

Standard Model Precision Measurements with two Z bosons and two jets in ATLAS.

A Dissertation

Presented to

The Faculty of the Graduate School of Arts and Sciences  
Brandeis University

Department of Physics

Professor Gabriella Sciolla, Advisor

In Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy

by  
Prajita Bhattacharai

May 2023

This dissertation, directed and approved by Prajita Bhattacharai's Committee, has been accepted and approved by the Faculty of Brandeis University in partial fulfillment of the requirements for the degree of:

**DOCTOR OF PHILOSOPHY**

Wendy Cadge, Dean  
Graduate School of Arts and Sciences

Dissertation Committee:

Professor Gabriella Sciolla, Department of Physics, Brandeis University  
Professor Aram Apyan, Department of Physics, Brandeis University  
Dr. Alessandro Tricoli, Brookhaven National Laboratory

Copyright by  
Prajita Bhattacharai

2023

## Acknowledgements

I would not be able to get to this milestone without the people who supported me professionally and personally during my Ph.D. I am sincerely grateful for all the positive impacts that you have had on my career and life.

First, I must thank my doctoral advisor **Gabriella Sciolla**. Thank you for making me a part of your research group and providing me with several opportunities to grow professionally as a researcher. Reaching this milestone would not be possible without your continuous support and mentorship throughout the past 5.5 years. I have learned much from you, from physics to teaching, research proposal writing, and job applications. You always inspire me in many ways.

Second, I want to thank my other two members of the committee, **Aram Apyan** and **Alessandro Tricoli**. I am honored to have you on my committee. Thank you for reviewing my dissertation and giving me insightful feedback. In addition to the review, I must thank Aram for the discussions in the past year about vector boson scattering. I have learned a lot from you on this topic. I must also thank Alessandro for his guidance on this dissertation and the measurement presented here. Your comments have constantly challenged me to think about physics and measurement at the most fundamental level. You have always helped me realize gaps in my understanding and inspired me to be a better physicist.

Next, I would like to thank my supervising post-doctoral researcher **Max Goblirsch**. You have taught me so much, from the fundamentals to the technical and statistical aspects of experimental research in Particle Physics. I succeeded as a student with your kind, patient, and diligent mentorship. Thank you for mentoring me in the past five years, involving me in exciting projects in several areas of ATLAS, and always making time to help, even with a full schedule.

Thank you to **Professor Craig Blocker** for temporarily coming out of retirement to help me review the Theory chapter of this thesis. Thank you as well to **Guglielmo Frattari** and **Jiayi Chen** for helping me review different parts of this thesis.

Next, I would like to thank all members of the **VBS ZZjj Differential** analysis team for their essential contribution to different aspects of this measurement and for being patient, as I made many mistakes along the way. Special thanks to **Matteo D'Uffizi** for working side by side on this measurement from its initial conception to the finalization.

Thanks to the ITk mentors I had during my year at Brookhaven National Lab. First, I express my sincere gratitude to **David Lynn** for supervising my work on ITk, including the ATLAS qualification task. Thanks also to **Stefania Stucci**, **Phil Kuczewski**, and **Laura Bergsten** for all the guidance during this period.

Next, I would like to thank some additional ATLAS mentors who significantly impacted my career. Thanks to **Peter Jenni** for your optimism in my abilities and for helping me figure out the next steps of my career as a researcher. Thanks to **Valentina Cairo** and **Gabriel Facini** for guiding my work and involvement in Tracking CP.

I want to thank some additional mentors within the Sciolla group. First, thank you **Hannah Herde** for your continuous mentorship in different aspects of my career, from ITk to preparing talks and networking. I am most grateful for your help in facing one of my biggest fear, public speaking. Thank you **Zach Schillaci** for all your technical support from git to Athena.

To my two undergraduate mentors, **David Probst** and **Jonathan Kessler**, thank you for believing in my abilities and setting me on this path of graduate school.

I am grateful to be a part of the research group where I have met many of my close friends. Each of you has been an integral part of my support system in graduate school. Thank you to **Laura Bergsten** for being my only friend for a year while living in Long Island, always taking me on adventures and introducing me to numerous new things. Thank you **Jiayi Chen** for sharing the experience of graduate school and Zumba sessions and always believing in me. Thank you **Zach Schillaci** for your joyful friendship and time with **Pinot** and for introducing me to the stock market. Thank you **Francesca Capocasa** for all the good times, food and celebrations. Thank you to **Sagar Addepalli** for being the best flatmate in the past year, for many delicious meals, and for listening to my countless complaints and Taylor Swift playlist.

I also want to thank a few other friends from Brandeis, **Aldric Rosario**, **Salman Alam**, **Sarvesh Uplap**, **Guangfeng Yu** and **Francis Rivera** for going through the first year experience of countless hours spent solving problem sets and preparing for qualification exams together. Thanks to Aldric and Salman for always saying "yes" to an adventure and good food. Thanks to **Gess Kelly** for her friendship and optimism in me.

To my family, **Mom**, **Dad**, **Aagaman**, and **Aashma**, thank you for all your love and support.

To my three special friends, whom I consider my family in the US, **Anu, Gargi**, and **Sapana**, thank you for your friendship and support.

Finally, to **Dinesh**, thank you for all your unconditional support and for taking care of me while I wrote this thesis.

## ABSTRACT

Standard Model Precision Measurements with two Z bosons and two jets in ATLAS.

A dissertation presented to the Faculty of the  
Graduate School of Arts and Sciences of Brandeis University  
Waltham, Massachusetts

By Prajita Bhattacharai

This thesis presents the differential cross-section measurements as a function of several kinematic observables, sensitive to vector boson scattering for two Z bosons production in association with two jets ( $ZZ^*jj$ ), where each  $Z$  boson is reconstructed in same-flavor opposite-charge two leptons final state. The electroweak  $ZZ^*jj$  production includes the rare triple and quartic self-couplings of gauge bosons, whose scattering amplitude at high energies is regularized by the Standard Model  $H \rightarrow ZZ^*$  processes. The analysis is performed using the proton-proton collision data collected by the ATLAS experiment during LHC Run-2 at  $\sqrt{s} = 13$  TeV center-of-mass collision energy, corresponding to an integrated luminosity of  $139\text{ fb}^{-1}$ . Several theories Beyond the Standard Model are expected to modify the electroweak  $ZZ^*jj$  cross-sections at high energies. Certain phase space of some kinematic observables describing the electroweak  $ZZ^*jj$  process is sensitive to these new physics modifications. Therefore, performing differential cross-section measurements as a function of these kinematic observables for the electroweak  $ZZ^*jj$  processes is crucial. Given the low statistics in Run-2, the cross-sections corrected for detector effects are measured in an electroweak-enhanced phase space and compared to the state-of-the-art Standard Model predictions. The differential cross-sections are also used to constrain anomalous quartic gauge couplings using a dimension-8 Effective Field Theory formalism.

# Table of Contents

Acknowledgements . . . . .	iv
Abstract . . . . .	vi
List of Tables . . . . .	xi
List of Figures . . . . .	xiii
List of Abbreviations . . . . .	xix
<b>I Introduction</b>	<b>1</b>
<b>II Theory</b>	<b>3</b>
1 The Standard Model . . . . .	4
1.1 Symmetries . . . . .	4
1.2 Particles and Fields . . . . .	5
1.3 Theoretical Formulation of the Standard Model . . . . .	7
2 Limitations of the Standard Model . . . . .	21
3 Phenomenology of Proton-Proton Collisions . . . . .	23
4 Electroweak Diboson Physics . . . . .	28
<b>III Experimental Setup</b>	<b>33</b>
5 The Large Hadron Collider . . . . .	34
6 ATLAS Detector . . . . .	37
6.1 ATLAS Coordinate System . . . . .	37
6.2 Inner Detector . . . . .	39
6.3 Calorimeters . . . . .	40
6.4 Muon Spectrometer . . . . .	43
7 Physics Object Reconstruction . . . . .	44

7.1	Trigger . . . . .	45
7.2	Tracks and Vertex Reconstruction . . . . .	45
7.3	Electron Reconstruction . . . . .	48
7.4	Muon Reconstruction . . . . .	51
7.5	Jet Reconstruction . . . . .	53
8	Future Upgrades . . . . .	57
8.1	High-Luminosity LHC . . . . .	57
8.2	ATLAS Upgrades . . . . .	58
<b>IV</b>	<b>Analysis Overview</b>	<b>60</b>
9	Goals . . . . .	60
10	Detector-Level Object Selection . . . . .	61
10.1	Electrons . . . . .	61
10.2	Muons . . . . .	61
10.3	Jets . . . . .	62
10.4	Overlap Removal . . . . .	63
11	Trigger . . . . .	64
12	Event Selection . . . . .	67
12.1	Signal Region . . . . .	68
13	Fiducial Phase Space . . . . .	72
14	Datasets and Monte Carlo Simulation . . . . .	75
14.1	LHC Dataset . . . . .	75
14.2	Monte Carlo Samples . . . . .	76
14.3	Event Weights . . . . .	81
15	Definition of Measured Observables . . . . .	83

<b>V Analysis Strategy</b>	<b>85</b>
16 Background Estimation . . . . .	85
16.1 Data Driven Estimate of Non-Prompt Background . . . . .	85
17 Unfolding . . . . .	111
17.1 Method Overview . . . . .	111
17.2 Binning Optimization . . . . .	113
17.3 Method Validation . . . . .	116
17.4 Bias and Optimization . . . . .	118
18 Uncertainties on the Measurement . . . . .	122
18.1 Theoretical Uncertainties . . . . .	122
18.2 Experimental Uncertainties . . . . .	124
18.3 Unfolding Uncertainties . . . . .	127
18.4 Background Uncertainties . . . . .	131
18.5 Statistical Uncertainties . . . . .	131
18.6 Breakdown of Uncertainties . . . . .	131
<b>VI Results</b>	<b>134</b>
19 Detector Level Measurements . . . . .	135
20 Unfolded Differential Cross-sections . . . . .	139
21 Effective Field Theory ReInterpretation . . . . .	142
21.1 Introduction . . . . .	142
21.2 Event Generation . . . . .	146
21.3 Statistical Strategy for Limit Setting . . . . .	146
21.4 Results . . . . .	151
<b>VII Conclusion and Outlook</b>	<b>152</b>
22 Conclusion . . . . .	152

23	Outlook . . . . .	153
<b>References</b>		<b>154</b>

<b>Appendices</b>		<b>161</b>
A	Personal Contribution . . . . .	162
A.1	Contribution to $ZZ^*(\rightarrow 4\ell)jj$ Measurement . . . . .	162
A.2	Contribution to the ATLAS Experiment . . . . .	162
B	Additional Study on Unfolding Bias . . . . .	165
C	VBS-Suppressed Region . . . . .	170
C.1	Systematics . . . . .	170
C.2	Detector Level Measurements . . . . .	171
C.3	Unfolded Cross-sections . . . . .	171

# List of Tables

1	Properties of SM gauge bosons. [1]	6
2	Summary of different interactions of fermions under different gauge theory. The check mark identifies the interaction each fermion participates in.	7
3	Electroweak quantum numbers of the SM half-integer spin fermions (quarks and leptons) arranged in a left-handed $SU(2)$ doublet and right-handed $SU(2)$ singlet. The down-type left-handed quarks in $SU(2)_L$ quark doublets $d'$ , $s'$ and $b'$ are linear combinations of $d$ , $s$ and $b$ quarks [2].	8
4	Definition of the baseline and signal electrons.	62
5	Definition of the baseline and signal muons.	62
6	Definition of the baseline and signal jets.	63
7	Overlap removal used in the analysis. An object removed in one step does not enter into the subsequent step.	63
8	Trigger menu used in the analysis for event preselection	65
9	Details of event selection.	70
10	Details of the kinematic pre-selection applied to the particle-level electrons, muons, and jets. The required kinematic cuts are applied to the dressed leptons.	73
11	Details of the kinematic selections applied to form a particle-level quadruplet and a particle-level dijet in the fiducial volume.	74
12	List of signal MC samples used in the analysis. Each process consists of three different generation campaigns corresponding to the data-taking conditions of the ATLAS Run2 data-taking periods.	79
13	List of background MC samples used in the analysis. Each process consists of three different generation campaigns corresponding to the data-taking conditions of the ATLAS Run2 data-taking periods.	80

14	List of MC samples used in the estimation and validation of the data-driven non-prompt background estimation. . . . .	81
15	Binning for all unfolded observables in VBS-Enhanced and VBS-Suppressed regions. . . . .	115
16	Breakdown of the relative systematic uncertainties (%) for each bin of $m_{jj}$ in the VBS-Enhanced region. . . . .	132
17	Eighteen dimension-8 operators from the Eboli model giving anomalous quartic gauge vertices in several multi-boson processes. . . . .	144
18	The eight genuine QGC dimension-8 operators constrained by the measurement. . . . .	145
19	Expected and observed limits based on the unfolded Asimov and measured data, respectively, for the selected dimension-8 Wilson coefficients in Eboli Model. The limits are obtained by simultaneously fitting unfolded differential cross-sections of the $m_{jj}$ and $m_{4\ell}$ , including the overflow contribution in the VBS-Enhanced region. . . . .	151
20	Breakdown of the relative systematic uncertainties (%) for each bin of $m_{jj}$ in the VBS-Suppressed region. . . . .	170

# List of Figures

1	The seventeen fundamental particles of the SM include three generations of twelve fermions, four gauge bosons, and the scalar Higgs bosons. [3]	5
2	Phenomenology of di-Z boson production in association with two jets in a proton-proton collider.	24
3	Parton distribution functions $xf_q(x, Q^2)$ for reference momentum transfer $Q_0^2 = 10 \text{ GeV}^2$ (left) and $Q_0^2 = 10^4 \text{ GeV}^2$ (right) [4].	26
4	Cross-sections for longitudinal component of the electroweak $VVjj$ production as a function of $\sqrt{s}$ [5]. These calculations were made before the discovery of the SM Higgs; thus, the calculations in the second figure use an SM-like Higgs with 120 GeV mass. For the SM Higgs with a mass of 125.10 GeV, the regularization is similar in shape.	29
5	Typical diagrams of LO $qq$ and $gg$ induced QCD $\alpha_S^2 \alpha_{EWK}^2$ production of $ZZ^*jj$ . The two $Z \rightarrow \ell\ell$ vertices each contribute an additional electroweak coupling of $\alpha_{EWK}$ .	29
6	Typical diagrams for LO $gg$ loop induced the QCD $\alpha_S^4 \alpha_{EWK}^2$ production of $ZZ^*jj$ . The two $Z \rightarrow \ell\ell$ vertices each contribute an additional electroweak coupling of $\alpha_{EWK}$ .	30
7	Feynman diagrams at LO for the EWK $\alpha_{EWK}^4$ production of $ZZ^*jj$ . The two $Z \rightarrow \ell\ell$ vertices each contribute an additional electroweak coupling of $\alpha_{EWK}$ .	31
8	A detailed layout of multiple-steps that goes into proton acceleration before entering the main LHC ring [6].	35
9	A detailed schematic of the ATLAS detector with all its sub-detectors [7].	38
10	A schematic of the right-handed ATLAS coordinate system [8].	39
11	A schematic of the inner detector of ATLAS showing the IBL, pixel detectors, SCT, and TRT [9].	41
12	A schematic of the electromagnetic and hadronic calorimeters in ATLAS [7].	42

13	A schematic of the different components of the muon spectrometer in ATLAS [7]. . . . .	43
14	Simplified representation of various particles traversing through the various layers of the ATLAS sub-detectors and leaving unique signatures [10]. . . . .	44
15	Trigger and data acquisition system in ATLAS [11]. . . . . . . . . . .	46
16	Schematic showing the five-track parameters [12]. . . . . . . . . . .	47
17	Schematic showing the two techniques of track reconstruction, primary inside-out and secondary outside-in [13]. . . . . . . . . . .	48
18	Distributions showing the identification (left) and isolation (right) efficiencies for electrons in data as a function of their $E_T$ [14]. . . . . . . . . .	50
19	Schematic of four different types of muons reconstructed using several layers of sub-detectors [15]. . . . . . . . . . .	52
20	Schematic of particle flow jet reconstruction [16]. . . . . . . . . . .	54
21	Resolution of jet transverse momentum for only cluster-based jets (LC+JES) and particle flow jets as a function of $p_T$ (left) and $\eta$ (right) [16]. . . . .	55
22	A schematic of different steps of jet calibration [17]. . . . . . . . . .	55
23	Timeline of LHC operation starting from 2011 to the planned HL-LHC upgrade. Taken from <a href="https://hilumilhc.web.cern.ch/content/hl-lhc-project">https://hilumilhc.web.cern.ch/content/hl-lhc-project</a> . . . . .	57
24	Schematic layout of ITK [18]. . . . . . . . . . .	59
25	Trigger efficiency as a function of $m_{4\ell}$ in events with a quadruplet and a dijet in data and SM prediction corresponding to the 2015-2016 data-taking period. . . . .	66
26	Centrality dependence for yield, EWK selection efficiency, and EWK significance. . . . . . . . . . .	69
27	Event display of a candidate $pp \rightarrow ZZjj \rightarrow e^+e^-\mu^+\mu^-jj$ recorded by the ATLAS experiment in Run-2 2017 data-taking period [19]. . . . . . . . .	71
28	Total integrated luminosity collected during data taking period in Run-2 [20]. . . . .	76
29	Various steps in MC sample generation. . . . . . . . . . .	77
30	Pile-up distributions in different Run-2 data-taking period. [20] . . . . .	78

31	Figure showing the decay angle $\theta_{\ell 1 \ell 2}^*$ ( $\theta_{\ell 3 \ell 4}^*$ ) of the negative lepton in the primary (secondary) pair's rest frame. [21]. . . . .	84
32	A schematic of the non-prompt lepton from semi-leptonic decays of b-hadrons. Jet activities surround the non-prompt muon, and the muon track does not point to the hard scatter interaction point. . . . .	86
33	An overview of the non-prompt background estimation. . . . .	87
34	Origins of leptons in the signal region in events with a quadruplet and a dijet. The lepton origin is classified by the IFF classifier tool. Only leptons that are part of the signal quadruplet are shown. . . . .	89
35	Origins of non-prompt electrons (left) and muons (right) in the signal region in events with a signal quadruplet and a dijet. The events are normalized to the total number of non-prompt electrons (left) and non-prompt muons (right). . . . .	89
36	Fraction of non-prompt electrons and muons in the $Z + jets$ and $t\bar{t}$ control regions as a function of their $p_T$ . . . . .	91
37	Sources of non-prompt electrons in $Z + jets$ (left) and $t\bar{t}$ (right) control regions. . . . .	91
38	Sources of non-prompt muons in $Z + jets$ (left) and $t\bar{t}$ (right) control regions. . . . .	92
39	Origins of non-prompt electrons (left) and muons (right) in the combined control region. . . . .	93
40	Additional baseline electrons as a function of $p_T$ in control regions. . . . .	94
41	Additional baseline muons as a function of $p_T$ in control regions. . . . .	95
42	Fake efficiencies of non-prompt electrons measured in the combined control region from data as a function of their $p_T$ , $\eta$ , and $n_{jets}$ . . . . .	99
43	Fake efficiencies of non-prompt muons measured in the combined control region from data as a function of their $p_T$ , $\eta$ , and $n_{jets}$ . . . . .	100
44	Fake efficiencies of electrons measured in the combined control region from data as a function of their $p_T$ and $\eta$ in two slices of $n_{jets}$ . . . . .	100

45	Fake efficiencies of muons measured in the combined control region from data as a function of their $p_T$ and $\eta$ in two slices of $n_{jets}$ . . . . .	101
46	Uncertainties on the fake efficiency of the fake electrons measured in the combined control region from data as a function of their $p_T$ , $\eta$ , and $n_{jets}$ . . . . .	102
47	Uncertainties on the fake efficiency of the fake muons measured in the combined control region from data as a function of their $p_T$ , $\eta$ , and $n_{jets}$ . . . . .	103
48	Sources of non-prompt electrons and muons in the different-flavor and the same-charge validation regions. . . . .	105
49	Comparisons of data to the total MC prediction and non-prompt background in the different flavor (top) and same charge (bottom) validation regions as a function of $m_{4\ell}$ . . . . .	107
50	Comparisons of data to the total MC prediction and non-prompt background in different-flavor (left) and same-charge (right) validation regions as a function of three kinematic observables used to parametrize the fake efficiencies in the combined control region. . . . .	108
51	MC prediction and fake-factor method estimate of the non-prompt background as a function of $m_{4\ell}$ in the VBS-Enhanced region. Black bands represent the systematic uncertainties from the fake factor method. . . . .	109
52	Detector level SM predictions of signal and background processes along with the non-prompt backgrounds estimated from the fake-factor method as a function of $m_{4\ell}$ (left) and $p_{T,4\ell}$ (right) in the VBS-Enhanced region. . . . .	110
53	Unfolding inputs from SM MC as a function of $m_{jj}$ in the VBS-Enhanced region. . . . .	114
54	MC technical closure test of the unfolding procedure for $m_{jj}$ . The detector-level MC distribution (in blue) is unfolded (black) with the nominal SM unfolding inputs and compared to the particle-level distribution (in red) from the same MC. . . . .	116

55	Injection test with dimension-8 <i>FT0</i> EFT operator. . . . .	117
56	Unfolding validation using physics variation where parton-initiated QCD (left) or the EWK (right) process cross-sections are varied. . . . .	118
57	A step-by-step overview of the data-driven closure test to get the unfolding bias. . . . .	120
58	Unfolding bias (left) and statistical uncertainty (right) with up to 4 unfolding iterations as a function of $m_{jj}$ in VBS-Suppressed region. . . . .	121
59	MC-based unfolding bias in each bin of $m_{jj}$ in the VBS-Enhanced region using Gaussian toys. The distribution shows the relative difference between unfolded and true values for toys as a function of the true value in each bin. The number of unfolding iterations is varied, and the final bias is chosen as the one-sigma intersection with the average value of the toy truth distribution (light-green region intersecting with the black curve). . . . .	129
60	The left plot shows detector-level predictions of $m_{jj}$ in VBS-Enhanced region predicted by SHERPA, MADGRAPH and MADGRAPH re-weighted to match the SHERPA lineshape. The right plot shows re-weighted MADGRAPH5 detector-level (blue) distribution, which is unfolded (black) using unfolding inputs from nominal-SHERPA and compared to the particle-level re-weighted MADGRAPH5 distribution (red). The lower panel of the right plot showing the ratio between re-weighted particle-level and unfolded-level MADGRAPH5 gives the QCD modeling uncertainties. . . . .	130
61	Uncertainties as a function of $m_{jj}$ in the VBS-Enhanced region. . . . .	133
62	Detector level distributions in the VBS-Enhanced region. . . . .	137
63	Detector level distributions in the VBS-Enhanced region. . . . .	138
64	Unfolded differential cross-sections in the VBS-Enhanced region. . . . .	140
65	Unfolded differential cross-sections in the VBS-Enhanced region. . . . .	141

66	Covariance matrix for $m_{jj} + m_{4\ell}$ unfolded differential cross-sections in the VBS-Enhanced region. . . . .	149
67	An example of setting constraints on dimension-8 $\mathcal{O}_{T,0}$ EFT operator with one-dimensional $m_{jj} + m_{4\ell}$ Asmiov-unfolded differential cross-sections. . . . .	150
68	An example of setting constraints on dimension-8 $\mathcal{O}_{T,0}$ EFT operator with one-dimensional $m_{jj} + m_{4\ell}$ unfolded differential cross-sections measured from data. . . . .	151
69	MC-toy-based unfolding bias and statistical uncertainty as a function of several numbers of iterations in each bin of $m_{jj}$ distribution in the VBS-Enhanced region. . . . .	167
70	MC-toy-based unfolding bias in each bin of $m_{jj}$ in the VBS-Enhanced region using Gaussian toys after subtracting the contribution of the fiducial fake events from both nominal and toy MC predictions. . . . .	168
71	Fraction of fake fiducial events as a function of $p_T$ and $\eta$ of the leading (x-axis) and the sub-leading jets (y-axis) in the VBS-Enhanced region. . . . .	169
72	Uncertainties as a function of $m_{jj}$ in the VBS-Suppressed region. . . . .	171
73	Detector-level distributions in the VBS-Suppressed region. . . . .	172
74	Detector-level distributions in the VBS-Suppressed region. . . . .	173
75	Unfolded differential cross-sections in the VBS-Suppressed region. . . . .	175
76	Unfolded differential cross-sections in the VBS-Suppressed region. . . . .	176

## List of Abbreviations

- ATLAS: A Torodial LHC Apparatus
- aQGC: anomalous Quartic Gauge Couplings
- BSM: Beyond the Standard Model
- C: Charge conjugation
- CERN: Conseil Européen pour la Recherche Nucléaire
- CR: Control Region
- CSC: Cathode Strip Chambers
- ECAL: Electromagnetic Calorimeter
- EFT: Effective Field Theory
- EWK: Electroweak
- fJVT: forward Jet Vertex Tagger
- FSR: Final State Radiation
- GRL: Good Run List
- H: Weak Hypercharge
- HCAL: Hadronic Calorimeter
- HF: Heavy Flavor
- HGTD: High Granularity Timing Detector
- HL-LHC: High Luminosity Large Hadron Collider
- HLT: High-level Trigger

- I: Weak Isospin
- IBL: Insertable B-Layer
- ID: Inner Detector
- IP: Interaction Point
- IFF: Isolation and Fake Forum
- JVT: Jet Vertex Tagger
- $\mathcal{L}_{\mathcal{SM}}$ : Lagrangian
- LAr: Liquid Argon
- LB: Luminosity Block
- LEP: Large Electron Positron Collider
- LF: Light Flavor
- LH: Left Handed
- LHC: Large Hadron Collider
- MC: Monte Carlo
- MS: Muon Spectrometer
- P: Parity
- PS: Proton Synchrotron
- PDF: Parton Distribution Function
- Q: Electric Charge
- QGC: Quartic Gauge Coupling

- QED: Quantum Electrodynamics
- QCD: Quantum Chromodynamics
- $(\mathcal{P})$ : Poincare group
- RF: Radio frequency
- RH: RightHanded
- SCT: Semiconductor Tracker
- SF: Scale factors
- SF-OC: Same-flavor, Opposite-charged
- SM: Standard Model
- SMEFT: Standard Model Effective Field Theory
- SR: Signal Region
- SPS: Super Proton Synchrotron
- T: Time-reversal
- TGC: Triple Gauge Coupling
- TRT: Transition Radiation Tracker
- TTVA: Track-to-vertex association
- VBS: Vector Boson Scattering
- VR: Validation Region
- VRSC: Same Charge Validation Region
- VRDF: Different Flavor Validation Region

- VEV: Vacuum Expectation Value

# Chapter I: Introduction

Particle Physics investigates the fundamental nature of the universe. The Standard Model (SM), the fundamental theory of Particle Physics, provides a theoretical formulation that explains all known elementary particles, their interactions, and three of the four fundamental forces observed in nature: strong, electromagnetic, and weak forces. Fifty years after its formulation, the parameters predicted by the SM have been measured experimentally with high precision. The experimental discovery of the Higgs boson in 2012 established the SM as a complete and highly successful theory. However, the lack of description of the fourth fundamental force, gravity, and other experimentally evident phenomena, such as the existence of dark matter, suggest that the SM provides only an incomplete description of nature. Still, experimental evidence of new physics Beyond the Standard Model (BSM) has yet to be observed. The current primary objective of the Large Hadron Collider (LHC) at CERN is to look for experimental evidence of new physics which might explain or resolve some of the shortcomings of the SM.

New physics searches can be broadly categorized into two types, direct and indirect. The direct search focuses on finding experimental evidence of new physics signatures, such as new particles. In contrast, the indirect approach focuses on precisely measuring the parameters of the SM-predicted processes, looking for deviations compared to the state-of-the-art theoretical predictions. One critical phenomenon of the SM is the vector boson scattering (VBS) in final states involving multiple vector bosons, which are force-mediating particles for the electroweak (EWK) force. VBS processes are mediated by the rare triple and quartic self-couplings between the vector bosons, whose SM amplitudes interfere destructively with the Higgs-mediated processes. Several BSM theories modify either the strength of the vector boson self-couplings or that of the Higgs-mediated processes, thus, altering the extent of the interference and, consequently, the cross-sections from the predicted values. As many

new physics particles are expected to exist at high energies, such deviations are expected to appear at higher energies that have not been probed experimentally yet.

This thesis presents an indirect approach to new physics searches in one of the VBS-sensitive multiboson final states. The measurement analyzes the data collected by the ATLAS experiment at the LHC from 2015-2018 to measure the VBS-sensitive production of two  $Z$  bosons in association with two jets, where each  $Z$  boson decays into a pair of same-flavor opposite-charge (SF-OC) leptons. The quartic self-coupling of the vector bosons in  $ZZ^*(\rightarrow 4\ell)jj$  final state is experimentally accessible with the collected LHC dataset for the first time. The measurement uses a model-independent approach to look for anomalous quartic gauge couplings. Thus, measurements presented in this thesis are at the frontier of Particle Physics, pushing the boundaries of new physics searches through an indirect approach.

The theory of SM, its shortcomings, and the  $ZZ^*(\rightarrow 4\ell)jj$  process are discussed in Chapter *II*. The LHC and ATLAS experiments are then introduced in Chapter *III*. Chapters *IV* and *V* discuss the details of the measurement. The final results are presented in Chapter *VI*.

## **Chapter II: Theory**

This chapter describes the theoretical framework of the measurements discussed in this thesis. Section 1 introduces the Standard Model of Particle Physics and concepts relevant to this thesis. Section 2 discusses the outstanding problems with the SM, thus, motivating the experimental measurement. Section 3 discusses the phenomenology of the proton-proton collisions, and Section 4 discusses the physics of two  $Z$  bosons production in association with two jets.

# 1 The Standard Model

The SM of Particle Physics is a mathematical framework based on quantum field theory, incorporating quantum mechanics and special relativity. The SM describes all known fundamental particles in nature and their interactions. It consists of two sets of particles with intrinsic angular momentum: half-integer-spin fermions that are fundamental constituents of matter particles and force-carrying integer-spin bosons. The seventeen fundamental particles of the SM and their properties, such as mass, charge, and intrinsic spin, are shown schematically in Figure 1. Two textbooks on Particle Physics, Mark Thomson’s Modern Particle Physics [22], and Halzen & Martin’s Quarks & Leptons [2] guide the discussion written in this section.

## 1.1 Symmetries

The fundamental particles of the SM and their interactions can be described by constructing a general renormalizable Lagrangian ( $\mathcal{L}_{SM}$ ) that respects certain sets of given symmetries. The Lagrangian of the SM is independent of the reference frame, naturally respecting the complete external symmetries of special relativity, the Poincare group ( $\mathcal{P}$ ). Thus, the SM is invariant under spacetime translations, boosts, and rotations. Additionally, by construct of the Lagrangian, the SM respects an internal local gauge symmetry  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ . The  $SU(3)_C$  symmetry is associated with the Quantum Chromodynamics (QCD) discussed in more detail in Section 1.3.3. The  $SU(2)_L \otimes U(1)_Y$  gauge symmetry discussed in 1.3.4 is associated with the unified electroweak theory that combines Quantum Electrodynamics (QED) and the weak interactions.

According to Noether’s theorem, a quantity is conserved for each continuous transformation that leaves the Lagrangian invariant [23]. Several interesting physical quantum numbers are conserved as a consequence of the symmetries respected by the SM. The  $SU(3)_C$  in QCD conserves color charge. Weak isospin ( $I$ ) and weak hypercharge ( $Y$ ) are the quantum num-

bers associated with the  $SU(2)_L$  and  $U(1)_Y$  gauge groups, respectively. At low energies the  $SU(2)_L \otimes U(1)_Y$  symmetry is spontaneously broken as discussed in Section 1.3.4. The  $SU(2)_L$  group follows a chiral structure where the gauge fields couple explicitly to the left-handed (LH) chiral fermions states and the right-handed (RH) chiral anti-fermions states.

The SM also respects CPT symmetry, a combination of three additional discrete symmetries, charge conjugation (C), parity (P), and time-reversal (T). The charge-conjugation transformation transforms particles to anti-particles by reversing the quantum numbers, whereas the parity transformation transforms left-handed particles to right-handed particles and vice-versa.

## 1.2 Particles and Fields

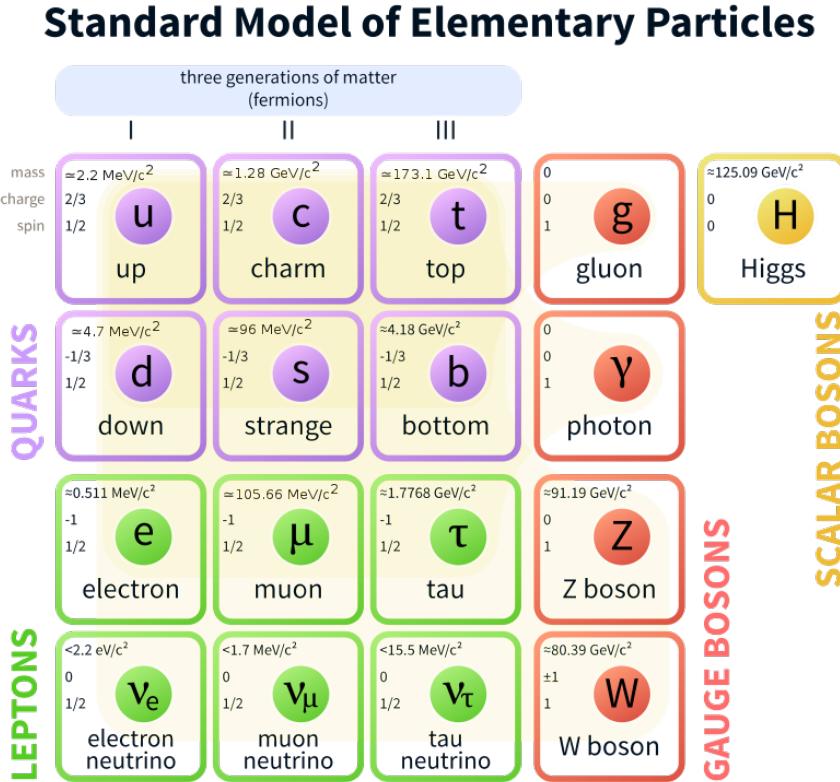


Figure 1: The seventeen fundamental particles of the SM include three generations of twelve fermions, four gauge bosons, and the scalar Higgs bosons. [3]

The twelve of the seventeen fundamental particles shown in Figure 1 are half-integer-spin fermions. The fermions can be distinguished further into two categories, leptons and quarks, each having three generations of particles with similar properties. For each fermion, there exists an anti-fermion with the same additive quantum numbers but with opposite signs. The four spin-1 bosons shown in Table 1 are collectively called the gauge bosons. Quanta of these gauge fields mediate the electromagnetic, weak, and strong interactions and are invariant under various local gauge transformations [24]. As summarized in Table 2, different fermions participate in different interactions, and a gauge coupling parameter governs the interaction strength.

Massless photon ( $\gamma$ ) mediates the electromagnetic interaction, whereas the massive  $W$  and  $Z$  bosons mediate weak interaction. The electric charge ( $Q$ ), which is conserved in all interactions, is related to the isospin and hypercharge by  $Q = I_3 + \frac{Y}{2}$ , where  $I_3$  is the third component of the weak isospin. As a consequence of the chiral structure of  $SU(2)_L$ , each generation of fermion contains a left-handed doublet with  $I_3 = \pm\frac{1}{2}$  and a right-handed singlet carrying  $I_3 = 0$  as shown in Table 3.

Table 1: Properties of SM gauge bosons. [1]

Interaction Type		Particle	$Q$	Mass [GeV]	Symmetry Group
Electroweak	Electromagnetic	Photon ( $\gamma$ )	0	0	$SU(2) \otimes U(1)$
	Weak	$W^\pm$	$\pm 1$	80.4	
		$Z$ boson	0	91.2	
Strong		gluons (g)	0	0	$SU(3)$

Each generation of lepton, electron ( $e$ ), muon ( $\mu$ ) and tau ( $\tau$ ), is accompanied by a neutral particle called neutrino ( $\nu$ ) with same lepton flavor ( $\nu_e, \nu_\mu$  and  $\nu_\tau$ ). In SM, anti-neutrinos have an opposite lepton flavor quantum number than neutrinos. The lepton flavor is conserved in all interactions.

The quarks are categorized further into two categories, the up-type quarks with  $+\frac{2}{3}$  charge and the down-type quarks with  $-\frac{1}{3}$  charge. Up ( $u$ ), charm ( $c$ ), and top ( $t$ ) are the

Table 2: Summary of different interactions of fermions under different gauge theory. The check mark identifies the interaction each fermion participates in.

Particles		Strong $SU(3)$	Electromagnetic $U(1)$	Weak $SU(2)$
Quarks	$u, c, t$ $d, s, b$	✓	✓	✓
Leptons	$e, \mu, \tau$ $\nu_e, \nu_\mu, \nu_\tau$	- -	✓ -	✓ ✓

first, second, and third generation of the up-type quarks, while the down ( $d$ ), strange ( $s$ ) and bottom ( $b$ ) are the three generations of the down-type quarks. The quarks interact strongly with one another by strong interaction mediated by the massless neutral gluons, which follow from  $SU(3)$  gauge symmetry by exchange of color charges. Each quark can have either one of the three color charges (red, blue, and green), whereas an anti-quark can have either an anti-red, anti-blue or anti-green color charge. There are eight gluons in the SM with color charges formed by a linear combination of color with an anti-color charge excluding the colorless combination. Since gluons have a color charge, they interact with other gluons strongly. Only color-neutral hadronic states formed by a combination of quarks and gluons are observed experimentally.

The charge-neutral Higgs boson is the only spin-0 scalar particle in the SM. It gives masses to the three weak bosons ( $W^\pm$  and  $Z$ ), all quarks, and charged leptons through Spontaneous Symmetry Breaking, which is discussed in Section 1.3.4.

### 1.3 Theoretical Formulation of the Standard Model

Relativistic quantum field theory is the theoretical framework of the SM that describes elementary particles and their interactions. This section introduces the framework.

Table 3: Electroweak quantum numbers of the SM half-integer spin fermions (quarks and leptons) arranged in a left-handed  $SU(2)$  doublet and right-handed  $SU(2)$  singlet. The down-type left-handed quarks in  $SU(2)_L$  quark doublets  $d'$ ,  $s'$  and  $b'$  are linear combinations of  $d$ ,  $s$  and  $b$  quarks [2].

Particle Types	First	Second	Third	$I_3$	Y	Q
Leptons	$\begin{pmatrix} e \\ \nu_e \end{pmatrix}_L$	$\begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}_L$	$\begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}_L$	$-\frac{1}{2}$ $\frac{1}{2}$	-1 -1	-1 0
	$e_R$	$\mu_R$	$\tau_R$	0	-2	-1
Quarks	$\begin{pmatrix} u \\ d' \end{pmatrix}_L$	$\begin{pmatrix} c \\ s' \end{pmatrix}_L$	$\begin{pmatrix} t \\ b' \end{pmatrix}_L$	$\frac{1}{2}$ $-\frac{1}{2}$	$\frac{1}{3}$ $\frac{1}{3}$	$\frac{2}{3}$ $-\frac{1}{3}$
	$u_R$	$c_R$	$t_R$	0	$\frac{4}{3}$	$\frac{2}{3}$
	$d_R$	$s_R$	$b_R$	0	$-\frac{2}{3}$	$-\frac{1}{3}$

### 1.3.1 Lagrangian of the Standard Model

The compact Lagrangian density given in Equation 1.1 describes the dynamics of the SM and is invariant under the local gauge transformation of the  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$  symmetry group:

$$\mathcal{L}_{SM} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\not{D}\psi + h.c. + |D_\mu\phi|^2 - V(\phi) + \bar{\psi}_i y_{ij}\psi_j\phi + h.c. \quad (1.1)$$

The first term,  $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ , describes the dynamics of the gauge boson interactions. The term,  $i\bar{\psi}\not{D}\psi + h.c.$ , describes the interaction of the fermion and anti-fermion fields, respectively. The term  $|D_\mu\phi|^2$  describes the couplings between the Higgs boson and gauge bosons, whereas the term  $V(\phi)$  represents the Higgs potential and its self-interactions. The term  $\bar{\psi}_i y_{ij}\psi_j\phi$  generates masses for fermions based on their Yukawa couplings  $y_{ij}$  to the Higgs field. Similarly, the last term  $h.c.$  generates masses for anti-fermions. The SM lagrangian

density, as shown by Equation 1.1, is a combination of complex interactions that follow different types of local gauge symmetries, and concise derivations of different interactions are discussed individually in the following few sections.

### 1.3.2 Quantum Electrodynamics

Quantum electrodynamics describes electromagnetic interaction. The Dirac Lagrangian density ( $\mathcal{L}_{Dirac}$ ) describes the free propagation of a fermion in a vacuum as:

$$\mathcal{L}_{Dirac} = \bar{\psi} i\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi, \quad (1.2)$$

where  $\psi$  is the fermionic spinor,  $\gamma^\mu$  represents the Dirac matrices with  $\mu$  being the Lorentz index running from 0 to 3, and  $m$  is the mass of the fermion. The Lagrangian in Equation 1.2 is invariant under a  $U(1)$  global gauge transformation,

$$\psi \rightarrow \psi' = e^{iq\alpha} \psi, \quad (1.3)$$

where  $q$  is a parameter of the transformation itself and  $\alpha$  is a real phase factor. However, under the local gauge transformation of form

$$\psi \rightarrow \psi' = e^{iq\alpha(x)} \psi, \quad (1.4)$$

where  $\alpha$  depends on  $x = (x_0, x_1, x_2, t)$ , the Dirac Lagrangian in Equation 1.2 is not invariant.

To make the Lagrangian of Equation 1.2 invariant, a gauge field  $A_\mu$  with the following transformation properties is introduced,

$$A_\mu \rightarrow A_\mu - \partial_\mu \alpha. \quad (1.5)$$

$A_\mu$  couples to fermionic fields  $\psi(x, t)$  with strength  $q$ . A covariant derivative specific to the

local gauge transformation is defined as:

$$D_\mu = \partial_\mu - iqA_\mu. \quad (1.6)$$

The quantity  $q$  can be interpreted as the electric charge  $-e$  of fermion, which gives the coupling strength of QED. With these substitutions, the Dirac Lagrangian in Equation 1.2 changes to

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi, \quad (1.7)$$

which is invariant under  $U(1)$  gauge transformation respecting the  $U(1)$  gauge symmetry.

The gauge field  $A_\mu$  can be interpreted as the photon field and is related to the electromagnetic field tensor by

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \quad (1.8)$$

The gauge invariant kinetic term of the photon,  $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ , can be inserted into the Lagrangian in Equation 1.7, which gives the full Lagrangian of QED invariant under  $U(1)$  gauge transformation:

$$\mathcal{L}_{QED} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(i\gamma^\mu D_\mu - m)\psi. \quad (1.9)$$

$\mathcal{L}_{QED}$  in Equation 1.9 is the full Lagrangian for QED, and the electromagnetic phenomena can be derived by solving for the equations of motion applying the Lorentz gauge condition  $\partial_\mu A^\mu = 0$ . A small dimensionless coupling constant  $\alpha$  determines the relative strength of the QED interactions. The probability of particle interactions and decay rates can be calculated as the square modulus of a perturbative series in powers of  $\alpha$ . The value of  $\alpha$  depends on the energy scale because of the additional contributions from the vacuum polarization of virtual quark and virtual lepton loop corrections. For a given momentum transfer squared  $Q^2$  of the exchanged photon, the effective coupling constant depends on the  $\alpha_0 = \frac{e^2}{4\pi}$  with

zero momentum transfer as,

$$\alpha(Q^2) = \frac{\alpha_0(Q^2 = 0)}{1 - \Delta\alpha(Q^2)}, \quad (1.10)$$

with the small value of  $\alpha_0 \approx \frac{1}{137.036}$  [25].

### 1.3.3 Quantum Chromodynamics

Quantum Chromodynamics defines the interaction between the quarks, requiring  $SU(3)$  gauge transformation on the quark field with color charge  $j$  (red, blue, or green). The Dirac Lagrangian for a quark can be written similarly to that of an electron with a small modification to include all possible colors of quark field  $q_j$  as

$$\mathcal{L} = \bar{q}_j(i\gamma^\mu\partial_\mu - m)q_j. \quad (1.11)$$

The generators of the  $SU(3)$  group are eight linearly independent traceless Gell-Mann matrices that do not commute with each other such that

$$[T_a, T_b] = if_{abc}T_c, \quad (1.12)$$

where  $f_{abc}$  is the structure constant of  $SU(3)$ . The local  $SU(3)$  gauge transformation is

$$q(x) \rightarrow e^{i\alpha_a(x)T_a}q(x), \quad (1.13)$$

where  $T_a = \frac{\lambda_a}{2}$ , and  $a = 1, 2 \dots 8$ . To understand the source of gauge invariance in the Lagrangian of Equation 1.11, an infinitesimal transformation of the color field is considered as

$$q(x) \rightarrow [1 + i\alpha_a(x)T_a]q(x) \ni \partial_\mu q \rightarrow (1 + i\alpha_a T_a)\partial_\mu q + iT_a q \partial_\mu \alpha_a. \quad (1.14)$$

The last term  $iT_a q \partial_\mu \alpha_a$  breaks the gauge invariance. Similarly to QED, eight gauge fields corresponding to each  $a$ ,  $G_\mu^a$  are introduced with following transformation properties:

$$G_\mu^a \rightarrow G_\mu^a - \frac{1}{g_s} \partial_\mu \alpha_a - f_{abc} \alpha_b G_\mu^c. \quad (1.15)$$

These gauge fields  $G_\mu^a$  are the gluon fields. Similarly to QED, the covariant derivative is defined as

$$D_\mu = \partial_\mu + i g_s \frac{\lambda_a}{2} G_\mu^a, \quad (1.16)$$

where  $g_s$  is the coupling strength of the gluon fields to the quark fields.

The Lagrangian density in Equation 1.11 is then

$$\mathcal{L} = \bar{q}_j (i\gamma^\mu D_\mu - m) q_j. \quad (1.17)$$

Similarly to QED, a gauge-invariant kinetic term  $-\frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu}$ , dependent on the field strength tensor  $G_{\mu\nu}^a$  is added to Equation 1.17 to give the full QCD Lagrangian. The kinetic terms allow self-interaction within the gluon fields, which is an important feature of QCD. The field strength tensor  $G_{\mu\nu}^a$  is defined as

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f_{abc} G_\mu^b G_\nu^c. \quad (1.18)$$

Therefore, the complete  $SU(3)$  gauge invariant Lagrangian describing the quarks and gluons interaction is

$$\mathcal{L}_{QCD} = \bar{q}_j (i\gamma^\mu D_\mu - m) q_j - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu}. \quad (1.19)$$

Similarly to QED, the probability of particle interactions and decay rates in QCD can be calculated as the square-modulus of a perturbative series in powers of the strong coupling constant  $\alpha_S = \frac{g_S^2}{4\pi}$  which depends on the energy scale. The energy-dependent effective strong

coupling is

$$\alpha_S(q^2) = \frac{12\pi}{(33 - 2n_f) \ln(q^2/\Lambda_{QCD})}, \quad (1.20)$$

where  $n_f$  is the number of types of quarks with masses lower than the energy scale, and  $\Lambda_{QCD} \approx 0.2$  GeV is the *Landau pole* below which the coupling constant increases and diverges [26]. There are two important experimental consequences of the running of the strong coupling constant. First, quarks and gluons cannot be isolated at low energies, leading to *color confinement*. Second, at high energies, the strong interaction reaches *asymptotic freedom*, and the exchange of gluons requires minimum energy, causing an abundance of gluon-induced QCD processes at hadron colliders [26].

### 1.3.4 Electroweak Theory

Massive gauge bosons mediate weak interactions. A typical example of weak interactions is radioactive beta decay, where a neutron turns into a proton by emitting an electron and an anti-electron-neutrino. The Dirac equation formulated in 1930s could explain the motions of electrons via relativistic quantum mechanics. However, the nuclear decay processes were still a mystery. Fermi developed the first theory of weak interaction to explain beta decay using a four-fermion-interaction vertex. The formulation successfully describes the beta decay at low energies when the interaction energy is much smaller than the  $W$  boson mass. In the meantime, QED following the  $U(1)$  symmetry was formulated to explain the electromagnetic interaction. Experimental evidence suggested that an exchange of spin-1 massive particles mediated weak interactions. It was challenging to develop a local gauge invariant mathematical theory, including spin-1 massive gauge bosons, which conserves unitarity at high energies.

During the 1960s, Glashow, Salam, and Weinberg (GWS) worked independently and made different contributions to formulate a theoretical model of weak interactions following a local gauge invariance of  $SU(2)_L \otimes U(1)_Y$  [27] [28] [29]. The theory postulates the existence of four massless gauge bosons, two electrically-charged and two electrically neutral,

which mediate unified electromagnetic and weak interactions. However, the observed short range of weak force could only be explained with massive gauge bosons of electroweak interactions. Therefore, implying the underlying symmetry of weak interactions is broken by some mechanism, which was later understood through the Higgs Mechanism discussed in Section 1.3.5.

Experimental observations suggest weak interactions violate parity by only affecting the left-handed fermion and right-handed anti-fermion fields. Thus, the unified electroweak theory is described by  $SU(2)_L \otimes U(1)_Y$  gauge interactions. Similarly to the electric charge  $Q$  conserved in QED by  $U(1)$  symmetry, the weak hypercharge [ $Y = 2(Q - I_3)$ ] related to the electric charge and the weak isospin ( $I_3$ ) is conserved by the  $U(1)_Y$  symmetry. The fermion fields are represented by the left-handed doublets  $\chi_L$  and the right-handed singlets  $\psi_R$ , introduced in Table 3. The doublet and singlet field for the first generation of leptons are

$$\chi_L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \quad \text{and} \quad \psi_R = e_R, \text{ respectively.}$$

Similarly, the doublet and singlet field for the first generation of quarks are

$$\chi_L = \begin{pmatrix} u \\ d \end{pmatrix}_L \quad \text{and} \quad \psi_R = u_R \text{ or } d_R, \text{ respectively.}$$

The Lagrangian for these fermion fields should be invariant under local gauge transformation corresponding to both  $SU(2)_L$  and  $U(1)_Y$  symmetry as,

$$\chi_L \rightarrow e^{i\beta(x)Y + i\alpha_a(x)\tau_a} \chi_L, \quad (1.21)$$

$$\psi_R \rightarrow e^{i\beta(x)Y} \psi_R, \quad (1.22)$$

where  $\beta(x)$  and  $\alpha(x)$  are the local phase transformation for  $U(1)_Y$  and  $SU(2)_L$  symmetry groups respectively. Weak hypercharge operator  $Y$  and Pauli matrices  $\tau_{a=1,2,3}$  are generators of  $U(1)_Y$  and  $SU(2)_L$  symmetry groups respectively. Similarly to the formulation in QED

and QCD discussed in Section 1.3.2 and 1.3.3, four new field strength tensors  $B_{\mu\nu}$  and  $W_{\mu\nu}^a$  corresponding respectively to the  $U(1)_Y$  and  $SU(2)_L$  transformations are introduced. The  $SU(2)_L \otimes U(1)_Y$  gauge-invariant Lagrangian for a massless fermion and massless gauge fields is:

$$\mathcal{L}_0 = \bar{\chi}_L \gamma^\mu [i\partial_\mu - g \frac{\tau_a}{2} W_\mu^a + \frac{g'}{2} B_\mu] \chi_L + \bar{\psi}_R \gamma^\mu [i\partial_\mu + g' B_\mu] \psi_R - \frac{1}{4} W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \quad (1.23)$$

where similarly to QED and QCD, field strength tensors are defined in terms of the covariant derivative to preserve gauge-invariance in kinetic terms as

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \quad (1.24)$$

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g \epsilon^{abc} W_\mu^b W_\nu^c. \quad (1.25)$$

The non-Abelian part of the  $SU(2)_L$  transformation is represented by the last term of Equation 1.25, which gives the quartic and triple self-interactions between the gauge bosons with coupling strength  $g$ .

The electroweak Lagrangian in Equation 1.23 contains two terms, one of which gives rise to the charged-current interaction with the two  $SU(2)$  boson fields

$$W_\mu^\pm = \frac{W_\mu^1 \mp i W_\mu^2}{\sqrt{2}}, \quad (1.26)$$

via exchange of the  $W^\pm$  bosons and the neutral current interactions via exchange of the two neutral gauge boson fields  $W_\mu^3$  and  $B_\mu$ .

The Lagrangian for the charged-current interaction for the first generation of quarks and leptons is

$$\mathcal{L}_{CC} = \frac{g}{2\sqrt{2}} \{ W_\mu^\dagger [\bar{u} \gamma^\mu (1 - \gamma_5) d + \bar{\nu}_e \gamma^\mu (1 - \gamma_5) e] + h.c. \}. \quad (1.27)$$

The  $SU(2)_L$  charged-current interaction Lagrangian for the next two generations follows the

same formalism, establishing the universality of the quark and lepton interactions as a direct consequence of the gauge symmetry.

The neutral-current Lagrangian is given by

$$\mathcal{L}_{NC} = \sum_j \bar{\psi}_j \gamma^\mu \{ A_\mu [\frac{\tau_3}{2} \sin \theta_W + g' Y \cos \theta_W] + Z_\mu [\frac{\tau_3}{2} \cos \theta_W - g' Y \sin \theta_W] \} \psi_j, \quad (1.28)$$

where the two neutral gauge fields  $Z_\mu$  and  $A_\mu$  associated with  $Z$  boson and photon governing the weak neutral and electromagnetic interactions are obtained from an arbitrary linear combination of the  $W_\mu^3$  and  $B_\mu$  fields as

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix}. \quad (1.29)$$

The following condition is imposed to obtain QED from  $A_\mu$ :

$$\begin{aligned} g \sin \theta_W &= g' \cos \theta_W = e \\ Y &= Q - T_3, \end{aligned} \quad (1.30)$$

where  $T_3 = \frac{\tau_3}{2}$  is the weak isospin and  $\theta_W$  is the Weinberg mixing angle. The Weinberg angle has been measured experimentally with high precision and can be expressed in terms of the two  $SU(2)_L$  coupling  $g'$  and  $U(1)_Y$  coupling  $g$  as:

$$\begin{aligned} \sin \theta_W &= \frac{g'}{\sqrt{g^2 + g'^2}} \\ \cos \theta_W &= \frac{g}{\sqrt{g^2 + g'^2}}. \end{aligned} \quad (1.31)$$

Similarly to QED and QCD, the probability of particle interactions and decay rates in electroweak interactions can be calculated as the square-modulus of a perturbative series in powers of the weak coupling constant  $\alpha_{EWK}$ . The value of  $\alpha_{EWK}$  depends on the energy scale and the type of weak interaction, charged-current, neutral-current, or the non-Abelian

self-couplings of the gauge bosons.

### 1.3.5 Higgs Mechanism

The Lagrangian in Equation 1.23 describes the electroweak interactions only for massless fermions and massless gauge bosons, which contradicts the experimental observations. The mass origin of the fermions and gauge bosons is discussed in this section. Massive gauge bosons in the Lagrangian 1.23 can be accommodated through the Brout-Englert-Higgs (BEH) mechanism by introducing a complex scalar field  $\phi$  in the spinor representation of  $SU(2)_L$  doublet [30] as,

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}. \quad (1.32)$$

A new term in the SM Lagrangian  $\mathcal{L}_{BEH}$  depending on this scalar field can be defined as

$$\mathcal{L}_{BEH} = (D_\mu \phi)^\dagger (D^\mu \phi) - \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \quad (1.33)$$

which is invariant under  $SU(2)_L \otimes U(1)_Y$  transformations. The first term  $(D_\mu \phi)^\dagger (D^\mu \phi)$  describes the kinematic of the new fields, and the BEH potential  $V(\phi)$  is given by the second term as

$$V(\phi) = \lambda (\phi^\dagger \phi)^2 - \mu^2 \phi^\dagger \phi, \quad (1.34)$$

where the term  $\lambda (\phi^\dagger \phi)^2$  describes the quartic self-interactions of the scalar fields. The values of  $\lambda$  must be positive to have an energetically bounded theory.

For  $\mu^2 > 0$ , the scalar field develops a nonzero Vacuum Expectation Value (VEV) which spontaneously breaks the symmetry. Due to the symmetry of  $V(\phi)$  an infinite number of degenerate states exists with the potential  $v$  only depending on the combination of  $\phi^\dagger \phi$  [31] with minimum energy satisfying  $\phi^\dagger \phi = \frac{v^2}{2}$ . This minimum energy requirement reduces one of the four degrees of freedom of the complex scalar field  $\phi$ . A gauge transformation can eliminate the three remaining degrees of freedom. We can choose  $\phi$  by eliminating the upper

component and the imaginary part of the lower component of the complex scalar field as,

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} , \quad (1.35)$$

where a real scalar field, the Higgs field ( $H$ ), emerges as the excitation from the vacuum state. This choice of the minimum spontaneously breaks the gauge symmetry [32].

After substituting  $\phi$  in Equation 1.33, the kinetic term of the Lagrangian takes the form

$$\begin{aligned} \mathcal{L}_{BEH, Kinetic} &= \frac{\lambda}{2} v^4 \\ &+ \frac{1}{2} \partial_\mu H \partial^\mu H - \lambda v^2 H^2 + \frac{\lambda}{\sqrt{2}} v H^3 + \frac{\lambda}{8} H^4 \\ &+ \frac{1}{4} (v + \frac{1}{\sqrt{2}} H)^2 (W_\mu^1 \quad W_\mu^2 \quad W_\mu^3 \quad B_\mu) \begin{pmatrix} g^2 & 0 & 0 & 0 \\ 0 & g^2 & 0 & 0 \\ 0 & 0 & g^2 & gg' \\ 0 & 0 & gg' & g^2 \end{pmatrix} \begin{pmatrix} W^{1\mu} \\ W^{2\mu} \\ W^{3\mu} \\ B^\mu \end{pmatrix}, \end{aligned} \quad (1.36)$$

where the first line is the vacuum energy density, a constant value that can be ignored in the case of QFT. The second line describes the triple and quartic self-interactions of the Higgs field, as well as the quadratic term including the Higgs field, whose coefficient gives the Higgs mass,  $m_H = \sqrt{2\lambda v^2}$ . Further simplifying the last line gives two sets of kinetic terms. The first set consists of interactions of the  $SU(2)$  gauge boson fields with the Higgs field. The second set consists of terms that are quadratic in the gauge boson fields. The coefficient related to the first set of terms gives the strength of Higgs and gauge boson interactions. While the coefficient related to the later set of terms gives the mass eigenvalues for the  $SU(2)$  gauge bosons as,

$$\begin{aligned} m_{W^1}^2 &= m_{W^2}^2 = \frac{1}{4} g^2 v^2 \\ m_{W^3}^2 &= \frac{1}{4} (g^2 + g') v^2 \\ m_B &= 0 \end{aligned} \quad (1.37)$$

As bosons are spin-1 particles, the three now-massive gauge bosons  $W^\pm$  and  $Z$  each

acquire an additional degree of freedom, the *longitudinal* polarization state with their spin aligned perpendicular to the direction of their motion. The initial  $SU(2)_L$  Lagrangian in Equation 1.33 started with a total of twelve degrees of freedom, four related to the complex scalar field  $\phi$  and eight related to the four massless gauge fields. After the Higgs mechanism, the real scalar Higgs field now has one degree of freedom. The other three scalar degrees of freedom have been eaten to give each of the  $W^\pm$  and  $Z$  bosons their mass and the longitudinal polarization component. This phenomenon in the Higgs mechanism is called the Electroweak Symmetry Breaking (EWSB) mechanism. As discussed above, the EWSB mechanism is at the heart of the SM by which the gauge boson gets the mass which also gives rise to the longitudinal polarization of the massive vector bosons. This thesis summarizes a measurement with an experimental sensitivity to a such important property of the theory.

The last remaining piece in the SM is adding the fermion mass to the Lagrangian. A simple Lagrangian with the fermion mass can be written as

$$\mathcal{L}_{mass\ fermion} = -m(\bar{\chi}_L \psi_R + \bar{\psi}_R \chi_L). \quad (1.38)$$

This term violates  $SU(2)_L$  gauge symmetry because the left-handed fermions are doublets, and the right-handed are singlets. After adding a scalar complex field  $\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}$  the Lagrangian takes the form

$$\mathcal{L}_{Yukawa,\ \ell} = \frac{G_\ell v}{\sqrt{2}}(\bar{\chi}_L \psi_R + \bar{\psi}_R \chi_L) - \frac{G_\ell}{\sqrt{2}}(\bar{\chi}_L \psi_R + \bar{\psi}_R \chi_L)H, \quad (1.39)$$

with arbitrary parameters  $G_{\ell=e,\mu,\tau}$ . The constant in the first term  $\frac{G_\ell v}{\sqrt{2}}$  represents the mass of the fermions, whereas the second term gives the interaction of fermions with the Higgs field.

Similarly, the mass terms for quarks follow but including the down-type quarks, the parameters corresponding to  $G_\ell$  are matrices  $G_q^{ij}$  for the quark families  $i, j$  and up-type and

down-type quarks as

$$\mathcal{L}_{Yukawa, Q} = -G_d^{ij}(\bar{u}_i, \bar{d}_i)_L \phi d_{jR} - G_u^{ij}(\bar{u}_i, \bar{d}_i)_L \phi u_{jR} + h.c. \quad (1.40)$$

The final Standard Model Lagrangian is the sum of the QED (Equation 1.9), QCD (Equation 1.19), electroweak interactions including the self-interactions of vector bosons (Equation 1.23), Higgs potential, Higgs self-interactions, and Higgs-vector boson interactions (Equation 1.33), and the Higgs-fermion Yukawa coupling (Equations 1.39 and 1.40), which in compact form is written as Equation 1.1. The final electroweak theory of SM with massive gauge bosons tightly constrains the W, Z, and Higgs interactions, their masses, and the self-couplings of gauge bosons in terms of a few parameters. Any deviations from predictions indicate the presence of physics beyond the SM, thus, losing the mathematical underpinnings of the theory. Therefore, this thesis targets precision measurement of the electroweak sector to detect possible deviations caused by BSM effects.

## 2 Limitations of the Standard Model

Many discoveries have experimentally validated the Standard Model's predictions since the 20<sup>th</sup> century. The breakthrough discovery of the Higgs boson in 2012 at the LHC validated the last piece of the theory [33] [34]. Many predicted parameters, such as production cross-sections and decay branching ratios for several processes, have been measured with high precision. No statistically significant discrepancy from theory has been observed except for the controversial  $W^\pm$  boson mass measurement from the CDF-II Collaboration [35]. Despite the spectacular success of the theory, experimental evidence suggests that the theory is incomplete. SM has the following limitations:

- It fails to provide an explanation for dark matter, whose existence is experimentally supported by astrophysical observations such as galactic rotation curves and gravitational lensing [36].
- The CP violation allowed in the SM cannot explain the amount of matter/anti-matter asymmetry observed in the universe.
- The strengths of the four fundamental forces are different by many orders of magnitude. The hierarchy of such interactions has yet to be understood.
- The simplest formulation of the SM discussed in this thesis assumes the neutrinos to be massless left-handed particles. However, recent experimental results suggest that the neutrino masses must be non-zero to generate the observed neutrino oscillation [37]. There are two possibilities to accommodate the neutrino masses in theory by adding either a Dirac or Majorana mass term, which remains an open question.
- It fails to explain the gravitational force.
- Some recent experimental measurements, such as the measurement of the anomalous magnetic dipole moment of a muon, *Muon g-2* [38], and measurements in *B*-physics [39]

show evidence of deviation from the SM predictions.

These limitations suggest that the SM is an effective field theory, only describing an approximation of our universe, thus, motivating the experimental searches for new physics beyond the Standard Model. Experimentally, there are two ways to look for BSM physics: direct searches and indirect precision measurements. Experimental signatures of BSM-predicted particles, such as their invariant masses, are searched directly by direct searches. This thesis focuses on an indirect approach, where precisely measured SM-predicted differential cross-sections are compared with state-of-the-art theoretical predictions looking for hints of deviation caused by the BSM physics.

### 3 Phenomenology of Proton-Proton Collisions

The main results discussed in this thesis are differential cross-sections for di-Z boson production in association with two jets in a proton-proton collider at the center of mass energy of  $\sqrt{s} = 13$  TeV. The differential cross-section measured gives the production probability of two  $Z$  bosons and two jets as a function of their kinematic properties resulting from  $p$ - $p$  interactions at a given LHC collision energy and luminosity.

Protons are composite particles made up of quarks and gluons. Collisions happen between different constituents of the protons, *partons*. The partons carry only a fraction of the total momentum  $x_i$ ; thus, a partonic cross-section is  $\sqrt{\hat{s}} = \sqrt{s x_1 x_2}$ . Figure 2 schematically shows the di-Z boson production in association with two jets from  $p$ - $p$  collision. The parton interaction that produces the physics of interest ( $ZZ^*jj$ ) with large momentum transfer is *hard scattering*. The additional partons of the two protons that interact in the hard interaction process lead to minor energy deposits in the detector referred to as *underlying events*. Since the  $p$ - $p$  collision happens in bunches, other protons interact, generating additional outgoing particles and energy deposits in the detector, referred to as *pile-up*.

The differential cross-section  $d\sigma$  for two particles is given by:

$$d\sigma = \frac{|\mathcal{M}|^2}{F} dQ, \quad (3.1)$$

where  $F$  is the incident flux, and  $dQ$  represents the Lorentz invariant phase space factor. The scattering amplitude  $\mathcal{M}$  is the matrix element calculated from the Lagrangian density of the SM using a perturbative expansion [40].

The cross-section of a hard scattering process with two initial-state protons  $p_1$  and  $p_2$  producing the final state  $X$  is given by:

$$d\sigma_{p_1 p_2 \rightarrow X} = \int dx_1 dx_2 \sum_{q_1, q_2} f_{q_1}(x_1, \mu_F) f_{q_2}(x_2, \mu_F) d\sigma_{q_1 q_2 \rightarrow X}(x_1, x_2, \mu_F, \mu_R), \quad (3.2)$$

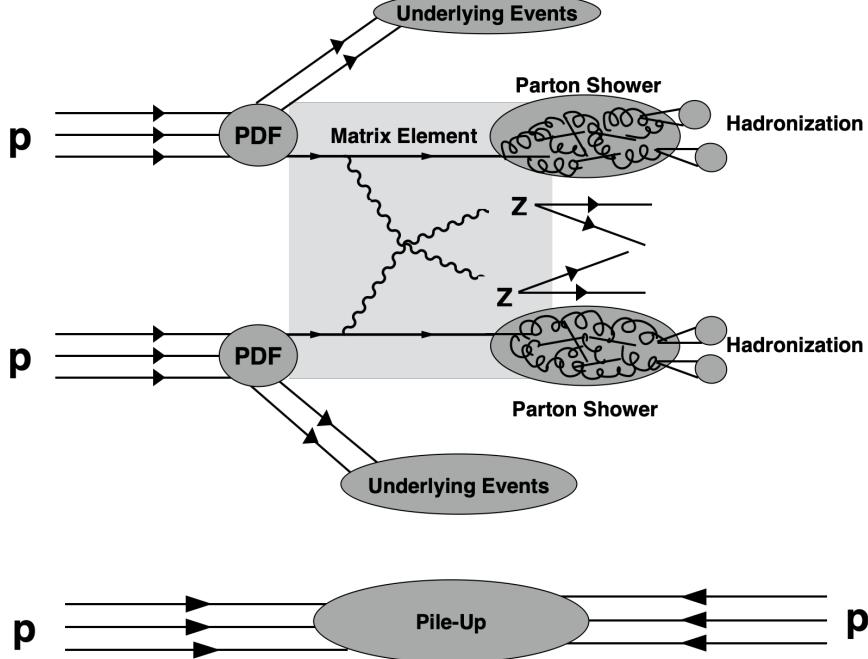


Figure 2: Phenomenology of di- $Z$  boson production in association with two jets in a proton-proton collider.

where  $q_1$  and  $q_2$  are the partons of the protons, and  $d\sigma_{q_1 q_2 \rightarrow X}(x_1, x_2, \mu_F, \mu_R)$  is the partonic cross-section. The functions  $f_{q_1}(x_1, \mu_F)$  and  $f_{q_2}(x_2, \mu_F)$  are the parton distribution functions (PDF) representing the density of the partons  $q$  inside a proton carrying the longitudinal momentum fraction  $x$ .

The partonic cross-section is calculated perturbatively as an expansion in terms of the strong coupling constant  $\alpha_S$  as,

$$d\sigma_{q_1 q_2 \rightarrow X} = \alpha_S^k \sum_{m=0}^n c_m \alpha_S^m. \quad (3.3)$$

The coefficient  $c_m$  depends on the center-of-mass energy, and theoretical calculations usually contain a finite number of coefficients. Leading order (LO) calculations include one term ( $n = 0$ ), whereas next-to-leading order (NLO) and next-to-next-to-leading order (NNLO) contains two ( $n = 1$ ) and three ( $n = 2$ ) terms, respectively. The theoretical calculations relevant to this thesis are generally calculated at NLO.

The higher-order terms in the series contain additional virtual loop<sup>1</sup> corrections and real emissions<sup>2</sup> of quarks and gluons. When calculating scattering amplitudes for processes involving virtual loops beyond the LO, the integration over the loop momentum introduces singularities. These *ultraviolet singularities* are associated with the high momentum particle that participates in the virtual loops. The divergences are controlled via the renormalization procedure, where the singularities are absorbed by the redefinition of the strong coupling constant  $\alpha_S$  to keep the predictions finite. The renormalization process is energy-dependent, and the predicted cross-sections from theoretical calculations depend on an energy-dependent *renormalization scale*  $\mu_R$ . Moreover, additional *infrared singularities* arise from real emissions of soft or collinear gluons. The infrared divergences either cancel out as predicted by the Kinoshita-Lee-Nauenberg theorem [41] or are fixed by introducing the *factorization scale*  $\mu_F$ , where the PDFs and fragmentation functions are redefined. Thus, any finite order prediction of cross-section depends on two energy scales  $\mu_R$  and  $\mu_F$ . The scale dependence is reduced when higher order terms in the perturbative series are included and vanishes when including all perturbative orders. In practice, these scales are fixed to the energy scale of the process being evaluated. The residual scale dependence is taken as uncertainties on the predicted cross-sections, which are estimated by varying the values of  $\mu_R$  and  $\mu_F$  in Monte Carlo simulations.

The PDFs used in Equation 3.2 are determined experimentally using data from deep-inelastic-scattering, jet production, and Drell-Yan events [42] [43]. As shown in Figure 3, a PDF of a parton depends on the reference value of the momentum transfer  $Q_0^2$ . The differences are driven by modifications of partons' momenta resulting from the emission of gluons from quarks and the splitting of gluons to  $q\bar{q}$  pairs. A PDF at any value of  $Q^2$  can be calculated using the PDF at reference scale  $Q_0^2$ . The factorization scale  $\mu_F$  determines

---

<sup>1</sup>virtual loop corrections consists of radiation of gluon from a quark which is absorbed internally by the same or different quark. Thus, creating a loop of radiation with additional QCD vertex. The particles in virtual loop corrections do not appear in the final state.

<sup>2</sup>real emission consists of adding an additional QCD vertex through radiation of quarks or gluons. Particles from real emissions appear in the final state.

the threshold whether the perturbative corrections modify the PDF or are included in the partonic cross-sections  $d\sigma_{q_1 q_2}$  [40].

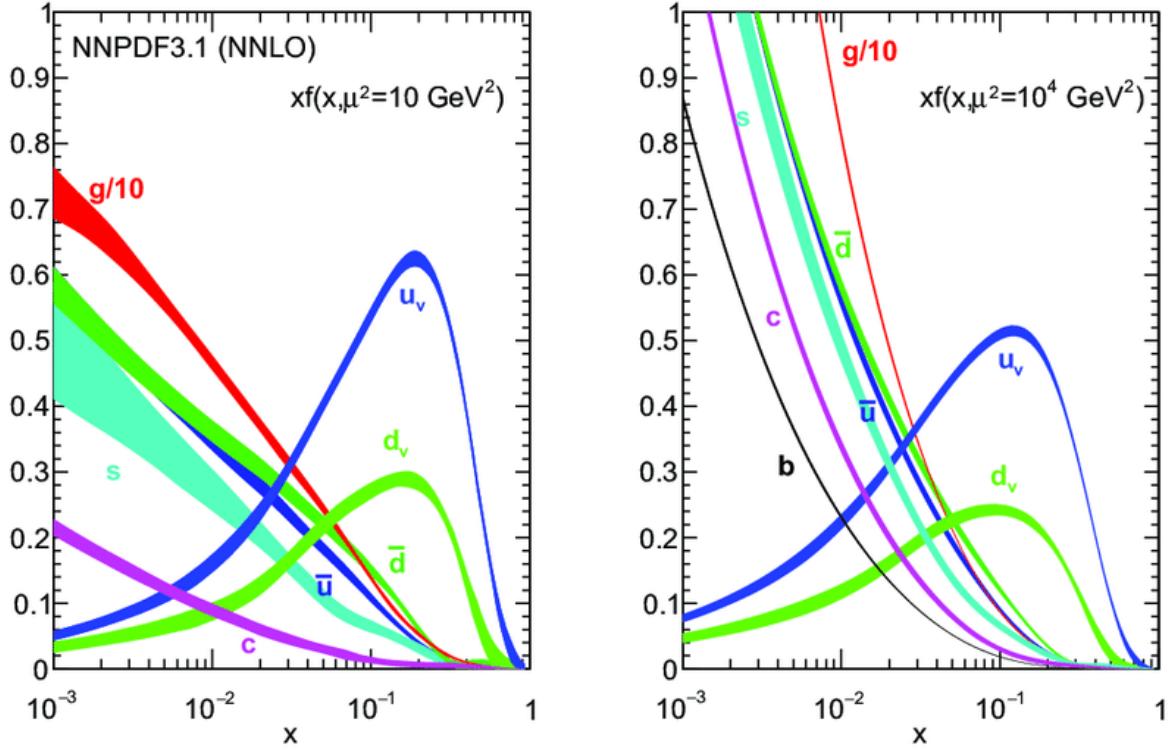


Figure 3: Parton distribution functions  $xf_q(x, Q^2)$  for reference momentum transfer  $Q_0^2 = 10 \text{ GeV}^2$  (left) and  $Q_0^2 = 10^4 \text{ GeV}^2$  (right) [4].

Any particles with a color charge involved in the interaction or produced during the hard scattering radiate gluons, which further emit QCD radiation forming showers of color particles, also known as *parton shower*. During parton showering, the energy of each parton is split among the radiated particles. Below an energy scale named pole of the QCD running coupling ( $\lambda_{QCD}$ ), the bounding potential of the strong force intervenes, and the partons are bound into a colorless state of stable and unstable hadrons. This process is named *hadronization* and leads to the formation of collimated sprays of charged and neutral hadrons in the detector called *jets*. The matrix element generating the hard-scattering process can describe a few hard QCD emissions. However, dedicated parton showering algorithms are used to

describe multiple QCD emissions. The parton shower represents an approximate perturbative treatment of higher-order QCD corrections. The parton showering algorithm calculates dominant contributions associated with soft or collinear parton splitting. The hadronization process is simulated using non-perturbative models. There are two phenomenological models for hadronization, string and cluster models. String models are based on the assumption that the potential energy between two quarks increases linearly as their spatial separation increases. The cluster model is motivated by considering central objects, hadrons, as a color-neutral cluster of quarks.

The theoretical predictions of an event shown in Figure 2 are calculated using Monte Carlo (MC) simulations which include matrix element calculations for hard scattering, the parton showering, the effect of the underlying events, hadronizations, and pile-up. A comprehensive overview of the methods used in MC simulation is discussed in Ref [44].

## 4 Electroweak Diboson Physics

At the LHC, two types of physics processes contribute to the production of di- $Z$  bosons in association with two jets [ $ZZ^*(\rightarrow 4\ell)jj$ ]: the QCD production at the order  $\alpha_S^{>2}\alpha_{EWK}^4$  and the EWK production at order  $\alpha_{EWK}^{>6}$  [45]. Figures 5 and 6 show the Feynman diagrams at leading order for parton-initiated and gluon loop-initiated QCD  $ZZ^*(\rightarrow 4\ell)jj$  processes, respectively, whereas Figure 7 shows the Feynman diagrams at leading order for the EWK  $ZZ^*(\rightarrow 4\ell)jj$  production [46]. The EWK production consists of two sets of interactions: the vector boson scattering processes involving either triple (Figure 7a) or quartic (Figure 7b) self-interactions of the gauge-bosons, and the processes featuring the Higgs bosons (Figure 7c and 7d).

Figure 4a shows the cross-sections as a function of the center of mass energies ( $\sqrt{s}$ ) for the longitudinally polarized vector boson scattering, corresponding to the self-interactions of the gauge-bosons. The longitudinal cross-sections grow quadratically at high values of  $\sqrt{s}$ , eventually violating the unitarity bounds. The unitarity violation is observed in cross-sections for longitudinal vector boson scattering in all  $V_L V_L \rightarrow V_L V_L$  processes. Figure 4b shows the longitudinal cross-sections for the electroweak  $V_L V_L \rightarrow V_L V_L$  processes, including the amplitudes from the vector bosons self-interactions and  $H \rightarrow V_L V_L$  processes. As shown, the precise interference between the Higgs-featured and VBS processes restores the unitarity at high energies for all multi-boson final states [47].

As discussed in Section 1.3.5, the massive  $W$  and  $Z$  bosons get their masses via the BEH mechanism through EWSB. As a consequence of EWSB, the  $W$  and  $Z$  bosons acquire an additional degree of freedom (the longitudinal polarization mode) whose scattering interfere with the Higgs-featured processes. Thus, measuring the cross-sections of electroweak production of the di- $Z$  bosons in association with two jets provides a direct probe of the EWSB, which is at the heart of the SM [45]. As the unitarity is restored at high energies, the cross-section of the electroweak  $ZZ^*(\rightarrow 4\ell)jj$  process is sensitive to possible BSM modifications

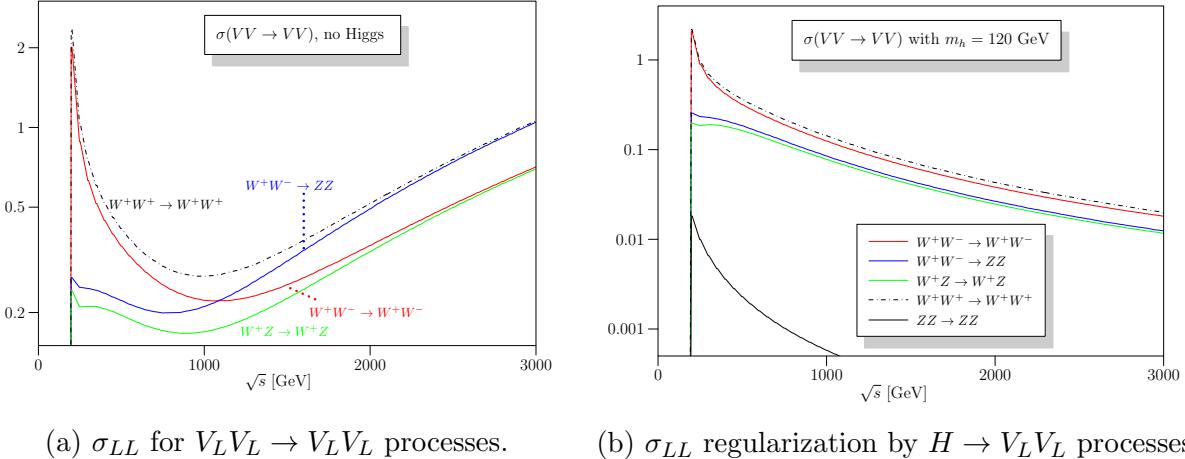


Figure 4: Cross-sections for longitudinal component of the electroweak  $VVjj$  production as a function of  $\sqrt{s}$  [5]. These calculations were made before the discovery of the SM Higgs; thus, the calculations in the second figure use an SM-like Higgs with 120 GeV mass. For the SM Higgs with a mass of 125.10 GeV, the regularization is similar in shape.

at high energies. Therefore, measuring the cross-sections of the electroweak  $ZZ^*(\rightarrow 4\ell)jj$  processes differentially as a function of kinematically sensitive observables is essential.

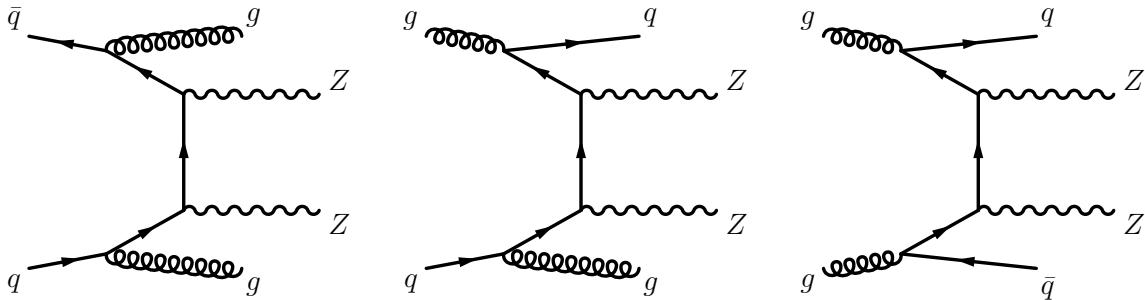


Figure 5: Typical diagrams of LO  $qq$  and  $gg$  induced QCD  $\alpha_S^2 \alpha_{EWK}^2$  production of  $ZZ^*jj$ . The two  $Z \rightarrow \ell\ell$  vertices each contribute an additional electroweak coupling of  $\alpha_{EWK}$ .

The triple and quartic self-interactions of the gauge bosons arise from the square of the non-Abelian structure of  $SU(2)$  in the kinetic term  $\frac{1}{4}W_{\mu\nu}^a W_a^{\mu\nu}$  of the EWK Lagrangian in Equation 1.23. Implementing the values of the field strength tensor  $W_{\mu\nu}^a$  from Equation 1.25, the relations of  $W_\mu^\pm$  fields in Equation 1.26, and the relations of neutral gauge fields in

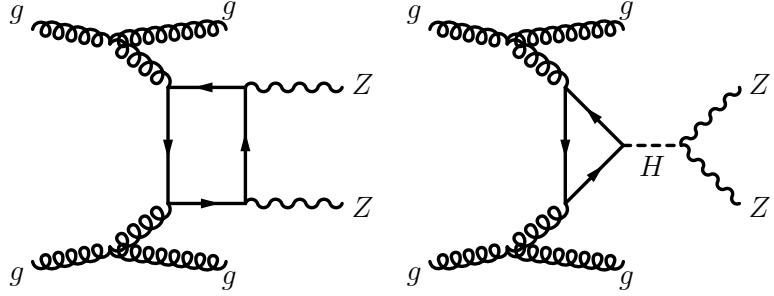


Figure 6: Typical diagrams for LO  $gg$  loop induced the QCD  $\alpha_S^4 \alpha_{EWK}^2$  production of  $ZZ^*jj$ . The two  $Z \rightarrow \ell\ell$  vertices each contribute an additional electroweak coupling of  $\alpha_{EWK}$ .

Equation 1.29, the triple and quartic self interaction terms become,

$$\mathcal{L}_3 = ie_{V=\gamma,Z}[W_{\mu\nu}^+ W^{-\mu} V^\nu - W_{\mu\nu}^- W^{+\mu} V^\nu + W_\mu^+ W_\nu^- V^{\mu\nu}], \quad (4.1)$$

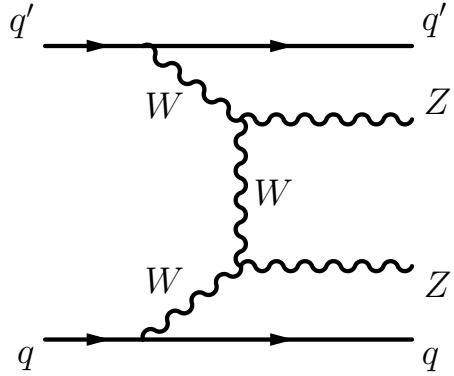
$$\begin{aligned} \mathcal{L}_4 = & e_W^2 [W_\mu^- W^{+\mu} W_\nu^- W^{+\nu} - W_\mu^- W^{-\mu} W_\nu^+ W^{+\nu}] \\ & + e_{V=\gamma,Z}^2 [W_\mu^- W^{+\mu} V_\nu V^\nu - W_\mu^- V^\mu W_\nu^+ Z^\nu] \\ & + e_\gamma e_Z [2W_\mu^- W^{+\mu} Z_\nu A^\nu - W_\mu^- Z^\mu W_\nu^+ A^\nu - W_\mu^- A^\mu W_\nu^+ Z^\nu], \end{aligned} \quad (4.2)$$

where  $e_\gamma = g \sin \theta_W$ ;  $e_W = \frac{e_\gamma}{2\sqrt{2} \sin \theta_W}$  and  $e_Z = e_\gamma \cot \theta_W$  are the precise coupling strengths for vector boson self-interaction. Both triple and quartic neutral couplings, such as  $ZZZ$  or  $ZZZZ$  are absent in the SM.

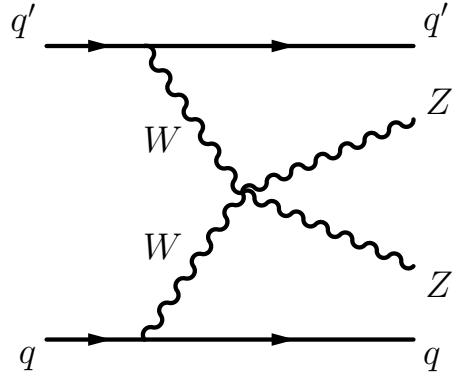
Similarly, the couplings of Higgs to vector bosons are also predicted precisely by the BEH mechanism in Equation 1.36 as:

$$\mathcal{L}_{HVV} = \frac{m_W^2}{v^2} W_\mu^+ W^{-\mu} h^2 + \frac{m_Z^2}{v^2} Z_\mu Z^\mu h^2. \quad (4.3)$$

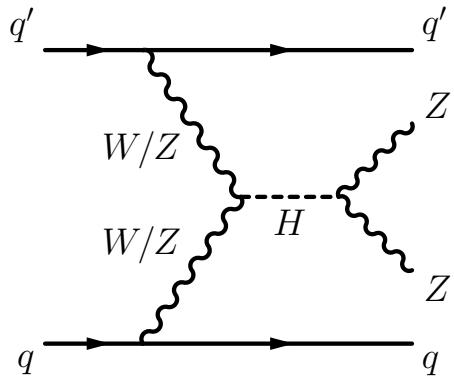
The EWK production of  $ZZ^*(\rightarrow 4\ell)jj$  is extremely sensitive to any possible anomalous triple gauge couplings (aTGC), anomalous quartic gauge couplings (aQGC), or anomalous Higgs to vector boson couplings [48] [49] [50]. Therefore, it is imperative to probe the high energy behavior of the EWK production of  $ZZ^*(\rightarrow 4\ell)jj$  to seek possible deviations from



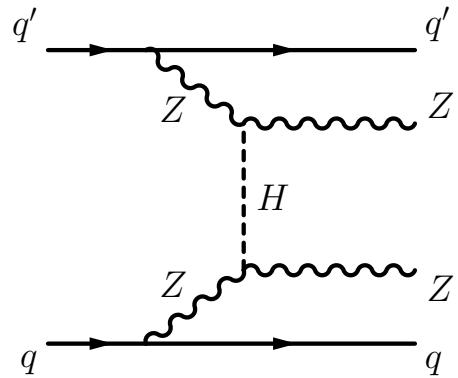
(a) ZZjj production with two triple gauge coupling (TGC) vertices.



(b) ZZjj production with a quartic gauge coupling (QGC) vertex.



(c) s-channel Higgs ZZjj Production.



(d) t-channel Higgs ZZjj Production.

Figure 7: Feynman diagrams at LO for the EWK  $\alpha_{EWK}^4$  production of  $ZZ^*jj$ . The two  $Z \rightarrow \ell\ell$  vertices each contribute an additional electroweak coupling of  $\alpha_{EWK}$ .

the physics processes Beyond the Standard Model.

The EWK  $ZZ^*(\rightarrow 4\ell)jj$  production with each  $Z$  boson decaying to a pair of SF-OC lepton pairs is an extremely rare process. Thus, the QCD background processes dominate the  $ZZ^*(\rightarrow 4\ell)jj$  events recorded by the ATLAS detector during Run-2 [19]. The electroweak production of  $ZZ^*jj$  shown in Figure 7 has some characteristic kinematic properties. In each electroweak process, the initial state quark radiates a vector boson which scatters either by self-interactions or Higgs and produces the two final state  $Z$ -bosons, each of which decays to the SF-OC leptons. The initial state partons continue in almost the same direction as their parent protons, forming jets on the opposite side of the ATLAS detector with a large spatial gap and no additional hadronic activity from the hard scattering between the two jets [51]. As the final state jets originate directly from the initial partons in the hard interaction, they are highly energetic, resulting in two jets with high momenta and a significant invariant mass [51]. Due to momentum conservation, the two  $Z$ -bosons are produced centrally with respect to the dijet. The decay of the two  $Z$  bosons into SF-OC muons or electrons defines the final signature of the VBS- $ZZ^*(\rightarrow 4\ell)jj$ -like event.

The previous ATLAS measurement of  $ZZ(\rightarrow 4\ell)jj$  used a multi-variate analysis exploiting these kinematic properties to measure the fiducial cross-sections for purely electroweak  $ZZjj$  process [19]. Due to limited statistics in the final state, it is impossible to measure differential cross-sections for the purely electroweak  $ZZ^*(\rightarrow 4\ell)jj$  process using the LHC Run-2 dataset. Therefore, a VBS-Enhanced phase space, which includes a high fraction of events from the EWK production based on their unique kinematic properties, is defined. The differential cross-sections presented in this thesis are measured in this electroweak-enhanced phase space.

## Chapter III: Experimental Setup

The European Organization for Nuclear Research, CERN, in Geneva, Switzerland, is home to the world’s largest particle accelerator, the Large Hadron Collider. The measurements presented in this thesis correspond to the processes at the frontier of high-energy collisions. The collisions at the relevant high-energy scale are only possible through large particle accelerators, which reduces the energy loss through synchrotron radiation. The LHC detectors are large in size to effectively measure and stop the high energy particles from collisions. There are currently eight experiments analyzing the data from the LHC, among which ATLAS and CMS are the two largest multipurpose experiments. They analyze the collected data to perform SM precision measurements and direct searches for new physics. This thesis analyzes data collected by the ATLAS experiment between 2015-2018.

This chapter gives a description of the LHC in Section 5, the ATLAS experiment in Section 6, details on physics object reconstruction in Section 7, and plans for future upgrades in Section 8.

## 5 The Large Hadron Collider

The LHC was designed to collide hadrons, either protons or heavy ions, at instantaneous luminosity up to  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  and center-of-mass energies ( $\sqrt{s}$ ) up to 14 TeV [52]. The LHC, the most powerful particle accelerator, is the successor of the Large Electron Positron Collider (LEP), which collided electrons and positrons at  $\sqrt{s}$  up to 250 GeV [53]. The LHC reuses the same tunnel system as LEP with a circumference of 27 km built 50 to 175 meters underground at the French-Swiss border outside Geneva, Switzerland.

The first run of the LHC, Run-1, started in 2010 at  $\sqrt{s} = 7 \text{ TeV}$ , which was later increased to 8 TeV. Run-1 lasted until 2013, after which LHC was shut down for two years of planned upgrades to enhance the accelerator's and the detector's performances. LHC resumed its operation for Run-2 from 2015 to 2018 at the  $\sqrt{s} = 13 \text{ TeV}$ . The data collected during the Run-2 period is analyzed for this thesis.

Accelerating the proton to the desired center-of-mass energy is a multi-step process shown schematically in Figure 8. First, the protons are created from hydrogen gas by removing the electrons through ionization with an intense electric field. A proton beam is then formed in the LINAC-2 linear accelerator and injected into the circular PS Booster, which accelerates the 50 MeV proton to energies of 1.4 GeV [54]. The beams are then injected into the Proton Synchrotron (PS), accelerating them to energies of 25 GeV [54]. The proton beams are then injected into the Super Proton Synchrotron (SPS) to further accelerate at 450 GeV energies and injected into the main LHC rings. The two opposite proton beams reach the desired final energy of 6.5 TeV using radio frequency (RF) acceleration cavities [54]. The accelerated proton beams are maintained for several hours of data-taking.

The final RF accelerated proton beams are in evenly-spaced discrete bunches, each consisting of about  $10^{11}$  protons. The bunches are separated by 25 ns spacing [54]. The proton beams are guided in the tunnel by a magnetic field using superconducting dipole and quadruplet magnets. The main LHC ring comprises 1232 dipole magnets that provide a

## CERN's Accelerator Complex

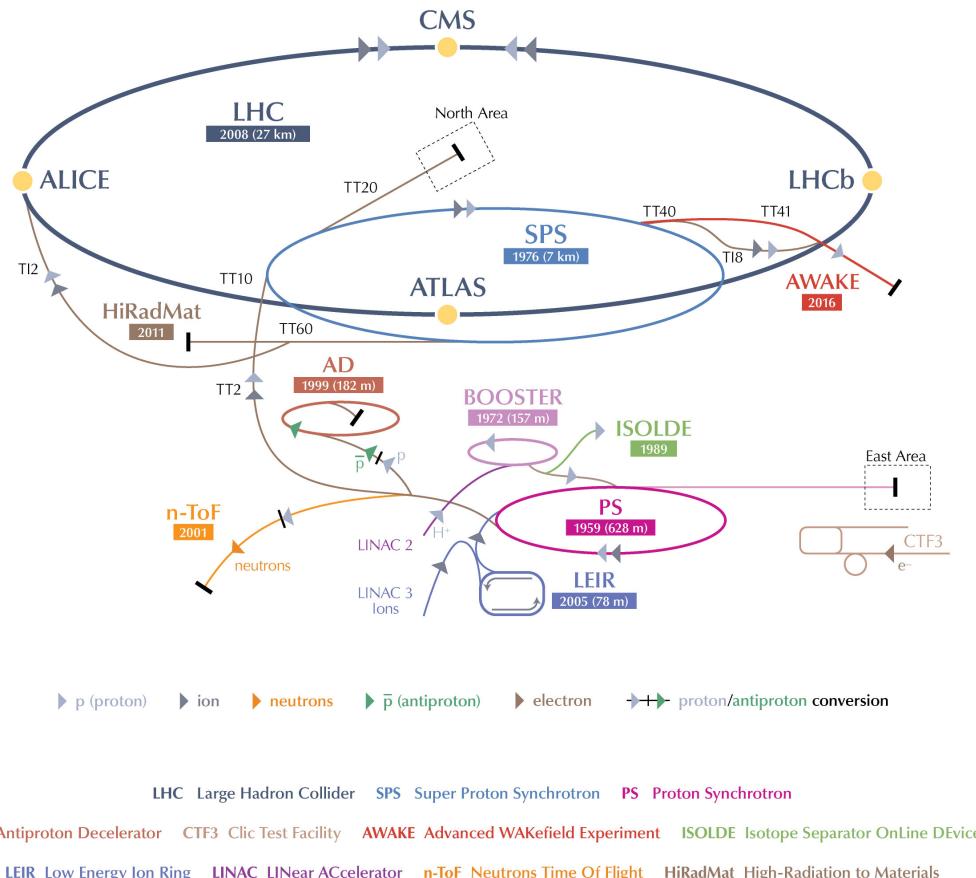


Figure 8: A detailed layout of multiple-steps that goes into proton acceleration before entering the main LHC ring [6].

strong magnetic field of 8 T to bend the beams and about 392 quadruplet magnets to focus the beams in the transverse plane [54]. The superconducting magnets are cooled down to 1.9 K, which requires two vacuum systems to hold the cryomagnets and the helium distribution lines. To avoid unnecessary interactions, the beams are accelerated and maintained in an ultra-high vacuum of  $10^{-13}$  atm [54].

The two opposite beam lines meet in four interaction points, creating proton-proton collisions. The collisions are recorded by the LHC’s four main detectors: ATLAS, CMS, LHCb, and ALICE. ALICE is designed to investigate the heavy-ion collisions and the quark-gluon plasma, whereas the LHCb experiment is designed to study flavor physics [54].

## 6 ATLAS Detector

A Toroidal LHC ApparatuS (ATLAS) is a general-purpose detector of LHC that detects events from proton-proton, and heavy ion collisions [7]. It is 44 meters long and 25 meters wide cylindrical-shaped detector built around LHC Interaction Point 1 [7]. ATLAS has multiple concentric sub-detectors layered around the beamline, providing forward-backward symmetric coverage. The two proton beams collide at the center of the detector producing outgoing particles from hard scattering, underlying events, and pile-up. The outgoing particles interact with the detector material leaving tracks and energy deposits in several layers of the sub-detectors.

The sub-detector closest to the beamline is the *Inner Detector (ID)*, which measures the trajectories of the charged particle and plays a crucial role in identifying the physical position of the hard-scattering, also known as the *primary or hard-scattering vertex*. The ID is surrounded by a solenoid magnet that provides a 2 T magnetic field to bend the particle trajectories for momentum measurements [7]. Outside the solenoid lies the *electromagnetic calorimeter (ECAL)* and then the *hadronic calorimeter (HCAL)*, which measure the energy of electromagnetic and hadronic physics objects, respectively. The outermost layer of the ATLAS detector is the *Muon Spectrometer (MS)* that identifies the muons and provides a secondary measure of their momentum. The MS is embedded inside a toroidal magnetic field that provides a magnetic field up to 3.5 T [7]. Figure 9 shows a schematic of the ATLAS detector with all its sub-detectors.

### 6.1 ATLAS Coordinate System

ATLAS measurements use a right-handed coordinate system with the nominal interaction point as the origin. The beamline is along the cylindrical symmetry axis of the detector, which defines the longitudinal  $z$ -axis. The transverse  $xy$ -plane is perpendicular to the beam direction, where  $x$ -axis points to the center of the LHC ring and  $y$ -axis points upwards

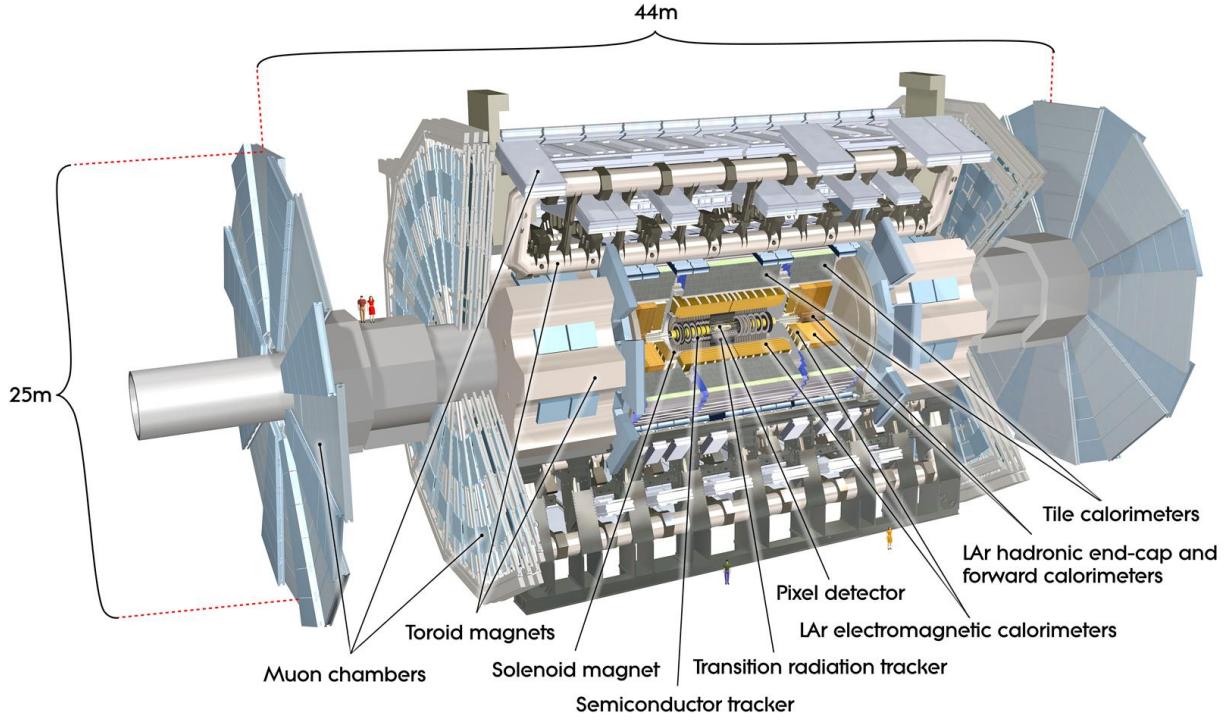


Figure 9: A detailed schematic of the ATLAS detector with all its sub-detectors [7].

towards the surface. Figure 10 shows a schematic of the ATLAS coordinate system. The angle measured around the beamline in  $xy$ -plane gives the azimuthal angle  $\phi$ , whereas the angle measured with respect to the  $z$ -axis gives the polar angle  $\theta$ . Transverse momentum ( $p_T$ ) is particle's momentum in the  $xy$ -plane, defined as

$$p_T = \sqrt{p_x^2 + p_y^2} = p \sin \theta. \quad (6.1)$$

*Rapidity* ( $y$ ) defined in terms of a particle's energy ( $E$ ) and momentum ( $p$ ) is a commonly used collider physics quantity defined as,

$$y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right). \quad (6.2)$$

Particles with larger momentum along the  $z$ -axis have larger values of rapidity, whereas particles with larger momentum values in the transverse plane have smaller values of rapidity.

For particles with negligible mass, the rapidity approaches a purely angular variable called *pseudorapidity* ( $\eta$ ) defined as

$$\eta = \frac{1}{2} \ln \left( \frac{|\vec{p}| + p_z}{|\vec{p}| - p_z} \right) = -\ln \left[ \tan \left( \frac{\theta}{2} \right) \right]. \quad (6.3)$$

Higher values of rapidity and pseudorapidity refer to the forward region of the detector. The ATLAS detector has full coverage in  $\phi$  and coverage up to  $|\eta| < 4.9$ , corresponding to  $1.3^\circ < \theta < 178.7^\circ$  [7].

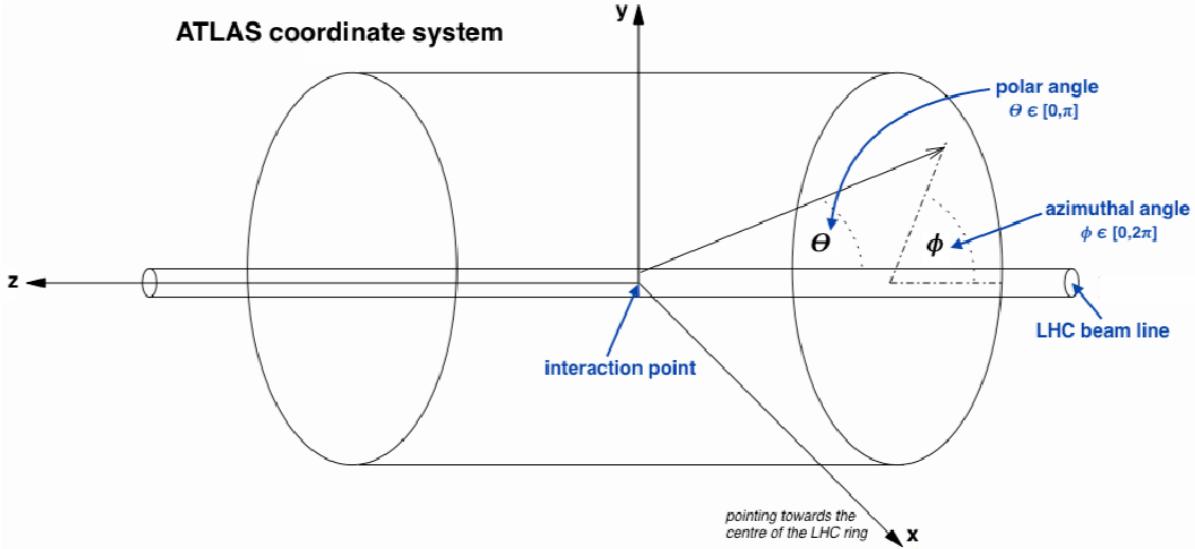


Figure 10: A schematic of the right-handed ATLAS coordinate system [8].

## 6.2 Inner Detector

The inner detector is the innermost sub-detector of ATLAS and is responsible for tracking charged particles' trajectories and identifying the hard scattering vertex. Closest to the interaction point is the Insertable B-Layer (IBL) [55], which was installed between Run-1 and Run-2 to improve tracking at higher pile-up. The IBL is highly granular, consisting of roughly 12 million silicon pixel sensors with a size of  $50 \times 250 \mu m^2$  [55]. Located 3.3 cm from the beamline, the IBL can reconstruct tracks with  $|\eta| < 2.5$  [55].

Three layers of silicon-pixel detectors for a total of 1,744 pixel sensors, each with pixels of size  $50 \times 400 \mu\text{m}^2$ , surround the IBL [56]. The slightly larger pixel size compared to the IBL is optimal for the pile-up at a distance larger than 5 cm from the interaction point. The pixel detector provides coverage up to  $|\eta| < 2.5$  with a single point spatial resolution between 5 and 12  $\mu\text{m}$  [56]. Surrounding the pixel detector is the Semiconductor Tracker (SCT) consisting of five layers of silicon microstrip detectors with a mean strip pitch of 80  $\mu\text{m}$  in the barrel region, and varying pitch of 57 – 94  $\mu\text{m}$  in the end-cap regions [57].

At a distance about 50 cm from the beamline lies the outermost layer of the ATLAS inner detector, the Transition Radiation Tracker (TRT), with 370,000 straw tubes each with a diameter of 4 mm [58]. Each TRT straw tube is filled with an Argon-based gas mixture and has a 31  $\mu\text{m}$  diameter tungsten wires [58].

Figure 11 shows the different parts of the inner detector and their distances from the interaction point.

### 6.3 Calorimeters

ATLAS has two calorimeters, electromagnetic and hadronic, designed to measure the energy of charged and neutral particles up to  $|\eta| \leq 4.9$  [7]. When interacting with a material, an electron loses its energy by photon emission, producing a pair of oppositely charged electrons, which could each radiate a photon, creating an electromagnetic shower in the detector. Similarly, in the hadronic calorimeters, the hadronic particles also result in a shower of particles through multiple scattering. The calorimeters measure the energy of the particles by reconstructing the electromagnetic and hadronic showers. The calorimeters are designed to fully absorb the shower of all particles except muons and neutrinos. Thus, materials with high radiation length ( $X_0$ ) and high interaction length ( $\lambda$ ) are chosen respectively to build the electromagnetic and hadronic calorimeters.

The accordion-shaped electromagnetic calorimeter lies outside the solenoid surrounding the ID consisting of an alternate layer of lead absorber plates and highly granular liquid-

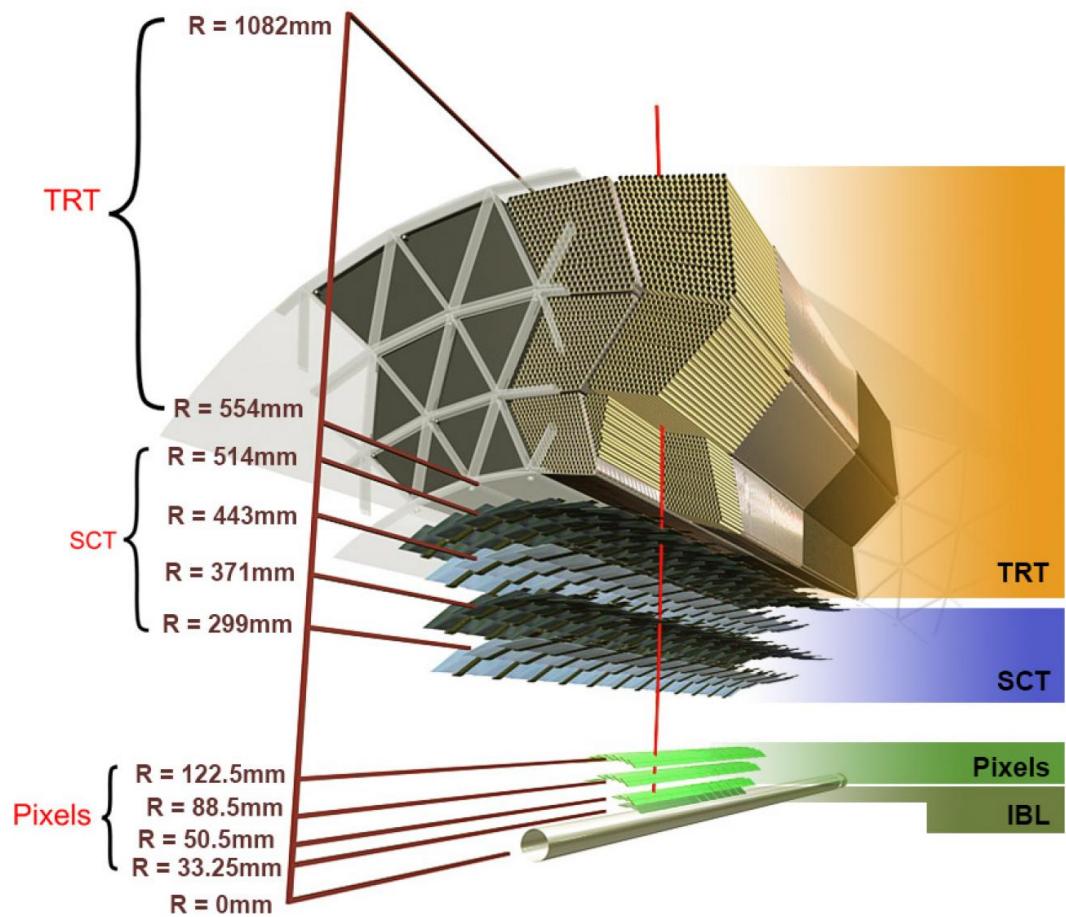


Figure 11: A schematic of the inner detector of ATLAS showing the IBL, pixel detectors, SCT, and TRT [9].

argon (LAr) cells to precisely measure the energies of electrons and photons. It comprise of one barrel section in  $|\eta| < 1.475$  region and two end-caps in  $1.375 < |\eta| < 3.2$  region [59]. The calorimeter's central region ( $|\eta| < 2.5$ ) is designed to identify electrons and photons with high precision.

The hadronic calorimeter surrounds the ECAL and consists of steel absorbers and active scintillator tiles in  $|\eta| < 1.7$  region. The end-cap regions ( $1.5 < |\eta| < 3.2$ ) consist of copper absorbers and active LAr detectors. The forward region ( $3.2 < |\eta| < 4.9$ ) comprises the tungsten absorber followed by active LAr detectors [60].

Figure 12 shows the layout of the ATLAS calorimeters. The calorimeters are segmented into cells which consist of an alternating layer of passive materials to absorb the particle and active layers to measure the energy. Object reconstruction in calorimeters starts with the formation of *topological clusters* of cells with significant energy deposits.

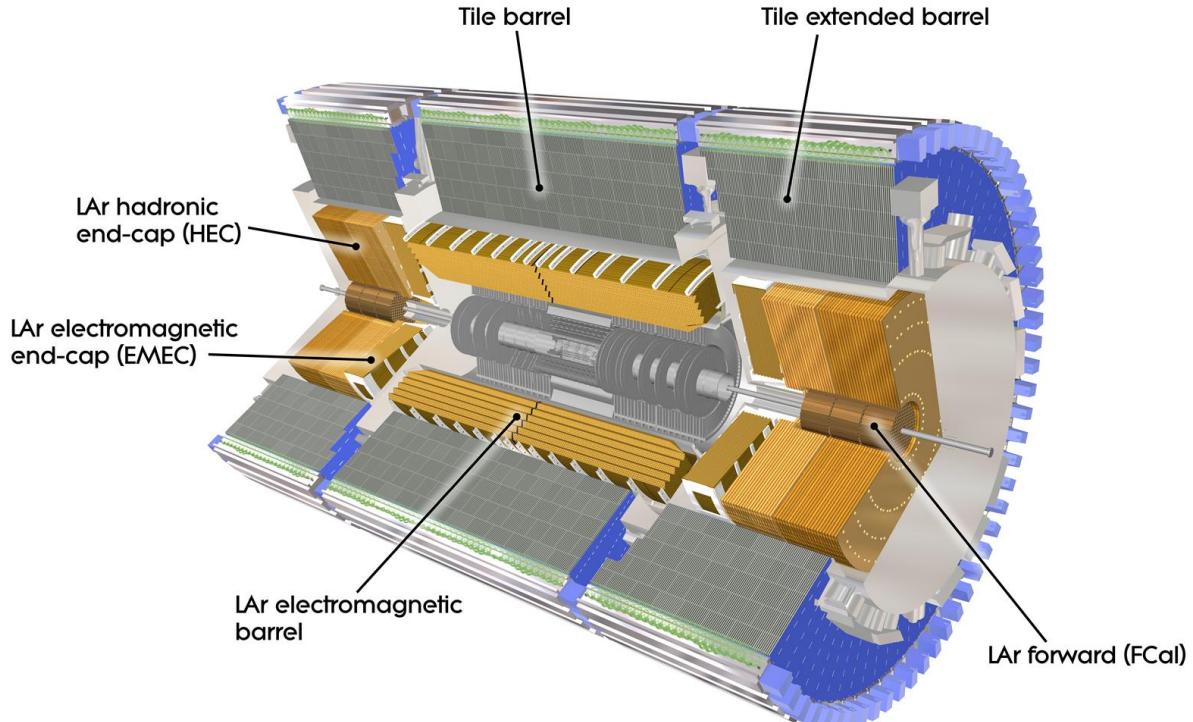


Figure 12: A schematic of the electromagnetic and hadronic calorimeters in ATLAS [7].

## 6.4 Muon Spectrometer

Muons are deeply penetrating charged particles that leave minimum ionizing deposits in the calorimeter. The muon spectrometer, the outermost part of the ATLAS detector, identifies the muons and gives an additional measure of muon's momentum by tracking its trajectories which are deflected in the magnetic field provided by the superconducting toroidal magnets [7]. The MS tracks muon with  $p_T > 3$  GeV in  $|\eta| < 2.7$  range [7]. As shown in Figure 13, the muon spectrometer comprises four types of detectors. The three stations of Monitored Drift Tubes (MDT) occupy the  $|\eta| < 2.0$  region, while the Cathode Strip Chambers (CSC) cover the  $2.0 < |\eta| < 2.7$  region [7]. Resistive Plate Chambers (RPC) ( $|\eta| < 1.05$ ) and the Thin-gap Chambers (TGC) in  $|\eta| = 1.05$  range comprises the trigger system in MS [7].

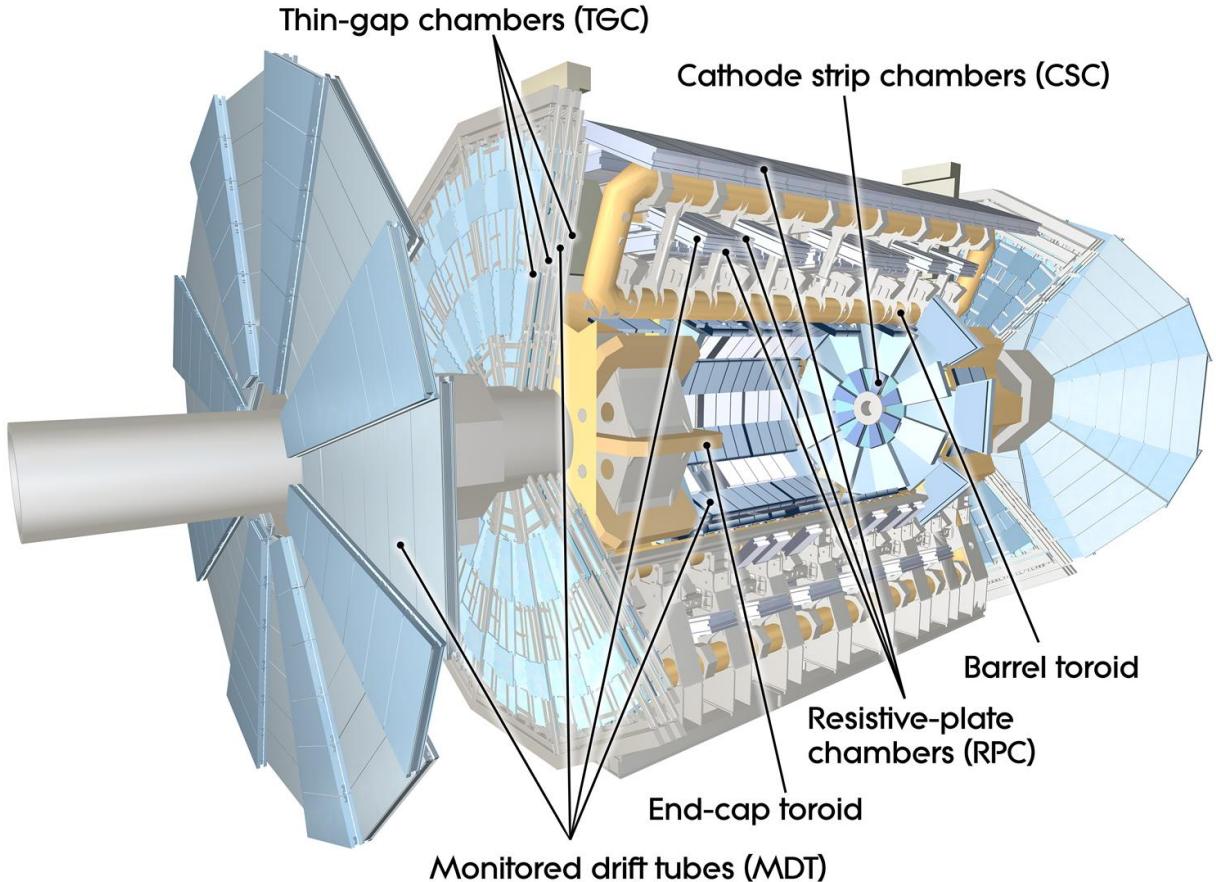


Figure 13: A schematic of the different components of the muon spectrometer in ATLAS [7].

## 7 Physics Object Reconstruction

Different particles leave unique signatures in different sub-detectors of ATLAS. Figure 14 shows a simplified representation of various particles passing through different sub-detectors and leaving various signatures. Physics object reconstruction is the process of interpreting these signals to extract meaningful information about each particle. This section discusses the detail of reconstruction relevant to this thesis.

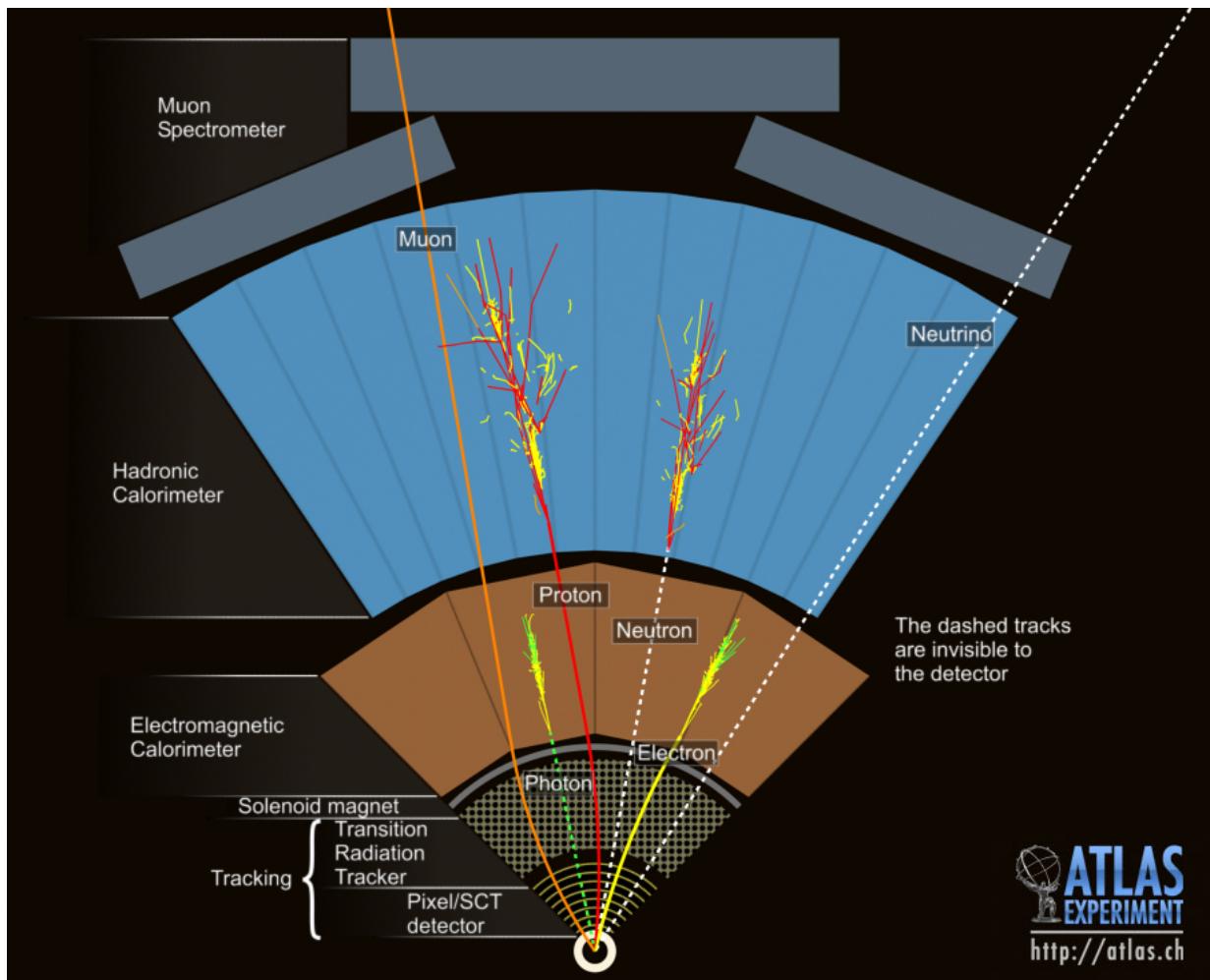


Figure 14: Simplified representation of various particles traversing through the various layers of the ATLAS sub-detectors and leaving unique signatures [10].

## 7.1 Trigger

The first step of particle and event reconstruction is to select interesting high-energy events from a pool of lower-energy signals from minimally scattering protons. The 25 ns bunch crossing frequency yields a large amount of data making it impossible to store all events. ATLAS trigger system identifies the events interesting for physics measurements to store permanently.

The ATLAS trigger consists of two levels: a Level 1 (L1) trigger integrated into the hardware and a high-level software trigger (HLT) [61]. The L1 trigger is based on custom-built electronics, which uses signals from the calorimeters and the muon trigger system (TGC and RPC) to identify event features such as electrons, photons, jets, taus, and missing energy. The L1 trigger reduces the 40 MHz incoming collision data rate (corresponding to 25 ns bunch crossing) by 400 to a more manageable 100 kHz output [61]. The events accepted by the L1 trigger define regions of interest (ROI), and HLT algorithms are run on these events to select ones with candidate physics objects passing dedicated kinematic requirements. The software-based HLT trigger further reduces the data rate by almost a factor of 100 to 1.5 kHz [7]. With the combination of the L1 and HLT trigger system, the data rate is reduced by 400,000, and the selected events with a data readout of 1.5 GB per second are stored permanently. Figure 15 shows the schematic of ATLAS’s trigger and data acquisition system.

The physics object reconstruction discussed below converts the raw data output stored in permanent storage to physics objects used in physics analyses.

## 7.2 Tracks and Vertex Reconstruction

Tracking charged particles is a critical step in reconstruction. Tracks encode charged particles’ momentum and trajectory, playing an essential role in particle identification and primary vertex reconstruction. As the inner detector is closest to the beamline and comprises minimally ionizing detector material with high granularity, it plays the main role in track

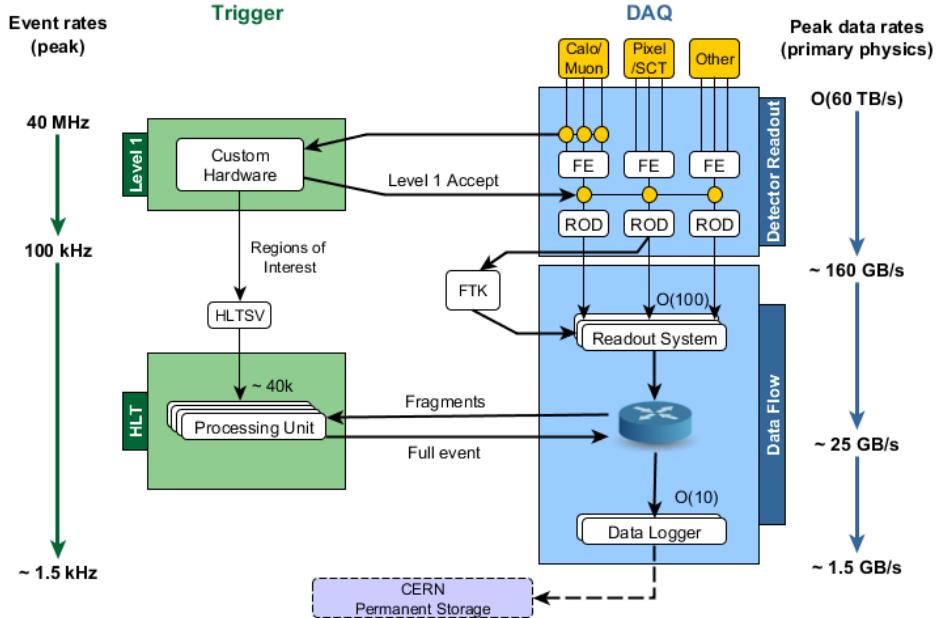


Figure 15: Trigger and data acquisition system in ATLAS [11].

reconstruction. A charged particle passing through different layers of ID leaves a signal via ionization. As the ID solenoidal field is homogenous, the resulting trajectory is circular in the  $xy$  plane. Five parameters shown in Figure 16 define charged particle tracks,

- $q/p_T$ : the ratio of charge and transverse momentum defining the curvature
- $d_0$ : the distance of the closest approach to the primary vertex in  $xy$ -plane defining the transverse impact parameter
- $z_0$ : the longitudinal impact parameter along the  $z$ -axis
- $\phi_0$ : the azimuthal angle
- $\theta_0$ : the polar angle of the particle direction at the closest point of approach [62].

As shown in Figure 17, track reconstruction used in Run-2 consists of two different approaches, the primary *inside-out* approach and the secondary *outside-in* approach [62]. The first step in the inside-out track reconstruction is the space point and drift circle formation,

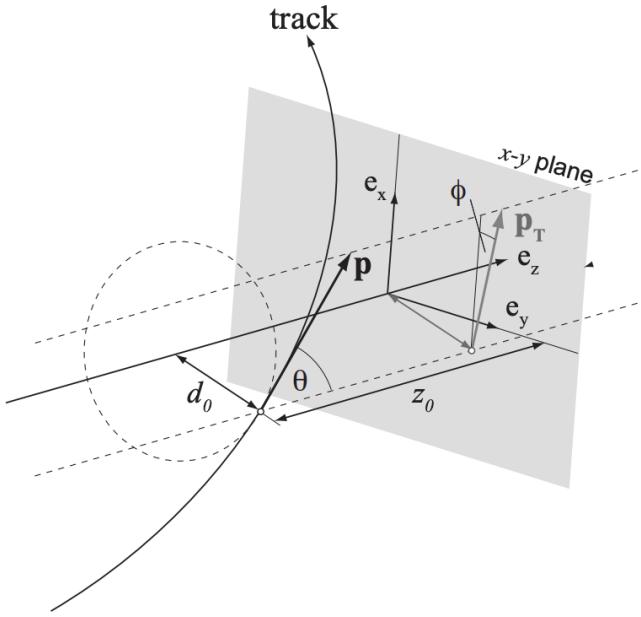


Figure 16: Schematic showing the five-track parameters [12].

formed respectively, by the clusters of signals from the silicon detectors and drift-circles hits in the TRT. Second, track seeds are formed from a collection of three silicon-detector space points and extrapolated to the outer layers by including the compatible clusters in the track trajectory. Once the track is formed, an ambiguity resolution algorithm is applied to reassign shared clusters to the track with a better match, and the final track candidate is fitted using a global  $\chi^2$  method. The last step of inside-out track reconstruction is adding compatible TRT drift holes and refitting the tracks.

The inside-out method is optimal for charged particles originating from the primary vertex. However, the outside-in approach is needed for charged particles generated away from the primary vertex, such as electrons from reconverted photons or pions from  $K_S$  decays. In this approach, the track pattern recognition starts at the TRT in the regions of interest flagged by the electromagnetic calorimeter and backtracks to the silicon detectors.

Tracks of the charged particle are extrapolated inward to the beamline and are assigned to vertices [63]. In most ATLAS analyses, including the one presented in this thesis, the

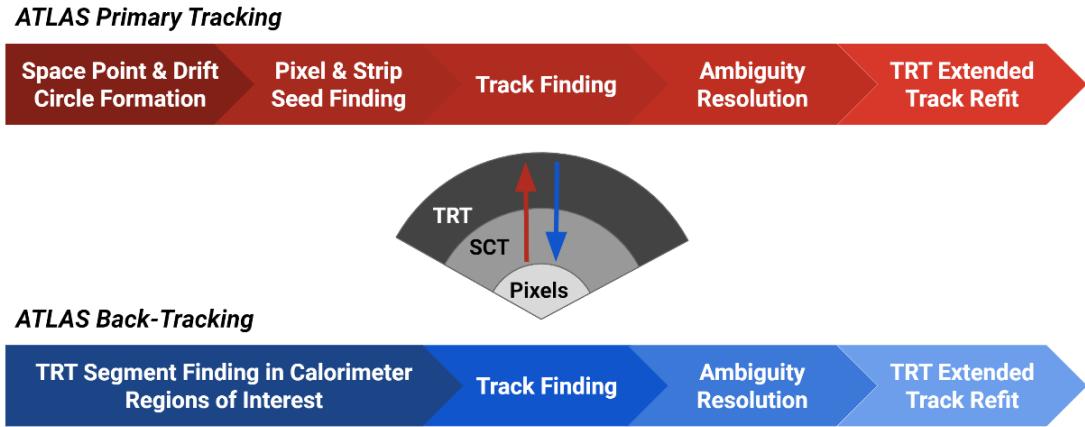


Figure 17: Schematic showing the two techniques of track reconstruction, primary inside-out and secondary outside-in [13].

space-point with the highest quadrature sum of track  $p_T$  ( $\sum_{track} p_T^2$ ) is identified as the primary vertex of an event.

### 7.3 Electron Reconstruction

When interacting with a material, electrons produce a photon by bremsstrahlung radiation, the process defining the interaction of a charged particle with the electric field of an atomic nucleus. Any energetic photon, either from a physics process or bremsstrahlung, can turn into a pair of  $e^+e^-$ , which can again radiate another set of photons, thus, giving rise to an electromagnetic shower. For given energy and material, electromagnetic showers have a characteristic penetration depth.

In ATLAS, electrons are reconstructed by combining the tracking information from the ID and the energy deposits in nearby cells of the calorimeter, energy clusters [14]. The clusters are formed if the energy deposit exceeds four times the expected deposits from the pile-up. The reconstruction efficiency for high-energy electrons with transverse energy ( $E_T$ ) more significant than 15 GeV is 97 – 99% [14]. *Prompt electrons* originate from the hard scattering process and are the primary interest of physics analysis. However, the detector has electrons from *non-prompt sources* including jets, misidentification, and pile-up. It is

essential to identify and isolate the prompt electrons in an event and measure the efficiency for identification and isolation as a function of particles' kinematic properties, such as  $\eta$  or  $E_T$ . Limited by the ID coverage, only electrons within the  $|\eta| < 2.47$  range can be reconstructed and identified as prompt in ATLAS.

The electron identification is based on a multivariate-likelihood (LH) technique which takes information from tracking detectors and calorimeters as input. The identification tool is trained to separate signal and background using simulated  $Z \rightarrow e^+e^-$  and  $J/\psi \rightarrow e^+e^-$  events. The LH tool provides four *working points* (VeryLoose, Loose, Medium, and Tight) at different values of the LH discriminant to cover various needs of several ATLAS analyses. The analysis presented in this thesis uses electrons satisfying the Loose identification working point with at least one hit in the IBL.

Prompt electrons originating from W, Z, or H decays are characterized by low activity around them in the  $\eta - \phi$  plane. An isolation requirement is applied to the electron candidates to select ones from the hard scattering. Calorimeter and track-based requirements on isolation variables are defined to quantify the isolation. The variables are based on the amount of activity around an isolation cone of the candidate electron. Calorimeter-based isolation relies on the variable  $E_{T,cone}^{iso}$ , the sum of transverse energies inside a  $\Delta R$  cone<sup>3</sup> of 0.2 for the electron candidate. Similarly, the track-based isolation variable is  $p_{T,cone}^{iso}$ , the sum of the transverse momentum of the electron candidate within a  $p_T$ -dependent  $\Delta R$ , which is defined as,

$$\Delta R = \min \left( \frac{10 \text{ GeV}}{p_T}, \Delta R_{max} \right), \quad (7.1)$$

where the maximum cone size is  $\Delta R_{max} = 0.2$ . Similar to the identification, several working points are available for electron isolation. The measurement in this thesis uses the *Loose\_VarRad* isolation working point which requires  $E_{T,cone}^{iso} < 0.3$  and  $p_{T,cone}^{iso} < 0.15$ . Figure 18 shows the electron identification and isolation efficiencies as a function of their  $E_T$ . The Loose working point has the highest identification and isolation efficiencies and is the

---

<sup>3</sup> $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$  gives the geometrical distance between any two physics object.

optimal working point for the measurement because of the fully reconstructable clean final state.

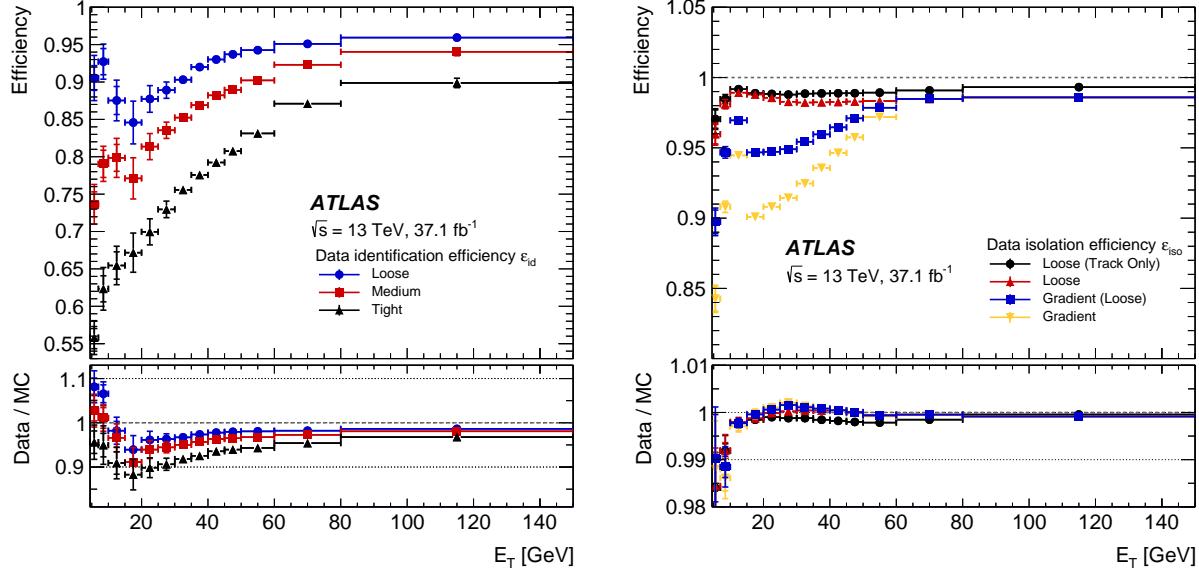


Figure 18: Distributions showing the identification (left) and isolation (right) efficiencies for electrons in data as a function of their  $E_T$  [14].

The total electron efficiency is defined as the product of the electron reconstruction, identification, isolation, and trigger efficiencies as

$$\epsilon_{total} = \epsilon_{reco} \times \epsilon_{id} \times \epsilon_{iso} \times \epsilon_{trigger}. \quad (7.2)$$

Each of the efficiency terms is evaluated on data and MC. *Scale Factors (SF)* defined as the ratio of the measured efficiency in data and the efficiency simulated in MC are derived and applied to the simulation to match the one observed in the data. Typically, SFs are close to one, and systematic uncertainties related to the electron reconstruction, identification, isolation, trigger efficiency, and the different scale factors are considered in the measurement.

## 7.4 Muon Reconstruction

The rate of bremsstrahlung radiation is inversely proportional to the square of a particle’s mass. Since muons are about 200 times heavier than electrons, they primarily interact with the detector material through ionization. Therefore, muons are minimally ionizing particles that do not create electromagnetic shower in the calorimeters and pass through all layers of the ATLAS detector. Hence, muon detection relies on track measurements from the inner detector and muon spectrometer. As shown in Figure 19, four types of muons are defined based on the type of sub-detectors used during a muon reconstruction:

- **Combined muons:** muons reconstructed from a global refit of ID and MS tracks;
- **Segment-tagged muons:** muons reconstructed from a fitted ID track and MS segment track;
- **Calo-tagged muons:** muons reconstructed using ID track matched to the minimum ionizing energy deposits in the calorimeters;
- **Standalone Muons:** muons reconstructed solely from MS tracks.

Similar to the electron reconstruction discussed in Section 7.3, reconstructed muons from the hard scatter are identified and isolated from the muons originating from secondary sources. Muon identification working points are developed by applying quality requirements in the simulated  $t\bar{t}$  events where a  $W$  from top-quark decays to a muon and a neutrino. The quality cuts require at least one-pixel hit, five SCT hits, and not more than three pixel or SCT holes. At least 10% of TRT hits are included in the fit for the  $0.1 < \eta < 0.9$  range with full TRT coverage [64]. The following three variables are used to define an identification *working point* for the reconstructed muons [64]:

- $q/p$  significance, defined as the absolute difference between the ratio of muon’s charge and momentum of the muons measured in the ID and the MS divided by the uncertainties added in quadrature,

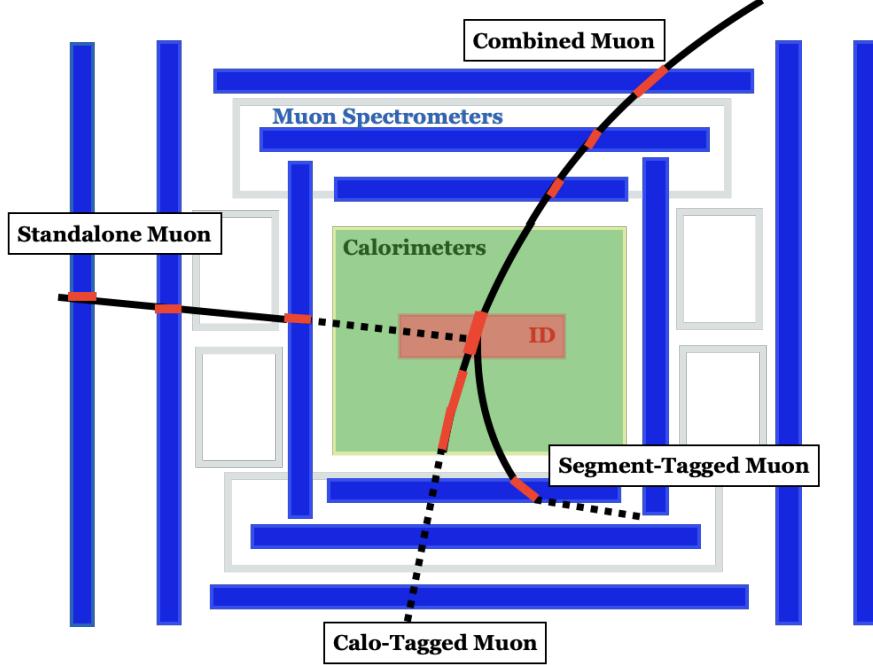


Figure 19: Schematic of four different types of muons reconstructed using several layers of sub-detectors [15].

- $\rho'$ , defined as the absolute difference of transverse momentum measurements in the ID and MS divided by the combined track's  $p_T$ , and
- normalized  $\chi^2$ , of the combined track.

The latter two quantities are only used in the case of combined muons. Four identification working points (Loose, Medium, Tight, and High- $p_T$ ) are defined. This measurement uses the Loose identification point, which comprises all four types of muons and is developed specifically for processes with four leptons in the final state [64].

To evaluate the total reconstruction efficiency of muons,  $Z \rightarrow \mu^+ \mu^-$  and  $J/\psi \rightarrow \mu^+ \mu^-$  events are used. The reconstruction efficiency in the region with ID coverage  $|\eta| < 2.5$  is obtained by using the tag-and-probe method, whereas, for muon with  $|\eta| > 2.5$ , it is estimated by evaluating SFs based on a double ratio of data and MC in  $Z \rightarrow \mu\mu$  events [65].

As discussed for electrons, muons are required to meet isolation requirements. A more robust isolation technique based on the particle flow algorithm is utilized for muons. As the particle flow algorithm for muons relies on utilizing the information from the ID, calorimeters,

and the MS, it uses both charged and neutral particle flow objects. The first isolation variable consists of the sum of the transverse momentum of the muon candidate within a  $p_T$ -dependent  $\Delta R$  cone with a maximum size of  $\Delta R = 0.3$ . The second neutral particle flow isolation variable is constructed using the sum of the transverse energy of the neutral particle flow objects within the cone size of  $\Delta R = 0.2$ . Muons used in the measurement satisfy *PflowLoose\_VarRad* isolation working point, which requires both charged and neutral particle flow isolation variables to be less than 0.16.

Similarly to the electrons, the total efficiency for muon reconstruction and selection is the product of reconstruction, identification, isolation, and trigger efficiencies. The systematic uncertainties on each part of muon reconstruction and different scale factors are propagated to the final cross-section measurement.

## 7.5 Jet Reconstruction

Due to color confinement in QCD, a quark or gluon cannot exist on its own and goes through hadronization to form a collimated color-neutral stream of particles, *jets*. Generally, jets penetrate through the electromagnetic calorimeter and get fully absorbed by the material in the hadronic calorimeter. ATLAS jet reconstruction relies on the formation of *topo-clusters*, energy deposits in the calorimeter cells using a sequential recombination algorithm. Most commonly, the jets are reconstructed using a type of sequential algorithm, the anti- $k_T$  algorithm. At first, pseudojets are formed by iteratively combining nearby particles based on a specific metric distance to other particles from the pseudojet. The iteration is continued over a set of nearby particles until the pseudojets' metric distance is smaller than that between the pseudojet and the nearby particle. For the anti- $k_T$  algorithm, the separation distance metric is based on their transverse momenta [66]. A radius parameter for jets determines the angular size of resulting jets in the  $y - \phi$  plane.

The analysis reported here uses jets reconstructed using the anti- $k_T$  algorithm with a particle flow implementation [16] [66]. The particle flow reconstruction technique was first

implemented by ATLAS reconstruction in Run–2. It combines information from the inner detector tracks and the calorimeter’s topo-clusters. It has advantages over simple cluster-based algorithms because the tracking detector has a better momentum resolution for lower-energy charged particles, extending the reconstruction to include particles with topo-cluster energy deposits lower than the noise thresholds [16]. The particle flow algorithm removes cluster contributions from pile-up using track-to-vertex association.

Figure 20 shows a schematic of the particle flow jet reconstruction algorithm used in ATLAS. First, ID tracks and topo clusters are reconstructed. If an ID track matches the topo-cluster, the tracks are used for jet reconstruction, and the cluster is modified by subtracting the energy of the used tracks. In the case of neutral jets or jets outside the geometrical acceptance of ID, only the topo-clusters are used as input for jet reconstruction.

Figure 21 shows the improvement in jet transverse momenta resolution when using the particle flow algorithm over the cluster-based algorithm as a function of jet’s  $p_T$  (left) and  $\eta$  (right). The distributions show the resolution is improved significantly for low- $p_T$  and central jets having low  $\eta$  with particle flow algorithm.

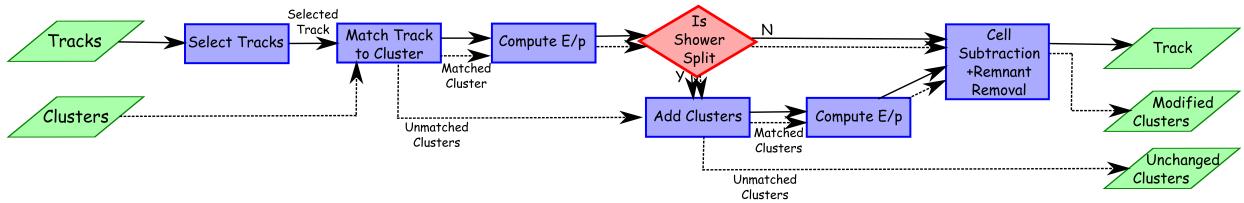


Figure 20: Schematic of particle flow jet reconstruction [16].

Jet energies are highly susceptible to pile-up density, and jet sub-structure could interact differently with various sub-detectors. Therefore, the reconstructed jets must be calibrated before being used in the measurement. Figure 22 schematically shows different steps of jet calibration, which is discussed in detail in Ref [17]. The first step in jet calibration is an event-by-event subtraction of pile-up-related average energy density ( $\rho$ ) from jet candidate energy. The pile-up is first corrected as a function of  $p_T$ , then as a function of average pile-up density ( $\mu$ ) and the number of primary vertex per event ( $N_{PV}$ ). The second set of

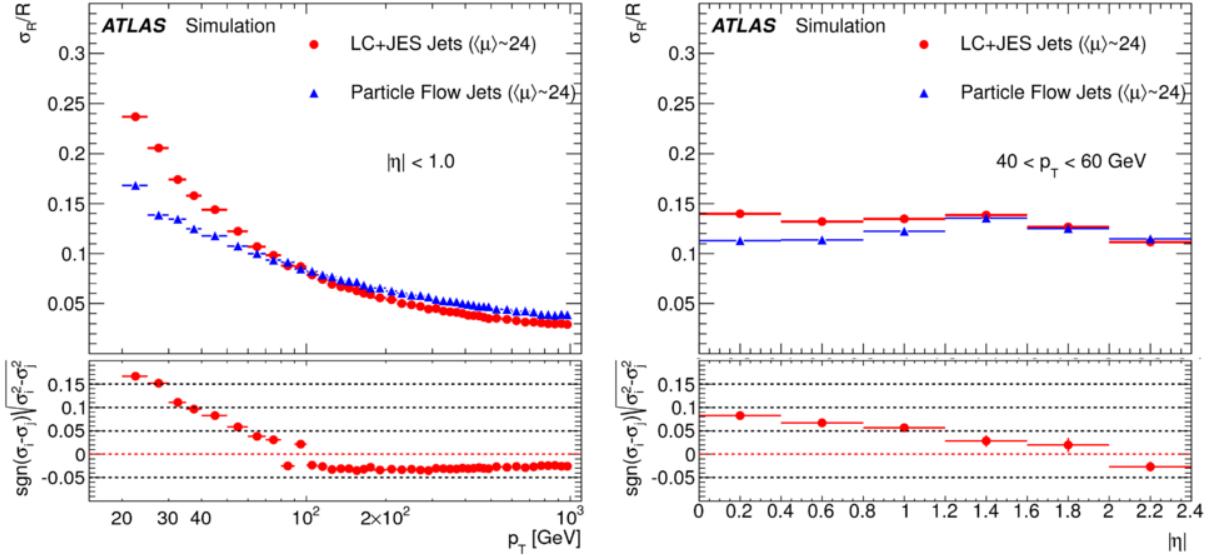


Figure 21: Resolution of jet transverse momentum for only cluster-based jets (LC+JES) and particle flow jets as a function of  $p_T$  (left) and  $\eta$  (right) [16].

jet calibration accounts for the detector response, correcting for the interaction with the detector's passive material and possible flavor dependence on the detector response. *In-situ* calibration is the final part of the jet calibration, which accounts for the differences in jet response measured in data and MC.

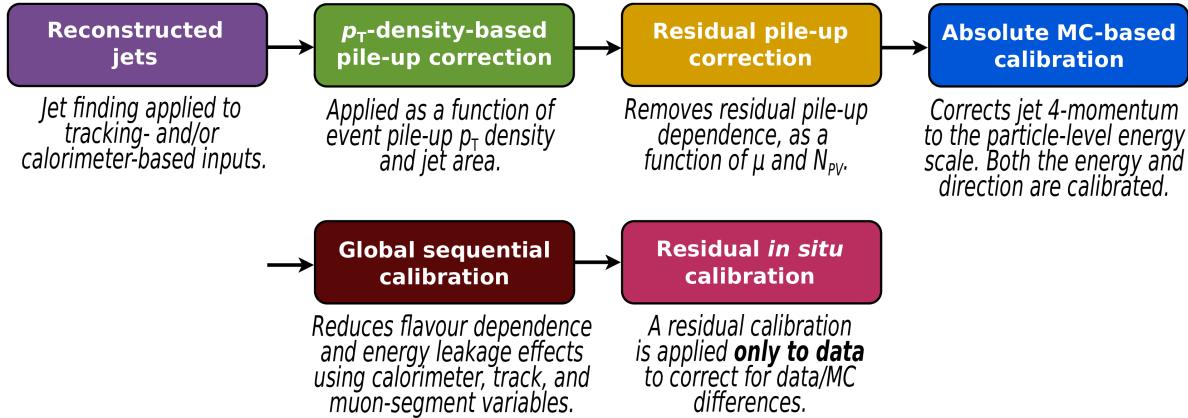


Figure 22: A schematic of different steps of jet calibration [17].

The analysis uses a *Jet Vertex Tagger* (JVT) variable to differentiate hard-scattering jets from pile-up jets [67]. The JVT variable is a multivariate combination of two variables; first, the *jet-vertex-fraction* defined as a fraction of the total momentum of tracks in jet associated

with the primary vertex, and second, the number of reconstructed primary vertices in an event ( $N_{PV}$ ) [67]. To identify the jets from hard scattering outside the ID acceptance in  $2.5 < |\eta| < 4.5$  range, a *forward Jet Vertex Tagger* (fJVT) discriminant is used, which exploits the correlations between the topological clusters from the cells in the calorimeters with significant energy deposits [68].

The systematic uncertainties related to each step of jet reconstruction, calibration, and JVT/fJVT tagging are propagated to the final measurements of the differential cross-sections.

# 8 Future Upgrades

## 8.1 High-Luminosity LHC

The planned High-Luminosity Large Hadron Collider (HL-LHC) is expected to operate starting in mid-2029. The primary goals of the HL-LHC are to collect large quantities of high-quality data needed to study rare SM processes such as Higgs self-interaction, Higgs couplings to lighter particles, the longitudinal component of vector boson scattering processes, and to extend the direct BSM searches beyond the current reach of LHC. The HL-LHC upgrade aims to increase the center-of-mass energy of proton-proton collisions to  $\sqrt{s} = 14$  TeV and the instantaneous luminosity up to  $\mathcal{L} = 7.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  [69]. Figure 23 shows the complete operation of LHC starting in 2011 to the planned decade-long HL-LHC program.

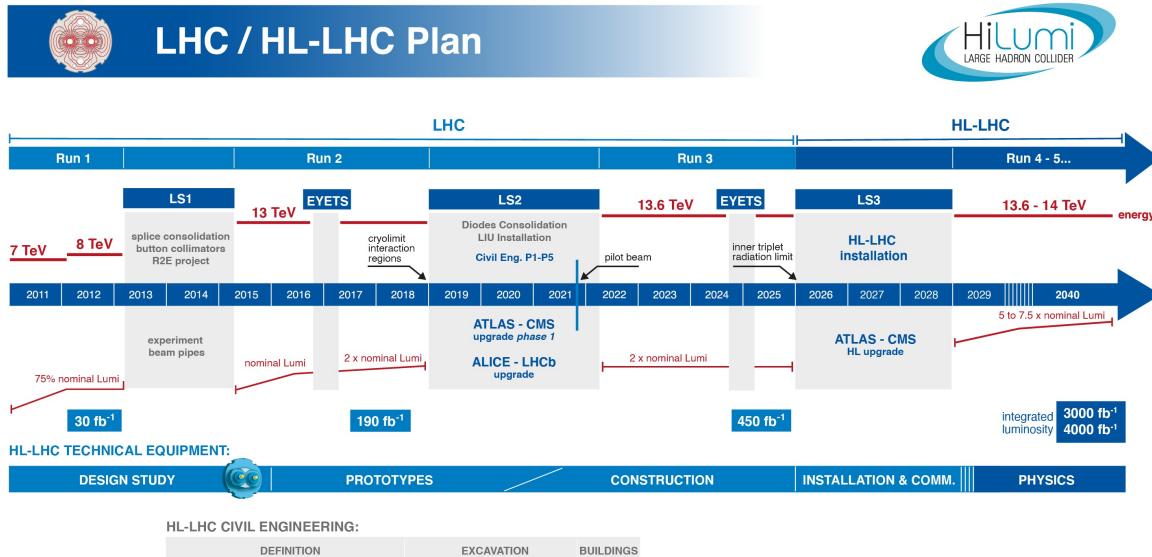


Figure 23: Timeline of LHC operation starting from 2011 to the planned HL-LHC upgrade. Taken from <https://hilumilhc.web.cern.ch/content/hl-lhc-project>.

## 8.2 ATLAS Upgrades

The higher center-of-mass energy collisions and about 200 interactions per bunch crossing at the HL-LHC give rise to several detector challenges, such as higher detector occupancy, harsher radiation conditions, and higher particle fluxes [69]. The ATLAS detector will upgrade several sub-systems to meet the challenges of the HL-LHC. The most significant upgrade is replacing the current ID with all-Silicon Inner Tracking (ITk) detector [69]. Other upgrades include the muon system upgrades, such as the replacement of some MDT chambers in the inner barrel region [70], the trigger and data acquisition system upgrade to meet challenges from higher detector occupancy [71], as well as upgrading the electronics of several other sub-systems [69]. A new High Granularity Timing Detector (HGTD) will also be inserted in the end-cap regions to supplement the tracking system [69].

The ITk consists of Silicon pixel and strip detectors to increase granularity and radiation hardness with less material in the detector. Figure 24 shows the ITk layout with 5 inner layers of pixel detector and four outer layers of strips detector. The tracking for ITk is extended in the forward region up to  $|\eta| < 4.0$  [72].

At the HL-LHC, the ATLAS experiment is expected to record at least ten times more data than Run-2, making the precision measurements of the rare vector boson scattering process crucial. Experimental uncertainties related to jet identification and reconstruction are significant sources of total systematic uncertainties in Run-2 VBS measurements. Compared to Run-2, at the HL-LHC, the pile-up jet rejection efficiency for the forward jets is expected to increase dramatically due to extended  $\eta$  coverage from the ITk [73]. Moreover, the timing information from the HGTD in the HL-LHC is expected to further improve the forward pile-up jet rejection efficiency up to approximately 30% [74]. Therefore, the HL-LHC program is critical to thoroughly probe the VBS measurements with extremely small cross-sections and two jets in the forward regions. Section 23 discusses the potential of VBS measurements at the HL-LHC in more detail.

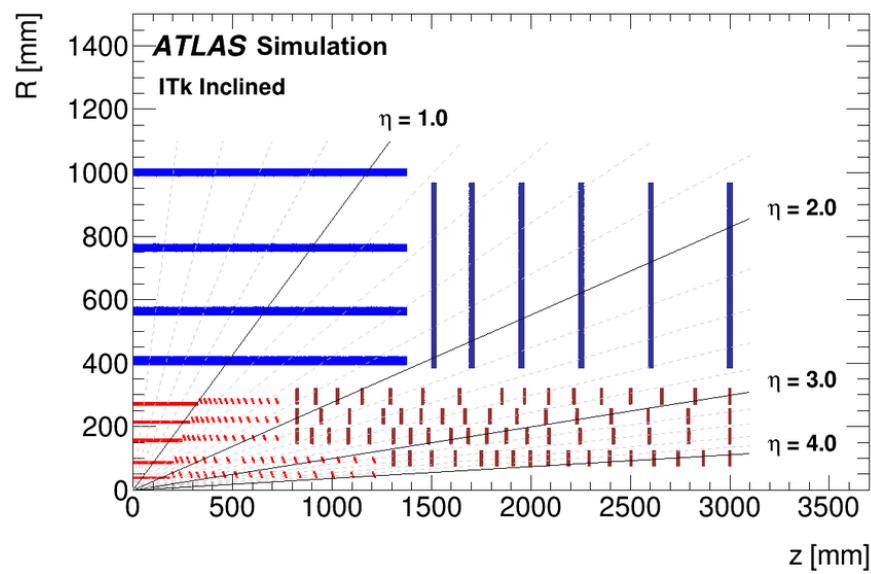


Figure 24: Schematic layout of ITk [18].

## Chapter IV: Analysis Overview

### 9 Goals

The primary goal of this analysis is to measure the differential cross-section of  $ZZ^*(\rightarrow 4\ell)jj$  processes as a function of several kinematic observables sensitive to the EWK production mode. The clean final state provides an invaluable avenue to study the rare electroweak production mode, which is experimentally accessible for the first time with LHC Run-2 statistics. The differential cross-sections are measured in a VBS-Enhanced region within a fiducial phase space that falls under the physical acceptance of the detector. For simpler re-interpretation in the future without ATLAS detector simulations, the differential cross-sections are measured at the particle-level using an unfolding technique, which removes the detector effects, such as limited efficiency and resolution. The measured cross-sections are then compared to the most precise SM predictions. As discussed in Section 4, the electroweak production of  $ZZ^*(\rightarrow 4\ell)jj$  consists of contributions from VBS and is sensitive to possible BSM effects. Therefore, the unfolded differential cross-sections are used to constrain parameters of BSM physics, modifying the quartic self-interactions of the vector bosons as shown in Figure 7b.

## 10 Detector-Level Object Selection

This section summarizes the detector-level selections applied to the three physics objects, electrons, muons, and jets used in the measurement. These kinematic requirements ensure that only high-quality physics objects relevant to the  $ZZ^*(\rightarrow 4\ell)jj$  production are selected for the final measurement. Each physics object considered in the analysis is categorized as *baseline* or *signal*. Physics objects satisfying a set of kinematic selections or looser identification criteria are categorized as *baseline*. In contrast, the baseline objects that pass either stricter kinematic selections or additional isolation and track-to-vertex association (TTVA) requirements are categorized as *signal objects*.

### 10.1 Electrons

Baseline electron objects are required to satisfy the kinematic selections of  $p_T > 7$  GeV and  $|\eta| < 2.47$  and a loose likelihood identification working point of *LHVeryLoose*. Some baseline electron candidates are reconstructed by matching the pile-up tracks to the calorimeter cluster deposit. A loose vertex association requirement of  $|z_0 \sin \theta| < 0.5$  mm is applied to the baseline electron candidates to avoid these wrongfully reconstructed electrons.

Signal electrons are required to pass a more stringent loose likelihood identification, *LHLooseBL*, which requires at least one hit in the innermost layer of the pixel detector (IBL). The signal electrons are distinguished by tightening the impact parameter significance of the baseline electrons to  $d_0/\sigma_{d0} < 5$  and requiring an additional isolation working point identification of *LooseVarRad* defined in Section 7.3. Table 4 summarizes the several kinematic selections imposed to define the baseline and signal electrons.

### 10.2 Muons

Baseline muons are required to satisfy  $|\eta| < 2.7$ ,  $p_T > 5$  GeV, a loose impact parameter requirements of  $|z_0 \sin \theta| < 0.5$  mm, and *Loose* identification working point. The signal muons

Table 4: Definition of the baseline and signal electrons.

Selection Category	Baseline	Signal
Kinematic cuts	$p_T > 7 \text{ GeV}$ $ \eta  < 2.47$	$p_T > 7 \text{ GeV}$ $ \eta  < 2.47$
Identification	LHVeryLoose	LHLooseBL
Vertex Association	$ z_0 \sin \theta  < 0.5 \text{ mm}$	$ z_0 \sin \theta  < 0.5 \text{ mm}$
Isolation Working Point	—	PflowLooseVarRad
Impact Parameter Significance	—	$d_0/\sigma_{d_0} < 5$

are identified from the baseline muons by requiring additional isolation identification of *PflowLooseVarRad* defined in Section 7.4 and tightening the TTVA requirements to  $d_0/\sigma_{d_0} < 3$ . Table 5 summarizes the baseline and signal muons selection requirements.

Table 5: Definition of the baseline and signal muons.

Selection Category	Baseline	Signal
Kinematic cuts	$p_T > 5 \text{ GeV}$ Calo-tagged $p_T > 15 \text{ GeV}$ $ \eta  < 2.7$	$p_T > 5 \text{ GeV}$ Calo-tagged $p_T > 15 \text{ GeV}$ $ \eta  < 2.7$
Identification	Loose	Loose
Vertex Association	$ z_0 \sin \theta  < 0.5 \text{ mm}$	$ z_0 \sin \theta  < 0.5 \text{ mm}$
Isolation Working Point	—	PflowLooseVarRad
Impact Parameter Significance	—	$d_0/\sigma_{d_0} < 3$

### 10.3 Jets

As discussed in Section 4, the jets from the EWK  $ZZ^*jj$  production are highly energetic; thus, reconstructed jets are required to have  $p_T > 30 \text{ GeV}$ . The jet energy scale and resolution calibration discussed in Section 7.5 is only valid for jets within  $|\eta| < 4.5$  region. Therefore, the baseline jets are required to be in the  $|\eta| < 4.5$  region. Baseline jets in  $|\eta| < 2.4$  satisfying the *Tight* working point of the *Jet-vertex-tagger (JVT)* tool, and in  $|\eta| > 2.5$  satisfying the *Tight* working point of *forward-jet-vertex-tagger (fJVT)* tool are classified as signal jets. Table 6 summarizes the details of baseline and signal jets selection.

A particular type of jets, the *b-jets*, containing b-hadrons initiated from a b-quark, is also used for the background estimation. The b-jets reconstruction relies on multivariate analysis

(MVA), utilizing the fact that b-hadrons have a long mean lifetime (about 1 picosecond), leading to a displaced secondary vertex in the detectors. A tagger discussed in detail in Ref [75] with b-tagging efficiency of 77% is used in the analysis to identify b-jets.

Table 6: Definition of the baseline and signal jets.

Selection Category	Baseline	Signal
Kinematic cuts	$p_T > 30 \text{ GeV}$ $ \eta  < 4.5$	$p_T > 30 \text{ GeV}$ $ \eta  < 4.5$
Jet-Vertex-Tagger	– –	$ \eta  < 2.4 \text{ JVT ("Tight")}$ $ \eta  > 2.5 \text{ fJVT ("Tight")}$

## 10.4 Overlap Removal

In order to avoid double-counting, an *overlap removal* procedure is applied to remove physics objects reconstructed from the same detector signal. The measurement uses a lepton-favored overlap removal which selects leptons over jets. Overlap removal is an iterative process in which only objects surviving all previous steps are used in the subsequent steps. Table 7 summarizes the overlap removal steps, where the  $\Delta R$  is the angular separation between objects calculated using rapidity.

Table 7: Overlap removal used in the analysis. An object removed in one step does not enter into the subsequent step.

Remove Object	Accept Object	Overlap Criteria
Electron	Electron	Share a track or have overlapping calorimeter cluster. Keep electron with higher $p_T$
Muon	Electron	Share ID track, and the muon is calo-tagged
Electron	Muon	Share ID track
Jet	Electron	$\Delta R_{e-jet} < 0.2$
Jet	Muon	$\Delta R_{\mu-jet} < 0.2$ /ghost-associated and $N_{jet \text{ tracks}} < 3$

## 11 Trigger

Due to the presence of four fully reconstructed leptons in the final state, the data events and predicted detector-level events are preselected using a logical OR of different single and double-lepton triggers. The trigger menu varies according to the data-taking run periods to reflect the changes in the high-level trigger system, which are required to cope with increasing data rates. Additionally, trigger matching is required for the selected events. The trigger matching selects a subset of preselected events in which at least one lepton of the quadruplet is matched to one of the fired triggers. Table 8 shows the trigger menu used by the analysis per different data periods using either electrons, muons, or mixed electron-muon triggers. The  $HLT_*$  string specifies the high-level trigger menu used where the "e" or "mu" substring specifies the type of object used to fire the trigger, and the attached number specifies the minimum  $p_T$  threshold for the object. The substring  $lh^*$  attached to  $HLT_*$  stands for the likelihood identification working point for the electrons, and the  $ivar^*$  specifies the isolation working point used for either object. The string  $L1^*$  indicates the use of either calorimeters or MS L1 trigger, and the string  $noL1$  infers that no L1 triggers were fired.

Figure 25 shows the efficiency of trigger selection in events with a signal quadruplet and a dijet as a function of the invariant mass of the four-lepton system ( $m_{4\ell}$ ) in data and total SM prediction for 2015-2016 data-taking period. The trigger efficiency is the ratio of events passing detector-level event selection and trigger requirements to the total number of events passing detector-level event selection requirements. The requirements for detector-level event selection are discussed in Section 12. For both data and prediction, the trigger efficiency is 100% except for the first bin. Scale factors are applied to MC events to account for possible trigger efficiency differences between data and MC events. The scale factors are defined as a fraction of trigger efficiency for MC to that of data and retrieved from the ATLAS supported tool *TrigGlobalEfficiencyCorrectionTool*<sup>4</sup>.

---

<sup>4</sup><https://gitlab.cern.ch/atlas/athena/tree/21.2/Trigger/TrigAnalysis/TrigGlobalEfficiencyCorrection>

Period	Leptons	Triggers
2015	Electron	HLT_e24_lhmedium_L1EM20VH HLT_e60_lhmedium HLT_e120_lhloose HLT_2e12_lhvloose_L12EM10VH
	Muon	HLT_mu20_iloose_L1MU15 HLT_mu50 HLT_2mu10 HLT_mu18_mu8noL1
	Mixed	HLT_e7_lhmedium_mu24 HLT_e17_lhloose_mu14
2016	Electron	HLT_e26_lhtight_nod0_ivarloose HLT_e60_lhmedium_nod0 HLT_e140_lhloose_nod0 HLT_2e17_lhvloose_nod0
	Muon	HLT_mu26_ivarmedium HLT_mu50 HLT_2mu14 HLT_mu22_mu8noL1
	Mixed	HLT_e7_lhmedium_nod0_mu24 HLT_e17_lhloose_nod0_mu14
2017	Electron	HLT_e26_lhtight_nod0_ivarloose HLT_e60_lhmedium_nod0 HLT_e140_lhloose_nod0 HLT_2e24_lhvloose_nod0
	Muon	HLT_mu26_ivarmedium HLT_mu50 HLT_2mu14 HLT_mu22_mu8noL1
	Mixed	HLT_e17_lhloose_nod0_mu14 HLT_e26_lhmedium_nod0_mu8noL1
2018	Electron	HLT_e26_lhtight_nod0_ivarloose HLT_e60_lhmedium_nod0 HLT_e140_lhloose_nod0 HLT_2e24_lhvloose_nod0
	Muon	HLT_mu26_ivarmedium HLT_mu50 HLT_2mu14 HLT_mu22_mu8noL1
	Mixed	HLT_e17_lhloose_nod0_mu14 HLT_e26_lhmedium_nod0_mu8noL1

Table 8: Trigger menu used in the analysis for event preselection

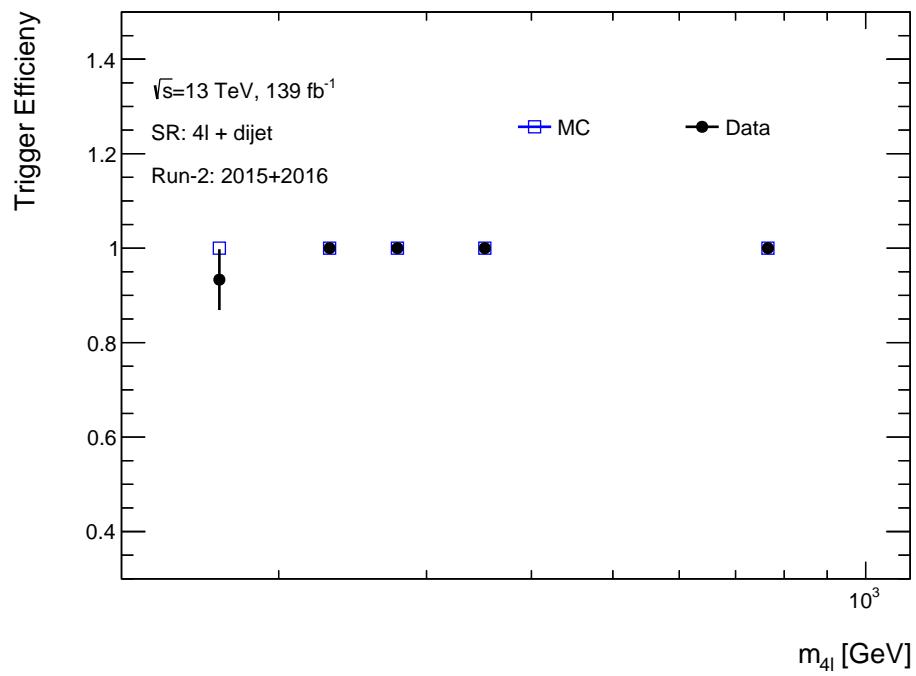


Figure 25: Trigger efficiency as a function of  $m_{4\ell}$  in events with a quadruplet and a dijet in data and SM prediction corresponding to the 2015-2016 data-taking period.

## 12 Event Selection

A  $ZZ^*(\rightarrow 4\ell)jj$  event at the detector level consists of a lepton quadruplet formed by two SF-OC lepton pairs from each  $Z$  boson decay and a dijet from the initial state partons. The quadruplet is formed in events with four prompt leptons (electrons or muons), where the leading and sub-leading leptons satisfy  $p_T > 20$  GeV to ensure a high trigger efficiency. Due to momentum conservation, the prompt leptons are separated from each other. Therefore, all prompt leptons in an event must have  $\Delta R > 0.05$  to reduce contributions from mis-reconstruction while keeping leptons from possible boosted production scenarios. Additionally, each SF-OC lepton pair's invariant mass is required to satisfy  $m_{\ell\ell} > 5$  GeV to suppress the contamination from lower resonance backgrounds. A quadruplet is formed from the two SF-OC lepton pairs whose invariant masses are closest and next closest to the mass of the Z-boson ( $m_Z$ ). Based on these requirements, the quadruplets can be of the following three types:

- $4e$ : events with two  $e^+e^-$  pairs.
- $4\mu$ : events with two  $\mu^+\mu^-$  pairs.
- $2e2\mu$  or  $2\mu2e$ : events where one of the pair is  $e^+e^-$  and other is  $\mu^+\mu^-$

Once the quadruplet is formed, the leading-lepton pair is defined as the one with a higher absolute rapidity value, i.e.,  $|y_{ij}|$ . This additional identification is motivated by consistently defining the leading and sub-leading pairs at both particle and detector levels to reduce the resolution-induced bin migrations, which need to be corrected by the unfolding procedure. The quadruplets with all four leptons passing the signal lepton criteria of the TTVA and isolation are the *signal quadruplet* defining the signal region. The quadruplets with one or more lepton failing either the isolation or TTVA requirement are the *non-signal quadruplets*. They are used in the non-prompt background estimation discussed in Section 16.1. The invariant mass of the quadruplet ( $m_{4\ell}$ ) must be greater than 130 GeV to exclude the events

from on-shell  $H \rightarrow ZZ^*$  production, which are measured extensively by ATLAS analyses focused on Higgs measurements.

As discussed in Section 4, the final state jets in electroweak production of  $ZZ^*(\rightarrow 4\ell)jj$  come from the initial state quarks on the opposite sides of the interaction point. Thus, the dijet is selected by requiring two signal jets defined in Section 10.3 from the opposite side of the detector, i.e.,  $\eta_{leading\ jet} \times \eta_{sub-leading\ jet} < 0$ . To maximize the probability of selecting an event from EWK  $ZZ^*(\rightarrow 4\ell)jj$  production, a requirement of significant rapidity difference between the jets of  $\Delta y_{jj} > 2$  and a large invariant mass of  $m_{jj} > 300$  GeV are imposed on the dijet selection.

## 12.1 Signal Region

As discussed in Section 4, the two  $Z$  bosons in the electroweak production of  $ZZ^*jj$  are produced centrally with respect to the dijet. Thus, the signal region of the analysis is defined based on the centrality ( $\zeta$ ) of the di- $Z$ boson production in an event. Centrality depends on the rapidity of the quadruplet and the rapidity of the dijet as:

$$\zeta = \frac{|y_{quadruplet} - 0.5 * (y_{leading\ jet} + y_{sub-leading\ jet})|}{|y_{leading\ jet} - y_{sub-leading\ jet}|}. \quad (12.1)$$

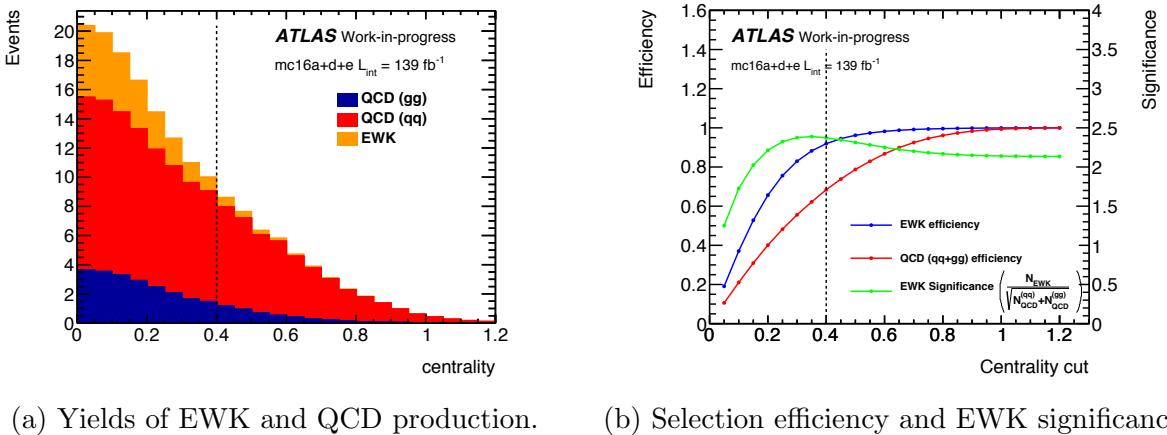
Figure 26a shows the predicted MC distribution of centrality for the three main production modes of  $ZZ^*(\rightarrow 4\ell)jj$ . The significance of the EWK component over the inclusive parton initiated and  $gg$ -loop initiated QCD production is defined as,

$$s = \frac{N_{EWK}}{\sqrt{N_{QCD}^{(qq)} + N_{QCD}^{(gg)}}}, \quad (12.2)$$

where  $N_{EWK}$ ,  $N_{QCD}^{(qq)}$  and  $N_{QCD}^{(gg)}$  are the numbers of events from electroweak, parton-initiated QCD and gluon-loop initiated QCD productions, respectively. The chosen cut value on the centrality maximizes the EWK significance while maintaining a good selection efficiency of

EWK events. Figure in 26b shows the efficiency and significance for various cut values of centrality.

A VBS-Enhanced signal region is defined based on events with a quadruplet, a dijet, and  $\zeta < 0.4$ . The low value of the centrality and the requirements for a signal dijet ensures that the events in this signal region originate in a more significant fraction from the electroweak production of  $ZZ^*(\rightarrow 4\ell)jj$ . A VBS-Suppressed control region is also defined based on events with a quadruplet, a dijet, and  $\zeta > 0.4$ . These events mainly originate from the QCD production of  $ZZ^*(\rightarrow 4\ell)jj$  and are used to optimize the analysis strategies. Table 9 summarizes all selections applied to select  $ZZ^*(\rightarrow 4\ell)jj$  detector-level events.



(a) Yields of EWK and QCD production.

(b) Selection efficiency and EWK significance.

Figure 26: Centrality dependence for yield, EWK selection efficiency, and EWK significance.

Figure 27 illustrates a typical signal event with two  $Z$ -bosons produced in association with two jets. The event display corresponds to an event recorded during Run Number 340368 of the 2017 data-taking period. The two light-yellow cones on two opposite sides of the detector with a large rapidity gap represent the reconstructed dijet of the event with an invariant mass of  $m_{jj} = 2228 \text{ GeV}$ . In this event, one of the SF-OC lepton pairs is formed from  $e^+e^-$  representing  $Z \rightarrow e^+e^-$  decay, and the other is formed from  $\mu^+\mu^-$  corresponding to  $Z \rightarrow \mu^+\mu^-$  decay, which are represented respectively by the pairs of green and red tracks. The invariant mass of the four-lepton system is  $m_{4\ell} = 605 \text{ GeV}$ . The significant rapidity separation between the two jet cones on the opposite sides of the ATLAS detector and

Table 9: Details of event selection.

Event Selection	Cut	Requirement
Event Preselection	Trigger Vertex	Fire at least one lepton trigger At least one vertex with 2 or more tracks
Quadruplet Selection	Lepton Kinematics Lepton Separation Pair Requirement  Minimal $\Delta m_Z$  ZZ Mass	$p_T > 20$ GeV for two leading leptons $\Delta R_{ij} > 0.05$ between leptons in quadruplet Two SF-OC lepton pairs $m_{\ell\ell} > 5$ GeV quadruplet with smallest $ m_{12} - m_Z  +  m_{34} - m_Z $ Leading Pair: pair with highest $ y_{ij} $ $m_{4\ell} > 130$ GeV
Quadruplet Categorisation	Signal Quadruplet non-signal Quadruplet	Quadruplet with all <b>signal leptons</b> Quadruplet with $\geq 1$ <b>baseline-not-signal lepton</b>
Dijet Selection	Different Detector Sides Rapidity Separation Leading Jet $p_T$ Dijet Mass Dijet	$\eta_{leading\ jet} \times \eta_{sub-leading\ jet} < 0$ $\Delta y_{jj} > 2$ $p_{T,\ leading\ jet} > 40$ GeV $m_{jj} > 300$ GeV Both jets required to pass either JVT or FJVT
Event Categorisation	VBS-Enhanced Region VBS-Suppressed Region	signal quadruplet & dijet and centrality ( $\zeta$ ) $< 0.4$ signal quadruplet & dijet and centrality ( $\zeta$ ) $> 0.4$

centrally produced two  $Z$  bosons defines the characteristic feature of the EWK production of  $ZZ(\rightarrow 4\ell)jj$ .

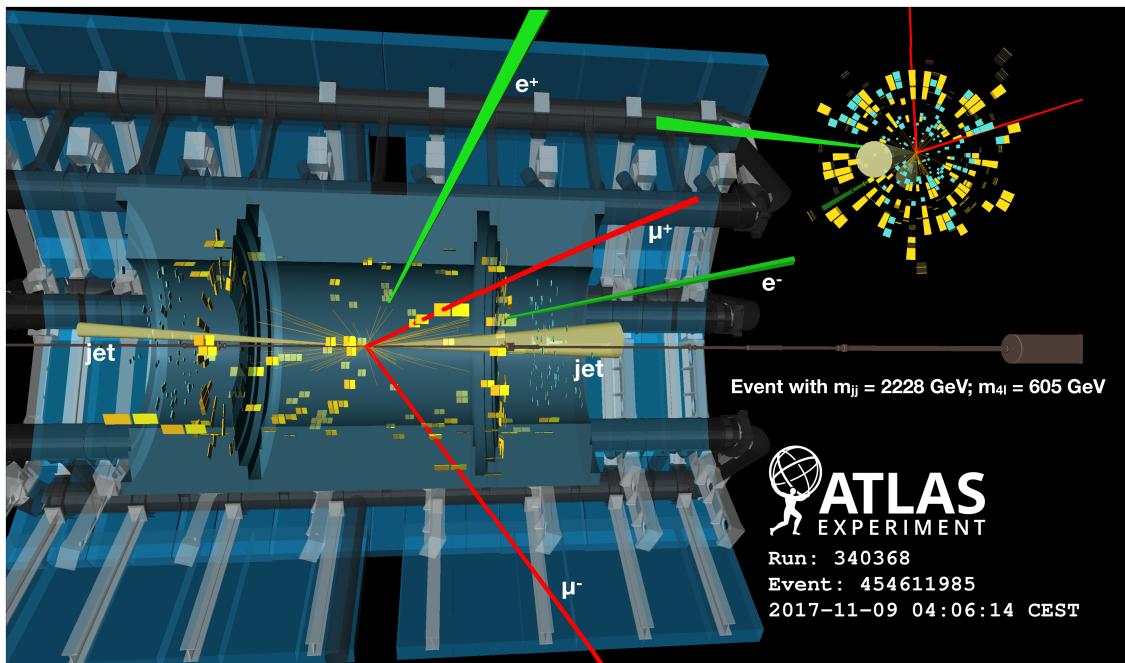


Figure 27: Event display of a candidate  $pp \rightarrow ZZjj \rightarrow e^+e^-\mu^+\mu^-jj$  recorded by the ATLAS experiment in Run-2 2017 data-taking period [19].

## 13 Fiducial Phase Space

The ATLAS detector has limited acceptance in the physical phase space, and the selected objects and events are reconstructed within this acceptance phase space. Thus, a fiducial phase space imitating the detector acceptance is defined using physics objects at the particle level to measure the unfolded differential cross-sections. The particle-level signal events are essential to derive the inputs needed to correct the detector effects. Moreover, the measured unfolded cross-sections are compared to the SM-predicted particle-level cross-sections evaluated from these events. Thus, it is essential to carefully select the particle-level signal events in the defined fiducial phase space. This section summarizes the kinematic requirements defining the fiducial phase space of the analysis and the selection of the particle-level signal events.

The kinematic selections for the fiducial phase space are close to the detector-level object and event selections. The kinematic requirements applied are motivated by the nature of the electroweak production of  $pp \rightarrow ZZ^*(\rightarrow 4\ell)jj$  [ $\ell = e, \mu$ ], where the two SF-OC lepton pairs from two  $Z$  bosons are produced centrally with respect to the highly energetic dijet. Moreover, the fiducial phase space contains no leptons from tau decays. Both fiducial-level electrons and muons are required to be dressed. The dressing procedure accounts for the energy losses of leptons through photon emissions via bremsstrahlung. Dressed leptons are constructed by adding the four-momenta of nearby photons within the lepton's small  $\Delta R < 0.1$  cone. Several kinematic cuts summarized in Table 10 are applied individually to the muons, electrons, and jets to ensure the selected particle-level objects fall within the detector's acceptance before defining the events. Motivated by the discussion of physics object reconstruction in Section 7, each electron is required to have  $p_T > 7$  GeV and  $|\eta| < 2.47$ , whereas the muons satisfy  $p_T > 5$  GeV and  $|\eta| < 2.7$ .

Event quadruplets are formed from the events with at least four leptons by requiring two SF-OC lepton pairs, with leading and sub-leading lepton  $p_T > 20$  GeV and angular

Table 10: Details of the kinematic pre-selection applied to the particle-level electrons, muons, and jets. The required kinematic cuts are applied to the dressed leptons.

Selections	Electrons	Muons	Jets
$p_T$	$> 7 \text{ GeV}$	$> 5 \text{ GeV}$	$> 30 \text{ GeV}$
$ \eta $	$< 2.47$	$< 2.7$	$< 4.5$

separation between any two leptons to satisfy  $\Delta R > 0.05$ . The invariant mass of any SF-OC lepton pair is required to satisfy  $m_{\ell\ell} > 5 \text{ GeV}$ . These particle-level selections have similar motivation to those defined for the detector-level requirements in Section 12. In any event with more than two SF-OC lepton pairs, the quadruplet is formed by choosing the two pairs that minimize the distance to the  $Z$  resonance pole. Once the quadruplet is formed, the leading-lepton pair is defined as the one with a higher absolute rapidity value, i.e.,  $|y_{ij}|$ . Finally, an additional criterion on the invariant mass of the quadruplet of  $m_{4\ell} > 130 \text{ GeV}$  is imposed.

Similarly, the event dijet is constructed from the two leading jets with the opposite sign of pseudo-rapidity ( $\eta$ ) to imitate the detector-level VBS dijet production, where jets are reconstructed on the opposite side of the detector. Similar to detector-level, the particle-level jets are required to satisfy  $|n| < 4.5$ ,  $p_{T, \text{leading jet}} > 40 \text{ GeV}$ , and  $p_{T, \text{sub-leading jet}} > 30 \text{ GeV}$ . The dijet is required to have a significant rapidity separation of  $|\Delta y_{jj}| > 2$  and high invariant mass of  $m_{jj} > 300 \text{ GeV}$  to resemble dijet produced in electroweak  $ZZ^*(\rightarrow 4\ell)jj$  production. Table 11 summarizes the requirements to select the particle-level quadruplet and dijet in an event.

Table 11: Details of the kinematic selections applied to form a particle-level quadruplet and a particle-level dijet in the fiducial volume.

Selections	Cut
Lepton Kinematics	$P_{T, \text{leading lepton}} > 20 \text{ GeV}$ $P_{T, \text{sub-leading lepton}} > 20 \text{ GeV}$
Pair Requirement	$\Delta R_{\ell i, \ell j} > 0.05$ SF-OC with $m_{\ell\ell} > 5 \text{ GeV}$
Quadruplet Requirement	2 pair candidates with smallest $ m_{12} - m_Z  +  m_{34} - m_Z $ Leading pair: pair with highest $ y_{ij} $ Sub-leading pair: pair with lowest $ y_{ij} $ $m_{4\ell} > 130 \text{ GeV}$
Di-jet Requirement	$p_{T, \text{leading jet}} > 40 \text{ GeV}$ $ \Delta y_{jj}  > 2$ $m_{jj} > 300 \text{ GeV}$

## 14 Datasets and Monte Carlo Simulation

### 14.1 LHC Dataset

The measurement uses the LHC collision data, the ATLAS Run-2 dataset collected by the ATLAS experiment during its operation in 2015, 2016, 2017, and 2018. This dataset corresponds to proton-proton collisions at the center-of-mass energy of  $\sqrt{s} = 13$  TeV and total integrated luminosity of  $139 \pm 2.4$   $\text{fb}^{-1}$  measured by the LUCID-2 detector [76] [77]. The LUCID-2 detector is a Cherenkov detector with 16 photo-multipliers (PMT) that provides counts of particles from each colliding bunch. The counts integrated over time periods of about 60 seconds are called luminosity blocks (LB). The instantaneous luminosity is given by

$$\mathcal{L} = \frac{R}{\sigma_{vis}}, \quad (14.1)$$

where  $R$  is the event rate and  $\sigma_{vis}$  is the calibration constant measured during the LHC's special runs, which took place at the beginning of each year of data taking. The calibration constant accounts for the non-linear response of the LUCID-2 detector. Uncertainty on the integrated luminosity is obtained from the combination of the measurements of each year of the LHC run.

Each data-taking run period of the LHC is further divided into sub-periods of one to three weeks that vary in beam and detector conditions. The dataset used in physics analyses is required to satisfy a series of data quality checks discussed in detail in Ref [20]. The data passing these requirements collectively form a Good Run List (GRL) and consists of several LB. Figure 28 shows the total integrated luminosity delivered by LHC in the green distribution, recorded by the ATLAS experiment in the yellow distribution and part of the GRL in the blue distribution. The plateaus correspond to the end-of-year shutdowns of LHC, and the slopes correspond to the increasing instantaneous luminosity in different data-taking periods.

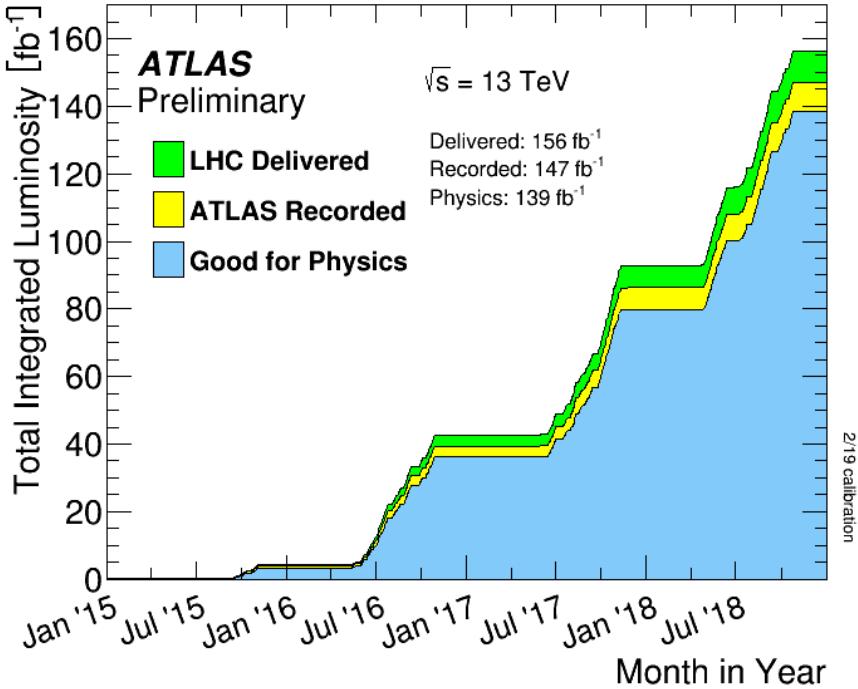


Figure 28: Total integrated luminosity collected during data taking period in Run-2 [20].

## 14.2 Monte Carlo Samples

As discussed in Section 3, the  $pp \rightarrow ZZ^*(\rightarrow 4\ell)jj$  events are simulated by MC generators. The first step in generating SM prediction is *event generation* which incorporates the matrix element calculations for the hard-scatter  $pp \rightarrow ZZ^*(\rightarrow 4\ell)jj$  process, the parton showering, hadronization, and the effect of the underlying events. The generated events are then *simulated* to interact with the ATLAS material using the Geant4 simulation toolkit following the description in Ref [78]. The energy deposits of the simulated events in the detectors are then *digitized* where the simulated signal hits are overlayed with cavern background events hits and additional hits from soft QCD events to simulate the effect of the pile-up. Finally, the events are *reconstructed* using the same procedure to reconstruct data objects using detector geometry corresponding to the data-taking period. The final simulations are available in terms of *physics derivations*, which are used in the final analysis. Figure 29 shows a schematic overview of the MC generation.

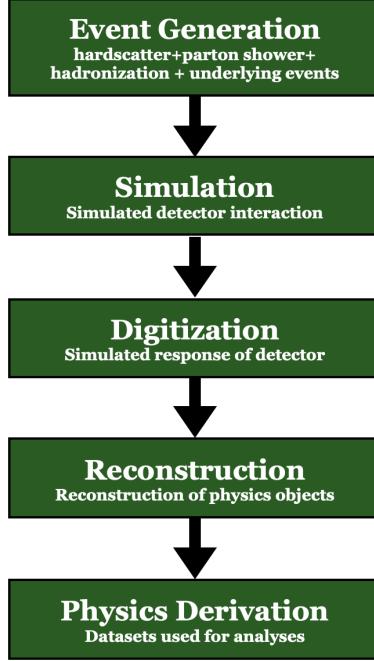


Figure 29: Various steps in MC sample generation.

Each physics process is simulated using different generation campaigns corresponding to the different conditions of Run-2 ATLAS data-taking periods. As shown in Figure 30, the pile-up distribution is different for the different data-taking periods. The MC-generated events are modified to correctly simulate the effect of pile-up distribution to simulate that of the data.

### 14.2.1 Signal Samples

As discussed in Section 4, two types of interactions, QCD and EWK, give us  $pp \rightarrow ZZ^*(\rightarrow 4\ell)jj$  final state. The two types of QCD process, quark induced  $qqZZ$  [ $qq \rightarrow ZZ^*(\rightarrow 4\ell)jj$ ] and gluon induced  $ggZZ$  [ $gg \rightarrow ZZ^*(\rightarrow 4\ell)jj$ ] are simulated using the SHERPA 2.2.2 MC generator. The parton initiated  $qqZZ$  samples corresponding to Figure 5 are generated with NLO accuracy in QCD up to one additional parton emission and LO accuracy for up to three additional parton emissions. The loop-induced  $ggZZ$  samples emerging at NNLO in  $\alpha_S$  corresponding to Figure 6 are generated using LO-accurate matrix elements for up to one additional parton emission [79]. The generator uses an NNPDF3.0NNLO PDF set evaluated

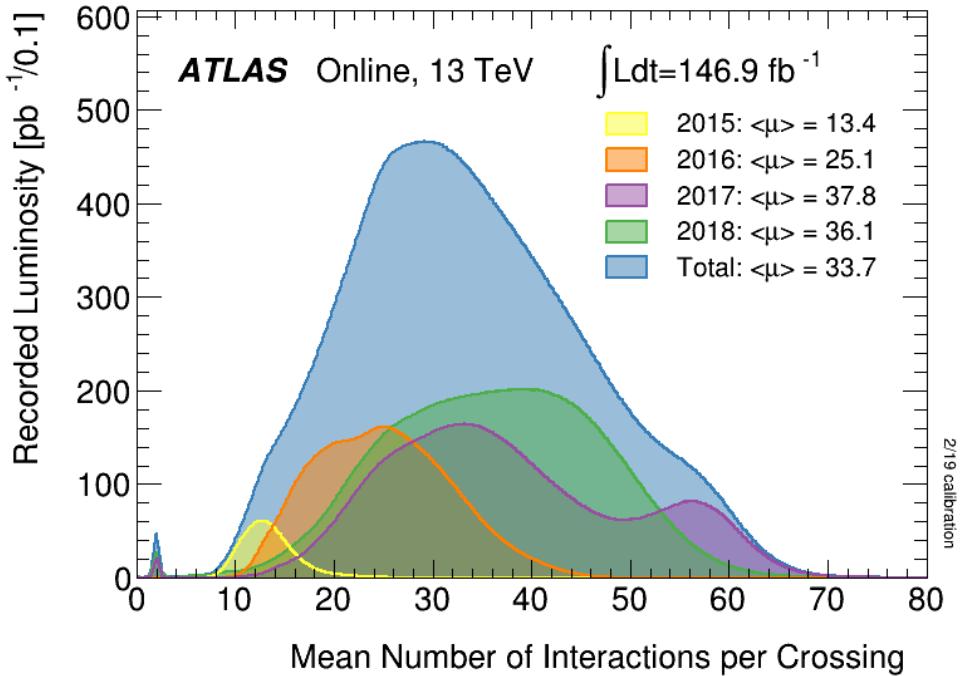


Figure 30: Pile-up distributions in different Run-2 data-taking period. [20]

using different measurements from several experiments, such as deep-inelastic inclusive cross-sections measurement from HERA-II, the combined charm data from HERA, jet production, vector boson rapidity and transverse momentum measurements from ATLAS, CMS and LHCb, total cross sections of top quark pair production from ATLAS and CMS and W+c data from CMS [80]. Parton showering is done by SHERPA’s internal algorithm based on Catani–Seymour dipole factorization matrix element [81]. The matrix element calculations are matched and merged using the *ME + PS@NLO* prescription [82]. The matching process requires the hard jets to match the parton-level quarks and gluons from matrix elements, whereas the merging allows the MC generators to merge two or more low-mass jets.

An alternative MADGRAPH5 samples produced at NLO accuracy for up to one additional parton emission and LO accuracy for up to three additional parton emissions [83] are also used in the measurement for the parton induced  $qqZZ$  samples. The generator uses the A14NNPDF23LO PDF set, and the ME is interfaced with PYTHIA8 for parton showering, merging, and matching [84].

The EWK production  $qqZZjj$  [ $qq \rightarrow ZZ^*(\rightarrow 4\ell)jj$ ] is simulated using a POWHEGV2 generator using an MSTW2008 PDF set with NLO accuracy in QCD and interfaced with PYTHIA8 for parton showering and hadronization [85]. An alternative sample at LO accuracy is also used in the measurement from MADGRAPH5 with A14NNPDF23LO PDF set and PYTHIA8 showering [83]. The POWHEGV2 NLO prediction of electroweak  $qqZZjj$  does not contain the contribution from electroweak triboson  $VZZ$  processes where two vector bosons decay leptonically and one decay hadronically. The contribution from these electroweak triboson processes is predicted using the Sherpa 2.2.2 MC generator at LO accuracy for up to two additional parton emissions and added to the POWHEGV2 predictions. Table 12 summarizes the signal MC used in the measurement.

Process	Description	Generator	PDF	Accuracy
<hr/>				
QCD $qqZZ$ $q\bar{q} \rightarrow ZZ^*(\rightarrow 4\ell)jj$	inclusive	SHERPA2.2.2 MADGRAPH	NNPDF3.0NNLO A14NNPDF23LO	$0, 1j@NLO + 2, 3j@LO$
<hr/>				
QCD $ggZZ$ loop $gg \rightarrow ZZ^*(\rightarrow 4\ell)jj$	$m_{4\ell} > 130$ GeV	SHERPA2.2.2	NNPDF3.0NNLO	$0, 1j@LO$
<hr/>				
EWK $qqZZjj$ $q\bar{q} \rightarrow ZZ^*(\rightarrow 4\ell)jj$	$m_{4\ell} > 130$ GeV	POWHEGV2 MADGRAPH	MSTW2008 A14NNPDF23LO	$\geq 2j$ (EWK) @ NLO QCD $\geq 2j$ (EWK) @LO
<hr/>				
EWK $VZZ$ $q\bar{q} \rightarrow VZZ^* \rightarrow 4\ell jj$		SHERPA2.2.2	NNPDF3.0NNLO	$1, 2j@LO$

Table 12: List of signal MC samples used in the analysis. Each process consists of three different generation campaigns corresponding to the data-taking conditions of the ATLAS Run2 data-taking periods.

### 14.2.2 Background Samples

In addition to the QCD and EWK production discussed above, two other processes, triboson ( $WWZ$ ,  $WZZ$ ,  $ZZZ$ ) and  $Z$ -bosons production in association with a top quark pair ( $t\bar{t}Z$ ), also contributes to the  $4\ell jj$  final state. The triboson processes are modeled with SHERPA2.2.2 generator at NLO accuracy in QCD for zero or one additional parton emission and LO

accuracy for up to two additional parton emissions. The triboson samples only include the fully leptonic decays of the vector bosons. Therefore, there is no overlap between the background triboson and the signal EWK  $qqZZjj$  samples. The  $t\bar{t}Z$  processes are modeled by SHERPA2.2.0 generator at LO accuracy with up to one additional parton emission using the MEPS@LO set-up [86]. The same algorithms as in the QCD  $qqZZ$  sample generation are used for parton showering, matching, and merging. The MC simulation of the triboson and  $t\bar{t}Z$  samples are subtracted directly from the data. Table 13 summarizes the details of these samples.

Process	Description	Generator	PDF	Accuracy
$pp \rightarrow W^{(*)}W^{(*)}Z^{(*)} \rightarrow 4\ell 2\nu$		SHERPA2.2.2		
$pp \rightarrow W^{(*)}Z^{(*)}Z^{(*)} \rightarrow 5\ell 1\nu$	inclusive	SHERPA2.2.2	NNPDF3.0NNLO	$0, 1j@NLO + 2j@LO$
$pp \rightarrow Z^{(*)}Z^{(*)}Z^{(*)} \rightarrow 6\ell$		SHERPA2.2.2		
$pp \rightarrow t\bar{t} + Z(\rightarrow 2\ell)$	$m_{ll} > 5$ GeV	SHERPA2.2.0	NNPDF3.0NNLO	LO

Table 13: List of background MC samples used in the analysis. Each process consists of three different generation campaigns corresponding to the data-taking conditions of the ATLAS Run2 data-taking periods.

### 14.2.3 Samples for Non-prompt Background

In addition to the triboson and  $t\bar{t}Z$  samples, the analysis has additional backgrounds coming from events with one or more non-prompt or fake leptons. These non-prompt backgrounds are estimated using a data-driven method discussed in detail in Section 16.1. MC samples are used to develop and validate the data-driven non-prompt background estimation procedure. Three sources of events could contribute as a source for non-prompt background events. The first type of events is from a Z-boson production in association with jets  $pp \rightarrow Z^*(\rightarrow 2\ell) + jets$ , which is simulated for both three or more leptons using SHERPA2.2.1. The subdominant process is events from  $t\bar{t} \rightarrow 2\ell + X$  production in which both top quarks decay semileptonically, which is simulated with POWHEG+PYTHIA8 and uses the A14NNPDF23LO PDF set [87]. The third type of non-prompt backgrounds arises

from the  $WZ$  production in which both bosons decay leptonically  $pp \rightarrow WZ \rightarrow 3\ell 1\nu$  and is simulated using **SHERPA2.2.2**. Table 14 summarizes the different processes and MC generators used in various studies related to the data-driven fake factor method to estimate the non-prompt background. A simulated non-prompt background event in the  $4\ell jj$  final state could comprise an event with a quadruplet formed from one or more fake leptons from a non-prompt source, such as misidentification and other remaining leptons from these three physics processes.

Process	Description	Generator	PDF	Accuracy
$pp \rightarrow Z^*(\rightarrow 2e) + jets$				
$pp \rightarrow Z^*(\rightarrow 2\mu) + jets$	inclusive	SHERPA2.2.2	NNPDF3.0NNLO	$NLO + 2j, LO + 4j$
$pp \rightarrow Z^*(\rightarrow 2\tau) + jets$				
$pp \rightarrow t\bar{t} \rightarrow 2\ell + X$	inclusive	POWHEG+PYTHIA8	A14NNPDF23LO	LO
$pp \rightarrow WZ \rightarrow 3\ell 1\nu$	inclusive	SHERPA2.2.2	NNPDF3.0NNLO	$NLO + 1j, LO + 3j$

Table 14: List of MC samples used in the estimation and validation of the data-driven non-prompt background estimation.

### 14.3 Event Weights

The predictions from the MC generators are often generated with a higher effective luminosity than the data to reduce the statistical uncertainties in simulations. Therefore, the raw predictions of the MC are completely unscaled and cannot be compared to the data recorded by the detector directly. Different weights need to be accounted for to scale each generated MC event.

The MC-generated events beyond LO have a generator event weight which is needed once a sufficient number of events are simulated to attain the correct cross-section distribution. Some generated events have negative weight to account for the cancellations caused by interference that arises during matrix element calculation. The generator event weight also accounts for the Sudakov form factors associated with QCD emissions in parton showering.

Each MC-generated event first needs to be scaled based on its generator event weight and normalized by the total sum of all the generated event weights. Once the event weight is multiplied by the cross-section and the integrated luminosity of the data, the distribution is correctly normalized.

A higher-order cross-section calculation might be available for some physics processes. Such predictions are re-weighted by scaling the event weight using the available *k-factor*. Moreover, the simulations of certain physics processes are computationally intensive, and MC generators might impose kinematic filters to simulate these processes. The efficiency of such filters,  $\epsilon_{filt}$ , are also considered in the event weight.

A set of corrections in the event weight are related to the detector measurements. As discussed in Section 7, the *scale factors (SF)* correct the efficiencies in MC related to the reconstruction, identification, isolation, and trigger to match that of the measured data. These scale factors are applied to the event weight. Finally, the pile-up re-weighting corrects the MC event weights to match the distribution of the average number of interactions per bunch crossing observed in the data and is also considered in the event weight.

Therefore, each event weight in MC can be represented as,

$$w_{event} = \frac{\sigma \cdot \text{k-factor} \cdot \epsilon_{filt} \cdot w_{generator} \cdot L_{data} \cdot w_{reco}}{\sum w_{generator}}, \quad (14.2)$$

where  $\sigma$  is the cross-section of the process,  $w_{generator}$  is the generator event weight,  $L_{data}$  is the integrated luminosity from data, and  $w_{reco}$  accounts for all detector measurement related scale factors and pile-up re-weighting in detector-level predictions.

## 15 Definition of Measured Observables

The primary results of this thesis are the unfolded differential cross-sections of the following eleven different kinematic observables:

- $m_{4\ell}$ : invariant mass of the four-lepton system (or 2  $Z$ -bosons),
- $m_{jj}$ : invariant mass of the dijet system,
- $p_{T,4\ell}$ : transverse momentum of the four-lepton system,
- $p_{T,jj}$ : transverse momentum of the dijet system,
- $p_{T,4\ell jj}$ : transverse momentum of the four-leptons and the dijet system,
- $s_{T,4\ell jj}$ : scalar transverse momentum of the four-leptons and the dijet system
- $\Delta\phi_{jj}^{signed}$ : difference in the azimuthal angle between the two jets, ordered according to their rapidity, i.e.

$$\Delta\phi_{jj}^{signed} = \begin{cases} \phi(j_1) - \phi(j_2) & \text{if } y_{j_1} > y_{j_2} \\ \phi(j_2) - \phi(j_1) & \text{otherwise,} \end{cases}$$

- $\Delta y_{jj}$ : the absolute value of rapidity difference between the leading and the sub-leading jets,
- $\zeta$ : centrality of the system,
- $\cos\theta_{\ell 1 \ell 2}^*$ : cosine of the decay angle of the negative lepton of the leading pair in the pair's rest frame as shown by Figure 31, and
- $\cos\theta_{\ell 3 \ell 4}^*$ : cosine of the decay angle of the negative lepton of the sub-leading pair in the pair's rest frame as shown by Figure 31.

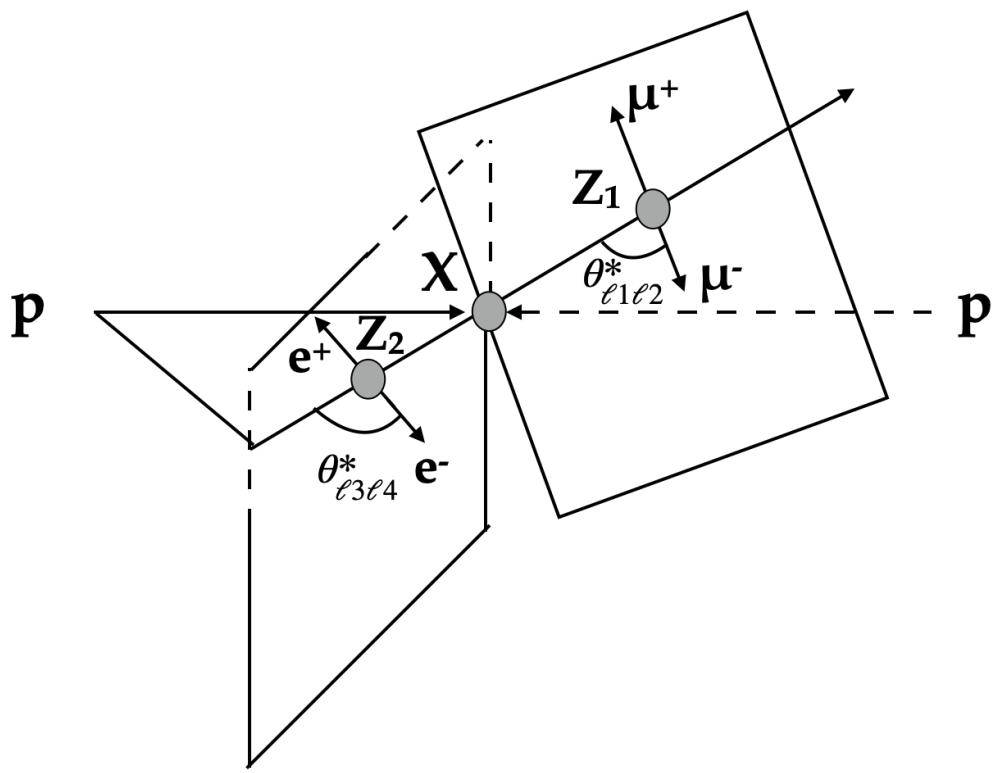


Figure 31: Figure showing the decay angle  $\theta_{\ell 1 \ell 2}^*$  ( $\theta_{\ell 3 \ell 4}^*$ ) of the negative lepton in the primary (secondary) pair's rest frame. [21].

## Chapter v: Analysis Strategy

### 16 Background Estimation

In addition to the three signal physics processes (parton initiated  $qqZZ$ , gluon-loop initiated  $ggZZ$  and EWK  $qqZZjj$ ), background processes such as  $t\bar{t}V$ ,  $VVV$  and non-prompt background comprising of *non-prompt* leptons also contribute to the  $ZZ^*(\rightarrow 4\ell)jj$  final state. As discussed in Section 14.2, the events originating from  $t\bar{t}V$  and  $VVV$  processes are simulated using MC generators and subtracted from data. The non-prompt backgrounds are estimated using a data-driven approach discussed below.

#### 16.1 Data Driven Estimate of Non-Prompt Background

*Non-prompt leptons* originate from a non-hard scatter source, either from a secondary interaction, such as a jet decay or from charged tracks misidentification. Figure 32 shows an example of non-prompt lepton production, where a b-jet from hard scattering produces a muon whose track does not point towards the interaction vertex and is surrounded by jet activities. The signal lepton criteria of isolation and TTVA discussed in Section 10 discards most of the non-prompt leptons. However, some non-prompt leptons pass the signal criteria and, in association with other prompt leptons, form a quadruplet in the signal region, giving rise to the *fake background* or *non-prompt background* events in the analysis. The origins of non-prompt leptons are discussed in detail in Section 16.1.1.

The non-prompt backgrounds could be predicted using the MC for  $Z(\rightarrow \ell\ell) + jets$ ,  $t\bar{t} + X$  and  $WZ$  processes where one or more non-prompt leptons in association with the prompt leptons form a signal quadruplet. However, it is difficult to precisely predict the statistically limited contribution of non-prompt backgrounds using MC generators. It is also challenging

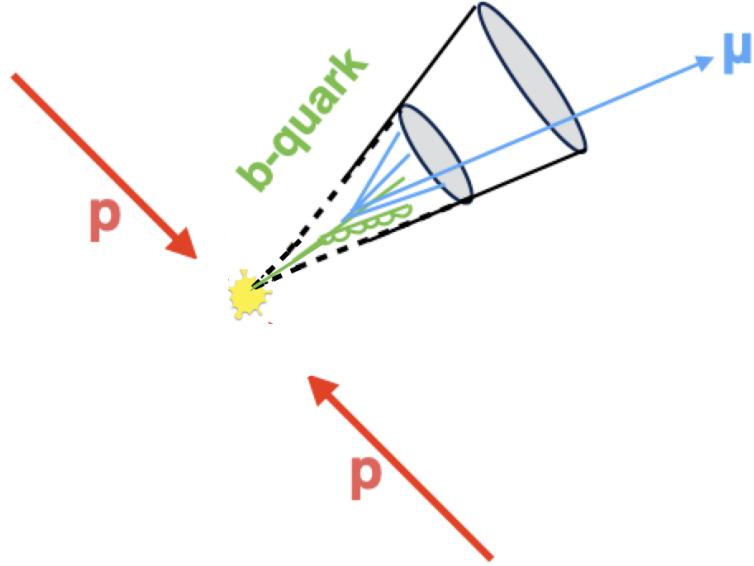


Figure 32: A schematic of the non-prompt lepton from semi-leptonic decays of b-hadrons. Jet activities surround the non-prompt muon, and the muon track does not point to the hard scatter interaction point.

to precisely model the non-prompt leptons originating from the reconstruction effects, such as misidentification and photon conversion. Therefore, the non-prompt backgrounds are estimated using an entirely data-driven technique discussed in this Section.

Figure 33 illustrates the non-prompt background estimation process. The *fake factor*( $F$ ), which is defined as the ratio between the number of non-prompt leptons that pass the signal requirements to the number of non-prompt leptons that fail the signal requirements, is estimated from a combined control region (CR). The combined control region is formed by combining two independent control regions  $Z+jets$ , and  $t\bar{t}$ , which are enriched in non-prompt leptons, and the combination is discussed in Section 16.1.2. The non-prompt background is estimated from the non-signal quadruplets by applying the fake factors to each *anti-signal lepton*, baseline lepton that fails the TTVA or the isolation requirements. Section 16.1.3 discusses the technical aspects of the fake factor method, and Section 16.1.4 discusses the fake efficiencies used to derive the fake factors. The non-prompt background estimated

from the fake-factor method is first validated in fake-enriched validation regions discussed in Section 16.1.5 and estimated for the signal region, which is discussed in Section 16.1.6.

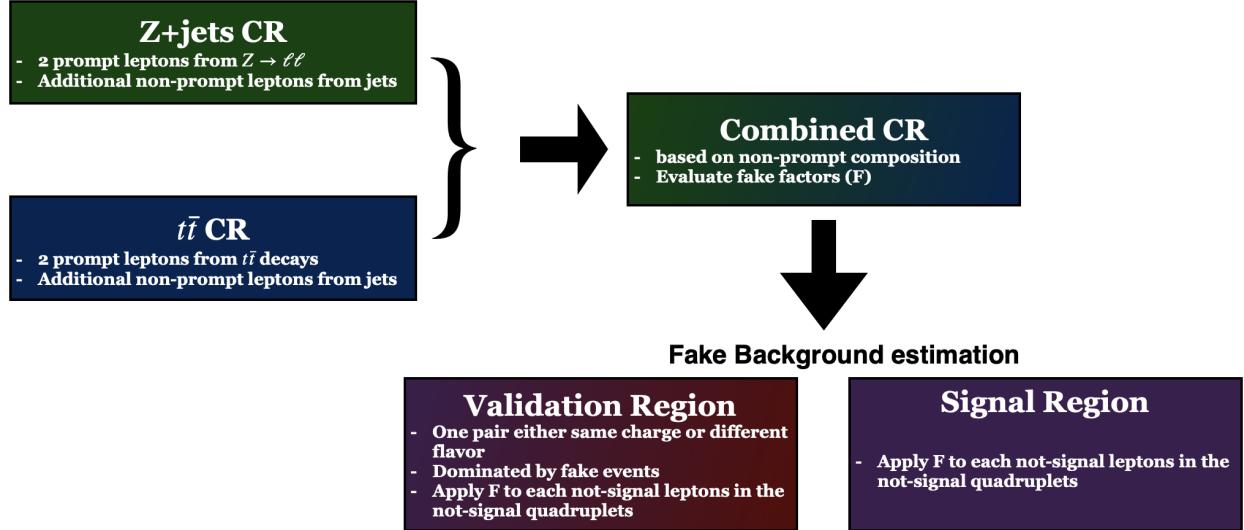


Figure 33: An overview of the non-prompt background estimation.

### 16.1.1 Lepton Composition

The non-prompt background MC predictions provide essential insights into the origin of the non-prompt leptons. A classification tool<sup>5</sup> developed by the ATLAS Isolation and Fake Forum (IFF) identifies the true origin of the leptons, which is studied to understand the composition of non-prompt leptons in various phase-space regions of the analysis. The tool has the following classification of truth origin for a non-prompt lepton:

- *Unknown or KnownUnknown*: leptons with insufficient truth-level information to be classified by the tool.
- *IsoElectron*: electrons originate either from the hard scatter or a boson decay. These electrons are treated as prompts in signal and background control regions.
- *ChargeFlipIsoElectron*: electrons whose charge is mismeasured at detector-level and is classified as non-prompts.

<sup>5</sup><https://gitlab.cern.ch/atlas/athena/-/tree/21.2/PhysicsAnalysis/AnalysisCommon/TruthClassification>

- *PromptMuon*: muons originate from either the hard scatter or a boson decay. These muons are treated as prompts in signal and background control regions.
- *PromptPhotonConversion*: non-prompt electrons originating from photon conversion.
- *TauDecay*: leptons originating from tau decays are treated as prompt leptons.
- *BHadronDecay*: leptons originating from hadrons containing a b-quark. These types of leptons are one of the primary sources of non-prompt leptons.
- *CHadronDecay*: leptons originating from hadrons containing a c-quark.
- *LightFlavourDecay*: leptons originating from mesons and lighter hadrons.

Figure 34 shows the origin of all leptons that are part of the quadruplet in the events with a signal quadruplet and a dijet. Most of the leptons in these regions are prompt and predominantly originate from  $ggZZ$ ,  $qqZZ$ , and EWK  $qqZZjj$  processes. The leptons that are classified *Unknown/KnownUnknown* mainly originate from  $t\bar{t}Z(\rightarrow \ell\ell)$  and  $VVV$  physics processes. Due to the technical implementation of these physics processes in MC simulations, information on the intermediary bosons for these samples is unavailable, thus, failing the identification of lepton’s origin. The *Unknown/KnownUnknown* leptons are treated as prompt leptons in the signal region. This treatment relies on the fact that  $\Delta R$  between the *Unknown/KnownUnknown* classified truth leptons and reconstruction level lepton is observed to be close to 0. The *Unknown/KnownUnknown* classified leptons are treated as non-prompt leptons in the background control regions.

Figure 35 shows the predicted fraction of non-prompt electrons (left) and non-prompt muons (right) in the events with a signal quadruplet and a dijet. The non-prompt leptons originating from *b*-hadrons or *c*-hadrons are collectively called *heavy flavor (HF)* non-prompt leptons, whereas all other non-prompt leptons are categorized as *light flavor (LF)*. About 50% of non-prompt electrons in the signal region originate from heavy flavor sources, whereas more than 90% of non-prompt muons originate from the heavy flavor decays.

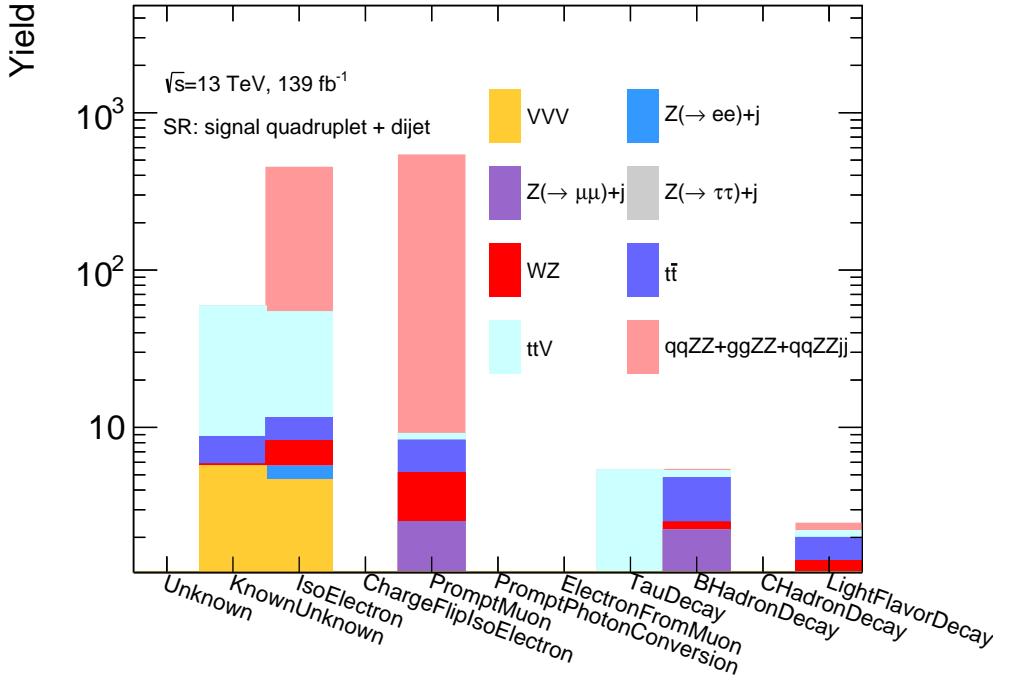


Figure 34: Origins of leptons in the signal region in events with a quadruplet and a dijet. The lepton origin is classified by the IFF classifier tool. Only leptons that are part of the signal quadruplet are shown.

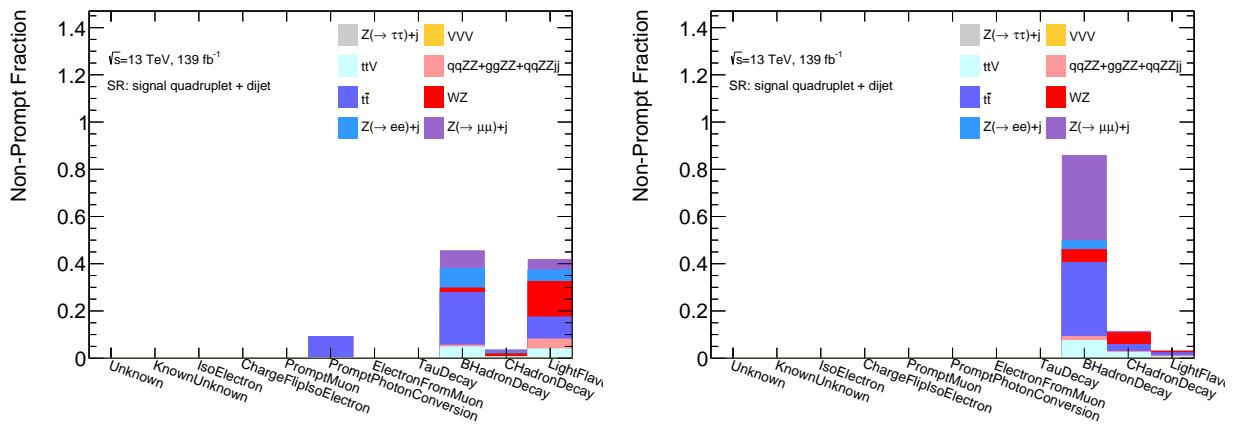


Figure 35: Origins of non-prompt electrons (left) and muons (right) in the signal region in events with a signal quadruplet and a dijet. The events are normalized to the total number of non-prompt electrons (left) and non-prompt muons (right).

### 16.1.2 Control Regions

The fake factors are measured from data in the fake enriched background control region. The control region is formed by combining two independent control regions, the  $Z + jets$  control region and the  $t\bar{t}$  control region. An event in the control regions consists of a prompt lepton pair from a physics process such as a boson or top-quark decay and additional leptons from non-prompt sources. Both control regions use a single or di-lepton trigger similar to the signal region and require the leading and sub-leading leptons in an event to satisfy  $p_{T, \text{leading lepton}} > 20 \text{ GeV}$  and  $p_{T, \text{sub-leading lepton}} > 15 \text{ GeV}$ . An event in the  $Z + jets$  CR consists of an SF-OC prompt-lepton pair from the  $Z$  boson decay with an invariant mass of  $76 \text{ GeV} < m_{\ell\ell} < 106 \text{ GeV}$ , and additional leptons. Additionally, no events can have missing transverse energy higher than 50 GeV to suppress the contamination from the  $WZ$  process. Similarly, the  $t\bar{t}$  CR consists of events with different flavor prompt-lepton pairs and additional leptons. An event in the  $t\bar{t}$  CR requires at least one b-tagged jet to reduce the  $WZ$  contamination. The b-tagging in the  $t\bar{t}$  CR is performed by a flavor tagging tool described in Ref [88].

Figure 36 shows the fractions of additional electrons (left) and muons (right) originating from non-prompt sources as a function of their  $p_T$  in the  $Z + jets$  CR and the  $t\bar{t}$  CR. The fake fraction is defined as a ratio between the number of additional-baseline leptons (not from  $Z$  boson or top-quark decay) that are classified as non-prompt to the total number of additional-baseline leptons. A high fraction ( $\geq 80\%$ ) of baseline electrons originate from non-prompt sources in both  $Z + jets$  CR and  $t\bar{t}$  CR. More than 95% of the low- $p_T$  baseline muons are from non-prompt sources in both control regions. These distributions show that most of the additional leptons in either control region are expected to be from non-prompt sources, thus, motivating the control regions to evaluate the fake factors.

The control regions have a unique non-prompt lepton composition as shown by Figures 37 and 38. More than 80% of the non-prompt electrons in the  $Z + jets$  CR originate from the light flavor decays, but about 60% are from the light flavor decays in the  $t\bar{t}$  CR. Similarly,

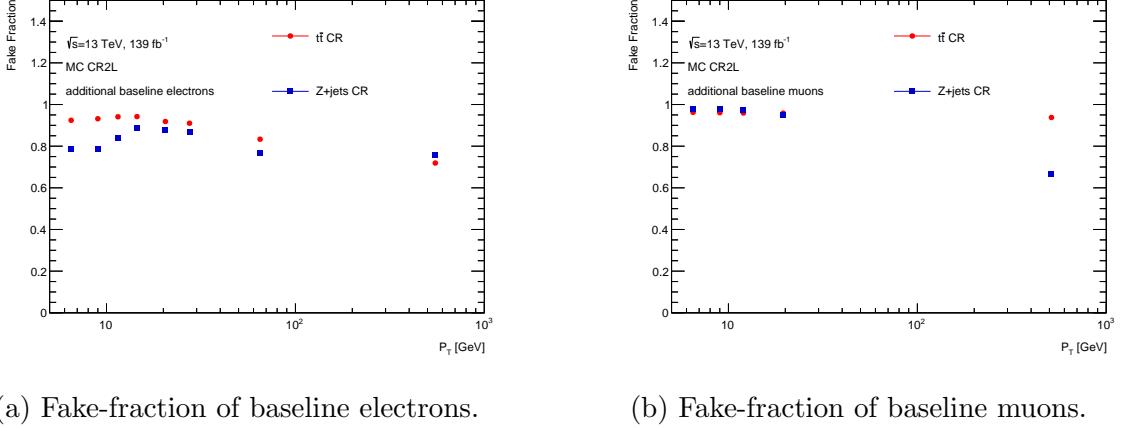


Figure 36: Fraction of non-prompt electrons and muons in the  $Z + jets$  and  $t\bar{t}$  control regions as a function of their  $p_T$ .

about 80% of the non-prompt muons in the  $Z + jets$  CR originate from the heavy flavor, whereas more than 90% are from the heavy-flavor decays in the  $t\bar{t}$  CR. The non-prompt compositions of the signal region shown in Figure 35 are different from either control region. The non-prompt leptons from different sources have different probabilities of satisfying the signal lepton requirements. Thus, the two independent control regions are combined to form a single control region whose non-prompt lepton composition matches that of the signal region.

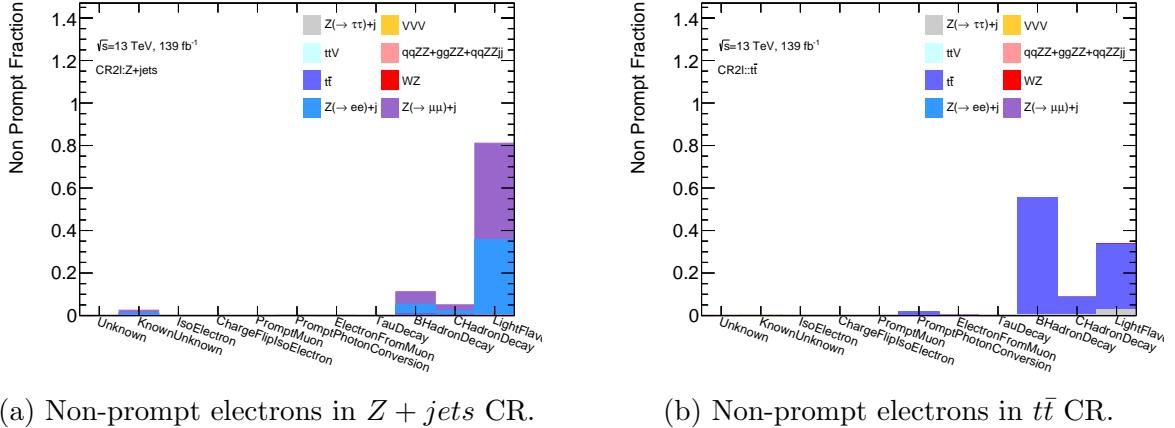


Figure 37: Sources of non-prompt electrons in  $Z + jets$ (left) and  $t\bar{t}$ (right) control regions.

The  $b$ -jet requirement applied to suppress the prompt-lepton contamination from the WZ process in  $t\bar{t}$  CR ensures the presence of at least one jet in all events. Therefore, events

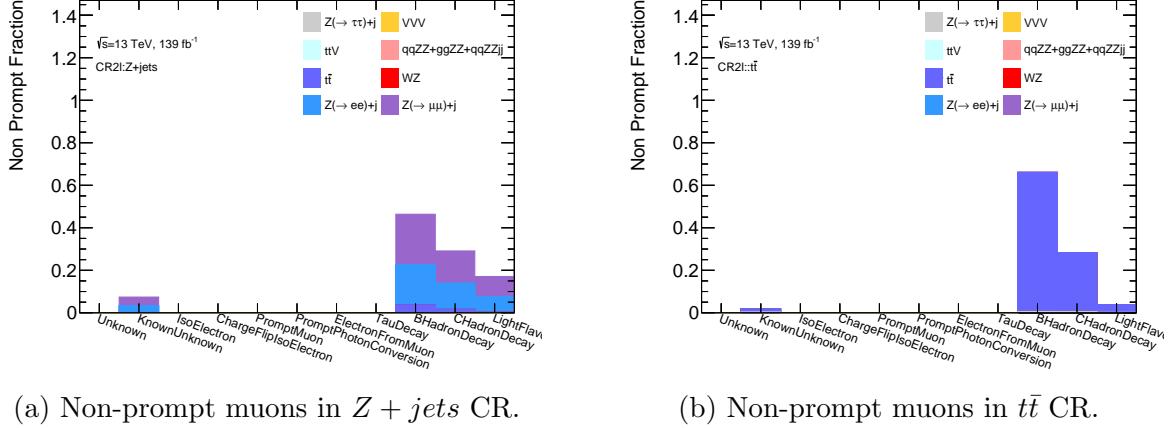


Figure 38: Sources of non-prompt muons in  $Z + jets$ (left) and  $t\bar{t}$ (right) control regions.

without jets in the combined control region only consist of  $Z + jets$  events. The two control regions are first weighted and combined for the events with the jets to match the heavy flavor composition of the  $n_{jet} > 0$  events in the signal region. The combination weights are evaluated by solving the following equation,

$$\frac{[w \cdot N_{Z+jets} \cdot f_{HF,Z+jets}] + [(1-w) \cdot N_{t\bar{t}} \cdot f_{HF,t\bar{t}}]}{[w \cdot N_{Z+jets} + (1-w) \cdot N_{t\bar{t}}]} = f_{HF,SR}, \quad (16.1)$$

where  $N$  is the total yield in the control region,  $f_{HF}$  is the ratio of the non-prompt leptons from heavy-flavor decays to total non-prompt leptons, and  $w$  is the combination weight to be determined.

As the composition of non-prompt electrons and muons are different in different regions, the weights are evaluated separately for electrons and muons and estimated to be  $w_\mu = 0.26$  and  $w_e = 0.06$ . Figure 39 shows the composition of the non-prompt electrons and muons in the combined control region, which is formed by a weighted combination of the  $Z + jets$  CR and the  $t\bar{t}$  CR.

Figures 40a and 40b show the distributions of additional baseline electrons as a function of their  $p_T$  in the  $Z + jets$  CR and  $t\bar{t}$  CR, respectively, whereas, Figure 40c show the same for the combined control region. For  $Z + jets$  CR at low  $p_T$ , additional baseline electrons are overestimated in MC by about 20%, thus showing that the precision of MC to estimate the

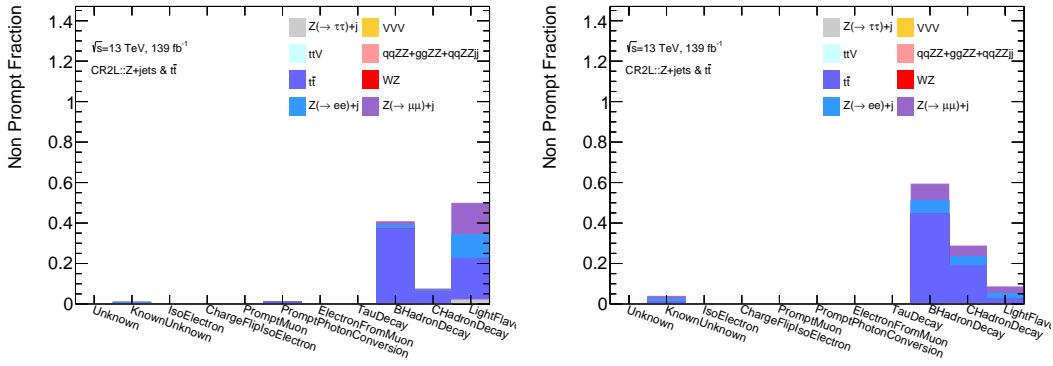


Figure 39: Origins of non-prompt electrons (left) and muons (right) in the combined control region.

non-prompt leptons is limited. Similarly, Figures 41a, 41b and 41c show the distributions of additional baseline muons as a function of their  $p_T$  in the three control regions. In the  $Z + jets$  CR, the low  $p_T$  additional muons mainly originate from  $Z \rightarrow \ell\ell$  process, whereas the high  $p_T$  events, originate more significantly from  $t\bar{t}$  and WZ processes.

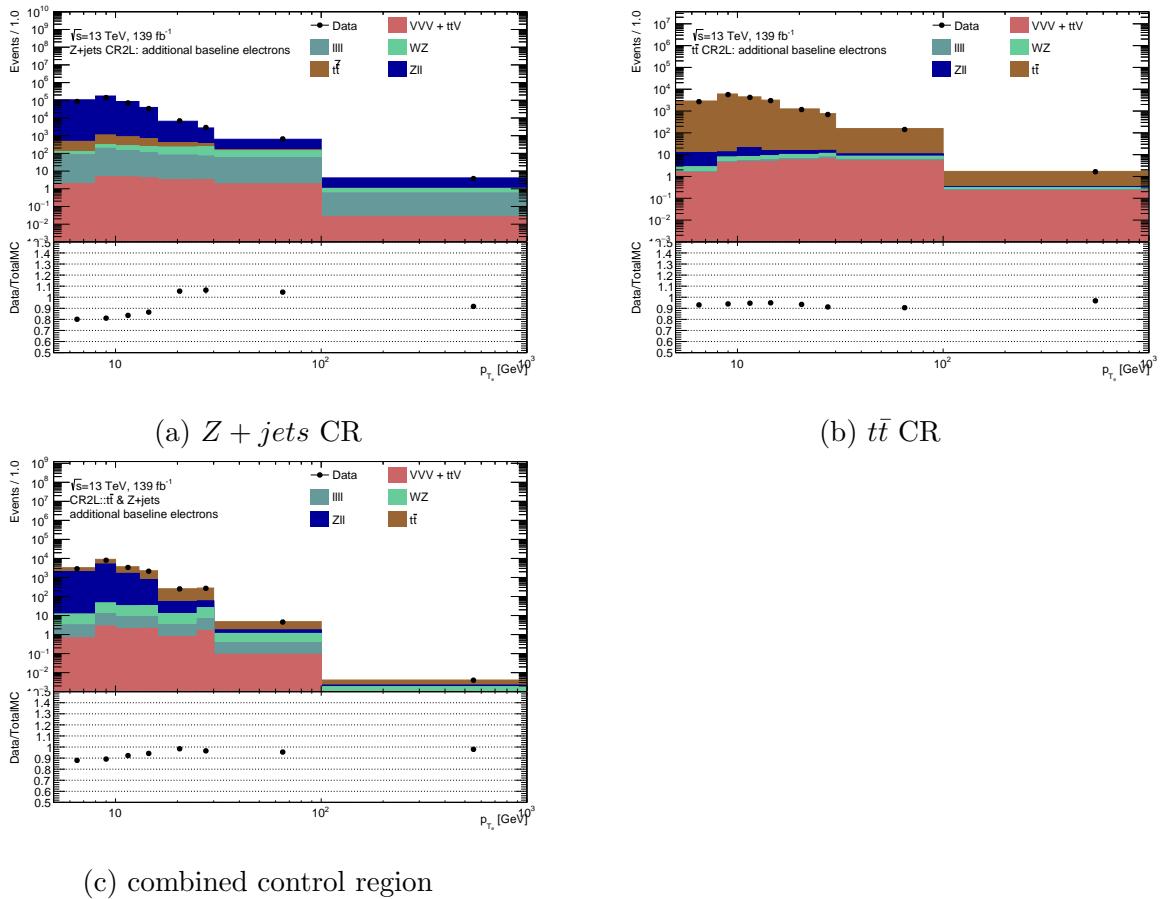


Figure 40: Additional baseline electrons as a function of  $p_T$  in control regions.

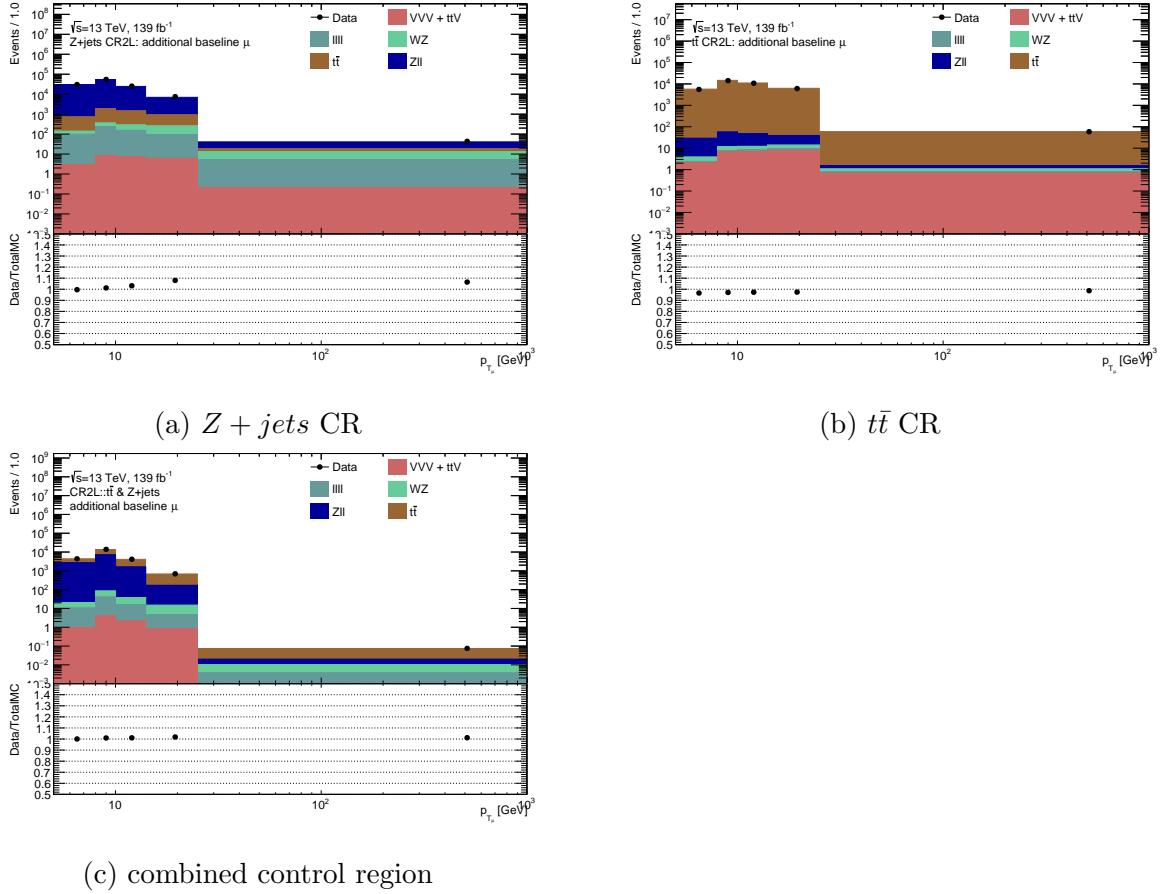


Figure 41: Additional baseline muons as a function of  $p_T$  in control regions.

### 16.1.3 Fake Factor Strategy

The *fake factor tool (FF tool)* is used to estimate the non-prompt backgrounds [89]. The tool takes as input the *fake efficiency* ( $f$ ), which is defined as the ratio between the number of non-prompt leptons that passes the signal selection to the total number of the non-prompt baseline leptons

$$f = \frac{N_{\text{non-prompt signal leptons}}}{N_{\text{non-prompt baseline leptons}}}, \quad (16.2)$$

where the fake efficiency is calculated in the combined control region.

For a simple case of a signal region with one signal lepton, the observed number of signal lepton ( $N^T$ ) and baseline-anti-signal lepton ( $N^L$ ) can be estimated in terms of the number of prompt or real baseline leptons ( $N_R^B$ ) and the number of non-prompt or fake baseline leptons ( $N_F^B$ ) as

$$N^T = rN_R^B + fN_F^B, \quad (16.3)$$

and

$$N^L = (1 - r)N_R^B + (1 - f)N_F^B, \quad (16.4)$$

where  $r$  is the *real efficiency* defined as the ratio between the number of the signal to the baseline prompt leptons,  $N_{\text{prompt signal leptons}}$  and  $N_{\text{prompt baseline leptons}}$ , respectively, such that

$$r = \frac{N_{\text{prompt signal leptons}}}{N_{\text{prompt baseline leptons}}}. \quad (16.5)$$

Equations 16.3 and 16.4 can be written as a  $2 \times 2$  matrix equation as

$$\begin{pmatrix} N^T \\ N^L \end{pmatrix} = \begin{pmatrix} r & f \\ 1 - r & 1 - f \end{pmatrix} \begin{pmatrix} N_R^B \\ N_F^B \end{pmatrix}. \quad (16.6)$$

The number of classified non-prompt baseline leptons can be estimated by ignoring the

higher-order term of the fake efficiency as

$$N_F^B = \frac{1}{r-f} [r(N^T + N^L) - N^T]. \quad (16.7)$$

Therefore, the predicted number of classified non-prompt signal lepton is

$$N_F^T = \frac{f}{r-f} [r(N^T + N^L) - N^T]. \quad (16.8)$$

The fake factor method assumes the  $r \rightarrow 1$  limit, which simplifies Equation 16.8. However, since the real efficiency of any measurement is less than one, this approximation overestimates the non-prompt background. To account for this overestimation, the number of genuine baseline-anti-signal prompt leptons ( $N_R^L$ ) are measured in MC and subtracted to get the final background yield as

$$N_F^T = \frac{f}{1-f} [N^L - N_R^L], \quad (16.9)$$

where, the coefficient

$$F = \frac{f}{1-f}, \quad (16.10)$$

is the fake factor  $F$ . The method makes a typically safe assumption that the real anti-signal prompt leptons are modeled precisely in MC. As the fake efficiency  $f$  is estimated from data in the combined control region, the fake factor background estimation method does not rely on any efficiencies or predicted yield in the tight signal region.

This method can be extended to the four-lepton signal region where there are four baseline leptons, of which one or more could be non-prompt. Corresponding to the permutation of individual leptons to be either signal or baseline-anti-signal, there are  $2^4 = 16$   $\{N^{TTTT}, N^{TTTL}, N^{TTLL}, \dots, N^{LLLL}\}$  observations to consider. The analysis considers  $N^{TTTT}$  the signal region, while the background is estimated from quadruplets with at least one baseline-anti-signal lepton.

#### 16.1.4 Fake Efficiency and Systematics

The fake efficiency defined in Section 16.1.3 is evaluated from the combined control region using the total number of additional leptons from data as

$$f = \frac{N_{Data}^{Signal} - N_{MC}^{Prompt\ Signal}}{N_{Data}^{Baseline} - N_{MC}^{Prompt\ Baseline}}, \quad (16.11)$$

where  $N_{Data}^{Signal}$  and  $N_{Data}^{Baseline}$  are the numbers of signal and baseline leptons, respectively, from data. Similarly,  $N_{MC}^{Prompt\ Signal}$  and  $N_{MC}^{Prompt\ Baseline}$  are the numbers of signal and baseline leptons, respectively, from MC that originate from prompt lepton sources. These subtracted prompt leptons from MC prediction account for the possible additional leptons that could originate from prompt sources.

Figures 36, 40, and 41 show that the fake-fraction and the total yield of the additional leptons are dependent on their transverse momentum  $p_T$ . Therefore, the fake efficiency evaluated using Equation 16.11 depends on the lepton  $p_T$ . Because of the low resolution of the detector in forward regions, a higher number of non-prompt leptons are expected in this region; thus, the fake efficiency depends on the leptons' pseudorapidity ( $\eta$ ). Additionally, since the non-prompt leptons predominantly originate from jets, the fake efficiency also depends on the number of jets ( $n_{jets}$ ) in an event.

Figures 42 and 43 show the fake efficiencies for electrons and muons, respectively, as a function of  $p_T$  (top-left),  $\eta$  (top-right) and  $n_{jets}$  (bottom-center). For electrons, the fake efficiency first decreases with  $p_T$  up to 20 GeV, then increases. Since the high- $p_T$  muons are most likely to originate from a prompt source, fake efficiency typically decreases as a function of  $p_T$  for the muons. The dependencies on  $\eta$  for both muons and electrons are most likely from lower detector resolution in high  $\eta$  regions, which causes a higher number of misidentifications and lower efficiency for TTVA.

As discussed in Section 10.4, the lepton-favored overlap removal used in the analysis rejects jets if they overlap with leptons. Due to the  $b - jet$  requirement in  $t\bar{t}$  CR, the

$n_{jet} = 0$  events only consist of contributions from the  $Z + jets$  CR, which does not have an explicit event requirement on the number of jets. The probability of non-prompt leptons passing the isolation requirement is higher in events with no jets or surrounding hadronic activity. Therefore, as observed, higher fake efficiency is expected in events without jets.

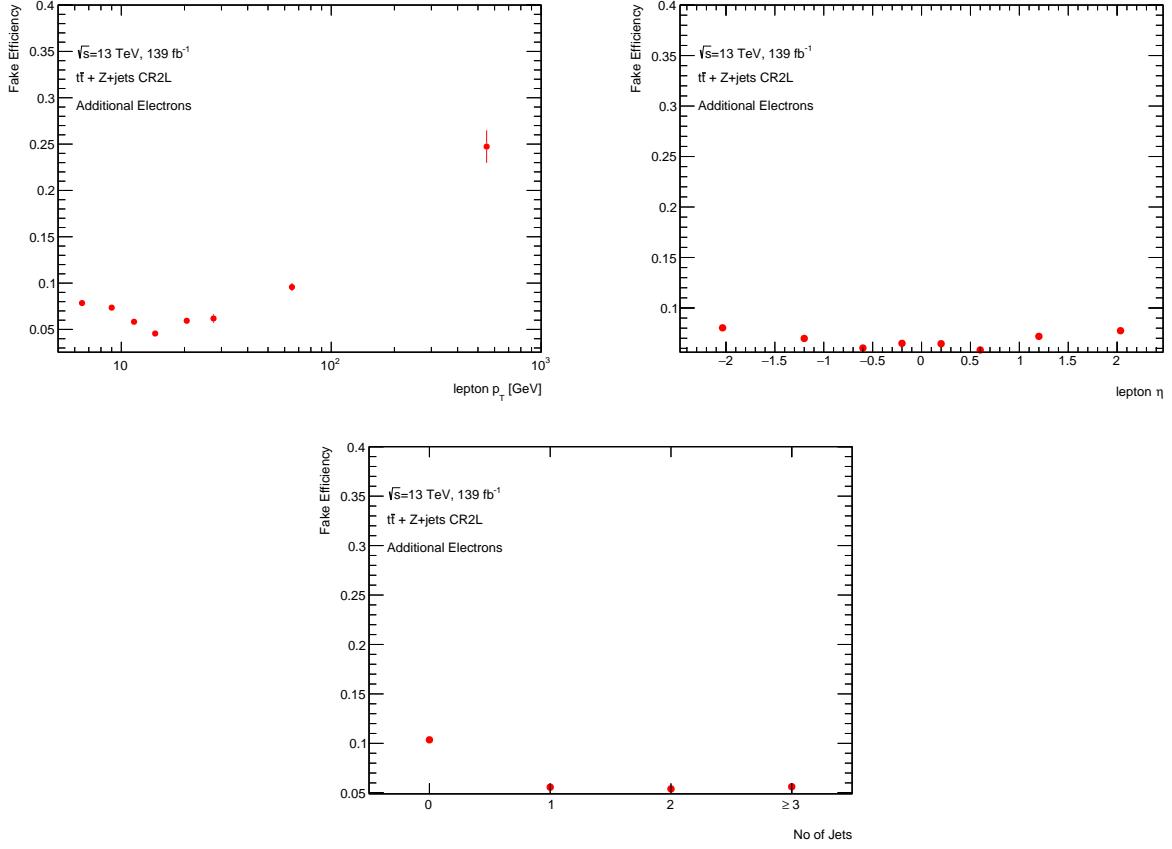


Figure 42: Fake efficiencies of non-prompt electrons measured in the combined control region from data as a function of their  $p_T$ ,  $\eta$ , and  $n_{jets}$ .

The final fake efficiencies in the data-driven non-prompt background estimate are parametrized in three-dimensional distributions of  $p_T$ ,  $\eta$ , and  $n_{jets}$ . Only two bins ( $n_{jet} = 0$  and  $n_{jet} > 0$ ) are used for the number of jets to reduce statistical fluctuations. Figures 44a and 44b show the fake efficiency of electrons as a function of  $p_T$  and  $\eta$  for  $n_{jet} = 0$  and  $n_{jet} > 0$  bins, respectively. Similar distributions are shown in Figures 45a and 45b for muons.

The fake efficiency distributions' binomial errors are propagated as the statistical un-

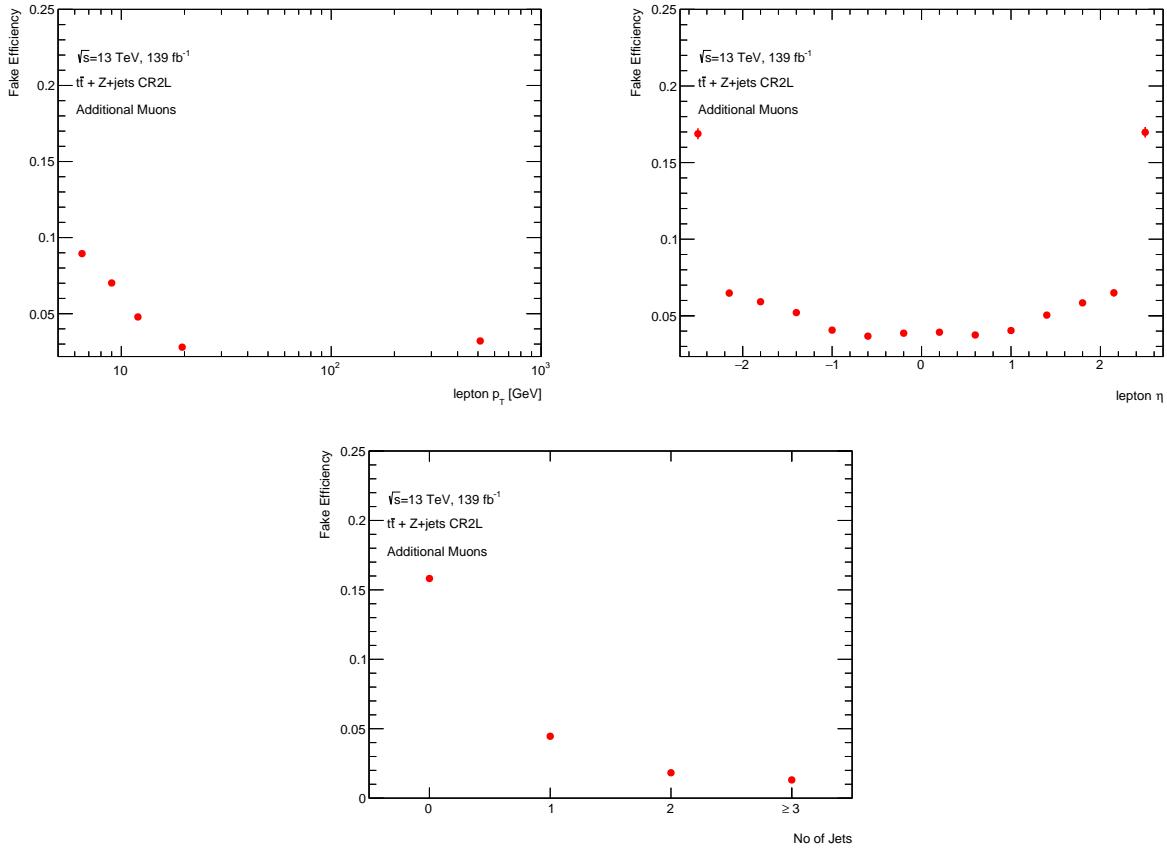


Figure 43: Fake efficiencies of non-prompt muons measured in the combined control region from data as a function of their  $p_T$ ,  $\eta$ , and  $n_{jets}$ .

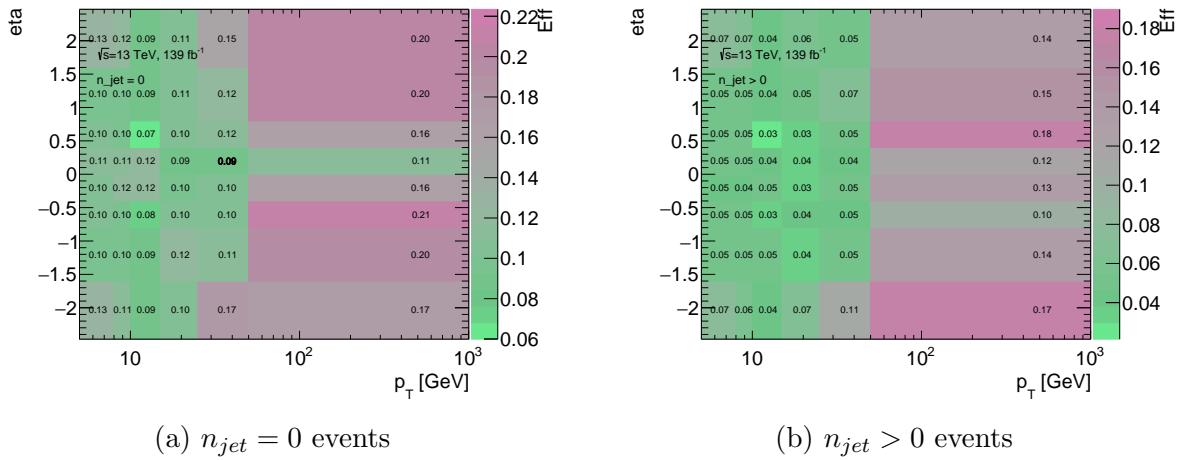


Figure 44: Fake efficiencies of electrons measured in the combined control region from data as a function of their  $p_T$  and  $\eta$  in two slices of  $n_{jets}$ .

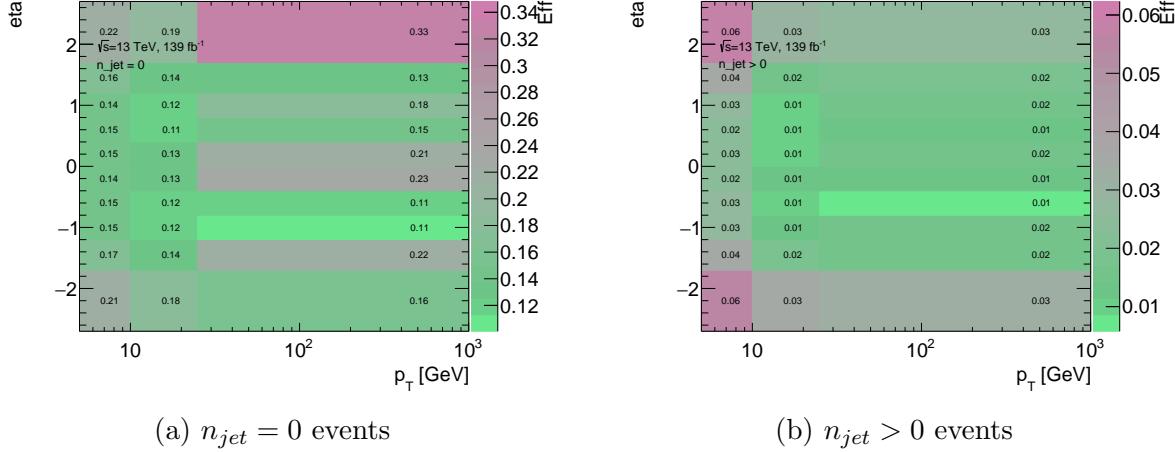


Figure 45: Fake efficiencies of muons measured in the combined control region from data as a function of their  $p_T$  and  $\eta$  in two slices of  $n_{jets}$ .

certainties on the fake estimate. The subtracted prompt component of Equation 16.11 is estimated using MC predictions. As discussed in Section 3, the prediction relies on the PDF, the energy-dependent QCD factorization and renormalization scale, and the strong coupling constant ( $\alpha_S$ ). Therefore, the theory uncertainties on these three parameters are propagated as systematic uncertainties of the fake efficiency.

For each theory uncertainty, a variation-applied fake efficiency is evaluated by separately varying the numerator and denominator of the fake efficiency in Equation 16.11. The difference between the variation-applied fake efficiency and the nominal fake efficiency is considered as a systematic uncertainty. Figures 46 and 47 show the statistical and systematic uncertainties on the fake efficiencies for electrons and muons, respectively, as a function of their  $p_T$ ,  $\eta$ , and  $n_{jets}$  calculated in the combined control region. For both electrons and muons, the statistical uncertainty is dominant.

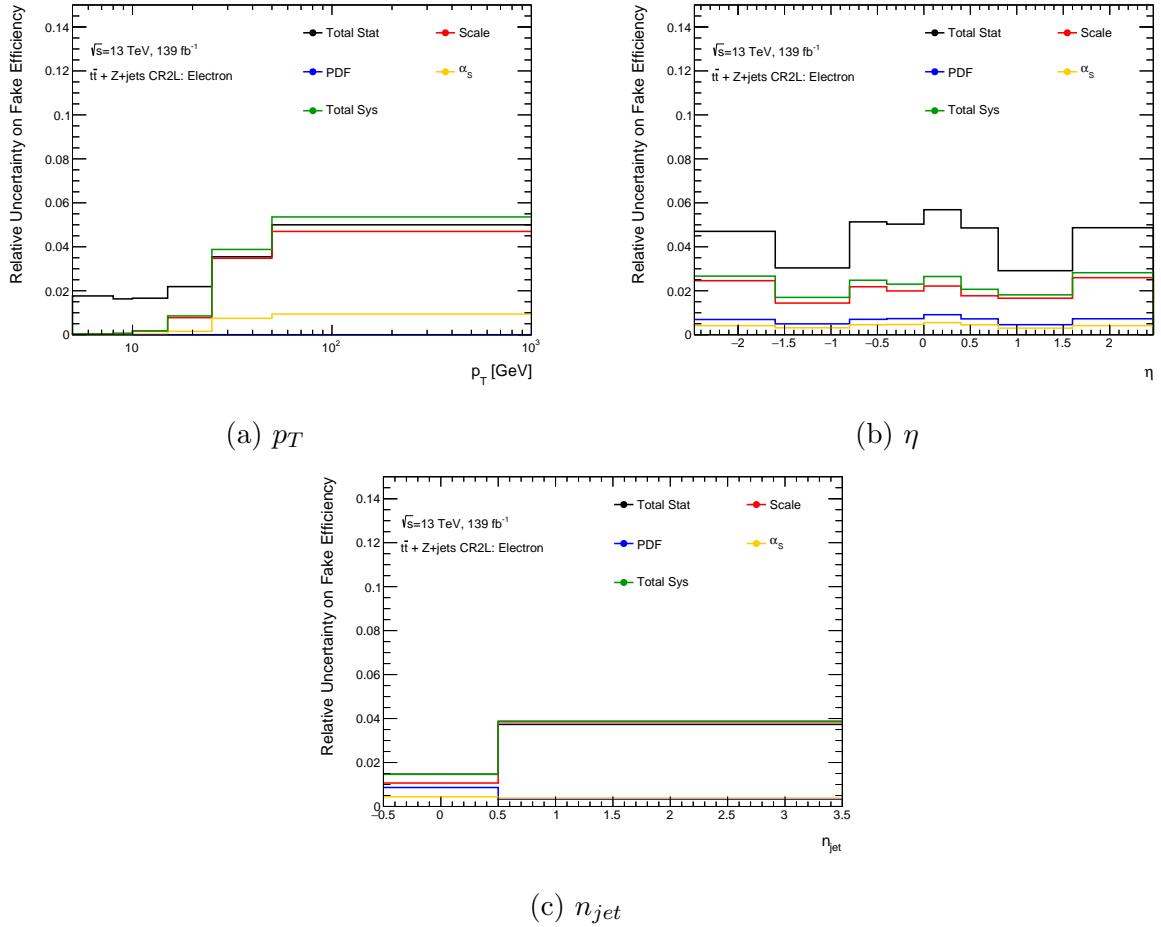


Figure 46: Uncertainties on the fake efficiency of the fake electrons measured in the combined control region from data as a function of their  $p_T$ ,  $\eta$ , and  $n_{jets}$ .

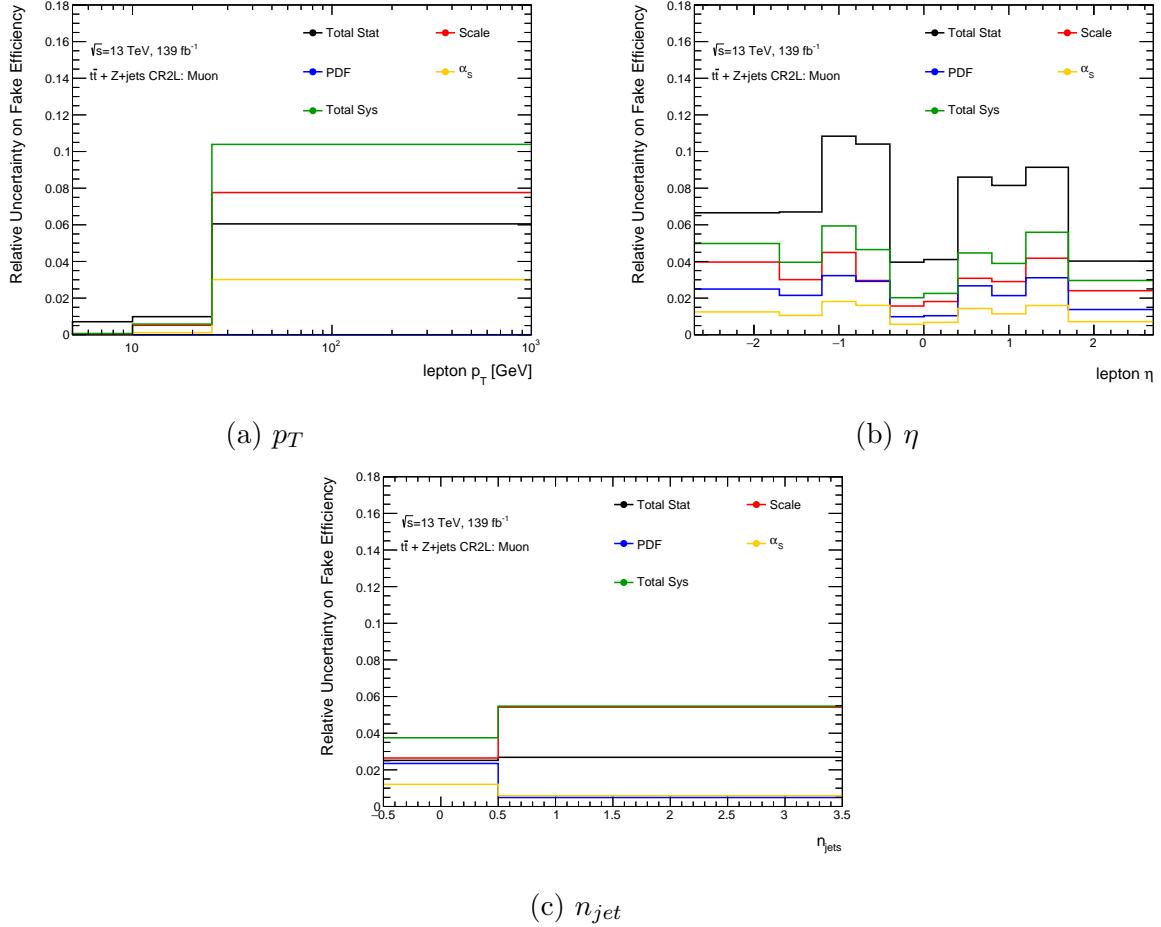


Figure 47: Uncertainties on the fake efficiency of the fake muons measured in the combined control region from data as a function of their  $p_T$ ,  $\eta$ , and  $n_{jet}$ .

### 16.1.5 Method Validation

Before implementing the fake-factor method to estimate the non-prompt background in the signal region, the method is validated in two separate validation regions:

1. Different-flavor validation region (VRDF): one pair in the quadruplet must have two different-flavor leptons;
2. Same-charge validation region (VRSC): one pair in the quadruplet must have two same-charge leptons.

The low statistics in both regions result from requiring one of the pairs in the quadruplet to be either same-charge or different-flavor. Therefore, only signal quadruplets are required in the validation regions events without any dijet requirement. The validation regions' quadruplets are formed by requiring the same kinematic criteria as that of the signal region discussed in Section 12. The trigger requirement, object selection, and overlap removal are identical to the signal region. Additionally, events in the VRDF are required not to have any b-tagged jet to reduce the contribution from  $t\bar{t}Z$  processes. Reducing the  $t\bar{t}Z$  component further reduces the significant modeling uncertainties related to the  $t\bar{t}Z$  process.

By imposing either a same-charge or a different-flavor lepton pair, the event yield in validation regions is dominated by events where at least one lepton in the quadruplet is a non-prompt-signal lepton known as the non-prompt background in the signal region. The events also originate from other physics processes, such as  $qqZZ$ ,  $qqZZjj$ ,  $ggZZ$ ,  $t\bar{t}Z$ , and  $VVV$  whose contribution is predicted by the same MC generators as that of the signal region.

Figures 48a and 48b show the non-prompt compositions in the different-flavor validation region for non-prompt electrons and muons, respectively. Similarly, Figures 48c and 48d show the non-prompt compositions of the non-prompt electrons and muons, respectively, in the same-charge validation region. The non-prompt compositions in the two validation regions are different from that of the signal region shown in Figure 35 or that of the background control regions composition shown in Figures 37 and 38. Therefore, to validate the non-

prompt background estimation strategy, observing a good correspondence between data and a combination of the total MC prediction with the non-prompt background yield in both validation regions is imperative.

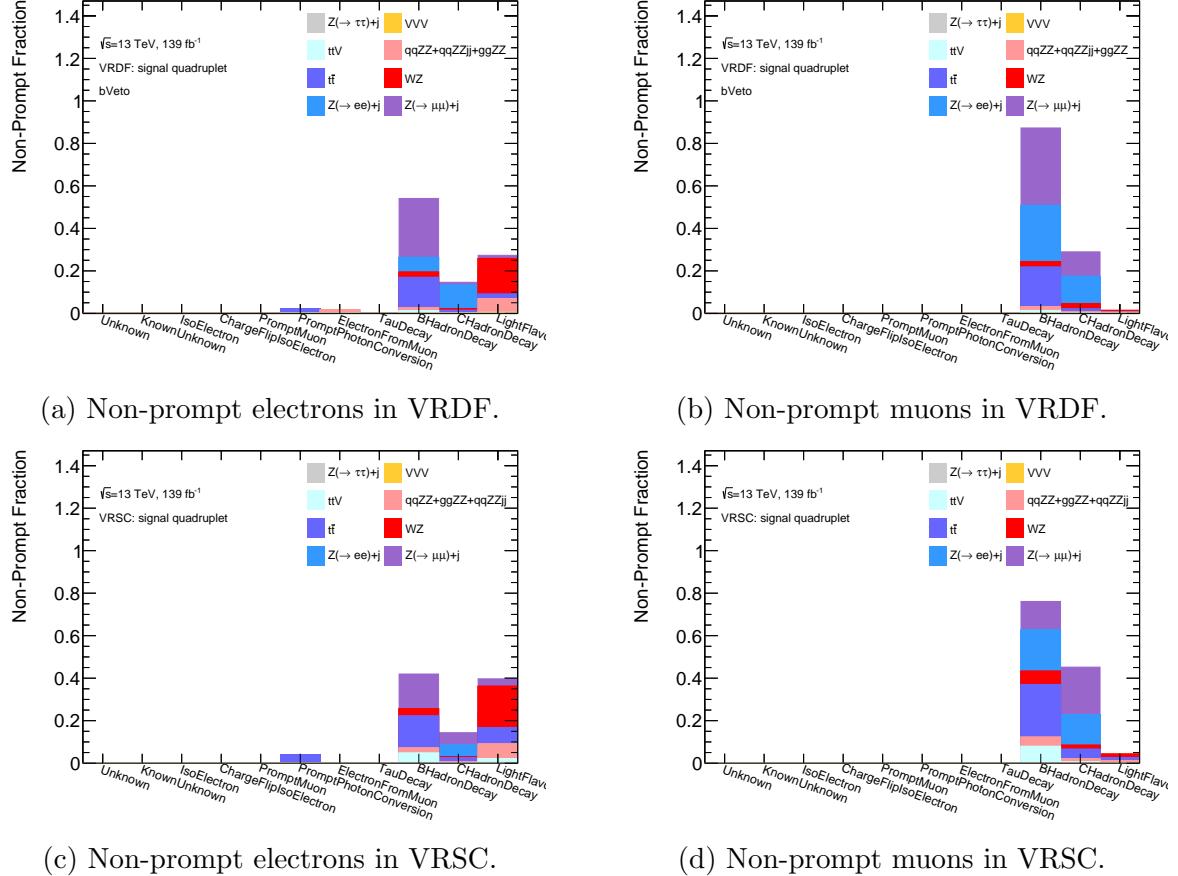
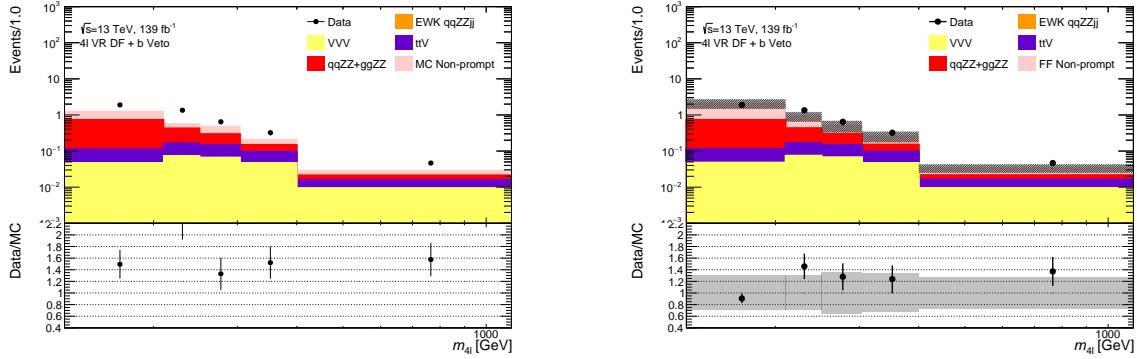


Figure 48: Sources of non-prompt electrons and muons in the different-flavor and the same-charge validation regions.

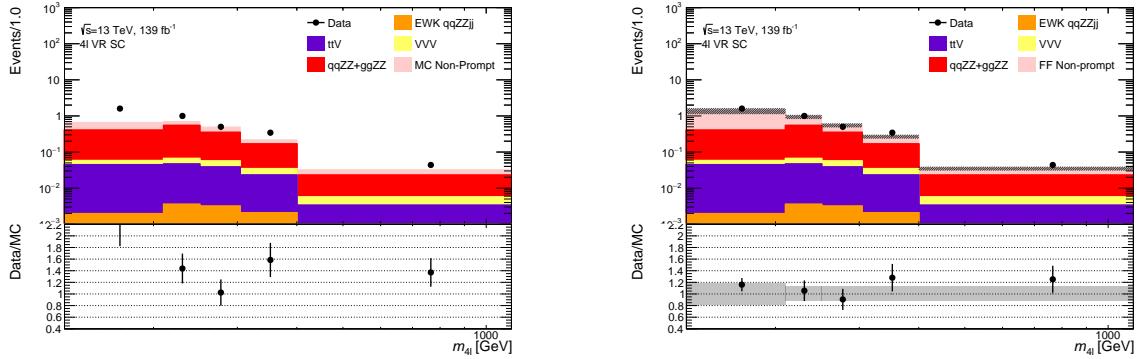
The non-prompt backgrounds for the validation regions are estimated by applying the fake factor to each baseline-anti-signal leptons in the non-signal quadruplet, as discussed in Section 16.1.3. Figure 49a shows the data and the predicted MC yield in VRDF as a function of  $m_{4\ell}$  where the non-prompt backgrounds are estimated from  $Z + jets$ ,  $t\bar{t}$ , and  $WZ$  MC predictions. Figure 49b shows the same, but the non-prompt background events are estimated using the fake factor method. Similarly, Figures 49c and 49d show the yields as a function of  $m_{4\ell}$  in VRSC. Both regions have similar characteristics, and the non-prompt background dominates the event yield with some contribution from other physics processes.

The systematic gray bands in Figures 49b and 49d include the total systematic and statistical uncertainties from the fake factor method, as well as the uncertainties on PDF, QCD scale, and strong coupling ( $\alpha_S$ ) on the  $qqZZ$ ,  $qqZZjj$  and  $ggZZ$  MC prediction. The bands also include the uncertainties in the cross-section measurements of the  $ttZ$  and  $VVV$  processes. The treatment of the systematic theoretical uncertainties is the same as that of the signal region, which will be discussed in Section 18.1. Other experimental uncertainties related to the lepton reconstruction and identification, trigger, and luminosity discussed in Section 18.2 are assumed to be negligible for the validation regions. For most bins in both regions, the data and MC yield are compatible within the  $1\sigma$  band of the total uncertainties. Additionally, the agreement between data and MC simulation is better when the reducible events are estimated using the fake factor method, thus, fully validating the method. Moreover, the estimate from the fake factor method is robust in validation regions where the non-prompt lepton composition is different from that of the phase space where the fake factors were derived.

The data and MC yield comparisons for three kinematic observables,  $p_T$  of leading lepton,  $\eta$  of leading lepton, and the number of jets  $n_{jets}$  are shown by Figures 50a, 50c and 50e, respectively in VRDF. Similarly, Figures 50b, 50d and 50f, show the data and MC yield for  $p_T$  of leading lepton,  $\eta$  of leading lepton and the number of jets  $n_{jets}$ , respectively in VRSC. The data and MC prediction are compatible in most bins within total uncertainties in each validation region.



(a) VRDF: MC predicted non-prompt background. (b) VRDF: FF estimate of non-prompt background.



(c) VRSC: MC predicted non-prompt background. (d) VRSC: FF estimate of non-prompt background.

Figure 49: Comparisons of data to the total MC prediction and non-prompt background in the different flavor (top) and same charge (bottom) validation regions as a function of  $m_{4\ell}$ .

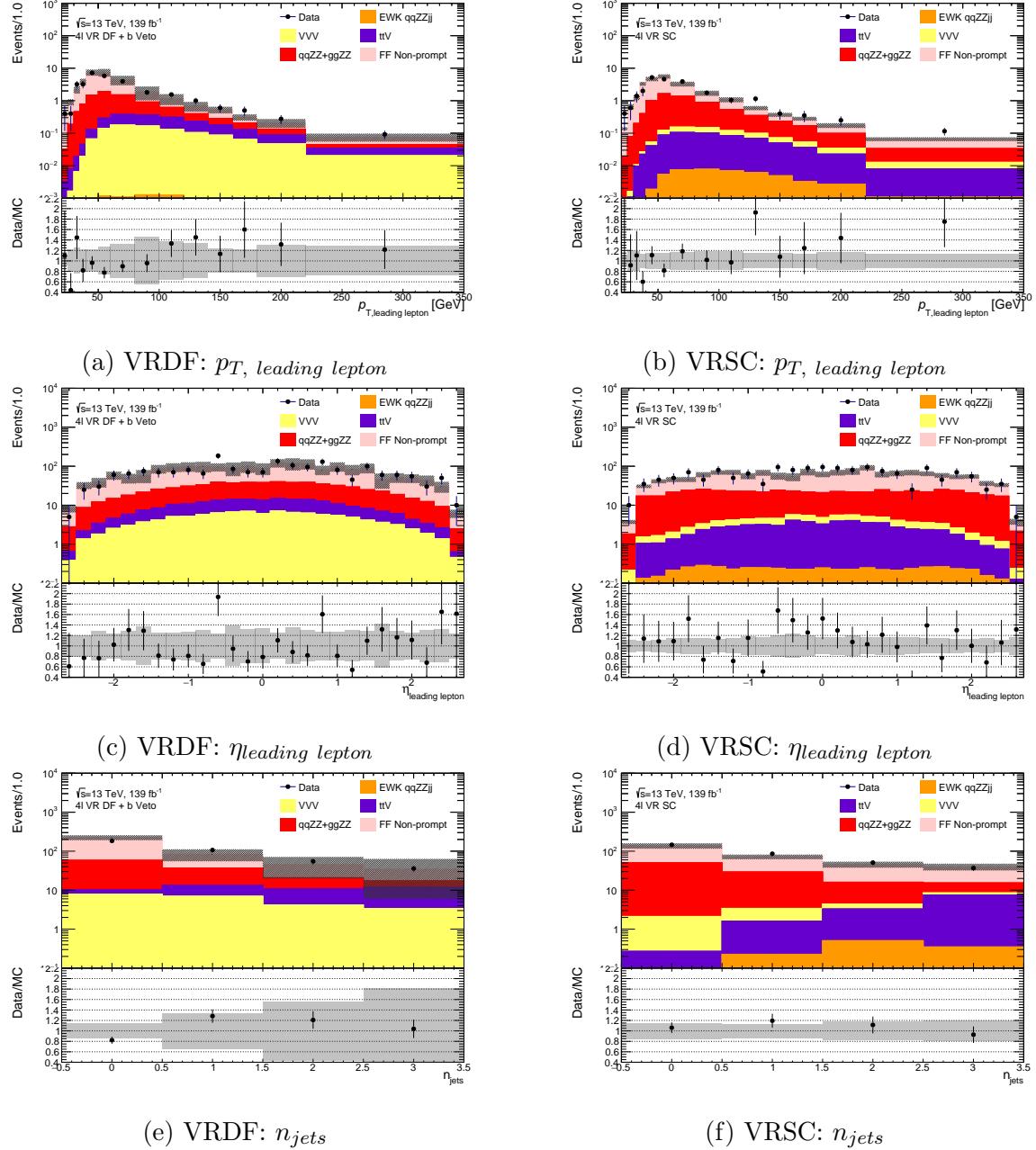


Figure 50: Comparisons of data to the total MC prediction and non-prompt background in different-flavor (left) and same-charge (right) validation regions as a function of three kinematic observables used to parametrize the fake efficiencies in the combined control region.

### 16.1.6 Signal Region Estimation

Similar to the validation regions, the background in the signal region is estimated by applying the fake factor to the baseline-anti-signal leptons in the non-signal quadruplets, as discussed in Section 16.1.3. Distributions in Figure 51 compare the non-prompt background predicted from MC and estimated from the fake-factor method in the VBS-Enhanced region as a function of  $m_{4\ell}$ . The data-driven estimate of the non-prompt background closely agrees with that predicted by the MC.

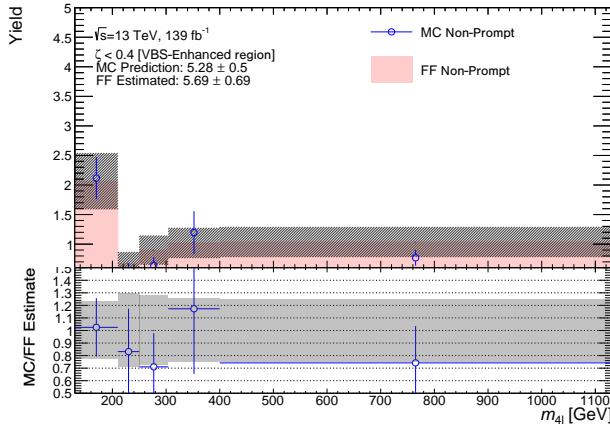


Figure 51: MC prediction and fake-factor method estimate of the non-prompt background as a function of  $m_{4\ell}$  in the VBS-Enhanced region. Black bands represent the systematic uncertainties from the fake factor method.

Figure 52 shows the total SM detector-level predictions and estimated non-prompt background as a function of  $m_{4\ell}$  and  $p_{T,4\ell}$  in the VBS-Enhanced region. The signal MC predictions include the combination of EWK  $qqZZjj$  and QCD parton and gluon-loop initiated  $qqZZ ggZZ$  samples, respectively. The *FF Non-Prompt*,  $t\bar{t}V$ , and  $VVV$  distributions show the three background processes, where *FF Non-Prompt* is the estimated non-prompt lepton backgrounds using the data-driven fake factor method. The ratio panels show the predicted high values of signal-to-background ratios.

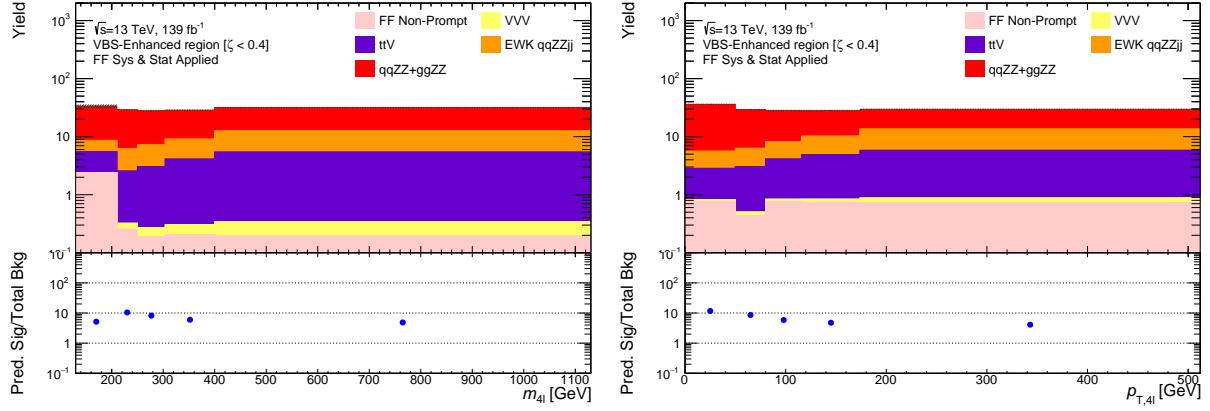


Figure 52: Detector level SM predictions of signal and background processes along with the non-prompt backgrounds estimated from the fake-factor method as a function of  $m_{4\ell}$ (left) and  $p_{T,4\ell}$  (right) in the VBS-Enhanced region.

# 17 Unfolding

The main results of this thesis are differential cross-section measurements at the particle level. The inclusive detector-level cross-section for a given physics process  $p_1 p_2 \rightarrow X$  is

$$\sigma_{p_1 p_2 \rightarrow X}^{\text{detector level}} = A \times \epsilon \times \sigma_{p_1 p_2 \rightarrow X}^{\text{particle-level}}, \quad (17.1)$$

where  $\sigma_{p_1 p_2 \rightarrow X}^{\text{particle-level}}$  is the *true cross-section* of the physics process predicted by the theory,  $A$  is the *detector acceptance* and  $\epsilon$  is the *reconstruction efficiency*. The physical layout of the ATLAS detector does not cover all areas of the phase space, and the term  $A$  accounts for the limited acceptance of the ATLAS detector. Several parts of the detector have several reconstruction efficiencies, which are accounted for by the factor  $\epsilon$ . The detector-level cross-section is measured experimentally in terms of the number of particles in a given final state ( $N$ ) and integrated Luminosity  $L$  as  $\sigma_{p_1 p_2 \rightarrow X}^{\text{detector level}} = \frac{N}{L}$ . The *true* particle-level inclusive cross-section can be estimated by correcting for detector acceptance and detector efficiency on the measured cross-section  $\sigma_{p_1 p_2 \rightarrow X}^{\text{detector level}}$ . Additionally, for differential cross-sections, where the cross-section is measured in different bins of the kinematic observables, further correction is needed to correct the resolution-induced migration between nearby bins.

This Chapter discusses the unfolding technique in detail. Section 17.1 gives an overview of the unfolding algorithm, and Section 17.2 discusses the optimal binning for unfolding. Section 17.3 validates the unfolding method, and Section 17.4 discusses the bias from unfolding and the attempts to optimize the bias.

## 17.1 Method Overview

The analysis uses an *iterative Bayesian unfolding* algorithm based on Baye's theorem [90] [91] using the ATLAS-supported *RooUnfold* package [92]. Bayes' theorem formulates a mathematical relation to obtain a probability of an effect  $E$  caused by several independent causes

$C_i$ , given the initial probability of the causes  $P(C_i)$  and the conditional probability of the  $i - th$  cause to produce the effect  $P(E|C_i)$  as

$$P(C_i|E) = \frac{P(E|C_i).P(C_i)}{\sum_j P(E|C_j).P(C_j)}. \quad (17.2)$$

The obtained probability depends on the prior probability of the cause and the conditional probability of cause and effect, where, in experimental Particle Physics, the cause is the actual particle-level value, and the effect is the measured one. The prior dependency is reduced by using an iterative technique, where the outcome of the previous iteration is used as a prior for the subsequent iteration.

For a single iteration, the algorithm can be summarized as

$$U_i = \frac{1}{\epsilon_i} \times \sum_j^{reco\ bins} (R_j - F_j).f_i \cdot \frac{M_{ji}T_i}{\sum_k^{truth\ bins} M_{jk}T_k}, \quad (17.3)$$

where  $U_i$  is the unfolded yield in the target bin  $i$ ,  $T_i$  is the predicted particle-level yield in particle bin  $i$ ,  $R_j$  is the observed detector-level yield in reco bin  $j$  and  $F_j$  is the subtracted detector-level reducible background yield.  $M_{ij}$  is the migration matrix element from particle-level bin  $j$  to detector-level bin  $i$ .

Based on the discussion above, conceptually, three corrections from the SM MC prediction need to be applied to estimate the unfolded yield. The three unfolding inputs are:

- **Reconstruction efficiency ( $\epsilon$ )**, which accounts for the limited efficiency of the detector. It is defined as the ratio between the number of events that pass both detector- and fiducial-level selections and the number of events that pass only the fiducial-level selection.
- **Fiducial fraction ( $f$ )**, which corrects for events that are outside the fiducial region selected as part of the detector-level events due to limited detector resolution. An example of such an event would be a signal  $ZZ^*(\rightarrow 4\ell)jj$  event where one of the jets

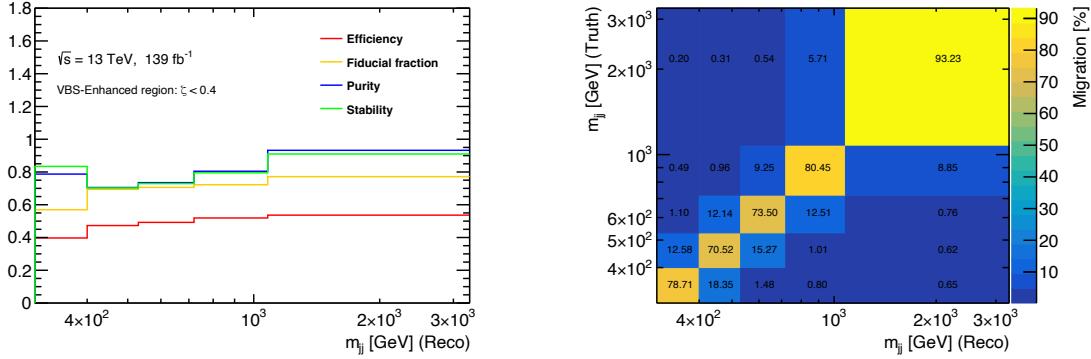
originates from pile-up instead of hard-scatter. It is defined as the ratio between the number of events that pass both detector- and fiducial-level selections to the number of events passing only the detector-level selection.

- ***Migration matrix*** ( $M_{ij}$ ), is a two-dimensional matrix that accounts for events migrated from particle-level bin  $j$  to detector-level bin  $i$ . The migration matrix corrects the probability of bin migration. It is measured in MC by comparing particle- and detector-level distributions for events that pass both fiducial- and detector-level selections. Bin migrations result from resolution effects and smearing of the reconstructed distributions. The diagonal component of the migration matrix is related to the *fiducial purity*, which corresponds to the fraction of detector-level events that originate from the same bin at the particle level. A similar distribution *stability* defined as the fraction of particle-level events reconstructed in the same detector-level bin is also closely related to the diagonal component of the migration matrix.

Figures 53a and 53b show the first two unfolding inputs along with purity and stability, and the migration matrix for  $m_{jj}$ , respectively, in the VBS-Enhanced region. The reconstruction efficiency is less than 50% resulting from the poor efficiency of jet reconstruction. The fiducial fraction and purity are smaller in lower bins of  $m_{jj}$ , which mainly corresponds to the contribution from pile-up jets faking the event selection. The normalized migration matrix shown in the second plot with the particle-level prediction in  $y - axis$  and the detector-level prediction in  $x - axis$  is diagonal.

## 17.2 Binning Optimization

Optimal binning is required to perform the unfolding procedure effectively for all kinematic observables. Two factors drive the choice of binning: the necessity to have large enough bin statistics to maintain the Gaussian approximation used in the statistical uncertainty propagation while preserving the shape of the differential distributions and the necessity to



(a) reconstruction efficiency, fiducial fraction, (b) Normalized migration matrix showing percentage of bin migration  
purity, and stability.

Figure 53: Unfolding inputs from SM MC as a function of  $m_{jj}$  in the VBS-Enhanced region.

minimize large bin migrations and statistical uncertainties from unfolding. Therefore, each bin must have at least 15 events in the VBS-Suppressed region and at least 20 events in the VBS-Enhanced signal region.

To maintain a good performance of the unfolding, each bin for the kinematic observable has at least 70% purity except for  $p_{T,4\ell jj}$  where at least 50% purity is required. Moreover, every bin width must be equal to or greater than the observable resolution in the same bin for each observable. The resolution in each particle-level bin is evaluated from MC by comparing the difference of particle and detector-level yield for events that pass both fiducial- and detector-level event selections. The difference is fitted using Gaussian approximation, and twice the resulting standard deviation is taken as the bin resolution. Table 15 shows the final bin choices for all kinematic observables used in differential cross-section measurement.

Table 15: Binning for all unfolded observables in VBS-Enhanced and VBS-Suppressed regions.

Observable	Region	Binning
$m_{jj}$ [GeV]	VBS-Enhanced VBS-Suppressed	[300, 400, 530, 720, 1080, 3280] [300, 410, 600, 178]
$m_{4\ell}$ [GeV]	VBS-Enhanced VBS-Suppressed	[130, 210, 250, 304, 400, 1130] [130, 226, 304, 752]
$p_{T,4\ell}$ [GeV]	VBS-Enhanced VBS-Suppressed	[0, 50, 80, 116, 174, 512] [0, 76, 140, 424]
$p_{T,jj}$ [GeV]	VBS-Enhanced VBS-Suppressed	[0, 52, 82, 116, 172, 524] [0, 80, 146, 448]
$p_{T,4\ell jj}$ [GeV]	VBS-Enhanced VBS-Suppressed	[0, 20, 42, 64, 298] [0, 36, 70, 254]
$s_{T,4\ell jj}$ [GeV]	VBS-Enhanced VBS-Suppressed	[70, 240, 320, 420, 580, 1410] [70, 330, 500, 1210]
$ \Delta y_{jj} $	VBS-Enhanced VBS-Suppressed	[2, 3.08, 3.74, 4.32, 5.06, 7.4] [2, 2.94, 3.78, 5.4]
$\Delta\phi_{jj}^{signed}$	VBS-Enhanced VBS-Suppressed	$[-\pi, -2.1, 0, 2.1, \pi]$ $[-\pi, 0, \pi]$
$\cos\theta_{\ell i \ell j}^*$	VBS-Enhanced VBS-Suppressed	[-1, -0.5, 0, 0.5, 1] [-1, 0, 1]
$\zeta$	VBS-Enhanced VBS-Suppressed	[0, 0.06, 0.12, 0.18, 0.26, 0.4] [0.4, 0.5, 0.64, 1.02]

## 17.3 Method Validation

The unfolding method is validated using three tests: MC-based closure test, injection test, and test with physics variations.

### 17.3.1 MC Validation Test

The first validation of the unfolding technique is with the SM MC. A detector-level distribution for a kinematic observable is predicted from SM and unfolded using the unfolding inputs from the same MC. Figure 54 shows an example of the MC-based closure test for  $m_{jj}$  in the VBS-Enhanced region. The detector-level MC prediction is unfolded using the inputs from the same MC, and the resulting unfolded distribution is compared with the particle-level prediction, as shown in the lower panel. Since both detector-level prediction and unfolding inputs are derived from the same MC, a perfect closure is observed when comparing the unfolded- and particle-level predictions.

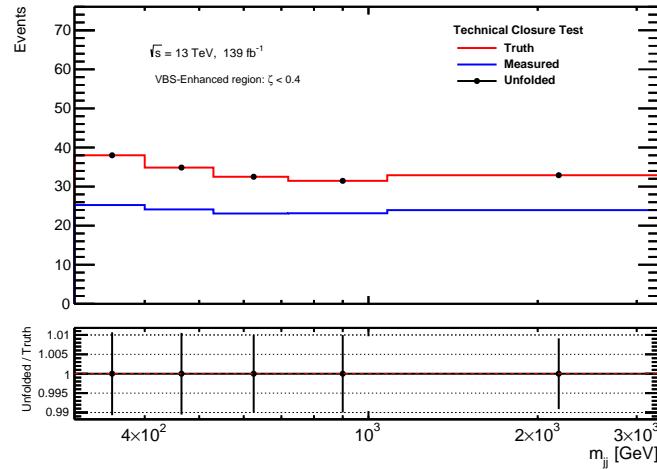


Figure 54: MC technical closure test of the unfolding procedure for  $m_{jj}$ . The detector-level MC distribution (in blue) is unfolded (black) with the nominal SM unfolding inputs and compared to the particle-level distribution (in red) from the same MC.

### 17.3.2 Injection Test

The analysis uses a model-independent EFT approach which will be discussed in Section 21 to constrain the effect of BSM physics. Therefore, it is essential to test the ability of the unfolding algorithm to uncover the accurate particle-level prediction from data containing BSM physics via injection test. In an injection test, a BSM physics contribution is added to the SM detector-level prediction, unfolded with the nominal SM unfolding inputs, and compared with the BSM-added particle-level distribution. Figure 55a shows an injection test for  $m_{jj}$  in the VBS-Enhanced region where a BSM contribution is added to the SM MC. The BSM contribution is from linear and quadratic contributions of an  $FT0$  dimension-8 EFT operator, which modifies the quartic gauge electroweak vertex shown in Figure 7b. The value of the coefficient corresponding to the injected  $FT0$  dimension-8 EFT operator is at about 75% of the obtained upper limit when fitting the measured cross-section, which is discussed in detail in Section 21. Figure 55b shows the result of the injection test. The BSM-added detector-level MC prediction is unfolded using nominal SM MC unfolding inputs and compared against the BSM-added particle-level distribution. A small non-closure of about 5% in the last bin of  $m_{jj}$  is observed, which is well within the statistical uncertainties of the unfolded distribution.

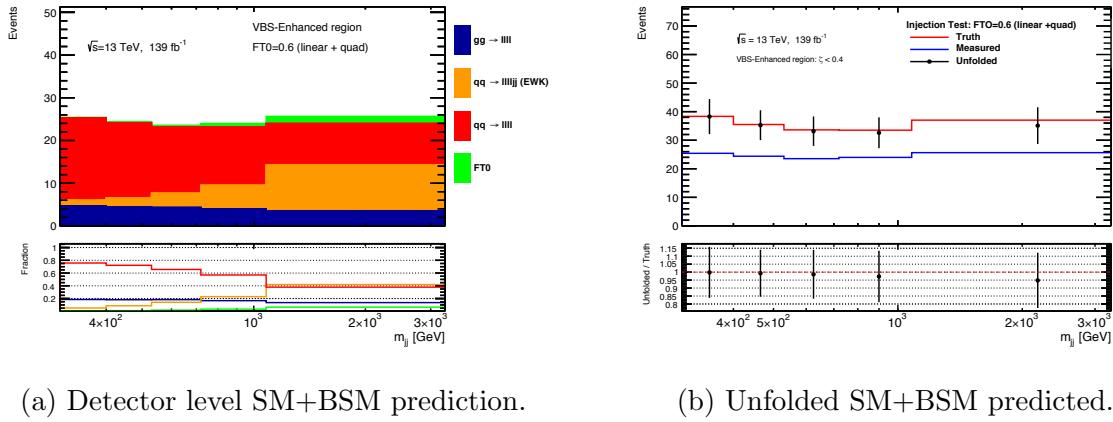


Figure 55: Injection test with dimension-8  $FT0$  EFT operator.

### 17.3.3 Physics Variation

From the previous ATLAS electroweak  $ZZjj$  analysis, a slight enhancement on the central value of the EWk  $ZZjj$  cross-section was measured [19]. The final unfolding validation tested the ability of the algorithm to recover the actual shape of particle-level distribution if a physics process cross-section was different from the SM prediction. First, as shown by Figure 56a, the cross-section for parton-initiated QCD  $qqZZ$  is varied by a factor equal to the expected statistical uncertainty of  $\pm 15\%$ . The varied detector-level distribution is then unfolded using the nominal SM MC unfolding inputs and compared with the varied fiducial-level prediction. Figure 56b shows the same test where the EWK  $qqZZjj$  cross-section is varied by  $\pm 11\%$  based on the enhanced cross-section observed in the previous measurement. In both cases, a non-closure of about 1% is observed, well below the statistical uncertainties from unfolding.

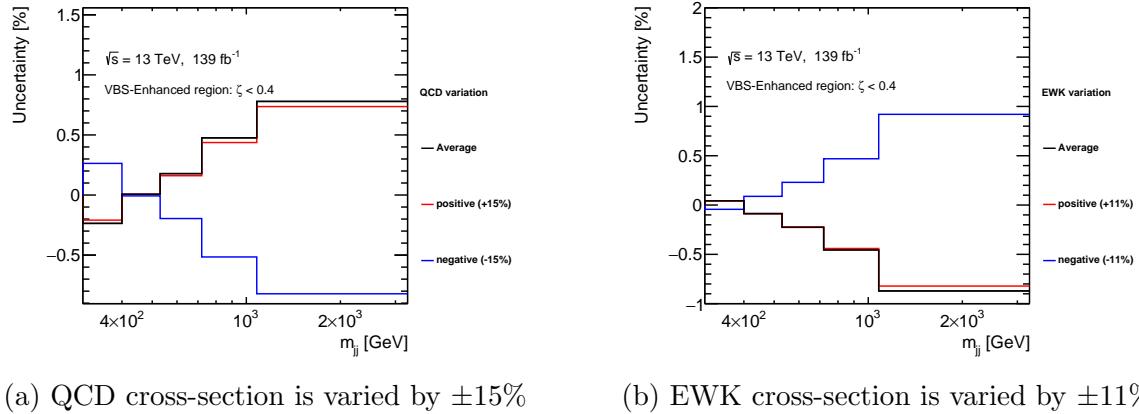


Figure 56: Unfolding validation using physics variation where parton-initiated QCD (left) or the EWK (right) process cross-sections are varied.

## 17.4 Bias and Optimization

The unfolded procedure relies on a prior value depending on the SM MC, which naturally biases the unfolded cross-sections. With each iteration of unfolding, the algorithm improves the knowledge of the prior, thus, reducing the unfolding bias. However, with an increasing

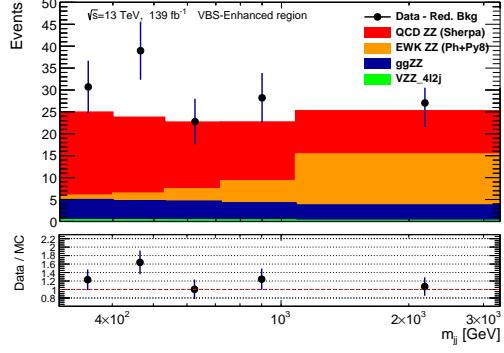
number of iterations, the iteration correction for bin migrations amplifies the statistical fluctuations in data, resulting in larger values of statistical uncertainties. Therefore, a finite number of iterations is chosen, and the resulting unfolding bias is taken as the systematic uncertainty for the measurement.

For optimization, the unfolding bias is evaluated by the *data-driven closure test*, where a pseudo dataset is developed using the ratio of observed background-subtracted data and SM detector-level prediction. First, for each observable, the data to MC ratio is smoothed using Friedman’s Super Smoother technique [93], fixing the endpoints to the ratio value in the first and the last bins. A re-weighting function for each observable is developed to reweigh the SM fiducial- and detector-level predictions. The reweighed pseudo-detector-level prediction is then unfolded with the nominal unfolding inputs from SM and compared with the reweighed pseudo-particle-level prediction to get the final unfolding bias.

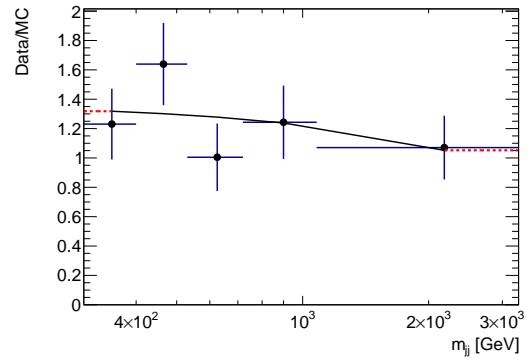
Figure 57 shows a step-by-step procedure for the data-driven closure test for  $m_{jj}$  in the VBS-Enhanced region. The detector-level SM prediction and collected data with their ratio are shown in Figure 57a, and Figure 57b shows the smooth ratio of Data to MC prediction. Figure 57c shows the nominal and pseudo-detector-level predictions. Finally, Figure 57d compares the pseudo-unfolded distribution to the pseudo-particle-level prediction, and the lower panel shows the final unfolding bias obtained from the data-driven closure test. As shown by the ratio panel of Figure 57d, unfolding bias is observed to be about 10%.

The bias observed in Figure 57d is obtained using a single iteration for Bayesian unfolding. The data-driven closure test was repeated for several iterations in the VBS-Suppressed region to reduce the unfolding bias. The resulting unfolding bias and systematic uncertainties up to 4 iterations are shown by Figures 58a and 58b, respectively. As expected, the unfolding bias decreases, whereas the statistical uncertainty increases with the higher number of iterations. A single iteration is the optimal measurement choice to balance the statistical and the unfolding bias uncertainties.

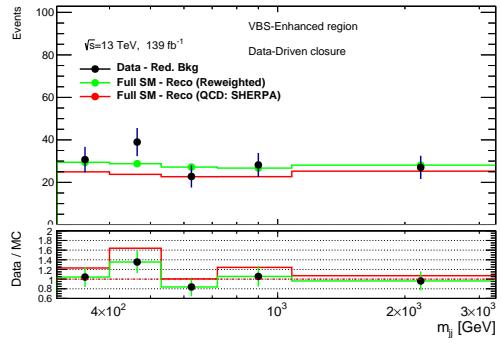
Unfolding bias is the largest source of the systematic uncertainty of the analysis and is



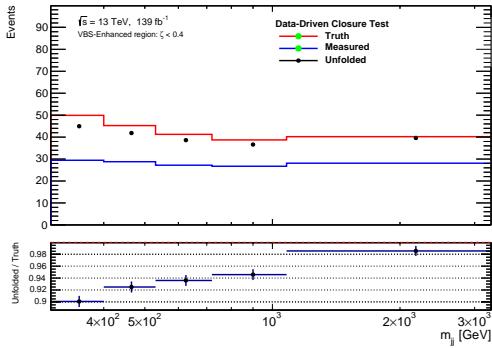
(a) Data and MC for  $m_{jj}$



(b) Smoothed ratio of Data and MC.



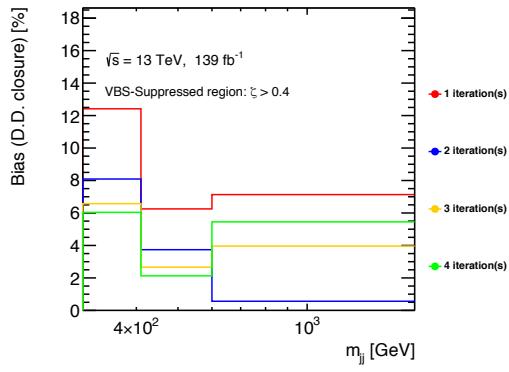
(c) Nominal SM (red) detector-level yield and re-weighted detector-level yield (green).



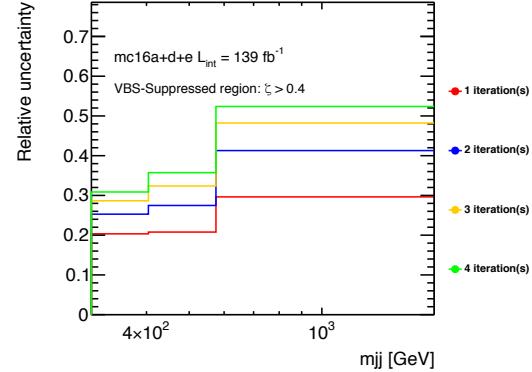
(d) Unfolding bias.

Figure 57: A step-by-step overview of the data-driven closure test to get the unfolding bias.

estimated using an MC-driven toy method discussed in Section 18.3. The observed significant bias is caused by detector-level pile-up jets at lower  $p_T$  or higher  $\eta$  that are not part of the fiducial phase space but are present in the detector-level measurement. Compared to the efficiencies for high  $p_T$  or central jets, the jet-vertex-tagger and forward-jet-vertex-tagger have lower efficiency in selecting the hard scattering jets at lower  $p_T$  or higher  $\eta$  region. Thus, resulting in contamination from *fiducial-fake-event*. The additional MC-based studies on the unfolding bias are summarized in Appendix B.



(a) Unfolding bias



(b) Statistical uncertainty

Figure 58: Unfolding bias (left) and statistical uncertainty (right) with up to 4 unfolding iterations as a function of  $m_{jj}$  in VBS-Suppressed region.

## 18 Uncertainties on the Measurement

The differential cross-section measurements discussed in this thesis are affected by three sources of systematic uncertainties: experimental sources, theoretical sources, and intrinsic systematics related to the unfolding process. The statistical uncertainty of the measurements is dominant as data statistics limit the cross-section measurements. This section discusses the source of theoretical, experimental, and unfolding uncertainties and the propagation of statistical and systematic uncertainties to the final unfolded cross-section measurements.

### 18.1 Theoretical Uncertainties

The following sources of theoretical uncertainties are considered in the measurement.

- **Uncertainties on QCD Scale**

As discussed in Section 3, the theoretical predictions of cross-sections depend on the factorization scale ( $\mu_F$ ) and renormalization scale ( $\mu_R$ ) [94]. To account for this dependence, a QCD scale uncertainty is evaluated by scaling  $\mu_F$  and  $\mu_R$  independently using on-the-fly variations provided by the MC generators. The variations constitute of six-point variations of  $\mu_F$  and  $\mu_R$  from  $-50\%$  to  $+100\%$  around their nominal values of 1, such that  $\{\mu_R = 0.5, \mu_F = 0.5\}$ ,  $\{\mu_R = 0.5, \mu_F = 1.0\}$ ,  $\{\mu_R = 1.0, \mu_F = 0.5\}$ ,  $\{\mu_R = 1.0, \mu_F = 2.0\}$ ,  $\{\mu_R = 2.0, \mu_F = 1.0\}$ , and  $\{\mu_R = 2.0, \mu_F = 2.0\}$ . The final uncertainty is evaluated as the absolute envelope of the six variations. The QCD scale uncertainties are individually evaluated for  $qqZZ$ ,  $ggZZ$ , and EWK  $qqZZjj$  samples.

- **Uncertainties on PDF and  $\alpha_S$**

The cross-sections also depend on the choice of the PDF used by the MC generators. Thus, the PDF uncertainty for Sherpa and MADGRAPH5 samples that use NNPDF3.0NNLO is evaluated using the prescription described in Ref [80] using on-the-fly variation weights. The PDF variations include a set of 100 internal varia-

tions, two additional variations from the nominal PDF re-weighted to the alternative MMHT2014nnlo [95], and CT14nnlo [96] PDF sets and variations of the strong coupling constant by  $\pm 0.001$  around the nominal value of  $\alpha_S$  is 0.118. The total uncertainty is taken as the absolute envelope of all standard deviations of 100 internal variations and the two alternate PDF variations, added in quadrature with the envelope of the  $\alpha_S$  variations,

$$\sigma_{PDF}^{NNPDF3.0NNLO} = \sqrt{[max(\sigma_{std. dev. int.}, |\sigma_{MMHT2014nnlo}|, |\sigma_{CT14nnlo}|)]^2 + \sigma_{\alpha_S}^2}. \quad (18.1)$$

The PDF uncertainty is evaluated individually for  $qqZZ$ ,  $ggZZ$ , and **MADGRAPH5** EWK  $qqZZjj$  samples.

The electroweak EWK  $qqZZjj$  samples generated by POWHEG-V2 do not have on-the-fly variations to evaluate the PDF uncertainty. Therefore, PDF uncertainty from the **MADGRAPH5** sample is taken as the PDF uncertainty for POWHEG-V2 EWK  $qqZZjj$  samples.

- **Uncertainties on  $gg \rightarrow ZZ^{(*)}$  NLO Corrections**

The uncertainty is related to the NLO QCD k-factor applied to the  $ggZZ$  sample [97]. The NLO QCD k-factors applied are evaluated differentially as a function of the  $m_{4\ell}$ .

- **$t\bar{t}V$  and  $VVV$  cross-sections**

The experimental uncertainties on recently published cross-section measurements of the  $t\bar{t}V$  [98] and  $WZZ$  [99] processes by ATLAS are propagated for the analysis. On the entire  $t\bar{t}V$  process, a flat conservative variation of 15% is applied, taken from the cross-section measurement of  $t\bar{t}Z$ . Similarly, for  $VVV$  conservative 10% variation taken from the cross-section measurement of  $WWZ$  is applied to the total  $VVV$  samples.

As shown above, the theoretical uncertainties are process specific and are evaluated separately for each MC sample. The theory uncertainties need to be propagated to the unfolded

cross-section measurements. For each theory uncertainty, variation-applied particle- and detector-level predictions are built by substituting the varied distribution for the selected process instead of the nominal one. The variation-applied detector-level prediction is unfolded using the unfolding inputs from nominal SM predictions. For each variation, the difference of variation-applied unfolded result from the variation-applied truth MC prediction gives the systematic uncertainty. In general, the theoretical variations significantly affect the predicted particle-level and detector-level predictions; however, they have a negligible impact on the shape of the distributions. Since the variation is applied to both the detector and the particle-level yields, the resulting uncertainties from theory systematics on the unfolded cross-sections are small.

## 18.2 Experimental Uncertainties

The experimental uncertainties arise from the measurement of the energy and momentum scales of the reconstructed objects and the uncertainties on object reconstruction, identification, and selection efficiencies. The following category summarizes the sources of experimental uncertainties.

### Jet Related Uncertainties

The analysis requires two jets in the final state. Therefore, jet-related uncertainties resulting from jet reconstruction, calibration, and selection uncertainties are the most significant sources of experimental systematic uncertainties.

#### • Jet Reconstruction Uncertainty

The jet-related uncertainties associated with reconstruction and different steps of calibration discussed in Section 7.5 are provided by ATLAS-supported tool *JetUncertainties*<sup>6</sup>.

The tool provides several configurations for jet-related uncertainties adjusted to the various needs of several analyses. The measurement uses *GlobalReduction\_FullJER*

---

<sup>6</sup><https://twiki.cern.ch/twiki/bin/view/AtlasProtected/JetUncertainties>

configuration with a total of 36 uncertainties, each with upward and downward components, corresponding to  $36 \times 2$  variations, 20  $\times 2$  of which are related to JES, and  $13 \times 2$  are related to JER.  $6 \times 2$  of the  $36 \times 2$  variations are related to the  $\eta$  intercalibration procedure,  $4 \times 2$  to the pile-up energy subtraction step in jet calibration, and  $8 \times 2$  to the in-situ calibration of jets. Additional  $1 \times 2$  variations arise separately each from the flavor composition, flavor response, a single particle response at high  $p_T$ , and possible punch-through<sup>7</sup> effects.

- **JVT and fJVT Uncertainties**

Additional sets of jet uncertainties ( $1 \times 2$ ) arising each from the efficiencies of jet vertex selections, JVT, and fJVT cut requirements are also considered in the analysis.

An envelope of the up and down variation for each of the 13 JER uncertainties are added in quadrature to get the total effect of JER systematics. The variations related to the JES and jet calibration are single-sided, and an average in modulus is taken as the uncertainty for each of these sources. The total uncertainty from jets is the sum quadrature of uncertainty from total JER and all other sources.

### **Lepton Related Uncertainties**

The following categories define the lepton-related uncertainties in the analysis

- **Electron Efficiencies**

The electron efficiency uncertainty consists of uncertainties on the trigger, identification, reconstruction, and isolation efficiencies of electrons. These uncertainties are provided by an ATLAS-supported tool *ElectronEffciencyCorrection*<sup>8</sup>. There are a total of 61 nuisance parameters related to electron efficiencies, each with upward and downward components corresponding to  $61 \times 2$  variations. 34  $\times 2$  out of 61 is related to uncertainties in identification efficiency, 25  $\times 2$  related to the reconstruction efficiencies,

---

<sup>7</sup>particles' probabilities of passing through the calorimeters

<sup>8</sup><https://gitlab.cern.ch/atlas/athena/-/tree/21.2/PhysicsAnalysis/ElectronPhotonID/ElectronEfficiencyCorrection>

and a single nuisance parameter ( $1 \times 2$ ) each from the isolation efficiency and trigger efficiency scale factors.

- **Muon Efficiencies**

Similar to the electrons, muon efficiency uncertainty consists of variations on the trigger, identification, reconstruction, and isolation efficiencies of muons, which are provided by another ATLAS-supported tool *MuonEfficiencyCorrections*<sup>9</sup>. In total, there are  $10 \times 2$  nuisance parameters, sets of two ( $2 \times 2$ ) variations corresponding to trigger efficiency scale factors, sets of four ( $4 \times 2$ ) related to the identification and reconstructed efficiency, two sets of two ( $2 \times 2$ ) each corresponding to the isolation efficiency and track-to-vertex association efficiency.

- **Electron Scale and Resolution**

The electron scale and resolution uncertainty is accounted for by two sets of nuisance parameters corresponding to  $2 \times 2$  variations.

- **Muon Scale and Resolution**

For muons resolution and scale uncertainties, there are 5 sets of nuisance parameters,  $2 \times 2$  corresponding to the muon momentum resolution as measured separately by the inner detector and the muon spectrometer. One set of nuisance parameters ( $1 \times 2$ ) corresponds to the uncertainties on the muon momentum scale, and two sets of  $2 \times 2$  are associated with the uncertainties in the measurement of muon sagitta due to the ID alignment.

## Other Experimental Uncertainties

- **Pile-up Re-weighting**

As discussed in Section 14.3, the MC predictions are re-weighted to match the pile-up profile of data. A single  $1 \times 2$  nuisance parameter accounts for upward and downward variations in the factors used for pile-up re-weighting.

---

<sup>9</sup>21`https://gitlab.cern.ch/atlas/athena/-/tree/21.2/PhysicsAnalysis/MuonID/MuonIDAnalysis/MuonEfficiencyCorrection`

- **Luminosity**

As discussed in Section 14.1, the uncertainty in the collected integrated luminosity of  $139\text{ fb}^{-1}$  is  $\pm 1.7\%$ , which is applied as a flat variation to both particle and detector-level yields.

The experimental uncertainties affect all detector-level MC predictions and the estimate of the non-prompt backgrounds. A total variation-applied detector-level prediction is constructed for each variation, including the contributions from signal physics processes ( $qqZZ + ggZZ + EWK\ qqZZjj$ ), MC predicted background physics processes ( $ttV + VVV$ ) and variation-applied non-prompt background estimate. The variation-applied background (MC + fakes) is first subtracted from the variation-applied total MC prediction and then unfolded using the unfolding inputs from the nominal SM prediction. The individual systematic uncertainty corresponds to the difference between the variation-applied and nominal unfolded distributions for each variation.

### 18.3 Unfolding Uncertainties

The following two uncertainties are intrinsic to the unfolding process itself and are included in the uncertainties for the unfolded differential cross-sections.

- **Unfolding Bias**

The inherent unfolding bias can be estimated either using the data-driven method discussed in Section 17.4 or using an MC-toy-based method. It is the most significant source of systematic uncertainty for the measurement. Both methods yield similar results. However, the MC-toy-based method is free from the influence of statistical fluctuation in data and is, thus, chosen as the systematic uncertainty on the intrinsic unfolding bias.

First, ten thousand toys are constructed from the particle-level MC distribution, independently fluctuating each bin of the truth distribution with a random number

extracted from a Gaussian with mean zero and variance of  $N_{truth}^{bin}$ . A pseudo-detector-level dataset is created by *folding* using the nominal response from MC. Folding is a technique to create a detector-level distribution from the MC by applying the detector's acceptance and efficiency corrections, adding detector resolution-induced bin migration between the truth and detector-level distributions, and including the contribution of the fiducial fakes. The pseudo dataset is then unfolded using the nominal unfolding inputs and plotted as the average bias of the toys as a function of the toy's particle-level yield. The final unfolding bias is the value where this distribution intersects with the 68.3% confidence interval of the toys' actual value. Figure 59 shows the unfolding bias evaluated in each bin of  $m_{jj}$  in the VBS-Enhanced region, estimated using the MC-toy-based approach. The unfolding bias ranges from 8 – 15% in different bins.

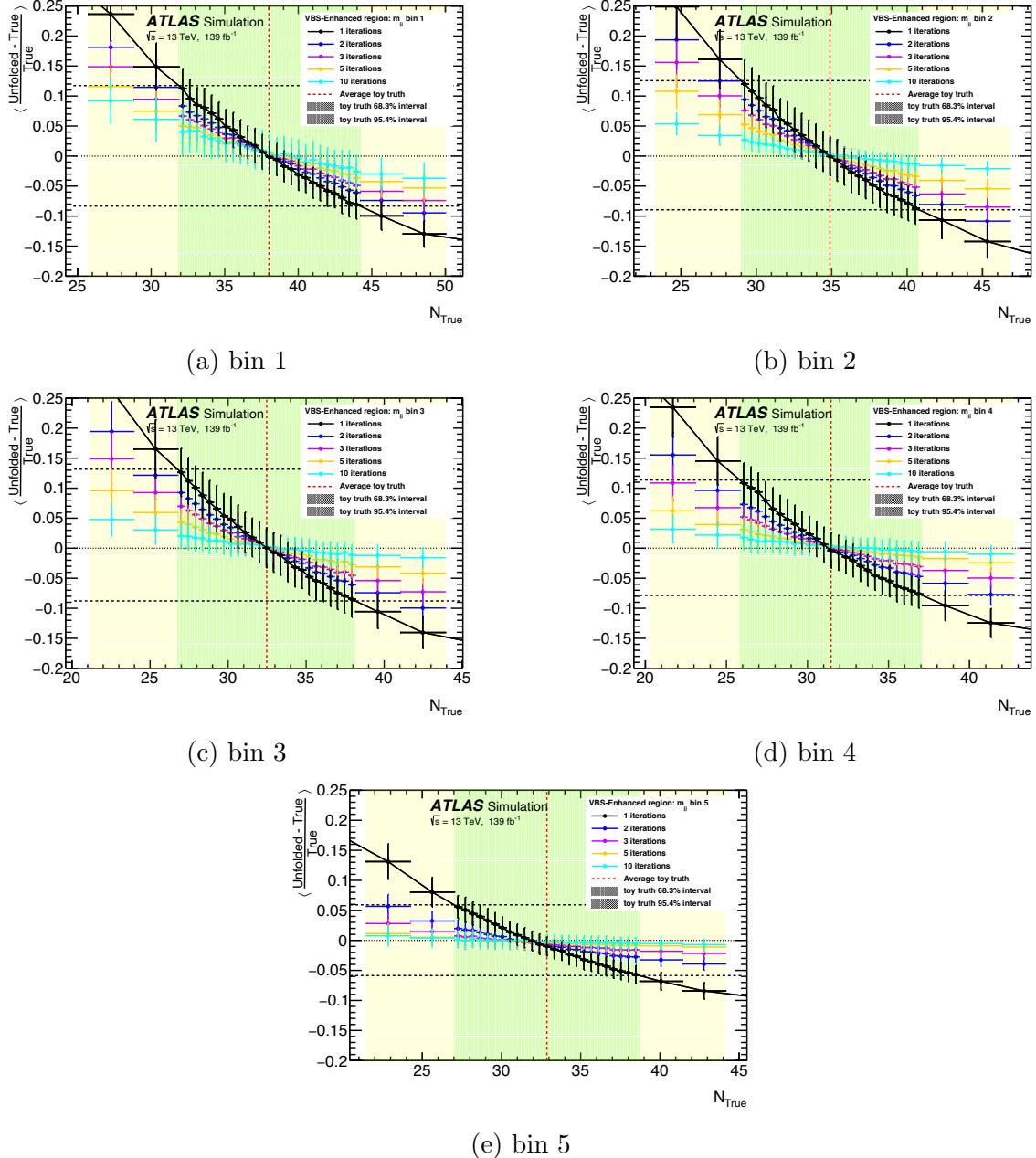


Figure 59: MC-based unfolding bias in each bin of  $m_{jj}$  in the VBS-Enhanced region using Gaussian toys. The distribution shows the relative difference between unfolded and true values for toys as a function of the true value in each bin. The number of unfolding iterations is varied, and the final bias is chosen as the one-sigma intersection with the average value of the toy truth distribution (light-green region intersecting with the black curve).

- **QCD  $qqZZ$  Modeling Uncertainty**

There are known differences in predictions from different generators caused by differences in parton shower and hadronization modeling. Therefore, the second source of unfolding systematics is required to account for the differences in the unfolding input modeling, which is estimated by varying the prediction of the dominant QCD  $qqZZ$  process. To avoid double-counting, the uncertainty of the unfolding method covered by the unfolding bias, an alternative  $qqZZ$  sample predicted by MADGRAPH5 is first re-weighted to match the nominal-SHERPA lineshape. The re-weighted detector-level distribution is unfolded using the inputs from nominal-SHERPA and compared with the re-weighted-MADGRAPH5 particle-level distribution. The relative difference between these two distributions is taken as modeling systematic uncertainty. Figure 60 shows the estimation of the modeling uncertainty for  $m_{jj}$  in the VBS-Enhanced region. The ratio panel of the right plot shows the QCD modeling uncertainties, which range from 2 – 4% varying in different bins.

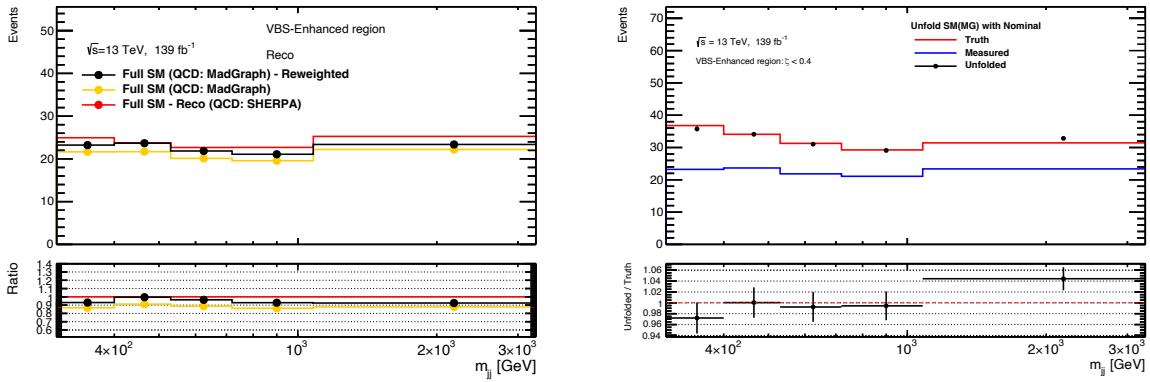


Figure 60: The left plot shows detector-level predictions of  $m_{jj}$  in VBS-Enhanced region predicted by SHERPA, MADGRAPH and MADGRAPH re-weighted to match the SHERPA line-shape. The right plot shows re-weighted MADGRAPH5 detector-level (blue) distribution, which is unfolded (black) using unfolding inputs from nominal-SHERPA and compared to the particle-level re-weighted MADGRAPH5 distribution (red). The lower panel of the right plot showing the ratio between re-weighted particle-level and unfolded-level MADGRAPH5 gives the QCD modeling uncertainties.

## 18.4 Background Uncertainties

There are additional sources of uncertainties from the data-driven estimate of the non-prompt background. The statistical and systematic uncertainties on the fake efficiencies discussed in Section 16.1.4, estimated in the combined control region, are propagated to the final unfolded cross-sections. First, the variation-applied non-prompt background is calculated and subtracted from the nominal detector-level prediction for each variation. The subtracted altered distribution is then unfolded with nominal unfolding inputs. The difference between the altered-unfolded distribution and the nominal-unfolded distribution gives the impact of the background uncertainties on the unfolded cross-section measurements.

## 18.5 Statistical Uncertainties

The statistical uncertainty from the recorded data needs to be propagated to the estimated unfolded yield. Equation 17.3 gives the unfolded yield for a target bin  $i$  with a single iteration. As the background subtracted detector yields are filled event by event, the detector-level distribution is uncorrelated between any two bins. However, as shown by the equation, an unfolded yield in one single bin depends on all detector-level bins due to the resolution effects via bin migration. Therefore, the statistical uncertainty on the unfolded yield is a combination of the uncertainties in detector-level bins and uncertainties on the migration probabilities, which takes the covariance between the detector-level bins into account. The statistical uncertainties at the unfolded level are evaluated by the *RooUnfold* package, which propagates both of these uncertainties.

## 18.6 Breakdown of Uncertainties

Table 16 shows the impact of systematic uncertainties in the VBS-Enhanced region for each bin of  $m_{jj}$ . In most bins, the unfolding bias is the dominant source of systematic uncertainty, followed by the jet systematics.

Bin $m_{jj}$ [GeV]	[300, 400)	[400, 530)	[530, 720)	[720, 1080)	[1080, 3280)
QCD MC modelling	2.91	0.05	0.77	0.56	4.21
Jet	<b>7.49</b>	<b>7.64</b>	<b>9.15</b>	<b>9.08</b>	<b>9.02</b>
Trigger	0.13	0.14	0.13	0.45	0.53
Leptons	1.70	1.69	1.64	2.36	2.60
PRW	0.38	0.58	0.83	0.88	0.62
Theory ( $qqZZ$ )	2.74	2.28	2.63	2.01	0.73
Theory (EWK $qqZZjj$ )	0.08	0.05	0.07	0.14	0.96
Theory ( $ggZZ$ )	0.19	0.10	0.13	0.13	0.51
MC Bkg. ( $t\bar{t}V+VVV$ )	2.67	2.64	2.43	1.84	1.17
Fake Bkg. (stat + syst)	3.32	3.11	2.12	2.68	2.81
Luminosity	1.32	1.34	1.34	1.36	1.42
Unfolding Bias	<b>10.03</b>	<b>10.75</b>	<b>10.97</b>	<b>9.61</b>	<b>5.89</b>
Total	13.99	14.18	15.07	14.08	12.42

Table 16: Breakdown of the relative systematic uncertainties (%) for each bin of  $m_{jj}$  in the VBS-Enhanced region.

Figure 61a schematically shows different types of uncertainties affecting the measurements in the VBS-Enhanced region as a function of  $m_{jj}$ . The solid black line represents the total statistical uncertainty which is dominant in all bins. Total systematic uncertainty, represented by the dashed black line, is the combination of all systematic uncertainties. Depending on the bin, it is dominated either by unfolding bias or jet-related uncertainties. Figure 61b shows the impact of different categories of the jet-related systematic uncertainties on the unfolded differential cross-sections. In most bins of  $m_{jj}$ , the dominant jet uncertainties are from the pile-up energy correction step in the jet calibration. The uncertainties from jet eta-dependent calibration and jet energy resolution are also significant. Overall, the jet reconstruction uncertainties have about 8 – 9% effect on each bin of the unfolded cross-sections.

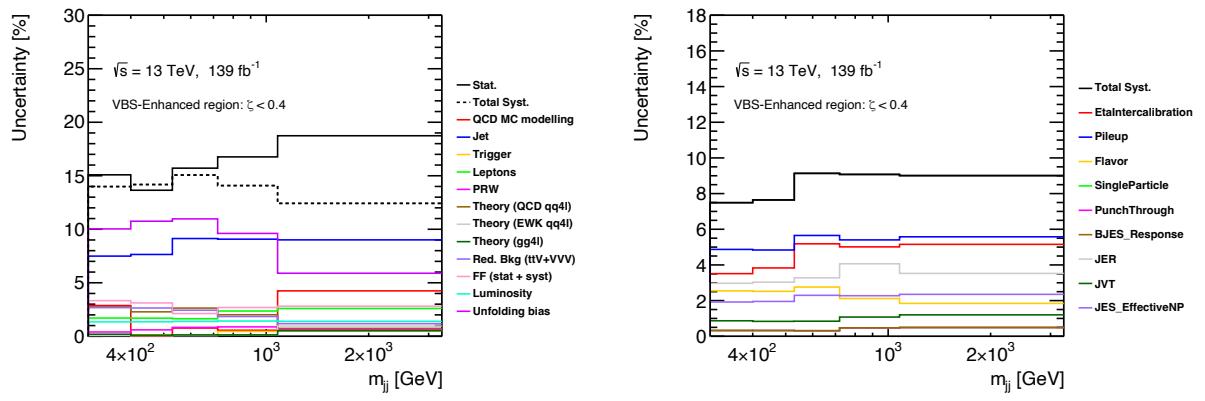


Figure 61: Uncertainties as a function of  $m_{jj}$  in the VBS-Enhanced region.

## Chapter VI: Results

This chapter presents the main results of this thesis. Section 19 presents the detector-level measurements of the eleven kinematic observables introduced in Section 15, whereas Section 20 presents the unfolded differential cross-sections. Section 21 reinterprets the unfolded cross-sections shown in Section 20 to constrain parameters of physics beyond the SM affecting the quartic gauge vertices of electroweak  $ZZ^*jj$  production. The results presented in this thesis are preliminary and not the final measurement for this analysis from the ATLAS experiment.

## 19 Detector Level Measurements

Figure 62 shows the measured detector-level data and predicted detector-level yields for six kinematic observables:

- invariant mass of the dijet system [ $m_{jj}$ ],
- invariant mass of the two  $Z$  bosons [ $m_{4\ell}$ ],
- transverse momentum of the two tagging jets [ $p_{T,jj}$ ],
- transverse momentum of the two  $Z$  bosons [ $p_{T,4\ell}$ ],
- transverse momentum of the two  $Z$  bosons and two tagging jets [ $p_{T,4\ell jj}$ ], and
- scalar transverse momentum of the two  $Z$  bosons and dijet [ $s_{T,4\ell jj}$ ].

Similarly, Figure 63 shows the detector-level data and predictions for remaining kinematic observables:

- cosine of the decay angle of the negative lepton of the leading (sub-leading) pair in the pair's rest frame [ $\cos \theta_{\ell 1(3)\ell 2(4)}^*$ ],
- signed difference between the azimuthal angle of two jets [ $\Delta\phi_{jj}^{signed}$ ],
- rapidity difference between two jets [ $\Delta y_{jj}$ ], and
- centrality of the system [ $\zeta$ ].

For each of these distributions, the measured data (black dot) is compared with the state-of-the-art SM predictions, where the two QCD signals  $qqZZ$  (red) and  $ggZZ$  (blue), and the two MC predicted backgrounds,  $VVV$  (yellow) and  $ttV$  (purple) are SHERPA predictions. The contribution from non-prompt backgrounds (light pink) is estimated using the data-driven method. The  $qqZZjj$  electroweak signal is obtained from POWHEG V2, and the electroweak

production of triboson and two jets ( $VZZjj$ ) is obtained from SHERPA. The vertical-solid error bars on the data points represent the statistical uncertainty in the measured data, and the dashed black band in each bin represents the impact of the total theoretical and experimental uncertainties on the predicted detector-level yields. The impact on the bin yield from the total systematic uncertainties and the statistical precision each ranges from 15 to 20%, depending on the bins and the distributions.

The lower panels in these distributions show the ratio of data yields to the total SM predicted yields, which are compatible within the total uncertainties. Some discrepancies are observed for some distributions. For instance, the yield is underpredicted in the second and the third bin of  $m_{jj}$  and  $m_{4l}$ , respectively. However, these differences are statistically insignificant. Moreover, a slight but statistically insignificant asymmetry is also observed in the measured distribution of  $\Delta\phi_{jj}^{signed}$ , which is expected to be symmetric in the SM.

A simple chi-squared per degree of freedom ( $\chi^2/NDF$ ) is estimated to quantify whether the measured data agree with the SM prediction. As the data yield is event counts, the  $\chi^2/NDF$  is computed using the residual difference in each bin between the unweighted data yield and the weighted MC prediction yield. The respective statistical and systematic uncertainties are also considered in the computation of the  $\chi^2/NDF$ . The reported values of  $\chi^2/NDF$  for each distribution in Figures 62 and 63 show statistically good agreement between the measured data and SM predictions. Most values of  $\chi^2/NDF$  are smaller than one suggesting the systematic uncertainties could be overestimated in these distributions. The experimental uncertainties related to physics object reconstruction are estimated in a different phase space than the analysis's, which could contribute to the overestimation of these systematic uncertainties.

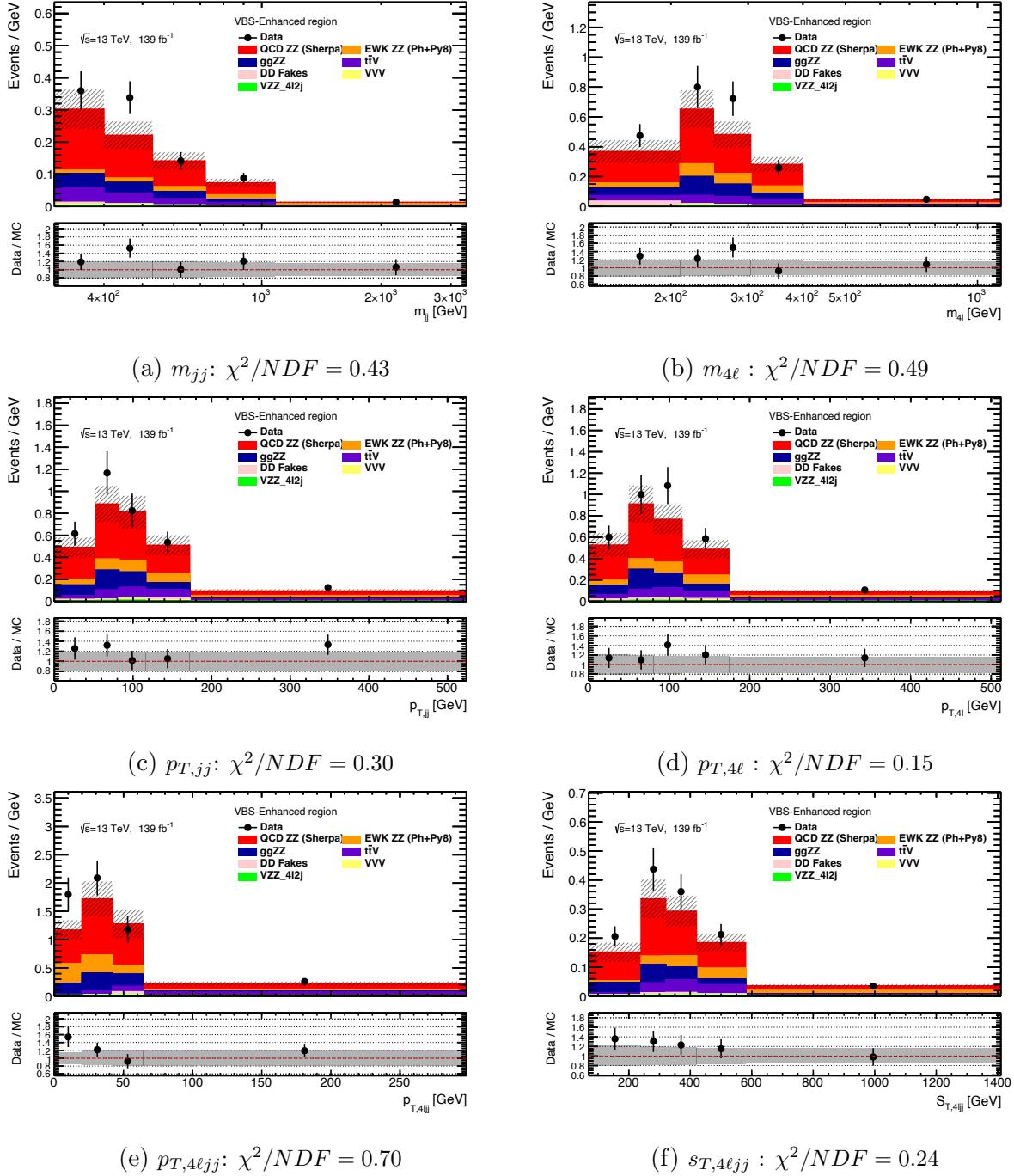


Figure 62: Detector level distributions in the VBS-Enhanced region.

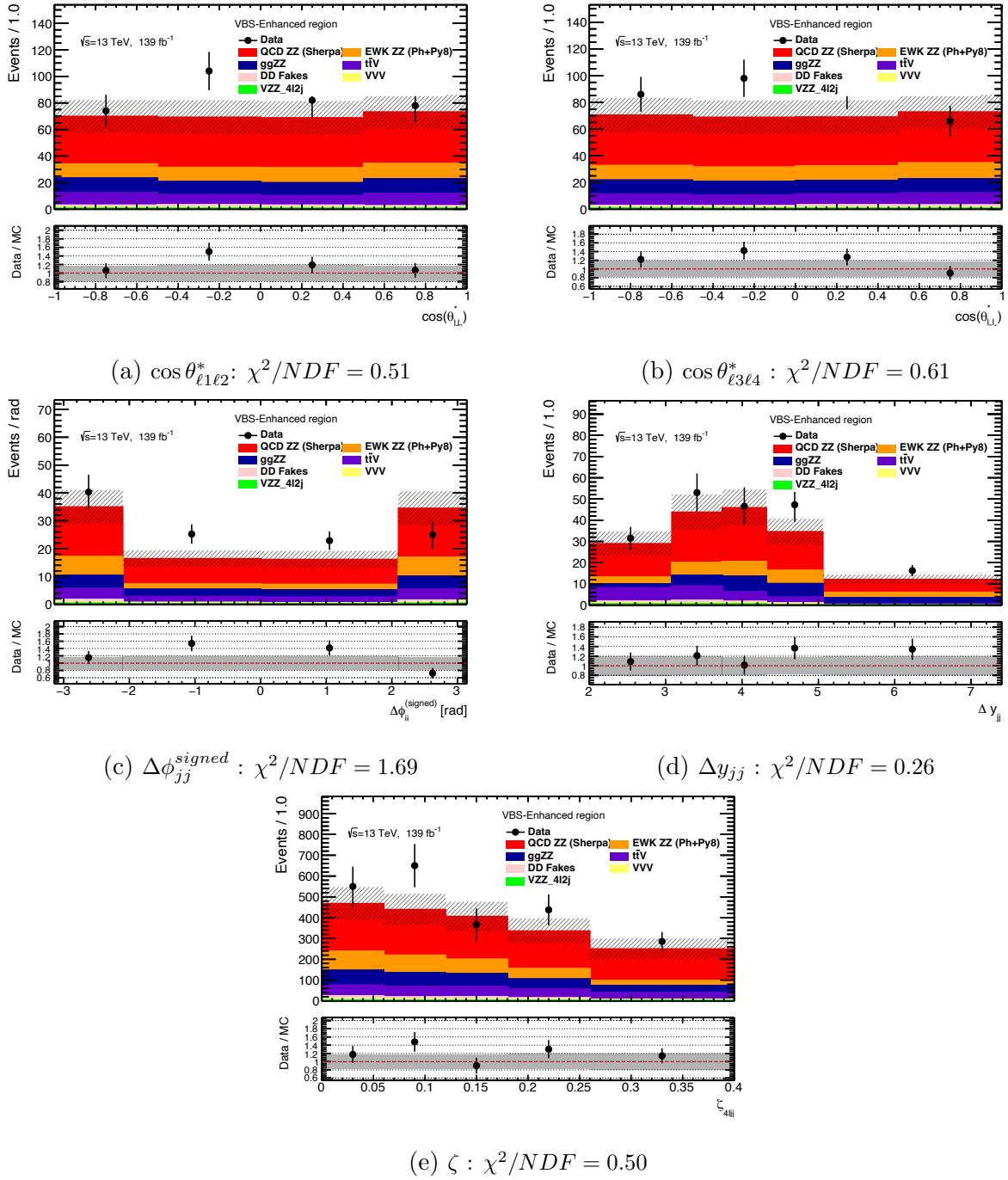


Figure 63: Detector level distributions in the VBS-Enhanced region.

## 20 Unfolded Differential Cross-sections

The background-subtracted data is unfolded using the iterative Bayesian Unfolding discussed in Section 17. The unfolded differential cross-sections, the main results of this thesis, are obtained by multiplying the inverse of the integrated luminosity to the background-subtracted unfolded yield in each bin. The unfolded differential cross-sections for eleven kinematic observables are shown in Figures 64 and 65.

Each distribution of unfolded differential cross-section is compared to two different state-of-the-art particle-level SM predictions, one in which the QCD  $qqZZ$  contribution is predicted by the SHERPA generator, and the other where the QCD  $qqZZ$  contribution is predicted by MADGRAPH generator. The light-red and light-blue bands are the fiducial-level systematics on the SHERPA and MADGRAPH particle-level cross-sections, respectively. The vertical error bars on the unfolded cross-sections are the statistical uncertainty. In contrast, the black bands represent the total systematic uncertainties on unfolded cross-sections from theoretical, experimental, and unfolding sources. The unfolded cross-sections are limited by statistical precision in all distributions and all bins. For some distributions, one or two events are found in the overflow bin resulting from bin migrations. These events are added to the content of the last bin of the unfolded distributions.

Generally, for all distributions, the data is well modeled by the MC simulations within  $2\sigma$  of the uncertainty band. Two p-values are determined by comparing unfolded cross-sections to the two predicted cross-sections to quantify the agreement between the experimentally measured unfolded data and the SM-predicted cross-sections. The p-values are calculated based on a technique discussed in Ref [100] by taking the residual and uncertainties of two weighted histograms. For all kinematic observables, the reported p-values obtained by comparing to either generator are more significant than 0.05. Therefore, in the analyzed LHC Run-2 dataset, for the  $ZZ^*(\rightarrow 4\ell)jj$  process, all differential cross-sections in the VBS-Enhanced region are concluded to agree with the SM predictions.

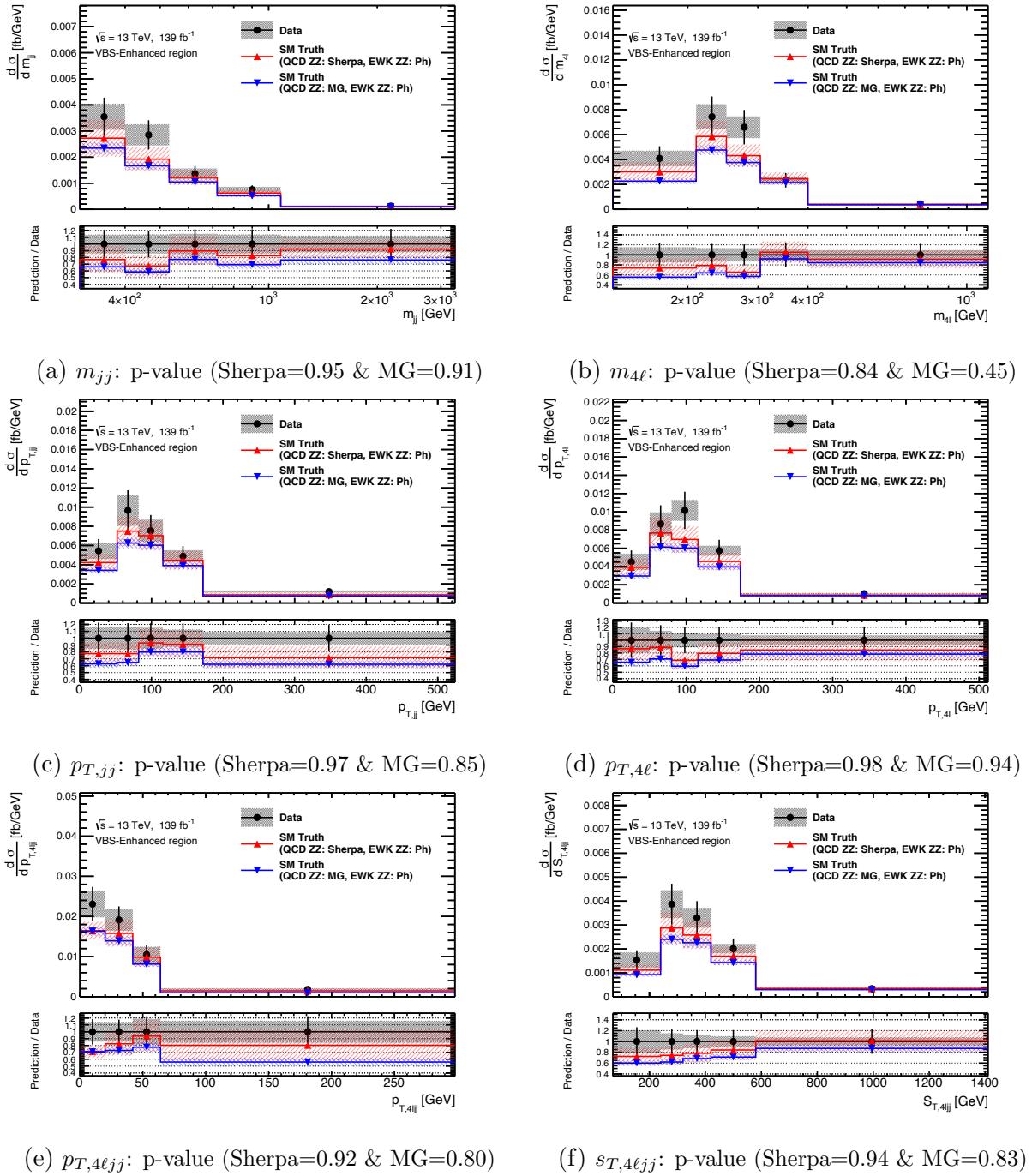


Figure 64: Unfolded differential cross-sections in the VBS-Enhanced region.

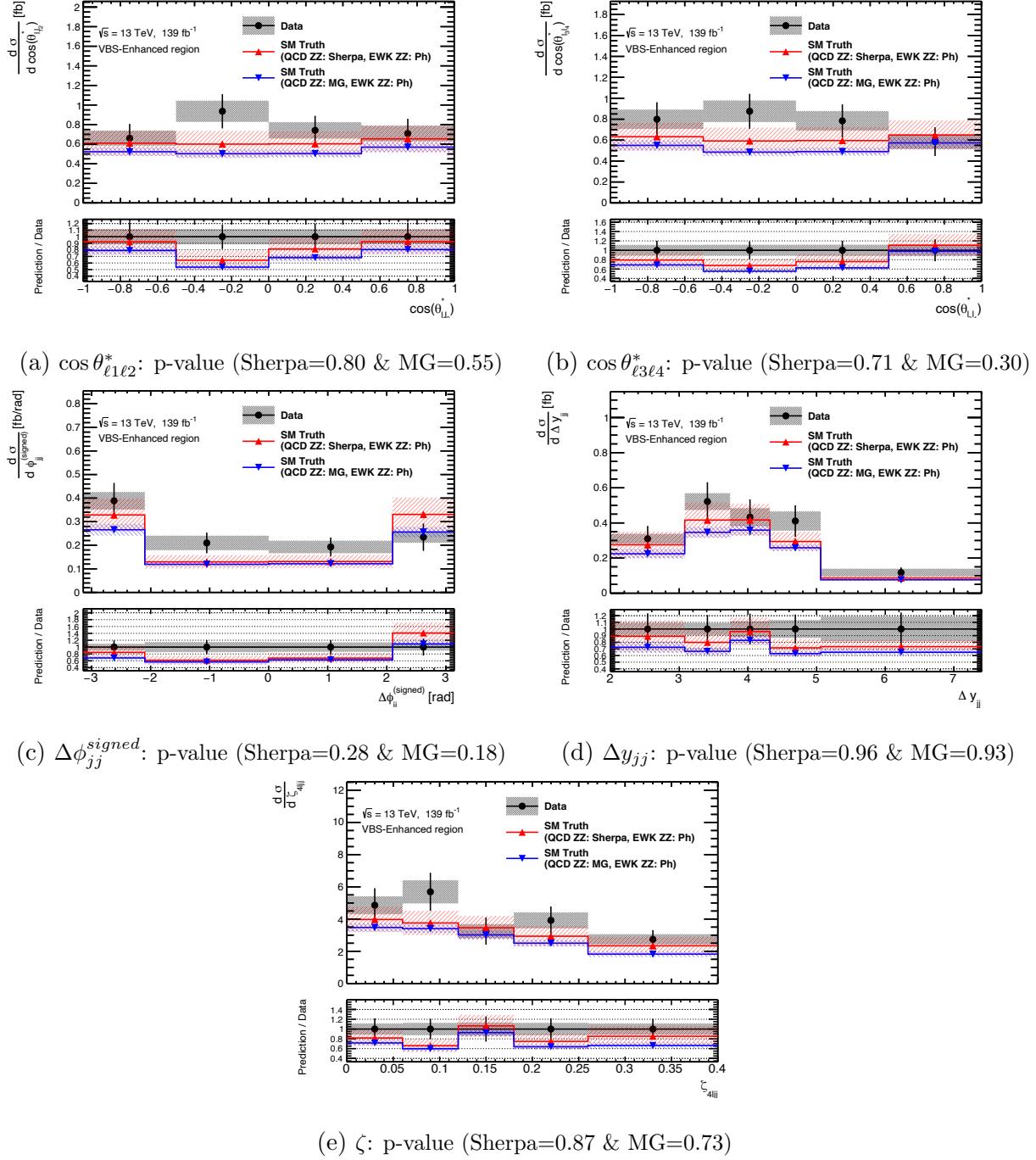


Figure 65: Unfolded differential cross-sections in the VBS-Enhanced region.

# 21 Effective Field Theory ReInterpretation

The unfolded cross-sections from data are reinterpreted to constrain the possible effects of new physics using a model-independent approach. The Effective Field Theory, along with the model and operators used, are briefly introduced in Section 21.1. The simulation of the EFT MC samples is discussed in Section 21.2. The statistical fit to constrain the contributions from BSM physics, including statistical and systematic uncertainties, is discussed in Section 21.3, and finally, Section 21.4 presents the obtained results.

## 21.1 Introduction

Similarly to the Fermi theory developed by Fermi to describe the beta decay before the formulation of the electroweak theory, the effects of BSM physics with any new heavy resonances and short range can be described with a model-independent EFT approach at low energy scales. Due to its large mass, the potential new resonance can affect processes below the cut-off scale  $\Lambda$  only through a virtual propagator, thus modifying the cross-sections for low-energy physics processes. The Lagrangian, including the new physics beyond the cut-off scale  $\Lambda$ , can be written through a Standard Model Effective Field Theory (SMEFT) formalism, where new physics describing operators are built with higher dimensions in the energy of SM fields. The SM Lagrangian is dimension four in energy, and the SMEFT Lagrangian conserving the  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$  symmetry is constructed by adding new interactions through the SM field operators with dimensions greater than four ( $d > 4$ ) as,

$$\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \sum_{d \geq 5} \sum_i \frac{c_i^d}{\Lambda^{d-4}} \mathcal{O}_i^d, \quad (21.1)$$

where  $\mathcal{O}_i^d$  is the higher dimension operator describing new physics with dimensionless coupling constants  $c_i^d$ , also known as the Wilson coefficients [101].

Any SMEFT Lagrangian given by Equation 21.1 should reduce to the SM at low energy

scales and is required to respect the unitarity bound such that the amplitude of any EFT process cannot grow too fast with a given energy scale. There are one dimension-five operator and 20 dimension-seven operators that meet all requirements of SMEFT [101] [102]. However, these operators violate the conservation of either baryon or lepton numbers, which to date are experimentally observed to be conserved. Moreover, higher dimensions EFT operators are suppressed by the larger order of magnitude of the large cut-off energy scale. Therefore, in practice, only dimension-six and dimension-eight operators are considered to affect the low-energy processes measured at the LHC.

The main results of the analysis are the unfolded differential cross-sections in electroweak-enhanced phase space. Thus, the EFT constrains the operators that affect the electroweak processes shown by the Feynman diagrams in Figure 7. There is a total of 59 dimension-six SMEFT operators, and 26 of them affect the electroweak processes by modifying either the self-interactions of the gauge bosons, the Higgs-vector boson interactions,  $Z \rightarrow \ell\ell$  vertices or the purely fermionic fields affecting either the four-leptons or lepton-quark interactions. However, based on an initial sensitivity study using SM predicted Asimov<sup>10</sup> dataset, it was observed that the constraints on these operators were more stringent in processes with more significant statistics such as inclusive four lepton measurements presented in Ref [103] or a global SMEFT fit using LEP, ATLAS, and CMS data discussed in Ref [104]. Therefore, in this thesis, only dimension-8 operators affecting the electroweak  $pp \rightarrow ZZ^*(\rightarrow 4\ell)jj$  processes will be considered.

The quartic vector boson self-interactions are experimentally accessible with the LHC Run-2 dataset for the first time in the  $ZZ^*jj$  process. There are EFT operators that modify the QGC vertex shown in Figure 7b, resulting in anomalous Quartic Gauge Couplings (aQGC). The aQGC operators relevant to the measurement are defined by the Eboli Model discussed in detail in Ref [49]. Table 17 shows the 18 dimension-8 operators from the Eboli model that give aQGC for processes involving multi-boson final states either by modify-

---

<sup>10</sup>SM predicted detector-level distributions where the bin errors are the Poissonian errors derived from the weighted event counts.

ing the SM electroweak interactions or by introducing the SM forbidden neutral couplings such as  $ZZZZ$ ,  $ZZZA$ ,  $ZZAA$ ,  $ZAAA$  and  $AAAA$ . Among these 18 operators, eight shown in Table 18 formed by the combination of different field strength tensors only affect the quartic gauge-self interactions of the vector bosons scattering without any impact to the triple self-interactions or Higgs-mediated processes. The measurement constraints these eight *genuine-QGC operators*.

Table 17: Eighteen dimension-8 operators from the Eboli model giving anomalous quartic gauge vertices in several multi-boson processes.

	WWWW	WWZZ	ZZZZ	WWAZ	WWAA	ZZZA	ZZAA	ZAAA	AAAA
$\mathcal{O}_{S,0}, \mathcal{O}_{S,1}$	x	x	x						
$\mathcal{O}_{M,0}, \mathcal{O}_{M,1}, \mathcal{O}_{M,6}, \mathcal{O}_{M,7}$	x	x	x	x	x	x	x		
$\mathcal{O}_{M,2}, \mathcal{O}_{M,3}, \mathcal{O}_{M,4}, \mathcal{O}_{M,5}$		x	x	x	x	x	x		
$\mathcal{O}_{T,0}, \mathcal{O}_{T,1}, \mathcal{O}_{T,2}$	x	x	x	x	x	x	x	x	x
$\mathcal{O}_{T,5}, \mathcal{O}_{T,6}, \mathcal{O}_{T,7}$		x	x	x	x	x	x	x	x
$\mathcal{O}_{T,8}, \mathcal{O}_{T,9}$			x		x	x	x	x	x

The amplitude of a process, including the aQGC EFT operators, depends on the SMEFT matrix element  $\mathcal{M}_{SMEFT}$  which can be written as,

$$\mathcal{M}_{SMEFT} = \mathcal{M}_{SM} + \sum_i \frac{c_i}{\Lambda^4} \mathcal{M}_i, \quad (21.2)$$

where  $\mathcal{M}_{SM}$  is the SM matrix element and  $\mathcal{M}_i$  is the matrix element of the EFT operator  $\mathcal{O}_i$ . Similarly, the cross-section depends on the square of the matrix element  $\mathcal{M}_{SMEFT}$ , which is

$$|\mathcal{M}_{SMEFT}|^2 = |\mathcal{M}_{SM}|^2 + 2 \sum_i \frac{c_i}{\Lambda^4} Re(\mathcal{M}_{SM}^* \mathcal{M}_i) + \sum_{i,j} \frac{c_i c_j}{\Lambda^8} Re(\mathcal{M}_i^* \mathcal{M}_j). \quad (21.3)$$

In the measurement, only one EFT operator is constrained per fit. Therefore there is no need to consider the interference between the two EFT operators represented by the last

Table 18: The eight genuine QGC dimension-8 operators constrained by the measurement.

Operators	Definition	Wilson Coefficient
$\mathcal{O}_{T,0}$	$Tr[\hat{W}_{\mu\nu}\hat{W}^{\mu\nu}] \times Tr[\hat{W}_{\alpha\beta}\hat{W}^{\alpha\beta}]$	$f_{T0}$
$\mathcal{O}_{T,1}$	$Tr[\hat{W}_{\alpha\nu}\hat{W}^{\mu\beta}] \times Tr[\hat{W}_{\mu\beta}\hat{W}^{\alpha\nu}]$	$f_{T1}$
$\mathcal{O}_{T,2}$	$Tr[\hat{W}_{\alpha\mu}\hat{W}^{\mu\beta}] \times Tr[\hat{W}_{\beta\nu}\hat{W}^{\nu\alpha}]$	$f_{T2}$
$\mathcal{O}_{T,5}$	$Tr[\hat{W}_{\mu\nu}\hat{W}^{\mu\nu}] \times B_{\alpha\beta}B^{\alpha\beta}$	$f_{T5}$
$\mathcal{O}_{T,6}$	$Tr[\hat{W}_{\alpha\nu}\hat{W}^{\mu\beta}] \times B_{\mu\beta}B^{\alpha\nu}$	$f_{T6}$
$\mathcal{O}_{T,7}$	$Tr[\hat{W}_{\alpha\mu}\hat{W}^{\mu\beta}] \times B_{\beta\nu}B^{\nu\alpha}$	$f_{T7}$
$\mathcal{O}_{T,8}$	$B_{\mu\nu}B^{\mu\nu}B_{\alpha\beta}B^{\alpha\beta}$	$f_{T8}$
$\mathcal{O}_{T,9}$	$B_{\alpha\mu}B^{\mu\beta}B_{\beta\nu}B^{\nu\alpha}$	$f_{T9}$

term. Thus, simplifying Equation 21.3 to,

$$|\mathcal{M}_{SMEFT}|^2 = |\mathcal{M}_{SM}|^2 + 2 \sum_i \frac{c_i}{\Lambda^4} Re(\mathcal{M}_{SM}^* \mathcal{M}_i) + \sum_{i,j} \frac{c_i^2}{\Lambda^8} |\mathcal{M}_i|, \quad (21.4)$$

where the second term is the *linear-only EFT term* representing the interference in the matrix element between SM and EFT, and the last term is the *quadratic term* representing the pure EFT-only contribution. The analysis uses a full EFT model considering effects from both the interference and the quadratic term.

The SMEFT predicted cross-section for a single operator involving an aQGC vertex is thus given as

$$\sigma_{SMEFT}^{pred} = \sigma_{SM}[1 + c.A_{Int} + c^2.B_{Quad}], \quad (21.5)$$

where  $A_{Int}$  and  $B_{Quad}$  are the corresponding relative linear and quadratic EFT contributions. The same is true when comparing the differential cross-sections, where cross-sections are

compared in each bin.

## 21.2 Event Generation

The EFT samples are generated for each dimension-8 EFT operator using **MADGRAPH5** at leading order using the NNPDF3.0NLO PDF set and ignoring all the QCD vertices. All Feynman diagrams contributing to the inclusive  $pp \rightarrow ZZ^*jj$  process are first generated using **MADGRAPH5** and interfaced with **PYTHIA8** to simulate parton showering, hadronization, and multiparton interaction. Some generator-level kinematic selections on jets such as  $\eta_j < 4.5$ ,  $p_{T,j} > 15$  GeV, and  $m_{jj} > 10$  GeV are applied to speed up the generation process. Each  $Z$ -boson is then decayed leptonically ( $Z \rightarrow \ell\ell$ ) using **MADSPIN**. For each EFT operator, the cross-sections for linear-only and quadratic-only terms are generated by setting the value of the relevant Wilson coefficient to 1 and the value of all other Wilson coefficients to 0. The differential cross-section in bin  $i$  for any value of the Wilson coefficient is then estimated by utilizing the following relation,

$$\frac{d\sigma^i}{dX}(c) = \frac{d\sigma_{SM}^i}{X} + c \cdot \frac{d\sigma_{Int,c=1}^i}{dX} + c^2 \cdot \frac{d\sigma_{Quad,c=1}^i}{dX}. \quad (21.6)$$

## 21.3 Statistical Strategy for Limit Setting

A profile likelihood technique [105] is used to constrain each Wilson coefficient at 95% confidence level. The method relies on building a chi-square ( $\chi^2$ ) statistic that measures the compatibility between the differential cross-sections predicted by SMEFT ( $\vec{\sigma}_{pred}$ ) to the unfolded differential cross-sections either from the predicted Asimov data or measured data ( $\vec{\sigma}_{data}$ ). For a given value of a selected Wilson Coefficient  $c_X$ , the  $\chi^2$  is given by

$$\chi^2(c_X, \vec{\theta}) = [\vec{\sigma}_{data} - \vec{\sigma}_{pred}(c_X) - \sum_{\theta} \theta \cdot \vec{e}_{\theta}]^T \mathbf{C}^{-1} [\vec{\sigma}_{data} - \vec{\sigma}_{pred}(c_X) - \sum_{\theta} \theta \cdot \vec{e}_{\theta}], \quad (21.7)$$

where  $\mathbf{C}$  is the total statistical and systematic covariance matrices on measured differential cross-sections ( $\vec{\sigma}_{data}$ ) and  $\theta$  is a vector including all nuisance parameters, where each component corresponds to a particle-level uncertainty on SM prediction with an absolute magnitude of  $e_\theta$ .

The likelihood function gives a conditional probability, encoding the probability of observing the data under the assumption of the SMEFT model, and is constructed as

$$\mathcal{L}(c_X, \vec{\theta}) = \frac{1}{\sqrt{(2\pi)^k \cdot |\mathbf{C}|}} \cdot e^{-\frac{1}{2}\chi^2(c_X, \vec{\theta})} \cdot \prod_{\theta} \mathcal{G}(\theta), \quad (21.8)$$

where  $k$  is the number of measurements equal to the length of  $\vec{\sigma}_{data}$ .  $\mathcal{G}(\theta)$  is the Gaussian-constrained nuisance parameter  $\theta$  with a mean of zero and standard deviation of one.

The profile likelihood ratio test statistic for a SMEFT model depending on a single parameter  $c$  is defined as

$$q(c) = -2 \ln \frac{\mathcal{L}(c, \hat{\vec{\theta}}(c))}{\mathcal{L}(\hat{c}, \hat{\vec{\theta}}(c))}, \quad (21.9)$$

where  $\hat{c}$  and  $\hat{\vec{\theta}}(c)$  in the denominator are the unconditional maximum likelihood estimators of  $c$  and  $\vec{\theta}(c)$ , while  $\hat{\vec{\theta}}(c)$  is the value of  $\vec{\theta}$  which maximizes the likelihood for a given value of  $c$ . Following the definition of the test statistic in Equation 21.9, likelihood in Equation 21.8 and  $\chi^2$  in Equation 21.7 and implementing the logarithmic relations, the test statistics simplifies to a summation of two terms: the  $\Delta\chi^2$  and difference in the sum of quadrature over the nuisance parameters as,

$$q(c) = \Delta\chi^2 + \Delta \sum_i \Theta^2. \quad (21.10)$$

Low test statistic values closer to zero infer a good agreement between the measured unfolded cross-sections and the SMEFT-predicted cross-sections. According to Wilk's theorem, the test statistic distribution in Equation 21.9 asymptotically approaches the  $\chi^2$  distribution for one degree of freedom [106]. Therefore, the 95% confidence limits (CL) for a Wilson

coefficient are constructed assuming Wilk's theorem by excluding values of  $c$  for which

$$\int_0^{q(c)} \chi^2(d.o.f = 1) dq > 95\%. \quad (21.11)$$

In 95% C.L. and one degree of freedom, the upper limit of the q-value should be 3.84.

In the measurement, a one-dimensional unfolded differential cross-section distribution of  $m_{jj}$  appended to the last four bins of  $m_{4\ell}$  distribution is used to constrain the dimension-8 Eboli EFT operators. This choice of combined 1D observable is motivated to maximize the sensitivity by using two observables but the inability to construct a full 2D observable due to low statistics. Both  $m_{jj}$  and  $m_{4\ell}$  distributions describe the same event and using a full combined distribution overconstrains the limits obtained from the fit. Therefore, the first bin of  $m_{4\ell}$  with no sensitivity to the EFT contribution is removed.

### Covariance of the Unfolded Cross-sections (C)

The covariance used in Equations 21.7 and 21.8 is a total sum of the statistical covariance matrix shown in Figure 66a and the systematic covariance matrix shown in Figure 66b. The statistical covariance is estimated by the *Bootstrapping method* discussed in detail in Ref [107], which generates a pseudo-detector-level dataset following a Poisson distribution with a mean one. Ten thousand toy replicas are generated by varying the content of each bin and unfolded using unfolding inputs from the nominal SM prediction. The correlation between the unfolded replica-distributions gives the covariance shown in Figure 66a.

The systematic covariance matrix includes all sources of theoretical, experimental, and unfolding uncertainties discussed in Section 18. The covariance is constructed by assuming each systematic uncertainty to be 100% correlated between bins and uncorrelated from any other systematic uncertainty. For instance, if the  $n^{th}$  systematic uncertainty of the  $i^{th}$  bin is the difference between the nominal and variation-applied yield  $\delta_{n,i}$ , then the systematic covariance in bin  $ij$  is given by summing the product of the difference over all various sources

of uncertainties,

$$C_{ij}^{syst.} = \sum_n \delta_{n,i} \delta_{n,j}. \quad (21.12)$$

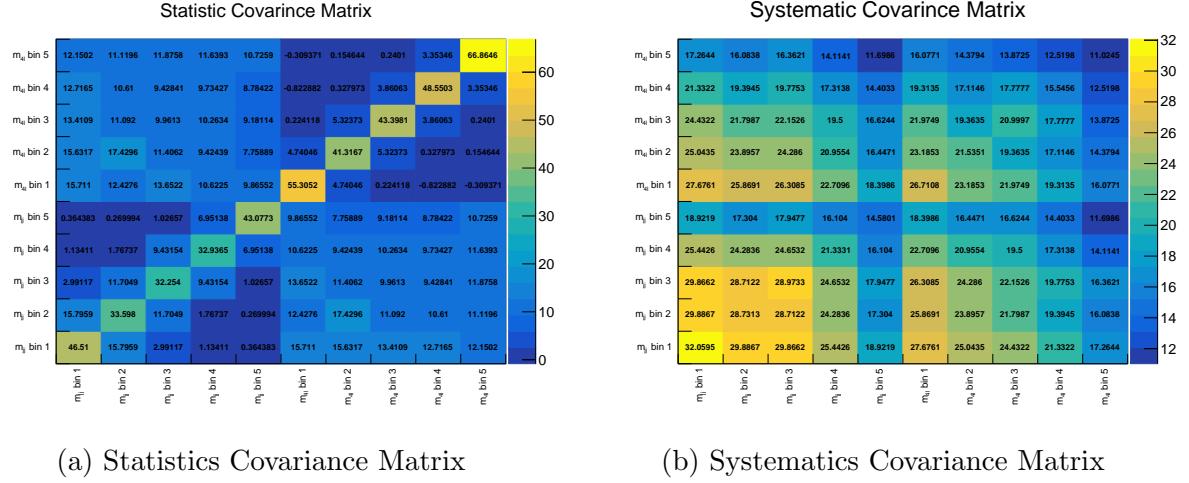


Figure 66: Covariance matrix for  $m_{jj} + m_{4\ell}$  unfolded differential cross-sections in the VBS-Enhanced region.

### Nuissance Parameters ( $\theta$ )

The following particle-level uncertainties are considered as nuisance parameters in the profile likelihood fit,

- **QCD Scale:** The same scale variations discussed in Section 18.1 are considered for the parton-initiated QCD  $qqZZ$ , gluon-loop initiated QCD  $ggZZ$  and EWK  $q q Z Z j j$  samples.
- **QCD PDF and  $\alpha_S$ :** The same set of PDF and  $\alpha_S$  at variations discussed in Section 18.1 are considered at the particle-level for the parton-initiated QCD  $qqZZ$ , gluon-loop initiated QCD  $ggZZ$  and EWK  $q q Z Z j j$  samples.
- **ggZZ NLO Re-weighting:** Uncertainty on the k-factor for the  $ggZZ$  sample is also taken into account.
- **QCD Modeling Uncertainties:** This uncertainty accounts for the modeling differences in the QCD  $qqZZ$  samples and is estimated by taking the difference between

SHERPA and MADGRAPH predictions.

Figure 67 shows an example of setting constraints on a single dimension-8  $\mathcal{O}_{T,0}$  operator with the cut-off energy scale set to 1 TeV. The example uses the unfolded differential cross-sections from the Asimov dataset. Figure 67a shows the likelihood scan as a function of different values of  $f_{T0}$  Wilson coefficient for two cases; first, only including the statistical covariances, and second, including both statistical and systematic covariances for the unfolded differential cross-sections. The two respective curves represent different test statistic ( $q$ ) values for different values of the Wilson coefficient. The intersection between the curves and the upper value of  $q = 3.84$  gives the 95% confidence limit for the  $f_{T0}$  Wilson coefficient to be  $[-0.898, 0.847]$ . Figure 67b shows the three distributions; SM predicted particle-level yield with fiducial-level theoretical uncertainties, the SMEFT particle-level prediction when the value of  $f_{T0}$  Wilson coefficient equals 0.847, the upper limit from the fit, as well as the unfolded SM Asimov data with statistical, experimental, theoretical, and unfolding uncertainties. Enhancements of about 20 – 25% on the last bin of  $m_{jj}$  and  $m_{4\ell}$  are observed when the EFT contribution is maximum allowed by the expected upper limit value for  $f_{T0}$ . Similarly, Figures 68a and 68b show the same but when using the unfolded differential cross-sections measured from data.

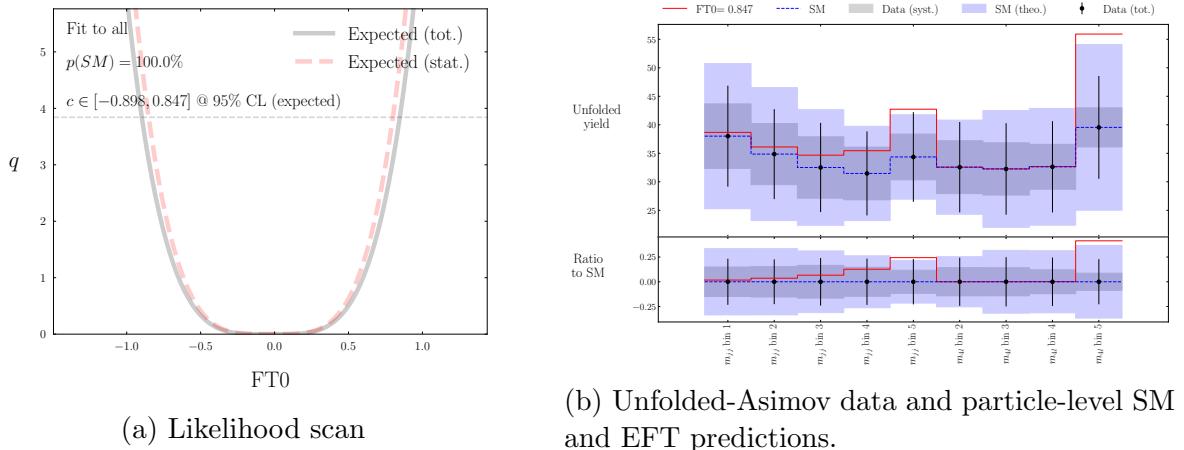


Figure 67: An example of setting constraints on dimension-8  $\mathcal{O}_{T,0}$  EFT operator with one-dimensional  $m_{jj} + m_{4\ell}$  Asimov-unfolded differential cross-sections.

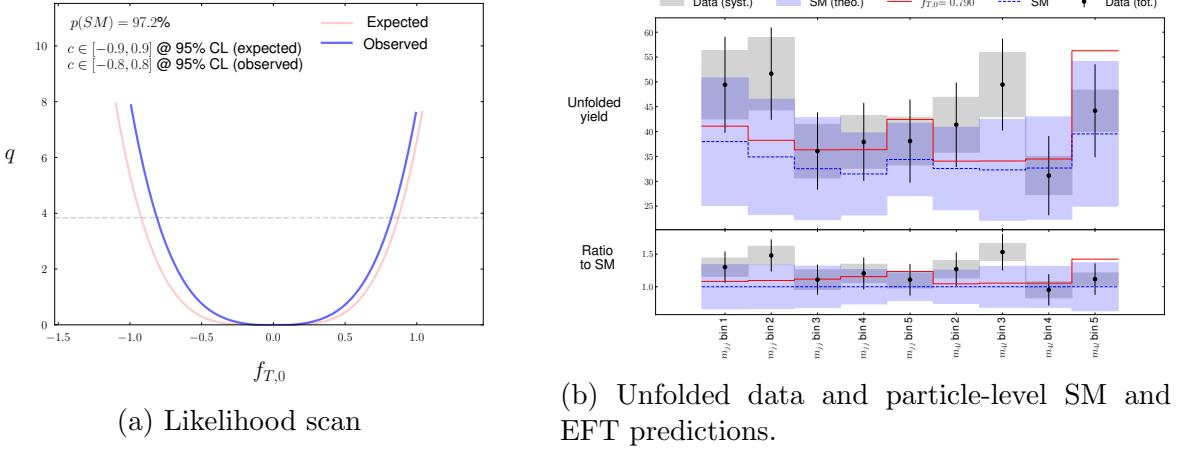


Figure 68: An example of setting constraints on dimension-8  $\mathcal{O}_{T,0}$  EFT operator with one-dimensional  $m_{jj} + m_{4\ell}$  unfolded differential cross-sections measured from data.

## 21.4 Results

The expected and observed 95% confidence limits on the Wilson coefficients associated with the 8 genuine QGC operators are shown in Table 19. The limits are obtained by fitting the unfolded differential cross-sections from SM prediction and measured data. The observed limits from data cross-sections are compatible with the expected limits obtained using the SM-predicted cross-sections. Therefore, it is concluded that the effects of new physics are not observed in the reinterpreted LHC Run-2 dataset. The obtained limits are three times larger than those obtained from CMS using detector-level  $ZZjj$  measurements [45].

Table 19: Expected and observed limits based on the unfolded Asimov and measured data, respectively, for the selected dimension-8 Wilson coefficients in Eboli Model. The limits are obtained by simultaneously fitting unfolded differential cross-sections of the  $m_{jj}$  and  $m_{4\ell}$ , including the overflow contribution in the VBS-Enhanced region.

Wilson Coefficient	Expected [Asimov]	Observed [Data]
$f_{T0}$	$[-9.0, 8.5] \times 10^{-1}$	$[-8.4, 7.9] \times 10^{-1}$
$f_{T1}$	$[-1.1, 1.1]$	$[-1.0, 1.0]$
$f_{T2}$	$[-2.3, 2.2]$	$[-2.2, 2.1]$
$f_{T5}$	$[-2.3, 2.2]$	$[-2.2, 2.1]$
$f_{T6}$	$[-3.6, 3.6]$	$[-3.3, 3.3]$
$f_{T7}$	$[-7.7, 7.4]$	$[-7.2, 6.9]$
$f_{T8}$	$[-1.9, 1.9]$	$[-1.8, 1.8]$
$f_{T9}$	$[-4.1, 4.1]$	$[-3.9, 3.9]$

# Chapter VII: Conclusion and Outlook

## 22 Conclusion

Vector boson scattering is a critical phenomenon in the electroweak sector of the Standard Model of Particle Physics. Vector boson scattering processes include rare triple and quartic self-couplings of the electroweak gauge bosons whose production rate at high energies is sensitive to possible modifications from physics beyond the Standard Model. The presence of clean signature of two same-flavor, opposite-sign lepton pairs in the  $ZZ^*(\rightarrow 4\ell)jj$  final state with minor contributions from background processes offers an excellent avenue to study the high-energy behavior of the vector boson scattering. Thus, the LHC Run-2 dataset collected by the ATLAS experiment during the 2015 – 2018 data-taking period corresponding to the integrated luminosity of  $139\text{ fb}^{-1}$  is analyzed to study VBS sensitive  $ZZ^*jj$  processes. However, given the low production cross-section of electroweak  $ZZ^*jj$  and small branching ratio of  $Z \rightarrow e^+e^-(\mu^+\mu^-)$ , these processes are statistically limited with the current dataset. Thus, unfolded differential cross-section measurements of  $ZZ^*(\rightarrow 4\ell)jj$  in an electroweak enhanced phase-space are measured as a function of eleven kinematic observables and compared to the state-of-the-art Standard Model predictions. The measured differential cross-sections agree with the theoretical predictions within the experimental and statistical uncertainties. The quartic gauge-self-coupling in  $ZZ^*jj$  final state is experimentally accessible for the first time with the ATLAS Run-2 datasets. The unfolded differential cross-sections are then used to put competitive constraints on beyond the Standard Model effects to the quartic gauge self-coupling of the vector bosons, using a model-independent effective field theory approach.

## 23 Outlook

With Run-2 datasets, the electroweak production of several multiboson processes such as VBS same-sign  $WW$  [108], VBS  $WWW$  [109], VBS  $WZ$  [110], and VBS  $ZZ$  [19] were experimentally observed for the first time with the ATLAS experiment. These VBS measurements are still statistically dominated and could gain higher precision from more extensive statistics. Run-3 started in July of 2022 and is expected to continue until 2025 at the highest to date center-of-mass-energy of proton-proton collisions,  $\sqrt{s} = 13.6$  TeV [111]. In Run-3, the ATLAS experiment is expected to record more than twice the dataset of Run-2, corresponding to an integrated luminosity of  $300\ fb^{-1}$ . Run-3 statistics are crucial to study the cross-sections of any VBS processes differentially. The Run-3 datasets are expected to make the differential cross-sections measurement of the fully electroweak  $ZZ^*(\rightarrow 4\ell)jj$  process statistically feasible and put stronger constraints on the BSM parameters causing anomalous self-interactions of the gauge bosons.

However, the HL-LHC discussed briefly in Section 8.1 is expected to start the golden era for the vector boson scattering measurements. The ATLAS experiment is expected to record about ten times more data with a more precise reconstruction of the forward jets, essential physics objects defining the VBS processes, through the extended  $\eta$  coverage of the inner tracker and the additional timing information from the high granularity timing detectors. With the extensive statistics and the unprecedented proton-proton collisions at a center-of-mass energy of  $\sqrt{s} = 14$  TeV at the HL-LHC, the scattering of the longitudinally polarized vector bosons is expected to be within experimental reach [112]. As discussed in Section 4, the self-interactions of the longitudinally polarized vector bosons are regularized by the Higgs-mediated processes to restore the unitarity at high energies. The ultimate goal of the Standard Model electroweak multiboson measurements is to experimentally measure the cross-sections of VBS processes for longitudinally polarized vector bosons.

## References

- [1] P.A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020.
- [2] Francis Halzen and Alan Martin. *Quarks & Leptons: An introductory course in modern particle physics*. John Wiley & Sons, New York, USA, 1984.
- [3] Wikipedia contributors. Standard Model of Elementary Particles. [https://en.wikipedia.org/wiki/Standard\\_Model](https://en.wikipedia.org/wiki/Standard_Model), 2019. [Online; accessed 05-November-2022].
- [4] Luigi Del Debbio. Parton distributions in the LHC era. *EPJ Web of Conferences*, 175:01006, Jan 2018.
- [5] Ana Alboteanu, Wolfgang Kilian and Jürgen Reuter. Resonances and unitarity in weak boson scattering at the LHC. *J. High Energ. Phys.*, 2008(11):010–010, Nov 2008.
- [6] Fabienne Marcastel. CERN’s Accelerator Complex. La chaîne des accélérateurs du CERN, 2013. General Photo.
- [7] The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(08):S08003, Aug 2008.
- [8] Sebastian Murk. Trigger Algorithms and Monte Carlo Event Generation for Dijet Searches in ATLAS and Beyond, 2017.
- [9] The ATLAS Collaboration. Alignment of the ATLAS Inner Detector in Run-2. *Eur. Phys. J. C*, 80(12), 2020.
- [10] Joao Pequenao. Event Cross Section in a computer generated image of the ATLAS detector., 2008.
- [11] The ATLAS Collaboration. DAQ approved plots, 2020.
- [12] T. G. Cornelissen et al. Updates of the ATLAS Tracking Event Data Model (Release 13). Technical report, CERN, Geneva, 2007.
- [13] ATLAS Tracking Combined Performance Group. ATLAS Track Reconstruction. <https://atlassoftwaredocs.web.cern.ch/trackingTutorial/>, 2021. [Online; accessed 15-November-2022].
- [14] The ATLAS Collaboration. Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at  $\sqrt{s} = 13$  TeV. *Eur. Phys. J. C*, 79(8), Aug 2019.
- [15] Sebastien Rettie. Muon identification and performance in the ATLAS experiment. Technical report, CERN, Geneva, 2018.
- [16] The ATLAS Collaboration. Jet reconstruction and performance using particle flow with the ATLAS Detector . *Eur. Phys. J. C*, 77(7), Jul 2017.

- [17] The ATLAS Collaboration. Jet energy scale and resolution measured in proton–proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *Eur. Phys. J. C*, 81(8):689, 2021.
- [18] The ATLAS Collaboration. Technical Design Report for the ATLAS Inner Tracker Pixel Detector. Technical report, CERN, Geneva, 2017.
- [19] The ATLAS Collaboration. Observation of electroweak production of two jets and a  $Z$ -boson pair, 2020.
- [20] The ATLAS Collaboration. ATLAS data quality operations and performance for 2015–2018 data-taking. *Journal of Instrumentation*, 15(04):P04003–P04003, Apr 2020.
- [21] Yanyan Gao et al. Spin determination of single-produced resonances at hadron colliders. *Phys. Rev. D*, 81(7), Apr 2010.
- [22] Mark Thomson. *Modern particle physics*. Cambridge University Press, New York, 2013.
- [23] E. Noether. Invariant variation problems. *Transport Theory and Statistical Physics*, 1(3):186–207, Jan 1971.
- [24] Jose Bernabeu. *Fundamental Physics and Physics Education Research*, chapter Symmetries in the Standard Model, pages 3–16. Springer International Publishing, Cham, 2021.
- [25] The OPAL Collaboration. Measurement of the running of the QED coupling in small-angle Bhabha scattering at LEP. *Eur. Phys. J. C*, 45(1):1–21, Jan 2006.
- [26] Alexandre Deur, Stanley J. Brodsky and Guy F. de Téramond. The QCD running coupling. *Progress in Particle and Nuclear Physics*, 90:1–74, Sep 2016.
- [27] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579–588, 1961.
- [28] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.
- [29] A. Salam and J.C. Ward. Electromagnetic and weak interactions. *Physics Letters*, 13(2):168–171, 1964.
- [30] K. Nguyen. The Higgs Mechanism, 2009.
- [31] M. E. Peskin and D. V. Schroeder. *An Introduction to Quantum Field Theory*. Westview PRESS, Colorado, USA, 1995.
- [32] W. Buchmüller and C. Lüdeling. Field Theory and Standard Model, 2006.
- [33] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, Sep 2012.

- [34] The ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012.
- [35] The CDF II Collaboration. High-precision measurement of the W boson mass with the CDF II detector. *Science*, 376(6589):170–176, 2022.
- [36] Richard Massey, Thomas Kitching and Johan Richard. The dark matter of gravitational lensing. *Reports on Progress in Physics*, 73(8):086901, Jul 2010.
- [37] Y. Ashie et al. Evidence for an Oscillatory Signature in Atmospheric Neutrino Oscillations. *Phys. Rev. Lett.*, 93:101801, Sep 2004.
- [38] Alex Keshavarzi et al. Muon g-2: A review. *Nuclear Physics B*, 975:115675, 2022.
- [39] R. Aaij et al. Tests of Lepton Universality Using  $B^0 \rightarrow K_S^0 \ell^+ \ell^-$  and  $B^+ \rightarrow K_S^{*+} \ell^+ \ell^-$  Decays. *Phys. Rev. Letters*, 128(19), May 2022.
- [40] Peter Skands. QCD for Collider Physics, 2011.
- [41] Ikuo Ito. IR Divergences in QCD and Kinoshita-Lee-Nauenberg Theorem. *Progress of Theoretical Physics*, 67(4):1216–1231, Apr 1982.
- [42] Alekhin Sergey, Melnikov Kirill and Petriello Frank. Fixed target Drell-Yan data and NNLO QCD fits of parton distribution functions. *Phys. Rev. D*, 74:054033, Sep 2006.
- [43] Sergey Alekhin et al. The PDF4LHC Working Group Interim Report, 2011.
- [44] Andy Buckley et al. General-purpose event generators for LHC physics. *Physics Reports*, 504(5):145–233, 2011.
- [45] The CMS Collaboration. Evidence for electroweak production of four charged leptons and two jets in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *Physics Letters B*, 812:135992, 2021.
- [46] B. Jager, A. Karlberg and G. Zanderighi. Electroweak ZZjj production in the Standard Model and beyond in the POWHEG-BOX V2. *J. High Energ. Phys.*, 141, 2014.
- [47] A. Denner and T. Hahn. Radiative corrections to  $W+W^- \rightarrow W+W^-$  in the electroweak standard model. *Nuclear Physics B*, 525(1-2):27–50, 1998.
- [48] Tao Han et al. New physics signals in longitudinal gauge boson scattering at the LHC. *J. High Energ. Phys.*, 2010(3), Mar 2010.
- [49] E. S. Almeida, O. J. P. Éboli and M. C. Gonzalez-Garcia. Unitarity constraints on anomalous quartic couplings. *Phys. Rev. D*, 101:113003, Jun 2020.
- [50] Jung Chang et al. WW Scattering in the Era of Post Higgs Discovery. *Phys. Rev. D*, 87(9), May 2013.

- [51] P. H. Williams. A Z-monitor to calibrate Higgs production via vector boson fusion with rapidity gaps at the LHC. *Eur. Phys. J.*, C 26:429–440, 2003.
- [52] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008.
- [53] CERN. *LEP design report*. CERN, Geneva, 1984.
- [54] CERN. LHC Guide. CERN-Brochure-2017-002-Eng, 2017.
- [55] M. Capeans et al. ATLAS Insertable B-Layer Technical Design Report. Technical report, CERN, 2010.
- [56] The ATLAS Collaboration. Technical Design Report for the ATLAS Inner Tracker Pixel Detector. Technical report, CERN, Geneva, 2017.
- [57] The ATLAS Collaboration. Technical Design Report for the ATLAS Inner Tracker Strip Detector. Technical report, CERN, Geneva, 2017.
- [58] The ATLAS Collaboration. Particle Identification Performance of the ATLAS Transition Radiation Tracker. Technical report, CERN, Geneva, 2011.
- [59] The ATLAS Collaboration. *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical design report. ATLAS. CERN, Geneva, 1996.
- [60] The ATLAS Collaboration. *ATLAS tile calorimeter: Technical Design Report*. Technical design report. ATLAS. CERN, Geneva, 1996.
- [61] The ATLAS collaboration. Operation of the ATLAS trigger system in Run 2. *Journal of Instrumentation*, 15(10):P10004, Oct 2020.
- [62] The ATLAS Collaboration. Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2. *Eur. Phys. J. C*, 77(10), 2017.
- [63] Federico Meloni. Primary vertex reconstruction with the ATLAS detector. Technical Report 12, CERN, Geneva, 2016.
- [64] The ATLAS Collaboration. Muon reconstruction performance of the ATLAS detector in proton-proton collision data  $\sqrt{s} = 13$  TeV. *Eur. Phys. J. C*, 76(5), May 2016.
- [65] The ATLAS Collaboration. Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data. *Eur. Phys. J. C*, 74(11), Nov 2014.
- [66] Matteo Cacciari, Gavin P. Salam and Gregory Soyez. The anti-kt jet clustering algorithm. *J. High Energ. Phys.*, 2008(04):063, Apr 2008.
- [67] The ATLAS Collaboration. Tagging and suppression of pileup jets with the ATLAS detector. Technical report, CERN, Geneva, 2014.
- [68] The ATLAS Collaboration. Forward jet vertex tagging using the particle flow algorithm. Technical report, CERN, Geneva, 2019.

- [69] The ATLAS Collaboration. *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2020.
- [70] The ATLAS Collaboration. Technical Design Report for the Phase-II Upgrade of the ATLAS Muon Spectrometer. Technical report, CERN, Geneva, 2017.
- [71] Mengqing Wu. FELIX: Readout upgrade for the ATLAS Trigger DAQ system in HL-LHC. *Nuclear Instruments and Methods in Physics*, 1048:167994, 2023.
- [72] The ATLAS Collaboration. Technical Design Report for the ATLAS Inner Tracker Strip Detector. Technical report, CERN, Apr 2017.
- [73] The ATLAS Collaboration. Expected Tracking Performance of the ATLAS Inner Tracker at the HL-LHC. Technical report, CERN, Geneva, 2019.
- [74] The ATLAS Collaboration. Tagging and suppression of pile-up jets in the forward region using timing information with the ATLAS detector at  $\sqrt{s} = 14$  TeV at HL-LHC. Technical report, CERN, Geneva, 2022.
- [75] The ATLAS Collaboration. ATLAS b-jet identification performance and efficiency measurement with  $t\bar{t}$  events in pp collisions at  $\text{sqrt}s = 13$  TeV. *Eur. Phys. J. C*, 79(11), Nov 2019.
- [76] G. Avoni et al. The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS. *Journal of Instrumentation*, 13(07):P07017, Jul 2018.
- [77] The ATLAS Collaboration. Luminosity determination in  $pp$  collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector at the LHC. Technical report, CERN, Jun 2019.
- [78] S. Agostinelli et al. GEANT4—a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.
- [79] Enrico Bothmann et al. Event generation with Sherpa 2.2. *SciPost Physics*, 7(3), Sep 2019.
- [80] Richard D. Ball et al. Parton distributions for the LHC run II. *J. High Energ. Phys.*, 2015(4), Apr 2015.
- [81] S. Schumann and F. Krauss. A parton shower algorithm based on Catani-Seymour dipole factorisation. *J. High Energ. Phys.*, 2008(03):038, Mar 2008.
- [82] Stefan Höche et al. QCD matrix elements + parton showers. The NLO case. *J. High Energ. Phys.*, 2013(27), 2013.
- [83] 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. *J. High Energ. Phys.*, 2014(7), Jul 2014.
- [84] Torbjörn Sjöstrand, Stephen Mrenna and Peter Skands. A brief introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178(11):852–867, Jun 2008.

- [85] Barbara Jäger, Alexander Karlberg and Giulia Zanderighi. Electroweak ZZjj production in the Standard Model and beyond in the POWHEG-BOX V2. *J. High Energ. Phys.*, 2014(3), Mar 2014.
- [86] Stefano Catani et al. QCD Matrix Elements + Parton Showers. *J. High Energ. Phys.*, 2001(11):063–063, Nov 2001.
- [87] Torbjörn Sjöstrand et al. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.
- [88] The ATLAS Collaboration. ATLAS  $b$ -jet identification performance and efficiency measurement with  $t\bar{t}$  events in  $pp$  collisions at  $\sqrt{s} = 13$  TeV. *Eur. Phys. J.*, 79(11), 2019.
- [89] J. Maurer, J. Reichert and E. Varnes. Technical discussion of FakeBkgTools. FF, 2018.
- [90] Giulio D’Agostini. A multidimensional unfolding method based on Bayes’ Theorem. Technical report, DESY, Hamburg, 1994.
- [91] G. D’Agostini. Improved iterative Bayesian unfolding, 2010.
- [92] Lydia Brenner et al. Comparison of unfolding methods using RooFitUnfold. *International Journal of Modern Physics A*, 35(24):2050145, 2020.
- [93] Jerome H. Friedman. A VARIABLE SPAN SMOOTHING. *Journal of American Statistical Association*, Oct 1984.
- [94] Enrico Bothmann, Marek Schönherr and Steffen Schumann. Reweighting QCD matrix-element and parton-shower calculations. *Eur. Phys. J. C*, 76(11), Oct 2016.
- [95] L. A. Harland-Lang et al. Parton distributions in the LHC era: MMHT 2014 PDFs. *Eur. Phys. J. C*, 75(5), May 2015.
- [96] Sayipjamal Dulat et al. New parton distribution functions from a global analysis of quantum chromodynamics. *Phys. Rev. D*, 93(3), Feb 2016.
- [97] Fabrizio Caola et al. QCD corrections to vector boson pair production in gluon fusion including interference effects with off-shell Higgs at the LHC. *J. High Energ. Phys.*, 2016(7), Jul 2016.
- [98] The ATLAS Collaboration. Measurement of the  $t\bar{t}Z$  and  $t\bar{t}W$  cross sections in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *Phys. Rev. D*, 99:072009, Apr 2019.
- [99] The ATLAS Collaboration. Evidence for the production of three massive vector bosons with the ATLAS detector. *Physics Letters B*, 798:134913, 2019.
- [100] N. D. Gagunashvili. Comparison of weighted and unweighted histograms. *Proceedings of Science*, 2006.

- [101] Ilaria Brivio, Yun Jiang and Michael Trott. The SMEFTsim package, theory and tools. *J. High Energ. Phys.*, 2017(12), Dec 2017.
- [102] Landon Lehman. Extending the standard model effective field theory with the complete set of dimension-7 operators. *Phys. Rev. D*, 90(12), Dec 2014.
- [103] The ATLAS Collaboration. Measurements of differential cross-sections in four-lepton events in 13 TeV proton-proton collisions with the ATLAS detector. *J. High Energ. Phys.*, 2021(7), Jul 2021.
- [104] John Ellis, Christopher W. Murphy, Verónica Sanz and Tevong You. Updated global SMEFT fit to Higgs, diboson and electroweak data. *J. High Energ. Phys.*, 2018(6), Jun 2018.
- [105] Glen Cowan, Kyle Cranmer, Eilam Gross and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71(2), Feb 2011.
- [106] Wilks, S. S. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *Annals Math. Statist.*, 9(1):60–62, 1938.
- [107] The ATLAS Collaboration. Evaluating statistical uncertainties and correlations using the bootstrap method. Technical report, CERN, Geneva, 2021.
- [108] The ATLAS Collaboration. Observation of electroweak production of a same-sign  $W$  boson pair in association with two jets in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *Phys. Rev. Lett.*, 123(16), Oct 2019.
- [109] The ATLAS Collaboration. Observation of  $WW\bar{W}$  Production in  $pp$  Collisions at  $\sqrt{s} = 13$  TeV with the ATLAS Detector. *Phys. Rev. Lett.*, 129(6), Aug 2022.
- [110] The ATLAS Collaboration. Observation of electroweak  $W^\pm Z$  boson pair production in association with two jets in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *Physics Letters B*, 793:469–492, Jun 2019.
- [111] The ATLAS Collaboration. ATLAS Experiment records “first physics” at new high-energy frontier. CERN, Jul 2022 [Online].
- [112] The ATLAS Collaboration. Prospects for the measurement of the  $W^\pm W^\pm$  scattering cross section and extraction of the longitudinal scattering component in  $pp$  collisions at the High-Luminosity LHC with the ATLAS experiment. Technical report, CERN, Geneva, 2018.

# Appendices

Supplementary materials to the main body of this thesis are presented in the following appendices. Appendix A summarizes my contributions to the measurement presented in this thesis as well as to the ATLAS experiment. The largest source of systematic uncertainty on the results is the unfolding bias and Appendix B summarizes the studies conducted to understand the source of this bias. Finally, Appendix C presents the detector-level measurements and the differential unfolded cross-sections measured in the VBS-suppressed region of the phase space.

## A Personal Contribution

### A.1 Contribution to $ZZ^*(\rightarrow 4\ell)jj$ Measurement

The measurements presented in this thesis are the result of the effort of the ATLAS "Differential VBS  $ZZ^*jj$ " analysis team. As one of the two primary analyzers in this team, I have contributed to several aspects of this measurement, from the initial formation to the current stage of finalization and publication. I contributed significantly to defining the analysis phase space discussed in Chapter *IV*. The phase space defined in the preceding ATLAS analysis [19], which observed the electroweak  $ZZjj$  process was not optimal for differential measurements. Therefore, I led the studies performed to modify the kinematic selections to increase the acceptance, defined the isolation working point to maintain optimal signal-selection and background-rejection probabilities, and established novel pair sorting strategy to reduce the bin migration in unfolding. Moreover, I developed and maintained the main analysis framework, which implements the latest recommendations from combined performance groups for physics object reconstruction discussed in Chapter *III*, and performs the kinematic selections discussed in Chapter *IV*. This framework applies all the scale factors and event weights in MC events and the final events used in the measurement. Additionally, I estimated the non-prompt backgrounds of the analysis using the background estimation technique discussed in Section 16 and contributed to selecting the relevant systematics discussed in Section 18 of Chapter *V*. Finally, I validated the novel next-to-leading-order POWHEG  $qqZZjj$  MC sample used as the primary sample for the electroweak production of  $ZZ^*(\rightarrow 4\ell)jj$ . I checked the compatibility between the measured data and SM prediction.

### A.2 Contribution to the ATLAS Experiment

I have been a member of the ATLAS collaboration since 2017 and have contributed to the three critical areas of the experiment, detector development, detector performance, and

physics analyses.

**Detector Development:** I spent my first year at Brandeis working at Brookhaven National Lab, where I contributed to the prototype development of the all-silicon inner tracker for the High-Luminosity LHC. During this period, I assembled the first three prototypes of the ITk strip detector using a semi-automated loading setup consisting of an Aeoreetch robotic arm, cameras, and an alignment system. I developed 3D-printed tooling pins used in the alignment of loading the sensors on staves. I led comprehensive thermal and mechanical tests of the first stave prototype, the thermo-mechanical stave, using IR imaging and laser metrology, respectively. These tests validated the cooling system designs for the ITk stave core structure and the stability of the mechanical design. During this year, I learned the fundamentals of Particle Physics detectors, their development, and their operations.

**Physics Analysis:** Apart from the  $ZZ^*(\rightarrow 4\ell)jj$  analysis presented in this thesis, I have worked on two other ATLAS analyses involving four leptons in the final state, analyzing the full Run-2 dataset. After my Ph.D. candidacy and ATLAS qualification, I joined the inclusive four-lepton measurement analysis team, whose goal was to inclusively measure the differential unfolded cross-sections of the Standard Model four-lepton process. In my first year, I worked on identifying the suitable lepton isolation working points, studied the impact of including electrons and muons originating from tau leptons in the unfolded differential cross-sections, and obtained the most precise measurement to date of the branching ratio of  $Z \rightarrow 4\ell$ .

Since early 2021, I have been a key member of the on-shell  $ZZ$  CP and polarization analysis team, where the main goal is to extract the fraction of two  $Z$  bosons simultaneously longitudinally polarized and search for additional CP violation. Like the  $ZZ^*(\rightarrow 4\ell)jj$  analysis, I contributed to background estimation, phase space optimization, and event selection. Additionally, I derived the differential unfolded cross-sections with all relevant systematic and statistical uncertainties used in the CP violation search using an effective field theory approach.

**Detector Performance:** The training of a particle physicist is incomplete without understanding the detector’s performance. Therefore, in 2021 I joined the ATLAS Tracking Combined Performance group to validate the Run-3 tracking reconstruction software’s performance in early Run-3 data and different types of MC simulation.

Similarly to  $ZZ^*(\rightarrow 4\ell)jj$  measurement, most ATLAS physics analyses use the vertex with the highest value of the sum-squared of track’s transverse momenta ( $\sum_{tracks} p_T^2$ ) as the hard scattering vertex of the measurement. However, in processes with softer leptons and invisible tracks (including photons),  $\sum_{tracks} p_T^2$  is inadequate to identify the hard scatter vertex. Therefore, I am currently working on developing an alternative algorithm that is suitable for a variety of different physics processes.

The experiences I have gained in different areas of Particle Physics have significantly shaped my discussion of the measurement presented in this thesis.

## B Additional Study on Unfolding Bias

As discussed in Section 17.4, the bias from the unfolding process is the largest source of the systematic uncertainty for the measurement. Additional studies were conducted to understand the underlying source of this bias. Figure 69 shows the unfolding bias and statistical uncertainty on the unfolded yield as a function of the increasing number of unfolding iterations for each bin of  $m_{jj}$  in the VBS-Enhanced region. The bias is evaluated using the MC-toy-based method introduced in Section 17.4. The total uncertainty (black) is always the smallest for a single iteration, further assuring the choice of iteration is optimal. As expected, the bias decreases, and the statistical uncertainty increases with the increasing number of unfolding iterations.

The unfolding bias is expected to converge to a value of zero with a higher number of iterations. However, as observed in Figure 69, the rate of convergence of unfolding bias to zero is lower, suggesting that the fiducial fakes present in the detector-level distributions are not fully corrected by the unfolding method. The fiducial fraction, as shown by Figure 53, is usually between 60 – 80% in this measurement. To confirm that a high fraction of fiducial fakes cause the unfolding bias, these are subtracted manually from the MC predictions of nominal and toy distributions. The MC-toy-based unfolding bias estimate is repeated, and Figure 70 shows the resulting unfolding bias in each bin of  $m_{jj}$  in the VBS-Enhanced region. Compared to the nominal unfolding bias shown in Figure 59, Figure 70 has a smaller bias in each bin, and the convergence of the bias to zero is faster.

The differential measurements of the  $ZZ^*(\rightarrow 4\ell)jj$  process are statistically limited, so it is impossible to directly subtract the fiducial fakes from data without degrading the statistical precision and adding a new source of systematic uncertainty from the subtraction of fiducial fake events. However, it is imperative to understand the origin and topology of the fiducial fake events to reduce their impact without degrading the unfolding performance. Figures 71a and 71b show the fake fraction, the fraction of detector-level events passing detector-

level selection but failing the particle-level selection as a function of  $p_T$  and  $\eta$ , respectively, for the leading (x-axis) and the sub-leading (y-axis) jets. More significant fractions of fakes are observed in low- $p_T$  and high  $\eta$  region, which is likely related to the worse resolution of jet reconstruction in low- $p_T$  and smaller efficiency of fJVT tagging in forward regions. The large fraction of fiducial fakes is understood to originate either from migrations outside the fiducial volume due to jet resolution effects or from wrongfully selecting events with pile-up jets. Stronger kinematic selections were applied to the leading and sub-leading jets in an attempt to reduce the bias, but this resulted in the degradation of unfolding performance due to low statistics. Therefore, the nominal bias shown in Section 18.3 was deemed optimal for the measurement.

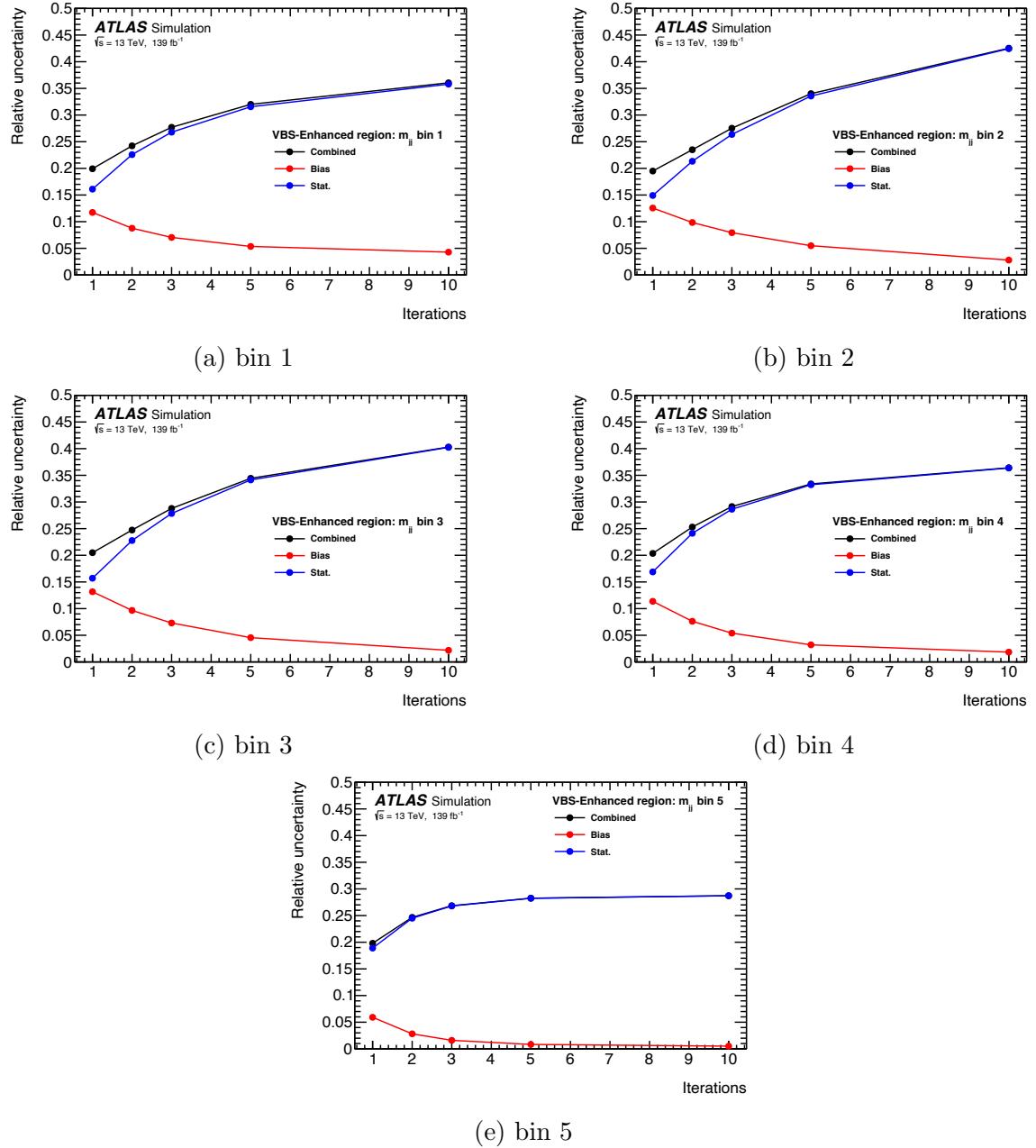


Figure 69: MC-toy-based unfolding bias and statistical uncertainty as a function of several numbers of iterations in each bin of  $m_{jj}$  distribution in the VBS-Enhanced region.

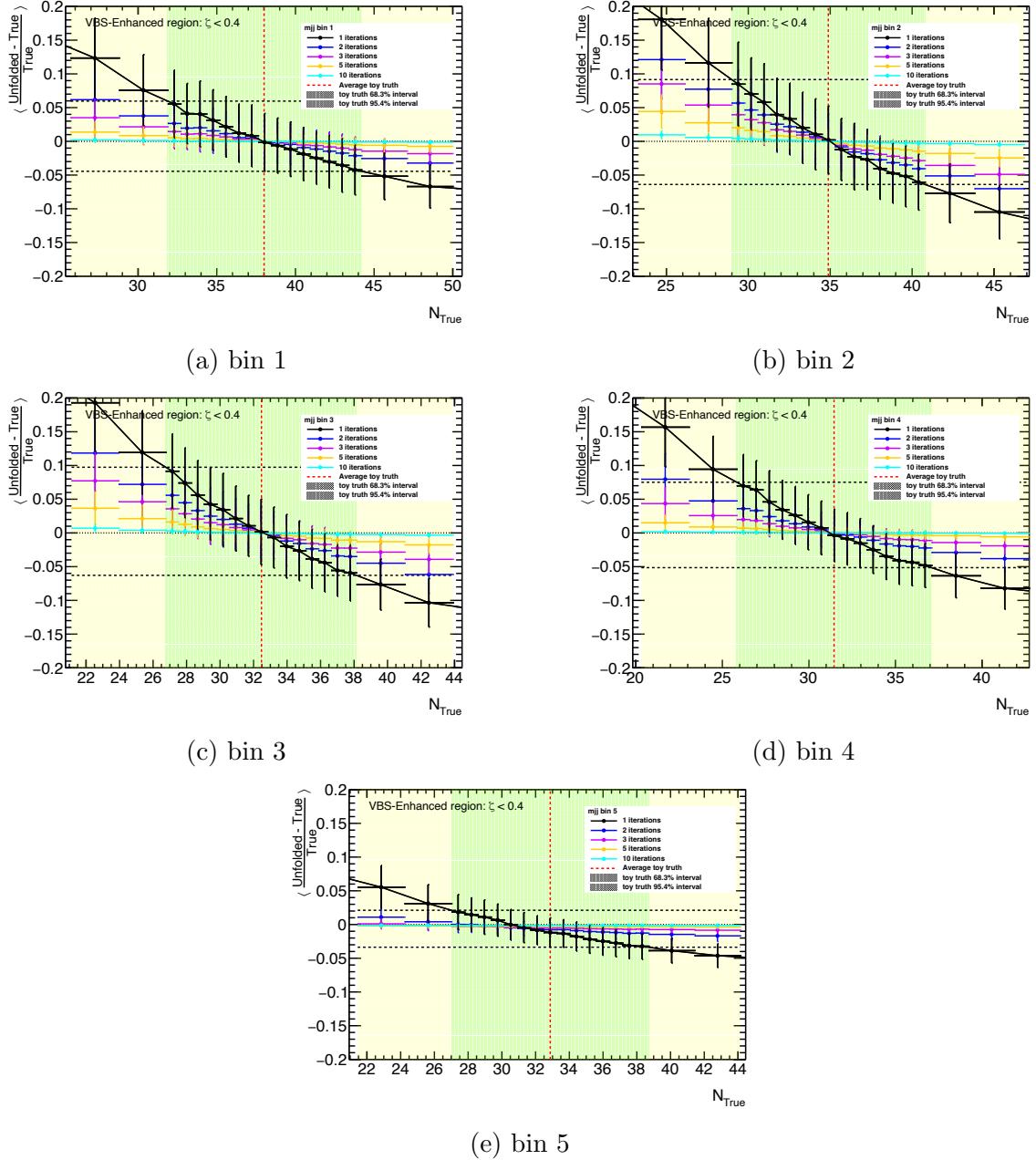


Figure 70: MC-toy-based unfolding bias in each bin of  $m_{jj}$  in the VBS-Enhanced region using Gaussian toys after subtracting the contribution of the fiducial fake events from both nominal and toy MC predictions.

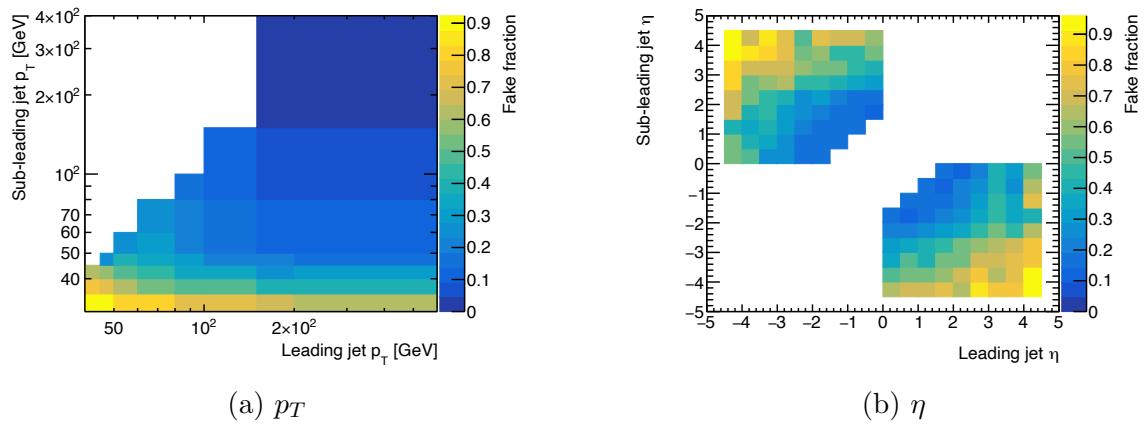


Figure 71: Fraction of fake fiducial events as a function of  $p_T$  and  $\eta$  of the leading (x-axis) and the sub-leading jets (y-axis) in the VBS-Enhanced region.

## C VBS-Suppressed Region

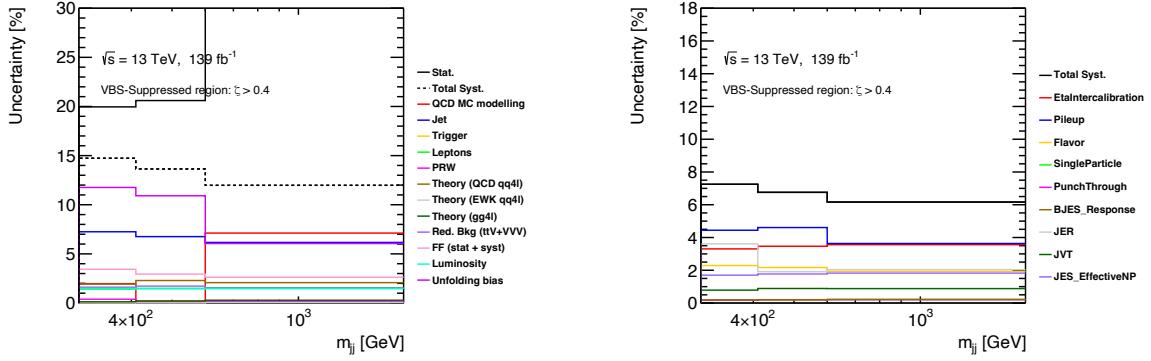
This section summarizes the results of the detector-level yield and the unfolded differential cross-sections measured in the VBS-Suppressed region. The systematics affecting these results are also discussed briefly.

### C.1 Systematics

The same systematic uncertainties discussed in Chapter *V* also impact the measurements in the VBS-Suppressed region. Table 20 shows the impact of several systematic uncertainties on the unfolded differential cross-sections in each bin of  $m_{jj}$  for the VBS-Suppressed region. Like the VBS-Enhanced region, unfolding bias followed by the jet-related uncertainties and QCD MC modeling uncertainty are the most significant sources of systematic uncertainties. Figure 72a shows the impact of the statistical and different systematic uncertainties on the unfolded cross-sections as a function of  $m_{jj}$ . The unfolded cross-sections are statistically limited. Figure 72b shows the breakdown of the jet-related uncertainties, where uncertainties related to the punch through calibration step are dominant.

Table 20: Breakdown of the relative systematic uncertainties (%) for each bin of  $m_{jj}$  in the VBS-Suppressed region.

Bin $m_{jj}$ [GeV]	[300, 410)	[410, 600)	[600, 1780)
QCD MC modelling	1.97	0.16	<b>7.12</b>
Jet	<b>7.26</b>	<b>6.75</b>	<b>6.16</b>
Trigger	0.03	0.06	0.08
Leptons	1.42	1.49	1.56
PRW	0.39	0.06	0.16
Theory ( $qqZZ$ )	1.93	2.29	2.07
Theory (EWK $qqZZjj$ )	0.02	0.02	0.04
Theory ( $ggZZ$ )	0.13	0.22	0.28
MC Bkg. ( $t\bar{t}V+VVV$ )	1.61	1.72	1.52
Fake Bkg. (stat + syst)	3.43	2.95	2.62
Luminosity	1.48	1.36	1.45
Unfolding Bias	<b>11.76</b>	<b>10.95</b>	<b>6.06</b>
Total	14.74	13.64	11.98



(a) Statistical and systematic uncertainties.

(b) Breakdown of jet-related uncertainties.

Figure 72: Uncertainties as a function of  $m_{jj}$  in the VBS-Suppressed region.

## C.2 Detector Level Measurements

Figures 73 and 74 show the measured data yields compared to the SM detector-level predictions as a function of the eleven kinematic observables in the VBS-Suppressed region. Distributions are statistically limited, and the impact of theoretical and experimental systematic uncertainties shown in black bands ranges from about 20 to 30% depending on bins and distributions. Similarly to the VBS-Enhanced region, a  $\chi^2/NDF$  for each distribution is also reported. These values suggest a good agreement between the measured data and the SM prediction.

## C.3 Unfolded Cross-sections

Figures 75 and 76 show the measured unfolded differential cross-sections as a function of the eleven kinematic observables in the VBS-Suppressed region. The unfolded differential cross-sections (black) are compared with two different state-of-the-art SM particle-level predictions, first, where the QCD  $q\bar{q}ZZ$  is generated by SHERPA (red) and second, where the QCD  $q\bar{q}ZZ$  is generated by MADGRAPH (blue). The vertical error bars show the statistical uncertainties of the unfolded cross-sections, and the black bands represent the total systematic uncertainties. Similarly, the dashed-red and dashed-blue bands represent the fiducial-level uncertainties on the SHERPA and MADGRAPH particle-level predictions, respectively.

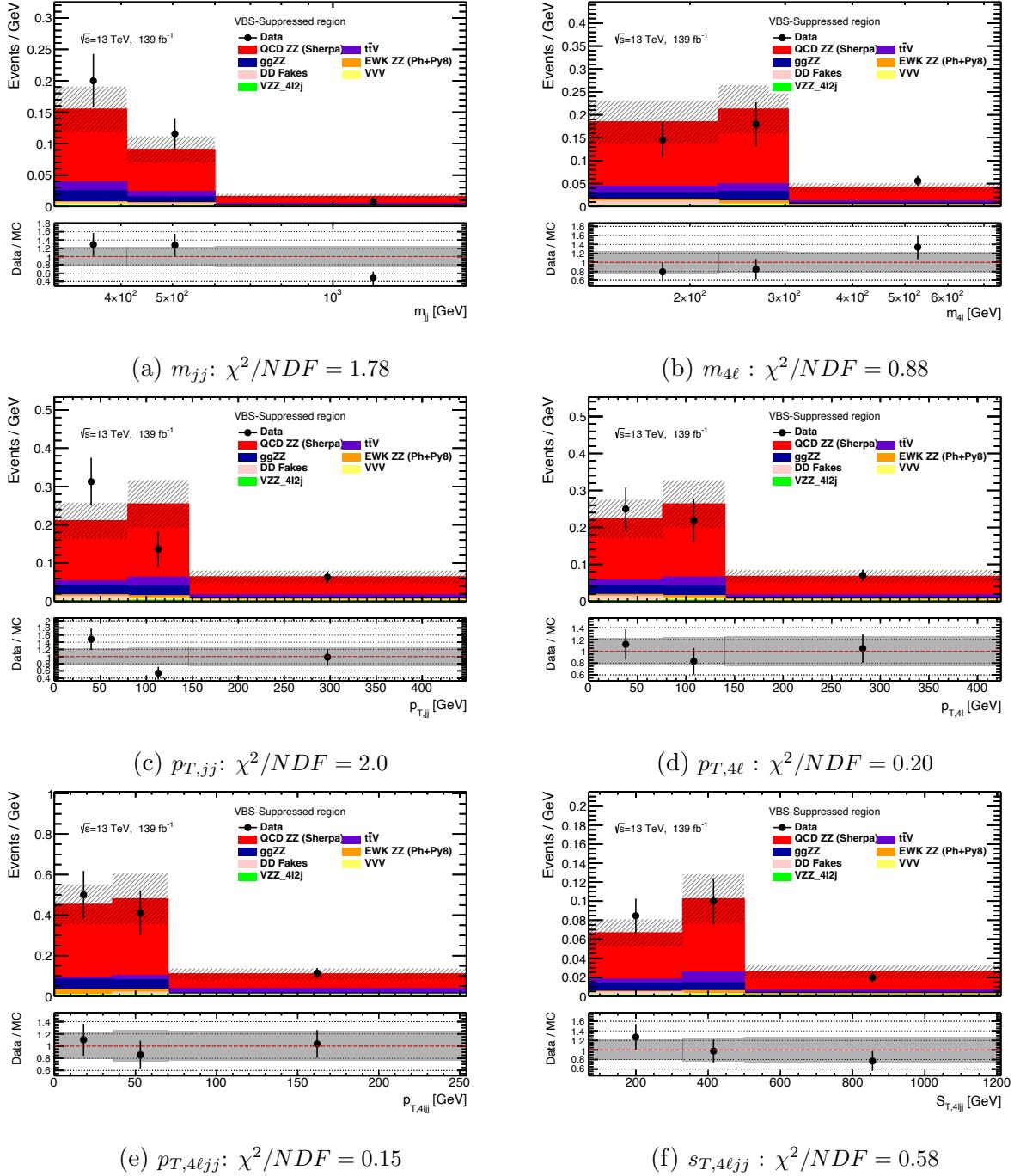


Figure 73: Detector-level distributions in the VBS-Suppressed region.

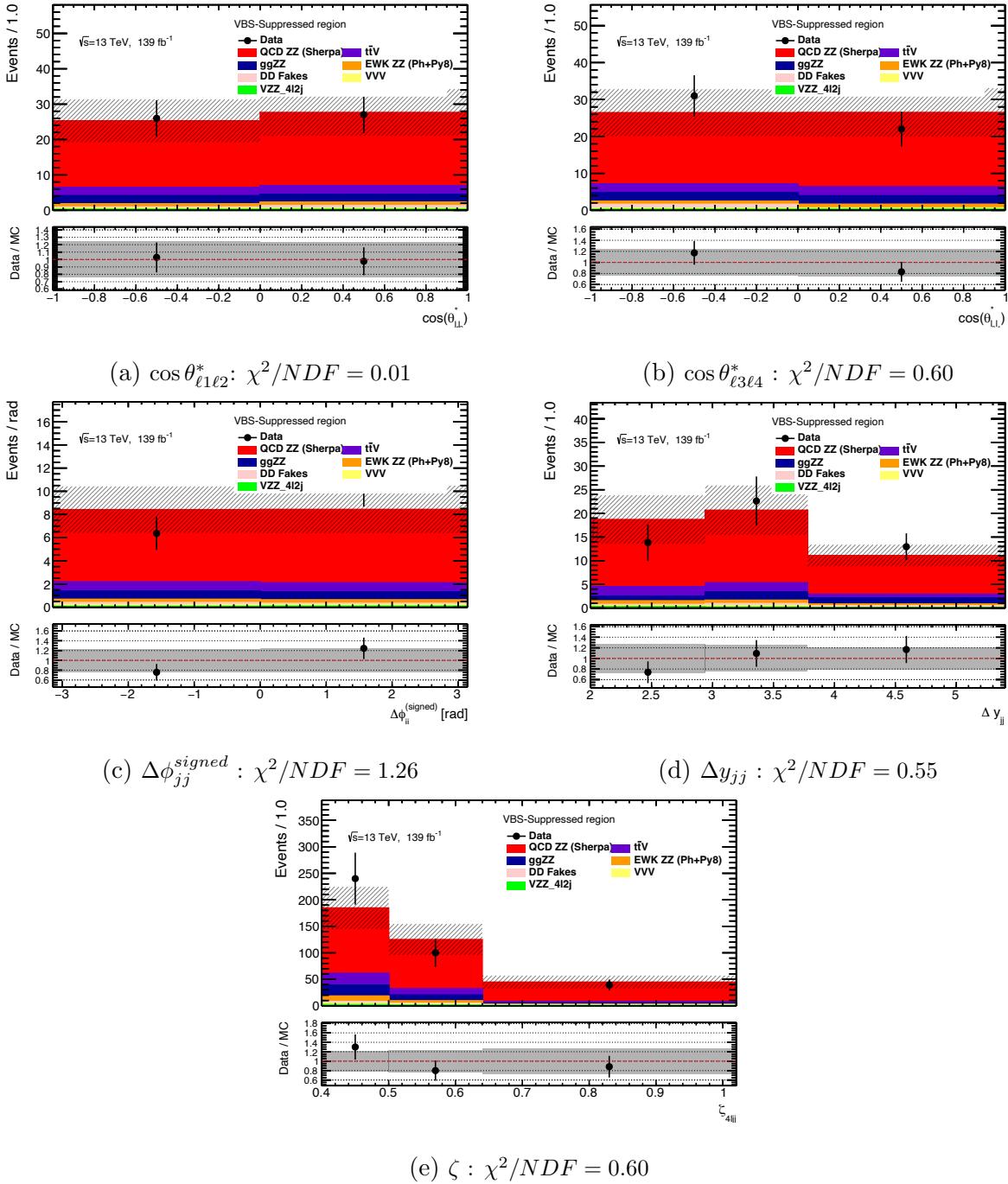


Figure 74: Detector-level distributions in the VBS-Suppressed region.

Similarly to the VBS-Enhanced region, the p-values comparing the unfolded cross-sections with the two SM particle-level predicted cross-sections are estimated. The reported p-values of more than 0.05 and the ratios of particle-level yields to the unfolded-data yields suggest no significant statistical deviations of the measured differential cross-sections from the SM predicted values. Therefore, in the analyzed LHC Run-2 dataset, for the  $ZZ^*(\rightarrow 4)\ell jj$  process, all differential cross-sections in the VBS-Suppressed region are concluded to agree with the SM predictions.

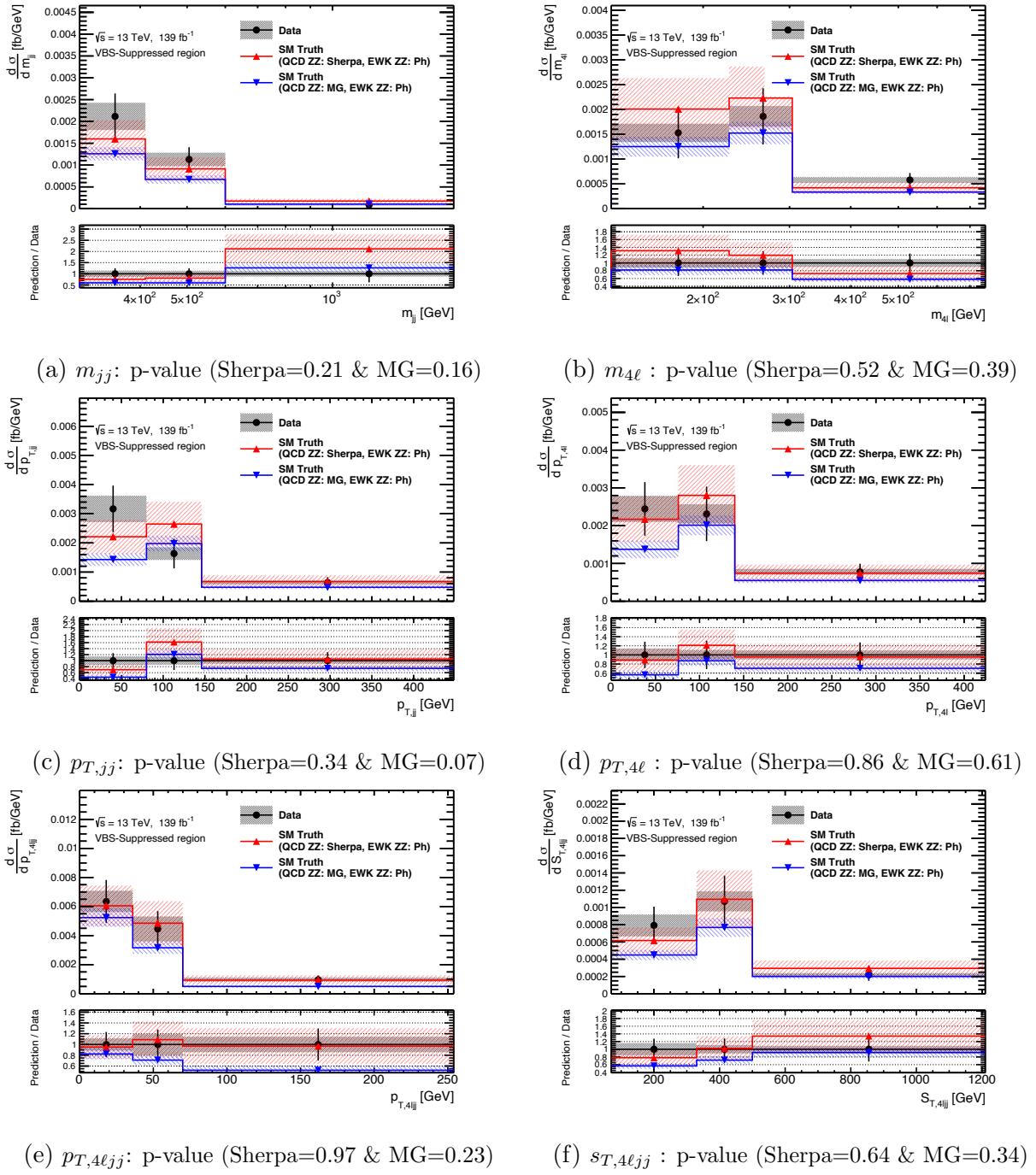


Figure 75: Unfolded differential cross-sections in the VBS-Suppressed region.

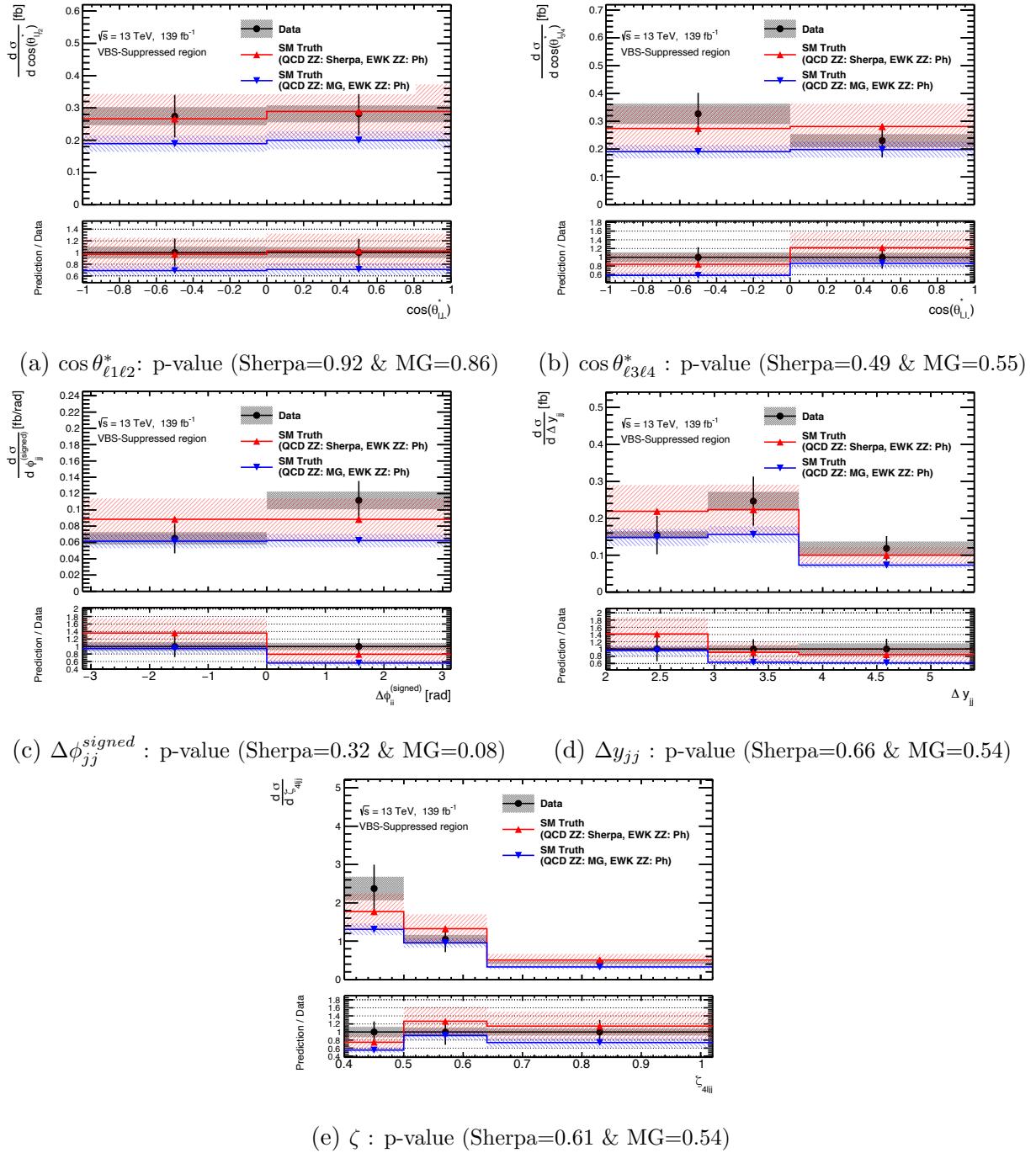


Figure 76: Unfolded differential cross-sections in the VBS-Suppressed region.