



School of Physics and Astronomy בית הספר לפיזיקה ולאסטרונומיה
The Raymond and Beverly Sackler הפקולטה למדעים מדויקים
Faculty of Exact Sciences ע"ש ריימונד ובברלי סאקלר
Tel Aviv University אוניברסיטת תל אביב

Research Thesis for a Degree "Doctor of Philosophy"

Search for the Decay of the Higgs Boson to a Pair of Charm Quarks

Submitted to the Senate of Tel Aviv University by

Guy Koren

The work was carried out under the supervision of

Prof. Erez Etzion

July, 2022

In loving memory of my grandfather

Itzhak Kopilovich

1929-2020

*a man of extraordinary technical talent and an endless thirst for knowledge.
I can not think of a single person who would have appreciated this work as much as he.*

Acknowledgments

In the recent past years, I had the privilege of working in a field I love and taking part in one of humankind's most extraordinary science experiments. The completion of this thesis marks the end of this adventure, and I would like to take the time to thank and acknowledge a number of dear individuals whose guidance and support made it all possible.

First and foremost, I would like to thank my supervisor, *Prof. Erez Etzion*. Erez is the one who first introduced me to particle physics, and it is his passion and love of the field that drew me to pursue this path. Then, upon taking me as his student, Erez walked the fine line between being an adviser and a friend and provided me with much support and advice, both professionally and personally. Thank you for making this journey possible, I will cherish these years and always look back at them with a big smile on my face. I would also like to thank *Dr. Liron Barak* for the many hours of discussion and consultation, and for being a role model of creativity, dedication, and hard work. To *Meny Raviv-Moshe*, the heart, soul, and working hands of our lab, for the vast technical knowledge you provided me with, for the deep conversations on family and life, and for constantly making me laugh. To *Hadar Cohen*, who was much more than an "office neighbor" during this period, and came through as a true friend, and a face I look forward to seeing every day. I would also like to thank my colleagues at Tel Aviv University: *Yan Benhammou*, *Gideon Bella*, *Abner Soffer*, and *Adi Ashkenazi*, each for their special contribution and role they had in my journey.

To my colleagues at CERN and in the VH(cc) analysis team: *Jonathan Shlomi*, *Eilam Gross*, *Tristan Du Pree*, *Marko Stamenkovic*, *Maria Mironova*, *Antonio Costa*, *Tom Neep*, *Zijun Xu*, *Tao Wang*, and particularly to *Andy Chisholm* and *Hanna Arnold*, who were my mentors within the analysis team, and made sure to always make themselves available to further explain, discuss, and listen, and to do so with much patience and generosity. You two have taught me a great deal.

To my loving family, and especially to my parents, *Mira* and *Zvi Koren*. I thank you from the bottom of my heart for the unconditional love and infinite support, for the guidance you provide me, and for the feeling that you always have my back no matter what. There are truly no words that can express my gratitude to you, or describe the special role you have in my life and in the person I grew up to be. I love you two very very much.

Last, but very far from least, I want to thank my daughter *Noga*, that at only 1 year of age, has already taught me so much, and had become a beacon of light in my life. And to the love of my life, *Hagar*. Thank you. Thank you for constantly pushing

me forward and encouraging me, for your wise advice, and for the fresh point of view you provide. Thank you for being a pillar of strength, and for walking by my side and looking after me, not only throughout this academic journey, but day in and day out through every aspect of our lives. I love you very very much, and could not have done this without you.

Abstract

This thesis presents a search for the decay of the Higgs boson to a pair of charm and anti-charm quarks, performed using 139 fb^{-1} of data collected from proton-proton collisions at center-of-mass energy of 13 TeV, recorded by the ATLAS detector at CERN during Run-2 of the LHC between 2015-2018. The search targets events in which the Higgs boson is produced in association with an electroweak gauge boson, where the gauge boson subsequently decays leptonically. By searching in final states containing leptons, which have relatively clean signatures, the analysis is able to suppress the contribution from the dominant QCD background at the LHC. To identify jets originating from charm quarks, the analysis makes use of multivariate *charm-tagging* algorithms, which rely on the unique characteristics of c - (and b -) hadron decays to distinguish between c -jets and b -jets or jets of lighter flavor. The search is performed in a cut-based approach where a fit to the invariant mass of the two jets with highest transverse momentum is used to extract the signal strength $\mu_{VH(cc)}$. This corresponds to the ratio of the measured cross-section times branching fraction of the process, over the Standard Model prediction. No evidence for the process is observed, and an observed (expected) 95% confidence level upper limit of $26 \left(31^{+12}_{-8} \right)$ is set, which corresponds to an observed (expected) constraint on the charm Yukawa coupling modifier of $|\kappa_c| < 8.5(12.4)$ at 95% confidence level. This result is the first direct constraint on κ_c .

Contents

Introduction	1
I Theory	8
1 The Standard Model of Particle Physics	9
1.1 Fundamental Concepts: Fields, Lagrangians, and Symmetries	9
1.2 Particles and Forces	11
1.3 Quantum-Chromo-Dynamics (QCD)	15
1.4 Electroweak Theory (EW)	17
1.4.1 Charged Electroweak Interaction	19
1.4.2 Neutral Electroweak Interaction	20
1.4.3 Electroweak gauge bosons-self Interactions	23
1.5 The Higgs Mechanism	24
1.5.1 Spontaneous Symmetry Breaking	24
1.5.2 Vector Boson Masses	26
1.5.3 Fermion Masses	28
1.5.4 The SM Higgs Boson	28
1.6 Limitations of the SM	29
2 LHC Phenomenology	33
2.1 Collider Physics	33
2.1.1 Proton-Proton Collisions	33
2.1.2 Collider Kinematics	35
2.1.3 Luminosity	37

2.2	Higgs Physics at the LHC	37
2.2.1	Main Higgs Boson Production Mechanisms	38
2.2.2	Higgs Boson Decay	40
2.2.3	Higgs Coupling Modifiers - The κ Framework	42
II	Experiment	44
3	The Large Hadron Collider and the ATLAS Experiment	45
3.1	Large-Hadron-Collider (LHC)	45
3.1.1	Layout and Operation	45
3.1.2	Run-2 Performance	47
3.2	The ATLAS Experiment	49
3.2.1	Coordinate System	50
3.2.2	Inner Detector	52
3.2.3	Calorimeters	54
3.2.4	Muon Spectrometer	59
3.2.5	The Trigger and Data Acquisition System	61
4	Reconstruction of Physics Objects	63
4.1	Tracks and Primary Vertices	63
4.2	Electrons and Photons	66
4.2.1	Reconstruction	66
4.2.2	Identification	67
4.2.3	Isolation	68
4.2.4	Calibration	69
4.3	Muons	69
4.3.1	Reconstruction	69
4.3.2	Identification	70
4.3.3	Isolation	71
4.3.4	Calibration	71
4.4	Jets	72
4.4.1	Jet Formation	72
4.4.2	Jet Reconstruction	73
4.4.3	Jet Calibration	76
4.4.4	Pileup Jets Suppression	80
4.5	Hadronic Taos	81
4.6	Missing Transverse Momentum	82

5	Flavor Tagging	84
5.1	Flavor-Tagging Algorithms	85
5.1.1	Low-Level Algorithms	86
5.1.2	High-Level Algorithms	88
5.2	Flavor-Tagging Performance	91
5.3	Flavor-Tagging Working Points Calibration	93
III	VH(cc)	95
6	The VH(cc) Analysis - Introduction	96
6.1	Motivation	96
6.2	Overview of the Analysis	97
7	Data and MC Samples	100
7.1	Data Sample	100
7.2	Simulation Samples	100
7.2.1	Modeling of the VH Signal Process	102
7.2.2	Modeling of the Background Processes	102
7.2.3	Corrections to Simulated Events	104
8	Selection and Categorization	105
8.1	Trigger Selection	105
8.2	Object Selection	106
8.2.1	Overlap Removal	107
8.3	Event Selection and Categorization	108
8.3.1	0-lepton Channel Specific Selection	110
8.3.2	1-lepton Channel Specific Selection	110
8.3.3	2-lepton Channel Specific Selection	111
8.4	Signal Regions and Control Regions	112
8.4.1	0 c -tag Control Region	112
8.4.2	Top Control Region	112
8.4.3	High ΔR Control Region	113
8.5	Estimation of the Multijet Background	113
8.5.1	Multijet Background in the 0-lepton Channel	113
8.5.2	Multijet Background in the 1-lepton Channel	119
8.5.3	Multijet Background in the 2-lepton Channel	121

9	Flavor Tagging in $VH(cc)$	124
9.1	Flavor Tagging Strategy in $VH(cc)$	124
9.2	Working Point Optimization	126
9.3	Calibration of the c -tagging Efficiency of c -jets	128
9.4	Truth-tagging	131
9.4.1	ΔR Correction for V +jets	132
9.4.2	Residual non-closure correction	135
10	Statistical Analysis	138
10.1	Definition of the Likelihood Function	138
10.2	Test Statistic and Measurement of the Signal Strength	140
10.3	Exclusion and CL_s limits	141
10.4	Fit Model	142
11	Systematic Uncertainties	145
11.1	Nuisance Parameters	145
11.2	Experimental Systematic Uncertainties	147
11.2.1	Luminosity and Pileup	147
11.2.2	Leptons	147
11.2.3	Jets	148
11.2.4	Missing Transverse Momentum	150
11.2.5	Flavor Tagging	151
11.3	Modeling Systematic Uncertainties	152
11.3.1	$VH(cc)$ signal	154
11.3.2	$VH(bb)$	154
11.3.3	V +jets	154
11.3.4	Top background	161
11.3.5	Diboson	166
11.3.6	Multijet	168
12	Results	172
12.1	Fit performance	172
12.1.1	Post-fit $m_{c\bar{c}}$ distributions	173
12.1.2	Nuisance parameters pulls and correlations	185
12.2	Signal strength measurement and exclusion limits	189
12.2.1	Nuisance parameter ranking	193
12.3	Limits on the c -quark Yukawa coupling y_c	195

<i>Contents</i>	x
13 Concluding Remarks	200
Concluding Remarks	200
13.1 Future prospects	201
Appendix A Acceptance ratios tables	204
Bibliography	209

List of Figures

1.1	The particle content of the Standard Model.	14
1.2	Feynman diagrams for the QCD gauge interaction terms.	16
1.3	Feynman diagrams for the charged weak interaction terms.	20
1.4	Feynman diagrams for the neutral electroweak interaction terms.	23
1.5	Feynman diagrams for the electroweak bosons self-interaction terms. . . .	24
1.6	The one-dimensional potential $V(\phi) = -\frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4$	25
1.7	Feynman diagrams for the interaction between the $W^\pm$ and Z bosons and the Higgs boson.	27
1.8	Feynman diagram for the interaction between the SM fermions and the Higgs boson.	28
1.9	Feynman diagrams for the Higgs boson triple and quartic self-interaction. .	29
1.10	Quantum loop corrections to the Higgs boson self energy.	31
2.1	Examples of Parton Distribution Functions at two energy scales.	34
2.2	Feynman diagrams for the main Higgs boson production modes at the LHC. .	38
2.3	Cross-section as a function of pp center-of-mass energy for the main Higgs production modes at the LHC.	39
2.4	Feynman diagrams for the $gg \rightarrow ZH$ contribution to the total VH produc- tion mode.	40
2.5	Feynman diagrams for the decay of the Higgs to $\gamma\gamma$ and gg	41
2.6	Measured Higgs coupling strength modifiers as a function of mass.	43
3.1	The CERN accelerator complex.	46
3.2	Projected plan for the operation of the LHC and HL-LHC.	48
3.3	Integrated luminosity as a function of time during Run-2 of the LHC. . . .	49
3.4	Average number interactions per bunch crossing during Run-2 of the LHC. .	50

3.5	A schematic view of the ATLAS detector.	51
3.6	Transversal view of particle paths inside the ATLAS detector.	51
3.7	Schematic illustrations of the ID and it's sub-components.	52
3.8	A cut-away view of the ATLAS calorimeter system.	55
3.9	Examples for the leading order Feynman diagrams of the two main processes taking place in an EM shower.	56
3.10	Schematic view of the ATLAS Muon Spectrometer.	60
4.1	Illustration of the superclustering algorithm for electrons and photons. . .	67
4.2	Illustration of the topo-clustering algorithm restricted to 2-D.	74
4.3	A simulated parton-level event clustered with four different jet algorithms. .	76
4.4	Calibration stages for EM-scale jets.	77
4.5	The average jet energy response and the signed difference between the truth and reconstructed jet η	79
4.6	The two-dimensional JVT likelihood and the distribution of JVT for pileup and hard-scatter jets.	81
5.1	Schematic diagram of tracks with positive and negative impact parameters and the distribution of d_0/σ_{d_0} for different flavors of jets.	87
5.2	Distributions of the final MV2c10 and DL1 output scores.	92
5.3	The light- and c -jet rejection versus the b -tagging efficiency for low- and high-level flavor-tagging algorithms.	93
6.1	Leading order Feynman diagrams for the 0-, 1-, and 2-lepton channels in the VH(cc) analysis.	98
7.1	Leading order Feynman diagrams for the single-top process.	103
8.1	Postfit distributions of $\min \left(\Delta\phi(E_T^{\vec{miss}, jets}) \right)$ for the data and electroweak MC backgrounds.	118
8.2	Comparison of the multijet, Z+jets, and W+jets $m_{c\bar{c}}$ shape in the 0-lepton channel.	119
8.3	Post-fit m_T^W distribution in the 1 c -tag signal regions used to extract the multijet normalization.	122
8.4	Post-fit m_T^W distribution in the 2 c -tag signal regions used to extract the multijet normalization.	123
9.1	VH(cc) analysis categorization in events with 3 jets.	125
9.2	Summary of the VH(cc) analysis strategy.	126

9.3	Working point optimization for the simultaneous c -tagging and b -tagging veto.	127
9.4	Calibration of the $VH(cc)$ c -tagging efficiency of c -jets.	130
9.5	Truth- and direct-tagging $m_{c\bar{c}}$ distributions for $t\bar{t}$ and $Z+cc$ events.	133
9.6	Truth- and direct-tagging comparisons as a function of the minimum ΔR between a c -tagged b -jet and its closest jet. The solid black curve on the bottom pad shows the fourth-polynomial function fitted to the ratio between TT and DT. The total number of DT (Yield DT) and TT (Yield TT) events used in each plot is also specified.	134
9.7	Truth- and direct-tagging comparisons as a function of the minimum ΔR between a c -tagged c -jet and its closest jet. The solid black curve on the bottom pad shows the fourth-polynomial function fitted to the ratio between TT and DT. The total number of DT (Yield DT) and TT (Yield TT) events used in each plot is also specified.	134
9.8	Truth- and direct-tagging comparisons as a function of the minimum ΔR between a c -tagged light-jet and its closest jet. The solid black curve on the bottom pad shows the fourth-polynomial function fitted to the ratio between TT and DT. The total number of DT (Yield DT) and TT (Yield TT) events used in each plot is also specified.	134
9.9	Truth- and direct-tagging comparisons before and after applying the minimum ΔR correction.	135
9.10	Correction for the residual TT vs DT non-closure.	137
11.1	p_T response distribution for PYTHIA (blue) and HERWIG.	150
11.2	The p_T response variation for b -, c -, and light-jets.	150
11.3	Feynman diagrams for the production of a vector boson in association with jets.	157
11.4	Normalized $m_{c\bar{c}}$ distributions of the individual V +jets flavor components.	158
11.5	Normalized $m_{c\bar{c}}$ distributions of the $V+hf$, $V+mf$, and $V+l$ categories.	159
11.6	Ratio of alternative over nominal normalized $m_{c\bar{c}}$ shapes for the V +jets sub-processes in high- p_T^V high ΔR CR.	162
11.7	Ratio of alternative over nominal normalized $m_{c\bar{c}}$ shapes after the latter had been reweighted by Systematic 1.	163
11.8	$m_{c\bar{c}}$ distribution of the $t\bar{t}$ process flavor components in the 1-lepton, 1 c -tag, 3-jet signal region.	164
11.9	Fits to the $\Delta R(j_1, j_2)$ distribution in the 1-lepton signal region.	170
11.10	Multijet $m_{c\bar{c}}$ shape uncertainty in the analysis signal regions.	171

12.1	Post-fit $m_{c\bar{c}}$ distributions in the 0-lepton signal region. (a) show the 1 c -tag, 2-jet SR, (b) show the 1 c -tag, 3-jet SR, (c) show the 2 c -tag, 2-jet SR, and (d) show the 2 c -tag, 3-jet SR.	174
12.2	Post-fit $m_{c\bar{c}}$ distributions in the 0-lepton high ΔR control region. (a) show the 1 c -tag, 2-jet category, (b) show the 1 c -tag, 3-jet category, (c) show the 2 c -tag, 2-jet category, and (d) show the 2 c -tag, 3-jet category.	175
12.3	Post-fit $m_{c\bar{c}}$ distributions in the 1-lepton signal region. (a) show the 1 c -tag, 2-jet SR, (b) show the 1 c -tag, 3-jet SR, (c) show the 2 c -tag, 2-jet SR, and (d) show the 2 c -tag, 3-jet SR.	176
12.4	Post-fit $m_{c\bar{c}}$ distributions in the 1-lepton high ΔR control region. (a) show the 1 c -tag, 2-jet category, (b) show the 1 c -tag, 3-jet category, (c) show the 2 c -tag, 2-jet category, and (d) show the 2 c -tag, 3-jet category.	177
12.5	Post-fit $m_{c\bar{c}}$ distribution in the 1-lepton 0 c -tag CR. (a) show the 2-jet category and (b) show the 3-jet category.	178
12.6	Post-fit $m_{c\bar{c}}$ distributions in the 0-lepton (a) and 1-lepton (b) top CR.	178
12.7	Post-fit $m_{c\bar{c}}$ distributions in the 2-lepton high- p_T^V signal region. (a) show the 1 c -tag, 2-jet SR, (b) show the 1 c -tag, 3+ jet SR, (c) show the 2 c -tag, 2-jet SR, and (d) show the 2 c -tag, 3+ jet SR.	179
12.8	Post-fit $m_{c\bar{c}}$ distributions in the 2-lepton low- p_T^V signal region. (a) show the 1 c -tag, 2-jet SR, (b) show the 1 c -tag, 3+ jet SR, (c) show the 2 c -tag, 2-jet SR, and (d) show the 2 c -tag, 3+ jet SR.	180
12.9	Post-fit $m_{c\bar{c}}$ distributions in the 2-lepton high- p_T^V high ΔR control region. (a) show the 1 c -tag, 2-jet category, (b) show the 1 c -tag, 3+ jet category, (c) show the 2 c -tag, 2-jet category, and (d) show the 2 c -tag, 3+ jet category.	181
12.10	Post-fit $m_{c\bar{c}}$ distributions in the 2-lepton low- p_T^V high ΔR control region. (a) show the 1 c -tag, 2-jet category, (b) show the 1 c -tag, 3+ jet category, (c) show the 2 c -tag, 2-jet category, and (d) show the 2 c -tag, 3+ jet category.	182
12.11	Post-fit $m_{c\bar{c}}$ distributions in the 2-lepton 0 c -tag control region. (a) show the 2-jet, high- p_T^V category, (b) show the 3+ jet, high- p_T^V category, (c) show the 2-jet, low- p_T^V category, and (d) show the 3+ jet, low- p_T^V category.	183
12.12	Post-fit $m_{c\bar{c}}$ distributions in the 2-lepton top $e\mu$ control region. (a) show the 2-jet, high- p_T^V category, (b) show the 3+ jet, high- p_T^V category, (c) show the 2-jet, low- p_T^V category, and (d) show the 3+ jet, low- p_T^V category.	184
12.13	Monte Carlo modeling-related nuisance parameter pulls and constraints.	186
12.14	Detector-related nuisance parameter pulls and constraints.	187
12.15	Correlation matrix from the unconditional combined fit to Data.	190

12.16	Fitted values for the $VH(cc)$, $VW(cl)$ and $VZ(cc)$ signal strengths in the fit to the full dataset.	191
12.17	Observed and expected limits for the $VHcc$ signal at 95% CL.	192
12.18	Background subtracted mass spectrum from the fit to data in the combined analysis signal region.	193
12.19	Impact of systematic uncertainties for the 10 largest contributions on the fitted $VH(cc)$ signal-strength parameter.	194
12.20	The κ_V and κ_c coupling modifiers in the $VH(cc)$ Feynman diagram.	197
12.21	Parametrization of the signal strength as a function of the charm quark Yukawa coupling modifier $\mu(\kappa_c)$	198
12.22	Expected and observed profile likelihood scans for κ_c	199

List of Tables

1.1	The four fundamental forces.	12
1.2	Electroweak charges for the lepton and quark fields.	22
2.1	Higgs boson decay branching ratios.	41
5.1	Input variables used by the high-level flavor-tagging algorithms.	90
7.1	Signal and background processes and their corresponding nominal and alternative MC generators.	101
7.2	Heavy flavor filters used in the V +jets MC samples.	102
8.1	Summary of the signal region event selection.	111
8.2	Summary of the analysis signal and control regions	114
8.3	Pre-fit event yields in the 0-lepton channel. For each region, the entries highlighted in color represent process with large relative event yields. . . .	115
8.4	Pre-fit event yields in the 1-lepton channel. For each region, the entries highlighted in color represent process with large relative event yields. . . .	115
8.5	Pre-fit event yields in the 2-lepton signal regions.	116
8.6	Pre-fit event yields in the 2-lepton high ΔR control regions.	117
8.7	Pre-fit event yields in the 2-lepton 0 c -tag and Top CRs.	118
8.8	Summary of differences between the isolated (SR) and inverted isolation regions (MJ-CR).	120
8.9	Fraction of QCD (multijet) backgrounds in the signal region.	121
9.1	Efficiencies of the c -tagging (including the b -tag veto) working point. . . .	127
10.1	Summary of the fit range and binning.	143

11.1	Fractions of the individual Z+jets and W+jets components in the signal regions.	156
11.2	V+jets normalizations.	160
11.3	Normalizations of the top backgrounds.	165
11.4	Summary of the shape uncertainties considered for the top backgrounds. .	166
11.5	The relative contribution of the different diboson components in the three analysis channels.	167
12.1	Values of the free-floating background normalization parameters obtained from the likelihood fit to data.	185
12.2	Measured signal strength of the signal processes in the unconditional fit to data.	191
12.3	The observed and expected 95% CL upper limit on the $VH(cc)$ signal strength.	192
12.4	Breakdown of contributions to the uncertainty in the fitted values of $\mu_{VH(cc)}$, $\mu_{VW(cl)}$ and $\mu_{VZ(cc)}$	196
12.5	Expected and observed 68% CL and 95% CL limits on κ_c	198
A.1	Acceptance ratio uncertainties between the 3(+)-jet and 2(+)-jet categories.	205
A.2	Acceptance ratio uncertainties between individual components in backgrounds which group together multiple processes/flavors.	206
A.3	Acceptance ratio uncertainties between the low- and high- p_T^V regions. . .	206
A.4	Acceptance ratio uncertainties between control and signal regions.	207
A.5	Acceptance ratio uncertainties between lepton channels.	208

Introduction

*"Stands at the sea,
wonders at wondering: I
a universe of atoms
an atom in the universe"*

—Richard Feynman

The most amazing fact about our universe, at least in my mind, is that everything is made from the same fundamental building blocks. These building blocks, which are a small set of elementary particles, can (in principle) describe all phenomena of nature, from the largest galaxy clusters to a single grain of sand. Understanding the nature of these particles and the ways in which they interact with each other is at the heart of particle physics. Today, the description of elementary particles is given within the theory known as the *Standard Model of Particle Physics* (SM). This theory, established during the mid-1960's and early-1970's, has been subject to rigorous testing and its predictions were able to account for almost all experimental results. Yet despite its success, the SM is known to be incomplete; first, Gravitational interactions are not accounted for by the theory, and one must resort to General Relativity in order to describe them. Moreover, observations such as the existence of Dark Matter, matter and anti-matter asymmetry, and neutrino oscillations are examples of phenomena that can not be explained within the framework of the SM. Attempts to resolve these discrepancies are made by developing theories that extend the Standard Model, and seek their evidence in major experimental operations. At the time of writing these lines, no evidence of physics beyond the Standard Model (BSM) has been found yet.

The European Organization for Nuclear Research, CERN¹, located outside of Geneva, Switzerland, is the world's largest laboratory for research of fundamental particle physics. Since its establishment in 1954, scientists at CERN have been operating physics experiments based on particle accelerators and detectors, with the goal of witnessing evidence of new physics processes, and measuring the current theories to higher precision. The state-of-the-art generation of experimental machinery includes the Large-Hadron-Collider (LHC) [1], a ring-shaped particle accelerator used to conduct collisions between protons at higher energies than ever before. The LHC started operating in 2010, and in July of 2012 the discovery of the Higgs boson, considered to be the last missing part in the SM, was announced by the ATLAS [2] and CMS [3] experiments [4, 5].

After the discovery of the Higgs boson, attention has turned to more detailed measurements of its properties as a means of testing the predictions of the Standard Model. One such class of predictions is the couplings of the Higgs to other particles, which dictates the rate at which it will decay to the lighter SM fermions. At the time of writing this work, direct evidence for the Higgs boson couplings to the Z and W bosons, τ lepton, and the Yukawa coupling to third generation quarks have already been obtained [6–10]. More recently, evidence for the decay of the Higgs to a pair of muons has been reported [11, 12], yet not at the sufficient significance level required to claim an observation of this process. In generations other than the third, measurements of the Yukawa coupling of the Higgs boson to quarks are difficult at hadron colliders, due to small branching fractions, large backgrounds, and challenges in jet flavor identification.

The work presented in this thesis describes a search for the decay of the Higgs boson to a pair of charm and anti-charm quarks. According to the SM, only 2.9% of the Higgs bosons produced at the LHC will decay to charm quarks. This low branching ratio makes the $H \rightarrow c\bar{c}$ decay mode susceptible to contributions from new physics, as the expected SM rate can be potentially altered by small or moderate BSM coupling.

The search presented here assumes a scenario in which the Higgs is co-produced with an electroweak (EW) gauge boson, either a W or a Z. The leptonic decays of the W and Z bosons lead to relatively clean signatures that can be used to significantly suppress the contributions from background processes and allow for an efficient triggering strategy. Events are selected in three channels, based on the decay of the gauge boson, namely the 0-, 1-, and 2-lepton channels targeting the $ZH \rightarrow \nu\bar{\nu}c\bar{c}$, $WH \rightarrow l\nu c\bar{c}$, $ZH \rightarrow l\bar{l}c\bar{c}$ signatures, respectively. To identify jets originating from the hadronization

¹From French: *Conseil Européen pour la Recherche Nucléaire*.

of charm quarks, the search relies on multivariate charm-tagging algorithms.

A first iteration of this analysis was published in 2018 [13], with 36 fb^{-1} of data and considering only the case where the associated vector is a Z boson decaying into two charged leptons, $Z \rightarrow l\bar{l}$. The search yielded an observed (expected) upper limit on the cross-section time branching ratio of the process of $\sigma(pp \rightarrow ZH) \times BR(H \rightarrow c\bar{c}) < 110 \left(150^{+80}_{-40}\right) \times \text{SM expectation}$.

More recently, a second iteration of the search has been completed [14]. The current version is performed using the full 139 fb^{-1} collected by ATLAS during Run-2 of the LHC, considers all three decay modes of the vector boson, and includes major refinements to various aspects of the analysis strategy, such as the event categorization, flavor-tagging scheme, fit model, as well as general updates and improvements to the software framework used. It is this second iteration of the search which is depicted in detail in this thesis.

Structure of the thesis

This thesis is divided into three parts, structured as follows:

■ Part 1 - *Theory*:

The first part contains a description of the theoretical framework relevant to this work.

- Chapter 1 begins with a short introduction to some of the "prerequisite" fundamental concepts, followed by an overview of the Standard Model covering the particles, interactions, and the mechanism of spontaneous symmetry breaking, from which emerges the Higgs boson.
- Chapter 2 describes the phenomenological aspects relevant to collider physics, and particularly to the LHC, where the colliding particles are protons. It then focuses on the phenomenology of the Higgs boson in the LHC, with an overview of the dominant production modes and favorable decays mode.

■ Part 2 - *Experiment*:

The second part is dedicated to the experimental aspects which are at the heart of any physics analysis performed within the ATLAS collaboration.

- Chapter 3 describes the LHC machine and the ATLAS detector, detailing its different sub-system technologies and means of operation.
- Chapter 4 focuses on the reconstruction of physics objects, describing the techniques and algorithms by which the electronic signals recorded in the detector are interpreted as final state particles studied by the various physics analyses.
- Chapter 5 presents a general overview of the *flavor-tagging* methodology in ATLAS. Flavor-tagging refers to the techniques designed to distinguish between jets initiated by heavy-flavored quarks (b - or c -jets) to jets initiated by light (u , d , s) quarks or gluons. The unique characteristics of heavy-flavored jets are discussed first, followed by a description of the common flavor-tagging algorithms used in ATLAS, the assessment of their performance, and calibration.

■ **Part 3 - $VH(cc)$:**

The third and final part of this thesis describes in detail the various aspects of the $VH(cc)$ analysis, searching for the Higgs decay into charm quarks. Unlike the two previous parts, which are more general and are most likely relevant to all physics analyses in ATLAS, this part outlines the specific work performed by the author as part of the analysis team.

- Chapter 6 first introduces the motivation for searching for this particular process and gives a general overview of the analysis.
- Chapter 7 presents the data sample and MC simulation samples used to model the signal and background processes.
- Chapter 8 discusses the selection of $VH(cc)$ candidate events. This includes the triggers used in the search, the criteria imposed on reconstructed physics objects, and the selection made on kinematic variables at the event level, designed to enhance the number of signal events over background events. Then the categorization of events into signal and control regions is described, and lastly, the estimation of the QCD initiated ("multijet") background is depicted.
- Chapter 9 presents specific details regarding flavor-tagging within the $VH(cc)$ analysis. First, the flavor-tagging strategy is introduced, then the optimization of the c -tagging working point is presented, followed by a detailed description of the method for calibrating the c -tagging efficiency of c -jets. The chapter ends with a description of a technique known as *truth-tagging*, which

is used in the analysis to enhance the statistical power of simulated backgrounds where the majority of events fail to pass the c -tagging criteria.

- Chapter 10 is dedicated to the statistical analysis and fit strategy. The definition of the likelihood function and test statistic are first presented, followed by a discussion on the fit model used in the search.
- Chapter 11 discusses the systematic uncertainties accounted for in the search. It begins with a brief introduction to the notion of nuisance parameters and how they are accounted for in the fit, then continues to present the two classes of systematics considered: experimental uncertainties, related to the measurement and interpretation of physical quantities in the detector, and modeling systematic which accounts for the uncertainty on the predictions of the signal and background processes.
- Chapter 12 presents the final results. The fit performance is first analyzed in terms of the behavior of the nuisance parameters post-fit and the fit model's ability to successfully describe the data. The measured signal strength and corresponding 95% CL upper limit on the cross-section times branching ratio of the signal process are then reported. Lastly, the results of the fit are interpreted within the Higgs coupling modifiers framework (known also as κ framework) and a first direct constraint on the Yukawa charm coupling is reported.

Authors personal contribution

I have been a member of the analysis team from its very beginning and as such have contributed both to the first iteration of the analysis and to the current search which is detailed in this work. My personal contributions and the studies I have conducted are summarized below:

■ First iteration, $Z(l\bar{l})H(c\bar{c})$, 36 fb^{-1} , 2016-2018:

My contribution to the analysis's first iteration was relatively limited, as during the time I was also occupied with my qualification work (which is briefly described below). In addition to the items listed below, I have performed studies characterizing the background composition of events with zero reconstructed charged leptons (i.e. $Z \rightarrow \nu\bar{\nu}$ decay channel). However, as during 2017 ATLAS

migrated to a new software release, there was pressure from the collaboration to finalize analyses still using the old software. As such, the analysis team decided to focus the efforts on the $Z \rightarrow l\bar{l}$ decay channel which was in the most mature state at the time. Therefore, these studies were eventually not directly used in the ATLAS first $VH(cc)$ publication.

- I have studied the jet-energy-scale (JES) and jet-energy-resolution (JER) response of c -jets. Prior to this analysis, the JES response of c -jets had not been thoroughly studied in ATLAS, and in terms of the JES uncertainties c -jets were classified as light-jets and assigned a different uncertainty component than b -jets. As this analysis was the first in ATLAS to target final states with c -jets, it was necessary to examine the JES response of c -jets and decide on the proper treatment accordingly. This study was conducted in collaboration with the ATLAS Jet-Etmiss (missing transverse energy) group, which is responsible for the combined performance of jets and E_T^{miss} in ATLAS.
- I have worked on the estimation of the QCD ("multijet") background arising from jets faking leptons. In the case of $Z \rightarrow l\bar{l}$ events, this background was expected to be negligible, and indeed the studies showed that it amounts to $\sim 1\%$ of the total background and can be neglected in the final fit. Based on this study it was also decided to ignore the multijet background in the current version of the analysis, for events containing two reconstructed charged leptons.

■ **Second iteration, $V(l\bar{e}p)H(c\bar{c})$, 139 fb^{-1} , 2018-2021:**

In the second iteration of the analysis I have been involved in almost every aspect of the research. In particular:

- I was involved in the development and implementation of the new software framework.
- I was responsible for the calibration of the c -tagging efficiency of c -jets, which is a correction applied to c -tagged jets in simulation to account for differences with respect to the efficiency evaluated on data. This work was done in collaboration with the flavor-tagging (FTAG) group in ATLAS.
- I was responsible for the multijet background estimation and its associated uncertainties. This background is considered in the fit only for events with a

single charged lepton (i.e. $W \rightarrow l\nu$ decay channel) and is evaluated using a data-driven method.

- I was involved in the implementation of the *truth-tagging* technique and in the studies to characterize its agreement with the standard *direct-tagging* approach.
- I contributed greatly to modeling of the background processes. This includes the splitting of the dominant V +jets background into the different flavor component groups, and the derivation of the shape and acceptance uncertainties of the V +jets, $t\bar{t}$, and diboson backgrounds.
- I was involved in the design of the fit model, and in various studies regarding the treatment of nuisance parameters, understanding their post-fit pulls and correlations, and the impact of correlating them between different regions and categories.
- I was also a co-author of the analysis internal documentation [15]. This document details all aspects of the analysis, and is intended to be reviewed as part of the analysis approval process within the collaboration. As such, a substantial part of the information presented in this work is based on this document.

■ Qualification work: sTGC on-detector cooling system, 2017-2018:

My qualification task in ATLAS was the development and implementation of the sTGC on-detector cooling system. The sTGC (small Thin-Gap-Chambers) is a muon detector mounted on the New-Small-Wheel, which, at the time of writing this work, is being installed on the ATLAS detector [16]. The readout electronic boards of the sTGC are intended to operate in a harsh environment with large particle activity, as such, they are required to be cooled to maintain good operating conditions. The on-detector cooling relies on the ATLAS water cooling system, with cold water running in Copper pipes. The readout boards are placed in Cooper cooling cages, which are soldered to pipes. The thermal coupling between the electronic components to the cage surface is achieved via thermal foam. Within this project my responsibilities were the validation of the thermal heat transfer in measurements, set up of the cooling system to be used during assembly, and definition of leak test and heat contact integrity methods for the cooling pipe preparations.

Part I

Theory

The Standard Model of Particle Physics

If you are reading these lines I assume you have some level of familiarity with particle physics and the Standard Model. Nevertheless, one can not write a thesis detailing research performed in the Large-Hadron-Collider without including introductory theoretical background. The Standard Model (SM) of particle physics seems like an adequate place to start.

The SM is the theory that aims to describe the fundamental building blocks making up our universe, and the ways in which they interact. These building blocks are point-like entities called particles, which can be categorized into two general classes, depending on their spin. Particles that have integer spin ($s = 0, 1, 2, \dots$) are called *bosons* and particles that have half-integer spin ($s = \frac{1}{2}, \frac{3}{2}, \dots$) are referred to as *fermions*. But before we embark on our journey and discuss the nature and properties of these particles as described by the SM, it will be useful to first review some of the fundamental concepts on which this theory lies.

1.1 | Fundamental Concepts: Fields, Lagrangians, and Symmetries

The mathematical framework used to describe and analyze the physics of elementary particles is *Quantum Field Theory* (QFT). This framework ties together the theory of quantum mechanics with Einstein's theory of special relativity. In the context of QFT, space-

time is populated with mathematical constructs called *fields*, $\phi(x)$.¹ These fields carry out certain quantum numbers, and we can use them to create or annihilate particles with these quantum numbers at the space-time point $x \equiv x^\mu = (\vec{x}, t)$. In this picture, we can think of the physical particles as local excitation in the corresponding field. For example, there exists an electron field, and when this field is excited at a certain point in space-time the result is what we call an *electron*. For a comprehensive review of QFT, one can refer to [17]. Note that one consequence of this approach is that the elementary particles are point-like entities, with no associated physical dimensions.

The dynamics of the quantum fields described above are governed by another mathematical construct called *Lagrangian density*, $\mathcal{L}$ (in the following I will refer to it simply as "Lagrangian"). The Lagrangian of a theory is a function of the fields it consists of and their derivatives with respect to space-time, such that it can be written as $\mathcal{L}(\phi(x), \partial^\mu \phi(x))$. The *principle of least action* [18] dictates that the "action" (which is the space-time integral of $\mathcal{L}$) remains unchanged with respect to variations of the fields. This in turn leads to the Euler-Lagrange equation, from which the equation of motion (EOM) of the field can be derived:

$$\frac{\partial \mathcal{L}}{\partial \phi} = \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \right), \quad (1.1)$$

where the x dependence of the field ϕ is omitted. Note that Eq. (1.1) must be satisfied for each of the fields included in the theory.

One of the most fundamental concepts in particle physics is that of *symmetry*. Unlike many of the notions in this chapter, symmetry is a concept we also come across in our day-to-day lives, when describing things that have similar and balanced proportions. Yet physicists often use a more precise definition; to a physicist, a symmetry refers to a set of transformations under which a physical system remains unchanged or invariant. The importance of this concept comes from Noether's theorem [19], which relates between the (continuous) symmetries of a system to conserved quantities. For example, invariance under space (time) translations results in the conservation of momentum (energy). A particular type of symmetry, which is a key ingredient in the construction of the SM, is *gauge* symmetry. To better understand gauge symmetry, let us consider as

¹We typically use $\phi(x)$ to denote a *scalar* field (associated with spin-0). A spin- $\frac{1}{2}$ field (*spinor*) is most commonly denoted by $\psi(x)$, and a spin-1 field (*vector*) with $A^\mu(x)$. To avoid further confusion, throughout the text I will mention explicitly the relevant type of field I am referring to.

an example a massive complex scalar (i.e. spin-0) field ϕ , with mass given by m . The corresponding Lagrangian is:

$$\mathcal{L} = (\partial^\mu \phi) (\partial_\mu \phi^*) - m^2 \phi \phi^*. \quad (1.2)$$

This Lagrangian is invariant with respect to a rotation in an angle α in the complex plane:

$$\phi(x) \rightarrow \phi(x) e^{i\alpha} \quad ; \quad \phi^*(x) \rightarrow \phi^*(x) e^{-i\alpha}. \quad (1.3)$$

Therefore, this Lagrangian is said to be symmetric under *global* $U(1)$ transformations, with global denoting that α is the same for any space-time point x , and $U(1)$ is the one-dimensional group of unitary transformations. Note however, that if we take α to be dependent on x , the $U(1)$ symmetry is broken. Therefore the Lagrangian of Eq. (1.2) does not represent a *local* $U(1)$ symmetry. In order to make this Lagrangian invariant under local $U(1)$ transformations, we can introduce a new *vector* field $A_\mu(x)$ which transforms as:

$$A_\mu(x) \rightarrow A_\mu(x) - \frac{1}{q} \partial_\mu \alpha(x), \quad (1.4)$$

and replace the derivative ∂_μ in Eq. (1.2) with the *covariant* derivative D_μ defined as:

$$D_\mu = \partial_\mu + iqA_\mu(x), \quad (1.5)$$

where q is the coupling strength between the scalar field ϕ and the vector field A_μ . With these modifications, the Lagrangian is now symmetric under local $U(1)$ transformations. The field A_μ we've introduced is known as a *gauge field*, and the theory now described by the Lagrangian is known as a *gauge theory*. The terminology *gauge* implies that one can use different configurations of the fields in the theory and obtain identical observables. The number of gauge fields required to restore a given local symmetry depends on the number of generators in the gauge symmetry group. In our example, the $U(1)$ symmetry group has one generator, and hence a single gauge field was sufficient to restore the local gauge symmetry.

In the picture described above, the gauge fields are commonly thought of as interaction mediators that couple to the original fields in the theory.

1.2 | Particles and Forces

To the best of our current knowledge, there are four known fundamental forces in the universe: the strong nuclear force (SI), the weak nuclear force (WI), electromagnetism

Force	Relative Strength	Range [m]
Strong	1	10^{-15}
Electromagnetic	$\frac{1}{137}$	∞
Weak	10^{-6}	10^{-18}
Gravity	10^{-39}	∞

Table 1.1: The four fundamental forces, their strength compared to the strong interaction at their respective appropriate scale, and the range of the interaction.

(EM), and gravity. The first three can be represented within the framework of quantum field theory, and are encapsulated in the Standard Model. At the moment, there is no known way to reconcile between gravity and QFT, hence gravity is not included explicitly in the SM. Gravity, however, is much weaker than the other forces, and it is not relevant to the types of particle interaction that current experiments are able to probe. This is illustrated in Table 1.1 which lists the four forces, together with their relative strength², and the range of the interaction [20].

In the picture of the SM, the forces act between spin- $\frac{1}{2}$ fermions. The SM fermions can be further classified as *quarks* or *leptons* based on their quantum numbers and the charges they carry. Both leptons and quarks come in three generations (or families) which are essentially copies of the first. Each quark generation is comprised of an *up-type* quark, with an electric charge³ of $+\frac{2}{3}$, and a *down-type* quark with an electric charge of $-\frac{1}{3}$. The up-type quarks are, from first generation to third, the up (*u*), charm (*c*), and top (*t*), and the down type quarks are the down (*d*), strange (*s*), and bottom (*b*). Each generation also contains a pair of leptons, one that is charged with -1 and its neutral counterpart, called a neutrino - ν . Again, from first generation to third, they are named electron (*e* and ν_e), the muon (μ and ν_μ), and tau (τ and ν_τ). For each of these fermions there exist its *anti-fermion* counterpart, which has a similar mass but opposite-sign charge.

The forces in the SM are mediated by spin-1 gauge bosons (also called *vector bosons*).

²The "strength" of an interaction is somewhat an ambiguous term, as it depends on the energy scale at which the interaction is probed. This property of the forces is known as *running constant coupling* and more details will be given later in this chapter

³Throughout this work electric charge is given in units of the elementary charge e , defined as the amount of charge carried by a single electron

Each force corresponds to a local symmetry of the SM Lagrangian. As we already saw, keeping the Lagrangian invariant under a local symmetry requires the inclusion of additional gauge fields. The gauge bosons "carrying" the interactions are merely the physical representation of these gauge fields. The full gauge symmetry group of the SM is given by:

$$SU(3)_C \times SU(2)_L \times U(1)_Y. \quad (1.6)$$

The terms in Eq (1.6) represent the different interactions in the following way:

The description of the strong nuclear interaction is based on a gauge theory called **Quantum-Chromo-Dynamics (QCD)**, with an underlying symmetry group of $SU(3)_C$. Only particles that carry the strong charge, also referred to as *color* charge, are subject to strong interactions. The term "color" arises from the fact that the strong charge is found to exist in three states labeled as red (r), green (g), and blue (b). In an analogy to the components of white light, colorless states are formed when the three states are combined ($r + g + b$), or with a combination of a color and anti-color pair ($r + \bar{r}$), ($g + \bar{g}$), or ($b + \bar{b}$). It is an empirical fact that only colorless states are observed in nature. The gauge bosons mediating the strong interaction are eight massless *gluons*, (g).

The description of electromagnetism and weak nuclear interactions in the SM is unified based on the **GWS electroweak** theory. In this model, proposed by Glashow [21], Weinberg [22], and Salam [23], the symmetry group is $SU(2)_L \times U(1)_Y$. Particles that are subject to weak interactions are charged under *weak isospin*, which corresponds to the $SU(2)_L$ part of the symmetry group. Three W gauge bosons are associated with $SU(2)$ weak isospin transformation labeled W_1, W_2, W_3 . The charge related to the $U(1)_Y$ part of the symmetry group is called *weak hypercharge*, and it is mediated by a single gauge boson B .

At this point you might ask yourself: "what about the electromagnetic charge we are all familiar with from our daily lives, and that we learned about in our classical physics course". The answer is that the EM charge only emerges after the unification of the weak and electromagnetic forces breaks. This happens via a mechanism known as *electroweak symmetry breaking* (EWSB), proposed by Higgs [24], Brout, and Englert [25]. As a result of EWSB the three W and the B bosons are transformed into a massless *photon* (γ) responsible for mediating the EM interaction, and three massive gauge bosons, the charged $W^\pm$ and neutral Z which propagate the weak interaction.

Another feature of EWSB is that it requires the addition of a complex scalar field to the theory. This scalar field is manifested into a massive spin-0 boson, known as the *Higgs* boson. It is through coupling to the Higgs field that the gauge bosons and fermions in the SM acquire their mass. The details of the EWSB mechanism are discussed later in this section. But for now, with the inclusion of the Higgs boson, we have our current picture of the SM particles, which are summarized together with their properties in Fig. 1.1.

Three Generations of Matter (Fermions)					
	I	II	III		
mass→	2.4 MeV/c ²	1.27 GeV/c ²	171.2 GeV/c ²	0	±125 GeV/c ²
charge→	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	0	0
spin→	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	0
name→	u up	c charm	t top	γ photon	H Higgs- boson
Quarks	4.8 MeV/c ²	104 MeV/c ²	4.2 GeV/c ²	0	
	$-\frac{1}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	0	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
	d down	s strange	b bottom	g gluon	
Leptons	<2.2 eV/c ²	<0.17 MeV/c ²	<15.5 MeV/c ²	91.2 GeV/c ²	
	0	0	0	0	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
	ν_e electron neutrino	ν_μ muon neutrino	ν_τ tau neutrino	Z⁰ weak force	
Bosons (Forces)	0.511 MeV/c ²	105.7 MeV/c ²	1.777 GeV/c ²	80.4 GeV/c ²	
	-1	-1	-1	±1	
	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	
	e electron	μ muon	τ tau	W[±] weak force	

Figure 1.1: The particle content of the Standard Model.

1.3 | Quantum-Chromo-Dynamics (QCD)

Quantum-Chromo-Dynamics (QCD) is the field theory associated with the non-abelian $SU(3)_C$ symmetry group. It describes the strong nuclear interactions binding together the atomic nuclei. There are 8 generators associated with this symmetry group, corresponding to the gauge boson fields called *gluons* and labeled G_a^μ (with $a = 1, \dots, 8$). Among the SM fermions only quarks participate in strong interactions as only they are charged with the color charge related to this interaction. However, in QCD the force-carrying gluons also carry the color charge, which means that they are also subject to the strong force. We will soon see how this is in fact a consequence of the non-abelian structure of the $SU(3)_C$ symmetry group.

We can write the Lagrangian of QCD in the compact form:

$$\mathcal{L}_{QCD} = -\frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu} + \sum_f \bar{q}_f i \gamma_\mu D^\mu q_f. \quad (1.7)$$

In this expression q_f is a quark field with flavor $f = u, d, c, s, t, b$, and $G_{\mu\nu}^a$ is the kinetic tensor, which accounts for the propagation of the gluon gauge fields in space and time. It is defined as

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f^{abc} G_\mu^b G_\nu^c, \quad (1.8)$$

where g_s is the *strong coupling constant* and f^{abc} are the *structure constants* of the gauge group. The covariant derivative for QCD takes the form of:

$$D^\mu = \left(\partial^\mu - i g_s \frac{\lambda^a}{2} G_a^\mu \right), \quad (1.9)$$

where λ^a are the eight *Gell-Mann* matrices:

$$\begin{aligned} \lambda^1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda^2 &= \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda^3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ \lambda^4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, & \lambda^5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, & \lambda^6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\ \lambda^7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, & \lambda^8 &= \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \end{aligned} \quad (1.10)$$

Plugging Eq. (1.8) and Eq. (1.9) into Eq. (1.7) we see the the QCD Lagrangian in its extended form:

$$\begin{aligned}
\mathcal{L}_{\text{QCD}} = & -\frac{1}{4} \left(\partial_\mu G_\nu^a - \partial_\nu G_\mu^a \right) \left(\partial^\mu G_a^\nu - \partial^\nu G_a^\mu \right) + \sum_f \bar{q}_f i \gamma_\mu \partial^\mu q_f \\
& - g_s f^{abc} (\partial_\mu G_{a\nu}) G_b^\mu G_c^\nu - \frac{g_s^2}{4} f^{abc} f^{ade} G_b^\mu G_c^\nu G_{d\mu} G_{e\nu} \\
& + g_s G_a^\mu \sum_f \bar{q}_f^a \gamma_\mu \left(\frac{\lambda^a}{2} \right)_{\alpha\beta} q_f^\beta.
\end{aligned} \tag{1.11}$$

In this expression, the first row represents the kinetic term of the gluon and quark fields. The second row represents the self-interacting terms of the gluon fields. Note that these terms are proportional to the structure constants f^{abc} . Since only in non-abelian symmetry groups do we have non-vanishing structure constants $f^{abc} \neq 0$, we can now understand how strong interactions between gluons are a consequence of the non-commutativity of the $SU(3)_C$ group. The last row represents the gluon-quark interaction introduced by the gauge terms of the covariant derivative of Eq. (1.9). The Feynman diagrams associated with the QCD gauge interaction terms are presented in Fig. 1.2.

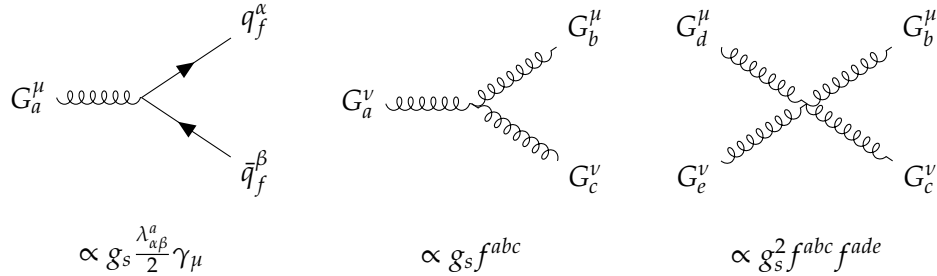


Figure 1.2: Feynman diagrams for the QCD gauge interaction terms.

Note that the gluon-quark-quark interaction terms in Eq. (1.11) are only allowed for quarks with the same flavor, and quark mixing is forbidden in the strong interaction.

It is an empirical fact that no colored states are observed in nature. This feature of QCD is known as *color confinement*. As a consequence, quarks are only found in color-singlet bound states known as *hadrons*. Hadrons can be formed, for example, as a combination of three valence quarks with colors $r + g + b$ (Baryons), or a combination of a quark and anti-quark pair with $r + \bar{r}$, $g + \bar{g}$, or $b + \bar{b}$ (Mesons). In addition to the valence quarks, hadrons also contain *sea quarks*, which are virtual quark+anti-quark pairs

that are constantly created and annihilated from gluons inside the hadron. More recently additional hadronic states consisting of four quarks (tetraquark) and five quarks (pentaquark), have been observed [26, 27].

Another feature of QCD is that as the distance between quarks decreases the strength of the interaction between them becomes smaller. This phenomenon is known as *asymptotic freedom*. At low momentum transfers (large distances) the strong coupling constant α_s becomes large, making it impossible to use perturbation theory for QCD predictions. However, at high momentum transfers (small distances), perturbation theory holds due to the small value of α_s , and in this regime QCD predictions are well tested.

1.4 | Electroweak Theory (EW)

The electroweak interaction is felt by all fermions in the SM. It is a *chiral* theory, meaning it distinguishes between left- and right-handed components of the fermion fields. Chirality, or "handedness", is an intrinsic property of fermions. To better understand the notion of *chirality* we can first examine the concept of *helicity*, which is defined as the projection of a particle's spin onto its momentum. A particle can have either right-handed helicity where its spin orientation and direction of motion coincide, or left-handed helicity where these are in opposite directions. The helicity of a massive particle depends on the observer frame of reference, as one can always find a frame in which the momenta is pointing in the opposite direction. However, for massless particles traveling at the speed of light, no such frame exists and they are always in a state of definite helicity. Hence for massless particles helicity is equivalent to chirality. For massive particles, only chirality is invariant of choice of reference frame.

The symmetry group describing the electroweak interaction is $SU(2)_L \times U(1)_Y$. The $SU(2)_L$ part of the symmetry group corresponds to the *weak isospin* which acts on the left-handed components of the lepton and quark fields. The $U(1)_Y$ part represents the *weak hypercharge* acting on both the left- and right-handed components of these fields.

The electroweak Lagrangian can be written as:

$$\begin{aligned} \mathcal{L}_{EW} = & -\frac{1}{4}W_{\mu\nu}^i W_i^{\mu\nu} - \frac{1}{4}B_{\mu\nu} B^{\mu\nu} \\ & + \sum_f \bar{l}_L^f i\gamma_\mu D^\mu l_L^f + \bar{e}_R^f i\gamma_\mu D^\mu e_R^f + \bar{\nu}_R^f i\gamma_\mu D^\mu \nu_R^f \\ & + \sum_f \bar{q}_L^f i\gamma_\mu D^\mu q_L^f + \bar{u}_R^f i\gamma_\mu D^\mu u_R^f + \bar{d}_R^f i\gamma_\mu D^\mu d_R^f \end{aligned} \quad (1.12)$$

Similarly to Eq. (1.8), here $W_i^{\mu\nu}$ and $B^{\mu\nu}$ represents the kinetic gauge tensors associated to the $SU(2)_L$ and $U(1)_Y$ groups respectively. They can be written explicitly as:

$$\begin{aligned} W_i^{\mu\nu} &= \partial_\mu W_\nu^i - \partial_\nu W_\mu^i + g\epsilon^{ijk}W_\mu^jW_\nu^K \\ B^{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \end{aligned} \quad (1.13)$$

The index i represents the three generators associated with $SU(2)_L$, ϵ^{ijk} is the Levi-Civita anti-symmetric tensor, which is the structure constant of the symmetry group, and g is the weak isospin coupling constant. B^μ is the gauge field associated with the single generator of $U(1)_Y$. The L and R indices in Eq. (1.12) are used to distinguish between left- and right-handed fermions. Since $SU(2)_L$ acts only on left-handed fermions it is convenient to write the left-handed leptons and quarks as doublets:

$$\begin{aligned} l_L^f &= \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix} , \quad \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix} , \quad \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix} \\ q_L^f &= \begin{pmatrix} u_L \\ d_L \end{pmatrix} , \quad \begin{pmatrix} c_L \\ s_L \end{pmatrix} , \quad \begin{pmatrix} t_L \\ b_L \end{pmatrix} , \end{aligned} \quad (1.14)$$

and the right handed components as singlets:

$$\begin{aligned} \nu_R^f &= \nu_{eR} , \nu_{\mu R} , \nu_{\tau R} ; \quad e_R^f = e_R , \mu_R , \tau_R \\ u_R^f &= u_R , c_R , t_R ; \quad d_R^f = d_R , s_R , b_R. \end{aligned} \quad (1.15)$$

The covariant derivatives appearing in Eq. (1.12) are defined according to handedness of the fermion fields on which they act:

$$\begin{aligned} D^\mu l_L^f &= \left(\partial^\mu - ig \frac{\sigma_i}{2} W_i^\mu - ig' \frac{Y}{2} B^\mu \right) l_L^f \\ D^\mu e_R^f &= \left(\partial^\mu - ig' \frac{Y}{2} B^\mu \right) e_R^f \\ D^\mu \nu_R^f &= \left(\partial^\mu - ig' \frac{Y}{2} B^\mu \right) \nu_R^f. \end{aligned} \quad (1.16)$$

The same definition holds for quark fields replacing l_L^f with q_L^f and e_R, ν_R with u_R, d_R . g' is the coupling constant related to $U(1)_Y$ and Y is the weak-hypercharge. The value

of Y associated with each field is listed in Table 1.2. The σ_i appearing in the first line of Eq. (1.16) are the three *Pauli matrices* generating the $SU(2)_L$ group transformations:

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (1.17)$$

We have introduced four gauge bosons related to the EW interaction, three W_i^μ and a single B^μ . Note however, that these do not correspond yet to the *photon* and weak gauge bosons $W^\pm, Z$ that we observe in nature. We will soon see how one obtains such gauge boson fields when splitting the interaction terms in the EW Lagrangian into charged and neutral interactions.

1.4.1 | Charged Electroweak Interaction

Electroweak charged interactions take place between particles of different flavor. Namely, they involve either a charged lepton and a neutrino or an up-type and a down-type quarks. These interactions are introduced by the off-diagonal terms of the $SU(2)_L$ generators, σ^1 and σ^2 . The corresponding part of the EW Lagrangian is:

$$\begin{aligned} \mathcal{L}_{int}^C = & \frac{g}{2} \sum_f \bar{l}_L^f \gamma_\mu (\sigma_1 W_1^\mu + \sigma_2 W_2^\mu) l_L^f \\ & + \frac{g}{2} \sum_f \bar{q}_L^f \gamma_\mu (\sigma_1 W_1^\mu + \sigma_2 W_2^\mu) q_L^f. \end{aligned} \quad (1.18)$$

If we define $\sigma^\pm \equiv \frac{1}{2} (\sigma^1 \pm i\sigma^2)$ we can obtain the gauge fields associated to the $W^\pm$ bosons:

$$W_\pm^\mu = \frac{1}{\sqrt{2}} (W_1^\mu \mp iW_2^\mu), \quad (1.19)$$

and Eq. (1.18) becomes:

$$\mathcal{L}_{int}^C = -\frac{g}{\sqrt{2}} \left(\sum_{l=e,\mu,\tau} \bar{\nu}_{lL} \gamma^\mu W_\mu^+ l_L + \sum_{\substack{q=u,c,t \\ q'=d,s,b}} \bar{q}_L \gamma^\mu W_\mu^+ V_{qq'} q_L' \right) + h.c., \quad (1.20)$$

where $V_{qq'}$ are terms in the Cabibbo-Kobayashi-Maskawa (CKM) unitary matrix, which accounts for the mixing of quark flavors:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \quad (1.21)$$

The Feynman diagrams associated with the charged weak interaction terms are presented in Fig. 1.3

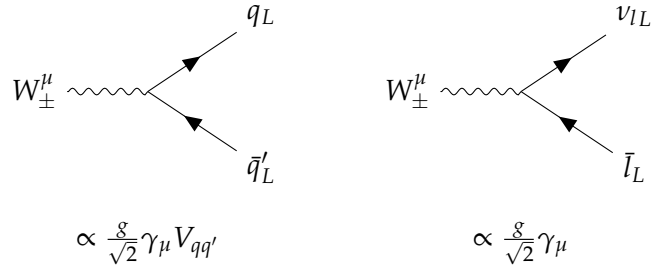


Figure 1.3: Feynman diagrams for the charged weak interaction terms.

Eq. (1.20) reveals some essential features of the electroweak theory:

1. Only left-handed particles take part in charged EW interactions. Consequently, parity is violated by these interactions.
2. The charged weak interaction between leptons is both diagonal and universal. Diagonality refers to the fact that the interaction couples each charged lepton to a single neutrino of the same flavor. Universality implies that the coupling of the $W^\pm$ to the leptonic flavors e, μ, τ is equal.
3. Charged weak interaction between quarks is neither diagonal nor universal. Diagonality is broken by V_{CKM} which allows for interactions between different generations of quarks. The distinct values of the individual terms of V_{CKM} implies that the coupling between $W^\pm$ and different flavors of quarks is non-universal.

1.4.2 | Neutral Electroweak Interaction

Neutral EW interaction correspond to the diagonal interaction terms of Eq. (1.12) introduced by the $SU(2)_L$ generator σ^3 as well as the interaction term of right-handed fermion fields. Thus, the neutral interaction terms of the EW Lagrangian are given by:

$$\begin{aligned}
\mathcal{L}_{int}^N = & \sum_f \bar{l}_L^f \gamma_\mu \left(\frac{g}{2} \sigma_3 W_3^\mu + \frac{g'}{2} Y B^\mu \right) l_L^f \\
& + \sum_f \bar{q}_L^f \gamma_\mu \left(\frac{g}{2} \sigma_3 W_3^\mu + \frac{g'}{2} Y B^\mu \right) q_L^f \\
& + \sum_f \frac{g'}{2} Y B^\mu \left(\bar{\nu}_R^f \gamma_\mu \nu_R^f + \bar{e}_R^f \gamma_\mu e_R^f \right) + \sum_f \frac{g'}{2} Y B^\mu \left(\bar{u}_R^f \gamma_\mu u_R^f + \bar{d}_R^f \gamma_\mu d_R^f \right).
\end{aligned} \tag{1.22}$$

We can define Ψ as a vector composed of all the left- and right- handed quark and lepton fields, and introduce a new tensor T_3 that acts on these fields in the following way:

$$\begin{aligned}
T_3 \left(\nu_L^f \right) &= T_3 \left(u_L^f \right) = +\frac{1}{2} \\
T_3 \left(e_L^f \right) &= T_3 \left(d_L^f \right) = -\frac{1}{2} \\
T_3 \left(\nu_R^f \right) &= T_3 \left(e_R^f \right) = T_3 \left(u_R^f \right) = T_3 \left(d_R^f \right) = 0.
\end{aligned} \tag{1.23}$$

This makes it possible to write Eq. (1.22) in the compact form:

$$\mathcal{L}_{int}^N = \bar{\Psi} \gamma_\mu \left(g T_3 W_3^\mu + g' \frac{Y}{2} B^\mu \right) \Psi. \tag{1.24}$$

Yet Eq. (1.24) still do not reveal the relation to the Weak and Electromagnetic interactions realized in nature. In order to obtain such a description it is necessary to first rotate the fields B^μ and W_3^μ by an angle θ_W :

$$\begin{pmatrix} B^\mu \\ W_3^\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & -\sin \theta_W \\ \sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} A^\mu \\ Z^\mu \end{pmatrix}, \tag{1.25}$$

which transforms Eq. (1.24) into:

$$\begin{aligned}
\mathcal{L}_{int}^N = & \bar{\Psi} \gamma_\mu \left(g \sin \theta_W T_3 + g' \cos \theta_W \frac{Y}{2} \right) A^\mu \Psi \\
& + \bar{\Psi} \gamma_\mu \left(g \cos \theta_W T_3 - g' \sin \theta_W \frac{Y}{2} \right) Z^\mu \Psi.
\end{aligned} \tag{1.26}$$

The rotation angle θ_W is known as the *Weinberg angle* or *weak mixing angle*. We wish that the field A^μ will correspond to the photon mediating EM interactions. Therefore we require that the term proportional to A^μ matches the coupling between the photon

Fermion	Hypercharge Y	T_3	EM charge Q	Weak charge $Q_Z \cdot (\sin \theta_W \cos \theta_W)$
ν_L^f	-1	$+\frac{1}{2}$	0	$\frac{1}{2}$
e_L^f	-1	$-\frac{1}{2}$	-1	$-\frac{1}{2} + \sin^2 \theta_W$
ν_R^f	0	0	0	0
e_R^f	-2	0	-1	$\sin^2 \theta_W$
u_L^f	$+\frac{1}{3}$	$+\frac{1}{2}$	$+\frac{2}{3}$	$\frac{1}{2} - \frac{2}{3} \sin^2 \theta_W$
d_L^f	$+\frac{1}{3}$	$-\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{1}{2} + \frac{1}{3} \sin^2 \theta_W$
u_R^f	$+\frac{4}{3}$	0	$+\frac{2}{3}$	$-\frac{2}{3} \sin^2 \theta_W$
d_R^f	$-\frac{2}{3}$	0	$-\frac{1}{3}$	$\frac{1}{3} \sin^2 \theta_W$

Table 1.2: Electroweak charges for the lepton and quark fields.

and the electron field which is described by a quantum field theory called *Quantum-Electrodynamics* and takes the form:

$$\mathcal{L}_{int}^{EM} = e \bar{\Psi} \gamma_\mu Q \Psi A^\mu. \quad (1.27)$$

Here e represents the EM coupling strength and Q is the EM charge. For the two terms to match it is sufficient to satisfy:

$$\begin{aligned} g \sin \theta_W &= g' \cos \theta_W = e \\ Q &= T_3 + \frac{Y}{2}. \end{aligned} \quad (1.28)$$

This enables us to further simplify the neutral Weak interaction Lagrangian, and write it in the elegant form:

$$\mathcal{L}_{int}^N = e \bar{\Psi} \gamma_\mu (Q A^\mu + Q_Z Z^\mu) \Psi, \quad (1.29)$$

where we have defined the Q_Z as the *weak charge*:

$$Q_Z \equiv \frac{1}{\sin \theta_W \cos \theta_W} (T_3 - Q \sin^2 \theta_W). \quad (1.30)$$

We now recognize in Eq (1.29) the two interaction terms corresponding to the photon and Z-boson, that match what we empirically know to exist in nature.

Table 1.2 summarizes the values of Y , T_3 , Q and Q_Z for the different lepton and quark fields. One conclusion that emerges from Table 1.2 is that $Q(\nu_R^f) = Q_Z(\nu_R^f) = 0$.

This implies that right-handed neutrinos do not take part in EM and Weak interactions. Moreover, as we already saw that leptons of any kind do not participate in strong interactions, we have no description of right-handed neutrinos within the SM.

The numeric value of θ_W is a free parameter of the theory, which is to be determined from experiments such as described in [28]. Empirically we find $\sin^2 \theta_W = 0.23146 \pm 0.00012$.

The Feynman diagrams associated with the neutral electroweak interaction terms are in Fig. 1.4.

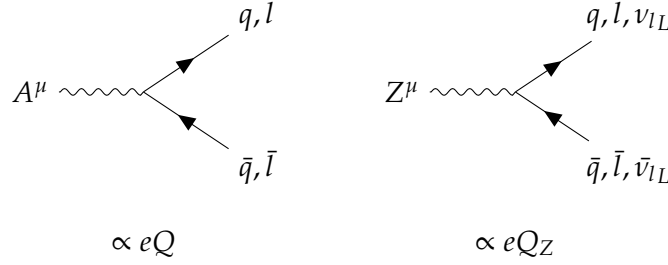


Figure 1.4: Feynman diagrams for the neutral electroweak interaction terms.

1.4.3 | Electroweak gauge bosons-self Interactions

The covariant derivative in Eq. (1.12) also give rise to self-interaction terms between the EW gauge bosons. These are given by:

$$\begin{aligned}
 \mathcal{L}_{int}^{Gauge} = & ig \sin \theta_W \left(W_{\mu\nu}^+ W_-^\mu A^\nu - W_{\mu\nu}^- W_+^\mu A^\nu + A_{\mu\nu} W_+^\mu A_-^\nu \right) \\
 & + ig \cos \theta_W \left(W_{\mu\nu}^+ W_-^\mu Z^\nu - W_{\mu\nu}^- W_+^\mu Z^\nu + Z_{\mu\nu} W_+^\mu Z_-^\nu \right) \\
 & + \frac{g^2}{2} (2\eta^{\mu\nu}\eta^{\rho\sigma} - \eta^{\mu\rho}\eta^{\nu\sigma} - \eta^{\mu\sigma}\eta^{\nu\rho}) \times \\
 & \left[\frac{1}{2} W_\mu^+ W_\nu^+ W_\rho^- W_\sigma^- - W_\mu^+ W_\nu^- (A_\rho A_\sigma \sin^2 \theta_W + Z_\rho Z_\sigma \cos^2 \theta_W + 2A_\rho Z_\sigma \cos \theta_W \sin \theta_W) \right],
 \end{aligned} \tag{1.31}$$

where we have adopted the definition of the kinetic tensors for the new, rotated fields we have introduced in Sections 1.4.1 and 1.4.2:

$$\begin{aligned}
 W_\pm^{\mu\nu} &= \partial_\mu W_\nu^\pm - \partial_\nu W_\mu^\pm \\
 Z^{\mu\nu} &= \partial_\mu Z_\nu - \partial_\nu Z_\mu \\
 A^{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu.
 \end{aligned} \tag{1.32}$$

Note that the terms in the first two lines of Eq. (1.31) correspond to interactions involving three bosons, and those in the third and fourth lines are four-boson interaction terms. The Feynman diagrams describing the gauge bosons self-interaction are shown in Fig 1.5

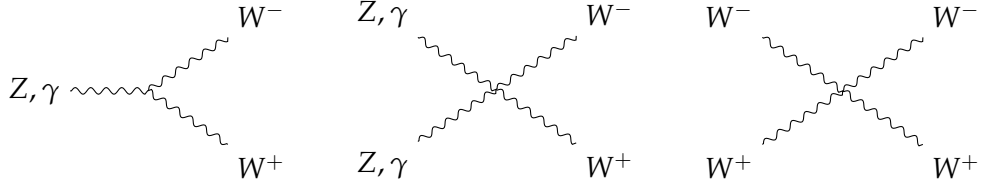


Figure 1.5: Feynman diagrams for the electroweak bosons self-interaction terms.

1.5 | The Higgs Mechanism

One consequence of the left-handed fermion fields being $SU(2)_L$ doublets, while the right-handed components are singlets, is that it is impossible to include an explicit mass term for the fermions in the SM Lagrangian. An explicit mass term in the form $m_f \bar{\psi}\psi$ will mix the left- and right-handed components, which violates the weak isospin symmetry. A similar problem occurs if we try to include an explicit mass term for the EW gauge bosons, as such a term will violate the local $SU(2)_L \times U(1)$ gauge invariance of the SM Lagrangian. Yet, since the fermions and EW gauge bosons we observe in nature are massive, we must come up with an alternative way of generating mass terms for the particles. This is done through the concept of *spontaneous symmetry breaking* by introducing additional scalar field(s) to the theory.

Note that this section describes the *theoretical* aspect related to the Higgs mechanism and the way masses are generated in the SM. The reader can find an overview of the phenomenology of the Higgs boson in the LHC, including the favorable production mechanisms and decay channels, in Sec. 2.2.

1.5.1 | Spontaneous Symmetry Breaking

Before we see how mass terms are generated in the SM, let us first demonstrate the concept of spontaneous symmetry breaking by examining a more simple scenario. Consider the Klein-Gordon Lagrangian of a scalar field ϕ with a mass term μ , and a field self-interaction term λ :

$$\mathcal{L}_\phi = \frac{1}{2} \partial^\mu \phi \partial_\mu \phi - \frac{1}{2} \mu^2 \phi^2 - \frac{1}{4} \lambda \phi^4. \quad (1.33)$$

$\mathcal{L}_\phi$ is invariant under the transformation $\phi \rightarrow -\phi$. For the potential $V(\phi) = \frac{1}{2} \mu^2 \phi^2 + \frac{1}{4} \lambda \phi^4$ to have a finite minimum, λ must be positive. However, there is no such restriction for μ^2 . For $\mu^2 > 0$ the potential $V(\phi)$ has a minimum at $\phi = 0$, as shown in Fig 1.6(a). This minimum corresponds to the ground state, also known as the vacuum $\langle 0|\phi|0\rangle$. However, if we chose $\mu^2 < 0$, the associated term in the Lagrangian can no longer be interpreted as a mass term and, as seen in Fig 1.6(b), the potential has a minima at:

$$\phi = \pm v = \pm \left| \sqrt{\frac{-\mu^2}{\lambda}} \right|. \quad (1.34)$$

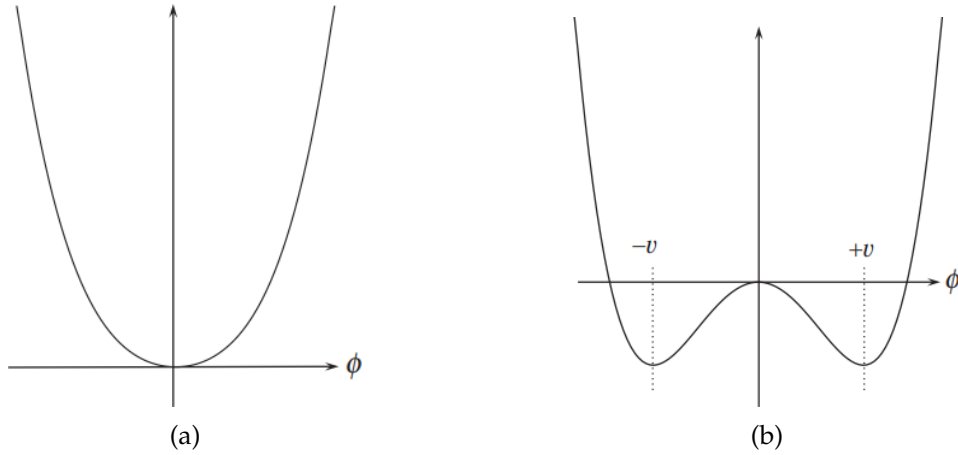


Figure 1.6: The one-dimensional potential $V(\phi) = -\frac{1}{2} \mu^2 \phi^2 + \frac{1}{4} \lambda \phi^4$, with $\lambda > 0$ and $\mu^2 > 0$ (a) and $\mu^2 < 0$ (b).

We can obtain the particle states, which are described by excitations of the field, by choosing the ground state to be at $\phi = +v$ and expanding around the minima $\phi(x) = v + \eta(x)$. The Lagrangian then takes the form:

$$\mathcal{L}_\eta = \frac{1}{2} \partial^\mu \eta \partial_\mu \eta - \lambda v^2 \eta^2 - \lambda v \eta^3 - \frac{1}{4} \lambda (\eta^4 - v^4). \quad (1.35)$$

From comparing with the Lagrangian of Eq. (1.33) we see that the term proportional to η^2 can be interpreted as a mass term $m_\eta = \sqrt{-2\mu^2}$. The higher-order terms in η are the self-coupling terms, corresponding to triple and quartic couplings.

Notice that the new Lagrangian $\mathcal{L}_\eta$ is no longer invariant under $\eta \rightarrow -\eta$. The symmetry of the Lagrangian was broken by our choice of the vacuum state $\phi = +v$. This is an example of spontaneous symmetry breaking.

1.5.2 | Vector Boson Masses

In order to provide masses to the EW gauge bosons in the SM, we extend the example above to the spontaneous breaking of the $SU(2)_L \times U(1)_Y$ local gauge symmetry. This requires the introduction of a complex scalar $SU(2)_L$ doublet field Φ :

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \quad (1.36)$$

This scalar field is defined to have hypercharge $Y = +1$. The gauge-invariant Lagrangian term of this scalar is:

$$\mathcal{L}_H = (D^\mu \Phi)^\dagger (D_\mu \Phi) - \mu^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2, \quad (1.37)$$

where, similarly to the left-handed quark and lepton fields which are also doublets of $SU(2)_L$, the covariant derivative acting on Φ takes the form:

$$D_\mu \Phi = \left(\partial_\mu - ig \frac{\sigma_i}{2} W_\mu^i - ig' \frac{Y}{2} B_\mu \right) \Phi. \quad (1.38)$$

For $\mu^2 < 0$ the potential has a minima satisfying:

$$\Phi^\dagger \Phi = \frac{1}{2} (\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = \frac{v^2}{2} = -\frac{\mu^2}{2\lambda}. \quad (1.39)$$

Repeating the steps in the previous example, we chose the physical vacuum to be at $\phi_1 = \phi_2 = \phi_4 = 0$, $\phi_3 = v$ and expand around this minima $\Phi = \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}$. We will soon show that $h(x)$ is a real massive scalar field that corresponds to the Higgs boson. Using this parametrization of $\Phi(x)$ the kinetic term of Eq. (1.37) becomes:

$$(D^\mu \Phi)^\dagger (D_\mu \Phi) = \frac{1}{2} \partial^\mu h \partial_\mu h + (v + h)^2 \left(\frac{g^2}{4} W_\mu^\pm W_\mu^\pm + \frac{g^2}{8 \cos^2 \theta_W} Z_\mu Z^\mu \right). \quad (1.40)$$

One can identify the terms proportional to v^2 as mass terms of the $W^\pm$ and Z bosons:

$$m_W^2 = \frac{1}{4}g^2v^2, \quad m_Z^2 = \frac{1}{8\cos^2\theta_W}g^2v^2, \quad m_A = 0. \quad (1.41)$$

The expressions for the EW gauge boson masses depend on g , v , and $\cos\theta_W$ (or alternatively g'), which are free parameters of the SM that must be determined experimentally. Experiments have confirmed the existence of two massive spin-1 bosons related to the weak sector and found their masses to be $m_W = 80.363 \pm 0.020$ GeV and $m_Z = 91.1876 \pm 0.0021$ GeV [28]. Notice that as expected, the photon field A^μ remains massless.

It is also possible to explain the masses acquired by the $W^\pm$ and Z boson in terms of *Goldstone's theorem* [29], which states that for every broken continuous symmetry appears an associated massless scalar field (known as a *Goldstone boson*). Here we interpret the masses as if the $W^\pm$ and Z fields acquire longitudinal field components after absorbing the Goldstone boson degrees of freedom.

The terms proportional to vh and h^2 in Eq (1.40) corresponds to interactions involving the $W^\pm$, Z bosons and either one or two Higgs bosons. It can also be noticed that these interaction terms are proportional to the mass of the vector boson:

$$\mathcal{L}_{hVV} \propto \frac{m_V^2}{v}, \quad \mathcal{L}_{hhVV} \propto \frac{m_V^2}{v^2}. \quad (1.42)$$

The Feynman diagrams describing the interaction between the EW gauge bosons and the Higgs boson are presented in Fig 1.7

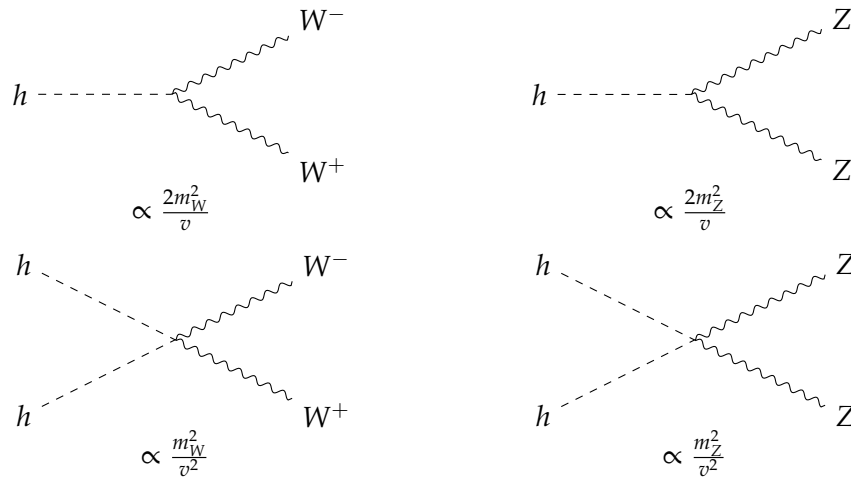


Figure 1.7: Feynman diagrams for the interaction between the $W^\pm$ and Z bosons and the Higgs boson.

1.5.3 | Fermion Masses

The masses of the EW gauge boson emerge directly from the covariant derivative acting on the scalar field expanded around it's minima. Fermion masses in the SM are generated via the *Yukawa* interaction which takes place between the fermions and the scalar field Φ . The Yukawa terms in the Lagrangian take the form:

$$\mathcal{L}_{Yuk} = - \sum_{f=e,\mu,\tau} y_f \bar{l}_L^f \Phi e_R^f - \sum_{f=d,s,b} y_f \bar{q}_L^f \Phi d_R^f - \sum_{f=u,c,t} y_f \bar{q}_L^f \Phi^* u_R^f + \text{h.c.}, \quad (1.43)$$

with y_f the Yukawa coupling of a fermion with flavor f to the Higgs field. After spontaneous symmetry breaking the Yukawa terms translates into:

$$- \frac{y_f}{\sqrt{2}} v (\bar{f}_R f_L + \bar{f}_L f_R) - \frac{y_f}{\sqrt{2}} h (\bar{f}_R f_L + \bar{f}_L f_R). \quad (1.44)$$

The first term in Eq. (1.44) can be associated with a mass term for the fermion f with a mass $m_f = v \frac{y_f}{\sqrt{2}}$. The second term represents the interaction of the fermion with the Higgs boson itself, which is demonstrated by the Feynman diagram in Fig. 1.8. Notice that similarly to the vector bosons, the coupling of the Higgs to the fermions is proportional to the fermions mass.

Similar to the case of the vector bosons, the masses of the fermions in the SM also depend on free parameters of the theory - the Yukawa couplings y_f and the vacuum expectation value (vev) v . The measured mass of each fermion is listed in Fig. 1.1.

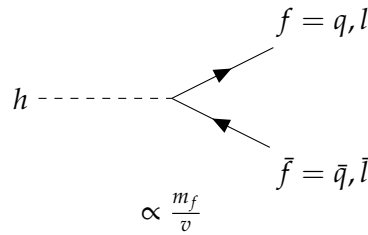


Figure 1.8: Feynman diagram for the interaction between the SM fermions and the Higgs boson.

1.5.4 | The SM Higgs Boson

We identify $h(x)$, the excitation of the scalar field around the vev as the physical Higgs boson. From our choice to expand in the lower ($T_3 = -\frac{1}{2}$) components of the doublet Φ

we can see that the Higgs boson is electromagnetically neutral, as according to Eq. (1.28) the EM charge is given by: $Q = T_3 + \frac{Y}{2} = +\frac{1}{2} - \frac{1}{2} = 0$. After the EW symmetry is broken and the scalar Φ field acquires a vacuum expectation value, the potential becomes:

$$V(\Phi) \rightarrow V(h) = \lambda v^2 h^2 + \lambda v h^3 + \frac{1}{4} \lambda h^4 + \text{const.}, \quad (1.45)$$

where the const. term includes the terms which do not directly depend on h . We can identify the first term as a mass term for the Higgs boson $m_h^2 = 2\lambda v^2$. The other two terms in Eq. (1.45) are the triple and quartic self-interaction vertices, the Feynman diagrams for which are shown in Fig 1.9. The kinetic term of the Higgs boson is obtained directly from the covariant derivative of Eq. (1.40).

The Higgs boson was discovered at the LHC in 2012 [4, 5]. Its discovery marks the observation of all particles predicted by the SM. The mass of the Higgs boson, which is another free parameter of the SM, was found to be $m_h = 125$ GeV.

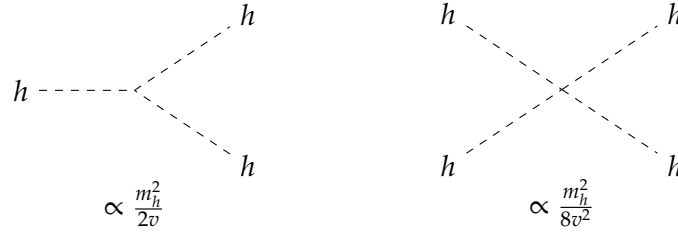


Figure 1.9: Feynman diagrams for the Higgs boson triple and quartic self-interaction.

1.6 | Limitations of the SM

The Standard Model can explain all observed phenomena in collider physics in a wide range of energies [30]. Its particle content includes and describes all visible matter in the universe. Yet despite the good agreement between the SM predictions and observations, there are several experimental signatures, mostly from non-collider experiments, which can not be explained in the framework of the SM. These signatures hint at *new physics*, that underlies beyond the Standard Model (BSM). This section provides an overview of (some) of the open questions.

Gravity The most striking evidence that the SM is not a complete theory that accounts for all observed phenomena in the universe, is the fact that it does not include a description of gravity. For gravity to be included in the framework of the SM in a similar man-

ner to the other forces, an additional force-carrying particle is required, the stipulated *graviton*, responsible for mediating the gravitational interaction. Such a particle is yet to be observed. The current theory of gravity is general relativity, which is essentially not a quantum theory. Constructing a quantum theory of gravity has proven to be a difficult task. For an in-depth overview of the problems that arise from the quantization of the gravitational field the reader can refer to [31].

Dark Matter and Dark Energy Since the mid-1930s we know that ordinary matter, described by the SM, accounts for $\sim 5\%$ of the universe [32]. From the remaining contributions, $\sim 22\%$ comes from *Dark Matter* (DM), which is the name given to the fraction of mass not bound up in the luminous stars, gas, or other known forms. A number of astrophysical and cosmological observations point to the existence of DM. One example is the velocity distribution of stars as they orbit the galactic center, which is not consistent with the theoretical prediction taking into account only ordinary matter. DM was introduced to solve this problem as an additional massive component that would modify the gravitational properties of galaxies and reconcile between the prediction and measurements. At the time of writing this work, Dark Matter is yet to be directly detected, but a number of its properties are understood from cosmological constraints. For example, in order to be non-relativistic and form large galactic structures at the time of decoupling from the early universe plasma, we know that DM candidates must be more massive than neutrinos, which stay relativistic throughout the expansion of the universe. Moreover, because DM is undetected in current astronomical observations we know that it can only interact via the weak and gravitational interaction [33]. The remaining $\sim 73\%$ is accounted for by *Dark Energy*, which is responsible for accelerating the expansion of the universe. Little is known on the general properties of Dark Energy, and if particles are in fact responsible for it, their nature remains a mystery.

Matter / Anti-Matter Asymmetry Lepton and baryon number are conserved in the SM, meaning that any SM process must contain the same number of (leptons – anti-leptons) and of (baryons – anti-baryons) in the initial and final state. This implies that at the Big Bang a similar amount of matter and anti-matter particles were created, and we should expect them to remain in equal proportions at all times. This is of course in contradiction with what we currently see in our universe, which is comprised of matter and not anti-matter. Therefore, there should be some mechanism acting in the early universe that violates baryon and lepton number, in order to explain the prevalence of matter in the universe. The matter/anti-matter asymmetry is also related to CP-violation, which is a well-known feature of weak interactions. However the amount of

CP-violation allowed in the SM is not enough to account for the apparent asymmetry, a fact that also hints at the existence of BSM physics [34].

The Hierarchy Problem As mentioned in Section 1.5.4, the Higgs mass is a free parameter of the SM that is to be determined experimentally. Yet the Higgs mass is subject to higher-order corrections from quantum loops in the Higgs boson propagation, such as the ones illustrated in Fig 1.10, which imply:

$$m_h^2 = m_{h,0}^2 + \Delta m_h^2, \quad (1.46)$$

where $m_h^2 = 2\lambda v^2 = (125 \text{ GeV})^2$ corresponds to the measured Higgs mass, and $m_{h,0}^2$ is what's known as the *bare* mass. The correction term Δm_h^2 from a fermion f takes the form:

$$\Delta m_h^2 = -\frac{y_f}{16\pi^2} \left[2\Lambda^2 + \mathcal{O} \left(m_f^2 \ln \left(\frac{\Lambda}{m_f} \right) \right) \right], \quad (1.47)$$

where Λ is a cut-off energy scale at which the SM no longer gives a valid description of nature. If no other physics exists up to a cut-off scale equal to the Planck mass $\Lambda \sim M_P \approx 10^{18} \text{ GeV}$ where gravitational effects become relevant, the correction terms of Eq. (1.47) become huge. In that case however, the fact that the observed Higgs mass is $m_h = 125 \text{ GeV}$ implies an exceptionally precise cancellation between $m_{h,0}^2$ and Δm_h^2 . This high level of *fine-tuning* between the bare mass and the corrections hints that new physics hides in an energy scale lower than the Planck mass.

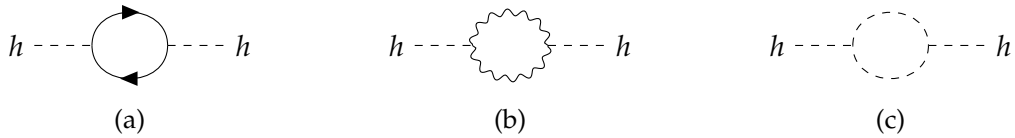


Figure 1.10: Examples of quantum loop corrections to the Higgs boson self energy from a fermion (a), a boson (b), and a scalar (c).

Neutrino Masses We already saw that the masses of fermions in the SM are generated through the Yukawa interaction. As Eq. (1.43) illustrates, Yukawa terms in the Lagrangian couple between the scalar field and the left-handed and right-handed components of the fermion fields. Because the SM does not contain right-handed neutrino fields, the neutrinos remain massless even after the spontaneous symmetry breaking. Yet the observation of *neutrino oscillations* [35] contradicts the massless neutrino hypothesis. Neutrino Oscillations is a phenomena in which neutrinos (supposedly) change

their flavor between $(\nu_e, \nu_\mu, \nu_\tau)$. It is described using the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix, which is a unitary matrix that relates between the flavor basis of neutrinos (in which they have a definite flavor) and the mass basis (in which they have a definite mass). I will not go into the details of this mechanism, which is explained in e.g. [36]. I will simply state that in order for this mechanism to work, it requires that at least two of the three types of neutrinos have different, non-zero masses. Hence, an extension is required for the SM to contain non-zero mass terms for the neutrinos.

LHC Phenomenology

This chapter provides an introduction to the physics relevant for particle collider experiments, and particularly the LHC. Section 2.1 aims to answer practical questions such as what happens when two protons are brought to collide in very high energies? What observables can we use to test our current theories? and what can we learn from the data collected? Section 2.2 discusses specific aspects of Higgs physics, such as the main production modes and most promising decay channels to study at the LHC.

2.1 | Collider Physics

The collisions in the LHC take place between highly energetic protons. A proton is the lightest type of Baryon, composed of three valence quarks of flavors uud . Since the proton is a composite particle, the actual interactions in the LHC occur between its parton (quark and gluon) constituents. This fact complicates the analysis of a pp collision compared to say, a collision between two elementary particles such as e^+e^- .

2.1.1 | Proton-Proton Collisions

When a proton is accelerated to gain a certain momentum, it is its constituents that carry this net momentum. We can thus write the momenta of the partons in a hard scatter:

$$p_1^\mu = x_1 P_1^\mu \quad , \quad p_2^\mu = x_2 P_2^\mu, \quad (2.1)$$

where x_i is the fraction of the total proton momentum P_i^μ carried by the scattering partons. The momenta fractions x_1 and x_2 are distributed probabilistically and indepen-

dently. Or, in other words, the probability of finding momentum fraction x_1 in one proton is independent of what happens in the other proton. This probability depends on the energy transfer in the collision Q , and is given by the *Parton Distribution Functions* (PDFs) $f_{q/p}(x_1, Q^2)$. In order to determine the PDFs at a certain Q^2 we must rely on experimental measurements, such as the deep-inelastic scattering data obtained at the Hadron-Electron ring Accelerator (HERA) at DESY, Hamburg [37]. From there the evolution of the PDFs with Q^2 can be calculated using the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [38–40]. Figure 2.1 shows the PDFs at two different energy scales obtained by the NNPDF collaboration [41]. The x -axis in the plot represents the fraction x of the proton momentum carried by a parton, either a quark or a gluon. Note that the y -axis portrays the product of $xf_{q/p}(x, Q^2)$, also known as *parton momentum distributions*, which is a measure of the probability to find a parton with flavor q carrying a momentum fraction x , when probing the proton at an energy scale Q . The colored curves illustrate how this probability (i.e. the PDFs) change with x for gluons and quarks of different flavor. One can immediately see from the figure that in addition to the valence quarks, marked by u_v and d_v in the plot, a substantial fraction of the proton momentum is contained in the gluons and sea quarks. Moreover, as we probe the proton at higher Q , the valence quarks tend to carry a smaller fraction of the proton momentum.

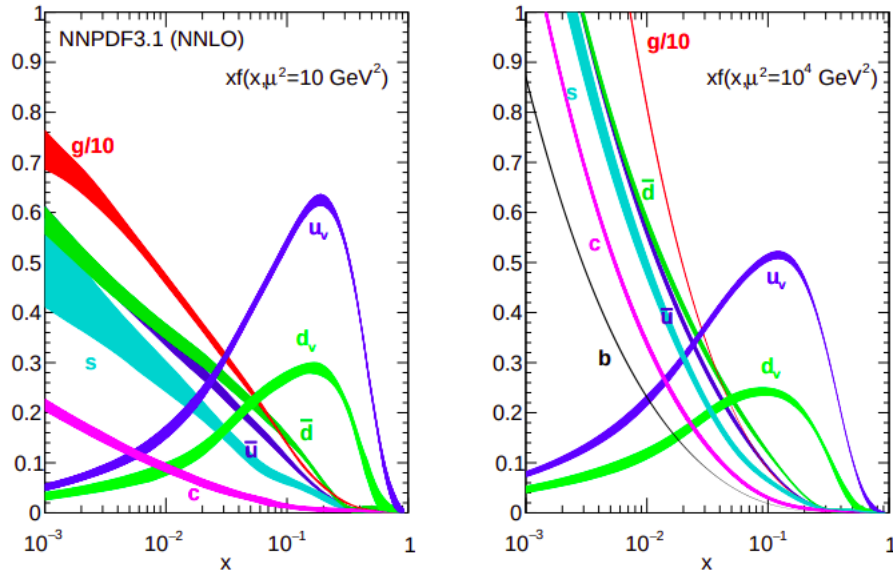


Figure 2.1: Examples of Parton Distribution Functions obtained in NNLO NNPDF3.1 global analysis at energy scales $Q^2 = 10 \text{ GeV}^2$ (left) and $Q^2 = 1 \text{ TeV}^2$ (right).

With the above discussion in mind, we can start making predictions on the outcome

of pp collisions in the LHC. One observable we will be interested in is the cross-section σ , which can be thought of as a measure for the probability of a given process to occur. We will say more about σ later in the section, but for now we will just state that it depends on the particles involved in the process and their kinematic properties. The expression for the total cross-section in pp collisions is thus given by:

$$\sigma_{pp \rightarrow X} = \sum_{i,j} \int_0^1 dx_1 f_{q_i/p_1}(x_1, \mu_F^2) \int_0^1 dx_2 f_{q_j/p_2}(x_2, \mu_F^2) \hat{\sigma}_{q_i q_j \rightarrow X}(x_1 p_1, x_2 p_2, \mu_F^2), \quad (2.2)$$

where the sum runs on partons of types i, j that exist inside the proton and contribute to the hard scatter process, which is described by the parton-level cross-section $\hat{\sigma}_{q_i q_j \rightarrow X}(x_1 p_1, x_2 p_2, \mu_F^2)$. Both the PDFs and the parton-level cross-section are given for an energy transfer equal to the *factorization scale* $Q^2 = \mu_F^2$. The factorization scale defines the cut-off scale below which we use the non-perturbative PDFs to determine the initial conditions of the partons participating in the hard process. The regime above the factorization scale is governed by short-distance processes that can be calculated using perturbation theory in QFT. The fact that these two regimes factorize and can be separated from each other is crucial to being able to compute anything at all using perturbation theory in the LHC.

2.1.2 | Collider Kinematics

In the basic picture of a pp collision at the LHC, the two protons arrive in the interaction point with back-to-back momenta, conventionally taken to be along the $\hat{z}$ direction. Naively one could think that in such collisions the center-of-mass frame and the laboratory (or detector) frame coincide. However, as we already concluded, the hard scatter interactions we are interested in take place between the partons inside the proton. We have also seen that the momentum carried by each of the colliding partons is determined probabilistically and independently from each other. We, as experimental physicists, measure and have access only to the properties in the lab frame. But what we are interested in, is the physics in the collision (parton) center-of-mass frame, which is boosted along $\hat{z}$ with respect to the lab frame by some

$$\beta = \frac{|\vec{p}_1 + \vec{p}_2|}{E_1 + E_2} = \frac{|x_1 - x_2|}{x_1 + x_2}, \quad (2.3)$$

where $x_{1/2}$ are the momentum fraction introduced in Eq. (2.1). β is random and unknown on an event-by-event basis. Thus in order to relate the two reference frames, we choose to work with Lorentz-invariant variables.

The momentum in the transverse direction, $\vec{p}_T = (p_x, p_y)$ is not affected by the boost, and so is the azimuthal angle $\phi = \arctan\left(\frac{p_x}{p_y}\right)$. However the energy E and polar angle θ are not boost invariant, so instead we define the transverse energy $E_T = \sqrt{p_T^2 + m^2}$ and rapidity $y = \frac{1}{2} \ln \frac{E+p_z}{E-p_z}$. Note that y by itself is not boost-invariant and transforms as:

$$y \rightarrow y' = y + \frac{1}{2} \ln \frac{1+\beta}{1-\beta}, \quad (2.4)$$

but the difference in rapidity between two objects (particles) is boost-invariant. In the ultra-relativistic limit, which holds for all the objects we measure, we have $E_T \rightarrow p_T$, and $y \rightarrow \ln \cot \frac{\theta}{2} \equiv \eta$. We prefer to use η (which is called *pseudo-rapidity*) instead of y as it is simpler to measure, easily related to θ and preserves the property that $\Delta\eta$ is invariant under a boost. We can also define the angular separation between any two objects as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. We can now describe a particle measured in the detector using boost-invariant kinematic variables:

$$p^\mu = (p_T \cosh \eta, p_T \cos \phi, p_T \sin \phi, p_T \sinh \eta). \quad (2.5)$$

Using our definition of Eq. (2.5) we can estimate the invariant mass of two particles as:

$$m_{12}^2 = (E_1 + E_2)^2 - (\vec{p}_1 + \vec{p}_2)^2 = 2p_{T1}p_{T2} [\cosh(\eta_1 - \eta_2) - \cos(\phi_1 - \phi_2)]. \quad (2.6)$$

This will be particularly handy when trying to reconstruct the decay of a heavy particle, such as a vector boson, a Higgs boson, or even a postulated new particle. This is due to the fact that decays of heavy particles are associated with a resonant excess in the invariant mass distribution, following the shape of the Breit-Wigner distribution [42].

Another standard variable we use is *missing transverse momentum*, defined as $\vec{p}_T^{\text{miss}} = -\sum \vec{p}_T$. To understand its role consider that in the initial state the net momentum in the transverse direction is zero, and so we expect to have a total of $\sum \vec{p}_T = 0$ in the final state as well. But if that is not the case and $\vec{p}_T^{\text{miss}} \neq 0$, it can point to the existence of a neutrino in the event, as neutrinos rarely interact with matter and do not leave a signal in our detectors.

2.1.3 | Luminosity

As mentioned above, in particle physics, the cross-section σ is often used as a measure for the probability of different processes. The value of σ for a given process can be calculated from its Feynman diagrams. It will depend on the details of the process, such as the particles involved, their momenta, and the type of interaction. σ is typically measured in units of *barn*, with $1 \text{ b} = 10^{-28} \text{ m}^2$. But for a given process or phenomenon, the cross-section alone is not enough to determine the probability of witnessing it in a collider experiment. For that, the concept of *Luminosity*, $\mathcal{L}$, is introduced. $\mathcal{L}$ is defined through the rate R of events of some type and the cross-section for them to happen:

$$R = \mathcal{L} \cdot \sigma. \quad (2.7)$$

One can then deduce the total number of events in a given time interval δt simply by integrating the above:

$$N = \sigma \int_t^{t+\delta t} \mathcal{L} dt, \quad (2.8)$$

which introduces the concept of integrated luminosity $L = \int \mathcal{L} dt$. Note that $\mathcal{L}$ (and L) is something we can control in an experiment, whereas σ is determined by the underlying physics, i.e. a property of "nature". Knowing the integrated luminosity is key to making theory interpretations of an experimental event count. When L is known, we can extract information on the cross-section of a given process from the number of events in which the process was observed. Conversely, when searching for new phenomena with some unknown cross-section, a higher integrated luminosity increases the chances of discovery. L is typically given in units of fb^{-1} .

2.2 | Higgs Physics at the LHC

To study the Higgs boson at the LHC one must of course consider the theoretical aspects discussed in Section 1.5. In addition, the fact that in the LHC the initial state particles are hadrons also has a direct implication on some of our considerations and choices. These are manifested both in the frequent ways the Higgs is produced at the LHC, and in the decay channels we search for it.

2.2.1 | Main Higgs Boson Production Mechanisms

The four most dominant production modes of the Higgs boson from pp collisions at the LHC are: gluon-gluon fusion (ggF), vector boson fusion (VBF), associated production with a vector boson (VH), and associated production with heavy quarks ($t\bar{t}H$ and $b\bar{b}H$). The leading order Feynman diagrams for these processes are illustrated in Figure 2.2.

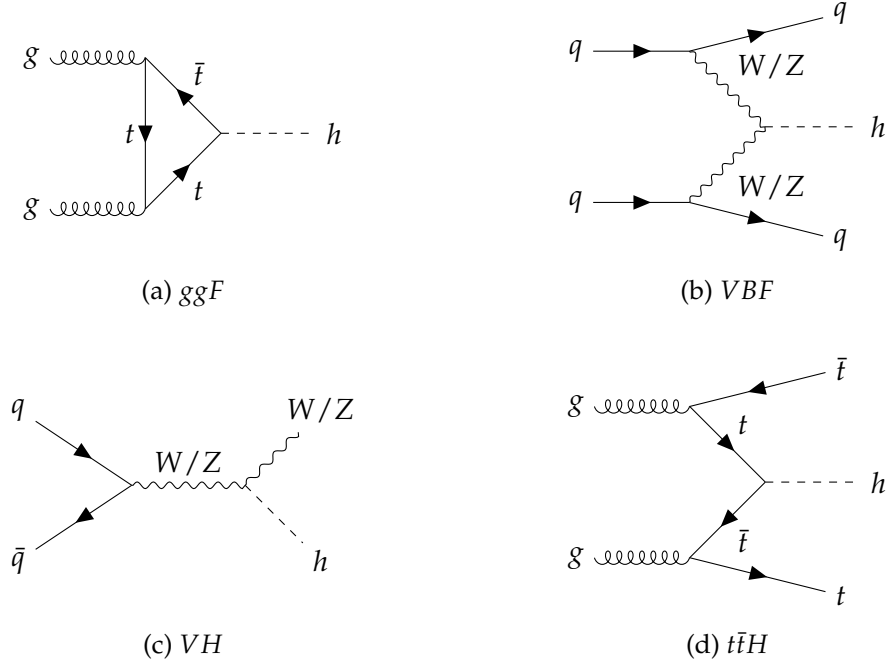


Figure 2.2: Feynman diagrams for the main production modes of the Higgs boson at the LHC: (a) gluon-gluon fusion, (b) vector boson fusion, (c) associated production with a vector boson, (d) associated production with a $t\bar{t}$ pair.

In the following, a brief description of the dominant Higgs production mechanisms is presented. Figure 2.3 shows the cross-section of these production modes as a function of the center-of-mass energy at the LHC for a Higgs boson with mass $m_H = 125$ GeV. The cross-section values quoted in the discussion below, as well as the plot presented in Figure 2.3 are based on the work of [43].

gluon-gluon Fusion The most dominant production mechanism of the Higgs at the LHC is via the ggF process, where two merging gluons create a triangle loop of virtual quarks, resulting in the creation of a Higgs boson. The large momentum fraction carried by gluons in pp collision at the 13 TeV center-of-mass energy, is the reason this is the dominant production mode, with a cross-section of about $\sigma_{ggF} \approx 44$ pb, despite it being generated via a loop. Due to the large Yukawa coupling of the Higgs to heavy quarks,

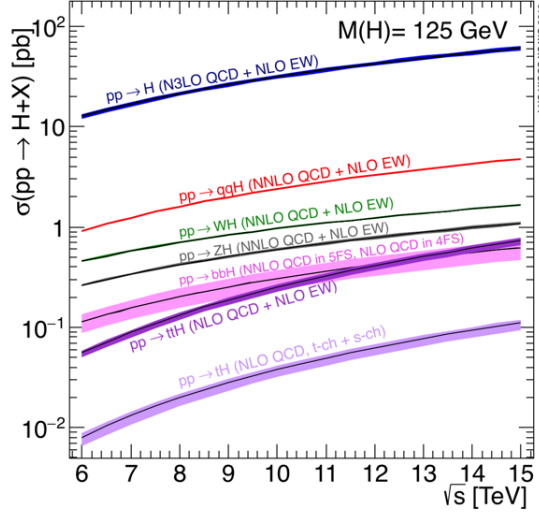


Figure 2.3: Cross-section as a function of pp collision center-of-mass energy for the four dominant Higgs production modes at the LHC.

the contribution to the virtual loop comes mainly from top quarks, and to a lesser extent from bottom quarks, while contributions from 1st and 2nd generation quarks are highly suppressed. Note that out of the four main production modes ggF is the only one where, in leading order, the final state contains only a Higgs boson with no additional particles.

Vector Boson Fusion The second leading production mode, VBF , occurs when a scattering of two quarks is mediated by a pair of $W^\pm$ or Z boson, which then merge and create a Higgs boson. Therefore, this production mode can be used to probe the coupling of the Higgs to vector bosons. The topology of the VBF event is typically characterized by the existence of two additional jets in the forward and backward regions of the detector, initiated from the scattered quarks. The cross-section for this process is almost an order of magnitude smaller than for ggF with $\sigma_{VBF} \approx 4$ pb.

Associated Production with a Vector Boson The creation of a Higgs together with a vector boson (sometimes also referred to as *Higgsstrahlung*) is of particular importance to the work outlined in this thesis, which relies on this production mode to search for the decay of $H \rightarrow c\bar{c}$.

The VH process is initiated from a quark and anti-quark pair which interact to form a virtual off-shell $W^\pm$ or Z boson, which in turn goes on-shell by radiating a Higgs boson. For the case where the vector boson is a Z , additional Feynman diagrams contribute to

the total cross-section, coming from $gg \rightarrow ZH$ through top-quark loops, as shown in Figure 2.4

The VH production is the third most prominent production mode at the LHC with cross-sections of $\sigma_{WH} \approx 1.4$ pb and $\sigma_{ZH} \approx 0.9$ pb.

For a VH process the final state comprises of a Higgs and a vector boson with back-to-back momenta in the transverse plane. By selecting events where the W/Z decays leptonically, one can greatly reduce the huge QCD background, making it possible to target the decay of the Higgs to b - or c -quarks.

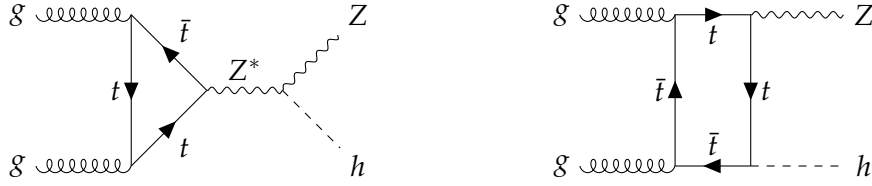


Figure 2.4: Feynman diagrams for the $gg \rightarrow ZH$ contribution to the total VH production mode.

Associated Production with Heavy Quarks The production mode where the Higgs is created together with heavy quarks, $t\bar{t}H$ and $b\bar{b}H$ is the fourth most dominant at the LHC. It is initiated by two gluons that each split into a pair of $t\bar{t}$ or $b\bar{b}$ quarks. Two of the quarks then merge to create the Higgs boson, while the other two are emitted and also reach the detector. The cross-sections for these production mechanisms are $\sigma_{t\bar{t}H} \approx 0.51$ pb and $\sigma_{b\bar{b}H} \approx 0.48$ pb. The $t\bar{t}H$ process offers the best sensitivity for measuring the top Yukawa coupling.

2.2.2 | Higgs Boson Decay

The total width of the Higgs boson is $\Gamma_h = 4.1$ MeV, which translates to a mean lifetime of $\tau_h = 1.6 \times 10^{-22}$ secs [28]. Therefore a Higgs boson produced in the LHC decays very close to the interaction point and is only indirectly observed through its decay products.

As discussed in Section 1.5, the Higgs coupling to gauge bosons and fermions is proportional to their mass, hence the Higgs will most often decay to the heaviest particles, as long the phase-space allows it. Table 2.1 shows the decay branching ratios of a $m_H = 125$ GeV Higgs boson [44].

The dominant decay mode is $H \rightarrow b\bar{b}$, as the bottom quark is the heaviest particle in the SM to obey the condition $m_H > 2m_b$. The decay into heavier particles such as W and

Decay Channel	Branching Ratio	Rel. Uncertainty
$H \rightarrow b\bar{b}$	5.77×10^{-1}	+3.2% -3.3%
$H \rightarrow W^+W^-$	2.15×10^{-1}	+4.3% -4.2%
$H \rightarrow gg$	8.57×10^{-2}	+10.2% -10.0%
$H \rightarrow \tau^+\tau^-$	6.32×10^{-2}	+5.7% -5.7%
$H \rightarrow c\bar{c}$	2.91×10^{-2}	+12.1% -12.2%
$H \rightarrow ZZ$	2.64×10^{-2}	+4.3% -4.2%
$H \rightarrow \gamma\gamma$	2.28×10^{-3}	+5.0% -4.9%
$H \rightarrow Z\gamma$	1.54×10^{-3}	+9.0% -8.8%
$H \rightarrow \mu^+\mu^-$	2.19×10^{-4}	+6.0% -5.9%

Table 2.1: The Higgs boson decay branching ratios and their relative uncertainties, for a SM Higgs boson with mass $m_H = 125$ GeV [44].

Z bosons is suppressed by the available phase-space of a $m_H = 125$ GeV, and can only take place as long as one of the bosons is off-shell. Despite the fact that the Higgs doesn't couple directly to massless particles, Table 2.1 shows non-zero branching ratios to a pair of photons/gluons. This is due to second-order diagrams where the Higgs couples to the heavy particles (mainly top-quark or W boson) such as displayed in Figure 2.5.

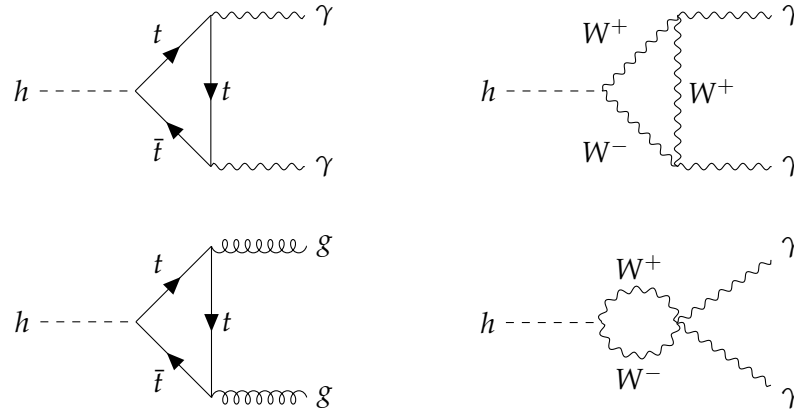


Figure 2.5: Feynman diagrams for the decay of the Higgs to $\gamma\gamma$ and gg mediated by top-quark and W boson loops.

From an experimental point of view, there are drawbacks to observing the three leading decay channels in the LHC. The $H \rightarrow b\bar{b}$ decay channel suffers from large QCD background due to top-quarks decaying to bottom quarks, or QCD production of $b\bar{b}$ pairs through gluon splitting. A similar, but a more severe problem occurs for $H \rightarrow gg$,

as a pair of gluons will have a very similar signature to the soft $2 \rightarrow 2$ QCD interactions which are the most abundant process in the LHC. For the decay to a pair of W bosons, one can in principle overcome the QCD background by searching for leptonic decays of the W . However, since these decays include neutrinos which can not be detected, the mass resolution of the decay products in this channel is rather poor (approximately 20%).

On the other hand, the decay channels in which the Higgs was first observed were $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^*$ with the two Z decaying to four leptons ($ZZ^* \rightarrow 4\ell$). Both of these decay channels have a small branching ratio, but they provide a very clean experimental signature, and thus they are often referred to as "golden channels".

As for the decay of the Higgs to a pair of charm quarks, which is the focus of this thesis, like the other hadronic final states, this channel also suffers from the large QCD background. In addition, the fact that the $H \rightarrow c\bar{c}$ branching ratio is almost 20 times smaller than $H \rightarrow b\bar{b}$ makes it a very challenging process to study at the LHC.

2.2.3 | Higgs Coupling Modifiers - The κ Framework

The compatibility of experimental results with the SM expectation is often expressed within the κ framework [45], where one uses the ratio to the SM prediction rather than absolute coupling. In practice this is done by the introduction of the coupling-strength modifiers defined as $\kappa_F \frac{m_F}{v}$ for a fermion with mass m_F and $\sqrt{\kappa_V} \frac{m_V}{v}$ for a boson with mass m_V . Measuring a κ parameter different from 1 points to deviations from the SM prediction. The measured value of each κ parameter is shown in Figure 2.6 as a function of the mass of the particle. So far no tension has been observed between the measured values and SM expectation.

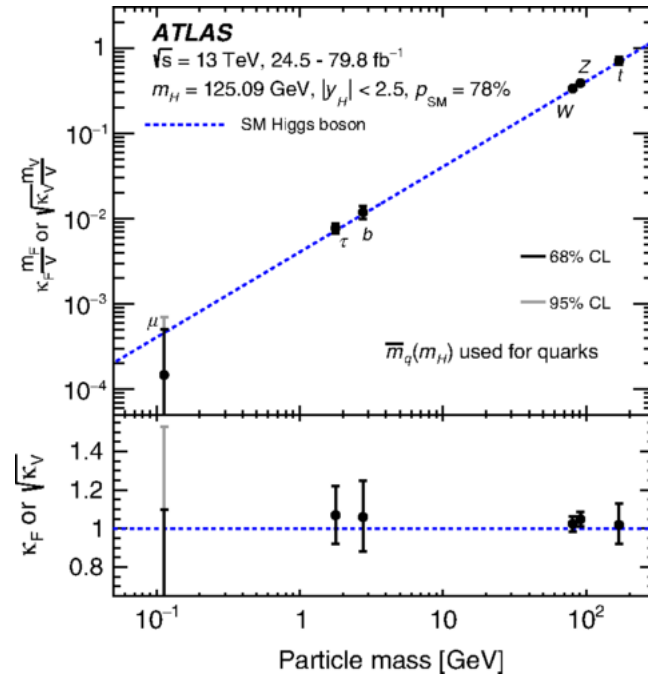


Figure 2.6: Measured Higgs coupling strength modifiers as a function of the particle mass [46]. The vertical axis in the plot is defined differently for fermions and EW bosons. All measured couplings are consistent with the SM expectation, which is represented in the plot by the dashed blue line.

Part II

Experiment

The Large Hadron Collider and the ATLAS Experiment

The work described in this thesis relies on pp collisions data collected by the ATLAS experiment, at the LHC. Section 3.1 includes an introduction to the LHC. Various aspects of the ATLAS experiment, such as the detector layout, its sub-systems, and means of operation, are discussed in Section 3.2.

3.1 | Large-Hadron-Collider (LHC)

The Large Hadron Collider (LHC) is a high energy pp and heavy ions accelerator operating at the European Organization for Nuclear Research (CERN), located outside of Geneva, Switzerland [1].

3.1.1 | Layout and Operation

The LHC occupies a circular tunnel with a circumference of 27 km and a depth between 45 - 170 meters underground. The tunnel was constructed between 1984 and 1989 and was originally used for the Large Electron-Positron collider (LEP) [47], which operated for eleven years until November 2000 when it was switched off. In early 2001 work began on dismantling the LEP collider and making room for its successor, the LHC.

The LHC collides beams of protons at center-of-mass energy of $\sqrt{s} = 13$ TeV. To bring the protons to such energies, they are accelerated through a number of phases that successively increase their energy. This acceleration chain is based on a series of

lower energy accelerators, which comprise the CERN accelerator complex, illustrated in Figure 3.1.

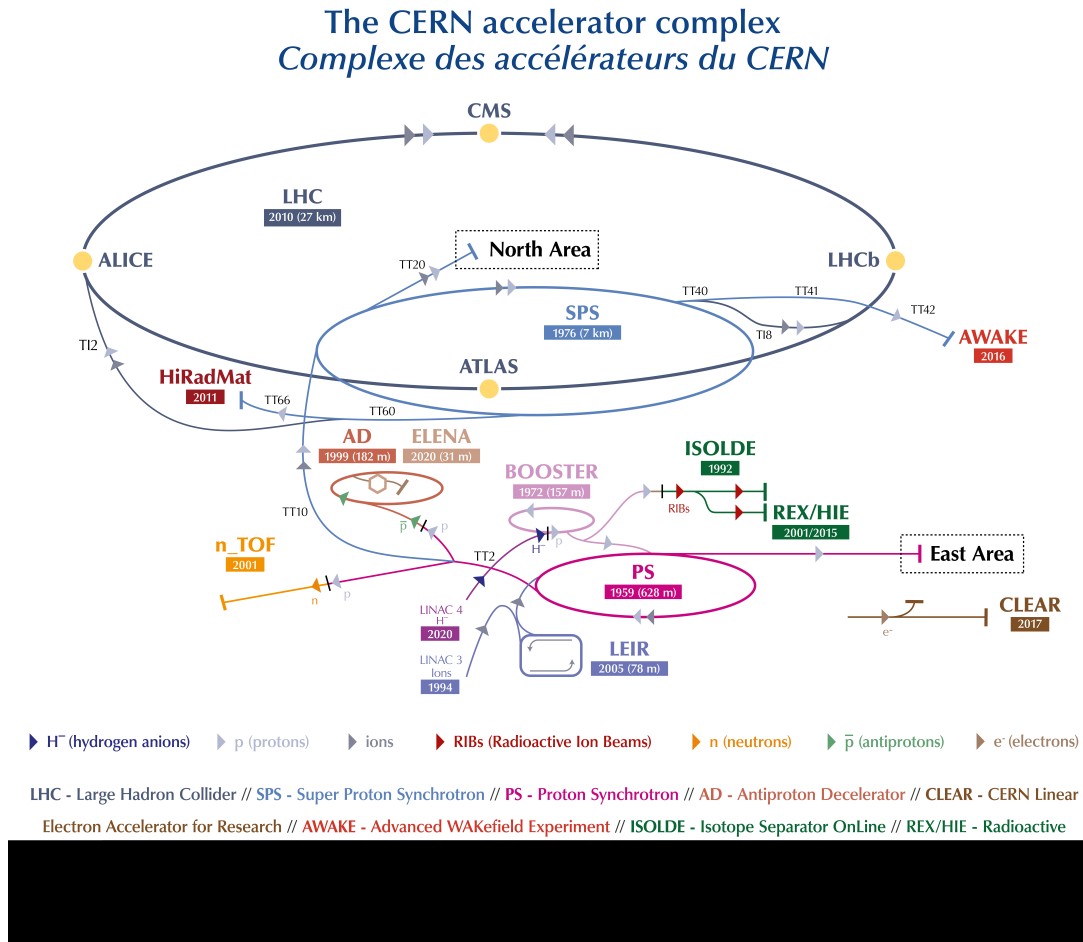


Figure 3.1: The CERN accelerator complex. The LHC (dark blue) is the last ring in a chain of particle accelerators [48].

The acceleration process begins with Hydrogen atoms, which are ionized in an electric field resulting in protons. The protons are sent through the linear accelerator LINAC2, where they are collected into bunches of approximately 1.15×10^{11} protons and accelerated to an energy of 50 MeV. They are then injected into the Proton Synchrotron Booster (PSB), which accelerates them to an energy of 1.4 GeV. Next, the protons reach the Proton Synchrotron (PS) that brings them to an energy of 26 GeV before injecting them into the Super Proton Synchrotron (SPS) where they are accelerated up to 450 GeV. From the SPS the protons are finally injected into the LHC, where they are accelerated to their final energy of 6.5 TeV.

The acceleration inside the LHC is based on superconducting radio-frequency (RF) cavities. Each beam uses 8 RF cavities operating at 4.5 K, providing an acceleration field of 5 MV/m at 400MHz [49]. This results in an acceleration time of around 20 minutes to reach 6.5 TeV. The LHC uses 1232 superconducting dipole magnets in order to bend the beam, and 392 quadrupole magnets for beam focusing. The magnets are made of superconducting coils of Niobium-Titanium (NbTi) alloy that are cooled via super-fluid Helium at 1.4 K. The LHC magnets provide a magnetic field up to 8.3 T, which is required to maintain a proton beam with an energy of 6.5 TeV within the curvature of the tunnel radius [50].

When entering the LHC, the bunches of protons are injected in both directions, clockwise and anti-clockwise. Their paths meet in four different points along the ring, where the collisions occur. Around each collision point, a particle detector is built, set to measure and analyze the collision products. Two of these detectors, ATLAS [2] and CMS [3], are "general purpose" detectors, designed to be sensitive to a wide range of experimental signatures for new physics searches, as well as performing SM precision measurements. Although they share a similar layout, they are based on different detector technologies. The other two experiments are ALICE [51], used mainly for heavy-ion collisions, LHCb [52], designed for detecting bottom and charm mesons in the forward direction to study flavor physics and CP violation.

The LHC started collecting pp data in 2009 with a center of mass energy of $\sqrt{s} = 7$ TeV, which was then increased between 2011-2012 to $\sqrt{s} = 8$ TeV. The data-taking period between 2009-2012 is referred to as Run-1, and it was followed by a long shutdown (LS1) where the LHC machine and the detectors were repaired and upgraded (2013-2015). In 2015 operation resumed for Run-2 with $\sqrt{s} = 13$ TeV. Run-2 ended in 2018 when another shutdown (LS2) started for further maintenance work and upgrades. In the current LHC prospect, illustrated in Figure 3.2, LS2 is planned to end in 2022 when Run-3 (2022-2025) begins. It will be followed by yet another long shutdown (LS3), after which the High-Luminosity LHC (HL-LHC) is set to be commissioned [53].

3.1.2 | Run-2 Performance

In Section 2.1.3 we have touched upon the concept of Luminosity, which can be used to quantify the performance and amount of data collected by a collider. During Run-2, the period in which the data analyzed in this thesis was collected, the LHC provided an integrated luminosity of 156 fb^{-1} , of which 139 fb^{-1} of data was recorded and found suitable for physics analysis in ATLAS. The integrated luminosity as a function of time

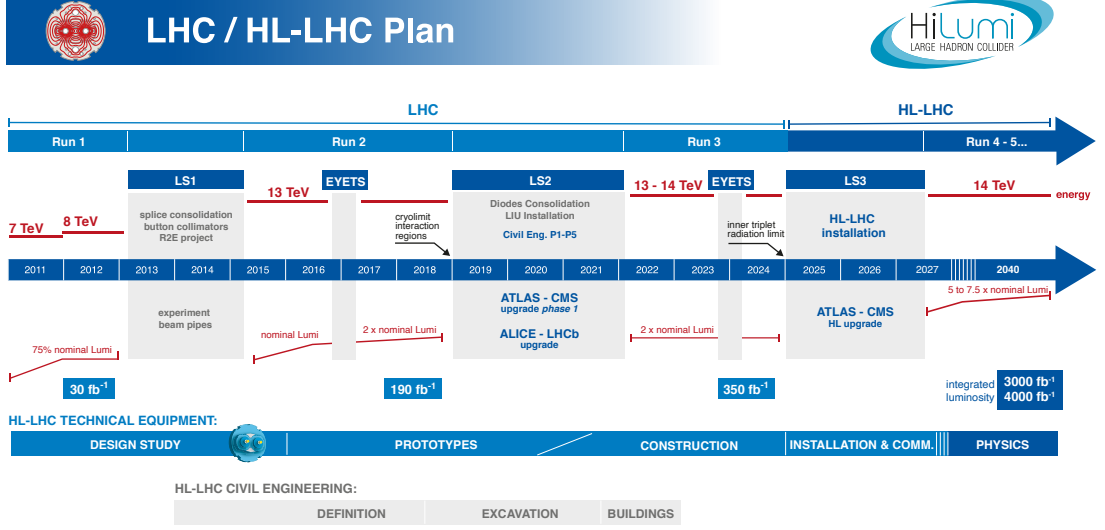


Figure 3.2: Projected plan for the operation of the LHC and HL-LHC from 2011 until 2040 and beyond [53].

is shown in Figure 3.3.

Another quantity that is related to Luminosity is *pileup*. Pileup refers to additional low p_T inelastic collisions, that are difficult to disentangle from the hard scattering pp interactions we are interested in. Pileup can be categorized into *in-time pileup*, resulting from multiple pp interactions within the same bunch crossing, and *out-of-time pileup* caused from interactions in the previous and subsequent bunch crossings. A common way to quantify the effect of pileup is through the average number of interactions per bunch crossing, $\langle \mu \rangle$, which can be calculated as:

$$\mu = \frac{\mathcal{L}_{bunch} \sigma_{inel}}{f}, \quad (3.1)$$

where $\mathcal{L}_{bunch}$ is the instantaneous per bunch luminosity, σ_{inel} the total pp inelastic cross section, corresponding to 80 mb at $\sqrt{s} = 13$ TeV, and f is the revolution frequency of the LHC. The total and per-year distributions of $\langle \mu \rangle$ during Run-2 are shown in Figure 3.4.

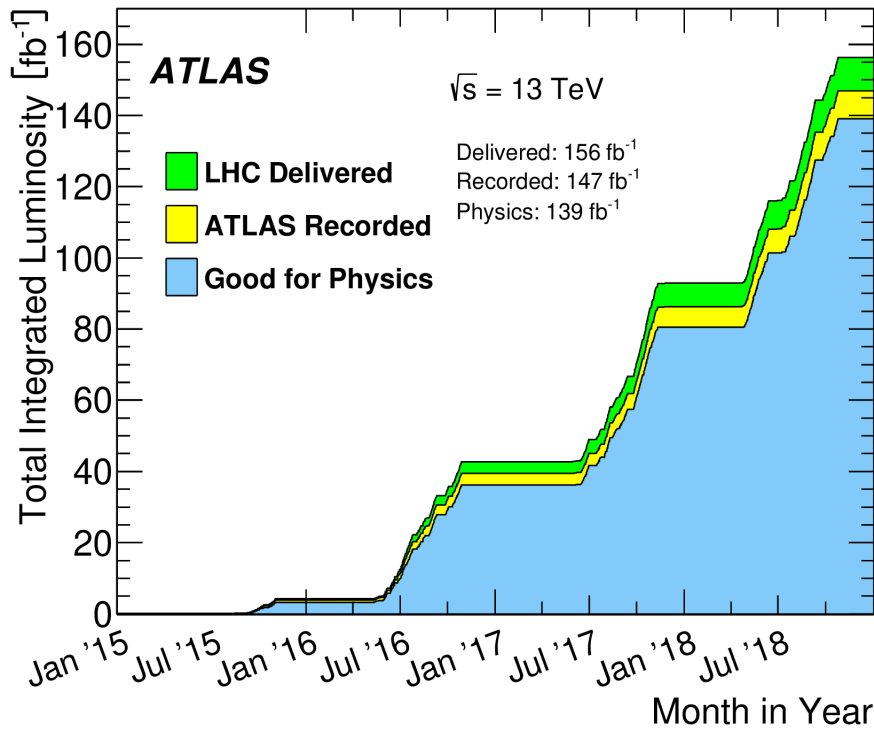


Figure 3.3: Integrated luminosity as a function of time delivered by the LHC (green), collected by ATLAS (yellow), and found suitable for physics analysis (blue).

3.2 | The ATLAS Experiment

The ATLAS (A Toroidal LHC ApparatuS) detector is a general-purpose particle detector located 100 m underground on Interaction Point 1 of the LHC ring. It has a cylindrical shape extending for 45 meters, with a diameter of 25 meters and weighing about 7000 tonnes. A schematic sketch of the ATLAS detector is shown in Figure 3.5. It comprises of several sub-detector systems, built as different coaxial layers around the beam pipe, with the collision point at the center of the detector. Each sub-detector utilizes different techniques to detect and measure the properties of particles moving through it.

The major sub-detector systems in ATLAS are the Inner Detector (ID), the calorimeter system, and the muon spectrometer. The role each of the sub-detector systems plays in the measurement is illustrated in Figure 3.6. Charged particles, like electrons and protons, interact with the material in the ID, while neutral particles pass through it leaving no detectable signs. The calorimeter system uses two techniques to measure the energy deposited by the particles passing through it. Photons and electrons give birth to EM

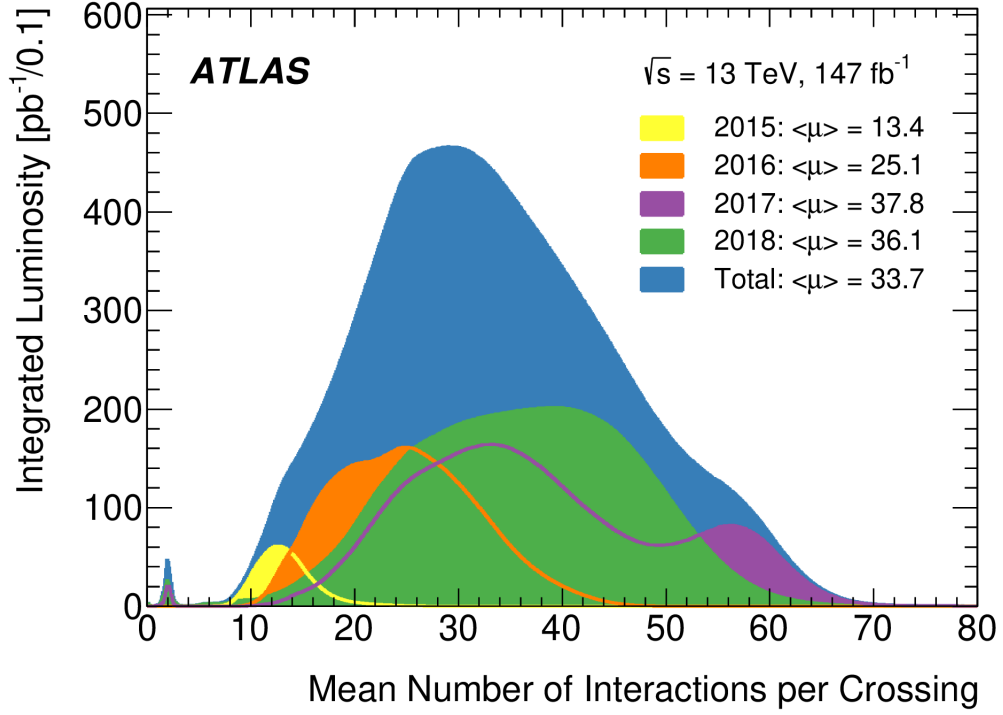


Figure 3.4: Average number interactions per bunch crossing, $\langle\mu\rangle$, per year during Run-2 of the LHC.

showers in the EM calorimeter, and hadrons are detected in the hadronic calorimeter. Muons barely interact with any of the material in the ID and calorimeters. They make it all the way out to the muon spectrometer where they can be detected. Neutrinos escape undetected as they almost do not interact with the material, and their existence is deduced from the missing transverse momentum, as defined in Section 2.1.2.

Before discussing the sub-detector systems in more detail, let us review the coordinate system used in the measurements of all properties.

3.2.1 | Coordinate System

ATLAS uses a right-handed coordinate system with the interaction point (IP) defining the origin, and the beam pipe along the $\hat{z}$ -axis. The $x - y$ plane defines the transverse direction to the beam pipe, with the $\hat{x}$ -axis pointing to the center of the LHC ring and the $\hat{y}$ -axis pointing upwards. The azimuthal angle ϕ is measured as usual around the $\hat{z}$ -axis in the $x - y$ plane while the polar angle θ is defined with respect to the beam axis.

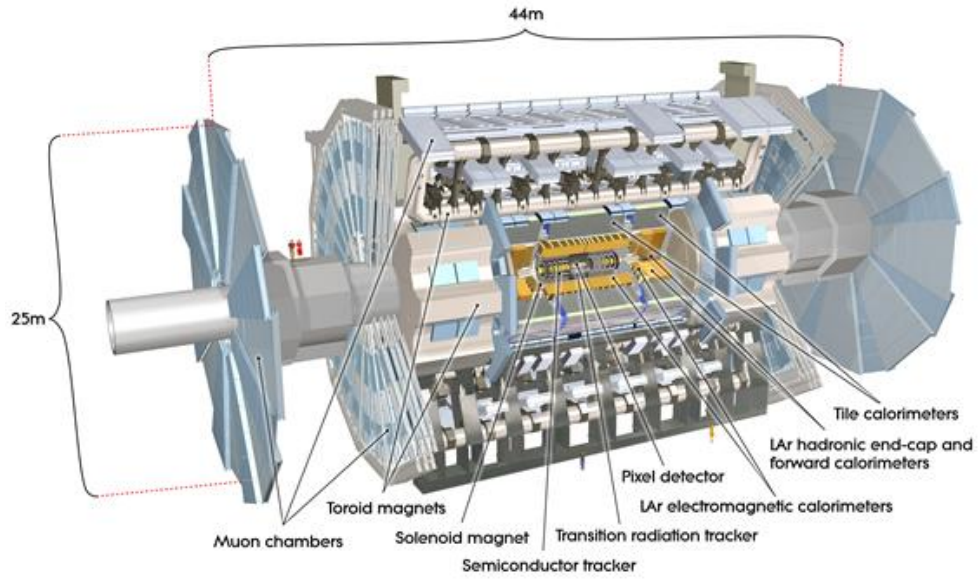


Figure 3.5: A schematic view of the ATLAS detector.

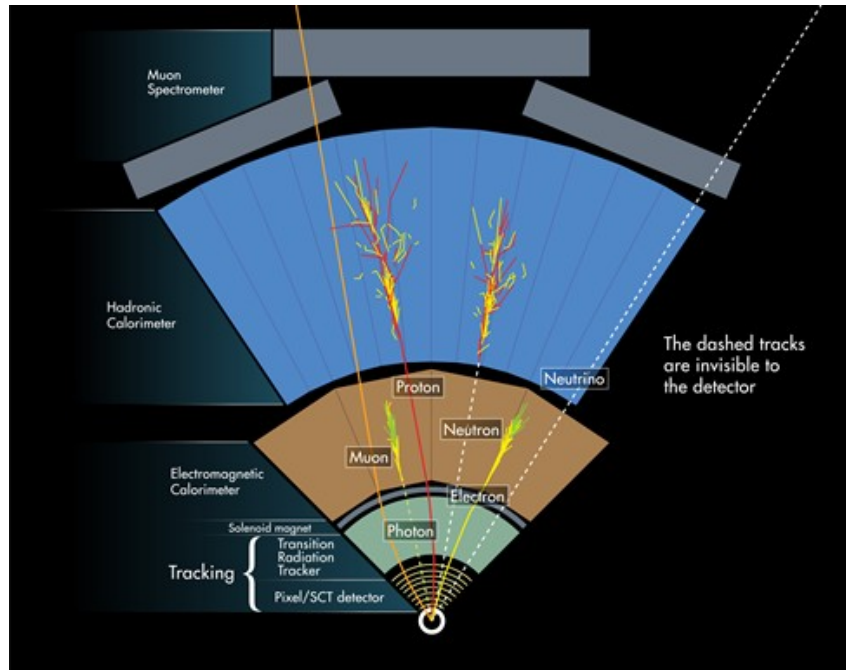


Figure 3.6: Transversal view of the particle paths inside the different sub-systems of the ATLAS detector [54].

Section 2.1.2 explains the importance of working with boost-invariant variables. Hence, let us remind here that the pseudo-rapidity $\eta = \ln \cot \frac{\theta}{2}$ is often used instead

of θ , and that angular distance in the $\phi - \eta$ plane is defined using $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

The complete detector is split into Barrel Region (BR) sections and End-Cap (EC) Regions sections: in the former the detector layers are positioned on cylinders around the beam axis, in the latter, these are positioned in wheels perpendicular to the beam pipe.

3.2.2 | Inner Detector

As the name suggests, the Inner Detector (ID) is the sub-detector located closest to the beamline. It is designed to provide precise track measurements of charged particles, with excellent momentum resolution and primary and secondary vertex identification within the pseudorapidity range $|\eta| \leq 2.5$. The ID comprises of four tracking sub-systems, all immersed in a 2T magnetic field parallel to the beam axis, generated by a superconducting solenoid magnet. Figure 3.7(a) shows a scheme of the ID and Figure 3.7(b) shows radial position of each (barrel) a transverse section of the barrel Inner Detector.

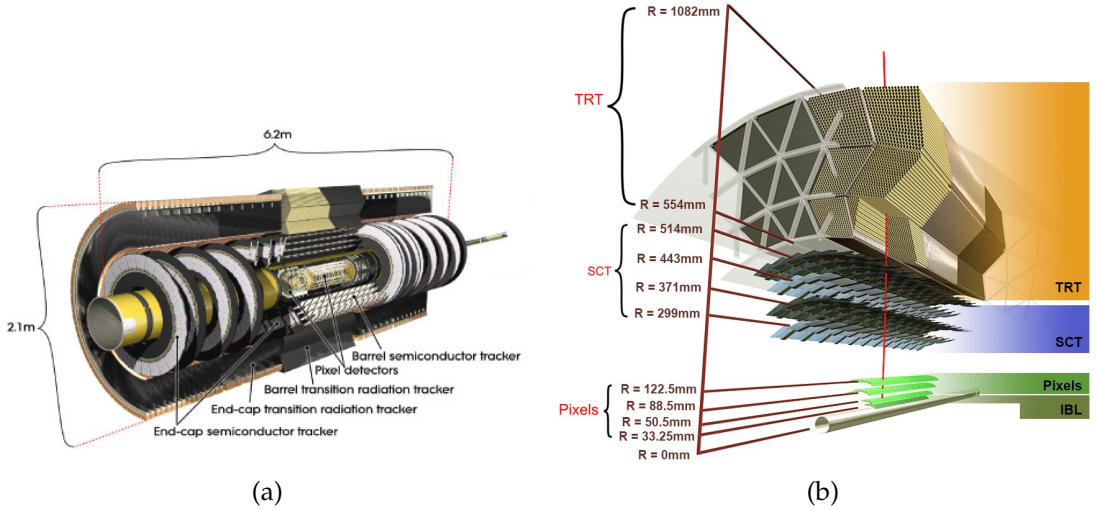


Figure 3.7: (a) Schematic illustration of the ID and its sub-components. (b) The four sub-detector layers and structural element of the barrel ID along the radial direction.

3.2.2.1 | Insertable B-Layer (IBL)

The IBL is the innermost component of the Inner Detector [55]. It was added during LS1 between Run-1 and Run-2 with the aim of providing measurements closest to the interaction point. The IBL plays a critical role in the identification of jets originating from

heavy flavored (i.e. b - and c -) quarks, which relies on precise vertex reconstruction to find the displaced vertex of the b - and c - hadron decays (see Chapter 5 for more details).

The IBL consists of 14 staves instrumented with planar and 3D silicon pixel sensor technology along 332 mm on each side from the center of the ATLAS detector. The working principle of a silicon detector is based on a silicon p-n junction operating in reverse bias voltage. The size of each pixel is $50 \times 250 \mu\text{m}^2$ providing a resolution on the longitudinal impact parameter, z_0 of approximately $200 - 80 \mu\text{m}$ for $0.4 < p_T < 20 \text{ GeV}$.

3.2.2.2 | Pixel Detector

The ATLAS Pixel detector is composed of 1744 pixel modules arranged in three barrel layers and two end-caps with three disks each [56]. The pixels correspond to silicon sensors with a size of $50 \times 400 \mu\text{m}^2$, operating with the same working principle as the sensors of the IBL. In total, the Pixel detector contains approximately 80 million readout channels, each corresponding to a pixel. The resolution of a pixel is $10 \mu\text{m}$ in $r\phi$ and $115 \mu\text{m}$ in z in the barrel layers (r in the end-cap disks).

3.2.2.3 | Semiconductor Tracker (SCT)

Like the IBL and the Pixel detector, the SCT relies on silicon semiconductor technology [57]. It consists of four cylindrical layers in the barrel and nine disk layers in each of the end-caps. The barrel sector has 2112 modules, each comprised of four silicon-strip sensors with strips parallel to the beamline at a pitch of $80 \mu\text{m}$ and thickness of $285 \mu\text{m}$. Two rectangular sensors are paired together and contain a total of 768 strips of approximately 12 cm in length, which provide 2D measurements. An additional pair is glued back-to-back to the first one, rotated at an angle of 40 mrad. This small angle offset allows for the measurement of the third coordinate. The same back-to-back arrangement is used for the 988 modules in each of the end-cap regions. However, the end-cap modules use trapezoid sensors with strip pitch of variable size ($80 \mu\text{m}$ on average), and are mounted in such a way that the strips have oriented radially. The SCT provides a hit resolution of $17 \mu\text{m}$ in $r\phi$ in both the end-cap and barrel. The resolution in z is $580 \mu\text{m}$ in the barrel and similarly, the resolution in r is $580 \mu\text{m}$ in the end-cap.

3.2.2.4 | Transition Radiation Tracker (TRT)

The TRT is the outermost part of the ATLAS Inner Detector [58]. Unlike the IBL, Pixel, and SCT detector which relies on silicon semi-conductor technologies, the TRT is a transition radiation detector that uses gas ionization for particle detection. It is composed of

straw tubes of 4mm in diameter filled with a gas mixture of 70% Xe, 27% CO₂, and 3% O₂. In the center of each drift tube there is a gold-plated tungsten wire of 31 μm diameter. The tungsten wire is kept at ground level potential while the tube walls are kept at -1.5 kV. When a charged particle traverses a straw, it ionizes the gas and produces electrons that drift toward the wire where they are collected and read out. The tubes are embedded in polymer fibers and foils which provide a transition material boundary. In traversing the boundary, a relativistic charged particle may emit radiation with a probability that depends on the relativistic factor $\gamma = \frac{E}{m}$. The transition radiation photons are absorbed in the straw, producing high-threshold signals. Since the emission is much more likely for an electron than for a pion or a muon, it is more probable that an electron track is associated with higher energy hits, providing a good discrimination power for electron identification. The TRT is composed of a barrel and two end-caps covering together the pseudorapidity range $|\eta| \leq 2.0$. The barrel region contains 52544 straw tubes of 1.5 m length oriented in parallel to the beam axis, and covering the radius range from 0.5 m to 1.1 m. Each end-cap contains 122880 straw tubes of length 0.4 m, which are disposed perpendicularly to the beam axis, and covering the range $0.8 \text{ m} < |z| < 2.7 \text{ m}$.

3.2.3 | Calorimeters

The goal of a calorimeter is to fully contain and measure the energy of an incident particle. When particles enter the calorimeter they initiate a shower of secondary particles. The characteristic of the particle shower depends on the type and energy of the incident particles, as well as the properties of the calorimeter. For an accurate energy measurement, it is important that all interacting particles will be fully absorbed in the calorimeter material, and deposit their entire energy.

In general, particle physics calorimeters can be divided into two categories: *sampling calorimeters* and *homogeneous calorimeters*. A sampling calorimeter comprises of alternating layers of a heavy absorber material and an active material. The goal of the absorber is to interact with the traversing particles and gradually decrease their energy by initiating the showers of secondary particles. When the particles pass through and interact with the active material, an electrical signal is formed which can be detected, and is used for measuring the particles energy. In a homogeneous calorimeter, a single active material is used for both the absorption and energy measurement of the particle.

Calorimeters can also be classified by the type of particles they are designed to detect. Electromagnetic calorimeters aim to measure mainly electrons and photons by means of their EM interactions in dense, high Z materials. Hadronic calorimeters are

used for measuring hadrons, which deposit their energy via both nuclear and EM interaction with the calorimeter material.

The calorimeter system in ATLAS is illustrated in Figure 3.8. An EM calorimeter based on liquid Argon with a lead absorber is used in the barrel and end-cap regions. A hadronic calorimeter based on scintillating tiles is used in the barrel, and in the end-caps the hadronic calorimeter comprises of liquid-Argon with Copper as an absorber. In the forward region of the detector (up to $|\eta| \leq 4.9$) an additional forward calorimeter is used. All calorimeters in ATLAS are sampling calorimeters, and they will be described in more detail shortly.

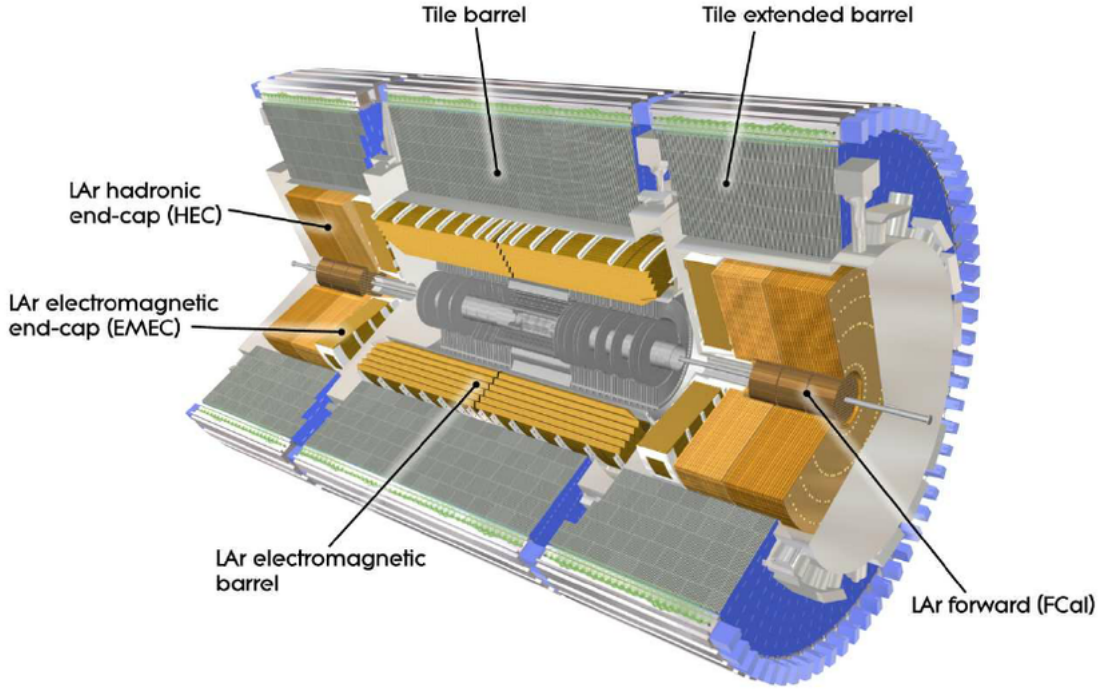


Figure 3.8: A cut-away view of the ATLAS calorimeter system [2].

The energy resolution of a calorimeter can be parameterized as:

$$\frac{\sigma_E}{E} = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C, \quad (3.2)$$

where S is the stochastic term, which accounts for fluctuation in the number of particles produced in the shower. N is the noise term that is related to the pileup and electronic noise, and is independent of the measured energy. The constant term, C , accounts for

systematic effects such as miscalibration and defects or inactive material in the calorimeter. The $\oplus$ sign indicates a quadratic sum between the contributions of the three terms.

3.2.3.1 | EM Calorimeters

The EM calorimeter is designed to measure the energy and position of (mainly) electrons, positrons, and photons. While traversing through the calorimeter material, the $e^\pm/\gamma$ interacts with its atoms. High energy electrons and positrons mainly interact by emitting bremsstrahlung radiation, while for photons the dominant process is e^+e^- pair production. An example of leading order Feynman diagram for these two processes is shown in Figure 3.9. The secondary particles created in these interactions, also interact with the material, such that a cascade of particles - a "shower" - is formed in the direction of the initial particle's motion. One can qualitatively describe this process as if in each step the interacting particle loses some of its energy to the new particles. As the typical energy of the particles in the shower drops, other processes such as ionization and scattering are also becoming relevant. The *critical energy* E_c is defined as the energy level at which the electron cross-section for emitting bremsstrahlung radiation is smaller than that of ionization in the material. E_c depends on the material used, and it sets the energy scale at which the shower stops, as most of the energy loss happens through ionization and no additional EM particles are generated. The measurement of the particle energy is then based on the fact that the energy released in the detector material by the charged particles of the shower, mainly through ionization and excitation, is proportional to the energy of the incident particle.

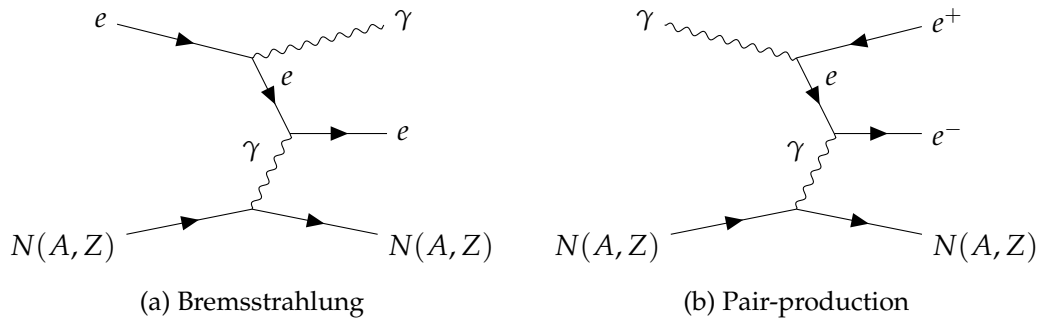


Figure 3.9: Examples for the leading order Feynman diagrams of the two main processes taking place in an EM shower. In (a) Bremsstrahlung radiation is emitted in the interaction between a high-energy electron and a nucleus. (b) shows electron-positron pair-production from a high-energy photon.

EM showers are also typically described by a length scale which is the *radiation length*, X_0 , defined as the mean distance an electron will advance in the material be-

fore losing $\frac{1}{e}$ of its energy through Bremsstrahlung radiation. It is equivalent to $\frac{7}{9}$ of the mean free path for a photon to undergo pair production. X_0 is a function of the density and atomic number Z of the material.

The LAr EM Calorimeter

The EM calorimeter in ATLAS is positioned outside of the solenoid magnet generating the magnetic field for the inner tracking detector. It is a sampling calorimeter composed of layers of lead as an absorber and liquid-Argon (LAr) as the active material [59]. The layers are constructed in an accordion shape to achieve maximal coverage in ϕ while maintaining a relatively fast readout time. The EM calorimeter is divided into a barrel part (EMB) which extends to $|\eta| < 1.475$, and two end-cap components (EMEC) covering the range $1.375 < |\eta| < 3.2$, each nested in its own cryostat. A *pre-sampler* made of a thin layer (11 mm of depth) of instrumented Argon is positioned between the solenoid and the EMB with the purpose of sampling the shower prior to interaction with the calorimeter material, thus compensating for energy losses in the ID and the solenoid. The barrel sector comprises of two half-barrels, each consisting of 1024 accordion shape absorbers interleaved with readout electronics. It extends between $1.5 \text{ m} < r < 1.97 \text{ m}$ and it is segmented into three detection layers:

- The innermost layer (L1), also known as the *strip layer*, with a thickness of up to $4.4X_0$, and segmentation of 0.003×0.1 along the $\eta - \phi$ plane. The high granularity in η plays a critical role in discriminating between π^0 (which decay $\sim 99\%$ of times via $\pi^0 \rightarrow \gamma\gamma$) and prompt photons.
- The middle layer (L2) is where most of the energy in the shower is deposited, with a thickness of $22X_0$. The cells in L2 have a $\eta - \phi$ segmentation of 0.025×0.025 .
- The back layer (L3) is a thin layer of about $2X_0$, and a cell size of $\eta \times \phi = 0.05 \times 0.025$. It is used to measure energy leakage into the hadronic calorimeter.

The EMEC consists of two wheels, one on each side of the EMB. Each wheel contains 768 absorbers interleaved with readout electronics, and extends radially from $30 \text{ cm} < r < 208 \text{ cm}$ and longitudinally between $3.6 \text{ m} < |z| < 4.2 \text{ m}$.

3.2.3.2 | Hadronic Calorimeters

Hadronic calorimeters in high-energy particle experiments are used to measure the energy and position of *jets*. Jets are the experimental signatures of partons. It is a collimated spray of hadrons formed due to the strong force confinement when a colored particle is e.g. "ripped" from a hadron in a hard scatter, or created in a decay of a heavier particle. Jet formation and various properties of jets are discussed in more detail in Section 4.4.

Hadronic energy loss typically occurs through strong nuclear interactions, in which secondary particles - mainly protons, neutrons, $\pi^\pm$ and π^0 - are formed and give rise to hadronic showers. The hadronic showers are typically more complicated and more difficult to model compared to EM showers, for a number of reasons:

- The number of high-energy particles in the hadronic shower is smaller than in an EM shower, which leads to large fluctuations.
- A non-negligible fraction of the initial energy is absorbed as nuclear binding energy and target recoil, and thus remains undetected.
- At every step of the shower approximately 1/3 of the energy goes into an electromagnetic shower, mainly due to production of π^0 which immediately decay to a pair of photons.
- Due to the invisible energy component in hadronic showers, the detection efficiency of a calorimeter for EM energy depositions is usually better than for hadronic energy depositions. This means that either the energy scale of the hadronic calorimeter must be independently determined, or the calorimeter must be built specially to compensate for this lost signal.

Similarly to X_0 in the EM shower, the hadronic shower is also characterized by a typical interaction length, λ , which is the mean free path required to reduce by a factor of $\frac{1}{e}$ the number of relativistic hadrons in the medium. The hadronic interaction length depends on the absorbing material with $\lambda \approx 35A^{\frac{1}{3}} \text{ g cm}^{-2}$, and is typically $\mathcal{O}(10)$ times longer than X_0 , requiring hadronic calorimeters to be thick compared to EM calorimeters. The typical cutoff energy of the hadronic shower is the pion production threshold, below which most of the energy is deposited through ionization of the material.

The ATLAS Hadronic Calorimeter

The ATLAS hadronic calorimeter is located outside of the EM calorimeter. It is a non-compensating calorimeter that makes use of different technologies depending on the respective detector region. The *Tile calorimeter* uses steel as an absorber and plastic scintillating tiles as active material [60]. The tiles are oriented radially with respect to the beam pipe. A charged particle traversing the tile will excite its atoms, resulting in scintillation photons from the de-excitation of the atoms. These photons are collected by wave-length shifting fibers and then measured by photomultiplier tubes. The Tile calorimeter is composed of a barrel section 5.8 m in length and covering up to $|\eta| < 1$ and two extended barrels in the range $0.8 < |\eta| < 1.7$, with 2.8 m in length each. Both systems have an inner radius of 2.28 m and an outer radius of 4.25 m, which amounts to an average radial depth of 7.4λ .

In the end-cap region, the hadronic *end-cap calorimeter* (HEC) is based on LAr as active material with copper as an absorber [59]. It comprise of two wheels each consisting of a series of copper plates and 8.5 mm LAr gaps. In the front wheel there are 24 copper plates 25 mm thick each, and in the rear wheel there are 16 plates each 50 mm thick. The HEC covers the pseudorapidity range of $1.5 < |\eta| < 3.2$ and has a depth of approximately 12λ .

Located in the same cryostats as the HEC, an additional *forward calorimeter* (FCal) is used to cover the range of $3.1 < |\eta| < 4.9$ [61]. The FCal is segmented into three 45 cm layers. The first layer is designed to measure electromagnetic particles, using Copper as an absorber. The outer two layers are hadronic layers with tungsten plates as an absorbing material. In all three detection layers, the absorber plates are placed in between LAr gaps which function as the active material. The FCal has a total depth of approximately 10λ and $200X_0$.

3.2.4 | Muon Spectrometer

Muons are *minimum-ionizing particles* (MIPs) and thus are expected to pass through the Inner Detector and Calorimeter systems without being completely absorbed in the material. Therefore the Muon Spectrometer (MS) is the outermost detector system in ATLAS, with the implied assumption that muons are the only particles that can reach that far from the interaction point. Aside from measuring the muon momenta and trajectory, the MS also participates in the ATLAS trigger system and therefore utilizes two types

of detector technologies: fast detectors for the trigger system and precise detectors to measure the muon spatial position accurately. A schematic overview of the muon spectrometer including the different detection sub-systems employed is presented in Figure 3.10. The barrel part consists of 3 layers at 5 m, 7.5 m, and 10 m from the beamline, and is divided into 16 azimuthal sections. The two end-caps are built of three wheels on each side at 7.4 m, 14 m, and 21.5 m from the interaction point. To determine the muons momentum, a magnetic field is required to bend their trajectory and measure their tracks. Three large air-core superconducting toroid magnets, one in the barrel and one in each of the end-caps, supply the MS with a magnetic field of approximately 0.5 T in the region of $|\eta| < 2.7$.

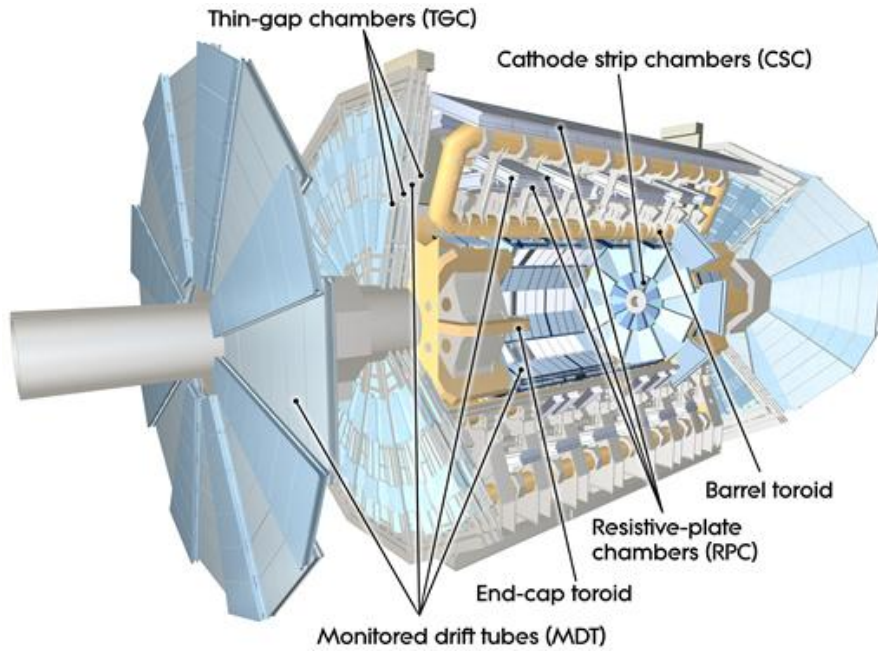


Figure 3.10: Schematic view of the ATLAS Muon Spectrometer. The different sub-detector systems are indicated [62].

3.2.4.1 | Monitored Drift Tubes

The Monitored Drift Tubes (MDT) are used for precision tracking measurements in both the barrel and end-cap regions. The MDT consists of three to eight layers of aluminum tubes, 30 mm in diameter, filled with a mixture of 93% Ar and 7% CO₂ at a pressure of 3 bar. A single MDT module has a resolution of 80 μm , but a chamber consisting of a few layers of tubes reaches a resolution of 35 μm .

3.2.4.2 | Cathode Strip Chambers

While the MDT performs well in the central region of the detector, the high-pressure requirements of the Ar gas make them inadequate for regions of high particle multiplicity. Therefore in the forward $2 < |\eta| < 2.7$ region, Cathode Strip Chambers (CSC) complement the precision measurements of tracks. The CSC are multi-wire proportional chambers composed of radial anode wires and cathode strips perpendicular to the wires. The resolution of the CSC is $40\ \mu\text{m}$ in the bending plane ($\hat{r} - \hat{z}$) and $5\ \mu\text{m}$ in the non-bending plane.

3.2.4.3 | Resistive Plate Chambers

The Resistive Plate Chambers (RPC) provide triggers in the barrel region of $|\eta| < 1.05$. Two resistive plates, made of phenolic-melamine plastic laminate, are kept parallel to each other at a distance of 2 mm by insulating spacers and a layer of gas. An electric field between the plates allows avalanches to form along the ionizing tracks towards one of the plates, from which the signal is readout.

3.2.4.4 | Thin Gap Chambers

Triggering in the end-cap ($1.05 < |\eta| < 2.4$) region is provided by the Thin Gap Chambers (TGC). These are multi-wire proportional chambers filled with a gas mixture of CO_2 and nC_5H_{12} (n-pentane). Inside each chamber, there are anode wires made of gold plated tungsten, and cathode strips which are perpendicular to the wires. When a charged particle passes through the gas, it ionizes it and creates a negative (positive) signal in the wires (strips).

3.2.5 | The Trigger and Data Acquisition System

Given the high instantaneous luminosity of the LHC, the trigger and data acquisition (TDAQ) system play a crucial role in deciding whether an event is kept or not for offline physics analysis. With a bunch spacing of 25 ns, ATLAS generates data at a rate of 40 MHz, corresponding to $\mathcal{O}(1\text{TB}/\text{sec})$. This is vastly more data than can be written to disk and the majority of the data will contain information of no interest. The trigger system decides, in real-time, whether to record or reject the data of a collision. The triggered and recorded information is called an event. By selecting events passing predefined hardware and software criteria, the recording rate is reduced to $\sim 1\ \text{kHz}$. There are several channels of potential trigger chains which run in parallel. For instance, events with deposited energy above some threshold may be saved for further analysis. In a

specific data analysis, the triggers are assigned in order to select the events that are interesting with respect to the phenomena one is interested in.

The ATLAS trigger system is composed of a hardware-based *Level-1 Trigger* and a software-based *High-Level-Trigger*:

Level-1 Trigger

The first step in the trigger, Level-1 (L1) is based on hardware. It considers inputs from the Calorimeter and Muon Spectrometer systems and identifies regions-of-interests (ROIs) in the detectors with high p_T physic objects. The physics objects are classified depending on the origin of the incoming signal to jets, electrons/photons, muons, and missing transverse energy (E_T^{miss}). The information in the ROI is then passed on to the next trigger level. After the L1 selection the rate is reduced to a maximum of 100 kHz.

High-Level Trigger

The next step in the trigger is the High-Level Trigger (HLT) which only considers an event that passes L1. The HLT is a computing system connected by fast dedicated networks. Sophisticated selection algorithms are applied to the ROIs of L1, and to additional tracking information from the ID, within an average processing time of 200 ms. The output rate is reduced in the HLT from the 100 kHz of L1, to an output rate of approximately 1kHz. All events passing the HLT are written to disk.

Reconstruction of Physics Objects

In Section 3.2 we introduced the various sub-system comprising the ATLAS detector and discussed some of their technical and mechanical aspects. We also mentioned which sub-system is responsible to detect and identify which type of particles (see for example Figure 3.6). In this chapter, we will explain in more details how the electronic signals collected from the detector are used to define the physics objects later studied in physics analyses. Section 4.1 explains how the primary vertex in the collision is determined and how charged particles tracks are identified. This is the basis for the reconstruction of all physics objects. We then focus on the procedures used for building the individual particles, starting with electrons and photons in Section 4.2, followed by muons in Section 4.3, the reconstruction of jets formed in the hadronization of quarks and gluons is described in Section 4.4, τ -leptons in Section 4.5, and missing transverse momentum, which is helpful for determining the presence of a neutrino in the event, in Section 4.6.

Aside from photons and τ -leptons, the VH(cc) analysis makes use of all physics objects presented in this chapter. Yet here a more general description is given, and additional details related to the definitions and choices within the VH(cc) analysis can be found in Chapter 8

4.1 | Tracks and Primary Vertices

Tracks represent the curved trajectories of charged particles as they move in the presence of a magnetic field. They are crucial for measuring the charge and momentum of charged particles, which can be deduced from the curvature of the track. In addition, measuring the track with good accuracy also contributes to identifying the position of

interaction vertices, and suppressing signals from soft interactions and pileup.

In ATLAS, tracks are reconstructed from energy depositions (referred to as *hits*) in the layers of the Inner Detector [63]. The main tracking procedure starts from an *inside-out* strategy, where the track is built from hits in the innermost pixel layer. This is done by finding *track seeds*, created as sets of three space points in the pixel, strip, or in a combination of both silicon detectors. A combinatorial Kalman filter [64] is then used to combine track seeds together and form the track. This however, results in a number of ambiguous track candidates, due to an incorrect assignment of the hits or overlapping tracks with shared hits. To solve these ambiguities, each track candidate is assigned a score based on various requirements including, momentum, number of hits, their origin (most precise hits coming from the Pixel layer have a higher score), and the presence of holes (defined as a missing hit on an active module where one was expected based on the particle trajectory). Track candidates with a score below some threshold are discarded, while highly ranked tracks are then extended and combined with the measurements of the TRT.

After the inside-out tracking is complete, a reversed *outside-in* approach follows with the goal of reconstructing tracks that may have missed some hits in the inner silicon layers. Here the tracking is initiated by TRT track segments in regions seeded from the EM calorimeter. These are then matched with hits in the Pixel and SCT layers which were not assigned in the inside-out phase. TRT segments without any silicon hit are also kept and used for tasks as photon conversion reconstruction.

Finally, the resulting tracks from both inside-out and outside-in strategies form the final collection of reconstructed tracks. These tracks are described using a reference point (usually taken to be the *beamspot position*, defined as the average position of the pp interactions) and five track parameters:

- The polar angle θ and the azimuthal angle ϕ .
- The *transverse impact parameter* d_0 , defined as the closest point in the transverse plane to the reference position, and the *longitudinal impact parameter* z_0 , defined as the closest point in the longitudinal plane to the reference position.
- The ratio $\frac{q}{p}$ between the measured charge and momentum of the track.

Once tracks are created, *primary vertices* (PV), defined as the position where the hard scattering between the protons occurred, can be determined. As a result of pileup con-

ditions, several interaction vertices can be reconstructed in each bunch-crossing. The PV is defined as the vertex with the largest collision energy, and knowledge of its position is fundamental for a physics analysis.

The PV reconstruction procedure in ATLAS involves two steps: *vertex finding* and *vertex fitting*. It is explained in detail in [65] and can be summarized by the following steps:

- A set of tracks satisfying the following selection criteria is defined:
 - $p_T > 400$ MeV, $|\eta| < 2.5$, $|d_0| < 4$ mm, $\sigma(d_0) < 5$ mm, $\sigma(z_0) < 10$ mm
 - Number of SCT hits ≥ 4 , number of silicon (SCT or Pixel) hits ≥ 9 , no Pixel holes
 - If $|\eta| > 1.65$ silicon hits ≥ 11 and SCT holes ≤ 1
 - IBL hits + B-layer (closest pixel layer) hits ≥ 1
 - No more than 1 shared pixel hit or 2 shared SCT hits

where $\sigma(d_0)$ and $\sigma(z_0)$ denote the uncertainties on d_0 and z_0 .

- A seed position for the first vertex is selected. The transverse position of the first seed is taken as the center of the beam spot. The z-coordinate of the seed is the mode of the z-coordinates of tracks at their respective points of closest approach to the reconstructed center of the beam spot.
- The tracks and the seed are used to estimate the best vertex position with a fit. The fit is an iterative procedure where in each iteration less compatible tracks are down-weighted and the vertex position is recomputed.
- After the vertex position is determined, tracks that are incompatible with the vertex are removed from it and allowed to be used in the determination of another vertex.
- The procedure is repeated with the remaining tracks in the event.

All vertices with at least two associated tracks are retained as primary vertex candidates. The number of such candidates in the event is labeled as N_{PV} and can indicate the pileup conditions and activity within the event. The primary vertex, corresponding to

the hard scatter interaction, is the candidate with the highest sum of squared transverse momenta from the associated tracks, calculated as:

$$\sum_{\text{track} \in PV} p_T^2. \quad (4.1)$$

All other reconstructed vertices are considered as pileup vertices.

4.2 | Electrons and Photons

The reconstruction of both electrons¹ and photons relies on clusters of energy depositions in the EM calorimeter, due to the similar signature of the EM showers they produce. In the case of an electron one can expect to find one or more ID tracks which can be associated with the EM cluster, and therefore used to distinguish it from a photon, for which the only signal should be that of the calorimeter. Yet the picture is somewhat more complex due to processes such as the emission of bremsstrahlung photons from an electron or photon conversion to an electron-positron pair inside the ID material.

4.2.1 | Reconstruction

The procedure of electron and photon reconstruction [66] starts from clusters of energy deposits measured in topologically connected EM and hadronic calorimeter cells, denoted *topo-clusters*. The scheme by which topo-cluster are defined is explained later in Section 4.4.2.1 when discussing the reconstruction of jets. For electron and photon reconstruction, these topo-clusters are matched to ID tracks which are re-fitted using a Gaussian-Sum-Filter algorithm [67] to account for energy losses through bremsstrahlung. The track-cluster matching is done by extrapolating the track position into the middle layer of the EM calorimeter, and requiring $|\Delta\eta| < 0.5$ and $-10 < q \cdot (\phi_{\text{track}} - \phi_{\text{clus}}) < 0.05$, with q referring to the reconstructed charge of the track. The asymmetric requirement on $q \cdot (\phi_{\text{track}} - \phi_{\text{clus}})$ accounts for cases when tracks miss some energy radiated from photons that clusters measure. The algorithm also builds photon conversion vertices from two opposite-charge tracks forming a vertex consistent with that of a massless particle, or single tracks without any hits in the innermost sensitive layers. The conversion vertices are also matched to the selected topo-clusters by applying requirements on $\Delta\eta$ and $\Delta\phi$, which vary depending on the position of the tracks and if it consists of silicon/TRT hits.

¹Unless stated otherwise, throughout this discussion I will not make a distinction between e^- and e^+ and use the term "electrons" to refer to both

The next step is the reconstruction of *superclusters*, which proceeds independently for photons and electrons in two stages for each, and is illustrated in Figure 4.1: in the first, the topo-clusters are iterated one by one, in order of descending E_T , and are tested to be superclusters seeds. An electron supercluster seed is required to have $E_T > 1$ GeV and must be matched to a track with at least four silicon hits. A photon supercluster seed must fulfill $E_T > 1.5$ GeV without any requirements on track or conversion vertex matching. The next stage is the supercluster *satellite* finding, based on a $\Delta\eta \times \Delta\phi$ window around the seed barycenter. For both electrons and photons, a topocluster is considered a satellite if it falls within a range of $\Delta\eta \times \Delta\phi = 0.075 \times 0.125$, as these cases tend to represent secondary EM showers originating from the same initial electron or photon. For electrons only, satellites are also considered if they fall within a broader range of $\Delta\eta \times \Delta\phi = 0.125 \times 0.300$ from the seed. For photons with conversion vertices made up only of tracks containing silicon hits, a cluster is added as a satellite if its best-matched (electron) track belongs to the conversion vertex matched to the seed cluster. Superclusters are formed by all seed clusters and their associated satellites. The final step involves associating the EM calorimeter cells to a given supercluster.

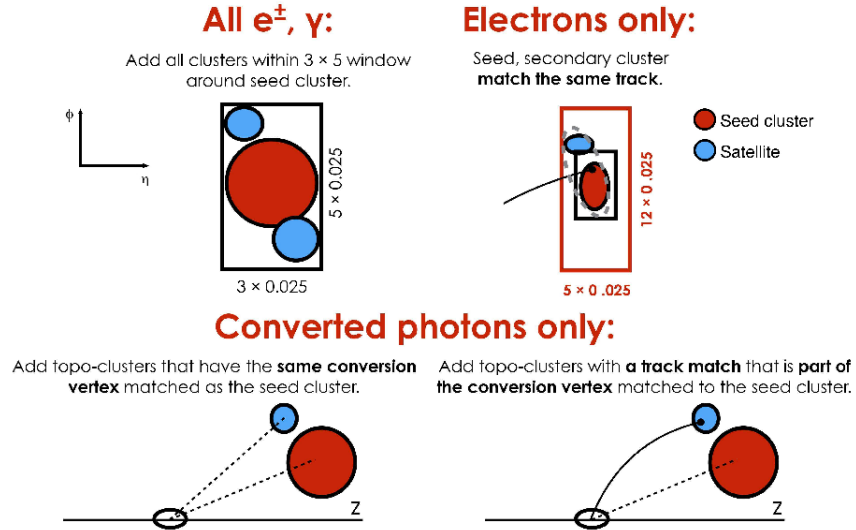


Figure 4.1: Illustration of the superclustering algorithm for electrons and photons. Seed clusters are shown in red, satellite clusters in blue [66].

4.2.2 | Identification

Different identification and quality requirements are defined in order to improve the quality of electrons and photons, and to reduce the background arising from semilep-

tonic decays of heavy hadrons - in the case of electrons, or $\pi^0 \rightarrow \gamma\gamma$ decays and hadronic jets - in the case of photons.

The identification of electrons relies on a likelihood discriminant constructed from quantities based on the ID track and lateral/longitudinal development of the EM shower [68]. Samples of MC simulation of di-jet, $Z \rightarrow e^+e^-$, and $W \rightarrow ev_e$ are used to determine likelihood function and extract electron efficiencies in bins of E_T and η . Based on these efficiencies, three working points are defined as *Tight*, *Medium*, and *Loose*, corresponding to 98%, 90%, and 80% efficiency, respectively.

Photon identification is based on a set of cuts on several discriminating variables related to the longitudinal and lateral shower development in the EM calorimeter, and the shower leakage to the hadronic calorimeter [69]. These variables take advantage of the fact that prompt photons tend to produce a narrower shower and smaller leakage to the HCal compared to jets, and that $\pi^0 \rightarrow \gamma\gamma$ decays are often characterized by two local energy maxima in the first calorimeter layer. The photon identification criteria are optimized using MC simulation of γ +jet, $\gamma\gamma$, $H \rightarrow \gamma\gamma$ as well as sample of background photons in QCD di-jets events. Two working points are defined as *Loose* and *Tight* corresponding to identification efficiencies of 97% and 85%, respectively for photons with $E_T > 20$ GeV.

4.2.3 | Isolation

The typical signature of a prompt electron is characterized by a small activity in the $\Delta\eta \times \Delta\phi$ region around the electron. Therefore, to further improve the rejection of background from photon conversions and fake electrons, requirements on the isolation of electron candidates are made using several calorimeter-based and track-based isolation variables [68]. Two such variables, relevant for selecting the electrons used in the analysis described later in this thesis, are the calorimeter-based $TopoE_T^{\text{cone20}}$, defined by summing the transverse energy from topo-clusters within a cone of size $\Delta R = 0.2$ around the candidate electron cluster, excluding the contribution from pileup:

$$TopoE_T^{\text{cone20}} = \sum_{c \in \text{cone}} E_T^c - E_T^{\text{clusters}} - E_T^{\text{pileup}}, \quad (4.2)$$

and the track based $p_T^{\text{varcone20}}$, constructed from summing the transverse momenta of tracks with $p_T > 1$ GeV, within a cone of $\Delta R = \min\left(\frac{10\text{GeV}}{p_T}, 0.2\right)$ around the electron candidate trajectory:

$$p_T^{\text{varcone20}} = \sum_{t \in \text{cone}} p_T^t - p_T^{\text{tracks}}. \quad (4.3)$$

Based on the isolation variables, isolation efficiencies are computed using simulations samples enriched with $Z \rightarrow e^+e^-$ and $J/\psi \rightarrow e^+e^-$, and are used to define different operating points for electron isolation. For the VH(cc) analysis described in this thesis, the two relevant working points are [70]:

FCLoose: A combination of calorimeter- and track-based isolation targeted to keep a value of 99% isolation efficiency by requiring both $TopoE_T^{\text{cone20}}/p_T < 0.2$ and $p_T^{\text{varcone20}}/p_T < 0.15$.

HighPtCaloOnly: A tighter, calorimeter-based isolation criteria targeted to keep 95% isolation efficiency by requiring $TopoE_T^{\text{cone20}} < \max(0.015 \cdot p_T, 3.5\text{GeV})$

Background photons are usually accompanied by nearby hadronic activity, whereas prompt photons are well isolated. Similar calorimeter-based and track-based isolation criteria as described above for electrons are also used to determine photon efficiency working points [69].

4.2.4 | Calibration

Electrons and photons are also calibrated to account for differences between their energy in MC simulation and actual data. The calibration is based on a multivariate algorithm trained on MC, a residual correction for non-uniformities in the calorimeter response, and adjustment for the data/MC agreement using $Z \rightarrow e^+e^-$ and $Z \rightarrow l^+l^-\gamma$ events. The corrections are applied on the MC simulation [71].

4.3 | Muons

Muons are reconstructed from track hits in the ID and in the Muon Spectrometer. The track reconstruction is first performed separately in the two sub-systems, and the tracks are later combined to construct muon candidates.

4.3.1 | Reconstruction

The reconstruction procedure of muon tracks in the ID follows the same methodology described in Section 4.1 used for all other tracks. Muon reconstruction in the MS starts with a search for hit patterns inside each muon chamber to form segments [72]. Precision measurements in the MDT and CSC provide information on the muon trajectory

in the bending plane, while the RPC and TGC hits measure the coordinate orthogonal to the bending plane. Muon track candidates are then built by fitting together hits from segments in different layers. The combination starts from segments generated in the middle layers of the detector, and is then extended to the inner and outer layers. At least two matching segments are required to build a muon candidate track, except in the transition region between the end-cap and barrel, where a single high-quality segment can be used to build a track. The hits associated with each track candidate are fitted using a global χ^2 fit, and candidates with a bad score above some threshold are discarded, or re-fitted excluding the hits providing large contributions to the χ^2 .

The selected MS tracks are then combined with ID tracks using algorithms that consider information provided by the ID, MS, and calorimeters. Four muon categories are defined depending on the sub-detectors used for reconstruction:

- **Combined (CB) muons** : a combined track is formed with a global fit that uses the hits from both the ID and MS sub-detectors. During the global fit procedure, MS hits may be added to or removed from the track to improve the fit quality.
- **Segment-tagged (ST) muons** : a track in the ID is classified as a muon if, once extrapolated to the MS, it is associated with at least one local track segment in the MDT or CSC chambers. ST muons are used when muons cross only one layer of MS chambers, either because of their low p_T or because they fall in regions with reduced MS acceptance.
- **Calorimeter-tagged (CT) muons** : a track in the ID is identified as a muon if it can be matched to an energy deposit in the calorimeter compatible with a minimum-ionizing particle. These muons are useful to recover muon acceptance in regions of the detector where the MS is only partially instrumented due to cabling and services for the ID and the calorimeter.
- **Extrapolated (CT) muons** : the muon trajectory is reconstructed based only on the MS track and a loose requirement on compatibility with originating from the interaction point. These muons are mainly intended to extend the muon reconstruction acceptance to the $2.5 < |\eta| < 2.7$ region not covered by the ID.

4.3.2 | Identification

After reconstruction, identification of prompt muons is performed to distinguish them from background muons, mainly arising from π and K decays. Several variables can

be constructed to suppress background muons and impose quality requirements on the track in the ID and MS. For CB tracks, the variables used in muon identification are the normalized χ^2 of the combined track fit, the q/p *significance*, defined as the absolute value of the difference between the ratio of the charge and momentum of the muons measured in the ID and MS divided by the sum in quadrature of the corresponding uncertainties, and ρ' , defined as the absolute value of the difference between the p_T measurements in the ID and MS divided by the p_T of the combined track [72]. Based on these quantities four muon operating points are defined as *Loose*, *Medium*, *Tight*, and *High- p_T* , with identification efficiencies that are determined in $Z \rightarrow \mu^+\mu^-$ MC events, and ranging from 98.1% (*Loose*) to 89.9% (*Tight*). The *High- p_T* working point is specially designed to improve the momentum resolution for muon tracks above 100 GeV.

4.3.3 | Isolation

Prompt muons from the decay of a W , Z , or a Higgs boson are well isolated, whereas background muons from semi-leptonic decays of heavy-flavored hadrons are typically characterized by near-by hadronic activity. To distinguish between them, isolation criteria are applied to muons based on the same variables used for electrons and described in Section 4.2.3. The relevant muon isolation operating points for the VH(cc) analysis described in this thesis are [70]:

- **FixedCutLoose:** A combination of calorimeter- and track-based isolation targeted to keep a value of 99% muon isolation efficiency by requiring both $TopoE_T^{\text{cone20}}/p_T < 0.3$ and $p_T^{\text{varcone30}}/p_T < 0.15$. Note that in contrast to Eq. (4.3), the track-isolation variable in this case refers to a maximum cone size of $\Delta R = 0.3$.
- **HighPtTrackOnly:** A tighter, track-based only isolation criteria targeted to keep 95% muon isolation efficiency by requiring $p_T^{\text{cone20}} < 1.25$ GeV, with p_T^{cone20} defined similarly to Eq. (4.3) in a cone with a fixed radius of $\Delta R = 0.2$.

The efficiencies for the isolation working points are determined in data and simulation of $Z \rightarrow \mu^+\mu^-$ and $J/\psi \rightarrow \mu^+\mu^-$ events.

4.3.4 | Calibration

Muon momentum calibration is used to apply the corrections to the simulated muon p_T reconstructed in the ID and MS sub-detectors, to precisely describe the measurement of

the same quantities in data. The ID and MS correction parameters are extracted from data using a binned maximum-likelihood fit with templates derived from simulation, comparing the invariant mass distribution for $Z \rightarrow \mu^+ \mu^-$ and $J/\psi \rightarrow \mu^+ \mu^-$ candidate events [72].

4.4 | Jets

The most abundant particles created in the LHC collisions are quarks and gluons. However, due to QCD confinement states with a color charge different from zero are not observed in nature. As a consequence, partons can not be detected on their own, and the experimental signature of quarks and gluons in the detector is an object called a *jet*. Jets make up a remarkably rich topic, from the theoretical models describing their formation, through the reconstruction algorithms used to define them, and the measurement and association of their energy. In the following discussion I will not manage to touch upon all the concepts related to jets, but only summarize some of their key features. One particular topic that is omitted from this chapter is that of *flavor-tagging*. Flavor-tagging refers to the identification of jets initiated by heavy-favored quarks, mainly b –, but also c –quarks. Naturally, flavor-tagging is of major importance to the analysis described in this thesis, which targets decays of the Higgs to charm quarks. Therefore, Chapter 5 will be dedicated solely to the topic of flavor-tagging.

4.4.1 | Jet Formation

To illustrate the mechanism in which jets are formed, let us start with a single parton, traveling along some direction. Even though, as already mentioned, stable states of single partons do not exist in nature, we can think of this parton as just been produced in a hard interaction, or resulting from the decay of a heavier particle. The parton has a probability of emitting another parton, which is given by [73]:

$$dS = \frac{2\alpha_S C_F}{\pi} \frac{dE}{E} \frac{d\theta}{\theta} \frac{d\phi}{2\pi}, \quad (4.4)$$

with C_F is the QCD color factor corresponding to $C_F = \frac{1}{2}$ for $q \rightarrow qg$, $C_F = \frac{4}{3}$ for $g \rightarrow qq$ and $C_F = 3$ for $g \rightarrow gg$ splitting. Equation (4.4) tells us that gluons tend to shower more than quarks, and also that the probability of emitting a parton diverges in two cases:

- **Infrared (soft) divergences** for emission of low energy partons with $E \rightarrow 0$

- **Collinear divergences** for partons emitted in small angles $\theta \rightarrow 0$ with respect to the original parton direction.

This implies the parton will mostly radiate soft partons along the axis of its motion. The radiated gluons and quarks continue to iteratively emit more partons, and the result is a process known as *parton branching*, which is the first step in jet-formation. It should be noted, that the parton branching is a perturbative procedure, which can be re-summed to all orders of the perturbation series, under certain assumptions.

When the parton shower has evolved long enough, the energy of the partons is reduced, such that low momentum transfers occur, and the phase of *hadronization* begins. During the hadronization, partons are combined into color-singlet states, thus forming the hadrons. The hadronization procedure is non-perturbative and can only be described using phenomenological models, such as the Lund string model [74] or the cluster model [75]. Nevertheless, the *local parton-hadron duality* [76] ensures that the flow of quantum numbers at the hadron-level follows approximately the corresponding flow at parton-level.

4.4.2 | Jet Reconstruction

A jet is not a uniquely defined object, but rather the output of a clustering algorithm which groups together jet constituents according to their kinematic properties. In ATLAS there are two main approaches for jet building, which differ in the jet constituents considered as input to the clustering algorithms. In the first approach, EMTopo jets use topologically connected clusters of energy depositions in the calorimeter cells (topo-clusters, similar to those mentioned in Section 4.2.1 in the context of electron reconstruction) as the jet building blocks. More recently, an approach called *Particle-flow* (PFlow) has been developed where both calorimeter and track information is used to form the constituents of the jet [77]. In the following discussion I will concentrate on EMTopo jets, as this is the class used in the VH(cc) analysis described in this thesis.

4.4.2.1 | Topo-clustering

The first step in building EMTopo jets is combining signals from close-by calorimeter cells into an object called a topo-cluster. A topo-cluster is seeded by a cell with an energy deposit which is significantly above the noise level. In the *420 scheme* used in ATLAS during Run-2, the requirement for the seeding cell was that energy deposited is at least 4 standard deviation (4σ) above the (electronic+pileup) noise distribution. Once seeded,

the topo-cluster is extended by adding all neighbor cells with a significance of at least 2σ . Finally, the boundary of the topo-cluster is determined by adding all nearest neighbor cells (equivalent to a threshold of 0σ). This procedure is illustrated (in 2-D) in Figure 4.2.

As the name EMTopo suggests the energy of the cells used in the topo-clustering procedure is measured at the electromagnetic scale, where the energy of electrons and photons is reconstructed correctly, but not that of hadrons due to lack of corrections for the non-compensating nature of the ATLAS calorimeter. This effect is later compensated by using jet calibrations as described in Section 4.4.3

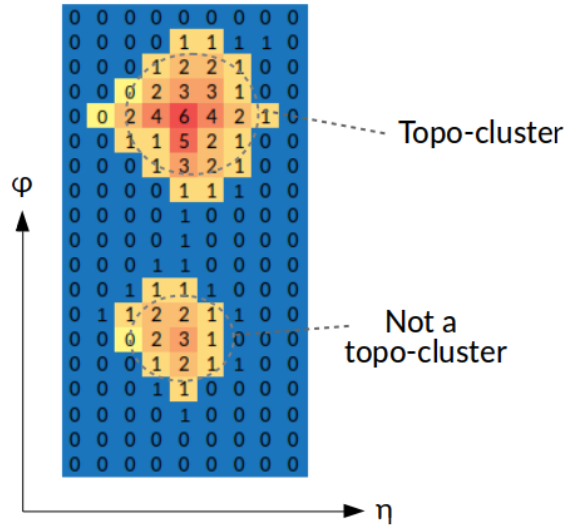


Figure 4.2: Illustration of the topo-clustering algorithm restricted to 2-D. The number in each cell represents the significance of the cell signal above the distribution of electronic+pileup noise.

4.4.2.2 | Jet Algorithms

As already mentioned, jets are defined by the means of a clustering algorithm. Each such algorithm is a set of mathematical rules, which dictates how to group together the jet constituents (topo-clusters in our case) to form the object defined as a jet. It is clear that even if one starts from the same initial constituents, using different clustering algorithms can result in a distinct collection of jets. Hence whenever discussing a measurement or a physics analysis that uses jets, it is necessary to specify the algorithms by which they were reconstructed. The jets used in the VH(cc) analysis described in this thesis were reconstructed using the anti- k_t algorithm [78], which is the standard, and most common clustering algorithm in ATLAS. The anti- k_t algorithm belongs to a wider family of algorithms known as *sequential recombination algorithms*, that use a bottom-

up approach, starting from the individual topo-clusters and trying to successively "re-solve" the QCD branchings. Before diving into the details of the anti- k_t algorithm, it is worth also mentioning another family of algorithms, called Cone algorithms, where a top-down approach is used to find the energy flow within a cone of some radius R [79].

The anti- k_t algorithm

In sequential recombination algorithms the distance between two constituents i, j is parametrized as:

$$\begin{aligned} d_{ij} &= \min \left(p_{T_i}^{2p}, p_{T_j}^{2p} \right) \frac{\Delta R_{ij}^2}{R^2} \\ d_{iB} &= p_{T_i}^{2p}, \end{aligned} \quad (4.5)$$

where R is the pre-defined distance parameter of the algorithm, and $\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ is the angular distance between the objects. d_{iB} represents the distance between object i and the beam line. For the anti- k_t algorithm the power in the definition of d_{ij} is $p = -1$. Setting the power to $p = 0, 1$ defines the Cambridge/Aachen [80] and k_t [81] algorithms respectively.

Using the measures of Equation (4.5) the anti- k_t algorithms proceeds in the following iterative steps:

1. Compute d_{ij} and d_{iB} for all i, j pairs and find the minimum.
2. If the minimum is a d_{ij} , merge object i and j and return to step 1.
3. If the minimum is a d_{iB} , consider object i as a final jet and remove it from the list of remaining items and return to step 1.
4. Continue until no particles are left.

The negative sign of the power p in the anti- k_t algorithm implies that the jets will form around the hardest objects in the collection. This is due to the fact that for close-by objects with high p_T , d_{ij} is extremely small, and will thus be combined by the algorithm first. The contribution of softer particles is usually added in later stages. The anti- k_t algorithm results in jets with a typical round and symmetric shape around the jet axis, as

can be seen in Figure 4.3, which shows a comparison of the jet collection obtained by applying different algorithms on the same initial simulated event. The distance parameter R regulates the area of the jet, as it determines how much two far objects are penalized by the $\left(\frac{\Delta R_{ij}^2}{R^2}\right)$ term during the clustering. In ATLAS the typical choice for standard quark/gluon-initiated jets is $R = 0.4$. However, some analyses make use of large- R jets with $R = 1.0$, which is characteristic of boosted heavy particles (e.g. top-quark, W , Z) decaying into quarks.

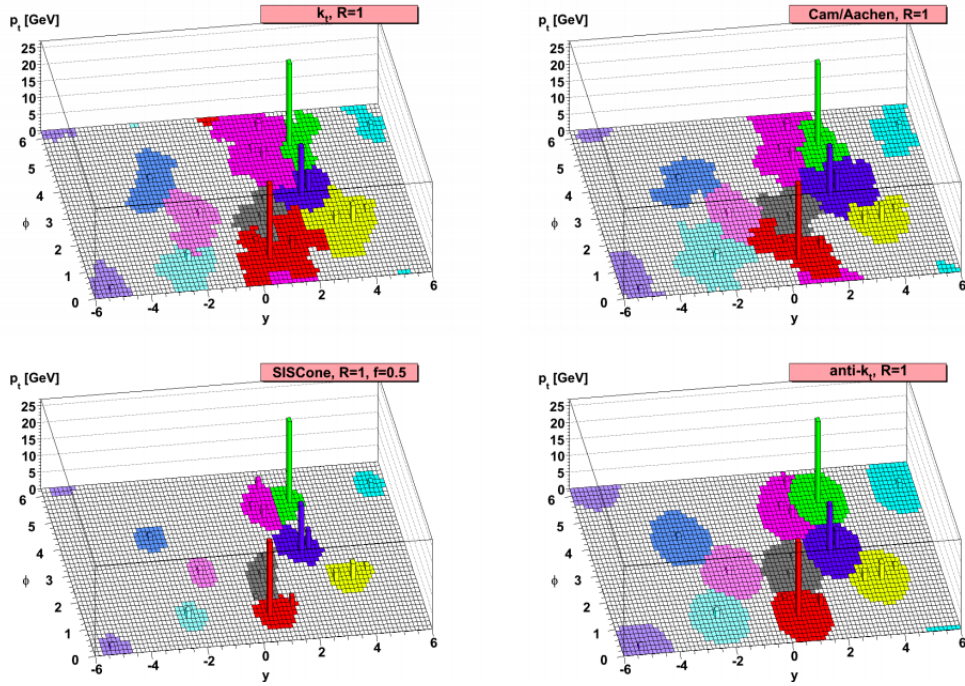


Figure 4.3: A simulated parton-level event clustered with four different jet algorithms [78].

4.4.3 | Jet Calibration

As previously mentioned, the jet reconstruction procedure detailed in Section 4.4.2 starts from energy clusters measured at the EM scale. As a consequence, the energy of electromagnetic showers is expected to be accurately measured, but because of the non-compensating nature of the calorimeter system, energy depositions initiated by hadrons are not properly accounted for due to the undetectable energy created during their interactions. Other effects, such as pileup and energy leaks, can also cause energy loss in the final reconstructed jet. The jet energy scale (JES) calibration consists of a number of

steps applied to the reconstructed jets, to correct for several effects and assign them with meaningful energy and momentum that correspond to that of the initial partons [82].

The JES calibration steps are illustrated by the flowchart shown in Figure 4.4 and are briefly described in the following:

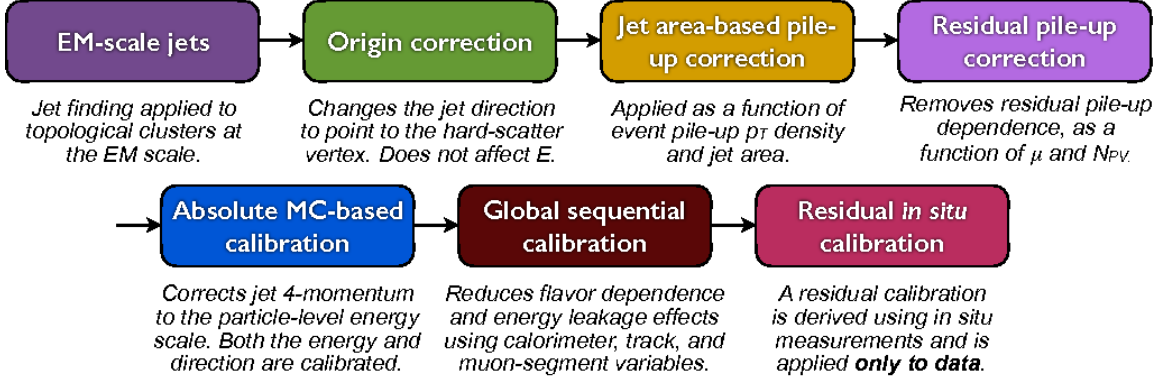


Figure 4.4: Calibration stages for EM-scale jets [82].

Jet Origin Correction The result of applying the clustering algorithm on the energy topo-clusters is a collection of jets at the EM scale. An *origin correction* changes the 4-vector of the jets to make them point back to the primary vertex instead of the detector center. This correction only affects the direction of the jets and acts primarily to improve their η resolution.

Pileup Correction Next, the effects of the in-time and out-of-time pileup are accounted for. This is done in two steps. First, an *area-based* method subtracts the per-event pileup contribution to the jet p_T . The pileup contribution is calculated from ρ - the median p_T density of jets in the $\eta - \phi$ plane - and from the jet area A . Next, an additional *residual correction* is applied to the jet based on the number of primary vertices N_{PV} and the number of interactions per bunch-crossing μ . Following the two pileup correction steps, the p_T of the jet is given by:

$$p_T^{\text{corr}} = p_T^{\text{reco}} - \overbrace{\rho \cdot A}^{\text{area-based}} - \underbrace{\alpha \cdot (N_{PV} - 1) - \beta \cdot \mu}_{\text{residual}}, \quad (4.6)$$

where p_T^{reco} refers to the EM-scale p_T of the reconstructed jet. The coefficients α and β in the residual correction term are extracted from linear fits to the jet p_T as a function of

N_{PV} and μ , and parameterize the p_T dependency on these variables.

Absolute MC-based Jet Energy and η Calibration This step corrects the reconstructed jet 4-momentum to the particle level-energy scale and accounts for biases in the jet η reconstruction. The correction relies on comparing the energy of reconstructed jets to *truth* jets, i.e. jets obtained by applying the same anti- k_t clustering algorithm on simulated particles, rather than the energy topo-clusters. The *jet energy response*, defined as:

$$\mathcal{R} = \frac{E^{\text{reco}}}{E^{\text{truth}}}, \quad (4.7)$$

is calculated in bins of p_T and η using reconstructed and truth jets that are geometrically matched within $\Delta R = 0.3$. The $\mathcal{R}$ distribution in each bin is fitted to a Gaussian, and the (inverse of the) resulting mean is considered as the correction applied to the reconstructed jet. Figure 4.5(a) show the jet energy response distribution as a function of η_{det} (the jet η pointing to the geometric center of the detector) for jets with different truth energies. One can see from the plot that low energy jets tend to have a lower p_T response, and as such are assigned a larger correction factor. Gaps and transition regions between calorimeter sub-detectors result in a lower energy response due to absorbed or undetected particles. This is evident in the figure for the barrel–end-cap ($|\eta_{\text{det}}| \sim 1.4$) and end-cap–forward ($|\eta_{\text{det}}| \sim 3.1$) transition regions. As a bias is also observed between η^{reco} and η^{truth} , an additional correction is applied based on the difference between the two, and is shown in Figure 4.5(b). The correction is largest in jets that encompass two calorimeter regions with different energy responses caused by changes in calorimeter geometry or technology. This artificially increases the energy of one side of the jet with respect to the other, altering the reconstructed jet four-momentum. Again, the barrel–end-cap and end-cap–forward transition regions can be clearly seen in the figure.

Global Sequential Calibration (GSC) Following the previous jet calibrations, residual dependencies of the JES on longitudinal and transverse features of the jet are observed, as well as effects due to the jet particle composition and the internal jet energy distribution. These are addressed by applying a sequence of sub-calibration known as the *Global Sequential Calibration* (GSC). Five observables are identified to improve the resolution of the JES, for each an independent jet 4-momentum correction is derived as a function of p_T^{truth} and $|\eta|$ by inverting the reconstructed jet response. The five observables are:

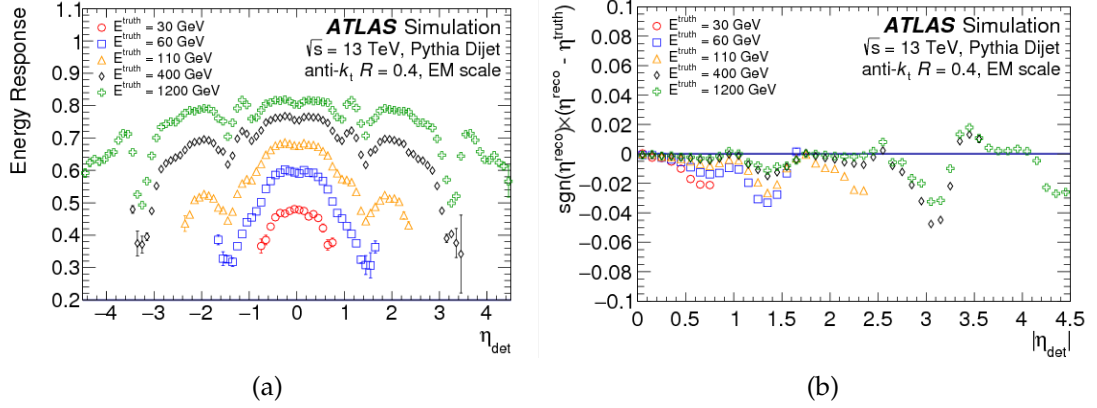


Figure 4.5: The average jet energy response (a) and the signed difference between the truth jet and reconstructed jet η (b) shown as a function of η_{det} for jets of a truth energy $E^{\text{truth}} = 30, 60, 110, 400, 1200$ GeV.

- f_{Tile0} , the fraction of jet energy measured in the first layer of the hadronic Tile calorimeter.
- f_{LAR3} , the fraction of jet energy measured in the third layer of the EM LAr calorimeter.
- n_{trk} , the number of tracks with $p_T > 1$ GeV associated with the jet.
- $\mathcal{W}_{\text{trk}}$, the average p_T weighted transverse distance in the $\eta - \phi$ plane between the jet axis and the all track of $p_T > 1$ GeV associated with to the jet.
- n_{segments} , the number of muon track segments associated with the jet.

In-situ Correction The last stage of the jet calibration accounts for remaining differences between data and MC simulation. Such differences arise from the imperfect description of the detector response and detector material in MC simulation, as well as in the simulation of the hard scatter, pileup, jet formation, and electromagnetic and hadronic interactions with the detector. Two classes of corrections are considered. The first, known as η -*intercalibration*, corrects the average response of forward jets to that of well-measured central jets using dijet events. The second class of corrections is based on balancing the p_T of jets that are recoiling against a well-measured reference object, which can be a photon, Z-boson, and multijet systems. The in-situ response is defined as the ratio of reconstructed jet p_T to the reference object p_T . The correction factor applied to the reconstructed jet is then obtained by inverting the ratio of in-situ response in data and MC:

$$c = \frac{R_{\text{in-situ}}^{\text{data}}}{R_{\text{in-situ}}^{\text{MC}}}. \quad (4.8)$$

The correction is derived as a function of the jet p_T , and also as a function of the jet η in the η –intercalibration.

The jet energy resolution (JER) defined as $\frac{\sigma_{p_T}}{p_T}$ is also obtained in-situ from balancing the jet p_T in dijets events [83].

4.4.4 | Pileup Jets Suppression

As discussed in Section 3.1.2, at the high LHC collision rate, pileup contamination can have a significant impact on the interpretation of the hard scatter events. In the context of jets, pileup can add electronic signals that modify the reconstructed jet energy, shape, and position. However, this effect is mitigated through the pileup correction described in the previous sub-section. On the other hand, pileup can also result in additional fake jets, not associated with the hard-scatter vertex. To discriminate between these *pileup jets* and the hard-scatter jets targeted by physics analyses a discriminant called *Jet Vertex Tagger* (JVT) had been developed [84]. The JVT is constructed from a two-dimensional likelihood based on two variables. The first, known as corrJVF (with "JVF" standing for "Jet-Vertex Fraction") is defined as:

$$\text{corrJVF} = \frac{\sum_k p_T^{\text{trk}_k} (PV_{HS})}{\sum_l p_T^{\text{trk}_l} (PV_{HS}) + \langle p_T^{\text{PU}} \rangle}, \quad (4.9)$$

where the numerator and first term in the denominator is the scalar p_T sum of tracks associated with the jet that originate from the high-scatter vertex PV_{HS} . The second term in the denominator $\langle p_T^{\text{PU}} \rangle$ is proportional to the average scalar p_T sum of tracks associated with the jet but originating from pileup vertices. corrJVF is constructed such that its value is expected to be near one for hard-scatter jets, and zero for pileup jets.

The second variable is R_{p_T} which is defined as follows:

$$R_{p_T} = \frac{\sum_k p_T^{\text{trk}_k} (PV_{HS})}{p_T^{\text{jet}}}. \quad (4.10)$$

Here the numerator is again the scalar p_T sum of tracks associated with the jet and originating from the high-scatter vertex. The denominator now is the calibrated jet p_T

(including the pileup subtraction). The R_{p_T} distribution is expected to peak at zero for pileup jets.

R_{p_T} and corrJVF are used to construct a two-dimensional likelihood based on a k-nearest neighbor (k-NN) algorithm trained on a sample of hard-scatter and pileup jets. The final JVT likelihood score is presented in Figure 4.6(a) as a function of R_{p_T} and corrJVF. The JVT likelihood is constructed such that hard-scatter jets, which are associated with large values of R_{p_T} and corrJVF, will tend to have a JVT likelihood score closer to one. This is also illustrated in Figure 4.6(b), which show distributions of the JVT likelihood score for pileup jets and hard-scatter jets.

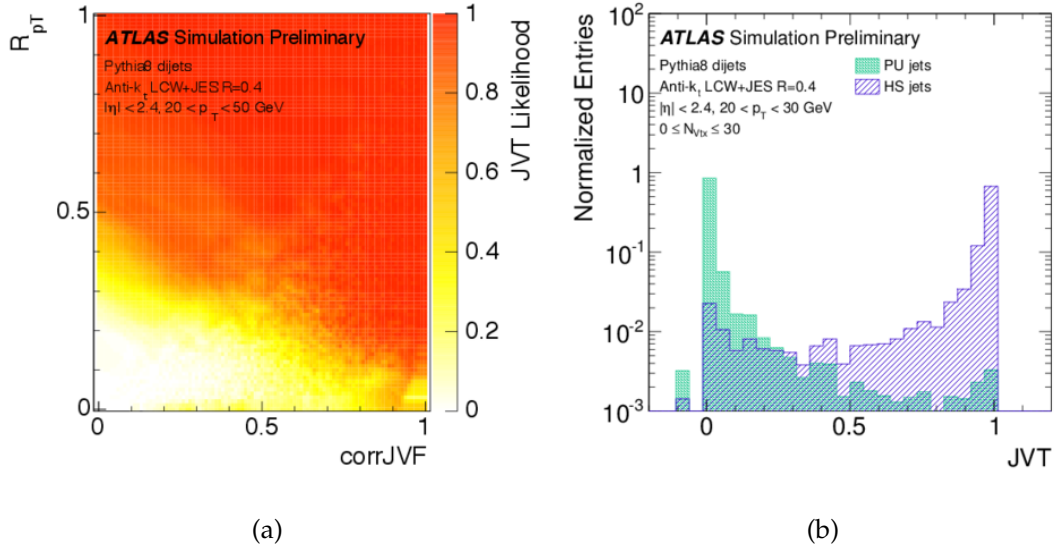


Figure 4.6: The two-dimensional JVT likelihood as a function of corrJVF and R_{p_T} (a) and the distribution of JVT for pileup and hard-scatter jets (b). Values closer to one indicate a higher probability of the jet being a hard-scatter jet [84].

4.5 | Hadronic Taus

Tau particles ($\tau^\pm$) are the third generation charged leptons. With a mass of 1.78 GeV [28] they are the heaviest among the lepton family. With a mean lifetime of $(290.3 \pm 0.5) \times 10^{-15}$ sec, a typical τ generated in pp collisions in the LHC will travel a distance of $c\tau = 87\mu\text{m}$ before decaying, prior to reaching the detector. The leptonic $\tau \rightarrow l\nu_l\nu_\tau$ decays, which occur in $\sim 35\%$ of the time, yield a lepton (muon or electron) in the final state which is very hard to distinguish from a prompt one. The other 65% of the τ decays are to hadrons, resulting in a narrow jet of particles in the calorimeter, associated with a

number of ID tracks [85]. Therefore τ reconstruction and identification in ATLAS refers to the hadronic components of the decays, which are characterized with one or three $\pi^\pm$ (called *prongs*) and in some of the cases to additional π^0 which immediately decay to a pair of photons.

The reconstruction of taus starts from anti- k_t jets with $R = 0.4$ that satisfy $p_T > 10$ GeV and $|\eta| < 2.5$. To increase the reconstruction efficiency and reduce pileup effects, the τ production vertex is identified and is required to have either one or three tracks associated with it. The tracks are required to be within an angular distance of $\Delta R = 0.2$ from the τ -jet candidate direction. The impact parameter of tracks with respect to the τ vertex is required to satisfy: $|d_0| < 1.0$ mm and $|\Delta z_0 \sin \theta| < 1.5$ mm.

τ identification serves to distinguish between τ -jets and QCD jets using a boosted-decision-tree (BDT) algorithm that considers several discriminating variables, such as track multiplicity, collimated energy depositions, and the displacement of the τ vertex. The BDT is trained separately for 1-prong and 3-prong decays using simulation samples of $Z/\gamma^* \rightarrow \tau^+ \tau^-$. Three working points labeled *Loose*, *Medium*, and *Tight* are defined.

4.6 | Missing Transverse Momentum

The concept of missing transverse momentum was introduced in Section 2.1.2. It plays a key role in estimating the energy/momentum of *invisible* particles generated in the LHC collisions. By invisible particles we refer to particles that hardly interact with the detector material, thus leaving no detectable signal as they pass through it. Neutrinos are the most obvious example of invisible particles, but not only them, as e.g. dark-matter candidates are also expected to escape detection.

The conservation of momentum in the transverse plane requires the total transverse momentum in the event to be zero. The imbalance, E_T^{miss} , is reconstructed using two contributions [86]. The first comes from objects produced in the primary interaction, evaluated from fully reconstructed and calibrated particles and jets. The second contribution is a *soft term*, accounting for charged particles tracks that are associated with the PV but not with any of the reconstructed hard objects. The $\vec{E}_T^{\text{miss}}$ vector is calculated from the vectorial sum of missing transverse momentum terms:

$$\begin{aligned}
\vec{E}_T^{\text{miss}} &= - \left(\sum_{\text{electrons}} \vec{p}_T^e + \sum_{\text{photons}} \vec{p}_T^\gamma + \sum_{\tau\text{-leptons}} \vec{p}_T^\tau + \sum_{\text{muons}} \vec{p}_T^\mu + \sum_{\text{jets}} \vec{p}_T^{\text{jet}} + \sum_{\text{soft-track}} \vec{p}_T^{\text{soft}} \right) \\
&= \vec{E}_T^{\text{miss,hard}} + \vec{E}_T^{\text{miss,soft}}.
\end{aligned}
\tag{4.11}$$

The particles and jets included in the hard term must reflect the selections of each physics analysis, which means that different analyses may end up using different $\vec{E}_T^{\text{miss}}$ vectors. It is also common to refer to the magnitude of the missing transverse momentum, which is denoted simply as:

$$E_T^{\text{miss}} = |\vec{E}_T^{\text{miss}}|. \tag{4.12}$$

The total transverse energy $\sum E_T$ is computed for each event as the scalar sum of the transverse momentum of objects entering the $\vec{E}_T^{\text{miss}}$ calculation:

$$\sum E_T = p_T^e + p_T^\gamma + p_T^\tau + p_T^\mu + p_T^{\text{jet}} + p_T^{\text{soft}}. \tag{4.13}$$

This quantity is used to quantify the hardness of the hard-scatter event in the transverse plane.

Missing transverse momentum can also be constructed based on inner detector tracks only, and in this case it is referred to as $E_{T,\text{trk}}^{\text{miss}}$ or equivalently p_T^{miss} . This provides a robust estimation of the E_T^{miss} in the event while being less sensitive to effects from pileup. However, since p_T^{miss} is based mostly on tracks of charged particles, it can not account for neutral particles.

Flavor Tagging

The subject of this thesis is a search for the decay of the Higgs boson to a pair of charm quarks. However, as all types of quarks undergo the steps of parton-branching and hadronization that result in a jet (except top-quarks which decay via $t \rightarrow Wb$ before they hadronize), one might wonder how is it possible to distinguish between jets initiated by charm-quarks, and those initiated by any other flavor. The answer to this question is a technique called *flavor-tagging*, which is the subject of this chapter. The idea is to exploit the distinctive properties of heavy hadron decays and use them to construct classifiers that can tell, with some probability, whether a given jet was initiated by a bottom quark, a charm quark, or a light-flavored quark or gluon. Note that throughout this chapter the term "heavy flavor" refers to bottom and charm, and no distinction is made between the up, down, and strange flavors which are collectively referred to as *light*. In addition, jets initiated by gluons are also considered to be light-flavored, unless stated otherwise.

Historically, flavor-tagging started from the identification of jets containing b -hadrons (b -jets), as their high mass and long lifetime make it easier to distinguish them from the other flavors. Moreover, in ATLAS it is of high necessity to be able to identify b -jets specifically, and b -tagging plays a crucial role in many areas of the physics program [87]. To name a few examples: in the Higgs sector, the largest branching ratio is to a pair of b -quarks with $\sim 60\%$ of the decays, and b -jets are also present in the decays of the top in the $t\bar{t}H$ production mode. In addition, many searches for new physics focus on scenarios where the BSM particles couple to third-generation quarks (e.g. [88, 89]). Also, in the LHC the production cross sections for a $t\bar{t}$ pair and a single top-quark are among the highest of the SM processes. Hence the presence of a b -quark in top-decays makes b -tagging essential for a variety of SM precision measurements, and for a reliable estimation of top-initiated SM background processes.

The lower masses and shorter lifetimes of c -hadrons compared to b -hadrons make the properties of c -jets somewhat "in-between" b - and light-jets, and hence the task of identifying them is more difficult. An additional complication is due to the c -hadrons present in the typical decay chain of b -hadrons, which should be distinguished from the hadrons formed by a prompt charm quark. As a result, c -tagging, i.e. the identification of jets initiated by c -hadrons (c -jets), is a rather novel development only made possible in recent years thanks to improvements in machine learning algorithms and techniques. In fact, the first iteration of the VH(cc) analysis [13], published in 2018, is the first official ATLAS result to make use of c -tagging.

This chapter is sectioned as follows: The algorithms used to perform flavor-tagging of jets in ATLAS are described in Section 5.1. The assessment of the algorithms and the flavor-tagging performance is then explained in Section 5.2, and Section 5.3 presents the flavor-tagging calibration, which accounts for possible differences between applying these algorithms to data and simulation. Given the above discussion, even though b - and c -tagging share the same methodology and use similar algorithms, in the following sections I find it easier to introduce the key concepts of flavor-tagging by the means of b -jets, and when necessary highlight how the picture changes for c -jets.

5.1 | Flavor-Tagging Algorithms

B -hadrons have a number of characteristics that can be used to distinguish b -jets from light flavor jets. Given a lifetime of the order of 1.5 ps ($\langle c\tau \rangle \approx 450\mu\text{m}$), measurable b -hadrons have a significant mean flight length, $\langle l \rangle = \beta\gamma c\tau$, in the detector before decaying, generally leading to at least one vertex displaced from the hard-scatter collision point. Their large mass of order > 5 GeV leads to high decay product multiplicity, which in $\sim 20\%$ of the times include a charged lepton (e or μ) that can be associated with the jet. Moreover, since the CKM matrix suppresses transitions from third to first generation ($|V_{bc}| \gg |V_{bu}|$), b -hadrons decay into c -hadrons, a tertiary vertex from the subsequent decay of the c -hadron may also be present within the jet cone. Another feature that can be exploited is the hard fragmentation of b -quarks, which leads to $\sim 70\%$ of the jet momentum being carried by the b -hadron and can be associated to the secondary, displaced vertex.

In ATLAS the b -tagging strategy includes two stages. First, each of the b -jets characteristics mentioned above is used to define a *low-level algorithm*. In the second stage, the results of the low-level b -tagging algorithms are combined in *high-level algorithms*

consisting of multivariate classifiers. The output of a high-level classifier is a single number (a "score"). A jet with a score greater than some pre-defined threshold value is considered to be b -tagged.

The development and commissioning of the flavor-tagging algorithms in ATLAS, as well as their calibration, is the responsibility of the Flavour-Tagging (FTAG) working group [90]. The algorithms are trained using MC simulation of $t\bar{t}$ events as well as a BSM scenario of Z' decaying to $b\bar{b}$. The $t\bar{t}$ event topology ensures the presence of two b -jets from the decays of the top, and a sufficient amount of c - and light-jets from subsequent decays of the W . The Z' sample is used to optimize the b -tagging performance of high p_T jets. Unlike in real data, in MC the *truth* flavor of the jet is known, which is a crucial ingredient for the learning procedure of the algorithms. Jet flavor labels are assigned exclusively in the following way:

1. If a weakly decaying b -hadron with $p_T > 5$ GeV is found within $\Delta R < 0.3$ of the jet axis, the jet is labeled a b -jet.
2. If no b -hadron is found, but a weakly decaying c -hadron ($p_T > 5$ GeV) is found within $\Delta R < 0.3$ of the jet axis, then the jet is labeled as a c -jet.
3. Otherwise, if a τ -lepton is found within $\Delta R < 0.3$ of the jet axis, the jet is labeled a τ -jet.
4. If any one hadron or τ -lepton matches more than one jet, the closest jet is chosen as its parent.
5. All unlabeled jets after steps 1 through 4 are labeled as light-jets.

5.1.1 | Low-Level Algorithms

Low-level b -tagging algorithms fall into two broad categories. The first approach is based on the large impact parameters of tracks originating from the b -hadron decay, and the second category relies on explicitly reconstructing the secondary (and tertiary) vertices within the jet. Information on muons in the jet, produced by the semi-leptonic decays of the b - and c -hadrons is also considered in a dedicated algorithm.

Impact Parameter Based: IP2D and IP3D The IP2D and IP3D are complementary impact parameter-based algorithms [91]. They both make use of the signed transverse impact parameter significance of tracks, d_0/σ_{d_0} , to construct discriminating variables.

IP3D also considers the longitudinal impact parameter significance, $z_0 \sin \theta / \sigma_{z_0 \sin \theta}$. As illustrated in Figure 5.1(a), the sign of the impact parameter is considered positive if the primary vertex is in front of the secondary vertex (relative to the jet direction) and negative otherwise. Figure 5.1(b) show the distribution of signed d_0 significance of tracks from b -, c -, and light-jets. The plot shows that b -jets tend to have large positive values, due to the long lifetime and decay length of b -hadrons. Only tracks passing a number of quality criteria based on their hit pattern in the detector (labeled "Good") are used to construct the d_0 / σ_{d_0} distribution in Figure 5.1(b). More details on the track-quality categorization used by the IP2D and IP3D algorithms can be found in [92]. The IP2D and IP3D algorithms rely on constructing log-likelihood ratios (LLR) taking as inputs probability density functions (PDFs) for b -, c -, and light-flavor jet probabilities on a per-track basis.

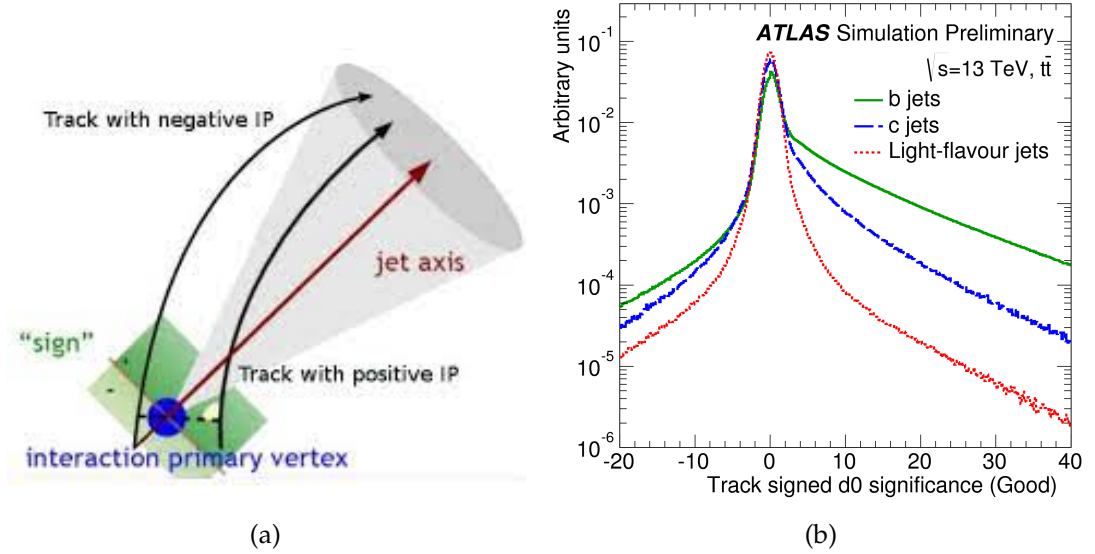


Figure 5.1: (a) Schematic diagrams of a jet associated with tracks of positive and negative impact parameter [93]. (b) Distribution of the signed transverse impact parameter significance d_0 / σ_{d_0} of tracks from b -jets (green), c -jets (blue), and light-jet (red). Only tracks from the "Good" category, defined in [92], are used to construct the distributions.

Secondary Vertex Finding: SV1 The secondary vertex tagging algorithm, SV1, reconstructs a single displaced secondary vertex in a jet [94]. The reconstruction starts from identifying the possible two-track vertices built with all tracks associated with the jet, while rejecting tracks that are compatible with the decay of long-lived particles (K_s or Λ), photon conversions, or hadronic interactions with the detector material. The remaining

secondary vertices and associated tracks are then used to construct eight discriminating observables which are used as input to the high-level taggers. These observables include the number of tracks associated with the SV1 vertex, the invariant mass of the secondary vertex, its energy fraction (defined as the total energy of all the tracks associated with the secondary vertex divided by the energy of all the tracks associated with the jet), and the three-dimensional decay length significance (i.e. decay length divided by its uncertainty).

Multi-Vertex Finding: JetFitter The topological multi-vertex algorithm, `JetFitter` (JF), attempts to reconstruct the full b -hadron decay chain based on the topological structure of the weak b - and c -hadron decays inside the jet [95]. It uses a modified Kalman filter with the assumption that the b - and c - vertices are located on a common line. As with SV1, several discriminating variables are constructed and used as input to the high-level taggers. `JetFitter` also plays a role in the discrimination between b - and c -jets, as it can exploit the topological differences between decays of bottom and charm hadrons. In charm hadron decays one expects to find only a single secondary vertex with intermediate charged decay multiplicity and a decay distance comparable to that of b -hadrons. Therefore, additional c -tagging variables are built by using exclusively the jets where only one displaced vertex is reconstructed by `JetFitter`. These include the number, invariant mass, and energy of the particles associated with the secondary vertex as well as their rapidity, computed with respect to the jet axis and the vector defined between the primary and secondary vertex.

Soft Muon Tagger Algorithm: The Soft Muon Tagger (SMT) is based on the reconstruction of muons coming from semi-leptonic decays of heavy-flavor hadrons. These muons tend to have smaller p_T compared to muons coming from electroweak boson decays (hence the label "soft"), and large transverse momentum relative to the jet axis, p_T^{rel} . The SMT considers muons that are matched to the closest jet within $\Delta R = 0.4$. For the selected muon-jet pairs the ΔR , d_0 , p_T^{rel} and three variables related to the quality of the muon track are used in a boosted decision tree (BDT) to define the SMT discriminant.

5.1.2 | High-Level Algorithms

The high-level taggers are multivariate classifiers that take as input the observables provided from the low-level algorithm, as well as kinematic properties of the jet (p_T and η), and the jet flavor. The output of the high-level taggers is a score which is assigned on a per-jet basis. One can then set a threshold value for the high-level tagger score,

above which a jet is considered as b -tagged. A specific threshold value is referred to as a b -tagging *working point* (WP) and in practice, as explained in Section 5.2, is determined such that the desired fraction of b -jets are accepted.

For the data collected during Run-2, two main high-level algorithms were used in ATLAS, named *MV2c10* and *DL1*. In the VH(cc) analysis, c -tagging was performed through a combination of the two: *DL1* was used to select c -jets, while *MV2c10* to better discriminate against jets that are b -tagged. More details on the specific flavor tagging selection choices in the VH(cc) analysis are discussed in Chapter 9.

The full list of input variables considered by the high-level algorithms discussed below is presented in Table 5.1.

MV2 tagger: The MV2 algorithm is a BDT discriminant trained using the ROOT Toolkit for Multivariate Data Analysis (TMVA) [96]. For the training of the algorithm b -jets are considered as signal and the background is composed of a mixture of c - and light-jets. A number of variations exist for the MV2 algorithm which differ in the fraction of c -jets in the training background sample. For the work presented in this thesis, b -tagging was done with a version known as *MV2c10* [92], in which the background is composed of 7% c -jets and 93% light-jets. The distribution of the final *MV2c10* b -tagging discriminant for b -, c -, and light-jets is shown in Figure 5.2(a). The plot shows that the final *MV2c10* BDT output score (labeled D_{MV2}) peaks near $D_{MV2} \sim 1$ for b -jets, whereas c -jets have a more balanced distribution, and light-jets tend to have values close to $D_{MV2} \sim -1$. Therefore, by selecting jets with high D_{MV2} scores one can obtain a sample with a large fraction of b -jets.

MV2 can be also used for the purpose of c -tagging, as was the case in the first iteration of the VH(cc) analysis [97]. This was done by combining two variants of the *MV2* algorithm. *MV2c100* is trained with b -jets as signal and only c -jets as background, and *MV2cl100* has c -jets as signal and only light-jets as background. Both variants use the standard *MV2* input variables and the additional c -tagging variables of Table 5.1. In order to c -tag, a 2-D cut is defined by setting threshold values on the output of both *MV2* variants.

DL1 tagger: The second high-level b -tagging algorithm, *DL1*, is based on a deep neural network (NN) trained using Keras [98] with the Theano [99] backend and the Adam optimizer [100]. The NN has a three-dimensional output corresponding to the probabil-

Input	Variable	Description
Kinematics	p_T	Jet p_T
	η	Jet $ \eta $
IP2D/IP3D	$\log(P_b/P_{\text{light}})$	Likelihood ratio between the b -jet and light-flavour jet hypotheses
	$\log(P_b/P_c)$	Likelihood ratio between the b - and c -jet hypotheses
	$\log(P_c/P_{\text{light}})$	Likelihood ratio between the c -jet and light-flavour jet hypotheses
SV1	$m(\text{SV})$	Invariant mass of tracks at the secondary vertex assuming pion mass
	$f_E(\text{SV})$	Energy fraction of the tracks associated with the secondary vertex
	$N_{\text{TrkAtVtx}}(\text{SV})$	Number of tracks used in the secondary vertex
	$N_{2\text{TrkVtx}}(\text{SV})$	Number of two-track vertex candidates
	$L_{xy}(\text{SV})$	Transverse distance between the primary and secondary vertex
	$L_{xyz}(\text{SV})$	Distance between the primary and the secondary vertex
	$S_{xyz}(\text{SV})$	Distance between the primary and the secondary vertex divided by its uncertainty
JETFITTER	$\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})(\text{SV})$	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex.
	$m(\text{JF})$	Invariant mass of tracks from displaced vertices
	$f_E(\text{JF})$	Energy fraction of the tracks associated with the displaced vertices
	$\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})(\text{JF})$	ΔR between jet axis and vectorial sum of momenta of all tracks attached to displaced vertices
	$S_{xyz}(\text{JF})$	Significance of average distance between PV and displaced vertices
	$N_{\text{TrkAtVtx}}(\text{JF})$	Number of tracks from multi-prong displaced vertices
	$N_{2\text{TrkVtx}}(\text{JF})$	Number of two-track vertex candidates (prior to decay chain fit)
JETFITTER c -tagging	$N_{1\text{-trk vertices}}(\text{JF})$	Number of single-prong displaced vertices
	$N_{\geq 2\text{-trk vertices}}(\text{JF})$	Number of multi-prong displaced vertices
	$L_{xyz}(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$	Distance of 2^{nd} or 3^{rd} vertex from PV
	$L_{xy}(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$	Transverse displacement of the 2^{nd} or 3^{rd} vertex
	$m_{\text{Trk}}(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$	Invariant mass of tracks associated with 2^{nd} or 3^{rd} vertex
	$E_{\text{Trk}}(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$	Energy fraction of the tracks associated with 2^{nd} or 3^{rd} vertex
	$f_E(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$	Fraction of charged jet energy in 2^{nd} or 3^{rd} vertex
	$N_{\text{TrkAtVtx}}(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$	Number of tracks associated with 2^{nd} or 3^{rd} vertex
	$Y_{\text{trk}}^{\text{min}}, Y_{\text{trk}}^{\text{max}}, Y_{\text{trk}}^{\text{avg}}(2^{\text{nd}}/3^{\text{rd}}\text{vtx})(\text{JF})$	Min., max. and avg. track rapidity of tracks at 2^{nd} or 3^{rd} vertex

Table 5.1: Input variables used by the MV2 and DL1 algorithms. The JetFitter c -tagging variables are not used by the standard MV2 b -tagging algorithms [87].

ity of a jet to be a b -jet, c -jet, or light-jet. The three probabilities are then used to construct the final DL1 b -tagging discriminant:

$$D_{\text{DL1}} = \ln \left(\frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_{\text{light}}} \right), \quad (5.1)$$

where p_b , p_c , p_{light} represent the b -, c -, and light-jet probabilities, and f_c is the effective c -jet fraction in the background sample. A value of $f_c = 8\%$ was found to optimize the performance of the DL1 b -tagger and is used in the work presented in this thesis. The distribution of the final DL1 b -tagging discriminant for b -, c -, and light- jets is shown in Figure 5.2(b). Here too, one can see that b -jets tend to have large values of the final b -tagging discriminant, and as mentioned above in the context of MV2, this feature can be used to discriminate them from light- and c -jets.

The three-dimensional output of the DL1 NN offers flexibility when constructing the final discriminant. First, the c -jet fraction in the background can be chosen after the algorithm is trained in order to optimize its performance. Moreover, and much more critical to the work described in this thesis, it makes it almost trivial to adapt the algorithm for c -tagging, as one can use the three output probabilities to define the discriminant:

$$D_{\text{DL1}}^{c\text{-tag}} = \ln \left(\frac{p_c}{f_b \cdot p_b + (1 - f_b) \cdot p_{\text{light}}} \right), \quad (5.2)$$

where now f_b refers to the fraction of b -jets in the training sample. For the VH(cc) analysis the specific values of f_b and the threshold cut on the $D_{\text{DL1}}^{c\text{-tag}}$ were optimized to maximize the signal efficiency as described in Section 9.2.

5.2 | Flavor-Tagging Performance

The performance of a flavor-tagging algorithm is quantified by the *tagging efficiency* of signal and background jets, which are the probability to correctly identify a signal jet (i.e. to b -tag a b -jet/ c -tag a c -jet) or to mistakenly tag a background jet. The tagging efficiencies for each jet flavor, ϵ_b , ϵ_c , and ϵ_{light} are first evaluated using MC simulation, and then calibrated to account for possible differences when applying the algorithms to data (as will be explained in more detail shortly). For b -tagging, ϵ_b is also used to define four working points, by tuning the cut values on the tagger output score such that they correspond to signal efficiencies of $\epsilon_b = 60\%$, 70% , 77% , and 85% . Choosing to work with

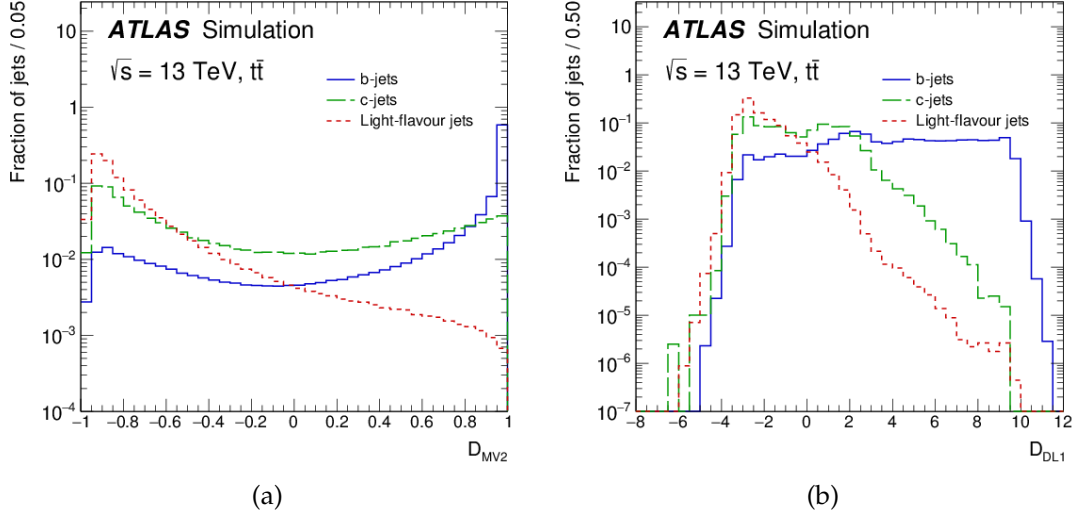


Figure 5.2: Distribution of the final MV2c10 BDT output score (a) and D_{DL1} (evaluated with $f_c = 8\%$) (b) for b -(blue), c -(green), and light-jets (red) [101].

a more loose working point (i.e. with a higher tagging efficiency) has the advantage of successfully identifying more of the signal jets, however, as a consequence a larger portion of the background jets is wrongfully selected, resulting in high levels of contamination and bad purity of the selected sample. This effect is illustrated in Figure 5.3 which shows the b -tagging efficiency on the x -axis and the light- (Figure 5.3(a)) and c -jet (Figure 5.3(b)) rejection on the y -axis, for the individual low-level and high-level algorithms. The background rejection is defined as the inverse of the efficiency (i.e. $1/\epsilon_{light}$ and $1/\epsilon_c$ for b -tagging) and is often used to characterize the ability of a working point to disqualify background jets. As seen in the plot, one can not independently set both the b -tagging efficiency and the background rejection, as there is a trade-off between the two; in order to obtain a clean b -jets sample with as little light- and c -jets contamination as possible, one may set the b -tagging efficiency of the algorithm to smaller values. On the other hand, increasing the b -tagging efficiency of an algorithm will result in identifying more b -jets and wrongfully selecting more light- and c -jets. Figure 5.3 also show, as expected, the improved performance from combining the low-level algorithms into high-level taggers, as MV2 (solid blue) and DL1 (dashed red) demonstrate higher background rejection for all values of b -tagging efficiency. The bottom pad in the plots show the ratio between each algorithm and the blue curve of MV2.

In principle, c -tagging working points can be defined in a similar manner using ϵ_c . It should be noted though, that c -tagging efficiencies are significantly lower than

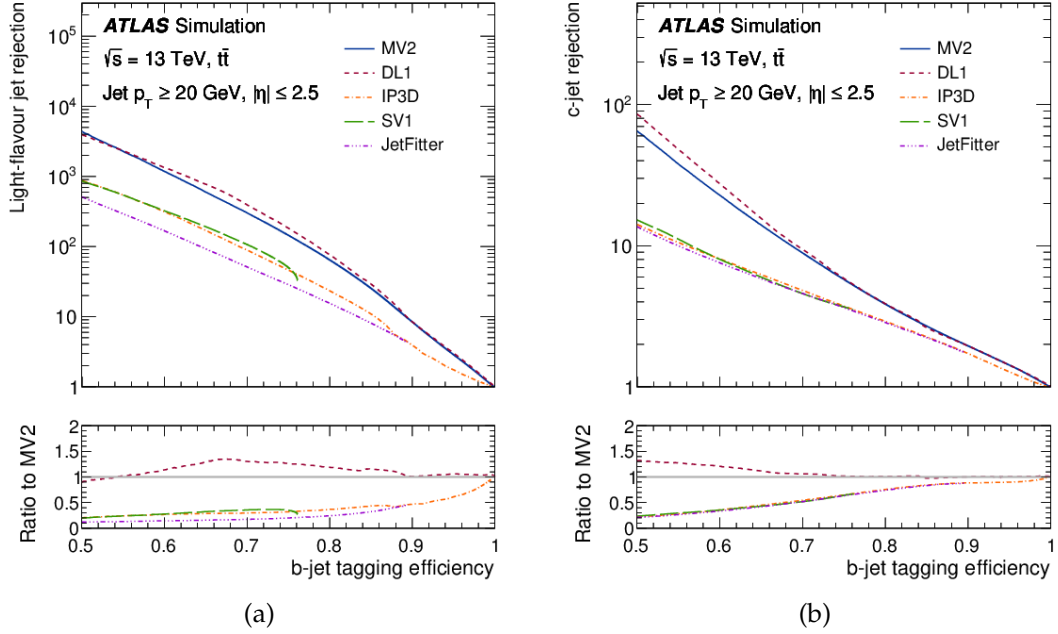


Figure 5.3: The light- (a) and c -jet (b) rejection versus the b -tagging efficiency for the IP3D, SV1, JetFitter, MV2, and DL1 b -tagging algorithms, evaluated in $t\bar{t}$ events [92]. The bottom pad in the plots show the ratio between each algorithm and the blue curve of MV2.

those of b -tagging, typically ranging between 20%-40%. This makes sense given that c - and b -tagging relies on similar features of the jet, which are less pronounced in c -jets compared to b -jets. Therefore, instead of a c -tagging working point based on a fixed ϵ_c value, the c -tagging working points used in VH(cc) were specifically configured¹ as to maximize the selection of $VH \rightarrow c\bar{c}$ signal events over background events, with the goal of increasing the analysis sensitivity (see Section 9.2 for more details).

5.3 | Flavor-Tagging Working Points Calibration

As already mentioned, the tagging efficiencies for each jet flavor are initially evaluated using the MC simulation samples used for the algorithm training. However, differences may arise between the efficiency measured in simulation to that observed in real data. To account for these differences a calibration procedure is performed, which results in correction scale factors (SF) applied to the MC jets, defined as:

¹Because each flavor-tagging working point requires its own calibration and maintenance, the FTAG group in ATLAS aims to keep a small number of official working points, which are used by all analyses. However, when work on VH(cc) began it was the only analysis to use c -tagging algorithm, making it possible to configure the c -tagging working point according to the analysis-specific needs.

$$\text{SF} = \frac{\epsilon^{\text{data}}}{\epsilon^{\text{MC}}}. \quad (5.3)$$

The correction scale factors are dependent on the simulated jet truth flavor and its kinematics. A dedicated calibration procedure is performed for each jet flavor using events in data enriched with that particular flavor.

For b -jets, the calibration uses di-leptonic $t\bar{t}$ events in data (i.e. events where both W bosons decay leptonically) [102]. The event selection used, based on a BDT trained on kinematic variables, leads to a sample very pure in b -jets. 2-D probability density functions of the tagging discriminant are built as a function of the leading and sub-leading jet p_T and are used in a likelihood fit to extract the tagging efficiency in data.

The tagging efficiency of light-jets is measured in data using events containing a Z boson with additional production of jets (referred to as Z +jets) [103]. The Z +jets process contains a higher fraction of light-flavored jets compared to events containing a top-quark, where a b -jet from the $t \rightarrow b$ decay is always present. The light-jet calibration uses the *negative-tag* method, where a sample enriched in light-jets is achieved by requiring jets to have a negative signed impact parameter significance.

The tagging efficiency of c -jets is measured in data with semi-leptonic $t\bar{t}$ events (one W boson decaying leptonically and the other W decaying hadronically). Since I contributed much work to the c -jets calibration, it will be described in greater detail in Section 9.3 in the context of c -tagging within the VH(cc) analysis.

In addition, as the c -tagging efficiency in simulation depends on the various properties of the MC generator, an extra calibration known as MC/MC scale factors is applied to account for differences between simulations. They are derived by computing the ratios of the flavor-tagging efficiencies in alternative and nominal $t\bar{t}$ samples, and are applied as a correction to Eq (5.4):

$$\text{SF} = \frac{\epsilon^{\text{data}}}{\epsilon^{\text{MC}, \text{nominal}}} \times \frac{\epsilon^{\text{MC}, \text{nominal}}}{\epsilon^{\text{MC}, \text{alternative}}}. \quad (5.4)$$

Part III

VH(cc)

The $VH(cc)$ Analysis - Introduction

The previous chapters covered general topics which are, to some extent, relevant to any physics analysis performed at the LHC. From this point forwards, this thesis focuses on the research at the center of my Ph.D. work, which is the direct search for the decay of the Higgs boson to a pair of charm and anti-charm quarks. As the search targets the VH production mode of the Higgs boson, it is conveniently denoted in this thesis as the $VH(cc)$ analysis. In the following sections, I will describe the various aspects of the analysis, with emphasis on the parts where I had a significant contribution. But first, in the current chapter I will give a broad overview of the analysis, including the motivation for performing the search, described in Section 6.1, and the basic strategy which is illustrated in Section 6.2.

6.1 | Motivation

Following the discovery of the SM Higgs boson in 2012, attention has turned to more detailed measurements of its properties as a means of testing the predictions of the standard model. One such class of predictions is given by the Yukawa couplings between the Higgs and SM fermions, which according to the theory are expected to be proportional to the fermions mass. Deviations from the SM-predicted values may point towards new and exotic physics and a more complex mechanism of mass generation than the one present in the SM. That being said, the Yukawa couplings of the third generation quarks [9, 10] and leptons [8] have been measured and found consistent with the SM prediction. More recently, evidence for the decay of the Higgs to a pair of muons has been reported [11, 12], yet not at the sufficient significance level required to claim observation of this process. As for the first generation of leptons and (first and) second

generations of quarks, no experimental evidence for their coupling to the Higgs has yet been found.

The discussion above illustrates why it would be beneficial to measure the $H \rightarrow c\bar{c}$ decay. First, it is a direct probe of y_c , the Yukawa charm quark coupling, which is one of the largest contributions to the total Higgs width Γ_H yet to be established experimentally. Its measurement is required to further establish the non-universal pattern of the Yukawa couplings. Second, the fact that it is so small ($y_c \sim 5 \times 10^{-3}$) means that it may be susceptible to significant modifications in a variety of BSM physics scenarios (e.g. [104–108]).

6.2 | Overview of the Analysis

Measuring the $H \rightarrow c\bar{c}$ decay in a hadron collider such as the LHC is a very challenging task, for three main reasons. First, the branching ratio of the process is rather small, with $BR(H \rightarrow c\bar{c}) \approx 2.91\%$, approximately 20 times smaller than the decay to a bottom anti-bottom pair. Second, the abundance of hadronic final state events from plain QCD scatterings impose a large background, and also requires jet triggers to be *prescaled*. This means that only a portion of the events is actually being recorded, and the rest are discarded. Prescaling is necessary to maintain a reasonable data recording rate, but it results in lower statistical power and thus loss of sensitivity. Third, due to the low c -tagging efficiency, the majority of Higgs events that do decay into a charm-anti-charm pair will not be identified correctly. As for the smallness of $BR(H \rightarrow c\bar{c})$ there is nothing that we as physicists can do, and we must accept that this is fully dictated by nature. For the other two reasons mentioned though, we can improve the sensitivity to the process with some of the choices we make. For example, we search for Higgs bosons produced in the VH mode, and target the events where the vector boson decays leptonically. This both reduces the QCD background and allows for efficient triggering. Other choices are reflected in various aspects of the analysis, which are listed briefly in the discussion below and described in detail in the following sections.

The event topologies of the $V(\rightarrow \text{leptons})H(\rightarrow c\bar{c})$ processes considered in the analysis contain two c -jets, and either zero, one, or two charged leptons. Note, that here and throughout this part, a charged lepton denotes either e or μ , due to complications in reconstructing τ leptons which decay prior to reaching the detector (see Section 4.5).

Because of the differences in the final state objects and background composition, the

events in the analysis are categorized depending on the number of charged leptons. The 0-lepton channel targets the $Z(\rightarrow \nu\nu)H(\rightarrow c\bar{c})$ process, the 1-lepton channel targets the $W(\rightarrow l\nu)H(\rightarrow c\bar{c})$, and the 2-lepton channel targets the $Z(\rightarrow ll)H(\rightarrow c\bar{c})$ process. The leading-order Feynman diagrams for the three channels are shown in Figure 6.1.

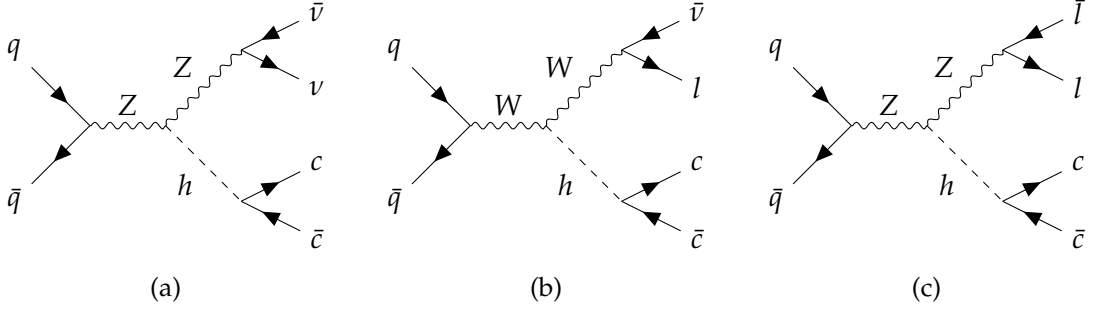


Figure 6.1: Leading order Feynman diagrams for (a) the 0-lepton channel, (b) the 1-lepton channel, and (c) the 2-lepton channel in the VH(cc) analysis.

In addition to the three leptonic channels, events passing the trigger selection and satisfying basic quality requirements are categorized based on the number of jets in the event, and the p_T of the vector boson denoted p_T^V . The $H \rightarrow c\bar{c}$ candidate is reconstructed from the two leading p_T jets in the event. Due to the low c -tagging efficiency, events are considered if at least one of the two leading p_T jets is c -tagged.

The main background process in all three channels is the production of a vector boson in association of jets (denoted V +jets), with W +jets and Z +jets dominating the 1-lepton and 2-lepton channels respectively, and a combination of the two making up the majority of the background in the 0-lepton channel. Other background processes considered in the analysis are $t\bar{t}$, single-top production, and di-boson. These backgrounds, as well as the $VH(\rightarrow b\bar{b})$ process and the $VH(\rightarrow c\bar{c})$ signal, are modeled using MC simulation. In addition to the processes mentioned above, in the 1-lepton channel exists a non-negligible *multijet* contribution to the background, coming from QCD production of jets, subsequently faking leptons and/or E_T^{miss} , or decaying to non-prompt leptons. The multijet component is modeled with a data-driven method.

The invariant mass of the $H \rightarrow c\bar{c}$ candidate, $m_{c\bar{c}}$, is used as a signal versus background discriminant. A joint binned maximum-profile likelihood fit to $m_{c\bar{c}}$ is used to extract the signal yield and the background normalization. The parameter of interest, μ_{VHcc} , common to all channels, is the signal strength, defined as the ratio of the measured signal yield to the SM prediction. The analysis strategy is validated by the simultaneous measurement of the di-boson processes in which one of the vectors decays to

at least one charm quark, $VW(\rightarrow cl)$ and $VZ(\rightarrow c\bar{c})$, where V is a W or Z and l is a down-type quark.

Data and MC Samples

This chapter describes the data sample used to perform the analysis in addition to the MC simulation used to model both the signal and background processes. Information on the data sample is presented in Section 7.1, Section 7.2 details the signal and various background processes considered in the analysis, and the MC simulation used for their estimation.

7.1 | Data Sample

The $VH(cc)$ analysis presented in this thesis uses 139 fb^{-1} of pp collision data collected at center-of-mass energy $\sqrt{s} = 13 \text{ TeV}$ by the ATLAS detector from 2015 to 2018. The used dataset was recorded with the LHC beams and all relevant ATLAS sub-detectors operating in normal conditions, and with a combination of E_T^{miss} , electron, and muon triggers described in Section 8.2.

7.2 | Simulation Samples

The $VH(\rightarrow c\bar{c})$ signal process as well as the SM background processes (with exception of the QCD background in the 1-lepton channel) are modeled using MC simulation. These simulated events are generated at least at next-to-leading order (NLO), processed using the full ATLAS detector simulation based on GEANT4 [109], and reconstructed using the same algorithms used to process the data.

Because one can not rely on a specific MC generator to accurately describe the true nature of the data, two sets of simulation samples, produced with different MC gener-

ators and/or configurations, are used to estimate each process. The first, referred to as the *nominal* MC sample, is considered as the baseline prediction on the background and signal processes used in the fit. The second, referred to as the *alternative* sample, is used to derive modeling uncertainties on the normalization, shape, and relative acceptance into the analysis categories, as detailed in Section 11.3. The nominal and alternative MC samples used in the analysis are summarized in Table 7.1, and are briefly described in the discussion below. Full details on the simulation on each sub-process can be found in the internal documentation of the analysis [15] as well as in [110].

Process	Nominal Generator	Alternative Generator	$\sigma \times \text{BR}$ [pb]
$qq \rightarrow ZH \rightarrow \nu\bar{\nu}c\bar{c}$	POWHEG+PYTHIA 8	POWHEG+HERWIG 7	4.42×10^{-3}
$qq \rightarrow W^+H \rightarrow l^+\nu c\bar{c}$			8.15×10^{-3}
$qq \rightarrow W^-H \rightarrow l^-\bar{\nu}c\bar{c}$			5.19×10^{-3}
$qq \rightarrow ZH \rightarrow l^+l^-c\bar{c}$			2.22×10^{-3}
$gg \rightarrow ZH \rightarrow \nu\bar{\nu}c\bar{c}$			7.10×10^{-4}
$gg \rightarrow ZH \rightarrow \nu\bar{\nu}c\bar{c}$			3.59×10^{-4}
$qq \rightarrow ZH \rightarrow \nu\bar{\nu}b\bar{b}$	POWHEG+PYTHIA 8	POWHEG+HERWIG 7	8.90×10^{-2}
$qq \rightarrow W^+H \rightarrow l^+\nu b\bar{b}$			1.64×10^{-1}
$qq \rightarrow W^-H \rightarrow l^-\bar{\nu}b\bar{b}$			1.04×10^{-1}
$qq \rightarrow ZH \rightarrow l^+l^-b\bar{b}$			4.48×10^{-2}
$gg \rightarrow ZH \rightarrow \nu\bar{\nu}b\bar{b}$			1.43×10^{-2}
$gg \rightarrow ZH \rightarrow \nu\bar{\nu}b\bar{b}$			7.23×10^{-3}
$Z \rightarrow \nu\bar{\nu} + \text{jets}$	SHERPA 2.2.1	MADGRAPH+PYTHIA 8	1914
$W \rightarrow l\nu + \text{jets}$			20080
$Z \rightarrow l^+l^- + \text{jets}$			2107 ($m_{ll} > 40 \text{ GeV}$)
$t\bar{t}$ (semi-leptonic)	POWHEG+PYTHIA 8	POWHEG+PYTHIA 8 + ISR/FSR,	452
$t\bar{t}$ (di-leptonic)		POWHEG+HERWIG 7 MADGRAPH5_aMC@NLO+PYTHIA 8	87
Single-top Wt -channel	POWHEG+PYTHIA 8	POWHEG+PYTHIA 8 + ISR/FSR,	68
Single-top s -channel		POWHEG+HERWIG 7	66
Single-top t -channel		MADGRAPH5_aMC@NLO+PYTHIA	33
$qq \rightarrow WW \rightarrow l\nu q\bar{q}'$	SHERPA 2.2.1	POWHEG+PYTHIA 8	50
$qq \rightarrow WZ \rightarrow l\nu q\bar{q}$			11
$qq \rightarrow WZ \rightarrow \nu\bar{\nu} q\bar{q}'$			6.8
$qq \rightarrow WZ \rightarrow l^+l^- q\bar{q}'$			3.4
$qq \rightarrow ZZ \rightarrow \nu\bar{\nu} q\bar{q}$			4.3
$qq \rightarrow ZZ \rightarrow l^+l^- q\bar{q}$			2.2
$gg \rightarrow WW \rightarrow l\nu q\bar{q}$			2.1
$gg \rightarrow ZZ \rightarrow \nu\bar{\nu} q\bar{q}$			0.44
$gg \rightarrow ZZ \rightarrow l^+l^- q\bar{q}$			0.22

Table 7.1: Signal and background processes and their corresponding nominal and alternative MC generators. Full details on the simulation of each sub-process can be found in the internal documentation of the analysis [15] as well as in [110].

7.2.1 | Modeling of the VH Signal Process

The VH signals include three main Higgs production modes: $ZH \rightarrow \nu\nu c\bar{c}$, $ZH \rightarrow llc\bar{c}$ and $WH \rightarrow l\nu c\bar{c}$, where $l = e, \mu, \tau$. The ZH production modes are furthermore split into two contributions depending on the partonic initial state: $q\bar{q} \rightarrow ZH$ and $gg \rightarrow ZH$.

WH signal samples are normalized to the production cross-section at NNLO (QCD) and NLO (EW). The inclusive cross-section of ZH production is calculated at NNLO (QCD) and NLO (EW), the cross-section of gluon induced ZH production is then calculated at NLO (QCD), and quark induced production is taken as the difference between the two in order to avoid double counting. All cross-sections and branching ratios are obtained from the LHC Higgs Cross Section Working Group prescriptions for a Higgs boson mass of $m_H = 125$ GeV. The value of 0.02891 is used for the SM $H \rightarrow c\bar{c}$ BR, while the value of 0.5824 is used for the SM $H \rightarrow b\bar{b}$ BR [43].

7.2.2 | Modeling of the Background Processes

Production of a Vector Boson in Association with Jets

The production of W and Z bosons in association with jets represent the dominant background for the analysis. The simulation samples considers only processes where the vector boson decays leptonically, namely: $Z \rightarrow l^+l^-$, $W^- \rightarrow l^-\bar{\nu}$ or $W^+ \rightarrow l^+\nu$, and $Z \rightarrow \nu\bar{\nu}$ with $l = e, \mu, \tau$. To provide better statistical precision for W and Z boson production in association with heavy flavor quarks, the V +jets samples are generated applying different filters to select the flavor composition of the jets produced in association with the V boson. The filters used for these samples are outlined in Table 7.2.

Filter	Description
BFilter	at least 1 b -hadron with $p_T > 0$ GeV and $ \eta < 4$ at least 1 b -hadron with $p_T > 5$ GeV and $ \eta < 2.9$
CFilterBVeto	at least 1 c -hadron with $p_T > 4$ GeV and $ \eta < 3$ veto events which pass the BFilter
CVetoBVeto	veto events which pass the BFilter or the CFilterBVeto

Table 7.2: Heavy flavor filters used in the V +jets MC samples.

The V +jets backgrounds are subdivided into different components depending on the flavors of the two leading p_T jets in the event. The different sub-process considered are $V+bb$, $V+bc$, $V+bl$, $V+cc$, $V+cl$, and $V+ll$, with separate categories considered for $V = W/Z$ and b, c, l represent b -, c -, or light-jets.

Top Quark Backgrounds

The modeling of top-quark pair production within the phase space of the analysis differs between the 0-/1-lepton and 2-lepton channel due to the differing definitions of the V -boson system. The V -boson system in the 0-/1-lepton channel typically represents a single W -boson from a single top-quark decay, meanwhile in the 2-lepton channel the V -boson candidate is constructed from the l^+l^- -system, which corresponds to two leptons from two different top-quarks. This difference is reflected in the different flavor compositions of the $t\bar{t}$ process for the 0+1-lepton and 2-lepton channels, as the possibility of a hadronically decaying W -bosons in the 0-/1-lepton channel yields a larger charm flavor sub-component.

Both nominal and alternative samples are filtered at the event generation stage using truth level information, such that are required to have either one W boson decaying leptonically (referred to as *non-all-had*), or both W bosons decaying leptonically (referred to as *di-lepton*).

The production of a single-top quark is considered in the analysis through three sub-processes: s -channel, t -channel, and Wt -channel. Leading order Feynman diagrams for these processes are presented in Figure 7.1. Single top samples in the t - and s -channels are generated applying a lepton filter to require the leptonic decay of the W boson, while no such filter is applied in the Wt -channel samples.

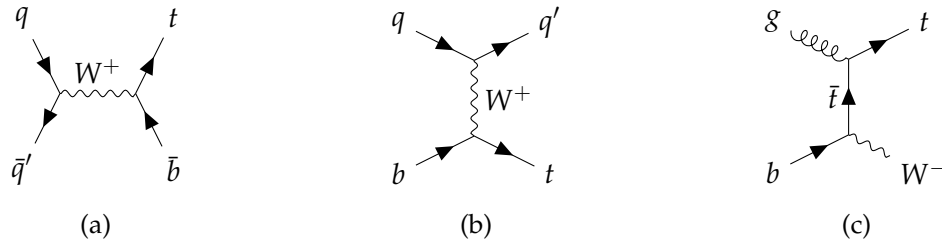


Figure 7.1: Leading order Feynman diagrams for the single-top process in (a) the s -channel, (b) the t -channel, and (c) the Wt -channel.

Vector Boson Pairs Production

The diboson contribution consists of final states arising from WW , WZ , and ZZ events. The processes that can give a significant contribution in the analysis phase space are those where one vector decays hadronically via $Z(\rightarrow c\bar{c})$ or $W(\rightarrow cl)$, and the other boson decays to leptons via $Z(\rightarrow \nu\bar{\nu})$, $W(\rightarrow l\nu)$, and $Z(\rightarrow l\bar{l})$. The topology of these diboson events is very similar to that of the $VH(\rightarrow c\bar{c})$ signal process. The main differ-

ence is that the invariant mass of the di-jet system is expected to peak under the Z or W mass, rather than m_H . Therefore the strength of the $VW(\rightarrow cl)$ and $VZ(\rightarrow c\bar{c})$ processes is measured simultaneously with $VH(\rightarrow c\bar{c})$ as a validation for the analysis strategy, as detailed in Chapter 10

7.2.3 | Corrections to Simulated Events

In order to properly model the pp collision data, each MC event is corrected by a series of associative calibration/correction scale factors, which include:

MC Event Weights: Generator dependent weights stemming from the matrix element amplitude/phase space calculation and parton shower procedure.

Luminosity Scaling: The effective luminosity of MC the sample is dictated by $\mathcal{L}_{MC}^{\text{eff}} = N_{MC}/\sigma_{MC}$, where N_{MC} is the number of produced MC events and σ_{MC} is the process specific cross section. Each event is therefore scaled to the luminosity of the data by applying a weight equal to $\mathcal{L}_{\text{data}}/\mathcal{L}_{MC}^{\text{eff}}$.

Pileup Reweighting: Event by event scale factor applied to correct the randomly generated average number of interactions per bunch crossing of each MC event, to that of the data distribution specific to each period (which are shown in Figure 3.4). The correction scale factors are applied using the CP::PileupReweightingTool documented in [111].

Flavor Tagging Scale Factors: MC-to-data scale factors designed to correct the efficiency of b -, c -, and light-jets as explained in Section 5.3.

Trigger Scale Factors: Correction scale factors designed to compensate for prescaling used during data collection, trigger efficiency variation, and data-to-MC efficiency correction. Full details can be found in [112].

Object Reconstruction Calibration Factors: Calibration scale factors for the reconstructed physics object described in Chapter 4, and used in the analysis (jets, electrons, muons, E_T^{miss}).

Selection and Categorization

This chapter describes the selection of $VH(\rightarrow c\bar{c})$ candidate events. Such events contain two c -jets from the decay of the Higgs boson, and depending on the leptonic channel, either a pair of charged leptons with same-flavor and opposite-sign (2-lepton), one charged lepton with additional E_T^{miss} (1-lepton), or a large amount of E_T^{miss} with zero charged leptons (0-lepton). Even though it was mentioned already, I remind the reader that in the context of this analysis a charged lepton refers to electrons and muons only.

The triggers used to select the physics objects comprising the events are described in Section 8.1, with further criteria applied to these objects detailed in Section 8.2. Section 8.3 focuses on the selection of events within each of the analysis channels, and their categorization into different signal and control regions is summarized in Section 8.4. Finally, the estimation of the QCD multijet background in each leptonic channel is discussed in Section 8.5.

All selection criteria in the analysis are summarized in Table 8.1

8.1 | Trigger Selection

Virtually all the events in the LHC contain jets, either from prompt partons in the collision or due to pileup jets entering the event. Leptons on the other hand, are much rarer and therefore it is much more efficient to use them for triggering. For this reason the $VH(cc)$ analysis makes use of E_T^{miss} , electron, and muon triggers to select candidate events, rather than jet triggers. In the following, a short summary of the triggers used to select events in each channel is given. For a detailed description the reader may refer to [15, 112].

Events in the 0-lepton channel are triggered using the lowest unscaled E_T^{miss} trigger in each data collection period. In the 1-lepton channel events where the charged lepton is an electron are triggered by the lowest unscaled single electron trigger in each data collection period. However, due to inefficiencies in single muon triggers, events containing a muon use the same E_T^{miss} triggers as the 0-lepton channel. In the 2-lepton channel events are selected using the lowest unscaled single electron or lowest unscaled single muon triggers in each data-taking period.

8.2 | Object Selection

Physics objects are reconstructed using the standard reconstruction procedures detailed in Chapter 4. For events to be considered in the analysis, the physics object they contain must satisfy additional requirements based on their identification, isolation, and kinematic properties, as specified below.

Two quality criteria are defined for leptons selected in the analysis, referred to as *VH-loose* leptons and *VH-signal* leptons. To qualify as *VH-loose* leptons, electron candidates are required to pass the *Loose* identification working point criteria, specified in Section 4.2.2, and the *FCLoose* isolation working point criteria defined in Section 4.2.3. In addition, selected electrons must have $p_T > 7$ GeV, $|\eta| < 2.47$, and to reject tracks from pileup they must also fulfill $d_0/\sigma_{d_0} < 5$ and $|\Delta z_0 \sin \theta| < 0.5$ mm. Similarly, muon candidates are required to pass the criteria imposed by the *Loose* identification and *FixedCutLoose* isolation working points described in Sections 4.3.2 and 4.3.3, and must have $p_T > 7$ GeV, $|\eta| < 2.7$, $d_0/\sigma_{d_0} < 3$ and $|\Delta z_0 \sin \theta| < 0.5$ mm.

The 1-lepton channel is susceptible to large contamination from QCD events where a jet is mistakenly identified as a lepton. Therefore more stringent requirements are considered for selecting *VH-signal* leptons, and in addition to the criteria described above electron candidates must also pass the *Tight* identification and *FixedCutHighPtCaloOnly* working points criteria, and have a p_T of at least 27 GeV. Muons are required to fulfill the criteria of *Medium* identification and *FixedCutHighPtTrackOnly* isolation working points, and have $p_T > 25$ GeV with $|\eta| < 2.5$.

Jets in the analysis are reconstructed in the *EMTopo* scheme (i.e. from topological calorimeter clusters in the EM scale, as explained in Section 4.4.2) using the anti- k_t algorithm with a radius parameter of $R = 0.4$. Jets are classified as either *signal jets* if they fall in the central $|\eta| < 2.5$ region, or *forward jets* within $2.5 < |\eta| < 4.5$. Only signal

jets are used for the reconstruction of the Higgs boson, and they must satisfy $p_T > 20$ GeV. Forward jets are only considered as additional jets in the event, which in the case of a $VH(cc)$ event may arise from initial-state (ISR) or final-state (FSR) radiation. As the forward region is more prone to pileup contamination forward jets are required to have $p_T > 30$ GeV. The JVT discriminant described in Section 4.4.4 is used to suppress jets arising from pileup in the central region. As such, signal jets are only considered if their JVT score is above 0.59.

A *muon-in-jet* correction is also applied to signal jets in case a reconstructed muon is found within $\Delta R = 0.4$ from the jet. This correction is targeted for heavy flavored jets with a muon from a semi-leptonic decay biasing the jet energy reconstruction. The muon-in-jet correction improves the $m_{c\bar{c}}$ resolution of the Higgs candidate by $\sim 6\%$ [15].

The missing transverse momentum E_T^{miss} plays a key role in the 0-lepton and 1-lepton channels, where the decay of the vector boson contains two or one neutrinos. E_T^{miss} is reconstructed as described in Section 4.6 from the negative vectorial sum of p_T of physics objects and an additional soft-term accounting for tracks associated with the PV but not with any of the objects in the event.

τ -leptons are reconstructed with the procedure described in Section 4.5. However, as the analysis does not consider τ -leptons as one of the physics objects in interest, events with a reconstructed τ -lepton passing the *Medium* identification working point, which corresponds to 60% efficiency for τ s having three prongs ($\pi^\pm$) and 75% efficiency for one prong, are not selected.

8.2.1 | Overlap Removal

The reconstruction of the various physics objects relies on information from several sub-detector systems, and as such the same signal might be used to build up different objects. To avoid any biases and double counting an overlap removal procedure is applied on the pre-selected objects, based on the following steps:

- **τ -electron:** If $\Delta R(\tau, e) < 0.2$, the τ -lepton is removed.
- **τ -muon:** If $\Delta R(\tau, \mu) < 0.2$, the τ -lepton is removed, with the exception that if the τ has $p_T > 50$ GeV and the muon is not a combined muon, then the τ -lepton is not removed.

- **electron-muon:** If a combined muon shares an ID track with an electron, the electron is removed. If a calorimeter-tagged muon shares an ID track with an electron, the muon is removed.
- **electron-jet:** If $\Delta R(\text{jet}, e) < 0.2$ the jet is removed. For any surviving jets, if $\Delta R(\text{jet}, e) < \min(0.4, 0.04 + 10 \text{ GeV}/p_T^e)$, the electron is removed.
- **muon-jet:** If $\Delta R(\text{jet}, \mu) < 0.2$ or the muon ID track is associated with the jet, then the jet is removed if the jet has less than three associated tracks with $p_T > 500$ MeV or both of the following conditions are met: the p_T ratio of the muon and jet is larger than 0.5 ($p_T^\mu/p_T^{\text{jet}} > 0.5$) and the ratio of the muon p_T to the sum of p_T of tracks with $p_T > 500$ MeV associated to the jet is larger than 0.7. For any surviving jets, if $\Delta R(\text{jet}, \mu) < \min(0.4, 0.04 + 10 \text{ GeV}/p_T^\mu)$, the muon is removed.
- **τ -jet:** If $\Delta R(\tau, \text{jet}) < 0.2$, the jet is removed.

8.3 | Event Selection and Categorization

Events passing the trigger selection and satisfying basic quality requirements are assigned to either the 0-lepton, 1-lepton, or 2-lepton channel, depending on the number of charged leptons in the event. In all three channels events are required to have at least two signal jets, the most energetic of which must have a $p_T > 45$ GeV. Events with additional jets, which can be either signal or forward are also considered and are used to define the following exclusive jet multiplicity categories: events containing only two signal jets comprise the 2-jet category, three jets the 3-jet category, and events with three or more jets make up the 3+-jet category. In the 0-lepton and 1-lepton channels, the 2- and 3-jet categories are used and events with four or more jets are rejected to suppress the background from $t\bar{t}$. In the 2-lepton channel, where the high jet multiplicity regions result in a 5 – 6% gain in the expected signal significance, the 2-jet and 3+-jet categories are used.

The Higgs candidate is reconstructed from the two leading p_T jets and its invariant mass, $m_{c\bar{c}}$, is required to be greater than 50 GeV. Events are categorized based on the number of c -tags among the two leading p_T jets: 0 c -tag, 1 c -tag, 2 c -tag. In addition, a b -tagging veto is applied on any c -tagged jet as well as on the third jet in the event, requiring for these jets to fail the b -tagging criteria used in the $VH(bb)$ analysis [113]. The b -tagging veto requirement is imposed in order to achieve orthogonality between the signal regions of the $VH(cc)$ and $VH(bb)$ analyses, as it ensures that the events

selected into the $VH(cc)$ signal region have less than 2 b -tagged jets. More details on the flavor-tagging strategy in the analysis and on the b -tagging veto in particular are given in Section 9.1.

The reconstruction of the vector boson depends on the specific leptonic channel. In the 0-lepton channel the 4-vector of the Z boson is taken simply as the missing transverse momentum vector E_T^{miss} , in the 1-lepton channel the W boson is defined as the vectorial sum of the E_T^{miss} and the reconstructed lepton, and in the 2-lepton channel the Z boson corresponds to the di-lepton system. The best signal sensitivity is obtained in regions of high p_T of the vector boson, denoted p_T^V . Therefore a single p_T^V category is considered in the 0-lepton and 1-lepton channel for $p_T^V > 150$ GeV, referred to as the high- p_T^V region. In the 2-lepton channel, in addition to the high- p_T^V region, the signal sensitivity is increased by including a low- p_T^V region with $75 \text{ GeV} < p_T^V < 150 \text{ GeV}$. This p_T^V categorization is used in the definition of the signal regions considered in the final fit. However, in a number of auxiliary studies conducted during the development stages of the analysis a different p_T^V categorization was adopted in which the $p_T^V > 150$ GeV region was further split at $p_T^V = 250$ GeV. Therefore, the reader may come across specific places in this thesis where the definition of the high- p_T^V region refers to $p_T^V > 250$ GeV, and an additional medium- p_T^V category designates the region of $150 \text{ GeV} < p_T^V < 250 \text{ GeV}$. Whenever this is the case I will mention it explicitly to avoid confusion.

Two jets originating from a Higgs or vector boson resonance, or produced via gluon splitting are expected to have a lower ΔR value in the p_T^V regimes considered in the analysis compared to the majority of background processes. Therefore, the angular separation between the two leading jets, $\Delta R(j_1, j_2)$, can be used to further reduce the level of background. The upper selection of $\Delta R(j_1, j_2)$ was optimized based on the $VH(cc)$ signal significance, defined as $\frac{S}{\sqrt{S+B}}$, with S indicating signal and B indicating background events. The $\Delta R(j_1, j_2)$ optimization was performed separately in the low-, medium, and high- p_T^V regions, and yielded the following selection criteria:

- low- p_T^V ($75 \text{ GeV} < p_T^V < 150 \text{ GeV}$) : $\Delta R(j_1, j_2) \leq 2.3$
- medium- p_T^V ($150 \text{ GeV} < p_T^V < 250 \text{ GeV}$) : $\Delta R(j_1, j_2) \leq 1.6$
- high- p_T^V ($p_T^V > 250 \text{ GeV}$) : $\Delta R(j_1, j_2) \leq 1.2$

8.3.1 | 0-lepton Channel Specific Selection

Events in the 0-lepton channel must have 0 VH-loose leptons. Further requirements are applied on the scalar sum of p_T (S_T) of the jets in the event. For 2-jet events S_T must be greater than 120 GeV, and for 3-jet events S_T must be greater than 150 GeV. To remove QCD background events that enter the 0-lepton channel due to jet energy mis-measurements, a selection is applied based on the azimuthal difference between the E_T^{miss} and the nearest jet. This is motivated by the fact that fake E_T^{miss} tends to be aligned with the mis-measured jet. The following requirements, referred to as *anti-QCD* cuts are thus applied:

- $|\Delta\phi(\vec{E}_T^{\text{miss}}, \vec{p}_T^{\text{miss}})| < 90^\circ$
- $|\Delta\phi(\text{jet}_1, \text{jet}_2)| < 140^\circ$
- $|\Delta\phi(\vec{E}_T^{\text{miss}}, h)| > 120^\circ$, with h indicating the reconstructed Higgs candidate.
- $\min(|\Delta\phi(\vec{E}_T^{\text{miss}}, \text{jets})|) > 20^\circ$ for 2-jet events, $> 30^\circ$ for 3-jet events.

The remaining multijet fraction after the implementation of the anti-QCD cuts is at the level of less than 1%, and therefore no multijet background component is considered in the 0-lepton channel.

To remove the contribution of non-collision background, an additional cut is applied requiring the missing transverse momentum to satisfy $p_T^{\text{miss}} > 30$ GeV.

8.3.2 | 1-lepton Channel Specific Selection

In addition to the criteria common to the 0-, 1-, and 2-lepton channels, additional event selection are implemented in the 1-lepton channel. To reduce the contamination of QCD multijet events, a tighter requirement is applied on the identification and isolation of selected leptons, and events are required to have exactly one VH-signal lepton (e or μ) and no additional VH-loose leptons. The QCD background is further reduced by only considering the $p_T^V > 150$ GeV region and by applying a cut on $E_T^{\text{miss}} > 30$ GeV for events where the lepton is an electron. An addition criteria is applied on the transverse mass of the W -boson candidate, defined as $m_T^W = \sqrt{2p_T^l E_T^{\text{miss}} (1 - \cos(\Delta\phi(l, E_T^{\text{miss}}))}$, which is required to be smaller than 120 GeV.

8.3.3 | 2-lepton Channel Specific Selection

Selected events in the 2-lepton channel are required to have exactly two VH-loose leptons of the same flavor, and at least one of which must also satisfy the VH-signal lepton requirement. For selected di-muon events the two muons must also be of opposite sign (i.e. μ^- and μ^+). A similar criteria is not applied for di-electron events due to the higher rate of charge mis-identification. To suppress backgrounds with a non-resonant lepton pair such as di-leptonic $t\bar{t}$, the invariant mass of the di-lepton system m_{ll} is required to be in the $81 \text{ GeV} < m_{ll} < 101 \text{ GeV}$ range consistent with a Z-boson.

Common Selections	
Jets	≥ 2 signal jets
c -tagging	1 or 2 c -tagged signal jets
b -tagging veto	≥ 3 signal or forward jets
Leading jet p_T	$> 45 \text{ GeV}$
$m_{c\bar{c}}$	$> 50 \text{ GeV}$
$\Delta R(j_1, j_2)$	$< \begin{cases} 2.3 : 75 < p_T^V < 150 \text{ GeV} \\ 1.6 : 150 < p_T^V < 250 \text{ GeV} \\ 1.2 : p_T^V > 250 \text{ GeV} \end{cases}$
0-lepton	
Trigger	lowest un-prescaled E_T^{miss} trigger
Leptons	0 VH-loose leptons
E_T^{miss}	$> 150 \text{ GeV}$
p_T^{miss}	$> 30 \text{ GeV}$
S_T	> 120 (2-jet), $> 150 \text{ GeV}$ (3-jet)
$ \Delta\phi(E_T^{\text{miss}}, p_T^{\text{miss}}) $	$< 90^\circ$
$\min(\Delta\phi(E_T^{\text{miss}}, \text{jets}))$	$> 20^\circ$ (2-jet), $> 30^\circ$ (3-jet)
$ \Delta\phi(E_T^{\text{miss}}, h) $	$> 120^\circ$
$ \Delta\phi(\text{jet}_1, \text{jet}_2) $	$< 140^\circ$
p_T^V regions	$[150, \infty] \text{ GeV}$
1-lepton	
Trigger	e channel: un-prescaled single electron μ channel: above see 0-lepton triggers
Leptons	1 VH-signal lepton > 1 VH-loose lepton veto
E_T^{miss}	$> 30 \text{ GeV}$ (e channel only)
m_T^W	$< 120 \text{ GeV}$
p_T^V regions	$[150, \infty] \text{ GeV}$
2-lepton	
Trigger	un-prescaled single lepton
Leptons	2 VH-loose leptons (≥ 1 VH-signal lepton)
m_{ll}	Same flavor, opposite-charge for $\mu\mu$ $\in [81, 101] \text{ GeV}$
p_T^V regions	$[75, 150], [150, \infty] \text{ GeV}$

Table 8.1: Summary of the signal region event selection in the 0-, 1- and 2-lepton channels.

8.4 | Signal Regions and Control Regions

The selection criteria described above and summarized in Table 8.1 are used to define sixteen different signal regions. In addition, twenty-eight control regions (CRs) are defined to better constrain the modeling of the background processes, and are described in the following. A summarized description of the signal and control regions used in the analysis is given in Table 8.2, and the pre-fit event yields in each region for the three channels are shown in Tables 8.3-8.7.

8.4.1 | 0 c -tag Control Region

As the name suggests, the 0 c -tag control region is defined simply by selecting events where none of the two leading jets are c -tagged. Imposing this selection results in a region with high purity of V +light flavored jets events. The 0 c -tag CR is considered only for the 1-lepton and 2-lepton channels. The former is dominated by the W +light jets background, and the latter is enriched with Z +light jets events.

8.4.2 | Top Control Region

The contribution of the $t\bar{t}$ background in the analysis signal region is largely reduced due to the requirement of having events with less than two b -tagged jets. However, in the 3-jet category in the 0- and 1-lepton channels, as well as in the low- p_T^V region in the 2-lepton channel, a non-negligible $t\bar{t}$ and single-top component remains. Therefore, a dedicated top control region is defined in which the modeling of these processes can be better determined. In the 0- and 1-lepton channels, where the $t\bar{t}$ component is mainly from semi-leptonic decays, the control region is considered only for 3-jet events. It is defined by inverting the b -tagging veto requirement on the third jet in the event (i.e. by selecting events in which the third leading p_T jet is b -tagged). Due to limited statistics in events with 2 c -tagged jets, this control region comprises of only the 1 c -tag category.

In the 2-lepton channel the top background is coming mainly from di-leptonic $t\bar{t}$ events. A high purity $t\bar{t}$ control region can be defined by selecting events with different flavored leptons ($e\mu$ or μe). This eliminates most of the backgrounds and the signal, which are all expected to contain same flavor lepton pairs from a Z boson decay. The $e\mu$ control region in the 2-lepton channel comprised of only 1 c -tag events. For each p_T^V and jet multiplicity category a dedicated $e\mu$ control region is defined.

8.4.3 | High ΔR Control Region

The High ΔR control region is designed to constrain the V +jets background normalization and shape. In particular, as the typical shape of the $m_{c\bar{c}}$ distribution in the analysis signal region is steeply falling after the Higgs mass, this control region is aimed to help with the modeling near the Higgs mass and at high $m_{c\bar{c}}$. For each signal region a corresponding high ΔR CR is defined using the same selection except for $\Delta R(j_1, j_2)$, which extends from the signal region to $\Delta R(j_1, j_2) = 2.5$. The motivation for applying this cut-off is that from $\Delta R(j_1, j_2) > 2.5$ the $m_{c\bar{c}}$ distribution in the CR is starting to drop off and the statistics becomes limited.

8.5 | Estimation of the Multijet Background

Background from QCD-induced multijet events arises either from jet energy mis-measurements, or from "fake" leptons. The former can lead to large E_T^{miss} values and is therefore the dominant multijet source in the 0-lepton channel. The latter is the result of hadronic jets mistakenly identified as leptons and/or non-prompt electrons and muons from semi-leptonic decays of heavy flavored hadron. Due to the difficulty of accurately modeling QCD processes in MC simulation, data-driven approaches are used to estimate this background in each of the analysis channels. In the 0-lepton and 2-lepton channels, the analysis selection rejects the vast majority of multijet events, and the estimation is used to verify that the multijet contribution is indeed negligible and can be neglected in the final fit. The level of multijet contamination in the 1-lepton channel, although much smaller than the dominant V +jets process, is not negligible compared to some of the other backgrounds. Therefore a robust procedure is necessary to estimate the multijet background component in the 1-lepton channel, which is included in the final fit along with the MC-simulated electroweak backgrounds.

The estimation of the multijet background was one of my personal responsibilities in the analysis. I took part in the development of the methods described below, and was in charge of their implementation and validation within the analysis framework.

8.5.1 | Multijet Background in the 0-lepton Channel

The anti-QCD cuts specified in Section 8.3.1 greatly suppress the multijet background in the 0-lepton channel. To evaluate the remaining contribution and ensure that this background can be safely neglected in the fit, a study is performed by removing the

Region description	Channel	p_T^V	$\Delta R(j_1, j_2)$	c -tags	Jet multiplicity
SR	0-lepton	$p_T^V > 150$ GeV	$\Delta R(j_1, j_2) < \begin{cases} 1.6 : p_T^V < 250 \text{ GeV} \\ 1.2 : p_T^V > 250 \text{ GeV} \end{cases}$	1 c -tags	2-jet 3-jet
				2 c -tags	2-jet 3-jet
	1-lepton	$p_T^V > 150$ GeV	$\Delta R(j_1, j_2) < \begin{cases} 1.6 : p_T^V < 250 \text{ GeV} \\ 1.2 : p_T^V > 250 \text{ GeV} \end{cases}$	1 c -tags	2-jet 3-jet
				2 c -tags	2-jet 3-jet
	2-lepton	$75 < p_T^V < 150$ GeV	$\Delta R(j_1, j_2) < 2.3$	1 c -tags	2-jet 3+-jet
				2 c -tags	2-jet 3+-jet
		$p_T^V > 150$ GeV	$\Delta R(j_1, j_2) < \begin{cases} 1.6 : p_T^V < 250 \text{ GeV} \\ 1.2 : p_T^V > 250 \text{ GeV} \end{cases}$	1 c -tags	2-jet 3+-jet
				2 c -tags	2-jet 3+-jet
0 c -tag CR	1-lepton	$p_T^V > 150$ GeV	$\Delta R(j_1, j_2) < \begin{cases} 1.6 : p_T^V < 250 \text{ GeV} \\ 1.2 : p_T^V > 250 \text{ GeV} \end{cases}$	0 c -tags	2-jet 3-jet
				0 c -tags	2-jet 3+-jet
	2-lepton	$p_T^V > 150$ GeV	$\Delta R(j_1, j_2) < \begin{cases} 1.6 : p_T^V < 250 \text{ GeV} \\ 1.2 : p_T^V > 250 \text{ GeV} \end{cases}$	0 c -tags	2-jet 3-jet
				0 c -tags	2-jet 3-jet
Top CR (b-tagged 3rd jet)	0-lepton	$p_T^V > 150$ GeV	$\Delta R(j_1, j_2) < \begin{cases} 1.6 : p_T^V < 250 \text{ GeV} \\ 1.2 : p_T^V > 250 \text{ GeV} \end{cases}$	1 c -tags	3-jet
	1-lepton	$p_T^V > 150$ GeV	$\Delta R(j_1, j_2) < \begin{cases} 1.6 : p_T^V < 250 \text{ GeV} \\ 1.2 : p_T^V > 250 \text{ GeV} \end{cases}$	1 c -tags	3-jet
Top CR (diff. flavor leptons)	2-lepton	$75 < p_T^V < 150$ GeV	$\Delta R(j_1, j_2) < 2.3$	1 c -tags	2-jet 3+-jet
		$p_T^V > 150$ GeV	$\Delta R(j_1, j_2) < \begin{cases} 1.6 : p_T^V < 250 \text{ GeV} \\ 1.2 : p_T^V > 250 \text{ GeV} \end{cases}$	0 c -tags	2-jet 3-jet
High ΔR CR	0-lepton	$p_T^V > 150$ GeV	$\Delta R(j_1, j_2) \in \begin{cases} [1.6, 2.5] : p_T^V < 250 \text{ GeV} \\ [1.2, 2.5] : p_T^V > 250 \text{ GeV} \end{cases}$	1 c -tags	2-jet 3-jet
				2 c -tags	2-jet 3-jet
	1-lepton	$p_T^V > 150$ GeV	$\Delta R(j_1, j_2) \in \begin{cases} [1.6, 2.5] : p_T^V < 250 \text{ GeV} \\ [1.2, 2.5] : p_T^V > 250 \text{ GeV} \end{cases}$	1 c -tags	2-jet 3-jet
				2 c -tags	2-jet 3-jet
	2-lepton	$75 < p_T^V < 150$ GeV	$\Delta R(j_1, j_2) \in [2.3, 2.5]$	1 c -tags	2-jet 3+-jet
				2 c -tags	2-jet 3+-jet
		$p_T^V > 150$ GeV	$\Delta R(j_1, j_2) \in \begin{cases} [1.6, 2.5] : p_T^V < 250 \text{ GeV} \\ [1.2, 2.5] : p_T^V > 250 \text{ GeV} \end{cases}$	1 c -tags	2-jet 3+-jet
				2 c -tags	2-jet 3+-jet

Table 8.2: Summary of the analysis signal and control regions.

cut on $\min(|\Delta\phi(E_T^{\vec{miss}}, \text{jets})|)$ and fitting the data and MC background distributions of this variable. As the data distribution is steeply falling at low $(|\Delta\phi(E_T^{\vec{miss}}, \text{jets})|)$, an ex-

	1 tag 2 jet SR	1 tag 3 jet SR	2 tag 2 jet SR	2 tag 3 jet SR	1 tag 2 jet dRCR	1 tag 3 jet dRCR	2 tag 2 jet dRCR	2 tag 3 jet dRCR	1 tag 3 jet topCR
data	65490	58212	1462	1432	54178	49795	1180	1159	4656
VHcc	7.64 ±0.88	5.88 ±0.72	1.06 ±0.19	0.72 ±0.13	1.70 ±0.21	1.80 ±0.27	0.15 ±0.03	0.13 ±0.03	0.15 ±0.02
Total Bkg.	56641 ±6191	52475 ±4653	1241 ±149	1298 ±140	43820 ±2864	41784 ±3076	888 ±81	917 ±85	4706 ±319
VHbb	58 ±12	41.2 ±8.6	2.48 ±0.53	1.60 ±0.33	11.9 ±2.5	10.9 ±2.4	0.45 ±0.10	0.35 ±0.09	4.03 ±0.88
W+jets	21411 ±3314	18771 ±3121	372 ±78	351 ±77	20204 ±2184	17741 ±2096	327 ±52	294 ±42	359 ±69
Wmftau	9560 ±2228	8919 ±2169	142 ±39	136 ±38	8152 ±977	8279 ±1452	152 ±35	143 ±31	237 ±63
Wlf	10412 ±2407	8475 ±2196	79 ±25	60 ±22	11451 ±1952	8819 ±1509	109 ±36	74 ±24	102 ±27
Whf	1438 ±473	1377 ±455	150 ±63	155 ±63	601 ±80	644 ±90	66 ±14	77 ±16	19.5 ±8.7
Z+jets	28711 ±5201	22118 ±3321	630 ±124	531 ±109	21432 ±1844	19241 ±2233	487 ±61	451 ±72	573 ±102
Zl	11469 ±4248	8529 ±2013	77 ±35	59 ±22	8179 ±1296	7208 ±1173	56 ±17	51 ±15	112 ±32
Zmf	14005 ±2910	11048 ±2573	208 ±54	164 ±47	10888 ±1277	9934 ±1855	156 ±29	144 ±35	414 ±97
Zhf	3237 ±735	2542 ±599	345 ±106	307 ±96	2365 ±300	2099 ±412	275 ±51	256 ±61	47 ±12
ttbar	3802 ±505	9192 ±924	138 ±22	336 ±41	1340 ±162	3573 ±281	53.6 ±8.6	140 ±17	3680 ±294
Diboson	2477 ±183	2157 ±178	96 ±11	74.2 ±8.0	471 ±38	841 ±68	10.6 ±1.7	19.9 ±2.3	46.6 ±4.6
single top	181 ±13	196 ±22	2.7 ±1.1	5.0 ±1.1	361 ±34	376 ±29	9.1 ±1.9	12.36 ±0.90	45.6 ±5.6

Table 8.3: Pre-fit event yields in the 0-lepton channel. For each region, the entries highlighted in color represent process with large relative event yields.

	1 tag 2 jet SR	1 tag 3 jet SR	2 tag 2 jet SR	2 tag 3 jet SR	0 tag 2 jet 0tagCR	0 tag 3 jet 0tagCR	1 tag 2 jet dRCR	1 tag 3 jet dRCR	2 tag 2 jet dRCR	2 tag 3 jet dRCR	1 tag 3 jet topCR
data	83947	86316	1897	2277	996916	920678	61582	64084	1168	1425	13797
VHcc	7.23 ±0.81	5.47 ±0.79	1.01 ±0.17	0.68 ±0.13	13.0 ±1.4	10.8 ±1.5	1.54 ±0.17	1.48 ±0.23	0.15 ±0.03	0.11 ±0.03	0.14 ±0.03
Total Bkg.	74930 ±8863	83617 ±8451	1691 ±238	2269 ±256	960576 ±31692	931939 ±56337	52398 ±3326	56713 ±4907	1008 ±96	1339 ±116	14625 ±882
VHbb	57 ±15	39 ±11	2.33 ±0.62	1.57 ±0.44	337 ±89	245 ±68	11.0 ±2.9	8.9 ±2.5	0.41 ±0.12	0.28 ±0.09	3.56 ±1.00
W+jets	54649 ±8653	46570 ±7943	1056 ±227	985 ±219	809064 ±30370	671734 ±53991	42331 ±3144	39974 ±4814	729 ±92	740 ±106	909 ±171
Wmftau	32081 ±7056	27242 ±6647	479 ±127	419 ±120	123569 ±7958	104598 ±13634	26318 ±2583	24901 ±4379	387 ±71	379 ±88	650 ±161
Wlf	18398 ±4825	15602 ±4185	125 ±47	110 ±42	674206 ±29293	557823 ±52231	13955 ±1779	12963 ±1978	101 ±25	93 ±26	196 ±52
Whf	4171 ±1340	3726 ±1182	453 ±182	457 ±179	11289 ±967	9312 ±1022	2058 ±220	2110 ±282	241 ±54	268 ±54	63 ±23
Z+jets	1477 ±138	1470 ±167	30.1 ±5.6	33.3 ±5.7	24402 ±1432	23675 ±2803	1241 ±119	1384 ±133	32.8 ±4.6	30.0 ±3.8	27.7 ±4.7
ttbar	12922 ±1568	30615 ±2787	484 ±70	1133 ±130	75751 ±8746	183903 ±15701	4805 ±467	12379 ±686	196 ±24	496 ±47	13532 ±865
Diboson	2966 ±284	2502 ±258	82.8 ±9.3	72.1 ±8.0	21393 ±1496	19719 ±1708	490 ±47	976 ±89	10.1 ±1.4	22.9 ±2.8	44.4 ±4.8
MultiJet	2446 ±1059	1885 ±683	22 ±15	26 ±21	26287 ±1083	28015 ±1183	2699 ±970	978 ±635	18.7 ±6.9	15.1 ±6.6	
single top	413 ±35	537 ±71	12.8 ±1.4	18.0 ±2.3	3342 ±181	4648 ±302	820 ±64	1013 ±68	20.7 ±4.4	34.6 ±3.5	107 ±14

Table 8.4: Pre-fit event yields in the 1-lepton channel. For each region, the entries highlighted in color represent process with large relative event yields.

	1 tag 2 jet high SR	1 tag 2 jet low SR	1 tag 3 jet high SR	1 tag 3 jet low SR	2 tag 2 jet high SR	2 tag 2 jet low SR	2 tag 3 jet high SR	2 tag 3 jet low SR
data	7074	42448	10812	39808	189	1133	302	1009
VHcc	1.53 ± 0.20	2.90 ± 0.38	2.12 ± 0.51	3.09 ± 0.70	0.22 ± 0.04	0.39 ± 0.08	0.26 ± 0.07	0.33 ± 0.09
Total Bkg.	6686 ± 1084	37892 ± 6144	9953 ± 1443	37508 ± 6092	168 ± 29	968 ± 173	256 ± 51	967 ± 206
VHbb	13.2 ± 3.3	24.5 ± 6.1	15.8 ± 4.9	21.9 ± 6.8	0.57 ± 0.15	1.00 ± 0.26	0.62 ± 0.20	0.81 ± 0.26
W+jets	4.17 ± 0.77	23 ± 12	6.2 ± 3.0	31.9 ± 8.3		0.15 ± 0.05	0.05 ± 0.02	
Z+jets	6319 ± 1084	35689 ± 6143	9309 ± 1442	34541 ± 6089	152 ± 29	873 ± 172	232 ± 50	862 ± 205
Zl	2351 ± 846	13176 ± 4979	3410 ± 739	12461 ± 3028	16.4 ± 7.2	95 ± 41	24.7 ± 8.3	91 ± 30
Zmf	3195 ± 654	18097 ± 3392	4745 ± 1203	17952 ± 5087	48 ± 12	265 ± 60	71 ± 21	274 ± 85
Zhf	773 ± 176	4416 ± 1200	1155 ± 298	4129 ± 1428	87 ± 25	513 ± 156	137 ± 45	497 ± 185
ttbar	29.9 ± 3.2	1025 ± 85	98.2 ± 5.6	1603 ± 116	1.16 ± 0.12	41.6 ± 4.6	3.15 ± 0.30	57.5 ± 6.6
Diboson	312 ± 29	1054 ± 93	508 ± 52	1220 ± 129	14.9 ± 2.1	49.8 ± 9.3	20.1 ± 2.7	44.2 ± 6.0
single top	7.20 ± 0.93	74.6 ± 6.8	15.8 ± 1.6	89.8 ± 8.9	0.04 ± 0.03	2.87 ± 0.76	0.27 ± 0.08	2.26 ± 0.35

Table 8.5: Pre-fit event yields in the 2-lepton signal regions. The notions "high" and "low" in the title row refer to the high- and low- p_T^V regions. For each region, the entries highlighted in color represent process with large relative event yields.

ponential function is chosen to model the multijet background. The normalization and shape parameters of the multijet component are determined in the fit. Because of the pre-fit normalization differences between EW MC backgrounds and data, an additional floating parameter accounting for the Z/W +jets background normalization is used. A fit is performed in each of the 0-lepton signal regions, and two examples are shown in Figure 8.1(a) for the 1 c -tag 2-jet and Figure 8.1(b) for the 1 c -tag 3-jet regions. The plots in Figure 8.1 are generated using data collected between the years 2015-2016, and correspond only to the medium- p_T^V region between $150 < p_T^V < 250$ GeV. In the 2-jet category the plot shows very few events at low $(|\Delta\phi(E_T^{\vec{miss}}, \text{jets})|)$, which implies that the vast majority of multijet events are removed by the other anti-QCD cuts. In the 3-jet category the multijet contribution occupies the low region of $\min(|\Delta\phi(E_T^{\vec{miss}}, \text{jets})|)$, as evident by the excess of data events compared to the MC backgrounds. The red curve in the plots represents the sum of the evaluated multijet background and the EW MC background, scaled by the normalization factor extracted from the fit. In Figure 8.1(b), one can identify an additional red curve, corresponding to the exponential function describing the multijet background. Because of the minuscule number of multijet events in the

	1 tag 2 jet high dRCR	1 tag 2 jet low dRCR	1 tag 3 jet high dRCR	1 tag 3 jet low dRCR	2 tag 2 jet high dRCR	2 tag 2 jet low dRCR	2 tag 3 jet high dRCR	2 tag 3 jet low dRCR
data	5848	7076	10132	7858	107	187	267	208
VHcc	0.38 ± 0.06	0.33 ± 0.04	0.85 ± 0.25	0.38 ± 0.08	0.04 ± 0.01	0.04 ± 0.01	0.05 ± 0.02	0.03 ± 0.01
Total Bkg.	5216 ± 389	6201 ± 409	9264 ± 1204	7157 ± 1114	131 ± 15	163 ± 19	230 ± 45	183 ± 42
VHbb	3.32 ± 0.84	2.69 ± 0.67	5.3 ± 1.8	2.44 ± 0.73	0.12 ± 0.03	0.14 ± 0.04	0.16 ± 0.06	0.08 ± 0.03
W+jets	3.49 ± 0.69	3.1 ± 1.5	2.74 ± 0.99	4.3 ± 1.6		0.02 ± 0.01	0.07 ± 0.04	0.01 ± 0.00
Z+jets	5105 ± 389	5929 ± 409	8850 ± 1203	6637 ± 1114	127 ± 15	153 ± 19	220 ± 45	166 ± 42
Zl	1861 ± 243	2058 ± 257	3113 ± 457	2262 ± 377	13.9 ± 3.4	14.9 ± 3.7	23.1 ± 6.4	16.8 ± 4.6
Zmf	2611 ± 295	3093 ± 307	4720 ± 1073	3592 ± 1003	38.7 ± 6.8	46.5 ± 7.7	72 ± 19	57 ± 17
Zhf	634 ± 71	778 ± 83	1017 ± 297	783 ± 301	75 ± 13	92 ± 17	125 ± 40	92 ± 38
t $\bar{t}$ bar	33.7 ± 3.4	188 ± 15	92.3 ± 5.2	339 ± 16	1.20 ± 0.12	7.87 ± 0.71	3.05 ± 0.30	12.18 ± 0.75
Diboson	64.0 ± 7.0	63.5 ± 5.5	301 ± 32	159 ± 17	1.50 ± 0.22	1.89 ± 0.26	6.84 ± 0.93	4.16 ± 0.52
single top	7.02 ± 0.76	14.8 ± 2.0	13.4 ± 1.3	16.4 ± 1.3	0.37 ± 0.25	0.16 ± 0.10	0.39 ± 0.08	0.46 ± 0.06

Table 8.6: Pre-fit event yields in the 2-lepton high ΔR control regions. The notions "high" and "low" in the title row refer to the high- and low- p_T^V regions. For each region, the entries highlighted in color represent process with large relative event yields.

2-jet category, a similar curve can not be seen in Figure 8.1(a). The χ^2 and Kolmogorov-Smirnov (KS) goodness-of-fit test values are also specified on the plot. Only statistical uncertainties are considered for these fits.

The amount of multijet events in the 3-jet category is then evaluated by integrating the fitted multijet parameters in the $\min(|\Delta\phi(E_T^{\text{miss}}, \text{jets})|) > 30^\circ$ region, and it is found to be less than 0.1% of the total background in each of the signal regions.

As a final test, the $m_{c\bar{c}}$ shape of the multijet background is estimated by subtracting the electroweak MC from the data in a region defined by $\min(|\Delta\phi(E_T^{\text{miss}}, \text{jets})|) < 25^\circ$. This region effectively acts as a control region as it is orthogonal to the analysis signal region and is enriched with multijet events, as seen in Figure 8.1(b). The shape of the multijet $m_{c\bar{c}}$ distribution is compared to that of Z+jets and W+jets (in the signal region). Figure 8.2 shows an example of such comparison conducted in the 1 c -tag 3-jet region. This test is used as validation that the shapes of the multijet and electroweak backgrounds are similar. Si and thus any remaining multijet contamination can be absorbed by these backgrounds in the main fit.

	1 tag 2 jet high topCR	1 tag 2 jet low topCR	1 tag 3 jet high topCR	1 tag 3 jet low topCR	0 tag 2 jet high 0tagCR	0 tag 2 jet low 0tagCR	0 tag 3 jet high 0tagCR	0 tag 3 jet low 0tagCR
data	30	1122	101	1727	109028	614935	156891	562101
VHcc					2.70 ±0.34	5.24 ±0.64	4.4 ±1.1	6.8 ±1.4
Total Bkg.	36.8 ±2.7	1075 ±84	104.3 ±5.1	1634 ±113	102591 ±4007	557517 ±16317	148097 ±10676	535293 ±61379
VHbb		0.02 ±0.01	0.01 ±0.00	0.02 ±0.01	77 ±19	145 ±36	99 ±31	144 ±44
W+jets	1.51 ±0.38	5.0 ±1.0	4.8 ±2.0	15.2 ±4.7	50.5 ±4.7	205 ±29	92.5 ±7.0	373 ±35
Z+jets	1.54 ±0.28	4.6 ±1.2	2.23 ±0.35	8.2 ±2.0	100039 ±4003	542499 ±16301	142757 ±10668	512728 ±61365
Zl	0.40 ±0.13	3.0 ±1.2	0.67 ±0.16	2.68 ±0.87	82625 ±3735	447965 ±14703	117066 ±9845	421525 ±57961
Zmf	0.98 ±0.25	0.88 ±0.29	1.40 ±0.31	3.8 ±1.6	15280 ±1413	82872 ±6854	22642 ±4064	80512 ±19917
Zhf	0.17 ±0.04	0.72 ±0.15	0.16 ±0.06	1.64 ±0.78	2134 ±281	11662 ±1600	3049 ±590	10691 ±3093
t $\bar{t}$ bar	27.6 ±2.6	996 ±84	86.0 ±4.6	1536 ±112	196 ±19	6208 ±452	723 ±33	10949 ±632
Diboson	0.02 ±0.00	0.28 ±0.07	0.10 ±0.02	0.88 ±0.17	2175 ±178	7869 ±566	4326 ±416	10418 ±1125
single top	6.12 ±0.56	69.9 ±5.9	11.14 ±0.93	74 ±11	53.6 ±3.9	590 ±28	99.0 ±6.2	681 ±61

Table 8.7: Pre-fit event yields in the 2-lepton 0 c -tag and Top CRs. The notions "high" and "low" in the title row refer to the high- and low- p_T^V regions. For each region, the entries highlighted in color represent process with large relative event yields.

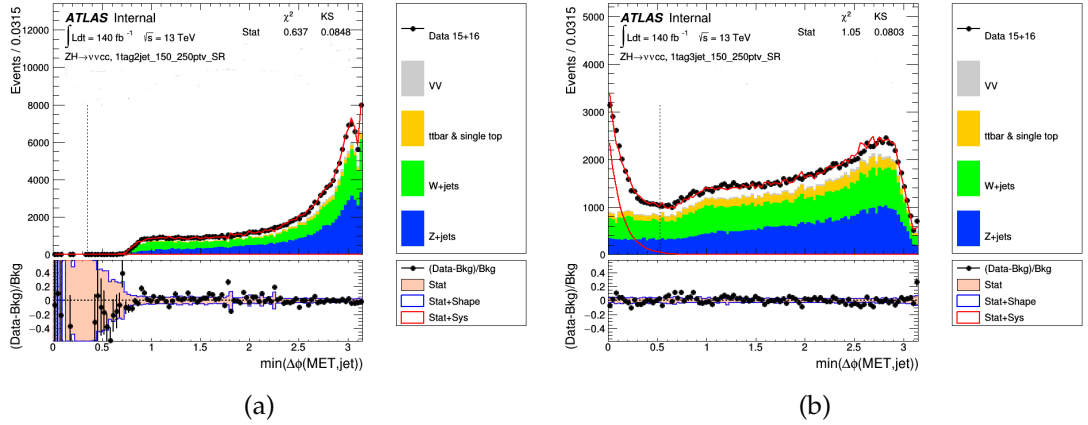


Figure 8.1: Postfit distributions of $\min(|\Delta\phi(E_T^{\text{miss}}, \text{jets})|)$ for the data and electroweak MC backgrounds in the 1 c -tag 2-jet (a) and 1 c -tag 3-jet regions (b). An exponential model is used to estimate the contribution of the QCD multijet process. The dashed gray vertical line indicates the cut value of 20° (2-jet) and 30° (3-jet) below which events in the 0-lepton channel are rejected.

Similar plots to Figures 8.1 and 8.2 showing the remaining signal regions, as well as additional information on the results of the 0-lepton multijet estimate can be found in

the internal documentation of the analysis [15].

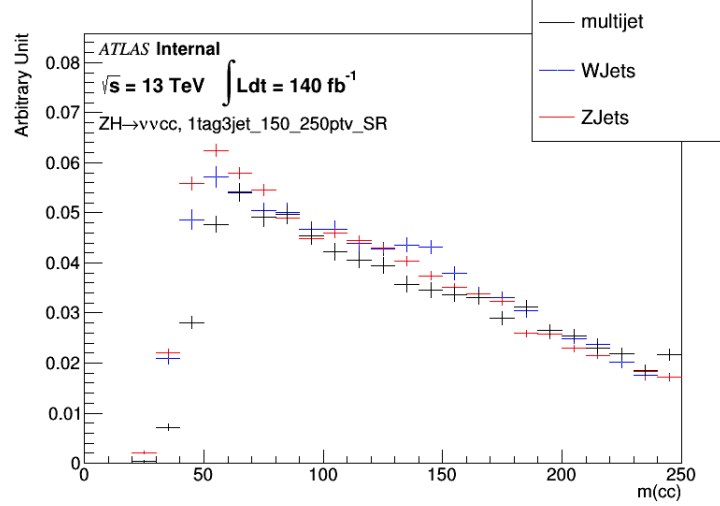


Figure 8.2: Comparison of the multijet (black), Z+jets (red), and W+jets (blue) $m_{c\bar{c}}$ shape in the 1 c -tag 3-jet region of the 0-lepton channel.

8.5.2 | Multijet Background in the 1-lepton Channel

In the 1-lepton channel the contribution of the QCD background is greatly reduced by only considering the $p_T^V > 150 \text{ GeV}$ region, and by imposing, on top of the VH-loose lepton definition used to separate events between the 0-, 1-, and 2-lepton channels, additional tighter identification and isolation criteria on the selected leptons. Nevertheless, the multijet contamination in this channel is at a level that can not be neglected and ignored in the final fit, and thus a robust procedure is used to estimate both the normalization and shape of this background component.

The contribution of the multijet background in the analysis signal region is estimated using data in a multijet enriched control region (MJ-CR), defined by inverting the isolation cuts on the selected leptons. The MJ-CR is used only for the purpose of the multijet background estimation and is not included in the analysis final fit. Table 8.8 summarizes the isolation cuts applied on electrons and muons in both the signal region and the MJ-CR (also referred to as *inverted isolation region*). Note that the normalization and shape of the multijet background in the analysis control regions (i.e. 0-tag CR, Top-CR, and high- ΔR CR) are estimated in a similar manner as described below for the signal regions. However to avoid confusion, in the following discussion the term signal and control region refers only to whether the leptons in the event pass the criteria of Table 8.8.

	Isolated region (signal region)	Inverted isolation region (MJ-CR)
Electron	FCLoose $TopoE_T^{\text{cone20}} < \max(0.015 \cdot p_T, 3.5\text{GeV})$	FCLoose $TopoE_T^{\text{cone20}} > \max(0.015 \cdot p_T, 3.5\text{GeV})$
Muon	FixedCutLoose $p_T^{\text{cone20}} < 1.25\text{ GeV}$	FixedCutLoose $p_T^{\text{cone20}} > 1.25\text{ GeV}$

Table 8.8: Summary of differences between the isolated (SR) and inverted isolation regions (MJ-CR) used for the estimation of the MJ background. The isolation variables $TopoE_T^{\text{cone20}}$ and p_T^{cone20} are defined in Sections 4.2.3 and 4.3.3.

A multijet template is derived by subtracting the electroweak backgrounds MC from data in the MJ-CR. As the statistics in the 2 c -tag MJ-CR are poor, the templates derived in the 1 c -tag control region are considered for both the 1- and 2- c -tag regions. As different isolation criteria are used for electrons and muons, and also because the sources of fake electrons and fake muons are different, separate templates are used for events with an electron and a muon. Overall four different templates are used: $(e, \mu) \times (2\text{-jet}, 3\text{-jet})$.

The templates derived in the inverted isolation control region determine the shape of the multijet background. To estimate the multijet normalization, the templates are fitted to data in each of the analysis signal regions together with the electroweak backgrounds. The fit is conducted to the m_T^W distribution as this variable shows good discrimination between QCD and electroweak events. However, m_T^W only has a marginal separation ability between the different electroweak processes (namely the $t\bar{t}$ and W +jets background components). Hence, a single template is used for all the electroweak background processes. A fit is performed in every signal region of the analysis separately. For the purpose of the fit events are further split depending if the selected lepton is an electron (denoted e -mode) or a muon (denoted μ -mode). This choice is found to result in the best overall modeling of the multijet background. For the extraction of the multijet normalization, the selection cut on $m_T^W < 120\text{ GeV}$ is not applied both in the MJ-CR where the m_T^W templates are derived and in the signal region where the fit is conducted. In addition to the statistical uncertainty related to the fitted parameters, in both the e - and μ - modes, a number of sources of systematic uncertainties are considered, which can impact both the normalization and $m_{c\bar{c}}$ shape of the multijet background. They are described later in Section 11.3.6 together with the systematic uncertainties related to the modeling of the MC-based electroweak backgrounds.

Post-fit distributions of data, electroweak, and QCD backgrounds are presented in Figures 8.3 and 8.4 for the 1 c -tag and 2 c -tag signal regions. Overall, a good post-fit

agreement is seen in the plots across the signal regions for both the e - and μ -modes. In the 2 c -tag regions, the statistical errors are higher due to the smaller number of data events. From the results of the fits, normalization scale factors are extracted and assigned to the multijet template and the electroweak background in each signal region. The corresponding fractions of multijet events compared to the total background are presented in Table 8.9. The table shows that the estimated number of multijet events in the e - and μ -modes is in the order of up to 2 – 3% of the total background in each signal region. In the 2 c -tag 2-jet e -mode (Figure 8.4(a)) and the 2 c -tag 3-jet μ -mode (Figure 8.4(d)), the fit assigns a very small normalization factor to the multijet background, resulting in effectively no multijet events in these categories. In these cases, the normalization used in the final likelihood fit is taken as half of the upper error, and the uncertainty is symmetrized around this value.

Additional auxiliary plots showing the distributions of $m_{c\bar{c}}$ and m_T^W in the MJ-CR, and the resulting normalization scale factors from the m_T^W fit imposed on other variables can be found in the internal documentation of the analysis [15], together with similar fits used to estimate multijet background in the 0 c -tag and high ΔR control regions.

Signal region		MJ fraction, e -mode [%]	MJ fraction, μ -mode [%]
1-tag	2-jet	$2.52^{+1.66}_{-0.70}$	$3.26^{+2.41}_{-1.76}$
	3-jet	$2.66^{+2.33}_{-0.51}$	$1.82^{+0.77}_{-0.92}$
2-tag	2-jet	$1.08^{+1.08}_{-1.08}$	$1.83^{+4.38}_{-1.04}$
	3-jet	$3.36^{+4.03}_{-3.31}$	$1.53^{+1.53}_{-1.53}$

Table 8.9: Fraction of QCD (multijet) backgrounds in the signal region. The quoted errors are a combination of the statistical component from the m_T^W fit and the contribution from systematic sources described in Section 11.3.6.

8.5.3 | Multijet Background in the 2-lepton Channel

In the 2-lepton channel QCD multijet background is highly suppressed by requiring an event with two isolated leptons, and a di-lepton invariant mass m_{ll} close to that of a Z boson. Any residual QCD background is estimated from data via a fit to the sidebands of the same-sign m_{ll} distribution¹. A linear fit model is assumed for the multijet back-

¹The study described here was performed in the first iteration of this analysis [13]

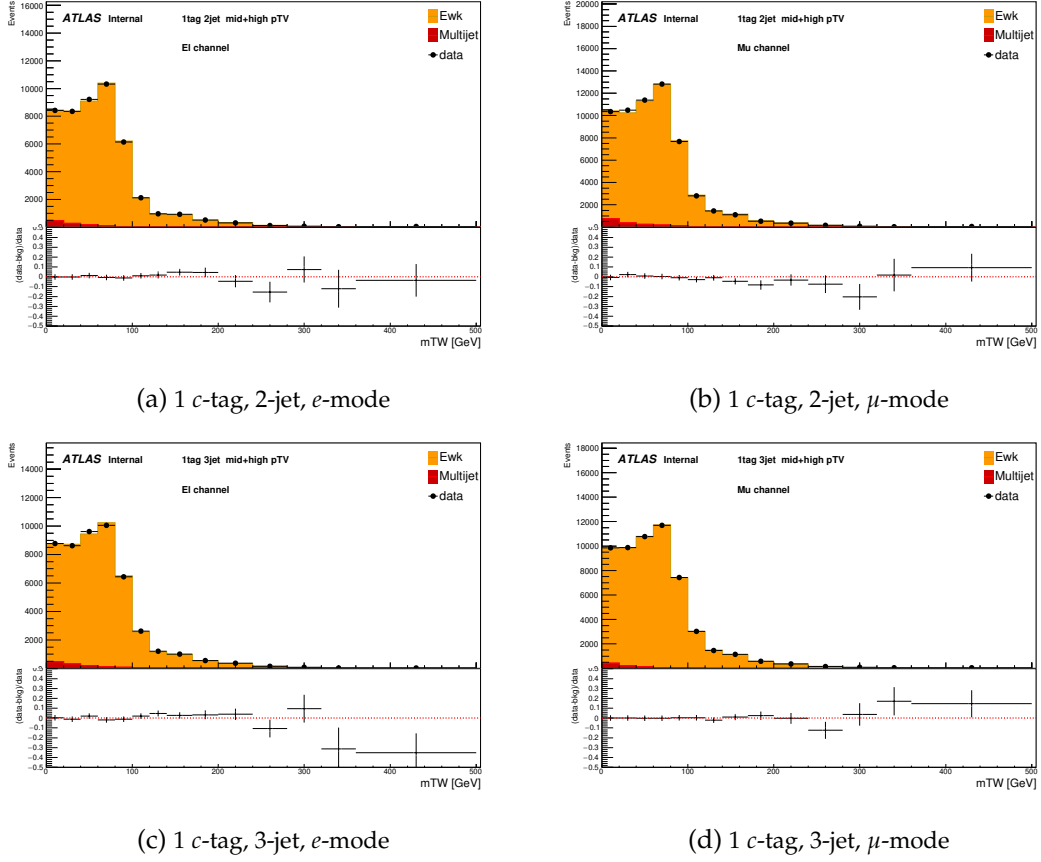


Figure 8.3: Post-fit m_T^W distribution in the 1 c -tag signal regions after applying QCD and electroweak normalization scale factors.

ground as a first-order approximation. Multijet events are not expected to produce a peak near the Z mass, hence the same-sign m_{ll} distribution is fitted with the exclusion of the region $60 \text{ GeV} < m_{ll} < 110 \text{ GeV}$. The multijet yield is determined by integrating the fitted parameters in the signal region, and subtracting the contribution from electroweak backgrounds, derived by repeating the procedure with MC simulation of the backgrounds. In all signal regions, the resulting multijet contribution amounts to less than 1%. To ensure that the resulting multijet contamination does not vary significantly with the selection of exclusion region in the fit, an additional test is performed, in which this region is altered. Eight different points are tested for the low-mass and high-mass edges of the exclusion window. The conclusion from this test is that modifying the exclusion region in the fit has no effect on the multijet contribution compared to the expected background.

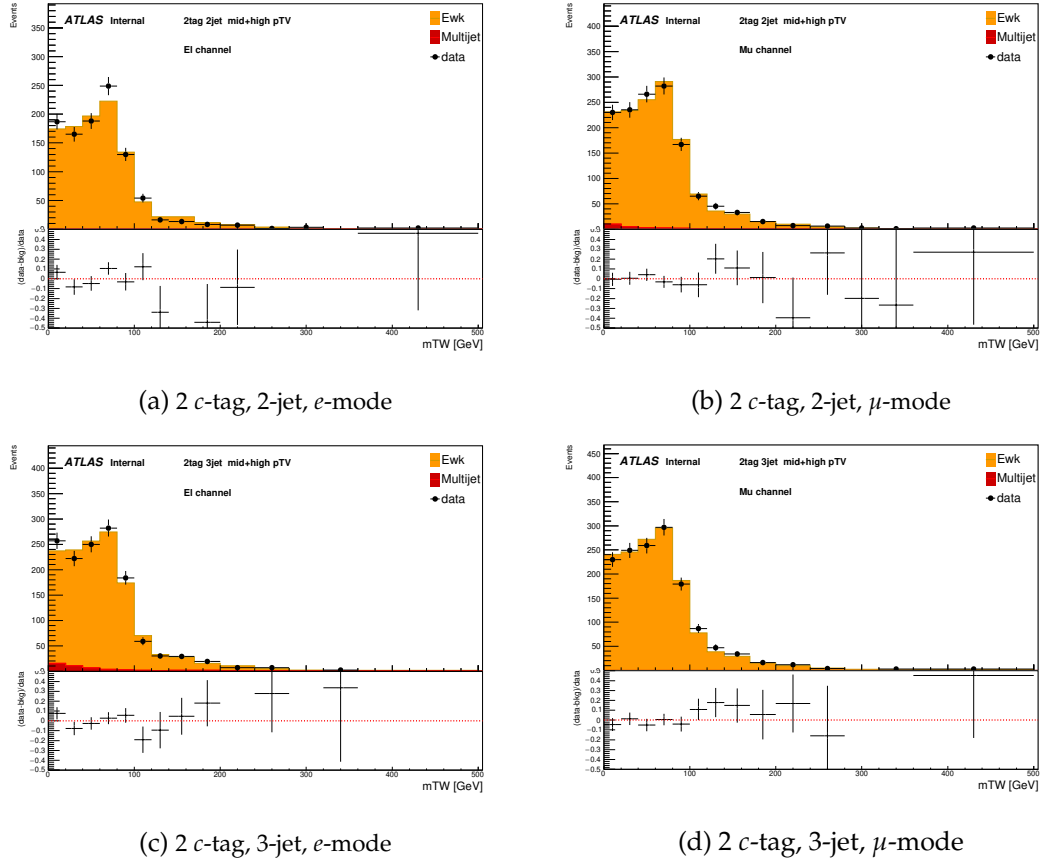


Figure 8.4: Post-fit m_T^W distribution in the 2 c -tag signal regions after applying QCD and electroweak normalization scale factors.

More information can be found in the internal documentation of the analysis first iteration [114].

Flavor Tagging in $VH(cc)$

The general concept of flavor-tagging was already introduced in Chapter 5. This chapter aims to give more details on the aspects of flavor-tagging specifically relevant to the $VH(cc)$ analysis. Section 9.1 discusses the general flavor-tagging strategy in the analysis, including the algorithms used and the way in which we achieve an orthogonal signal region with respect to the $VH(bb)$ analysis. Due to low efficiency in tagging c -jets, our c -tagging working point (WP) was specifically designed to maximize the analysis sensitivity, using the procedure described in Section 9.2. A detailed description of the c -tagging efficiency calibration of c -jets, a task that I was responsible for, is presented in Section 9.3. Finally, Section 9.4 introduces the concept of *truth-tagging*, which is a technique developed to overcome the low statistical power of our background MC samples after applying the c -tagging selection.

9.1 | Flavor Tagging Strategy in $VH(cc)$

One of the more crucial aspects of this analysis is being able to identify the c -jets arising from the decay of the Higgs boson. This is done by applying a c -tagging algorithm on the two leading p_T signal jets, used for the reconstruction of the Higgs candidate $m_{c\bar{c}}$. In the case of the $VH(cc)$ analysis, the c -tagging algorithm used is the neural network based DL1 algorithm described in Section 5.1.2, with the c -tagging discriminant of Eq (5.2).

Another challenge in this analysis is achieving orthogonality with respect to the $VH(bb)$ analysis. Here orthogonality means that no events are shared between the signal regions of the two analyses. Since the $VH(cc)$ analysis is rather similar to $VH(bb)$ in

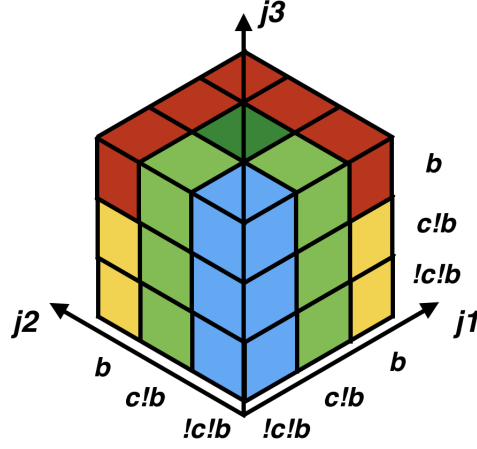


Figure 9.1: $VH(cc)$ analysis categorization in events with 3 jets, red cubes represent events with 2 b -tags defining the $VH(bb)$ signal region, dark green cubes represent events with 2 c -tags and light green cubes events with 1 c -tag both considered for the $VH(cc)$ signal region, blue cubes are events failing both b -tagging and c -tagging requirements, yellow cubes correspond to event failing the $VH(cc)$ signal region tag selection as they include a b -tagged jet among the two leading jets in the event.

terms of the object and event selection, designing a proper flavor-tagging strategy is necessary in order to remove the overlap between the analyses, achieve orthogonality, and reduce the $VH(\rightarrow b\bar{b})$ contamination in the $VH(cc)$ analysis. The $VH(bb)$ signal region contains events with exactly two b -tagged jets. Therefore, the easiest way to remove the overlap between the two analyses is to make sure that events entering the $VH(cc)$ signal region always contain less than two b -tagged jets. This is done by including a b -tagging veto in the c -tagging requirement. In other words, in order for a jet to be c -tagged in the analysis, it must both pass the DL1 c -tagging criteria and fail the b -tagging criteria of the MV2c10 algorithm at 70% efficiency working point, which is the requirement for b -tagging jets in the $VH(bb)$ analysis. Events with three or more jets require special treatment due to the possibility of finding 2 b -tags among one of the leading jets and the third jet. This effect is illustrated in Figure 9.1. The b -tagging veto is therefore applied also on all non-leading jets in the event. This extra requirement guarantees less than 2 b -tagged jets for each event in the analysis, effectively achieving orthogonality with $VH(bb)$. The effect of this extra selection cut was studied and shown to have a marginal impact in the 2-lepton channel while mostly removing $t\bar{t}$ background in the 0- and 1-lepton channels. A sketch summarizing the analysis flavor-tagging strategy is shown in Figure 9.2

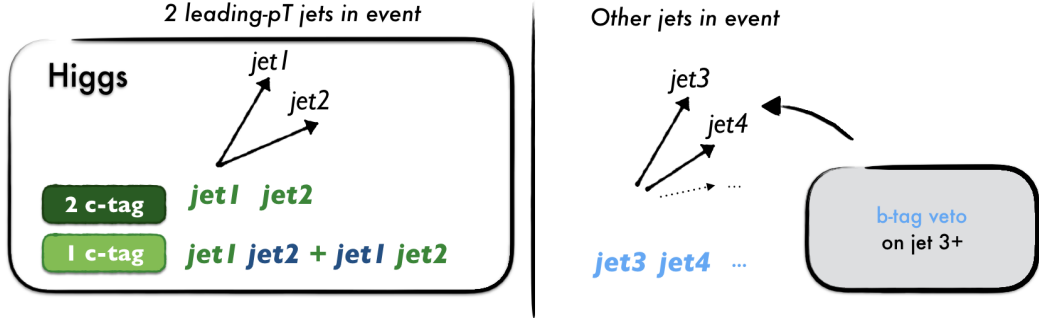


Figure 9.2: $VH(cc)$ analysis strategy: events are categorized based on the number of c -tags in the two leading p_T jets used to reconstruct the Higgs boson candidate and an additional b -tag veto is applied on all remaining jets in the event. Here green indicates c -tagged jets and light blue indicate jets which are not b -tagged.

9.2 | Working Point Optimization

As mentioned above, c -tagging in the analysis is performed using the DL1 algorithm. In this methodology, for a jet to be c -tagged and identified as c -jet it must satisfy a condition imposed by the $D_{DL1}^{c\text{-tag}}$ discriminant:

$$D_{DL1}^{c\text{-tag}} = \ln \left(\frac{p_c}{f_b \cdot p_b + (1 - f_b) \cdot p_{light}} \right) < \text{cut}, \quad (9.1)$$

where, recalling from Sec 5.1.2, p_b , p_c , and p_{light} represent the probability of the jet being a b -, c - or light-jet, f_b is the fraction of b -jets in the training sample, and cut is the threshold value of the discriminant above which jets are considered to be c -tagged. Note that a large value of f_b will increase the discrimination of c -jets and b -jets while a low value will predominantly increase the discrimination of c -jets and light-jets.

Recall that the three probabilities p_i are output of the algorithm determined on a per-jet basis, and we have virtually no control over their values. On the other side, f_b and the cut value must be determined by the user, and one must select their values to fully define the tagging criteria.

An optimization study is performed to choose the optimal values of f_b and cut maximizing the sensitivity of the $VH(cc)$ analysis. A 2-D grid is defined by varying the f_b and cut values, and for each combination the sensitivity of the signal is estimated using two different figures of merit considering only statistical uncertainties: the expected limit at 95% confidence level (CL) is used in the 0-lepton and 1-lepton channels, and the

significance (defined as $z = \frac{S}{\sqrt{S+B}}$, with S and B the number of signal and background events, respectively) in the 2-lepton channel. In the case of the 2-lepton channel the 95% CL expected limit was also tested and shown to lead to very similar results as z . The b -tagging veto on the MV2c10 70% efficiency WP is included in the c -tagging selection criteria. The optimization is first performed separately in each leptonic channel to make sure that the optimized values are compatible between the channels.

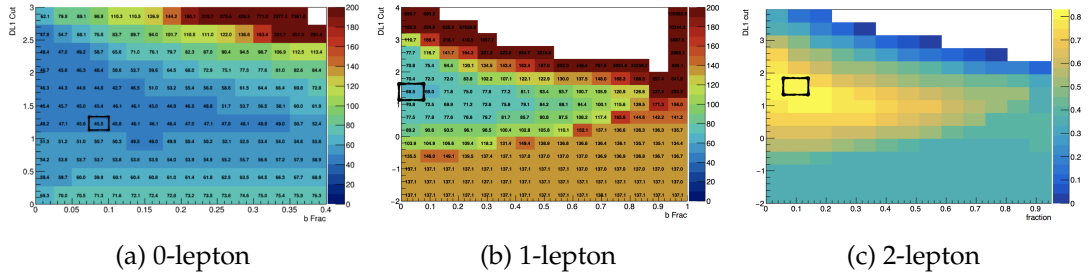


Figure 9.3: Working point optimization for the simultaneous c -tagging and b -tagging veto in the (a) 0 – lepton, (b) 1 – lepton, and (c) 2 – lepton channels. The optimization is performed with respect to the expected 95% CL limit considering statistical uncertainty only in the case of the 0-lepton and 1-lepton channel, and with respect to the signal significance in the case of the 2-lepton channel. The black rectangle corresponds to the optimal value.

Figure 9.3 shows the result of the grid optimization on the working point cut selection and the b -fraction f_b of the $D_{DL1}^{c\text{-tag}}$ definition. The three optimizations across all the channels converge towards the same working point and the optimum lies in a stable region. Therefore, a unique working point is defined for all three channels and the optimal value for the cut selection is found to be $D_{DL1}^{c\text{-tag}} > 1.3$ with a b -fraction value of $f_b = 0.08$. This low optimal value for f_b indicates that the signal is more sensitive to the rejection of light-jets than b -jets. This can be explained by the default b -tag veto requirement removing up to 70% of b -jets before the optimization. The performance of the optimal working point of the c -tagging including the b -tagging veto working point is summarized in Table 9.1. Due to the similarities between τ -jets and c -jets signature in the detector, their c -tagging performance is similar.

Requirement	Efficiency			
$D_{DL1}^{c\text{-tag}} \geq 1.3$ and $f_b=0.08$ and $MV2c10 \leq 0.83$	c -jets	b -jets	light-jets	τ -jets
	27%	8%	1.6%	25%

Table 9.1: Efficiencies of the c -tagging (including the b -tag veto) working point chosen by the optimization study.

9.3 | Calibration of the c -tagging Efficiency of c -jets

After a flavor-tagging working point has been defined it is typically characterized by its corresponding tagging efficiency. This number can be estimated from MC, where the *truth* flavor of jets is known, simply by measuring the fraction of signal jets passing the working points criteria. However, as our simulation does not perfectly describe reality, differences may arise between tagging efficiency measured in MC to that observed in data. To account for these differences, a calibration procedure is implemented to derive efficiency scale factors, which are then applied to simulated jets. Each jet flavor has its own dedicated calibration and set of scale factors. The procedures of calibrating the tagging efficiency of b - and light-jets were summarized in Section 5.3. As the c -tagging calibration of c -jets was one of my responsibilities in the analysis, it will be discussed in this section in more details.

The tagging efficiency of c -jets is measured in data using semi-leptonic $t\bar{t}$ events. That is, both top and anti-top decay to a b -jet and a W -boson, one W decays to a lepton and neutrino, and the other W decays hadronically. The physics objects in each event are thus one lepton (either e or μ), the E_T^{miss} from the neutrino, and four jets, two of which are b -jets from top decays, and the other two are either a pair of light-jets, or light+charm from the W . A kinematic likelihood fitter (KLFitter) [115] is used to determine which of the four jets in the event are from the W decay and which are the b -jets from the top quark. The KLFitter assigns the jets to the different objects and chooses the permutation (i.e which jet is assigned to which object) with the maximum likelihood compared to the known topology of the top decay in semi-leptonic channel. Only events with a likelihood score > -49 , equivalent to 68% of the $t\bar{t}$ events, are considered in the analysis. In addition, the two jets identified as b -jets from the top decays must pass the tightest b -tagging requirement of the DL1 algorithm equivalent to 60% b -tagging efficiency.

The c -tagging efficiency is determined from a χ^2 fit applied to the jets classified by the KLFitter as coming from the W decay. Considering that W -boson decays into a b -quark are suppressed by the CKM matrix, in approximately half of the events the W will decay via $W \rightarrow cs$, which implies a c -jet component corresponding to $\sim 25\%$ of the jets considered in the fit.

The jets associated with the W -boson decay are categorized into four exclusive bins, depending on whether they pass or not different flavor-tagging criteria. The four categories correspond to jets which are: b -tagged by the MV2c10 60% WP, b -tagged by the MV2c10 70% WP, c -tagged by the DL1 WP used in VH(cc), or not tagged by any of

the mentioned flavor-tagging criteria. The efficiency scale factors for the four categories are extracted simultaneously in the fit. The b -tagging working points are included in this procedure for consistency with the method used to derive b -tagging efficiency scale factors for c -jets [116]. However, due to the b -tagging veto implemented in the $VH(cc)$ c -tagging criteria, the inclusion of the b -tagging categories has a very small impact on the extracted c -tagging scale factors. This has been cross-checked by performing a similar calibration with the b -tagging categories removed. The resulting scale factors were consistent within $< 0.5\%$.

As one does not expect to find two c -jets originating from the decay of a single W boson, events where both jets are c -tagged are rejected to simplify the fit. The b - and light-jet c -tagging efficiency scale factors, obtained from the methods described in Section 5.3, are applied to the MC jets prior to performing the fit. Events are further categorized based on the p_T of the leading and subleading jets, and the fit is performed by comparing the fractions of tagged jets in data and simulation. The χ^2 function minimized in the fit is defined as:

$$\begin{aligned} \chi^2 = \sum_i \sum_j & \left[\frac{N_{\text{data}}^{60}(i, j) - p(i, j) \left(c^{60}(i) N_{\text{MC},c}^{60}(i, j) + N_{\text{MC},b/l}^{60}(i, j) \right)}{N_{\text{data}}^{60}(i, j)} \right]^2 \\ & + \left[\frac{N_{\text{data}}^{70}(i, j) - p(i, j) \left(c^{70}(i) N_{\text{MC},c}^{70}(i, j) + N_{\text{MC},b/l}^{70}(i, j) \right)}{N_{\text{data}}^{70}(i, j)} \right]^2 \\ & + \left[\frac{N_{\text{data}}^{\text{ctag}}(i, j) - p(i, j) \left(c^{\text{ctag}}(i) N_{\text{MC},c}^{\text{ctag}}(i, j) + N_{\text{MC},b/l}^{\text{ctag}}(i, j) \right)}{N_{\text{data}}^{\text{ctag}}(i, j)} \right]^2 \\ & + \left[\frac{N_{\text{data}}^{\text{untag}}(i, j) - p(i, j) \left(c^{\text{untag}}(i) N_{\text{MC},c}^{\text{untag}}(i, j) + N_{\text{MC},b/l}^{\text{untag}}(i, j) \right)}{N_{\text{data}}^{\text{untag}}(i, j)} \right]^2, \end{aligned} \quad (9.2)$$

where:

- i - is the jet p_T bin of tagged jet associated with W -decay
- j - is the jet p_T bin of untagged jet associated with W -decay
- $N_{\text{data}}^{60} / N_{\text{data}}^{70} / N_{\text{data}}^{\text{ctag}} / N_{\text{data}}^{\text{untag}}$ - Number of data events which are b -tagged by the MV2c10 60% efficiency WP / b -tagged by the MV2c10 70% efficiency WP / c -tagged by the $VH(cc)$ WP / not tagged

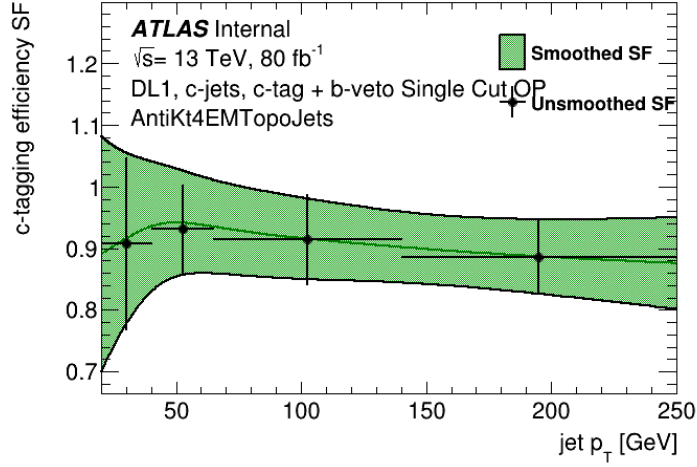


Figure 9.4: Calibration of the $VH(cc)$ c -tagging efficiency of c -jets. Raw scale factors (SF) obtained from the fit and their associated uncertainties are shown in the black crosses. The smoothed scale factors are given by the dark green curve, and their uncertainties by the light green band.

- $N_{MC,c}^{60} / N_{MC,c}^{70} / N_{MC,c}^{ctag} / N_{MC,c}^{untag}$ - Number of MC events where simulated c -jet is b -tagged by the MV2c10 60% efficiency WP / b -tagged by the MV2c10 70% efficiency WP / c -tagged by the $VH(cc)$ WP / not tagged
- $N_{MC,b/l}^{60} / N_{MC,b/l}^{70} / N_{MC,b/l}^{ctag} / N_{MC,b/l}^{untag}$ - Number of MC events where simulated b -jet or light-jet is b -tagged by the MV2c10 60% efficiency WP / b -tagged by the MV2c10 70% efficiency WP / c -tagged by the $VH(cc)$ WP / not tagged
- $p(i,j)$ - floating normalization parameters for each p_T bin.
- $c^{60}(i) / c^{70}(i) / c^{ctag}(i) / c^{untag}(i)$ - the main floating parameters for the MV2c10 60% b -tagging efficiency WP / MV2c10 70% b -tagging efficiency WP / c -tagging WP / untagged scale factors. They are interpreted as the correction to tagging efficiency $\frac{\epsilon^{data}}{\epsilon^{MC}}$, with $\epsilon^{data(MC)}$ the tagging efficiency measured on data (simulation).

The resulting simulation-to-data c -tagging efficiency scale factors for c -jets are presented in Figure 9.4. Before the scale factors are used in the analysis, they are smoothed in jet p_T in order to generate continuously varying scale factors that prevent distortions in the distributions of variables. The smoothing is done using a local polynomial kernel estimator with a smoothing parameter of 0.4.

Several sources of systematic uncertainties are considered for the c -tagging calibration of c -jets. These include uncertainties on the modeling of the $t\bar{t}$ process, uncertainty on the b - and light-jet c -tagging efficiency, and experimental uncertainties similar to those considered in the main fit of the analysis and described in Section 11.2. The dominant experimental uncertainties here are related to the reconstruction of physics objects, jet energy scale, and pileup modeling. The uncertainty from each source is evaluated by making the corresponding change to the MC and re-deriving the scale factors for each tagging interval and jet p_T bin. The total systematic uncertainty for each bin is evaluated by adding the individual contributions in quadrature.

9.4 | Truth-tagging

The high rejection for light flavor and b -jets and modest c -jet efficiency of the c -tagging algorithm leads to a large statistical uncertainty on the simulated background events. To avoid the consequences of a simulated background description with large statistical uncertainty, a parameterized efficiency approach known as *truth tagging* (TT) is used. Instead of placing a direct requirement on the c -tagging discriminant, which is referred to as *direct tagging* (DT), events are weighted based on their probability of passing the c -tagging requirement of the analysis. This method has the advantage that all events passing the analysis selection (except from flavor-tagging selection) are used.

The event weight is calculated using efficiency maps, which parameterize the c -tagging efficiency as a function of jet p_T and η . The efficiency maps are derived for each flavor separately. The c -tagging WP includes by default a b -tagging veto and therefore the efficiency maps are calculated accounting for (c -tagging + b -tagging veto). Concerning the b -tagging veto on the third (or more) jets in the event, this veto is applied using direct tagging. Therefore it is not considered in the truth tagging weight computation. The TT weights for the 1 and 2 c -tag categories are thus given by:

$$\begin{aligned} w_{2 \text{ } c\text{-tag}} &= c_1 \epsilon_1 \times c_2 \epsilon_2 \\ w_{1 \text{ } c\text{-tag}} &= c_1 \epsilon_1 \times (1 - c_2 \epsilon_2) + c_2 \epsilon_2 \times (1 - c_1 \epsilon_1), \end{aligned} \tag{9.3}$$

where ϵ_1 and ϵ_2 are the efficiencies for the first and second jets, which are extracted from the efficiency maps depending on the flavor, p_T , and η of each jet. c_1 and c_2 are correction factors which depend on the ΔR between the jets. These factors, which will be described in more details shortly, are applied on a per jet basis in V +jets samples

(i.e. for the other backgrounds $c_i = 1$). They are implemented to correct for differences between the $m_{c\bar{c}}$ distribution obtained with direct-tagging and truth-tagging.

Any event can be assigned a TT weight for either the 1 or 2 c -tag regions. To avoid introducing correlations between the 1 and 2 c -tag categories the events are split randomly based on the parity of the event number - even events are assigned to the 1 c -tag region and odd events to the 2 c -tag region. The TT weight is multiplied by 2 to correct for this split.

9.4.1 | ΔR Correction for V +jets

When comparing the $m_{c\bar{c}}$ and $\Delta R(j_1, j_2)$ distributions for direct and truth tagging, good closure is observed for $t\bar{t}$ samples, as shown in Figure 9.5(a). This is expected given the efficiency maps that are derived from $t\bar{t}$ events. However, for V +jets, a non-closure is observed, particularly at low values of $\Delta R(j_1, j_2)$, which corresponds to low values of $m_{c\bar{c}}$. An example can be seen in Figure 9.5(b). For events with a small separation between jets, there are two effects that could lead to differences between the truth tagging and direct tagging:

1. At low $\Delta R(j_1, j_2)$, fragmentation in the nearby jet can impact the c -tagged jet, which for example can mean that the c -jet is misidentified as a b -jet if a b -hadron leaked in the c -jet substructure. Generally, additional tracks or vertices can also impact the flavor-tagging efficiency. This would have a clear effect in the direct tagging, but is not accounted for in truth tagging.
2. Similarly, at low ΔR between the two jets, the clusters in the calorimeter of the jet under tagging consideration can be affected by the energy deposit of a nearby jet, which then shifts the jet axis. This would impact the performance of the `JetFitter` algorithm, which assumes that the jet axis is on the flight path of the heavy hadron. Again, this would affect the direct tagging but would not be accounted for in the truth tagging.

In order to mitigate this discrepancy, a $\Delta R(j_1, j_2)$ dependent correction factor is applied. This correction factor is applied per jet, and depending on the true flavors of the jet considered for flavor tagging and its closest jet. The correction factor is derived as a function of the minimum ΔR between the c -tagged jet and any of the other jets in the event. A sample of $Z(\rightarrow ll)$ +jets events is used to derive the corrections, with only minimum requirements on reconstructed lepton and jets, no kinematic selection applied,

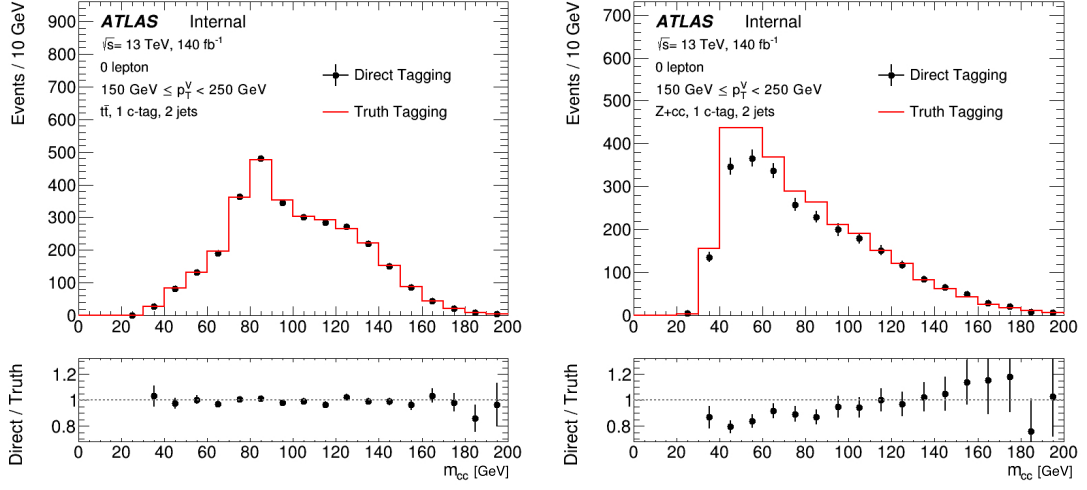
(a) $0 - \text{lepton}, t\bar{t}$ (b) $0 - \text{lepton}, Z+cc$

Figure 9.5: Truth-tagging (red line) and direct-tagging (black points) m_{cc} distributions for (a) $t\bar{t}$ and (b) $Z+cc$ events in the 0-lepton channel, 1 c -tag, 2-jet, and $150 < p_T^V < 250$ GeV region.

and inclusive in $p_T^V > 75$ GeV. For simplicity, the leading jet in the events is used to derive the correction. A closure test was performed using the sub-leading jet and the results are in agreement.

Figure 9.6 shows the ΔR distribution in the Z +jets sample used to derive the correction, for the case of the c -tagged jet being a true b -jet, and for the closest jet to be a b -, c - or light-jet respectively. There is good closure between direct and truth-tagging for all flavors of the closest jets. The same three plots are shown in Figure 9.7 for the case of the c -tagged jet being a true c -jet. Here there are clear discrepancies visible with an overall overestimation of the truth tagging by 10 – 20%, mostly observed when the closest jet is a true b -jet or c -jet. For a nearby light-jet the effect is smaller but still visible at low $\Delta R(j_1, j_2)$. For a c -tagged light-jet, a similar overall discrepancy of 10 – 20% is observed which is shown in Figure 9.8.

The ΔR correction is derived by considering the ratio of truth-tagging to direct-tagging as a function of $\Delta R(j_1, j_2)$ and fitting a fourth-order polynomial. In turn, this correction is then applied per jet when calculating the TT weight of Equation (9.3). The fitted fourth-order polynomials used to derive the corrections are shown as the solid curve on the bottom pad of Figures 9.6–9.8. The total number of DT and TT events is also specified in the plots as Yield DT and Yield TT.

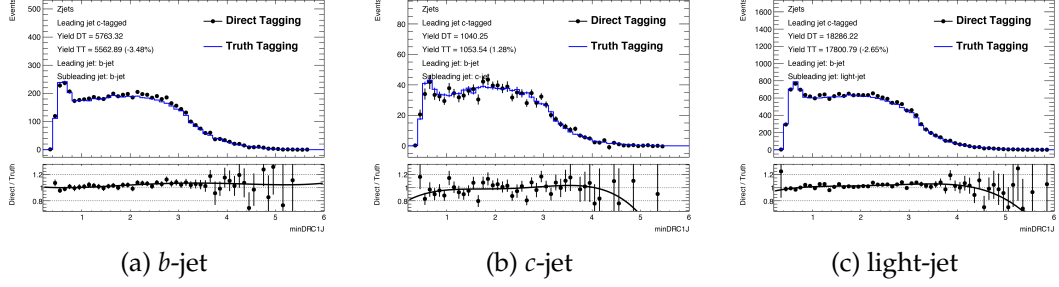


Figure 9.6: Truth- and direct-tagging comparisons as a function of the minimum ΔR between a c -tagged b -jet and its closest jet. The solid black curve on the bottom pad shows the fourth-polynomial function fitted to the ratio between TT and DT. The total number of DT (Yield DT) and TT (Yield TT) events used in each plot is also specified.

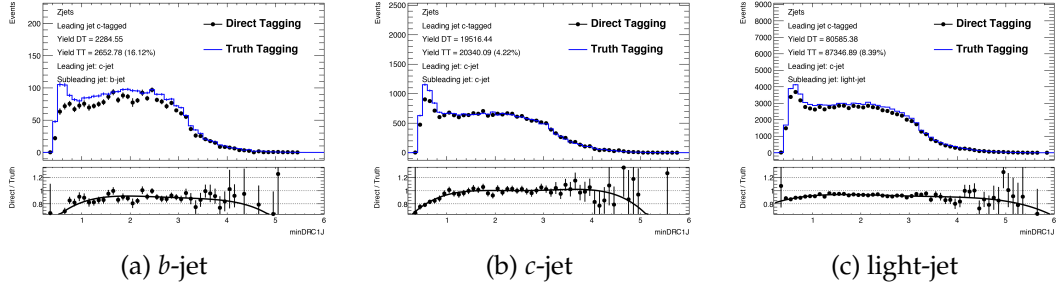


Figure 9.7: Truth- and direct-tagging comparisons as a function of the minimum ΔR between a c -tagged c -jet and its closest jet. The solid black curve on the bottom pad shows the fourth-polynomial function fitted to the ratio between TT and DT. The total number of DT (Yield DT) and TT (Yield TT) events used in each plot is also specified.

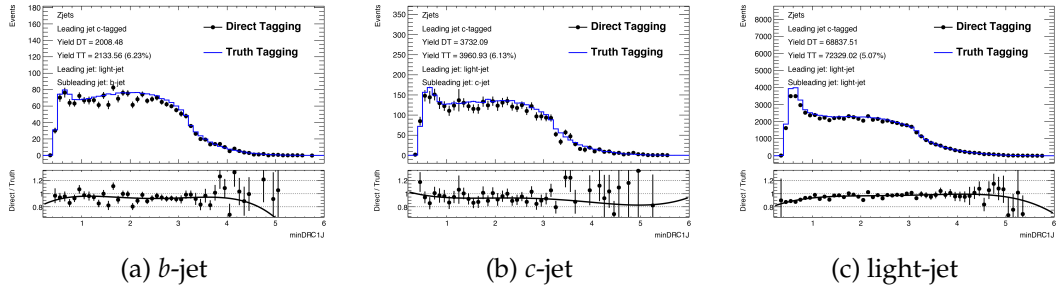


Figure 9.8: Truth- and direct-tagging comparisons as a function of the minimum ΔR between a c -tagged light-jet and its closest jet. The solid black curve on the bottom pad shows the fourth-polynomial function fitted to the ratio between TT and DT. The total number of DT (Yield DT) and TT (Yield TT) events used in each plot is also specified.

The closure between truth-tagging and direct-tagging after applying the ΔR based correction was subsequently tested for the different channels. Good closure was observed in all three channels. Comparisons before and after applying the ΔR correction are shown in Figure 9.9, for the Z+cc sample which showed a significant non-closure

before the correction.

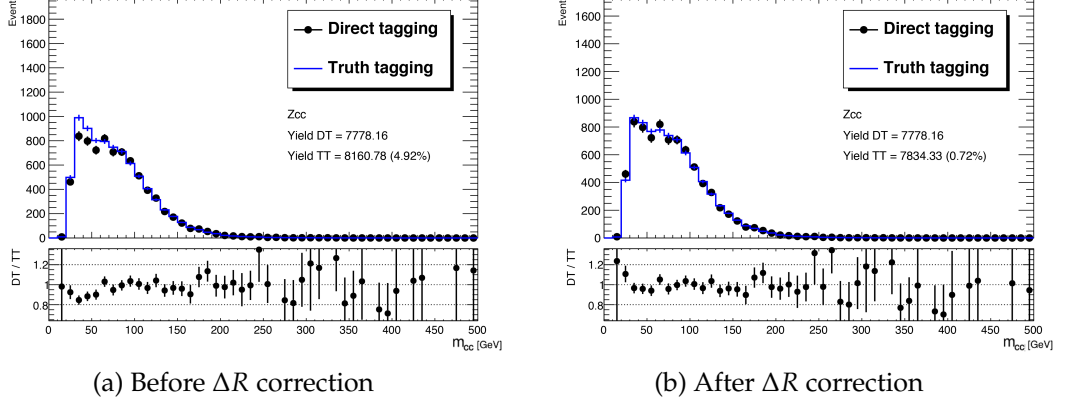


Figure 9.9: Truth- and direct-tagging comparisons of the $m_{c\bar{c}}$ distribution of the Z+cc sub-process in the 2-lepton channel (a) before and (b) after applying the minimum ΔR correction. The total number of DT (Yield DT) and TT (Yield TT) events used in each plot is also specified.

9.4.2 | Residual non-closure correction

In order to mitigate a possible systematic non-closure between truth- and direct-tagging, an additional normalization correction is applied to the TT $m_{c\bar{c}}$ distribution. The DT/TT correction is predicted with a likelihood fit, using the following parametrization:

$$\begin{aligned}
 N(0\text{-tag})_{ij}^{DT} &= [1 - C \cdot \epsilon_i^T] \cdot [1 - C \cdot \epsilon_j^T] \cdot N(\text{pre-tag})_{ij} \\
 N(1\text{-tag})_{ij}^{DT} &= (C \cdot \epsilon_i^T \cdot [1 - C \cdot \epsilon_j^T] + C \cdot \epsilon_j^T \cdot [1 - C \cdot \epsilon_i^T]) \cdot N(\text{pre-tag})_{ij} \\
 N(2\text{-tag})_{ij}^{DT} &= [C \cdot \epsilon_i^T] \cdot [C \cdot \epsilon_j^T] \cdot N(\text{pre-tag})_{ij} \\
 N(0\text{-tag})_{ij}^{TT} &= [1 - \epsilon_i^T] \cdot [1 - \epsilon_j^T] \cdot N(\text{pre-tag})_{ij} \\
 N(1\text{-tag})_{ij}^{TT} &= (\epsilon_i^T \cdot [1 - \epsilon_j^T] + \epsilon_j^T \cdot [1 - \epsilon_i^T]) \cdot N(\text{pre-tag})_{ij} \\
 N(2\text{-tag})_{ij}^{TT} &= \epsilon_i^T \cdot \epsilon_j^T \cdot N(\text{pre-tag})_{ij}.
 \end{aligned} \tag{9.4}$$

The pre-tag yield, $N(\text{pre-tag})_{ij}$, is constant and is derived from the direct-tagging distribution, with the indices ij representing the flavors of the two leading jets used to reconstruct the $m_{c\bar{c}}$ distribution. The total yields in each tag region, $N(0/1/2\text{-tag})_{ij}^{DT/TT}$, are calculated from the $m_{c\bar{c}}$ TT and DT distributions. The fit includes three free parameters: a global DT/TT efficiency correction factor C , which is independent of the jet flavor and is common to both jets, and the tagging efficiencies for the two jets comprising the Higgs candidate ϵ_i^T and ϵ_j^T , assumed to be the same for both direct- and truth-tagging.

An independent fit is performed for each sample where truth-tagging is used and in each of the leptonic channels. Due to the lower statistics in the 2 c -tag category of the DT samples, the DT 2 c -tag category is not used in the fit to extract the DT/TT corrections. Using the fitted parameters, the ratio between direct- and truth-tagging yields can be obtained from the model for each of the tag categories:

$$\begin{aligned}
 \left(\frac{DT}{TT}\right)_{ij}^{0\text{-tag}} &= \frac{[1 - C \cdot \epsilon_i^T] \cdot [1 - C \cdot \epsilon_j^T]}{[1 - \epsilon_i^T] \cdot [1 - \epsilon_j^T]} \\
 \left(\frac{DT}{TT}\right)_{ij}^{1\text{-tag}} &= \frac{(C \cdot \epsilon_i^T \cdot [1 - C \cdot \epsilon_j^T] + C \cdot \epsilon_j^T \cdot [1 - C \cdot \epsilon_i^T])}{(\epsilon_i^T \cdot [1 - \epsilon_j^T] + \epsilon_j^T \cdot [1 - \epsilon_i^T])} \\
 \left(\frac{DT}{TT}\right)_{ij}^{2\text{-tag}} &= \frac{[C \cdot \epsilon_i^T] \cdot [C \cdot \epsilon_j^T]}{\epsilon_i^T \cdot \epsilon_j^T}.
 \end{aligned} \tag{9.5}$$

The ratios given in Equation (9.5) are used as the final DT/TT correction applied to the truth-tagging template in the 0, 1, and 2 c -tag categories. The results of the likelihood fit are presented in Figure 9.10 for the 0-, and 2-lepton channels in the $Z+cc$ and the 1-lepton channel in $W+cc$. The plots show the ratio of TT/DT yields calculated directly from the MC prediction (black), and after applying the correction derived in the model (red). Note that the corrections applied to the TT templates correspond to one over the red circles. The ratio between TT/DT with and without the correction (i.e. the ratio of the red and black circles) is also shown in the plot in blue.

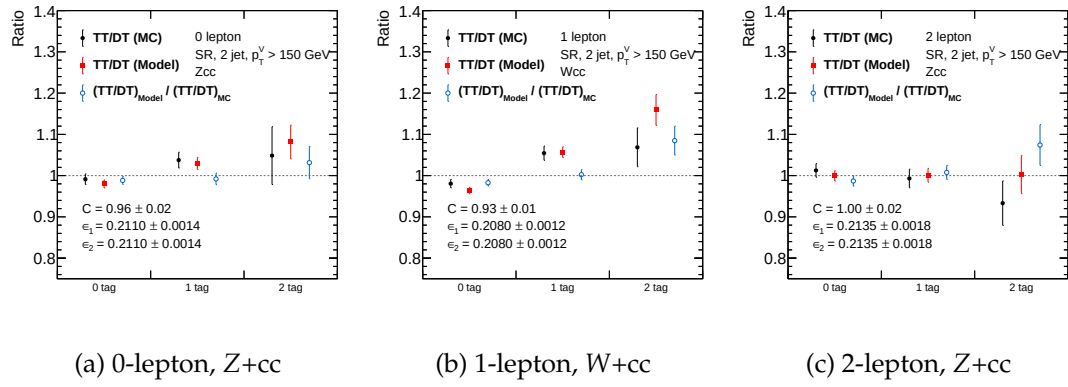


Figure 9.10: Ratio of TT and DT yields in the different tag categories for 0-lepton, 1-lepton, and 2-lepton. Black circles show the TT/DT ratio obtained directly from the MC prediction, red circles show the same ratio calculated using the model, and a comparison between the two is shown in blue. The correction for TT vs DT residual non-closure obtained from likelihood fits to the pre-tag and tag categories is equal to one over the red dot in the plot. The fitted parameters in each fit are also given in the figure

Statistical Analysis

The following chapter describes the statistical methods used in the analysis to extract the strength of the signal process from the data. In order to measure the $VH(cc)$ signal yield, a binned likelihood fit is performed to the invariant mass distribution of the Higgs candidate, $m_{c\bar{c}}$. The fit is done simultaneously to all analysis regions and all three lepton sub-channels. The technical implementation is done within the `WorkSpaceMaker` framework building upon `HistFactory` [117] together with `RooFit` [118] and `RooStats` [119]. The likelihood function is defined in Section 10.1, the test statistic used to estimate the signal strength is described 10.2, and the fit model is presented in Section 10.4.

10.1 | Definition of the Likelihood Function

The definition of the likelihood function is given in the following step-by-step. The foundation is a counting experiment in each bin using a set of Poissonian probability terms

$$\mathcal{L}(\mu) = \prod_{i \in \text{bins}} \text{Pois}(N_i | \mu s_i + b_i) = \prod_{i \in \text{bins}} \frac{(\mu s_i + b_i)^{N_i}}{N_i!} e^{-(\mu s_i + b_i)}, \quad (10.1)$$

where N_i is the number of measured data events in bin i and s_i and b_i are expected signal and background counts, estimated from MC simulation. The parameter μ is referred to as the *signal strength* and is a common scale factor to the signal yield in all bins. It can be interpreted as the ratio of the measured $\sigma \times \text{BR}$ for the signal process divided by its (SM) expectation.

$$\mu = \frac{\sigma \times \text{BR}}{\sigma_{\text{SM}} \times \text{BR}_{\text{SM}}}. \quad (10.2)$$

One of the goals of the fit is the determination of this parameter, also known as the *Parameter of Interest* (POI), and its uncertainty.

Systematic uncertainties are parameterized as a set of nuisance parameters (NPs) labeled $\vec{\theta}$. The NPs can modify the expected signal and background yields such that $\{s_i, b_i\} \rightarrow \{s_i(\vec{\theta}), b_i(\vec{\theta})\}$. Their effect is determined beforehand in auxiliary measurements as described in Chapter 11, and they are implemented in the likelihood via a set of Gaussian penalty terms, $\mathcal{L}_{\text{Aux}}$, with mean zero and a variance of unity.

$$\mathcal{L}_{\text{Aux}} = \prod_{\theta \in \vec{\theta}} \frac{1}{\sqrt{2\pi}} e^{-\theta^2/2}. \quad (10.3)$$

These terms enable to constrain the effect of each NP to the certainty with which its nominal value is known. That is, as the fit "moves" the value of θ away from the central value 0, the likelihood is "penalized" with a factor smaller than one.

NP controlling the overall normalization of the main backgrounds, referred to as *floating normalization factors*, do not correspond to a Gaussian constraint, and their value is solely determined from data in the fit.

The problem of limited simulation statistics for the background processes is addressed by introducing γ -parameters γ_i in each bin, that are allowed to modify the expected background yield as $b_i(\vec{\theta}) \rightarrow \gamma_i b_i(\vec{\theta})$. They are constrained by multiplying the likelihood with:

$$\mathcal{L}_{\text{BkgStat}}(\vec{\gamma}) = \prod_{i \in \text{bins}} \text{Gauss}(\beta_i | \gamma_i \beta_i, \sqrt{\gamma_i \beta_i}), \quad (10.4)$$

with $\beta_i = \frac{1}{\sigma_{\text{rel}}^2}$, where σ_{rel} is the relative statistical uncertainty on the expected total background yield.

Combining Equations (10.1), (10.3), and (10.4), the full likelihood can be (schematically) written as:

$$\mathcal{L}(\mu, \vec{\theta}, \vec{\gamma}) = \prod_{i \in \text{bins}} \text{Pois}(N_i | \mu s_i(\vec{\theta}) + \gamma_i b_i(\vec{\theta})) \times \prod_{\theta \in \vec{\theta}} \frac{1}{\sqrt{2\pi}} e^{-\theta^2/2} \times \prod_{i \in \text{bins}} \text{Gauss}(\beta_i | \gamma_i \beta_i, \sqrt{\gamma_i \beta_i}). \quad (10.5)$$

10.2 | Test Statistic and Measurement of the Signal Strength

A maximum likelihood procedure is used to determine the POI and its uncertainty. The used test-statistic is the log-likelihood ratio, constructed as

$$q_\mu = -2 \ln(\lambda(\mu)) = -2 \ln \left(\frac{\mathcal{L}(\mu, \hat{\hat{\theta}}, \hat{\hat{\gamma}})}{\mathcal{L}(\hat{\mu}, \hat{\hat{\theta}}, \hat{\hat{\gamma}})} \right), \quad (10.6)$$

where $\{\hat{\mu}, \hat{\theta}, \hat{\gamma}\}$ are the values that maximize the likelihood overall and $\{\hat{\hat{\theta}}, \hat{\hat{\gamma}}\}$ the ones that maximize $\mathcal{L}$ for a fixed value of μ . Maximization of the likelihood with the signal strength parameter fixed to a certain value of μ is called a *conditional fit*. An *unconditional fit* refers to a fit performed with the signal strength not fixed, but rather free to be determined by data (i.e. "floated").

The likelihood ratio $\lambda(\mu)$ is defined in the range $[0, 1]$. A value of $\lambda(\mu)$ close to 1 describes a good agreement between the data and a specific hypothesis of μ , whereas values near 0 are typical of disagreement. Therefore, as the value of q_μ gets larger the incompatibility between the data and the given hypothesis of μ increases. This (in)compatibility is often quantified using the *p-value*:

$$p_\mu = \int_{q_{\mu, obs}}^{\infty} f(q_\mu | \mu) dq_\mu, \quad (10.7)$$

with $q_{\mu, obs}$ the observed value of q_0 in data, and $f(q_\mu | \mu)$ is the probability density function (PDF) of q_μ assuming a signal strength value of μ . The p-value can be interpreted as the probability, under the hypothesis μ , of finding a test statistic as incompatible with the data as the observed q_μ . In this analysis, the asymptotic approximation of a non-central χ^2 distribution for the PDF of the test-statistic [120] is used, leading to the relation $p_\mu = \sqrt{q_\mu}$. The results are expressed in terms of the significance Z , which is related to the p-value by:

$$Z = \Phi^{-1}(1 - p_\mu), \quad (10.8)$$

with Φ^{-1} the inverse Gaussian cumulative distribution function. A special case is that of the *background-only* hypothesis which assumes $\mu = 0$. If the data is incompatible with the background-only hypothesis, it may indicate the presence of the sought signal process. A p-value of $p_0 = 2.87 \times 10^{-7}$, equivalent to a significance of $Z = 5$ (i.e. a devi-

ation of 5σ with respect to the background-only hypothesis) is the common benchmark for rejecting the background-only hypothesis and declaring an observation or discovery.

10.3 | Exclusion and CL_s limits

In case the observed signal in data is not sufficient to reject the background-only hypothesis, exclusion limits can be set on the magnitude of μ the analysis is sensitive to. A common metric used by LHC experiments to exclude a given hypothesis (i.e. a given value of μ) is known as CL_s [121] and is given by:

$$CL_s = \frac{p_\mu}{1 - p_0}. \quad (10.9)$$

The quantity p_μ (p_0) is the p-value quantifying the compatibility of the data with the signal-plus-background (background-only) hypothesis. The largest value of μ satisfying $CL_s \geq 0.05$ is considered an upper limit on a given signal hypothesis, and all models that assume a larger μ value are excluded at 95% confidence level (CL). Given our interpretation of μ as in Equation (10.2) the 95% CL upper limit corresponds to the largest possible signal cross-section time branching ratio that is still consistent with the observed data and therefore can not be excluded.

The advantage of using the CL_s metric rather than the bare p-value is that the former is more robust against statistical fluctuations in the number of events observed in data. For example, if a search has little sensitivity to a specific signal model, the expected number of signal events is minimal with respect to the expected number of background events. In this case, a sufficiently large downward fluctuation in the observed number of events may lead to exclusion of signal hypotheses that the search is not a-priori sensitive to. With the CL_s method downward fluctuations in data as those described above will lead to larger values of p_0 , thus leading to larger values of CL_s that avoid premature exclusion of a signal model.

In practice, the *observed* 95% CL upper limit is determined by generating a number of signal+background pseudo-experiments with different values of μ , and scanning for the μ that satisfies $CL_s = 0.05$. It is also useful to compute the *expected* 95% CL upper limit, which can testify to the sensitivity of the search. For this, a similar procedure is performed only the interpretation of p_μ and p_0 in Eq. (10.9) is changed to reflect the compatibility of the hypotheses with the *Asimov dataset* [121]. The Asimov dataset is

constructed from the nominal simulated dataset, with all NPs restricted to their nominal (prior) value.

10.4 | Fit Model

The fit model uses the Higgs candidate invariant mass, $m_{c\bar{c}}$, as the final discriminant separating the $VH(cc)$ signal from the background. Binned histograms of $m_{c\bar{c}}$ are produced for data and for simulated signal and background events, with the mass range $50 < m_{c\bar{c}} < 210$ GeV. The lower limit of the fit range is motivated by the larger background modeling uncertainties observed at $m_{c\bar{c}} \leq 50$ GeV. In addition, low $m_{c\bar{c}}$ also corresponds to a lower $\Delta R(j_1, j_2)$ in the case of the V +jets background, and is subject to a higher level of disagreement between direct and truth tagging. In the high ΔR CR, which is aimed to constrain the background at high mass, the $m_{c\bar{c}}$ range extends from 100 GeV to 350 GeV. The fit binning is chosen with the following considerations:

- Minimum bin width of 10 GeV, which is close to the experimental resolution of the Higgs candidate invariant mass distribution.
- Minimize the number of bins without any data and simulation events.
- Ensure that there is no bin with a MC relative statistical uncertainty above an arbitrary threshold of 20%.

This results in a choice of a bin width of 10 GeV in all analysis signal regions, with exception of 2-lepton, 2 c -tag, high- p_T^V , where due to low statistics 15 GeV are used. A bin width of 10 GeV is also used in the 0-lepton and 1-lepton Top CR, and 20 GeV bins are used in the high ΔR CR. The 0 c -tag CR and the 2-lepton $t\bar{t}$ CR are comprised of a single bin. Table 10.1 summarizes the fit range and binning adopted in each signal and control region of the analysis.

Three parameters-of-interest are considered in the fit model, and are determined simultaneously in the minimization of the likelihood. One POI describes the signal strength of $VH(cc)$, and the other two account for the signal strengths of the diboson processes where one vector decays leptonically and the other is either a Z decaying to a pair of charm quarks or W decaying to a charm quark and light-quark (denoted $VZ(cc)$ and $VW(cl)$). The similarity between $VH(cc)$ and the diboson processes is reflected in the same final-state objects and in the fact that the jets originate from a resonant decay

Region	Fit range	Bin width
Signal region	50 – 210 GeV	10 GeV (15 GeV in 2-lepton, 2 c -tag, high- p_T^V)
0 c -tagging CR		Single bin
0- & 1-lepton Top CR		10 GeV
2-lepton $t\bar{t}$ CR		Single bin
High ΔR CR	100 – 350 GeV	20 GeV

Table 10.1: Summary of the fit range and binning adopted in each of the analysis signal and control regions.

of a heavier particle. This enables to use the measurement of the diboson processes as a validation of the analysis procedure. Due to c -tagging requirement in the event selection, the di-jet flavor composition of the diboson processes is highly enriched in c -jets, allowing to extract signal strengths for the separate flavor components rather than a flavor inclusive measurement.

The various MC processes described in Section 7.2 are sub-divided in the fit into sub-processes based on the flavors of the two leading p_T jets. The motivation for this further categorization is that jets having a different flavor composition are often the result of different physics (sub) processes, and as such tend to be described by an $m_{c\bar{c}}$ spectrum with a different shape. Therefore, events originating from a similar physics process should be grouped together and be subject to the same normalization and modeling uncertainties. The final processes and sub-categories included in the fit are listed in the following. A more detailed discussion of the considerations and reasoning behind this particular choice of categorization is given in Section 11.3 in the context of the modeling uncertainties.

- Signal: $VH \rightarrow Vc\bar{c}$: with $V \rightarrow ll, l\nu, \nu\nu$
- $VH \rightarrow Vb\bar{b}$ with $V \rightarrow ll, l\nu, \nu\nu$
- Diboson:
 - $VZ \rightarrow Vc\bar{c}$ signal, with $V \rightarrow ll, l\nu, \nu\nu$
 - $VW \rightarrow Vcl$ signal, with $V \rightarrow ll, l\nu, \nu\nu$
 - VV background
- V +jets, split into flavor compositions as:

- V+heavy flavors (V+hf): bb and cc
- V+mixed flavors (V+mf): cl , bl , and bc
- V+light flavors (V+l): l
- W+ τ flavors (W+ τ): $l\tau$, $c\tau$, and $b\tau$ (considered only for W in the 0-lepton channel)
- $t\bar{t}$ and single-top Wt -channel, split into flavor compositions as:
 - Top(bq): bl , bc , bb , $b\tau$
 - Top(lq): cl , l , cc
- single-top s -, t -channel
- multijet (1-lepton channel only)

Systematic Uncertainties

This chapter describes the various sources of systematic uncertainties which affect the measurement of the $VH(cc)$ signal strength. We have already mentioned in Section 10.1 that the systematic uncertainties are manifested as nuisance parameters (NPs) in the likelihood function, and Section 11.1 gives more details regarding the implementation and interpretation of these nuisance parameters within the framework of the fit. The different sources of systematic uncertainties, which can be grouped into two main categories based on their origin, are discussed in the next two sections. Experimental systematics, described in Section 11.2, are associated with the uncertainty with which physical quantities (e.g. energies) are measured in the detector, and to the ambiguities in the reconstruction procedures of physics objects. Modeling systematics on the other hand, are related to the uncertainty in the MC predictions of the signal and background processes, and are presented in Section 11.3.

11.1 | Nuisance Parameters

Systematic uncertainties are parameterized in the fit model as nuisance parameters (NP). The prior knowledge on each NP, θ_i , is reflected by a Gaussian penalty term added to the likelihood, with the convention of mean zero and a variance of unity ($\text{Gauss}(0|\theta_i, 1)$). As the fit shifts the central value of a NP from zero, the likelihood is multiplied by a smaller factor. In this way, these terms enable to constrain the effect of each NP to the degree of certainty with which its nominal value is known.

The *pull* of a nuisance parameter is a typical metric to quantify how far it was shifted by the fit from its expected value. For a normalized nuisance parameter with expecta-

tion value $\theta_0 = 0$ and variance $\sigma_0^2 = 1$ the pull is defined as:

$$pull(\theta) = \frac{\hat{\theta} - \theta_0}{\sigma_0} = \hat{\theta}, \quad (11.1)$$

where $\hat{\theta}$ is the value of the NP determined in the likelihood fit. A pull equal to $+1(-1)$ corresponds to a NP shifted by $+\sigma(-\sigma)$. One generally expects that for a robust fit model and well-behaved minimization procedure, the pull average of all NPs is close to zero, with relatively few NP having a large pull. It is also helpful to examine the uncertainty on the NP pulls. If the post-fit uncertainty on a given NP is smaller than unity, it implies that the fit is able to extract more information on that NP than what was provided by the initial auxiliary measurement. In this case that specific NP is said to be *constrained* by the data in the fit.

Several types of nuisance parameters are considered in the analysis, which can affect either the normalization of the $m_{c\bar{c}}$ distribution, its shape, or both. **Shape uncertainty NPs** directly vary the bin content of the templates considered in the fit model. For a given template, the modified bin content is calculated from linear interpolation between the nominal distribution and an alternative template, obtained from external studies such as those described in the next sections. **Normalization uncertainty NPs** affect only the normalization of one or several components of the fit model. Two types of normalization NPs are implemented within the fit: *constrained normalization parameters* and *free normalization parameters*. Similar to the shape uncertainty NPs, the constrained normalization parameters are associated to external constraints derived from independent studies, implemented as Gaussian penalty terms as described above. Free normalization parameters (also referred to as "floating normalizations") on the other hand, are free to vary in the fit with no likelihood penalty. Such NPs are only used to model the normalization of the dominant backgrounds, where the event yield is high enough to determine the NP value with sufficient statistical accuracy. Specifically in the fit model described here, free normalizations are considered for the V +jets and Top-related ($t\bar{t}$ and single-top Wt -channel) backgrounds.

The effects of certain systematic uncertainties may be correlated between different processes or categories. A fully correlated uncertainty is described by a single NP that is pulled by the fit in the same direction and magnitude in all regions in which the uncertainty is applied. Alternatively, a systematic uncertainty may be allowed to move independently in different categories or processes in the fit. Such uncertainties are referred to as "de-correlated", and are technically implemented using multiple NPs. Ex-

perimental uncertainty sources are typically correlated, as the effects they account for are mostly common to all processes and categories. The de-correlation of uncertainties is more common in modeling systematics, as e.g. the predictions of two MC generators may be subject to differences based on the details of the specific physics process or the region of phase-space.

11.2 | Experimental Systematic Uncertainties

Experimental systematic uncertainties are related to mis-measurements effects associated with the ATLAS detector. These include the uncertainty on the energy scale, the efficiency in identification, reconstruction, and trigger selection of different physic objects, as well as the uncertainties related to flavor-tagging and the luminosity and pileup condition during data-taking. In ATLAS, the Combined-Performance (CP) working groups are responsible for assessing these uncertainties via independent auxiliary analyses. As such, with the exception of a few special cases, I and my fellow colleagues working on the $VH(cc)$ analysis were not involved in the derivation of the systematic uncertainties described in this section. Therefore these uncertainties are only described here briefly, and the reader may refer to the internal documentation of the analysis for more detail [15]. The two exceptions are the uncertainties related to c -tagging and to the jet-energy scale applied on c -jets, both of which required adaption and validation of the existing methods used in ATLAS before they could be applied in the $VH(cc)$ analysis.

11.2.1 | Luminosity and Pileup

The uncertainty on the measurement of the integrated luminosity during the data collecting period between 2015-2018 is 1.7. It is derived following the methodology described in [122] in dedicated running periods in each year.

To correct for possible differences in the distribution of pileup vertices between data and simulation, an uncertainty is applied on the re-weighting of pileup distributions. This uncertainty is derived by varying the nominal value of $\{\mu\}$ (the average number of interactions per bunch crossing) to obtain a change associated with $\pm\sigma$ [111].

11.2.2 | Leptons

The procedures used for the reconstruction of electrons and muons in the analysis are described in detail in Chapter 4. These procedures include derivation of scale factors

for the leptons reconstruction and identification efficiency, isolation, and energy calibration. Systematic uncertainties are evaluated by varying each of the corresponding scale factors.

For electrons, the systematic uncertainties considered contain both statistical and systematic errors. As such, three NP are included which account for the uncertainties in the reconstruction, identification, and isolation efficiencies. Two additional uncertainty components are included to account for the electron energy scale and resolution, and a single uncertainty component is considered for the electron trigger efficiency.

Eight independent NPs are considered for the uncertainty in muon reconstruction efficiency, isolation, and track-to-vertex association. These include two nuisance parameters for each source of uncertainty, one for the statistical and one for the systematic component. In addition, for muon reconstruction, separate uncertainties are considered for muons above and below a p_T of 15 GeV, due to different processes used for the scale factors derivation at these two p_T regions ($J/\Psi \rightarrow \mu\mu$ at low p_T and $Z \rightarrow \mu\mu$ at higher p_T). Three more NPs are considered to account for the muon momentum scale and two more for the momentum resolution, one related to the measurement in the Inner Detector and one to that of the Muon Spectrometer. Two NPs are also considered for the statistical and systematic errors on the muon trigger efficiency.

11.2.3 | Jets

Jet-related uncertainties are associated with the various steps of the jet energy scale (JES) and resolution (JER) calibration, described in Section 4.4.3. A total of 20 independent nuisance parameters are considered. Of these, 16 are derived from in-situ analyses and account for statistical, detector, modeling, η -intercalibration non-closure, and pileup corrections. An additional uncertainty component is dedicated to high- p_T jets, as the standard JES calibration is derived with jets up to a p_T of 2 TeV. In addition, because the in-situ uncertainties rely on analyses performed in data samples with a specific jet flavor composition, three additional uncertainties are associated with jet flavor, known as flavor composition, flavor response, and bJES uncertainties. The first two are related to the fraction and response of gluon-initiated jets compared to light-jets. They are derived with the exclusion of heavy flavored jets, and are meant to be applied solely to light-jets. Instead, the bJES uncertainty is dedicated to account for the energy response of heavy-flavored jets. However, prior to the first iteration of the $VH(cc)$ analysis it was not clear whether c -jets should be assigned the bJES uncertainty (which is derived using only b -jets), or be treated as light-jets and assigned the flavor response and flavor com-

position uncertainties. Therefore, a study was performed (for which I was responsible) with the goal of answering this very question.

11.2.3.1 | p_T response of c -jets

To better understand the response of c -quarks in the detector, a MC based study of the c -jets p_T response is performed. The p_T response, defined as $\frac{p_T^{\text{reco}}}{p_T^{\text{truth}}}$, is evaluated using samples of di-jets events from two MC generators:

- PYTHIA 8 [123] using the A14 tune [124] and the NNPDF23LO [125] parton density function set (PDF). This is the nominal MC generator used to derive the JES calibration itself.
- HERWIG++ [126] with the AUET2 [127] tune and CT10 PDF [128].

For each MC generator, the response is evaluated for c -, b -, and light-jets separately. The response for each flavor is then compared between the two simulations, as differences in the simulation configuration can result in a variation in the response distribution.

The comparison is done with respect to mean values of the p_T response distribution, provided from a fit to the response distribution and analysis of the fit results. The p_T response is modeled using a Gaussian function, and the fit range is taken as $\pm 5 \times \sigma$. 16 p_T bins in the range of 45 – 2500 GeV are used, with an independent fit performed in each bin.

Prior to the fit, the two distributions are re-weighted in each p_T slice, such that they have similar normalization equal to unity. The mean extracted from the result of the Gaussian fit is compared between PYTHIA (μ_P) and HERWIG (μ_H). The variation in the average response is estimated for each flavor separately and is taken as:

$$\delta = \frac{|\mu_P - \mu_H|}{(\mu_P + \mu_H)/2}. \quad (11.2)$$

As an example, Figure 11.1 shows the response distributions of PYTHIA (blue) and HERWIG (red) in the individual bin $p_T \in [160, 210]$ GeV. The mean, extracted from the fit, and the RMS (given as the error) are shown in the plot.

The response variations δ , defined in Eq. (11.2) of the different flavors are plotted together in Figure 11.2.

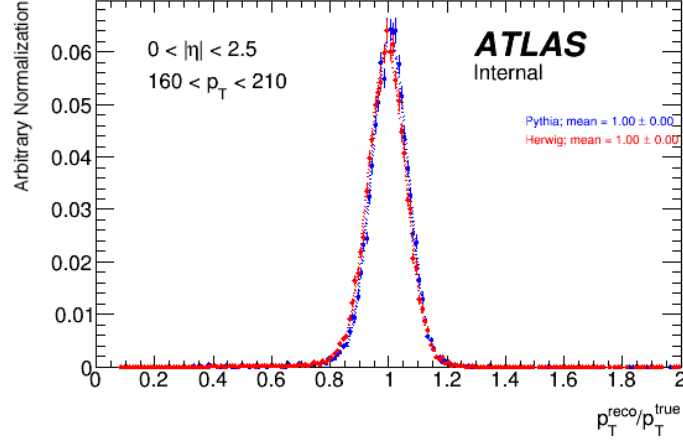


Figure 11.1: p_T response distribution for PYTHIA (blue) and HERWIG (Red) for jets with $p_T \in [160, 210]$ GeV. The mean of each distribution (extracted from the fit) and the RMS (given as the error) are shown in the plot.

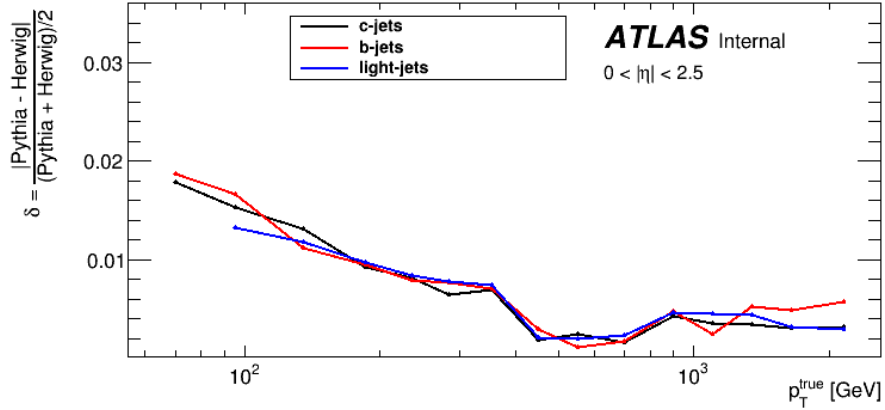


Figure 11.2: The p_T response variation (defined in Eq. (11.2)) for b - (red), c - (black), and light- (blue) jets. Due to a technical problem in the HERWIG sample the response variation of light-jets in the first p_T bin is not included in the plot.

The conclusions of this study suggest that a similar treatment may be assigned to c - and b -jets in terms of the flavor-related JES uncertainties. Hence c -jets are applied the b JES uncertainty component.

11.2.4 | Missing Transverse Momentum

The calculation of E_T^{miss} , presented in Eq. (4.11), takes into account a "hard term" representing the sum of jets and other particles in the event, and a "soft term" which accounts for additional tracks associated with the PV. As such, the uncertainty related to the hard-

term is propagated to the E_T^{miss} calculation through the uncertainties of the other physics object mentioned in this chapter. Three uncertainty components are considered which are related to the scale, resolution, and efficiency of the E_T^{miss} soft term. They are derived based on a comparison of MC simulation to data.

Systematic uncertainties related to the E_T^{miss} trigger efficiency are also taken into account, considering three effects. The first is the statistical error on the dataset used to derive the efficiency scale factors. As the E_T^{miss} trigger SF are derived using W +jets events, the second uncertainty is related to the differences arising from the use of different physics processes, namely Z +jets and $t\bar{t}$. The third systematic source accounts for the kinematic dependence E_T^{miss} trigger efficiency on the scalar sum of all final state jets.

11.2.5 | Flavor Tagging

Flavor-tagging uncertainties are considered in the analysis for effects related to both direct-tagging and truth-tagging. They are discussed in the following

11.2.5.1 | Direct-Tagging

Direct-tagging systematic uncertainties are related to the efficiency scale factors derived for each flavor of jets and are described in Section 9.3 (for c -jets) and Section 5.3 (for b - and light-jets). The individual c -tagging systematics contain up to 40 elements for each jet flavor, which include statistical, experimental, and modeling uncertainties. In an effort to alleviate the burden of accounting for all the variations, a principal component analysis is performed to reduce the number of uncertainty components from ~ 160 to a total of ~ 30 nuisance parameters. The reduced set of NP is a combination of the various components, and thus the information of the origin of the systematic uncertainty is lost.

11.2.5.2 | Truth-Tagging

The systematic uncertainties related to truth-tagging efficiencies are expected to be covered by the systematics of the direct-tagging calibration scale factors. In addition, two systematic uncertainties are added to account for the residual non-closure between truth and direct-tagging.

The first systematic uncertainty accounts for the ΔR correction applied to the truth-tagging V +jets processes. A total of 14 NPs are used, with two NPs (one for each jet-multiplicity category) assigned for each of the seven V +jets sub-processes included in the fit. The $\pm\sigma$ uncertainty is taken as the full size of the correction.

The second systematic uncertainty is related to the residual DT/TT normalization correction. The $\pm\sigma$ uncertainty is taken as the full size of the correction. To allow maximal granularity in the fit to undo the correction, independent uncertainty components are assigned in the 0, 1, and 2 c -tag categories of each sub-process.

11.3 | Modeling Systematic Uncertainties

MC simulation is used to estimate the nominal shape and normalization of the $m_{c\bar{c}}$ template of the signal and various background processes. Modeling uncertainties account for possible differences that may arise from the use of alternative MC generators, due to e.g. a different order of perturbation theory corrections or a different hadronization model. Moreover, variations of the factorization (μ_F) and renormalization scale (μ_R) in the nominal generators are also taken into account.

The modeling systematic uncertainties can have an effect either on the overall normalization of processes, the shape of the $m_{c\bar{c}}$ distribution, or the *acceptance*, defined as the fraction of events passing the full event selection in a given analysis category.

The normalization of the dominant V +jets and ($t\bar{t}$ + single-top Wt -channel) backgrounds are left unconstrained and are directly extracted from data in the fit. For the signal and the remaining backgrounds processes, the overall normalization and associated uncertainties are obtained from calculations of the cross-sections and branching ratios, and from the ratio between the total number of events (i.e the sum in all analysis regions) in the nominal and alternative MC samples.

The $m_{c\bar{c}}$ shape uncertainty is derived from the ratio between normalized distributions of the nominal and alternative generators. The shape variation is evaluated by fitting the ratio to a functional form, usually taken as a 3rd order polynomial which was found to successfully describe the ratio for the majority of the processes considered. The fit is initially performed in each analysis region separately. However, if for a given process the variations derived in two (or more) analysis regions are found to be similar, a common uncertainty is derived by re-doing the fit using a combination of the regions.

Uncertainties on the $m_{c\bar{c}}$ shape resulting from variation of the renormalization (μ_R) and factorization (μ_F) scale are also considered. These uncertainties are implemented by re-running the nominal MC simulation with these parameters modified by a factor of 2 (up-variation) and $\frac{1}{2}$ (down-variation). As a result, an independent shape variation is assigned for each process and in each c -tag/ p_T^V -jet-multiplicity region.

Acceptance uncertainty accounts for variations in the number of events falling in a certain analysis category and the possible migration from a category A to a category B . Such variations may arise from differences between simulation samples in the distributions of kinematic variables involved in the event selection and categorization. These uncertainties give an additional degree of freedom for the normalization of one category to change with respect to the other, therefore, they are only considered between categories that are grouped together and assigned a common normalization. For example, in most cases the same normalization factor is applied to a process in the 2-jet and in the 3-jet category, and as such the acceptance uncertainty between these two categories can mitigate modeling differences in ISR/FSR. Acceptance uncertainties are not implemented however between the 1 and 2 c -tag categories, as differences between the tagging performances in distinct MC simulations are expected to be covered by the flavor-tagging systematics. The acceptance uncertainties, also known as *acceptance ratios* or *extrapolation uncertainties*, are calculated as *double-ratios*:

$$\frac{\left(\frac{n_{\text{ctg. } A}}{n_{\text{ctg. } B}} \right)_{\text{alternative}}}{\left(\frac{n_{\text{ctg. } A}}{n_{\text{ctg. } B}} \right)_{\text{nominal}}}, \quad (11.3)$$

where $n_{\text{ctg. } A}$ and $n_{\text{ctg. } B}$ represents the number of selected events in categories A and B . The double-ratios of Eq. (11.3) are calculated both for nominal vs alternative MC generators and for the nominal MC with vs without scale variations of (μ_F) and (μ_R) . The final acceptance uncertainty (i.e. the prior of the associated nuisance parameter) is taken as the sum in quadrature of all contributions. It is initially calculated separately in each leptonic channel and in some cases separately in the signal and control regions. If the values agree within the statistical uncertainty then the prior extracted in the channel/region with the highest purity is used. In case of discrepancies, different priors may be assigned in different channels/regions.

The following sub-section details the modeling studies and resulting uncertainties for the different processes. The nominal and alternative MC simulation samples used to derive the uncertainties are listed in Table 7.1.

11.3.1 | $VH(cc)$ signal

The signal process modeling uncertainties are derived split by the production process, i.e. separately for the $qq \rightarrow ZH$, $qq \rightarrow WH$, and $gg \rightarrow ZH$.

Two components to the $VH(cc)$ signal normalization uncertainty are implemented. The first, are theory uncertainties on the $H \rightarrow c\bar{c}$ branching ratio and the WH and ZH production cross-sections. Secondly, an uncertainty on the total number of events accepted in the analysis phase-space is derived from a comparison of the nominal generator to the alternative generator and the scale variations.

In addition, acceptance ratios are introduced between the different analysis categories, namely between the high- and low- p_T^V regions (in the 2-lepton channel) and between the 3(+) jet and 2-jet regions. The values used as priors for the acceptance ratio nuisance parameters are presented in the tables of App. A for the signal process as well as for the background processes described in the following sub-sections.

The uncertainty on the $m_{c\bar{c}}$ shape, even though derived separately for the three production processes, is correlated to a single NP controlling the $VH(cc)$ shape in the fit. The shape variations are derived inclusively in the number of jets and c -tags, and separately for the low- and high- p_T^V regions.

11.3.2 | $VH(bb)$

For the $VH, H \rightarrow b\bar{b}$ background, a normalization uncertainty is applied based on the uncertainty of the signal strength observed in the measurement of this process by ATLAS [9]. A separate uncertainty is applied to the $ZH(bb)$ and $WH(bb)$ components. In addition, acceptance ratios between categories and $m_{c\bar{c}}$ shape variations are considered following the same methodology described above for the signal process.

11.3.3 | V +jets

Production of a vector boson in association with jets is the main background process in all three channels of the analysis. The treatment of this background in the fit and the design of a robust modeling approach can have a considerable impact on the final outcome of the search described in this thesis. Therefore, an extensive amount of effort was put into the V +jets modeling strategy, in which I had a substantial part.

The normalization and modeling systematics of the V +jets background are deter-

mined separately for the case the vector boson is a Z or a W . For simplicity, and given the similar considerations in the modeling of both backgrounds, they are described together in the discussion below and are jointly referred to as V +jets.

11.3.3.1 | Flavor split

The V +jets backgrounds are sub-divided into different components based on the flavor of the two leading jets. Six sub-process are considered for both W and Z production, namely $V+bb$, $V+bc$, $V+bl$, $V+cc$, $V+cl$, and $V+l$ (in which both jets are light-flavored). In the case the vector boson is a W , there is a non-negligible contribution in the 0-lepton channel from events where one of the Higgs-candidate jets originates from a hadronically decaying τ lepton. Hence three additional categories, $W(\tau_h\nu) + b$, $W(\tau_h\nu) + c$ and $W(\tau_h\nu) + l$ are considered in the 0-lepton channel.

The relative fraction of the sub-processes described above with respect to the total W/Z +jets background in the signal region is listed in Table 11.1. One can see from the table that in the 1 c -tag region the leading processes are $V+cl$ and $V+l$, in roughly equal proportion, with secondary contribution from $V+cc$. The 2 c -tag region on the other hand, is mainly comprised of $V+cc$ events, followed by $V+cl$, and then $V+l$. The components involving a b -jet, as well as events with a hadronic τ -jet in the 0-lepton channel are sub-dominant.

The picture reflected from Table 11.1 is generally similar in the high ΔR CR, with a larger portion of $W(\tau_h\nu)$ events compared to the SR. As one can expect, in the 0 c -tag CR the vast majority of events are of $V+l$.

The strategy for the final fit and the derivation of the V +jets background modeling is to group similar physics processes together and assign them with a common normalization. The acceptance ratios and shape uncertainties for each group are then described using the most dominant background in that group. The main assumption is that the shape of the $m_{c\bar{c}}$ distribution will tend to be similar for flavor compositions originating from the same physics, thus grouping them together can help the fit determine their common normalization. In addition, similar processes may be subject to similar modeling issues and should therefore be accounted together.

The Feynman diagrams contributing to the V +jets backgrounds are illustrated in Figure 11.3. The flavor compositions are split into three sub-process categories: di-jet pairs composed of heavy, same-flavor jets, are mostly generated through gluon split-

SR, $p_T^V > 150$ GeV		1 c-tag		2 c-tag	
Channel	process	2-jet	3(+)-jet	2-jet	3(+)-jet
0-lepton	$Z + cc$	$(8.1 \pm 0.1)\%$	$(8.1 \pm 0.1)\%$	$(48.1 \pm 0.8)\%$	$(49.4 \pm 0.9)\%$
	$Z + cl$	$(38.2 \pm 0.3)\%$	$(39.9 \pm 0.3)\%$	$(24.0 \pm 0.2)\%$	$(24.5 \pm 0.2)\%$
	$Z + l$	$(41.0 \pm 0.1)\%$	$(39.3 \pm 0.2)\%$	$(13.5 \pm 0.1)\%$	$(12.6 \pm 0.1)\%$
	$Z + bb$	$(2.6 \pm 0.1)\%$	$(2.6 \pm 0.1)\%$	$(5.2 \pm 0.1)\%$	$(4.9 \pm 0.1)\%$
	$Z + bc$	$(1.4 \pm 0.1)\%$	$(1.5 \pm 0.1)\%$	$(4.1 \pm 0.1)\%$	$(4.0 \pm 0.1)\%$
	$Z + bl$	$(8.8 \pm 0.1)\%$	$(8.6 \pm 0.1)\%$	$(5.1 \pm 0.1)\%$	$(4.7 \pm 0.1)\%$
2-lepton	$Z + cc$	$(8.6 \pm 0.2)\%$	$(8.6 \pm 0.1)\%$	$(49.7 \pm 1.0)\%$	$(49.0 \pm 0.7)\%$
	$Z + cl$	$(39.5 \pm 0.3)\%$	$(41.1 \pm 0.2)\%$	$(23.2 \pm 0.3)\%$	$(24.5 \pm 0.2)\%$
	$Z + l$	$(37.6 \pm 0.2)\%$	$(36.7 \pm 0.1)\%$	$(11.5 \pm 0.1)\%$	$(11.5 \pm 0.1)\%$
	$Z + bb$	$(3.1 \pm 0.1)\%$	$(2.9 \pm 0.1)\%$	$(6.1 \pm 0.1)\%$	$(5.5 \pm 0.1)\%$
	$Z + bc$	$(1.6 \pm 0.1)\%$	$(1.7 \pm 0.1)\%$	$(4.4 \pm 0.1)\%$	$(4.6 \pm 0.1)\%$
	$Z + bl$	$(9.5 \pm 0.1)\%$	$(9.0 \pm 0.1)\%$	$(5.1 \pm 0.1)\%$	$(4.9 \pm 0.1)\%$
0-lepton	$W + cc$	$(5.4 \pm 0.1)\%$	$(5.7 \pm 0.1)\%$	$(37.3 \pm 1.0)\%$	$(39.5 \pm 1.1)\%$
	$W + cl$	$(39.6 \pm 0.4)\%$	$(41.3 \pm 0.4)\%$	$(28.9 \pm 0.4)\%$	$(29.9 \pm 0.4)\%$
	$W + l$	$(41.3 \pm 0.3)\%$	$(41.6 \pm 0.3)\%$	$(15.5 \pm 0.2)\%$	$(15.7 \pm 0.2)\%$
	$W + bb$	$(1.3 \pm 0.1)\%$	$(1.3 \pm 0.1)\%$	$(3.0 \pm 0.1)\%$	$(3.0 \pm 0.1)\%$
	$W + bc$	$(0.6 \pm 0.1)\%$	$(0.7 \pm 0.1)\%$	$(2.1 \pm 0.1)\%$	$(2.3 \pm 0.1)\%$
	$W + bl$	$(4.3 \pm 0.1)\%$	$(4.4 \pm 0.1)\%$	$(3.0 \pm 0.1)\%$	$(3.0 \pm 0.1)\%$
	$W(\tau_h \nu) + l$	$(6.8 \pm 0.4)\%$	$(4.6 \pm 0.2)\%$	$(5.5 \pm 0.4)\%$	$(3.0 \pm 0.2)\%$
	$W(\tau_h \nu) + c$	$(0.6 \pm 0.2)\%$	$(0.3 \pm 0.1)\%$	$(4.6 \pm 0.5)\%$	$(3.3 \pm 0.5)\%$
	$W(\tau_h \nu) + b$	$(0.1 \pm 0.1)\%$	$(0.1 \pm 0.1)\%$	$(0.2 \pm 0.1)\%$	$(0.2 \pm 0.1)\%$
1-lepton	$W + cc$	$(6.4 \pm 0.1)\%$	$(6.6 \pm 0.1)\%$	$(42.4 \pm 0.6)\%$	$(43.1 \pm 0.6)\%$
	$W + cl$	$(52.1 \pm 0.3)\%$	$(52.4 \pm 0.3)\%$	$(35.9 \pm 0.3)\%$	$(36.0 \pm 0.3)\%$
	$W + l$	$(35.3 \pm 0.1)\%$	$(34.9 \pm 0.1)\%$	$(12.8 \pm 0.1)\%$	$(12.5 \pm 0.1)\%$
	$W + bb$	$(1.3 \pm 0.1)\%$	$(1.2 \pm 0.1)\%$	$(2.9 \pm 0.1)\%$	$(2.7 \pm 0.1)\%$
	$W + bc$	$(1.0 \pm 0.1)\%$	$(1.0 \pm 0.1)\%$	$(3.3 \pm 0.1)\%$	$(3.2 \pm 0.1)\%$
	$W + bl$	$(3.9 \pm 0.1)\%$	$(3.8 \pm 0.1)\%$	$(2.6 \pm 0.1)\%$	$(2.5 \pm 0.1)\%$

Table 11.1: Fractions of the individual Z+jets and W+jets components in the signal regions (SR) in the $p_T^V > 150$ GeV category and their statistical uncertainty. For the 0-lepton and 1-lepton channels the 3 jets category is used, while for 2-lepton the 3+ jets category is used.

ting in case the vector boson is a W , or a combination of gluon splitting and fusion of b/c quarks for $V=Z$, as seen in the top row of Figure 11.3. Therefore the first sub-process category is referred to as $V+\mathbf{hf}$ ("heavy flavor") and contains the $V+bb$ and $V+cc$ backgrounds. The normalization of backgrounds with a combination of light and heavy flavor jets, namely $V+bc$, $V+bl$, and $V+cl$, is determined jointly as $V+\mathbf{mf}$ ("mixed flavor"), with the contributing Feynman diagrams shown in the bottom row of Figure 11.3. The third category consists of the $V+l$ process, which is generated through a combination of each of the diagrams of Figure 11.3, replacing the indicated b/c quark flavors with u, d, s . In the 0-lepton channel a special treatment is required for events where the W boson decays into hadronic τ and a neutrino. If the (leading) additional jet in the event is a b - or a c -jet, the event is grouped in the mixed flavor category ($W+\mathbf{mf}$), and if it is a light-jet the event is floated together with $W+l$. The motivation behind this approach

is that had the hadronic τ -jet been correctly identified and removed during the overlap removal procedure, these are the categories the events would end up in.

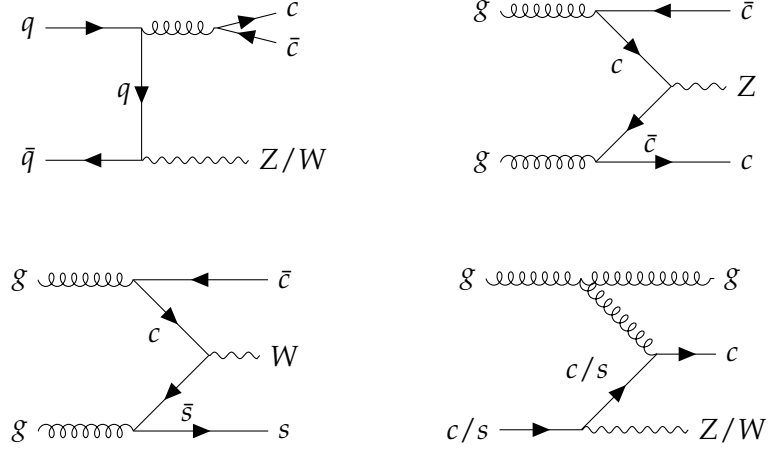


Figure 11.3: Feynman diagrams for the production of a vector boson in association with jets. The processes in the top row contribute to the $V+hf$ categories, while the bottom two diagrams give rise to the $V+mf$ categories. The $V+l$ category is generated from a combination of the four diagrams shown, with proper modification of the indicated flavors

The normalized $m_{c\bar{c}}$ shapes of the individual components in the Z +jets and W +jets background are presented in Figure 11.4 for the combined (1+2) c -tag, (2+3) jet, high- p_T^V region. For the W +jets components, one can see in the plot a more pronounced distinction between the processes in the heavy flavor category ($W+cc$ and $W+bb$) and those in the mixed flavor category ($W+cl$, $W+bc$, $W+bl$). This can be explained by the fact that the $W+hf$ processes are generated almost exclusively by the gluon splitting diagram, whereas the additional Feynman diagram contributing to the $Z+hf$ production can result in similar kinematic properties to the processes in the $Z+mf$ category.

Figure 11.5 shows the normalized $m_{c\bar{c}}$ shapes after grouping the flavor components in the three V +jets sub-process categories. As one can expect from the relative fractions in Table 11.1, the shape of each category is driven by the most dominant background, namely $V+cc$ in $V+hf$, $V+cl$ in $V+mf$, and $V+light$ jets in the $V+l$ category.

11.3.3.2 | Normalization uncertainties

In the fit, all Z +jets and W +jets background normalizations are left free to float. For the $V+hf$ ($V+mf$) category, the constraining power of the background mainly comes from the 2 c -tag (1 c -tag) SR and high ΔR CR. The $Z+hf$ and $Z+mf$ floating normalizations are de-correlated between the two p_T^V categories, in particular to allow additional freedom

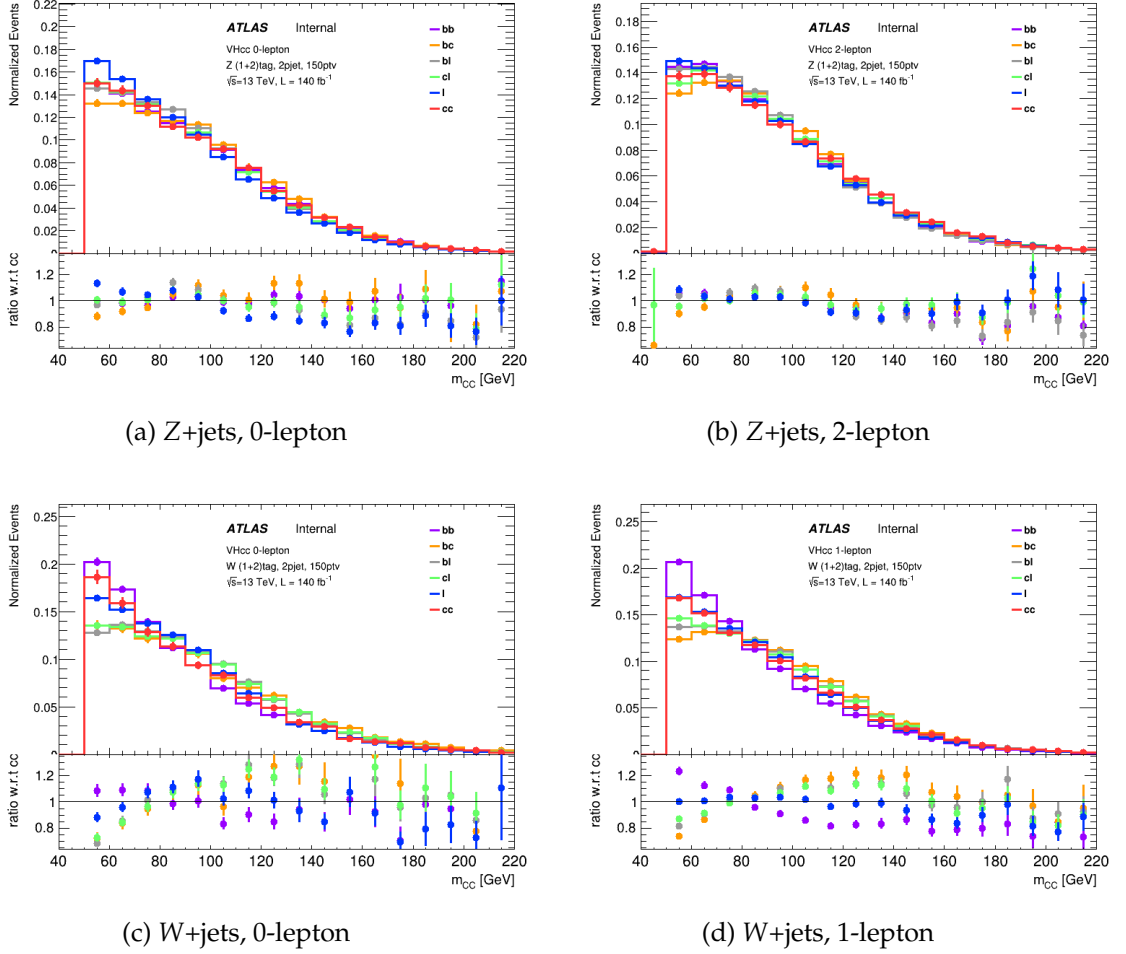


Figure 11.4: Normalized $m_{c\bar{c}}$ distributions of the individual V +jets flavor components for (a) a Z boson in the 0-lepton channel, (b) a Z boson in the 2-lepton channel, (c) a W boson in the 0-lepton channel, and (d) a W boson in the 1-lepton channel. All plots correspond to the combined (1+2) c -tag, (2+3) jet, high- p_T^V region.

for the low- p_T^V category in the 2-lepton channel to change with respect to the high- p_T^V categories in the 0-lepton and 2-lepton channels.

The $V+l$ backgrounds are constrained using dedicated 0 c -tag control regions, comprised of a single-bin. There are four such CRs in the 2-lepton channel to constrain $Z+l$ (one in each p_T^V and jet multiplicity region) and two in the 1-lepton channel (one in each jet multiplicity region) to constrain $W+l$. The $Z/W+l$ normalizations in the 0-lepton channel are determined from the factors obtained using the other two channels. Due to the large available statistics in the 0 c -tag region, each of the CRs is associated with its own floating normalization, i.e. de-correlated for each p_T^V and jet multiplicity bin.

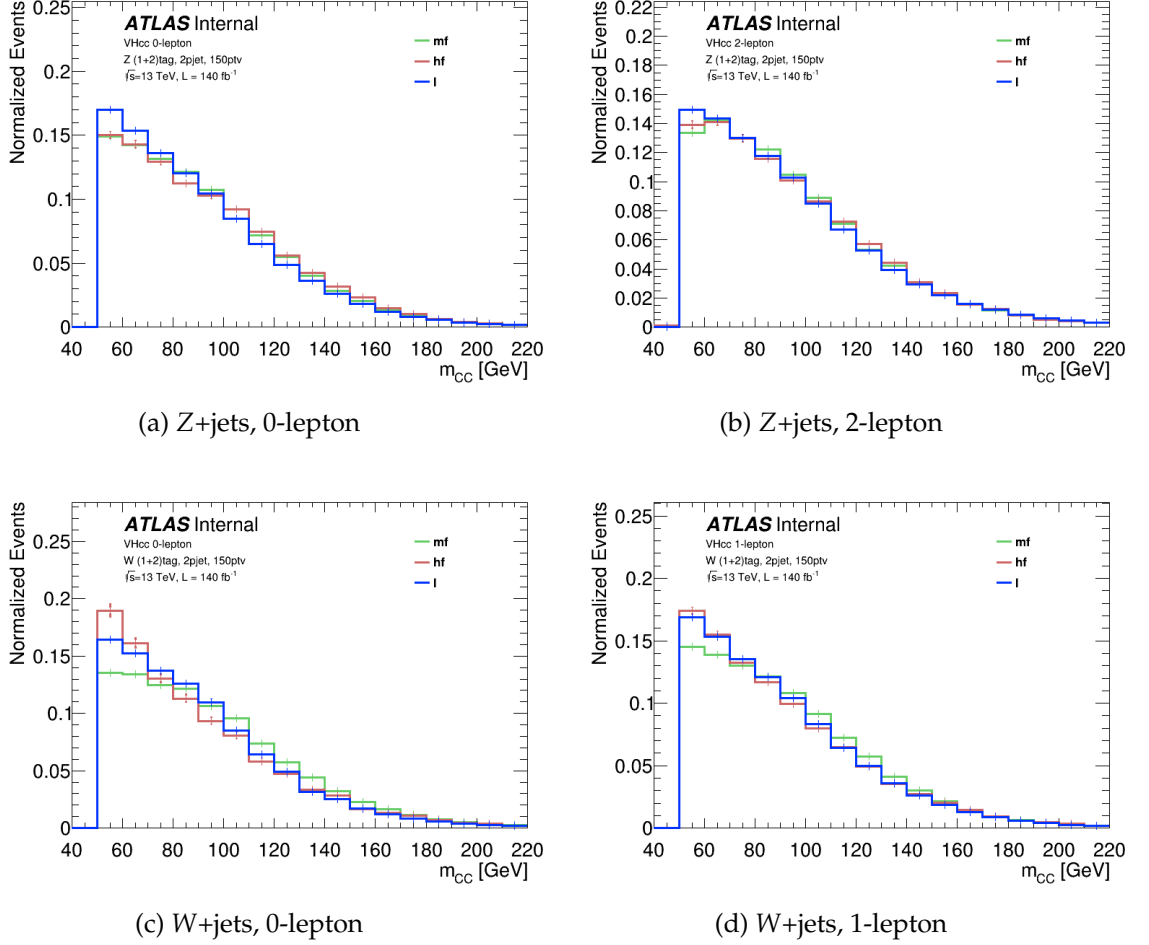


Figure 11.5: Normalized $m_{c\bar{c}}$ distributions of the $V+hf$ (red), $V+mf$ (green) and $V+l$ (blue) categories for (a) a Z boson in the 0-lepton channel, (b) a Z boson in the 2-lepton channel, (c) a W boson in the 0-lepton channel, and (d) a W boson in the 1-lepton channel. All plots correspond to the combined (1+2) c -tag, (2+3) jet, high- p_T^V region.

A summary of the normalization uncertainties of the W+jets and Z+jets backgrounds are presented in Table 11.2. The Z+jets backgrounds are negligible in the 1-lepton channel and the W+jets backgrounds are negligible in the 2-lepton channel. Hence the corresponding normalization uncertainties are not assigned in these channels.

11.3.3.3 | Acceptance ratios

Acceptance ratios for the V +jets background processes are calculated between jet multiplicity categories, the 0/1/2-lepton channels, between the flavor compositions of the $V+hf$ and $V+mf$ categories, and between the signal region and the Top CR. Acceptance

Channel	Background	Normalization Uncertainty
0-lepton and 2-lepton	$Z + hf(bb, cc)$	Floating, de-correlated in p_T^V
	$Z + mf(bc, bl, cl)$	Floating, de-correlated in p_T^V
	$Z + l$	Floating, de-correlated in p_T^V and jet multiplicity
0-lepton and 1-lepton	$W + hf(bb, cc)$	Floating
	$W + mf(bc, bl, cl, c\tau_h, b\tau_h)$	Floating
	$W + l(l, l\tau_h)$	Floating, de-correlated in jet multiplicity

Table 11.2: V +jets normalizations.

ratios between SR and the high ΔR CR are not explicitly derived for the V +jets processes, as normalization differences between these regions are accounted for by the V +jets $m_{c\bar{c}}$ shape uncertainties, as described in the following sub-section.

The values presented in the tables of App. A represent the prior values assigned to the acceptance uncertainty NPs in the fit. They are calculated using (1+2) c -tag events with the squared sum of three terms such as in Eq. (11.3), corresponding to the alternative MC generator and to variations of the regularization and factorization scales. Out of the three terms, the most dominant for the V +jets processes is typically the alternative generator. Varying μ_R generally has a smaller impact, though in a few cases it is of a similar size to that of the alternative MC. The term arising from variations of μ_F has a negligible effect on the final uncertainty. Note that even though different priors may be assigned in the SR and in the high ΔR CR, the NP themselves are correlated, such that they can be pulled in the fit in the same magnitude and direction.

11.3.3.4 | Shape uncertainty

Aside from accounting for any shape effects in the $m_{c\bar{c}}$ distributions, the $m_{c\bar{c}}$ shape uncertainty of the V +jets processes also takes into account the acceptance uncertainty between the SR and the high ΔR CR. This is achieved using two different sets of shape systematics. The first, labeled as *Systematic 1*, is derived from a fit to the ratio between the alternative and nominal MC in the high ΔR CR only, and is applied as a shape and normalization uncertainty to the signal and CR in a correlated way. This allows to extrapolate the shape difference from the control region to the signal region, and also mitigate acceptance effects between the two regions.

The second systematic, labeled *Systematic 2*, is a residual uncertainty added to introduce additional freedom to the $m_{c\bar{c}}$ shape in the signal region. This systematic is derived by reweighting the nominal (SHERPA 2.2.1) $m_{c\bar{c}}$ distributions by *Systematic 1*. Then the ratio between the reweighted SHERPA $m_{c\bar{c}}$ template and the alternative (MADGRAPH+PYTHIA

8) distribution is calculated and parametrized, and used as the second systematic uncertainty.

The alternative to nominal MC $m_{c\bar{c}}$ ratios for the V +jets sub-processes in the high- p_T^V region are presented in Figure 11.6. The plots show a comparison between the ratios calculated in the 1- and 2- c -tag and 2- and 3(+)-jets categories. If the ratios are compatible between the regions they are combined and fitted together to form a single shape. The fitted black curve corresponding to *Systematic 1* is also shown. The ratio distributions exhibit large statistical fluctuation at low $m_{c\bar{c}}$, due to the limited number of events in the high ΔR CR with $m_{c\bar{c}} \lesssim 100$ GeV. However, these fluctuations have little impact on the resulting *Systematic 1*, as the fit gives a larger weight to data points with small statistical errors.

The plots in Figure 11.6 show that the alternative to nominal MC ratios are consistent between the 1 c -tag and 2 c -tag regions. Therefore, to increase the statistical significance of the dataset the shape uncertainties are derived using a joined (1+2) c -tag category. The ratios of Z +hf, W +hf, Z +mf, and W +mf are compatible between the 2-jet and 3(+)-jet categories, hence a single shape is used to describe each of these processes. For Z +l and W +l however, the ratios derived in the 2-jet and 3(+)-jet regions are incompatible, and a dedicated shape is used in each category. The Z +jets shapes are all derived using the 0-lepton channel, and the W +jets shapes using the 1-lepton channel. In addition, a separate set of shapes are also derived for Z +jets in the low- p_T^V region of the 2-lepton channel.

Figure 11.7 shows similar plots for the residual signal region shape uncertainties (*Systematic 2*). In most cases, these residual shape uncertainties are small, with exception of Z +l and W +l 2-jet shapes where a noticeable trend is observed, going in the opposite direction of *Systematic 1*. This is the intended behavior of this approach, i.e. if the high ΔR CR constrains the $m_{c\bar{c}}$ shapes in the SR in a particular way, the residual SR uncertainty is able to correct the $m_{c\bar{c}}$ shape in the signal region according to data in the SR.

11.3.4 | Top background

The backgrounds originating from the decays of top quarks, i.e. $t\bar{t}$ and single-top, are considered together in the derivation of the modeling systematics.

Single-top processes are a minor background in the 0-lepton and 1-lepton channels,

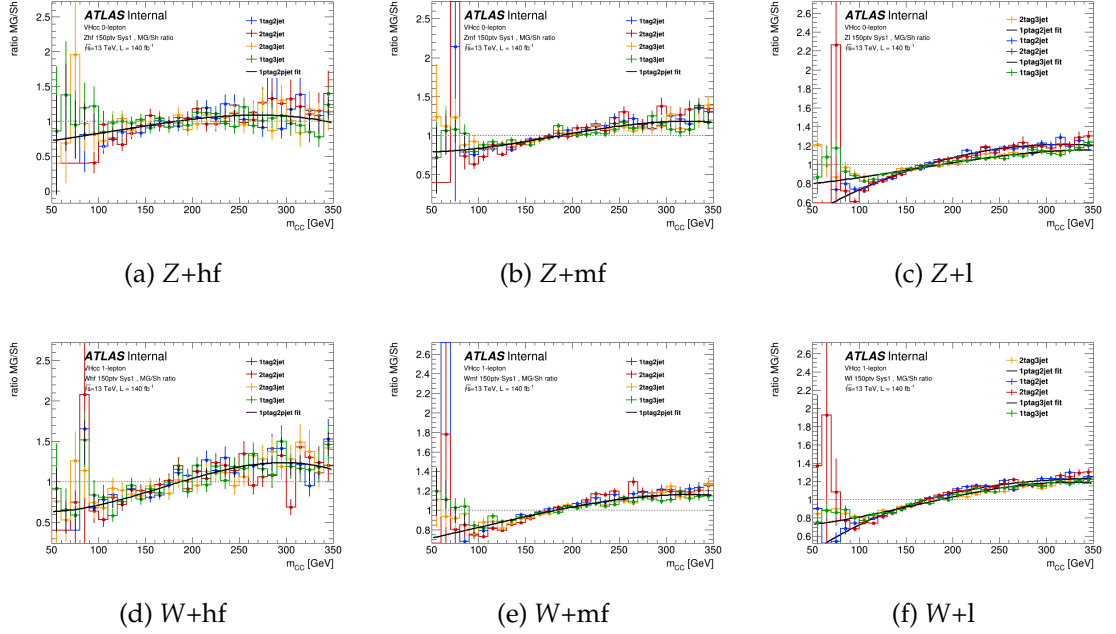


Figure 11.6: Ratio of alternative (MADGRAPH+PYTHIA 8) over nominal (SHERPA 2.2.1) normalized $m_{c\bar{c}}$ shapes for the V +jets sub-processes in high- p_T^V high ΔR CR. The fitted black curve(s) corresponding to the shape uncertainty used in the final fit (*Systematic 1*) are also shown in the plot.

with the main component coming from the production in the Wt -channel, whereas the t -channel and s -channel are sub-dominant, with a contribution of less than 1% of the total background. As a consequence, for all three channels a cross-section uncertainty is included in the fit, but only for the Wt -channel component modeling systematic uncertainties are considered.

In the 0-lepton and 1-lepton channels the top backgrounds are constrained using the top CR, which is constructed by inverting the b -tagging veto requirement on the third jet in an event, as described in Section 8.4.2. One control region is introduced for each channel, and only the 1 c -tag events are used. In the 2-lepton channel the contribution of the top backgrounds is small, and directly constrained from a dedicated $e\mu$ control region, also described in Section 8.4.2.

11.3.4.1 | Flavor composition and normalization

In the 0-lepton and 1-lepton channels, the top background events are categorized into two groups based on the flavor composition of the leading two jets. The **Top(bq)** category corresponds to events where the Higgs is reconstructed from at least one b -jet

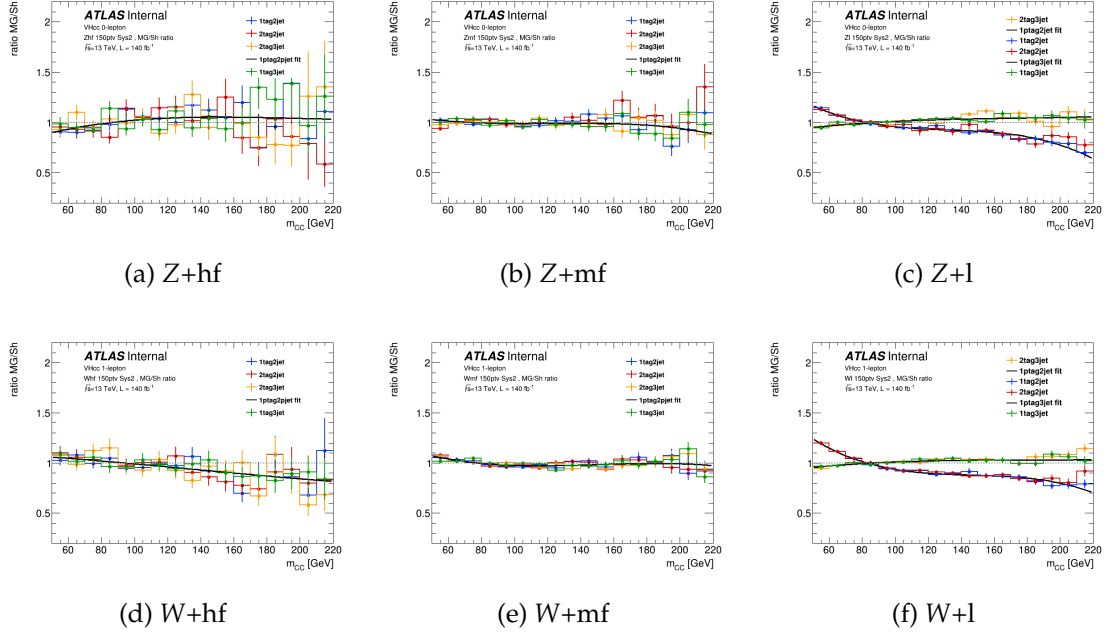


Figure 11.7: Ratio of alternative (MADGRAPH+PYTHIA 8) over nominal (Sherpa 2.2.1) normalized $m_{c\bar{c}}$ shapes after the latter had been reweighted by *Systematic 1*. The ratios are derived for the V +jets sub-processes in the high- p_T^V signal region. The fitted black curve(s) corresponding to the shape uncertainty used in the final fit (*Systematic 2*) are also shown in the plot.

coming from a top quark decay. It contains di-jet flavors of bb , bc , bl , and $b\tau$, and is characterized by a $m_{c\bar{c}}$ distribution with a broad peak at ~ 130 GeV. The **Top(lq)** category on the other hand comprises events where both jets are produced from the decay of the W -boson. It contains di-jet flavors of cl , ll , and cc , and peaks under the W mass at 80 GeV. One may notice that the cc component enters the Top(lq) category even though it is impossible for a W boson to decay to a pair of charm quarks. This choice is motivated by the $m_{c\bar{c}}$ shape of this component which also peaks under the W mass and is more similar to that of the Top(lq) processes. This is illustrated in Figure 11.8(a) which shows the normalized $m_{c\bar{c}}$ distributions of $t\bar{t}$ in the SR. For simplicity, the plot only shows the 1-lepton channel, 1 c -tag, 3-jet region, but the picture is similar in the other categories as well. One possible explanation for the shape of cc component is a bias in the flavor identification of c -jets in MC, where light-jets are mistakenly labeled as c -jets. In any case, as the cc component accounts for a small fraction of the total top background events (less than 1% in the 1 c -tag category and less than 3% in the 2 c -tag category) this choice has a negligible impact on the modeling of the top processes. The flavor composition of the $t\bar{t}$ backgrounds is shown in Figure 11.8(b), again, for the 1-lepton channel, 1 c -tag, 3-jet SR. Note also that the plots in Figure 11.8 only show events with $150 < p_T^V < 250$ GeV.

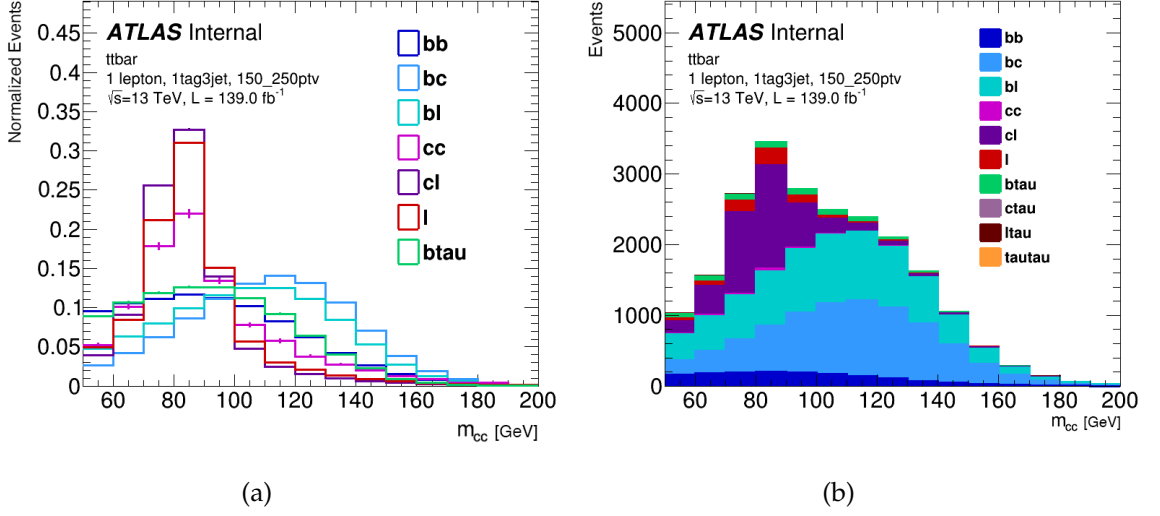


Figure 11.8: $m_{c\bar{c}}$ distribution of the $t\bar{t}$ process flavor components in the 1-lepton, 1 c -tag, 3-jet signal region. Normalized distributions are shown in (a) to demonstrate the shape differences between the various flavor components. The flavor composition is presented in (b) to illustrate the relative contribution of each component to the total $t\bar{t}$ background. Only events with $150 < p_T^V < 250$ GeV are considered in the plots.

Based on this categorization two floating normalizations are defined in the 0-lepton and 1-lepton channels, one for Top(bq) and the other for Top(lq). The normalization factor is shared between the 0- and 1-lepton channels and between the $t\bar{t}$ and single-top Wt -channel processes. The motivation for this approach is that the Top(bq) component is peaked close to the Higgs mass, and the Top(lq) is peaked underneath the Diboson signals. Floating the different top components separately adds flexibility to the fit by decoupling the two regions in $m_{c\bar{c}}$, and thus disentangling the signals.

In the 2-lepton channel the $t\bar{t}$ background is constrained using the top $e\mu$ control region, which is introduced into the fit as four single-bin regions (i.e. a control region for each p_T^V and jet multiplicity category). Each control region has its separate floating normalization, due to the large available statistics. The normalizations of the $t\bar{t}$ background in the 2-lepton channel are de-correlated from that of the Top(bq) and Top(lq) in the 0- and 1-lepton channels. The contribution from single-top processes is very small in the 2-lepton channel hence the normalization derived in the other channels is applied to single-top Wt . A summary of the normalizations of the top background is presented in Table 11.3.

Channel	Background	Normalization Uncertainty
0- & 1-lepton	Top(bq): $t\bar{t}$ & single-top Wt (bb,bc,bl,b τ)	Floating
	Top(bq): $t\bar{t}$ & single-top Wt (cl,ll,cc)	Floating
2-lepton	$t\bar{t}$	Floating, de-correlated in jet multiplicity and p_T^V
0-, 1-, & 2-lepton	single-top s-channel	cross-section uncertainty
	single-top t-channel	cross-section uncertainty

Table 11.3: Normalizations of the top backgrounds.

11.3.4.2 | Acceptance ratios

In the 2-lepton channel acceptance ratios are not considered as the normalizations are floated separately in each region. In the 0-lepton and 1-lepton channels, acceptance ratios are calculated between the signal and control region, between the different jet multiplicity categories, between the $t\bar{t}$ and single-top Wt processes, and between the 0- and 1-lepton channels. The acceptance ratios are again calculated using Eq. (11.3), considering two alternative MC simulations (POWHEG+HERWIG 7 which account for parton shower variation, and MADGRAPH5_aMC@NLO+PYTHIA 8 accounting for matrix element uncertainty) and the nominal MC with additional ISR/FSR (accounting for variation of μ_R). The description and treatment of the acceptance ratios are given in the discussion below, and the resulting priors are listed in the tables of App A.

The SR/CR acceptance ratio is mainly driven by the alternative MC generators. Two NP are used, one for the Top(bq) and another for the Top(lq) component. In addition, within each flavor component different priors are assigned to the $t\bar{t}$ and single-top Wt processes.

A single NP is assigned for the acceptance uncertainty between jet multiplicity regions. However, the calculated ratios are again different for the $t\bar{t}$ and single-top processes, and as such different priors are used for these processes.

The acceptance uncertainty between Wt and $t\bar{t}$ is mainly driven by the parton shower variation (HERWIG 7). The uncertainty is only applied to the Top(bq) component, as the $m_{c\bar{c}}$ distribution of the $t\bar{t}$ (lq) and single-top Wt (lq) has a very similar shape, making it impossible for the fit to distinguish between the two.

The acceptance uncertainty between the 0- and 1-lepton channels is de-correlated between the Top(bq) and Top(lq) components.

11.3.4.3 | Shape uncertainties

Uncertainties on the $m_{c\bar{c}}$ shape are considered only in the 0-lepton and 1-lepton channels, as the $e\mu$ control regions in the 2-lepton channels are comprised of a single bin and thus effectively has no shape. Each of the three uncertainty sources mentioned above in the context of the acceptance ratios is associated with its own shape uncertainty, derived from a fit to the ratio with respect to the nominal MC prediction. The fit is performed using data in the 1-lepton channel, and the derived shapes are applied to both 0- and 1-lepton as they are in agreement between the channels. The shapes are derived separately for the bq and lq flavor components, inclusively in the number of jets, and using a single shape for the SR and top-CR. Even though different shapes are considered for $t\bar{t}$ and single-top Wt , the shape uncertainties on these two processes are correlated and are governed by a single NP. This makes a total of six nuisance parameters associated with the $m_{c\bar{c}}$ shape uncertainty of the top processes, as summarized in Table 11.4. Note that given the non-negligible contribution of $t\bar{t}$ (bq) in the high ΔR CR, the shape of this component is also derived and applied in the high ΔR CR.

Shape uncertainty	Processes	Regions derived in
POWHEG+HERWIG 7	Top(bq)	SR + top-CR + high ΔR CR
	Top(lq)	SR + top-CR
MADGRAPH5_aMC@NLO+PYTHIA 8	Top(bq)	SR + top-CR + high ΔR CR
	Top(lq)	SR + top-CR
POWHEG+PYTHIA 8 + ISR/FSR	Top(bq)	SR + top-CR + high ΔR CR
	Top(lq)	SR + top-CR

Table 11.4: Summary of the shape uncertainties considered for the top backgrounds.

11.3.5 | Diboson

Similarly to the other backgrounds, modeling uncertainties for the diboson processes are derived by comparison to the alternative (POWHEG + PYTHIA 8) MC generator and to the nominal MC with variations of μ_R and μ_F .

11.3.5.1 | Normalization uncertainty

Four different diboson components are considered:

- ZZ , where one of the Z -bosons decays leptonically and one hadronically
- $W_{lep}Z_{had}$, where the W -boson decays leptonically and the Z hadronically

- $Z_{lep}W_{had}$, where the Z-boson decays leptonically and the W-boson decays hadronically
- WW, where one of the W-bosons decays leptonically and one hadronically

These processes are dominant in different channels, as summarized in Table 11.5.

Channel	ZZ	$W_{lep}Z_{had}$	$Z_{lep}W_{had}$	WW
0-lepton	20.8%	8.1%	33.0%	38.1%
1-lepton	0.8%	17.9%	1.1%	80.2%
2-lepton	40.7%	0.0%	59.3%	0.0%

Table 11.5: The relative contribution of the different diboson components in the three analysis channels.

An overall normalization uncertainty is derived by comparing the total number of events in the analysis signal regions generated using the nominal MC simulation, the alternative MC simulation, and the nominal MC with variations of the regularization and factorization scales. For the derivation of the normalization uncertainty the $W_{lep}Z_{had}$, $Z_{lep}W_{had}$ processes are correlated and are associated to a single nuisance parameter with a prior of 12%. A single normalization factor is also applied to each of the WW (10%) and ZZ (5%) processes. The normalization uncertainty is applied across all analysis regions and to all flavor combinations of the leading jets.

11.3.5.2 | Acceptance ratios

Acceptance ratios are considered between the jet multiplicity categories, the low- and high- p_T^V regions (in the 2-lepton channel), and between the three leptonic channels. They are derived separately for the WW, ZZ, and WZ components, yet within each process the same uncertainty is applied to all di-jet flavor combinations. Both the p_T^V and jet multiplicity acceptance ratios are de-correlated between $W_{lep}Z_{had}$ and $Z_{lep}W_{had}$ (though the same prior is used for both components). The acceptance ratios are not considered in the high ΔR CR due to the small number of events in that region.

11.3.5.3 | Shape uncertainties

For the derivation of the diboson shape uncertainties, the different components are then further split based on the leading jets flavors into signal and background, grouped according to the processes of interest:

- $VZ(cc)$: $qq \rightarrow ZZ(cc), gg \rightarrow ZZ(cc), W_{lep}Z(cc)$
- $VW(cl)$: $qq \rightarrow WW(cl), gg \rightarrow WW(cl), Z_{lep}W(cl)$
- VV background: all remaining diboson components

As for the other backgrounds, the shape uncertainties are derived from a fit to the normalized ratio of alternative over nominal MC generators. A polynomial shape is used to parameterize the shape of the ratio in the fit. A different fit is used to extract the shape of the signal and background diboson processes. However, after the shape is derived, the signal and background sub-processes are correlated again in the main fit, such that they are associated with a single nuisance parameter. The same shape is applied inclusively for the 2-jet and 3(+)-jet categories. The shape uncertainties in the 2-lepton low- p_T^V category are derived from a standalone fit and use a different shape than that of the high- p_T^V .

11.3.6 | Multijet

The multijet background estimation (described in detail in Section 8.5) uses data in a dedicated control region (referred to as inverted-isolation region, or MJ-CR) to obtain an $m_{c\bar{c}}$ template for the multijet process. The template is then fitted in the SR where the multijet normalization is determined. In this process, a number of uncertainty sources are considered that have an impact on the shape of the $m_{c\bar{c}}$ distribution, its normalization, or both. The systematic uncertainties that impact the shape are described first, since most of these are also considered for the normalization.

11.3.6.1 | Shape uncertainties

The modeling systematic uncertainties considered on the shape of the multijet $m_{c\bar{c}}$ distribution in the SR and high ΔR CR are:

- **Electron trigger (*e*-mode only)**: the impact of the choice of electron trigger on the multijet estimate is evaluated, as this may introduce a bias in the inverted isolation region. Instead of using the combination of triggers mentioned in Section 8.1, simply the lowest p_T trigger is used.
- **Reduced MJ-CR**: the uncertainty related to extrapolation from the MJ-CR to the SR is evaluated by considering a reduced MJ-CR. The reduced MJ-CR is defined

using additional cuts on the isolation criteria of Table 8.8. In the e -mode, this is done by additionally requiring $TopoE_T^{\text{cone20}} < 11$ GeV, and in the μ -mode, $p_T^{\text{cone20}} < 4.5$ GeV. The additional cuts are designed for keeping about half of data events in the full inverted isolation region.

- **Top and W scale factors in the MJ-CR:** the impact of using altering the normalization of the Top and W+jets processes in the electroweak background subtraction procedure is considered. The normalization scale factors applied on the Top and W+jets backgrounds are extracted from an additional fit to the $\Delta R(j_1, j_2)$ distribution, because m_T^W does a poor job differentiating these two processes (which is the reason for using a combined "electroweak" template to describe the background in the template fit). The multijet shape is evaluated with and without applying the normalization factors, and the difference in shape is taken as the systematic uncertainty. In addition, the same normalization factors are used in the signal region before merging the Top and W components to create the electroweak background component. This can change their relative proportions and modify the shape of the final electroweak template used in the fit, which in turn can impact the resulting multijet normalization scale factors. The fits to the $\Delta R(j_1, j_2)$ distribution are presented in Figure 11.9.

The uncertainties described above are estimated separately in the e - and μ -channels, and separately in the SR and high ΔR CR. Figure 11.10 shows the contribution of the different systematic sources to the $m_{c\bar{c}}$ shape uncertainty in the SR. From the plots, one can see that the main shape uncertainty comes from using a reduced MJ-CR and from using an alternative electron trigger. Applying Top and W normalization scale factors in the control region has a very small impact on the derived shape uncertainties.

11.3.6.2 | Normalization uncertainties

The sources of systematic uncertainty that have an impact on the $m_{c\bar{c}}$ shape are also considered as uncertainties on the multijet normalization. The impact on the multijet normalization is indirectly driven by changes to the m_T^W templates, and to the relative yield in the signal regions. The individual contributions to the normalization systematic uncertainty in the signal region are added in quadrature, and then added in quadrature to the statistical error from the m_T^W fit to give the overall normalization uncertainty, calculated separately in each signal region and for the e - and μ -modes. In the high ΔR CR the systematic component is not considered and only the statistical error is taken as the uncertainty on the normalization.

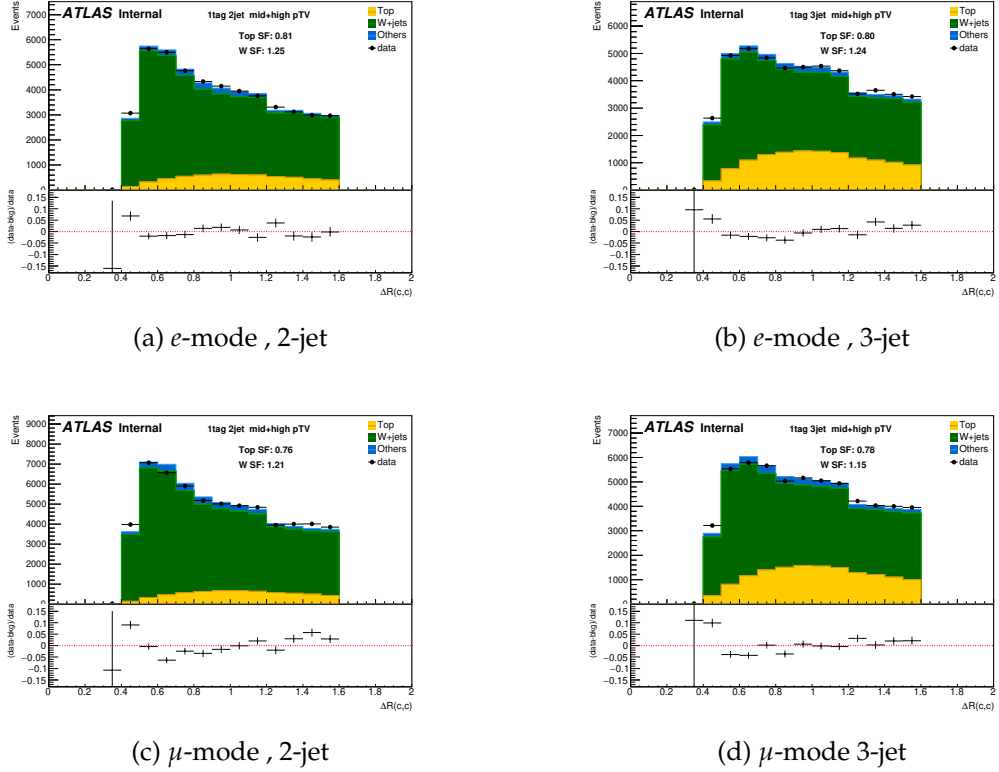


Figure 11.9: Fits to the $\Delta R(j_1, j_2)$ distribution in the analysis signal region. The normalization of the Top and W+jets background components are floated in the fit, and the resulting scale factors are used to evaluate a systematic uncertainty. The normalization of the other background processes is fixed to the prediction from MC.

In addition to the sources already described above in the context of the shape uncertainty, the following sources are considered exclusively as normalization uncertainties:

- **Including the $E_T^{\text{miss}} < 30 \text{ GeV}$ region in the template fit (*e-mode only*):** this induces a significant change to the m_T^W distribution both for the multijet component derived from the inverted isolation region in data, and for the electroweak background components estimated from simulation.
- **Using a different variable:** Instead of m_T^W , the azimuthal difference between the lepton and E_T^{miss} , $\Delta\Phi(\text{lep.}, E_T^{\text{miss}})$, is used as the variable in the template fit.
- **Applying the $m_T^W < 120 \text{ GeV}$ cut:** this cut, which is not applied in the nominal multijet background extraction, is now applied both in the SR and in the MJ-CR.

The combination of these sources gives rise to systematic normalization uncertainty

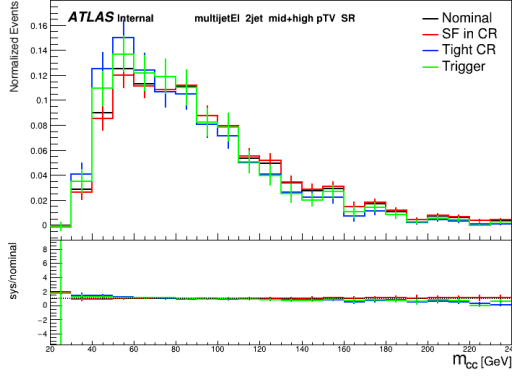
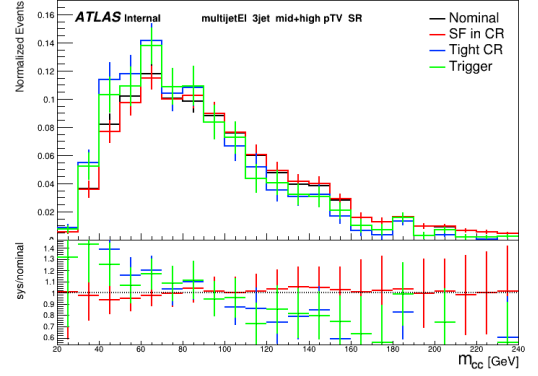
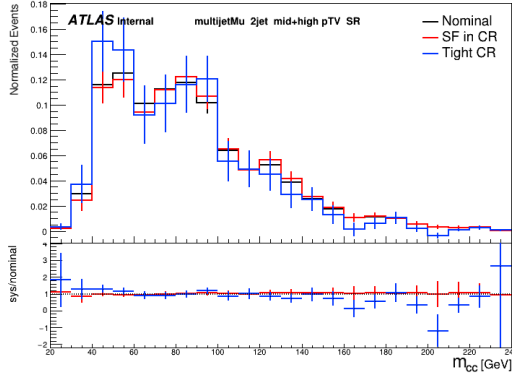
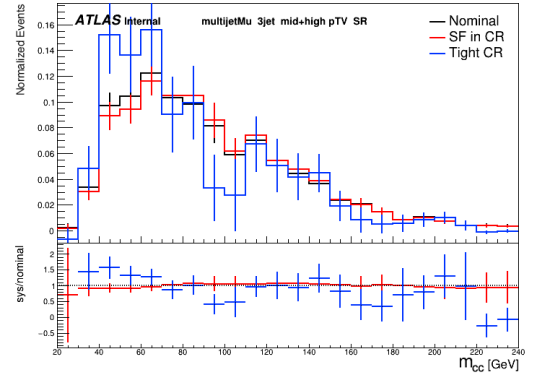
(a) e -mode , 2-jet(b) e -mode , 3-jet(c) μ -mode , 2-jet(d) μ -mode 3-jet

Figure 11.10: Multijet $m_{c\bar{c}}$ shape uncertainty in the analysis signal regions. The templates in the plots are normalized to unity.

on the multijet background. Out of the systematic sources, the largest contribution to the uncertainty comes from the inclusion of events with $E_T^{\text{miss}} < 30$ GeV (in the e -mode only), followed by the reduced MJ-CR and the use of $\Delta\Phi$ (lep., E_T^{miss}) in the fit.

The total normalization uncertainty, which includes also the statistical component, is shown in Table 8.9 of Section 8.5. In the 1 c -tag region, the contributions of the systematic and statistical sources to the total uncertainty are of comparable size. In the 2 c -tag category however, where the data yields are much lower, the statistical component is the more dominant.

Results

The following chapter presents the results of the global likelihood fit. The baseline fit is performed across the 44 analysis regions to extract simultaneously three parameters of interest, corresponding to the signal strengths of $VH(cc)$, $VZ(cc)$, and $VW(cl)$.

First, the overall performance of the fit is analyzed in Section 12.1, by examination of the post-fit $m_{c\bar{c}}$ distribution, the resulting nuisance parameters pulls and their correlations. Then, Section 12.2 shows the measured signal strength parameters, and the corresponding 95% CL upper limit on the $VH(cc)$ process cross-section time branching ratio. Finally, in Section 12.3 the fit results are interpreted within the κ framework, which allows setting the first direct upper limit on the Yukawa coupling of the charm quark.

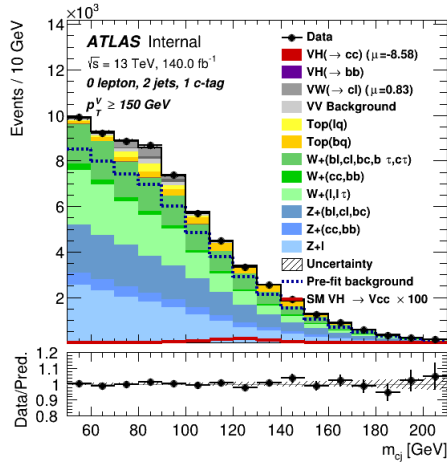
12.1 | Fit performance

Before looking at the result of the signal strength measurement, the performance of the fit must be analyzed and understood in terms of the behavior of the nuisance parameters. At this stage, where the analysis is said to be *blinded*, one can try and spot suspicious behavior which may be reflected in NPs being very pulled/constrained by the fit, groups of NPs with high correlations between them, or regions with large systematic differences between data and MC. At this stage, it is also possible to make changes to the fit model (or to any stage of the analysis chain) based on the observed fit performance. It is however worth pointing out that observing a behavior like described above does not necessarily imply there is an issue with the fit, but further investigation or understanding is needed. Once the analysis team is confident there are no underlying

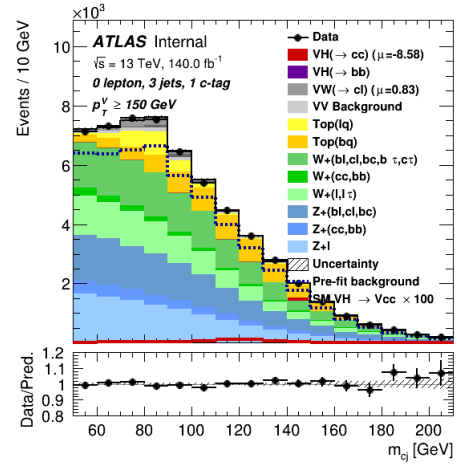
issues or biases in the fit model, further changes to the fit procedure are frozen, and the measured signal strength is *unblinded*.

12.1.1 | Post-fit $m_{c\bar{c}}$ distributions

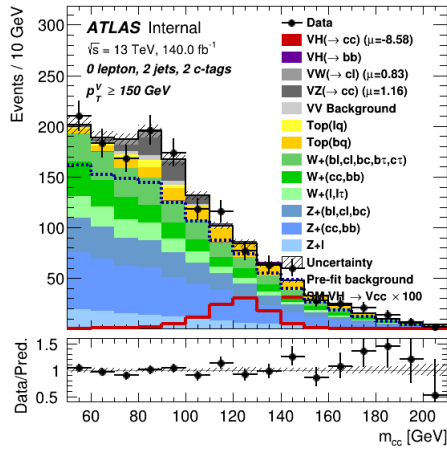
Figures 12.1-12.12 show the post-fit $m_{c\bar{c}}$ distributions after the unconditional fit to data. The plots are obtained from the combined (i.e. all three leptonic channels) fit, by setting the nuisance parameters to their best fit value. The x-axis in the figures is labeled m_{cc} for distribution in the 2 c -tag category, m_{cj} for distributions in the 1 c -tag category, and m_{jj} in the 0 c -tag CR. Overall, a good agreement between measured data and simulation can be observed.



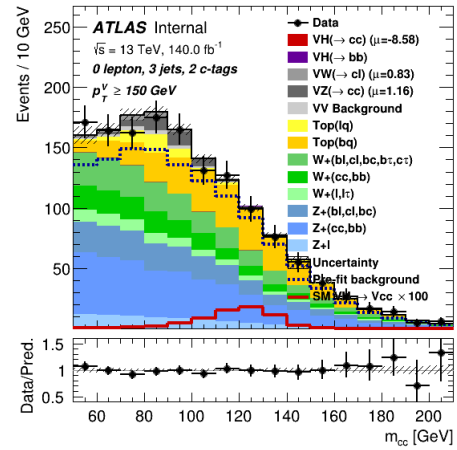
(a) 0-lepton, 1 c-tag, 2-jet, SR



(b) 0-lepton, 1 c-tag, 3-jet, SR



(c) 0-lepton, 2 c-tag, 2-jet, SR



(d) 0-lepton, 2 c-tag, 3-jet, SR

Figure 12.1: Post-fit $m_{c\bar{c}}$ distributions in the 0-lepton signal region. (a) show the 1 c-tag, 2-jet SR, (b) show the 1 c-tag, 3-jet SR, (c) show the 2 c-tag, 2-jet SR, and (d) show the 2 c-tag, 3-jet SR.

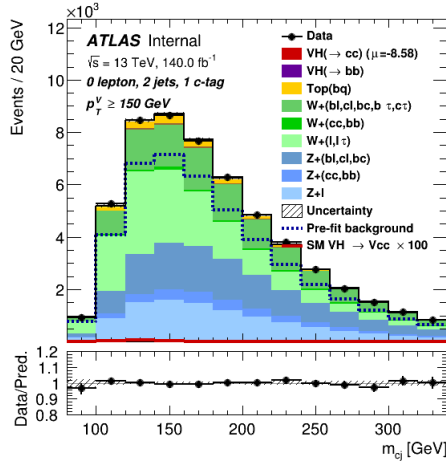
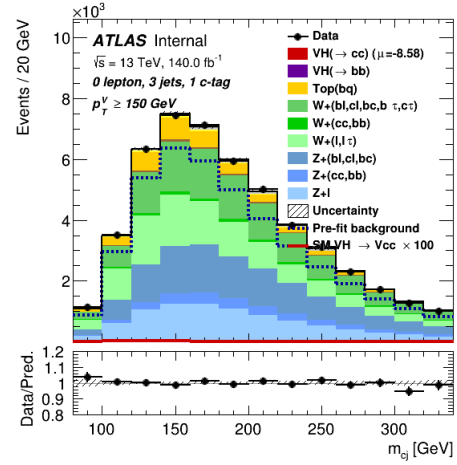
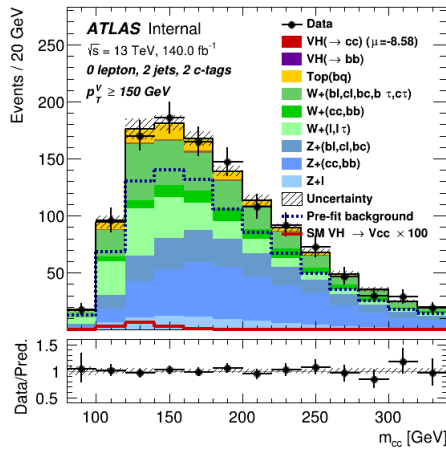
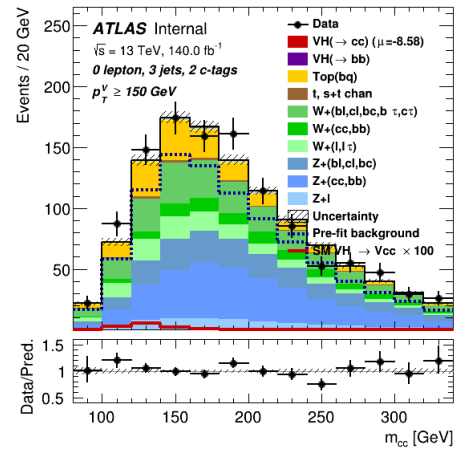
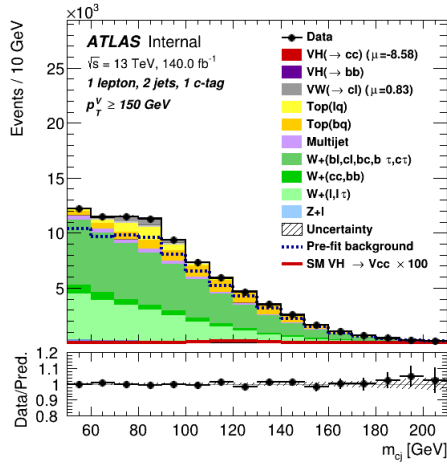
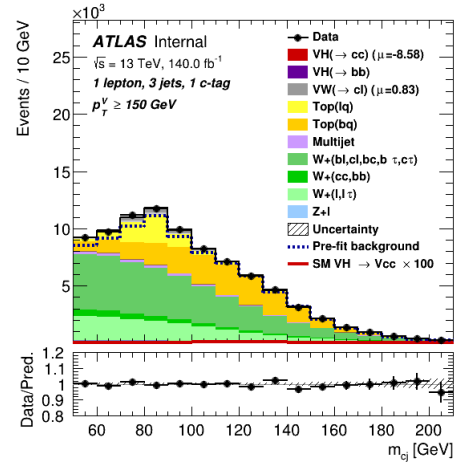
(a) 0-lepton, 1 c-tag, 2-jet, high ΔR CR(b) 0-lepton, 1 c-tag, 3-jet, high ΔR CR(c) 0-lepton, 2 c-tag, 2-jet, high ΔR CR(d) 0-lepton, 2 c-tag, 3-jet, high ΔR CR

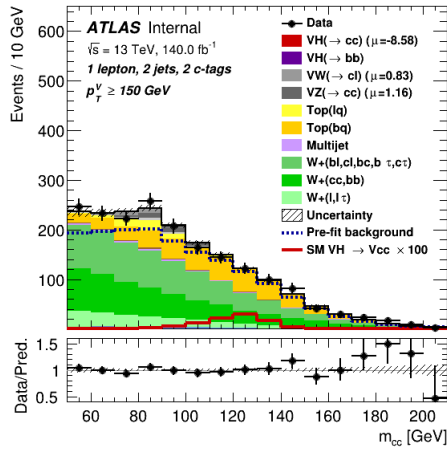
Figure 12.2: Post-fit $m_{c\bar{c}}$ distributions in the 0-lepton high ΔR control region. (a) show the 1 c-tag, 2-jet category, (b) show the 1 c-tag, 3-jet category, (c) show the 2 c-tag, 2-jet category, and (d) show the 2 c-tag, 3-jet category.



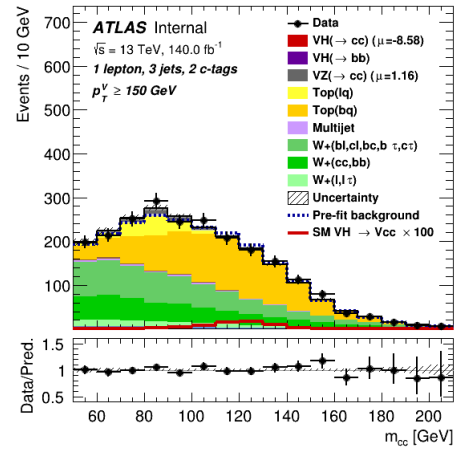
(a) 1-lepton, 1 c-tag, 2-jet, SR



(b) 1-lepton, 1 c-tag, 3-jet, SR



(c) 1-lepton, 2 c-tag, 2-jet, SR



(d) 1-lepton, 2 c-tag, 3-jet, SR

Figure 12.3: Post-fit $m_{c\bar{c}}$ distributions in the 1-lepton signal region. (a) show the 1 c-tag, 2-jet SR, (b) show the 1 c-tag, 3-jet SR, (c) show the 2 c-tag, 2-jet SR, and (d) show the 2 c-tag, 3-jet SR.

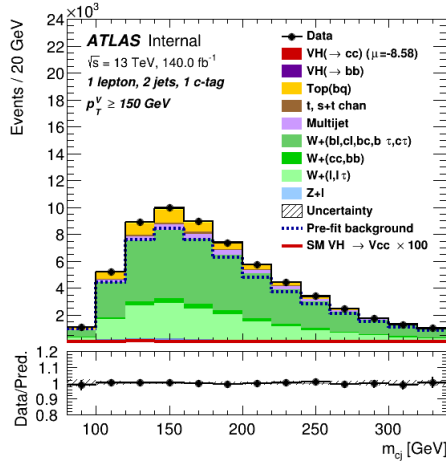
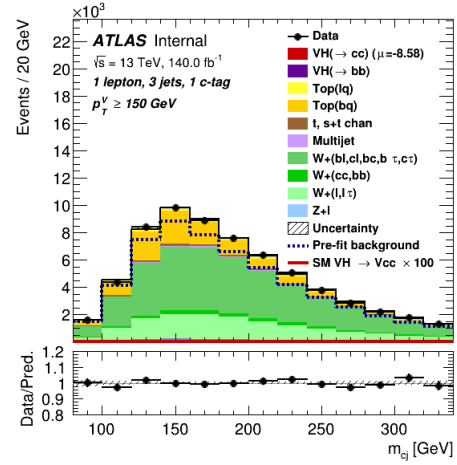
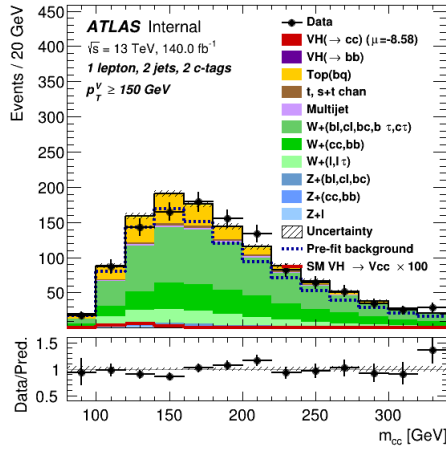
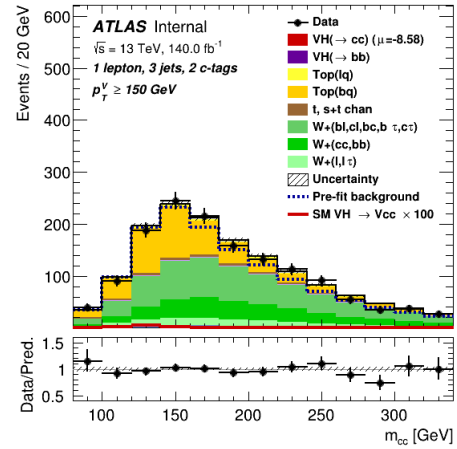
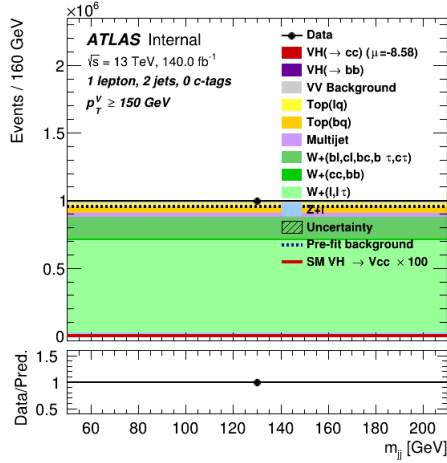
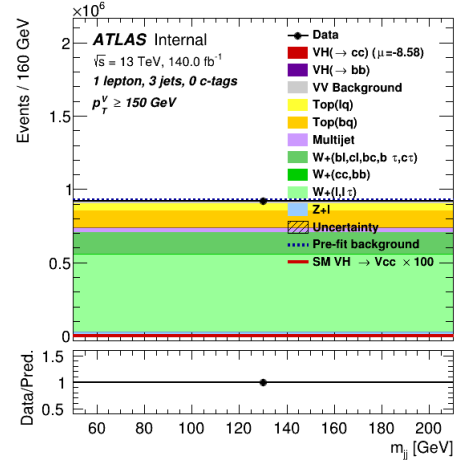
(a) 1-lepton, 1 c-tag, 2-jet, high ΔR CR(b) 1-lepton, 1 c-tag, 3-jet, high ΔR CR(c) 1-lepton, 2 c-tag, 2-jet, high ΔR CR(d) 1-lepton, 2 c-tag, 3-jet, high ΔR CR

Figure 12.4: Post-fit $m_{c\bar{c}}$ distributions in the 1-lepton high ΔR control region. (a) show the 1 c-tag, 2-jet category, (b) show the 1 c-tag, 3-jet category, (c) show the 2 c-tag, 2-jet category, and (d) show the 2 c-tag, 3-jet category.

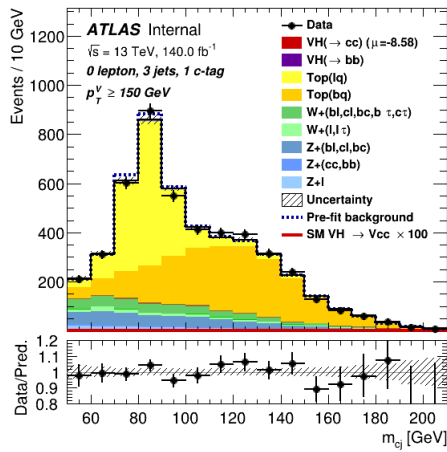


(a) 1-lepton, 2-jet, 0 c-tag CR

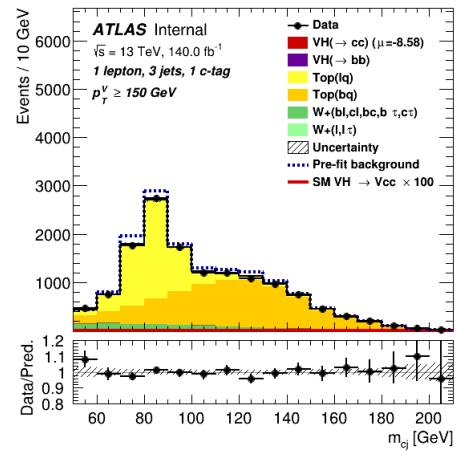


(b) 1-lepton, 3-jet, 0 c-tag CR

Figure 12.5: Post-fit $m_{c\bar{c}}$ distribution in the 1-lepton 0 c-tag CR. (a) show the 2-jet category and (b) show the 3-jet category.



(a) 0-lepton, 1 c-tag, 3-jet, top CR



(b) 1-lepton, 1 c-tag, 3-jet, top CR

Figure 12.6: Post-fit $m_{c\bar{c}}$ distributions in the 0-lepton (a) and 1-lepton (b) top CR.

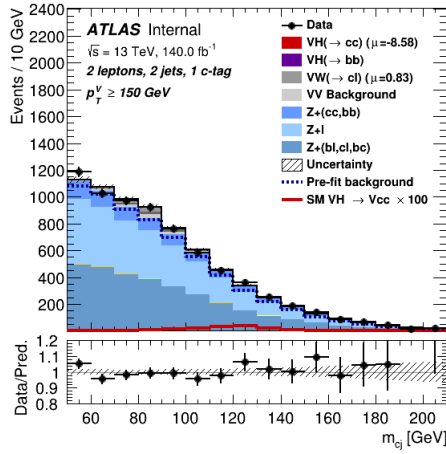
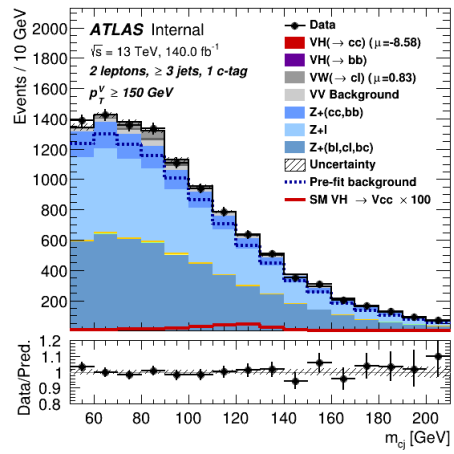
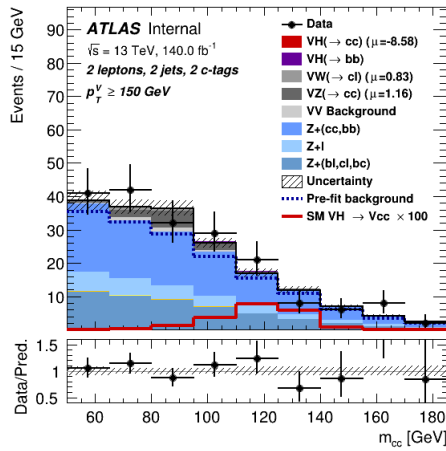
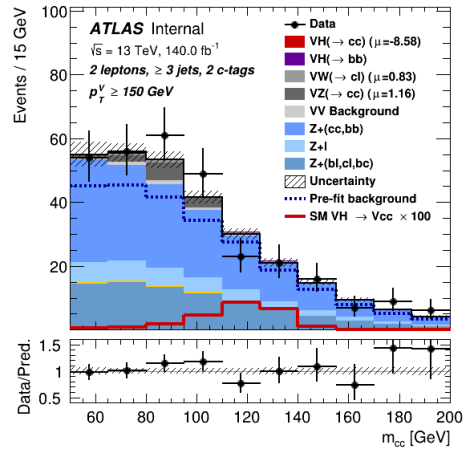
(a) 2-lepton, 1 c-tag, 2-jet, high- p_T^V SR(b) 2-lepton, 1 c-tag, 3+ jet, high- p_T^V SR(c) 2-lepton, 2 c-tag, 2-jet, high- p_T^V SR(d) 2-lepton, 2 c-tag, 3+ jet, high- p_T^V SR

Figure 12.7: Post-fit $m_{c\bar{c}}$ distributions in the 2-lepton high- p_T^V signal region. (a) show the 1 c-tag, 2-jet SR, (b) show the 1 c-tag, 3+ jet SR, (c) show the 2 c-tag, 2-jet SR, and (d) show the 2 c-tag, 3+ jet SR.

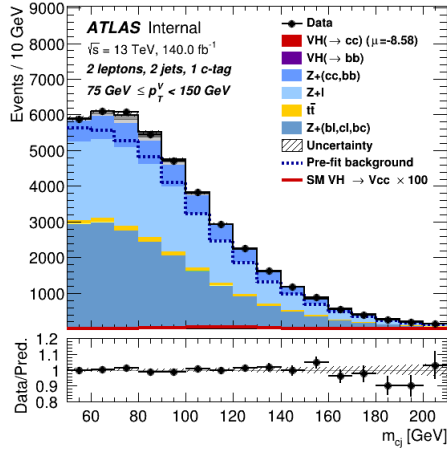
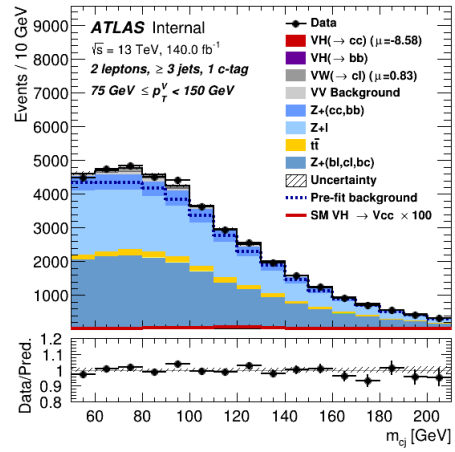
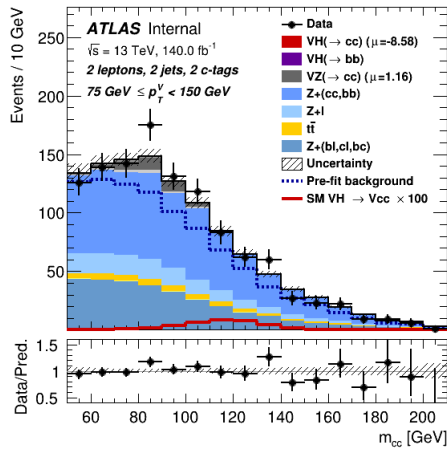
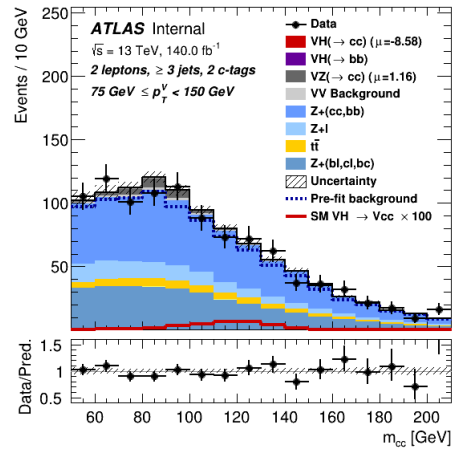
(a) 2-lepton, 1 c-tag, 2-jet, low- p_T^V SR(b) 2-lepton, 1 c-tag, 3+ jet, low- p_T^V SR(c) 2-lepton, 1 c-tag, 2-jet, low- p_T^V SR(d) 2-lepton, 1 c-tag, 3+ jet, low- p_T^V SR

Figure 12.8: Post-fit $m_{c\bar{c}}$ distributions in the 2-lepton low- p_T^V signal region. (a) show the 1 c-tag, 2-jet SR, (b) show the 1 c-tag, 3+ jet SR, (c) show the 2 c-tag, 2-jet SR, and (d) show the 2 c-tag, 3+ jet SR.

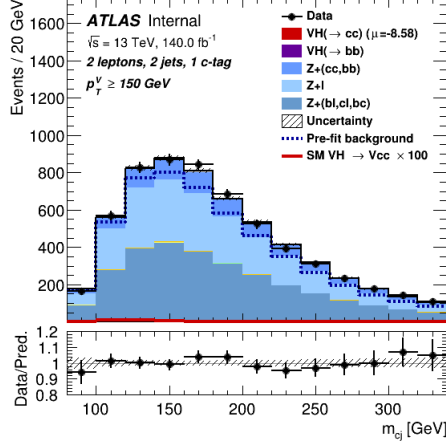
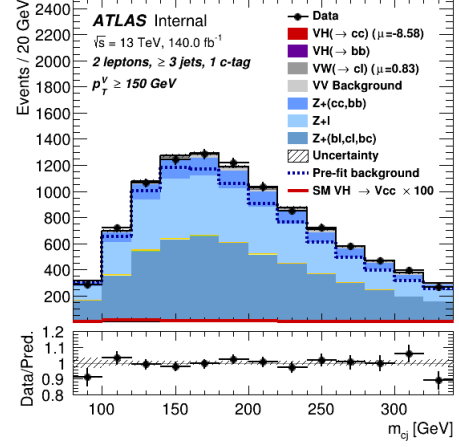
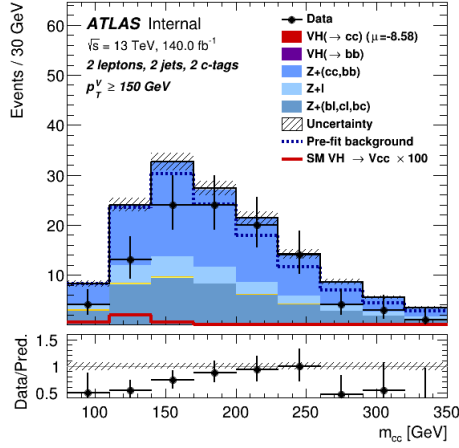
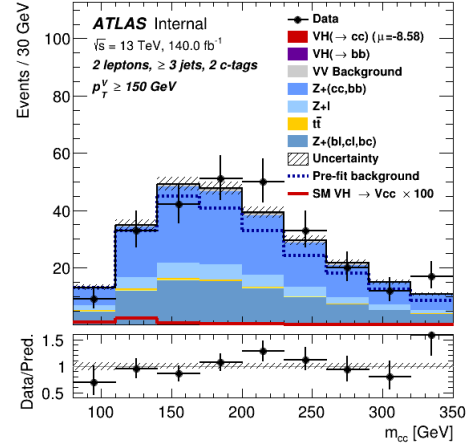
(a) 2-lepton, 1 c-tag, 2-jet, high- p_T^V high ΔR CR(b) 2-lepton, 1 c-tag, 3+ jet, high- p_T^V high ΔR CR(c) 2-lepton, 2 c-tag, 2-jet, high- p_T^V high ΔR CR(d) 2-lepton, 2 c-tag, 3+ jet, high- p_T^V high ΔR CR

Figure 12.9: Post-fit $m_{c\bar{c}}$ distributions in the 2-lepton high- p_T^V high ΔR control region. (a) show the 1 c-tag, 2-jet category, (b) show the 1 c-tag, 3+ jet category, (c) show the 2 c-tag, 2-jet category, and (d) show the 2 c-tag, 3+ jet category.

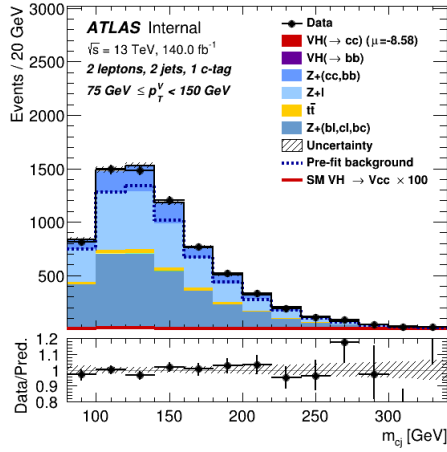
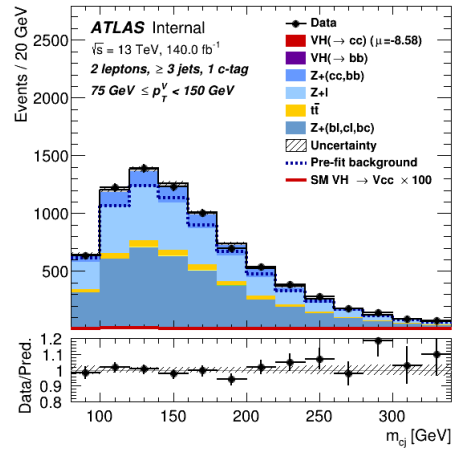
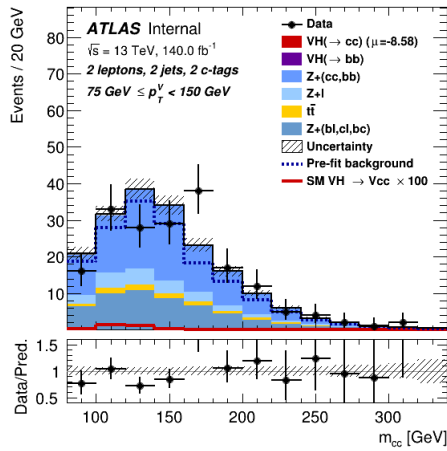
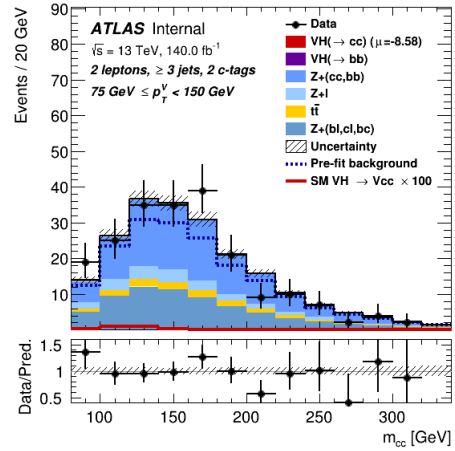
(a) 2-lepton, 1 c-tag, 2-jet, low- p_T^V high ΔR CR(b) 2-lepton, 1 c-tag, 3+ jet, low- p_T^V high ΔR CR(c) 2-lepton, 2 c-tag, 2-jet, low- p_T^V high ΔR CR(d) 2-lepton, 2 c-tag, 3+ jet, low- p_T^V high ΔR CR

Figure 12.10: Post-fit $m_{c\bar{c}}$ distributions in the 2-lepton low- p_T^V high ΔR control region. (a) show the 1 c-tag, 2-jet category, (b) show the 1 c-tag, 3+ jet category, (c) show the 2 c-tag, 2-jet category, and (d) show the 2 c-tag, 3+ jet category.

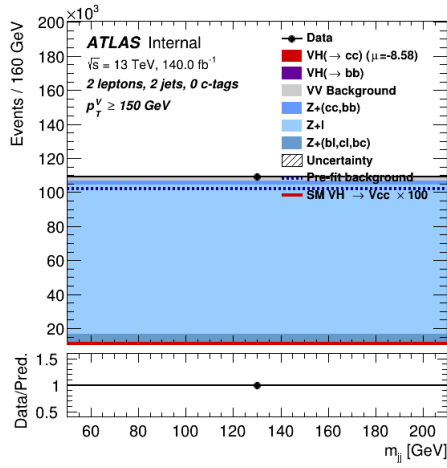
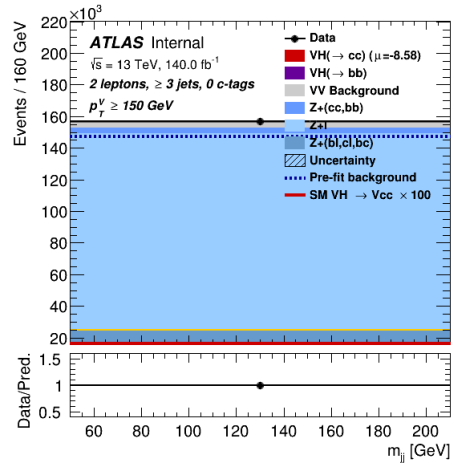
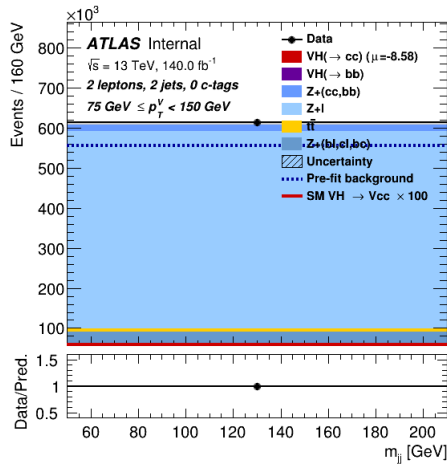
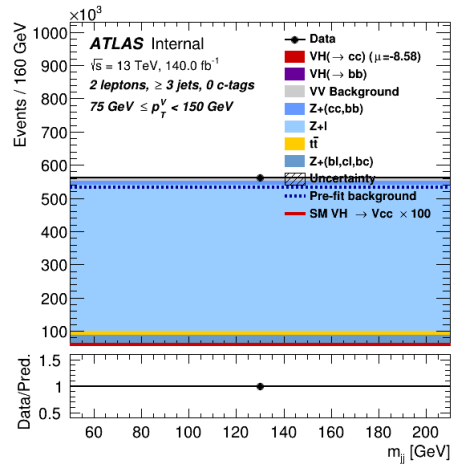
(a) 2-lepton, 2-jet, high- p_T^V , 0 c-tag CR(b) 2-lepton, 3+ jet, high- p_T^V , 0 c-tag CR(c) 2-lepton, 2-jet, low- p_T^V , 0 c-tag CR(d) 2-lepton, 3+ jet, low- p_T^V , 0 c-tag CR

Figure 12.11: Post-fit $m_{c\bar{c}}$ distributions in the 2-lepton 0 c-tag control region. (a) show the 2-jet, high- p_T^V category, (b) show the 3+ jet, high- p_T^V category, (c) show the 2-jet, low- p_T^V category, and (d) show the 3+ jet, low- p_T^V category.

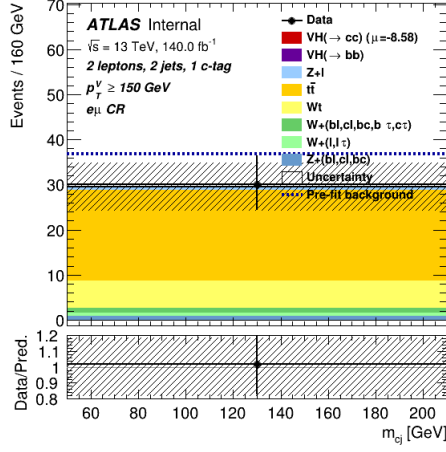
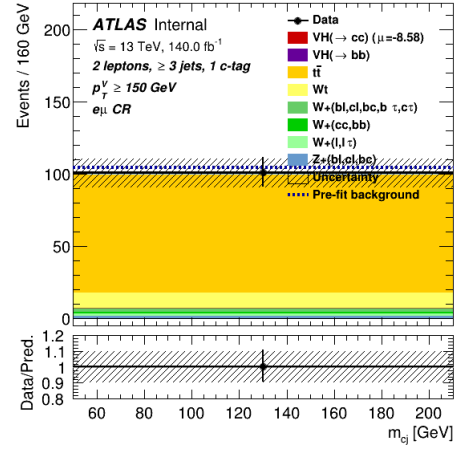
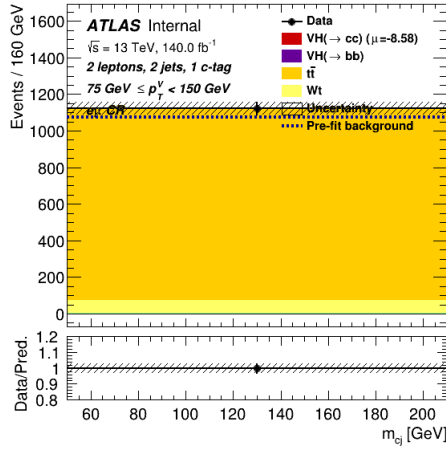
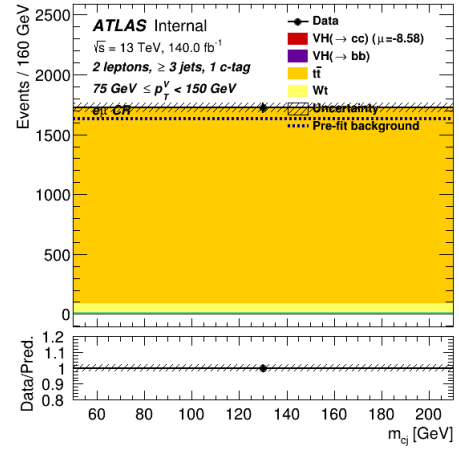
(a) 2-lepton, 1 c-tag, 2-jet, high- p_T^V , $e\mu$ CR(b) 2-lepton, 1 c-tag 3+ jet, high- p_T^V , $e\mu$ CR(c) 2-lepton, 1 c-tag, 2-jet, low- p_T^V , $e\mu$ CR(d) 2-lepton, 1 c-tag 3+ jet, low- p_T^V , $e\mu$ CR

Figure 12.12: Post-fit $m_{c\bar{c}}$ distributions in the 2-lepton top $e\mu$ control region. (a) show the 2-jet, high- p_T^V category, (b) show the 3+ jet, high- p_T^V category, (c) show the 2-jet, low- p_T^V category, and (d) show the 3+ jet, low- p_T^V category.

12.1.2 | Nuisance parameters pulls and correlations

The fitted values of the floating background normalization factors are shown in Table 12.1. The table shows that the normalization factors of the V +jets background are generally larger than 1, while for the top background they are below 1. This implies that the MC pre-fit prediction underestimates the number of V +jets events and overestimates that of $t\bar{t}$ events within the analysis phase-space.

Figures 12.13 and 12.14 show the NP pulls, defined in Equation 11.1, obtained from the combined fit to data (red) and to the Asimov dataset (black). Given the large number of nuisance parameters in the fit, they are grouped in the plot based on the source of the uncertainty and/or the process it is related to.

Background	p_T^V	Jets	Value
Top(bq)			0.91 ± 0.06
Top(lq)			0.94 ± 0.08
$t\bar{t}$ (2-lepton)	$p_T^V > 150$ GeV	2	0.76 ± 0.22
		3	0.96 ± 0.13
	$75 < p_T^V < 150$ GeV	2	1.08 ± 0.08
		3	1.06 ± 0.07
W+hf			1.16 ± 0.35
W+mf			1.28 ± 0.14
W+l		2	1.02 ± 0.04
		3	0.97 ± 0.05
Z+hf	$p_T^V > 150$ GeV		1.19 ± 0.22
	$75 < p_T^V < 150$ GeV		1.25 ± 0.25
Z+mf	$p_T^V > 150$ GeV		1.10 ± 0.15
	$75 < p_T^V < 150$ GeV		1.11 ± 0.15
Z+l	$p_T^V > 150$ GeV	2	1.07 ± 0.03
		3	1.08 ± 0.05
	$75 < p_T^V < 150$ GeV	2	1.12 ± 0.04
		3	1.07 ± 0.06

Table 12.1: Values of the free-floating background normalization parameters obtained from the likelihood fit to data. The uncertainties represent the combined statistical and systematic uncertainties. Unless otherwise stated, normalization parameters are correlated across all p_T^V and jet multiplicity analysis regions.

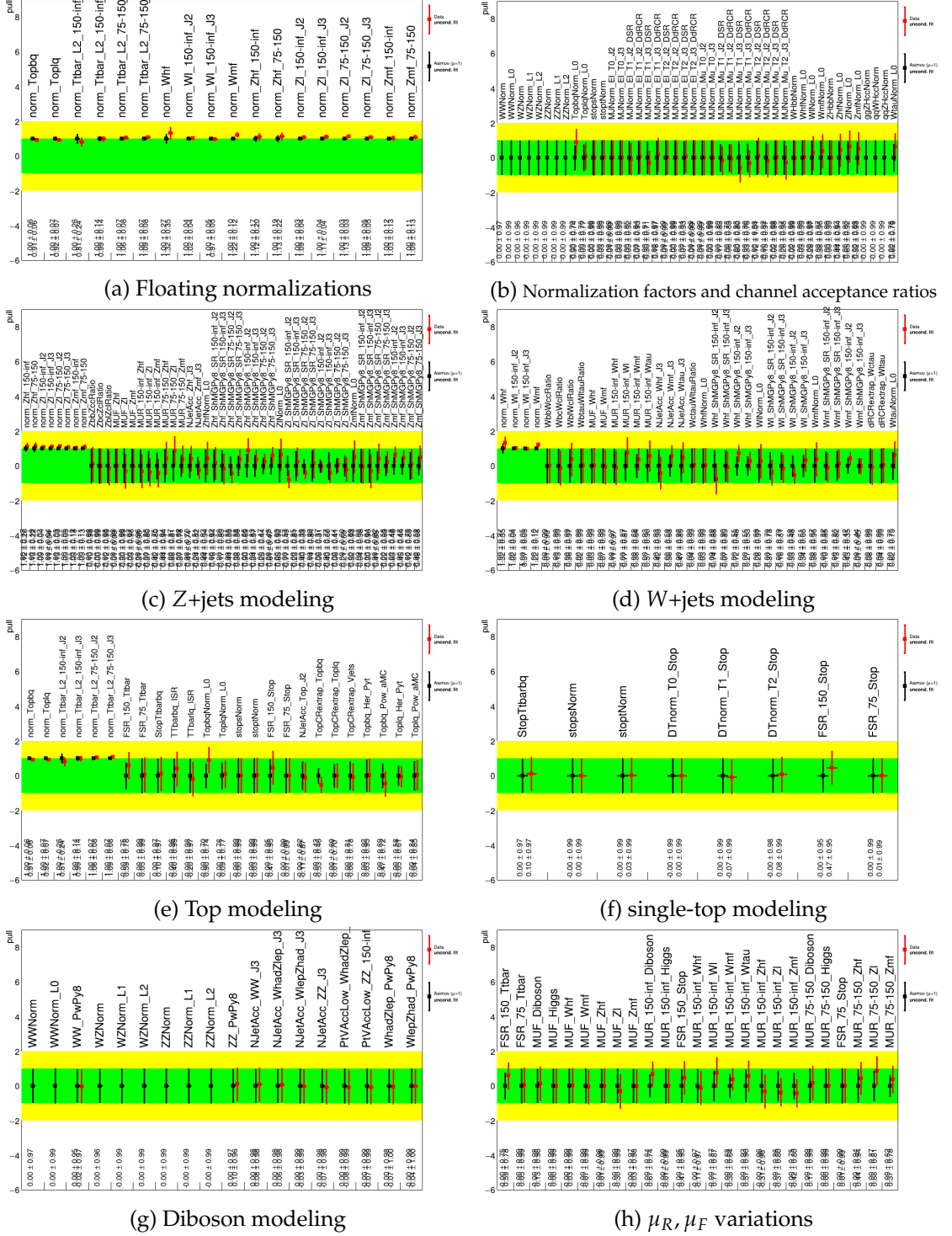
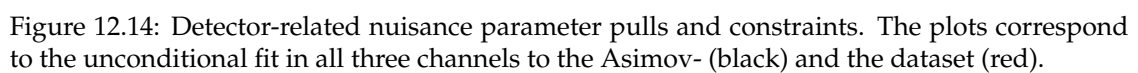


Figure 12.13: Monte Carlo modeling-related nuisance parameter pulls, constraints, and post-fit scale factors for the freely floating normalization parameters. The plots correspond to the unconditional fit in all three channels to the Asimov- (black) and the dataset (red).



In general, the constraints on the nuisance parameters derived in the fit to data agree well with those derived in the fit to the Asimov dataset. The floating normalizations of the major backgrounds can be constrained well by the data, in particular, due to the introduction of the control regions.

The Top(bq/lq) floating normalizations for the 0-lepton and 1-lepton channel are constrained to a 1σ -bands of around 7% for both flavor components, and a value slightly below 1 in the fit to data. The $t\bar{t}$ floating normalizations for the 2-lepton channel are constrained from the top $e\mu$ control region, with the largest 1σ -band in the 2-jet $p_T^V > 150$ GeV category.

The W+hf and W+mf floating normalizations are constrained for the signal and high ΔR control region, from the 1-lepton and 0-lepton channels. The constraint on the W+hf normalization is slightly larger in data compared to Asimov, with constraints of 35% and 27% respectively. The W+l normalizations are constrained from the 0 c -tag CR in the 1-lepton channel, with a 1σ -bands band of roughly 5%. The Z+hf and Z+mf floating normalizations are constrained for the signal and high ΔR control region in the 2-lepton and 0-lepton channel, where the low p_T^V category in 2-lepton has a dedicated floating normalization. For both p_T^V categories, Z+hf and Z+mf normalizations are constrained to an order of 20% and 13% respectively. Again, the Z+l normalizations are well constrained to approximately 5% from the 0 c -tag control region, in this case in 2-lepton.

The other experimental and modeling NPs are also reasonably well behaved, but there are some that are worth noting:

The V+jets modeling NPs are quite well constrained, and show some pulls, in particular for the $m_{c\bar{c}}$ shape NPs. The constraints are expected, due to the high ΔR control region, and because the shape NPs also have a normalization effect. The pulls are within 1σ , and thus not a major cause for concern.

The Top modeling NPs are generally quite well behaved. The TopCRextrap_Topbq NP, which is the acceptance uncertainty between SR and top CR, is constrained. This is expected from the large amount of Top background events in some of the signal regions and a relatively large prior on this NP, of the order of 10%. The TTbarNorm_LO_Topbq NP, which describes the acceptance between the 0-lepton and 1-lepton channel, is pulled to 1σ .

A number of the jet-related experimental NPs are strongly pulled, e.g. JER_DataVsMC and JER_EffectiveNP_5. The source of these pulls has been studied by de-correlating

the NP between channels and categories. The pull on JER_DataVsMC is driven by a few specific regions within the 1 c -tag category: 0-lepton 3-jet SR and high ΔR CR, 1-lepton 2-jet SR, and 2-lepton low- p_T^V 2-jet SR. For JER_EffectiveNP_5 no particular region has been found to drive the pull. A test has been performed by removing the JER NP from the fit. A negligible impact has been found on the POIs of the $VH(cc)$ and the two diboson processes.

The flavor tagging NPs are expected to be slightly constrained in this analysis. There is a small pull on the FT_C0 NP, which is associated with the c -tagging of c -jets. This has been studied by de-correlating this NP into the different channels, and the upwards pull was seen for all three channels, with the largest pull in the 1-lepton channel. The strong pull on the FT_L0 NP (associated with the c -tagging of light-jets) is driven by the 0-lepton channel. The FT_T0 NP (associated to the c -tagging of τ -jets) is pulled to -1σ , which is caused by the 0-lepton channel, where there is a relatively large τ -jet contribution.

Figure 12.15 shows the correlation matrix of the nuisance parameters and POIs obtained from the unconditional fit to data. Only nuisance parameters that have at least one correlation of magnitude 0.5 or higher with another nuisance parameter appear in the plot.

The correlations between the signal strength of $VH(cc)$ and the two diboson signal strengths are positive, 17% for $VW(cl)$ and 16% for $VZ(cc)$, whereas there is an anti-correlation of 17% between the two diboson signal strengths.

The correlation matrix is mostly diagonal, however it is possible to identify blocks of correlated NPs. As such, there are correlations among the floating normalization NPs, as well as between these NPs and the flavor tagging NPs, in particular the c -tagging efficiency NP for light jets (FT_L0) and c -jets (FT_C0). There are also correlations among the floating normalizations and the TT_dR NPs, which are expected as both have a normalization effect. Some anti-correlations can be observed between the different the V +jets modeling shape uncertainties, as well as between their jet-multiplicity acceptance ratios.

12.2 | Signal strength measurement and exclusion limits

The measured values of the POIs (i.e. the signal strengths of the $VH(cc)$, $VZ(cc)$, and $VW(cl)$ processes) obtained from the best fit to data are presented in Table 12.2. In

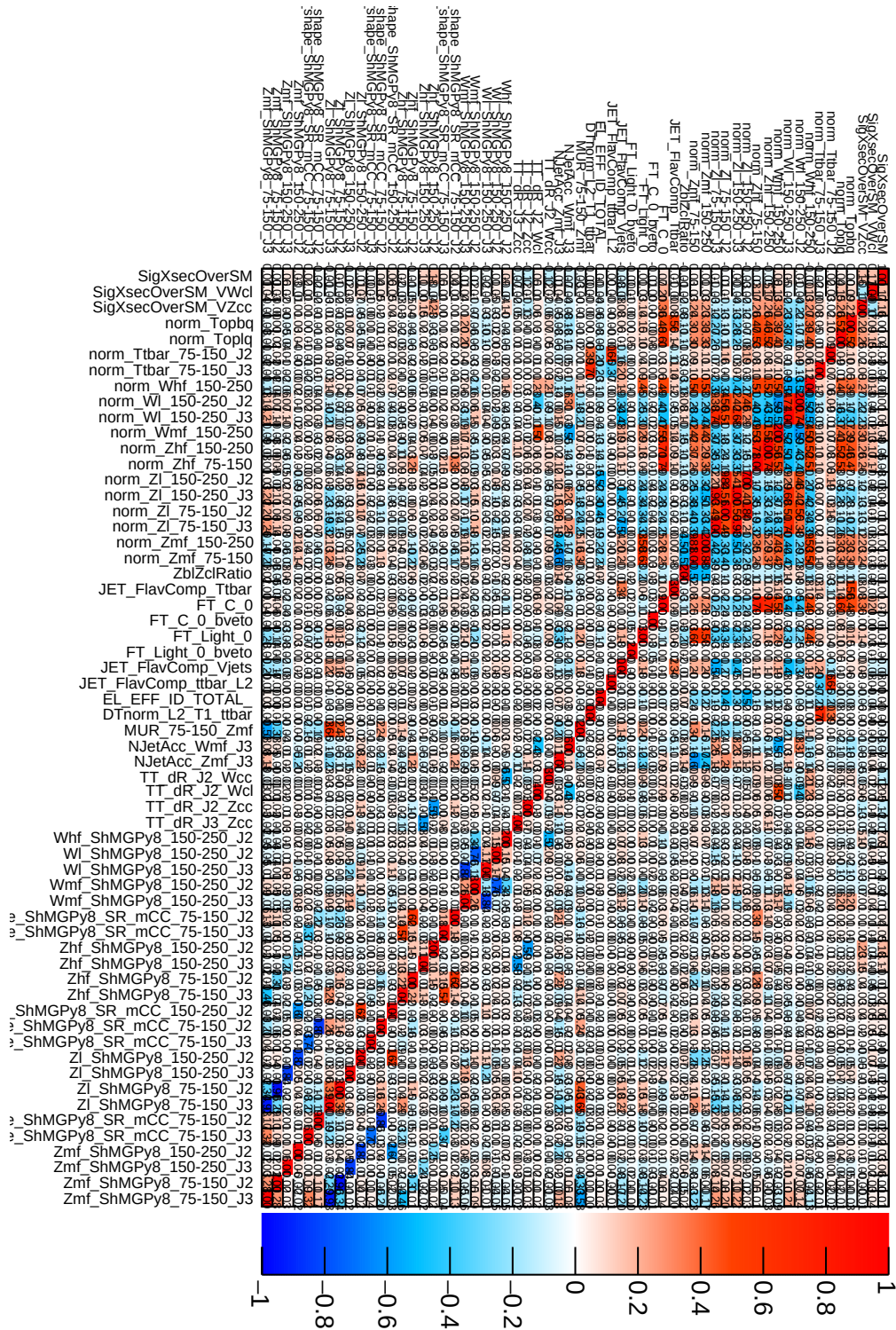


Figure 12.15: Correlation matrix from the unconditional combined fit to Data. Only nuisance parameters that have at least one correlation of magnitude 0.50 or higher with another nuisance parameter are shown.

Figure 12.16(a) the $VH(cc)$ signal strength is shown for the baseline combined fit, as well as in a five POI fit setup where the $VH(cc)$ signal strength is de-correlated between the lepton channels (i.e. each lepton channel has a separate $VH(cc)$ POI). Figure 12.16(b) shows the measured signal strength of the two signal diboson processes.

Parameter of Interest	Measured value
$\mu_{VH(cc)}$	-8.58 ± 15.10
$\mu_{VZ(cc)}$	1.16 ± 0.48
$\mu_{VW(cl)}$	0.83 ± 0.24

Table 12.2: Measured signal strength in the unconditional fit of the $VH(cc)$, $VZ(cc)$, and $VW(cl)$ signal processes.

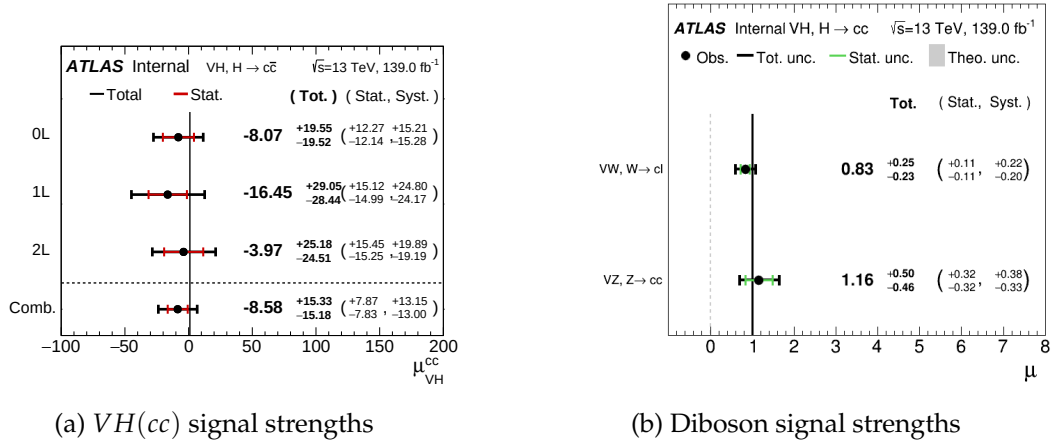


Figure 12.16: Fitted values for the (a) $VH(cc)$ signal strength and (b) $VW(cl)$ and $VZ(cc)$ signal strengths in the fit to the full dataset, as well as de-correlated between channels for the $VH(cc)$ signal strengths.

No excess over the SM background-only prediction is observed for the $VH(cc)$ process, and an observed (expected) 95% CL upper limit of $26 \left(31^{+12}_{-8} \right)$ is set. Table 12.3 shows the limits in the combined fit as well as in the individual channels fit. The limit of each individual channel corresponds to a fit performed using only data of that specific leptonic channel. One can see from the table that the 0-lepton channel yields the best standalone limit on the $VH(cc)$ signal process, with comparable sensitivity to that obtained in the combined fit. This is a result of a negative pull in the 0-lepton fitted signal strength due to the more complex background composition in this channel that can hardly be fitted without the 1- and 2-lepton channels to control the W +jets and Z +jets backgrounds.

Figure 12.17 shows the limit in the combined fit and in the case where the $VH(cc)$

95% CL upper limit on the signal strength		
Channel	Expected Limit	Observed Limit
0-lepton	$43.4^{+17.0}_{-12.1}$	26.4
1-lepton	$61.9^{+24.3}_{-17.3}$	53.5
2-lepton	$54.4^{+21.3}_{-15.2}$	56.9
0+1+2-lepton	$31.1^{+12.1}_{-8.7}$	26.0

Table 12.3: The observed and expected 95% CL upper limit on the $VH(cc)$ signal strength for the three analysis channels and in the combined fit.

signal strength is de-correlated between the lepton channels.

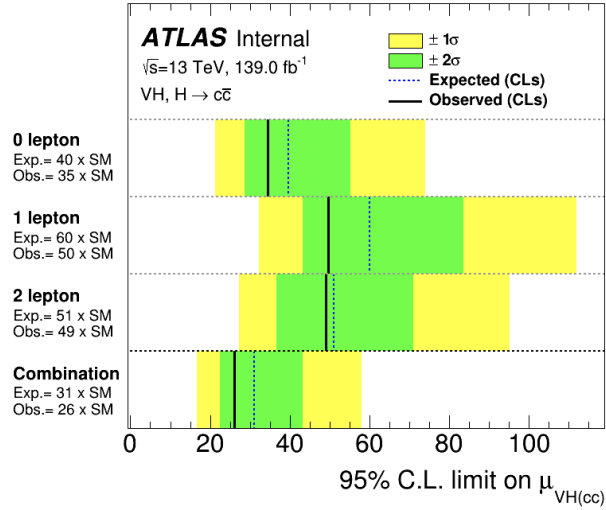


Figure 12.17: Observed and expected limits for the $VHcc$ signal at 95% CL, the result in the channels correspond to the results obtained from a combined fit with the channels POI decorrelated.

The so-called *background-subtracted* plot is presented in Figure 12.18, where the invariant mass spectrum of the data is shown after the background events are removed. The remaining contributions in the plot are the two diboson signal processes in gray, and the $VH(cc)$ signal, scaled by their observed signal strengths. The plot corresponds to a combination of all the analysis signal regions, i.e. 0+1+2-lepton, 1+2 c -tags, 2+3 jets (> 2 jets for the 2-lepton channel), $p_T^V > 150$ GeV ($p_T^V > 75$ GeV for the 2-lepton channel) categories.

The observed (expected) significance of the $VZ(cc)$ and $VW(cl)$ processes are 2.6(2.2) and 3.8(4.6) standard deviations respectively. These results, which correspond to the

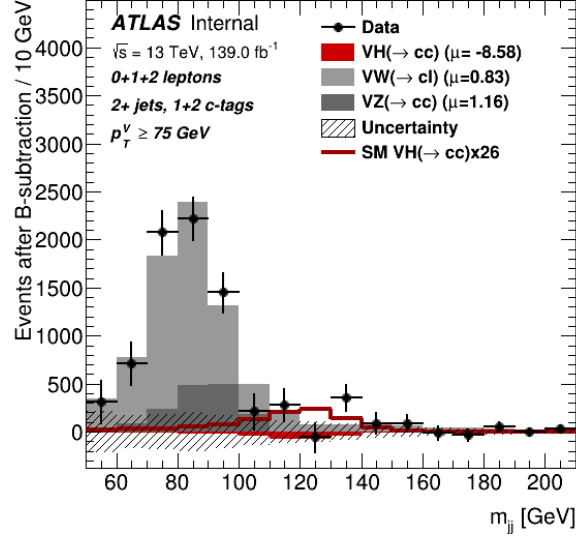


Figure 12.18: Background subtracted mass spectrum from the unconditional 3 POI fit to data in the combined analysis signal region. The $VZ(cc)$, $VW(cl)$, and $VH(cc)$ processes are scaled to their observed signal strength from the fit.

first measurements achieved using c -tagging, are in good agreement with the SM predicted values, and serve as validation of the analysis procedure.

12.2.1 | Nuisance parameter ranking

In order to quantify the impact that a NP θ_i has on the fitted signal-strength parameter, a *likelihood scan* is performed by fixing the value to the scanning point and refitting all other NPs $\theta_{j \neq i}$. The scan is continued until the (negative) log-likelihood difference with respect to the nominal fit, $\Delta \ln(\mathcal{L})$, reaches $\frac{1}{2}$. This point marks the 1σ impact of this NP on the signal strength and the scan is repeated starting from the nominal position into the other direction to get -1σ impact. The result of this scan for the NPs with the highest impact is summarized in decreasing order.

Figure 12.19 shows the NP ranking for the $VH(cc)$ signal strength derived from the data. The highest-ranked NP is the residual SR shape for the $Z+hf$ component in the high- p_T^V 2-jet signal region, and in general, the NPs associated with the extrapolation from the high ΔR and top control region are ranked highly. This is expected as they provide the freedom to move the distributions in the signal regions and are not constrained by control regions. Only one floating normalization shows up in the ranking, $W+mf$. The $VW(cl)$ and $VZ(cc)$ signal strengths also have large impacts, due to their

correlation with the $VH(cc)$ signal.

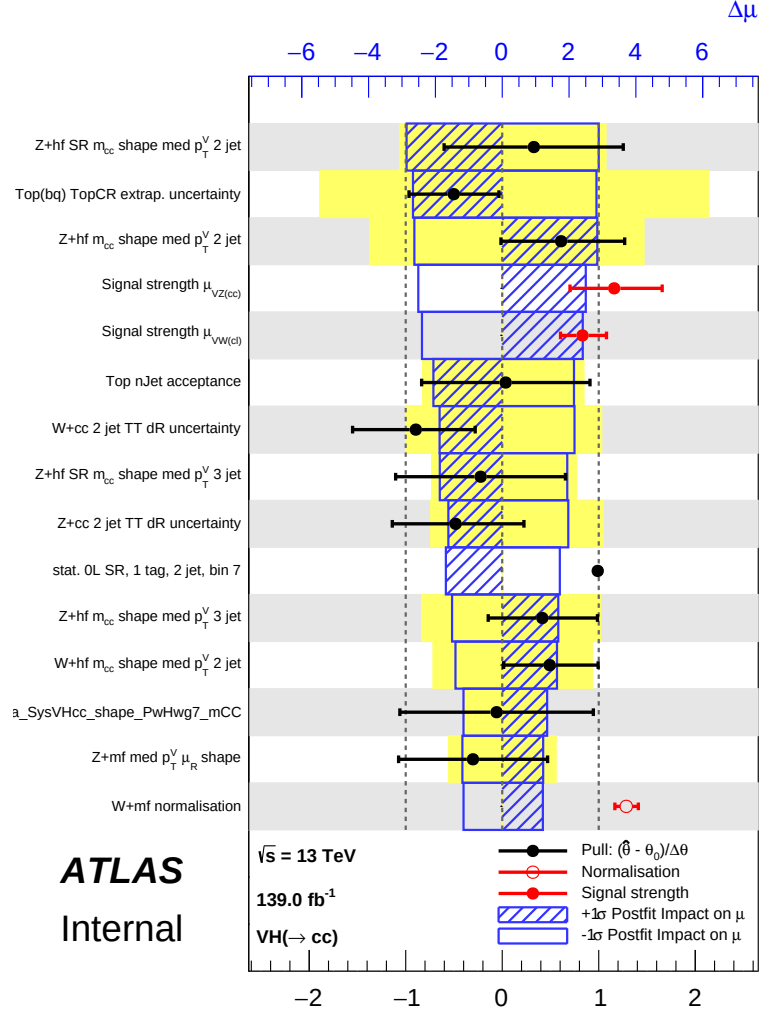


Figure 12.19: Impact of systematic uncertainties for the 10 largest contributions on the fitted $VH(cc)$ signal-strength parameter $\hat{\mu}$ sorted in decreasing order. The boxes show the variations of $\hat{\mu}$, referring to the top x -axis, when fixing the corresponding individual nuisance parameter to its post-fit value modified upwards or downwards by its post-fit uncertainty, i.e. $\hat{\theta} \pm \sigma_{\hat{\theta}}$, and repeating the fit. The impact of up- and down-variations can be distinguished via the dashed and plane box fillings. The filled circles show the deviation of the fitted value for each nuisance parameter, $\hat{\theta}$, from their nominal input value θ_0 expressed in standard deviations with respect to their nominal uncertainties $\Delta\theta$ (bottom x -axis). The error bars show the post-fit uncertainties on $\hat{\theta}$ with respect to their nominal uncertainties.

12.2.1.1 | $VH(cc)$ Breakdown of Uncertainties

Table 12.4 shows the breakdown of systematic uncertainties into categories of similar origin and their respective impact on the fitted $VH(cc)$ signal strength $\hat{\mu}$, derived on data. The impact of a group is obtained by running the fit with all NPs belonging to it fixed and comparing the uncertainty on the signal strength (obtained from similar likelihood scans as described in Section 12.2.1) $\sigma_{\hat{\mu}'}$ with the uncertainty from the nominal fit $\sigma_{\hat{\mu}}$ as:

$$\text{impact} = \sqrt{\sigma_{\hat{\mu}}^2 - \sigma_{\hat{\mu}'}^2}. \quad (12.1)$$

The total statistical impact is defined as the uncertainty of the fit with all NPs fixed to their nominal expectation value except for the floating normalizations, and the total systematic impact is calculated as the quadratic difference of the statistical impact to the nominal uncertainty. The sum in quadrature of the individual contributions differs from the total uncertainty due to correlations between the nuisance parameters.

For the $VH(cc)$ signal strength the statistical uncertainty is of the same order of magnitude as the systematic uncertainty, with the systematic uncertainty being slightly larger.

12.3 | Limits on the c -quark Yukawa coupling y_c

The results of the fit can be interpreted within the κ framework (introduced in Section 2.2.3). In the context of the $VH(cc)$ analysis the Higgs-charm and Higgs-vector boson coupling modifiers, κ_c and κ_V , appear explicitly as represented in Figure 12.20.

The $VH(cc)$ signal strength can then be re-parametrized in the likelihood as:

$$\mu(\kappa_i) = \frac{\text{BR}_{H \rightarrow c\bar{c}}(\kappa_i) \sigma_{pp \rightarrow VH}(\kappa_i)}{\text{BR}_{H \rightarrow c\bar{c}}^{\text{SM}}(\kappa_i) \sigma_{pp \rightarrow VH}^{\text{SM}}(\kappa_i)} = \frac{\kappa_V^2 \kappa_c^2}{\kappa_H^2}, \quad (12.2)$$

where $\text{BR}_{H \rightarrow c\bar{c}}$ is the branching ratio of the Higgs boson decay to a pair of charm and anti-charm quarks, $\sigma_{pp \rightarrow VH}$ is the cross section of the VH production mode in pp collisions, and κ_H^2 represents the modification to the total Higgs width from the contribution of the various decay channels, which can be written as:

Source of uncertainty		$\mu_{VH(cc)}$	$\mu_{VW(cq)}$	$\mu_{VZ(cc)}$
Total		15.3	0.24	0.48
Statistical		10.0	0.11	0.32
Systematics		11.5	0.21	0.36
Statistical uncertainties				
Data statistics only		7.8	0.05	0.23
Floating normalisations		5.1	0.09	0.22
Theoretical and modelling uncertainties				
$VH(\rightarrow c\bar{c})$		2.1	< 0.01	0.01
Z +jets		7.0	0.05	0.17
Top-quark		3.9	0.13	0.09
W +jets		3.0	0.05	0.11
Diboson		1.0	0.09	0.12
$VH(\rightarrow b\bar{b})$		0.8	< 0.01	0.01
Multi-Jet		1.0	0.03	0.02
Simulation statistics		4.2	0.09	0.13
Experimental uncertainties				
Jets		2.8	0.06	0.13
Leptons		0.5	0.01	0.01
$E_{\text{T}}^{\text{miss}}$		0.2	0.01	0.01
Pileup and luminosity		0.3	0.01	0.01
Flavour tagging	c -jets	1.6	0.05	0.16
	b -jets	1.1	0.01	0.03
	light-jets	0.4	0.01	0.06
	τ -jets	0.3	0.01	0.04
Truth-flavour tagging	ΔR correction	3.3	0.03	0.10
	Residual non-closure	1.7	0.03	0.10

Table 12.4: Breakdown of contributions to the uncertainty in the fitted values of $\mu_{VH(cc)}$, $\mu_{VW(cl)}$ and $\mu_{VZ(cc)}$. The sum in quadrature of different sources of uncertainty may differ from the total due to correlations. In the case that the up and down systematic variations have different values, the mean of the absolute values is shown.

$$\kappa_H^2 = \frac{\sum_j \Gamma_j^{\text{SM}} \times \kappa_j^2}{\Gamma_H^{\text{SM}}}, \quad (12.3)$$

with Γ_{SM} the SM width for various processes of the Higgs decay.

If we consider the simplest model where we assume:

- All coupling modifiers other than κ_c are set to their SM value (i.e. $\kappa_i = 1, \forall i \neq c$)
- The coupling modifier only affects the Higgs boson decays (i.e. neglecting the contribution to the VH production through $gg \rightarrow ZH$, which can be manifested by replacing the top-quarks with charm-quarks in the loop of the right diagram in Figure 2.4)

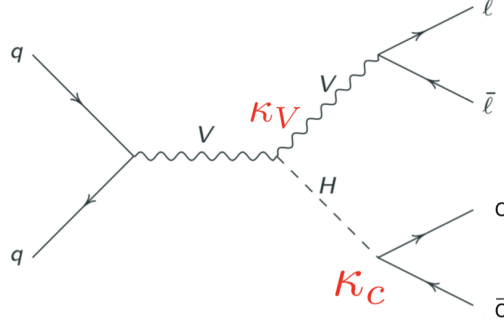


Figure 12.20: Feynman diagram of the $VH(cc)$ process for the 2-lepton channel, red represents the two vertices where the κ_V and κ_c coupling modifiers appear explicitly.

- Only SM decays are considered (i.e. neglecting BSM contributions that can modify the total Higgs width)

then the signal strength parametrization in Eq (12.2) becomes:

$$\mu(\kappa_i) = \frac{\kappa_c^2}{\text{BR}_{H \rightarrow c\bar{c}}\kappa_c^2 + (1 - \text{BR}_{H \rightarrow c\bar{c}})}. \quad (12.4)$$

Eq. (12.4) shows that at a small value of the signal strength, the constraint on κ_c scales as $\mu(\kappa_c) \approx \kappa_c^2$ while at very high values of the signal strength the constraint on κ_c reaches an asymptotic limit defined by $\mu(\kappa_c) \rightarrow \frac{1}{\text{BR}_{H \rightarrow c\bar{c}}}$. Therefore, any measured signal strength above this threshold will have no constraining power over κ_c . This effect is illustrated in Figure 12.21.

In order to extract a limit on the value of κ_c the $VH(cc)$ signal strength is re-parameterized directly in the fit according to equation Eq. (12.4) and a profile likelihood scan is performed. The results of the profile likelihood scan for a fit with 3 POIs (κ_c , $\mu_{VZ(cc)}$ and $\mu_{VW(cl)}$) are shown in Figure 12.22 for the individual channels and the combined fit to data. The 68% and 95% expected and observed CL limits are summarized in Table 12.5. An observed (expected) limit of $\kappa_c < 8.5$ (12.4) at 95% CL is obtained in the combined fit.

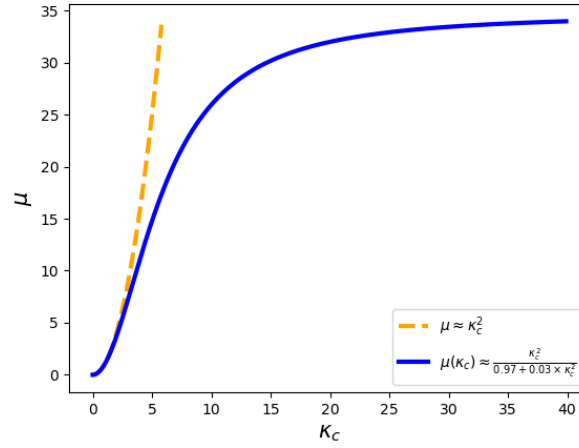


Figure 12.21: Parametrization of the signal strength as a function of the charm quark Yukawa coupling modifier $\mu(\kappa_c)$ in the context of a scenario where all other coupling modifiers are considered to unity, as in Eq (12.4).

Channel	Expected		Observed	
	68% CL	95% CL	68% CL	95% CL
0-lepton	[-6.91, 6.94]	-	[-2.72, 2.75]	-
1-lepton	[-14.00, 14.03]	-	[-6.55, 6.58]	-
2-lepton	[-9.25, 9.28]	-	[-12.53, 12.51]	-
0+1+2-lepton	[-4.88, 4.91]	[-12.32, 12.38]	[-3.5, 3.5]	[-8.5, 8.5]

Table 12.5: Expected and observed 68% CL and 95% CL limits on κ_c in the 3 POI fit setup. The per-channel limits are obtained using standalone fits where only the data of the corresponding channel is considered.

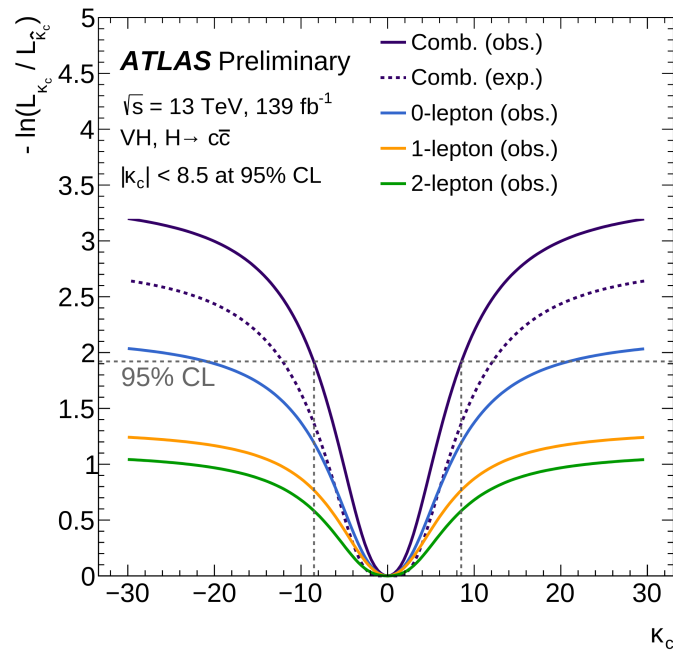


Figure 12.22: Expected and observed profile likelihood scans for κ_c . The per-channel likelihoods are obtained using a five-POI fit, in which each channel has a separate $VH(cc)$ parameter of interest.

Concluding Remarks

This thesis has presented a search for the $VH(V \rightarrow \text{lep}, H \rightarrow c\bar{c})$ process, in which a Higgs boson is generated in association with an electroweak vector boson, the vector boson subsequently decays into leptons and the Higgs into a pair of charm and anti-charm quarks. Measuring, or at the very least constraining the $H \rightarrow c\bar{c}$ decay rate is beneficial for our understanding of the origin of masses in the SM, as the decay of the Higgs to second generation fermions has yet to be established experimentally. In addition, modifications to the Yukawa charm coupling y_c are possible through a variety of BSM scenarios, thus it is an intriguing place to look for signs of new and exotic physics.

The data analyzed in this work were collected from pp collisions at center-of-mass energy $\sqrt{s} = 13$ TeV, and correspond to integrated luminosity of 139 fb^{-1} which were recorded by the ATLAS experiment during Run-2 of the LHC between 2015-2018.

The search uses three channels, $ZH \rightarrow \nu\bar{\nu}c\bar{c}$, $WH \rightarrow l\nu c\bar{c}$, and $ZH \rightarrow l\bar{l}c\bar{c}$, and relies on a multivariate c -tagging algorithm to identify signal events containing charm quarks. Events passing the analysis selection criteria are further split into 16 signal regions, categorized by the number of identified c -tagged jets, the number of jets, and the p_T of the reconstructed vector boson.

The dominant background processes are the production of W/Z boson in association with jets (V +jets), and $t\bar{t}$. The contribution of these background sources is modeled with the help of dedicated control regions. These include 16 high ΔR control regions enriched with V +jets, 6 top control regions, and 6 0 c -tag control region with high purity of V +light flavored jets.

The invariant mass distribution of the Higgs candidate, $m_{c\bar{c}}$, is used as a signal-

vs-background discriminant in the final fit. The analysis strategy is validated with the study of diboson production, which is found to be consistent with the SM prediction, with observed (expected) significances of $2.6\sigma(2.2\sigma)$ for the $VZ(cc)$ process and $3.8\sigma(4.6\sigma)$ for the $VW(cl)$ process.

No excess of $VH(cc)$ events over the expected background was found, and an observed (expected) limit of $26 \left(31^{+12}_{-8}\right)$ times the predicted SM cross-section times branching-ratio was set. This result, which corresponds to the most stringent limit on the process to date, is further interpreted within the Higgs coupling modifiers (κ) framework. In the considered model, which takes into account only effects on the Higgs boson width, the analysis is able to set a first direct constraint on the charm Yukawa coupling modifier and determine an observed (expected) limit of $|\kappa_c| < 8.5(12.4)$ at 95% CL.

The two main challenges in this search are the small rate of the sought signal process and the low efficiency of the c -tagging algorithm. The former is dictated by nature and there is nothing we physicists can do to increase the abundance of this process. As for the latter, there is obviously room for improving the c -tagging algorithm performance, yet this is not a simple task. The properties of c -hadron decays, which are somewhat "in-between" those of b - and light-hadrons, impose an inherent difficulty as one must distinguish c -jets from the heavier b -jets on one side, and from light-jets on the other side. Nevertheless, many efforts have been dedicated to maximizing the search sensitivity, both in terms of optimization of the c -tagging working point, and in terms of the analysis strategy reflected in the event selection and categorization, and in the background modeling and fit model.

13.1 | Future prospects

The results of the search detailed in this work had been published and presented at LHCP 2021 conference [14]. Since then the analysis team has been focusing on a statistical combination of the results with those of the $VH(bb)$ analysis, to report a two-dimensional limit on the Higgs coupling modifiers of the b - and c -quarks κ_b and κ_c . This combination is made relatively straightforward by virtue of the orthogonality between the signal regions of the two analyses. In fact, this combination was in mind when we decided to implement a b -tag veto in the c -tagging discriminant and designed the signal regions to be orthogonal. At the time of writing these lines, this work is at final stages of approval within the collaboration.

In parallel, work has begun within the Higgs group on the so-called " $VH(bb/cc)$ " legacy analysis, which targets simultaneously the $VH(bb)$ process, $VH(cc)$ process, and $VH(bb)$ in the boosted regime. For the latter, one imposes that the vector boson is generated with very large transverse momentum, and requires the Higgs to be reconstructed from a single large- R jet, containing two small- R jets inside. This search will be performed using the full Run-2 dataset, and is planned to include a number of improvements/modifications to the jet reconstruction algorithms, flavor-tagging scheme, and the general software framework.

Very preliminary studies have been made also on the prospects of $H \rightarrow c\bar{c}$ in Run-3 of the LHC and in the High-Luminosity LHC. First, if we naively assume similar c -tagging efficiency and acceptance as in the current analysis, the signal strength scales with luminosity as:

$$\mu \rightarrow \mu \times \sqrt{\frac{139 \text{ fb}^{-1}}{\mathcal{L}}}. \quad (13.1)$$

This translates to an expected limit of $\sim 17 \times \text{SM}$ by the end of Run-3 with 300 fb^{-1} and $\sim 5.5 \times \text{SM}$ with the foreseen 3000 fb^{-1} at the HL-LHC. Note that as the statistical significance of a search is given by

$$z = \frac{S}{\sqrt{B}} \propto \frac{\epsilon_{\text{signal}}^2}{\epsilon_{\text{bkg}}}, \quad (13.2)$$

the most promising way to increase the sensitivity is to improve the c -tagging efficiency of c -jets (and the rejection of b - and light-jets). From a simple toy-study based on MC, where we replace the c -tagging requirement by re-weighting events to the best-case scenario (i.e. 100% efficiency for c -jets and 0% for b - and light-jets) one obtains a factor of $\sim \times 4$ improvement to the expected limit. When the c -tagging performance is assumed to be at its optimum, the limiting aspect to the sensitivity becomes the background from $Z \rightarrow c\bar{c}$.

However, one must also consider that flavor-tagging relies heavily on tracking, particularly in the vicinity of the beamline, and in the conditions expected in the HL-LHC, with very large intensity and particle activity near the beamline, increasing the c -tagging efficiency will not be a straight forward task. Additional improvement to the analysis can be achieved by employing a Multi-Variate-Analysis (MVA) rather than the current cut-based search using $m_{c\bar{c}}$ as the final discriminant. Very preliminary internal studies suggest this can lead to a $\sim 10 - 15\%$ improvement to the analysis sensitivity.

The Standard Model of particle physics is our best description of the physics of elementary particles. It defines the building blocks of our universe and the ways in which they interact. The predictions of the Standard Model have been tested through many experiments, and it has so far been able to successfully explain all observed phenomena in collider physics. Yet we also know that the current picture of the SM is not complete, as it fails to account for a number of physical phenomena observed in non-collider experiments. Precision measurements of the Higgs boson couplings and decay can provide an effective way for further probing the SM, and for searching for deviations originating from BSM physics. It will be fascinating to see in the coming years how much the LHC-era will improve our understanding of nature.

Acceptance ratios tables

This appendix contains tables with the values extracted and used as priors for the acceptance ratio uncertainty nuisance parameters described in Section 11.3. The acceptance ratios between the 3(+)-jet and 2-jet categories are shown in Table A.1, flavor composition acceptance ratios in Table A.2, acceptance ratios between high- and low- p_T^V regions in Table A.3, CR to SR acceptance ratios in Table A.4 and acceptance ratios between the different leptonic channels in Table A.5.

3(+)-jet/2-jet acceptance ratios		
Process	Derived in	Value
$qq \rightarrow WH$	1-lepton, SR	$(11 \pm 1)\%$
$qq \rightarrow ZH$	0-lepton, SR	$(8 \pm 1)\%$
	2-lepton, high- p_T^V , SR	$(12 \pm 1)\%$
$gg \rightarrow ZH$	0-lepton, SR	$(19 \pm 2)\%$
	2-lepton, high- p_T^V , SR	$(56 \pm 3)\%$
Z+hf	0-lepton, SR	$(10 \pm 2)\%$
	2-lepton, high- p_T^V , SR	$(15 \pm 2)\%$
	2-lepton, low- p_T^V , SR	$(25 \pm 2)\%$
	0-lepton, high ΔR CR	$(16 \pm 2)\%$
	2-lepton, high- p_T^V , high ΔR CR	$(28 \pm 2)\%$
	2-lepton, high- p_T^V , high ΔR CR	$(37 \pm 3)\%$
Z+mf	0-lepton, SR	$(10 \pm 1)\%$
	2-lepton, high- p_T^V , SR	$(15 \pm 1)\%$
	2-lepton, low- p_T^V , SR	$(21 \pm 2)\%$
	0-lepton, high ΔR CR	$(14 \pm 1)\%$
	2-lepton, high- p_T^V , high ΔR CR	$(20 \pm 1)\%$
	2-lepton, high- p_T^V , high ΔR CR	$(25 \pm 2)\%$
W+hf	1-lepton, SR	$(8 \pm 2)\%$
	1-lepton, high ΔR CR	$(11 \pm 2)\%$
W+mf	1-lepton, SR	$(10 \pm 1)\%$
	1-lepton, high ΔR CR	$(14 \pm 1)\%$
ZZ	2-lepton, high- p_T^V , SR	$(11 \pm 3)\%$
WW	1-lepton, SR	$(7 \pm 2)\%$
WZ	0-lepton, SR	$(9 \pm 1)\%$
$t\bar{t}$ (bq)	1-lepton, SR	$(7 \pm 1)\%$
$t\bar{t}$ (lq)	1-lepton, SR	$(7 \pm 1)\%$
single-top Wt (bq)	1-lepton, SR	$(9 \pm 1)\%$
single-top Wt (lq)	1-lepton, SR	$(9 \pm 1)\%$

Table A.1: Acceptance ratio uncertainties between the 3(+)-jet and 2(+)-jet categories. The priors are initially derived in the channel with highest purity for each process. In cases where the values derived in different channels are not in agreement, different priors are used as indicated in the table. The quoted errors represent the statistical uncertainty associated with the calculation of each value.

Flavor composition acceptance ratios		
Processes	Derived in	Value
$Z+bb/Z+cc$	0-lepton, 2- & 3-jet , SR & high ΔR CR	$(20 \pm 2)\%$
$Z+bl/Z+cl$	0-lepton, 2- & 3-jet , SR & high ΔR CR	$(18 \pm 2)\%$
$Z+bc/Z+cl$	0-lepton, 2- & 3-jet , SR & high ΔR CR	$(6 \pm 1)\%$
$W+bb/W+cc$	1-lepton, 2- & 3-jet , SR	$(4 \pm 1)\%$
	1-lepton, 2- & 3-jet , high ΔR CR	$(10 \pm 2)\%$
$W+bl/W+cl$	1-lepton, 2- & 3-jet , SR & high ΔR CR	$(31 \pm 1)\%$
$W+bc/W+cl$	1-lepton, 2- & 3-jet , SR & high ΔR CR	$(33 \pm 1)\%$
$W\tau_h+c/W+cl$	0-lepton, 2- & 3-jet , SR & high ΔR CR	$(11 \pm 5)\%$
$W\tau_h+b/W+cl$	0-lepton, 2- & 3-jet , SR & high ΔR CR	$(27 \pm 5)\%$
$W\tau_h+l/W+l$	0-lepton, 2- & 3-jet , SR & high ΔR CR	$(8 \pm 2)\%$
$\text{single-top}(bq)/t\bar{t}(bq)$	1-lepton, Top CR	$(12 \pm 1)\%$

Table A.2: Acceptance ratio uncertainties between individual components in backgrounds which group together multiple processes/flavors. The priors are initially derived in the channel with highest purity for each process. In cases where the values derived in different channels are not in agreement, different priors are used as indicated in the table. The quoted errors represent the statistical uncertainty associated with the calculation of each value.

$\text{low-}p_T^V/\text{high-}p_T^V$ acceptance ratios		
Processes	Derived in	Value
$qq \rightarrow ZH$	2-lepton, 2- & 3+ -jet SR	$(2 \pm 1)\%$
$gg \rightarrow ZH$	2-lepton, 2- & 3+ -jet SR	$(5 \pm 1)\%$
ZZ	2-lepton, 2- & 3+ -jet SR	$(4 \pm 2)\%$
WZ	2-lepton, 2- & 3+ -jet SR	$(4 \pm 1)\%$

Table A.3: Acceptance ratio uncertainties between the low- and high- p_T^V regions. The quoted errors represent the statistical uncertainty associated with the calculation of each value.

CR/SR acceptance ratios			
Regions	Process	Derived in	Value
High ΔR CR/SR	0-lepton, 2- & 3-jet	$W\tau_h+c$	$(18 \pm 8)\%$
	0-lepton, 2- & 3-jet	$W\tau_h+l$	$(5 \pm 3)\%$
Top-CR/SR	Z+hf	0-lepton, 3-jet	$(10 \pm 7)\%$
	Z+mf	0-lepton, 3-jet	$(10 \pm 4)\%$
	Z+l	0-lepton, 3-jet	$(17 \pm 7)\%$
	W+hf	0-lepton, 3-jet	$(10 \pm 8)\%$
	W+mf	0-lepton, 3-jet	$(13 \pm 4)\%$
	W+l	0-lepton, 3-jet	$(5 \pm 4)\%$
	$t\bar{t}(bq)$	1-lepton, 3-jet	$(9 \pm 1)\%$
	$t\bar{t}(lq)$	1-lepton, 3-jet	$(9 \pm 1)\%$
	single-top $Wt(bq)$	1-lepton, 3-jet	$(16 \pm 2)\%$
	single-top $Wt(lq)$	1-lepton, 3-jet	$(16 \pm 2)\%$

Table A.4: Acceptance ratio uncertainties between control and signal regions. The priors are initially derived in the channel with highest purity for each process. In cases where the values derived in different channels are not in agreement, different priors are used as indicated in the table. The quoted errors represent the statistical uncertainty associated with the calculation of each value.

Lepton channels acceptance ratios			
Process	Channels	Derived in	Value
Z+hf	0-lepton/2-lepton	2- & 3(+)-jet , high- p_T^V , SR	$(4 \pm 2)\%$
		2- & 3(+)-jet , high- p_T^V , high ΔR CR	$(8 \pm 2)\%$
Z+mf	0-lepton/2-lepton	2- & 3(+)-jet , high- p_T^V , SR	$(5 \pm 1)\%$
		2- & 3(+)-jet , high- p_T^V , high ΔR CR	$(6 \pm 1)\%$
Z+l	0-lepton/2-lepton	2- & 3(+)-jet , high- p_T^V , SR	$(5 \pm 1)\%$
		2- & 3(+)-jet , high- p_T^V , high ΔR CR	$(2 \pm 1)\%$
W+hf	0-lepton/1-lepton	2- & 3-jet , SR	$(1 \pm 2)\%$
		2- & 3-jet , high ΔR CR	$(6 \pm 4)\%$
W+mf	0-lepton/1-lepton	2- & 3-jet , SR	$(3 \pm 1)\%$
		2- & 3-jet , high ΔR CR	$(3 \pm 1)\%$
W+l	0-lepton/1-lepton	2- & 3-jet , SR	$(6 \pm 1)\%$
		2- & 3-jet , high ΔR CR	$(3 \pm 1)\%$
Top(bq)	0-lepton/1-lepton	2- & 3-jet , SR & Top CR	$(4 \pm 1)\%$
Top(lq)	0-lepton/1-lepton	2- & 3-jet , SR & Top CR	$(4 \pm 1)\%$
ZZ	1-lepton/0-lepton	2- & 3-jet , SR	$(4 \pm 2)\%$
	2-lepton/0-lepton	2- & 3(+)-jet , SR	$(2 \pm 1)\%$
WW	0-lepton/1-lepton	2- & 3-jet , SR	$(3 \pm 1)\%$
WZ	1-lepton/0-lepton	2- & 3-jet , SR	$(2 \pm 1)\%$
	2-lepton/0-lepton	2- & 3(+)-jet , SR	$(4 \pm 1)\%$

Table A.5: Acceptance ratio uncertainties between lepton channels. The quoted errors represent the statistical uncertainty associated with the calculation of each value.

Bibliography

- [1] Lyndon Evans and Philip Bryant. “LHC Machine”. In: *Journal of Instrumentation* 3 (Aug. 2008), S08001–S08001. DOI: [10.1088/1748-0221/3/08/s08001](https://doi.org/10.1088/1748-0221/3/08/s08001). URL: <https://doi.org/10.1088/1748-0221/3/08/s08001>.
- [2] ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *Journal of Instrumentation* 3 (Aug. 2008), S08003–S08003. DOI: [10.1088/1748-0221/3/08/s08003](https://doi.org/10.1088/1748-0221/3/08/s08003). URL: <https://doi.org/10.1088/1748-0221/3/08/s08003>.
- [3] CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *Journal of Instrumentation* 3 (Aug. 2008), S08004–S08004. DOI: [10.1088/1748-0221/3/08/s08004](https://doi.org/10.1088/1748-0221/3/08/s08004). URL: <https://doi.org/10.1088/1748-0221/3/08/s08004>.
- [4] ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Physics Letters B* 716.1 (Sept. 2012), pp. 1–29. DOI: [10.1016/j.physletb.2012.08.020](http://dx.doi.org/10.1016/j.physletb.2012.08.020). URL: <http://dx.doi.org/10.1016/j.physletb.2012.08.020>.
- [5] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Physics Letters B* 716.1 (Sept. 2012), pp. 30–61. DOI: [10.1016/j.physletb.2012.08.021](http://dx.doi.org/10.1016/j.physletb.2012.08.021). URL: <http://dx.doi.org/10.1016/j.physletb.2012.08.021>.
- [6] ATLAS Collaboration. “Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8

- TeV with the ATLAS detector". In: *Phys. Rev. D* 91 (Jan. 2015), p. 012006. DOI: [10.1103/PhysRevD.91.012006](https://doi.org/10.1103/PhysRevD.91.012006). URL: <https://link.aps.org/doi/10.1103/PhysRevD.91.012006>.
- [7] ATLAS Collaboration. "Observation and measurement of Higgs boson decays to WW^* with the ATLAS detector". In: *Phys. Rev. D* 92 (July 2015), p. 012006. DOI: [10.1103/PhysRevD.92.012006](https://doi.org/10.1103/PhysRevD.92.012006). URL: <https://link.aps.org/doi/10.1103/PhysRevD.92.012006>.
- [8] ATLAS Collaboration. "Cross-section measurements of the Higgs boson decaying into a pair of τ -leptons in proton-proton collisions at $\sqrt{s}=13$ TeV with the ATLAS detector". In: *Phys. Rev. D* 99.7 (Apr. 2019). DOI: [10.1103/physrevd.99.072001](https://doi.org/10.1103/physrevd.99.072001). URL: <http://dx.doi.org/10.1103/PhysRevD.99.072001>.
- [9] ATLAS Collaboration. "Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector". In: *Physics Letters B* 786 (Nov. 2018), pp. 59–86. DOI: [10.1016/j.physletb.2018.09.013](https://doi.org/10.1016/j.physletb.2018.09.013). URL: <http://dx.doi.org/10.1016/j.physletb.2018.09.013>.
- [10] ATLAS Collaboration. "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector". In: *Physics Letters B* 784 (Sept. 2018), pp. 173–191. DOI: [10.1016/j.physletb.2018.07.035](https://doi.org/10.1016/j.physletb.2018.07.035). URL: <http://dx.doi.org/10.1016/j.physletb.2018.07.035>.
- [11] ATLAS Collaboration. "A search for the dimuon decay of the Standard Model Higgs boson with the ATLAS detector". In: *Physics Letters B* 812 (Jan. 2021), p. 135980. DOI: [10.1016/j.physletb.2020.135980](https://doi.org/10.1016/j.physletb.2020.135980). URL: <http://dx.doi.org/10.1016/j.physletb.2020.135980>.
- [12] CMS Collaboration. "Evidence for Higgs boson decay to a pair of muons". In: *Journal of High Energy Physics* 2021.1 (Jan. 2021). DOI: [10.1007/jhep01\(2021\)148](https://doi.org/10.1007/jhep01(2021)148). URL: [http://dx.doi.org/10.1007/JHEP01\(2021\)148](http://dx.doi.org/10.1007/JHEP01(2021)148).
- [13] ATLAS Collaboration. "Search for the Decay of the Higgs Boson to Charm Quarks with the ATLAS Experiment". In: *Phys. Rev. Lett.* 120 (May 2018), p. 211802. DOI: [10.1103/PhysRevLett.120.211802](https://doi.org/10.1103/PhysRevLett.120.211802). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.120.211802>.

- [14] ATLAS Collaboration. *Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays to charm quarks with the ATLAS detector*. Tech. rep. Geneva: CERN, June 2021. URL: <https://cds.cern.ch/record/2771724>.
- [15] ATLAS Collaboration. *Search for $H \rightarrow c\bar{c}$ decays in associated VH production with the ATLAS experiment: Internal Supporting Document*. Tech. rep. Geneva: CERN, Jan. 2020. URL: <https://cds.cern.ch/record/2706457>.
- [16] ATLAS Collaboration. *New Small Wheel Technical Design Report*. Tech. rep. ATLAS New Small Wheel Technical Design Report. June 2013. URL: <https://cds.cern.ch/record/1552862>.
- [17] Michael E Peskin and Daniel V Schroeder. *An introduction to quantum field theory*. Includes exercises. Boulder, CO: Westview, 1995. URL: <https://cds.cern.ch/record/257493>.
- [18] Alberto Rojo and Anthony Bloch. *The principle of least action: history and physics*. Cambridge: Cambridge University Press, 2018. DOI: [10.1017/9781139021029](https://doi.org/10.1017/9781139021029). URL: <https://cds.cern.ch/record/2312316>.
- [19] Emmy Noether. “Invariant variation problems”. In: *Transport Theory and Statistical Physics* 1 (Jan. 1971), pp. 186–207. DOI: [10.1080/00411457108231446](https://doi.org/10.1080/00411457108231446). eprint: <https://doi.org/10.1080/00411457108231446>. URL: <https://doi.org/10.1080/00411457108231446>.
- [20] David J Griffiths. *Introduction to elementary particles; 2nd rev. version*. Physics textbook. New York, NY: Wiley, 2008. URL: <https://cds.cern.ch/record/111880>.
- [21] Sheldon L Glashow. “Partial-symmetries of weak interactions”. In: *Nuclear Physics* 22 (Feb. 1961), pp. 579–588. DOI: [https://doi.org/10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2). URL: <http://www.sciencedirect.com/science/article/pii/0029558261904692>.
- [22] Steven Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (Nov. 1967), pp. 1264–1266. DOI: [10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264>.

- [23] Abdus Salam. “Weak and Electromagnetic Interactions”. In: *Conf. Proc. C 680519* (May 1968), pp. 367–377. DOI: [10.1142/9789812795915_0034](https://doi.org/10.1142/9789812795915_0034).
- [24] Peter W Higgs. “Broken symmetries, massless particles and gauge fields”. In: *Physics Letters* 12 (Sept. 1964), pp. 132–133. DOI: [https://doi.org/10.1016/0031-9163\(64\)91136-9](https://doi.org/10.1016/0031-9163(64)91136-9). URL: <http://www.sciencedirect.com/science/article/pii/0031916364911369>.
- [25] François Englert and Robert Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (Aug. 1964), pp. 321–323. DOI: [10.1103/PhysRevLett.13.321](https://link.aps.org/doi/10.1103/PhysRevLett.13.321). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321>.
- [26] LHCb Collaboration. “Observation of the resonant character of the $Z(4430)^-$ state”. In: *Phys. Rev. Lett.* 112 (June 2014), p. 222002. DOI: [10.1103/PhysRevLett.112.222002](https://doi.org/10.1103/PhysRevLett.112.222002). arXiv: [1404.1903](https://arxiv.org/abs/1404.1903) [hep-ex].
- [27] LHCb Collaboration. “Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \rightarrow J/\psi K^- p$ Decays”. In: *Phys. Rev. Lett.* 115 (Aug. 2015), p. 072001. DOI: [10.1103/PhysRevLett.115.072001](https://doi.org/10.1103/PhysRevLett.115.072001). arXiv: [1507.03414](https://arxiv.org/abs/1507.03414) [hep-ex].
- [28] Particle Data Group. “Review of Particle Physics”. In: *Progress of Theoretical and Experimental Physics* 2020 (Aug. 2020). 083C01. DOI: [10.1093/ptep/ptaa104](https://doi.org/10.1093/ptep/ptaa104). eprint: <https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf>. URL: <https://doi.org/10.1093/ptep/ptaa104>.
- [29] Jeffrey Goldstone, Abdus Salam, and Steven Weinberg. “Broken Symmetries”. In: *Phys. Rev.* 127 (Aug. 1962), pp. 965–970. DOI: [10.1103/PhysRev.127.965](https://doi.org/10.1103/PhysRev.127.965). URL: <https://link.aps.org/doi/10.1103/PhysRev.127.965>.
- [30] ATLAS Collaboration. *Standard Model Summary Plots March 2021*. Tech. rep. Geneva: CERN, Mar. 2021. URL: <https://cds.cern.ch/record/2758261>.
- [31] Dieter R Brill and Robert H Gowdy. “Quantization of general relativity”. In: *Reports on Progress in Physics* 33 (May 1970), pp. 413–488. DOI: [10.1088/0034-4885/33/2/301](https://doi.org/10.1088/0034-4885/33/2/301). URL: <https://doi.org/10.1088/0034-4885/33/2/301>.

- [32] Planck Collaboration. “Planck 2018 results”. In: *Astronomy and Astrophysics* 641.A6 (Sept. 2020). DOI: [10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910). URL: <http://dx.doi.org/10.1051/0004-6361/201833910>.
- [33] Martijn Mulders and Claude Duhr, eds. *Proceedings, 2018 European School of High-Energy Physics (ESHEP 2018): Maratea, Italy, June 20 – July 3 2018*. Vol. 6/2019. CERN Yellow Reports: School Proceedings. Geneva: CERN, 2019. DOI: [10.23730/CYRSP-2019-006](https://doi.org/10.23730/CYRSP-2019-006).
- [34] Andrei D Sakharov. “Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe”. In: *Soviet Physics Uspekhi* 34 (May 1991), pp. 392–393. DOI: [10.1070/pu1991v034n05abeh002497](https://doi.org/10.1070/pu1991v034n05abeh002497). URL: <https://doi.org/10.1070/pu1991v034n05abeh002497>.
- [35] Super-Kamiokande Collaboration. “Evidence for Oscillation of Atmospheric Neutrinos”. In: *Phys. Rev. Lett.* 81 (Aug. 1998), pp. 1562–1567. DOI: [10.1103/PhysRevLett.81.1562](https://doi.org/10.1103/PhysRevLett.81.1562). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.81.1562>.
- [36] Carlo Giunti and Chung W. Kim. *Fundamentals of Neutrino Physics and Astrophysics*. 2007. ISBN: 978-0-19-850871-7.
- [37] H1 and ZEUS Collaborations. *Combination of Measurements of Inclusive Deep Inelastic $e^\pm p$ Scattering Cross Sections and QCD Analysis of HERA Data*. June 2015. DOI: [10.48550/ARXIV.1506.06042](https://arxiv.org/abs/1506.06042). URL: <https://arxiv.org/abs/1506.06042>.
- [38] Guido Altarelli and Giorgio Parisi. “Asymptotic freedom in parton language”. In: *Nuclear Physics B* 126 (Aug. 1977), pp. 298–318. DOI: [https://doi.org/10.1016/0550-3213\(77\)90384-4](https://doi.org/10.1016/0550-3213(77)90384-4). URL: <https://www.sciencedirect.com/science/article/pii/0550321377903844>.
- [39] Yuri L Dokshitzer. “Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics”. In: *Sov. Phys. JETP* 46 (Apr. 1977), pp. 641–653.
- [40] Vladimir N Gribov and Lev N Lipatov. “Deep inelastic $e p$ scattering in perturbation theory”. In: *Sov. J. Nucl. Phys.* 15 (Apr. 1972), pp. 438–450.

- [41] NNPDF Collaboration. “Parton distributions from high-precision collider data”. In: *The European Physical Journal C* 77.10 (Oct. 2017). DOI: [10.1140/epjc/s10052-017-5199-5](https://doi.org/10.1140/epjc/s10052-017-5199-5). URL: <http://dx.doi.org/10.1140/epjc/s10052-017-5199-5>.
- [42] Gregory Breit and Eugene Wigner. “Capture of Slow Neutrons”. In: *Phys. Rev.* 49 (Apr. 1936), pp. 519–531. DOI: [10.1103/PhysRev.49.519](https://doi.org/10.1103/PhysRev.49.519). URL: <https://link.aps.org/doi/10.1103/PhysRev.49.519>.
- [43] LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. 2017. arXiv: [1610.07922](https://arxiv.org/abs/1610.07922) [hep-ph].
- [44] LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*. Ed. by Sven Heinemeyer. CERN Yellow Reports: Monographs. Geneva: CERN, July 2013. DOI: [10.5170/CERN-2013-004](https://doi.org/10.5170/CERN-2013-004). URL: <http://cds.cern.ch/record/1559921>.
- [45] LHC Higgs Cross Section Working Group. *LHC HXSWG interim recommendations to explore the coupling structure of a Higgs-like particle*. 2012. arXiv: [1209.0040](https://arxiv.org/abs/1209.0040) [hep-ph].
- [46] ATLAS Collaboration. “Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13 \text{ TeV}$ collected with the ATLAS experiment”. In: *Phys. Rev. D* 101 (Sept. 2019), 012002. 48 p. DOI: [10.1103/PhysRevD.101.012002](https://doi.org/10.1103/PhysRevD.101.012002). arXiv: [1909.02845](https://arxiv.org/abs/1909.02845). URL: <https://cds.cern.ch/record/2688596>.
- [47] Stephen Myers. *The LEP Collider, from design to approval and commissioning*. John Adams’ memorial lecture. Delivered at CERN, 26 Nov 1990. Geneva: CERN, 1991. DOI: [10.5170/CERN-1991-008](https://doi.org/10.5170/CERN-1991-008). URL: <https://cds.cern.ch/record/226776>.
- [48] Esma Mobs. “The CERN accelerator complex - 2019. Complexe des accélérateurs du CERN - 2019”. In: (July 2019). General Photo. URL: <https://cds.cern.ch/record/2684277>.

- [49] Daniel Boussard and Trevor Paul R Linnecar. *The LHC Superconducting RF System*. Tech. rep. Geneva: CERN, Dec. 1999. URL: <https://cds.cern.ch/record/410377>.
- [50] Vittorio Parma and Lucio Rossi. *Performance of the LHC magnet system*. Tech. rep. Sept. 2009, 6 p. URL: <https://cds.cern.ch/record/1204578>.
- [51] ALICE Collaboration. “The ALICE experiment at the CERN LHC”. In: *Journal of Instrumentation* 3 (Aug. 2008), S08002–S08002. DOI: [10.1088/1748-0221/3/08/s08002](https://doi.org/10.1088/1748-0221/3/08/s08002). URL: <https://doi.org/10.1088/1748-0221/3/08/s08002>.
- [52] LHCb Collaboration. “The LHCb Detector at the LHC”. In: *Journal of Instrumentation* 3 (Aug. 2008), S08005–S08005. DOI: [10.1088/1748-0221/3/08/s08005](https://doi.org/10.1088/1748-0221/3/08/s08005). URL: <https://doi.org/10.1088/1748-0221/3/08/s08005>.
- [53] Isabel Béjar Alonso et al. *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*. Ed. by Isabel Béjar Alonso et al. CERN Yellow Reports: Monographs. Geneva: CERN, 2020. DOI: [10.23731/CYRM-2020-0010](https://cds.cern.ch/record/2749422). URL: <https://cds.cern.ch/record/2749422>.
- [54] ATLAS Collaboration. “How ATLAS detects particles: diagram of particle paths in the detector”. Jan. 2013. URL: <https://cds.cern.ch/record/1505342>.
- [55] ATLAS Collaboration. *ATLAS Insertable B-Layer Technical Design Report*. Tech. rep. Sept. 2010. URL: <https://cds.cern.ch/record/1291633>.
- [56] ATLAS Collaboration. *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*. Tech. rep. Geneva: CERN, Sept. 2017. URL: <https://cds.cern.ch/record/2285585>.
- [57] ATLAS Collaboration. *Technical Design Report for the ATLAS Inner Tracker Strip Detector*. Tech. rep. Geneva: CERN, Apr. 2017. URL: <https://cds.cern.ch/record/2257755>.
- [58] ATLAS Collaboration. *Particle Identification Performance of the ATLAS Transition Radiation Tracker*. Tech. rep. Geneva: CERN, Sept. 2011. URL: <https://cds.cern.ch/record/1383793>.

- [59] ATLAS Collaboration. *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, Dec. 1996. URL: <https://cds.cern.ch/record/331061>.
- [60] ATLAS Collaboration. *The ATLAS Tile Calorimeter*. Tech. rep. Geneva: CERN, Mar. 2015. DOI: [10.1109/ANIMMA.2015.7465554](https://cds.cern.ch/record/2004868). URL: <https://cds.cern.ch/record/2004868>.
- [61] ATLAS Collaboration. “The ATLAS Forward Calorimeter”. In: *Journal of Instrumentation* 3 (Feb. 2008), P02010–P02010. DOI: [10.1088/1748-0221/3/02/p02010](https://doi.org/10.1088/1748-0221/3/02/p02010). URL: <https://doi.org/10.1088/1748-0221/3/02/p02010>.
- [62] ATLAS Collaboration. “Computer generated image of the ATLAS Muons subsystem”. Mar. 2008. URL: <https://cds.cern.ch/record/1095929>.
- [63] ATLAS Collaboration. *Performance of the reconstruction of large impact parameter tracks in the ATLAS inner detector*. Tech. rep. Geneva: CERN, July 2017. URL: <https://cds.cern.ch/record/2275635>.
- [64] Rudolf E Kalman. “A New Approach to Linear Filtering and Prediction Problems”. In: *Journal of Basic Engineering* 82 (Mar. 1960), pp. 35–45. DOI: [10.1115/1.3662552](https://asmedigitalcollection.asme.org/fluidsengineering/article-pdf/82/1/35/5518977/35_1.pdf). eprint: https://asmedigitalcollection.asme.org/fluidsengineering/article-pdf/82/1/35/5518977/35_1.pdf. URL: <https://doi.org/10.1115/1.3662552>.
- [65] ATLAS Collaboration. “Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC”. In: *The European Physical Journal C* 77.5 (May 2017). DOI: [10.1140/epjc/s10052-017-4887-5](https://dx.doi.org/10.1140/epjc/s10052-017-4887-5). URL: <http://dx.doi.org/10.1140/epjc/s10052-017-4887-5>.
- [66] ATLAS Collaboration. “Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data”. In: *Journal of Instrumentation* 14 (Aug. 2019), P12006. 70 p. DOI: [10.1088/1748-0221/14/12/P12006](https://cds.cern.ch/record/2684552). arXiv: [1908.00005](https://arxiv.org/abs/1908.00005). URL: <https://cds.cern.ch/record/2684552>.
- [67] ATLAS Collaboration. *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*. Tech. rep. Geneva: CERN, May 2012. URL: <https://cds.cern.ch/record/1449796>.

- [68] ATLAS Collaboration. *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*. Tech. rep. Geneva: CERN, June 2016. URL: <https://cds.cern.ch/record/2157687>.
- [69] ATLAS Collaboration. *Photon identification efficiency measurements with the ATLAS detector using LHC Run 1 data*. Tech. rep. Geneva: CERN, Aug. 2014. URL: <https://cds.cern.ch/record/1747242>.
- [70] ATLAS Collaboration. *IsolationSelectionTool < AtlasProtected < TWiki*. <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/IsolationSelectionTool>.
- [71] ATLAS Collaboration. “Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data”. In: *Journal of Instrumentation* 14.03 (Mar. 2019), P03017–P03017. DOI: [10.1088/1748-0221/14/03/p03017](https://doi.org/10.1088/1748-0221/14/03/p03017). URL: <http://dx.doi.org/10.1088/1748-0221/14/03/P03017>.
- [72] ATLAS Collaboration. “Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV”. In: *The European Physical Journal C* 81 (Dec. 2020), 578. 44 p. DOI: [10.1140/epjc/s10052-021-09233-2](https://doi.org/10.1140/epjc/s10052-021-09233-2). arXiv: [2012.00578](https://arxiv.org/abs/2012.00578). URL: <https://cds.cern.ch/record/2746302>.
- [73] Richard K Ellis, William J Stirling, and Bryan R Webber. *QCD and collider physics*. Vol. 8. Cambridge University Press, Feb. 2011. ISBN: 978-0-511-82328-2, 978-0-521-54589-1.
- [74] Bo E Andersson et al. “Parton fragmentation and string dynamics”. In: *Physics Reports* 97 (July 1983), pp. 31–145. DOI: [https://doi.org/10.1016/0370-1573\(83\)90080-7](https://doi.org/10.1016/0370-1573(83)90080-7). URL: <https://www.sciencedirect.com/science/article/pii/0370157383900807>.
- [75] Giuseppe Marchesini et al. “HERWIG 5.1 - a Monte Carlo event generator for simulating hadron emission reactions with interfering gluons”. In: *Computer Physics Communications* 67 (Jan. 1992), pp. 465–508. DOI: [https://doi.org/10.1016/0010-4655\(92\)90055-4](https://doi.org/10.1016/0010-4655(92)90055-4). URL: <https://www.sciencedirect.com/science/article/pii/0010465592900554>.
- [76] Yu L Dokshitzer, Valery A Khoze, and S.I Troyan. “On the concept of local parton-hadron duality”. In: *Journal of Physics G: Nuclear and Particle Physics* 17 (Oct. 1991),

- pp. 1585–1587. DOI: [10.1088/0954-3899/17/10/017](https://doi.org/10.1088/0954-3899/17/10/017). URL: <https://doi.org/10.1088/0954-3899/17/10/017>.
- [77] ATLAS Collaboration. “Jet reconstruction and performance using particle flow with the ATLAS Detector”. In: *The European Physical Journal C* 77.7 (July 2017). DOI: [10.1140/epjc/s10052-017-5031-2](https://doi.org/10.1140/epjc/s10052-017-5031-2). URL: <http://dx.doi.org/10.1140/epjc/s10052-017-5031-2>.
- [78] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. “The anti-ktjet clustering algorithm”. In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. DOI: [10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063). URL: <http://dx.doi.org/10.1088/1126-6708/2008/04/063>.
- [79] Gavin P. Salam. “Towards jetography”. In: *The European Physical Journal C* 67.3-4 (May 2010), pp. 637–686. DOI: [10.1140/epjc/s10052-010-1314-6](https://doi.org/10.1140/epjc/s10052-010-1314-6). URL: <http://dx.doi.org/10.1140/epjc/s10052-010-1314-6>.
- [80] Yu L Dokshitzer et al. “Better jet clustering algorithms”. In: *Journal of High Energy Physics* 1997.08 (Aug. 1997), pp. 001–001. DOI: [10.1088/1126-6708/1997/08/001](https://doi.org/10.1088/1126-6708/1997/08/001). URL: <http://dx.doi.org/10.1088/1126-6708/1997/08/001>.
- [81] Stephen D Ellis and Davison E Soper. “Successive combination jet algorithm for hadron collisions”. In: *Phys. Rev. D* 48 (Oct. 1993), pp. 3160–3166. DOI: [10.1103/PhysRevD.48.3160](https://doi.org/10.1103/PhysRevD.48.3160). URL: <https://link.aps.org/doi/10.1103/PhysRevD.48.3160>.
- [82] ATLAS Collaboration. “Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Phys. Rev. D* 96 (Oct. 2017), p. 072002. DOI: [10.1103/PhysRevD.96.072002](https://doi.org/10.1103/PhysRevD.96.072002). URL: <https://link.aps.org/doi/10.1103/PhysRevD.96.072002>.
- [83] ATLAS Collaboration. “Jet energy resolution in proton-proton collisions at $\sqrt{s} = 13$ TeV recorded in 2010 with the ATLAS detector”. In: *The European Physical Journal C* 73.3 (Mar. 2013). DOI: [10.1140/epjc/s10052-013-2306-0](https://doi.org/10.1140/epjc/s10052-013-2306-0). URL: <http://dx.doi.org/10.1140/epjc/s10052-013-2306-0>.

- [84] ATLAS Collaboration. *Tagging and suppression of pileup jets with the ATLAS detector*. Tech. rep. Geneva: CERN, May 2014. URL: <https://cds.cern.ch/record/1700870>.
- [85] ATLAS Collaboration. *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*. Tech. rep. Geneva: CERN, Nov. 2015. URL: <https://cds.cern.ch/record/2064383>.
- [86] ATLAS Collaboration. “Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV”. In: *The European Physical Journal C* 78.11 (Nov. 2018). DOI: [10.1140/epjc/s10052-018-6288-9](https://doi.org/10.1140/epjc/s10052-018-6288-9). URL: <http://dx.doi.org/10.1140/epjc/s10052-018-6288-9>.
- [87] ATLAS Collaboration. “ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV”. In: *The European Physical Journal C* 79.11 (Nov. 2019). DOI: [10.1140/epjc/s10052-019-7450-8](https://doi.org/10.1140/epjc/s10052-019-7450-8). URL: <http://dx.doi.org/10.1140/epjc/s10052-019-7450-8>.
- [88] ATLAS Collaboration. “Search for phenomena beyond the Standard Model in events with large b-jet multiplicity using the ATLAS detector at the LHC”. In: *The European Physical Journal C* 81 (Jan. 2021), p. 11. DOI: [10.1140/epjc/s10052-020-08730-0](https://doi.org/10.1140/epjc/s10052-020-08730-0). arXiv: [2010.01015](https://arxiv.org/abs/2010.01015) [hep-ex].
- [89] ATLAS Collaboration. “Searches for third-generation scalar leptoquarks in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector”. In: *Journal of High Energy Physics* 06 (June 2019), p. 144. DOI: [10.1007/JHEP06\(2019\)144](https://doi.org/10.1007/JHEP06(2019)144). arXiv: [1902.08103](https://arxiv.org/abs/1902.08103) [hep-ex].
- [90] *FlavourTagging < AtlasProtected < TWiki*. <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/FlavourTagging>.
- [91] ATLAS Collaboration. *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*. Tech. rep. Geneva: CERN, July 2017. URL: <https://cds.cern.ch/record/2273281>.
- [92] ATLAS Collaboration. “Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV”. In: *Journal of High Energy Physics*

- 08 (Aug. 2018), p. 089. DOI: [10.1007/JHEP08\(2018\)089](https://doi.org/10.1007/JHEP08(2018)089). arXiv: [1805.01845](https://arxiv.org/abs/1805.01845) [hep-ex].
- [93] ATLAS Collaboration. *Calibration of light-flavour jet b-tagging rates on ATLAS data at $\sqrt{s} = 13$ TeV*. Tech. rep. Geneva: CERN, Sept. 2017. DOI: [10.22323/1.314.0745](https://doi.org/10.22323/1.314.0745). URL: <https://cds.cern.ch/record/2284133>.
- [94] ATLAS Collaboration. *Secondary vertex finding for jet flavour identification with the ATLAS detector*. Tech. rep. Geneva: CERN, June 2017. URL: <https://cds.cern.ch/record/2270366>.
- [95] ATLAS Collaboration. “A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS”. In: *Journal of Physics: Conference Series* 119 (July 2008), p. 032032. DOI: [10.1088/1742-6596/119/3/032032](https://doi.org/10.1088/1742-6596/119/3/032032). URL: <https://doi.org/10.1088/1742-6596/119/3/032032>.
- [96] Andreas Hocker et al. *TMVA - Toolkit for Multivariate Data Analysis with ROOT: Users guide*. *TMVA - Toolkit for Multivariate Data Analysis*. Tech. rep. TMVA-v4 Users Guide: 135 pages, 19 figures, numerous code examples and references. Geneva: CERN, Mar. 2007. URL: <https://cds.cern.ch/record/1019880>.
- [97] ATLAS Collaboration. *Search for the decay of the Higgs boson to charm quarks with the ATLAS experiment*. Tech. rep. Geneva: CERN, July 2017. URL: <https://cds.cern.ch/record/2274809>.
- [98] Francois Chollet et al. *Keras: the Python deep learning API*. 2015. URL: <https://github.com/fchollet/keras>.
- [99] The Theano Development. *Theano: A Python framework for fast computation of mathematical expressions*. 2016. arXiv: [1605.02688](https://arxiv.org/abs/1605.02688) [cs.SC].
- [100] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: [1412.6980](https://arxiv.org/abs/1412.6980) [cs.LG].
- [101] ATLAS Collaboration. “ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV”. In: *The European Physical Journal C* 79 (Nov. 2019), p. 970. DOI: [10.1140/epjc/s10052-019-7450-8](https://doi.org/10.1140/epjc/s10052-019-7450-8). arXiv: [1907.05120](https://arxiv.org/abs/1907.05120) [hep-ex].

- [102] Andrew Stuart Bell et al. *Calibration of b -tagging using di-leptonic $t\bar{t}$ events produced in pp collisions at $\sqrt{s} = 13$ TeV and a combinatorial likelihood approach*. Tech. rep. Geneva: CERN, Nov. 2016. URL: <https://cds.cern.ch/record/2231568>.
- [103] Laura Barranco Navarro et al. *Negative tag mistag rate calibration*. Tech. rep. Geneva: CERN, Sept. 2019. URL: <https://cds.cern.ch/record/2689901>.
- [104] Daniel Egana-Ugrinovic, Samuel Homiller, and Patrick Meade. “Higgs bosons with large couplings to light quarks”. In: *Phys. Rev. D* 100.11 (Dec. 2019). DOI: [10.1103/PhysRevD.100.115041](https://doi.org/10.1103/PhysRevD.100.115041). URL: <http://dx.doi.org/10.1103/PhysRevD.100.115041>.
- [105] Shaouly Bar-Shalom and Amarjit Soni. “Universally enhanced light-quarks Yukawa couplings paradigm”. In: *Phys. Rev. D* 98.5 (Sept. 2018). DOI: [10.1103/PhysRevD.98.055001](https://doi.org/10.1103/PhysRevD.98.055001). URL: <http://dx.doi.org/10.1103/PhysRevD.98.055001>.
- [106] Cédric Delaunay et al. “Enhanced Higgs boson coupling to charm pairs”. In: *Phys. Rev. D* 89 (Feb. 2014), p. 033014. DOI: [10.1103/PhysRevD.89.033014](https://doi.org/10.1103/PhysRevD.89.033014). URL: <https://link.aps.org/doi/10.1103/PhysRevD.89.033014>.
- [107] Francisco J Botella et al. “What if the masses of the first two quark families are not generated by the standard model Higgs boson?” In: *Phys. Rev. D* 94 (Dec. 2016), p. 115031. DOI: [10.1103/PhysRevD.94.115031](https://doi.org/10.1103/PhysRevD.94.115031). URL: <https://link.aps.org/doi/10.1103/PhysRevD.94.115031>.
- [108] Diptimoy Ghosh, Rick Sandeepan Gupta, and Gilad Perez. “Is the Higgs Mechanism of Fermion Mass Generation a Fact? A Yukawa-less First-Two-Generation Model”. In: *Physics Letters B* 755 (Apr. 2016), pp. 504–508. DOI: [10.1016/j.physletb.2016.02.059](https://doi.org/10.1016/j.physletb.2016.02.059). arXiv: [1508.01501](https://arxiv.org/abs/1508.01501) [hep-ph].
- [109] Stefano Agostinelli et al. “GEANT4—a simulation toolkit”. In: *Nucl. Instrum. Meth. A* 506 (July 2003), pp. 250–303. DOI: [10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [110] Konie Al Khoury et al. *Data/Monte Carlo Modelling Studies for the Standard Model $VH, H \rightarrow b\bar{b}$ Analysis*. Tech. rep. Geneva: CERN, Sept. 2019. URL: <https://cds.cern.ch/record/2690042>.

- [111] William Buttinger. *Using Event Weights to account for differences in Instantaneous Luminosity and Trigger Prescale in Monte Carlo and Data*. Tech. rep. Geneva: CERN, May 2015. URL: <https://cds.cern.ch/record/2014726>.
- [112] Luca Ambroz et al. *Object selections for SM Higgs boson produced in association with a vector boson in which $H \rightarrow b\bar{b}$ and V decays leptonically with Run-2 data*. Tech. rep. Geneva: CERN, Sept. 2019. URL: <https://cds.cern.ch/record/2690090>.
- [113] Konie Al Khoury et al. *Measurement of $VH, H \rightarrow b\bar{b}$ with the ATLAS detector*. Tech. rep. Geneva: CERN, Oct. 2019. URL: <https://cds.cern.ch/record/2692011>.
- [114] Eilam Gross et al. *Search for $H \rightarrow c\bar{c}$ decays in associated ZH production*. Tech. rep. Geneva: CERN, Dec. 2016. URL: <https://cds.cern.ch/record/2237398>.
- [115] Johannes Erdmann et al. “A likelihood-based reconstruction algorithm for top-quark pairs and the KLFFitter framework”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 748 (June 2014), pp. 18–25. DOI: <https://doi.org/10.1016/j.nima.2014.02.029>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900214001855>.
- [116] ATLAS Collaboration. *Measurement of the c -jet tagging efficiency in $t\bar{t}$ events using a likelihood approach using pp collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS detector*. Tech. rep. Geneva: CERN, Dec. 2020. URL: <https://cds.cern.ch/record/2747223>.
- [117] Kyle Cranmer et al. *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*. Tech. rep. New York: New York U., Jan. 2012. URL: <https://cds.cern.ch/record/1456844>.
- [118] Wouter Verkerke and David Kirkby. *The RooFit toolkit for data modeling*. 2003. arXiv: [physics/0306116](https://arxiv.org/abs/physics/0306116) [physics.data-an].
- [119] Lorenzo Moneta et al. “The RooStats Project”. In: *PoS ACAT2010* (Sept. 2010), p. 057. DOI: [10.22323/1.093.0057](https://doi.org/10.22323/1.093.0057). arXiv: [1009.1003](https://arxiv.org/abs/1009.1003) [physics.data-an].

- [120] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *The European Physical Journal C* 71.2 (Feb. 2011). DOI: [10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0). URL: <http://dx.doi.org/10.1140/epjc/s10052-011-1554-0>.
- [121] Alexander L Read. “Presentation of search results: the CLs technique”. In: 28 (Sept. 2002), pp. 2693–2704. DOI: [10.1088/0954-3899/28/10/313](https://doi.org/10.1088/0954-3899/28/10/313). URL: <https://doi.org/10.1088/0954-3899/28/10/313>.
- [122] The ATLAS collaboration. “Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC”. In: (June 2019).
- [123] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z Skands. “A Brief Introduction to PYTHIA 8.1”. In: *Comput. Phys. Commun.* 178 (June 2008), pp. 852–867. DOI: [10.1016/j.cpc.2008.01.036](https://doi.org/10.1016/j.cpc.2008.01.036). arXiv: [0710.3820](https://arxiv.org/abs/0710.3820) [hep-ph].
- [124] Stefano Frixione, Giovanni Ridolfi, and Paolo Nason. “A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction”. In: *Journal of High Energy Physics* 2007 (Sept. 2007), pp. 126–126. DOI: [10.1088/1126-6708/2007/09/126](https://doi.org/10.1088/1126-6708/2007/09/126). URL: <https://doi.org/10.1088/1126-6708/2007/09/126>.
- [125] Richard Corke and Torbjorn Sjostrand. “Interleaved Parton Showers and Tuning Prospects”. In: *Journal of High Energy Physics* 03 (Mar. 2011), p. 032. DOI: [10.1007/JHEP03\(2011\)032](https://doi.org/10.1007/JHEP03(2011)032). arXiv: [1011.1759](https://arxiv.org/abs/1011.1759) [hep-ph].
- [126] Manuel Bähr et al. “Herwig++ physics and manual”. In: *The European Physical Journal C* 58.4 (Dec. 2008), pp. 639–707. DOI: [10.1140/epjc/s10052-008-0798-9](https://doi.org/10.1140/epjc/s10052-008-0798-9). URL: <https://doi.org/10.1140/epjc/s10052-008-0798-9>.
- [127] ATLAS Collaboration. *New ATLAS event generator tunes to 2010 data*. Tech. rep. ATL-PHYS-PUB-2011-008. Geneva: CERN, Apr. 2011. URL: <https://cds.cern.ch/record/1345343>.
- [128] Hung-Liang Lai et al. “New parton distributions for collider physics”. In: *Phys. Rev. D* 82 (Oct. 2010), p. 074024. DOI: [10.1103/PhysRevD.82.074024](https://doi.org/10.1103/PhysRevD.82.074024). URL: <https://link.aps.org/doi/10.1103/PhysRevD.82.074024>.

תקציר

העבודה הזו מתארת חיפוש אחר הדעיכה של בוזון ההיגס לזוג של קוורק ואנטי קוורק מסוג "קסם" (charm), באמצעות 139 fb^{-1} של מידע הנאסף מהתנגשויות פרוטון-פרוטון באנרגיית מרכז מסה של 13 TeV בגלאי ATLAS שבמאיץ ההדרונים הגדול שב-CERN. החיפוש מתמקד באירועים בהם בוזון ההיגס נוצר ביחד עם בוזון כיול אלקטרו-חלש, כאשר ההתפרקות של בוזון הכיול היא ללפטונים. היות וללפטונים חתימה יחסית נקיה בגלאי, על ידי חיפוש במצבים סופיים הכוללים לפטונים ניתן להקטין את ההשפעה של אירועי QCD, אשר מהווים את הרקע הדומיננטי במאיץ ההדרונים. על מנת לזהות סילונים שנוצרו מקוורקים מסוג "קסם", נעשה שימוש באלגוריתמים מרובי-משתנים, המנצלים את התכונות היחודיות של התפרקויות הדרונים מסוג charm (ומסוג bottom), כדי להבדיל בין סילוני charm לסילוני bottom או לסילונים הנוצרים מקוורקים קלים יותר. החיפוש עושה שימוש בהתאמה סטטיסטית להתפלגות המסה של צמד הסילונים בעלי התנע הרב ביותר במישור המאונך להתנגשות, ממנה נגזרת עוצמת הסיגנל $\mu_{VH(cc)}$, המייצגת את היחס בין הקצב המדוד של התהליך לקצב שנחזה על ידי המודל הסטנדרטי. בחיפוש לא נצפו ראיות לתהליך ובהתאם לכך הוצבו חסמים עליונים על הערכים האפשריים של $\mu_{VH(cc)}$ ושל קבוע צימוד Yukawa בין ההיגס וקוורק ה"קסם", κ_c . תוצאה זו מהווה את החסם הישיר הראשון על κ_c .



School of Physics and Astronomy **בית הספר לפיזיקה ולאסטרונומיה**
The Raymond and Beverly Sackler **הפקולטה למדעים מדויקים**
Faculty of Exact Sciences **ע"ש ריימונד ובברלי סאקלר**
Tel Aviv University **אוניברסיטת תל אביב**

חיבור מדעי המוגש כחלק מתואר "דוקטור לפילוסופיה"

חיפוש אחר התפרקויות של בוזון ההיגס לצמדי קוורקים מסוג קסם

מוגש לסנאט אוניברסיטת תל אביב על ידי

גיא קורן

העבודה נעשתה בהנחיית
פרופסור ארז עציון

ינואר 2022