these decays involve intermediate meson resonances with known masses: $\rho(770)$ for $\mathrm{h}^{\pm}\pi^0$, $\mathrm{a}_1(1260)$ for both $\mathrm{h}^{\pm}\pi^0\pi^0$ and $\mathrm{h}^{\pm}\mathrm{h}^{\mp}\mathrm{h}^{\pm}$, and $\rho(1450)$ for $\mathrm{h}^{\pm}\mathrm{h}^{\mp}\mathrm{h}^{\pm}\pi^0$. In each of these cases, the τ_{h} candidate is only kept if the mass is within a few GeV of the mass of the corresponding resonance. Additionally, all τ_{h} candidates, with the exception of those reconstructed in DMs 5 or 6, are are required to have charge ± 1. The algorithm selects a single τ_{h} candidate for each PF jet. As a result, if there are multiple valid candidates the τ_{h} candidate with the highest p_{T} is chosen.

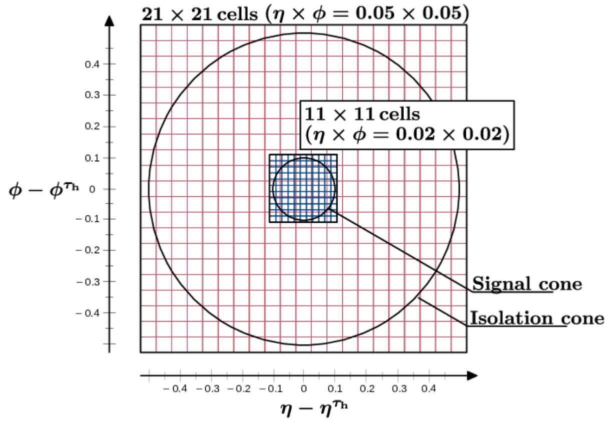

Figure 2.10: *The layout of the signal and isolation grids for the DeepTau network. These contain the particle-level input variables and are processed using convolutional layers*[34]

Identification using neural network

The DeepTau neural network[34] was developed to improve the discrimination of reconstructed τ_h decays originating from genuine tau leptons against τ_h candidates originating from (quark or gluon) jets, electrons, and muons. The network can be categorized as a deep, multiclass, convolutional network. Information from particles in the spatial vicinity of the τ_h decay is used to assign four discriminating scores corresponding to the probability that it is a jet, an electron, a muon, or a genuine τ_h.

The convolutional layers use a 3×3 filter to reduce the two grids of $\eta - \phi$ space, shown in Fig 2.10, to a single output cell. The smaller grid contains the signal cone of the τ_h candidate, which is defined by the HPS algorithm. The finer spacing of this inner grid is able to resolve the dense core of high-p_T quark and gluon jets. The grids are overlapping, meaning information from the inner grid is also considered by the outer grid. In each cell, variables from seven different types of particles are considered: PF electrons, muons, photons, charged hadrons, and neutral hadrons, as well as electrons and muons reconstructed with dedicated algorithms. Additionally, the network takes into account many high level variables which are processed separately from the convolutional layers.

The loss function is a weighted sum of tau cross entropy (H_τ) and multiple focal loss components ($\bar{F}_{cmb}$ and $\bar{F}_\alpha$)[34]. It is defined as

$$L\left(y^{true}, y^{pred}\right) = k_\tau H_\tau + \left(k_e + k_\mu + k_j\right) \cdot \bar{F}_{cmb} + k_F \sum_{\alpha \in \{e,\mu,j\}} k_\alpha \cdot \bar{F}_\alpha, \tag{2.15}$$

where k represents various scale factors, y^{true} is the generator truth value, and y^{pred} is the neural network prediction value with $0 \leq y^{pred} \leq 1$. A discriminator, $D_{e,\mu,j}$, is calculated from these prediction values and is defined as

$$D_{e,\mu,j} = \frac{y_\tau}{(y_\tau + y_{e,\mu,j})}, \tag{2.16}$$

where y_τ is the probability the candidate is a tau and $y_{e,\mu,j}$ is the probability that the candidate is an electron, muon, or jet.

Chapter 3

Analysis

The properties of the W boson have been thoroughly studied by nearly every high energy particle physics experiment since its discovery at CERN in 1983[35;36]. Despite this scrutiny, no exclusive decays of the W boson to hadrons have been observed. Analyses have typically focused on the leptonic decay modes which occur roughly 33% of the time for the W boson. The current world standard values for the branching fractions of the dominant decay modes of the W boson are given in Table 3.1[2].

All other decays of the W boson are to hadrons. These occur roughly twice as often as the leptonic decays, however there are a large number of possible unique final states due to the complex hadronization process. A study of simulated W boson decays to hadrons found that there would be an average of 30 final state particles that would be visible to a detector[37]. It is not feasible to accurately reconstruct this many final state particles in the challenging hadronic environment of proton-proton collisions, however the lower multiplicity states are still statistically significant given the abundance of W boson events produced at

Mode	Branching Fraction (%)
eν	10.71 ± 0.16
$\mu\nu$	10.63 ± 0.15
$\tau\nu$	11.38 ± 0.21
hadrons	67.41 ± 0.27

Table 3.1: *Branching fractions for dominant decay modes of the W boson*

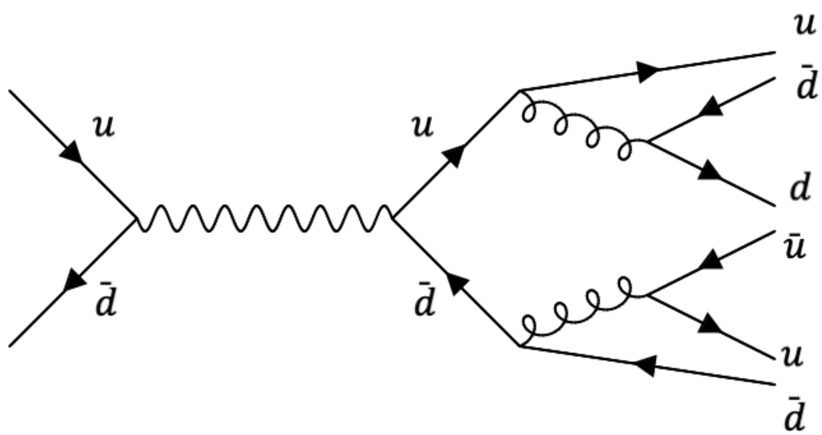

Figure 3.1: *Example Feynman diagram for the decay of a W boson to three charged pions*

the LHC.

This analysis focuses on the three body decay of the W boson to a triplet of charged pions, which provides a clean signature of high p_{T} tracks. An example of a leading order diagram for this decay is shown in Fig. 3.1.

$$W^{\pm} \rightarrow \pi^{\pm}\pi^{\mp}\pi^{\pm} \tag{3.1}$$

This decay presents the best opportunity to probe the exclusive hadronic decays of the W boson for potential new physics. In addition, the methods developed in this analysis will be of use as we continue to explore rare processes at the LHC and beyond.

While there exists no exact theoretical prediction for the $W^{\pm} \rightarrow \pi^{\pm}\pi^{\pm}\pi^{\mp}$, it is expected to be of the same order of magnitude as $Z \rightarrow \pi^{\pm}\pi^{\pm}\pi^{0}$.[37;38] A previous search for this decay placed the first upper limit on the branching fraction at 1.01×10^{-6} at 95% confidence level[39] using $77\,\mathrm{fb}^{-1}$ of proton-proton collision data collected during 2016 and 2017 by the CMS detector at $\sqrt{s} = 13\mathrm{TeV}$.

This note reports a search for $W^{\pm} \rightarrow \pi^{\pm}\pi^{\pm}\pi^{\mp}$ decays from W bosons produced inclusively at two different center-of-mass energies. The proton-proton collision data was collected by the CMS detector and corresponds to integrated luminosities of $137.6\,\mathrm{fb}^{-1}$ at $\sqrt{s} = 13\,\mathrm{TeV}$

Data Taking Era	**Integrated Luminosity [fb^{-1}]**
`2016preVFP`	19.5
`2016postVFP`	16.8
`2017`	41.5
`2018`	59.8
`2022preEE`	7.98
`2022postEE`	26.7
`2023preBPix`	17.8
`2023postBPix`	9.45

Table 3.2: *Integrated luminosity for each data taking period*

and 61.9 fb^{-1} at $\sqrt{s}$ =13.6 TeV.

This analysis builds upon the experience of previous studies and exploits advancements in reconstruction and identification algorithms, specifically using the DeepTau version 2.1 (v2.1) and DeepTau version 2.5 (v2p5) neural networks[40]. Further improvements include a more realistic next-to-next-to-leading order model of the signal process, a more robust background estimation method using data, and the use of machine learning tools to improve the sensitivity to signal. A larger dataset is used to exploit the increase in statistics.

3.1 Simulated samples and data

3.1.1 Data

This search analyzes 137.1 fb^{-1} of proton-proton collision data gathered by the CMS Collaboration from 2016 to 2018 at the LHC at $\sqrt{s}$ =13 TeV as well as 61.9 fb^{-1} gathered in 2022 and 2023 at $\sqrt{s}$ =13.6 TeV. The luminosity per data taking period is summarized in Table 3.2.

For each of the data taking periods in Table 3.2, only events which have been certified for physics analyses were used. This means that every subdetector was operating within specific performance limits. CMS aggregates the events which have been collected into different datasets based on the which trigger caused the event to be recorded. One of these datasets includes events which satisfied the requirements for any of the triggers targeting hadronic tau candidates. As described in Sec. 2.3.8, charged pions are used in the reconstruction of tau candidates. Given that there are no triggers dedicated to charged pions, the tau dataset will provide the most events matching a final state of three charged pions.

3.1.2 Triggers

In the Run 3 data taking eras, the following trigger was used for both data and simulation:

- `HLT_DoubleMediumDeepTauPFTauHPS35_L2NN_eta2p1`

The following trigger was only available in the 2023 data taking eras and was used in addition to the one mentioned above to increase sensitivity towards the W to three pions decay channel. This trigger has lower p_{T} thresholds than the other ditau triggers and requires both tau candidates to be reconstructed in a single-prong decay mode. There is also a requirement that the ditau mass be between 5 and 80 GeV.

- `HLT_DoubleMediumDeepTauPFTauHPS30_L2NN_eta2p1_OneProng_M5to80`

In 2018 the triggers used for data were:

- `HLT_DoubleMediumChargedIsoPFTau40_Trk1_TightID_eta2p1_Reg`
- `HLT_DoubleTightChargedIsoPFTau35_Trk1_TightID_eta2p1_Reg`
- `HLT_DoubleTightChargedIsoPFTau40_Trk1_eta2p1_Reg`
- `HLT_DoubleMediumChargedIsoPFTauHPS35_Trk1_eta2p1_Reg`

For 2018 simulation, only the following trigger was used:

- `HLT_DoubleMediumChargedIsoPFTauHPS35_Trk1_eta2p1_Reg`

For 2017, the triggers used for data and simulation were:

- `HLT_DoubleMediumChargedIsoPFTau40_Trk1_TightID_eta2p1_Reg`
- `HLT_DoubleTightChargedIsoPFTau35_Trk1_TightID_eta2p1_Reg`

- HLT_DoubleTightChargedIsoPFTau40_Trk1_eta2p1_Reg

For 2016, the triggers used for data and simulation were:

- HLT_DoubleMediumIsoPFTau35_Trk1_eta2p1_Reg
- HLT_DoubleMediumCombinedIsoPFTau35_Trk1_eta2p1_Reg

Scale factors are applied to simulation samples to correct for the difference in trigger performance between data and simulation. These scale factors also account for the change in HLT paths during 2018 data-taking. These scale factors depend on the p_T and decay mode of the tau candidate and are determined by a dedicated group for the recommended p_T ranges of tau candidates which are 5 GeV above the trigger thresholds.

3.1.3 Signal Simulations

The kinematic requirements from the ditau trigger limit the selection of two of the three pions to only those with $p_T > 35$. In order for the decay products of the W boson to have this large transverse momentum the initial W boson must have been produced with some initial momentum, or boost. This momentum is balanced by the presence of an initial state jet which must be modeled as when when producing a simulation of a signal event.

The event generators capable of producing next-to-leading order (NLO) and next-to-next-to-leading order (NNLO) events are not capable of performing the decay to three pions that we require in this analysis. To solve this, we produce a full signal event in three steps: first, we generate a prompt W boson at NNLO and decay only to the initial decay products, in this case a muon and a muon neutrino. These are all saved to a file in Les Houches (LHE) format. This ensures the W boson is produced with accurate properties. Next, we manually change the decay products in the LHE file to three pions, ensuring that charge is conserved. Finally, we use PYTHIA8 to shower these pions and simulate the full event. A more detailed description of the signal generation is provided in Appendix 6

A comparison of the p_T distributions for LO and NNLO signals is shown in Fig. 3.2, normalized to one. The NNLO signal produces fewer events with low p_T due to the presence

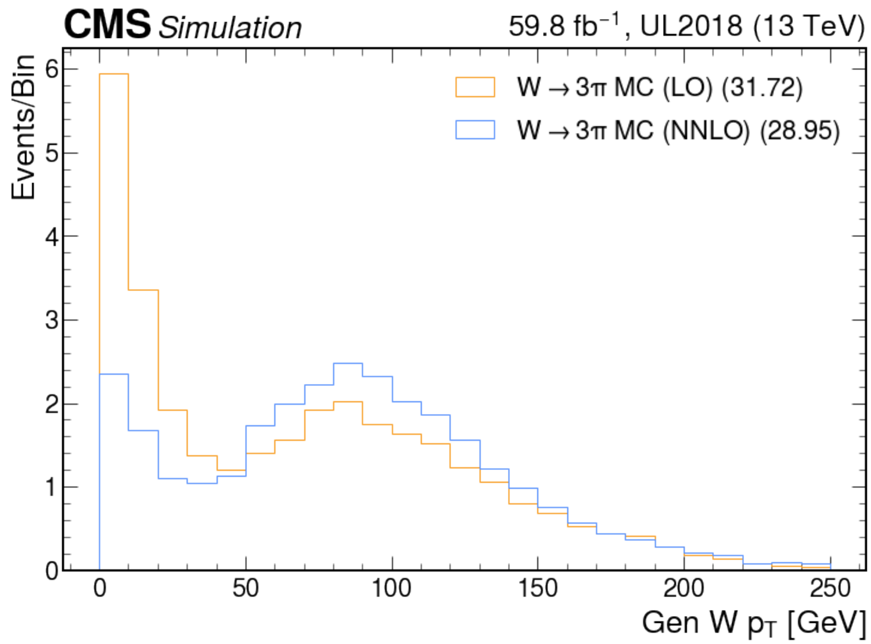

Figure 3.2: *Comparison of W boson p_T distributions in LO (orange) and NNLO (blue) signal samples*

of initial state jets capable of providing the recoil for the W boson. Ultimately, we relied on this NNLO signal sample as it provides the most accurate event description available.

3.1.4 Background simulations

Other than the QCD background, all other processes were simulated using Monte Carlo processes. All of the simulated background samples were produced through the official CMS MC central production. Like in our signal generation, `PYTHIA` is used with the CP5 tune to simulate the hadronization and showering of particles. Both Drell-Yan and $t\bar{t}$ events are modeled at next-to-leading order by `MADGRAPH` and `POWHEG`, respectively . To simulate the CMS detector's response to these events, `Geant4` is used.

3.2 Objects and event reconstruction

3.2.1 Charged pion reconstruction

Our event reconstruction begins with the reconstruction of specific physics objects by the particle-flow (PF) algorithm[24]. This algorithm combines information from every sub-detector to reconstruct charged and neutral hadrons, photons, electrons, and muons. These are the building blocks used to reconstruct more complex objects. Charged pions from the decay of a W boson can be found using two of these more complex objects: hadronic tau leptons and isolated tracks.

Tau candidates

The HPS algorithm can be viewed at a basic level as a charged pion reconstruction algorithm. Tau candidates are reconstructed by the HPS algorithm from charged and neutral pions into specific decay modes, as described in Sec. 2.3.8. The pions from the prompt decay of a W boson will be isolated and therefore match the signature for the single-prong hadronic tau decay with zero neutral pions present (DM0).

Figure 3.3 shows the decay mode for taus in the signal sample that were matched at generator level to a pion coming from a W boson and those that were not matched. The most abundant source of taus matched at generator level are those classified as DM0. This agrees with physical intuition because we expect only three isolated charged pions in our final state. Any additional neutral pions would likely be from pileup activity. In addition to the kinematic requirements applied by the HPS algorithm and described in Sec. 2.3.8, tau candidates are required to have $p_{\mathrm{T}} > 18\,\mathrm{GeV}$ and $|\eta| < 2.3$.

Furthermore, tau candidates are discriminated against electrons, muons, and jets by the DeepTau neural network. If DM0 tau candidates are used as pion candidates, this discriminating score can be used separate our signal process from background processes. This is especially useful for the QCD multijet background which is expected to dominate for this decay.

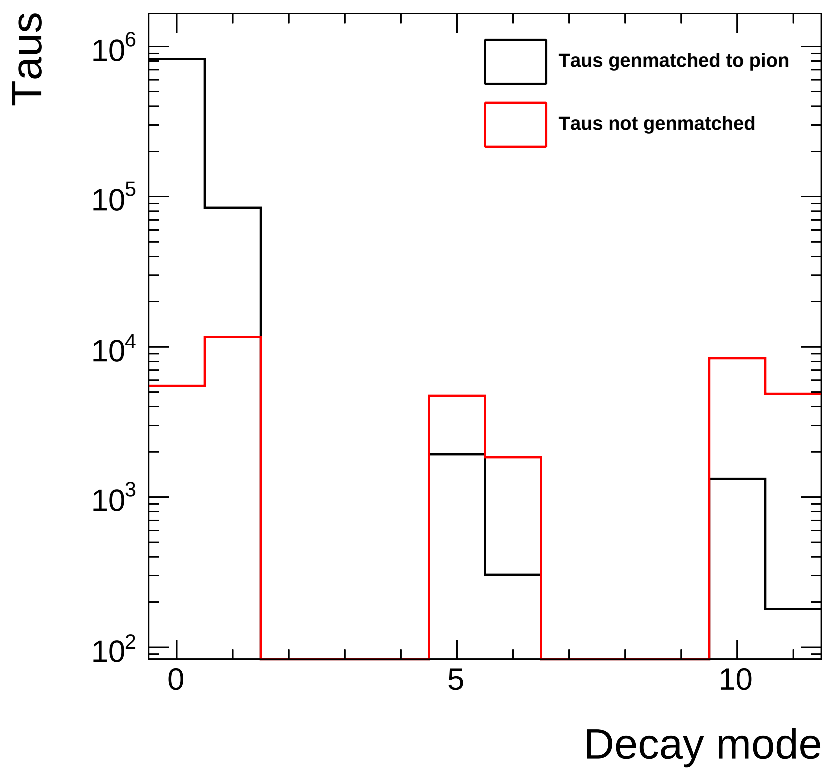

Figure 3.3: *Decay mode of Tau objects in the signal sample*

Several working points (WPs) are developed for specific efficiency values of this numerical discriminator. For the discriminator against jets and electrons eight WPs are available, ranging from “VVVLoose” to “VVTight”. For the discriminator against muons four WPs are available, ranging from “VLoose” to “Tight”. Figure 3.4 shows the distribution of these WPs for events in the signal sample. Only tau objects passing the loosest working point versus electrons, muons, and jets are stored for offline analysis.

From isolated track objects

Charged pions from the decay of a W boson will also be reconstructed as a more general class of objects called isolated tracks. Charged pions are reconstructed as isolated track objects if they satisfy the following requirements:

- $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $|\eta| < 2.5$,
- for tracks with $p_{\mathrm{T}} < 15\,\mathrm{GeV}$, the transverse ($d_{xy}$) and longitudinal ($d_z$) impact parameters are required to be less than 0.2 and 0.1, respectively,
- for tracks with $p_{\mathrm{T}} < 25\,\mathrm{GeV}$, the sum of the p_{T} of the charged hadrons in a cone of $\Delta R < 0.3$ around the track is required to be $< 5\,\mathrm{GeV}$,
- for tracks with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$, the sum of the p_{T} of charged hadrons in a cone of $\Delta R < 0.3$ around the track must be $< 0.2 \times p_{\mathrm{T}}$.

The variables measured for isolated track objects are not nearly as robust as those for tau candidates. Other than basic kinematic properties (p_{T}, η, ϕ), the most useful property is the relative particle flow isolation with $\Delta\beta$ pileup subtractions. The absolute and relative isolation are defined in Eq.3.2 and Eq.3.3, respectively, as[24]:

$$I_{absolute} = \sum_{\substack{charged \\ |d_z|<0.2}} p_{\mathrm{T}} + \max\Big(0, \sum p_{\mathrm{T}}^{h^0} + \sum p_{\mathrm{T}}^{\gamma} - \Delta\beta \sum_{\substack{charged \\ |d_z|>0.2}} p_{\mathrm{T}}\Big), \tag{3.2}$$

$$I_{relative} = \frac{I_{absolute}}{p_{\mathrm{T}}}, \tag{3.3}$$

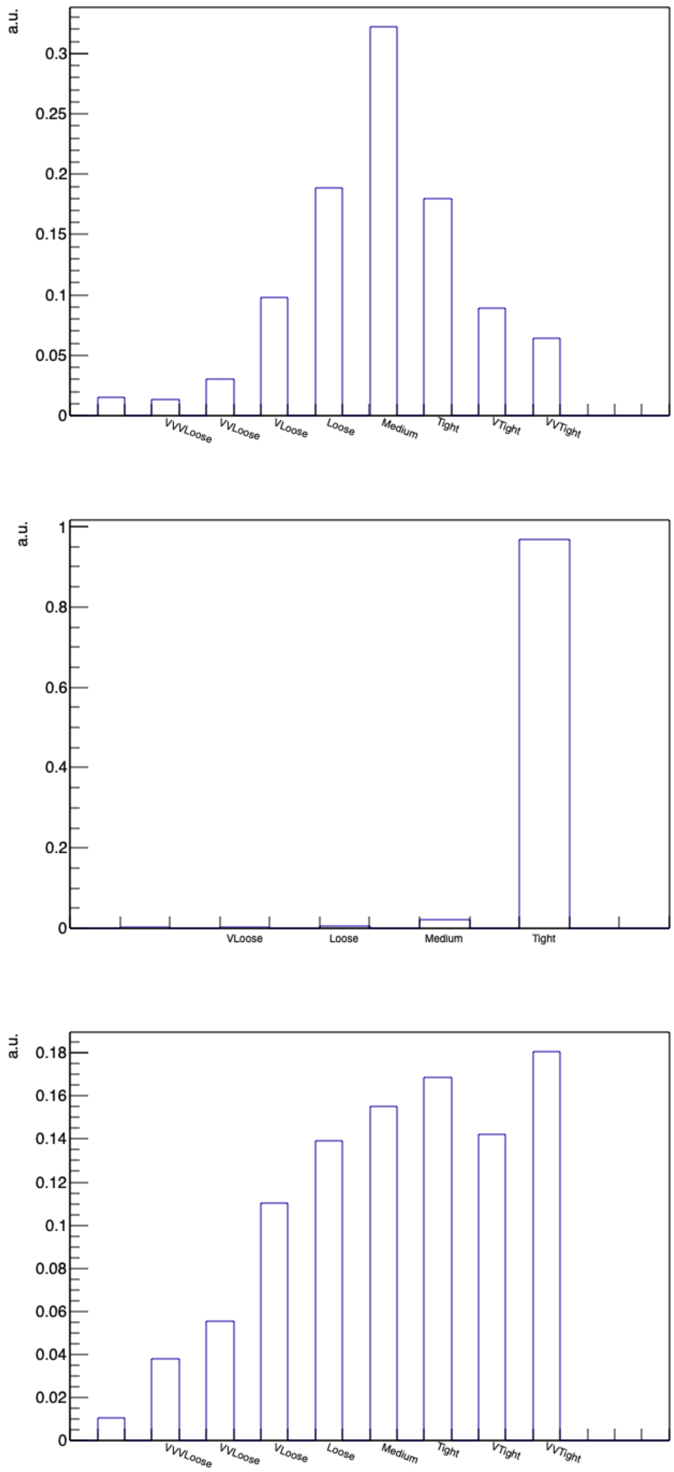

Figure 3.4: *Normalized distribution of DeepTauv2p1 discriminator WPs versus electrons (top), muons (middle), and jets (bottom) in signal sample before any selection*

where $p_{\mathrm{T}}^{h^0}$ is the p_{T} from neutral hadrons, p_{T}^{γ} is the p_{T} from photons, and $\Delta\beta$ is an estimate of the predicted contributions from pileup. These which are subtracted from the contributions from neutral particles. Simulations of proton-proton collisions showed that neutral particles carry roughly half the energy of charged particles so a value of $\Delta\beta = 0.5$ is used to represent that fraction.

This variable is rudimentary compared to the HPS decay mode and DeepTau discriminants which are calculated for tau candidates. The potential advantage from using isolated track objects is due to the 8 GeV lower p_{T} threshold. Figure 3.5 shows the p_{T} of the generated pions in events which satisfied the ditau trigger requirements. The lowest p_{T} pion distribution peaks at much lower values than the other two pions. Many of these pions would not be reconstructed as either isolated tracks or tau candidates, however the isolated track collection would make a larger portion of the distribution accessible. Figure 3.6 shows the p_{T} range available for both tau candidate and isolated track object reconstruction overlaid on the generated level p_{T} distribution for the lowest p_{T} pion. In the following section these considerations are studied empirically to determine which source of charged pions maximizes the signal yield and efficiency.

3.2.2 Pion triplet reconstruction

With these two sources of charged pions in mind, there are several different ways to arrive at a final state of three charged pions. These could be narrowed down to four scenarios when considering the limitations from triggers. The only triggers related to isolated tracks are exclusively for the calibration and commissioning of the Hadronic barrel and endcap calorimeters.

On the other hand, there are a wide variety of tau triggers available for use. The p_{T} thresholds on the single tau triggers are much too high to be useful for this analysis. Ditau triggers, however, do exist with low enough p_{T} thresholds for our objects to pass. In the 2016, 2017, 2018, and 2022 data taking years the ditau trigger with the lowest p_{T} thresholds required at least two hadronic tau candidates with $p_{\mathrm{T}} > 35$ GeV. A special trigger was

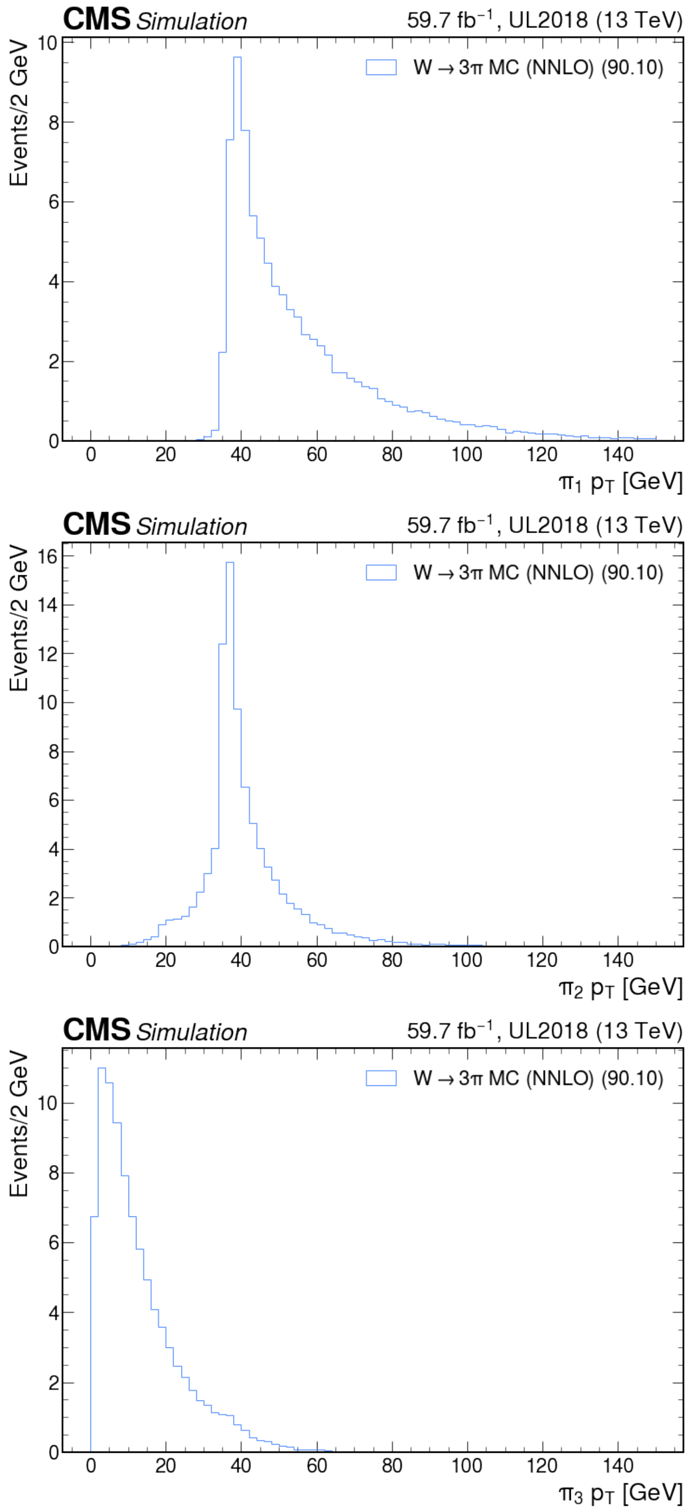

Figure 3.5: *Generated p_T of the highest (top), second highest (middle) and lowest (bottom) p_T pion after a ditau trigger is applied. The trigger requires at least two hadronic tau candidates with $p_T > 35\,\mathrm{GeV}$.*

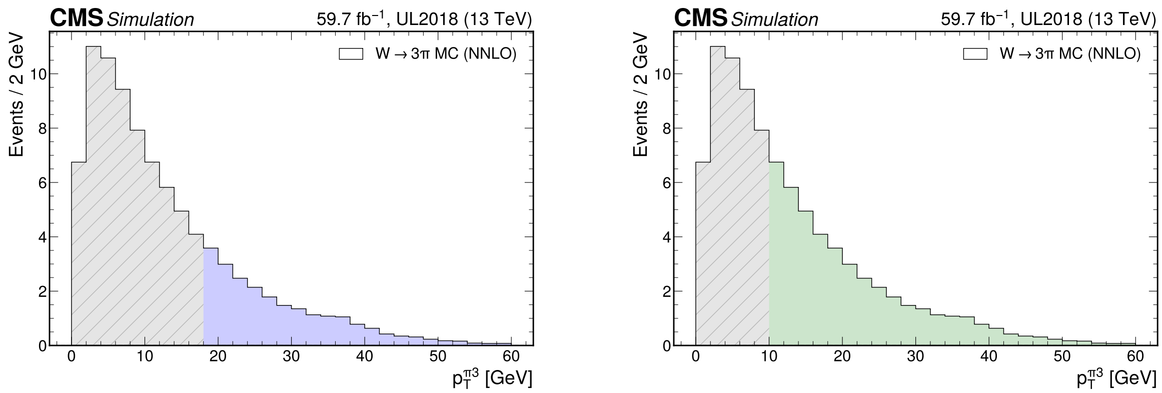

Figure 3.6: *Generated p_T of the lowest p_T pion after a ditau trigger is applied. The trigger requires at least two hadronic tau candidates with $p_T > 35\,\mathrm{GeV}$. In these plots the blue and green shaded areas demonstrate the p_T range in which tau candidates and isolated track objects, respectively, are reconstructed.*

introduced for data taking in 2023 which requires at least two single-prong, hadronic tau candidates with $p_{\mathrm{T}} > 35\,\mathrm{GeV}$. Additionally, the invariant mass of the ditau system is required to be between 5 and $80\,\mathrm{GeV}$. The exact triggers used are described above in section 3.1.2.

The final state object combinations that we consider are therefore:

- Three taus
- Three isolated tracks
- Two taus and one isolated track
- Three highest p_{T} objects that are either taus or isolated tracks

Requiring a ditau trigger guarantees that at least two hadronic tau candidates will be present in each event. This motivates the use of either the first or third strategies because it would give access to the additional tau variables. For completeness, however, the other two options are studied as well.

In the latter two schemes where we are using both collections to form a triplet there are often cases where the same object is classified as both a tau and an isolated track. When this happens we can choose to use the kinematic variables of either one. They differ due to the way that tau objects are formed. As described in Sec. 2.3.8, the p_T, η, and ϕ of a tau candidate take into account the hadronic activity in an isolation region surrounding the lead track.

For the case of two taus and one isolated track, we specifically used the isolated track variables, regardless of if the object was also found in the tau collection. The last scenario where the collections were fully mixed and an object was present in both collections was split further into two cases: one where the tau properties were used and one where the track properties were used.

Using the signal sample only, we performed two studies with these configurations. First, we measured the efficiency of reconstructing a valid triplet consisting of three unique pions of total charge $\pm$ 1. We required that the two highest p_T objects had $p_T > 35$ GeV (the lowest ditau trigger thresholds) and $|\eta| < 2.1$, regardless of if they were taus or isolated tracks. The third object was required to have $|\eta| < 2.3$. The p_T cut was determined by what type of object it was: > 18 GeV for a tau and greater than > 10 GeV for an isolated track.

Additionally, all taus were required to be reconstructed in decay mode 0. Isolated tracks were required to pass the basic quality cuts which included being of high purity, being closest to the primary vertex of interest than other vertices, and not being from a lost track. To avoid the potential overlap of tau object isolation cones, we required that all three objects be separated from each other by at least $\Delta R > 0.5$.

The number of valid triplets was compared to the number of generator level triplets that would pass the same p_T and η cuts as described above, using 10 GeV for the third pion. In events where more than one triplet could be formed which satisfied all of the above requirements the triplet containing the three highest p_T pions were chosen.

The efficiencies for each scenario are as follows:

- **All Taus:** 48.3%
- **All Isolated Tracks:** 63.0%
- **2 Taus + 1 Track:** 73.8%

- **Mixture, using Tau properties:** 77.8%
- **Mixture, using Track properties:** 77.8%

From these efficiencies it is evident that there is a significant gain in signal efficiency from selecting final state pions from both the tau and isolated track collections as opposed to limiting the selection to only one or the other. The scenarios using a mixture of both collections yielded the highest efficiency. However, it should be noted that these could include events with less than two tau candidates which would therefore not pass the trigger requirements. It is therefore not clear how these events could be included in the analysis. Nonetheless, we continued to consider these scenarios in our studies.

Next, we considered how often the pions used to construct the triplet selected by these different schemes was matched at generator level to the three pions originating from the W boson. An event with a triplet having all three pions matched was classified as a "Correct Triplet" event. If the triplet chosen was not matched at generator level to the three pions originating from the W boson, the event was classified as a "Wrong Triplet" event. Lastly, events in which no triplet could be formed from the available objects was referred to as a "No Triplet Found" event. The percentages of these different event categories for each triplet scenario are summarized in Table 3.3.

Scenario	Correct Triplet (%)	Wrong Triplet (%)	No Triplet Found (%)
All Taus	43.9	4.4	51.7
All Isolated Tracks	60.8	2.2	37.0
2 Taus + 1 Track	70.4	3.4	26.2
Mixture (Tau)	72.2	5.6	22.2
Mixture (Track)	72.4	5.5	22.2

Table 3.3: *The percentage of events in the signal sample with all three pions matched to generator level pions compared with the percentage of events with a triplet having at least one pion not matched and the percentage of events in which no triplet could be formed.*

The scenarios in which pions were chosen from either source (2 Taus + 1 Track and Mixture) saw the correct triplet selected more often then if exclusively taus or isolated tracks were used. The comparably large percentage (51.7%) of not finding a valid triplet in an event when only using tau objects is due to the fact that tau candidates are only reconstructed for $p_{\mathrm{T}} > 18\,\mathrm{GeV}$.

When the two collections are fully mixed the correct triplet is found slightly more often than explicitly looking for two taus and an isolated track. However, in those cases we also select the incorrect triplet much more often. If there was more than one possible triplet formed for these events it could still be possible to reconstruct the correct triplet by applying specific selection cuts. In the preceding studies only the triplet with the three highest p_{T} pions was considered. An additional study was performed to consider how often the triplet which had all three pions matched to generated pions from a generated W boson was reconstructed but not the first one selected. These events are a percentage of the events which were categorized as "Wrong Triplet" in the preceding study. Table 3.4 summarizes the results.

Scenario	Correct Triplet Present (%)
All Taus	34.1
All Isolated Tracks	63.6
2 Taus + 1 Track	64.7
Mixture (Tau)	71.4
Mixture (Track)	69.6

Table 3.4: *Finding the correct triplet when the wrong one was chosen but more than one triplet was able to be reconstructed.*

Based on these studies the scenario of choosing two pions from the tau collection and the third from the isolated track collection (2 Taus + 1 Track) works the best for this analysis. This scheme has both a high signal selection efficiency and high signal purity. An additional consideration is that requiring a ditau trigger assures that at least two hadronic tau candidates will be present in the event. This scenario benefits from both the enhanced discrimination provided by tau candidates and the increase in signal yield from the $10\,\mathrm{GeV}$ p_{T} threshold on the isolated tracks.

Kinematic properties of tau objects

Due to the nature of tau reconstruction, the p_{T}, η, and ϕ variables are reflective of the entire hadronic jet. The p_{T}, η, and ϕ of the lead track comprising the tau candidate can be calculated by multiplying by the ratio of the tau candidate variable (p_{T}, η, or ϕ) to the lead track variable. This ratio is provided for each tau candidate.

In the case of prompt, isolated pions from the decay of a W boson the lead track variables will be very close to that of the tau candidate. For reconstructed pions from QCD multijet events however, there can be large differences because any extra electromagnetic activity around the isolated track is included during tau reconstruction. Additional separation between the signal and backgrounds can therefore be gained by using information from the tau candidate as opposed to the lead track.

3.2.3 Pion triplet selection

Signal selection on tracks and taus

This section describes the cuts applied to the individual pions in the pion triplet. The offline selection for our three pions is based on the p_{T} requirements from the available triggers. Different selection criteria are applied to the pions based on whether they are found in the tau collection or the isolated tracks collection.

The two tau objects must be reconstructed by HPS in the one prong decay mode with no additional neutral pions (DM0), have $p_{\mathrm{T}} > 35\,\mathrm{GeV}$, and $|\eta| < 2.1$. Both taus are required to pass the VVLoose DeepTauv2p1 (DeepTauv2p5 in 2022 and 2023 data taking eras) working point versus electrons, the Tight working point versus muons, and the Medium working point versus jets. This working point versus jets reduces the contribution from QCD multijet events.

The isolated track is required to pass the basic quality cuts described in 3.2.2 and to have $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.5$. We also require that the isolated track has a relative isolation as defined in 3.2 of < 0.1. The isolation requirement is specifically to reduce contribution from

QCD multijet background. The three pions from W decay are expected to be prompt, so all objects, taus and isolated tracks, are required to have $|d_{xy}| < 0.045$ cm and $|d_z| < 0.2$ cm. This also serves to reduce the contributions from pileup interactions which will originate further from the primary vertex.

As mentioned in Sec. 3.1.2, the trigger scale factors are only provided for tau candidates with p_{T} 5 GeV above the trigger thresholds. The signal yield would be significantly reduced if the p_{T} requirements on the tau candidates were raised accordingly. To compensate we used the same samples and methods used in the centrally produced scale factors to derive trigger scale factors for the p_{T} region covered by this analysis.

Event cuts

This section describes the requirements applied to the whole event. We expect a clean final state with three charged pions so events are discarded which have additional leptons passing standard loose selection criteria, shown in Table 3.5. This criteria is predetermined by dedicated working groups. This serves to reduce backgrounds in which genuine tau leptons are present but the pions selected for the triplet are jets from initial state radiation. Consider for example a Drell-Yan event involving the decay of a Z boson to two tau leptons with two additional jets. If one tau decays to an electron or a muon and the other decays to a single charged pion the genuine pion and two additional jets could be reconstructed as the triplet. In this example there would be an extra electron or muon from the leptonic decay of the other tau lepton. If this extra electron or muon satisfied the loose criteria in Table 3.5, the event can be removed. This is known as a "lepton veto" and is effective at reducing the contribution from Drell-Yan processes which contribute non-negligibly.

Cut	Electrons	Muons
p_{T} [GeV]	> 10	> 10
$\|\eta\|$	< 2.5	< 2.4
$\|d_{xy}\|$ [cm]	< 0.045	< 0.045
$\|d_z\|$ [cm]	< 0.2	< 0.2

Table 3.5: *Loose requirements for lepton veto*

Additionally, electrons are required to pass `mvaFall17V2noIso_WP90`, have less than 2 lost hits, pass the conversion veto, and have a combined relative particle flow isolation (`pfRelIso03_all`) of less than 0.3.

Muons are required to pass the medium working point of the cut-based ID and have a combined relative particle flow isolation (`pfRelIso04_all`) of < 0.3.

Signal selection on pion triplet

Once three pions which satisfy the various requirements described above have been successfully identified triplets can be formed using these objects. Each triplet consists of two pions reconstructed as tau objects and one reconstructed as an isolated track. The total charge of the triplet must be $\pm$ 1 in accordance with charge conservation from the W boson. To avoid overlapping tau cones, we require that the tau objects are separated from each other and from the isolated track by $\Delta R > 0.5$. This reflects the size of the isolation cone used during tau reconstruction.

Various corrections are applied to the tau objects based specifically on the its properties. These include energy scales and DeepTau ID scale factors. As a result, it is still important that the p_{T}, η, and ϕ variables are not contaminated with particle flow objects from another nearby tau object as this could result in the application of an incorrect correction. This ΔR separation requirement does not remove a significant amount of signal events.

At this stage, we can measure the individual efficiency of each pion in reconstructed triplet. This efficiency is defined for each pion as the fraction of events in which the object (Tau or Isolated Track) can be matched to within $\Delta R < 0.1$ to one of the generated pions originating from the generated W boson. These efficiencies are shown in Table 3.6.

Pion	Efficiency
1 (Tau)	0.9963
2 (Tau)	0.9945
3 (IsoTrack)	0.9215

Table 3.6: *Efficiencies of matching the three pions in a reconstructed triplet to the generated pions originating from a generated W boson.*

The W boson mass peaks sharply at 80 GeV for W bosons produced at rest. However, due to the p_T requirements from the available ditau triggers, the invariant mass of boosted W bosons that we are actually able to reconstruct has a large tail. This is because we are selecting pions from the high end of their p_T distributions.

Ordering of triplets

In events which more than one triplet can be reconstructed from the available objects, the various triplets are sorted by the following properties, in this order: DeepTau score of highest p_T tau, DeepTau score of second highest p_T tau, isolation of isolated track as defined in Eq. 3.2, and finally the scalar p_T sum of the three objects.

3.3 Estimation of backgrounds

3.3.1 QCD multijet

The dominant background in this analysis is from QCD multijet. The proton-proton collisions at the CERN LHC frequently produce quark and gluon jets, both of which yield large amounts of charged pions. These types of events are highly complex and can not be accurately simulated with enough statistics to be used for an estimation. As a result the QCD multijet background is estimated from data. The estimate for the QCD multijet background in the signal region is estimated from three control regions.

In the first control region all of the selection requirements are the same as the signal region except the absolute value of the triplet charge (|Q|) is required to be 3. This region will be dominated by QCD multijet events. The second control region requires |Q| to be 1, but has relaxed isolation requirements. Events in this control region must fail either the DeepTauVsJet cut or the third pion isolation cut. This ensures that there is no overlap with the signal region. The final control has the same relaxed isolation requirements but requires |Q| to be 3. These selection requirements are summarized in Table 3.7

In each control region the QCD multijet background is estimated by taking the difference

Cut	**Signal Region**	**Control Region 1**	**Control Region 2**	**Control Region 3**
$\lvert Q \rvert$	1	3	1	3
$\lvert d_{xy} \rvert$	< 0.045	< 0.045	< 0.045	< 0.045
$\lvert d_z \rvert$	< 0.2	< 0.2	< 0.2	< 0.2
DeepTauVsEl	$\geq$ VVLoose	$\geq$ VVLoose	$\geq$ VVLoose	$\geq$ VVLoose
DeepTauVsMu	$\geq$ Tight	$\geq$ Tight	$\geq$ Tight	$\geq$ Tight
DeepTauVsJet	$\geq$ Medium	$\geq$ Medium	$\geq$ VVLoose	$\geq$ VVLoose
3rd Pion Iso	< 0.4	< 0.4	< 5	< 5

Table 3.7: *Summary of cuts in signal and control regions*

of observed data and the simulated MC processes. The second and third control regions are used to calculate a parameter which scales the QCD multijet estimate in a $|Q| = 3$ region to a $|Q| = 1$ region. The CMS statistical analysis tool, COMBINE[41], is used to calculate the maximum likelihood estimate for this parameter. The COMBINE tool performs this fit by evaluating the negative-log-likelihood function

$$-\ln \mathcal{L}(\hat{\vec{\mu}}, \hat{\vec{\nu}}), \tag{3.4}$$

where $\mathcal{L}$ is the likelihood function, $\hat{\vec{\mu}}$ is the parameter to be estimated, and $\hat{\vec{\nu}}$ are the values of nuisance parameters.

This parameter is treated as a floating normalization and is applied to the the QCD multijet estimate in the third control region as it is fit with the QCD multijet estimate in the second control region. The maximum likelihood estimate for this parameter is then applied to the QCD multijet estimate in the first control region, yielding an estimate for the QCD multijet background in the signal region.

This method is limited by the statistics in data as each control region needs to be well populated with events in order for the COMBINE to make an accurate estimate of the parameter. Therefore to increase the statistics, the data from all data taking eras is added together for the QCD multijet estimate in each control region. This is then scaled by the luminosity fraction for a given data taking era with respect to the total luminosity. These factors are shown in Table 3.8. The scaled QCD multijet estimation is fit individually in each data-taking era yielding eight unique fitting parameters. In this procedure the shape of the QCD multijet background in the signal region is taken from a phase space where the triplet charge (Q) is required to be ± 3.

3.3.2 Drell-Yan

Drell-Yan (DY) processes, specifically $Z \to \tau\bar{\tau}$ events, contribute to the background in two main ways. First, in the case of $Z \to \tau\bar{\tau}$ events, we can reconstruct two genuine pions and then misidentify an additional jet as the third pion. The other possibility applies to

Era	Luminosity Ratio Factor
`2016 preVFP`	0.098
`2016 postVFP`	0.084
`2017`	0.208
`2018`	0.300
`2022 preEE`	0.040
`2022 postEE`	0.134
`2023 preBPix`	0.089
`2023 postBPix`	0.047

Table 3.8: *Scaling factor used to normalize QCD multijet estimate for each data taking era.*

$Z \to ee$ or $Z \to \mu\bar{\mu}$ events. In these events electrons, muons, and additional jets can all be misidentified as pions. Some combination of misidentified leptons and misidentified jets will combine to mimic our signal. We estimate this background from simulations modeled at next-to-leading order.

Several cuts serve to reduce the contribution from DY events. First, the hadronic tau candidates are required to pass the VVLoose DeepTauv2p1 working point versus electrons and the Tight working point versus muons. This targets $Z \to e\bar{e}$ and $Z \to \mu\bar{\mu}$ events in which one or both of the leptons was misidentified as a hadronic tau candidate. Next, the hadronic tau candidates are required to pass the Medium DeepTauv2p1 working point versus jets. The lepton veto of loose electrons and muons also reduces the DY backgrounds in situations where two or more jets are misidentified as our signal pions, leaving an extra lepton present.

3.3.3 $t\bar{t}$

Events involving the decay of a top quark pair ($t\bar{t}$) can contribute to the background in several ways. These events are typically characterized by at least four final state objects. Two of these are bottom quark jets, which are produced in association with a W boson during the decay of the top quark. The other objects come from the decay of the W boson and so can be a lepton-neutrino pair, a hadronic tau jet with associated an neutrino, or a quark jet.

As $t\bar{t}$ events are characterized by large amounts of hadronic activity the isolation requirements on our final state pions serve to reduce this background. Genuine tau leptons from the decay of the associated W bosons could be combined with an isolated track associated with the bottom quark jet to form a triplet. These tracks from the bottom quark jet will tend to be less isolated than those from the electroweak decay of a W boson and so the isolation requirement on the isolated track serves to reduce the contribution of these events.

We estimate this background from simulation after applying corrections based on the p_{T} of the top quark[42]. The corrections compensate for the top quark p_{T} spectrum in MC being higher than what was measured in data. These corrections are described in section 3.4.6. The fully hadronic, fully leptonic, and semi-leptonic final states are modeled separately.

3.3.4 Diboson

Diboson processes can produce a wide variety of final states due to the many possible decay channels of the W and Z boson and the presence of two decaying objects. These backgrounds were estimated from simulation.

The different final states considered and which combination of bosons they correspond to are summarized below:

- **2 W Bosons:** 2 leptons and 2 neutrinos, 1 lepton and 1 neutrino and 2 quarks, 4 quarks
- **2 Z Bosons:** 2 leptons and 2 neutrinos, 2 leptons and 2 quarks, 2 neutrinos and 2 quarks, 4 leptons, 4 quarks
- **1 W and 1 Z Boson:** 3 leptons and 1 neutrino, 1 lepton and 3 neutrinos, 2 quarks and 2 leptons, 1 lepton and 1 neutrino and 2 quarks

3.3.5 Others

Very minor contributions to the total background from triboson, $t\bar{t}$ with associated vector boson (W or Z), single top, and W+jets processes are estimated from simulation. In particular, no single top or W+jets events passed our full signal selection.

Variable	Requirement
|Q|	1
$\|d_{xy}\|$	< 0.045
$\|d_z\|$	< 0.2
DeepTauVsEl	$\geq$ VVLoose
DeepTauVsMu	$\geq$ Tight
DeepTauVsJet	$\geq$ Medium
3rd Pion Iso	< 0.4

Table 3.9: *Summary of cuts in signal region*

3.3.6 Observables

The cuts described in the sections above are summarized in Table 3.9. The phase space delineated by these requirements is called the signal region because a large percentage of the simulated W decays to three charged pions satisfy these conditions. When selecting events for the analysis the first triplet which meets the necessary requirements is chosen, when ordered as described in Section 3.2.3. Only one reconstructed triplet is used for each event.

The invariant mass of three reconstructed charged pions is an obvious choice for an observable. In signal events with a real W boson decaying to three charged pions this distribution will be sharply peaked at the known mass of the W boson, 80.4 GeV. Other processes contributing to the signal region will have broad distributions over the invariant mass of the three reconstructed charged pions. Unfortunately, these broad distributions, particularly that of the QCD multijet background (see Section 3.3.1), overlap with the signal peak. This is partly due to the kinematic requirements on the two pions reconstructed as hadronic tau candidates. Figure 3.7 shows the signal and background invariant mass distributions in the signal region. In this plot, the signal yield is increased by a factor of 10 in order for it to be visible. The contributions for the various backgrounds are estimated as described in Section 3.3.

Other variables considered as a choice of observable are shown in Fig 3.8. Like in Fig 3.7, the signal yield is again scaled by a factor of 10 for visibility. These provide only a limited discriminating power on their own but could be combined with the invariant mass to construct a new, more powerful observable.

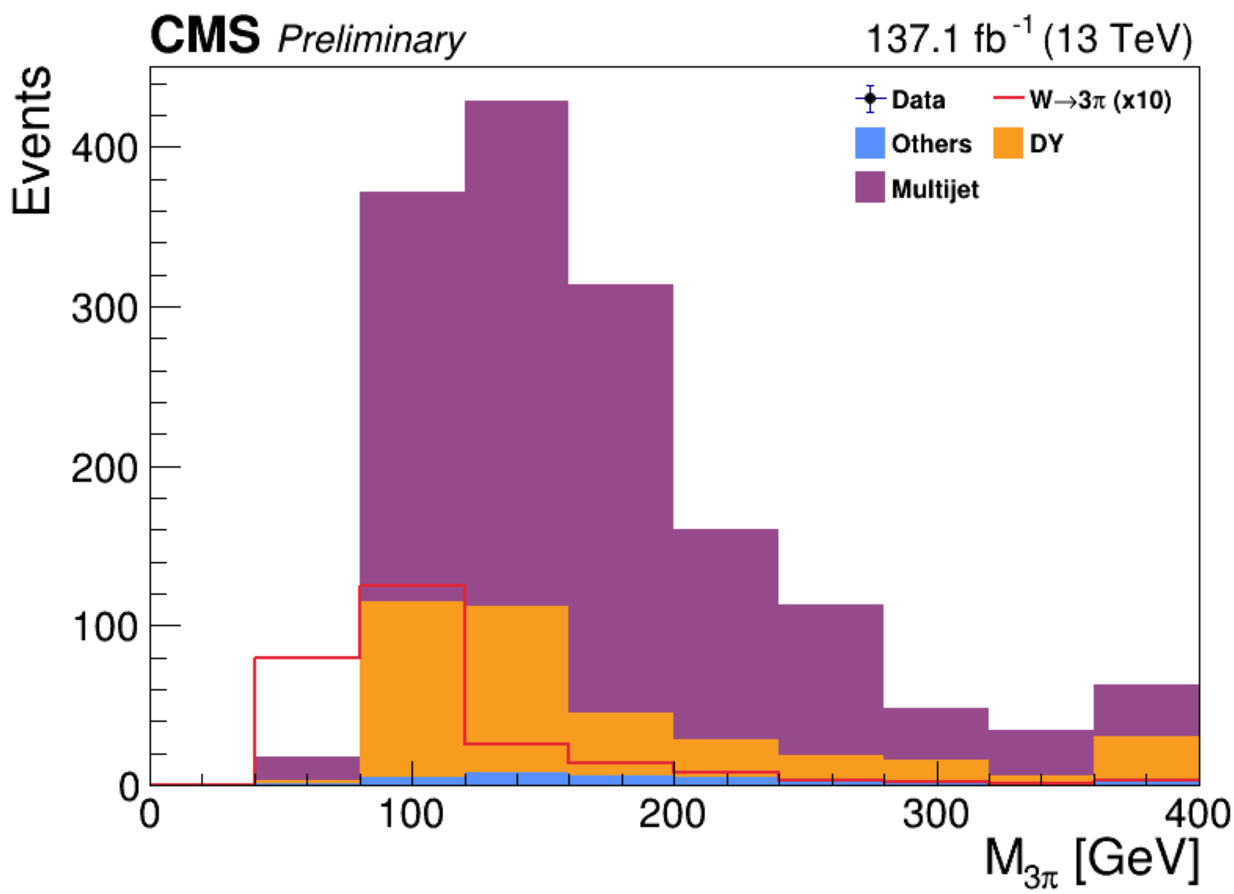

Figure 3.7: *Invariant mass of the three reconstructed charged pions for signal (red), QCD multijet (magenta), DY (orange), and others (blue).*

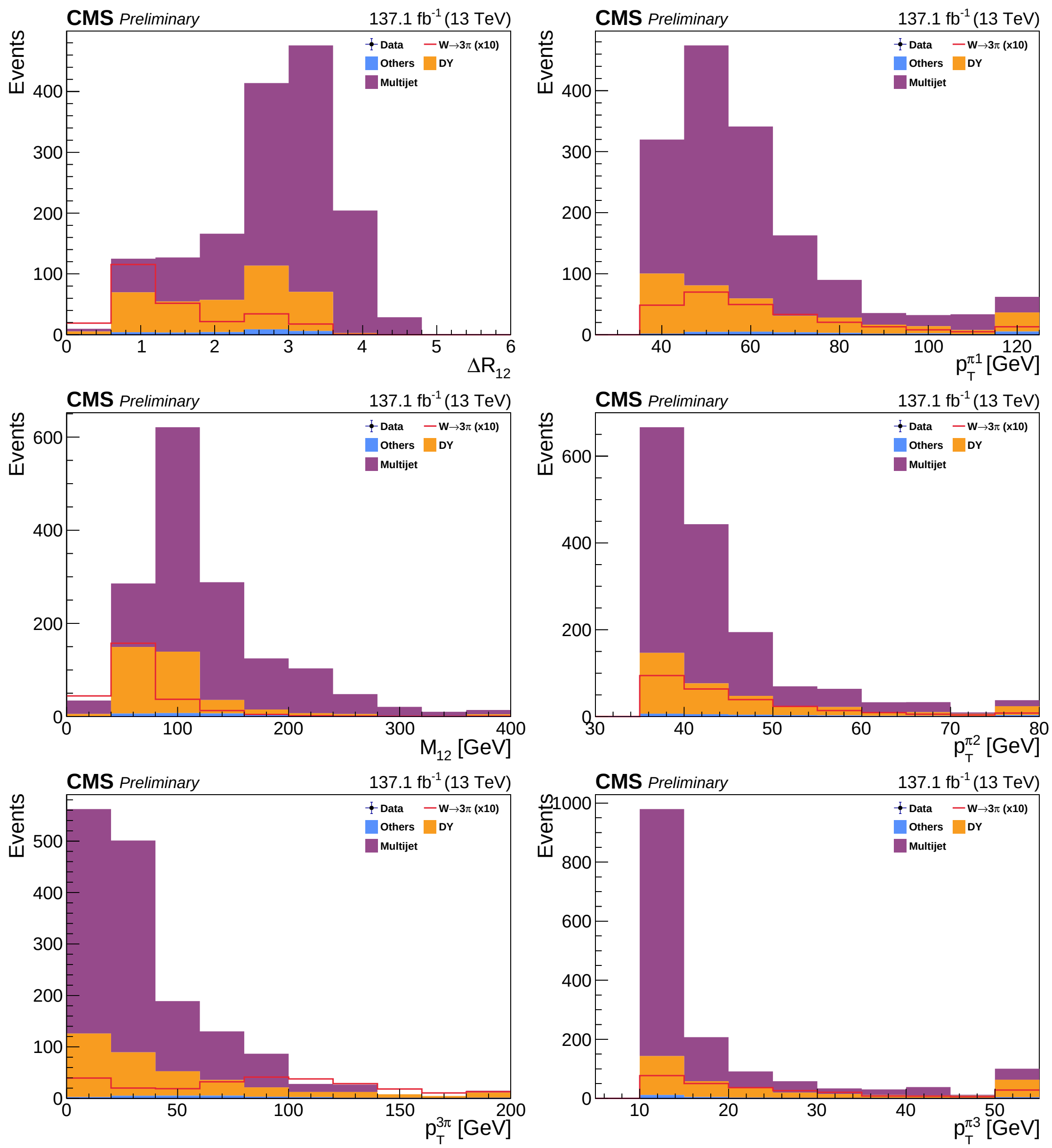

Figure 3.8: *Signal (red), QCD multijet (magenta), DY (orange), and others (blue) plotted for different variables of the reconstructed pion triplet. The signal yield has been scaled up by a factor of 10 in each plot.*

Layer	Number of nodes	Activation function
Input	11	-
1st Hidden	33	ReLU
2nd Hidden	11	ReLU
3rd Hidden	11	ReLU
Output	3	softmax

Table 3.10: *Summary of architecture*

3.4 Neural Network

This section describes the procedure to setup, train, and validate the multiclass neural network (NN) used in this analysis. The task of the NN is to assign probabilities to individual pion triplets which correspond to the likelihood that the triplet was produced as the result of a specific physical process. Three classes are considered: signal (W boson to 3 charged pions), QCD multijet, and DY. The signal discriminating score is used as the discriminating observable for the combined maximum-likelihood fit.

3.4.1 Architecture

The NN used in this analysis is a fully connected, feed-forward, multiclass model built using the Keras API[43]. It consists of five layers: an input layer, three hidden layers, and an output layer. The model is summarized in the Table 3.10. In total, there are 938 trainable parameters.

The activation function used in the hidden layers is the rectified linear unit activation function, known as ReLU. This function returns the input if it is greater than zero, and returns zero otherwise. A plot of this distribution is shown in Fig 3.9. The softmax activation function is used in the final layer and produces a probability distribution such that the outputs sum to 1.

3.4.2 Input variables

The NN was trained on 11 different kinematic variables of the pion triplet. These variables are:

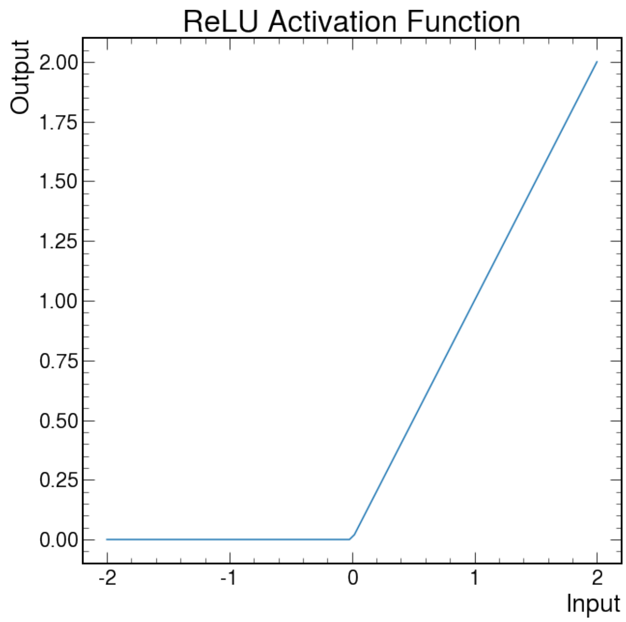

Figure 3.9: *The ReLU activation function*

- p_{T} of first, second, and third pions
- ΔR separation between all pion pairs
- invariant mass of all pion pairs
- p_{T} of reconstructed W boson
- invariant mass of reconstructed W boson

Other variables were considered but this set of variables was chosen due to the visible separation between the different classes. Each variable is scaled prior to training such that the mean of the total distribution is 0 and the standard deviation is 1. This ensures that all variables are treated equally by the network and no information is learned simply from the physical values. The scaled distributions for the 11 variables are shown in Fig. 3.10.

3.4.3 Training dataset

The network was trained on representative events from the three different classes. For the signal class, a separate set of MC samples from those used in the analysis were generated for

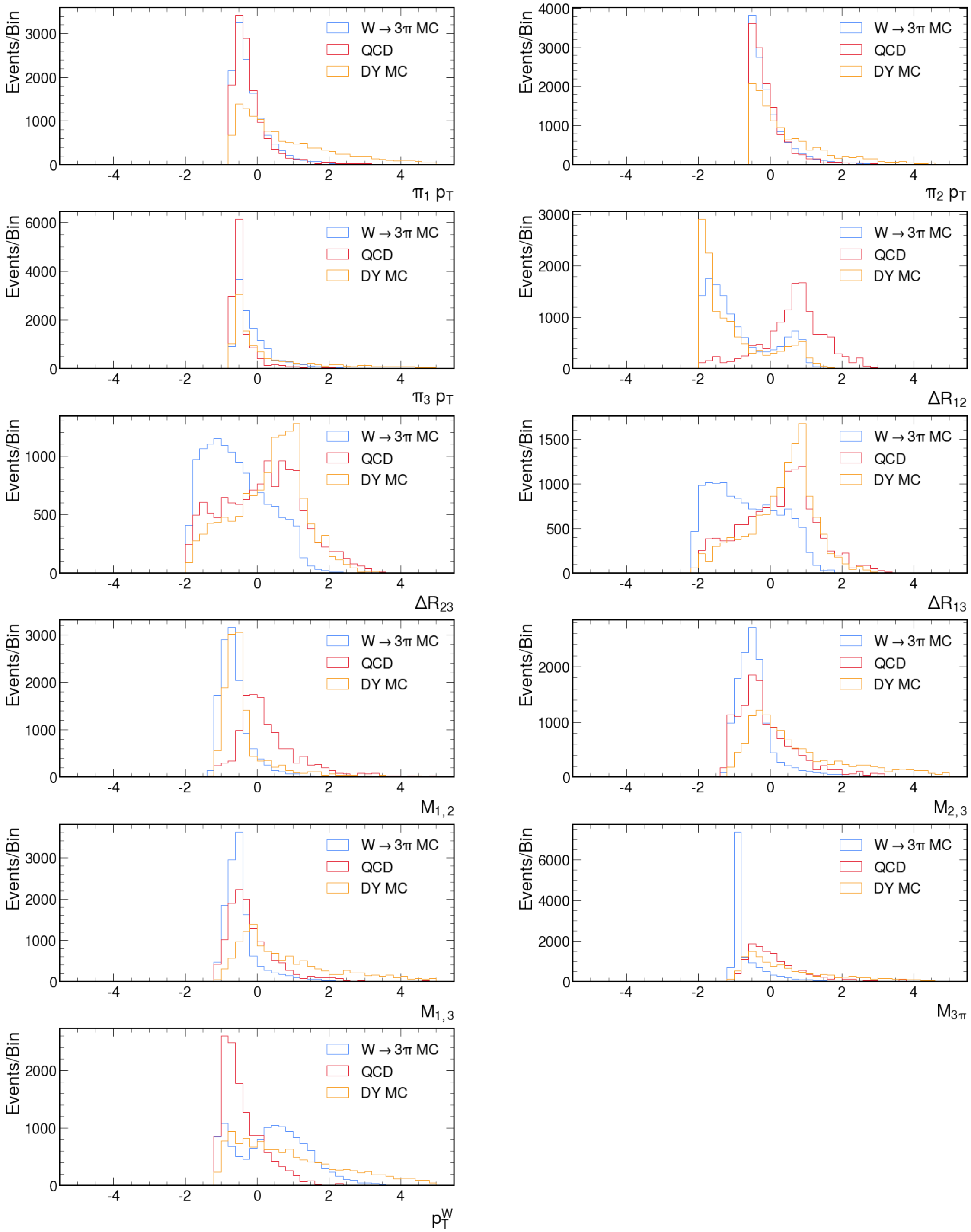

Figure 3.10: *The 11 variables used for neural network training. The distributions here have been scaled such that the mean is at 0 and the variance is 1, so the x-axis has no physical meaning.*

each era. These were produced with the specifications described in Sec. 3.1.3. For the DY class, MC samples were also used. The network was trained directly on the samples used in the analysis. For the QCD class, the network was trained on events from data for which the total charge of the triplet was ±3.

Events for training were selected in the same way as the analysis. The triplets trained on consisted of two single-prong taus and an isolated track. Table 3.11 summarizes the selection requirements for the three classes. In the cases where more than one triplet pass this criteria, a single triplet is chosen per event for training according to the ordering described in Sec. 3.2.3.

Cut	**Signal**	**QCD**	**Drell-Yan**
|Q|	1	3	1
$\|d_{xy}\|$	< 0.045	< 0.045	< 0.045
$\|d_z\|$	< 0.2	< 0.2	< 0.2
DeepTauVsEl	$\geq$ VVLoose	$\geq$ VVLoose	$\geq$ VVLoose
DeepTauVsMu	$\geq$ Tight	$\geq$ Tight	$\geq$ Tight
DeepTauVsJet	$\geq$ VVLoose	$\geq$ VVLoose	$\geq$ Medium
3rd Pion Iso	< 5	< 5	< 0.4

Table 3.11: *Summary of cuts for NN training classes*

To avoid a bias in applying the network on events which were used for training, a two-fold approach was employed. The total training dataset was split into two parts, with two networks trained, one on each half of the data. Then each network is applied on the other half of the data. The unweighted number of events for each class available for each network is shown in Table 3.12. Weights are applied to the classes such that the sum of the weighted events of each class is equal. This prevents the network from learning information purely from the ratios of the classes.

3.4.4 Training procedure

In order to properly evaluate the performance of the network 5% of events are witheld completely from training. Of the remaining 95%, 75% are used for training with the remaining used as a validation set. The training set is randomly split into batches of 4096 events. The

Class	“Even” Network	“Odd” Network
Signal	21992	21916
QCD	2134	2114
Drell-Yan	5891	5768

Table 3.12: *Number of unweighted events available for each class*

No. of hidden layers	3
Hidden layer activation function	ReLU
Regularization	Dropout, 20%
Loss function	categorical crossentropy
Optimizer	Adam [44]
Learning Rate	1e-3
Batch Size	4096
Early Stopping	75 epochs
Validation Split	25%

Table 3.13: *Summary of hyperparameters which can be adjusted during training in order for the network to converge.*

trainable parameters are adjusted after each batch has been processed by the network. Once all events in the training set have been used the network is evaluated on the validation set. This defines a single epoch.

The objective of the NN training is to find a minimum of the loss function. For this multiclass network the loss function used was categorical crossentropy. Training continues until a predetermined number of epochs or the validation loss has not decreased for 75 epochs, whichever comes first. The latter of these conditions is a likely indicator that the loss function has reached a minimum. This is therefore a preferable stopping condition to a fixed number of epochs, especially because inherent randomness in the batches will cause natural variations in the number of epochs necessary to reach a minimum.

If the network becomes too specific to the features of the training sample it will perform poorly on the validation and testing samples. This is known as overtraining and techniques used to prevent overtraining are called regularization. Dropout layers are one form of regularization which this network uses. In each of the first two hidden layers, 25% of the nodes are deactivated each epoch.

The loss and accuracy for the training and validation sets is record after each epoch and

Figure 3.11: *Training (blue) and validation (orange) loss shown for each epoch as an example of a training which has converged.*

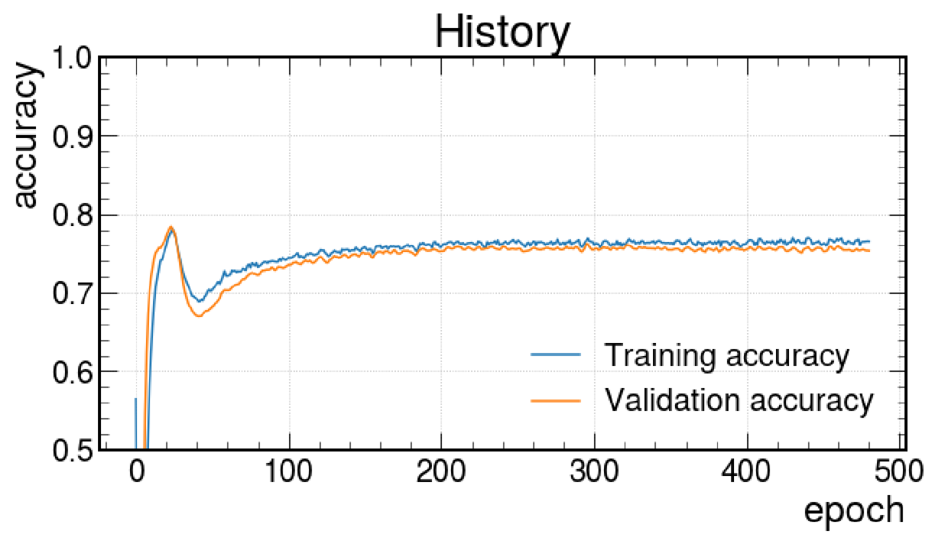

Figure 3.12: *Training (blue) and validation (orange) accuracy shown for each epoch as an example of a training which has converged.*

inspected following once training is complete. Examples of these loss and accuracy plots are shown in Fig 3.11 and Fig 3.12, respectively. These plots do not show the loss and accuracy for the exact networks used in this analysis. Instead they are examples of networks which have converged. The training and validation curves are similar throughout the training and reach the same minimum value. This indicates that the network is fully optimized for the architecture and hyperparameters.

3.4.5 Performance of neural network

Once a set of hyperparameters is found which allows the training and validation losses to converge to a minimum value the performance of the network is evaluated on the testing data set. These events were not included in the training and therefore provide an unbiased look at the performance of the network. Events in the test dataset are labeled by which class they are known to originate so discrimination scores an event receives can be compared against the true label.

The network outputs three discriminant scores, one for each class. These can be visualized in a few different ways. First, in Fig. 3.13 the three discriminants are viewed in one dimension. The discriminant score for both the training and testing data set is plotted for comparison to ensure the distributions are similar. The performance is inspected in two dimensions in Fig. 3.14 and Fig. 3.15. These plots are not of the exact network used in the analysis but are representative of good discrimination.

The convergence of the training loss and validation loss to a similar value and the comparable discrimination of training and testing data for all three classes implies that the network is well trained and can be used in the analysis.

The distributions of $Q = \pm 1$ and $Q = \pm 3$ events over the neural network signal discriminant are compared in Fig. 3.16. The disagreements in shape over the neural network signal discriminant seen between the control regions will be taken into account with a systematic uncertainty.

The agreement between data and MC for the neural network signal discriminant was checked in a DY enriched control region. Events in this control region were required pass the single muon trigger, `HLT_IsoMu24`, and to have exactly two oppositely charged muons and exactly one additional tau candidate. The tau candidate was required to pass the VVLoose DeepTau discriminator versus jets.

Figure 3.17 shows the DY MC sample compared with data for the neural network signal discriminant and the visible dimuon mass. The visible dimuon mass was required to be $> 80\,\mathrm{GeV}$ and $< 100\,\mathrm{GeV}$ when plotting the neural network signal discriminant. This

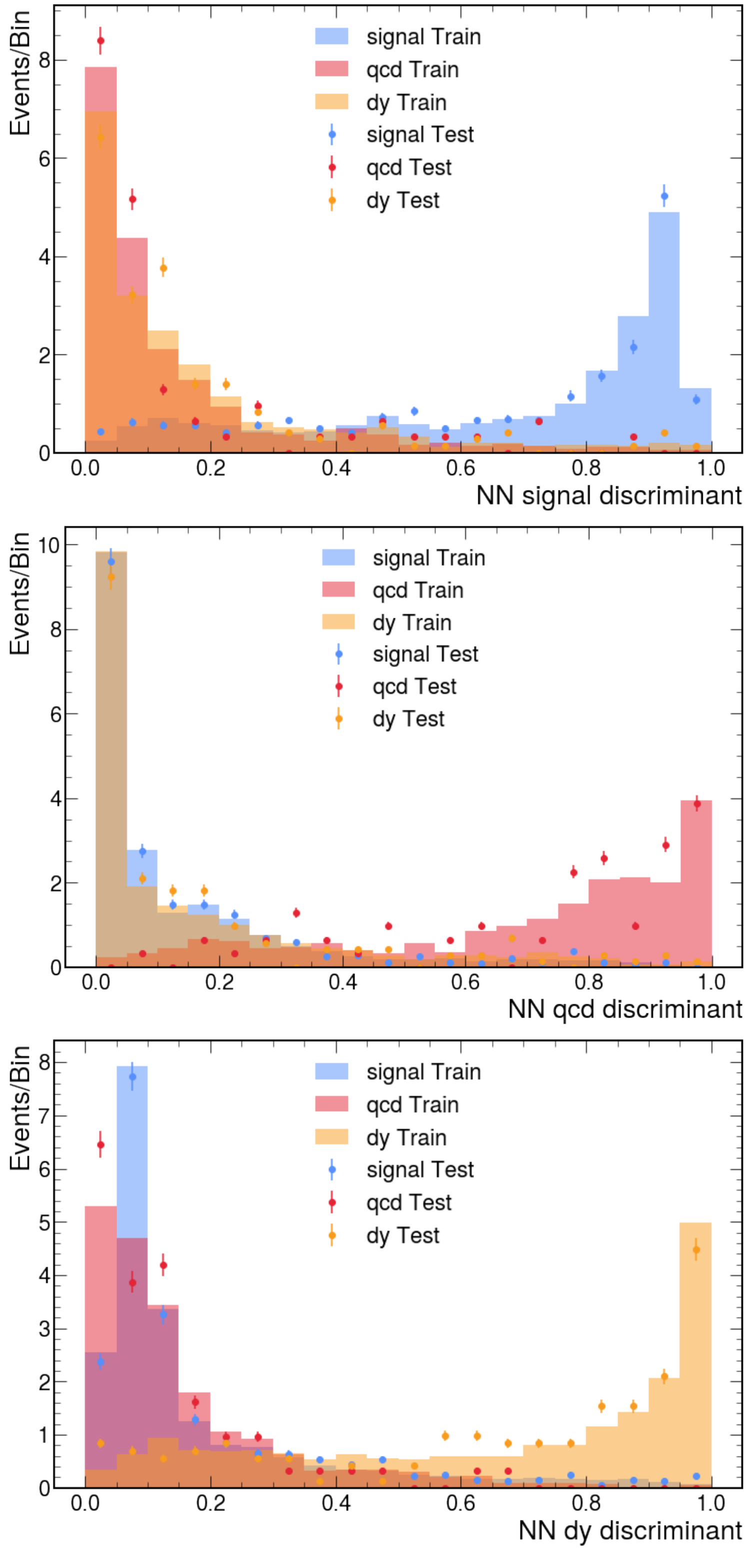

Figure 3.13: *The neural network discriminants for signal (top), qcd (middle) and DY (bottom). In each plot the filled histograms*

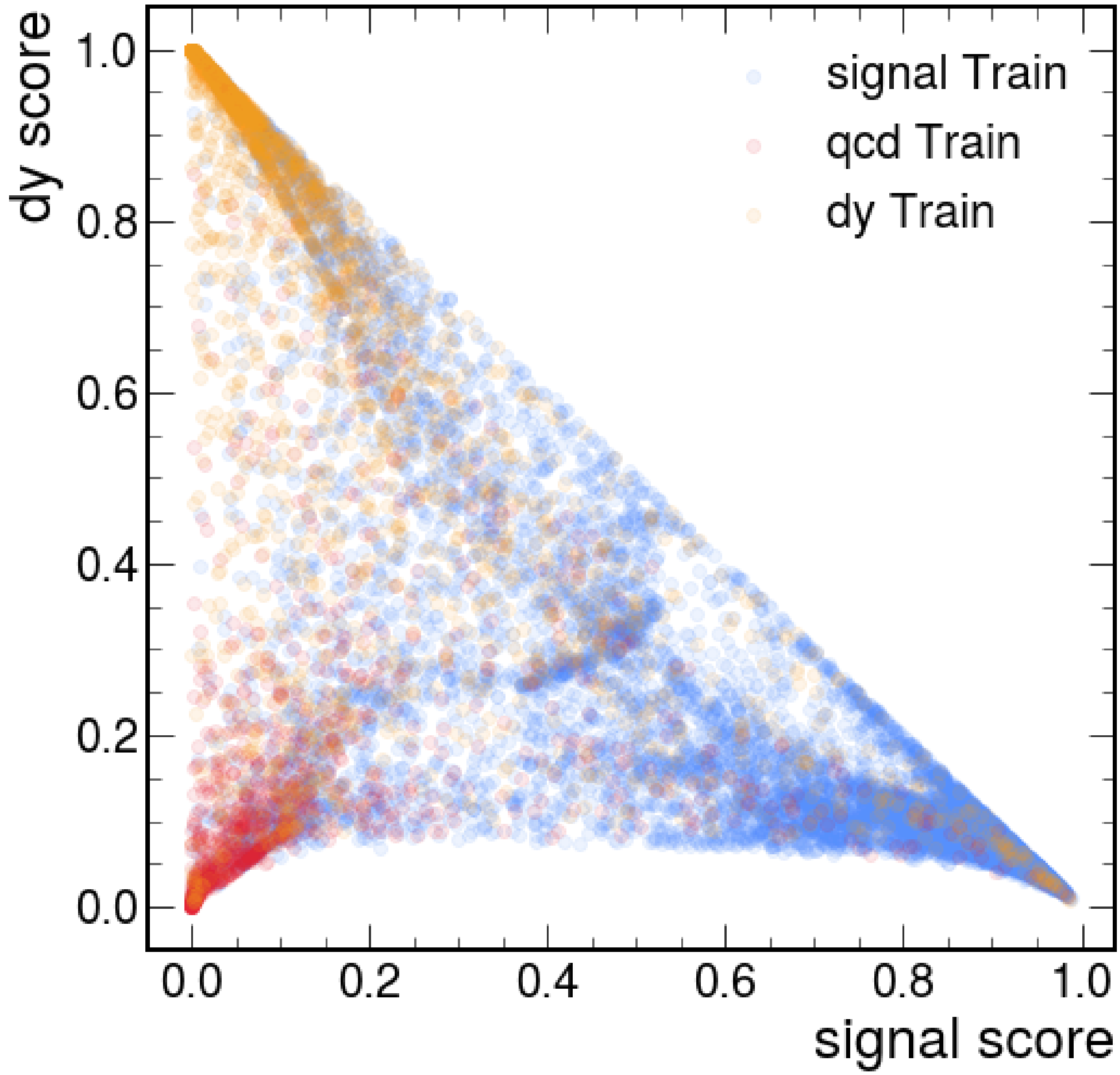

Figure 3.14: *The neural network signal discriminant score is plotted against the DY score. The different classes are shown in blue (signal), red (qcd), and orange (DY).*

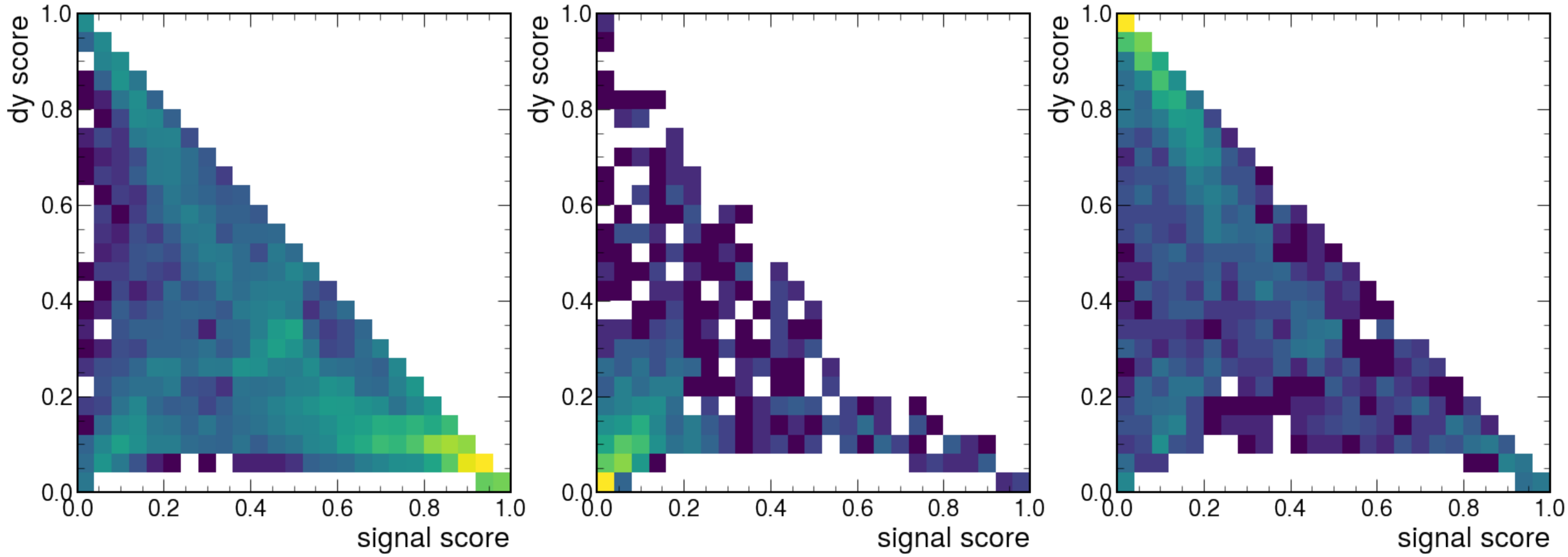

Figure 3.15: *The neural network signal discriminant score is plotted against the DY score for signal (left), qcd (middle), and DY (right)*

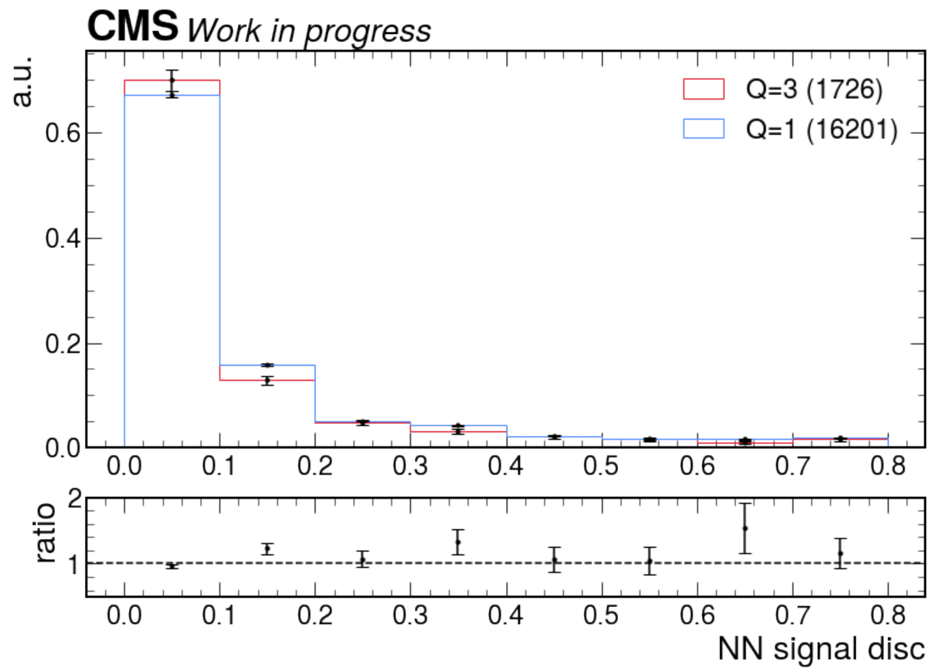

Figure 3.16: *Comparison of shape for QCD multijet background in control regions. Both distributions have been normalized to a integral of one.*

ensures the events in data are from $Z \to \mu\bar{\mu}$ decays. The shapes agree to within statistical uncertainties.

3.4.6 MC Reweighting

$t\bar{t}$ NNLO

This correction takes the form of a weight which is a function of the p_{T} of both top quarks present in the event. This correction is necessary because the p_{T} of the top quarks measured during Run 1 was found to be softer than what was predicted in MC simulations. This was further verified during Run 2. Some of this discrepancy can be explained by the absence of higher-order calculations, however the exact origin is unknown.

The specific parametrization of the scale factor function depends on if next-to-leading order (NLO) or next-to-next-to-leading order (NNLO) simulations are being used. The $t\bar{t}$ samples listed in section 3.1 are of NNLO, so we use the following scale factor:

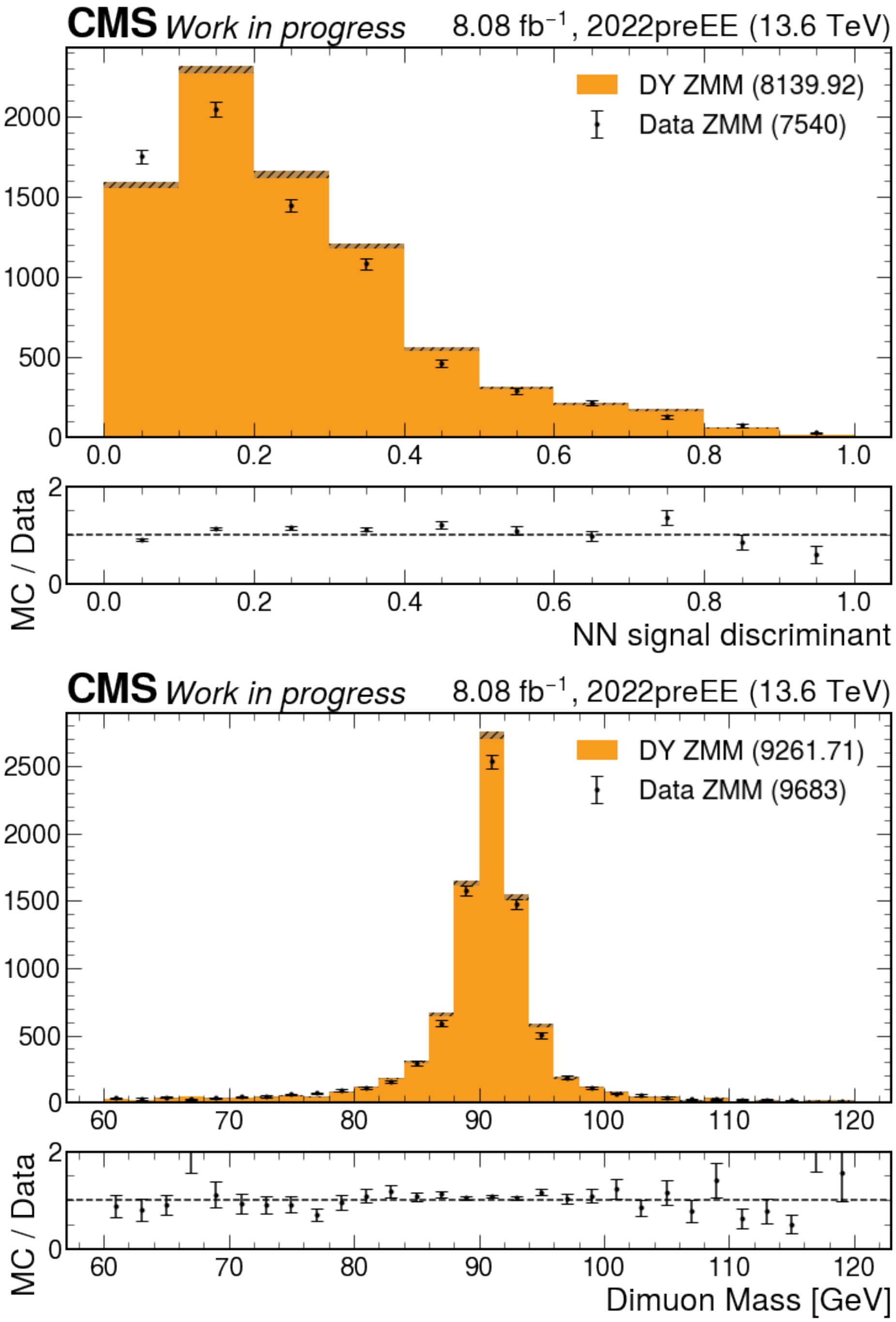

Figure 3.17: *Data-MC agreement for neural network signal discriminant (top) and visible dimuon mass (bottom) in a $Z \to \mu\bar{\mu}$ control region. In the top plot the visible mass of the dimuon system has been restricted to be between* 80 *and* 100 GeV

$$SF(t^{p_{\mathrm{T}}}) = 0.103 \cdot e^{-0.0118 \cdot p_{\mathrm{T}}} - 0.000134 \cdot p_{\mathrm{T}} + 0.973 \tag{3.5}$$

This scale factor is calculated for both top quarks present in the event, and then the two are combined as described in equation 3.6:

$$SF(t_1^{p_{\mathrm{T}}}, t_2^{p_{\mathrm{T}}}) = \sqrt{SF(t_1^{p_{\mathrm{T}}}) \times SF(t_2^{p_{\mathrm{T}}})} \tag{3.6}$$

Pileup reweighting

Particles originate from many primary vertices during collisions, however all those not associated with the vertex from which the selected pions originated are considered "pileup" (PU). In simulation, the number of PU interactions is determined from a distribution with a specific mean. The mean value also comes from a distribution, which must be weighted to match that of data[28].

3.5 Uncertainties

A maximum likelihood fit to the neural network discriminant score includes the various systematic uncertainties detailed below. In this analysis, however, statistical uncertainties dominate over the detailed systematics.

3.5.1 Tau object uncertainties

The uncertainty in the energy scale of tau leptons is considered by scaling the energy of tau leptons and propagating the effects to the event yield. This is derived separately for DM0 and DM1 tau leptons. As our background is dominated by QCD multijet events this uncertainty does not have a significant effect.

For uncertainties in the identification of hadronic tau decays, we use those provided by the TAU physics object group. Unique uncertainties exist for each DeepTau working point. These are calculated in independent η bins for the case of electrons and muons misidentified as tau leptons. For electrons, there are two η bins: [0, 1.5] and [1.5, 2.3]. For muons, there are 5 bins in η with edges [0, 0.4, 0.8, 1.2, 1.7, 2.3].

There are two statistical uncertainties considered for tau objects. Both are split by DM.

Uncertainties in the efficiency of the trigger are also considered by varying the p_{T} of the tau leptons. These are also derived separately for DM0 and DM1 tau leptons.

This analysis only uses DM0 tau leptons, so when uncertainties are split by DM, the DM1 option is not considered.

3.5.2 Uncertainties in background estimation

Data-driven background

We estimate the QCD multijet background using data and so the shape uncertainty is related to the available data statistics. Each bin of QCD multijet background is calculated independently from scaling the corresponding bin in our shape control region by the ratio of the other two control regions. We therefore calculate an up and a down uncertainty for each bin in the fit. These correspond to the positive and negative product of the shape control region bin and the square root of ratio of the number of events in the other two control regions.

Pileup

A shape uncertainty is added to account for the uncertainty in the reweighting of MC samples to match the pileup distribution. This was described in section 3.4.6.

Parton shower

We consider the effect of variations in the modeling of initial-state radiation (ISR) and final-state radiation (FSR). This is done by scaling up and down the effect of ISR and FSR and then examining the effect to the event yield.

QCD scale uncertainty

The effect of the QCD scale is considered for separately for three groups of MC samples, corresponding to if the final state particles originated from single boson processes, diboson processes, or top quark processes.

Cross section

The measured cross sections for the $W \to 3\pi$ and DY MC samples have an associated uncertainty. These are considered as shape uncertainties

Underlying Event

The underlying event (UE) refers to the extra particles that are not directly produced from the hard scattering of protons. The specific `PYTHIA` tune was modified corresponding to the "up" and "down" variations for $t\bar{t}$ and single-top samples. The result was an overall 1.1% uncertainty on the luminosity.

3.5.3 Others

The normalization of all simulated processes are affected by uncertainties in the luminosity measurement. We consider correlated uncertainties affecting 2016, 2017 and 2018. An additional correlated is considered between 2017 and 2018. The data taking years of 2018, 2022, and 2023 have individual uncertainties on the luminosity. In total, the uncertainties per year are 1.2%, 2.3%, 2.5%, 1.4%, and 1.4% for 2016, 2017, 2018, 2022, and 2023 respectively[45–47].

As mentioned above, this analysis is dominated by statistical uncertainties. These are implemented bin-by-bin with the `autoMCstat` feature available from COMBINE[41].

3.6 Results

A binned maximum likelihood fit is performed using the COMBINE tool on the distributions of the neural network signal discriminant. The normalization and shape uncertainties listed in

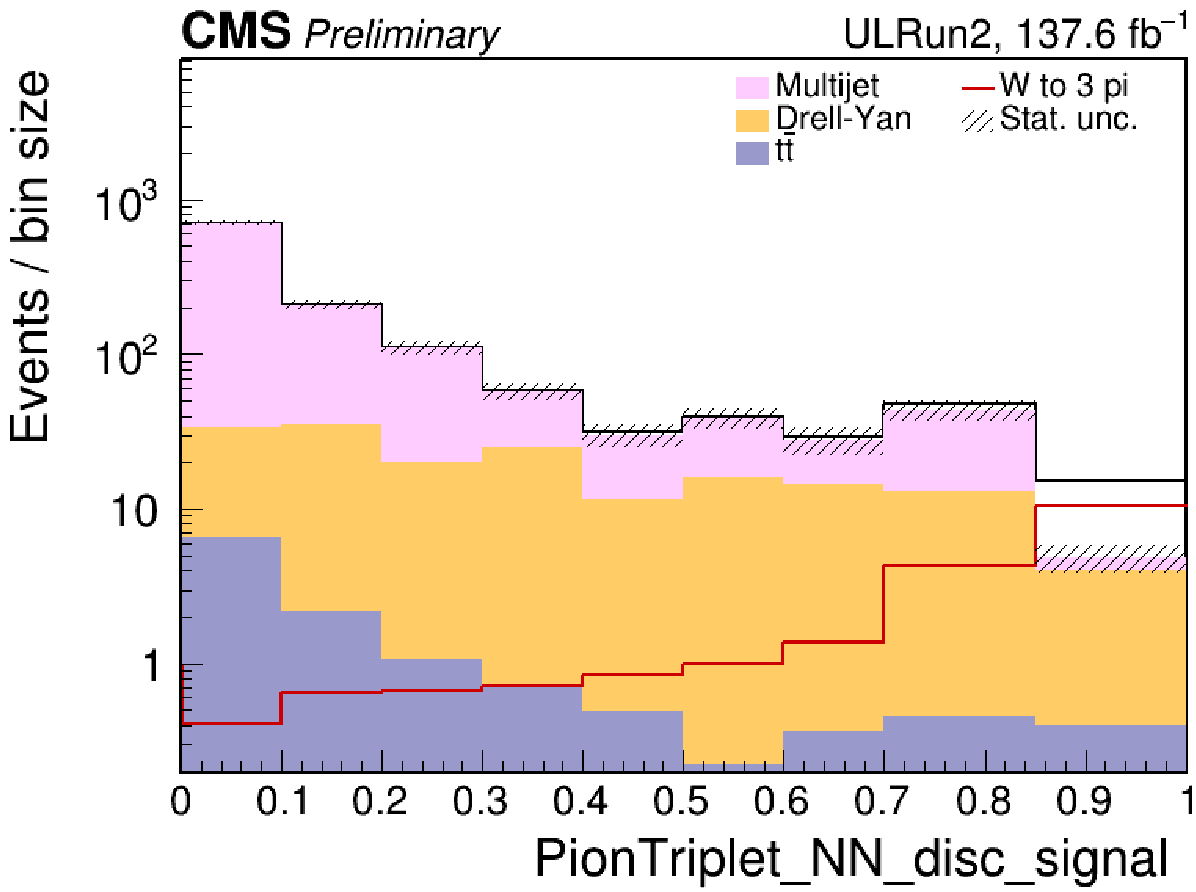

Figure 3.18: *Neural network signal discriminant plotted for the combination of all Run 2 data taking eras*

Sec 3.5 are considered as nuisance parameters. An Asimov dataset is used which assumes the data distribution exactly matches the estimation of backgrounds. The blinded distribution for the neural network signal discriminant is shown in Fig. 3.18.

The expected upper limits on the branching fraction of a W boson to three charged pions are set at at 95% confidence limit using the asymptotic approximation[48]. The expected limits are extracted individually for each data taking era and then combined for Run 2. These limits are shown in Table 3.14 and Table 3.15.

Era	Expected Limit x10^{-6}
2016 preVFP	2.3125
2016 postVFP	2.1484
2017	1.6406
2018	1.3398

Table 3.14: *Expected limits on the branching fraction for decay of W boson to three charged pions for each data taking era*

Era	Expected Limit x10^{-6}
Run 2	0.8164

Table 3.15: *Expected limits on the branching fraction for decay of W boson to three charged pions for combinations of data taking eras*

Chapter 4

HCAL Operations Support

This chapter details contributions to the continued operation of the detector, specifically HCAL. An analysis of the radiation damage to the HF calorimeter is described in Section 4.1. Using this procedure, corrections to the energy response were derived. Section 4.2 describes the synchronization framework used to update calibrations related data for HCAL. Significant improvements were made to this framework as it was deployed for data-taking in 2022.

4.1 Radiation damage calibrations to HF Long fibers

As mentioned in Section 2.2.4, the HF calorimeter sits outside of the tracker acceptance and therefore experiences a large amount of particle flux. An annotated diagram of the HF components is shown in Fig 4.1. It is therefore necessary to measure the effects of radiation on the quartz fibers of HF and apply corrections if the energy response has degraded. The L and S fibers are evaluated separately as a result of the different types of particle showers to which they are sensitive. Z boson decays to electron-positron pairs are used for the L fiber calibration because of the large production cross section and the relatively background-free EM signature. The strategy uses electrons which deposit energy in HF to reconstruct the invariant mass of Z bosons candidates and compare it to Drell-Yan Monte Carlo (MC) simulations[49].

Electrons from the decay of a Z boson have high-p_{T} and small isolation. If these electrons shower in HF, they do not pass through the tracker and are therefore not reconstructed as PF electrons (see Section 2.3). Instead, energy deposits in the L and S fibers are labeled as PF photons and PF hadrons, respectively. These PF objects are then clustered into PF jets using the methods described in Section 2.3.7. As a result, HF electrons are reconstructed using PF jets. First, the highest energy PF jet constituent is selected to define the initial kinematic variables for the HF electron. Next, other jet constituents are added as a 4-vector sum to the HF electron if they are within a distance $\Delta R = 0.15$[49] until none remain. Electron jets are expected to be more isolated than hadronic jets and so the energy in an isolation cone of $0.15 < \Delta R < 0.30$ must be less than 55% of the HF electron energy. Additionally, the following requirement is placed on the L and S fiber energy deposited by the HF electron

$$\frac{2S}{L-S} < 1.2, \tag{4.1}$$

where $2S$ and $L - S$ are approximations for the hadronic and EM energy of the electron, respectively.

Events are selected which have exactly one isolated electron in ECAL with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$ and at least one HF electron, reconstructed as described above, with $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ and $2.964 < |\eta| < 5.191$. If there are multiple HF electrons in an event, the highest p_{T} is chosen. The invariant mass of the Z boson candidate is reconstructed from the ECAL and HF electrons. To isolate the effects of radiation on the long fibers, the HF electron 4-vector includes only the long fiber energy. Finally, the di-electron invariant mass distribution is plotted in different regions of η and ϕ and compared between data and MC.

It is convenient at this point to define integer ranges of η and ϕ, referred to as $i\eta$ and $i\phi$, respectively. HF covers the area $|i\eta| = 29 - 41$, with each $|i\eta|$ spanning an angular distance of $\Delta\eta = 0.087$. There are 72 $i\phi$ sections, each covering the same angular distance as an the $|i\eta|$ sections, $\Delta\phi = 5^{\circ}$ or $\frac{\pi}{36}$.

Separate distributions of the invariant Z boson mass are made for each $i\eta$ region, inclusive

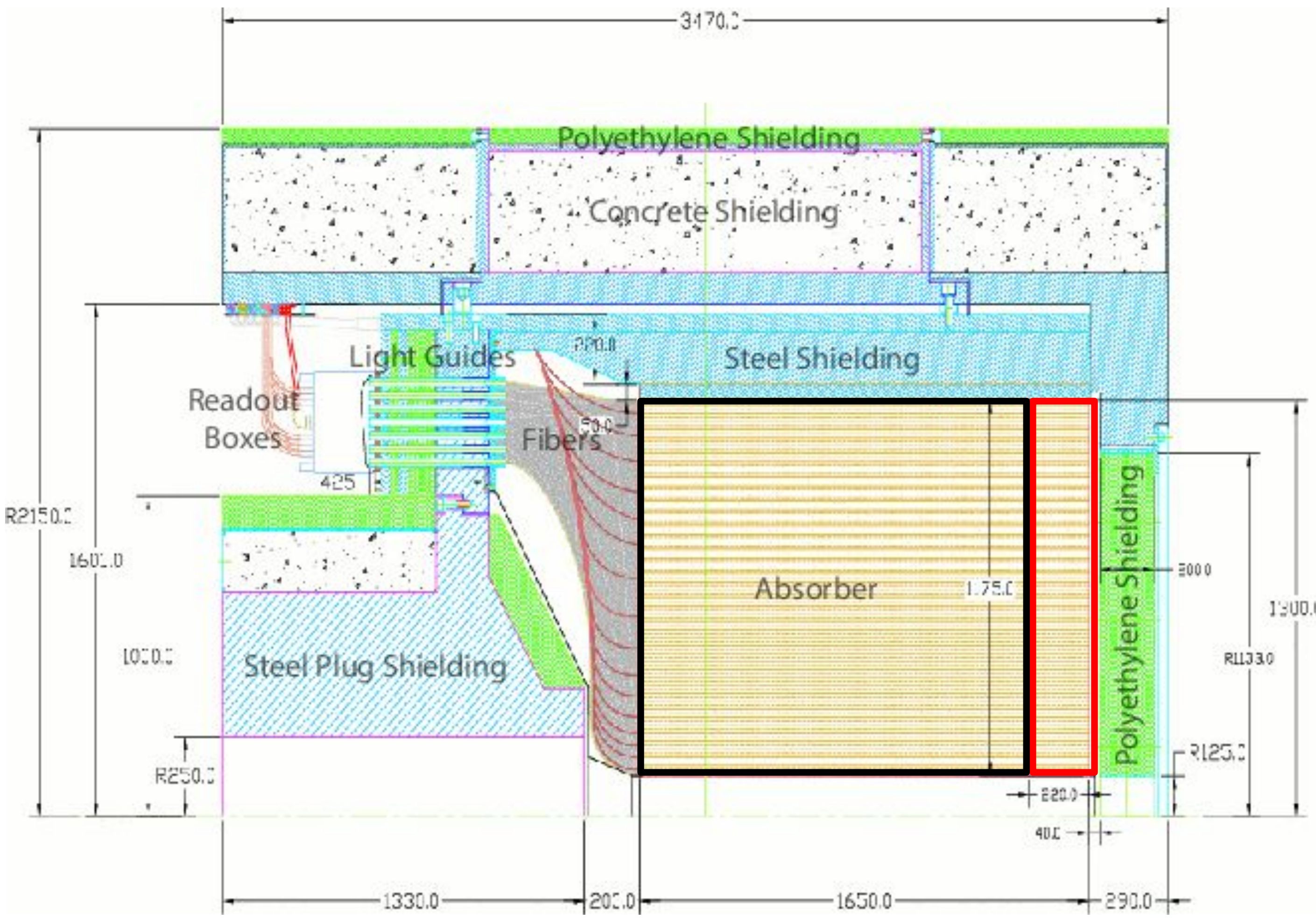

Figure 4.1: *Annotated diagram of the CMS hadron forward calorimeter. The yellow region represents the quartz fibers embedded in the steep absorber. The portion of this absorber outlined in black contains both long and short fibers, while the portion outlined in red contains only long fibers*[1].

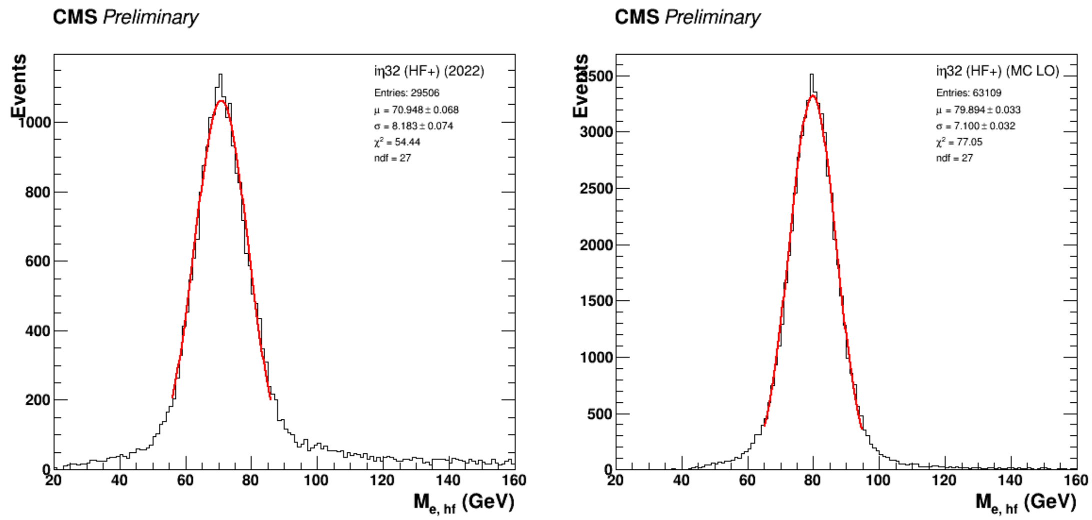

Figure 4.2: *Di-electron mass distribution for $i\eta = 32$ in 2022 data (left) and MC (right). The red curve is a Gaussian fit with the parameters shown on the plots.*

over $i\phi$. The same is done for each $i\phi$ region, inclusive over the HF $i\eta$ regions. An example of this distribution for 2022 data and MC is shown in Figure 4.2. The mean in both data and MC is lower than the nominal Z boson mass of 91.1880 GeV because only the energy from the L fibers is used to calculate the 4-vector of the HF electron[2]. The distributions are fitted with a Gaussian function and the mean is extracted, yielding a single value for each $|i\eta|$ and $i\phi$ region which can be compared between data and MC. The value of invariant Z boson mass in each $|i\eta|$ and $i\phi$ region is plotted in Figure 4.3.

As can be seen in Figure 4.3, there is a significant difference between the reconstructed di-electron mass in data and MC. This difference varies based on the $i\eta$ value but it is relatively constant across all $i\phi$ values. Therefore, the corrections will need to be derived for each $i\eta$ region. The corrected L fiber (L_{corr}) energy will take the following form,

$$L_{corr} = L \times \alpha(\eta) \times \frac{MC(\eta)}{Data(\eta)}, \tag{4.2}$$

where L is the original, uncorrected L fiber energy, $\alpha(\eta)$ is an η-dependent scaling factor, and $\frac{MC(\eta)}{Data(\eta)}$ is the ratio of the reconstructed invariant Z mass in MC to data.

The $\alpha(\eta)$ scaling factors are derived empirically. The di-electron mass distribution is

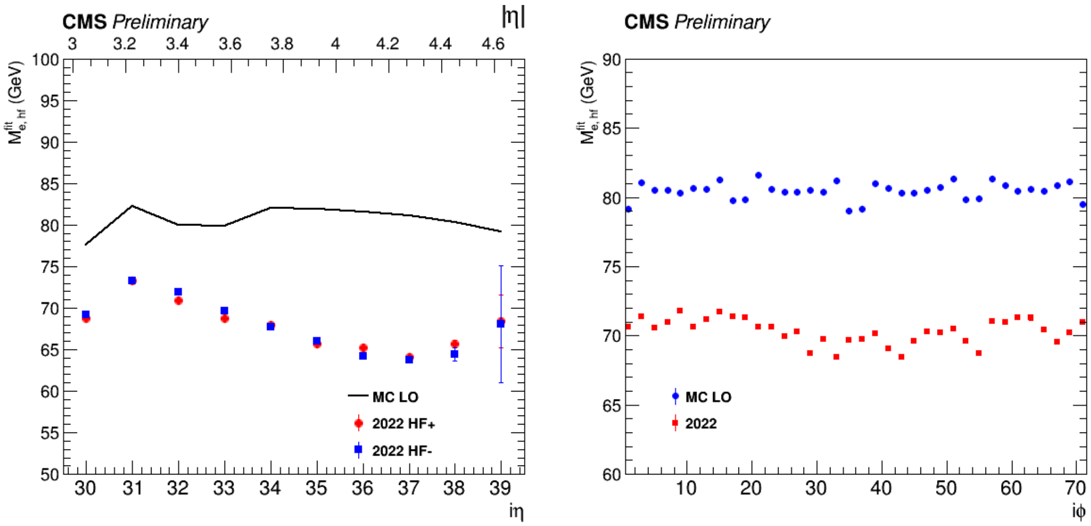

Figure 4.3: *The mean of the di-electron mass distribution plotted in each $i\eta$ (left) and $i\phi$ (right) for 2022 data and Drell-Yan MC.*

plotted for several different values of α. Each distribution is fitted with a Gaussian and the peak values are then fitted to a line as a function of α. This equation can then be solved for when the di-electron mass peak equals the MC value. Consequently, a unique correction to the L fiber energy is calculated for each $i\eta$ region. Figure 4.4 shows the reconstructed Z mass as a function of $|i\eta|$ and $i\phi$, with the radiation damage corrections applied. The lack of statistics at large values of $i\eta$ reduces the accuracy of the corrections. The technique described here was used to derive corrections during the 2023 data-taking run.

4.2 Online-to-Offline (O2O) framework for HCAL

At CMS, non-event data that results from alignment and calibration algorithms is referred to as “Conditions”. These change over time and are used in the offline event reconstruction algorithms. Each sub-detector is responsible for propagation of these Conditions from the detector to the CMS Condition Database[50]. These are known as the “online” and “offline” locations, respectively. The CMS Condition Database stores the bulk data as binary arrays called “Payloads”. Each Payload is associated with a time frame called an “Interval Of Validity” (IOV), which uses the CMS Run numbers to define when that specific Condition

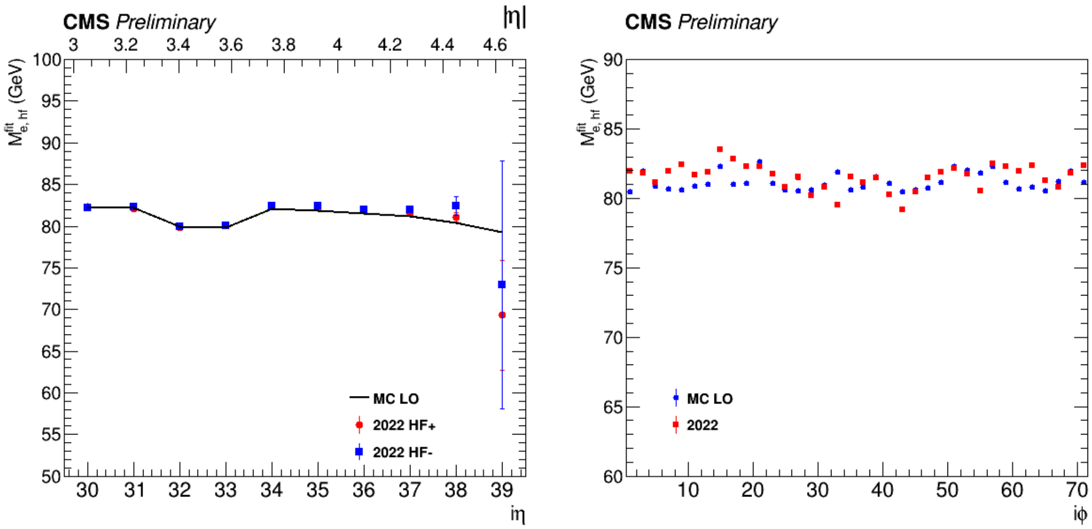

Figure 4.4: *After radiation damage corrections have been applied to the L fiber energy in 2022 data, the mean of the di-electron mass distribution plotted in each $i\eta$ (left) and $i\phi$ (right) now matches what is expected from Drell-Yan MC. A lack of statistics at large $i\eta$ (38, 39) limits the accuracy of the corrections in those regions*

was valid. Finally, successive Payloads and IOVs are stored together in an object called a "Tag".

The set of Conditions relevant to the HCAL sub-detector is managed by a framework of scripts called "Online-to-Offline" (O2O). The input to these scripts is a list of Tags and an associated Run number. These Tags contain the current, valid set of Conditions, specified by the Payload at the IOV associated with the provided Run number. The O2O scripts compare the valid Payloads on each Tag to the latest Conditions stored in the Condition Database. If they are different, the valid Payload is appended to the Tag in the Conditions Database. In this way, the online state of the detector is constantly monitored and kept synchronized with the Condition Database, ensuring the accurate Conditions are available for offline event reconstruction. Figure 4.5 shows an example of how the input and target tags in the are compared in the Conditions Database prior to appending.

The development and maintenance of this framework is essential to keeping HCAL properly calibrated for data taking. Several improvements were made to the scripts and workflow. First, a series of protections were added. These include explicit checks for the following details:

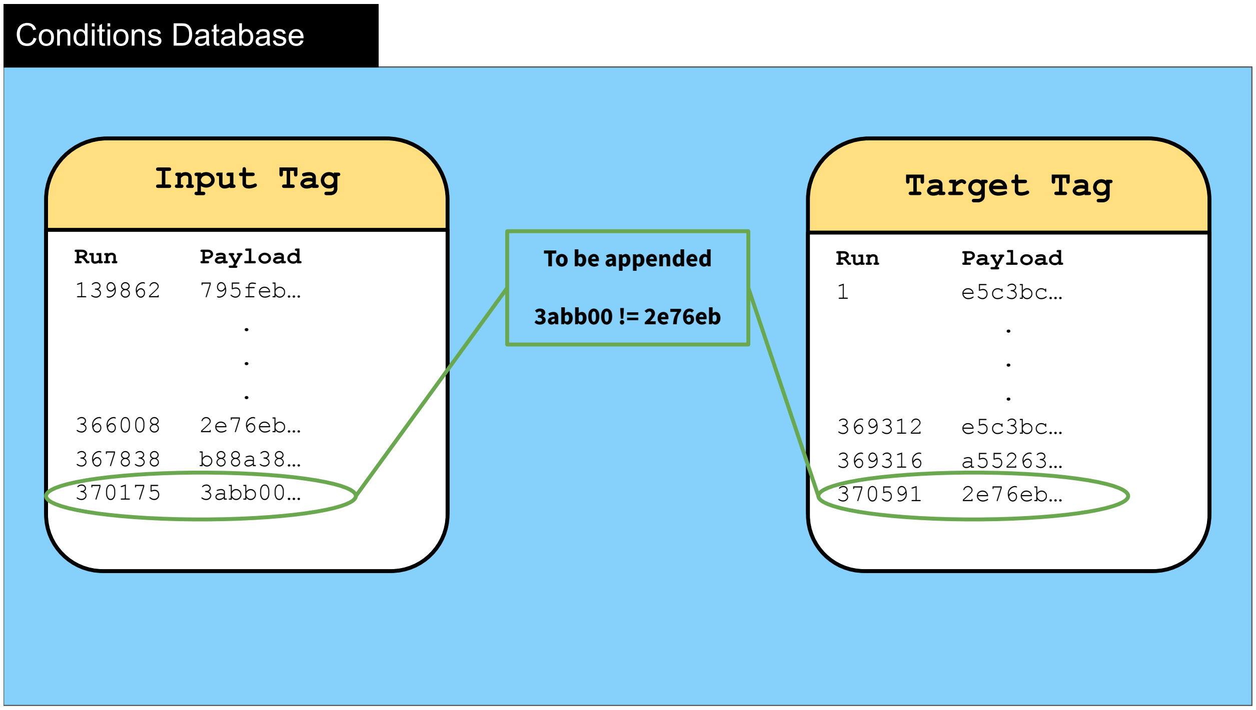

Figure 4.5: *An example of the appending procedure in O2O. In this example, the most recent payload on the input tag (left) will be appended to the target tag (right) because they are not the same.*

- necessary files and directories exist
- user has necessary permissions for the files and directories
- user has authentication to access the Conditions Database
- the specified set of Tags exists
- appropriate proxies are present to allow for web connections

Several different people located in various regions around the globe are in charge of the different Conditions and need to be aware of any changes being made. Functionality was added to the scripts which allow emails to be send, alerting the necessary parties to any important actions. A bot account was created as well to provide updates in a dedicated channel in a messaging application.

A procedure was implemented which allows the scripts to perform a “Dry Run”. In this scenario the scripts are run exactly as normal however they stop prior to making an appends to the Conditions Database. Instead, an organized description of the expected actions is send via email and instant message to the relevant people. During data taking this proved to be very useful as many mistakes or bugs were caught prior to execution of the scripts. The usage of this procedure is now standard. Lastly, the scripts were optimized using parallel processing to ensure that they would execute in less than a minimum amount of time.

Chapter 5

Conclusion

This thesis presents a search for the rare decay of a W boson to three charged pions. A novel approach to reconstruct the triplet of three pions from the combination of two tau candidates and one isolated track objects is used thereby increasing the signal yield. To handle the increase in background contamination from, a multiclass neural network was trained which successfully converges to a minimum value and achieves excellent separation of the three classes. A more robust method is developed using this neural network to better characterize the large and complex QCD multijet background. Advancements in reconstruction and trigger algorithms are taken advantage of to further increase the sensitivity to the signal. Finally, all available data is used to lower the statistical uncertainties in the extraction of the expected upper limits on the branching fraction for this rare decay.

The combined expected upper limit for all $199.5\,\mathrm{fb}^{-1}$ of data is $\mathcal{B}(W \to 3\pi) < 0.XXXX$. This is the most stringent limit on this decay to date and is beginning to reach the lower bounds of the speculative theory estimates on this decay. Ongoing advancements in tau candidate identification and discrimination as well as the addition of the remainder of the Run 3 data to be collected by the CMS experiment at the CERN LHC should make this decay a promising one to study.

Bibliography

[1] Aram "Hayrapetyan and others". Development of the CMS detector for the CERN LHC run 3. *JINST*, 19(05):P05064, 2024. doi: 10.1088/1748-0221/19/05/P05064. URL https://cds.cern.ch/record/2870088. arXiv: 2309.05466 Number: CMS-PRF-21-001, CERN-EP-2023-136, CMS-PRF-21-001-003 tex.collaboration: CMS.

[2] S. Navas and others. Review of particle physics. *Physical Review D: Particles and Fields*, 110(3):030001, 2024. doi: 10.1103/PhysRevD.110.030001. tex.collaboration: Particle Data Group.

[3] Mark Thomson. Modern particle physics. In *Modern particle physics*. Cambridge University Press, Cambridge, United Kingdom ;, 2013. ISBN 978-1-107-03426-6. tex.lccn: 2013002757.

[4] F. (Franz) Mandl and G. (Graham) Shaw. Quantum field theory. In *Quantum field theory*. Wiley, Hoboken, N.J, 2nd edition. edition, 2010. ISBN 978-0-471-49683-0. tex.lccn: 2010000255.

[5] Eugene D. Commins and Philip H. Bucksbaum. Weak interactions of leptons and quarks. In *Weak interactions of leptons and quarks*. Cambridge University Press, Cambridge [Cambridgeshire] ;, 1983. ISBN 978-0-521-23092-6. tex.lccn: 82004452.

[6] Roel Aaij et al. Observation of $J/\psi p$ resonances consistent with pentaquark states in $\Lambda_b^0 \to J/\psi K^- p$ decays. *Phys. Rev. Lett.*, 115:072001, 2015. doi: 10.1103/PhysRevLett.115.072001. URL https://cds.cern.ch/record/2033887.

[7] Ewa Lopienska. The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022, 2022. URL https://cds.cern.ch/record/2800984. General Photo.

[8] Oliver Sim Brüning, Paul Collier, P Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004. doi: 10.5170/CERN-2004-003-V-1. URL https://cds.cern.ch/record/782076.

[9] CMS Collaboration. CMS luminosity - public results, 2024. URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[10] Tai Sakuma and Thomas McCauley. Detector and event visualization with SketchUp at the CMS experiment. *Journal of Physics: Conference Series*, 513(2):022032, June 2014. doi: 10.1088/1742-6596/513/2/022032. URL https://dx.doi.org/10.1088/1742-6596/513/2/022032.

[11] I. Neutelings. Graphics with tikz in latex, 2025. URL https://tikz.net/category/physics/particle-physics/.

[12] CMS Tracker Group of the CMS Collaboration. The CMS Phase-1 pixel detector upgrade. *Journal of Instrumentation*, 16(02):P02027, February 2021. doi: 10.1088/1748-0221/16/02/P02027. URL https://dx.doi.org/10.1088/1748-0221/16/02/P02027.

[13] S. Chatrchyan and others. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008. doi: 10.1088/1748-0221/3/08/S08004.

[14] Jay Richard Dittmann. Hcal tower depth segmentation figures, 2024. URL https://cms-docdb.cern.ch/cgi-bin/DocDB/ShowDocument?docid=13342.

[15] G Baiatian and others. Design, performance and calibration of the CMS forward calorimeter wedges. Technical Report CMS-NOTE-2006-044, CERN, Geneva, 2008. URL https://cds.cern.ch/record/951395.

[16] C. M. S. Collaboration. Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays. *Journal of Instrumentation*, 5(03):T03021, March 2010. doi:

10.1088/1748-0221/5/03/T03021. URL https://dx.doi.org/10.1088/1748-0221/5/03/T03021.

[17] M. Abbas and others. Performance of a triple-GEM demonstrator in pp collisions at the CMS detector. *Journal of Instrumentation*, 16(11):P11014, November 2021. doi: 10.1088/1748-0221/16/11/P11014. URL https://dx.doi.org/10.1088/1748-0221/16/11/P11014. Publisher: IOP Publishing.

[18] N. Ellis. Trigger and data acquisition, 2010. URL https://arxiv.org/abs/1010.2942. _eprint: 1010.2942.

[19] Joao Varela. The CMS trigger system. In *Proceedings of international europhysics conference on high energy physics — PoS(HEP2005)*, volume 021, page 392, 2007. doi: 10.22323/1.021.0392.

[20] Somnath Choudhury. Performance of the High-Level Trigger System at CMS in LHC Run-2. *IEEE Transactions on Nuclear Science*, 68(8):2035–2042, 2021. doi: 10.1109/TNS.2021.3087618.

[21] The CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *Journal of Instrumentation*, 9(10):P10009, October 2014. doi: 10.1088/1748-0221/9/10/P10009. URL https://dx.doi.org/10.1088/1748-0221/9/10/P10009.

[22] Wolfgang Adam, Boris Mangano, Thomas Speer, and Teddy Todorov. Track Reconstruction in the CMS tracker. Technical report, CERN, Geneva, 2006. URL https://cds.cern.ch/record/934067.

[23] R Frühwirth, Wolfgang Waltenberger, and Pascal Vanlaer. Adaptive Vertex Fitting. Technical report, CERN, Geneva, 2007. URL https://cds.cern.ch/record/1027031.

[24] A.M. Sirunyan and others. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017. doi: 10.1088/1748-0221/12/10/P10003. URL https://cds.cern.ch/record/2270046. _eprint: 1706.04965.

[25] Wolfgang Adam, R Frühwirth, Are Strandlie, and T Todor. Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC. Technical report, CERN, Geneva, 2005. URL https://cds.cern.ch/record/815410.

[26] Gavin P. Salam. Towards jetography. *The European Physical Journal C*, 67(3–4): 637–686, May 2010. ISSN 1434-6052. doi: 10.1140/epjc/s10052-010-1314-6. URL http://dx.doi.org/10.1140/epjc/s10052-010-1314-6. Publisher: Springer Science and Business Media LLC.

[27] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-ktjet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, April 2008. ISSN 1029-8479. doi: 10.1088/1126-6708/2008/04/063. URL http://dx.doi.org/10.1088/1126-6708/2008/04/063. Publisher: Springer Science and Business Media LLC.

[28] A.M. Sirunyan and others. Pileup mitigation at CMS in 13 TeV data. *Journal of Instrumentation*, 15(09):P09018, September 2020. doi: 10.1088/1748-0221/15/09/P09018. URL https://dx.doi.org/10.1088/1748-0221/15/09/P09018.

[29] Daniele Bertolini, Philip Harris, Matthew Low, and Nhan Tran. Pileup per particle identification. *Journal of High Energy Physics*, 2014(10), October 2014. ISSN 1029-8479. doi: 10.1007/jhep10(2014)059. URL http://dx.doi.org/10.1007/JHEP10(2014)059. Publisher: Springer Science and Business Media LLC.

[30] David Krohn, Matthew D. Schwartz, Matthew Low, and Lian-Tao Wang. Jet cleansing: Separating data from secondary collision induced radiation at high luminosity. *Physical Review D*, 90(6), September 2014. ISSN 1550-2368. doi: 10.1103/physrevd.90.065020. URL http://dx.doi.org/10.1103/PhysRevD.90.065020. Publisher: American Physical Society (APS).

[31] CMS Collaboration. Performance of tau-lepton reconstruction and identification in CMS. *Journal of Instrumentation*, 7(01):P01001–P01001, January 2012. ISSN 1748-

0221. doi: 10.1088/1748-0221/7/01/p01001. URL http://dx.doi.org/10.1088/1748-0221/7/01/P01001. Publisher: IOP Publishing.

[32] The CMS collaboration. Reconstruction and identification of τ lepton decays to hadrons and ν_τ at cms. *Journal of Instrumentation*, 11(01):P01019, jan 2016. doi: 10.1088/1748-0221/11/01/P01019. URL https://dx.doi.org/10.1088/1748-0221/11/01/P01019.

[33] Albert M Sirunyan et al. Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13\,\text{TeV}$. *Journal of Instrumentation*, 13(10):P10005, oct 2018. doi: 10.1088/1748-0221/13/10/P10005. URL https://dx.doi.org/10.1088/1748-0221/13/10/P10005.

[34] Armen Tumasyan and others. Identification of hadronic tau lepton decays using a deep neural network. *JINST*, 17:P07023, 2022. doi: 10.1088/1748-0221/17/07/P07023. _eprint: 2201.08458.

[35] Geoffrey T J Arnison et al. Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540 GeV$. *Phys. Lett. B*, 122:103–116, 1983. doi: 10.5170/CERN-1983-004.123. URL https://cds.cern.ch/record/142059.

[36] M others Banner. Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN $\overline{p}$p collider. *Phys. Lett. B*, 122:476–485, 1983. doi: 10.1016/0370-2693(83)91605-2. URL https://cds.cern.ch/record/142759.

[37] Michelangelo Mangano and Tom Melia. Rare exclusive hadronic W decays in a t-tbar environment, 2014. URL https://arxiv.org/abs/1410.7475. arXiv: 1410.7475 [hep-ph].

[38] Tom Melia. Exclusive hadronic W decay: $W \to \pi\gamma$ and $W \to \pi^+\pi^+\pi^-$. *Nucl. Part. Phys. Proc.*, 273-275:2102–2106, 2016. doi: 10.1016/j.nuclphysbps.2015.09.341. URL https://cds.cern.ch/record/2263121.

[39] Albert M Sirunyan and others. Search for W boson decays to three charged pions. *Physical Review Letters*, 122:151802, 2019. doi: 10.1103/PhysRevLett.122.151802. URL http://cds.cern.ch/record/2655339. arXiv: 1901.11201 Number: CMS-SMP-18-009, CERN-EP-2019-001, CMS-SMP-18-009-004 tex.collaboration: CMS.

[40] Armen Tumasyan and others. Identification of hadronic tau lepton decays using a deep neural network. *JINST*, 17:P07023, 2022. doi: 10.1088/1748-0221/17/07/P07023. arXiv: 2201.08458 [hep-ex] Number: CMS-TAU-20-001, CERN-EP-2021-257 tex.collaboration: CMS.

[41] Aram Hayrapetyan et al. The CMS statistical analysis and combination tool: COMBINE. Submitted to *Comput. Softw. Big Sci.*, 2024.

[42] Michał Czakon, David Heymes, Alexander Mitov, Davide Pagans, Ioannis Tsinikos, and Marco Zaro. Top-pair production at the LHC through NNLO QCD and NLO EW. *Journal of High Energy Physics*, 2017(10), October 2017. doi: 10.1007/jhep10(2017)186. URL https://doi.org/10.1007/JHEP10(2017)186. Publisher: Springer Science and Business Media LLC.

[43] François Chollet and others. Keras, 2015. URL https://keras.io.

[44] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017. URL https://arxiv.org/abs/1412.6980.

[45] A. M. Sirunyan et al. Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$TeV in 2015 and 2016 at CMS. *Eur. Phys. J. C*, 81:800, 2021. doi: 10.1140/epjc/s10052-021-09538-2.

[46] 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, CERN, 2018. URL https://cds.cern.ch/record/2621960.

[47] CMS Collaboration. CMS luminosity measurement for the 2018 data-taking period at

$\sqrt{s}$ = 13TeV. CMS Physics Analysis Summary CMS-PAS-LUM-18-002, CERN, 2019. URL https://cds.cern.ch/record/2676164.

[48] The ATLAS and CMS Collaborations and The LHC Higgs Combination Group. Procedure for the LHC Higgs boson search combination in Summer 2011. Technical report, CERN, Geneva, 2011. URL https://cds.cern.ch/record/1379837.

[49] Albert M Sirunyan et al. Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13$ TeV. *JINST*, 15:P05002, 2020. doi: 10.1088/1748-0221/15/05/P05002. URL https://cds.cern.ch/record/2691403.

[50] Giacomo Maria Govi, Salvatore Di Guida, Miguel Ojeda-Sandonis, Andreas Pfeiffer, and Roland Sipos. The CMS condition database system. Technical Report 4, CERN, Geneva, 2015. URL https://cds.cern.ch/record/2121260. Number: CMS-CR-2015-073 tex.collaboration: CMS.

Chapter 6

Appendix A

For the initial generation of the W boson, we used two existing next-to-next-to-leading order gridpacks. One gridpack produces exclusively $W^+ \to \bar{\mu}^+ \nu_\mu$ while the other generates exclusively $W^- \to \mu^- \bar{\nu}_\mu$. The cross section is 111 400 pb for the inclusive W^+ production and 81 800 pb for inclusive W^- production. These gridpacks generate the W boson to the highest level of accuracy available. This step produces a file in LHE format with only the initially generated particles present. This can be up to five particles: the W boson, the muon, the neutrino, and up to two additional quarks or gluons. The muon and neutrino are manually removed from this file before passing it as an input to `PYTHIA8`. The `PYTHIA8` settings used to produce this signal sample are shown in Table 6.1. Unfortunately, the `PYTHIA8` generator ignores spin effects that would arise from the W boson having a spin of 1. Effectively, the decay proceeds as if the W boson was a scalar boson rather than a vector boson. Future work on this analysis would be encouraged to properly implement the W boson decay as a vector boson to three charged pions in a higher order event generator.

Decay Settings
24:onMode = off
24:AddChannel = on 1.0 100 211 211 -211
SpaceShower:pTmaxMatch = 1
TimeShower:pTmaxMatch = 1
ParticleDecays:allowPhotonRadiation = on
TimeShower:QEDshowerByL = off
TimeShower:QEDshowerByOther = off
BeamRemnants:hardKTOnlyLHE = on
BeamRemnants:primordialKThard = 2.225001
SpaceShower:dipoleRecoil = 1
Pythia8 Common Settings
Tune:preferLHAPDF = 2
Main:timesAllowErrors = 10000
Check:epTolErr = 0.01
Beams:setProductionScalesFromLHEF = off
SLHA:minMassSM = 1000.
ParticleDecays:limitTau0 = on
ParticleDecays:tau0Max = 10
ParticleDecays:allowPhotonRadiation = on
Pythia8 CP5 Settings
Tune:pp 14
Tune:ee 7
MultipartonInteractions:ecmPow = 0.03344
MultipartonInteractions:bProfile = 2
MultipartonInteractions:pT0Ref = 1.41
MultipartonInteractions:coreRadius = 0.7634
MultipartonInteractions:coreFraction = 0.63
ColourReconnection:range = 5.176
SigmaTotal:zeroAXB = off
SpaceShower:alphaSorder = 2
SpaceShower:alphaSvalue = 0.118
SigmaProcess:alphaSvalue = 0.118
SigmaProcess:alphaSorder = 2
MultipartonInteractions:alphaSvalue = 0.118
MultipartonInteractions:alphaSorder = 2
TimeShower:alphaSorder = 2
TimeShower:alphaSvalue = 0.118
SigmaTotal:mode = 0
SigmaTotal:sigmaEl = 21.89
SigmaTotal:sigmaTot = 100.309
PDF:pSet = LHAPDF6:NNPDF31_nnlo_as_0118

Table 6.1: *Signal generation settings*