Probing polarised production of two same-charge $W^{\pm}$ bosons produced in association with two jets with the ATLAS detector

Luka Nedic

Supervisors: Dr. Karolos Potamianos, Prof. Chris Hays

15 January 2025

Abstract

The Standard Model's electroweak symmetry breaking mechanism describes the nature of longitudinally polarised weak bosons. The Standard Model also predicts a special class of interactions involving weak bosons, which include quartic and trilinear vertices, referred to as Vector Boson Scattering (VBS). Experimental measurements of polarised VBS test both the gauge structure of the electroweak sector and the electroweak symmetry breaking mechanism. Furthermore, the cross-section for longitudinally polarised VBS is highly sensitive to the Higgs couplings and the gauge-boson self-couplings; the cross-section for longitudinally polarised VBS diverges without exact cancellations from interactions with the Higgs boson.

This thesis presents an experiment probing the nature of polarised VBS at the Large Hadron Collider. The experiment measures the fiducial cross-section for the electroweak production of two polarised same-sign W bosons in association with two jets. The cross-section is measured in the leptonic decay channels of the W bosons. The measurement is performed using 139 fb^{-1} of proton-proton collision data at a centre-of-mass energy $\sqrt{s} = 13$ TeV recorded by the ATLAS detector.

The experiment measures the electroweak production of two-same sign W bosons, with at least one longitudinally polarised boson in the W boson and proton-proton centre-of-mass frames, with an observed (expected) significance of 3.73 (4.20) and 3.13 (3.59) standard deviations, respectively. The experiment also measures, at a 95% confidence upper limit, the observed (expected) cross-sections for the electroweak production of two longitudinally polarised same-sign W bosons in the W boson and proton-proton centre-of-mass frames as 0.617 (0.710) fb and 0.569 (0.648) fb, respectively.

Acknowledgements

I want to express my sincerest gratitude to my supervisor Karolos Potamianos. Thank you for your dedication and wisdom over the past four years. Starting my postgraduate studies during the pandemic could have been a lot more difficult had it not been for your time, support and supervision. I have really enjoyed working with you, and I am very grateful to have been your first student! I hope we get to work together again at some point in the future. I would also like to thank my second supervisor, Chris Hays, for your support throughout my studies and for your time and effort in reviewing my thesis.

To my parents, Maja and Dusko Nedic, my brother, Aleksa, and my grandparents, thank you for all your support. I feel immensely fortunate to have such a caring and supportive family. You have inspired me throughout my life, and I will always be incredibly thankful for everything you have done for me; my achievements are our achievements. I would also like to thank my girlfriend Seren. Your love and support has been incredible and I could not imagine having to finish this thesis without you. I will always be grateful for you.

A special thank you to my closest friends, Richard and Kate. I appreciate all of your support since all the way back to the start of my undergraduate studies in Durham. I would not be here without you. Finally, I would like to acknowledge all the friends I made in Oxford. I have been very fortunate to meet lots of amazing people who made my time in Oxford feel special. The memories I have made will stay with me forever.

Contents

1 The Standard Model and Vector Boson Scattering

At the present time, the most accurate and unified theory describing the phenomenology of particle physics is the Standard Model. The theory has been extensively tested and is foremost consistent with experimental data: almost all of its predictions have now been verified. The model is also incomplete: gravity, dark matter, neutrino masses, and the matter/anti-matter asymmetry of the universe are all unaccounted for by the model. There exist many untested Beyond-the-Standard Model (BSM) theories, but there is no indication as to whether any of these models will be confirmed a priori; the near future of theory will be shaped by experimental discovery.

To motivate the experiment presented by this thesis, this chapter will outline the theoretical foundations of the Standard Model, beginning with a general overview of the model and its history. Following this, this chapter will give a detailed outline of the pertinent parts of the theory tested by this experiment: the Electroweak theory and the Higgs mechanism. The final section of this chapter will apply the theory to the phenomenology of vector boson scattering.

1.1 The Standard Model

The Standard Model is an $SU(3)_C \times SU(2)_L \times U(1)_Y$ renormalisable gauge theory, defined explicity in Section 1.2. The theory describes the interactions between the fundamental constituents of matter with three of the four fundamental forces: the electromagnetic, weak and strong forces. The model is characterised by 17 fundamentally distinct particles, some of which have multiple types based on electric or colour charge. Each of these particles has been experimentally detected, with the last being the Higgs boson discovered in 2012 [1]. The theoretical foundations of the model were developed through the 20th century based on a number of experimental and theoretical discoveries.

1.1.1 History

The origins of the Standard Model begin in the early 20th century with the development of Quantum Field Theory (QFT). Early attempts to unify classical quantum mechanics and special relativity led to the postulation of antiparticles, first experimentally observed in 1932 [2]. Antiparticles introduce pair production and annihilation, which requires multi-particle states. When combined with quantum mechanics, the changing numbers of particles within these multi-particle states result in paradoxical predictions, often attributed to violations of probability-conserving currents. The solution is to introduce quantum fields as a method of modelling systems with many particles, which ultimately led to the development of QFT.

Following QFT, during the mid-20th century, came the development of Quantum Electrodynamics [3] (QED). QED applies gauge invariance to model electromagnetic interactions. Initially, after its development, QED was found to give divergences for perturbative calculations [4]. It was discovered, however, that these divergences could be regulated, leading to a crucial component of the Standard Model: renormalisation. After the success of QED, the concept of gauge invariance was extended to non-abelian groups with the

development of Yang–Mills theory [5]. The Standard Model is an $SU(3) \times SU(2) \times U(1)$ Yang-Mills gauge theory that combines Quantum Chromodynamics (QCD) with the Glashow–Weinberg–Salam theory [6, 7] (GWS) as a unified theory.

Finally, from experiment, it is understood that the weak force is short-range, which requires massive gauge bosons. Yang-Mills theories however, predict massless gauge bosons. The success of gauge invariance in modelling QED and QCD led to the development of the Higgs mechanism, which introduces mass terms to gauge bosons through spontaneous symmetry breaking [8, 9, 10]: the final component of the Standard Model.

1.1.2 Overview

Within the Standard Model, there are two general classifications of particles, categorised by the particle's spin: fermions and bosons. Fermions are integer-odd spin-half particles and are described as the building blocks of matter. Each fermion has an associated anti-fermion, differing by having an opposite charge. Fermions can be further categorised into two sub-classifications: quarks and leptons. Both interact through the weak force and electromagnetism. Quarks, however, differ by also interacting with the strong force and carrying a colour charge.

There are, in total, 12 discovered fermions: six quarks and six leptons. Quarks and leptons can each be grouped into pairs, and the pairs are categorised into three generations. The properties of the quark and lepton pairs are identical between generations except for their mass; higher-generation fermions are heavier. The apparent symmetry between the two types of fermions is not understood. At the current time, the symmetry could be the manifestation of some larger undiscovered symmetry or entirely coincidental. The Standard Model itself does not impose any conditions on the number of fermions, and it is currently a parameter of experiment.

Bosons are integer-even spin particles and play an intricate role in the interactions of matter. As with fermions, bosons can be further classified into two categories: gauge bosons and scalar bosons. The gauge bosons are spin-1 particles and are interpreted as the carriers of the fundamental forces. Within the Standard Model, each force has an associated set of gauge bosons, and the interactions of each force with matter are described by the interactions of matter with the respective gauge bosons of the force. There are in total 12 experimentally observed gauge bosons: 8 gluons, the $W^{\pm}$ and Z bosons, and the photon.

Scalar bosons are spin-0 particles. There is currently only one experimentally confirmed scalar boson: the Higgs boson. Unlike the gauge bosons, the Higgs is not interpreted as a force carrier but rather as a consequence of the Higgs mechanism that introduces mass to the weak gauge bosons. At the time of writing, there is still significant ongoing research to understand the properties of the Higgs and whether the discovered scalar boson is the Higgs boson of the Standard Model. A full overview of the Standard Model particles and their measured experimental properties is shown in Fig. 1.

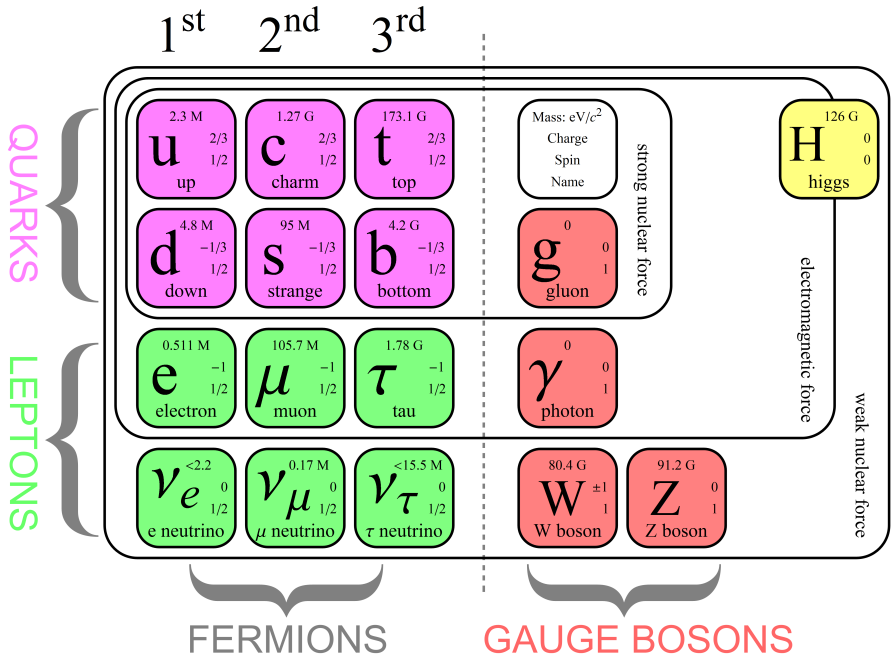


Figure 1: The fundamental particles of the Standard Model and their properties. Figure taken from [11].

1.2 Gauge Theory

One of the major foundations of the Standard Model is gauge theory, a special class of field theories invariant under local symmetry transformations. Within the Standard Model, gauge theories are used to describe the interactions of the fundamental forces of nature. The underlying concepts are crucial to the theory being tested by this experiment. To begin with, this section will overview the fundamentals of gauge theory and its application to the Standard Model. The theory will be presented in terms of classical fields prior to quantisation. Whilst the quantisation procedure is central to deriving predictions, the phenomenology arising from gauge theory motivating this experiment can be understood in terms of classical fields. It is, therefore, noted that the theory presented in this and the subsequent sections must be quantised and renormalised to calculate physical predictions. The relevant literature related to the quantisation process will be referenced.

1.2.1 Abelian Gauge Theory and QED

Global symmetries lead to conserved currents and conservation laws as per Noether's theorem [12]; if the equations of motion of a system are invariant under a set of global transformations, then there exists a corresponding conservation law. The converse states that the conservation laws observed in a system imply that the equations of motions describing the system remain the same under some transformation, such as a rotation. The existence of conserved quantities has a profound meaning on the physics describing a system, manifesting as global symmetries.

The main requirement of a gauge theory is invariance of the Lagrangian under local symmetry transformations. A local symmetry is a stricter condition and imposes a much

larger set of symmetries on the system; global symmetries are a subset of a local symmetry. A local symmetry also introduces additional degrees of freedom into a system and could be viewed as an over-parametrisation. This section will demonstrate, however, analogously to global symmetries, that imposing local symmetries as a condition has a profound effect on the physical system and leads to a physical description of massless vector bosons. The converse is that massless vector bosons imply local gauge symmetries. To illustrate this concept, this section will overview the construction of an abelian gauge theory leading to QED.

The Lagrangian for a free Dirac fermion is given by,

$$\mathcal{L}_0 = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi, \tag{1}$$

where $\psi(x) = \psi$ is a Dirac field and m is the mass of the Dirac fermion. This Lagrangian is invariant under global $U(1)$ transformations of the fields ψ, given by,

$$\psi(x) \to \psi'(x) = e^{iq\theta}\psi(x), \tag{2}$$

where q, θ are constants. However, this Lagrangian is not invariant under local transformations: a space-time dependence on the parameter $\theta \to \theta(x)$ breaks the invariance at the partial derivatives. To formulate the Lagrangian to be invariant under local symmetry transformations, the solution is to introduce a spin-1 gauge field $A_\mu(x)$ which transforms as,

$$A_\mu(x) \to A'_\mu(x) = A_\mu(x) + \frac{1}{e}\partial_\mu\theta(x), \tag{3}$$

where the parameter $\theta(x)$ characterising the local symmetry transformation is absorbed into the gauge field $A_\mu(x)$ as a gauge function. The terms gauge invariance and invariance under local symmetry transformation are, in this context, synonymous. By defining the covariant derivative as,

$$D_\mu\psi(x) = [\partial_\mu - ieqA_\mu(x)]\psi(x), \tag{4}$$

which transforms as,

$$D_\mu\psi(x) \to (D_\mu\psi)' = e^{iq\theta(x)}D_\mu\psi(x), \tag{5}$$

the Lagrangian can be formulated in a gauge invariant way,

$$\mathcal{L} = i\bar{\psi}\gamma^\mu D_\mu\psi - m\psi\psi. \tag{6}$$

This Lagrangian can be expanded and rewritten in terms of the original Lagrangian for a free Dirac fermion plus an additional term identified as an interaction between the Dirac fermions ψ and the gauge field A_μ, given by

$$\mathcal{L} = \mathcal{L}_0 + \underbrace{eqA_\mu\bar{\psi}\gamma^\mu\psi}_{\text{interaction term}}. \tag{7}$$

For the gauge field A_μ to correspond to a propagating field, it must have a kinetic term. The only additional gauge invariant combinations of fields which can be added to the Lagrangian, which include A_μ and its partial derivatives, are permutations of the anti-symmetric field tensor defined as,

$$F_{\nu\mu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{8}$$

This tensor is more commonly known as the electromagnetic field tensor. Adding the gauge invariant kinetic term, otherwise known as the Maxwell equations, the full QED Lagrangian is given by,

$$\mathcal{L}_{\text{QED}} = i\bar{\psi}\gamma^\mu D_\mu \psi - m\psi\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} = \mathcal{L}_0 + \underbrace{eqA_\mu\bar{\psi}\gamma^\mu\psi}_{\text{EM interaction}} - \underbrace{\frac{1}{4}F_{\mu\nu}F^{\mu\nu}}_{\text{Maxwell equations}}. \tag{9}$$

Contextually, constructing a $U(1)$ gauge invariant Lagrangian led to a theoretical description of electromagnetic interactions: conditioning a local symmetry introduced a massless gauge boson. A detailed overview of the quantisation is discussed in [13, 14].

1.2.2 Yang-Mills Theories

The concept of constructing a gauge invariant Lagrangian can be further expanded beyond $U(1)$ symmetries to model more complex phenomenology and field structures. The Yang-Mills theories are a generalisation of the previous section, extending the local symmetries to the non-abelian $SU(n)$ group. The $SU(n)$ group is a symmetry group with matrices satisfying,

$$S \in SU(n) \rightarrow S^{-1} = S^\dagger, \det S = 1, \tag{10}$$

where $\dagger$ denotes the Hermitian conjugate operation. The group is a type of Lie group and forms a Lie algebra such that each $S \in SU(n)$ can be expressed in terms of $n^2 - 1$ independent parameters, defined as,

$$S = \exp(i\theta^a T^a), \tag{11}$$

where T^a are the group generators with commutation relation,

$$[T^a, T^b] = if^{abc}T^c. \tag{12}$$

The procedure for introducing a gauge invariant Lagrangian follows analogously to the abelian case. Each group generator introduces a gauge field A^a_μ, which likewise defines the covariant derivative. The final part is to include gauge invariant kinetic terms for the propagation of the gauge fields. Likewise, akin to QED, the kinetic energy terms are permutations of a field tensor, defined as,

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + gf^{abc}A^b_\mu A^c_\nu. \tag{13}$$

Yang-Mills theories form the basis of QCD and the Electroweak theory. A major difference between an abelian gauge theory and a Yang-Mills theory is the additional term present in the field tensor of Eq. 13. This term introduces self-interactions between gauge bosons and leads to a broader set of phenomenological predictions, which include vector boson scattering, the focus of the experiment presented in this thesis. Literature for quantising Yang-Mills theories is described in Ref. [5, 15], with a detailed overview given by [16].

1.3 Electroweak Theory

This thesis details an experimental measurement of the scattering of longitudinal polarised $W^{\pm}$ bosons, the mediators of the weak force. Within the Standard Model, the weak force is modelled by the electroweak theory and the Higgs mechanism. This section will give an overview of the electroweak theory and its structure, highlighting the pertinent aspects of the theory describing the scattering of vector bosons.

The electroweak theory, also referred to as the Glashow-Weinberg-Salam [6, 7] (GWS) model, is a unified model of electromagnetism and the weak force. The theory is a chiral[1] gauge theory reflecting the maximally parity-violating nature of the weak interactions, first experimentally observed by the Wu Experiment in 1956 [17]. Within the theory, the chiral components of the fermion fields are treated independently. Left/right-handed fermions are described by $SU(2)$ doublets/singlets as shown in Eq. 14.

$$\begin{array}{cc} \overbrace{\begin{pmatrix}\nu_e\\ e\end{pmatrix}_L \begin{pmatrix}\nu_\mu\\ \mu\end{pmatrix}_L \begin{pmatrix}\nu_\tau\\ \tau\end{pmatrix}_L}^{\text{Left-handed leptons}} & \overbrace{\begin{pmatrix}u\\ d\end{pmatrix}_L \begin{pmatrix}c\\ s\end{pmatrix}_L \begin{pmatrix}t\\ b\end{pmatrix}_L}^{\text{Left-handed quarks}} \\ \underbrace{e_R, \mu_R, \tau_R}_{\text{Right-handed leptons}} & \underbrace{u_R, d_R, c_R, s_R, b_R, t_R}_{\text{Right-handed quarks}} \end{array} \tag{14}$$

The theory is a Yang-Mills theory and is invariant under $SU(2) \times U(1)$ gauge transformations. Each of the sub-groups, $SU(2)$ and $U(1)$, introduce a conserved quantity and their respective gauge fields: for $SU(2)$ the weak-isospin T_3 and a triplet of Yang-Mills fields, W^a_μ where $a = 1, 2, 3$, and for $U(1)$ the weak hypercharge Y and single gauge field B_μ. The values of the weak-isospin and hypercharge for the different chiral fermions are shown in Figure 2, and the full electroweak Lagrangian for a general fermion $\psi_{L/R}$ is given by,

$$\mathcal{L}_{EW} = \underbrace{i\bar{\psi}_L \gamma^\mu D_\mu \psi_L}_{\text{left-handed terms}} + \overbrace{i\bar{\psi}_R \gamma^\mu D_\mu \psi_R}^{\text{right-handed terms}} - \underbrace{\frac{1}{4} F^{a\mu\nu} F^a_{\nu\mu} - \frac{1}{4} G^{\mu\nu} G_{\nu\mu}}_{\text{Gauge boson terms}} \quad , \tag{15}$$

where σ^a are the Pauli matrices, which are the group generators of $SU(2)$, and $F^{a\mu\nu}$, $G^{\mu\nu}$ are field tensors as defined in equations 13 and 8. The covariant derivative acting on left-handed fermions is given by the equation,

$$D_\mu \psi_L = (\partial_\mu - i\frac{1}{2} g \sigma^a W^a_\mu - ig' Y B_\mu)\psi_L \tag{16}$$

and the covariant derivative for right-handed fermions is given by,

$$D_\mu \psi_R = (\partial_\mu - ig' Y B_\mu)\psi_R. \tag{17}$$

Within the theory, the physical fields that mediate the electromagnetic and weak inter-

[1] The theory treats left-handed and right-handed particles differently

	Generation			$\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$			$\mathrm{SU}(3)_c$
	1^{st}	2^{nd}	3^{rd}	T_3	Y	q	colour
Leptons	$\begin{pmatrix}\nu_e \\ e^-\end{pmatrix}_L$	$\begin{pmatrix}\nu_\mu \\ \mu^-\end{pmatrix}_L$	$\begin{pmatrix}\nu_\tau \\ \tau^-\end{pmatrix}_L$	$+\frac{1}{2}$	$-\frac{1}{2}$	0	–
				$-\frac{1}{2}$	$-\frac{1}{2}$	-1	–
	e_R^-	μ_R^-	τ_R^-	0	-1	-1	–
Quarks	$\begin{pmatrix}u \\ d\end{pmatrix}_L$	$\begin{pmatrix}c \\ s\end{pmatrix}_L$	$\begin{pmatrix}t \\ b\end{pmatrix}_L$	$+\frac{1}{2}$	$+\frac{1}{6}$	$+\frac{2}{3}$	rgb
				$-\frac{1}{2}$	$+\frac{1}{6}$	$-\frac{1}{3}$	rgb
	u_R	c_R	t_R	0	$+\frac{2}{3}$	$+\frac{2}{3}$	rgb
	d_R	s_R	b_R	0	$-\frac{1}{3}$	$-\frac{1}{3}$	rgb

Figure 2: The Standard Model fermions and their associated quantum numbers. Figure adapted from [18].

actions are combinations of the gauge fields of weak isospin and weak hypercharge. The gauge fields the transforms orthogonally as,

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \tag{18}$$

and

$$\begin{pmatrix}W_\mu^3 \\ B_\mu\end{pmatrix} = \begin{pmatrix}\cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W\end{pmatrix}\begin{pmatrix}Z_\mu \\ A_\mu\end{pmatrix}. \tag{19}$$

where θ_W is defined as the weak mixing angle. Transforming the gauge fields also recovers the electric charge associated with the coupling strength of the electromagnetic interaction and is given by,

$$q = T_3 + \frac{Y}{2}. \tag{20}$$

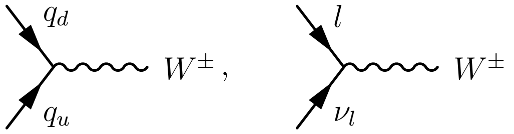
The abelian symmetry group associated with the weak hypercharge allows for the choice of any Y such that the fermion electric charges are correctly accounted for. Substituting the transformations into the Lagrangrian, the electroweak theory predicts three fermionic interactions. Defining the left-handed doublet for an arbitrary fermion as,

$$\psi_L = \begin{pmatrix}\psi_L^1 \\ \psi_L^2\end{pmatrix}, \tag{21}$$

the theory predicts a charge current interaction,

$$\mathcal{L}^{CC}_{int} = \frac{g}{\sqrt{2}}(\bar{\psi}^1_L\gamma^\mu\psi^2_L W^+_\nu + \bar{\psi}^2_L\gamma^\mu\psi^1_L W^-_\mu)$$

$$\Rightarrow$$

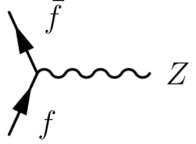


(22)

where q_u, q_d refer to the up and down quarks in the quark doublet pair, and l, ν_l refer to a lepton and lepton-neutrino doublet pair. The theory also predicts a neutral current interaction,

$$\mathcal{L}^Z = \frac{g}{\cos\theta_W}[(T_3 + \sin^2\theta_W)\bar{\psi}_L\gamma^\mu\psi_L + \sin^2\theta_W\bar{\psi}_R\gamma^\mu\psi_R]Z_\mu$$

$$\Rightarrow$$

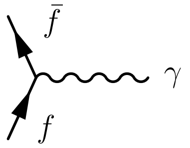


(23)

and the familiar electromagnetic interaction from QED,

$$\mathcal{L}^{\text{EM}} = eq\bar{\psi}\gamma^\mu\psi A_\mu$$

$$\Rightarrow$$



(24)

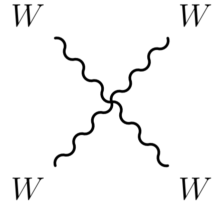
The full phenomenology of fermionic weak interactions requires the Higgs mechanism. The electroweak theory presented here describes interactions in terms of massless fermions and gauge bosons. The inclusion of mass is not inconsequential and introduces additional phenomenology, however the general structure described here for fermionic interactions remains generally true. The electroweak theory also predicts trilinear and quartic interaction between the W and Z gauge bosons: a consequence of the non-abelian gauge structure. Expanding the kinetic term for the field tensors of the weak-isospin gauge field, the interacting Lagrangian is given by

$$\mathcal{L}_{int} = -\frac{1}{2}g\epsilon^{abc}(\partial_\mu W^a_\nu - \partial_\nu W^a_\mu)W^{b\mu}W^{c\nu} - \frac{1}{4}g^2\epsilon^{abc}\epsilon^{ajk}W^b_\mu W^c_\nu W^{j\mu}W^{k\nu}. \tag{25}$$

As with the fermionic interactions, inserting the gauge field transformations leads to a collection of gauge boson interaction terms. The crucial parts of the interactions can be segmented as follows: quartic self-interaction terms such as,

$$\mathcal{L}_{WWWW} = \frac{1}{2}g^2(W^-_\mu W^-_\mu W^+ W^{+\nu} - W^-_\mu W^{+\mu}W^-_\nu W^{+\nu})$$

$$\Rightarrow$$

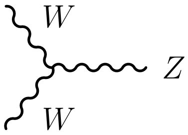


(26)

and trilinear interactions such as,

$$\mathcal{L}^Z = \frac{g}{\cos\theta_W}[(T_3 + \sin^2\theta_W)\bar{\psi}_L\gamma^\mu\psi_L + \sin^2\theta_W\bar{\psi}_R\gamma^\mu\psi_R]Z_\mu \tag{27}$$

$\Rightarrow$



Akin to the above discussion of fermionic interactions, the full model requires the Higgs mechanism, with the general structure of the diagrams presented remaining the same. The trilinear and quartic diagrams shown above form the basis of a set of interactions referred to as vector boson scattering.

1.4 Spontenous Symmetry Breaking

The masses of the $W^\pm$ and Z gauge bosons have been measured to be $m_W = 80.4$ GeV and $m_Z = 91.2$ GeV. However, gauge theories, such as the electroweak theory, predict massless gauge bosons. Furthermore, the chiral structure of the electroweak theory also prohibits fermionic mass terms, which would mix chiral components and violate gauge invariance. Fermions experimentally are also observed to have masses, as shown in Fig. 1. Whilst the electroweak theory correctly describes the nature of the weak force and electromagnetism, the theory models massless particles contradictory to experiment. In order to introduce mass terms into the electroweak theory without explicitly breaking the gauge structure, an additional mechanism is required: spontaneous symmetry breaking. This section will introduce spontaneous symmetry breaking and its application to gauge theories leading to the Higgs mechanism.

1.4.1 Nambu-Goldstone Theorem

Spontaneous symmetry breaking [19, 20] is a mechanism wherein the physical states of a system are asymmetric to symmetries manifest in the equations of motions describing the fields. Within field theory, the spectrum of particles is built from the ground state. If the ground state does not share the symmetries present in the Lagrangian, then neither will the physical spectrum of particles. This is referred to as symmetry breaking. The original symmetry remains in the Lagrangian but does not manifest in the physical states.

This section will introduce the Goldstone model to illustrate the spontaneous symmetry breaking of a continuous global symmetry. The subsequent section will then apply the same mechanism to the electroweak theory, leading to the Higgs mechanism or the spontaneous symmetry breaking of local gauge symmetry. The Lagrangian of the Goldstone model describes a complex scalar field, ϕ, and is defined as,

$$\mathcal{L} = \partial_\mu\phi\partial^\mu\phi^* - V(\phi), \tag{28}$$

where the potential $V(\phi)$ is defined as,

$$V(\phi) = -\mu^2\phi\phi^* + \lambda(\phi\phi^*)^2, \tag{29}$$

and μ and λ are constants. The Lagrangian has a continuous global $U(1)$ symmetry such that a transformation,

$$\phi(x) \to \phi'(x) = e^{i\theta}\phi(x), \tag{30}$$

leaves the Lagrangian invariant. However, the system's physical states can still be asymmetric, as can be seen by finding the ground state. The Hamiltonian energy density for this system is given by,

$$\mathcal{H} = \partial_0\phi\partial_0\phi^* + \nabla\phi\nabla\phi^* + V(\phi). \tag{31}$$

The Hamiltonian is minimised with respect to the complex scalar field ϕ by considering a constant ϕ and minimising $V(\phi)$. There are two possible configurations of minima. If $\mu^2 < 0$ as shown in the left of figure 3, then there is a minimum at 0 and the ground state respects the global symmetry with,

$$\langle 0|\psi|0\rangle = \psi_0 = 0. \tag{32}$$

However, if $\mu^2 > 0$, there are an infinite number of degenerate non-zero minima. The system must choose a minimum; "spontaneous" refers to the system choosing one of many degenerate minima. The non-zero minimum implies a non-zero vacuum expectation value v such that,

$$\langle 0|\psi|0\rangle = \psi_0 = \frac{v}{\sqrt{2}}e^{i\theta} \neq 0, \tag{33}$$

where θ is an arbitrary parameterisation of the degeneracy. In this configuration, the ground state is not invariant under a $U(1)$ transformation. After choosing a vacuum

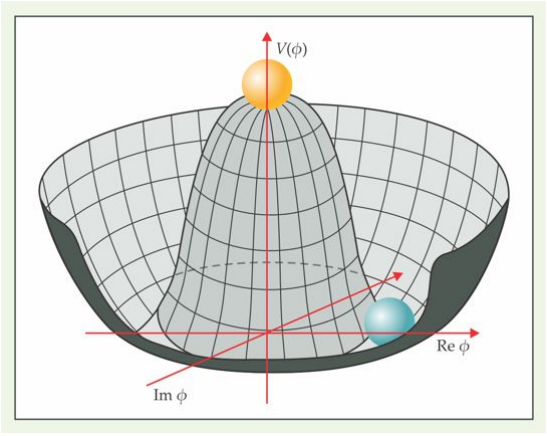


Figure 3: An illustration of the Higgs potential for two real scalar fields ϕ_1 and ϕ_2. Figure taken from [21].

expectation value, to determine the spectrum of particles, the theory must be quantisied. The procedure for quantising a complex scalar field is to expand the field in terms of creation and annihilation operators. Shifting the field such that,

$$\phi(x) = \nu + \varepsilon(x) \Rightarrow \langle 0|\varepsilon|0\rangle = 0, \tag{34}$$

allows for expanding the field around vacuum expectation value, and the spectrum of particles is understood in terms of $\varepsilon(x)$. To illustrate the implications on the Goldstone

model, the shifted complex scalar field can be parameterised as,

$$\psi(x) = \frac{1}{\sqrt{2}}(\varepsilon(x) + \nu)e^{i\frac{\sigma(x)}{v}}, \tag{35}$$

where ε, σ are real scalar fields. Substituting this into the Lagrangian and expanding, the Lagrangian can be expressed as

$$\mathcal{L} = \frac{1}{2}\partial_\mu\varepsilon\partial^\mu\varepsilon + \frac{1}{2}\partial_\mu\sigma\partial^\mu\sigma - \lambda\nu^2\varepsilon^2 + \text{interaction}. \tag{36}$$

The Lagrangian in Eq. 28 is significantly different. The $U(1)$ symmetry is no longer present and an additional massless scalar field, $\sigma(x)$ has appeared, in the Lagrangian: this is referred to as a Nambu-Goldstone boson. This result can be generalised as the Goldstone theorem, where a Lagrangian invariant under a continuous symmetry with an asymmetric ground state implies the existence of a massless Nambu-Goldstone boson.

1.4.2 The Higgs Mechanism

The Higgs mechanism is fundamental to the Standard Model and consolidates the current best description of electroweak physics. The mechanism occurs when a local gauge symmetry is spontaneously broken. As with global symmetry breaking, spontaneous symmetry breaking of a local symmetry generates massless Goldstone bosons. However, the difference is the additional gauge freedom coming from a local symmetry allows for the massless Goldstone bosons to be absorbed by a gauge transformation. This gives rise to the masses of the gauge bosons. The symmetry breaking of the electroweak theory, $SU(2) \times U(1)$, is referred to as Electroweak Symmetry Breaking (EWSB). This section will overview the Higgs mechanism, following a similar procedure as detailed in the previous section for the Goldstone model but now for a gauge invariant Lagrangian. The focus of this section is to highlight the pertinent parts of the Higgs mechanism and the interactions crucial to vector boson scattering.

The Standard Model Lagrangian is parameterised by a doublet complex scalar field ϕ, defined as,

$$\phi = \begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi_1 + i\phi_2\\ \phi_3 + i\phi_4\end{pmatrix}. \tag{37}$$

The $SU(2) \times U(1)$ gauge invariant Lagrangian for the complex scalar field ϕ is defined as,

$$\mathcal{L} = (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi^\dagger\phi) + \mathcal{L}_{\text{gauge}}, \tag{38}$$

where $\mathcal{L}_{\text{gauge}}$ are the kinetic terms associated with the gauge fields, D_μ is the covariant derivative, defined as,

$$D_\mu = \partial_\mu + igW_\mu^a\frac{\sigma}{2} + ig'YB_\mu, \tag{39}$$

and $V(\phi^\dagger\phi)$ is the potential added to break the symmetry,

$$V(\phi^\dagger\phi) = -\mu^2\phi^\dagger\phi + \frac{1}{2}\lambda(\phi^\dagger\phi)^2 \quad , \quad \mu^2 > 0. \tag{40}$$

Analogously to the Goldstone model, the ground state of this system is found by finding the minima of the Hamiltonian energy density, which occurs for constant field configurations that minimise the potential $V(\phi^\dagger\phi)$. There are an infinite number of degenerate field configurations satisfying,

$$\phi^\dagger\phi = \phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2 = -\frac{\mu^2}{2\lambda} = \frac{v^2}{2}. \tag{41}$$

When a symmetry is spontaneously broken, the ground state field axes can be freely chosen and here are conveniently given by,

$$\phi_0 = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v\end{pmatrix}, \tag{42}$$

where $v = \frac{\mu}{\sqrt{\lambda}}$. The complex scalar field can now be parameterised as,

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi_1 + i\phi_2\\ v + h(x) + i\phi_4\end{pmatrix} = e^{\frac{i}{v}\pi^a(x)\sigma^a}\begin{pmatrix}0\\ \frac{1}{\sqrt{2}}(v + h(x))\end{pmatrix}. \tag{43}$$

Once again, the complex scalar field $\phi(x)$ is expanded around the ground state configuration by a scalar field $h(x)$. In the Goldstone model, this complex scalar field would introduce three massless scalar fields for each $\pi_a(x)$ and a scalar $h(x)$ with mass m_h. However, this Lagrangian is invariant under local $SU(2)$ transformations. This means that two representations of fields, $\phi(x)$ and $\phi'(x)$, related by a transformation $U \in SU(2)$,

$$\phi(x) \to U(x)\phi(x) = \phi'(x), \tag{44}$$

are equivalent. By choosing a gauge such that $U(x) = e^{-\frac{i}{v}\pi^a(x)\sigma^a}$, referred to as the unitary gauge, the Goldstone bosons can be absorbed by the gauge transformation, such that the complex scalar field ϕ_U can be written as,

$$\phi_U = \begin{pmatrix}0\\ \frac{1}{\sqrt{2}}(v + h(x))\end{pmatrix}. \tag{45}$$

The Lagrangian for the Higgs field in the unitary gauge can now be written as,

$$\begin{aligned}\mathcal{L}^U_{\text{Higgs}} = {}& \phi_U^\dagger(\partial_\mu + igW_\mu^a\frac{\sigma}{2} + ig'YB_\mu)\phi_U(\partial_\mu + igW_\mu^a\frac{\sigma}{2} + ig'YB_\mu)\\ & - \lambda(\phi_u^\dagger\phi_u - \frac{\nu}{2})^2 + \mathcal{L}_{\text{gauge}},\end{aligned} \tag{46}$$

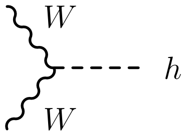
where the potential $V(\phi^\dagger\phi)$ has been shifted such that the vacuum expectation value is non-zero. Expanding the Lagrangian and inserting the transformations given in Eqs. 18 and 19 relating the physical fields of the weak force to the gauge fields leads to mass terms for the $W^\pm$ and Z_μ fields,

$$\mathcal{L}_{\text{mass}} = \frac{1}{4}g^2\nu^2W_\mu^-W^{+\mu} + \frac{1}{8}(g^2 + g'^2)\nu^2Z_\mu Z^\mu, \tag{47}$$

while the electromagnetic field A_μ remains massless. The scalar Higgs fields also introduce interactions with the gauge bosons, such as,

$$\mathcal{L}_{WWH} = g m_W W^-_\mu W^{+\mu} h \tag{48}$$

$\Rightarrow$



Interactions of this form involving a Higgs boson form an important contribution to vector boson scattering. The Higgs mechanism also gives rise to masses for the fermions by introducing Yukawa couplings into the Lagrangian.

1.5 Vector Boson Scattering

The electroweak theory predicts, at leading order (LO), unique interactions involving only gauge bosons, a consequence of its non-abelian gauge structure. Furthermore, EWSB introduces additional interactions between scalar and gauge bosons. A special class of these interactions is referred to as Vector Boson Scattering (VBS). The general definition of a VBS event is one in which two incoming gauge bosons scatter, leading to two outgoing gauge bosons, $VV \to VV$. Gauge bosons are intermediate particles, so in the context of an experiment, such an event could occur through the interaction of two parent particles, which radiate gauge bosons. There are several interactions which can lead to VBS events, as shown in Fig. 4. These involve trilinear interactions with s-, t-, and u-channel exchanges of γ/Z, quartic self-interactions, and s- and t-channel exchanges with the Higgs boson, h. Crucially, VBS events are highly sensitive to both the gauge structure of the electroweak theory and to EWSB. Experimental research of these interactions simultaneously tests two major components of the Standard Model.

The recent discovery of a scalar boson, the Higgs boson, at the Large Hadron Collider (LHC) has promoted further experimental interest in VBS. Present experimental research is centered around precision measurements of the Higgs, aimed at understanding if there are small BSM contributions to its interactions. In addition, VBS processes indirectly probe the nature of the Higgs and its interactions with gauge bosons. An overview of the current experimental status of VBS measurements is outlined in Chapter 4.

VBS events can be further partitioned by the polarisation configurations of the incoming and outgoing gauge bosons, such that $V_\lambda V_\lambda \to V_\lambda V_\lambda$ where $\lambda = L/T$ for a longitudinally/transversely polarised gauge boson. The purely longitudinal VBS configuration, $V_L V_L \to V_L V_L$, is of particular experimental interest as without the contributions from the Higgs boson, the theoretical cross-section diverges. This section will first begin by providing an overview of polarisation and a discussion of experimental observation in the context of W boson decay. Subsequently, the section will then discuss purely longitudinal VBS for $W^\pm W^\pm \to W^\pm W^\pm$ as the main motivation of this thesis.

1.5.1 Polarisation

Gauge theories predict spin-1 massless gauge bosons, denoted by the 4-vector field, A^μ. The solutions to the vector fields can be expanded in terms of a set of orthogonal basis

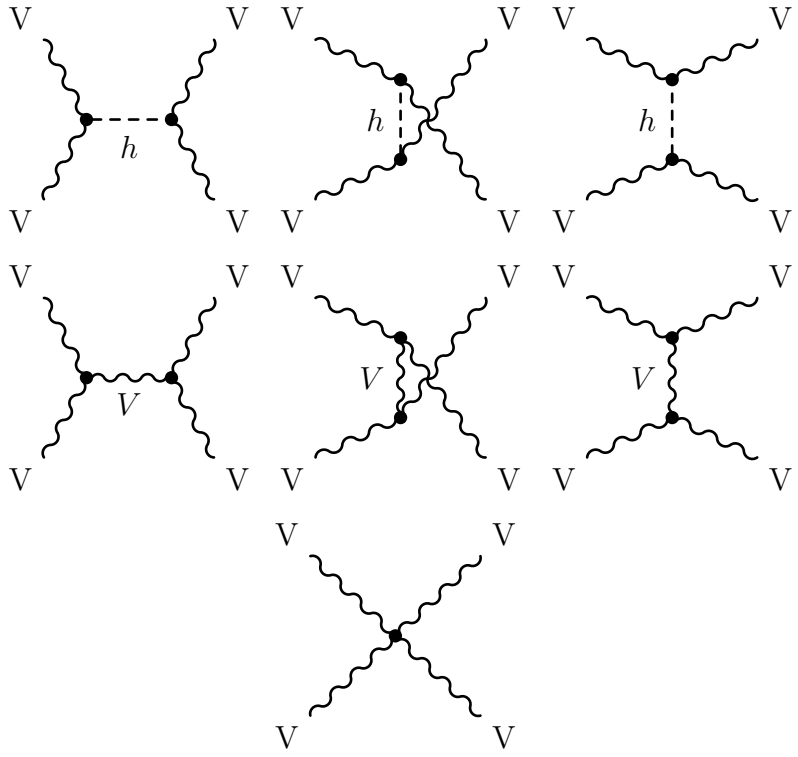


Figure 4: Generalised Feynman diagrams illustrating all of the possible leading-order VBS diagrams where $V = W/Z/\gamma$ bosons and h represents a Higgs boson.

vectors ϵ^i representing the polarisation states of the boson. The gauge freedom associated with each vector field allows for a choice of gauge, which then imposes conditions on the basis vectors and, thus, on the degrees of freedom of the field. The full set of physical solutions for massless vector field A^μ can be described in terms of two basis vectors, and hence two polarisation states ϵ^i with $i = 1, 2$. This could be inferred a priori, knowing that, for example, the photon has two polarisation states.

Each massless gauge boson has two transverse polarisation states. The two transverse states correspond to the two possible spin S_z configurations of the vector boson. The two transverse configurations are referred to as left- and right-handed, with $S_z = -1/1$, respectively. However, the electroweak gauge bosons are massive. By the Higgs mechanism, the Nambu-Goldstone bosons are absorbed by the electroweak gauge bosons, which acquire a mass and, consequently, an additional degree of freedom. This additional degree of freedom for gauge bosons manifests as a longitudinal polarisation. The electroweak gauge bosons have three possible polarisations: two transverse ($S_z = \pm 1$) and one longitudinal mode ($S_z = 0$).

Accordingly, polarisation is inextricably linked with both gauge theory and EWSB. Experimentally identifying the polarisation of a massive gauge boson, however, is not a trivial task. The polarisation must be inferred from angular distributions of the gauge bosons decay products. To show this, this section will derive an expression for the W^- boson decay rate to illustrate how polarisation affects the gauge boson's decay products. Consider the Feynman diagram for the leptonic decay of a W^- boson, as shown in figure

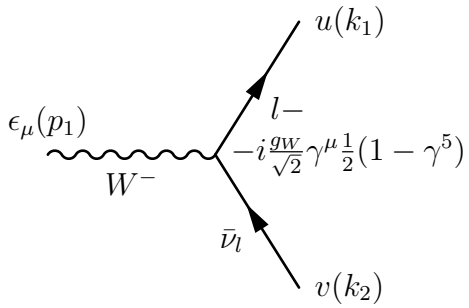


Figure 5: A Feynman diagram illustrating the leptonic decay of a W boson.

5. The matrix element for this decay is given by,

$$-i\mathcal{M}_{fi} = \epsilon_\mu(p_1)\bar{u}(k_1)\left[-i\frac{g_W}{\sqrt{2}}\gamma^\mu\frac{1}{2}(1-\gamma^5)\right]v(k_2), \tag{49}$$

where ϵ^μ refers to the gauge boson polarisation vector and $\bar{u}(k_1)$, $v(k_2)$ are the outgoing Dirac spinors. The matrix element can be rewritten in terms of a scalar product of the polarisation four-vector and the weak charged current,

$$\mathcal{M}_{fi} = \frac{g_W}{\sqrt{2}}\epsilon_\mu(p_1)j^\mu \quad \text{where} \quad j^\mu = \bar{u}(k_1)\gamma^\mu\frac{1}{2}(1-\gamma^5)\nu(k_2) \tag{50}$$

The decay is hence dependent on polarisation. Consider now the decay of the W^- boson in its rest frame as shown in Fig. 6. In this frame of reference, the polarisation vectors

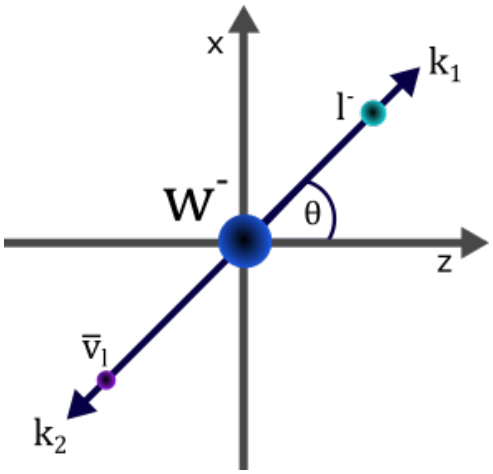


Figure 6: A graphical illustration of the kinematics of a W^- boson decay, in the W boson rest frame.

for the W boson are given by,

$$\epsilon_L^\mu = \frac{1}{\sqrt{2}}(0,1,-i,0), \quad \epsilon_0^\mu = (0,0,0,1), \quad \epsilon_R^\mu = -\frac{1}{\sqrt{2}}(0,1,i,0) \tag{51}$$

and the outgoing momentum of the two leptons is given by,

$$k_1 = \left(\frac{m_w}{2}, \frac{m_w}{2}\sin\theta, 0, \frac{m_w}{2}\cos\theta\right), \quad k_2 = \left(\frac{m_w}{2}, -\frac{m_w}{2}\sin\theta, 0, -\frac{m_w}{2}\cos\theta\right) \tag{52}$$

where θ is the angle of decay of the leptons. The calculation can be simplified by taking the ultra-relativistic limit. In this limit, the outgoing leptons can be considered massless and so only the left/right-handed particles/anti-particles participate in the decay. The expression for the weak current simplifies to,

$$j^{\mu} = \bar{u}_L(k_1)\gamma^{\mu}\nu_R(k_2) = 2E(0, -\cos\theta, -i, \sin\theta) \tag{53}$$

where $u_L(k_1)$ and $\nu_R(k_2)$ are the left- and right-handed helicity states of the Dirac spinors, respectively. Iterating over the different polarisation configurations, for longitudinally polarised W^- boson decay, the matrix element is given by,

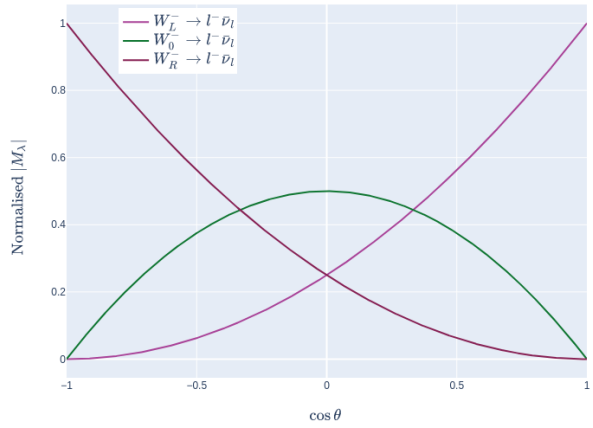


Figure 7: The $\cos\theta$ distribution for W^- decay, with longitudinal polarisation and right- and left-handed transverse polarisation shown separately.

$$-i\mathcal{M}_0 = i\frac{1}{\sqrt{2}}g_W m_W \sin\theta, \tag{54}$$

and likewise for each transverse polarisation,

$$-i\mathcal{M}_L = -i\frac{1}{\sqrt{2}}g_W m_W(1+\cos\theta), \quad -i\mathcal{M}_R = -i\frac{1}{\sqrt{2}}g_W m_W(1-\cos\theta) \tag{55}$$

Finally, the differential decay rate is proportional to the square of the matrix element,

$$\frac{d\Gamma}{d\Omega} \propto |M|^2, \tag{56}$$

where $d\Omega$ represents the angular distributions of the Lorentz-invariant final states. This differential decay rate is directly proportional to the angular distributions of the decaying leptons. Figure 7 shows the dependence on $\cos\theta$ for each of the three possible polarisation states. The overlap in the distributions makes it difficult to conclusively identify the polarisation of a single gauge boson, however polarisation could be inferred by an experiment measuring this distribution. In this thesis, it is referred to as the $\cos\theta$ distribution.

1.5.2 Polarised Vector Boson Scattering

There are a number of different VBS processes, each permutations of the electroweak gauge bosons. These can be further sub-categorised by the different possible polarisation configurations. Of all the processes, scattering involving only longitudinally polarised gauge bosons is of foremost experimental interest. Whilst all VBS processes share the property of being an important test of the gauge structure of the Standard Model, longitudinally polarised scattering is also an intrinsic test of the EWSB. This is because the Standard Model cross-section for longitudinal VBS would violate unitarity without contributions coming from the Higgs boson. Furthermore, this cancellation is exact, and so any small deviation in the Standard Model Higgs couplings would theoretically restore a divergence in the cross-section: a nonphysical prediction.

This thesis details an experimental measurement of polarised vector boson scattering of two same-charge W bosons, $W^{\pm}W^{\pm} \to W^{\pm}W^{\pm}$, with a focus on longitudinal polarisation. This section will outline the Standard Model theoretical predictions for the purely longitudinal VBS mode of this process. The calculation will demonstrate how unitarity-violating terms are cancelled by considering all of the Standard Model predictions and consequently show why this process is extremely sensitive to EWSB. The calculation of the theoretical predictions presented follows the literature [22]. The arguments presented in this section are valid for a number of other VBS processes, with the theoretical predictions outlined by [22].

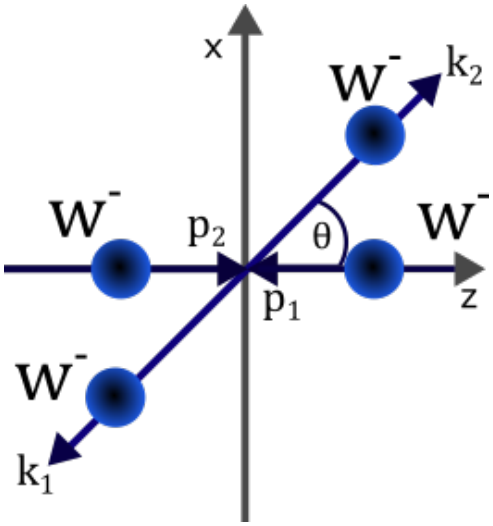


Figure 8: An illustration of the kinematics of a $W^-W^- \to W^-W^-$ scattering event, showing the definition of the angle θ.

The section will calculate the scattering amplitude in the Centre-of-Mass (CoM) frame of two incoming $W^{\pm}$ bosons, as shown in Fig. 8. The important variables related to the prediction are outlined below. The incoming and outgoing momenta of the gauge bosons are represented by,

$$\begin{aligned} p_1^{\mu} &= (E, 0, 0, -p), & p_2^{\mu} &= (E, 0, 0, p) \\ k_1^{\mu} &= (E, -p\sin\theta, 0, -p\cos\theta), & k_2^{\mu} &= (E, p\sin\theta, 0, p\cos\theta) \end{aligned} \tag{57}$$

where

$$E^2 = m_W^2 + p^2. \tag{58}$$

The longitudinal polarisation vectors of the incoming and outgoing gauge bosons in this reference frame are represented by,

$$\begin{aligned} \epsilon_1^\mu(0) &= \frac{1}{m_W}(-p, 0, 0, E), & \epsilon_2^\mu(0), &= \frac{1}{m_W}(-p, 0, 0, -E), \\ \epsilon_3^{*\mu}(0) &= \frac{1}{m_W}(p, -E\sin\theta, 0, -E\cos\theta), & \epsilon_4^{*\mu}(0) &= \frac{1}{m_W}(p, E\sin\theta, 0, E\cos\theta), \end{aligned} \tag{59}$$

and the two transverse configurations are given by,

$$\begin{aligned} \epsilon_1^\mu(\pm) &= \sqrt{2}(0, -1, \pm i, 0), & \epsilon_2^\mu(\pm) &= \sqrt{2}(0, 1, \pm i, 0) \\ \epsilon_3^{*\mu}(\pm) &= \sqrt{2}(0, -\cos\theta, \mp i, \sin\theta) & \epsilon_4^{*\mu}(\pm) &= \sqrt{2}(0, \cos\theta, \mp i, -\sin\theta). \end{aligned} \tag{60}$$

As previously mentioned, this section will only show the calculation of longitudinal scattering, so the above vectors for transverse polarisation are not used in the calculation but are included for completion.

The Mandelstam variables referenced in the calculation below are defined as,

$$s = (p_1 + p_2)^2 = 4E^2, \tag{61}$$

$$t = (p_1 - k_1)^2 = -4p^2 \sin^2\frac{\theta}{2}, \tag{62}$$

$$u = (p_1 - k_2)^2 = -4p^2 \cos^2\frac{\theta}{2}. \tag{63}$$

There are seven Feynman diagrams contributing to $W^\pm W^\pm \to W^\pm W$, as shown in figure 9. Akin to the W boson decay in the previous section, the matrix element can be expressed as the scalar product of the incoming and outgoing polarisation vectors with the interaction current of each VBS diagram, denoted by,

$$-i\mathcal{M}_i = -i\epsilon^\mu(\lambda_1)\epsilon^\nu(\lambda_2)\mathcal{M}_{\mu\nu\rho\sigma}\epsilon_3^{*\rho}(\lambda_3)\epsilon_4^{*\sigma}(\lambda_4), \tag{64}$$

where $\mathcal{M}_i$ refers to the interaction amplitude of a Feynman diagram. The scattering amplitude for a given polarisation configuration can be found by summing of all of the individual matrix element contributions contracted with the appropriate polarisation vectors. For purely longitudinal scattering, the polarisation vectors are defined in Eq. 59. For the quartic interaction shown in the Feynman diagram in Fig. 9c, the interaction current is given by,

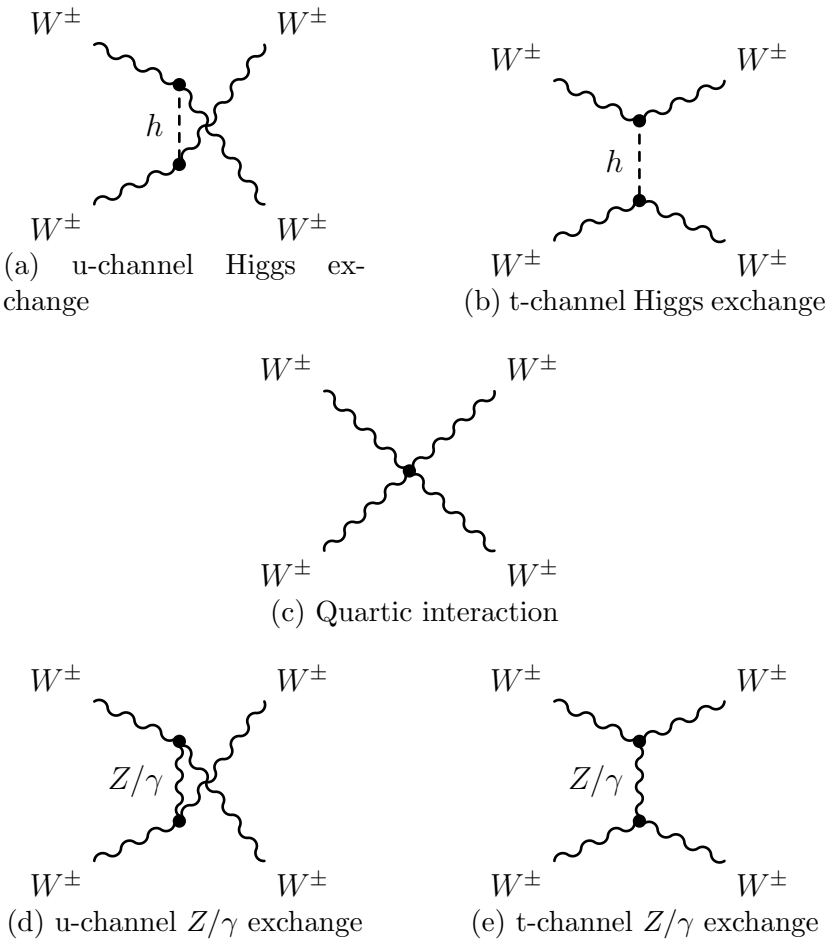
$$\mathcal{M}_{\mu\nu\rho\sigma} = ig^2[2g_{\rho\sigma}g_{\mu\nu} - g_{\rho\mu}g_{\sigma\nu} - g_{\rho\nu}g_{\sigma\mu}], \tag{65}$$

and hence the matrix-element for this contribution is as follows,

$$-i\mathcal{M}_\times = g^2[2(\epsilon_1\cdot\epsilon_2)(\epsilon_3^*\cdot\epsilon_4^*) - (\epsilon_1\cdot\epsilon_3^*)(\epsilon_2\cdot\epsilon_4^*) - (\epsilon_1\cdot\epsilon_4^*)(\epsilon_2\cdot\epsilon_3^*)]]. \tag{66}$$

where $\mathcal{M}_\times$ refers to the matrix element contribution for quartic interaction gauge boson interaction. The currents for the remaining interactions are detailed in [22], with the matrix elements for each of the contributions given by [22],

$$-i\mathcal{M}_H = \frac{-g^2 m_W^2}{t - m_H^2}\left[2(\epsilon_1\cdot\epsilon_3^*)(\epsilon_2\cdot\epsilon_4^*)\right], \tag{67}$$



(a) u-channel Higgs exchange

(b) t-channel Higgs exchange

(c) Quartic interaction

(d) u-channel Z/γ exchange

(e) t-channel Z/γ exchange

Figure 9: Feynman diagrams illustrating all of the leading-order VBS diagrams for $W^\pm W^\pm \to W^\pm W^\pm$. The process has no s-channel diagrams as they are not possible without violating charge conservation.

where $\mathcal{M}_H$ refers to the matrix elements coming from the u- and t-channel exchange of the Higgs boson, shown in Figs. 9a and 9b, and

$$\begin{aligned} -i\mathcal{M}_{Z\gamma} =& g^2 \left[\frac{\cos^2\theta_W}{t-m_Z^2} + \frac{\sin^2\theta_W}{t}\right] \\ &\times 8[(\epsilon_3^*\cdot p_1)(\epsilon_4^*\cdot p_2)(\epsilon_1\cdot\epsilon_2) + (\epsilon_3^*\cdot p_1)(\epsilon_2\cdot k_2)(\epsilon_1\cdot\epsilon_4^*)] \\ &\times(-4)[(\epsilon_3^*\cdot p_1)(\epsilon_4^*\cdot p_2)(\epsilon_1\cdot(p_2+k_2)) + (\epsilon_4^*\cdot p_2)(\epsilon_1\cdot\epsilon_3^*)(\epsilon_2\cdot(p_1+k_1)] \\ &\times[(\epsilon_1\cdot\epsilon_3^*)(\epsilon_2\cdot\epsilon_4^*)((p_1+k_1)\cdot(p_2+k_2))], \end{aligned} \tag{68}$$

where $\mathcal{M}_{Z\gamma}$ refers to the matrix elements coming from the u- and t-channel exchange of the Z and γ bosons, as shown in Figs. 9e and 9d. The purpose of outlining this calculation in this section is to explicitly show the cancellation of unitarity violating terms. In order to demonstrate this, the calculation is approximated by taking the ultra-relativistic limit, where $s >> m_W^2$. For a physical prediction, the scattering amplitude for any process should not depend on any positive power of s as this would violate unitarity. For purely longitudinal scattering, $W_L^\pm W_L^\pm \to W_L^\pm W_L^\pm$, the matrix element contributions, are given

by,

$$-i\mathcal{M}_{Z\gamma} = \frac{g^2}{4}[\cancel{\frac{s^2}{m_W^4}} - \cancel{\frac{2st}{m_W^4}} - \cancel{\frac{4s}{m_W^2}} + \mathcal{O}(s^0)], \tag{69}$$

$$-i\mathcal{M}_{\times} = \frac{g^2}{4}[-\cancel{\frac{s^2}{m_W^4}} + \cancel{\frac{2st}{m_W^4}} + \cancel{\frac{3s}{m_W^2}} + \mathcal{O}(s^0)], \tag{70}$$

$$-i\mathcal{M}_{H} = \frac{g^2}{4}[\cancel{\frac{s}{m_W^2}} + \mathcal{O}(s^0)], \tag{71}$$

where the relation $s + t + u = 4m_W^2$ has been used to remove u. The total scattering amplitude for this VBS process is found by taking the sum of all the individual matrix elements. Equations 69, 70, and 71 demonstrate how unitarity violating terms scaling with s cancel when summing over the contributions. The cancellation of gauge boson contributions is shown in blue, and Higgs contribution in red. The total matrix element is given by [22],

$$-i\mathcal{M}_{\text{tot}} = -\frac{g^2}{4}\frac{m_H^2}{m_W^2}\left[\frac{t}{t - m_H^2} + \frac{u}{u - m_H^2}\right]. \tag{72}$$

The total cross section is given by [22],

$$\frac{d\sigma}{d\Omega} = \frac{1}{64\pi^2 s}|\mathcal{M}_{\text{tot}}|^2. \tag{73}$$

The Standard Model prediction for this process is unitary, and the terms scaling with s cancel completely when the different contributions are summed together, giving a physical cross-section. The converse is that the individual contributions must be studied together; they cannot be separated. This highlights an important consequence of gauge invariance.

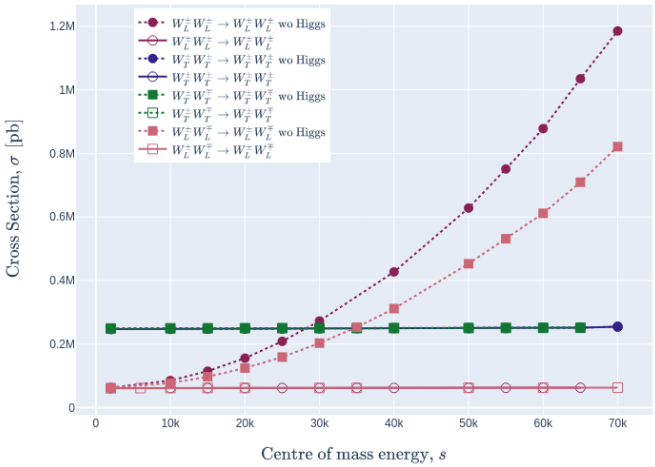


Figure 10: The predicted cross-sections as a function of centre-of-mass energy for various VBS processes with and without a Standard Model Higgs boson ("wo Higgs" in the figure). Cross-sections calculated with MadGraph5_aMC@NLO [23].

Evidently it would not be possible to construct a gauge invariant Lagrangian without one

of these interactions. Moreover, one of the divergences is cancelled by the Higgs contribution. Without a Standard Model Higgs, the cross-section for longitudinally-polarised gauge boson scattering would diverge with CoM energy, but this is not true for the remaining polarisation modes. Figure 10 shows the numerically estimated cross-section for $W_\lambda W_\lambda \to W_\lambda W_\lambda$, where $\lambda = L/T$, at various CoM energies. The figure shows that only longitudinally polarised VBS requires a Standard Model Higgs to ensure a physical cross-section. Furthermore, the cancellation is exact. Any deviations from the Standard

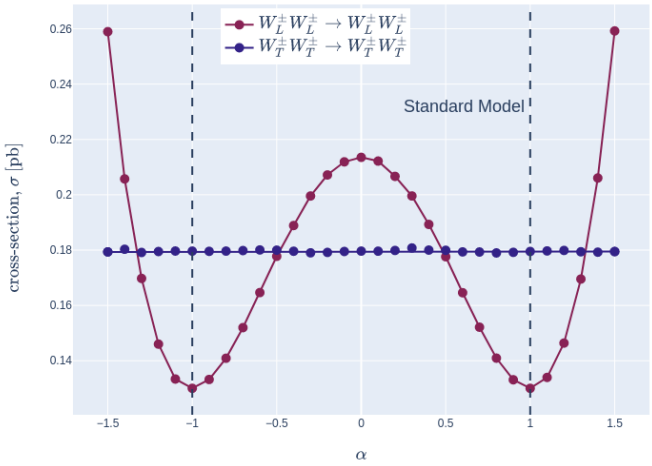


Figure 11: The Standard Model cross-section prediction for purely longitudinal and purely transverse $W^\pm W^\pm \to W^\pm W^\pm$ VBS for various values of α, which scales the Higgs boson coupling to gauge bosons. Cross-sections calculated with MadGraph5_aMC@NLO [23].

Model couplings restore this divergence. To explicitly see the impact of this, consider the gauge boson interacting part of Standard Model Lagrangian, but now with the coupling constants modified by a parameter α, as shown in Equation 74.

$$\mathcal{L}^{\alpha}_{\text{Higgs}} = \frac{1}{8}(2vh\alpha + h^2)[2g^2 W^-_\mu W^{+\mu} + (g^2 + g'^2) Z_\mu Z^\mu] - \lambda\nu h^3 - \frac{1}{4}\lambda h^4. \tag{74}$$

The predicted cross-section for transversely polarised and longitudinally polarised $W_L^\pm W_L^\pm \to W_L^\pm W_L^\pm$ at a fixed CoM energy for various values of α is shown in Fig. 11. From the figure, it can be seen that any anomalies in the Higgs couplings to the electroweak gauge bosons would directly change only the predictions of longitudinal scattering.

2 Experimental Apparatus

The experiment presented in this thesis is performed at CERN using data produced at the Large Hadron Collider (LHC) and collected by the ATLAS experiment. The measurement uses proton-proton collision data collected over a time period from 2015 to the end of 2018 by the ATLAS detector. It should be noted that this time period is before the start of this thesis. This means that the major body of work performed for this thesis did not contribute directly to the experimental design or the data collection but rather to data analysis to measure physical properties. Some of the work performed by the author contributed to the planned upgrade of the experiment, which is expected to begin operation towards the start of 2029. This is discussed in the next chapter. This chapter gives a detailed overview of the experimental setup and the methods used to collect the data. The purpose of this chapter is to outline the essential aspects of the experiment. The experiment highly influences the analysis, and the aim of this chapter is to outline the core aspects of the experiment that are most relevant to the analysis.

As this experiment is performed at CERN, this chapter begins by giving a brief introduction to CERN and the organisation. The data used in this measurement is produced by the LHC. This chapter briefly introduces the LHC and the relevant collider physics and terminology. The remainder of the chapter focuses on the ATLAS detector, the experiment used for the data collection. As previously mentioned, the data analysis is highly specific to this experiment. This chapter gives a detailed overview of the global aspects of the ATLAS detector. The first part outlines the experiment itself by discussing each of the different sub-detectors. The chapter then discusses the detector instrumentation and performance capabilities of each sub-detector. Finally, the chapter introduces the ATLAS reconstruction procedures, summarising the flow of data from raw signal measurements to data ready for physics analysis. A more detailed overview of the ATLAS tracking procedure is provided, as this is also highly relevant to the subsequent chapter.

2.1 CERN

The European Organisation for Nuclear Research, CERN [24] ("Conseil Européen pour la Recherche Nucléaire") is an international organisation operating the world's largest laboratory for scientific research. The idea of establishing the organisation first originated in the mid-1940s. Coming out of the Second World War, a European laboratory was suggested by scientists as an opportunity to encourage scientific collaboration between European countries and as a way to stop the brain drain to America.

The establishment of a European Council for Nuclear Research was first formally proposed at a UNESCO intergovernmental meeting in Paris in December 1951. An agreement was signed two months later, forming a provisional council to produce the CERN Convention. By June 1953, the final draft of the CERN Convention was signed by 12 Member States, and the organisation was officially born. The organisation has since grown substantially, with 23 member states and worldwide scientific collaboration at the time of writing.

The CERN site was established in 1954 and is located on the France–Switzerland border. Shortly after its formation, the first particle accelerator at CERN was built in

1957: the 600 MeV Synchrocyclotron (SC) [25]. Since then, the organisation has built a host of particle accelerators and experiments that have led to groundbreaking discoveries in particle physics. The first major discovery came in 1973 with the discovery of neutral currents with the Gargamelle bubble chamber [26], the first experimental evidence of a unified electroweak theory. Following this discovery, a decade later, in 1983, the W and Z bosons were discovered at CERN by the UA1 [27] and UA2 [28] detectors at the proton synchrotron (SPS). More recently, in 2012, CERN confirmed the discovery of a light scalar boson, deemed to be the Higgs boson of the Standard Model, by the ATLAS [29] and CMS [30] experiments at the LHC.

2.2 Large Hadron Collider

The LHC is the world's largest and most powerful particle accelerator. The accelerator is located at CERN, lying at a depth of 100 metres underground on the border between France and Switzerland. The machine itself is approximately 27 kilometres in circumference and was installed in the existing tunnel constructed by CERN in the 1980s for the Large Electron Positron collider (LEP). Plans to construct the LHC were first approved by the CERN Council in December 1994 [31, 32], with the accelerator beginning operation on the 10th of September 2008.

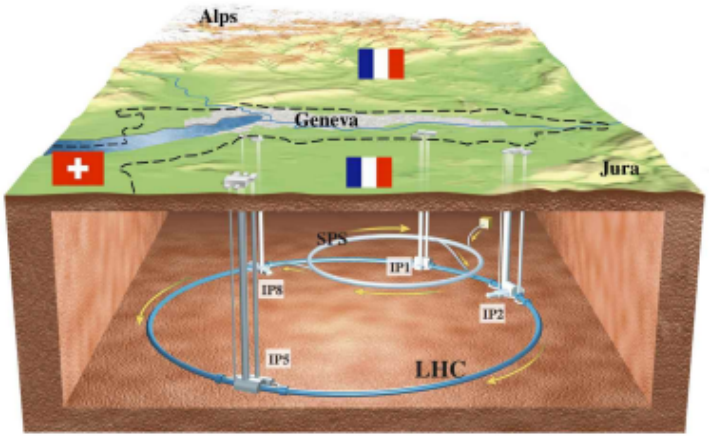


Figure 12: A graphical illustration of the LHC located underground at CERN, showing four of the eight interaction points of the LHC, which host the ATLAS, CMS, LHCb and ALICE experiments. Figure taken from [31].

Two key parameters describe the performance of a particle accelerator: luminosity and CoM energy. For a given process with cross-section σ, the collision rate $\dot{N}$ is directly proportional to the cross-section and machine luminosity L, and is given by,

$$\frac{dN}{dt} = L\sigma. \tag{75}$$

The total number of events N, generated for a given process is proportional to the integrated luminosity L_{int} and the cross-section σ, by

$$N = L_{\text{int}}\sigma. \tag{76}$$

For rare physics searches with very small cross-sections, a high luminosity is required to provide a production rate and total integrated luminosity such that there is a sufficient number of events to produce a measurement. Analogously, to discover new heavy particles, the centre-of-mass energy must be high enough for the accelerator to produce the particles. The LHC was initially designed to push the frontiers of experimental high-energy particle physics by providing collisions at the unexplored TeV energy scale and a groundbreaking luminosity.

LHC parameter	Table1	Table2
Centre-of-mass energy, E_{CM} [TeV]	7	13
Bunch Separation [ns]	50	25
Peak Luminosity [$\mathrm{cm}^{-2}\mathrm{s}^{-1}$]	7×10^{33}	2×10^{34}
Integrated Luminosity [fb^{-1}]	33	190
Interactions per crossing	≈ 21	≈ 55

Table 1: Parameters detailing the performance of the LHC during Run-1 and Run-2. Values taken from [33].

The LHC is a two-ring-superconducting proton-proton (p-p) and heavy ion (Pb) accelerator. The LHC collides bunches of protons at a frequency of 40 MHz with a designed peak luminosity of 10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ and centre-of-mass energy of 14 TeV. The accelerator collides hadrons, primarily protons (p), to avoid the energy loss inherent in electron colliders. The choice of particles to collide is a significant parameter of a particle accelerator. For circular electron accelerators such as the LEP, the primary limiting factor of centre-of-mass energy is synchrotron radiation: radiation emitted by accelerating charged particles. Since the emitted radiation is inversely proportional to the fourth power of the mass of a particle, the much larger masses of hadrons drastically reduce the energy loss during acceleration. This means the LHC can achieve far greater centre-of-mass energies than its predecessor, LEP, and crucially access the TeV energy range. The caveat is that hadrons are composite particles consisting of quarks and gluons. This significantly increases the complexity of theoretically modelling detector collisions and experimental data analysis.

The LHC is home to four major experiments: two high-luminosity general purpose detectors, ATLAS [34] and CMS [35], one flavour-physics experiment LHCb [36], and one heavy ion experiment, ALICE [37]. The performance of the LHC over its current time span is detailed in Table 1. Throughout Run-2, the accelerator exceeded its initial target peak luminosity by a factor of 2. The measurement presented in this thesis uses data produced by the LHC and collected by the ATLAS detector during Run-2. Since the discovery of the Higgs boson in 2012, the LHC is now entering a new era of precision Higgs physics. The LHC's current history and projected timeline is shown in Fig. 13. At the time of writing, the LHC is collecting data at the start of Run-3. Run-3 is designed as a transition phase for the LHC upgrade to the High-Luminosity Large Hadron Collider (HL-LHC), which will deliver record-breaking luminosity and open up the possibility of precision physics measurements of processes such as longitudinally polarised VBS. However, Run-3 will still double the integrated luminosity collected during Run-2.

2.3 The ATLAS Detector

The data analysed in this thesis was collected by the ATLAS detector. This section gives a global overview of the experiment, giving a detailed outline of the detector and experimental setup, from collisions to experimental data. The ATLAS detector is a general-purpose detector designed to take advantage of the high energy and immense luminosity provided by the LHC. The experiment was constructed over the course of the beginning of the century and is installed at the LHC IP1, where it began operation in autumn 2009 [38]. Following the first operation, over the subsequent decade, the detector has collected data over two periods: Run-1 and Run-2 (see Fig. 13). During Run-1 and Run-2, the ATLAS detector recorded 21 and 147 fb^{-1} of p-p collision data, respectively, at a CoM energies of $\sqrt{s} = 7, 8$ and 13 TeV [39]. The data subject to analysis in this experiment is the 147 fb^{-1} dataset collected during Run-2.

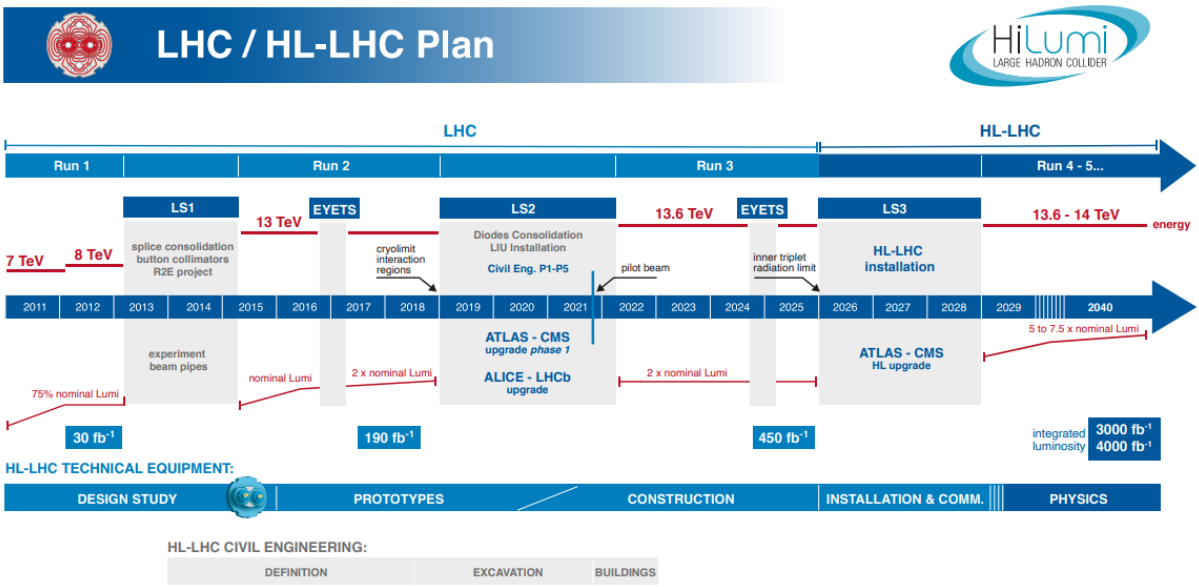


Figure 13: The current and projected LHC/HL-LHC timeline with the respective shutdown periods and planned upgrades for the HL-LHC and the ATLAS/CMS/ALICE detectors. Figure taken from [40].

The ATLAS detector is composed of four major sub-detectors surrounding the beamline: the Inner Detector, the Electromagnetic and Hadronic Calorimeters, and the Muon Spectrometers, ordered by distance from the beamline. Each of the four sub-detectors works co-dependently to measure the kinematic properties of high-energy collisions. The detectors are assembled radially along the beam line surrounding the collision point. Each detector can be decomposed into two sub-components: barrel and end-cap components. The barrel components are defined as components of the detector orientated concentrically around the beamline, while the end-cap refers to components located perpendicular to the barrels, orientated transversely to the beamline. The design provides maximal hermetic coverage for each of the different sub-systems.

The ATLAS magnetic system supports the detector sub-components, which are constructed within the detector and function to immerse the entire detector in a magnetic field. The ATLAS magnet system consists of a central solenoid surrounding the inner tracker and three massive toroidal magnets surrounding the muon spectrometers. The magnet systems bend the trajectories of charged particles. The magnet system works in conjunction with the inner tracker and muon spectrometers to measure the momentum and charge of charged particles. The ATLAS detector also contains three small forward detectors installed around the detector, which are primarily designed for monitoring and measuring luminosity. Finally, the data collection is managed by the Trigger and Data Acquisition system.

The ATLAS detector was designed to capitalise on the unprecedented luminosity provided by the LHC for physics at the TeV scale. The benchmark process for establishing the performance specification of the detector was the production of a Standard Model Higgs boson [38]. The design specification is outlined in great detail in the ATLAS experiment technical design report [38]. This section summarises each of the individual sub-systems and describes the capabilities and limitations of the experiment.

The incredibly harsh radiation environment of the LHC is a technical challenge shared by each detector sub-system. The components of the detector have been designed with some overlapping shared design principles. Each component is developed to be fast, with radiation-hard electronics and high granularity to reduce the effects of pile-up [38]. The detector components are also designed with maximal hermetic coverage to increase the detector's capabilities in reconstructing the full-event kinematics of p-p collisions. Finally, the materials used in the detector have been carefully chosen with many structural and physics-based requirements unique to each sub-system, with the details outlined in the design report [38].

2.3.1 The ATLAS Coordinate System

This subsection describes the conventional coordinate system used to describe the detector. The ATLAS experiment uses a right-handed Cartesian coordinate system [38], as shown in Fig. 14. The origin of the coordinate system is taken as the nominal IP, with the z-axis pointing along the beamline. The x-y plane is defined transverse to the z-axis, with the positive x-axis directed towards the centre of the LHC and the positive y-axis pointing upwards to the surface. The x-y plane is often described in terms of polar coordinates, r-ϕ, where $r = x^2+y^2$ and ϕ, the azimuthal angle around the beamline, measured anti-clockwise from the positive x-axis, see Fig. 14. Spherical coordinates are also defined, with the polar coordinate θ, taken around the x-axis, measured anti-clockwise from the positive z-axis. The polar coordinate θ is often instead described in terms of the pseudorapidity η, defined as,

$$\eta = -\log\tan(\theta/2) \in [0, \infty). \tag{77}$$

The pseudorapidity is an approximation of rapidity y, defined for a particle with energy E and momentum p_z along the z-axis, as,

$$y = \frac{1}{2}\log\left(\frac{E + p_z}{E - p_z}\right), \tag{78}$$

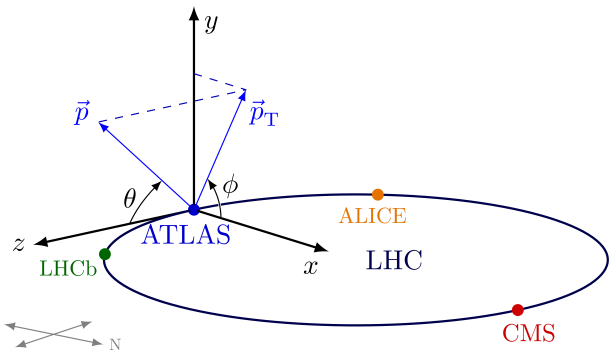


Figure 14: The ATLAS coordinate system as detailed by [38]. The coordinate system lies on the nominal IP. Figure taken from [38].

where the approximation converges to an equality for massless particles. The rapidity is preferred as the differences in rapidity are Lorentz invariant, unlike the differences in polar coordinate θ. The transverse momentum p_T is defined as the magnitude of the momentum in the x-y plane, and the missing transverse energy E_T^{miss} is defined as the magnitude of the momentum imbalance.

2.3.2 The Inner Detector

The ATLAS Inner Detector (ID) is a tracking detector responsible for measuring the trajectories of charged particles emerging from the collision region. A schematic layout of the ID is shown in Fig. 20. The inner detector is composed of three independent sub-detectors: the Pixel Detector (PD), the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT).

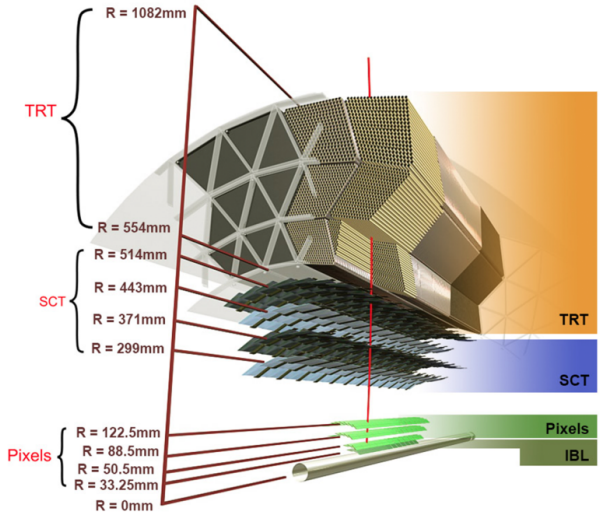


Figure 15: A cross-section of the ATLAS Inner Detector illustrating the different sub-detectors and their respective sizes radially, R, from the beamline. Figure taken from [41].

The Pixel Detector is the innermost detector situated around the beamline. The

sensors providing the measurements for the pixel detector are the pixel sensors. The pixels are silicon semiconductors. Charged particles passing through a pixel sensor induce electron-hole pairs. The induced charges are collected via a potential difference applied across the pixel, resulting in an electric signal. The Pixel Detector for Run-1 consisted of three barrel layers. An additional layer, the IBL [42], was installed during LS1 and has been operational since Run-2. The Run-1 PD is composed of 1744 identical semiconductor pixel sensors, 19×63 mm^2 in size. Each pixel sensor is composed of 47232 individual pixels with a nominal size of 50×400 μm^2. There are a total of 112 barrel staves and 48 end-cap sectors on which the 1744-pixel sensor modules are distributed. The IBL is composed of 14 carbon composite staves, and each stave supports 20 pixel sensor modules. Each module uses pixels of size 50×250 μm^2. The IBL has a total of approximately 12 million pixels [42].

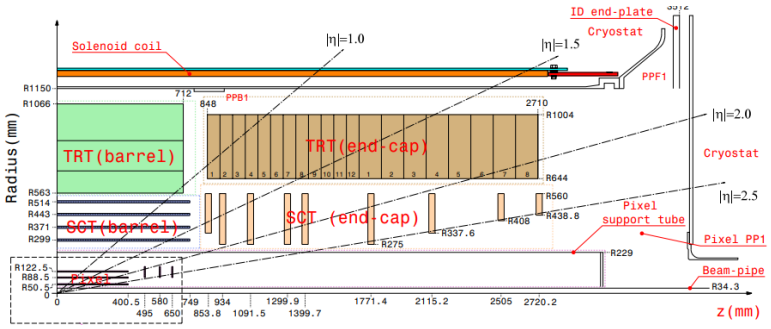


Figure 16: An illustration of the active medium of each of the sub-detectors in the Inner Detector radially and along the beamline. Figure taken from [41].

The SCT is the second innermost detector surrounding the PD. The SCT is designed so that charged particles passing through the detector are measured at four unique space-points. The main modules providing the measurements for the SCT are silicon microstrips. The SCT consists of four coaxial cylindrical layers in the barrel region and two end-caps with nine disk layers each. There are, in total 4, different types of silicon strip sensors used in the detector, 3 of which are used for the end-cap layers and one in the barrel. In total, the SCT consists of 4088 modules.

The outermost detector of the ID is the TRT. The TRT [43] is unlike the previously described detectors and uses thin-walled polyimide drift tubes (straw). Each drift tube is 4 mm in diameter, and there is a gold-plated tungsten wire in the centre of each tube. The tubes are filled with an Argon-based gas mixture [44]. When a charged particle passes through the tube, the particle ionises the gas, and the free electrons drift towards the wire, resulting in a signal. The TRT consists of 73 layers of straws in the barrel and 160 straw planes in the end-cap. In the barrel, the straws are aligned parallel to the beam axis, while in the end-caps, the straws are aligned radially. There are in total approximately 300000 straws.

In combination, the three detectors contribute to pattern recognition, tracking performance and electron identification. Measurements of particle trajectories allow for high-resolution measurements of the transverse momentum of the charged particles. Reconstructed tracks can also be used to determine the positions of primary and secondary

vertices, allowing for the identification of particle decays within the ID. The fiducial region of the PD and SCT is $|\eta| < 2.5$, while the TRT further extends the tracking radially within $|\eta| < 2$. The TRT additionally provides electron identification.

2.3.3 ATLAS Calorimeters

Surrounding the ID is the ATLAS calorimeter system. The system contributes to identifying and measuring electrons, photons, jets and missing transverse energy. The calorimeters are designed to completely stop and absorb almost all particles from the collision point.

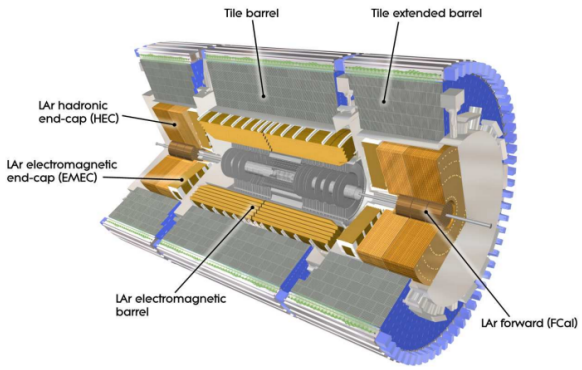


Figure 17: An illustration of the ATLAS Calorimeter system and the sub-detectors. Figure taken from [41].

ATLAS uses two types of calorimeter, one with liquid argon as the active medium (LAr calorimeter), and the other with a scintillator as the active medium (tile calorimeter). Both types of calorimeter are designed to measure the energy of incoming particles through ionisation. The calorimeters are designed with a dense material which acts as an absorber to reduce the energy of particles through scattering. The scattering induces electromagnetic or hadronic showers. These showers then pass through and ionise an active medium interspersed between the absorbers. The energy of the particle is inferred from the total ionisation.

The calorimeter system is composed of two main sub-systems: electromagnetic and hadronic calorimeters. The electromagnetic calorimeters are lead liquid argon calorimeters designed with an accordion geometry, as shown in Fig. 18. The absorbers are lead plates with a thickness of 1.13-2.22 mm whilst liquid argon is the active medium between the plates, chosen for its linear behaviour, stability of response over time and radiation-hardness [41]. The accordion geometry is designed to provide full coverage in ϕ and uniform performance in terms of linearity and resolution.

The calorimeters are contained within a system of three cryostats: one in the barrel, housing the electromagnetic barrel calorimeter, and two in the end-caps, housing the ElectroMagnetic End-cap Calorimeter (EMEC). The LAr calorimeters are composed of several layers. In the precision measurement region, ($|\eta| < 2.5$), covered by both the barrel and the EMEC, the calorimeters consist of three layers (see Fig. 18). In the high-η

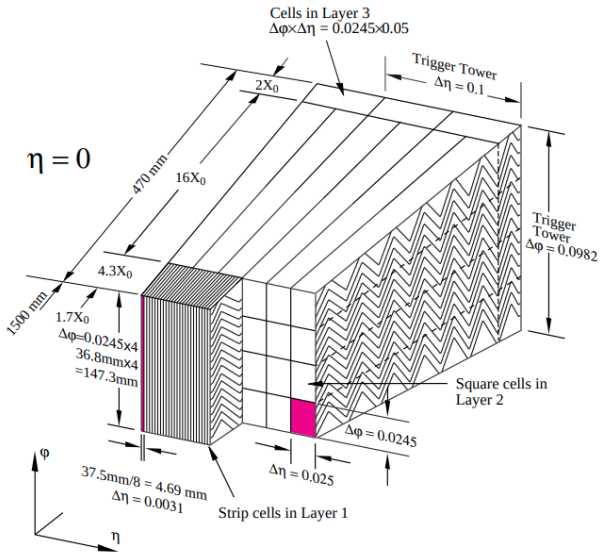


Figure 18: A slice of the electromagnetic calorimeter, showing the division of the three layers of the sub-detector in η and ϕ. Figure taken from [41].

regions ($2.5 < |\eta| < 3.2$), the calorimeters consist of two layers. The first layer is designed with a finer granularity to measure the position of the incoming particles accurately. The second layer collects the majority of the energy from the outgoing showers, and the third layer collects the tail end of the shower.

The second component of the ATLAS calorimeter system is the hadronic calorimeter. The hadronic calorimeter is composed of three sub-detectors: Tile calorimeters, LAr hadronic end-cap calorimeter and the LAr forward calorimeter. The three detectors are installed surrounding the electromagnetic calorimeters and provide almost hermetic converge up to $|\eta| < 4.9$. The central region of the ATLAS detector ($|\eta| < 1.7$) is covered by tile calorimeters [45]. Tile calorimeters are a type of sampling calorimeter which uses steel plates as the absorbing material and plastic scintillating tiles as the active medium. Scintillation light emitted by the tiles when ionised by hadronic showers and is transmitted by wavelength shifting fibres to photomultiplier tubes, resulting in a readout [45]. Tile calorimeters were chosen to maximise radial coverage of the ATLAS detector for the lowest cost [41].

The tile calorimeters are supported by the LAr hadronic end-cap Calorimeter (HEC) and the LAr forward calorimeter (FCal), which extend the hadronic calorimetry coverage up to $|\eta| < 4.9$. The HEC uses copper plates as the absorbing material and consists of a wheel in each end-cap, positioned behind the EMEC and housed in the same cryostat. Analogously to the barrel calorimeter, the wheel is divided into layers, with four layers in total. The thickness of the plates varies from 25 mm to 2.03 m across the layers. The FCal is a copper-tungsten LAr calorimeter housed in the EMEC cryostats, positioned behind the HEC and designed to cover the region closest to the beam ($3.1 < |\eta| < 4.9$). As well as extending the coverage of the hadronic calorimeters, the FCal is designed to also extend the coverage of the electromagnetic calorimeters. The FCal is composed of three modules in each end cap: one copper module designed for electromagnetic measurements

and two tungsten modules optimised for hadronic measurements.

2.3.4 Muon Spectrometer

The outermost component of the ATLAS detector is the muon spectrometer, surrounding the calorimeters. The muon spectrometer is a tracking detector primarily designed to identify and measure the charge and momentum of muons[2]. The muon spectrometer is immersed in a magnetic field produced by an air-core toroid system. The system is designed to generate a powerful magnetic field with a high bending power necessary for high-precision momentum measurements whilst minimising detector material to reduce multiple-scattering effects.

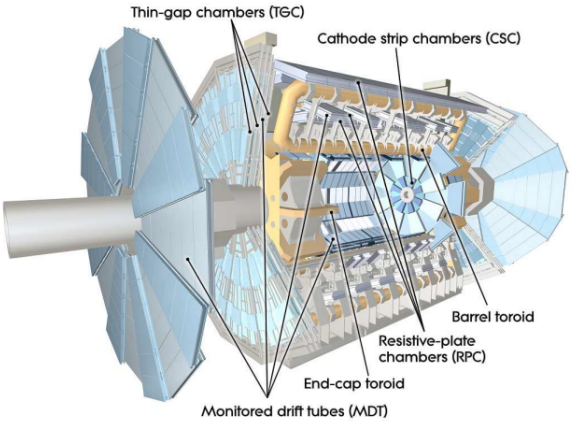


Figure 19: An illustration of the ATLAS Muon Spectrometer system and each of the sub-detectors. Figure taken from [41].

The muon spectrometer consists of multiple sub-detectors. Precision-tracking measurements over the $|\eta| < 2.7$ range are provided by Monitored Drift Tube chambers (MDTs). The drift tubes are filled with an $Ar - CO_2$ mixture at an absolute pressure of 3 bar. As with the TRT, muons passing through the drift tubes ionise the gas. The resulting electrons are collected by a central tungsten-rhenium wire across a 3080 V potential difference. The chambers themselves are composed of three to eight layers of drift tubes and achieve an average resolution of 35 μm per chamber.

The efficiency of an MDT system would be limited in the innermost end-cap layer in the forward region ($2 < |\eta| < 2.7$). The MDTs can operate at a maximum counting rate of 150 Hz/cm^2, which is exceeded in this region. Rather than MDTs, cathode strip chambers (CSCs) are used in this region to provide a much higher rate capability. The CSCs are multiwire proportional chambers with arrays of cathodes split into strips perpendicular and parallel to the wires. The CSCs operate similarly to the drift tubes in an $Ar - CO_2$ gas mixture. Muons passing through the CSCs ionise the gas, and electrons flow to the wire, away from the cathodes, while the positive ions flow towards the cathode strips. The track position is measured by the charges induced on the cathode strips.

[2]And any other charged particles that pass through the calorimeters

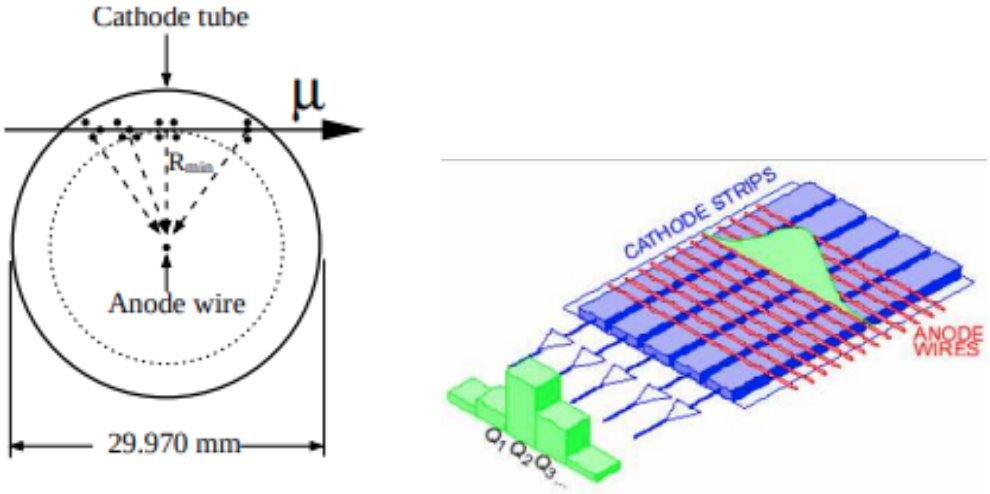


Figure 20: A graphical illustration of the drift tubes and cathode strip detector components. Figure taken from [41].

Additionally, the muon spectrometer is designed with a trigger system with the ability to identify muons passing through the spectrometer within a few nanoseconds. The trigger system is composed of two trigger chambers, Resistive Plate Chambers (RPC) in the barrel region ($|\eta| < 1.05$) and Thin Gap Chambers (TGC) in the end-caps ($1.05 < |\eta| < 2.4$). Both chambers are designed to deliver track information within $15 - 25\ ns$ by measuring the coordinates of the muon trajectory, both in the bending plane, η and in the non-bending plane, ϕ. The trigger system complements the precision-tracking chambers. The MDTs determine the coordinates of the muon in the bending plane. This measurement is then matched to the readout of the trigger chamber, and the non-bending coordinate of the trigger is taken as an additional measurement.

The RPCs are gaseous parallel electrode-plate detectors. The detector is composed of two parallel resistive plates separated by a distance of 2 mm, with an electric field applied across the plates. Muons passing through the gas (a mixture of $C_2H_2F_4$/Iso-C_4H_{10}/SF_6) ionise the gas, and a signal is read out at the anode. The TGCs are multi-wire proportional chambers with a wire-to-cathode distance of 1.4 mm and a wire-to-wire distance of 1.8 mm [41], as shown in Fig. 21. Muons passing through the TGCs ionise the gas mixture of CO_2 and n-C_5H_{12}, with the ions passing to the wires. The high electric field around the TGC and the small distance between wires leads to a high-resolution time measurement.

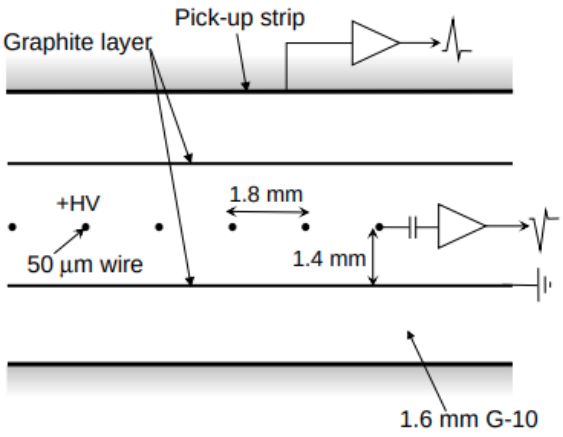


Figure 21: A graphical illustration of the multi-wire proportional chambers. Figure taken from [41].

2.3.5 Forward Detectors

The ATLAS detector also has three small detector systems installed in the forward region of the detector. The location of the three systems, relative to the ATLAS IP, are shown in Fig. 22. The first in order of distance, located at ±17 m from the IP, is the LUminosity measurement using Cerenkov Integrating Detector (LUCID), a dedicated detector for ATLAS luminosity monitoring. LUCID detects inelastic p-p scattering in the very forward direction. The detector provides online monitoring of the instantaneous luminosity and measurements of the total integrated luminosity.

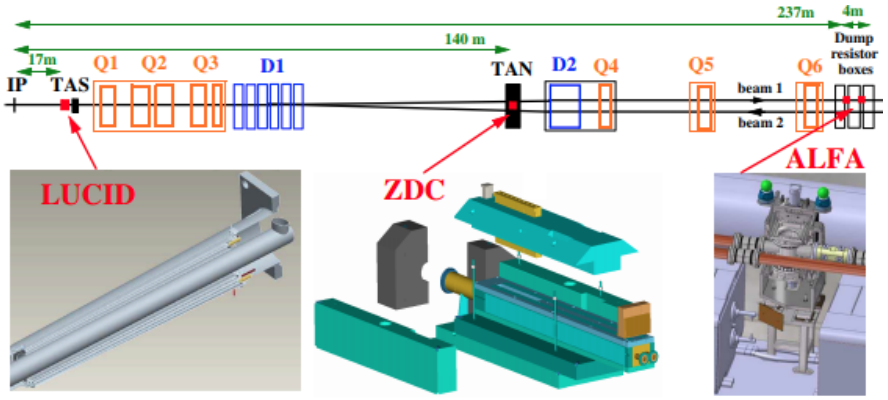


Figure 22: The ATLAS forward detectors and their distance from the interaction point. Figure taken from [41].

The second detector, located ±140 m from the IP, is the Zero-Degree Calorimeter (ZDC). The ZDC serves the purpose of detecting forward neutrons in order to determine the centrality of heavy-ion collisions. The ZDC module consists of layers of quartz rods and tungsten plates designed to detect particles in the extremely forward region, $|\eta| > 8.3$. The final detector system, located approximately ±240 m from the IP, is the Absolute

Luminosity For ATLAS (ALFA) detector. ALFA is composed of scintillating-fibre trackers located inside Roman pots [46] up to 1 mm from the beam line. The primary purpose of the detector is to measure elastic p-p scattering in the forward region, which, by the optical theorem [47], can be used to extract the total cross-section and luminosity at the ATLAS detector. These measurements are used to complement and calibrate the LUCID detector.

2.3.6 Trigger and Data Acquisition

The Trigger and Data Acquisition (TDAQ) system is responsible for filtering and collecting the measured experimental data to be used for physics analysis. The LHC provides collisions at a frequency of 40 MHz, and the ATLAS detector has approximately 100 million readout channels. Collecting the data from every collision would require data storage and processing beyond the current technical and computational capabilities. The ATLAS detector, therefore, employs the TDAQ system to filter and collect data. The TDAQ system uses a two-stage trigger system [48] to collect data, illustrated graphically in Fig. 23.

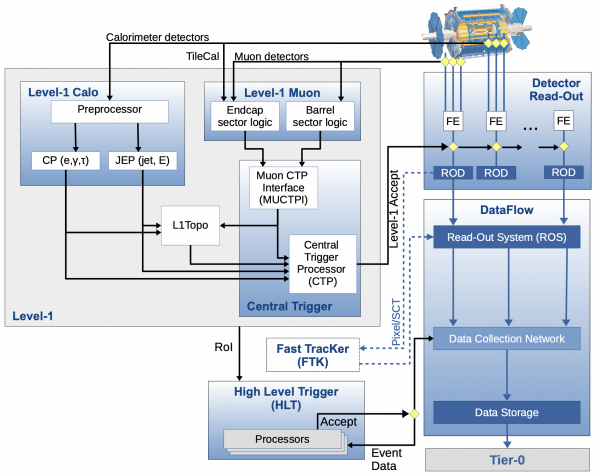


Figure 23: A graphical illustration of the ATLAS trigger and DAQ, showing the flow and processing of data collected by the ATLAS detector. Figure taken from [48].

The first-stage trigger is the Level-1 (L1) trigger. The L1 trigger is a hardware-based system that selects events using a limited amount of detector information. The inputs to the trigger are reduced granularity raw data provided directly from the muon triggers (the RPC and TGC) and from the entire calorimeter systems [48]. The L1 trigger selects events by considering event-level quantities, properties of physics objects or topological requirements, such as angular distances [48]. The trigger is designed to search for high transverse-momentum muons, electrons, photons, jets, τ leptons decaying into hadrons, and a large missing energy. The second-stage trigger is the High-Level Trigger (HLT). The HLT is software-based [48] and uses a two-stage reconstruction sequence to select events. Events selected by the L1 are first reconstructed using dedicated fast trigger algorithms to reject events. Events passing this first selection are then reconstructed with a more

precise set of algorithms similar to the offline event reconstruction and are subject to final selection criteria. Events selected by the HLT are sent to permanent storage for offline reconstruction and are exported to the Tier-0 computing facility [49] at CERN. The final data is subject to a data quality procedure described in Ref. [50].

2.4 Reconstruction

Collisions of physics interest are selected by the trigger and primarily originate from the leading parton-parton scattering interaction, referred to as the hard-scattering process. After selection, the trigger stores the detector output to be processed for physics analysis. The readout is processed to identify and measure the kinematics of particles resulting from the collision, a procedure referred to as reconstruction. The focus of this section is to discuss the general reconstruction methods relevant to the analysis presented in this thesis.

2.4.1 Tracking

Tracking, short for track reconstruction, is one of the major reconstruction components in measuring the full kinematics of an event. Track reconstruction refers to the process of determining the trajectories, referred to as tracks, of charged particles which have passed through the ATLAS tracking system. The ATLAS detector is immersed in a magnetic field such that the trajectories of charged particles passing through the detector are curved. The primary purpose of the tracking is to identify and measure the charge and the momentum of charged particles, information which can be deduced from the curvature of a track. Furthermore, the reconstructed tracking information is also used by almost every other reconstruction process and is paramount to the full reconstruction chain.

A track describes the trajectory of a charged particle measured by the detector. A track is parameterised using five parameters and a reference point, typically the nominal interaction point, using a perigee representation, [51] as shown in Fig. 25. The first two parameters of the track are the transverse (d_0) and longitudinal (z_0) impact parameters, which are defined respectively as the transverse and longitudinal distance closest to the interaction point. The remaining parameters are the azimuthal angle ϕ and the polar angle θ of the track momentum, and the ratio q/p of the charge of the track q, divided by the magnitude of its momentum p.

Clustering

Charged particles passing through the ID deposit energy in the Pixel and SCT sub-detectors. The deposited charge is collected across multiple adjacent pixels and strips. The first step of the track reconstruction chain is a pre-processing stage, which groups the raw detector output from adjacent pixel and strip components to assemble clusters. The clusters are then converted to three-dimensional *space-points*, representing the location where a charged particle passes through the active medium of the ID. Space-points are constructed in the SCT from pairs of clusters on both sides of the strip module and in the pixel from individual clusters. In a dense environment, multiple charged particles can leave charge deposits in the same module. A cluster created by the charge deposit

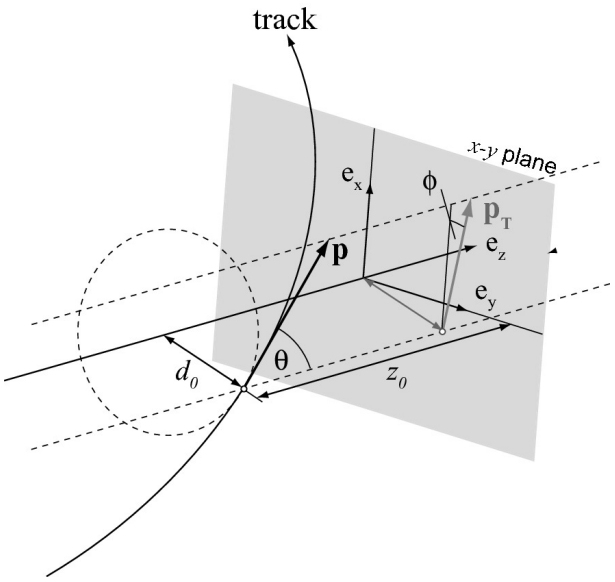


Figure 24: An illustration of an ATLAS-defined track and the corresponding track parameters. Figure taken from [52].

of a single particle is referred to as a single-particle cluster, and a cluster created by the charged deposit of multiple particles is referred to as a merged cluster [51].

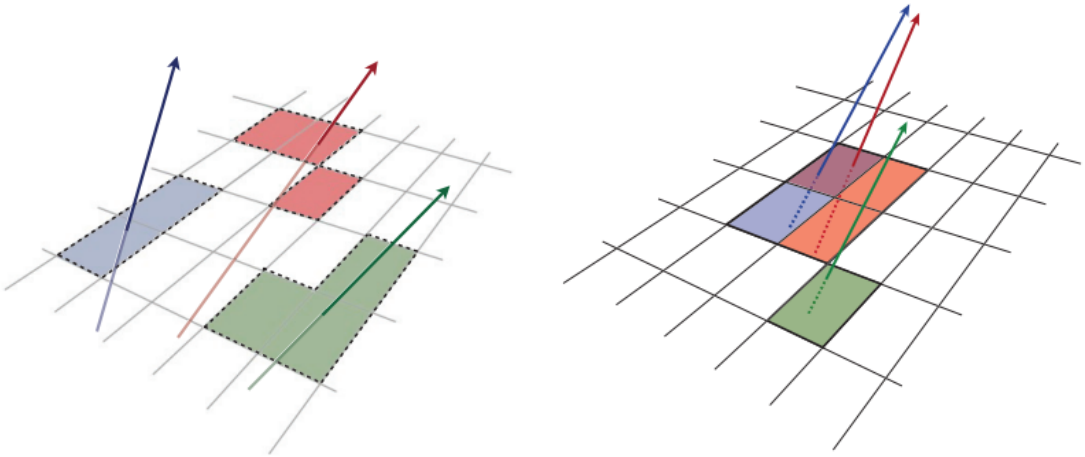


Figure 25: The left plot shows single-particle clusters whilst the right plot shows a merged particle cluster. Figure taken from [51].

Track Seeding and Formation

The formation of clusters and space-points is a preliminary stage of the track reconstruction chain. The challenge of track reconstruction is one primarily of combinatorics: sets of space-points must be correctly associated with the track of the charged particle. ATLAS track reconstruction begins by considering all possible combinations of space-points which could feasibly originate from a charged particle. This approach minimises the number of

correct tracks which are missed during reconstruction. From this collection of tracks, the track reconstruction algorithms then act to select the correct tracks and reject the incorrect tracks, referred to as fake tracks.

The first step of this chain is track seeding. Track seeding iteratively searches for combinations of three space-points compatible with a charged particle's track. Each set of three space-points is referred to as a track seed. Seeds are constructed from three space-points to maximise the number of tracks potentially considered while also allowing for a crude estimate of the track momentum and the impact parameters. One of the significant challenges with the combinatorics of tracking is the computational cost of reconstruction. The track seeding algorithm applies selection criteria on the track seed momentum and impact parameters in order to reject low-quality seeds. The momentum and impact parameters are estimated by assuming a uniform magnetic field with helical tracks relative to the centre of the interaction region.

In addition to seeding selection criteria, the track seeding algorithm also assigns an order of priority, depending on the sub-detector, which records the space-points of the seed. The priority is assigned to seeds generated in the SCT, then the Pixel Detector, and then mixed-detector seeds. Seeds must also have an additional "confirmation" space-point, which is compatible with the estimated track of the seed before being subject to the next stage of track reconstruction. One of the major challenges of track reconstruction is the dense, high pile-up environment of the LHC.

Track Candidates and Ambiguity Resolution

After the formation of track seeds, the track candidates are formed by extrapolating the track seeds with additional space-points in both the Pixel Detector and the SCT. The construction of track candidates is performed using a combinatorial Kalman filter [51]. The Kalman filter acts to create multiple track candidates from a seed if there are multiple additional space-points compatible with the track. After the construction of track candidates, all possible permutations of feasible tracks have been considered by the tracking algorithms. Many of the candidates consist of multiple incorrectly assigned space-points and are thus 'fake tracks'. In order to resolve any overlaps between track candidates and to reject fake tracks, an ambiguity resolution algorithm is applied to the track candidates.

The ambiguity solver acts to score track candidates based on a range of track quality criteria. Tracks are scored on the quality of the clusters reflecting the intrinsic resolution of the corresponding sub-detectors. The track score is also reduced if the track passes through an active medium of the detector without registering a space-point, referred to as a hole. Further, the score considers a χ^2 of the track fit and penalises track candidates with a poor fit. Finally, the logarithm of the track momentum is chosen to promote high-p_T tracks and suppress the number of low-p_T tracks, which typically have incorrectly assigned clusters. A lower-quality track which shares a larger number of associated hits with higher-quality tracks will be rejected, and the shared hits assigned to the track will be retained.

The final stage of ambiguity solving is dealing with clusters present in multiple track candidates which have passed the first selection. The accepted tracks are permitted to

have a limited number of shared clusters to compensate for the high-density environments. The treatment of clusters first depends on the type of cluster. A merged cluster should be assigned to multiple track candidates, but a single particle cluster assigned to multiple candidates, referred to as a shared cluster, is a strong indicator of an incorrect assignment. The ambiguity solver works to iteratively process tracks with shared clusters between two or more tracks. Clusters are removed if either of the tracks fails the shared cluster selection criteria, and the tracks are reprocessed without the cluster. The final tracks passing the ambiguity resolution stage are subject to a set of track parameter cuts. An important aspect of this stage is correctly identifying merged and shared clusters. To improve the identification efficiency, an artificial neural network is applied to identify merged clusters [51].

2.4.2 Electron Reconstruction

The identification and reconstruction of stable particles such as electrons, muons and photons is one of the foundations of the ATLAS experiment. This subsection outlines the aspects of ATLAS reconstruction of electrons following [53, 54], with the reader referred to the references for a more detailed description of the reconstruction process and the Run-2 performance.

Electrons produced in the ATLAS detector are measured in the central region of the detector, encompassed by the electromagnetic calorimeters. An electron passing through the central region will first traverse the inner detector before being absorbed by the calorimeters. The reconstruction of electrons is centred on the matching of reconstructed tracks in the ID with calorimeter-measured localised clusters of energy deposits. This in itself can be challenging as electrons passing through the detector medium can interact and lose energy due to bremsstrahlung. Radiated photons can also convert into electron-positron pairs. This process can occur within the ID and can potentially lead to multiple tracks associated with the same cluster in the EM calorimeter. Moreover, hadronic jets and converted photons can leave electron-like signatures in the detector and be incorrectly identified as electrons. The ATLAS electron reconstruction algorithms are designed to address the challenges of identifying prompt electrons.

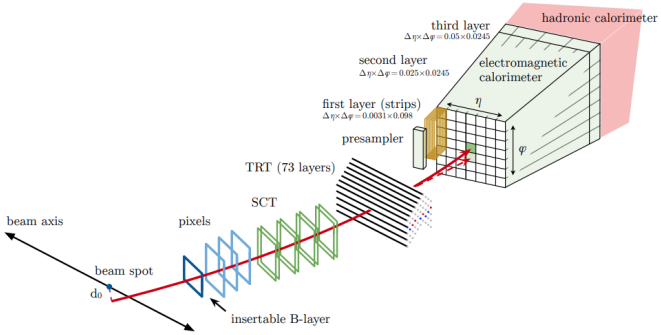


Figure 26: The event signature of an electron passing through the ID and being absorbed by the EM calorimeters. Figure taken from [53].

Electron Candidates

The first stage starts with the formation of electron seed-clusters to isolate localised electron-like energy deposits in the EM calorimeters. Seeds are formed by dividing the EM calorimeters in the $\eta \times \phi$ plane into a grid of 200×256 elements, referred to as towers, with a size $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ [53], shown in Fig. 26. For each of these elements, the energy in the cells in the longitudinal layer, corresponding to the three calorimeter layers and the pre-sampler, is summed to give the tower energy. A sliding window algorithm [55] is then applied, which traverses the EM calorimeters by iteratively moving through the constructed towers in the η or ϕ direction and searching for localised energy deposits. Seed clusters are formed from energy deposits with a summed transverse energy greater than 2.5 GeV, and overlapping seeds are removed with the higher energy seeds prioritised. The efficiency of the seed-cluster algorithm ranges but performs at an efficiency above 95% for $E_T > 7$ GeV and rises to above 99% for $E_T > 15$ GeV [53].

After cluster seed formation, the second stage of electron reconstruction proceeds with track association. The reconstructed tracks are loosely matched to clusters using the distance in the $\eta - \phi$ plane of the tracks extrapolated through to the EM calorimeters. A secondary fitting procedure is applied to tracks matched to a cluster and passing selection criteria on the track parameters. The fitting is performed using a Gaussian-sum filter (GSF) [56], which is based on the Kalman filter. The GSF accounts for the effects of bremsstrahlung, which reduces the electron momentum, along with multiple scattering [53]. The final stage of electron reconstruction is the formation of electron candidates. This stage matches track candidates to cluster seeds to construct electron candidates.

When forming candidates, if there are multiple tracks matched to a cluster, the reconstruction procedure applies an algorithm to select a primary track. The track is chosen based on the cluster-track distance in η and ϕ, and on the track parameters [53]. The candidate selection also applies criteria to remove candidates that appear to be photons. Each candidate must be associated with a track with at least four hits in the SCT, with the track having no association to a vertex from a photon conversation. Candidates are also selected based on the electron energy and transverse momentum, the presence of a pixel hit and secondary vertex information [53, 56]. The final stage of candidate selection determines the cluster size from the seed clusters. The energy of the clusters is determined by adding the energy of cells around seed clusters, with the final energy calibrated to reconstruct the electron energy correctly.

Electron Identification

After the reconstruction of electron candidates, the electron reconstruction chain applies an electron identification procedure to determine if the electron is truly a prompt electron. The procedure applies a likelihood-based (LH) identification, considering inputs from the tracking system and the calorimeters. These inputs include shower shapes, TRT information, track-cluster matching quantities, track parameters and variables parameterising the effects of bremsstrahlung [56]. A detailed list of the parameters and the fit can be found in [53, 56]. ATLAS electron identification provides four working points (WP): VeryLoose, Loose, Medium, and Tight. The WPs are defined for fixed values of the LH discriminant, each with an increasing threshold of the discriminant. The WP are chosen

to cover the broad range of physics requirements necessary for different analyses. The ATLAS recommenced WP for analyses is Medium and is applied in this experiment.

2.4.3 Muon Reconstruction

This subsection discusses the ATLAS reconstruction of Muons for Run-2 following the Refs. [57, 58]. Muons leave a unique signature in the detector. The energy lost due to bremsstrahlung for muons is far smaller compared to electrons due to the significantly larger muon mass. This means that muons traverse the entire through the detector material and are very minimally ionising. Muons are reconstructed using data collected by the ID, the calorimeters and, of course, the MS. The muon reconstruction chain begins by reconstructing tracks in the MS. The algorithm first identifies hit patterns inside each of the muon chambers to form segments. In the MDT chambers, the segments are reconstructed by a Hough transform [59], which searches for hits and fits short line segments in each layer. Meanwhile, the RPC and TGC hits are used to measure the azimuthal coordinates of the track. The segments in the CSC detector are reconstructed by a separate combinatorial search in both the η and the ϕ plane.

After the formation of segments, muon track candidates are constructed by fitting hits from segments in each of the different layers. This is done by a segment-seeded combinatorial search [58], which begins by using the segments constructed in the middle layers of the detector as seeds and then extends the search to segments in the inner and outer layers. Preliminary track candidates are then formed by fitting hits using a loose constraint on the IP and assuming a parabolic trajectory to approximate a muon bending in a magnetic field. Candidates are selected based on the fit quality and are subject to an overlap removal, which handles cases where several track candidates use the same segment. A final sχ^2 global fit is performed that accounts for the effects of interactions with the detector material and detector misalignments [57]. The track candidates are then subject to a final set of selection criteria based on the χ^2 fit.

After the formation of muon track candidates in the MS, the reconstruction procedure is divided into four primary strategies to combine the information from individual sub-detectors to reconstruct the muon track. This leads to the definition of a number of muon types: combined (CB), segment-tagged (ST), calorimeter-tagged (CT) and muon-spectrometer extrapolated (ME) [57]. The first type of reconstructed muon is referred to as a combined muon and is formed by matching an MS track to a track reconstructed independently in the ID. The tracks are matched using an outside-in pattern recognition [58] where the muon track from the MS is extrapolated inward to search for a matching track in the ID; the extrapolation takes into account the energy loss in the calorimeters. The combined track is formed with a global refit using the hits from both sub-detectors.

A second type of muon, referred to as a segment-tagged muon, is reconstructed by extrapolating an ID track to a reconstructed MS segment. A matched track is identified as an ST muon, and the muon parameters are inferred from the ID track. This type of muon passes through only one layer of the MS, typically due to its low p_T. Analogously, CT muons are reconstructed by extrapolating an ID track to the calorimeter and matching to energy deposits consistent with those of a minimally ionising particle. CT muons have large background contributions but are used to recover acceptance in the regions of the

MS where there is a lower acceptance. To reduce the background, CT muons require an ID track with a minimum p_T of 5 GeV, as opposed to 2 GeV used by the other strategies. Lastly, for muons with MS tracks which cannot be matched to an ID track, the parameters are extrapolated to the beamline and reconstructed as an ME muon. The parameters of the muon are defined from the IP. The primary purpose of these muons is to exploit the full acceptance of the MS, which partly extends beyond the acceptance of ID.

Muon Identification

Reconstructed muons are subject to an identification procedure, which applies selection criteria on the muon track parameters, the track fit, and a number of variables that test the compatibility of the measurements made by the different subdetectors [57], as outlined in [57, 58]. The aim of the identification procedure is to reduce the contribution of non-prompt muons, coming primarily from pion and kaon decays. Muons are identified by four different working points (WPs), which apply specific selection criteria. The WPs are defined to cover the extensive range of requirements needed for a variety of physics analyses. The four WPs are referred to as Medium, Loose, Tight, and High-p_T, where Loose, Medium, and Tight are ordered inclusive categories such that the looser criteria are a subset of the subsequent stricter criteria.

Medium muons are recommended as the default general-purpose selection for physics analyses in ATLAS. The criteria are only applied to CB or ME muon candidates. Each track is required to have a minimum of 3 hits in at least two MDT layers [58]. In addition, a requirement is applied to the compatibility between ID and MS momentum measurements. This selection is applied to suppress the background contribution coming from hadrons misidentified as muons. The Loose muon identification working point has a looser selection criterion than the Medium but is still designed to provide good-quality muon tracks. The loose WP considers all muon types. CT and ST muons are restricted, however, to the $|\eta| < 0.1$ region. The other WPs are not used in this analysis but are described in Ref. [58].

2.4.4 Jet Reconstruction

All particles which can be observed and detected are QCD colour-neutral. Particles carrying a QCD colour charge, namely quarks and gluons, cannot exist in a free unbound colour state due to colour confinement. Moreover, the LHC is a proton-proton collider, and parton collisions often produce colour-charged particles. These particles hadronise almost immediately to form bound states of colourless hadrons. The ATLAS detector observes hadrons resulting from a high-momentum quark or gluon production as jets. Formally, a jet is defined as a collimated spray of hadrons, which is the product of the hadronisation of quarks and gluons. The reconstruction of jets and their kinematic properties is crucial for a wide range of physics analyses. In addition, good jet reconstruction is crucial for missing transverse energy reconstruction. This section briefly outlines the ATLAS jet reconstruction algorithms and tagging procedures used by this experiment. Further detail can be found in the cited literature. The tagging of jets refers to the identification of the primary source of a jet. The ATLAS Run-2 b-jet identification performance is outlined

in [60]. ATLAS has several b-jet identification algorithms. In this experiment, the DL1r tagger [61] is used for b-jet tagging.

The objective of a jet reconstruction algorithm is to collate groups of hadrons and assign a collective measurement of the kinematic properties of the jet. The general definition of a jet is fairly broad but one of the crucial requirements is that it must be possible to model jets theoretically. This sets important constraints on the definition of a jet. In order to make comparisons with theoretical cross-sections, the jet reconstruction algorithm must ensure the jet definition is well-defined in terms of perturbative QCD. There are two conditions: infrared safety and collinear safety [62]. Infrared safety ensures that the jet properties are independent of soft parton emission. Collinear safety ensures the jet properties do not change if a parton is replaced with two partons with the same total energy as the original. Provided the jet reconstruction algorithms apply a definition of a jet that is safe in terms of perturbative QCD, order-by-order calculations in perturbation theory can be performed to model experimental data.

In the ATLAS detector, jets traverse the ID and EM calorimeters, and are fully absorbed by the hadronic calorimeters. A typical jet signature consists of energy deposits in both types of calorimeter and tracks in the ID. The first step of jet reconstruction is a preliminary stage to reconstruct the tracks in the ID and the energy deposits in the calorimeters. The track reconstruction procedure remains as described in the earlier sections of this chapter. The ATLAS calorimeters are segmented laterally and longitudinally, allowing for the three-dimensional reconstruction of particle showers. The reconstruction is based on the formation of topological cell clusters (topo-clusters) in the ATLAS calorimeters [63]. The function of the clustering algorithm is to reconstruct and separate different continuous energy showers, hadronic and electromagnetic, and to suppress noise contributions [63, 64]. A complete overview of the topo-clustering procedure is beyond the scope of this thesis, with a detailed description given in [63].

Jets reconstruction in ATLAS with the anti-k_T algorithm [65]. A detailed explanation of the algorithm is beyond the scope of this thesis and is outlined in the referenced literature. Crucially, the algorithm can reconstruct jets with both tracks and topo-clusters as physics objects and is perturbative-safe. The reconstruction procedure used in this analysis follows an approach referred to as particle flow [64], which uses combined measurements of tracks and topo-clusters as inputs to the anti-k_T algorithm. There are several benefits to using measurements from both sub-detectors. The ID has a higher track momentum resolution for low-energy charged particles compared to the energy resolution of the calorimeters and can reconstruct tracks down to 400 MeV, which would fail to pass the noise threshold when reconstructing the calorimeter clusters [63]. However, the calorimeters provide better energy resolution compared to the track resolution for high-energy particles. The two sub-detectors, therefore, working in combination, can provide the most optimal event reconstruction.

The caveat of the particle flow approach is the potential risk of double-counting. The ATLAS particle flow algorithm applies a boolean decision as to whether to use measurements from the tracker or the calorimeter, depending on the instrumental resolution. If a track measurement is used, the algorithm will extrapolate the track to identify the corresponding signal in the calorimeters correctly. The energy deposits are subtracted from the calorimeter measurement to avoid double counting. This approach, therefore,

hinges on the ability to accurately remove the energy deposits of tracked particles without removing energy deposited by other particles. The particle flow algorithm is applied to jets within the acceptance of the ATLAS ID ($|\eta| < 2.5$), while in the forward regions, the reconstruction uses only calorimeter information. Further discussion on the calibration of ATLAS reconstructed jets is given by [66].

2.4.5 E_T^{miss} Reconstruction

This subsection discusses the ATLAS reconstruction of the missing transverse energy (E_T^{miss}), as described in Ref. [67]. Protons are collided by the LHC with a net zero transverse momentum. The sum of the momenta of the constituent particles coming from the collision by conservation of momentum is also zero. However, the ATLAS detector cannot detect all types of particles, such as neutrinos. This leads to a non-zero missing transverse energy. The E_T^{miss} is a crucial measurable quantity as it characterises the transverse momentum carried by undetected particles. This experiment measures a VBS process in the leptonic decay channel. The final state consists of two neutrinos and has a large E_T^{miss}.

Reconstructing the E_T^{miss} is challenging as it requires the reconstruction of all of the physics objects in an event. There are two contributions to the met: hard- and soft-particles. The hard-particles describe fully reconstructed particles and jets, and the soft-particles describe reconstructed charged-particle tracks. The missing energy in the $x(y)$ direction $E_{x(y)}^{\text{miss}}$, is defined as [67],

$$E_{x(y)}^{\text{miss}} = -\sum_{i \in \text{hard}} p_{x(y),i} - \sum_{j \in \text{soft}} p_{x(y),j} \tag{79}$$

where $E_T^{\text{miss}} = |(E_x^{\text{miss}}, E_y^{\text{miss}})|$. The challenge for reconstructing the E_T^{miss} is that the reconstruction process for various physics objects is performed independently. This can lead to situations where the same detector signal is used to reconstruct multiple physics objects, leading to double counting. For example, the same calorimeter signal is used to reconstruct an electron and a jet. Therefore, the E_T^{miss} reconstruction procedure applies strict selection criteria for physics objects and an explicit ambiguity resolution to ensure that the detector signal is not double-counted. A detailed overview of the ambiguity resolution procedure is not necessary for the remainder of this thesis, and the reader is referred to [67] for more detail.

3 ITk Low-p_T Track Reconstruction

The HL-LHC [68] is expected to begin operation in early 2028, generating an average of 200 proton-proton collisions per beam-crossing, with a centre-of-mass energy of 14 TeV. The HL-LHC aims to deliver an integrated luminosity of 3000 fb^{-1}, more than ten times the luminosity accumulated in the first phase of the LHC. To adapt to the incredibly harsh radiation environment and the significant increase in luminosity, the ATLAS experiment will undergo substantial upgrades that will be installed according to the LHC timeline, see Figure 13. As part of the upgrades, the ATLAS experiment will receive a new all-silicon Inner Tracker (ITk) [69], replacing the current Inner Detector. Part of the work undertaken by the author contributed to the implementation of low-p_T tracking to the track reconstruction algorithms of the ITk. This section describes the motivation for this work and outline the study's results, with a complete description provided in [70].

Reconstructing the tracks of low-p_T particles has already been studied for the ATLAS Inner Detector with nominal pileup for the LHC in [71] and is known to pose challenges, both algorithmic and computational. However, the application of low-p_T tracking could be beneficial for a number of analyses and applications, potentially even VBS. Low-p_T tracking was used in early low-μ runs by the Physics Modelling Group (PMG) for generator inputs, e.g. forward energy flow and low-p_T jets and photons, and is also used to validate early tracking performance. Low-p_T tracking is used for photon-induced physics searches as it can help to better reconstruct the underlying event and allow for more accurate selections of signal and background. In addition, low-p_T tracking could potentially improve the identification of hadronic decays and jets originating from charm quarks and may benefit some SUSY models, such as those with a small mass splitting between the chargino and the LSP neutralino [72].

3.1 Inner Tracker

The ITk will be composed of a pixel and a strip sub-detector. The inner pixel detector consists of five barrel layers and pixel end-caps, surrounded by the strip detector with four barrel layers and six end-cap disks. This layout, shown schematically in Fig. 27, has been designed to maintain the same performance as the current ID but for a higher pileup environment.

The most significant upgrade with respect to the ID is the increase in acceptance, from $|\eta| < 2.5$ to $|\eta| < 4$. For physics at the HL-LHC, this is particularly important as it improves pileup suppression. The ITK will also provide better impact parameter resolutions compared to the ID due to the smaller pixel pitch in the innermost pixel layers. The pixel sensors have dimensions 25×100 or 50×50 μm^2 in the ITk [73] compared to 50×250 μm^2 for the IBL [42]. The ITk also sees improved resolution for transverse momentum measurements due to the better silicon strip sensor resolution compared to the TRT. Finally, the tracking efficiency for the ITk is expected to be similar to that of the ID. For particles with $p_T > 1$ GeV, the ITk tracking efficiency at $\mu = 200$ is within 5% of the Run-2 efficiency at $\mu = 38$.

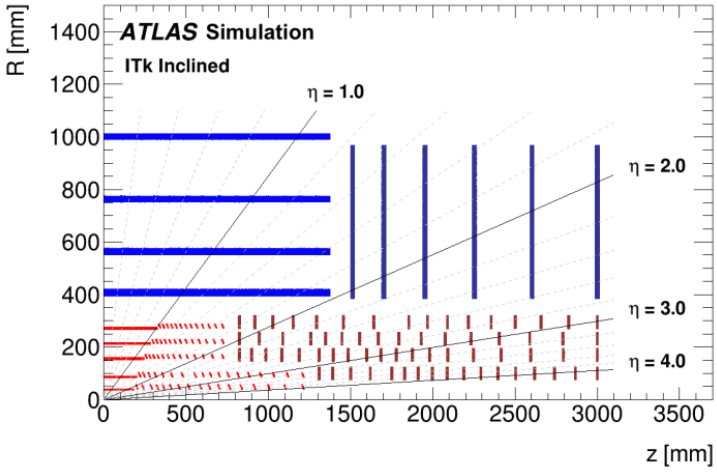


Figure 27: A cross-section of the ITk illustrating the active region of the detector. Figure taken from [69].

3.2 Track Reconstruction

The ITk track reconstruction algorithms follow a similar structure to the ID tracking, albeit with no TRT and with adapted track selection parameters. The ITk track reconstruction can be described as a sequential chain of five distinct algorithms: clustering, space-point formation, track seeding, candidate finding and finally, ambiguity solving. Clustering forms clusters from input hit data provided by the Strip and Pixel detectors. Clusters are used to generate space-points, which are subsequently passed to a seed finder. The seed finder searches for potential track candidates by searching for combinations of three space-points that could resemble a track in both the strip and pixel layers. A Kalman filter is then applied to the seeds to determine the track candidate. The candidates then undergo an ambiguity resolution, which resolves disputes between track candidates sharing clusters and removes any fake or duplicate tracks.

The work presented here examined the adaptation of the track reconstruction algorithms to reconstruct the tracks of lower p_T particles. In order to illustrate the tracking performance, this section first explicitly outlines the metrics used to benchmark tracking. Foremost, the effectiveness of track reconstruction can be gauged with two crucial performance criteria: track efficiency and fake rate. Tracking efficiency is the fraction of prompt particles which are associated with tracks passing a track quality selection [73]. The track efficiency is calculated using the probability that a track matches a truth particle, P_{match}, defined as,

$$P_{\text{match}} = \frac{2N^{\text{pix}}_{\text{common}} + N^{\text{strip}}_{\text{common}}}{N2^{\text{pix}}_{\text{track}} N^{\text{strip}}_{\text{track}}}, \tag{80}$$

where $N^{\text{pix/strip}}_{\text{common}}$ is the number of pixel/strip detector clusters associated with the track and the particle being matched, and $N^{\text{pix/strip}}_{\text{track}}$ is the number of pixel/strip detector hits

assigned to the track. The tracking efficiency ϵ [61] is then defined as,

$$\epsilon = \frac{N_{\text{reco}}((\text{selected, matched})}{N_{\text{truth}}((\text{selected})} \tag{81}$$

where N_{reco} is the number of selected reconstructed tracks matched to a selected truth particle with $P_{\text{match}} > 0.5$ and N_{truth} is the number of selected truth particles. The second important parameter is referred to as the fake rate. Fake tracks occur either from the incorrect assignment of clusters or from (fake) noise hits. The measure used for the fake rate is typically the ratio of the number of reconstructed tracks to the number of generated true particles.

3.3 Low-p_T Tracking

To outline some of the general challenges of low-p_T tracking, this section illustrates examples of simulated tracks with high and low transverse momentum. Simulation of the ITk and track reconstruction algorithms shown in this section are performed within the ATLAS software framework [74]. The simulated events are produced with Monte Carlo generators. Each event is simulated by propagating the corresponding particles through the ITk detector geometry to generate hits. The simulation output is then digitised and submitted for track reconstruction.

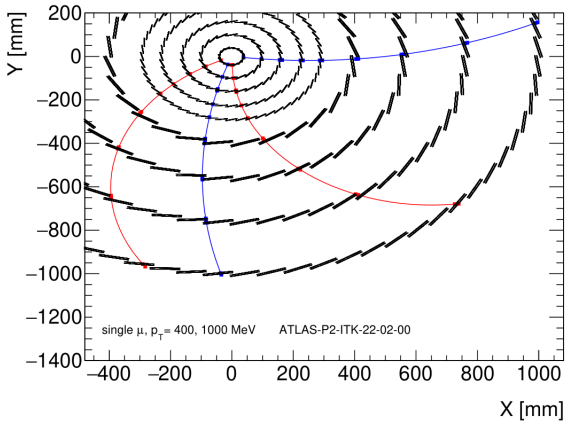


Figure 28: An x-y slice at $z = 0$ of the ITk illustrating interpolated track curvature from hits in sensitive detector elements for single muons with $p_T = 400$ MeV (red) and $p_T = 1000$ MeV (blue). The black components illustrate the barrel layers of the strip and pixel detector.

The default ITk tracking software reconstructs the tracks of charged particles down to a p_T of 900 MeV. A track p_T cut is applied when forming track candidates, which rejects tracks with a p_T lower than 900 MeV. In addition, a looser track p_T cut is applied at the seeding stage on the seed p_T estimate, which rejects seeds that are likely to lead to a low-p_T track. For the Run 2 ID, the track reconstruction algorithms apply a looser selection and reconstruct the p_T of particles down to 400 MeV [51]. The minimum track p_T cut is introduced to maintain reasonable computation times due to pileup. The challenge of

reconstructing tracks below this p_T threshold is one of combinatorics. In a magnetic field, charged particle curvature scales inversely with p_T. Higher-p_T particles propagate with a much smaller curvature and leave near linear hits. By contrast, lower-p_T particles have a much more significant curvature in the detector, shown in the example in Fig. 28.

Reconstructing high-p_T tracks is a task of finding clusters of hits resembling a straight line. In order to reconstruct the tracks of low-p_T particles, the seed search road must be broadened. In a high pileup environment, however, the potential seed combinations grow exponentially with lower p_T. This introduces a greater number of seeds and consequently track candidates, resulting in more complex track fitting and ambiguity solving. The HL-LHC is expected to operate in a pileup environment of up to $\mu = 200$, so reconstructing low-p_T tracks would require vast computational effort, as is shown later in this chapter. This is one of the primary reasons for the increase in the track p_T threshold for the ITk compared to the ID. Furthermore, the increased complexity of tracking low-p_T particles also typically leads to a reduction in tracking efficiency and an increase in the fake rate.

3.4 Tracking Configuration

Whilst low-p_T tracking is currently not viable for all data taking at the HL-LHC, it could still be applied to smaller subsets of events where previous selections limit the number of events. This chapter presents the work done to investigate low-p_T tracking for the ITk. The section begins by illustrating a simple implementation of low-p_T tracking to show some of the considerations when reconstructing lower-p_T tracks.

Cuts	Value	Stage
Track min-p_T	[0.9, 0.4, 0.4] GeV	After Candidate Finding
Track min clusters	[9, 8, 7]	After Candidate Finding
Track min silicon hits not shared	[7, 6, 5]	After Candidate Finding
Track max hits shared	2	After Candidate Finding
Track min pixel hits	1	After Candidate Finding
Track max holes	2	After Candidate Finding
Track pixel max holes	2	After Candidate Finding
Track silicon max holes	2	After Candidate Finding
Track max double holes	1	After Candidate Finding
Track max primary impact parameter	[2.0, 2.0, 10.0] mm	After Candidate Finding
Track max Z impact parameter	200 mm	After Candidate Finding
Seed min-p_T	0.4 GeV	During Seeding
Seed max primary impact parameter	2.0 mm	During Seeding
Seed max z impact parameter	200 mm	During Seeding

Table 2: The default ITk track cuts. Here values tabulated in a list refer to η-dependent cuts, i.e. $[a, b, c]$ applies the cut value a for $|\eta| \leq 1$, b for $1 < |\eta| < 2.6$ and c for $|\eta| \geq 2.6$.

The track reconstruction algorithms are configured using tracking cuts. The track reconstruction algorithms apply track requirements at various stages of the track reconstruction chain to control the fake rate and reduce the processing time. The current default ITk tracking cuts are given in Table 2. A simple implementation of low-p_T tracking is to keep the default tracking cuts and lower the track and seed p_T cut, given in Table

3. This simple low-p_T tracking mode is denoted by LowPt-X, where X refers to the tracking and seed p_T cut in MeV. To show the effect of lowering just the p_T cut, the LowPt-X tracking mode is simulated on a single muon sample. Single muon samples provide an upper bound for track reconstruction efficiency as a single muon is the simplest event the detector can be tasked to reconstruct. In theory, the tracking algorithms should be able to reconstruct the tracks of all single muon samples. Any inefficiencies in reconstructing single-muons indicate possible problems with the track reconstruction algorithms or the tracking cuts.

Low-p_T-X	Value	Default Value
Track min-p_T	[X, X, X] MeV	[900, 400, 400] MeV
Seed min-p_T	X MeV	900 MeV
Remaining	Default ITk	

Table 3: A simple low-p_T tracking configuration that keeps the default cuts and only changes the track and seed minimum p_T to the value X. The tabulated values in a list refer to η-dependent cuts, i.e. $[a, b, c]$ applies the cut value a for $|\eta| \leq 1$, b for $1 < |\eta| < 2.6$ and c for $|\eta| \geq 2.6$.

The seed, candidate and track efficiency of the simple low-p_T mode reconstructing a single-muon sample ($\mu = 0$), with a p_T cut of 200 MeV, is shown in Fig. 29. In the low-p_T region ($p_T < 900$ MeV), the tracking efficiency begins to drop exponentially with decreasing p_T. Single-muons with a very low-p_T ($p_T < 400$ MeV) are reconstructed with a significantly lower tracking efficiency, approaching 60%. For a single muon with no pileup, this is indicative of a problem with the track reconstruction algorithms. From Fig. 29, it can also be seen that the differences between the candidate and track efficiencies, compared to the track efficiency, also grow inversely with p_T. This suggests that the default ITk tracking cuts are too stringent and not suitable for low-p_T reconstruction.

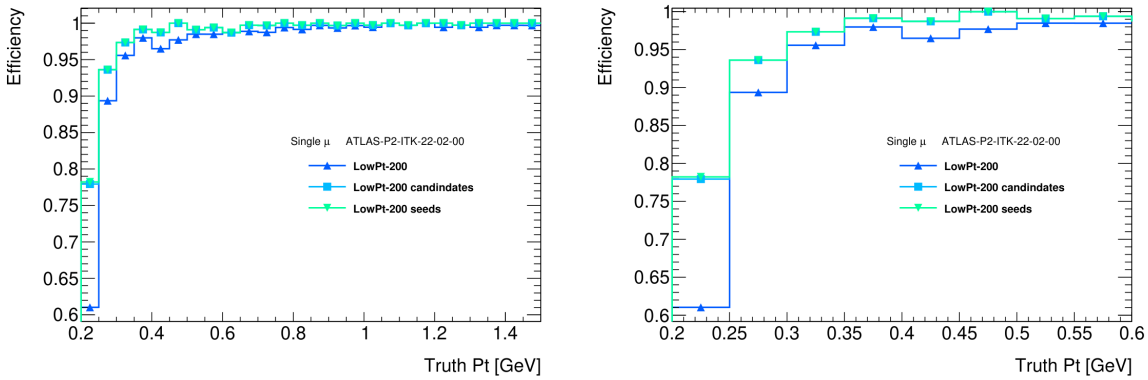


Figure 29: A comparison of the tracking, seed and candidate efficiencies for a single muon sample reconstructed using the Low-p_T-200 tracking configuration.

3.5 Tracking Cuts

Tracking cuts play an important role in tracking reconstruction efficiency. Figure 29 shows a significant drop in efficiency in the $p_T \in (200, 400)$ MeV region. As tracking cuts are primarily applied after candidate finding, the decreases in the track efficiency compared to the seed and candidate efficiency suggest that the decrease is at least partly due to the tracking cuts. The simple low-p_T mode demonstrated in this section applies the default tracking cuts and only lowers the track and seed p_T cut. However, the default ITk configuration has been optimised for tracks above 900 MeV. To illustrate this explicitly, this section investigates an important track selection parameter, the minimum hit requirement. The default requirement is a minimum of 9 hits per track in the low-$|\eta|$ region ($|\eta| < 1$).

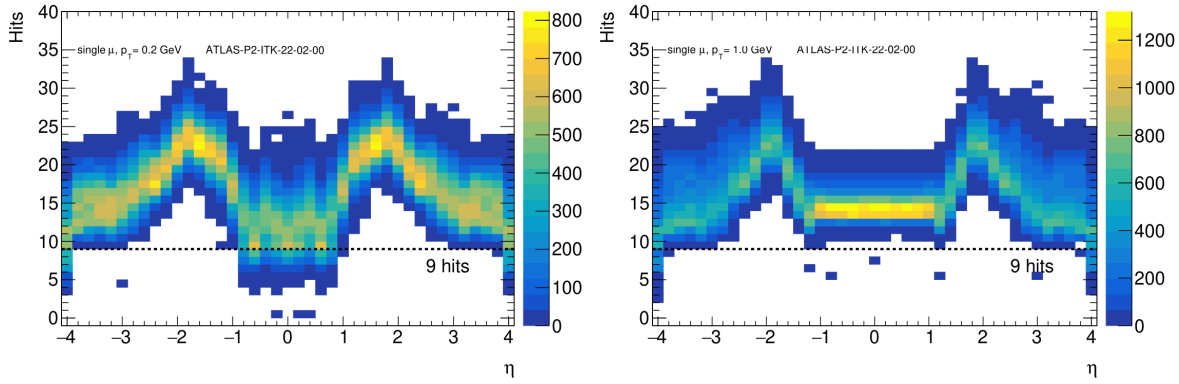


Figure 30: The number of hits as a function of η for single $p_T = 200$ (left) and $p_T = 1000$ (right) MeV muons propagating through the ITk.

Figure 30 compares the number of hits generated by a single muon sample passing through the ITk geometry with 200 and 1000 MeV. The figure shows that a significant proportion of 200 MeV muons will leave fewer than nine hits in the detector. In order to reconstruct the tracks for lower p_T charged particles, the tracking cuts should be adjusted accordingly.

Low-p_T-Modified-X	Value	Default Value
Track min-p_T	[X, X, X] MeV	[900, 400, 400] MeV
Seed min-p_T	X MeV	900 MeV
Track min clusters	[6, 5, 4]	[9, 8, 7]
Track min silicon hits not shared	[6, 5, 4]	[7, 6, 5]
Remaining	Default ITk	

Table 4: A low-p_T tracking configuration that keeps the default cuts except for the listed changes to the track or seed minimum p_T, or to the minimum number of clusters or unshared hits. The tabulated values in a list refer to η-dependent cuts, i.e. $[a, b, c]$ applies the cut value a for $|\eta| \leq 1$, b for $1 < |\eta| < 2.6$ and c for $|\eta| \geq 2.6$.

The tracking cuts for a modified low-p_T reconstruction mode with reduced hit requirements are shown in Tab. 4. The modified mode is denoted by Low-pT-Modified-X where X analogously refers to the track and seed-p_T cut in MeV. A comparison of the efficiency between the original low-p_T and modified low-pT mode is shown in Fig. 31. The figure shows an increase in track efficiency, most notably in the $p_T \in (200, 400)$ MeV region. Crucially, this highlights the need to adapt the tracking cuts to fully utilise the tracking reconstruction algorithms for low-p_T tracks. A large part of the presented work in this chapter investigated the effects of various tracking cuts on low-p_T track reconstruction and searched for the best tracking configuration aimed at maximising the efficiency at low-p_T.

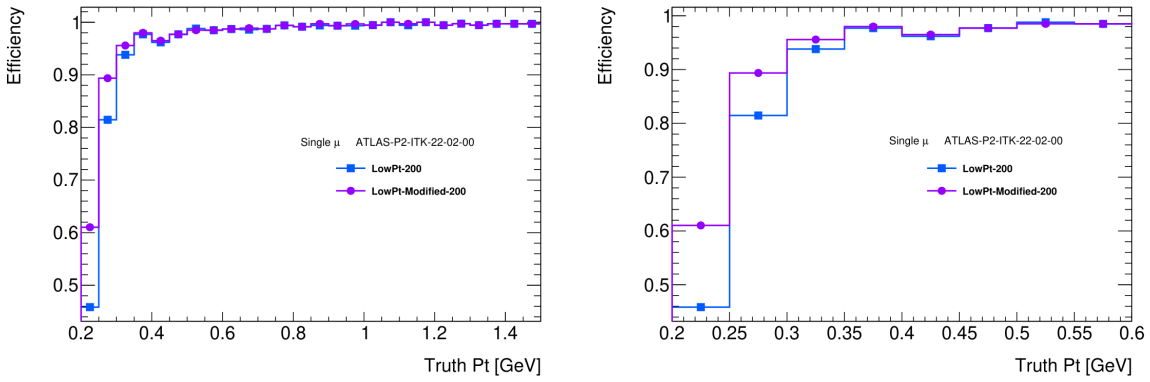


Figure 31: A comparison of the tracking efficiency for a single muon sample reconstructed using the Low-p_T-200 and Low-p_T-Modified-200 modes.

3.6 Seed Momentum Resolution

As discussed previously, drops in efficiency for a single muon sample can also highlight problems with the track reconstruction algorithms. Both Figs. 29 and 31 show a decrease in the seed efficiency for the bins approaching the minimum seed p_T threshold. A potential explanation for this drop is that during seeding, seeds are rejected on initial estimates of the impact parameters and track p_T. These estimates are made assuming perfect helical trajectories in a uniform magnetic field. However, these estimates are rough as they are constructed only from three space-points. This means that the resolution on seed p_T can be quite large, and if the p_T of charged particles is close to the minimum seed p_T, there is a high likelihood that the true track p_T is underestimated and the seed is falsely rejected. This effect also occurs after candidate finding, when the track p_T cut is applied, but the effect is much smaller as the track-p_T resolution is far better than the initial seed-p_T estimate.

To further demonstrate this, Fig. 32 shows the tracking efficiency off the simple low-p_T tracking mode and default ITk reconstruction mode for a $t\bar{t}$ sample with $\mu = 200$ sample. The default mode has a minimum seed p_T of 900 MeV, whilst this low-p_T mode has a cut of 500 MeV. For the default ITk tracking cuts, as the charged particle p_T decreases to the 900 MeV threshold, there is a significant drop in the tracking efficiency. However, for the

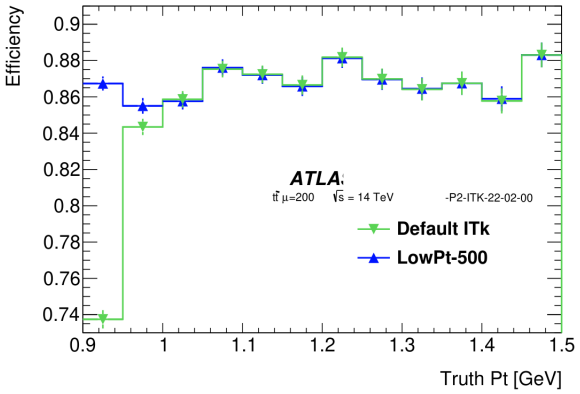


Figure 32: A comparison of the tracking efficiency for a $t\bar{t}$ sample with $\mu = 200$ reconstructed using the Default ITk and Low-p_T-500 modes.

low-p_T mode, this decrease is not present above this threshold. The difference in the two reconstruction modes is the seed and track p_T cut. No seed momentum resolution effect is visible for the low-p_T mode as the seed p_T cut is well below the default ITk cut. This increase in efficiency also comes with a caveat. The tracking cuts are designed to reduce and maintain reasonable processing times. Relaxing the tracking cuts, particularly for high pileup samples, leads to significantly longer processing times.

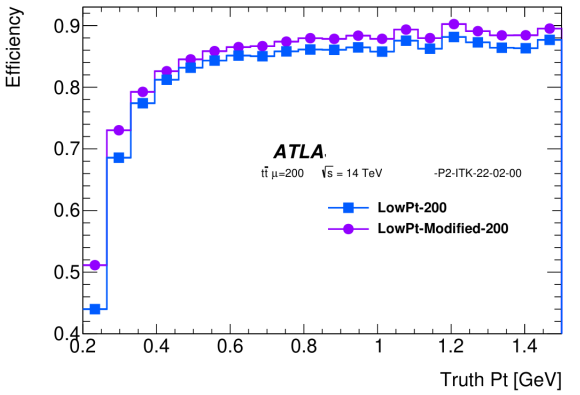


Figure 33: A comparison of the tracking efficiency for a $t\bar{t}$ $\mu = 200$ sample reconstructed using the Low-p_T-200 and Low-p_T-Modified-200 modes.

Further efficiency improvements are seen with the looser hit requirements in the Low-pT-Modified-X configuration. Figure 33 shows the tracking efficiency for the reconstruction of a $\mu = 200$ $t\bar{t}$ sample using the simple and modified low-p_T tracking modes. The modified low-p_T mode has a higher track reconstruction efficiency across all bins. The cost of this increase in efficiency is shown in Figure 34. The figure shows an increase of approximately 10% in CPU time for the configuration with modified cluster and hit requirements. In practice, such an increase would be a vast increase in the total processing time over the scope of the number of events expected at the HL-LHC. This illustrates the careful decisions which need to be made when selecting the track parameters. An increase

of 10% could, however, be feasible if the track reconstruction is performed on a smaller subset of events or events with a much lower pileup.

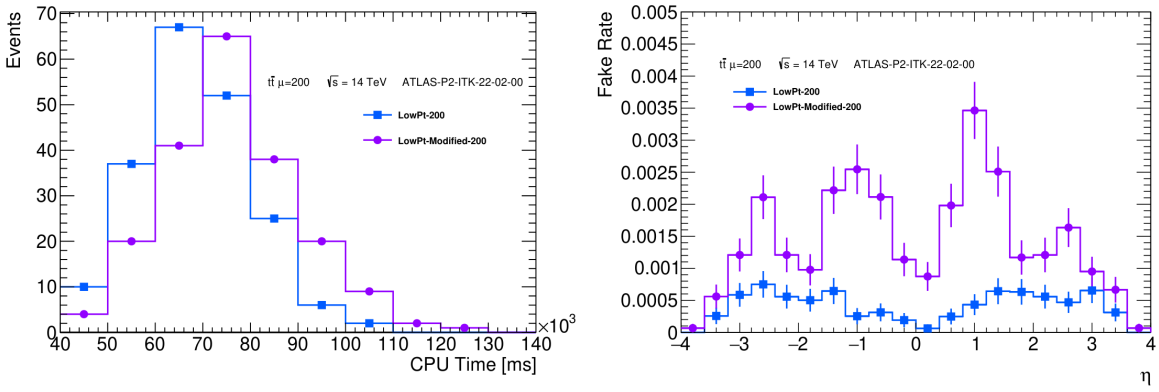


Figure 34: A comparison of the processing time per event (left) and the fake rate (right) for a $t\bar{t}$ with $\mu = 200$ sample reconstructed using the Low-p_T-200 and Low-p_T-Modified-200 modes.

Figure 34 depicts a secondary caveat to relaxing the tracking cuts: the fake rate. The figure shows a significant increase in the fake rate with looser tracking cuts. However, this is a lesser issue as the ITk track reconstruction algorithms perform extremely well in identifying fake tracks. The fake rate is still very low despite the looser tracking cuts.

3.7 Second Pass Reconstruction

As discussed, there are a number of challenges to low-p_T track reconstruction. The author implemented a low-p_T track reconstruction mode for the ITk as a second pass reconstruction. Low-p_T tracking as a second pass was first introduced in [71] for the Inner Detector. A second pass low-p_T reconstruction refers to running a low-p_T configuration on a reduced space-point set after initially running the default ITk track reconstruction. Performance-wise, in high-pile up, running with a single or second pass configuration is not expected to make a massive difference with the same tracking configuration. The benefit, however, of running with a second pass is increased flexibility. Tracking cuts can be modified solely for the low-p_T mode without affecting the default mode. This means the effect on CPU time when running with relaxed tracking cuts is expected to be less significant with a second pass. The implementation of the secondary pass reconstruction included an upper bound on the seed p_T in order to focus the reconstruction on low-p_T tracks and limit the reconstruction time.

3.7.1 Low-μ Low-p_T Reconstruction Mode

The first implemented low-p_T track reconstruction mode is referred to as a low-μ low-p_T tracking mode. The tracking cuts for the low-μ low-p_T reconstruction mode are shown in Tab. 5. As the name suggests, this mode is intended to be used only for low-μ datasets, for example, minimum bias low-luminosity measurements [75]. Here, the effects

of pileup are significantly reduced, and the efficiency can be maximised without extensive computational costs. The low-μ mode is also implemented as a single pass, as the pileup effects are negligible.

Low-p_T Low-μ	Value	Default Value
Track min-p_T	[200, 200, 200] MeV	[900, 400, 400] MeV
Track min clusters	[6, 5, 4]	[9, 8, 7] MeV
Track min silicon hits not shared	[6, 5, 4]	[7, 6, 5]
Seed min-p_T	200 MeV	900 MeV
Remaining	Default ITk	

Table 5: The low-p_T low-μ tracking configuration. The tabulated values in a list refer to η-dependent cuts, i.e. $[a, b, c]$ applies the cut value a for $|\eta| \leq 1$, b for $1 < |\eta| < 2.6$ and c for $|\eta| \geq 2.6$.

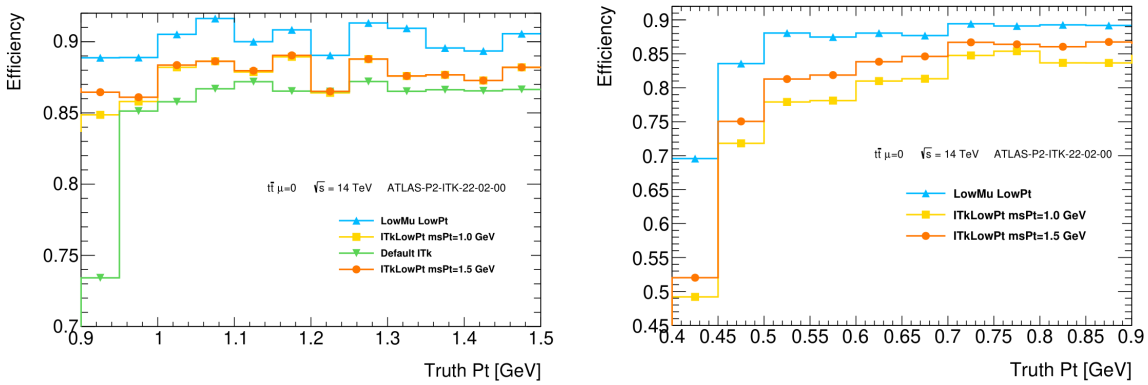


Figure 35: A comparison of the tracking efficiency for a $t\bar{t}$ $\mu = 0$ sample reconstructed using the low-p_T low-μ, general purpose low-p_T mode with max seed-p_T = 1.0, 1.5 GeV and the default ITk.

Figure 35 compares the tracking efficiency of the low-μ low-p_T reconstruction mode to that of the default ITk tracking, the general purpose low-p_T mode (described in Sec. 3.7.2), and the low-p_T mode with max-seed-p_T cuts of 1.0 or 1.5 GeV. The reconstructed sample is a $t\bar{t}$ $\mu = 0$ sample. Whilst the reconstruction is simulated with zero pileup, the sample itself has significant hadronic activity and should provide a reasonable benchmark for the performance expected for a minimum bias or pion low-μ sample. The figure shows that the efficiency is highest for the low-μ mode, largely due to the reduced silicon hit requirements. The trade-off is an increase in CPU time, as shown in Fig. 36, but this increase is not significant when compared to the processing time of the equivalent $\mu = 200$ sample. Likewise, the fake rate for this sample is negligible.

3.7.2 General Purpose Low-p_T Reconstruction Mode

This author also implemented a general purpose low-p_T reconstruction mode targeted at analyses that would benefit from low-p_T tracking. The tracking cuts for the general

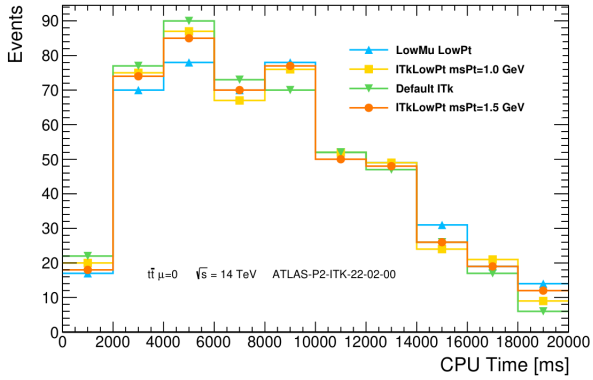


Figure 36: A comparison of the processing time per event for a $t\bar{t}$ $\mu = 0$ sample reconstructed using the low-p_T low-μ mode, the general purpose low-p_T mode with maximum seed-p_T = 1.0 or 1.5 GeV and the default ITk tracking.

purpose low-p_T reconstruction second pass are shown in Table 6. This mode is intended to run on high-pileup data, balancing computation time and efficiency. The minimum track p_T threshold for this mode is lowered to 400 MeV, and the remaining ITk tracking cuts are kept as default. From the studies shown in this thesis, this was the lowest feasible p_T where the tracking efficiency would not significantly degrade.

General purpose low-p_T	Value	Default Value
Track min-p_T	[400, 400, 400] MeV	[900, 400, 400] MeV
Seed min-p_T	400 MeV	900 MeV
Remaining	Default ITk	

Table 6: The general purpose low-p_T mode tracking cuts. The tabulated values in a list refer to η-dependent cuts, i.e. $[a, b, c]$ applies the cut value a for $|\eta| \leq 1$, b for $1 < |\eta| < 2.6$ and c for $|\eta| \geq 2.6$.

Figure 37 compares the tracking efficiency of the low-p_T reconstruction mode to that of the default ITk, the general purpose low-p_T mode with maximum seed-p_T cuts of 1.0 GeV (applied during the secondary pass) and the same low-p_T configuration but run as a single pass. The efficiency for the general purpose mode is the highest, with very marginal differences in the overall efficiency when running as a single pass. However, the second-pass reconstruction has the longest computation time, as shown in Fig. 38. The seed-p_T cut reduces the computation time compared to the other modes. The fake rate is minimal for the different reconstruction modes, with no distinct mode dominating.

The difference in computation time between the reconstruction modes may or may not be significant depending on the analysis. If the processing time is critical, for example, for a very large dataset, then the general purpose low-p_T mode with maximum seed p_T of 1 GeV is the most applicable. By contrast, if efficiency is the critical parameter, then the general purpose low-p_T mode should be applied, with the scope to further increase the efficiency by relaxing some of the tracking cuts during the second pass.

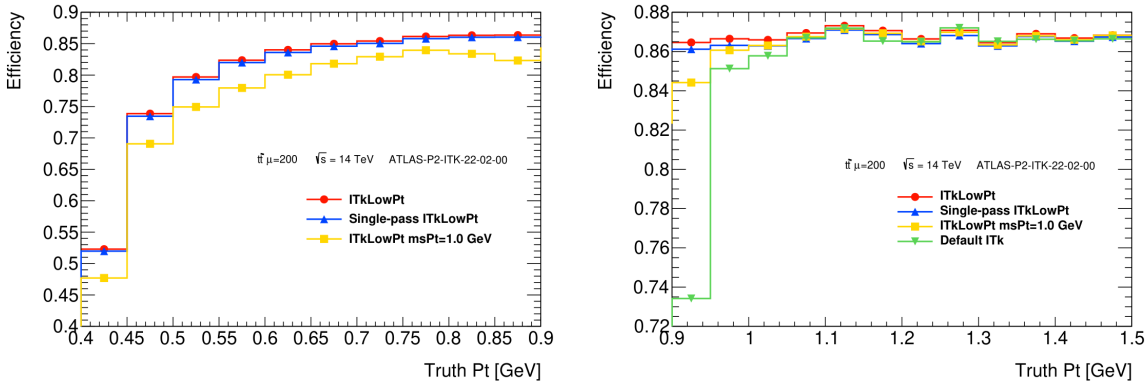


Figure 37: A comparison of the tracking efficiency for a $t\bar{t}$ with $\mu = 200$ sample reconstructed using the general purpose low-p_T mode, the general purpose low-p_T mode with maximum seed-$p_T = 1.0$ GeV, the general purpose low-p_T mode run as a single pass, and the default ITk tracking.

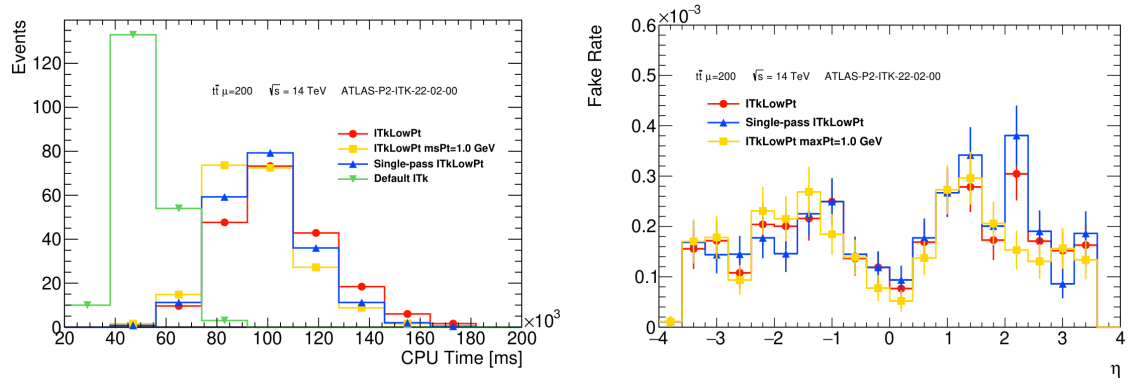


Figure 38: A comparison of the processing time per event (left) and the fake rate (right) for a $t\bar{t}$ $\mu = 200$ sample reconstructed using the general purpose low-p_T mode, the general purpose low-p_T mode with maximum seed $p_T = 1.0$ GeV, the general purpose low-p_T mode run as a single pass, and the default ITk tracking.

3.8 Sensitivity to Exclusive $W^{\pm}W^{\mp}$ Production in Photon Scattering

One of the primary use cases for low-p_T tracking is photon physics, in particular, WW pair production from the interaction of photons. Low-p_T tracking can improve the background rejection for photon physics by better reconstructing the underlying event. $\gamma\gamma \to WW$ pair production is a SM VBS process that occurs through trilinear and quartic gauge-boson interactions, as shown in Figure 39. The measurement of photon-induced WW pair production, therefore, corresponds to a direct test of the $SU(2) \times U(1)$ electroweak gauge structure. Moreover, the process can be used to test anomalous gauge couplings in the context of an effective field theory [76]. A $\gamma\gamma \to WW$ measurement has been published by ATLAS, showing agreement with Standard Model predictions and rejecting the background-only hypothesis at a significance of 8.4σ [76]. The low-p_T studies presented in

this chapter contributed to the analysis of a prospective measurement of photon-induced pair production at the HL-LHC [77] and referred to in this section as the HL-LHC analysis. The studies contributed by providing tracking efficiencies and impact parameter resolutions at low-p_T in high-pileup events to test the application of low-p_T tracking to this analysis. This section briefly discusses the aspects of the physics relevant to low-p_T, with the reader referred to [77] for more detail. .

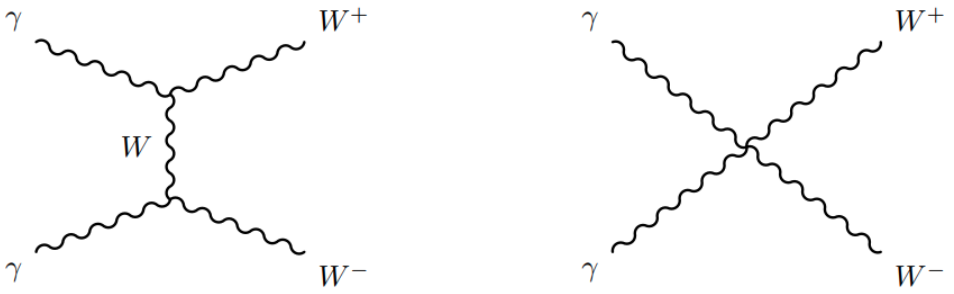


Figure 39: Feynman diagrams depicting photon-induced WW production.

At the ATLAS experiment, the WW final state proceeds via $pp \to p^* W^{\pm} W^{\mp} p^*$ where p^* denotes a final-state proton which either stays intact or fragments after emitting a photon. This leads to three production modes: elastic, single-dissociative and double-dissociative. In all three cases, the final state is typically composed of one jet each in the forward and the backward regions of the detector along the z-axis. The HL-LHC analysis is performed in the fully leptonic final state, analogously to the ATLAS measurement [76]. Whilst it has a smaller branching ratio compared to the semi-leptonic and fully hadronic decay channels, the signal-to-background ratio is better for fully leptonic decays.

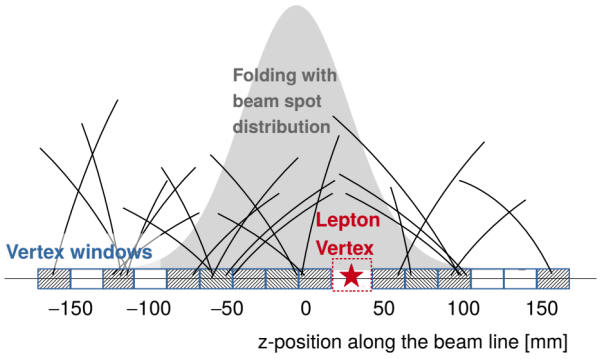


Figure 40: A graphical depiction of the track-veto window. Tracks within the lepton vertex window are vetoed (in this depiction, no such tracks are seen). Figure taken from [77].

The electroweak nature of the $\gamma\gamma \to WW$ process, and in particular the lack of colour exchange, means that the event signature of a typical event has highly suppressed activity in the central region of the detector. This can be used to reduce major background contributions coming from quark and gluon-induced WW production, which is typically produced with significant hadronic activity from the underlying event. The analysis

achieves this by applying a selection criterion that requires events to have no additional charged particle tracks in the vicinity of the interaction vertex. As shown in Figure 40, the analysis rejects tracks within a symmetric z-window around the dilepton vertex, referred to as the track veto window. The criterion requires the number of charged particle tracks to be equal to zero, excluding the two charged leptons in the window. The window size is an optimisable parameter of the analysis.

The HL-LHC tracking requires a minimum track p_T of 900 MeV to pass the track selection criteria. The track veto window is designed to reject soft hadronic activity associated with the underlying event of the dominant QCD-induced backgrounds. However, with a high track p_T cut, many of the soft tracks may not be reconstructed. In practice, this would lead to a larger number of background events falsely identified as signal. For the Run 2 analysis, this was not as problematic because the track p_T threshold is far lower. For this HL-LHC analysis, low-p_T tracking can mitigate these problems and improve the background rejection efficiency. This work contributed to the prospective study of [77] by providing tracking efficiencies for $\bar{t}t$ production at $\mu = 200$. The track efficiencies were used as inputs to a truth-level study of the QCD backgrounds at the HL-LHC. The low-p_T tracking is implemented after an initial subset of events is selected in the analysis.

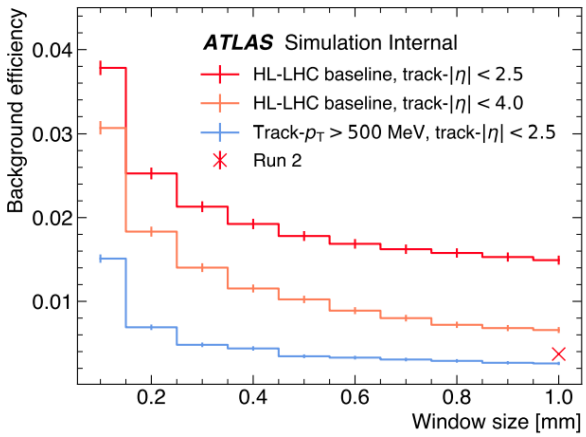


Figure 41: The efficiency for background events to pass the track veto in HL-LHC collisions as a function of track-veto window size. The low-p_T reconstruction mode with a minimum p_T of 0.5 GeV is compared to the default ITk tracking with the current tracking acceptance ($|\eta| < 2.5$) and the expected acceptance of the ITk ($|\eta| < 4$). Also shown is the Run-2 efficiency and track-veto window size. The figure is taken from [77].

The results of the truth level study are shown in Fig. 41. The figure shows the background efficiency as a function of the track veto window size. The background efficiency is defined as the number of background events passing the track veto divided by the total number of background events. The figure shows that the background efficiency is reduced by a factor of 2-5 when requiring a minimum p_T of 500 MeV, where the factor depends on the veto window size. The Run 2 window size and background efficiency are shown in the figure for comparison. The background rejection efficiency is significantly worse without low-p_T tracking.

4 Vector Boson Scattering at the LHC

Gauge bosons are produced at the LHC as intermediate particles and are observed through their constituent decay products. Processes involving gauge boson self-interactions like VBS can occur when two incoming partons radiate gauge bosons, which subsequently interact, as shown in Fig. 43. These processes are incredibly rare as they typically involve multiple orders of the electroweak coupling constant. As a result of the low cross-sections, VBS and similar processes were mostly inaccessible experimentally until the development of the LHC. Figure 42 shows the Standard Model cross-section measurements for various processes at the LHC. The figure shows that the cross-sections for measurements with multiple gauge bosons in final states, shown towards the right of the figure, are by far the smallest. VBS falls under the category of $VVjj$, which has the lowest average cross-section. Even with the massive luminosity of the LHC, VBS measurements, particularly of polarisation, are still relatively restricted and require a higher luminosity experiment like the HL-LHC for precision measurements.

The low-cross section is challenging, but fortunately, VBS events have a unique event topology due to the electroweak nature of the interaction. Due to this unique topology, it is experimentally possible to isolate a VBS phase space where VBS events are the predominant production mode. There are still, however, several challenges with the experimental study of VBS. For example, there are a number of different production mechanisms leading to gauge bosons in the final state. These can be both with and without VBS contributions, and many of these mechanisms cannot be modelled separately as they would violate gauge invariance. This chapter focuses on the general modelling aspects of VBS and illustrate a typical VBS event topology at the LHC. The discussion presented here is in the general context of VBS, but with the notion that the modelling aspect will thereafter be applied to polarised same-electric charge $W^{\pm}W^{\pm}$ VBS.

Polarisation introduces further complexities to VBS measurements. Polarised VBS has an even smaller cross-section as it represents a subset of the VBS process. Polarised measurements also require MC predictions of polarised events. As polarisation is not a Lorentz-invariant quantity, this can introduce a number of difficulties at the generator level. The development of polarised predictions, particularly at NLO, is still an active area of research. There are several MC generators capable of producing LO polarised predictions, though there are still challenges. Defining polarisation is not always possible. Polarisation is undefinable for a number of diagrams that contribute to electroweak production. In order to produce polarised predictions, suitable approximations must be made. This chapter demonstrates the necessity of polarised samples for experimental VBS measurements, the method of producing polarised samples, and the theoretical aspects relevant to the modelling.

4.1 General Overview

At the LHC, gauge bosons can scatter by any of the possible processes shown in Fig. 44, resulting in two outgoing gauge bosons. Following the scattering, the two outgoing partons hadronise and are detected as two jets. The final state of a VBS event contains two gauge bosons and two jets, denoted by $VVjj$. Due to their short lifetimes, the gauge

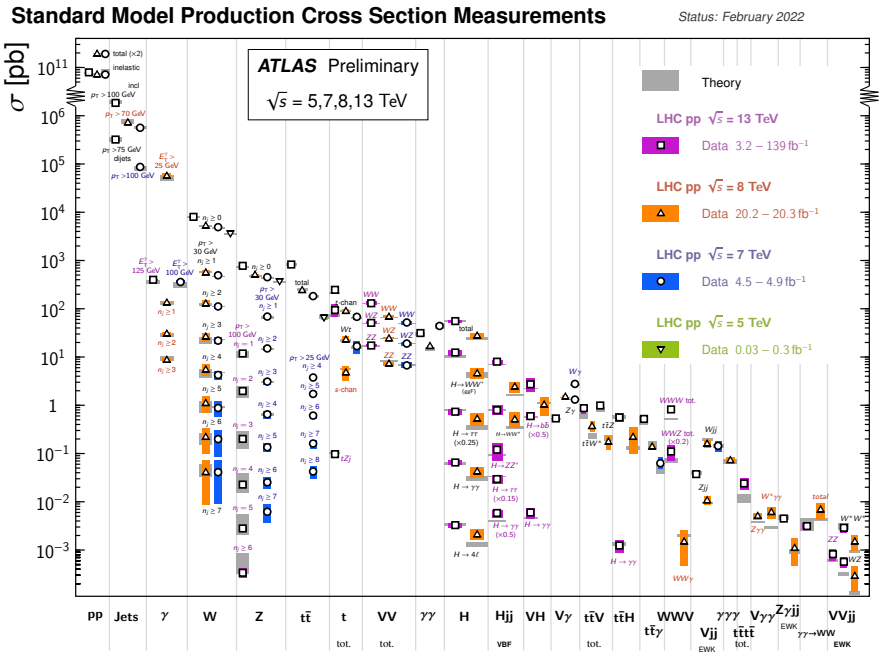


Figure 42: The measured cross-sections for Standard Model processes. Towards the right of the figure lie the processes with multiple gauge bosons in the final state. VBS is denoted by $VVjj$ and has the lowest average cross-section. Figure taken from [78].

bosons are studied through their respective decay channels. For events with two gauge bosons, there are three possible decay channels by which events can be sub-categorised: the leptonic, semi-leptonic and hadronic decay channels. The measured final state will consist of two jets and two pairs of fermions.

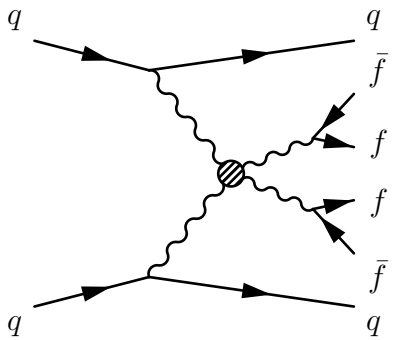


Figure 43: A Feynman diagram illustrating a VBS scattering process at the LHC, where the blob characterises each of the possible VBS diagrams shown in Fig. 44.

4.1.1 $VVjj$ Production

There are multiple production mechanisms that can produce a final state of $VVjj$ at the LHC. At leading order, these mechanisms can be sub-divided based on the order of the coupling constants. There are two categories: electroweak and strong production. At

tree-level, electroweak production includes all purely electroweak diagrams with coupling constants of $\mathcal{O}(\alpha_{\text{EM}}^6)$. Analogously, strong production includes all diagrams with QCD vertices, with coupling constants of $\mathcal{O}(\alpha_S^2\alpha_{\text{EM}}^4)$. Classifying the samples by order of the coupling constants is gauge invariant such that the samples can be modelled separately. At LO, the total cross-section for $VVjj$ is given by,

$$\begin{aligned}\sigma_{VVjj} &= |\mathcal{M}^{EW} + \mathcal{M}^{QCD}|^2 \\ &= |\mathcal{M}^{EW}|^2 + |\mathcal{M}^{QCD}|^2 + 2Re(\mathcal{M}^{EW}\mathcal{M}^{*QCD}) \\ &= \sigma_{VVjj}^{EW} + \sigma_{VVjj}^{QCD} + \sigma_{VVjj}^{INT},\end{aligned} \tag{82}$$

where σ_{VVjj} receives contributions coming from the electroweak and strong production modes and an additional contribution coming from their interference.

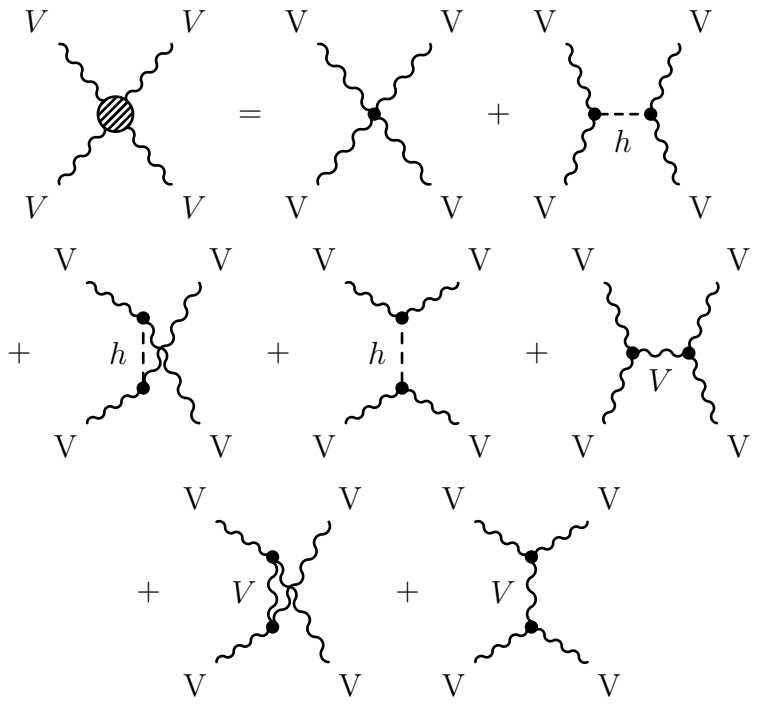


Figure 44: The tree-level Feynman diagrams corresponding to VBS.

4.1.2 Electroweak-induced $VVjj$ Production

Figs. 45 and 46 show some of the tree-level Feymann diagrams contributing to EW $VVjj$ production. The electroweak production mode can be further sub-categorised into three production modes: VBS production, non-VBS production, and triboson production. VBS production consists of all of the diagrams, which include the VBS sub-diagrams shown in Fig. 44. These diagrams include the quartic gauge boson self-interactions and the trilinear interactions through the s- and t-channel exchanges with the Higgs and Z/γ boson.

The electroweak non-VBS production mode consists of all diagrams with a final state of $VVjj$ without the VBS vertex, for example, as shown in Fig. 45. The intermediate

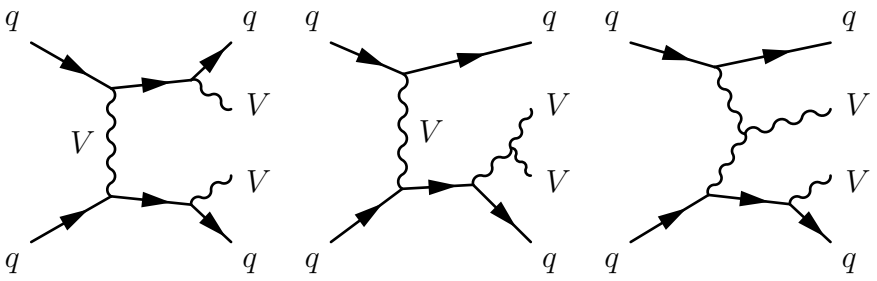


Figure 45: Feynmans diagrams illustrating EW-induced $VVjj$ production, with resonant gauge-bosons and without the VBS vertex.

gauge bosons are inferred through their decay products, and it is possible, as shown in Fig. 46, for these products to result from processes other than gauge boson decay. This is referred to as non-resonant production. The electroweak non-VBS production mode can, therefore, be further subdivided into two categories: resonant and non-resonant production.

The electroweak VBS and non-VBS production modes must be modelled together to preserve gauge invariance. The non-VBS contribution is referred to as an irreducible background. Finally, triboson production is the process in which three gauge bosons are produced. The $VVjj$ final state can occur in this process if one of the gauge bosons decays hadronically. The triboson production mode can be modelled independently while maintaining gauge invariance, and it is typically highly suppressed in a VBS phase space.

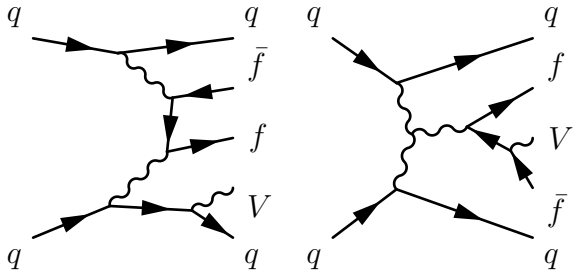


Figure 46: Feynman diagrams illustrating EW-induced $VVjj$ production, with a single resonant gauge boson.

4.1.3 QCD-induced $VVjj$ Production

Figures 47 demonstrates some of the Feynman diagrams for QCD-induced $VVjj$ production. There are two types of production modes. The first occurs when incoming quarks interact through a colour exchange with a gluon. The subsequent quarks radiate gauge bosons. These interactions can also include the gauge-boson self-interaction vertex but not the VBS vertex. The second type of interaction involves gluons in the initial state and does not include self-interaction vertices.

The QCD-induced $VVjj$ production mode typically has a much larger cross-section compared to electroweak production due to the colour factors, the larger number of diagrams, and the addition of gluons in the final state. Whilst the cross-section of the QCD

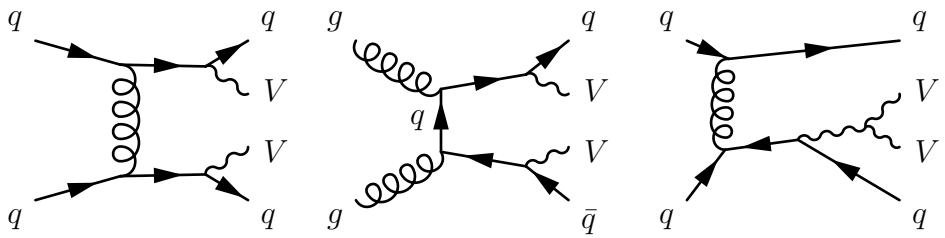


Figure 47: Feynman diagrams illustrating QCD-induced $VVjj$ production.

production mode is generally dominant, the processes show significant kinematic differences due to the colour exchange during the interaction. Thus, a VBS phase space can be defined where the QCD production process is suppressed.

4.1.4 Interfering QCD-EW $VVjj$ Production

Interference between QCD and electroweak VVjj production occurs when diagrams have identical final state quarks. The electroweak process is a colour singlet process, so the interference only occurs between diagrams where the quark propagation in the two diagrams is different. This means that the interference is non-zero only between different channels, for example, between the s and t-channel for contributions with the same external quarks. In this thesis, the QCD-EW interference (INT) $VVjj$ production is referred to as interference-induced production.

4.2 VBS Topology

VBS interactions have a unique event topology owing to the neutral colour exchange of the interaction. The unique event topology defines a kinematic phase space where VBS events are the predominant production mode. This means that often, for VBS measurements, some of the major backgrounds entering the VBS signal region are other VBS processes. VBS events are typically observed with a large dijet invariant mass m_{jj} and rapidity difference, Δy_{jj}. This is due to the colourless t-channel exchange that produces high-energy jets in the forward and backward regions of the detector. A simulated VBS event in the ATLAS detector is shown in Fig. 48.

To illustrate the physics behind this unique topology, first consider the scattering amplitude for a VBS process. During scattering, the gauge bosons are produced as virtual particles radiated by two incoming quarks. The scattering amplitude will, therefore, contain two propagator terms for the intermediate bosons. The amplitude for the process is proportional to the propagator terms, denoted by [79, 80]

$$|M_{fi}|^2 \propto \frac{p_1 \cdot p_2 \cdot p_3 \cdot p_4}{(q_1^2 - m_V^2)^2 (q_2^2 - m_V^2)^2} \tag{83}$$

where $p_{[1,2]}$ are the momenta of the incoming quarks, $p_{[3,4]}$ are the momenta of the outgoing quarks, $q_{[1,2]}$ are the momenta carried by the gauge bosons. To maximise this expression, $q_{1,2}$ must be minimised while $p_{3,4}$ maximised. This predicts that there will be a very small

deflection of the incoming and outgoing quarks, leading to a large Δy_{jj} and that the jets will have high momentum, resulting in a high m_{jj}. Furthermore, due to the absence of any colour exchange between the two incoming quarks, there is drastically reduced hadronic activity in the central region of the detector for VBS processes relative to QCD-induced $VVjj$. As such, VBS events are often described to have a very clean event signature, a characteristic common to purely electroweak processes.

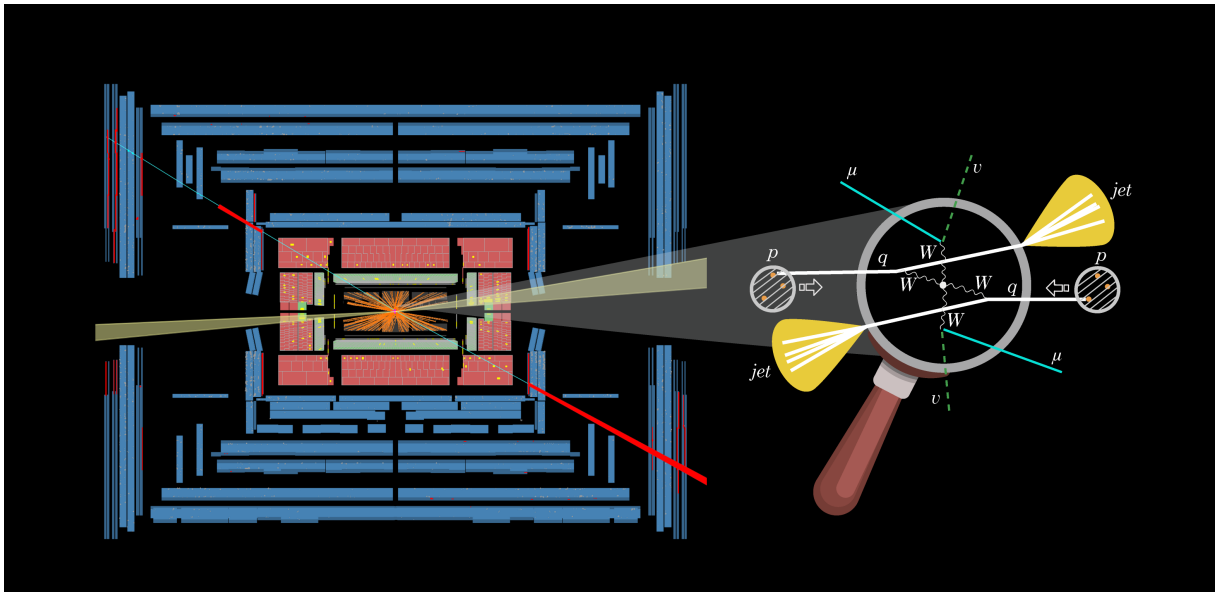


Figure 48: A projection of a VBS event in the ATLAS detector, showing jets in the forward and backward regions of the detector, with two muons detected between the tagging jets. Figure taken from [81].

The subject of this thesis is the experimental study of purely leptonic VBS. The branching ratios for hadronic decays of the W and Z bosons are approximately 70% [82], which means that the cross-section for the fully-leptonic decay channel is the smallest. However, semi-leptonic and hadronic decay channels consist of multi-jet final states. The number of final state processes with multi-jet backgrounds is very high, and so these channels are often less sensitive to VBS. Furthermore, the leptonic decay channel has a much cleaner event signature due to the reduced hadronic activity in the central region of the detector. There are no additional jets associated with the decay of the final state gauge bosons. As such, even with the smaller cross-section, the leptonic decay channel is experimentally preferable.

Figure 49 illustrates a typical VBS event in the leptonic decay channel, showing the characteristic presence of two high-energy jets in the forward and backward region of the detector, with the two leptons from the decay lying between the tagging jets. The gauge bosons in a VBS signal region carry a high transverse momentum, which is likewise propagated to the constituent decay products. At the LHC, VBS events involving the scattering of two W bosons in the leptonic decay channel are typically observed with a large missing transverse energy coming from the two undetected neutrinos. This can also be true for some VBS events involving Z bosons, where the Z bosons decay into a neutrino

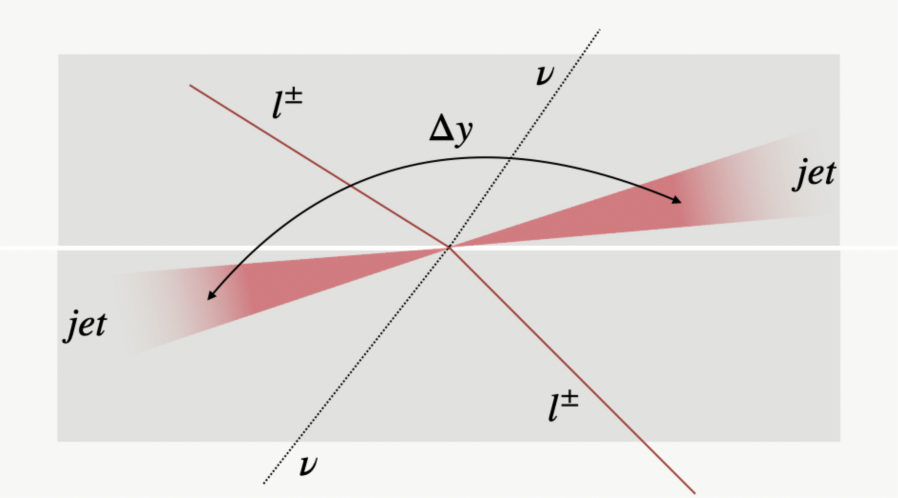


Figure 49: A graphical depiction of a VBS topology with jets in the forward and backward regions of the detector, with the two leptons associated with gauge boson decays, lying between the tagging jets. Figure taken from [80].

and antineutrino. The large E_T^{miss}, therefore, is often used as an additional criterion to define a leptonic VBS phase space.

4.3 Polarisation

Polarised gauge bosons are produced at the LHC as intermediate particles and studied through their decay products. For leptonic decays, the effects of gauge boson polarisation manifest in angular distributions of the leptons. The polarisation fractions for polarised gauge boson production can be extracted in the absence of detector cuts from the lepton angular distributions [83, 84, 85]. In practice, however, this is not experimentally feasible, and polarised samples are necessary to model and extract polarisation from data.

Polarised samples can be generated by replacing the gauge boson propagator with a polarised gauge boson propagator. The cross-section of the unpolarised process is then given by the sum of the polarised cross-section plus some non-zero interference between the polarisation amplitudes. The interference effects can be modelled by comparing the unpolarised and polarised samples. However, for processes such as VBS, producing polarised samples is more complicated. Defining polarisation is non-trivial for interactions with multiple intermediate gauge bosons due to contributions coming from non-resonant gauge bosons. The remainder of this chapter outlines the generation of polarised templates.

4.3.1 Polarisation Measurement

As discussed above, polarisation information can be extracted when the full angular distribution of the constituent charged lepton can be measured. Consider the production of a single polarised W boson, which decays leptonically. The polarised matrix element of

the decay is given by,

$$\mathcal{M}^{decay}_{\nu,\lambda} = \frac{ig}{2\sqrt{2}}\bar{u}(k)\epsilon^{*\mu}_{\lambda}\gamma_{\mu}(1-\gamma^{5})\nu(k') \tag{84}$$

where $\bar{u}(k)$ and $\nu(k')$ are Dirac spinors, and k, k' are the outgoing momenta of the two leptons in a given reference frame. In the W boson rest frame, the angle θ is defined as the angle of decay between the direction of motion of the W boson in the laboratory frame and the charged lepton in the W rest frame. Consider the differential cross-section in terms of θ after integrating out the azimuthal angular dependence:

$$\frac{d\sigma(\theta)}{d\cos\theta} = \int d\phi \frac{d\sigma(\theta,\phi)}{d\cos\theta d\phi} \tag{85}$$

In this case, the W boson decay can be considered in one plane and the differential cross-section is given by,

$$\frac{1}{\sigma}\frac{d\sigma}{d\cos\theta} = \frac{3}{8}(1\mp\cos\theta)^2 f_L + \frac{3}{8}(1\pm\cos\theta)^2 f_R + \frac{3}{4}\sin^2\theta f_0 \tag{86}$$

where $f_{L/R/0}$ are the polarisation fractions of the left/right/longitudinally polarised W boson, and the sign corresponds to a $W^{\pm}$. The normalisation is chosen here such that the fractions obey,

$$\frac{1}{\sigma}\int_{-1}^{1} d\cos\theta \frac{d\sigma}{d\cos\theta} = f_L + f_R + f_0 = 1 \tag{87}$$

The differential cross-section is composed of the contributions of each of the polarisations fractions. From the expression, as an example, it can be seen that the contribution to the differential cross section is zero from longitudinally polarised bosons when $\theta = 0, \pi$. This is where the charged lepton decays along the W boson direction of motion, which would violate spin conservation. Experimentally, polarisation information can be extracted from the differential cross section by considering the expectation value of the $\cos\theta$ distribution,

$$\langle\cos\theta\rangle = \frac{1}{\sigma}\int_{-1}^{1}\cos\theta\, d\cos\theta\frac{d\sigma}{d\cos\theta} = \mp\frac{1}{2}(f_L - f_R) = \pm(\frac{1}{2} - f_L - f_0). \tag{88}$$

The polarisation fractions can directly be calculated by measuring the $\cos\theta$ and $\cos^2\theta$ distributions, given by [84, 85]:

$$\begin{aligned} f_0 &= 2 - 5\left\langle\cos^2\theta\right\rangle \\ f_L &= -\frac{1}{2} \mp \langle\cos\theta\rangle + \frac{5}{2}\left\langle\cos^2\theta\right\rangle \\ f_R &= -\frac{1}{2} \pm \langle\cos\theta\rangle + \frac{5}{2}\left\langle\cos^2\theta\right\rangle \end{aligned} \tag{89}$$

Equivalently, the polarisation fractions could be extracted from the differential distribution by projection onto the Legendre polynomials [83]. The differential cross-section, now

in terms of the total solid angle, is given by:

$$\begin{aligned}\frac{1}{\sigma}\frac{d\sigma}{d\cos\theta d\phi} =& \frac{3}{16\pi}\Big[(1+\cos^2\theta) + \frac{A_0}{2}(1-3\cos^2\theta) + A_1\sin 2\theta\cos\phi \\ &+ A_2\sin^2\theta\cos 2\phi + A_3\sin\theta\cos\theta + A_4\cos\theta \\ &+ A_5\sin\theta\sin\phi + A_6\sin 2\theta\sin\phi + A_7\sin^2\theta\sin 2\phi\Big]\end{aligned} \tag{90}$$

where A_i are process-dependent angular coefficients. When integrated over the full phase space in $\phi \in [0, 2\pi)$, this recovers the differential cross-section given by Eq. 86. After introducing the ϕ dependence, the differential cross-section is no longer given in terms of the individual contributions coming from each polarisation fraction. This is because the ϕ dependence itself introduces interference effects between the individual polarisation fractions, which are only zero when integrated over the full phase space of ϕ.

Crucially, the above methods for extracting the polarisation of a gauge boson from its decay products require a full inference of the angular kinematics. Practically, this would require a detector capable of providing acceptance across the entire angular phase-space, which is not possible. Furthermore, the expressions are only valid in the absence of cuts. Cuts applied to the leptonic kinematic variables will directly affect the shape of the differential cross-section, and the expression in 86 will no longer hold true. In an experimental setting, polarised samples must be generated to measure the polarised cross-section from unpolarised data.

4.3.2 Polarised Cross-Sections

To illustrate the production of polarised samples, this subsection demonstrates the relationship between polarised and unpolarised cross-sections. To begin, consider the general expression for the unpolarised matrix element $\mathcal{M}$, in the unitary gauge for a single intermediate gauge boson produced at the LHC,

$$\mathcal{M} = \mathcal{M}_\mu^{prod}\frac{1}{k^2 - m_V^2 + i\Gamma m_V}\left(-g^{\mu\nu} + \frac{k^\mu k^\nu}{m_V^2}\right)\mathcal{M}_\nu^{decay}, \tag{91}$$

where $\mathcal{M}^{prod}$ and $\mathcal{M}_\nu^{decay}$ are the individual matrix elements for production and decay. The numerator of the propagator can be expressed in terms of the gauge boson polarisation vectors, using the completeness relation given by [86],

$$\left(-g^{\mu\nu} + \frac{k^\mu k^\nu}{m_V^2}\right) = \sum_{\lambda=1}^{4}\epsilon_\lambda^\mu(k)\epsilon_\lambda^{*\nu}(k) \tag{92}$$

where the polarisation vectors are defined as [83],

$$\begin{aligned}\epsilon_L^\mu &= \frac{1}{\sqrt{Q^2}}(k, 0, 0, E) \\ \epsilon_\pm^\mu &= \frac{1}{\sqrt{2}}(0, \mp 1, -i, 0) \\ \epsilon_A^\mu &= \sqrt{\frac{Q^2 - m_V^2}{Q^2 m_V^2}}(E, 0, 0, k),\end{aligned} \tag{93}$$

in the frame in which the gauge boson propagates along the z-axis with momentum k, energy E and invariant mass $\sqrt{Q^2} = \sqrt{E^2 - k^2}$. The polarisation vectors are defined with the following properties,

$$k_\mu \cdot \epsilon^\mu_\lambda(k) = 0, \tag{94}$$

and,

$$\epsilon^\mu_\lambda(k) \cdot \epsilon^*_{\mu,\lambda'}(k) = -\delta_{\lambda\lambda'}. \tag{95}$$

There are two transverse polarisation configurations $\epsilon_\pm$, one longitudinal configuration ϵ_L from the additional degree of freedom gained from absorbing the Goldstone bosons, and a non-physical auxiliary polarisation mode ϵ_A that occurs only for off-shell gauge bosons. It should also be noted that the polarisation vectors are not Lorentz-invariant, and as a consequence, neither are polarised cross-sections. Substituting this expression into the expression for the unpolarised matrix element gives

$$\mathcal{M} = \frac{1}{k^2 - m_V^2 + i\Gamma m_V} \sum_\lambda^4 \mathcal{M}^{prod}_\mu \epsilon^{*\mu}_\lambda(k) \epsilon^\nu_\lambda(k) \mathcal{M}^{decay}_\nu. \tag{96}$$

The unpolarised matrix element can thus be expressed in terms of the polarisation vectors. Implicitly, the polarisation vectors can each be absorbed by the matrix elements for the production and decay to define the respective polarised matrix element of each process,

$$\mathcal{M}^{prod}_{\mu,\lambda} = \mathcal{M}^{prod}_\mu \epsilon^{*\mu}_\lambda(k) \quad , \quad \mathcal{M}^{decay}_{\nu,\lambda} = \epsilon^\nu_\lambda(k) \mathcal{M}^{decay}_\nu \tag{97}$$

with polarisation λ. The matrix element for the intermediate vector boson is, therefore more compactly expressed as a sum of the individual polarised matrix elements, summed for each contribution,

$$\mathcal{M} = \frac{1}{k^2 - m_V^2 + i\Gamma m_V} \sum_\lambda^4 \mathcal{M}^{prod}_{\mu,\lambda} \mathcal{M}^{decay}_{\nu,\lambda} = \sum_{\lambda=1}^4 \mathcal{M}^F_\lambda \tag{98}$$

Furthermore, this means the squared amplitude is given by,

$$|\mathcal{M}|^2 = \sum_{\lambda=3}^1 |\mathcal{M}^F_\lambda|^2 + \sum_{\lambda\neq\lambda'} \mathcal{M}^F_\lambda \mathcal{M}^{*F}_{\lambda'} \tag{99}$$

The amplitude contains interference contributions coming from the different polarisation configurations. As discussed in the previous section, these terms are generally non-zero, except when the squared amplitude is integrated over the azimuthal angle [83]. Therefore, in a detector setting, it is also necessary to consider the interference between the polarised samples as an additional contribution when modelling the data.

The production of polarised templates, therefore, manifests by replacing the numerator of the gauge boson propagator,

$$\sum_\lambda \epsilon^\mu_\lambda \epsilon^{*\nu}_\lambda \to \epsilon^\mu_\lambda \epsilon^{*\nu}_\lambda. \tag{100}$$

It should also be noted that the individual polarised contributions are not Lorentz-invariant. This statement is true in general for any cross-section with polarised initial/final-states. Within field theories, in order to generate frame-independent scattering amplitudes, the full space of Lorentz indices must be contracted and summed for each helicity configuration. When simulating polarised samples, the samples are generated for a specified reference frame. The sum of the contributions and interference is frame-independent, while the individual contributions are frame-dependent. When generating samples, a given frame of reference can be chosen to increase the sensitivity to a given polarisation. For MC generators, this can sometimes be a technical restriction as Lorentz invariance is typically used to compute matrix elements for unpolarised scattering [87].

4.3.3 Polarised VBS

The previous section explained how to model the polarisation for a single intermediate gauge boson at the LHC. The process involved using the completeness relation to express the matrix element as a product of polarised production and decay, as in Eq. 96. The same concept can be extrapolated to processes involving multiple intermediate gauge bosons decaying to fermions. The matrix element for n gauge bosons can be expressed as,

$$\begin{aligned}\mathcal{M}_{res} = \sum_{\lambda_1,\ldots,\lambda_n} & \mathcal{M}^{prod}_{\mu\ldots\nu}(k_1,\ldots,k_n)\epsilon^{\mu}_{\lambda_1}(k_1)\ldots\epsilon^{\nu}_{\lambda_n}(k_n) \\ & \times \epsilon^{*\gamma}_{\lambda_1}(k_1)\ldots\epsilon^{*\eta}_{\lambda_n}(k_n)\mathcal{M}^{decay}_{\gamma}(k_1)\ldots\mathcal{M}^{decay}_{\eta}(k_n) \\ & \times \frac{1}{(k_1^2 - m_V^2 + i\Gamma m_V)\ldots(k_n^2 - m_V^2 + i\Gamma m_V)},\end{aligned} \tag{101}$$

where the sum is performed over the polarisations λ_i, with $k_1 \ldots k_n$ as the momenta of the gauge bosons, $\mathcal{M}^{prod}_{\mu\ldots\nu}(k_1,\ldots,k_n)$ as the matrix element for the production, and $\mathcal{M}^{decay}_{\eta}(k_i)$ the individual amplitudes for each decay. Using this expression, the polarisation vectors $\epsilon^{\mu}_{\lambda_i}$ can be absorbed into the matrix element for production and decay such that the total amplitude can be expressed in terms of the individual polarisation contributions.

There are, however, situations where this decomposition does not hold. VBS predictions are conjointly modelled with non-VBS contributions. The non-VBS contributions include diagrams where one or more of the final state fermion pairs do not originate from the decay of a single gauge boson, as shown in Fig. 50. For these non-resonant contributions, the method of deconstructing the matrix element into the polarised production and decay, as in Eq. 101, is not valid.

These diagrams cannot simply be ignored as this would violate gauge invariance, but it is also unclear how to include these diagrams in polarised predictions. This means that polarised templates of a definite polarisation can only be constructed for contributions with resonant gauge bosons where Eq. 101 is valid. As such, the only way to continue without directly violating gauge invariance is to introduce a suitable approximation to account for the non-resonant diagrams. Two common approximations exist: the Narrow-Width Approximation (NWA) and the On Shell Projection (OSP) method.

The NWA approximation replaces the denominator of the electroweak gauge boson

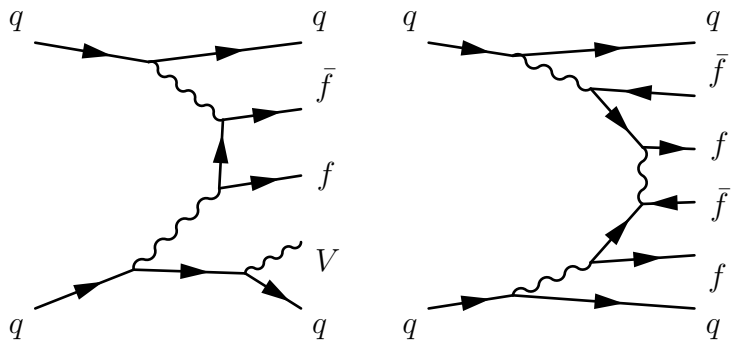


Figure 50: Feynman diagrams showing EW-induced $VVjj$ production, with a single-resonant gauge boson (left) and non-resonant production (right).

propagator by the following approximation,

$$\frac{1}{k^2 - m_V^2 + i\Gamma m_V} \to \frac{\pi\delta(k^2 - m_V^2)}{\Gamma m_V} \tag{102}$$

where δ is the delta function. The NWA ensures that only on-shell contributions remain. The second common approximation is the OSP method. The method works by projecting the numerator of the gauge boson propagator to the on-shell momentum. To demonstrate this, consider the full matrix element, which can be expressed as the sum of the resonant and non-resonant contributions:

$$\mathcal{M} = \mathcal{M}_{res} + \mathcal{M}_{nonres} \tag{103}$$

Consider the case of two intermediate gauge bosons as in a VBS interaction. Using Eq 101, the resonant matrix element $\mathcal{M}_{res}$. can be written as,

$$\begin{aligned}
\mathcal{M}_{res} = &\sum_{\lambda_1,\lambda_2} \mathcal{M}^{prod}_{\mu\nu}(k_1,k_2)\epsilon^{\mu}_{\lambda_1}(k_1)\epsilon^{\nu}_{\lambda_2}(k_2) \\
&\times \epsilon^{*\gamma}_{\lambda_1}(k_1)\epsilon^{*\eta}_{\lambda_n}(k_n)\mathcal{M}^{decay}_{\gamma}(k_1)\mathcal{M}^{decay}_{\eta}(k_2) \\
&\times \frac{1}{(k_1^2 - m_V^2 + i\Gamma m_V)(k_2^2 - m_V^2 + i\Gamma m_V)} \\
=&\frac{\sum_{\lambda_1,\lambda_2} \mathcal{M}^{prod}_{\lambda_1\lambda_2}(k_1,k_2)\mathcal{M}^{decay}_{\lambda_1}(k_1)\mathcal{M}^{decay}_{\lambda_2}(k_2)}{(k_1^2 - m_V^2 + i\Gamma m_V)(k_2^2 - m_V^2 + i\Gamma m_V)}
\end{aligned} \tag{104}$$

where in the final line, the polarisation vectors have been absorbed to define the polarised production and decay matrix elements of each gauge boson i with polarisation λ_i:

$$\begin{aligned}
\mathcal{M}^{prod}_{\lambda_1\lambda_2}(k_1,k_2) &= \mathcal{M}^{prod}_{\mu\nu}(k_1,k_2)\epsilon^{\mu}_{\lambda_1}(k_1)\epsilon^{\nu}_{\lambda_2}(k_2) \\
\mathcal{M}^{decay}_{\lambda_i}(k_i) &= \epsilon^{*\gamma}_{\lambda_1}(k_i)\mathcal{M}^{decay}_{\gamma}(k_i)
\end{aligned} \tag{105}$$

In the OSP method, the momenta of the gauge bosons in the numerator of the propagators are replaced with their on-shell projection $\bar{k}_i$ while the denominator stays the same. The

matrix element becomes [83],

$$\begin{aligned}\mathcal{M} &= \frac{\sum_{\lambda_1,\lambda_2} \mathcal{M}^{prod}_{\lambda_1\lambda_2}(k_1,k_2)\mathcal{M}^{decay}_{\lambda_1}(k_1)\mathcal{M}^{decay}_{\lambda_2}(k_2)}{(k_1^2 - m_V^2 + i\Gamma m_V)(k_2^2 - m_V^2 + i\Gamma m_V)} + \mathcal{M}_{nonres} \\ &\to \frac{\sum_{\lambda_1,\lambda_2} \mathcal{M}^{prod}_{\lambda_1\lambda_2}(\bar{k}_1,\bar{k}_2)\mathcal{M}^{decay}_{\lambda_1}(\bar{k}_1)\mathcal{M}^{decay}_{\lambda_2}(\bar{k}_2)}{(k_1^2 - m_V^2 + i\Gamma m_V)(k_2^2 - m_V^2 + i\Gamma m_V)} = \mathcal{M}_{OSP}\end{aligned} \tag{106}$$

The two approaches are similar, but the OSP partially includes off-shell effects since the denominator is unchanged. The approximations both have an error of the order $\mathcal{O}(\Gamma/M)$ and have relatively high accuracy when the decay width Γ, is much smaller than the mass of the resonant particle. For both methods, in order to recover some of the off-shell kinematics, the mass of each intermediate gauge boson is smeared according to a Breit-Wigner (BW) distribution with the kinematics of the outgoing particles adjusted to the smearing. The NWA approximation is used by MC generators MadGraph and Sherpa, which are the generators used in this experiment.

To summarise, there are a number of considerations in modelling the polarisation of a process such as VBS. Polarisation samples are necessary to extract polarisation information from unpolarised data. Polarisation is not a Lorentz invariant quantity, and neither are the cross-section predictions. Polarisation can be measured in different frames of reference, and this can be used to enhance the sensitivity to a given polarisation. For processes involving multiple intermediate gauge bosons, the polarisation is undefined for contributions coming from non-resonant production. To omit these diagrams in a gauge invariant manner, there are two common approximations: the NWA or the OSP method. Both force the gauge boson propagator terms to be on-shell, removing diagrams with non-resonant production and ultimately allowing for the generation of polarised templates.

5 Experimental Methods and Analysis

This chapter outlines the aims and objectives of this experiment. The major work of this thesis is dedicated to the experimental modelling and data analysis of polarisation. The measurement is performed using proton-proton collision data generated by the LHC at $\sqrt{s} = 13$ TeV and collected by the ATLAS experiment over the period of 2015-2018. The work of this thesis did not contribute to the data collection and is solely focused on the physics of polarisation. The analysis was performed in conjunction with the first ATLAS measurement of polarised $W^{\pm}W^{\pm}$ production via vector-boson scattering. However, it differs from the ATLAS measurement in several respects, for example, by the multivariate techniques. ATLAS measurement is referred to as the polarisation analysis/measurement.

The analysis is built upon the ATLAS differential, and fiducial cross-section measurement of inclusive and electroweak $W^{\pm}W^{\pm}jj$ VBS [88], referred to as the differential analysis/measurement. Aspects of the experimental analysis described by this thesis are inherited from the differential analysis and remain unchanged. The work of the previous result is adapted for the measurement of polarisation. This chapter describes the inherited aspects of the differential analysis and some additional work implemented for the polarisation measurement.

5.1 Introduction to $W^{\pm}W^{\pm}jj$ Production

The experimental analysis of data described in this thesis measures the cross-sections of longitudinally polarised EW-induced $W^{\pm}W^{\pm}jj$ production. The measurement is performed using Run-2 data consisting of $\sqrt{s} = 13$ TeV proton-proton collisions generated by the LHC and collected by the ATLAS detector. The cross-sections is measured in the W boson leptonic decay channels. For proton-proton collisions, the full final state of the process consists of two same-charge leptons in association with two jets and large missing transverse energy. To motivate this experimental analysis of this process, this section outlines the advantages of studying polarisation through same-electric charge $W^{\pm}W^{\pm}jj$ VBS.

Experimental measurements of polarised VBS are a crucial test of the electroweak symmetry-breaking mechanism of the Standard Model. Out of all VBS processes, $W^{\pm}W^{\pm}jj$ is often referred to as the "golden channel" for VBS measurements. This is mainly because the process has the highest EW- to QCD-induced production ratio. The QCD-induced production mode is typically one of the largest and most difficult backgrounds to constrain for VBS experiments. The ratio of the cross-sections for each of the different VBS processes and their respective EW- and QCD-induced production modes is shown in Table 7. The table illustrates that the electroweak to QCD ratio of $W^{\pm}W^{\pm}jj$ is almost an order of magnitude larger than the other VBS processes.

The ratio of the cross-section for the QCD-induced VBS is respectively smaller for $W^{\pm}W^{\pm}jj$ due to the same-electric charge requirement. The requirement eliminates many of the potential QCD diagrams that are present in other VBS processes. Figure 51 shows some of the major contributions to QCD-induced $VVjj$, where the diagrams crossed out in red do not contribute to QCD-induced $W^{\pm}W^{\pm}jj$. The same-electric charge requirement removes the quark-gluon and gluon-gluon contributions to the QCD-induced production

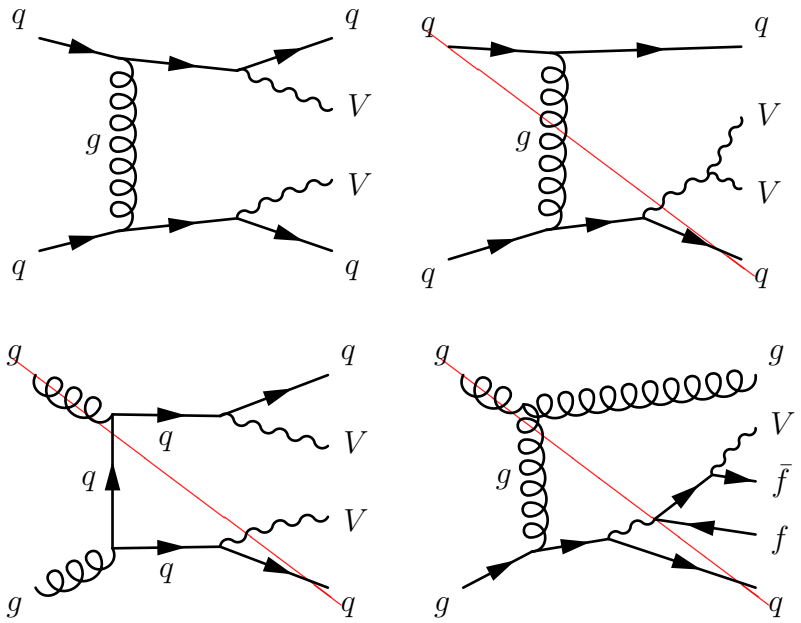


Figure 51: Feynman diagrams for QCD-induced $VVjj$ production at the LHC. The red-lines indicate diagrams present for QCD-induced $VVjj$, but not for QCD-induced $W^{\pm}W^{\pm}jj$.

and suppresses contributions coming from antiquark-quark annihilation.

Process	σ^{EW}	σ^{QCD}	$\frac{\sigma^{\mathrm{EW}}}{\sigma^{\mathrm{QCD}}}$
$pp \to ZZjj \to \nu_l\nu_l\nu'_l\nu'_ljj$	1.68	1.38	1.22
$pp \to W^{\pm}Zjj \to l^{\pm}\nu_l\nu'_l\nu'_ljj$	7.64	15.5	0.49
$pp \to W^{\pm}W^{\pm}jj \to l^{\pm}\nu_l l'^{\pm}\nu'_ljj$	3.97	0.346	11.5
$pp \to W^{\pm}W^{\mp}jj(ZZjj) \to l^{\pm}\nu_l l'^{\mp}\nu'_ljj$	12.3	21.8	0.56
$pp \to W^{\pm}Zjj \to l^{\pm}\ l^{\mp}l'^{\mp}\nu'_ljj$	2.34	4.38	0.53
$pp \to ZZjj \to l^{\pm}\ l^{\mp}l'^{\pm}l'^{\mp}jj$	0.098	0.1	0.98

Table 7: A comparison and the ratio of the cross-sections in a VBS phase-space for EW-induced (σ_{EW}) and QCD-induced (σ_{QCD}) $VVjj$ production. Shown for different VBS processes and their lepton channels. Table adapted from [18].

The same-electric charge requirement also has an effect on the electroweak diagrams. The s-channel electroweak contributions, including s-channel VBS diagrams, are not part of $W^{\pm}W^{\pm}$ VBS, as illustrated by Fig. 52. This process is, however, still highly sensitive to EWSB as the cross-section for longitudinal-longitudinal $W_L^{\pm}W_L^{\pm}$ is still predicted to diverge without Higgs contributions. Further, the same-charge requirement suppresses electroweak diagrams that do not involve VBS contributions and non-resonant gauge bosons. This itself is advantageous as it acts to improve the sensitivity to the electroweak gauge coupling [89].

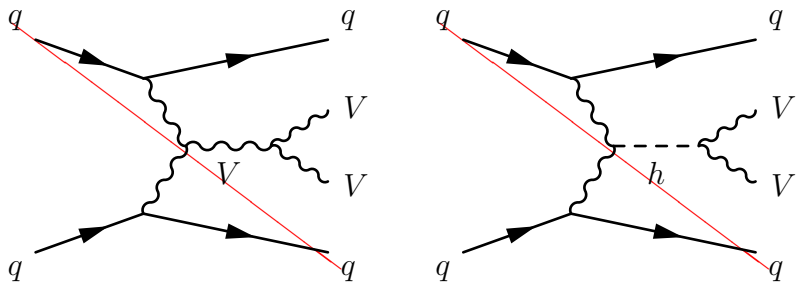


Figure 52: Feynman diagrams for s-channel EW-induced $VVjj$ production at the LHC. The red-lines indicate diagrams present for QCD-induced $VVjj$, but not for QCD-induced $W^{\pm}W^{\pm}jj$.

5.2 Analysis Objectives and Overview

The focus of this experiment is to measure the cross-sections of longitudinally polarised $W^{\pm}W^{\pm}jj$. This experiment measures the cross-sections for,

1. longitudinal-longitudinal $W_L^{\pm}W_L^{\pm}jj$ production,
2. longitudinal-any (LX), $W_L^{\pm}W_X^{\pm}jj$ production.

Each measurement is performed in the WW and pp CoM frame, associated with the respective definition of polarisation. In total, the experiment performs four cross-section measurements. This section first discusses the existing polarisation measurements to outline this experiment's objectives. At the time of writing, the most relevant result is the polarised cross-section measurements of $W^{\pm}W^{\pm}jj$ using Run-2 data reported by CMS [90]. The collaboration recorded polarised cross-section measurements with at least one longitudinally polarised W boson with an observed significance of 2.3σ and reported a 95% confidence level upper limit of 1.17 fb for purely longitudinally polarised scattering.

The measurement performed by CMS is similar to the work presented in the thesis. However, there are some significant differences. The fiducial region of the CMS result is different. This analysis also applies other multivariate analysis techniques and a newer modelling of the polarised samples. However, as the available dataset of both measurements is similar and both results are largely dominated by statistical uncertainty, the measurement of this experiment is comparable to the CMS result. This analysis performs analogous measurements, with the aim to measure $W^{\pm}W^{\pm}jj$ VBS with at least one longitudinally polarised W boson in the WW CoM frame to $> 3\sigma$ significance (corresponding to evidence of this process).

5.2.1 Analysis Structure

The cross-section is measured in a VBS fiducial region, designed to be enriched with $W^{\pm}W^{\pm}jj$ events. The experiment itself is implemented using an analysis framework for processing ATLAS reconstructed data: simulated and experimental. The analysis is built upon an ATLAS internal framework, the Common Analysis Framework (CAF) [91], which was developed using ATLAS data analysis software (AthAnalysis [74]). The

polarisation analysis was implemented with AthAnalysis (21.8). The remainder of this chapter discusses each component of the analysis associated with modelling and event selection.

5.3 Samples

The Standard Model predictions for event yields in this experiment are modelled by a combination of Monte Carlo (MC) event generators and data-driven estimates. The term samples refers to sets of MC events based on theoretically calculated processes. MC generators simulate particle physics events and are widely used in experimental particle physics. This section outlines the MC-generated samples used in this analysis and the settings used for each sample's generation. The background samples in the analysis are mostly inherited from the differential analysis. The signal samples are updated to polarised EW samples for the polarisation experiment.

All of the samples in this analysis are generated using the ATLAS simulation internal framework [92] (AthGeneration-21.6.27). The framework acts as an interface for producing samples with each MC generator. The framework also provides event simulation of the ATLAS detector using GEANT4 [93] and can run the full ATLAS event reconstruction chain for simulated data. Each sample is additionally overlaid with minimum-bias events generated with Pythia 8.186 [94] to simulate the pile-up effects. Minimum-bias events refer to the broad set of events passing the minimum-bias trigger, which selects inelastic collisions with no bias towards a particular final state [75].

Three MC event generators are used for sample production: MadGraph5 aMC@NLO [95, 23], Sherpa [96] and Powheg-Box [97, 98, 99]. Each generator can generate fixed-order tree-level and NLO hard-scatter events and can be interfaced with parton-shower MC generators for hadronisation. Samples generated with Madgraph and Powheg are interfaced with the Pythia generator [100] for parton-showering. The MC generator Sherpa has a built-in parton shower algorithm [101] following the ME+PS@NLO prescription [102]. The samples are also generated using the NNPDF sets [103] with the PDF and scale variations to estimate theoretical uncertainties.

5.3.1 Background Samples

The backgrounds in this analysis are modelled with a combination of MC and data-driven techniques; the latter is discussed in subsequent sections. This subsection outlines the background MC samples used in this analysis and their generation. The complete sample list is shown in Table 8, which illustrates the background sample processes, the MC event generators used to produce the sample, the order at the matrix element (ME) level and the PDF set used for the ME calculations.

The first set of background samples is referred to as the diboson $VVjj$ backgrounds. The EW- and QCD-induced production modes are generated separately by fixing the order of the EW coupling constant α_{EM}. For the QCD samples, the order is fixed to $\mathcal{O}(\alpha_{\text{EM}}^4)$ and for the electroweak samples, $\mathcal{O}(\alpha_{\text{EM}}^6)$. The diboson QCD background samples are all produced with the Sherpa 2.2.2 generator. The matrix elements are computed at NLO by Sherpa using Comix [104] and OneLoops [105] and are merged with a sample with one

additional jet calculated at NLO and samples with up to three additional jets calculated at LO, following the MEPS@NLO prescription [102]. The EW $VVjj$ samples are generated at LO with a combination of MadGraph5_aMC@NLO 2.6.2, interfaced with Pythia8.235 using the dipole recoil scheme[3] and Sherpa 2.2.2.

Process	ME Generator + Parton Shower	Order	PDF Set
Diboson			
EW $WZjj$	MadGraph5_aMC@NLO2.6.2+Pythia8.235	LO	NNPDF3.0NLO
EW $ZZjj$	Sherpa2.2.2	LO	NNPDF3.0NNLO
EW $V\gamma jj$	MadGraph5_aMC@NLO2.6.5+Pythia8.240	LO	NNPDF3.0NLO
QCD $WZjj$	Sherpa2.2.12	0,1j@NLO; 2,3j@LO	NNPDF3.0NNLO
QCD $VVjj$	Sherpa2.2.2	0,1j@NLO; 2,3j@LO	NNPDF3.0NNLO
QCD $V\gamma jj$	Sherpa2.2.11	0,1j@NLO; 2,3j@LO	NNPDF3.0NNLO
Triboson			
VVV	Sherpa2.2.1 (leptonic) Sherpa2.2.2 (one $V \to jj$)	+0,1j@LO	NNPDF3.0nnlo
Other prompt			
tZq	MadGraph5_aMC@NLO2.3.3.p1 + Pythia8.212	LO	NNPDF2.3LO
$t\bar{t}V$	MadGraph5_aMC@NLO2.3.3.p0 + Pythia8.210	NLO	NNPDF3.0NLO

Table 8: The MC background samples with the MC and PS generator, the process order and parton shower scheme, and the PDF set generated with the sample detailed for each background [89, 88].

In addition to the diboson samples, triboson processes (VVV) with four final state leptons and two jets are considered as background. The triboson samples are generated excluding pairs of same-charge leptons. Triboson production diagrams with pairs of same-charge leptons are included in EW $W^{\pm}W^{\pm}jj$ samples. The triboson backgrounds are simulated at LO with Sherpa2.2.2 with up to one additional jet, where one of the gauge bosons decays hadronically.

The third major category of background samples, referred to as "Other prompt" in Table 8, corresponds to processes that can produce two same-charge prompt leptons, which are not $VVjj$ or triboson processes. Two processes are considered for these samples, both involving top quark production. The first sample is top quark pair production in association with a gauge boson, $t\bar{t}V$. This sample is generated at NLO with MadGraph5_aMC@NLO 2.3.3 and showered with Pythia 8.210. The second sample is the production of a single top quark in association with a Z boson, tZq. This sample is generated at LO with MadGraph5_aMC@NLO 2.3.3 and interfaced with Pythia 8.212.

5.3.2 Signal Samples

The signal samples are the biggest difference between the MC samples in the differential and polarisation measurements. The unpolarised $W^{\pm}W^{\pm}jj$ samples inherited from the

[3]It was found in the differential analysis the modelling of radiation in the central region of rapidity for radiation gave over-predictions compared to data. The mismodeling is mitigated using the dipole-recoil scheme [106].

differential measurement are separated by the EW coupling constant. The EW-induced production at $\mathcal{O}(\alpha_{\text{EM}}^6)$, QCD-induced production at $\mathcal{O}(\alpha_{\text{EM}}^4)$ and interference-induced production at $\mathcal{O}(\alpha_{\text{EM}}^5)$. These samples are referred to as inclusive $W^\pm W^\pm jj$ production. Each sample is simulated at LO accuracy with Madgraph5_aMC@NLO 2.6.7 and is parton-showered with Pythia 8.244 using the dipole-recoil scheme [107].

Process	ME Generator + parton shower	Order	PDF Set
Signal			
EW $W^\pm W^\pm jj$	MadGraph5_aMC@NLO2.6.7 + Pythia8.244	LO	NNPDF3.0NLO
QCD $W^\pm W^\pm jj$	MadGraph5_aMC@NLO2.6.7 + Pythia8.244	LO	NNPDF3.0NLO
INT $W^\pm W^\pm jj$	MadGraph5_aMC@NLO2.6.7 + Pythia8.244	LO	NNPDF3.0NLO
EW $W_\lambda^\pm W_\lambda^\pm jj$	MadGraph5_aMC@NLO2.9.5 + Pythia8.245	LO	NNPDF3.0NLO

Table 9: The MC signal samples with the MC and PS generator, the process order and parton shower scheme, and the PDF set generated with the sample detailed for each background.

The polarisation samples, denoted by $W_\lambda^\pm W_\lambda^\pm jj$ are the new samples generated for this analysis. The polarised $W_\lambda^\pm W_\lambda^\pm jj$ EW samples are generated analogously at $\mathcal{O}(\alpha_s^6)$ but require fixed polarisation states $\lambda = L, T$, of the intermediate gauge bosons. In total, there are three unique gauge boson polarisation configurations: purely longitudinal, LL, mixed polarisation, LT and purely transverse, TT. The samples are generated with MadGraph5_aMC@NLO 2.9.5 at LO accuracy and interfaced with Pythia 8.245 for showering and hadronisation. The polarisation samples are produced with the NWA to remove off-shell contributions. The off-shell contributions are treated as an additional background and estimated by subtracting the on-shell contributions from the inclusive EW sample. In order to define polarisation, the samples must be generated in a given reference frame. Only the summed contribution is frame-independent. The samples are generated in two reference frames, the incoming parton-parton (pp) centre-of-mass (CoM) frame and the CoM frame of the two intermediate W bosons WW. Cross-section measurements are performed for both CoM frames.

5.4 Data

The experimental data used in this measurement was collected by the ATLAS detector from 2015 to 2018 during Run-2 of the LHC. The dataset consists of proton-proton collisions at $\sqrt{s} = 13$ TeV produced by the LHC operating with a bunch spacing on 25 ns. The total integrated luminosity of the dataset totals to 139.0 fb^{-1}. Events used in this experiment are required to pass a predefined set of data quality requirements. This analysis follows the ATLAS recommendations [108], which are defined by Good Run Lists (GRLs). GRLs are references to luminosity blocks where the recommended data quality requirements are passed [109]. The ATLAS collaboration publishes GRLs, which are used in this analysis to select the experimental data. The GRLs used are shown in Table 10.

GRLs
data15_13TeV.periodAllYear_DetStatus-v89-pro21-02_Unknown_PHYS_StandardGRL_All_Good_25ns.xml
data16_13TeV.periodAllYear_DetStatus-v89-pro21-01_DQDefects-00-02-04_PHYS_StandardGRL_All_Good_25ns.xml
data17_13TeV.periodAllYear_DetStatus-v99-pro22-01_Unknown_PHYS_StandardGRL_All_Good_25ns_Triggerno17e33prim.xml
data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL_All_Good_25ns_Triggerno17e33prim.xml

Table 10: The GRLs used to select the experimental data in this analysis [108].

5.5 Object Reconstruction and Selection

The ATLAS reconstruction software can reconstruct several physics objects for data analysis. This analysis applies the same criteria to reconstructed analysis objects as the differential analysis. Initial testing showed that changing the selection criteria would not improve the analysis's sensitivity to polarisation. The selection criteria are applied in two stages to electrons, muons, jets and the E_T^{miss}. First, a baseline pre-selection is applied to all analysis objects. The objects are then subject to an overlap removal. The remaining objects are then required to pass the second stage of criteria, which defines the signal region. The selection criteria are implemented in the experiment using the analysis framework.

5.5.1 Electrons

Electrons are reconstructed using data from the EM calorimeters and the Inner Tracker [110]. Candidate electrons are formed by matching tracks reconstructed in the Inner Tracker to clusters of energy deposits in the EM calorimeters. Electrons are identified based on a likelihood (LH) discriminant [53]. A likelihood function is constructed from various quantities measured by the inner detector and calorimeters, such as shower shape, track quality, and track-to-calorimeter-cluster matching. There are four operating points for the identification criteria, referred to as VeryLoose, Loose, Medium, and Tight, corresponding to increasing thresholds for the LH discriminant [53].

The baseline criteria for electrons are summarised in the central column of Table 11. Baseline electrons are used to reject additional soft electrons in an event that could come from backgrounds such as photon conversions. The tracks matched to the electrons are subject to cuts on the transverse and longitudinal impact parameters that ensure the electron track originates from the primary vertex of the interaction. Baseline electrons are required to pass the Loose identification criteria.

The signal criteria for electrons are summarised in the right column of Table 11. Signal electrons are selected with $p_T > 27$ GeV and electrons within detector acceptance are required to fall outside of the EM calorimeter transition region ($1.37 \leq |\eta| \leq 1.52$). The signal electrons must pass the Tight identification criteria and the Gradient isolation criteria [110]. The isolation criteria are applied to isolation variables that quantify the activity near lepton tracks and energy deposits from particles nearby. In addition, the signal criteria require electrons to pass the selection applied by the Electron Charge ID Selector Tool (ECIDS). The ECIDS is a charge selector tool which uses boosted decision trees to reject candidate electrons whose charge has been potentially misidentified [110].

Parameter	Baseline	Signal Selection
Identification	LooseLH	TightLH
Kinematic Acceptance	$p_T > 4.5$ GeV	$p_T > 27$ GeV
Geometrical Acceptance	$\|\eta\| < 2.47$	$\|\eta\| < 2.47$ and $\|\eta\| \notin [1.37, 1.52]$
Longitudinal Impact parameter	$\|z_0 \times \sin\theta\| < 0.5$ mm	$\|z_0 \times \sin\theta\| < 0.5$ mm
Transverse Impact parameter	$\|\frac{d_0}{\sigma d_0}\| < 5$	$\|\frac{d_0}{\sigma d_0}\| < 5$
Isolation Requirement	None	Gradient
		Pass ECIDS

Table 11: A comparison of the baseline and signal region selection criteria applied to electrons in the differential and polarisation analyses.

5.5.2 Muons

Muons are reconstructed from the tracks measured by the muon Spectrometers and matched, if possible[4], to a corresponding track in the Inner Detector [57, 58]. The muon selection criteria in this analysis are based on the recommended requirements of the Muon Combined Performance group [58] and are provided by the Muon Selection Tool. Analogolously to electrons, muons are selected based on various identification criteria denoted by: Tight, Medium, Loose, VeryLoose, HighPt, LowPt [58]. The default ATLAS recommendation is Medium [58].

The baseline cuts for muons are shown in the central column of Table 12. Similarly to electrons, selection criteria are also applied to the impact parameters of the track associated with the muon in the ID to ensure the muons originate from the primary vertex. In addition, baseline muons are subject to the Loose identification criteria.

Signal muons are required to have a $p_T > 27$ GeV with the remaining selection criteria detailed in the right column of Table 12. Signal muons must pass the identification criteria, Medium, and also, in addition, the Isolation Requirement FixedCutPflowTight [57]. Analogously, the isolation criteria are selected based on calorimeter and tracking variables. The isolation criteria are applied to reduce the contribution of muons coming from b-jets and in-flight decays [88].

5.5.3 Jets

Jets are reconstructed using a particle flow algorithm [64] using data from the ID and the calorimeters. The jets are reconstructed using the anti-k_T algorithm [65] with a resolution parameter of $R = 0.4$. The baseline cuts require jets with a $p_T > 20$ GeV and $|\eta| \leq 4.5$. The p_T requirement is increased to $p_T > 25$ GeV for signal jets. In addition, jets originating from pile-up interactions are reduced using the Jet-Vertex-Tagger algorithm

[4]Some Muons fall outside the acceptance of the ID but within the MS. For these muons, the tracks are extrapolated to the beamline [57].

Parameter	Baseline	Signal Selection
Identification	Loose	Medium
Kinematic Acceptance	$p_T > 3$ GeV	$p_T > 27$ GeV
Geometrical Acceptance	$\|\eta\| < 2.7$	$\|\eta\| < 2.5$
Longitudinal Impact Parameter	$\|z_0 \times \sin\theta\| < 1.5$ mm	$\|z_0 \times \sin\theta\| < 0.5$ mm
Transverse Impact Parameter	$\|\frac{d_0}{\sigma d_0}\| < 15$	$\|\frac{d_0}{\sigma d_0}\| < 3$
Isolation Requirement	None	FixedCutPflowTight

Table 12: A comparison of the baseline and signal region selection criteria applied to muons in the differential and polarisation analyses.

[111]. The algorithm is applied to jets reconstructed with a $p_T < 60$ GeV. In addition, a b-tagging algorithm is applied to jets during jet selection. Jets with a b quark are identified with the 85% efficiency working point of the DL1r heavy-flavour tagger [112].

5.5.4 Missing Transverse Energy

$W^{\pm}W^{\pm}jj$ signal events consist of two neutrinos in the final state. As such, signal events typically have a large E_T^{miss}. The reconstructed objects passing the above baseline selection, in particular baseline objects from the primary vertex, are defined in order to improve the E_T^{miss} resolution. Within ATLAS, the E_T^{miss} is reconstructed from two contributions: hard and soft objects [67]. Hard objects are composed of fully reconstructed particles and jets. Soft objects correspond to charged-particle tracks associated with the hard-scatter vertex. In addition, when calculating E_T^{miss}, an internal ambiguity resolution is applied to prevent double counting [67]. The reconstruction of each particle is performed independently, which means there is potential for the same detector signal to be used more than once.

5.5.5 Overlap Removal

The full event reconstruction consists of multiple algorithms, which use detector data to reconstruct the physics objects of the event. Each algorithm processes the detector readout independently so that if one of the algorithms is updated, the remaining algorithms do not need to be updated to reflect these changes. The drawback to this approach is that duplicate objects can be reconstructed from the same detector measurements. For this reason, an overlap removal (OR) procedure is to resolve any reconstruction ambiguities.

The overlap removal procedure applied in this thesis follows the standard ATLAS recommended procedure [113]. A brief description is given in this subsection. Overlap removal must discriminate between the case of duplication and the case of two close-by physics objects. A common measure of the proximity of objects is defined as $\Delta R = \sqrt{\Delta\phi^2 + \Delta y}$, where $\Delta\phi$ is the difference in the azimuthal angle and Δy is the difference in the rapidity of the two objects. This variable is used since both ϕ and y are Lorentz

invariant under a boost along the beamline. The OR procedure is applied iteratively following a prescribed priority list.

The first stage of the OR procedure is applied to lepton object pairs. There are two possibilities for duplicate leptons: overlapping electrons and overlapping electron muon pairs. The case of duplicate muons is not considered as they are removed during reconstruction by the TrackSelectorProcessorTool [113]. Electron duplicates can occur if two electrons are reconstructed with an overlapping EM cluster or a shared track. The recommendation is to remove the softer electron if the electron is closer than 0.05 in ΔR to the leading electron and is of the same charge or shares the same track. A duplicate muon electron pair is possible if the muon radiates a high-energy photon via bremsstrahlung. The recommendation is, if the two objects are closer than 0.01 in ΔR or share the same ID track, to keep the muon, providing the muon has been tagged in the muon chambers. Otherwise, the electron is kept.

The second stage of the OR procedure is applied primarily between electron and light-flavour jets. The OR removal procedure aims to remove jets that are identical to reconstructed prompt electrons and preserve heavy-flavour jets with semi-leptonic decays and light-flavour jets that can fake an electron. The recommended procedure for an overlapping jet-electron is if the electron-jet is closer than 0.2 in ΔR, then the jet is removed. A secondary pass is then applied, and the electron is removed for surviving jets, with electron jet pairs with a distance between $0.2 < \Delta R < 0.4$.

5.6 Signal Region Definition

Table 13 shows the signal region selection. The selection can be subdivided into three categories: lepton, jet and E_T^{miss} cuts. The signal region requires the presence of exactly two same electric charge signal leptons (electrons or muons). The leptons are required to have $p_T > 27$ GeV. Additional cuts are applied to reduce the background contribution coming from Drell–Yan events, $Z/\gamma \to ee$, where the charge of one of the electrons is misidentified. To suppress this background, in the ee-channel, leptons are required to have an $|\eta| \leq 1.37$. In these events, a Z veto is also applied such that events with a dilepton mass of $|m_{ee} - m_Z| < 15$ GeV are rejected.

To suppress the background contribution from processes with more than two leptons in the final state, events with a third pre-selected lepton that survives overlap removal are rejected. Furthermore, events are also rejected if the third lepton does not survive overlap removal but, together with one of the signal leptons with opposite charge and same flavour, has a dilepton mass with $|m_{ll} - m_Z| < 15$ GeV. Along with the lepton selection, a selection is applied to the missing transverse energy due to the two final-state neutrinos.

The final set of signal selection criteria is for the signal jet objects. For a VBS event, there are expected to be two high-energy jets, and as such, the selection requires at least two jets, with the leading jet $p_T \geq 65$ GeV and the sub-leading jet $p_T \geq 35$ GeV. The two jets are selected with the largest p_T, referred to as the tagging jets, and are used to define the dijet invariant mass m_{jj}, and the dijet rapidity difference Δy_{jj}. The jet selection also rejects any events identified by the DL1r heavy-flavour tagger containing a b-jet in order to suppress the contributions from top processes. Finally, the selection

$W^{\pm}W^{\pm}jj$ Signal Region	
Leptons	
Number	2 signal leptons with equal electric charge
Lepton p_T	$p_T > 27$ GeV
Lepton η	$\|\eta\| < 2.5$ for muons $\|\eta\| < 2.47$ for electrons excluding $1.37 \leq \|\eta\| \leq 1.52$; in the ee-channel $\|\eta\| \leq 1.37$
Dilepton mass, m_{ll}	$m_{ll} \geq 20$ GeV $\|m_{ee} - m_Z\| > 15$ GeV
Jets	
Number	At least 2 jets
Leading p_T	$p_T \geq 65$ GeV
Sub-leading p_T	$p_T \geq 35$ GeV
Dijet Mass, m_{jj}	$m_{jj} \geq 500$ GeV
B-veto	b-jet veto using the DL1r tagger
Dijet Rapidity	$\|\Delta y_{jj}\| > 2$
Missing Transverse Energy	
E_T^{miss}	$E_T^{miss} \geq 30$ GeV

Table 13: The signal region defined for both the polarised and differential analyses. Cuts applied here are to signal objects passing the object selection.

requires a large dijet invariant mass $m_{jj} \geq 500$ GeV and rapidity difference $|\Delta y_{jj}| > 2$, which act to suppress the QCD-induced VBS background.

5.7 Backgrounds

There are a number of different processes that can directly or indirectly produce the electroweak-induced $W^{\pm}W^{\pm}jj$ final state of $l^{\pm}l^{\pm}+jj+E_T^{\text{miss}}$. These processes are referred to as backgrounds and must be modelled to represent the expected event yields in the signal region accurately. Some backgrounds contribute by directly reproducing the same final state, while others enter the signal region due to detector limitations. To extract the number of polarised $W^{\pm}W^{\pm}jj$ events observed in data, the background contribution to the signal region must also be correctly accounted for. The backgrounds are estimated through a combination of MC simulations and data-driven methods and can be grouped into the following categories,

1. $W^{\pm}W^{\pm}jj$ QCD and interference,

2. $W^{\pm}Zjj$ QCD and EW,

3. non-prompt lepton backgrounds,

4. photon-conversion ($V\gamma$),

5. electron charge misidentification,

6. other prompt.

This section gives a detailed description of each of the above backgrounds and a discussion of their modelling. Table 14 shows the predicted event yields for each of the different processes. The biggest background contribution arises from $W^{\pm}Zjj$ production. The second largest contribution is from the non-prompt lepton background.

Process	Predicted Event Yields
$W^{\pm}Zjj$ QCD	83 ± 9
$W^{\pm}Zjj$ EW	14.9 ± 1.2
$W^{\pm}W^{\pm}jj$ QCD	24 ± 7
$W^{\pm}W^{\pm}jj$ INT	7.6 ± 1.6
Non-promt	56 ± 12
$V\gamma$	11 ± 5
Charge misid.	10 ± 4
Other prompt	7.1 ± 2.8
$W_L^{\pm}W_L^{\pm}jj$ EW	18.3 ± 0.9
$W_L^{\pm}W_T^{\pm}jj$ EW	59 ± 3
$W_T^{\pm}W_T^{\pm}jj$ EW	125 ± 6

Table 14: The predicted background and signal event yields in the signal region for 139.0 fb^{-1}. Table adapted from the differential analysis [89, 88].

5.7.1 $W^{\pm}Zjj$ Production

The dominant background entering the signal region is $W^{\pm}Z/\gamma jj$ production. The modelling of this background is described in Chapter 6.

5.7.2 Non-prompt Background

The second largest source of background events arises from non-prompt leptons passing the lepton signal selection. Non-prompt leptons are defined as leptons that do not originate from the primary collision; they are also referred to as fake leptons. Non-prompt leptons can appear from the misidentification of jets as leptons or are produced by heavy-flavour hadron decays. The main contributing processes to the non-prompt background are W+jets and semi-leptonic $t\bar{t}$ processes. They are overall the second largest background, accounting for 12% of the predicted event yield. The background is estimated using a data-driven technique referred to as the fake factor method.

Fake Factor Method

The fake factor method predicts the number of background events in the signal region using data in an independent region with a modified signal lepton selection, and scaling by a "fake factor" to recover the non-prompt contribution in the signal region. The modified lepton selection is referred to as the background lepton selection. The background leptons are chosen to have properties similar to a signal lepton; however, they would still fail the overall signal lepton selection due to the inversion of the identification or isolation criteria.

The fake factor is defined as the ratio of the number of non-prompt leptons passing the signal lepton selection $N_{\text{sig}}^{\text{non-prompt}}$, divided by the number of non-prompt leptons passing the background lepton selection $N_{\text{bkg}}^{\text{non-prompt}}$:

$$F = \frac{N_{\text{sig}}^{\text{non-prompt}}}{N_{\text{bkg}}^{\text{non-prompt}}} \tag{107}$$

The fake factors are estimated from data and extracted from a single-lepton dataset enriched with non-prompt leptons. The non-prompt contribution is calculated by subtracting the MC estimated prompt lepton contribution from the data, with the fake factors defined as,

$$F = \frac{N_{\text{sig}} - N_{\text{sig}}^{\text{prompt}}}{N_{\text{bkg}} - N_{\text{bkg}}^{\text{prompt}}} \tag{108}$$

where $N_{\text{sig/big}}$ is the total number of events with a lepton passing the signal/background selection and $N_{\text{sig/bkg}}^{\text{prompt}}$ is the MC-predicted number of prompt leptons passing the signal/background selection. The number of background events entering the signal region is estimated using an independent region that requires one signal lepton and one background lepton. The number of events with a non-prompt lepton in the signal region, $N_{\text{sig+sig}}^{\text{non-prompt}}$, is estimated to be the number of data events in the independent region, $N_{\text{sig+bkg}}$, minus the MC-predicted contribution from prompt leptons in this region, $N_{\text{sig+bkg}}^{\text{prompt}}$, multiplied by the fake factor:

$$N_{\text{sig+sig}}^{\text{non-prompt}} = F \times (N_{\text{sig+bkg}} - N_{\text{sig+bkg}}^{\text{prompt}}) \tag{109}$$

For the different-flavour lepton channels, the expression is given instead as,

$$N_{\text{sig+sig}}^{\text{non-prompt}} = F_e \times (N_{\text{sig}_\mu\text{+bkg}_e} - N_{\text{sig}_\mu\text{+bkg}_e}^{\text{prompt}}) + F_\mu \times (N_{\text{sig}_e\text{+bkg}_\mu} - N_{\text{sig}_e\text{+bkg}_\mu}^{\text{prompt}}) \tag{110}$$

where $F_{e/\mu}$ are the fake factors for electrons and muons.

Fake factor Estimation

The differential and polarisation analyses use a sample enriched in non-prompt leptons to estimate the fake factors using data. The sample is selected to contain at least one signal or background lepton candidate and one jet. As the lepton contribution is primarily non-prompt, the selection is referred to as the dijet selection. The dijet region is defined in Table 15. The signal and background leptons, defined in Table 16, differ in their identification and isolation requirements.

	Dijet Selection
Leptons	
Number	At least one signal or background lepton
Lepton-p_T	$p_T > 27$ GeV
Jets	
Number	At least 1 jets
jet p_T	$p_T \geq 25(30)$ GeV for $\lvert\eta\rvert < 2.5$ ($2.5 < \lvert\eta\rvert < 4.5$)
Dilepton-jet phi, $\phi_{l,j}$	$\lvert\phi_{l,j}\rvert > 2.8$
B-veto	b-jet veto using the DL1r tagger
Missing Transverse Energy	
E_T^{miss}	$E_T^{miss} + m_T \leq 50$ GeV

Table 15: The dijet region defined for both the polarised and differential analysis. Cuts applied here are to signal objects passing the object selection.

In the case of the fake-factor estimation, the backgrounds to the non-prompt sample are sources of prompt leptons, of which the largest is W+jets production. In order to reduce the number of prompt-leptons in the dijet region, the dilepton-jet pair is selected to be back-to-back in the azimuthal plane, with the dilepton-jet azimuthal difference, $\phi_{l,j}$, required to be $|\phi_{l,j}| > 2.8$. This reduces the number of prompt leptons from W+jets production since the charged lepton from the W boson decay is generally not in the same direction as the W boson. This background is further suppressed in the dijet region with the requirement that the sum of the missing transverse energy and the transverse mass of the lepton m_T, be less than 50 GeV.

The fake factors are estimated from two separate regions. The signal and background electrons are required to pass ECIDS in order to suppress charge misidentification. The jet must pass the criteria given in Table 15, with the exception of the B-veto. Two separate jet regions are defined, one with the B-veto and one without. The nominal fake factors used by both analyses are calculated as the average of the two estimates in these regions.

MC samples are used to estimate the prompt-lepton background in the dijet region. The background samples consist of W/Z + jets, single-top, $t\bar{t}$, and single photon production. The data in the dijet region are required to be on the GRL and selected by the trigger "HLT_e12_lhvloose_nod0_L1EM10VH" for data containing an electron and the trigger "HLT_mu14" for data containing a muon. The data from different triggers are combined by weighting each event with the trigger prescales.

Electron Selection		
Parameter	Signal Electron	Background Electron
Identification	TightLH	MediumLH
Isolation Requirement	Gradient	
		Fail LHTight or fail ID isolation
Kinematic Acceptance	$p_T > 27$ GeV	$p_T > 27$ GeV
Geometrical Acceptance	$\lvert\eta\rvert < 1.37$ and $(1.52 < \lvert\eta\rvert < 2.47)$	$\lvert\eta\rvert < 1.37$ and $(1.52 < \lvert\eta\rvert < 2.47)$
Longitudinal Impact Parameter	$\lvert z_0 \times \sin\theta\rvert < 0.5$ mm	$\lvert z_0 \times \sin\theta\rvert < 0.5$ mm
Transverse Impact parameter	$\lvert\frac{d_0}{\sigma d_0}\rvert < 5$	$\lvert\frac{d_0}{\sigma d_0}\rvert < 5$
ECIDS	Pass ECIDS	Pass ECIDS
Muon Selection		
Parameter	Signal Muon	Background Muon
Kinematic Acceptance	$p_T > 27$ GeV	$p_T > 27$ GeV
Geometrical Acceptance	$\lvert\eta\rvert < 2.5$	$\lvert\eta\rvert < 2.5$
Longitudinal Impact Parameter	$\lvert z_0 \times \sin\theta\rvert < 0.5$ mm	$\lvert z_0 \times \sin\theta\rvert < 0.5$ mm
Transverse Impact Parameter	$\lvert\frac{d_0}{\sigma d_0}\rvert < 3$	$\lvert\frac{d_0}{\sigma d_0}\rvert < 10$
		Fail ID $\lvert\frac{d_0}{\sigma d_0}\rvert$ or fail ID isolation
Identification	Medium	Medium
Isolation Requirement	FixedCutPflowTight	FixedCutPflowLoose

Table 16: A comparison of the background and signal lepton definitions.

5.7.3 Electron Charge Misidentification

A portion of the backgrounds entering the $W^{\pm}W^{\pm}$ signal region arise from the misidentification of electron charge. Events with opposite sign leptons e.g. from processes producing Z bosons such Drell-Yan, can contaminate the signal region if the charge of one of the leptons is incorrectly identified. For the ATLAS detector, the charge misidentification rate of muons is very small [114, 115]. As such the backgrounds associated with muon charge misidentification are insignificant and not considered in either analysis. By contrast, the electron charge misidentification rate is non-negligible and therefore, the background contributions must be considered.

Electron charge is measured from the curvature of the reconstructed track associated with the electron. Charge misidentification can occur if the track is falsely reconstructed or the track is correctly reconstructed, but the wrong track is associated with the electron.

The major source of the charge misidentification comes from the interaction of the electron with detector material. These interactions create secondary particles, photons and electron-positron pairs, which induce additional hits around the primary electron track. This can lead to either a situation where these additional hits are incorrectly associated with the hits from the primary track, leading to an incorrect track, or the hits can be reconstructed as an additional track, which ends up being falsely associated with the electron.

Measurement of the Charge Misidentification

The charge misidentification rate is measured using MC-simulated $Z \to ee$ samples. The misidentification probability can be calculated by taking the ratio of the number of misidentified electrons to the total number of electrons in a given phase space. The MC events, however, must first be corrected using a data-driven method. This is because detector simulation cannot precisely model detector interactions, which leads to charge misidentification, and the probabilities cannot be extracted from just the MC. To correct for the differences in the charge misidentification probability distributions for data, ϵ_{data}, and for MC ϵ_{MC}, the following scale factors are applied:

$$SF_{\text{wrong}} = \frac{\epsilon_{\text{data}}}{\epsilon_{\text{MC}}}, \quad SF_{\text{correct}} = \frac{1-\epsilon_{\text{data}}}{1-\epsilon_{\text{MC}}}, \tag{111}$$

where SF_{wrong} is the scale factor for an electron with an incorrectly reconstructed charge and SF_{correct} for a correctly reconstructed charge. The distributions of the scale factors, SF_{wrong} and SF_{correct}, are calculated and provided by the Egamma Combined Performance (CP) group [110]. The CP group uses a data-driven method referred to as the sideband method to measure the scale factors from data. The method involves isolating a $Z \to ee$ region, with two independent regions used to estimate and subtract the non-Z backgrounds in the main region [110]. From the scale factors, it is generally observed that charge misidentification rates are overestimated by MC, as the scale factor SF_{wrong} is typically below one.

The MC samples are weighted by scale factors using the truth information and the distributions provided by the CP group:

$$w_{SF} = SF_{l_0} \times SF_{l_1}, \tag{112}$$

where SF_{l_i} is the respective scale factor for the leading/sub-leading lepton. The charge misidentification probabilities are then calculated using opposite-sign electrons in a $Z \to ee$ MC sample, weighted with the scale factors. The probabilities are calculated in an independent region selecting leptons which pass the TightLH identification, the Gradient isolation criteria, and the ECIDS selection tool. The corrected MC samples are used to obtain a per-electron charge misidentification probability that can be applied to data. The charge misidentification probability distribution is calculated in a binned distribution in terms of p_T and η, defined as,

$$\epsilon_{\text{mis}} = \frac{N_{\text{mis}}}{N_{\text{tot}}} \tag{113}$$

where N_{mis} is the number of misidentified-charge leptons and N_{tot} is the total number of leptons.

Charge Misidentification Background Estimate

The background yields entering the signal due to charge misidentification are estimated by weighting data events with opposite-sign leptons by the charge misidentification probability. The weights applied to each event are:

$$w = \frac{\epsilon_{l_0}(1-\epsilon_{l_1}) + \epsilon_{l_1}(1-\epsilon_{l_0})}{(1-\epsilon_{l_0})(1-\epsilon_{l_1}) + \epsilon_{l_0}\epsilon_{l_1}}, \tag{114}$$

where ϵ_{l_i} is the charge misidentification probability for the corresponding lepton. The weight is the probability for either lepton to have a misidentified charge divided by the probability for neither or both leptons to have a misidentified charge. The weights are applied to all leptonic channels, with the muon charge misidentification probability set to $\epsilon_\mu = 0$.

Electron Energy Correction

The energy resolution worsens for charge misidentified electrons. Interactions with detector material lead to a significant energy loss. This means that the reconstructed energy of charge misidentified electrons is far lower than the truth energy of the electron. In this experiment, a correction is applied in $|\eta|$ and p_T to account for this effect. The correction is inherited from the differential analysis. The reader is referred to Refs. [89, 88, 115, 116] for a detailed explanation of the electron energy correction and the associated sources of uncertainty.

5.7.4 Photon Conversion Background

A prompt photon can be misidentified as an electron, referred to as a photon conversion [117]. This leads to another source of backgrounds from processes which can generate an electroweak gauge boson and a photon ($V\gamma$) if the gauge boson decays leptonically and the photon is misidentified as an electron. This background is modelled using MC [89, 88].

5.7.5 Other Prompt Backgrounds

A prompt lepton background refers to a background which enters the signal region by producing prompt leptons. The main prompt lepton backgrounds are the QCD-induced $W^\pm Zjj$ and the $W^\pm W^\pm jj$ processes. The "other" prompt lepton backgrounds refer collectively to ZZ production, triboson processes VVV, and top quark processes, namely tZq and $t\bar{t}V$. These backgrounds are all modelled with MC and combined account for 1.6% of the total predicted event yield in the signal region [89, 88].

5.8 Validation Regions

Validation regions are independent regions with overlapping criteria with the signal region that are designed to test modelling. Different backgrounds dominate the event yields in each validation region. In this analysis, the validation regions are used to test the

modelling of the multivariate analysis techniques on unseen data. The two validation regions are defined in Table 17 and referred to as,

1. the low Δy_{jj} region,
2. the top-fakes region.

The low ΔY_{jj} validation region has a very similar selection criteria to the signal region but with the ΔY_{jj} condition inverted. The region is largely dominated by QCD-induced background processes, particularly $W^{\pm}Zjj$ but also the non-prompt and charge misidentified backgrounds. The top-fakes validation region requires events with exactly one b-jet; the b-jet must not be the leading jet. The region is used to test the modelling on the non-prompt background.

Validation Regions		
	Low Δy_{jj} Region	Top-fakes Region
Leptons		
Number	2 signal-leptons with equal electric charge	2 signal-leptons with equal electric charge
Lepton-p_T	$p_T > 27$ GeV	$p_T > 27$ GeV
Lepton-η	$\lvert\eta\rvert < 2.5$ for muons $\lvert\eta\rvert < 2.47$ for electrons excluding $1.37 \leq \lvert\eta\rvert \leq 1.52$ in the ee-channel $\lvert\eta\rvert \leq 1.37$	$\lvert\eta\rvert < 2.5$ for muons $\lvert\eta\rvert < 2.47$ for electrons excluding $1.37 \leq \lvert\eta\rvert \leq 1.52$ in the ee-channel $\lvert\eta\rvert \leq 1.37$
Dilepton Mass, m_{ll}	$m_{ll} \geq 20$ GeV $\lvert m_{ee} - m_Z\rvert > 15$ GeV	$m_{ll} \geq 20$ GeV $\lvert m_{ee} - m_Z\rvert > 15$ GeV
Jets		
Number	At least 2 jets	At least 2 jets
Leading p_T	$p_T \geq 65$ GeV	$p_T \geq 65$ GeV
Sub-leading p_T	$p_T \geq 35$ GeV	$p_T \geq 35$ GeV
Dijet Mass, m_{jj}	$m_{jj} \geq 500$ GeV	$m_{jj} \geq 500$ GeV
B-jet	b-jet veto using the DL1r tagger	$n_{b_{\text{jet}}} = 1$
Dijet rapidity	$\lvert\Delta y_{jj}\rvert < 2$	**Leading jet is not a b-jet** **Dijet rapidity dropped**
Missing Transverse Energy		
E_T^{miss}	$E_T^{miss} \geq 30$ GeV	E_T^{miss} **selection dropped**

Table 17: The definitions of the low Δy_{jj} and top-fakes validation regions.

6 $W^{\pm}Zjj$ Background

The dominant background, contributing 22% of the total event yield [89, 88] in this experiment, is the $W^{\pm}Zjj$ process. This background can mimic the $l^{\pm}l^{\pm}jj + E_T^{\text{miss}}$ final state if the opposite-charge lepton associated with the Z boson decay is undetected. A lepton can be missed if the lepton falls outside of the geometric acceptance of the detector or is not successfully identified. In both the polarised and differential analysis, the background is constrained using a combination of MC simulations and data-driven methods. The two analyses follow a parallel strategy in modelling the background, however the MC sample has been updated in the polarisation measurement. The work presented in this chapter outlines the update to the modelling of this background.

The $W^{\pm}Zjj$ background receives two contributions coming from its electroweak and QCD-induced production modes. The QCD production mode is the dominant of the two in the experiment signal region. The two contributions can be separated in a gauge-invariant way and modelled using MC generators, similarly to $W^{\pm}W^{\pm}jj$ process. However, it has also been found that MC generators typically predict larger cross-sections than what is experimentally observed in data for the QCD-induced process. In order to model the $W^{\pm}Zjj$ background correctly, a normalisation factor is applied to the corresponding sample. The factor is constrained by data in a dedicated $W^{\pm}Zjj$ control region. Another known feature of MC generators is the mismodelling of the dijet invariant mass for QCD-induced Vjj and $VVjj$ processes [118]. A data-driven technique is applied to reweight the m_{jj} distribution to represent data better. The rate mismodelling and the shape reweighting will be discussed in detail in this chapter.

During the internal ATLAS review stage of the polarisation analysis, the board recommended updating the QCD-induced $W^{\pm}Zjj$ MC generator to Sherpa 2.2.12, which was expected to improve the m_{jj} modelling. The suggestion was made as the newer version of the Sherpa generator should improve some of the mismodelling in the MC predictions. The sample generated with the previous version, Sherpa 2.2.2, used in the differential analysis, is referred to as the differential $W^{\pm}Zjj$ QCD sample. The author integrated the new sample into the analysis and updated the m_{jj} reweighting and the theoretical uncertainties.

6.1 WZ Control Region

As an introduction, this chapter demonstrates the mismodelling of the $W^{\pm}Zjj$ QCD sample by comparing the MC predictions with data. The comparison is performed in a $W^{\pm}Zjj$ control region. This experiment uses the control region to constrain the $W^{\pm}Zjj$ QCD normalisation factor, further discussed in Chapter 10. The region is designed to be kinematically adjacent to the signal region and is dominated by $W^{\pm}Zjj$. The selection criteria are chosen to ensure that the kinematics of the $W^{\pm}Zjj$ events closely match those in the signal region. The control region is defined in Table 18. The major differences with the signal region are highlighted in bold.

The primary difference between the two regions is that the third-lepton veto requirement is dropped in the $W^{\pm}Zjj$ control region. In the signal region, events with a third additional baseline lepton surviving overlap removal are rejected. This condition is in-

verted in the $W^{\pm}Zjj$ control region, with the region requiring a third-lepton passing the signal selection, but with a loosened p_T requirement: $p_T > 15$ GeV. Furthermore, the additional criteria in the signal region to reject ee events with a dilepton mass close to Z mass, $|m_{ee} - m_Z| > 15$ GeV, is also dropped in the $W^{\pm}Zjj$ control region. The selection also applies a trilepton invariant mass cut of $m_{lll} > 106$ GeV to suppress the contributions of Z+jets and $Z\gamma$, and a slightly loosened m_{jj} selection.

	$W^{\pm}Z$ Control Region
Leptons	
Number	2 signal-leptons with equal electric charge 1 veto-lepton with opposite charge
Lepton-$\mathbf{p_T}$	$p_T > 27$ GeV veto-lepton $p_T > 15$ GeV
Lepton-η	$\|\eta\| < 2.5$ for muons $\|\eta\| < 2.47$ for electrons excluding $1.37 \leq \|\eta\| \leq 1.52$
Dilepton Mass, $\mathbf{m_{l^{\pm}l^{\pm}}}$ (same-charge pair)	$m_{l^{\pm}l^{\pm}} \geq 20$ GeV
Trilepton Mass, $\mathbf{m_{lll}}$	$m_{lll} \geq 106$ GeV
Jets	
Number	At least 2 jets
Leading p_T	$p_T \geq 65$ GeV
Sub-leading p_T	$p_T \geq 35$ GeV
Dijet Mass, $\mathbf{m_{jj}}$	$m_{jj} \geq 200$ GeV
B-veto	b-jet veto using the DL1r tagger
Dijet rapidity	$\|\Delta y_{jj}\| > 2$
Missing Transverse Energy	
E_T^{miss}	$E_T^{miss} \geq 30$ GeV

Table 18: The $W^{\pm}Z$ control region, defined for both the polarised and differential analyses. Cuts applied here are to signal objects passing the object selection. The differences with respect to the signal region are highlighted in bold.

6.2 Mismodelling of QCD-induced $W^{\pm}Zjj$

As discussed previously, this section illustrates the mismodelling of the $W^{\pm}Zjj$ QCD sample by comparing the MC prediction to experimental data in a control region dominated by $W^{\pm}Zjj$. The cross-sections and the m_{jj} distributions for QCD-induced VBS and VBF are known to be difficult to model accurately using MC generators in VBS or VBF phase space. Both kinds of mismodelling are problematic for the analysis. The former will misrepresent the event yields in the signal region, while the latter is used as

a discriminant to extract the signal yield. This section's focus is to explicitly illustrate both types of mismodelling within the MC.

6.2.1 $W^{\pm}Zjj$ QCD Cross-Section Mismodelling

Sample	Events
$W^{\pm}Zjj$ QCD	821.2 ± 2.6
$W^{\pm}Zjj$ EWK	141.4 ± 0.5
$W^{\pm}Zjj$ INT	2.34 ± 0.07
$ZZjj$	60.6 ± 0.4
Z + jets	27.7 ± 3.5
$t\bar{t}V$	30.7 ± 0.5
$t\bar{t}$	11.0 ± 0.7
$Z\gamma$	5.4 ± 2.6
Total MC Prediction	1100 ± 5
Data	832 ± 29

Table 19: The MC-predicted event yields in the $W^{\pm}Zjj$ control region compared to data. Table adapted from [89, 88].

Table 19 gives the MC-predicted event yields and the data yield in the $W^{\pm}Zjj$ control region. The total event yield predicted by the MC is 32% larger than that of the data, a significant overestimation. The dominant process in the $W^{\pm}Zjj$ control region is the QCD-induced $W^{\pm}Zjj$, which contributes 75% to the predicted MC event yield. The sub-dominant process is the EW-induced $W^{\pm}Zjj$ process. Given the size of the mismodelling and dominance of the QCD $W^{\pm}Zjj$ contribution, this process must contribute significantly, if not completely, to the overall mismodelling seen in the region. Figure 53 compares the data to the MC prediction for several kinematic variables.

6.2.2 Mismodelling of the $W^{\pm}Zjj$ QCD Dijet Mass Distribution

In addition to the rate mismodelling, the $W^{\pm}Zjj$ QCD sample also exhibits shape mismodelling in the m_{jj} distribution. Figure 53 shows that the m_{jj} distribution has a steeper slope in the data compared to the MC. To illustrate this more clearly, Fig. 54 shows a comparison of the data with the processes other than $W^{\pm}Zjj$ QCD subtracted from the data, divided by the QCD-induced $W^{\pm}Zjj$ MC. The division acts to normalise the prediction and highlight the difference in the shapes of the distributions. The red line is a line of best fit for the binned data. The line has a negative slope, which shows that the observed data has a steeper m_{jj} distribution than the predicted MC.

To correct the mismodelling, a data-driven reweighting technique is applied to the $W^{\pm}Zjj$ QCD sample. The reweighting function is derived by fitting the normalised data shown in Fig. 54 and applied as additional event weights to the sample MC. This procedure is applied in both the differential and polarisation analyses, albeit with different implementations due to the different samples. The reweighting is particularly important

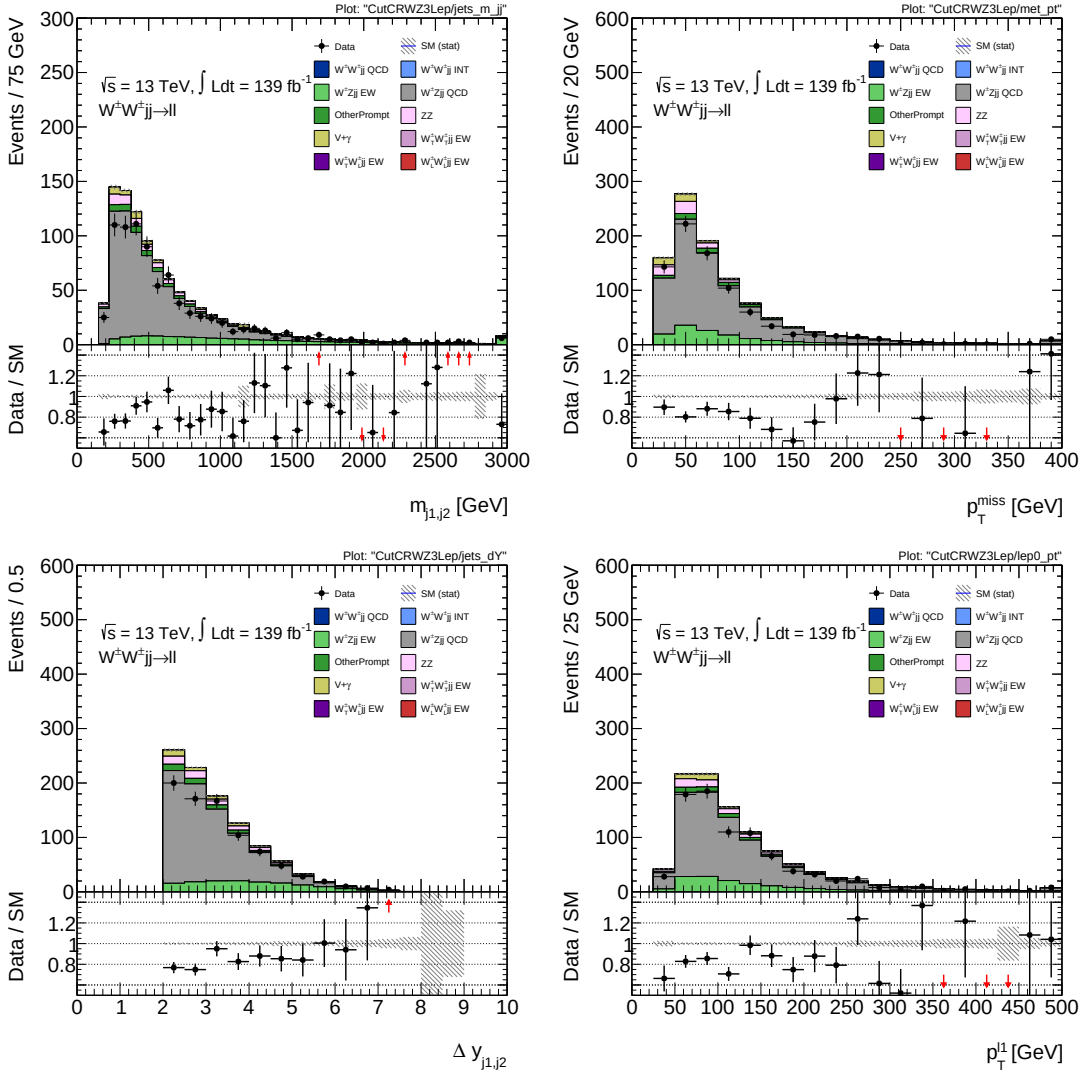


Figure 53: A comparison of the MC predicted m_{jj}, p_T^{miss}, Δy_{jj} and the leading leptonp_T distribution to data in the $W^{\pm}Zjj$ control region.

for the differential analysis as the m_{jj} distribution is used to extract the differential cross-section. For the polarisation analysis, the m_{jj} distribution is used as an input to multivariate classifiers and must likewise be modelled correctly.

6.3 Reweighting the $W^{\pm}Zjj$ QCD Dijet Mass Distribution

The focus of this section is to describe the derivation of the reweighting function applied to the $W^{\pm}Zjj$ QCD sample. The reweighting function is derived by performing a fit on the normalised variable shown in Fig. 54 and applied as event weights to the $W^{\pm}Zjj$ QCD MC. The fit is performed on a subset of the $W^{\pm}Zjj$ control region with suppressed EW-induced $W^{\pm}Zjj$ contributions. The reweighting is a normalised shape reweighting and preserves the event yields. The event yields are constrained separately by the normalisation factor applied to the QCD-induced $W^{\pm}Zjj$.

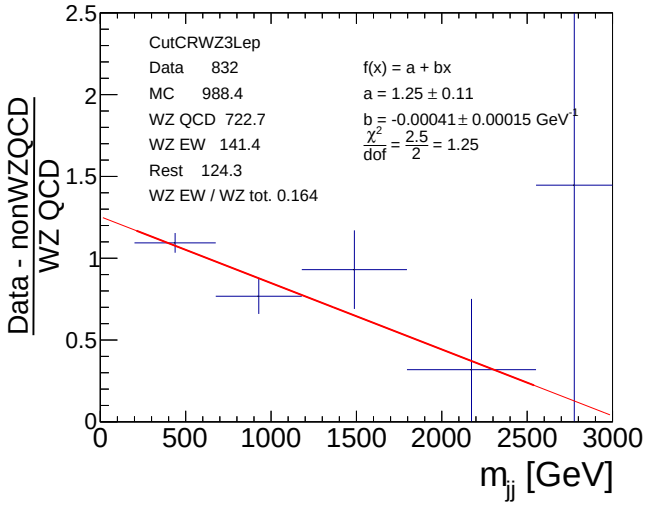


Figure 54: The ratio of the corrected data to the $W^{\pm}Zjj$ QCD MC as a function of m_{jj}. The corrected data corresponds to the observed yield minus the MC prediction of all processes other than $W^{\pm}Zjj$ QCD. The y-axis of the histogram is the data with the non-$W^{\pm}Zjj$ QCD MC subtracted, normalised to the $W^{\pm}Zjj$ QCD MC.

6.3.1 $W^{\pm}Zjj$ High-Centrality Control Region

Electroweak-induced $W^{\pm}Zjj$ production constitutes 15% of the total $W^{\pm}Zjj$ event yield in the $W^{\pm}Zjj$ control region. The reweighting is solely applied to the $W^{\pm}Zjj$ QCD sample. In order to exclude any effects in the m_{jj} distribution from potential mismodelling in the electroweak $W^{\pm}Zjj$, the reweighting function is derived in a region of the $W^{\pm}Zjj$ control region with suppressed electroweak contributions. This phase space is defined by the differential analysis to have additional selection criteria.

There are a number of variables that can be used to discriminate between the electroweak and QCD production modes. The electroweak production mode has no colour exchange between the two partons and is therefore expected to have no additional central jets ($n_{\text{jets, central}} = 0$), where central refers to the rapidity region spanned by the two selected jets (the 'tagging jets'). In addition, due to the t-channel gauge-boson propagators, the cross-section is maximal for low momentum exchanges between the two incoming partons. This means that the outgoing W and Z bosons are expected to lie centrally between the two tagging jets. The differential analysis tested several different variables corresponding to the centrality of the individual gauge bosons and diboson system, as well as the number of central jets, $n_{\text{jets, central}}$.

The control region was selected using the ratio of the $W^{\pm}Zjj$ electroweak to total yields, f_{EW}:

$$f_{EW} = \frac{N_{WZ,\text{EW}}}{N_{WZ,\text{tot}}}. \tag{115}$$

The distributions of the kinematic variables were also compared to the ratio $N_{WZ,\text{QCD}}/f_{EW}$. The region was choosen to minimise f_{EW} whilst maximising $N_{WZ,\text{QCD}}/f_{EW}$ to ensure that the electroweak contribution is minimal and the $N_{WZ,\text{QCD}}$ yield is maximal for the extraction of the reweighting function. The variable chosen is the diboson, WZ system

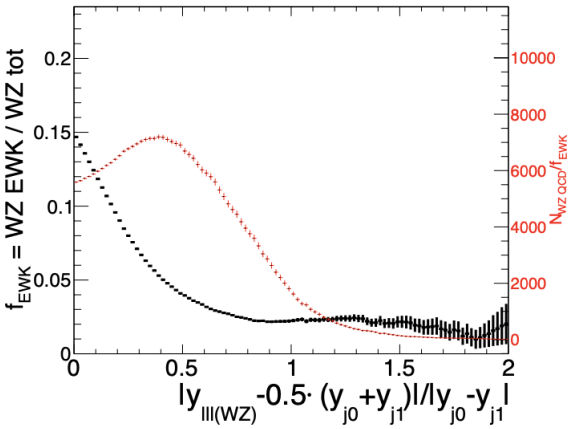


Figure 55: The ratios f_{EW} (defined in Eq. 115) and $N_{WZ,\mathrm{QCD}}/f_{EW}$ as a function of the WZ-system centrality (defined in Eq. 116). Both ratios are evaluated using MC. Figure taken from [89, 88].

centrality, defined as,

$$\left| \frac{y_{lll(WZ)} - 0.5(y_{j0} + y_{j1})}{y_{j0} - y_{j1}} \right|, \tag{116}$$

where $y_{lll(WZ)} = 0.5 \times (y_{lW} + y_{ll(Z)})$ is defined as the average of the rapidities of the visible products of the $W^{\pm}$ and Z bosons: $y_{ll(Z)}$, is the rapidity of the Z boson associated with two same-flavour opposite-charge leptons with the closest dilepton invariant mass to the Z mass and $y_{l(W)}$ is the rapidity of the lepton associated with the W boson. Likewise, $y_{j0(1)}$ corresponds to the rapidity of the (sub)-leading tagging jet. The ratios f_{EW} and $N_{WZ,\mathrm{QCD}}/f_{EW}$ are plotted in Fig. 55. An additional control region uses the latter variable, with a selection requiring a value greater than 0.5. This defines the $W^{\pm}Zjj$ High Centrality control region. This threshold value is chosen by examining Fig. 55 which shows a peak at approximately 0.5 in this ratio.

6.3.2 Reweighting Function Derivation

The reweighting function is calculated by fitting an exponential function to the binned distribution of the ratio of the data minus the non-$W^{\pm}Zjj$-QCD MC to the $W^{\pm}Zjj$ QCD MC, as shown in Fig. 56. The fit performed is a χ^2 fit and is implemented using the ROOT library [119]. The χ^2 fit is a maximum likelihood fit in the presence of Gaussian uncertainties, which assumes that each bin in the fitted histogram corresponds to an independent measurement of a Gaussian random variable. The likelihood function $\mathcal{L}$ is defined as the product of a set of Gaussian distributions denoted by $\mathcal{G}(x|\mu,\sigma)$ with mean μ and standard deviation σ, such that,

$$\mathcal{L}(y,x,\sigma|f(x|a,b)) = \prod_{i}^{\text{bins}} \mathcal{G}(y_i|\mu,\sigma) = \prod_{i}^{\text{bins}} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(y_i - f(x_i|a,b))^2}{2\sigma_i^2}}, \tag{117}$$

where $f(x|a,b)$ is the fitted function with parameters a, b, and x, y, σ are input vectors corresponding to the histogram central value, content and uncertainty for each bin i

respectively. Taking the negative log-likelihood of Eq. 117 simplifies the expression such that,

$$\begin{aligned} -2\log\Big[\mathcal{L}(y,x,\sigma|f(x|a,b))\Big] &= -2\log\Big[\prod_{i}^{\text{bins}}\mathcal{G}(y_i|\mu,\sigma)\Big] \\ &= \sum_{i}^{\text{bins}} -2\log\big(\frac{1}{\sqrt{2\pi\sigma_i^2}}\big) + \frac{(y_i - f(x_i|a,b))^2}{\sigma_i^2} \\ &= \sum_{i}^{\text{bins}} -2\log\big(\frac{1}{\sqrt{2\pi\sigma_i^2}}\big) + \chi^2(y_i, x_i, \sigma_i, f), \end{aligned} \tag{118}$$

which is minimised when minimising the χ^2 value. The minimisation is calculated using the Minuit algorithm [120] to find the best-fit parameters and to estimate their uncertainty. The χ^2 fit is performed in the interval $200 < m_{jj} < 2500$ GeV. The statistical uncertainty is the limiting factor to the fit. The interval and binning are optimised to reduce the statistical uncertainty whilst maximising the number of data points. The chosen fit function is an exponential function, defined by,

$$f_{rew}(x) = ae^{-xb}, \tag{119}$$

where a and b are the parameters estimated by the likelihood fit. Over the course of testing, a variety of different fit functions were trialled, and it was found that the exponential function gives the lowest χ^2/dof, where dof represents the number of degrees of freedom.

The final fit function is shown in Fig. 56. In the figure, the blue crosses depict the bin values, with the bin widths and uncertainty indicated by the horizontal/vertical lines, respectively. The red line shows the fitted exponential function for the best fit parameters, $a = 1.24 \pm 0.22$ and $b = 0.00038 \pm 0.00030$, and the outer band shows the statistical uncertainty on the fit (described in the next section). The figure also shows the event yields in the $W^{\pm}Zjj$ High Centrality control region. As aforesaid, this region is highly dominated by $W^{\pm}Zjj$ QCD process.

6.3.3 Reweighting Function Statistical Uncertainty

The primary source of uncertainty for each of the fit function parameters a, b originates from the limited event yield in each bin of the fitted histogram. The statistical uncertainty is estimated by propagating the error on each of the fit function parameters a and b. The variance of the function of random variable X can be approximated with a Taylor expansion as,

$$\text{Var}(f(X)) \approx f'(\text{E}[X])^2\text{Var}(X), \tag{120}$$

where $\text{E}[X]$ is the expected value of X. The variance of the product of two random variables, X and Y is likewise given by,

$$\text{Var}(XY) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y). \tag{121}$$

Assuming now that the parameters a, b are Gaussian random variables with uncertainties given by their standard deviation σ_a, σ_b respectively, then the statistical uncertainty on

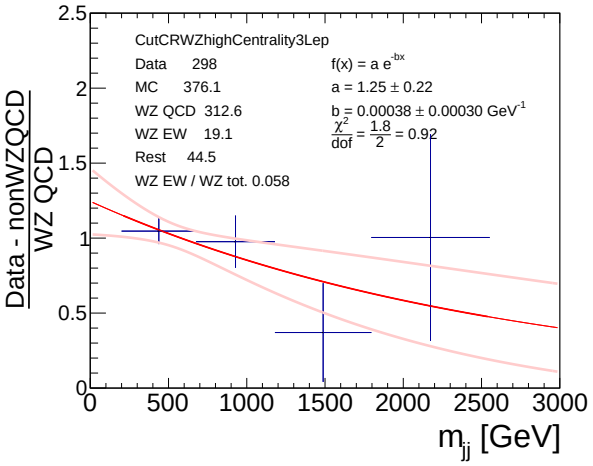


Figure 56: The fitted exponential function in the $W^{\pm}Zjj$ High Centrality control region. The binned histogram given by the blue lines shows the data minus the non-$W^{\pm}Zjj$-QCD MC normalised to the $W^{\pm}Zjj$ QCD MC. The horizontal lines indicate the bin widths and the vertical lines show the statistical uncertainties. The red line shows the fitted exponential function and the pink band shows the statistical uncertainty, found by propagating the uncertainty of the fit parameters, as in Eq. 122.

the reweighting function can be taken as the square root of the variance:

$$\begin{aligned}\sigma_f^2 = \mathrm{Var}(ae^{-bx}) &= \sigma_a^2\left(\frac{\partial f}{\partial a}\right)^2 + \sigma_b^2\left(\frac{\partial f}{\partial b}\right)^2 + 2\sigma_{ab}\left(\frac{\partial f}{\partial a}\right)\left(\frac{\partial f}{\partial b}\right) \\ &= e^{-2xb}[\sigma_a^2 + a^2b^2\sigma_b^2 - 2ab\sigma_a\sigma_b\rho_{ab}],\end{aligned} \tag{122}$$

where σ_{ab} is the covariance, given by $\sigma_{ab} = \sigma_a\sigma_b\rho_{ab}$, where ρ_{ab} is the correlation between the two parameters. The statistical uncertainty given by Eq. 122 is plotted in Fig. 56. The uncertainty band clearly does not preserve the normalisation of the reweighting function. In this current form, this is potentially problematic. The m_{jj} distribution is used to constrain the normalisation factor applied to the $W^{\pm}Zjj$ QCD MC. The statistical uncertainty on the reweighting function should not affect the uncertainty on the event yields, as this is only a normalised shape correction. By not preserving the normalisation, this uncertainty could pull the $W^{\pm}Zjj$ QCD normalisation factor and potentially bias the signal extraction. Therefore, to try to preserve the normalisation, the statistical uncertainty should also be in the form of an exponential function. Fortunately, the correlation between the two parameters is approximately equal to one ($\rho_{ab} = 0.91 \approx 1$). By making the approximation that the parameters a and b are correlated, the uncertainty on the reweighting can be expressed as,

$$\begin{aligned}\sigma_f &\approx e^{-xb}\sqrt{[\sigma_a^2 + a^2b^2\sigma_b^2 - 2ab\sigma_a\sigma_b]} \\ &= e^{-xb}[\sigma_a + ab\sigma_b],\end{aligned} \tag{123}$$

or equivalently, this is equal to re-fitting one of the parameters a or b for a fixed value of the up or down variation of the other. With this approximation, the uncertainty is also

now in the form of an exponential. The implementation of the approximation following Eq. 123 is shown in Fig. 57. As an estimate, this is a reasonable approximation of the statistical uncertainty on the reweighting function and, moreover, will better preserve the normalisation of the shape correction. The uncertainty will still introduce a small effect on the event yields, but this should not be significant enough to affect the constraint on the $W^{\pm}Zjj$ QCD normalisation.

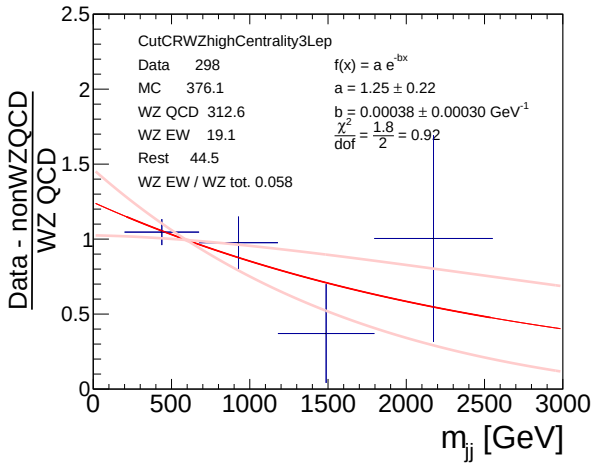


Figure 57: The fitted exponential function in the $W^{\pm}Zjj$ High Centrality control region, as in Fig. 56 but now with uncertainty bands as described by Eq. 123.

6.3.4 Reweighting Function Theory Uncertainty

In addition to the statistical uncertainty, the theory uncertainties in the $W^{\pm}Zjj$ QCD MC prediction need to be considered. As in the previous section, these uncertainties should preserve the normalisation. Uncertainties due to missing higher-order terms are estimated by varying the renormalisation and factorisation scales. Analogously, the uncertainties due to the PDF and the α_s choice are also considered by varying the PDF sets and using different values of α_s. Both types of theory uncertainty have a defined prescription as to how to propagate the variations, described in Chapter 10.

The magnitude of the theory uncertainties is much smaller than that of the statistical uncertainity. For this reason, the theory uncertainties are estimated by taking the envelope of each source respectively. This does not follow the recommended prescription in combining the variations for the PDF and α_s variations; however, due to the size of the uncertainties, this is a reasonable approximation and maintains the exponential form in order to preserve the normalisation.

6.3.5 The Reweighting Function

To conclude, the final set of uncertainties on the reweighting function, statistical and theoretical, are shown in Fig. 58. As discussed previously, the biggest uncertainty is the statistical uncertainty. Of the theory uncertainties, the scale uncertainties are slightly lar-

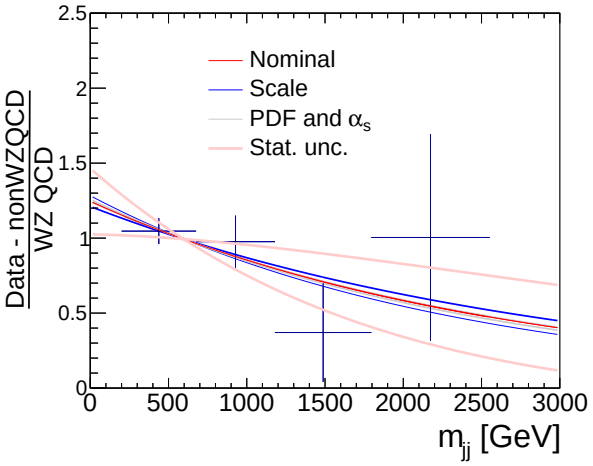


Figure 58: The nominal exponential fitted function and the envelope of the theory variations of the $W^{\pm}Zjj$ QCD MC

ger than the PDF uncertainties, but both are small relative to the statistical uncertainty. The total event yields in $W^{\pm}Zjj$ control control regions are shown in Table 20.

The table shows differences of up to 1% in the event yield from changing the reweighting factor for each theory uncertainty. Given these small changes in yields, the uncertainty in the reweighting factor can be factorised from the normalisation uncertainty.

Variation	WZ CR
Nominal (M_{jj}-reweighted)	720.5 ± 3.0
MjjRewPDFDown	720.5 ± 3.0
MjjRewPDFUp	720.8 ± 3.0
MjjRewScaleDown	722.2 ± 3.0
MjjRewScaleUp	719.0 ± 2.9
MjjRewStatDown	712.0 ± 2.9
MjjRewStatUp	727.9 ± 3.0
NoMjjRew	727.8 ± 3.0

Table 20: The total event yields in the $W^{\pm}Zjj$ control region for the up/down variations for each of the uncertainties, and without reweighting.

To illustrate the effect of the reweighting, a template profile likelihood stat-only fit is performed akin to the signal extraction fit discussed in Chapter 10. The fit is configured with two regions, the signal region and $W^{\pm}Zjj$ control region. The two regions constrain the normalisations of QCD-induced $W^{\pm}Zjj$ production. The input to the signal region is Asimov data (see Chapter 10), while the $W^{\pm}Zjj$ control region is fit to data. Figure 59 shows the m_{jj} distribution without and with (on the left/right respectively) the reweighting post-fit in the $W^{\pm}Zjj$ control region. From the figure, it is clear that the reweighting improves the shape of the m_{jj} distribution in comparison to the data.

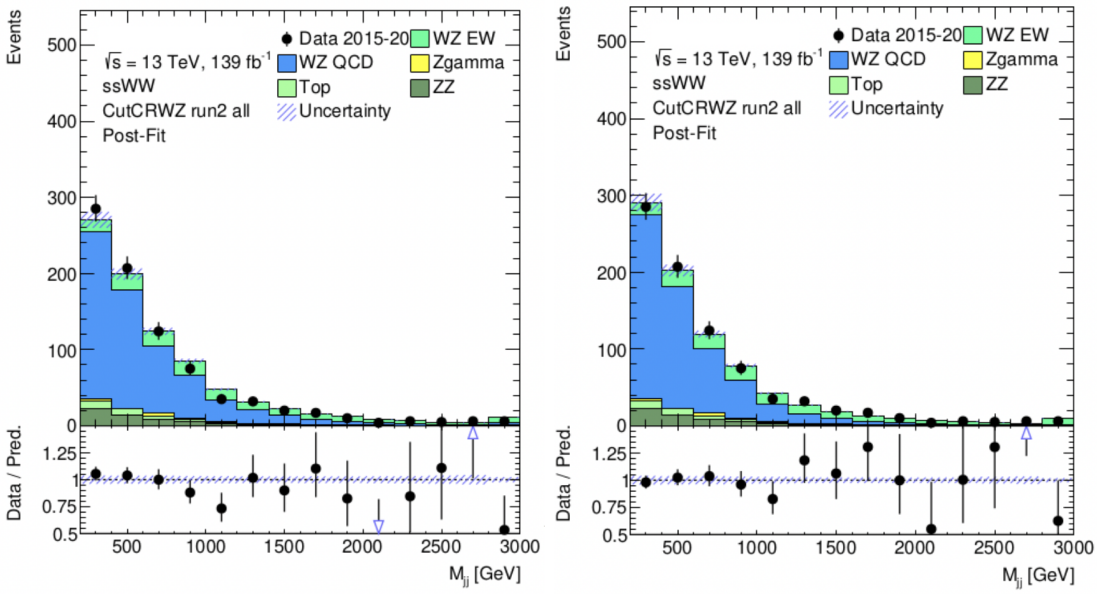


Figure 59: A comparison of the fitted $W^{\pm}Zjj$ control region for the fit performed without (left) and with (right) the reweighting function applied to the $W^{\pm}Zjj$ QCD MC.

7 Polarisation Classifiers

The measurement of polarised electroweak $W^{\pm}W^{\pm}jj$ production is the primary focus of this experiment, with an emphasis on longitudinal polarisation. This chapter and the subsequent chapter present the work investigating the application of a combination of multivariate analysis techniques (MVAs). These techniques are designed to isolate each polarisation fraction individually and from the analysis backgrounds to construct regions of data with a high statistical sensitivity to polarisation. This chapter begins by discussing the challenges of a polarisation measurement to motivate the multivariate methods. Subsequent sections will then introduce the machine learning concepts and nomenclature applied in the presented work. The remaining sections outline the training data and procedures, an analysis of the input features, and the results and validation of the multivariate models. This chapter focuses on the polarisation discrimination aspect of EW-induced $W^{\pm}W^{\pm}jj$ production. The subsequent chapter follows a similar structure but presents the work discriminating the inclusive EW $W^{\pm}W^{\pm}jj$ signal from the analysis backgrounds.

7.1 Motivation

Identifying the polarisation of gauge bosons in a VBS process is challenging. In the leptonic decay channel, polarisation manifests through the angular distributions of the leptons. The first challenge arises due to the process consisting of multiple polarised gauge bosons. For individual measurements of polarisation, the leptons must be individually distinguishable. However, due to the same-charge requirement, same-flavour decays are indistinguishable in this experiment. An analysis of only opposite flavour channels is possible, but this reduces the event yield by half. Since the uncertainty on this measurement is dominantly statistical, this is currently not a feasible option. For this reason, this measurement focuses on the combined gauge boson polarisation modes, with the primary objective of identifying processes with only longitudinally polarised $W^{\pm}$ bosons.

For electroweak $W^{\pm}W^{\pm}jj$ production, there are three possible polarisation configurations: longitudinal-longitudinal (LL), longitudinal-transverse (LT) and transverse-transverse (TT). The leptons are indistinguishable, so the transverse-longitudinal (TL) fraction is experimentally identical to LT. The purely longitudinal scattering mode has the highest interest for experimental study due to its association with the EWSB mechanism. However, this mode only accounts for approximately 9% of the total electroweak $W^{\pm}W^{\pm}jj$ event yield. Scattering with one longitudinal gauge boson constitutes a larger proportion of the total electroweak yield, but polarisation effects are diluted as the kinematic distributions contain a mix of polarisations. To motivate the application of MVAs, this section highlights polarisation effects by comparing the kinematic distributions of the different polarisation configurations.

Figure 60 shows the normalised distributions of the dijet $\Delta\phi_{jj}$ and ΔY_{jj} distributions, which effectively represent the momentum transfer in the scattering event. The figure shows that longitudinal events are predicted to have larger dijet $\Delta\phi_{jj}$ and $\Delta\eta_{jj}$. The differences in the jet deflection suggest different momentum exchanges in the η-ϕ plane owing to polarisation of the gauge bosons.

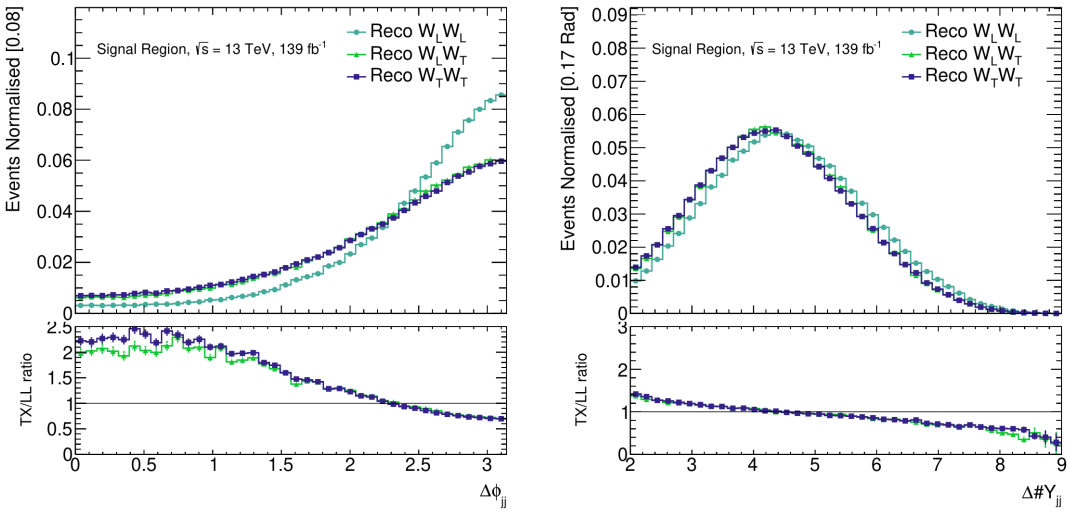


Figure 60: The normalised $\Delta\phi_{jj}$ and ΔY_{jj} distributions, shown on the left and right respectively, for the polarised electroweak $W^{\pm}W^{\pm}jj$ MC samples in the WW CoM.

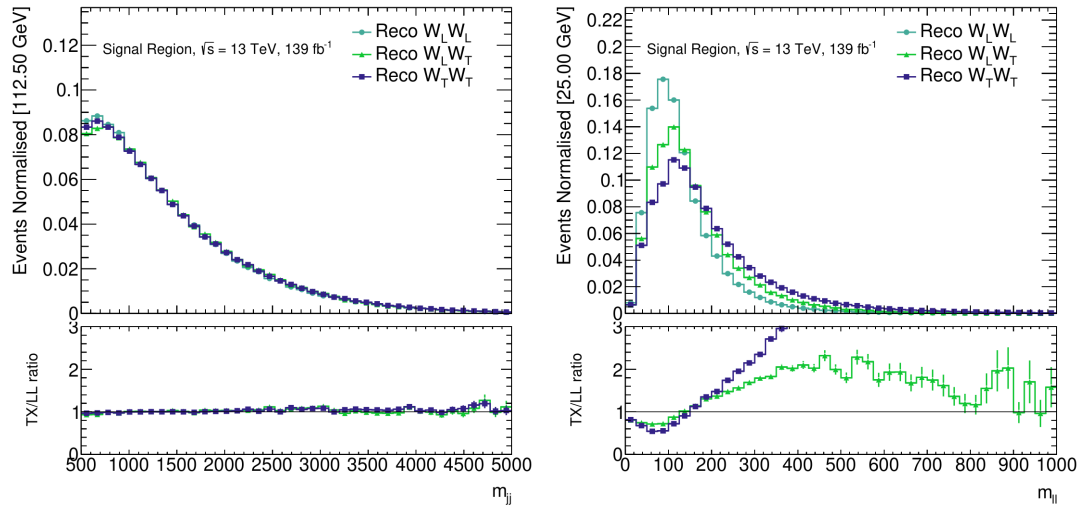


Figure 61: The normalised m_{jj} and m_{ll} distributions, shown on the left and right respectively, for the polarised electroweak $W^{\pm}W^{\pm}jj$ MC samples in the WW CoM.

Figure 61 shows the normalised distributions of the dijet m_{jj} and dilepton m_{ll} invariant mass. The dijet invariant mass distribution shows almost no differences as a result of polarisation. However, the dilepton mass distribution shows the dilepton system is observed with larger transverse momentum for transverse $W^{\pm}$ boson production. Polarisation effects are also present in the dilepton $\Delta\phi_{ll}$ and $\Delta\eta_{ll}$ distributions, shown in Fig. 62.

Further polarisation effects can be seen in the η-ϕ plane for both the leading and sub-leading leptons in p_T, as shown in Fig. 64. The leading and sub-leading distributions show leptons associated with longitudinal gauge bosons are predicted to be measured more in the detector's forward and backward regions. Moreover, the leading and sub-

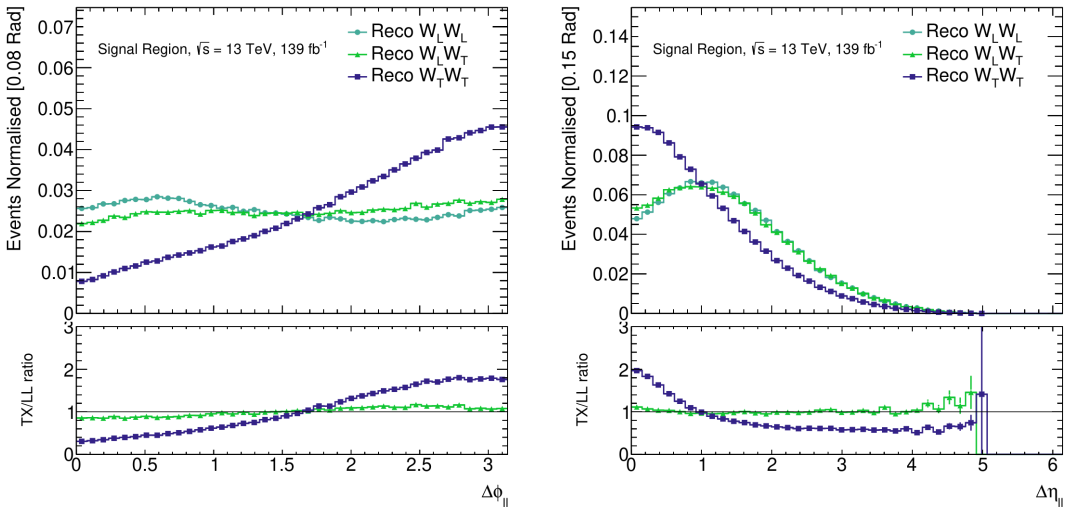


Figure 62: The normalised $\Delta\phi_{ll}$ and $\Delta\eta_{ll}$ distributions, shown on the left and right respectively, for the polarised electroweak $W^{\pm}W^{\pm}jj$ MC samples in the WW CoM.

leading leptons are predicted to be typically measured with a smaller p_T for longitudinal scattering, compared to transverse scattering.

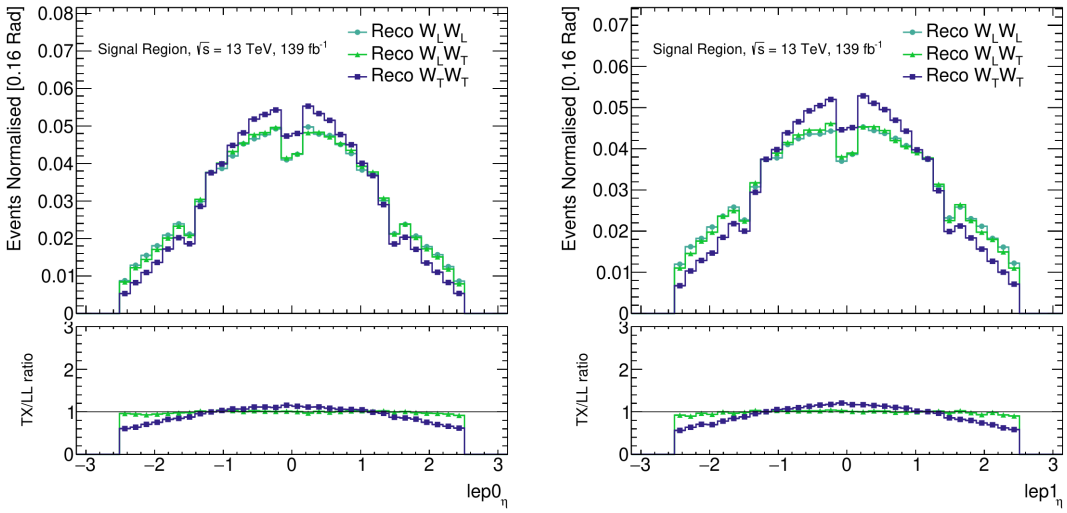


Figure 63: The normalised distributions of the leading and sub-leading leptons η, shown on the left and right respectively, for the polarised electroweak $W^{\pm}W^{\pm}jj$ MC samples in the WW CoM.

While there are kinematic differences between polarisation configurations, the challenge in measuring longitudinal polarisation lies in the large component of transverse scattering: TT is predicted to constitute 62% of the total $W^{\pm}W^{\pm}jj$ electroweak event yield. The transverse scattering component can be considered a background for experimental measurements of longitudinal polarisation. Furthermore, the predicted event yields of LT and LL are smaller than the number of background $W^{\pm}Zjj$ events in the

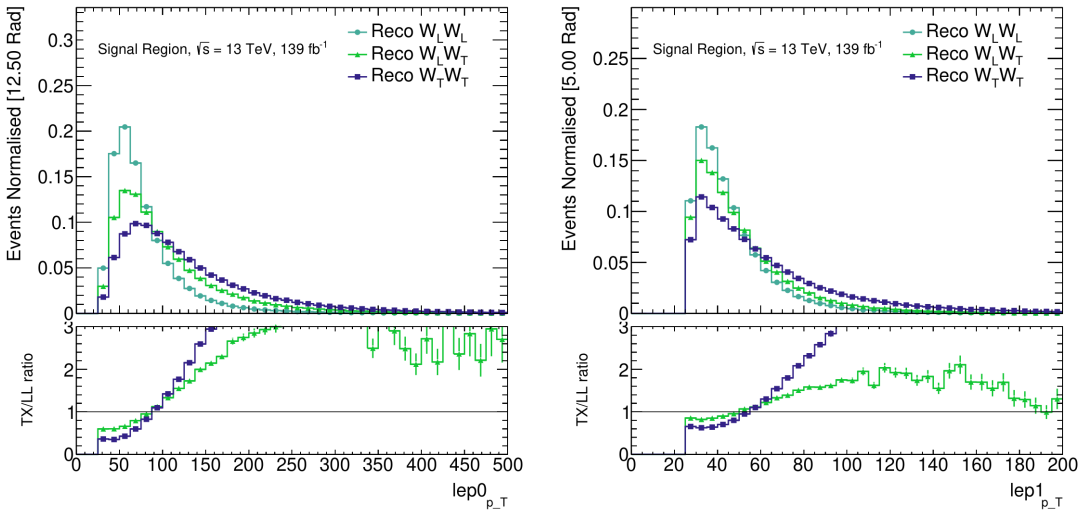


Figure 64: The normalised distributions of the leading and sub-leading leptons p_T, shown on the left and right respectively, for the polarised electroweak $W^{\pm}W^{\pm}jj$ MC samples in the WW CoM.

analysis signal region.

The contribution from the transverse scattering component or other backgrounds could be reduced by adjusting the signal region. As an example, lowering the p_T threshold would proportionally increase the number of longitudinally polarised events. However, this is not feasible, as this would lead to much larger background contributions from fake leptons and would not improve the measurement. Any selection criteria to reduce the TT contribution would also reduce the available statistics for LL and LT. This is also not practical for a measurement already dominated by statistical uncertainty.

To address the low event yields of longitudinally polarised gauge bosons, multivariate analysis techniques are used to exploit the discussed differences in the kinematics associated with polarisation. There are two main challenges to measuring longitudinal polarisation: the dominance of the transverse scattering mode and the larger event yields associated with the analysis backgrounds. The work demonstrated in this thesis will illustrate the implementation of a combination of classification algorithms, addressing each of these challenges. One classifier is implemented to separate the inclusive electroweak $W^{\pm}W^{\pm}jj$ signal from the backgrounds, and a second classifier is used to isolate the polarisation fractions.

7.2 Machine Learning

The MVAs studied and applied in this work are formally subsets of the field of machine learning. The nomenclature and training procedures outlined in this and the subsequent chapters are concepts unique to this field. For this reason, this section provides a global overview of machine learning concepts. This introduces the necessary theory to follow the work presented in this chapter and follows Refs. [121, 122]. The major work presented here applies various machine learning methodologies to polarisation classification. An

extensive overview is therefore not warranted, and the reader is referred to the provided citations for a detailed description of machine learning and deep neural networks.

7.2.1 Introduction

Machine Learning (ML) is a broad subset of applied statistics that uses data to numerically estimate complex mathematical functions to solve many different tasks and problems. The "machine" refers to the use of computational numerical methods, and "learning" refers to using an example dataset to find a generalised function that can be applied to solve a given problem. A ML algorithm is a set of rules and processes applied to estimate a function from an example dataset. The input variables of the ML function are often referred to as features, which are a selection of variables that the function has been trained to process to give a desired output. The overall machine learning function is also often referred to as a model.

In this thesis, ML is used to solve two types of problems: classification and regression. Classification is the task of finding a function, $f(\bar{x}) : \mathcal{R}^n \to \{1, \ldots, k\}$, to map a set of features $\bar{x} \in \mathcal{R}^n$ to a numerically encoded output vector $y = \{1, \ldots, k\}$ of k classifiable objects. A common use of classification is for image recognition. For these types of classification problems, the input features are a vector of pixels, and the output vector encodes whether the image contains a given set of objects. Similarly, regression refers to the more general task of finding a function $f(\bar{x}) : \mathcal{R}^n \to \mathcal{R}^n$, to map a set of features $\bar{x} \in \mathcal{R}^n$ to a set of output variables $\bar{y} \in \mathcal{R}^n$. A regression is performed in this work to map measurable kinematic input variables by the experiment to the $\cos\theta$ distribution associated with each polarised gauge boson of a VBS event.

The general act of learning refers to the algorithm's ability to adjust the predictions of the ML function to match a desired output. This is typically achieved by defining the problem with a loss function that computes the prediction error. The algorithm iteratively adjusts the loss function using the data to minimise the error. There are several different types of ML algorithms, but the focus of this thesis is on supervised learning. Supervised learning refers to learning with an example dataset where each data point is associated with a target output value. A supervised learning algorithm uses the loss function to compute the error in the prediction of the ML function to the target output value for a given data point. The performance of the ML algorithm is also additionally quantified with metrics. The metric is problem-dependent but is typically a designated function used to evaluate the accuracy of a model.

The major challenge of any ML algorithm is that the model should also perform well on previously unseen data, referred to as generalisation. The loss function and metric that quantify the performance of the model should indicate that the model performs equally well on unobserved data as on the data used for learning. To ensure the model performs well on unseen data, the available dataset can be partitioned into a training and a validation set. Over the course of learning (also referred to as training), the loss function is simultaneously computed for the validation set. The two primary goals of any machine learning algorithm are to reduce the training error and to ensure the difference in the training and validation error remains small.

The case where the algorithm fails to reduce the loss on the training data is referred

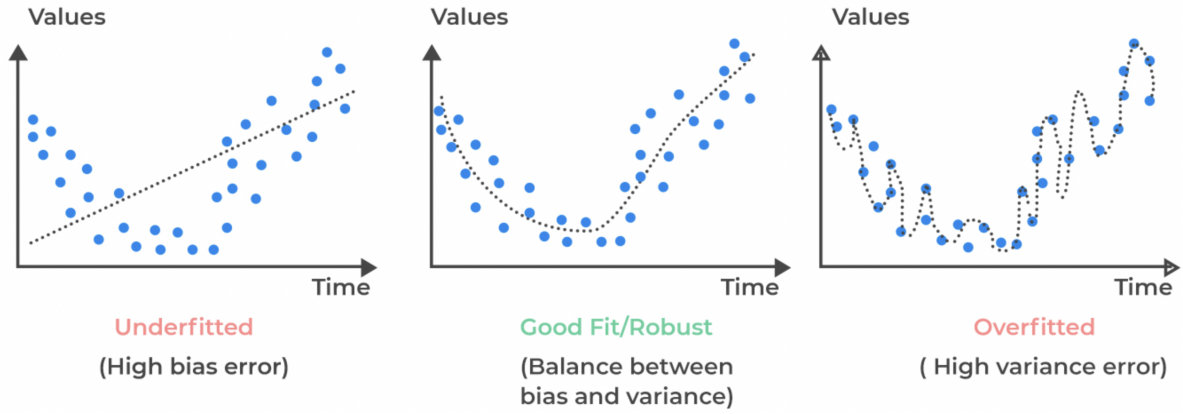


Figure 65: An illustration of underfitting and overfitting of example training-set data. The fitted function is indicated by the dashed line. Figure taken from [123].

to as underfitting, implied on the left of Fig. 65. The case where the testing error is significantly larger than the training error is referred to as overfitting, implied by the right plot of Fig. 65. Many factors can cause overfitting or underfitting, not least the algorithm used to train the model itself. One of the significant factors is referred to as the capacity of the model. The capacity is a broad term used to describe the available degrees of freedom in the fitted function. As an example, consider an algorithm which is trying to learn to regress a function on a given dataset. A polynomial of order k could be used to fit the data. A model of order $k = 1$ will only be able to learn a linear fit for the data and will underfit if the trained dataset is not linear. The converse is a polynomial with many orders will be far more likely to overfit and fail to generalise. Capacity is also sometimes referred to as complexity.

For each ML algorithm, several free parameters dictate the network's capacity and protocols for training. These parameters are referred to as hyperparameters. During training, trialling an extensive set of hyperparameters is common to find the most suitable model. As discussed, when training a model, to avoid overfitting, the error is also computed on an unseen dataset[5]. When trialling a model over many different sets of hyperparameters, the best models are often chosen based on the performance of the networks using this unseen dataset. However, by trialling, the unseen data biases the set of parameters. The dataset is, therefore, often partitioned into three separate datasets: training, testing and validation. The validation dataset monitors the loss on an unseen dataset and compares the algorithm's performance for the different hyperparameters. The testing set is used to give a completely unbiased estimate of the model's performance on unseen data.

These concepts can, in practice, be applied to many different ML algorithms. However, the focus of this chapter is on deep learning. The difficulty of machine learning exponentially grows with the number of dimensions in the data, referred to as the curse of dimensionality [121]. Classical ML models often fail to adapt to higher dimensions suitably. Deep learning models, however, do not suffer from the same issues and can generalise to higher dimensional data. These models are generally very flexible and can

[5]Unseen data refers to data that has not been used during training. The training process is blinded to this dataset.

be applied to various problems, including classification and regression, making them very suitable for the work of this thesis.

7.2.2 Deep Learning

Deep learning [121, 122] is a subset of ML that uses multi-layered neural networks, also commonly referred to as Deep Neural Networks (DNN) and feed-forward neural networks. Neural networks are computational graphs which are used as the approximation function. An illustration of an example DNN is shown in Fig. 66. The structure is loosely inspired by neurons in the brain. Each layer consists of a set of neurons, also referred to as nodes. The neurons have an input, which is the output of the previous layer, and likewise provide an output.

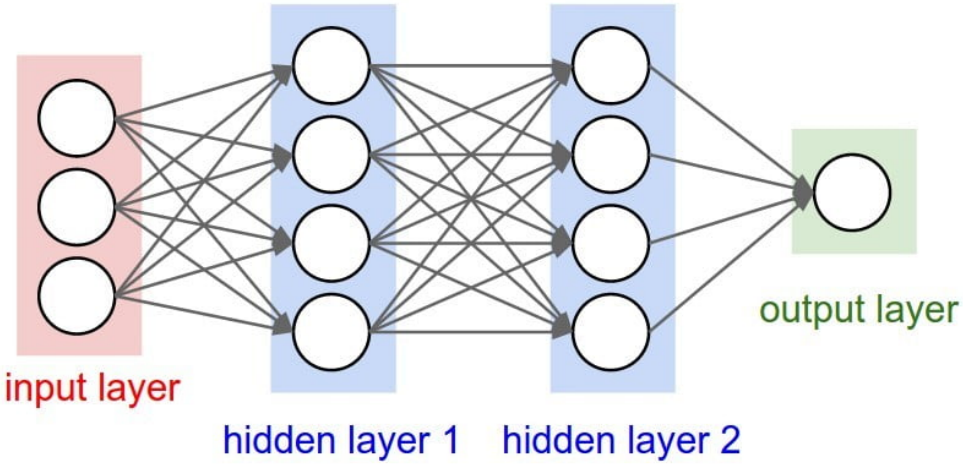


Figure 66: An example of a two-layered deep neural network. Figure taken from [124].

The DNN starts with an input layer representing the function's input features, $\bar{x}$. The input features pass through the network to a final output layer representing the model's output, y. The full network can be expressed as $y = f(\bar{x})$. Feed-forward describes this kind of network structure as the information moves in one direction, from the input to the output. Mathematically, a DNN is implemented as a composition of functions. The DNN in Fig. 66 would be expressed as $y = f^{\text{out}}(f^{\text{h2}}(f^{\text{h1}}(\bar{x})))$ where $f^{\text{out,h2,h1}}$ present the functions denoting the output and hidden layers 2 and 1.

The number of layers is referred to as the depth of the composition chain. Each hidden layer comprises a number of nodes represented by a vector. The number of nodes in a layer is sometimes referred to as the width of the network. A DNN's capacity is highly dependent on the width and depth. Depending on the task, a network may be more suitable with a higher width or depth. The mathematical representation of a hidden layer is formally denoted by $\phi(\bar{x} \cdot \bar{w} + \bar{b})$ where $\bar{x} \in \mathcal{R}^n$ is the input vector to the layer, $\bar{w} \in \mathcal{R}^{n \times m}$ is a matrix of weights, $\bar{b} \in \mathcal{R}^m$ is a bias vector and ϕ is referred to as the activation function of the layer.

The activation function is one of the crucial components of a DNN. A non-linear activation function introduces the capability of the network to model non-linear data. It

has been shown that DNNs consisting of multiple hidden layers with non-linear activation functions are universal approximators to any n dimensional continuous function [125]. This property makes them highly suitable for ML. The non-linear nature does, however, introduce additional complications. Unlike other simpler ML models that can minimise the error by solving a set of linear equations, a deep neural network must learn through iterative gradient-based optimisation. With this approach, there is no guarantee of convergence, and the training itself is highly sensitive to the initialisation of the parameters.

Gradient-based optimisation is a method of minimising the loss function with respect to the predictions of the network. The method is applied iteratively to reduce the error and traverse to a local minima of the loss function. At each iteration, the gradient is calculated for each individual weight and the weight is adjusted by subtracting the negative of the gradient times a constant scale factor, referred to as the learning rate. By subtracting the gradient, the weight is adjusted in the direction that decreases the fastest. The gradient is typically calculated using backpropagation, which utilises the chain rule starting from the output and moving backwards through the network. The reader is referred to [121, 122] for a detailed description. The end goal of any deep learning algorithm is to adjust the weights in each layer such that for a given set of inputs, the network makes an accurate classification.

7.3 Polarisation Classifiers

This analysis uses MVAs to separate longitudinally polarised events from the $W^{\pm}W^{\pm}jj$ backgrounds and from $W^{\pm}W^{\pm}jj$ events with only transverse gauge bosons. The purely transverse mode can be considered as the biggest background in this experiment. The isolation of longitudinal signal events is implemented through the use of classification algorithms. There are multiple potential classification configurations that could achieve the experiment's objectives. This thesis will present the training of a combination of two classification algorithms. The first classifier separates the inclusive $W^{\pm}W^{\pm}jj$ electroweak signal from the associated backgrounds, with the second separating the individual polarisation fractions. In this thesis, the former is referred to as the electroweak classifier and the latter as the polarisation classifier.

As mentioned, other classification configurations could potentially achieve similar results. For example, a single MVA could be trained to isolate scattering with longitudinal gauge bosons from TT and the backgrounds. However, this approach is expected to be less sensitive to polarisation. This is because the kinematic variables which provide the best discrimination between the inclusive electroweak $W^{\pm}W^{\pm}jj$ signal and the analysis backgrounds are generally very different to the kinematics discriminating polarisation. An approach with a single classifier may be less sensitive as the algorithm must simultaneously differentiate between two very different classification problems. A combination of classifiers is chosen to maximise discrimination with respect to polarisation. The caveat to this approach is that the polarisation classifier is trained only on polarised signal samples. This essentially means that the classifier is trained assuming a very high purity $W^{\pm}W^{\pm}jj$ electroweak signal. As the polarisation classifier is not trained on the analysis backgrounds, the model's performance will be highly dependent on the effectiveness of the electroweak classifier in isolating the electroweak signal.

7.3.1 Network Targets

The classifiers documented in this section are multi-class classifiers that predict the probability a given event is one of the three polarisation fractions: LL, LT, or TT. This is an extension of binary classification[6]. The multi-class approach aims to ensure the model is sensitive to each of the polarisation fractions. The additional benefit of this approach is the same model can be used to extract the cross-sections for both LL and LX. The ATLAS polarisation analysis uses a different configuration and applies two separate binary classification algorithms: one for LL and one for LX. The ATLAS approach is expected to be slightly more sensitive but requires training twice the number of models. As discussed in Chapter 4, polarisation is not a Lorentz invariant quantity. The cross-sections for polarised production are measured in both the WW and *pp* CoM frames. A polarisation classifier is also trained for each reference frame. In the final measurement, the same electroweak classifier is used for measurements in both reference frames, as the combined samples are Lorentz invariant.

7.3.2 Training Procedure

The training procedure is split into two stages: a first and second pass. The first pass is performed with the complete set of input variables. DNNs are highly effective at identifying features in data relevant to a given classification problem. However, they have the downside of being essentially a "black box". Prior to training, it is unclear which input features beyond physics intuition are beneficial to the classification algorithm. Moreover, even after training, the relationships and correlations between the input variables and the final classification are unknown. Whilst being highly effective in adapting to a given classification problem, the DNNs have the downside of not giving much insight into the physics itself. For this reason, this section details the implementation of an input feature selection procedure referred to as INVASE. This input feature selection model will subsequently use the best-performing networks trained in the first pass to give insight into the most important features associated with polarisation.

The input feature selection procedure selects a subset of the most relevant inputs. This subset is implemented as the nominal set of input variables, and the final networks used in this measurement with this subset are trained in the second pass. A smaller set of input features is preferable. A larger number of input variables increases the complexity of the network, which can lead to overfitting. Only the relevant variables necessary to distinguish polarisation should used as inputs. In addition, the polarisation classifiers are trained entirely using MC. The modelling of the classifiers with respect to data is dependent on the accuracy of the MC predictions. Generally, a smaller set of input variables is less prone to mismodelling. The first pass will also serve the purpose of establishing a rough upper bound on the network performance. Dropping input features is expected to have a small effect on the overall performance.

The best-performing classifiers of the second pass training are used for the final measurement. The following sections document the training procedure and the phase space of the hyperparameters tested. A detailed overview of the training for each pass is provided.

[6]Binary classifiers predict the probability an event is either signal or background.

The best-performing networks at each pass are also validated using the analysis's control regions to compare the modelling with experimental data against unseen MC. This is an important test to ensure there is no major mismodelling or overfitting by the networks.

7.3.3 Samples

The polarisation classifiers are trained on the simulated polarised electroweak $W^{\pm}W^{\pm}jj$ signal samples. A separate classifier is trained for the WW and *pp* CoM frames. Table 30 quantifies the raw number of events for training and the number of events expected in the data for each process.

Training Samples		
Sample	Raw Events	Predicted Events
WW CoM frame		
$W_L^{\pm}W_L^{\pm}jj$	215008	18.29
$W_L^{\pm}W_T^{\pm}jj$	170387	58.88
$W_T^{\pm}W_T^{\pm}jj$	456456	124.50
pp CoM frame		
$W_L^{\pm}W_L^{\pm}jj$	19673	11.49
$W_L^{\pm}W_T^{\pm}jj$	21471	67.84
$W_T^{\pm}W_T^{\pm}jj$	46320	123.06

Table 21: The samples used for training the polarisation classifier. For each CoM frame, the raw number of training events and the number of events expected in the data signal region are shown.

The number of raw events in the WW CoM frame is significantly larger than in the *pp* CoM frame. Generating fully reconstructed samples is very computationally expensive. It was decided to prioritise the WW CoM frame within the bounds of the available resources. This reference frame is expected to give better sensitivity to longitudinal scattering. The number of transverse-only scattering events is almost the same between the two frames. However, the WW CoM frame has a larger number of LL events. This is advantageous to the LL measurement as the statistical uncertainty is substantial. The increased number of LL events in the WW CoM frame may also benefit the LX measurement. The classifiers are expected to perform better at isolating LL events from the TT background.

7.3.4 Input Variables

The list of input features tested for the polarisation DNNs is given in Table 22. A total of 32 input variables are initially trialled during the first stage of training. The variables are split into two categories: low-level and high-level. Low-level variables refer to the kinematic properties of physics objects directly reconstructed from the detector. High-level variables are compositions of low-level variables. A sufficiently complex DNN could

potentially learn to reconstruct the necessary relationships from just the low-level variables with equal performance compared to DNNs trained using both. However, the size and scope of such networks are not immediately apparent before training. As this work focuses on maximising the sensitivity to polarisation within the bounds of the computational capacity, the high-level variables are also included in the training.

Input Variables		
Name	Description	Scaling
		Low-level variables
lep0 p_T	Leading lepton transverse momentum, p_T	$\log_{10}(x)$
lep1 p_T	Sub-leading lepton transverse momentum, p_T	$\log_{10}(x)$
lep0 η	Leading lepton η	x
lep1 η	Sub-leading lepton η	x
lep0 type	PID of the leading lepton η	x
lep1 type	PID of the sub-Leading lepton η	x
lep1 ϕ	Recalibrated sub-Leading lepton azimuthal angle [126], ϕ	x
lep0 E	Leading lepton energy, e	$\log_{10}(x)$
lep1 E	Sub-leading lepton energy, e	$\log_{10}(x)$
jet0 p_T	Leading jet transverse momentum, p_T	$\log_{10}(x)$
jet1 p_T	Sub-leading jet transverse momentum, p_T	$\log_{10}(x)$x
jet0 y	Leading jet rapidity, y	x
jet1 y	Sub-leading jet rapidity, y	x
jet0 ϕ	Recalibrated sub-Leading jet azimuthal angle [126], ϕ	x
jet1 ϕ	Recalibrated sub-Leading jet azimuthal angle [126], ϕ	x
E_T^{miss}	Missing transverse energy, p_T^{miss}	x
E_T^{miss}-ϕ	Recalibrated missing transverse energy azimuthal angle [126], ϕ	x
		High-level variables
lep0 zep	Leading lepton Zeppenfeld variable [127]	$\sqrt{x}$
lep1 zep	Sub-leading lepton Zeppenfeld variable [127]	$\sqrt{x}$
leps $\Delta\eta$	Difference in the leading and sub-leading leptons η, $\Delta\eta_{ll}$	x
leps charge	Charge of the leading and sub-leading leptons	x
lep0 m_T	Leading lepton transverse mass, m_T	$\sqrt{x}$
lep1 m_T	Sub-leading lepton transverse mass, m_T	$\sqrt{x}$
leps m_{ll}	Invariant mass of leading and sub-leading leptons, m_{ll}	$\log_{10}(x)$
leps m_T^{ll}	Transverse mass of the dilepton system, m_T^{ll}	$\sqrt{x}$
leps p_T	Sum leading and sub-leading leptons p_T	$\log_{10} x$
leps ΔR	Distance of the dilepton $\Delta\eta$ and $\Delta\phi$	x
leps m_{o1}	Early-projected massless invariant mass of dilepton system, m_{o1}	$\sqrt{x}$
jets $\Delta\eta$	Difference in the leading and sub-leading jet η, $\Delta\eta_{ll}$	x
jets $\Delta\phi$	Difference in the leading and sub-leading jet ϕ, $\Delta\phi_{ll}$	x
jets ΔR	Distance of the dijet $\Delta\eta$ and $\Delta\phi$	x
jets m_{jj}	Invariant mass of leading and sub-leading jet, m_{jj}	$\log_{10}(x)$
lepton-jet p_T ratio	Lepton-jet p_T ratio	x
lepton-jet min(ΔR)	Minimum ΔR between the leptons and jets	x

Table 22: The full set of input features tested for the polarisation classifiers.

The right-hand column of Table 22 lists the scaling functions applied to some variables to produce a distribution with a mean of zero and a standard deviation of one. The scaling functions are applied to avoid network bias introduced by variables with large values. For example, p_T variables can be several orders of magnitude larger than angular variables.

The large event weights would dominate in the network and dilute the impact of the angular variables. However, as discussed, the angular variables are often more sensitive to polarisation. The normalisation ensures that each variable is treated uniformly.

7.4 Training Methods

This section outlines the training methods and network architectures applied in this chapter and the subsequent chapter. The section describes the methods used for training, the network structures and initialisation, and the procedure applied to optimise the network hyperparameters. The following subsections will each provide a global overview of the specific methods, with relevant literature cited for the reader. As explained earlier, the analysis in this thesis primarily applies these methods to study polarisation, so a detailed overview is not pertinent to the understanding of this chapter.

7.4.1 Cross-Validation

During the training process, the full dataset available for learning is typically partitioned into a training, testing and validation sets. The training and validation are used to train and optimise the network hyperparameters, while the testing set is used as an unbiased estimate of the model performance. Generally, half of the available data is used for training, and the remaining half is evenly split for testing and validation. This procedure is suitable for most ML methods but is limiting when the available dataset is small, as is the case for the training documented here. Furthermore, to ensure the measurement remains unbiased, partitioning the dataset in this manner would mean only a quarter of the available MC data could be used to extract the signal strength. Reducing the available MC statistics will increase the uncertainty, so this approach is not feasible.

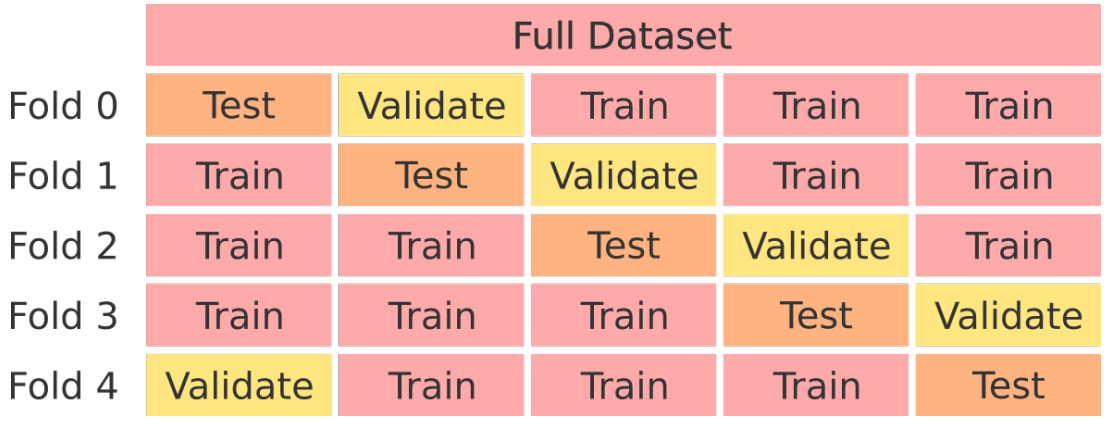


Figure 67: An illustration of the cross-validation procedure and the partitioning of training, testing and validation data for each fold.

Alternatively, the ratios between the testing, training and validation sets could be altered. Increasing the testing data would reduce the available data for training and reduce the model's performance, so this approach is also not feasible. For this reason, this thesis employs a method inspired by cross-validation. Cross-validation is a procedure

in which the available data is partitioned into n even-sized datasets referred to as folds. The model is trained for multiple-fold configurations during cross-validation, where one fold is used for validation and the remaining for training. The method is generally used to test the consistency of the model for a given set of hyperparameters across the entire dataset. If the model performs consistently for a given set of hyperparameters, then the model will likely generalise to the given dataset.

This thesis applies five-fold cross-validation, which partitions the dataset into five folds as shown in Fig. 67: three for training, one for testing, and one for validation. The cross-validation procedure remains the same, whereby a total of five networks are trained, each with different folds for training, testing and validation. Each of the five networks is used for the measurement, and the signal extraction is performed using the MC prediction over the test fold. The advantage of this method is that, in combination with the five networks, it gives an unbiased estimate of the full MC dataset. In practice, this method is implemented by partitioning the data by event number such that in the signal region, the corresponding network prediction is given only on the testing MC. The downside of this method is that it requires training multiple networks. The performance of each of the networks will be different across different folds. Selecting the best network is not only a question of choosing the best network on an individual fold but also the model most consistent across all five folds.

7.4.2 Adam Optimiser

The minimisation algorithm used to train the DNNs in this thesis is the ADAptive Moment estimation optimiser (Adam) [128]. Classical Stochastic Gradient Descent (SGD) maintains a fixed learning rate throughout each training iteration and applies the same learning rate to each weight. Adam is an SGD method which uses adaptive learning rates. This rate is adjusted using the average of the first- and second-order moments of the gradient. These moments are calculated from an exponential moving average of the gradient and the squared gradient. The optimiser is parameterised with three hyperparameters: two parameters β_1 and β_2, which control the exponential decay of these moving averages, and a parameter ϵ, which is a small constant introduced for numerical stability. The Adam optimiser has been shown to perform better than similar SGD methods for various machine learning problems [128]. In the initial stages of testing, different optimisers were trialled, and Adam optimiser was found to be the most consistent.

7.4.3 Bayesian Optimisation

The method for hyperparameter optimisation applied in this thesis is Bayesian Optimisation [129]. Hyper-parameter optimisation is the task of finding the best set of parameters, i.e. those which minimise or maximise the performance of the network on a given metric. Most classic hyperparameter optimisation methods involve random searches or grid scans, which evaluate the performance of the networks for a large set of parameters. Often, these methods work well but can be very computationally expensive. Bayesian optimisation is different and assumes the parameter set over a given metric is Gaussian distributed. Each evaluation of a given set of parameters improves the approximation of distribution, which allows the algorithm to find the regions of parameter space worth exploring. A detailed

description can be found in [129]. Formally, the task of Bayesian Optimisation is as follows: given a function $f : \mathcal{X} \to \mathcal{R}$, find the value x' which minimises or maximises f, i.e.,

$$x' = \arg\min_{x \in \mathcal{X}} f(x) \text{ or } \arg\max_{x \in \mathcal{X}} f(x). \tag{124}$$

An assumption is made for the optimisation based on a probabilistic belief that the function f is Gaussian distributed. Using a Gaussian prior, Bayesian optimisation constructs a posterior distribution after randomly trialled data points. With each data point, the posterior distribution is updated and improved. The choice of $x \in \mathcal{X}$ used to evaluate f is selected by the minimisation of an acquisition function representing the loss of evaluating f for a given value of x. The task of optimisation is then to minimise the expectation value of the acquisition function with respect to the posterior distribution. Practically, Bayesian optimisation translates the task of minimising or maximising a computationally expensive black-box function f to minimising a much simpler acquisition function.

7.4.4 Initialisation

The initialisation of the network weights is a crucial component of network convergence. A detailed discussion of NN weight initialisation is beyond the scope of this thesis. This section instead describes the initialisation strategy chosen for this measurement, referencing the relevant literature. Network depth is an important parameter to model generalisation and the ability of the model to learn complex relationships. However, it has been shown that there exists a scale which limits the maximum depth information can propagate through a network [130, 131]. The limit depends on the choice of hyperparameters and the initialisation. It has also been shown that there exists a specific choice of initialisation, referred to as the edge of chaos, which maximises this limit, allowing for a deeper propagation of information through the network [130, 131]. The choice of initialisation is also dependent on the activation function. The activation functions tested in this thesis are the ELU [132] (Exponential Linear Units) and swish activation function [133]. This analysis follows the edge of chaos initialisation recommended for the swish activation function [134] and the ELU activation [135].

7.4.5 Network Architecture

The network architecture is an essential component of network convergence. The architectures presented in this work are dense, fully-connected networks. These networks are composed of hidden layers with a fixed number of n nodes, as illustrated in Fig. 66. Various other network structures were initially trialled, such as those with varying widths in the hidden layers and more complex network structures following a particle-based model described in [136]. However, the initial trials showed no evident performance gain between the different structures. For simplicity, it was decided to train dense, fully-connected networks. These networks are parameterised by only two variables, significantly reducing the complexity of the hyperparameter optimisation. There may still be some improvement with more complex network structures, as in [136]. However, this improvement is expected to be marginal.

7.4.6 Categorical cross-entropy

The categorical cross-entropy loss function is used to optimise the networks. Cross-entropy is a measure of similarity between two probability distributions. Categorical cross-entropy determines a probability distribution across multiple classes, in this case, polarisation categories. It is an extension to binary cross-entropy, which predicts an event being signal or background. A higher entropy suggests a greater similarity between the two distributions. The loss function is defined as

$$L(y, \hat{y}) = -\sum_{i}^{n} y_i log(\hat{y}_i), \tag{125}$$

where y is the true distribution, $\hat{y}$ is the predicted distribution, and the sum is performed over the n classes. The loss function is the negative of the cross-entropy such that minimising the loss maximises the similarity between the distributions. The loss function is used in conjunction with the "softmax" function, which normalises the output of a network to a probability distribution. The softmax function is defined as,

$$\sigma(x)_i = \sum \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_n}}, \tag{126}$$

where i is the output for a given class, x an input vector and n is the total number of classes.

7.4.7 ROC Curve and AUC

To evaluate the performance of the networks, this analysis applies the Area Under Curve (AUC) metric [137]. This metric is used generally to evaluate the performance of binary classification algorithms. As suggested by the name, the metric is the integrated area under the Receiver Operating Characteristic (ROC) curve [137], illustrated in Fig. 68. The ROC curve is the false positive rate (FPR) as a function of the true positive rate (TPR), where

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}, \tag{127}$$

and

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{128}$$

with (T/F)(N/P) denote a true/false and negative/positive. For example, in an experimental context, a false positive refers to an event incorrectly identified as signal, and a true negative refers to an event correctly identified as background. The ROC curve is calculated by varying the classification threshold. The classification threshold refers to the minimum probability for an event to be classed as a signal event. Lowering the threshold will increase the number of signal events identified correctly and the number of background events falsely identified as signal. The ROC curve describes the performance of the binary classification algorithm at various classification thresholds. A better-performing classifier will typically see a larger TPR for a given FPR. The AUC summarises the performance of the network at each of the thresholds. The benefit of the AUC is that it is scale-invariant

and has no dependence on the classification threshold [137]. A larger AUC implies a better classification model. During the initial testing stage, it was found that the AUC was more correlated with statistical significance, as opposed to other potential metrics such as the accuracy, albeit marginally.

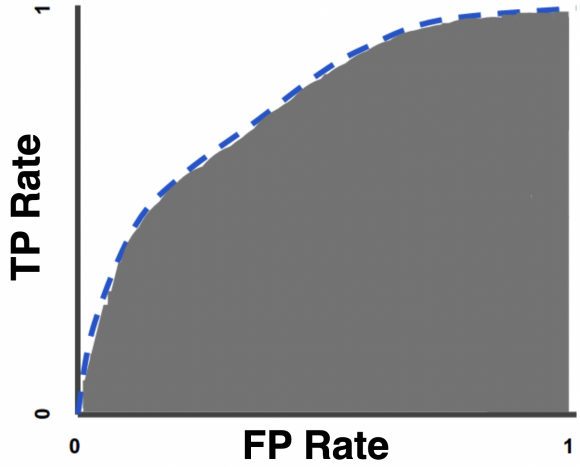


Figure 68: An example ROC curve, where the grey region is the integrated area of the curve (referred to as the AUC).

The AUC can also be extended for multi-class classification, where the ROC is calculated individually for each class, and the AUC is given as the average across all classes.

7.5 Instance-wise Variable Selection

Each input variable selected for the initial training offers a different level of discrimination to polarisation. As discussed in Sec. 7.1, some of the variables, such as the dijet invariant mass m_{jj} are sensitive to inclusive electroweak $W^{\pm}W^{\pm}jj$ production but not to polarisation. Many of the variables listed in Table 22 may not contribute towards the network's ability to distinguish between polarisation configurations. To reduce the potential for over-fitting and mismodelling, unnecessary input variables are removed. The input variables are selected using a model referred to as INstance-wise VAriable SElection (INVASE) [138]. INVASE attempts to quantify the relevance of features in each sample. Understanding which variables are most pertinent to the model will help illuminate the underlying physics of the models and quantify the importance of each input variable.

7.5.1 Overview

This sub-section briefly overviews the structure of the INVASE model following [138] and its implementation in this work. In this case, the model aims to discover the relevant input variables for discriminating polarisation. INVASE is inspired by actor-critic models. In this context, the actor is responsible for finding the optimal set of input variables by testing the impact of different variables and comparing the predictions to a critic with the complete set of inputs. The technical implementation is detailed in Fig. 69. The model

consists of three networks: a selector (the actor) and a baseline and predictor network (the critic).

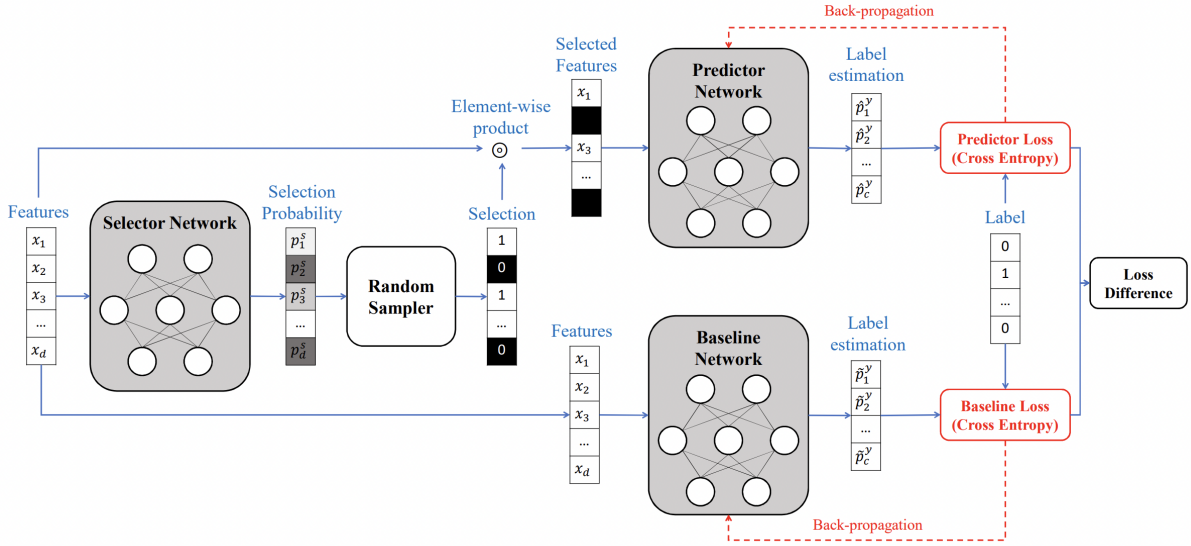


Figure 69: A graphical depiction of the INVASE model, showing the relationships between the actor, critic and baseline network. In the figure, the actor is referred to as the selector network, and the critic as the predictor network. Figure taken from [138].

The selector network returns a probability a given feature should be selected. The selected features are used as inputs to the predictor network. The baseline network is used for comparison and uses the full set of input variables. The input variable selection is implemented by optimising the selector network with a loss function that minimises the Kullback-Leibler (KL) divergence[7] between the baseline and the predictor. The loss function also includes an additional factor, which penalises the model for predicting variables with a high probability. This factor is scaled by a hyperparameter, λ, denoted as the INVASE parameter. The optimisation process must balance minimising the KL divergence without inducing a significant penalty for selecting too many input features. The number of selected input features scales with the parameter λ. During the training process, the actor will learn which input features are most important to minimise the difference between the baseline and predictor networks. The reader is referred to Ref. [138] for a detailed description.

7.6 First-Pass Training

The first stage of the training procedure for the polarisation classifiers determines a baseline network to be used for the input feature selection model. The networks in the first pass are trained with the full set of input variables shown in Table 22. A large number of networks are tested during this stage. The training and network size is configured with a preselected set of hyperparameters. The hyperparameters are chosen and optimised using

[7]The KL divergence measures dissimilarity between two probability distributions [139]

Bayesian optimisation. This section begins by outlining the network architectures and a list of hyperparameters tested for the polarisation classifiers. Subsequently, the section details the results of the Bayesian optimisation procedure.

7.6.1 Network Architecture

The network architectures tested for both first pass and second passes are dense, fully-connected networks implemented with Keras [140]. The architecture additionally applies dropout [141] and batch normalisation [142] layers as regularisation methods. The reader is referred to the references for a detailed description of each. The networks are defined with the following structure,

1. Input layer, N_i
2. Repeated N_l hidden-layers,
   (a) Batch normalisation layer
   (b) Dense layer with N_n nodes
   (c) Activation layer
   (d) Dropout layer with α_d
3. Output Layer,

where N_i, N_l, and N_n are integer values representing the number of input features, hidden layers and nodes per layer in the network and α_d the dropout rate. The structure of the hidden layers remains the same and is repeated N_l times. Each hidden layer consists of a batch normalisation layer connected to a dense layer with N_n nodes. The activation function is then applied across these layers and is fed as an input to the dropout layer.

7.6.2 Hyperparameters

Hyperparameters configure the network architecture. The full list of hyperparameters trialled for the polarisation classifiers in the first pass for each reference frame is shown in Table 23. The parameters highlighted in bold are different between the reference frames. The parameter ranges tested for the *pp* CoM frame are chosen to be slightly smaller, reflecting the smaller dataset available for training. These networks are also tested with a slightly higher dropout rate. The table is also split by parameters configuring the network and the parameters related to the training procedure. The bottom set of hyperparameters is associated with the training through the Adam optimiser. A broad range of parameter values are tested for the training networks. The range of batch sizes was selected following the recommendations of [143]. In general, larger batch sizes lead to faster training but have been shown to lead to poorer generalisation.

The hyperparameters are optimised using Bayesian Optimisation. The AUC calculated for the validation data is the objective metric maximised during the optimisation procedure. The training data are partitioned following the cross-validation procedure. However, full cross-validation training was not applied during Bayesian Optimisation. Given the

Hyperparameters		
Parameter	Range (WW)	Range (*pp*)
Inputs	32	32
Layers	[5, 12]	**[3, 10]**
Nodes	[60, 140]	**[50, 120]**
Activation Function	ELU, swish	ELU, swish
Output Function	softmax	softmax
Number of outputs	3	3
Dropout	[0.1, 0.4]	**[0.15, 0.5]**
β_1	[0.001, 0.99]	[0.001, 0.99]
β_2	[0.001, 0.99]	[0.001, 0.99]
ϵ	$[1 \times 10^{-3}, 1 \times 10^{-8}]$	$[1 \times 10^{-3}, 1 \times 10^{-8}]$
Learning rate	$[1 \times 10^{-1}, 1 \times 10^{-5}]$	$[1 \times 10^{-1}, 1 \times 10^{-5}]$
Batch Size	[32, 512]	[32, 512]

Table 23: The full ranges of hyperparameters tested for the first-pass training of the polarisation classifiers. Differences in architecture between the WW and the *pp* CoM frame are shown in bold. The horizontal line in the table indicates the separation between parameters associated with the network architecture and parameters associated with the Adam optimiser.

extensive set of hyperparameters, many iterations are required to find an optimal parameter space. For this reason, only the first iteration of cross-validation is performed. Repeating the optimisation procedure across each of the folds would be incredibly computationally expensive. The full cross-validation training procedure is only applied to the final best-performing model with the optimal set of hyperparameters. Initial testing found that the network performance did not vary sufficiently during cross-validation to warrant performing the optimisation procedure separately for each training fold.

The training procedure also applies an early stopping condition. The condition is triggered if the evaluated validation loss does not decrease for 12 consecutive training epochs. If the validation loss fails to decrease in this period, then the training is stopped. This condition is applied to avoid overfitting the training data. The loss function during training is also weighted with the predicted event weights to be representative of the expected event yields in the signal region. This means that events with large event weights will have a more significant impact on the network training. Finally, the Adam optimiser trains the polarisation classifiers with the categorical cross-entropy loss function.

7.6.3 Optimisation Results

The highest performing networks with respect to the validation AUC for the first iteration of cross-validation are shown in Table 24. During the first pass, approximately 6000 networks were trained and tested in the optimisation procedure. The highest AUC for each reference frame is,

1. WW CoM: 0.7706 ± 0.0004
2. *pp* CoM: 0.757 ± 0.002

The performance with respect to the AUC in the WW centre-of-mass frame is better due to the larger training dataset and network size. The networks in the *pp* CoM frame with the highest AUC were trained with a larger dropout rate, reflecting the smaller training dataset. The table also shows the total number of training epochs before the training procedure was halted due to the early stopping condition. The models in the *pp* CoM frame were trained with a faster learning rate and a similar batch size, leading to the difference in the number of training epochs.

Best Parameters		
Parameter	WW CoM frame	*pp* CoM frame
Inputs	32	32
Layers	10	3
Nodes	124	77
Activation Function	elu	elu
Output Function	Softmax	Softmax
Number of outputs	3	3
Dropout	0.187549	0.253217
Batch Normalisation	True	True
β_1	0.516165	0.501201
β_2	0.005000	0.993067
ϵ	0.00000011	0.00000011
Learning rate	0.001256	0.005584
Batch Size	311	307
Epochs	117	49

Table 24: The best set of hyperparameters associated with the first-pass training for the polarisation classifiers in each CoM frame.

7.7 Input Feature Selection

The input variable selection procedure uses the INVASE model to select an optimal subset of input variables for the polarisation classifier.

7.7.1 Training Procedure

The network architectures of the predictor and baseline networks follow analogously to the networks described in the first pass, with the hyperparameters for both given in Table 24. The hyperparameters for the actor are chosen following the models detailed in [138] and are shown in Table 25. Initial preliminary testing found that larger actor networks struggled to converge, with smaller networks being more suitable. The only parameter which is varied during training is λ. The actor is trained using the Adam optimiser and a stopping condition equivalent to that used in the baseline network training. As with the first pass, the input feature selection is only performed for the first cross-validation iteration.

Actor Hyperparameters	
Parameter	Value
Inputs	38
Layers	2
Nodes	76
Activation Function	elu
Output Function	Sigmoid
Number of outputs	38
Batch Normalisation	True
β_1	0.9
β_2	0.99
ϵ	1×10^{-7}
Learning rate	0.001
Batch Size	311
λ	[0.000001, 0.5]

Table 25: The set of hyperparameters tested for the actor network used in the INVASE model associated with the polarisation classifiers.

7.7.2 Results

A total of 13 individual INVASE models were trained per reference frame for each classifier, differing only by the λ parameter. The final output of each model returns a set of relevant input variables. The first step is for the actor to provide a probability that an input variable should be selected. In this analysis, variables are selected if the probability is greater than 0.5. The input variables are ranked based on the number of selections. Variables with more selections have been chosen by the actor for larger values of λ and

are, therefore, deemed more important to the classification. The ranking of the input variables is illustrated in Fig. 70 for the WW and *pp* CoM frames and gives some interesting insight into the physics of polarisation.

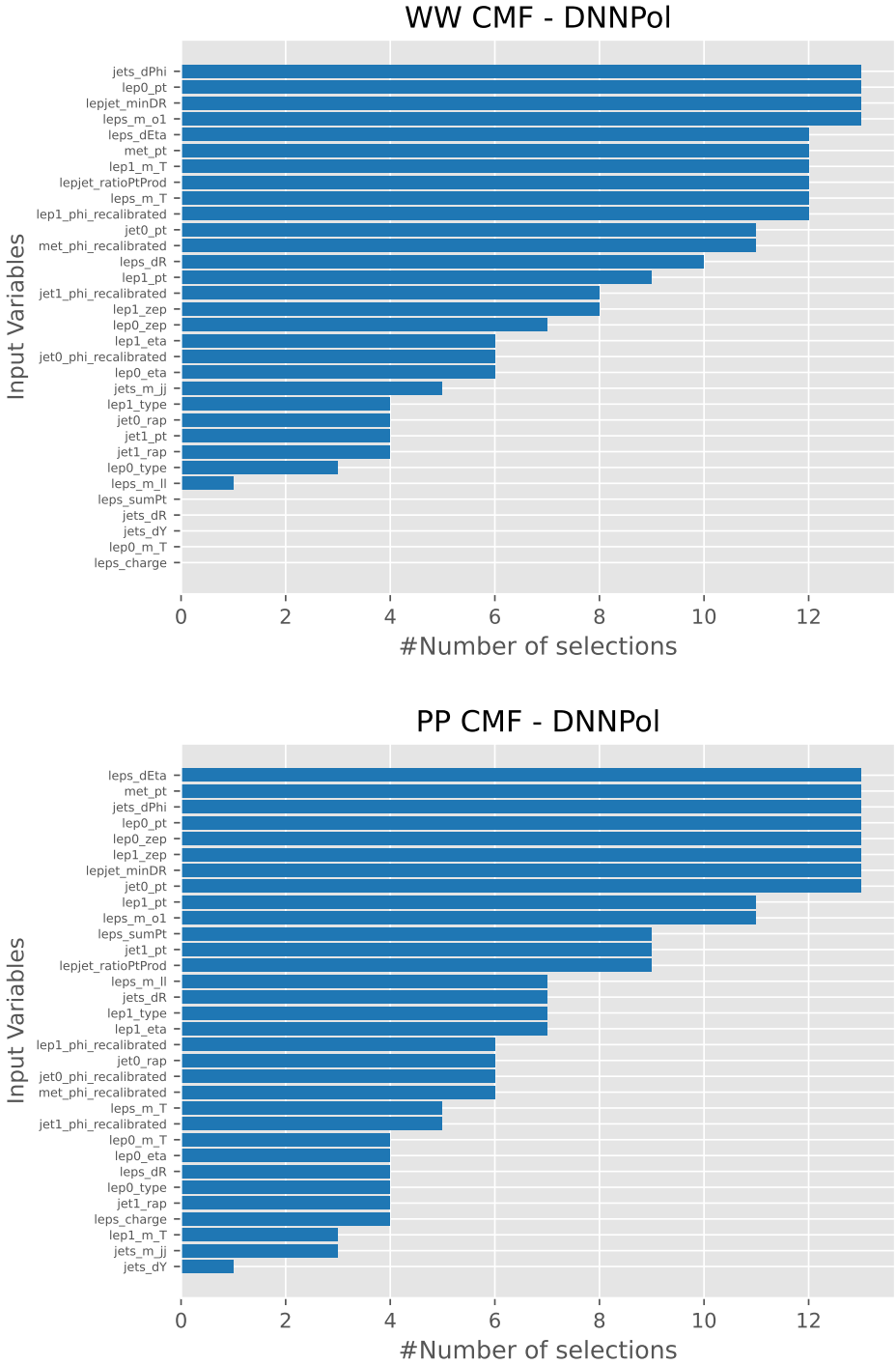


Figure 70: The ranking of the best input features in each reference frame following the training of the INVASE models.

Both reference frames share features in the ranking. The leading lepton p_T and the $\Delta\phi_{jj}$ distributions are the two of the most important variables for both reference frames and are selected for every iteration of the input feature selection. Figure 71 shows the normalised distribution for each variable in both centre-of-mass frames. The figures demonstrate that both variables are highly discriminating in both reference frames, consistent with the physics discussion in the introduction of this chapter.

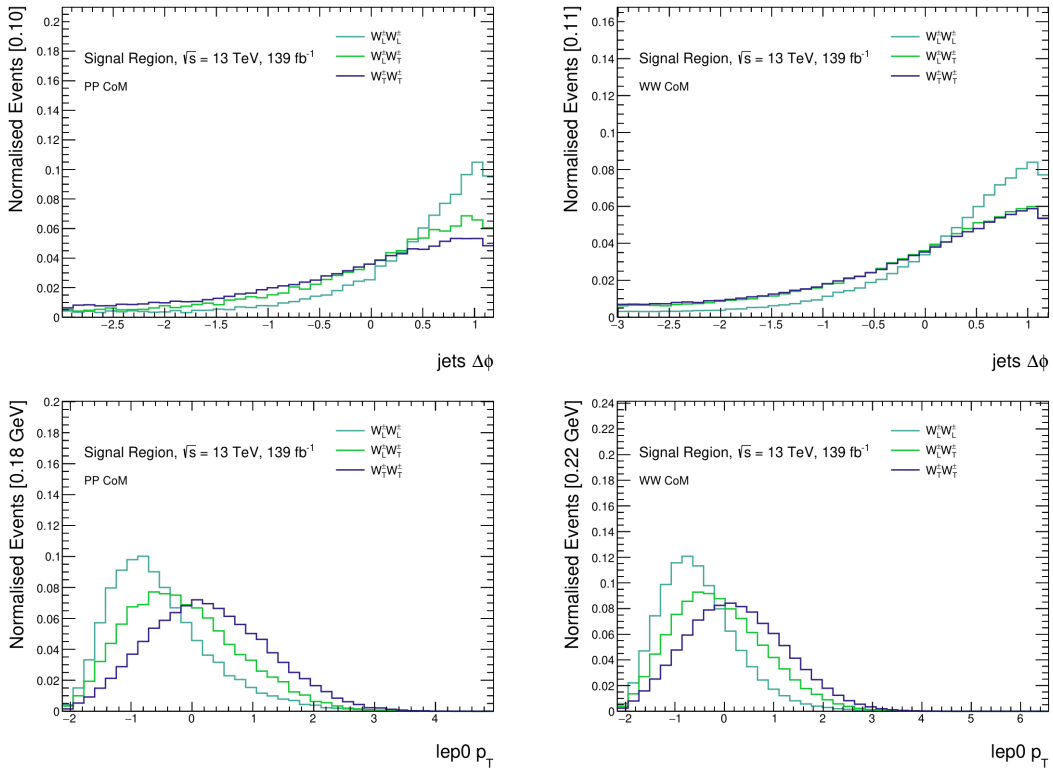


Figure 71: The normalised jet $\Delta\phi$ and leading lepton p_T distribution in the pp (left) and WW (right) CoM frames.

There are, however, clear differences in the rankings in the two frames, owing to the physics of polarisation. The most evident differences originate from the lepton angular distributions, which vary between the reference frames. As an example, the lepton ΔR_{ll} variable is ranked significantly lower for the networks trained in the pp CoM frame. Figure 72 shows the lepton ΔR_{ll} and lepton $\Delta\phi_{ll}$ distributions. The figure demonstrates the differences in the distributions for polarisation in both reference frames. The polarisation separation is lower in the pp CoM frame; the ΔR_{ll} ranking is also lower. This effect originates from the differences in the lepton $\Delta\phi_{ll}$ distributions shown in Fig. 72.

The ranking of the Zeppenfeld variables for the leading and sub-leading leptons is seen to be higher in the pp CoM frame. Figure 73 shows the sub-leading lepton Zeppenfeld variable for each polarisation configuration in both reference frames. The figure demonstrates a better polarisation separation polarisation in the pp CoM frame compared to the WW CoM frame.

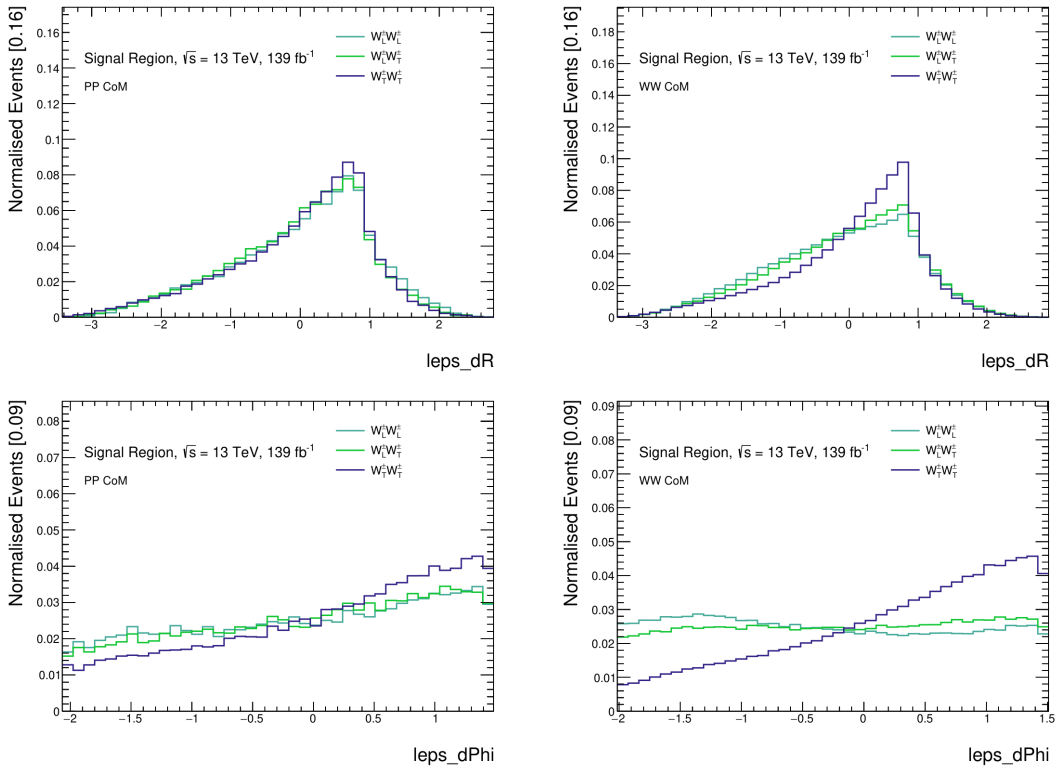


Figure 72: The normalised lepton ΔR and leading lepton $\Delta\phi$ distribution for the pp (left) and WW (right) CoM frames.

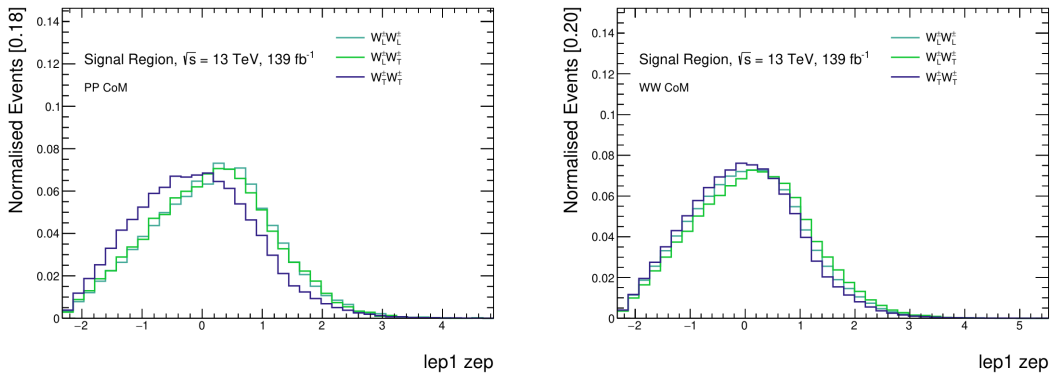


Figure 73: The normalised sub-leading Zeppenfeld distribution for the pp (left) and WW (right) CoM frames.

7.8 Second Pass Training

Following the input variable selection, the network training is repeated but with a subset of features selected by choosing a value of λ with a high-performing validation AUC on the critic. The final input variables selected are shown in Table 26. Both reference frames share a similar set of inputs. Most low-level variables are chosen, with only a few of the most discriminating high-level variables included. Both reference frames use a set of 21

inputs, reduced from the original 32.

Input Variables			
Name	WW CoM frame	*pp* CoM frame	Scaling
			Low-level variables
lep0 p_T	X	X	$\log_{10}(x)$
lep1 p_T	X	X	$\log_{10}(x)$
lep0 η	X	X	x
lep1 η	X	X	x
lep0 type	-	-	x
lep1 type	-	-	x
lep1 ϕ	X	X	x
jet0 p_T	X	X	$\log_{10}(x)$
jet1 p_T	-	X	x
jet0 y	X	-	x
jet1 y	X	-	x
jet0 ϕ	X	X	x
jet1 ϕ	X	X	x
E_T^{miss}	X	X	x
E_T^{miss}-ϕ	X	X	$\log_{10}(x)$
			High-level variables
lep0 zep	X	X	$\sqrt{x}$
lep1 zep	X	X	$\sqrt{x}$
leps $\Delta\eta$	X	X	x
leps charge	-	-	x
lep0 m_T	-	-	$\sqrt{x}$
lep1 m_T	-	-	$\sqrt{x}$
leps m_{ll}	-	X	$\log_{10}(x)$
leps m_T^{ll}	X	X	$\sqrt{x}$
leps p_T	-	-	$\log_{10} x$
leps ΔR	X	-	x
leps m_{o1}	X	X	$\sqrt{x}$
jets $\Delta\eta$	-	-	x
jets $\Delta\phi$	X	X	x
jets ΔR	-	X	x
jets m_{jj}	-	-	$\log_{10}(x)$
lepton-jet p_T ratio	X	-	x
lepton-jet min(ΔR)	X	X	x

Table 26: The final sets of input variables used for the WW and *pp* CoM frames polarisation classifiers. The variables marked with 'X' are selected, and the variables marked with '-' are dropped for the second-pass training.

7.8.1 Procedure

The training procedure is repeated as in the first pass. The network architectures tested are described in Section 7.6.1. The hyperparameter ranges are given in Table 27. The values highlighted in bold indicate the differences between each reference frame. Each classifier is trained using the Adam Optimiser, which minimises the categorical cross-entropy loss function. The classifiers are also trained with an early stopping condition to halt training after no improvement in the validation loss for 12 epochs.

7.8.2 Optimisation Results

During the second pass-training, approximately 4000 models were trained and tested. The models with the highest scoring validation AUC are used as the final networks for this experimental measurement. The best-performing set of hyperparameters for each reference frame are given in Table 28. The polarisation classifier in the WW (PP) CoM frame is referred to as DNNPol (DNNPolPP). Each of the outputs for the classifiers are

Hyperparameters		
Parameter	Range (WWcmf)	Range (PPcmf)
Inputs	21	21
Layers	[7, 10]	**[3, 5]**
Nodes	[60, 140]	**[50, 120]**
Activation Function	elu, swish	elu, swish
Output Function	Softmax	Softmax
Number of outputs	3	3
Dropout	[0.15, 0.3]	**[0.2, 0.4]**
β_1	[0.001, 0.99]	[0.001, 0.99]
β_2	[0.001, 0.99]	[0.001, 0.99]
ϵ	$[1 \times 10^{-3}, 1 \times 10^{-8}]$	$[1 \times 10^{-3}, 1 \times 10^{-8}]$
Learning rate	$[1 \times 10^{-1}, 1 \times 10^{-5}]$	$[1 \times 10^{-1}, 1 \times 10^{-5}]$
Batch Size	[200, 400]	[200, 400]

Table 27: The phase-space of hyperparameters tested for the second-pass training. The bold indicates the differences between the parameters trialled in the WW and *pp* CoM frame. The horizontal line divides the parameters associated with the network architecture and the Adam optimiser.

subscripted by the target polarisation category, for example, $\text{DNNPol}_{\text{LL}}$. Characteristically, the networks trained on data in the *pp* CoM frame have a slightly higher dropout rate due to a smaller training dataset. After the second pass, the training procedure for the selected networks is repeated with the full cross-validation procedure. The final validation AUC for the full training procedure is given in Table 29. The table shows consistent performance across the folds, with slight variations in the best-performing epoch.

The results of the final training showed improved performance compared to the bigger networks trained in the first pass. There are a couple of possible reasons for the improvement. First, the optimisation procedure for the second pass was performed for more trials and a smaller range of parameters. This means that the optimiser in the second pass had a larger scope to fine-tune the hyperparameters. The network sizes also largely remained the same but with a reduced set of input features. Dropping some of the weaker discriminant variables while keeping the same network size may have improved network sensitivity to polarisation.

The total output of the polarisation classifiers for the WW CoM frame is given in Fig. 74. Each subplot shows the normalised predictions for the DNNPol networks trained in the cross-validation procedure. The predictions are made using the entire dataset, i.e. the training, testing and validation data. As observed with the validation AUC in Table 29, the predictions shown in Fig. 74 are consistent across each cross-validation fold, with minor variations in the shape.

The normalised distributions in Fig. 74 also demonstrate a clear separation in polarisation. Each TT distribution clearly separates TT events and events with longitudinally polarised gauge bosons. The classifiers have the most difficulty distinguishing mixed po-

Best Parameters		
Parameter	WW CoM frame	*pp* CoM frame
Name	DNNPol	DNNPolPP
Inputs	21	21
Layers	8	4
Nodes	104	120
Activation Function	swish	ELU
Output Function	softmax	softmax
Number of outputs	3	3
Dropout	0.226127	0.287081
β_1	0.291029	0.366648
β_2	0.024093	0.024751
ϵ	0.0000001	0.0000001
Learning rate	0.001042	0.001458
Batch Size	243	229

Table 28: The hyperparameters of the best polarisation classifiers in the WW and *pp* CoM after the second pass.

Validation AUC					
	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4
pp CoM frame	0.755	0.752	0.752	0.750	0.756
WW CoM frame	0.776	0.778	0.776	0.777	0.776

Table 29: The validation AUC of the polarisation classifiers used in this experimental measurement. The statistical uncertainty on the AUC is of the order of magnitude $10^{-4}(10^{-2})$ in the WW (PP) CoM frame.

larisation events, demonstrated by the central output of the networks, LT. These network features are a consequence of the event weighting applied during training. As discussed, the training data are weighted to the signal region event yields. As the TT polarisation fraction dominates the event yields, the classifiers are more influenced to learn to classify TT events. The networks perform most poorly in classifying LT events. However, the networks have greater difficulty separating LT from LL than TT events. This feature is desirable for the measurement of mixed scattering LX.

Figure 75 shows the normalised training, testing, and validation $\text{DNNPol}_{\text{TT}}$ distributions for the different polarisation samples in each cross-validation fold. Only the TT network prediction is shown, as this is the distribution used in the final measurement. From the figures, no significant overfitting is present in the network predictions. The models show good generalisation across all of the folds. Some minor overfitting is apparent in the deviations between the data sets beyond the statistical uncertainty.

Figure 76 demonstrates the full DNNPolPP normalised network predictions across

each fold. The distributions shown in the figure have similar features. In this frame, there are slightly larger variations across the folds. This is due to the smaller training dataset, which leads to a larger statistical uncertainty. The classifiers in the *pp* CoM frame show slightly improved classification of LT events. The fraction of LT events is bigger in the *pp* CoM frame, and the networks are more influenced to learn how to classify LT events. However, it is also apparent that the DNNPolPP classifiers struggle to distinguish LL and LT events. Moreover, the general classification is overall slightly worse in the *pp* CoM frame. This could partially be a polarisation effect in the *pp* CoM frame; however, the performance is limited due to the smaller network sizes and data available for training.

Figure 77 shows the normalised $\text{DNNPolPP}_{\text{TT}}$ distribution for the same samples as Fig. 76. Similarly to the previous discussion, the figures show no major overfitting. Some small overfitting is present in the predictions, and the overfitting magnitude is slightly larger than that in the WW CoM frame. This is because the dataset available for training in the *pp* CoM frame is significantly smaller.

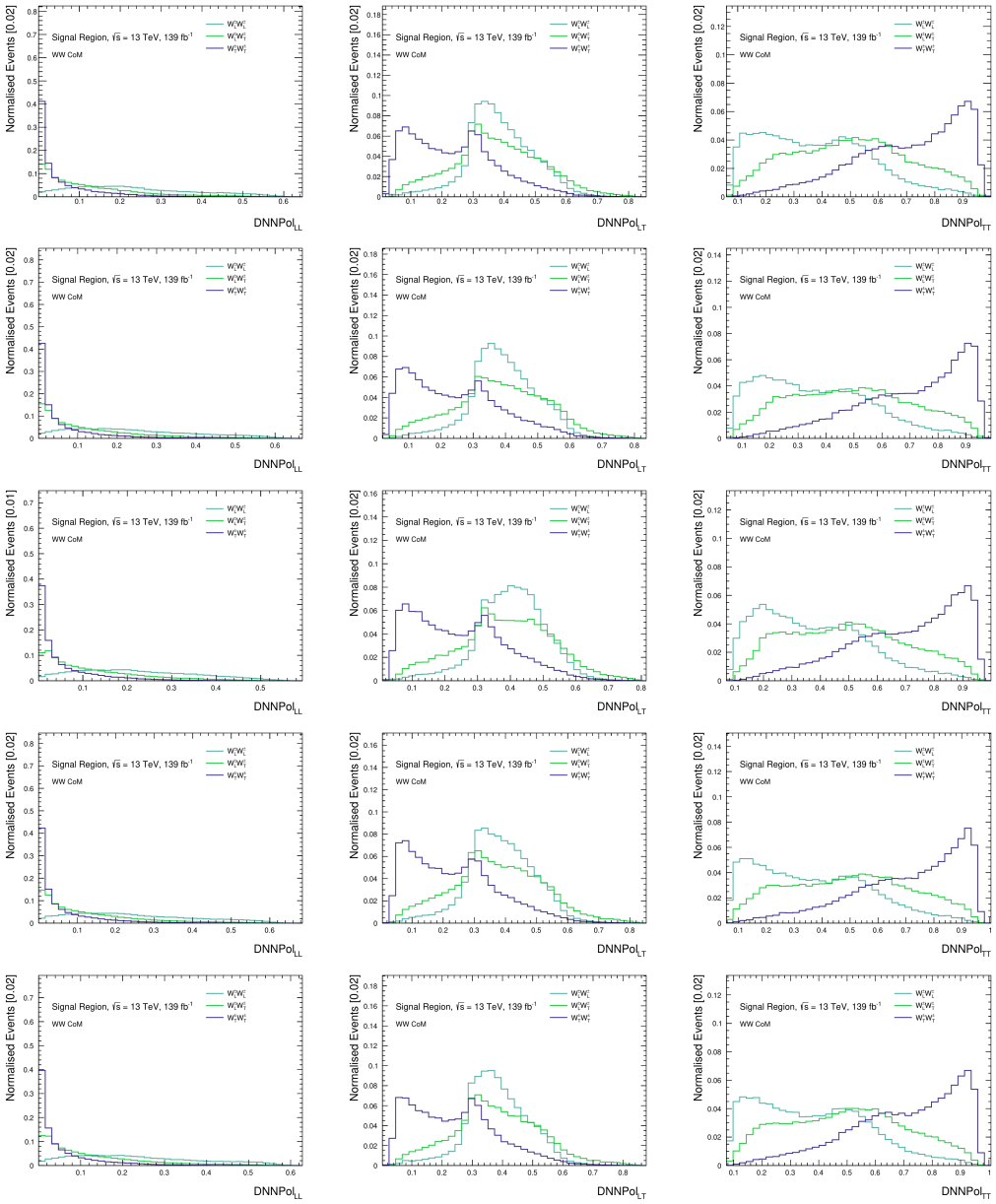


Figure 74: The normalised predictions of the final polarisation classifiers in the WW CoM frame after the full cross-validation procedure. The name of each output is subscripted by the polarisation configuration. The hyperparameters of the networks are given in Table 28, labelled as DNNPol.

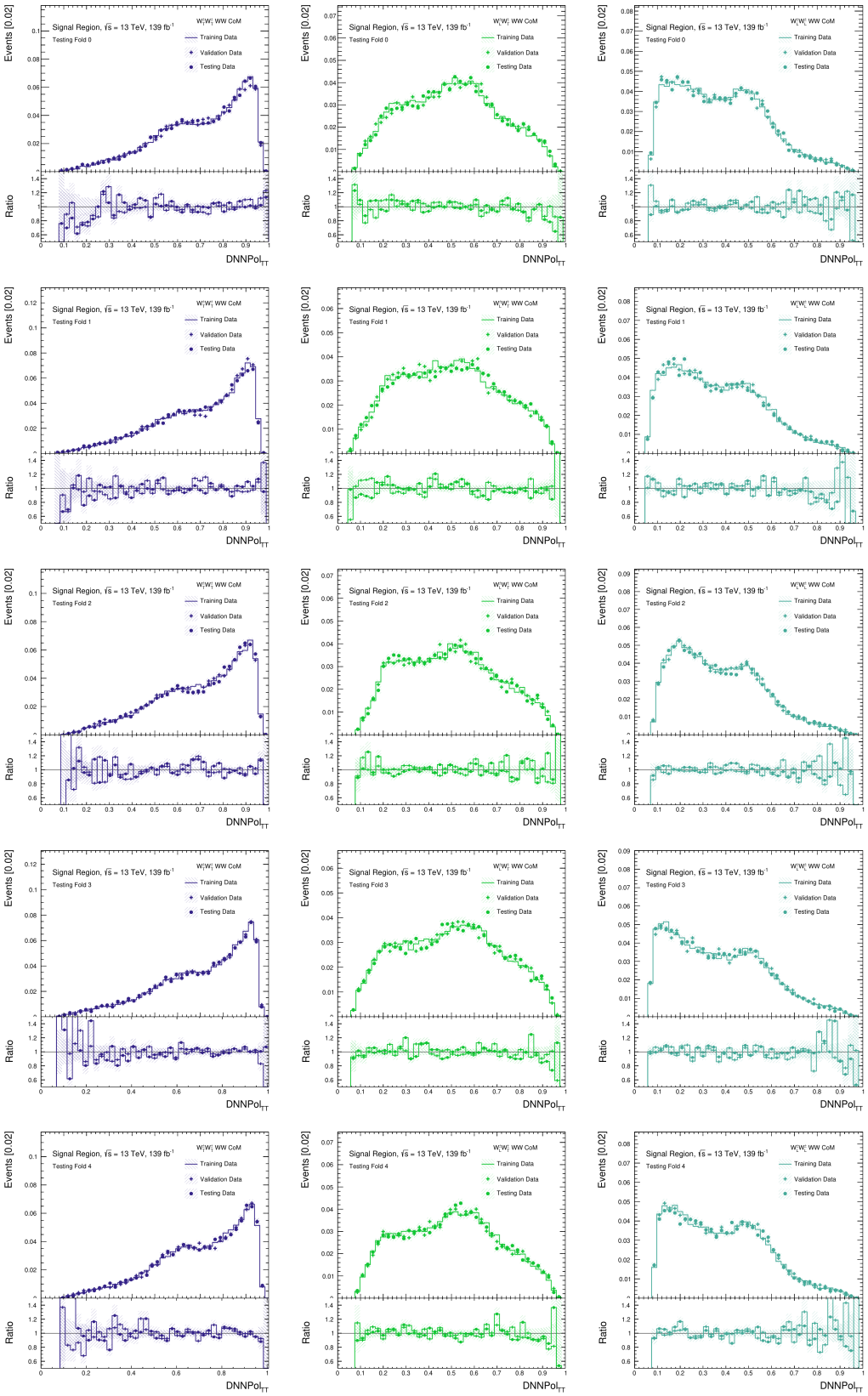


Figure 75: The normalised predictions of the polarisation classifiers in the WW CoM frame, shown for the network TT output for the polarisation samples split into training, testing and validation data. The hyperparameters of the networks are given in Table 28 labelled as DNNPol.

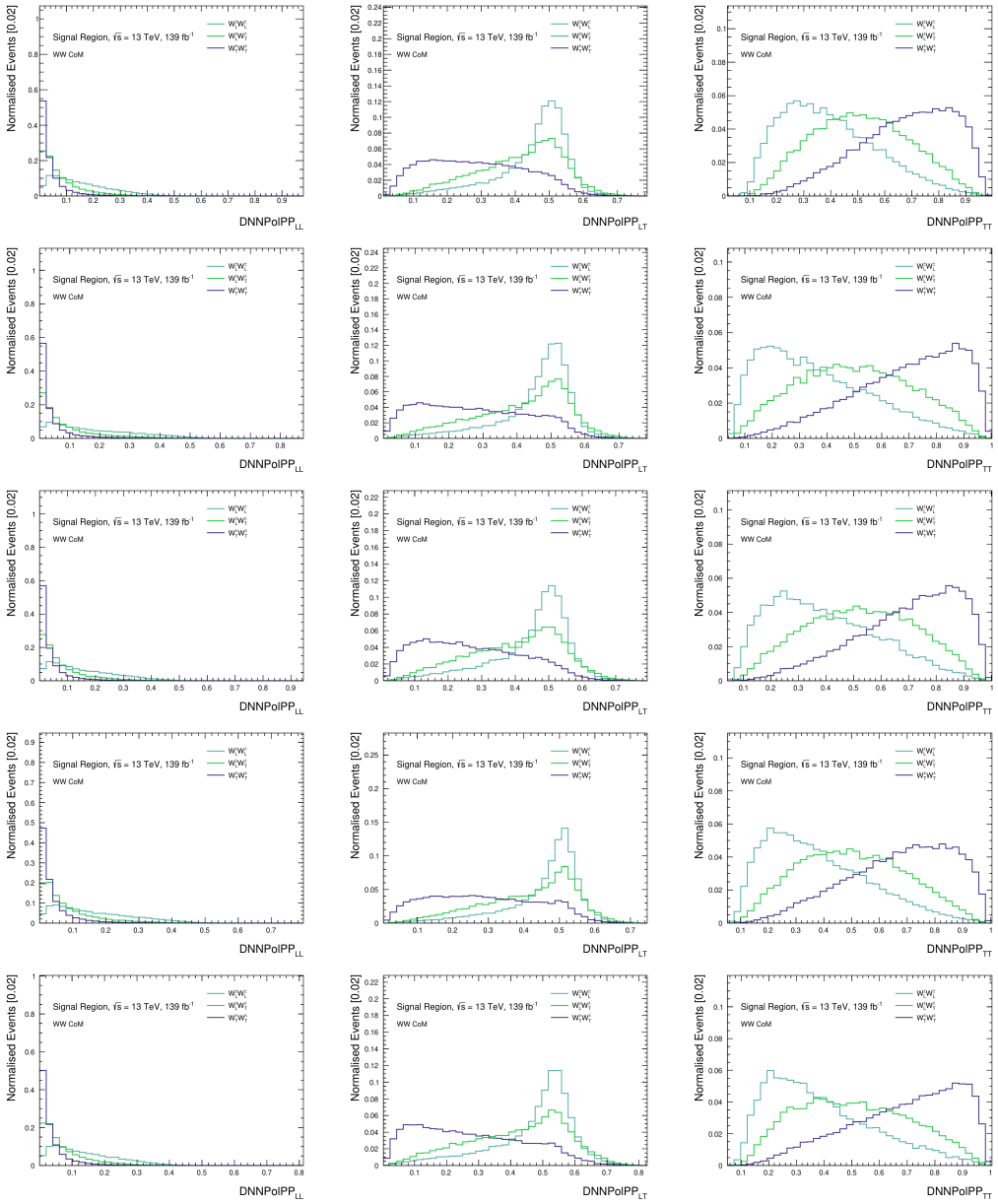


Figure 76: The normalised predictions of the final polarisation classifiers in the *pp* CoM frame after the full cross-validation procedure. The name of each output is subscripted by the polarisation configuration. The hyperparameters of the networks are given in Table 28 labelled DNNPolPP.

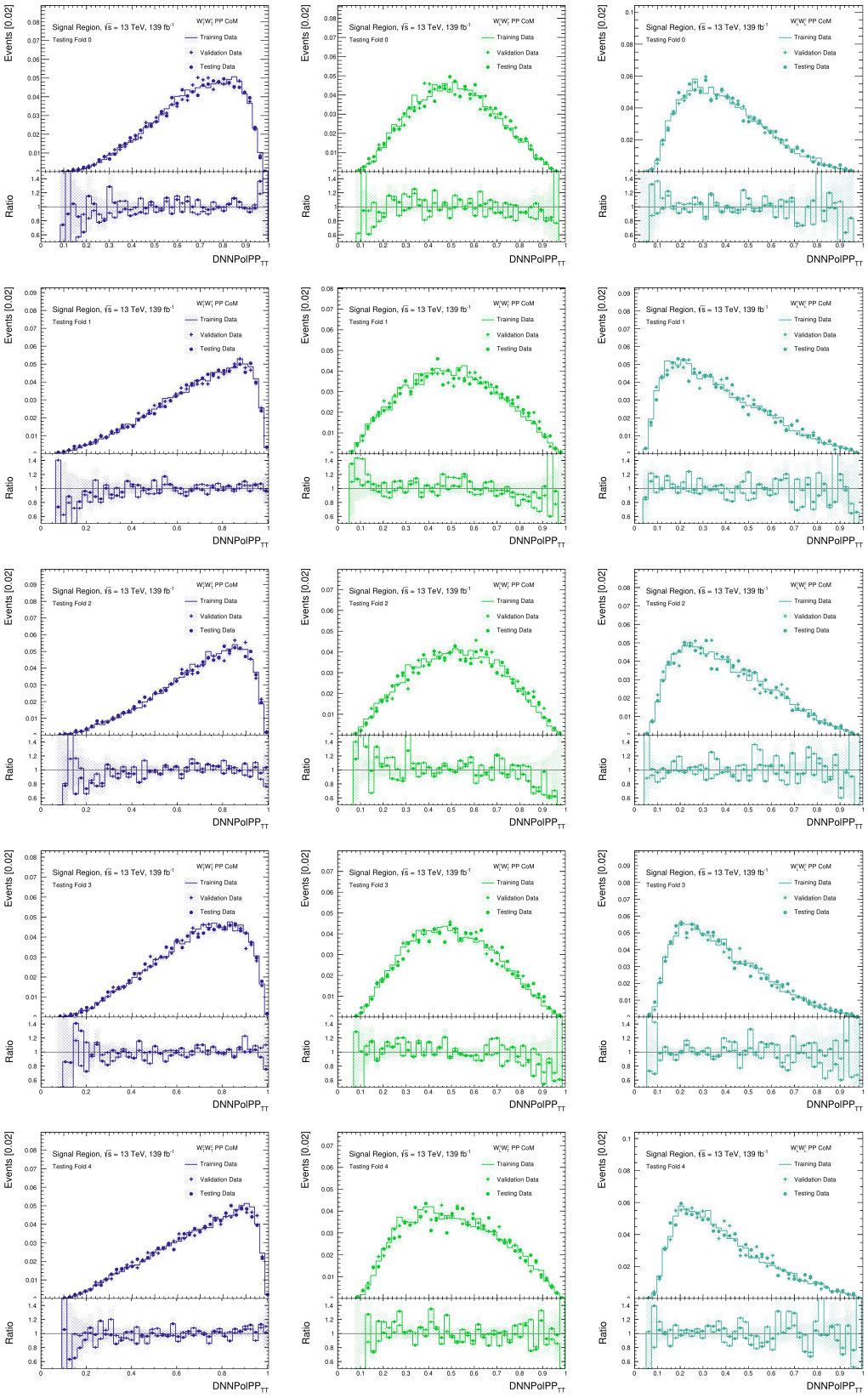


Figure 77: The normalised predictions of the polarisation classifiers in the *pp* CoM frame, shown for the network TT output for the polarisation samples split into training, testing and validation data. The hyperparameters of the networks are given in Table 28 labelled as DNNPolPP.

7.9 Validation

This section illustrates comparisons of the network MC predictions for the inclusive electroweak $W^{\pm}W^{\pm}jj$ MC versus data in the analysis signal region. This section also demonstrates the network predictions in the validation regions of the analysis. The validation procedure is an important test to check for any mismodelling. Comparisons between data and background in the validation regions test the modelling in regions unseen by the network. For these comparisons, only statistical uncertainties are shown.

7.9.1 Classifiers Distributions

The section illustrates the model predictions against experimental data in three validation regions: the low-Δy_{jj}, top-fakes and WZ validation regions. Figures 78, 79 and 80 show the $\text{DNNPol}_{\text{TT}}$ distributions against data in each validation region. For each validation region, the output is shown in each fold of the cross-validation procedure.

The predictions in the low-Δy_{jj} validation region test the modelling of the network distribution for backgrounds dominated by QCD processes. For this analysis, the dominant backgrounds are the QCD-induced $W^{\pm}Zjj$ background, the non-prompt lepton background and the charge-flip background. The figures show good agreement between data and the background predictions, with no evident mismodelling. The distributions show consistent modelling between the folds within the bounds of statistical uncertainty. The non-prompt lepton background and the top-quark background dominate the top-fakes validation region. Figure 79 shows a slight excess in the data. However, the modelling of the network output distribution is good. Finally, the prediction in the WZ validation region, used in this experiment to extract the normalisation factor for the QCD-induced $W^{\pm}Zjj$ background, is shown in Fig. 80. The region is dominated primarily by QCD-induced $W^{\pm}Zjj$ events. There is an excess in the MC predictions compared to the data discussed in Chapter 6. However, the general agreement, excluding this rate mismodelling, is good. Overall, all of the models show no pronounced mismodelling in unseen regions.

Finally, the WW CoM frame classifier prediction for the signal region of the analysis is shown in Fig. 81 for $\text{DNNPol}_{\text{TT}}$ distribution. The distributions show good shape agreement between the data and the signal and background modelling within statistical uncertainty. However, there are discrepancies in individual bins and an underestimation of the overall predicted number of events against data. There appears to possibly be an underestimation of the number of TT events, as the mismodelling seems to be present in the distribution for $\text{DNNPol}_{\text{TT}}$ values close to one.

Similarly, for the DNNPolPP classifier, the modelling in the low-Δy_{jj}, top-fakes region and the WZ validation region is shown in Figs. 82, 83 and 84 for the $\text{DNNPolPP}_{\text{TT}}$ distribution. From the distributions, there appears to be good agreement between the modelling of the experiment backgrounds and the measured data. There is no pronounced mismodelling in either of the validation regions, and the general distributions appear to agree with the data.

The prediction in the signal region of the analysis for the $\text{DNNPol}_{\text{TT}}$ distribution is shown in Fig.85 respectively. As with the classifiers in the WW CoM, there appears to be good agreement between the experimental data and the background modelling. The overall modelling of the signal and background with the classifiers appears to reasonably

agree with the data. There are some discrepancies and what appears to be an underestimation of data. The figures suggest there could be some potential mismodelling in the TT samples.

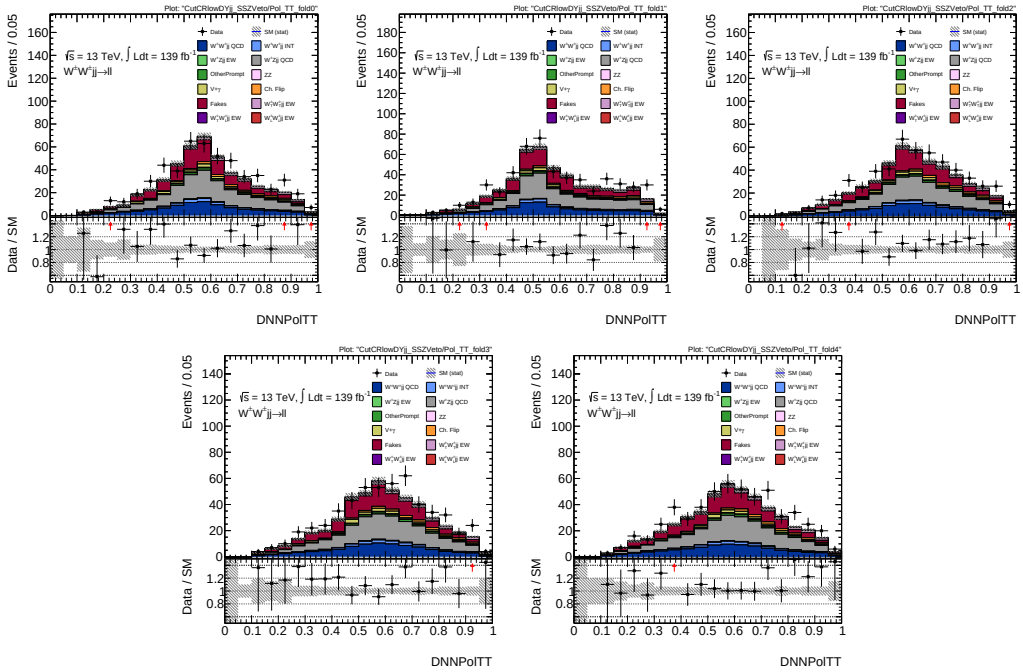


Figure 78: The predicted event yields in the low-ΔY_{jj} validation region against data for the polarisation classifier $\text{DNNPol}_{\text{TT}}$. The uncertainty shown here is statistical. The hyperparameters of the network are given in Table 28.

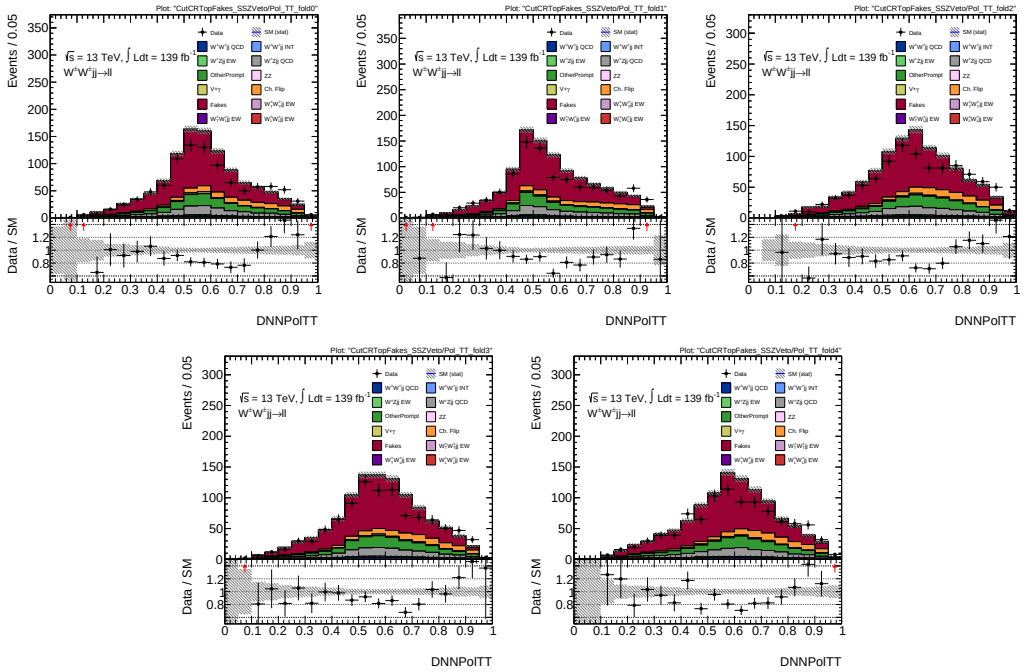


Figure 79: The predicted event yields in the top-fakes control region against data for the polarisation classifier $\text{DNNPol}_{\text{TT}}$. The uncertainty shown here is statistical. The hyperparameters of the network are given in Table 28.

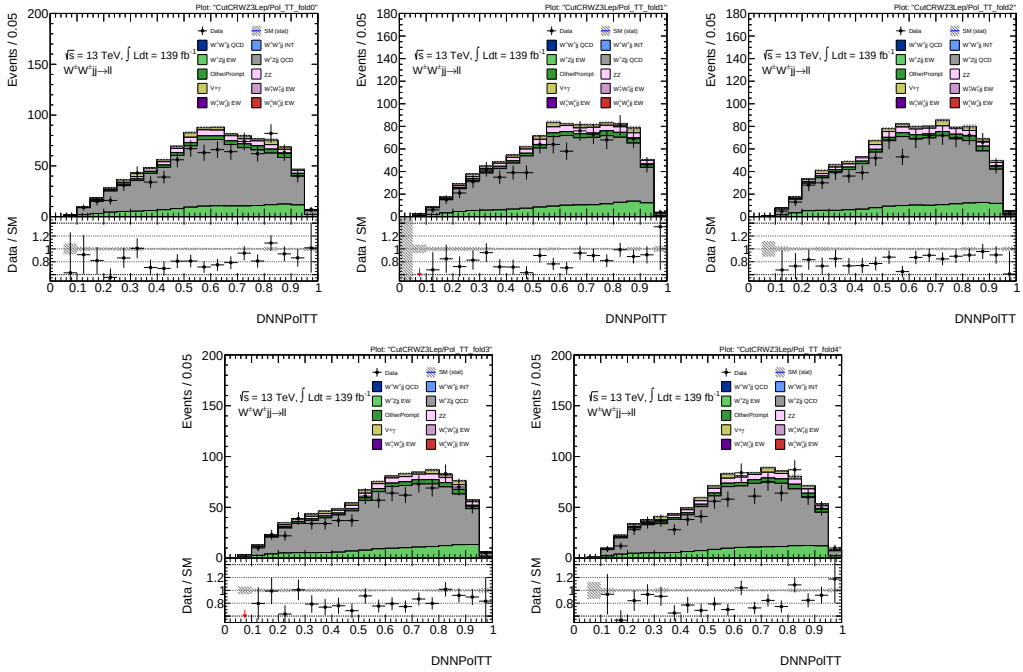


Figure 80: The predicted event yields in the WZ validation region against data for the polarisation classifier $\text{DNNPol}_{\text{TT}}$. The uncertainty shown here is statistical. The hyperparameters of the network are given in Table 28.

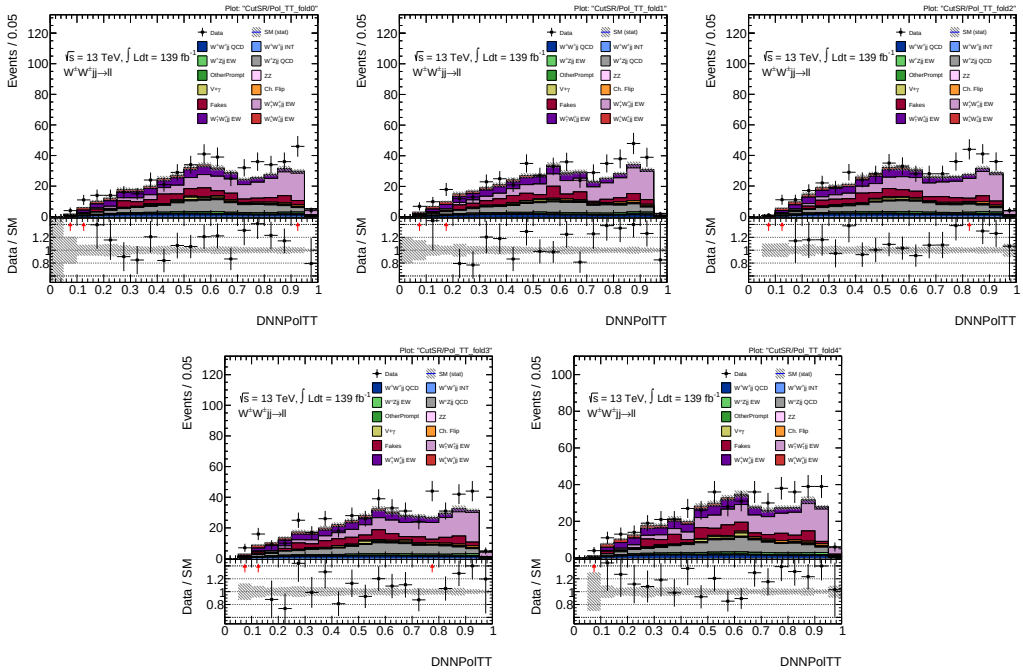


Figure 81: The predicted event yields in the signal region against data for the polarisation classifier $\text{DNNPol}_{\text{TT}}$. The uncertainty shown here is statistical. The hyperparameters of the network are given in Table 28.

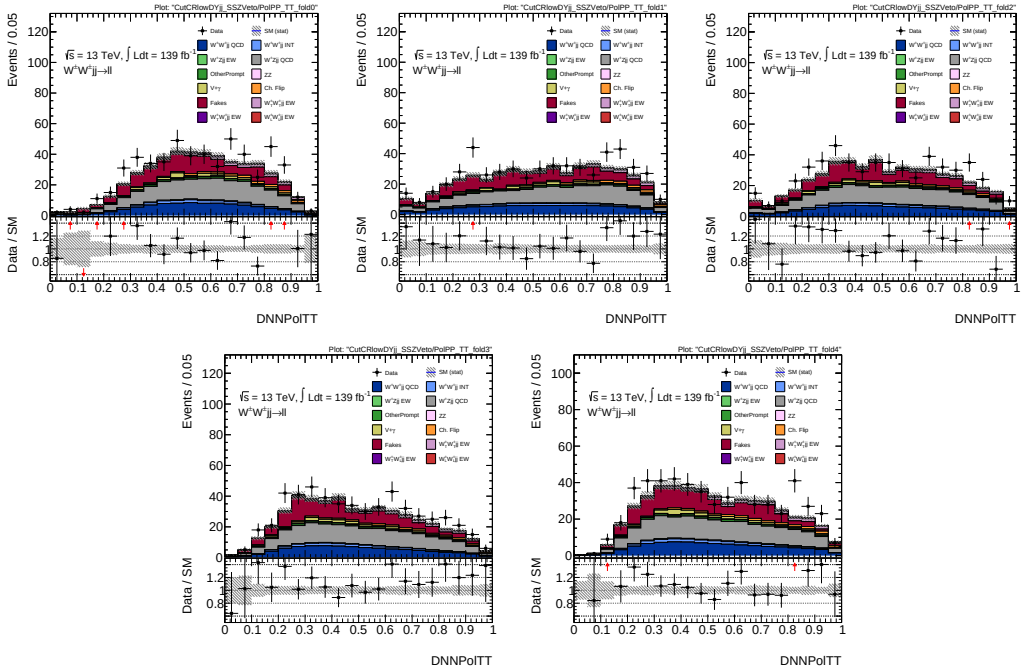


Figure 82: The predicted event yields in the low-ΔY_{jj} validation region against data for the polarisation classifier $\text{DNNPolPP}_{\text{TT}}$. The uncertainty shown here is statistical. The hyperparameters of the network are given in Table 28.

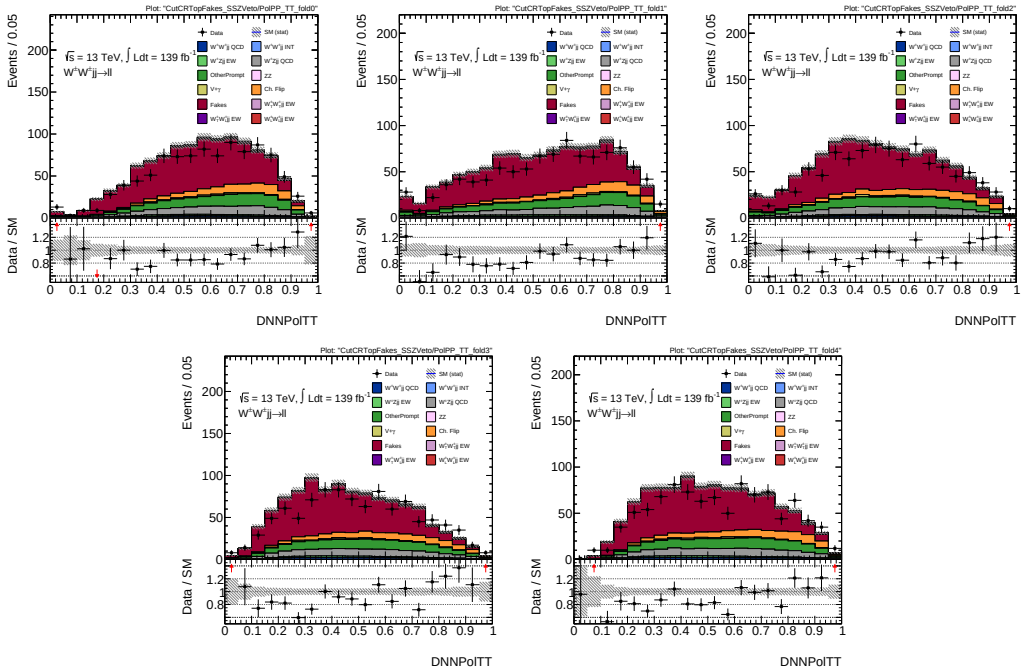


Figure 83: The predicted event yields in the top-fakes validation region against data for the polarisation classifier $\text{DNNPolPP}_{\text{TT}}$. The uncertainty shown here is statistical. The hyperparameters of the network are given in Table 28.

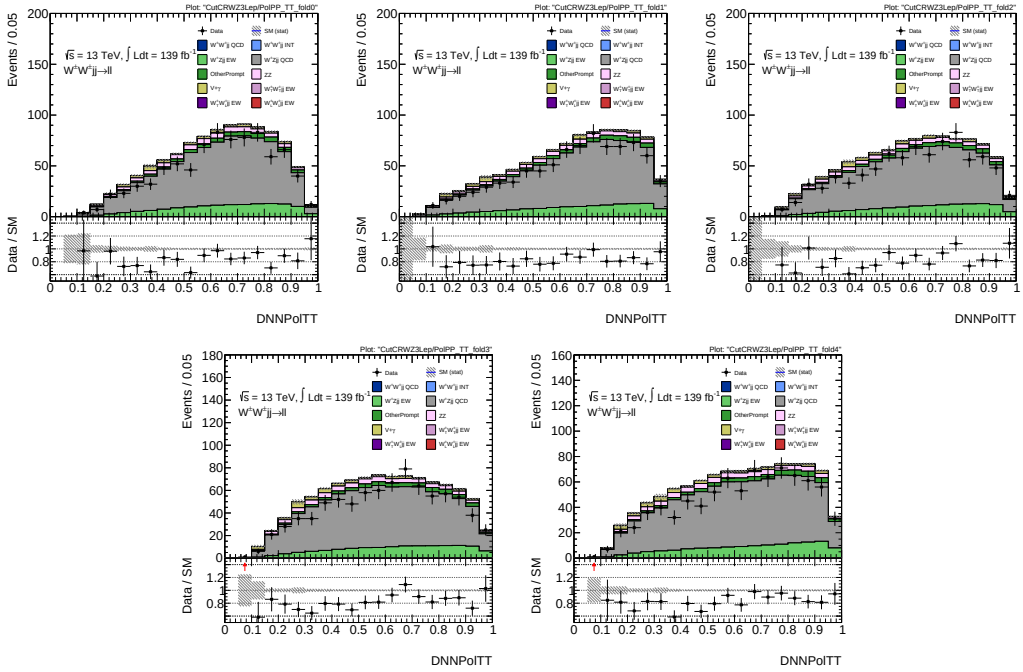


Figure 84: The predicted event yields in the WZ validation region against data for the polarisation classifier $\text{DNNPolPP}_{\text{TT}}$. The uncertainty shown here is statistical. The hyperparameters of the network are given in Table 28.

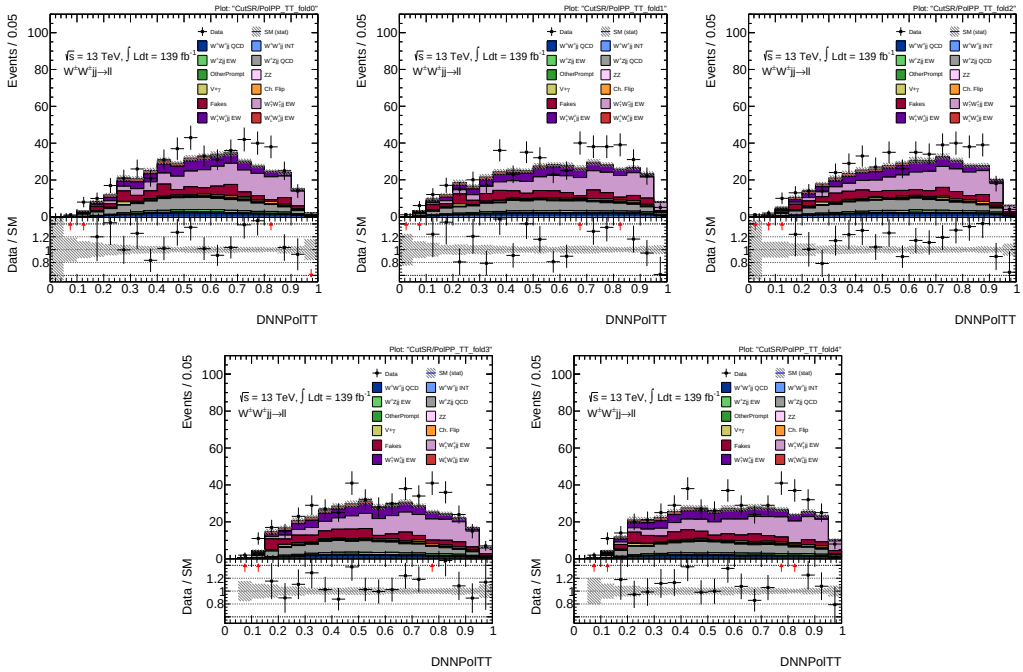


Figure 85: The predicted event yields in the signal control region against data for the polarisation classifier $\text{DNNPolPP}_{\text{TT}}$. The uncertainty shown here is statistical. The hyperparameters of the network are given in Table 28.

8 Electroweak Classifiers

This measurement employs a series of multivariate analysis techniques to enhance sensitivity to polarisation. As introduced in the previous chapter, this analysis applies two classification algorithms using DNNs. The first algorithm is applied to separate the inclusive EW $W^{\pm}W^{\pm}jj$ signal from the analysis backgrounds, and a second algorithm distinguishes the EW $W^{\pm}_{\lambda}W^{\pm}_{\lambda}jj$ polarisation configurations. In this thesis, the algorithms are referred to as the electroweak and polarisation classifiers. Chapter 7 detailed the training procedure of the polarisation classifier, and this chapter will analogously describe the training of the electroweak classifier.

The reader is referred to Section 7.2 for an introduction to the machine learning concepts and nomenclature used in this chapter. The structure of this chapter will follow analogously to Chapter 7. The training methods and procedures for the electroweak classifiers are described in Section 7.3.2. This chapter begins by providing a brief motivation for implementing an electroweak classifier and an introduction to the kinematics of EW-induced $W^{\pm}W^{\pm}jj$ events. The chapter will then outline the training procedure for the electroweak classifier, separating the training into a first and second pass. The chapter concludes with the validation results.

8.1 Motivation

The background estimation procedures in this analysis are separated into Monte Carlo and data-driven methods. The events used to train the classifiers are generated with MC, except for those modelling the charge-flip background. The charge-flip events are generated using the data-driven method described in Chapter 5.7.3. The modelling of the fake-leptons background is not suitable for training. This background is estimated by subtracting prompt events in the signal region, leading to events with negative event weights. This is impractical as the event weights are used to scale the training data to the expected event yields in the signal region, and the effect of scaling the loss function with negative values is unclear.

This section discusses the kinematics of inclusive EW $W^{\pm}W^{\pm}jj$ production and the analysis backgrounds in order to highlight kinematic variables a DNN could utilise to discriminate the EW signal. The first part of this section will focus on the most prominent contributors, namely the $VVjj$ backgrounds. The most significant background contributions in the signal region are the QCD- and interference-induced $W^{\pm}W^{\pm}jj$ production and the EW- and QCD-induced $W^{\pm}Zjj$ production. The QCD-induced contribution is the largest for both processes. Other prompt contributions coming from diboson and triboson production or from top quarks are also included in the training. However, these backgrounds have much smaller contributions to the total event yield and low event yields in the MC.

8.1.1 $VVjj$ Backgrounds

Figure 86 compares the MC predictions for the polarised EW $W^{\pm}W^{\pm}jj$ samples to those for the $VVjj$ backgrounds. The figure shows the normalised distributions of the dijet

invariant mass and the ΔR distribution. As expected of a VBS process, the electroweak $W^{\pm}W^{\pm}jj$ polarised events show distributions with a larger dijet invariant mass. This is because the electroweak interaction is colour-neutral, and the cross-section is maximised by minimising the momentum exchange between the incoming partons due to the electroweak gauge boson propagator term. The QCD production mode diagrams include a colour exchange between the incoming partons and additional interactions with gluons. The momentum exchange between the incoming partons is larger than for the electroweak interaction, leading to a smaller dijet invariant mass.

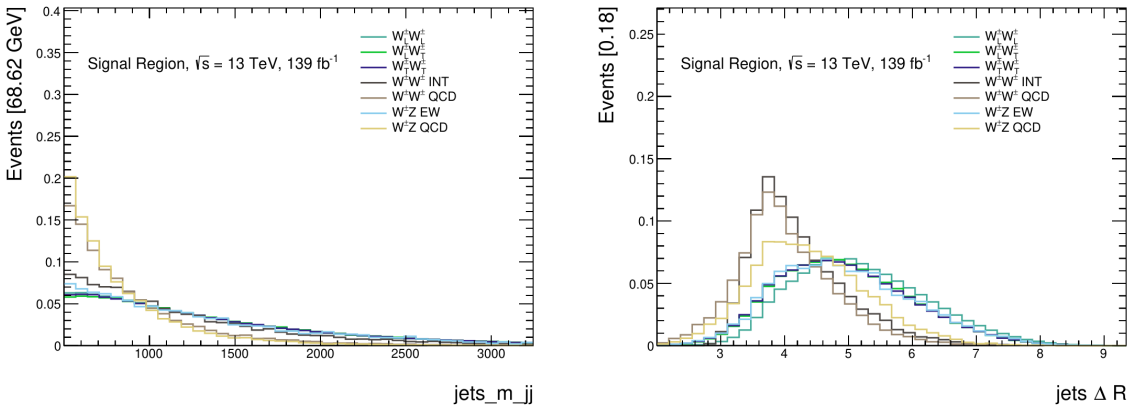


Figure 86: The normalised di-jet invariant mass m_{jj} and di-jet ΔR distributions shown on the left and right, respectively, for the MC predictions of the polarised $W^{\pm}W^{\pm}$ signal samples compared to the $VVjj$ backgrounds.

As discussed in Chapter 4, VBS processes share some distinctive features in their event topologies. The diagrams contributing to each production mode of $W^{\pm}Zjj$ are very similar to those of $W^{\pm}W^{\pm}jj$ within their charge requirements. The similarities between each process are exemplified in the dijet invariant mass distribution shown in Fig. 86. The distribution shows similarities between the QCD- and EW-induced $W^{\pm}Zjj$ and $W^{\pm}W^{\pm}jj$ production modes. This indicates that discriminating the EW-induced $W^{\pm}Zjj$ background from the EW signal will be challenging for the electroweak classifier.

Figure 87 compares the normalised distributions of the rapidity of the leading jet in p_T and the minimum ΔR between the leptons and the jets. Compared to EW production, the leading jet rapidity shows evident shape differences with the QCD- and interference-induced production. Kinematic variables involving the leptons are typically more sensitive to polarisation effects but can also discriminate the EW signal. The lepton-jet minimum ΔR distribution shown in Fig. 87 shows separation between the QCD- and EW-induced $W^{\pm}W^{\pm}jj$ production modes as well as small polarisation discrimination.

For the electroweak classifier, input features are preferable without large polarisation effects. The classifier is trained using the combined polarised samples as the inclusive electroweak process. Due to the event weighting, the classifier is expected to prioritise learning features to separate the QCD- and EW-induced VBS processes. However, polarisation effects can dilute the separation between QCD- and EW-induced production. This is illustrated in the lepton $\Delta\eta$ distribution of Fig. 88. The distribution shows an overlap

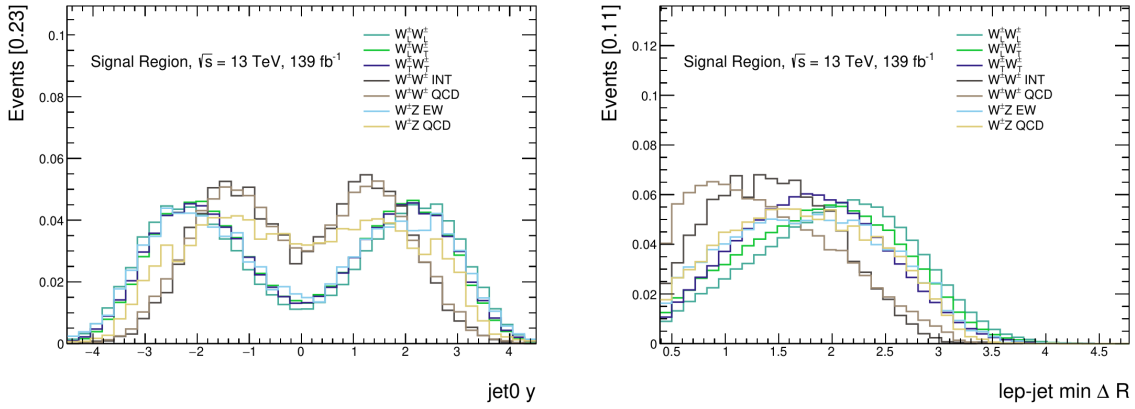


Figure 87: The normalised distributions of leading jet rapidity and lepton-jet min ΔR shown on the left and right, respectively, for the MC predictions of the polarised $W^{\pm}W^{\pm}$ signal samples compared to the $VVjj$ backgrounds.

between the $W^{\pm}Zjj$ QCD production mode and the fully transverse $W^{\pm}W^{\pm}jj$ polarisation mode. As the event yields of the EW signal consist predominately of transverse polarisation, the lepton $\Delta\eta$ distribution is not expected to be a useful input feature.

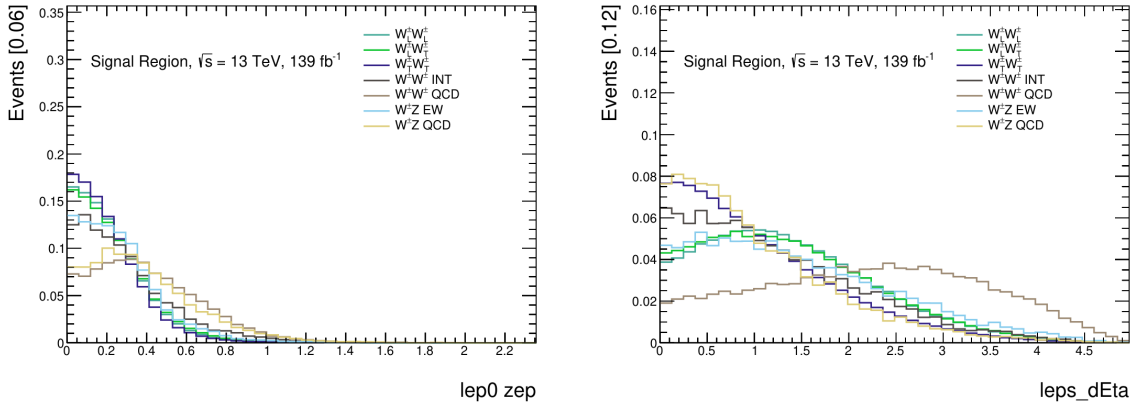


Figure 88: The normalised distributions of leading lepton Zeppenfeld variable and dilepton $\Delta\eta_{ll}$ shown on the left and right, respectively, for the MC predictions of the polarised $W^{\pm}W^{\pm}$ signal samples and the $VVjj$ backgrounds.

8.1.2 Prompt and Charge-Flip Backgrounds

The remaining backgrounds considered for training arise from prompt and charge flip backgrounds. The prompt backgrounds include any processes which can produce two same-sign leptons. This includes diboson processes, namely ZZ, triboson processes with VVV, excluding the same-sign triboson contribution, and top production through $t\bar{t}V$ and tZq. The prompt contribution makes up less than 2% of the total event yield in the SR. The charge-flip background arises from the misidentification of electron charge. The main processes contributing to this background are Z boson production through Drell-Yan and

dileptonic $t\bar{t}$ production with electrons. The event yield contributed by the charge-flip background is similarly of order 2%. Both of these backgrounds have very small event yields relative to contributions from the $VVjj$ backgrounds and the signal events. This means the relative contribution to learning from these backgrounds is expected to be small.

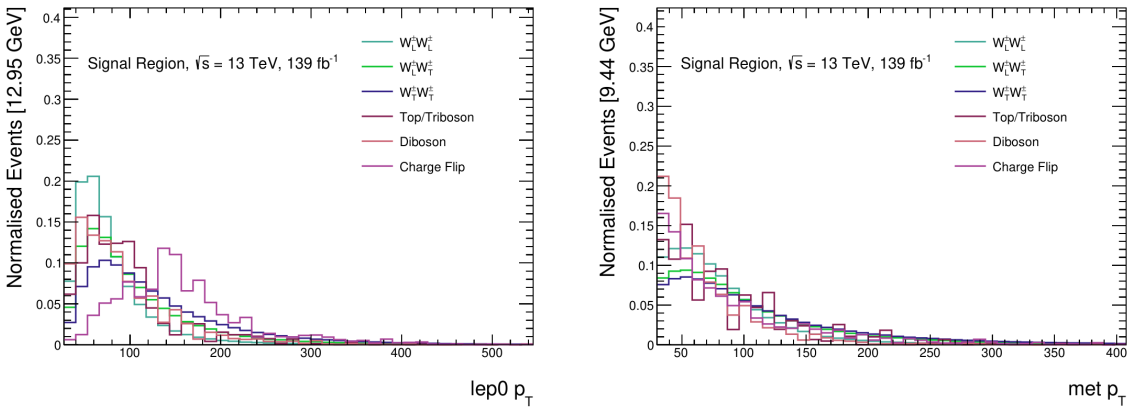


Figure 89: The normalised distributions of leading lepton p_T and the E_T^{miss} shown on the left and right, respectively, for the MC predictions of the polarised $W^{\pm}W^{\pm}$ signal samples compared to the $VVjj$ backgrounds.

Figure 89 shows the normalised distributions for the leading lepton p_T and the E_T^{miss}, comparing the prompt and charge flip backgrounds to the polarised EW $W^{\pm}W^{\pm}jj$ signal. The polarised EW signal samples show an E_T^{miss} distribution with a higher mean than those of the prompt and charge-flip backgrounds. The selection criteria in the signal region require high-p_T leptons, which results in a large E_T^{miss}.

The leading lepton pT distribution was shown in the previous chapter to be a good discriminator for polarisation. However, the separation between the prompt backgrounds and the electroweak signal is small. This variable could still be useful for an electroweak classifier as the charge-flip background has a significantly large mean p_T distribution. It is unclear which variables will be most impactful to the classification algorithm beyond intuition, which motivates an input feature selection procedure.

Figure 90 shows the normalised distributions of the sub-leading lepton type and the sub-leading lepton Zeppenfeld variable [127]. In the distribution, a lepton type of 6 corresponds to an electron/positron and a lepton type of 8 to a muon/anti-muon. The charge-flip background primarily includes processes which produce prompt electrons. The normalised lepton-type distribution shows far more events produced in the electron channel than the electroweak signal. This is also observed in the diboson and prompt background. Another interesting discriminating variable is the Zeppenfeld variable, shown in Fig. 90 for the sub-leading lepton. The variable shows a separation between the polarised electroweak signal samples and the prompt and charge-flip backgrounds.

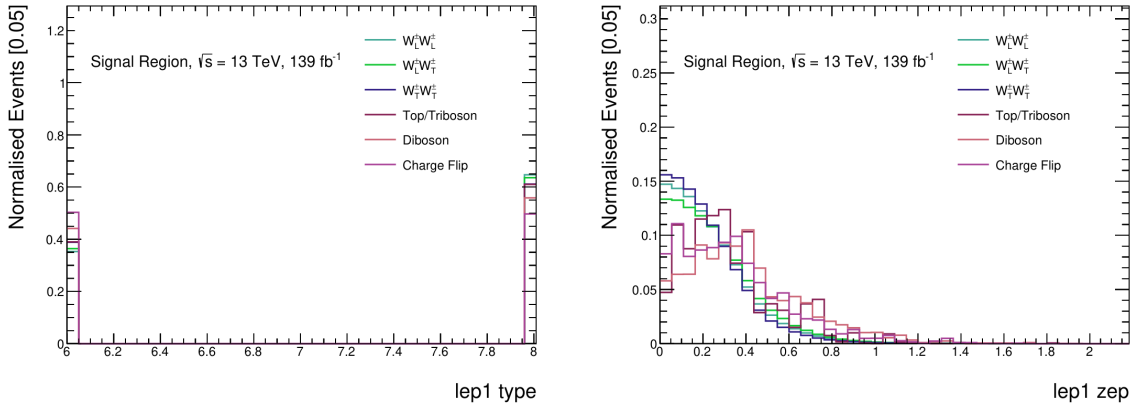


Figure 90: The normalised distributions of leading and sub-leading lepton types shown on the left and right, respectively, for the MC predictions of the polarised $W^{\pm}W^{\pm}$ signal samples compared to the $VVjj$ backgrounds.

8.2 Electroweak Classifier

8.2.1 Network Description

The electroweak classifier is applied to isolate the inclusive EW $W^{\pm}W^{\pm}jj$ signal to maximise the effectiveness of the polarisation classifier. The two classifiers are implemented together to generate a two-dimensional distribution. The electroweak classifier constrains the analysis backgrounds to one region of the distribution. The polarisation classifier separates the polarisation configurations in the region dominated by the electroweak $W^{\pm}W^{\pm}jj$ signal. The same electroweak classifier is used for all measurements in this experiment. The electroweak classifier is trained by combining polarisation samples in the WW CoM frame. The summed distributions are equivalent to the unpolarised $W^{\pm}W^{\pm}jj$ EW signal and are Lorentz invariant.

8.2.2 Training Procedure

The training procedure for the electroweak classifiers is the same as described in Section 7.3.2.

8.2.3 Samples

Table 30 provides the raw events and the corresponding predicted event yields used for training. The nominal and differential samples $W^{\pm}Zjj$ QCD, discussed in Chapter 6, are combined for the training data since the distributions of the input variables show negligible differences.

8.2.4 Input Variables

The list of input features tested for the electroweak classifiers is given in Table 22. A total of 32 input variables were trialled during training. As with the polarisation classifiers, the

Training Samples		
Sample	Raw Events	Predicted Events
$W^{\pm}W^{\pm}jj$ (WW CoM)		
$W_L^{\pm}W_L^{\pm}jj$	215008	18.4
$W_L^{\pm}W_T^{\pm}jj$	170387	58.6
$W_T^{\pm}W_T^{\pm}jj$	456456	124.2
Backgrounds		
$W^{\pm}Zjj$ QCD	48922	57.9
$W^{\pm}Zjj$ EW	8255	14.9
$W^{\pm}W^{\pm}jj$ QCD	25094	24.0
$W^{\pm}W^{\pm}jj$ INT	78981	7.59
Charge-Flip	8231	10.0
Diboson	2781	2.50
Top + Triboson	1132	5.02

Table 30: The training data split into signal and background samples used to train the electroweak classifier.

input features are scaled to be normally distributed. Scaling ensures that input features are treated equally and are not biased to large values.

8.3 First Pass Training

8.3.1 Trialled Hyperparameters

The network architectures trialled in the first pass training are the same as for the polarisation classifier, described in Section 7.6.1. The network hyperparameters are optimised using Bayesian optimisation; see Section 7.4.3. The set of parameters tested for the first pass electroweak classifiers are shown in Table 31. A slightly more extensive set of networks is tested compared to the polarisation classifiers in the WW CoM, reflecting the larger training dataset. Each electroweak classifier is trained by minimising the binary cross-entropy loss function. The activation function for the output is the sigmoid function, which normalises the network output to give a probability distribution. The optimisation procedure is only applied for the first cross-validation fold. Only the final best-selected network is trained with the complete cross-validation method. The networks are optimised using the Adam optimiser. Each hyperparameter associated with the optimiser is included in the search range.

8.3.2 Optimisation Results

The hyperparameters of the best-performing network with respect to the validation AUC for the first-pass training procedure of the electroweak classifier are shown in Table 32.

Trialled Hyperparameters	
Parameter	Range
Layers	[7, 13]
Nodes	[60, 150]
Activation Function	ELU, swish
Output Function	sigmoid
Number of outputs	1
Dropout	[0.1, 0.4]
β_1	[0.001, 0.99]
β_2	[0.001, 0.99]
ϵ	$[1 \times 10^{-3}, 1 \times 10^{-8}]$
Learning rate	$[1 \times 10^{-1}, 1 \times 10^{-5}]$
Batch Size	[32, 512]

Table 31: The trialled ranges of hyperparameters tested for the electroweak classifier.

During the first pass, approximately 8000 networks were trained and tested in the optimisation procedure. The highest recorded AUC for the model shown in the table is 0.8203 ± 0.0001. The network size and parameters are similar to the polarisation classifiers in the WW CoM. The dataset used for training the electroweak classifier is slightly larger. The similarity in the network sizes reflects the similarities in the size of the training datasets.

Best Parameters	
Parameter	Electroweak Classifier
Inputs	32
Layers	8
Nodes	109
Activation Function	ELU
Output Function	Sigmoid
Number of outputs	1
Dropout	0.206738
β_1	0.750892
β_2	0.540354
ϵ	0.0000001
Learning rate	0.009860
Batch Size	247
Epochs	74

Table 32: The hyperparameters associated with the electroweak classifier with the highest validation AUC following the first-pass training.

The entire cross-validation procedure is performed for the model hyperparameters shown in Table 32. This step is performed as an additional validation step to test the training methodology. The validation AUC of the trained model for each of the five folds is shown in Table 33. The table shows consistent performance across each fold, with only a slight decrease for Fold 1.

Validation AUC					
	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4
	0.8220	0.8189	0.8229	0.8242	0.8226

Table 33: The validation AUC for the full cross-validation procedure of an electroweak classifier trained with the hyperparameters described in Table 32. The statistical uncertainty on the AUC is of order 10^{-4}.

8.4 Input Feature Selection

Similarly to Section 7.7, this section details the input variable selection procedure applied after the first-pass training. The procedure applies the INVASE model to find an optimal subset of input variables for the electroweak classifier. A detailed description of the INVASE architecture is given in Sec. 7.5. The model is parameterised with the INVASE parameter λ, which controls the strength of the input feature selection.

8.4.1 Training Procedure

The network parameters chosen in the first pass training are used as the hyperparameters for the predictor and baseline networks of the INVASE model. The hyperparameters are detailed in Table 32. Likewise, Table 34 outlines the hyperparameters selected for the actor. As discussed in Section 7.7, the actor size is selected based on initial testing. A slightly bigger network is applied to the electroweak classifier owing to the larger dataset. The actor is trained using the Adam optimiser. The parameters of the optimiser are chosen as the default parameters to simplify the training procedure. The only varied parameter is the number of nodes in the network and the parameter . An early stopping condition is applied during training, monitoring the actor validation loss. The training procedure is halted after 12 epochs of no improvement to the validation loss.

8.4.2 Results

A total of 10 individual INVASE models were trained for the electroweak classifier, each differing primarily by the λ parameter. As outlined in Sec. 7.5, the actor chooses the final set of outputs. The actor predicts the probability that a variable should be selected for a given set of events. Variables predicted with a mean probability greater than 0.5 are chosen as input features. The INVASE model is controlled by the λ parameter. A larger parameter value leads to a smaller number of input features and a weaker convergence between the critic and baseline networks.

Actor Hyperparameters	
Parameter	Value
Inputs	32
Layers	2
Nodes	76, 152
Activation Function	elu
Output Function	sigmoid
Number of outputs	32
Batch Normalisation	True
β_1	0.9
β_2	0.99
ϵ	1×10^{-7}
Learning rate	0.001
Batch Size	247
λ	[0.000001, 0.5]

Table 34: The hyperparameters of the actor-network used during the INVASE procedure for the electroweak classifiers.

The input variable ranking, described in Sec. 7.7 is shown in Fig. 91. The variables are ranked based on the number of selections for each of the INVASE models, where the λ parameter of a variable increases with the number of selections.

The ranking is significantly different compared to the polarisation classifiers' ranking, shown in Fig. 70. The lepton kinematic variables, which are discriminating in terms of polarisation, are lower ranked in Fig. 91. By contrast, the jet kinematic variables are ranked higher. The event weighting of the electroweak classifier prioritises discriminating the electroweak signal from the QCD $VVjj$ backgrounds. Figure 92 shows the dijet invariant mass distribution and the dijet difference in rapidity for the training samples of the electroweak classifier. The distributions show that each variable is highly discriminating between the QCD $VVjj$ backgrounds and the electroweak signal.

In addition, other variables, such as the lepton type, are ranked higher for the electroweak classification. Figure 93 shows the normalised distributions for the leading and sub-leading lepton types. The variable shows good separation for some backgrounds, such as the background associated with charge-misidentified electrons. The lepton kinematic variables are expected to be ranked lower due to polarisation effects. As the samples are combined, large polarisation effects dilute the ability of the classifier to discriminate the electroweak distributions from the analysis backgrounds.

8.5 Second Pass Training

Following the input variable selection, the network training is repeated with a selected subset of input features. The subset is chosen by selecting a value of λ with the highest performing validation AUC for the critic network when training the INVASE procedure.

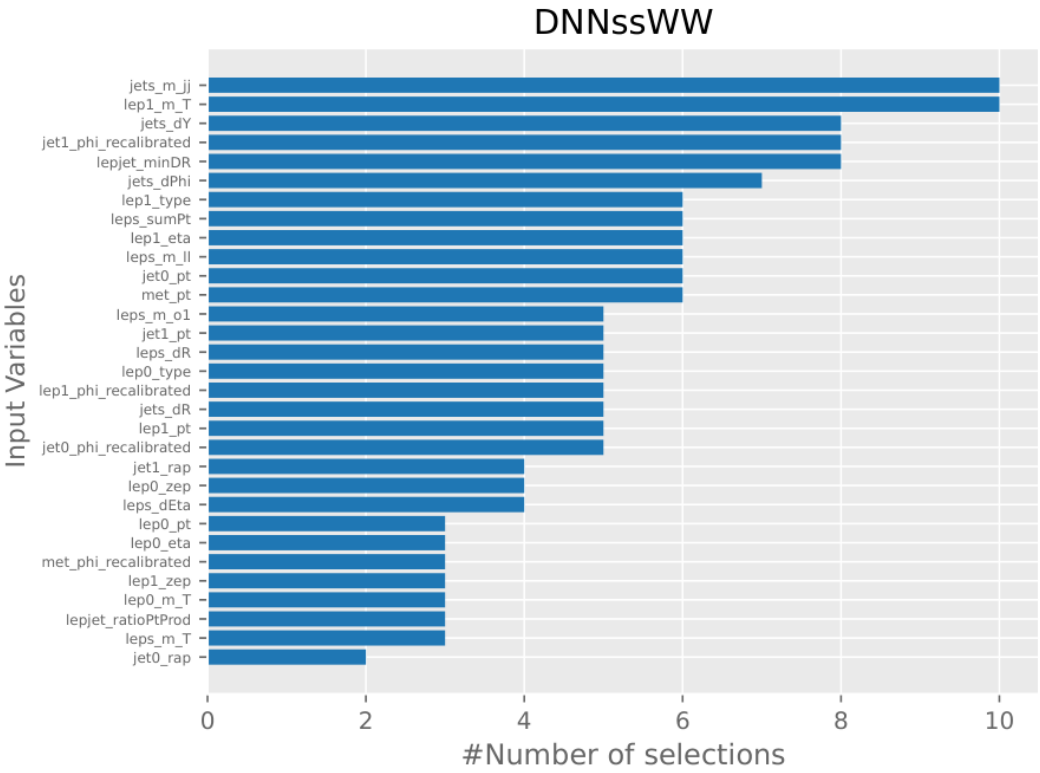


Figure 91: The ranking of the input features following the training of the INVASE models.

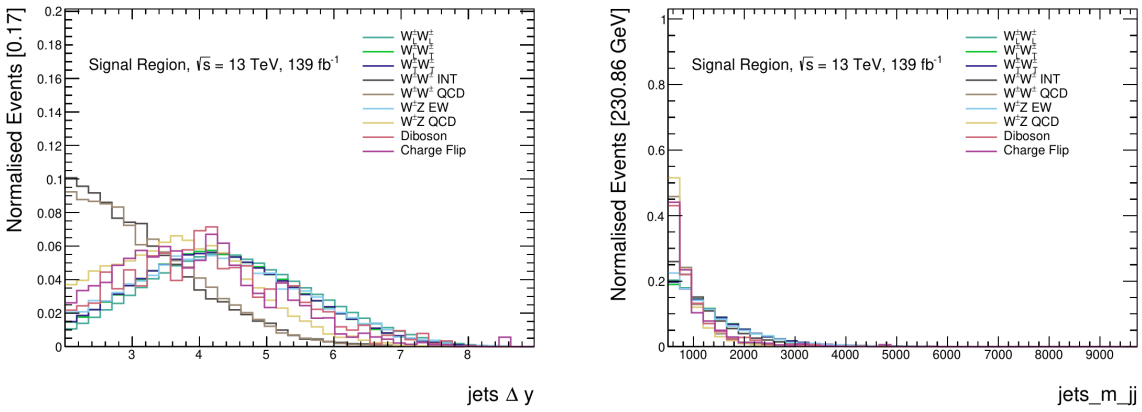


Figure 92: The normalised dijet ΔY and m_jj distributions.

However, in addition to this subset, further variables are included based on the previous chapter's ranking. The final set of input variables consists of 21 inputs and is illustrated in Table 36. Including additional input features reflects the convergence issues observed in the INVASE models. Table 36 also shows a more extensive selection of high-level variables compared to the polarisation classifiers. A sufficiently complex DNN should be able to learn the relevant relations between the low-level variables. The inclusion of a larger number of high-level variables suggests the networks trained for the electroweak

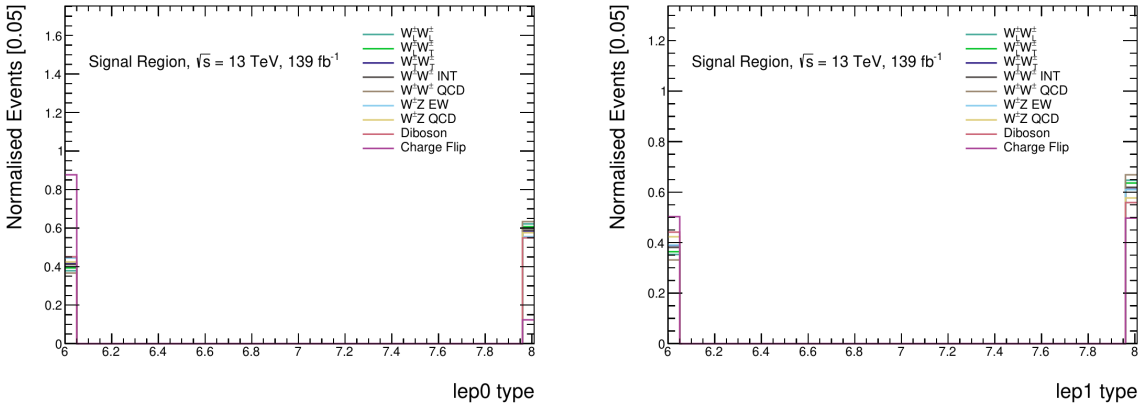


Figure 93: The normalised distributions of the leading and sub-leading lepton types.

classifiers are slightly lacking in complexity.

Hyperparameters	
Parameter	Range
Inputs	21
Layers	[7, 12]
Nodes	[60, 140]
Activation Function	elu, swish
Output Function	sigmoid
Number of outputs	13
Dropout	[0.15, 0.3]
β_1	[0.001, 0.99]
β_2	[0.001, 0.99]
ϵ	$[1 \times 10^{-3}, 1 \times 10^{-8}]$
Learning rate	$[1 \times 10^{-1}, 1 \times 10^{-5}]$
Batch Size	[200, 400]

Table 35: The range of hyperparameters tested during the second-pass training.

8.5.1 Trialled Hyperparameters

The same training procedure is repeated as in the first pass. The network architectures are as described in Sec. 7.6.1. The hyperparameter ranges trialled in the second pass are shown in Table 35. The classifiers are each trained using the Adam optimiser. The electroweak classifiers are minimised with the binary cross-entropy loss function. As in the previous section, the classifiers are also trained with an early stopping condition. The condition will stop training if there is no improvement in the validation loss for 12 consecutive epochs.

Input Variables		
Name	EW Classifier	Scaling
		Low-level variables
lep0 p_T	-	$\log_{10}(x)$
lep1 p_T	-	$\log_{10}(x)$
lep0 η	X	x
lep1 η	X	x
lep0 type	X	x
lep1 type	X	x
lep1 ϕ	-	x
jet0 p_T	X	$\log_{10}(x)$
jet1 p_T	X	x
jet0 y	-	x
jet1 y	-	x
jet0 ϕ	X	x
jet1 ϕ	X	x
E_T^{miss}	X	x
E_T^{miss}-ϕ	-	$\log_{10}(x)$
		High-level variables
lep0 zep	-	$\sqrt{x}$
lep1 zep	X	$\sqrt{x}$
leps $\Delta\eta$	-	x
leps charge	-	x
lep0 m_T	X	$\sqrt{x}$
lep1 m_T	X	$\sqrt{x}$
leps m_{ll}	-	$\log_{10}(x)$
leps m_T^{ll}	-	$\sqrt{x}$
leps p_T	X	$\log_{10} x$
leps ΔR	X	x
leps m_{o1}	X	$\sqrt{x}$
jets ΔY	X	x
jets $\Delta\phi$	X	x
jets ΔR	X	x
jets m_{jj}	X	$\log_{10}(x)$
lepton-jet p_T ratio	-	x
lepton-jet min(ΔR)	X	x

Table 36: The input feature selection of the electroweak classifier.

8.5.2 Results

During the second pass training, an approximate total of 4000 models were trained and tested. The models with the highest scoring validation AUC are selected to be used as the final networks for this measurement. The best-performing set of hyperparameters for the electroweak classifier is given in Table 37. The full cross-validation procedure is performed for this set of hyperparameters, and the classifier is referred to as DNNssWW for the remainder of this thesis. The final validation AUC for the complete training procedure is given in Table 38. The table shows consistent performance across each of the folds, with slight variation in the best-performing epoch. Moreover, the results show slight improvements and better consistency across the folds to the best network trained in the first pass. A similar improvement was also seen for the polarisation classifiers and

discussed in Section 7.8.2. Dropping some weaker discriminant variables improved the classification performance over all folds.

Best Parameters	
Parameter	EW Classifier
Name	DNNssWW
Inputs	21
Layers	9
Nodes	117
Activation Function	elu
Output Function	sigmoid
Number of outputs	1
Dropout	0.180000
β_1	0.017908
β_2	0.022979
ϵ	0.0000001
Learning rate	0.001170
Batch Size	278

Table 37: The hyperparameters of the electroweak classifier used in this measurement, referred to as DNNssWW.

Validation AUC					
	Fold 0	Fold 1	Fold 2	Fold 3	Fold 4
	0.8203	0.8210	0.8251	0.8250	0.8239

Table 38: The highest validation AUC of the DNNssWW model. The hyperparameters are given in Table 37. The statistical uncertainty on the AUC is of order 10^{-4}.

The total output of the electroweak classifier for each fold is shown in Fig. 94. The subplots illustrate the normalised predictions for the DNNssWW networks trained in the cross-validation procedure. The predictions are made using the entire training, testing, and validation dataset. Reflective of Table 38, the predictions shown in Fig. 94 show consistent performance across each fold.

Figure 95 illustrate the normalised DNNssWW network predictions for each fold of the cross-validation procedure respectively as functions of the training, testing and validation data for each polarisation sample on the given fold. The network predictions show a high classification probability for the electroweak $W^{\pm}W^{\pm}jj$ signal samples. The figures also show no significant overfitting, indicative of good generalisation. This is to be expected as signal samples account for the largest majority of the training data.

Figure 96 shows the same set of distributions but for the VBS backgrounds. The distributions show the electroweak classifier is very capable of separating the QCD-induced

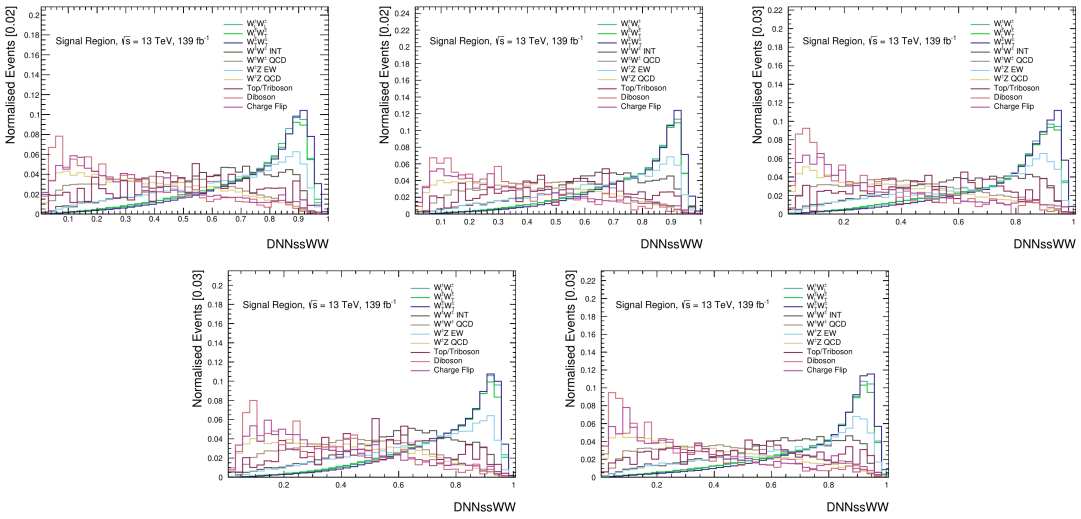


Figure 94: The DNNssWW network predictions for each training fold. The predictions are made with the full training dataset.

$VVjj$ backgrounds and most adapted at classifying the $W^{\pm}Zjj$ QCD background. This is primarily a feature of the event weighting as the $W^{\pm}Zjj$ QCD is the most considerable background of this analysis. In addition, as predicted, the classifier appears to misidentify the $W^{\pm}Zjj$ electroweak production mode. As a similar electroweak process, the kinematic distributions are very similar to the electroweak signal, so the classifier cannot classify this background correctly. The figure also shows some minor overfitting, particularly for the backgrounds. This is expected as the training dataset for the backgrounds is smaller.

Finally, Fig. 97 shows the normalised distribution for the prompt lepton backgrounds and the background associated with the charge misidentification. The classifier performs well at identifying the diboson and charge-flip background. Both of these backgrounds are constrained relatively well by including the lepton type in input features. The background coming from the top and triboson contributions are classified adequately. The available statistics for training these backgrounds are a major limiting factor.

Furthermore, these backgrounds have small event yields and small event weights. For the overall limitations, the separation is reasonably good. The prompt and charge-flip backgrounds also account for only a fraction of the training data. The predicted distributions show more significant overfitting. Once again, this is primarily a feature of the available data. Additional regularisation techniques could improve the observed overfitting but likely at the expense of the overall network performance.

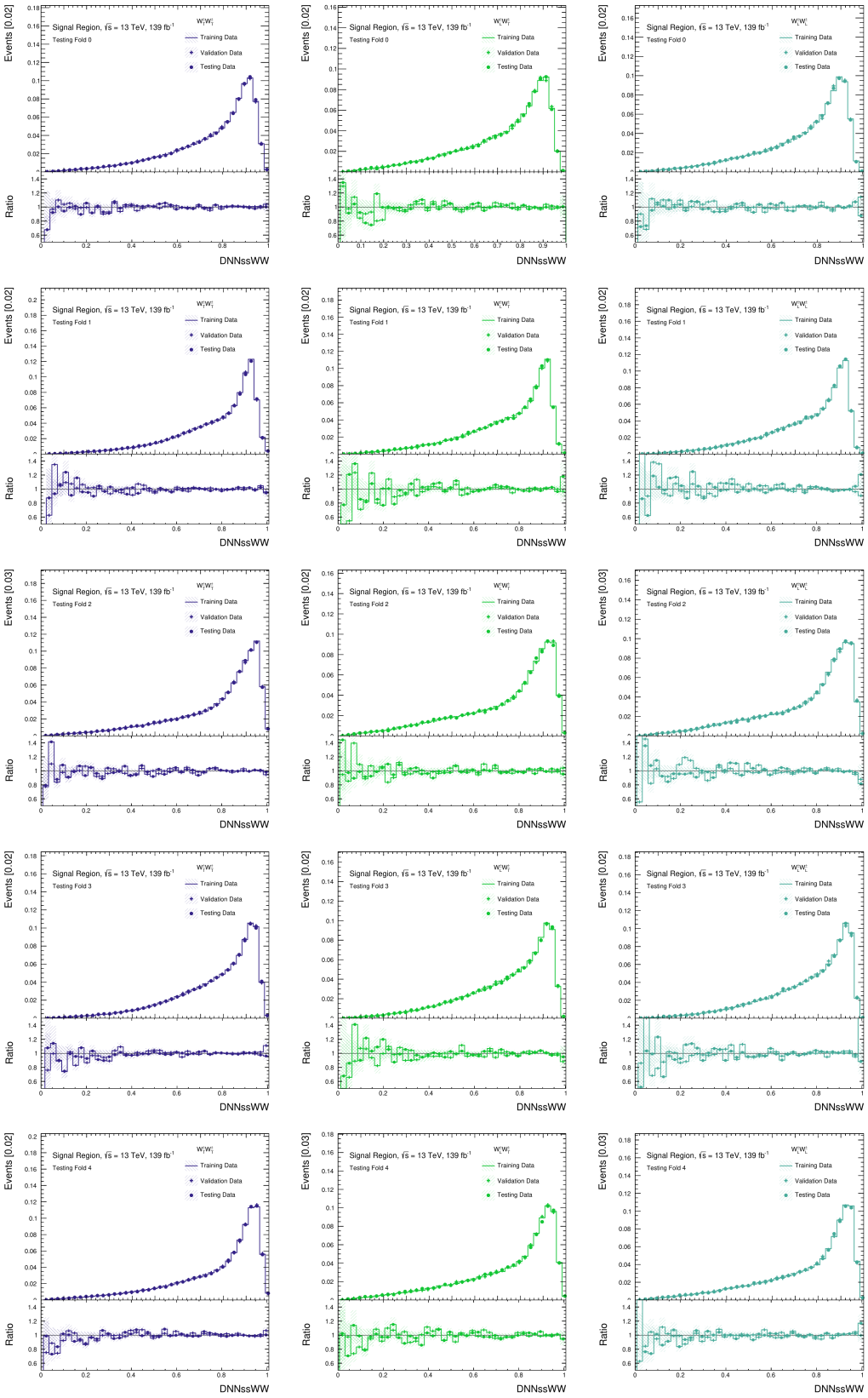


Figure 95: The normalised predictions of the final electroweak classifier after the full cross-validation procedure for the polarised signal samples. The hyperparameters of the networks are given in Table 37, labelled as DNNssWW.

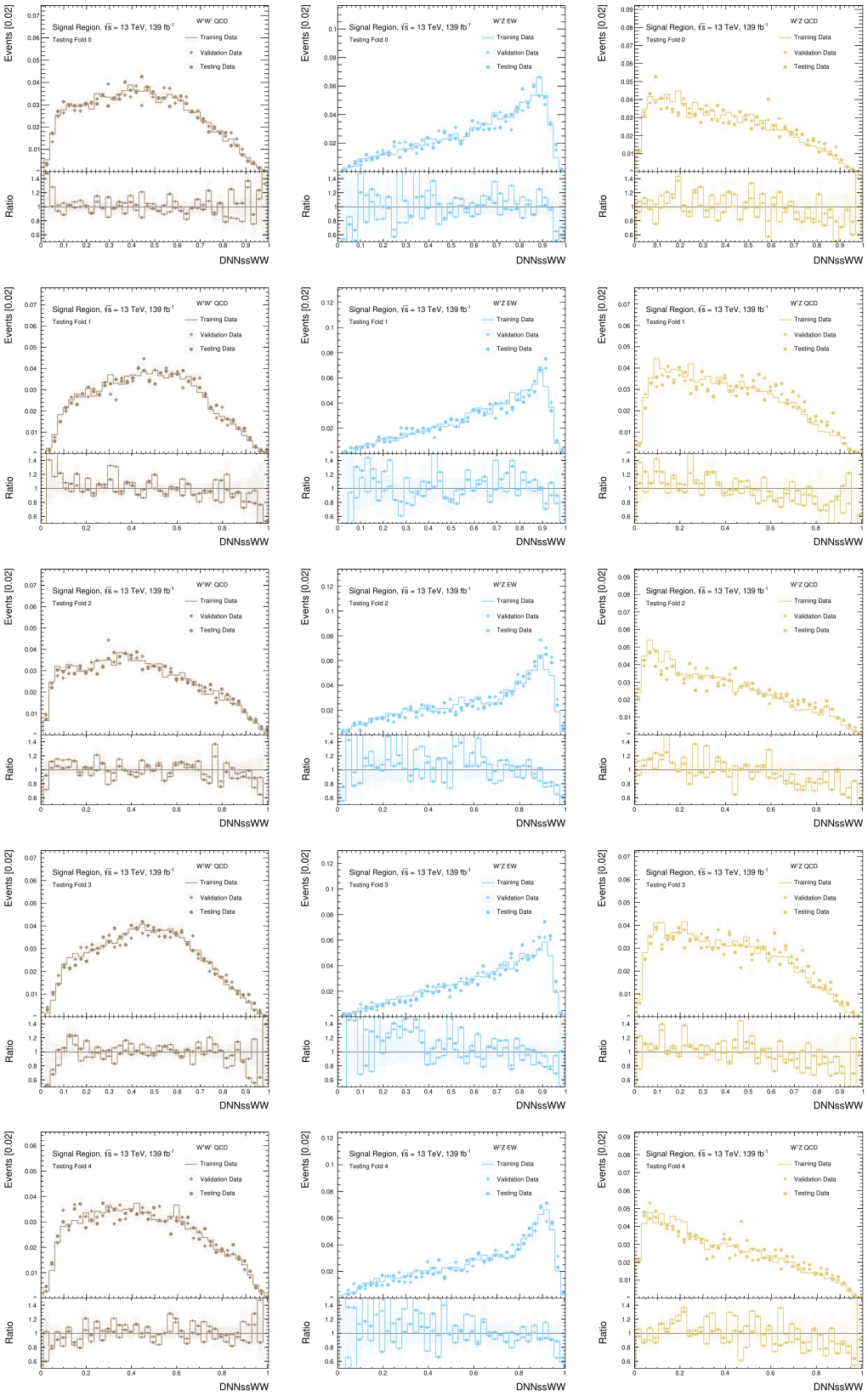


Figure 96: The normalised predictions of the final electroweak classifier after the full cross-validation procedure for the $VVjj$ backgrounds. The hyperparameters of the networks are given in Table 37, labelled as DNNssWW.

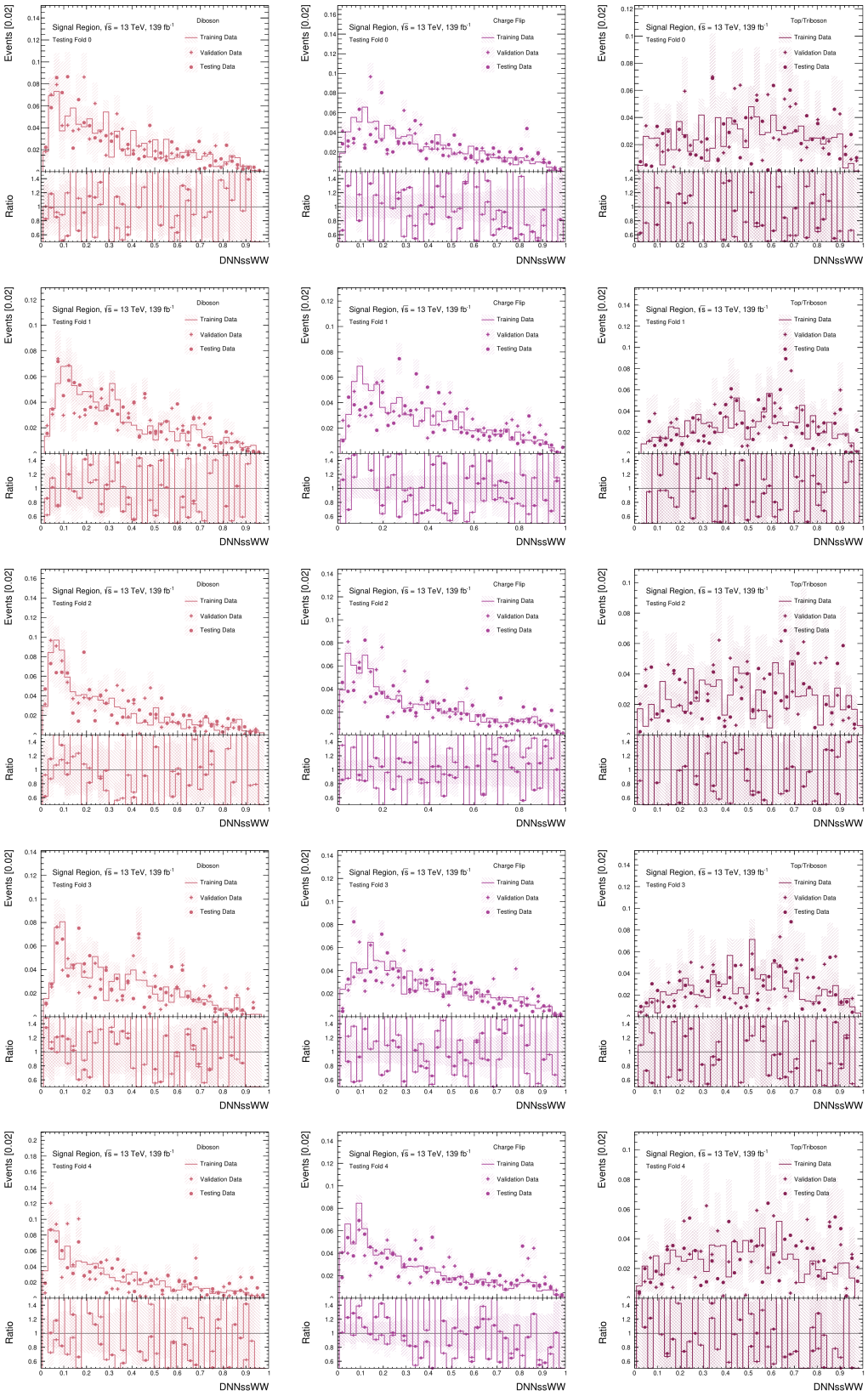


Figure 97: The normalised predictions of the final electroweak classifier after the full cross-validation procedure for the prompt and charge-flip backgrounds. The hyperparameters of the networks are given in Table 37, labelled as DNNssWW.

8.6 Validation

This section will compare the network MC predictions for the inclusive EW $W^{\pm}W^{\pm}$ signal sample to data in the analysis signal region. The section will also test the network predictions in the validation regions of the analysis as a test of network generalisation for unseen data. As in Section 7.9, the uncertainty shown in the following plots is only statistical.

This section will first compare the DNNssWW predictions to experimental data in the analysis validation regions. Figures 98, 99 and 100 show the network predictions corresponding to the classification probability of an electroweak signal event for each of the analysis modelled backgrounds compared to experimental data in the three validation regions: low-Δy_{jj}, top-fakes and WZ. The predictions in the low Δy_{jj} region illustrate consistent modelling compared to data. The consistent modelling is indicative of good model generalisation. The distributions also show consistent modelling between each fold.

Similarly, Fig. 99 shows the modelling in the top-fakes control region. As discussed in Section 7.9, there is a small excess in the number of background events. However, the overall shape predictions of the classifier appear to be consistent with the data. This region is dominated by the non-prompt lepton contribution and the top quark background, backgrounds largely unseen by the classifier. The distributions show good shape agreement with data and do not indicate any significant mismodelling.

Finally, the network predictions in the WZ validation region are shown in Fig. 100. The $W^{\pm}Zjj$ QCD-induced events dominate the region. As discussed in Chapter 6, there is an excess in the MC predictions to the data. However, the predictions in the control region show good agreement in the overall predictions, excluding this rate mismodelling. As with the other validation regions, the models all show no significant mismodelling and relatively consistent predictions between each fold.

The predictions of the electroweak classifier in the analysis signal region are shown in Fig. 101. The figures show a clear separation between the electroweak signal and the analysis backgrounds. The event modelling appears to be in reasonable agreement with the data. However, as in Sec. 7.9, an excess in data is observed. The excess appears to be for data predicted to be largely electroweak signal. This agrees with the observations seen in Sec. 7.9 where the excess appeared for TT events.

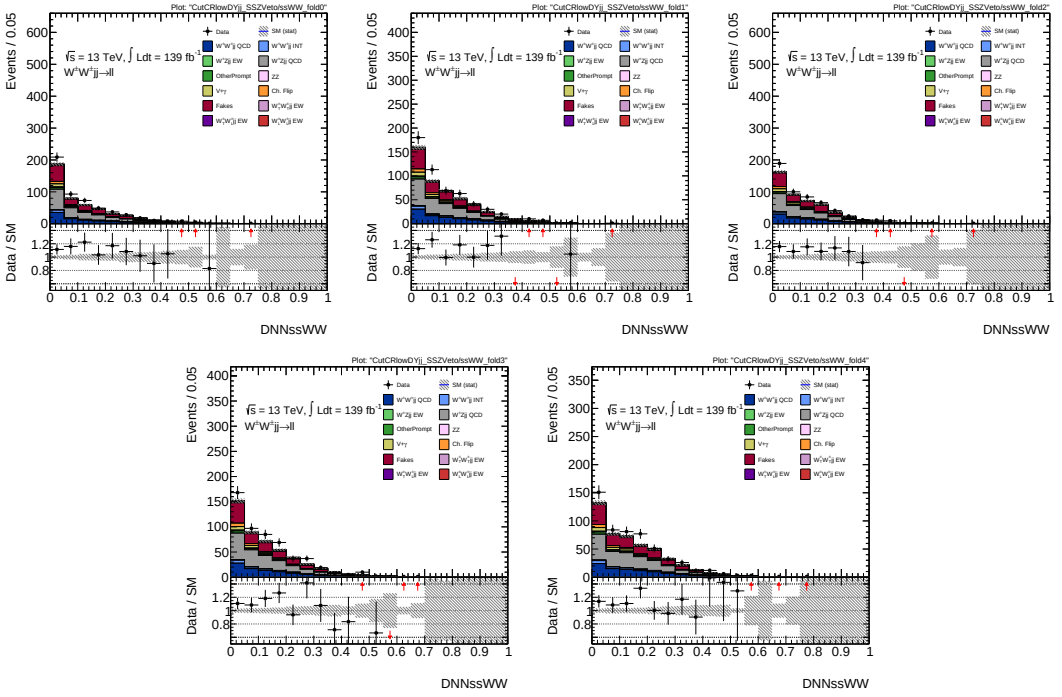


Figure 98: The predicted event yields in the low-ΔY_{jj} validation region compared to data for the electroweak classifier DNNssWW. The uncertainty shown here is statistical. The hyperparameters of the network are given in Table 37.

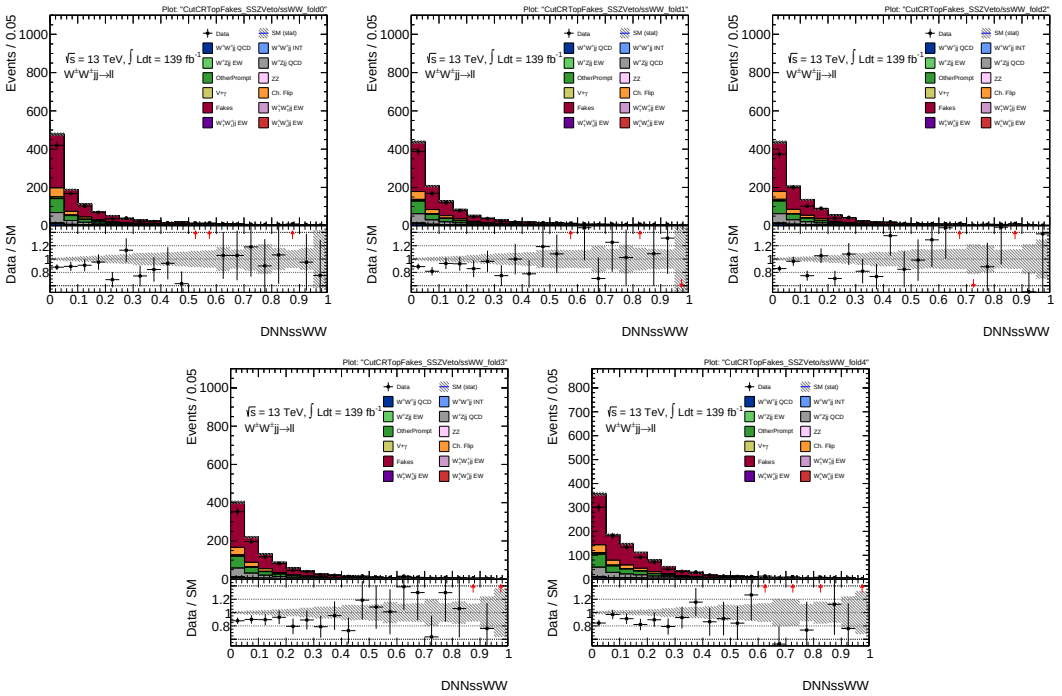


Figure 99: The predicted event yields in the top-fakes validation region compared to data for the electroweak classifier DNNssWW. The uncertainty shown here is statistical. The hyperparameters of the network are given in Table 37.

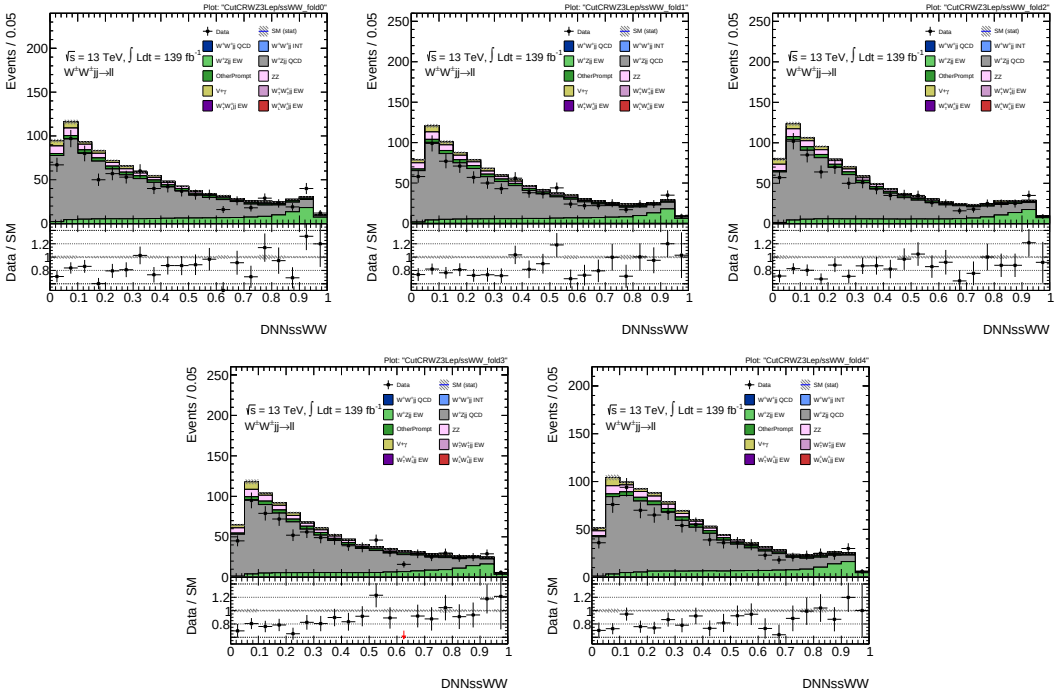


Figure 100: The predicted event yields in the WZ validation region compared to data for the electroweak classifier DNNssWW. The uncertainty shown here is statistical. The hyperparameters of the network are given in Table 37.

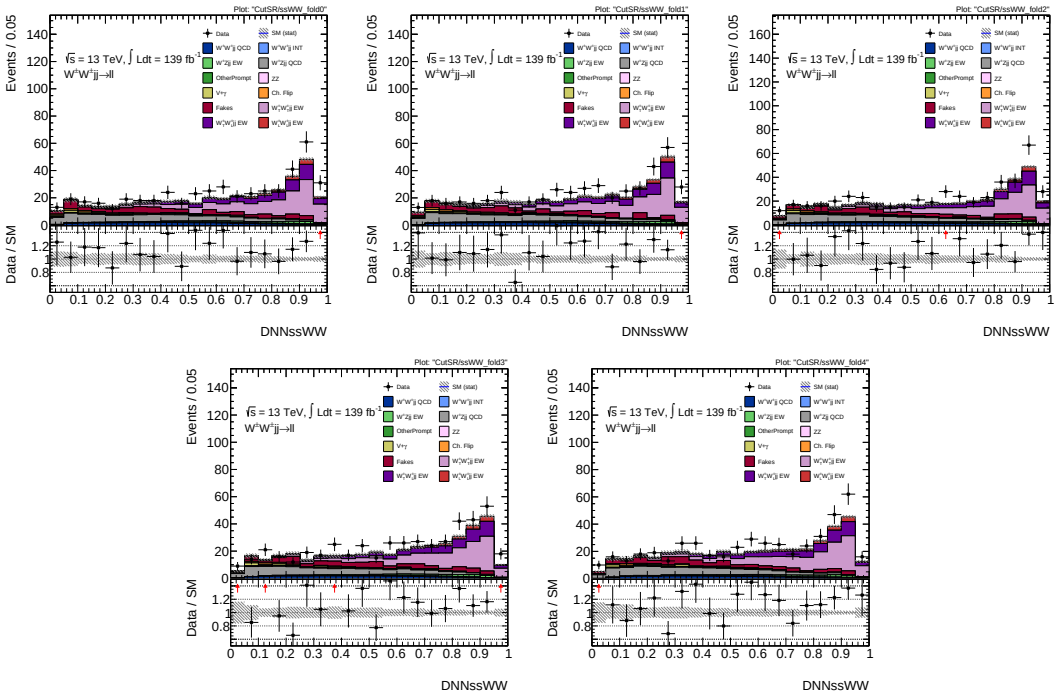


Figure 101: The predicted event yields in the signal region compared to data for the electroweak classifier DNNssWW. The uncertainty shown here is statistical. The hyperparameters of the network are given in Table 37.

9 The $\cos\theta$ Regression

The angular distributions of the leptons associated with EW-induced $W^{\pm}W^{\pm}jj$ production depend on the polarisation of the parent W bosons. These distributions are intrinsically correlated with each lepton's $\cos\theta$ distribution. Experimental measurements of $\cos\theta$ directly probe gauge boson polarisation. To measure $\cos\theta$ from experimental data, it is necessary to reconstruct the momentum of the parent gauge boson, which requires momentum measurements of all of its decay products. For the leptonic decay channel, measurements of this nature are currently experimentally impossible as the neutrino momentum cannot be directly measured. The ATLAS detector reconstructs the missing transverse energy; however, the distribution of the E_T^{miss} between the two W bosons is unknown.

Whilst for this process, the $\cos\theta$ distributions cannot directly be measured for processes event cannot be measured, there is still intrinsic polarisation data which can be collected experimentally: the lepton kinematics and the MET. It may be possible to approximate $\cos\theta$ value from this experimental data. This section illustrates the application of multivariate regression techniques using DNNs to estimate each lepton's $\cos\theta$ value. The regression aims to develop a physics-based model intrinsically connected to polarisation.

The benefit of a regression-based model over more commonly used classifiers is that a regression aims to model a physics-based quantity instead of the discrete separation of signal and background. As discussed in the previous chapters, the use of classifiers is a black-box approach. A classifier has a given set of inputs, which leads to a predicted set of outputs, and the relationships between the two are difficult to understand. Although this approach is effective in enhancing signal, the approach itself gives very little insight into the underlying physics. By contrast, the relationships between the inputs and outputs of a regression are better understood. The regression model can be compared with the true data to provide meaningful insight into the underlying physics.

The application of regression techniques to approximate the $\cos\theta$ value for polarised $W^{\pm}W^{\pm}jj$ was first explored in [144]. This section follows a similar methodology, training a regression to estimate the $\cos\theta$ value of the two leptons in a $W^{\pm}W^{\pm}$ VBS event. The work presented here primarily differs from Ref. [144] by the training datasets. The training data used in Ref. [144] had access to the transverse gauge boson helicity at the truth level. Generating samples with left and right-handed polarisation states is not gauge invariant [87] and is not applicable in this analysis.

The section starts with a discussion on the production of polarised W bosons at LHC and demonstrates that $W^{\pm}$ boson production is predominately left-handed at the LHC. The fractions of the various polarisation states will impact the shape of $\cos\theta$. The section details the modelling of the input variables. The regression is trained using truth data, but the implementation in the analysis is for reconstructed data as an input. The section examines detector effects that may affect the $\cos\theta$ regression. The latter part of the chapter will document the training and modelling of the regressions.

9.1 Polarisation at the LHC

Processes generating gauge bosons at the LHC have varied contributions from each of the different polarisation states. The ratio of the polarisation fractions is highly process-dependent and indicative of the underlying physics. It has been shown that W bosons produced at the LHC with a large transverse momentum, p_T are predominately left-handed [84, 85, 145]. Gauge boson polarisation foremost affects the angular distributions of the leptons and the $\cos\theta$ distribution. It is crucial to understand the underlying physics dictating the $\cos\theta$ distribution as a means to validate the data used in this section.

To illustrate the polarisation effects of individual processes at the LHC, this subsection will analyse polarised $W^{\pm}$ boson production in association with jets for high-p_T. The underlying physics is more constructive to demonstrate in the context of a single gauge boson, but the discussion presented could be extrapolated to VBS production processes, albeit with a much greater difficulty. This subsection illustrates the modelling of polarised VBS processes and shows that the dominant production mode of W bosons is also left-handed.

9.1.1 W + jets

This section will begin by examining the polarisation of high-p_T $W^{\pm}$ boson production in association with jets (W+jets). It has been experimentally determined that at the LHC, W+jets predominately generates left-handed W bosons [145]. A full description of the polarisation of W+jets requires complex helicity state computations beyond the scope of analytical calculation. This section instead demonstrates a heuristic-based physics analysis of polarisation following the arguments presented by [84]. The analysis will illustrate the source of the asymmetry in the polarisation of $W^{\pm}$ boson production, which manifests from the parton distribution functions in proton-proton collisions and the left-handed nature of the electroweak interaction.

Following the arguments of Ref. [84], consider the case of W+jets production at very low-p_T. This the W boson moves in the direction of the beamline with a transverse momentum $p_T \approx 0$. The motion is one-dimensional and simplifies angular momentum conservation. There are two LO order contributions, $u\bar{d} \to W^+$ and $d\bar{u} \to W^-$. The PDFs of pp collisions show the valence quarks typically carry a higher momentum fraction x than the sea quarks. In the case of W production, it is a reasonable approximation to assume that the $W^{\pm}$ boson travels in the direction of the valence quark, as shown in Fig. 102·

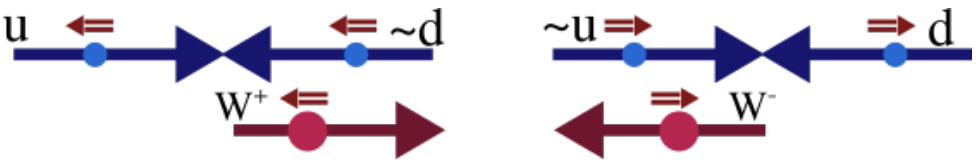


Figure 102: An graphical depiction of W+jets production for $p_T^W \approx 0$. The direction of motion of the particles is illustrated with single-lined arrows, while the spin alignment is portrayed with double-lined arrows.

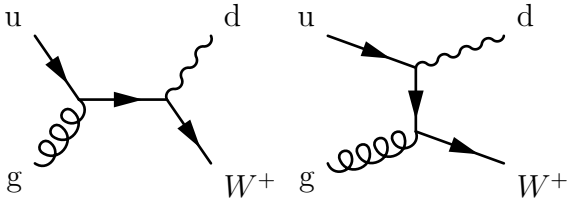


Figure 103: Feynman diagrams showing the leading order s- and t-channel diagrams for $ug \to W^+$

Approximating the quarks to be massless fixes the quarks as left-handed and anti-quarks as right-handed. In this limit, where the $W^{\pm}$ has $p_T \approx 0$, angular momentum conservation requires both W^+ and W^- production to generate left-handed W bosons. If the sea quark carries a larger momentum fraction x, this will result in the production of a right-handed W boson. From the PDFs, this situation is far less likely, so $W^{\pm}$ bosons produced with a low-p_T are expected to be predominately left-handed. The origin of this asymmetry is due to the left-handed nature of the electroweak interaction and the PDFs of pp collision.

Now consider the case of W^+ boson production (the W^- case can be analysed in the same way) with very high-p_T. In this case, there are now two directions of motion, which complicates the angular momentum conservation discussed in the $p_T^W \approx 0$ case. In the ultra high-p_T limit, it can be assumed the W^+ boson moves perpendicularly to the beam line in the pp reference frame. In this situation, the longitudinal scattering mode is negligible. This is due to the equivalence theorem, which equates the amplitude of emission and absorption of a longitudinally polarised gauge boson to the amplitude of emission and absorption of a Goldstone boson [146]. Since a Goldstone boson does not couple to massless fermions, the longitudinal contribution can be assumed to be negligible at high energy. There are 3 LO contributions to W+jets production at the LHC:

$$ug \to W^+ d, \quad u\bar{d} \to W^+ g, \quad g\bar{d} \to W^+ \bar{u}$$

The cross-section for each contribution is highly dependent on the respective PDFs required for each contribution. Since the momentum fraction x carried by a parton scales proportionally to the transverse momentum of the outgoing $W^{\pm}$ boson. For high p_T, it can be assumed that x is large [84]. The hierarchy of the respective PDFs at the LHC for this process is given by,

$$q(x) >> g(x) >> \bar{q}(x) \tag{129}$$

This hierarchy also introduces a hierarchy for the three LO contributions, such that,

$$\frac{d\sigma(ug \to W^+)}{dp_T^W} >> \frac{d\sigma(u\bar{d} \to W^+ g)}{dp_T^W} >> \frac{d\sigma(g\bar{d} \to W^+ \bar{u})}{dp_T^W} \tag{130}$$

Considering the dominant process, with two leading order diagrams, an s- and t-channel contribution, are shown in Fig. 103. After a leptonic decay of the $W^+ \to l^+ \upsilon_l$, the cross-section for each diagram is proportional to the decay products, as given by [84],

$$d\sigma_s \propto (p_d \cdot p_\nu)^2 \quad , \quad d\sigma_t \propto (p_u \cdot p_l)^2 \tag{131}$$

where $d\sigma_s$ is the cross-section of the s-channel diagram in Fig. 103 and $d\sigma_t$ the cross-section of the t-channel diagram. In ultra-high p_T limit, where the scattering of the W^+ boson can be approximated as perpendicular to the beamline. The expressions for the s-channel and t-channel diagrams are given by [84],

$$d\sigma_s \propto (1 - \cos\theta)^2 \tag{132}$$

and for t-channel

$$d\sigma_t \propto \frac{1}{4}(1 + \cos\theta)^2 \tag{133}$$

The expressions can be identified as the angular distributions for left-handed W^+ bosons for the s-channel and the right-handed distribution for the t-channel. However, the t-channel expression carries an additional factor of $\frac{1}{4}$, which introduces a polarisation asymmetry in the ultra-high p_T^W limit. The case of the s-channel decay can be understood in terms of angular momentum conservation. Here, the off-shell u-quark decays into a W^+d pair. The W^+ cannot be right-handed as such a decay would violate angular momentum conservation along the Wd axis. The remaining processes generate polarisation states equally and net right-handed gauge boson. However, the dominant process is predominately left-handed.

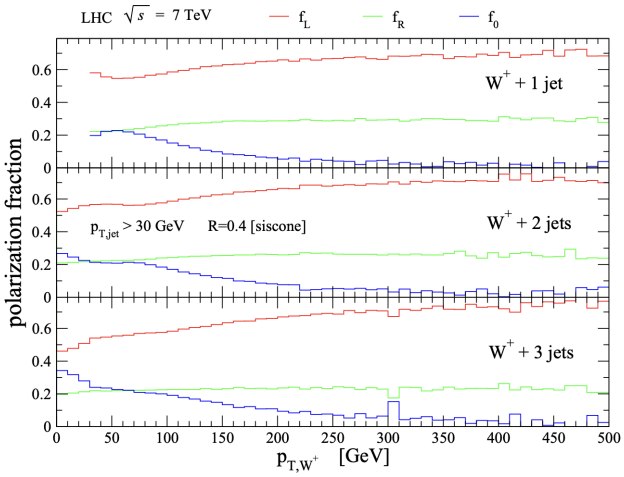


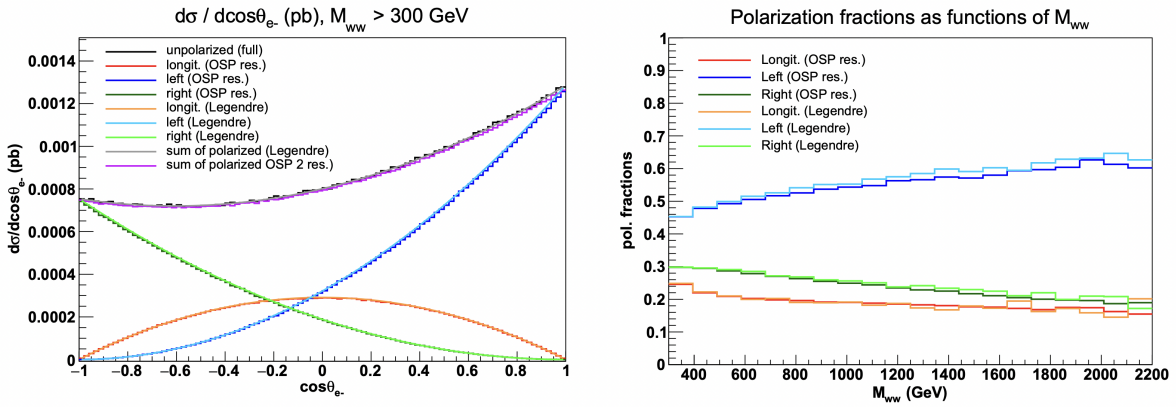
Figure 104: The ratio of polarisation fractions, f_L, f_R and f_0, corresponding to left- and right-handed, and longitudinally polarised W^+ gauge bosons, for W^++ jets production with 1, 2, 3 jets as a function of the W boson p_T. Figure taken from [84]

Figure 104 demonstrates MC simulation of W+jets and the ratio of the polarisation fractions ($f_l + f_R + f_0 = 0$) produced for p-p collisions as a function of the W boson p_T. As the p_T increases, the left-handed fraction f_l increases.

9.1.2 VBS

The production mechanisms for $W^\pm$ bosons in electroweak VBS are more complicated. A similar heuristic examination of helicity and polarisation is far more complex due to the vector nature of the two gauge bosons. This section illustrates instead numerical

simulations of polarised VBS. Polarised $W^{\pm}$ production for VBS processes at the LHC has been studied in Refs. [83, 147].



(a) $\cos\theta_e$ (b) Polarization fractions

Figure 105: The differential cross-section distribution with respect to the $\cos\theta$ distribution of the e^- decay for $W^{\pm}W^{\mp}jj$ production and the diboson invariant mass M_{WW}. The distributions are shown in terms of the polarisation states of the W^- boson, extracted using Legendre polynomials and the OSP method [83]. Figure taken from [83].

Figure 105 shows the differential cross-section as a function of $\cos\theta$ and the polarisation fractions associated with polarised $W^{\pm}W^{\mp}jj$ production. Figure 105(b) shows that W^- gauge bosons are produced as predominately left-handed in VBS process at the LHC. Figure 105(a) shows the impact of the asymmetry on the $\cos\theta$ distribution. The differential cross-section is predicted to be larger for left-handed W^- production. The total transverse $\cos\theta$ distribution is therefore asymmetric. Figure 106 shows the distributions for the same process but for the W^+ gauge boson. The transverse gauge boson $\cos\theta$ distribution shows the same asymmetry. W^+ gauge bosons are also predominately produced as left-handed at the LHC in VBS processes.

This chapter models the $\cos\theta$ distribution for polarised $W^{\pm}W^{\pm}jj$ production. The distributions shown in this section are over a much larger phase space. Moreover, for $W^{\pm}W^{\mp}jj$ production, the leptons are distinguishable by charge. This chapter demonstrates the impact of selection criteria on the $\cos\theta$ distribution and the impact of p_T order required for a process with indistinguishable leptons.

9.2 Input Variables

A total of 16 variables are used for the regression. The chosen variables are the kinematic variables of the leptons, jets, and the met, summarised in Tab. 39. No further composite higher-level variables are included. The lepton four-momenta are included as these are used in the $\cos\theta$ calculation. Likewise, the E_T^{miss} and the leading and sub-leading jet p_T are also included to give a complete picture of the event kinematics. The regression is trained on truth data to maximise the available statistics for training. The caveat is that the input data to the regression in the experiment is detector-simulated and experimental data. Therefore, it is essential to examine the difference between the input variables at

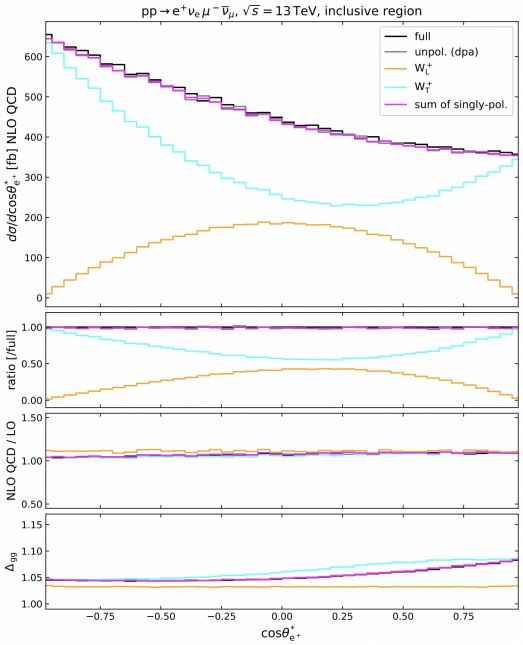


Figure 106: The differential cross-section distribution with respect to the $\cos\theta$ distribution for $W^{\pm}W^{\mp}jj$ production. The distributions are shown in terms of the polarisation states of the W^{+} boson. Figure taken from [147].

the truth level and after detector simulation to understand the effect on the regression. Training a regression using ATLAS simulated data is possible as the truth record is maintained during simulation.

A comparison of the truth and detector simulated input variables for each polarisation mode is shown for the jets in Figs. 108, 110, 112, for the leptons in Figs. 109, 111, 113 and for the E_T^{miss} in Fig. 107. Each distribution is normalised to illustrate the input shape differences and shown in the signal region. From the figures, it can be seen there are no major differences between the truth and reconstructed data. The biggest differences can be seen for the lepton η measurement. The region between $1.37 < |\eta| < 1.52$ is the transition region between the EM calorimeter barrel and end-cap. Particles passing through this region interact with both components of the calorimeter, which often leads to inconsistencies in reconstructed measurements. In the signal selection, signal electrons are excluded in this region, which leads to the shape differences seen in Figs. 109, 111, 113.

The smaller differences come in the form of p_T resolution effect impacting the jet, leptons and E_T^{miss} kinematic distributions. The shape differences exhibited for each of the three polarisation configurations are limited. There will be some differences between the regression predictions with the reconstructed samples and the truth sample. However, the difference should be uniform across the three polarisation fractions.

Variable	Definition
lep0 p_T	Transverse momentum of leading lepton in p_T
lep0 η	η coordinate of leading lepton in p_T
lep0 ϕ	ϕ coordinate of leading lepton in p_T
lep0 e	Energy of leading lepton in p_T
lep1 p_T	Transverse momentum of sub-leading lepton in p_T
lep1 η	η coordinate of sub-leading lepton in p_T
lep1 ϕ	ϕ coordinate of sub-leading lepton in p_T
lep1 e	Energy of leading sub-lepton in p_T
E_T^{miss}	Missing transverse momentum
E_T^{miss}-ϕ	ϕ coordinate of the missing energy
jet0 p_T	Transverse momentum of leading jet in p_T
jet0 η	η coordinate of leading jet in p_T
jet0 ϕ	ϕ coordinate of leading jet in p_T
jet1 p_T	Transverse momentum of sub-leading jet in p_T
jet1 η	η coordinate of sub-leading jet in p_T
jet1 ϕ	ϕ coordinate of sub-leading jet in p_T

Table 39: The input variables used to train the $\cos\theta$ regressions.

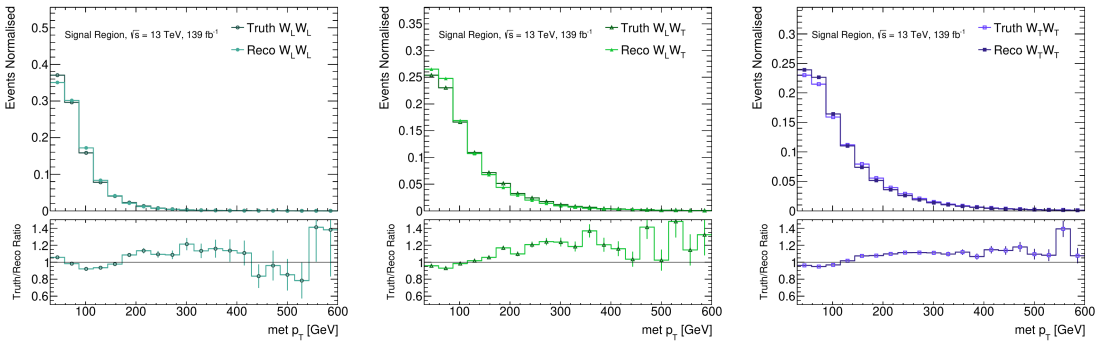


Figure 107: The E_T^{miss} distribution used as inputs to the regression, simulated and at truth level, for each polarised $W^{\pm}W^{\pm}jj$ sample in the signal.

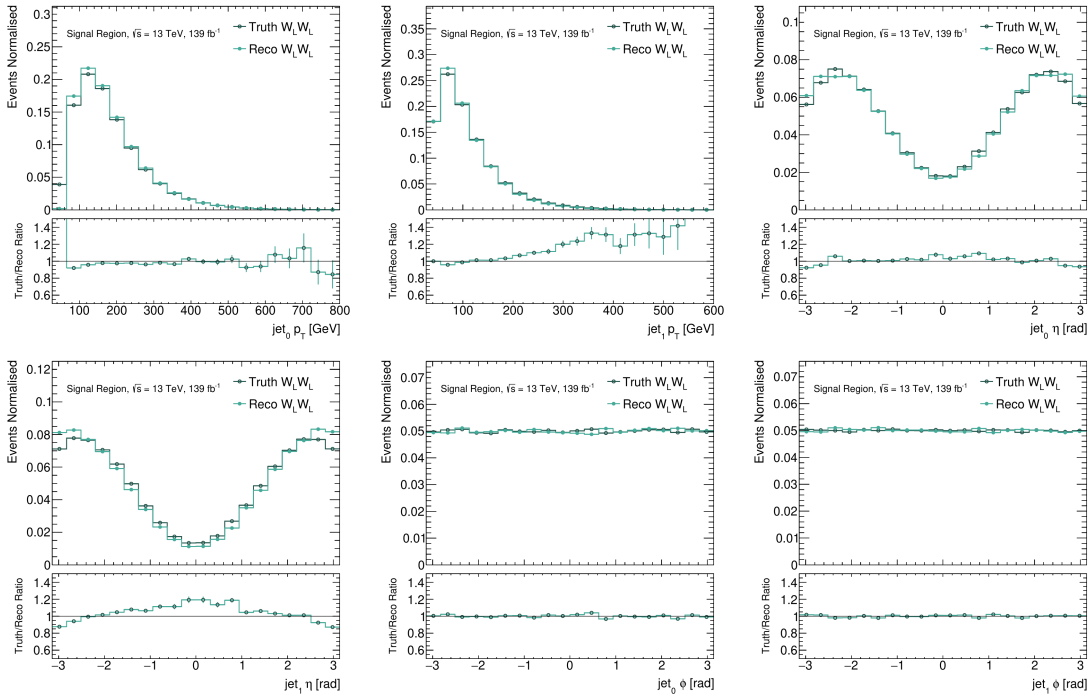


Figure 108: The jet variables, used as inputs to the regression, simulated and truth level, for the polarised $W_L^{\pm}W_L^{\pm}jj$ sample in the signal region.

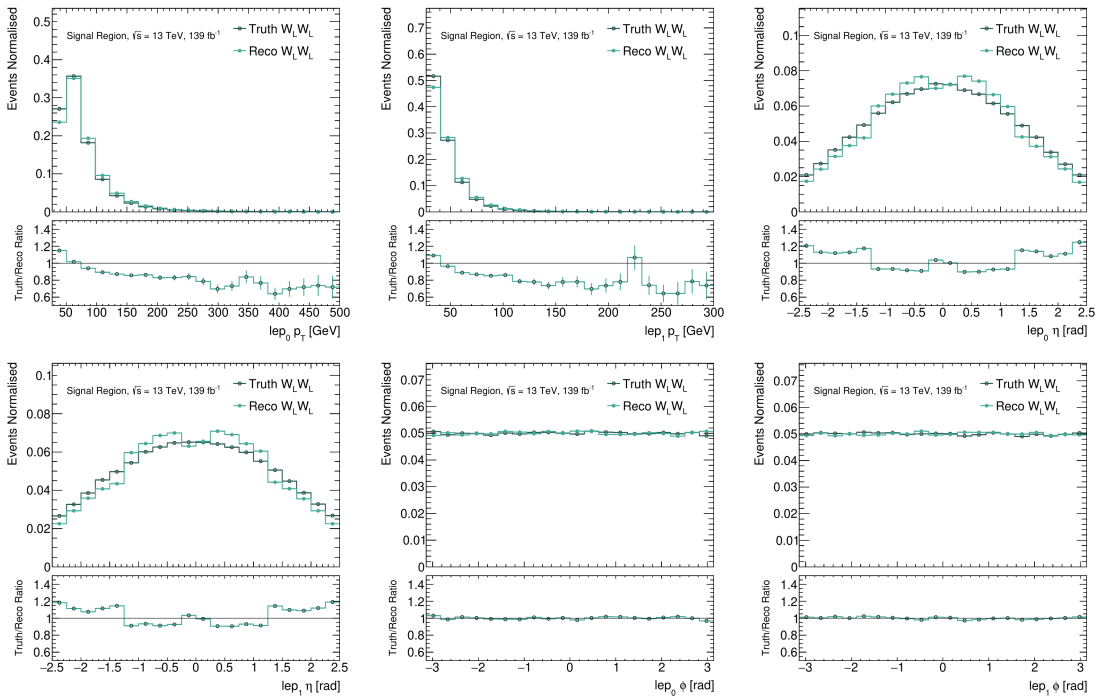


Figure 109: The lepton variables, used as inputs to the regression, simulated and at truth level, for the polarised $W_L^{\pm}W_L^{\pm}jj$ sample in the signal region.

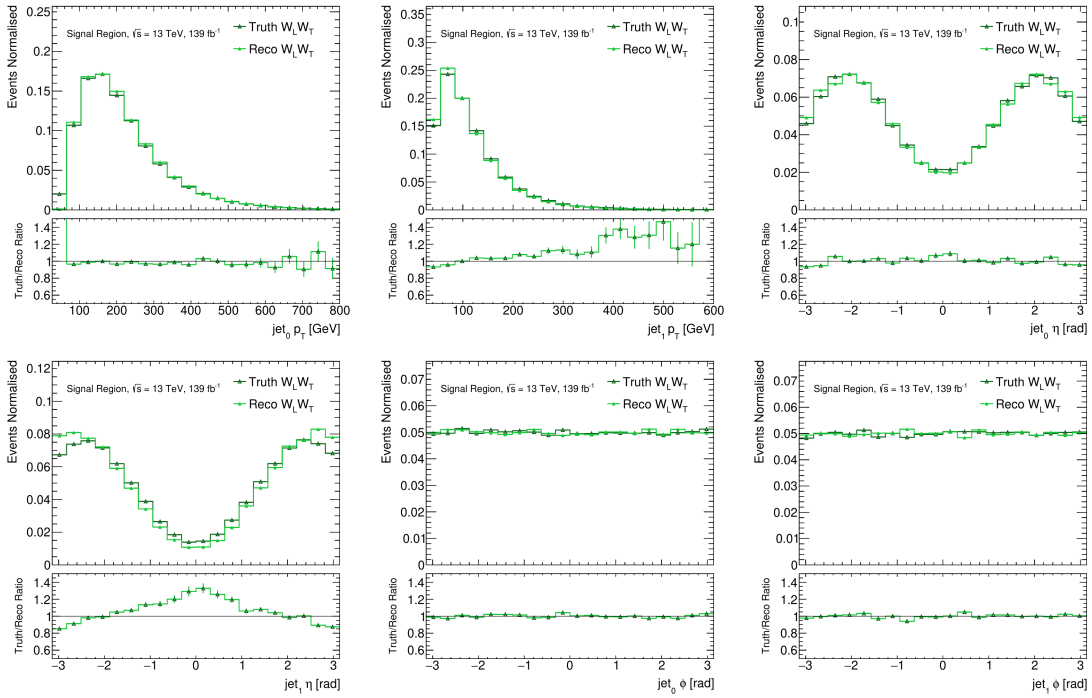


Figure 110: The jet variables, used as inputs to the regression, simulated and at truth level, for the polarised $W_L^{\pm}W_T^{\pm}jj$ sample in the signal region.

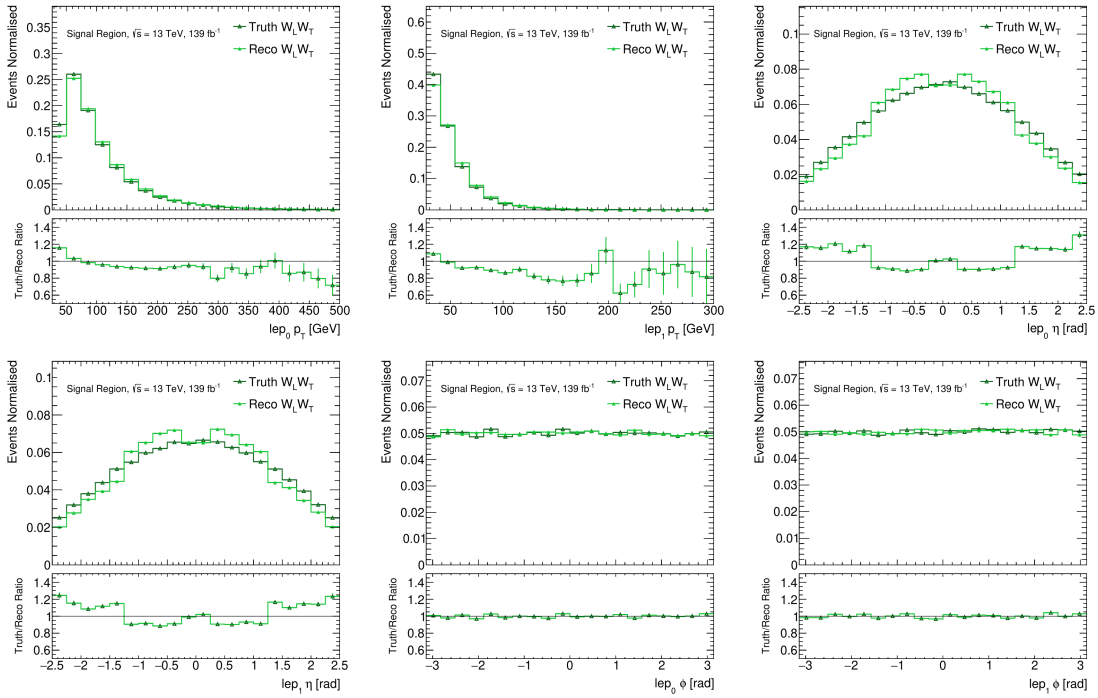


Figure 111: The lepton variables, used as inputs to the regression, simulated and at truth level, for the polarised $W_L^{\pm}W_L^{\pm}jj$ sample in the signal region.

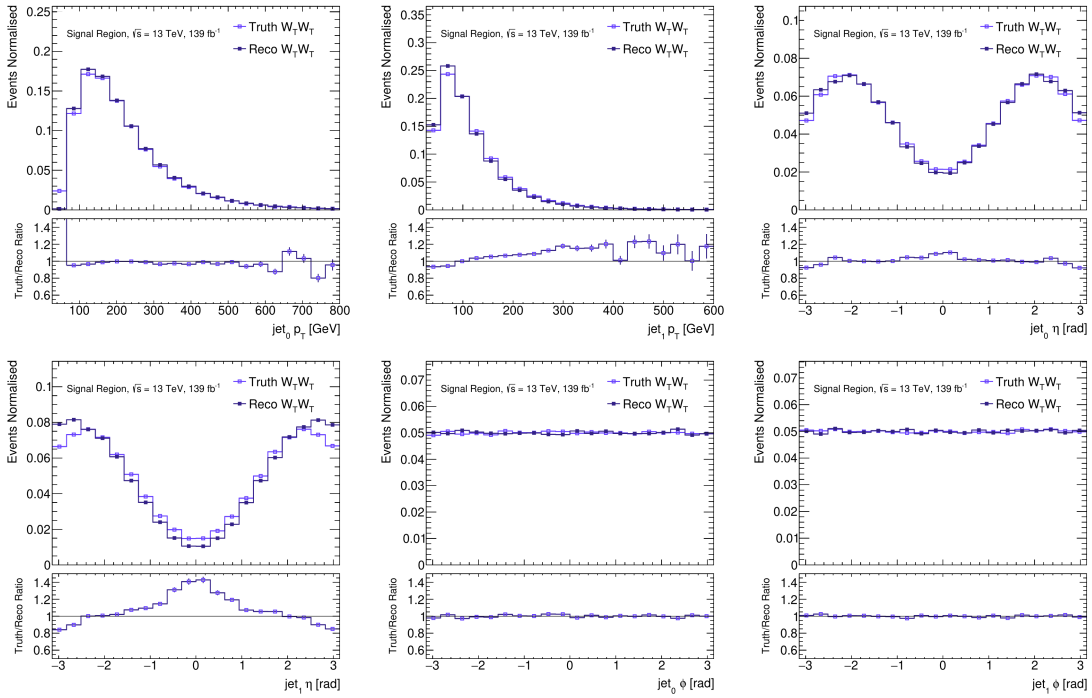


Figure 112: The jet variables, used as inputs to the regression, simulated and at truth level, for the polarised $W_T^{\pm}W_T^{\pm}jj$ sample in the signal region.

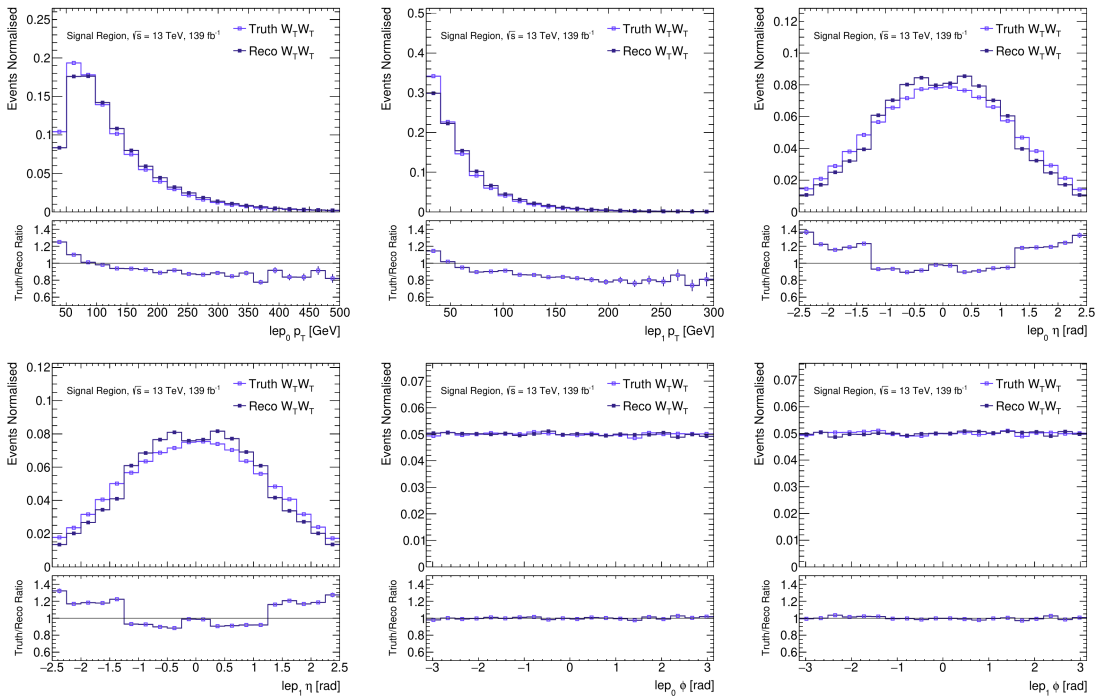


Figure 113: The lepton variables, used as inputs to the regression, simulated and at truth level, for the polarised $W_L^{\pm}W_T^{\pm}jj$ sample in the signal region.

9.3 Output Variables: Modelling $\cos\theta$

A regression is trained to predict the $\cos\theta$ value of the leading and sub-leading lepton in p_T. In this chapter, the term $\cos\theta$ distribution refers to the joint distributions of $\cos\theta$ for each lepton. Intrinsically, the $\cos\theta$ value is highly correlated to W boson polarisation. Each of the three polarisation configurations gives a different unique $\cos\theta$ distribution. This section will discuss the modelling of the distributions for each polarisation.

9.3.1 Definition and computation

For a lepton l originating from the decay of its parent W, the $\cos\theta$ value is defined as the cosine of the angle between the direction of motion of parent W, p_T^W, and the direction of motion of the lepton in the W boson rest frame, p_T^l, defined by,

$$\cos\theta = \frac{p_T^W \cdot p_T^l}{|p_T^l||p_T^W|}. \tag{134}$$

The variable depends on p_T^W and is frame-dependent. The $\cos\theta$ distributions shown in this section are computed from the truth record of the MC samples. The reference frame in which the distribution is calculated is in the CoM frame of the W bosons. A reference frame is chosen to ensure the two $\cos\theta$ values associated with each lepton are calculated on the same set of axes. Furthermore, the leptons are distinguished by ordering in terms of p_T. A Rivet [148] routine is implemented to compute the $\cos\theta$ value using the following procedure.

1. Select the two neutrinos from the event truth record.
2. From the neutrinos, select the parent W bosons and order by the bosons by lepton p_T.
3. Boost into the WW centre of mass frame. For each W, to calculate the unit direction $p_T^W/|p_T^W|$.
4. For each W, boost into the rest frame and calculate the $\cos\theta$ as per equation 134.

9.3.2 $\cos\theta$ distributions

Two factors can affect the shape of the $\cos\theta$ distribution: reference frame and event cuts. The former is evident from the definition in Eqn. 134 and knowing that polarisation is not Lorentz invariant. Similarly, as discussed in Chapter 4, the polarisation fractions and the $\cos\theta$ distribution are dependent on event cuts. The angular dependence of the cross-section in ϕ is only removed when integrating over the full-phase space.

The same-electric charge requirement indirectly introduces an additional dependence on the $\cos\theta$ distribution associated with this experiment. In the measurement, leptons are selected to be either muons or electrons and, due to the same-electric charge requirement, are indistinguishable. To provide meaningful labelling of the input variables to the regression, the leptons are ordered by p_T prior to calculating the $\cos\theta$ value. The ordering introduces an additional p_T dependence on the shape.

To illustrate the effect of the event cuts and p_T ordering, this section will first compare the $\cos\theta$ distributions before and after signal region cuts. The distributions shown will be for $e\mu$ events where the leptons can be distinguished, and the p_T ordering is not required. This will independently show the effect of the signal region cuts. Subsequently, the section will then show the $\cos\theta$ distribution but ordered by p_T. The distributions will again be shown before and after signal region cuts.

9.3.3 $e\mu$ $\cos\theta$ distributions

The $\cos\theta$ distributions for $e\mu$ events pre- and post-signal region cuts are shown in Figs. 114 and 115. In these plots, no cut refers to the distributions before the experiment signal region cuts. Furthermore, $\cos_e\theta$ refers to the $\cos\theta$ distribution associated with electrons and $\cos_\mu\theta$, the distribution associated with muons. The effect of the cuts exponentially scales with $\cos\theta$ approaching -1, shown by the ratio plots in Fig. 116. The figures show that the event cuts predominately affect the polarisation configurations with transversely polarised W bosons. For longitudinal scattering, backwards scattering with $\cos\theta \approx -1$ is highly suppressed.

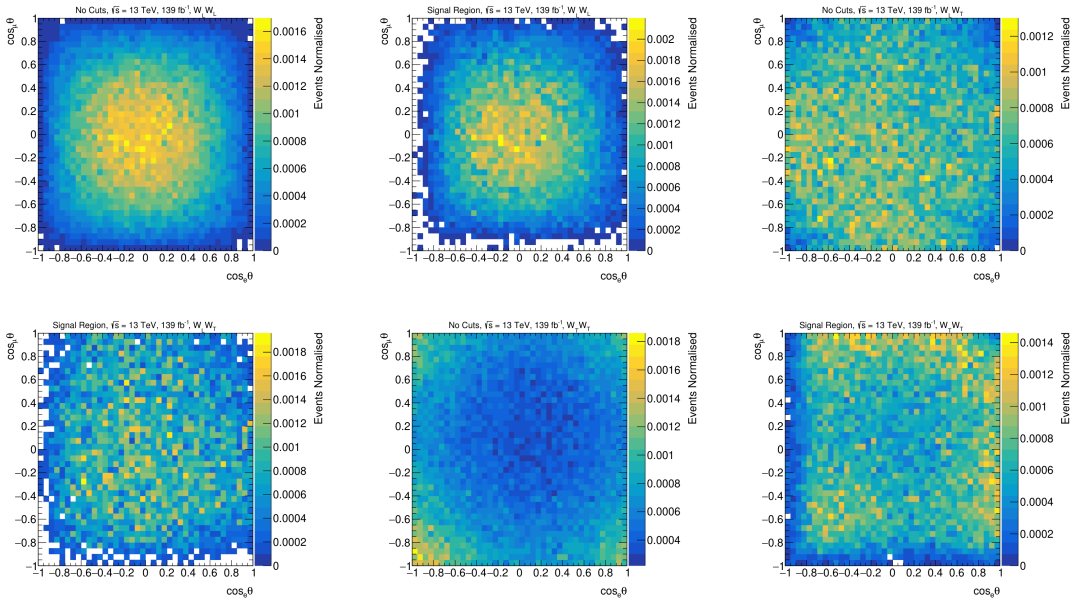


Figure 114: A comparison of the $\cos\theta$ distributions for each of the three polarisation fractions, prior to signal region cuts, referred to in the plots as no cuts, and post signal region cuts.

The primary cut affecting the shape of the $\cos\theta$ distribution is the Lepton p_T cut. The leptons selected in the signal region are associated with high p_T W bosons. The leptons which decay with a $\cos\theta < 0$, in the rest frame of the W, decay partially backwards, against the direction of motion to the W. This means that in the reference frame of the detector, the leptons have a lower p_T and are less likely to meet the minimum lepton p_T cut. As the $\cos\theta$ value approaches $\cos\theta \approx -1$, the leptons are projected almost entirely

against the direction of motion of the W, and this is the region in which the lepton p_T cut most significantly changes the $\cos\theta$ distribution.

To see the polarisation effects, consider partitioning the $e\mu$ events by charge: $e^-\mu^-$ and $e^+\mu^+$. Figure 115 shows an asymmetry in the transverse distribution, which is due to the asymmetry of W^+ and W^- production at the LHC. The individual $\cos\theta$ distributions for $e^-\mu^-$ and $e^+\mu^+$ are shown in Figs. 117 and 118. In both figures, it can be observed that the LHC is preferential to generating left-handed W bosons. The p_T cuts events with $\cos\theta \approx -1$, for $e^-\mu^-$ this predominately affects right-handed W^- boson production. By contrast, for $e^+\mu^+$, the p_T cuts will predominately affect the production of left-handed W^+ bosons.

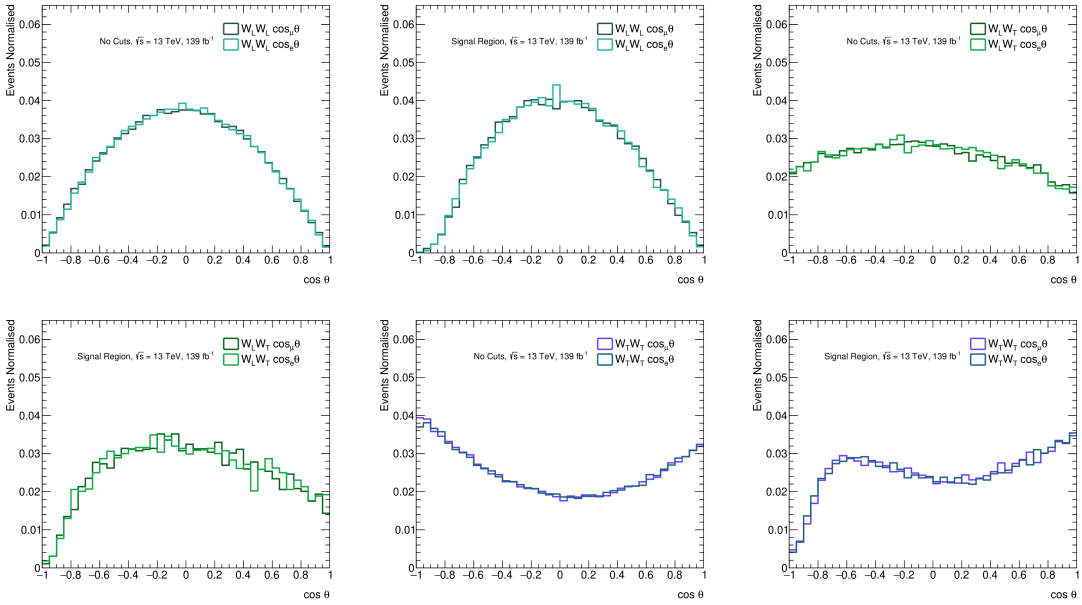


Figure 115: The individual $e\mu$ $\cos\theta$ distributions for each of the three polarisation fractions, prior to signal region cuts, referred to in the plots as no cuts, and post signal region cuts.

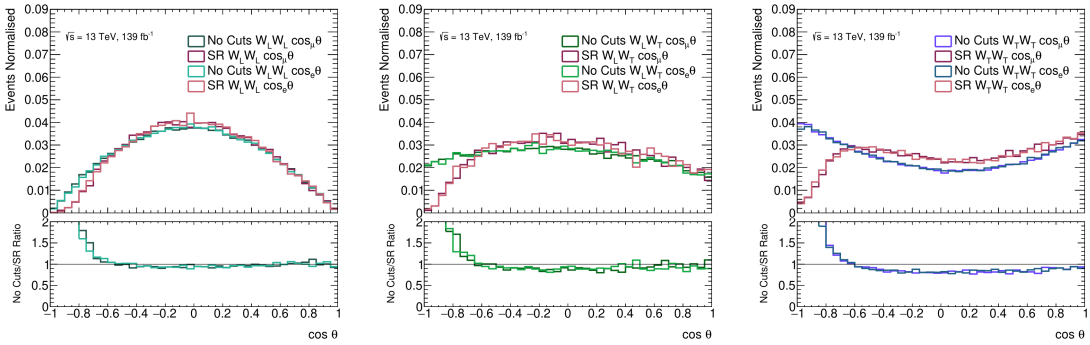


Figure 116: The p_T ordered $\cos\theta$ distributions for each of the three polarisation fractions, prior to signal region cuts, referred to in the plots as no cuts, and post signal region cuts.

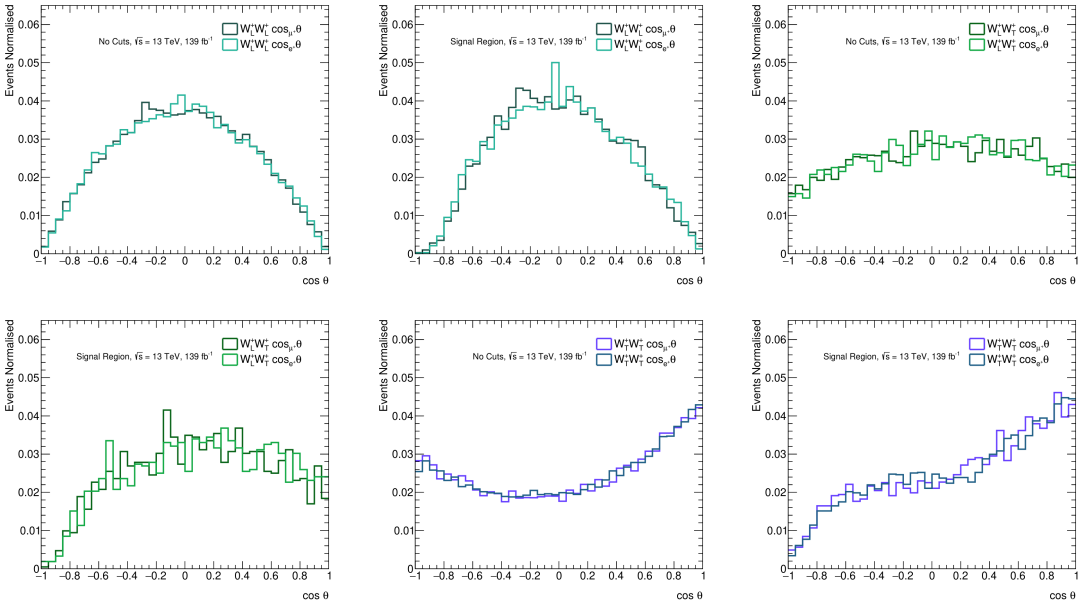


Figure 117: A comparison of the individual $e^+\mu^+$ $\cos\theta$ distributions for each of the three polarisation fractions, prior to signal region cuts, referred to in the plots as no cuts, and post signal region cuts.

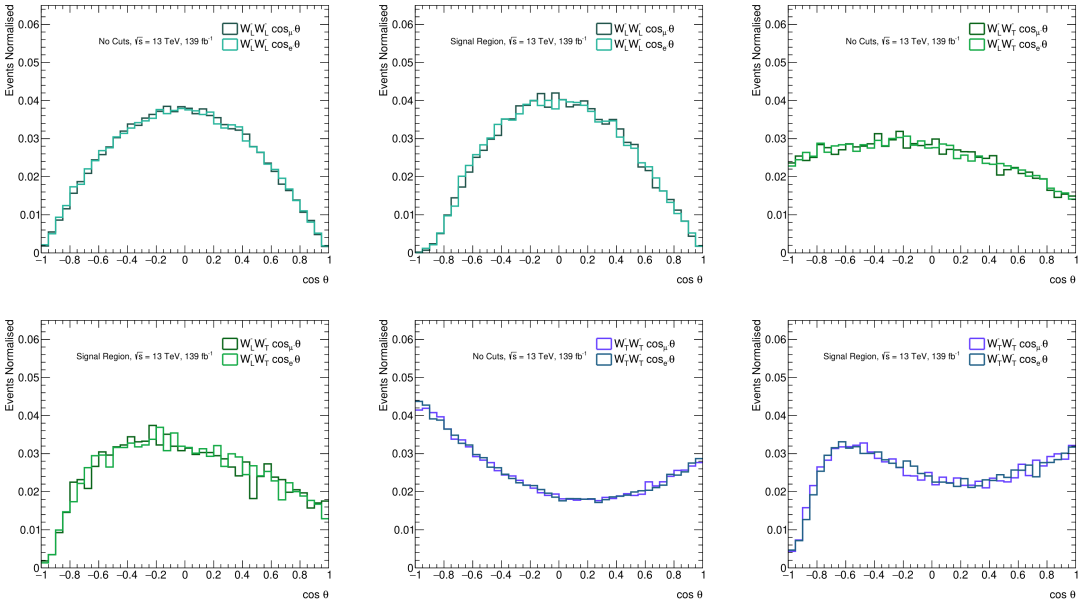


Figure 118: A comparison of the individual $e^-\mu^-$ $\cos\theta$ distributions for each of the three polarisation fractions, prior to signal region cuts, referred to in the plots as no cuts, and post signal region cuts.

9.3.4 p_T ordered $\cos\theta$ distributions

The leptons in the remainder of this chapter are ordered by p_T where the leading lepton $\cos\theta$ distribution is denoted by $\cos_0\theta$ and the sub-leading by $\cos_1\theta$. A comparison of the $\cos\theta$ distributions with p_T ordering are shown in Figs. 119 and 120 before and after signal region cuts. Comparing with Figs. 115 and 120, it is apparent that p_T ordering introduces an asymmetry between the two $\cos\theta$ distributions for each lepton.

Fig. 121 shows that the effect of the cuts is similar as discussed previously, predominately reducing the number of events with $\cos\theta \approx -1$. However, the cuts are not uniform for the two lepton $\cos\theta$ values. In all three distributions, it can be observed that the cuts, in particular the lepton p_T cut, affect the sub-leading $\cos_1\theta$ value more than the leading, demonstrated by Fig. 121. The effect of the cuts predominately changes the shape of the sub-leading lepton $\cos_1\theta$ as this lepton is the minimum requirement to pass the signal region cuts. The leading will pass the selection criteria by definition if the sub-leading lepton passes the minimum p_T cut.

The data used to train the output of the regression is shown in Fig. 119 and 120. Each of the distributions is unique and reflective of the W boson polarisation. The most significant differences in the distributions are between the purely transverse and purely longitudinal polarisation. The aim of the regression is to try to approximate these distributions.

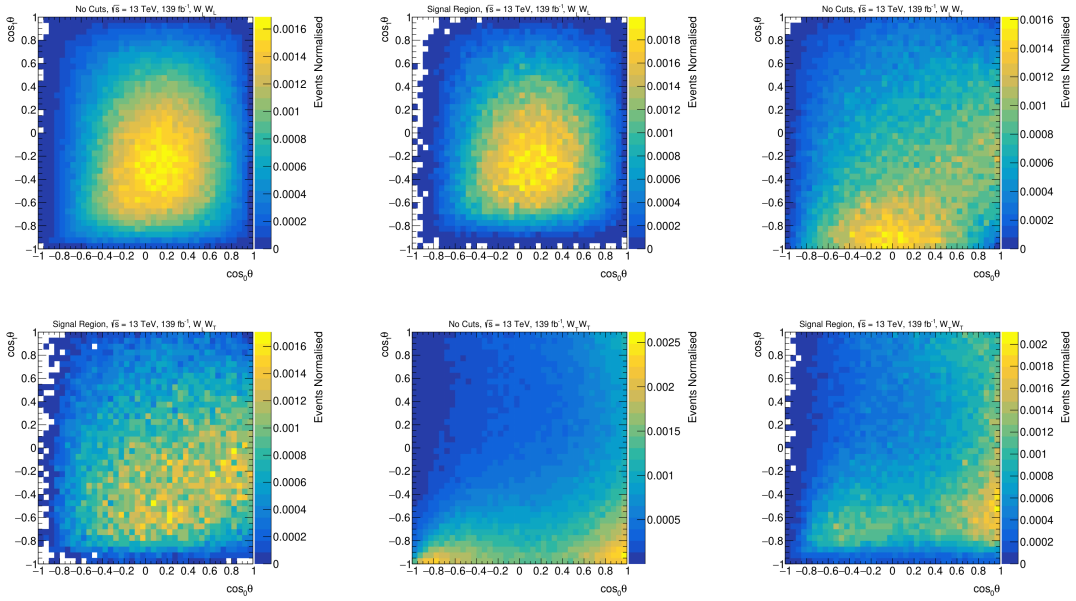


Figure 119: A comparison of the p_T ordered $\cos\theta$ distributions for each of the three polarisation fractions, prior to signal region cuts, referred to in the plots as no cuts, and post signal region cuts.

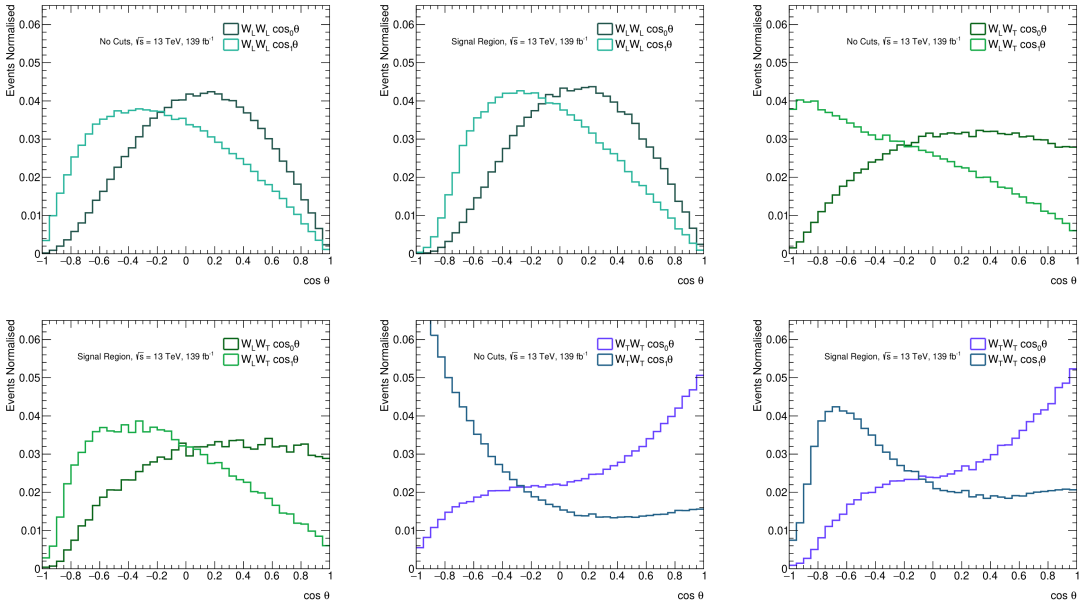


Figure 120: A comparison of the p_T ordered $\cos\theta$ distributions for each of the three polarisation fractions, prior to signal region cuts, referred to in the plots as no cuts, and post signal region cuts.

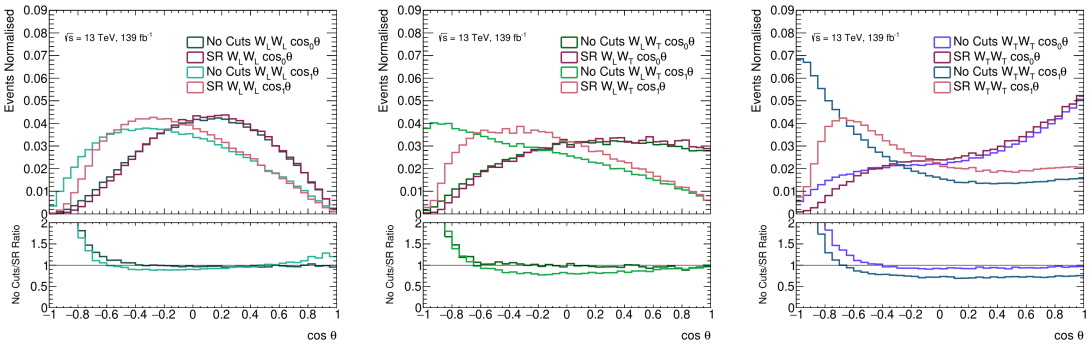


Figure 121: A comparison of the p_T ordered $\cos\theta$ distributions for each of the three polarisation fractions, prior to signal region cuts, referred to in the plots as no cuts, and post signal region cuts.

9.4 Regression Training

In this work, a regression is trained for each of the different polarisation configurations: LL, LT and TT. Each regression predicts the $\cos\theta$ value of the leading and sub-leading lepton. At the analysis stage, the polarisation of a sample is unknown. Within the analysis, the regressions are treated as kinematic distributions. Each sample will have a unique prediction from the regression. The kinematic differences in the samples will manifest as differences in the prediction, which can be used to identify polarisation or as additional input variables to a classifier. The regression in this section is trained using the Python package Keras [140]. The remainder of this section documents the regression

training procedure.

9.4.1 Training Samples

A separate set of polarised samples is generated to train the regression. The samples are generated using the same routine as the simulated samples described in Sec. 5.3.2, but only at the truth level. The total number of truth events and the data used for training, testing and validation is summarised in Tab. 40. The regression is not trained with the cross-validation method discussed in Sec. 7.4.1. The training dataset is significantly larger for the regression. The simulated data is also not used for training and is unseen for the regression, so there is no motivation for cross-validation.

Truth Dataset				
Sample	Total Events	Signal Region	Training	Testing & Validation
$W_L^{\pm}W_L^{\pm}jj$	2073929	620570	310285	155142
$W_L^{\pm}W_T^{\pm}jj$	1069630	298698	149349	74674
$W_T^{\pm}W_T^{\pm}jj$	2973336	800039	400020	200010

Table 40: A breakdown of the number of events available in truth-level samples used to train the regression.

9.4.2 Loss Functions

During the training process, multiple loss functions were also tested. Each loss function has unique properties and responds differently to the training data. Three primary loss functions were tested: the Mean Square Error (MSE), Log Cosh (LC) and Cosine Similarity (CS) loss functions. The MSE loss is defined as,

$$L(y,\hat{y})_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \tag{135}$$

where y is a vector of the true values, $\hat{y}$ is a vector of the predicted values and N is the total number of predictions. The mean square error is the loss function most commonly used for regression problems. The loss is the average of the square of the error over the training sample set. By squaring the error, the MSE loss function significantly penalises predictions with a high error. The MSE error is typically recommended for data sets with minimal noise and with normally distributed data. The LC loss is defined as,

$$L(y,\hat{y})_{\text{LC}} = \sum_{i=1}^{N}\log(\cosh(\hat{y}_i - y_i)). \tag{136}$$

The Log Cosh error acts similarly to the MSE in that for small errors, the LC function is approximately equal to $\frac{1}{2}(\text{error})^2$. However, for larger errors, the function is approximately equal to $|\text{error}| - log(2)$. The LC is less sensitive to outliers and single data points with

large deviations and will not avoid heavily weighting outliers. This makes the LC more appropriate for data sets with noise. Finally, the CS loss function is defined as,

$$L(y, \hat{y})_{\text{CS}} = -\sum_{i=1}^{N} \frac{\hat{y}_i \cdot y_i}{|y_i||\hat{y}_i|} \tag{137}$$

This loss function is unique in that it provides similarity measures between two predictions rather than an error. The loss function calculates only the alignment of predictions rather than the magnitude, which is useful to maximise the proximity between predictions and targets.

9.4.3 Network Architecture and Hyperparameters

Multiple different network architectures were tested for the regression. Each architecture can be characterised by multiple hyperparameters, which determine the size and complexity of the network. More complex networks increase the variance but are also more prone to overfitting. During testing, it was found that the performance of the networks did not vary massively between different architectures. As with the classifiers, dense, fully connected networks are trained to simplify the optimisation procedure. Figure 122 shows the network structure chosen for the regressions. The structure consists of an input layer with 16 kinematic inputs, a batch normalisation layer, subsequent hidden layers and an output layer. Table 41 shows the hyperparameter ranges tested for the networks. The hyperparameters are optimised using Bayesian optimisation, with the target metric being the MSE of the validation set.

Best CS Models	
Parameters	Range
Activation	[ELU, RELU, swish] [149]
Output Activation	tanh
Layers	[1, 8]
Nodes	[16, 320]
L1	[0, 0.01]
L2	[0, 0.01]
Learning rate	[0.0001, 0.01]
Dropout	True/False ([0, 0.2])
Batch Size	(64, 10000)

Table 41: The range of hyperparameters tested during Bayesian optimisation. L1 and L2 parametrise the size of the L1 and L2 regularisation parameter [150] applied to each hidden layer. Dropout layers were also tested, indicated by true if dropout is used. The range of the dropout parameters is also given.

The output activation function was limited to just the tanh function to ensure the regression was bounded between $[-1, 1]$. The networks were also tested with and without

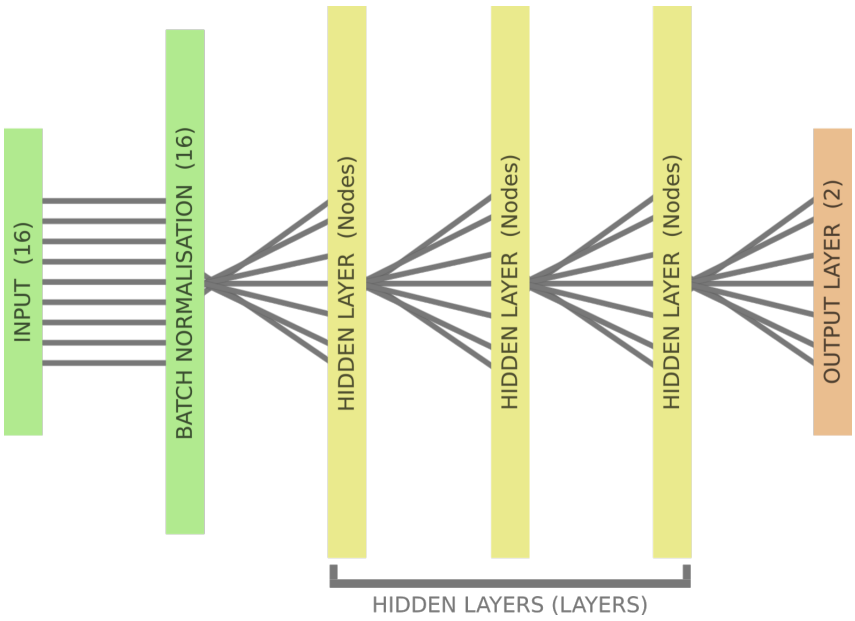


Figure 122: A graphical depiction of the network structure used by the regressions. The terms "Nodes" and "Layers" refer to the attributes in Table 41

dropout layers between the dense hidden layers, referenced in the table. With the batch normalisation layer and L1 and L2 regularisation [150], the dropout layers introduced too much noise into the regression and were dropped for the final regressions. In addition, the SELU activation function was initially tested; however, technical issues with implementing the networks in the analysis framework led to dropping the activation function.

9.5 Regression Selection

After hyperparameter optimisation, the next step is to select the final regression networks. The networks are trialled to optimise the MSE on the validation set. However, as discussed, the MSE significantly penalises values with a high error. Similarly, the networks with the highest cosine similarity may not equate to the networks with the best MSE. To get a fair comparison of each of the different networks, a selection criteria is chosen to rank the regressions. The method used to rank the regression in this section proceeds as follows:

1. Select the best 100 regressions based on MSE.
2. Order by the average of bin-by-bin Euclidean distance between the two predictions
3. Drop the networks where the Euclidean distance between the truth and corresponding simulated sample falls below a threshold.

The selection criteria ensure the most accurate models are chosen for the ranking. The models are ranked by the bin-by-bin Euclidean distance between the distribution of the regression and the true distribution. The binned comparison is unbiased towards a specific loss function and is slightly more lenient as it compares the average per bin. The

final criteria drops any networks where the distance between the regression prediction of the simulated and truth samples differ significantly. These criteria are chosen to ensure only models that generalise to simulated data are ranked.

9.5.1 Results and Validation

The hyperparameters of best regressions using the ranking are shown in Tab. 42. A comparison of truth, training and testing for each regression output is shown in Fig. 123 for each polarisation. The figures show no significant overfitting present in the regressions. Furthermore, the regressions each appear to model the truth $\cos\theta$ value of the leading lepton better than the sub-leading lepton. Figure 124 shows a comparison of the regression and truth two-dimensional $\cos\theta$ distributions. The figures show reasonable convergence to the true $\cos\theta$ value in some regions of the distribution. However, it is apparent that the predicted and true distributions are very different.

Best Models			
Parameters	Regression LL1	Regression LT1	Regression TT1
Activation	ELU	ELU	ELU
Layers	3	4	2
Nodes	160	192	92
L1	000376	0.005000	0.000
L2	0.000556	0.000	0.00049
Learning rate	0.002580	0.0001	0.0001
Batch Size	4360	4771	747

Table 42: The hyperparameters of the selected regression models. All three models were trained with the cosine similarity loss function and no dropout layers.

The regression conservatively overpredicts the mean. This indicates that the networks potentially lack the variance necessary for convergence. However, larger networks were trialled and showed limited improvement. This is likely a feature of the physics. Polarisation information is lost with the neutrinos and is only available partly in the E_T^{miss}. The regression appears to struggle to understand how to assign the E_T^{miss} distribution between the leptons. As a result, the regression essentially has to make an educated guess, and for events with the highest uncertainty, the regression convergences towards the mean. Predicting the mean will give the lowest average error for incorrect predictions.

Figure 125 shows the regression tested with simulated samples. Comparing Figs. 123 and 125, the regression prediction with the truth training sample used for training aligns most closely to the corresponding simulated sample. The detector effects appear to only marginally affect the shape of the regression. The modelling of the regressions with data is shown in Fig. 126 in the experiment signal region. As has been observed for the validation of the classifiers, there is a slight excess in data compared to the MC predictions. However, Fig. 126 shows no significant mismodelling by the regressions.

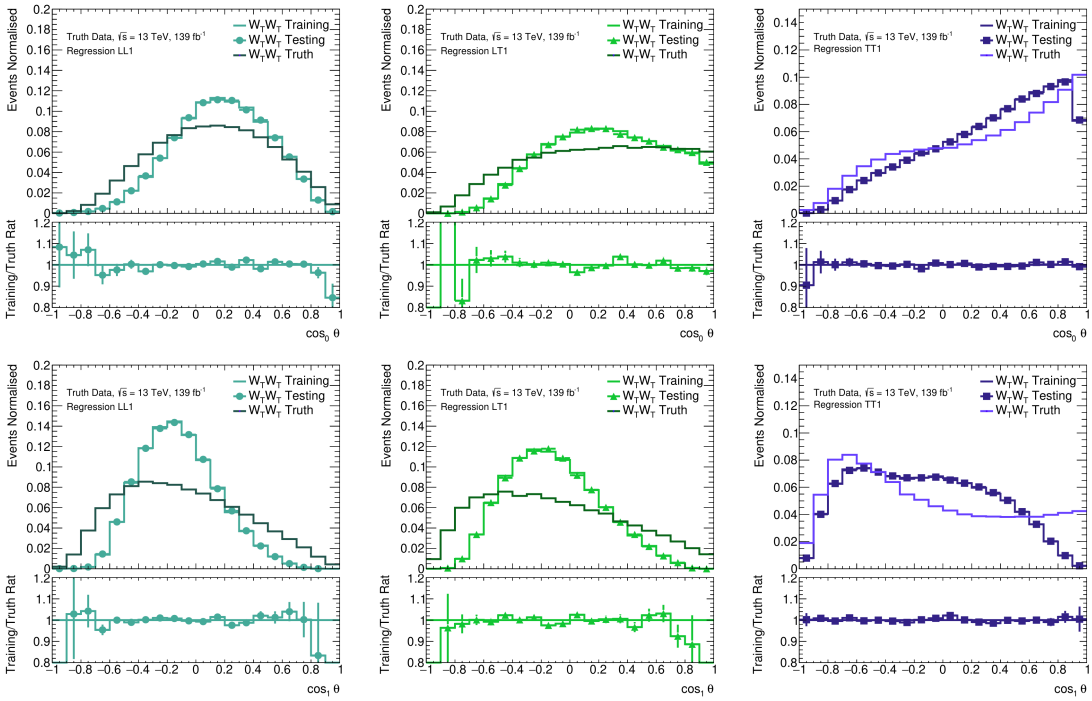


Figure 123: The training and testing $\cos\theta$ prediction by the regression, and the true $\cos\theta$ value, for the leading and sub-leading lepton. The regressions shown are detailed in Tab. 42.

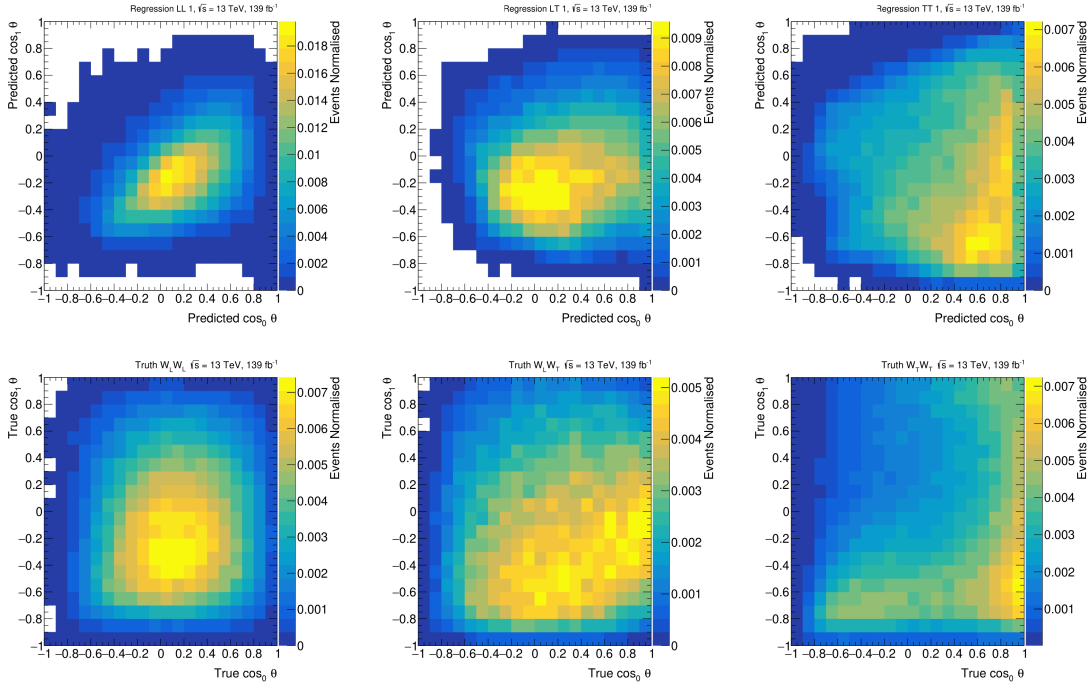


Figure 124: The regression $\cos\theta$ prediction using the testing data, and the true $\cos\theta$ value, for the leading and sub-leading lepton. The regressions shown are detailed in Tab. 42.

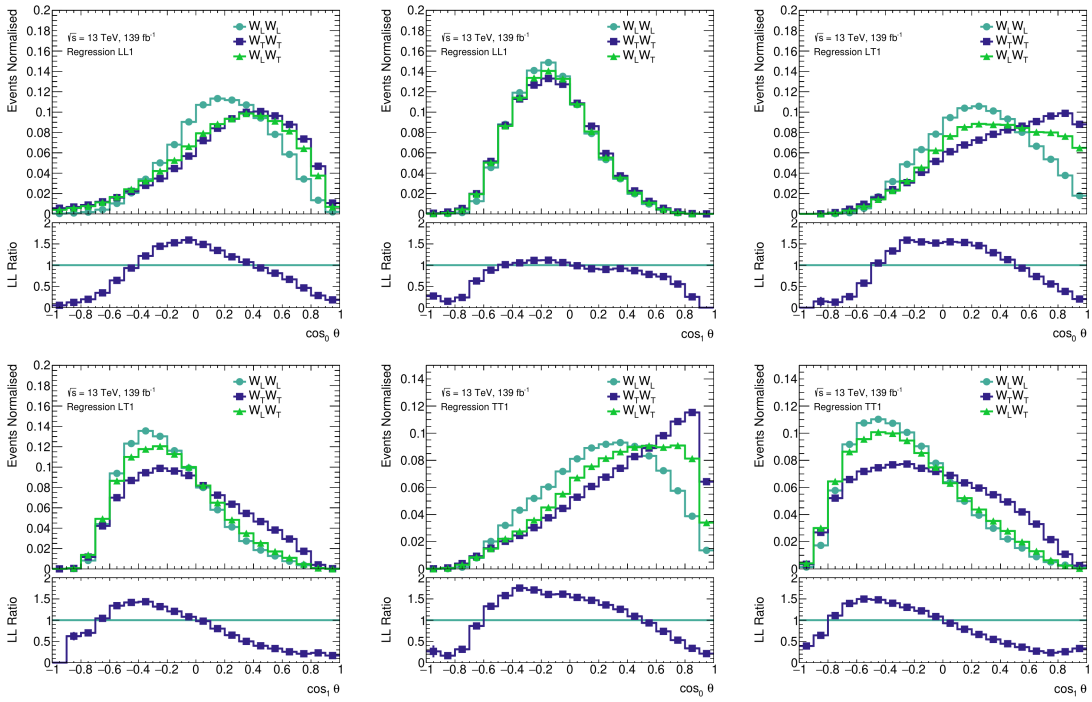


Figure 125: The regression $\cos\theta$ prediction for each simulated polarisation sample, Tab 30. The regressions shown are detailed in Tab. 42.

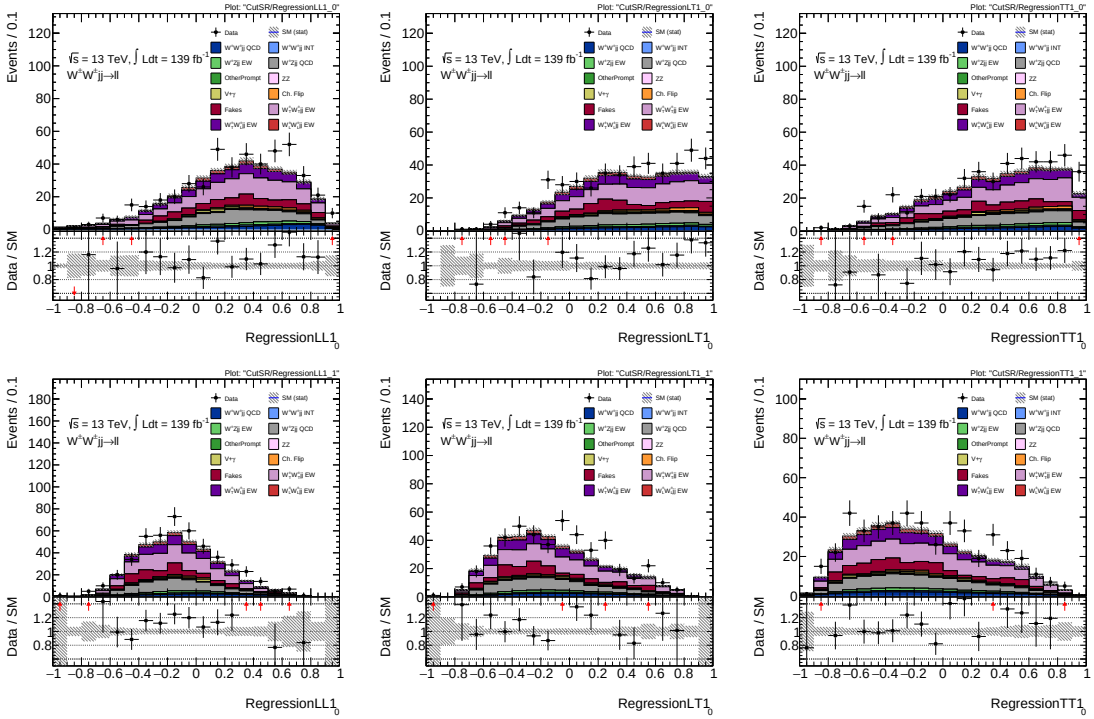


Figure 126: The regression $\cos\theta$ prediction in the signal region compared to data. The regressions shown are detailed in Tab. 42.

9.5.2 Polarisation Classifiers

The aim of performing the $\cos\theta$ regression was to model a kinematic variable very intrinsically connected to polarisation. Figure 125 shows a reasonable separation between the regression predictions for the different polarisation configurations. As a result, the regressions trained in this section were tested as additional kinematic input variables to the polarisation classifiers. The regressions were tested in the input feature selection procedure, detailed in Sec. 7.7. Each output for the regressions shown in Tab. 42 is included as an input.

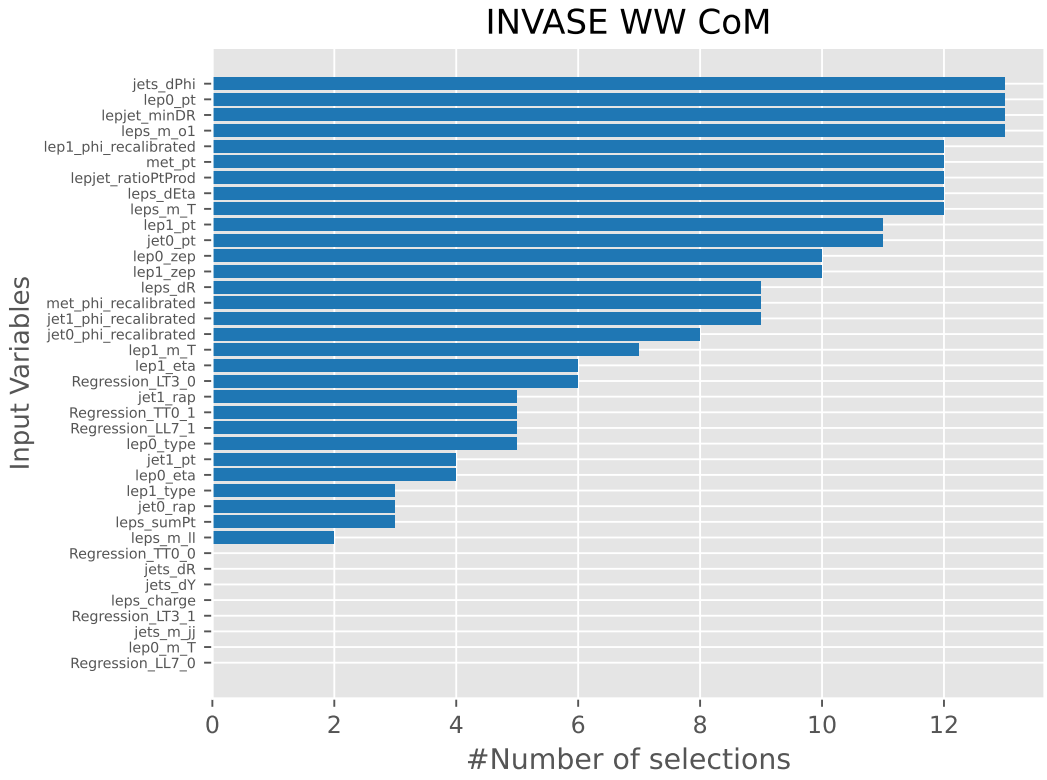


Figure 127: The ranking of the input variables for the polarisation classifier in the WW CoM with the regression models in Tab. 42 included as inputs.

Figure 127 shows the ranking of the input features of the polarisation classifiers in the WW CoM frame with the regression. The figure shows that the regressions are ranked very low. Each of the regression models is selected a few times, but only for a single output. The ranking shows that the regression models do not add much to the discrimination of polarisation. As the differences in the regression predictions shown in Fig. 125 originate from the differences in the kinematic variables for the polarisation samples, the ranking implies classifiers do not gain intrinsic information from the regression variables, and the kinematic variables are sufficient to classify polarisation. Moreover, the addition of the regressions to the polarisation classifiers introduced computational issues within the analysis framework by significantly increasing the processing time. For this reason, the regressions are not used in the final measurement of this experiment.

9.6 Idealised Regression

As discussed in the results, the loss of polarisation information associated with the missing neutrinos significantly impacts the ability of the regression to regress the $\cos\theta$ distribution. To illustrate this explicitly, this section will show an idealised regression, which includes neutrino momenta. The hyperparameters of TT regression, labelled by RegressionTT1 in Tab. 42, are used to perform the idealised regression, with 24 input variables where the additional eight correspond to the four-momenta of the truth neutrinos. The idealised regression is also trained with the MSE.

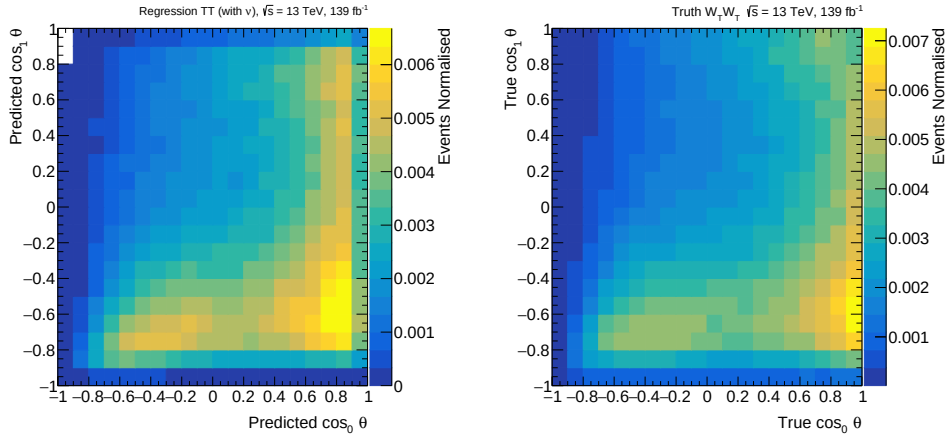


Figure 128: The predicted $\cos\theta$ distribution by the idealised regression, with the neutrino momenta as inputs, compared to the true $\cos\theta$ value.

Figures 128 and 129 show the idealised regression predictions and the truth predictions. The figures show a much clearer convergence from the regression and only minor deviations at the tail of the predictions.

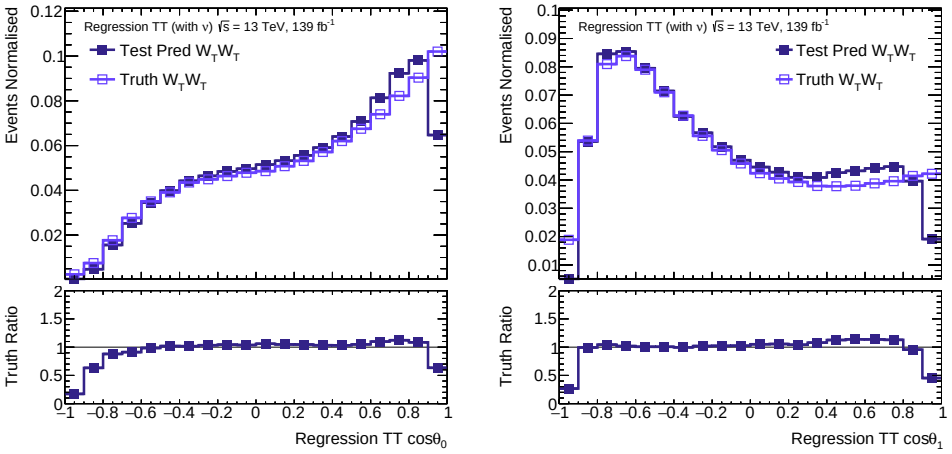


Figure 129: The prediction for leading and sub-leading $\cos\theta$ distribution by the idealised regression, with the neutrino momenta as inputs, compared to the true $\cos\theta$ value.

10 Experiment Measurement

This experiment measures the fiducial cross-sections of polarised electroweak-induced $W^{\pm}W^{\pm}jj$ production. The cross-section measurements are performed for electroweak production with at least one longitudinal gauge boson LX and for purely longitudinal production LL in the WW and pp CoM frame. The significance of the signal hypothesis is also calculated for each measurement.

The sources of uncertainty associated with this experiment can be subdivided into two categories: statistical and systematic. Statistical uncertainties originate from having finite experimental data, which limits the precision of a measurement. Systematic uncertainties describe the remaining uncertainties. They originate from having an "imperfect" experiment with finite resolutions on detector measurements and limited accuracy of theoretical predictions. This section will discuss the modelling of the systematic uncertainties for this measurement, which includes a detailed overview of each systematic and its impact.

The sources of systematic uncertainties can be divided into three categories: experimental, theoretical and MC statistics. The experimental uncertainties encompass the effects of the precision of the ATLAS detector and have a subsequent impact on the reconstruction and event selection of physics objects. In addition, the experimental uncertainties include the uncertainties on the data-driven methods used to estimate the charge-misidentification and non-prompt lepton backgrounds and also the uncertainty associated with the LHC luminosity. The theoretical uncertainties characterise the limited orders available in perturbation theory in the theoretical predictions of cross-sections. Finally, the MC uncertainties describe the statistical uncertainty of the MC generator. The systematic uncertainties in this analysis are propagated as shape differences in observables and as normalisation factors on the event yield. The modelling of the experimental uncertainties remains unchanged from the differential.

10.1 Theory Modelling

This section will describe the general modelling of the theoretical predictions. There are a number of sources of theoretical uncertainty associated with the fiducial cross-section calculation. The MC signal samples are generated at LO with a finite precision. The calculation of partonic cross-sections has a number of uncertainties arising from the limited precision in perturbation theory. Furthermore, to account for higher-order effects, this experiment applies an EW and QCD NLO correction to the signal samples. Each correction has an associated uncertainty. The subsequent section will discuss the NLO correction and the propagation of the uncertainties. Theoretical uncertainties are considered for each of the three polarisation samples. Uncertainties are also estimated for the QCD-induced $W^{\pm}Zjj$ background.

10.1.1 Partonic Cross-Sections

This section gives a general outline of the physics describing the calculation of partonic cross-sections to motivate some of the sources of theoretical uncertainty. The reader is referred to [13] and [151] for a detailed overview of the concepts behind renormalisation

and factorisation. The physics of proton-proton collisions occurs at different energy scales. The hard scatter collision is a high-energy physics process involving elementary point-like particles. In the case of proton-proton collisions, these particles are partons. Before scattering, however, the partons are color confined and the physics is described by low-energy QCD. The physics of the collisions, before, during and after, occurs at different energy scales. The factorisation theorem can express the cross-section for proton-proton collisions, [151] as,

$$\sigma(pp \to X) = \sum_{i,j} \int \int dx_1 dx_2 \int f_i(x_i, \mu_F) f_j(x_2, \mu_F) d\sigma_{i,j\to X}(x_i, x_j, \mu_R, \mu_F) \tag{138}$$

where the sum is performed over the partons of each proton i and j, $f_{i/j}(x_i, \mu_F)$ are the parton distribution functions (PDFs) which represent the probability of observing the parton with momentum fraction $x_{i/j}$, $d\sigma_{i,j\to X}$ is the hard-scatter cross-section and $u_{F/R}$ are the factorisation and renormalisation scales respectively. The theorem factorises the total hadronic cross-section as the convolution of the PDFs, describing the low-energy physics and the cross-section of the hard scattering between the constituent partons. Crucially, this separates the low-energy phenomena, bound by colour confinement in the non-perturbative QCD regime, from the high-energy physics, where the partons are asymptotically free and can be modelled with perturbative QCD. While PDFs cannot be calculated with perturbative QCD, they are universal and process-independent.

Cross-section calculations in the perturbative regime often lead to divergences. There are two main types of singularities: Ultraviolet (UV) and Infrared (IR). The UV divergences arise due to unconstrained momenta in closed-loop diagrams. IR divergences are generally associated with massless particles. IR singularities appear due to massless particles approaching the low momentum limit, referred to as a soft divergence and or when a massless particle radiates a massless particle, referred to as collinear divergences. Both types of singularities can be resolved with appropriate regularisation methods. In the case of UV divergences, the additional parameters introduced by the regularisation scheme are absorbed into the measurable physical constants. The UV divergences are absorbed into the renormalised masses and coupling constant and introduce a dependence on the renormalisation scale μ_R. A similar procedure is applied to deal with IR singularities whereby factorisation absorbs IR singularities into the PDFs and introduces a dependence on the factorisation scale μ_F.

The renormalisation scale characterises the physics at high energy and asymptotic freedom through the running of the coupling constant α_s. For high-energy interactions, the effective coupling is small, and perturbative QCD can be used to model the interactions. The factorisation scale describes the boundary between the physical scales describing low and high-energy phenomena. By the factorisation theorem, the cross-section can be separated into the physics of perturbative and non-perturbative QCD. The factorisation scale qualitatively represents this boundary. The caveat to the factorisation and renormalisation is the introduction of nonphysical scales into the physical observable. The cross-section for a process should be invariant to a choice of scale. The sensitivity to the renormalisation and factorisation scale is only reduced by performing calculations at higher orders of perturbation theory.

Multiple sources of theoretical uncertainty are associated with calculating finite-order partonic cross-sections. There are uncertainties associated with the missing higher-order terms in the perturbative expansion. There are also uncertainties associated with PDFs and the value of the strong coupling constant α_s. The procedure for estimating each uncertainty is applied following ATLAS recommendations outlined by the Physics Modelling Group [152]. The subsequent sections will discuss uncertainty estimation with scale and PDF variations.

10.1.2 Scale Variations

The theoretical uncertainties associated with the missing higher-order terms are estimated with scale variations. Scale variations vary the renormalisation and factorisation scales by up to a factor of two. The scale variations are generated for the $W^{\pm}W^{\pm}jj$ and $W^{\pm}Zjj$ MC samples as additional event weights. The uncertainty associated with the scale variations is calculated as recommenced by [152] using the 7-point scale variation method. The variations are generated pairwise, with up and down variations of μ_R and μ_F by factors of 2. Off-diagonal variations in different directions are not applied as variations [152]. The uncertainty is estimated by taking the envelope of the uncertainty for each variation, defined by,

$$\max\left[\mathcal{O}(\mu_{R,i},\mu_{F,i}) - \mathcal{O}(\mu_{R,0},\mu_{F,0})\right] \tag{139}$$

where $\mathcal{O}$ is the observable of interest, i is the index of a given scale variation and 0 is the index of the central scale [152]. The nominal factorisation and renormalisation scales are chosen to be at $\mu_R = \mu_F = \sqrt{p_T^{j_1} p_T^{j_2}}$, where $p_T^{j_i}$ is the transverse momentum of the leading and sub-leading jet, $i = 1$ and $i = 2$.

10.1.3 PDF Variations

PDFs are universal and process-independent and can be determined by experiment. PDFs are typically obtained by fitting experimental data, often from deep inelastic scattering experiments. When calculating PDFs, there are several sources of uncertainty [152, 153], experimental and theoretical. The experimental uncertainties are associated with the measured dataset used for the PDF fits. There are also uncertainties related to the functional form of the fit. There are additional theoretical uncertainties due to missing higher-order terms in the DGLAP evolution, which predict the evolution of PDFs at different energy scales. The uncertainty associated with the PDFs is estimated by varying the internal parameters of the nominal PDF used by the MC generator and by varying the nominal PDF set [153].

PDF uncertainties are estimated for the MC polarised signal samples and the WZ QCD samples. The polarised signal samples are generated using the NNPDF3.0NLO [154] PDF set at NLO in QCD and with $\alpha_s = 0.118$, alongside the prescribed internal PDF set variations. The $W^{\pm}Zjj$ QCD samples are generated using the NNPDF3.0NNLO [154]. The recommendation for calculating the PDF uncertainties with the PDF sets is

to take the mean of the associated variations, with standard deviation given by,

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(\sigma_i - \sigma_0)}, \tag{140}$$

where σ_i is the nominal PDF set. The PDF uncertainties for the QCD-induced $W^{\pm}Zjj$ background are estimated using these recommendations. However, the theoretical uncertainties associated with the PDFs are propagated differently for the signal sample due to the NLO correction. Section 10.2.4 discusses this in further detail.

10.2 Polarised NLO corrections

The modelling of polarised vector boson production processes at NLO is an ongoing area of active research. As discussed in Chapter 4, defining polarisation for vector bosons can be non-trivial. At LO, electroweak induced $W^{\pm}W^{\pm}jj$ production includes contributions from diagrams with non-resonant $W^{\pm}$ bosons. The polarised templates used in this experiment are generated at LO using MadGraph5_aMC@NLO, interfaced with Pythia. The generator applies the NWA to negate non-resonant diagrams where the definition of polarisation is unclear. Generating NLO predictions introduces several additional complications at the generator level. NLO effects can be propagated by generating samples at LO and applying a correction to account for NLO effects. However, at present, complete NLO corrections are unknown for polarised $W^{\pm}W^{\pm}jj$. Complete NLO corrections are, however, available for unpolarised $W^{\pm}W^{\pm}jj$ production [155] and are applied in the differential analysis.

During the development of the work, a newer version of the MC generator Sherpa was released with the ability to simulate polarised vector boson scattering with NLO QCD effects [156]. In this experiment and the polarisation analysis, the newer generator version is used to extract an NLO QCD correction factor and is applied to the MadGraph samples. Introducing NLO corrections is crucial to improving the accuracy of theoretical predictions and accounting for higher-order polarisation effects. The modelling of NLO corrections presented here is not the author's work and is inherited from the polarisation analysis. At the time of writing, research on the ATLAS polarisation measurement is still ongoing. A citation is currently unavailable, except for the internal documentation, which is only accessible to members of the CERN collaboration.

10.2.1 NLO $W^{\pm}W^{\pm}jj$ production

As an introduction, this subsection will outline the general theoretical aspects of NLO predictions for $W^{\pm}W^{\pm}jj$ production. The outline follows the Ref. [155], with the reader referred to the literature for a more detailed overview of the NLO corrections associated with unpolarised $W^{\pm}W^{\pm}jj$. As discussed in Chapter 4, the hadronic process,

$$pp \to W^{\pm}W^{\pm}jj \to l^{\pm}\nu l^{\pm}\nu jj \tag{141}$$

can be divided into three gauge-invariant production modes at LO: electroweak, QCD, and the interference associated with the electroweak and QCD diagrams. The LO production modes are defined at fixed orders of the electroweak and QCD coupling constants.

Electroweak production refers to the order $\mathcal{O}(\alpha^6)$, QCD production at $\mathcal{O}(\alpha^4\alpha_s^2)$ and the interference at $\mathcal{O}(\alpha^5\alpha_s)$. The tree-level VBS diagrams are only present in the electroweak production mode.

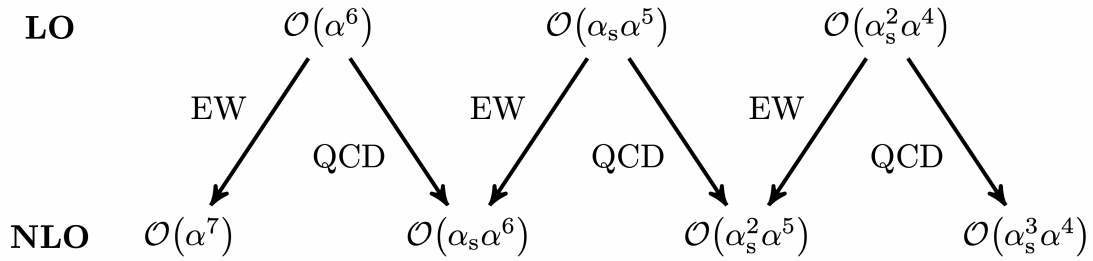


Figure 130: The contributing orders at LO and NLO for $W^\pm W^\pm jj$ production. Figure taken from [155].

At NLO, diagrams receive contributions from both real and virtual emissions in QCD and EW. There are four possible contributes at fixed orders, $\mathcal{O}(\alpha^7)$, $\mathcal{O}(\alpha^6\alpha_s)$, $\mathcal{O}(\alpha^5\alpha_s^2)$, $\mathcal{O}(\alpha^3\alpha_s^4)$. As shown in Fig. 130, each order is associated with the application of NLO EW or QCD corrections to the LO diagrams. For the orders $\mathcal{O}(\alpha^6\alpha_s)$ and $\mathcal{O}(\alpha^5\alpha_s^2)$, a separation into electroweak and QCD-induced production is no longer possible. The NLO $\mathcal{O}(\alpha^6\alpha_s)$ sample contains diagrams with NLO QCD effects applied to the LO electroweak diagrams and NLO EW effects applied to the LO interference contributions.

10.2.2 Polarised NLO $W^\pm W^\pm jj$ production

The Sherpa MC generator can simulate polarised cross sections for vector boson production processes [156]. To meaningfully assign polarisation to vector bosons, the NWA approximation is used by the generator, which replaces the $W^\pm$ boson propagator with the delta function,

$$\frac{1}{q^2 - m_W^2 + i\Gamma_W m_W} \to \frac{\pi\delta(q^2 - m_W^2)}{\Gamma_W m_W}, \tag{142}$$

which forces all intermediate $W^\pm$ bosons as on-shell. This suppresses contributions from single-resonant and non-resonant diagrams. However, not all the propagators are replaced with this approximation. The propagator for gauge bosons decaying hadronically remains unchanged in the NWA. As an example, consider a leading-order s-channel diagram for electroweak induced $W^\pm W^\pm jj$ production with the hadronic decay of a $W^\pm$ illustrated in Fig. 131. For this and other s-channel diagrams, the propagator term associated with the W boson, which decays hadronically, can lead to divergences. At leading-order, these divergences are mitigated by the m_{jj} selection criteria, which stays clear of the pole.

Fig. 131 illustrates a similar s-channel diagram but with higher-order QCD effects. At NLO in QCD, real emissions can lead to additional jets. This means the m_{jj} selection criteria may not be applied to the jet associated with the W boson decays, and the divergence arising from the propagator is unconstrained. To avoid potential divergences associated with the s-channel diagrams, the s-channel diagrams are dropped for the production of polarised NLO QCD samples. The contribution from the NLO s-channel diagrams is expected to be small but not negligible and is accounted for in the correction.

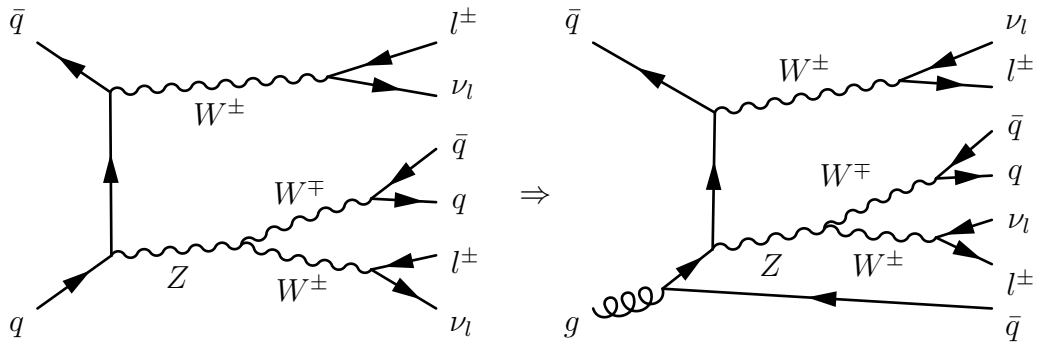


Figure 131: The figure on the left shows an s-channel diagram for $W^{\pm}W^{\pm}jj$ production. The figure on the right shows an equivalent s-channel diagram with higher-order QCD effects.

10.2.3 Derivation of the NLO QCD correction

There are a number of considerations that need to be made when deriving the NLO QCD corrections to be applied to the polarised LO MadGraph samples. To derive the correction, multiple electroweak-induced samples are generated with Sherpa [126]:

1. $W^{\pm}W^{\pm}jj$: Sherpa polarised electroweak production at LO.
2. $W^{\pm}W^{\pm}jj$ (t-chan): Sherpa polarised electroweak production at LO with missing s-channel diagrams.
3. $W^{\pm}W^{\pm}jj + 0,1j$ (t-chan): Sherpa polarised electroweak production at NLO QCD.
4. $l^{\pm}\nu l^{\pm}\nu jj + 0,1j$ (t-channel): Sherpa unpolarised electroweak production at NLO QCD with missing s-channel diagrams.
5. $l^{\pm}\nu l^{\pm}\nu jj + 0,1j$: Sherpa unpolarised electroweak production at NLO QCD.

where $+0,1j$ refers to the additional jets associated with the higher-order QCD corrections. The correction factor is derived for each polarisation sample by comparing the polarised and unpolarised samples,

$$\text{corr} = \left(\frac{W^{\pm}W^{\pm}jj + 0,1j \text{ (t-chan)}}{W^{\pm}W^{\pm}jj}\right)_{\text{pol}} \cdot \left(\frac{l^{\pm}\nu l^{\pm}\nu jj + 0,1j}{W^{\pm}W^{\pm}jj + 0,1j \text{ (t-chan)}}\right)_{\text{unpol}} \tag{143}$$

The correction associated with the NLO QCD polarisation effects (the first term in Eqn. 143) is calculated by comparing the NLO prediction with the missing s-channel to the LO Sherpa sample. The author's work verified there are negligible differences between the Sherpa and MadGraph predictions at LO for the polarised samples. As such, the Sherpa LO sample is sufficient to extract the correction factor. Furthermore, the effects of the missing s-channel diagrams at LO are seen to be negligible [126]. This means an additional correction factor is not required to correct the missing s-channel diagrams at

LO, so the correcting factor is derived sufficiently by comparing the NLO correction with the missing s-channel with the LO sample.

The correction associated with the missing s-channel (the second term in Eqn. 143) is derived by comparing the unpolarised NLO sample to the sum of the polarised NLO samples with the missing s-channel diagrams. A comparison of the unpolarised NLO sample with the missing s-channel and the polarised NLO sample with the missing s-channel showed negligible differences [126]. The difference in the samples arises from the NWA used to define polarisation. However, as the effect of the NWA is negligibly small, the correction is derived by comparing the polarised NLO samples with the missing s-channel contributions.

To reflect the multivariate nature of this analysis, correction factors are applied for the full phase space of kinematic variables, referred to as the multivariate correction. The multivariate correction is implemented using deep neural networks, as described in [157]. The reader is referred to [157] detailed description of the training methodology and the application of binary classifiers to phase-space reweighting. The reader is further referred to [126] for a detailed description of the training procedure applied to derive the correction factors used by this experiment.

10.2.4 Theoretical Uncertainties

The LO polarised MadGraph signal samples and the NLO polarised Sherpa signal samples are generated with PDF, scale and α_s variations to estimate the theoretical uncertainty of the predictions. To avoid potential double counting, the theory uncertainties assigned in this experiment are derived solely from the higher-order Sherpa sample [126]. The following systematic variations were generated for the NLO Sherpa sample,

1. PDF variations,
2. Alternate PDF sets: PDF sets differing by their parametrisation and the input data used in the fit,
3. Scale Variations,
4. α_s variations.

The NLO effects are applied through the multivariate correction. Estimating the effect of each variation requires training an additional DNN for upwards of 50 systematic variations [126]. As this is impractical, the effect of the theoretical uncertainties is instead propagated by studying the effect of the variations on a single kinematic distribution. The effect of the uncertainties was found to be most significant for the m_{jj} [126]. As effects are most dominant for this distribution, propagating the effect of the variations by the m_{jj} is expected to cover the effects for the full kinematic space. The uncertainties are propagated by the m_{jj} as a reweighting factor defined as,

$$r(m_{jj}) = \frac{(l^{\pm}\nu l^{\pm}\nu jj + 0,1j)^{\text{sys}}}{(l^{\pm}\nu l^{\pm}\nu jj + 0,1j)^{\text{nom}}} \tag{144}$$

where "sys" refers to a given systematic variation and "nom" is the nominal variation. The final theoretical uncertainties are calculated by taking the envelope for each contribution. This is the recommended procedure for the scale and alternative PDF, but not for the PDF and α_s variations. This is not expected to impact the results as the dominant source of uncertainty is the scale uncertainty. Moreover, the PDF and α_s variations are seen to be small, so the envelope is a reasonable approximation.

10.2.5 NLO Electroweak Correction

At the time of writing, no NLO-EW corrections are known for polarised electroweak-induced $W^{\pm}W^{\pm}jj$ production. This experiment follows the strategy of the polarisation analysis and applies the unpolarised higher-order electroweak correction to the polarised samples. Full unpolarised NLO corrections have been calculated by [155]. The differential measurement implements the NLO correction by fitting the m_jj dependant differential cross-section for the LO $W^{\pm}W^{\pm}jj$ EW sample, with the NLO-EW predictions provided by [155]. The fit function is defined as [89, 88],

$$f(m_{jj}) = p_0 + p_1 \log \frac{m_{jj}}{GeV} + p_2 \log^2 \frac{m_{jj}}{GeV}, \tag{145}$$

where following the fit, $p_0 = 0.511$, $p_1 = 0.176$ and $p_2 = -0.018$. Figure 132 illustrates a comparison of the leading order electroweak $W^{\pm}W^{\pm}jj$ sample at fixed order $\mathcal{O}(\alpha^6)$, and the NLO-EW predictions at fixed order $\mathcal{O}(\alpha^7)$, with the calculated fit function shown by the black line. The strategy applied in this experiment differs from the CMS polarisation measurement. The CMS measurement only applies a NLO-EW correction to the transverse polarisation mode $W_T^{\pm}W_T^{\pm}jj$ [90]. The publication suggests that the effects at NLO are expected to impact the TT polarisation configuration most dominantly because polarised NLO calculations for electron-positron $W^{\pm}W^{\mp}$ production [158] show smaller impacts for production with longitudinally polarised $W^{\pm}$ bosons. As it is unknown how the NLO EW corrections affect polarisation, this experiment applies the NLO correction to each sample and takes a conservative 100% uncertainty on the correction.

10.2.6 Combined Correction

It is not immediately clear how the higher-order QCD and EW corrections should be combined. The corrections applied to the polarised LO electroweak samples constitute the NLO process at different fixed orders of the coupling constant. The combination of the correction follows the approach of [159] and is inherited from the polarisation analysis. When combining the effects of each correction, the total correction can be implemented either by adding or multiplying the individual correction. These combinations can be expressed as [126],

$$d\sigma^{+}_{\mathrm{NLO}} = d\sigma_{\mathrm{LO}}(1 + \delta_{\mathrm{QCD}} + \delta_{\mathrm{EW}}), \tag{146}$$

and,

$$d\sigma^{\times}_{\mathrm{NLO}} = d\sigma_{\mathrm{LO}}(1 + \delta_{\mathrm{QCD}})(1 + \delta_{\mathrm{EW}}), \tag{147}$$

where $\delta_{\mathrm{QCD(EW)}}$ is the relative correction associated with the NLO-QCD(EW) correction, $d\sigma_{\mathrm{LO}}$ represents the differential cross-section for the LO sample and $d\sigma^{+,\times}_{\mathrm{NLO}}$ the NLO differential cross-section for the additive and multiplicative correction respectively. Following,

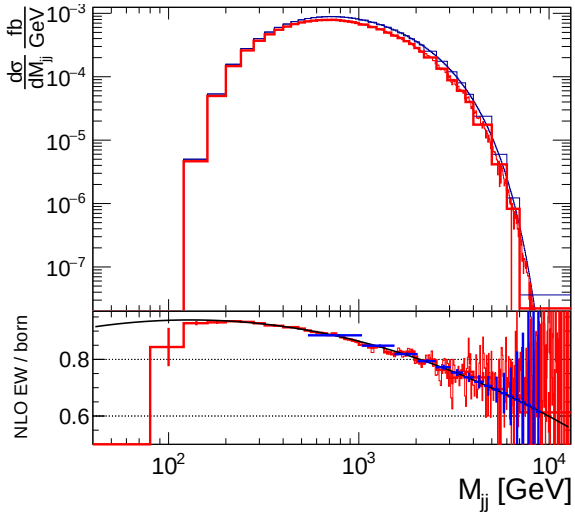


Figure 132: The differential cross-section as a function of m_{jj} for the EW LO prediction (blue histogram) and the NLO-EW prediction (red histogram) provided by the authors of [155]. The blue line shows the line of best fit, and the red line shows the analytical correction function. The fit reweighting function (black line), Eq. 145 is shown in the ratio of the two distributions, derived in [89, 88]. Figure taken from [89].

[159], the multiplicative correction $d\sigma^{\times}_{\text{NLO}}$ is taken as the nominal, and the difference in the two corrections is taken as an uncertainty for the combined QCD-EW correction.

10.3 Theoretical Predictions

10.3.1 Fiducial Region

Experimentally, the available phase space for data collection is limited by detector acceptance and data collection. For this experimental measurement, the theoretical cross-section is calculated in a fiducial region designed to match the experiment signal region as closely as possible. The predicted number of events in the fiducial region should be almost identical to the signal region. The cross-section is calculated at the particle level, after parton showering, but before detector simulation. The fiducial region is defined in Table 43.

Similarly to the signal region, an overlap removal procedure is applied for jets and electrons. The particle level jets are clustered with electrons, excluding neutrinos and muons, parallel with detector jet definitions. An electron and jet are defined to overlap if lepton-jet $\Delta R_{jl} < 0.4$. If the ratio of the $p^e_T/p^j_T < 0.5$, then the electron is assumed to originate from a hadronic decay within the jet and is removed. If the ratio $p^e_T/p^j_T > 0.5$, then the jet is removed.

10.3.2 Predicted Fiducial Cross-Sections

The theory predicted fiducial cross-sections are shown in Table 44. The cross-sections are calculated with the full NLO correction applied with the theoretical uncertainties applied

	$W^{\pm}W^{\pm}jj$ Fiducial Region
Leptons	
Number of leptons	2 prompt-leptons (e, μ) with equal electric charge
Number of τ	W boson decay with τ are vetoed.
Lepton-p_T	$p_T > 27$ [GeV]
Lepton-η	$\lvert\eta\rvert < 2.5$
Leptons ΔR_{ll}	$\Delta R_{ll} > 0.3$
Dilepton mass m_{ll}	$m_{ll} > 20$ [GeV]
Z-veto (ee channel)	$\lvert m_{ee} - m_Z\rvert > 15$ [GeV]
Jets	
Number	At least 2 jets
Leading p_T	Leading jet-$p_T \geq 65$ [GeV]
Sub-leading p_T	Sub-leading jet-$p_T \geq 35$ [GeV]
Jet-η	$\lvert\eta\rvert < 4.5$
Dijet mass, m_{jj}	$m_{jj} \geq 500$ [GeV]
Dijet rapidity	$\lvert\Delta y_{jj}\rvert > 2$
Missing transverse energy	
MET	$E_T^{miss} \geq 30$ [GeV]

Table 43: The fiducial region defined for both the polarised and differential analysis. Cuts applied here are applied at the particle level.

as described above. These fiducial cross-sections are taken from the polarisation analysis [126] and not the work of the author.

	fiducial cross section (fb)
LLWW	$0.272 \pm 0.0014(stat.) \pm 0.017(theo.)$
LXWW	$1.123 \pm 0.0045(stat.) \pm 0.074(theo.)$
LLpp	$0.172 \pm 0.0009(stat.) \pm 0.011(theo.)$
LXpp	$1.175 \pm 0.0052(stat.) \pm 0.077(theo.)$

Table 44: Summary of predicted next-to-leading order fiducial cross section for LL/LX in WW/pp center-of-frames with statistical uncertainties and theoretical uncertainties. Table taken from [126].

10.4 Distribution Bin Optimisation

This section will outline the binning optimisation procedure applied to these measurements. The same procedure is applied as in the polarisation analysis [126] and is not the

work of the author. The sensitivity to polarisation is partially dependent on the binning of the fitted histograms. A finer binning can improve the significance of a measurement but requires sufficient statistics in the data and the MC. Too many bins can lead to large statistical uncertainties. A binning optimisation is applied to find the optimal binning that balances, maximising the sensitivity to polarisation and ensuring sufficient statistics in each bin. In this experiment, the optimisation procedure is performed separately for each measurement.

10.4.1 Significance Estimate

The binning optimisation procedure tests a variety of different binning configurations to find the best, and for each test, the significance is calculated. Running the full fitting procedure for each test would be very computationally expensive and time-consuming. Instead, the optimisation procedure uses a simplified model to estimate the significance. The significance is calculated from a logarithmic likelihood (log-likelihood) function. The significance is calculated by taking the ratio of the log-likelihood functions corresponding to the background only and signal hypothesis [160]. Before defining the likelihood function, this section first introduces the parameters of the function. The number of expected events, as predicted by the MC, for a given bin i, is defined as,

$$n^i = S^i_{LL} + S^i_{LX} + S^i_{TT} + \sum_{bkg} b^i \tag{148}$$

where S^i_{XX} denotes the number of MC polarised signal events in bin i for a given polarisation configuration where $\mathrm{XX} = [LL, LT, TT]$ and $\sum_{bkg} b^i$ is the sum of the expected background events for the bin. In this fitting setup, the Asimov dataset refers to the data where the number of observed events is equal to the number of expected events. The number of predicted events is parameterised in terms of normalisation factors referred to as signal strength factors and a set of nuisance parameters. The full set of nuisance parameters is not considered as this would be very computationally expensive. Instead, a single nuisance parameter, $\bar{a}$, is applied to each signal and background prediction as an overall scale uncertainty. Within this setup, the number of predicted events for each measurement is defined as [126],

$$m^i(\mu, \mu_T, \bar{a}) = \begin{cases} \mu a_0 S^i_{LL} + \mu_T(a_1 S^i_{LT} + a_2 S^i_{TT}) + \sum_{bkg} a_{i+2} b^i & \text{LL measurement} \\ \mu(a_0 S^i_{LL} + a_1 S^i_{LT}) + \mu_T a_2 S^i_{TT} + \sum_{bkg} a_{i+2} b^i & \text{LX measurement} \end{cases} \tag{149}$$

where μ is the signal strength for the parameter of interest, which is measurement dependant, μ_T is a normalisation factor also to allow the other polarised predictions to float, and $\bar{a}$ is a vector compromising of the scale uncertainties for each sample. The simplified

log-likelihood is defined as [126],

$$\log(\mathcal{L}(\mu,\mu_T,\bar{a})) = \left[\prod_{i=0}^{N_{\text{bins}}} f_{Pois}(n^i|m^i) \prod_{i=0}^{3+N_{\text{bkg}}} f_{Pois}\left(\frac{1}{\Delta a_j^2}\middle|\frac{a_j}{\Delta a_j^2}\right)\right]$$
$$= \sum_{i=0}^{N_{\text{bins}}}(n^i\log(m^i)-m^i) + \sum_{i=0}^{3+N_{\text{bkg}}}\left(\frac{1}{\Delta a_j^2}\log\left(\frac{a_j}{\Delta a_j^2}\right)-\frac{a_j}{\Delta a_j^2}\right) + C(n,\Delta\bar{a}) \tag{150}$$

where $\Delta\bar{a}$ corresponds to the relative uncertainty associated with the scale factors and $C(n,\Delta\bar{a})$ is a coefficient of constant terms. The significance of discovery is calculated as a one-sided hypothesis test, which tests against the background-only hypothesis [160]. The significance is calculated for the Asimov dataset corresponding to $\mu=\mu_T=\bar{a}=1$. The log-likelihood ratio of the background only and the signal plus background hypothesis defines the test statistic [160],

$$q_0 = -2\log\left(\frac{\mathcal{L}(0,\hat{\mu}_T,\hat{a}}{\mathcal{L}(1,1,\bar{1}}\right) \tag{151}$$

where $\hat{\mu}_T$ and $\hat{a}$ are found by minimising the likelihood function with respect to $\mu=0$. The discovery p-value is given by,

$$p_0 = 1-\Phi(q_0), \tag{152}$$

and the significance as,

$$Z = \Phi^{-1}(1-p_0) = \sqrt{q_0}, \tag{153}$$

where Φ is the standard normal Gaussian CDF. The relative uncertainties are chosen following [126].

10.4.2 Optimisation Procedure

The distributions used as inputs to the fit are two-dimensional distributions of the $\text{DNNPol}_{\text{TT}}$ prediction on the x-axis and DNNssWW on the y-axis; examples are shown in Fig. 133. The binning optimisation procedure starts by splitting the distribution along the DNNssWW axis. The split distributions are then projected along the same axis to form multiple one-dimensional distributions of $\text{DNNPol}_{\text{TT}}$. The optimisation procedure tests various dissections but requires at least 50 background or transversely polarised $W^{\pm}W^{\pm}jj$ events in each region. The procedure starts by projecting the two-dimensional distribution into a single histogram and increases the number of splits until the total significance does not improve.

After splitting, the binning is optimised for each individual distribution. The procedure starts by requiring each bin to contain a minimum of 10 background or TT events. Bins are merged until the minimum number of background events is achieved. The merging procedure is repeated, going left to right and right to left along the distribution axis. The higher significance estimate is taken as the final. Finally, a merging procedure is performed for bin edges, which have the smallest effect on significance. Bins which only increase the total significance by a maximum of 0.01 are merged. The final binning optimisation results for each CoM frame and each network are shown in Table 45.

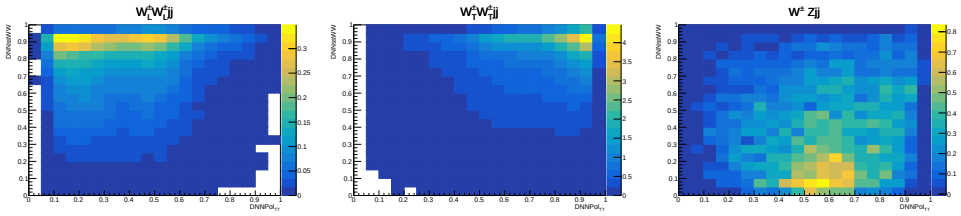


Figure 133: The DNNssWW vs DNNPolTT network predictions for the LL and TT signal samples, and the $W^{\pm}Zjj$ QCD-induced background. The distribution shows the number of expected events in each bin.

Binning		
WW CoM	DNNssWW	$\text{DNNPol}_{\text{TT}}$
LX	[0.0, 0.3, 0.6, 1.0]	[0.0, 0.35, 0.75, 1.0], [0.0, 0.3, 0.55, 0.8, 1.0], [0.0, 0.3, 0.45, 0.55, 0.65, 0.7, 0.85, 1.0]
LL	[0.0, 0.2, 0.4, 1.0]	[0.0, 0.4, 0.7, 1.0], [0.0, 0.35, 0.6, 1.0], [0.0, 0.2, 0.3, 0.45, 0.7, 1.0]
pp CoM	DNNssWW	$\text{DNNPolPP}_{\text{TT}}$
LX	[0.0, 0.5, 0.8, 1.0]	[[0.0, 0.55, 0.8, 1.0], [0.0, 0.35, 0.45, 0.65, 0.8, 1.0], [0.0, 0.4, 0.55, 0.65, 0.85, 1.0]
LL	[0.0, 0.2, 0.55, 1.0]	[0.0, 0.3, 0.45, 1.0], [0.0, 0.25, 0.6, 1.0], [0.0, 0.25, 0.35, 0.45, 0.65, 1.0]

Table 45: The bin edges of the distributions selected by the bin optimisation procedure for each measurement.

10.5 Statistical Model

The measurement performed in this analysis can be described in its simplest form as a counting experiment. As an introduction, this section will give a short overview of a simple counting model. The experiment aims to measure the number of events observed for a physics process, referred to as the signal. The number of expected signal and background events is predicted with a combination of MC and data-driven modelling. For simplicity, consider event predictions only using MC. Explicitly, the MC predicts N^{pred} events, which can be expressed as,

$$N^{\text{pred}} = S + B, \tag{154}$$

where S is the number of signal events predicted, and B is the number of background events predicted. The number of observed events is denoted by N^{obs}. In a counting experiment, the number of observed events is Poisson distributed with a mean λ. Assuming the modelling is correct, the mean should be equivalent to N^{pred}. However, to extract the fraction of signal events, it is common to parameterise the mean in terms of a continuous parameter μ. The parameter is referred to as the signal strength parameter, with $\mu = 0$ corresponding to the background-only hypothesis. With this parameterisation, the mean can be expressed as,

$$\lambda(\mu) = \mu S + B. \tag{155}$$

The aim of the measurement is to determine, given N^{obs} observed events, what value of μ best matches the data. Statistical this is equivalent to finding the value μ, which maximises the likelihood $\mathcal{L}$ of observing N^{obs} events. For the simple counting experiment,

the likelihood is expressed as

$$\mathcal{L}(N^{\text{obs}}|\mu) = \text{Pois}(N^{\text{obs}}|\mu S + B). \tag{156}$$

It is common to take the logarithm of the likelihood function and negate the function, referred to as the Negative Log Likelihood (NLL). This often simplifies the expression analytically.

The counting experiment outlined in this section can be described as a "perfect" experiment. In practice, experiments are not perfect and have finite resolutions on instrumentation. Similarly, theoretical predictions will also have associated uncertainties, for example, from missing higher-order terms in perturbative expansions. These different sources of uncertainties can affect the final predictions and should be included in the model. Consider the following example. The counting experiment measures the transverse momentum of an electron, and the experiment applies selection criteria to reject electrons with a low-p_T. The background acceptance is, therefore, sensitive to the uncertainty on the p_T measurement.

To illustrate how to include this in the experiment model correctly, assume that the calibration on the p_T measurement can be parameterised by α, and the effect on the background prediction is found to be given by $\theta(\alpha)$. A series of auxiliary measurements, $\vec{x}$, can be made to calibrate the uncertainty on the electron p_T. The likelihood function can now be expressed as

$$\mathcal{L}(N^{\text{obs}}, \vec{x}|\mu, \alpha) = \text{Pois}(N^{\text{obs}}|\mu S + B(1 + \theta(\alpha)) \cdot \mathcal{L}(\vec{x}|\alpha). \tag{157}$$

where $\mathcal{L}(\vec{x}|\alpha)$ is the likelihood of observing $\vec{x}$ for the continuous parameter α. The final likelihood is just a product of likelihood functions for each parameter. In practice, however, the auxiliary measurements required to calibrate the uncertainty are complex and often dependent on additional parameters. Furthermore, for a large experiment such as ATLAS, including the full likelihood function for each auxiliary measurement is impractical. For this reason, the likelihood associated with each uncertainty is instead typically estimated with a constraint term. Following the same example, the uncertainty on the electron p_T is measured to be $\pm\sigma$. The experiment calculates that the background yields vary by $\theta^{\pm}$ for $\pm\sigma$. Introducing a parameter α to constrain the uncertainty and assuming α is unit Gaussian distributed, the predicted background yield can be estimated as a function of $\theta(\alpha)$, after interpolating the calculated values $\theta^{\pm}$. The effect of the uncertainty can therefore, be incorporated into the likelihood function as

$$\mathcal{L}(N^{\text{obs}}|\mu, \alpha) = \text{Pois}(N^{\text{obs}}|\mu S + B(1 + \theta(\alpha)) \cdot \mathcal{G}(0|\alpha, 1). \tag{158}$$

where $\mathcal{G}$ is the Gaussian distribution and $\theta(\pm 1) = \theta^{\pm}$. The likelihood function representing the auxiliary measurement has been replaced by a constraint term in the form of a Gaussian. This approximation is reasonable if the likelihood associated with the uncertainty is in the Gaussian regime, which is assumed to be true for this example. In the terminology of ATLAS, this uncertainty is referred to as being profiled. Two important considerations must be made when using this approximation. Firstly, the nature of the constraint term should reflect its likelihood function. Secondly, the interpolation scheme used to model the impact on the predicted event yields to form a continuous function

should also be chosen carefully. Overall, the overarching goal of this example is to illustrate the general features of the likelihood model used in this experiment and to show how sources of uncertainties are propagated to the measurement.

10.5.1 Likelihood Fit

The statistical model used in this measurement is implemented using TRExFitter [161], which is an ATLAS internal framework for binned template profile likelihood fits. The framework is a wrapper for building models with HistFactory [162]. In the terminology used in this section, a region refers to a phase space of data defined by an event selection. Regions are required by definition to have disjoint event selections [162]. The term sample in this context refers to a set of scattering processes which can be added together incoherently [162]

The measurement in this experiment is performed using a binned maximum likelihood function. The binned likelihood function is expressed as,

$$\mathcal{L}(\vec{n} \mid \mu, \vec{\theta}) = \prod_{r \in \text{ regions}} \prod_{b \in \text{ bins in r}} \text{Pois}\big(n_b \mid e_b(\vec{b}, \lambda, \vec{\alpha}, \vec{\phi})\big) \cdot \prod_{j \in \vec{\theta}/\vec{\phi}} f_j(j) \tag{159}$$

where,

$\vec{\theta} = \{\lambda\} \cup \vec{\alpha} \cup \vec{\gamma} \cup \vec{\phi}/\mu$

n_b the number of observed events in bin b in region r,

e_b the number of expected events in bin b in region r,

$\vec{\gamma}$ set of nuisance parameters encoding MC statistical uncertainty,

$\vec{\alpha}$ set of nuisance parameters encoding the systematic uncertainties,

$\vec{\phi}$ set of unconstrained normalisation factors,

λ nuisance parameter associated with the luminosity.

The likelihood function is a product of Poisson distributions representing each bin of the fitted distribution. The likelihood function parameterises two types of systematic uncertainty, $\vec{\gamma}$ and $\vec{\alpha}$, respectively representing the MC statistical uncertainty in each bin and the theoretical and experimental systematic uncertainties. The final parameter, λ, corresponds to the uncertainty of the experiment luminosity. Each source of uncertainty is profiled and constrained with the recommended constraint term. The constraint terms for each source of uncertainty are defined as,

$$\begin{aligned} f_\gamma(\vec{\gamma}) &= \prod_{r \in \text{ regions}} \prod_{b \in \text{ bins in r}} \text{Pois}(m_b \mid \gamma_b \cdot \tau_b) \\ f_\lambda(\vec{\gamma}) &= \mathcal{G}(1 \mid \lambda, \Delta\lambda) \\ f_\alpha(\vec{\alpha}) &= \prod_{\alpha_i \in \alpha} \mathcal{G}(0 \mid \alpha_i, 1) \end{aligned}$$

where $\mathcal{G}$ is the Gaussian distribution, m_b is the summed number of observed MC events in bin b associated with the MC uncertainty, and τ_b is the number of expected events in bin

b before scaling by γ_b. The $\vec{\gamma}$ nuisance parameters are associated with the MC statistical uncertainty. The parameters characterise the uncertainty on the predicted number of MC events as being different from the true rate. The constraint term represents an auxiliary measurement where a new MC sample is generated, which fluctuates by $\gamma_b \cdot \tau_b$ [162]. A Gaussian distribution constrains the remaining uncertainties. The nuisance parameters $\vec{\alpha}$ represent the measurement of the $\pm\sigma$ variation in the uncertainty.

The number of expected events is given by,

$$v_b(\gamma_b, \lambda, \vec{\alpha}, \vec{\phi}) = \lambda \cdot \gamma_B \sum_{s \in \text{samples}} \left(v_b^s + \sum_{\alpha_i \in \vec{\alpha}} \sigma^s(\alpha_i)\right) \prod_{\phi_j \in \vec{\phi}^s} \phi_j \prod_{\alpha_i \in \vec{\alpha}} \eta(\alpha_i) \tag{160}$$

where v_b^s is the predicted number of events for a given sample. The experimental and theoretical uncertainties are divided into normalisation factors, η and shape effects, σ. This is to avoid negative expectations from down pulls in the systematic uncertainties [89, 88]. For each input, the effect of the event yields is known for the $\pm\sigma$. To generate continuous functions of α_i, each source of systematic uncertainty is interpolated and extrapolated. The interpolation and extrapolation scheme follows the recommendations of [162].

10.5.2 Measurement Configurations

The likelihood function is configured with three different experimental regions,

1. Signal region,
2. Low-m_{jj} region,
3. WZ control region; constrains the QCD-induced $W^{\pm}Zjj$ normalisation factor.

The binned distributions provided as inputs to the function in the signal region are the DNN outputs, binned as described in Section 10.4.2. A single binned, m_{jj} distribution is used for the WZ and low-m_{jj} control regions. The likelihood function aims to separately extract the signal strength factors associated with,

1. longitudinal production $W_L^{\pm}W_L^{\pm}jj$
2. production with at least one longitudinal gauge boson $W_L^{\pm}W_L^{\pm}jj + W_L^{\pm}W_T^{\pm}jj$

The function is configured with the following unconstrained normalisation factors $\vec{\phi}$,

1. μ signal strength factor scaling the signal sample,
2. $\mu_{T\Lambda}$ scale factor scaling the polarised non-signal sample, where $\Lambda = \{^{T \text{ for LX}}_{X \text{ for LL}}$,
3. μ_{WZ} scale factor for the WZ QCDbackground.

where μ corresponds to the signal strength factor, which scales either the LL or LX signal sample.

10.5.3 Signifance Calculation

The first part of this experiment will test for the discovery of polarised electroweak-induced $W^{\pm}W^{\pm}jj$ signal. The experiment will perform a hypothesis test to determine the significance of the signal hypothesis. In the field of particle physics, the accepted threshold for discovery is a significance exceeding 5σ, while a significance greater than 3σ is the threshold for evidence. The general template for likelihood-based statistical tests implemented in this measurement is detailed in Ref. [160]. For the discovery of new physics, the null hypothesis, H_0, is defined as the background-only hypothesis associated with the expected modelling of the known processes. The alternate hypothesis, H_1, is defined as the signal hypothesis associated with modelling the signal and background. The test for discovery calculates the p-value of the background-only hypothesis. The p-value is the probability of finding data with equal or greater incompatibility with a given hypothesis [160]. The p-value is related to the significance of the signal hypothesis by,

$$Z = \Phi^{-1}(1 - p_0), \tag{161}$$

where Φ^{-1} is the inverse of the standard normal Gaussian CDF, with the thresholds for rejecting the background-only hypothesis as discussed above.

For likelihood models as used in this experiment, the recommended test statistic for discovery is the likelihood ratio, defined as,

$$t_\mu = -2\log\left(\frac{\mathcal{L}(\mu, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\theta})}\right), \tag{162}$$

where $\hat{\mu}$ and $\hat{\theta}$ are maximum likelihood estimators of μ and $\vec{\theta}$ respective and, $\hat{\hat{\theta}}$ is the conditional maximum likelihood estimator of $\vec{\theta}$ for a fixed value of μ. In a test for discovery, the background-only hypothesis H_0 is a model in which $\mu = 0$. The alternate hypothesis, i.e. for the observation of signal, is for $\mu > 0$. The test statistic for discovery, denoted by q_0, is defined as,

$$q_0 = \begin{cases} -2\log\left(\frac{\mathcal{L}(\prime, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\theta})}\right) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0 \end{cases}. \tag{163}$$

In the asymptotic limit, the significance is given by,

$$Z = \Phi^{-1}(1 - p_0) = \sqrt{q_0}. \tag{164}$$

10.5.4 Upper Limit Calculations

This experiment calculates upper limits on the signal strength parameters for purely longitudinal production. For the case of finding upper limits, the recommended hypothesis test is reversed, with the signal hypothesis assigned as the null hypothesis H_0, tested against the background-only hypothesis, H_1. The test statistic for the upper limit calculation is denoted by q_μ and is defined as,

$$q_\mu = \begin{cases} -2\log\left(\frac{\mathcal{L}(\mu,\hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu},\hat{\theta})}\right) & \hat{\mu} < \mu \\ 0 & \hat{\mu} < 0 \end{cases}. \tag{165}$$

In the asymptotic limit, the p-value is calculated as [160],

$$p_0 = 1 - \Phi(\sqrt{q_\mu}) \tag{166}$$

The upper limit is calculated by the threshold at which the p_0 value is rejected. The accepted threshold in upper limits calculations is a value of 0.05. The value of μ can be calculated to find the value at the limit, with this μ corresponding to a 95% upper confidence limit on the signal-strength parameter.

10.6 Cross-section extraction

The final part of this experiment will extract the fiducial cross-sections from the data. The predicted fiducial cross-section is calculated from the theory predictions as,

$$\sigma^{\text{fid}}_{\text{pred}} = \sigma^{\text{tot}}_{\text{pred}} \cdot \mathcal{A}, \tag{167}$$

where $\sigma^{\text{tot}}_{\text{pred}}$ is the total cross-section calculated by the MC generator, integrating over the full-phase space, and $\mathcal{A}$ is the acceptance denoting the number of events which pass the fiducial selection. The acceptance is calculated by dividing the sum of events weights, passing the event selection with the total sum of weights. The uncertainty on the total cross-section is given by,

$$\Delta\sigma^{\text{fid}}_{\text{pred}} = \sigma^{\text{tot}}_{\text{pred}} \cdot \Delta\mathcal{A}. \tag{168}$$

Uncertainties associated with the MC generator and the calculation of total cross-section $\sigma^{\text{tot}}_{\text{pred}}$ cancel with $\sigma^{\text{fid}}_{\text{pred}}$. Theoretical and statistical uncertainties affect the predicted fiducial cross-section. The effect of the theory uncertainties is calculated by reweighting $\mathcal{A}$ with event weights associated with the theory variations.

The expected and observed fiducial cross-section refers to the measured cross-sections with a given dataset. The expected cross-section is calculated using the Asimov dataset, corresponding to finding the conditional maximum likelihood estimators associated with $\mu = 1$. The Asimov dataset is exactly the expected Standard Model prediction. The observed measurement is performed using the measured experimental data. The expected and observed cross-sections are affected by all systematic sources of uncertainty. However, for the theoretical uncertainty, only shape effects are considered, and the uncertainties are normalised to the nominal. The signal strength parameter relates the measured and predicted cross-section by,

$$\mu = \frac{\sigma^{\text{tot}}_{\text{meas}}}{\sigma^{\text{tot}}_{\text{pred}}} \tag{169}$$

such that,

$$\sigma^{\text{fid}}_{\text{meas}} = \sigma^{\text{tot}}_{\text{pred}} \cdot \mu, \tag{170}$$

$$\Delta\sigma^{\text{fid}}_{\text{meas}} = \sigma^{\text{tot}}_{\text{pred}} \cdot \Delta\mu, \tag{171}$$

The theory uncertainties are not propagated as the uncertainty on $\sigma^{\text{tot}}_{\text{pred}}$ is entirely anti-correlated with the theory uncertainty associated with $\sigma^{\text{tot}}_{\text{meas}}$.

10.6.1 Nuisance Parameters

An important cross-check of the fit results is to examine the impact on the nuisance parameters. Nuisance parameter pull, in units of standard deviation, is defined as,

$$\text{Pull}(\theta) = \frac{\hat{\theta} - \theta_0}{\Delta\theta} \tag{172}$$

where $\hat{\theta}$ is the maximum likelihood estimate of θ, and θ_0 and $\Delta\theta$ are Pre-fit nominal value and uncertainty. After the fit, deviations from the central value are indicated by the pull, which shows that features in the data have been constrained by a nuisance parameter. Large pulls, particularly beyond one standard deviation, potentially suggest discrepancies between data and prediction. Likewise, if post-fit uncertainty is significantly reduced, then this shows that the nuisance parameter is overconstrained by the data. The nuisance parameters should remain close to $\pm\sigma$ post-fit. The second parameter useful in assessing the impact of the nuisance parameters is the impact. The impact is calculated by, calculating the nominal post-fit values, obtaining $\hat{\mu}$. For each nuisance parameter, the best-fit values pre and post-fit values are shifted by pre- and post-fit upper and lower variation, and the fit is reformed to find μ_F^{up} and μ_F^{down}.The impact is then used to rank the nuisance parameters pre- and post-fit by the maximum of (μ_F^{up} - $\hat{\mu}$, μ_F^{down} - $\hat{\mu}$). The ranking shows the impact of the parameters on the best-fit values for μ

10.7 Experimental Uncertainties

Systematics associated with the experiment are accounted for following the recommendations of the ATLAS Combined Performance groups. The uncertainties in the reconstruction and calibration of physics objects are evaluated individually by each performance group. The computation of the uncertainties involves a careful treatment of correlations so that the uncertainties provided by each group can be treated as uncorrelated. The experimental uncertainties affect the predicted MC signal and background event yields, as well as the event kinematics. This analysis implements the propagation of the experimental uncertainties within the CAF Framework. Experimental uncertainties associated with the calibration of the object energy and momentum, both scale and resolution, are applied to all physics objects used in this analysis (leptons, jets, and MET). Furthermore, systematics accounting for reconstruction, identification, isolation and selection efficiencies are implemented as scale factors for the signal and background MC estimates. Finally, this measurement considers the uncertainties associated with the fake-factor measurements and the charge misidentification probabilities described in Refs. [89, 88].

Table 46 shows the systematics associated with electron reconstruction, which includes the electron energy and momentum calibration, the uncertainty in the trigger efficiency, the uncertainty in the reconstruction efficiency, and the uncertainty associated with the identification and isolation efficiencies for the analysis working points [110]. The systematics are not uniform across different regions of detectors; however, to simplify the fitting procedure, this analysis applies a combined variation for each source of uncertainty. This is an overestimate, but the relative impact of the electron uncertainties is small. An exception to this is made for the identification efficiencies, where a simplified model with 16

Electron Experimental Systematics			
Variation Name	Target	Comment	Nuisance Parameters
EG_SCALE_ALL	Electron Kinematics	Energy scale uncertainty	1 NP (up & down variations)
EG_SCALE_AF2	Electron Kinematics	Energy scale uncertainty	1 NP (up & down variations)
EG_RESOLUTION_ALL	Electron Kinematics	Energy resolution uncertainty	1 NP (up & down variations)
EL_EFF_Trigger_Total_1NPCOR_PLUS_UNCOR	Scale Factors	Trigger efficiency uncertainty	1 NP (up & down variations)
EL_EFF_Reco_Total_1NPCOR_PLUS_UNCOR	Scale Factors	Reconstruction efficiency uncertainty	1 NP (up & down variations)
EL_EFF_ID_TotalCorrUncertainty	Scale Factors	Identification efficiency uncertainty	1 NP (up & down variations)
EL_EFF_ID_CorrUncertaintyNP	Scale Factors	Identification efficiency uncertainty split	16 NP (up & down variations)
EL_EFF_ID_SIMPLIFIED_UncorrUncertaintyNP	Scale Factors	Identification efficiency uncertainty split	18 NP (up & down variations)
EL_EFF_Iso_Total_1NPCOR_PLUS_UNCOR	Scale Factors	Isolation efficiency uncertainty	1 NP (up & down variations)

Table 46: Experimental uncertainties for electron energy/momentum calibration, reconstruction, identification and isolation efficiencies [110, 89].

NP is used, which describes the uncertainty in identification efficiency in different $\eta \times p_T$ regions [110].

Muon Experimental Systematics			
Variation Name	Target	Comment	Nuisance Parameters
MUON_ID	Muon Kinematics	Momentum resolution uncertainty ID	1 NP (up & down variations)
MUON_MS	Muon Kinematics	Momentum resolution uncertainty MS	1 NP (up & down variations)
MUON_SCALE	Muon Kinematics	Momentum scale uncertainty	1 NP (up & down variations)
MUON_SAGITTA_RHO	Muon Kinematics	Charge dependent momentum scale uncertainty	1 NP (up & down variations)
MUON_SAGITTA_RESBIAS	Muon Kinematics	Charge dependent momentum scale uncertainty	1 NP (up & down variations)
MUON_EFF_TrigStatUncertainty	Scale Factors	Trigger efficiency uncertainty	1 NP (up & down variations)
MUON_EFF_TrigSystUncertainty	Scale Factors	Trigger efficiency uncertainty	1 NP (up & down variations)
MUON_EFF_RECO_STAT	Scale Factors	Reconstruction and identification efficiency uncertainty	1 NP (up & down variations)
MUON_EFF_RECO_SYS	Scale Factors	Reconstruction and identification efficiency uncertainty	1 NP (up & down variations)
MUON_ISO_STAT	Scale Factors	Isolation efficiency uncertainty	1 NP (up & down variations)
MUON_ISO_SYS	Scale Factors	Isolation efficiency uncertainty	1 NP (up & down variations)
MUON_TTVA_STAT	Scale Factors	Track-to-vertex association efficiency uncertainty	1 NP (up & down variations)
MUON_TTVA_SYS	Scale Factors	Track-to-vertex association efficiency uncertainty	1 NP (up & down variations)

Table 47: Experimental uncertainties for muon energy/momentum calibration, reconstruction, identification and isolation efficiencies [57, 89].

The experimental uncertainties associated with the muon reconstruction and trigger are given in Tab. 47. For reconstructed muons, uncertainties related to the momentum resolution are accounted for in the ID and the MS. Additional systematics are also included for the track-to-vertex association efficiency. As with the electron systematics, the impact of the muon systematics is expected to be small, and combined variations reflecting the total uncertainty are applied in this measurement.

Table 48 illustrates the experimental uncertainties associated with jet reconstruction and tagging. The uncertainties in the jet include jet energy scale and resolution calibration [64], uncertainties associated with the jet vertex tagger [111], and the uncertainties related to the efficiency of the b-taggers for the b-jet veto window [112]. Finally, uncertainties to the b-jet veto are considered as the scale factors applied to MC backgrounds to improve agreement between data and MC [89, 88, 112]. Table 49 shows the systematics associated with the E_T^{miss}. The uncertainties associated with the physics objects used to reconstruct the E_T^{miss} are included for the E_T^{miss} calculation, but only the uncertainties related to the unused tracks.

Jet Experimental Systematics			
Variation Name	Target	Comment	Nuisance Parameters
JET_BJES_Response	Jet Kinematics	Energy scale uncertainty on b-jets	1 NP (up & down variations)
JET_EffectiveNP_Detector	Jet Kinematics	Energy scale uncertainty split	2 NP (up & down variations)
JET_EffectiveNP_Mixed	Jet Kinematics	Energy scale uncertainty split	3 NP (up & down variations)
JET_EffectiveNP_Modelling	Jet Kinematics	Energy scale uncertainty split	4 NP (up & down variations)
JET_EffectiveNP_Statistical	Jet Kinematics	Energy scale uncertainty split	5 NP (up & down variations)
JET_EtaIntercalibration_Modeling	Jet Kinematics	Energy scale uncertainty on eta-intercalibration	1 NP (up & down variations)
JET_EtaIntercalibration_TotalStat	Jet Kinematics	Energy scale uncertainty on eta-intercalibration	1 NP (up & down variations)
JET_EtaIntercalibration_NonClosure_highE	Jet Kinematics	Energy scale uncertainty on eta-intercalibration	1 NP (up & down variations)
JET_EtaIntercalibration_NonClosure_negEta	Jet Kinematics	Energy scale uncertainty on eta-intercalibration	1 NP (up & down variations)
JET_EtaIntercalibration_NonClosure_posEta	Jet Kinematics	Energy scale uncertainty on eta-intercalibration	1 NP (up & down variations)
JET_Flavor_Composition	Jet Kinematics	Energy scale uncertainty on flavour composition	1 NP (up & down variations)
JET_Flavor_Response	Jet Kinematics	Energy scale uncertainty on samples' flavour response	1 NP (up & down variations)
JET_Pileup_OffsetMu	Jet Kinematics	Pile-up dependent energy scale uncertainty	1 NP (up & down variations)
JET_Pileup_OffsetNPV	Jet Kinematics	Energy scale uncertainty on pile-up (NPV dependent)	1 NP (up & down variations)
JET_Pileup_PtTerm	Jet Kinematics	Energy scale uncertainty on pile-up (p_T term)	1 NP (up & down variations)
JET_Pileup_RhoTopology	Jet Kinematics	Energy scale uncertainty on pile-up	1 NP (up & down variations)
JET_PunchThrough_MC16	Jet Kinematics	Energy scale uncertainty for punch-through jets	1 NP (up & down variations)
JET_SingleParticle_HighPt	Jet Kinematics	Energy scale uncertainty from high-p_T jets	1 NP (up & down variations)
JET_JER_DataVsMC_MC16	Jet Kinematics	Energy resolution uncertainty, each for both MC and pseudo-data	1 NP (up & down variations)
JET_JER_EffectiveNP	Jet Kinematics	Energy resolution uncertainty, each for both MC and pseudo-data	11 NP (up & down variations)
JET_JER_EffectiveNP_12restTerm	Jet Kinematics	Energy resolution uncertainty, each for both MC and pseudo-data	1 NP (up & down variations)
JET_JvtEfficiency	Scale Factor	JVT efficiency uncertainty	1 NP (up & down variations)
FT_EFF_Eigen_B	Scale Factor	b-tagging efficiency uncertainties	1 NP (up & down variations)
FT_EFF_Eigen_C	Scale Factor	for b jets, 3 for c jets and 4 for light jets	1 NP (up & down variations)
FT_EFF_Eigen_L	Scale Factor		1 NP (up & down variations)
FT_EFF_Eigen_extrapolation	Scale Factor	b-tagging efficiency uncertainty on the extrapolation to high-p_T jets	1 NP (up & down variations)
FT_EFF_Eigen_extrapolation_from_charm	Scale Factor	b-tagging efficiency uncertainty on tau jets	1 NP (up & down variations)

Table 48: Experimental uncertainties for jet energy/momentum calibration, reconstruction, identification and isolation efficiencies [64, 111, 112, 60, 89].

E_T^{miss} Experimental Systematics			
Variation Name	Target	Comment	Nuisance Parameters
MET_SoftTrk_ResoPara	E_T^{miss} Kinematics	uncertainity on the longitudinal resolution	1 NP (up & down variations)
MET_SoftTrk_ResoPrep	E_T^{miss} Kinematics	uncertainity on the transverse resolution	1 NP (up & down variations)
MET_SoftTrk_Scale	E_T^{miss} Kinematics	uncertainity on the longitudinal scale	1 NP (up & down variations)
MET_JetTrk_Scale	E_T^{miss} Kinematics	scale uncertainity due to tracks in jets	1 NP (up & down variations)

Table 49: Experimental uncertainties for electron energy/momentum calibration, reconstruction, identification and isolation efficiencies [67, 89].

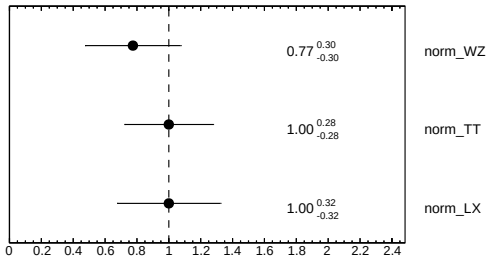
10.8 Asimov Measurement

This section will summarise the expected results of this experiment, corresponding to the fit results with the Asimov dataset. For the Asimov dataset, the value of the signal strength parameter is fixed to $\mu = 1$, along with the normalisation factor constraining the other polarisation fraction $\mu_{T\lambda} = 1$. For the expected fit, the WZ normalisation factor μ_{WZ} is still constrained with experimental data in the WZ control region. This ensures the $W^{\pm}Zjj$ QCD-induced background is correctly normalised in the signal region.

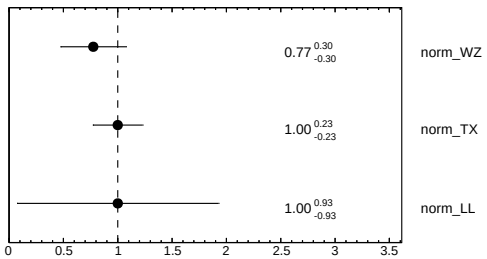
Before detailing the results, the section will first introduce some of the terminology used to describe the results of particle physics experiments. Pre-fit results refer to nominal values of the unconstrained normalisation and nuisance parameters. The nuisance parameters $\vec{\theta}$ are normalised to a central value of 0 and have a value of ± 1 corresponding to the uncertainty of $\pm\sigma$. Post-fit values results refer to the maximum likelihood estimators of the normalisation factors and the nuisance parameters.

10.8.1 Expected Significance

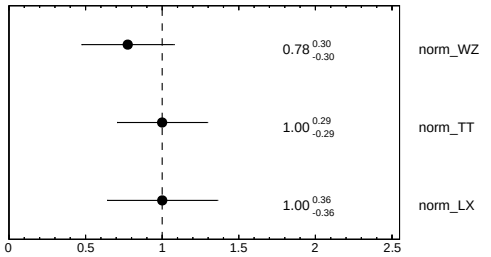
The expected significance, calculated with Asimov data, in the WW CoM frame, is $4.20(1.11)\sigma$ and in the pp CoM frame 3.59(0.68) for the LX(LL) measurement, respectively. The expected significance indicates sensitivity to show evidence of polarised $W^{\pm}W^{\pm}jj$ production with a longitudinally polarised gauge boson for the WW and pp CoM frame. The post-fit normalisation factors and uncertainties are shown in Fig. 134. The uncertainty on the LL measurement is very large due to limited statistics.



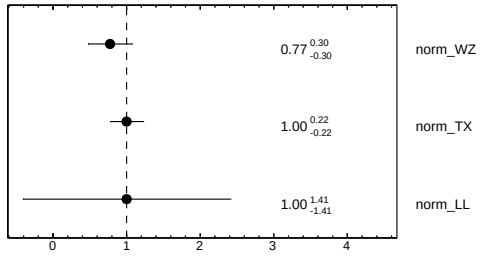
(a) Asimov LX fit in the WW CoM frame.



(b) Asimov LL fit in the WW CoM frame.



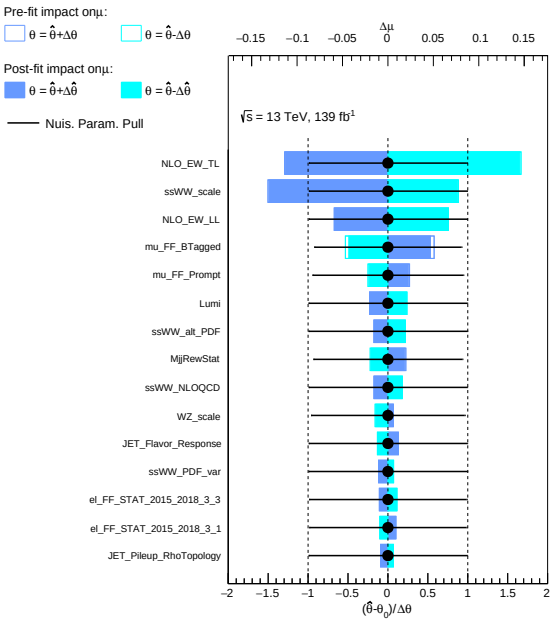
(c) Asimov LX fit in the pp CoM frame.



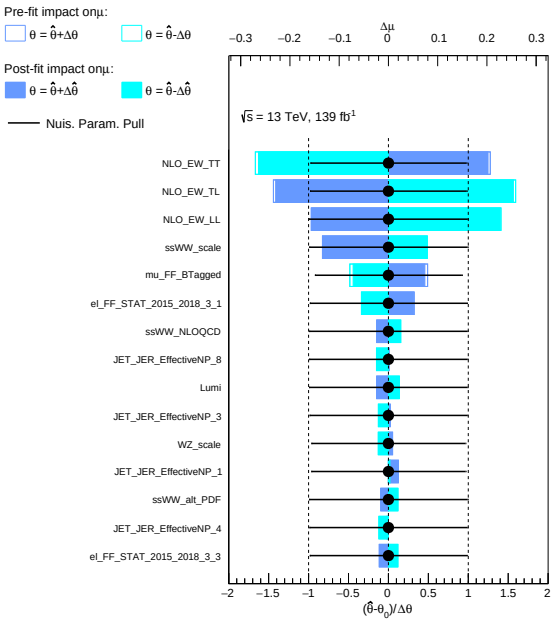
(d) Asimov LL fit in the pp CoM frame.

Figure 134: Values of the post-fit normalisation factors and their uncertainties, including the normalisation effects of the theoretical uncertainties.

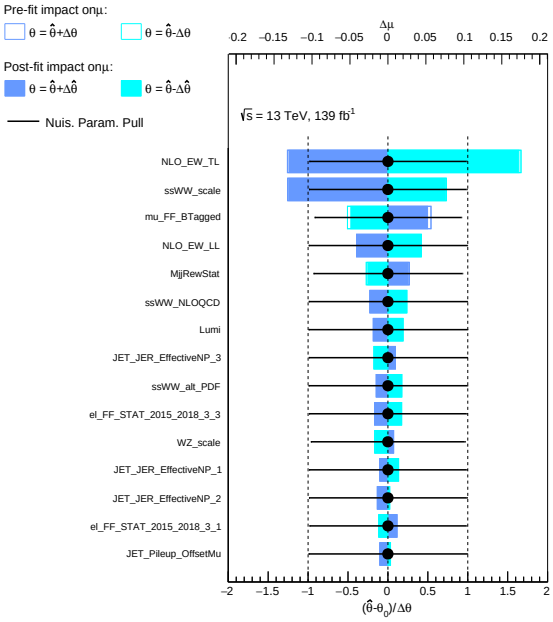
The impact ranking of the NP for each measurement is shown in Fig. 135 for the signal strength parameters and Fig. 136 for the normalisation factor on the traverse fraction. The ranking shows the theoretical uncertainties that most significantly impact the signal strength parameters. For all measurements, the dominant source arises from the NLO EW correction and scale uncertainty. For the LX measurement, only the NLO EW corrections on LL and LT have large impacts. This is because the LX events are well constrained from TT, so the normalisation effects of the TT background have minor effects on the measurement. The LL events are far less constrained, so the NLO TT and TL corrections have much larger impacts.



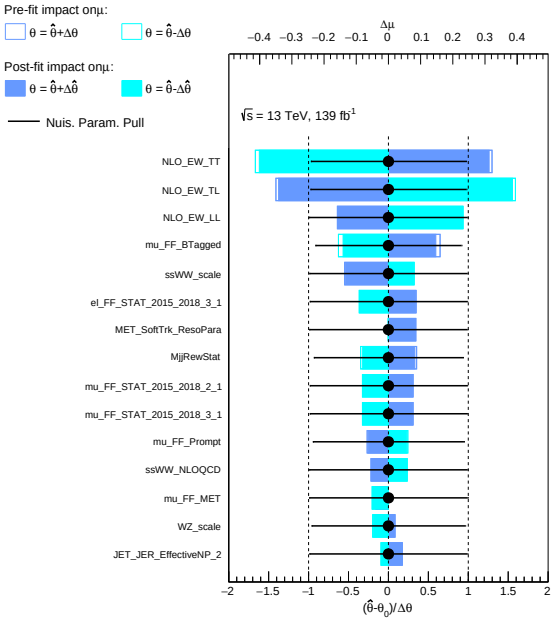
(a) μ_{LX} for the Asimov LX fit in the WW CoM frame.



(b) μ_{LL} for the Asimov fit LL in the WW CoM frame.

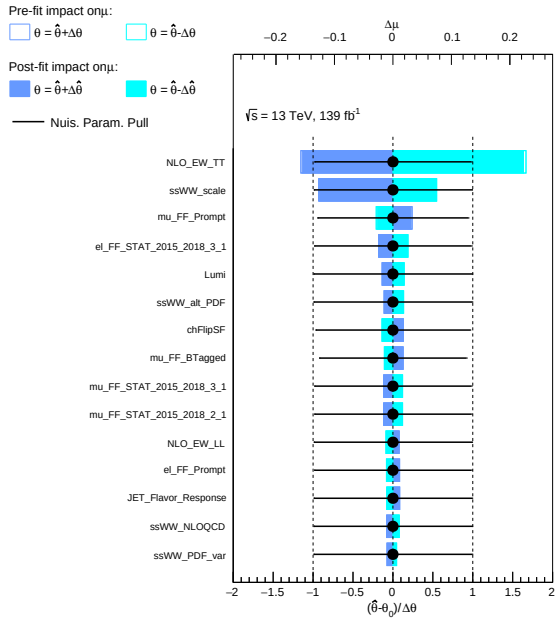


(c) μ_{LX} for the Asimov LX fit in the pp CoM frame.

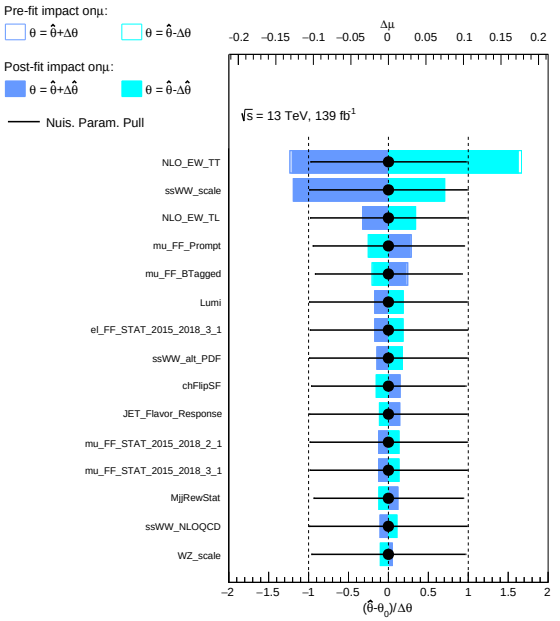


(d) μ_{LL} for the Asimov LL fit in the pp CoM frame.

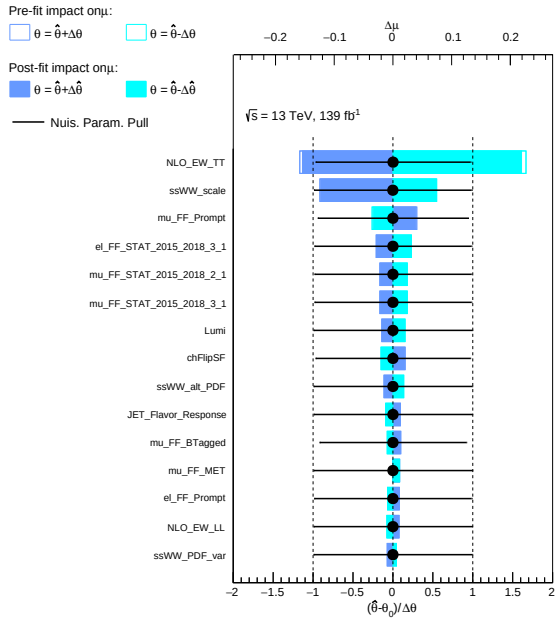
Figure 135: NP impacts on signal strength factor $\mu_{L\lambda}$, where λ = [X, L].



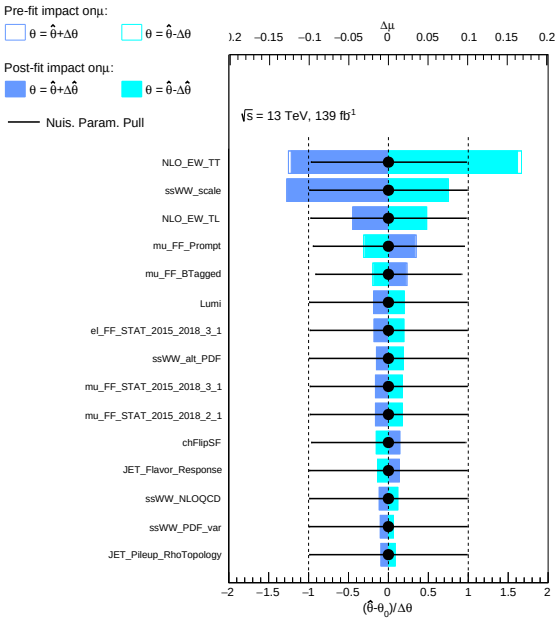
(a) μ_{TT} for the Asimov LX fit in the WW CoM frame.



(b) μ_{TX} for the Asimov fit LL in the WW CoM frame.



(c) μ_{TT} for the Asimov LX fit in the pp CoM frame.



(d) μ_{TX} for the Asimov LL fit in the pp CoM frame.

Figure 136: NP impacts on the unconstrained normalisation factor $\mu_{T\lambda}$, where λ = [X, T].

10.8.2 Expected Cross-section

The expected cross-section is extracted as described in Sec. 10.6. The fit in this section is performed with the normalised theory uncertainties. In the LX Asimov measurement, the best-fit value for the signal strength parameter in the WW CoM and *pp* CoM frame,

$$\mu^{\text{exp-meas}}_{LX(\text{WW})} = 1.000 \pm 0.268\ (\text{tot.}) = 1.000 \pm 0.256\ (\text{stat.}) \pm 0.081\ (\text{syst.}),$$
$$\mu^{\text{exp-meas}}_{LX(pp)} = 1.000 \pm 0.303\ (\text{tot.}) = 1.000 \pm 0.289\ (\text{stat.}) \pm 0.092\ (\text{syst.}),$$

where the uncertainties correspond to total, statistical and systematic uncertainties. A full breakdown of the uncertainties for the WW and *pp* measurement is provided in Table 51. The contribution from each uncertainty is calculated by fixing the associated NP to the best-fit value and re-performing the fit. The uncertainty is then subtracted from the total uncertainty on the cross-section in quadrature. This ensures the sum in quadrature of the grouped systematics and statistical uncertainty corresponds to the total uncertainty on the cross-section. The uncertainty is dominantly statistical, with the impact of the experimental and modelling systematics significantly smaller. The largest contribution from the systematics comes from the non-prompt background and the fake-factor estimation. This background is the largest in the signal region and is not constrained by either of the DNNs.

For the LL Asimov measurement, the best-fit value for the signal strength parameter in the WW and *pp* CoM frame is,

$$\mu^{\text{exp-meas}}_{LL(\text{WW})} = 1.000 \pm 0.844\ (\text{tot.}) = 1.000 \pm 0.837\ (\text{stat.}) \pm 0.116\ (\text{syst.}),$$
$$\mu^{\text{exp-meas}}_{LL(pp)} = 1.000 \pm 1.325\ (\text{tot.}) = 1.000 \pm 1.300\ (\text{stat.}) \pm 0.252\ (\text{syst.}).$$

The LL measurements are also dominated by statistical uncertainty due to the very low event yields. The statistical uncertainty is 130% on longitudinal scattering in the *pp* CoM frame and 83% in the WW CoM frame. Table 50 shows the predicted and expected measured fiducial cross-sections. The expected upper limit on the fiducial cross section for LL in the WW CoM at a 95% confidence level is 0.710 (fb) and in the *pp* CoM frame 0.648 (fb). Figure 137 and 138 shows the ranked impact of the nuisance parameters for the cross-section extraction.

	Fiducial cross-section (fb)	Breakdown (fb)	Predicted cross-section
$\sigma^{\text{exp-meas}}_{LX(\text{WW})}$	1.123 ± 0.301 (tot.)	1.123 ± 0.288 (stat.) ± 0.09 (syst.)	1.123 ± 0.07 (.tot)
$\sigma^{\text{exp-meas}}_{LL(\text{WW})}$	0.272 ± 0.230 (tot.)	0.272 ± 0.228 (stat.) ± 0.032 (syst.)	0.272 ± 0.02 (.tot)
$\sigma^{\text{exp-meas}}_{LX(\text{PP})}$	1.175 ± 0.356 (tot.)	1.175 ± 0.340 (stat.) ± 0.095 (syst.)	1.175 ± 0.08 (.tot)
$\sigma^{\text{exp-meas}}_{LL(\text{PP})}$	0.172 ± 0.228 (tot.)	0.172 ± 0.224 (stat.) ± 0.043 (syst.)	0.172 ± 0.01 (.tot)

Table 50: The expected fiducial cross-sections for LX and LL in the WW and *pp* CoM frame. The cross-sections are calculated by multiplying the signal strength parameters μ and the uncertainty $\Delta\mu$ with the predicted cross-section.

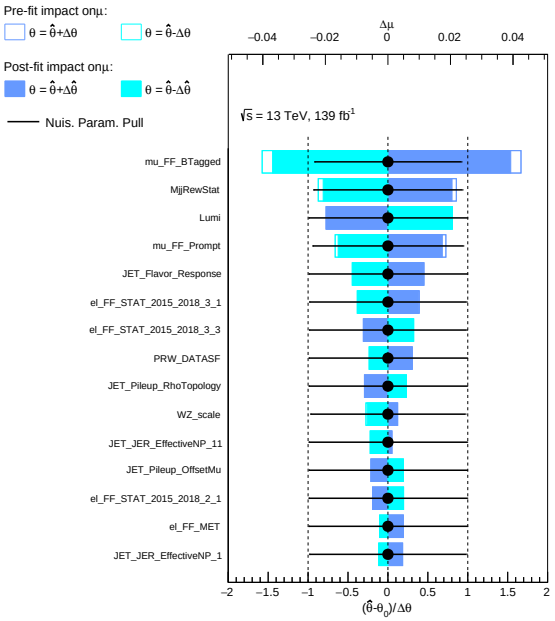
Source	Impact [%]
Experimental	
Electron calibration	0.3
Muon calibration	0.5
Jet energy scale and resolution	1.6
$E_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution	0.2
b-tagging inefficiency	0.2
Fake leptons	4.2
Charge misidentified leptons.	0.2
Pileup modelling	0.7
Modelling	
$W^{\pm}W^{\pm}jj$ NLO correction combination	0.09
$W^{\pm}W^{\pm}jj$ NLO EW	0.5
$W^{\pm}W^{\pm}jj$ NLO QCD	0.4
$W^{\pm}W^{\pm}jj$ α_s	0.01
$W^{\pm}W^{\pm}jj$ Alternate PDF	0.02
$W^{\pm}W^{\pm}jj$ PDF	0.02
$W^{\pm}W^{\pm}jj$ Scale	0.2
WZ scale, PDF, α_s	0.9
WZ reweighting	1.8
Model statistical	1.3
Experimental and modelling	5.6
Luminosity	2.0
Data statistical	26.2
Total	26.8

(a) LX WW CoM measurement

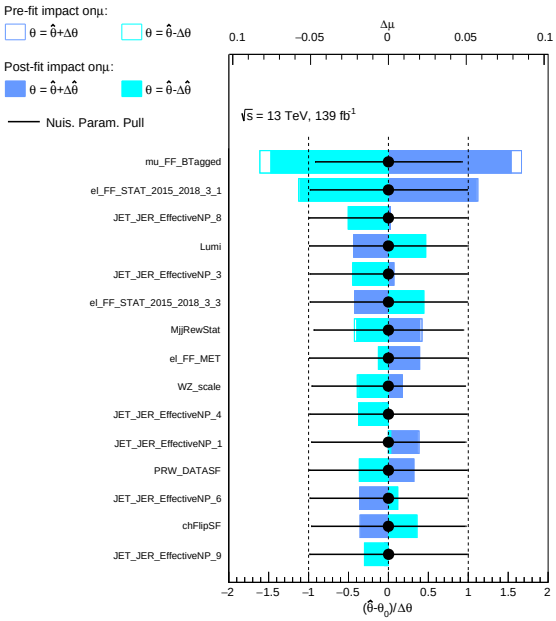
Source	Impact [%]
Experimental	
Electron calibration	0.4
Muon calibration	0.6
Jet energy scale and resolution	2.5
$E_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution	0.7
b-tagging inefficiency	0.2
Fake leptons	5.0
Charge misidentified leptons.	0.5
Pileup modelling	0.8
Modelling	
$W^{\pm}W^{\pm}jj$ NLO correction combination	0.2
$W^{\pm}W^{\pm}jj$ NLO EW	0.8
$W^{\pm}W^{\pm}jj$ NLO QCD	1.0
$W^{\pm}W^{\pm}jj$ α_s	0.02
$W^{\pm}W^{\pm}jj$ Alternate PDF	0.03
$W^{\pm}W^{\pm}jj$ PDF	0.04
$W^{\pm}W^{\pm}jj$ Scale	0.4
WZ scale, PDF, α_s	0.7
WZ reweighting	2.7
Model statistical	1.3
Experimental and modelling	6.6
Luminosity	1.9
Data statistical	29.5
Total	30.3

(b) LX pp CoM measurement

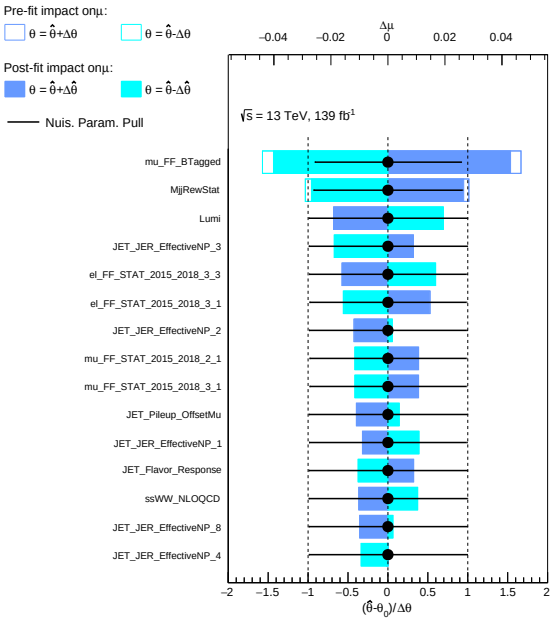
Table 51: Breakdown of the uncertainty on the $W^{\pm}W^{\pm}jj$ LX cross-section measurement in the WW and pp CoM frame using Asimov data.



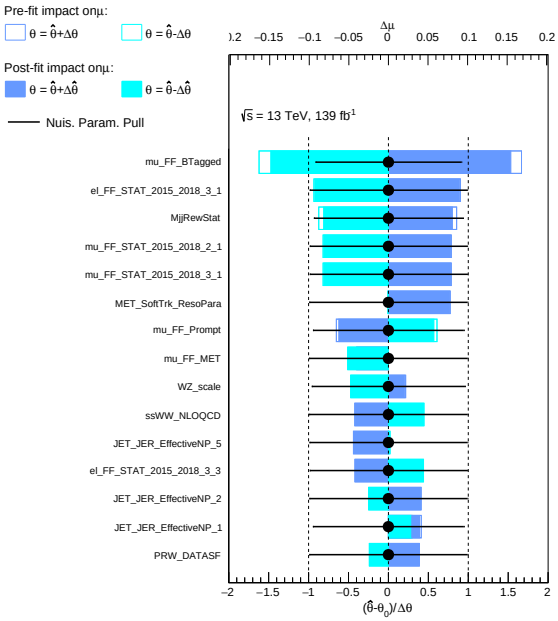
(a) μ_{LX} for the Asimov LX fit in the WW CoM frame.



(b) μ_{LL} for the Asimov fit LL in the WW CoM frame.

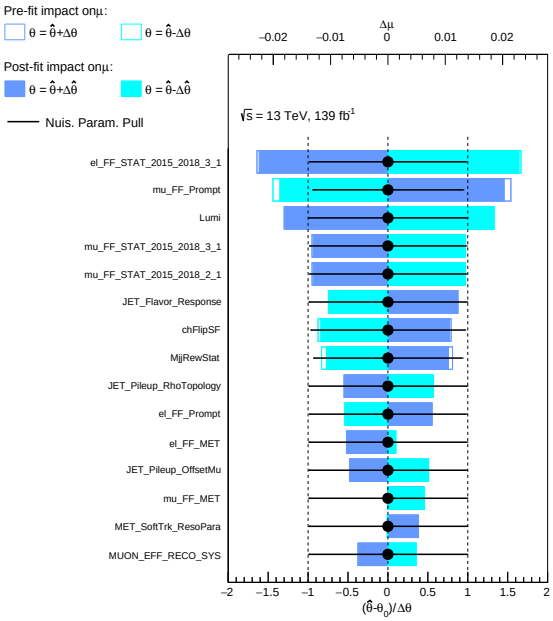


(c) μ_{LX} for the Asimov LX fit in the pp CoM frame.

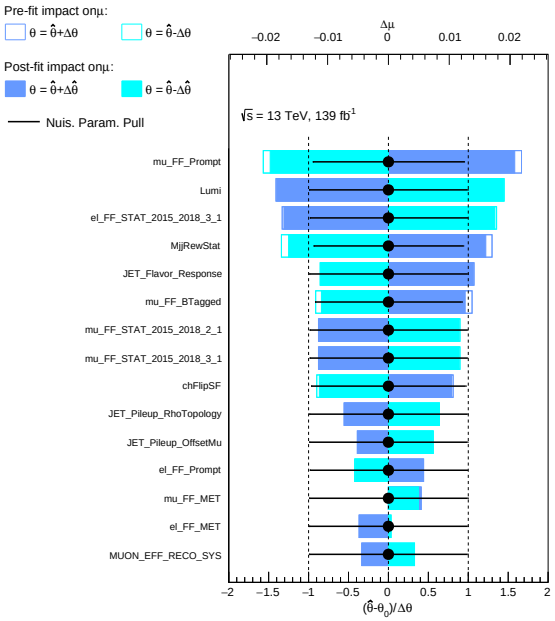


(d) μ_{LL} for the Asimov LL fit in the pp CoM frame.

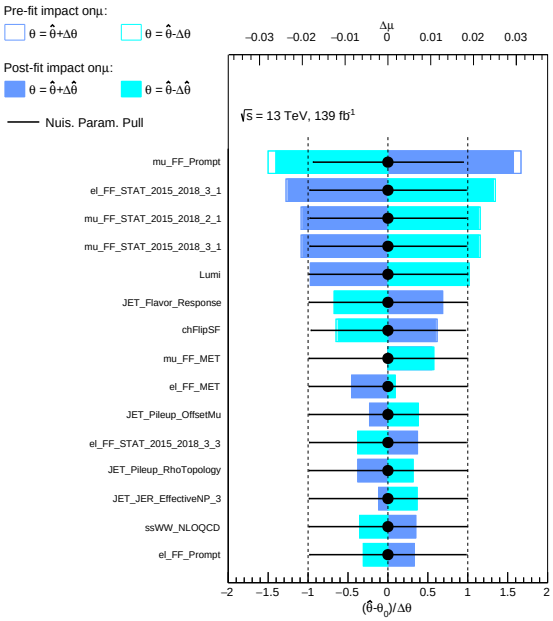
Figure 137: NP impacts on signal strength factor $\mu_{L\lambda}$ for the cross-section extraction, where $\lambda = [\mathrm{X}, \mathrm{L}]$.



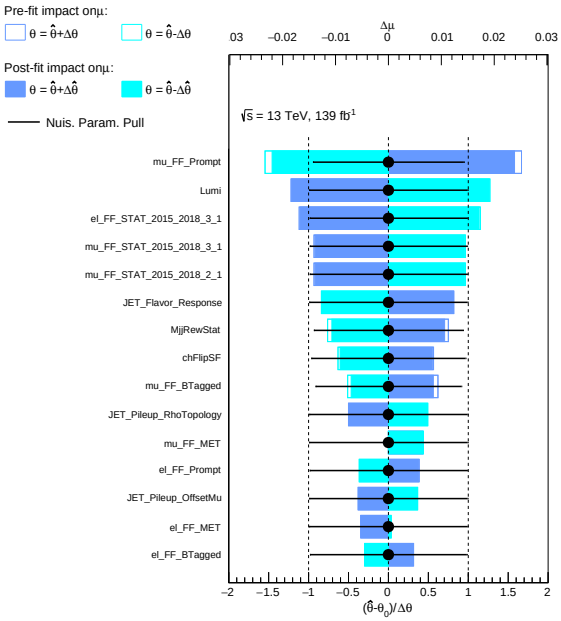
(a) μ_{TT} for the Asimov LX fit in the WW CoM frame.



(b) μ_{TX} for the Asimov fit LL in the WW CoM frame.



(c) μ_{TT} for the Asimov LX fit in the pp CoM frame.



(d) μ_{TX} for the Asimov LL fit in the pp CoM frame.

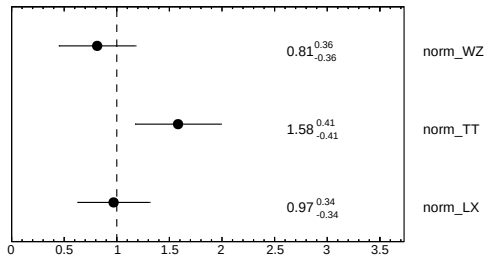
Figure 138: NP impacts on the unconstrained normalisation factor $\mu_{T\lambda}$ for the cross-section extraction, where $\lambda = [\mathrm{X}, \mathrm{T}]$.

10.9 Observed Measurement

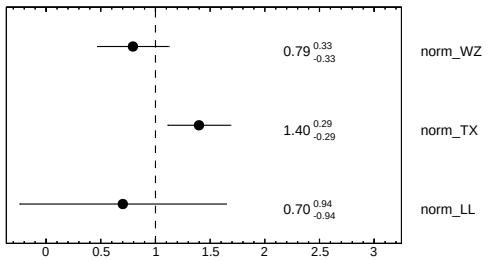
10.9.1 Observed Significance

This section will summarise the observed results of this experiment, corresponding to the fit results with the measured data. The post-fit unconstrained normalisation factors are shown in Fig. 139. The pre-fit summary plots show an excess in data, which is constrained by the fit to the TT polarisation fraction. The pre- and post-fit distributions for the LX and LL measurements in the WW CoM frame are shown in Figs. 140 and 141, and Figs. 142 and 143 respectively. The same plots for the *pp* CoM frame are shown in Figs. 144 and 145, and Figs. 146 and 147. In the WW CoM frame, for the LX measurement, the most significant excess in data is seen to be in the region where $0.6 < \text{DNNssWW} < 1.0$ for bins with $0.7 < \text{DNNPol}_{\text{TT}} < 1.0$. These bins are dominated by the TT signal, and as a result, the unconstrained normalisation factor associated with the TT fraction is constrained to the excess.

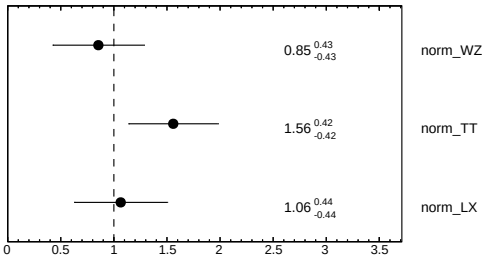
The NP impact ranking is shown in Figs. 148 and 149 for the signal strength parameters and the unconstrained normalisation factor. The ranking is consistent with the ranking seen for the Asimov dataset. The figures show slightly larger NP pulls in the *pp* CoM frame, which suggests slightly larger discrepancies in the data. There are no significantly large pulls except for the LX measurement in the *pp* CoM frame shown in Fig. 148c. The observed significance for the LX(LL) measurements in the WW and *pp* CoM frame are $3.73(0.74)\sigma$ and $3.13(0.17)\sigma$ respectively.



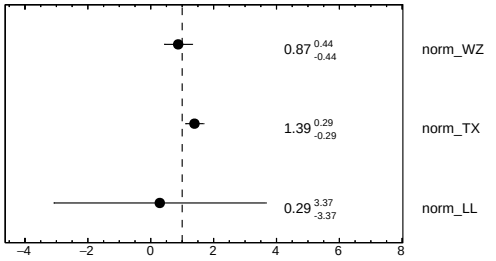
(a) Observed LX fit in the WW CoM frame.



(b) Observed LL fit in the WW CoM frame.



(c) Observed LX fit in the *pp* CoM frame.



(d) Observed LL fit in the *pp* CoM frame.

Figure 139: Values of the post-fit normalisation factors and their uncertainties, including the normalisation effects of the theoretical uncertainties for the observed measurement.

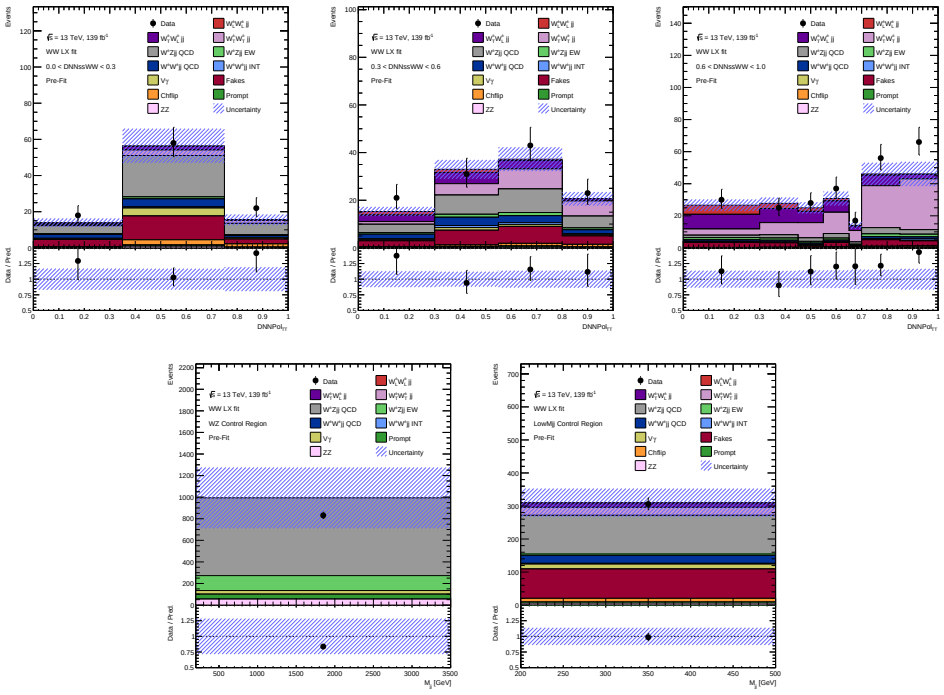


Figure 140: Pre-fit distributions for the observed LX measurement in the WW CoM frame. The uncertainties include the full theoretical uncertainties, including normalisation.

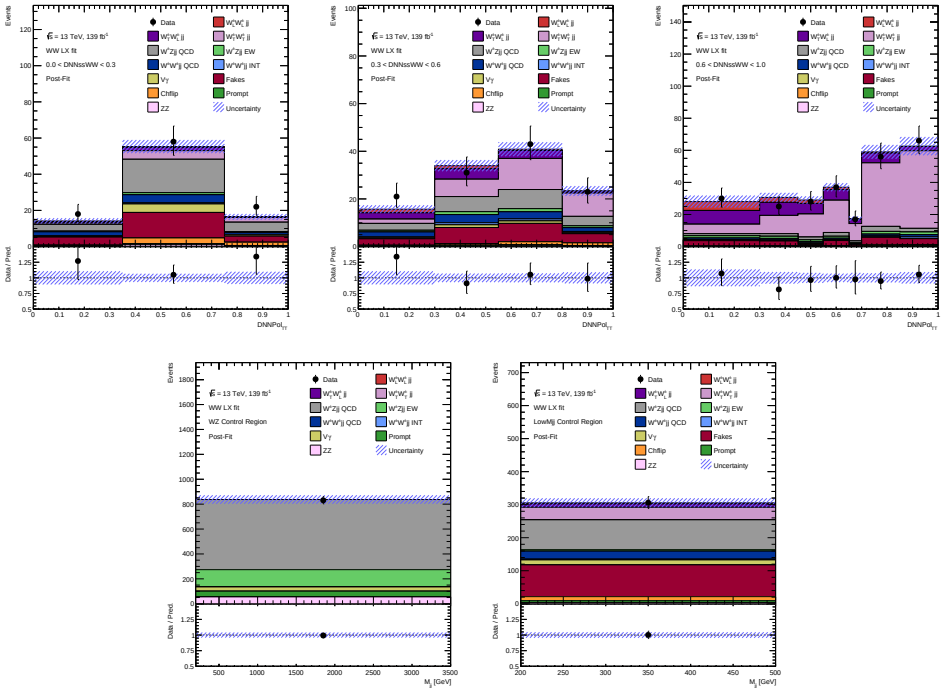


Figure 141: Post-fit distributions for the observed LX measurement in the WW CoM frame. The uncertainties include the full theoretical uncertainties, including normalisation.

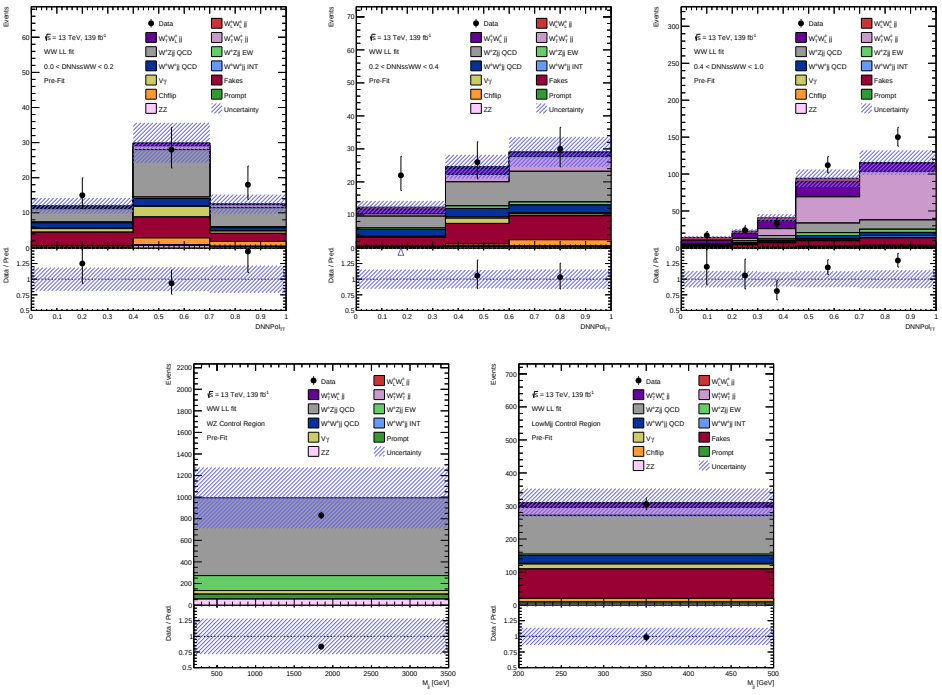


Figure 142: Pre-fit distributions for the observed LL measurement in the WW CoM frame. The uncertainties include the full theoretical uncertainties, including normalisation.

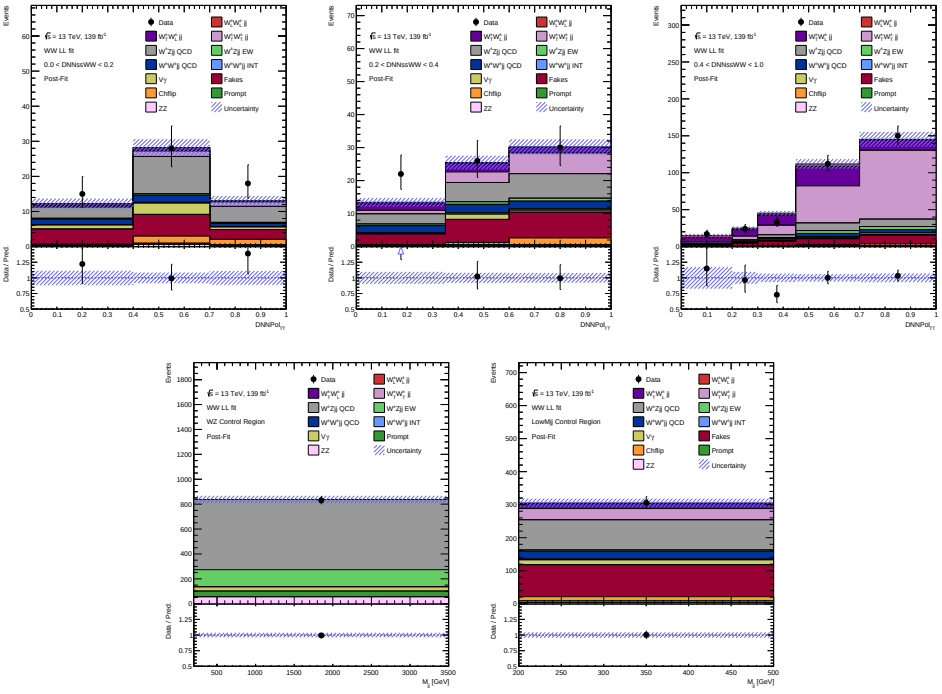


Figure 143: Post-fit distributions for the observed LL measurement in the WW CoM frame. The uncertainties include the full theoretical uncertainties, including normalisation.

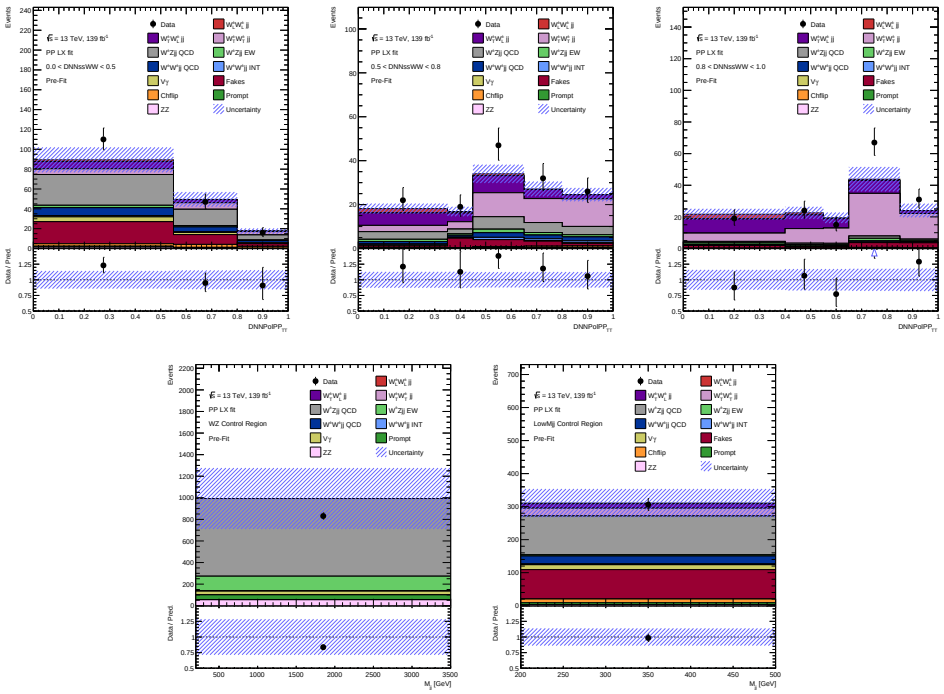


Figure 144: Pre-fit distributions for the observed LX measurement in the *pp* CoM frame. The uncertainties include the full theoretical uncertainties, including normalisation.

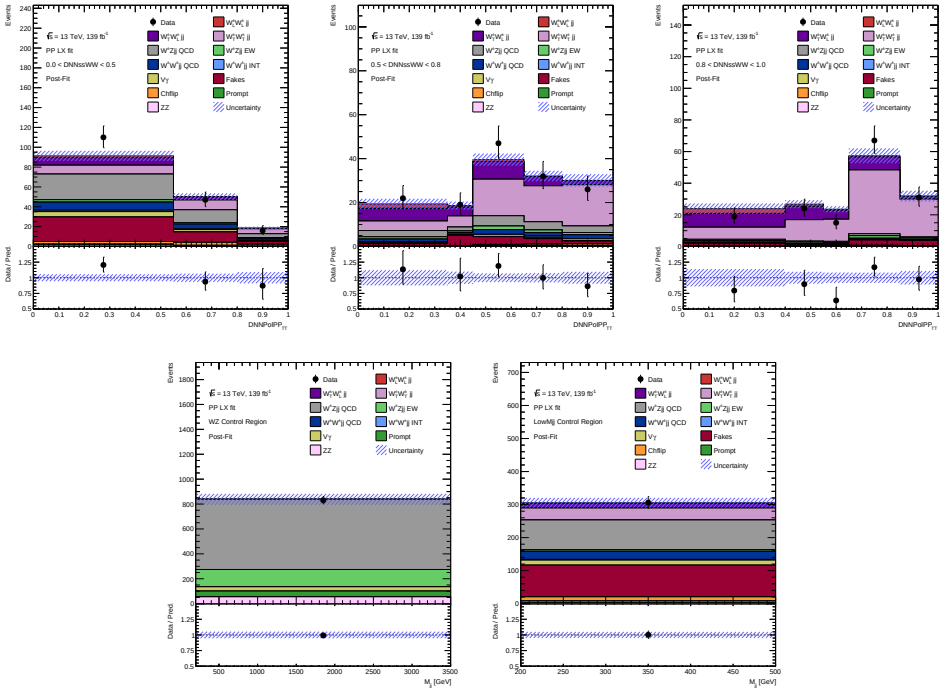


Figure 145: Post-fit distributions for the observed LX measurement in the *pp* CoM frame. The uncertainties include the full theoretical uncertainties, including normalisation.

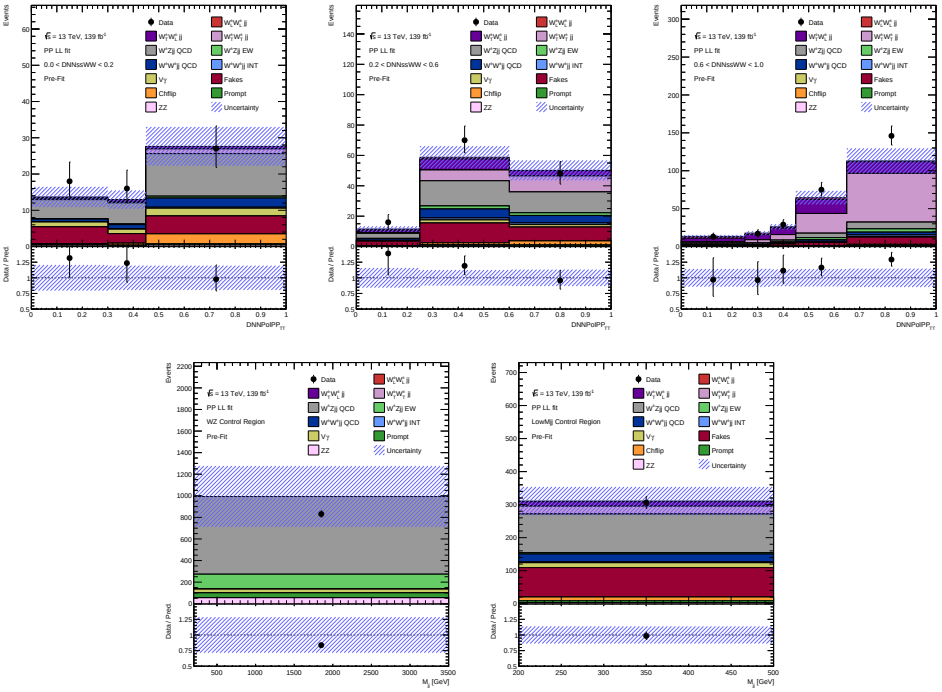


Figure 146: Pre-fit distributions for the observed LL measurement in the *pp* CoM frame. The uncertainties include the full theoretical uncertainties, including normalisation.

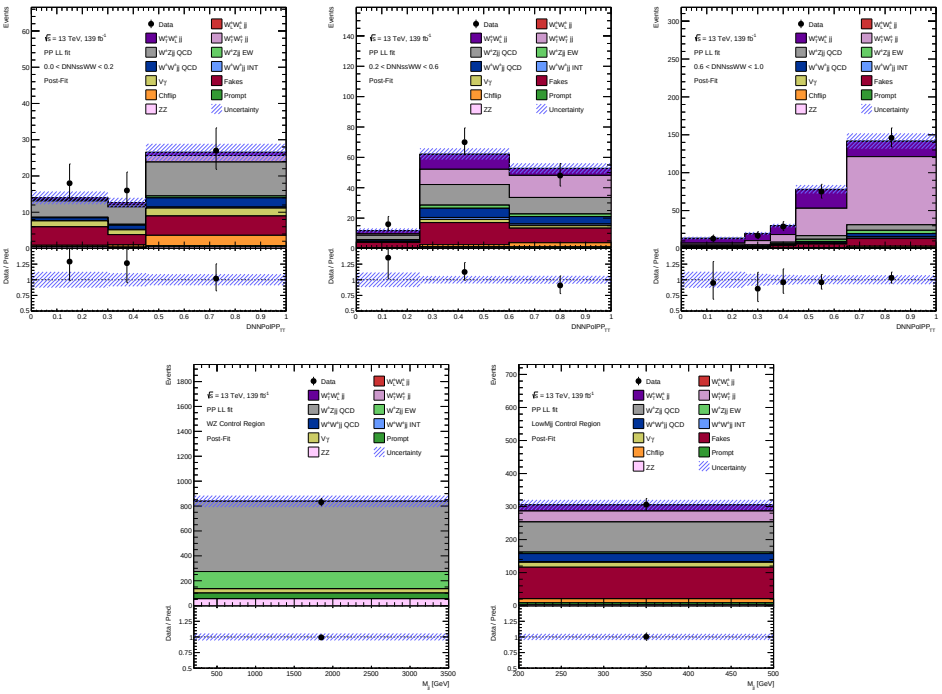
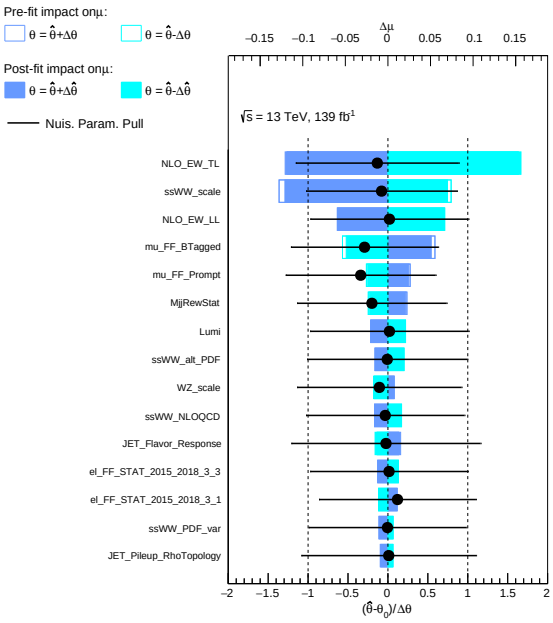
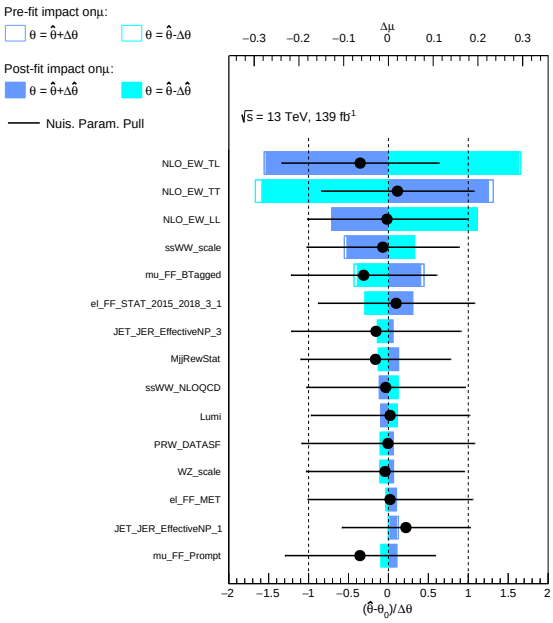


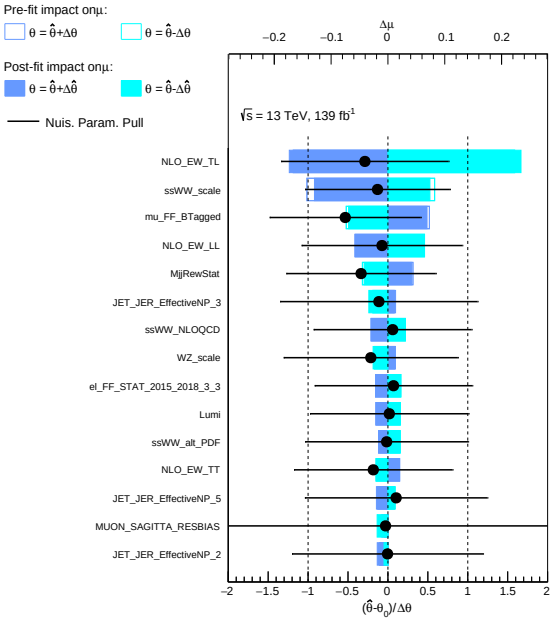
Figure 147: Post-fit distributions for the observed LL measurement in the *pp* CoM frame. The uncertainties include the full theoretical uncertainties, including normalisation.



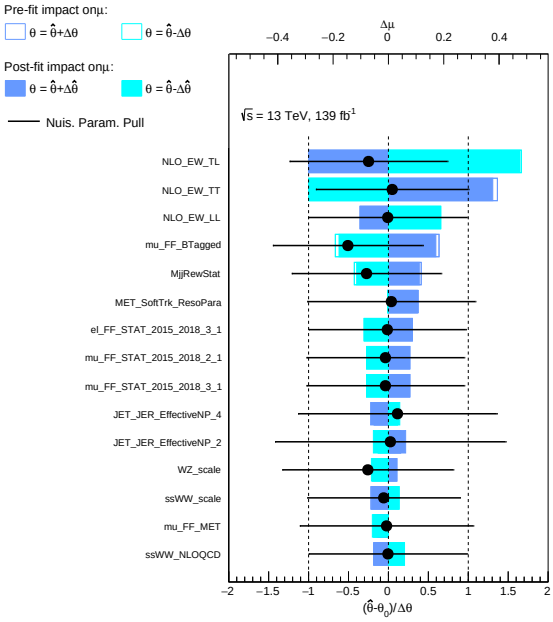
(a) μ_{LX} for the observed LX fit in the WW CoM frame.



(b) μ_{LL} for the observed fit LL in the WW CoM frame.

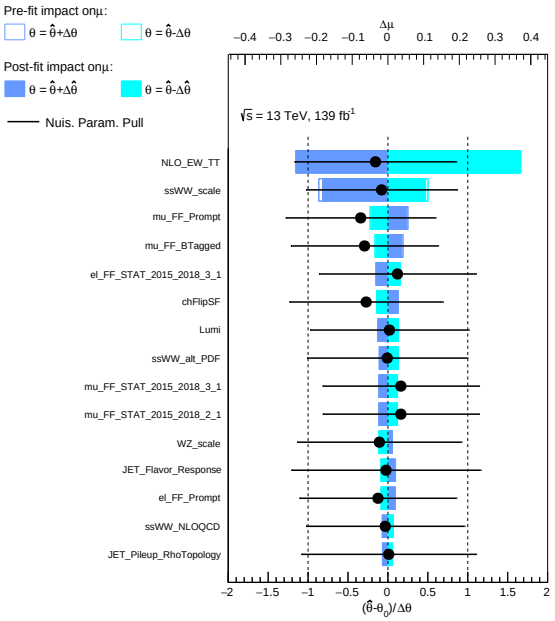


(c) μ_{LX} for the observed LX fit in the *pp* CoM frame.

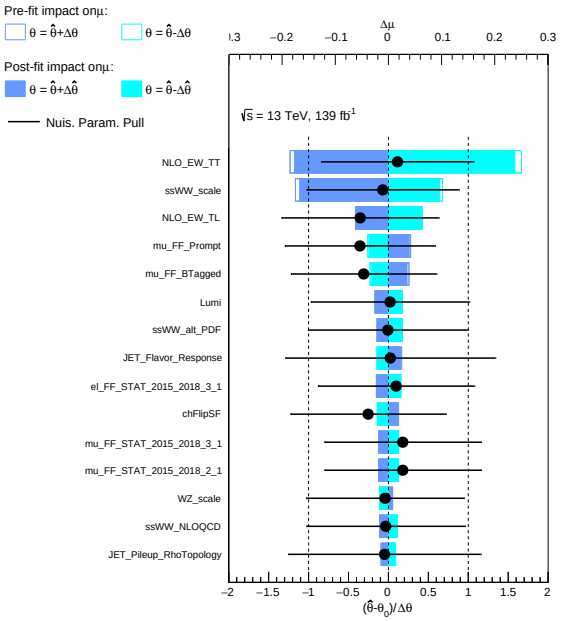


(d) μ_{LL} for the observed LL fit in the *pp* CoM frame.

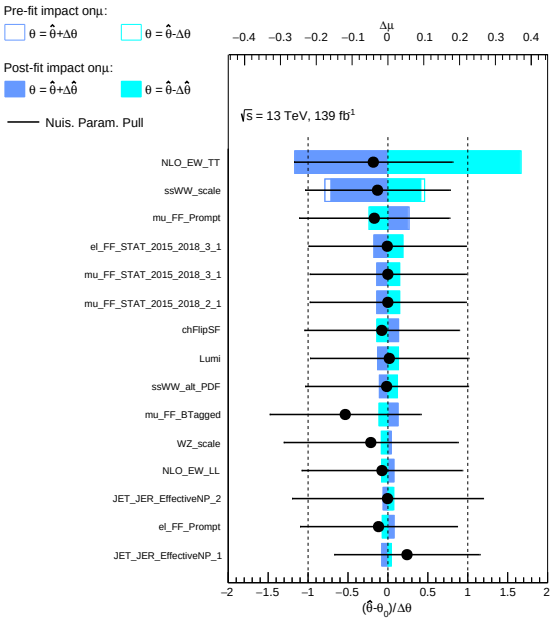
Figure 148: NP impacts on signal strength factor $\mu_{L\lambda}$, where λ = [X, L].



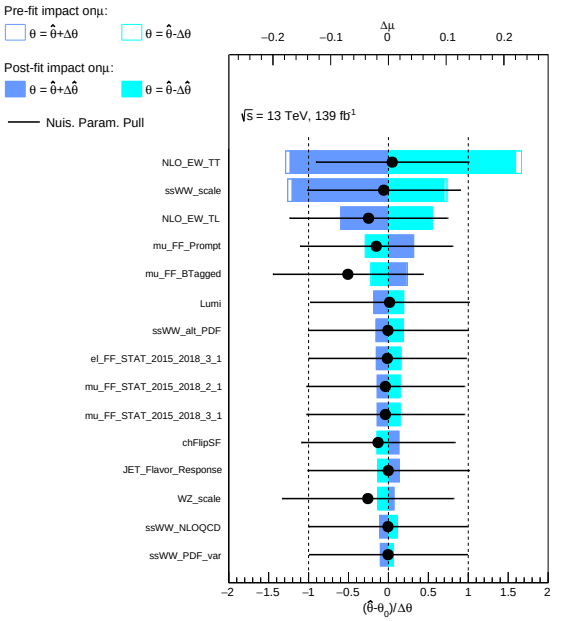
(a) μ_{TT} for the observed LX fit in the WW CoM frame.



(b) μ_{TX} for the observed fit LL in the WW CoM frame.



(c) μ_{TT} for the observed LX fit in the pp CoM frame.



(d) μ_{TX} for the observed LL fit in the pp CoM frame.

Figure 149: NP impacts on the unconstrained normalisation factor $\mu_{T\lambda}$, where λ = [X, T].

10.9.2 Observed Cross-Section

The observed cross-section is extracted as described in Sec. 10.6. In the LX Asimov measurement, the best-fit value for the signal strength parameter in the WW CoM is,

$$\mu^{\text{obv-meas}}_{LX(\text{WW})} = 0.943 \pm 0.283 \text{ (tot.)} = 0.943 \pm 0.268 \text{ (stat.)} \pm 0.090 \text{ (syst.)},$$

and in the *pp* CoM frame,

$$\mu^{\text{obv-meas}}_{LX(pp)} = 0.992 \pm 0.341 \text{ (tot.)} = 0.992 \pm 0.320 \text{ (stat.)} \pm 0.117 \text{ (syst.)},$$

where the uncertainties correspond to total, statistical and systematic uncertainties. A full breakdown of the uncertainties for the WW and *pp* LX measurements is provided in Tab. 52.

Source	Impact [%]
Experimental	
Electron calibration	0.4
Muon calibration	0.5
Jet energy scale and resolution	2.4
$E_{\text{T}}^{\text{miss}}$ scale and resolution	0.2
b-tagging inefficiency	0.2
Fake leptons	4.4
Charge misidentified leptons.	0.06
Pileup modelling	0.8
Modelling	
$W^{\pm}W^{\pm}jj$ NLO correction combination	0.1
$W^{\pm}W^{\pm}jj$ NLO EW	0.6
$W^{\pm}W^{\pm}jj$ NLO QCD	0.4
$W^{\pm}W^{\pm}jj$ α_s	0.02
$W^{\pm}W^{\pm}jj$ Alternate PDF	0.02
$W^{\pm}W^{\pm}jj$ PDF	0.02
$W^{\pm}W^{\pm}jj$ Scale	0.3
WZ scale, PDF, α_s	0.8
WZ reweighting	2.0
Model statistical	6.8
Experimental and modelling	9.0
Luminosity	1.9
Data statistical	26.8
Total	26.8

(a) LX WW CoM measurement

Source	Impact [%]
Experimental	
Electron calibration	0.5
Muon calibration	1.5
Jet energy scale and resolution	4.0
$E_{\text{T}}^{\text{miss}}$ scale and resolution	1.1
b-tagging inefficiency	0.3
Fake leptons	5.5
Charge misidentified leptons.	0.3
Pileup modelling	0.8
Modelling	
$W^{\pm}W^{\pm}jj$ NLO correction combination	0.3
$W^{\pm}W^{\pm}jj$ NLO EW	1.0
$W^{\pm}W^{\pm}jj$ NLO QCD	1.4
$W^{\pm}W^{\pm}jj$ α_s	0.02
$W^{\pm}W^{\pm}jj$ Alternate PDF	0.04
$W^{\pm}W^{\pm}jj$ PDF	0.05
$W^{\pm}W^{\pm}jj$ Scale	0.4
WZ scale, PDF, α_s	1.2
WZ reweighting	3.2
Model statistical	8.1
Experimental and modelling	11.7
Luminosity	1.9
Data statistical	32.0
Total	30.3

(b) LX *pp* CoM measurement

Table 52: Breakdown of the uncertainty on the $W^{\pm}W^{\pm}jj$ LX cross-section measurement in the WW and *pp* CoM frame using observed data.

The breakdown of uncertainties shows no significant increase in the uncertainty associated with the systematics, and the results are consistent with the Asimov fit shown in Tab. 51. However, the systematic uncertainties appear to be slightly larger in the *pp* CoM frame. This is because the excess in data is less constrained by the μ_{TT} parameter in the *pp* CoM frame. The networks in the WW CoM frame are better at separating the TT polarisation fraction, and the fit attributes more of the discrepancy with data to the unconstrained normalisation factor associated with TT.

For the LL observed measurement, the best-fit value for the signal strength parameter in the WW CoM is,

$$\mu^{\text{obv-meas}}_{LL(\text{WW})} = 0.550 \pm 0.838 \text{ (tot.)} = 0.550 \pm 0.812 \text{ (stat.)} \pm 0.207 \text{ (syst.)},$$

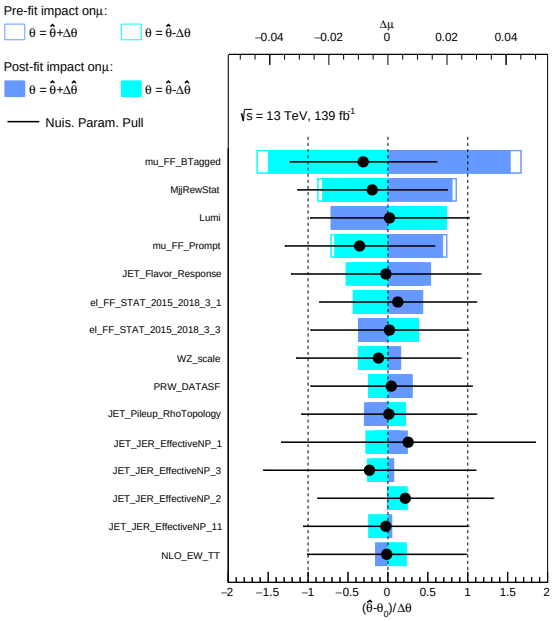
and in the *pp* CoM frame,

$$\mu^{\text{obv-meas}}_{LL(pp)} = 0.138 \pm 4.130 \text{ (tot.)} = 0.138 \pm 3.918 \text{ (stat.)} \pm 1.305 \text{ (syst.)}.$$

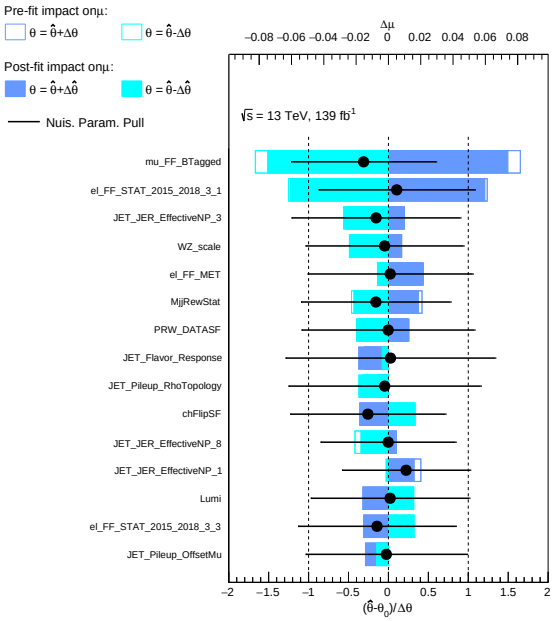
Table 53 shows the predicted and observed fiducial cross-sections. The observed upper limit on the fiducial cross section for LL in the WW CoM at a 95% confidence level is 0.617 (fb) and in the *pp* CoM frame 0.569 (fb). Figures 150 and 151 show the ranked impact of the nuisance parameters for the cross-section extraction. The ranking is consistent with the expected measurement.

	Fiducial cross-section (fb)	Breakdown (fb)	Predicted cross-section
$\sigma^{\text{obv-meas}}_{LX(\text{WW})}$	1.056 ± 0.318 (tot.)	1.123 ± 0.301 (stat.) ± 0.101 (syst.)	1.123 ± 0.07 (.tot)
$\sigma^{\text{obv-meas}}_{LL(\text{WW})}$	0.150 ± 0.228 (tot.)	0.150 ± 0.221 (stat.) ± 0.056 (syst.)	0.272 ± 0.02 (.tot)
$\sigma^{\text{obv-meas}}_{LX(\text{PP})}$	1.163 ± 0.401 (tot.)	1.163 ± 0.376 (stat.) ± 0.137 (syst.)	1.175 ± 0.08 (.tot)
$\sigma^{\text{obv-meas}}_{LL(\text{PP})}$	0.024 ± 0.710 (tot.)	0.024 ± 0.674 (stat.) ± 0.224 (syst.)	0.172 ± 0.01 (.tot)

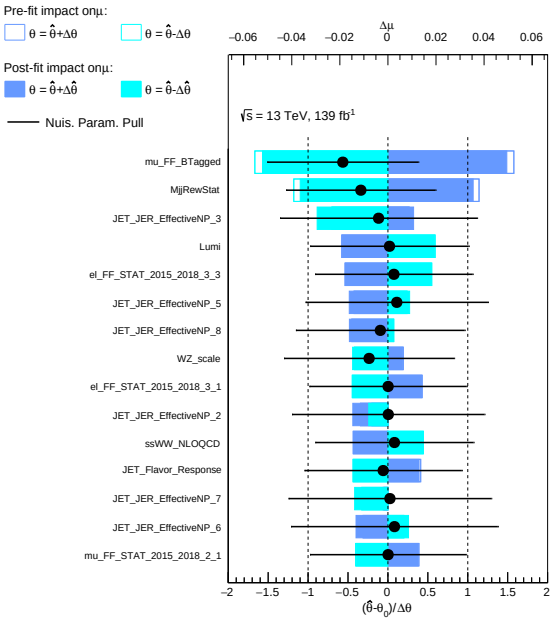
Table 53: The observed fiducial cross-sections for LX and LL in the WW and *pp* CoM frame. The cross-sections are calculated by multiplying the signal strength parameters μ and the uncertainty $\Delta\mu$ with the predicted cross-section.



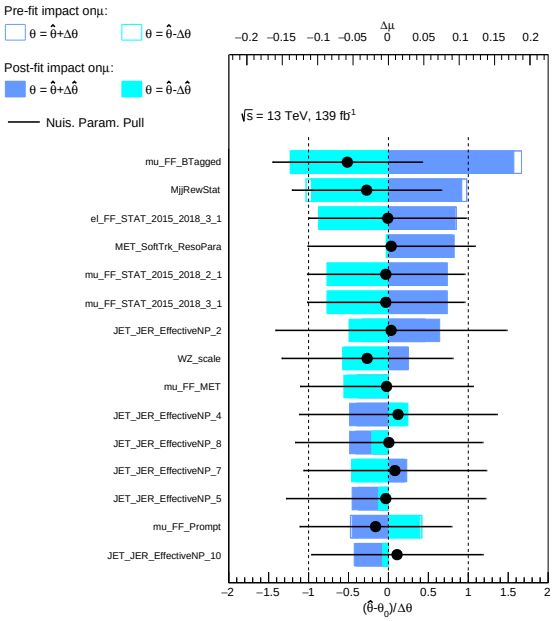
(a) μ_{LX} for the observed LX fit in the WW CoM frame.



(b) μ_{LL} for the observed fit LL in the WW CoM frame.

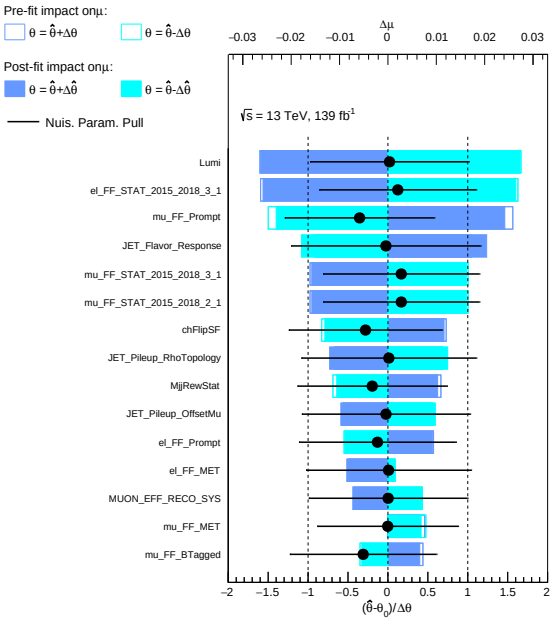


(c) μ_{LX} for the observed LX fit in the *pp* CoM frame.

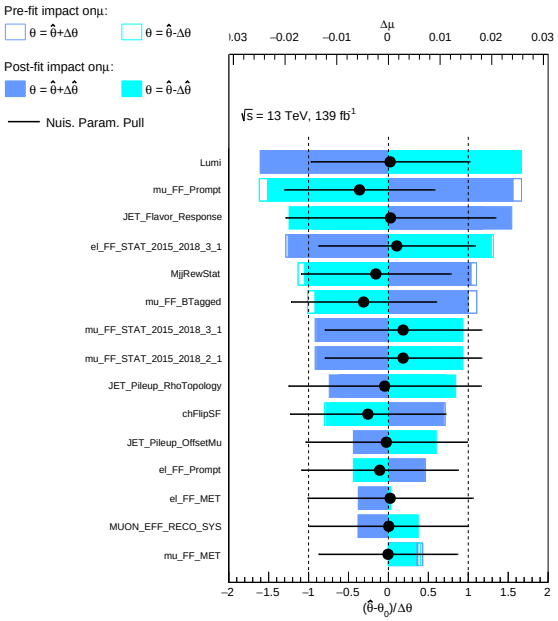


(d) μ_{LL} for the observed LL fit in the *pp* CoM frame.

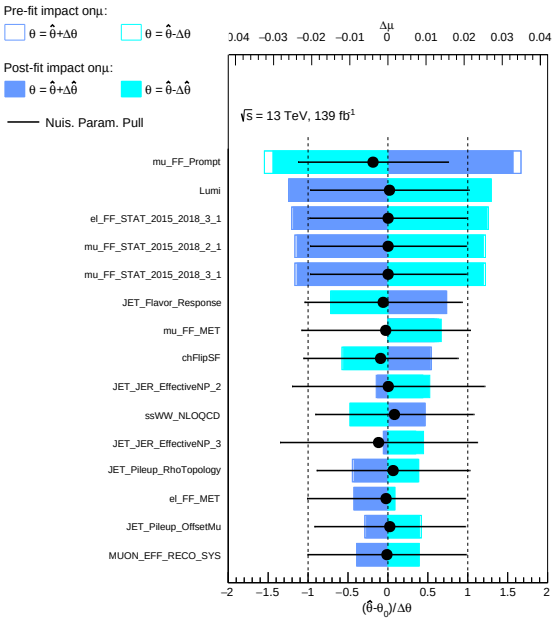
Figure 150: NP impacts on signal strength factor $\mu_{L\lambda}$ for the cross-section extraction, where λ = [X, L].



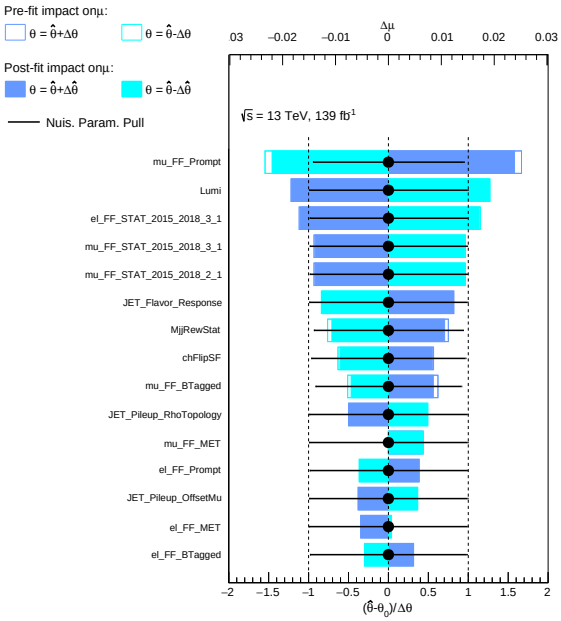
(a) μ_{TT} for the observed LX fit in the WW CoM frame.



(b) μ_{TX} for the observed fit LL in the WW CoM frame.



(c) μ_{TT} for the observed LX fit in the pp CoM frame.



(d) μ_{TX} for the Asimov LL fit in the pp CoM frame.

Figure 151: NP impacts on the unconstrained normalisation factor $\mu_{T\lambda}$ for the cross-section extraction, where $\lambda = [\mathrm{X}, \mathrm{T}]$.

10.10 Analysis of results

The results of this experiment measure an observed (expected) significance for the LL and LX measurement in the WW CoM frame of 3.73 (4.20) and 0.74 (1.11) standard deviations, respectively. The measurement in *pp* Com frame found the observed (expected) significance for the LL and LX measurements to be 3.13 (3.59) and 0.17(0.68) standard deviations, respectively. A comparison of the expected, observed, and predicted fiducial cross-section for each measurement is shown in Tab. 54. The experiment found the observed (expected) cross-section upper bounds at a 95% confidence level for LL EW production, in the WW and *pp* CoM frame, to be 0.617 (0.710) fb and 0.569 (0.648) fb. The results suggest evidence of EW-induced $W^{\pm}W^{\pm}jj$ production with a longitudinally polarised gauge boson.

	Observed cross-section (fb)	Expected cross-section (fb)	Predicted cross-section
$\sigma^{\text{obv-meas}}_{LX(\text{WW})}$	1.056 ± 0.318	1.123 ± 0.301	1.123 ± 0.07
$\sigma^{\text{obv-meas}}_{LL(\text{WW})}$	0.150 ± 0.228	0.272 ± 0.230	0.272 ± 0.02
$\sigma^{\text{obv-meas}}_{LX(\text{PP})}$	1.163 ± 0.401	1.175 ± 0.356	1.175 ± 0.08
$\sigma^{\text{obv-meas}}_{LL(\text{PP})}$	0.024 ± 0.710	0.172 ± 0.228	0.172 ± 0.01

Table 54: The observed, expected and predicted fiducial cross-sections for LX and LL in the WW and *pp* CoM frame.

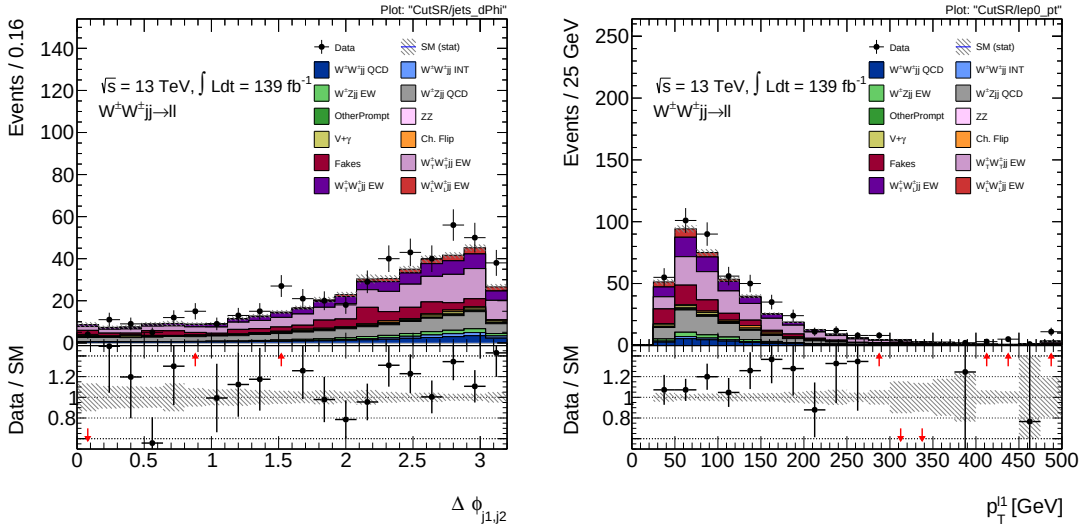


Figure 152: The $\Delta\phi_{jj}$ and lep$_0$-p_T distributions in the signal region compared to data.

However, there are also some apparent discrepancies in the modelling. There is a normalisation effect; the total predicted event yields are smaller than the data. The fit results also suggest an excess in TT events observed in the data larger than one standard

deviation. The origin of this mismodelling is not immediately clear. The observed excess is more significant than that seen in the differential measurement [89, 88]. The biggest difference in the experiments is in the application of the NLO corrections. The differential analysis is performed with the inclusive LO signal samples, and the NLO corrections are implemented as systematic uncertainties. This measurement applies the NLO corrections to the polarisation samples and applies a 100% uncertainty on the correction. The NLO EW correction reduces significantly the predicted event yields of the signal samples, so the mismodelling is larger in this experiemnt.

Both classifiers identify the excess in data to largely originate from TT events, as seen in Fig. 140, and also observed in Figs. 81 and 101. To illustrate why, Fig. 152 shows the $\Delta\phi_{jj}$ and lep$_0$-p_T distributions in the signal region compared to data. Both of these variables are highly ranked for the electroweak and polarisation classifiers. The figure shows the an excess in data in the high-$\Delta\phi_{jj}$ region ($\Delta\phi_{jj} > 2.5$) and in the high-p_T region (lep$_0$-$p_T > 75$ GeV). These are kinematic regions typically associated with the EW-induced $W^{\pm}W^{\pm}jj$ signal, which may be one of the reasons why the classifiers are predicting parts of the excess as TT events. However, the multivariate analysis makes drawing any evidential conclusions beyond intuition difficult.

There are a number of ways this experiment could be improved. The measurements are dominated by statistical uncertainty and are very large, particularly for the purely longitudinal measurement. Foremost, the most considerable improvements to polarised VBS measurements will come with larger datasets for analysis. The results also show the necessity for polarised NLO predictions. It is clear from the results there is a large discrepancy with the NLO EW, which can be resolved with polarised EW predictions at NLO. The experiment could also potentially be repeated with a smaller subset of variables used in the multivariate analysis. The accuracy of the classifiers would slightly decrease but may make analysing the origin of the mismodelling easier.

Crucially, however, the results do indicate the first evidence of EW-induced $W^{\pm}W^{\pm}jj$ production with longitudinally polarised $W^{\pm}$ bosons in both the WW and pp CoM frame. Moreover, the experiment sets important limits on the cross-section for purely longitudinal production. The sensitivity achieved in this experiment to polarisation was only possible as a result of the deep neural network techniques. Without the MVAs, the ATLAS polarisation measurement found the expected significance of the LX measurement in the WW CoM to be approximately 2.1σ, significantly below the threshold of evidence. The MVA techniques implemented in this thesis doubled the experimental sensitivity to polarisation. Furthermore, the results also show improved sensitivity when compared to the CMS result. This partially originates from the differences in the fiducial regions but primarily from the MVA techniques. The CMS result applied a combination of BDTs, however this experiment implemented DNNs. The DNNs show an improved ability to discriminate polarisation.

This experiment marks an important step in understanding the EWSB mechanism of the Standard Model. As both theory and experiment improve, the nature of polarised VBS will be uncovered in the future. Moreover, with Run-3 and the HL-LHC, the data available for VBS measurements will significantly grow and with that, the ability to probe polarisation more accurately. The future is bright for polarisation!

11 Sensitivity to VBS production in $W^{\pm}W^{\pm}jj$ at Future Hadron Colliders

To conclude this thesis, this chapter presents a study investigating the prospects of measuring polarised same-sign $W^{\pm}W^{\pm}jj$ at a future collider. The initial study contributed to the 2021 US Snowmass process [163, 164]. Following the Snowmass process, the work detailed in Ref. [163] has been updated and improved. At the time of writing, the work is still ongoing. This chapter presents part of the study with the author's main contributions.

11.1 Introduction

Vector boson scattering (VBS) processes are important probes of the non-Abelian gauge structure of the electroweak (EW) interactions and the EW symmetry breaking mechanism. The unitarity of the tree-level amplitude of the scattering of longitudinally polarised gauge bosons at high energies [165, 166, 167] *could* be restored by the Higgs-like boson observed at the CERN LHC with a mass of about 125 GeV [168, 169, 170]. However, modifications of the cross sections of processes involving scattering of longitudinally polarised gauge bosons with respect to the Standard Model (SM) expectations are predicted in physics beyond the SM models via the presence of additional new resonances or via modifications of the Higgs boson couplings to gauge bosons [171, 172, 173].

The VBS processes at proton-proton (pp) colliders are characterised by the presence of two gauge bosons in association with a forward/backward pair of jets. The goal of this note is to study the prospects for measuring the polarised scattering of same-electric-charge (same-sign) $W^{\pm}W^{\pm}jj$ production at a future high energy pp machine [174, 175]. In particular, our studies focus on future hadron colliders operating at a center-of-mass energy ($\sqrt{s}$) of 27, 50 and 100 TeV (the latter of which corresponding to the envisioned FCC-hh collider [174]) with an integrated luminosity of 30 ab^{-1}. The leptonic decay mode of EW production of same-sign $W^{\pm}W^{\pm}jj$, where both W bosons decay into electrons or muons (including those coming from tau leptons) and their respective (anti-) neutrinos, is a promising final state as the ratio between the strong (QCD-induced) and electroweak production of $W^{\pm}W^{\pm}jj$ is small. The ATLAS and CMS Collaborations have measured EW $W^{\pm}W^{\pm}jj$ production at $\sqrt{s} = 8$ and 13 TeV [176, 177, 178, 179, 180].

The W boson can be polarised either longitudinally (W_L) or transversely (W_T). This leads to three contributions $W_L{}^{\pm}W_L{}^{\pm}jj$, $W_L{}^{\pm}W_T{}^{\pm}jj$ (which includes $W_T{}^{\pm}W_L{}^{\pm}jj$), and $W_T{}^{\pm}W_T{}^{\pm}jj$ to the overall $W^{\pm}W^{\pm}jj$ production. In this study, the candidate $W^{\pm}W^{\pm}jj$ events contain exactly two leptons with the same electric charge, missing transverse momentum, and two jets with a large rapidity separation and a high dijet invariant mass. The helicity eigenstates are defined in the WW center-of-mass reference frame which has a higher contribution from longitudinally polarised W bosons compared to the center-of-mass frame of the colliding partons. Studies of changes in polarization fractions and in kinematic distributions arising from defining the helicity eigenstates in different reference frames at the LHC are reported in Ref. [181]. A maximum-likelihood fit is performed using the distribution of events as a function of a Boosted Decision Tree (BDT) variable optimised to

distinguish between signal and background processes that can mimic the signal signature, to extract the cross sections of the $W_L{}^{\pm}W_L{}^{\pm}jj$, $W_L{}^{\pm}W_T{}^{\pm}jj$, and $W_T{}^{\pm}W_T{}^{\pm}jj$ contributions.

A first measurement attempt of the production cross sections of polarised scattering in the same-sign $W^{\pm}W^{\pm}jj$ process was performed by the CMS Collaboration using a data sample corresponding to an integrated luminosity of 137 fb^{-1} at $\sqrt{s} = 13$ TeV [90], with an observed significance of about one standard deviation for the $W_L{}^{\pm}W_L{}^{\pm}jj$ component. There have been a number of studies focusing on techniques to maximise the sensitivity to the purely longitudinal contributions at the High-Luminosity LHC (HL-LHC) [182, 183, 184, 185]. The CMS Collaboration recently projected the result in Ref. [90] to an integrated luminosity of 3000 fb^{-1}, expected at the end of the HL-LHC, to obtain an expected significance of about four standard deviations for $W_L{}^{\pm}W_L{}^{\pm}jj$ production [186]. Beyond the LHC, studies focusing on the prospects of measuring longitudinally polarised ZZ scattering at a future high-energy muon collider were performed in Ref. [187]. This study follows up in more detail on a contribution to the 2021 US Snowmass process in which a sensitivity to $W_L{}^{\pm}W_L{}^{\pm}jj$ at a future 100 TeV pp collider was projected to be around 17% [163].

11.2 Event Simulation

Signal events are modeled at leading order in QCD using the Monte Carlo event generator MadGraph5_aMC@NLO version 3.4.1 [188, 87] with the NNLO hessian NNPDF3.1 PDF set [189] interfaced with PYTHIA version 8.306 [190] for parton showering. Events where either both W bosons are longitudinally polarized ($W_L{}^{\pm}W_L{}^{\pm}jj$ events), both W bosons are transversely polarized ($W_T{}^{\pm}W_T{}^{\pm}jj$ events), or those where one of the W bosons is longitudinally polarized and the other is transversely polarized ($W_L{}^{\pm}W_T{}^{\pm}jj$ events) are generated separately.

Background processes considered in this analysis include the production of $W^{\pm}W^{\pm}jj$ via the strong interaction, the production of $WZjj$ via the electroweak and strong interactions, and the production of tl^+l^-j, referred to as tZq. These processes are simulated with MadGraph5_aMC@NLO as well.

The Delphes [191] program, using a generic FCC detector card [192], is then used to simulate detector effects for all samples. Jets are clustered from the reconstructed objects using FastJet [193] with the anti-kt clustering algorithm [194], and a distance parameter of 0.4. For the validation of the signal samples, an inclusive $W^{\pm}W^{\pm}jj$ electroweak sample was generated using the same generator and settings as well as Delphes for detector effects, and good agreement was observed with respect to the sum of the polarized signal samples. Detector-specific backgrounds such as those in which lepton charge is misidentified or those in which jets mimic leptons were not considered in this analysis as they are highly detector-specific.

11.3 Event Selection

Signal ($W^{\pm}W^{\pm}jj$) events are characterised by the presence of two high transverse momentum (p_T) same-charge leptons (electrons or muons, denoted by ℓ), large missing transverse momentum (E_T^{miss}) and two forward/backward jets with large di-jet invariant mass

(M_{jj}). In this analysis, events with τ leptons are only considered when the τ decays leptonically. For simplicity, the event selection was optimised for $\sqrt{s} = 100$ TeV to maximise the background rejection and enhance the signal, and then applied to all $\sqrt{s}$ values under study. To eliminate leptons from Z bosons, the invariant mass of a pair of electrons or muons was required to be away from the Z mass shell by at least 10 GeV. Table 55 gives a summary of all object and event selection cuts applied in this analysis. The distributions of events as a function of the separation in azimuthal angle ϕ between the two leading jets in p_T ($\Delta\phi_{jj}$) are shown in Fig. 153 for all the three center-of-mass energies.

Selection type	Requirement
Number of leptons	Exactly 2 same-charge leptons
Lepton p_T	$p_T \geq 15$ GeV
Number of jets	≥ 2
Jet p_T	$p_T \geq 50$ GeV
Di-lepton invariant mass	$M_{ll} \geq 60$ GeV
$Z-$veto	$\lvert M_{ll} - M_Z \rvert > 10$ GeV
Di-jet invariant mass	$M_{jj} \geq 2$ TeV
Missing transverse momentum	$E_T^{miss} \geq 50$ GeV

Table 55: Selection criteria used for $W^{\pm}W^{\pm}jj$ events.

11.4 Theory uncertainties

Three primary sources of theoretical uncertainties are associated with perturbative QCD matrix-element calculations. The first and most significant is due to the missing higher-order terms in perturbative expansions. To estimate this uncertainty, samples are generated with variations of the renormalisation (μ_R) and factorisation (μ_F) scales. Variations up to a factor of 2 are performed in the up and down direction, with the final uncertainty evaluated by taking the 7-point scale variations [152]. In addition to the scale variations, PDF and α_S uncertainties also contribute to the theoretical uncertainty coming from perturbative QCD. PDFs are obtained by fitting experimental data. The theoretical uncertainty associated with the choice in PDF is evaluated by varying the internal parameters of the PDF following the prescription from Ref [153]. Finally, the third source of theoretical uncertainty is associated with the determination of strong coupling constant, α_S. The coupling constant is determined experimentally using fixed-order calculations in perturbation theory. The coupling constant has both experimental and theoretical uncertainties due to the truncated fixed-order calculations. The α_S uncertainties are evaluated at two different α_S values, and the uncertainty is calculated following the recommendations of [153].

Table 56 summarises the effect of theory uncertainties on their respective sample as a relative percentage of the total event yield. The systematics have the largest impact on the 27 TeV samples, with the effects decreasing with centre-of-mass energy. The largest source of uncertainty for all centre-of-mass energies arises from the scale uncertainties on the QCD-induced WZ background.

CoM [TeV]	W_LW_L	W_LW_T	W_TW_T	WW QCD	tZq	WZ EW	WZ QCD
100	4.92	5.32	5.40	13.0	10.4	5.48	30.8
50	10.8	9.99	10.1	18.4	10.3	11.1	37.6
27	14.7	15.4	15.8	23.7	10.4	16.3	52.0

Table 56: Relative percentage effect on the event yields of the theory systematics for each sample.

11.5 Sensitivity to $W^{\pm}W^{\pm}jj$ polarizations

To separate the three $W^{\pm}W^{\pm}jj$ polarisations from background processes, two Boosted Decision Trees (BDTs) were trained using XGBoost [195]. The first BDT separates the $W^{\pm}W^{\pm}jj$ signal from the background processes (SB), and the second BDT separates the three $W^{\pm}W^{\pm}jj$ polarisations (Pol). Several kinematic variables were used as input to the BDTs. This subset of variables for training was chosen from a larger set based on performance. Additionally, a correlation coefficient matrix was generated. The top 15 of the initial 25 variables were kept while avoiding keeping multiple variables with high correlations with respect to each other. Hyperparameters, including the learning rate, maximum depth, subsample ratio, and number of estimators, were tuned via a random grid search algorithm [196]. Simulated MC events were split into training and testing samples used for validation.

The Pol BDT is a multi-class classifier consisting of three outputs. Each output corresponds to a classification probability associated with one of the polarisation fractions. The SB BDT is a binary classifier, with a single output giving the classification probability of an inclusive $W^{\pm}W^{\pm}jj$ event. The two BDTs are combined within the analysis framework to generate a three-dimensional histogram. The x and y dimensions denote the predicted probability of observing a longitudinal-longitudinal and transverse-transverse event coming from the output of the polarisation BDT, and the z dimension is the probability of an inclusive $W^{\pm}W^{\pm}jj$ signal event coming from the SB BDT. As the sum predicted probability of the Pol BDT is equal to one, only two outputs are necessary for complete sensitivity to polarisation. The three-dimensional BDT is "unrolled" to form a one-dimensional histogram. The unrolling translates each individual bin in the three-dimensional histogram to a bin in a one-dimensional histogram.

The sensitivity to the $W^{\pm}W^{\pm}jj$ polarisation contributions is estimated using a profile likelihood fit as implemented in the RooFit [197] and RooStats [198] packages. We consider three signal strength parameters, μ_{LL}, μ_{LT} and μ_{TT} corresponding to each of the $W^{\pm}W^{\pm}jj$ polarization states. In addition to the theoretical uncertainties, we assume a normalisation uncertainty of 2% for all the processes due to uncertainties in the integrated luminosity. The uncertainties associated with a limited number of simulated events are also included in the profile likelihood fit. We do not consider any other sources of systematic uncertainties in this study.

The sensitivity is measured for two distributions: $\Delta\phi_{jj}$ and the three-dimensional BDT. The $\Delta\phi_{jj}$ distribution is used as a baseline benchmark for the sensitivity to polarisation and as a comparison to illustrate the improvements when applying multivariate analysis techniques such as BDTs. Before fitting, the BDT distributions are also subject

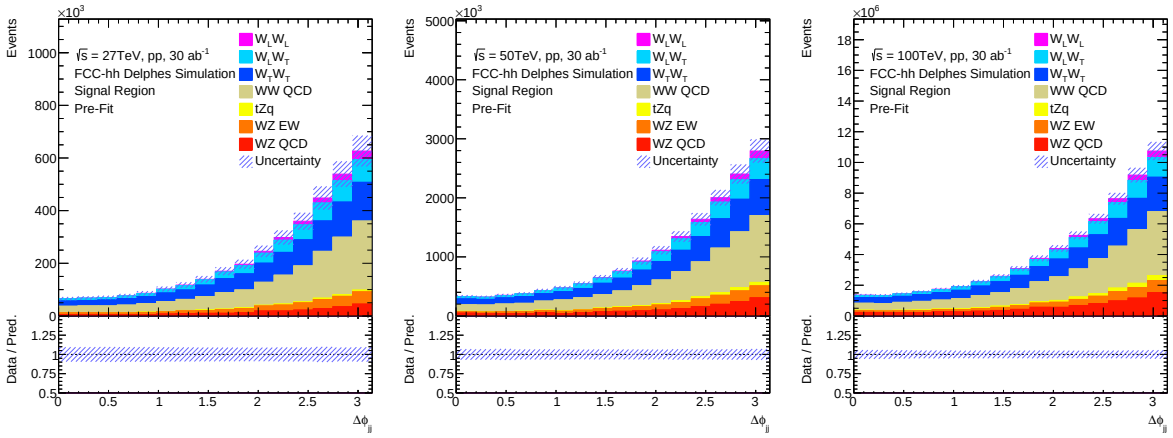


Figure 153: The pre-fit distribution of events as a function of $\Delta\phi_{jj}$.

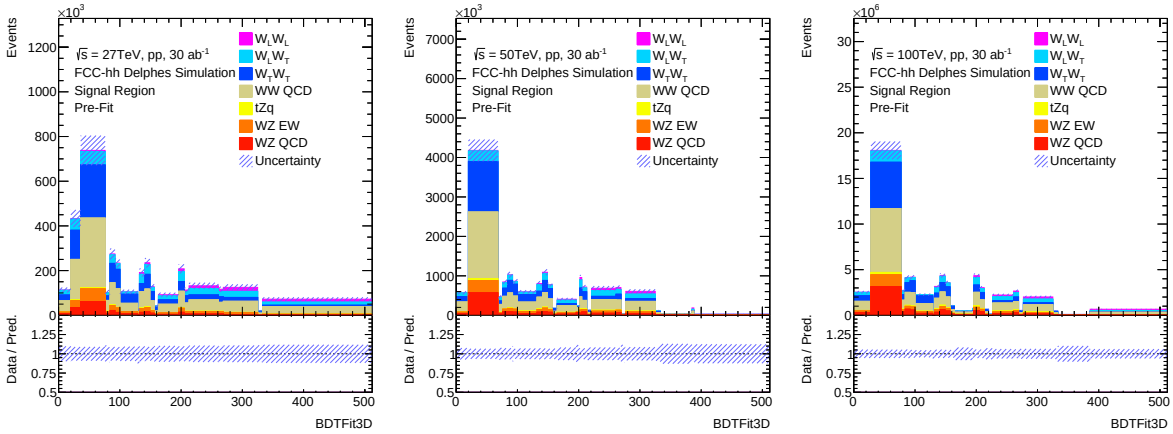


Figure 154: The pre-fit distribution of events as a function of the unrolled BDT (BDTFit3D).

to a binning optimisation procedure. The procedure requires a fixed minimum of background and signal events per bin, as well as a minimum of 200 events per background. The procedure then iteratively merges bins with no major improvements in the significance, estimated as suggested by Ref. [199]. The pre-fit distributions for $\Delta\phi_{jj}$ and the BDT following the binning optimisation procedure are shown in Figs. 153 and 154.

Table 57 shows the signal strengths μ, defined as the ratio to the SM cross-section, and the corresponding uncertainties for each $W^{\pm}W^{\pm}jj$ polarization state for all $\sqrt{s}$ values under study. The values are expected to be unity because the Asimov dataset is used in the fit.

Polarization	Signal Strength: BDT		
	$\sqrt{s} = 27$ TeV	$\sqrt{s} = 50$ TeV	$\sqrt{s} = 100$ TeV
μ_{LL}	1 ± 0.20	1 ± 0.15	1 ± 0.13
μ_{LT}	1 ± 0.12	1 ± 0.085	1 ± 0.080
μ_{TT}	1 ± 0.12	1 ± 0.069	1 ± 0.062
	Signal Strength: $\Delta\phi_{jj}$		
μ_{LL}	1 ± 1.02	1 ± 0.62	1 ± 0.40
μ_{LT}	1 ± 0.45	1 ± 0.42	1 ± 0.14
μ_{TT}	1 ± 0.33	1 ± 0.26	1 ± 0.12

Table 57: Signal strengths μ after the $\Delta\phi_{jj}$ and BDT fit for each $W^{\pm}W^{\pm}jj$ polarization state and all $\sqrt{s}$ values under study.

11.6 Conclusions

Cross-section measurements of the polarised scattering of same-sign $W^{\pm}W^{\pm}jj$ production using the leptonic decay mode at a future high energy pp collider with an integrated luminosity of 30 ab^{-1} should be possible with a relative precision around 20% at $\sqrt{s} =$ 27 TeV, 15% at $\sqrt{s} = 50$ TeV and 13% at $\sqrt{s} = 100$ TeV for the purely longitudinal contribution and better for the other contributions. The uncertainty on the normalisation factors shown in Table 57 is predominately statistical.

Table 57 illustrates significantly better sensitivity to polarisation with the multivariate analysis as opposed to fitting a single kinematic distribution. At a future collider, constraining the purely longitudinal fraction will still be challenging due to the overwhelmingly larger event yields from the other polarisation configurations and will likely require multivariate techniques. Crucially, however, the results show significant improvements in the sensitivity to polarisation when compared to the experiment presented in this thesis. The projected sensitivity to the purely longitudinal scattering is at a minimum, a factor of four better than the sensitivity achieved in the ATLAS experiment.

References

[1] The ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, September 2012.

[2] Carl D. Anderson. The positive electron. *Phys. Rev.*, 43:491–494, Mar 1933.

[3] Paul A. M. Dirac. Quantum theory of emission and absorption of radiation. *Proc. Roy. Soc. Lond. A*, 114:243, 1927.

[4] F. J. Dyson. The radiation theories of Tomonaga, Schwinger, and Feynman. *Phys. Rev.*, 75:486–502, Feb 1949.

[5] C. N. Yang and R. L. Mills. Conservation of isotopic spin and isotopic gauge invariance. *Phys. Rev.*, 96:191–195, Oct 1954.

[6] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc. C*, 680519:367–377, 1968.

[7] Sheldon L. Glashow. The renormalizability of vector meson interactions. *Nuclear Physics*, 10:107–117, 1959.

[8] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

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

[10] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964.

[11] Matic Lubej. https://www.physik.uzh.ch/groups/serra/StandardModel.html.

[12] E. Noether. Invariante variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1918:235–257, 1918.

[13] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.

[14] Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 3 2014.

[15] L.D. Faddeev and V.N. Popov. Feynman diagrams for the Yang-Mills field. *Physics Letters B*, 25(1):29–30, 1967.

[16] R. L. Arnowitt and S. I. Fickler. Quantization of the Yang-Mills field. *Phys. Rev.*, 127:1821–1829, Sep 1962.

[17] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental test of parity conservation in beta decay. *Phys. Rev.*, 105:1413–1415, Feb 1957.

[18] Stefanie Todt. The $W^{\pm}W^{\pm}$ Scattering Process, 2022. Presented 02 Dec 2022.

[19] Yoichiro Nambu. Quasi-particles and gauge invariance in the theory of superconductivity. *Phys. Rev.*, 117:648–663, Feb 1960.

[20] J Goldstone. Field theories with "superconductor" solutions. *Nuovo Cimento*, 19:154–164, 1961.

[21] Joseph Lykken and Maria Spiropulu. The future of the Higgs boson. *Physics Today*, 66(12):28–33, 12 2013.

[22] V. Barger, Kingman Cheung, T. Han, and R. J. N. Phillips. Strong w^+w^+ scattering signals at pp supercolliders. *Phys. Rev. D*, 42:3052–3077, Nov 1990.

[23] R. Frederix, S. Frixione, V. Hirschi, D. Pagani, H.-S. Shao, and M. Zaro. The automation of next-to-leading order electroweak calculations. *Journal of High Energy Physics*, 2018(7), July 2018.

[24] CERN. https://home.cern/. [Online; accessed 1-February-2024].

[25] CERN. The Synchrocyclotron. https://home.cern/science/accelerators/synchrocyclotron, 1957. [Online; accessed 3-February-2024].

[26] F.J. Hasert *et al.* Search for elastic muon-neutrino electron scattering. *Physics Letters B*, 46(1):121–124, 1973.

[27] G. *et al.* Arnison. Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV. *Phys. Lett. B*, 122:103–116, 1983.

[28] M. *et al.* Banner. Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider. *Phys. Lett. B*, 122:476–485, 1983.

[29] The ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1–29, 2012.

[30] The CMS Collaboration. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012.

[31] O Brüning, H Burkhardt, and S Myers. The Large Hadron Collider. *Prog. Part. Nucl. Phys.*, 67:705–734, 2012. Published in "Progress in Particle and Nuclear Physics" (2012), 10.1016/j.ppnp.2012.03.001.

[32] Lyndon R Evans. LHC Status and Plans. 1998.

[33] C Kilby and on behalf of the ATLAS Collaboration. The design and performance of the ATLAS Inner Detector trigger for Run 2 LHC Collisions at $\sqrt{s}$=13 tev. *Journal of Physics: Conference Series*, 762(1):012029, oct 2016.

[34] *ATLAS: technical proposal for a general-purpose pp experiment at the Large Hadron Collider at CERN*. LHC technical proposal. CERN, Geneva, 1994.

[35] *Technical proposal*. LHC technical proposal. CERN, Geneva, 1994. Cover title : CMS, the Compact Muon Solenoid : technical proposal.

[36] *LHCb : Technical Proposal*. CERN, Geneva, 1998.

[37] *ALICE: Technical proposal for a Large Ion collider Experiment at the CERN LHC*. LHC technical proposal. CERN, Geneva, 1995.

[38] The ATLAS Collaboration. The ATLAS experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(08):S08003, aug 2008.

[39] The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider: A Description of the Detector Configuration for Run 3. Technical report, CERN, Geneva, 2023.

[40] CERN. High Luminosity LHC Project. https://hilumilhc.web.cern.ch/content/hl-lhc-project. [Online; accessed 18-February-2024].

[41] ATLAS Collaboration. Experiment Briefing: Keeping the ATLAS Inner Detector in perfect alignment. General Photo, 2020.

[42] The ATLAS Collaboration. ATLAS Insertable B-Layer Technical Design Report Addendum. Technical report, 2012. Addendum to CERN-LHCC-2010-013, ATLAS-TDR-019.

[43] A Vogel. ATLAS Transition Radiation Tracker (TRT): Straw Tube Gaseous Detectors at High Rates. Technical report, CERN, Geneva, 2013.

[44] The ATLAS Collaboration. The ATLAS experiment at the CERN Large Hadron Collider: a description of the detector configuration for Run 3. *Journal of Instrumentation*, 19(05):P05063, May 2024.

[45] Ana Maria Henriques Correia. The ATLAS Tile Calorimeter. Technical report, CERN, Geneva, 2015.

[46] S. *et al.* Khalek. The ALFA Roman Pot detectors of ATLAS. *Journal of Instrumentation*, 11(11):P11013–P11013, November 2016.

[47] Giulia Pancheri and Yogendra N. Srivastava. Introduction to the physics of the total cross section at lhc: A review of data and models. *The European Physical Journal C*, 77(3), March 2017.

[48] The ATLAS collaboration. Operation of the ATLAS trigger system in Run 2. *Journal of Instrumentation*, 15(10):P10004–P10004, October 2020.

[49] Markus Elsing, Luc Goossens, Armin Nairz, and Guido Negri. The ATLAS Tier-0: Overview and operational experience. *Journal of Physics: Conference Series*, 219(7):072011, apr 2010.

[50] The ATLAS Collaboration. ATLAS data quality operations and performance for 2015–2018 data-taking. *Journal of Instrumentation*, 15(04):P04003–P04003, April 2020.

[51] The ATLAS Collaboration. Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2. *The European Physical Journal C*, 77(10), October 2017.

[52] Cecilia Tosciri. *Physics Object Reconstruction*, pages 49–69. Springer International Publishing, 2021.

[53] The ATLAS Collaboration. Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13$ tev. *The European Physical Journal C*, 79(8), August 2019.

[54] The ATLAS Collaboration. Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data. Technical report, CERN, Geneva, 2016.

[55] W Lampl, S Laplace, D Lelas, P Loch, H Ma, S Menke, S Rajagopalan, D Rousseau, S Snyder, and G Unal. Calorimeter Clustering Algorithms: Description and Performance. Technical report, CERN, Geneva, 2008.

[56] The ATLAS Collaboration. Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung. Technical report, CERN, Geneva, 2012.

[57] The ATLAS Collaboration. Muon reconstruction and identification efficiency in atlas using the full Run 2 pp collision data set at $\sqrt{s} = 13$ tev. *The European Physical Journal C*, 81(7), July 2021.

[58] The ATLAS Collaboration. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$=13 tev. *The European Physical Journal C*, 76(5), May 2016.

[59] J. Illingworth and J. Kittler. A survey of the hough transform. *Computer Vision, Graphics, and Image Processing*, 44(1):87–116, 1988.

[60] The ATLAS Collaboration. ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ tev. *The European Physical Journal C*, 79(11), November 2019.

[61] The ATLAS Collaboration. Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment. Technical report, CERN, Geneva, 2017.

[62] S. Catani and D. Zeppenfeld. Jet Algorithms. In *1st Les Houches Workshop on Physics at TeV Colliders*, pages 132–140, 5 2000.

[63] The ATLAS Collaboration. Topological cell clustering in the atlas calorimeters and its performance in lhc run 1. *The European Physical Journal C*, 77(7), July 2017.

[64] The ATLAS Collaboration. Jet reconstruction and performance using particle flow with the ATLAS Detector. *The European Physical Journal C*, 77(7), July 2017.

[65] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, April 2008.

[66] The ATLAS Collaboration. Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ tev with the ATLAS detector. *The European Physical Journal C*, 81(8), August 2021.

[67] The ATLAS Collaboration. Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, 78(11), November 2018.

[68] O. Aberle *et al.* *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2020.

[69] The ATLAS Collaboration. Technical Design Report for the ATLAS Inner Tracker Pixel Detector. Technical report, CERN, Geneva, 2017.

[70] Luka Nedic and Karolos Potamianos. Low-pT reconstruction for the ITk. Technical report, CERN, Geneva, 2022.

[71] William Patrick McCormack *et al.* Low-p_{T} tracking for ATLAS in nominal LHC pileup. Technical report, CERN, Geneva, 2020.

[72] The ATLAS Collaboration. Search for direct pair production of sleptons and charginos decaying to two leptons and neutralinos with mass splittings near the W-boson mass in $\sqrt{s} = 13$ tev pp collisions with the ATLAS detector. *Journal of High Energy Physics*, 2023(6), June 2023.

[73] The ATLAS Collaboration. Expected tracking and related performance with the updated ATLAS Inner Tracker layout at the High-Luminosity LHC. Technical report, CERN, Geneva, 2021.

[74] ATLAS Collaboration. Software and computing for run 3 of the atlas experiment at the lhc, 2024.

[75] Michael Leyton and on behalf of the ATLAS Collaboration. Minimum Bias and Underlying Event Measurements with ATLAS. pages 11–19, 2012.

[76] The ATLAS Collaboration. Observation of photon-induced w^+w^- production in pp collisions at $\sqrt{s} = 13$ tev using the ATLAS detector. *Physics Letters B*, 816:136190, May 2021.

[77] Savannah Clawson *et al.* Sensitivity to exclusive WW production in photon scattering at the High Luminosity LHC. Technical report, CERN, Geneva, 2021.

[78] The ATLAS Collaboration. Standard Model Summary Plots February 2022. Technical report, CERN, Geneva, 2022.

[79] Michael Rauch. Vector-Boson Fusion and Vector-Boson Scattering, 2016.

[80] Shalu Solomon. The first measurement of the differential cross-section of electroweak $W^{\pm}W^{\pm}jj$ production at 13 TeV with the ATLAS detector, 2022. Presented 13 Jan 2023.

[81] ATLAS Collaboration. ATLAS Feature: Unraveling Nature's secrets – vector boson scattering at the LHC. General Photo, 2020.

[82] M. Tanabashi *et al.* Review of Particle Physics. *Phys. Rev. D*, 98:030001, Aug 2018.

[83] Alessandro Ballestrero *et al.* W boson polarization in vector boson scattering at the LHC. *Journal of High Energy Physics*, 2018(3), March 2018.

[84] Z. Bern *et al.* Left-handed W bosons at the LHC. *Physical Review D*, 84(3), August 2011.

[85] W. J. Stirling and E. Vryonidou. Electroweak gauge boson polarisation at the lhc. *Journal of High Energy Physics*, 2012(7), July 2012.

[86] A. Kadeer, J. G. Körner, and U. Moosbrugger. Helicity analysis of semileptonic hyperon decays including lepton-mass effects. *The European Physical Journal C*, 59(1):27–47, November 2008.

[87] Diogo Buarque Franzosi, Olivier Mattelaer, Richard Ruiz, and Sujay Shil. Automated predictions from polarized matrix elements. *Journal of High Energy Physics*, 2020(4), April 2020.

[88] The ATLAS Collaboration. Measurement and interpretation of same-sign w boson pair production in association with two jets in pp collisions at $\sqrt{s} = 13$ tev with the atlas detector. *Journal of High Energy Physics*, 2024(4), April 2024.

[89] Aram Apyan *et al.* Support note for measurement of fiducial and differential $W^{\pm}W^{\pm}jj$ production cross-sections at *sqrts* =13 TeV using the full Run-2 dataset. Technical report, CERN, Geneva, 2021.

[90] The CMS Collaboration. Measurements of production cross sections of polarized same-sign w boson pairs in association with two jets in proton-proton collisions at $\sqrt{s}$=13 tev. *Physics Letters B*, 812:136018, January 2021.

[91] ATLAS Collaboration. The ATLAS Common Analysis Framework. https://atlas-caf.web.cern.ch/.

[92] The ATLAS Collaboration. The ATLAS Simulation Infrastructure. *The European Physical Journal C*, 70(3):823–874, September 2010.

[93] S. Agostinelli *et al.*. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.

[94] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. A brief introduction to pythia 8.1. *Computer Physics Communications*, 178(11):852–867, June 2008.

[95] J. Alwall *et al.* The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7), July 2014.

[96] Enrico Bothmann *et al.* Event generation with Sherpa 2.2. *SciPost Physics*, 7(3), September 2019.

[97] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.

[98] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.

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

[100] Christian Bierlich *et al.* A comprehensive guide to the physics and usage of PYTHIA 8.3, 2022.

[101] S Schumann and F Krauss. A parton shower algorithm based on catani-seymour dipole factorisation. *Journal of High Energy Physics*, 2008(03):038–038, March 2008.

[102] Stefan Höche, Frank Krauss, Marek Schönherr, and Frank Siegert. A critical appraisal of nlo+ps matching methods. *Journal of High Energy Physics*, 2012(9), September 2012.

[103] Richard D. Ball *et al.* The path to proton structure at 1% accuracy: NNPDF Collaboration. *The European Physical Journal C*, 82(5), May 2022.

[104] Tanju Gleisberg and Stefan Höche. Comix, a new matrix element generator. *Journal of High Energy Physics*, 2008(12):039–039, December 2008.

[105] F. Cascioli, P. Maierhöfer, and S. Pozzorini. Scattering amplitudes with open loops. *Phys. Rev. Lett.*, 108:111601, Mar 2012.

[106] Baptiste Cabouat and Torbjörn Sjöstrand. Some dipole shower studies. *The European Physical Journal C*, 78(3), March 2018.

[107] The ATLAS Collaboration. Modelling of the vector boson scattering process $pp \to W^{\pm}W^{\pm}jj$ in Monte Carlo generators in ATLAS. Technical report, CERN, Geneva, 2019.

[108] ATLAS Collaboration. Good Run Lists for Analysis Run-2. https://twiki.cern.ch/twiki/bin/viewauth/%20AtlasProtected/GoodRunListsForAnalysisRun2l.

[109] The ATLAS Collaboration. ATLAS data quality operations and performance for 2015–2018 data-taking. *Journal of Instrumentation*, 15(04):P04003–P04003, April 2020.

[110] The ATLAS Collaboration. Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data. *Journal of Instrumentation*, 14(12):P12006–P12006, December 2019.

[111] The ATLAS Collaboration. Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ tev using the atlas detector. *The European Physical Journal C*, 76(11), October 2016.

[112] The ATLAS Collaboration. Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run. Technical report, CERN, Geneva, 2017.

[113] D Adams *et al.* Recommendations of the Physics Objects and Analysis Harmonisation Study Groups 2014. Technical report, CERN, Geneva, 2014.

[114] The ATLAS Collaboration. Search for heavy Majorana neutrinos with the ATLAS detector in pp collisions at $\sqrt{s} = 8$ TeV. *Journal of High Energy Physics*, 2015(7), July 2015.

[115] Jose Antonio Fernandez Pretel. Estimation of the electron charge misidentification background for the same-sign WW measurement with the ATLAS detector at $\sqrt{s} =$ 13 TeV, 2020. Presented 03 Apr 2020.

[116] Carsten Bittrich *et al.* Support note for measurement of electroweak $W^{\pm}W^{\pm}jj$ production at $\sqrt{s} = 13$ TeV. Technical report, CERN, Geneva, 2018.

[117] The ATLAS Collaboration. Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016. *The European Physical Journal C*, 79(3), March 2019.

[118] The ATLAS Collaboration. Measurement of the cross-section for electroweak production of dijets in association with a Z boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Physics Letters B*, 775:206–228, December 2017.

[119] B Bellenot, Ph Canal, O Couet, G Ganis, P Mato, L Moneta, A. Naumann, and D. Piparo. Root 6 and beyond: Tobject, c++14 and many cores. *Journal of Physics: Conference Series*, 664(6):062006, dec 2015.

[120] Fred James and Matthias Winkler. MINUIT User's Guide. 6 2004.

[121] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[122] Michael A. Nielsen. *Neural Networks and Deep Learning*. Determination Press, 2015. http://neuralnetworksanddeeplearning.com/.

[123] Overfitting and methods of addressing it. https://analystprep.com/study-notes/cfa-level-2/quantitative-method/overfitting-methods-addressing/. Accessed: 2024-04-15.

[124] What's a deep neural network? https://www.bmc.com/blogs/deep-neural-network/. Accessed: 2024-04-15.

[125] George V. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2:303–314, 1989.

[126] Aram Apyan *et al.* Polarization measurements of the same-sign WW vector boson scattering at $\sqrt{s} = 13$ TeV with the ATLAS detector. Technical report, CERN, Geneva, 2023.

[127] D. Rainwater, R. Szalapski, and D. Zeppenfeld. Probing color-singlet exchange in z+2-jet events at the cern lhc. *Physical Review D*, 54(11):6680–6689, December 1996.

[128] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[129] Roman Garnett. *Bayesian Optimization*. Cambridge University Press, 2023.

[130] Ben Poole, Subhaneil Lahiri, Maithra Raghu, Jascha Sohl-Dickstein, and Surya Ganguli. Exponential expressivity in deep neural networks through transient chaos, 2016.

[131] Samuel S. Schoenholz, Justin Gilmer, Surya Ganguli, and Jascha Sohl-Dickstein. Deep information propagation, 2017.

[132] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus), 2016.

[133] Prajit Ramachandran, Barret Zoph, and Quoc V. Le. Searching for activation functions, 2017.

[134] Soufiane Hayou, Arnaud Doucet, and Judith Rousseau. On the selection of initialization and activation function for deep neural networks, 2018.

[135] Soufiane Hayou, Arnaud Doucet, and Judith Rousseau. On the impact of the activation function on deep neural networks training, 2019.

[136] Junho Lee, Nicolas Chanon, Andrew Levin, Jing Li, Meng Lu, Qiang Li, and Yajun Mao. Polarization fraction measurement in same-sign $w^{\pm}w^{\pm}$ scattering using deep learning. *Physical Review D*, 99(3), February 2019.

[137] Classification: ROC and AUC. https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc.

[138] Jinsung Yoon, James Jordon, and Mihaela van der Schaar. INVASE: Instance-wise variable selection using neural networks. In *International Conference on Learning Representations*, 2019.

[139] James M. Joyce. *Kullback-Leibler Divergence*, pages 720–722. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[140] François Chollet. keras. https://github.com/fchollet/keras, 2015.

[141] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014.

[142] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015.

[143] Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang. On large-batch training for deep learning: Generalization gap and sharp minima, 2017.

[144] Jacob Searcy, Lillian Huang, Marc-André Pleier, and Junjie Zhu. Determination of the $w^{\pm}w^{p}m$ polarization fractions in $pp \to w^{\pm}w^{\pm}jj$ using a deep machine learning technique. *Physical Review D*, 93(9), May 2016.

[145] The ATLAS Collaboration. Measurement of the polarisation of W bosons produced with large transverse momentum in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS experiment. Measurement of the polarisation of W bosons produced with large transverse momentum in pp collisions at sqrt(s) = 7 TeV with the ATLAS experiment. *Eur. Phys. J. C*, 72:2001, 2012.

[146] Michael S. Chanowitz and Mary K. Gaillard. The TeV physics of strongly interacting W's and Z's. *Nuclear Physics B*, 261:379–431, 1985.

[147] Ansgar Denner and Giovanni Pelliccioli. Polarized electroweak bosons in $W^{+}W^{-}$ production at the LHC including NLO QCD effects. *Journal of High Energy Physics*, 2020(9), September 2020.

[148] Andy Buckley *et al.* Rivet user manual. *Computer Physics Communications*, 184(12):2803–2819, December 2013.

[149] Abien Fred Agarap. Deep Learning using Rectified Linear Units (ReLU), 2019.

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

[151] John C. Collins, Davison E. Soper, and George Sterman. Factorization of hard processes in qcd, 2004.

[152] The ATLAS Collaboration. Pmg systematic uncertainty recipes. https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/PmgSystematicUncertaintyRecipes.

[153] Jon Butterworth *et al.* PDF4LHC recommendations for LHC Run II. *Journal of Physics G: Nuclear and Particle Physics*, 43(2):023001, January 2016.

[154] Richard D. Ball *et al.* Parton distributions for the LHC run II. *Journal of High Energy Physics*, 2015(4), April 2015.

[155] Benedikt Biedermann, Ansgar Denner, and Mathieu Pellen. Complete NLO corrections to $W^{\pm}W^{\pm}$ scattering and its irreducible background at the LHC. *Journal of High Energy Physics*, 2017(10), October 2017.

[156] Mareen Hoppe, Marek Schönherr, and Frank Siegert. Polarised cross sections for vector boson production with SHERPA, 2023.

[157] Anders Andreassen and Benjamin Nachman. Neural networks for full phase-space reweighting and parameter tuning. *Phys. Rev. D*, 101:091901, May 2020.

[158] A. Denner and S. Pozzorini. One-loop leading logarithms in electroweak radiative corrections: I. results. *The European Physical Journal C*, 18(3):461–480, January 2001.

[159] Ansgar Denner and Giovanni Pelliccioli. NLO EW and QCD corrections to polarized ZZ production in the four-charged-lepton channel at the LHC. *Journal of High Energy Physics*, 2021(10), October 2021.

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

[161] TRExFitter. Trexfitter documentation. https://trexfitter-docs.web.cern.ch/trexfitter-docs.

[162] Kyle Cranmer, George Lewis, Lorenzo Moneta, Akira Shibata, and Wouter Verkerke. HistFactory: A tool for creating statistical models for use with RooFit and RooStats. Technical report, New York U., New York, 2012.

[163] Aram Apyan *et al.* Sensitivity to longitudinal vector boson scattering in $W^{\pm}W^{\pm}jj$ at future hadron colliders. *arXiv:2203.07994*, 2022.

[164] Alberto Belloni *et al.* Report of the topical group on electroweak precision physics and constraining new physics for snowmass 2021, 2022.

[165] M. J. G. Veltman. Second Threshold in Weak Interactions. *Acta Phys. Pol. B*, 8:475, 1977.

[166] Benjamin W. Lee, C. Quigg, and H. B. Thacker. The strength of weak interactions at very high-energies and the Higgs boson mass. *Phys. Rev. Lett.*, 38:883, 1977.

[167] Benjamin W. Lee, C. Quigg, and H. B. Thacker. Weak interactions at very high-energies: the role of the Higgs boson mass. *Phys. Rev. D*, 16:1519, 1977.

[168] The ATLAS Collaboration. Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1, 2012.

[169] S. Chatrchyan *et al.* Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 716:30, 2012.

[170] S. Chatrchyan *et al.* Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV. *JHEP*, 06:081, 2013.

[171] Domenec Espriu and Brian Yencho. Longitudinal WW scattering in light of the Higgs boson discovery. *Phys. Rev. D*, 87:055017, 2013.

[172] Jung Chang, Kingman Cheung, Chih-Ting Lu, and Tzu-Chiang Yuan. WW scattering in the era of post-Higgs-boson discovery. *Phys. Rev. D*, 87:093005, 2013.

[173] Seung J. Lee, Myeonghun Park, and Zhuoni Qian. Probing unitarity violation in the tail of the off-shell Higgs boson in $V_L V_L$ mode. *Phys. Rev. D*, 100:011702, 2019.

[174] A. Abada *et al.* FCC-hh: The Hadron Collider: Future Circular Collider Conceptual Design Report Volume 3. *Eur. Phys. J. ST*, 228:755, 2019.

[175] CEPC Conceptual Design Report: Volume 1 - Accelerator. 9 2018.

[176] Georges Aad *et al.* Evidence for electroweak production of $W^{\pm}W^{\pm}jj$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Phys. Rev. Lett.*, 113:141803, 2014.

[177] Vardan *et al.* Khachatryan. Study of vector boson scattering and search for new physics in events with two same-sign leptons and two jets. *Phys. Rev. Lett.*, 114:051801, 2015.

[178] Albert M Sirunyan *et al.* Observation of electroweak production of same-sign W boson pairs in the two jet and two same-sign lepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 120:081801, 2018.

[179] The ATLAS Collaboration. Observation of electroweak production of a same-sign W boson pair in association with two jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. Lett.*, 123:161801, 2019.

[180] Albert M Sirunyan *et al.* Measurements of production cross sections of WZ and same-sign WW boson pairs in association with two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 809:135710, 2020.

[181] Alessandro Ballestrero, Ezio Maina, and Giovanni Pelliccioli. Different polarization definitions in same-sign WW scattering at the LHC. *Phys. Lett. B*, 811:135856, 2020.

[182] Jacob Searcy, Lillian Huang, Marc-André Pleier, and Junjie Zhu. Determination of the WW polarization fractions in $pp \to W^{\pm}W^{\pm}jj$ using a deep machine learning technique. *Phys. Rev. D*, 93:094033, 2016.

[183] Junho Lee *et al.* Polarization fraction measurement in same-sign WW scattering using deep learning. *Phys. Rev. D*, 99(3):033004, 2019.

[184] P. Azzi *et al.* Standard Model Physics at the HL-LHC and HE-LHC. *arXiv:1902.04070*, 2018.

[185] Jennifer Roloff, Viviana Cavaliere, Marc-André Pleier, and Lailin Xu. Sensitivity to longitudinal vector boson scattering in semileptonic final states at the HL-LHC. *Phys. Rev. D*, 104:093002, 2021.

[186] CMS Collaboration. Prospects for the measurement of vector boson scattering production in leptonic $\mathrm{W^{\pm}W^{\pm}}$ and WZ diboson events at $\sqrt{s} = 14$ TeV at the High-Luminosity LHC. Technical report, CERN, Geneva, 2021.

[187] Tianyi Yang *et al.* Longitudinally polarized ZZ scattering at a muon collider. *Phys. Rev. D*, 104:093003, 2021.

[188] J. Alwall *et al.* The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07(79), 2014.

[189] R.D. Ball *et al.* Parton distributions with LHC data. *Nuclear Physics B*, 867(244), 2013.

[190] T. Sjöstrand *et al.* An Introduction to PYTHIA 8.2. *Computer Physics Communications*, 191(159), 2015.

[191] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. DELPHES 3, A modular framework for fast simulation of a generic collider experiment. *JHEP*, 02:057, 2014.

[192] Michele Selvaggi. Delphes fcc-hh card. https://github.com/delphes/delphes/blob/master/cards/FCC/FCChh.tcl. Accessed: 2022.

[193] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet user manual. *Eur. Phys. J. C*, 72:1896, 2012.

[194] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-kt jet clustering algorithm. *JHEP*, 04:063, 2008.

[195] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *CoRR*, abs/1603.02754, 2016.

[196] F. Pedregosa *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[197] Wouter Verkerke and David P. Kirkby. The RooFit toolkit for data modeling. *eConf*, C0303241:MOLT007, 2003.

[198] Lorenzo Moneta, Kevin Belasco, Kyle S. Cranmer, S. Kreiss, Alfio Lazzaro, Danilo Piparo, Gregory Schott, Wouter Verkerke, and Matthias Wolf. The RooStats Project. *PoS*, ACAT2010:057, 2010.

[199] Formulae for Estimating Significance. Technical report, CERN, Geneva, 2020.