Dissertation

zur Erlangung des akademischen Grades
Doctor rerum naturalium
(Dr. rer. nat.)

Longitudinal $W^{\pm}$ Boson Polarization in Same-Charged $W^{\pm}W^{\pm}$ Scattering in the ATLAS Experiment

Max Vincent Stange
geboren am 06.05.1996 in Hoyerswerda

TECHNISCHE UNIVERSITÄT DRESDEN
INSTITUT FÜR KERN- UND TEILCHENPHYSIK
FACHRICHTUNG PHYSIK
FAKULTÄT MATHEMATIK UND NATURWISSENSCHAFTEN

Eingereicht am 30. Januar 2025

1. Gutachter:	Prof. Dr. Michael Kobel
2. Gutachter:	Dr. Frank Siegert
3. Gutachter:	Prof. Dr. Karolos Potamianos

Abstract

The Standard Model of particle physics forms the theoretical basis of modern physics alongside the general theory of relativity. Precision measurements are used to search for flaws in this theory in order to discover physics beyond the Standard Model. One of the most interesting physical processes is the longitudinally polarized scattering of two same-charged $W^{\pm}$ bosons. Polarized $W^{\pm}$ boson scattering offers a unique opportunity to explore the electroweak symmetry breaking of the Standard Model.

The investigation of $W^{\pm}$ boson scattering requires particle collisions at center-of-mass energies that can only be achieved at the Large Hadron Collider (LHC). The final states of the scattering process are measured by the CMS and ATLAS detector. The $W^{\pm}W^{\pm}$ scattering process was observed for the first time in 2018 by the CMS collaboration in Ref. [1] and confirmed one year later by the ATLAS collaboration in Ref. [2]. Due to the short lifetime of the $W^{\pm}$ bosons, the leptonic decay products of the bosons were investigated in both analyses. Based on these analyses, the Run 2 data recorded at the LHC from 2015 to 2018 are examined for the polarization states of the scattered $W^{\pm}$ bosons. The same-charged $W^{\pm}$ boson scattering is the first possibility to study the polarization in vector boson scattering (VBS). In 2021, the CMS analysis of longitudinally polarized $W^{\pm}$ boson scattering was published in Ref. [3]. However, this analysis did not provide evidence above the $3\,\sigma$ significance threshold for the existence of a longitudinal polarization state in VBS.

This thesis is a major contribution to the ATLAS analysis of the Run 2 data aiming for this evidence. In this thesis, the first experimental application of approximate polarized next-to-leading-order corrections on same-charged $W^{\pm}$ bosons scattering is achieved. These corrections employ state-of-the-art calculations of the polarized cross-sections and significantly improve the Standard Model prediction. Based on the Standard Model prediction, deep neural networks (DNNs) are trained to study the $W^{\pm}$ boson polarization in the measured data. The DNNs are optimized for the polarization discrimination and validated against data in a control region. The decision-making of the DNNs is investigated to achieve a better understanding of the kinematic information learned by the networks. The DNN output for the measured data and the Standard Model prediction is analyzed by a profile likelihood fit to extract the significance and cross-section of the longitudinal polarization state. The likelihood fit requires the estimation of the theoretical uncertainty of the polarization states which is implemented in this thesis. Furthermore, a new strategy is developed to investigate and constrain the impact of these uncertainties. For the final application of the profile likelihood fit, a two-dimensional bin optimization algorithm is designed for the histograms of the DNN output to maximize the expected significance.

The contribution of the $W^{\pm}W^{\pm}$ scattering with at least one longitudinally polarized $W^{\pm}$ boson is measured with a significance of $3.4\,\sigma$ and a fiducial cross-section of $1.04 \pm 0.35\,\text{fb}$. Thus, this ATLAS analysis provides the first evidence for the longitudinal polarization in VBS. The measured fiducial cross-section is in agreement with the Standard Model prediction of $1.35\,\text{fb}$.

For the scattering of two longitudinally polarized $W^{\pm}$ bosons, a significance of $0.082\,\sigma$ is observed. The corresponding fiducial cross-section is found to be smaller than $0.50\,\text{fb}$ at 95% confidence level. This upper limit covers the Standard Model prediction of $0.33\,\text{fb}$ and constraints the fiducial cross-sections predicted by alternative theories.

Zusammenfassung

Das Standardmodell der Teilchenphysik bildet neben der allgemeinen Relativitätstheorie die theoretische Grundlage der modernen Physik. Mit Hilfe von Präzisionsmessungen wird nach Schwachstellen in dieser Theorie gesucht, um Physik jenseits des Standardmodells zu entdecken. Einer der interessantesten physikalischen Prozesse ist die longitudinal polarisierte Streuung von zwei gleichgeladenen $W^{\pm}$-Bosonen. Die polarisierte $W^{\pm}$-Bosonenstreuung bietet eine einzigartige Möglichkeit, die elektroschwache Symmetriebrechung des Standardmodells zu erforschen.

Die Untersuchung der $W^{\pm}$-Bosonenstreuung erfordert Teilchenkollisionen bei Schwerpunktsenergien, die nur am Large Hadron Collider (LHC) erreicht werden können. Die Endzustände des Streuprozesses werden mit dem CMS- und ATLAS-Detektor gemessen. Der $W^{\pm}W^{\pm}$-Streuprozess wurde 2018 zum ersten Mal von der CMS-Kollaboration in Ref. [1] beobachtet und ein Jahr später von der ATLAS-Kollaboration in Ref. [2] bestätigt. Aufgrund der kurzen Lebensdauer der $W^{\pm}$-Bosonen wurden in beiden Analysen die leptonischen Zerfallsprodukte der Bosonen untersucht. Basierend auf diesen Analysen werden die von 2015 bis 2018 am LHC aufgenommenen Run 2-Daten auf die Polarisationszustände der gestreuten $W^{\pm}$-Bosonen untersucht. Die gleichgeladene $W^{\pm}$-Bosonenstreuung ist die erste Möglichkeit, die Polarisation in der Vektor-Bosonen-Streuung (VBS) zu untersuchen. Im Jahr 2021 wurde die CMS-Analyse der longitudinal polarisierten $W^{\pm}$-Bosonenstreuung in Ref. [3] veröffentlicht. Diese Analyse lieferte jedoch keinen Beweis oberhalb der $3\,\sigma$-Signifikanzschwelle für die Existenz eines longitudinalen Polarisationszustandes in VBS.

Diese Arbeit ist ein wichtiger Beitrag zur ATLAS-Analyse der Run 2-Daten, die auf diesen Nachweis abzielt. In dieser Arbeit wird die erste experimentelle Anwendung von approximativen polarisierten Korrekturen nächstführender Ordnung auf gleichgeladene $W^{\pm}$-Bosonenstreuung erreicht. Diese Korrekturen verwenden modernste Berechnungen der polarisierten Wirkungsquerschnitte und verbessern die Vorhersage des Standardmodells erheblich. Auf der Grundlage der Vorhersage des Standardmodells werden tiefe neuronale Netze (DNNs) trainiert, um die Polarisation des $W^{\pm}$-Bosons in den gemessenen Daten zu untersuchen. Die DNNs werden für die Polarisationsunterscheidung optimiert und anhand von Daten in einer Kontrollregion validiert. Die Entscheidungsfindung der DNNs wird untersucht, um ein besseres Verständnis der von den Netzen erlernten kinematischen Informationen zu erreichen. Die DNN-Ausgabe für die gemessenen Daten und die Vorhersage des Standardmodells wird durch einen Profil-Likelihood-Fit analysiert, um die Signifikanz und den Querschnitt des longitudinalen Polarisationszustands zu extrahieren. Der Likelihood-Fit erfordert die Abschätzung der theoretischen Unsicherheit der Polarisationszustände, welche in dieser Arbeit implementiert wird. Außerdem wird eine neue Strategie entwickelt, um die Auswirkungen dieser Unsicherheiten zu untersuchen und zu begrenzen. Für die endgültige Anwendung der Profil-Likelihood-Fit wird ein zweidimensionaler Bin-Optimierungsalgorithmus für die Histogramme der DNN-Ausgabe entwickelt, um die erwartete Signifikanz zu maximieren.

Der Beitrag der $W^{\pm}W^{\pm}$-Streuung mit mindestens einem longitudinal polarisierten $W^{\pm}$-Boson wird mit einer Signifikanz von $3.4\,\sigma$ und einem Wirkungsquerschnitt von 1.04 ± 0.35 fb gemessen. Damit liefert diese ATLAS-Analyse den ersten Nachweis für die longitudinale Polarisation in VBS. Der gemessene Wirkungsquerschnitt stimmt mit der Vorhersage des Standardmodells von 1.35 fb überein.

Für die Streuung von zwei longitudinal polarisierten $W^{\pm}$-Bosonen wird eine Signifikanz von $0,082\,\sigma$ beobachtet. Der entsprechende Wirkungsquerschnitt ist kleiner als $0,50$ fb bei einem Konfidenzintervall von 95%. Diese Obergrenze deckt die Vorhersage des Standardmodells von $0,33$ fb ab und schränkt die von alternativen Theorien vorhergesagten Wirkungsquerschnitte ein.

Contents

1 **Introduction** 3

I **Background and Methodology** 5

2 **Experimental Setup** 9
2.1 The Large Hadron Collider . . . 9
2.1.1 Injection Chain of the LHC . . . 10
2.1.2 Magnet System . . . 11
2.1.3 Vacuum System . . . 11
2.1.4 Run 2 Performance . . . 11
2.2 The ATLAS Experiment . . . 12
2.2.1 Detector Coordinate System . . . 14
2.2.2 Layout of the ATLAS Detector . . . 14
2.2.3 Object Reconstruction . . . 16
2.2.4 Trigger System . . . 17

3 **Theoretical Foundations** 19
3.1 Spins in Relativistic Theory . . . 19
3.1.1 Lorentz Group and Lorentz Algebra . . . 19
3.1.2 Representations of the Lorentz Group . . . 20
3.1.3 Spin and Helicity in the Poincare Algebra . . . 21
3.2 Quantum Field Theory . . . 21
3.2.1 Lagrange Density of Free Particles . . . 22
3.2.2 U(1) Symmetry of Quantum Electrodynamics . . . 24
3.2.3 SU(3) Symmetry of Quantum Chromo Dynamics . . . 25
3.2.4 SU(2) Symmetry of Weak Interaction . . . 26
3.2.5 Electroweak Unification in $\mathrm{U}(1)_Y \otimes \mathrm{SU}(2)$ . . . 27
3.2.6 Higgs Mechanism . . . 27
3.2.7 Summary of the Standard Model . . . 31
3.3 Polarized Same-Charged $W^\pm W^\pm$ Scattering . . . 32
3.3.1 Polarized Intermediate $W^\pm$ Bosons . . . 32
3.3.2 Motivating the Higgs Boson by Same-Charged $W_L^\pm W_L^\pm$ Scattering . . . 34
3.3.3 Same-Charged $W^\pm W^\pm$ Scattering at LHC . . . 35
3.3.4 Polarization Reference Frame . . . 39

4 **Simulation of the Standard Model Prediction** 41
4.1 Hard Subprocess . . . 41
4.2 Parton Shower . . . 43
4.3 Multi-Jet Merging . . . 44
4.4 Hadronization . . . 44
4.5 Hadron and Tau Decay . . . 45
4.6 QED Radiation . . . 45
4.7 Underlying Event . . . 45
4.8 Event Levels of Simulated Data . . . 45

5 Neural Networks **47**
5.1 Basic Components of Neural Networks . . . 47
5.2 Activation Functions . . . 48
5.3 Loss Functions . . . 49
5.4 Training Neural Networks . . . 50
5.4.1 Stochastic Gradient Descent . . . 50
5.4.2 Adam Optimization . . . 51
5.4.3 Back Propagation for FFNs . . . 52
5.5 Network Hyperparameters . . . 53
5.6 Generalization . . . 53
5.6.1 Overtraining and Undertraining . . . 54
5.6.2 Regularization . . . 54
5.7 Model Evaluation by the ROC-AUC . . . 56
5.8 k-Fold Method . . . 57
5.9 Access Feature Importance via SHAP . . . 59

II Polarization Analysis **63**

6 Object and Event Selection **67**
6.1 Pre-selection . . . 67
6.2 Object Selection . . . 67
6.2.1 Electron Selection . . . 68
6.2.2 Muon Selection . . . 69
6.2.3 Jet Selection . . . 69
6.2.4 Reconstruction of the Missing Transverse Momentum . . . 70
6.2.5 Overlap Removal . . . 70
6.3 Event Selection . . . 71
6.3.1 Signal Region . . . 71
6.3.2 Low-m_{jj} Control Region . . . 72
6.3.3 $W^{\pm}Z$ Control Region . . . 72
6.3.4 Fiducial Signal Region . . . 72

7 Standard Model prediction **75**
7.1 $W^{\pm}W^{\pm}jj$ Signal Simulation . . . 75
7.1.1 $W^{\pm}W^{\pm}jj$-EW Polarization by MADGRAPH . . . 75
7.1.2 $W^{\pm}W^{\pm}jj$-EW Polarization by SHERPA . . . 75
7.1.3 Polarization Interference . . . 76
7.1.4 $W^{\pm}W^{\pm}jj$-INT and $W^{\pm}W^{\pm}jj$-QCD Background . . . 76
7.2 Background Estimation . . . 76
7.2.1 $W^{\pm}Z$ Background . . . 76
7.2.2 Non-Prompt Background . . . 77
7.2.3 Charge-Flip Background . . . 79
7.2.4 $V\gamma$, ZZ, Z+jets, Top, and Tri-boson Background . . . 80

8 Higher-Order Corrections on $W^{\pm}W^{\pm}jj$-EW **81**
8.1 Approximate NLO-QCD Correction . . . 81
8.1.1 Comparison of Simulation by Sherpa and MadGraph . . . 81
8.1.2 Correction Procedure . . . 83
8.1.3 Comparison of the Sherpa Simulation Setups . . . 85
8.1.4 Multivariate Correction . . . 90
8.1.5 Technical Validation of the Multivariate Correction . . . 93
8.1.6 Validation of the Multivariate Correction with MadGraph . . . 98
8.1.7 One-dimensional Correction . . . 99
8.1.8 Discussion of Theory Uncertainties . . . 102
8.2 NLO-EW Correction . . . 103
8.3 Combination of the NLO Corrections . . . 105

9 **Training of Discriminator DNNs** **107**
9.1 Training Input 107
9.2 Network Architecture 111
9.3 DNN Optimization Algorithm 111
9.3.1 Optimization with OPTUNA 111
9.3.2 Backward Elimination of Training Variables 112
9.3.3 Combined Optimization Strategy 113
9.3.4 Optimization Results 114

10 **Validation of DNN Discriminators** **123**
10.1 Validation of the 5-Fold Method 123
10.2 DNN Validation with Measured Data 129
10.2.1 Validation in Low-m_{jj} Control Region 129
10.2.2 Validation in the Signal Region 131
10.3 Interpretation of the Classifier Decision-Making 132
10.3.1 Dataset Choice for SHAP Sampling 132
10.3.2 Evaluation of Feature Impact 134

11 **Uncertainties** **141**
11.1 Experimental Uncertainties 141
11.2 Modeling Uncertainties 142
11.2.1 $W^{\pm}W^{\pm}jj$-EW Polarization Theory Uncertainties 142
11.2.2 NLO Correction Uncertainty of $W^{\pm}W^{\pm}jj$-EW 148
11.2.3 Theory Uncertainties of the $W^{\pm}Z$ Background 148
11.2.4 Reweighting Uncertainty of $W^{\pm}Z$ Background 149

12 **Statistical Evaluation and Results** **151**
12.1 Introduction of the Likelihood Fit 151
12.2 Network Application Procedure 153
12.3 Optimization of the Histogram Binning 153
12.3.1 Significance estimation 154
12.3.2 Rebinning 155
12.3.3 Optimization Results 156
12.4 Application of Scale Uncertainties 157
12.4.1 Kinematic Impact of Scale Systematics 158
12.4.2 Maximal Shape Effects in DNN Score Distribution 158
12.4.3 Resulting Correlation Scheme 164
12.5 Expected Results 165
12.5.1 Expected Significance 165
12.5.2 Expected Fiducial Cross-section 172
12.6 Observed Results 176
12.6.1 Observed Significance 176
12.6.2 Observed Fiducial Cross-Section 183
12.7 Results with Polarization Modeling by MadGraph 184
12.8 Comparison with CMS Results 185
12.9 Projection for Future LHC Runs 185

13 **Conclusion** **189**

III Appendix 191

A **Monte Carlo Event Generator Samples** **193**

B **Approximate NLO-QCD Correction** **197**

C **DNN Optimization** **213**

D Data Closure Test **225**

E Interpretation of the Classifier Decision-Making **231**

F Uncertainties **235**

G Statistical Evaluation **239**
G.1 Expected Results 240
G.2 Observed Results 242

Conventions

Units: The following thesis uses natural units by setting the speed of light c and the reduced Planck constant $\hbar$ to 1. Thus, the energy unit eV is also used for masses and momenta.

Summation convention: The Einstein summation convention is used for contravariant and covariant indices. In order to distinguish the summation convention from indices that simply occur twice, it is only used for indices that are Greek letters. For example the term $a_\mu b^\mu$ is a scalar product of two four-vectors a and b while $a_i b^i$ is the product of their i-th component.

Vectors: Vector-type variables are written in bold type $\boldsymbol{k}$. Vectors in three-dimensional space, such as momentum or velocity, are indicated by a vector arrow $\vec{v}$.

Order of measured objects: Events measured at the ATLAS experiment contain multiple particles or jets. A set of these objects is sorted in descending order according to their momentum transversal to the beam axis. The object with the largest transverse momentum is referenced as the “leading” object, followed by the “subleading” object.

Chapter 1

Introduction

Modern physics is based on two fundamental theories: the theory of General Relativity and the Standard Model of particle physics. These two theories cover two distinct aspects of our physical world. General Relativity describes the concept of gravity and is based on the geometrical idea of a curved spacetime. The theory of General Relativity is thus a description of our universe, which primarily takes place on the size scales of planetary systems and galaxies. On the other hand, the Standard Model of particle physics describes the interactions of the smallest components of our world. This theory covers the electromagnetic, weak and strong interactions of elementary particles. The Standard Model is the unification of special relativity and quantum mechanics and comprehensively describes all known fundamental interactions apart from gravity. The theory has existed since the 1970s but had a fundamental weakness when it was first conceived: Although the interactions of particles could be accurately predicted, these particles were massless in theory. Since particles are experimentally known to have masses, theorists were investigating different methods to introduce particle masses into the Standard Model. The most successful approach was the Higgs mechanism, which postulates the existence of a new particle: the Higgs boson. Although this theory became increasingly popular, the existence of a Higgs boson could not be confirmed experimentally until the 21st century. The hope of discovering this new particle lay in high-energy physics and motivated the construction of the Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN). The LHC accelerates hadrons and collides them at previously unattainable center-of-mass energies. The experimental setup of the LHC and one of its most important detectors is described in chapter 2. In 2012, the experiments at CERN led to the discovery of the Higgs boson. This discovery was one of the most important physical breakthroughs of the 21st century so far because it finally filled a gap in one of the most fundamental theories of physics. The resulting Standard Model with Higgs mechanism is summarized in chapter 3.

Although the discovery of the Higgs boson was a massive breakthrough for modern physics, various physical problems remain unresolved. The Standard Model does not explain the matter-antimatter asymmetry in the universe, neutrino oscillations, or the dark matter and dark energy known from astronomic measurements. Therefore, although the Standard Model is a very well tested theory, it has to be replaced by other theories in the future. The particle physics community searches for flaws in the theory and possibilities for physics beyond the Standard Model. Since neither the Standard Model nor current experiments raise hope of discovering further particles, scientific research focuses on precision measurements. If differences between experimental data and the Standard Model prediction were discovered, new theories could emerge that would change our understanding of the universe.

Since the recently discovered Higgs boson has a central role in the Standard Model, many of the precision measurements focus on further investigating the Higgs mechanism. One of the most promising approaches is the polarization measurement in the scattering of $W^{\pm}$ or Z bosons, called vector bosons. The polarization of a particle is given by the orientation of its spin relative to its direction of motion. Particles with spin 1 can have a transverse polarization and, if not massless, longitudinal polarization. Since the vector bosons gain their mass by interacting with the Higgs field, their longitudinal polarization directly results from the Higgs mechanism. Moreover, without a Higgs boson, the scattering of two longitudinally polarized bosons would violate a fundamental physics concept, even if a different source for the boson masses were available. Without a Higgs boson, the longitudinally polarized vector boson scattering would violate unitarity, leading to unphysical scattering probabilities of more than 100%. The interaction with the Higgs boson avoids the unitarity violation. The polarization measurement in the scattering of vector bosons thus tests not only the relevance of the Higgs mechanism for particle masses but also the necessity of the Higgs boson to restore the fundamental concepts of physics. Therefore, the polarized vector boson scattering is one of the most promising processes for investigating the Higgs mechanism and searching for physics beyond the Standard Model.

The LHC is the only experiment capable of producing the necessary center-of-mass energy to study vector boson scattering. The corresponding measurements are performed using the ATLAS and CMS detectors. The design of the LHC and the ATLAS detector is explained in chapter 2. Possible candidates of vector boson scattering are the oppositely charged $W^{\pm}W^{\mp}$-, the same-charged $W^{\pm}W^{\pm}$-, the $W^{\pm}Z$- and ZZ-scattering. A polarization analysis of the oppositely charged $W^{\pm}W^{\mp}$ scattering is not feasible currently due to the high background contribution from other processes. Two ATLAS analyses [4][5] and one CMS analysis [6] have studied the polarization of $W^{\pm}Z$ boson pairs. The polarization of a ZZ boson pair was investigated in the ATLAS analysis in Ref. [7]. These analyses showed that a polarization analysis in the case of two vector bosons is possible in principle. However, they did not directly investigate the scattering of the vector bosons, which has only been possible for the same-charged $W^{\pm}W^{\pm}$ scattering so far.

The Run 2 data recorded from 2015 to 2018 at the CMS detector was analyzed for the polarization in same-charged $W^{\pm}W^{\pm}$ scattering [3]. Two reference frames were used to define the polarization: the center-of-mass frame of the two interacting $W^{\pm}W^{\pm}$ (WW-cmf) and the partonic center-of-mass frame (pp-cmf). For the scattering of two longitudinally polarized $W^{\pm}$ bosons, a 95% confidence interval upper limit of 1.17 fb in the WW-cmf and 1.06 fb in the pp-cmf was obtained. When measuring events with at least one longitudinal $W^{\pm}$ boson, a significance of $2.3\,\sigma$ was achieved in the WW-cmf and $2.6\,\sigma$ in the pp-cmf. However, these significances are not sufficient to claim evidence nor observation.

This thesis analyzes the Run 2 data of the ATLAS experiment to provide the evidence of longitudinal polarization in $W^{\pm}W^{\pm}$ scattering and test new techniques for future analyses. The existing analysis of unpolarized same-charged $W^{\pm}$ boson scattering in Ref. [8] is used as a baseline for the polarization analysis. The existing analysis strategy is extended by the Standard Model prediction of the polarization states and an extensive machine learning strategy to extract the polarization information from data. Since the leptonic decays of the $W^{\pm}$ bosons are studied, a significant part of the kinematic information about the polarized $W^{\pm}$ boson is lost, as the emitted neutrino is not detectable in the ATLAS detector. Therefore, neural networks extract the remaining information from the entire measured phase space.

For the sake of clarity, this thesis is split into two parts. The first part contains the theoretical background necessary to understand the analysis strategy. In the second part, the polarization analysis is described with a focus on the author's personal contribution to the ATLAS analysis:

In chapter 8, recent event generator features and theory calculations are used to correct the individual predictions of the polarization states for next-to-leading-order effects. A state-of-the-art hyperparameter optimization strategy is applied to train neural networks with an optimal discriminative power for the polarization states. In chapter 10, the resulting networks are validated and their decision-making process is investigated. The theory uncertainties of the polarization prediction are implemented in chapter 11. In chapter 12, a new strategy is presented to study and constrain the impact of these theory uncertainties. This final chapter provides also the expected and observed analysis results of the polarization measurement.

Part I

Background and Methodology

The first part of this thesis provides the theoretical background of the polarization analysis. The investigated data originates from the ATLAS experiment at the LHC. An overview of this experimental setup and the reconstruction of the measured objects is provided in chapter 2. Subsequently, the relevant theoretical framework of particle physics is summarized in chapter 3. It introduces the Standard Model and the corresponding Higgs mechanism and highlights the importance of polarization in vector boson scattering. The chapter ends with an explanation of polarization in same-charged $W^{\pm}$ boson scattering produced at the LHC. The measurement of the polarized vector boson scattering requires the Standard Model prediction of the polarized signal and the corresponding background processes. Therefore, the concept of event generators is described in chapter 4 to provide an overview of how the Standard Model prediction is achieved by simulations. For the polarization measurement, neural networks extract the available information from the entire phase space. The concept and functionality of fully-connected feedforward networks are explained in chapter 5. This description covers the network architecture, the training process, the generalization error, and the investigation of the network performance.

Chapter 2

Experimental Setup

Very high interaction energies are necessary to study processes like the scattering of two $W^{\pm}$ bosons. These energies are achieved by the collision of two high-energetic particles or nuclei. At the time of this thesis, the biggest and most powerful accelerator is the Large Hadron Collider (LHC) [9] at the research complex of the European Organization for Nuclear Research (CERN) at the Swiss-French border. Several detectors are available at the LHC ring to measure the resulting collision. This thesis focuses on the measurement at the ATLAS detector [10]. The following sections briefly introduce the setup of the LHC and ATLAS detector, followed by a section summarizing how the objects relevant to this analysis are reconstructed in the ATLAS experiment.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [9] is a ring accelerator with a 26.7 km circumference. It is built in the underground 45 m–170 m below the surface in the tunnel of the former LEP accelerator operating between 1989 and 2000. The LHC accelerates and collides beams of protons and beams of heavy ions for various experiments. Experiments besides the ATLAS experiment are CMS, LHCB, TOTEM, and ALICE, which are not studied in this thesis. The LHC is built to reach a center-of-mass energy of 14 TeV for proton-proton collisions. The LHC consists of two superconducting accelerator rings to collide two protons at this energy. The two proton beams are separated by 194 mm and only share about a 130 m long pipe along the interaction regions.

The number of events

$$N_{\text{event}} = L\sigma_{\text{event}} \tag{2.1}$$

of a specific event type produced at the LHC is determined by the corresponding cross-section σ_{event} and the luminosity L of the accelerator. The luminosity

$$L = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi\epsilon_n\beta^*} F \tag{2.2}$$

depend on the following properties of the collider:

- Number of particles per bunch N_b
- Number of bunches per beam n_b
- Revolution frequency f_{rev}
- Relativistic gamma factor γ_r
- Normalized transverse beam emittance ϵ_n
- Beta function at collision point β^*
- Geometric luminosity reduction factor F due to the beam crossing angle at the collision point

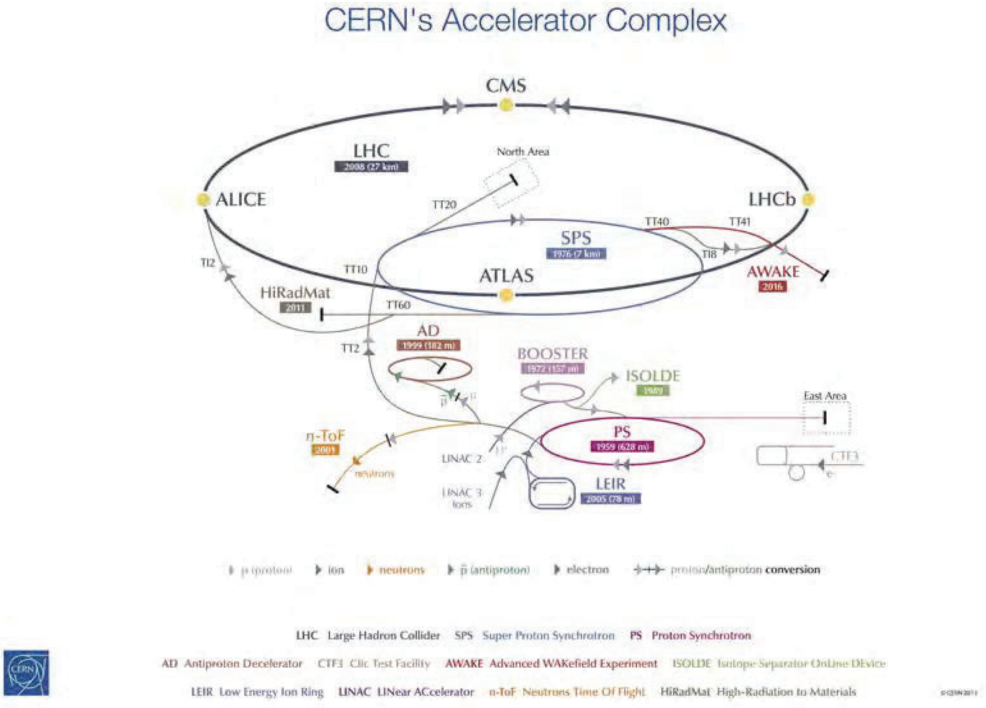

CERN's Accelerator Complex
CMS
LHC
2008 (27 km)
North Area
ALICE
LHCb
TT20
TT40
TT41
SPS
1976 (7 km)
TI8
TI2
TT10
ATLAS
AWAKE
2016
HiRadMat
2011
TT60
AD
1999 (182 m)
TT2
BOOSTER
1972 (157 m)
ISOLDE
1989
East Area
n-ToF
2001
PS
1959 (628 m)
CTF3
e⁻
neutrons
LINAC 2
LINAC 3
Ions
LEIR
2005 (78 m)
p (proton)
ion
neutrons
p̄ (antiproton)
electron
proton/antiproton conversion
LHC Large Hadron Collider
SPS Super Proton Synchrotron
PS Proton Synchrotron
AD Antiproton Decelerator
CTF3 Clic Test Facility
AWAKE Advanced WAKefield Experiment
ISOLDE Isotope Separator OnLine DEvice
LEIR Low Energy Ion Ring
LINAC LINear ACcelerator
n-ToF Neutrons Time Of Flight
HiRadMat High-Radiation to Materials
© CERN 2013

Figure 2.1: Overview of the CERN accelerator complex during Run 2 taken from Ref. [15].

The LHC is designed for 2808 proton bunches per beam with 25 ns bunch spacing [9]. Thus, a luminosity of $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ can be reached [9]. In order to achieve this impressive luminosity and collision energy, the LHC was built with a cutting-edge vacuum and magnet system and a multi-step injection chain that pre-accelerates the used proton beams.

2.1.1 Injection Chain of the LHC

The LHC is the final step of an accelerator chain relying on several pre-accelerators. The acceleration of the proton beam starts at the linear accelerator Linac2 [11] followed by the ring accelerators Proton Synchrotron Booster [12], Proton Synchrotron [13], and Super Proton Synchrotron [14] before the beam is injected in the LHC. An overview of the entire CERN accelerator complex during Run 2 of the LHC is shown in figure 2.1.

Linac2

The origin of the protons colliding in the LHC is a bottle of hydrogen gas. An electrical field ionizes the hydrogen atoms before injecting them into the linear accelerator Linac2 [11]. Linac2 consists of cylindrical conductors accelerating the protons and quadrupole magnets focusing the beam. The resulting pulsed proton beam enters the subsequent Proton Synchrotron Booster at a proton energy of 50 MeV. For the next runs of the LHC Linac2 is replaced by Linac4.

Proton Synchrotron Booster

The proton beam leaving the Linac2 accelerator is split vertically by a series of pulsed magnets. The separated beam enters the four rings of the Proton Synchrotron Booster (PSB) [12]. The Booster rings have a radius of 25 m and consist of 32 dipole and 48 quadrupole magnets to accelerate and preserve the proton bunches. At the end of the acceleration, the proton bunches of the four PSB rings are synchronized and recombined by a system of vertical kickers. The PSB was initially designed to produce beams with 800 MeV proton energy. After several upgrades until the early 2000, the PSB can reach 1.4 GeV to fulfill the requirements for the LHC injection chain.

Proton Synchrotron

The PSB is followed by the Proton Synchrotron (PS) [13][16] which was after its first run in 1959, one of the most powerful colliders. The PS has a circumference of 628 meters and is assembled on a floating floor to minimize the impact of external disturbance. The accelerator consists of 100 straight sections separated by 100 main magnets. At the straight sections, additional auxiliary magnets and radio-frequency cavities accelerating the protons are installed. The same magnetic elements are used to focus and bend the beam, which caused many technical challenges at the time the PS was built. After numerous uses in various experiments and for pre-acceleration for other accelerators, the PS, which accelerates protons to up to 26 GeV, is an integral part of the LHC injection chain.

Super Proton Synchrotron

The last step of the injection chain is the Super Proton Synchrotron (SPS) [14], the second-largest accelerator at CERN. 1317 magnets are installed at the SPS, including 744 main dipole magnets to bend the proton beam. Similar to the previous steps of the injection chain, these magnet systems consist of mainly conventional room-temperature electromagnets. One of the greatest achievements of the SPS as a stand-alone accelerator was the discovery of the $W^{\pm}$ and Z bosons. Today, with an acceleration to 450 GeV, it is the last step in the LHC injection chain.

2.1.2 Magnet System

Since the LHC has to accelerate two counter-rotating proton beams, the magnet system described in [9] provides opposite magnetic fields in two separate rings. To reduce space consumption and costs, a ”twin-bore” design was chosen for almost all magnet systems. The superconducting magnets are based on NbTi Rutherford cables. With the cooling below 2 K provided by superfluid helium, these cables can operate at magnetic fields above 8 T. In total, 1232 main dipole magnets, about 3800 single aperture, and 1000 twin-aperture corrector magnets are installed along the ring. The resulting magnet system of the LHC consists of a complex combination of dipole, quadrupole, sextupole, and octupole magnets to deflect, correct, focus, and defocus the beams.

2.1.3 Vacuum System

The LHC operates three vacuum systems described in Ref. [9]. The first two are the insulation vacuum systems for the cryomagnets and the helium distribution. At cryogenic temperatures, these systems operate at a pressure of 10^{-6} mbar. The third vacuum system provides the vacuum for the beam pipelines. This vacuum has a gas density that is equivalent to $10^{15}\,\mathrm{H_2\,m^{-3}}$ to enable a beam lifetime of 100 hours. An even purer vacuum of $10^{13}\,\mathrm{H_2\,m^{-3}}$ is generated in the interaction regions, which further reduces the potential background for the experiments.

2.1.4 Run 2 Performance

The LHC’s outstanding design and injection chain are also reflected in its luminosity. The LHC runs until 2016 can be split into main periods: Run 1, mainly from 2011 to 2012, and Run 2, in 2015 to 2016. During Run 1 the proton beam operated at an energy of 3.5 TeV to 4 TeV [17]. In the second period, Run 2, this energy was increased to 6.5 TeV [18]. The integrated luminosity over the individual years of Run 1 and Run 2 can be seen in figure 2.2. The commissioning years 2011 for Run 1 and 2015 for Run 2 show a significantly lower luminosity than their corresponding production years in 2012 and 2016 to 2018. From the comparison of the slopes of the integrated luminosities in figure 2.2 and the reached peak

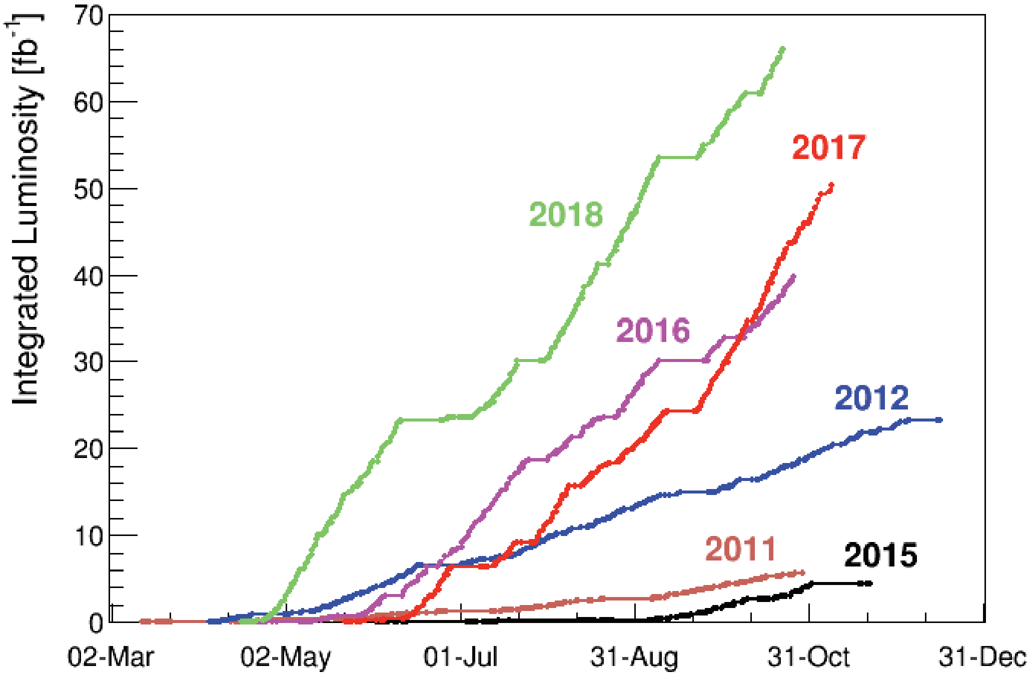


Figure 2.2: The LHC's integrated luminosity of Run 1 and Run 2 periods. The figure is taken from Ref. [18]

luminosities in figure 2.3 one can see that the Run 2 of the LHC provided a significant higher luminosity than during Run 1. The higher luminosity is the result of various improvements done after Run 1. In particular, the higher brightness in the LHC injection chain and the β^* function at the collision point, which is smaller than intended in the initial design of the LHC, should be emphasized. Thus Run 2 produced a total integrated luminosity of 160 fb^{-1} despite technical difficulties such as air ingress into the beam vacuum system during the winter shutdown 2017 [18].

In summary, the Run 2 data analyzed in this thesis was produced with a record center-of-mass energy of 13 TeV and an outstanding integrated luminosity of 160 fb^{-1}. For the ATLAS experiment, an integrated Run 2 luminosity of 140 fb^{-1} is recorded and passes the standard data-quality selection explained in Ref. [19]. Therefore, the Run 2 dataset offers promising possibilities for studying high-energy physics and searching for physics beyond the Standard Model. This thesis analyzes the 140 fb^{-1} Run 2 dataset of the ATLAS detector.

2.2 The ATLAS Experiment

The LHC offers a unique opportunity to study collisions at an unprecedented center-of-mass energy and luminosity. In order to fully utilize these possibilities, the four independent detectors ATLAS, CMS, ALICE, and LHCb are installed along the LHC ring. The data analyzed in this thesis originates from the ATLAS detector shown in figure 2.4. With a length of 45 m and a height of 22 m ATLAS is the largest detector ever built for a particle collider. It is built to handle the high collision rate of the LHC and the emitted radiation doses while performing particle measurements with high precision. The following sections give a short overview of the ATLAS detector, the object reconstruction, and the trigger system.

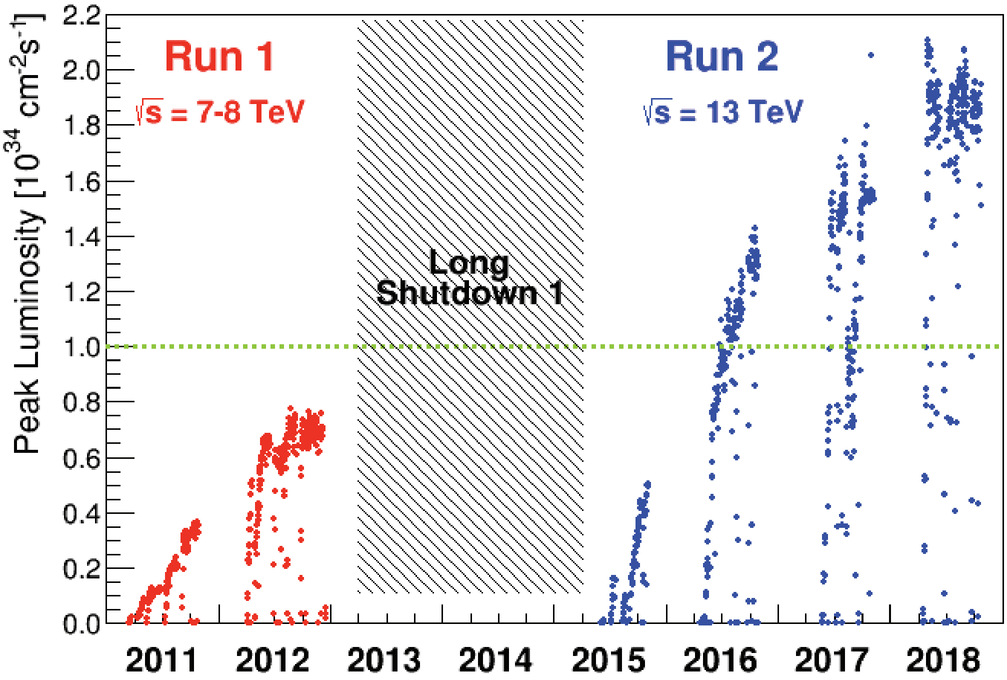


Figure 2.3: The peak luminosity during Run 1 and Run 2 of the LHC taken from Ref. [18]

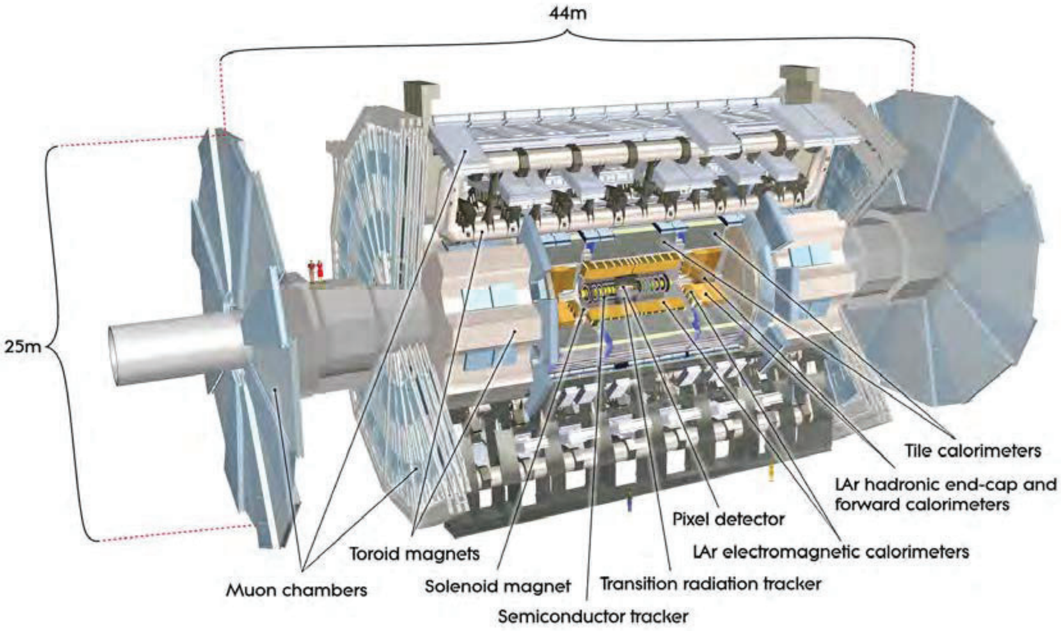

44m
25m
Tile calorimeters
LAr hadronic end-cap and forward calorimeters
Pixel detector
LAr electromagnetic calorimeters
Toroid magnets
Muon chambers
Solenoid magnet
Transition radiation tracker
Semiconductor tracker

Figure 2.4: Schematic overview of the ATLAS detector and its main components. The image is taken from Ref. [20].

2.2.1 Detector Coordinate System

The ATLAS detector is built around the beam pipeline with the collision point at its center. Thus, the detector has a rotation symmetry with respect to the beam pipeline. The coordinate system used to describe particles originating from the collision is shown in figure 2.5. In the Cartesian coordinate system the x-axis is directed towards the center of the collider and the y-axis points upwards. Therefore, the x- and y-axes are transverse to the beam pipeline. The z-axis points counterclockwise in the direction of the beam.

To simplify the description of particles emitted from the collision the spherical coordinates ϕ and θ are introduced. ϕ is the polar angle with respect to the x-axis in a plane orthogonal to the beam pipeline. The azimuth angle θ is measured with respect to the z-axis. In most cases, the pseudo-rapidity

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

is used instead of the angle θ to describe the azimuth orientation. The pseudo-rapidity η contains the same geometrical information, but in the relativistic limit $E \gg mc^2$ or for massless particles, the differences in η are invariant under boosts in the direction of the beam pipeline. Therefore, using η instead of θ enables an invariant description of the event kinematic. Based on this coordinate system, the overall angular separation can be measured by

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{2.4}$$

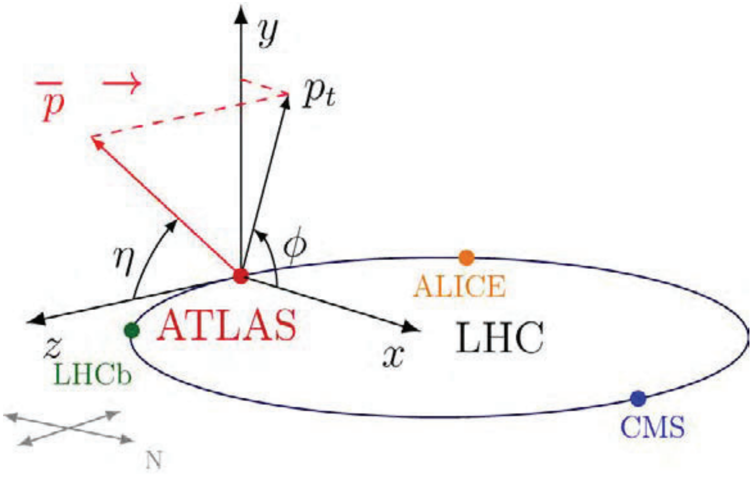


Figure 2.5: The coordinate system of the ATLAS detector with respect to the LHC. The image is taken from Ref. [21].

2.2.2 Layout of the ATLAS Detector

The ATLAS detector is built of different layers of detector systems around the collision point. From the inside to the outside, the inner detector, the electromagnetic calorimeter, the hadronic calorimeter, and the muon chamber follow one another. The choice of these systems and their sequence was made to ensure the optimal reconstruction of electrons, muons, photons, and jets. The description of the ATLAS detector system is based on Ref. [22], [23], and [24].

Inner Detector

After the proton collision within the beam pipeline the emerging particles enter the inner detector. This detector part has a length of 5.3 m and a diameter of 2.5 m covering the pseudo-rapidity range of $|\eta| < 2.5$. The inner detector measures the track of charged particles and consists of three layers of subdetectors. Closest to the beam pipeline, four silicon pixel layers are installed. The innermost layer is the Insertable B-Layer (IBL), which was installed for Run 2 to maintain the required performance for the higher collision rate. The IBL is followed by the B-Layers, Layer 1 and Layer 2, which were already in operation during Run 1. The pixels have a size of $50\,\mu\text{m} \times 250\,\mu\text{m}$ for the IBL and $50\,\mu\text{m} \times 400\,\mu\text{m}$ for the outer pixel layers, thus offering very high granularity. This is crucial to achieve high vertex precision and enable the reconstruction of secondary vertices. The pixel detectors are followed by the semiconductor tracker, which consists of four layers of silicon microstrip trackers supporting the precision tracking of the pixel detector. The outermost subdetector is the transition radiation tracker. This tracker consists of layers of 4 mm diameter straw tubes filled with gas. These layers of tubes allow the continuous tracking of charged particles, improving pattern recognition, that is crucial for electron identification. A homogeneous magnetic field is generated in track detectors to reconstruct the momentum of charged particles. This magnetic field is provided by a system of superconducting magnets.

Electromagnetic Calorimeter

The inner detector is followed by the electromagnetic calorimeter. This calorimeter is a liquid argon detector using lead plates as absorber material with a high showering rate for electrons, positrons, and photons via pair production and bremsstrahlung. The barrel region is covered in the range $|\eta| < 1.475$ with a 4 mm gap at $\eta = 0$. The end-caps are covered on both sides by an inner wheel at $2.5 < |\eta| < 3.2$ and an outer wheel at $1.375 \leq |\eta| \leq 3.2$. The kapton electrodes of the electromagnetic calorimeter are arranged in an accordion geometry. This geometry provides an optimal azimuth coverage in ϕ. To reconstruct the full energy of electrons and photons, the electromagnetic calorimeter has a thickness of about 24 radiation lengths in the barrel region and 26 radiation lengths at the end-caps. Within a radiation length, the kinematic energy of the electrons is reduced on average by a factor of $1/e$. Therefore, electrons and photons are expected to store their full energy within this calorimeter.

Hadronic Calorimeter

The hadronic calorimeter encloses the electromagnetic calorimeter to measure the energy of jets. The hadronic calorimeter consists of the tile calorimeter, the end-cap calorimeter, and the forward calorimeter. The tile calorimeter barrel covers $|\eta| < 1.0$ and is extended by two barrels at $0.8 < |\eta| < 1.7$. The tile calorimeter is divided into 64 modules of scintillator tiles with steel as absorber material. The scintillator layers add up to a thickness of approximately 7.4 interaction lengths. The interaction length is the mean distance between two inelastic interactions of the hadron. The end-cap calorimeter has a geometrical coverage of $1.5 < |\eta| < 1.7$ and thus overlaps with the tile and forward calorimeter. Each end-cap is covered by a 25 mm thick inner wheel and a 50 mm outer wheel. The wheels are made of copper plates with liquid argon gaps for the active material. The forward calorimeter covers the end-caps close to the beam axis. On each end-cap, this liquid argon calorimeter consists of one copper module for electromagnetic measurements followed by tungsten modules for hadronic interactions. The high-density design of the forward calorimeter reaches a thickness of about ten interaction lengths to minimize punch-through effects into the muon system.

Muon Chamber

The only detectable particles that escape the hadronic calorimeter are muons. These are measured in the muon chamber, which makes up a major part of the volume of the ATLAS detector. A very precise measurement of the muon tracks is provided by the monitored drift tubes. Approximately 12000 precision-mounted alignment sensors monitor their deformation and relative position. In the region of $2 < |\eta| < 2.7$, cathode strip chambers with higher granularity are installed. Air-cone toroid magnets provide a strong magnetic field to reconstruct the muon momentum. Eight large magnets are installed in the barrel region of the muon chamber, and the additional magnet systems at the end caps also consist of eight coils each.

2.2.3 Object Reconstruction

The presented structure of the ATLAS detector allows a sophisticated object reconstruction. The following sections provide a brief overview of the reconstruction of electrons, muons, and jets, all of which are relevant to the polarization measurement in this thesis. Therefore, the reconstruction of these objects enables the object and event selection in chapter 6 to define a kinematic region for polarization measurement.

Electron Reconstruction

The electron reconstruction is described in detail in Ref. [25]. An electron passing the inner detector is measured by its hits in the tracking layers. These hits are assembled into clusters. The clusters in the different tracking layers are combined via pattern recognition, ambiguity resolution, and an extension by the transition radiation tracker. A track candidate is considered if it has a transverse momentum above 400 MeV. An optimized Gaussian-sum filter is applied to the clusters to account for energy losses. For an electron candidate, the resulting track in the inner detector has to be matched to a cluster in the electromagnetic calorimeter (EM cluster). This EM cluster is constructed by summing the energies collected in the three calorimeter layers. A sliding-window algorithm is used to seed electromagnetic cluster candidates for the energy deposits. This seeding process is repeated for all elements of the calorimeter and overlaps are removed by selecting the clusters with higher transverse energy. These clusters are then matched with the track candidates in the inner detector. At least one track has to be matched to the electromagnetic cluster to be reconstructed as electron. In the case of several matching tracks, a selection algorithm is performed based on the distance in η and ϕ to the cluster and the number of hits in the silicon layers.

Since also photons lead to an electromagnetic shower in the electromagnetic calorimeter, their reconstruction also relies on the EM cluster algorithm. The electromagnetic cluster of a photon can either be assigned to a conversion vertex or has no associated track in the inner detector. More details about the different reconstruction of electrons and photons are provided by Ref. [26].

Muon Reconstruction

As described in Ref. [27], muons are reconstructed by their hits in the inner detector and the muon chamber combined with information from the calorimeters. Hits in the muon chamber are combined into preliminary track candidates with parabolic trajectory due to the magnet field in the muon system. Based on this preliminary track, a global χ^2 fit removes outliers and includes hits not considered. Overlaps between tracks are solved by removing lower-quality tracks. Including information from the rest of the ATLAS detector five different types of muon reconstructions are used:

- **Combined (CB)**
  For CB muons, a combined track fit is performed for the hits in the inner detector and the muon chambers. Additionally, also the energy loss in the calorimeters is taken into account.

- **Inside-out combined (IO)**
  IO muons are reconstructed by extrapolating tracks in the inner detector to match them with at least three hits in the muon chamber.

- **Muon-spectrometer extrapolated (ME)**
  ME muons cannot be matched to any track in the inner detector and their track is thus extrapolated to the beamline.

- **Segment-tagged (ST)**
  If a track in the inner detector can be matched to at least one hit in the muon chamber, an ST muon will be reconstructed. The parameters of ST muons are directly taken from the track in the inner detector.

- **Calorimeter-tagged (CT)**
  For CT muons, no hits in the muon chamber are matched. They are reconstructed by extrapolating the track in the inner detector to energy deposits in the calorimeters. The transverse momentum of the inner detector tracks of CT muons must exceed 5 GeV instead of the 2 GeV required for reconstructing the other muon types.

Jet Reconstruction

Jets consist of a group of particles originating from emitted gluons or quarks. They cause a bundle of tracks in the inner detector and extensive showers in the electromagnetic and hadronic calorimeter. To cluster the signature of individual particles to a jet, the anti-k_t algorithm [28] is used. This algorithm was introduced as an alternative to the k_t and Cambridge/Aachen inclusive jet finding algorithm and defines the default jet reconstruction in the ATLAS experiment. The anti-k_t algorithm combines entities measured in the detector according to their minimal relative distance. By repeating this procedure until no entity is left to combine the full jet of particles is reconstructed. For the reconstruction of jets used in this analysis, a radius parameter of $R = 0.4$ is chosen for the anti-k_t algorithm. For Run 2 analyses, the resulting jet reconstruction is improved by the particle-flow (`Pflow`) algorithm [29]. The `Pflow` algorithm combines information from the calorimeters with the track information from the inner detector. This algorithm updates the energy deposition in the calorimeters by charged particles with the more precise momentum reconstruction in the inner detector. Thus, the energy and angular resolution, reconstruction efficiency, and pile-up stability are significantly improved [30].

2.2.4 Trigger System

With a bunch crossing every 25 ns, the ATLAS detector measures collisions at a rate of 40 MHz. Since this rate exceeds the technical possibilities of data storage, the measured data is processed by a trigger chain. This trigger chain consists of the Levl-1 trigger and the High-Level trigger and reduces the mean rate of recorded data down to 1.2 kHz. The trigger chain is designed to significantly reduce the data rate with a high efficiency of selecting interesting events. The following summary of the trigger system is based on the detailed descriptions in Ref. [31].

Level-1 (L1) Trigger L1 is a hardware-based trigger system using information from the calorimeters and muon chamber with reduced granularity. The Cluster Processor of the L1 trigger identifies electrons, photons, and τ candidates above a kinematic threshold. The jet candidates and the missing transverse momentum are selected by the Jet/Energy-sum Processor. Their output is fed into the central trigger processor, which decides on the event selection based on the output of the various calorimeter trigger systems. This selection is made according to the kinematics of individual objects, the overall event kinematics, and topological requirements. The resulting event selection reduces the event rate down to the maximum ATLAS detector red-out rate of 100 kHz. In addition to the event selection, the L1 trigger system also identifies Regions-of-Interest (ROI). These ROI are selected in the spherical geometry of ϕ and η and are investigated by the next trigger stage.

High-Level Trigger (HLT) The software-based HLT is the second stage of the trigger system. This trigger system is executed on a computing farm of 40000 Processing Units designed for decisions within a few hundred milliseconds. The decision chain consists of an early event rejection in a fast trigger algorithm and a second rejection step that uses CPU-intensive reconstruction algorithms. This decision chain uses the kinematic information in the ROI chosen by the L1 trigger in combination with a reconstruction algorithm of features measured in the overall ATLAS detector. The HLT is mainly based on the offline software Athena [32] and reduces the event rate to 1.2 kHz. This rate corresponds to a data rate of 1.2 GB/s and is permanently stored for physics analyses.

Chapter 3

Theoretical Foundations

The polarization analysis in this thesis probes the fundamental concepts of the Standard Model. Section 3.1 introduces the concept of spin and helicity that are directly connected to the polarization of a particle. Afterwards, section 3.2 provides a short introduction to quantum field theory with a focus on the interactions of the Standard Model. In section 3.3, the polarization in electroweak same-charged $W^\pm W^\pm$ scattering is explained, and its importance within the Standard Model is highlighted.

3.1 Spins in Relativistic Theory

The mathematical description of spins is based on group theory. The following introduction of the spin is based on the detailed explanation of group theory in Ref. [33]. This section summarizes the Lorentz Group, the corresponding spin representations, and the resulting helicity which is the quantity studied in this analysis.

3.1.1 Lorentz Group and Lorentz Algebra

The Lorentz group is a SO(3,1) Lie group of the Lorentz transformations in special relativity. The description of special relativity is based on the Minkowski metric $\eta_{\mu\nu}$ defining the invariant line element

$$\begin{aligned} ds^2 &= \eta_{\mu\nu} dx^\mu dx^\nu \\ &= dt^2 - dx^2 - dy^2 - dz^2 \end{aligned} \tag{3.1}$$

with $x^0 = t$, $x^1 = x$, $x^2 = y$, and $x^3 = z$ in cartesian coordinates. The values of the covariant and contravariant Minkowski metric are $\eta_{00} = \eta^{00} = 1$, $\eta_{ii} = \eta^{ii} = -1$ for $i = 1, 2, 3$, and $\eta_{ij} = \eta^{ij} = 0$ for $i \neq j$. This line element ds^2 is invariant under Lorentz transformations Λ^μ_ν applied on the coordinates. Thus, the Minkowski metric does not change under Lorentz transformation

$$\eta_{\alpha\beta} = \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta \eta_{\mu\nu}. \tag{3.2}$$

The Lorentz transformation is introduced by the infinitesimal transformation $\delta^\mu_\nu + A^\mu_\nu$. Applying this infinitesimal transformation to the metric tensor

$$\begin{aligned} \eta_{\alpha\beta} &= (\delta^\mu_\alpha + A^\mu{}_\alpha)(\delta^\nu{}_\beta + A^\nu{}_\beta)\eta_{\mu\nu} \\ &= \eta_{\alpha\beta} + A^\mu{}_\alpha \eta_{\mu\beta} + A^\nu{}_\beta \eta_{\alpha\nu} + \mathcal{O}(A^\mu{}_\alpha A^\nu{}_\beta) \end{aligned} \tag{3.3}$$

requires $A^\mu{}_\alpha \eta_{\mu\beta} = -A^\nu{}_\beta \eta_{\alpha\nu}$. This requirement leads to a fully off-diagonal A^μ_ν with six independent generators:

$$A^0{}_i = A^i{}_0 \qquad \text{for}\, i = 1, 2, 3 \tag{3.4}$$

$$A^i{}_j = -A^j{}_i \qquad \text{for}\, i, j = 1, 2, 3\, \text{and}\, i \neq j \tag{3.5}$$

The antisymmetric spatial components $A^i{}_j$ represent spatial rotations J_a in three dimensions. The symmetric mixed time-space components $A^0{}_i$ and $A^i{}_0$ correspond to the three orthogonal

Lorentz boosts K_a which describe a linear relative motion. Therefore, an arbitrary Lorentz transformation

$$\Lambda = e^{i \sum_i \theta_i J_i + i \sum_j \eta_j K_j}. \tag{3.6}$$

is the combined application of the six group generators J_i and K_j and describes rotations of the angles θ_i and Lorentz boosts with the rapidities η_j. The corresponding commutators

$$[J_i, J_j] = i \sum_k \epsilon^{ijk} J_k, \tag{3.7}$$

$$[J_i, K_j] = i \sum_k \epsilon^{ijk} K_k, \tag{3.8}$$

and

$$[K_i, K_j] = -i \sum_k \epsilon^{ijk} J_k \tag{3.9}$$

of these group generators form the Lorentz algebra.

3.1.2 Representations of the Lorentz Group

In the previous section, the SO(3,1) Lorentz algebra was introduced with the generators of rotations J_i and Lorentz boosts K_i. These generators do not commute with each other. However, the commutation properties can be simplified by introducing a new set of generators

$$J_i^{\pm} = \frac{1}{2}(J_i \pm iK_i) \tag{3.10}$$

of the SO(3,1). The resulting commutators

$$[J_i^+, J_j^+] = i \sum_k \epsilon^{ijk} J_k^+, \tag{3.11}$$

$$[J_i^-, J_j^-] = i \sum_k \epsilon^{ijk} J_k^-, \tag{3.12}$$

and

$$[J_i^+, J_j^-] = 0 \tag{3.13}$$

define two distinct SU(2) algebras. Thus, the complexity of the SO(3,1) Lorentz algebra decomposes into a SU(2) $\otimes$ SU(2) algebra motivating the spinor representations of the Lorentz algebra. The spinor representations of the Lorentz algebra can be labeled by (j^-, j^+) with the half-integers j^- and j^+. The dimension of this representation is given by $(2j^- + 1)(2j^+ + 1)$ resulting from the two independent SU(2) algebras. In the following, three special spinor representations are discussed which correspond to different particle spin states in the Standard Model of particle physics.

The Scalar $(0, 0)$ Representation

In this representation, the generators J_i^- and J_i^+ are zero and thus the rotation and Lorentz boost generators are zero. This representation is a one-dimensional scalar which is not affected by rotations and Lorentz boosts [34].

The $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$ Representations

The spinorial representations $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$ are the two-dimensional representations of spin-$\frac{1}{2}$ particle states. The group elements in this representation are called left-handed Weyl spinors ψ_L and right-handed Weyl spinors ψ_R with

$$\psi_L \in \left(\frac{1}{2}, 0\right), \qquad \psi_R \in \left(0, \frac{1}{2}\right). \tag{3.14}$$

The Vectorial $(\frac{1}{2},\frac{1}{2})$ Representation

The four-dimensional $(\frac{1}{2},\frac{1}{2})$ representation corresponds to spin-1 particles. One can show that the four-dimensional elements of the $(\frac{1}{2},\frac{1}{2})$ representation can be interpreted as complex four-vectors V_μ affected by the Lorentz transformation. More details about the interpretation of the individual components of the four-vector V_μ are presented in section 3.2.1.

3.1.3 Spin and Helicity in the Poincare Algebra

For the description of particle states, the Lorentz transformation has to be extended by translations

$$x'^\mu = \Lambda^\mu_{\ \nu} x^\nu + a^\mu. \tag{3.15}$$

The resulting Poincare algebra provides the mathematical basis for describing the particles. The particle properties are defined via Casimir operators which commute with all generators $A_{\mu\nu}$ of the Poincare group. The product of the momentum operators P^μ leads to the Casimir operator $P_\mu P^\mu$. The application on a particle state $P_\mu P^\mu \left|p\right\rangle = m^2 \left|\vec{p}, m\right\rangle$ returns the mass m of the particle. Since the Casimir operator commutes with $A_{\mu\nu}$, the mass of the particle is Lorentz invariant. The second Casimir operator of the Poincare group $W_\mu W^\mu$ results from the Pauli-Lubanski operator

$$W_\sigma = -\frac{1}{2}\epsilon_{\mu\nu\rho\sigma} A^{\mu\nu} P^\rho\,, \tag{3.16}$$

where $\epsilon_{\mu\nu\rho\sigma}$ is the four-dimensional extension of the Levi-Civita symbol. For a particle state $\left|\vec{p}, m, s\right\rangle$ with spin s, the application of the Casimir operator $W_\mu W^\mu \left|\vec{p}, m, s\right\rangle = m^2 s(s+1)\left|\vec{p}, m, s\right\rangle$ shows the Lorentz invariance of the particle spin s. Thus, the particle mass and spin are fundamental particle properties in the Poincare algebra that do not depend on the reference frame.

Massive particles can have $(2s+1)$ spatial spin states $\vec{s}$. Since these spatial spin states are not Lorentz invariant, their definition relies on the choice of a reference frame. The most commonly used spin basis are the canonical and helicity basis. This thesis focuses on the helicity basis: The reference system rotated in the direction of the particle momentum and the system is boosted into the rest system of the particle [35]. The resulting spin states in the helicity basis are defined by the eigenstates of the helicity operator

$$h = \frac{\vec{p}}{|\vec{p}|}\cdot\vec{s}. \tag{3.17}$$

The helicity is the projection of a particle spin on its direction of motion. The possible eigenstate values of the helicity are $h = 0, \pm\frac{1}{2}, \pm 1, ...$ Particles with a positive helicity are called right-handed, and particles with a negative helicity are called left-handed. The concept of helicity is directly connected to the definition of polarization. The polarization of vector bosons measured in this thesis is further explained in section 3.2.1. Since helicity and polarization are not Lorentz invariant, the choice of the reference frame is discussed in section 3.3.4.

3.2 Quantum Field Theory

The previous section introduces the concept of scalars, Weyl spinors, and vector representations. These representations of the Lorentz algebra are directly associated with particles and their spin property. Their kinematics in quantum field theory is described by the minimization of the resulting action $S = \int_{t_1}^{t_2} \mathcal{L} dt$ of the defined Lagrangian density $\mathcal{L}$. For a field ϕ, the Lagrangian has to fulfill the Euler-Lagrange equation

$$\frac{\partial\mathcal{L}}{\partial\phi} - \partial_\mu \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} = 0 \tag{3.18}$$

to give a stationary action $S = \int_{t_1}^{t_2} \mathcal{L} dt$.

The following introduction into the Standard Model und the corresponding Higgs mechanism is based on Ref. [36] and Ref. [37].

3.2.1 Lagrange Density of Free Particles

Free Scalar Particles

The relativistic motion of a free spin-0 particle is described by the Lagrangian

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu \phi \partial^\mu \phi^* - m^2 \phi \phi^*\right) , \tag{3.19}$$

of the complex scalar field $\phi(\vec{r}, t)$ where m corresponds to the rest mass of the particle. The complex conjugate ϕ^* ensures real values for the Lagrange density. Since this Lagrangian contains only Lorentz scalars, it remains invariant under Lorentz transformations and the resulting kinematics is thus in agreement with special relativity. The kinematic of the scalar field can be derived by inserting the partial derivatives

$$\frac{\partial \mathcal{L}}{\partial \phi^*} = -m^2 \phi \qquad \text{and} \qquad \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} = \partial^\mu \phi \tag{3.20}$$

into the Euler-Lagrange equation. Thus, the field has to fulfill the wave equation

$$\partial_\mu \partial^\mu \phi - m^2 \phi = 0 \tag{3.21}$$

to minimize the action S. This wave equation is an analogy to the energy-momentum-relation $E^2 = p^2 + m^2$ of special relativity with $E = i\partial_0$ and $\vec{p} = -i\vec{\nabla}$.

Free Spin-$\frac{1}{2}$ Particles

A spin-$\frac{1}{2}$ particle is given by the complex four-component Dirac spinor ψ. This spinor decomposes into the left- and right-handed Weyl spinors ψ_L and ψ_R introduced in section 3.1.2. To ensure a real scalar Lagrange density, the adjoint spinor

$$\bar{\psi} = \psi^\dagger \gamma^0 \tag{3.22}$$

is introduced by the Hermitian conjugate of the Dirac spinor and the time-like component of the gamma matrices γ^μ. The gamma matrices are 4x4 matrices based on the Pauli matrices:

$$\gamma^0 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}, \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} \quad (i = 1, 2, 3), \quad \gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix} \tag{3.23}$$

The kinematic of the free spin-$\frac{1}{2}$ particle is given by the Lagrangian

$$\mathcal{L} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi. \tag{3.24}$$

By substituting the partial derivatives

$$\frac{\partial \mathcal{L}}{\partial \bar{\psi}} = i\gamma^\mu \partial_\mu \psi - m\psi \qquad \text{and} \qquad \frac{\partial \mathcal{L}}{\partial(\partial_\mu \bar{\psi})} = 0 \tag{3.25}$$

into the Euler-Lagrange equation, one can directly derive the Dirac equation

$$i\gamma^\mu \partial_\mu \psi - m\psi = 0. \tag{3.26}$$

The Dirac equation is solved by the Dirac-spinor wave functions

$$\psi_1 = \sqrt{E+m} \begin{pmatrix} 1 \\ 0 \\ \frac{p_z}{E+m} \\ \frac{p_x + ip_y}{E+m} \end{pmatrix} e^{i(\vec{p}\vec{x} - Et)} \qquad \text{and} \qquad \psi_2 = \sqrt{E+m} \begin{pmatrix} 0 \\ 1 \\ \frac{p_x - ip_y}{E+m} \\ \frac{-p_z}{E+m} \end{pmatrix} e^{i(\vec{p}\vec{x} - Et)} \tag{3.27}$$

for spin-$\frac{1}{2}$ particles and

$$\psi_3 = \sqrt{E+m} \begin{pmatrix} \frac{p_z}{E+m} \\ \frac{p_x + ip_y}{E+m} \\ 1 \\ 0 \end{pmatrix} e^{-i(\vec{p}\vec{x} - Et)} \qquad \text{and} \qquad \psi_2 = \sqrt{E+m} \begin{pmatrix} \frac{p_x - ip_y}{E+m} \\ \frac{-p_z}{E+m} \\ 0 \\ 1 \end{pmatrix} e^{-i(\vec{p}\vec{x} - Et)} \tag{3.28}$$

for spin-$\frac{1}{2}$ anti-particles. The distinction between particles and antiparticles arises from their corresponding energy. While spin-$\frac{1}{2}$ particles possess a positive energy, spin-$\frac{1}{2}$ antiparticles correspond to solutions of the Dirac equation with a negative energy. In the Feynman-Stückelberg interpretation in equation (3.28), these negative energy particles move backward in time, thus representing antiparticles with positive energy propagating forward in time. The antiparticles have the same properties, like mass and spin, as the corresponding particles, except their charges have opposite signs.

Free Spin-1 Particles

The kinematic of a free spin-1 field V_μ is given by the Lagrangian

$$\mathcal{L} = -\frac{1}{4}G_{\mu\nu}G^{\mu\nu} + \frac{1}{2}m^2 V_\mu V^\mu \tag{3.29}$$

with the field strength tensor

$$G_{\mu\nu} = \partial_\mu V_\nu - \partial_\nu V_\mu . \tag{3.30}$$

From this Lagrangian, the equations of motion

$$\left(\partial_\mu \partial^\mu + m^2\right) V^\nu = 0 \qquad \text{and} \qquad \partial_\mu V^\mu = 0 \tag{3.31}$$

can be derived. These equations are solved by a set of plane waves [38]:

$$V^\mu(\vec{x}) = \int \frac{d^3\vec{p}}{(2\pi)^3}\frac{1}{2E}\sum_{\lambda=0}^{2}\left(\epsilon^\mu_\lambda(E,\vec{p})a_\lambda(E,\vec{p})e^{-ip_\nu x^\nu} + \epsilon^{*\mu}_\lambda(E,\vec{p})a^\dagger_\lambda(E,\vec{p})e^{ip_\nu x^\nu}\right) \tag{3.32}$$

a and $a^\dagger$ are the bosonic annihilation and creation operators, and ϵ^μ_λ represent the different polarization states of the field. Massive and massless spin-1 fields have two transvere polarizations $\vec{\epsilon}_1(E,\vec{p})$ and $\vec{\epsilon}_2(E,\vec{p})$. $\vec{\epsilon}_1(E,\vec{p})$ is an arbitrary vector $\hat{\vec{p}}_\perp$ orthogonal to the direction of motion and thus $\vec{\epsilon}_2(E,\vec{p})$ is given by the vector $\hat{\vec{p}} \times \hat{\vec{p}}_\perp$. In the case of a motion in z direction, one can choose the explicit form $\epsilon^\mu_1 = (0,1,0,0)$ and $\epsilon^\mu_2 = (0,0,1,0)$. From these two linear polarization states the circular polarizations

$$\epsilon^\mu_\pm = \frac{1}{\sqrt{2}}\left(\epsilon^\mu_1 \pm i\epsilon^\mu_2\right) \tag{3.33}$$

can be derived. The left- and right-handed circular polarizations ϵ^μ_- and ϵ^μ_+ correspond to the helicity states. For massive spin-1 fields the additional longitudinal polarization $\epsilon^\mu_0(\vec{p})$ arises. If the direction of motion aligns with the z axis, the longitudinal polarization can be written as

$$\epsilon^\mu_0(\vec{p}) = \frac{1}{m}(|\vec{p}|,0,0,E). \tag{3.34}$$

Since the longitudinal polarization state is proportional to $\frac{p^\mu}{m}$, it becomes dominant at high energies. The sum of the different polarization states fulfills the completeness relation:

$$\sum_{\lambda=0}^{2}\epsilon^\mu_\lambda(\vec{p})\epsilon^{\nu*}_\lambda(\vec{p}) = -g^{\mu\nu} + \frac{1}{m^2}p^\mu p^\nu \tag{3.35}$$

The momentum space propagator

$$i\frac{-g^{\mu\nu} + \frac{1}{m^2}p^\mu p^\nu}{p_\alpha p^\alpha - m^2 + i\epsilon} \tag{3.36}$$

is associated with the internal spin-1 V_μ line in Feynman diagrams [38].

If not stated otherwise, the indices L and T in this thesis refer to the longitudinal and transverse polarization states.

3.2.2 U(1) Symmetry of Quantum Electrodynamics

So far, the kinematic of free particles was discussed. The interaction between the particles is introduced by gauge symmetries. Symmetries play a central role in the description of physics. According to Noether theorem each continuous symmetry results in a conservation law [39]. Prominent examples are the continuous time, translation, and rotation symmetry, which imply the conservation of energy, momentum, and angular momentum. Therefore, such symmetries are deeply connected with the most fundamental laws of physics.

These existing symmetries motivate the introduction of other laws of physics as consequences of corresponding gauge symmetries. A simple example of a local gauge symmetry is the local rotation in the complex plane by an angle $\alpha(\vec{r}, t)$. For a Dirac field ψ of a spin-$\frac{1}{2}$ particle, this means the transformation $\psi \to \psi' = e^{iq\alpha}\psi$ with a position- and time-dependent α. However, this transformation applied in the spin-$\frac{1}{2}$ Lagrangian density leads to

$$\begin{aligned}\mathcal{L}' &= i\bar{\psi}'\gamma^\mu\partial_\mu\psi' - m\bar{\psi}'\psi' \\ &= ie^{-iq\alpha}\bar{\psi}\gamma^\mu\partial_\mu(e^{iq\alpha}\psi) - me^{-iq\alpha}\bar{\psi}e^{iq\alpha}\psi \\ &= i\bar{\psi}\gamma^\mu\left(\partial_\mu + iq\partial_\mu\alpha\right)\psi - m\bar{\psi}\psi \\ &\neq \mathcal{L}\end{aligned} \tag{3.37}$$

and thus requires additional modifications to the Lagrangian density. The derivative ∂_μ is extended by an additional field B_μ to the covariant derivative $D_\mu = \partial_\mu + iqB_\mu$ to establish this invariance. Here q represents a constant which, as shown below, corresponds to the electrical charge of the spin-$\frac{1}{2}$ particle. If the field B_μ behaves under U(1) transformation like

$$B_\mu \to B'_\mu = B_\mu - \partial_\mu\alpha, \tag{3.38}$$

the Lagrange density

$$\begin{aligned}\mathcal{L}' &= i\bar{\psi}'\gamma^\mu D'_\mu\psi' - m\bar{\psi}'\psi' \\ &= i\bar{\psi}'\gamma^\mu\left(\partial_\mu + iqB'_\mu\right)\psi' - m\bar{\psi}'\psi' \\ &= ie^{-iq\alpha}\bar{\psi}\gamma^\mu\left(\partial_\mu + iqB_\mu - iq\partial_\mu\alpha\right)(e^{iq}\psi) - me^{-iq\alpha}\bar{\psi}e^{iq\alpha}\psi \\ &= i\bar{\psi}\gamma^\mu\left(\partial_\mu + iqB_\mu\right)\psi - m\bar{\psi}\psi \\ &= \mathcal{L}\end{aligned} \tag{3.39}$$

shows the desired additional gauge invariance. Thus, this U(1) invariance could only be guaranteed by introducing a new field B_μ. The kinematics of this new vector field must now be introduced into the Lagrangian

$$\mathcal{L} = i\bar{\psi}\gamma^\mu D_\mu\psi - m\bar{\psi}\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{3.40}$$

by the field strength tensor

$$F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu\,. \tag{3.41}$$

This Lagrangian provides a complete description of a complex spin-$\frac{1}{2}$ particle ψ under U(1) invariance and the corresponding vector field.

If one compares field B_μ, which was added to ensure the local gauge invariance, with classical physical theories, it becomes apparent that this must be the electromagnetic field. The field strength tensor in equation (3.41) can also be found in Maxwell's equations and the required transformation in equation (3.38) also corresponds to Maxwell's electromagnetism. Thus, the field B_μ can be identified with the electromagnetic field, where the constant q corresponds to the electric charge e. This electric charge corresponds thereby to the conservation quantity which follows according to Noether's theorem from the introduced continuous symmetry.

Hence, the Lagrangian density in equation (3.40) represents a theory of quantum electrodynamics (QED) for a complex spin-$\frac{1}{2}$ particle ψ with the electric charge q and an electromagnetic vector field B_μ, which is associated with the photon.

3.2.3 SU(3) Symmetry of Quantum Chromo Dynamics

The previous section showed that electromagnetism can be derived by introducing a U(1) gauge symmetry into the Lagrangian density. The same strategy of local gauge invariances is used to introduce the other interactions known from modern particle physics. In order to explain why quarks form stable protons and neutrons, the strong interaction of quantum chromo dynamics (QCD) is introduced. The quarks get three additional degrees of freedom represented by the three color charges: red, blue, and green. A SU(3) gauge invariance is introduced to describe this new interaction with three charge states. The generators of this SU(3) gauge group correspond to 3×3 matrices. The SU(3) has eight generators which can be represented by the Gell-Mann matrices

$$\begin{aligned} \lambda^1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda^2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\ \lambda^4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \lambda^5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad \lambda^6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \\ \lambda^7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \lambda^8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \end{aligned} \tag{3.42}$$

The algebra defined by the Gell-Mann matrices is given by the commutation relation:

$$[\lambda^i, \lambda^j] = 2i \sum_k f^{ijk} \lambda^k, \tag{3.43}$$

with the structure constant f^{ijk} of the SU(3). In SU(3), the Dirac field ψ with color charge states is transformed by

$$\psi \to \psi' = e^{ig_S \sum_{j=1}^{8} \alpha_j \frac{1}{2} \lambda^j} \psi \tag{3.44}$$

with eight position- and time-dependent parameters α_j.

The invariance of the Lagrangian density $\mathcal{L}$ is ensured by covariant derivative $D_\mu = \partial_\mu + ig_S \sum_{j=1}^{8} G_\mu^j \frac{\lambda^j}{2}$ which consists of the sum over eight fields G_μ^j. Similar to equation (3.39) the derivative has to fulfill

$$\left(\partial_\mu + ig_S \sum_j^8 G_\mu^{j\prime} \frac{\lambda^j}{2} \right) \left(e^{ig_S \sum_{k=1}^{8} \alpha_k \frac{1}{2} \lambda^k} \psi \right) = e^{ig_S \sum_{k=1}^{8} \alpha_k \frac{1}{2} \lambda^k} \left(\partial_\mu + ig_S \sum_j^8 G_\mu^j \frac{\lambda^j}{2} \right) \psi \tag{3.45}$$

to enable the SU(3) invariance. This requirement results in the vector field transformation:

$$G_\mu^{j\prime} = G_\mu^j - \partial_\mu \alpha_j - g_S \sum_{lk}^8 f^{lkj} G_\mu^l \alpha_k \tag{3.46}$$

The derivatives of the functions α_j are similar to the transformation in the U(1) symmetry, but the second term is a direct consequence of the non-commutating Gell-Mann matrices. The eight fields G_μ^j represent the gluons transmitting the strong interaction. Due to the non-commutating Gell-Mann matrices, the SU(3) of QCD is a nonabelian gauge theory with self-interactions of the gluons.

The kinematics of a spin-$\frac{1}{2}$ particle ψ in QCD is determined by the Lagrangian

$$\mathcal{L} = i\bar{\psi}\gamma^\mu D_\mu \psi - m\bar{\psi}\psi - \frac{1}{4} \sum_j^8 G_{\mu\nu}^{(j)} G^{\mu\nu\,(j)} \tag{3.47}$$

with the field strength tensor [40]

$$G_{\mu\nu}^{(j)} = \partial_\mu G_\nu^j - \partial_\nu G_\mu^j + g_S \sum_k \sum_l f^{jkl} G_\mu^k G_\nu^l. \tag{3.48}$$

The last term of the field tensor leads to products of three and four vector boson fields in the Lagrangian. The products correspond to the triple and quartic self-coupling of the gluons.

3.2.4 SU(2) Symmetry of Weak Interaction

An invariance that has long been assumed for the laws of physics is parity. Parity is the spatial inversion $\vec{x} \to -\vec{x}$. The QED and QCD described in the previous sections are indeed invariant under this transformation. In 1957, however, Wu and her collaborators were able to show that this does not apply to all of physics. They studied the decay of polarized ^{60}Co in a magnetic field. The magnetic moment of the cobalt nuclei was aligned with the strong external magnetic field and the direction of the electron resulting from the $^{60}\text{Co} \to ^{60}\text{Ni}^* + e^- + \bar{\nu}_e$ decay was measured. Since the polarization and the magnetic field are axial vectors, but the momentum is not, only the momentum changes under parity transformation. Thus, if parity invariance applies to this decay process a specific momentum direction of the resulting electron would be as likely as its opposite direction. However, this was not the case, so Wu and her collaborators showed that the weak interaction involved was not invariant under parity transformation.[37]

Therefore, the gauge invariance on which the weak interaction is based must reflect this parity violation. The weak interaction acts on weak isospin doublets with the weak eigenstates of the particles. These weak eigenstates of the particles are connected to their mass eigenstates by the CKM mixing matrix for quarks and the PMNS mixing matrix for neutrinos. Due to the parity violation, the weak interaction only acts on left-handed particle doublets (or right-handed antiparticle doublets)

$$L \in \left\{ \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L, \begin{pmatrix} u \\ d' \end{pmatrix}_L, \begin{pmatrix} c \\ s' \end{pmatrix}_L, \begin{pmatrix} t \\ b' \end{pmatrix}_L \right\} \tag{3.49}$$

and does not affect the right-handed particle singlets (or left-handed antiparticle singlets)

$$R \in \left\{ e_R^-, \mu_R^-, \tau_R^-, u_R, d_R, c_R, s_R, t_R, b_R \right\}. \tag{3.50}$$

The left-handed doublets are also called isospin doublets. Their corresponding weak charge of left-handed particles is given by $I_3^W = \frac{1}{2}$ for neutrinos and up-type quarks and $I_3^W = -\frac{1}{2}$ for charged leptons and down-like quarks.

The SU(2) gauge theory for the weak interaction provides the transformation of the weak isospin doublets

$$L \to L' = e^{ig_W \sum_{j=1}^{3} \alpha_j \frac{1}{2} \sigma_j} L \tag{3.51}$$

using the Pauli matrices σ_j.

Three gauge fields are introduced into the covariant derivative $D_\mu = \partial_\mu + ig_W \sum_j^3 W_\mu^j \frac{\sigma_j}{2}$ to maintain the gauge invariance of the Lagrangian. Similar to the SU(3) gauge theory in the previous section, the transformation of these fields is given by

$$W_\mu^j \to W_\mu^{j\prime} = W_\mu^j - \partial_\mu \alpha_j - g_W \sum_{lk}^{3} \epsilon^{lkj} W_\mu^l \alpha_k \,, \tag{3.52}$$

but the permutation relation of the Pauli matrices must be used instead of the Gell-Mann matrices. By adding the field strength tensor

$$W_{\mu\nu}^{(j)} = \partial_\mu W_\nu^j - \partial_\nu W_\mu^j + g_W \sum_k \sum_l \epsilon^{jkl} W_\mu^k W_\nu^l \,, \tag{3.53}$$

the Lagrangian

$$\mathcal{L} = i\bar{\psi}\gamma^\mu D_\mu \psi - m\bar{\psi}\psi - \frac{1}{4} \sum_j^3 W_{\mu\nu}^{(j)} W^{\mu\nu\,(j)} \tag{3.54}$$

describes the kinematic of left-handed particle (right-handed antiparticle) doublets, right-handed particle (left-handed antiparticle) singlets, and weak gauge bosons [40].

In the case of the isospin doublet of electron and electron-neutrino, the interaction with the weak field is given by

$$\mathcal{L}_{\text{W-INT}} = i\frac{g_W}{2}\gamma^\mu \begin{pmatrix} \bar{\nu}_L & \bar{e}_L \end{pmatrix} \begin{pmatrix} W_\mu^3 & W_\mu^1 - iW_\mu^2 \\ W_\mu^1 + iW_\mu^2 & -W_\mu^3 \end{pmatrix} \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}. \tag{3.55}$$

Thus, the fields $W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp iW_\mu^2\right)$ connect the electron and muon parts of the doublets in the weak interaction. The resulting Lagrangian of the interaction with the weak fields

$$\begin{aligned}\mathcal{L}_{\text{W-INT}} &= i\frac{g_W}{2}\gamma^\mu \begin{pmatrix}\bar{\nu}_L & \bar{e}_L\end{pmatrix}\begin{pmatrix}W_\mu^3 & \sqrt{2}W_\mu^+ \\ \sqrt{2}W_\mu^- & -W_\mu^3\end{pmatrix}\begin{pmatrix}\nu_L \\ e_L\end{pmatrix} \\ &= ig_W\gamma^\mu\left(\frac{1}{2}\bar{\nu}_L W_\mu^3 \nu_L - \frac{1}{2}\bar{e}_L W_\mu^3 e_L + \frac{1}{\sqrt{2}}\bar{\nu}_L W_\mu^+ e_L + \frac{1}{\sqrt{2}}\bar{e}_L W_\mu^- \nu_L\right)\end{aligned} \tag{3.56}$$

consists of a W^+ boson with a positive electrical charge, a W^- boson with a negative electrical charge, and an electrical neutral W^3 boson.

3.2.5 Electroweak Unification in $\mathbf{U(1)_Y \otimes SU(2)}$

The electrical neutral gauge boson W^3 of the weak interaction introduced in the previous section only couples on left-handed particles (right-handed antiparticles). Experiments have shown the existence of a neutral boson taking part in weak interactions, the Z boson. However, in contradiction to the described W^3 boson, the experimentally found Z boson couples to both left-handed and right-handed particles. This flaw of the theory can be solved by unifying the U(1) symmetry of QED with the SU(2) symmetry of the weak interaction. In the Glashow-Salam-Weinberg (GSW) electroweak model the U(1) of QED is replaced by a $\mathrm{U}(1)_Y$ symmetry

$$\psi \to \psi' = e^{ig'\frac{Y}{2}\alpha}\psi \tag{3.57}$$

with the weak hypercharge $Y = 2(Q + I_3^W)$. The invariance under the new $\mathrm{U}(1)_Y$ symmetry leads to the corresponding covariant derivation $D_\mu = \partial_\mu + ig'\frac{Y}{2}B_\mu$. The gauge field B_μ of $\mathrm{U}(1)_Y$ is combined with the W_μ^3 gauge field of SU(2) to give the photon field

$$A_\mu = B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W \tag{3.58}$$

and the Z field

$$Z_\mu = -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W. \tag{3.59}$$

θ_W is the electroweak mixing angle which will be introduced by the Higgs mechanism in the next section.

3.2.6 Higgs Mechanism

The previous sections derived the Lagrangian of a massive spin-$\frac{1}{2}$ particle assuming local gauge invariance under transformations from different gauge groups. The invariance was achieved by introducing massless gauge fields. For the weak interaction, this conflicts with experiments showing that the W^+, W^-, and Z bosons have a mass. Directly adding mass terms like $\frac{1}{2}m^2 B_\mu B^\mu$ would violate the initial gauge invariance. An additional field, the Higgs field ϕ, is introduced to achieve the particle masses without gauge violation. This Higgs field has a potential $V(\phi)$ and breaks the $\mathrm{U}(1)_Y \otimes \mathrm{SU}(2)$ symmetry. The approach of Ref. [37] is followed where the Higgs mechanism is introduced for the simpler U(1) symmetry, requiring a complex scalar Higgs field. Afterwards, the Standard Model Higgs doublet is explained, breaking the electroweak $\mathrm{U}_Y(1) \otimes \mathrm{SU}(2)$ symmetry.

Higgs Mechanism with U(1) Symmetry Breaking

To explain the concept behind the Higgs mechanism the simple case of the U(1) symmetry of section 3.2.2 is used. This symmetry results in the massless gauge field B_μ. The complex scalar field

$$\phi = \frac{1}{\sqrt{2}}\left(\phi_1 + i\phi_2\right) \tag{3.60}$$

is introduced with two real field components ϕ_1 and ϕ_2. The Higgs field has the potential:

$$V(\phi) = \mu^2\phi\phi^* + \lambda\left(\phi\phi^*\right)^2 \tag{3.61}$$

The parameters μ^2 and λ are free real parameters with $\mu^2 < 0$ and $\lambda > 0$. The shape of this potential is shown in figure 3.1. The kinematic of the scalar ϕ is described by the Lagrangian

$$\mathcal{L} = (D_\mu\phi)^* D^\mu\phi - V(\phi) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{3.62}$$

The interaction with B_μ originates from the covariant derivative $D_\mu = \partial_\mu + iqB_\mu$. The ground state, also called the vacuum state, of the field ϕ fulfills:

$$\phi\phi^* = \frac{1}{2}\left(\phi_1^2 + \phi_2^2\right) = -\frac{\mu^2}{2\lambda} =: v \tag{3.63}$$

with v denoting the vacuum expectation value of the potential $V(\phi)$. By choosing a discrete minimum like $\phi_1 = v$ and $\phi_2 = 0$, the initial U(1) is broken since the groundstate is not invariant under a rotation in complex numbers. The consequences of this symmetry breaking arise when the field is expanded around its ground state. The field around the chosen groundstate is given by $\phi = \frac{1}{\sqrt{2}}(v + \eta + i\xi)$ with the real scalar fields η and ξ. By inserting this parametrization of the field ϕ into the Higgs potential

$$V(\phi) = -\frac{1}{4}\lambda v^4 + \lambda v^2\eta^2 + \lambda v\eta^3 + \frac{1}{4}\lambda\eta^4 + \lambda v\eta\xi^2 + \frac{1}{2}\lambda\eta^2\xi^2 + \frac{1}{4}\lambda\xi^4 \tag{3.64}$$

and the derivative term

$$(D_\mu\phi)^* D^\mu\phi = \frac{1}{2}\left(\partial_\mu - iqB_\mu\right)(v + \eta - i\xi)\left(\partial^\mu + iqB^\mu\right)(v + \eta + i\xi) \tag{3.65}$$

the Lagrangian in equation (3.62) takes the form:

$$\begin{aligned}\mathcal{L} = {} & \underbrace{\frac{1}{2}\partial_\mu\eta\partial^\mu\eta - \lambda v^2\eta^2}_{\text{massive } \eta} + \underbrace{\frac{1}{2}\partial_\mu\xi\partial^\mu\xi}_{\text{massless } \xi} - \underbrace{\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}q^2v^2B_\mu B^\mu}_{\text{massive } B_\mu \text{ field}} \\ & + \frac{1}{4}\lambda v^4 - \underbrace{\lambda v\eta^3 + \lambda v\eta\xi^2}_{\text{triple self-coupling}} - \underbrace{\frac{1}{4}\lambda\eta^4 - \frac{1}{2}\lambda\eta^2\xi^2 - \frac{1}{4}\lambda\xi^4}_{\text{quartic self-coupling}} + \underbrace{qv\partial_\mu\xi B^\mu}_{\text{direct coupling to } B_\mu} \\ & - \underbrace{q\partial_\mu\eta B^\mu\xi + q\partial_\mu\xi B^\mu\eta + q^2vB_\mu B^\mu\eta}_{\text{triple coupling to } B_\mu} + \underbrace{\frac{1}{2}q^2B_\mu B^\mu\eta\eta + \frac{1}{2}q^2B_\mu B^\mu\xi\xi}_{\text{quartic coupling to } B_\mu}\end{aligned} \tag{3.66}$$

As already illustrated in equation (3.66), the Lagrangian can be separated into different terms. The choice of the vacuum state results in a massive real field η, a massless real Goldstone field ξ, and a massive field B_μ. Thus, the U(1) symmetry breaking give the previously massless field B_μ the mass $m_B = qv$ and also the η boson field obtaina the mass $m_\eta = \sqrt{2\lambda v^2}$. The interactions of the complex field ϕ are given by triple and quartic self-couplings, and triple and quartic couplings to the B_μ field. In addition to these triple and quartic couplings, the term $qv\partial_\mu\xi B^\mu$ arises. This term represents a direct coupling between the fields ξ and B_μ. Thus, the spin-1 gauge field B_μ can directly transform into the spin-0 Goldstone field ξ. Furthermore, a problem regarding the degrees of freedom arises. Before the symmetry breaking, the Lagrangian had two spin degrees of freedom in ϕ_1 and ϕ_2 and four spin degrees of freedom in the two transverse polarized states of the spin-1 B_μ. After the U(1) symmetry breaking, the B_μ becomes massive and thus gains a longitudinal polarization state providing an additional degree of freedom. These two problems can be solved by an appropriate U(1) gauge transformation. For small ξ, the complex field becomes in first order

$$\phi = \frac{1}{\sqrt{2}}(v + \eta)e^{i\frac{\xi}{v}}. \tag{3.67}$$

With the transformation

$$\begin{aligned}\phi \to \phi' &= e^{-i\frac{\xi}{v}}\phi \\ &= e^{-i\frac{\xi}{v}}\frac{1}{\sqrt{2}}(v + \eta)e^{i\frac{\xi}{v}} \\ &= \frac{1}{\sqrt{2}}(v + \eta)\end{aligned} \tag{3.68}$$

and

$$B_\mu \to B'_\mu = B_\mu + \frac{1}{gv}\partial_\mu \xi, \tag{3.69}$$

the Goldstone field ξ can be entirely removed from the Lagrangian in equation (3.66). Therefore, the resulting Lagrangian describes a massive scalar field η and a massive gauge field B_μ whose longitudinal component is associated with a Goldstone field.

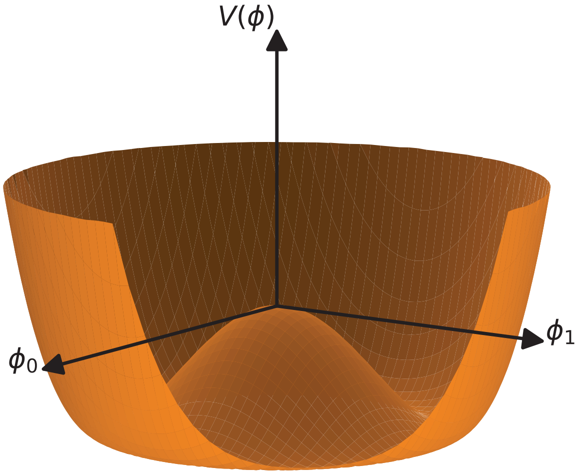


Figure 3.1: The potential $V(\phi) = \mu^2\phi\phi^* + \lambda\,(\phi\phi^*)^2$ of a complex scalar field ϕ breaking the U(1) symmetry.

Standard Model Higgs Mechanism

To introduce masses for the W^+, W^-, and Z boson, the Higgs mechanism is applied for the $\mathrm{U}_Y(1) \otimes \mathrm{SU}(2)$ symmetry. Thus, the Standard Model Higgs field is given by a complex electroweak doublet

$$\phi = \begin{pmatrix} \phi_+ \\ \phi_0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{3.70}$$

As in section 3.2.6, the dynamics of the Higgs field is given by the Lagrangian

$$\mathcal{L} = (D_\mu\phi)^\dagger D^\mu\phi - V(\phi) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{3.71}$$

with the Higgs potential

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda\left(\phi^\dagger\phi\right)^2. \tag{3.72}$$

The minimum $\phi^\dagger\phi = -\frac{\mu^2}{2\lambda}$ of the potential $V(\phi)$ represents the vacuum state. As in section 3.2.6, one specific groundstate is chosen by $\phi_3 = v$ and $\phi_1 = \phi_2 = \phi_4 = 0$ with the vacuum expectation value $v^2 = -\frac{\mu^2}{\lambda}$. The choice of this specific groundstate breaks the electroweak $\mathrm{U}_Y(1) \otimes \mathrm{SU}(2)$ symmetry. For the expansion of the field

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ v + h + i\phi_4 \end{pmatrix} \tag{3.73}$$

around the chosen ground state, the field components ϕ_1, ϕ_2, and ϕ_4 can be removed by a suitable gauge transformation similar to section 3.2.6. The Goldstone bosons associated with the scalar fields ϕ_1, ϕ_2, and ϕ_4 become the longitudinal components of the Z and $W^\pm$ bosons. This unitarity gauge transformation leads to the Higgs doublet

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix} \tag{3.74}$$

of the Salam-Weinberg model. The real lower component $v+h$ of this Higgs doublet has the electric charge $Q=0$ and the weak isospin $I_W^{(3)} = -\frac{1}{2}$. Thus, the weak hypercharge of the Higgs boson is $Y=1$ in the $\mathrm{U}_Y(1) \otimes \mathrm{SU}(2)$ symmetry. The resulting covariant derivative of the Higgs field

$$\begin{aligned} D^\mu \phi &= \begin{pmatrix} \partial_\mu + i\frac{1}{2}g' B_\mu + i\frac{1}{2}g_W W_\mu^3 & i\frac{1}{\sqrt{2}}g_W W_\mu^+ \\ i\frac{1}{\sqrt{2}}g_W W_\mu^- & \partial_\mu + i\frac{1}{2}g' B_\mu - i\frac{1}{2}g_W W_\mu^3 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v+h \end{pmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} i\frac{1}{\sqrt{2}}g_W W_\mu^+ (v+h) \\ \left(\partial_\mu + i\frac{1}{2}g' B_\mu - i\frac{1}{2}g_W W_\mu^3\right)(v+h) \end{pmatrix} \end{aligned} \tag{3.75}$$

introduces the interactions between the Higgs field and the electroweak gauge bosons. In the final Lagrangian, these interactions arise in the term

$$\begin{aligned} (D_\mu \phi)^\dagger D^\mu \phi =& \frac{1}{2}\partial_\mu h \partial^\mu h + \frac{1}{4} g_W^2 W_\mu^- W^{+\mu} (v+h)^2 \\ &+ \frac{1}{8}\left(g_W W_\mu^3 - g' B_\mu\right)\left(g_W W^{3\mu} - g' B^\mu\right)(v+h)^2 \end{aligned} \tag{3.76}$$

that also contains the mass terms of the bosons. The mass $m_W = \frac{1}{2} g_W v$ of the W^+ and W^- bosons is therefore directly proportional to the weak coupling strength g_W and the vacuum expectation value v of the Higgs field. The last term in equation (3.76) introduces a mass term for the fields B_μ and W_μ^3. As motivated in section 3.2.5, the combination of these two fields results in the photon field $A_\mu = B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W$ and the Z boson field $Z_\mu = -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W$ known from experiments. Since it is known from experiments that the photon is massless and the Z boson is massive, one can directly identify the photon field

$$A_\mu = \frac{g_W W_\mu^3 + g' B_\mu}{\sqrt{g_W^2 + g'^2}} \tag{3.77}$$

with $m_A = 0$ and

$$Z_\mu = \frac{g_W W_\mu^3 - g' B_\mu}{\sqrt{g_W^2 + g'^2}} \tag{3.78}$$

with $m_Z = \frac{1}{2} v \sqrt{g_W^2 + g'^2}$. The mass of the Higgs field arises directly from the Higgs potential $V(\phi) = \mu^2 (v+h)^2 + \lambda (v+h)^4$. With the vacuum expectation value $v = -\frac{\mu^2}{\lambda}$, the mass of the Higgs boson is given by $m_H = \sqrt{2\lambda v^2}$. Since this mass depends on the free parameter λ that does not occur in the masses of the electroweak gauge bosons, the value of the Higgs mass is not predicted by theory and has to be measured in experiments.

The Fermion Masses

So far, the masses of the fermions were introduced by simply adding the term $-m\bar{\psi}\psi$. However, the term $\bar{\psi}\psi = \left(\bar{\psi}_R \psi_L + \bar{\psi}_L \psi_R\right)$ is not invariant under the $\mathrm{U}_Y(1) \otimes \mathrm{SU}(2)$ symmetry since the SU(2) transformation only affects the left-handed particles (right-handed antiparticles). This problem can be solved by using the Higgs field ϕ. The Higgs doublet ϕ is combined with the left-handed fermion doublet L. Under an infinitesimal SU(2) transformation, the product $\bar{L}\phi$ transforms as

$$\begin{aligned} \bar{L}\phi \to \bar{L}'\phi' &= \bar{L}\left(I - i g_W \sum_j^3 \epsilon_j \frac{1}{2}\sigma_j\right)\left(I + i g_W \sum_k^3 \epsilon_k \frac{1}{2}\sigma_k\right)\phi \\ &= \bar{L}\phi + \mathcal{O}(\epsilon_l^2) \end{aligned} \tag{3.79}$$

and is therefore invariant under the SU(2) gauge transformation. By multiplying also with the right-handed singlet R, the resulting term $\bar{L}\phi R$ becomes invariant under the $\mathrm{U}_Y(1) \otimes \mathrm{SU}(2)$ gauge transformation. Due to this invariance, the Yukawa term $-g_f\left(\bar{L}\phi R + \bar{R}\phi^\dagger L\right)$ is added to the Lagrangian to generate the fermion masses. In the Salam-Weinberg model, the Yukawa mass term of the weak doublet e.g. of the electron and its neutrino becomes

$$\begin{aligned}\mathcal{L}_{m_e} &= -\frac{g_e}{\sqrt{2}}\left(\begin{pmatrix}\bar{\nu}_e & \bar{e}\end{pmatrix}_L \begin{pmatrix}0 \\ v+h\end{pmatrix} e_R + \bar{e}_R \begin{pmatrix}0 & v+h\end{pmatrix}\begin{pmatrix}\nu_e \\ e\end{pmatrix}_L\right) \\ &= -\frac{g_e}{\sqrt{2}} v\left(\bar{e}_R e_L + \bar{e}_L e_R\right) - \frac{g_e}{\sqrt{2}} h\left(\bar{e}_R e_L + \bar{e}_L e_R\right) \\ &= -\frac{g_e}{\sqrt{2}} v\bar{e}e - \frac{g_e}{\sqrt{2}} h\bar{e}e.\end{aligned} \tag{3.80}$$

This Yukawa coupling term can generate the mass $m_f = \frac{g_f v}{\sqrt{2}}$ for the charged leptons e^-, mu^-, and τ^- and the down-type quarks d', s', and b'. In addition, these particles couple to the Higgs field with the coupling strength g_f proportional to their mass. The neutrinos are massless in the Standard Model, but the masses of the up-type quarks u, c, and t are still missing. To generate masses also in the upper components of the fermion doublets, the conjugate Higgs field

$$\phi_c = -i\sigma_2\phi^* = \begin{pmatrix}-\phi_3 + i\phi_4 \\ \phi_1 - i\phi_2\end{pmatrix} \tag{3.81}$$

is used. The field ϕ_c transforms under SU(2) in the same way as ϕ and thus preserves the gauge invariance in $\bar{L}\phi_c R$. This Yukawa leads to the up-type quark mass terms:

$$\mathcal{L}_{m_u} = -\frac{g_u}{\sqrt{2}} v\bar{u}u - \frac{g_u}{\sqrt{2}} h\bar{u}u \tag{3.82}$$

So, the masses of the up-type quarks and their coupling to the Higgs field are equivalent to the Yukawa coupling of the lower components of the fermion doublets.

3.2.7 Summary of the Standard Model

As presented in the previous sections, the Standard Model of particle physics is a relativistic quantum field theory with a $\mathrm{U}_Y(1) \otimes \mathrm{SU}(2) \otimes \mathrm{SU}(3)$ symmetry. The Standard Model Lagrangian

$$\mathcal{L} = i\bar{\psi}\gamma^\mu D_\mu\psi - \frac{g_f}{\sqrt{2}} v\bar{\psi}\psi - \frac{g_f}{\sqrt{2}} h\bar{\psi}\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + (D_\mu\phi)^\dagger D_\mu\phi - V(\phi) \tag{3.83}$$

describes the kinematics of massive fermions ψ and boson fields. The covariant derivative

$$D_\mu = \partial_\mu + ig'\frac{Y}{2}B_\mu + ig_W\sum_j^3 W_\mu^j\frac{\sigma_j}{2} + ig_S\sum_j^8 G_\mu^j\frac{\lambda^j}{2} \tag{3.84}$$

introduces the interaction with gauge bosons. The strong interaction is transmitted by the gluons G_μ^j coupling to all particles with a color charge. The bosons B_μ and W_μ^j represent the electroweak part of the Standard Model. Their combination gives the physical boson fields of the Photon A_μ for the electromagnetic interaction and the Z, W^+, and W^- boson of the weak interaction. The field tensor

$$\begin{aligned}F_{\mu\nu} =& \partial_\mu B_\nu - \partial_\nu B_\mu \\ &+ \sum_j^3\left(\partial_\mu W_\nu^j - \partial_\nu W_\mu^j + g_W\sum_k\sum_l \epsilon^{jkl}W_\mu^k W_\nu^l\right) \\ &+ \sum_j^8\left(\partial_\mu G_\nu^j - \partial_\nu G_\mu^j + g_S\sum_k\sum_l f^{jkl}G_\mu^k G_\nu^l\right)\end{aligned} \tag{3.85}$$

generates additional self-coupling interactions between specific gauge bosons. The masses of the Standard Model particles are introduced by the Higgs field ϕ connected to the Higgs potential $V(\phi)$. The minimum of this potential breaks the $\mathrm{U}_Y(1) \otimes \mathrm{SU}(2)$ symmetry and generates the gauge boson masses via the covariant derivative $(D_\mu\phi)^\dagger D_\mu\phi$ and the fermion masses via Yukawa couplings.

The Standard Model of particle physics applies to several generations of fermions. The first generation is the electron, electron-neutrino, up-quark, and down-quark. Two further generations of fermions are known from experiments with the same charges but higher masses than those of the first generation. The complete list of particles known in the Standard Model and their properties are listed in table 3.1.

Table 3.1: The Standard Model Particles with their corresponding masses, spin, and charges taken from Ref. [41]

Particle	**Mass (GeV)**	**Spin**	**Electric charge**	**Weak Isospin**	**Color Charge**
Quarks					
Up	2.2×10^{-3}	1/2	+2/3	+1/2	Yes
Down	4.7×10^{-3}	1/2	-1/3	-1/2	Yes
Charm	1.3	1/2	+2/3	+1/2	Yes
Strange	93.4×10^{-3}	1/2	-1/3	-1/2	Yes
Top	172.7	1/2	+2/3	+1/2	Yes
Bottom	4.2	1/2	-1/3	-1/2	Yes
Leptons					
Electron	0.511×10^{-3}	1/2	-1	-1/2	No
Muon	105.66×10^{-3}	1/2	-1	-1/2	No
Tau	1.777	1/2	-1	-1/2	No
Electron Neutrino	$< 2 \times 10^{-9}$	1/2	0	+1/2	No
Muon Neutrino	$< 0.19 \times 10^{-3}$	1/2	0	+1/2	No
Tau Neutrino	$< 18.2 \times 10^{-3}$	1/2	0	+1/2	No
Bosons					
Photon	0	1	0	0	No
$W^\pm$	80.377	1	±1	±1	No
Z^0	91.1876	1	0	0	No
Gluon	0	1	0	0	Yes
Higgs	125.25	0	0	-1/2	No

3.3 Polarized Same-Charged $W^\pm W^\pm$ Scattering

This thesis studies the scattering of same-charged polarized $W^\pm$ bosons. As described in section 3.2.6, the $W^\pm$ bosons gain their mass and thus their longitudinal polarization from the electroweak symmetry breaking. Therefore, the measurement of the $W^\pm$ boson polarization states is very sensitive to the Higgs mechanism and potential physics beyond the Standard Model.

3.3.1 Polarized Intermediate $W^\pm$ Bosons

Due to the high mass of $W^\pm$ boson, they decay quickly in lighter particles. Thus, the kinematic of the $W^\pm$ boson is experimentally accessible via its decay products. The $W^\pm$ boson can decay hadronically or leptonically, with the focus of this thesis being on the leptonic decay channels. The amplitude of a single intermediate W^+ boson

$$\mathcal{M} = \mathcal{M}^{\text{prod}}_\mu \frac{i}{k_\alpha k^\alpha - m^2 + i\Gamma m}\left(-g^{\mu\beta} + \frac{k^\mu k^\beta}{m^2}\right)\left(-\frac{ig}{2\sqrt{2}}\bar{\psi}_l \gamma_\beta (1-\gamma^5)\psi_{\nu_l}\right). \tag{3.86}$$

depends on the four-momentum k_μ, the mass m, and the decay width Γ of the W^+ boson [42]. $\mathcal{M}^{\text{prod}}_\mu$ is the matrix element of the corresponding $W^\pm$ boson production, and $\bar{\psi}_l$ and ψ_{ν_l} are the spinors of the antilepton-neutrino-pair originating from the W^+ boson decay. The completeness relation in equation (3.35) can be used to introduce vector boson polarization into equation (3.86). For an intermediate spin-1 field, the completeness relation becomes

$$\sum_{\lambda=0}^{3} \epsilon^\mu_\lambda \epsilon^{\nu *}_\lambda = -g^{\mu\nu} + \frac{1}{m^2} k^\mu k^\nu. \tag{3.87}$$

Similarly to the polarization of free spin-1 fields described in section 3.2.1, an intermediate $W^{\pm}$ boson whose momentum $\vec{\kappa}$ aligns with the z axis has the longitudinal polarization

$$\epsilon_0^{\mu} = \frac{1}{\sqrt{E^2 - \kappa^2}}(\kappa, 0, 0, E) \tag{3.88}$$

but additional obtains the auxiliary polarization

$$\epsilon_3^{\mu} = \sqrt{\frac{E^2 - \kappa^2 - m^2}{(E^2 - \kappa^2)\, m^2}}(E, 0, 0, \kappa). \tag{3.89}$$

For on-shell bosons with $E^2 - \kappa^2 = m^2$, this auxiliary polarization vanishes and the longitudinal polarization is equivalent to equation (3.34) for free spin-1 fields. With the on-shell W^+ boson production and decay amplitude,

$$\mathcal{M}_{\lambda}^{\mathcal{P}} = \mathcal{M}_{\mu}^{\text{prod}}\epsilon_{\lambda}^{\mu} \qquad \text{and} \qquad \mathcal{M}_{\lambda}^{\mathcal{D}} = \frac{-ig}{2\sqrt{2}}\bar{\psi}_l \epsilon_{\lambda}^{\mu *}\gamma_{\mu}(1-\gamma^5)\psi_{\nu_l}, \tag{3.90}$$

the matrix element of equation (3.86) can be written as

$$\mathcal{M} = \sum_{\lambda=0}^{2}\mathcal{M}_{\lambda}^{\mathcal{P}}\frac{i}{k_{\alpha}k^{\alpha} - m^2 + i\Gamma m}\mathcal{M}_{\lambda}^{\mathcal{D}} = \sum_{\lambda=0}^{2}\mathcal{M}_{\lambda}^{\mathcal{F}}. \tag{3.91}$$

Thus, the unpolarized amplitude $\mathcal{M}$ is the sum of the individual polarization amplitudes $\mathcal{M}_{\lambda}^{\mathcal{F}}$. However, for the calculation of the cross-section, the polarization states cannot be fully separated in the squared amplitude

$$\underbrace{|\mathcal{M}|^2}_{\text{coherent sum}} = \underbrace{\sum_{\lambda=0}^{2}\left|\mathcal{M}_{\lambda}^{\mathcal{F}}\right|^2}_{\text{incoherent sum}} + \underbrace{\sum_{\lambda\neq\lambda'}\mathcal{M}_{\lambda}^{\mathcal{F}*}\mathcal{M}_{\lambda'}^{\mathcal{F}}}_{\text{interference terms}}. \tag{3.92}$$

The resulting interference must be considered as an additional contribution to the transverse and longitudinal cross-sections. The interference contribution only vanishes for an integral over the full azimuthal angle ϕ of the decay leptons [42]. Therefore, as soon as specific lepton selection criteria are applied, this interference component can generally no longer be neglected.

With an integral over ϕ, the differential cross-section

$$\frac{1}{\frac{d\sigma(X)}{dX}}\frac{d\sigma(\theta, X)}{d\cos\theta dX} = \frac{3}{8}(1 \mp \cos\theta)^2 f_L(X) + \frac{3}{8}(1 \pm \cos\theta)^2 f_R(X) + \frac{3}{4}\sin^2\theta f_0(X) \tag{3.93}$$

of a decaying $W^{\pm}$ boson depends on the polar decay angle θ in the $W^{\pm}$ boson rest-frame and the remaining phase space variables denoted by X [42]. The upper sign represents the solution for the W^+ and the lower sign for the W^- boson. The right-handed (f_R), left-handed (f_L), and longitudinal (f_0) polarization fractions depend on the additional phase space kinematic X. Due to the absence of interference, these three polarization fractions add up to one.

In this thesis, the leptonic decay channels of the scattered $W^{\pm}$ bosons are investigated. Since the resulting neutrinos do not interact with the ATLAS detector, the decay angles of the $W^{\pm}$ bosons cannot be measured. Thus, a direct $W^{\pm}$ boson polarization measurement via the decay angle dependency in equation (3.93) is not possible. For the polarization measurement, the individual components of the squared amplitude in equation (3.92) are simulated[1] to predict the contributions of different polarization states. The predicted kinematics of the polarization states are compared to measured data to extract the polarization fractions. Since the neutrino kinematics are accessible in the simulated predictions, equation (3.93) can be used to validate the $W^{\pm}$ boson decay angle for the polarization simulation.

---

[1] The concept of simulation is explained in chapter 4

3.3.2 Motivating the Higgs Boson by Same-Charged $W_L^\pm W_L^\pm$ Scattering

In section 3.2.6, the Higgs boson is motivated by the lack of particle masses in the Standard Model. However, a Standard Model without the Higgs boson would have additional fundamental problems. These problems arise from the interaction of the electroweak bosons. One example is the longitudinal scattering $W_L^\pm W_L^\pm \to W_L^\pm W_L^\pm$. The leading-order Feynman diagrams of this scattering process are shown in figure 3.2. Considering only the quartic gauge coupling of the $W^\pm$ bosons in figure 3.2a lead to an amplitude $\mathcal{M}_{\text{quartic}} \sim \epsilon_0\epsilon_0\epsilon_0\epsilon_0$ proportional to the longitudinal polarization vector ϵ_0. Since ϵ_0^μ is proportional to $\frac{p^\mu}{m_W}$ the amplitude $\mathcal{M}_{\text{quartic}}$ is proportional to s^2 with s being the squared center-of-mass energy of the interacting $W^\pm$ bosons. Thus, the amplitude of this interaction would diverge at high energies violating the principle of unitarity. The unitarity principle is equivalent to the requirement that the probabilities of a specific initial state leading to different final states should always add up to one. Proving if this requirement is fulfilled would require calculating an infinite number of Feynman diagrams. Therefore, to validate unitarity for $2 \to 2$ processes, a commonly accepted criterion is that the amplitude of the leading-order diagrams should become asymptotically flat with increasing energy [43]. Since the amplitude of the quartic vertex $\mathcal{M}_{\text{quartic}}$ diverges with s^2 the $W_L^\pm W_L^\pm \to W_L^\pm W_L^\pm$ scattering would violate this fundamental requirement.

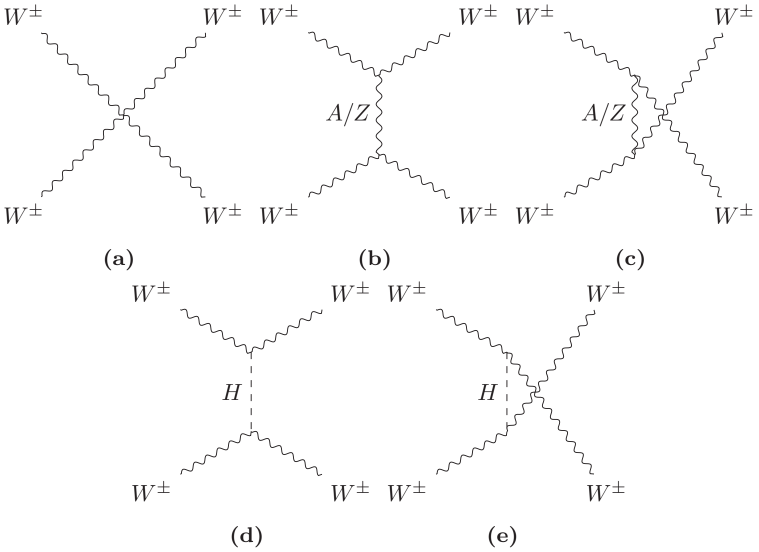


Figure 3.2: Leading-order Feynman diagrams for the $W^\pm W^\pm \to W^\pm W^\pm$ scattering

As shown in figures 3.2b and 3.2c, a Standard Model without Higgs would still have additional $W_L^\pm W_L^\pm \to W_L^\pm W_L^\pm$ Feynman diagrams with the exchange of a photon or a Z boson. These extra terms can cover some divergence of the quartic vertex but an amplitude

$$\mathcal{M}_{\text{gauge}} = -g_W^2 \frac{s}{4m_W^2} + \mathcal{O}(s^0) \tag{3.94}$$

remains that still increases with increasing energy [43]. Thus, only considering the quartic vertex and the exchange of a photon or Z boson would lead to a theory with unitarity violation for $W_L^\pm W_L^\pm \to W_L^\pm W_L^\pm$ scattering. Introducing the exchange of an electrical neutral scalar particle H solves this problem. The scalar particle H in figures 3.2d and 3.2e couples to the $W^\pm$ boson with the coupling strength g_{HWW}. This interaction generates the additional term

$$\mathcal{M}_H = g_{HWW}^2 \frac{s}{m_W^2} + \mathcal{O}(s^0) \tag{3.95}$$

that can only cancel out the divergence of $\mathcal{M}_{\text{gauge}}$ for $g_{HWW} = g_W m_W$ [43]. This requirement is exactly fulfilled by the Standard Model Higgs introduced in section 3.2.6. With the

Standard Model Higgs, the amplitude at high energies $s \gg m_H^2$ becomes

$$\mathcal{M}_{\text{gauge}} + \mathcal{M}_H = -g_W^2 \frac{m_H^2}{4m_W^2} \tag{3.96}$$

finally fulfilling the unitarity requirement for $W_L^\pm W_L^\pm \rightarrow W_L^\pm W_L^\pm$ [43].

Thus, the longitudinal scattering of $W^\pm$ bosons plays a crucial role in the Higgs mechanism of the Standard Model. The $W^\pm$ bosons gain their mass trough the electroweak symmetry breaking, and the resulting longitudinal component $W_L^\pm$ originates directly from the Goldstone bosons of the Higgs field. Independently of the Higgs mechanism described in section 3.2.6, the unitarity principle applied on longitudinal scattering $W_L^\pm W_L^\pm \rightarrow W_L^\pm W_L^\pm$ motivates a scalar particle with properties equivalent to the Standard Model Higgs boson. This makes the longitudinal scattering $W_L^\pm W_L^\pm \rightarrow W_L^\pm W_L^\pm$ one of the most interesting processes to study the electroweak symmetry breaking of the Higgs mechanism.

3.3.3 Same-Charged $W^\pm W^\pm$ Scattering at LHC

Since no experimental setup is available which is able to directly create two colliding $W^\pm$ boson beams, the $W^\pm$ bosons have to be produced in a collision. The energy necessary to produce an interacting $W^\pm$ pair becomes achievable with the proton-proton collisions at the LHC. The $W^\pm$ bosons are emitted by quarks of the colliding protons and their leptonic or hadronic decay products are measured in the ATLAS detector. Since an analysis of the hadronic decay channels would be impeded by a variety of multi-jet backgrounds from proton-proton collisions, this thesis focuses on the purely leptonic decay channels of the $W^\pm$ bosons. The reconstruction of the outgoing $W^\pm$ boson polarizations from their decay products is not straightforward and is the main challenge tackled in this thesis. However, even if a perfect tagging of the longitudinal polarization would be possible for the decaying $W^\pm$ bosons the study of $W_L^\pm W_L^\pm \rightarrow W_L^\pm W_L^\pm$ scattering is still not fully achieved because the polarization of the incoming bosons cannot be pre-determined. At this point, a significant advantage of the same-charged $W^\pm W^\pm \rightarrow W^\pm W^\pm$ scattering over other vector boson scattering processes arises. As shown in figure 3.3 taken from Ref. [43], at high center-of-mass energy, the final state $W_L^\pm W_L^\pm$ originates almost exclusively from the $W_L^\pm W_L^\pm$ initial state. Other $VV \rightarrow V_L V_L$ processes are significantly affected by helicity-flips, but for same-charged $W^\pm W^\pm \rightarrow W^\pm W^\pm$ scattering at the high center-of-mass energy achieved at the LHC, the polarization of the final state is similar to the initial state polarization [43]. Therefore, the measurement of $W^\pm W^\pm \rightarrow W_L^\pm W_L^\pm$ at the LHC is almost equivalent to the measurement of $W_L^\pm W_L^\pm \rightarrow W_L^\pm W_L^\pm$.

This thesis studies the same-charged $W^\pm W^\pm$ scattering via $pp \rightarrow l^\pm l'^\pm \nu\nu' jj$ processes at the LHC. Figure 3.4a shows the typical $W^\pm W^\pm$ scattering process where the circle is a placeholder for the $W^\pm W^\pm \rightarrow W^\pm W^\pm$ diagrams given in figure 3.2. Due to the hadronic production and the leptonic decay a variety of additional diagrams can lead to the same final state. These Feynman diagrams can be grouped by the number of their electroweak couplings $\alpha_{EW} \in \{g_e^2/(4\pi),\, g_W^2/(4\pi)\}$ with $g_e := g'$ and strong couplings $\alpha_S = g_S^2/(4\pi)$. At leading-order, the amplitude coupling orders $\mathcal{O}(\alpha_{EW}^3)$ shown in figure 3.4 and $\mathcal{O}(\alpha_{EW}^2 \alpha_S)$ shown in figure 3.5 contribute to the $pp \rightarrow l^\pm l'^\pm \nu\nu' jj$ process. Thus, the cross-section of the $W^\pm W^\pm jj$ scattering with two jets in the final state can be split into

$$\begin{aligned} \sigma_{W^\pm W^\pm jj} &\sim \left| \mathcal{M}_{\mathcal{O}(\alpha_{EW}^3)} + \mathcal{M}_{\mathcal{O}(\alpha_{EW}^2 \alpha_S)} \right|^2 \\ &\sim \left| \mathcal{M}_{\mathcal{O}(\alpha_{EW}^3)} \right|^2 + \left| \mathcal{M}_{\mathcal{O}(\alpha_{EW}^2 \alpha_S)} \right|^2 + 2Re\left(\mathcal{M}_{\mathcal{O}(\alpha_{EW}^3)} \mathcal{M}^*_{\mathcal{O}(\alpha_{EW}^2 \alpha_S)} \right). \end{aligned} \tag{3.97}$$

The resulting cross-section

$$\sigma_{W^\pm W^\pm jj} = \sigma_{W^\pm W^\pm jj\text{-EW}} + \sigma_{W^\pm W^\pm jj\text{-QCD}} + \sigma_{W^\pm W^\pm jj\text{-INT}} \tag{3.98}$$

consists of the pure electroweak contribution $\sigma_{W^\pm W^\pm jj\text{-EW}}$ at order $\mathcal{O}(\alpha_{EW}^6)$, a QCD-induced production $\sigma_{W^\pm W^\pm jj\text{-QCD}}$ at order $\mathcal{O}(\alpha_{EW}^4 \alpha_S^2)$, and the interference of electroweak and QCD-induced production $\sigma_{W^\pm W^\pm jj\text{-INT}}$ at order $\mathcal{O}(\alpha_{EW}^5 \alpha_S)$. As shown in figure 3.4j, the pure electroweak contribution $W^\pm W^\pm jj$-EW does not only include diagrams with $W^\pm$ boson

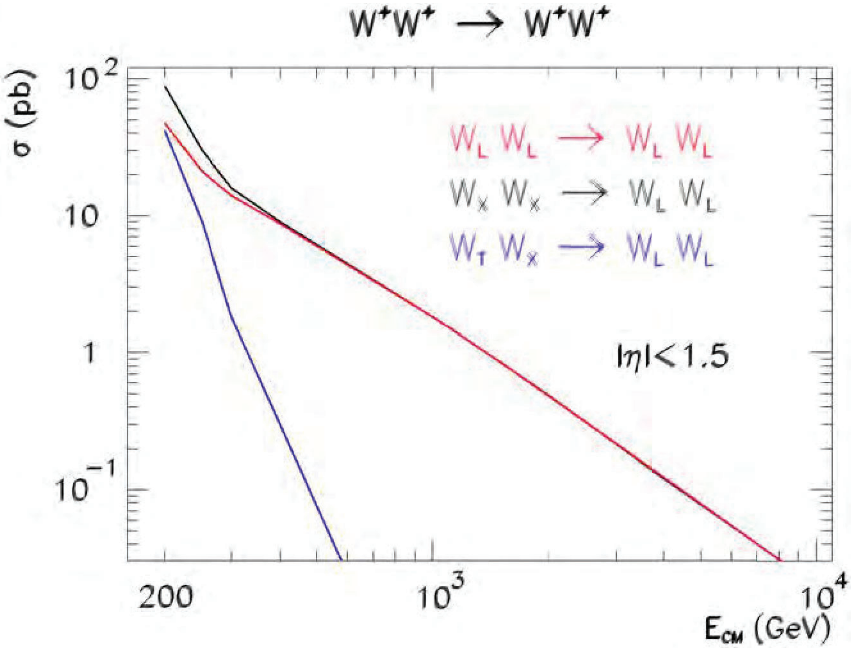


Figure 3.3: "Total $W^+W^+ \rightarrow W_L^+W_L^+$ scattering cross-sections in the SM as a function of the center-of-mass energy. Shown are the individual contributions of different initial polarization states to the final state consisting of purely longitudinal $W_L^+W_L^+$ pairs. Subscript X denotes any polarization (T or L). Assumed are two on-shell, unpolarized, colliding W^+ beams. A cut on the scattering angle that corresponds to pseudorapidity of ±1.5 with respect to the incoming W direction was applied. Results of MadGraph calculations." [43]

scattering. However, since the $W^\pm W^\pm jj$-EW process cannot be further split in a gauge invariant way, $\sigma_{W^\pm W^\pm jj\text{-EW}}$ is taken as the signal for the $W^\pm W^\pm jj$ polarization analysis [44]. $W^\pm W^\pm jj$-QCD and $W^\pm W^\pm jj$-INT have to be considered as irreducible backgrounds leading to the same final states.

Electroweak $W^\pm W^\pm jj$-EW Contribution

Figure 3.4 shows an exemplary set of Feynman diagrams contributing to $W^\pm W^\pm jj$-EW. The incoming quark lines in figures 3.4a to 3.4c have no color exchange and are directly connected to the outgoing quark lines. Therefore, the resulting final jets are expected in the forward regions giving them a significant separation in Δy_{jj} and a high invariant mass m_{jj} [44]. The resulting jet kinematic differs significantly from the kinematic originating from the s-channel-like diagrams in figures 3.4d to 3.4g. Since their jets originate from $W^\pm$ boson decay their invariant mass m_{jj} is expected to follow a Breit-Wigner distribution peaking at the $W^\pm$ boson mass and is therefore significantly smaller than for the t/u-channel-like Feynman diagrams in figures 3.4a to 3.4c. Indeed, the analysis performed in this thesis chooses a high m_{jj} selection, giving a clear preference for the t/u-channel-like Feynman diagrams. Although the choice of this selection is primarily motivated by the suppression of backgrounds, the explicit preference for t/u-channels also offers an advantage for polarization reconstruction. The quark lines in the t/u-channel-like Feynman diagrams are directly connected to the initial $W^\pm$ bosons without any additional interaction at leading-order. Thus, polarization information of the emitted $W^\pm$ boson can be accessed via the final jet kinematics [43]. Since no significant impact of helicity-flips is expected for $W^\pm W^\pm \rightarrow W^\pm W^\pm$ this improves also the polarization reconstruction for the leptonically decaying $W^\pm$ bosons.

In addition to these fully-resonant diagrams, a group of single-resonant and non-resonant diagrams shown in figures 3.4h to 3.4j contributes to the $W^\pm W^\pm jj$-EW process. However, these non-resonant diagrams are suppressed by a factor of Γ_W/m_W typically reducing their contribution to a few percent at maximum [44]. This suppression is crucial for the polarization definition in $W^\pm W^\pm jj$-EW. As described in section 3.3.1 an additional un-

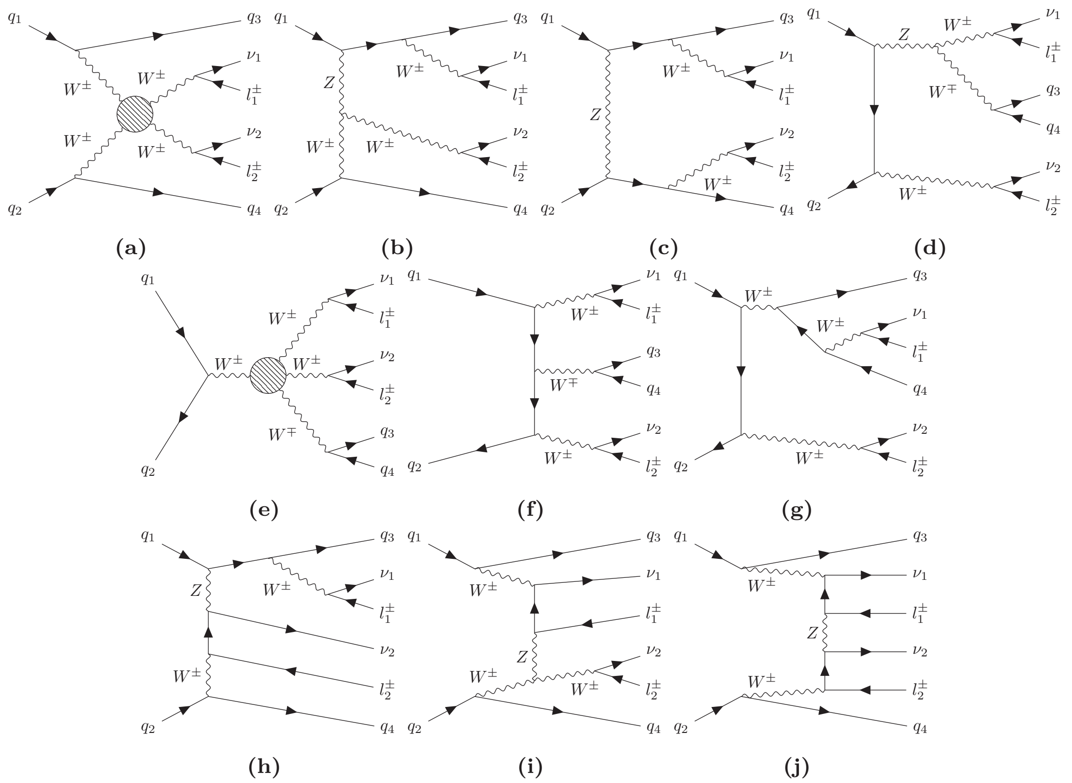


Figure 3.4: A selection of possible Feynman diagrams for $W^\pm W^\pm jj$-EW with only electroweak interactions. This selection includes fully-resonant t/u-channel-like diagrams in figures 3.4a to 3.4c, fully-resonant s-channel-like diagrams in figures 3.4d to 3.4g, single-resonant diagrams in figures 3.4h and 3.4i, and a non-resonant diagram in figure 3.4j.

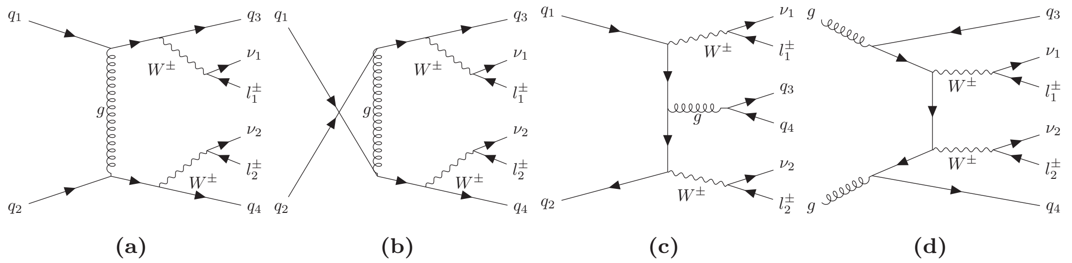


Figure 3.5: A selection of possible Feynman diagrams for $W^\pm W^\pm jj$-QCD at the coupling order $\mathcal{O}(\alpha_{EW}^4 \alpha_S^2)$.

physical auxiliary polarization arises for intermediate off-shell bosons. Thus, only for the fully-resonant diagrams a physical meaningful polarization can be assigned. The commonly used solution for this problem is to drop the single-resonant and non-resonant diagrams and assign the polarization for the resonant boson decays in the remaining diagrams. However, dropping these diagrams leads to gauge violation. This thesis uses polarization templates generated by Sherpa and MadGraph that have different approaches to deal with this gauge violation:

- **Narrow-Width Approximation (NWA)** used by Sherpa [45]:
  For the leptonically decaying $W^\pm$ bosons, the denominator of the $W^\pm$ propagator is replaced by a delta function:

$$\frac{1}{q^2 - m_W^2 + i\Gamma_W m_W} \rightarrow \frac{\pi\delta\left(q^2 - m_W^2\right)}{\Gamma_W m_W} \tag{3.99}$$

  The propagator denominator of the other intermediate bosons is not replaced by a delta function and only their width Γ_V is set to 0:

$$\frac{1}{q^2 - m_V^2 + i\Gamma_V m_V} \rightarrow \frac{1}{q^2 - m_V^2} \tag{3.100}$$

This can lead to a divergence of the propagator of hadronically decaying $W^\pm$ bosons in s-channel-like diagrams (figures 3.4d to 3.4g). At leading-order, this is avoided by the $m_{jj} > 200\,\text{GeV}$ selection criterion in the signal and Low-m_{jj} control region introduced in section 6.3.

- **Neighborhood Restriction** used by MadGraph [46]:
  The violation of the gauge invariance is constrained by limiting the invariant mass of the intermediate bosons to the neighborhood of their literature mass. A $|m_V - k^\mu k_\mu| \leq$ `bwcutoff`$\cdot\Gamma_V$ selection is applied to only consider on-shell-like contributions [47][48][49].

With these approximations, the single-resonant and non-resonant diagrams can be dropped without the calculation being affected by gauge violations. However, the effects of the diagram selection combined with the NWA or the neighborhood restriction must be considered for the polarization analysis. Therefore, the effects of this approximation are studied in section 8.1 and a corresponding correction is derived.

QCD-Induced $W^\pm W^\pm jj$-QCD Contribution

As shown in figure 3.5, the $W^\pm W^\pm jj$-QCD contributions including QCD interactions lead to the same final states as the $W^\pm W^\pm jj$-EW diagrams. Due to the additional strong couplings, diagrams with gluons as incoming particles, outgoing jets, and intermediate color interactions also contribute to the $W^\pm W^\pm jj$-QCD background. A characteristic property of most of the $W^\pm W^\pm jj$-QCD contributions is the strong color interaction connecting their incoming gluons/quarks. Thus, their final state jets have a much more central distribution than the t/u-channel-like diagrams in figure 3.4 [44]. Therefore, excluding events with hadronic activity in the central region can significantly constrain this background.

Interference between Electroweak and QCD-Induced Production

The $W^\pm W^\pm jj$-INT background arises from the interference between the fully electroweak and the QCD-induced contributions. These contributions have a coupling order $\mathcal{O}(\alpha_{EW}^5 \alpha_S)$ and originate solely from the interference of $W^\pm W^\pm jj$-EW and $W^\pm W^\pm jj$-QCD diagrams with identical external quarks. Since the phase space used in this thesis favors the t/u-channel-like diagrams of the $W^\pm W^\pm jj$-EW process, their interference with the $W^\pm W^\pm jj$-QCD process is expected to give the dominant interference contribution. The resulting interferences can be split into three categories discussed in Ref. [44]:

- **Different flavored quark lines**
  When the incoming quarks are of different flavor they have to be connected via a gluon for $W^\pm W^\pm jj$-QCD as in figure 3.5a. This leads to the exchange of a color octet between the two quarks. Since the quarks in t/u-channel-like diagrams of $W^\pm W^\pm jj$-EW are always connected via a color singlet the resulting interference vanishes exactly.

- **Identical quark flavor lines**
  In this case the $W^\pm W^\pm jj$-EW interfere with $W^\pm W^\pm jj$-QCD diagrams with flipped quark lines as shown in figure 3.5b. This leads to a non-vanishing contribution to the interference. However, due to the color connection between the quarks in the $W^\pm W^\pm jj$-QCD process their jet kinematic prefers a phase space that significantly differs from the jet kinematic of the t/u-channel-like diagrams of $W^\pm W^\pm jj$-EW. Thus, this interference does not fully vanish but is significantly suppressed.

- **Quark-antiquark lines**
  Quark-antiquark pair diagrams like figure 3.5c interfere with $W^\pm W^\pm jj$-EW. This interference is not vanishing but suppressed by the jet kinematics for the same reasons discussed for the case of the identical quark flavor lines.

Due to the limited number of possible diagram combinations and the differences in the jet phase space, the interference $W^{\pm}W^{\pm}jj$-INT is expected to introduce a significantly lower background for the polarization analysis than the QCD-induced $W^{\pm}W^{\pm}jj$-QCD contribution.

3.3.4 Polarization Reference Frame

The polarization of a particle depends directly on its momentum and is therefore not Lorentz-invariant. Therefore, the $W^{\pm}$ boson polarization is not strictly defined but depends on the kinematic reference frame. For the scattering of two bosons, the center-of-mass system of the two bosons is best motivated by the theory of the Standard Model. As described in section 3.3.2, the cancelation of the Higgs contributions and the pure gauge boson interactions is one of the main motivations of polarization measurements in vector boson scattering. According to Ref. [50], this unitarity cancelation at tree-level becomes maximal in the $W^{\pm}W^{\pm}$ center-of-mass frame ($W^{\pm}W^{\pm}$-cmf). Thus, the $W^{\pm}W^{\pm}$-cmf is the natural choice to study the polarization in diboson scattering. This reference frame was already chosen for the ATLAS polarization analyses of the $W^{\pm}Z$ [4] and ZZ production [7]. Therefore, the $W^{\pm}W^{\pm}$-cmf is the nominal reference frame of the polarization measurement in this thesis.

The existing CMS polarization analysis [3] for same-charged $W^{\pm}W^{\pm}jj$-EW scattering provided results in the $W^{\pm}W^{\pm}$-cmf but also in the center-of-mass frame of the two initial partons (pp-cmf). Therefore, an additional polarization measurement is performed in the pp-cmf to provide values comparable to the CMS experiment. However, the main focus of this thesis remains on the polarization in the $W^{\pm}W^{\pm}$-cmf.

Chapter 4

Simulation of the Standard Model Prediction

The Standard Model of particle physics comprehensively describes the known interactions in proton-proton collisions at the LHC. However, an analytic calculation of the corresponding Standard Model prediction would include an infinite amount of Feynman diagrams for a vast amount of particles and is thus impossible. In order to make a theoretical prediction factorization is used combined with perturbation theory. The overall problem is split into subprocesses in different energy scales. In figure 4.1, this factorization of the event simulation is visualized for the simulation of a $t\bar{t}H$ event. This chapter briefly summarizes the detailed explanations in Ref. [51], which is recommended for a deeper understanding of the event generation. The factorization steps covered in the following sections are:

- **Hard subprocess**: (red blobs in figure 4.1)
  Primary scattering process at the collision energy
- **Parton shower**: (red lines in figure 4.1)
  Evolution of the emitted partons down to an energy scale of 1 GeV
- **Multi-jet merging**:
  Combination method for the hard subprocess calculation and the subsequent parton shower
- **Hadronization**: (light green blobs in figure 4.1)
  Transition of colored partons in the parton shower to color-neutral hadrons
- **Hadron and tau decay**: (dark green blobs in figure 4.1)
  Decay of unstable hadrons and leptons
- **QED radiation**: (yellow lines in figure 4.1)
  QED-induced radiation of particles
- **Underlying event**: (purple blob and lines in figure 4.1)
  Secondary interactions in addition to the hard scattering

4.1 Hard Subprocess

The number of predicted events $N_{\text{event}} = L\sigma_{\text{event}}$ depends on the luminosity L of the detector and the cross-section of the event σ_{event} determined by the hard subprocess. At the LHC, the hard subprocess describes the initial interaction of the partons a and b. This interaction involves a large momentum transfer, creating the final state f with n particles. The resulting $ab \rightarrow f$ process can be described by perturbation theory using Feynman diagrams. The corresponding matrix element $\mathcal{M}_{ab\rightarrow f}$ is calculated by summing over the single Feynman

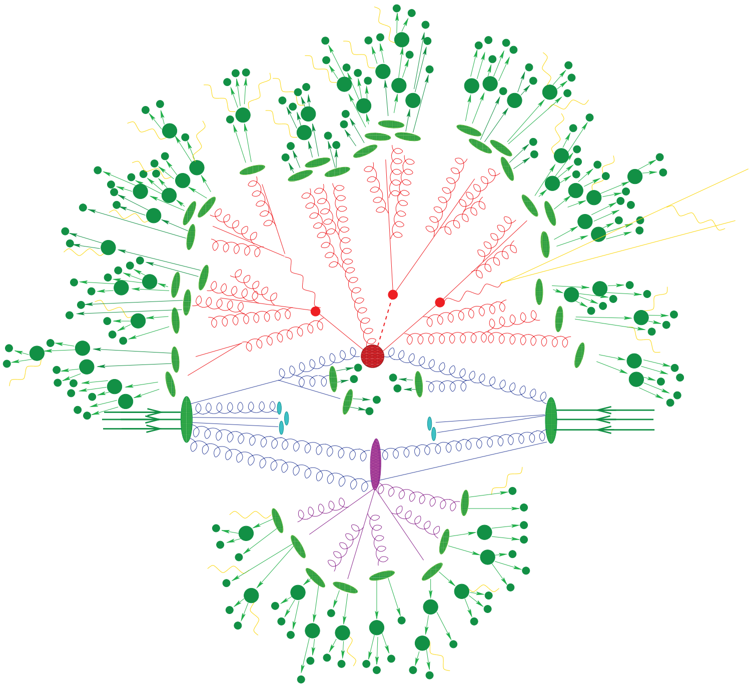


Figure 4.1: Representation of a $t\bar{t}H$ event produced by the SHERPA event generator taken from Ref. [52]. The simulation includes the hard interaction (big red blob), the decay of the top quarks and the Higgs boson (small red blobs), photon radiations (yellow), the QCD radiations (red), parton hadronization (light green blobs), and the final decay of the hadrons (dark green). In addition, a secondary interaction (purple) is included in the represented simulation.

diagrams. By integrating over the phase space element $d\Phi_n$ of the final state particles, the partonic cross-section is obtained:

$$\sigma_{\text{parton}} = \int d\Phi_n |\mathcal{M}_{ab\to f}(\Phi_n)|^2 \tag{4.1}$$

The differential phase space element of the final state is defined by

$$d\Phi_n = \prod_i^n \frac{d^3p_i}{(2\pi)^3 2E_i} \cdot (2\pi)^4 \delta\left(P_a + P_b - \sum_i^n P_i\right) \tag{4.2}$$

and fulfills the energy momentum conservation $P_a + P_b = \sum_i^n P_i$. Due to the infinite number of Feynman diagrams, the partonic cross-section is written as perturbation series

$$\sigma_{\text{parton}} = \sum_{k=0}^{\infty} \sigma_{\text{parton}}^{(k)} \left(\mathcal{O}(\alpha^{min}\alpha^k)\right) \tag{4.3}$$

by expanding it into contributions with different order in the coupling strength $\alpha \in \{\alpha_S, \alpha_{EW}\}$. The lowest order of the coupling strength for the studied process is given by α^{min}. The impact of $\sigma_{\text{parton}}^{(k)}$ decreases with increasing coupling order k of electroweak and high-energetic QCD interactions. Therefore, stopping the perturbation series at a specific coupling order offers the possibility to limit the calculation on dominant Feynman diagrams. The contribution $\sigma_{\text{parton}}^{(0)}$ with the lowest number of vertices is called leading-order (LO). When the Feynman diagrams with one additional vertex are taken into account, the calculation is called next-to-leading-order (NLO). The partonic NLO cross-section

$$\begin{aligned}\sigma_{\text{parton}}^{\text{NLO}} &:= \sigma_{\text{parton}}^{0} + \sigma_{\text{parton}}^{1} \\ &= \sigma_{\text{parton}}^{0} + \sigma_{\text{parton}}^{V} + \sigma_{\text{parton}}^{I} + \int d\Phi_1 \left(\sigma_{\text{parton}}^{R} - \sigma_{\text{parton}}^{S}\right)\end{aligned} \tag{4.4}$$

includes virtual emissions in σ^V_{parton} via internal loops and real emissions in σ^R_{parton} via radiation of additional particles in the final state. The partonic cross-section of the real emissions is integrated over the phase space element $d\Phi_1$ of the additionally emitted particle. Since σ^V_{parton} diverges towards negative infinity and σ^R_{parton} diverges towards positive, the auxiliary terms σ^I_{parton} and σ^S_{parton} are introduced. These auxiliary terms are constructed in such a way that they prevent the individual divergences and cancel each other out with $\sigma^I_{\text{parton}} = \int d\Phi_1 \sigma^S_{\text{parton}}$. This enables the individual calculation of the real and virtual NLO contributions without changing the total NLO cross-section.

Further divergences originate from high momenta in loop integrals. These ultraviolet divergences can be resolved by introducing a renormalization scale μ_R into the matrix element calculation [53]. The matrix element couplings are evaluated at the specific energy scale defined by μ_R.

At the LHC, the interacting partons a and b are part of the protons h_1 and h_2 accelerated at the LHC. Their momentum is be defined by the fractions x_a and x_b of the momentum of their corresponding parent hadron. The parton distribution function (PDF) $f^h(x, \mu_F)$ gives the probability density of partons inside the proton. The PDFs are derived by a global to fit to measured data. The choice of the specific PDF is associated with systematic uncertainties, which are discussed in more detail in section 11.2. For a given hadronic center-of-mass energy s, the cross-section

$$\sigma = \sum_{a,b} \int_0^1 dx_a dx_b \int d\Phi_n f_a^{h_1}(x_a, \mu_F) f_b^{h_2}(x_b, \mu_F) \cdot \frac{1}{2x_a x_b s} |\mathcal{M}_{ab\to f}(\Phi_n; \mu_F, \mu_R)|^2 \tag{4.5}$$

of the hard subprocess is defined by the integral over the momentum fractions carried by the interacting partons and the integral over the final state phase space Φ. For the factorization into parton density and matrix element calculation, a factorization scale μ_F has to be chosen. This factorization scale μ_F separates the perturbative QCD calculation of the matrix element at high energy scales and the non-perturbative QCD description at the low-energy scales.

Due to the high dimensionality of the integral in equation (4.5), Monte Carlo techniques are used to calculate the integral in equation (4.5). Regardless of the number of dimensions, the statistical uncertainty of Monte Carlo techniques decreases with the increasing number of generated events N_{MC} by the factor of $1/N_{MC}$.

A variety of event generators can simulate the hard subprocess. The ones used in this thesis are MadGraph [46], Sherpa [54], and Powheg [55].

4.2 Parton Shower

The hard subprocess describes the differential cross-section of the scattering at a high energy scale. The subsequent evolution of jets and their inner structure is derived by the parton shower. The shower is described as a Markov process in which further partons are emitted in succession, with each emission depending only on the current state and not explicitly on the previous emissions. This iterative approach describes the evolution of a state with few partons at the energy scale of the hard subprocess to a state with many partons at low energy. The momentum transfer of the parton shower decreases in each step down to the order of 1 GeV where the QCD becomes strongly interacting and the perturbation theory breaks down. The initial state of the parton shower starts with the final state of the hard subprocess. The shower algorithm is implemented as subsequent collinear splittings ordered according to a shower-dependent evolution scale. The initial parton is split into two partons whose direction of motion is almost parallel to the momentum direction of the initial parton. The resulting state with one additional parton is the starting point for further collinear splittings. The evolution of parton momenta is continued until no branchings are produced above the minimal energy scale of the shower algorithm. To connect the parton shower algorithm with the hard subprocess a resummation scale has to be chosen. This resummation scale represents the energy at which partons emit other partons. This starting condition of the shower evolution ensures that any emission of the parton shower is distinct from emission calculated in the hard subprocess and thus avoids double counting. So far, only the collinear emission of partons has been discussed, not the emission of soft gluons. This contribution can be considered by performing the collinear shower algorithm with the opening angle as the evolution scale. An alternative and more common approach is based color dipoles.

In addition to the final state parton shower originating from the hard subprocess, colliding protons emit partons also before the hard collision. The emitted partons can participate in the subsequent hard interaction. This initial state radiation is simulated in a backward evolution. Starting with a highly energetic parton participating in the hard subprocess, the parton is traced back to the low energy scale of its initial emission from the proton. Thus, a similar algorithm as for the final state radiation is used but with evolution backward in time.

In this thesis, the default shower algorithms in Herwig [56] and Pythia [57] are used for hard subprocesses generated by MadGraph. For processes generated by Powheg, the Pythia's default shower algorithm is used. For events generated by Sherpa the parton shower is directly provided by this event generator.

4.3 Multi-Jet Merging

As described in the previous sections, the simulation of final state radiations is possible in the hard subprocess or by the parton shower evolution. While matrix element calculations provide an excellent prediction for well-separated hard emissions, the parton shower algorithm is more applicable for collinear and soft partons. Thus, a clever combination of these two methods can provide a more precise prediction for the emission of additional partons. To correctly combine these two methods, one has to consider that the matrix element calculation gives the inclusive probability of at least n partons in a state calculated at the lowest order of the strong coupling α_S. The parton shower, on the other hand, provides an exclusive probability of exactly n partons by an approximate calculation to all orders in α_S. Thus, a naive combination of matrix element calculation and parton shower can lead to double-counting for the parton shower evolution.

One procedure for merging the matrix element calculation and the parton shower is the CKKW algorithm [58] which is a key element of the Sherpa event generator. This algorithm is based on a phase space separation into a region of hard jet production and a region for soft jet evolution. The matrix element calculation is interpreted as a core process with a subsequent series of shower branchings. The shower history within the matrix element calculation is required for a correct merging with a subsequent shower evolution. This shower history is reconstructed by a clustering algorithm running the shower evolution backward. For a final state with n partons, the most probable previous state with $n-1$ partons is determined. A merging scale Q_{cut} must be chosen to define resulting shower histories of the generator. Hard parton emissions above the merging scale Q_{cut} are taken from the calculation of the matrix element. For the description of soft and collinear parton emission with $Q < Q_{cut}$, the parton shower algorithm is used. The resulting algorithm can consider higher-order QCD effects of real emissions by merging matrix element calculations of different jet multiplicity and the subsequent parton showers. This method has been extensively studied and successfully applied in producing individual $W^{\pm}$ and Z bosons and their corresponding pair productions.

Therefore, the merging algorithm is used for the simulation of the same-charged $W^{\pm}W^{\pm}jj$-EW contribution in section 7.1 and for the correction for higher-order QCD effects in section 8.1.

4.4 Hadronization

The parton shower describes the parton and jet evolution down to an energy scale of 1 GeV. At this energy scale, the QCD perturbation theory breaks down as $\alpha_S \rightarrow 1$. Thus, phenomenological models have to be used to describe how hadrons are formed. The hadronization process from the colored partonic state to a color neutral hadronic final state can be described by the string or cluster model:

String Model: At high distances, the strong interaction leads to a linear confinement. Color flux tubes between partons describe the linear potential between two color charges. The endpoints of the color tubes represent the quarks, while gluons are given as kinks on the tube. If the two partons move away from each other, the tube is stretched with its tension increasing with length. Thus, the energy stored in the tube increases linearly. As soon as enough energy has been stored in the tube, a new quark-antiquark pair is formed, to which the original partons are now connected according to their color charge. This procedure forms color-neutral parton systems that can further break down if their invariant mass is high enough. This process is repeated until color-neutral hadrons are formed.

Cluster Model: The preconfinement property of parton showers allows the clustering of the shower into color-neutral sets of partons. In this step, gluons are represented by color-anticolor lines connected at the gluon vertices. The clusters within the parton shower are formed by combining nearby color-anticolor partners into pairs. In the next step, these clusters decay into hadrons. A two-body decay is performed under consideration of the flavors and kinematics leading to individual color-neutral hadrons.

4.5 Hadron and Tau Decay

After hadronization, unstable particles decay during their travel through the ATLAS detector. The decay of heavy hadrons is simulated by combining matrix element calculations with experimental results. Different decay algorithms are possible for the hadron decay depending on the used hadronization algorithm, the hadrons selected for the simulation, and the choice of considered decay modes. In addition to unstable hadrons, only the decay of τ leptons has to be considered since muons are stable enough to travel through the whole detector. Its semileptonic decay channel dominates the decay of τ leptons into a neutrino and a virtual $W^{\pm}$ boson that decays hadronically.

4.6 QED Radiation

Besides the QCD-induced radiation of particles, QED radiation must also be considered in the event generation. The most common strategy is using the same shower algorithm used for the QCD-induced radiations. Especially for processes with QED and QCD radiations using the same shower algorithm is the preferred method to simulate both simultaneously. An alternative and less used method to model the QED radiation is the Yennie-Frautschi-Suura formalism, which is described in more detail in Ref. [51].

4.7 Underlying Event

In addition to the hard subprocess and the initial state radiation further interactions and particles can originate from the initial proton-proton collision. This additional activity is described in the underlying event with contributions from multiple parton interaction (MPI). These processes are softer than the hard subprocess and usually lead to a higher multiplicity and total transverse energy of the event. For more details about the underlying event and the MPI simulation, Ref. [51] is recommended.

4.8 Event Levels of Simulated Data

The simulation process described in this chapter aims to generate data distributions predicted by the Standard Model. Depending on the use case, the generated events are taken from different stages of the simulation process to prevent the computational effort of the following simulation steps. In this thesis three event levels are used:

- **Parton-level:**
  Events at parton-level originate from the simulation of the hard process. They consist

of fundamental particles, such as leptons, quarks, and gluons. In section 8.2, theory calculations on parton-level are used to implement a NLO-EW correction for the $W^{\pm}W^{\pm}jj$-EW polarization states.

- **Particle-level:**
  After the subsequent simulation steps described above, the event is considered to be on particle-level. The event consists of long-living leptons and hadrons. The particle-level represents the event state an idealized detector would measure. Simulations on particle-level are used to define the fiducial signal region in section 6.3.4, derive approximate NLO-QCD corrections in section 8.1, and implement theory uncertainties in section 11.2.1.
- **Reconstruction-level:**
  The last step of the simulation chain is the simulation of the measurement at the ATLAS detector. This simulation is performed by GEANT4 [59] followed by the object reconstruction described in section 2.2.3. The resulting reconstruction-level provides the final prediction of data measured at the ATLAS detector. Therefore, the reconstruction-level is the default and most commonly used event level in this thesis.

Chapter 5

Neural Networks

In many analyses, such as the search for same-charged $W^{\pm}W^{\pm}jj$ scattering, individual measured variables are examined to see how well they separate signal and background processes. Signal regions are defined using physically motivated selection criteria, and the data is analyzed in a sensitive variable. However, this procedure is only applicable if such a sensitive variable exists and is measurable. For the measurement of the longitudinal polarization in $W^{\pm}W^{\pm}$ scattering, however, most of the information about the original $W^{\pm}$ boson kinematics is lost via the leptonic decay. Only their decay products leave the beam pipeline and the resulting neutrinos cannot be measured in the ATLAS detector. Thus, this analysis faces the problem that no variable exists to improve the sensitivity to the longitudinal polarized $W^{\pm}W^{\pm}$-EW scattering. Therefore, multiple physical variables are combined by neural networks to boost the significance. For this thesis, only binary classifiers are trained, used as sensitive variables in chapter 9 and as multivariate reweighting algorithms in section 8.1.

The basic concept of neural networks is inspired by the human brain. The human brain consists of about 10^{11} neurons. Each neuron has about 1000 dendrites and an axon that branches towards the end [60]. The axon branches can couple to the synaptic regions of other dendrites. The neuron gets stimulated by electro-chemical processes at the dendrite, which can increase or decrease the electric potential within the neuron. As soon as the voltage exceeds a threshold value, an electrical impulse is transmitted via the axon to other neurons. The key to information processing in the brain is the synapses, which either amplify or inhibit the signal and grow with frequent use and regress with infrequent use. Since these biological concepts are capable of performing very complex tasks, they are transferred to artificial neural networks to outperform the sequential rule-based algorithms for information processing [61].

5.1 Basic Components of Neural Networks

An artificial neuron model is designed to transfer the principles of the brain to machine processes. The artificial neuron, also called a node, is connected to other neurons given by the network architecture. Analogous to what happens at synapses in the biological equivalent, the output x_i of the ith neuron is combined with an associated weight w_{ij}, a bias θ_j, and a propagation function h_j before it enters the jth neuron [62]. The bias θ is comparable to the electrical threshold of a biological neuron. In most applications, as in this thesis, the resulting input signal $s_j = h_j\left(\vec{x}, \vec{w_j}, \theta_j\right)$ is calculated using a scalar product as the propagation function $h_j\left(\vec{x}, \vec{w_j}, \theta_j\right) = \sum_i x_i w_{ij} + \theta_j$. The resulting inner activation is now given to an activation function a_j to calculate the output y_j of the neuron. Thus, the output of the j-th neuron results in

$$y_j = a_j\left(\sum_i x_i w_{ij} + \theta_j\right) \tag{5.1}$$

when using a linear propagation function [61].

A neural network is a cluster of neurons defined by the connections between them. A suitable network topology must be chosen to process information with a neural network.

Since neural networks are, in principle, directed graphs, they can be constructed into any graph. However, neurons are often grouped into layers whose members are connected to the same subsets of neurons in the network. For simplification, the neurons of one layer usually also have the same propagation and activation function [61].

In this thesis, the use of neural networks refers to modeling a function f, which maps the measured values x to a value $y = f(x)$. This function can be, for example, a physical probability density used for classification. For feedforward neural networks (FFNs), the information flows only in one direction through the network without any feedback connections. This information flow gives an order to the layers of the network, starting with the input layer, which for example receives the event variables, followed by several hidden layers, which are finally followed by the output layer [63].

In figure 5.1, a model for binary classification used in this thesis is shown. It takes 24 input variables ranging from the transverse momentum and energy of the leading lepton to the invariant mass of the two jets. The input layer is directly connected to a normalization layer, normalizing the distribution of these variables to a mean of 0 and a standard deviation of 1. The normalization layer is applied to ensure that all input variables can have the same impact on the model's prediction even if their values are in very different orders of magnitude. The normalization layer is fully-connected to the first hidden layer, followed by another hidden layer. The second hidden layer is connected to the output layer, which consists of a single neuron representing the score of the binary classification. Adding more neurons to the output layer could extend the model to a multi-class classifier. Although the shown network is relatively small, with two layers of 60 neurons each, it already has 5221 trainable parameters. By adding more layers and neurons per layer the number of trainable parameters can easily grow into the millions.

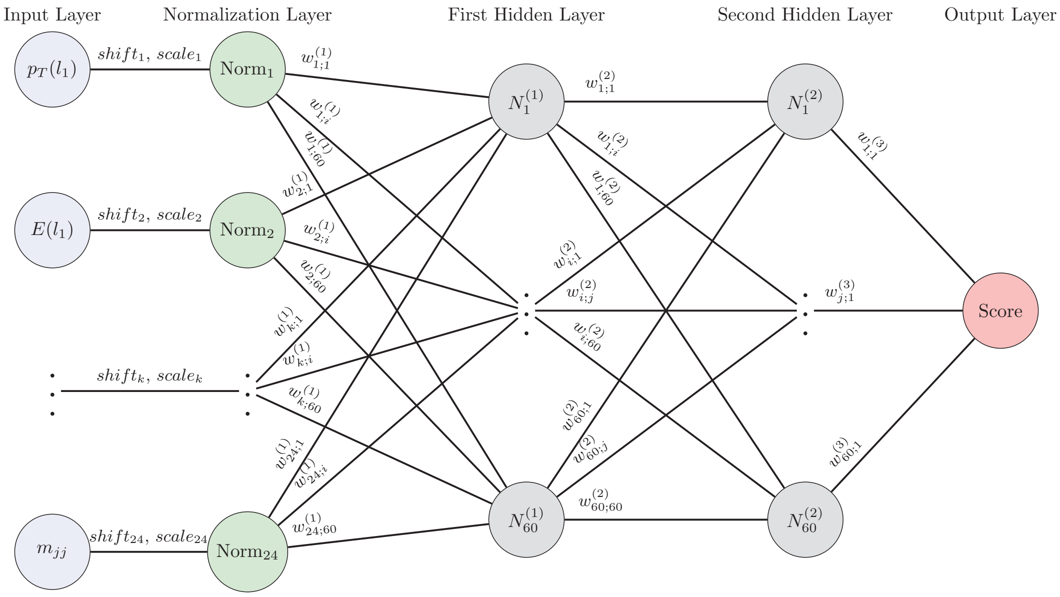

Input Layer
Normalization Layer
First Hidden Layer
Second Hidden Layer
Output Layer
$p_T(l_1)$
$E(l_1)$
m_{jj}
$shift_1$, $scale_1$
$shift_2$, $scale_2$
$shift_k$, $scale_k$
$shift_{24}$, $scale_{24}$
Norm_1
Norm_2
Norm_{24}
$w^{(1)}_{1;1}$
$w^{(1)}_{1;i}$
$w^{(1)}_{1;60}$
$w^{(1)}_{2;1}$
$w^{(1)}_{2;i}$
$w^{(1)}_{2;60}$
$w^{(1)}_{k;1}$
$w^{(1)}_{k;i}$
$w^{(1)}_{k;60}$
$w^{(1)}_{24;1}$
$w^{(1)}_{24;i}$
$w^{(1)}_{24;60}$
$N^{(1)}_1$
$N^{(1)}_{60}$
$w^{(2)}_{1;1}$
$w^{(2)}_{1;i}$
$w^{(2)}_{1;60}$
$w^{(2)}_{i;1}$
$w^{(2)}_{i;j}$
$w^{(2)}_{i;60}$
$w^{(2)}_{60;1}$
$w^{(2)}_{60;j}$
$w^{(2)}_{60;60}$
$N^{(2)}_1$
$N^{(2)}_{60}$
$w^{(3)}_{1;1}$
$w^{(3)}_{j;1}$
$w^{(3)}_{60;1}$
Score

Figure 5.1: Neural network with normalization layer, two hidden layers, and an output layer with one neuron.

5.2 Activation Functions

Without the activation function introduced in equation (5.1) the neural network would represent a linear combination of the input features. To handle complex non-linear problems the non-linearity introduced by the activation function is required. The default recommendation for neural networks is the rectified linear unit (ReLU) function

$$f_{\mathrm{ReLU}}(x) = \max(0, x) \tag{5.2}$$

representing a constant activation of zero for negative neuron input and a linear activation for positive input [63].

An alternative to the ReLU function is given by the Swish function

$$f_{\text{Swish}}(x) = \frac{x}{1 + e^{-x}}. \tag{5.3}$$

As can be seen in figure 5.2 the Swish function is very close to the ReLU function but is non-monotonic and smooth. In Ref. [64], the Swish function was proven to outperform the ReLU function, especially for deep networks.

The activation function is not only required to introduce non-linearity into the network but can also be used to restrict the output of a neuron. Such a restriction is often the case for the output layer. For a binary classification the Sigmoid function

$$f_{\text{Sigmoid}}(x) = \frac{1}{1 + e^{-\beta x}} \tag{5.4}$$

is very useful in the output layer since it restricts the network output to (0, 1) and thus enables an interpretation as probability. The parameter β can be chosen as part of the training progress, but the choice $\beta = 1$ is suitable for most of the applications [63].

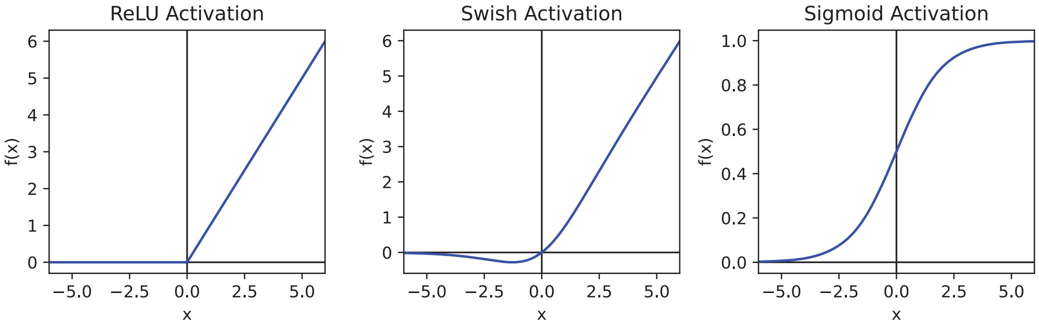


Figure 5.2: The ReLU, Swish, and Sigmoid function as examples for the neurons activation function.

5.3 Loss Functions

For the supervised training of a network, a loss function must be chosen to quantify the agreement between the network output and the target values for the training data. In general, the maximum likelihood approach is used for this purpose [63]. Thus, the loss function is given by the cross-entropy between the training data distribution and the distribution of the network output.

The loss function derived by the maximum likelihood is dependent on the model and the task it should perform. For the binary classifications used in this thesis, the true classifications in the training dataset $y_i \in \{0, 1\}$ are compared to the network score $p_i \in (0, 1)$ assigned to these events. As shown in Ref. [65], the binary cross-entropy (BCE) for the binary classification is given by the BCE loss

$$L_{\log}(\boldsymbol{y}, \boldsymbol{p}) = -\sum_{i}^{N} \left(y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i) \right). \tag{5.5}$$

As shown in figure 5.3, minimizing the BCE loss for each event results in the correct classification. Thus, the minimization of the mean BCE loss in equation (5.5) leads to an overall improved agreement between the true classification $\boldsymbol{y}$ of the training data and the corresponding network output $\boldsymbol{p}$.

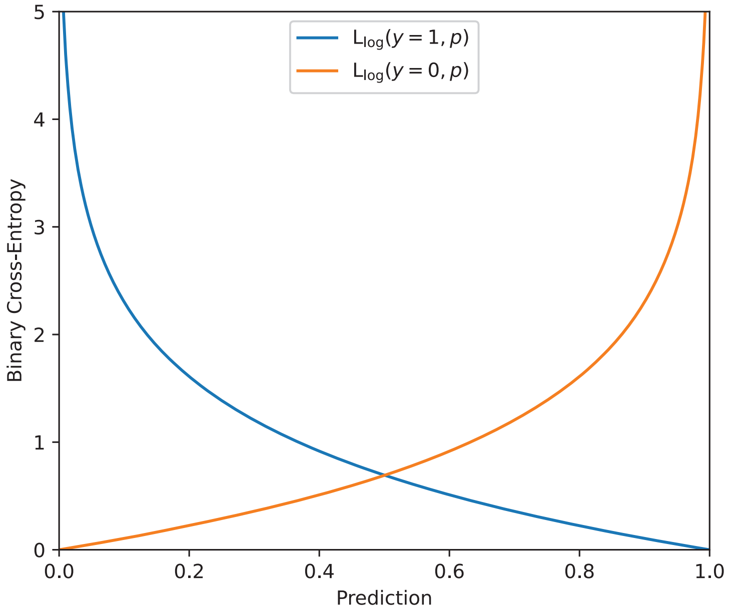


Figure 5.3: The BCE loss used for the binary classification with $y_i \in \{0, 1\}$

5.4 Training Neural Networks

As described in the previous section, a network is trained by minimizing a corresponding loss function. Since networks usually consist of thousands to millions of trainable parameters and require huge sets of training data, the minimization process is not trivial.

5.4.1 Stochastic Gradient Descent

The training of a neural network begins with a random initialization of its trainable parameters $\boldsymbol{\theta}$. Starting with this random initialization the minimum of the loss function $L(\boldsymbol{x}, \boldsymbol{\theta})$ is searched for the given training data $\boldsymbol{x}$ by the method of gradient descent [63]. By calculating the derivative $\nabla_{\boldsymbol{\theta}} L(\boldsymbol{x}, \boldsymbol{\theta})$ a new set of training parameter

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} - \eta \nabla_{\boldsymbol{\theta}} L(\boldsymbol{x}, \boldsymbol{\theta}_{t-1}) \tag{5.6}$$

can be chosen which results in a lower training loss. The step size in the training parameter space is given by the learning rate η, which is generally set to a small value. Repeating this gradient descent moves the set of trainable parameters towards the nearest minimum in the loss function. The global minimum does not necessarily have to be found, as a sufficiently small loss can already lead to a well-performing network. However, poorly performing minima should be avoided.

For additive loss functions like equation (5.5) the computational cost of gradient descent scales linearly with the size of the training dataset. A large training set is required for sufficient training, but this would blow up the training effort. This problem can be solved by using the stochastic gradient descent. Instead of using the whole training dataset to calculate the next step in the training parameter space only a subset, a so-called minibatch, of the training data is used. This mini-batch is drawn evenly from the training set and usually consists of 1 to several hundred events. By only computing the expectation of gradient descent on a small minibatch, the trainable parameters can updated more rapidly leading to a much faster convergence for most optimization algorithms. Also increasing the training set while introducing redundancy does not affect the convergence time of the training process if a fixed minibatch size is used. In addition, the use of minibatches also introduces some noise in the training process which decreases the probability of stopping in small local minima [63].

5.4.2 Adam Optimization

Directly using the stochastic gradient descent to optimize the training parameter does not result in a smooth training behavior. Due to the small size of the minibatches the derivative $\nabla_{\boldsymbol{\theta}} L(\boldsymbol{x}, \boldsymbol{\theta})$ shows large variations dependent on the minibatch. Thus, the direction of the updating in the training parameter space has huge fluctuations and does not smoothly lead to the minimum. This problem is solved by the *Adaptive Moment Estimation* optimizer (Adam) [66]. The Adam optimization combines the momentum method with *Root Mean Square Propagation* (RMSProp) which are described below. All of these methods rely on the calculation of the gradient descent

$$\boldsymbol{g}_t = \nabla_{\boldsymbol{\theta}} L(\boldsymbol{x}, \boldsymbol{\theta}_{t-1}). \tag{5.7}$$

Momentum Optimization

In Ref. [67] the idea was proposed to assign momentum to the training process. Instead of updating the position in the training parameter space, the momentum

$$\boldsymbol{m}_t = \alpha \boldsymbol{m}_{t-1} - \eta \boldsymbol{g}_t \tag{5.8}$$

is updated. α is an exponential decay factor determining the contribution of previous momenta. The training itself starts with $\boldsymbol{m}_0 = \mathbf{0}$. In each training step the parameter $\boldsymbol{\theta}$ are updated by

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{m}_t. \tag{5.9}$$

With this approach, each training step becomes dependent on the previous one, resulting in a significantly smoother and faster convergence toward the minimum of the loss function.

RMSProp

The idea of RMSProp was initially published in a Coursera lecture [68] and then taken up by later papers. So far the constant step size η resulted in a constant learning rate during the entire training. With RMSProp, this learning rate is adapted for each training parameter during the training process. The exponential moving average v_t of the element-wise product over θ $\boldsymbol{g}_t \odot \boldsymbol{g}_t$ is introduced which gets updated in each training step by

$$\boldsymbol{v}_t = \beta \boldsymbol{v}_{t-1} + (1 - \beta) \boldsymbol{g}_t \odot \boldsymbol{g}_t \tag{5.10}$$

with the exponential decay parameter β [69]. At the beginning of the training $\boldsymbol{v}_t$ is initialized to $\boldsymbol{v}_0 = \mathbf{0}$ [66]. For the next training step

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} - \frac{\eta}{\sqrt{\boldsymbol{v}_t} + \epsilon} \odot \boldsymbol{g}_t \tag{5.11}$$

the gradient descent is element-wise divided by $\sqrt{\boldsymbol{v}_t}$. Usually, a very small parameter ϵ is introduced to avoid a divergence. The RMSProp method increases the learning rate for a specific training parameter if the corresponding gradient descent is rather small over the previous training steps. This leads to faster convergence in flat regions of the loss function and smaller training steps in regions with higher loss changes.

Adam Algorithm

These two approaches are combined in the algorithm of the Adam optimizer introduced in Ref. [66]. The full algorithm is described in table 5.1 and the list of necessary parameters for the optimization is given in table 5.2. These parameters have to be chosen before the training and determine the training process.

The algorithm in table 5.1 shows that the Adam optimization is not only a combination of the RMSProp with momentum. The procedure is extended by the calculation of the bias-corrected moments $\hat{\boldsymbol{m}}_t$ and $\hat{\boldsymbol{v}}_t$. This bias arises if one calculates the exponential moving momentum of a variable k_t by

$$\begin{aligned} k_t &= \beta k_{t-1} + (1-\beta) f(g_t) \\ &= (1-\beta) \sum_{i=1}^{t} \beta^{t-i} f(g_i). \end{aligned} \tag{5.12}$$

The expected value of this moving average results in

$$\begin{aligned} \mathbb{E}[k_t] &= \mathbb{E}\left[(1-\beta) \sum_{i=1}^{t} \beta^{t-i} f(g_i)\right] \\ &= \mathbb{E}[f(g_t)](1-\beta) \sum_{i=1}^{t} \beta^{t-i} \\ &= \mathbb{E}[f(g_t)](1-\beta^t) \end{aligned} \tag{5.13}$$

if g_i is stationary. This bias correction of $\hat{\boldsymbol{m}}_t$ and $\hat{\boldsymbol{v}}_t$ results in more stable training as shown in Ref. [66].

Table 5.1: Algorithm of the Adam optimizer [66]

Initialization
$\boldsymbol{m}_0 = \boldsymbol{0}$
$\boldsymbol{v}_0 = \boldsymbol{0}$
Updating in the t-th training step
$\boldsymbol{g}_t = \nabla_{\boldsymbol{\theta}} L(\boldsymbol{x}, \boldsymbol{\theta}_{t-1})$
$\boldsymbol{m}_t = \beta_1 \boldsymbol{m}_{t-1} + (1-\beta_1)\boldsymbol{g}_t$
$\boldsymbol{v}_t = \beta_2 \boldsymbol{v}_{t-1} + (1-\beta_2)\boldsymbol{g}_t \odot \boldsymbol{g}_t$
$\hat{\boldsymbol{m}}_t = \boldsymbol{m}_t/(1-\beta_1^t)$
$\hat{\boldsymbol{v}}_t = \boldsymbol{v}_t/(1-\beta_2^t)$
$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} - \frac{\eta}{\sqrt{\boldsymbol{v}_t}+\epsilon} \odot \boldsymbol{m}_t$

Table 5.2: Parameters of the Adam optimization and their recommended defaults [66]

Parameter	Description	Recommended default value
η	Learning rate	0.001
$\beta_1 \in [0,1)$	Exponential decay for first moment estimate	0.9
$\beta_2 \in [0,1)$	Exponential decay for second moment estimate	0.999
ϵ	Stabilizer for the division operation	10^{-8}

The Adam algorithm is a simple, computationally efficient optimization procedure requiring little memory. Ref. [66] has shown that the Adam method is robust and well-performing across different fields of machine learning. Therefore, this algorithm is used for the training of neural networks in section 8.1 and chapter 9 with a focus on the correct choice of the optimizer parameters in table 5.2.

5.4.3 Back Propagation for FFNs

As described in the previous sections, the core idea of training a network is to optimize the trainable parameters according to the derivative $\nabla_{\boldsymbol{\theta}} L(\boldsymbol{x}, \boldsymbol{\theta})$ of the loss function. However, the calculation of this derivative for a specific weight or bias in a multilayer feedforward network has not yet been explained. Since the network usually has thousands of connections between the neurons, this problem seems quite complicated but can be solved using the chain rule for derivatives. For an FFN with n layers, the output of the j-th neuron in the k-th layer is given by

$$y_j^{(k)} = a^{(k)} \left(\sum_{i}^{N^{(k)}} w_{ji}^{(k)} y_i^{(k-1)} + \theta_j^{(k)} \right). \tag{5.14}$$

To optimize the training weights $w_{ji}^{(n)}$ and $\theta_j^{(n)}$ in the output layer of the network the derivative

$$\frac{\partial L}{\partial w_{ji}^{(n)}} = \frac{\partial L}{\partial y_j^{(n)}} \frac{\partial y_j^{(n)}}{\partial w_{ji}^{(n)}} \tag{5.15}$$

has to be calculated for the gradient descent. If the loss and the activation functions are differentiable and their corresponding derivatives are known this calculation is straightforward. For the updating of the weight $w_{ji}^{(n-1)}$ in the $(n-1)$-th layer, the chain rule can be used to calculate the derivative

$$\begin{aligned} \frac{\partial L}{\partial w_{ji}^{(n-1)}} &= \sum_{l}^{N^{(n)}} \frac{\partial L}{\partial y_l^{(n)}} \frac{\partial y_l^{(n)}}{\partial w_{ji}^{(n-1)}} \\ &= \sum_{l}^{N^{(n)}} \frac{\partial L}{\partial y_l^{(n)}} \frac{\partial y_l^{(n)}}{\partial y_j^{(n-1)}} \frac{\partial y_j^{(n-1)}}{\partial w_{ji}^{(n-1)}} . \end{aligned} \tag{5.16}$$

In the same way, the derivatives for the remaining weights and biases in the front layers can be calculated by multiplying and summing the partial derivatives $\partial y_i^{(k)} / \partial y_j^{(k-1)}$. This procedure is called back-propagation since it describes the flow of information from the loss function backwards through the network.

For the training process, the derivative of the loss function has to be calculated for every trainable parameter. Thus, the same partial derivatives would be calculated several times leading to a potential waste of computational power. By storing the results of these subexpressions the computational effort for the partial derivatives scales only linearly with the number of connections between the neurons. More details about the technical implementation via graphs and tensor operations can be found in Ref. [63].

5.5 Network Hyperparameters

The previous section explained the training procedure to optimize the trainable parameters. However, there exists an additional set of parameters that is not adapted during training, called hyperparameters. These hyperparameters affect the architecture of the network and its training behavior. Examples of hyperparameters are the number of layers and nodes per layer defining the network's size, and the choice of the activation function in each layer. To train a network using the Adam algorithm, the step size η, the stabilizer ϵ, and the exponential decays β_1 and β_2 must be chosen before the training. If regularization, explained in section 5.6.2, is used, their parameters are also counted as hyperparameters for the training process. All in all, there can be many possible hyperparameters when training a machine learning algorithm. Since they are not adapted during training but can significantly impact the training progress, choosing suitable hyperparameters is crucial for successful training. Therefore, an extensive algorithm for hyperparameter optimization is used in chapter 9. In most cases, the optimization of hyperparameters focuses on the generalization of the network, which is explained in the next section.

5.6 Generalization

So far, the overall architecture of a neural network and its training process have been described. During the training, the network optimizes its trainable parameters to minimize the loss for the training data. However, the application of machine learning generally focuses rather on the design and training of a model that has a good performance on data not used for the training. Thus, the generalization error is introduced in addition to the training error represented by the training loss. This generalization error indicates the model performance on unseen inputs.

5.6.1 Overtraining and Undertraining

The generalization error [63] can be measured on a validation dataset not used for training if the training and validation datasets

- are independent of each other
- and follow the same underlying probability distribution.

Since the trainable parameters are adapted to the training dataset, the model is expected to perform better on the training dataset than on the validation dataset. Thus, successful training is usually defined by a distinct minimization of the training loss while keeping the gap to the validation loss as small as possible. An insufficient minimization of the training loss is called undertraining since the trainable parameters cannot adapt well enough to adequately solve the problem. If minimizing the training loss results in an increasing validation loss, the trainable parameters can adapt well to the training data but only with a poor generalization which is called overtraining. The risk of over- and undertraining is determined by the capacity of the model. The capacity of a network is characterized by the variety of functions it can depict. A capacity that is too low leads to undertraining since the network is not able to fit the underlying probability density of the training data. On the other hand, a model with a too-high capacity tends to learn the noise in the training dataset instead of the underlying probability density, leading to overtraining. Therefore, an adequate choice of network capacity has to be made to enable successful training with good generalization.

5.6.2 Regularization

The model performance can be improved by introducing preferences into the training process. Although these regularizations can restrict the minimization of the training loss, they improve the generalization of the model [63]. Various regularization methods commonly used to modify the training algorithm exist. In this thesis, early stopping, dropout, and batch normalization are used to improve the generalization of the neural networks.

Early Stopping

As explained in the previous section, the model's actual performance is not necessarily improved by a smaller training loss since the generalization can worsen due to overtraining. Thus reaching the minimum in the training loss does not correspond to the optimal performance for later application. This fact raises the question of when one should stop the training process. This problem can be solved by monitoring the loss on the validation data during the training process. The trainable parameters are adapted to the minibatches using the stochastic gradient descent. One run over all minibatches of the training dataset is called a training epoch. At the end of an epoch, the loss is evaluated for the validation dataset before the next training epoch is started. In figure 5.4, the training process of an exemplary neural network trained on ATLAS Open Data [70] is shown. The model is chosen to have a capacity that is too high for the given training data to illustrate the problem better. While the training loss decreases continuously, the validation loss decreases at the beginning of the training but increases again after a specific training duration. During the training process, the trainable parameters are stored after each epoch. This allows to restore the model with the best validation loss after the training, resulting in a model with the optimal generalization. Since the validation loss is expected to have a U-shape the training is stopped if the validation loss has not improved for a fixed number of epochs, called patience. This approach of stopping the training before the training loss has converged and restoring the best-performing state is referred to as early stopping.

Early stopping is one of the most commonly used forms of regularization to reduce overtraining. Besides its simplicity and effectiveness, it can be implemented without affecting the architecture of the model or the training procedure. With early stopping the effective capacity of the model is adapted by restraining the training parameter space to the neighborhood of their initialization [63].

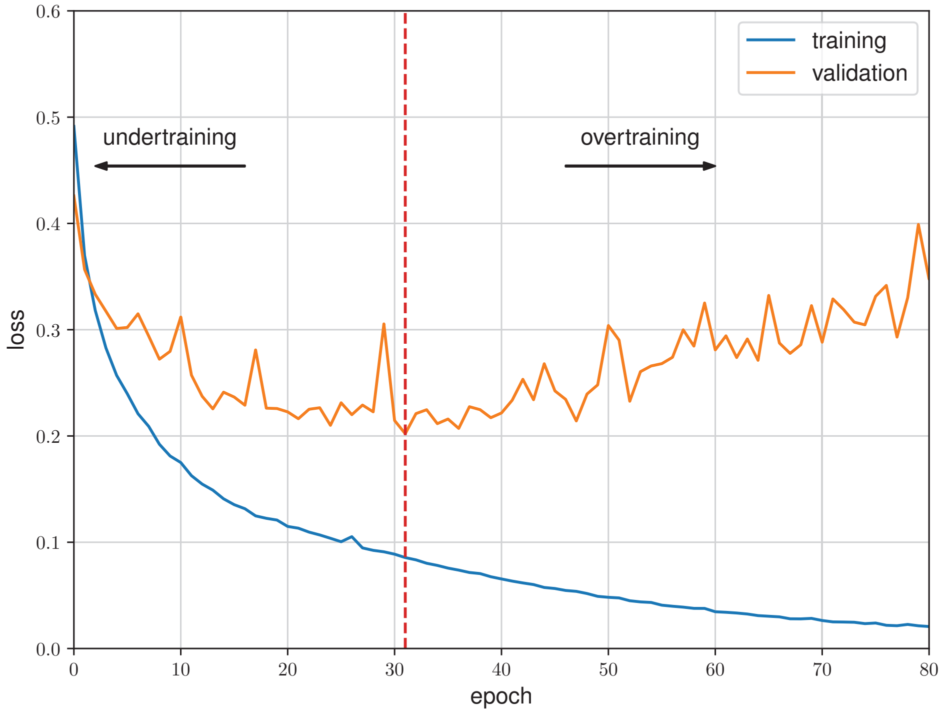


Figure 5.4: Training and validation loss of a network trained to separate Higgs ($H \to ZZ \to l_0^+ l_0^- l_1^+ l_1^-$) and background events with four leptons in the final state. The training data was provided by the ATLAS Open Data [70]

Dropout

Another approach to avoid overtraining is to train several models of different architecture or on different training datasets and average over their predictions. Unfortunately, this is very computationally intensive since they all have to be trained and evaluated on the validation data and require well-performing hyperparameters. A more efficient approach to introduce this averaging is dropout [71]. During the training, neurons get randomly dropped leading to thinned sub-networks. For the application on data not used for training, all neurons are active. Thus, a model with n neurons can be interpreted as a combination of 2^n potential sub-networks. By defining a dropout rate p_{drop}, the activation of a neuron in the k-th layer during the training is given by

$$y_j^{(k)} = a^{(k)} \left(\frac{1}{1 - p_{\text{drop}}} \sum_i^{N^{(k)}} w_{ji}^{(k)} r_i^{(k)}(p_{\text{drop}}) y_i^{(k-1)} + \theta_j^{(k)} \right) . \tag{5.17}$$

The Bernoulli random variable $r_i^{(k)}(p_{\text{drop}}) \in \{0, 1\}$ has the probability p_{drop} of being 0 and gets evaluated once per minibatch. Since this would reduce the mean neuron input during the training by $(1 - p_{\text{drop}})$ the correction factor $1/(1 - p_{\text{drop}})$ is applied [71]. This ensures that the expected output of a neuron is the same for the training with dropout and the later application without dropout.

Dropout introduces some noise into the neural network while training. Each neuron in a network trained with dropout is expected to work with a randomly chosen set of connected neurons and cannot rely only on the connection to one specific neuron. Therefore, this neuron has to learn a feature of all the incoming neurons and thus becomes more robust.

The exact rate of dropout that is helpful depends heavily on the amount of training data and the network size. However, it was shown in Ref. [71] that a higher number of neurons is favored for networks trained with dropout than for training without dropout.

Batch Normalization

During training, all trainable parameters get updated simultaneously according to their corresponding gradient descent. Thus, the parameters of a neuron in the k-th layer get updated according to the current input to this neuron. However, the input of the neuron is also changed due to the updating of the parameters in the previous layers. This change in the input distribution of the neurons during training is referred to as an internal covariate shift. In Ref. [72], batch normalization is introduced to solve this problem. Batch normalization is the approach to gain a stable neuron input distribution by normalizing its mean and variance. For each minibatch $\mathbb{B}$ the inner activation

$$x_j^{(k)} = \sum_i^{N^{(k)}} w_{ji}^{(k)} y_i^{(k-1)} + \theta_j^{(k)} \tag{5.18}$$

of the neuron is evaluated to calculate its corresponding mean

$$\mu_j^{(\mathbb{B},k)} = \frac{1}{N_{\mathbb{B}}} \sum^{N_{\mathbb{B}}} x_j^{(k)} \tag{5.19}$$

and variance

$$\sigma_j^{(\mathbb{B},k)} = \sqrt{\frac{1}{N_{\mathbb{B}}} \sum^{N_{\mathbb{B}}} \left(x_j^{(k)} - \mu_j^{(\mathbb{B},k)} \right)^2}. \tag{5.20}$$

Thus, a normalized inner activation

$$\hat{x}_j^{(k)} = \frac{x_j^{(k)} - \mu_j^{(\mathbb{B},k)}}{\sqrt{\left(\sigma_j^{(\mathbb{B},k)}\right)^2 + \epsilon}} \tag{5.21}$$

can be calculated for the corresponding minibatch. The parameter ϵ is used to ensure numerical stability. To account for this restriction of the inner neuron activation two new trainable parameters $\gamma_j^{(k)}$ and $\beta_j^{(k)}$ are introduced in the calculation of the neuron output:

$$y_j^{(k)} = a^{(k)} \left(\gamma_j^{(k)} \hat{x}_j^{(k)} + \beta_j^{(k)} \right) \tag{5.22}$$

With these new trainable parameters, the network is able to represent the same functions as without batch normalization but with improved training behavior. The mean and variance of the inner neuron activation can be modified directly during the training instead of being a complicated result of the trainable parameters of all previous layers. Since the impact of the bias $\theta_j^{(k)}$ in equation (5.18) becomes negligible due to the batch normalization this training parameter can be dropped [63]. The purpose of the bias can be fully covered by the trainable bias $\beta_j^{(k)}$ of the batch normalization.

The use of batch normalization improves the training process and enables higher step sizes speeding up the training. Since the problem of internal covariate shift is significantly reduced the training of deeper networks becomes more stable [72].

During the training, the moving averages

$$\mu_j^{(k)} \leftarrow m\mu_j^{(k)} + (1-m)\mu_j^{(\mathbb{B},k)} \tag{5.23}$$

and

$$\sigma_j^{(k)} \leftarrow m\sigma_j^{(k)} + (1-m)\sigma_j^{(\mathbb{B},k)} \tag{5.24}$$

get updated after each training step with the moving average momentum m.

5.7 Model Evaluation by the ROC-AUC

The networks trained in this thesis focus on binary classification. A metric for the network performance has to be chosen to optimize the hyperparameters of the classifier. If the goal is to separate two datasets, e.g., signal and background events, the receiver operating characteristic (ROC) [73] is often used. Assuming two categories in the data labeled $p \in \{0, 1\}$, the classifier is trained to assign a score y between 0 and 1 to each sample. To receive a purely binary classification from this score a threshold m_{thresh} can be chosen to split the data at $y = m_{thresh}$. For this split, a true-positive rate

$$tp(m_{thresh}) = \frac{N(p = 1 \wedge y \geq m_{thresh})}{N(p = 1)} \tag{5.25}$$

can be derived. The true-positive rate gives the proportion of samples with $p = 1$ that are correctly assigned to this category by $y \geq m_{thresh}$. The false-positive rate

$$fp(m_{thresh}) = \frac{N(p = 0 \wedge y \geq m_{thresh})}{N(p = 0)}, \tag{5.26}$$

on the other hand, represents the proportion of samples with $p = 0$ that are falsely sorted into this category. A successful classification is characterized by a high true-positive rate tp and a small false-positive rate fp. Varying the threshold m_{thresh} can lead to different values for fp and tp representing the ROC graph (fp, tp). The point ROC $= (0, 1)$ corresponds to a perfect classifier classifying all samples correctly. By setting m_{thresh} e.g. to 0 all samples end up in the category $y \geq m_{thresh}$ resulting in ROC $= (1, 1)$. ROC $= (0, 0)$ can be reached by setting m_{thresh} higher than all scores y seen in the dataset. By varying the threshold between 0 and 1 the ROC can be displayed as function $tp(fp)$. In figure 5.5, the ROC curves for different classification score distributions are shown. The shown score distributions were sampled only for illustration purposes and were not produced by actual networks. The straight line from $(0, 0)$ to $(1, 1)$ represents the ROC curve of a random classifier that is expected to produce the same value for the true-positive and false-positive rate for a given m_{thresh}. Thus, the ROC curve of a useful classifier with $tp > fp$ tends towards the upper left corner of the perfect classification ROC $= (0, 1)$. The area under the curve (AUC) is used to compare the ROC curve of multiple networks. The maximal ROC AUC of a classifier is one, representing a perfect classifier. The lower bound for a meaningful classifier is given by a ROC AUC of 0.5, indicating on average random classification. The ROC AUC can be interpreted as the probability that the classifier correctly ranks two randomly chosen samples of the two categories $p = 0$ and $p = 1$.

One can argue whether comparing networks by the ROC AUC gives an advantage over comparing their BCE losses. If the purpose of the classification is merely to separate two categories, the ROC AUC is the better metric. In figure 5.5, three example distributions are shown with their corresponding binary cross-entropy (BCE) and ROC AUC. The separation between the two data classes improves from example 1 to example 3 but the BCE worsens. This can be explained by the BCE in equation (5.5) depending primarily on the difference between the output score y and the true label p for each sample. However, the ROC curve indicates the models ability to correctly rank the samples in the output score independently of their exact value. Thus, the ROC AUC is a more direct metric of how well the two distributions are separated from each other.

5.8 *k*-Fold Method

In this thesis, the evaluation of the network performance and the subsequent network application are done via the so-called *k*-fold method. This method is based on the choice of distinct datasets:

- **Training dataset**
  The data used to adapt the trainable parameters of the network
- **Validation dataset**
  The data not used for training but to evaluate the network performance
- **Test dataset**
  The data used neither for the training nor for the evaluation of the model

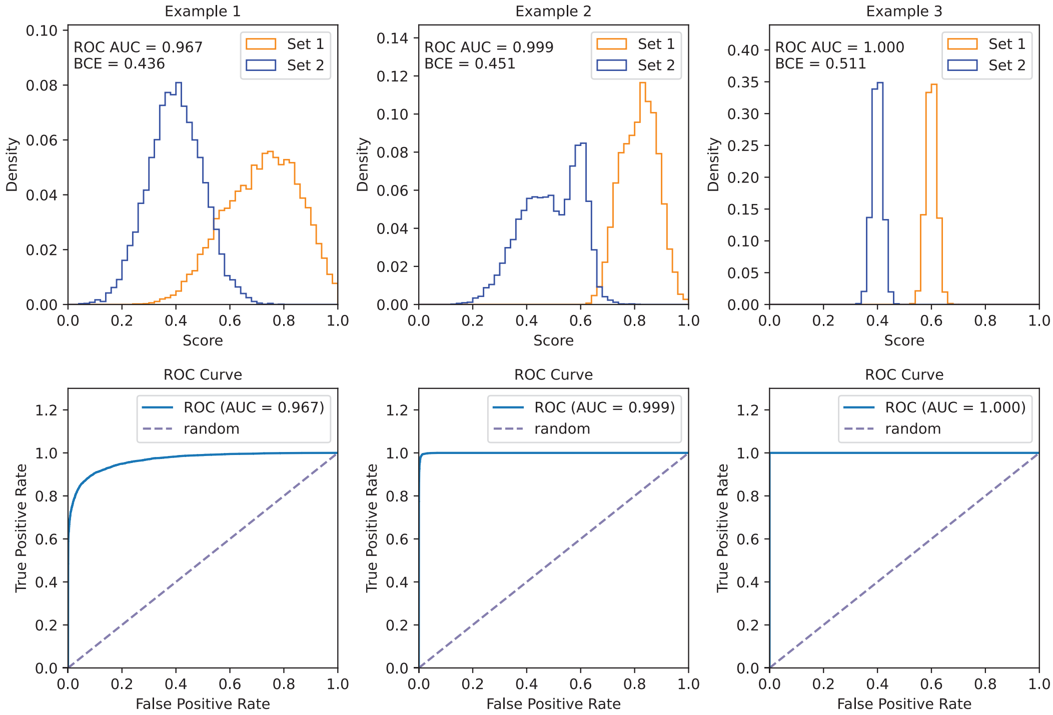


Figure 5.5: Three exemplary distributions of classifier scores separating two datasets. For each of the examples, the binary cross-entropy and the ROC-AUC are calculated. The ROC curves are shown below the distributions.

In this thesis, the available data is split into 60% training, 20% validation, and 20% test data. The distribution of the data is achieved by an initial split into five sets (folds) of equal size. The definition of a fold is based on the event number assigned to each measured or simulated event. The event number serves as an index and is independent of the physical properties of the event. Therefore, splitting the data by $((\text{EventNumber} - i_{\text{Fold}}) \mod 5) = 0$ with the fold index $i_{\text{Fold}} \in [0, 1, 2, 3, 4]$ leads to five uniform and reproducible folds. Three of these folds are used for training, one for validation and one as test dataset.

The performance of the model is evaluated on the validation dataset to find suitable hyperparameters. However, the evaluation of the network performance is limited by the size of the validation dataset. A common approach to better estimate the network performance is the k-fold cross-validation. As described above the available data is split into five folds. In the k-fold method, five networks are trained with overlapping training folds and non-overlapping validation datasets. Thus, each of the networks is evaluated on statistically independent validation data. This allows the calculation of a mean network performance and the estimation of the corresponding statistical uncertainty of the performance. Therefore, k-fold cross-validation will be used for the hyperparameter optimization in chapter 9.

Besides the network performance evaluation, the k-fold method is crucial in this analysis to avoid biases in the model application. When applying DNN algorithms, the different performances on training and unseen data must be considered. Even if overtraining has been avoided, classification on training data is expected to perform slightly better than on unseen data. Thus, if the DNN predictions on the training dataset are used to compare to predictions on measured data, an additional systematic mismodeling is expected. In order to avoid this mismodeling, the k-fold method is used in this analysis with five folds. As shown in figure 5.6 and described above, the data is split into five sets: three used for training, one used for validation, and one test set. By iterating these sets, the DNN algorithm is trained five times with non-overlapping validation and test sets. The test dataset is neither used for training nor evaluating the training performance and is therefore not involved in any part of the training or hyperparameter optimization. Thus, the DNN predictions obtained using testing datasets can be assumed to be identically distributed as the predictions on measured data. To get predictions for all events in the dataset, the prediction is calculated using the DNN where the respective events were part of the testing dataset. Thus, predictions are made only on events not used for the respective training or optimization. This method will be used to apply a multivariate correction in section 8.1 and for the application of the polarization discriminators trained in chapter 9. The consistency of DNNs trained on different folds is investigated for the polarization discriminators in section 10.1.

The k-fold application method is an established strategy for multivariate ATLAS analyses. For example, the $W^{\pm}Z$ polarization analysis in Ref. [4] uses a 2-fold method to apply the discriminator networks. This thesis chooses 5 instead of 2 folds to increase the fraction of data used for training.

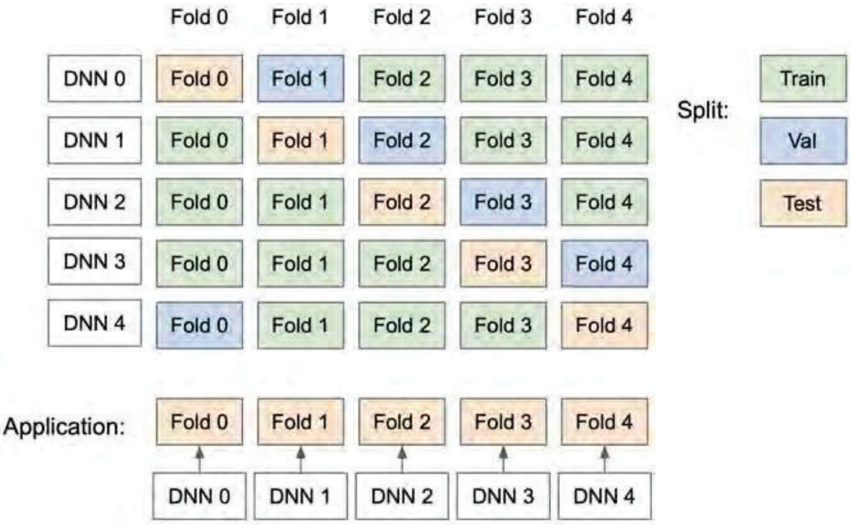


Figure 5.6: The concept of the k-fold method with $k = 5$ folds and the split intro training, validation and test data. The final DNNs are applied to their corresponding test dataset.

5.9 Access Feature Importance via SHAP

Neural networks are so-called black boxes since their number of trainable parameters is too large to understand the decisions they make in detail. Nevertheless, understanding the behavior of these black boxes can further validate their functionality and enable conclusions to be drawn about the concepts learned during training. Drawing a direct connection between the model input and its corresponding decision becomes possible with SHapley Additive exPlanations (SHAP) published in Ref. [74]. The following explanation of SHAP values is based on this publication.

A much simpler explanation model g_x is used to better understand the decision made by a complex model f. This explanation model g_x explains the complex decision $f(\boldsymbol{x})$ locally for one specific vector $\boldsymbol{x}$. For the explanation model, simplified inputs $\boldsymbol{x}'$ are defined which are connected to the original input via a mapping function $\boldsymbol{x} = h_x(\boldsymbol{x}')$. For N input features the simplified input $\boldsymbol{x}' \in \{0,1\}^N$ indicates whether a feature in x is included in the model. The explanation model for an input $\boldsymbol{x}$ tries to model $g_x(\boldsymbol{z}') \approx f(h_x(\boldsymbol{z}'))$ close to the corresponding input value with $\boldsymbol{z}' \approx \boldsymbol{x}'$. To have an explanation model with interpretable parameters, an additive feature attribution model

$$g_x(\boldsymbol{z}') = \phi_0 + \sum_{i=1}^{N} \phi_i z'_i \tag{5.27}$$

is used with the trainable impact parameters $\phi_i \in \mathbb{R}$. Figure 5.7 shows for one individual network application how these parameters ϕ_i add up to the final decision of the SHAP model. The shown model, $\text{DNN}_{W^\pm W^\pm}$, is trained in chapter 9 to distinguish $W^\pm W^\pm jj$-EW processes from background contributions. The concept of this SHAP model is explained in the following.

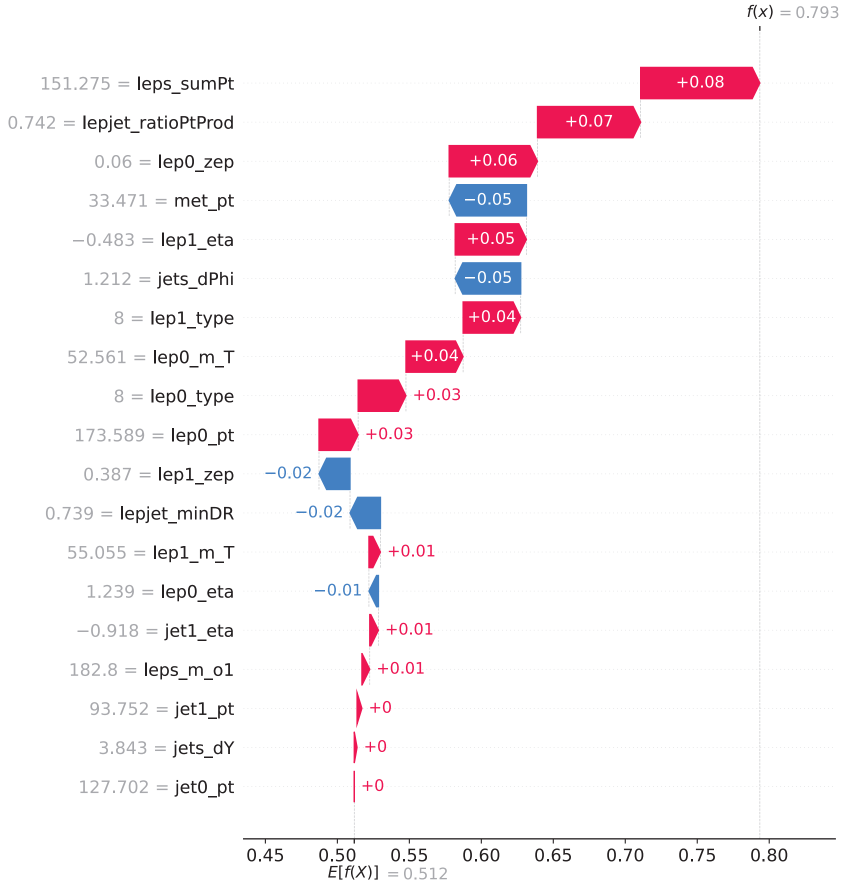


Figure 5.7: The SHAP values of $\text{DNN}_{W^\pm W^\pm}$ trained in chapter 9 are calculated for an exemplary event. The input features are listed with their value for the specific event and are ranked according to their impact on the specific network decision. The output value $E[f(X)]$ is the expectation value of the network output.

Since multiple definitions of g_x are possible, three additional requirements are chosen for this additive explanation model:

- **Local accuracy**
  The output of the explanation model g_x for the corresponding input x has to match the decision of the complex model:

$$f(\boldsymbol{x}) = g_x(\boldsymbol{x}') \tag{5.28}$$

- **Missingness**
  The simplified input $\boldsymbol{x}'$ indicates the feature presence and the parameter ϕ_i the feature impact on the decision. The property of missingness requires missing feature presence with $x_i' = 0$ to have a vanishing impact $\phi_i = 0$:

$$x_i' = 0 \Rightarrow \phi_i = 0 \tag{5.29}$$

- **Consistency**
  If a model $f(x)$ is slightly changed to a model $f'(x)$ and the impact of a specific feature increases independently of the other features, the feature impact ϕ_i should not decrease

for $f'(x)$. With $\boldsymbol{z}' \setminus i$ denoting $z'_i = 0$ this property can be written as:

$$\begin{aligned} f'(h_x(\boldsymbol{z}')) - f'(h_x(\boldsymbol{z}' \setminus i)) \geq f(h_x(\boldsymbol{z}')) - f(h_x(\boldsymbol{z}' \setminus i)) \qquad \forall \boldsymbol{z}' \in \{0,1\}^N \\ \Rightarrow \phi_i(f', x) \geq \phi_i(f, x) \end{aligned} \tag{5.30}$$

These three conditions are only fulfilled for Shapley values [75] developed for game theory. The Shapley regression values can be used to calculate the impact parameters ϕ_i of the additive model in equation (5.27) to describe a complex model f while fulfilling the requirements in equations (5.28) to (5.30). The Shapley regression model requires a retraining of the model f on the subsets $S \subset F$ of the original set of training features F. To get the impact of the feature x_i, a model $f_{S\cup\{i\}}$ trained on a set including this feature is compared to a model f_S not trained on this feature. Due to the interplay of various features for the model prediction, this comparison has to be done for all possible feature subsets $S \subseteq F \setminus \{i\}$. The final Shapley regression value is the weighted sum of the differences in the output of the retrained models:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \left(f_{S\cup\{i\}}(x_{S\cup\{i\}}) - f_S(x_S) \right) \tag{5.31}$$

However, most models use a fixed set of input features, which cannot be reduced without changing the network. Therefore, a conditional approximation function $E[f(\boldsymbol{z})|\boldsymbol{z}_S]$ is introduced that approximates the original model $f(\boldsymbol{z})$ for a reduced set of input features $\boldsymbol{z}_S$. The Shapley values of this approximation function are called SHAP values and follow the principles of local accuracy, missingness, and consistency. They are used in section 10.3 to investigate the decision-making of the classifier networks trained in this thesis.

Part II

Polarization Analysis

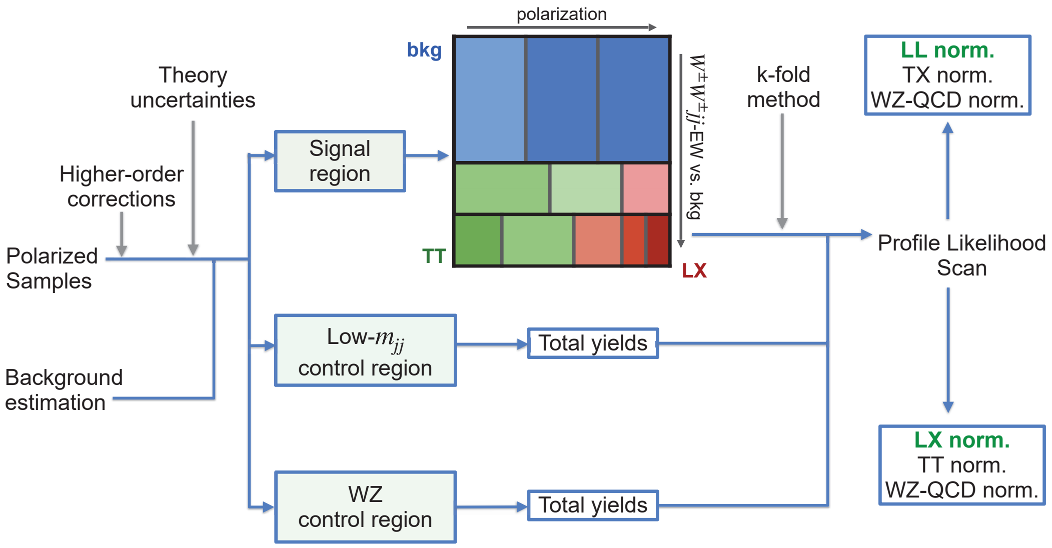

polarization
bkg
Theory uncertainties
Higher-order corrections
Polarized Samples
Background estimation
Signal region
TT
LX
$W^\pm W^\pm jj$-EW vs. bkg
k-fold method
LL norm.
TX norm.
WZ-QCD norm.
Profile Likelihood Scan
Low-m_{jj} control region
Total yields
WZ control region
Total yields
LX norm.
TT norm.
WZ-QCD norm.

Figure 5.8: Visualization of the analysis strategy

The strategy of the polarization analysis is visualized in figure 5.8. The actual order of the associated chapters partially deviates from this visualization to account for the connections between the analysis steps. This second part of the thesis starts with chapter 6, which introduces the selection criteria for objects and events measured at the ATLAS detector. These selection criteria are chosen according to the differential $W^\pm W^\pm jj$ analysis in Ref. [8] and lead to three distinct phase spaces: the signal region, the Low-m_{jj} control region, and the $W^\pm Z$ control region. The Standard Model prediction in these phase spaces is summarized in chapter 7. Due to the matching phase space definitions, the estimation of the background contributions is consistent with the differential $W^\pm W^\pm jj$ analysis in Ref. [8]. The polarized $W^\pm W^\pm jj$-EW predictions, on the other hand, have not been studied in ATLAS experiments so far. The use of these polarized predictions requires a correction for higher-order calculations. The corresponding corrections for higher-order electroweak and QCD effects are extracted and applied in chapter 8. After correcting the prediction of the Standard Model, multivariate analysis techniques are used to access the polarization information of the events measured in the signal region. The optimization and training of the neural networks is described in chapter 9, followed by a validation of the network output and a study of the decision-making in chapter 10. The uncertainties relevant to the polarization analysis are summarized in chapter 11. Since the uncertainties of the background contributions are already known from Ref. [8], the chapter focuses on the theory uncertainties of the polarized prediction. The neural network scores are combined in the signal region to evaluate the measured data in chapter 12. This combined application produces a two-dimensional histogram separating backgrounds, transverse polarization, and longitudinal polarization. A bias from the training process of the networks is avoided by the k-fold application method. In order to constrain the background contributions, the total number of events in the Low-m_{jj} and $W^\pm Z$ control region enter the profile likelihood fit in addition to the distribution in the signal region. The likelihood fit extracts the normalization of the longitudinal polarization together with the normalization of the transverse polarization state and the $W^\pm Z$-QCD background. The measurement is performed for the single boson polarization LX and the double boson polarization LL.

Chapter 6

Object and Event Selection

The amount of data generated at the ATLAS detector goes far beyond the statistical power of other experiments. Since only a fraction of this data is expected to contain events of scientific interest, a sophisticated chain of selection criteria has been established. This chapter highlights the selection criteria for the $W^{\pm}W^{\pm}jj$-EW polarization measurement. The chosen selections are identical to those used in the published unpolarized $W^{\pm}W^{\pm}jj$-EW analysis [8]. This paper has already shown that the chosen criteria are very well suited for the $W^{\pm}W^{\pm}jj$-EW measurement and thus provide a solid basis for the polarization measurement.

6.1 Pre-selection

Since the measured collisions originate from the crossing of proton bunches at the LHC, multiple proton collisions can happen simultaneously. This results in multiple collision vertices for each bunch crossing. Thus, a choice has to be made which is the primary vertex of a collision event used for analysis. This decision is made in favor of the vertex with the highest sum of transverse track momenta squared. The remaining vertices are the so-called pileup and are handled as the background of the primary measurement. A primary vertex is only considered for the analysis if it has at least three associated tracks.

The choice of a primary vertex is followed by the object reconstruction explained in section 2.2.3 and the trigger chain described in section 2.2.4. Different trigger criteria are used for electrons and muons, whereby the requirements for object quality increase with decreasing transverse momentum. The triggers used during the individual years of Run 2 are listed in table 6.1 and further described in Ref. [76] and [77]. If at least one electron or muon originating from the primary vertex passes one of these triggers, the event is recorded for potential analyses.

Table 6.1: Electron and muon trigger used during Run 2 of the ATLAS detector [76][77].

Time period	electron trigger	muon trigger

2015	`HLT_e24_lhmedium_L1EM20VH`
	`HLT_e60_lhmedium`
	`HLT_e120_lhloose`
		`HLT_mu20_iloose_L1MU15`
		`HLT_mu50`

2016 - 2018	`HLT_e26_lhtight_nod0_ivarloose`
	`HLT_e60_lhmedium_nod0`
	`HLT_e140_lhloose_nod0`
		`HLT_mu26_ivarmedium`
		`HLT_mu50`

6.2 Object Selection

A collision at the ATLAS detector can lead to many particles emitted into the detector. These particles are measured within the ATLAS detector and reconstructed into object candidates. The object selection defines which of these reconstructed objects, which can be electrons, muons, or jets, are taken into account for the analysis.[1] The general reconstruction of these objects within the ATLAS detector is described in section 2.2.3. The object selection is divided into so-called baseline and signal objects (electrons, muons, or jets). As the name indicates signal objects fulfill the tight selection criteria to be studied in the signal region of the analysis. Baseline objects, on the other hand, are defined by looser requirements and are used to check for additional leptons or jets in the event. In addition to the directly measured objects, the missing transverse momentum p_T^{miss} (MET) is defined to enable sensitivity to the occurrence of neutrinos. In the final step, an overlap removal is performed to avoid the possibility of multiple reconstructions of the same physical object.

6.2.1 Electron Selection

Electrons originating from a collision in the ATLAS detector first pass through the inner detector and leave a track due to their charge. Subsequently, the electron is stopped in the adjacent electromagnetic calorimeter by forming an energy cluster. Thus, such a cluster associated with a track in the inner detector indicates an electron and further selection criteria can be applied. The summary of the applied selection criteria for baseline and signal electrons is listed in table 6.2.

First, it must be ensured that the electron track originates from the collision. For this purpose, the point *pca* of the track closest to the proton beam is determined. The transversal and longitudinal impact parameters are measured for this point *pca*. The transversal parameter d_0 corresponds to the transversal distance between *pca* and the beam, and the longitudinal impact parameter z_0 is the distance along the z-axis between *pca* and the primary vertex. The resulting selection of $|d_0/\sigma_{d_0}| < 5$ and $|z_0 \times \sin\theta| < 0.5\,\mathrm{mm}$ ensures a high probability that the electron track originates from the collision at the primary vertex.

The quality of the electron reconstruction is determined by a likelihood-based identification using measurements from the tracking and calorimeter system. The analysis of the shower shape and the matching between track and shower leads to the four operating points `VeryLooseLH`, `LooseLH`, `MediumLH`, and `TightLH`, which are described in detail in Ref. [25]. The `LooseLH` identification is required for baseline electrons, while signal electrons have to fulfill the more stringent requirements of the `TightLH` identification.

For geometrical acceptance, electrons are expected to hit the detector in the $|\eta| < 2.47$ range. The detector crack at $1.37 \leq |\eta| \leq 1.52$ is excluded for signal electrons. Since the signal electrons are expected to originate from a $W^{\pm}$ boson decay, a transverse momentum higher than 27 GeV is required. In contrast, baseline electrons are selected with $p_{\mathrm{T}} > 4.5\,\mathrm{GeV}$.

To enforce a high object quality and reduce potential backgrounds the selection of signal electrons is expended by three additional requirements. The first one is the isolation which is defined by the activity in the cone $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ around the track of the electron candidate. The resulting electron-isolation working points are described in Ref. [25], with the `Gradient` working point being applied for the signal electrons in this analysis.

For a further distinction from photons, the `Author` quantity is used. This quantity is based on the presence of pixel hits, the ratio of energy and momentum, the transverse momentum of the track, and secondary vertex information. The resulting `Author` value corresponds to the degree of unambiguity in deciding whether the reconstructed object corresponds to an electron or a photon. `Author` = 1 is required for signal electrons, corresponding to a reconstruction exclusively as an electron.

The last requirement for signal electrons is motivated by the existence of the charge-flip background described in section 7.2.3. Since no single kinematic variable of the measurement is sufficient to distinguish between charge-flip electrons and electrons originating directly from the collision, a boosted decision tree is trained in Ref. [78]. The resulting Electron Charge ID Selector (ECIDS) was trained to distinguish between these electron types. Thus, a threshold on the ECIDS output for signal electrons significantly reduces the charge-flip background.

---

[1] Photons are not used in this analysis

Table 6.2: The selection criteria for baseline and signal electrons as described in Ref. [8].

	baseline	signal
Identification:	`LooseLH`	`TightLH`
Kinematic acceptance:	$p_{\mathrm{T}} > 4.5\,\mathrm{GeV}$	$p_{\mathrm{T}} > 27\,\mathrm{GeV}$
Geometrical acceptance:	$\lvert\eta\rvert < 2.47$	$\lvert\eta\rvert < 2.47$, excluding $1.37 \leq \lvert\eta\rvert \leq 1.52$
Longitudinal impact parameter:	$\lvert z_0 \times \sin\theta\rvert < 0.5\,\mathrm{mm}$	$\lvert z_0 \times \sin\theta\rvert < 0.5\,\mathrm{mm}$
Transverse impact parameter:	$\lvert\frac{d_0}{\sigma_{d_0}}\rvert < 5$	$\lvert\frac{d_0}{\sigma_{d_0}}\rvert < 5$
Isolation requirement:	-	`Gradient`
`Author` requirement:	-	1
Charge-flip rejection:	-	`ECIDS`

6.2.2 Muon Selection

In contrast to electrons, muons penetrate all layers of the ATLAS detector. Thus, the muon reconstruction [27] requires information from the track in the inner detector, hits in the electromagnetic and hadronic calorimeter, and their trace in the muon chamber. This information is combined into an identification criterion to distinguish muons originating directly from the collision and from hadron decay. The identification working point for signal muons is `Medium`, which has suitable identification efficiency and purity while the corresponding systematic uncertainties remain small. For the baseline muons, the `Loose` working point is used, which has a higher efficiency but is also affected by higher systematic uncertainties.

With $|z_0 \times \sin\theta| < 0.5\,\mathrm{mm}$ and $|\frac{d_0}{\sigma_{d_0}}| < 3$ the selection of signal muons is significantly stricter than for baseline muons that only have to fulfill $|z_0 \times \sin\theta| < 1.5\,\mathrm{mm}$ and $|\frac{d_0}{\sigma_{d_0}}| < 15$. This stricter selection based on the impact parameters leads to a higher probability that a signal muon originates from the collision at the primary vertex than in the case of baseline muons.

The kinematic acceptance of signal muons is with $p_{\mathrm{T}} > 27\,\mathrm{GeV}$ identical to signal electrons. The baseline muons, on the other hand, are accepted with a lower transverse momentum of $p_{\mathrm{T}} > 3\,\mathrm{GeV}$ than baseline electrons. The geometric acceptance of $|\eta| < 2.7$ for baseline muons is also greater than for electrons. Signal muons must be measured within $|\eta| < 2.5$.

The last difference between signal and baseline muons is their isolation. Signal muons are required to fulfill the `FixedCutPflowTight` isolation criteria. This isolation decreases the activity around the signal muon track compared to the objects selected as baseline muons.

A summary of all the object selections applied for muons is listed in table 6.3.

Table 6.3: The selection criteria for baseline and signal muons as described in Ref. [8].

	baseline	signal
Identification:	`Loose`	`Medium`
Kinematic acceptance:	$p_{\mathrm{T}} > 3\,\mathrm{GeV}$	$p_{\mathrm{T}} > 27\,\mathrm{GeV}$
Geometrical acceptance:	$\lvert\eta\rvert < 2.7$	$\lvert\eta\rvert < 2.5$
Longitudinal impact parameter:	$\lvert z_0 \times \sin\theta\rvert < 1.5\,\mathrm{mm}$	$\lvert z_0 \times \sin\theta\rvert < 0.5\,\mathrm{mm}$
Transverse impact parameter:	$\lvert\frac{d_0}{\sigma_{d_0}}\rvert < 15$	$\lvert\frac{d_0}{\sigma_{d_0}}\rvert < 3$
Isolation requirement:	-	`FixedCutPflowTight`

6.2.3 Jet Selection

The jets are reconstructed by the anti-k_t [28] and `Pflow` algorithm [29] based on information from the calorimeters and the inner detector. To increase the signal purity additional requirements are chosen for the jets summarized in table 6.4. Analog to electrons and muons, jets must also be verified whether they originate from the primary collision. Therefore, the jet vertex tagging (JVT) [79] is used for signal jets with $p_T < 60\,\text{GeV}$ and $|\eta| < 2.4$ to check whether they can be matched to the primary vertex of the collision.

As described in section 3.3.3, the two jets resulting from the $W^{\pm}W^{\pm}jj$ scattering are expected to be in the forward regions of the detector. Thus, the jets have a significantly larger geometrical acceptance with $|\eta| < 4.5$ than the selected leptons. A transverse momentum of $p_{\mathrm{T}} > 20\,\text{GeV}$ is required for baseline jets, and signal jets are selected with $p_{\mathrm{T}} > 25\,\text{GeV}$.

Table 6.4: The selection criteria for baseline and signal jets as described in Ref. [8].

	baseline	signal
Clustering:	anti-k_t algorithm [28] with $R = 0.4$	anti-k_t algorithm [28] with $R = 0.4$
Kinematic acceptance:	$p_{\mathrm{T}} > 20\,\text{GeV}$	$p_{\mathrm{T}} > 25\,\text{GeV}$
Geometrical acceptance:	$|\eta| < 4.5$	$|\eta| < 4.5$
Vertex matching:	-	JVT for $p_T < 60\,\text{GeV}$ and $|\eta| < 2.4$

6.2.4 Reconstruction of the Missing Transverse Momentum

Since the leptonic decay of $W^{\pm}$ bosons is studied, sensitivity to neutrino occurrence is crucial for this analysis. Before the two protons collide their transverse momentum is exactly zero. Due to momentum conservation the transverse momenta of all emitted particles should cancel out, but neutrinos can not be measured within the ATLAS detector. Therefore, if a significant transverse momentum is missing for complete cancellation, this is a clear indication of neutrino activity. This missing transverse momentum p_T^{miss} [80] is calculated by the transverse momenta of the baseline leptons and jets combined with soft-hadronic activity. An internal overlap removal procedure avoids the double-counting of deposits.

6.2.5 Overlap Removal

The presented lepton and jet selection can lead to double-counting of objects measured in the ATLAS detector. An overlap removal is performed to prevent an object from being reconstructed multiple times. This overlap removal is based on Ref. [81] and updated according to Ref. [82]. This procedure removes reconstructed objects in the following order:

- **Electron vs. electron**
  If two reconstructed electrons share the same track in the inner detector or have overlapping clusters in the electromagnetic calorimeter the electron with the higher p_T is kept.

- **Electron vs. muon**
  If a muon and an electron share the same track in the inner detector and the muon is tagged in the muon chamber the electron is removed. If the muon is not tagged in the muon chamber it is removed and the electron is kept.

- **Electron vs. jet**
  The jet is removed if an electron overlaps a jet within $\Delta R < 0.2$. In the case of a ΔR greater than 0.2 but smaller than $\min(0.4,\, 0.04 + \frac{10\,\text{GeV}}{p_T^e})$ the jet is kept and the electron is removed.

- **Muon vs. jet**
  For overlapping jets and muons, the jet is removed in each of the following cases:
  - The ΔR between the two objects is smaller than 0.2.
  - The jet has less than three associated tracks with $p_T > 500\,\text{MeV}$.
  - The ratio of transverse momenta $\frac{p_T^{\mu}}{p_T^{j}}$ is larger than 0.5 and the ratio of the transverse muon momentum and the jet tracks with $p_T > 500\,\text{MeV}$ is greater than 0.7.

If none of these criteria is satisfied and ΔR is smaller than $\min(0.4,\, 0.04 + \frac{10\,\text{GeV}}{p_T^e})$ the jet is kept and the muon is removed.

6.3 Event Selection

Based on the object selection described in the previous section, specific events can be selected for the polarization analysis. This event selection is done by defining kinematic regions for the measurement. The main region for the polarization measurement is the so-called signal region. To constrain background contributions for the statistical fit a Low-m_{jj} and $W^\pm Z$ control region is used.

6.3.1 Signal Region

This polarization analysis of $W^\pm W^\pm jj$-EW scattering focuses on the leptonic decays of the $W^\pm$ bosons. Thus, two signal leptons with identical charges are required in the signal region. To minimize the contamination from $W^\pm Z$ and ZZ events the event is discarded if an additional baseline lepton is reconstructed. For the case of two electrons, the background originating from Drell-Yan processes is suppressed by a Z-peak veto at $|m_{ee} - m_Z| > 15\,\text{GeV}$ and a tighter geometry selection with $|\eta| < 1.37$. The impact of low-mass Drell-Yan processes is reduced by an invariant mass requirement $m_{ll} \geq 20\,\text{GeV}$ on the leptons.

Due to the neutrinos expected from the leptonic $W^\pm$ boson decay, a minimal missing transverse momentum of $p_T^{miss} \geq 30\,\text{GeV}$ is required.

As already discussed in section 3.3.3, two highly energetic jets in the forward regions of the detector are expected for the $W^\pm W^\pm jj$-EW scattering. Thus, at least two signal jets are required in the signal region and have to satisfy a high invariant mass requirement with $m_{jj} \geq 500\,\text{GeV}$ and a geometrical separation criteria with $|\Delta y_{jj}| > 2$. In addition, the highest jet p_T has to be greater than 65 GeV and the second highest p_T jet has to fulfill $p_T > 35\,\text{GeV}$.

To reduce the contamination of top quark processes with $t \to Wb$ b-tagging is used. The neural network method described in Ref. [83] is applied. In particular, the neural network DL1r of the 85% efficiency working point is used in this analysis. If any baseline jet with $|\eta| < 2.5$ and matched by JVT is b-tagged by this neural network the event is dropped.

The event selection for the $W^\pm W^\pm jj$-EW signal region is summarized in table 6.5. Since the signal purity of this region was optimized in the unpolarized $W^\pm W^\pm jj$-EW analysis [8] it offers an excellent baseline for the polarization measurement in this signal.

Table 6.5: The event selection defining the signal region used in Ref. [8].

Exactly two signal leptons with the same electrical charge No additional baseline leptons
In the case of two electrons they have to fulfill $|\eta| < 1.37$ and $|m_{ee} - m_Z| > 15\,\text{GeV}$
$m_{ll} \geq 20\,\text{GeV}$
$p_T^{miss} \geq 30\,\text{GeV}$
At least two signal jets The highest p_T jet has to satisfy $p_T > 65\,\text{GeV}$. The second highest p_T jet has to fulfill $p_T > 35\,\text{GeV}$.
$m_{jj} \geq 500\,\text{GeV}$
$|\Delta y_{jj}| > 2$
No jet is b-tagged using the DL1r tagger with the 85% efficiency working point

6.3.2 Low-m_{jj} Control Region

The signal region is a region of high signal purity and is, therefore, the most suitable kinematic region to extract the polarization. However, the statistical fit of the polarization is limited by various backgrounds that are introduced in section 7.2. The Low-m_{jj} control region is constructed to constrain these backgrounds in the fit. The definition of this region is consistent with the signal region definition in table 6.5 but with a lower m_{jj} selection of $200\,\text{GeV} < m_{jj} < 500\,\text{GeV}$. The invariant di-jet mass m_{jj} is the most discriminant variable for the $W^{\pm}W^{\pm}jj$-EW signal against the corresponding backgrounds. The resulting Low-m_{jj} control region is thus kinematically very similar to the signal region but with a significantly lower $W^{\pm}W^{\pm}jj$-EW contribution. Thus, including this region in the fit constrains the normalization of the backgrounds.

6.3.3 $W^{\pm}Z$ Control Region

The fully leptonic decay of $W^{\pm}Z$ results in the dominant background in the signal region. Unfortunately, its simulation by Monte Carlo event generators is known from Ref. [8] to have poor modeling. Therefore, the shape and normalization of the corresponding background simulation are corrected using data measured in the ATLAS experiment. For this purpose, the $W^{\pm}Z$ control region is kinematically close to the signal region but with a third signal lepton with $p_T > 15\,\text{GeV}$. To reduce the contribution from $Z+\text{jets}$ and $Z\gamma$ a tri-lepton invariant mass of $m_{lll} > 106\,\text{GeV}$ is required. The remaining selection criteria listed in table 6.6 are consistent with the merged Low-m_{jj} control and signal region definition. Thus, the resulting $W^{\pm}Z$ region offers the opportunity to determine the $W^{\pm}Z$ background with a kinematic very similar to the $W^{\pm}Z$ background expected for the other regions.

Table 6.6: The event selection defining the $W^{\pm}Z$ control region used in Ref. [8].

Two signal leptons and an additional third signal lepton with $p_T > 15\,\text{GeV}$
In the case of two electrons they have to fulfill $\lvert\eta\rvert < 1.37$ and $\lvert m_{ee} - m_Z\rvert > 15\,\text{GeV}$
At least two leptons satisfy $m_{ll} \geq 20\,\text{GeV}$
$m_{lll} \geq 106\,\text{GeV}$
$p_T^{miss} \geq 30\,\text{GeV}$
At least two signal jets The highest p_T jet has to satisfy $p_T > 65\,\text{GeV}$. The second highest p_T jet has to fulfill $p_T > 35\,\text{GeV}$.
$m_{jj} \geq 200\,\text{GeV}$
$\lvert\Delta y_{jj}\rvert > 2$
No jet is b-tagged using the DL1r tagger with the 85% efficiency working point

6.3.4 Fiducial Signal Region

The selections described so far are applied to objects and events reconstructed by the ATLAS detector. Predicting these regions requires a simulation of the reconstruction by the ATLAS detector and its corresponding acceptance. The introduction of a fiducial signal region at the particle-level simplifies the theoretical investigation of the analysis. The particle-level fiducial signal region is kinematically close to the reconstruction-level signal region in table 6.5. The main selection criteria are listed in table 6.7. The leptons are selected with $p_T > 27$ GeV and $|\eta| < 2.5$ and dressed with prompt photons within $\Delta R < 0.1$. The jets are reconstructed by the anti-k_t algorithm with $R = 0.4$. If the resulting jet overlaps with an electron within $\Delta R < 0.4$, the electron is removed for $p_{T,e}/p_{T,\text{jet}} < 0.5$. The jet is removed if the electron carries more than 50% of the transverse jet momentum.

The resulting reference signal region makes it easier for theorists to contribute to and interpret the analysis results. In section 8.2, the fiducial signal region definition is used by the authors of Ref. [50] to provide higher-order electroweak calculations of the polarized $W^{\pm}W^{\pm}jj$-EW signal. For the final statistical evaluation in sections 12.5 and 12.6, the fiducial signal region defines the fiducial cross-sections which are one of the main results of the polarization analysis: The signal strength of the polarized $W^{\pm}W^{\pm}jj$-EW processes is extracted from the reconstruction-level signal region. The resulting signal strength is applied on the fiducial cross-section expected in the fiducial signal region. The particle-level definition of this region makes it easier for theorists to compare the predictions of their models with the measurement results.

Table 6.7: The particle-level event selection of the fiducial signal region.

Exactly two same-charged leptons with $p_T > 27$ GeV and $\lvert\eta\rvert < 2.5$ Veto on events with taus from $W^{\pm}$ in ME calculation
$m_{ll} > 20$ GeV
$\lvert m_{ee} - m_Z\rvert > 15$ GeV in the ee-channel
$p_T^{miss} \geq 30$ GeV
Two jets with $\lvert\eta\rvert < 4.5$ leading and subleading jets satisfying $p_T > 65$ GeV and $p_T > 35$ GeV Veto on b-jets
$m_{jj} \geq 500$ GeV
$\Delta y_{jj} > 2$

Chapter 7

Standard Model prediction

The analysis in this thesis aims to probe whether a longitudinal polarization predicted by the Standard Model exists. This goal is only achievable by directly comparing the Standard Model prediction and measured data. Thus, correct modeling of the prediction is crucial for the analysis. The background predictions in section 7.2 are already known from the unpolarized $W^{\pm}W^{\pm}jj$-EW analysis [8]. The only updated background is the $W^{\pm}Z$-QCD contribution. For the $W^{\pm}W^{\pm}jj$-EW signal in section 7.1, on the other hand, new polarized templates are required. Details about the Monte Carlo generated samples not used in Ref. [8] are listed in table A.1.

7.1 $W^{\pm}W^{\pm}jj$ Signal Simulation

The $W^{\pm}W^{\pm}jj$ process is simulated by Monte Carlo event generators. The list of simulated samples is given in table A.2. Two event generators, MadGraph [46] and Sherpa [54], are used to simulate the polarized $W^{\pm}W^{\pm}jj$-EW signal. The additional contribution from polarization interference is estimated separately for these two simulation approaches. The QCD-induced contribution $W^{\pm}W^{\pm}jj$-QCD and the interference between electroweak and QCD-induced production $W^{\pm}W^{\pm}jj$-INT result in a significant background to the $W^{\pm}W^{\pm}jj$-EW polarization signal.

7.1.1 $W^{\pm}W^{\pm}jj$-EW Polarization by MadGraph

The MadGraph event generator can simulate polarized vector boson scattering at leading-order accuracy. The PDF set `NNPDF3.0nlo` is used for the generation of distinct polarization templates. One unpolarized sample, the LL, TL, and TT polarization states in the $W^{\pm}W^{\pm}$ center-of-mass system and the LL, TL, and TT polarization states in the partonic center-of-mass system are produced. The polarization samples are validated to follow the expected $W^{\pm}$ boson decay distribution in equation (3.93) and the sum of the polarizations is consistent with the unpolarized $W^{\pm}W^{\pm}jj$-EW cross-section. Since the simulation has leading-order accuracy, the samples will be reweighted for higher-order effects in chapter 8.

7.1.2 $W^{\pm}W^{\pm}jj$-EW Polarization by Sherpa

With the Sherpa 3.0 version, the simulation of cross-sections with polarized intermediate particles becomes available [45]. The `PDF4LHC21_40_pdfas` PDF set is used to generate polarized Sherpa samples (DSID 700965-700969, and 701228-701231). In contrast to the polarization simulation by MadGraph, the Sherpa simulations are not distinct polarization samples. The polarization information is provided eventwise by additional weights, allowing the simulation of all polarization states studied in the analysis to be performed in just one run. The advantage of the Sherpa polarization simulation is the availability of jet-merging described in section 4.3 to introduce higher-order QCD effects. Thus, a $W^{\pm}W^{\pm}jj + 0, 1j$-EW sample is generated that merges the leading-order $W^{\pm}W^{\pm}jj$-EW simulation with the $W^{\pm}W^{\pm}jjj$-EW contribution which has one additional jet. To enable the simulation of polarization the narrow-width approximation is applied. This narrow-width approximation requires dropping the s-channel-like diagrams to avoid divergences at higher-order QCD. In section 8.1, this approximation is described in more detail and the corresponding correction for the missing s-channels is derived. A NLO-EW correction is implemented in section 8.2. The validation of the polarization modeled by Sherpa is done in section 8.1.1 against the simulation by the MadGraph event generator.

7.1.3 Polarization Interference

The impact of the interference $\sum_{\lambda\neq\lambda'}\mathcal{M}_{\lambda}^{\mathcal{F}*}\mathcal{M}_{\lambda'}^{\mathcal{F}}$ between the different polarization states is expected to be small. The estimation of this background depends on the event generator used for the polarization simulation. For MadGraph, the interference is calculated by the difference

$$N_{\text{int}} = N_{\text{unpol}} - \sum_{\text{pol}} N_{\text{pol}} \tag{7.1}$$

between the unpolarized sample (DSID 506202) and the sum of the polarized contributions.

For the polarization simulation by the Sherpa event generator, the interference contribution is directly provided by an event weight as for the individual polarization states. Thus, Sherpa directly models the polarization interference.

7.1.4 $W^{\pm}W^{\pm}jj$-INT and $W^{\pm}W^{\pm}jj$-QCD Background

The $W^{\pm}W^{\pm}jj$-INT ($\mathcal{O}(\alpha_{EW}^5)$) and $W^{\pm}W^{\pm}jj$-QCD ($\mathcal{O}(\alpha_{EW}^4)$) contribution is modelled by the MadGraph event generator. These simulated samples were already used and validated in the unpolarized $W^{\pm}W^{\pm}jj$-EW measurement in Ref. [8]. The theoretical background of these distributions is described in detail in section 3.3.3, and the sample details are listed in table A.2.

7.2 Background Estimation

Several processes contribute to the background in the signal region defined in section 6.3.1. Their shape and number of events are modeled using event generator simulations and kinematic information extracted from measured data. The different background contributions are described in the following sections and sorted in decreasing order based on the number of events resulting from the background. Since these backgrounds were already studied in the unpolarized $W^{\pm}W^{\pm}jj$-EW analysis [8] only a brief overview is given in the following.

7.2.1 $W^{\pm}Z$ Background

The events originating from $W^{\pm}Zjj \to l^{\pm}\nu l^{\pm}l^{\mp}jj$ are the most dominant background in the signal region. This background is not entirely avoided by the two-lepton requirement of the signal region since one of the leptons originating from the Z boson can be outside the detector acceptance. If one of the leptons is not reconstructed or fails the baseline lepton selection, the $W^{\pm}Zjj \to l^{\pm}\nu l^{\pm}l^{\mp}jj$ process can fulfill the signal requirements of two same-charged leptons, two jets, and a missing transverse momentum.

The $W^{\pm}Z$ background comprises the electroweak scattering $W^{\pm}Z$-EW and the strong scattering $W^{\pm}Z$-QCD. In table A.3, the event generator samples are listed that are used to model this background. While the $W^{\pm}Z$-EW MadGraph sample offers a good prediction for the $W^{\pm}Z$-EW contribution, the $W^{\pm}Z$-QCD is known from Ref. [8] to be poorly modeled by Monte Carlo simulations. By moving from the $W^{\pm}Z$-QCD Sherpa2.2.2 sample used in the unpolarized $W^{\pm}W^{\pm}jj$-EW analysis [8] to the newer $W^{\pm}Z$-QCD Sherpa2.2.12 sample, the mismodeling has slightly improved. However, the prediction by the $W^{\pm}Z$-QCD Sherpa2.2.12 sample still needs to be corrected using the same data-driven approach as in Ref. [8]. The implementation and shape correction of the Sherpa2.2.12 sample are discussed in Ref. [84].

To improve the prediction of the $W^{\pm}Z$-QCD background, the $W^{\pm}Z$ control region introduced in section 6.3.3 is used. Since the m_{jj}-distribution is known to be most affected by the mismodeling but is crucial for the separation of $W^{\pm}W^{\pm}jj$-EW and backgrounds, the event generator prediction is corrected by measured data in this kinematic variable. The non-$W^{\pm}Z$-QCD processes are subtracted from the measured data in the m_{jj}-histogram for the shape correction. The remaining data distribution is normalized to unity and divided by the normalized $W^{\pm}Z$-QCD prediction. The resulting ratio represents the m_{jj} shape difference between prediction and data. The shape of the correction is smoothed by an exponential function $f(x) = ae^{-bx}$. This function is fit to the ratio of (data − non-$W^{\pm}Z$-QCD)/$W^{\pm}Z$-QCD in the range of $200\,\text{GeV} < m_{jj} < 2500\,\text{GeV}$. This analytic fit is applied on the $W^{\pm}Z$-QCD background in the signal and Low-m_{jj} control region to correct the shape of the prediction. For this shape correction three different uncertainties are considered[1]:

- PDF and α_S choice:
  The reweighting function is derived for different PDF and α_S variations. The envelope of the resulting fits is chosen as uncertainty.
- Higher order effects:
  The impact of higher-order effects is estimated by the envelope of fit functions for different scale variations.
- Statistic:
  The statistical uncertainty of the measured data in the $W^{\pm}Z$ control region and the limited number of Monte Carlo events generated for the prediction.

In addition to the shape, the normalization of the $W^{\pm}Z$-QCD background has to be improved. The normalization correction is done by including the $W^{\pm}Z$ control region in the final fitting procedure in section 12.1 with a free fit parameter $\mu_{\text{WZ-QCD}}$ on the normalization of the $W^{\pm}Z$-QCD background. For this fit, only the total number of events in the $W^{\pm}Z$ control region is considered since the exponential reweighting function already extracts the shape information.

7.2.2 Non-Prompt Background

Another essential but poorly modeled background originates from so-called non-prompt leptons. Non-prompt leptons do not originate directly from the primary collision but are the product of hadron decays or jets misidentified as leptons. Since the modeling by Monte Carlo simulations is very poor, the data-driven fake factor method is used to extract this prediction. The fake factor method for the $W^{\pm}W^{\pm}jj$-EW analysis is developed in Ref. [85] and summarized in the following.

A new lepton selection Anti-ID is introduced for the fake factor method, which has looser object quality requirements than the signal leptons described in section 6.2. Anti-ID selected electrons must pass only the `MediumLH` identification and have no isolation requirement. The Anti-ID muon selection requires the `FixedCutPflowLoose` isolation and a transverse impact parameter of $|\frac{d_0}{\sigma_{d_0}}| < 10$. To avoid an overlap between the Anti-ID and signal object selection the Anti-ID leptons must also fail at least one of the tighter signal object selection criteria. Leptons passing this Anti-ID selection criteria are significantly more likely to be non-prompt than leptons passing the signal object selection.

By this additional object selection, the fake factor

$$F = \frac{f}{\bar{f}} \tag{7.2}$$

can be defined with the probability f of a non-prompt lepton passing the signal selection and the probability $\bar{f}$ of a non-prompt lepton being Anti-ID selected. The fake factor F is measured in a single lepton control region enriched with non-prompt leptons. Exactly one signal or Anti-ID selected lepton is required in this region with a transverse momentum of

---

[1] More details about PDF, α_S, and scale variations are given in section 11.2

$p_T > 27\,\text{GeV}$. The events also have to have at least one signal jet with $p_T > 25(30)\,\text{GeV}$ for $|\eta < 2.5|(2.5 < \eta < 4.5)$. The tagging jet must be back-to-back with the lepton with $|\Delta\phi_{lj}| > 2.8$. To suppress non-prompt $W^{\pm}$+jets events, the sum of the missing transverse momentum p_T^{miss} and the transverse mass of the lepton and p_T^{miss} system is required to be smaller than 50 GeV. The fake factor is averaged over non-prompt control regions with and without a veto on b-tagged jets. However, the systematic uncertainties of the fake factor method are derived from the non-prompt control region with a b-jet veto to be more consistent with the signal region selection.

In this control region, the number of events with signal-selected leptons N_{signal} and Anti-ID selected leptons $N_{\text{Anti-ID}}$ are extracted from data. By subtracting the number of prompt events $N_{\text{signal}}^{\text{prompt}}$ and $N_{\text{Anti-ID}}^{\text{prompt}}$ predicted by event generators, the actual number of non-prompt events in this control region can be calculated. This enables a direct extraction of the fake factor

$$F(p_T) = \frac{N_{\text{signal}} - N_{\text{signal}}^{\text{prompt}}}{N_{\text{Anti-ID}} - N_{\text{Anti-ID}}^{\text{prompt}}} \tag{7.3}$$

dependent on the lepton p_T. Since the non-prompt rates of electrons and muons differ significantly, two distinct fake factor F_e and F_μ are extracted from the non-prompt control region.

Another region is defined as the transfer of non-prompt rates to the signal region. This region has the same selection criteria as the signal region, but one of the two leptons must be Anti-ID selected. As in the non-prompt control region, the number of prompt events is subtracted using samples produced by event generators. The resulting difference is the expected non-prompt contribution with one signal and one Anti-ID lepton. By applying the fake factor, this contribution can be projected to the non-prompt background

$$\begin{aligned} N_{\text{signal; signal}}^{\text{non-prompt}} = {} & F_e(p_T)\left(N_{\text{signal}\,e;\,\text{Anti-ID}\,e} - N_{\text{signal}\,e;\,\text{Anti-ID}\,e}^{\text{prompt}}\right) \\ & + F_e(p_T)\left(N_{\text{signal}\,\mu;\,\text{Anti-ID}\,e} - N_{\text{signal}\,\mu;\,\text{Anti-ID}\,e}^{\text{prompt}}\right) \\ & + F_\mu(p_T)\left(N_{\text{signal}\,e;\,\text{Anti-ID}\,\mu} - N_{\text{signal}\,e;\,\text{Anti-ID}\,\mu}^{\text{prompt}}\right) \\ & + F_\mu(p_T)\left(N_{\text{signal}\,\mu;\,\text{Anti-ID}\,\mu} - N_{\text{signal}\,\mu;\,\text{Anti-ID}\,\mu}^{\text{prompt}}\right) \end{aligned} \tag{7.4}$$

in the signal region with two signal-selected leptons. For the fake factor application, the p_T of the Anti-ID selected lepton is used.

For the calculation of the fake factor, a variety of experimental uncertainties is considered:

- Prompt subtraction:
  The prompt subtraction is varied by 5% to estimate a modeling uncertainty on the fake factor.

- Jet flavor composition:
  The fake factor is extracted separately from the non-prompt control region once with $N_{b\text{-jet}} = 0$ and $N_{b\text{-jet}} > 0$. The mean fake factor is chosen as the nominal strategy, while the two individual fake factors represent the uncertainty for the jet flavor composition in the non-prompt control region.

- p_T^{miss} reconstruction:
  To account for uncertainties in the missing momentum reconstruction the $p_T^{miss} + m_T$ selection of the control region is varied by 5 GeV.

- Trigger bias:
  The fake factor is re-estimated for trigger selections with different identification and isolation working points.

- Statistical uncertainty of the fake factor: The statistical uncertainty of the non-prompt control region is propagated to the fake factor calculation.

7.2.3 Charge-Flip Background

A major contribution of events with two leptons measured in the ATLAS detector originates from the $Z \to e^+e^-$ decay. Most of these events are rejected by the same-charge requirement of the signal region, but some $Z \to e^+e^-$ events pass this selection due to charge-flip. Charge-flip is a process causing electrons to be reconstructed with an opposite charge. For highly energetic electrons, charge misidentification can occur due to an indeterminable curvature of the track. However, in most cases, charge-flip results from bremsstrahlung emitted by the electron. As shown in figure 7.1 the electron (positron) emits a hard photon that carries most of the momentum of the initial electron (positron). This photon produces a high energetic positron (electron) and a low energetic electron (positron) via pair production. The two soft electrons (positrons) get lost in the detector, and only the opposite-charged positron (electron) is measured in the detector.

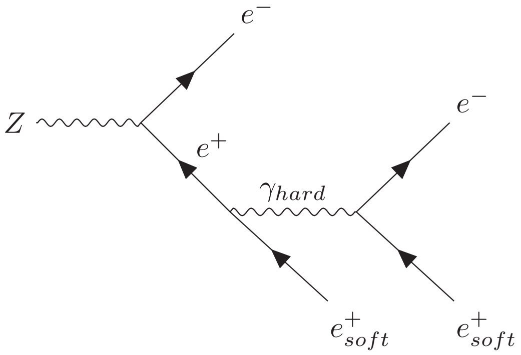


Figure 7.1: Feynman diagram of an electron charge-flip

The data-driven estimation of the charge-flip background used for this analysis was developed in Ref. [86]. A modified signal region with oppositely charged leptons is used to model this background. The data in this opposite-charged signal region is multiplied by a factor

$$w = \frac{\epsilon_{l1}(1-\epsilon_{l2}) + \epsilon_{l2}(1-\epsilon_{l1})}{(1-\epsilon_{l1})(1-\epsilon_{l2}) + \epsilon_{l1}\epsilon_{l2}} \tag{7.5}$$

dependent on the charge-flip probabilities ϵ of the individual leptons. The probabilities are derived by a MadGraph simulation of $Z \to ee$ in a dedicated charge-flip control region and corrected by additional scale factors derived by the *Egamma Combined Performance (CP) group* in Ref. [87]. In this $Z \to ee$ charge-flip control region, an invariant mass of $m_{ll} > 20\,\text{GeV}$ is required, and events have to pass the `ECIDS` charge-flip rejection. The electrons have to have `TightLH` isolation, `Gradient` isolation, a transverse momentum of $p_T > 20\,\text{GeV}$ and geometrical acceptance of $|\eta| < 2.47$ excluding $1.37 \leq |\eta| \leq 1.52$. In this region, the charge-flip probabilities $\epsilon(p_T, \eta)$ are calculated as a function of the electron p_T and η. For muons, the charge-flip probability is set to exactly 0.

By multiplying the data in the opposite-charged signal region by the charge-flip factor w, a very good estimation of charge-flip events in the signal region is provided. However, an additional energy correction is required since the charge-flip process leads to energy losses. Thus, the transverse momentum of the electrons is corrected by

$$p_T^{\text{corrected}} = \frac{p_T}{\alpha} + dE. \tag{7.6}$$

The momentum is scaled by the correction factor α and a Gaussian smearing is applied by the random value dE. The value of α and the Gaussian distribution for dE are calculated using the MadGraph simulation for $Z \to ee$ in different bins of $|\eta|$.

The uncertainty of the charge-flip background originate from the statistical uncertainty in the opposite-charged signal region and the systematic uncertainty of the scale factors provided by the *Egamma Combined Performance (CP) group*. The uncertainties of the scale factors are explained in Ref. [87] and [78].

7.2.4 $V\gamma$, ZZ, Z+jets, Top, and Tri-boson Background

The remaining background consists of contributions of $V\gamma$, ZZ, Z+jets, top, and tri-boson events. These backgrounds are known from the unpolarized $W^{\pm}W^{\pm}jj$-EW analysis [8] to have a minor impact compared to the backgrounds described in the previous sections. The event generator samples used to model these backgrounds are listed in tables A.4 to A.7. To avoid an overlap with the data-driven estimates of non-prompt and charge-flip contributions, only prompt and non-charge-flip events are considered in the signal region.

The uncertainties of these backgrounds are estimated by conservative normalization uncertainties implemented in Ref. [8].

Chapter 8

Higher-Order Corrections on $W^{\pm}W^{\pm}jj$-EW

The polarized templates are simulated by MadGraph with leading-order accuracy with subsequent parton shower (LO+PS). The Sherpa event generator also includes higher-order QCD effects via multi-jet merging. However, the multi-jet merging required additional approximation to model the polarization states. Thus, both samples require additional corrections to model the effects at higher-order QCD ($\alpha_S^2\alpha_{EW}^4 \rightarrow \alpha_S^3\alpha_{EW}^4$) correctly. These effects are studied on particle-level by comparing different Sherpa samples and deriving a multivariate higher-order correction.

For higher-order electroweak effects ($\alpha_S^2\alpha_{EW}^4 \rightarrow \alpha_S^2\alpha_{EW}^5$), the authors of Ref. [50] provided a calculation on parton-level. These calculation results can directly introduce the next-to-leading-order electroweak (NLO-EW) correction into the analysis.

8.1 Approximate NLO-QCD Correction

The recent Sherpa version 3.0 enables the simulation of the cross-section of polarized intermediate vector bosons [45]. As described in section 4.3, the Sherpa event generator provides multi-jet merging including the radiation of additional jets in the matrix element calculation. The multi-jet merging is used for a study of the higher-order QCD effects of real emissions on the polarized $W^{\pm}W^{\pm}jj$-EW scattering. In Ref. [45], the multi-jet merging with one additional QCD radiation in the matrix element is shown to cover the dominant NLO-QCD effects and is therefore referred to as approximate NLO-QCD+PS. The approximate NLO-QCD+PS study is performed on particle-level samples without detector simulation to reduce the computational effort. The correction procedure is explained in section 8.1.2 and requires multiple samples listed in table 8.1. The resulting higher-order QCD corrections are applied to the nominal reconstruction-level polarization samples generated with detector simulation. For the application of the correction, the event-wise particle-level information is used, which is additionally stored in the reconstruction-level polarization samples. Therefore, the event kinematic of the correction application is consistent with the kinematic used for the extraction.

8.1.1 Comparison of Simulation by Sherpa and MadGraph

The polarization simulation by the Sherpa event generator became available after the samples generated by MadGraph were validated and established in the $W^{\pm}W^{\pm}jj$-EW polarization analysis. Therefore, the Sherpa prediction is introduced into the analysis by a validation against the MadGraph prediction. This comparison has to be done at LO-QCD+PS since MadGraph cannot perform multi-jet merging for vector boson scattering. As in the rest of this chapter, this comparison is performed for particle-level samples to avoid the computational effort of the detector simulation.

Table 8.1: Sherpa $W^{\pm}W^{\pm}jj$-EW samples to study the approximate NLO-QCD+PS effects.

DSID	sample name
	Sherpa $W^{\pm}W^{\pm}jj$
700961	Sh_300b1_WlvWlvjj_mm_LO_pol
700962	Sh_300b1_WlvWlvjj_pp_LO_pol
	Sherpa $W^{\pm}W^{\pm}jj$ (t-ch.)
700963	Sh_300b1_WlvWlvjj_mm_LO_MinTChannel_pol
700964	Sh_300b1_WlvWlvjj_pp_LO_MinTChannel_pol
	Sherpa $W^{\pm}W^{\pm}jj + 0, 1j$ (t-ch.)
700965	Sh_300b1_WlvWlvjj_mm_MinTChannel_pol_highMjj
700966	Sh_300b1_WlvWlvjj_pp_MinTChannel_pol_highMjj
700968	Sh_300b1_WlvWlvjj_mm_MinTChannel_pol_lowMjj
700969	Sh_300b1_WlvWlvjj_pp_MinTChannel_pol_lowMjj
	Sherpa $l^{\pm}\nu l^{\pm}\nu jj + 0, 1j$ (t-ch.)
700971	Sh_300b1_llvvjj_ss_MinTChannel
	Sherpa $l^{\pm}\nu l^{\pm}\nu jj + 0, 1j$
700970	Sh_300b1_llvvjj_ss

The comparison is made in the particle-level phase space defined in table 8.2. This phase space is equivalent to the signal region definition merged with the Low-m_{jj} control region and covers the kinematic region relevant to the polarization measurement. The number of events predicted by MadGraph and Sherpa at LO-QCD+PS are listed in table 8.3. The simulation with MadGraph predicts slightly lower cross-sections than Sherpa. The predicted TL fraction by Sherpa is a bit higher and the TT fraction is a bit lower than in the prediction by MadGraph. The fractions are calculated relative to the sum of the longitudinal and transverse polarization states.

Table 8.2: Particle-level selection corresponding to the SR and Low-m_{jj} region selection.

Exactly two signal leptons with $p_T > 27$ GeV and the same electrical charge with $\lvert\eta\rvert < 2.5$ for muons and with $\lvert\eta\rvert < 2.47$ excluding $1.37 \leq \lvert\eta\rvert \leq 1.52$ for electrons with $\lvert\eta\rvert < 1.37$ in the ee channel
$m_{ll} > 20$ GeV
$\lvert m_{ee} - m_Z\rvert > 15$ GeV in the ee-channel
$E_{\mathrm{T}}^{\mathrm{miss}} \geq 30$ GeV
$m_{jj} \geq 200$ GeV
at least two jets leading and subleading jets satisfying $p_T > 65$ GeV and $p_T > 35$ GeV
$\Delta y_{jj} > 2$

The most prominent shape differences between the predictions emerge in the subleading jet p_T, the di-jet invariant mass m_{jj}, and the jet separation ΔR_{jj}. Figure 8.1 shows the comparison in these variables for the polarization in the $W^{\pm}W^{\pm}$ and partonic center-of-mass frame. The overall shapes agree but differ in the kinematic tails. This difference in the jet kinematic can be explained by the different showering algorithms used in the simulations. However, the general agreement between the Sherpa and the MadGraph generated sample

Table 8.3: Comparison of the cross-sections, polarization fractions, and number of generated Monte Carlo events for different MadGraph and Sherpa samples. The shown events are selected by the particle-level phase space defined in table 8.2.

			WW-cmf			pp-cmf		
	MC events	total σ	LL	LT+TL	TT	LL	LT+TL	TT
MG unpol	73261	2.880 fb	-	-	-	-	-	-
MG pp-cmf	149321	2.925 fb	-	-	-	0.177 fb 6.07%	1.030 fb 35.20%	1.718 fb 58.73%
MG WW-cmf	1426669	2.896 fb	0.277 fb 9.58%	0.874 fb 30.18%	1.744 fb 60.24%	-	-	-
Sh $W^\pm W^\pm jj$	1269306	2.933 fb	0.285 fb 9.60%	0.909 fb 30.59%	1.778 fb 59.81%	0.182 fb 6.09%	1.072 fb 35.91%	1.732 fb 58.00%

is sufficient considering the expected number of events in the signal region. Thus, for the application in this thesis, the polarization simulation in Sherpa is successfully validated against the MadGraph simulation.

8.1.2 Correction Procedure

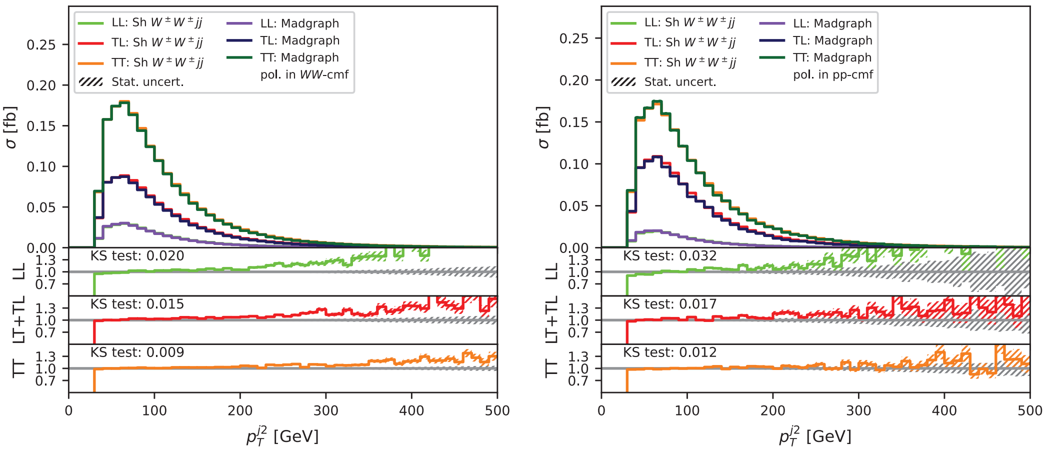
The $W^\pm W^\pm jj$-EW scattering is already described in detail in section 3.3.3. However, it is worth revisiting the most essential points to clarify the principle of the higher-order QCD corrections better. For the vector boson polarization simulation, the Sherpa event generator uses a narrow-width approximation (NWA) to suppress the contribution of the single-resonant diagrams (figures 3.4h and 3.4i) and non-resonant diagrams (figure 3.4j). This approximation lead to on-shell $W^\pm$ bosons and thus divergences for the hadronically decaying $W^\pm$ boson in s-channel-like diagrams (figures 3.4d to 3.4g) for $m_{jj} \to m_{W^\pm}$. This divergence is avoided at LO by the $m_{jj} > 200\,\text{GeV}$ selection in the signal region and Low-m_{jj} control region. However, when including additional QCD emissions in the matrix element calculation, divergences reappear if the two leading jets do not originate from the same $W^\pm$ boson as shown in figure 8.2. In order to avoid this divergence, the s-channel-like Feynman diagrams are dropped for the polarization simulation at approximate NLO-QCD+PS. This results in an additional correction factor at approximate NLO-QCD+PS to correct for the missing s-channel diagrams. In this paper, dropping the s-channels is referred to as t-channel approximation, but this also includes the remaining u-channel-like Feynman diagrams.

To ensure a valid correction for approximate NLO-QCD+PS effects several Sherpa simulations of $W^\pm W^\pm jj$-EW listed in table 8.1 are compared on particle-level in the following section:

- $W^\pm W^\pm jj$ vs. $W^\pm W^\pm jj$ (t-ch.):
  Effects of the t-channel approximation on the polarization states at LO-QCD+PS
- $W^\pm W^\pm jj$ (t-ch.) vs. $W^\pm W^\pm jj + 0, 1j$ (t-ch.):
  Impact of including one additional QCD emission in the matrix element calculation while the t-channel approximation is applied
- $W^\pm W^\pm jj + 0, 1j$ (t-ch.) vs. $l^\pm \nu l^\pm \nu jj + 0, 1j$ (t-ch.):
  Impact of the NWA at approximate NLO-QCD+PS. This comparison can only be done for the unpolarized contribution
- $l^\pm \nu l^\pm \nu jj + 0, 1j$ (t-ch.) vs. $l^\pm \nu l^\pm \nu jj + 0, 1j$:
  Unpolarized contribution of s-channel-like diagrams for the simulation with multi-jet merging

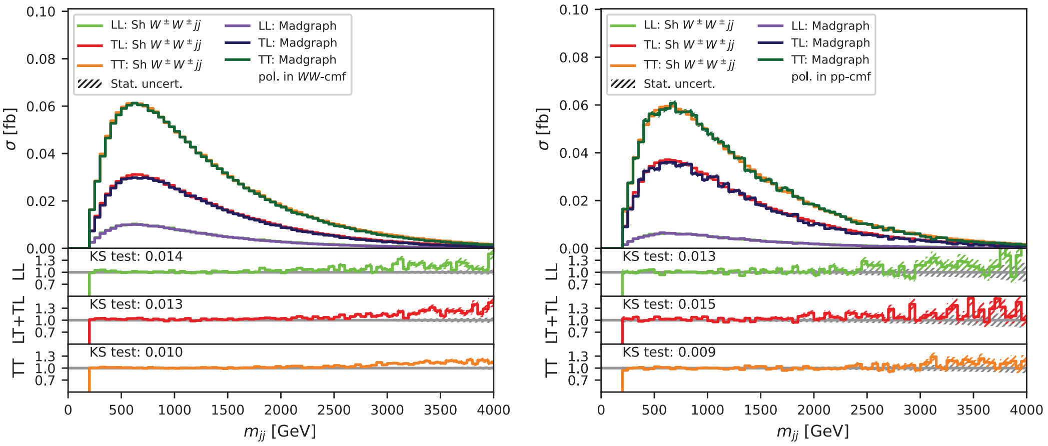
The resulting correction factor for a polarized sample at LO-QCD+PS

$$\text{corr}_{\text{jet-merged}} = \left(\frac{W^\pm W^\pm jj + 0, 1j\ \text{(t-ch.)}}{W^\pm W^\pm jj} \right)_{pol} \cdot \left(\frac{l^\pm l^\pm \nu\nu jj + 0, 1j}{W^\pm W^\pm jj + 0, 1j\ \text{(t-ch.)}} \right)_{unpol} \tag{8.1}$$



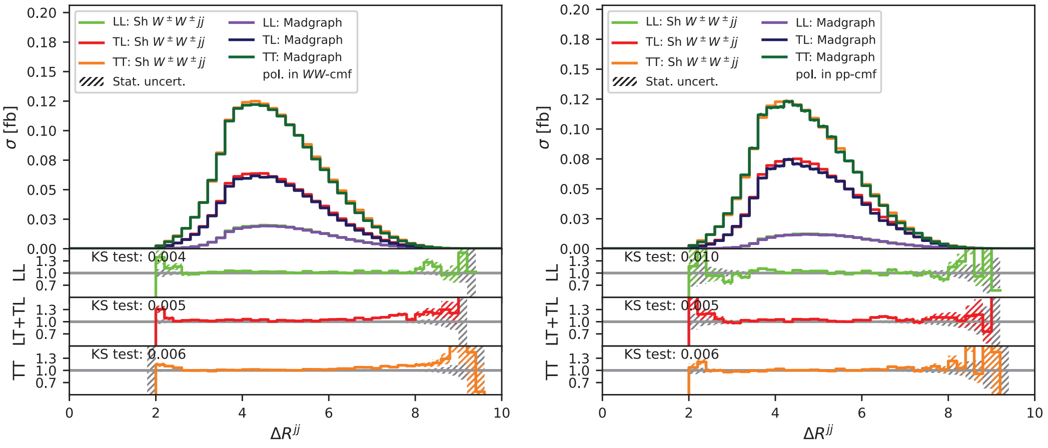
(a) WW-cmf: sublead. jet p_T

(b) pp-cmf: sublead. jet p_T



(c) WW-cmf: jet m_{jj}

(d) pp-cmf: jet m_{jj}



(e) WW-cmf: jet ΔR_{jj}

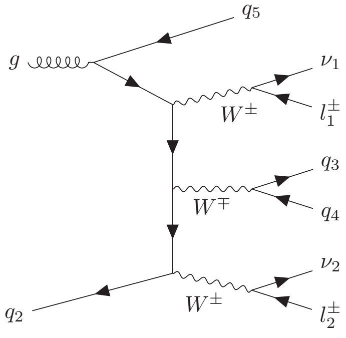
(f) pp-cmf: jet ΔR_{jj}

Figure 8.1: Comparison of the differential polarized cross-sections simulated by SHERPA $W^{\pm}W^{\pm}jj$ and MADGRAPH at particle-level in the phase space defined in table 8.2. The polarizations are defined in the WW-cmf for figures 8.1a, 8.1c and 8.1e and in the pp-cmf for figures 8.1b, 8.1d and 8.1f.

corrects for the additional QCD radiation and the approximations necessary for the polarized simulation. The product in equation (8.1) already includes the correction

$$\mathrm{corr}_{\text{s-ch.}} = \left(\frac{l^{\pm}l^{\pm}\nu\nu jj + 0,1j}{W^{\pm}W^{\pm}jj + 0,1j\ \text{(t-ch.)}} \right)_{unpol} \tag{8.2}$$

for the NWA and the missing s-channel diagrams at approximate NLO-QCD+PS. The validity of these two corrections will be studied in the subsequent sections.



(a)

Figure 8.2: A s-channel-like Feynman diagram for $W^{\pm}W^{\pm}jj$-EW at NLO-QCD.

8.1.3 Comparison of the Sherpa Simulation Setups

Since the measurement by the detector can cause kinematic migrations, the set of events passing the particle-level selection is not entirely equivalent to the set of events passing the reconstruction-level selection. These migrations can be considered by a multivariate correction which is extracted from a significantly looser particle-level phase space. The selection on particle-level in table 8.4 covers almost all events expected to potentially migrate into the reconstruction-level signal region and Low-m_{jj} region. The η selection of leptons is kept tight since it is applied in the object selection. A looser η selection on leptons could lead to a different set of leptons used for the subsequent event selection, and therefore, the lepton η is kept tight. The p_T selection of the leptons is also applied as object selection, but as the leptons are sorted by p_T, a higher p_T requirement can be introduced by a subsequent event selection. Therefore, a loose p_T selection can be used as a baseline and later tightened by an additional event selection. The resulting cross-sections for the different Sherpa samples are listed in table 8.5. While the cross-section differs between the different Sherpa simulations, the polarization fractions are not affected.

Table 8.4: Loose particle selection used to derive the higher-order QCD correction.

Exactly two signal leptons with $p_T > 20$ GeV and the same electrical charge with $\|\eta\| < 2.5$ for muons and with $\|\eta\| < 2.47$ excluding $1.37 \leq \|\eta\| \leq 1.52$ for electrons
$m_{ll} > 15$ GeV
at least two jets leading and subleading jets satisfying $p_T > 40$ GeV and $p_T > 20$ GeV

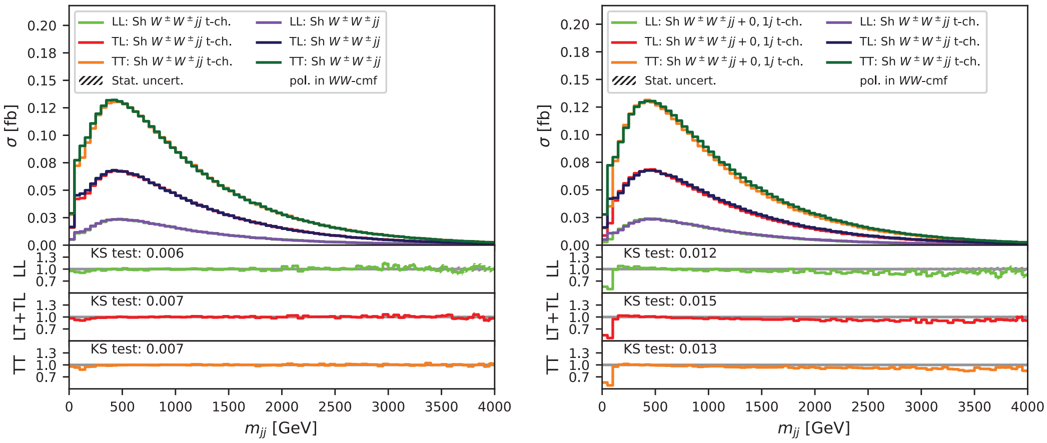
Since the polarization analysis follows a multivariate approach to use the full phase space information for signal discrimination, the Sherpa sample comparison has to be done in a variety of kinematic variables. The lepton and jet rapidity η, transverse momentum p_T and azimuthal angle relative to the leading lepton ($\phi - \phi^{l1}$) and the missing transverse momentum p_T^{miss} are studied to check the kinematic of individual objects. In addition, also high-level variables like the transverse masses ($m_T^{l1,MET}$, $m_T^{l2,MET}$, and $m_T^{ll,MET}$), angle separations

Table 8.5: Comparison of the cross-sections, polarization fractions, and number of generated Monte Carlo events for different Sherpa samples. The shown events are selected by the loose particle-level event selection given in table 8.4.

			WW-cmf			pp-cmf		
	MC events	total σ	LL	LT+TL	TT	LL	LT+TL	TT
Sh $W^{\pm}W^{\pm}jj$	2708209	6.261 fb	0.650 fb	1.950 fb	3.669 fb	0.416 fb	2.302 fb	3.556 fb
			10.36%	31.10%	58.53%	6.64%	36.69%	56.67%
Sh $W^{\pm}W^{\pm}jj$ t-ch.	2714037	6.207 fb	0.644 fb	1.935 fb	3.638 fb	0.413 fb	2.286 fb	3.524 fb
			10.36%	31.13%	58.52%	6.63%	36.74%	56.63%
Sh $W^{\pm}W^{\pm}jj + 0,1j$ t-ch.	2682967	5.935 fb	0.625 fb	1.849 fb	3.465 fb	0.402 fb	2.197 fb	3.345 fb
			10.53%	31.13%	58.34%	6.76%	36.96%	56.28%
Sh $l^{\pm}\nu l^{\pm}\nu jj + 0,1j$ t-ch.	2706926	6.051 fb	-	-	-	-	-	-
Sh $l^{\pm}\nu l^{\pm}\nu jj + 0,1j$	1278108	7.706 fb	-	-	-	-	-	-

(ΔR, Δy, $\Delta\phi$), the invariant di-lepton and di-jet mass, and transverse momentum of the di-lepton system are investigated. Therefore, the following comparison covers almost all kinematic variables considered for the neural network training in section 9.1. A more detailed explanation of the kinematic variables is provided in section 9.1. The following comparison of the differential cross-sections focuses on the polarization defined in the WW-cmf as the polarization in the pp-cmf leads to the same conclusions.

As shown in figure 8.3a, dropping the s-channel diagrams at LO slightly reduces the number of events with $m_{jj} < 200\,\text{GeV}$. In s-channel-like diagrams, the two jets originate from the decay of a $W^{\pm}$ boson. Their invariant mass m_{jj} and their separation into Δy_{jj} and ΔR_{jj} is therefore significantly lower than for the two tagging jets in t-channel-like diagrams. The differential cross-sections in m_{jj}, Δy_{jj}, and ΔR_{jj} for the polarizations in both reference frames are given in figure B.1 in the appendix. For other kinematic variables, almost no impact is seen from dropping the s-channels at LO.

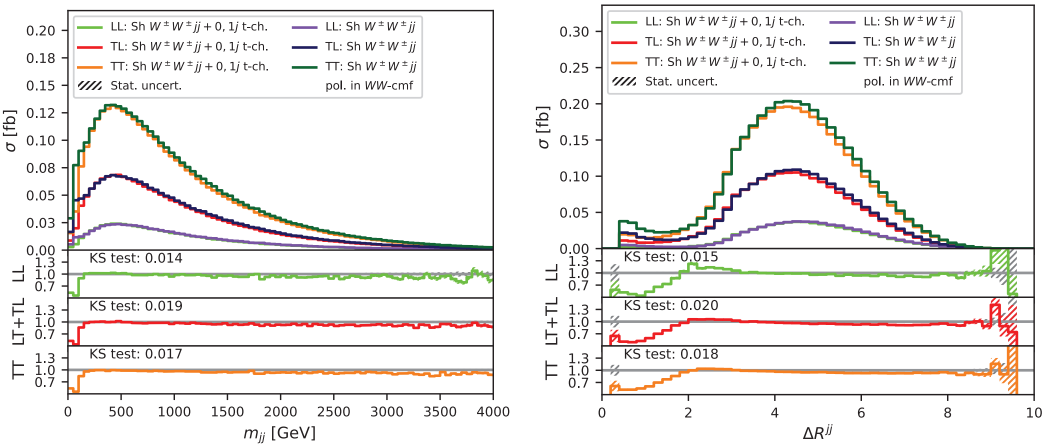


(a) $W^{\pm}W^{\pm}jj$ vs. $W^{\pm}W^{\pm}jj$ (t-ch.) **(b)** $W^{\pm}W^{\pm}jj$ (t-ch.) vs. $W^{\pm}W^{\pm}jj + 0,1j$ (t-ch.)

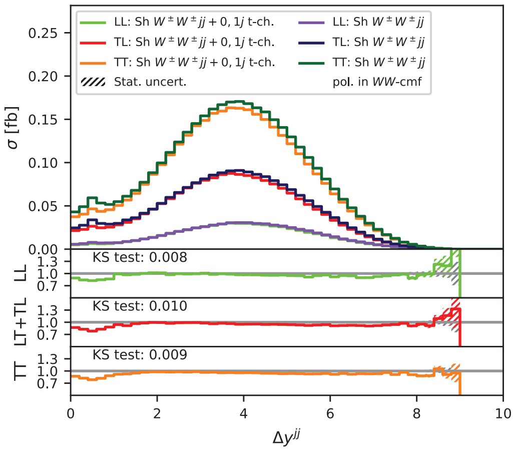
Figure 8.3: Comparison of the differential cross-sections simulated by Sherpa $W^{\pm}W^{\pm}jj$ vs. $W^{\pm}W^{\pm}jj$ (t-ch.) in figure 8.3a and $W^{\pm}W^{\pm}jj$ (t-ch.) vs. $W^{\pm}W^{\pm}jj + 0,1j$ (t-ch.) in figure 8.3b on particle-level in the phase space defined in table 8.4

A much more dominant effect results from the consideration of an additional jet. The comparison of LO-QCD+PS and approximate NLO-QCD+PS, both with t-channel approximation, is shown in figure 8.3b. The cross-section of events with $m_{jj} < 100\,\text{GeV}$ is drastically suppressed by the additional jet emission. This cross-section drop originates from the $m_{jj} > 100\,\text{GeV}$ selection at Sherpa generator-level. This selection criterion is applied after the matrix element calculation. Due to smearing effects from the parton shower, events with m_{jj} below 100 GeV are possible in the generated sample. By multi-jet merging, an additional jet is included in the matrix element calculation. In this case, the generator-level m_{jj} selection is applied on the two jets with the highest p_T and thus considers also the third jet. The additional jet can easily have a higher p_T value but is expected to have an overall smaller Δy_{jj} separation and therefore a smaller m_{jj} than expected from the two jets produced at LO-QCD+PS. Thus, the jet-merged sample is expected to be more affected by the generator-level m_{jj} selection.

Since the approximate NLO-QCD+PS effects have a significantly higher impact on the process kinematic than the t-channel approximation at LO-QCD+PS, combining both effects into one common correction. The shape differences between $W^{\pm}W^{\pm}jj$ and $W^{\pm}W^{\pm}jj + 0,1j$ (t-ch.) are summarized by the Kolmogorov Smirnov values in figure 8.5a. As expected by the previous comparison steps, the main difference arises in the di-jet kinematics of m_{jj}, $\Delta\phi_{jj}$ and ΔR_{jj} shown in figure 8.4. The differences between $W^{\pm}W^{\pm}jj$ and $W^{\pm}W^{\pm}jj{+}0,1j$ (t-ch.) will be corrected by the first factor of the $\text{corr}_{\text{jet-merged}}$ correction factor in equation (8.1).



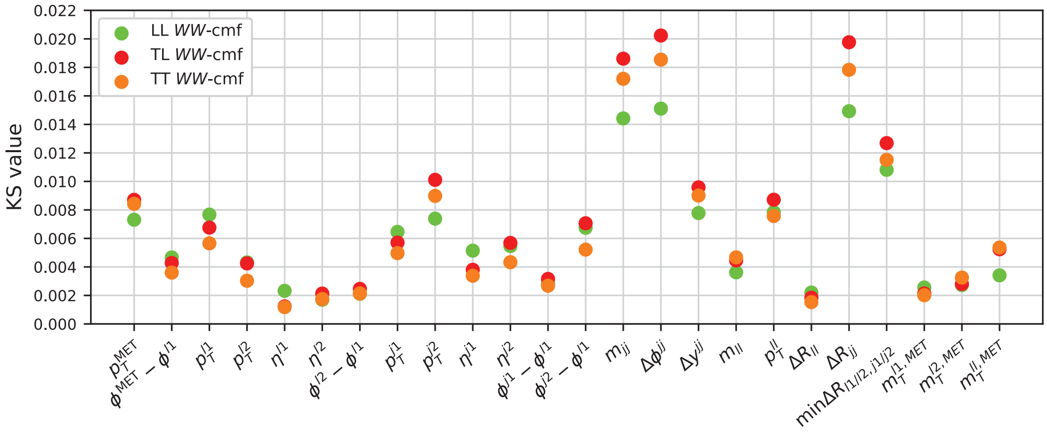
(a) m_{jj} **(b)** ΔR_{jj}



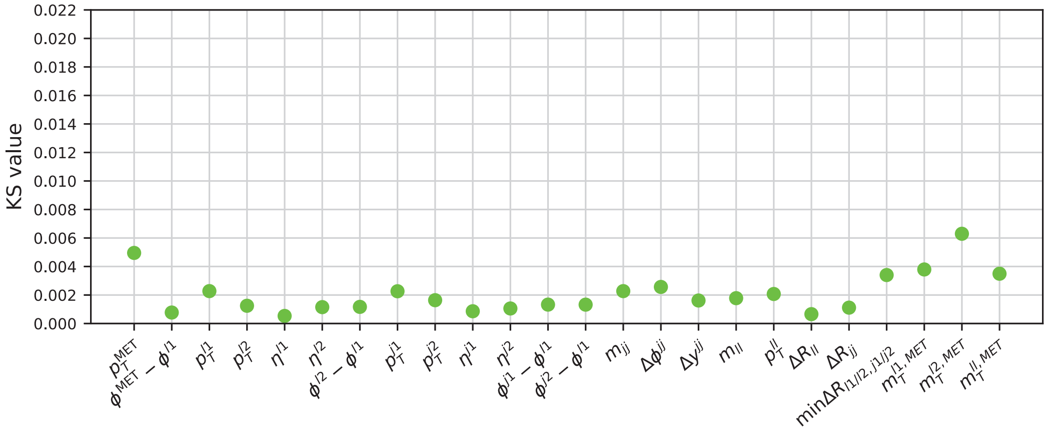
(c) Δy_{jj}

Figure 8.4: Comparison of the differential cross-sections simulated by Sherpa $W^{\pm}W^{\pm}jj$ and $W^{\pm}W^{\pm}jj + 0,1j$ (t-ch.) on particle-level in the phase space defined in table 8.4

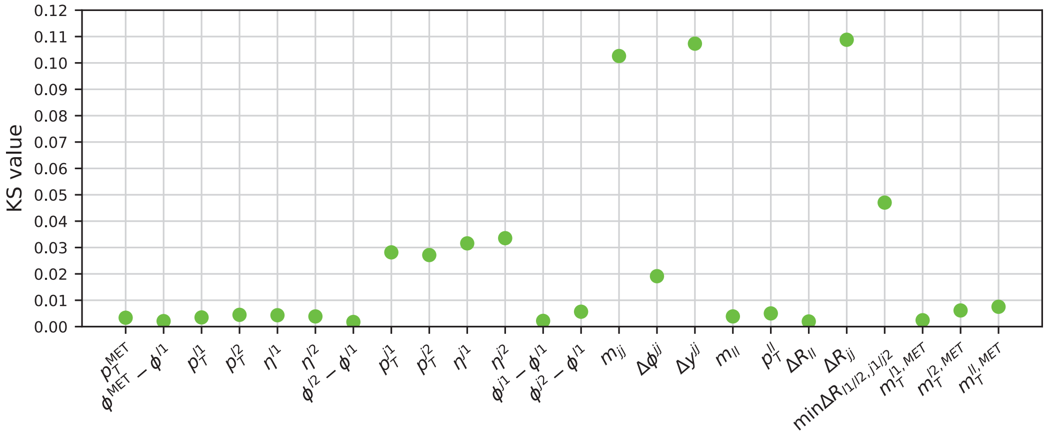
A correction for higher-order QCD effects on the different polarizations has to be combined with a correction for the used approximations at approximate NLO-QCD+PS. The impact of the NWA at approximate NLO-QCD+PS is negligible with the t-channel approximation as indicated by the small Kolmogorov Smirnov values in figure 8.5b. The direct comparison of $W^{\pm}W^{\pm}jj + 0,1j$ (t-ch.) and $l^{\pm}\nu l^{\pm}\nu jj + 0,1j$ in figure 8.5c shows significant differences in the jet-kinematics. As shown in figure 8.6, events with low m_{jj} and low Δy_{jj} are dropped by excluding the s-channels. At approximate NLO-QCD+PS, the leading jets in s-channel-like diagrams do not necessarily originate directly from the colliding hadrons. Thus, their m_{jj} and Δy_{jj} can have significantly lower values than for t-channel-like contri-



(a) SHERPA $W^{\pm}W^{\pm}jj$ vs $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.)



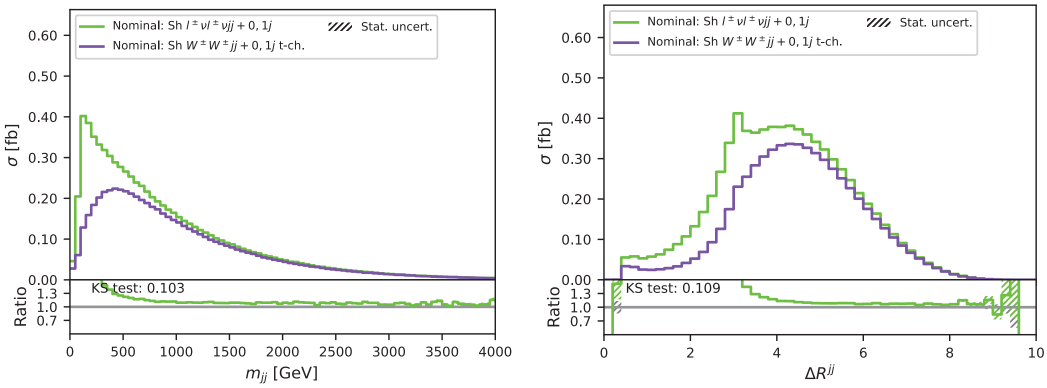
(b) SHERPA $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) vs $l^{\pm}l^{\pm}\nu\nu jj+0,1j$ (t-ch.)



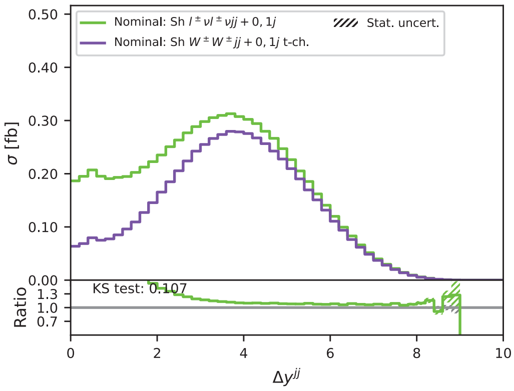
(c) SHERPA $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) and $l^{\pm}l^{\pm}\nu\nu jj+0,1j$

Figure 8.5: The Kolmogorov Smirnov values for SHERPA $W^{\pm}W^{\pm}jj$ vs $W^{\pm}W^{\pm}jj$ (t-ch.), $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) vs $l^{\pm}l^{\pm}\nu\nu jj+0,1j$ (t-ch.), and $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) and $l^{\pm}l^{\pm}\nu\nu jj+0,1j$ on particle-level in the phase space defined in table 8.4. The polarization is defined in the WW-cmf.

butions. However, it is less obvious why a peak-like structure arises for ΔR_{jj} in figure 8.6b. Therefore, a short study will be done in the following to show that no deeper physical correlation is required to produce this peak.



(a) m_{jj} **(b)** ΔR_{jj}



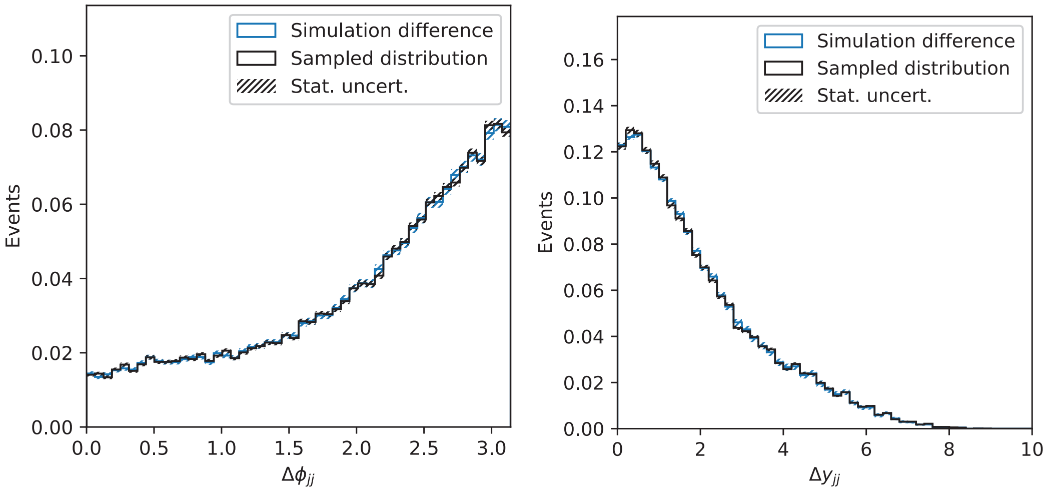
(c) Δy_{jj}

Figure 8.6: Comparison of the differential cross-sections simulated by Sherpa $W^\pm W^\pm jj + 0, 1j$ (t-ch.) and $l^\pm l^\pm \nu\nu jj + 0, 1j$ on particle-level in the phase space defined in table 8.4.

ΔR_{jj} Peak for s-Channel Contributions at Higher-Order QCD

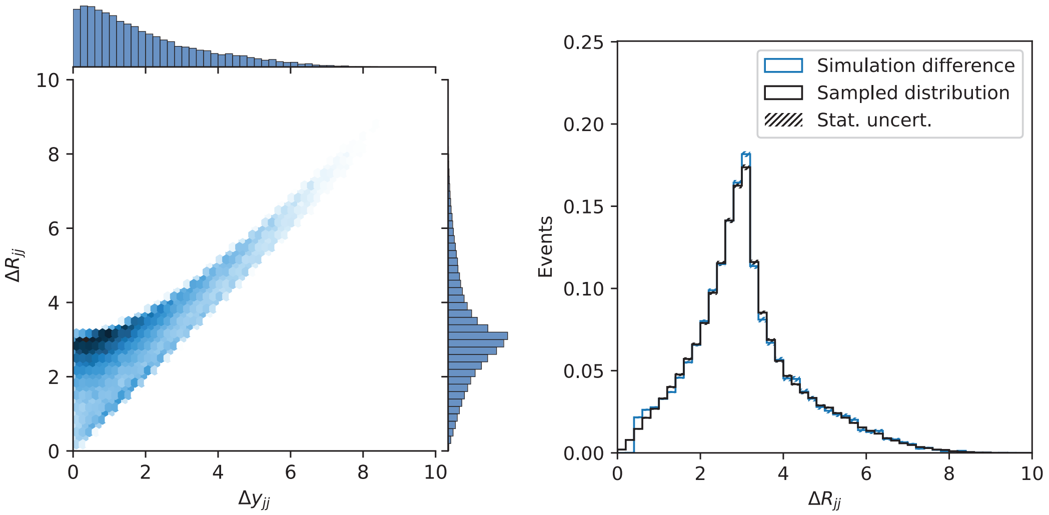
The difference between $W^\pm W^\pm jj$ (t-ch.) and $l^\pm \nu l^\pm \nu jj + 0, 1j$ in the Δy_{jj} and $\Delta \phi_{jj}$ distribution is investigated to study the additional s-channel-like contribution. 100000 tuples of Δy_{jj} and $\Delta \phi_{jj}$ values are sampled following these differential differences between $W^\pm W^\pm jj + 0, 1j$ (t-ch.) and $l^\pm l^\pm \nu\nu jj + 0, 1j$. The resulting distributions are shown in figure 8.7. The random values are created without any correlation between Δy_{jj} and $\Delta \phi_{jj}$.

The tuples can be used to calculate the resulting $\Delta R_{jj} = \sqrt{(\Delta y_{jj})^2 + (\Delta \phi_{jj})^2}$ distribution. The scatter plot figure 8.8a shows the origin of the peak-like structure in ΔR_{jj}. The increasing probability of s-channel-type contributions for high separation in $\Delta \phi_{jj}$ leads to the increase for ΔR_{jj} until $\Delta R_{jj} \approx \pi$. For values higher than $\Delta R_{jj} \approx \pi$, a stronger separation in Δy_{jj} is required. However, a strong separation in Δy_{jj} is not very likely for s-channel-like diagrams. The preferred small separation in Δy_{jj} causes the rapid probability decrease for events with $\Delta R_{jj} \gtrapprox \pi$ and thus the resulting peak structure. The sampled distribution in figure 8.8b is in an excellent agreement with the actual difference between $W^\pm W^\pm jj$ (t-ch.) and $l^\pm \nu l^\pm \nu jj + 0, 1j$. The difference between the sampled distribution and the actual additional s-channel-type contributions at ΔR_{jj} close to 0 arises from the required minimal separation of two jets, which was not considered when generating the tuples. Thus, it is confirmed that no additional correlation is required to explain the peak in ΔR_{jj}.



(a) $\Delta\phi_{jj}$ **(b)** Δy_{jj}

Figure 8.7: The difference between $W^{\pm}W^{\pm}jj$ (t-ch.) and $l^{\pm}\nu l^{\pm}\nu jj + 0, 1j$ in the Δy_{jj} and $\Delta\phi_{jj}$ distribution in the kinematic region defined in table 8.4 at particle-level. These simulations differences are sampled by 100000 tuples of Δy_{jj} and $\Delta\phi_{jj}$.



(a) Δy_{jj} vs. ΔR_{jj} **(b)** ΔR_{jj}

Figure 8.8: The difference between $W^{\pm}W^{\pm}jj$ (t-ch.) and $l^{\pm}\nu l^{\pm}\nu jj+0, 1j$ in the Δy_{jj} and $\Delta\phi_{jj}$ distributions is sampled by 100000 tuples of Δy_{jj} and $\Delta\phi_{jj}$. Figure 8.8a shows the resulting two-dimensional distribution in Δy_{jj} and ΔR_{jj}. In figure 8.8b, the sampled ΔR_{jj} distribution is compared to the difference between $W^{\pm}W^{\pm}jj$ (t-ch.) and $l^{\pm}\nu l^{\pm}\nu jj + 0, 1j$.

8.1.4 Multivariate Correction

Since the polarization measurement is a multivariate analysis, a reweighting in the full phase space for the approximate NLO-QCD+PS effects represents the most accurate solution. Another advantage of a multivariate correction is the ability to cover migrations on particle-level into the reconstruction-level fit region.

As described in Ref. [88], a binary classifier trained with binary cross-entropy can be used for reweighting in the entire phase space. To apply this method both categories must have the same support in the phase space. This is already confirmed by the sample comparison in section 8.1.3: All the simulations listed in table 8.1 cover the same kinematic ranges in the phase space given by the loose particle-level event selection in table 8.4.

The simulations by the Monte Carlo event generator provide a probability density $p(x)$ for polarized $W^{\pm}W^{\pm}jj$-EW events for given phase space points x. The required correction is multiplied by the initial distribution in order to reweight the probability $p_{\text{initial}}(x)$ to the target probability $p_{\text{target}}(x)$ for each point x. The multivariate correction approximates the ratio of the probabilities $p_{\text{initial}}(x)$ and $p_{\text{target}}(x)$ given by the two simulations for an event x. The resulting ratio

$$\frac{p_{\text{target}}(x)}{p_{\text{initial}}(x)} \approx \frac{\text{DNN}(x)}{1-\text{DNN}(x)} \tag{8.3}$$

can be directly derived by the output of the classifier trained to separate these two samples. The DNN is trained for the binary classification with the training labels $y_{\text{initial}} = 0$ and $y_{\text{target}} = 1$. If an event can be assigned to both distributions with equal probability, the DNN output is 0.5, resulting in a reweighting factor of 1. Events that can be assigned more to the initial distribution with $\text{DNN}(x) < 0.5$ are weighted down, and events with $\text{DNN}(x) > 0.5$ are weighted up.

To derive both correction factors in equation (8.1) sets of classifiers are trained. The classifier $\text{DNN}^{pol}_{0,1j}$ is trained to separate Sherpa $W^{\pm}W^{\pm}jj$ and $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) for each of the polarization states in the WW-cmf and pp-cmf. For the correction of $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) to $l^{\pm}l^{\pm}\nu\nu jj+0,1j$ the neural network $\text{DNN}_{\text{s-ch.}}$ is trained on unpolarized events. In contrast to the classifiers trained in chapter 9, the event weights are applied during the training without normalizing the two categories to the same sum of weights. Thus, the DNNs are also expected to learn the overall cross-section differences of the compared samples. With these networks the correction $\text{corr}_{\text{jet-merged}}$ in equation (8.1) can be written as

$$\text{corr}^{\text{multi-var}}_{\text{jet-merged}} = \frac{\text{DNN}^{pol}_{0,1j}}{1-\text{DNN}^{pol}_{0,1j}} \cdot \frac{\text{DNN}_{\text{s-ch.}}}{1-\text{DNN}_{\text{s-ch.}}}, \tag{8.4}$$

and the correction for the missing s-channels in equation (8.2) is given by

$$\text{corr}^{\text{multi-var}}_{\text{s-ch.}} = \frac{\text{DNN}_{\text{s-ch.}}}{1-\text{DNN}_{\text{s-ch.}}}. \tag{8.5}$$

Train the Correction Networks

The particle-level training data was created without τ lepton channels. Thus, the lepton kinematics are not used for the training to avoid a bias due to the missing τ lepton decays. However, the correction quality is not expected to decrease significantly since the dominant effects investigated in section 8.1.3 are found in the jet phase space. Despite this limitation of the phase space, the multivariate correction still provides very good agreement, as shown in section 8.1.5.

The DNNs are trained on the particle-level kinematics listed in table 8.6 together with their non-linear scaling to result in Gaussian-like distributions. The number of available training events is listed in table 8.5. The hyperparameters of the networks are chosen by the OPTUNA optimization procedure described in section 9.3.1. As this algorithm was developed in Ref. [89] specifically for the polarization discrimination in this analysis, it is described in detail in chapter 9 for the discriminator networks. This algorithm is reused to select suitable hyperparameters for the multivariate higher-order QCD correction. The hyperparameter space used for the optimization is listed in table 8.7. The optimization is performed by searching for a set of hyperparameters that maximizes or minimizes a chosen optimization metric. Since the approximation of the probability fraction in equation (8.3) relies on the training on the binary cross-entropy, the validation loss is also chosen as the optimization metric.

Training a network on a given dataset always introduces a bias towards this training dataset. As described in section 5.8, the k-fold method can be used to avoid this bias by using an iterating split into training, validation, and test data. Since the $\text{DNN}^{pol}_{0,1j}$ is trained on Sherpa samples and will only be applied to the MadGraph samples, no bias is introduced. Thus, no test dataset is required for training and hyperparameter optimization

Table 8.6: Variables used to optimize and train the DNNs for the higher-order QCD correction. The variables were scaled for the training and application.

Kinematics	Descriptions	scaling
p_T^{j1}	p_T of the leading jet	$\log_{10}(x)$
η^{j1}	η of the leading jet	x
p_T^{j2}	p_T of the subleading jet	$\log_{10}(x)$
η^{j2}	η of the subleading jet	x
ΔR_{jj}	ΔR between the two leading jets	x
Δy_{jj}	Δy between the two leading jets	x
$\Delta\phi_{jj}$	$\Delta\phi$ between the two leading jets	x
m_{jj}	Invariant mass of the two leading jets	$\log_{10}(x)$

Table 8.7: DNN hyperparameters for the multivariate higher-order QCD correction. For the tunable hyperparameters, the search space and the sampling distribution are listed.

Fixed hyperparameters		
Hyperparameter	Value	
Optimizer	Adam	
Loss function	binary cross-entropy	
Activation function	swish	
Optimized hyperparameters		
Hyperparameter	Seach space	Parameter sampling
Number of layers	$[2, 6]$	Uniform
Neurons per layer	$[32, 128]$	Uniform
Dropout rate	$[0.0, 0.5]$	Uniform
Batch size	$[32, 512]$	Logarithmic
Learning rate	$[10^{-5}, 10^{-2}]$	Logarithmic
Adam β_1	$[10^{-4}, 0.99]$	Logarithmic
Adam $1-\beta_2$	$[10^{-5}, 0.9999]$	Logarithmic
Adam ϵ	$[10^{-10}, 1.0]$	Logarithmic

and the resulting networks can be applied on MadGraph without the need for a k-fold application.

In contrast, the $\text{DNN}_{\text{s-ch.}}$ is also applied to the Sherpa $W^{\pm}W^{\pm}jj + 0, 1j$ (t-ch.) sample for the polarization measurement. Therefore, the 5-fold method (k-fold method with $k = 5$) is required for the reweighting with $\text{DNN}_{\text{s-ch.}}$. For the polarization discriminator DNNs in chapter 9, a split into a training, validation, and test dataset will be used. However, for the reweighting $\text{DNN}_{\text{s-ch.}}$, a training and validation set without an additional test set is sufficient. The main bias toward the one validation set used for the hyperparameter optimization is prevented by retraining the best network with a different seed after the optimization. The remaining bias is negligible since this network is only used for the reweighting of the Standard Model prediction and not applied on measured data. Consequently, the networks are trained on 80% of the data and the remaining 20% are used for validation. For the application to Madgraph, using only one of the five trained networks is sufficient. For application on the Sherpa polarization sample, the 5-fold method has to be used to avoid bias.

Since the $\text{DNN}_{0,1j}^{pol}$ networks are expected to be quite similar for all polarizations, their optimization is only done for LL in WW-cmf. The same hyperparameters are used to train networks for the other polarization states and only the number of training epochs is updated. The optimization results are given in table 8.8. The corresponding training progress and the fits to the validation loss during the optimization are given in figures B.2 and B.3.

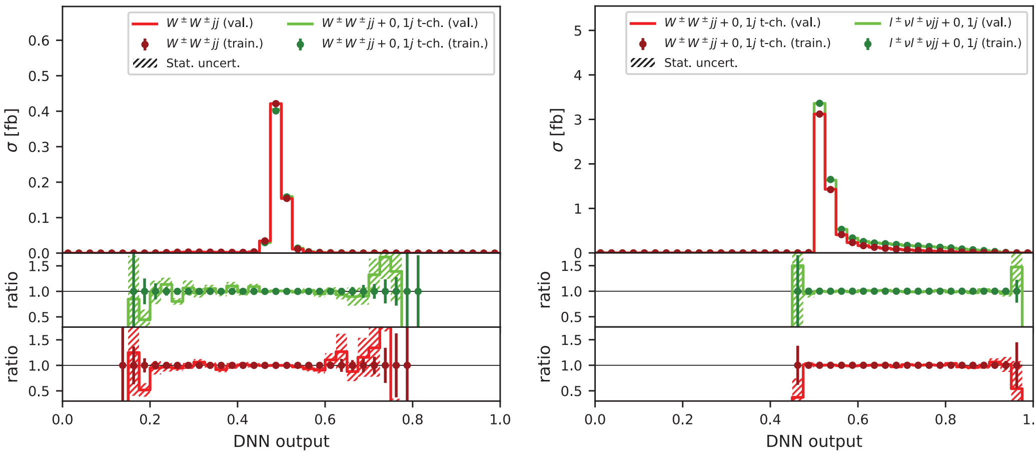
The output distribution of $\text{DNN}_{0,1j}^{\text{LL WW-cmf}}$ and $\text{DNN}_{\text{s-ch.}}$ are shown in figure 8.9. As shown in the appendix in figure B.4, the output distribution of $\text{DNN}_{0,1j}^{\text{LL WW-cmf}}$ is consistent with the output of $\text{DNN}_{0,1j}^{pol}$ trained for the other polarization states. The classification by $\text{DNN}_{0,1j}^{pol}$ peaks slightly below 0.5 since the Sherpa $W^{\pm}W^{\pm}jj$ and $W^{\pm}W^{\pm}jj + 0, 1j$ (t-ch.) sample have very similar kinematic and $W^{\pm}W^{\pm}jj$ has a slightly higher cross-section in table 8.5. Therefore, the correction factors given by $\text{DNN}_{0,1j}^{pol}/(1-\text{DNN}_{0,1j}^{pol})$ is expected to be marginally

Table 8.8: Hyperparameters chosen for the neural network training. The shown hyperparameters were chosen in the final step of the hyperparameter optimization.

Hyperparameter	$\text{DNN}_{0,1j}^{pol}$	$\text{DNN}_{\text{s-ch.}}$
Number of layers	6	4
Neurons per layer	45	74
Dropout	0.0536796	0.0650895
Activation function	swish	swish
Batch size	996	840
Learning rate	9.56887e-05	0.000273763
Adam beta_1	0.680364	0.980238
$1-$ Adam beta_2	0.000155779	0.00256898
Adam ϵ	2.2657e-05	1.83744e-06
epochs	LL in WW-cmf: 77 TL in WW-cmf: 49 TT in WW-cmf: 83 LL in pp-cmf: 39 TL in pp-cmf: 35 TT in pp-cmf: 56	96

less than 1. This is expected from the comparison of $W^{\pm}W^{\pm}jj$ and $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) in figure 8.4, which showed good agreement in most of the phase space except for the lower values at $m_{jj} < 200\,\text{GeV}$.

The $\text{DNN}_{\text{s-ch.}}$ provides a significantly stronger separation. For most of the events, the classification peaks at 0.5. For the phase space region with significant s-channel contributions, the network prefers the classification as $l^{\pm}\nu l^{\pm}\nu jj+0,1j$. This higher network score leads to upward correction that covers the missing s-channels in $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) at low m_{jj} and low Δy_{jj}.



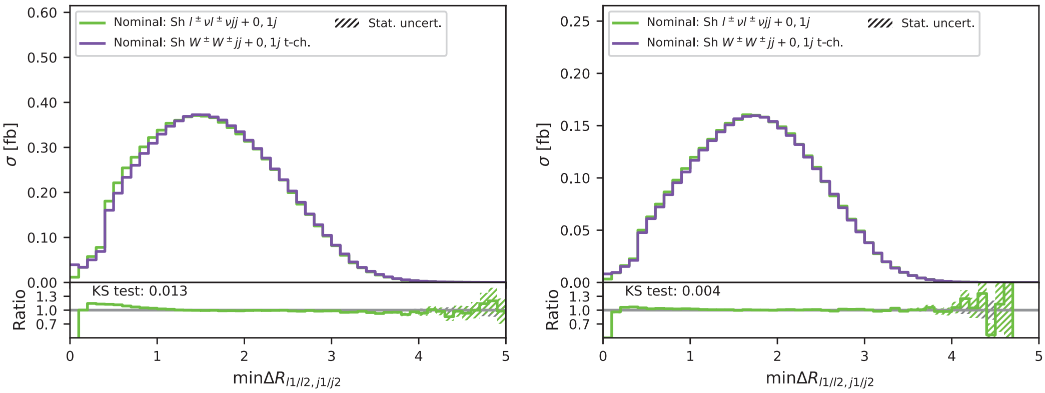
(a) $\text{DNN}_{0,1j}$ for LL in WW-cm

(b) $\text{DNN}_{\text{s-ch.}}$

Figure 8.9: The classification output of $\text{DNN}_{0,1j}^{\text{LL}\,WW\text{-cmf}}$ and $\text{DNN}_{\text{s-ch.}}$ evaluated on their training and validation data.

8.1.5 Technical Validation of the Multivariate Correction

The multivariate higher-order QCD corrections are tested in the loose kinematic region given in table 8.4. To avoid a bias the networks are applied using the k-fold method on the corresponding validation data used during training. An overview of the shape effects of the multivariate correction is given by the Kolmogorov Smirnov values in figure 8.11 before (dots) and after (crosses) the correction. As indicated by the Kolmogorov-Smirnov values before and after the correction, the closure is improved in the full phase space which would not be possible with a correction in only one variable. The closure in individual kinematic variables in the jet phase space after the multivariate correction is given in figures B.5 to B.7. After correcting $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) for missing s-channels, only the $min(\Delta R_{l1/l2,j1/j2})$ distribution in figure 8.10a shows significant differences to the $l^{\pm}l^{\pm}\nu\nu jj+0,1j$ distribution. However, the next step of the closure test will show that this discrepancy mostly vanishes when using a tighter event selection.

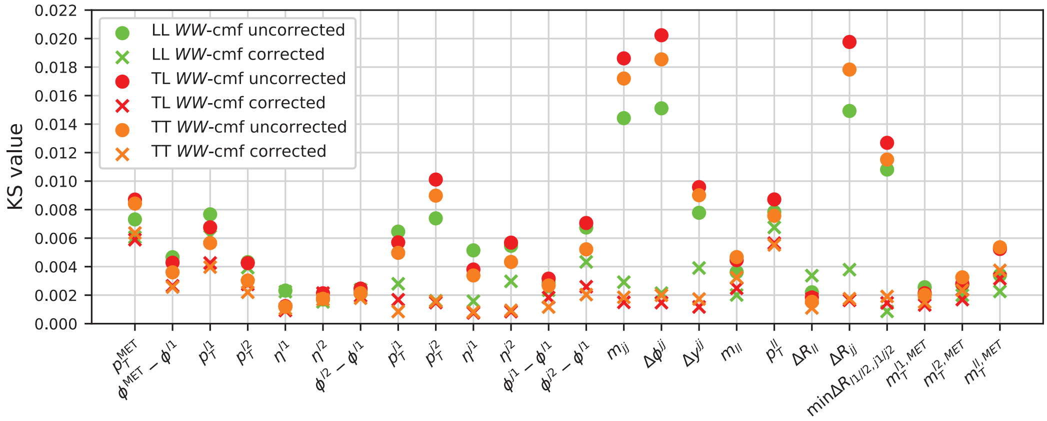


(a) Loose event selection defined in table 8.4 **(b)** Tight event selection defined in table 8.2

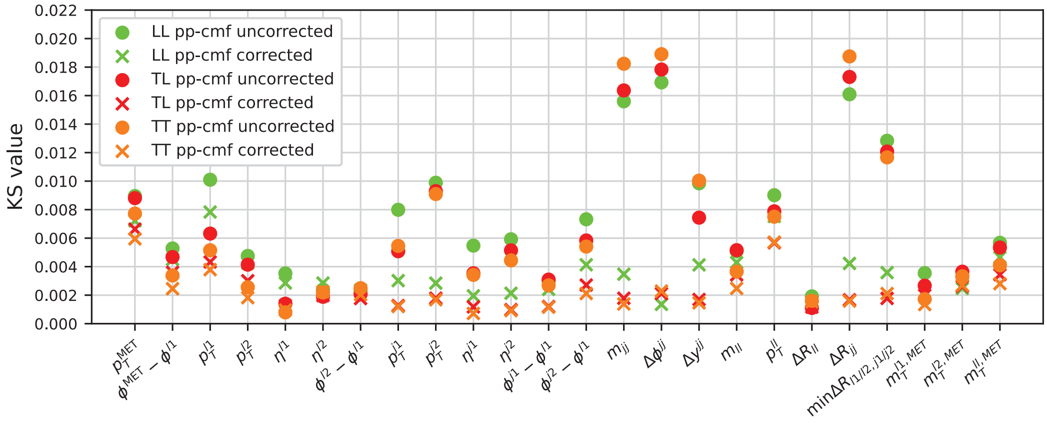
Figure 8.10: Comparison of the differential cross-sections simulated by Sherpa $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) and $l^{\pm}l^{\pm}\nu\nu jj+0,1j$ in $min(\Delta R_{l1/l2,j1/j2})$ on particle-level. The Sherpa $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) is corrected by the DNN$_{\text{s-ch.}}$. The networks are applied in the k-fold method to the corresponding validation data used during their training to avoid a bias toward the training data. The comparison is performed for the two event selections defined in table 8.4 and table 8.2.

Since the shown phase space is much broader than the one used for the polarization measurement, a second validation is done in the phase space defined in table 8.2 corresponding to the particle-level selection of the signal and Low-m_{jj} region. Closure in this region would imply that the multivariate corrections work sufficiently well in the phase spaces relevant to the polarization measurement. The shape improvement from the correction by DNN$^{pol}_{0,1j}$ in the WW-cmf is shown in figure 8.12. The results in the pp-cmf are equivalent and therefore given in the appendix in figure B.8. The agreement between the $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) and the reweighted $W^{\pm}W^{\pm}jj$ sample is excellent in the entire phase space. The comparison plots in individual variables of the lepton and jet phase space are given in figures B.9 to B.12. The effects of the correction by DNN$_{\text{s-ch.}}$ are shown in figure 8.13. The low Kolmogorov-Smirnov values indicate an excellent shape agreement in the entire kinematic phase space. This closure is confirmed by the plots of individual kinematic variables given in figures B.13 and B.14. As shown in figure 8.10b, the agreement in $min(\Delta R_{l1/l2,j1/j2})$ is significantly improved by the tighter event selection.

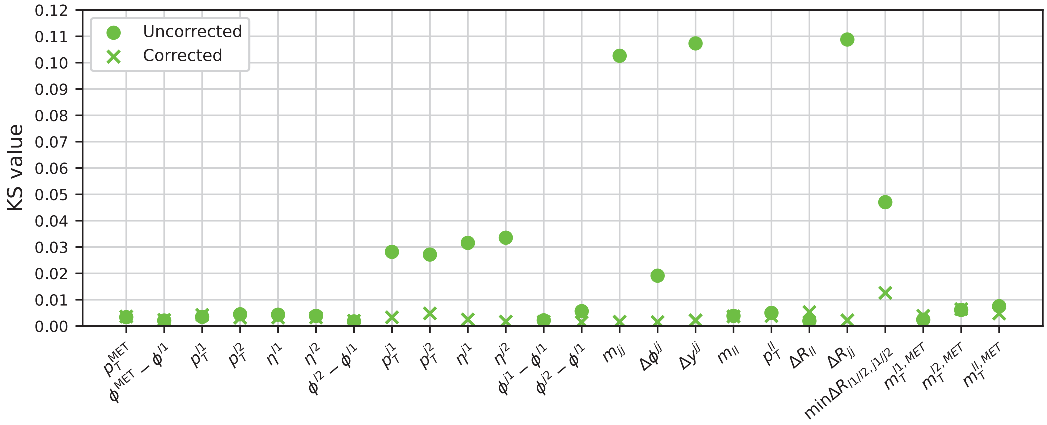
Therefore, the multivariate correction corrects the shapes in the entire phase space. The agreement with the expected distributions is almost perfect in the signal and Low-m_{jj} region. The remaining differences are tiny and thus not expected to impact the statistically limited polarization analysis. Thus, the closure test validates the multivariate correction procedure to correctly introduce the effects of an additional real QCD emission in the matrix element and the contribution from missing s-channels.



(a) KS values for Sherpa $W^{\pm}W^{\pm}jj$ vs $W^{\pm}W^{\pm}jj$ (t-ch.) (polarization in WW-cm)

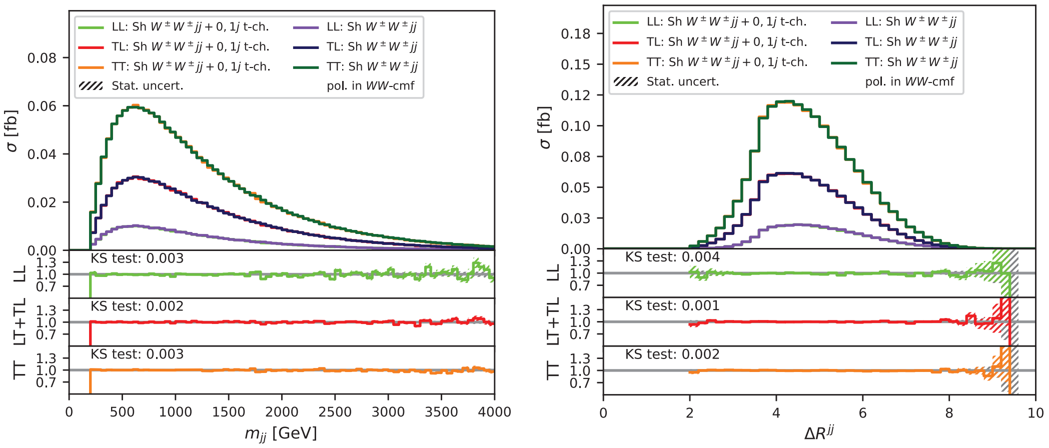


(b) KS values for Sherpa $W^{\pm}W^{\pm}jj$ vs $W^{\pm}W^{\pm}jj$ (t-ch.) (polarization in pp-cmf)



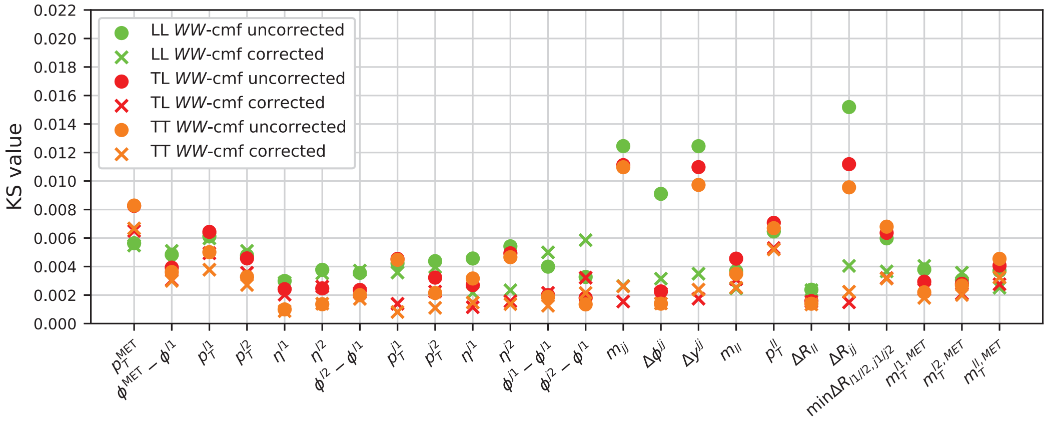
(c) KS values for Sherpa $W^{\pm}W^{\pm}jj + 0,1j$ (t-ch.) and $l^{\pm}l^{\pm}\nu\nu jj + 0,1j$

Figure 8.11: Comparison of the differential cross-sections simulated by Sherpa $W^{\pm}W^{\pm}jj$, $W^{\pm}W^{\pm}jj$ (t-ch.), and $l^{\pm}l^{\pm}\nu\nu jj + 0,1j$ on particle-level in the loose region defined in table 8.4 before (dots) and after (crosses) the mutlivariate correction. In figures 8.11a and 8.11b $W^{\pm}W^{\pm}jj$ is corrected by the $\text{DNN}^{pol}_{0,1j}$ and in figure 8.11c $W^{\pm}W^{\pm}jj$ (t-ch.) is corrected by $\text{DNN}_{\text{s-ch.}}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training.



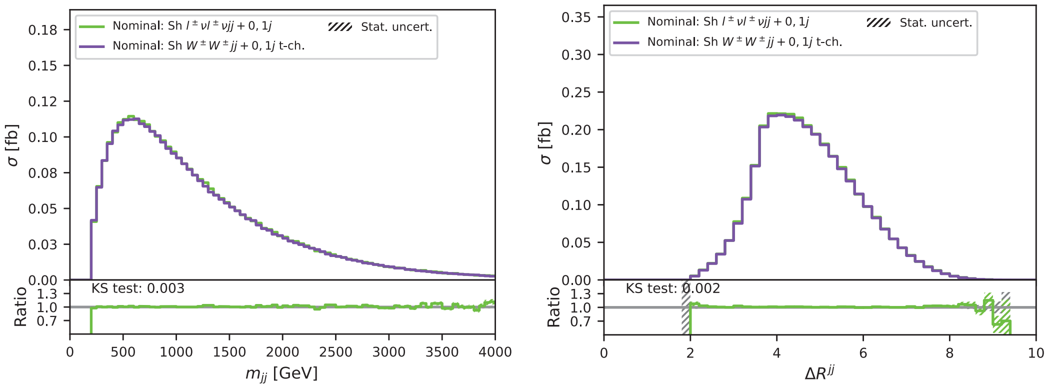
(a) m_{jj}

(b) ΔR_{jj}

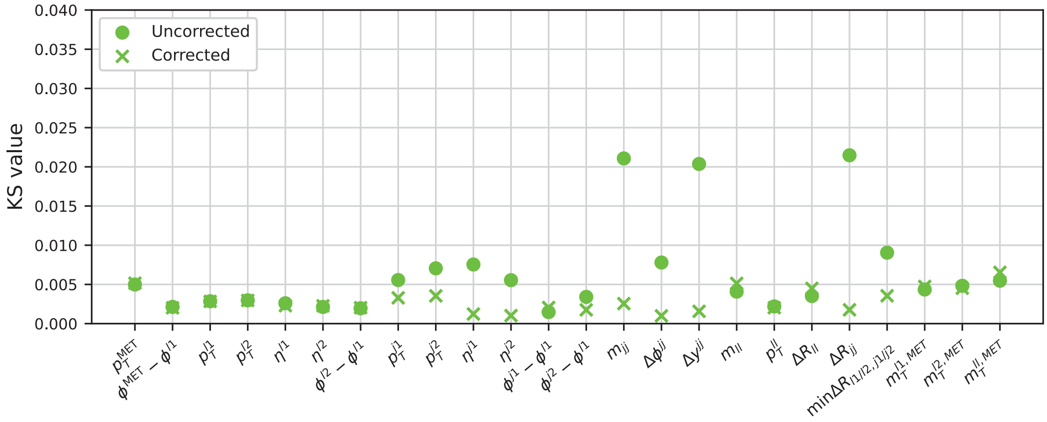


(c) KS values before (dots) and after (crosses) the multivariate correction

Figure 8.12: Comparison of $W^\pm W^\pm jj$ and $W^\pm W^\pm jj$ (t-ch.) with polarization defined in the WW-cmf simulated by Sherpa on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The Sherpa $W^\pm W^\pm jj$ is corrected by the $\text{DNN}^{pol}_{0,1j}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training. The Kolmogorov-Smirnov values give an overview of the shape improvements in the tested variables.



(a) m_{jj} **(b)** ΔR_{jj}



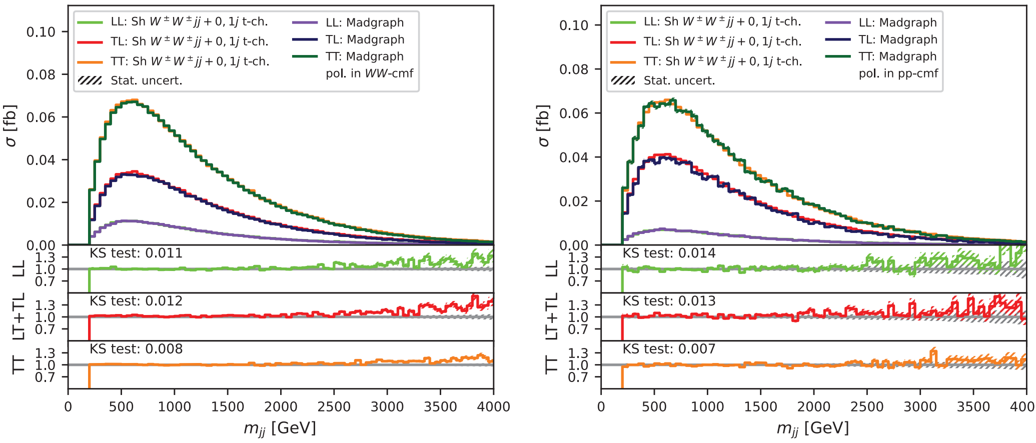
(c) KS values before (dots) and after (crosses) the multivariate correction

Figure 8.13: Comparison of the differential cross-sections simulated by Sherpa $W^\pm W^\pm jj + 0, 1j$ (t-ch.) and $l^\pm l^\pm \nu\nu jj + 0, 1j$ on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The Sherpa $W^\pm W^\pm jj + 0, 1j$ (t-ch.) is corrected by the $\text{DNN}_{\text{s-ch.}}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training. The Kolmogorov-Smirnov values give an overview of the shape improvements in the tested variables.

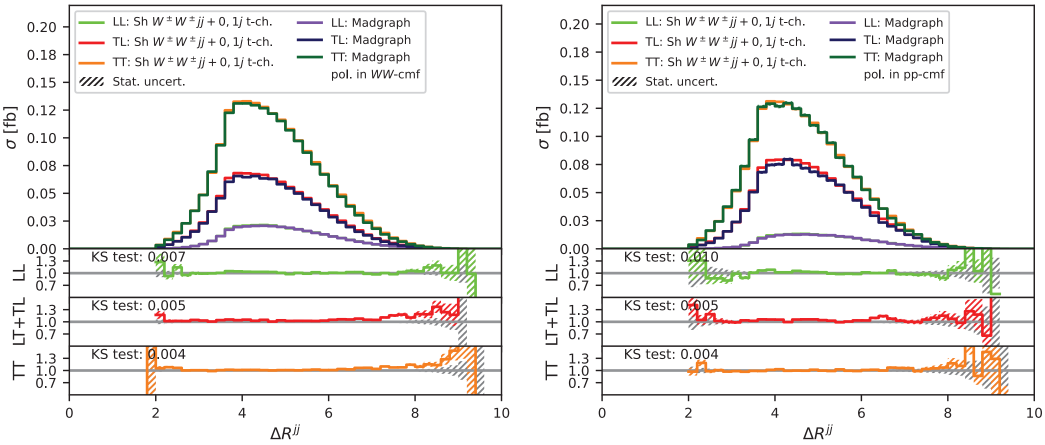
8.1.6 Validation of the Multivariate Correction with MadGraph

In the previous section, the validity of the multivariate correction for Sherpa simulations is validated. However, the higher-order correction is also applied to the polarization predictions generated by MadGraph. Therefore, a second test is done to validate that the multivariate correction can be applied independently of the generator of the polarization sample. The kinematic region in table 8.2 equivalent to the signal and Low-m_{jj} region selection is used for this validation on particle-level.

The target distribution of this closure test is the Sherpa $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) sample corrected by the $\text{corr}^{\text{multi-var}}_{\text{s-ch.}}$. The $\text{corr}^{\text{multi-var}}_{\text{s-ch.}}$ correction in equation (8.5) corrects the missing s-channels in Sherpa $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) as validated in section 8.1.5. Since the MadGraph event generator simulates the polarization at LO-QCD+PS, the combined multivariate correction $\text{corr}^{\text{multi-var}}_{\text{jet-merged}}$ is applied. The closure between the corrected MadGraph and Sherpa predictions is shown in figure 8.14. The agreement between the resulting distributions is as good as for the LO-QCD+PS simulation in section 8.1.1. Thus, the corrections extracted from Sherpa can be applied to MadGraph without any significant bias.



(a) WW-cmf: jet m_{jj} **(b)** pp-cmf: jet m_{jj}



(c) WW-cmf: jet ΔR_{jj} **(d)** pp-cmf: jet ΔR_{jj}

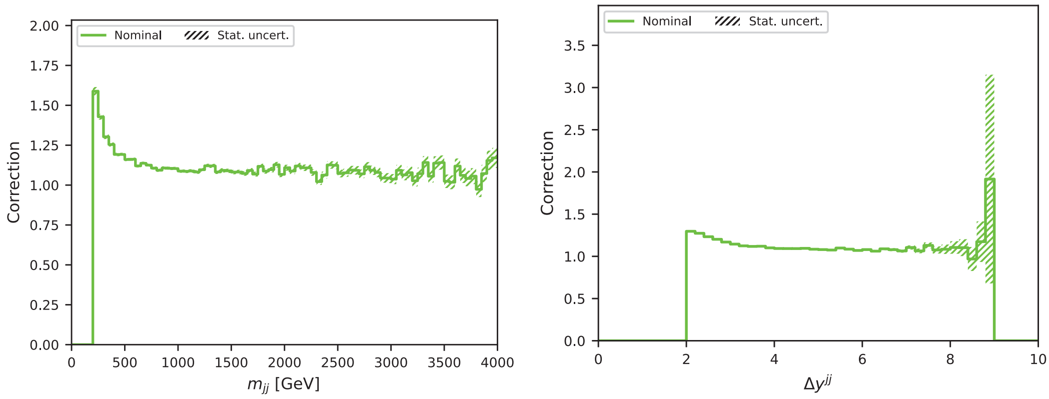
Figure 8.14: Comparison of the differential polarized cross-sections simulated by Sherpa $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) and MadGraph at particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The Sherpa sample is corrected by $\text{corr}^{\text{multi-var}}_{\text{s-ch.}}$ and the event weights of the MadGraph sample are multiplied with $\text{corr}^{\text{multi-var}}_{\text{jet-merged}}$. The polarizations are defined in the WW-cmf for figures 8.14a and 8.14c and in the pp-cmf for figures 8.14b and 8.14d.

8.1.7 One-dimensional Correction

The multivariate correction is expected to correct the entire jet phase space and is, therefore, taken as the nominal correction. To assign uncertainty to the multivariate correction procedure an alternative one-dimensional correction in only one kinematic variable is performed. The resulting difference to the multivariate correction provides an uncertainty for the correction procedure.

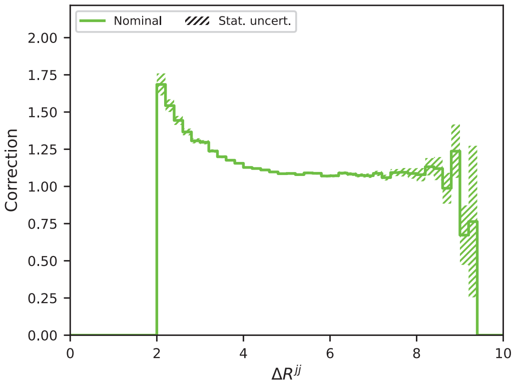
Choice of the Correction Variable

The corrections in different variables are compared to find the best candidate for the one-dimensional correction. Since a one-dimensional correction cannot handle multivariate migrations from outside the particle-level phase space into the reconstruction-level phase space, the kinematic phase space in table 8.2 is investigated to test the corrections. This particle-level selection is equivalent to the signal and Low-m_{jj} region used for the polarization measurement. To simplify the one-dimensional correction, the corrections for higher-order QCD effects and missing s-channels at approximate NLO-QCD+PS are done using the same kinematic variable. As already shown in figure 8.11, the main shape difference is seen in the m_{jj}, Δy_{jj}, and ΔR_{jj} distributions due to the missing s-channels at approximate NLO-QCD+PS. Since the difference between $W^{\pm}W^{\pm}jj$ and $W^{\pm}W^{\pm}jj + 0,1j$ (t-ch.) is significantly smaller than between $W^{\pm}W^{\pm}jj + 0,1j$ (t-ch.) and $l^{\pm}l^{\pm}\nu\nu jj + 0,1j$, the focus of this section is on finding the most suitable variable for the correction for the missing s-channels. The resulting binned corrections for missing s-channel contributions are shown in figure 8.15.



(a) Correction factor in m_{jj}

(b) Correction factor in Δy_{jj}



(c) Correction factor in ΔR_{jj}

Figure 8.15: The missing s-channel correction in equation (8.2) derived in m_{jj}, Δy_{jj}, and ΔR_{jj}. The corrections are extracted in the combined signal and Low-m_{jj} region defined in table 8.2.

The effects of bin-wise corrections in m_{jj}, Δy_{jj}, and ΔR_{jj} are summarized in figure 8.16. All the shown corrections significantly improve the agreement in the di-jet kinematic. In contrast to the multivariate correction, however, the one-dimensional corrections also worsen the agreement in some kinematics. The biggest differences are in the jet p_T and m_{jj} distributions after the correction in Δy_{jj}. The correction in m_{jj} and ΔR_{jj} lead to a similar agreement in the phase space and are therefore both valid options for the one-dimensional correction. Since the electroweak NLO correction in section 8.2 is implemented in m_{jj}, the technical implementation of the one-dimensional higher-order QCD correction in m_{jj} is more straightforward and therefore preferable.

Extraction of the One-dimensional Correction

To extract the one-dimensional correction, the particle-level phase space in table 8.2 corresponding to the signal and Low-m_{jj} region is used, except for the m_{jj} criteria to cover potential migrations. The correction $\mathrm{corr}_{\text{jet-merged}}$ in equation (8.1) and $\mathrm{corr}_{\text{s-ch.}}$ in equation (8.2) as function of m_{jj} are derived by the fraction of the corresponding histograms. The corrections are smoothed using a fit function to avoid jumps at the bin edges. The fit functions considered all have the term $ae^{-\alpha x}$ to describe the behavior at high m_{jj} values and an additional term to describe low m_{jj} range. The fit functions considered are:

$$f_1(m_{jj}) = ae^{-\alpha \frac{m_{jj}}{\mathrm{GeV}}} + \frac{b}{|\frac{m_{jj}}{\mathrm{GeV}} - c|} \tag{8.6}$$

$$f_2(m_{jj}) = ae^{-\alpha \frac{m_{jj}}{\mathrm{GeV}}} + \frac{b}{1 + e^{\beta \frac{m_{jj}}{\mathrm{GeV}}}} \tag{8.7}$$

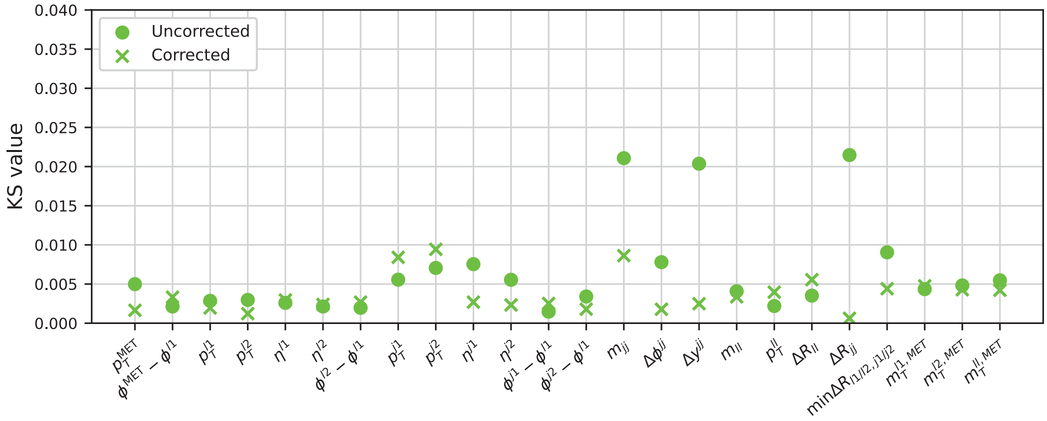
$$f_3(m_{jj}) = ae^{-\alpha \frac{m_{jj}}{\mathrm{GeV}}} + \frac{b}{c + e^{\beta \frac{m_{jj}}{\mathrm{GeV}}}} \tag{8.8}$$

The fit is done by minimizing the χ^2 value between the binned correction factors and the bin-wise integrals of the fit function:

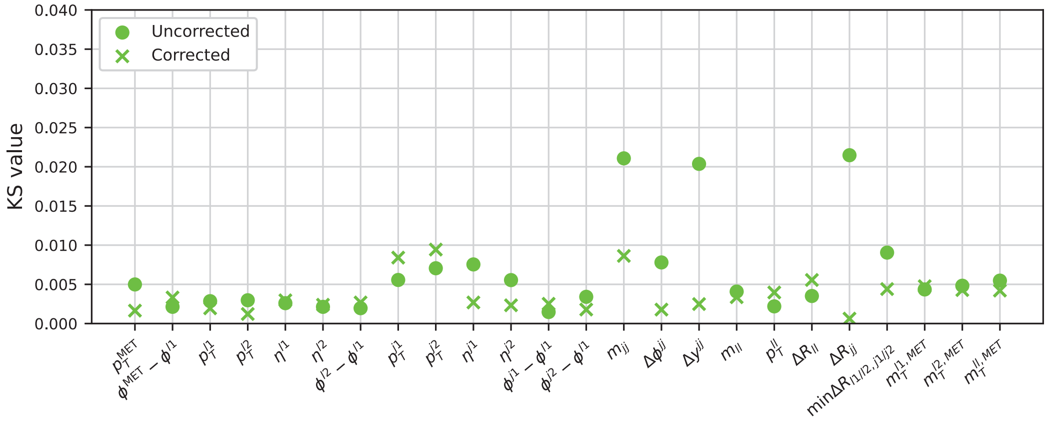
$$\chi^2_{fit} = \sum_i^{N_{\text{bins}}} \frac{1}{\Delta \mathrm{corr}_i} \left(\mathrm{corr}_i - \frac{\int_{\mathrm{bin}_i^{\text{left}}}^{\mathrm{bin}_i^{\text{right}}} f(x)dx}{\mathrm{bin}_i^{\text{right}} - \mathrm{bin}_i^{\text{left}}} \right)^2 \tag{8.9}$$

$\Delta\mathrm{corr}_i$ is the statistical uncertainty of the correction in the corresponding bin. The fit results for the correction in m_{jj} are shown in figure 8.17. Since there is no significant difference between the higher-order QCD corrections for different polarization states, only the result for LL in WW-cmf is shown in figure 8.17. The higher-order corrections for the other polarization states are given in the appendix in figure B.15. For the total higher-order QCD correction $\mathrm{corr}_{\text{jet-merged}}$, the function $f_1(m_{jj})$ results in excellent fit for all polarizations. For the missing s-channel correction at higher-order QCD $\mathrm{corr}_{\text{s-ch.}}$ the function $f_3(m_{jj})$ covers the correction shape over the full m_{jj} range very well. The p-value of only 5% originates from fluctuations in some bins and cannot be covered by the choice of a different fit function. Therefore, the fit function $f_1(m_{jj})$ is used for the total higher-order QCD correction of all polarizations and the fit function $f_3(m_{jj})$ is used for the missing s-channel correction at approximate NLO-QCD+PS. The resulting nominal fit parameters for the corrections are given in tables 8.9 and 8.10.

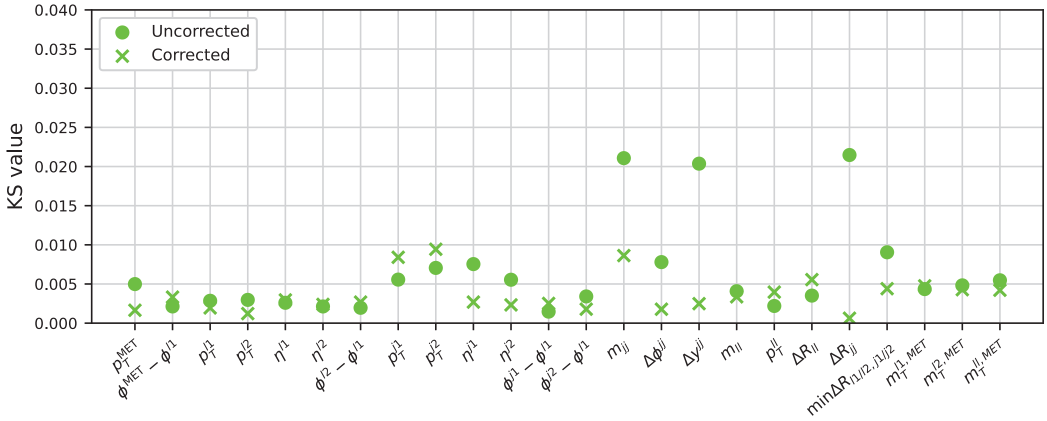
The corrections in m_{jj} represent a much simpler higher-order QCD correction than the multivariate approach in section 8.1.4. Therefore, the multivariate correction is used as the nominal correction. To estimate potential shape effects in the jet phase space not covered by the multivariate correction the multivariate correction is compared with the one-dimensional correction in m_{jj}. This one-dimensional correction covers only shape effects in m_{jj} and is extracted from a significantly tighter phase space than the multivariate correction. Therefore, the difference between the two corrections gives an estimate of missing shape effects not covered by the nominal multivariate correction and is applied as symmetrized uncertainty.



(a) KS values for correction in m_{jj}



(b) KS values for correction in Δy_{jj}



(c) KS values for correction in ΔR_{jj}

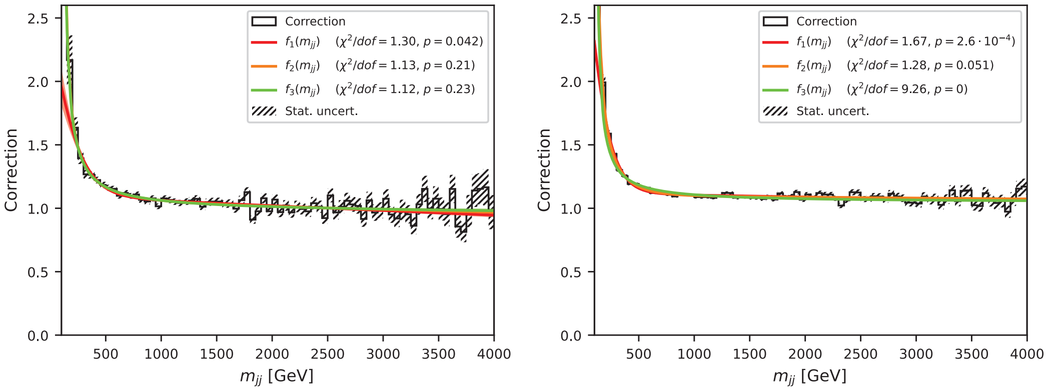
Figure 8.16: The Kolmogorov-Smirnov values for the comparison of $W^{\pm}W^{\pm}jj + 0, 1j$ (t-ch.) and $l^{\pm}l^{\pm}\nu\nu jj + 0, 1j$ in the combined signal and Low-m_{jj} region defined in table 8.2. The $W^{\pm}W^{\pm}jj + 0, 1j$ (t-ch.) distribution is corrected by a correction factor in m_{jj} (figure 8.16a), Δy_{jj} (figure 8.16b), and ΔR_{jj} (figure 8.16c). The correction factors are shown in figure 8.15.

Table 8.9: Fit parameters of $f_1(m_{jj})$ defined in equation (8.6) for the total higher-order QCD correction in m_{jj}.

polarization	$\chi^2_{\text{fit}}/\text{dof}$	p-value	a	α	b	c
LL WW-cmf	1.12	0.23	0.9852	$6.07 \cdot 10^{-6}$	74.37	99.11
LT+TL WW-cmf	0.77	0.93	0.9777	$4.61 \cdot 10^{-6}$	72.01	93.78
TT WW-cmf	0.70	0.98	0.9805	$6.864 \cdot 10^{-6}$	69.90	91.23
LL pp-cmf	1.13	0.21	0.9765	$2.275 \cdot 10^{-7}$	81.50	75.87
LT+TL pp-cmf	0.70	0.98	0.9861	$8.887 \cdot 10^{-6}$	73.75	91.29
TT pp-cmf	0.67	0.99	0.9731	$3.079 \cdot 10^{-6}$	69.32	94.20

Table 8.10: Fit parameters of $f_3(m_{jj})$ defined in equation (8.8) for the correction of missing s-channels at higher-order QCD.

$\chi^2_{\text{fit}}/\text{dof}$	p-value	a	α	b	c	β
1.28	0.05	1.101	$6.304 \cdot 10^{-6}$	0.7253	-1.52	$4.905 \cdot 10^{-3}$



(a) jet-merged corr. for LL in WW-cmf **(b)** Missing s-channel corr. at jet-merged

Figure 8.17: Comparison of different fit functions for the higher-order QCD correction factor $\text{corr}_{\text{jet-merged}}(m_{jj})$ for the LL polarization state in WW-cmf and the missing s-channel correction $\text{corr}_{\text{s-ch.}}(m_{jj})$. The fit functions are defined in equations (8.6) to (8.8).

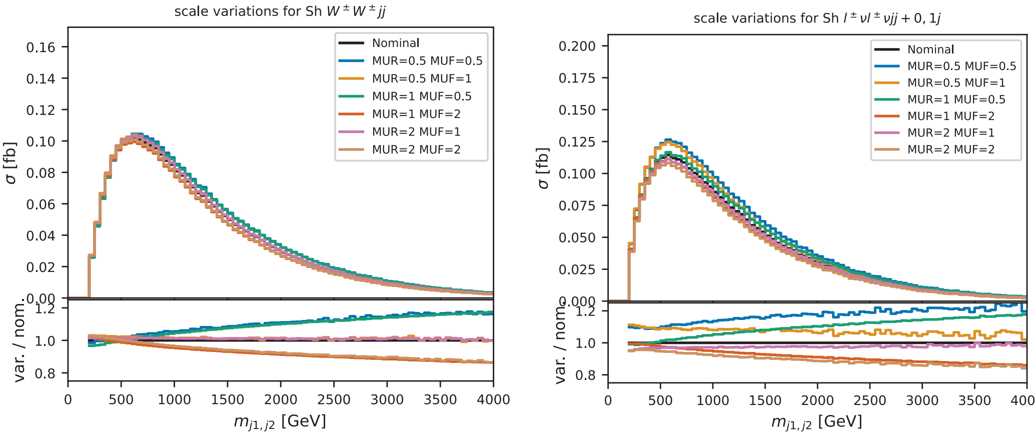
8.1.8 Discussion of Theory Uncertainties

The prediction by SHERPA event generator provide a variety of theory uncertainties. As described in chapter 4, the simulation depends on the choice of the PDF, the value of the strong coupling α_S, the renormalization and factorization scale, the resummation scale, and the merging scale.[1] The extraction of a generally applicable higher-order QCD correction would require considering these uncertainties. However, the same uncertainties also affect the polarization predictions to which the higher-order correction is applied. Therefore, considering the theory uncertainties for the polarization prediction and the higher-order QCD correction leads to a double-counting of the uncertainty. Thus, one must decide at which point the theory uncertainties are introduced.

The particle-level SHERPA samples listed in table 8.1 provide PDF, α_S, and renormalization and factorization scale variations via on-the-fly-weights to calculate the systematic uncertainties. In section 11.2 it is shown that the scale variations represent the dominant uncertainty. Thus, the strategy to apply the theory uncertainty has to give the best estimation of the scale uncertainty. A comparison of the renormalization and factorization scale variations on LO-QCD+PS and approximate NLO-QCD+PS is shown in figure 8.18. The scale variations at LO-QCD+PS do not represent a meaningful estimate of the missing higher-orders in the matrix element calculation. Therefore, the renormalization and factorization scale uncertainty is underestimated at LO-QCD+PS and has to be extracted from the approximate NLO-QCD+PS simulation. The theory uncertainties of the Sherpa $l^{\pm}l^{\pm}\nu\nu jj{+}0,1j$ sample are extracted in section 11.2. Since the theory uncertainties are taken from the target distribution of the higher-order QCD correction they are not additionally studied for the higher-order QCD correction.

---

[1] A detailed study of the theory uncertainties will follow in section 11.2



(a) $W^{\pm}W^{\pm}jj$ **(b)** $l^{\pm}l^{\pm}\nu\nu jj+0,1j$

Figure 8.18: The renormalization and factorization scale variations for the Sherpa $W^{\pm}W^{\pm}jj$ and $l^{\pm}l^{\pm}\nu\nu jj+0,1j$ simulation on particle-level. The events are selected by the combined signal and Low-m_{jj} region as defined in table 8.2.

8.2 NLO-EW Correction

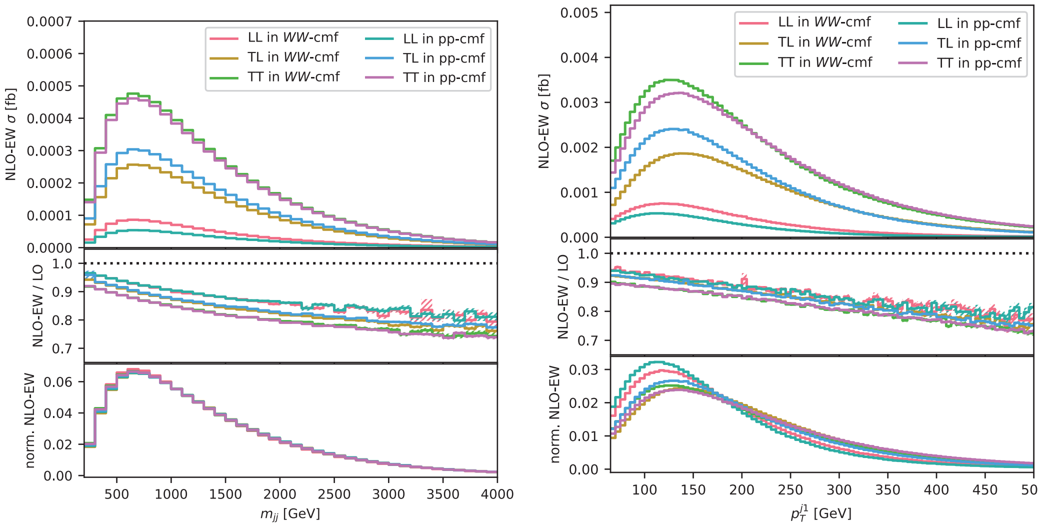
A calculation for higher-order electroweak effects was already available for the unpolarized differential $W^{\pm}W^{\pm}jj$-EW analysis. The authors of Ref. [90] performed a comparison of the m_{jj}-dependent differential cross-section of $W^{\pm}W^{\pm}jj$-EW at Born level ($\mathcal{O}(\alpha_{EW}^{6})$) and at NLO-EW (up to $\mathcal{O}(\alpha_{EW}^{7})$). However, this calculation was not done for individual polarization states. Already existing calculations for inclusive $W^{+}W^{-}$ [91] and ZZ [92] production show that the NLO-EW corrections can significantly differ between the polarizations. With Ref. [50], the first polarization-dependent higher-order calculation for same-charged $W^{+}W^{+}$ scattering was performed. The authors of Ref. [50] repeated this calculation in the parton-level phase space given in table 8.11. The calculation focuses on the $W^{+}W^{+}jj$ scattering with a subsequent decay into one positron, one anti-muon, and the corresponding neutrinos. The resulting phase space is consistent with the fiducial $W^{\pm}W^{\pm}jj$-EW signal region in section 6.3.4 merged with the Low-m_{jj} control region definition. The calculations were done with MoCaNLO and validated by a second calculation using BBMC. Both tools are known to simulate intermediate boson polarization correctly and are based on the Recola library [93]. The provided differential cross-section dependent on the invariant di-jet mass and the transverse momentum of the leading jet are shown in figure 8.19. This thesis uses the NLO-EW correction in m_{jj} as the nominal correction. The difference to the alternative correction in p_T^{j1} is used as systematic uncertainty of the application method to account for the multivariate approach of the polarization analysis. The NLO-EW correction is applied according to the particle-level event kinematic of the polarized reconstruction-level prediction. Therefore, the application of the NLO-EW correction is consistent with the application strategy of the higher-order QCD correction.

To extract a smooth NLO-EW correction, an analytic fit is performed to the ratio

$$\mathrm{corr}_{\text{NLO-EW}}^{pol}(o) = \frac{\left(\frac{d\sigma_{\text{NLO-EW}}^{pol}}{do}\right)}{\left(\frac{d\sigma_{\text{LO}}^{pol}}{do}\right)} \tag{8.10}$$

Table 8.11: Parton-level selection for the NLO-EW calculation provided by the authors of Ref. [50].

Exactly one positron and one antimuon with $p_\mathrm{T} > 27$ GeV and $\lvert\eta\rvert < 2.5$
$m_{ll} > 20$ GeV
$E_\mathrm{T}^\mathrm{miss} \geq 30$ GeV
Two jets with $\lvert\eta\rvert < 4.5$ leading and subleading jets satisfying $p_\mathrm{T} > 65$ GeV and $p_\mathrm{T} > 35$ GeV, respectively
$m_{jj} \geq 200$ GeV
$\Delta y_{jj} > 2$
Separation between electron and jets has to fulfill $\Delta R_{ej} > 0.4$



(a) differential in m_{jj} **(b)** differential in p_T^{j1}

Figure 8.19: LO and NLO-EW differential cross-sections in m_{jj} and p_T^{j1} provided by the authors of Ref. [50]. The calculation is done in the phase space listed in table 8.11 for polarization in the $W^\pm W^\pm$-cmf (solid line) and partonic center-of-mass frame (dotted line).

for each of the polarization combinations *pol* and the two kinematic observables $o = m_{jj}$ and $o = p_T^{j1}$. For the fit in m_{jj}, the same analytic formula

$$f(m_{jj}) = p_0 + p_1 \ln \frac{m_{jj}}{\mathrm{GeV}} + p_2 \ln^2 \frac{m_{jj}}{\mathrm{GeV}} \tag{8.11}$$

as for the unpolarized NLO-EW correction in Ref. [8] is used. The NLO-EW correction in p_T^{j1} is well described by the linear function

$$f(p_T^{j1}) = p_0 + p_1 \frac{p_T^{j1}}{\mathrm{GeV}}. \tag{8.12}$$

The resulting fits are shown in figure 8.20 and the corresponding fit parameters are listed in table 8.12 for m_{jj} and table 8.13 for p_T^{j1}. The linear fit describes the shape in p_T^{j1} very well and also the corrections in m_{jj} are correctly modeled by the analytic function in equation (8.11). Only in the case of the correction for TL polarization in m_{jj}, the $\chi^2_\mathrm{fit}/\mathrm{dof}$ is not close to 1, but the shape of the correction in figures 8.20a and 8.20b is nevertheless very well covered.

The fit uncertainties are also plotted in figure 8.20 but barely visible. Due to the low number of fit parameters and the excellent agreement in the full range, the uncertainty on the fit is minimal and is thus not propagated to the analysis.

The availability of the shown NLO-EW corrections significantly impacts the polarization analysis. Before the publication of Ref. [50], the unpolarized NLO-EW correction from Ref. [8] would have been used. Since there was no knowledge of how higher-order electroweak effects would affect the different polarization states in $W^{\pm}W^{\pm}jj$-EW, the unpolarized correction would have been applied to each polarization with a 100% uncertainty. Thus, polarized NLO-EW corrections avoid one of the most dominant systematic uncertainties. In addition to a vanishing uncertainty, the polarized NLO-EW correction further affects the expected significance. The NLO-EW correction for LL is significantly smaller than the unpolarized NLO-EW correction, while the correction is larger for TT. This difference in the NLO-EW correction improves the ratio of longitudinal over transverse polarization and therefore the expected significance compared to a potential application of an unpolarized NLO-EW correction.

Since the polarized NLO-EW corrections were published during the finalization of this analysis, they are only applied for the final statistical evaluation in sections 12.5 and 12.6. They are not included in the training of the neural networks in section 9.1 and the bin optimization in section 12.3. For the training data the unpolarized NLO-EW correction from the $W^{\pm}W^{\pm}jj$-EW analysis in Ref. [8] with the fit parameters $p_0 = 0.511$, $p_1 = 0.176$, and $p_2 = -0.0181$ is used. This correction is also applied for the bin optimization. Since the bin optimization in section 12.3 optimizes the expected significance, a 100% uncertainty on δ_{EW} for each polarization was considered.

Table 8.12: Fit parameters of equation (8.11) for the NLO-EW correction in m_{jj}.

Polarization	χ^2_{fit}/dof	p-value	p_0	p_1	p_2
Unpolarized	1.6184	0.011	0.65605	0.12671	-0.013974
LL in WW-cmf	0.9559	0.54	0.73713	0.10904	-0.01226
TL in WW-cmf	2.2844	$1.8 \cdot 10^{-5}$	0.75425	0.10076	-0.012169
TT in WW-cmf	1.1005	0.31	0.69365	0.11283	-0.01303
LL in pp-cmf	0.9522	0.55	0.77496	0.096275	-0.01117
TL in pp-cmf	1.7142	$4.8 \cdot 10^{-3}$	0.91299	0.050347	$-8.1952 \cdot 10^{-3}$
TT in pp-cmf	1.0630	0.37	0.80717	0.081635	-0.010942

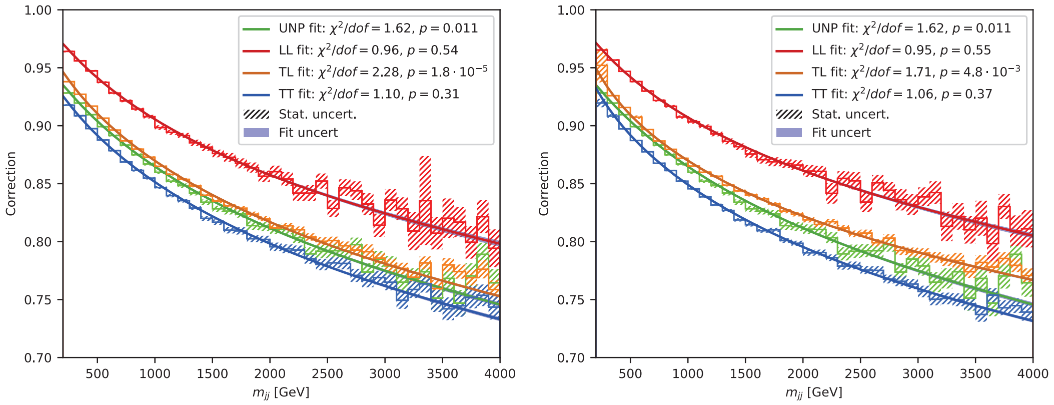
Table 8.13: Fit parameters of equation (8.12) for the NLO-EW correction in p_T^{j1}.

Polarization	χ^2_{fit}/dof	p-value	p_0	p_1
Unpolarized in WW-cmf	0.7122	0.98	0.94285	$-4.195 \cdot 10^{-4}$
LL in WW-cmf	0.8875	0.76	0.97074	$-4.063 \cdot 10^{-4}$
TL in WW-cmf	1.1325	0.19	0.95494	$-4.3927 \cdot 10^{-4}$
TT in WW-cmf	0.7859	0.93	0.92815	$-4.0401 \cdot 10^{-4}$
LL in pp-cmf	0.9387	0.64	0.96366	$-3.8534 \cdot 10^{-4}$
TL in pp-cmf	1.0252	0.42	0.95555	$-4.6191 \cdot 10^{-4}$
TT in pp-cmf	1.1160	0.22	0.92744	$-3.9677 \cdot 10^{-4}$

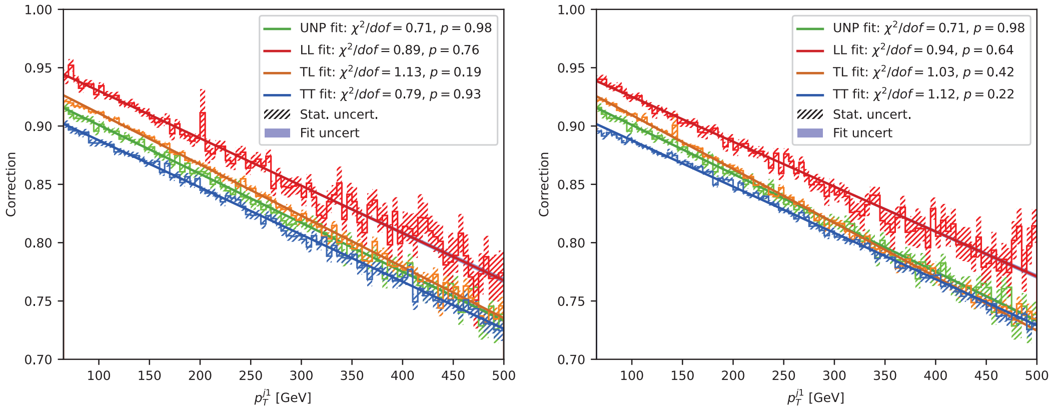
8.3 Combination of the NLO Corrections

To combine the corrections for higher-order QCD and EW effects, the approach of Ref. [92] is followed. The relative changes in the differential cross-section can be expressed as

$$\delta_{\text{EW}} = \frac{d\Delta\sigma_{\text{EW}}}{d\sigma_{\text{LO}}}, \qquad \text{and} \qquad \delta_{\text{QCD}} = \frac{d\Delta\sigma_{\text{QCD}}}{d\sigma_{\text{LO}}}. \tag{8.13}$$



(a) NLO-EW corr. in $W^{\pm}W^{\pm}$-cmf in terms of m_{jj} **(b)** NLO-EW corr. in pp-cmf in terms of m_{jj}



(c) NLO-EW corr. in $W^{\pm}W^{\pm}$-cmf in terms of p_T^{j1} **(d)** NLO-EW corr. in pp-cmf in terms of p_T^{j1}

Figure 8.20: The ratio of the NLO-EW and LO differential cross-sections in the phase space listed in table 8.11 provided by the authors of Ref. [50]. An analytic fit of equation (8.11) for m_{jj} and equation (8.12) for p_T^{j1} is performed for each of the resulting NLO-EW corrections in the $W^{\pm}W^{\pm}$-cmf and pp-cmf.

For the combination of these two relative corrections, additive and multiplicative approaches are possible:

$$d\sigma_{\mathrm{NLO}_+} = d\sigma_{\mathrm{LO}}(1 + \delta_{\mathrm{QCD}} + \delta_{\mathrm{EW}}) \tag{8.14}$$

$$d\sigma_{\mathrm{NLO}_\times} = d\sigma_{\mathrm{LO}}(1 + \delta_{\mathrm{QCD}})(1 + \delta_{\mathrm{EW}}) \tag{8.15}$$

At high lepton energies, the dominant EW corrections are expected to factorize with respect to the QCD corrections [92]. Thus, the multiplicative combination $d\sigma_{\mathrm{NLO}_\times}$ is the nominal correction. The difference between $d\sigma_{\mathrm{NLO}_\times}$ and $d\sigma_{\mathrm{NLO}_+}$ is used as an estimate for the difference to the mixed QCD-EW correction. It is symmetrized and used as the uncertainty of the combination method.

Chapter 9

Training of Discriminator DNNs

The signal region described in section 6.3.1 has been optimized to reach a sufficient purity in $W^{\pm}W^{\pm}jj$-EW events. To enable a more extensive evaluation of this signal with respect to the $W^{\pm}$ boson polarization additional analysis strategies must be used. Since the $W^{\pm}$ boson decays into a lepton and a neutrino and only the lepton can be reconstructed, much information regarding the original polarization is not accessible in the measurement. Thus, no single discriminant observable can separate the polarizations and multivariate analysis techniques must be used. For this, deep neural networks (DNN) implemented in Keras [94] extract sensitive variables for the polarization measurement from the entire event kinematic. The longitudinal $W^{\pm}W^{\pm}jj$-EW polarization is separated from the transverse polarization state and the other backgrounds by combining two networks. The $\text{DNN}_{W^{\pm}W^{\pm}}$ is trained to distinguish the $W^{\pm}W^{\pm}jj$-EW contribution from the Standard Model backgrounds. The DNN_{pol} enables the investigation of the polarization states. The polarization networks are trained to separate longitudinal single-boson polarization (LX) or double-boson polarization (LL) from the remaining transverse polarization states. Since the polarization is investigated in the partonic and $W^{\pm}W^{\pm}$ center-of-mass frame, four distinct DNN_{pol} are trained for the polarization analysis.

The motivation behind combining the output of two networks, $\text{DNN}_{W^{\pm}W^{\pm}}$ and DNN_{pol}, results from the strategy of the statistical analysis and is explained further in section 12.2.

9.1 Training Input

Since only supervised learning strategies are applied in this analysis, the training is done on the Standard Model predictions described in chapter 7. The MadGraph polarization samples are used for the polarization modeling. Since the main target of the analysis is the polarization measurement in the WW-cmf, 10 times more training events are generated in this reference frame to further improve the separation power of the trained networks.

When the training input was extracted, the particle-level kinematics were not available in the reconstruction-level polarization samples. Thus, a workaround is used to apply the higher-order corrections of chapter 8 that are derived for particle-level kinematics. For the higher-order QCD correction the one-dimensional higher-order QCD correction in m_{jj} is used which is extracted in section 8.1.7. This reweighting procedure is based on the particle-level invariant di-jet mass m_{jj} of the event. Since this particle-level information was not available in the training sample, the reconstruction-level m_{jj} kinematics is used. Since the higher-order correction does not affect the individual event kinematics and only corrects the event weight, this approximation is sufficient for the training data. For the higher-order electroweak correction of the training data the very recent theoretical calculations used in section 8.2 was not available. Therefore, the unpolarized NLO-EW correction of the differential $W^{\pm}W^{\pm}jj$-EW analysis in Ref. [8] is used. This NLO-EW correction on particle-level was provided by the authors of Ref. [90] for the phase space that is still used by the polarization analysis. The correction is described by the analytic function

$$f(m_{jj}) = 0.511 + 0.176 \ln \frac{m_{jj}}{\text{GeV}} - 0.0181 \ln^2 \frac{m_{jj}}{\text{GeV}}. \tag{9.1}$$

As for the higher-order QCD correction, this correction is applied to the reconstruction-level m_{jj} kinematics of the training data.

The data-driven approach in section 7.2.3 is used to create the training data for the charge-flip background. For the non-prompt background, the data-driven approach of section 7.2.2 uses a looser object selection and a significant subtraction of prompt events. The prompt subtraction would introduce a significant amount of negative event weights. The network will assign the events with negative weights to their opposite $W^{\pm}W^{\pm}jj$-EW category in order to reduce the loss. The measured data with positive weights, on the other hand, reduce the loss if they are assigned to the background category. In this way, the network does not consistently learn to sort non-prompt events to the background category, but learns the difference between the measured data and prompt subtraction of the data-driven background estimation. Therefore, the data-driven estimation of non-prompt background is not used for training. However, the non-prompt background could also be modelled by MadGraph+Pythia8 W+jets samples (DSID: 363600-363671) and a non-all-hadronic $t\bar{t}$ sample (DSID:410470) as described in the $W^{\pm}W^{\pm}jj$-EW analysis in Ref. [8]. Since this prediction would only contribute 170 simulated Monte Carlo events, which are also known to poorly model the non-prompt contribution, no major positive effect is expected from including them in the training. As such, the non-prompt background is not included in the DNN training.

Two Monte Carlo event-generated samples are available for the $W^{\pm}Z$-QCD background. As described in section 7.2.1, the $W^{\pm}Z$-QCD Sherpa2.2.2 sample used in the unpolarized $W^{\pm}W^{\pm}jj$-EW analysis [8] is replaced by a $W^{\pm}Z$-QCD Sherpa2.2.12 sample for the statistical analysis. Since the $W^{\pm}Z$-QCD is the dominant background in the signal region, both samples are used to increase the amount of training information available for this background. The respective event weights must be reduced to avoid doubling the predicted $W^{\pm}Z$-QCD contribution in the training data. Updating the event weights can be directly combined with an overall correction of the expected $W^{\pm}Z$-QCD contribution known from Ref. [8] to be overestimated. From the fit in the $W^{\pm}Z$ control region in Ref. [8] one expects 543 $W^{\pm}Z$-QCD events in the $W^{\pm}Z$ control region. Both, WZQCD_Sherpa222 and WZQCD_Sherpa2212, are scaled to have the same number of events in the signal region and their sum meets this expectation in the $W^{\pm}Z$ control region.

The $V\gamma$ samples are not used for the training due to the poor modeling shown in the $W^{\pm}W^{\pm}jj$-EW analysis [8]. Training on a sample with very poor modeling while using it to predict the distribution of unseen data could introduce an additional bias between the classification of measured data and the Standard Model prediction. A complete list of all processes included in the DNN training as well as the number of Monte Carlo events used to model them and the respective predicted number of events is given in table 9.1.

The total event weights are considered in the loss calculation to meet the expected ratios while training. As usual for binary classification, the weights within the two categories are normalized to the same sum to enable an even separation. Table 9.1 shows that the event weights of different processes can differ by several orders of magnitude. To avoid significant fluctuations between the training batches all events with a weight ten times higher than the medium weight are duplicated and their event weight is reduced by the number of duplications. This procedure was established in Ref. [89] to distribute events with high weights over multiple training batches instead of dominating one batch.

The DNN hyperparameters are optimized while initially using all variables listed in table 9.2. Some variables are pre-scaled for the training and application to get a more Gaussian-distributed input for the neural networks. This scaling improves the training process for the DNN algorithms. The distributions of the training input variables are shown in the appendix in figures C.1 to C.8. The angle ϕ is redefined with respect to the direction of the leading lepton to account for the rotational symmetry of the detector and the leading lepton ϕ is dropped. Thus, the DNN becomes independent of the exact event orientation in ϕ to reduce redundancies in the training input.

The training input can be divided into low-level and high-level variables. Low-level variables represent the kinematics of leptons or jets, and are measured directly by the ATLAS detector. From these low-level variables, more complex high-level variables can be derived. Generally, a neural network can reconstruct these high-level variables from the low-level variables by itself. However, the high-level variables offer an opportunity to introduce prior

Table 9.1: Processes used for training of the DNN classifiers with number of generated events and the number of events they predict for the signal region. For $W^\pm Z$-QCD, an additional scale factor was applied to achieve a better agreement with the expected data.

Process	Generated Events	Predicted Events
$W^\pm W^\pm jj$-EW LL WW-cmf	215008	18.29
$W^\pm W^\pm jj$-EW TL WW-cmf	170387	58.88
$W^\pm W^\pm jj$-EW TT WW-cmf	456456	124.50
$W^\pm W^\pm jj$-EW LL pp-cmf	19673	11.49
$W^\pm W^\pm jj$-EW TL pp-cmf	21471	67.84
$W^\pm W^\pm jj$-EW TT pp-cmf	46320	123.07
$W^\pm W^\pm jj$-EW MGH7	201488	206.52
$W^\pm W^\pm jj$-QCD	25375	24.05
$W^\pm W^\pm jj$-Int	79805	7.57
$W^\pm Z$-EW	8273	14.95
$W^\pm Z$-QCD Sherpa222	27800	$82.75/2 = 41.38 \rightarrow 28.50$
$W^\pm Z$-QCD Sherpa2212	22199	$76.79/2 = 38.40 \rightarrow 28.50$
top	2192	5.02
ZZ	2880	2.51
charge-flip	8231	10.10

physics knowledge about event kinematics into the training. Thus, the angular separations in Δy, $\Delta\phi$, and ΔR between different objects are provided. The η values of leptons and jets can be used to define the lepton Zeppenfeld variable

$$Z_l^\star = \left| \frac{\eta_l - \frac{1}{2}(\eta_{j1} - \eta_{j2})}{\eta_{j1} - \eta_{j2}} \right|. \tag{9.2}$$

For the di-lepton and di-jet systems, the invariant masses are calculated. In addition to the invariant masses, transverse mass projections are also taken into account. The definition of the massless early-projected transverse mass,

$$m_{o1}^{ll,MET} = \sqrt{(p_T^{l1} + p_T^{l2} + p_T^{miss})^2 - (\vec{p}_T^{\,l1} + \vec{p}_T^{\,l2} + \vec{p}_T^{\,miss})^2}, \tag{9.3}$$

and the mass-preserving transverse mass,

$$m_T^{ll,MET} = \sqrt{(E_T^{ll} + p_T^{miss})^2 - (\vec{p}_T^{\,l1} + \vec{p}_T^{\,l2} + \vec{p}_T^{\,miss})^2}, \tag{9.4}$$

with $E_T^{ll} = \sqrt{(m_{ll})^2 + \left(p_T^{ll}\right)^2}$ are described in Ref. [95]. The transverse mass is also calculated for the individual leptons, $m_T^{l1,MET}$ and $m_T^{l2,MET}$.

In order to improve the sensitivity to the polarization, Ref. [84] aims to reconstruct the decay angles of the two $W^\pm$ bosons using regression models. Two different training setups are studied for the regression models, and both are investigated in the following. The regression models are trained on the leptons, jets, and missing transverse momentum kinematics. Since the regression models are trained to regress the true $W^\pm$ decay angles, they can potentially contain information not directly available in the reconstruction-level phase space kinematic. Thus, the regression model output is considered as a potential training input for the classifiers. A more detailed discussion of the regression networks is omitted for this thesis, as they are removed from the training data by the following optimization.

The training data is split into train, validation, and test sets according to the event number in a ratio of 3:1:1. The test set is given by $((\text{EventNumber} - i_{\text{Fold}}) \mod 5) = 0$, events with $((\text{EventNumber} - i_{\text{Fold}} + 1) \mod 5) = 0$ belong to the validation set and the remaining events are used for training. The index i_{Fold} defines the different splits used for cross-validation in the hyperparameter optimization and for the 5-Fold method described in section 5.8.

Table 9.2: Variables used to optimize and train the binary classifier. The variables were pre-scaled for the training and application to have a more Gaussian-like distribution.

Kinematics	Descriptions	scaling
Low-level variables		
p_T^{l1}	p_T of the leading lepton	$\log_{10}(x)$
η^{l1}	η of the leading lepton	x
$l1$ type	flavor of the leading lepton	x
p_T^{l2}	p_T of the subleading lepton	$\log_{10}(x)$
η^{l2}	η of the subleading lepton	x
$\phi^{l2} - \phi^{l1}$	Redefined ϕ of the subleading lepton	x
$l2$ type	flavor of the subleading lepton	x
p_T^{j1}	p_T of the leading jet	$\log_{10}(x)$
η^{j1}	η of the leading jet	x
$\phi^{j1} - \phi^{l1}$	Redefined ϕ of the leading jet	x
p_T^{j2}	p_T of the subleading jet	$\log_{10}(x)$
η^{j2}	η of the subleading jet	x
$\phi^{j2} - \phi^{l1}$	Redefined ϕ of the subleading jet	x
p_T^{miss}	Missing transverse momentum	$\log_{10}(x)$
$\phi(p_T^{miss}) - \phi^{l1}$	Redefined ϕ of the missing transverse energy	x
High-level variables		
$Z_{l1}^{\star}$	Zeppenfeld variable of the leading lepton	$\sqrt{x}$
$Z_{l2}^{\star}$	Zeppenfeld variable of the subleading lepton	$\sqrt{x}$
$m_T^{l1,MET}$	Transverse mass of the leading lepton and p_T^{miss}	$\sqrt{x}$
$m_T^{l2,MET}$	Transverse mass of the subleading lepton and p_T^{miss}	$\sqrt{x}$
ΔR_{ll}	ΔR between the two leading leptons	x
$\Delta\eta_{ll}$	$\Delta\eta$ between the two leading leptons	$\sqrt{x}$
m_{ll}	Invariant mass of the two leading leptons	$\log_{10}(x)$
p_T^{ll}	p_T of the dilepton system	$\sqrt{x}$
$m_T^{ll,MET}$	Transverse mass of the dilepton system	$\sqrt{x}$
$m_{o1}^{ll,MET}$	Early-projected massless invariant mass of dilepton system and missing transverse energy	$\sqrt{x}$
ΔR_{jj}	ΔR between the two leading jets	x
Δy_{jj}	Δy between the two leading jets	x
m_{jj}	Invariant mass of the two leading jets	$\log_{10}(x)$
$\Delta\phi_{jj}$	$\Delta\phi$ between the two leading jets	x
$(p_T^{l1} \cdot p_T^{l2})/(p_T^{j1} \cdot p_T^{j2})$	p_T ratio of leptons and jets	$\log_{10}(x + 0.02)$
$min(\Delta R_{l1/l2,j1/j2})$	Minimal ΔR between the leptons and jets	x
Regression models		
$\text{Regression}_{\text{LL},1}(l1)$	Regressed $W^{\pm} \to \nu l1$ decay angle trained on LL events	x
$\text{Regression}_{\text{LL},1}(l2)$	Regressed $W^{\pm} \to \nu l2$ decay angle trained on LL events	x
$\text{Regression}_{\text{LL},2}(l1)$	Regressed $W^{\pm} \to \nu l1$ decay angle trained on LL events	x
$\text{Regression}_{\text{LL},2}(l2)$	Regressed $W^{\pm} \to \nu l2$ decay angle trained on LL events	x
$\text{Regression}_{\text{LT},1}(l1)$	Regressed $W^{\pm} \to \nu l1$ decay angle trained on LT events	x
$\text{Regression}_{\text{LT},1}(l2)$	Regressed $W^{\pm} \to \nu l2$ decay angle trained on LT events	x
$\text{Regression}_{\text{LT},2}(l1)$	Regressed $W^{\pm} \to \nu l1$ decay angle trained on LT events	x
$\text{Regression}_{\text{LT},2}(l2)$	Regressed $W^{\pm} \to \nu l2$ decay angle trained on LT events	x
$\text{Regression}_{\text{TT},1}(l1)$	Regressed $W^{\pm} \to \nu l1$ decay angle trained on TT events	x
$\text{Regression}_{\text{TT},1}(l2)$	Regressed $W^{\pm} \to \nu l2$ decay angle trained on TT events	x
$\text{Regression}_{\text{TT},2}(l1)$	Regressed $W^{\pm} \to \nu l1$ decay angle trained on TT events	x
$\text{Regression}_{\text{TT},2}(l2)$	Regressed $W^{\pm} \to \nu l2$ decay angle trained on TT events	x

9.2 Network Architecture

This thesis focuses on fully-connected feed-forward neural networks. The first layer of the networks is the normalization layer. This layer applies the scaling function corresponding to the individual variables in table 9.2 and normalizes the mean of the resulting distribution to 0 and the variance to 1. This normalization layer modifies the input variables to have similar distributions without losing information about the kinematic phase space.

The normalization layer is followed by a dense layer connected to a batch normalization layer. After the output of the dense layer is normalized, the activation function is applied. A dropout layer is connected after the activation function for further regularization during the training. This sequence of dense, batch normalization, activation, and dropout layers is repeated until the final dense layer, which provides the output of the network. More information about the functionality of a neural network and the individual layers is provided in chapter 5.

9.3 DNN Optimization Algorithm

Although focusing only on fully-connected neural networks, there is a wide variety of possible network hyperparameters and training procedures. In order to simplify the optimization, only networks with a constant number of neurons per layer are used, which means that only the number of layers and neurons per layer have to be optimized for the network size. Also, the activation function is not additionally optimized in the following but is set to the swish function [64] for all layers. Following the recommendation for the swish function in Ref. [96], the weights are initialized by the `truncated_normal` distribution with a variance of 2.952 normalized by the number of input neurons of the corresponding weight tensor, and a normal distribution with a standard deviation of 0.2 initializes the bias. The Adam optimizer performs the weight updates, using binary cross entropy for the loss function. Besides the neural network's size, a wide variety of training parameters still have to be optimized to gain a well-performing classifier. The accuracy of the training process can be optimized by varying the batch size. To reduce overtraining, dropout in each layer is introduced. For the Adam optimizer, the learning rate, β_1, β_2, and ϵ parameters are optimized. The optimization of the set of training variables and hyperparameters described in this chapter is implemented in the OPTIMA tool [97] and was developed in Ref. [89]. The metric for the optimization is the area under the ROC curve (AUC) since this metric is more directly related to the actual separation of the different processes than the binary cross entropy. The advantages of the AUC are described in more detail in section 5.7.

9.3.1 Optimization with OPTUNA

Due to the variety of hyperparameters to be optimized and the resulting very high number of possible combinations, a grid search for optimization is not feasible. In order to run this complex optimization, the open-source optimization framework OPTUNA [98] is used. This tool can be used to optimize complex problems with an efficient sampling strategy. For the sampling, the TPESampler [99] is used to find and test suitable hyperparameters. Using Ray Tune [100], several neural networks can be trained simultaneously and underperforming trials can be pruned by the ASHA-Scheduler [101] if their validation AUC is not improving fast enough. This pruning eliminates poorly performing networks at an early stage to save resources in the optimization. However, since in general small networks converge faster than larger networks, pruning could prevent the choice of larger networks. Thus, a minimum number of layers is required in the optimization. The hyperparameter search space spans 6 to 14 layers for the optimization of the neural networks trained to separate polarization states in the WW-cmf as well as the $W^{\pm}W^{\pm}jj$-EW vs background models, while 3 to 8 layers were chosen for the polarization discrimination in the pp-cmf due to the smaller MC sample. The full search space and sampling method is listed in table 9.3. When selecting new hyperparameter combinations during the optimization, the performance of previous trials is evaluated based on their most recent instead of their best epoch. This reduces the impact of random fluctuations of the validation AUC. A similar problem arises when choosing the best combination of hyperparameters at the end of the optimization. A fit to the validation AUC is performed for each of the training runs using the fit function

$$f_{loss}(x_{epoch}) = \frac{A}{(x_{epoch})^2} + \frac{B}{x_{epoch}} + Cx_{epoch} + D(x_{epoch})^2 + const. \tag{9.5}$$

For all performed trainings during the optimization, the fit function is applied in the last two-thirds of the training process or in the last 20 epochs, whichever provides more points for fitting. Trainings with less than 20 epochs are not considered.

The result of the optimization is the hyperparameter set of the training with the highest AUC in the fit. The location of the maximum determines the optimal number of epochs to train.

Table 9.3: DNN hyperparameters optimized for the discriminator DNNs. The search space and sampling distribution are listed for the tunable hyperparameters.

Fixed hyperparameters		
Hyperparameter	Value	
Optimizer	Adam	
Loss function	binary cross-entropy	
Activation function	swish	
Optimized hyperparameters		
Hyperparameter	Seach space	Parameter sampling
Number of layers	$[3, 8]$ for DNN$_{pol}$ in pp-cmf $[6, 14]$ for DNN$_{pol}$ in WW-cmf and DNN$_{W^{\pm}W^{\pm}}$	Uniform
Neurons per layer	$[32, 512]$	Uniform
Dropout rate	$[0.0, 0.5]$	Uniform
Batch size	$[32, 512]$	Logarithmic
Learning rate	$[10^{-5}, 10^{-2}]$	Logarithmic
Adam β_1	$[10^{-4}, 0.99]$	Logarithmic
Adam $1 - \beta_2$	$[10^{-5}, 0.9999]$	Logarithmic
Adam ϵ	$[10^{-10}, 1.0]$	Logarithmic

9.3.2 Backward Elimination of Training Variables

The method described so far allows the optimization of the hyperparameters of the neural network in a very efficient way. However, whether all input variables used were necessary or are dispensable for the training has not been tested so far. Reducing the set of training variables can reduce the occurrence of local minima in the training loss and, therefore, improve the training process [102]. Before dropping training variables, an initial hyperparameter optimization is performed on the complete set of variables to ensure well-performing hyperparameters. Afterward, each variable is tested for its importance in the training process. A combination of two different approaches is used for this:

Approach 1 Shuffle the values of the variables. The shuffling is done to assign random values for this variable while preserving the initial variable distribution.

Approach 2 Remove the variable from the dataset using the hyperparameters and number of training epochs as given by the best fit of the initial OPTUNA optimization.

Both approaches calculate the model predictions and determine the mean validation AUC on the five folds before and after the variable var$_i$ is dropped. When using approach 1, this procedure is repeated 10 times to reduce the statistical uncertainty. The resulting metric used to quantify the importance of a variable is the mean relative change of the validation AUC with respect to the training on the full variable set:

$$\delta_i = \frac{1}{5} \sum_{j=0}^{4} \frac{\mathrm{AUC}_j(\text{var-set} \setminus \text{var}_i)}{\mathrm{AUC}_j(\text{var-set})} - 1 \tag{9.6}$$

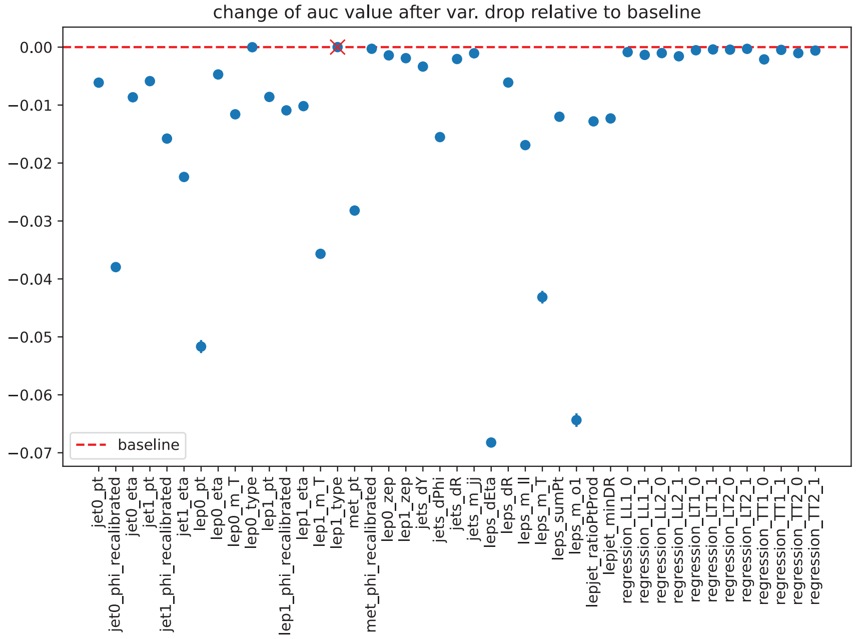


Figure 9.1: Optimization of input variables for the DNN trained to separate LL and TX in WW-cmf with shuffling. The mean relative AUC change is plotted after shuffling a variable of the variable set in table 9.2 10 times on each fold. The variable with the lowest performance decrease is dropped, and in the next step, the remaining variables are tested. This figure represents the first iteration of the variable optimization for DNN$_{pol}$ trained for LL in WW-cmf.

This first step of the variable optimization is shown in figure 9.1 for the DNN$_{pol}$ separating LL and TX in the WW-cmf. The relative change δ_i drops below 0 for several variables, but the removal of a significant number of variable results in a relative change close to 0. The variable with the highest δ_i is dropped, leading to the best performance improvement (lowest performance decrease).

After the variable is dropped, the model is retrained on the reduced variable set. The retraining is done with three different seeds for each of the five folds, which leads to a precise evaluation of the performance change in each fold after dropping the variable.

At the start of the variable optimization, approach 1 (shuffling of variables) is used to determine variable importance because it is computationally cheap and can identify variables that have no significant impact on the final decision. This approach is followed until no improvement by dropping a variable has been achieved for five consecutive iterations. Afterwards, the best-performing set of variables is chosen, and the procedure is continued with approach 2 (retraining). The retraining has significantly higher computational costs than the shuffling but can drop discriminant but highly correlated variables. This procedure is continued until the evaluated performance change no longer overlaps with the best-performing set within their standard deviations. For this stopping criterion, a patience of five iterations is used.

9.3.3 Combined Optimization Strategy

The final optimization procedure is done in three steps. First, a hyperparameter optimization is performed with 1000 networks trained on fold 0 to optimize the validation AUC of the neural network. The resulting set of hyperparameters is used to perform the variable optimization, reducing the set of training inputs. The variable optimization is followed by a final hyperparameter optimization with 3000 trials trained on fold 0 to find the best parameters for the reduced set of variables. However, due to the high number of trials, this validation performance can also be a coincidence. Therefore, the selected network is retrained on the five folds with a different seed. Due to the retraining the final networks become independent of the explicit selection during the optimization.

9.3.4 Optimization Results

The set of hyperparameters of the initial hyperparameter optimization is given in table 9.4. Due to the smaller training sets, the networks trained to separate the polarization in the pp-cmf tend to be smaller and thus have fewer trainable parameters. The shown sets of hyperparameters are then used for the variable optimization. The variable optimization reduces the set of variables to the sets given in table 9.6. The results of the subsequent hyperparameter optimization are listed in table 9.5. The set of chosen training variables and hyperparameters are discussed in the following.

Table 9.4: Hyperparameters derived from the pre-optimization. The pre-optimization was done on the full set of variables in table 9.2.

Hyperparameter	$W^{\pm}W^{\pm}jj$-EW vs bkg.	LL vs TX in WW-cmf	LX vs TT in WW-cmf	LL vs TX in pp-cmf	LX vs TT in pp-cmf
Number of layers	9	10	10	7	3
Neurons per layer	192	123	204	139	426
Dropout rate	0.353334	0.204865	0.388665	0.427303	0.448612
Activation function	swish	swish	swish	swish	swish
Batch size	333	283	399	226	337
Learning rate	0.00134618	0.000885491	0.00100069	0.00245006	0.00110225
Adam β_1	0.00191544	0.565954	0.00437381	0.620739	0.896171
Adam $1-\beta_2$	0.0274911	0.00304935	0.00481873	0.00564895	0.14664
Adam ϵ	1.50036e-08	1.87817e-10	1.60614e-05	6.31996e-10	5.18648e-09
epochs	54	52	106	46	28

Table 9.5: Hyperparameters chosen for the neural network training. The shown hyperparameters were chosen in the final step of the hyperparameter optimization.

Hyperparameter	$W^{\pm}W^{\pm}jj$-EW vs bkg.	LL vs TX in WW-cmf	LX vs TT in WW-cmf	LL vs TX in pp-cmf	LX vs TT in pp-cmf
Number of layers	8	9	7	3	4
Neurons per layer	491	296	322	211	96
Dropout rate	0.355903	0.381438	0.450564	0.472462	0.2293
Activation function	swish	swish	swish	swish	swish
Batch size	288	436	342	181	417
Learning rate	0.0021943	0.00158784	0.000330002	0.00143182	0.00136966
Adam $beta_1$	0.0710309	0.729623	0.874392	0.67804	0.410719
Adam $1-\beta_2$	0.0012125	0.00227713	0.000171442	0.00300309	0.261055
Adam ϵ	6.72079e-08	2.42132e-09	1.85018e-06	1.0855e-09	1.44148e-06
epochs	24	73	162	86	62

Training Variables

For all DNNs, a significant number of variables are dropped during the optimization. The relative change of the AUC on validation data is shown in figure 9.2 for the DNN$_{pol}$ trained for LL in WW-cmf. The optimization plots for the remaining networks are shown in figures C.9 and C.10. Since the network performances increase, only variables harmful to the training are dropped. The validation AUC increases until the first variables that provide information necessary for the classification are dropped. The resulting set of variables corresponds to the last iteration before the rapid decrease of the network performance starts. Thus, the chosen variables are expected to provide the smallest set of training variables that contain all the information required for the network training.

The resulting variable sets are summarized in table 9.6. The corresponding numbers of chosen low- and high-level variables are listed in table 9.7. The polarization DNNs trained in the WW-cmf tend more to use low-level variables than the polarization DNNs in the pp-cmf. This difference can be explained by the bigger network size in table 9.4 and the ten times

Table 9.6: List of variables used for the training. For each neural network, the variables kept after the variable optimization for the training are marked with X. The index 1 or 2 for the regression variables in the lower part of the table relates to two different hyperparameter choices for the regression DNN models.

Variable	ssWW vs bkg	LL vs TX in WW-cmf	LX vs TT in WW-cmf	LL vs TXf in pp-cmf	LX vs TT in pp-cmf
Low-level variables					
p_T^{l1}	X	X	X	X	X
η^{l1}	X	X	X	X	
$l1$ type	X				
p_T^{l2}		X	X	X	X
η^{l2}	X	X	X		
$\phi^{l2} - \phi^{l1}$		X	X		
$l2$ type	X				
p_T^{j1}	X	X	X	X	
η^{j1}		X	X	X	
$\phi^{j1} - \phi^{l1}$		X	X	X	
p_T^{j2}	X			X	X
η^{j2}	X	X	X		
$\phi^{j2} - \phi^{l1}$		X	X	X	
p_T^{miss}	X	X	X	X	X
$\phi(p_T^{miss}) - \phi^{l1}$					
High-level variables					
$Z_{l1}^{\star}$	X			X	X
$Z_{l2}^{\star}$	X			X	X
m_T^{l1}	X	X	X	X	
m_T^{l2}	X	X	X	X	X
$\Delta R_{l1,l2}$		X		X	X
$\Delta\eta_{l1,l2}$		X	X	X	X
$m_{l1,l2}$					
$p_T^{l1,l2}$	X				
$m_T^{l1,l2}$					
$m_{o1}^{l1,l2,MET}$	X	X	X		X
$\Delta R_{j1,j2}$					
$\Delta y_{j1,j2}$	X	X			X
$\Delta\phi_{j1,j2}$	X	X	X	X	X
$m_{j1,j2}$				X	
$(p_T^{l1} * p_T^{l2})/(p_T^{j1} * p_T^{j2})$	X	X	X		
$min(\Delta R_{l1/l2,j1/j2})$	X	X	X	X	X
Regression models					
$\text{Regression}_{\text{LL},1}(l1)$					
$\text{Regression}_{\text{LL},1}(l2)$					
$\text{Regression}_{\text{LL},2}(l1)$					
$\text{Regression}_{\text{LL},2}(l2)$					
$\text{Regression}_{\text{LT},1}(l1)$					
$\text{Regression}_{\text{LT},1}(l2)$					
$\text{Regression}_{\text{LT},2}(l1)$					
$\text{Regression}_{\text{LT},2}(l2)$					
$\text{Regression}_{\text{TT},1}(l1)$					
$\text{Regression}_{\text{TT},1}(l2)$					
$\text{Regression}_{\text{TT},2}(l1)$					
$\text{Regression}_{\text{TT},2}(l2)$					

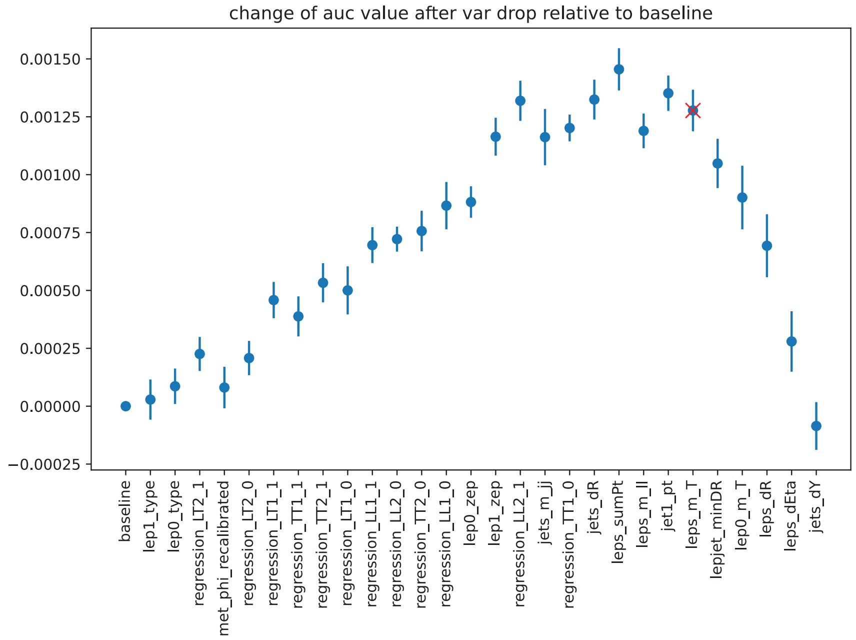


Figure 9.2: Optimization of input variables for the DNN$_{pol}$ trained for LL in WW-cmf. Plotted is the relative AUC change of the variable whose removal results in the best mean AUC change after shuffling/retraining. If a significant performance decrease has been seen for five iterations, the procedure is stopped and the variable set that is still consistent with the best-performing set is chosen (marked with x).

Table 9.7: The number of low- and high-level variables chosen for the discriminator networks. The detailed list of chosen variables is given in table 9.6.

DNN	Number of low-level variables	Number of high-level variables	Total Number of variables
DNN$_{W^\pm W^\pm}$	9	10	19
DNN$_{pol}$ for LL in WW-cmf	11	9	20
DNN$_{pol}$ for LX in WW-cmf	11	7	18
DNN$_{pol}$ for LL in pp-cmf	9	9	18
DNN$_{pol}$ for LX in pp-cmf	4	9	13

higher number of training events. The networks in the WW-cmf have a higher capacity and more data to learn the characteristics of the polarization states from the kinematics of individual objects. Since the networks in the pp-cmf have a significantly lower capacity and less training data, they benefit more from the physics information stored in high-level variables.

The two polarization networks DNN$_{pol}$ classifying the polarization in the WW-cmf have almost the same training variables. This agreement in the variable set is reasonable since the networks only differ by the classification of the mixed state LT. Similar kinematic variables are expected to be discriminant for the polarization classification for LL vs. TX and LX vs. TT. Thus, the agreement between the two training variable sets in the WW-cmf further validates the procedure for variable optimization. For the polarization networks in the pp-cmf, the sets of chosen variables mostly overlap but are not as consistent as those in the WW-cmf. The more prominent differences in the pp-cmf result from the choice of hyperparameters in table 9.4. The number of layers and nodes per layer differ significantly between the LL and LX classifiers in the pp-cmf. Due to these differences in the network size, different sets of training variables are preferred.

The lepton flavor is the first variable dropped during optimization for all the polarization networks. Since the lepton flavor originating from a $W^\pm$ boson decay does not depend on the $W^\pm$ polarization this information has no impact on the polarization DNNs. In addition, the output of the regression networks is dropped for all polarization networks. They don't provide an additional benefit to the existing low- and high-level variables and dropping them improves the network performance.

For the $\mathrm{DNN}_{W^\pm W^\pm}$, the invariant di-jet mass m_{jj} is dropped. Dropping this variable is surprising since the m_{jj} kinematic is one of the most discriminant variables to separate $W^\pm W^\pm jj$-EW from backgrounds. However, the p_T of the individual jets is kept together with their angular separation in Δy_{jj} and $\Delta\phi_{jj}$. Thus, the network can reconstruct the di-jet kinematic internally, making the external calculation of the mass superfluous.

The set of input variables of the networks represents the physics information it has learned. A deeper understanding of the learned information is derived in section 10.3 by studying the impact of the input values on the network decision.

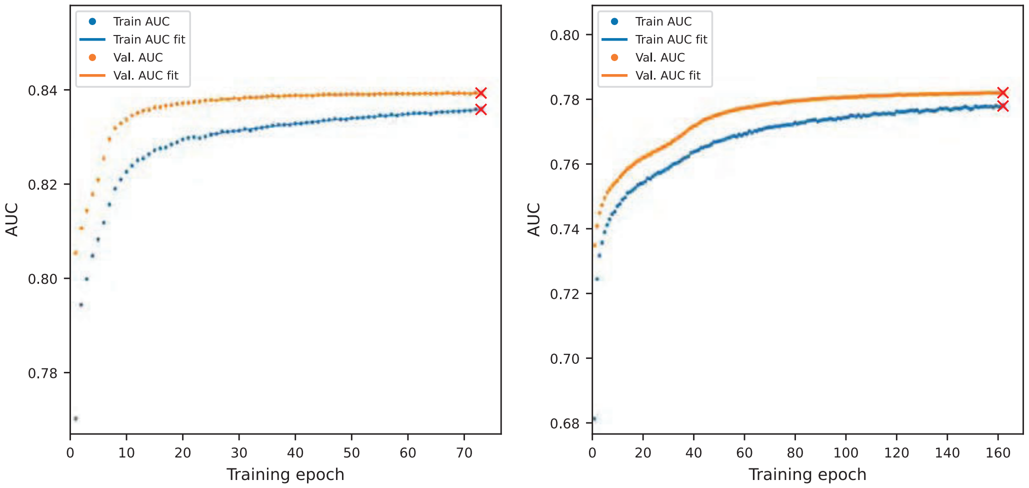
Optimization results for DNN_{pol}

The hyperparameter optimization with OPTUNA is repeated with 3000 trials for the resulting reduced set of training variables. The best-performing networks are selected by the maximum in the fit of equation (9.5) to the validation AUC. The training processes of the best-performing polarization networks are shown in figure 9.3 and the corresponding hyperparameters are listed in table 9.5. As expected for the lower number of training events for the polarization in the pp-cmf, the corresponding neural networks favor a lower number of layers, while for the polarization in the WW-cmf deeper networks perform better. In comparison to the initial choice of hyperparameters in table 9.4 before the variable optimization, the number of layers is reduced. Only for the LX classifier in pp-cmf, one additional layer is added but the number of nodes per layer is reduced by more than a factor of four. Although this has reduced the capacity of the networks, the number of training epochs has increased significantly. Thus, reducing the set of training variables leads to stabilized training convergence allowing more training epochs.

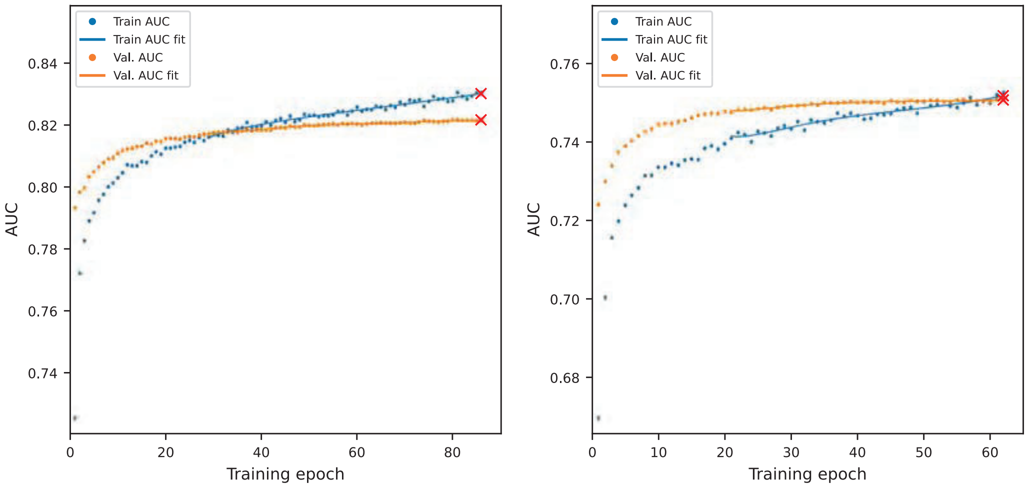
The output distributions of the resulting networks are shown in figure 9.4. The classification of validation data agrees well with the distribution seen for training data. The only significant differences can be seen for longitudinal events classified close to 0. Thus, more longitudinal events are classified as transversal as expected from the application on training data. However, the number of longitudinally polarized events classified as transversal is very small and the measurement of the longitudinal polarization is dominated by the $\mathrm{DNN}_{pol} > 0.5$ region. It is essential to highlight that even a more prominent difference between the distributions of validation and training data would only affect the classification performance but would not introduce a bias. Since the analysis uses the k-fold method described in section 5.8 the networks are only applied to the test dataset to predict the measured data.

For the LX in WW-cmf classification shown in figure 9.4b, an additional bump is seen, which does not occur for the LL classification in figure 9.4a. This bump corresponds to the mixed polarization state TL. The TL polarization state is kinematically closer to TT than the LL polarization. Therefore, the TL contribution peaks at $\mathrm{DNN}_{pol} \approx 0.6$ and not like the LL contribution at $\mathrm{DNN}_{pol} \approx 0.9$. The comparison of figure 9.4a and figure 9.4b indicates that the separation of LL and TX is simpler than for LX vs. TT. The simpler separation of LL and TX also explains why no TL bump is seen in figure 9.4a.

In the partonic center-of-mass frame in figures 9.4c and 9.4d, the same differences between the LL and LX classification occur. In comparison to figures 9.4a and 9.4b, the overall separation in the pp-cmf is not as good as in the WW-cmf. The lower number of training events and network capacity in the pp-cmf can partly explain the poorer separation. However, whether this can be resolved entirely with more data and larger networks is unknown. Moving from a polarization definition in the WW-cmf to the pp-cmf could reduce the kinematic differences between the polarization states. Since the analysis mainly aims for the polarization measurement in the WW-cmf the exact origin of the poorer separation in the pp-cmf is not further investigated.

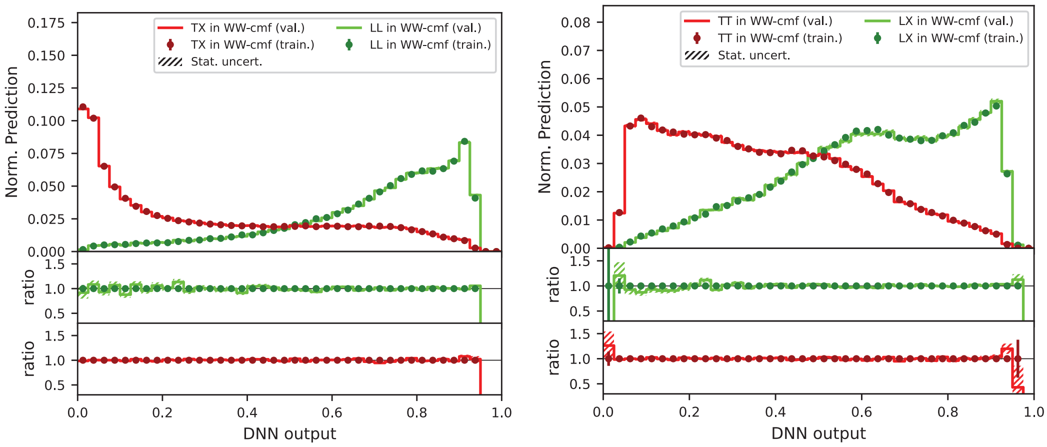


(a) DNN for LL in WW-cmf **(b)** DNN for LX in WW-cmf



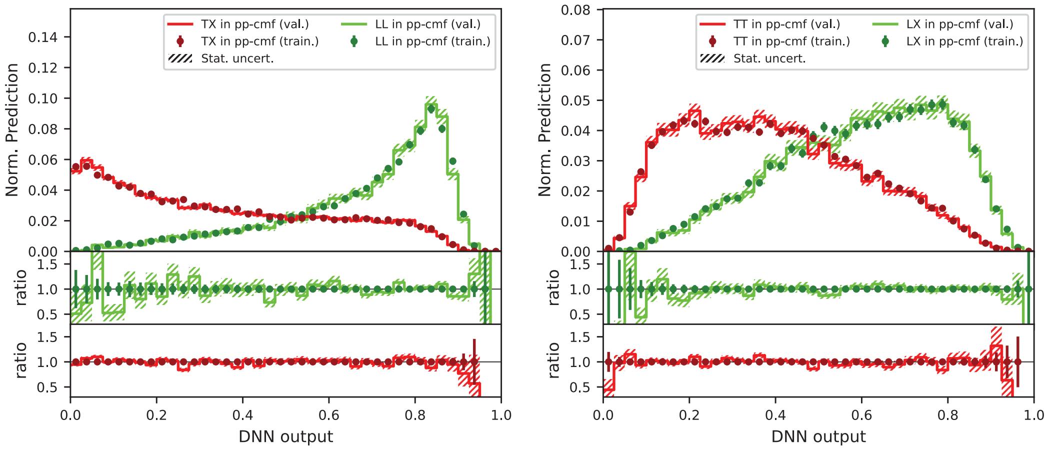
(c) DNN for LL in pp-cmf **(d)** DNN for LX in pp-cmf

Figure 9.3: The AUC during the training process of the polarization discriminator networks selected by the final hyperparameter optimization. The epoch with the highest AUC in the validation AUC fit is marked with a red cross.



(a) DNN for LL in WW-cmf

(b) DNN for LX in WW-cmf



(c) DNN for LL in pp-cmf

(d) DNN for LX in pp-cmf

Figure 9.4: The output distribution of the polarization DNNs trained on fold 0 and applied on the corresponding training and validation data.

Optimization results for $\mathbf{DNN}_{W^\pm W^\pm}$

Throughout the analysis, all networks were optimized several times whenever newer updates were available for the training data. However, during the previous optimization of the network $\text{DNN}_{W^\pm W^\pm}$, an error occurred that was detected too late and is therefore discussed in this section.

Like the polarization networks, the network separating $W^\pm W^\pm jj$-EW and background is also optimized for 3000 trials on the reduced variable set. The optimization algorithm selects the best network according to its maximum in the fit of equation (9.5) to the validation AUC. The fit is applied to avoid the selection of outliers. However, in this specific case, the fit to the validation AUC has failed as can be seen in figure 9.5. The fit function does not agree with the validation AUC and the fit maximum is above the validation AUC values. In addition, the training epoch selected from the optimization algorithm does not correspond to the best training epoch of the network leading to a network that is not trained to its optimum. Unfortunately, this error was not noticed until the network was implemented in the analysis and applied to the full Standard Model prediction, including all systematic uncertainties. Therefore, an update of this network must be well justified, as it is associated with a significant computational effort and leads to a delay in the analysis timeline.

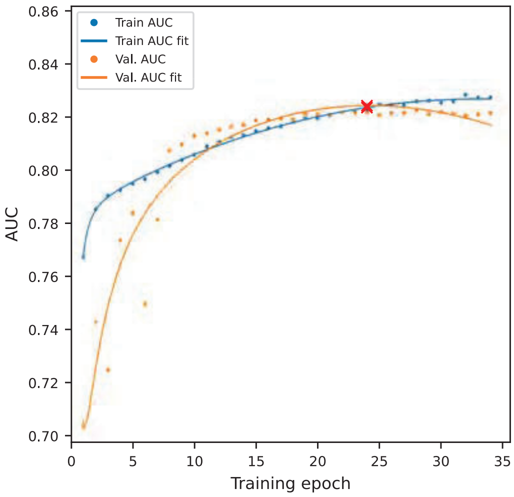
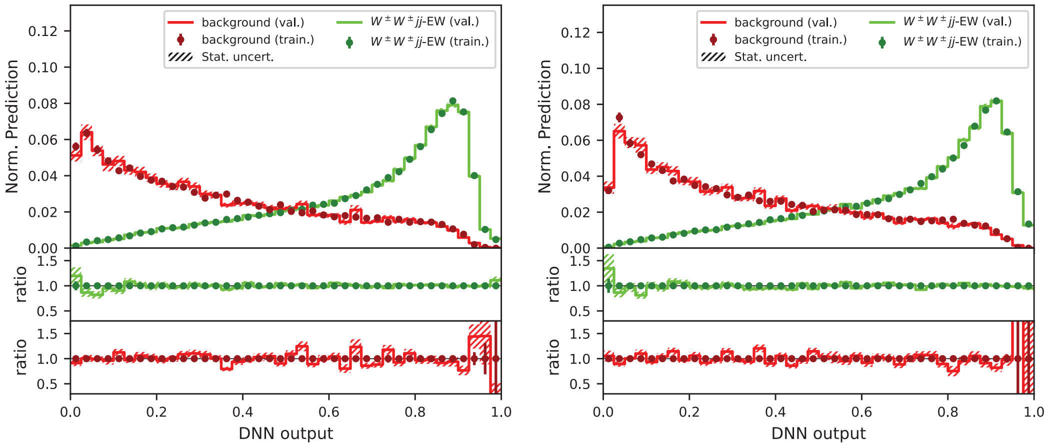


Figure 9.5: The AUC during the training process of $\text{DNN}_{W^\pm W^\pm}$ selected by the final hyperparameter optimization. The epoch with the highest AUC in the validation AUC fit is marked with a red cross.

The hyperparameters of the selected network are listed in table 9.5. The number of layers, batch size, and dropout are close to the results of the initial hyperparameter optimization in table 9.4. The number of nodes per layer has increased but this is also the case for most of the polarization networks. Therefore, the overall choice of the hyperparameters is reasonable[1]. Only the number of training epochs is significantly smaller than the other networks. However, the training of the chosen network was stopped after 34 epochs due to early stopping with a patience of five epochs. Thus, the chosen number of 24 training epochs is already close to the performance plateau and does not correspond to a significant undertraining of the network. Since the selected training setup is not optimal but still reasonable, the actual impact of alternative hyperparameters has to be discussed. The output distribution of chosen $\text{DNN}_{W^\pm W^\pm}$ is shown in figure 9.6a. The optimization algorithm in OPTIMA provides an additional set of hyperparameters which is selected according to the highest validation AUC value and does not depend on a fit. The network with the best AUC value is with six layers and 314 nodes per layer slightly smaller and is trained for 48 epochs until it reaches its optimal performance on validation data. The corresponding output distribution in figure 9.6b has a very similar shape to the network chosen by the failed fit.

---

[1] The hyperparameter being so reasonable facilitated the late discovery of the error in the optimization.



(a) Network selected by best AUC fit

(b) Network selected by best AUC value

Figure 9.6: The output distribution of DNN$_{W^\pm W^\pm}$ selected by the fit to the validation AUC in figure 9.6a and selected by the best validation AUC value in figure 9.6b. The networks are trained on fold 0 and applied to the corresponding training and validation data.

The validation AUC$_{\text{best-fit}} = 0.8158 \pm 0.0021$ of the network chosen by the fit is close to the validation AUC$_{\text{best-value}} = 0.8163 \pm 0.0028$ of the network chosen by the highest validation AUC.[2] Therefore, choosing different hyperparameters would not improve the separation of $W^\pm W^\pm jj$-EW and backgrounds.

Since the network's performance is not significantly affected by the failed fit in the optimization process, updating the network and rerunning the analysis is not reasonable. This decision is additionally supported by the expected impact of the DNN$_{W^\pm W^\pm}$ classification on the analysis results. The significance of the longitudinal polarization measurement depends primarily on the separation from the transverse polarization states. The DNN$_{W^\pm W^\pm}$ is used to split the signal region in three different regions with increasing $W^\pm W^\pm jj$-EW purity. These polarization networks, on the other hand, are used as sensitive differential distributions. Thus, minor changes in the DNN$_{W^\pm W^\pm}$ classification have no impact on the final analysis results.

---

[2] The AUC uncertainties are derived by the training and application on the five different folds used for the network training. Therefore, the uncertainties result from the differences of the different folds. Thus, the uncertainties of AUC$_{\text{best-fit}}$ and AUC$_{\text{best-value}}$ are correlated.

Chapter 10

Validation of DNN Discriminators

The neural networks trained in chapter 9 provide discriminant variables for the polarization measurement and are therefore a key element of this analysis. Three criteria on these models are validated in the following sections:

- **Validity between the training folds:**
  The networks trained on different folds provide a consistent output for data not used for training.
- **Consistency for measured data:**
  The networks behave similarly for the Standard Model prediction and measured data.
- **Decision-making based on known physics:**
  The decision of the networks should align with the predicted kinematics they are trained on.

The first two criteria are necessary to ensure a correct application of the networks. The last criterion verifies that the network behaves as expected and shows where the multivariate decision goes beyond the distributions of individual observables.

10.1 Validation of the 5-Fold Method

When applying DNN algorithms, the performance differences between training and unseen data must be considered. The k-fold method avoids a bias originating from the training data. The concept of this method is introduced in section 5.8. For the training of the neural networks in chapter 9, the data are split into five sets: three used for training, one used for validation, and one test set. The resulting neural networks are fully independent of their test sets since they were not used for training or optimization. The final DNN application is therefore a combination of five DNNs, each responsible for the corresponding test fold.

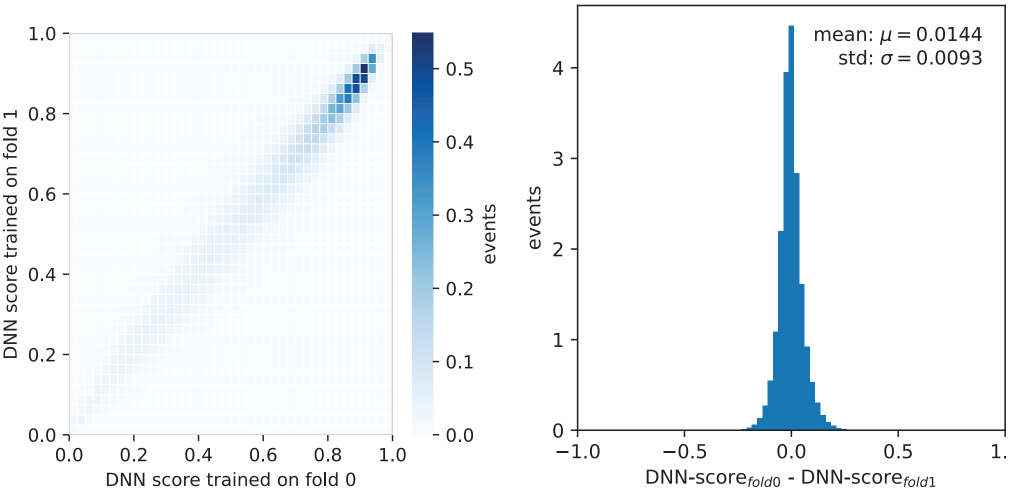
Since the training data overlap only partially and the training of a network is a statistical process, their classification decisions can differ. Even if the separation for the individual networks is excellent, the sum of the classified test datasets could be significantly less discriminant if the individual shapes are inconsistent. In order to ensure a consistent classification of unseen data, the consistency of individual decisions of DNNs trained on the different folds is tested. The DNNs are applied to common datasets, none of which are used for training, to investigate event-wise classification differences:

- DNN$_{pol}$:
  The polarization networks are trained on the polarized MadGraph samples. Therefore, the networks are applied on the unpolarized MadGraph $W^{\pm}W^{\pm}jj$-EW sample and the predictions of the individual backgrounds. Since these events are not used for training,

they can be directly used for a consistency test in the signal region without any bias from the training process.

- $\mathrm{DNN}_{W^\pm W^\pm}$:
  For the $W^\pm W^\pm jj$-EW and background separation only the polarization samples are not used for training. Therefore, only the different polarization samples are used for the closure test in the signal region. The classification for individual backgrounds is investigated in the Low-m_{jj} region not used during training.

To illustrate the concept of the test, $\mathrm{DNN}_{W^\pm W^\pm}$ trained on fold 0 and fold 1 are compared in detail. Their consistency is tested for the predicted $W^\pm W^\pm jj$-EW LL polarization state in WW-cmf. The comparison of the event-wise application is shown in the heatmap in figure 10.1a. Since the matrix is very diagonal, a strong correlation between the network decisions is indicated. This consistency is also validated by the classification difference plotted in figure 10.1b. For most events, the deviation is tiny and no systematic shift is visible. Thus, the $\mathrm{DNN}_{W^\pm W^\pm}$ trained on fold 0 and fold 1 consistently classify individual longitudinal polarized $W^\pm W^\pm jj$-EW signal events.

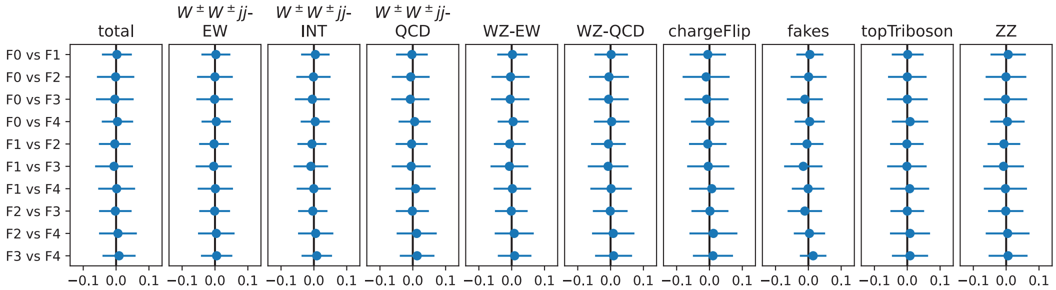


(a) Heatmap

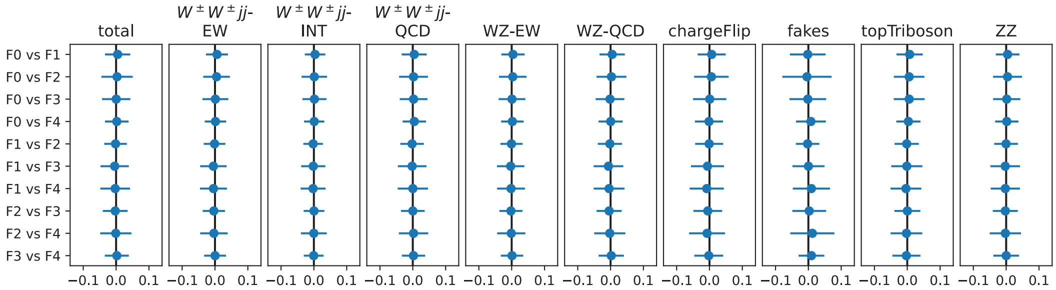
(b) Deviation

Figure 10.1: Deviation in the classification of LL polarization in WW-cmf events in the signal region for $\mathrm{DNN}_{W^\pm W^\pm}$ trained on fold 0 and fold 1. In figure 10.1a the correlation and in figure 10.1b the deviation of the classification is shown.

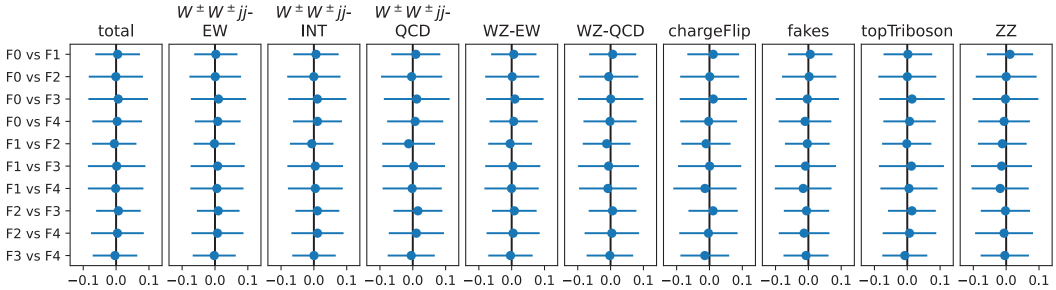
Such a detailed comparison of all trained networks and all considered processes would lead to an immense amount of correlation plots. To simplify the validation procedure, the mean deviations and the 50%-, 75%-, and 90%-quantile of the absolute deviations are studied in the following. For the neural networks trained for the polarization classification, the mean deviations are shown in figure 10.2. These figures show the event-wise deviation for the different processes passing the signal region selection. Since all mean deviations are very close to 0, no systematic shift is expected for the polarization neural networks trained on different folds. Therefore, their differences in the classification of individual events result from the non-deterministic training and are not a consequence of the different physical concepts they have learned. The 50%-, 75%-, and 90%-quantiles of the absolute deviations are shown in figure 10.3 to further investigate the event-wise differences. The classification of the networks trained on different folds differs only by less than 5% to 10% for most of the events. This excellent agreement is not limited to the $W^\pm W^\pm jj$-EW signal on which the polarization DNNs were trained but also applies to all other backgrounds. Thus, the classification of an unseen event is not very dependent on the neural network fold chosen for the classification. The remaining differences for individual events are expected to be canceled in the k-fold application.



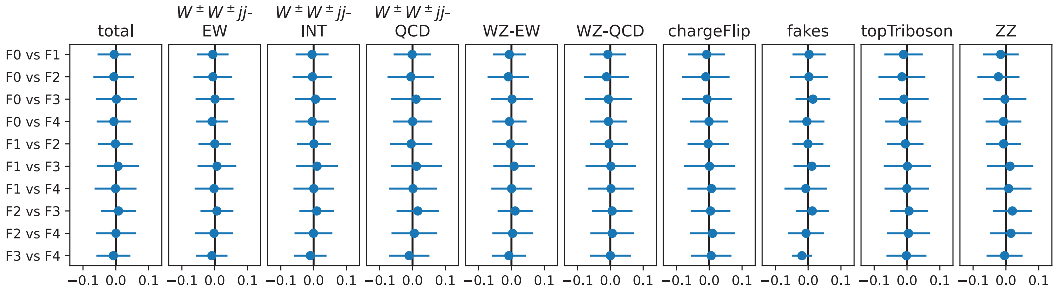
(a) LL in WW-cmf



(b) LX in WW-cmf

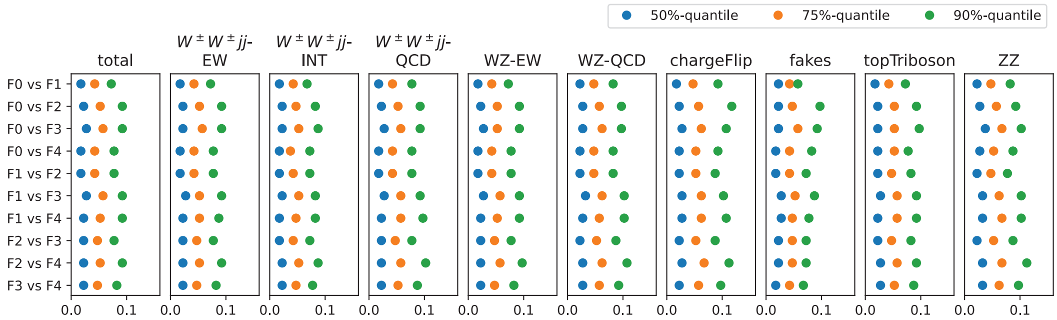


(c) LL in pp-cmf

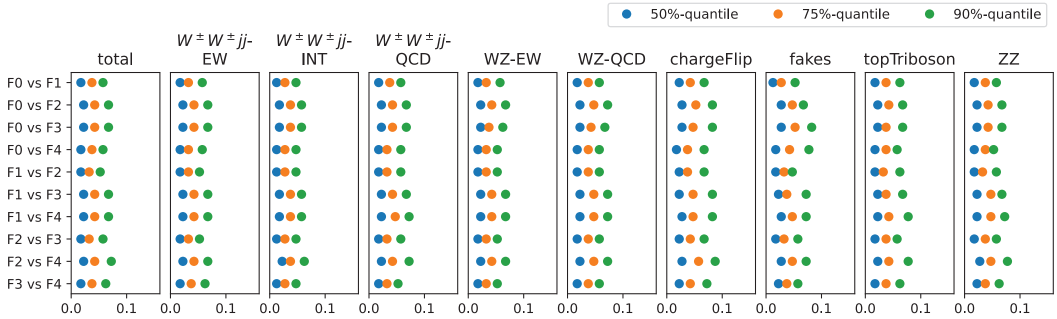


(d) LX in pp-cmf

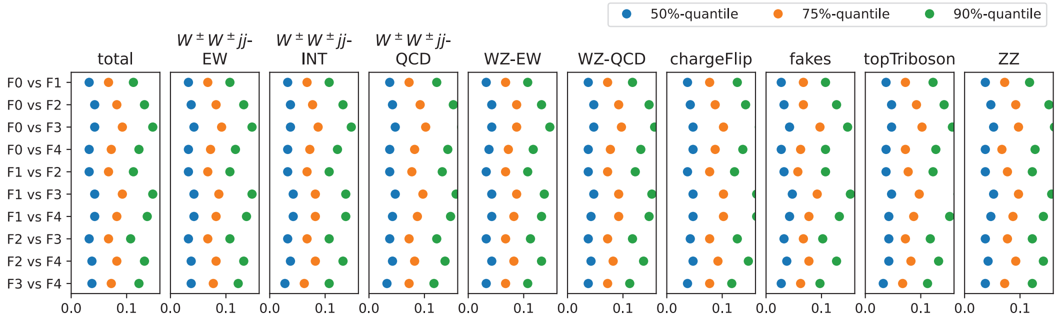
Figure 10.2: The deviation in the classification of different processes in the signal region for the polarization neural networks trained on different folds. The resulting mean deviation is shown together with the standard deviation of the classification differences.



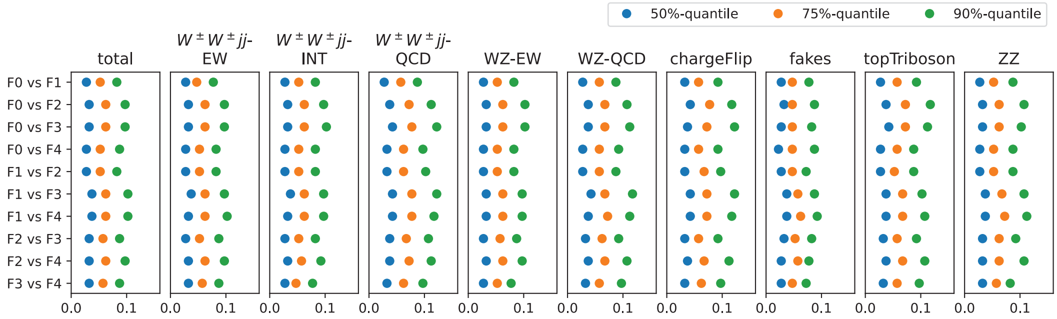
(a) LL in WW-cmf



(b) LX in WW-cmf



(c) LL in pp-cmf



(d) LX in pp-cmf

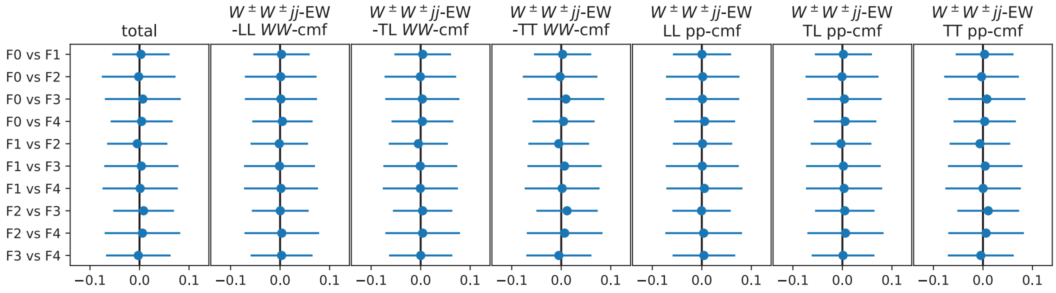
Figure 10.3: The deviation in the classification for the polarization neural networks trained on different folds. The resulting 50%-, 75%-, and 90%-quantiles of the absolute classification differences for different processes in the signal region is shown.

However, the figure 10.3 indicates significant differences between the consistencies of the different polarization DNN sets. The deviations for the polarization networks in the WW-cmf are expected to be significantly smaller than for the pp-cmf. The ten times higher number of training events in the WW-cmf leads to less statistical differences between the training folds. In addition, figure 10.3 also shows that the differences between the folds are more significant for the LL vs. TX networks than for the LX vs. TT networks. The exact reason for this is not known. One explanation could be the split of training events for the polarization classification. The ratio of raw training events $N_{LL}/N_{TX} \approx 0.3$ is significantly smaller than for $N_{LX}/N_{TT} \approx 0.8$. Thus, the polarization information for the LL vs. TX network is extracted more unevenly from the training data. The different binary split could explain why the LX vs. TT networks are statistically more robust between the different training folds.

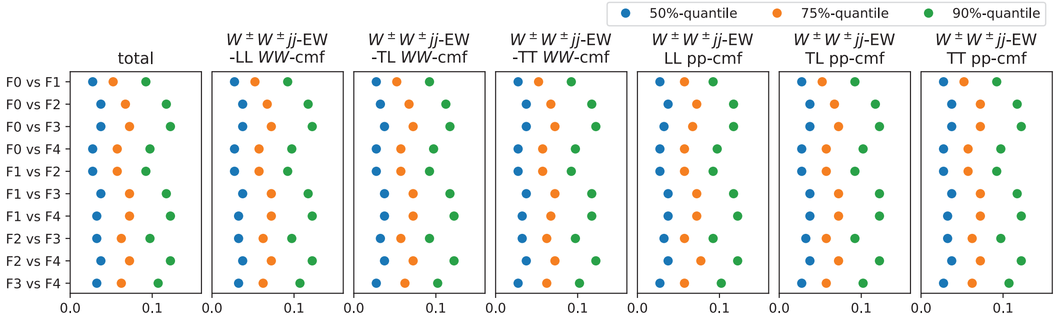
For $\mathrm{DNN}_{W^\pm W^\pm}$, only the polarization samples were not used for the training in the signal region. The mean deviation and the quantiles of the absolute deviation for these events are shown in figures 10.4a and 10.4b. The $\mathrm{DNN}_{W^\pm W^\pm}$ has no systematic bias in the classification in the signal region and the individual decisions only differ by less than 5% to 10%. Since the polarization samples cover the same kinematic phase space as the Standard Model background, this closure is representative of all events selected by the signal region.

In the Low-m_{jj} region, the test can be done with the full range of $W^\pm W^\pm jj$-EW signal and background predictions since networks are trained in the signal region. The absolute deviations in the Low-m_{jj} region shown in figure 10.4d are even smaller than in the signal region in figure 10.4b. The closure test in figure 10.5e which is performed in section 10.2 shows that the $\mathrm{DNN}_{W^\pm W^\pm}$ score in the Low-m_{jj} region is significantly shifted towards 0. Thus, the overall spread of the classification values is smaller and therefore also the deviation for individual classifications. The mean deviations shown in figure 10.4c indicate systematic differences between the training folds. The question arises whether these systematic shifts in the Low-m_{jj} region could also affect the measurement in the signal region. The pulls in figure 10.4c are consistent across the different Standard Model processes. Since no systematic shifts for the polarized $W^\pm W^\pm jj$-EW signal are found in the signal region, these shifts have to be a result of moving to the Low-m_{jj} region. This region was not included in the training and the networks trained on different folds thus differ mainly in their handling of the kinematic region with $m_{jj} < 500\,\mathrm{GeV}$. Since the Low-$m_{jj}$ region enters the final fit only via its total number of events, this bias has no impact.

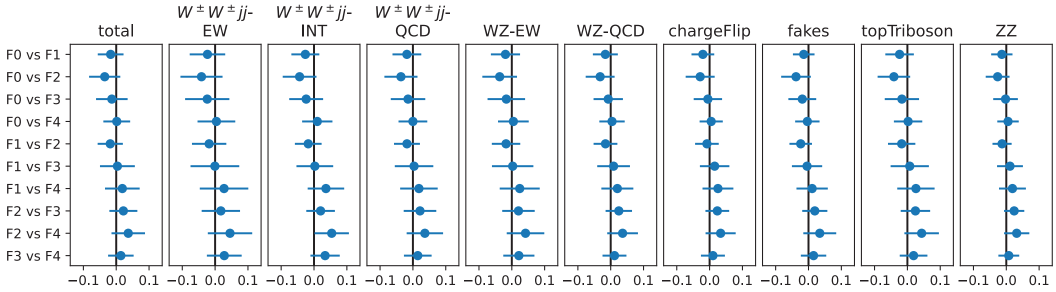
In summary, all the sets of networks used in the k-fold method are consistent between their folds. Thus, the resulting distributions that are summed in the k-fold method will be as discriminant as the individual distributions. In addition, this test shows that the networks trained on different folds have learned the same physical concepts, leading to consistent decisions for unseen data. The remaining differences are expected to originate from the non-deterministic training process. A further reduction of these deviations is not required for this analysis but could be reached by larger training datasets.



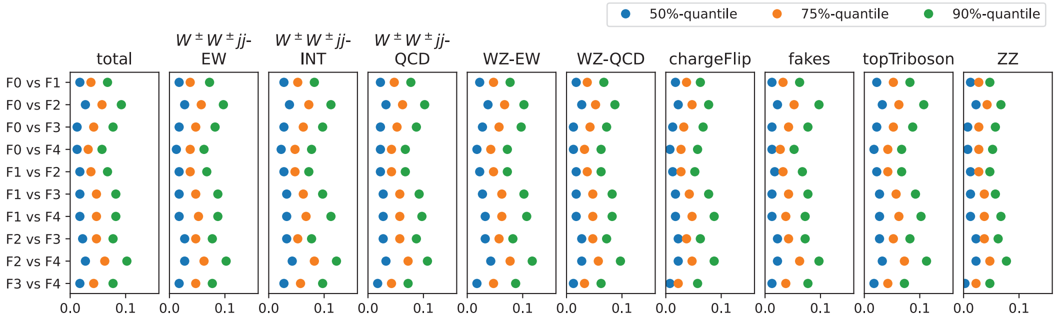
(a) Mean deviation in signal region



(b) Different quantiles of the absolute deviation in signal region



(c) Mean deviation in Low-m_{jj} region



(d) Different quantiles of the absolute deviation in Low-m_{jj} region

Figure 10.4: The deviation in the classification for the neural networks trained on different folds to separate $W^{\pm}W^{\pm}jj$-EW signal and background. In figures 10.4a and 10.4c the mean deviation is shown together with the standard deviation of the classification differences. The corresponding quantiles are given in figures 10.4b and 10.4d. The events in figures 10.4a and 10.4b pass the signal region selection and the events in figures 10.4c and 10.4d originate from the Low-m_{jj} control region.

10.2 DNN Validation with Measured Data

The neural networks for the polarization measurement are primarily trained on simulated events. Only for the charge-flip background, measured data with oppositely charged leptons are used. For the final polarization measurement, the DNNs will be applied to the background and signal predictions, resulting in the DNN distributions predicted by the Standard Model. These expected distributions are compared to the DNN output distributions for data measured by the ATLAS detector. Since the DNNs are trained on the Standard Model prediction, the question arises whether they behave differently on measured data. For example, a bias of the DNNs towards a specific simulation scheme could affect the statistical fit in the comparison with measured data.

The DNNs are applied on measured data and on Standard Model predictions not used for the training to test for such a bias. It is essential not to test distributions that are crucial for the final polarization measurement to avoid bias in the analysis strategy. For this reason, the Low-m_{jj} control region is investigated since this region is kinematically close to the signal region but enters the final fit only by its total number of events. Since the measured data in the signal region was already published in the differential unpolarized $W^{\pm}W^{\pm}jj$ measurement in Ref. [8], a validation with data in the signal region is also reasonable. Only the neural network that separates $W^{\pm}W^{\pm}jj$-EW signal and background is tested in the signal region to avoid pre-empting the polarization measurement.

In the differential unpolarized $W^{\pm}W^{\pm}jj$ measurement [8] the normalization of the $W^{\pm}Z$-QCD background and the $W^{\pm}W^{\pm}jj$-EW signal was fit to data. Since these fit results have already been published, they provide a good baseline for the following closure with measured data. The Sherpa2212 $W^{\pm}Z$-QCD is scaled by 0.69 to match the post-fit number of events in the WZ control region of the unpolarized $W^{\pm}W^{\pm}jj$ measurement in Ref. [8]. For the polarization samples, the post-fit $W^{\pm}W^{\pm}jj$-EW signal normalization of 1.14 is applied as scale factor to match the fit results in Ref. [8].

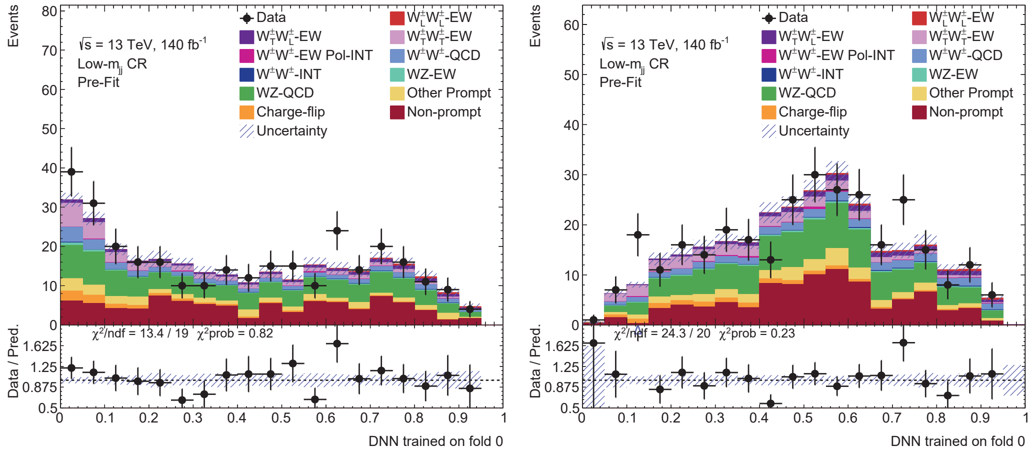
Since running over all systematic variations is computationally expensive, the following closure test considers only statistical uncertainties of the prediction.

10.2.1 Validation in Low-m_{jj} Control Region

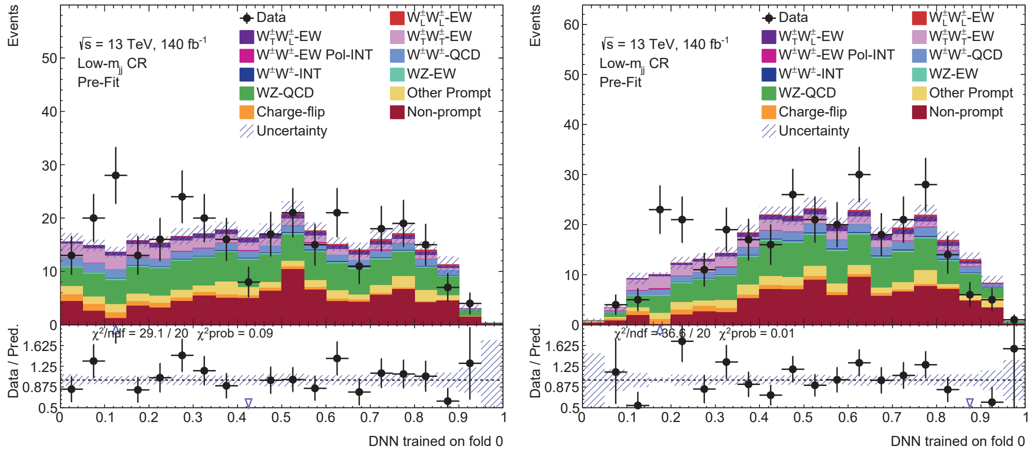
Since the Low-m_{jj} control region is not used for DNN training, the DNNs can be applied without the k-fold method without risking a bias. Thus, each DNN trained on a different fold can be validated on the full Standard Model prediction in the Low-m_{jj} control region. For all neural networks trained on fold 0, the classification of the predicted and measured data is shown in figure 10.5. The closure tests of the remaining folds are shown in figures D.1 to D.5. An overview of all the performed closure tests is given by the stat-only χ^2 values and the corresponding p-values listed in table 10.1. The distributions shown in the plots and the resulting χ^2 values indicate good agreement between the prediction of the Standard Model and the measured data. Since no significant shift is observed in any of the distributions, the DNNs do not appear sensitive to differences in the kinematics of prediction and data in the Low-m_{jj} control region. Due to the kinematic similarities between the Low-m_{jj} control region and the signal region, it can be assumed that this should also be the case in the signal region.

In addition to the closure between prediction and data, the overall shape of the distributions is reasonable. The polarization DNNs in figures 10.5a to 10.5d separate the LL and TT polarization states in the $W^{\pm}W^{\pm}jj$-EW signal contribution. The networks have thus learned general concepts of polarization kinematics that also apply in the Low-m_{jj} region, which does not overlap with the kinematic region used for training.

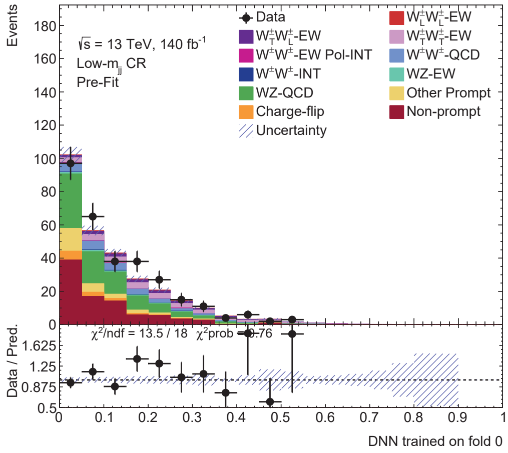
For the DNN$_{W^{\pm}W^{\pm}}$ in figure 10.5e, most of the events are classified to background. The bias towards background classification is expected since m_{jj} is a discriminant variable for $W^{\pm}W^{\pm}jj$-EW against the background. Nevertheless, a separation between $W^{\pm}W^{\pm}jj$-EW and background can still be seen in these histograms proving that the network can differ the contributions also in a kinematic region not used during training.



(a) DNN for LL in WW-cmf **(b)** DNN for LX in WW-cmf



(c) DNN for LL in pp-cmf **(d)** DNN for LX in pp-cmf



(e) DNN for $W^{\pm}W^{\pm}jj$-EW vs. bkg

Figure 10.5: DNNs trained on fold 0 applied in the Low-m_{jj} control region. According to the fit results from the unpolarized $W^{\pm}W^{\pm}jj$ measurement [8], the $W^{\pm}Z$-QCD contribution is scaled by 0.69 and the $W^{\pm}W^{\pm}jj$-EW signal is scaled by 1.14.

Table 10.1: The stat-only χ^2 values and the corresponding p-values for the DNN closure test with measured data in the Low-m_{jj} control region.

DNN	χ^2/N_{df}	p-value
LL in WW-cmf (fold 0)	0.71	0.82
LL in WW-cmf (fold 1)	0.76	0.75
LL in WW-cmf (fold 2)	0.86	0.64
LL in WW-cmf (fold 3)	0.89	0.60
LL in WW-cmf (fold 4)	0.56	0.93
LL in pp-cmf (fold 0)	1.46	0.09
LL in pp-cmf (fold 1)	1.14	0.31
LL in pp-cmf (fold 2)	0.94	0.53
LL in pp-cmf (fold 3)	1.25	0.20
LL in pp-cmf (fold 4)	0.67	0.86
LX in WW-cmf (fold 0)	1.22	0.23
LX in WW-cmf (fold 1)	1.05	0.40
LX in WW-cmf (fold 2)	1.44	0.09
LX in WW-cmf (fold 3)	1.66	0.03
LX in WW-cmf (fold 4)	1.04	0.41
LX in pp-cmf (fold 0)	1.83	0.01
LX in pp-cmf (fold 1)	1.27	0.19
LX in pp-cmf (fold 2)	1.68	0.03
LX in pp-cmf (fold 3)	0.97	0.50
LX in pp-cmf (fold 4)	1.96	0.01
$W^\pm W^\pm jj$-EW vs. bkg (fold 0)	0.75	0.76
$W^\pm W^\pm jj$-EW vs. bkg (fold 1)	1.1	0.33
$W^\pm W^\pm jj$-EW vs. bkg (fold 2)	0.73	0.80
$W^\pm W^\pm jj$-EW vs. bkg (fold 3)	0.74	0.79
$W^\pm W^\pm jj$-EW vs. bkg (fold 4)	1.06	0.39

10.2.2 Validation in the Signal Region

To avoid bias in the analysis strategy none of the polarization DNNs are tested in the signal region. The ensemble of DNNs separating $W^\pm W^\pm jj$-EW signal and background is applied in the k-fold method, resulting in one histogram that is not affected by a bias from the training and optimization procedure of the DNNs. The resulting distribution of the Standard Model prediction in figure 10.6 is in excellent agreement with the data in the signal region. The stat-only p-value of 11% supports this. Since there is no shift between the two distributions, the information extracted from the Standard Model prediction during training is consistent with the kinematic of the measured data.

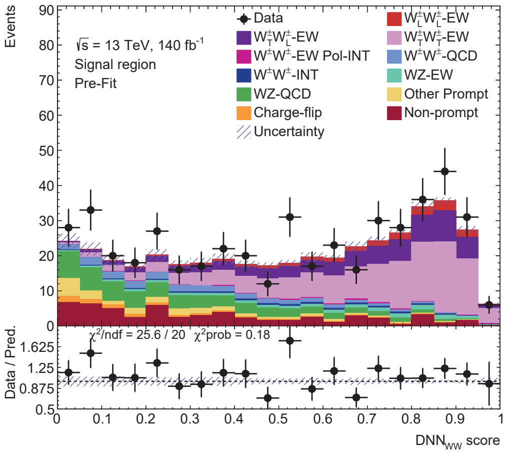


Figure 10.6: The DNNs trained on different folds to separate $W^\pm W^\pm jj$-EW and background are applied in the signal region. Each DNN is only applied on its corresponding fold and is not used during training and optimization. The resulting distributions are stacked. According to the fit results from the unpolarized $W^\pm W^\pm jj$ measurement [8], the $W^\pm Z$-QCD contribution is scaled by 0.69 and the $W^\pm W^\pm jj$-EW signal is scaled by 1.14.

10.3 Interpretation of the Classifier Decision-Making

The neural networks DNN$_{W^\pm W^\pm}$ and DNN$_{pol}$ are trained on the phase space kinematics of the Standard Model prediction. Thus, the decisions of the networks are based on the concepts of physics they have learned from the training data. The SHAP values introduced in section 5.9 allow the investigation of the decision-making of the networks. The Kernel SHAP method [74] is used for the event-wise calculation of the SHAP values. First, the setup of the Kernel method is examined to make the appropriate choice of the required data sets. This setup is then applied to the classifier networks trained in chapter 9.

10.3.1 Dataset Choice for SHAP Sampling

Due to the k-fold application introduced in section 5.8 an ensemble of networks is applied to fill the DNN histograms. Thus, the SHAP analysis of the feature importance is evaluated for each fold and combined afterward. Calculating the SHAP values requires a mask dataset to sample the feature distribution and an explanation dataset containing the individual events to be explained. The mask sample is extracted from the training dataset of the corresponding network following the recommendation in Ref. [74]. The explanation sample is taken from the validation dataset of the respective fold to be consistent with the unbiased test dataset used in the statistical evaluation of the polarization analysis.

Since the trained networks are binary classifiers, the mask and explanation sample are each equally split into the two categories used for the network training. The events in each of the two categories are randomly chosen according to their event weight. Thus, the chosen events are expected to follow the predicted distributions. Since the computational effort grows with the size of the mask and explanation sample, a reasonable number of events has to be chosen for both samples.

For the explanation dataset, 2000 events are chosen in total. This results in 1000 events per binary category, 200 events each per fold. Due to the increasing computational effort, a significantly larger explanation dataset is not reasonable.

The choice of the number of events in the mask data set used for Shapley sampling is associated with a trade-off. A larger mask dataset leads to a more accurate calculation of the individual SHAP values, but the computational effort increases significantly. Since there is no official recommendation for the size of the mask dataset, the impact of the number of mask events is studied. While keeping the explanation dataset fixed, the mask dataset is randomly sampled with different numbers of mask events N_{mask}. The mask events are taken from the training dataset of the corresponding fold with $N_{mask}/2$ events per binary category. To keep the computational effort reasonable the number of mask events is varied in the range $N_{mask} \in \{20, 40, 60, 80, 100\}$ leading to independent mask datasets of increasing size.

For $\mathrm{DNN}_{W^\pm W^\pm}$, the impact of N_{mask} on the mean absolute SHAP values for each feature is shown in figure 10.7. The shown uncertainties represent the statistical uncertainty of the explanation dataset and is therefore fully correlated between the runs with different N_{mask}. It becomes clear that already a very small mask dataset with only 20 events is sufficient to study the overall impact of the features on the network decision. After $N_{mask} = 40$ the calculation of the mean feature impact is very stable, and the difference between the individual runs are in the same magnitude as the statistical uncertainty of the explanation dataset. The impact of the increasing N_{mask} for the polarization DNNs has equivalent effects and is therefore shown in the appendix in figure E.1.

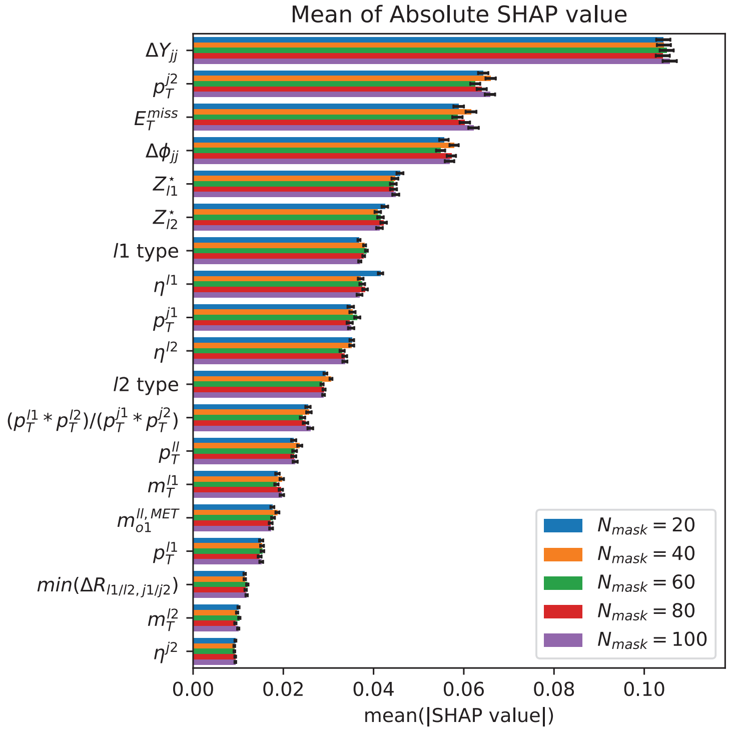


Figure 10.7: The mean absolute SHAP values of $\mathrm{DNN}_{W^\pm W^\pm}$ for different input features. The assigned uncertainties originate from the statistical uncertainty of the explanation dataset. The SHAP evaluation is done for independent mask datasets of different size.

In figures 10.7 and E.1 the impact of choosing a mask dataset with $N_{mask} > 40$ seems negligible for estimation of the mean feature impact. However, this may also result from individual differences in the SHAP value calculation averaging out due to the size of the explanation dataset. Therefore, for each time N_{mask} was increased by 20 events the change on the SHAP value calculation is evaluated for each event of the common explanation dataset. If the calculation of the SHAP values becomes more accurate with larger N_{mask}, the event-by-event differences ΔSHAP should become smaller with each increase in N_{mask}. The quantiles of the event-wise differences ΔSHAP of $\mathrm{DNN}_{W^\pm W^\pm}$ is shown in figure 10.8. The plots for the polarization networks are very similar and therefore moved to the appendix in figure E.2.

With decreasing feature importance also the absolute differences between the different SHAP runs decreases. This is reasonable and consistent with the differences seen in figures 10.7 and E.1. By comparing the quantiles in figure 10.8 for each feature used by the DNNs, the differences in the SHAP values decrease with increasing N_{mask}. However, for most features, the change from $N_{mask} = 60$ to $N_{mask} = 80$ has a similar event-wise impact as the change from $N_{mask} = 80$ to $N_{mask} = 100$. Thus, the number of mask events probably

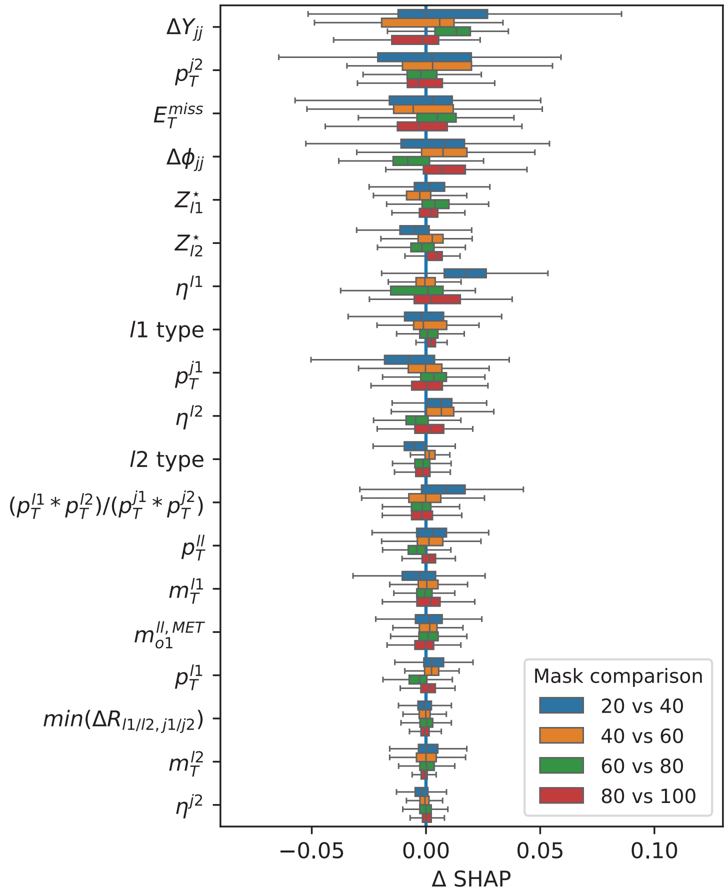


Figure 10.8: The SHAP values of DNN$_{W^\pm W^\pm}$ originating from mask datasets with increasing size are compared by calculating the event-wise difference. This is done for each time N_{mask} was increased by 20. For better visibility the individual differences are combined into a box plot of the quantiles and whiskers covering the full range.

has to be increased into the order of $N_{mask} = 200$ or higher to further improve the SHAP accuracy. Since the event-wise impact on the SHAP value calculation for $N_{mask} = 100$ is already below 0.02 for most of the events a much more computationally intensive calculation with $N_{mask} = 200$ is not reasonable.

Therefore, 100 mask events are chosen for each fold. The systematic uncertainty of per-event SHAP values arising from the limited number of mask events will not be studied further. However, this uncertainty is expected to be below 0.02 for most events and thus does not significantly affect the SHAP evaluation in the next section.

10.3.2 Evaluation of Feature Impact

SHAP values indicate the importance of an input feature for a specific decision of the DNN. Therefore, analyzing the SHAP values over the entire kinematic phase space gives an indication of the physical concepts that the DNN has learned during training. Since kinematic differences between the $W^\pm W^\pm jj$-EW signal and the backgrounds are better understood than the kinematic differences between the polarizations, DNN$_{W^\pm W^\pm}$ is a good starting point for studying the SHAP values.

Feature Impact for DNN$_{W^\pm W^\pm}$

The SHAP values evaluated for the 5-fold DNN$_{W^\pm W^\pm}$ ensemble are shown in figure 10.9a. The individual feature values are encoded in the color of the corresponding SHAP values, taking into account the non-linear scaling in table 9.2 used for training. The difference between the mean values in $W^{\pm}W^{\pm}jj$-EW signal $\bar{x}_1$ and background $\bar{x}_0$ are listed to enable a direct comparison with the feature distribution.

The DNN decision significantly depends on the separation of the jets in Δy_{jj} and $\Delta\phi_{jj}$. By classifying events with a high jet separation to be more signal-like the DNNs have learned one of the main signatures of the $W^{\pm}W^{\pm}jj$-EW signal. Also, the preference for higher missing transversal momentum is consistent with the existence of two neutrinos in the signal definition.

In general, a clearer separation in $\bar{x}_1 - \bar{x}_0$ results in the behavior of the neuronal network also following this distribution difference, as in the case of the Zeppenfeld variables of the leptons. Features without a strong separation between signal and background like the η distributions also don't show a clear correlation with the DNN decision-making, since they depend more on the values of the other features.

For the lepton types, muons are more preferred to be classified as signal than electrons. As the corresponding $(\bar{x}_1 - \bar{x}_0)/\sigma(x)$ differences indicate, the separation is stronger for the type of the leading lepton. The relevance of the lepton type can be explained by the charge-flip background that only affects electrons. On this point the mean absolute impact of each feature in plotted figure 10.9b becomes relevant. While this plot confirms the behavior in figure 10.9a again, the impacts on the decisions for different processes are also compared. Despite the limited number of evaluated events, one can see additional attention is paid to the η distribution and the type of leading lepton in the case of charge-flip background. If this is true, the model has learned that specific kinematic regions are more likely for charge-flip effects requiring a higher focus on the lepton type.

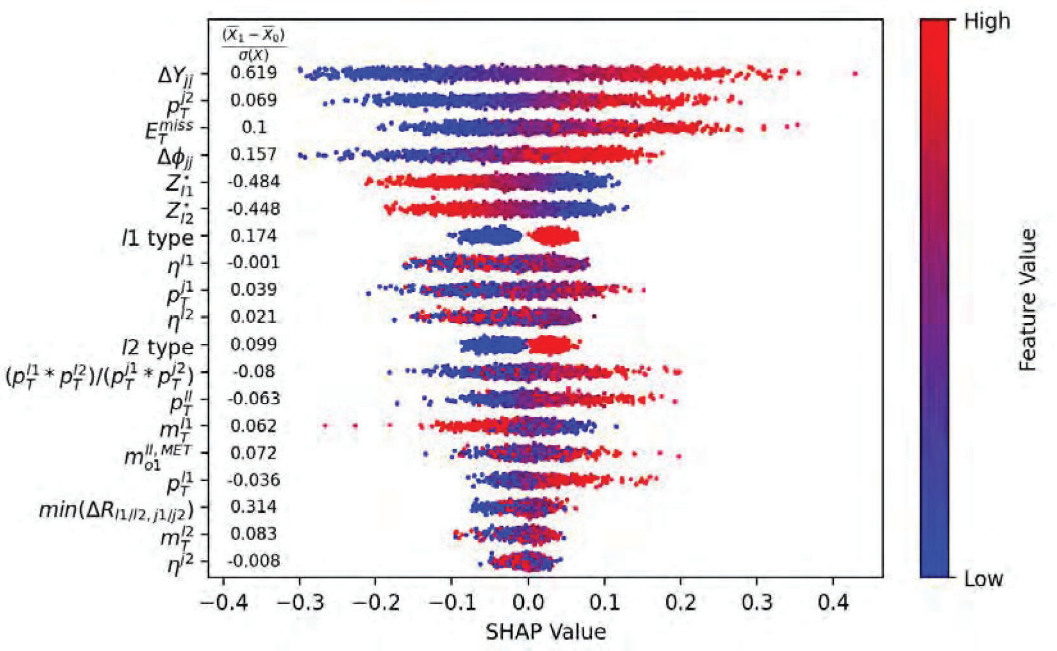
The overall distribution of the SHAP values shows that even if the ensemble of DNNs is a black box the resulting decision-making seems comprehensible based on the prior physics knowledge. Therefore, the DNNs separating $W^{\pm}W^{\pm}jj$-EW signal and background appear to be a reasonable encoding of the kinematic phase space used for training.

Feature Impact for Polarization Classification

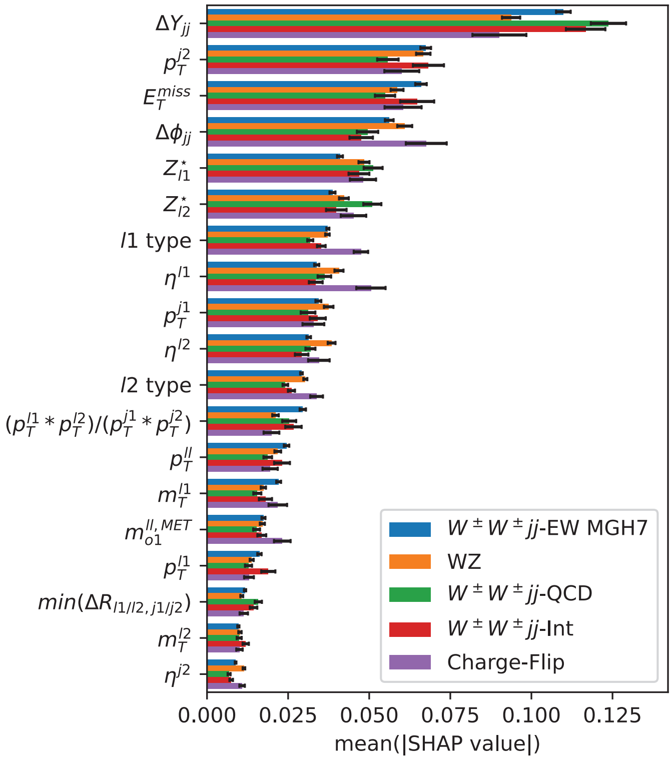
The SHAP values for the polarization DNNs are shown in figures 10.10 and 10.11. As one would expect all the polarization DNNs show a quite similar behavior in the event classification. All DNNs take a low lepton p_T and a low p_T^{miss} as an indicator for longitudinal polarization. This is also consistent with the difference between the mean transverse momenta listed in the figures. Due to the symmetry of the detector, the η values of individual objects are not linearly correlated with the DNN decision-making. This is also the case for most of the azimuth angles of the jets in respect to the leading lepton $\phi^j - \phi^{l1}$. However, for other angle differences like $\Delta\phi_{jj}$, Δy_{jj}, and $\Delta\eta_{ll}$ a correlation between feature value and impact on the decision is clearly visible. This is also reasonable, since the mean values of these angle differences differ significantly between longitudinal and transverse polarization states.

Nevertheless, a more prominent difference in the mean feature values does not necessarily result in a higher impact on the network decision. Although features with the highest difference $(\bar{x}_1 - \bar{x}_0)/\sigma(x)$ also have the greatest impact on the decision, there are also enough exceptions. One example is the p_T^{j1} in figure 10.10a which differs significantly between LL and TX in the pp-cmf but has a very minor impact on the final decision of the network. Similar effects can be seen when comparing the different DNNs. The $m_o^{ll,MET}$ feature has a significantly smaller value for longitudinal polarized events. The DNNs for the polarization defined in WW-cmf focus on this feature in figures 10.10a and 10.10b. However, the impact of $m_o^{ll,MET}$ in figure 10.10b on the decisions of the DNN trained on LX in pp-cmf is much less dominant, despite a similar value for $(\bar{x}_1 - \bar{x}_0)/\sigma(x)$.

Another very interesting behavior can be seen for the M_T^{l2} SHAP value distribution for all DNNs. Despite the fact that M_T^{l2} is on average a bit lower for LL/LX than for TX/TT events the DNN takes higher M_T^{l2} values as an important indicator for LX polarization. This counterintuitive behavior could be a result of the already favored low p_T^{l2} and low p_T^{miss} values for longitudinal polarization. Since M_T^{l2} is very correlated with p_T^{l2} and p_T^{miss} some kinematic information is already available for the DNN allowing a more complex evaluation of this feature.



(a) SHAP values



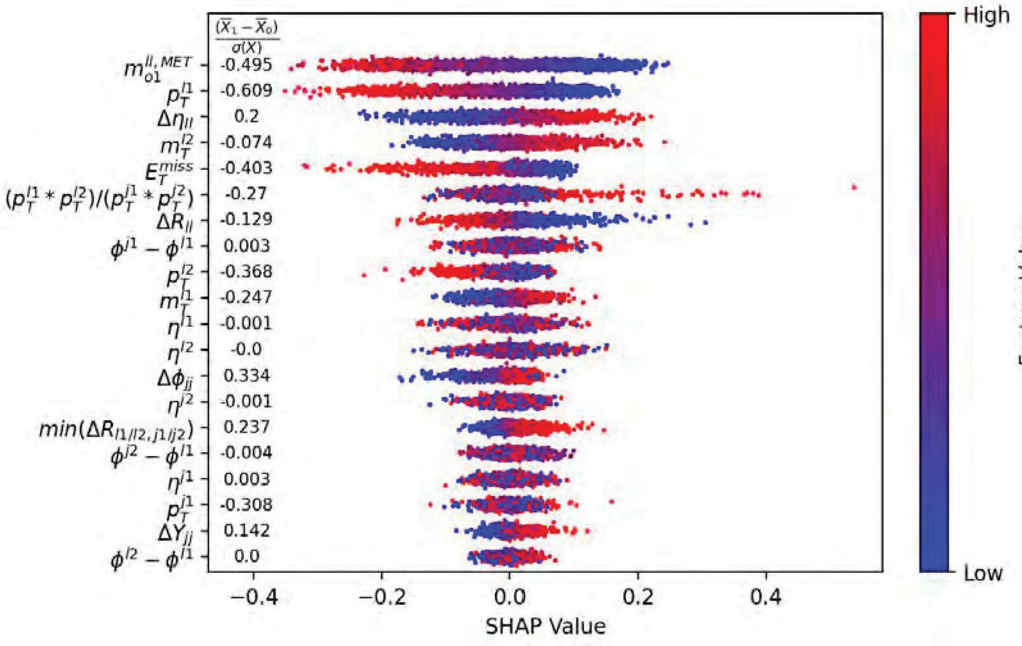
(b) Mean absolute SHAP value

Figure 10.9: The SHAP values are evaluated for DNN$_{W^\pm W^\pm}$. In figure 10.9a the individual SHAP values are plotted with a color encoding the value of the corresponding feature. In addition, it is also listed how much the mean values of the features in $W^\pm W^\pm jj$-EW signal $\bar{x}_1$ and background $\bar{x}_0$ differ. Figure 10.9b shows the mean absolute impact of each feature split into the 5 major contributions in the training data.

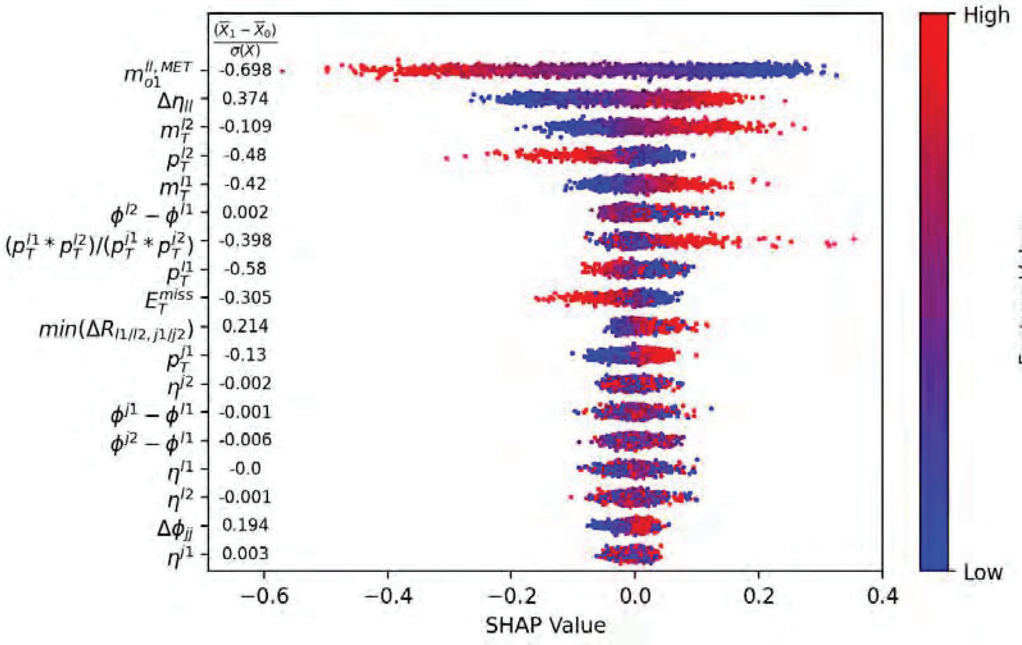
The mean impact of each feature is shown in figure 10.12. The ranking indicates that the polarization networks favor the lepton and missing momentum kinematic for the polarization classification. Jet kinematics have a lower impact on the network decision. Thus, the networks focus on the decay products of the $W^{\pm}$ bosons rather than on the jets associated with the initial emission of the $W^{\pm}$ bosons.

The feature impact is quite consistent across the polarization datasets. This is expected due to the kinematic overlap of the different polarizations. However, for some features there are visible differences of the mean SHAP values between the polarization states. In most of these cases the decision-making for the LL and TT contributions differs the most while the mixed polarization is somewhere between. Thus, the DNN seem to have found kinematic regions that are dominated by fully transversal or fully longitudinal polarization states. In these regions the DNN potentially changes its preferences within the set of input features.

In summary, it can be concluded that the individual polarization DNNs have learned similar concepts and their decisions can in general be explained by the physical expectations. However, one can also see that the decision-making of the networks goes far beyond the individual kinematic distributions and establishes complex connections between the input features. Therefore, a complete explanation of the DNN behavior is not achievable but the calculation of SHAP values allows justified assumptions to be made regarding the decision-making process, that will be used in section 12.4.

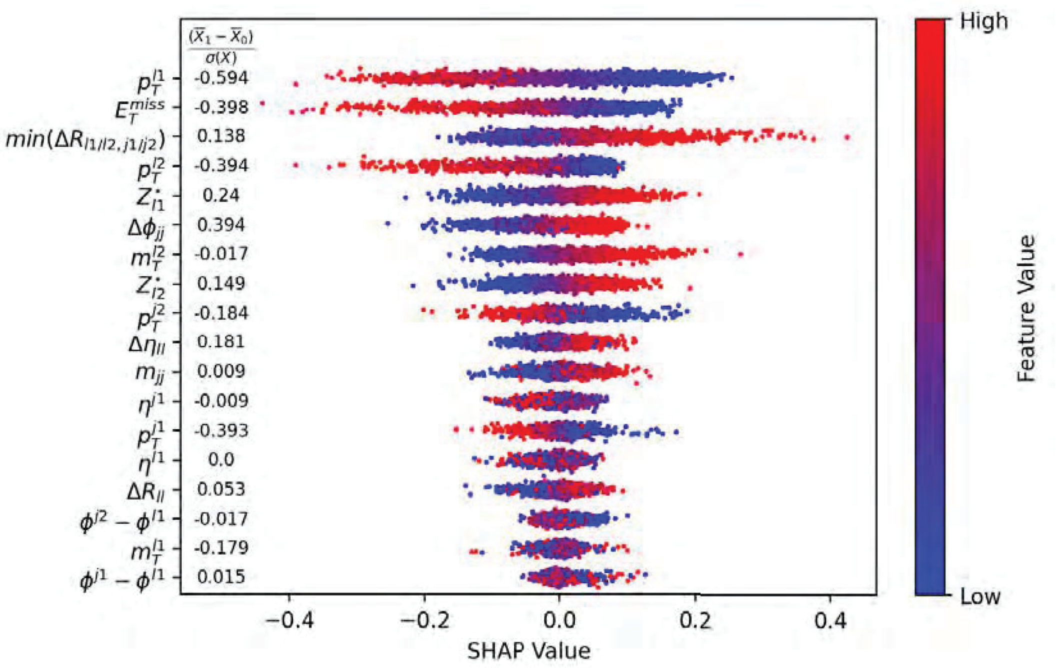


(a) LL in WW-cmf

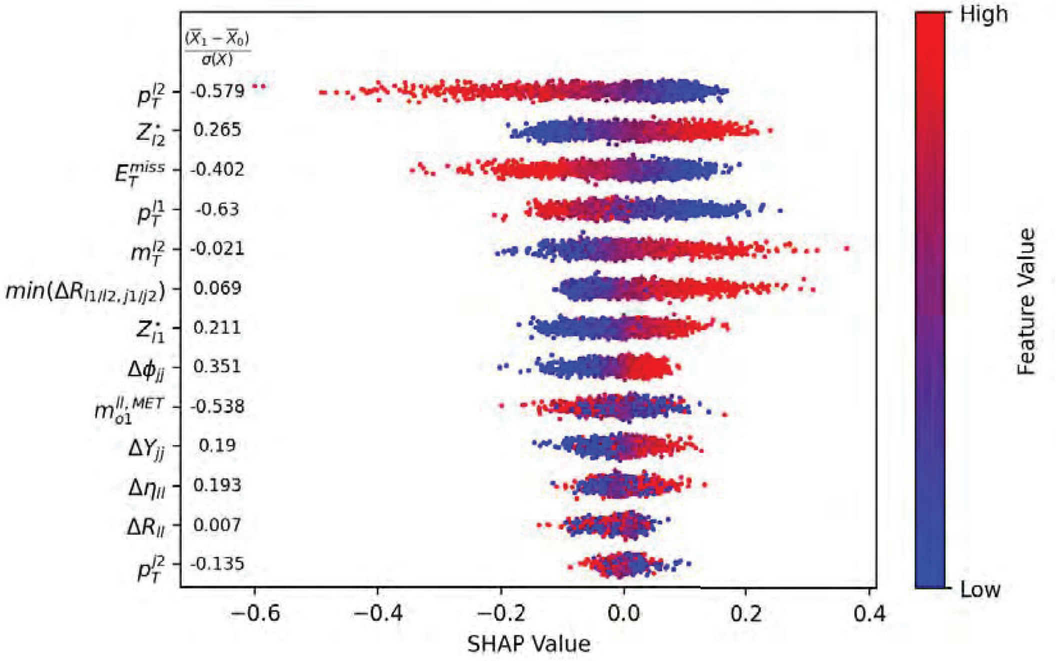


(b) LX in WW-cmf

Figure 10.10: The SHAP values of the polarization DNNs trained in WW-cmf are plotted with a color encoding the value of the corresponding feature. In addition, it is also listed how much the mean values of the features in LL/LX-polarization $\bar{x}_1$ and TX/TT-polarization $\bar{x}_0$ differ.

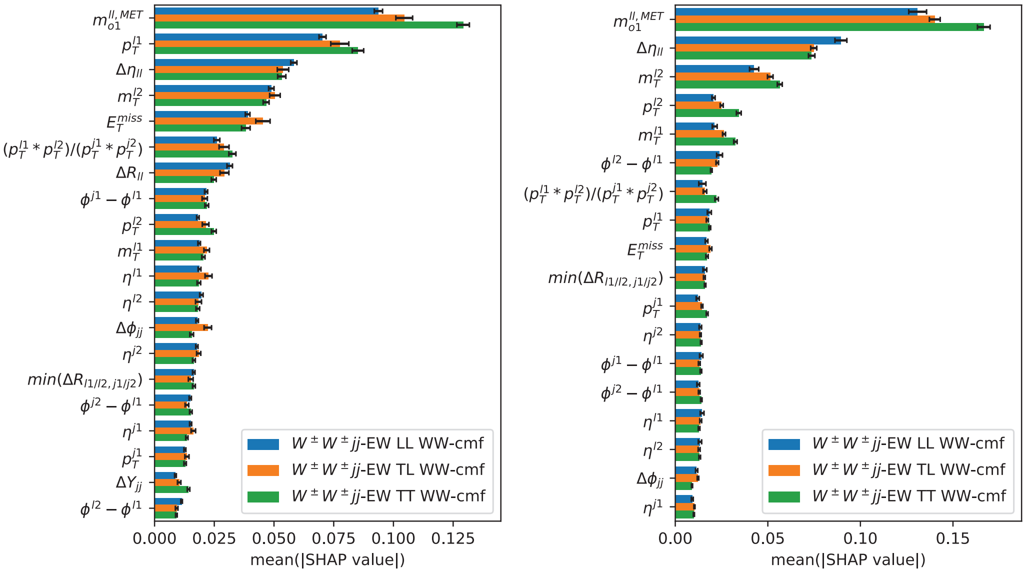


(a) LL in pp-cmf



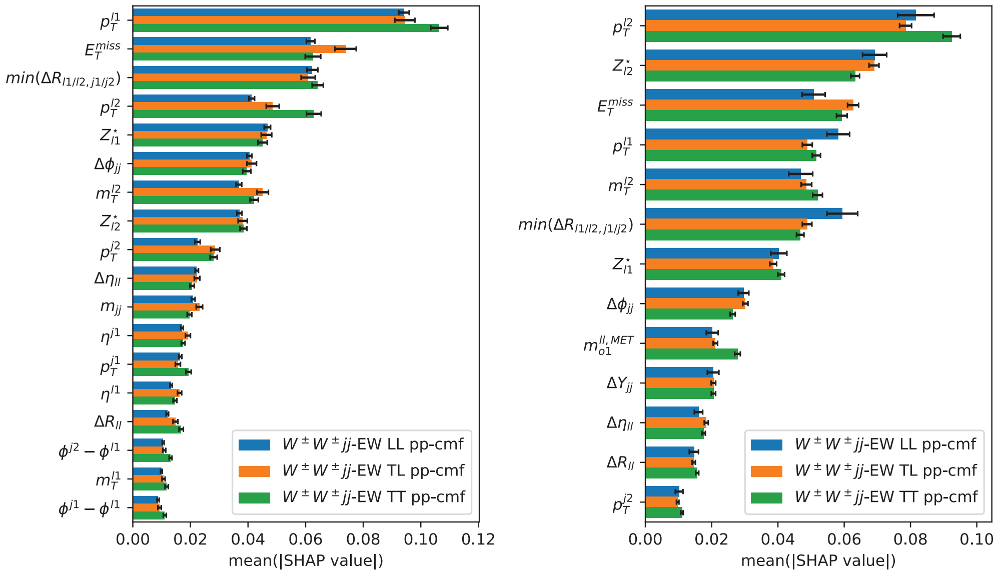
(b) LX in pp-cmf

Figure 10.11: The SHAP values of the polarization DNNs trained in pp-cmf are plotted with a color encoding the value of the corresponding feature. In addition, it is also listed how much the mean values of the features in LL/LX-polarization $\bar{x}_1$ and TX/TT-polarization $\bar{x}_0$ differ.



(a) LL in WW-cmf

(b) LX in WW-cmf



(c) LL in pp-cmf

(d) LX in pp-cmf

Figure 10.12: The mean absolute SHAP values of the polarization DNNs split into the LL, TL, and TT contributions in the corresponding reference frame.

Chapter 11

Uncertainties

Analyses at experiments like ATLAS have to deal with a variety of different uncertainty sources. The uncertainties can be categorized into experimental and modeling uncertainties. Experimental uncertainties originate from the measurement and reconstruction of the data. The modeling uncertainties, on the other hand, are the uncertainties of the Standard Model modeling which can be improved in future theoretical calculations. The modeling uncertainty includes the specific choices in the simulation process and the subsequent corrections to improve the modeling. Since the polarization analysis uses the same background, and object and event selection as the unpolarized $W^{\pm}W^{\pm}jj$-EW analysis in Ref. [8], most uncertainties are already derived and studied. The uncertainties new in this analysis affect the polarization prediction. Therefore, this chapter provides a short overview of the background uncertainties known from Ref. [8] and focuses mainly on the theory uncertainties of the polarization states.

11.1 Experimental Uncertainties

The experimental uncertainties of this polarization analysis have a variety of origins. Each event is affected by the uncertainty of the integrated luminosity [19] of the ATLAS measurement and a scale variation `PRW_DATASF` for the pileup reweighting [103]. For the lepton reconstruction, the four-momentum variation `EG_SCALE_ALL` [26] of the electron energy scale and the four-momentum resolution uncertainty `MUON_MS` [27] from the muon system are considered. For the jet reconstruction and b-tagging described in Ref. [30] and Ref. [83], a wide range of variations are considered. The uncertainty of the four-momentum is composed of

- `JET_EffectiveNP_Modelling` (1 to 4),
- `JET_EtaIntercalibration_Modeling`,
- `JET_Flavor_Composition`,
- `JET_Flavor_Response`,
- `JET_Pileup_OffsetMu`,
- `JET_Pileup_OffsetNPV`,
- and `JET_Pileup_RhoTopology`.

The uncertainty of the energy resolution of the jet four-momentum is derived by eleven `JET_JER_EffectiveNP` variations. For the uncertainty of the b-tagging efficiency, the scale factor variations

- `FT_EFF_Eigen_B`,
- `FT_EFF_Eigen_C`,
- `FT_EFF_Eigen_L`,

- `FT_EFF_Eigen_extrapolation`,
- and `FT_EFF_Eigen_extrapolation_from_charm`

are applied. The reconstruction of the missing transverse momentum is affected by the `MET_SoftTrk_ResoPara`, `MET_SoftTrk_ResoPerp`, and `MET_SoftTrk_Scale` variation.

In addition to the reconstruction and identification process, further uncertainties arise from the data-driven background modeling. For the data-driven estimation of the non-prompt background, the fake factor uncertainties are listed in section 7.2.2. The uncertainty of the charge-flip background in section 7.2.3 is provided by scale factor variations provided by the *Egamma Combined Performance (CP) group* [78]. Since these two backgrounds originate from a misclassification in the ATLAS detector, their corresponding uncertainties are classified as experimental.

11.2 Modeling Uncertainties

The dominant modeling uncertainties affect the $W^{\pm}W^{\pm}jj$-EW polarization states and the $W^{\pm}Z$-QCD background. For both processes, the modeling uncertainties are split into the uncertainties of the simulation and the uncertainties of the modeling corrections. The less significant modeling uncertainties of the prompt backgrounds described in section 7.2.4 are estimated by normalization uncertainties derived in Ref. [8].

11.2.1 $W^{\pm}W^{\pm}jj$-EW Polarization Theory Uncertainties

The simulation of the Standard Model prediction is described in chapter 4. In order to make reliable predictions for the proton-proton collisions at the LHC the knowledge of the quark and gluon structure within the proton is essential. This proton structure is given by the parton distribution functions (PDFs) for a specific energy scale. Since the PDFs are extracted by experimental measurements in combination with Standard Model predictions, they have a variety of uncertainties, which will be explained in this section. In addition to selecting the PDF, the strong coupling α_S and the renormalization, factorization, resummation, and merging scale have to be chosen in the simulation steps described in chapter 4.

The application of theory uncertainties was already shortly discussed for the higher-order QCD correction on $W^{\pm}W^{\pm}jj$-EW in section 8.1.8. This discussion concludes that the theory uncertainties are underestimated at leading-order QCD. Therefore, the Sherpa $l^{\pm}\nu l^{\pm}\nu jj+0,1j$ sample (DSID 700970) is considered for this uncertainty calculation. The Sherpa $l^{\pm}\nu l^{\pm}\nu jj+0,1j$ provides the target distribution for the higher-order QCD correction in section 8.1 and thus provides a more realistic uncertainty estimation than the LO-QCD polarization samples. The nominal PDF sets used for the Sherpa samples is provided by the PDF4LHCWorking Group [104]. The following uncertainties will be propagated to the Sherpa and MadGraph polarization samples used for the polarization measurement. The variations of the PDF, nominal PDF set, strong coupling, and renormalization and factorization scale are accessible in the Sherpa $l^{\pm}\nu l^{\pm}\nu jj+0,1j$ sample via so-called on-the-fly weights. For the resummation and merging scale variations, the alternative samples listed in table 11.1 were produced.

The considered theory uncertainties are:

- **PDF Variations**
  A PDF is obtained by a fit to experimental data in Ref. [104], resulting in uncertainties in the fitting procedure. The used PDF4LHC_40 set contains the nominal PDF and 40 PDFs with varied parametrization. The quadratic sum over the Hessian PDF set

  $$\delta^{\mathrm{PDF}}\sigma = \sqrt{\sum_{i=1}^{40}\left(\sigma_i - \sigma_0\right)^2} \tag{11.1}$$

  is expected to cover the internal PDF uncertainty on the cross-section σ. The index i indicate the different Hessian eigenvector directions with $i=0$ being the central PDF.

Table 11.1: Sherpa $l^{\pm}\nu l^{\pm}\nu jj+0,1j$ samples with resummation or merging scale variation.

DSID	sample name
Sherpa $l^{\pm}\nu l^{\pm}\nu jj+0,1j$ resummation scale MU_Q	
700998	Sh_300b1_llvvjj_ss_MUQ20_corrected
700999	Sh_300b1_llvvjj_ss_MUQ05_corrected
Sherpa $l^{\pm}\nu l^{\pm}\nu jj+0,1j$ merging scale CKKW	
701215	Sh_300_llvvjj_ss_CKKW30
701216	Sh_300_llvvjj_ss_CKKW15

- **Alternate PDF Sets**
  Since PDF sets differ by their parametrization and the input data used for the fit, the choice of the PDF set can affect the resulting theory prediction. Thus, the nominal PDF set is compared to CT18ZNNLO [105], CT18ANNLO [105], CT18XNNLO [105], MSHT20nnlo_as118 [106], NNPDF31_nnlo_as_0118_hessian [107], and NNPDF40_nnlo_as_01180_hessian [107]. The envelope of the differences between the nominal PDF and the alternative central PDFs covers the uncertainty of the PDF set choice [108].

- **Strong Coupling α_S**
  The strong coupling constant α_S is measured by combining multiple datasets. The strong coupling
  $$\alpha_S(m_Z^2) = 0.1180 \pm 0.0010 \tag{11.2}$$
  is given at the Z mass with its 68% confidence level [104]. This uncertainty is propagated to the polarization measurement via
  $$\delta^{\alpha_S}\sigma = \frac{1}{2}(\sigma(\alpha_S = 0.119) - \sigma(\alpha_S = 0.117)). \tag{11.3}$$
  The nominal value $\alpha_S(m_Z^2) = 0.1180$ used in PDF4LHC_40 is taken as the central value [104].

- **Renormalization and factorization scale variations**
  Scale variations are used to estimate the uncertainty originating from missing higher-order corrections. The renormalization scale μ_R and the factorization scale μ_F are pairwise varied. For the Sherpa sample, a pairwise 7-point variation is used $(\mu_R, \mu_F) \in \{(0.5, 0.5), (0.5, 1), (1, 0.5), (1, 1), (2, 1), (1, 2), (2, 2)\}$. The scale uncertainties are combined by their envelope [108].

- **Resummation scale**
  The resummation scale defines the upper cutoff scale for the evolution of the parton shower [108]. The MU_Q scale of Sherpa is varied between 0.5 and 2.

- **Merging scale**
  To avoid an overlap of jets simulated in the matrix element calculation and in the parton shower a merging scale CKKW is chosen. The hard jets are taken from the matrix element calculation and jets below the merging scale are simulated by the parton shower [58]. This merging scale CKKW is varied with (20^{+10}_{-5}) GeV.

The listed uncertainties are extracted from unpolarized Sherpa $l^{\pm}\nu l^{\pm}\nu jj+0,1j$ sample on particle-level. These uncertainties are then transferred to the polarized predictions used for the polarization measurement. Ideally, the transfer of uncertainties should happen in the entire phase space. Then, the uncertainties would be combined in the final distribution of the discriminator DNNs. The transfer of individual systematic variations in the full phase space could be achieved by the same multivariate reweighting approach used for the higher-order

QCD correction. However, the propagation of all systematic variations would require more than 50 individual networks.

To avoid applying this high number of DNNs, the effects of the systematic variations are studied in the combined signal and Low-m_{jj} region defined in table 8.2. A multivariate correction is unnecessary if the systematic variations only affect the shape of one specific variable. Kolmogorov Smirnov (KS) values quantify the shape effects in the different kinematic variables. The KS values are calculated for the nominal $l^{\pm}\nu l^{\pm}\nu jj + 0, 1j$ sample and its systematic variations on particle-level.

The renormalization and factorization scale variation in figure 11.1 has the dominant shape impact of all the systematic variations derived by on-the-fly weights. The internal PDF variations, the alternative PDF sets, and the α_S variations have a similar KS value distribution across the different kinematic variables but with a more minor shape impact. Thus, their KS-value distribution is moved into the appendix in figure F.1 and the focus remains on the renormalization and factorization scale variation. Since the variations are derived by on-the-fly weights the two distributions compared in the KS test contain the same events and only the event weights are varied. Thus, the KS values in figures 11.1 and F.1 are significantly lower than the expectation for two statistically independent distributions and their resulting p-values are not meaningful and therefore not discussed. However, the KS values indicate that the theory systematics from on-the-fly weights dominantly affect the m_{jj} kinematic. Thus propagating the theory uncertainties dependent on the particle-level m_{jj} is expected to cover their dominant effects on the entire phase space.

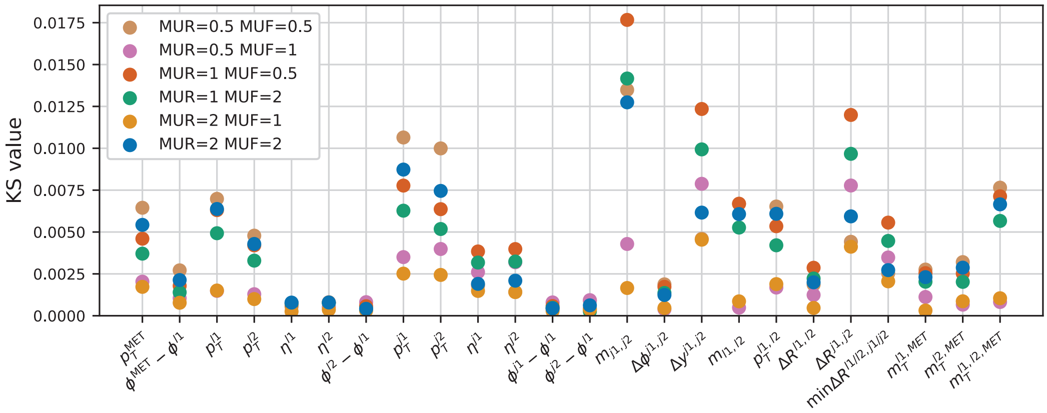


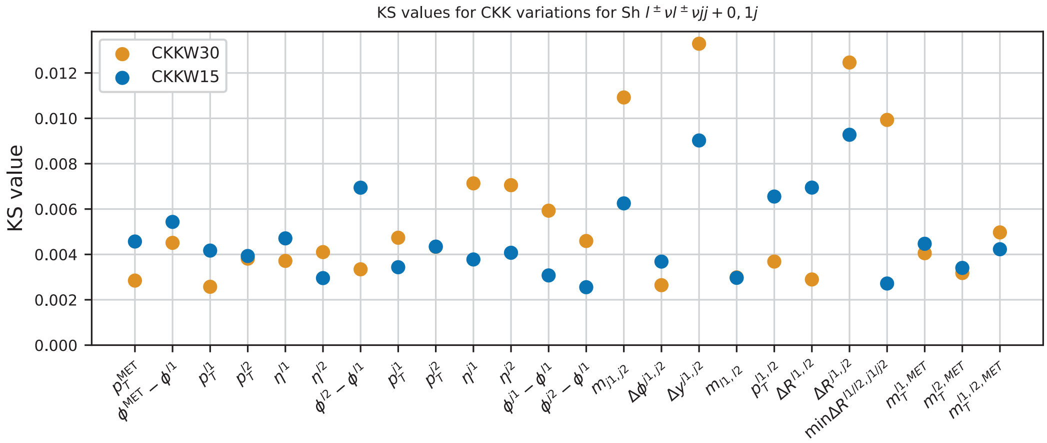
Figure 11.1: Kolmogorov Smirnov test between the nominal Sherpa $l^{\pm}\nu l^{\pm}\nu jj + 0, 1j$ sample and the corresponding renormalization and factorization scale variations. The kinematic distributions are derived on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2.

For the resummation and merging scale variations, alternative samples are produced. The resulting KS values in figure 11.2 are in the same magnitude as for the renormalization and factorization scale variations in figure 11.1. However, in contrast to figure 11.1, the resummation and merging scale variations are derived by statistically independent samples. Thus, the shown KS values correspond to almost no shape difference between the variations. This is confirmed by the corresponding p-values in figure F.2. Even though the shape effects are very small in all variables, the most affected distributions are m_{jj}, Δy_{jj}, and ΔR_{jj}. Since these variables are significantly correlated, the m_{jj} distribution is chosen for the uncertainty propagation to be consistent with the propagation of the on-the-fly weight variations.

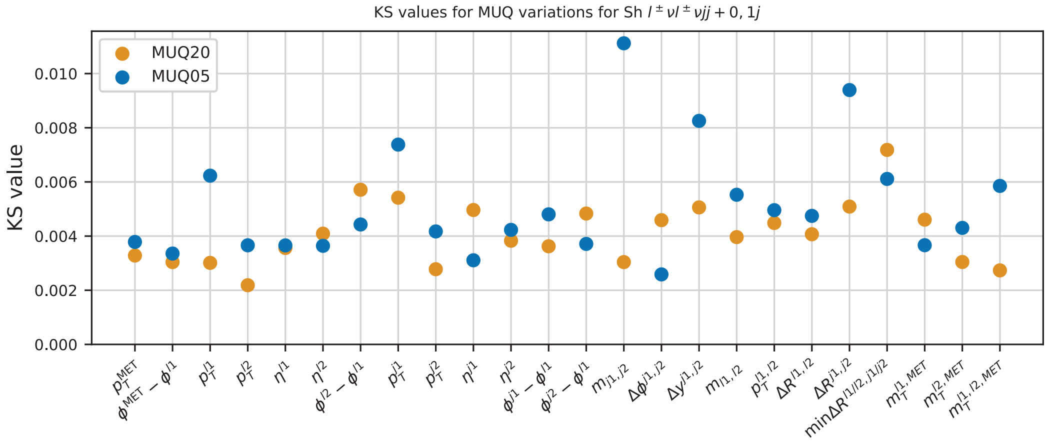
The theory uncertainties are extracted by the reweighting factor

$$r^{\text{sys}}(m_{jj}) = \frac{(l^{\pm}\nu l^{\pm}\nu jj + 0, 1j)^{\text{sys}}}{(l^{\pm}\nu l^{\pm}\nu jj + 0, 1j)^{\text{nominal}}}. \tag{11.4}$$

The correction factor is extracted in combined signal and Low-m_{jj} region defined in table 8.2 but without the m_{jj} selection criteria to cover potential migrations between the particle-level



(a) KS-values for merging scale variations



(b) KS-values for resummation scale variations

Figure 11.2: Kolmogorov Smirnov values for the comparison of the nominal SHERPA $l^\pm \nu l^\pm \nu jj + 0, 1j$ sample and the corresponding resummation and merging scale variations. The kinematic distributions are derived on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2.

and reconstruction-level phase spaces. Calculating this factor allows an early combination of the uncertainties reducing the necessary computational effort for the subsequent analysis. The uncertainties propagated by $r^{\text{sys}}(m_{jj})$ are combined via:

- Internal PDF variations: $\delta^{\text{PDF}}\sigma = \sqrt{\sum_{i=1}^{40} (\sigma_i - \sigma_0)^2}$
- Alternate PDF sets: envelope
- Strong Coupling α_S: $\delta^{\alpha_S}\sigma = \frac{1}{2}(\sigma(\alpha_S = 0.119) - \sigma(\alpha_S = 0.117))$
- Renormalization and factorization scale variations: envelope
- Resummation scale: individual handling of MUQ05 and MUQ20
- Merging scale: individual handling of CKKW15 and CKKW30

An analytic fit smoothes the resulting distributions. The functions

$$f_1^{sys}(x) = a + \sum_{i=1}^{5} c_i x^i \tag{11.5}$$

and

$$f_2^{sys}(x) = a + \sum_{i=1}^{3} c_i x^i + \sum_{i=4}^{6} \frac{c_i}{x^i} \tag{11.6}$$

are considered to model the shape effects of the combined theory uncertainty sets. The final fits are shown in figure 11.3 and their corresponding performances are listed in table 11.2. For the later application of the theory systematics only the range $150\,\text{GeV} < m_{jj} < 4000\,\text{GeV}$ is used and for m_{jj} outside this range the correction factor at corresponding edge of the m_{jj} range is applied. Thus, the fit in the first bin $100\,\text{GeV} \leq m_{jj} < 150\,\text{GeV}$ does not affect the final uncertainty calculation.

The fit function $f_1^{sys}(x)$ is chosen for the internal PDF, resummation scale, and merging scale variations. This function describes their shape very well and has a significantly more flat behavior in the first bin of $100\,\text{GeV} \leq m_{jj} < 150\,\text{GeV}$. The corresponding fit parameters are listed in table 11.3. For the alternative PDF sets, the α_S variations, and the renormalization and factorization scale variations, the function $f_2^{sys}(x)$ provides the best agreement and is therefore chosen with the fit parameters listed in table 11.4.

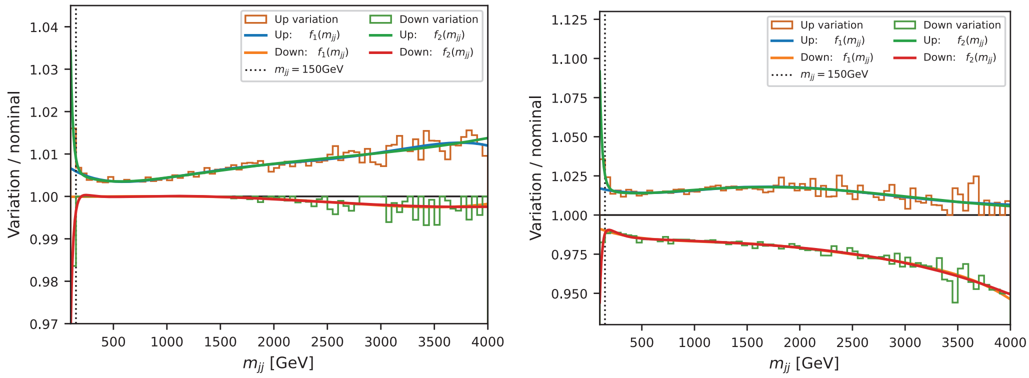
Table 11.2: The performance of the χ^2 fits of the functions in equations (11.5) and (11.6) to the combined systematic variations.

Systematic	$f_1^{sys}(x)$		$f_2^{sys}(x)$	
	$\chi^2_{\text{fit}}/\text{dof}$	p-value	$\chi^2_{\text{fit}}/\text{dof}$	p-value
CKKW15	1.39	0.016	1.36	0.022
CKKW30	1.10	0.26	1.12	0.23
MUQ05	1.28	0.056	1.24	0.079
MUQ20	1.08	0.30	1.08	0.30
PDF var down	0.033	1.00	0.029	1.00
PDF var up	0.033	1.00	0.029	1.00
alpha s down	$4.3 \cdot 10^{-3}$	1.00	$4.1 \cdot 10^{-3}$	1.00
alpha s up	$4.3 \cdot 10^{-3}$	1.00	$4.1 \cdot 10^{-3}$	1.00
alt PDF down	0.019	1.00	0.019	1.00
alt PDF up	0.040	1.00	0.039	1.00
scale down	0.083	1.00	0.044	1.00
scale up	0.24	1.00	0.21	1.00

Table 11.3: Fit parameters of $f_1^{sys}(x)$ in equation (11.5) for the combined relative uncertainties of the internal PDF, resummation scale, and the merging scale variations on particle-level Sherpa $l^\pm \nu l^\pm \nu jj + 0, 1j$.

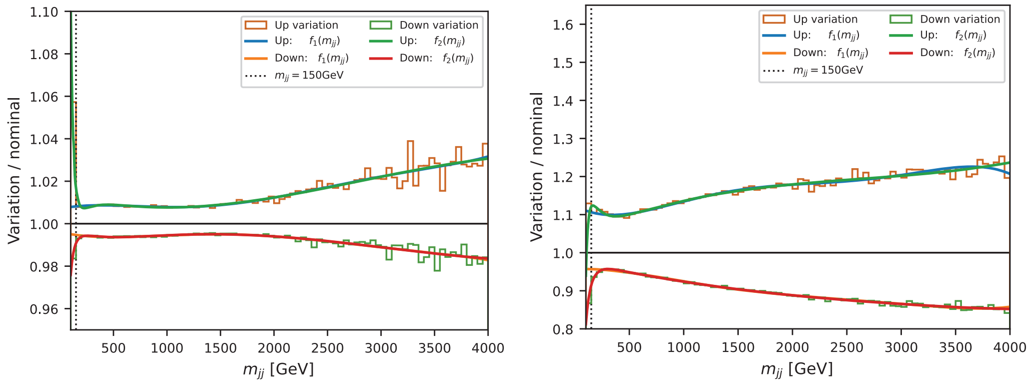
Systematic	a	c_1	c_2	c_3	c_4	c_5
CKKW30	0.9487	$7.258 \cdot 10^{-5}$	$5.828 \cdot 10^{-8}$	$-7.661 \cdot 10^{-11}$	$2.617 \cdot 10^{-14}$	$-2.878 \cdot 10^{-18}$
CKKW15	1.303	$-3.13 \cdot 10^{-4}$	$3.991 \cdot 10^{-7}$	$-2.41 \cdot 10^{-10}$	$6.573 \cdot 10^{-14}$	$-6.541 \cdot 10^{-18}$
MUQ20	0.9106	$1.258 \cdot 10^{-4}$	$-1.19 \cdot 10^{-7}$	$3.893 \cdot 10^{-11}$	$-2.691 \cdot 10^{-15}$	$-3.457 \cdot 10^{-19}$
MUQ05	1.141	$-4.555 \cdot 10^{-4}$	$6.176 \cdot 10^{-7}$	$-3.841 \cdot 10^{-10}$	$1.06 \cdot 10^{-13}$	$-1.062 \cdot 10^{-17}$
PDF var up	1.018	$3.172 \cdot 10^{-6}$	$-2.903 \cdot 10^{-9}$	$2.145 \cdot 10^{-12}$	$-2.093 \cdot 10^{-16}$	$-2.571 \cdot 10^{-21}$
PDF var down	0.9817	$-3.172 \cdot 10^{-6}$	$2.903 \cdot 10^{-9}$	$-2.145 \cdot 10^{-12}$	$2.093 \cdot 10^{-16}$	$2.571 \cdot 10^{-21}$

It is noticeable for the merging scale in figure 11.3e that both CKKW15 and CKKW30 are up-variations. This is surprising since the simulated CKKW30 sample has a lower cross-section than the nominal sample. To better understand this effect, the systematic variation of the merging scale is investigated in the loose particle-level region defined in table 8.4 without any di-jet selection. In this region, the distribution in Δy_{jj} in figure 11.4 is striking. In the range of $\Delta y_{jj} < 2$ the CKKW30 is indeed a significant down-variation. However, the $\Delta y_{jj} > 2$ selection of the signal region removes this part of the distribution and only the up-variation in the range of $\Delta y_{jj} > 2$ is left. Therefore, the CKKW30 sample results in an up-variation due to the shape dependency of this variation and the specific selection of the signal region definition.



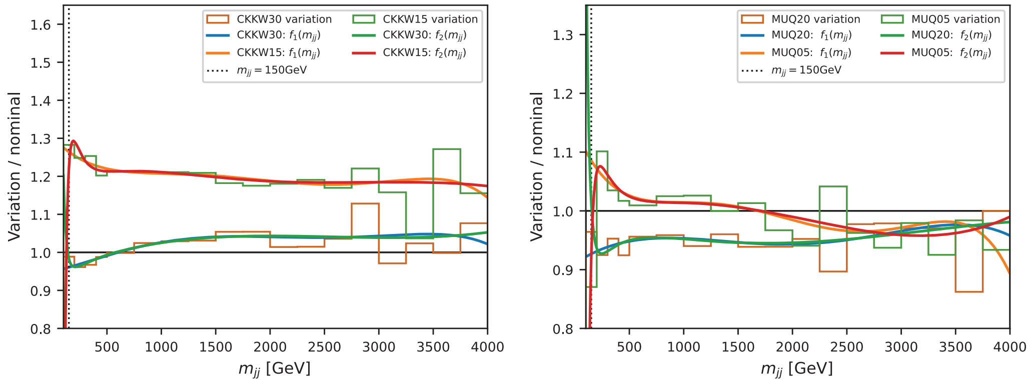
(a) α_S variations

(b) Alternate PDF sets



(c) PDF variations

(d) Renormalization and factorization scale variations



(e) Merging scale variations

(f) Resummation scale variations

Figure 11.3: The combinations of the relative systematics $r^{\text{sys}}(m_{jj})$ in equation (11.4) for particle-level Sherpa $W^{\pm}W^{\pm}jj + 0, 1j$. The analytic functions given by equations (11.5) and (11.6) are fit to the combined relative uncertainties. The kinematic distributions are derived on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2 but without the m_{jj} selection criteria.

Table 11.4: Fit parameters of $f_2^{sys}(x)$ in equation (11.6) for the combined relative uncertainties of the alternative PDF sets, the α_S variations, and the renormalization and factorization scale variations on particle-level Sherpa $W^\pm W^\pm jj + 0, 1j$.

Systematic	a	c_1	c_2	c_3	c_4	c_5	c_6
alpha s up	0.9811	$1.638 \cdot 10^{-5}$	$-4.452 \cdot 10^{-9}$	$4.745 \cdot 10^{-13}$	9.35	-1912	$1.527 \cdot 10^{5}$
alpha s down	1.019	$-1.638 \cdot 10^{-5}$	$4.453 \cdot 10^{-9}$	$-4.747 \cdot 10^{-13}$	-9.352	1912	$-1.527 \cdot 10^{5}$
alt PDF up	0.977	$3.827 \cdot 10^{-5}$	$-1.367 \cdot 10^{-8}$	$1.411 \cdot 10^{-12}$	16.04	-3322	$2.832 \cdot 10^{5}$
alt PDF down	0.9947	$-8.673 \cdot 10^{-6}$	$2.891 \cdot 10^{-9}$	$-8.622 \cdot 10^{-13}$	-6.738	2313	$-2.138 \cdot 10^{5}$
scale up	1.086	$1.077 \cdot 10^{-4}$	$-3.573 \cdot 10^{-8}$	$4.699 \cdot 10^{-12}$	-39.44	$1.438 \cdot 10^{4}$	$-1.201 \cdot 10^{6}$
scale down	0.9333	$-2.992 \cdot 10^{-5}$	$-1.934 \cdot 10^{-12}$	$5.139 \cdot 10^{-13}$	24.14	-4659	$1.089 \cdot 10^{5}$

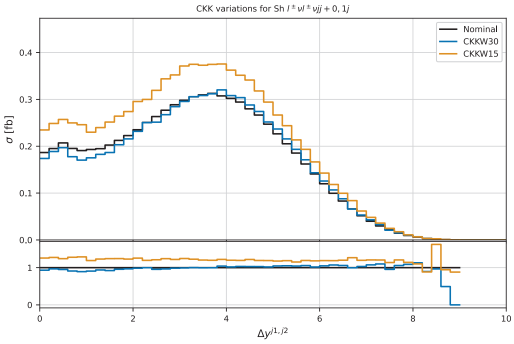


Figure 11.4: The merging scale variations compared to the nominal $l^\pm \nu l^\pm \nu jj + 0, 1j$ distribution in Δy_{jj} on particle-level. The events are selected by the loose particle-level region defined in table 8.4.

The smoothed systematic uncertainties are propagated to the reconstruction-level Sherpa and MadGraph samples described in section 7.1. The fit function $f^{sys}(m_{jj})$ is applied according to the particle-level m_{jj} of the corresponding event to be consistent with the particle-level extraction of the systematic. The most dominant impact is expected from the scale systematics shown in figures 11.3d to 11.3f. Their relative uncertainty is in the order of 10% to 20%, which significantly limits the accuracy of the theory prediction. The application of the scale uncertainties is further discussed in section 12.4.

11.2.2 NLO Correction Uncertainty of $W^\pm W^\pm jj$-EW

The approximate NLO-QCD and NLO-EW correction of the $W^\pm W^\pm jj$-EW polarization and their corresponding uncertainties are described in chapter 8. Two different application methods are implemented for each of the higher-order corrections. The difference between the two correction methods is taken as symmetrized application uncertainty. The QCD and electroweak correction combination is described in section 8.3. The difference between the two approaches of multiplying and adding the corrections represents an additional modeling uncertainty.

11.2.3 Theory Uncertainties of the $W^\pm Z$ Background

For the $W^\pm Z$ samples, internal PDF variations, alternative PDF sets, and renormalization and factorization scale variations are used to estimate the theory uncertainty. These uncertainties are directly provided by on-the-fly weights within the sample. Thus, the uncertainties can be combined in the final polarization DNN score without any further study of their impact on the remaining phase space. The NNPDF30_nnlo_as_0118 PDF set was used for the sample production with 100 internal PDF variations. The combined uncertainty of this PDF set is given by the standard deviation[1]

$$\delta^{\mathrm{PDF}}\sigma = \sqrt{\frac{1}{100}\sum_{i=1}^{100}(\sigma_i - \sigma_0)}. \tag{11.7}$$

The resulting combined internal PDF uncertainty is combined with the alternative PDF sets, CT14 and MMHT2014, via an envelope. For the renormalization and factorization scale variations, the 7-point variation of (μ_R, μ_F) is used similarly as for the polarization samples. The resulting scale variations are combined via an envelope [108].

11.2.4 Reweighting Uncertainty of $W^{\pm}Z$ Background

The m_{jj} shape of the $W^{\pm}Z$-QCD background is corrected by an analytic function derived in the $W^{\pm}Z$ control region. This procedure is described in section 7.2.1. The corresponding PDF, α_S, theory scale, and statistical uncertainties contribute to the modeling uncertainty of the $W^{\pm}Z$-QCD background.

---

[1] This differs from the Hessian PDF set of polarization samples which is summed up quadratically without the division by the number of variations.

Chapter 12

Statistical Evaluation and Results

In the previous chapters, the signal and background modeling were described, the signal was corrected for higher-order effects, neural networks were trained as discriminating variables, and the relevant uncertainties were estimated. Thus, all the analysis parts are ready to start the statistical evaluation of the longitudinal polarization.

This thesis aims to measure the LL and LX polarization state in the partonic and WW center-of-mass frame. The transverse polarizations TX and TT are considered as background of this measurement. Since the contribution of the transverse polarization state is only known from the same simulation as the longitudinal polarization state, its normalization is considered as an additional free parameter for the likelihood fit described in section 12.1. The additional free parameter motivates the combined application procedure of two networks described in section 12.2. For this application procedure, the binning of the resulting histograms is optimized in section 12.3 to maximize the expected significance. Based on the resulting histograms, the application of the scale uncertainties is discussed in section 12.4. The scale uncertainties are the dominant theory uncertainties and their application strategy is crucial for the sensitivity of the polarization measurement. Finally, the expected and observed fit results are presented in sections 12.5 and 12.6 and further discussed in sections 12.7 to 12.9.

Based on the implementation of the higher-order corrections, discriminator networks, and theory uncertainties presented in the previous chapters, the ATLAS $W^{\pm}W^{\pm}jj$-EW polarization analysis group [109] has processed the Standard Model prediction to create the fit input and a baseline fit configuration. The set of uncertainties is pruned to consider only uncertainties which affect the background contributions by more than 0.1%. The pruned list of systematic uncertainties and their corresponding category and sub-category are listed in table G.1.

12.1 Introduction of the Likelihood Fit

A profile likelihood fit extracts the significance and normalization of the longitudinal polarization state fit. The likelihood $\mathcal{L}$ compares the measured data with the prediction considering free-floating parameters and uncertainties. Since the distinction between bins in different histograms and regions is arbitrary, the term bin in the following refers to bins across all histograms and regions. The measurement in the ATLAS detector leads to n_i measured events in the bin i. The number of predicted events

$$\begin{aligned} m_i(\mu_L, \mu_T, \mu_{\text{WZ-QCD}}, \gamma_i, \boldsymbol{\theta}) =& \mu_L s_{L,i}(\boldsymbol{\theta}) + \mu_T \gamma_i s_{T,i}(\boldsymbol{\theta}) + \mu_{\text{WZ-QCD;i}} \gamma_i b_{\text{WZ-QCD}}(\boldsymbol{\theta}) \\ &+ \sum_{p \in B \setminus \{\text{WZ-QCD}\}} \gamma_i b_{p,i}(\boldsymbol{\theta}) \end{aligned} \tag{12.1}$$

depends on the number of predicted longitudinally polarized events s_L (LL/LX), transversally polarized events s_T (TX/TT), and background events b_p. The set of different background processes is given by B. The number of expected events depends on the chosen normalization μ_L, μ_T, and $\mu_{\text{WZ-QCD}}$ of the longitudinal $W^{\pm}W^{\pm}jj$-EW, transversal $W^{\pm}W^{\pm}jj$-EW, and $W^{\pm}Z$-QCD contribution. The statistical and systematic uncertainties of the prediction are considered by the nuisance parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$.

The likelihood fit is performed via the TRExFitter [110]. This tool is a wrapper around the HistFactory [111]. The systematic uncertainty on the nominal prediction $\sigma_{i,p}$ of the process p in the bin i is provided by the up and down variation, $\sigma^{+}_{i,p}$ and $\sigma^{-}_{i,p}$. The TRExFitter separates the normalization and shape impact of the nuisance parameter θ to avoid negative predictions in the fit. The shape effects $\sigma^{shape,+}_{i,p}$ and $\sigma^{shape,-}_{i,p}$ are introduced by the linear interpolation

$$\sigma_{i,p}(\theta) = \begin{cases} \sigma^{0}_{i,p} + \theta(\sigma^{shape,+}_{i,p} - \sigma^{0}_{i,p}) & \theta \geq 0 \\ \sigma^{0}_{i,p} + \theta(\sigma^{0}_{i,p} - \sigma^{shape,-}_{i,p}) & \theta < 0 \end{cases}. \tag{12.2}$$

For the normalization uncertainty the exponential interpolation

$$\eta_p(\theta) = \begin{cases} (\sigma^{+}_{p}/\sigma^{0}_{p})^{\theta} & \theta \geq 0 \\ (\sigma^{-}_{p}/\sigma^{0}_{p})^{\theta} & \theta < 0 \end{cases} \tag{12.3}$$

is used to avoid negative numbers of events. The resulting systematic uncertainties are introduced into the likelihood via a Gaussian constraint $f_{\text{Gauss}}(\theta_j)$. Thus, the shape and normalization of the processes can be pulled for the fit in consideration of the known constraints.

The statistical uncertainty of the Standard Model prediction is considered by the nuisance parameters $\boldsymbol{\gamma}$. The total predicted number of events in each bin is affected by one $\boldsymbol{\gamma}$ normalization factor. The γ_i is constrained by the relative statistical uncertainty δ^{stat}_{i} of the combined prediction in the corresponding bin. This constraint is embedded in a Poisson distribution of $1/(\delta^{stat}_{i})^2$ events to provide the correct relative uncertainty.

The resulting likelihood

$$\begin{aligned} \mathcal{L}\left(\mu_L, \mu_T, \mu_{\text{WZ-QCD}}, \boldsymbol{\gamma}, \boldsymbol{\theta}\right) = \prod_{i}^{N_{\text{bins}}} & f_{\text{Pois}}(n_i, m_i(\mu_L, \mu_T, \mu_{\text{WZ-QCD}}, \gamma_i, \boldsymbol{\theta})) \\ & \times f_{\text{Pois}}\left(\frac{1}{(\delta^{stat}_{i})^2} \middle| \frac{\gamma_i}{(\delta^{stat}_{i})^2}\right) \times \prod_{j}^{N_{\text{nuis}}} f_{\text{Gauss}}(\theta_j) \end{aligned} \tag{12.4}$$

is a combination of the agreement with measured data within a Poisson statistic and the constraints of the nuisance parameter. Maximizing the likelihood $\mathcal{L}$ provides the expectation values of the free floating normalizations $\hat{\mu}_L$, $\hat{\mu}_T$, and $\hat{\mu}_{\text{WZ-QCD}}$ and the nuisance parameters $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\gamma}}$.

The likelihood fit is used to determine the significance of the longitudinal polarization state LL and LX. The discovery significance is calculated against the null hypothesis that the longitudinal polarization state, LL or LX, does not exist. Therefore, the unconditional fit $\mathcal{L}(\hat{\mu}_L, \hat{\mu}_T, \hat{\mu}_{\text{WZ-QCD}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}})$ is compared to the background-only fit $\mathcal{L}(0, \hat{\hat{\mu}}_T, \hat{\hat{\mu}}_{\text{WZ-QCD}}, \hat{\hat{\boldsymbol{\gamma}}}, \hat{\hat{\boldsymbol{\theta}}})$ with the longitudinal signal strength fixed to $\mu_L = 0$. These two likelihoods are combined in the logarithmic likelihood ratio

$$q_{\mu} = \begin{cases} -2\ln\left(\dfrac{\mathcal{L}\left(\mu_L, \hat{\hat{\mu}}_T, \hat{\hat{\mu}}_{\text{WZ-QCD}}, \hat{\hat{\boldsymbol{\gamma}}}, \hat{\hat{\boldsymbol{\theta}}}\right)}{\mathcal{L}\left(\hat{\mu}_L, \hat{\mu}_T, \hat{\mu}_{\text{WZ-QCD}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}\right)}\right) & \text{for } \hat{\mu}_L \geq 0 \\ 0 & \text{for } \hat{\mu}_L < 0 \end{cases} \tag{12.5}$$

to define the test statistic q_{μ}. As described in Ref. [112], the discovery significance Z_0 of a positive signal can be approximated by the asymptotic formula

$$Z_0 = \sqrt{q_0}. \tag{12.6}$$

12.2 Network Application Procedure

In chapter 9, multiple networks are trained for specific classification purposes. Firstly, four 5-fold DNN$_{pol}$ sets are trained to provide a sensitive variable for the measurements of the $W^{\pm}W^{\pm}jj$-EW polarization states LL and LX in the partonic and WW center-of-mass frame. Secondly, one 5-fold DNN$_{W^{\pm}W^{\pm}}$ set separates the $W^{\pm}W^{\pm}jj$-EW process from the remaining backgrounds. Dividing the separation task into two outputs has a significant advantage over only one network output, separating the longitudinal polarization state from the transverse state and the other backgrounds. The two network outputs can be interpreted as classification tuple $(score_{pol}, score_{W^{\pm}W^{\pm}})$ leading to a two-dimensional histogram. The different regions of the two-dimensional histogram for the LX measurement are visualized in figure 12.1. The region close to $(1,1)$ is the purest in the longitudinal polarization. The region around $(1,0)$ is dominated by the transverse polarization and significantly constrains the free-floating normalization of the transverse state. For lower DNN$_{W^{\pm}W^{\pm}}$ scores, the contamination from other backgrounds increases, providing an additional constraint on their contribution. Therefore, the chosen application procedure provides a region pure in the longitudinal signal and additionally constrains the different backgrounds. The next section demonstrates how the two-dimensional histogram in $(score_{pol}, score_{W^{\pm}W^{\pm}})$ can be represented by a set of one-dimensional histograms.

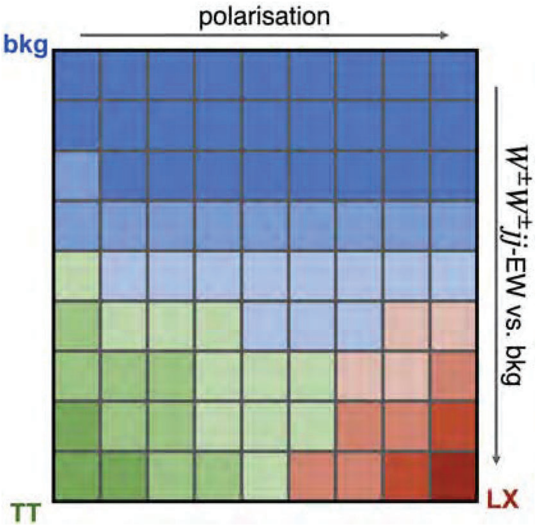


Figure 12.1: The DNN$_{W^{\pm}W^{\pm}}$ output is combined with the output of the polarization DNN$_{LX}$ to create a two-dimensional output space. The colors indicate the dominant proportion of either background contributions (blue), the TT polarization state (green) or the LL polarization state (green).

12.3 Optimization of the Histogram Binning

The application of the two multivariate outputs to gain a two-dimensional distribution was already used in the CMS $W^{\pm}W^{\pm}jj$-EW polarization analysis in Ref. [3]. For the binning of the corresponding histograms, the CMS analysis has chosen a 5x5 binning with similar number of longitudinally polarized events. The polarization analysis in this thesis improves this approach by choosing the binning with the best expected significance. However, the number of possible bin edges to test is high and performing the fit by the TRExFitter requires significant computation time. Thus, a computationally fast estimation of the significance is introduced.

12.3.1 Significance estimation

The expected significance is calculated using the Standard Model prediction as data distribution. The expected data

$$n_i = s_{L,i} + s_{T,i} + \sum_{p\in B} b_{p,i} \tag{12.7}$$

in the i-th bin is therefore the sum of the longitudinal polarization state $s_{L,i}$, the transverse polarization state $s_{T,i}$, and the different background contributions $b_{p,i}$. The resulting distribution is called Asimov data.

In general, various nuisance parameters must be considered in the fit to calculate the significance. Since this would result in a very complex and computationally expensive fit, the uncertainties of the predicted contributions are highly simplified. The predicted contributions are considered to have only one main normalization uncertainty each. This greatly reduces the complexity of the fit but also distorts the estimated significance because shape uncertainties and anti-correlations are no longer taken into account. However, since this estimated significance is only used for bin optimization, whereby the significance has to be calculated very frequently and thus as quickly as possible, this estimation is accepted.

The normalization uncertainties are introduced by the nuisance parameters $\boldsymbol{\omega}$ and their relative uncertainties $\Delta\omega_j$ for the corresponding contribution. The significance is calculated for the longitudinal polarization strength μ_L with a free-floating transverse polarization strength μ_T. The normalization of the $W^{\pm}Z$-QCD is fixed to the 0.74 expected from the $W^{\pm}Z$ control region in section 12.5. Thus, the number of predicted events

$$m_i(\mu_L, \mu_T, \boldsymbol{\omega}) := \begin{cases} \mu_L \omega_0 s_{\mathrm{LL};i} + \mu_T\left(\omega_1 s_{\mathrm{LT};i} + \omega_2 s_{\mathrm{TT};i}\right) + \sum_j^{N_{\mathrm{bkg}}} \omega_{j+2} b_{j,i} & \text{for LL meas.} \\ \mu_L\left(\omega_0 s_{\mathrm{LL};i} + \omega_1 s_{\mathrm{LT};i}\right) + \mu_T \omega_2 s_{\mathrm{TT};i} + \sum_j^{N_{\mathrm{bkg}}} \omega_{j+2} b_{j,i} & \text{for LX meas.} \end{cases} \tag{12.8}$$

depends on the parameters μ_L, μ_T, and $\boldsymbol{\omega}$. The contributions of the polarization states LL, TL, and TT in each bin i are given by $s_{\mathrm{LL};i}$, $s_{\mathrm{LT};i}$, and $s_{\mathrm{TT};i}$.

The significance is estimated by a logarithmic likelihood ratio fit for N_{bins} bins. The number of events in each is expected to be Poisson distributed. For the nuisance parameters $\boldsymbol{\omega}$, also a Poisson distribution is chosen to avoid negative normalizations in the fit. In order to gain the correct relative uncertainty for the normalization parameters, the corresponding Poisson distributions are calculated for $1/\Delta\omega_j^2$. The fact that this can lead to non-integer values for the Poisson distributions can be neglected for the time being, as the Poisson distributions do not need to be explicitly calculated to estimate the significance. For the significance estimation, non-zero values are assumed for $\Delta\omega_j$, n_i, and m_i. The non-zero event numbers n_i and m_i are ensured for all bins by an initial bin merging described in section 12.3.2.

In order to avoid the explicit calculation of the Poisson distributions[1] the likelihood

$$\begin{aligned} \mathcal{L}\left(\mu_L, \mu_T, \boldsymbol{\omega}\right) &= \prod_i^{N_{\mathrm{bins}}} f_{\mathrm{Poisson}}(n_i|m_i) \prod_j^{3+N_{\mathrm{bkg}}} f_{\mathrm{Poisson}}\left(\frac{1}{{\Delta\omega_j}^2}\middle|\frac{\omega_j}{{\Delta\omega_j}^2}\right) \\ &= \prod_i^{N_{\mathrm{bins}}} \frac{m_i^{n_i}}{n_i!} e^{-m_i} \prod_j^{3+N_{\mathrm{bkg}}} \frac{\left(\frac{\omega_j}{{\Delta\omega_j}^2}\right)^{\frac{1}{{\Delta\omega_j}^2}}}{\frac{1}{{\Delta\omega_j}^2}!} e^{-\frac{\omega_j}{{\Delta\omega_j}^2}} \end{aligned} \tag{12.9}$$

is not used directly to estimate the nuisance parameters. The use of the log-likelihood

$$\ln\left(\mathcal{L}\left(\mu_L, \mu_T, \boldsymbol{\omega}\right)\right) = \sum_i^{N_{\mathrm{bins}}} \left(n_i \ln(m_i) - m_i\right) + \sum_j^{3+N_{\mathrm{bkg}}} \left(\frac{1}{{\Delta\omega_j}^2} \ln\left(\frac{\omega_j}{{\Delta\omega_j}^2}\right) - \frac{\omega_j}{{\Delta\omega_j}^2}\right) + c(\boldsymbol{n}, \boldsymbol{\Delta\omega}) \tag{12.10}$$

simplifies the maximum likelihood fit since a constant summand containing the factorials can be split off, which no longer needs to be explicitly calculated.

---

[1] The calculation would be possible using the gamma function but is not necessary as shown in the following.

Table 12.1: Relative normalization uncertainties for different processes.

Process	Relative Normalization Uncertainty
$W^{\pm}W^{\pm}jj$-EW LL WW-cmf	0.103
$W^{\pm}W^{\pm}jj$-EW TL WW-cmf	0.103
$W^{\pm}W^{\pm}jj$-EW TT WW-cmf	0.045
$W^{\pm}W^{\pm}jj$-EW LL pp-cmf	0.107
$W^{\pm}W^{\pm}jj$-EW TL pp-cmf	0.105
$W^{\pm}W^{\pm}jj$-EW TT pp-cmf	0.047
$W^{\pm}W^{\pm}jj$-QCD	0.043
$W^{\pm}W^{\pm}jj$-INT	0.026
$W^{\pm}Z$-EW	0.029
$W^{\pm}Z$-QCD	0.315
$W^{\pm}Z$-INT	0.399
Top and triboson	0.065
ZZ	0.042
$V\gamma$	0.163
Non-prompt	0.203
Chargeflip	0.365

To calculate the significance, one compares the likelihood fit to the Asimov data $\mathcal{L}(\hat{\mu}_L, \hat{\mu}_T, \hat{\boldsymbol{\omega}}) = \mathcal{L}(1, 1, \mathbf{1})$ with the likelihood $\mathcal{L}(0, \hat{\hat{\mu}}_T, \hat{\hat{\boldsymbol{\omega}}})$ fit for $\mu_L = 0$. $\hat{\hat{\mu}}_T$ and $\hat{\hat{\boldsymbol{\omega}}}$ are calculated by minimizing equation 12.10. The resulting likelihoods can be combined into the likelihood ratio

$$\frac{\mathcal{L}\left(0, \hat{\hat{\mu}}_T, \hat{\hat{\boldsymbol{\omega}}}\right)}{\mathcal{L}\left(1, 1, \mathbf{1}\right)} = \prod_{i}^{N_{\text{bins}}} \left(\frac{m_i}{n_i}\right)^{n_i} e^{n_i - m_i} \prod_{j}^{3+N_{\text{bkg}}} \left(\frac{\left(\frac{\omega_j}{\Delta\omega_j{}^2}\right)}{\left(\frac{1}{\Delta\omega_j{}^2}\right)}\right)^{\frac{1}{\Delta\omega_j{}^2}} e^{\frac{1}{\Delta\omega_j{}^2} - \frac{\omega_j}{\Delta\omega_j{}^2}}. \tag{12.11}$$

By the corresponding log-likelihood ratio

$$\ln\left(\frac{\mathcal{L}\left(0, \hat{\hat{\mu}}_T, \hat{\hat{\boldsymbol{\omega}}}\right)}{\mathcal{L}\left(1, 1, \mathbf{1}\right)}\right) = \sum_{i}^{N_{\text{bins}}} n_i \left(\ln\left(\frac{m_i}{n_i}\right) + 1 - \frac{m_i}{n_i}\right) + \sum_{j}^{3+N_{\text{bkg}}} \frac{1}{\Delta\omega_j{}^2} \left(\ln\left(\omega_j\right) + 1 - \omega_j\right) \tag{12.12}$$

the test statistic

$$q_{0,A} = -2 \ln\left(\frac{\mathcal{L}\left(0, \hat{\hat{\mu}}_T, \hat{\hat{\boldsymbol{\omega}}}\right)}{\mathcal{L}\left(1, 1, \mathbf{1}\right)}\right) \tag{12.13}$$

is defined. Following the asymptotic formula of Ref. [112] for Asimov fits, the median significance of a discovery can be directly derived by the test statistic $q_{0,A}$:

$$\text{median}[Z_0|\mu_L = 1] = \sqrt{q_{0,A}} \tag{12.14}$$

Since the number of nuisance parameters is significantly reduced, the significance estimation is much faster than the nominal fitting strategy with TRExFitter. Therefore, a fast estimate of the significance is used to optimize the bin edges. The relative normalization uncertainties $\boldsymbol{\Delta\omega}$ listed in table 12.1 are estimated by the total uncertainties on the predicted number of events in the signal region. These numbers are based on an intermediate status of the polarization analysis and are therefore not entirely consistent with the uncertainties in the final statistical evaluation in sections 12.5 and 12.6. However, the uncertainty updates primarily affect the longitudinal polarization states and are of minor importance for the significance estimation due to the free parameters μ_L and μ_T.

12.3.2 Rebinning

The starting point of the bin optimization is a fine equidistant binning of the two-dimensional histogram with 20 bins along the $W^{\pm}W^{\pm}jj$-EW vs. background classification axis and 20 bins for the polarization classification.

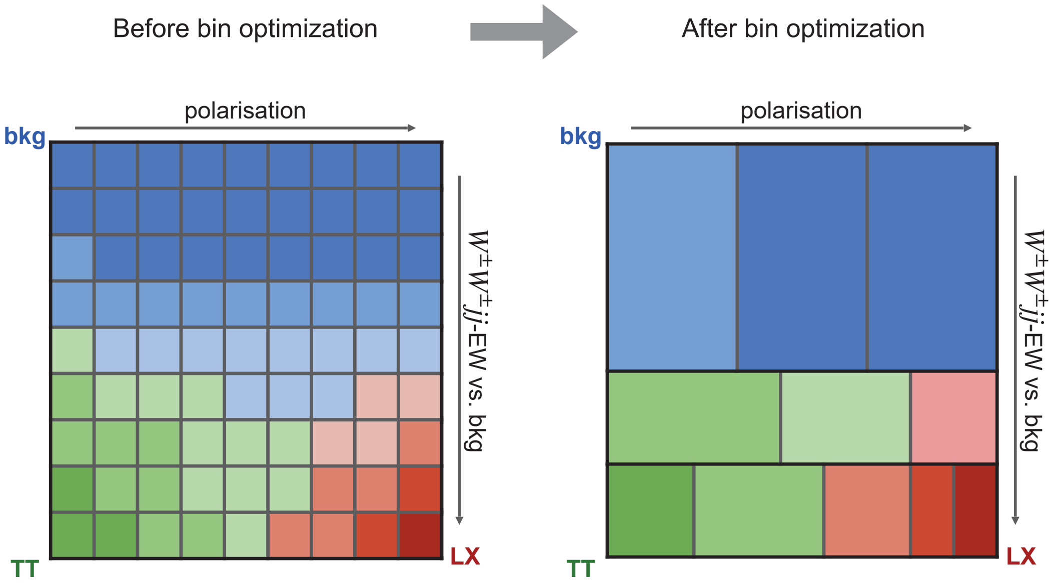


Figure 12.2: Visualization of the rebinning strategy. The two-dimensional histogram is split along the $\text{DNN}_{W^\pm W^\pm}$ output into several histograms. Each resulting histogram has specific optimized binning.

The rebinning procedure is visualized in figure 12.2. The two-dimensional histogram is split into several one-dimensional histograms used for the polarization measurement. This is done by splitting the histogram along the $W^\pm W^\pm jj$-EW vs. background axis at at least one value of the $\text{DNN}_{W^\pm W^\pm}$ classification. Each resulting histogram is merged along the $\text{DNN}_{W^\pm W^\pm}$ axis to a one-dimensional histogram. This procedure leads to several histograms of the polarization DNN classification with increasing $W^\pm W^\pm jj$-EW purity.

In order to measure the significance of the polarization, a minimal number of events per histogram bin is required. The fine bin edges of each histogram are merged until at least ten events originating from background or transverse polarized $W^\pm W^\pm jj$-EW contributions can be expected in each bin. Starting on one side of each one-dimensional histogram, a new bin edge is only accepted when the minimum number of events is reached and a further extension of the current bin would not increase the estimated significance. This merging is done for each histogram once from left to right and right to left, and the bin edges with the highest estimated significance are used.

Finally, to minimize the number of bins considered in the fit, the bin edges having the smallest effect on the expected significance are merged. This is performed for all one-dimensional histograms until each has at least three bins and the expected overall significance has been reduced by a maximum of 0.01 compared to the non-merged histograms.

This merging procedure is done for all possible splits along the $W^\pm W^\pm jj$-EW vs. background classification axis resulting in one-dimensional histograms with at least 50 events originating from background or transverse polarized $W^\pm W^\pm jj$-EW processes. The number of splits is increased until the estimated significance no longer improves.

12.3.3 Optimization Results

Different numbers of splits are listed in table 12.2 together with their estimated significance without consideration of shape uncertainties. Since no major significance increase is seen for using more than two splits, each two-dimensional histogram will be split into three one-dimensional histograms. The chosen splits and bin edges for the likelihood fit are given in table 12.3. When switching from the partonic to the WW reference frame, there are only minor changes to the binning. Differences occur primarily between the LL and LX measurements due to the changed impact of the TL polarization state. For all polarization measurements, the number of bins increases with increasing $\text{DNN}_{W^\pm W^\pm}$ score due to the higher $W^{\pm}W^{\pm}jj$-EW purity. Thus, the histogram with the highest $\text{DNN}_{W^{\pm}W^{\pm}}$ score is expected to have the most significant impact on the polarization measurement in sections 12.5 and 12.6.

Table 12.2: Estimated significance (without consideration of shape uncertainties) for different number of splits into one-dimensional histograms. The optimization algorithm is described in section 12.3

NN polarization classification	Number of splits in $\text{DNN}_{W^{\pm}W^{\pm}}$	Estimated significance
LL in WW-cmf	0	1.2457
	1	1.3627
	2	1.3766
	3	1.3737
LX in WW-cmf	0	3.3815
	1	4.2920
	2	4.4050
	3	4.3910
LL in pp-cmf	0	0.6518
	1	0.8309
	2	0.8399
	3	0.8362
LX in pp-cmf	0	2.5987
	1	3.7856
	2	3.8435
	3	3.8638

Table 12.3: The optimized splits in the $\text{DNN}_{W^{\pm}W^{\pm}}$ classification and the corresponding bin edges in polarization classification.

Polarization	Split in $\text{DNN}_{W^{\pm}W^{\pm}}$	Binning in polarization classification
LL in WW-cmf	$[0.0, 0.2, 0.6, 1.0]$	$[0.0, 0.4, 0.7, 1.0]$ $[0.0, 0.5, 0.8, 1.0]$ $[0.0, 0.4, 0.6, 0.7, 0.8, 1.0]$
LX in WW-cmf	$[0.0, 0.3, 0.7, 1.0]$	$[0.0, 0.3, 0.7, 1.0]$ $[0.0, 0.3, 0.5, 0.6, 0.7, 0.8, 1.0]$ $[0.0, 0.15, 0.35, 0.45, 0.55, 0.7, 1.0]$
LL in pp-cmf	$[0.0, 0.15, 0.5, 1.0]$	$[0.0, 0.55, 0.8, 1.0]$ $[0.0, 0.7, 0.8, 1.0]$ $[0.0, 0.4, 0.6, 0.7, 0.8, 1.0]$
LX in pp-cmf	$[0.0, 0.25, 0.6, 1.0]$	$[0.0, 0.45, 0.6, 1.0]$ $[0.0, 0.25, 0.4, 0.55, 0.75, 1.0]$ $[0.0, 0.15, 0.3, 0.4, 0.5, 0.6, 0.7, 1.0]$

12.4 Application of Scale Uncertainties

The dominant background of the longitudinal polarization measurement is the transverse polarization state. Thus, the uncertainties of the polarization states can significantly impact the likelihood fit. The highest uncertainties on the polarized contributions are the renormalization and factorization, merging, and resummation scale uncertainties described in section 11.2.1. Especially the renormalization and factorization can lead to up variations of about 20% and down variations of about 10%. As long as these uncertainties mainly affect the normalization of the $W^{\pm}W^{\pm}jj$-EW signal and not the shape, the effects on the significance of the polarization measurement should be small. Variations in the signal scaling are primarily covered by the free-floating fit parameter μ_T. However, variations in the shape of the transverse polarization are more critical. The significance will drop if the transverse polarization can vary so that its shape becomes more similar to the longitudinal polarization. Unfortunately, this is the case for the systematic scale uncertainties. According to the recommendations in Ref. [113], the theory allows every reasonable prediction within the envelopes of the scale uncertainties. Therefore, a detailed study is done in the following to understand the shape effects originating from the scale uncertainties. First, section 12.4.1 evaluates which shape effects from scale uncertainties can be expected in different kinematics. Then, in section 12.4.2, the expected impact on the output distributions of the neural networks is estimated. The resulting correlation scheme of the scale uncertainties is described in section 12.4.3.

12.4.1 Kinematic Impact of Scale Systematics

As described in chapter 4 and section 11.2.1, the scale uncertainties cover missing higher-order QCD effects and uncertainties in the showering evolution of jets. Therefore, one can assume that the resulting shape effects occur mainly within the jet phase space. This assumption is supported by the experience gained from the higher-order QCD corrections in section 8.1 and the study of the shape impact of theory uncertainties in section 11.2.1:
The correction for approximate NLO-QCD+PS effects studied the difference between Sherpa $W^{\pm}W^{\pm}jj$ and $l^{\pm}l^{\pm}\nu\nu jj+0,1j$ on particle-level. The Kolmogorov Smirnov values in figure 12.3 summarizes their shape differences in the particle-level phase space of the signal region and Low-m_{jj} control. This figure indicates that the studied higher-order QCD effects in section 8.1 dominantly affect the jet phase space.
The renormalization and factorization scale are supposed to cover the higher-order QCD effects that are not covered by the implemented correction. The Kolmogorov Smirnov values of the individual 7-point variations in figure 11.1 indicate that this uncertainty also affects dominantly the shapes in the jet phase space. Although not every possible theory form is covered by the 7-point variation, they indicate where the shape effects have the most significant impact. Therefore, the assumption that scale uncertainties mainly affect jet kinematics is very reasonable.

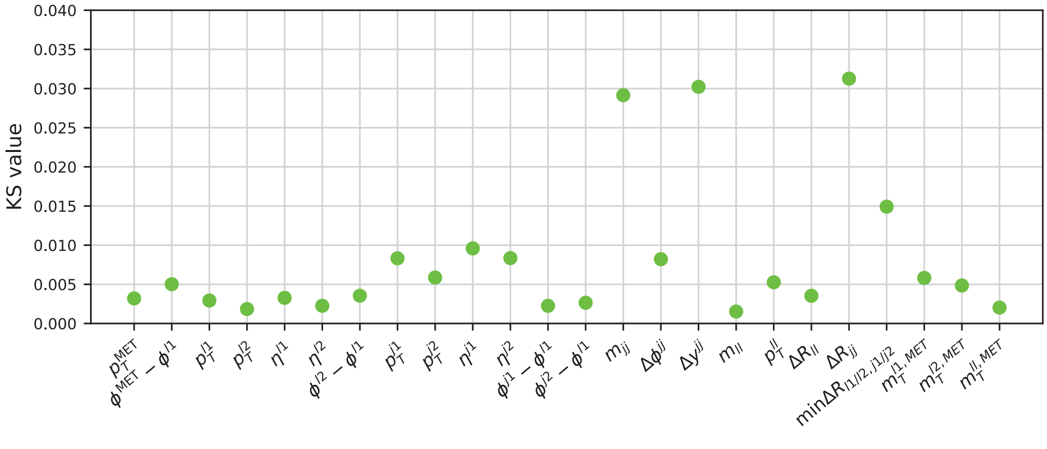


Figure 12.3: Comparison of Sherpa $W^{\pm}W^{\pm}jj$ and $l^{\pm}l^{\pm}\nu\nu jj+0,1j$ at particle-level in the phase space defined in table 8.2.

12.4.2 Maximal Shape Effects in DNN Score Distribution

Once the effects of scale uncertainties in the kinematic phase space have been specified to be dominant in the jet kinematics, it is necessary to clarify what effects this might have on the DNN distributions. In section 10.3, the individual decisions of the neural networks are studied to understand their dependency on the kinematic network input. This study has already shown that the decisions of the $\mathrm{DNN}_{W^{\pm}W^{\pm}}$ rely much more on the jet kinematics than the decisions of the polarization networks. Therefore, the polarization decision is expected to be less correlated with the jet kinematics. This is proven by calculating the Pearson correlation values between the DNN scores of the $W^\pm W^\pm jj$-EW signal and the corresponding event kinematics. The resulting Pearson correlations in figure 12.4 show that $\mathrm{DNN}_{W^\pm W^\pm}$ has significantly higher correlations in the jet phase space. The polarization networks are significantly more correlated with the kinematics in the lepton phase space and the missing transverse momentum. Nevertheless, a correlation with jet kinematics is also visible for the polarization networks, especially for the network trained on LL vs TX in pp-cmf. Therefore, a test is required to what extent this correlation can transfer a shape effect from the jet phase space to the shape of the resulting DNN score distribution.

The following procedure aims to estimate the maximal shape effect on the DNN score resulting from a shape variation in an individual jet kinematic. For a better understanding of this study, it is advisable to take a look at figure 12.5, which results from the procedure described in the following. The initial jet variable is split into six bins and the normalization in each bin i is multiplied by a parameter c_i. This parameter c_i can freely float within the envelope of the given scale uncertainty in this bin. Thus, c_i represents the possible theory variations in the specific jet kinematic. This is done for the η and p_T of the leading and subleading jet, the invariant di-jet mass m_{jj}, and the angular separations of the jets in $\Delta\phi_{jj}$, ΔR_{jj}, and Δy_{jj}. Six bins per variable, each with a corresponding parameter c_i, are sufficient for this test, as the main aim is to estimate the effect of general shape changes rather than small local fluctuations. These bins have the same width and are thus equally distributed over the kinematic range covered by the events in the signal region. The impact of the normalization parameters $\boldsymbol{c}$ can be directly transferred to the DNN distributions. A shape parameter s_{shape} is defined for the DNN distributions to find the most disadvantageous theory variation in c_i. The parameter

$$s_{\mathrm{shape}} := \left(\frac{\mathrm{DNN}^{\mathrm{sys}}_{\text{last bin}}(\boldsymbol{c})}{\mathrm{DNN}^{\mathrm{nom}}_{\text{last bin}}} - \frac{\mathrm{DNN}^{\mathrm{sys}}_{\text{first bin}}(\boldsymbol{c})}{\mathrm{DNN}^{\mathrm{nom}}_{\text{first bin}}} \right) \frac{\sum_i \mathrm{DNN}^{\mathrm{nom}}_i}{\sum_i \mathrm{DNN}^{\mathrm{sys}}_i(\boldsymbol{c})} \tag{12.15}$$

calculates the difference between the relative change in the last bin of the DNN distribution and the relative change in the first bin. To only measure shape effects, the difference is normalized by the change of the total number of events. For the polarization DNNs a high s_{shape} value would represent a down variation at low DNN scores and an up variation at high DNN scores. This kind of variation would modify the transverse polarization distribution to be more consistent with the sum of transversal and longitudinal polarized events. Thus, the maximal s_{shape} value for a given polarization DNN is expected to result in the lowest significance. For the DNN trained for $W^\pm W^\pm jj$-EW signal vs. background separation, such a clear connection between s_{shape} and the significance cannot be drawn. However, the s_{shape} value is still a suitable indicator for the shape effects on the $\mathrm{DNN}_{W^\pm W^\pm}$ score.

In section 12.3, a two-dimensional binning is chosen for each of the polarization measurements. To simplify the problem, only the projections on one of the two DNN scores are studied. For the $\mathrm{DNN}_{W^\pm W^\pm}$ the binning [0, 0.2, 0.6, 1] and for the polarization DNNs [0, 0.2, 0.4, 0.6, 0.8, 1] is used. The resulting bin widths are similar to the bin optimization results in table 12.3.

The s_{shape} value of the individual DNN distributions is now maximized for $W^\pm W^\pm jj$-EW signal events by varying the normalization parameters c_i for each of the jet kinematics within the scale uncertainty envelope. The shape effects s_{shape} for the individual neural network distributions are shown in figure 12.6. The renormalization and factorization scale can introduce the most significant shape impact since their envelope provides the widest range for the floating parameter c_i. For the $\mathrm{DNN}_{W^\pm W^\pm}$ score, shape changes in m_{jj} have the dominant impact on the distribution of the neural network score. The corresponding maximal shape variation is shown in figure 12.5. A down variation for m_{jj} close to 500 GeV and an up variation for all subsequent bins leads to a significant shape modification in the DNN score. This effect is expected due to the high positive Pearson correlation value for m_{jj} in figure 12.4.

As figure 12.6 indicates, the shape of the polarization DNN score distribution is unaffected by variations in m_{jj}. However, variations in leading jet p_T and the Δy_{jj} can change the shape of the polarization DNN distribution. As shown in figure 12.6, the shape effects on the

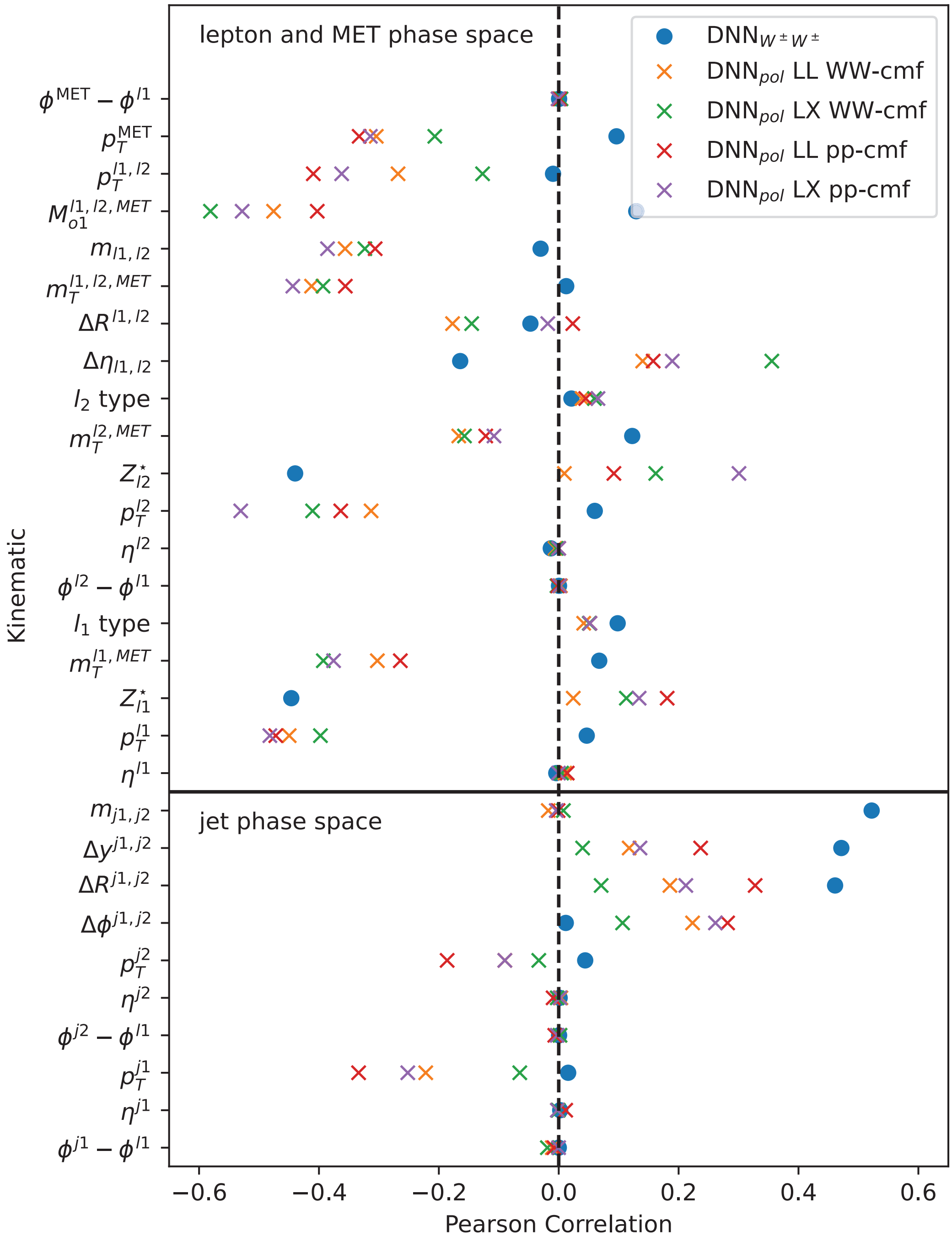


Figure 12.4: Pearson correlation between the DNN scores and the individual event kinematics.

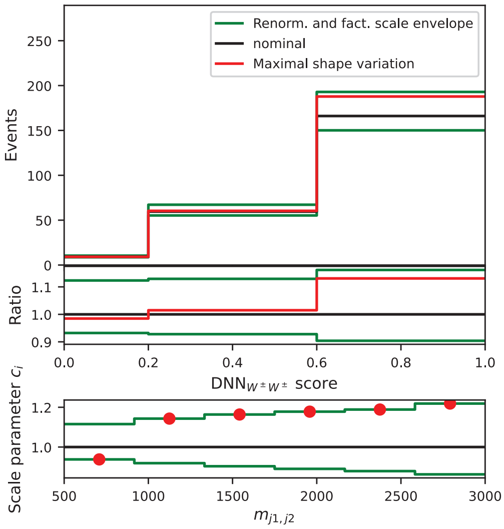


Figure 12.5: The distribution of the DNN score separating $W^{\pm}W^{\pm}jj$-EW and background and the systematic envelope of the renormalization and factorization scale systematic. The variation with the biggest shape effect s_{shape} defined in equation (12.15) originates from the shown pulls of c_i in the m_{jj} distribution.

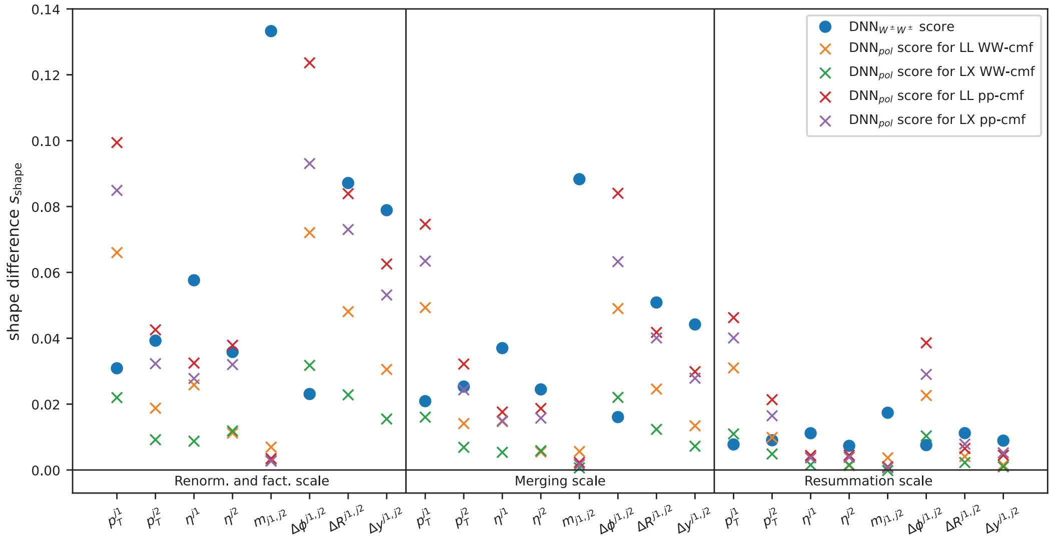
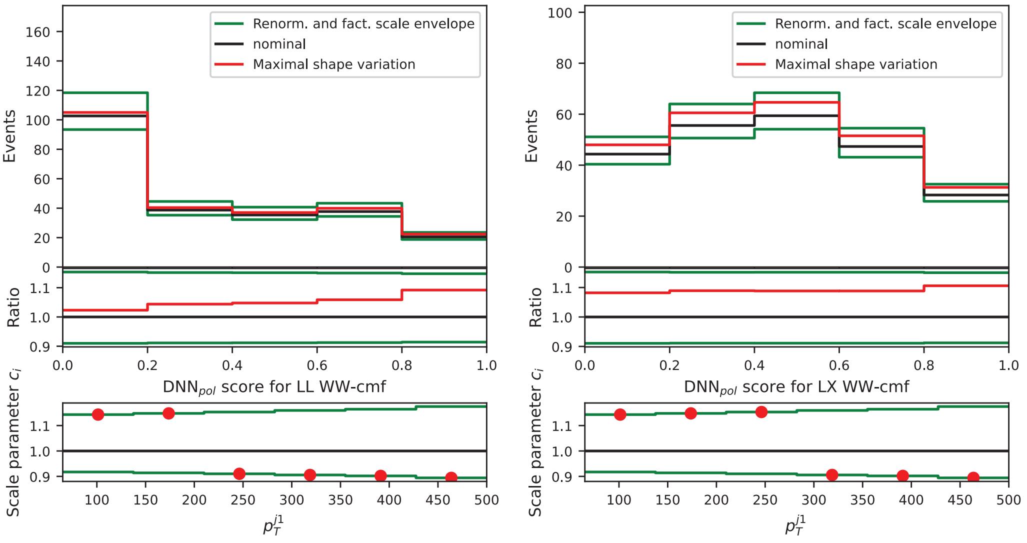


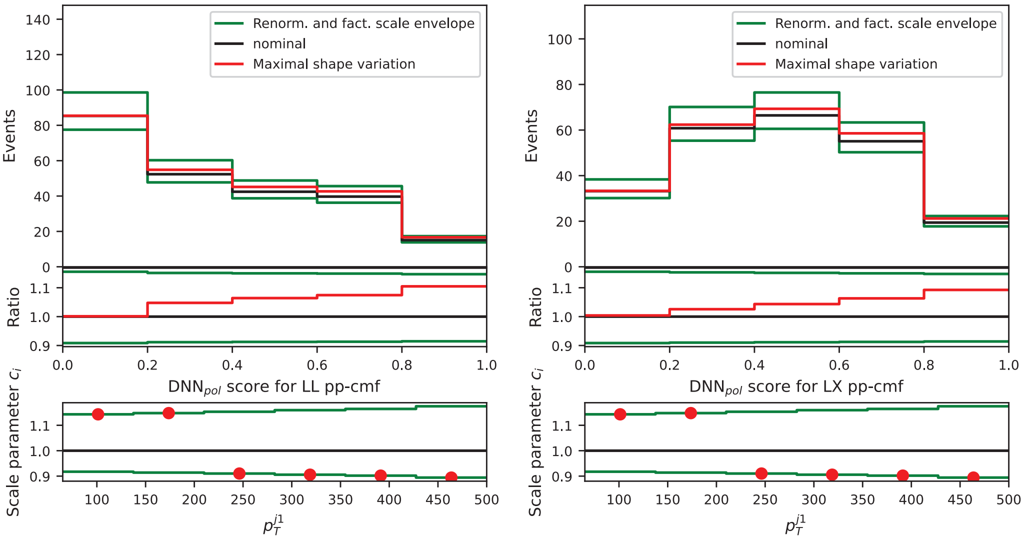
Figure 12.6: The maximal s_{shape} value defined in equation (12.15) for the different DNNs. The s_{shape} value is maximized within the envelope of different scale uncertainties and jet kinematics.

polarization DNNs in the WW-cmf are significantly smaller than for the partonic center-of-mass frame. The maximal shape differences are shown in figure 12.7 for the leading jet p_T and in figure 12.8 for $\Delta\phi_{jj}$. For the network trained for LX polarization in the WW-cmf, the shape is almost flat in figures 12.7b and 12.8b. Thus, the DNN score for the LX measurement in the WW-cmf is not affected by shape variations in the jet phase space. The shape effects for LL in WW-cmf are shown in figures 12.7a and 12.8a. The shape modifications in the jet phase space can lead to small shape effects in the DNN score of LL in WW-cmf. However, these shape effects are negligible compared to the DNN separating $W^{\pm}W^{\pm}jj$-EW and background. For the polarization measurement in the partonic center-of-mass system the shape effects in figures 12.7c, 12.7d, 12.8c and 12.8d become more prominent. Especially for the LL polarization in the pp-cmf, the shape effects become almost as prominent as for the $W^{\pm}W^{\pm}jj$-EW vs. background DNN in figure 12.5.



(a) LL in WW-cmf

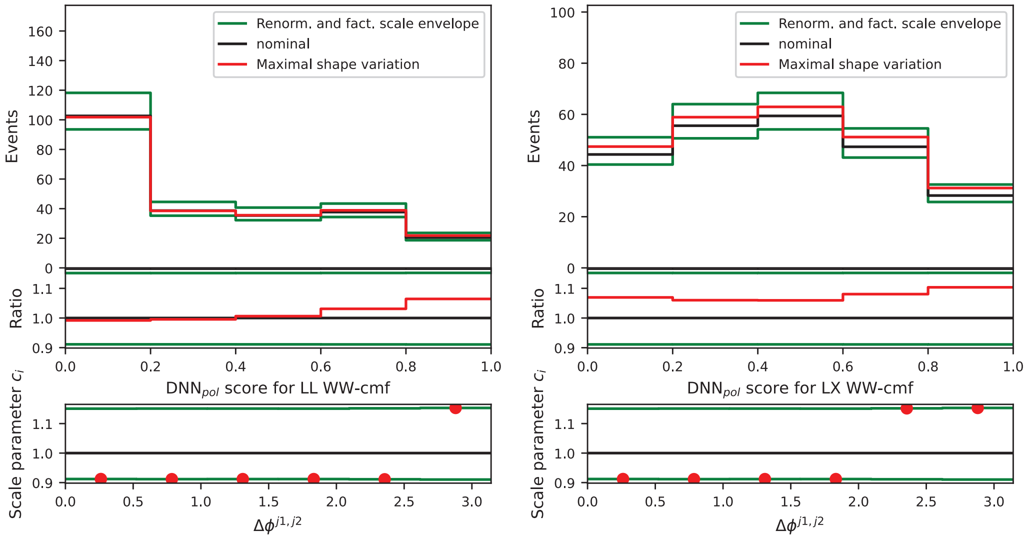
(b) LX in WW-cmf



(c) LL in pp-cmf

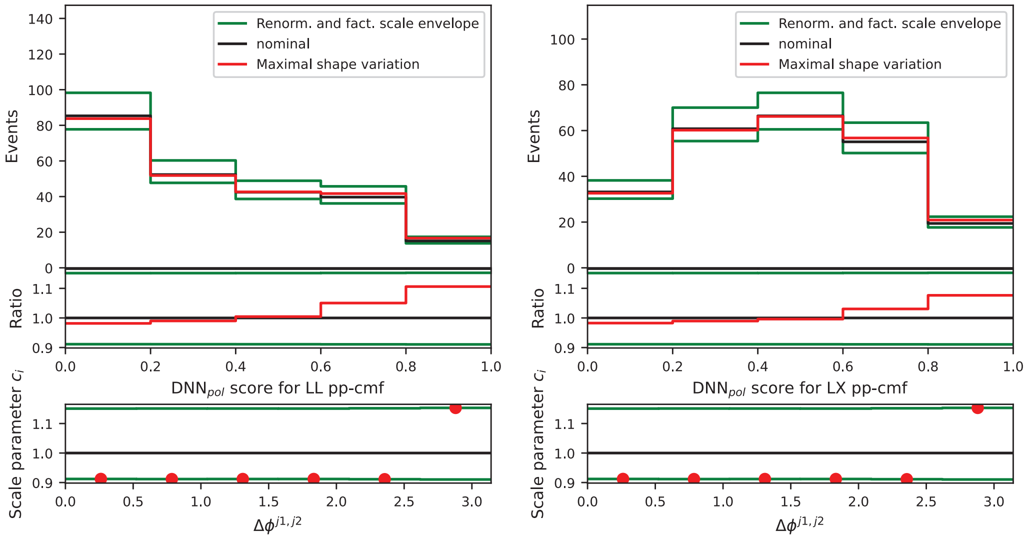
(d) LX in pp-cmf

Figure 12.7: The distribution of the polarization DNN scores and the systematic envelope of the renormalization and factorization scale systematic. The variation with the biggest shape effect s_{shape} defined in equation (12.15) originates from the shown pulls of c_i in the p_T^{j1} distribution.



(a) LL in WW-cmf

(b) LX in WW-cmf



(c) LL in pp-cmf

(d) LX in pp-cmf

Figure 12.8: The distribution of the polarization DNN scores and the systematic envelope of the renormalization and factorization scale systematic. The variation with the biggest shape effect s_{shape} defined in equation (12.15) originates from the shown pulls of c_i in the $\Delta\phi_{jj}$ distribution.

12.4.3 Resulting Correlation Scheme

For the polarization measurement in the WW-cmf the shape variations in the jet phase space have almost no effect. This is especially the case for the LX measurement, which is also expected to have the highest significance. Thus, a variation in the jet phase space is expected only to change the total number of events but not the shape in the LX polarization DNN. For the LL in WW-cmf measurement, the maximal shape effects are in the order of only a few percent. Since this is the maximal possible shape effect originating from the jet phase space, the variation can also be assumed to be approximately flat. This flat variation can be achieved by correlating all the bins in the individual polarization DNNs. This results in one nuisance parameter per scale uncertainty and polarization DNN.

For LX in pp-cmf and especially LL in pp-cmf, the use of a flat variation in the DNN score becomes arguable. However, the maximal shape effects from the jet phase space are still smaller than for the DNN score separating $W^{\pm}W^{\pm}jj$-EW and background. Since the overall sensitivity in the pp-cmf is already smaller than in the WW-cmf and the results in the WW-cmf represent the nominal measurement of this thesis, the same strategy as for the WW-cmf is used. Treating the scale uncertainties in the polarization DNN bins correlated also for the pp-cmf leads to a more straightforward analysis strategy with easier comparability with the nominal results in the WW-cmf.

The scale variations in the $\text{DNN}_{W^{\pm}W^{\pm}}$ score are shown to potentially have significant effects on the distribution shape. Therefore, the polarization histograms in table 12.3 that are split according to the $\text{DNN}_{W^{\pm}W^{\pm}}$ score are treated uncorrelated for the scale uncertainties. The decorrelation allows independent up and down variations of the total number of events in the different histograms. Thus, the fit can choose the most conservative systematic shape between the histograms. In addition, also the Low-m_{jj} region is decorrelated from the signal region to account for the shape effects in the jet phase space.

The study presented in this section allows a constraint of the theory scale uncertainties. For the polarization analysis in this thesis, strong indicators are found to locate the main impact of scale variations in the jet phase space. Based on this assumption, the scale uncertainties are correlated between the bins within the same histogram and decorrelated between the histograms and regions. The study designed in this section provides a new strategy for uncertainty handling for future analyses. Reasonable assumptions on the kinematic impact of theory uncertainties allow an estimation of the maximal shape impact. Understanding the maximum shape impact provides an excellent basis for the decision of potential correlations of the respective uncertainty.

12.5 Expected Results

The analysis strategy is defined before investigating the observed data distribution in the signal region. This procedure avoids fine-tuning the analysis to the specific characteristics in the measured data. Therefore, the fit strategy is investigated for the expected distributions. The prediction of the Standard Model is used as so-called Asimov data for the fitting instead of the data measured at the ATLAS detector. Measured data is only used in the $W^{\pm}Z$ control region since this region does not affect the $W^{\pm}W^{\pm}jj$-EW normalization but provides the correct normalization of the $W^{\pm}Z$-QCD background. Due to the $W^{\pm}W^{\pm}jj$-EW contribution in the signal and Low-m_{jj} control region, Asimov data is used. The Asimov data is constructed by the nominal prediction $m_i(\mu_L, \mu_T, \mu_{\text{WZ-QCD}}, \gamma_i, \boldsymbol{\theta})$ with $\mu_L = 1$, $\mu_T = 1$, $\gamma_i = 1$, and $\boldsymbol{\theta} = \mathbf{0}$. The normalization $\mu_{\text{WZ-QCD}}$ is set to 0.74 to provide Asimov data which is consistent with the data observed in the $W^{\pm}Z$ control region in figure 12.9. Since the $W^{\pm}Z$ control region is "unblinded" with measured data and the Low-m_{jj} control region and signal region are "blinded" by using Asimov data, this fit procedure is referred to as partially unblinded. The partially unblinded fit provides the expectation for the significance, fiducial cross-section, and impact of uncertainties.

The profile likelihood fit is used as described in section 12.1. The fit considers the distributions in the signal region, the Low-m_{jj} control region, and the $W^{\pm}Z$ control region defined in section 6.3. In the signal region, three polarization histograms are constructed with the binning derived in section 12.3. The Low-m_{jj} and $W^{\pm}Z$ control regions are included as 1-bin distributions considering only the total number of events to constrain the normalization of the backgrounds.

The main focus of the analysis is the polarization measurement in the WW-cmf. Therefore, the results in the partonic center-of-mass system are discussed but their corresponding figures are only shown in the appendix. All the shown results are derived for polarization simulated by Sherpa. The results for a polarization simulation by MadGraph are briefly discussed in section 12.7.

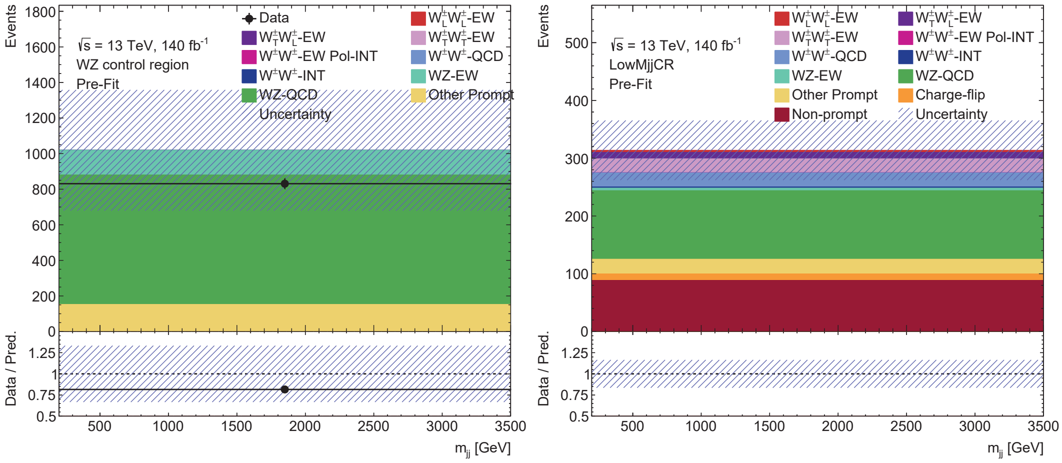
12.5.1 Expected Significance

The unblinded $W^{\pm}Z$ control region and the blinded Low-m_{jj} control region are shown in figure 12.9. The $W^{\pm}Z$ control region has the dominant impact on the $W^{\pm}Z$-QCD normalization while the Low-m_{jj} control region constrains especially the non-prompt background. The DNN distributions used for the LL measurement in the WW-cmf are shown in figure 12.10. The two splits at the $\text{DNN}_{W^{\pm}W^{\pm}}$ score leads to three histograms with increasing $W^{\pm}W^{\pm}jj$-EW purity. In each of these histograms, the polarized network separates the longitudinal and transverse polarization. Similar distributions are used to measure the LX polarization state and the polarizations in the pp-cmf. Since no pulls of the parameter of interest μ_L and μ_T, and the nuisance parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ are expected in the partially unblinded fit, the detailed discussion of the individual pre-fit and post-fit distributions is moved to the unblinded application for measured data in section 12.6. Thus, the current section focuses on the expected significances and impact of the systematic uncertainties.

Asimov Fit Results

The significances of the partially unblinded fit are listed in table 12.4. The expected significances are lower in the partonic center-of-mass frame. The corresponding normalizations are shown in figure 12.11a. Due to the use of Asimov data in the signal and Low-m_{jj} control region, the fit parameters of the polarization states are exactly at 1. The perfect agreement with the predicted polarization contribution is a consequence of the $W^{\pm}Z$-QCD contribution in the Asimov data being consistent with the data in the $W^{\pm}Z$ control region. Accordingly, the parameter $\mu_{\text{WZ-QCD}}$ is pulled to 0.74 for all partially unblinded fits.

The uncertainty of μ_L is significantly lower for the single boson polarization LX measurement than for the di-boson polarization LL. The lower uncertainty is a direct consequence of the additionally expected events of the LX polarization state. This difference is also reflected by the expected LL and LX significances in table 12.4. While the LX polarization is close



(a) $W^{\pm}Z$ control region

(b) Low-m_{jj} control region

Figure 12.9: The pre-fit distribution in the $W^{\pm}Z$ control region and Low-m_{jj} control region.

to discovery with $5\,\sigma$, the expected LL polarization is still significantly below evidence with $3\,\sigma$.

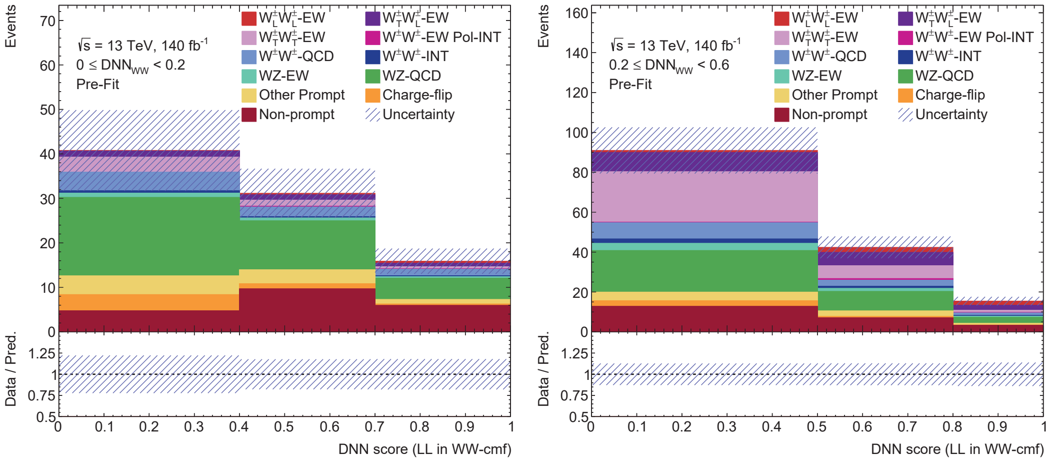
Table 12.4: The significances of the longitudinal polarization states extracted by the partially unblinded fits. The fits are derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74.

	WW-cmf	pp-cmf
LL significance	$1.512\,\sigma$	$0.937\,\sigma$
LX significance	$4.557\,\sigma$	$3.975\,\sigma$

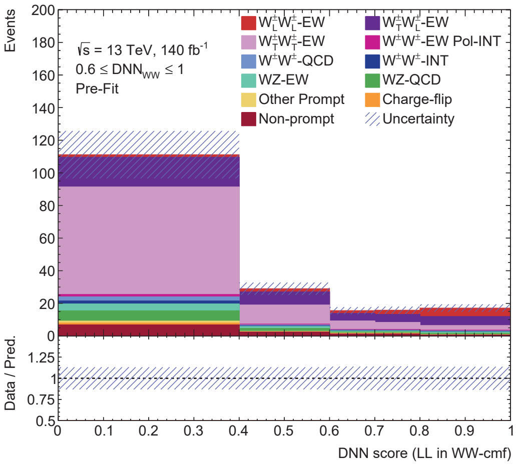
The significance and the uncertainties of the normalization also indicate a systematic difference between the WW-cmf and the pp-cmf. The difference for the di-boson polarization measurement originates dominantly from the significantly lower cross-section of LL in the pp-cmf. However, this does not explain the difference in the LX measurement which has consistent cross-sections in the two reference frames. As shown in section 9.3.4, the polarization networks provide a better separation of the polarization states in the WW-cmf than in the pp-cmf. The better separation directly leads to higher significance in the WW-cmf.

Discussion of Uncertainties

The main uncertainty of the polarization measurement is the low number of events expected for the $W^{\pm}W^{\pm}jj$-EW contribution. The number of events is limited by the run time of the ATLAS experiment and will improve in the next LHC runs. On the other hand, the systematic uncertainties are determined by the experimental setup, the analysis strategy, and the theoretical prediction. Thus, investigating their impact on the measurement leads to a better understanding of the analysis and shows potential for future improvements. As described in section 12.1, the significance of the likelihood fit originates from the ratio of the likelihood $\mathcal{L}(0, \hat{\hat{\mu}}_T, \hat{\hat{\mu}}_{\text{WZ-QCD}}, \hat{\hat{\gamma}}, \hat{\hat{\boldsymbol{\theta}}})$ with fixed signal strength $\mu_L = 0$ and the likelihood fit $\mathcal{L}(\hat{\mu}_L, \hat{\mu}_T, \hat{\mu}_{\text{WZ-QCD}}, \hat{\gamma}, \hat{\boldsymbol{\theta}})$ with free-floating normalization of the longitudinal polarization state. The significance drops if the data can be covered by the background-only hypothesis of $\mu_L = 0$ without significant pulls in the Gaussian and Poisson constrain terms of the nuisance parameters. The impact of the systematic uncertainties is investigated by two different methods:



(a) $\text{DNN}_{W^\pm W^\pm}$-score ≤ 0.2 **(b)** $0.2 < \text{DNN}_{W^\pm W^\pm}$-score ≤ 0.6



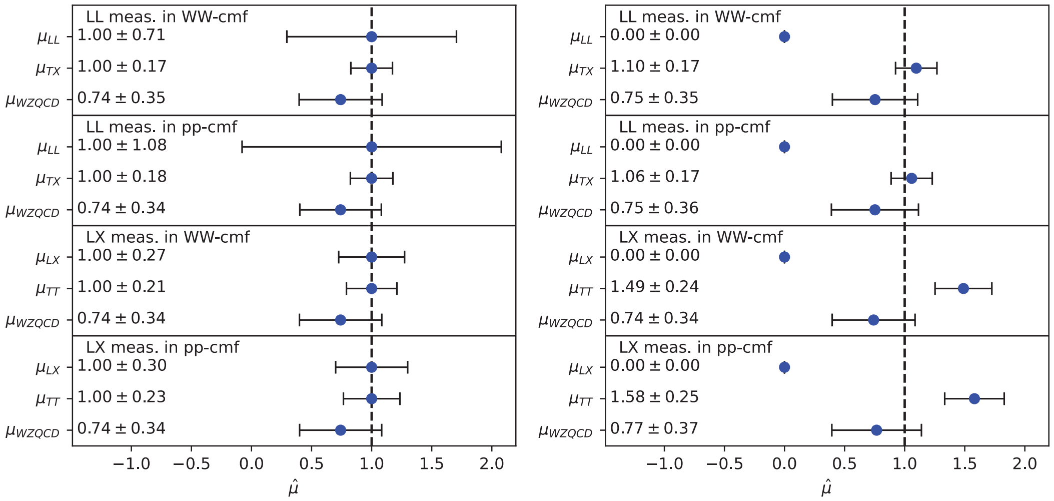
(c) $0.6 < \text{DNN}_{W^\pm W^\pm}$-score ≤ 1

Figure 12.10: The pre-fit signal region distributions for the LL in WW-cmf measurement. The polarization states are simulated by Sherpa. As described in section 12.3 the signal region is split according to the $\text{DNN}_{W^\pm W^\pm}$-score. The shown histograms result from the application of the DNN_{pol} trained for the LL polarization state in the WW-cmf.

- **Impact of nuisance parameter on signal strength**
  The impact of each individual nuisance parameter is investigated by fixing the parameter to its post-fit uncertainty $\pm\sigma$. The likelihood fit is repeated with this particular nuisance parameter fixed and the difference $\Delta\mu$ to the nominal fit result of μ_{LL}/μ_{LX} is calculated. The calculation of $\Delta\mu$ allows the ranking of the nuisance parameter impact on the signal strength. The resulting ranking plot for the LL and LX measurement in the WW-cmf is shown in figure 12.12.

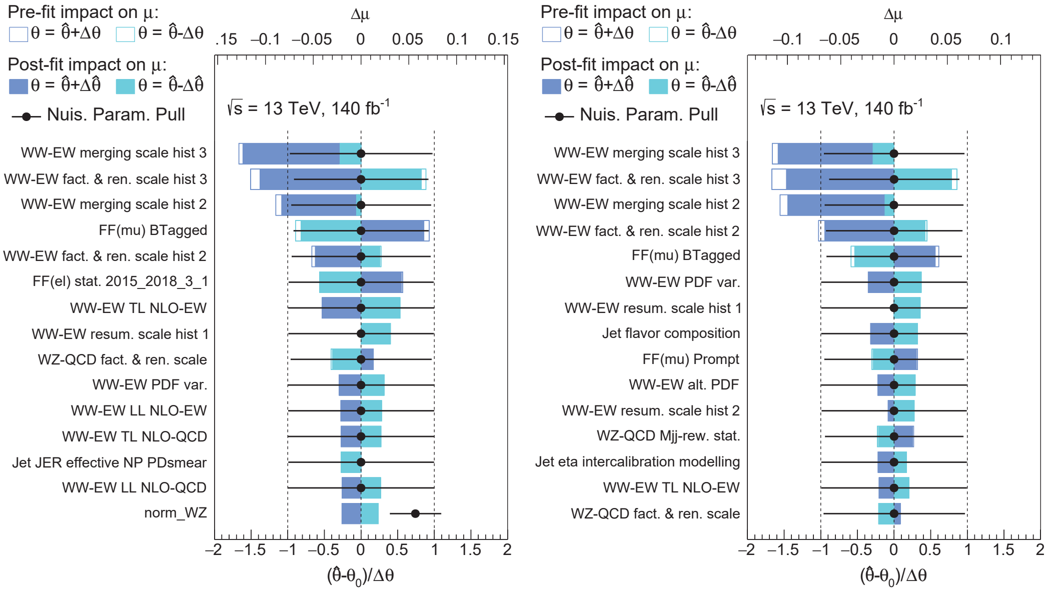
- **Background-only fit pulls**
  The pulls of the nuisance parameters in $\mathcal{L}(0, \hat{\hat{\mu}}_T, \hat{\hat{\mu}}_{\text{WZ-QCD}}, \hat{\hat{\gamma}}, \hat{\hat{\boldsymbol{\theta}}})$ is evaluated in a background-only fit with μ_{LL}/μ_{LX} fixed to 0. The normalization of the transverse polarization state and the $W^\pm Z$-QCD background is shown in figure 12.11b. The pulls of the nuisance parameters for the LL and LX fit in the WW-cmf can be found in figure 12.13.

Since the impact of the systematic uncertainties is similar between the two reference frames, the ranking and background-only pull plots for polarization in the pp-cmf are moved into the appendix in figures G.1 and G.2. The following discussion of the uncertainties of



(a) Fit normalizations **(b)** Fit normalizations with $\mu_{LL} = 0/\mu_{LX} = 0$

Figure 12.11: The μ_{LL}/μ_{LX}, μ_{TX}/μ_{TT}, and $\mu_{\text{WZ-QCD}}$ normalization resulting from the partially unblinded significance fits. In figure 12.11a, the normalization of the longitudinal polarization is free-floating while in figure 12.11b it is fixed to 0. The fit is derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74.

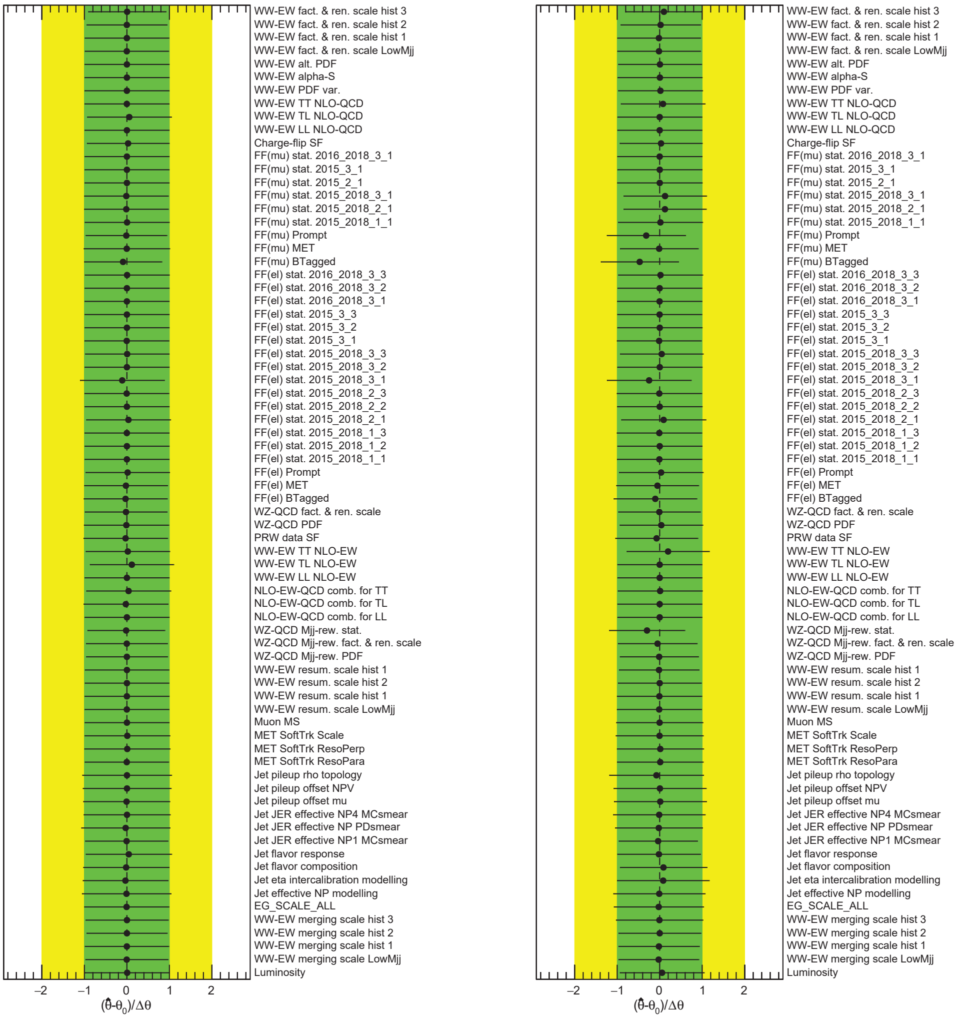


(a) Ranking of μ_{LL} measurement in WW-cmf **(b)** Ranking of μ_{LX} measurement in WW-cmf

Figure 12.12: The ranking plots for the partially unblinded significance fits in the WW-cmf. The fit is derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74.

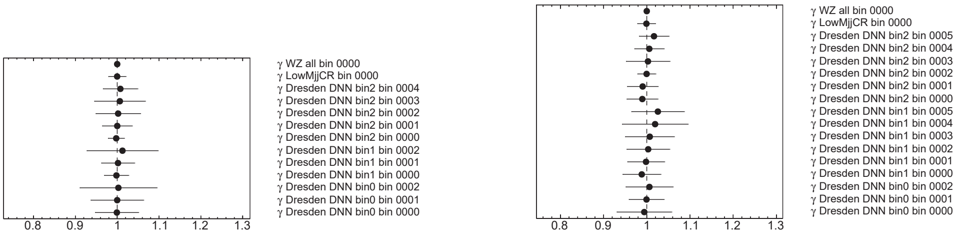
the measurement in the WW-cmf also applies to the pp-cmf.

The background-only fit of the LX measurement results in significantly stronger pulls of the nuisance parameters than for the LL measurement. The stronger pulls lead to lower likelihood values in the constraint terms and therefore the higher significance of the LX measurement. The deviation from the nominal $\boldsymbol{\theta} = 0$ is distributed over several nuisance parameter pulls for both measurements. The following paragraphs discus the impact of individual systematic uncertainties in more detail.



(a) $\boldsymbol{\theta}$ of LL fit in WW-cmf with $\mu_{LL} = 0$

(b) $\boldsymbol{\theta}$ of LX fit in WW-cmf with $\mu_{LX} = 0$



(c) γ of LL fit in WW-cmf with $\mu_{LL} = 0$

(d) γ of LX fit in WW-cmf with $\mu_{LX} = 0$

Figure 12.13: The pulls of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ for the partially unblinded background-only significance fits in the WW-cmf. The fit parameter μ_{LL}/μ_{LX} is fixed to 0. The fit is derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74.

Theory uncertainties: In figure 12.12, the nuisance parameters of the merging, and renormalization and factorization scale have the highest ranking. As shown in section 11.2.1, these two theory uncertainties have the dominant normalization uncertainty on the $W^{\pm}W^{\pm}jj$-EW signal. An up variation of these systematic uncertainties has to be compensated by decreasing the longitudinal signal strength. The one-sided impact of the merging scale is discussed in section 11.2.1. Since the merging, and renormalization and factorization scale directly affect the longitudinal signal strength, their impact in figure 12.12 does not necessarily represent their impact on the significance. Their relative uncertainty within the individual polarization histograms is flat. Since the theory scale uncertainties do not affect the shape of the $W^{\pm}W^{\pm}jj$-EW contribution within the individual histograms, their impact is already covered by the free-floating normalization on the transverse polarization. The minor impact of the theory scale uncertainties on the significance is proven by the pulls in the background-only fit in figures 12.13a and 12.13b where the corresponding post-fit parameters remain close to 0. The vanishing impact on the expected significance is a direct consequence of the correlation scheme derived in section 12.4. If the scale uncertainties were uncorrelated across the bins of the individual histograms, they would allow significant shape effects. These shape effects would result in a better agreement in the background-only fit and thus a lower significance of the longitudinal signal.
The same argumentation holds also for the remaining theory uncertainties of the $W^{\pm}W^{\pm}jj$-EW contribution. They partially occur in the ranking in figure 12.12 but are not pulled in figures 12.13a and 12.13b. Therefore, the theory uncertainties on the $W^{\pm}W^{\pm}jj$-EW signal do not affect the significance measurement.

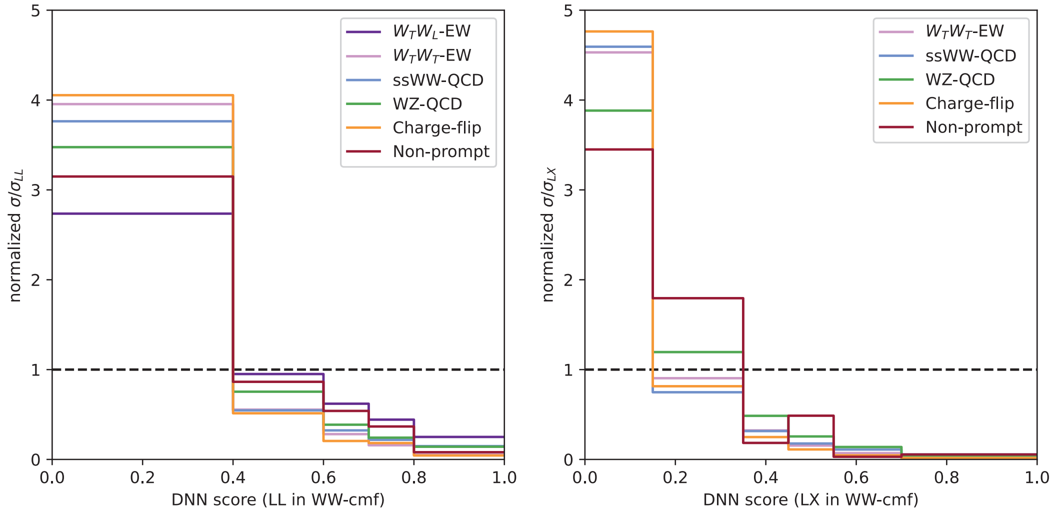
Fake factor method: The highest ranked nuisance parameters in figure 12.12, which are not $W^{\pm}W^{\pm}jj$-EW theory uncertainties, correspond to the b-tagging and the prompt subtraction in the fake factor method. Their impact is confirmed by the pulls in figures 12.13a and 12.13b. To explain the preference for non-prompt background uncertainties, the shape of the longitudinal signal is compared to the background shapes. The ratio of the normalized shapes is shown in figure 12.14 for the histogram with the highest $\text{DNN}_{W^{\pm}W^{\pm}}$ score which dominates the polarization measurement. The ratios indicate the potential of each background to cover the longitudinal signal distribution. The non-prompt contribution in the first bin is significantly closer to the shape of the longitudinal polarization than the other non-$W^{\pm}W^{\pm}jj$-EW backgrounds. Thus, a higher scaling of the non-prompt background improves the agreement with the Asimov data for the background-only fit. Therefore, the impact of systematic uncertainties could be improved by smaller non-prompt background uncertainties or a better separation between the non-prompt background and the longitudinal polarization. As explained in section 9.1, the non-prompt background is not used for the training of the discriminator networks due to the prompt subtraction in the fake factor method and the low number of generated events provided by Monte Carlo generators. If a future analysis can solve one of the two issues, the impact of the systematic non-prompt uncertainties will be reduced.

NLO correction: The ranking plot in figure 12.12a and both background-only pull plots in figures 12.13a and 12.13b indicate an impact from the systematic uncertainty of the NLO-EW correction. The NLO-EW correction is derived in section 8.2 for each individual polarization state. The NLO-EW uncertainty is given by the difference between the nominal reweighting in m_{jj} and an alternative reweighting in the leading lepton p_T. The correlation between discriminator networks and individual kinematics was studied in figure 12.4 in the previous section. The polarization networks show no correlation with m_{jj} but a significant correlation with the leading lepton p_T. Thus, using the leading lepton p_T for the NLO-EW correction instead of m_{jj} causes significant shape differences in the polarization DNN distribution shown in figures 12.15a and 12.15b. A multivariate NLO-EW correction in the full phase space is required to reduce this uncertainty in future analyses.

Reweighting of $W^{\pm}Z$ background: The last relevant pull for the LX background-only fit in figure 12.13b occurs for the statistical uncertainty of the m_{jj} shape reweighting of the $W^{\pm}Z$-QCD background. The limited number of events is the leading uncertainty of the $W^{\pm}Z$-QCD shape uncertainty. As shown in figures 12.15c and 12.15d, this uncertainty is flat within the polarization DNN score since these DNNs are not correlated with m_{jj}. However, the uncertainty leads to a significant shape effect across the phase space regions and histograms. The variation in the signal region and the Low-m_{jj} control region are anti-correlated and the impact of the variation increases with the $\mathrm{DNN}_{W^{\pm}W^{\pm}}$ score. Therefore, the statistical uncertainty of the m_{jj} reweighting offers the possibility to increase the $W^{\pm}Z$-QCD contribution in the histograms with high $W^{\pm}W^{\pm}jj$-EW purity with less impact on the phase spaces and histograms dominated by background. For future analyses, this uncertainty will be reduced by the higher integrated luminosity in the next LHC runs and potentially better shape modeling by Monte Carlo generators.

Bin-by-bin γ parameters: In addition to the pulls of the systematic nuisance parameters $\boldsymbol{\theta}$, the statistical bin-by-bin nuisance parameters $\boldsymbol{\gamma}$ in figures 12.13c and 12.13d are studied for the background-only fit. The $\boldsymbol{\gamma}$ parameters are pulled down for low polarization DNN scores and up for high polarization DNN scores to cover the longitudinal contribution. The impact of the bin-wise statistical modeling uncertainty can be reduced by higher number of events in the generated samples and the data-driven backgrounds.

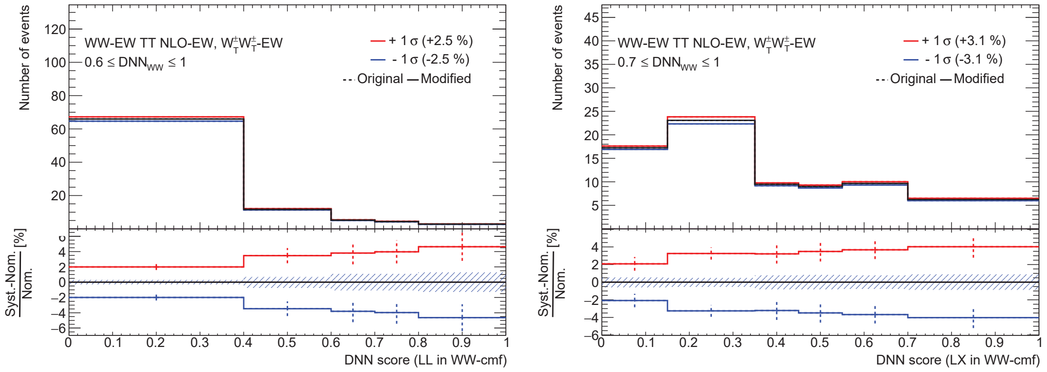
In conclusion, several systematic uncertainties affect the significance of the longitudinal measurement. However, the current polarization analysis is limited by the number of events in the measured data as the uncertainty breakdown in the subsequent section will show. The next runs of the LHC are expected to significantly increase the amount of available data, increasing the impact of the systematic uncertainties. The current section has shown the origin of the dominant systematic uncertainties, which could be improved for future analyses.



(a) $0.6 < \mathrm{DNN}_{W^{\pm}W^{\pm}}$-score ≤ 1 for μ_{LL} in WW-cmf measurement

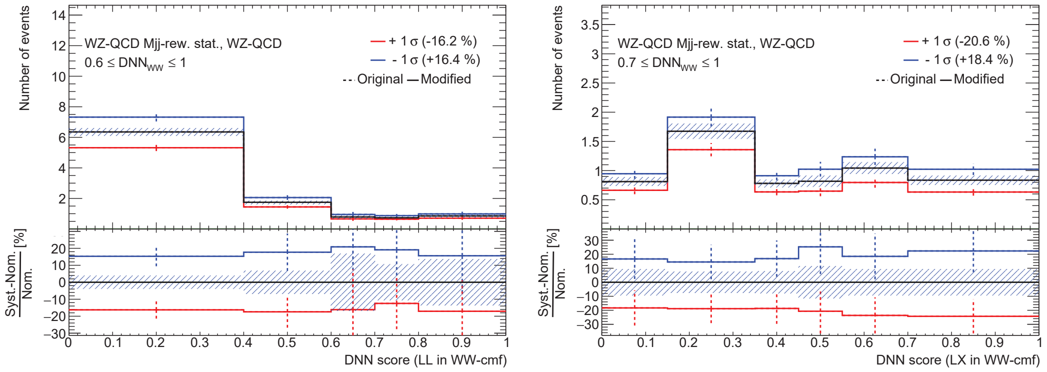
(b) $0.7 < \mathrm{DNN}_{W^{\pm}W^{\pm}}$-score ≤ 1 for μ_{LX} in WW-cmf measurement

Figure 12.14: The shape of the $W^{\pm}W^{\pm}jj$-QCD, $W^{\pm}Z$-QCD, charge-flip, and non-prompt background compared to the LL/LX signal of the polarization measurement in the WW-cmf. The shape differences are evaluated in the signal region histogram with the highest $\mathrm{DNN}_{W^{\pm}W^{\pm}}$ score.



(a) NLO-EW variation on TT for μ_{LL} in WW-cmf measurement

(b) NLO-EW variation on TT for μ_{LX} in WW-cmf measurement



(c) $W^{\pm}Z$-QCD stat. m_{jj} reweighting uncertainty for μ_{LL} in WW-cmf measurement

(d) $W^{\pm}Z$-QCD stat. m_{jj} reweighting uncertainty for μ_{LX} in WW-cmf measurement

Figure 12.15: The systematic uncertainty of the NLO-EW correction on the TT polarization state and the statistical uncertainty of the m_{jj} reweighting of the $W^{\pm}Z$-QCD background. The variations are shown in the signal region histogram with $0.6 <$ DNN$_{W^{\pm}W^{\pm}}$-score ≤ 1 for the LL measurement and $0.7 <$ DNN$_{W^{\pm}W^{\pm}}$-score ≤ 1 for the LX measurement.

12.5.2 Expected Fiducial Cross-section

The fitting strategy presented so far aims to determine the significance of the longitudinal polarization. The same profile likelihood strategy can be used to calculate the fiducial cross-section σ^{fid} in the particle-level signal region defined in section 6.3.4. The measured fiducial cross-section

$$\sigma^{fid}_{obs} = \hat{\mu} \cdot \sigma^{fid}_{pred} \tag{12.16}$$

is calculated by the predicted fiducial cross-section σ^{fid}_{pred} of the longitudinal polarization and the corresponding fit result for the normalization μ. The predicted fiducial cross-sections are listed in table 12.5. The fiducial cross-section fit of the normalization μ_L differs from the significance fit by the treatment of the theory signal uncertainties. An up variation of a pure normalization uncertainty in σ^{fid}_{pred} leads to an equivalent down variation in the corresponding fitted signal strength $\hat{\mu}_L$. Thus, the normalization effect of theory uncertainties cancels out in calculating the observed fiducial cross-section σ^{fid}_{obs}. Therefore, the signal theory uncertainties are normalized for the fiducial cross-section measurement on particle-level and only their shape impact is considered.

Asimov Fit Results

The expected normalizations $\hat{\mu}_L$ of the partially unblinded fit are listed in table 12.6 and the corresponding fiducial cross-sections are presented in table 12.7. The uncertainties are split

Table 12.5: The fiducial cross-sections of the longitudinal polarization states extracted by the partially unblinded fits. The fits are derived from measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74.

	fiducial cross-section
LL in WW-cmf	0.330 fb
LX in WW-cmf	1.354 fb
LL in pp-cmf	0.212 fb
LX in pp-cmf	1.420 fb

into their origin. The impact of the statistical uncertainty of the measured and Asimov data is derived by an additional fit considering only the statistical uncertainty of the prediction and data. The breakdown into the statistical modeling uncertainty, the systematic modeling uncertainty, the experimental uncertainty, and the luminosity uncertainty is directly provided by the TRExFitter. The TRExFitter estimates the impact of a group of uncertainties by dropping all nuisance parameters of the remaining uncertainty groups and reevaluating the fit. The allocation of nuisance parameters to these uncertainty groups and subgroups is shown in table G.1. The more detailed version of the uncertainty breakdown is given in table 12.8. The statistical uncertainty of the measured and Asimov data dominates the uncertainties and will decrease with the subsequent runs of the LHC.

Table 12.6: The expected normalization of the LL/LX polarization in the WW-cmf and pp-cmf. The uncertainties are split into the statistical uncertainty of the data, statistical modeling uncertainty, systematic modeling uncertainty, and luminosity uncertainty. The fit is derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74.

	Signal strength	Uncertainty breakdown				
		stat.	mod. stat.	mod. syst.	exp. syst.	lumi.
LL in WW-cmf	1.000 ± 0.695 (tot.)	±0.681	±0.142	±0.041	±0.079	±0.011
LX in WW-cmf	1.000 ± 0.251 (tot.)	±0.240	±0.052	±0.019	±0.048	±0.011
LL in pp-cmf	1.000 ± 1.103 (tot.)	±1.065	±0.221	±0.104	±0.148	±0.013
LX in pp-cmf	1.000 ± 0.276 (tot.)	±0.263	±0.056	±0.033	±0.053	±0.011

Table 12.7: The expected fiducial cross-section of the LL/LX polarization in the WW-cmf and pp-cmf. The uncertainties are split into the statistical uncertainty of the data, statistical modeling uncertainty, systematic modeling uncertainty, and luminosity uncertainty. The fit is derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74.

	Fid. cross-section [fb]	Uncertainty breakdown [fb]				
		stat.	mod. stat.	mod. syst.	exp. syst.	lumi.
LL in WW-cmf	0.330 ± 0.229 (tot.)	±0.225	±0.047	±0.013	±0.026	±0.004
LX in WW-cmf	1.354 ± 0.340 (tot.)	±0.325	±0.071	±0.025	±0.065	±0.015
LL in pp-cmf	0.212 ± 0.234 (tot.)	±0.226	±0.047	±0.022	±0.031	±0.003
LX in pp-cmf	1.420 ± 0.392 (tot.)	±0.373	±0.079	±0.047	±0.075	±0.015

Discussion of Uncertainties

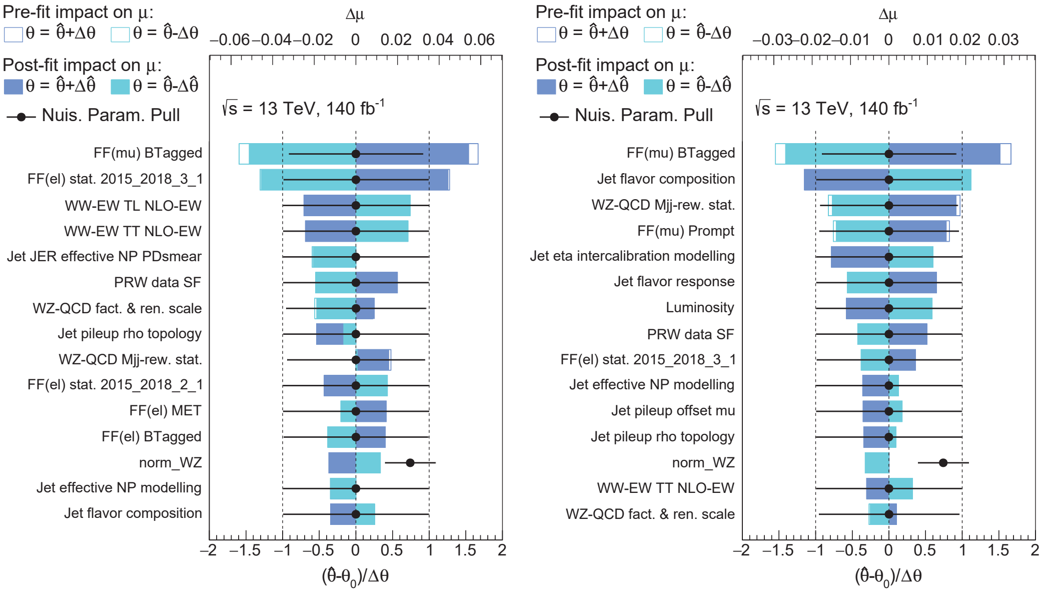
The discussion of the uncertainties on the fiducial cross-section is split into the uncertainty breakdown in tables 12.6 and 12.8 and the ranking plot in figure 12.16.

Uncertainty breakdown: The leading experimental uncertainty originates from the fake factor method used to model the non-prompt background. This experimental uncertainty will partially improve during the subsequent runs of the LHC due to the higher number of events in the control region used for the fake factor extraction.
The modeling uncertainty is dominated by the limited statistic. This uncertainty is derived by the γ nuisance parameters which are constrained by the number of simulated and data-driven events of the Standard Model prediction. More events can be requested for processes simulated by event generators to decrease this uncertainty. However, the statistical modeling uncertainty is dominated by the number of measured events scaled by the fake factor method. Therefore, also the modeling uncertainty is limited by the integrated luminosity of the LHC Run 2 dataset. In future analyses with more data, the importance of the polarized NLO correction uncertainties of the $W^{\pm}W^{\pm}jj$-EW contribution will increase.
The uncertainties of the normalization μ_L are higher in the partonic center-of-mass frame. The lower relative uncertainty in the WW-cmf results from the higher cross-section of the LL polarization state and the better separation by the polarization DNN. These differences are already discussed for the expected significances in the previous section.

Table 12.8: The impact of uncertainty sources on the expected fiducial cross-section of the LL/LX polarization in the WW-cmf and pp-cmf. The fit is derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74.

Source	Impact [%]			
	LL in WW-cmf	LX in WW-cmf	LL in pp-cmf	LX in pp-cmf
Experimental				
Electron reconstruction	0.32	0.15	0.38	0.37
Muon reconstruction	0.12	0.08	0.60	0.03
Jet reconstruction	1.93	2.92	3.91	3.12
p_T^{miss} reconstruction	0.64	0.14	2.23	0.24
Pileup modeling	1.96	0.90	2.74	0.92
Fake factor method	7.52	3.61	13.79	4.09
Charge-flip scale factor	1.00	0.08	1.26	0.11
Luminosity	1.18	1.11	1.33	1.08
Modeling				
Theory uncertainties of $W^{\pm}W^{\pm}jj$-EW	0.31	0.40	0.77	0.44
Approx. NLO correction unc. of $W^{\pm}W^{\pm}jj$-EW	3.87	0.71	8.44	2.43
Theory uncertainty of $W^{\pm}Z$-QCD	1.36	0.38	4.14	0.52
Reweighting uncertainty of $W^{\pm}Z$-QCD	0.80	1.65	4.58	2.13
Modeling statistic	14.51	5.23	22.09	5.60
Data statistic	68.13	24.01	106.52	26.27
Total	70.29	25.14	110.33	27.60

Ranking plot: The impact of the systematic uncertainties shown in table 12.8 is supported by the ranking plot in figure 12.16 for the measurement in the WW-cmf. Compared to the ranking in figure 12.12, the theory uncertainties are not listed on top. Since these uncertainties are normalized for the fiducial cross-section measurement, only their shape variation is considered. Due to their overall flat shape, the impact on the cross-section measurement is negligible. The remaining ranking is consistent with the results for the significance fit: The systematic uncertainty is dominated by the fake factor method followed by the application of the higher-order corrections and the statistical uncertainty of the $W^{\pm}Z$-QCD shape reweighting. The impact of these individual uncertainties is already explained in detail for the significance fit in section 12.5.1. The only significant change to the ranking for the significance fit is the increased impact of the jet reconstruction. The jet reconstruction is also the second most important experimental uncertainty in table 12.8 after the fake factor method. Therefore, as more data becomes available in the next LHC runs, an enhanced jet reconstruction could lead to improved sensitivity of future VBS polarization analyses.



(a) Ranking of μ_{LL} measurement in WW-cmf **(b)** Ranking of μ_{LX} measurement in WW-cmf

Figure 12.16: The ranking plots of the partially unblinded fiducial cross-section fits in the WW-cmf. The fit is derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74.

Upper Limit on LL

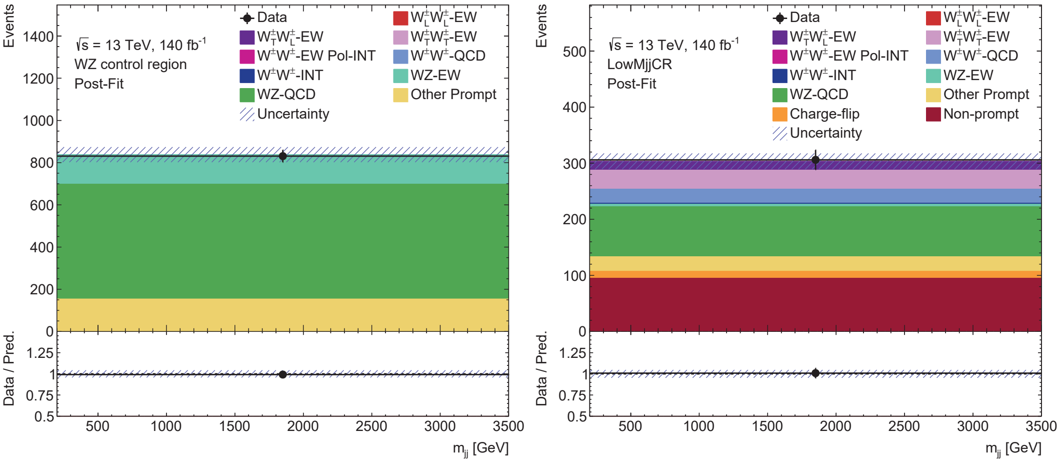
Since the uncertainty of the di-boson polarization state LL is in the same magnitude as the expected cross-section, an upper limit is calculated. The upper limit represents the maximum value of the one-sided 95% confidence level interval (CL) of the fiducial cross-section. This confidence interval is derived for the test statistic q_{μ} in equation (12.5). The resulting upper limits for the partially unblinded fit are listed in table 12.9. These upper limits exclude fiducial cross-sections of LL which are two to three times higher than the Standard Model prediction in table 12.5.

Table 12.9: The expected 95% CL upper limit of the LL polarization in the WW-cmf and pp-cmf. The fit is derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74.

	Fiducial cross-section upper limit [fb]
LL in WW-cmf	0.750
LL pp-cmf	0.660

12.6 Observed Results

The final step of this analysis is the application on measured data. The same fitting strategy is applied as in the previous sections but with the signal and Low-m_{jj} control region being unblinded. The fit results in the WW-cmf are analyzed in detail. For the polarization measurement in the pp-cmf, the significances and fiducial cross-sections are presented in this section, but the corresponding plots are moved to appendix G. Furthermore, the pre-fit event yields are listed in tables G.2 to G.5 and the post-fit event yields are listed in tables G.6 to G.9.



(a) $W^{\pm}Z$ control region

(b) Low-m_{jj} control region

Figure 12.17: The post-fit distribution for the LL in WW-cmf measurement in the $W^{\pm}Z$ control region and Low-m_{jj} control region.

12.6.1 Observed Significance

The $W^{\pm}Z$ and Low-m_{jj} control regions are included in the fit to constrain the background contributions. The corresponding number of events after the fit of the LL in WW-cmf significance is shown in figure 12.17. The likelihood fit results in an excellent agreement with the measured data in the control regions. The control region distributions for the remaining polarization measurements are equivalent and therefore given in figures G.3 to G.5 in the appendix. The polarization measurement is dominated by the signal region due to the high purity in $W^{\pm}W^{\pm}jj$-EW events and the application of the discriminator DNNs.

The observed significances (from a fully unblinded fit) and their expectations (from the partially unblinded fit in section 12.5) are listed in table 12.10. The observed significances are lower than expected. The significance for the LL polarization in the WW-cmf drops to almost 0. Only for the LX polarization in the WW-cmf, the measurement exceeds the $3\,\sigma$ threshold which are necessary to claim evidence. The observed normalizations of the longitudinal polarization states in figure 12.18 are consistent with the corresponding low significances. While the observed normalization of the transverse polarization is higher than expected, the longitudinal polarization states are scaled down by the fit. In the following paragraph, the observed normalizations and the low observed significance are explained by the differential data distributions measured in the signal region. This explanation is followed by a paragraph about the observed pulls of the nuisance parameters.

Explanation of the Observed Significances

The pre-fit and post-fit distributions for the LL in WW-cmf measurement are shown in figure 12.19. The total number of observed events exceeds the prediction in each histogram. The difference between the observed and expected number of events differs the most for

Table 12.10: The observed significances of the longitudinal polarization states. The signficances in brackets result from the partially unblinded fit in table 12.4.

	WW-cmf	pp-cmf
LL significance	$0.0825\,\sigma$ $(1.512\,\sigma)$	$0.742\,\sigma$ $(0.937\,\sigma)$
LX significance	$3.409\,\sigma$ $(4.557\,\sigma)$	$2.476\,\sigma$ $(3.975\,\sigma)$

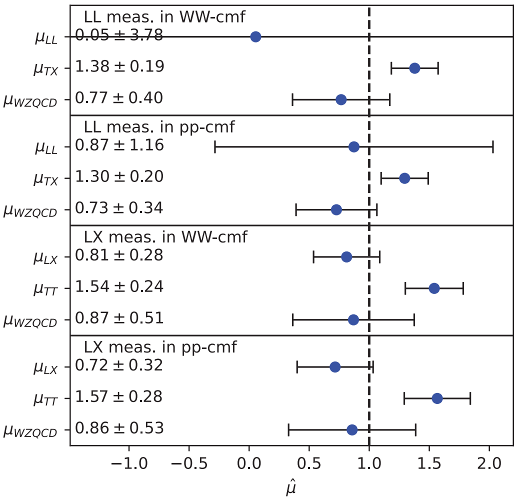
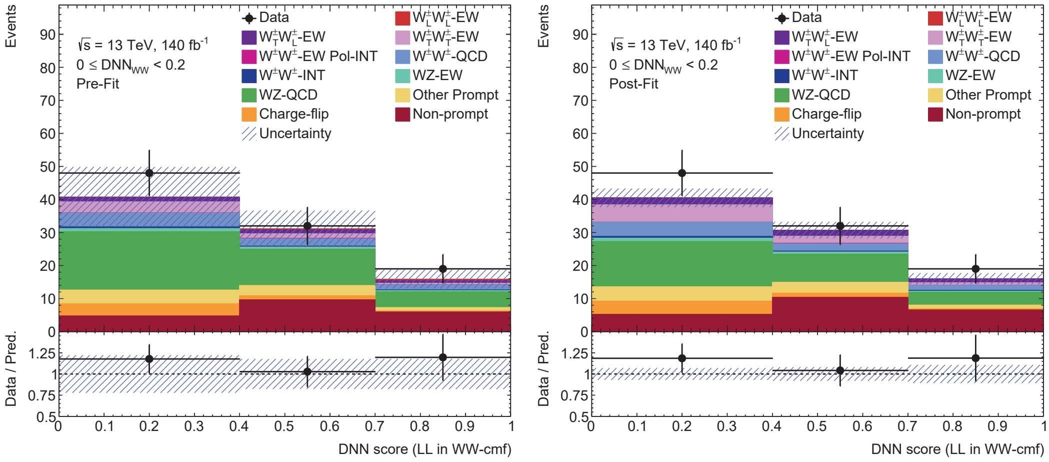


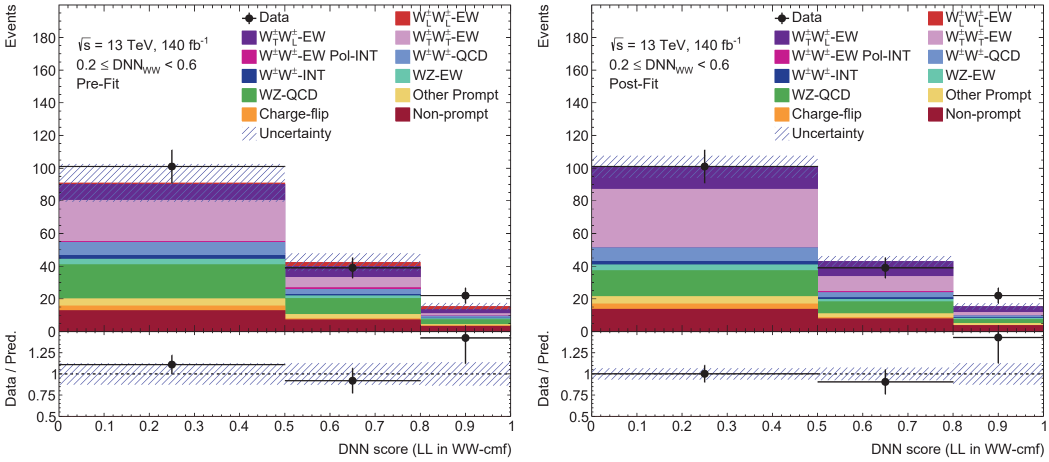
Figure 12.18: The μ_{LL}/μ_{LX}, μ_{TX}/μ_{TT}, and $\mu_{\text{WZ-QCD}}$ normalization resulting from the unblinded significance fits.

the histograms with the lowest and highest $\text{DNN}_{W^\pm W^\pm}$ score. An equivalent difference is also seen for the other polarization measurements since they use a similar splitting at the $\text{DNN}_{W^\pm W^\pm}$ score. The dominant impact on the polarization measurement originates from the histogram in figure 12.19e with $\text{DNN}_{W^\pm W^\pm}$-score > 0.6 which has the highest purity in $W^\pm W^\pm jj$-EW events. This histogram shows a clear trend in the ratio of data over Standard Model prediction. More data than predicted is measured for low DNN_{pol} values and less than predicted for high DNN_{pol} values. Thus, the TX polarization state is scaled up and the LL state is scaled down to match the observed data in the post-fit plot in figure 12.19f. This trend in the observed data explains the almost vanishing LL significance in the likelihood fit. Since the trend follows a smooth curve over the entire histogram, the question arises whether this is just a fluctuation in data or an actual mismodeling. As derived in the study of the network decision-making in section 10.3.2, the decisions by the polarization networks in the $W^\pm W^\pm$-cmf highly depend on the $m_{o1}^{ll,MET}$ kinematic. Furthermore, figure 12.4 shows that the $m_{o1}^{ll,MET}$ kinematic has one of the most dominant correlations with output score of the polarization networks. The pre-fit distribution of $m_{o1}^{ll,MET}$ in the signal region is shown in figure 12.20a. This observable has one of the most significant shape differences between the data and the Standard Model prediction, which also seems to be directly transferred to the distributions of the polarization DNNs. However, this trend is not observed in the Low-m_{jj} control region in figure 12.20b. Thus, a general mismodeling of $m_{o1}^{ll,MET}$ is unlikely.

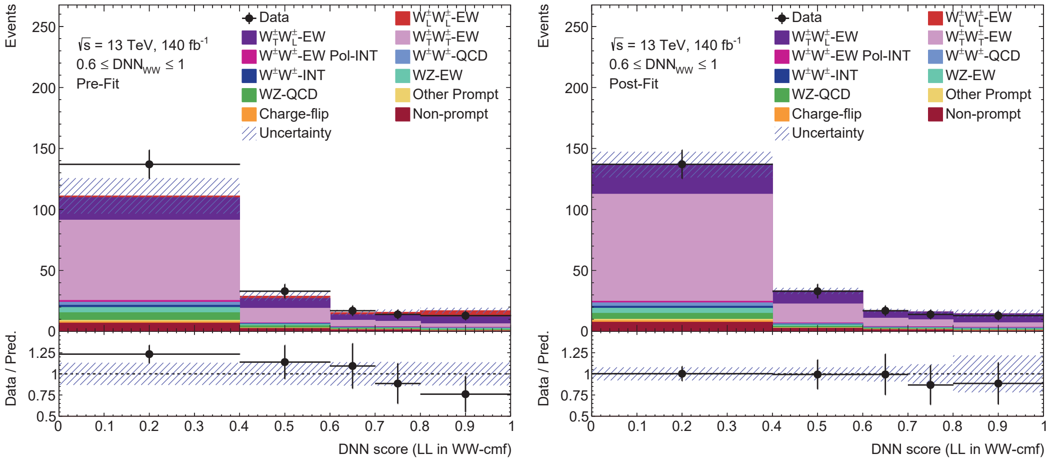
Furthermore, a general mismodeling is contradicted by the validation of the polarization DNNs against measured data in the Low-m_{jj} in section 10.2. In this control region, no significant deviation between observation and measurement is seen. Furthermore, for the signal region histograms figures 12.19a and 12.19c with $\text{DNN}_{W^\pm W^\pm}$-score < 0.6, the trend in the ratio of data to prediction is not visible and the number of measured events with a high polarization DNN score is even above the Standard Model prediction. Since the trend is only observed in histogram figure 12.19e $\text{DNN}_{W^\pm W^\pm}$-score > 0.6, a general mismodeling by polarization DNN can essentially be ruled out. Two options remain to explain the difference between data and Standard Model prediction:



(a) DNN$_{W^\pm W^\pm}$-score ≤ 0.2 pre-fit **(b)** DNN$_{W^\pm W^\pm}$-score ≤ 0.2 post-fit

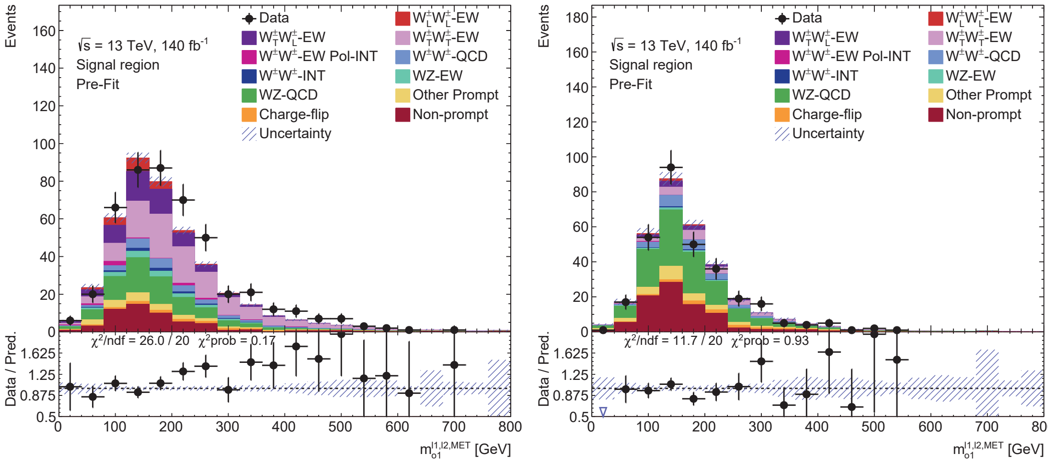


(c) $0.2 <$ DNN$_{W^\pm W^\pm}$-score ≤ 0.6 pre-fit **(d)** $0.2 <$ DNN$_{W^\pm W^\pm}$-score ≤ 0.6 post-fit



(e) $0.6 <$ DNN$_{W^\pm W^\pm}$-score ≤ 1 pre-fit **(f)** $0.6 <$ DNN$_{W^\pm W^\pm}$-score ≤ 1 post-fit

Figure 12.19: The pre- and post-fit signal region distributions for the LL in WW-cmf measurement. The polarization states are simulated by Sherpa. As described in section 12.3 the signal region is split according to the DNN$_{W^\pm W^\pm}$ score. The shown histograms result from the application of the DNN$_{pol}$ trained for the LL polarization state in the WW-cmf.



(a) Signal region

(b) Low-m_{jj} control region

Figure 12.20: Pre-fit comparison of the predicted and measured distribution in $m_{o1}^{ll,MET}$ in the signal and Low-m_{jj} control region.

1. A statistical fluctuation of the measured data caused a trend in the $m_{o1}^{ll,MET}$ distribution. The fact that this is the observable that correlates most strongly with polarization state is merely coincidence.

2. The Standard Model prediction overestimates the longitudinal polarization fraction. As a result, the difference between prediction and data only appears in the phase space region, which is dominated by the $W^{\pm}W^{\pm}jj$-EW contribution.

More data is required to make a justified decision between these two options. However, the following discussion of the measured significances and cross-sections remains valid for both cases.

The observed signal region distributions of the LX measurement in the WW-cmf are shown in figure 12.22 and are similar to the histograms of the LL measurement in figure 12.19. The similarities originate from the correlation between the network decisions shown in figure 12.21. Especially the networks trained in the same reference frame give similar scores. Due to the correlation between the two measurements in the WW-cmf, the same trend of the observed data appears in the histogram of the LX measurement with the highest $W^{\pm}W^{\pm}jj$-EW purity. To cover this trend in figure 12.22e the TT polarization state is scaled up and the LX contribution is reduced by the likelihood fit.

The histograms in the partonic center-of-mass frame are shown in figures G.6 and G.7 in the appendix. The LX in pp-cmf measurement shows a trend in the observed data for $\text{DNN}_{W^{\pm}W^{\pm}}$-score > 0.6 which is comparable to the polarization measurement in the WW-cmf. Therefore, the observed significance is also lower than expected. However, for the LL measurement in the pp-cmf, the ratio of data to prediction is relatively flat in the histogram with the highest $\text{DNN}_{W^{\pm}W^{\pm}}$ score. Therefore, the $0.742\,\sigma$ significance of the LL in pp-cmf is close to the expectation of $0.937\,\sigma$. Since the correlation between the network decisions decreases for different reference frames, the difference in the data distribution is reasonable. In addition, the LL in pp-cmf measurement also stands out in the $\text{DNN}_{W^{\pm}W^{\pm}}$ output values, which are used to split the signal region into histograms: The histogram with highest $W^{\pm}W^{\pm}jj$-EW purity starts for LL in pp-cmf at $\text{DNN}_{W^{\pm}W^{\pm}}$-score $= 0.5$. This is a looser selection than for the other polarization measurements leading to a higher number of events in this signal region histogram. The higher number of events could reduce the impact of statistical fluctuations.

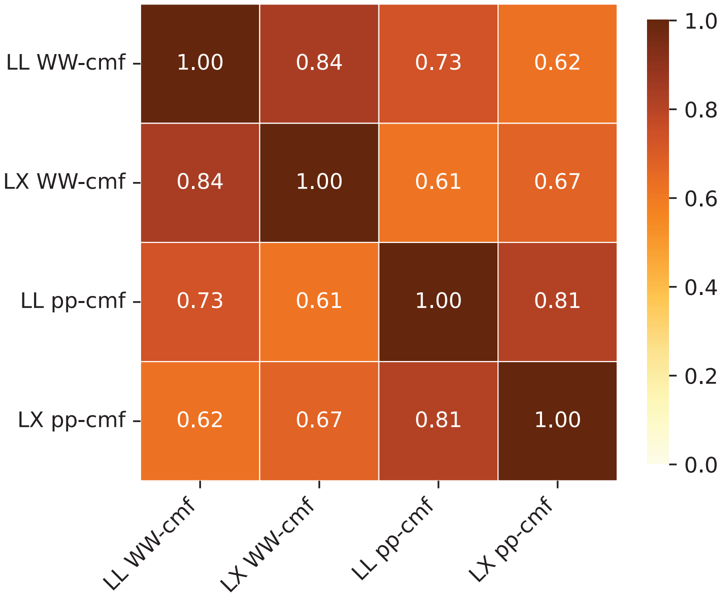
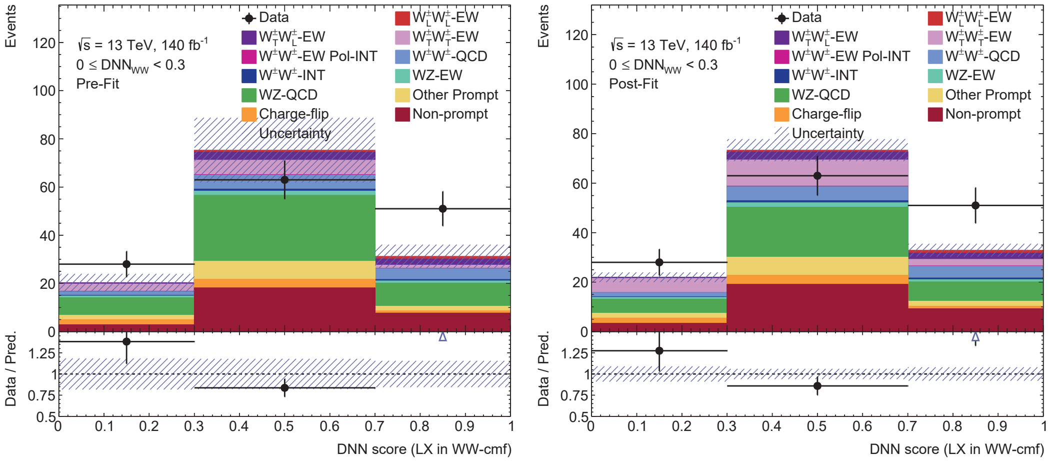


Figure 12.21: The correlation of the polarization DNN scores evaluated for $W^{\pm}W^{\pm}jj$-EW events in the signal region.

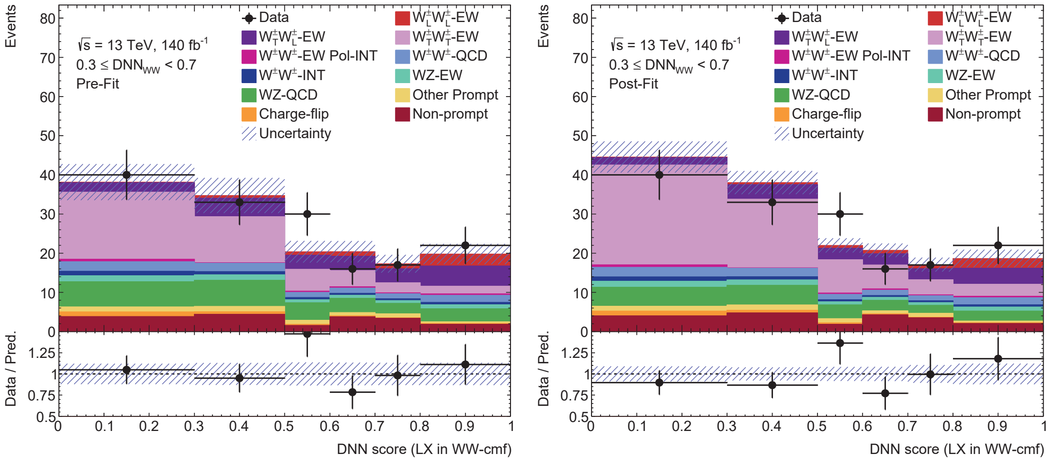
Pull of Nuisance Parameters

The unblinding of the signal region and Low-m_{jj} results in a pull of nuisance parameters. The pulls for the WW-cmf are shown in figure 12.23. The pulls in the pp-cmf are almost identical and therefore moved to the appendix in figure G.8. As observed in the background-only fit in section 12.5.1, the non-prompt background is pulled to a higher cross-section to cover the measured data. Besides the uncertainties of the fake factor method, the pulls in $\boldsymbol{\theta}$ are dominated by the theory scale nuisance parameters of the histogram with the lowest $\mathrm{DNN}_{W^{\pm}W^{\pm}}$ score. The Standard Model prediction in this histogram is below the measured number of events. Since the $W^{\pm}W^{\pm}jj$-EW contribution in this histogram is low, an additional up variation is required to narrow the gap to the measured data. The theory scale uncertainties have a significant normalization effect on the $W^{\pm}W^{\pm}jj$-EW contribution and are uncorrelated between the histograms. Therefore, the theory scale nuisance parameters are pulled up to further increase the $W^{\pm}W^{\pm}jj$-EW contribution in the histogram with the lowest $\mathrm{DNN}_{W^{\pm}W^{\pm}}$ score.

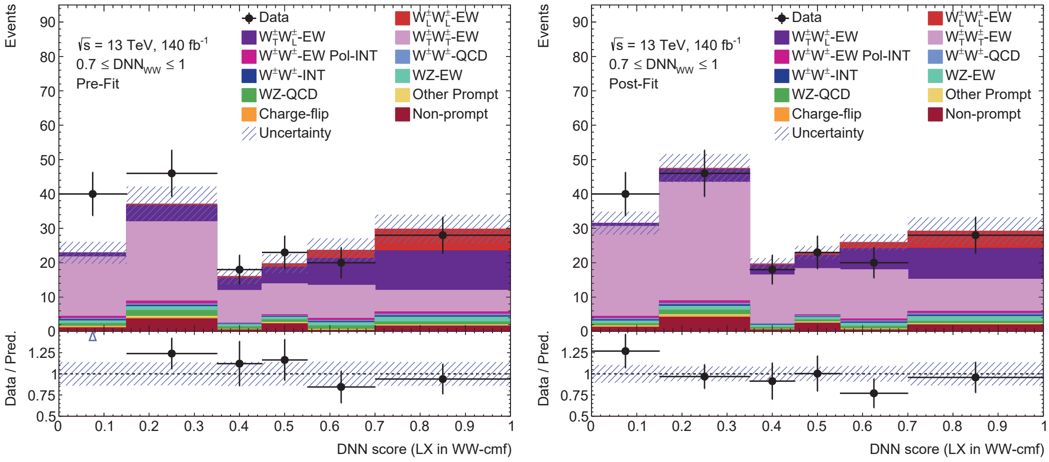
In summary, the measured significances are lower than expected. A statistical fluctuation in the measured data causes the drop of the significances. However, a significance of $3\,\sigma$ is observed for the single boson polarization LX in the center-of-mass frame of the two $W^{\pm}$ bosons. Thus, for the first time, this analysis can claim evidence for the existence of longitudinal polarization in same-charged electroweak $W^{\pm}W^{\pm}$ scattering.



(a) DNN$_{W^\pm W^\pm}$-score ≤ 0.3 pre-fit **(b)** DNN$_{W^\pm W^\pm}$-score ≤ 0.3 post-fit

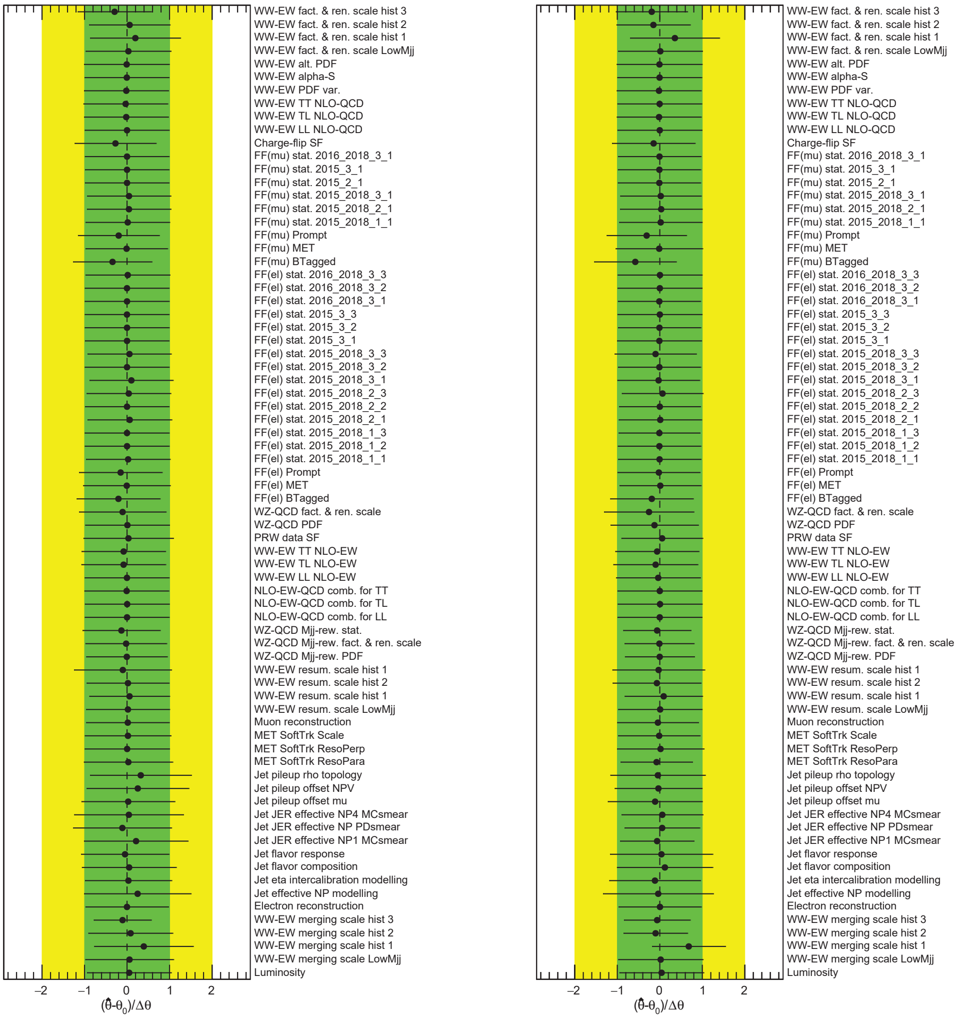


(c) $0.3 <$ DNN$_{W^\pm W^\pm}$-score ≤ 0.7 pre-fit **(d)** $0.3 <$ DNN$_{W^\pm W^\pm}$-score ≤ 0.7 post-fit



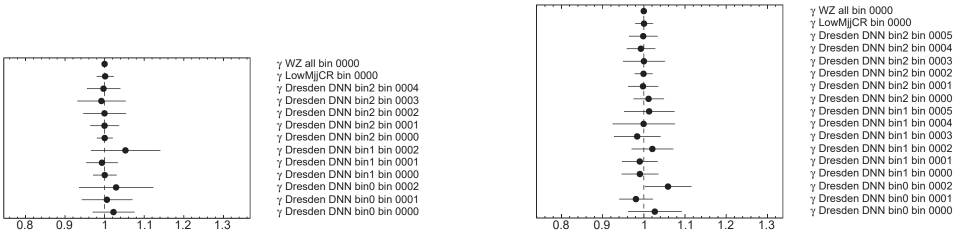
(e) $0.7 <$ DNN$_{W^\pm W^\pm}$-score ≤ 1 pre-fit **(f)** $0.7 <$ DNN$_{W^\pm W^\pm}$-score ≤ 1 post-fit

Figure 12.22: The pre- and post-fit signal region distributions for the LX in WW-cmf measurement. The polarization states are simulated by SHERPA. As described in section 12.3 the signal region is split according to the DNN$_{W^\pm W^\pm}$ score. The shown histograms result from the application of the DNN$_{pol}$ trained for the LX polarization state in the WW-cmf.



(a) $\boldsymbol{\theta}$ of LL fit in WW-cmf

(b) $\boldsymbol{\theta}$ of LX fit in WW-cmf



(c) γ of LL fit in WW-cmf

(d) γ of LX fit in WW-cmf

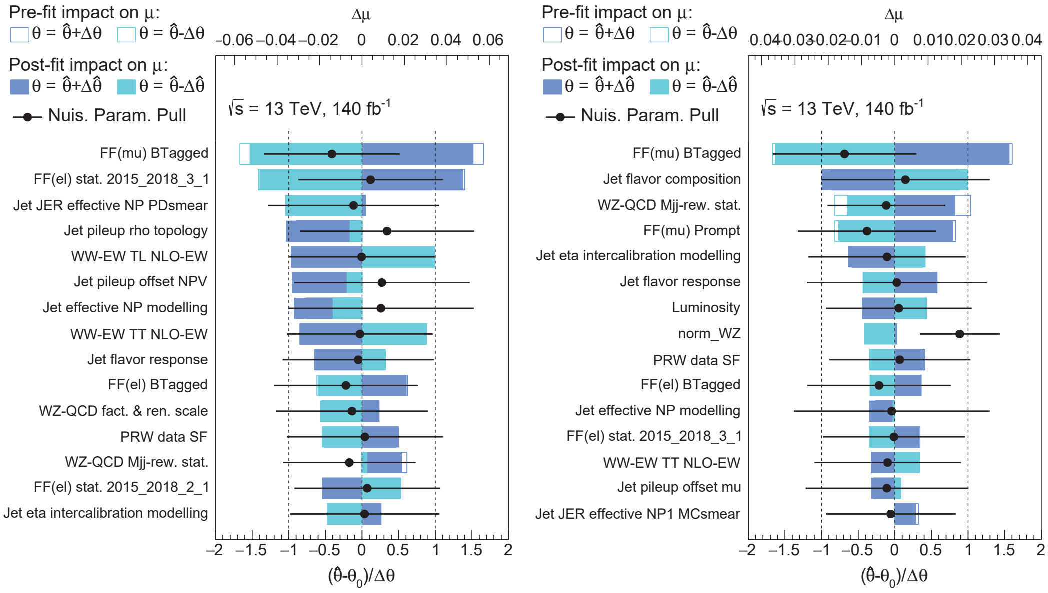
Figure 12.23: The pulls of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ for the partially unblinded background-only significance fits in the WW-cmf. The fit parameter μ_{LL}/μ_{LX} is fixed to 0. The fit is derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74.

12.6.2 Observed Fiducial Cross-Section

As described in section 12.5.2, only the shape impact of the signal theory uncertainties are considered for the fiducial cross-section measurement. The normalizations derived by the likelihood fit to measured data are listed in table 12.11. The measured normalizations are below the expected $\mu = 1$ normalization. The small normalizations are consistent with the low observed significances discussed in the previous section. However, considering the uncertainties, the observed normalizations still agree with the Standard Model prediction. The breakdown of the uncertainties is consistent with the fit of the expected normalization in table 12.6. The same applies to the ranking of the systematic uncertainties in figures 12.24 and G.9.

Table 12.11: The observed normalization of the LL/LX polarization in the WW-cmf and pp-cmf. The uncertainties are split into the statistical uncertainty of the data, statistical modeling uncertainty, systematic modeling uncertainty, and luminosity uncertainty.

	Signal strength	Uncertainty breakdown				
		stat.	mod. stat.	mod. syst.	exp. syst.	lumi.
LL in WW-cmf	0.032 ± 0.699 (tot.)	± 0.688	± 0.143	± 0.050	± 0.103	± 0.005
LX in WW-cmf	0.771 ± 0.255 (tot.)	± 0.246	± 0.054	± 0.019	± 0.051	± 0.010
LL in pp-cmf	0.848 ± 1.178 (tot.)	± 1.146	± 0.251	± 0.119	± 0.169	± 0.012
LX in pp-cmf	0.679 ± 0.296 (tot.)	± 0.285	± 0.064	± 0.042	± 0.059	± 0.008



(a) Ranking of μ_{LL} measurement in WW-cmf

(b) Ranking of μ_{LX} measurement in WW-cmf

Figure 12.24: The ranking plots for the measured fiducial cross-section fits in the WW-cmf.

The measured fiducial cross-section and the corresponding upper limit for LL are listed in table 12.12 and table 12.13. Due to the statistical fluctuation in the measured data, the measured upper limit of LL in WW-cmf is even lower than expected. This tight upper limit covers the Standard Model prediction of $\sigma^{fid}_{pred}(\mathrm{LL}) = 0.33\,\mathrm{fb}$ and rejects alternative theories with $\sigma^{fid}_{pred}(\mathrm{LL}) > 0.50\,\mathrm{fb}$. Therefore, in addition to the evidence of the longitudinal polarization in $W^{\pm}W^{\pm}jj$-EW scattering, this analysis provides further constraints for theories beyond the Standard Model.

Table 12.12: The observed fiducial cross-section of the LL/LX polarization in the WW-cmf and pp-cmf. The uncertainties are split into the statistical uncertainty of the data, statistical modeling uncertainty, systematic modeling uncertainty, and the luminosity uncertainty.

	Fid. cross-section [fb]	Uncertainty breakdown [fb]				
		stat.	mod. stat.	mod. syst.	exp. syst.	lumi.
LL in WW-cmf	0.011 ± 0.231 (tot.)	±0.227	±0.047	±0.017	±0.034	±0.002
LX in WW-cmf	1.044 ± 0.345 (tot.)	±0.333	±0.073	±0.026	±0.069	±0.013
LL in pp-cmf	0.180 ± 0.250 (tot.)	±0.243	±0.053	±0.025	±0.036	±0.003
LX in pp-cmf	0.964 ± 0.420 (tot.)	±0.405	±0.091	±0.060	±0.083	±0.012

Table 12.13: The measured 95% CL upper limit of the LL polarization in the WW-cmf and pp-cmf.

	Fiducial cross-section upper limit [fb]
LL in WW-cmf	0.501
LL in pp-cmf	0.669

12.7 Results with Polarization Modeling by MadGraph

The results in the previous section are derived with the polarization prediction simulated by Sherpa. To ensure the validity of the given results, the likelihood fit is performed also for the polarization prediction by MadGraph. The observed and expected significances and fiducial cross-sections are listed in tables 12.14 and 12.15. The expected significances are very consistent between the two generators. The expected significance of LX is slightly higher for Sherpa due to the higher TL cross-section discussed in section 8.1.1. This difference is also seen in the expected fiducial cross-sections in table 12.15.

A simulation by MadGraph leads to higher significances and fiducial cross-sections in the measured data. The MadGraph polarization prediction differs slightly from the polarization simulated by Sherpa. Thus, using different generators leads to a systematic difference in the fit results. However, the observed significances are similar. The nominal results of this thesis are provided by Sherpa which provides the lower significances. A similar difference can be seen for the fiducial cross-section. However, the cross-section uncertainties are almost identical and therefore not generator-dependent.

Since the simulation by Sherpa can perform jet-merging and therefore include higher-order QCD effects, Sherpa is chosen to produce the nominal results of this analysis. Future analyses could benefit from a more detailed comparison of the polarization modeling by Sherpa and MadGraph.

Table 12.14: The observed (expected) significances derived with polarization modeling by Sherpa and MadGraph.

	Sherpa	MadGraph
LL in WW-cmf	0.085 (1.512)	0.301 (1.503)
LX in WW-cmf	3.408 (4.557)	3.469 (4.349)
LL in pp-cmf	0.741 (0.938)	1.051 (0.904)
LX in pp-cmf	2.475 (3.975)	2.760 (3.780)

Table 12.15: The observed (expected) cross-section derived by polarization modeling by Sherpa and MadGraph.

	Sherpa	MadGraph
LL in WW-cmf	0.011 ± 0.231 (0.330 ± 0.232)	0.072 ± 0.231 (0.320 ± 0.228)
LX in WW-cmf	1.044 ± 0.345 (1.354 ± 0.340)	1.129 ± 0.350 (1.296 ± 0.340)
LL in pp-cmf	0.180 ± 0.250 (0.212 ± 0.234)	0.268 ± 0.259 (0.204 ± 0.234)
LX in pp-cmf	0.964 ± 0.421 (1.420 ± 0.392)	1.156 ± 0.434 (1.348 ± 0.391)

12.8 Comparison with CMS Results

In 2021, the CMS experiment published its $W^\pm W^\pm$ polarization analysis based on Run 2 LHC data in Ref. [3]. The object and event selection of the CMS analysis is similar to the ATLAS analysis. Thus, the expected 190 $W^\pm W^\pm jj$-EW events are close to the expected number of events presented in this thesis. The expected and observed ATLAS and CMS results are compared in table 12.16. The comparison of the expected results shows higher significances and more stringent 95% CL limits for the ATLAS experiment. The better expected performance of the ATLAS analysis can have multiple reasons. The modeling of the polarization interference, the two-dimensional bin optimization, the consideration of resummation and merging scale uncertainties, and especially the application of polarized higher-order corrections were not included in the CMS analysis. Another important difference between the two analyses is the rejection of non-prompt events. The predicted data in the signal region of the CMS analysis are only 36% $W^\pm W^\pm jj$-EW events but 40% non-prompt events with high uncertainty. Due to the better non-prompt rejection, the signal region of the ATLAS analysis has a $W^\pm W^\pm jj$-EW purity of 53% and non-prompt fraction of 13% resulting in a more minor total uncertainty. Therefore, the quality of the results presented in this thesis is not solely based on the new analysis techniques but is the result of the joint work within the ATLAS experiment.

Table 12.16: Comparison of the observed (expected) results of the longitudinally polarized $W^\pm W^\pm jj$-EW scattering measured at the ATLAS and CMS experiment.

	ATLAS	CMS
Significance LL in WW-cmf	$0.082\,\sigma$ $(1.5\,\sigma)$	not published
Significance LX in WW-cmf	$3.4\,\sigma$ $(4.6\,\sigma)$	$2.3\,\sigma$ $(3.1\,\sigma)$
95% CL upper limit of LL in WW-cmf	0.50 fb (0.75 fb)	1.17 fb (0.88 fb)
Significance LL in pp-cmf	$0.74\,\sigma$ $(0.94\,\sigma)$	not published
Significance LX in pp-cmf	$2.5\,\sigma$ $(4.0\,\sigma)$	$2.6\,\sigma$ $(2.9\,\sigma)$
95% CL upper limit of LL in pp-cmf	0.67 fb (0.66 fb)	1.06 fb (0.85 fb)

12.9 Projection for Future LHC Runs

Although this thesis represents significant progress in the study of longitudinal $W^\pm$ boson scattering, the observation of this process with more than $5\,\sigma$ is not achieved yet. Since the analysis is dominated by the statistical uncertainty of the measured data, the precision of polarization analyses will improve in the next LHC runs. The plan of the LHC project is summarized in figure 12.25. Parallel to the finalization of this thesis, Run 3 of the LHC is ongoing, and the first analyses were published with Run 3 data. Run 3 is already expected to provide a higher number of produced events than Run 2, but the subsequent High-Lumi-LHC (HL-LHC) phase will far exceed the current amount of measured data. The exact runtime of the HL-LHC has not yet been decided, but it could deliver a luminosity of up to $4000\,\text{fb}^{-1}$. Since measuring longitudinal vector boson scattering is a potential goal of the HL-LHC experiments, this thesis can provide valuable input for the HL-LHC planning discussions. The expected result of the partially unblinded fit in section 12.5 are projected onto measurements with luminosities of up to $4000\,\text{fb}^{-1}$. The measured and Asimov data are scaled up to the expected luminosities. The relative impact of the modeling, experimental, and luminosity uncertainties are kept constant, even though they result partially from the low measured number of events discussed in section 12.5.2. Thus, the following projection represents the lower limit of the expected significance.

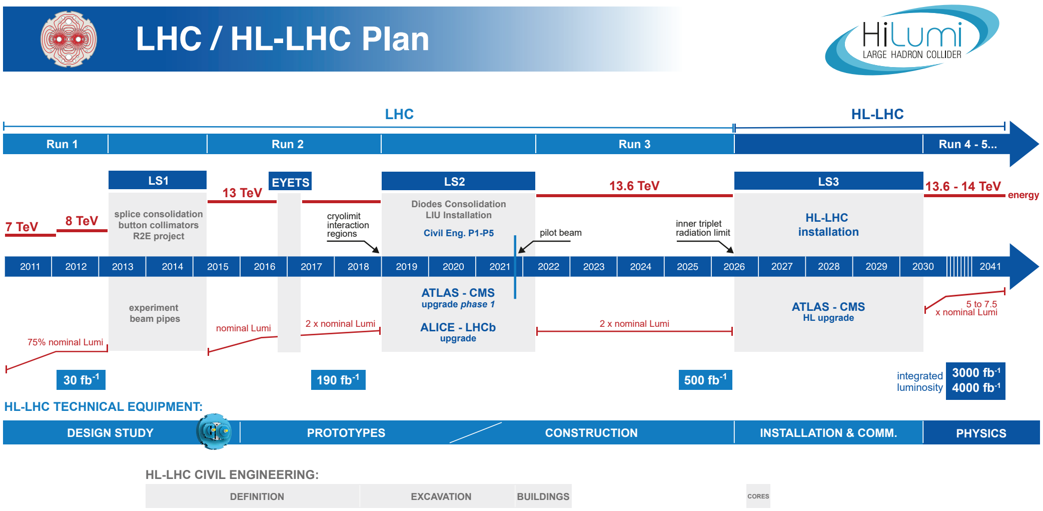
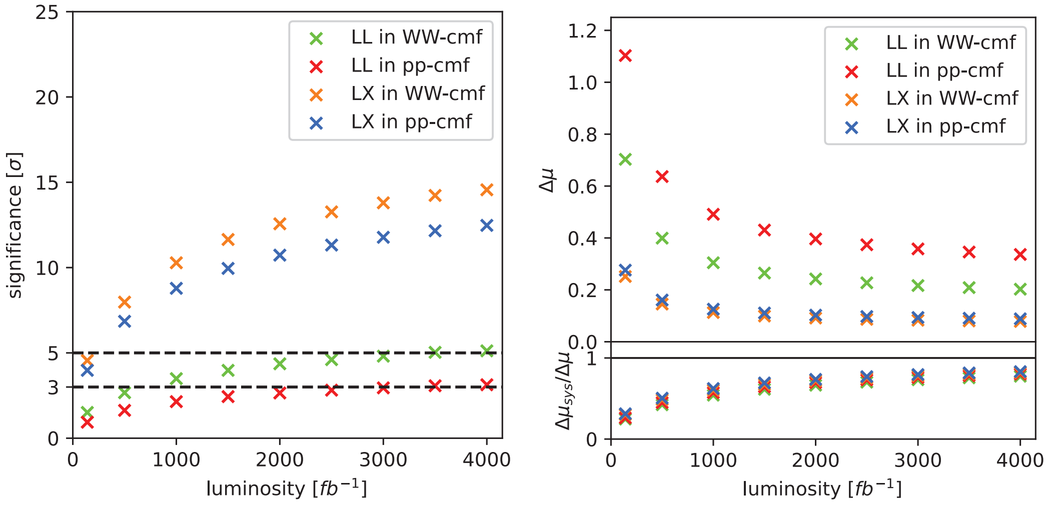

LHC / HL-LHC Plan
HiLumi
LARGE HADRON COLLIDER
LHC
HL-LHC
Run 1
Run 2
Run 3
Run 4 - 5...
LS1
EYETS
LS2
LS3
7 TeV
8 TeV
13 TeV
13.6 TeV
13.6 - 14 TeV
energy
splice consolidation
button collimators
R2E project
cryolimit
interaction
regions
Diodes Consolidation
LIU Installation
Civil Eng. P1-P5
pilot beam
inner triplet
radiation limit
HL-LHC
installation
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2041
experiment
beam pipes
ATLAS - CMS
upgrade phase 1
ALICE - LHCb
upgrade
ATLAS - CMS
HL upgrade
75% nominal Lumi
nominal Lumi
2 x nominal Lumi
2 x nominal Lumi
5 to 7.5
x nominal Lumi
30 fb^{-1}
190 fb^{-1}
500 fb^{-1}
integrated
luminosity
3000 fb^{-1}
4000 fb^{-1}
HL-LHC TECHNICAL EQUIPMENT:
DESIGN STUDY
PROTOTYPES
CONSTRUCTION
INSTALLATION & COMM.
PHYSICS
HL-LHC CIVIL ENGINEERING:
DEFINITION
EXCAVATION
BUILDINGS
CORES

Figure 12.25: Timeline of the LHC and the plan for the HL-LHC project presented in Ref. [114]

The expected significances for higher luminosities are shown in figure 12.26. With 500 fb^{-1}, the observation of the single boson polarization LX with more than 5 σ is expected. Thus, the Run 3 of the LHC could deliver new results for the polarization measurement. For the LL measurement in the WW-cmf, the Run 3 could provide evidence with 3 σ but for a 5 σ observation the HL-LHC with about 3000 fb^{-1} to 4000 fb^{-1} is required. Therefore, the potential discovery of the fully longitudinal vector boson scattering must be considered for the planning of the HL-LHC phase.

The expected uncertainties of the signal strength are shown in figure 12.26b. As soon as 500 fb^{-1} data is available, the uncertainties are expected to drop by almost a factor of two. As the $\Delta\mu_{sys}/\Delta\mu$ subplot of figure 12.26b indicates, the systematic uncertainties dominate the projected fit for a luminosity higher than 1000 fb^{-1}. Therefore, the total uncertainties of the signal strength are converging towards the systematic uncertainty. The dominance of systematic uncertainties results from the constant relative systematic uncertainty chosen for the projection. However, the statistical uncertainty of the data-driven backgrounds and the data-driven reweighting of the $W^{\pm}Z$ background will improve with more statistics. Therefore, the significance and relative uncertainties are only conservative estimates, and future analyses are expected to exceed the projections.



(a) Significance **(b)** Signal strength uncertainty

Figure 12.26: The projection of the expected significances and signal strength uncertainties of section 12.5.1 for LHC runs with higher luminosity. In figure 12.26b, the contribution of the systematic uncertainties is shown in the lower subplot.

Chapter 13

Conclusion

The longitudinally polarized $W^{\pm}W^{\pm}jj$-EW scattering is one of the most interesting processes to study electroweak symmetry breaking at the LHC. The longitudinal polarization states of the $W^{\pm}$ and Z boson arise from the components of the Higgs field, which would result in massless Goldstone bosons in an unbroken symmetry. Furthermore, a Standard Model without the Higgs boson predicts unphysical probabilities for the longitudinally polarized $W^{\pm}W^{\pm}jj$-EW scattering of more than 100%. The Higgs boson exchange leads to the necessary cancelation of Feynman diagrams to solve this unitarity violation. Therefore, experimental results for the polarized vector boson scattering (VBS) can confirm important Standard Model predictions and probe theories beyond the Standard Model.

This thesis is a major contribution to the first ATLAS analysis of the polarization in VBS. The investigated data originates from Run 2 proton-proton-collisions with an integrated luminosity of $140\,\text{fb}^{-1}$ and a center-of-mass energy of $13\,\text{TeV}$. The measurement employs state-of-the-art calculations of polarized cross-sections that only became available recently. For the first time, these predictions include approximate NLO-QCD and -EWK effects for the individual $W^{\pm}W^{\pm}jj$-EW polarization states. Based on the Standard Model prediction, neural networks are trained to use the kinematics of the entire phase space to extract the polarization information. Two sets of neural networks are optimized for a phase space enriched with $W^{\pm}W^{\pm}jj$-EW events. The first set separates the $W^{\pm}W^{\pm}jj$-EW contribution from the remaining backgrounds, and the second set is trained to distinguish the longitudinal and transverse polarization states. The resulting networks are extensively studied to prevent any application bias and provide state-of-the-art insights into the decision-making process. The two sets of networks are combined into a two-dimensional classification space that separates the longitudinal polarization state, the transverse polarization state, and the non-$W^{\pm}W^{\pm}jj$-EW backgrounds. A two-dimensional bin optimization algorithm is developed which maximizes the expected significance. Furthermore, this thesis presents a new approach to investigate and constrain the impact of theoretical uncertainties originating from missing higher-order calculations. This approach significantly increased the sensitivity of the polarization measurement.

A profile likelihood fit is performed to extract the longitudinally polarized $W^{\pm}W^{\pm}jj$-EW contribution in the measured data. Two separate fits are performed to measure the fully longitudinally polarized scattering (LL) and the scattering with at least one longitudinal polarized $W^{\pm}$ boson (LX). The polarization is defined in the $W^{\pm}W^{\pm}$ center-of-mass frame. This reference frame leads to the maximal unitarity cancelation in tree-level calculations and is thus of the highest interest [50]. The longitudinal polarization states are measured with an observation significance of $0.082\,\sigma$ for LL and $3.4\,\sigma$ for LX. The observed cross-sections and significances are below the expectations due to a statistical fluctuation in the data in the signal region. However, the results are still consistent with the prediction by the Standard Model. For the LX polarization state of the $W^{\pm}W^{\pm}jj$-EW scattering, a fiducial cross-section of $\sigma^{fid}_{obs}(\text{LX}) = 1.04 \pm 0.35\,\text{fb}$ is observed that is in agreement with the Standard Model prediction of $\sigma^{fid}_{pred}(\text{LX}) = 1.35\,\text{fb}$. This analysis provides evidence for the LX polarization state of the $W^{\pm}W^{\pm}jj$-EW process for the first time. The cross-section of the LL polarization state is found to be smaller than 0.50 fb at 95% confidence level. This upper limit covers the Standard Model prediction of $\sigma^{fid}_{pred}(\mathrm{LL}) = 0.33\,\mathrm{fb}$ and puts stringent constraints on theories beyond the Standard-Model that predict higher fiducial cross-sections than 0.50 fb.

The measurements presented in this thesis improve the experimental sensitivity to polarized same-charged $W^{\pm}$ boson scattering compared to the earlier publication by the CMS collaboration [3]. The observed and especially the expected significances and cross-sections exceed the ones published by CMS. Thus, the diligent implementation of the theory prediction, the detailed study of the theory uncertainties, and the application of state-of-the-art machine learning techniques presented throughout this thesis could improve the experimental sensitivity to polarized VBS and put ATLAS in a leading position in searches for this rare process.

Based on the presented analysis strategy, future analyses have to deal with challenges beyond the scope of this thesis. Simulating polarized VBS at higher-order would further improve the Standard Model prediction. Polarized predictions at higher-order would replace the used higher-order corrections and reduce the scale uncertainties that dominate the theoretical uncertainties. The recent version 3 of SHERPA is an important step in this direction, but a simulation of polarized VBS at NLO is not yet possible for any of the existing event generators. The reconstruction of the polarization information requires the use of multivariate techniques. In this thesis, neural networks are trained as binary classifiers to separate the polarization states. In the future, this could be replaced by polarization taggers trained on polarization fractions instead of binary categories. The concept of polarization taggers is recently studied in Ref. [115].

However, the main restriction for polarization analysis in VBS is the measured number of events. The analysis strategy presented in this thesis will allow the observation of the LX polarization state in the Run 3 of the LHC. The Run 3 is expected to deliver an additional $250\,\mathrm{fb}^{-1}$ of proton-proton collisions to the ATLAS experiment by the summer of 2026. Furthermore, the projection of the polarization analysis onto the planned High-Lumi-LHC phase shows the potential for the $5\,\sigma$ observation of fully longitudinally polarized $W^{\pm}W^{\pm}jj$-EW scattering. Therefore, this thesis represents an important step towards a bright future of polarization analyses that will potentially revolutionize our understanding of physics.

Part III

Appendix

Appendix A

Monte Carlo Event Generator Samples

Table A.1: Information about Monte Carlo generated samples used in this thesis which were not used in Ref. [8]

DSID	Generator	Cross-section [pb]	Filter efficiency
Polarized MadGraph sample for $W^{\pm}W^{\pm}jj$-EW			
506202	MadGraph(v.2.9.5.atlas2)+Pythia8(v.245p3.lhcb7)+EvtGen(v.1.7.0)	0.031062	1.0
506207	MadGraph(v.2.9.5.atlas2)+Pythia8(v.245p3.lhcb7)+EvtGen(v.1.7.0)	0.018046	1.0
506201	MadGraph(v.2.9.5.atlas2)+Pythia8(v.245p3.lhcb7)+EvtGen(v.1.7.0)	0.018454	1.0
506203	MadGraph(v.2.9.5.atlas2)+Pythia8(v.245p3.lhcb7)+EvtGen(v.1.7.0)	0.0019852	1.0
506204	MadGraph(v.2.9.5.atlas2)+Pythia8(v.245p3.lhcb7)+EvtGen(v.1.7.0)	0.0028838	1.0
506205	MadGraph(v.2.9.5.atlas2)+Pythia8(v.245p3.lhcb7)+EvtGen(v.1.7.0)	0.010732	1.0
506206	MadGraph(v.2.9.5.atlas2)+Pythia8(v.245p3.lhcb7)+EvtGen(v.1.7.0)	0.0093687	1.0
Polarized Sherpa sample for $W^{\pm}W^{\pm}jj$-EW			
700965	Sherpa(v.3.0.0beta1)	0.0023792	1.0
700966	Sherpa(v.3.0.0beta1)	0.0069042	1.0
700968	Sherpa(v.3.0.0beta1)	0.00079496	1.0
700969	Sherpa(v.3.0.0beta1)	0.0015661	1.0
701228	Sherpa(v.3.0.0beta1)	0.0053504	0.2226257
701229	Sherpa(v.3.0.0beta1)	0.015565	0.2216626
701230	Sherpa(v.3.0.0beta1)	0.0017874	0.2227125
701231	Sherpa(v.3.0.0beta1)	0.0035172	0.2230805
Polarized Sherpa for approximate NLO-QCD correction of $W^{\pm}W^{\pm}jj$-EW			
700961	Sherpa(v.3.0.0beta1)	0.003305	1.0
700962	Sherpa(v.3.0.0beta1)	0.0089404	1.0
700963	Sherpa(v.3.0.0beta1)	0.0032598	1.0
700964	Sherpa(v.3.0.0beta1)	0.0088539	1.0
700971	Sherpa(v.3.0.0beta1)	0.011628	1.0
700970	Sherpa(v.3.0.0beta1)	0.014854	1.0
Scale variations of polarized Sherpa sample for $W^{\pm}W^{\pm}jj$-EW			
700998	Sherpa(v.3.0.0beta1)	0.014881	1.0
700999	Sherpa(v.3.0.0beta1)	0.01547	1.0
Sherpa for $W^{\pm}Z$-QCD background			
700601	Sherpa(v.2.2.12.f290b9)	4.661	1.0

Table A.7: The DSID, physics name, and p-tag of the Monte Carlo generator samples used to model the Z+jets background.

DSID	sample name	p-tag
Zee_MGPy8EG		
363147	MGPy8EG_N30NLO_Zee_Ht0_70_CVetoBVeto	p3975
363148	MGPy8EG_N30NLO_Zee_Ht0_70_CFilterBVeto	p3975
363149	MGPy8EG_N30NLO_Zee_Ht0_70_BFilter	p3975
363150	MGPy8EG_N30NLO_Zee_Ht70_140_CVetoBVeto	p3975
363151	MGPy8EG_N30NLO_Zee_Ht70_140_CFilterBVeto	p3975
363152	MGPy8EG_N30NLO_Zee_Ht70_140_BFilter	p3975
363153	MGPy8EG_N30NLO_Zee_Ht140_280_CVetoBVeto	p3975
363154	MGPy8EG_N30NLO_Zee_Ht140_280_CFilterBVeto	p3975

363155	MGPy8EG_N30NLO_Zee_Ht140_280_BFilter	p3975
363156	MGPy8EG_N30NLO_Zee_Ht280_500_CVetoBVeto	p3975
363157	MGPy8EG_N30NLO_Zee_Ht280_500_CFilterBVeto	p3975
363158	MGPy8EG_N30NLO_Zee_Ht280_500_BFilter	p3975
363159	MGPy8EG_N30NLO_Zee_Ht500_700_CVetoBVeto	p3975
363160	MGPy8EG_N30NLO_Zee_Ht500_700_CFilterBVeto	p3975
363161	MGPy8EG_N30NLO_Zee_Ht500_700_BFilter	p3975
363162	MGPy8EG_N30NLO_Zee_Ht700_1000_CVetoBVeto	p3975
363163	MGPy8EG_N30NLO_Zee_Ht700_1000_CFilterBVeto	p3975
363164	MGPy8EG_N30NLO_Zee_Ht700_1000_BFilter	p3975
363165	MGPy8EG_N30NLO_Zee_Ht1000_2000_CVetoBVeto	p3975
363166	MGPy8EG_N30NLO_Zee_Ht1000_2000_CFilterBVeto	p3975
363167	MGPy8EG_N30NLO_Zee_Ht1000_2000_BFilter	p3975
363168	MGPy8EG_N30NLO_Zee_Ht2000_E_CMS_CVetoBVeto	p3975
363169	MGPy8EG_N30NLO_Zee_Ht2000_E_CMS_CFilterBVeto	p3975
363170	MGPy8EG_N30NLO_Zee_Ht2000_E_CMS_BFilter	p3975
ZmumuMGPy8EG		
363123	MGPy8EG_N30NLO_Zmumu_Ht0_70_CVetoBVeto	p3975
363124	MGPy8EG_N30NLO_Zmumu_Ht0_70_CFilterBVeto	p3975
363125	MGPy8EG_N30NLO_Zmumu_Ht0_70_BFilter	p3975
363126	MGPy8EG_N30NLO_Zmumu_Ht70_140_CVetoBVeto	p3975
363127	MGPy8EG_N30NLO_Zmumu_Ht70_140_CFilterBVeto	p3975
363128	MGPy8EG_N30NLO_Zmumu_Ht70_140_BFilter	p3975
363129	MGPy8EG_N30NLO_Zmumu_Ht140_280_CVetoBVeto	p3975
363130	MGPy8EG_N30NLO_Zmumu_Ht140_280_CFilterBVeto	p3975
363131	MGPy8EG_N30NLO_Zmumu_Ht140_280_BFilter	p3975
363132	MGPy8EG_N30NLO_Zmumu_Ht280_500_CVetoBVeto	p3975
363133	MGPy8EG_N30NLO_Zmumu_Ht280_500_CFilterBVeto	p3975
363134	MGPy8EG_N30NLO_Zmumu_Ht280_500_BFilter	p3975
363135	MGPy8EG_N30NLO_Zmumu_Ht500_700_CVetoBVeto	p3975
363136	MGPy8EG_N30NLO_Zmumu_Ht500_700_CFilterBVeto	p3975
363137	MGPy8EG_N30NLO_Zmumu_Ht500_700_BFilter	p3975
363138	MGPy8EG_N30NLO_Zmumu_Ht700_1000_CVetoBVeto	p3975
363139	MGPy8EG_N30NLO_Zmumu_Ht700_1000_CFilterBVeto	p3975
363140	MGPy8EG_N30NLO_Zmumu_Ht700_1000_BFilter	p3975
363141	MGPy8EG_N30NLO_Zmumu_Ht1000_2000_CVetoBVeto	p3975
363142	MGPy8EG_N30NLO_Zmumu_Ht1000_2000_CFilterBVeto	p3975
363143	MGPy8EG_N30NLO_Zmumu_Ht1000_2000_BFilter	p3975
363144	MGPy8EG_N30NLO_Zmumu_Ht2000_E_CMS_CVetoBVeto	p3975
363145	MGPy8EG_N30NLO_Zmumu_Ht2000_E_CMS_CFilterBVeto	p3975
363146	MGPy8EG_N30NLO_Zmumu_Ht2000_E_CMS_BFilter	p3975
ZtautauMadGraphPythia8		
361638	MadGraphPythia8EvtGen_A14NNPDF23LO_Ztautau_lowMll_Np0	p3975
361639	MadGraphPythia8EvtGen_A14NNPDF23LO_Ztautau_lowMll_Np1	p3975
361640	MadGraphPythia8EvtGen_A14NNPDF23LO_Ztautau_lowMll_Np2	p3975
361641	MadGraphPythia8EvtGen_A14NNPDF23LO_Ztautau_lowMll_Np3	p3975
361642	MadGraphPythia8EvtGen_A14NNPDF23LO_Ztautau_lowMll_Np4	p3975
361510	MadGraphPythia8EvtGen_A14NNPDF23LO_Ztautau_Np0	p3975
361511	MadGraphPythia8EvtGen_A14NNPDF23LO_Ztautau_Np1	p3975
361512	MadGraphPythia8EvtGen_A14NNPDF23LO_Ztautau_Np2	p3975
361513	MadGraphPythia8EvtGen_A14NNPDF23LO_Ztautau_Np3	p3975
361514	MadGraphPythia8EvtGen_A14NNPDF23LO_Ztautau_Np4	p3975

Table A.2: The DSID, physics name, and p-tag of the Monte Carlo generator samples used to model the polarized $W^{\pm}W^{\pm}jj$-EW signal and the $W^{\pm}W^{\pm}jj$-INT and $W^{\pm}W^{\pm}jj$-QCD contributions.

DSID	sample name	p-tag
EWSigPolMGPy8EG_LO		
506202	MGPy8_ssWWjj_lep_EW6_LO (unpolarized)	p4252
506207	MGPy8_ssWWjj_lep_TT_EW6_LO (TT in pp-cmf)	p4252
506201	MGPy8_ssWWjj_lep_TT_EW6_WWcmf_LO (TT in WW-cmf)	p4252
506203	MGPy8_ssWWjj_lep_LL_EW6_LO (LL in pp-cmf)	p4252
506204	MGPy8_ssWWjj_lep_LL_EW6_WWcmf_LO (LL in WW-cmf)	p4252
506205	MGPy8_ssWWjj_lep_TL_EW6_LO (TL in pp-cmf)	p4252
506206	MGPy8_ssWWjj_lep_TL_EW6_WWcmf_LO (TL in WW-cmf)	p4252
Sh_300b1_WlvWlvjj		
700965	Sh_300b1_WlvWlvjj_mm_MinTChannel_pol_highMjj	p4252
700966	Sh_300b1_WlvWlvjj_pp_MinTChannel_pol_highMjj	p4252
700968	Sh_300b1_WlvWlvjj_mm_MinTChannel_pol_lowMjj	p4252
700969	Sh_300b1_WlvWlvjj_pp_MinTChannel_pol_lowMjj	p4252
701228	Sh_300b1_WlvWlvjj_mm_TChannel_pol_highMjj_min1Tau	p4252
701229	Sh_300b1_WlvWlvjj_pp_TChannel_pol_highMjj_min1Tau	p4252
701230	Sh_300b1_WlvWlvjj_mm_TChannel_pol_lowMjj_min1Tau	p4252
701231	Sh_300b1_WlvWlvjj_pp_TChannel_pol_lowMjj_min1Tau	p4252
INTSigMGH7EG_LO		
500991	MGH7EG_LO_INT_ssWWjj	p4252
QCDSigMGH7EG_LO		
500990	MGH7EG_LO_EW4_ssWWjj	p4252

Table A.3: The DSID, physics name, and p-tag of the Monte Carlo generator samples used to model the $W^{\pm}Z$ background.

DSID	sample name	p-tag
WZEW_MGPy8EG		
364741	MGPy8EG_NNPDF30NLO_A14NNPDF23LO_lvlljjEW6_SFMinus	p4097
364742	MGPy8EG_NNPDF30NLO_A14NNPDF23LO_lvlljjEW6_SFPlus	p4097
364740	MGPy8EG_NNPDF30NLO_A14NNPDF23LO_lvlljjEW6_OFPlus	p4097
364739	MGPy8EG_NNPDF30NLO_A14NNPDF23LO_lvlljjEW6_OFMinus	p4097
WZQCD_Sherpa222		
364253	Sherpa_222_NNPDF30NNLO_lllv	p4097
WZQCD_Sherpa2212		
700601	Sh_2212_lllv	p4250

Table A.4: The DSID, physics name, and p-tag of the Monte Carlo generator samples used to model the top and tri-boson backgrounds.

DSID	sample name	p-tag
singletop_tV		
410658	PhPy8EG_A14_tchan_BW50_lept_top	p3975
410659	PhPy8EG_A14_tchan_BW50_lept_antitop	p3975
410644	PowhegPythia8EvtGen_A14_singletop_schan_lept_top	p3975
410645	PowhegPythia8EvtGen_A14_singletop_schan_lept_antitop	p3975
410648	PowhegPythia8EvtGen_A14_Wt_DR_dilepton_top	p3975
410649	PowhegPythia8EvtGen_A14_Wt_DR_dilepton_antitop	p3975
410560	MadGraphPythia8EvtGen_A14_tZ_4fl_tchan_noAllHad	p3975
ttV_aMcAtNloPythia8		
410155	aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_ttW	p3975
410218	aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_ttee	p3975
410219	aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_ttmumu	p3975
410220	aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_tttautau	p3975
410081	MadGraphPythia8EvtGen_A14NNPDF23_ttbarWW	p3975
ttZ_aMcAtNloPythia8		
410220	aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_tttautau	p3975
410218	aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_ttee	p3975
410219	aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_ttmumu	p3975

Table A.5: The DSID, physics name, and p-tag of the Monte Carlo generator samples used to model the $V\gamma$ backgrounds.

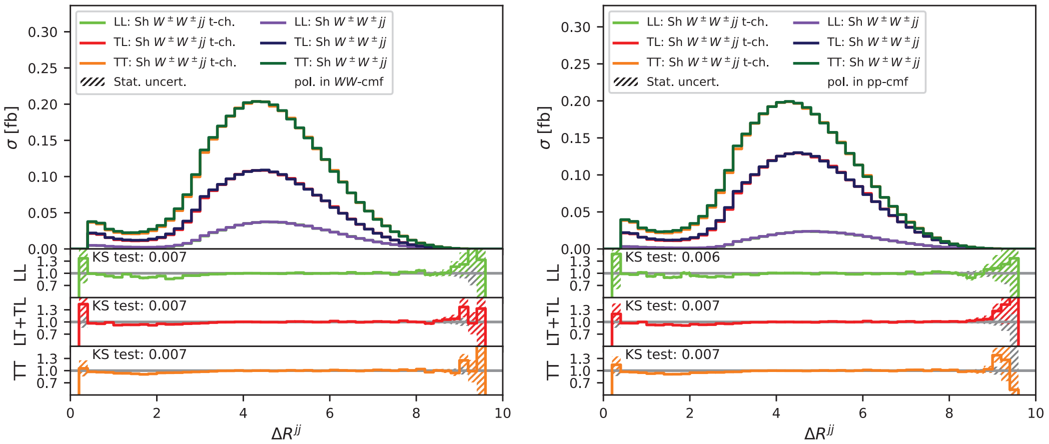
DSID	sample name	p-tag
WgammaSherpa222		
364521	Sherpa_222_NNPDF30NNLO_enugamma_pty_7_15	p3975
364522	Sherpa_222_NNPDF30NNLO_enugamma_pty_15_35	p3975
364523	Sherpa_222_NNPDF30NNLO_enugamma_pty_35_70	p3975
364524	Sherpa_222_NNPDF30NNLO_enugamma_pty_70_140	p3975
364525	Sherpa_222_NNPDF30NNLO_enugamma_pty_140_E_CMS	p3975
364526	Sherpa_222_NNPDF30NNLO_munugamma_pty_7_15	p3975
364527	Sherpa_222_NNPDF30NNLO_munugamma_pty_15_35	p3975
364528	Sherpa_222_NNPDF30NNLO_munugamma_pty_35_70	p3975
364529	Sherpa_222_NNPDF30NNLO_munugamma_pty_70_140	p3975
364530	Sherpa_222_NNPDF30NNLO_munugamma_pty_140_E_CMS	p3975
364531	Sherpa_222_NNPDF30NNLO_taunugamma_pty_7_15	p3975
364532	Sherpa_222_NNPDF30NNLO_taunugamma_pty_15_35	p3975
364533	Sherpa_222_NNPDF30NNLO_taunugamma_pty_35_70	p3975
364534	Sherpa_222_NNPDF30NNLO_taunugamma_pty_70_140	p3975
364535	Sherpa_222_NNPDF30NNLO_taunugamma_pty_140_E_CMS	p3975
363507	Sherpa_222_NNPDF30NNLO_WWZ_3l1v2j_EW6	p4250
363508	Sherpa_222_NNPDF30NNLO_WZZ_4l2j_EW6	p4250
363509	Sherpa_222_NNPDF30NNLO_WZZ_3l1v2j_EW6	p4250
ZgammaSherpa222		
364500	Sherpa_222_NNPDF30NNLO_eegamma_pty_7_15	p3975
364501	Sherpa_222_NNPDF30NNLO_eegamma_pty_15_35	p3975
364502	Sherpa_222_NNPDF30NNLO_eegamma_pty_35_70	p3975
364503	Sherpa_222_NNPDF30NNLO_eegamma_pty_70_140	p3975
364504	Sherpa_222_NNPDF30NNLO_eegamma_pty_140_E_CMS	p3975
364505	Sherpa_222_NNPDF30NNLO_mumugamma_pty_7_15	p3975
364506	Sherpa_222_NNPDF30NNLO_mumugamma_pty_15_35	p3975
364507	Sherpa_222_NNPDF30NNLO_mumugamma_pty_35_70	p3975
364508	Sherpa_222_NNPDF30NNLO_mumugamma_pty_70_140	p3975
364509	Sherpa_222_NNPDF30NNLO_mumugamma_pty_140_E_CMS	p3975
364510	Sherpa_222_NNPDF30NNLO_tautaugamma_pty_7_15	p3975
364511	Sherpa_222_NNPDF30NNLO_tautaugamma_pty_15_35	p3975
364512	Sherpa_222_NNPDF30NNLO_tautaugamma_pty_35_70	p3975
364513	Sherpa_222_NNPDF30NNLO_tautaugamma_pty_70_140	p3975
364514	Sherpa_222_NNPDF30NNLO_tautaugamma_pty_140_E_CMS	p3975

Table A.6: The DSID, physics name, and p-tag of the Monte Carlo generator samples used to model the ZZ background.

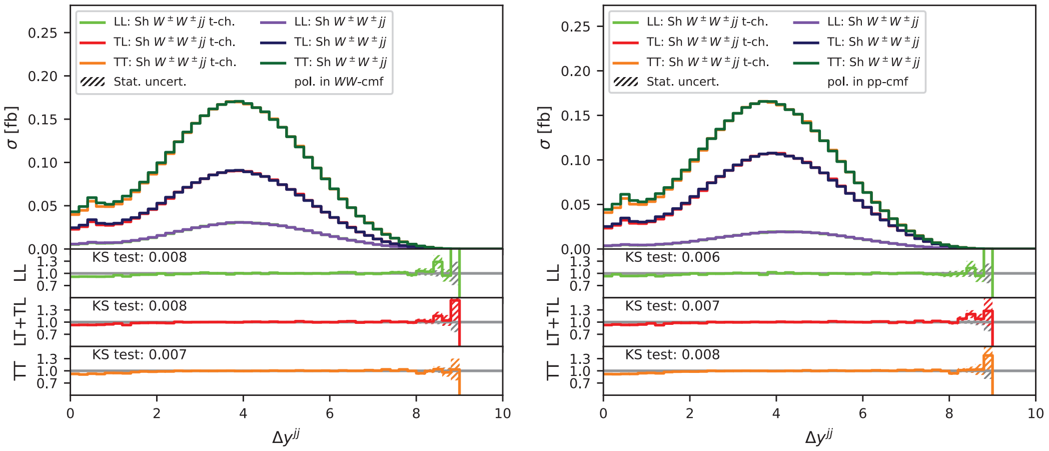
DSID	sample name	p-tag
ZZ_Sherpa		
363356	Sherpa_221_NNPDF30NNLO_ZqqZll	p4097
364250	Sherpa_222_NNPDF30NNLO_llll	p4097
364283	Sherpa_222_NNPDF30NNLO_lllljj_EW6	p4097

Appendix B

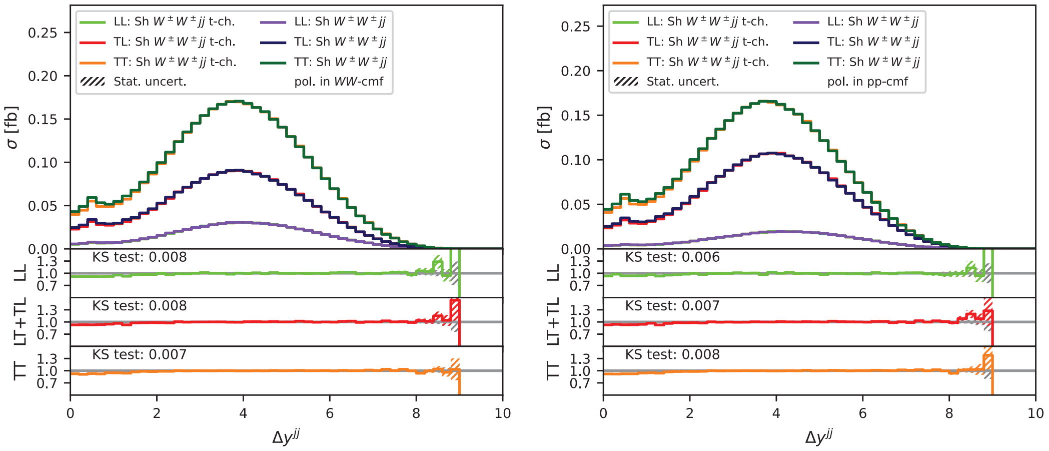
Approximate NLO-QCD Correction



(a) ΔR_{jj} in WW-cmf **(b)** ΔR_{jj} in pp-cmf

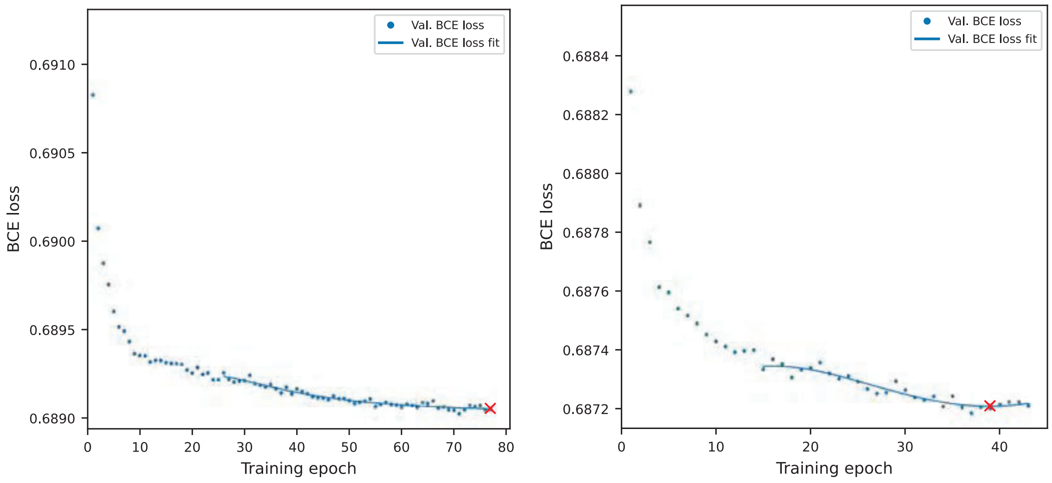


(c) Δy_{jj} in WW-cmf **(d)** Δy_{jj} in pp-cmf

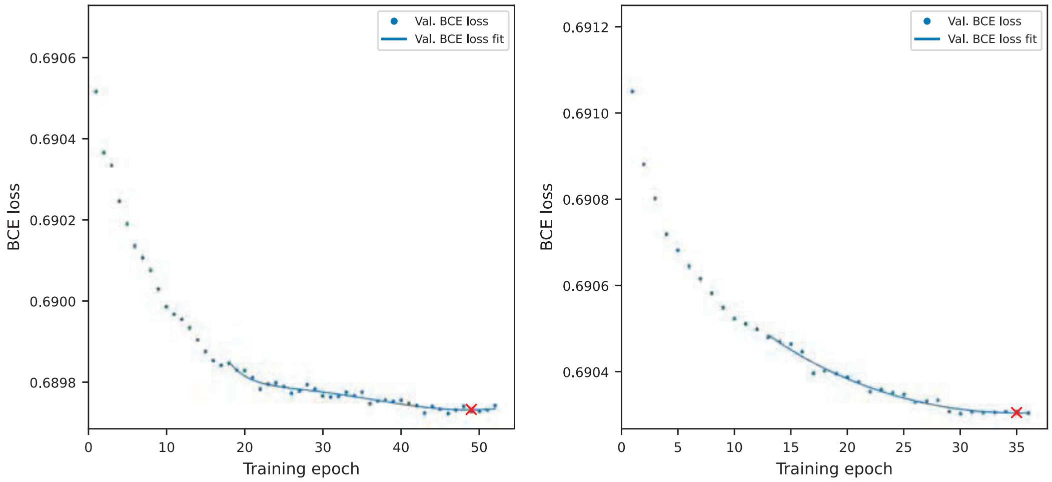


(e) Δm_{jj} in WW-cmf **(f)** Δm_{jj} in pp-cmf

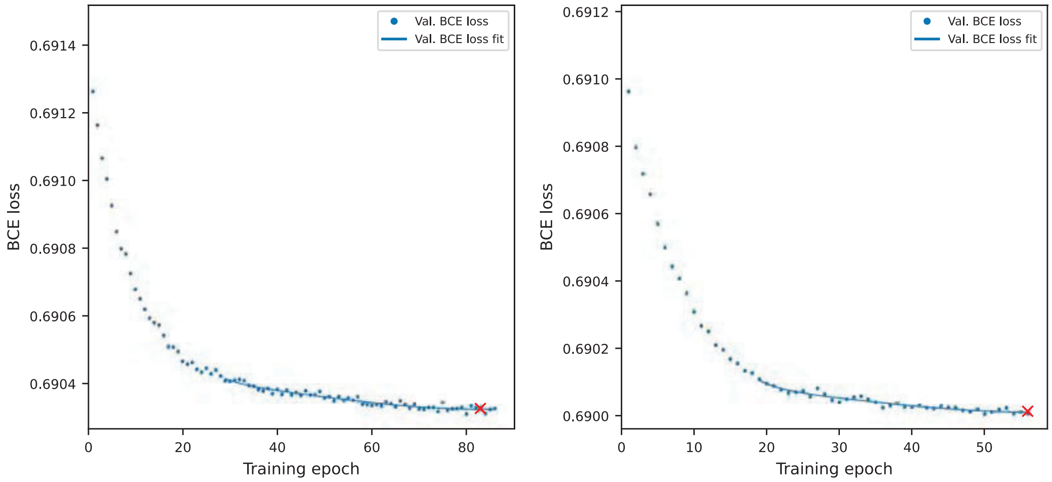
Figure B.1: Comparison of Sherpa $W^\pm W^\pm jj$ and $W^\pm W^\pm jj$ (t-ch.) on particle-level in the phase space defined in table 8.4. The comparison is done for polarization in the WW-cmf and pp-cmf.



(a) $\mathrm{DNN}_{0,1j}$ for LL in WW-cmf **(b)** $\mathrm{DNN}_{0,1j}$ for LL in pp-cmf



(c) $\mathrm{DNN}_{0,1j}$ for TL in WW-cmf **(d)** $\mathrm{DNN}_{0,1j}$ for TL in pp-cmf



(e) $\mathrm{DNN}_{0,1j}$ for TT in WW-cmf **(f)** $\mathrm{DNN}_{0,1j}$ for TT in pp-cmf

Figure B.2: The training process of $\mathrm{DNN}^{pol}_{0,1j}$ used for the approximate NLO-QCD correction. The DNNs are selected from the hyperparameter optimization by the minimum of the fit to the validation loss.

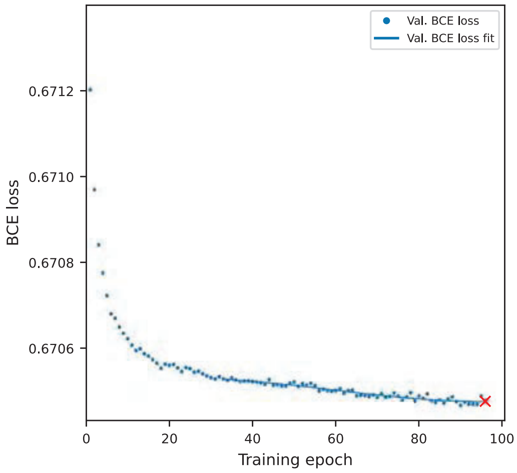
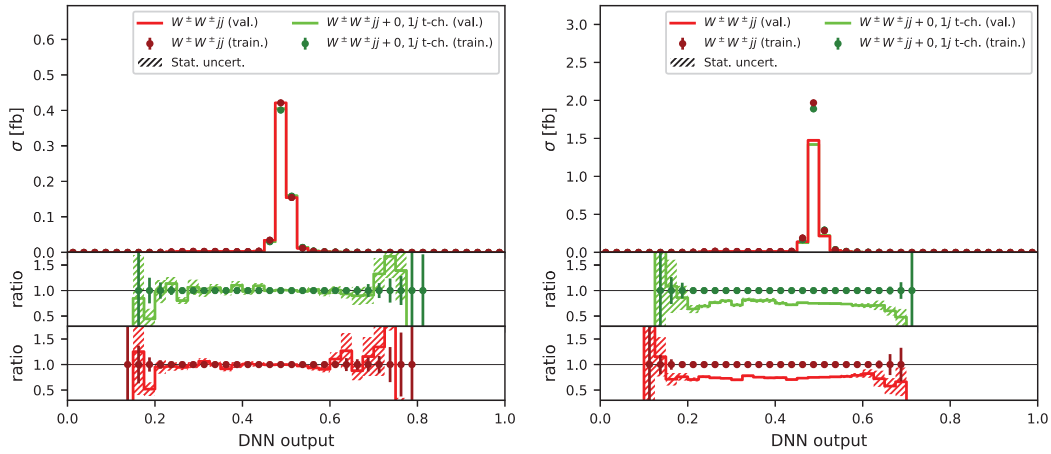
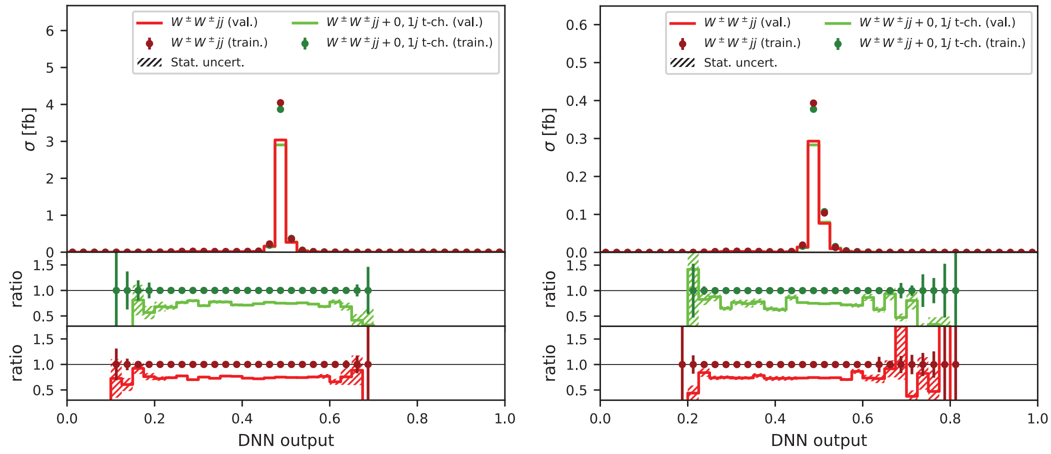


Figure B.3: The training process of $\mathrm{DNN}_{\text{s-ch.}}$ used for the approximate NLO-QCD correction. The DNNs are selected from the hyperparameter optimization by the minimum of the fit to the validation loss.



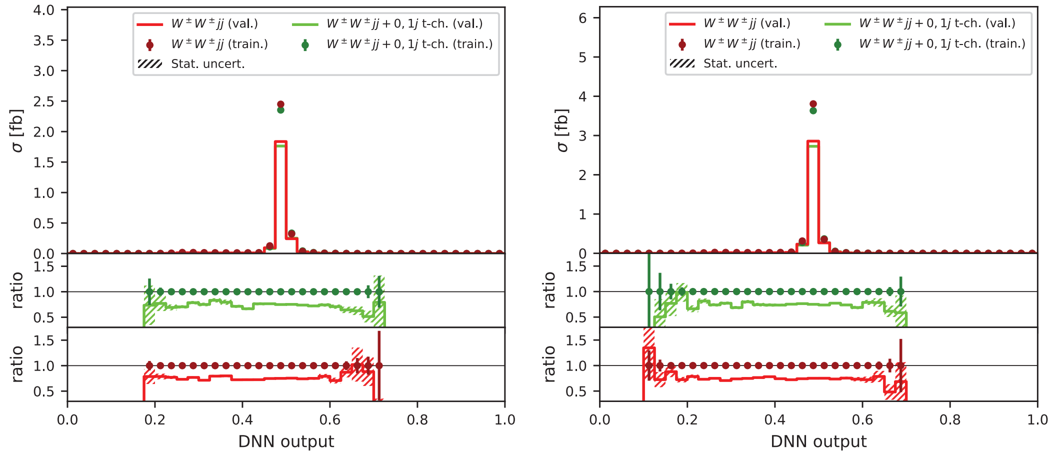
(a) $\mathrm{DNN}_{0,1j}$ for LL in WW-cmf

(b) $\mathrm{DNN}_{0,1j}$ for TL in WW-cmf



(c) $\mathrm{DNN}_{0,1j}$ for TT in WW-cmf

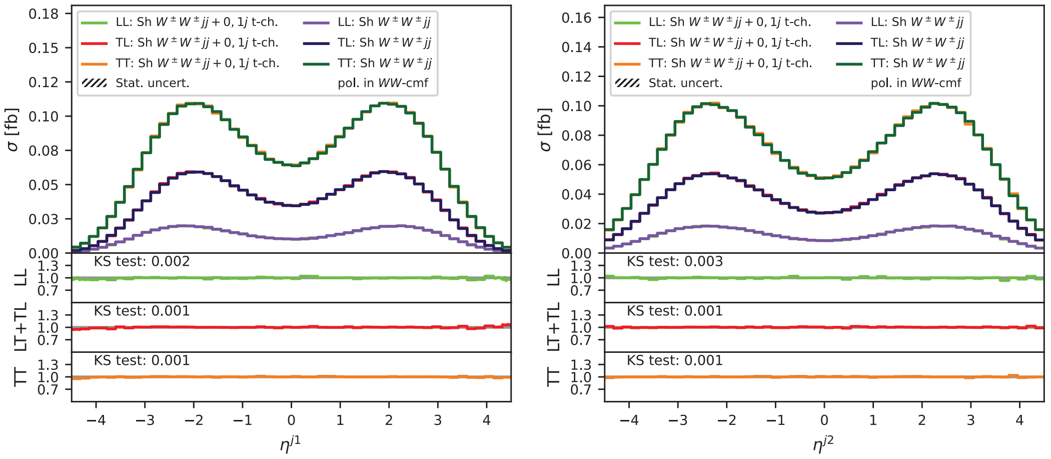
(d) $\mathrm{DNN}_{0,1j}$ for LL in pp-cmf



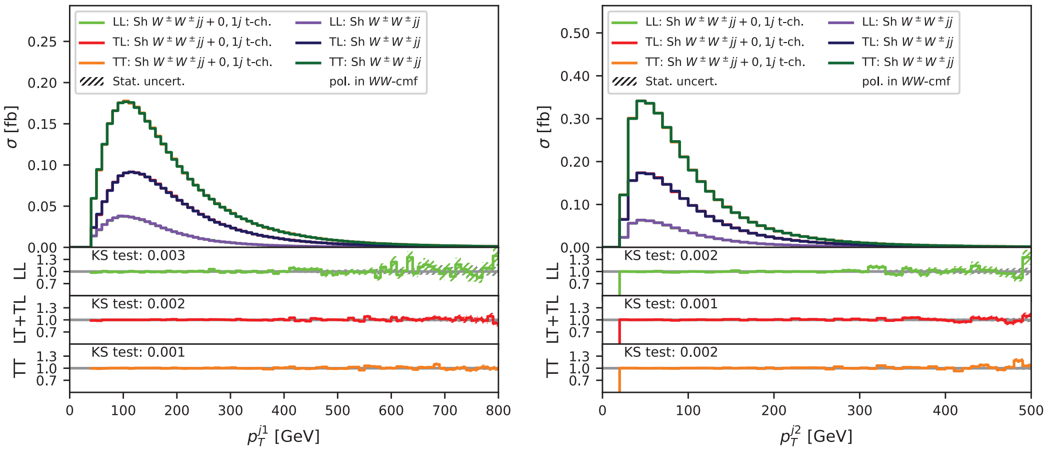
(e) $\mathrm{DNN}_{0,1j}$ for TL in pp-cmf

(f) $\mathrm{DNN}_{0,1j}$ for TT in pp-cmf

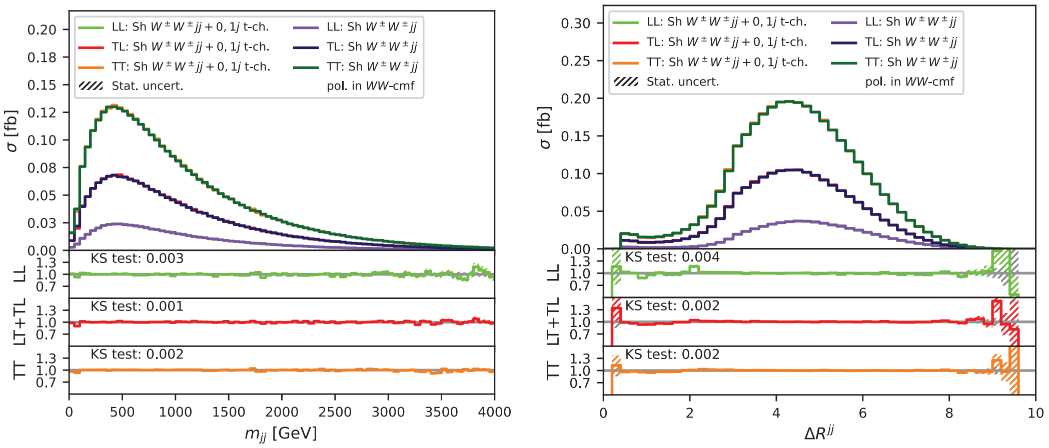
Figure B.4: The classification output of the $\mathrm{DNN}^{poll}_{0,1j}$ used for the multivariate higher-order QCD correction.



(a) lead. jet η **(b)** sublead. jet η

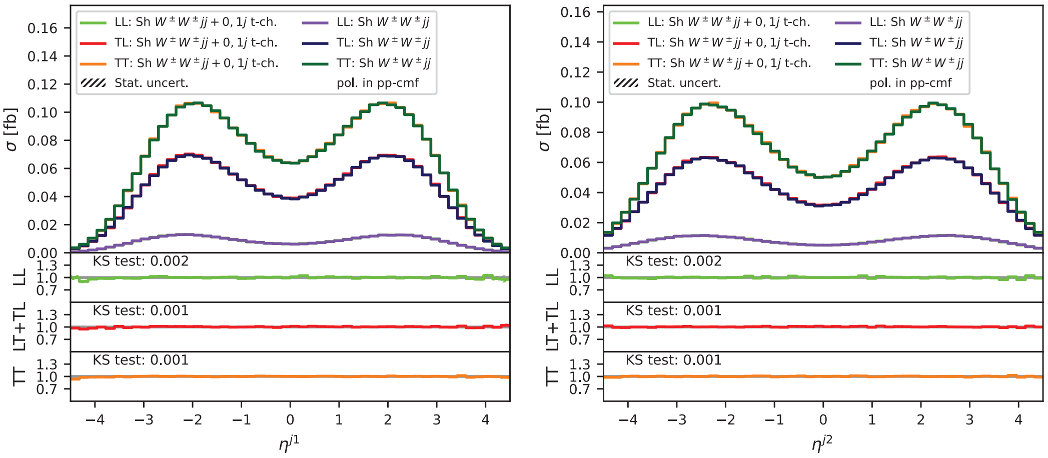


(c) lead. jet p_T **(d)** sublead. jet p_T

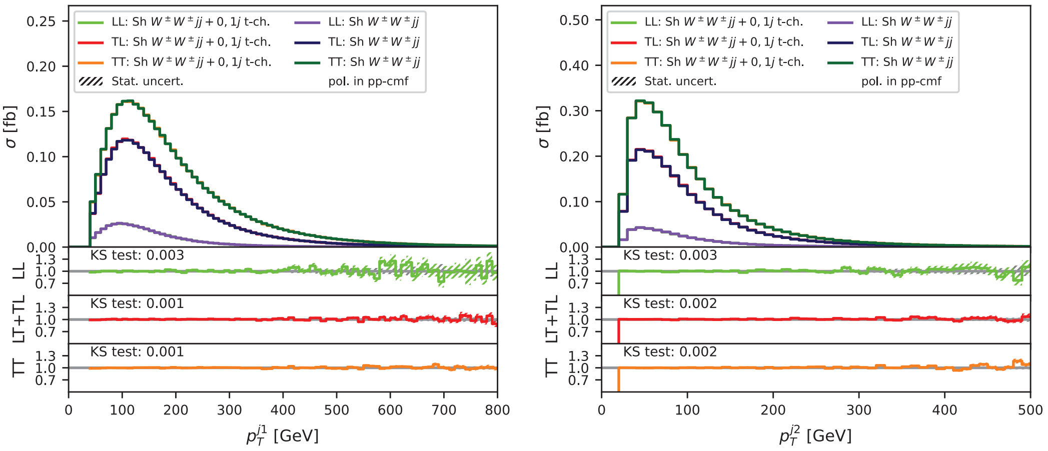


(e) m_{jj} **(f)** ΔR_{jj}

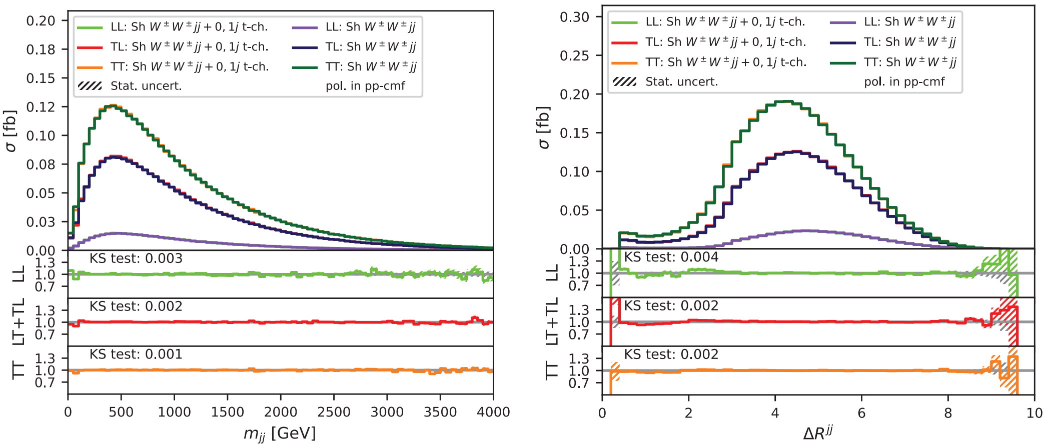
Figure B.5: Comparison of the polarization in WW-cmf simulated by Sherpa $W^\pm W^\pm jj$ and $W^\pm W^\pm jj$ (t-ch.) on particle-level in the phase space defined in table 8.4. The Sherpa $W^\pm W^\pm jj$ is corrected by the DNN$^{poll}_{0,1j}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training.



(a) lead. jet η **(b)** sublead. jet η

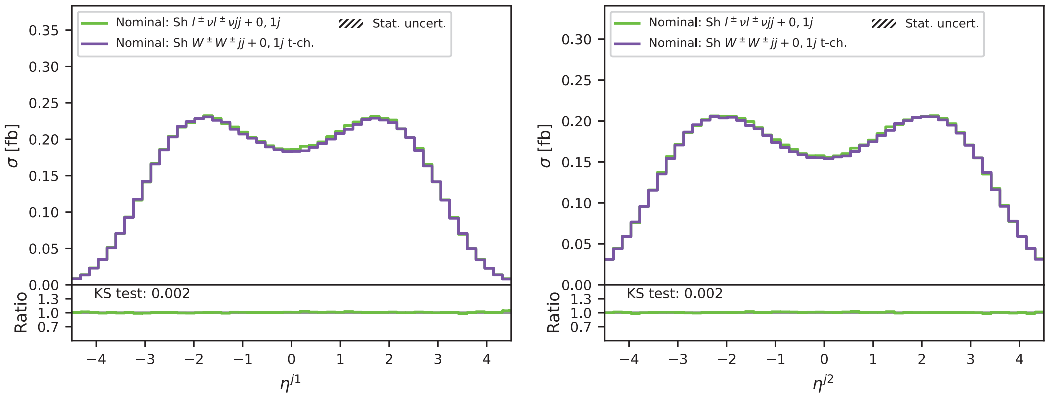


(c) lead. jet p_T **(d)** sublead. jet p_T



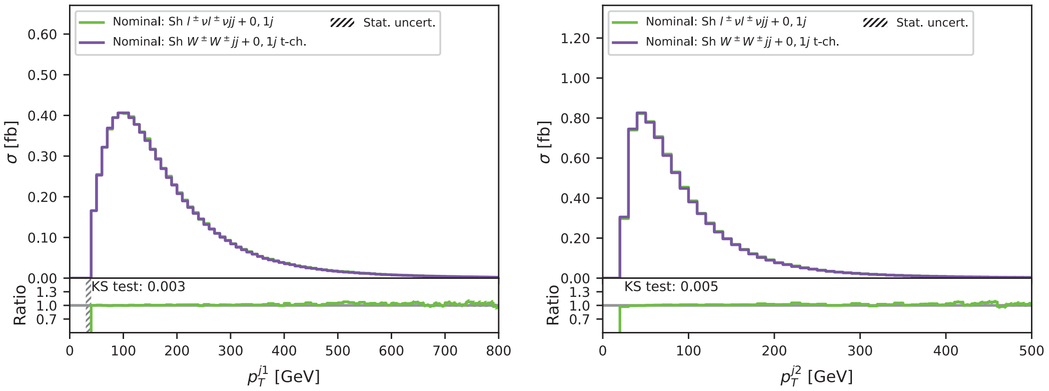
(e) m_{jj} **(f)** ΔR_{jj}

Figure B.6: Comparison of the polarization in pp-cmf simulated by SHERPA $W^\pm W^\pm jj$ and $W^\pm W^\pm jj$ (t-ch.) on particle-level in the phase space defined in table 8.4. The SHERPA $W^\pm W^\pm jj$ is corrected by the $\mathrm{DNN}^{poll}_{0,1j}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training.



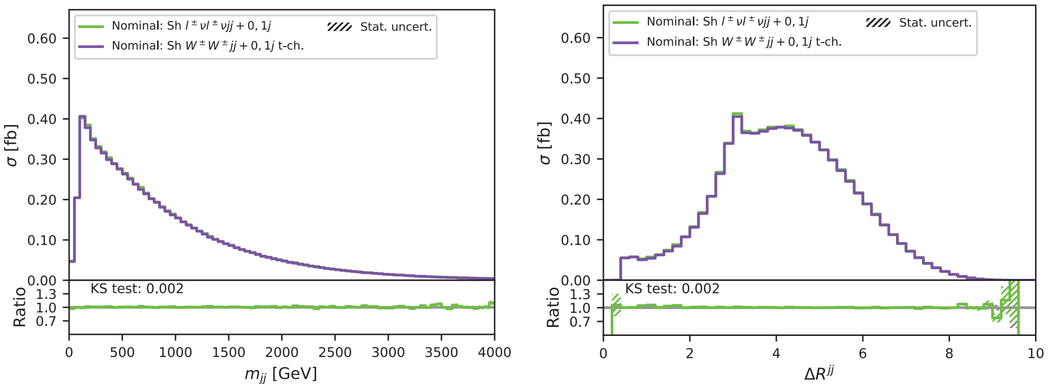
(a) lead. jet η

(b) sublead. jet η



(c) lead. jet p_T

(d) sublead. jet p_T



(e) m_{jj}

(f) ΔR_{jj}

Figure B.7: Comparison of Sherpa $W^\pm W^\pm jj + 0, 1j$ (t-ch.) and $l^\pm l^\pm \nu\nu jj + 0, 1j$ on particle-level in the phase space defined in table 8.4. The Sherpa $W^\pm W^\pm jj + 0, 1j$ (t-ch.) is corrected by the $\text{DNN}_{\text{s-ch.}}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training.

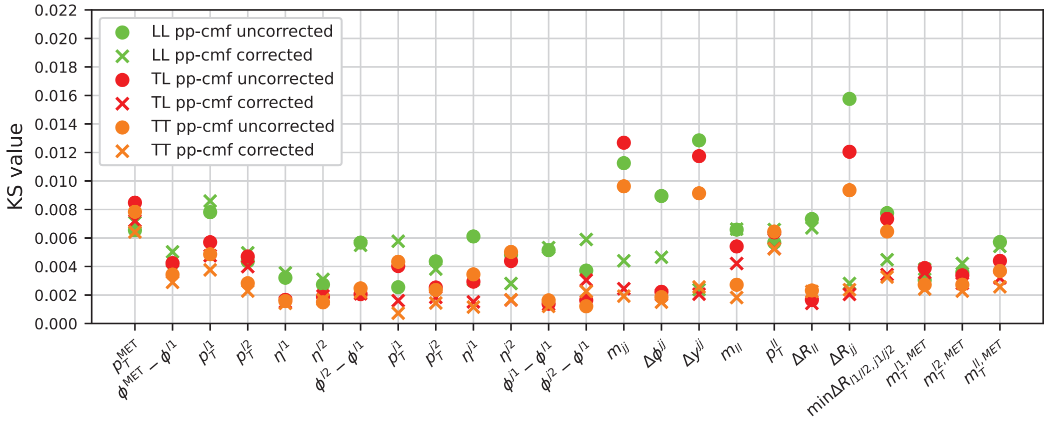
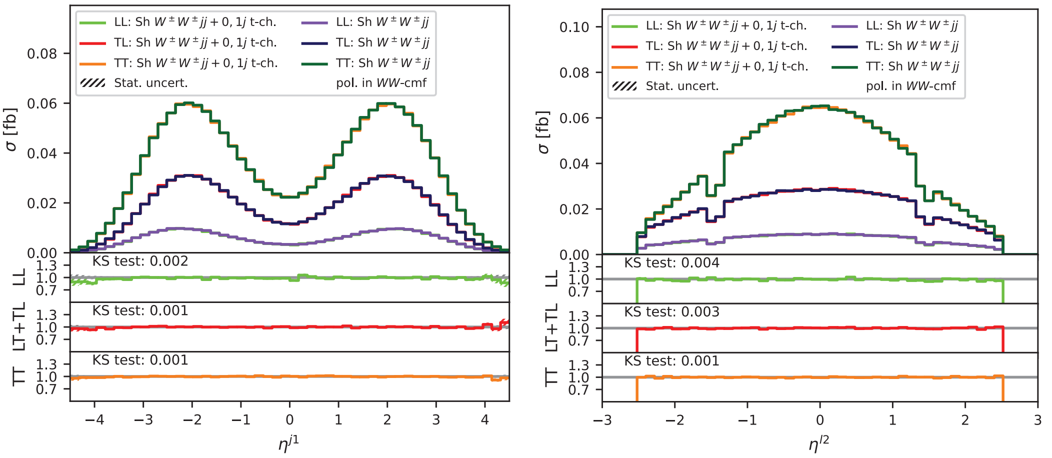
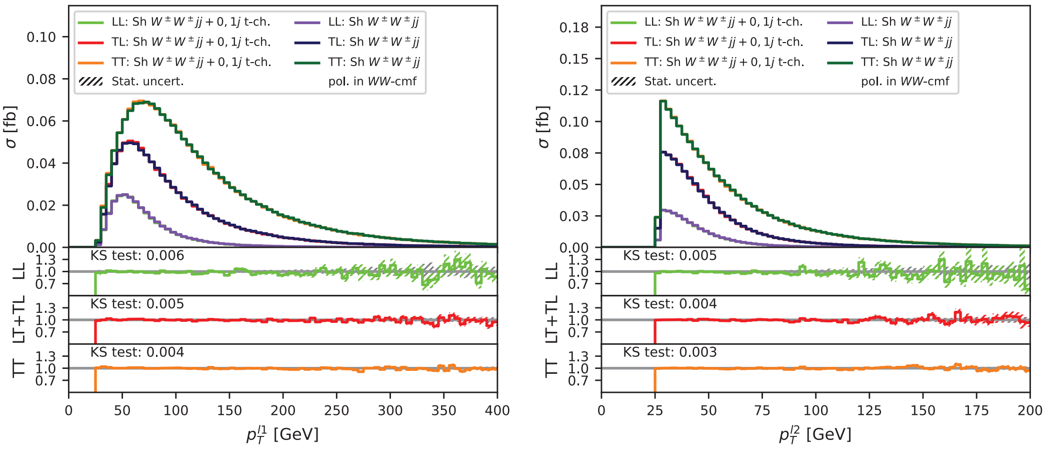


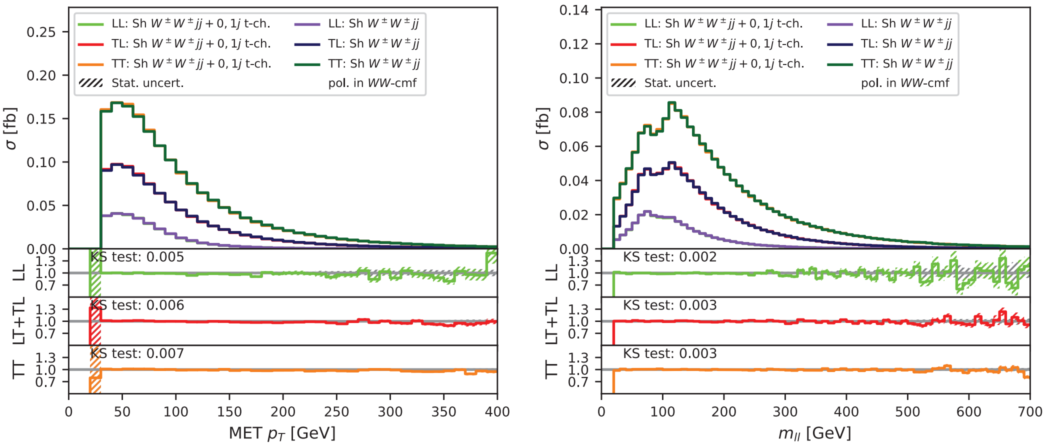
Figure B.8: Comparison of the polarization in pp-cmf simulated by Sherpa $W^\pm W^\pm jj$ and $W^\pm W^\pm jj$ (t-ch.) on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The Sherpa $W^\pm W^\pm jj$ is corrected by the DNN$^{poll}_{0,1j}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training. The Kolmogorov-Smirnov values give an overview of the shape improvements in the tested variables.



(a) lead. lep. η **(b)** sublead. lep. η

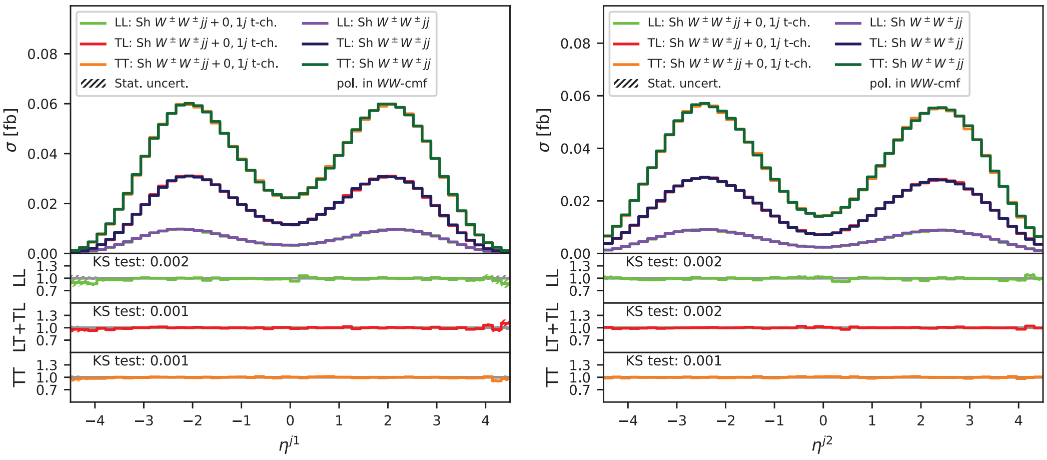


(c) lead. lep. p_T **(d)** sublead. lep. p_T

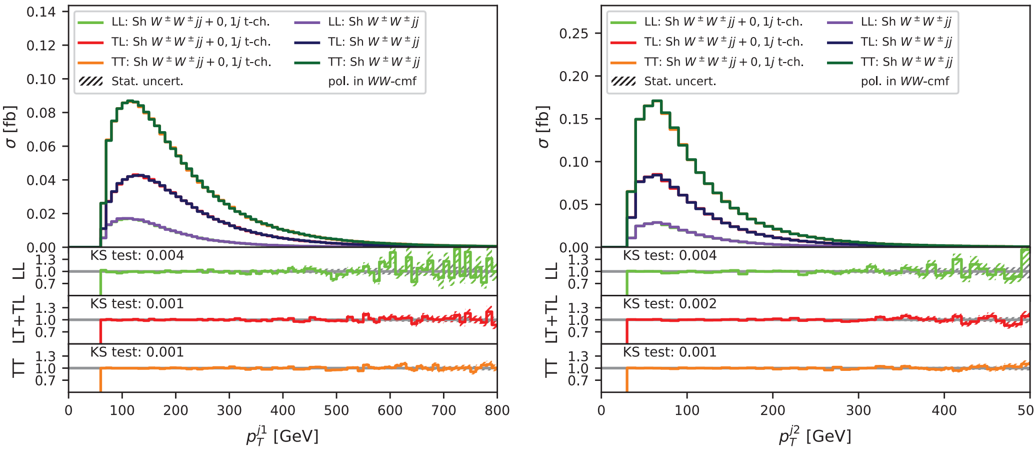


(e) p_T^{miss} **(f)** m_{ll}

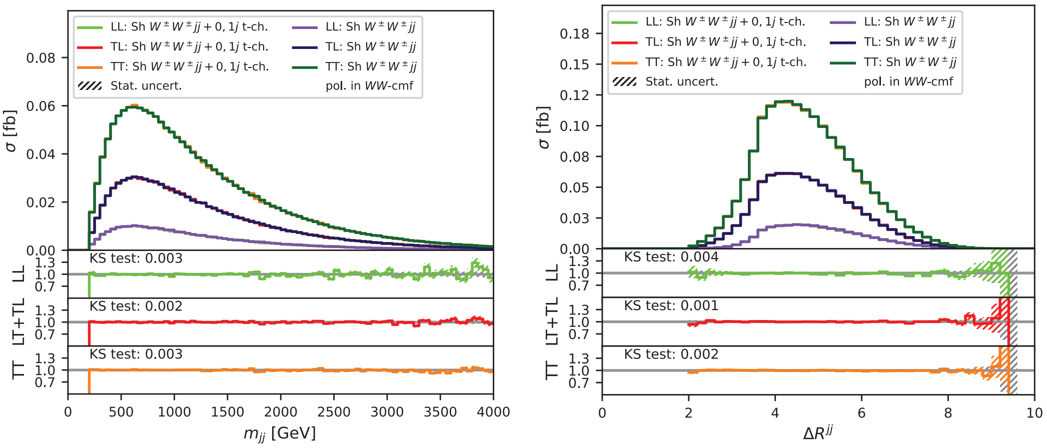
Figure B.9: Comparison of the polarization in WW-cmf simulated by SHERPA $W^\pm W^\pm jj$ and $W^\pm W^\pm jj$ (t-ch.) in lepton and p_T^{miss} kinematic on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The SHERPA $W^\pm W^\pm jj$ is corrected by the $\text{DNN}^{poll}_{0,1j}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training.



(a) lead. jet η **(b)** sublead. jet η

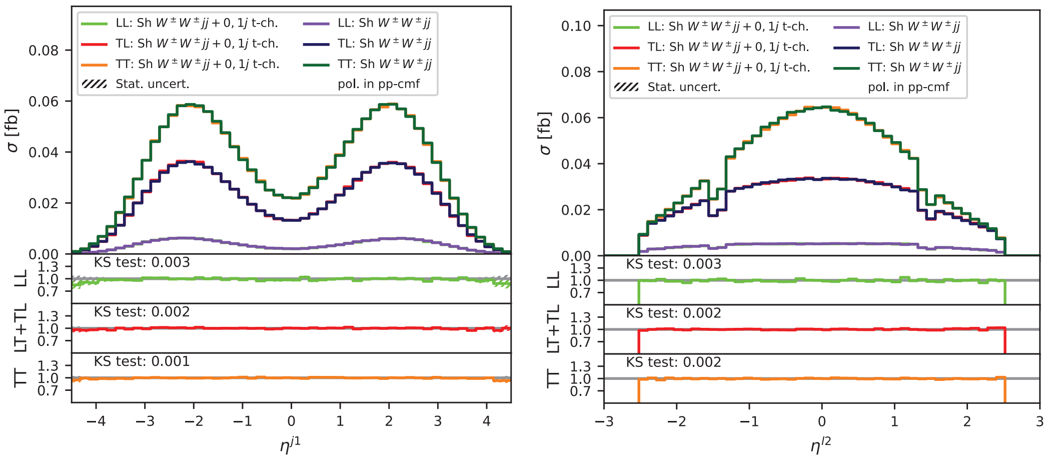


(c) lead. jet p_T **(d)** sublead. jet p_T



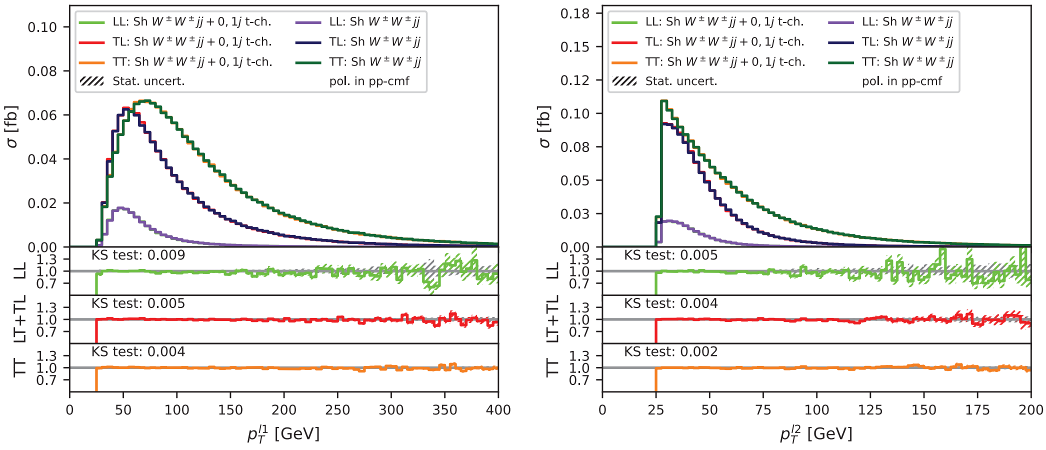
(e) m_{jj} **(f)** ΔR_{jj}

Figure B.10: Comparison of the polarization in WW-cmf simulated by Sherpa $W^\pm W^\pm jj$ and $W^\pm W^\pm jj$ (t-ch.) in jet kinematic on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The Sherpa $W^\pm W^\pm jj$ is corrected by the $\text{DNN}_{0,1j}^{poll}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training.



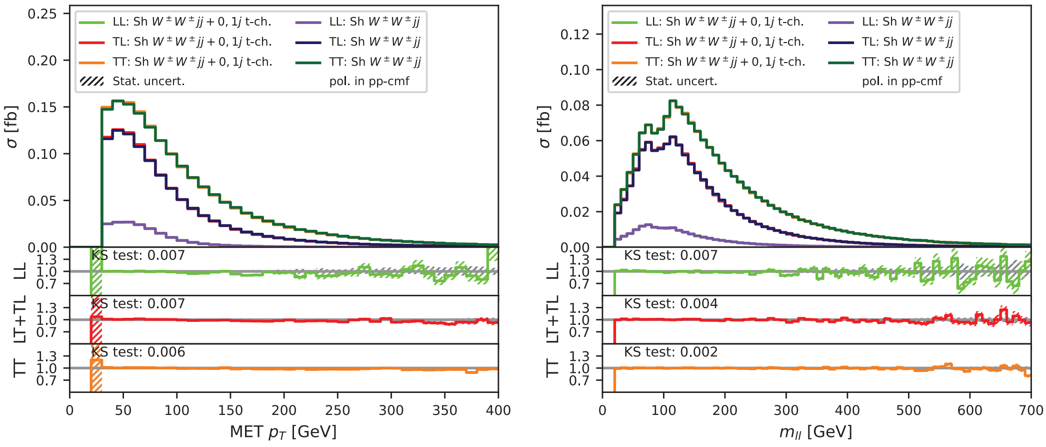
(a) lead. lep. η

(b) sublead. lep. η



(c) lead. lep. p_T

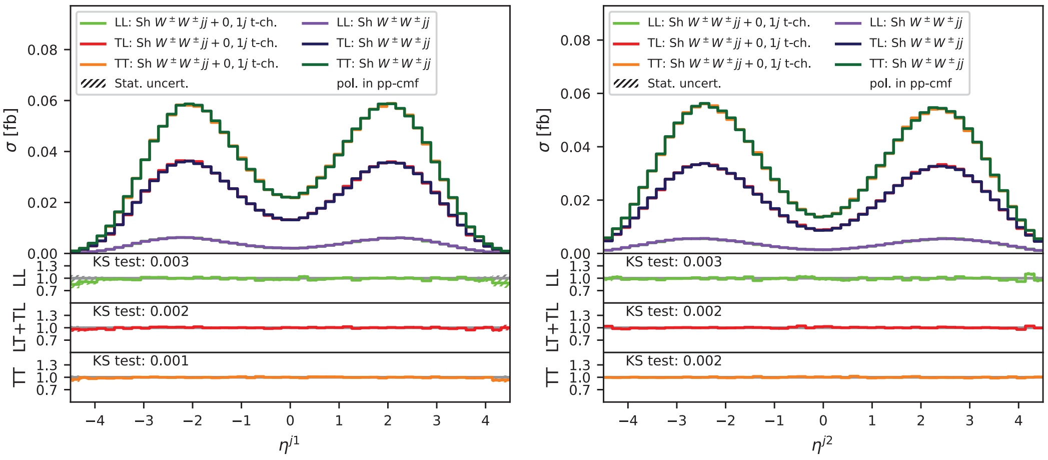
(d) sublead. lep. p_T



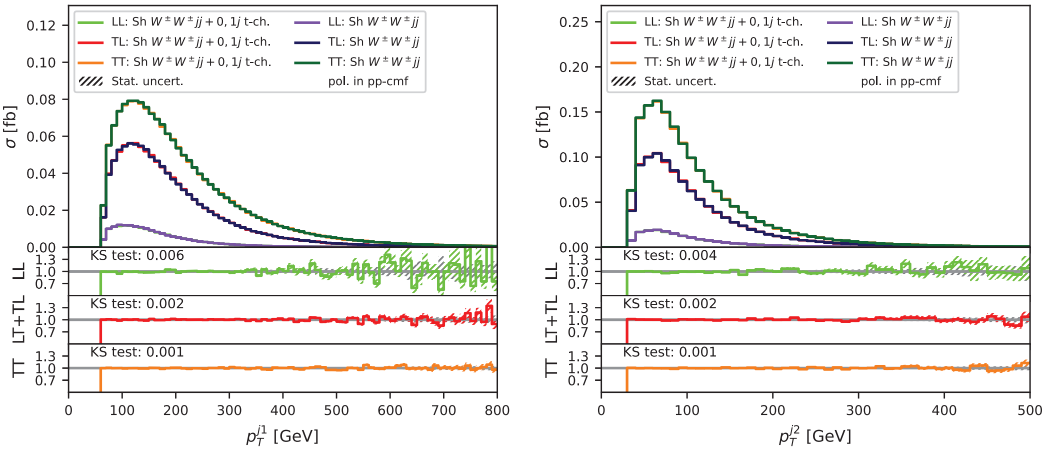
(e) p_T^{miss}

(f) m_{ll}

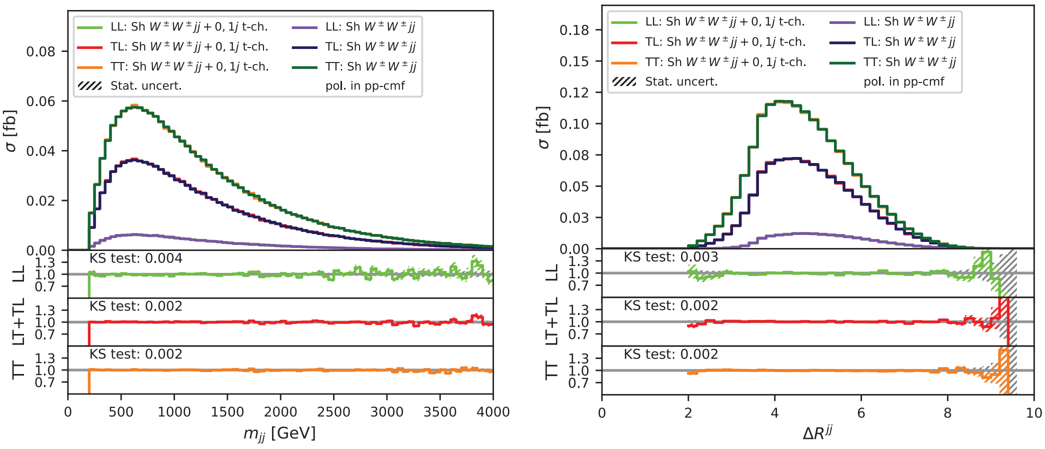
Figure B.11: Comparison of the polarization in pp-cmf simulated by Sherpa $W^\pm W^\pm jj$ and $W^\pm W^\pm jj$ (t-ch.) in lepton and p_T^{miss} kinematic on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The Sherpa $W^\pm W^\pm jj$ is corrected by the $\mathrm{DNN}_{0,1j}^{poll}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training.



(a) lead. jet η **(b)** sublead. jet η

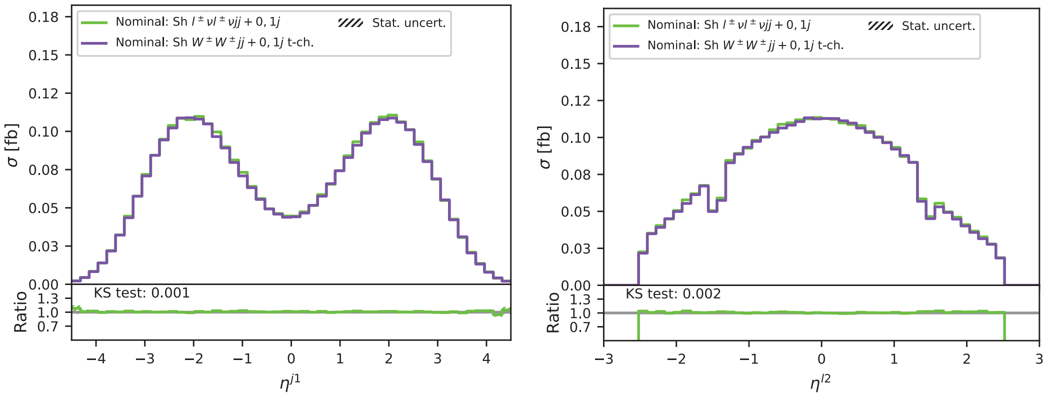


(c) lead. jet p_T **(d)** sublead. jet p_T

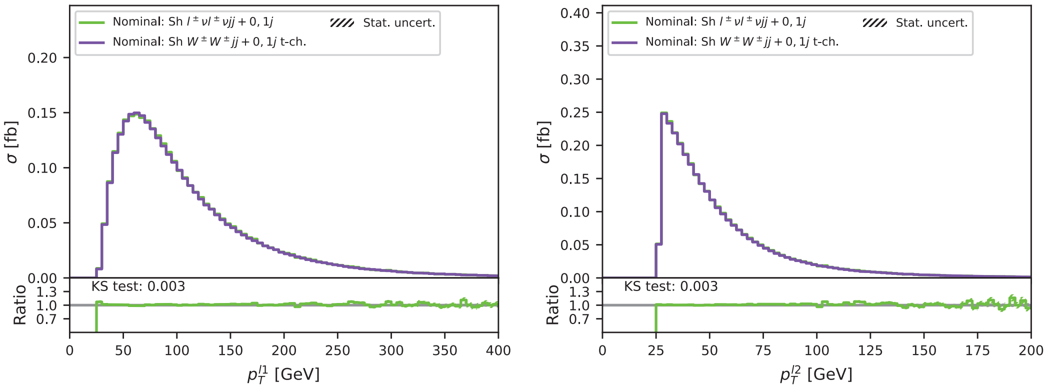


(e) m_{jj} **(f)** ΔR_{jj}

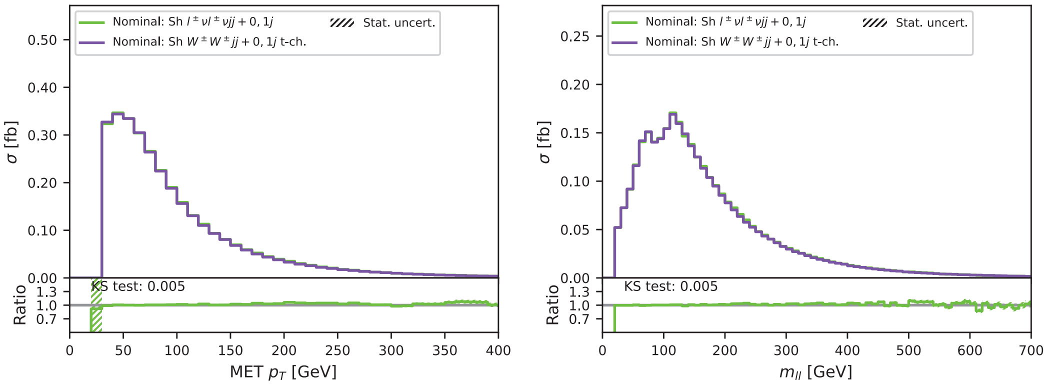
Figure B.12: Comparison of the polarization in pp-cmf simulated by Sherpa $W^\pm W^\pm jj$ and $W^\pm W^\pm jj$ (t-ch.) in jet kinematic on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The Sherpa $W^\pm W^\pm jj$ is corrected by the $\text{DNN}_{0,1j}^{poll}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training.



(a) lead. lep. η **(b)** sublead. lep. η

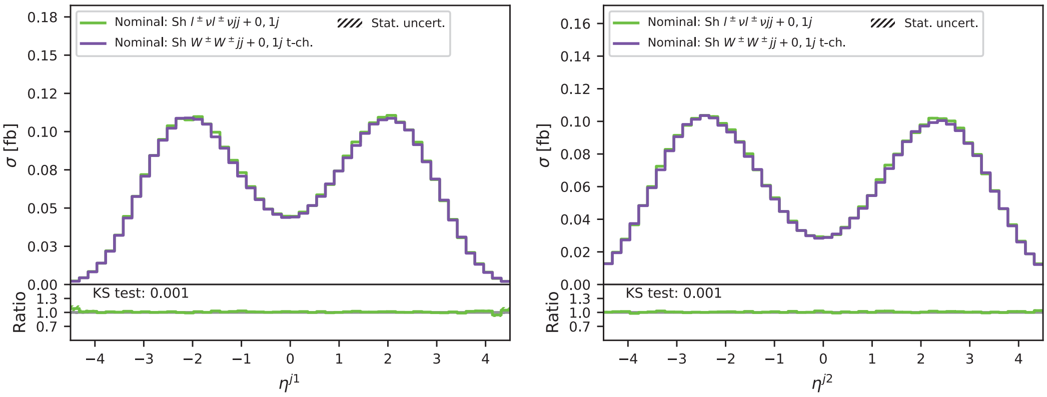


(c) lead. lep. p_T **(d)** sublead. lep. p_T

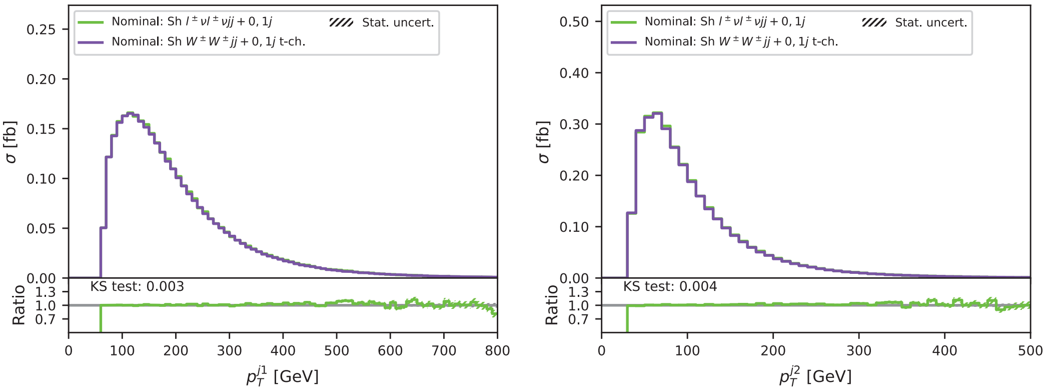


(e) p_T^{miss} **(f)** m_{ll}

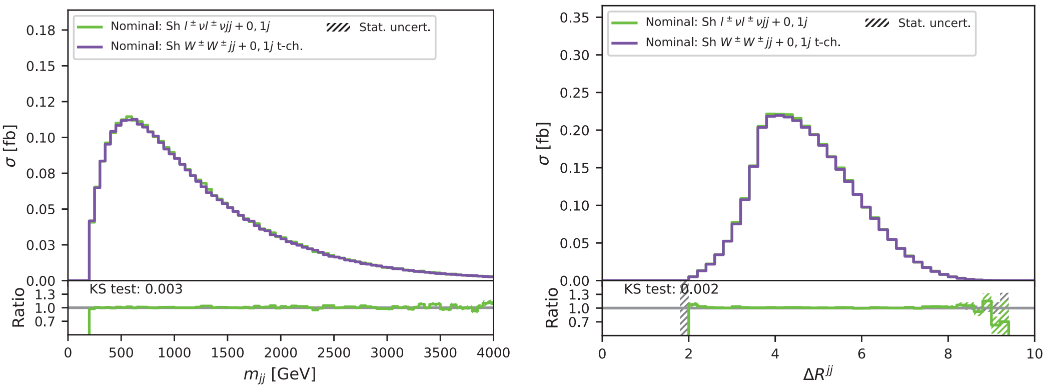
Figure B.13: Comparison of SHERPA $W^{\pm}W^{\pm}jj + 0, 1j$ (t-ch.) and $l^{\pm}l^{\pm}\nu\nu jj + 0, 1j$ in lepton and p_T^{miss} kinematic on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The SHERPA $W^{\pm}W^{\pm}jj+0, 1j$ (t-ch.) is corrected by the DNN$_{\text{s-ch.}}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training.



(a) lead. jet η **(b)** sublead. jet η

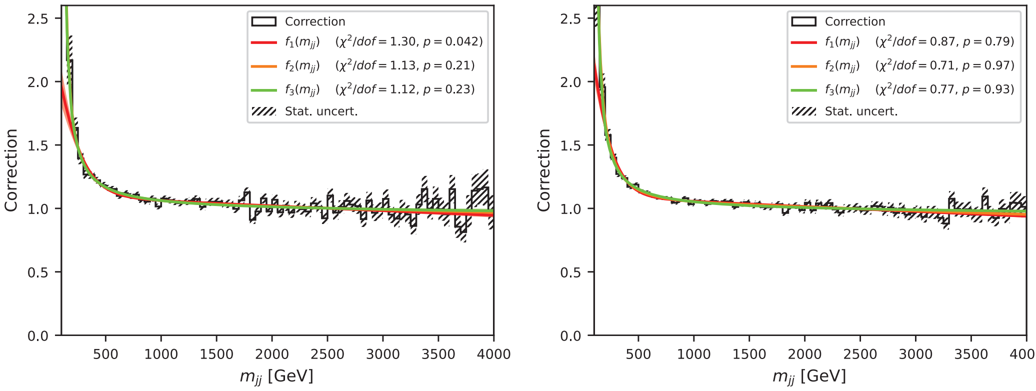


(c) lead. jet p_T **(d)** sublead. jet p_T



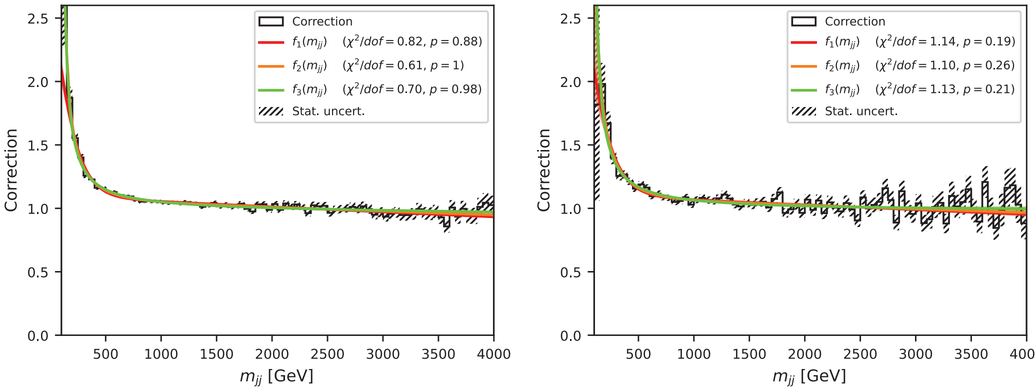
(e) m_{jj} **(f)** ΔR_{jj}

Figure B.14: Comparison of Sherpa $W^\pm W^\pm jj + 0, 1j$ (t-ch.) and $l^\pm l^\pm \nu\nu jj + 0, 1j$ jet kinematic on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The Sherpa $W^\pm W^\pm jj + 0, 1j$ (t-ch.) is corrected by the DNN$_{\text{s-ch.}}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training.



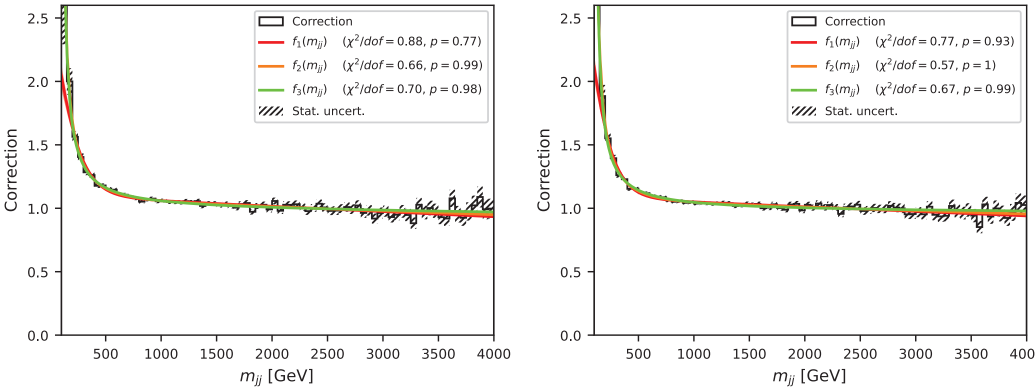
(a) jet-merged corr for LL in WW-cmf

(b) jet-merged corr for TL in WW-cmf



(c) jet-merged corr for TT in WW-cmf

(d) jet-merged corr for LL in pp-cmf



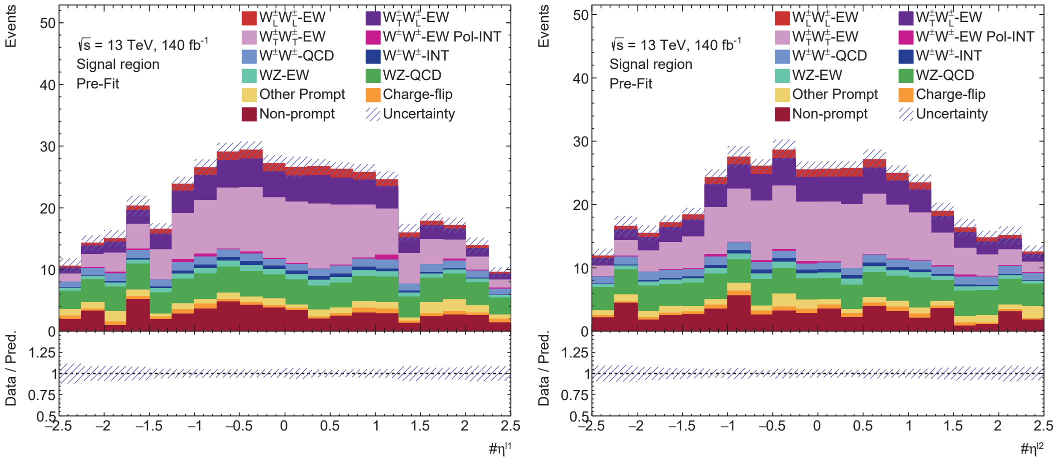
(e) jet-merged corr for TL in pp-cmf

(f) jet-merged corr for TT in pp-cmf

Figure B.15: Comparison of different fit functions for the higher-order QCD correction factor $\mathrm{corr}_{\text{jet-merged}}$ in m_{jj}. The correction is extracted from the combined signal and Low-m_{jj} region at particle-level.

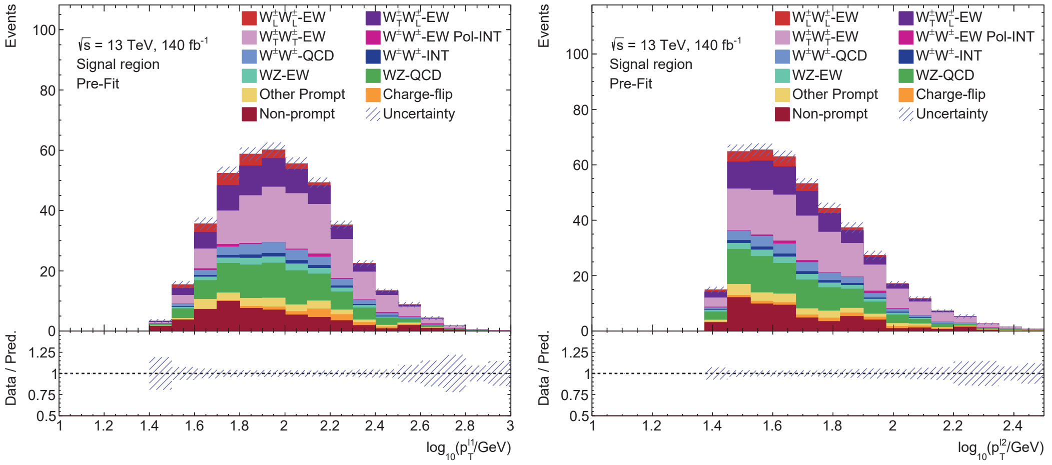
Appendix C

DNN Optimization



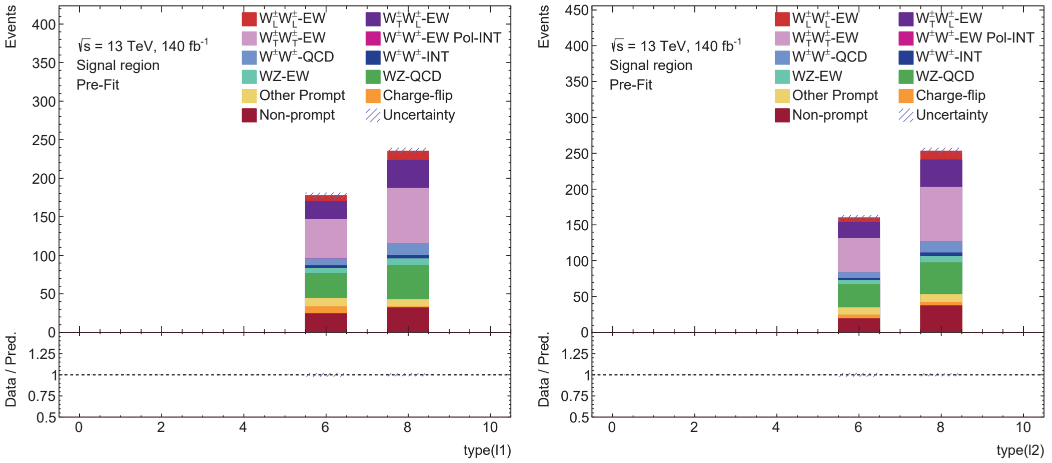
(a) lead. lepton η

(b) sublead. lepton η



(c) lead. lepton $log_{10}(p_T/\text{GeV})$

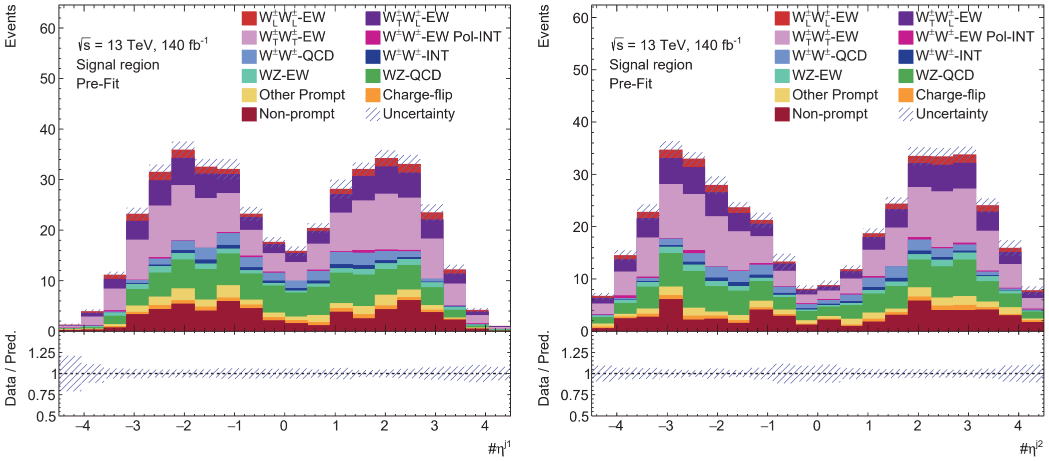
(d) sublead. lepton $log_{10}(p_T/\text{GeV})$



(e) lead. lepton type

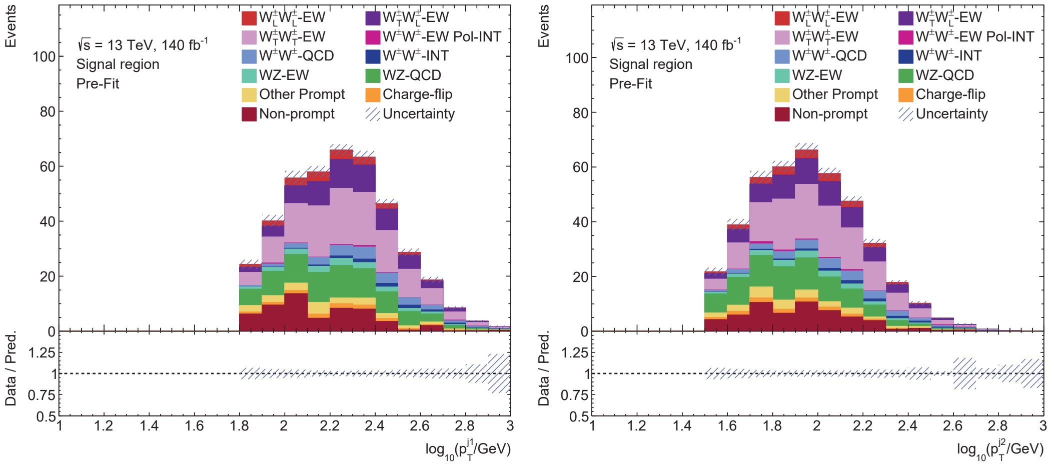
(f) sublead. lepton type

Figure C.1: Distributions of the training input in the signal region.



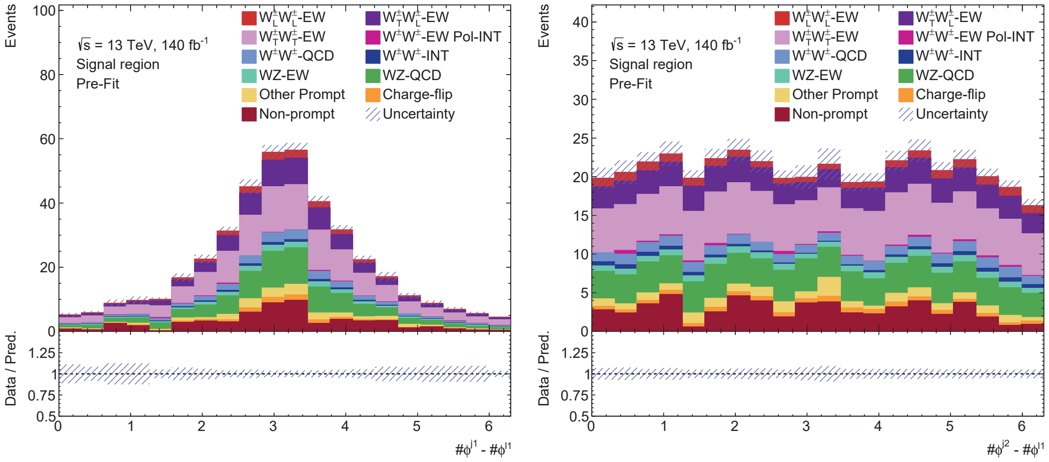
(a) lead. jet η

(b) sublead. jet η



(c) lead. jet $log_{10}(p_T/\mathrm{GeV})$

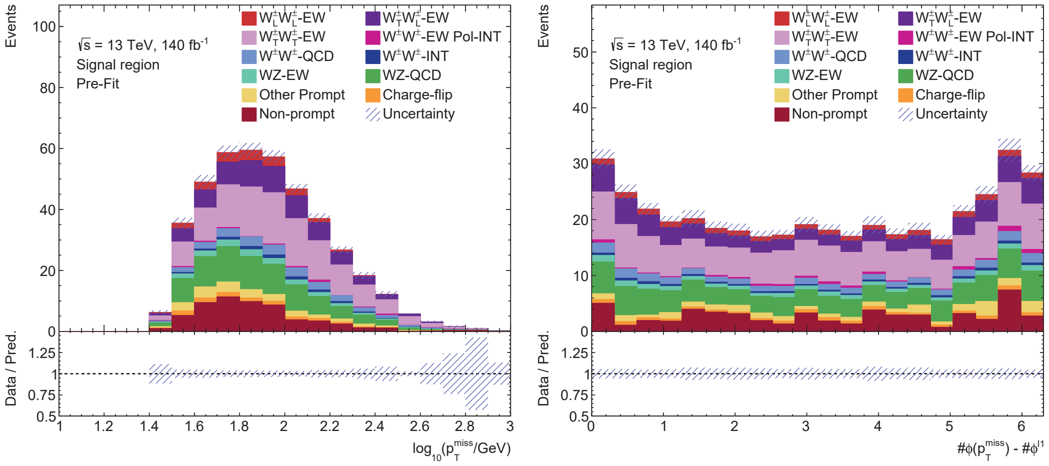
(d) sublead. jet $log_{10}(p_T/\mathrm{GeV})$



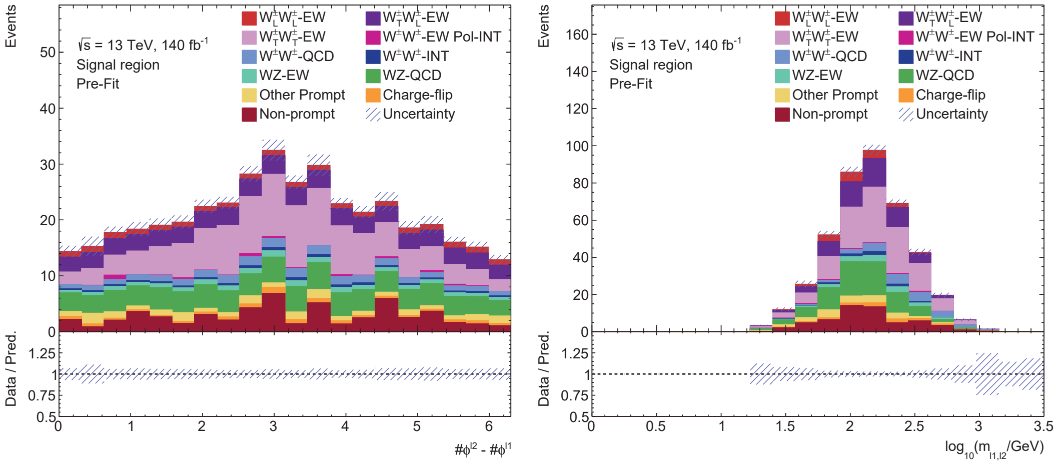
(e) lead. jet $\phi^{j1} - \phi^{l1}$

(f) sublead. jet $\phi^{j2} - \phi^{l1}$

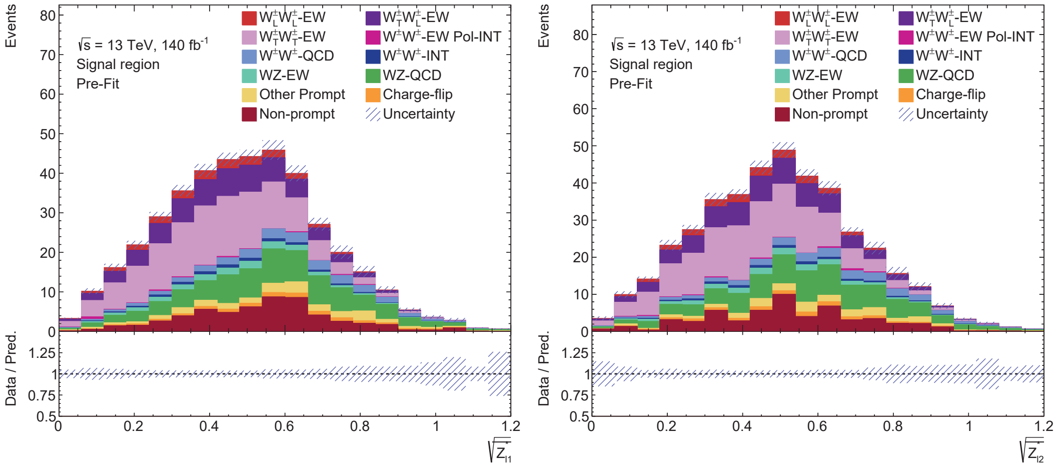
Figure C.2: Distributions of the training input in the signal region.



(a) Missing transverse momentum $log_{10}(p_T^{miss}/\mathrm{GeV})$ **(b)** $\phi(p_T^{miss}) - \phi^{l1}$

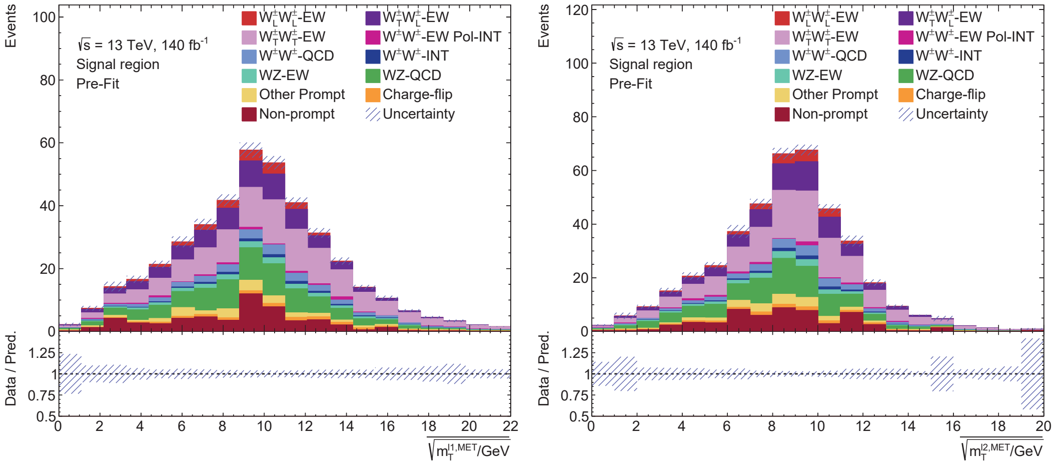


(c) sublead. lepton $\phi^{l2} - \phi^{l1}$ **(d)** lepton $log_{10}(m_{ll}/\mathrm{GeV})$

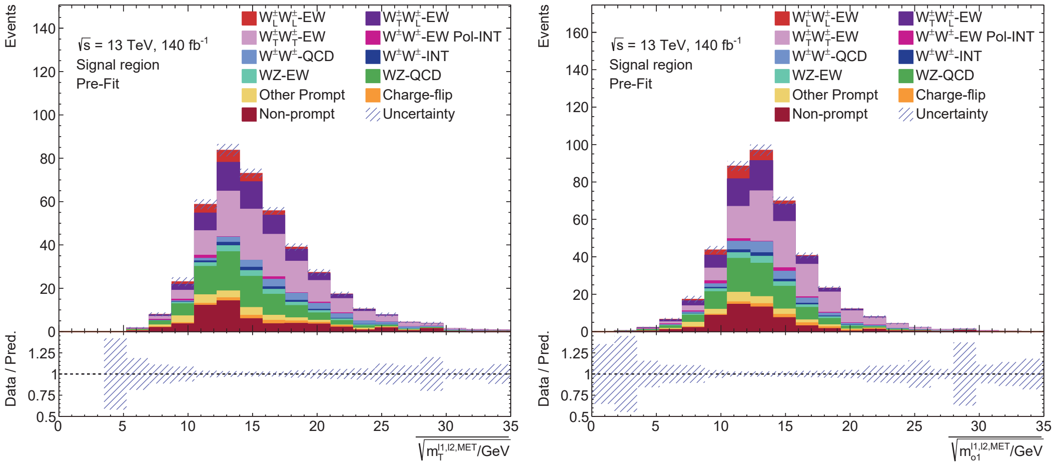


(e) lead. lepton $\sqrt{Z^{\star}/\mathrm{GeV}}$ **(f)** sublead. lepton $\sqrt{Z^{\star}/\mathrm{GeV}}$

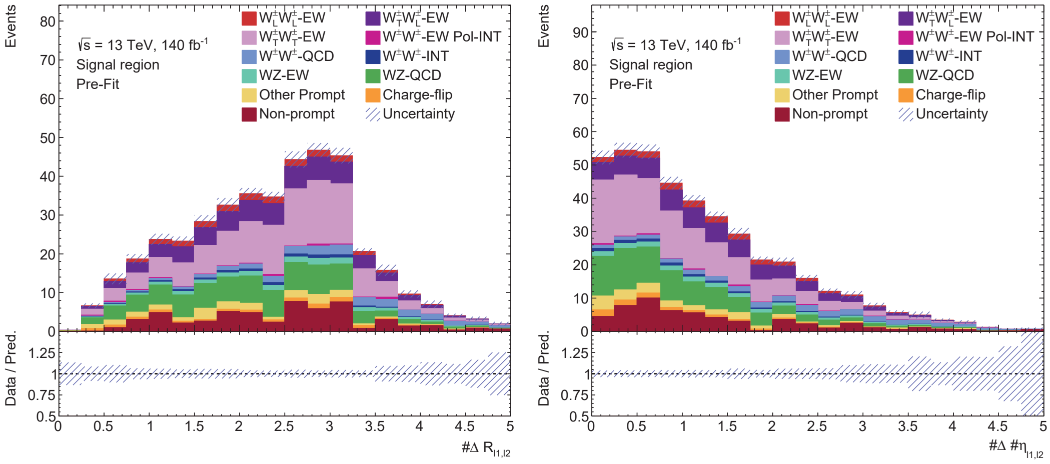
Figure C.3: Distributions of the training input in the signal region.



(a) lead. lepton $\sqrt{m_T^{l1,MET}}$/GeV **(b)** sublead. lepton $\sqrt{m_T^{l2,MET}}$/GeV

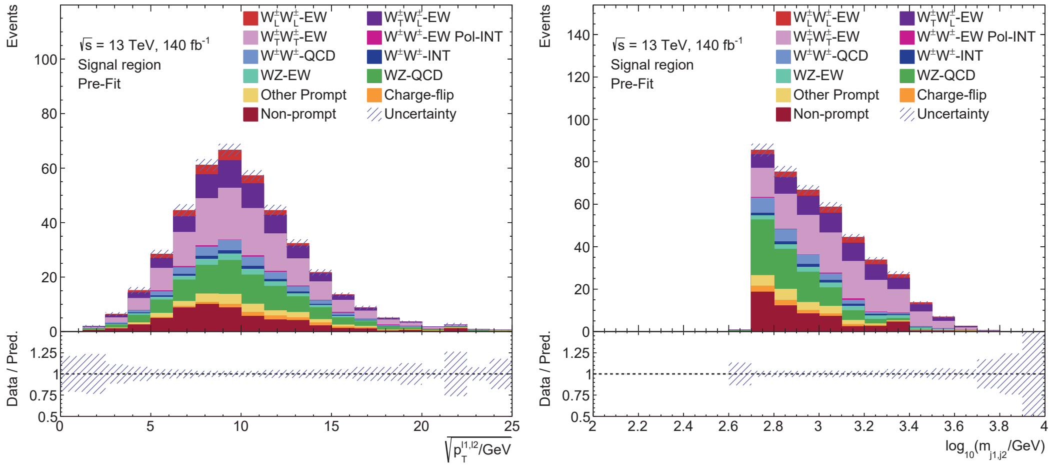


(c) lepton $\sqrt{m_T^{ll,MET}}$/GeV **(d)** lepton $\sqrt{m_{o1}^{ll,MET}}$/GeV



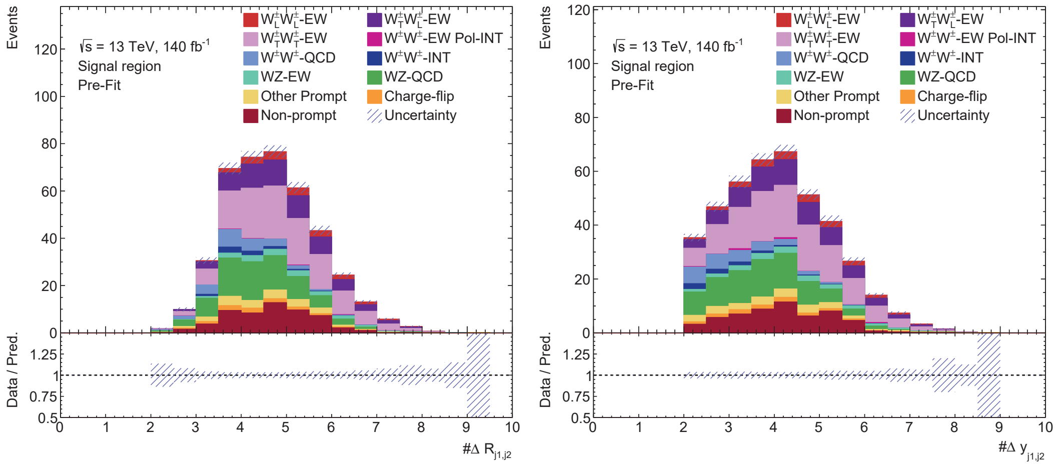
(e) lepton ΔR_{ll} **(f)** lepton $\sqrt{\Delta\eta_{ll}}$

Figure C.4: Distributions of the training input in the signal region.



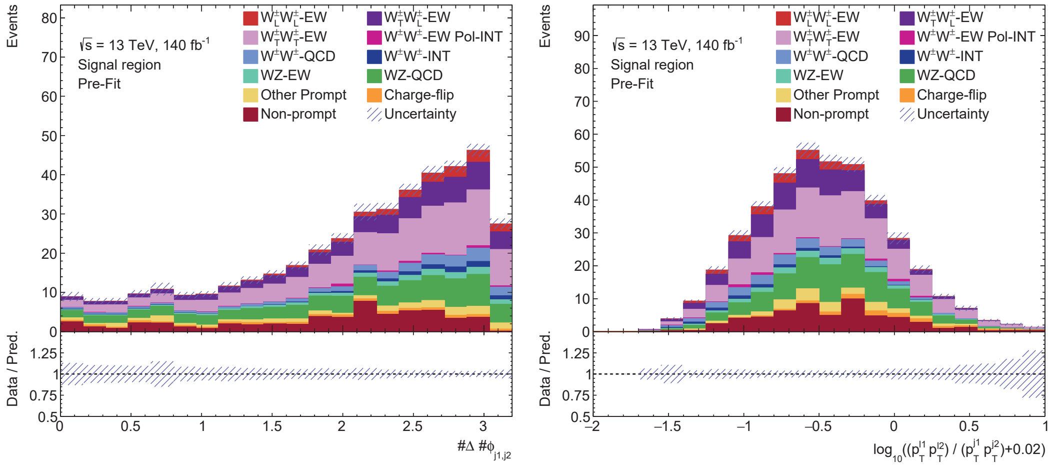
(a) lepton $\sqrt{p_T^{ll}/\text{GeV}}$

(b) jet $log_{10}(m_{jj}/\text{GeV})$



(c) jet ΔR_{jj}

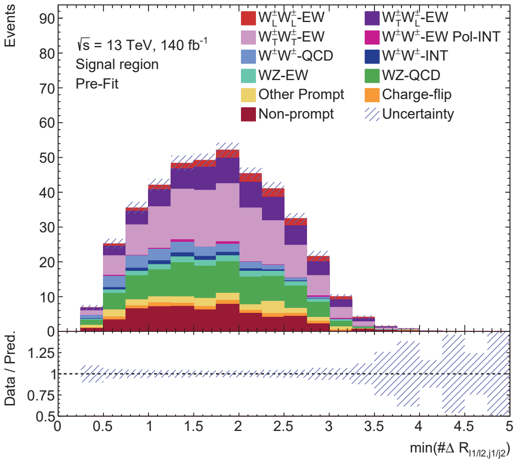
(d) jet Δy_{jj}



(e) jet $\Delta\phi_{jj}$

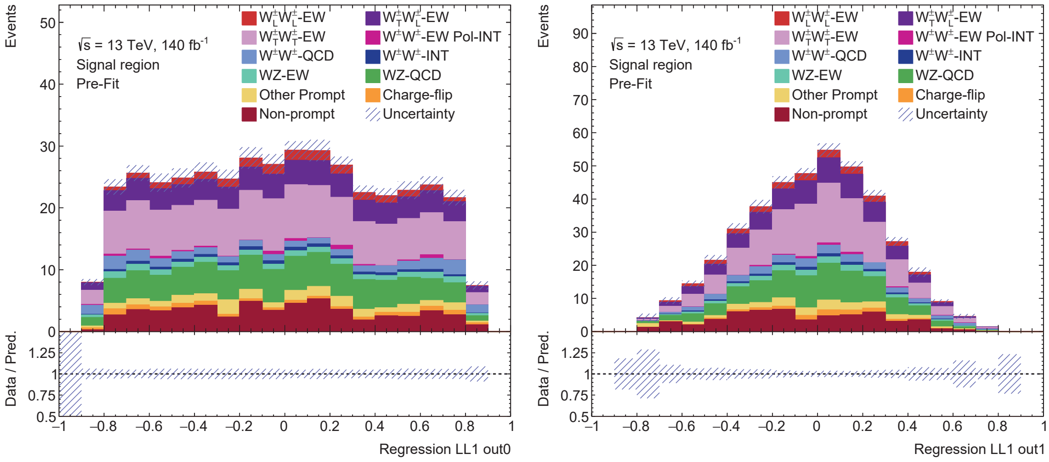
(f) lepton jet $log_{10}(\frac{p_T^{l1} \cdot p_T^{l2}}{p_T^{j1} \cdot p_T^{j2}}/\text{GeV} + 0.02)$

Figure C.5: Distributions of the training input in the signal region.

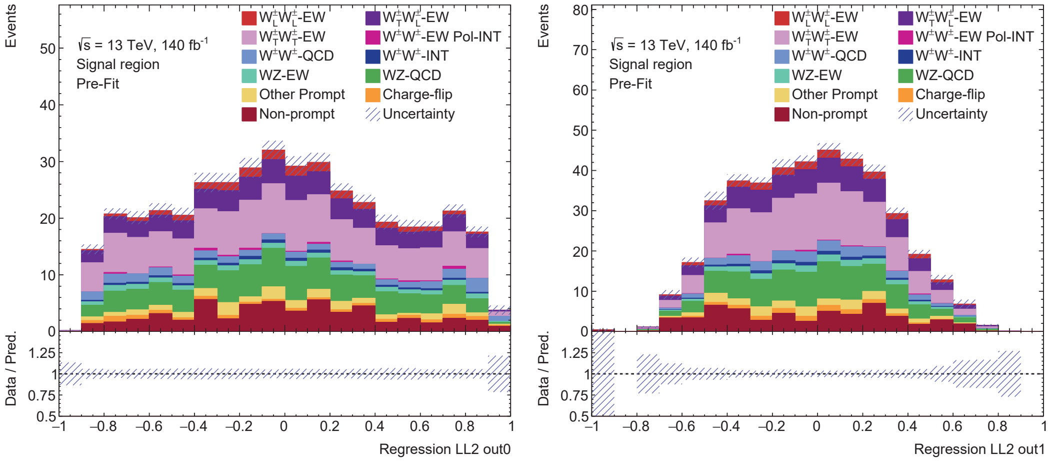


(a) lepton jet $min(\Delta R_{l1/l2,j1/j2})$

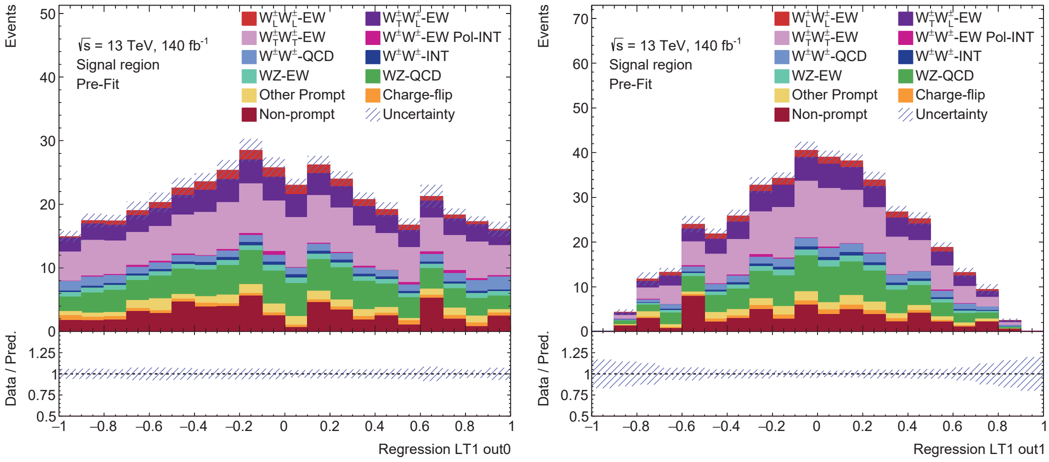
Figure C.6: Distributions of the training input in the signal region.



(a) Regression$_{\mathrm{LL},1}(l1)$ **(b)** Regression$_{\mathrm{LL},1}(l2)$

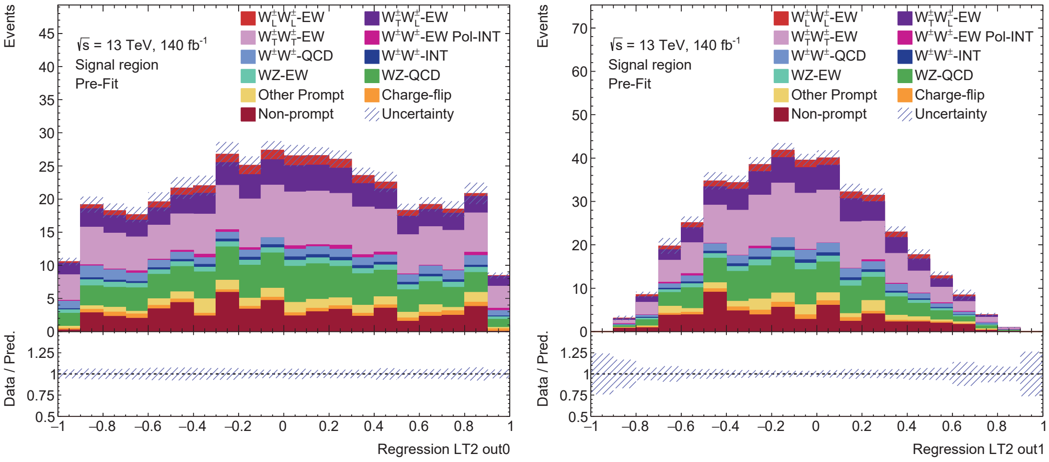


(c) Regression$_{\mathrm{LL},2}(l1)$ **(d)** Regression$_{\mathrm{LL},2}(l2)$

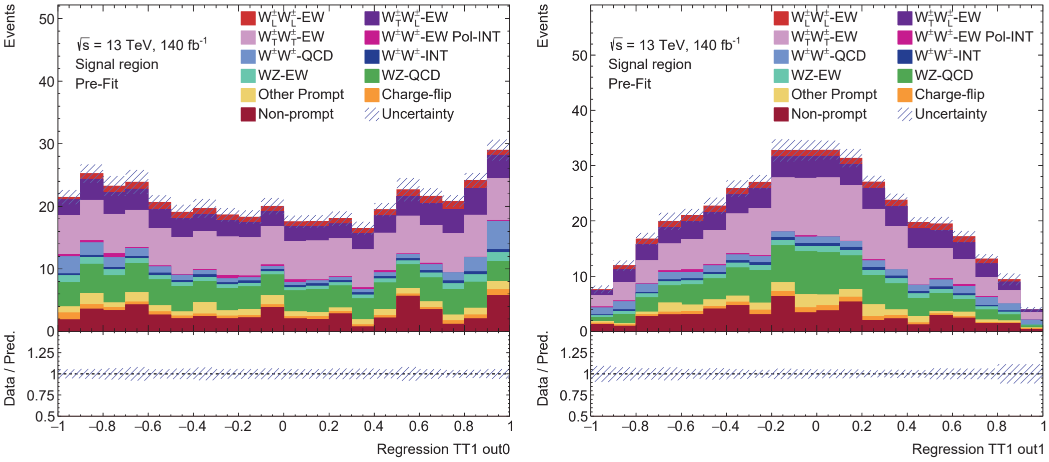


(e) Regression$_{\mathrm{LT},1}(l1)$ **(f)** Regression$_{\mathrm{LT},1}(l2)$

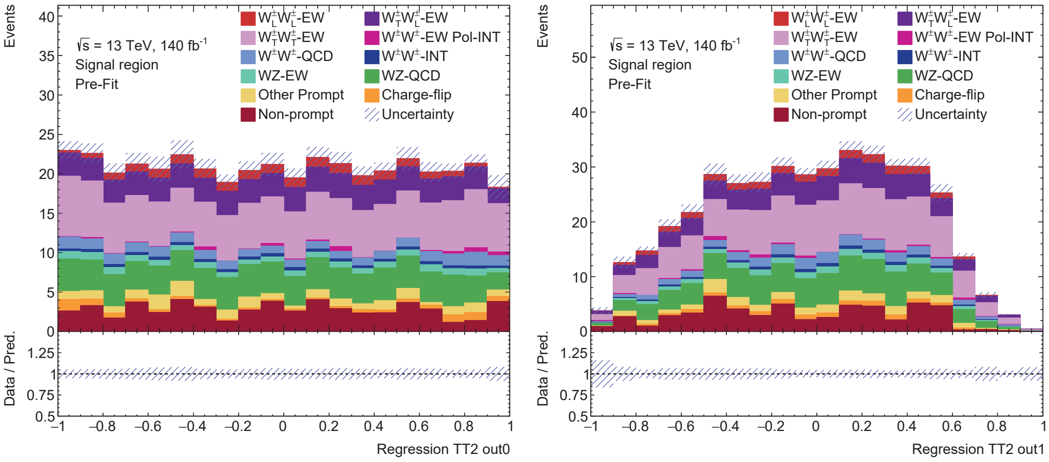
Figure C.7: Distributions of the training input in the signal region.



(a) $\text{Regression}_{\text{LT},2}(l1)$ **(b)** $\text{Regression}_{\text{LT},2}(l2)$

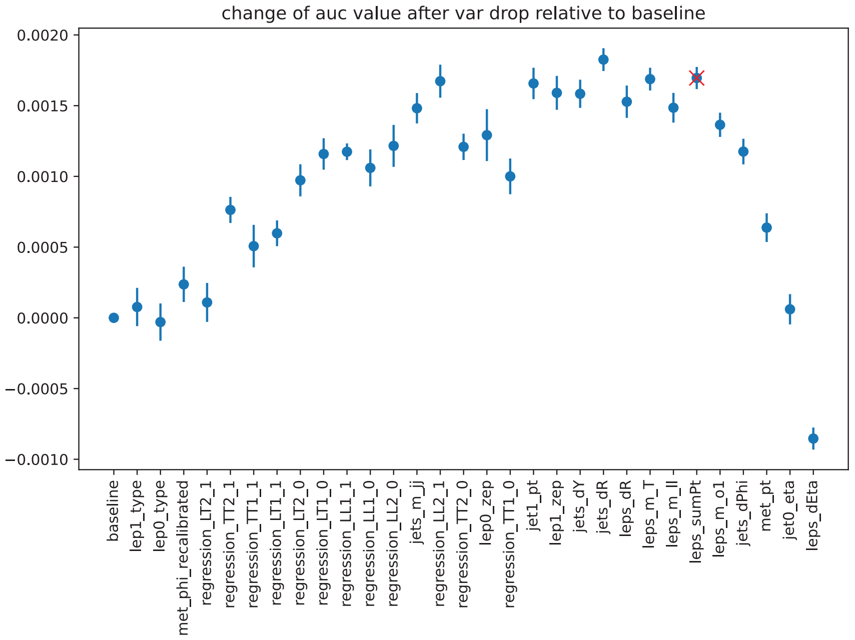


(c) $\text{Regression}_{\text{TT},1}(l1)$ **(d)** $\text{Regression}_{\text{TT},1}(l2)$

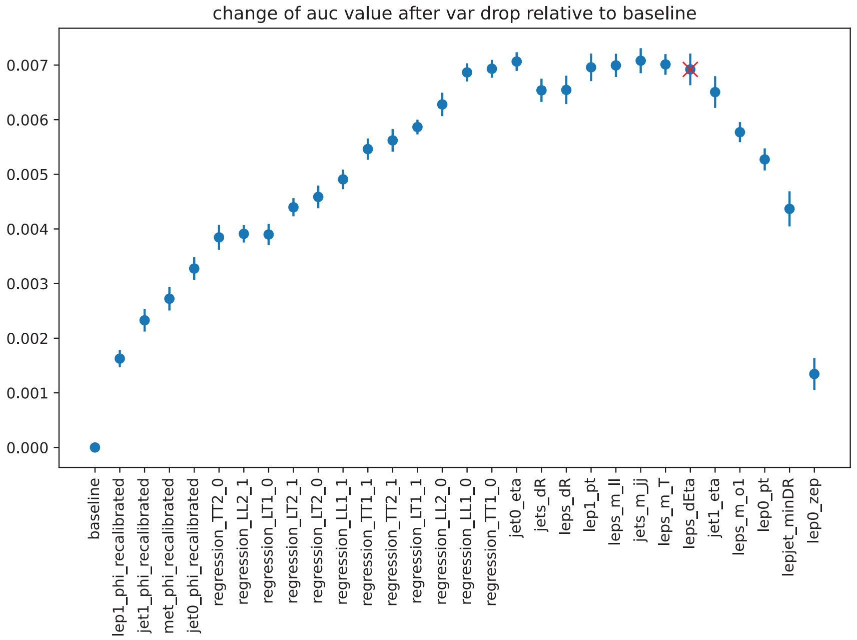


(e) $\text{Regression}_{\text{TT},2}(l1)$ **(f)** $\text{Regression}_{\text{TT},2}(l2)$

Figure C.8: Distributions of the training input in the signal region.

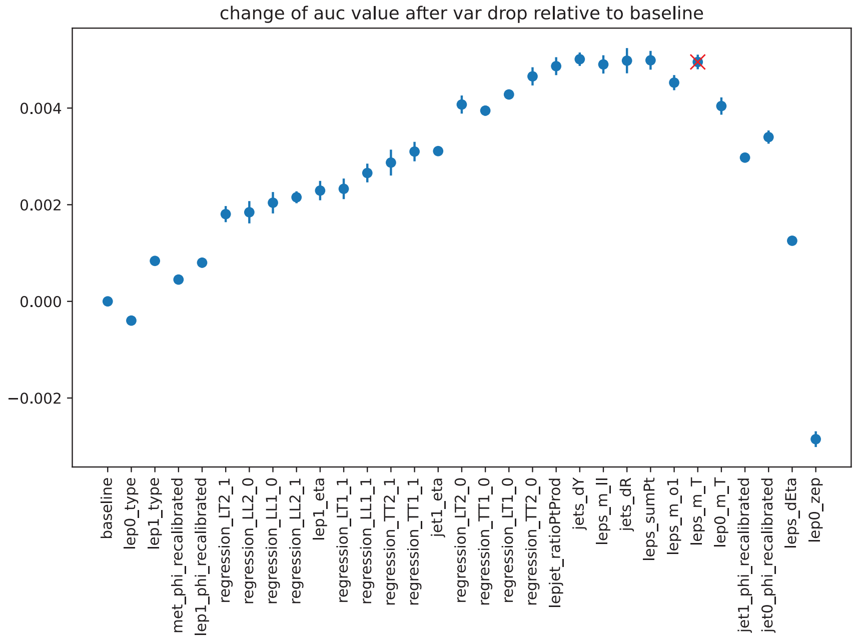


(a) DNN for LX in WW-cmf

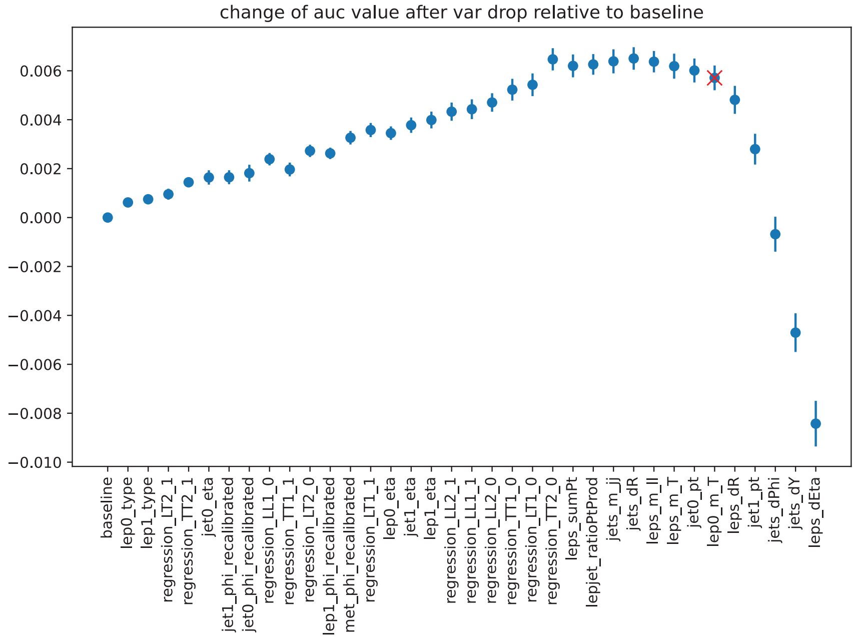


(b) DNN for $W^{\pm}W^{\pm}jj$-EW vs. bkg

Figure C.9: Optimization of input variables for DNN$_{W^{\pm}W^{\pm}}$ and DNN$_{pol}$ for LX in WW-cmf. Plotted is the relative AUC change of the variable whose removal results in the best mean AUC change after shuffling/retraining. If a significant performance decrease has been seen for five iterations, the procedure is stopped and the variable set that is still consistent with the best-performing set is chosen (marked with x).



(a) DNN for LL in pp-cmf

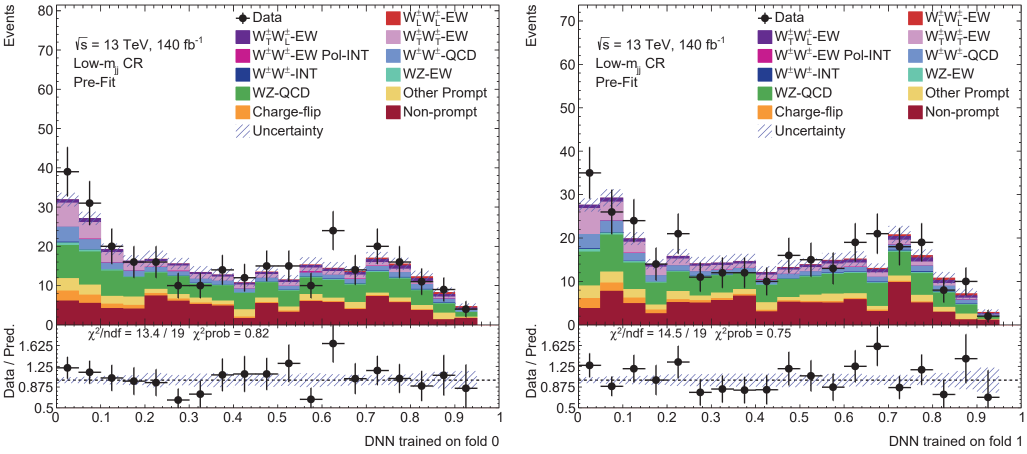


(b) DNN for LX in pp-cmf

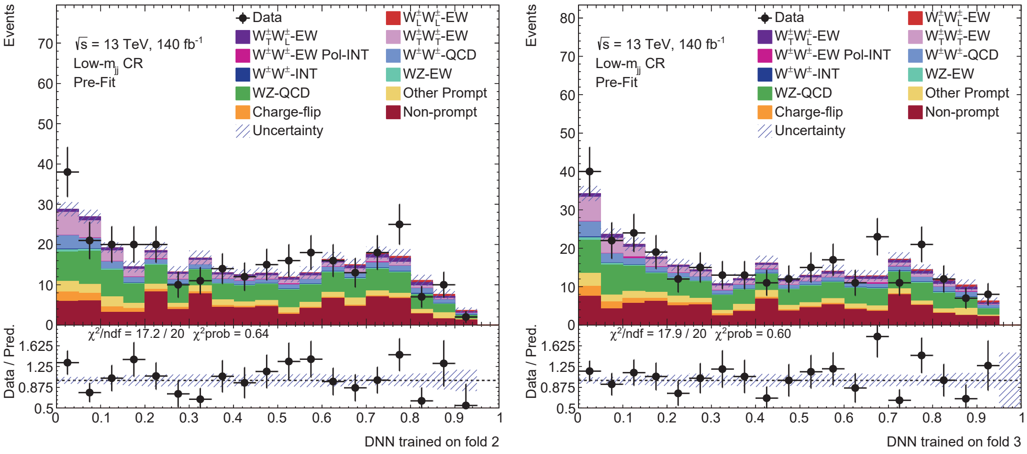
Figure C.10: Optimization of input variables for DNN$_{pol}$ for LL and LX in pp-cmf. Plotted is the relative AUC change of the variable whose removal results in the best mean AUC change after shuffling/retraining. If a significant performance decrease has been seen for five iterations, the procedure is stopped and the variable set that is still consistent with the best-performing set is chosen (marked with x).

Appendix D

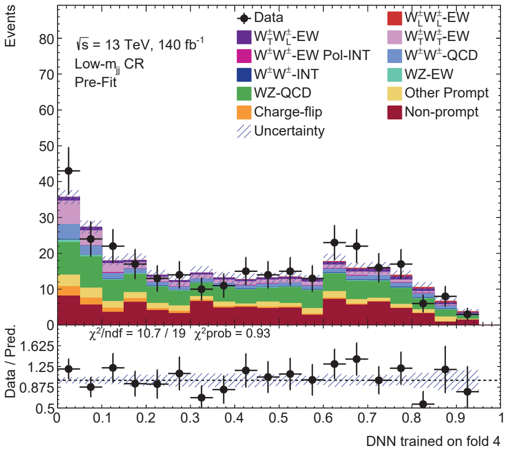
Data Closure Test



(a) Fold 0 **(b)** Fold 1

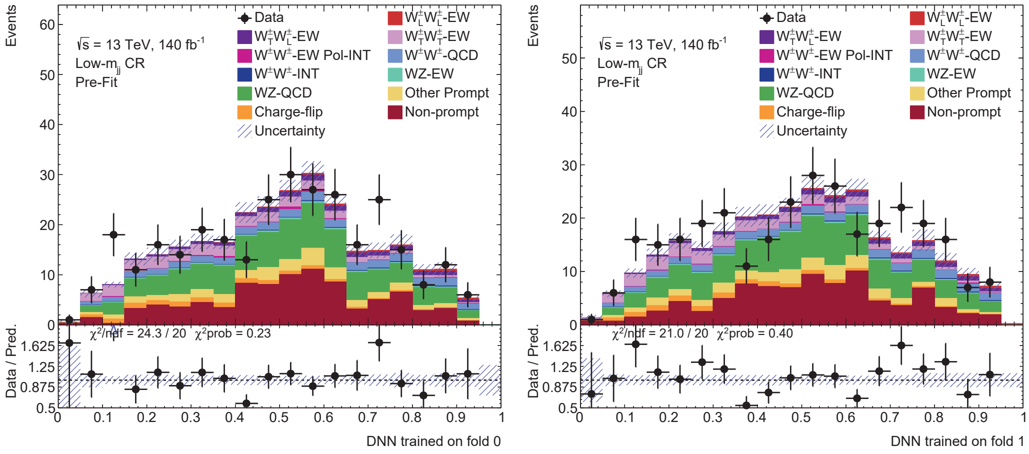


(c) Fold 2 **(d)** Fold 3



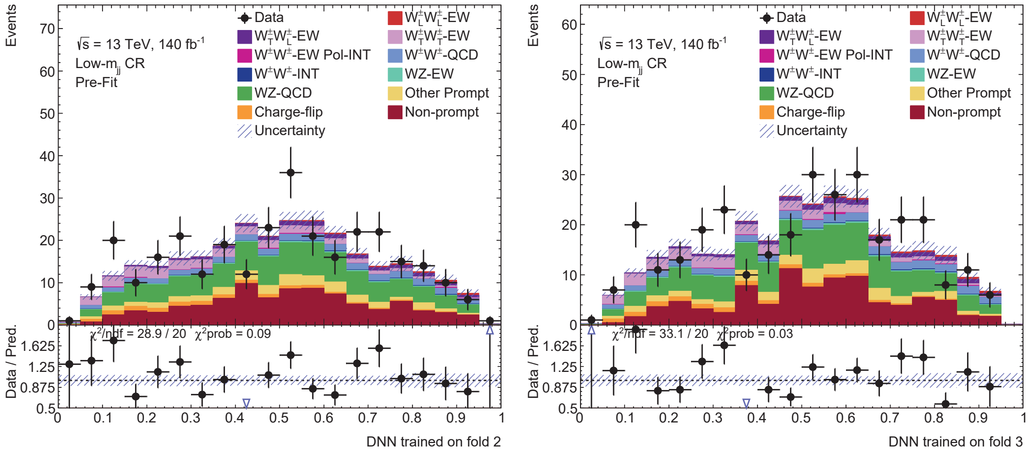
(e) Fold 4

Figure D.1: DNNs trained on different folds to seperate LL and TX in WW-cmf are applied in the Low-m_{jj} control region. According to the fit results from the unpolarized $W^\pm W^\pm jj$ measurement [8] the $W^\pm Z$-QCD contribution is scaled by 0.69 and the $W^\pm W^\pm jj$-EW signal is scaled by 1.14.



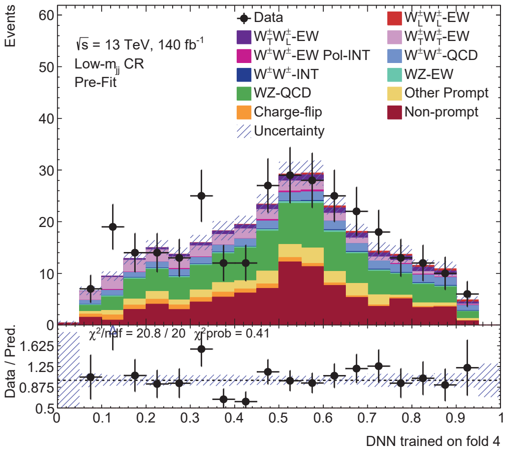
(a) Fold 0

(b) Fold 1



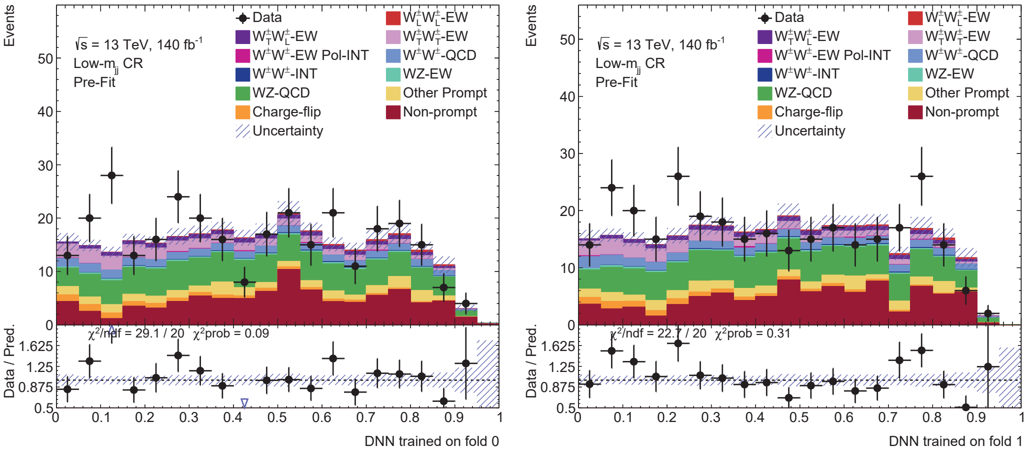
(c) Fold 2

(d) Fold 3

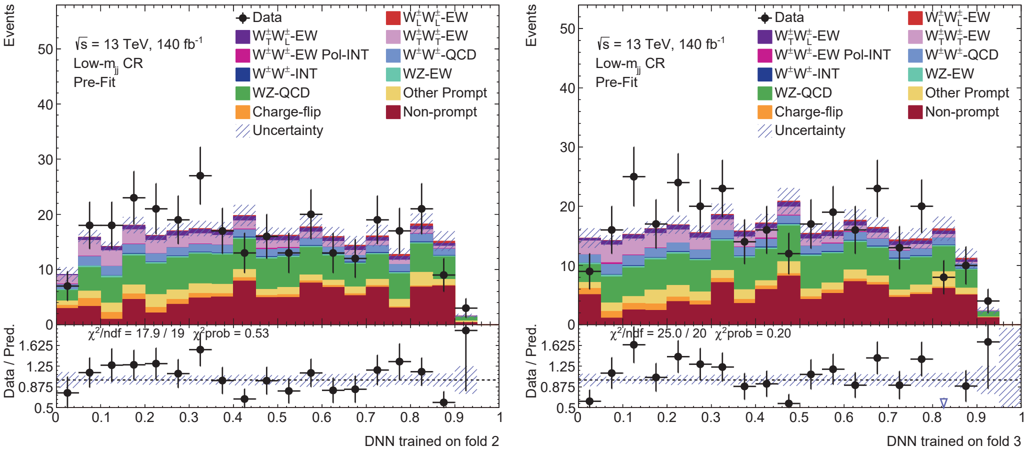


(e) Fold 4

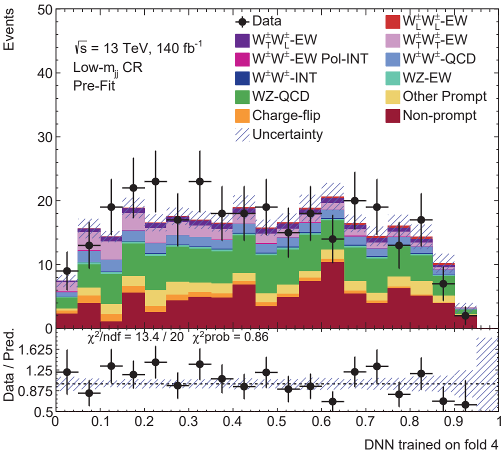
Figure D.2: DNNs trained on different folds to seperate LX and TT in WW-cmf are applied in the Low-m_{jj} control region. According to the fit results from the unpolarized $W^{\pm}W^{\pm}jj$ measurement [8] the $W^{\pm}Z$-QCD contribution is scaled by 0.69 and the $W^{\pm}W^{\pm}jj$-EW signal is scaled by 1.14.



(a) Fold 0 **(b)** Fold 1

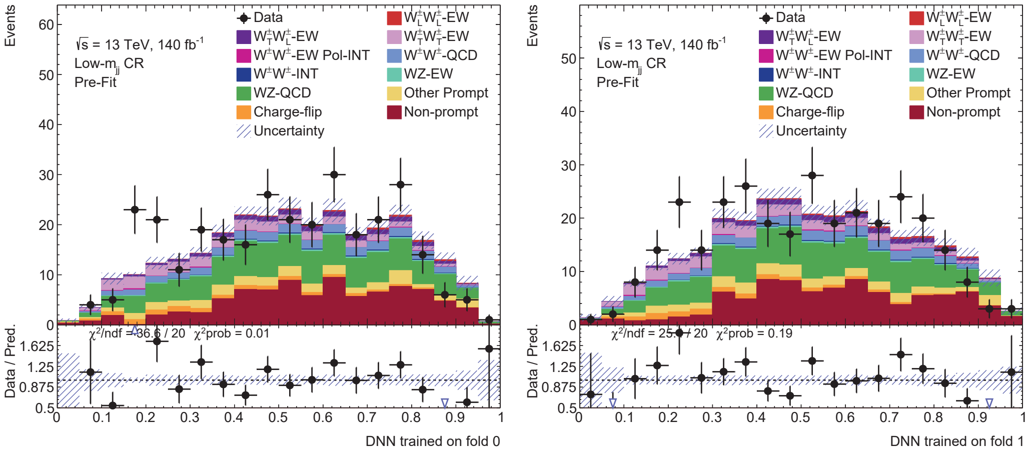


(c) Fold 2 **(d)** Fold 3



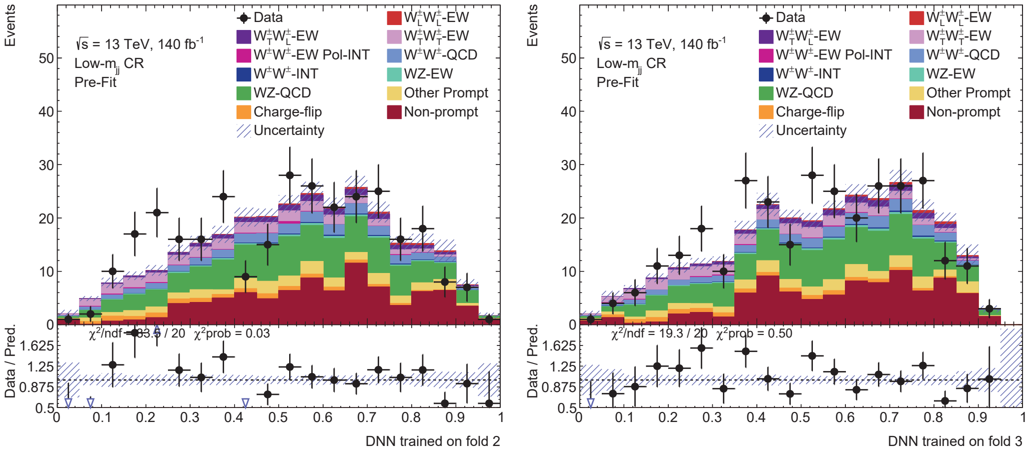
(e) Fold 4

Figure D.3: DNNs trained on different folds to seperate LL and TX in pp-cmf are applied in the Low-m_{jj} control region. According to the fit results from the unpolarized $W^\pm W^\pm jj$ measurement [8] the $W^\pm Z$-QCD contribution is scaled by 0.69 and the $W^\pm W^\pm jj$-EW signal is scaled by 1.14.



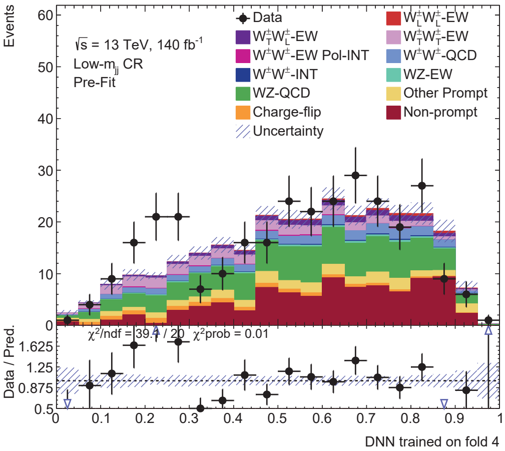
(a) Fold 0

(b) Fold 1



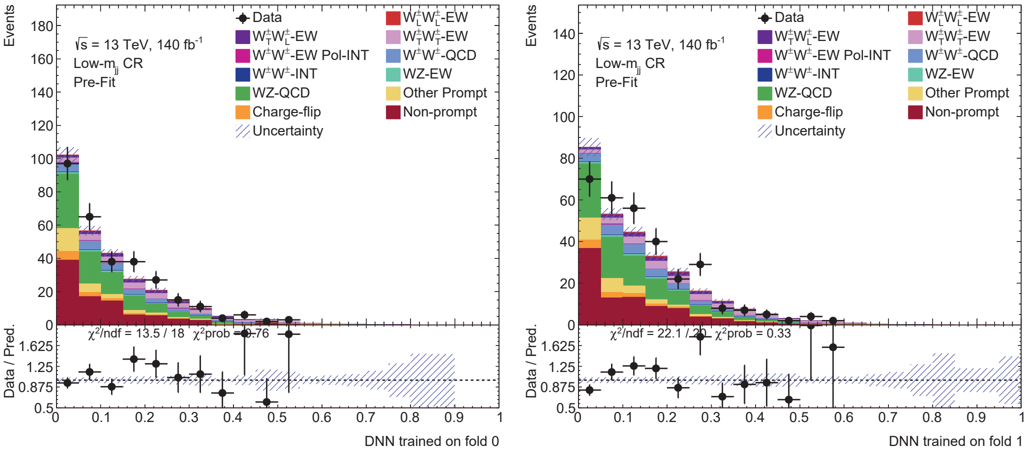
(c) Fold 2

(d) Fold 3

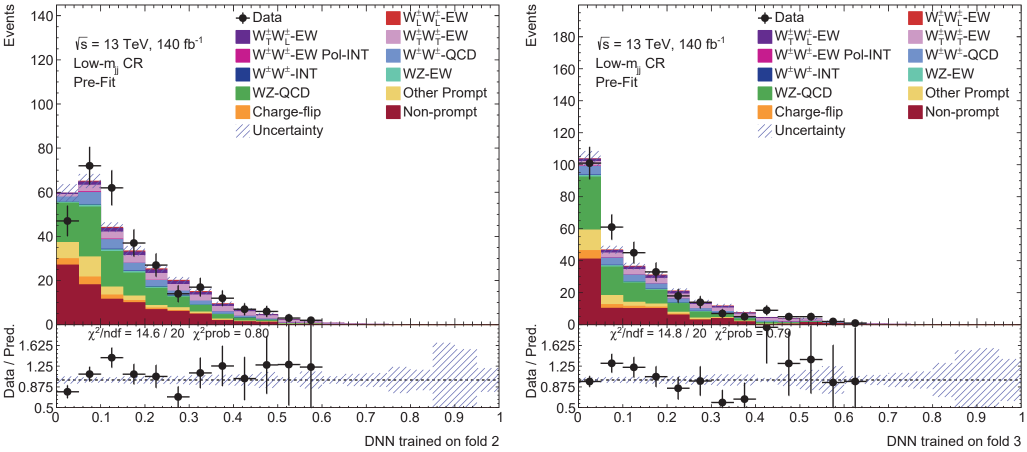


(e) Fold 4

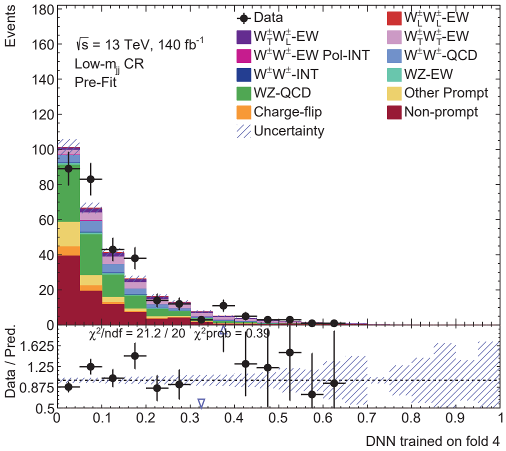
Figure D.4: DNNs trained on different folds to seperate LX and TT in pp-cmf are applied in the Low-m_{jj} control region. According to the fit results from the unpolarized $W^{\pm}W^{\pm}jj$ measurement [8] the $W^{\pm}Z$-QCD contribution is scaled by 0.69 and the $W^{\pm}W^{\pm}jj$-EW signal is scaled by 1.14.



(a) Fold 0 **(b)** Fold 1



(c) Fold 2 **(d)** Fold 3

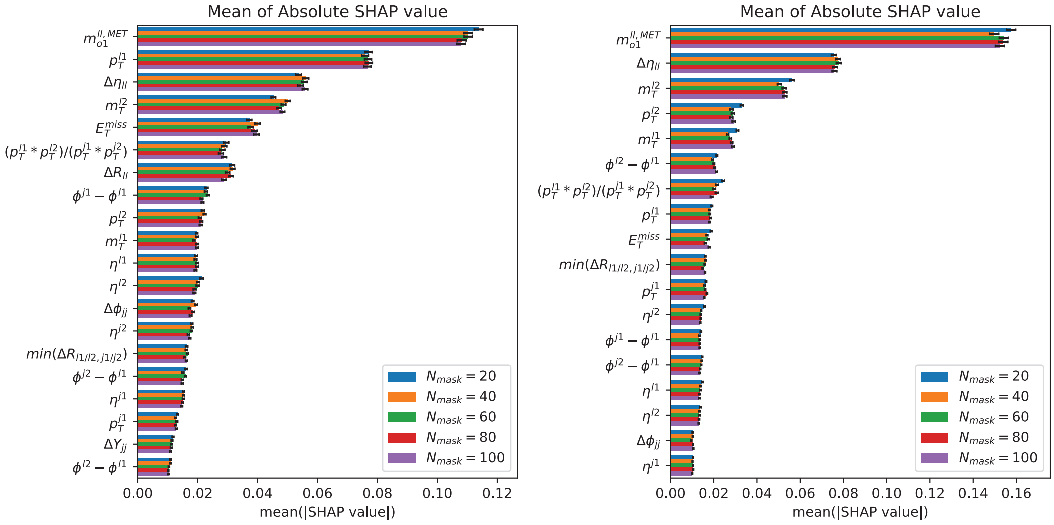


(e) Fold 4

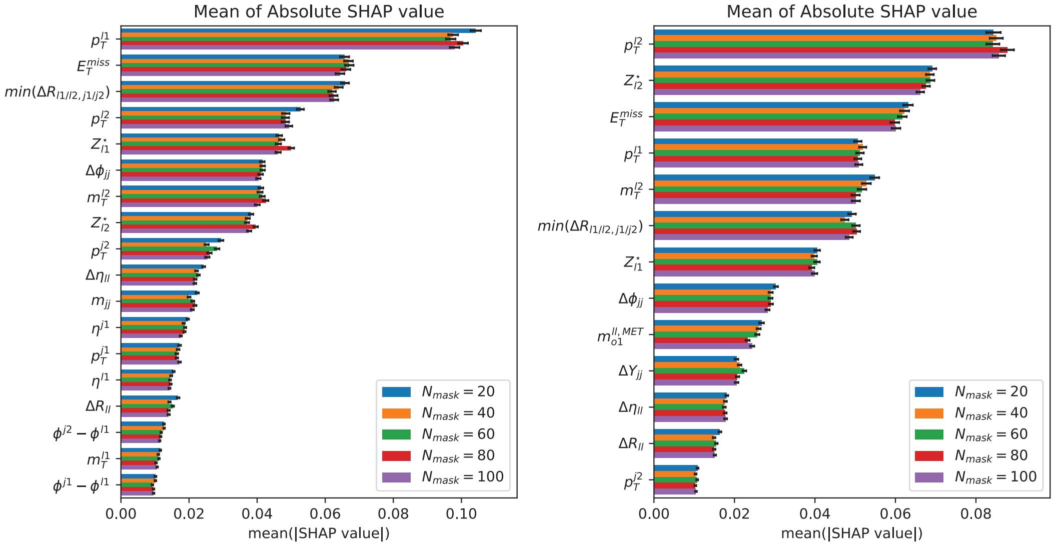
Figure D.5: DNNs trained on different folds to seperate $W^{\pm}W^{\pm}jj$-EW and background are applied in the Low-m_{jj} control region. According to the fit results from the unpolarized $W^{\pm}W^{\pm}jj$ measurement [8] the $W^{\pm}Z$-QCD contribution is scaled by 0.69 and the $W^{\pm}W^{\pm}jj$-EW signal is scaled by 1.14.

Appendix E

Interpretation of the Classifier Decision-Making

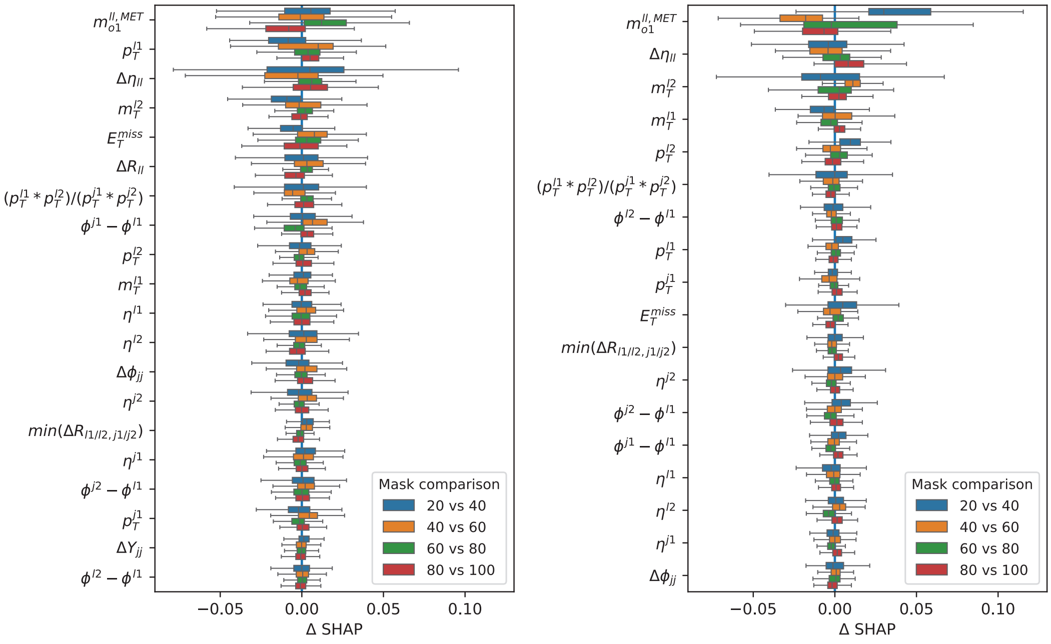


(a) LL in WW-cmf **(b)** LX in WW-cmf



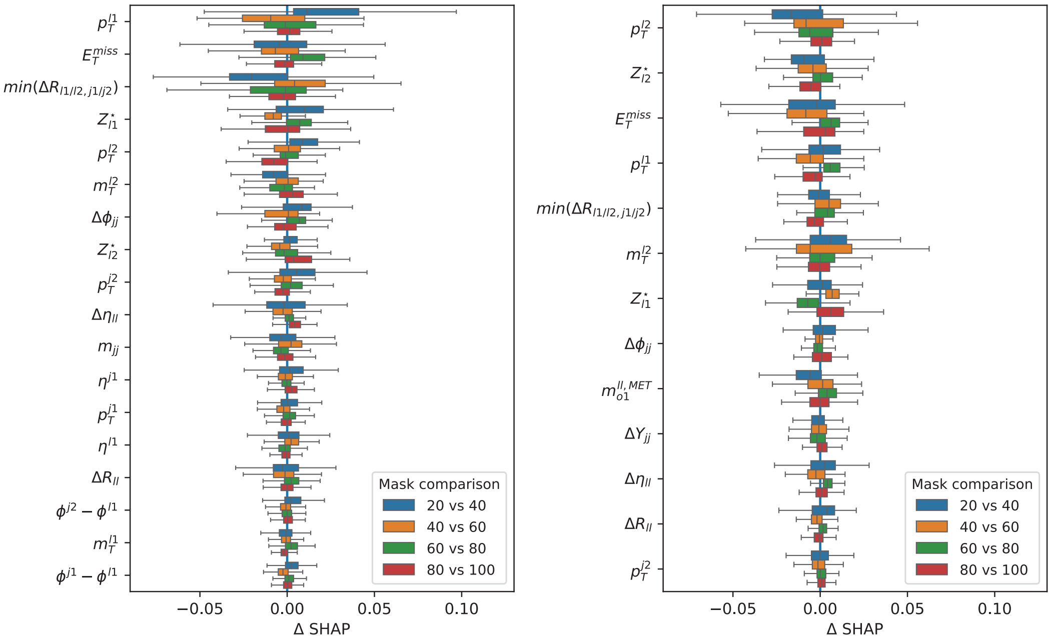
(c) LL in pp-cmf **(d)** LX in pp-cmf

Figure E.1: The mean absolute SHAP values of DNN$_{pol}$ for different input features. The assigned uncertainties originate from the statistical uncertainty of the explanation-dataset. The SHAP evaluation is done for independent mask-datasets of different size.



(a) LL in WW-cmf

(b) LX in WW-cmf



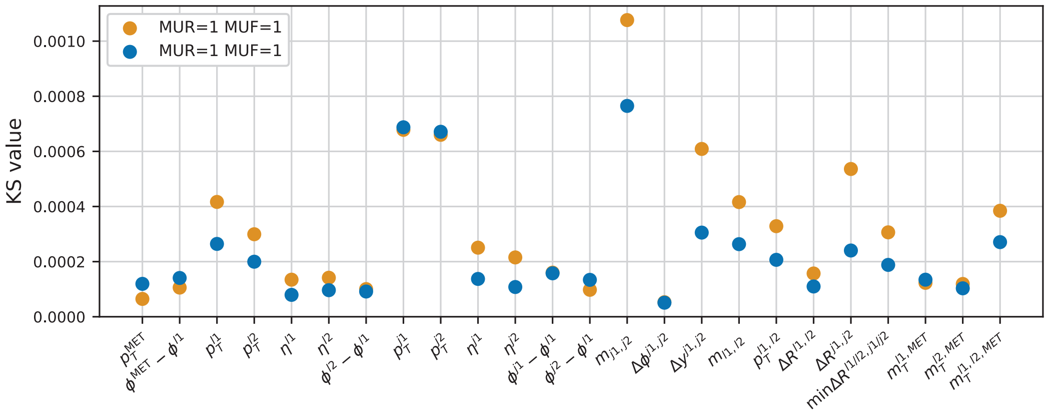
(c) LL in pp-cmf

(d) LX in pp-cmf

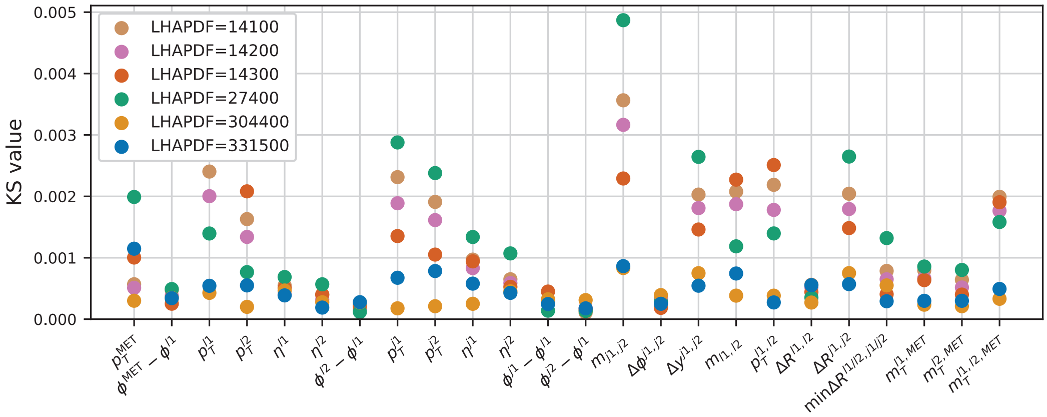
Figure E.2: The SHAP values of DNN$_{pol}$ originating from mask-datasets with increasing size are compared by calculate the event-wise difference. This is done for each time N_{mask} was increased by 20. For better visibility the individual differences are combined into a box plot of the quantiles and whiskers covering the full range.

Appendix F

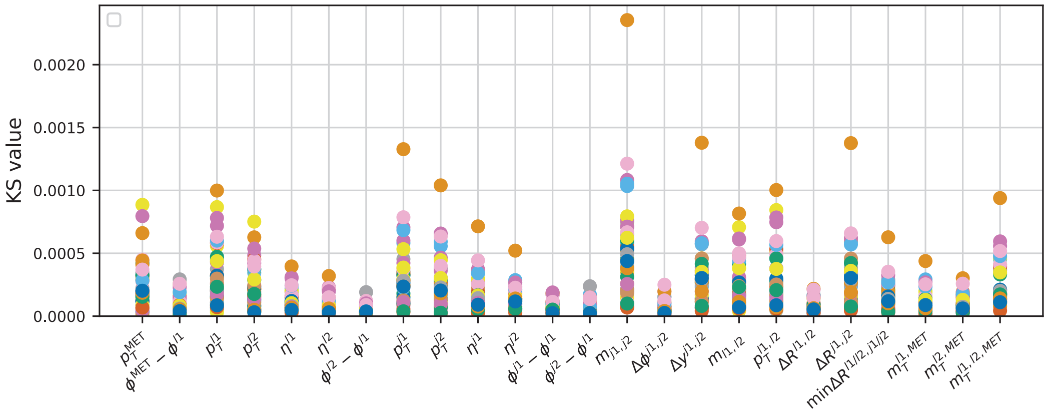
Uncertainties



(a) KS-values for α_S variations

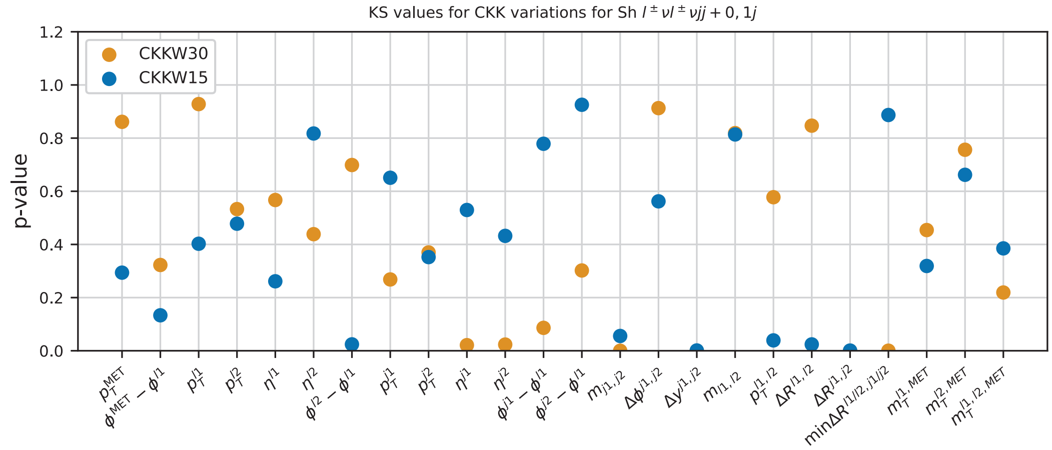


(b) KS-values for alternative PDF sets

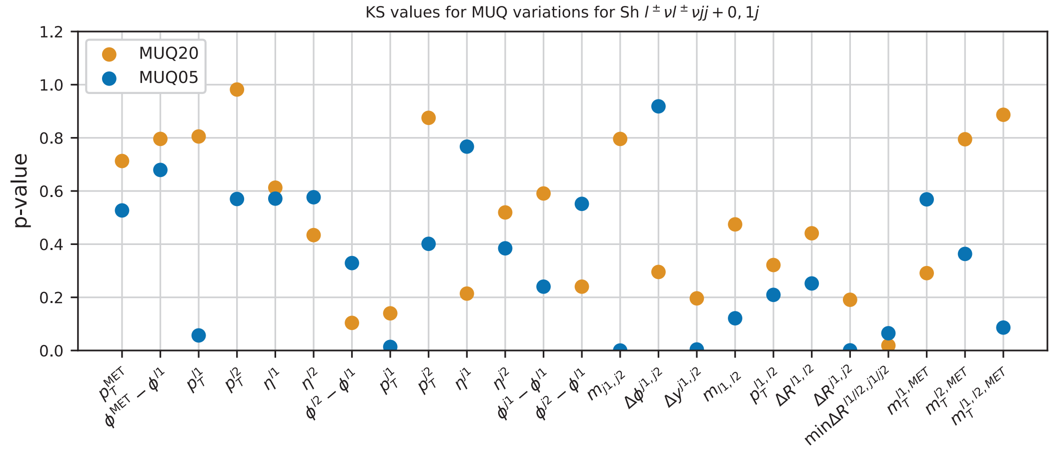


(c) KS-values for PDF variations

Figure F.1: Kolmogorov Smirnov test between the nominal Sherpa $l^{\pm}\nu l^{\pm}\nu jj + 0, 1j$ sample and the corresponding α_S, PDF set, and internal PDF variations. The kinematic distributions are derived on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2.



(a) KS p-values for merging scale variations



(b) KS p-values for resummation scale variations

Figure F.2: Kolmogorov Smirnov p-values for the comparison of the nominal Sherpa $l^{\pm}\nu l^{\pm}\nu jj+0,1j$ sample and the corresponding resummation and merging scale variations. The kinematic distributions are derived on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2.

Appendix G

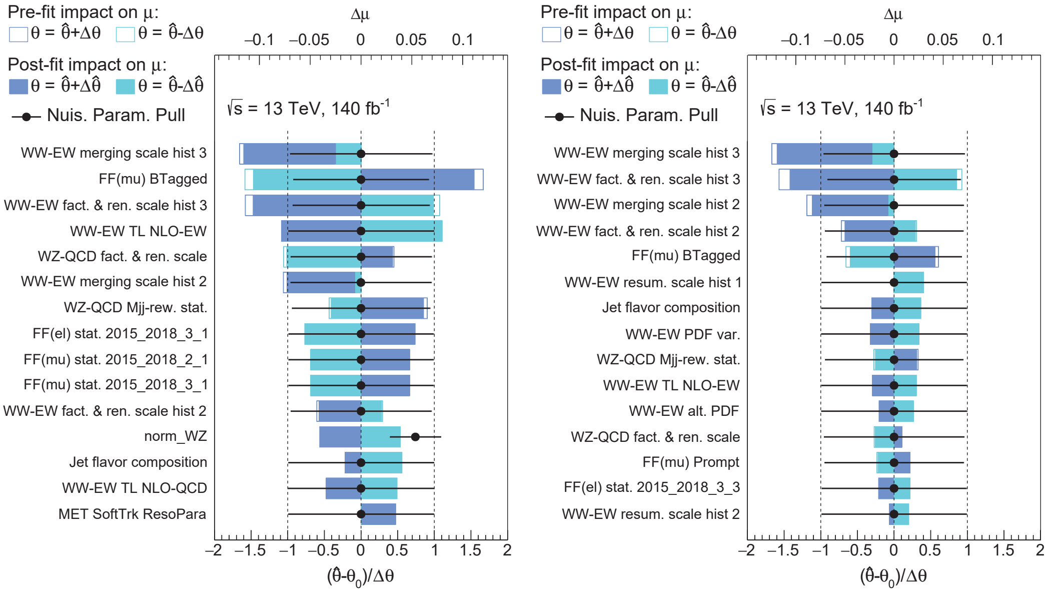
Statistical Evaluation

Table G.1: The pruned list of systematic variations and their corresponding category and sub-category.

Systematic variation	Main category	Sub-category
Luminosity	Lumi	Lumi
WW-EW LL NLO-QCD	Modeling	$W^{\pm}W^{p}mjj$-EW NLO correction
WW-EW TL NLO-QCD	Modeling	$W^{\pm}W^{p}mjj$-EW NLO correction
WW-EW TT NLO-QCD	Modeling	$W^{\pm}W^{p}mjj$-EW NLO correction
WW-EW TL NLO-EW	Modeling	$W^{\pm}W^{p}mjj$-EW NLO correction
WW-EW LL NLO-EW	Modeling	$W^{\pm}W^{p}mjj$-EW NLO correction
WW-EW TT NLO-EW	Modeling	$W^{\pm}W^{p}mjj$-EW NLO correction
NLO-EW-QCD comb. for TL	Modeling	$W^{\pm}W^{p}mjj$-EW NLO correction
NLO-EW-QCD comb. for LL	Modeling	$W^{\pm}W^{p}mjj$-EW NLO correction
NLO-EW-QCD comb. for TT	Modeling	$W^{\pm}W^{p}mjj$-EW NLO correction
WW-EW PDF var.	Modeling	$W^{\pm}W^{p}mjj$-EW theory unc.
WW-EW alt. PDF	Modeling	$W^{\pm}W^{p}mjj$-EW theory unc.
WW-EW alpha-S	Modeling	$W^{\pm}W^{p}mjj$-EW theory unc.
WW-EW fact. & ren. scale hist 1	Modeling	$W^{\pm}W^{p}mjj$-EW theory unc.
WW-EW merging scale hist 1	Modeling	$W^{\pm}W^{p}mjj$-EW theory unc.
WW-EW resum. scale hist 1	Modeling	$W^{\pm}W^{p}mjj$-EW theory unc.
WW-EW fact. & ren. scale hist 2	Modeling	$W^{\pm}W^{p}mjj$-EW theory unc.
WW-EW merging scale hist 2	Modeling	$W^{\pm}W^{p}mjj$-EW theory unc.
WW-EW resum. scale hist 2	Modeling	$W^{\pm}W^{p}mjj$-EW theory unc.
WW-EW fact. & ren. scale hist 3	Modeling	$W^{\pm}W^{p}mjj$-EW theory unc.
WW-EW merging scale hist 3	Modeling	$W^{\pm}W^{p}mjj$-EW theory unc.
WW-EW resum. scale hist 1	Modeling	$W^{\pm}W^{p}mjj$-EW theory unc.
WW-EW fact. & ren. scale LowMjj	Modeling	$W^{\pm}W^{p}mjj$-EW theory unc.
WW-EW merging scale LowMjj	Modeling	$W^{\pm}W^{p}mjj$-EW theory unc.
WW-EW resum. scale LowMjj	Modeling	$W^{\pm}W^{p}mjj$-EW theory unc.
WZ-QCD PDF	Modeling	$W^{\pm}Z$-QCD theory unc.
WZ-QCD fact. & ren. scale	Modeling	$W^{\pm}Z$-QCD theory unc.
WZ-QCD Mjj-rew. stat.	Modeling	$W^{\pm}Z$-QCD reweighting
WZ-QCD Mjj-rew. fact. & ren. scale	Modeling	$W^{\pm}Z$-QCD reweighting
WZ-QCD Mjj-rew. PDF	Modeling	$W^{\pm}Z$-QCD reweighting
Electron reconstruction	Experimental	Electron reconstruction
Jet effective NP modeling	Experimental	Jet reconstruction
Jet eta intercalibration modeling	Experimental	Jet reconstruction
Jet flavor composition	Experimental	Jet reconstruction
Jet flavor response	Experimental	Jet reconstruction
Jet pileup offset mu	Experimental	Jet reconstruction
Jet pileup offset NPV	Experimental	Jet reconstruction
Jet pileup rho topology	Experimental	Jet reconstruction
Jet JER effective NP PDsmear	Experimental	Jet reconstruction
Jet JER effective NP1 MCsmear	Experimental	Jet reconstruction
Jet JER effective NP4 MCsmear	Experimental	Jet reconstruction
Muon reconstruction	Experimental	Muon reconstruction
PRW data SF	Experimental	Pile-up
MET SoftTrk ResoPara	Experimental	MET reconstruction
MET SoftTrk ResoPerp	Experimental	MET reconstruction
MET SoftTrk Scale	Experimental	MET reconstruction
Charge-flip SF	Experimental	Charge-flip scale factor
FF(mu) stat. 2015_2018_1_1	Experimental	Fake factor method

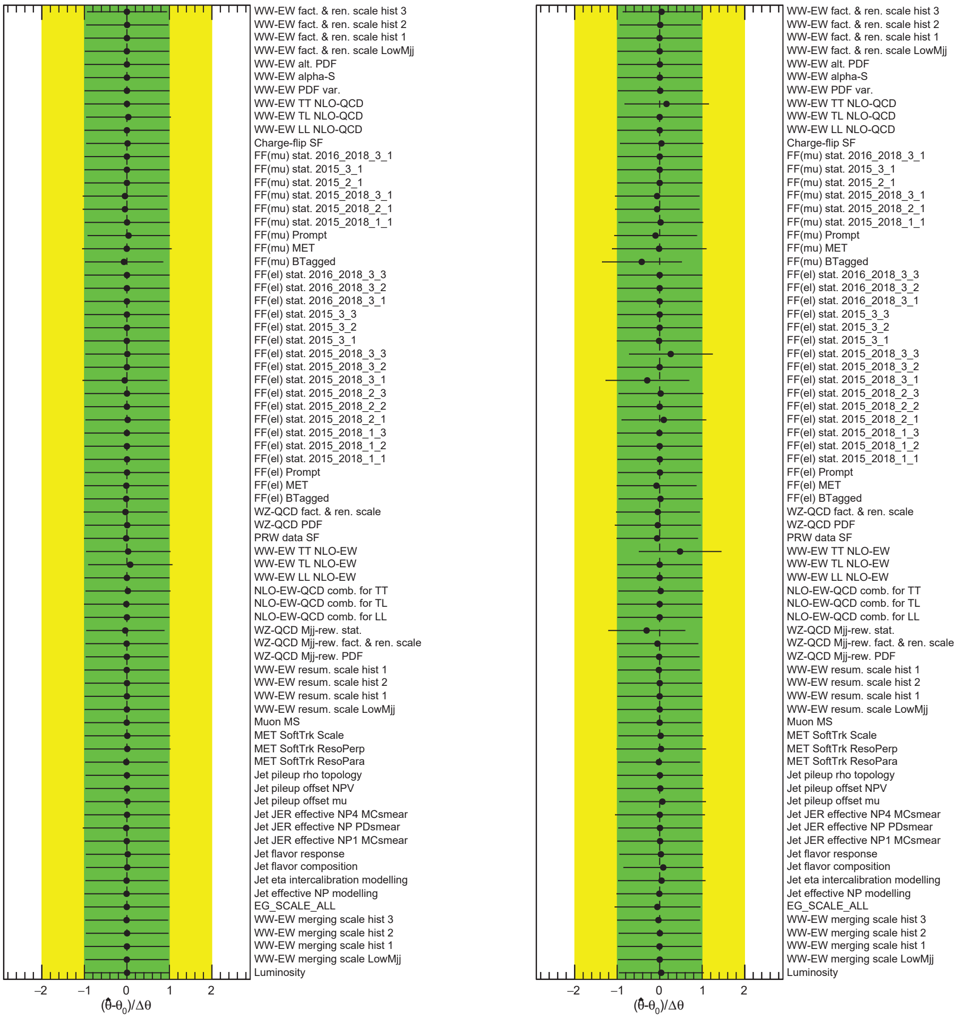
FF(mu) stat. 2015_2018_2_1	Experimental	Fake factor method
FF(mu) stat. 2015_2018_3_1	Experimental	Fake factor method
FF(mu) stat. 2015_2_1	Experimental	Fake factor method
FF(mu) stat. 2015_3_1	Experimental	Fake factor method
FF(mu) stat. 2016_2018_3_1	Experimental	Fake factor method
FF(el) stat. 2015_2018_1_1	Experimental	Fake factor method
FF(el) stat. 2015_2018_1_2	Experimental	Fake factor method
FF(el) stat. 2015_2018_1_3	Experimental	Fake factor method
FF(el) stat. 2015_2018_2_1	Experimental	Fake factor method
FF(el) stat. 2015_2018_2_2	Experimental	Fake factor method
FF(el) stat. 2015_2018_2_3	Experimental	Fake factor method
FF(el) stat. 2015_2018_3_1	Experimental	Fake factor method
FF(el) stat. 2015_2018_3_2	Experimental	Fake factor method
FF(el) stat. 2015_2018_3_3	Experimental	Fake factor method
FF(el) stat. 2015_3_1	Experimental	Fake factor method
FF(el) stat. 2015_3_2	Experimental	Fake factor method
FF(el) stat. 2015_3_3	Experimental	Fake factor method
FF(el) stat. 2016_2018_3_1	Experimental	Fake factor method
FF(el) stat. 2016_2018_3_2	Experimental	Fake factor method
FF(el) stat. 2016_2018_3_3	Experimental	Fake factor method
FF(mu) MET	Experimental	Fake factor method
FF(el) MET	Experimental	Fake factor method
FF(mu) Prompt	Experimental	Fake factor method
FF(el) Prompt	Experimental	Fake factor method
FF(mu) BTagged	Experimental	Fake factor method
FF(el) BTagged	Experimental	Fake factor method

G.1 Expected Results



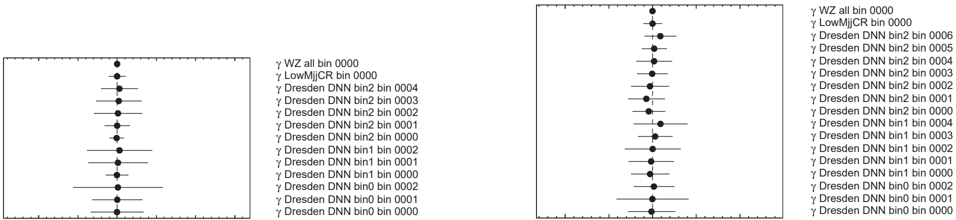
(a) Ranking of μ_{LL} measurement in pp-cmf **(b)** Ranking of μ_{LX} measurement in pp-cmf

Figure G.1: The ranking plots for the partially unblinded significance fits in the pp-cmf. The fit is derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74.



(a) $\boldsymbol{\theta}$ of LL fit in pp-cmf with $\mu_{LL} = 0$

(b) $\boldsymbol{\theta}$ of LX fit in pp-cmf with $\mu_{LX} = 0$



(c) γ of LL fit in pp-cmf with $\mu_{LL} = 0$

(d) γ of LX fit in pp-cmf with $\mu_{LX} = 0$

Figure G.2: The pulls of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ for the partially unblinded background-only significance fits in the pp-cmf. The fit parameter μ_{LL}/μ_{LX} is fixed to 0. The fit is derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74.

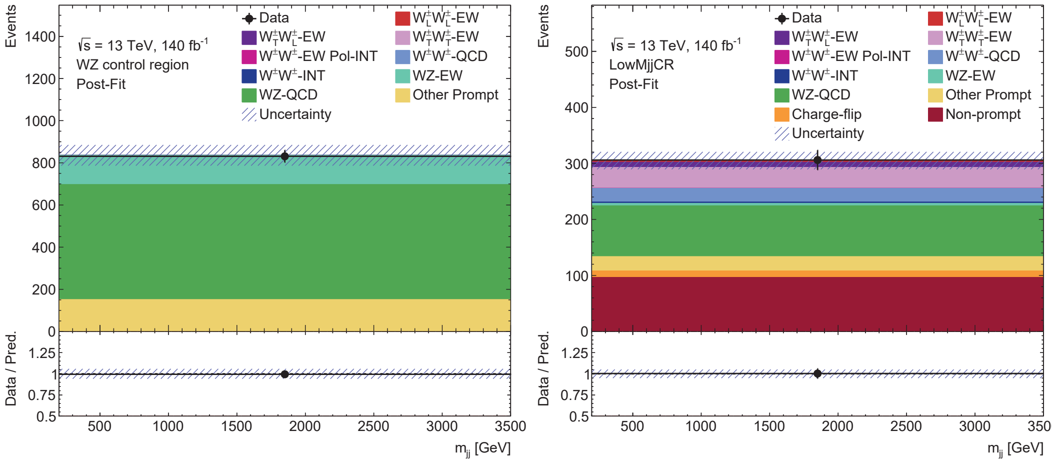
G.2 Observed Results

Table G.2: Pre-fit number of events in the regions used for the LL measurement in the WW-cmf. The signal region is split according to the output score of $\text{DNN}_{W^\pm W^\pm}$.

	$0 \leq \text{DNN}_{WW} < 0.2$	$0.2 \leq \text{DNN}_{WW} < 0.6$	$0.6 \leq \text{DNN}_{WW} \leq 1$	Low-m_{jj} CR	$W^\pm Z$ CR
$W^\pm_L W^\pm_L$-EW	0.99 ± 0.15	5.8 ± 0.8	13.3 ± 2.1	3.7 ± 0.6	$7{\cdot}10^{-5} \pm 1.4{\cdot}10^{-4}$
$W^\pm_T W^\pm_L$-EW	3.3 ± 0.5	18.7 ± 2.7	41 ± 7	11.1 ± 1.8	$6.1{\cdot}10^{-4} \pm 3.2{\cdot}10^{-4}$
$W^\pm_T W^\pm_T$-EW	5.2 ± 0.8	33 ± 5	90 ± 14	24.1 ± 3.9	$0.0019 \pm 7{\cdot}10^{-4}$
$W^\pm W^\pm$-EW Pol-INT	0.27 ± 0.05	1.36 ± 0.11	3.07 ± 0.17	0.14 ± 0.1	$3.1{\cdot}10^{-4} \pm 1.8{\cdot}10^{-4}$
$W^\pm W^\pm$-QCD	7.82 ± 0.31	12.5 ± 0.5	3.95 ± 0.17	24.6 ± 0.8	0.0057 ± 0.0025
$W^\pm W^\pm$-INT	1.02 ± 0.06	3.64 ± 0.08	2.98 ± 0.06	2.19 ± 0.07	0.0016 ± 0.0012
$W^\pm Z$-EW	1.98 ± 0.08	5.74 ± 0.19	7.35 ± 0.19	4.4 ± 0.12	139.5 ± 3.0
$W^\pm Z$-QCD	33 ± 15	34 ± 15	10 ± 5	120 ± 50	730 ± 340
Top, triboson	1.36 ± 0.16	2.24 ± 0.19	1.07 ± 0.12	6.11 ± 0.3	57.0 ± 1.6
ZZ	1.42 ± 0.07	0.79 ± 0.05	0.25 ± 0.027	3.76 ± 0.14	58.6 ± 0.7
$V\gamma$	5.5 ± 1.1	4.9 ± 0.9	1.0 ± 0.43	15.6 ± 2.7	36 ± 6
Charge-flip	5.1 ± 1.7	3.6 ± 1.4	1.4 ± 0.6	11.1 ± 4.5	0 ± 0
Non-prompt	20.3 ± 4.4	23 ± 5	12 ± 5	89 ± 14	0 ± 0
Total	88 ± 16	149 ± 18	189 ± 24	310 ± 50	1020 ± 340
Data	99	162	214	306	831

Table G.3: Pre-fit number of events in the regions used for the LX measurement in the WW-cmf. The signal region is split according to the output score of $\text{DNN}_{W^\pm W^\pm}$.

	$0 \leq \text{DNN}_{WW} < 0.3$	$0.3 \leq \text{DNN}_{WW} < 0.7$	$0.7 \leq \text{DNN}_{WW} \leq 1$	Low-m_{jj} CR	$W^\pm Z$ CR
$W^\pm_L W^\pm_L$-EW	1.95 ± 0.29	7.2 ± 1.1	10.9 ± 1.8	3.7 ± 0.6	$7{\cdot}10^{-5} \pm 1.4{\cdot}10^{-4}$
$W^\pm_T W^\pm_L$-EW	6.5 ± 0.96	23.2 ± 3.4	34 ± 6	11.1 ± 1.8	$6.1{\cdot}10^{-4} \pm 3.2{\cdot}10^{-4}$
$W^\pm_T W^\pm_T$-EW	10.4 ± 1.5	43 ± 6	75 ± 12	24.1 ± 3.9	$0.0019 \pm 7{\cdot}10^{-4}$
$W^\pm W^\pm$-EW Pol-INT	0.56 ± 0.07	2.03 ± 0.13	2.86 ± 0.15	0.14 ± 0.1	$3.1{\cdot}10^{-4} \pm 1.8{\cdot}10^{-4}$
$W^\pm W^\pm$-QCD	11.7 ± 0.5	10.39 ± 0.41	2.15 ± 0.11	24.6 ± 0.8	0.0057 ± 0.0025
$W^\pm W^\pm$-INT	1.73 ± 0.05	3.994 ± 0.099	1.92 ± 0.05	2.19 ± 0.07	0.0016 ± 0.0012
$W^\pm Z$-EW	3.31 ± 0.12	6.22 ± 0.17	5.54 ± 0.18	4.4 ± 0.12	139.5 ± 3.0
$W^\pm Z$-QCD	44 ± 20	27 ± 13	6.0 ± 3.1	120 ± 50	730 ± 340
Top, triboson	1.99 ± 0.2	2.0 ± 0.16	0.69 ± 0.11	6.11 ± 0.3	57.0 ± 1.6
ZZ	1.71 ± 0.08	0.62 ± 0.045	0.126 ± 0.02	3.76 ± 0.14	58.6 ± 0.7
$V\gamma$	7.6 ± 1.3	3.3 ± 0.8	0.62 ± 0.38	15.6 ± 2.7	36 ± 6
Charge-flip	6.5 ± 2.1	2.7 ± 1.2	0.96 ± 0.42	11.1 ± 4.5	0 ± 0
Non-prompt	29 ± 6	18.8 ± 4.5	9.3 ± 3.9	89 ± 14	0 ± 0
Total	127 ± 21	151 ± 18	149 ± 20	310 ± 50	1020 ± 340
Data	142	158	175	306	831



(a) $W^\pm Z$ control region

(b) Low-m_{jj} control region

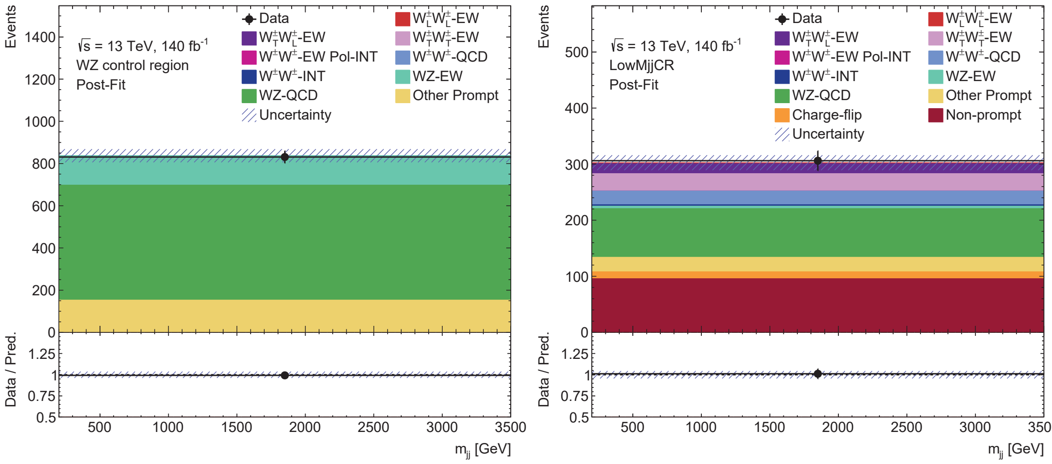
Figure G.3: The post-fit distribution for the LX in WW-cmf measurement in the $W^\pm Z$ control region and Low-m_{jj} control region.

Table G.4: Pre-fit number of events in the regions used for the LL measurement in the pp-cmf. The signal region is split according to the output score of $\mathrm{DNN}_{W^\pm W^\pm}$.

	$0 \leq \mathrm{DNN}_{WW} < 0.15$	$0.15 \leq \mathrm{DNN}_{WW} < 0.5$	$0.5 \leq \mathrm{DNN}_{WW} \leq 1$	Low-m_{jj} CR	$W^\pm Z$ CR
$W^\pm_L W^\pm_L$-EW	0.58 ± 0.09	2.89 ± 0.43	9.3 ± 1.5	2.22 ± 0.36	$5.2{\cdot}10^{-5} \pm 4.3{\cdot}10^{-5}$
$W^\pm_T W^\pm_L$-EW	2.49 ± 0.37	15.6 ± 2.3	56 ± 9	13.4 ± 2.1	$5.8{\cdot}10^{-4} \pm 3.3{\cdot}10^{-4}$
$W^\pm_T W^\pm_T$-EW	2.91 ± 0.44	22.4 ± 3.3	99 ± 16	23.3 ± 3.8	$0.002 \pm 8{\cdot}10^{-4}$
$W^\pm W^\pm$-EW Pol-INT	0.179 ± 0.042	0.58 ± 0.09	1.64 ± 0.17	0.013 ± 0.1	$2.5{\cdot}10^{-4} \pm 2.4{\cdot}10^{-4}$
$W^\pm W^\pm$-QCD	5.74 ± 0.22	12.2 ± 0.5	6.35 ± 0.27	24.6 ± 0.8	0.0057 ± 0.0025
$W^\pm W^\pm$-INT	0.722 ± 0.037	2.87 ± 0.07	4.05 ± 0.09	2.19 ± 0.07	0.0016 ± 0.0012
$W^\pm Z$-EW	1.38 ± 0.07	4.87 ± 0.18	8.82 ± 0.21	4.4 ± 0.12	139.5 ± 3.0
$W^\pm Z$-QCD	27 ± 12	34 ± 15	16 ± 8	120 ± 50	730 ± 340
Top, triboson	1.01 ± 0.13	2.07 ± 0.19	1.6 ± 0.14	6.11 ± 0.3	57.0 ± 1.6
ZZ	1.25 ± 0.07	0.82 ± 0.05	0.389 ± 0.034	3.76 ± 0.14	58.6 ± 0.7
$V\gamma$	4.47 ± 0.96	5.5 ± 1.1	1.3 ± 0.5	15.6 ± 2.7	36 ± 6
Charge-flip	4.2 ± 1.3	4.0 ± 1.5	2.0 ± 0.9	11.1 ± 4.5	0 ± 0
Non-prompt	17.9 ± 3.9	21 ± 5	16 ± 5	89 ± 14	0 ± 0
Total	70 ± 13	129 ± 17	224 ± 28	310 ± 50	1020 ± 340
Data	81	132	262	306	831

Table G.5: Pre-fit number of events in the regions used for the LX measurement in the pp-cmf. The signal region is split according to the output score of $\mathrm{DNN}_{W^\pm W^\pm}$.

	$0 \leq \mathrm{DNN}_{WW} < 0.25$	$0.25 \leq \mathrm{DNN}_{WW} < 0.6$	$0.6 \leq \mathrm{DNN}_{WW} \leq 1$	Low-m_{jj} CR	$W^\pm Z$ CR
$W^\pm_L W^\pm_L$-EW	1.19 ± 0.18	3.5 ± 0.5	8.0 ± 1.3	2.22 ± 0.36	$5.2{\cdot}10^{-5} \pm 4.3{\cdot}10^{-5}$
$W^\pm_T W^\pm_L$-EW	5.5 ± 0.8	19.7 ± 2.9	49 ± 8	13.4 ± 2.1	$5.8{\cdot}10^{-4} \pm 3.3{\cdot}10^{-4}$
$W^\pm_T W^\pm_T$-EW	7.0 ± 1.0	29.8 ± 4.4	88 ± 14	23.3 ± 3.8	$0.002 \pm 8{\cdot}10^{-4}$
$W^\pm W^\pm$-EW Pol-INT	0.39 ± 0.06	1.27 ± 0.1	2.78 ± 0.16	0.013 ± 0.1	$2.5{\cdot}10^{-4} \pm 2.4{\cdot}10^{-4}$
$W^\pm W^\pm$-QCD	9.79 ± 0.39	10.52 ± 0.43	3.95 ± 0.17	24.6 ± 0.8	0.0057 ± 0.0025
$W^\pm W^\pm$-INT	1.36 ± 0.05	3.3 ± 0.09	2.98 ± 0.06	2.19 ± 0.07	0.0016 ± 0.0012
$W^\pm Z$-EW	2.61 ± 0.11	5.11 ± 0.16	7.35 ± 0.19	4.4 ± 0.12	139.5 ± 3.0
$W^\pm Z$-QCD	39 ± 18	28 ± 13	10 ± 5	120 ± 50	730 ± 340
Top, triboson	1.7 ± 0.18	1.9 ± 0.17	1.07 ± 0.12	6.11 ± 0.3	57.0 ± 1.6
ZZ	1.58 ± 0.08	0.63 ± 0.05	0.25 ± 0.027	3.76 ± 0.14	58.6 ± 0.7
$V\gamma$	6.4 ± 1.2	4.1 ± 0.9	1.02 ± 0.43	15.6 ± 2.7	36 ± 6
Charge-flip	5.9 ± 1.9	2.8 ± 1.2	1.4 ± 0.6	11.1 ± 4.5	0 ± 0
Non-prompt	26 ± 5	17.3 ± 4.1	12.4 ± 4.5	89 ± 14	0 ± 0
Total	109 ± 19	128 ± 16	189 ± 24	310 ± 50	1020 ± 340
Data	126	135	214	306	831



(a) $W^\pm Z$ control region

(b) Low-m_{jj} control region

Figure G.4: The post-fit distribution for the LL in pp-cmf measurement in the $W^\pm Z$ control region and Low-m_{jj} control region.

Table G.6: Post-fit number of events in the regions used for the LL measurement in the WW-cmf. The signal region is split according to the output score of $\text{DNN}_{W^\pm W^\pm}$.

	$0 \leq \text{DNN}_{WW} < 0.2$	$0.2 \leq \text{DNN}_{WW} < 0.6$	$0.6 \leq \text{DNN}_{WW} \leq 1$	Low-m_{jj} CR	$W^\pm Z$ CR
$W^\pm_L W^\pm_L$-EW	0.1 ± 0.8	0.3 ± 4.3	1 ± 9	0.2 ± 2.7	$0 \pm 5{\cdot}10^{-5}$
$W^\pm_T W^\pm_L$-EW	5.1 ± 1.2	26.5 ± 3.9	55 ± 6	15.6 ± 3.3	$8.4{\cdot}10^{-4} \pm 2.2{\cdot}10^{-4}$
$W^\pm_T W^\pm_T$-EW	8.1 ± 1.9	46 ± 7	120 ± 12	34 ± 7	$0.00262 \pm 3.7{\cdot}10^{-4}$
$W^\pm W^\pm$-EW Pol-INT	0.275 ± 0.013	1.364 ± 0.041	3.07 ± 0.06	0.1425 ± 0.0032	$3.1{\cdot}10^{-4} \pm 2.9{\cdot}10^{-6}$
$W^\pm W^\pm$-QCD	8.04 ± 0.44	12.7 ± 0.6	3.99 ± 0.19	24.8 ± 1.1	0.0058 ± 0.0011
$W^\pm W^\pm$-INT	1.05 ± 0.07	3.66 ± 0.11	2.98 ± 0.06	2.2 ± 0.06	$0.0015 \pm 6{\cdot}10^{-4}$
$W^\pm Z$-EW	2.03 ± 0.09	5.82 ± 0.22	7.37 ± 0.19	4.43 ± 0.13	140.5 ± 3.3
$W^\pm Z$-QCD	26.1 ± 2.8	25.4 ± 3.3	7.9 ± 1.5	89 ± 12	544 ± 38
Top, triboson	1.4 ± 0.12	2.24 ± 0.1	1.08 ± 0.05	6.13 ± 0.22	57.3 ± 1.6
ZZ	1.44 ± 0.06	0.79 ± 0.023	0.253 ± 0.015	3.76 ± 0.09	58.7 ± 0.6
$V\gamma$	5.9 ± 0.9	5.1 ± 0.5	0.99 ± 0.12	16.5 ± 2.2	38 ± 6
Charge-flip	5.7 ± 1.7	4.0 ± 1.4	1.5 ± 0.7	12 ± 5	0 ± 0
Non-prompt	22.1 ± 2.9	25.1 ± 3.7	13.7 ± 3.6	95 ± 12	0 ± 0
Total	87 ± 5	159.3 ± 9.6	218 ± 15	303 ± 14	838 ± 36
Data	99	162	214	306	831

Table G.7: Post-fit number of events in the regions used for the LX measurement in the WW-cmf. The signal region is split according to the output score of $\text{DNN}_{W^\pm W^\pm}$.

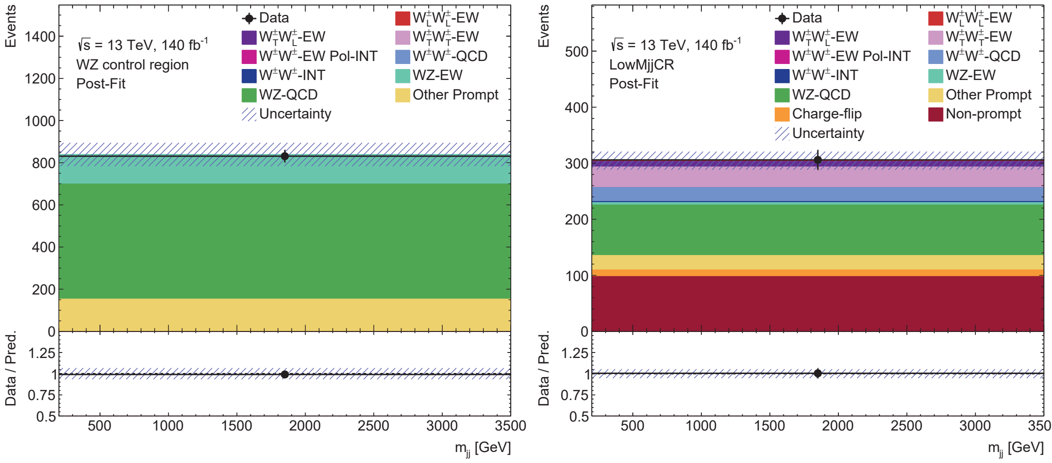
	$0 \leq \text{DNN}_{WW} < 0.3$	$0.3 \leq \text{DNN}_{WW} < 0.7$	$0.7 \leq \text{DNN}_{WW} \leq 1$	Low-m_{jj} CR	$W^\pm Z$ CR
$W^\pm_L W^\pm_L$-EW	1.9 ± 0.8	5.8 ± 1.9	8.5 ± 2.8	3.0 ± 1.1	$6{\cdot}10^{-5} \pm 1.1{\cdot}10^{-4}$
$W^\pm_T W^\pm_L$-EW	6.3 ± 2.5	18 ± 6	26 ± 8	9.1 ± 3.4	$5{\cdot}10^{-4} \pm 2{\cdot}10^{-4}$
$W^\pm_T W^\pm_T$-EW	18.9 ± 4.5	64.6 ± 10.0	112 ± 15	37 ± 8	$0.0029 \pm 5{\cdot}10^{-4}$
$W^\pm W^\pm$-EW Pol-INT	0.568 ± 0.018	2.03 ± 0.05	2.87 ± 0.05	0.1424 ± 0.0032	$3.1{\cdot}10^{-4} \pm 2.9{\cdot}10^{-6}$
$W^\pm W^\pm$-QCD	11.9 ± 0.6	10.4 ± 0.5	2.15 ± 0.11	24.6 ± 1.1	$0.0057 \pm 7{\cdot}10^{-4}$
$W^\pm W^\pm$-INT	1.75 ± 0.07	3.99 ± 0.13	1.918 ± 0.042	2.19 ± 0.06	$0.0016 \pm 5{\cdot}10^{-4}$
$W^\pm Z$-EW	3.34 ± 0.13	6.2 ± 0.19	5.52 ± 0.18	4.4 ± 0.12	139.2 ± 3.3
$W^\pm Z$-QCD	34 ± 5	20.6 ± 2.9	4.5 ± 0.9	91 ± 16	550 ± 50
Top, triboson	2.0 ± 0.13	2.0 ± 0.08	0.69 ± 0.05	6.12 ± 0.19	57.0 ± 1.5
ZZ	1.71 ± 0.06	0.617 ± 0.018	0.126 ± 0.008	3.76 ± 0.09	58.7 ± 0.6
$V\gamma$	7.6 ± 0.6	3.28 ± 0.41	0.61 ± 0.05	15.4 ± 1.8	35 ± 5
Charge-flip	6.8 ± 2.1	2.8 ± 1.2	1.02 ± 0.41	12 ± 5	0 ± 0
Non-prompt	31.6 ± 4.4	20.4 ± 3.0	10.6 ± 3.2	96 ± 12	0 ± 0
Total	128 ± 7	161 ± 10	177 ± 13	305 ± 16	840 ± 50
Data	142	158	175	306	831

Table G.8: Post-fit number of events in the regions used for the LL measurement in the pp-cmf. The signal region is split according to the output score of $\text{DNN}_{W^\pm W^\pm}$.

	$0 \leq \text{DNN}_{WW} < 0.15$	$0.15 \leq \text{DNN}_{WW} < 0.5$	$0.5 \leq \text{DNN}_{WW} \leq 1$	Low-m_{jj} CR	$W^\pm Z$ CR
$W^\pm_L W^\pm_L$-EW	0.6 ± 0.8	2.5 ± 3.4	8 ± 11	2.0 ± 2.8	$5{\cdot}10^{-5} \pm 7{\cdot}10^{-5}$
$W^\pm_T W^\pm_L$-EW	3.4 ± 0.8	20.1 ± 3.4	71 ± 8	17.7 ± 3.9	$7.6{\cdot}10^{-4} \pm 2.4{\cdot}10^{-4}$
$W^\pm_T W^\pm_T$-EW	3.98 ± 0.95	29 ± 5	127 ± 13	31 ± 7	$0.0026 \pm 4{\cdot}10^{-4}$
$W^\pm W^\pm$-EW Pol-INT	0.1875 ± 0.0097	0.578 ± 0.02	1.632 ± 0.038	$0.01296 \pm 2.9{\cdot}10^{-4}$	$2.5{\cdot}10^{-4} \pm 2.4{\cdot}10^{-6}$
$W^\pm W^\pm$-QCD	5.88 ± 0.34	12.2 ± 0.6	6.37 ± 0.29	24.7 ± 1.0	0.006 ± 0.0011
$W^\pm W^\pm$-INT	0.74 ± 0.05	2.87 ± 0.09	4.05 ± 0.09	2.2 ± 0.06	$0.0017 \pm 6{\cdot}10^{-4}$
$W^\pm Z$-EW	1.41 ± 0.08	4.88 ± 0.2	8.85 ± 0.22	4.42 ± 0.13	139.9 ± 3.3
$W^\pm Z$-QCD	21.5 ± 2.6	25.9 ± 3.3	12.7 ± 2.6	87 ± 10	544 ± 34
Top, triboson	1.02 ± 0.1	2.1 ± 0.12	1.59 ± 0.07	6.18 ± 0.23	57.3 ± 1.5
ZZ	1.3 ± 0.06	0.82 ± 0.024	0.387 ± 0.016	3.77 ± 0.09	58.7 ± 0.7
$V\gamma$	4.7 ± 0.7	5.6 ± 0.7	1.3 ± 0.09	15.9 ± 2.1	37 ± 6
Charge-flip	4.5 ± 1.4	4.2 ± 1.5	2.1 ± 0.9	12 ± 5	0 ± 0
Non-prompt	20.1 ± 2.5	23.4 ± 3.5	18.3 ± 4.2	96 ± 12	0 ± 0
Total	69.3 ± 4.3	134 ± 8	263 ± 16	303 ± 13	837 ± 31
Data	81	132	262	306	831

Table G.9: Post-fit number of events in the regions used for the LX measurement in the pp-cmf. The signal region is split according to the output score of $\mathrm{DNN}_{W^\pm W^\pm}$.

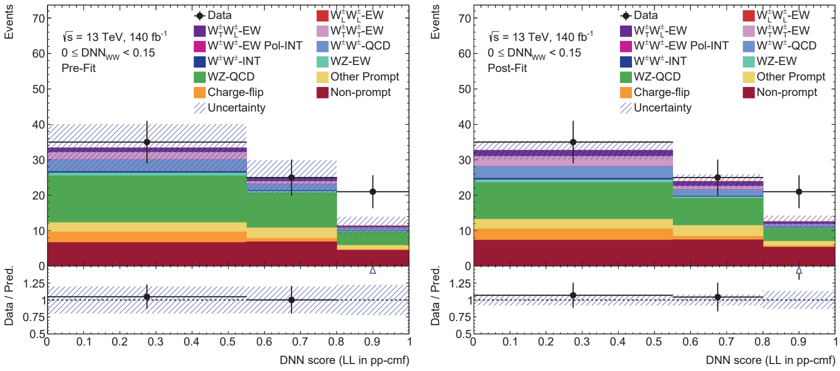
	$0 \leq \mathrm{DNN}_{WW} < 0.25$	$0.25 \leq \mathrm{DNN}_{WW} < 0.6$	$0.6 \leq \mathrm{DNN}_{WW} \leq 1$	Low-m_{jj} CR	$W^\pm Z$ CR
$W^\pm_L W^\pm_L$-EW	1.0 ± 0.5	2.5 ± 1.1	5.6 ± 2.4	1.6 ± 0.8	$3.8{\cdot}10^{-5} \pm 2{\cdot}10^{-5}$
$W^\pm_T W^\pm_L$-EW	4.6 ± 2.3	14 ± 6	34 ± 14	10 ± 5	$4.2{\cdot}10^{-4} \pm 2.2{\cdot}10^{-4}$
$W^\pm_T W^\pm_T$-EW	12.7 ± 3.3	46 ± 9	134 ± 20	37 ± 9	$0.0031 \pm 6{\cdot}10^{-4}$
$W^\pm W^\pm$-EW Pol-INT	0.399 ± 0.015	1.273 ± 0.033	2.78 ± 0.05	$0.01295 \pm 2.9{\cdot}10^{-4}$	$2.5{\cdot}10^{-4} \pm 2.4{\cdot}10^{-6}$
$W^\pm W^\pm$-QCD	10.1 ± 0.5	10.6 ± 0.5	3.99 ± 0.18	24.7 ± 1.0	$0.0057 \pm 8{\cdot}10^{-4}$
$W^\pm W^\pm$-INT	1.39 ± 0.06	3.32 ± 0.11	2.99 ± 0.06	2.2 ± 0.06	$0.0016 \pm 5{\cdot}10^{-4}$
$W^\pm Z$-EW	2.68 ± 0.13	5.14 ± 0.18	7.38 ± 0.19	4.42 ± 0.12	140.0 ± 3.1
$W^\pm Z$-QCD	31 ± 5	21.6 ± 3.7	8.2 ± 1.8	90 ± 17	550 ± 60
Top, triboson	1.75 ± 0.13	1.907 ± 0.096	1.08 ± 0.05	6.14 ± 0.22	57.3 ± 1.5
ZZ	1.62 ± 0.06	0.629 ± 0.02	0.252 ± 0.014	3.76 ± 0.09	58.7 ± 0.6
$V\gamma$	6.6 ± 0.8	4.13 ± 0.36	1.03 ± 0.14	15.8 ± 1.8	37 ± 5
Charge-flip	6.3 ± 2.0	3.0 ± 1.2	1.5 ± 0.7	12 ± 5	0 ± 0
Non-prompt	29.8 ± 4.1	19.3 ± 2.8	14.1 ± 3.6	98 ± 12	0 ± 0
Total	110 ± 6	134 ± 9	217 ± 15	305 ± 16	840 ± 60
Data	126	135	214	306	831



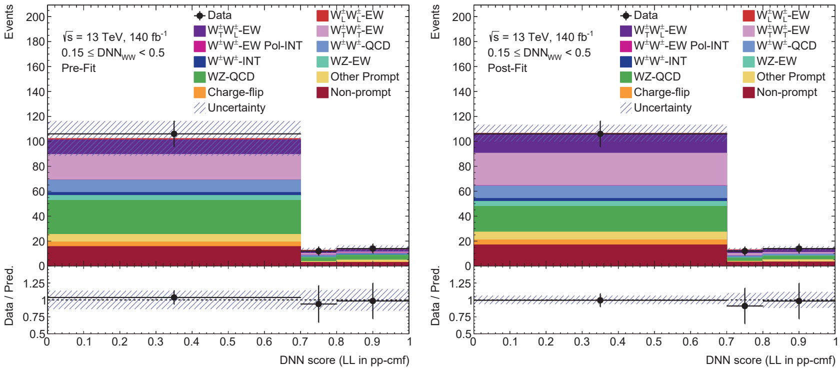
(a) $W^\pm Z$ control region

(b) Low-m_{jj} control region

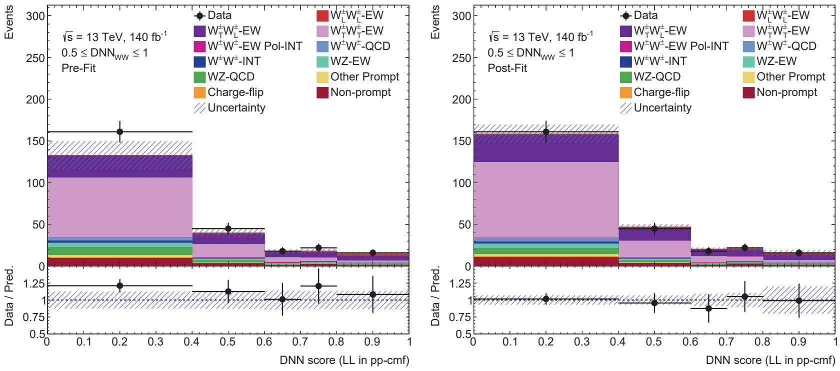
Figure G.5: The post-fit distribution for the LX in pp-cmf measurement in the $W^\pm Z$ control region and Low-m_{jj} control region.



(a) DNN$_{W^\pm W^\pm}$-score ≤ 0.15 pre-fit **(b)** DNN$_{W^\pm W^\pm}$-score ≤ 0.15 post-fit

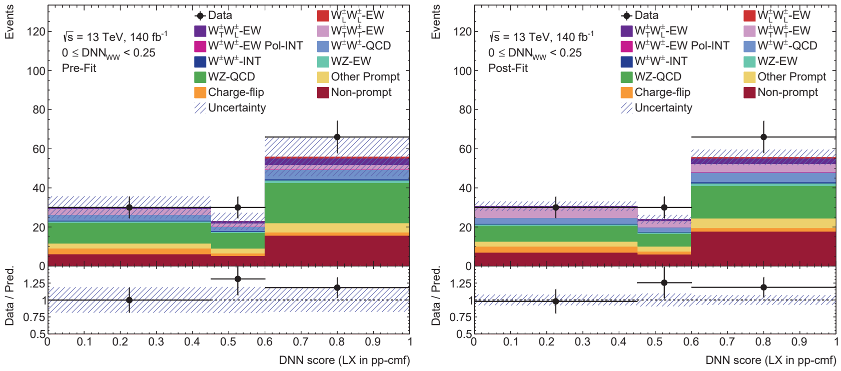


(c) $0.15 <$ DNN$_{W^\pm W^\pm}$-score ≤ 0.5 pre-fit **(d)** $0.15 <$ DNN$_{W^\pm W^\pm}$-score ≤ 0.5 post-fit

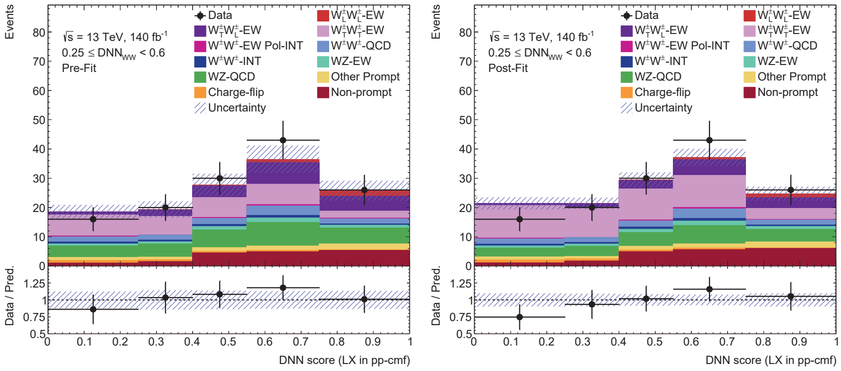


(e) $0.5 <$ DNN$_{W^\pm W^\pm}$-score ≤ 1 pre-fit **(f)** $0.5 <$ DNN$_{W^\pm W^\pm}$-score ≤ 1 post-fit

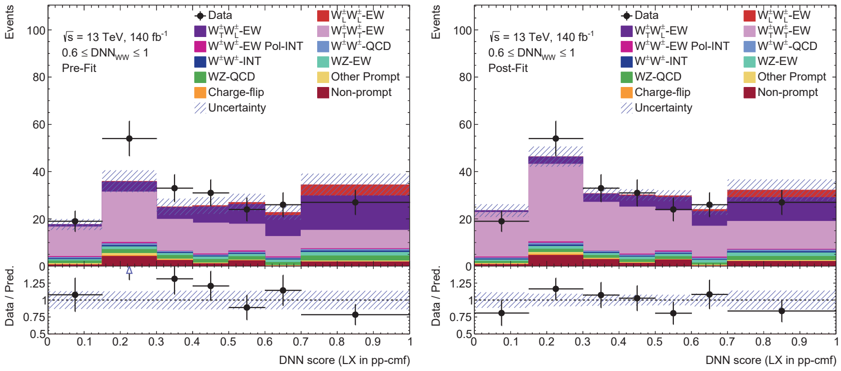
Figure G.6: The pre- and post-fit signal region distributions for the LL in pp-cmf measurement. The polarization states are simulated by SHERPA. As described in section 12.3 the signal region is split according to the DNN$_{W^\pm W^\pm}$ score. The shown histograms result from the application of the DNN$_{pol}$ trained for the LL polarization state in the pp-cmf.



(a) DNN$_{W^\pm W^\pm}$-score ≤ 0.25 pre-fit **(b)** DNN$_{W^\pm W^\pm}$-score ≤ 0.25 post-fit

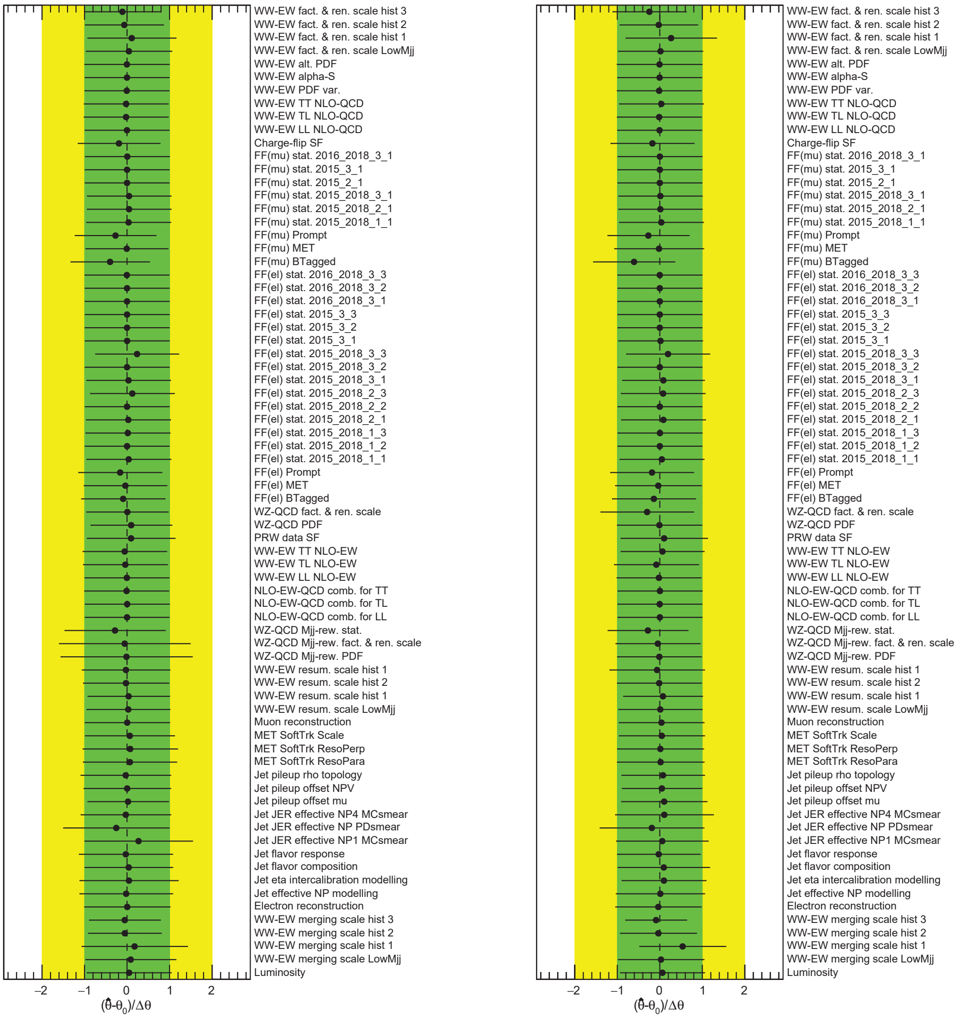


(c) $0.25 < $ DNN$_{W^\pm W^\pm}$-score ≤ 0.6 pre-fit **(d)** $0.25 < $ DNN$_{W^\pm W^\pm}$-score ≤ 0.6 post-fit



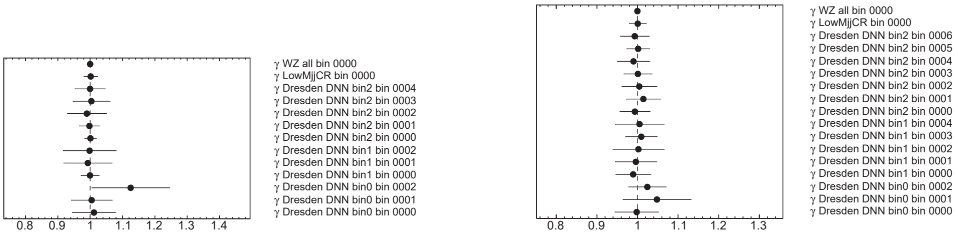
(e) $0.6 < $ DNN$_{W^\pm W^\pm}$-score ≤ 1 pre-fit **(f)** $0.6 < $ DNN$_{W^\pm W^\pm}$-score ≤ 1 post-fit

Figure G.7: The pre- and post-fit signal region distributions for the LX in pp-cmf measurement. The polarization states are simulated by SHERPA. As described in section 12.3 the signal region is split according to the DNN$_{W^\pm W^\pm}$ score. The shown histograms result from the application of the DNN$_{pol}$ trained for the LX polarization state in the pp-cmf.



(a) $\boldsymbol{\theta}$ of LL fit in pp-cmf

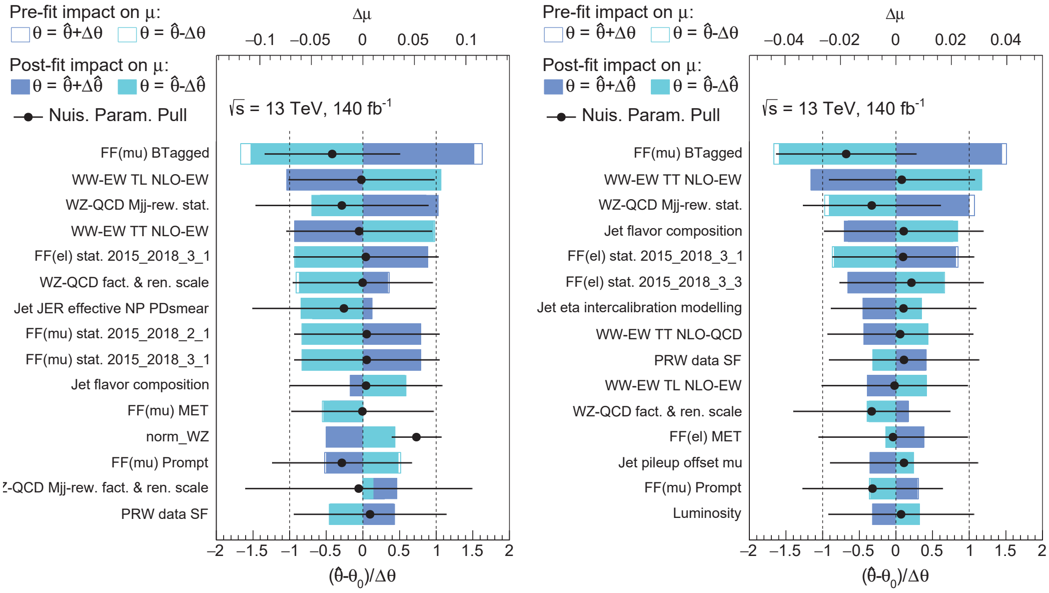
(b) $\boldsymbol{\theta}$ of LX fit in pp-cmf



(c) γ of LL fit in pp-cmf

(d) γ of LX fit in pp-cmf

Figure G.8: The pulls of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ for the partially unblinded background-only significance fits in the pp-cmf. The fit parameter μ_{LL}/μ_{LX} is fixed to 0. The fit is derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74.



(a) Ranking of μ_{LL} measurement in pp-cmf **(b)** Ranking of μ_{LX} measurement in pp-cmf

Figure G.9: The ranking plots for the measured fiducial cross-section fits in the pp-cmf.

List of Figures

2.1 Overview of the CERN accelerator complex during Run 2 taken from Ref. [15]. 10
2.2 The LHC's integrated luminosity of Run 1 and Run 2 periods. The figure is taken from Ref. [18] 12
2.3 The peak luminosity during Run 1 and Run 2 of the LHC taken from Ref. [18] 13
2.4 Schematic overview of the ATLAS detector and its main components. The image is taken from Ref. [20]. 13
2.5 The coordinate system of the ATLAS detector with respect to the LHC. The image is taken from Ref. [21]. 14

3.1 The potential $V(\phi) = \mu^2\phi\phi^* + \lambda\left(\phi\phi^*\right)^2$ of a complex scalar field ϕ breaking the U(1) symmetry. 29
3.2 Leading-order Feynman diagrams for the $W^{\pm}W^{\pm} \to W^{\pm}W^{\pm}$ scattering . . . 34
3.3 "Total $W^+W^+ \to W^+_L W^+_L$ scattering cross-sections in the SM as a function of the center-of-mass energy. Shown are the individual contributions of different initial polarization states to the final state consisting of purely longitudinal $W^+_L W^+_L$ pairs. Subscript X denotes any polarization (T or L). Assumed are two on-shell, unpolarized, colliding W^+ beams. A cut on the scattering angle that corresponds to pseudorapidity of ±1.5 with respect to the incoming W direction was applied. Results of MadGraph calculations." [43] 36
3.4 A selection of possible Feynman diagrams for $W^{\pm}W^{\pm}jj$-EW with only electroweak interactions. This selection includes fully-resonant t/u-channel-like diagrams in figures 3.4a to 3.4c, fully-resonant s-channel-like diagrams in figures 3.4d to 3.4g, single-resonant diagrams in figures 3.4h and 3.4i, and a non-resonant diagram in figure 3.4j. 37
3.5 A selection of possible Feynman diagrams for $W^{\pm}W^{\pm}jj$-QCD at the coupling order $\mathcal{O}(\alpha^4_{EW}\alpha^2_S)$. 37

4.1 Representation of a $t\bar{t}H$ event produced by the Sherpa event generator taken from Ref. [52]. The simulation includes the hard interaction (big red blob), the decay of the top quarks and the Higgs boson (small red blobs), photon radiations (yellow), the QCD radiations (red), parton hadronization (light green blobs), and the final decay of the hadrons (dark green). In addition, a secondary interaction (purple) is included in the represented simulation. . . . 42

5.1 Neural network with normalization layer, two hidden layers, and an output layer with one neuron. 48
5.2 The ReLU, Swish, and Sigmoid function as examples for the neurons activation function. 49
5.3 The BCE loss used for the binary classification with $y_i \in \{0, 1\}$ 50
5.4 Training and validation loss of a network trained to separate Higgs ($H \to ZZ \to l^+_0 l^-_0 l^+_1 l^-_1$) and background events with four leptons in the final state. The training data was provided by the ATLAS Open Data [70] 55
5.5 Three exemplary distributions of classifier scores separating two datasets. For each of the examples, the binary cross-entropy and the ROC-AUC are calculated. The ROC curves are shown below the distributions. 58

5.6 The concept of the k-fold method with $k = 5$ folds and the split intro training, validation and test data. The final DNNs are applied to their corresponding test dataset. 59

5.7 The SHAP values of $\mathrm{DNN}_{W^\pm W^\pm}$ trained in chapter 9 are calculated for an exemplary event. The input features are listed with their value for the specific event and are ranked according to their impact on the specific network decision. The output value $E[f(X)]$ is the expectation value of the network output. 60

5.8 Visualization of the analysis strategy 65

7.1 Feynman diagram of an electron charge-flip 79

8.1 Comparison of the differential polarized cross-sections simulated by Sherpa $W^\pm W^\pm jj$ and MadGraph at particle-level in the phase space defined in table 8.2. The polarizations are defined in the WW-cmf for figures 8.1a, 8.1c and 8.1e and in the pp-cmf for figures 8.1b, 8.1d and 8.1f. 84

8.2 A s-channel-like Feynman diagram for $W^\pm W^\pm jj$-EW at NLO-QCD. 85

8.3 Comparison of the differential cross-sections simulated by Sherpa $W^\pm W^\pm jj$ vs. $W^\pm W^\pm jj$ (t-ch.) in figure 8.3a and $W^\pm W^\pm jj$ (t-ch.) vs. $W^\pm W^\pm jj + 0,1j$ (t-ch.) in figure 8.3b on particle-level in the phase space defined in table 8.4 86

8.4 Comparison of the differential cross-sections simulated by Sherpa $W^\pm W^\pm jj$ and $W^\pm W^\pm jj + 0,1j$ (t-ch.) on particle-level in the phase space defined in table 8.4 87

8.5 The Kolmogorov Smirnov values for Sherpa $W^\pm W^\pm jj$ vs $W^\pm W^\pm jj$ (t-ch.), $W^\pm W^\pm jj+0,1j$ (t-ch.) vs $l^\pm l^\pm \nu\nu jj+0,1j$ (t-ch.), and $W^\pm W^\pm jj+0,1j$ (t-ch.) and $l^\pm l^\pm \nu\nu jj + 0,1j$ on particle-level in the phase space defined in table 8.4. The polarization is defined in the WW-cmf. 88

8.6 Comparison of the differential cross-sections simulated by Sherpa $W^\pm W^\pm jj+ 0,1j$ (t-ch.) and $l^\pm l^\pm \nu\nu jj + 0,1j$ on particle-level in the phase space defined in table 8.4. 89

8.7 The difference between $W^\pm W^\pm jj$ (t-ch.) and $l^\pm \nu l^\pm \nu jj+0,1j$ in the Δy_{jj} and $\Delta\phi_{jj}$ distribution in the kinematic region defined in table 8.4 at particle-level. These simulations differences are sampled by 100000 tuples of Δy_{jj} and $\Delta\phi_{jj}$. 90

8.8 The difference between $W^\pm W^\pm jj$ (t-ch.) and $l^\pm \nu l^\pm \nu jj+0,1j$ in the Δy_{jj} and $\Delta\phi_{jj}$ distributions is sampled by 100000 tuples of Δy_{jj} and $\Delta\phi_{jj}$. Figure 8.8a shows the resulting two-dimensional distribution in Δy_{jj} and ΔR_{jj}. In figure 8.8b, the sampled ΔR_{jj} distribution is compared to the difference between $W^\pm W^\pm jj$ (t-ch.) and $l^\pm \nu l^\pm \nu jj + 0,1j$. 90

8.9 The classification output of $\mathrm{DNN}_{0,1j}^{\mathrm{LL\,WW\text{-}cmf}}$ and $\mathrm{DNN}_{\text{s-ch.}}$ evaluated on their training and validation data. 93

8.10 Comparison of the differential cross-sections simulated by Sherpa $W^\pm W^\pm jj+ 0,1j$ (t-ch.) and $l^\pm l^\pm \nu\nu jj + 0,1j$ in $min(\Delta R_{l1/l2,j1/j2})$ on particle-level. The Sherpa $W^\pm W^\pm jj + 0,1j$ (t-ch.) is corrected by the $\mathrm{DNN}_{\text{s-ch.}}$. The networks are applied in the k-fold method to the corresponding validation data used during their training to avoid a bias toward the training data. The comparison is performed for the two event selections defined in table 8.4 and table 8.2. . . 94

8.11 Comparison of the differential cross-sections simulated by Sherpa $W^\pm W^\pm jj$, $W^\pm W^\pm jj$ (t-ch.), and $l^\pm l^\pm \nu\nu jj + 0,1j$ on particle-level in the loose region defined in table 8.4 before (dots) and after (crosses) the mutlivariate correction. In figures 8.11a and 8.11b $W^\pm W^\pm jj$ is corrected by the $\mathrm{DNN}_{0,1j}^{pol}$ and in figure 8.11c $W^\pm W^\pm jj$ (t-ch.) is corrected by $\mathrm{DNN}_{\text{s-ch.}}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training. 95

8.12 Comparison of $W^{\pm}W^{\pm}jj$ and $W^{\pm}W^{\pm}jj$ (t-ch.) with polarization defined in the WW-cmf simulated by SHERPA on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The SHERPA $W^{\pm}W^{\pm}jj$ is corrected by the $\mathrm{DNN}^{pol}_{0,1j}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training. The Kolmogorov-Smirnov values give an overview of the shape improvements in the tested variables. 96

8.13 Comparison of the differential cross-sections simulated by SHERPA $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) and $l^{\pm}l^{\pm}\nu\nu jj+0,1j$ on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The SHERPA $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) is corrected by the $\mathrm{DNN}_{\text{s-ch.}}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training. The Kolmogorov-Smirnov values give an overview of the shape improvements in the tested variables. 97

8.14 Comparison of the differential polarized cross-sections simulated by SHERPA $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) and MADGRAPH at particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The SHERPA sample is corrected by $\mathrm{corr}^{\text{multi-var}}_{\text{s-ch.}}$ and the event weights of the MADGRAPH sample are multiplied with $\mathrm{corr}^{\text{multi-var}}_{\text{jet-merged}}$. The polarizations are defined in the WW-cmf for figures 8.14a and 8.14c and in the pp-cmf for figures 8.14b and 8.14d. . . . 98

8.15 The missing s-channel correction in equation (8.2) derived in m_{jj}, Δy_{jj}, and ΔR_{jj}. The corrections are extracted in the combined signal and Low-m_{jj} region defined in table 8.2. 99

8.16 The Kolmogorov-Smirnov values for the comparison of $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) and $l^{\pm}l^{\pm}\nu\nu jj+0,1j$ in the combined signal and Low-m_{jj} region defined in table 8.2. The $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) distribution is corrected by a correction factor in m_{jj} (figure 8.16a), Δy_{jj} (figure 8.16b), and ΔR_{jj} (figure 8.16c). The correction factors are shown in figure 8.15. 101

8.17 Comparison of different fit functions for the higher-order QCD correction factor $\mathrm{corr}_{\text{jet-merged}}(m_{jj})$ for the LL polarization state in WW-cmf and the missing s-channel correction $\mathrm{corr}_{\text{s-ch.}}(m_{jj})$. The fit functions are defined in equations (8.6) to (8.8). 102

8.18 The renormalization and factorization scale variations for the SHERPA $W^{\pm}W^{\pm}jj$ and $l^{\pm}l^{\pm}\nu\nu jj+0,1j$ simulation on particle-level. The events are selected by the combined signal and Low-m_{jj} region as defined in table 8.2. . 103

8.19 LO and NLO-EW differential cross-sections in m_{jj} and p_T^{j1} provided by the authors of Ref. [50]. The calculation is done in the phase space listed in table 8.11 for polarization in the $W^{\pm}W^{\pm}$-cmf (solid line) and partonic center-of-mass frame (dotted line). 104

8.20 The ratio of the NLO-EW and LO differential cross-sections in the phase space listed in table 8.11 provided by the authors of Ref. [50]. An analytic fit of equation (8.11) for m_{jj} and equation (8.12) for p_T^{j1} is performed for each of the resulting NLO-EW corrections in the $W^{\pm}W^{\pm}$-cmf and pp-cmf. 106

9.1 Optimization of input variables for the DNN trained to separate LL and TX in WW-cmf with shuffling. The mean relative AUC change is plotted after shuffling a variable of the variable set in table 9.2 10 times on each fold. The variable with the lowest performance decrease is dropped, and in the next step, the remaining variables are tested. This figure represents the first iteration of the variable optimization for DNN_{pol} trained for LL in WW-cmf. 113

9.2 Optimization of input variables for the DNN_{pol} trained for LL in WW-cmf. Plotted is the relative AUC change of the variable whose removal results in the best mean AUC change after shuffling/retraining. If a significant performance decrease has been seen for five iterations, the procedure is stopped and the variable set that is still consistent with the best-performing set is chosen (marked with x). 116

9.3 The AUC during the training process of the polarization discriminator networks selected by the final hyperparameter optimization. The epoch with the highest AUC in the validation AUC fit is marked with a red cross. 118
9.4 The output distribution of the polarization DNNs trained on fold 0 and applied on the corresponding training and validation data. 119
9.5 The AUC during the training process of $\mathrm{DNN}_{W^\pm W^\pm}$ selected by the final hyperparameter optimization. The epoch with the highest AUC in the validation AUC fit is marked with a red cross. 120
9.6 The output distribution of $\mathrm{DNN}_{W^\pm W^\pm}$ selected by the fit to the validation AUC in figure 9.6a and selected by the best validation AUC value in figure 9.6b. The networks are trained on fold 0 and applied to the corresponding training and validation data. 121

10.1 Deviation in the classification of LL polarization in WW-cmf events in the signal region for $\mathrm{DNN}_{W^\pm W^\pm}$ trained on fold 0 and fold 1. In figure 10.1a the correlation and in figure 10.1b the deviation of the classification is shown. . . 124
10.2 The deviation in the classification of different processes in the signal region for the polarization neural networks trained on different folds. The resulting mean deviation is shown together with the standard deviation of the classification differences. 125
10.3 The deviation in the classification for the polarization neural networks trained on different folds. The resulting 50%-, 75%-, and 90%-quantiles of the absolute classification differences for different processes in the signal region is shown. . 126
10.4 The deviation in the classification for the neural networks trained on different folds to separate $W^\pm W^\pm jj$-EW signal and background. In figures 10.4a and 10.4c the mean deviation is shown together with the standard deviation of the classification differences. The corresponding quantiles are given in figures 10.4b and 10.4d. The events in figures 10.4a and 10.4b pass the signal region selection and the events in figures 10.4c and 10.4d originate from the Low-m_{jj} control region. 128
10.5 DNNs trained on fold 0 applied in the Low-m_{jj} control region. According to the fit results from the unpolarized $W^\pm W^\pm jj$ measurement [8], the $W^\pm Z$-QCD contribution is scaled by 0.69 and the $W^\pm W^\pm jj$-EW signal is scaled by 1.14. 130
10.6 The DNNs trained on different folds to separate $W^\pm W^\pm jj$-EW and background are applied in the signal region. Each DNN is only applied on its corresponding fold and is not used during training and optimization. The resulting distributions are stacked. According to the fit results from the unpolarized $W^\pm W^\pm jj$ measurement [8], the $W^\pm Z$-QCD contribution is scaled by 0.69 and the $W^\pm W^\pm jj$-EW signal is scaled by 1.14. 132
10.7 The mean absolute SHAP values of $\mathrm{DNN}_{W^\pm W^\pm}$ for different input features. The assigned uncertainties originate from the statistical uncertainty of the explanation dataset. The SHAP evaluation is done for independent mask datasets of different size. 133
10.8 The SHAP values of $\mathrm{DNN}_{W^\pm W^\pm}$ originating from mask datasets with increasing size are compared by calculating the event-wise difference. This is done for each time N_{mask} was increased by 20. For better visibility the individual differences are combined into a box plot of the quantiles and whiskers covering the full range. 134
10.9 The SHAP values are evaluated for $\mathrm{DNN}_{W^\pm W^\pm}$. In figure 10.9a the individual SHAP values are plotted with a color encoding the value of the corresponding feature. In addition, it is also listed how much the mean values of the features in $W^\pm W^\pm jj$-EW signal $\bar{x}_1$ and background $\bar{x}_0$ differ. Figure 10.9b shows the mean absolute impact of each feature split into the 5 major contributions in the training data. 136

10.10 The SHAP values of the polarization DNNs trained in WW-cmf are plotted with a color encoding the value of the corresponding feature. In addition, it is also listed how much the mean values of the features in LL/LX-polarization $\bar{x}_1$ and TX/TT-polarization $\bar{x}_0$ differ. 138

10.11 The SHAP values of the polarization DNNs trained in pp-cmf are plotted with a color encoding the value of the corresponding feature. In addition, it is also listed how much the mean values of the features in LL/LX-polarization $\bar{x}_1$ and TX/TT-polarization $\bar{x}_0$ differ. 139

10.12 The mean absolute SHAP values of the polarization DNNs split into the LL, TL, and TT contributions in the corresponding reference frame. 140

11.1 Kolmogorov Smirnov test between the nominal Sherpa $l^{\pm}\nu l^{\pm}\nu jj + 0, 1j$ sample and the corresponding renormalization and factorization scale variations. The kinematic distributions are derived on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. 144

11.2 Kolmogorov Smirnov values for the comparison of the nominal Sherpa $l^{\pm}\nu l^{\pm}\nu jj + 0, 1j$ sample and the corresponding resummation and merging scale variations. The kinematic distributions are derived on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. 145

11.3 The combinations of the relative systematics $r^{\text{sys}}(m_{jj})$ in equation (11.4) for particle-level Sherpa $W^{\pm}W^{\pm}jj + 0, 1j$. The analytic functions given by equations (11.5) and (11.6) are fit to the combined relative uncertainties. The kinematic distributions are derived on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2 but without the m_{jj} selection criteria. 147

11.4 The merging scale variations compared to the nominal $l^{\pm}\nu l^{\pm}\nu jj + 0, 1j$ distribution in Δy_{jj} on particle-level. The events are selected by the loose particle-level region defined in table 8.4. 148

12.1 The $\text{DNN}_{W^{\pm}W^{\pm}}$ output is combined with the output of the polarization DNN_{LX} to create a two-dimensional output space. The colors indicate the dominant proportion of either background contributions (blue), the TT polarization state (green) or the LL polarization state (green). 153

12.2 Visualization of the rebinning strategy. The two-dimensional histogram is split along the $\text{DNN}_{W^{\pm}W^{\pm}}$ output into several histograms. Each resulting histogram has specific optimized binning. 156

12.3 Comparison of Sherpa $W^{\pm}W^{\pm}jj$ and $l^{\pm}l^{\pm}\nu\nu jj + 0, 1j$ at particle-level in the phase space defined in table 8.2. 158

12.4 Pearson correlation between the DNN scores and the individual event kinematics. 160

12.5 The distribution of the DNN score separating $W^{\pm}W^{\pm}jj$-EW and background and the systematic envelope of the renormalization and factorization scale systematic. The variation with the biggest shape effect s_{shape} defined in equation (12.15) originates from the shown pulls of c_i in the m_{jj} distribution. . . 161

12.6 The maximal s_{shape} value defined in equation (12.15) for the different DNNs. The s_{shape} value is maximized within the envelope of different scale uncertainties and jet kinematics. 161

12.7 The distribution of the polarization DNN scores and the systematic envelope of the renormalization and factorization scale systematic. The variation with the biggest shape effect s_{shape} defined in equation (12.15) originates from the shown pulls of c_i in the p_T^{j1} distribution. 162

12.8 The distribution of the polarization DNN scores and the systematic envelope of the renormalization and factorization scale systematic. The variation with the biggest shape effect s_{shape} defined in equation (12.15) originates from the shown pulls of c_i in the $\Delta\phi_{jj}$ distribution. 163

12.9 The pre-fit distribution in the $W^{\pm}Z$ control region and Low-m_{jj} control region. 166

12.10 The pre-fit signal region distributions for the LL in WW-cmf measurement. The polarization states are simulated by SHERPA. As described in section 12.3 the signal region is split according to the DNN$_{W^\pm W^\pm}$-score. The shown histograms result from the application of the DNN$_{pol}$ trained for the LL polarization state in the WW-cmf. 167
12.11 The μ_{LL}/μ_{LX}, μ_{TX}/μ_{TT}, and $\mu_{\text{WZ-QCD}}$ normalization resulting from the partially unblinded significance fits. In figure 12.11a, the normalization of the longitudinal polarization is free-floating while in figure 12.11b it is fixed to 0. The fit is derived for measured data in the $W^\pm Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^\pm Z$-QCD contribution in the Asimov data is scaled by 0.74. 168
12.12 The ranking plots for the partially unblinded significance fits in the WW-cmf. The fit is derived for measured data in the $W^\pm Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^\pm Z$-QCD contribution in the Asimov data is scaled by 0.74. 168
12.13 The pulls of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ for the partially unblinded background-only significance fits in the WW-cmf. The fit parameter μ_{LL}/μ_{LX} is fixed to 0. The fit is derived for measured data in the $W^\pm Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^\pm Z$-QCD contribution in the Asimov data is scaled by 0.74. 169
12.14 The shape of the $W^\pm W^\pm jj$-QCD, $W^\pm Z$-QCD, charge-flip, and non-prompt background compared to the LL/LX signal of the polarization measurement in the WW-cmf. The shape differences are evaluated in the signal region histogram with the highest DNN$_{W^\pm W^\pm}$ score. 171
12.15 The systematic uncertainty of the NLO-EW correction on the TT polarization state and the statistical uncertainty of the m_{jj} reweighting of the $W^\pm Z$-QCD background. The variations are shown in the signal region histogram with $0.6 <$ DNN$_{W^\pm W^\pm}$-score ≤ 1 for the LL measurement and $0.7 <$ DNN$_{W^\pm W^\pm}$-score ≤ 1 for the LX measurement. 172
12.16 The ranking plots of the partially unblinded fiducial cross-section fits in the WW-cmf. The fit is derived for measured data in the $W^\pm Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^\pm Z$-QCD contribution in the Asimov data is scaled by 0.74. 175
12.17 The post-fit distribution for the LL in WW-cmf measurement in the $W^\pm Z$ control region and Low-m_{jj} control region. 176
12.18 The μ_{LL}/μ_{LX}, μ_{TX}/μ_{TT}, and $\mu_{\text{WZ-QCD}}$ normalization resulting from the unblinded significance fits. 177
12.19 The pre- and post-fit signal region distributions for the LL in WW-cmf measurement. The polarization states are simulated by SHERPA. As described in section 12.3 the signal region is split according to the DNN$_{W^\pm W^\pm}$ score. The shown histograms result from the application of the DNN$_{pol}$ trained for the LL polarization state in the WW-cmf. 178
12.20 Pre-fit comparison of the predicted and measured distribution in $m_{o1}^{ll,MET}$ in the signal and Low-m_{jj} control region. 179
12.21 The correlation of the polarization DNN scores evaluated for $W^\pm W^\pm jj$-EW events in the signal region. 180
12.22 The pre- and post-fit signal region distributions for the LX in WW-cmf measurement. The polarization states are simulated by SHERPA. As described in section 12.3 the signal region is split according to the DNN$_{W^\pm W^\pm}$ score. The shown histograms result from the application of the DNN$_{pol}$ trained for the LX polarization state in the WW-cmf. 181
12.23 The pulls of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ for the partially unblinded background-only significance fits in the WW-cmf. The fit parameter μ_{LL}/μ_{LX} is fixed to 0. The fit is derived for measured data in the $W^\pm Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^\pm Z$-QCD contribution in the Asimov data is scaled by 0.74. 182
12.24 The ranking plots for the measured fiducial cross-section fits in the WW-cmf. 183

12.25 Timeline of the LHC and the plan for the HL-LHC project presented in Ref. [114] 186

12.26 The projection of the expected significances and signal strength uncertainties of section 12.5.1 for LHC runs with higher luminosity. In figure 12.26b, the contribution of the systematic uncertainties is shown in the lower subplot. . . 187

B.1 Comparison of Sherpa $W^{\pm}W^{\pm}jj$ and $W^{\pm}W^{\pm}jj$ (t-ch.) on particle-level in the phase space defined in table 8.4. The comparison is done for polarization in the WW-cmf and pp-cmf. 198

B.2 The training process of $\text{DNN}^{pol}_{0,1j}$ used for the approximate NLO-QCD correction. The DNNs are selected from the hyperparameter optimization by the minimum of the fit to the validation loss. 199

B.3 The training process of $\text{DNN}_{\text{s-ch.}}$ used for the approximate NLO-QCD correction. The DNNs are selected from the hyperparameter optimization by the minimum of the fit to the validation loss. 200

B.4 The classification output of the $\text{DNN}^{poll}_{0,1j}$ used for the multivariate higher-order QCD correction. 201

B.5 Comparison of the polarization in WW-cmf simulated by Sherpa $W^{\pm}W^{\pm}jj$ and $W^{\pm}W^{\pm}jj$ (t-ch.) on particle-level in the phase space defined in table 8.4. The Sherpa $W^{\pm}W^{\pm}jj$ is corrected by the $\text{DNN}^{poll}_{0,1j}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training. 202

B.6 Comparison of the polarization in pp-cmf simulated by Sherpa $W^{\pm}W^{\pm}jj$ and $W^{\pm}W^{\pm}jj$ (t-ch.) on particle-level in the phase space defined in table 8.4. The Sherpa $W^{\pm}W^{\pm}jj$ is corrected by the $\text{DNN}^{poll}_{0,1j}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training. 203

B.7 Comparison of Sherpa $W^{\pm}W^{\pm}jj + 0,1j$ (t-ch.) and $l^{\pm}l^{\pm}\nu\nu jj + 0,1j$ on particle-level in the phase space defined in table 8.4. The Sherpa $W^{\pm}W^{\pm}jj + 0,1j$ (t-ch.) is corrected by the $\text{DNN}_{\text{s-ch.}}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training. 204

B.8 Comparison of the polarization in pp-cmf simulated by Sherpa $W^{\pm}W^{\pm}jj$ and $W^{\pm}W^{\pm}jj$ (t-ch.) on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The Sherpa $W^{\pm}W^{\pm}jj$ is corrected by the $\text{DNN}^{poll}_{0,1j}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training. The Kolmogorov-Smirnov values give an overview of the shape improvements in the tested variables. 205

B.9 Comparison of the polarization in WW-cmf simulated by Sherpa $W^{\pm}W^{\pm}jj$ and $W^{\pm}W^{\pm}jj$ (t-ch.) in lepton and p_T^{miss} kinematic on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The Sherpa $W^{\pm}W^{\pm}jj$ is corrected by the $\text{DNN}^{poll}_{0,1j}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training. 206

B.10 Comparison of the polarization in WW-cmf simulated by Sherpa $W^{\pm}W^{\pm}jj$ and $W^{\pm}W^{\pm}jj$ (t-ch.) in jet kinematic on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The Sherpa $W^{\pm}W^{\pm}jj$ is corrected by the $\text{DNN}^{poll}_{0,1j}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training. 207

B.11 Comparison of the polarization in pp-cmf simulated by Sherpa $W^{\pm}W^{\pm}jj$ and $W^{\pm}W^{\pm}jj$ (t-ch.) in lepton and p_T^{miss} kinematic on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The Sherpa $W^{\pm}W^{\pm}jj$ is corrected by the $\text{DNN}^{poll}_{0,1j}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training. 208

B.12 Comparison of the polarization in pp-cmf simulated by SHERPA $W^{\pm}W^{\pm}jj$ and $W^{\pm}W^{\pm}jj$ (t-ch.) in jet kinematic on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The SHERPA $W^{\pm}W^{\pm}jj$ is corrected by the $\text{DNN}^{poll}_{0,1j}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training. 209
B.13 Comparison of SHERPA $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) and $l^{\pm}l^{\pm}\nu\nu jj+0,1j$ in lepton and p_T^{miss} kinematic on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The SHERPA $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) is corrected by the $\text{DNN}_{\text{s-ch.}}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training. 210
B.14 Comparison of SHERPA $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) and $l^{\pm}l^{\pm}\nu\nu jj+0,1j$ jet kinematic on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. The SHERPA $W^{\pm}W^{\pm}jj+0,1j$ (t-ch.) is corrected by the $\text{DNN}_{\text{s-ch.}}$. To avoid a bias the networks are applied in the k-fold method to the corresponding validation data used during their training. 211
B.15 Comparison of different fit functions for the higher-order QCD correction factor $\text{corr}_{\text{jet-merged}}$ in m_{jj}. The correction is extracted from the combined signal and Low-m_{jj} region at particle-level. 212

C.1 Distributions of the training input in the signal region. 214
C.2 Distributions of the training input in the signal region. 215
C.3 Distributions of the training input in the signal region. 216
C.4 Distributions of the training input in the signal region. 217
C.5 Distributions of the training input in the signal region. 218
C.6 Distributions of the training input in the signal region. 219
C.7 Distributions of the training input in the signal region. 220
C.8 Distributions of the training input in the signal region. 221
C.9 Optimization of input variables for $\text{DNN}_{W^{\pm}W^{\pm}}$ and DNN_{pol} for LX in WW-cmf. Plotted is the relative AUC change of the variable whose removal results in the best mean AUC change after shuffling/retraining. If a significant performance decrease has been seen for five iterations, the procedure is stopped and the variable set that is still consistent with the best-performing set is chosen (marked with x). 222
C.10 Optimization of input variables for DNN_{pol} for LL and LX in pp-cmf. Plotted is the relative AUC change of the variable whose removal results in the best mean AUC change after shuffling/retraining. If a significant performance decrease has been seen for five iterations, the procedure is stopped and the variable set that is still consistent with the best-performing set is chosen (marked with x). 223

D.1 DNNs trained on different folds to seperate LL and TX in WW-cmf are applied in the Low-m_{jj} control region. According to the fit results from the unpolarized $W^{\pm}W^{\pm}jj$ measurement [8] the $W^{\pm}Z$-QCD contribution is scaled by 0.69 and the $W^{\pm}W^{\pm}jj$-EW signal is scaled by 1.14. 226
D.2 DNNs trained on different folds to seperate LX and TT in WW-cmf are applied in the Low-m_{jj} control region. According to the fit results from the unpolarized $W^{\pm}W^{\pm}jj$ measurement [8] the $W^{\pm}Z$-QCD contribution is scaled by 0.69 and the $W^{\pm}W^{\pm}jj$-EW signal is scaled by 1.14. 227
D.3 DNNs trained on different folds to seperate LL and TX in pp-cmf are applied in the Low-m_{jj} control region. According to the fit results from the unpolarized $W^{\pm}W^{\pm}jj$ measurement [8] the $W^{\pm}Z$-QCD contribution is scaled by 0.69 and the $W^{\pm}W^{\pm}jj$-EW signal is scaled by 1.14. 228
D.4 DNNs trained on different folds to seperate LX and TT in pp-cmf are applied in the Low-m_{jj} control region. According to the fit results from the unpolarized $W^{\pm}W^{\pm}jj$ measurement [8] the $W^{\pm}Z$-QCD contribution is scaled by 0.69 and the $W^{\pm}W^{\pm}jj$-EW signal is scaled by 1.14. 229

D.5 DNNs trained on different folds to seperate $W^{\pm}W^{\pm}jj$-EW and background are applied in the Low-m_{jj} control region. According to the fit results from the unpolarized $W^{\pm}W^{\pm}jj$ measurement [8] the $W^{\pm}Z$-QCD contribution is scaled by 0.69 and the $W^{\pm}W^{\pm}jj$-EW signal is scaled by 1.14. 230

E.1 The mean absolute SHAP values of DNN$_{pol}$ for different input features. The assigned uncertainties originate from the statistical uncertainty of the explanation-dataset. The SHAP evaluation is done for independent mask-datasets of different size. 232
E.2 The SHAP values of DNN$_{pol}$ originating from mask-datasets with increasing size are compared by calculate the event-wise difference. This is done for each time N_{mask} was increased by 20. For better visibility the individual differences are combined into a box plot of the quantiles and whiskers covering the full range. 233

F.1 Kolmogorov Smirnov test between the nominal Sherpa $l^{\pm}\nu l^{\pm}\nu jj+0,1j$ sample and the corresponding α_S, PDF set, and internal PDF variations. The kinematic distributions are derived on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. 236
F.2 Kolmogorov Smirnov p-values for the comparison of the nominal Sherpa $l^{\pm}\nu l^{\pm}\nu jj+0,1j$ sample and the corresponding resummation and merging scale variations. The kinematic distributions are derived on particle-level in the combined signal and Low-m_{jj} region defined in table 8.2. 237

G.1 The ranking plots for the partially unblinded significance fits in the pp-cmf. The fit is derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74. 240
G.2 The pulls of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ for the partially unblinded background-only significance fits in the pp-cmf. The fit parameter μ_{LL}/μ_{LX} is fixed to 0. The fit is derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74. 241
G.3 The post-fit distribution for the LX in WW-cmf measurement in the $W^{\pm}Z$ control region and Low-m_{jj} control region. 242
G.4 The post-fit distribution for the LL in pp-cmf measurement in the $W^{\pm}Z$ control region and Low-m_{jj} control region. 243
G.5 The post-fit distribution for the LX in pp-cmf measurement in the $W^{\pm}Z$ control region and Low-m_{jj} control region. 245
G.6 The pre- and post-fit signal region distributions for the LL in pp-cmf measurement. The polarization states are simulated by Sherpa. As described in section 12.3 the signal region is split according to the DNN$_{W^{\pm}W^{\pm}}$ score. The shown histograms result from the application of the DNN$_{pol}$ trained for the LL polarization state in the pp-cmf. 246
G.7 The pre- and post-fit signal region distributions for the LX in pp-cmf measurement. The polarization states are simulated by Sherpa. As described in section 12.3 the signal region is split according to the DNN$_{W^{\pm}W^{\pm}}$ score. The shown histograms result from the application of the DNN$_{pol}$ trained for the LX polarization state in the pp-cmf. 247
G.8 The pulls of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ for the partially unblinded background-only significance fits in the pp-cmf. The fit parameter μ_{LL}/μ_{LX} is fixed to 0. The fit is derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74. 248
G.9 The ranking plots for the measured fiducial cross-section fits in the pp-cmf. . 249

List of Tables

3.1 The Standard Model Particles with their corresponding masses, spin, and charges taken from Ref. [41] 32

5.1 Algorithm of the Adam optimizer [66] 52
5.2 Parameters of the Adam optimization and their recommended defaults [66] . 52

6.1 Electron and muon trigger used during Run 2 of the ATLAS detector [76][77]. 67
6.2 The selection criteria for baseline and signal electrons as described in Ref. [8]. 69
6.3 The selection criteria for baseline and signal muons as described in Ref. [8]. . 69
6.4 The selection criteria for baseline and signal jets as described in Ref. [8]. . . . 70
6.5 The event selection defining the signal region used in Ref. [8]. 71
6.6 The event selection defining the $W^{\pm}Z$ control region used in Ref. [8]. 72
6.7 The particle-level event selection of the fiducial signal region. 73

8.1 Sherpa $W^{\pm}W^{\pm}jj$-EW samples to study the approximate NLO-QCD+PS effects. 82
8.2 Particle-level selection corresponding to the SR and Low-m_{jj} region selection. 82
8.3 Comparison of the cross-sections, polarization fractions, and number of generated Monte Carlo events for different MadGraph and Sherpa samples. The shown events are selected by the particle-level phase space defined in table 8.2. 83
8.4 Loose particle selection used to derive the higher-order QCD correction. . . . 85
8.5 Comparison of the cross-sections, polarization fractions, and number of generated Monte Carlo events for different Sherpa samples. The shown events are selected by the loose particle-level event selection given in table 8.4. . . . 86
8.6 Variables used to optimize and train the DNNs for the higher-order QCD correction. The variables were scaled for the training and application. 92
8.7 DNN hyperparameters for the multivariate higher-order QCD correction. For the tunable hyperparameters, the search space and the sampling distribution are listed. 92
8.8 Hyperparameters chosen for the neural network training. The shown hyperparameters were chosen in the final step of the hyperparameter optimization. 93
8.9 Fit parameters of $f_1(m_{jj})$ defined in equation (8.6) for the total higher-order QCD correction in m_{jj}. 102
8.10 Fit parameters of $f_3(m_{jj})$ defined in equation (8.8) for the correction of missing s-channels at higher-order QCD. 102
8.11 Parton-level selection for the NLO-EW calculation provided by the authors of Ref. [50]. 104
8.12 Fit parameters of equation (8.11) for the NLO-EW correction in m_{jj}. 105
8.13 Fit parameters of equation (8.12) for the NLO-EW correction in p_T^{j1}. 105

9.1 Processes used for training of the DNN classifiers with number of generated events and the number of events they predict for the signal region. For $W^{\pm}Z$-QCD, an additional scale factor was applied to achieve a better agreement with the expected data. 109

9.2 Variables used to optimize and train the binary classifier. The variables were pre-scaled for the training and application to have a more Gaussian-like distribution. 110
9.3 DNN hyperparameters optimized for the discriminator DNNs. The search space and sampling distribution are listed for the tunable hyperparameters. . 112
9.4 Hyperparameters derived from the pre-optimization. The pre-optimization was done on the full set of variables in table 9.2. 114
9.5 Hyperparameters chosen for the neural network training. The shown hyperparameters were chosen in the final step of the hyperparameter optimization. 114
9.6 List of variables used for the training. For each neural network, the variables kept after the variable optimization for the training are marked with X. The index 1 or 2 for the regression variables in the lower part of the table relates to two different hyperparameter choices for the regression DNN models. . . . 115
9.7 The number of low- and high-level variables chosen for the discriminator networks. The detailed list of chosen variables is given in table 9.6. 116

10.1 The stat-only χ^2 values and the corresponding p-values for the DNN closure test with measured data in the Low-m_{jj} control region. 131

11.1 SHERPA $l^{\pm}\nu l^{\pm}\nu jj + 0, 1j$ samples with resummation or merging scale variation.143
11.2 The performance of the χ^2 fits of the functions in equations (11.5) and (11.6) to the combined systematic variations. 146
11.3 Fit parameters of $f_1^{sys}(x)$ in equation (11.5) for the combined relative uncertainties of the internal PDF, resummation scale, and the merging scale variations on particle-level SHERPA $l^{\pm}\nu l^{\pm}\nu jj + 0, 1j$. 146
11.4 Fit parameters of $f_2^{sys}(x)$ in equation (11.6) for the combined relative uncertainties of the alternative PDF sets, the α_S variations, and the renormalization and factorization scale variations on particle-level SHERPA $W^{\pm}W^{\pm}jj + 0, 1j$. 148

12.1 Relative normalization uncertainties for different processes. 155
12.2 Estimated significance (without consideration of shape uncertainties) for different number of splits into one-dimensional histograms. The optimization algorithm is described in section 12.3 157
12.3 The optimized splits in the $\text{DNN}_{W^{\pm}W^{\pm}}$ classification and the corresponding bin edges in polarization classification. 157
12.4 The significances of the longitudinal polarization states extracted by the partially unblinded fits. The fits are derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74. 166
12.5 The fiducial cross-sections of the longitudinal polarization states extracted by the partially unblinded fits. The fits are derived from measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74. . . 173
12.6 The expected normalization of the LL/LX polarization in the WW-cmf and pp-cmf. The uncertainties are split into the statistical uncertainty of the data, statistical modeling uncertainty, systematic modeling uncertainty, and luminosity uncertainty. The fit is derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74. 173
12.7 The expected fiducial cross-section of the LL/LX polarization in the WW-cmf and pp-cmf. The uncertainties are split into the statistical uncertainty of the data, statistical modeling uncertainty, systematic modeling uncertainty, and luminosity uncertainty. The fit is derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74. 173

12.8 The impact of uncertainty sources on the expected fiducial cross-section of the LL/LX polarization in the WW-cmf and pp-cmf. The fit is derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74. 174
12.9 The expected 95% CL upper limit of the LL polarization in the WW-cmf and pp-cmf. The fit is derived for measured data in the $W^{\pm}Z$ control region and Asimov data in the signal and Low-m_{jj} control region. The $W^{\pm}Z$-QCD contribution in the Asimov data is scaled by 0.74. 175
12.10 The observed significances of the longitudinal polarization states. The signficances in brackets result from the partially unblinded fit in table 12.4. 177
12.11 The observed normalization of the LL/LX polarization in the WW-cmf and pp-cmf. The uncertainties are split into the statistical uncertainty of the data, statistical modeling uncertainty, systematic modeling uncertainty, and luminosity uncertainty. 183
12.12 The observed fiducial cross-section of the LL/LX polarization in the WW-cmf and pp-cmf. The uncertainties are split into the statistical uncertainty of the data, statistical modeling uncertainty, systematic modeling uncertainty, and the luminosity uncertainty. 184
12.13 The measured 95% CL upper limit of the LL polarization in the WW-cmf and pp-cmf. 184
12.14 The observed (expected) significances derived with polarization modeling by Sherpa and MadGraph. 184
12.15 The observed (expected) cross-section derived by polarization modeling by Sherpa and MadGraph. 185
12.16 Comparison of the observed (expected) results of the longitudinally polarized $W^{\pm}W^{\pm}jj$-EW scattering measured at the ATLAS and CMS experiment. . . 185

A.1 Information about Monte Carlo generated samples used in this thesis which were not used in Ref. [8] 193
A.7 The DSID, physics name, and p-tag of the Monte Carlo generator samples used to model the Z+jets background. 193
A.2 The DSID, physics name, and p-tag of the Monte Carlo generator samples used to model the polarized $W^{\pm}W^{\pm}jj$-EW signal and the $W^{\pm}W^{\pm}jj$-INT and $W^{\pm}W^{\pm}jj$-QCD contributions. 195
A.3 The DSID, physics name, and p-tag of the Monte Carlo generator samples used to model the $W^{\pm}Z$ background. 195
A.4 The DSID, physics name, and p-tag of the Monte Carlo generator samples used to model the top and tri-boson backgrounds. 195
A.5 The DSID, physics name, and p-tag of the Monte Carlo generator samples used to model the $V\gamma$ backgrounds. 196
A.6 The DSID, physics name, and p-tag of the Monte Carlo generator samples used to model the ZZ background. 196

G.1 The pruned list of systematic variations and their corresponding category and sub-category. 239
G.2 Pre-fit number of events in the regions used for the LL measurement in the WW-cmf. The signal region is split according to the output score of $\mathrm{DNN}_{W^{\pm}W^{\pm}}$. 242
G.3 Pre-fit number of events in the regions used for the LX measurement in the WW-cmf. The signal region is split according to the output score of $\mathrm{DNN}_{W^{\pm}W^{\pm}}$. 242
G.4 Pre-fit number of events in the regions used for the LL measurement in the pp-cmf. The signal region is split according to the output score of $\mathrm{DNN}_{W^{\pm}W^{\pm}}$. 243
G.5 Pre-fit number of events in the regions used for the LX measurement in the pp-cmf. The signal region is split according to the output score of $\mathrm{DNN}_{W^{\pm}W^{\pm}}$. 243
G.6 Post-fit number of events in the regions used for the LL measurement in the WW-cmf. The signal region is split according to the output score of $\mathrm{DNN}_{W^{\pm}W^{\pm}}$. 244

G.7 Post-fit number of events in the regions used for the LX measurement in the WW-cmf. The signal region is split according to the output score of $\mathrm{DNN}_{W^{\pm}W^{\pm}}$. 244
G.8 Post-fit number of events in the regions used for the LL measurement in the pp-cmf. The signal region is split according to the output score of $\mathrm{DNN}_{W^{\pm}W^{\pm}}$. 244
G.9 Post-fit number of events in the regions used for the LX measurement in the pp-cmf. The signal region is split according to the output score of $\mathrm{DNN}_{W^{\pm}W^{\pm}}$. 245

Bibliography

[1] A. 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* $\sqrt{s} = 13\,\mathrm{TeV}$, Physical Review Letters **120** (2018), ISSN: 1079-7114, URL: `http://dx.doi.org/10.1103/PhysRevLett.120.081801` (cit. on pp. v, vii).

[2] M. Aaboud et al., *Observation of Electroweak Production of a Same-Sign W Boson Pair in Association with Two Jets in pp Collisions at* $\sqrt{s} = 13\,\mathrm{TeV}$ *with the ATLAS Detector*, Physical Review Letters **123** (2019), ISSN: 1079-7114, URL: `http://dx.doi.org/10.1103/PhysRevLett.123.161801` (cit. on pp. v, vii).

[3] Albert M Sirunyan et al., *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*, Phys. Lett. B **812** (2021) 136018, Submitted to Phys.Lett., arXiv: `2009.09429`, URL: `https://cds.cern.ch/record/2730856` (cit. on pp. v, vii, 4, 39, 153, 185, 190).

[4] Georges Aad et al., *Observation of gauge boson joint-polarisation states in* $W^{\pm}Z$ *production from pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Lett. B **843** (2023) 137895, arXiv: `2211.09435`, URL: `https://cds.cern.ch/record/2841241` (cit. on pp. 4, 39, 59).

[5] Georges Aad et al., *Studies of the Energy Dependence of Diboson Polarization Fractions and the Radiation-Amplitude-Zero Effect in* WZ *Production with the ATLAS Detector*, Phys. Rev. Lett. **133** (2024) 101802, arXiv: `2402.16365`, URL: `https://cds.cern.ch/record/2890051` (cit. on p. 4).

[6] Armen Tumasyan et al., *Measurement of the inclusive and differential WZ production cross sections, polarization angles, and triple gauge couplings in pp collisions at* $\sqrt{s} = 13$ *TeV*, JHEP **2207** (2022) 032, arXiv: `2110.11231`, URL: `https://cds.cern.ch/record/2786853` (cit. on p. 4).

[7] Georges Aad et al., *Evidence of pair production of longitudinally polarised vector bosons and study of CP properties in* $ZZ \to 4l$ *events with the ATLAS detector at* $\sqrt{s} = 13$ *TeV*, JHEP **2312** (2023) 107, arXiv: `2310.04350`, URL: `https://cds.cern.ch/record/2874981` (cit. on pp. 4, 39).

[8] *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*, tech. rep., CERN, 2023, URL: `https://cds.cern.ch/record/2859330` (cit. on pp. 4, 65, 67, 69–72, 75, 76, 80, 104, 105, 107, 108, 129, 130, 132, 141, 142, 193, 226–230).

[9] Lyndon R Evans and Philip Bryant, *LHC Machine*, JINST **3** (2008) S08001, This report is an abridged version of the LHC Design Report (CERN-2004-003), URL: `https://cds.cern.ch/record/1129806` (cit. on pp. 9–11).

[10] G Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003, Also published by CERN Geneva in 2010, URL: `https://cds.cern.ch/record/1129811` (cit. on p. 9).

[11] *Linear accelerator 2*, (2012), URL: `https://cds.cern.ch/record/1997427` (cit. on p. 10).

[12] K. HANKE, *PAST AND PRESENT OPERATION OF THE CERN PS BOOSTER*, International Journal of Modern Physics A **28** (2013) 1330019, eprint: `https://doi.org/10.1142/S0217751X13300196`, URL: `https://doi.org/10.1142/S0217751X13300196` (cit. on p. 10).

[13] Donald Cundy and Simone Gilardoni, *The Proton Synchrotron (PS): At the Core of the CERN Accelerators*, Adv. Ser. Direct. High Energy Phys. **27** (2017) 39, URL: `https://cds.cern.ch/record/2312566` (cit. on pp. 10, 11).

[14] *The Super Proton Synchrotron — home.cern*, `https://home.cern/science/accelerators/super-proton-synchrotron`, [Accessed 28-08-2024] (cit. on pp. 10, 11).

[15] Julie Haffner, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, (2013), General Photo, URL: `https://cds.cern.ch/record/1621894` (cit. on p. 10).

[16] *The Proton Synchrotron — home.cern*, `https://home.cern/science/accelerators/proton-synchrotron`, [Accessed 28-08-2024] (cit. on p. 11).

[17] R Alemany-Fernandez et al., *Operation and Configuration of the LHC in Run 1*, (2013), URL: `https://cds.cern.ch/record/1631030` (cit. on p. 11).

[18] Rende Steerenberg et al., *Operation and performance of the CERN Large Hadron Collider during proton Run 2*, (2019) MOPMP031, URL: `https://cds.cern.ch/record/2696126` (cit. on pp. 11–13).

[19] G. Aad et al., *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, The European Physical Journal C **83** (2023), ISSN: 1434-6052, URL: `http://dx.doi.org/10.1140/epjc/s10052-023-11747-w` (cit. on pp. 12, 141).

[20] Joao Pequenao, "Computer generated image of the whole ATLAS detector", 2008, URL: `https://cds.cern.ch/record/1095924` (cit. on p. 13).

[21] Giles Strong, *On the impact of selected modern deep-learning techniques to the performance and celerity of classification models in an experimental high-energy physics use case*, Machine Learning: Science and Technology **1** (2020) (cit. on p. 14).

[22] The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3** (2008) S08003, URL: `https://doi.org/10.1088%5C%2F1748-0221%5C%2F3%5C%2F08%5C%2Fs08003` (cit. on p. 14).

[23] A. Airapetian et al., *ATLAS detector and physics performance: Technical Design Report, 1*, Technical Design Report ATLAS, CERN, 1999, URL: `http://cds.cern.ch/record/391176` (cit. on p. 14).

[24] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, Journal of Instrumentation **13** (2018), ISSN: 1748-0221, URL: `http://dx.doi.org/10.1088/1748-0221/13/05/T05008` (cit. on p. 14).

[25] Morad Aaboud et al., *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **79** (2019) 639, arXiv: `1902.04655`, URL: `https://cds.cern.ch/record/2657964` (cit. on pp. 16, 68).

[26] G. Aad et al., *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data*, Journal of Instrumentation **14** (2019) P12006, ISSN: 1748-0221, URL: `http://dx.doi.org/10.1088/1748-0221/14/12/P12006` (cit. on pp. 16, 141).

[27] Georges Aad et al., *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV*, Eur. Phys. J., C **81** (2021) 578, arXiv: `2012.00578`, URL: `https://cds.cern.ch/record/2746302` (cit. on pp. 16, 69, 141).

[28] Matteo Cacciari, Gavin P Salam, and Gregory Soyez, *The anti-ktjet clustering algorithm*, Journal of High Energy Physics **2008** (2008) 063, ISSN: 1029-8479, URL: `http://dx.doi.org/10.1088/1126-6708/2008/04/063` (cit. on pp. 17, 69, 70).

[29] Morad Aaboud et al., *Jet reconstruction and performance using particle flow with the ATLAS Detector. Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466, arXiv: `1703.10485`, URL: `https://cds.cern.ch/record/2257597` (cit. on pp. 17, 69).

[30] Georges Aad et al., *Jet energy scale and resolution measured in proton-proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2021) 689, arXiv: `2007.02645`, URL: `https://cds.cern.ch/record/2722869` (cit. on pp. 17, 141).

[31] The ATLAS collaboration, *Operation of the ATLAS trigger system in Run 2*, Journal of Instrumentation **15** (2020) P10004, ISSN: 1748-0221, URL: `http://dx.doi.org/10.1088/1748-0221/15/10/P10004` (cit. on p. 17).

[32] ATLAS Collaboration, *Athena*, version 22.0.1, 2019, URL: `https://doi.org/10.5281/zenodo.2641997` (cit. on p. 17).

[33] Anthony Zee, *Group theory in a nutshell for physicists*, en, In a Nutshell, Princeton University Press, 2016 (cit. on p. 19).

[34] Michele Maggiore, *A modern introduction to quantum field theory*, Reprinted, Oxford University Press, 2006, ISBN: 9780198520733, URL: `http://slubdd.de/katalog?TN_libero_mab2` (cit. on p. 20).

[35] Elliot Leader, *Spin in Particle Physics*, Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology, Cambridge University Press, 2001 (cit. on p. 21).

[36] Gordon L. Kane, *Modern elementary particle physics*, Second edition, , ©2017, ISBN: 1107165083, URL: `http://slubdd.de/katalog?TN_libero_mab216448419` (cit. on p. 21).

[37] Mark Thomson, *Modern particle physics*, Cambridge University Press, c 2013, ISBN: 1107034264, URL: `http://slubdd.de/katalog?TN_libero_mab2` (cit. on pp. 21, 26, 27).

[38] Paul Langacker, *The standard model and beyond*, CRC Press, 2010, ISBN: 1420079069, URL: `https://katalog.slub-dresden.de/id/0-61049211X` (cit. on p. 23).

[39] Don B. Lichtenberg, *The standard model of elementary particles*, Bibliopolis, 1991, ISBN: 9788870882605, URL: `http://slubdd.de/katalog?TN_libero_mab2246056` (cit. on p. 24).

[40] E. A. Paschos and Emmanuel A. Paschos, *Electroweak theory*, 1. publ., Cambridge Univ. Press, 2007, ISBN: 9780521860987, URL: `http://slubdd.de/katalog?TN_libero_mab21564548` (cit. on pp. 25, 26).

[41] R. L. Workman et al., *Review of Particle Physics*, PTEP **2022** (2022) 083C01 (cit. on p. 32).

[42] Alessandro Ballestrero, Ezio Maina, and Giovanni Pelliccioli, *W boson polarization in vector boson scattering at the LHC*, Journal of High Energy Physics **2018** (2018), ISSN: 1029-8479, URL: `http://dx.doi.org/10.1007/JHEP03(2018)170` (cit. on pp. 32, 33).

[43] Michał Szleper, *The Higgs boson and the physics of WW scattering before and after Higgs discovery*, 2015, arXiv: `1412.8367 [hep-ph]`, URL: `https://arxiv.org/abs/1412.8367` (cit. on pp. 34–36).

[44] Michael Rauch, *Vector-Boson Fusion and Vector-Boson Scattering*, 2016, arXiv: `1610.08420 [hep-ph]`, URL: `https://arxiv.org/abs/1610.08420` (cit. on pp. 36, 38).

[45] Mareen Hoppe, Marek Schönherr, and Frank Siegert, *Polarised cross sections for vector boson production with SHERPA*, 2023, arXiv: `2310.14803 [hep-ph]` (cit. on pp. 37, 75, 81).

[46] 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** (2014) 079, arXiv: `1405.0301`, URL: `https://cds.cern.ch/record/1699128` (cit. on pp. 38, 43, 75).

[47] Diogo Buarque Franzosi, Olivier Mattelaer, Richard Ruiz, and Sujay Shil, *Automated predictions from polarized matrix elements*, Journal of High Energy Physics **2020** (2020), ISSN: 1029-8479, URL: `http://dx.doi.org/10.1007/JHEP04(2020)082` (cit. on p. 38).

[48] *FAQ 2014 : Questions : MadGraph5_aMC@NLO — answers.launchpad.net*, `https://answers.launchpad.net/mg5amcnlo/+faq/2014`, [Accessed 23-08-2024] (cit. on p. 38).

[49] *FAQ 2014 : Questions : MadGraph5_aMC@NLO — answers.launchpad.net*, `https://answers.launchpad.net/mg5amcnlo/+faq/2778`, [Accessed 23-08-2024] (cit. on p. 38).

[50] Ansgar Denner, Christoph Haitz, and Giovanni Pelliccioli, *NLO EW and QCD corrections to polarised same-sign WW scattering at the LHC*, 2024, arXiv: `2409.03620 [hep-ph]`, URL: `https://arxiv.org/abs/2409.03620` (cit. on pp. 39, 73, 81, 103–106, 189).

[51] Andy Buckley et al., *General-purpose event generators for LHC physics*, Phys. Rep. **504** (2011) 145, Comments: 226 pages, 30 figures, submitted to Physics Reports, arXiv: `1101.2599`, URL: `https://cds.cern.ch/record/1322340` (cit. on pp. 41, 45).

[52] T Gleisberg et al., *Event generation with SHERPA 1.1*, Journal of High Energy Physics **2009** (2009) 007, ISSN: 1029-8479, URL: `http://dx.doi.org/10.1088/1126-6708/2009/02/007` (cit. on p. 42).

[53] Michael Edward Peskin and Daniel V. Schroeder, *An introduction to quantum field theory*, [Nachdr.], Westview Press, [ca. 2007], ISBN: 0201503972, URL: `https://katalog.slub-dresden.de/id/0-1611445531` (cit. on p. 43).

[54] Enrico Bothmann et al., *Event generation with Sherpa 2.2*, SciPost Physics **7** (2019), ISSN: 2542-4653, URL: `http://dx.doi.org/10.21468/SciPostPhys.7.3.034` (cit. on pp. 43, 75).

[55] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, Journal of High Energy Physics **2010** (2010), ISSN: 1029-8479, URL: `http://dx.doi.org/10.1007/JHEP06(2010)043` (cit. on p. 43).

[56] Manuel Bähr et al., *Herwig++ physics and manual*, The European Physical Journal C **58** (2008) 639, ISSN: 1434-6052, URL: `http://dx.doi.org/10.1140/epjc/s10052-008-0798-9` (cit. on p. 44).

[57] Torbjörn Sjöstrand et al., *An introduction to PYTHIA 8.2*, Computer Physics Communications **191** (2015) 159, ISSN: 0010-4655, URL: `http://dx.doi.org/10.1016/j.cpc.2015.01.024` (cit. on p. 44).

[58] Stefan Höche, Frank Krauss, Steffen Schumann, and Frank Siegert, *QCD matrix elements and truncated showers*, Journal of High Energy Physics **2009** (2009) 053, ISSN: 1029-8479, URL: `http://dx.doi.org/10.1088/1126-6708/2009/05/053` (cit. on pp. 44, 143).

[59] S Agostinelli et al., *GEANT4–a simulation toolkikt. GEANT4. A Simulation toolkit*, Nucl. Instrum. Methods Phys. Res., A **506** (2003) 250, URL: `https://cds.cern.ch/record/602040` (cit. on p. 46).

[60] John Iovine, *Understanding neural networks*, Prompt, 2000, ISBN: 9780790611150, URL: `http://slubdd.de/katalog?TN_libero_mab2` (cit. on p. 47).

[61] Claudia Hagen, *Neuronale Netze zur statistischen Datenanalyse*, Als Ms. gedr., Shaker, 1997, ISBN: 9783826529726, URL: `http://slubdd.de/katalog?TN_libero_mab2` (cit. on pp. 47, 48).

[62] Michael Chester, *Neural networks a tutorial*, PTR Prentice Hall, 1993, ISBN: 9780133689037, URL: `http://slubdd.de/katalog?TN_libero_mab2` (cit. on p. 47).

[63] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*, `http://www.deeplearningbook.org`, MIT Press, 2016 (cit. on pp. 48–50, 53, 54, 56).

[64] Prajit Ramachandran, Barret Zoph, and Quoc V. Le, *Searching for Activation Functions*, 2017, arXiv: `1710.05941 [cs.NE]` (cit. on pp. 49, 111).

[65] Christopher M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag, 2006, ISBN: 0387310738 (cit. on p. 49).

[66] Diederik P. Kingma and Jimmy Ba, *Adam: A Method for Stochastic Optimization*, 2017, arXiv: `1412.6980 [cs.LG]` (cit. on pp. 51, 52).

[67] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams, *Learning representations by back-propagating errors*, Nature **323** (1986) 533, URL: `https://api.semanticscholar.org/CorpusID:205001834` (cit. on p. 51).

[68] Tijmen Tieleman and Geoffrey Hinton, *Lecture 6.5—RMSprop: Divide the gradient by a running average of its recent magnitude*, Coursera: Neural Networks for Machine Learning, 2012 (cit. on p. 51).

[69] Thomas Kurbiel and Shahrzad Khaleghian, *Training of Deep Neural Networks based on Distance Measures using RMSProp*, 2017, arXiv: `1708.01911 [cs.LG]` (cit. on p. 51).

[70] *Review of the 13 TeV ATLAS Open Data release*, tech. rep., CERN, 2020, URL: `https://cds.cern.ch/record/2707171` (cit. on pp. 54, 55).

[71] 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** (2014) 1929, URL: `http://jmlr.org/papers/v15/srivastava14a.html` (cit. on p. 55).

[72] Sergey Ioffe and Christian Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, CoRR **abs/1502.03167** (2015), arXiv: `1502.03167`, URL: `http://arxiv.org/abs/1502.03167` (cit. on p. 56).

[73] Tom Fawcett, *An introduction to ROC analysis*, Pattern Recognition Letters **27** (2006) 861, ROC Analysis in Pattern Recognition, ISSN: 0167-8655, URL: `https://www.sciencedirect.com/science/article/pii/S016786550500303X` (cit. on p. 57).

[74] Scott Lundberg and Su-In Lee, *A Unified Approach to Interpreting Model Predictions*, 2017, arXiv: `1705.07874 [cs.AI]`, URL: `https://arxiv.org/abs/1705.07874` (cit. on pp. 59, 132).

[75] L. S. Shapley, "17. A Value for n-Person Games", *Contributions to the Theory of Games, Volume II*, ed. by Harold William Kuhn and Albert William Tucker, Princeton University Press, 1953 307, ISBN: 9781400881970, URL: `https://doi.org/10.1515/9781400881970-018` (cit. on p. 61).

[76] G. Aad et al., *Performance of electron and photon triggers in ATLAS during LHC Run 2*, The European Physical Journal C **80** (2020), ISSN: 1434-6052, URL: `http://dx.doi.org/10.1140/epjc/s10052-019-7500-2` (cit. on p. 67).

[77] Georges Aad et al., *Performance of the ATLAS muon triggers in Run 2*, JINST **15** (2020) P09015, arXiv: `2004.13447`, URL: `https://cds.cern.ch/record/2716326` (cit. on p. 67).

[78] Christos Anastopoulos et al., *Electron identification and efficiency measurements in 2017 data*, tech. rep. ATL-COM-PHYS-2018-1727, CERN, 2019, URL: `https://cds.cern.ch/record/2652163` (cit. on pp. 68, 79, 142).

[79] G. Aad et al., *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** (2016), ISSN: 1434-6052, URL: `http://dx.doi.org/10.1140/epjc/s10052-016-4395-z` (cit. on p. 70).

[80] M. Aaboud et al., *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** (2018), ISSN: 1434-6052, URL: `http://dx.doi.org/10.1140/epjc/s10052-018-6288-9` (cit. on p. 70).

[81] D Adams et al., *Recommendations of the Physics Objects and Analysis Harmonisation Study Groups 2014*, tech. rep., CERN, 2014, URL: `https://cds.cern.ch/record/1743654` (cit. on p. 70).

[82] Will Buttinger, *Updated Overlap Removal Working Points*, tech. rep., 2015, URL: `https://indico.cern.ch/event/457238/` (visited on 10/11/2024) (cit. on p. 70).

[83] G. Aad et al., *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** (2019), ISSN: 1434-6052, URL: `http://dx.doi.org/10.1140/epjc/s10052-019-7450-8` (cit. on pp. 71, 141).

[84] Luka Nedic, "Probing polarised production of two same-charge $W^{\pm}$ bosons produced in association with two jets with the ATLAS detector", University of Oxford, 2025 (cit. on pp. 76, 109).

[85] 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", Presented 13 Jan 2023, Freiburg U., 2022, URL: `https://cds.cern.ch/record/2865127` (cit. on p. 77).

[86] 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", Presented 03 Apr 2020, Albert-Ludwigs University of Freiburg, 2020, URL: `https://cds.cern.ch/record/2791516` (cit. on p. 79).

[87] Georges Aad et al., *Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data*, JINST **14** (2019) P12006, arXiv: `1908.00005`, URL: `https://cds.cern.ch/record/2684552` (cit. on p. 79).

[88] Anders Andreassen and Benjamin Nachman, *Neural networks for full phase-space reweighting and parameter tuning*, Phys. Rev. D **101** (9 2020) 091901, URL: `https://link.aps.org/doi/10.1103/PhysRevD.101.091901` (cit. on p. 90).

[89] Erik Bachmann, "Polarization studies in same-sign W-boson pair production at the LHC with the ATLAS detector", Presented 12 Apr 2023, TUD Dresden University of Technology, 2023, URL: `https://cds.cern.ch/record/2872930` (cit. on pp. 91, 108, 111).

[90] Benedikt Biedermann, Ansgar Denner, and Mathieu Pellen, *Complete NLO corrections to W^+W^+ scattering and its irreducible background at the LHC*, Journal of High Energy Physics **2017** (2017), URL: `https://doi.org/10.1007%2Fjhep10%282017%29124` (cit. on pp. 103, 107).

[91] Ansgar Denner, Christoph Haitz, and Giovanni Pelliccioli, *NLO EW corrections to polarised W^+W^- production and decay at the LHC*, 2023, arXiv: `2311.16031 [hep-ph]`, URL: `https://arxiv.org/abs/2311.16031` (cit. on p. 103).

[92] 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** (2021), ISSN: 1029-8479, URL: `http://dx.doi.org/10.1007/JHEP10(2021)097` (cit. on pp. 103, 105, 106).

[93] Stefano Actis et al., *RECOLA: REcursive Computation of One-Loop Amplitudes*, Computer Physics Communications **214** (2017) 140, ISSN: 0010-4655, URL: `http://dx.doi.org/10.1016/j.cpc.2017.01.004` (cit. on p. 103).

[94] François Chollet et al., *Keras*, `https://keras.io` (cit. on p. 107).

[95] A. J. Barr et al., *Guide to transverse projections and mass-constraining variables*, Physical Review D **84** (2011), URL: `https://doi.org/10.1103%2Fphysrevd.84.095031` (cit. on p. 109).

[96] Soufiane Hayou, Arnaud Doucet, and Judith Rousseau, *On the Selection of Initialization and Activation Function for Deep Neural Networks*, 2018, arXiv: `1805.08266 [stat.ML]` (cit. on p. 111).

[97] URL: `https://gitlab.cern.ch/atlas-germany-dresden-vbs-group/optima` (visited on 11/15/2024) (cit. on p. 111).

[98] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework", *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019 (cit. on p. 111).

[99] URL: `https://optuna.readthedocs.io/en/stable/reference/samplers/generated/optuna.samplers.TPESampler.html` (visited on 01/15/2025) (cit. on p. 111).

[100] Richard Liaw et al., *Tune: A Research Platform for Distributed Model Selection and Training*, arXiv preprint arXiv:1807.05118 (2018) (cit. on p. 111).

[101] Liam Li et al., *A System for Massively Parallel Hyperparameter Tuning*, 2020, arXiv: `1810.05934 [cs.LG]` (cit. on p. 111).

[102] Robert May, Graeme Dandy, and Holger Maier, "Review of Input Variable Selection Methods for Artificial Neural Networks", *Artificial Neural Networks*, ed. by Kenji Suzuki, IntechOpen, 2011, chap. 2, URL: `https://doi.org/10.5772/16004` (cit. on p. 112).

[103] William Buttinger, *Using Event Weights to account for differences in Instantaneous Luminosity and Trigger Prescale in Monte Carlo and Data*, tech. rep., CERN, 2015, URL: `https://cds.cern.ch/record/2014726` (cit. on p. 141).

[104] Richard D Ball et al., *The PDF4LHC21 combination of global PDF fits for the LHC Run III**, Journal of Physics G: Nuclear and Particle Physics **49** (2022) 080501, ISSN: 1361-6471, URL: `http://dx.doi.org/10.1088/1361-6471/ac7216` (cit. on pp. 142, 143).

[105] Tie-Jiun Hou et al., *New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC*, Physical Review D **103** (2021), ISSN: 2470-0029, URL: `http://dx.doi.org/10.1103/PhysRevD.103.014013` (cit. on p. 143).

[106] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, The European Physical Journal C **75** (2015), ISSN: 1434-6052, URL: `http://dx.doi.org/10.1140/epjc/s10052-015-3397-6` (cit. on p. 143).

[107] Richard D. Ball et al., *Parton distributions for the LHC run II*, Journal of High Energy Physics **2015** (2015), ISSN: 1029-8479, URL: `http://dx.doi.org/10.1007/JHEP04(2015)040` (cit. on p. 143).

[108] URL: `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/PdfRecommendations` (visited on 06/10/2024) (cit. on pp. 143, 149).

[109] Aram Apyan et al., *Polarization measurements of the same-sign WW vector boson scattering at $\sqrt{s} = 13$ TeV with the ATLAS detector*, tech. rep., CERN, 2023, URL: `https://cds.cern.ch/record/2871465` (cit. on p. 151).

[110] CERN, *TRExFitter*, Accessed on 02-12-2024, URL: `https://trexfitter-docs.web.cern.ch/trexfitter-docs/` (cit. on p. 152).

[111] Kyle Cranmer, George Lewis, Lorenzo Moneta, Akira Shibata, and Wouter Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, tech. rep., New York U., 2012, URL: `https://cds.cern.ch/record/1456844` (cit. on p. 152).

[112] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, 2011, URL: `http://slubdd.de/katalog?TN_libero_mab2` (cit. on pp. 152, 155).

[113] Simone Amoroso et al., *Recommendations on the treatment of theoretical systematic uncertainties in statistical analysis of ATLAS data*, tech. rep., CERN, 2020, URL: `https://cds.cern.ch/record/2715689` (cit. on p. 158).

[114] CERN, *High Luminosity LHC Project*, Accessed on 16-12-2024, URL: `https://hilumilhc.web.cern.ch/content/hl-lhc-project` (cit. on p. 186).

[115] Michele Grossi, Massimiliano Incudini, Mathieu Pellen, and Giovanni Pelliccioli, *Amplitude-assisted tagging of longitudinally polarised bosons using wide neural networks*, The European Physical Journal C **83** (2023), ISSN: 1434-6052, URL: `http://dx.doi.org/10.1140/epjc/s10052-023-11931-y` (cit. on p. 190).

Danksagung

Tja, hier sind wir nun. Falls du bis hierhin wirklich diese Arbeit gelesen hast - ein riesiges Danke an dich! Ich hoffe, es war etwas Interessantes für dich dabei. Falls du primär hier bist, um zu schauen, wem ich für meine Arbeit danken möchte: Herzlich willkommen, freut mich, dass du vorbeischaust.

Seit meiner Bachelorarbeit bin ich in derselben Arbeitsgruppe am IKTP und Teil der ATLAS/CERN-Community. Die Forschungsthemen und vor allem die Menschen in diesem Feld haben mich von Anfang an begeistert. Durch die Arbeit in diesem Forschungsbereich durfte ich so viele hochintelligente, herzliche und inspirierende Menschen kennenlernen. Dieses weltoffene Forschungsgebiet, das sich um kollektiven Erkenntnisgewinn dreht, hat mich nachhaltig geprägt. Für die gewonnenen Eindrücke und die Möglichkeit, selbst etwas beizutragen, möchte ich mich bedanken.

Ohne die richtige fachliche Unterstützung wäre mein Beitrag jedoch nicht möglich gewesen. Aus diesem Grund möchte ich mich bei Michael Kobel, Frank Siegert und Joany Manjarres für die Betreuung während meiner Bachelor-, Master- und Doktorarbeit bedanken. Ihr standet mir bei fachlichen Fragen stets zur Seite und habt mir die Freiräume gelassen, um mich fachlich zu entfalten. Die Betreuung durch euch verlief immer auf Augenhöhe und war geprägt von Herzlichkeit und gegenseitiger Wertschätzung.

Diese Herzlichkeit spiegelt sich auch innerhalb der Arbeitsgruppe wider. Durch euch bin ich jeden Tag mit einem Lächeln zum Institut gelaufen und habe mich auf einen weiteren tollen Tag gefreut. Ich habe mich von Anfang an in dieser Arbeitsgruppe zu Hause gefühlt und werde die Arbeit mit euch sehr vermissen. Ein besonderer Dank gilt dabei Mareen Hoppe, Erik Bachmann, Jan-Eric Nitschke, Maren Buhring, Tim Herrmann, Orcun Kolay und Philip Sommer, die mich in den letzten Schritten meiner Doktorarbeit sehr unterstützt und mit ihrem Feedback dazu beigetragen haben. Diese Arbeit fußt auf eurem kombinierten Wissen zur Theorie des Standardmodells, zum maschinellen Lernen, zur Statistik, zum wissenschaftlichen Schreiben und zu Analysemethoden. Des Weiteren möchte ich mich auch bei Abhishek Nag bedanken, der einen stets für Aktivitäten außerhalb des Arbeitsalltags motiviert hat. Unter den ehemaligen alten Hasen der Arbeitsgruppe bin ich besonders Stefanie Todt und Carsten Bittrich dankbar, die mich in das Forschungsfeld eingeführt haben und von deren Expertise ich sehr profitieren konnte.

Für die Unterstützung außerhalb der akademischen Forschung möchte ich mich sehr bei meiner Familie bedanken. Meine Großeltern, Waltraut und Karl-Heinz Stange, haben mich von klein auf durch Basteln in der Werkstatt und regelmäßiges Üben in Bereichen von Geografie bis Mathematik gefördert. Meine Eltern, Katrin Stange und Jörn Stange, haben mich in allen Lebenslagen unterstützt. Sie haben mir viel von dieser Welt gezeigt, die Leidenschaft für Wissenschaft in mir entdeckt und gefördert und immer dafür gesorgt, dass ich ein sorgenfreies Leben führen kann. Der Mensch, der mich den Großteil meines Lebens am engsten begleitet hat, war meine Schwester Josephine Garritzmann. Sie hat mich mit großgezogen, war immer für mich da und war der perfekte Gegenpol, den ich in meiner Entwicklung gebraucht habe.

Zu guter Letzt möchte ich mich auch noch bei meinen Freunden bedanken. Zusammen feiern gehen, bis zum Sonnenaufgang auf dem Balkon philosophieren, Sorgen und Ängste teilen, gemeinsam Sport machen... Ihr seid Teil meiner Familie, und ohne euch würde ich mich nicht vollständig fühlen.

Selbständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.
Die vorliegende Dissertation wurde in der Zeit von Januar 2021 bis Januar 2025 im Institut für Kern- und Teilchenphysik unter der wissenschaftlichen Betreuung von Dr. Frank Siegert angefertigt.
Es haben keine früheren erfolglosen Promotionsverfahren stattgefunden. Ich erkenne die Promotionsordnung des Bereichs Mathematik und Naturwissenschaften an der Technischen Universität Dresden von 23.02.2011 an.

Max Vincent Stange
Dresden, 30.01.2025