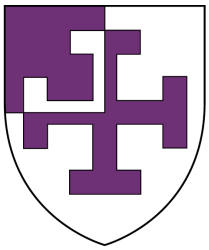
Study of a highly boosted Higgs boson decaying to b-jets final states with the ATLAS experiment at CERN





Maria Giovanna Foti

St Cross College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2021

To Oxford, the place where I could truly be myself. It has been an incredible honour, a real pleasure and a neverending love. You will always hold a special place in my heart.

Abstract

The observation of the decay of the Higgs boson to a $b\bar{b}$ pair in 2018 confirmed the Yukawa interaction and provided the first measurement of a Yukawa coupling to down-type quarks.

The most sensitive channel in the $H \to b\bar{b}$ measurement is given by the associated Higgs boson production with a vector boson V decaying leptonically.

The main focus is now the measurement of the Higgs boson production cross section as a function of its transverse momentum (p_T). This study is motivated by a wide range of Beyond the Standard Model (BSM) theories predicting an enhancement of $VH(H \to b\bar{b})$ events at high transverse momentum with respect to the SM expectations. At such high energy regimes, the Higgs boson is highly Lorentz-boosted and the $H \to b\bar{b}$ decay is reconstructed in the detector as a single object with a two-prong substructure.

This thesis describes the measurement of $VH(H \to b\bar{b})$ in the boosted regime performed inclusively, as well as in bins of the vector boson transverse momentum, for $p_T^V \in [250, 400[$ GeV and $p_T^V \in [400, \infty[$ GeV. The interpretation of these measurements using the Effective Field Theory (EFT) framework is also discussed. The results are based on the full LHC Run 2 dataset collected by the ATLAS experiment, corresponding to an integrated luminosity of 139 fb^{-1}.

A very important component of this study is the correct identification of b-jets. A new calibration of the b-jet tagging algorithm at very high transverse momentum, which reduced the uncertainties for high-p_T b-jets by 50%, is also included.

Acknowledgements

"No man is an island" and I am no exception. My experience as a DPhil student would have been much more arduous without all the people who encouraged me along this journey.

First and foremost, I would like to deeply thank my supervisor Daniela Bortoletto for her invaluable guidance, suggestions and support during these years. She entrusted me with scientific and personal freedom while always being there when I needed: she has never been too busy to discuss something, from the most important decisions to the tiniest details. It has been a privilege and an honour to be your student, and I will treasure the lessons you taught me for the future.

I would like to express my sincere gratitude to Valerio Dao, this thesis would not be the same without his precious contribution. He has been an incredible source of everything: knowledge, inspiration, advice, confidence, resources, patience and coffee! Thank you for being a generous leader and motivator, you deserve an equestrian statue.

Special thanks must go to all the members of the ATLAS $H \to b\bar{b}$ group and $VH \to b\bar{b}$ team, for being an great example of collaborative effort. In particular, I am very thankful to Elisabeth, for introducing me to the $VH \to b\bar{b}$ group in the best possible way and for always being available to clarify my doubts, and Hannah for leading the boosted analysis so well, along with Valerio. You have guided us towards this result, firm and

determined, but always being thoughtful and considerate. I always felt cared for as a group member and as a person, and I truly enjoyed working with you!

I would like to express my gratitude to the Bortoletto/Shipsey group for being my "scientific family". The discussions during our meetings have been extremely instructive and fruitful, and have always left me with food for thoughts. Particular thanks to Ian Shipsey, for always encouraging us to think outside the box, and Giacomo, for his continuous support and his capability to seamlessly cover so many layers of ATLAS knowledge. This list is not exhaustive but includes also Cecilia, Luca, Luigi M, Luigi V, Maria, Mariyan, Miha, Ricardo, Philipp, Siyuan and Yingjie.

During my DPhil I also worked in two different laboratories. In both of them I have been welcomed from the very beginning and immediately felt part of the team. Many thanks to the ATLAS Pixel Lab team at CERN and in particular to Andrea, for guiding me in my first hardware experience and for the support overseas, and to Carlos. Sincere thanks also to the OPMD group, especially to Richard and Kaloyan.

I would like to thank the CMS Bologna group for introducing me to Particle Physics, and a warm thank to my former supervisor Andrea Castro.

I would like to express my gratitude to the Clarendon Council, the Science and Technology Facilities Council and to St Cross College, for financially supporting my studies. I also want to thank Sue Geddes and Kim Proudfoot: their dedication has kept my mind focused on my research.

Finally, none of this would have been possible without the love and help of my friends, in Oxford and all over Europe, and my parents. Thank you Cri for always reminding me that: "si fa tutto".

Contents

Preface **1**

1 Theoretical Introduction **6**

1.1 Particles of the Standard Model 6

1.2 Interactions in the Standard Model 8

1.2.1 The Electroweak Theory 9

1.3 Spontaneous Symmetry Breaking and the Brout-Englert-Higgs mechanism 12

1.3.1 Masses of the Fermions 14

1.4 Higgs phenomenology 16

1.4.1 Higgs boson production at the LHC 17

1.4.2 Higgs boson decay modes 19

1.5 QCD phenomenology and cross section calculations at the LHC . . . 20

1.5.1 Monte-Carlo simulations and uncertainties 24

1.6 Limitations of the SM 26

1.6.1 Effective Field Theories 28

2 The Large Hadron Collider and the ATLAS experiment **30**

2.1 The Large Hadron Collider 30

2.1.1 The accelerator complex 31

2.1.2 Centre-of-mass energy 33

2.1.3 Luminosity . . . 34
2.1.4 Pile-up . . . 35
2.2 The ATLAS detector . . . 37
2.2.1 Coordinate System . . . 38
2.2.2 Magnet system . . . 39
2.2.3 Inner Detector . . . 40
2.2.4 Calorimetry . . . 43
2.2.5 Muon Spectrometer . . . 48
2.2.6 Trigger and Data Acquisition System . . . 49

3 Object reconstruction with the ATLAS experiment 51
3.1 ATLAS Simulation Infrastructure . . . 52
3.2 Detector objects . . . 54
3.2.1 Tracks . . . 54
3.2.2 Topological clusters . . . 55
3.3 High-level physics objects . . . 56
3.3.1 Electrons . . . 57
3.3.2 Photons . . . 58
3.3.3 Muons . . . 59
3.3.4 Jets . . . 61
3.3.4.1 Small-R jets . . . 63
3.3.4.2 Large-R jets . . . 67
3.3.4.3 Track jets . . . 71
3.3.5 Taus . . . 72
3.3.6 Missing transverse momentum and energy . . . 73
3.3.7 Overlap removal procedure . . . 75

4 *b*-Tagging **76**

4.1 Physical foundations of b-tagging algorithms . . . 76

4.2 b-Tagging algorithms in ATLAS . . . 78

4.2.1 Track-to-jet association for b-tagging algorithms . . . 78

4.2.2 Algorithms based on impact parameters . . . 79

4.2.3 Algorithm based on secondary vertex reconstruction . . . 81

4.2.4 Topological decay-chain reconstruction algorithm . . . 83

4.2.5 The `MV2` multivariate algorithm . . . 84

4.3 Usage of b-tagging in ATLAS physics analyses . . . 87

4.3.1 Performance of `MV2` . . . 87

4.3.2 Application of b-tagging to physics analysis . . . 88

4.4 Calibration of b-jets at very high transverse momentum . . . 91

4.4.1 Calibration strategy . . . 91

4.4.2 Samples, event and object selection . . . 94

4.4.3 Systematic uncertainties . . . 95

4.4.3.1 Track-related uncertainties . . . 96

4.4.3.2 Jet-related uncertainties . . . 98

4.4.3.3 Modelling uncertainties . . . 98

4.4.4 Results . . . 98

5 Measurement of $VH(H \to b\bar{b})$ at high transverse momentum **103**

5.1 Analysis strategy . . . 105

5.2 Data and Simulation Samples . . . 107

5.3 Object and Event Selection and Categorisation . . . 110

5.3.1 Object selection . . . 110

5.3.2 Event selection . . . 113

5.3.3 Event categorisation . . . 119

5.4 Optimisation of the analysis selection criteria: studies on the $\Delta R(b_1, b_2)$ in the 0-lepton channel . . . 127
5.5 Statistical analysis . . . 132
5.5.1 The VH-VZ fit . . . 136
5.6 Systematic Uncertainties . . . 138
5.6.1 Experimental Uncertainties . . . 138
5.6.2 Uncertainties on the modelling of signal and background processes 140
5.7 Analysis results . . . 151
5.8 Simplified Template Cross Section and Effective Field Theory Interpretation . . . 155

6 Future measurements of $VH(H \to b\bar{b})$ 162
6.1 The $VH(H \to b\bar{b})$ resolved analysis . . . 163
6.1.1 Analysis strategy . . . 163
6.1.2 Signal and Background processes . . . 164
6.1.3 Event selection and categorisation . . . 164
6.1.4 Multi Variate Analysis (MVA) . . . 165
6.1.5 Results . . . 168
6.2 Studies of the overlap of the $VH(H \to b\bar{b})$ resolved and boosted analyses 169
6.3 Studies of the combination strategy of the $VH(H \to b\bar{b})$ resolved and boosted analyses . . . 173
6.4 Studies of three b-jet events in the $VH(H \to b\bar{b})$ resolved analysis . . . 179
6.4.1 Event topology in the 3b-CR . . . 180
6.4.2 First studies for the addition of the 3b-CR . . . 182

7 Conclusions 188

A Future silicon pixel detectors: the ATLAS upgrade and beyond 191
A.1 Introduction on silicon detectors . . . 191

A.2 The ITk upgrade 193
A.3 The MALTA project 197

B Pull Plots for the $VH(H \to b\bar{b})$ boosted analysis **204**

Bibliography **209**

List of Figures

1.1 Particles of the SM 7

1.2 Higgs potential 13

1.3 Higgs boson production mechanisms at the LHC 17

1.4 Theoretical predictions for the Higgs boson production cross sections and decay rates at the LHC 19

1.5 Experimental measurements of Higgs boson production cross sections at the LHC 21

1.6 Experimental measurements of Higgs boson decays by the ATLAS experiment 22

1.7 Diagram of the main steps of the generation chain for MC simulations 25

2.1 The LHC tunnel 31

2.2 The LHC RF cavities 32

2.3 CERN's accelerator complex 32

2.4 Instantaneous and integrated luminosity delivered and collected by the ATLAS experiment 35

2.5 LHC Run 2 number of interactions per bunch crossing over the years 36

2.6 The ATLAS detector 37

2.7 ATLAS coordinate system 38

2.8 ATLAS magnet system 40

2.9 ATLAS ID 41

2.10 ATLAS Calorimetry 45
2.11 ATLAS Muon Spectrometer 48

3.1 Particles' experimental signatures inside the ATLAS detector 52
3.2 Jet reconstruction with the anti-k_t algorithm 62
3.3 JES correction, JER distribution and uncertainties on small-R jets . . 66
3.4 JVT pile-up discriminator output 67
3.5 Fractional large-R jet mass resolution for m^{calo}, m^{TA} and m^{comb} . . . 68
3.6 JES and JMS uncertainties on large-R jets 70
3.7 VR track jet radius vs p_T 71
3.8 $\vec{E}_T^{miss}$ resolution 74

4.1 Illustration of a b-jet 77
4.2 Track-to-jet association function, for b-tagging algorithms 79
4.3 Event display for a candidate $ZH \to \nu\nu b\bar{b}$ event 80
4.4 Impact parameter significance for jets of different flavour 82
4.5 Flavour tagging discriminant for the impact parameter based algorithms 82
4.6 Properties of reconstructed secondary vertices 83
4.7 Event display of a simulated B^0 decay 84
4.8 b-jet transverse momentum for $t\bar{t}$ and Z' events 85
4.9 `MV2` b-tagging output and rejection rate 87
4.10 b-tagging simulation-to-data scale factors and uncertainties 89
4.11 Jet p_T distributions for b-, c- and light-flavour jets 92
4.12 Simulation-based b-tagging efficiencies and uncertainties for `MV2` . . . 99
4.13 b-tagging simulation-based uncertainties for calorimeter jets 100
4.14 b-tagging simulation-based uncertainties for VR track jets 101

5.1 Effect of EFT operators on the Higgs boson p_T 104
5.2 Feynman diagrams for the LO quark initiated SM $VH(H \to b\bar{b})$ processes 106

5.3 Sketch of the $H \to b\bar{b}$ decay topology as a function of p_T 106
5.4 Signal and background processes of the $VH(H \to b\bar{b})$ boosted analysis 109
5.5 Jet energy corrections applied to the large-R Higgs boson candidate jet in the $VH(H \to b\bar{b})$ boosted analysis 113
5.6 Sketch of the boosted $VH(H \to b\bar{b})$ process 114
5.7 m_J distribution in the 0-lepton channel of the $VH(H \to b\bar{b})$ boosted analysis 115
5.8 m_J distribution in the 1-lepton channel of the $VH(H \to b\bar{b})$ boosted analysis 116
5.9 m_J distribution in the 2-lepton channel of the $VH(H \to b\bar{b})$ boosted analysis 118
5.10 p_T^V distributions in the 0-, 1- and 2-lepton channel of the $VH(H \to b\bar{b})$ boosted analysis 120
5.11 Graphical representation of the signal $VH(H \to b\bar{b})$ and $t\bar{t}$ background reconstruction in the $VH(H \to b\bar{b})$ boosted analysis 121
5.12 Distributions of the number of additional b-jets in the 0- and 1-lepton channel of the $VH(H \to b\bar{b})$ boosted analysis 122
5.13 Event distribution in the CRs of the 0- and 1-lepton channel of the $VH(H \to b\bar{b})$ boosted analysis 123
5.14 Distributions of the number of additional small-R calorimeter jets in the 0- and 1-lepton channel of the $VH(H \to b\bar{b})$ boosted analysis . . 124
5.15 Event distribution in the high purity and low purity SRs of the 0-lepton channel of the $VH(H \to b\bar{b})$ boosted analysis 125
5.16 Event distribution in the high purity and low purity SRs of the 1-lepton channel of the $VH(H \to b\bar{b})$ boosted analysis 126
5.17 $\Delta R\,(b_1, b_2)$ distributions for signal and background processes in the 0-lepton channel of the $VH(H \to b\bar{b})$ boosted analysis 128

5.18 Binned statistical significance as a function of the $\Delta R\,(b_1, b_2)$ selection in the 0-lepton channel of the $VH(H \to b\bar{b})$ boosted analysis 129
5.19 Selection efficiency for signal and background events as a function of $\Delta R\,(j_1, j_2)$ selection in the 0-lepton channel of the $VH(H \to b\bar{b})$ boosted analysis 129
5.20 Impact of the $\Delta R\,(b_1, b_2)$ selection on the event distribution in the 0-lepton channel of the $VH(H \to b\bar{b})$ boosted analysis 131
5.21 Log-likelihood ratio and p-value 135
5.22 Correlations between the two parameters of interest $\mu_{VH}^{b\bar{b}}$ and $\mu_{VZ}^{b\bar{b}}$ in the $VH(H \to b\bar{b})$ boosted analysis 137
5.23 m_J distribution in $t\bar{t}$ events for different flavour compositions in the 0-lepton channel of the $VH(H \to b\bar{b})$ boosted analysis 145
5.24 m_J distribution in $t\bar{t}$ events for ISR variations in the $VH(H \to b\bar{b})$ boosted analysis 148
5.25 m_J distribution in $t\bar{t}$ events for the PS variation in the $VH(H \to b\bar{b})$ boosted analysis 148
5.26 Event display for a candidate $WH \to b\bar{b}$ event 152
5.27 Background-subtracted m_J distribution and likelihood contours in the $VH(H \to b\bar{b})$ boosted analysis 153
5.28 Fitted values of $\mu_{VH}^{b\bar{b}}$ and $\mu_{VZ}^{b\bar{b}}$ of the $VH(H \to b\bar{b})$ boosted analysis . 156
5.29 VH STXS measurement results of the $VH(H \to b\bar{b})$ boosted analysis 158
5.30 Summary of the observed CL limits for the Wilson coefficients of the EFT operators in the $VH(H \to b\bar{b})$ boosted analysis 159
5.31 Effect of $p_T^V > 400$ GeV STXS bin on EFT operator 160

6.1 Signal yield distribution of $\Delta R(b_1, b_2)$ in the $VH(H \to b\bar{b})$ resolved analysis 166

6.2 Fraction of signal $VH(H \to b\bar{b})$ events selected by the resolved analysis and boosted analysis only, or both, as a function of p_T^V 170
6.3 Signal distribution of $\Delta R(b_1, b_2)$ vs p_T^V for resolved only, boosted only and overlap events 171
6.4 Signal distribution of $\Delta R(b_1, b_2)$, p_T^V, Higgs boson candidate mass and p_T for overlap events 172
6.5 Sketch of the possible different combination strategies for the $VH(H \to b\bar{b})$ boosted and resolved analyses 175
6.6 Statistical significances for the possible different combination strategies for the $VH(H \to b\bar{b})$ boosted and resolved analyses in the 0-lepton channel 177
6.7 E_T^{miss} distribution in the 3b-CR in the 0-lepton channel of the $VH(H \to b\bar{b})$ resolved analysis, for 150 GeV $\leq p_T^V <$ 400 GeV 181
6.8 m_{bb} and $\Delta R(b_1, b_2)$ distribution in the 3b-CR in the 0-lepton channel of the $VH(H \to b\bar{b})$ resolved analysis, for 150 GeV $\leq p_T^V <$ 400 GeV, for the *leading two b-jets* and *closest two b-jets* strategy 183
6.9 Comparison of the floating normalisations for the 0-lepton Asimov fit with and without the introduction of the 3b-CR for the $VH(H \to b\bar{b})$ resolved analysis 185
6.10 Conditional post-fit output distributions for the 0-lepton fit with the new 3b-CR in the $VH(H \to b\bar{b})$ resolved spectrum 186
6.11 m_{bb} distributions of $t\bar{t}$ events for the bc component in the 2 b-jet and 3b-CR region in the 0-lepton channel of the $VH(H \to b\bar{b})$ resolved analysis 187

A.1 Event display of a $t\bar{t}$ event in the ATLAS ITk at the HL-LHC 193
A.2 The ITk design 194
A.3 Schematic view of am ITk bare pixel module 195

A.4 Experimental setup for the ITk testing qualification in Oxford 195
A.5 50× magnified picture of the digital quad used for the Oxford testing qualification 196
A.6 Temperature of the digital quad during testing 197
A.7 Results of the tuning procedure of the ITk pixel digital quads 198
A.8 MALTA sensor 199
A.9 Results of the read-out test on two MALTA chips 201
A.10 Results of the threshold and noise tests on MALTA 202
A.11 Flow chart of the iterative masking procedure for MALTA 203

B.1 NP pulls, constraints and scale factors 205
B.2 Signal and background modelling-related NP pulls, constraints and post-fit normalisation scale factors 206
B.3 Detector-related NP pulls and constraints 207
B.4 Simulation statistics NP pulls 208

List of Tables

4.1 Input variables used by MV2 86

4.2 Selection criteria and rejection rates for the MV2 working points . . . 88

4.3 b-tagging simulation-based uncertainties for calorimeter jets 99

4.4 b-tagging simulation-based uncertainties for variable-R track jets . . . 102

5.1 MC generators for signal and background processes of the $VH(H \to b\bar{b})$ boosted analysis 108

5.2 Selection criteria for *baseline* electrons and muons in the $VH(H \to b\bar{b})$ boosted analysis 111

5.3 Selection criteria for *signal* electrons and muons in the $VH(H \to b\bar{b})$ boosted analysis 111

5.4 Event selection for the boosted $VH(H \to b\bar{b})$ analysis 119

5.5 Analysis categories of the $VH(H \to b\bar{b})$ boosted analysis 124

5.6 Summary of $\Delta R\,(b_1, b_2)$ selection studies in the 0-lepton channel of the $VH(H \to b\bar{b})$ boosted analysis 130

5.7 Floating normalisation values for the main backgrounds of the $VH(H \to b\bar{b})$ boosted analysis 154

5.8 Breakdown of the uncertainties in the $VH(H \to b\bar{b})$ boosted analysis 154

5.9 Wilson coefficients and corresponding EFT operators for the $VH(H \to b\bar{b})$ boosted analysis 159

6.1 Variables used for the multivariate discriminant in the $VH(H \to b\bar{b})$ resolved analysis 167
6.2 Event selection for the $VH(H \to b\bar{b})$ boosted-resolved overlap studies in the 0-lepton channel 169
6.3 Signal, background and data yields in the $3b$-CR in the 0-lepton channel of the $VH(H \to b\bar{b})$ resolved analysis, for 150 GeV $\leq p_T^V <$ 400 GeV . 180
6.4 Results of the studies for the introduction of the $3b$-CR in the 0-lepton channel of the $VH(H \to b\bar{b})$ resolved analysis 184

Preface

The studies presented in this thesis are the result of my work with the ATLAS Collaboration. The purpose of this manuscript is to present the main results achieved during this collaboration and describe my original contributions.

The major projects carried out during my DPhil are outlined in chronological order in what follows.

Calibration of b-jets at very high transverse momentum A critical component of analyses both in the Higgs sector (such as the study of the $H \to b\bar{b}$ decay) and searches for physics beyond the Standard Model is the correct identification of b-jets. For my ATLAS Qualification Task, I calibrated the b-jet tagging algorithm at very high transverse momentum working with the Flavour Tagging Group. This calibration led to a reduction of the uncertainties for high-p_T b-jets of 50%. The results have been incorporated in Reference [1] and are now widely used by the ATLAS Collaboration.

Measurement of the $VH(H \to b\bar{b})$ production at high transverse momenta I was one of the leading analysers of the $VH(H \to b\bar{b})$ boosted analysis [2], measuring the associated production of a vector boson ($V = W$, Z) with the Higgs boson decaying to a pair of b quarks at high transverse momentum and contributed to several crucial parts of the analysis, with a focus on the $ZH \to \nu\nu b\bar{b}$ (*0-lepton*) channel.

Firstly, I performed studies on jet substructure variables, exploiting the angular

separation between the two objects in which the decay products of the Higgs boson are reconstructed. This study highlighted that a substructure-based selection was not possible with the available data, but it should be used in the future when more data will become available.

Secondly, I was one of the main analysers studying the modelling of the background processes, leading the modelling studies for top pair production. I studied the topology of $t\bar{t}$ events in the analysis phase space and I evaluated the impact of the most important sources of systematic uncertainties affecting $t\bar{t}$. Thanks to the expertise gained during the $t\bar{t}$ studies for the $VH(H \to b\bar{b})$ boosted analysis, I became the analyser driving the modelling effort, supporting other team members performing studies on other processes.

I was also one of the main analysers developing the statistical analysis. Traditionally, the $VH(H \to b\bar{b})$ analysis had used the $VZ(Z \to b\bar{b})$ process as a benchmark to test any new features implemented in $VH(H \to b\bar{b})$, because of the similarities between the two processes. For boosted $VH(H \to b\bar{b})$ I designed and studied a two-parameter-of-interest fit, measuring simultaneously the $VZ(b\bar{b})$ and $VH(b\bar{b})$ signal strengths, to exploit their correlations. The fit became the default for the analysis, and I extensively studied its features and stability.

Measurements of WH and ZH production in the $H \to b\bar{b}$ decay channel I also contributed to the modelling studies of the "resolved" $VH(H \to b\bar{b})$ analysis [3], that reconstructs the Higgs boson candidate as a pair of two b-tagged small-radius jets.

Study of VH production in the $H \to b\bar{b}$ decay channel After the publication of the $VH(H \to b\bar{b})$ boosted and resolved analysis, the natural follow-up of this work is a $VH(H \to b\bar{b})$ analysis combining the "boosted" and "resolved"

$VH(H \to b\bar{b})$ analyses. Several challenges need to be addressed to achieve this goal, in terms of the object reconstruction, the b-tagging strategy and the treatment of resolved-boosted $VH(H \to b\bar{b})$ overlap events. I contributed to this effort by studying overlap events in the $ZH \to \nu\nu b\bar{b}$ channel and comparing different combination strategies, setting the grounds for the subsequent studies.

I also introduced a "$t\bar{t}$-dedicated" control region in the resolved $VH(H \to b\bar{b})$ analysis. This additional region, built by selecting events with three b-tagged jets in the resolved $VH(H \to b\bar{b})$ phase space, improves the $t\bar{t}$ background estimations, especially in the 0-lepton channel.

Detector Operations I contributed to the Inner Detector (ID) operations as an ID Control Room shifter and SCT Expert Shifter, for $\sim$ 50 hours, during the "Milestone weeks", organised in preparation for the start of the LHC Run 3 in 2022.

Development of silicon pixel detectors: MALTA The MALTA [4] pixel project focuses on Depleted Monolithic Active Pixel Sensors (DMAPS) produced in CMOS technology. The project started as an option for the outermost pixel layer of ITk [5]. I contributed to MALTA by developing a standardised Quality Assurance (QA) procedure to assess the performance of a chip through a series of short tests. The QA targets several features of the chips, such as the timing response, noise rate, pixel thresholds and pixel noise. The procedure is summarised in a document and is now regularly used on new chips.

Development of silicon detectors: the ITk pixel detector The increase in instantaneous luminosity and pile-up brought by the HL-LHC, along with the data rates and accumulated radiation damage, will render the current ATLAS inner tracker inoperable. The current ATLAS ID [6] will be replaced with a new all-silicon tracker, the ITk.

Many ATLAS institutions are now working to produce and test prototype ITk modules. Since testing assesses the modules' quality, it must be done correctly. Therefore institutions are asked to qualify as a testing site. I contributed to the qualification of the Oxford Physics Microstructure Detector facility by testing the read-out of pixel digital quad modules (RD53A) [7].

Outreach activities During my DPhil I took part in many outreach activities: I was a member of the Organising Committee of the Conference for Undergraduate Women in Physics (CUWiP) that took place in Oxford in 2018. I also took part in the ATLAS Guided Tours as an ATLAS underground guide, the CERN Open Days 2019 and the International Masterclasses in 2020 and 2021.

During my studies, I was also awarded the John Adam Prize in 2018 from the Oxford Sub-Department of Particle Physics, for "outstanding performance in the first year of postgraduate study for the year 2017/2018".

The thesis is structured as follows. Chapter 1 provides a theoretical introduction to some fundamental concepts in Particle Physics and the Higgs mechanism. Chapter 2 describes the LHC accelerator complex, as well as the ATLAS detector. Chapter 3 presents the physics objects widely used in the ATLAS Collaboration, with a focus on those used in the boosted and resolved $VH(H \to b\bar{b})$ analyses. Chapter 4 describes one of the most widely used b-tagging algorithm in ATLAS and my original work on the calibration of b-jets at very high transverse momentum. Chapter 5 describes the $VH(H \to b\bar{b})$ boosted analysis and includes my main contributions to the analysis development. Chapter 6 contains studies detailing possible analysis improvements for the next iteration of the $VH(H \to b\bar{b})$ analysis, merging the resolved and boosted regimes. I have produced all the plots in these last three chapters, except for the ones marked with the ATLAS label, created in collaboration with the analysis group and

published in the respective publications. Finally, Chapter 7 provides a summary and conclusions.

Chapter 1

Theoretical Introduction

The Standard Model of Particle Physics (SM) [8, 9] is the theoretical framework that describes the fundamental constituents of the universe and their interactions.

The SM remained incomplete until 2012, when the crucial discovery of the Higgs boson by the ATLAS and CMS collaborations [10, 11] was announced.

The nature of the Higgs boson, however, has to be yet fully understood through experimental measurements, that could potentially unveil effects of physics Beyond the Standard Model (BSM). The importance of the Higgs field lies in the fact that it is responsible for the mechanisms that generate the mass of elementary particles.

In this chapter, an introduction to the Standard Model, the Higgs boson and a model-independent tool to unveil new physics, are presented.

1.1 Particles of the Standard Model

The Standard Model describes the electromagnetic, weak and strong interactions between elementary particles[1], using the formalism of relativistic Quantum Field Theory (QFT), based on a $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetry [12], where, by Noether's theorem, each symmetry corresponds to a conserved quantity.

[1] A theory of gravitation is currently not included in the SM

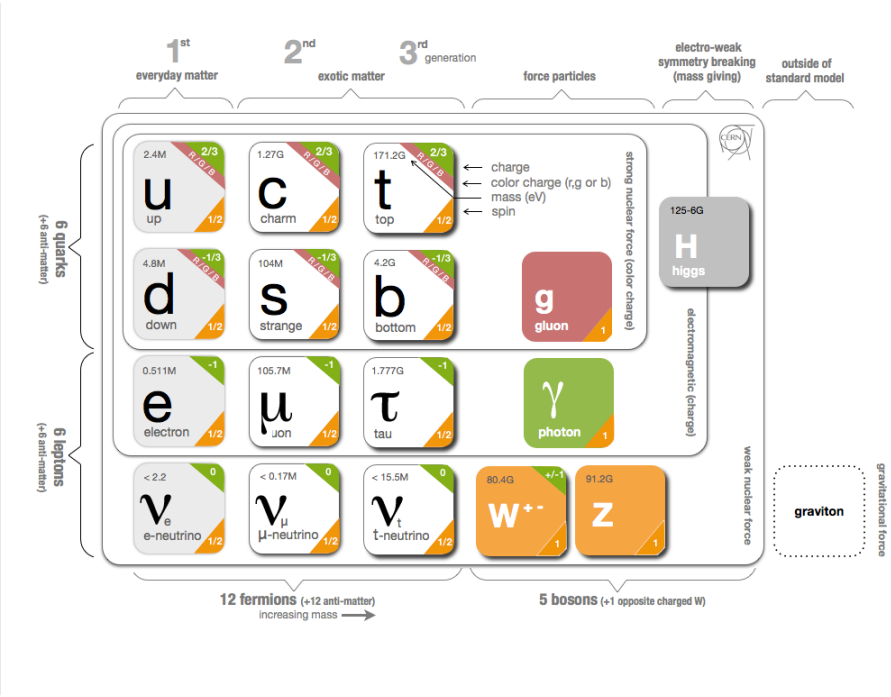


Figure 1.1: Infographic of the particles of the SM. Image source: [13].

In the SM, matter particles are described by half-integer spin *fermions*, while interactions are mediated by spin 1 *bosons*, which are the force-carrying particles. A summary of the main properties of the particles in the Standard Model can be found in Figure 1.1.

Fermions can be divided into three generations of leptons and quarks. Each fermion is associated with its anti-particle, that has the same mass, but opposite quantum numbers.

Each generation of leptons has an electrically charged lepton (e^-, μ^-, τ^-) and the respective neutrino (ν_e, ν_μ, ν_τ). Leptons, unlike quarks, cannot interact via the strong force. Furthermore, neutrinos can only interact weakly, since they do not carry electric charge.

Quarks carry a fractional electric charge and a strong colour charge. Gluons (g) are the mediators of the strong force [14] and are massless. Both gluons and quarks carry colour charge, the strong-interaction quantum number, that comes in

three colours: red, green and blue. Quarks are confined [15, 16, 17] in colour-singlets: mesons, composite states of a quark and an anti-quark, and baryons, composite states of three quarks or anti-quarks.

The $W^{\pm}$ and Z^0 bosons (often referred as W and Z) are the mediators of the weak force [18], while the photon (γ) is the mediator of the electromagnetic interaction.

Finally, in the SM, the Higgs mechanism generates the masses for the elementary (massive) particles.

1.2 Interactions in the Standard Model

Interactions are introduced in the Standard Model by requiring invariance of the Lagrangian under local gauge transformations in a symmetry group. Consider the free Dirac Lagrangian density:

$$\mathcal{L}_{Dirac} = \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi \tag{1.1}$$

where $\psi = \psi(x)$ is the Dirac spinor for a fermion (spin 1/2) of mass m, $\bar{\psi} = \psi^{\dagger}\gamma^0$ is the adjoint spinor, and γ^{μ} are the Dirac gamma matrices. If a $U(1)$ gauge transformation is applied, so that:

$$\psi \rightarrow \psi' = e^{i\alpha(x)}\psi \tag{1.2}$$

where $e^{i\alpha(x)}$ is a complex space-time dependent phase, the Dirac Lagrangian density acquires an additional term:

$$\mathcal{L}_{Dirac} \rightarrow \mathcal{L}'_{Dirac} = \mathcal{L}_{Dirac} - \bar{\psi}\gamma^{\mu}\partial_{\mu}\alpha(x) \tag{1.3}$$

In order to restore the $U(1)$ invariance of the Lagrangian, the partial derivative ∂_{μ} has to be replaced with the *covariant* derivative D_{μ}, with a precise transformation

relation under $U(1)$ when applied to spinor $\psi(x)$:

$$D_\mu \psi(x) \rightarrow D'_\mu \psi(x)' = e^{i\alpha(x)} D_\mu \psi(x) \tag{1.4}$$

It can be easily seen that Equation 1.4 is satisfied for:

$$D_\mu \psi(x) = \partial_\mu + ieA_\mu \tag{1.5}$$

with A_μ transforming under $U(1)$ as:

$$A_\mu \rightarrow A'_\mu = A_\mu - \frac{1}{e}\partial_\mu \alpha(x) \tag{1.6}$$

$U(1)$ is the symmetry group of quantum electrodynamics (QED), therefore in the case of QED, A_μ can be interpreted as the gauge field for the electromagnetic interaction (the photon), with interaction strength e. At this point, the full QED Lagrangian can be written as:

$$\mathcal{L}_{QED} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{\psi}(i\gamma^\mu D_\mu - m)\psi \tag{1.7}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the electromagnetic field tensor, $-\frac{1}{4} F_{\mu\nu} F^{\mu\nu}$ represents the kinetic term of the gauge field and the interaction between the gauge boson and the fermion is given by the term in D_μ.

1.2.1 The Electroweak Theory

The above discussion can be generalised to non-Abelian gauge groups[2], such as $SU(3)_C$ (the group of Quantum Chromodynamics - QCD), or the electroweak in-

[2]Non-Abelian groups are symmetry groups whose generators do not commute. For example, in the case of non-Abelian group $SU(2)_L$, its generators are Pauli matrices σ_i, for which the following commutation relation hold:

$$\left[\frac{\sigma_i}{2}; \frac{\sigma_j}{2}\right] = i\varepsilon_{ijk}\frac{\sigma_k}{2} \tag{1.8}$$

teraction, described by $SU(2)_L \times U(1)_Y$.

Let's discuss the $SU(2)$ group first. A local gauge $SU(2)$ transformation with generators σ^a (Pauli matrices) is defined as:

$$\psi \rightarrow \psi' = e^{i\alpha^a(x)\cdot\frac{\sigma^a}{2}} \tag{1.9}$$

In analogy to the $U(1)$ example, in order to preserve the invariance of the Lagrangian under a $SU(2)$ transformation, it is necessary to introduce *three* gauge fields, W^1_μ, W^2_μ, W^3_μ and their coupling g, with the appropriate transformation properties and the *covariant derivative* under $SU(2)$[3]:

$$D_\mu = \partial_\mu + ig\frac{\sigma_a}{2}W^a_\mu \tag{1.10}$$

so that the Lagrangian, containing the spinor term with the $SU(2)$ covariant derivative and the kinetic term for the W^a_μ $(a = 1, 2, 3)$ bosons, is invariant under a $SU(2)$ local gauge transformation (Equation 1.9).

The symmetry group of the weak interaction is $SU(2)_L$, where subscript L indicates that the weak charged-current interaction only couples with left-handed (LH) particle states (and right-handed (RH) antiparticle states). In this context, the LH particles and RH antiparticles are represented by weak isospin spinor doublets, with total "weak charge" (*weak isospin*) $I_W = \frac{1}{2}$ (and third component $I^3_W = \pm\frac{1}{2}$), while the RH particle and LH antiparticle states are placed in weak isospin singlets with $I_W = 0$ (and third component $I^3_W = 0$). As a consequence, the W^a_μ $(a = 1, 2, 3)$ gauge bosons only couple to (RH) LH (anti-) particle states. Weak charged-currents, corresponding

[3]The generalisation of the transformation rules for gauge fields W^1_μ, W^2_μ, W^3_μ requires the introduction of additional terms, due to the non-Abelian nature of the group. This is not done here for the sake of brevity, however, once the non-Abelian structure of the group is taken into account, the procedure is not *conceptually* different from the Abelian $U(1)$ case. A complete discussion, however, can be found in Chapter 4 of Ref. [12].

to the physical W bosons, can be obtained through linear combinations of these fields as:

$$W_\mu^\pm = \frac{1}{2}\left(W_\mu^1 \mp W_\mu^2\right) \tag{1.11}$$

As it has been said above, the $SU(2)_L$ symmetry implies that the W_μ^a $(a = 1, 2, 3)$ gauge bosons only couple to (RH) LH (anti-) particle states, and this would be true also for the purely-weak neutral-current (associated with gauge boson W_μ^3). However, experimental observations indicate that the Z boson couples to both left- and right-handed electrons.

Therefore, an additional symmetry group $U(1)_Y$ is introduced, with its gauge field B_μ, coupling g' and its "charge" (*weak hypercharge*):

$$Y = 2\left(Q - I_W^3\right) \tag{1.12}$$

where Q is the electromagnetic charge. The covariant derivative under $SU(2)_L \times U(1)_Y$ is then:

$$D_\mu = \partial_\mu + ig\frac{\sigma_a}{2}W_\mu^a + ig'\frac{Y}{2}B_\mu \tag{1.13}$$

In this model, the Glashow-Weinberg-Salam electroweak theory [19, 20, 21], the electromagnetic photon (A_μ) and the Z boson (Z_μ) result from a mixing of purely-weak boson W_μ^3, and the $U(1)_Y$ boson B_μ, trough the *weak mixing angle* (or *Weinberg angle*) θ_W $(\tan\theta_W = \frac{g'}{g})$:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} \tag{1.14}$$

and consequently [22]:

$$e = g \sin\theta_W \qquad e = g' \cos\theta_W \tag{1.15}$$

1.3 Spontaneous Symmetry Breaking and the Brout-Englert-Higgs mechanism

The description of the SM given so far requires that gauge bosons, as well as fermions, are massless, since adding particle masses in the Lagrangian would *explicitly* break local gauge symmetry that is at the core of the SM. However, this is manifestly in contrast with the experimental evidence, that shows that many particles have mass.

In the SM, masses arise from a mechanism that *spontaneously* breaks the symmetry: the Brout-Englert-Higgs mechanism [23, 24, 25, 26]. The generation of gauge boson masses is discussed first.

Consider the isospin doublet of complex scalar fields, with $I_W = \frac{1}{2}$ and $Y = 1$:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.16}$$

where ϕ^+ and ϕ^0 are a positively charged and a neutral field respectively and ϕ_1, ϕ_2, ϕ_3 and ϕ_4 real scalar fields.

The corresponding most general Lagrangian for the field can be written as:

$$\mathcal{L} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi) \tag{1.17}$$

where D_μ is the $SU(2)_L \times U(1)_Y$ covariant derivative in Equation 1.13. The first term in Equation 1.17 is the "kinetic" term for ϕ and $V(\phi)$ a potential (the Higgs potential), defined as:

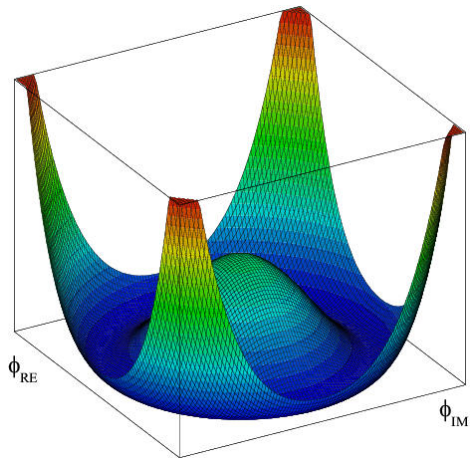


Figure 1.2: Graphic representation of the expected Higgs potential.

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda \left(\phi^\dagger \phi\right)^2 \tag{1.18}$$

where μ and λ are scalar constants. The vacuum state corresponds to the minima of the potential, found as:

$$\frac{\partial V}{\partial \left(\phi^\dagger \phi\right)} = -\mu^2 + 2\lambda \left(\phi^\dagger \phi\right) = 0 \tag{1.19}$$

therefore, *at the minimum of the potential*:

$$\left(\phi^\dagger \phi\right) = \frac{\mu^2}{2\lambda} = \frac{1}{2}\left(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2\right) \tag{1.20}$$

If $\mu^2 < 0$, Equation 1.20 has a unique minimum at $\phi_1 = \phi_2 = \phi_3 = \phi_4 = 0$. Conversely, for $\mu^2 > 0$ Equation 1.20 has real solutions (see Figure 1.2).

The solutions to Equation 1.20 are degenerate, since they are connected by an $SU(2)$ gauge transformation. The choice of the vacuum state breaks the symmetry of the Lagrangian, this is called *spontaneous symmetry breaking*. For example, one can choose $\phi_1^2 = \phi_2^2 = \phi_4^2 = 0$ and $\phi_3^2 = \frac{\mu^2}{\lambda} = v^2$. Expanding around this minimum and plugging this in Equation 1.16, one has[4]:

[4]This corresponds to a precise gauge choice, known as *unitary gauge*.

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.21}$$

where $h(x)$ is the Higgs field. Now substituting Equation 1.21 into Equation 1.17, in the unitary gauge one has:

$$\begin{aligned} \mathcal{L}_{Higgs} = & \frac{1}{2} \left(\partial_\mu h\right) \left(\partial^\mu h\right) + \frac{g^2}{4} \left(W_\mu^+ W^{\mu-}\right) \left(v + h\right) \\ & + \frac{1}{8} \left(g^2 + g'^2\right) Z_\mu Z^\mu \left(v + h\right) \\ & + \frac{\mu^2}{2} \left(v + h\right)^2 - \frac{\lambda}{4} \left(v + h\right)^4 \end{aligned} \tag{1.22}$$

where all the terms that are quadratic with respect to the fields: $W^\pm$, Z and h correspond to their mass terms:

$$m_W = \frac{1}{2} g v \qquad m_Z = \frac{1}{2} v \sqrt{g^2 + g'^2} = \frac{m_W}{\cos\theta_W} \qquad m_h = \sqrt{2\mu^2} = \sqrt{2\lambda v^2} \tag{1.23}$$

The gauge boson A_μ (the photon) does not appear here: it is not associated to any mass term and therefore remains massless.

1.3.1 Masses of the Fermions

The spontaneous symmetry breaking only results in the generation of mass terms for the electroweak bosons. Therefore, an extension of the SM is required, in order to give masses to the fermionic sector of the SM. An attractive feature of the Standard Model is that the same Higgs doublet which generates the $W^\pm$ and Z^0 masses is also sufficient to give masses to leptons and quarks.

The responsible mechanism is the Yukawa interaction, given by additional terms in the SM Lagrangian:

$$\mathcal{L}_{Yukawa} = -G_f \left(\bar{\psi}_L \phi \psi_R + \bar{\psi}_R \phi^\dagger \psi_L \right) \tag{1.24}$$

where G_f is the Yukawa coupling for a fermion f, $\psi_{L(R)}$ is the left (right) handed fermion isospin doublet (singlet) and ϕ is the complex scalar Higgs field. Examine here the case of the first generation of leptons:

$$\mathcal{L}_e = -\frac{G_e}{\sqrt{2}} \left[(\bar{\nu}_e, \bar{e})_L \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} e_R + \bar{e}_R \begin{pmatrix} 0 & v + h(x) \end{pmatrix} \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \right] = -\frac{G_e}{\sqrt{2}} \left[v + h(x) \right] (\bar{e}e) \tag{1.25}$$

the term purely proportional to $\bar{e}e$ is the *electron mass term*, so that:

$$m_e = \frac{G_e v}{\sqrt{2}} \tag{1.26}$$

and Equation 1.25 becomes:

$$\mathcal{L}_e = -m_e \bar{e}e - \frac{m_e}{v} \bar{e}e h(x) \tag{1.27}$$

with the second term being the interaction term between the electron and the Higgs field, that is also proportional to the electron mass. Note however that, since G_e is arbitrary, the actual mass of the electron is not predicted by the theory, but has to be determined experimentally. This formalism can equally be applied to the second and third generation of fermions, with a corresponding G_f term, required for each fermion flavour.

The quark masses are generated in the same way, but it is necessary to follow two steps. The first step consists in constructing a new Higgs doublet, in order to generate a mass for the upper member of a quark doublet, since the above formalism only gives masses to "down" type fermions. The additional term to the Yukawa Lagrangian is:

$$\mathcal{L}_{Yukawa,up} = -G_f \left(\bar{\psi}_L \widetilde{\phi}^c \psi_R + h.c. \right) \tag{1.28}$$

with $h.c.$ being the Hermitian conjugate and:

$$\widetilde{\phi}^c = \varepsilon_{2\times 2}\, \phi^* = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} = -\frac{1}{\sqrt{2}} \begin{pmatrix} v + h(x) \\ 0 \end{pmatrix} \tag{1.29}$$

This mechanism introduces the masses of "up" type fermions.

Finally, in the case of quarks, additional matrices need to be introduced, in order to eliminate extra - non diagonal - mass terms that would otherwise appear in the Lagrangian, and accounting for the difference between the weak-eigenstate basis and the mass-eigenstate basis. Once the additional matrices are introduced, it is possible to go from the mass-eigenstate basis (also referred to as physical basis) to the weak-eigenstate basis via the Cabibbo-Kobayashi-Maskawa (CKM) matrix [27].

The CKM matrix allows couplings between quarks in different weak isospin doublets through charged-current $W^{\pm}$ interactions, as observed experimentally.

1.4 Higgs phenomenology

In 2012, the ATLAS and CMS collaborations at CERN announced the observation a new particle in the mass region around 125 GeV [10, 11], consistent with the predictions for the SM Higgs boson. The discovery used the $H \to \gamma\gamma$, $H \to ZZ^* \to 4\ell$ and $H \to WW^* \to 2\ell\, 2\nu$[5] decay modes for their excellent experimental sensitivity.

In the SM the Higgs boson is a neutral scalar particle that interacts with fermions and gauge bosons trough a coupling constant that is proportional to their masses. For this reason, the Higgs boson is preferably produced in processes that involve

[5]Here ℓ indicates muons and electrons.

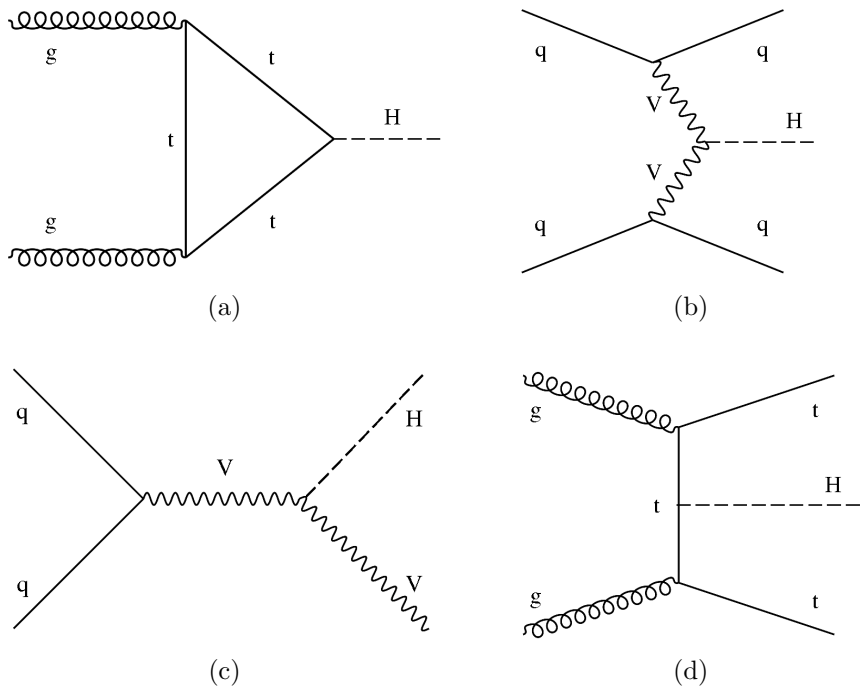


Figure 1.3: LO Feynman diagrams for the main Higgs boson production modes at the LHC: gluon-gluon fusion (1.3a), vector boson fusion (1.3b), associated production with a vector boson V (1.3c) and with a pair of top quarks (1.3d). Image source: [28].

heavy particles and it preferably decays into the more massive particles that are kinematically accessible. In the following, the Higgs boson production modes at the LHC and decays are discussed.

1.4.1 Higgs boson production at the LHC

At the LHC, four main mechanisms are studied for Higgs boson production in pp collisions (see Figure 1.3): *gluon-gluon fusion* (ggF), *vector-boson fusion* (VBF), the *associated production with a vector boson* $W^{\pm}$ or Z (VH or *Higgsstrahlung*), and the *production in association with top quarks* ($t\bar{t}H$). The cross sections for the individual production mechanisms depend on the energy in the centre-of-mass $\sqrt{s}$. The cross sections of the different production modes at $\sqrt{s} = 13$ TeV, shown in Figure 1.4a, is now briefly described.

Gluon-gluon fusion

At $\sqrt{s} = 13$ TeV the dominant production mode is gluon-gluon fusion, with a production cross section $\sigma_{ggF} = 48.6$ pb [29] for the SM Higgs boson[6]. This is due to the significant contribution from the gluon parton distributions (discussed in Section 1.5) [30] at these energy regimes. Since gluons are massless, the gluon-to-Higgs boson coupling happens via a loop of top quarks, while loops of lighter quarks are suppressed, as a consequence of the structure of the Yukawa couplings. Although ggF is the production mode with the highest cross section, when considering the $H \to b\bar{b}$ decay, it is experimentally extremely challenging to be studied, given the high contamination from non-resonant QCD processes in the final state.

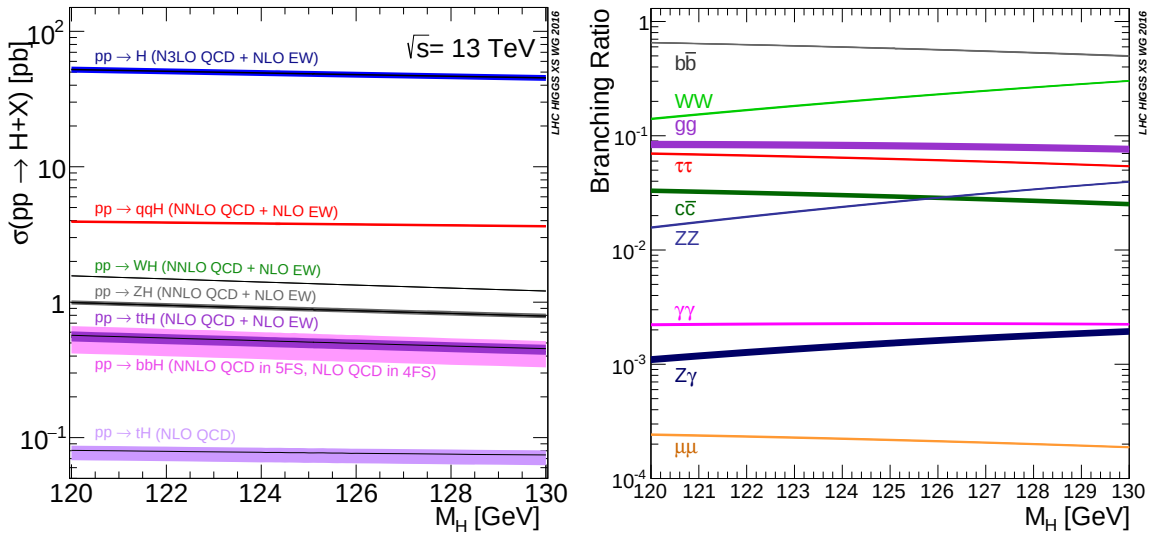
Vector-boson fusion

VBF production has the second-largest cross section: $\sigma_{VBF} = 3.77$ pb [29]. This process is characterised by the merging of two off-shell vector bosons produced in the scattering of two incoming quarks in the protons. VBF events have a clear experimental signature, since the final states are characterised by two forward jets, originated from the scattering of the two incoming quarks [31].

Associated production with a vector boson

In the VH production process, the Higgs boson is produced in association with a vector-boson W or Z. Although the cross section $\sigma_{VH} = 2.24$ pb [29] is $\sim$ 20 times lower than that of ggF production, this production mode offers a very clear experimental signature, when considering the leptonic decay of the vector boson V, so that it is possible to efficiently suppress backgrounds from non-resonant QCD processes. This production mode is the most sensitive for the study of the $H \to b\bar{b}$ decay, as it is discussed in Chapters 5 and 6.

[6] All the calculations presented in [29] assume $m_H = 125.09$ GeV.



(a) (b)

Figure 1.4: Standard Model Higgs boson production cross sections at $\sqrt{s} = 13$ TeV as a function of the Higgs boson mass (1.4a). Standard Model Higgs boson decay branching ratios as a function of the Higgs boson mass (1.4b). Image source: [29].

Associated production with a top quark pair

Another important production mode is $t\bar{t}H$, with cross section $\sigma_{t\bar{t}H} = 0.51$ pb [29]. The importance of this mode lies in the fact that it is possible to access directly to the top-Higgs Yukawa coupling.

1.4.2 Higgs boson decay modes

The predicted total width (Γ_H) of the SM Higgs boson is $\sim$ 4 MeV [32], and this implies a Higgs boson lifetime of O(10^{-22} s). For this reason, the Higgs boson can only be detected through its decay products.

It has already been seen that the Higgs boson's couplings to particles are proportional to the particles' masses. In the context of the decays, this means that the Higgs boson preferably decays to the heaviest particles that are kinematically allowed. Furthermore, as in the case of ggF production, loop diagrams allow the Higgs boson to "decay" to massless particles such as photons and gluons, via loops of top quarks

(dominantly).

Particle decays are usually described by a *branching ratio* (BR), that represents the probability for the parent particle P to decay into a certain final state X_i, calculated as:

$$\mathrm{BR}(P \to X_i) = \frac{\Gamma(P \to X_i)}{\Gamma_P} = \frac{\Gamma(P \to X_i)}{\sum_j \Gamma(P \to X_j)} \tag{1.30}$$

where $\Gamma(P \to X_i)$ and Γ_P are the *partial width* of the particle decay and the particle *total width* respectively.

The SM predictions for the Higgs boson decaying branching ratios as a function of the Higgs boson mass are shown in Figure 1.4b.

For a SM Higgs boson with a mass of 125 GeV, the dominant decay mode is $H \to b\bar{b}$, with a BR = 58.1 %. This decay is discussed in detail in this document, being the centre of the analyses described in Chapters 5 and 6.

The current measurements of Higgs boson production cross-sections, as well as the Higgs boson coupling values, are compatible with the Standard Model, as it can be seen in Figures 1.5 and 1.6 [33].

1.5 QCD phenomenology and cross section calculations at the LHC

In Section 1.1 quarks and gluons have been introduced along with mesons and baryons, i.e. hadrons. It has also been mentioned that both quarks and gluons can only be found as bound states in hadrons, and the process for which quarks and gluons (produced in the hard scattering, at the LHC) form hadrons is known as *hadronisation*.

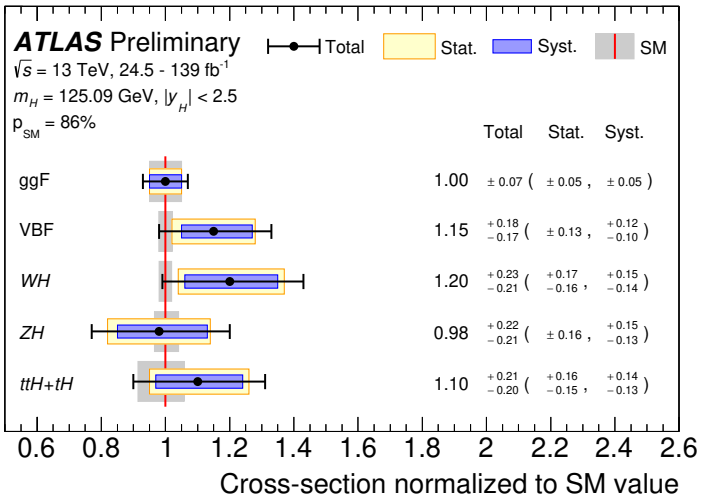


Figure 1.5: Cross sections for ggF, VBF, WH, ZH and $t\bar{t}H + tH$ normalised to their SM predictions, measured assuming SM values for the decay branching fractions. The black error bars, blue boxes and yellow boxes show the total, systematic, and statistical uncertainties in the measurements, respectively. The grey bands indicate the theory uncertainties in the SM cross-section predictions. In cases where several processes are combined, the combination assumes the relative fractions of each component to be as in the SM, within corresponding theory uncertainties. Cross sections are reported in the region of Higgs boson rapidity $|y_H| < 2.5$. Results are obtained in a simultaneous fit to the data, with the cross sections of each production mechanism as parameters of interest. Image source: [33].

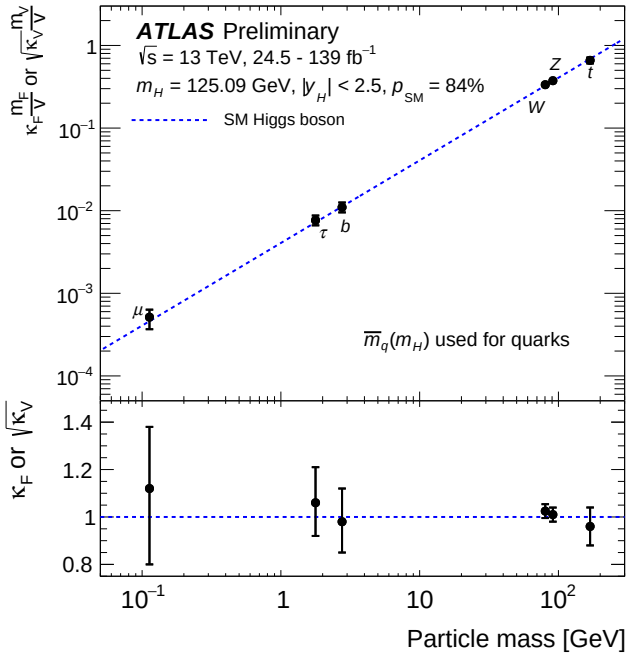


Figure 1.6: Reduced coupling-strength modifiers $\kappa_F \frac{m_F}{v}$ for fermions ($F = t, b, \tau, \mu$) and $\sqrt{\kappa_V}\frac{m_V}{v}$ for weak gauge bosons ($V = W, Z$) as a function of their masses m_F and m_V respectively, and the vacuum expectation value of the Higgs field $v = 246$ GeV. The results of the combination are obtained from a global likelihood fit to data with six degrees of freedom. The SM prediction for both cases is shown as a dotted line. Image source: [33].

The QCD coupling constant α_s has significant variations depending on the scale of energies at which it is measured. More specifically, it has large values for small energy scales (around 1 GeV), and decreases when the energy increases. At the typical energy scale of the LHC (> 100 GeV), α_s is sufficiently small to allow perturbative calculations to be used for quantitative predictions.

Even in this perturbative regimes, the convergence of the perturbative expansion requires higher-order corrections, whose *ultraviolet divergences* can be eliminated by introducing a redefinition of the fields, in order to absorb the divergent terms. This procedure, called *renormalisation*, requires the introduction of a *renormalisation* and a *factorisation scale*, μ_R and μ_F respectively [34, 35].

QCD calculations for processes at the LHC are challenging. In order to understand the particle production phenomenology at the LHC [36], it is necessary to introduce cross sections for hadronic collisions. The calculation of hadronic cross sections is based on the QCD *parton model*.

Consider now a high-energy pp collision, with a hadronic scattering taking place between partons (quarks or gluons) a and b, carrying fractions x_1 and x_2 of the momentum of the incoming protons respectively (initial state). The cross section $\sigma_{ab\to X}$ gives the probability that such initial state transitions into a final state X. The inclusive cross section for process $pp \to X$ can be expressed as:

$$\sigma_{pp\to X} = \sum_{a,b} \int_0^1 dx_1 \int_0^1 dx_2 f_a(x_1, \mu_F) f_b(x_2, \mu_F) \int d\sigma_{ab\to X}(x_1 P_1, x_2 P_2, \mu_R, \mu_F) \tag{1.31}$$

where the $f_i(x_i, \mu_F)$ are the *parton distribution functions* (PDFs) [37] and represent the probability for parton i to participate in the hard scattering interaction with momentum $p_i = x_i P_i$ (and P_i is the proton momentum). The PDF values depend on

the energy scale of the collisions, which are parametrised by the factorisation scale μ_F that can be interpreted as the scale that separates the long- and short-distance physics. These two scales can be separated thanks to factorisation theorems, allowing the non-perturbative QCD effects to be decoupled from the partonic cross-section.

1.5.1 Monte-Carlo simulations and uncertainties

In modern experimental particle physics, hence any ATLAS analysis, the physical processes as well as the particles' interaction with the detector are described using simulations. The interaction of particles with the ATLAS detector is briefly described in Section 3.1, while the *generation* of the physical processes is discussed in the following.

As shown in Equation 1.31, simulating a typical LHC process can require calculating multi-dimensional integrals. The most efficient way to compute such integrals is through random sampling techniques known as Monte Carlo methods (MC) [38].

Figure 1.7 shows the multiple steps of the event simulation:

Primary hard process generation

In the first stage of the simulation the hadronic cross section $\sigma_{pp\to X}$ of Equation 1.31 is calculated. This step contains the *matrix element* (ME) calculation (i.e. the calculation of $\sigma_{ab\to X}$ [40]), and involves PDFs, as well as the renormalisation and factorisation scales μ_R and μ_F.

Parton showers generation

Coloured quarks and gluons are created in the hard-scattering collision. These partons are too energetic to immediately form bound states and therefore produce a cascade of particles through the radiation of a gluon from a quark and via gluon splitting. This process is modelled with *parton shower* (PS) generators.

Simulation of the hadronisation

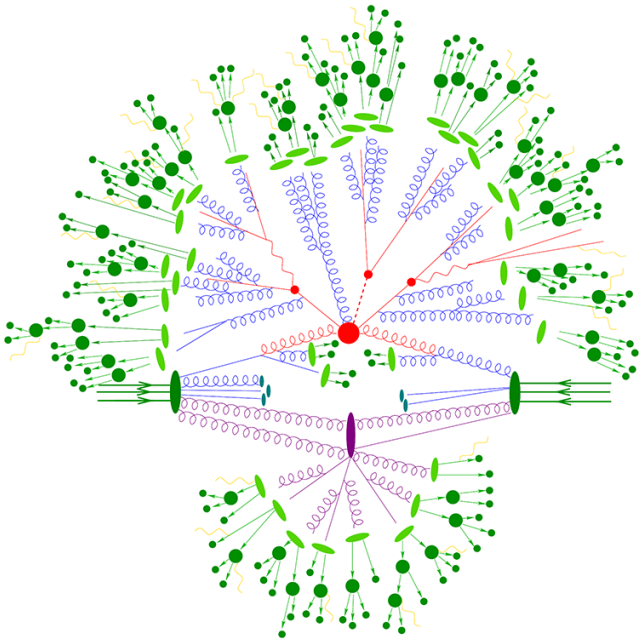


Figure 1.7: Schematic diagram of the main steps of an event generation for particle physics MC simulations. Different colours represent different steps: simulation of the hard process (red), parton showers (blue), hadrons (light green), hadrons' decay (dark green), simulation of the underlying events (purple). Image source: [39].

Once the evolution of the system reaches an energy in which the perturbative QCD regime can no longer be applied, phenomenological *hadronisation* models are used to convert the parton showers into outgoing hadrons [41].

Underlying event generation

The *underlying event* (UE) is the additional soft activity originating from the primary vertex, given by the partons that do not participate in the hard scatter.

Pile-up

Lastly, the simulation of pile-up collisions[7] is added on top of the event by overlaying simulated soft QCD events to the main simulated process, prior to the conversion of energy deposits to detector signals [42].

Each of these steps introduces specific assumptions and approximations that need

[7]Pile-up interactions result from the multiple proton-proton collisions occurring during a single LHC bunch crossing, as described in Section 2.1.4.

to be taken into account when working with simulations at the analysis stage. Typically, these effects are treated as different sources of *modelling uncertainties*, and are therefore part of the systematic uncertainties of physics analyses.

1.6 Limitations of the SM

Despite its great success, the Standard Model is not able to explain several anomalies in particle physics, indicating that the SM cannot be the complete picture. The main shortcomings are briefly explained in this section [43].

Neutrino masses

Despite being the most abundant particles just after photons, neutrinos remain the least understood ones within the SM. Indeed, several neutrino properties are still unknown, like their absolute mass, the number of species, or their exact nature. The SM predicts neutrinos as massless particles, however experiments [44, 45] have demonstrated the existence of neutrino oscillations, where neutrinos have been observed changing in one another, implying neutrinos as massive particles.

Dark Matter, Dark Energy

Several cosmological observations, such as those on galaxy rotations and gravitational lenses, suggest that a significant amount Dark Matter (DM) must exist in the Universe [27]. DM is estimated to compose up to $\sim 25\%$ of the universe, however its composition and the interactions with the SM particles remain unknown. The remaining 69% of the Universe energy, instead, consists of so-called Dark Energy, that appears to be associated with the vacuum in space. Its even distribution throughout the Universe is believed to lead to a repulsive force that tends to accelerate the expansion of the Universe. The rate of expansion and its acceleration can be measured by observations based on the Hubble law.

The hierarchy problem

One of the main arguments in support of an incomplete SM is the hierarchy problem, that highlights the discrepancy between the weak and gravitational energy scale [46]. The relatively small Higgs boson mass is a hint of *fine tuning* in the model, since first order corrections to the Higgs boson mass are found to be quadratically proportional to scale Λ, that is the cut-off energy scale at which the SM ceases to be valid. This introduces a divergence in the SM that prevents the SM to be valid up to the Plank scale. However, this divergence does not represent an issue if the HIggs boson mass corrections are of the same order as the mass itself, leading to an unnaturally *finely tuned* cancellation of terms. The hierarchy problem would require new physics at the TeV scale, for which DM is one the prominent candidates, or an extended theory, solving the divergence problem .

The strong CP problem

Another structural problem of the SM is the strong CP problem. Differently from weak interactions, strong interactions are observed to preserve P and CP symmetries, while theoretically their violation is not forbidden. The violation would arise as a non-zero parameter, θ_{QCD}. Experimental measurements of this parameter indicate $\theta_{QCD} < 10^{-9}$, and the reason why it takes this precise value is known as the strong CP problem. A proposed a solution to the strong CP problem [47] consists of introducing an additional broken global symmetry, introducing new light particles, known as *axions*.

This non-exaustive list includes some of the shortcomings of the SM, but it already evident how important it is to perform both direct searches for new physics BSM and indirect searches, thorugh precision measurements of the SM parameters, that could unveil potential discrepancies where new physics might be lying.

1.6.1 Effective Field Theories

Effective Field Theories (EFTs) are a general tool that can be used when phenomena are spread over different energy (length) scales. In these cases, an *effective* description can be valuable, to simplify calculations, or to allow a model-independent interpretation of physics results [48]. The EFT approach allows to describe the physics at a certain scale E, by approximating the more complex underlying physical theory that occurs at an energy scale $\Lambda \gg E$.

For example, the Fermi theory of weak interactions [49] is an EFT describing these interactions at energies below the W and Z masses [50]. Fermi's approach interprets the neutron decay $n \to p + e^- + \bar{\nu}$ as a point-like interaction among the four fermions. The energy scale (E) of the neutron decay is in the range of a few MeV, that is significantly smaller than the scale of the weak interaction $\Lambda \sim m_{W/Z} \sim 80/90$ GeV.

At the LHC, where the energy scale of processes is of a few 100 GeV, EFTs could shed light on BSM physics processes with much higher energy scale Λ ($\Lambda \gg 1$ TeV). Since this approach does not assume any specific BSM model, it is completely model-independent.

In the Standard Model Effective Field Theory (SMEFT), the EFT approach is used to expand the SM Lagrangian (of dimension $D = 4$) with additional operators, and can be written as:

$$\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \sum_i \frac{c_i^{(D)}}{\Lambda^{D-4}} \mathcal{O}_i^{(D)} \tag{1.32}$$

where $\mathcal{O}_i^{(D)}$ are dimension-D operators that are invariant under the $SU(3)_C \times SU(2)_L \times U(1)_Y$ symmetry of the SM and $c_i^{(D)}$ are the *Wilson coefficients* that parametrise the strength of the new interactions. In this framework, higher dimension operators are progressively suppressed by a larger power of the energy scale Λ, and therefore yield smaller and smaller effects.

In principle, however, operators of any dimension D can be added to the series, as long as they comply with the symmetries of the SM. Operators with dimension $D = 5$ violate the lepton number conservation [51], while dimension $D = 7$ operators violate the conservation of the baryon number minus lepton number $B - L$ [52], that are assumed to be conserved as a consequence of the SM symmetry. Operators with $D > 7$, instead, are further suppressed by the energy scale Λ.

Therefore, only operators with $D = 6$ $\mathcal{O}_i^{(6)}$ are considered in the SMEFT Lagrangian. At this order, 59 independent operators are involved in the effective Lagrangian [53], however only a restricted set affects Higgs boson physics.

Experimentally, the presence of BMS physics (and therefore the $\mathcal{O}_i^{(6)}$ operators that describe it) can affect the Higgs boson production cross-sections at high transverse momentum. Therefore, precise Higgs measurements at high-p_T can be used to constrain the SMEFT parameters, as in the case of the $VH(H \to b\bar{b})$ boosted analysis described in Chapter 5. In particular, the $VH(H \to b\bar{b})$ boosted analysis is sensitive to fourteen operators, shown in Table 5.9.

The EFT approach is particularly effective in the interpretation of fiducial measurements, such as those derived with the Simplified Template Cross Section framework, unfolded for the detector effects, that are described in detail in Chapter 5.

Chapter 2

The Large Hadron Collider and the ATLAS experiment

The work presented in this thesis has been conducted on data collected by the ATLAS experiment and originating from high energy proton-proton collisions delivered by the world's most powerful accelerator: the LHC. This chapter introduces the main features of the LHC and details the complex system of sub-detectors of the ATLAS experiment.

2.1 The Large Hadron Collider

The LHC [54], that began operations in 2008, is the largest and most powerful particle accelerator in the world. It is located at the European Organisation for Nuclear Research (CERN), by the Franco-Swiss border, just outside Geneva.

Constructed $\sim$ 100 m underground, in the tunnel that previously hosted the Large Electron-Positron Collider (LEP) [55], the LHC consists of a 27-kilometre ring (see Figure 2.1) where two high-energy particle beams travel extremely close to the speed of light, before they are made to collide. The beams travel in opposite directions in separate beam pipes - two tubes kept at ultrahigh vacuum - and are guided around



Figure 2.1: The LHC tunnel. The blue cylindrical structures contain the superconducting dipole magnets that bend the particle beams circulating inside the LHC. (September 2019, photo credits: MGF)

the accelerator ring by a strong magnetic field produced and maintained by superconducting electromagnets at a temperature of ~ 1.9 K. Particles circulating inside the LHC are accelerated by radiofrequency (RF) cavities located in a long straight section of the tunnel (see Figure 2.2).

2.1.1 The accelerator complex

The LHC is not the only accelerator at CERN, but a succession of machines (see Figure 2.3) accelerates the beams to increasingly higher energies [56]. Each machine boosts the energy of a beam of particles, before injecting the beam into the following machine in the sequence, up to the record energy of 6.5 TeV per beam. The machines in the complex can accelerate protons, as well as heavy ions, however, since the results reported in this thesis are obtained using data from proton-proton (pp) collisions, only the acceleration of proton beams will be described.

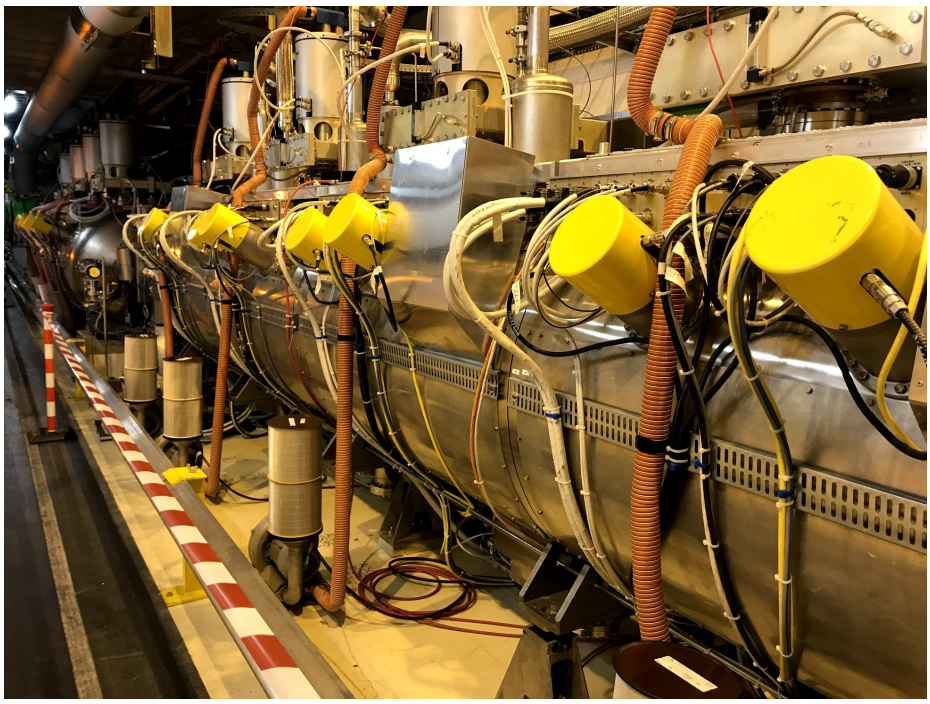


Figure 2.2: The LHC RF cavities. (September 2019, photo credits: MGF)

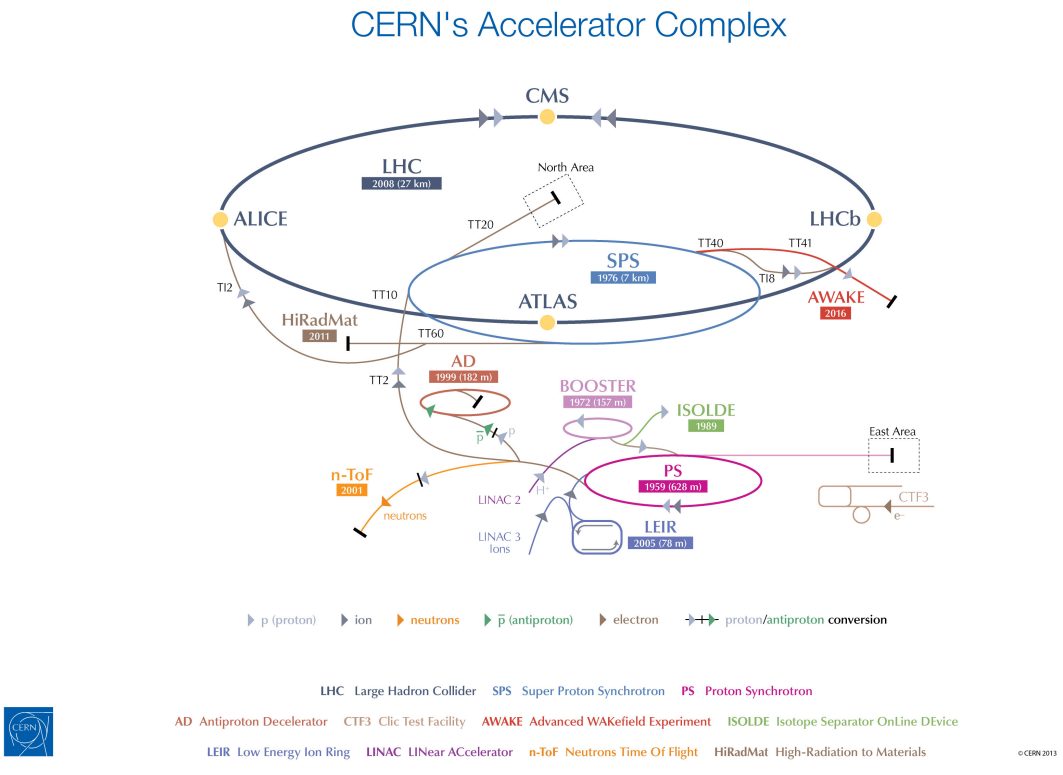


Figure 2.3: CERN's accelerator complex. Image source: [57].

The proton source is a simple tank of hydrogen gas that is ionised through an electric field, separating the atomic electrons from the protons. The first accelerator in the chain is Linac 2, that accelerates the protons to the energy of 50 MeV and then injects the beam into the Proton Synchrotron Booster (PSB), bringing the energy of the protons to 1.4 GeV. The booster is followed by the Proton Synchrotron (PS), which pushes the beam to 25 GeV and protons are then sent to the Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV.

The protons are finally transferred to the LHC, where they are split into the two separate beam pipes. It takes about 4 minutes to fill the LHC ring and 20 minutes for the protons to reach their maximum energy of 6.5 TeV (in Run 2). Under normal operating conditions, beams circulate for many hours inside the LHC beam pipes. The two beams are brought into collision inside four detectors - ALICE [58], ATLAS [59], CMS [60] and LHCb [61] - in bunches of $\sim 10^{11}$ protons, every 25 ns.

Three key-parameters of an accelerator, and the LHC specifically, are discussed in the following.

2.1.2 Centre-of-mass energy

The first key parameter of a collider is the delivered centre-of-mass energy, usually denoted as $\sqrt{s}$, using Mandelstam variable s. The centre-of-mass energy defines the accessible phase space for the production of resonances.

During the LHC Run 1, from 2009 to 2012, protons collided at $\sqrt{s} = 7$ TeV and 8 TeV. Long-shutdown 1 (LS1) followed in 2012-2015 and the superconducting beam pipe magnets were upgraded to reach $\sqrt{s} = 13$ TeV during Run 2, from 2015 to 2018.

The design centre-of-mass energy of $\sqrt{s} = 14$ TeV could be reached in Run 3, scheduled to begin in spring 2022.

2.1.3 Luminosity

The second key parameter of an accelerator is the delivered *luminosity*, that quantifies the rate of interactions per cross section (σ) of the particles in the colliding beams [8]. In other terms, for a given process, the number of expected interactions (N) is the product of the *integrated* luminosity (L) with the cross section σ of the process in question. The integrated luminosity is defined as the *instantaneous* luminosity over the time of operation of the machine:

$$N = \sigma L \tag{2.1}$$

$$L = \int \mathcal{L}\,(t)\,dt \tag{2.2}$$

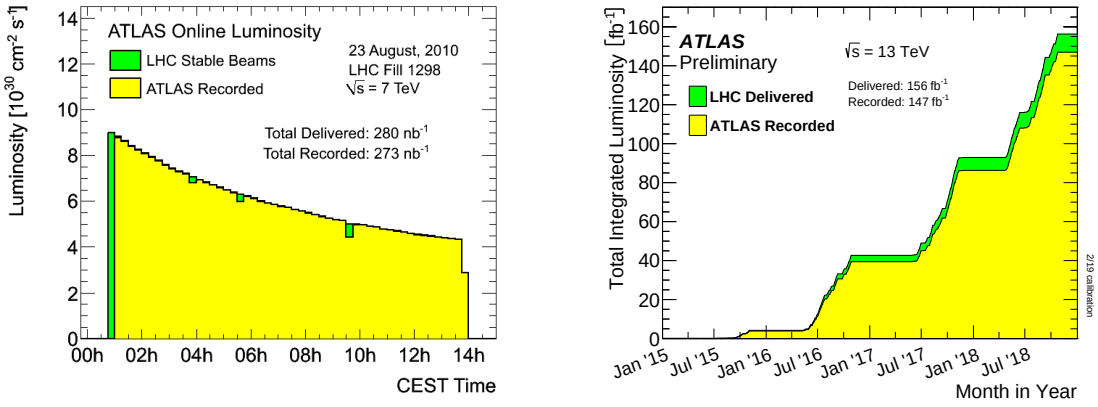
At a collider experiment, the instantaneous luminosity $\mathcal{L}$ can be calculated in terms of the number of particles in the colliding bunches n_1 and n_2, the collision frequency f and the root-mean-square beam sizes in the horizontal (σ_x) and vertical plane (σ_y), assuming a Gaussian beam profile and head-on collisions, as:

$$\mathcal{L} = f\frac{n_1 n_2}{4\pi\sigma_x\sigma_y} \tag{2.3}$$

The integrated luminosity is commonly measured in units of inverse cross section, thus inverse barns $[L] = \mathrm{b}^{-1}$, where 1 barn $= 10^{-24}$ cm^2. Consequently, the instantaneous luminosity is measured in $\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

After the start of collisions inside the LHC, the instantaneous luminosity decays exponentially as a consequence of the degradation of the circulating beams (see Figure 2.4a). Therefore, each LHC *fill* lasts for a limited period of time, usually about half a day, before the beams are dumped and a new fill is started. The peak instantaneous luminosity recorded by ATLAS was $2{\times}10^{34}$ $\mathrm{cm}^{-2}\mathrm{s}^{-1}$, twice the LHC design value.

Figure 2.4b shows the integrated luminosity collected during the LHC Run 2,



(a) (b)

Figure 2.4: ATLAS instantaneous luminosity profiles as measured online for LHC fills 1298 with 7 TeV centre-of-mass energy in 2010. Image source: [63] (2.4a). Integrated luminosity versus time delivered to (green) and recorded by (yellow) ATLAS during stable beams for pp collisions at $\sqrt{s}$ =13 TeV during the LHC Run 2. Image source: [64] (2.4b).

$L = (139.0 \pm 2.4)$ fb^{-1} [62].

2.1.4 Pile-up

The LHC high luminosity environment entails high collision rates, crucial for the experiments' scientific programme, which includes measurements and searches for new physics. Multiple pp collisions take place at every bunch-crossing, some of which are *interesting* hard-scattering events, contaminated by multiple soft QCD interactions.

These soft interactions form the so-called *pile-up* events, divided into *in-time* pile-up, when multiple collisions occur within the same bunch crossing, and *out-of-time* pile-up, generated by collisions of neighbouring bunch-crossings. Pile-up events are superimposed to the hard-scattering process, effectively degrading its reconstruction quality.

The average number of interactions per bunch crossing $\langle\mu\rangle$, used to describe this effect, is calculated as:

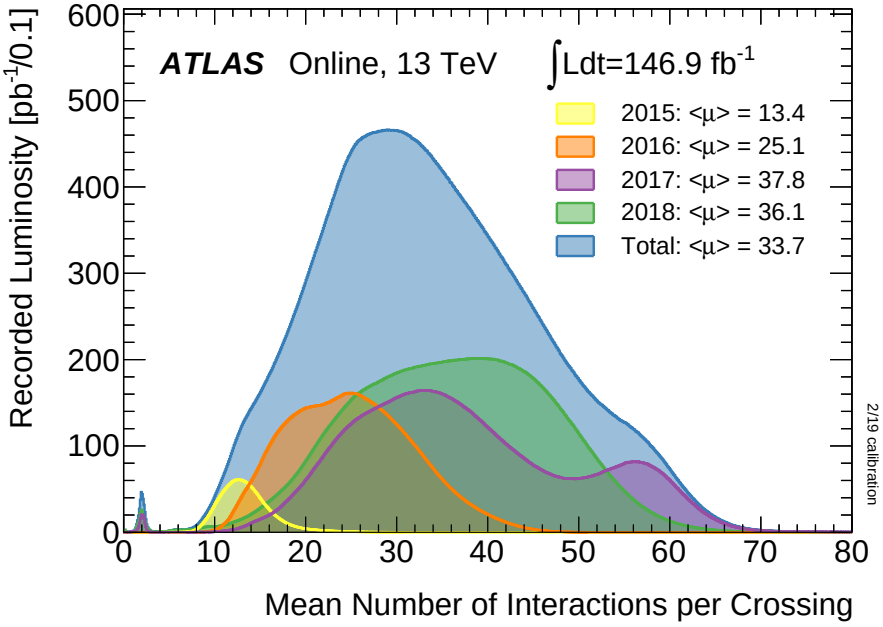


Figure 2.5: Luminosity-weighted distribution of the mean number of interactions per bunch crossing for the full Run 2 pp collision data at $\sqrt{s} = 13$ TeV. Image source: [64].

$$\langle \mu \rangle = \frac{\mathcal{L}_{bunch}\, \sigma_{inel}}{N_b\, f} \tag{2.4}$$

where $\mathcal{L}_{bunch}$ is the per-bunch instantaneous luminosity, σ_{inel} is the inelastic cross section for pp collisions, N_b is the number of bunches and f the collision frequency. The mean number of interactions per bunch crossing corresponds to the mean of the Poisson distribution of the number of interactions per bunch crossing calculated for each bunch.

The luminosity-weighted distribution of $\langle \mu \rangle$ for the Run 2 data taking is presented in Figure 2.5.

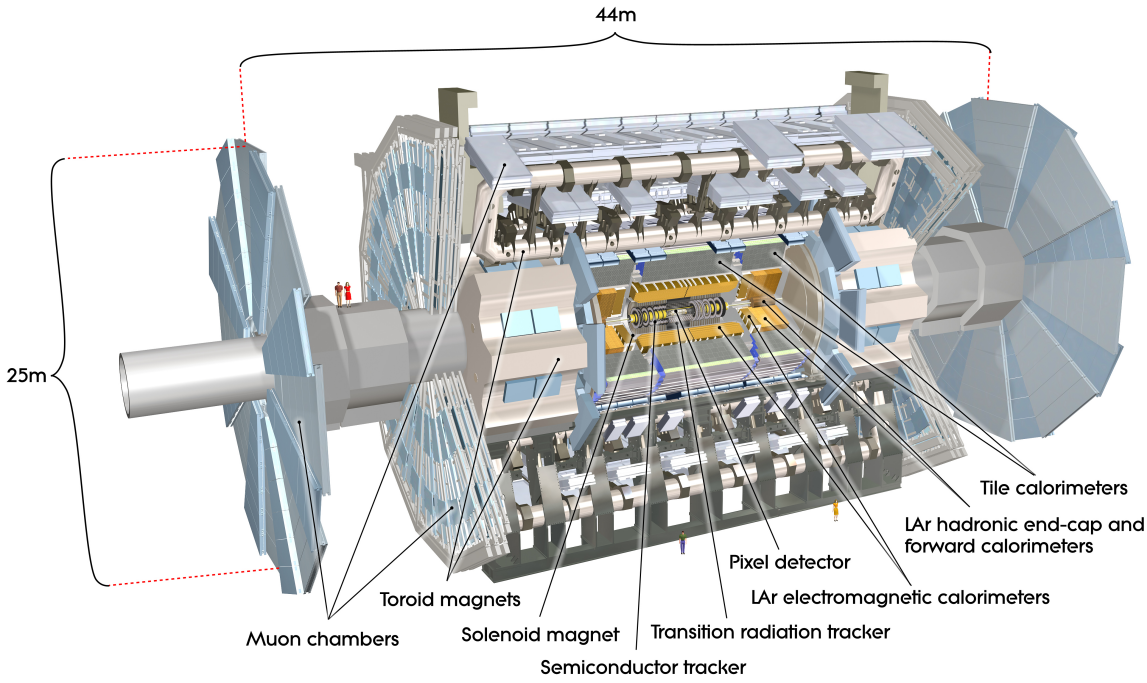


Figure 2.6: Schematic view of the ATLAS detector: the dimensions and main subsystems are indicated. Image source: [65].

2.2 The ATLAS detector

The ATLAS experiment (A Toroidal LHC ApparatuS) [59] is a general purpose detector devoted to the study of a wide range of physics processes, from precise SM measurements to BSM searches. ATLAS extends 46 m in length and 25 m in diameter, with a mass of 7000 tons (see Figure 2.6).

The detector has a forward–backward symmetric cylindrical geometry and a near 4π coverage in solid angle. It consists of an inner tracking detector surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic and hadron calorimeters, and a toroidal muon spectrometer.

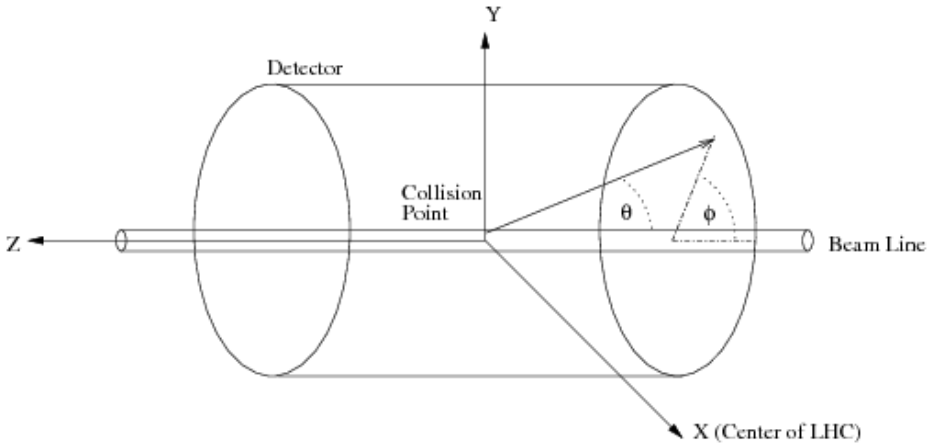


Figure 2.7: Illustration of ATLAS coordinate system. Image source: [66].

2.2.1 Coordinate System

As a reference frame, ATLAS uses a right-handed coordinate system, with its origin at the nominal interaction point in the centre of the detector, the z-axis along the beam pipe and the x–y plane transverse to the direction of the beam pipe (see Figure 2.7). The x-axis points from the interaction point to the centre of the LHC ring, and the y-axis points upward.

In the transverse plane cylindrical coordinates (r, ϕ) are used, with ϕ being the azimuthal angle around the z-axis and with respect to the x-axis, while the polar angle θ is defined in the y–z plane, with respect to the z-axis. The polar angle θ is often replaced by the *pseudorapidity* η, defined as:

$$\eta = -\ln \tan \frac{\theta}{2} \tag{2.5}$$

used at ultra-relativistic regimes ($p \gg m$ and $\theta \gg 1/\gamma_{Lorentz}$) as an approximation for the *rapidity* y:

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \tag{2.6}$$

where E is a particle energy and p_z is the component of its momentum along the

z-axis.

Rapidity y (as well as pseudorapidity η in the approximation mentioned above) is particularly advantageous to use since its intervals are invariant under a Lorentz-boost in the longitudinal (z) direction. Therefore, the angular separation ΔR between two objects (1, 2) can be conveniently defined in the η–ϕ space as:

$$\Delta R(1,2) = \sqrt{\Delta\phi_{12}^2 + \Delta\eta_{12}^2} = \sqrt{(\phi_1 - \phi_2)^2 + (\eta_1 - \eta_2)^2} \tag{2.7}$$

Finally, an extremely useful variable for the description of ATLAS events is the transverse momentum $p_T = p \sin\theta$, of particular interest at hadron colliders such as the LHC, since the momentum component along the z-axis is effectively unknown.

2.2.2 Magnet system

Magnetic fields are of fundamental importance for experiments at colliders, since they are used for the measurement of the momentum of the charged particles produced in the interactions. Charged particles bend inside a magnetic field, as a consequence of the Lorentz force, that is proportional to the component of the magnetic field perpendicular to the trajectory of the particle. The curvature radius is proportional to the particle momentum directly and inversely to the applied magnetic field, therefore an intense magnetic field is necessary for an accurate momentum measurement of high-p_T particles.

In the case of the ATLAS experiment, the magnetic field is created by a complex magnet system (see Figure 2.8) [67], consisting of four superconducting magnets, in a configuration aimed at restricting the magnetic field to the Inner Detector and the Muon Spectrometer.

The Central Solenoid provides a 2 T magnetic field along the z axis for the inner tracker, deflecting the particles along the ϕ coordinate. Three more magnets are

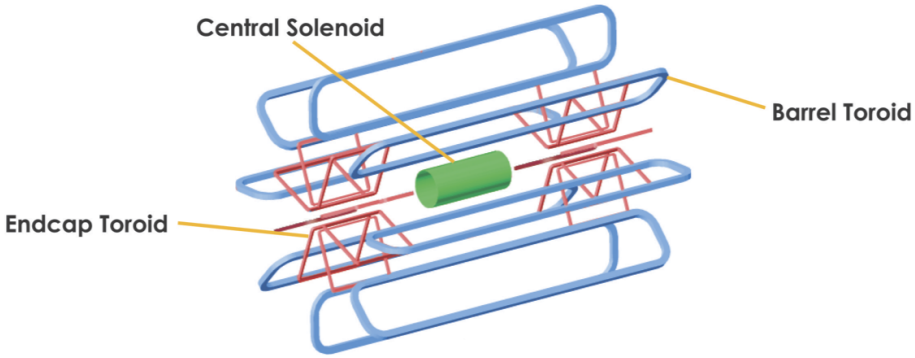


Figure 2.8: Illustration of the ATLAS magnet system. Image source: [67].

placed in the outer portion of the ATLAS detector, aiming at the deflection of muon trajectories in the η direction, for a precise measurement of their momentum. Such magnets are the majestic eight-coil Barrel Toroid and the two end-cap Toroids, for *central* and *forward* muons respectively.

2.2.3 Inner Detector

The ATLAS Inner Detector (ID) [6, 68] is responsible for the reconstruction of the tracks of charged particles, as well as the identification of interaction vertices. It is the closest system to the beam pipe and is surrounded by the calorimeter system. It extends from ~ 3 cm to ~ 1 m in radius, with a pseudorapidity coverage of $|\eta| \leq 2.5$. It consists of three distinct sub-detectors (see Figure 2.9): the Pixel Detector, the Semiconductor Tracker and the Transition Radiation Tracker.

Pixel Detector

The innermost sub-system of the ID is composed of silicon pixel detectors.

Its original layout comprised three barrel layers located at radii of 50.5, 88.5, and 122.5 mm with respect to the beam pipe, and two end-caps with 3 disc layers each at longitudinal distances of 49.5, 58.0 and 65.0 mm. In total 1744 modules read out 80 million pixels, with a typical pixel size of 50×400 μm^2 and sensor thickness of 250

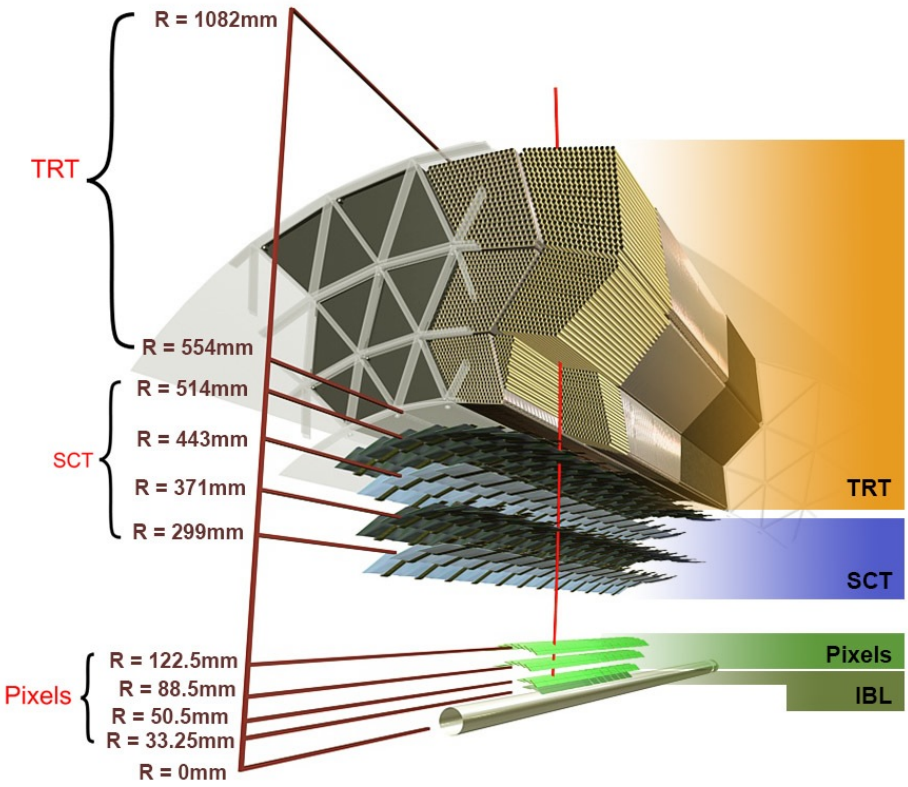


Figure 2.9: Illustration of the ATLAS ID. Image source: [69].

μm.

During the shut-down period before Run 2 (LS1) the pixel detector was improved, thanks to the installation of the Insertable B-Layer (IBL), a fourth pixel layer placed at radius 33.25 mm. The IBL added 12 million pixels, with size of 50 $\times$ 250 μm^2, enhancing the performance of the ID with respect to Run 1 in terms of vertex reconstruction and secondary vertex identification, crucial for the identification of b-hadrons.

The current resolution provided by the pixel detector is $\sim 10\mu$m in the $r-\phi$ plane and $\sim 67\mu$m in the z direction [70, 71].

SemiConductor Tracker

The second sub-system in the ID is the SemiConductor Tracker (SCT). It is composed of silicon microstrip modules [72] assembled in four concentric barrel layers and two

end-caps of nine disks each. It extends in radius from ~ 30 cm to ~ 51 cm. The barrel sector contains 2112 modules, whilst the end-caps are composed of 988 modules, all mapped to more than 6 million read-out channels.

Differently from pixel detectors, a single layer of strip detectors would only provide measurements along one spacial direction (ϕ). The measurement of the z coordinate is provided by a second layer of strips, glued back-to-back with the first one, at a *stereo* angle of 40 mrad. Each barrel layer and end-cap disk is composed of *stereo* strips in this configuration.

The SCT provides position measurements with a resolution of 17 μm in the $r-\phi$ plane and 580 μm along the z direction [73].

Transition Radiation Tracker

The final and most external sub-system in the ID is the Transition Radiation Tracker (TRT) [74], that is used both for track reconstruction and particle identification.

The TRT extends from ~ 55 cm up to ~ 1 m from the beam pipe and it a gaseous tracker, with 73 layers of cylindrical drift tubes (straws) of 4 mm in diameter, used as drift tubes for track position measurements. Each straw tube contains a gold-plated tungsten anode wire and it is filled with a gas mixture based on Xenon.

The space in between the straw tubes is filled with polyethylene, a plastic material, to induce transition radiation for charged particles, crucial for the particle identification, especially for the discrimination between electrons and charged pions in the energy range between 1 GeV and 100 GeV. When a relativistic charged particle traverses the boundary between two materials, it emits radiation with intensity that depends on its Lorentz boost $\gamma \propto E/m$, allowing for particle identification, for particles of the same momentum.

The TRT operates as a drift chamber: when a charged particle traverses the straw, it ionises the gas. The electrons drift toward the anode wire and cascade in the strong

electric field, producing a detectable signal. On each wire the signal is amplified, shaped and discriminated against two adjustable thresholds. The two thresholds allow for the simultaneous measurement of tracking information and identification of characteristic large energy deposits[1] due to the TR photons, efficiently absorbed by the Xenon-based gas mixture.

The TRT achieves a position resolution of 130 μm in the $r-\phi$ plane ($z-\phi$ in the end-caps).

As discussed in the previous section, the ID is immersed in a 2 T magnetic field generated by the central solenoid. By measuring the curvature of charged particle trajectories inside the magnetic field, the momentum of such particles is inferred. The momentum resolution σ_{p_T} for charged particle momenta p_T in the ATLAS ID is then [59]:

$$\frac{\sigma_{p_T}}{p_T} = 0.05\% \, p_T \, [GeV] \oplus 1\% \tag{2.8}$$

where symbol $\oplus$ indicates the sum in quadrature.

2.2.4 Calorimetry

The ATLAS calorimeters are located just outside the central solenoid. Their role is to accurately measure the energy of neutral and charged particles produced in the collisions, as well as the missing transverse momentum of events.

The calorimeters are hermetic along the ϕ coordinate and cover a pseudorapidity region up to $|\eta| < 4.9$. All calorimeters are *sampling* calorimeters consisting of multiple layers of *passive* and *active* materials. The passive layers are composed of dense

[1] The large energy deposits from transition radiation are measured as electrons generated via photoelectric effect by the TR photons on the anode wires.

materials with a high atomic number, used to generate a cascade of particles (*showers*) from the incoming particle, whilst the active layers measure the energy deposited by the showers, as ionisation energy and scintillation light. The particles' energy is then measured from the total energy of the electromagnetic and hadronic showers they originate, therefore calorimeters require large volumes, in order to contain fully the showers.

Calorimeters are essentially of two different types: *electromagnetic* calorimeters, measuring the energy of electrons and photons, and *hadronic* calorimeters, measuring the energy of hadronic showers (in combination with the energy deposits in the electromagnetic calorimeter).

The energy resolution σ_E of an electromagnetic calorimeter for a particle of energy E can be parametrised as [75]:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \tag{2.9}$$

The first term $a/\sqrt{E}$ is the *stochastic term* accounting for any Poisson statistics-related fluctuations, such as the intrinsic shower fluctuations and sampling fluctuations; the second term b/E is a *noise term*, accounting for the electronic noise in the read-out chain; and finally a *constant term* c incorporating instrumental effects, such as inefficiencies in the signal collection, or mechanical non-uniformities. The relative importance of the various terms depends on the energy of the incident particle.

The ATLAS full calorimetry system (shown in Figure 2.10) is rather complex and consists of five different sub-systems.

LAr electromagnetic barrel

The innermost layer of the calorimeter system is represented by the barrel electromagnetic calorimeter [77, 78]. Lead, arranged in an accordion structure, is used as the passive material and Liquid Argon (LAr) as the active material. The ionisation

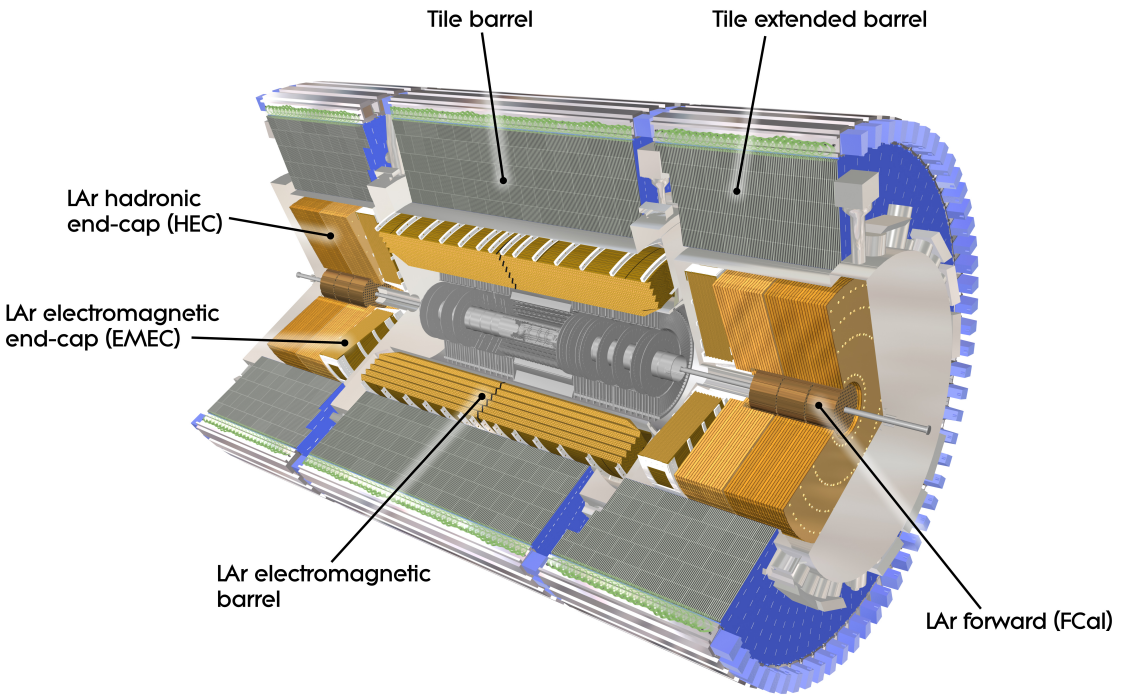


Figure 2.10: Cut-away view of the ATLAS calorimeter system. Image source: [76].

produced by the showers is measured on copper electrodes, arranged in a honeycomb structure. It is composed of three layers from 1.41 m to 1.96 m from the beam pipe, to observe the longitudinal development of the showers and determine their direction, and covers a pseudo-rapidity region of $|\eta| < 1.475$.

High-energy photons ($E_\gamma \gg 1$ MeV) interact with matter mainly through pair production and high-energy electrons through bremsstrahlung. The development of showers is parametrised in terms of the *radiation length* X_0[2], that is defined as the mean length to reduce the energy of an electron by a factor $1/e$ due to bremsstrahlung and, on a first approximation, the mean length for a photon to produce an electron-positron pair. The total thickness of the barrel electromagnetic calorimeter is about $22X_0$.

A single layer of instrumented liquid argon, the *pre-sampler*, is placed before the

[2]The radiation length can be approximated as $X_0 \sim 180\, A/Z^2$ [g/cm^2].

barrel electromagnetic calorimeter, to correct for the energy loss in the ID and the solenoid magnet.

LAr electromagnetic end-cap

The electromagnetic end-cap calorimeter (EMEC) is built using the same technology as the barrel LAr calorimeter described above and is composed of two coaxial wheels at each side of the barrel calorimeter, covering the pseudorapidity region $1.375 < |\eta| < 3.2$. Its radiation thickness is more than $24X_0$.

The energy resolution of the electromagnetic calorimeter (for the barrel and end-cap components inclusively) can be expressed as [59]:

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\% \tag{2.10}$$

where energy E is measured in GeV. The noise term (b/E in Equation 2.9) is found to be negligible [77].

TILE barrel

Outside the electromagnetic calorimeters there is the barrel hadronic tile calorimeter [79], extending from radius of 2.28 m to 4.25 m, covering $|\eta| < 1.0$, with two *extension* regions at $0.8 < |\eta| < 1.7$. The tile calorimeter is constructed alternating tiles of a steel (passive) absorber and a (active) plastic scintillator, read out by wavelength-shifting fibre optics connected to photo-multiplier tubes.

Similarly to electrons and photons, hadrons interact with matter creating a shower. The shower is however *hadronic* and is formed when hadrons deposit their energies through ionisation, as well as collisions mediated by the strong interaction. The development of hadronic showers in matter is characterised by the *nuclear interaction*

length λ defined, in analogy to the radiation length X_0 for electromagnetic showers, as the mean distance travelled by a hadron before interacting with a nucleus of the absorber. For a given material $\lambda \gg X_0$, hence why hadronic calorimeters are generally larger than electromagnetic calorimeters. The depth of tile hadronic calorimeter is $\sim 7\lambda$.

LAr hadronic end-cap

The LAr hadronic end-cap calorimeter [80] (HEC) is a sampling calorimeter using LAr as the active medium and copper plates as absorbers. It is located right behind the EMEC and consists of two independent wheels per side, with a pseudorapidity coverage of $1.5 < |\eta| < 3.2$.

Both the barrel TILE hadronic calorimeter and the LAr hadronic end-cap calorimeter yield an energy resolution of [59]:

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\% \tag{2.11}$$

LAr forward calorimeter

The final and most forward calorimetry sub-system is represented by the LAr forward calorimeter (FCal) [81, 82]. It is placed in a high-pseudorapidity region $3.1 < |\eta| < 4.9$, close to the beam pipe, where the particle density is extremely high.

It is composed of three layers per side and using LAr as the active material. In terms of passive absorber, the first layer is made of copper, while the other two are made of tungsten. Its depth is quantified as 10λ, and has a resolution of [59]:

$$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 10\% \tag{2.12}$$

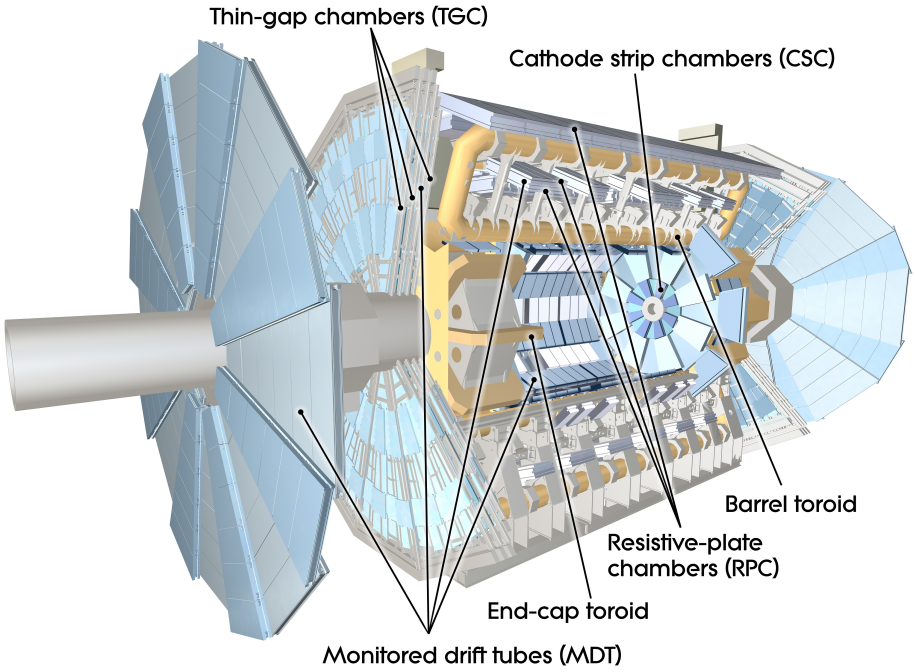


Figure 2.11: Overview of the ATLAS muon system. Image source: [83].

2.2.5 Muon Spectrometer

The Muon Spectrometer (MS) [84] is the outermost sub-system of the ATLAS detector and is designed to measure the trajectory and momentum of the charged particles escaping the calorimeters, i.e. muons. Parts of the MS are designed for triggering.

The MS extends from a radius of ∼5 m, to ∼10 m, and has a pseudo-rapidity coverage of $|\eta| < 1.5$ in the barrel region and up to $|\eta| < 2.7$ in the end-cap portions.

The MS performance heavily relies on the magnetic system described in Section 2.2.2 and measures the muon momenta with a resolution that varies from 2-3% at low-p_T (2 GeV $< p_T <$ 250 GeV), to 10%, for high-p_T muons ($\sim$ 1 TeV).

The MS is instrumented with two types of chambers (see Figure 2.11), the first one being tailored for high-precision tracking and the second one for fast triggering.

MS precision tracking chambers

The ATLAS chambers providing high-precision measurements for the muon momenta are of two types: the Monitored Drift Tubes (MDT) and the Cathode Strip Chambers (CSC).

The MDT [85] are used for high precision tracking in the barrel and end-cap region, covering up to $|\eta| < 2.7$ in pseudorapidity ($|\eta| < 2.0$ in the innermost end-cap layer). They are arranged in several layers, yielding a total spacial resolution of 35 μm.

The high occupancy region at $2.0 < |\eta| < 2.7$ is additionally covered by the CSC [86], that are multi wire proportional chambers, with cathodes segmented into strips. Their resolution is 60 μm along the r coordinate and 5 mm in the ϕ direction.

MS triggering chambers

The muon trigger system is given by two different sub-system: one covering the barrel region and the other serving the end-caps. Both sub-systems are designed to provide fast trigger signals for the traversing muons.

The barrel muon trigger is provided by three layers of Resistive Plate Chambers (RPC) [87], while the end-cap trigger system is given by very thin multi-wire chambers, the Thin Gap Chambers (TGC) [88].

2.2.6 Trigger and Data Acquisition System

The LHC delivers collisions with a time spacing of 25 ns, resulting in a rate of 40 MHz. The ATLAS experiment cannot record collisions at a similar rate, therefore it is necessary to use a Trigger and Data Acquisition System (TDAQ), to record only the collisions that are *interesting* for physics analyses.

The ATLAS trigger system [89] is designed to reject non-interesting data, dominated by low-p_T inelastic and diffractive collisions, while maintaining a high efficiency

in selecting the relevant physics processes, reducing the recorded event rate to 1 kHz.

The trigger system is built on two main levels of selection: a first hardware level-1 (L1) trigger and the final software-based high-level trigger (HLT).

L1

The L1 trigger is at hardware level and exploits quickly accessible coarse data from the calorimeters (L1Calo) and the MS (L1Muon), to identify high-p_T objects in dedicated $\eta - \phi$ regions of the detector: *Regions of Interest* (RoI). Calorimeters provide information on the clusters of energy deposits, as well as the missing transverse energy, while the muon spectrometer provides information on the transverse momentum (and position) of muons. The information on the RoIs is also stored. The L1 trigger reduces the event rate to 100 kHz.

HLT

The RoIs are then transferred to the HLT, to accelerate the reconstruction. The software integrates the RoI data with the full detector information from the sub-detectors, to perform a full reconstruction of the event and then run the trigger algorithms to finally select or reject the event. At this level more complex signatures can be determined, including events containing b-jets, hadronic taus and other combinations of physics objects.

A *trigger menu* defines the selection criteria for L1 and HLT, and the chosen trigger configuration depends on the final purpose for which the data is used, whether for physics analyses, efficiency and performance measurements, or detector calibration.

Chapter 3

Object reconstruction with the ATLAS experiment

Once the *pp* collisions take place at the core of the ATLAS detector, the particles generated by the collisions travel inside the detector volume[1], leaving different signatures in each of the sub-systems described in the previous Chapter, as shown in Figure 3.1.

The electronic signals produced inside the sub-detectors are reconstructed as *physics objects*, corresponding to the final-state particles and observables measured by ATLAS. The *reconstruction* procedure is divided into several steps, and allows the conversion from *detector-level* objects, such as tracks and calorimeter clusters, to *high-level physics* objects identified as particles (i.e.: electrons, muons, jets, etc).

Finally, objects are *calibrated*, to correct for detector effects. Calibrated objects are usually provided to the physics analyses following dedicated studies within the ATLAS Collaboration, however further optimisation can be done *ad hoc* at analysis level to further improve the selection and maximise the analysis sensitivity.

This chapter illustrates the main steps of the ATLAS simulation infrastructure

[1]Particles may also decay and their decay products travel inside the detector.

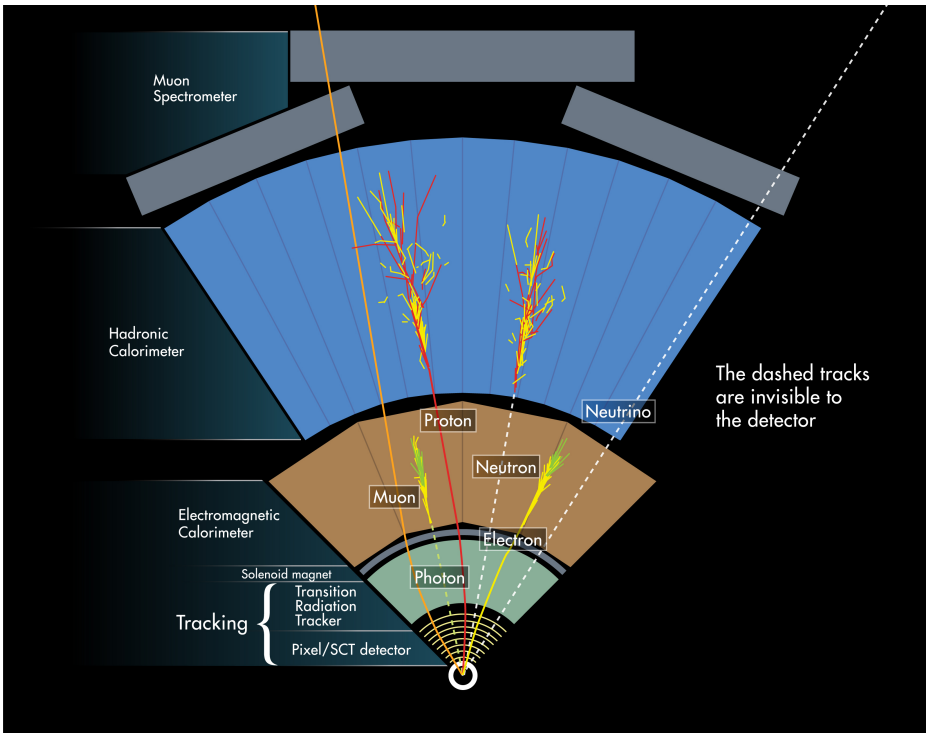


Figure 3.1: View of the particle paths and experimental signatures in the transverse plane of the ATLAS detector. Image source: [90].

and then describes the reconstruction algorithms and techniques characterising the physics objects, with a focus on the ones that are particularly relevant for the b-tagging studies and the $VH(H \to b\bar{b})$ analyses described in this thesis.

3.1 ATLAS Simulation Infrastructure

The importance of detailed simulation for the physics programme of the LHC experiments has been introduced in Section 1.5.1, where the role and main features of MC event generators have been discussed. The full ATLAS simulation chain [91] can be schematically divided into four fundamental steps:

Event generation This step, described in Section 1.5.1, involves the calculation of the matrix element for the hard scattering process, and also takes care of the parton shower and hadronisation simulation, thanks to the usage of dedicated

software. The information (called *truth*) of any stable particle propagating through the detector is stored as output.

Detector simulation Stable particles are propagated through the Geant4 [92] ATLAS detector simulation, where the detector is reproduced realistically, based on its material, geometry and response. The simulation is performed in steps, and the energy deposited in the sensitive portions of the detector is recorded as *hits*, containing the energy deposition, as well as the space-time coordinates. Truth-level information is stored for each step.

Detector simulation is the most computationally demanding task in the chain, particularly at the calorimetry stage. For this reason, a *fast* alternative (AFII) [93] to the full simulation (FullSim) chain is available and can be used by the Collaboration.

Digitisation The hits recorded in the active materials are converted into digits, used to mimic the output of the detector readout and to produce a *raw data object* file.

Reconstruction Finally, the detector response is reconstructed into physical objects using various algorithms that will be described in the following. The result of the reconstruction is stored in a special format, called Analysis Object Data (AOD), that can be read by analysis frameworks. The truth information is processed up to the reconstruction step, so as to be used by analysers for studies at generator level.

The ATLAS simulation chain, from the detector simulation and up to the physics object reconstruction is controlled by the Athena software [94].

3.2 Detector objects

In this section the key ingredients to build the high-level physics objects are discussed. When particles travel inside the detector, they usually leave small energy deposits in the tracking detectors (the ID and the muon chambers) or a significant amount of energy in the calorimeters. The two key ingredients for particle reconstruction algorithms are therefore *tracks* and calorimeter energy clusters, or *topological clusters* [95].

3.2.1 Tracks

The tracks of charged particles are reconstructed using the energy deposits they leave - via excitation and ionisation - when traversing the detector elements. Starting from the ID [96], the energy deposits are clustered together in points called *hits*, and then the track reconstruction proceeds using an *inside-out* strategy [97] based on a Kalman filter [98], starting from three hits in the silicon layer of the ID, and then proceeding outwards by adding successive hits compatible with the seed. This procedure allows to assign single hits to multiple tracks, resulting in possible ambiguities, that are solved by ranking tracks according to their *quality*. Additionally, a χ^2 fit is performed on the hits, so that poor-quality candidates are penalised. Finally, an *outside-in* reverse approach is used to combine the TRT hits with the tracks reconstructed in the silicon ID.

As a final step, the reconstructed tracks are kept in the final selection if they satisfy general quality requirements, on the number of hits in the silicon detectors (≥ 7), the number of *holes* (missing hits) in the pixel detector (≤ 1), the number of shared hits with another track (≤ 1), as well as kinematic requirements: $p_T > 500$ MeV and $|\eta| < 2.5$.

Primary vertex reconstruction

The reconstruction of the *primary vertex* [99, 100], where the pp hard scattering collision took place, is of paramount importance for physics analyses and is the result of an iterative procedure.

Once tracks have been computed, vertices are calculated as the intersection of two or more tracks. Firstly, a vertex seed position is identified from the reconstructed "high-quality" tracks. Then, tracks that are not compatible with the vertex position are removed and the vertex position is recomputed, until the vertex position passes quality requirements and is therefore fixed. The process is then repeated with the discarded tracks and other vertices are computed. Multiple vertices are reconstructed for a single bunch crossing, due to the high pile-up conditions of Run 2.

The primary vertex is selected as the vertex with the largest sum of squared transverse momenta ($\sum p_T^2$), where the sum includes only the tracks associated with that vertex. The selection has an efficiency $> 98\%$ [99] for reconstructing the correct primary vertex (where the hard scattering occurred) while the other vertices originate from pile-up and the underlying event.

3.2.2 Topological clusters

Clusters of topologically connected calorimeter cell signals, *topo-clusters*, are employed in the ATLAS experiment in the reconstruction of final states comprising isolated hadrons, jets and hadronically decaying τ-leptons [101]. In addition, topo-clusters are also used for the reconstruction of full-event observables such as the missing transverse momentum, and to reconstruct electrons and photons, due to their ability to better measure bremsstrahlung photons and electrons from photon conversion [102].

The topological clustering algorithm [95] forms clusters of neighbouring cells that have energies above the expected noise, yielding clusters that have a variable number

of cells. The expected noise is calculated as a quadrature sum of the electronic noise and the pile-up contribution, which depends on the instantaneous luminosity of the LHC.

The cluster growth starts at seed cells that have a signal-to-noise ratio in the cell above a high threshold ($t_{seed} = 4$). Seed cells form *proto-clusters*, that are ordered in descending order in signal-to-noise ratio. Then, the neighbouring cells of each seed cell are considered. If a neighbouring cell has not been used so far, and its signal-to-noise ratio is above a medium neighbour threshold $t_{neighbour} = 2$, the cell is added to the neighbour seed list and included in the proto-cluster. If a cell is adjacent to more than one proto-cluster, the proto-clusters are merged. If the signal-to-noise ratio is above the low cell threshold $t_{cell} = 0$ (and below $t_{neighbour}$), the cell is included only in the most significant adjacent proto-cluster. Once all seed cells have been processed, the original seed list is replaced by the neighbour seed list. This procedure is repeated until the seed list is empty. The low threshold at the perimeter ensures that tails of showers are not discarded, while the higher thresholds for seeds and neighbours suppress both electronics and pile-up noise.

Finally, a splitting step is applied to separate the proto-clusters according to their local energy maxima, to form *topo-clusters*.

3.3 High-level physics objects

In this section the high-level physics objects are discussed. Emphasis will be given to jets and missing transverse momentum, being the objects relevant for the b-tagging studies and the $ZH \to \nu\nu b\bar{b}$ channel of the $VH(H \to b\bar{b})$ analyses described in this thesis. The other physics objects will be briefly described for completeness. The specific object selection implemented for the analyses in this thesis will be discussed in Chapters 5 and 6.

3.3.1 Electrons

Electrons (and positrons) are reconstructed in the central region of the ATLAS detector by combining tracks reconstructed in the ID with topological clusters from the electromagnetic showers found in the electromagnetic calorimeters [102, 103]. In the forward region ($|\eta| > 2.5$), only calorimetric information from the EMEC and FCal is used.

An important aspect of reconstructed electrons is the discrimination between *prompt* and *non-prompt* electrons. The former are produced in the hard scattering, while the latter arise, for example, from semi-leptonic decays of heavy-flavour jets or photon conversion into e^+e^- pairs. Furthermore, the experimental signature of electrons can be faked, for example by mis-reconstructed pions or photons. *Identification* and *isolation* criteria have been developed to select electrons with high efficiency and purity.

Identification

The identification of electrons is performed using a likelihood-based method [104], combining track measurements (such as number of hits in each layer, track impact parameter and momentum loss in the ID) with calorimetric information (such as shower width and energy). Fixed values of the likelihood-based discriminant define different *operating points*, each having (p_T- and $|\eta|$-dependent) reconstruction efficiency and background rejection.

Isolation

Isolation criteria are defined to reject electrons with a significant amount of hadronic activity close-by and are therefore likely to be non-prompt. Given the experimental signature of electrons, isolation criteria [105] can use either track or calorimeter isolation properties.

The track-based isolation variable $p_T^{varcone0.2}$ is constructed from the sum of the transverse momenta of tracks with $p_T > 1$ GeV and within a cone of radius:

$$\Delta R = min\left(0.2, \frac{10\,\text{GeV}}{p_T^e}\right) \tag{3.1}$$

around a candidate electron of momentum p_T^e. To minimise pile-up contamination, only tracks from the primary vertex are used in the definition of this isolation variable.

Calorimeter-based isolation criteria are based on $E_T^{cone0.2}$, defined as the sum of transverse energies of topological clusters within a cone of size $\Delta R < 0.2$ from the candidate electron.

Different selection criteria applied to the isolation variables $p_T^{varcone0.2}$ and $E_T^{cone0.2}$, give rise to different operating points for electron isolation.

The reconstruction efficiencies for prompt electrons are measured in data using the J/Ψ and Z resonances [105] and correction scale factors are calculated in order for measurements on simulations to match those on data.

3.3.2 Photons

Energetic photons give origin to electromagnetic showers in the EM calorimeters, that have no matching tracks in the ID.

Prompt photons arising from the hard scattering can be distinguished from non-prompt ones, arising from hadronic jets, using selection criteria based on the shower shape [106]. Electromagnetic showers originated from prompt photons are typically narrower than the ones from hadronic jets and are associated with a smaller leakage in the hadronic calorimeters, indicating that no additional hadrons are present. Just as in the case of electrons, operating points are defined according to the selection

criteria.

Some photons will *convert* into e^+e^- pairs in the material traversed before reaching the calorimeters. Converted photons can be restructured from these tracks.

3.3.3 Muons

Muons are among the least interactive particles produced in pp collisions, traversing the full ATLAS detector. After leaving hits in the ID, muons deposit a small amount of energy in the calorimeters (for which they are *minimum ionising particles - MIP*), and then reach the muon spectrometer [107].

Firstly, they are reconstructed independently in the ID and MS, and then the information of the various sub-detectors is combined, in order to obtain the best performance. By combining differently the information from the involved sub-detectors, four types of muons are defined:

Combined muons: obtained through a global fit of the separate information coming from the ID and the MS. The procedure starts from the MS chambers and extrapolates inwards, towards the ID. Combined muons are those with the highest purity.

Segment-tagged muons: obtained by matching an ID track to at least one track segment in the MS. Low-p_T muons, or muons falling outside the MS acceptance usually belong to this category.

Calorimeter-tagged muons: obtained by matching an ID track to an energy deposit in the calorimeter system compatible with a MIP. This type of reconstruction is used outside the MS coverage or in regions of poor MS performance, for example in the central region of the detector ($|\eta| < 0.1$), due to the presence of cabling and supporting structures.

Extrapolated muons: obtained by only using information from the MS. These muons include for example those in the forward regions ($2.5 < |\eta| < 2.7$), outside the ID acceptance.

Identification

Once muons are reconstructed, prompt muons from hard scattering processes must be identified against the *background* muons, from in-flight decays of charged pions $\pi^{\pm}$ and kaons K. Background muons are characterised by a *kink* in the reconstructed track, where the in-flight decay took place.

Several variables discriminate between prompt muons and background non-prompt muons and selection criteria applied to these variables define different working points with the respective purity and rejections rate.

Isolation

Isolation criteria are defined to separate prompt muons decaying from short-lived bosons such as W, Z and H, and muons arising from the semileptonic decays of heavy-flavour hadrons. In analogy with electrons, isolation criteria are build using track and calorimeter information.

The track-based isolation variable $p_T^{varcone0.3}$ is constructed from the sum of the transverse momenta of the tracks found within a cone of radius:

$$\Delta R = min\left(0.3, \frac{10\,\text{GeV}}{p_T^{\mu}}\right) \tag{3.2}$$

around a candidate muon of transverse momentum p_T^{μ}.

Calorimeter-based isolation criteria are based on $E_T^{cone0.2}$, defined as the sum of transverse energies of the topological clusters within a cone of size $\Delta R < 0.2$ from the candidate muon.

Different selection criteria applied to $p_T^{varcone0.3}$ and $E_T^{cone0.2}$ define different operating points for muon isolation, each having certain reconstruction efficiency and background rejection.

The isolation efficiencies are measured in data and simulation on $J/\Psi \to \mu^+\mu^-$ and $Z \to \mu^+\mu^-$ events and scale factors are calculated in order to correct for simulation mis-modelling.

3.3.4 Jets

Due to confinement, quarks and gluons can only exist in bound states within hadrons. Therefore in ATLAS they can only be detected as *jets*, that are the product of the fragmentation and hadronisation of such partons [108]. Jets are collimated streams of colourless mesons and baryons observed in the calorimeters.

Reconstruction

Topo-clusters (see Section 3.2.2) can be combined into jets using many algorithms. Several criteria define a good jet reconstruction algorithm. One of the most important ones is *infrared and collinear safety* (IRC) [108], that guarantees that if one modifies an event with a collinear splitting or a soft emission, the set of hard jets found in the event should remain unchanged[2].

The anti-k_t jet clustering algorithm [109] is the main algorithm used for jet reconstruction in ATLAS. It is IRC safe and provides jets that are circular in the $y - \phi$ plane (see Figure 3.2).

[2]The importance of IRC safety for jet-clustering algorithms lies in three main reasons. Firstly, a hard *parton* (a *quark* or *gluon*) undergoes many collinear splittings as part of the fragmentation process. Additionally, there is always emission of soft particles in QCD events. Secondly, in fixed-order perturbative QCD calculations, soft emissions and collinear splittings are associated with divergent tree-level matrix elements, cancelled by corresponding divergent loop matrix elements only for IRC safe algorithms. Thirdly, detectors regularise collinear and infrared divergencies, because of finite resolution and non-zero momentum thresholds. The details of the regularisation depend on the experiment, making it quite difficult to connect experimental results for IRC unsafe algorithms to the expectations at hadron-level.

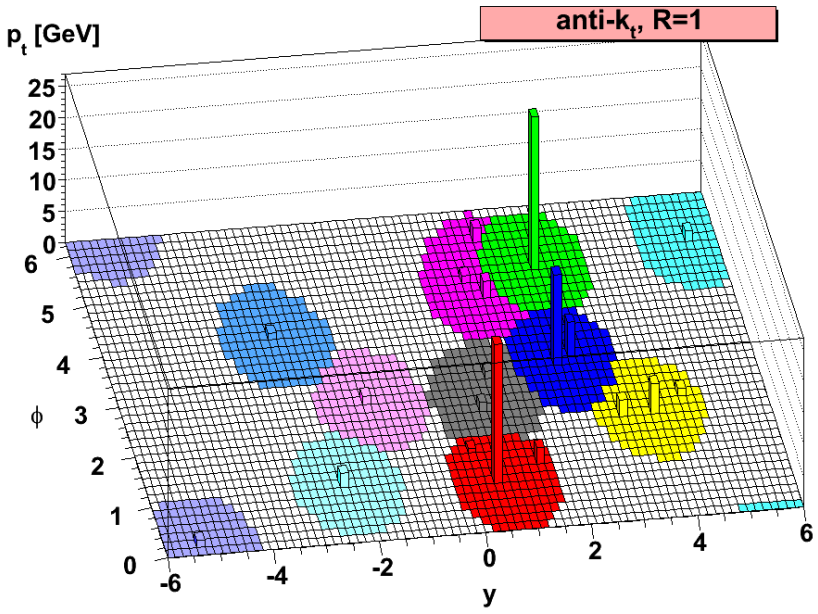


Figure 3.2: Simulation of jet reconstruction with the anti-k_t algorithm for a HERWIG-generated parton-level event. Image source: [109].

In the anti-k_t algorithm, each pair of topological clusters ij is associated with the following distance:

$$d_{ij} = min\left(\frac{1}{k_{T,i}^2}, \frac{1}{k_{T,j}^2}\right)\frac{\Delta_{ij}^2}{R}\,, \qquad \Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2\,, \tag{3.3}$$

where Δ_{ij}^2 is the angular distance between the two topological clusters, $k_{T,i(j)}$, $y_{i(j)}$ and $\phi_{i(j)}$ are the transverse momentum, rapidity and azimuth of topo-cluster i (j) respectively, and R is the *radius* of the jet cone. Values of $R = 0.4$ and $R = 1.0$ are the most used in ATLAS, with $R = 0.4$ defining *small-R* jets from quarks and gluons, and $R = 1.0$ defining *large-R* jets, usually containing the hadronic decays of high-p_T massive particles, such as top quarks and W, Z, and H bosons.

In the jet-clustering procedure, a second distance is defined:

$$d_{iB} = \frac{1}{k_{T,i}^2} \tag{3.4}$$

representing the distance (in the momentum space) between topo-cluster i and the beam axis.

The anti-k_t clustering algorithm works in steps, by first calculating d_{ij} and d_{iB} for all topo-clusters. Then, the smallest distance among d_{ij} and d_{iB} is found. If d_{ij} is the smallest, then the four-momenta of clusters i and j are combined into one. If d_{iB} is the smallest, cluster i is labelled a final jet and removed from the list of clusters. This process is repeated until all clusters are part of a jet.

3.3.4.1 Small-R jets

Small-R jets have a radius $R = 0.4$ and are usually originated by quarks and gluons.

Calibration

Once jets are clustered, their energy has to be *calibrated* to account for several experimental effects, such as the different response of electromagnetic and hadronic calorimeters, the presence of dead material and leakage, out-of-cone energy not included in the jet, and pile-up contamination. Thanks to the calibration, the jet energy scale (JES) of the reconstructed jet can be restored to the scale of *truth jets* (see Section 3.1), that are reconstructed from simulations using stable final-state particles as inputs (instead of calorimeter clusters). The calibration is derived using a combination of methods based on simulation, and data-driven techniques [110] that will be briefly explained here.

Origin correction

When jets are reconstructed, they point to the centre of the detector. This first correction step recomputes the jet four-momentum, to point to the hard scatter

primary vertex, while keeping the jet energy constant. This improves the jet η resolution.

Pile-up corrections

As it has been explained in Section 2.1.4, pile-up interactions can be *in-time* (when they occur within the bunch crossing of interest) or *out-of-time* (when they occur in neighbouring bunch crossings).

The contribution of pile-up to the jet energy is removed in two steps: first, the average pile-up contribution is subtracted from the jet's transverse momentum, based on the jet area [111]. Then, residual pile-up dependencies are removed with a second correction that depends on the number of interactions per bunch-crossing (μ).

Absolute JES correction

The four momenta of the reconstructed jets are corrected to the particle-level energy scale using simulation. The correction is E- and η-dependent, and is obtained through a comparison between reconstructed (*reco*) and *truth* jets, selected with a geometrical matching of $\Delta R = 0.3$.

Eta inter-calibration

This step uses well-measured *central* ($|\eta_{jet}| < 1.4$) jets, to derive a residual calibration for *forward* ($0.8 < |\eta_{jet}| < 4.5$) jets.

Global Sequential Calibration (GSC)

This correction reduces the fluctuations of the jet response due to energy leakage and, by exploiting shape variables and the energy of jets, improves the distinction between quark-initiated and gluon-initiated jets.

In-situ calibration

The last stage of the jet calibration corrects for differences between data and

simulation, arising from an incorrect detector description in the simulation. Jets in data are calibrated using the appropriate correction, derived from a range of studies on well understood and precisely measured physics processes, such as $Z \to \ell^+\ell^-$+*jets* and γ+*jets*, to constrain the jet p_T using the conservation of energy (*balance* method).

The total uncertainty associated with the complete JES calibration is $\approx 4.5\%$ at 20 GeV, $\approx 1\%$ at 200 GeV and $\approx 2\%$ at 2 TeV for an inclusive di-jet sample.

Another type of correction is applied to reconstructed jets, called *jet energy resolution* (JER). JER is defined as σ_{p_T}/p_T and it is also obtained in-situ, by comparing data and simulation [112]. The uncertainties associated with the JER calibration are $\approx 3\%$ at 20 GeV, and $< 1\%$ above 40 GeV on an inclusive di-jet sample.

The results of the calibration of small-R jets are shown in Figure 3.3.

Suppression of pile-up jets

The suppression of pile-up jets is fundamental for most physics analyses. After the dedicated corrections introduced above, spurious jets may remain.

The *jet-vertex-tagger* (JVT) is designed to discriminate hard scattering jets from pile-up jets [116]. It is constructed by combining variables based on tracking information (i.e. a weighted fraction of the transverse momentum sum of the tracks that are associated with the jet and originate from the hard scatter vertex) in a two-dimensional likelihood.

The output of the JVT discriminant is shown in Figure 3.4. Selection criteria applied to the discriminant define different working points with the respective JVT hard scatter efficiencies and pileup fake rates.

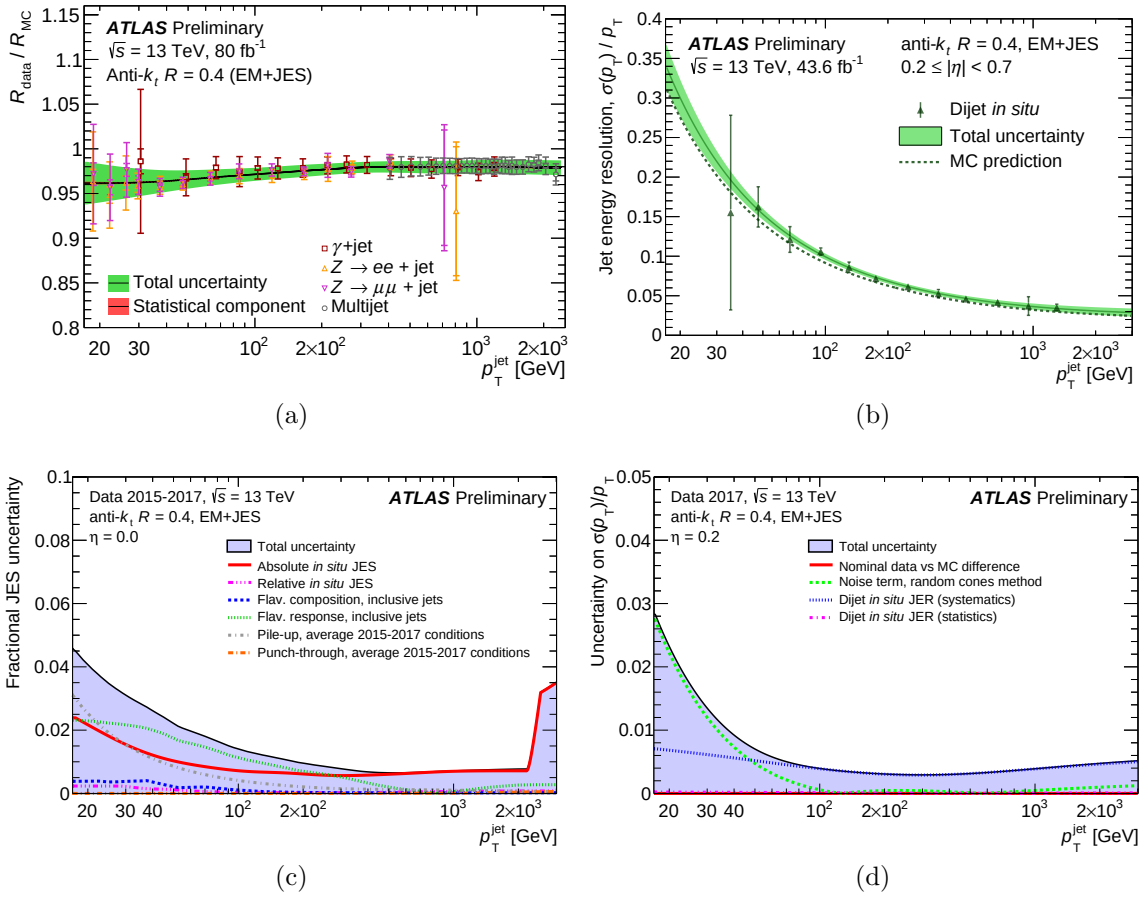


Figure 3.3: Data-to-simulation ratio of the average jet p_T response as a function of jet p_T for JES (3.3a). The relative jet energy resolution σ_{p_T}/p_T as a function of jet p_T (3.3b). JES (3.3c) and JER (3.3d) fractional uncertainties on small-R jets as a function of jet p_T. In Figure 3.3c, the uncertainty above 2 TeV is derived from single-particle response studies [113]. Image sources: [114] (3.3a, 3.3c), [115] (3.3b, 3.3d).

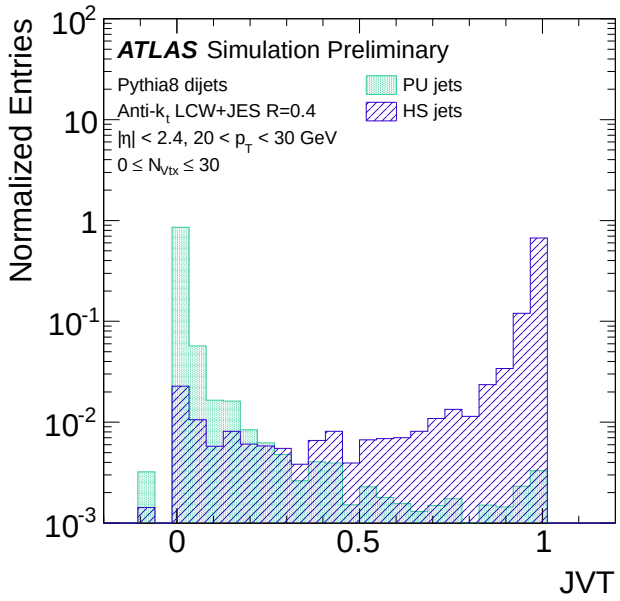


Figure 3.4: Distribution of JVT output for pile-up and hard scatter jets with $20 < p_T < 30$ GeV. Image source: [116].

3.3.4.2 Large-R jets

Large-R jets have radius $R = 1.0$, and usually contain the hadronic decays of high-p_T massive particles, such as top quarks and W, Z, and H bosons.

Large-R jet mass

The large-R jet mass is of crucial importance, since it is a powerful tool for distinguishing massive particle decays from non-resonant processes.

There are several ways in which the invariant mass of large-R jets can be reconstructed. The first and simplest way is through calorimeter information, using the energy and momentum of all associated topo-clusters [117]:

$$m^{calo} = \sqrt{\left(\sum_{i \in J} E_i\right)^2 - \left(\sum_{i \in J} \vec{p}_i\right)^2} \tag{3.5}$$

where m^{calo} is the *calorimeter* mass of large-R jet J, and the "i"s are the topo-cluster

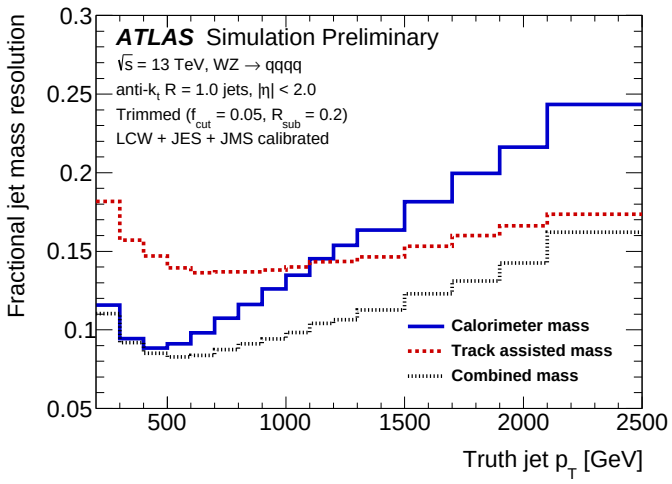


Figure 3.5: Fractional large-R jet mass resolution as a function of the truth jet mass transverse momentum, for three different mass definitions: calorimeter mass m^{calo}, track-assisted mass m^{TA} and combined mass m^{comb}. Image source: [118].

associated to it, with their energy E_i and momentum $\vec{p}_i$.

As a consequence of the Lorentz-boost, the angular spread in the decay products of a boosted massive particle scales as $1/p_T$ (see Equation 5.1), and for a sufficiently high Lorentz boost this spread is comparable with the calorimeter granularity. In order to avoid this resolution degradation (see Figure 3.5), tracking information is used to maintain the performance, beyond the granularity limit.

A straightforward method for combining tracking and calorimeter information is via the *track-assisted* mass (m^{TA}):

$$m^{TA} = \frac{p_T^{calo}}{p_T^{track}} m^{track} \tag{3.6}$$

where p_T^{calo} is the transverse momentum of the large-R jet, p_T^{track} and m^{track} are the transverse momentum and mass of the four-vector sum of tracks associated to the large-R jet. The ratio of the transverse momenta (p_T^{calo}/p_T^{track}) is used to correct for charged-to-neutral fluctuations.

The track-assisted mass has better resolution than the calorimeter mass at high-p_T, but not at lower-p_T (see Figure 3.5). Therefore, a weighted least-squares com-

bination of the two mass definitions is performed, such that the resolution of the *combined mass* measurement m^{comb} is always better than either of the two inputs [119] (see Figure 3.5):

$$m^{comb} = w_{calo}\, m^{calo} + w_{TA}\, m^{TA} \tag{3.7}$$

where w_{calo} and w_{TA} are determined from the expected mass resolutions σ_{calo} and σ_{TA} of the calorimeter and track-assisted measurements:

$$w_{calo} = \frac{\sigma_{calo}^{-2}}{\sigma_{calo}^{-2} + \sigma_{TA}^{-2}} \qquad w_{TA} = \frac{\sigma_{TA}^{-2}}{\sigma_{calo}^{-2} + \sigma_{TA}^{-2}} \tag{3.8}$$

Large-R jet grooming and calibration

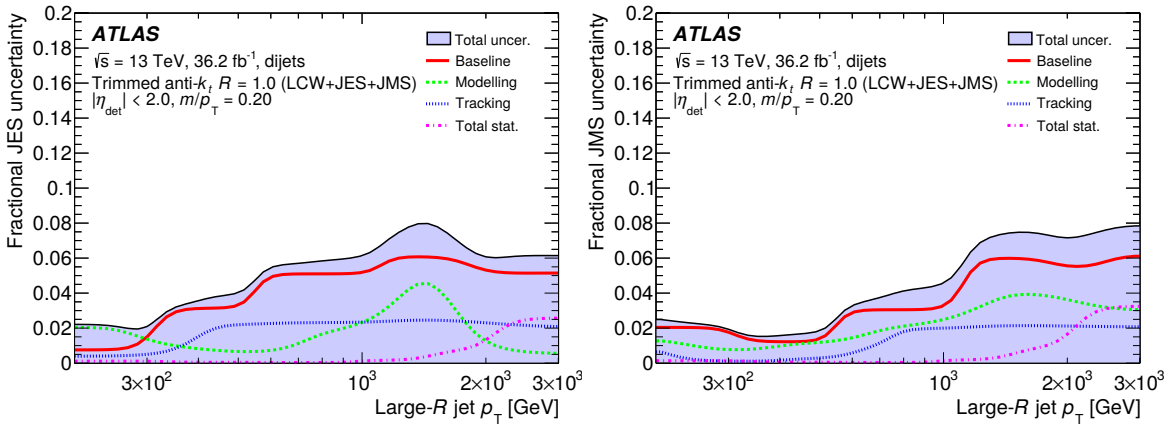
With a radius parameter $R = 1.0$ large-R jets are prone to collect energy depositions from soft particles originated from initial-state radiation, pile-up interactions or the underlying event[3]. In order to remove these contributions and reveal the hard substructure, several *grooming* algorithms have been developed [120]. In the analysis presented in this document, the grooming technique applied is *trimming* [121].

This is done by re-clustering the topo-clusters of the initial large-R jet using the k_T algorithm[4] into sub-jets with radius parameter $R_{sub} = 0.2$, and then removing any sub-jet that has a p_T less than 5% of the parent jet p_T. The remaining constituents form the trimmed jet.

After the grooming step, large-R jets are calibrated to account for several experimental effects. Thanks to the calibration, the jet energy scale (JES) and the jet mass scale (JMS) of the reconstructed jet can be restored to the scale of the matched *truth jet*. The calibration is derived using a combination of methods based on simulation, and data-driven techniques [119] that will be briefly explained here.

[3] The *underlying event* is the additional soft activity originating from the primary vertex.

[4] The k_T clustering algorithm can be obtained from Equations 3.3 and 3.4, by changing $k_{T,i(j)}^{-2} \rightarrow k_{T,i(j)}^{2}$.



(a) (b)

Figure 3.6: JES (3.6a) and JMS 3.6b) fractional uncertainties on large-R jets as a function of jet p_T. Image source: [119].

Simulation based E, η, m calibration

The trimmed large-R jets are calibrated to the energy scale of stable final state particles using corrections based on simulations to the jet energy, pseudorapidity and mass.

In-situ calibration

The residual *in-situ* correction, evaluated as data-to-simulation ratio, is applied to data, to account for the residual mis-modelling. The appropriate correction factor is chosen from the available studies of precisely measured physics processes (Z+*jets*, γ+*jets*). This correction was not applied to the large-R jets used in the $VH(H \to b\bar{b})$ boosted analysis in Chapter 5.

The total uncertainty associated with the JES and JMS calibrations used for the $VH(H \to b\bar{b})$ boosted analysis are shown in Figure 3.6. They are found to be $\sim 2\%$ at 300 GeV and to grow up to 8% at higher values of transverse momentum.

In addition to the scale uncertainties seen above, jet resolution uncertainties, that take into account the uncertainty on the jet mass resolution (JMR) and the

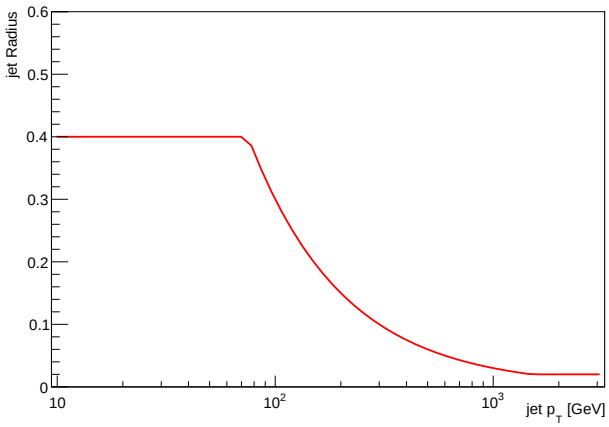


Figure 3.7: Sketch of the VR track jet radius as a function of the jet transverse momentum.

jet p_T resolution are applied. For JER, a Gaussian distribution with an absolute 2% uncertainty is used to smear the simulated distribution. The difference between the smeared and unsmeared distributions is taken as resolution uncertainty. As per the JMR uncertainty, a relative jet mass resolution smearing of 20% is applied to each jet. Both these numbers were derived from early Run 2 studies.

3.3.4.3 Track jets

Track jets can be reconstructed from clusters of tracks in a similar way to calorimeter jets with topo-clusters. The tracks used for the clustering of track jets pass the quality requirements mentioned in Section 3.2.1.

The track jets used in the $VH(H \to b\bar{b})$ boosted analysis are clustered using the anti-k_T algorithm with a variable radius parameter (VR track jets) [122, 123]. The variable radius R, which accounts for the collimation of jets at high momenta, is defined as:

$$R\left(p_T\right) = \frac{\rho}{p_T} \tag{3.9}$$

where the ρ parameter is set to 30 GeV. There are two additional parameters, $R_{min} =$

0.02 and $R_{max} = 0.4$, used to set the minimum and maximum variable jet radius, as shown in Figure 3.7.

Large-R jet to track jet association

As explained in Chapter 5, the $VH(H \to b\bar{b})$ boosted analysis uses large-R jets to reconstruct the $H \to b\bar{b}$ decay. Therefore, at least two *b-tagged* track jets[5] are required to be matched to the Higgs boson candidate large-R jet.

However, in dense hadronic environments, matching a track jet to a large-R jet using the angular separation between the two jets is challenging, since ambiguities may arise. This is solved thanks to *ghost association* [124].

This method involves adding the track jet four-momentum to the large-R jet, setting the track jet p_T to an infinitesimal amount. In this way, the track jet four-momentum only retains the spatial information on its direction ("*ghost*" track jet). At this point, the jet clustering algorithm for the large-R jet is rerun, including all the ghost track jets and the calorimeter topo-clusters. If the ghost track jet is contained in the re-clustered large-R jet, then the original track jet is said to be *ghost-associated* to the large-R jet.

The infinitesimal p_T of the ghosts and the resilience of the jet algorithm to soft particles ensures that there are no differences between the final jet and the one clustered without ghosts.

3.3.5 Taus

Tau particles ($\tau^{\pm}$) [27] are the heaviest leptons, with a mass of ≈ 1.78 GeV. The high mass implies a very small lifetime, that corresponds to an average decay length of 87 μm, well before reaching the IBL. Tau leptons can decay leptonically ($\mathrm{BR}(\tau \to \ell\nu_\ell\nu_\tau) = 35\%$) or hadronically in the remaining cases.

[5] b-tagging is explained in detail in Chapter 4.

Leptonic decays yield a charged lepton (electron or muon) in the final state that can not be easily distinguished from a prompt one and making their reconstruction and identification is challenging. Hadronically-decaying τ leptons can be typically identified by the presence either of a single charged pion (in 75% of the cases – *1-prong* signature) or three charged pions (in $\approx$ 22% of cases – *3-prong* signature). The charged pions are then reconstructed as jets of particles [125].

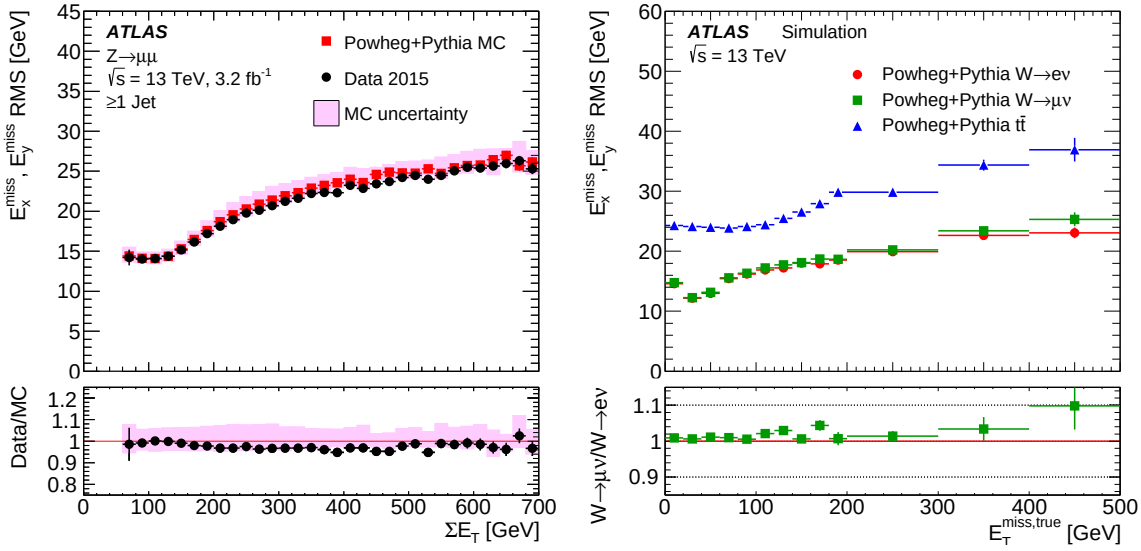
Several variables, including those based on the angular distance ΔR between tracks and the jet direction, and the tau vertex reconstruction, are fed into a BDT to identify jets as originating from hadronic tau decays [126]. Calibrated working points corresponding to different tau identification efficiency values are defined.

3.3.6 Missing transverse momentum and energy

So far, the experimental signatures of particles that can interact electromagnetically and strongly with the ATLAS detectors have been discussed. For particles that can only interact weakly, such as neutrinos and possible dark matter candidates, the interaction cross section with the sub-systems of the ATLAS detector is negligible.

These particles can however be indirectly detected thanks to the laws of conservation of energy and momentum. At hadron colliders, the total initial momentum of the colliding partons is unknown along the z-axis, but known (and null) in the transverse plane. Therefore, any transverse momentum imbalance can be attributed to weakly interacting particles, produced in the hard scattering collision.

The missing transverse momentum $\vec{E}_T^{miss}$ can be calculated by considering two contributions [127]: a *hard term*, corresponding to the sum of the transverse momenta of the reconstructed and calibrated physics objects described so far, and a *soft term*, accounting for additional charged-particle tracks in the ID, that are associated with the primary vertex, but not with any other physics object.



(a) (b)

Figure 3.8: $\vec{E}_T^{miss}$ resolution measured by $\mathrm{RMS}_{x(y)}^{miss}$ (3.8a) as a function of $\sum E_T$ determined for an exclusive $Z \to \mu\mu$ sample with at least one jet with $p_T > 20$ GeV in data and simulation, and (3.8b) as a function of the true transverse momentum $E_T^{miss,true}$ for simulated $W \to e\nu$, $W \to \mu\nu$ and $t\bar{t}$ samples. Image source: [127].

$$\vec{E}_T^{miss} = -\left(\sum_{hard} \vec{p}_T + \sum_{soft} \vec{p}_T\right) = -\left(\sum_{e} \vec{p}_T + \sum_{\mu} \vec{p}_T + \sum_{\tau} \vec{p}_T + \sum_{\gamma} \vec{p}_T + \sum_{jets} \vec{p}_T + \sum_{soft} \vec{p}_T\right) \tag{3.10}$$

$\vec{E}_T^{miss}$ is a vector, with magnitude E_T^{miss}, and its direction can be inferred from the momentum imbalance in the transverse plane and is usually indicated as a dashed line in event displays (see Figure 3.1).

The reconstruction and measurement of $\vec{E}_T^{miss}$ is challenging, due to the nature of this quantity, and the object selection can be refined in the context of each analysis, in order to optimise $\vec{E}_T^{miss}$.

Finally, the performance of the $\vec{E}_T^{miss}$ reconstruction is evaluated for scale and resolution (see Figure 3.8), and measured in events with null expected $\vec{E}_T^{miss}$ ($Z \to \ell^+\ell^-$, $\ell = e, \mu$), and events where $\vec{E}_T^{miss}$ is expected from the presence of at least one neutrino (W and top quark decays).

3.3.7 Overlap removal procedure

Physics objects are reconstructed from signals collected within different sub-detectors, and for this reasons the same signal can be used to reconstruct different objects. To avoid double (or multiple) counting, a hierarchy is defined for the physics object, prioritising muons and then, in order, electrons, photons, tau leptons and jets. Given some criteria, the object with lower priority is removed [128].

The overlap removal acts on the following pairs of physics objects:

- (e, τ_{had}): if $\Delta R(e, \tau_{had}) < 0.2$, τ_{had} is removed
- (μ, τ_{had}): if $\Delta R(\mu, \tau_{had}) < 0.2$, τ_{had} is removed
- (μ, e): if a combined μ shares tracks with an e, the e is removed
- (e, jet): if $\Delta R(e, jet) < 0.2$, the jet is removed
- (μ, jet): if $\Delta R(\mu, jet) < 0.2$ and $N^{jet}_{tracks} < 3$, the jet is removed
- (τ_{had}, jet): if $\Delta R(\tau_{had}, jet) < 0.2$, the jet is removed.

Chapter 4

b-Tagging

The identification of jets containing b-hadrons (b-tagging) is important for many physics programmes at the LHC. In the ATLAS experiment, b-tagging techniques play a major role in both precise SM measurements including the Higgs sector, and searches for new physics BSM.

This chapter illustrates the physical foundations of b-tagging, as well as the main b-tagging algorithms and their usage in ATLAS.

4.1 Physical foundations of *b*-tagging algorithms

The identification of b-jets relies on the properties of the b-hadrons originated in the hadronisation of b-quarks. The first one is the $\sim$ 1.5 ps lifetime, long compared to that of hadrons containing valence quarks of different flavours. This is a direct consequence of the $b \to c$ transition being "Cabibbo-suppressed", as it is governed by the off-diagonal term V_{cb} in the CKM matrix [129]. As a consequence, b-hadrons travel inside the detector leaving a measurable displacement with respect to the primary vertex, where the proton-proton collision took place. For example, a b-hadron of $p_T = 100$ GeV would travel on average 6 mm in the transverse plane inside the detector before decaying (see Figure 4.1).

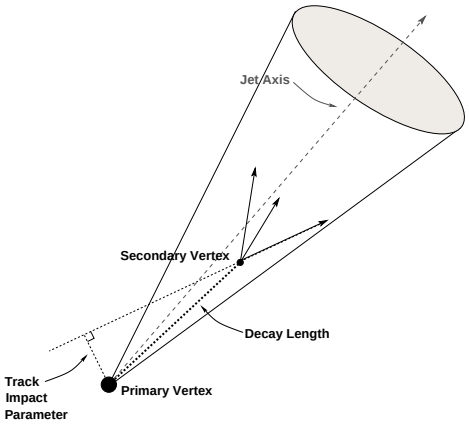


Figure 4.1: Illustration of the b-jet originated by a b-hadron decay. Image source: [130].

This displacement can be inferred using modern silicon tracker detectors that identify primary vertices and secondary vertices, from which long-lived hadrons decay to lighter particles.

Another property of b-hadrons involves their high mass. This produces b-jets with a large number of decay products and secondary particles with transverse momenta higher than those from light-flavoured and charmed hadron decays. Furthermore, b-hadrons have a high BR for semi-leptonic decays [27]:

$$BR(B \to X + \ell + \nu_\ell) \simeq 11\% \tag{4.1}$$

where B is the b-hadron, X is the (c-) "daughter" hadron, and ℓ and ν_ℓ are an electron or a muon and the respective neutrino.

Finally, b-hadrons decay preferably into c-hadrons ($|V_{cb}|^2 \gg |V_{ub}|^2$) [129], which subsequently decay. As a result, the typical topology of a b-jet is that of a decay chain with two consecutive vertices, the secondary vertex stemming from the b-hadron decay and a tertiary vertex from the decay of the c-hadron.

4.2 *b*-Tagging algorithms in ATLAS

The ATLAS experiment relies on multiple algorithms to perform b-tagging, exploiting the aforementioned properties of b-hadrons. Such algorithms can be divided into *low-level taggers*, that use directly the trajectories of charged particles reconstructed in the inner detector (ID) and then associated to jets, and *high-level taggers*, exploiting multivariate techniques and for which the output of low-level algorithms is used as an input. The output of high-level taggers is the variable used for the final discrimination (tagging) of the b-jets from jets of different flavours.

4.2.1 Track-to-jet association for *b*-tagging algorithms

It has been seen above that many low-level taggers use the trajectories of charged particles reconstructed in the ID and matched to jets, as the input for the b-tagging discrimination.

The track-to-jet association is based on the angular separation between the tracks and the jet, $\Delta R(track, jet)$ [131, 132], where the ΔR requirement is defined as a function of the jet p_T, to account for the collimation of the b-hadron decay products at higher p_T regimes (see Figure 4.2), through empirical formula:

$$\Delta R(track, jet) = 0.239 + e^{-1.220 - 1.64 \times 10^{-5} p_T\,[MeV]} \tag{4.2}$$

This ensures a high efficiency for the association of tracks from b-hadrons and suppresses the contamination from spurious tracks of different origin. Tracks can be associated only to one jet. If a track can be associated to more than one jet, it is associated to the closest jet in the ΔR space. Tracks (as well as jets) are also required to pass quality requirements that depend on the specific b-tagging algorithm.

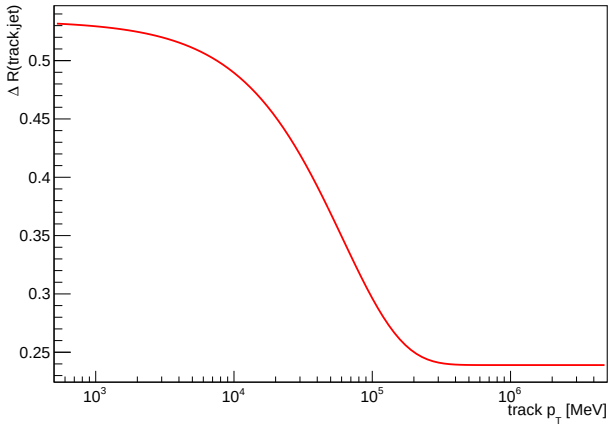


Figure 4.2: Track-to-jet association function used by b-tagging algorithms in the ATLAS experiment. The plot shows the $\Delta R(track, jet)$ as a function of the track p_T in MeV.

In the following, a detailed description of the b-tagging algorithm most widely used in the Run 2 ATLAS analyses, including the $VH(H \to b\bar{b})$ analyses presented in Chapters 5 and 6, is provided. The calibration of b-jets at very high transverse momentum described in Section 4.4 has been performed using the same algorithm.

4.2.2 Algorithms based on impact parameters

As discussed above, the long lifetime of b-hadrons results in a measurable displacement of the tracks originating from the secondary vertex to the primary vertex, as sketched in Figure 4.1 and shown in a candidate $ZH \to \nu\nu b\bar{b}$ event in Figure 4.3. A large impact parameter (IP) characterises displaced tracks and this property is used by the so-called *impact parameter based* low-level taggers. ATLAS exploits two complementary impact parameter based algorithms, that use the signed impact parameter significance of tracks [133, 134], as pictured in Figure 4.4.

The impact parameter significance is defined as the ratio between the impact parameter and its uncertainty. It can be calculated in the transverse plane (d_0/σ_{d_0}) or in the longitudinal plane ($z_0 \sin\theta/\sigma_{z_0 \sin\theta}$). The sign of the IP significance is determined

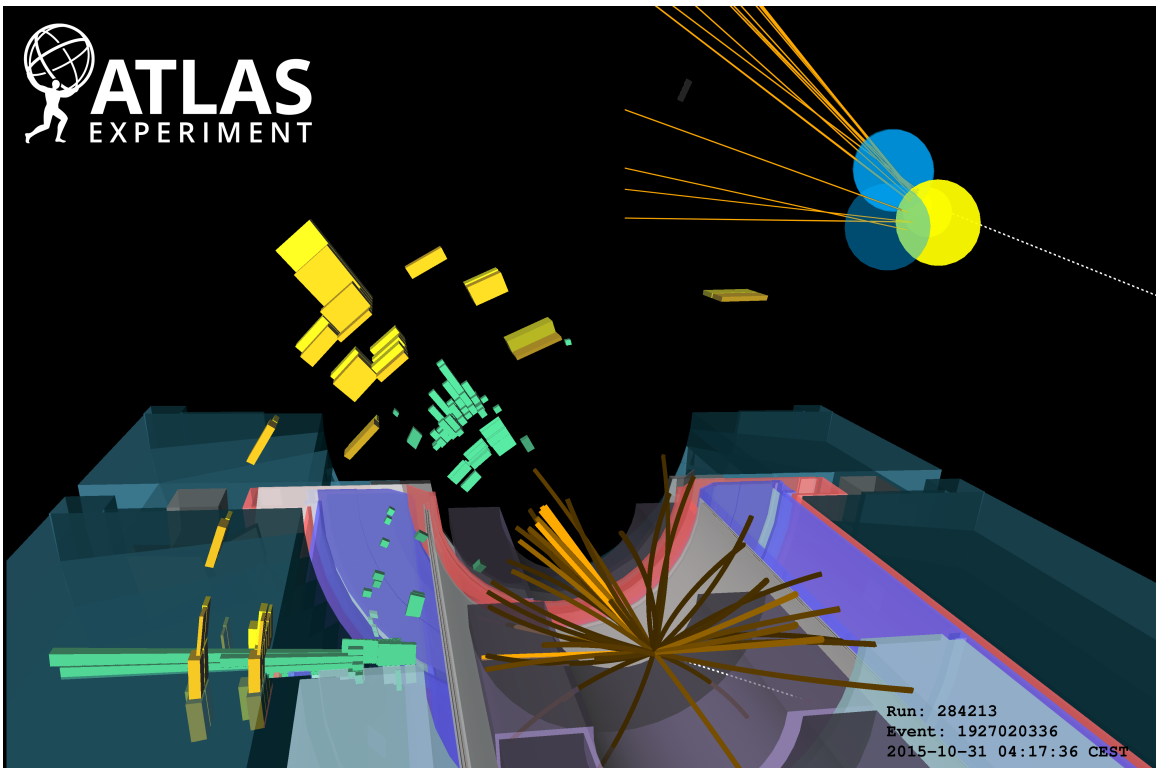


Figure 4.3: Candidate event for the process $ZH \rightarrow \nu\nu b\bar{b}$ (Run 284213, Event 1927020336). The missing transverse momentum is identified by a white dashed line. The decay products of the Higgs-boson candidate are reconstructed as two b-tagged jets built from tracking and calorimeter information. In the top right corner of the event display a zoom of the collision is presented. The structure of the primary vertex (yellow) and two secondary vertices for either of the b-tagged jets (blue) is highlighted. Image source: https://twiki.cern.ch/twiki/pub/AtlasPublic/EventDisplayRun2Physics/Run284213Evt1927020336_VHbb_0lepton_1.png.

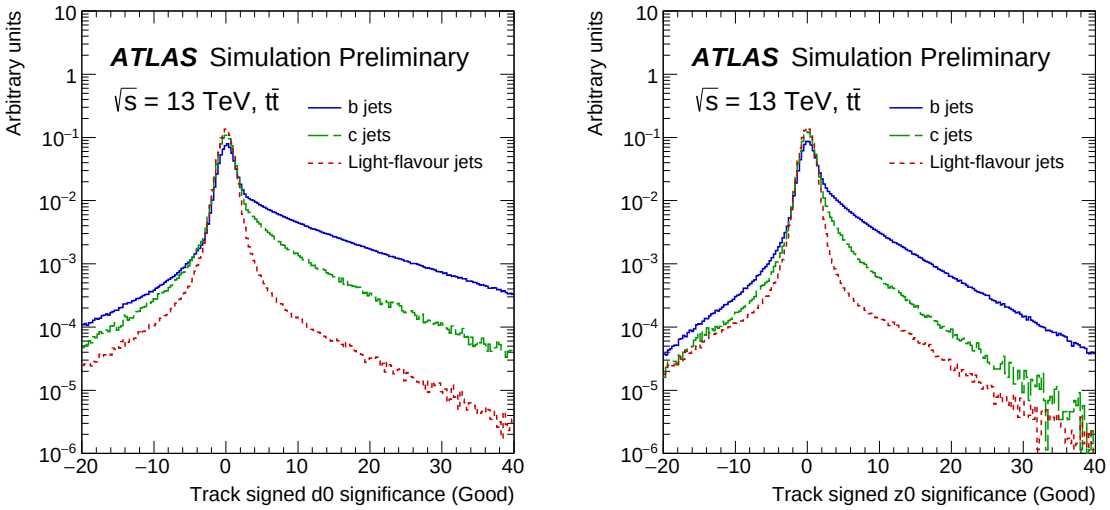
by the sign of the IP itself, on the basis that the decay point of the b-hadron must lie along its flight path. The sign is positive if the track extrapolation crosses the jet direction in front of the primary vertex, and negative otherwise. By construction, tracks from heavy flavour hadron decays tend to have a positive sign, as shown in Figure 4.4.

`IP2D` only exploits the transverse IP significance, while `IP3D` uses both the transverse and the longitudinal impact parameter significance of tracks in a two-dimensional template that also accounts for their correlation. Both algorithms exploit the probability density functions for the signed IP significance of these tracks, to define ratios of the b-jets and light-flavour jet probabilities, on a per-track basis. These ratios provide a log-likelihood discriminant for both algorithms (see Figure 4.5).

4.2.3 Algorithm based on secondary vertex reconstruction

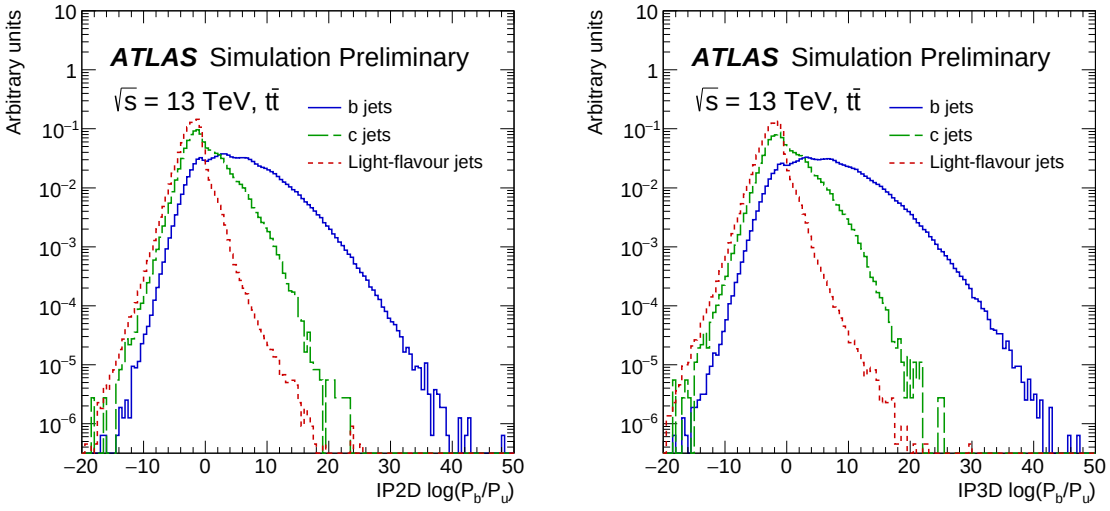
Displaced decays of b-hadrons originate *secondary vertices*, from which the tracks from the b-hadron decay products can be reconstructed (see Figure 4.3). The presence of a secondary vertex within a b-jet is exploited by a dedicated low-level tagger (`SV1`) [135], that attempts to reconstruct a *single* displaced secondary vertex in the jet.

This reconstruction is an iterative procedure, starting from identifying all the possible two-track vertices that can be built using the tracks associated with the jet, while rejecting those that are likely to originate from long-lived particles (such as K_S or Λ), photon conversions or hadronic interactions with the detector material. All tracks are then combined into one set of selected tracks, that is supplied to the vertex fitter. The fitting procedure [136, 137] aims to find only one secondary vertex from all the tracks in the set, and uses a χ^2 test either to assign tracks to the vertex, or to discard them. This procedure has been highly refined, optimised and perfected [138], to provide the highest b- (or c-) hadron vertex reconstruction efficiency with the smallest possible rate of reconstruction of vertices not coming from the heavy



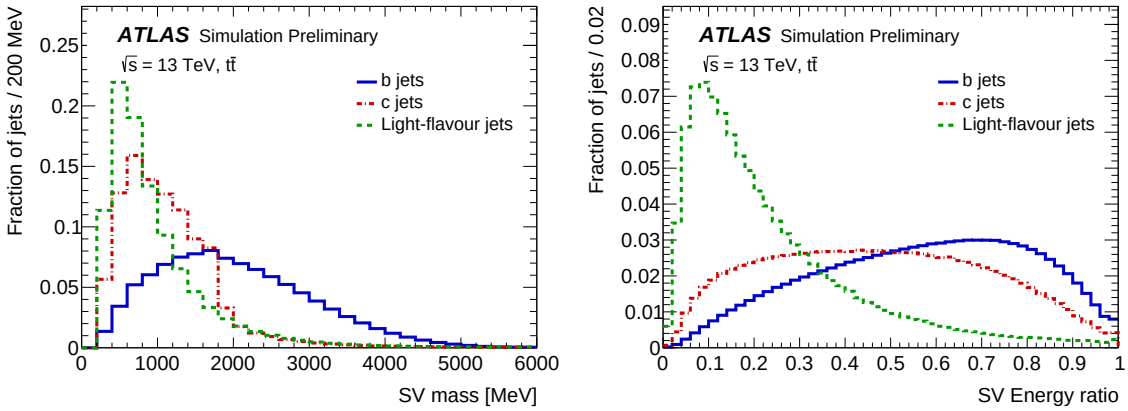
(a) (b)

Figure 4.4: Transverse (4.4a) and longitudinal (4.4b) signed impact parameter significance of tracks in $t\bar{t}$ events for b (solid blue), c (dashed green) and light flavour (dotted red) jets. Image source: [134].



(a) (b)

Figure 4.5: Log-likelihood ratio for the IP2D (4.5a) and IP3D (4.5b) b-tagging algorithms for b (solid blue), c (dashed green) and light flavour (dotted red) jets in $t\bar{t}$ events. Image source: [134].



(a) (b)

Figure 4.6: Distributions of the properties of the reconstructed secondary vertex, for b-jets (blue), c-jets (red) and light-flavoured jets (green): (4.6a) mass of the secondary vertex and (4.6b) "energy ratio" (defined as the total energy of the tracks associated with the secondary vertex divided by the energy of all the tracks associated with the jet). Image source: [135].

flavour hadron decays.

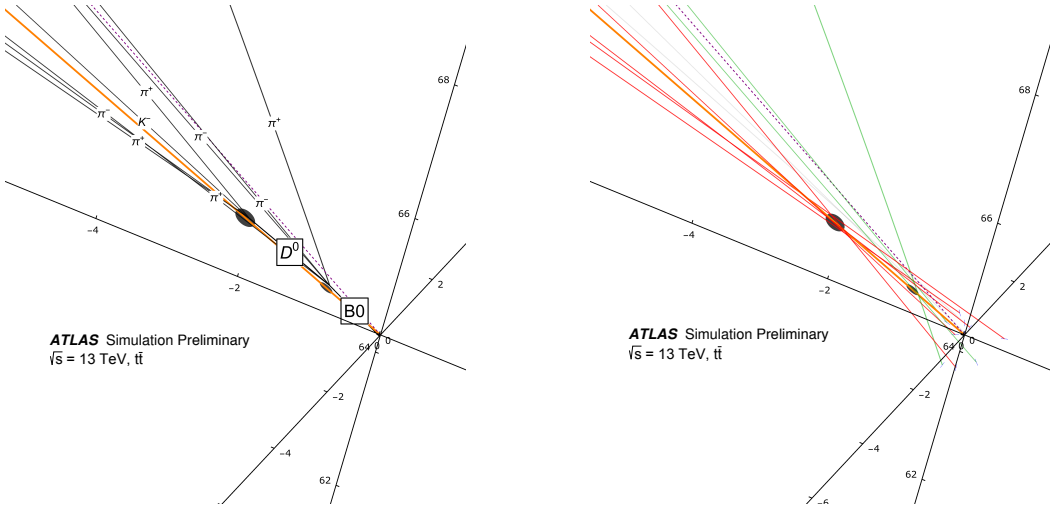
Once the secondary vertex and its tracks have been identified, a set of discriminating variables is constructed, to be used as inputs to the high-level taggers. The distributions for the secondary vertex mass and the energy fraction of the tracks associated with the secondary vertex (*energy ratio*) are shown in Figure 4.6.

4.2.4 Topological decay-chain reconstruction algorithm

The topological multi-vertex algorithm (`JetFitter`) [139, 140] exploits the topological structure of weak b- and c-hadron decays inside the jet and aims to reconstruct the full b-hadron decay chain (see Figure. 4.7).

The algorithm uses a modified Kalman filter [98] that finds a common line, approximating the b-hadron flight on which the primary, secondary (b-) and tertiary (c-) vertices lie.

This algorithm has been highly optimised [138] to increase the efficiency (especially that for tertiary vertex reconstruction) and minimising the effect of nearby



(a) (b)

Figure 4.7: Event display of a simulated B^0 decay showing the generated (4.7a) and reconstructed (4.7b) charged particle trajectories. Reconstructed particles are extrapolated to the point of closest approach to the beam collision point, including particles originating from the secondary (green) and the tertiary (red) vertex. The axes give the coordinates in mm. `JetFitter` performs a multi-vertex fit constraining all vertices to lie on a single flight axis (the orange line in both figures), initialised using the jet direction. Image source: [139].

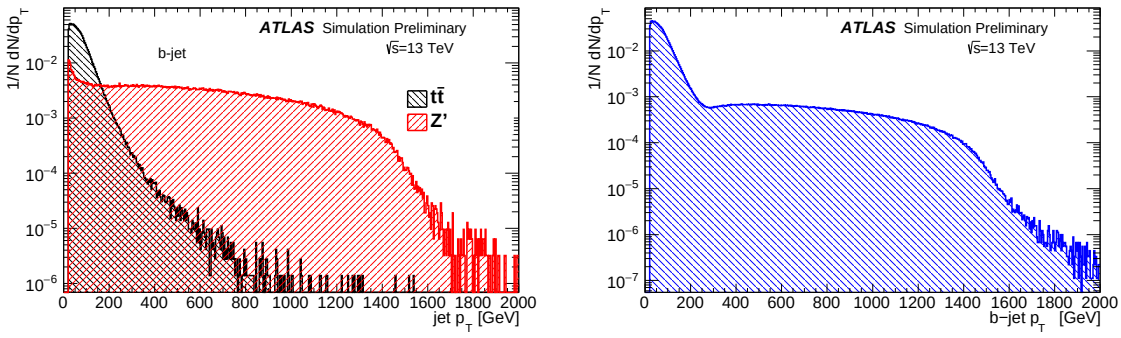
interactions not related to the weak b-hadron decay. Once the decay chain has been identified, a set of discriminating variables associated to the multiple vertices is constructed, to be used as inputs to the high-level taggers.

4.2.5 The `MV2` multivariate algorithm

The outputs of the low-level taggers described above are then combined into a final algorithm (`MV2`) [141, 1], that represents the last step for the jet to be "b-tagged" or not.

The `MV2` tagger is based on a boosted decision tree (BDT), whose input variables are detailed in Table 4.1. Two kinematic properties of the jets (p_T and $|\eta|$) are included, in order for the algorithm to exploit their correlations with other variables[1].

[1]In order to avoid differences in the kinematic distributions of signal (b-jets) and background (c-jets and light-flavour jets) to be used to discriminate the jet flavours, at training stage b-jets and c-jets are reweighed in p_T and $|\eta|$, so as to reproduce the spectrum of the light-flavour jets.



(a) (b)

Figure 4.8: Distribution of the b-jet transverse momentum for $t\bar{t}$ and Z' events (4.8a) and the *hybrid* sample used for the training of the multivariate-based taggers (4.8b). A similar strategy is applied to c- and light-flavour jets. Image source: [138].

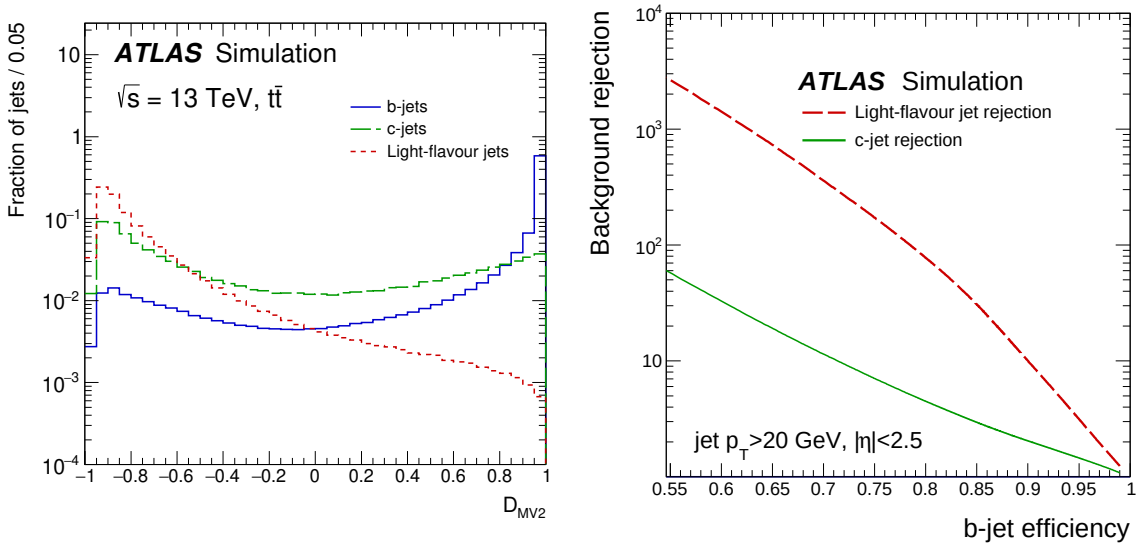
The `MV2` algorithm is trained using TMVA [142], on a "hybrid" $t\bar{t} + Z'$ sample. More details on the $t\bar{t}$ and Z' samples can be found in Section 4.4.2. The *hybrid* sample is build by selecting b-jets[2] from $t\bar{t}$ events, if the corresponding b-hadron p_T is below 250 GeV and from the Z' sample if above (see Figure 4.8), with an analogous strategy applied to c- and light-flavour jets.

The training is conducted considering b-jets as signal and an ensemble of c- and light-flavour jets as background. The fraction of c-jets in the background sample is set to 7%, allowing the charm rejection to be enhanced while still preserving a high light-flavour jet rejection. The output discriminant of the `MV2` algorithm for b-, c- and light-flavour jets, as well as its rejection rates, evaluated on $t\bar{t}$, are shown in Figure 4.9.

[2]The labelling scheme used to define the flavour of the jet (b- in this case) is performed by matching the jets to b-hadrons with $p_T > 5$ GeV within a cone of radius $\Delta R = 0.3$ around the jet axis.

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

Table 4.1: Input variables used by the MV2 algorithm. Table source: [1].



(a) (b)

Figure 4.9: MV2 output distributions for b-jets (blue), c-jets (red) and light-flavoured jets (green) (4.9a) and c-jets (green) and light-flavoured jets (red) MV2 rejection rates as a function of the b-jet tagging efficiency (4.9b). Image sources: [1, 141].

4.3 Usage of b-tagging in ATLAS physics analyses

4.3.1 Performance of MV2

The evaluation of a b-tagging algorithm is based on its identification (*tagging*) efficiency and the corresponding rejection power for other flavours (c- and light-). b-jets are selected based on fixed selection requirements on the algorithm discriminant (*working* or *operating points*), and the efficiency, as a function of the b-jet p_T, is determined as the ratio between the number of b-jets passing the selection and the number of total b-jets:

$$\epsilon_b(p_T) = \frac{N^{b-jets}_{passing\ selection}(p_T)}{N^{b-jets}_{total}(p_T)} \tag{4.3}$$

The rejection rate for c- (light-flavour) jets is defined as $1/\epsilon_c$ ($1/\epsilon_{light-flavour}$), where ϵ_c ($\epsilon_{light-flavour}$) is obtained from Equation 4.3 substituting the jet flavour accordingly. The selection on the MV2 algorithm discriminant determining the working points is

ϵ_b	MV2		
	Selection	Rejection	
		c-jet	Light-flavour jet
60%	> 0.94	23	1200
70%	> 0.83	8.9	300
77%	> 0.64	4.9	110
85%	> 0.11	2.7	25

Table 4.2: Selection, c-jet and light-flavour jet rejections for the different b-jet tagging working points for the `MV2` algorithm, evaluated on the baseline $t\bar{t}$ events. Table source: [1]

chosen to yield a b-tagging efficiency in the reference $t\bar{t}$ sample (see Section 4.4.2). At a given working point, the performance of two (or more) b-tagging algorithms can be compared by looking at the corresponding c- and light-flavour jet rejection rate, where the highest rejection rate indicates a better performing tagger. A summary of the selections used to define the `MV2` working points and the corresponding c- and light-flavour jet rejections is shown in Table 4.2. The choice of the used working point depends on the physics analysis. The $\epsilon_b = 70\%$ working point is used for the $VH(H \to b\bar{b})$ analyses presented in Chapters 5 and 6.

4.3.2 Application of *b*-tagging to physics analysis

After defining the working points, the p_T-dependent efficiency measured using collision data (ϵ_b^{data}) is compared to that obtained in simulation (ϵ_b^{sim}) and p_T-dependent data-to-simulation *scale factors* are defined to account for the difference between the two, due to imperfect event modelling and defects in the simulation. The procedure of calculating and applying the scale factors, as well as the measurement of the impact of the main sources of systematic uncertainties on the correction is usually referred to as *calibration*. The calibration adopted in the $VH(H \to b\bar{b})$ analyses is described in detail in [1]. In Figure 4.10 the b-tagging efficiency scale factors and uncertainties for the ϵ_b =70% working point are shown.

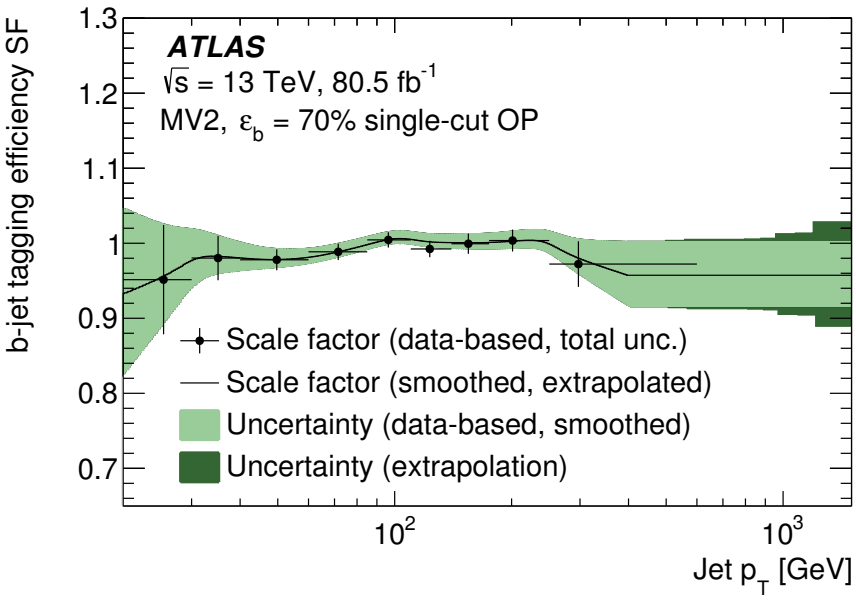


Figure 4.10: b-tagging efficiency simulation-to-data scale factors and uncertainties for the ϵ_b =70% working point for the MV2 algorithm as a function of the b-jet p_T. The η dependence is found to be negligible. Image source: [1].

In the ATLAS physics analyses that use b-tagging, a flavour-dependent weight is applied to each jet in the simulation to match the tagging rate measured in data. Therefore, once a working point is chosen, the product of the weights of each jet in the event is assigned as the event weight, to preserve the number of generated events. If the simulated jet is b-tagged the weight is simply the simulation-to-data scale factor:

$$w_{jet}(p_T) = SF_b(p_T) = \frac{\epsilon_b^{data}(p_T)}{\epsilon_b^{sim}(p_T)} \tag{4.4}$$

Conversely, if the jet is not tagged, the weight is calculated as:

$$w_{jet}(p_T) = \frac{1 - \epsilon_b^{data}(p_T)}{1 - \epsilon_b^{sim}(p_T)} = \frac{1 - SF_b(p_T)\epsilon_b^{sim}(p_T)}{1 - \epsilon_b^{sim}(p_T)} \tag{4.5}$$

Uncertainties on the scale factors are then assessed, in various kinematic regions and separately for different flavours. Finally, an additional uncertainty accounts for the *extrapolation* of the calibration to jets with p_T beyond the kinematic reach of the

data calibration, as discussed in Section 4.4. These uncertainties are then decomposed in each of the flavour categories into independent components, that physics analyses use as *nuisance parameters* in the statistical analysis [143] to obtain the final results.

4.4 Calibration of *b*-jets at very high transverse momentum

The necessity of having high-performance b-tagging algorithms across the full p_T spectrum at the ATLAS experiment has been discussed above, together with the physical foundations of b-tagging algorithms, their implementation and calibration. Finally, in Section 4.3.2, the usage of b-tagging in the ATLAS analyses has been explained.

At the increasingly high centre-of-mass energy of the pp collisions delivered by the LHC, one can expect events to have jets with increasingly high p_T, including those generated by b-hadrons. Furthermore, several BSM models predict final states with b-jets at very high p_T, above a few hundred GeV. Therefore, b-tagging algorithms need to be calibrated also at extremely high p_T.

At the LHC the main source of b-jets in data is provided by top quark decays (through $t \to bW$) and this constrains the b-jet transverse momenta below $\sim$ 400 GeV (see Figure 4.11). Therefore, in order to extend the calibration range beyond 400 GeV and up to the TeV scale, it is necessary to perform a fully simulation-based extrapolation, since all methods using data are statistically limited in the high p_T regime.

4.4.1 Calibration strategy

The simulation-based *extrapolation* to very high-p_T jets is conducted following two steps: firstly the *scale factor* from the highest p_T bin of the data-based calibration is directly applied to higher p_T jets. Secondly, additional *systematic uncertainties* δ_i^m, where i runs on the sources of uncertainty and m indicates the p_T bin of the calibration, are assessed from simulation; the final uncertainties for high p_T b-jets are

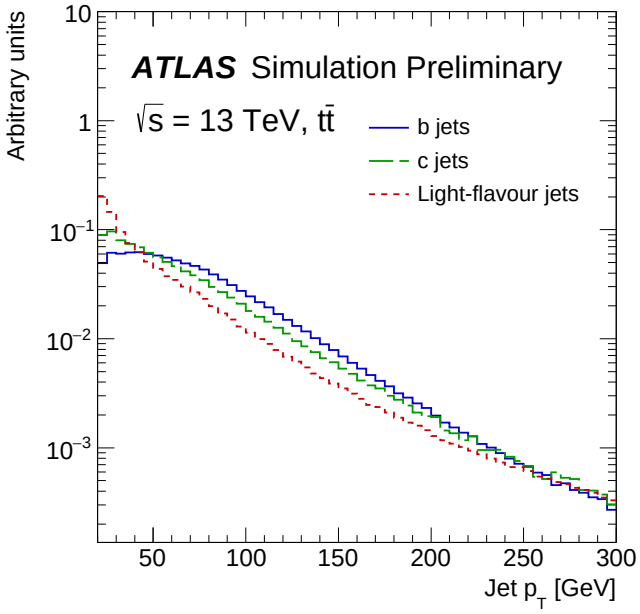


Figure 4.11: Kinematic distributions of p_T for b- (blue), c- (green) and light-flavour (red) jets in $t\bar{t}$ events at $\sqrt{s} = 13$ TeV. Image source: [134].

obtained by combining the *incremental*[3] simulation-based uncertainties in quadrature with the ones from the data based calibration at $p_T = 400$ GeV.

The method with which the final uncertainties are obtained will be now explained in detail. The b-tagging efficiency scale factor at high p_T, $SF(p_T)$, is defined with respect to the data-based calibration in a reference bin $SF_{DATA}(p_{T_{REF}})$, using a factor purely defined at simulation level, as:

$$SF(p_T) = SF_{DATA}(p_{T_{REF}}) \times \frac{SF_{SIM}(p_T)}{SF_{SIM}(p_{T_{REF}})} \tag{4.6}$$

The error on the scale factor, $SF(p_T)$, is obtained by applying the rules for the error propagation, where statistically independent errors (such those obtained using data and simulation) are summed in quadrature and statistically dependent errors (such as those obtained fully from simulation) are summed directly. The result is:

[3]The detailed explanation on *incremental* uncertainties follows shortly.

$$\Delta SF^2(p_T) = \left|\frac{\partial SF(p_T)}{\partial SF_{DATA}(p_{T_{REF}})}\right|^2 \Delta SF^2_{DATA}(p_{T_{REF}}) + \\ + \left[\frac{\partial SF(p_T)}{\partial SF_{SIM}(p_T)}\Delta SF_{SIM}(p_T) + \frac{\partial SF(p_T)}{\partial SF_{SIM}(p_{T_{REF}})}\Delta SF_{SIM}(p_{T_{REF}})\right]^2 \quad (4.7)$$

thus:

$$\Delta SF^2(p_T) = \left|\frac{SF_{SIM}(p_T)}{SF_{SIM}(p_{T_{REF}})}\right|^2 \Delta SF^2_{DATA}(p_{T_{REF}}) + \\ + \left[SF_{DATA}(p_{T_{REF}})\left(\frac{1}{SF_{SIM}(p_{T_{REF}})}\Delta SF_{SIM}(p_T) - \frac{SF_{SIM}(p_T)}{SF^2_{SIM}(p_{T_{REF}})}\Delta SF_{SIM}(p_{T_{REF}})\right)\right]^2 \quad (4.8)$$

Fixing the value of the scale factor as previously discussed: $SF_{DATA}(p_{T_{REF}}) = SF_{SIM}(p_T) = SF_{SIM}(p_{T_{REF}})$ and simplifying the equation accordingly, yields:

$$\Delta SF^2(p_T) = \Delta SF^2_{DATA}(p_{T_{REF}}) + [\Delta SF_{SIM}(p_T) - \Delta SF_{SIM}(p_{T_{REF}})]^2 \quad (4.9)$$

In the formula above $\Delta SF(p_T)$ is the total uncertainty on the scale factor in the extrapolation p_T range, $\Delta SF_{DATA}(p_{T_{REF}})$ is the (data-based) total uncertainty associated to the scale factor at $p_{T_{REF}}$ which is set to 400 GeV, and the final term $[\Delta SF_{SIM}(p_T) - \Delta SF_{SIM}(p_{T_{REF}})]$ is the incremental simulation-based uncertainty. Therefore, in each m-th bin of the extrapolation one has:

$$(\sigma^m)^2 = (\sigma^{400\,GeV}_{data})^2 + \sum_i \left[\delta^m_i - \delta^{400\,GeV}_i\right]^2 \quad (4.10)$$

where i runs on the different sources of simulation-based uncertainties. Finally, in the case when $\sigma^m < \sigma^n$ if $m > n$, the monotonic behaviour is enforced by fixing $\sigma^m = \sigma^n$,

to prevent artificial reductions of the uncertainties due to statistical fluctuations in simulated events.

In each p_T bin, the incremental MC-based uncertainties are calculated as the difference between the δ_i^m and the respective value at 400 GeV (δ_i^{400}). The result of this procedure is presented in Figure 4.10 above 400 GeV, where $\sigma_{data}^{400\,GeV}$ is marked by the light green band, while the incremental extrapolation uncertainty is in dark green, for the $\epsilon_b = 70\%$ working point.

The δ_i^m are determined as follows. In each of the m bins of the calibration, the measurement of the δ_i is performed following a three-step procedure. Firstly, the b-tagging efficiency for a given working point of the `MV2` b-tagging algorithm ($\epsilon_b^N(p_T)$) is measured in simulation. Subsequently, underlying quantities that are known to affect the b-tagging performance are *varied* in the simulation one by one and the b-jet tagging efficiency ($\epsilon_b^{v_i}(p_T)$) is recomputed in each case. Finally, the size of each contribution to the total uncertainty is determined as the difference between the nominal and variation efficiency. The total systematic uncertainties are then extracted as the sum in quadrature of the contributions from each independent source.

4.4.2 Samples, event and object selection

The study presented here uses simulated events reproducing 2017 data-taking conditions. In order to perform the study at p_T regimes above 400 GeV, a $Z' \to l\bar{l}$, $c\bar{c}$, $b\bar{b}$ sample is used, where the branching fractions of the decays are set to be one-third for each flavour (b, c, or light-flavour) of quark pairs, to give a p_T spectrum uniformly populated in jets of all flavours. The Z' events are simulated in proton-proton collisions at a centre-of-mass energy of 13 TeV with Pythia8 [144] and the cross-section of the hard scattering process is modified to broaden the natural width of the Z' resonance, resulting in a sample enriched in jets with transverse momenta up to the TeV scale. Track-related and jet-related systematic uncertainties are evaluated using

this sample (see Section 4.4.3).

In addition, samples with $t\bar{t}$ events from proton-proton collisions at a centre-of-mass energy of 13 TeV simulated with POWHEG [145] + PYTHIA8 [144] are included. Modelling systematic uncertainties are assessed on $t\bar{t}$ events simulated using an alternative MC generator, HERWIG7 [146] (as explained in Section 4.4.3). Only $t\bar{t}$ events with at least one leptonically-decaying W boson (from $t \to bW$) are considered in the study.

The ATLAS detector Geant4 [92] simulation and reconstruction software [147] is run on the simulated samples.

In order to exploit the statistical power of the samples, a very light selection is applied to events, based on the kinematic properties of jets and (jet-associated) tracks. Jets are reconstructed using the anti-k_T algorithm [148], with a radius parameter $R = 0.4$ and the selection applied requires events to contain jets with $p_T > 20$ GeV and $|\eta| < 2.5$, and tracks with $p_T > 500$ MeV and within the tracker acceptance in pseudorapidity ($|\eta| < 2.5$). Tracks are associated to the nearest jet by exploiting their angular separation, as seen in Section 4.2.1.

4.4.3 Systematic uncertainties

The uncertainties affecting the b-tagging efficiency considered for this study [1] are mainly of four types and consist of those affecting the reconstruction of tracks, of jets, the modelling of the b-hadrons and finally the interaction of long-lived b-hadrons with the detector material.

The first three types will be considered in detail in the following, while the interaction of long-lived b-hadrons with the detector material has been studied separately [149] and will be briefly described here. The origin of the latter uncertainty lies on the fact that only the interactions of b-hadron decay products with the detector material are simulated and not those of b-hadrons themselves. However, at increasingly high

p_T, an increasing fraction of b-hadron decays after the innermost layer of the pixel detector[4], which provides critical measurements for the b-tagging algorithms, and therefore differences in the b-jet tagging efficiency at high p_T are expected. The impact of this uncertainty on the MV2 efficiency is evaluated by comparing the b-tagging response in the standard ATLAS simulation, with a different simulation model, where the interaction of the detector material with the b-hadrons is added. The impact of this uncertainty is found to be 1% at 400 GeV increasing to 4% for jets $p_T > 1.5$ TeV.

Finally, at this stage no dedicated studies of samples enriched in high-energy b-jets, where collimated tracks from displaced decay vertices concur to create a challenging environment for the track reconstruction algorithm, are included.

4.4.3.1 Track-related uncertainties

As discussed in Section 4.2, track properties are of fundamental importance in the process of b-tagging the jets matched to them. Therefore, the sources of uncertainties that play a major role on the track reconstruction, must be examined.

Technically, the track-related systematic uncertainties are treated as nuisance parameters to be evaluated at different points of the nuisance parameter space: 0 corresponds to the *nominal* (unvaried) case and then the $\pm 1\sigma$ variations respectively are selected. This produces a set of varied events to which the MV2 algorithm is applied and for which the b-tagging efficiency is calculated. Finally, the difference in efficiency with respect to the unvaried case is extracted, to determine the size of tracking systematic uncertainties $\delta_{TRK,i}$.

The most significant sources of systematic uncertainties related to the tracking performance considered in the study can be grouped based on the quantities they affect:

Track reconstruction efficiency

[4]The fraction goes from about 5% of the b-hadrons within b-jets at jet $p_T = 150$ GeV and rises to be around 30% beyond jet $p_T = 500$ GeV.

This group of uncertainties assesses the knowledge of the inner detector (ID) and includes the effect of alignment, the presence of dead modules and additional material not accurately modelled in the simulation. It also tests different physics models used in the GEANT4 detector simulation. These differ in their description of the interaction between high-energy particles and nucleons.

Rate of fake tracks

Fake tracks are those reconstructed in the ID due to combinatorial effects, that cannot be matched to physical particles such as the b-hadrons. An estimate is made by assuming that the number of non-fake tracks is to first order proportional to the number of pile-up interactions (see Section 2.1.4), while any deviation from such linearity is considered to be due to fakes.

The uncertainty on the rate of fake tracks is obtained by varying the number of fake tracks within the uncertainties from dedicated studies, and the impact on the b-tagging efficiency is studied by randomly removing fake tracks in the simulation.

Tracks In Dense Environments (TIDE) [150, 96]

This uncertainty affects tracks within the core of high p_T jets, where the local track density is high, presenting particular challenges for the track reconstruction. Two additional uncertainties are considered, to take into account the different track reconstruction efficiency as well as the impact of fake tracks in such environment.

Impact parameter resolution

These uncertainties account for the difference in resolution between simulation and data on the transverse (d_0) and longitudinal (z_0) impact parameters (IPs).

4.4.3.2 Jet-related uncertainties

Jet-related systematic uncertainties $\delta_{JET,i}$ are extracted following the method explained above, where the impact on the b-tagging efficiency of the $\pm 1\sigma$ variation corresponding to each source is evaluated as the difference with respect to the nominal case. The uncertainties involve the *Jet Energy Scale* (JES) and the *Jet Energy Resolution* (JER). As described in Section 3.3.4.1, the first one is associated to the calibration of the response of the calorimeter. The latter, instead, is associated to the resolution in energy of the calorimeter.

4.4.3.3 Modelling uncertainties

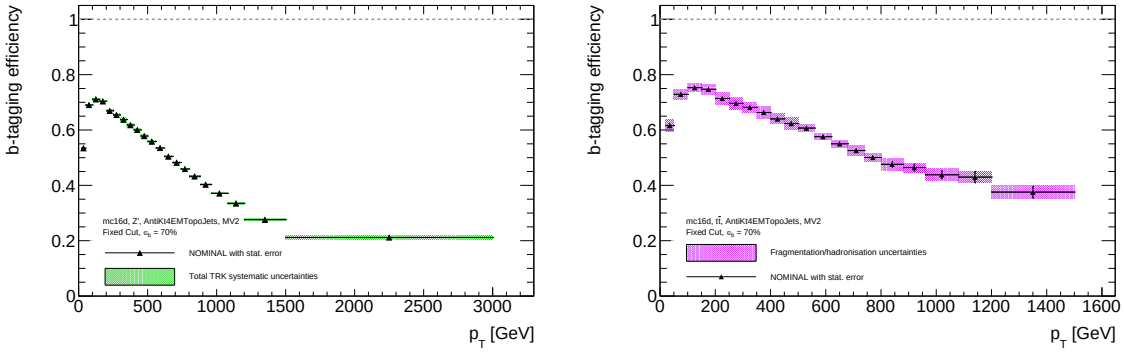
The third group of systematic uncertainties affects the simulation of the parton showers and the b-quark fragmentation function (see Section 1.5.1). The effect of this uncertainty (δ_{MOD}) is evaluated by comparing the b-tagging efficiency on $t\bar{t}$ events simulated with POWHEG+PYTHIA8 (nominal) and POWHEG+HERWIG7 (variation).

4.4.4 Results

The results are only shown for the $\epsilon_b = 70\%$ working point, which is widely used within the ATLAS collaboration. However, a full set of results for all the b-tagging working points was made available for the collaboration.

Another full set of results is derived for variable-radius (VR) track jets, explained in detail in Section 3.3.4.3, and used in the $VH(H \to b\bar{b})$ analysis at high p_T in Chapter 5. VR track jets are reconstructed as clusters of tracks (and not as calorimeter clusters), therefore track-related uncertainties take the place of the jet-related ones. For this reason, the results produced for this jet-topology only include track and modelling uncertainties (from Z' and $t\bar{t}$ events respectively).

Two examples of the b-tagging efficiencies can be found for reference in Figure 4.12. Figure 4.13 and Table 4.3 show the effect of systematic uncertainties on the



(a) (b)

Figure 4.12: b-tagging efficiencies for the `MV2` algorithm on Z' (4.12a) and $t\bar{t}$ (4.12b) events, including track-related and modelling uncertainties respectively.

	400 - 620 GeV	620 - 880 GeV	880 - 1080 GeV	1080 - 1200 GeV	1200 - 1500 GeV	1500 - 3000 GeV
δ_{TRK}	0.5% - 1%	0.5% - 1%	0.5% - 1%	1.2%	1.2%	4%
δ_{JET}	0.5%	1%	2%	2%	4%	7%
δ_{MOD}	3% - 4%	3% - 4%	5%	5%	6%	//
δ_{tot}	3.1% - 4.2%	3.2% - 4.2%	5.4% - 5.5%	5.5%	7.3%	$> 8.1\%$

Table 4.3: Summary of simulation-based uncertainties on the b-tagging efficiency as function of the b-jet transverse momentum on calorimeter jets, for the $\epsilon_b = 70\%$ operating point.

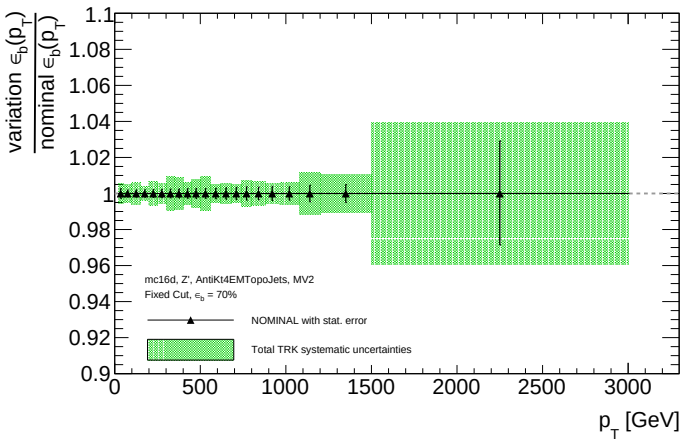
b-tagging efficiencies on calorimeter jets, as a function of the b-jet p_T.

Modelling uncertainties dominate across the whole p_T range, except for the very last bin, from 1.5 TeV to 3 TeV, where they cannot be calculated due to lack of statistics in the $t\bar{t}$ sample. Jet-related uncertainties are consistently $\leq 2\%$ up to 1.2 TeV. Finally, track-related uncertainties are ≤ 1.2 %, up to 1.5 TeV in p_T and give the smallest contribution to the total uncertainties.

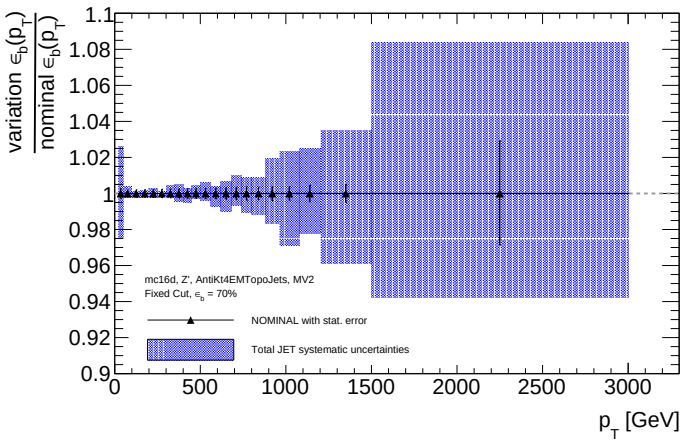
The total uncertainties are of 3.1% at b-jet $p_T = 400$ GeV and $> 8.1\%$ at 3 TeV.

After examining simulation-based uncertainties for calorimeter jets, results for VR track jets are presented for the same working point: $\epsilon_b = 70\%$. Figure 4.14 and Table 4.4 summarise the effect of systematic uncertainties on the b-tagging efficiencies on VR track jets, as a function of the b-jet p_T.

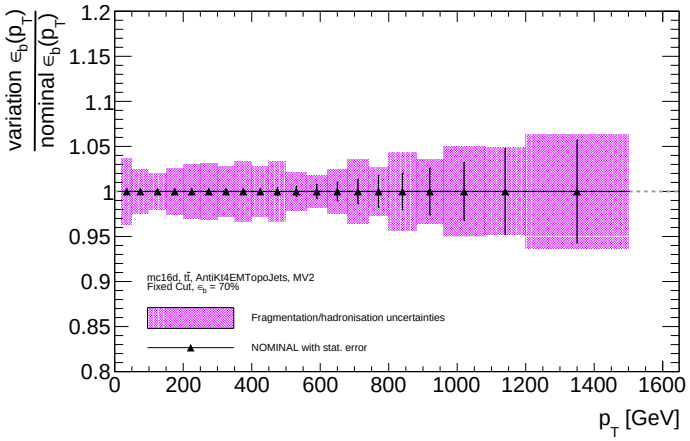
Also in this case modelling uncertainties dominate at any p_T range, up to 1.5



(a)

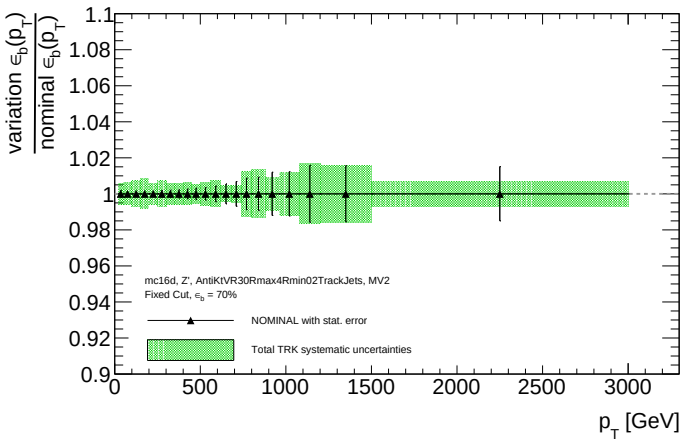


(b)

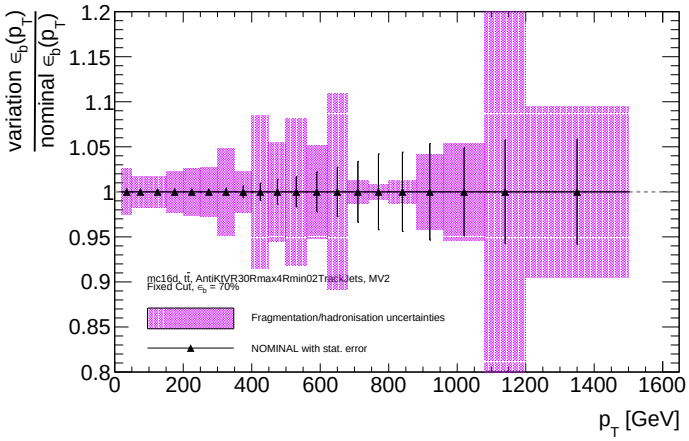


(c)

Figure 4.13: Impact of the track-related (4.13a), jet-related (4.13b) and modelling (4.13c) systematic uncertainties on the b-tagging efficiency, as function of the b-jet transverse momentum for calorimeter jets in Z' and $t\bar{t}$ events, for the $\epsilon_b = 70\%$ operating point. The ratio between variation and nominal efficiencies is shown, in order to highlight the size of the uncertainties.



(a)



(b)

Figure 4.14: Impact of the track-related (4.14a) and modelling (4.14b) systematic uncertainties on the b-tagging efficiency, as function of the b-jet transverse momentum for variable-radius track jets in Z' and $t\bar{t}$ events, for the $\epsilon_b = 70\%$ operating point. The ratio between variation and nominal efficiencies is shown, in order to highlight the size of the uncertainties.

	400 - 680 GeV	680 - 740 GeV	740 - 1080 GeV	1080 - 1200 GeV	1200 - 1500 GeV	1500 - 3000 GeV
δ_{TRK}	0.5%	0.5%	1.5%	2%	2%	$< 1\%$
δ_{MOD}	5% - 11%	2% - 6%	2% - 6%	20%	9%	//
δ_{tot}	5% - 11%	2.1% - 6%	2.5% - 6.2%	20%	9.2%	$> 1\%$

Table 4.4: Summary of simulation-based uncertainties on the b-tagging efficiency as function of the b-jet transverse momentum on variable-R track jets, for the $\epsilon_b = 70\%$ operating point.

TeV, while, track-related uncertainties are ≤ 2 %. Given the low data statistics at very high p_T regimes, modelling uncertainties (and therefore the total uncertainty) are effectively ∼6%-10% between 400 GeV and 1 TeV and ∼10%-20% above 1 TeV.

The calibration of b-jets at very high transverse momentum described in this Section (and included in [1]) laid the foundation for the calibration of b-jets at high transverse momentum, by providing additional uncertainties to the scale factors used by all the physics analyses exploiting b-tagging algorithms in the ATLAS collaboration. The study has since extended to other taggers and jet topologies.

Possible developments of the calibration, beyond the scope of my DPhil, involve the validation of the study at low p_T through the data-based efforts and an expansion of the study of modelling uncertainties, including additional sources such as initial and final state radiation. Further interesting improvements on the b-tagging of high-p_T jets include the study of non-isolated b-jets, such as those coming from the decay of highly boosted objects.

Chapter 5

Measurement of $VH(H \to b\bar{b})$ at high transverse momentum

The discovery of the Higgs boson [10, 11] in 2012 by the ATLAS and the CMS Collaborations was a major achievement in the understanding of the SM. Since then, substantial progress has been made in the measurement of the properties of the Higgs boson.

The observation of the decay of the Higgs boson to a $b\bar{b}$ pair is one of the most sought-after results of the LHC Run 2, since in the SM this decay mode has the largest branching ratio (BR $\sim$ 58%) [29]. The $H \to b\bar{b}$ decay is also the only process for which it is possible to measure the Yukawa coupling to down-type quarks, given the current experimental sensitivity. However, this measurement has several challenges at hadron colliders, the biggest of which is the small signal-to-background ratio, the main background being from non-resonant production of quarks and gluons (*QCD processes*).

The associated production of a Higgs boson and a vector boson (VH, $V = W^{\pm}, Z$) is critical for the measurement of $H \to b\bar{b}$. Despite having a cross section of only $\sigma_{VH} = 2.25$ pb, smaller than the ggF cross section $\sigma_{ggF} \sim 49$ pb for a Higgs boson

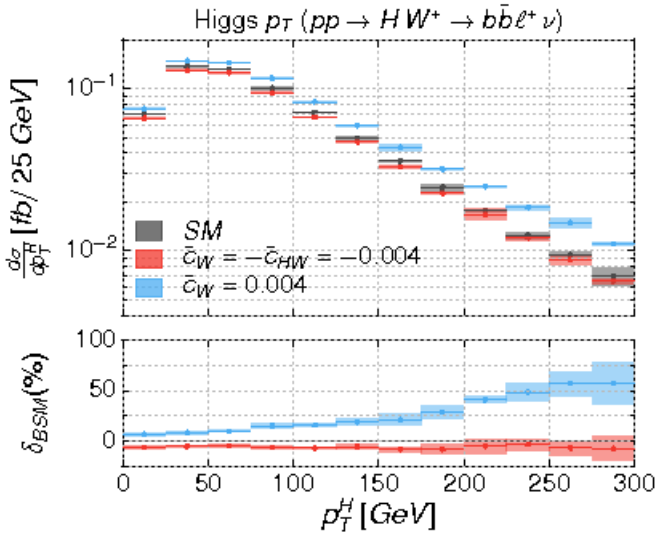


Figure 5.1: Comparison of differential distributions in the SM (black) and two EFT benchmarks (red and light blue), as a function of the Higgs boson p_T for $WH \to \ell\nu b\bar{b}$. Image source: [152].

of 125.09 GeV at $\sqrt{s} = 13$ TeV [29], the vector boson leptonic decay modes yield distinctive signatures that can be efficiently selected, while suppressing most of the background from QCD processes.

The $H \to b\bar{b}$ and VH processes were first observed in 2018 [151], and $VH(H \to b\bar{b})$ was measured to be consistent with the SM with a statistical significance of 4.9 standard deviations. The following step is to study the production cross section as a function of the transverse momentum (p_T) of the Higgs boson, since several BSM theories predict an enhancement of $VH(H \to b\bar{b})$ events at high p_T^H or p_T^V [152], as shown in Figure 5.1.

In this chapter, an overview of the $VH(H \to b\bar{b})$ analysis at high transverse momentum, the so-called "$VH(H \to b\bar{b})$ *boosted* analysis" of the ATLAS Experiment is presented [2]. The analysis is the high-p_T partner of the "$VH(H \to b\bar{b})$ *resolved* analysis" [3], described in Section 6.1.

After introducing the analysis strategy in Section 5.1, the data and simulation samples are summarised in Section 5.2. Then, the object and event selection and

categorisation used in the analysis are discussed in Section 5.3, while an excursus on the feasibility of substructure studies can be found in Section 5.4. The statistical analysis, with the simultaneous $VH - VZ$ fit is provided in Section 5.5. Section 5.6 is dedicated to the discussion of systematic uncertainties: along with the experimental ones, a detailed description of the modelling of the background and signal physics processes is given, with a particular focus on top-pair production. Finally, the results of the $VH(H \to b\bar{b})$ boosted analysis are given in Section 5.7, while in Section 5.8 an overview of the STXS measurement and its EFT interpretation is provided. All the plots and tables are either taken from the publication [2] (and therefore present the **ATLAS** label), or mine.

5.1 Analysis strategy

Three different signatures of vector boson decays are targeted by the $VH(H \to b\bar{b})$ boosted analysis, as shown in Figure 5.2. These are: $ZH \to \nu\bar{\nu}b\bar{b}$ (*0-lepton channel*), $WH \to \ell\nu b\bar{b}$ (*1-lepton channel*) and $ZH \to \ell^-\ell^+ b\bar{b}$ (*2-lepton channel*).

A Higgs boson decaying at high p_T^H is highly Lorentz-boosted. Therefore its decay products become so collimated that standard (i.e. *resolved*) reconstruction techniques begin to fail, as sketched in Figure 5.3. At high p_T^H, the angular separation of the decay products of the Higgs boson, b_1 and b_2, can be written as:

$$\Delta R\,(b_1, b_2) = 2\frac{m_H}{p_T^H} \tag{5.1}$$

where m_H is the mass of the decaying Higgs boson. Therefore, from $p_T^H > 250$ GeV, the Higgs boson candidate can be reconstructed as a single object (large-R jet, R=1.0, see Section 3.3.4.2) with a two-prong substructure.

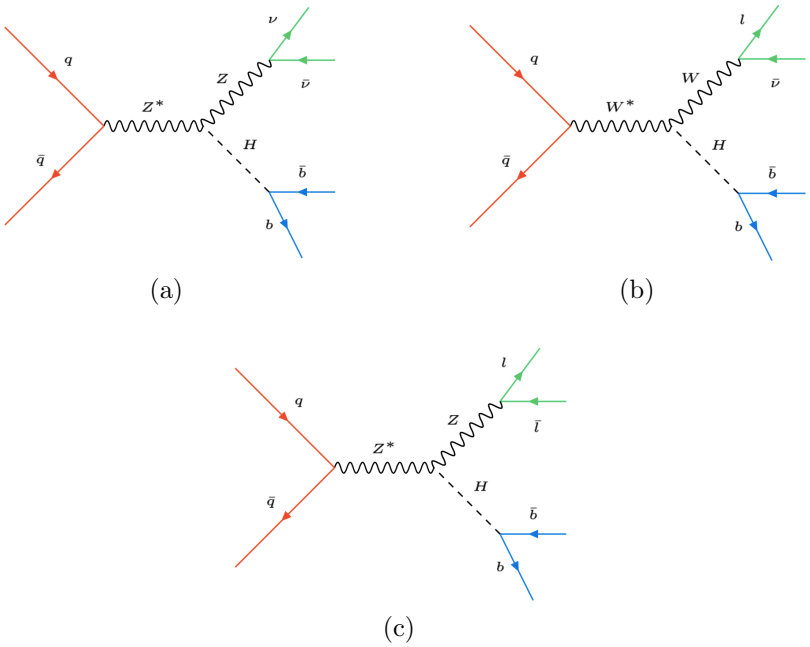


Figure 5.2: Feynman diagrams for the LO quark initiated SM $VH(H \to b\bar{b})$ processes, with V decaying leptonically, in the 0-lepton channel (5.2a), 1-lepton channel (5.2b) and 2-lepton channel (5.2c).

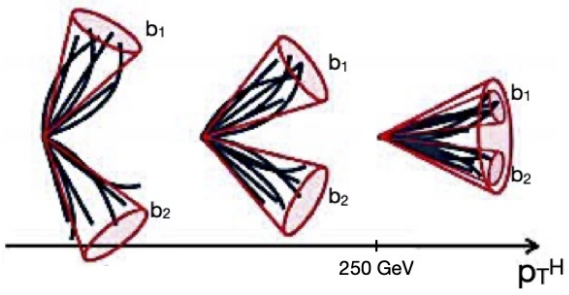


Figure 5.3: Sketch of the $H \to b\bar{b}$ decay topology, as a function of the Higgs boson transverse momentum.

5.2 Data and Simulation Samples

The $VH(H \to b\bar{b})$ boosted analysis is performed using the pp collision data collected by the ATLAS experiment in the full LHC Run 2, from 2015 to 2018, corresponding to an integrated luminosity of 139 fb^{-1} at a centre of mass energy $\sqrt{s} = 13$ TeV.

The majority of the SM background processes, as well as the $VH(H \to b\bar{b})$ signal are simulated with state-of-the-art Monte Carlo (MC) simulation packages such as Powheg [153], Pythia8 [144] and Sherpa [154]. All samples of simulated events are generated at least at next-to-leading-order (NLO). After the generation of the events, all samples are then interfaced with the ATLAS detector simulation [91] based on Geant4 [92].

All simulated processes are normalised using the most precise theoretical predictions currently available for their differential cross-sections. Table 5.1 summarises the signal and background processes simulated for the analysis, with the corresponding matrix element and parton shower generators used for each production.

A wide range of background processes enters the phase space of the $VH(H \to b\bar{b})$ boosted analysis, and the three channels show a very diverse background composition. A brief description of the signal and background processes (shown in Figure 5.4) is given in the following, while more details are given in Section 5.6.2.

Signal $VH(H \to b\bar{b})$

VH production at the LHC occurs through the annihilation of a quark-antiquark pair (*quark-initiated* diagrams), as shown in Figure 5.2. Additionally, at NLO, ZH production can be initiated by gluons, via a loop of top quarks. Both the quark- and gluon-induced processes contributing to the VH production have been taken into account, with the mass of the Higgs boson fixed at 125 GeV.

Top quark pair production $t\bar{t}$

Top quark pair production ($t\bar{t}$) is one of the most abundant background pro-

Process	Generator
Signal ($m_H = 125$ GeV, BR($H \to b\bar{b}$)= 58%)	
$qq \to WH \to \ell\nu b\bar{b}$	Powheg+Pythia8
$qq \to ZH \to \nu\nu b\bar{b}/\ell\ell b\bar{b}$	Powheg+Pythia8
$gg \to ZH \to \nu\nu b\bar{b}/\ell\ell b\bar{b}$	Powheg+Pythia8
Top quark ($m_t = 172.5$ GeV)	
$t\bar{t}$	Powheg+Pythia8
single top s-channel	Powheg+Pythia8
single top t-channel	Powheg+Pythia8
single top Wt	Powheg+Pythia8
Vector boson + jets	
$W \to \ell\nu$	Sherpa
$Z/\gamma^* \to \ell\ell$	Sherpa
$Z \to \nu\nu$	Sherpa
Diboson	
$qq \to WW$	Sherpa
$qq \to WZ$	Sherpa
$qq \to ZZ$	Sherpa
$gg \to VV$	Sherpa

Table 5.1: Signal and background processes with the corresponding generators used for the nominal samples. Table adapted from [2].

cesses in the 0-lepton and 1-lepton channels. In these channels a dedicated analysis region is constructed to estimate this process.

Single top

Single top is only present in the 0-lepton and 1-lepton channels, where it is significantly smaller with respect to $t\bar{t}$.

Vector boson + jets production (V+jets)

Along with $t\bar{t}$ production, vector bosons (W or Z) produced in association with jets are the other dominant background in the $VH(H \to b\bar{b})$ boosted analysis. *Z+jets* events can be found in the 0-lepton and 2-lepton channels, while *W+jets* are present in the 0-lepton and 1-lepton channel.

Diboson production

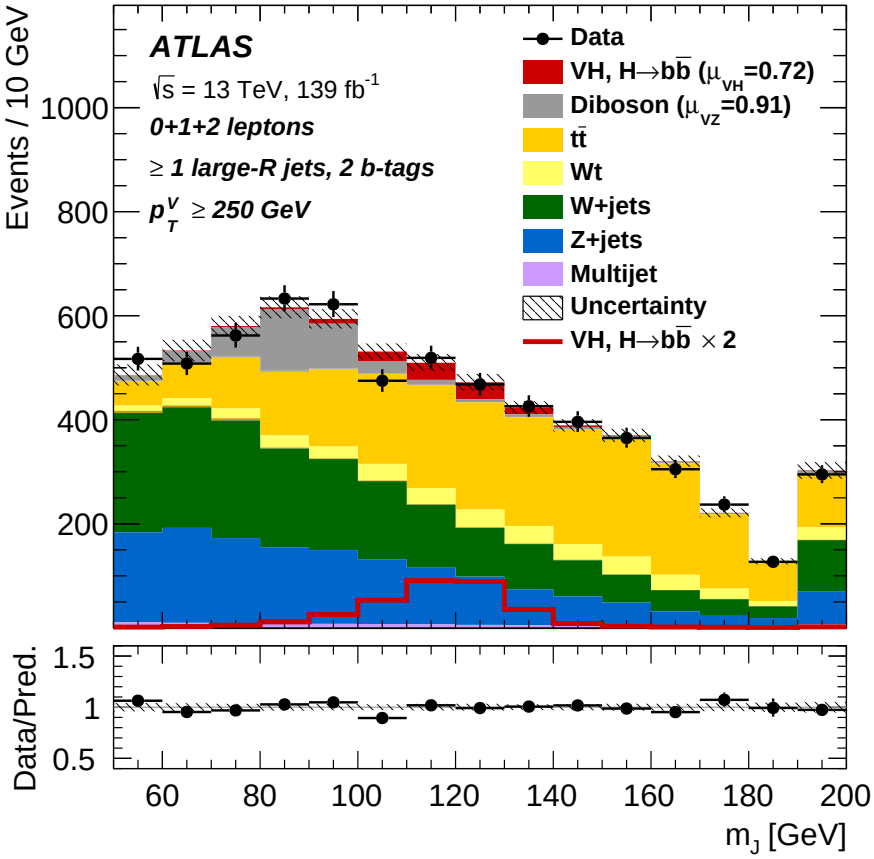


Figure 5.4: Large-R jet mass distribution combining all three lepton channels and regions. The Higgs boson signal is shown as a filled histogram in red color, normalised to the signal yield extracted from data (see Section 5.7). The size of the combined statistical and systematic uncertainty for the fitted signal and background is indicated by the hatched band.

The diboson background processes are composed of three distinct processes: WZ, ZZ and WW, with the first two that are more abundant than the third. They can be found in all three lepton channels.

Multijet background

The multi-jet background, produced by non-resonant QCD processes is largely suppressed by the requirement on the leptonic V decay, as well as the analysis selection criteria explained in Section 5.3.2. The residual contribution is only non-negligible in the 1-lepton channel, where it is evaluated using data-driven methods.

5.3 Object and Event Selection and Categorisation

The selection criteria of the $VH(H \to b\bar{b})$ boosted analysis aim at reconstructing a boosted Higgs boson candidate, associated to a leptonically decaying V. The number of stable[1] charged leptons in the leptonic V decay determines the analysis channel for which dedicated selection criteria, based on the kinematics and topology of signal events, are applied.

The reconstruction of the *Higgs boson candidate* as a large-R jet with a substructure of two b-tagged track jets matched to it defines the analysis and is therefore common to all three channels.

Finally, additional selection criteria are applied to suppress background events according to their expected topology.

5.3.1 Object selection

The definition of the physics objects is provided in Chapter 3. In the following the list of all the objects, along with their specific requirements, used in the $VH(H \to b\bar{b})$ boosted analysis is presented.

Electron and muons

Two different selections are applied to electrons and muons. The first one is used to define the three main analysis channels (0-, 1-, and 2-lepton), while the second one, more stringent, is used to define the requirements for the electrons and muons originating from the W and Z decay in the 1-lepton and the 2-lepton channel respectively.

The definition of the analysis channel is based on so-called *baseline* electrons and

[1] Electrons and muons, i.e. the leptons that are stable within the dimensions of the ATLAS experiment. τ leptons decay within the ATLAS experiment and are therefore not considered as stable.

	p_T	η	$\lvert d_0/\sigma(d_0)\rvert$	$\lvert z_0 \sin(\theta)\rvert$
Electrons	> 7 GeV	< 2.47	< 5	< 0.5 mm
Muons	> 7 GeV	< 2.7	< 3	< 0.5 mm

Table 5.2: Selection criteria for *baseline* electrons and muons used in the definition of the analysis lepton channel for the $VH(H \to b\bar{b})$ boosted analysis. Additionally, electrons are required to be isolated from other tracks and energy deposit clusters, and to pass loose likelihood selection criteria based on shower shapes (see Section. 3.3.1). Muons, on the other hand, are selected using the "loose" identification criterion of Ref. [107] and a loose track isolation.

	p_T	η	$\lvert d_0/\sigma(d_0)\rvert$	$\lvert z_0 \sin(\theta)\rvert$
Electrons				
1-lepton channel	> 27 GeV	< 2.47	< 5	< 0.5 mm
2-lepton channel	> 27 GeV	< 2.47	< 5	< 0.5 mm
Muons				
1-lepton channel	> 25 GeV	< 2.5	< 3	< 0.5 mm
2-lepton channel	> 27 GeV	< 2.5	< 3	< 0.5 mm

Table 5.3: Selection criteria for *signal* electrons and muons used for the identification of leptons from the W and Z decays in the $VH(H \to b\bar{b})$ boosted analysis. Additionally with respect to baseline electrons, signal electrons are selected using a tighter likelihood requirement, which also includes tracking and track-cluster matching variables, as well as using a tighter calorimeter-based isolation criterion (see Section. 3.3.1). Signal muons, on the other hand, are selected using the "medium" identification criterion of Ref. [107] and to have a tighter track isolation criterion.

muons, that are required to be *exactly* zero, one or two. Table 5.2 summarises the *baseline* selection criteria applied to electrons and muons, respectively.

The second type of more stringent criteria is used to define *signal* electrons and muons. A tighter selection ensures a higher rejection rate for QCD processes. Table 5.3 summarises the *signal* selection criteria applied to electrons and muons, respectively.

Jets

The $VH(H \to b\bar{b})$ boosted analysis uses three different anti-k_T jet collections, all explained in detail in Section 3.3.4: large-R jets with R=1.0 (see Section 3.3.4.2), small-R jet with R = 0.4 (see Section 3.3.4.1) and VR track-jets (see Section 3.3.4.3).

Large-R jets are used to reconstruct the Higgs boson candidate at high p_T^H. Given the relation in Equation 5.1, only large-R jets with $p_T > 250$ GeV are considered, in this way the angular separation between the two Higgs boson decay products is $\Delta R < 1.0$ and the full Higgs boson decay can be contained within the large-R jet. Large-R jet $|\eta| < 2.0$ is also required.

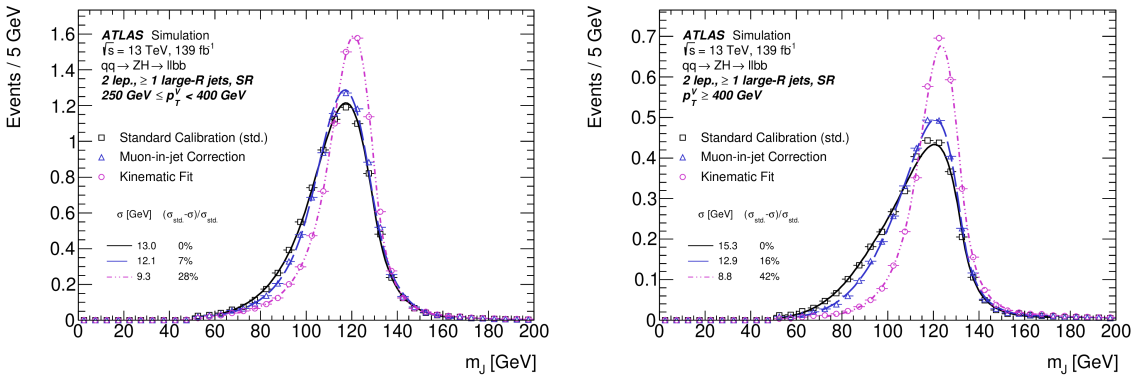
Since the *discriminant variable* used in the analysis is the large-R jet mass m_J, as shown in Section 5.5 the *combined mass* definition (Section 3.3.4.2) is used, which achieves the best performance. Additionally, $m_J > 50$ GeV is required.

Two corrections are applied to the large-R jet mass, in order to better set the scale of m_J and to improve its resolution. Since the targeted Higgs boson decay is $H \to b\bar{b}$, the presence of b-hadrons (and c-hadrons, from the b-hadron decay) is expected within the large-R jet. In Section 4.1 it has been explained how this type of hadrons have a large BR for semi-leptonic decays. The first correction accounts for the presence of muons from the b- or c-hadron semi-leptonic decays. Since these muons are not included in the "standard" large-R jets, this results in losses in the large-R jet energy. This correction, called *muon-in-jet*, adds back the four-momentum of the muon associated to a large-R jet.

The second correction is applied only in the 2-lepton channel, and improves the large-R jet energy resolution. It consists in performing a *kinematic fit* that constrains the $\ell^+\ell^- b\bar{b}$ system to be balanced in the transverse plain, and sets the dilepton mass to the Z boson mass ($m_Z \sim 91.2$ GeV [27]).

The large-R jet mass resolution improves by 5% to 10% after the first correction (depending on the lepton channel), while the second correction brings an additional improvement in the 2-lepton channel of up to 40%. The effect of these corrections is shown in Figure 5.5.

Small-R calorimeter jets are used in the $VH(H \to b\bar{b})$ boosted analysis to build the event E_T^{miss}, to estimate the multi-jet contribution in the 1-lepton channel, and



(a) (b)

Figure 5.5: Large-R jet mass comparison as additional corrections are applied to the jet energy scale for the signal $VH(H \to b\bar{b})$ in the 2-lepton channel and 250 GeV $\leq p_T^V <$ 400 GeV (5.5a) and $p_T^V \geq$ 400 GeV (5.5b) regions for the $qq \to ZH$ contribution.

for the event categorisation. Small-R jets are required to have $p_T > 30$ GeV and $|\eta| < 4.5$. Additionally, in order to reduce the number of small-R jets originating from pile-up interactions, small-R jets are required to pass the JVT requirement (see Section 3.3.4.1) if they are in the range $p_T < 120$ GeV and $|\eta| < 2.5$.

The final jet collection used in the $VH(H \to b\bar{b})$ boosted analysis is that of VR track-jets. VR track-jets are used in the analysis as the input of the b-tagging algorithm (described in Chapter 4), to correctly identify jets originating from the $H \to b\bar{b}$ decay. Only central ($|\eta| < 2.5$) VR track-jets with $p_T > 10$ GeV and with at least two tracks are considered.

5.3.2 Event selection

Common to all three lepton channels is the Higgs boson candidate selection. As it has already been discussed above, the Higgs boson is reconstructed as a large-R jet, with $p_T > 250$ GeV, $|\eta| < 2.0$ and $m_J > 50$ GeV. For events with more than one large-R jets, the one with the higher p_T is chosen as the Higgs boson candidate. Subsequently, the $H \to b\bar{b}$ decay is targeted by requiring the presence of at least two VR track-jets

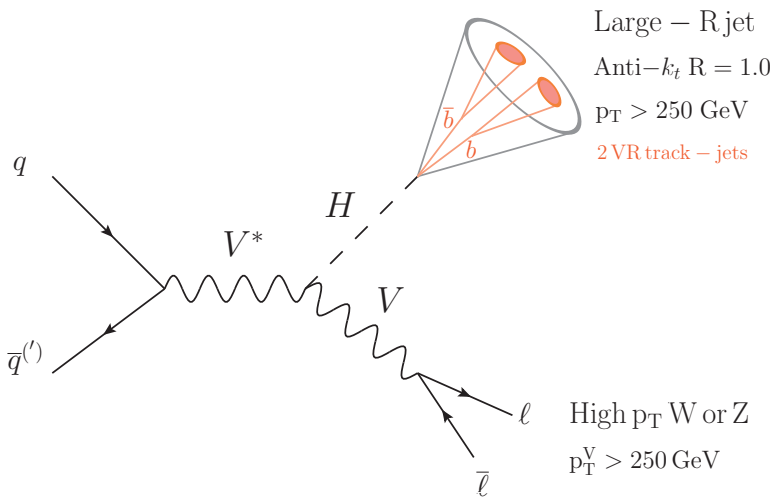


Figure 5.6: Sketch of the $VH(H \to b\bar{b})$ process at high Higgs boson transverse momentum.

with $p_T > 10$ GeV and $|\eta| < 2.5$, that are ghost-associated (see Section 3.3.4.3) to the large-R jets. Additionally, the two higher-p_T (*leading*) VR track jets are required to be b-tagged. The b-tagging selection in the analysis uses the ϵ_b =70% working point (see Section 4.3.1) of the `MV2` algorithm.

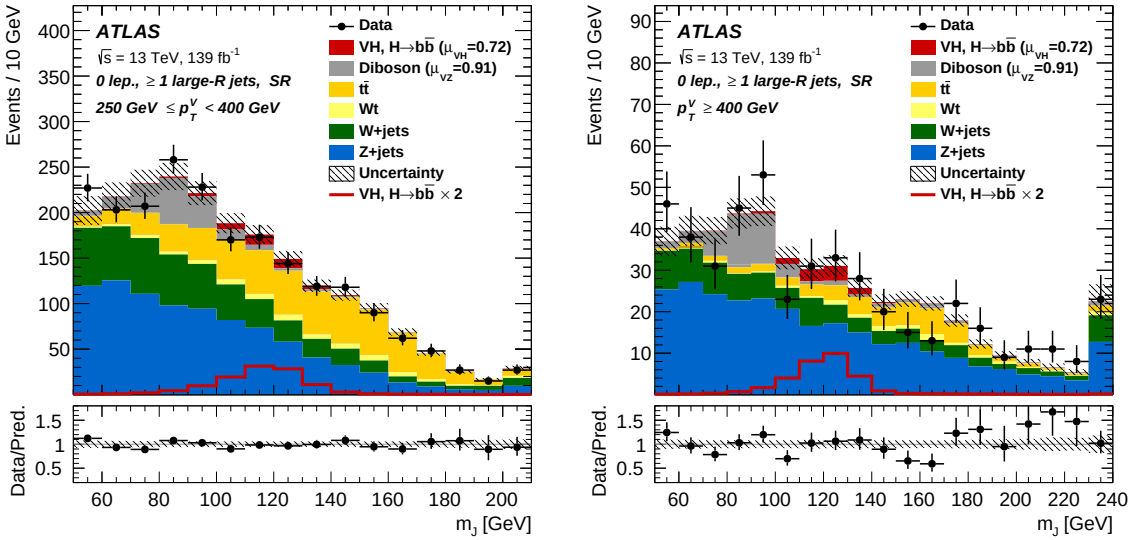
All three lepton channels require $p_T^V > 250$ GeV. In the $VH(H \to b\bar{b})$ boosted analysis p_T^V is often used as a proxy for p_T^H, since the resolution achieved for p_T^V, which is estimated using lepton momenta, is better than the p_T^H resolution, from the momentum resolution of the large-R jet. The p_T^V requirement also ensures that the triggers are fully efficient.

A sketch of the boosted $VH(H \to b\bar{b})$ process is shown in Figure 5.6.

The event selection criteria for each lepton channel are now presented.

0-lepton channel

Events in the 0-lepton channel (see Figure 5.7) are selected by a missing transverse energy E_T^{miss} trigger with thresholds varying from 70 GeV to 110 GeV, depending on



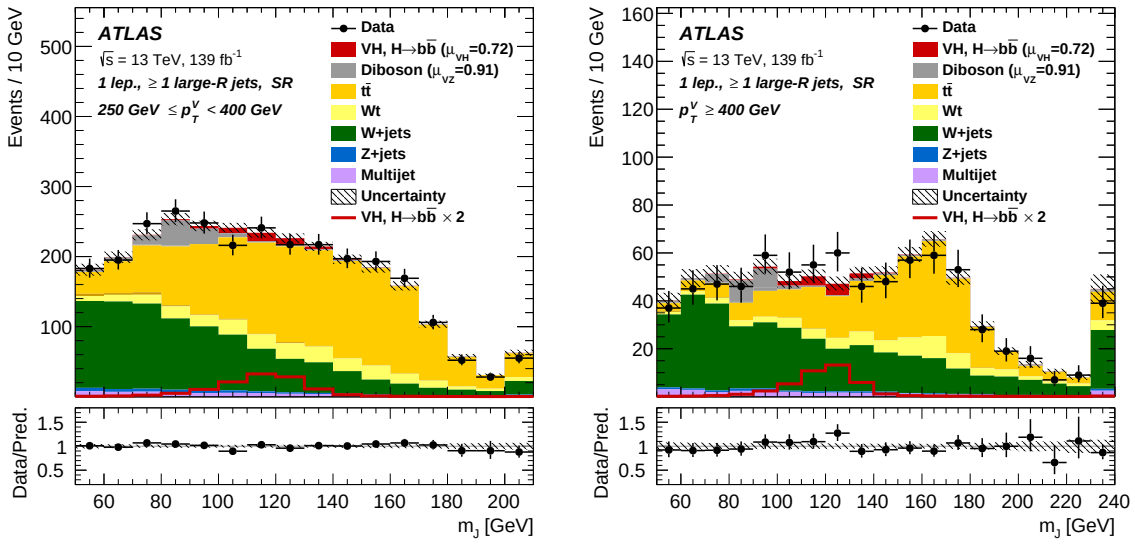
(a) (b)

Figure 5.7: Large-R jet mass distribution in the 0-lepton channel signal region, for 250 GeV $\leq p_T^V <$ 400 GeV (5.7a) and $p_T^V \geq$ 400 GeV (5.7b). The size of the combined statistical and systematic uncertainty for the fitted signal and background is indicated by the hatched band. The ratio of the data to the sum of the fitted signal and background is shown in the lower panel.

the data-taking period, to cope with increasing trigger rates at higher instantaneous luminosities. Additionally, events with *baseline* leptons are vetoed. Furthermore, since the 0-lepton channel targets $ZH \to \nu\bar{\nu}\, b\bar{b}$ events, the requirement on p_T^V translates into $E_T^{miss} \geq 250$ GeV, since E_T^{miss} is due to the $Z \to \nu\bar{\nu}$ decay.

A series of *anti-QCD cuts* is included in the selection. The multijet background in the 0-lepton channel is due to jet energy mis-measurements. As a result, a fake $\vec{E}_T^{miss}$ is expected to be aligned with the mis-measured jet, and opposite to the correctly reconstructed $\vec{E}_T^{miss}$. To reduce this background to a negligible level, three dedicated selection criteria are applied:

- the missing transverse momentum is not aligned with the Higgs-jet candidate: $\Delta\phi(\vec{E}_T^{miss}, H_{cand}) > 120°$.
- $\vec{E}_T^{miss}$ should be aligned with $\vec{E}_{T,trk}^{miss}$: $\Delta\phi(\vec{E}_T^{miss}, \vec{E}_{T,trk}^{miss}) < 90°$. Here $\vec{E}_T^{miss}$ is the missing energy transverse of the event, while $\vec{E}_{T,trk}^{miss}$ is the missing transverse



(a) (b)

Figure 5.8: Large-R jet mass distribution in the 1-lepton channel signal region, for 250 GeV $\leq p_T^V <$ 400 GeV (5.8a) and $p_T^V \geq$ 400 GeV (5.8b). The size of the combined statistical and systematic uncertainty for the fitted signal and background is indicated by the hatched band. The ratio of the data to the sum of the fitted signal and background is shown in the lower panel.

momentum calculated using only tracks reconstructed in the ID and matched to the primary vertex. Fake missing energy tends to diverge from the track-based transverse momentum.

- $\vec{E}_T^{miss}$ should be isolated from any calorimeter small-R jet (with small-R jet $p_T > 70$ GeV) and selection $\min[\Delta\phi(\vec{E}_T^{miss}, \textit{small-R jets})] > 30°$ is applied. Only small-R jets not overlapping with the Higgs-jet candidate are considered.

After these selections, the presence of multi-jet events in the 0-lepton channel is negligible.

1-lepton channel

In the 1-lepton channel (see Figure 5.8) two different trigger strategies are used for the electron- and muon- sub-channels.

For the muon sub-channel, the same trigger as in the 0-lepton channel is used,

being more efficient than single-muon triggers in the analysis phase space. This is due to the fact that muons do not enter in the online E_T^{miss} calculation and that uninstrumented regions affect the coverage of the muon spectrometer. Furthermore a *signal* muon with $p_T > 25$ GeV and no additional *baseline* leptons are required.

In case of the electron sub-channel, instead, a single-electron trigger is required. The lowest E_T threshold of these triggers varied with time from 24 GeV to 26 GeV. Furthermore, a veto on additional *baseline* leptons is imposed. The isolation requirements on the *signal* lepton remove most of the non-prompt lepton background, however an additional $E_T^{miss} > 50$ GeV requirement is applied in the electron sub-channel to reduce this background further.

In this channel, the reconstructed vector boson transverse momentum p_T^V corresponds to the vectorial sum of $\vec{E}_T^{miss}$ and of the charged lepton transverse momentum. A selection requiring $p_T^V > 250$ GeV is applied.

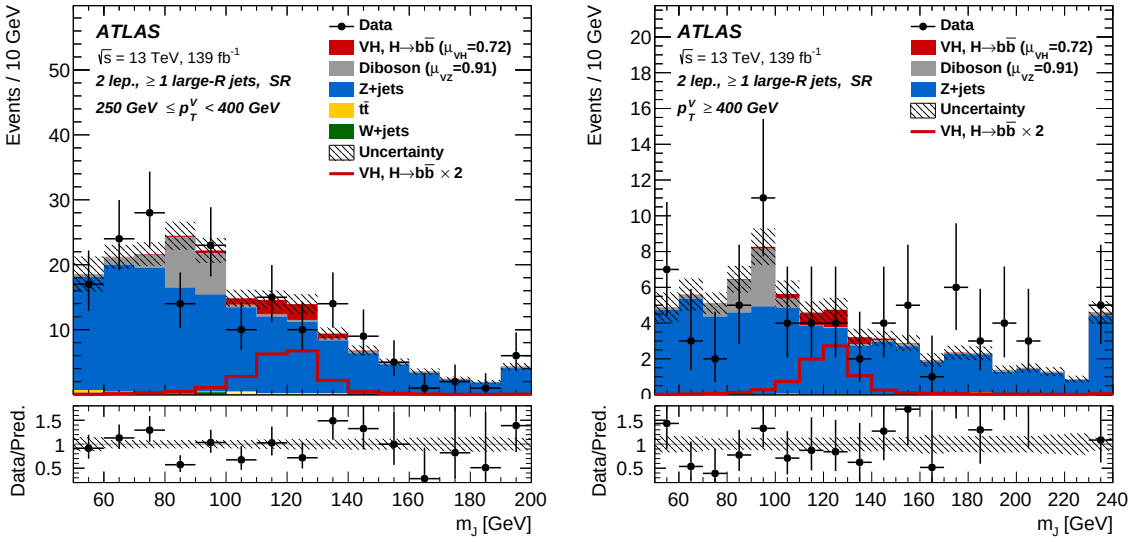
Finally, in order to reduce other backgrounds, such as top and W+*jets* productions, a further selection on the rapidity difference between the Higgs boson candidate and the vector boson (W in this case) is applied: $|\Delta y(V, H_{cand})| < 1.4$.

2-lepton channel

In the 2-lepton channel (see Figure 5.9), the same trigger strategy as in the 1-lepton channel is adopted.

The two lepton channel targets $ZH \to \ell^+\ell^- b\bar{b}$ events, therefore two same-flavour *baseline* leptons are required. Additionally, at least one *signal* lepton with $p_T > 27$ GeV is required. In the di-muon sub-channel the two muons are further required to be of opposite sign.

Similarly to the 1-lepton channel, the requirement on the rapidity difference between the Higgs boson candidate and the vector boson (Z in this case) is applied, to reduce the main Z+*jets* background process. In order to further suppress this back-



(a) (b)

Figure 5.9: Large-R jet mass distribution in the 2-lepton channel signal region, for 250 GeV $\leq p_T^V <$ 400 GeV (5.9a) and $p_T^V \geq$ 400 GeV (5.9b). The size of the combined statistical and systematic uncertainty for the fitted signal and background is indicated by the hatched band. The ratio of the data to the sum of the fitted signal and background is shown in the lower panel.

ground a requirement is applied to the lepton p_T imbalance[2], that is sensitive to the Z boson polarisation [155] and is therefore expected to differ between the ZH signal and the Z+*jets* background. Finally, the mass of the di-lepton system is required to be close to the Z boson mass ($m_Z \sim 91.2$ GeV [27]), with range values determined to optimise the analysis acceptance.

The full event selection is summarised in Table 5.4.

[2]The lepton p_T imbalance is defined as:

$$\frac{p_T^{\ell_1} - p_T^{\ell_2}}{p_T^Z}$$

Selection	0 lepton channel	1 lepton channel		2 leptons channel	
		e sub-channel	μ sub-channel	e sub-channel	μ sub-channel
Trigger	E_T^{miss}	Single electron	E_T^{miss}	Single electron	E_T^{miss}
Leptons	0 *baseline* leptons	1 *signal* lepton, $p_T > 27$ GeV	1 *signal* lepton, $p_T > 25$ GeV	2 *baseline* leptons among which ≥ 1 *signal* lepton, $p_T > 27$ GeV	
		no second *baseline* lepton		both leptons of the same flavour	
				-	opposite sign muons
E_T^{miss}	> 250 GeV	> 50 GeV	-	-	
p_T^V	$p_T^V > 250$ GeV				
large-R jets	at least one large-R jet, $p_T > 250$ GeV, $\|\eta\| < 2.0$				
Track-jets	at least two track-jets, $p_T > 10$ GeV, $\|\eta\| < 2.5$, matched to the leading large-R jet				
b-tagged jets	leading two track-jets matched to the leading large-R must be b-tagged (MV2c10, 70%)				
m_J	> 50 GeV				
$\min[\Delta\phi(\vec{E}_T^{miss}, \textit{small-R jets})]$	$> 30°$	-			
$\Delta\phi(\vec{E}_T^{miss}, H_{\text{cand}})$	$> 120°$	-			
$\Delta\phi(\vec{E}_T^{miss}, \vec{E}_{T,trk}^{miss})$	$< 90°$	-			
$\Delta y(V, H_{\text{cand}})$	-	$\|\Delta y(V, H_{\text{cand}})\| < 1.4$			
$m_{\ell\ell}$	-			66 GeV $< m_{\ell\ell} <$ 116 GeV	
Lepton p_T imbalance	-			$(p_T^{\ell_1} - p_T^{\ell_2})/p_T^Z < 0.8$	

Table 5.4: Event selection requirements for the boosted $VH(H \to b\bar{b})$ analysis channels and sub-channels.

5.3.3 Event categorisation

Events passing the selection criteria are classified in fourteen categories, defining fourteen different analysis regions. The analysis categorisation is useful to enhance the analysis sensitivity, by enhancing the signal-to-background (S/B) ratio, or by targeting specific background processes, otherwise difficult to estimate and control.

p_T^V splitting

The first categorisation is made on the p_T of the vector boson candidate (see Figure 5.10), since the signal-to-background ratio increases for large Higgs boson transverse momenta [156, 157]. Furthermore, BSM effects may be more pronounced in the high-p_T region and the splitting is therefore tailored towards a differential cross section measurement (see Section 5.8).

Based on their p_T^V, events fall into a medium-p_T^V region if 250 GeV $\leq p_T^V <$ 400 GeV, or a high-p_T^V region for $p_T^V \geq 400$ GeV.

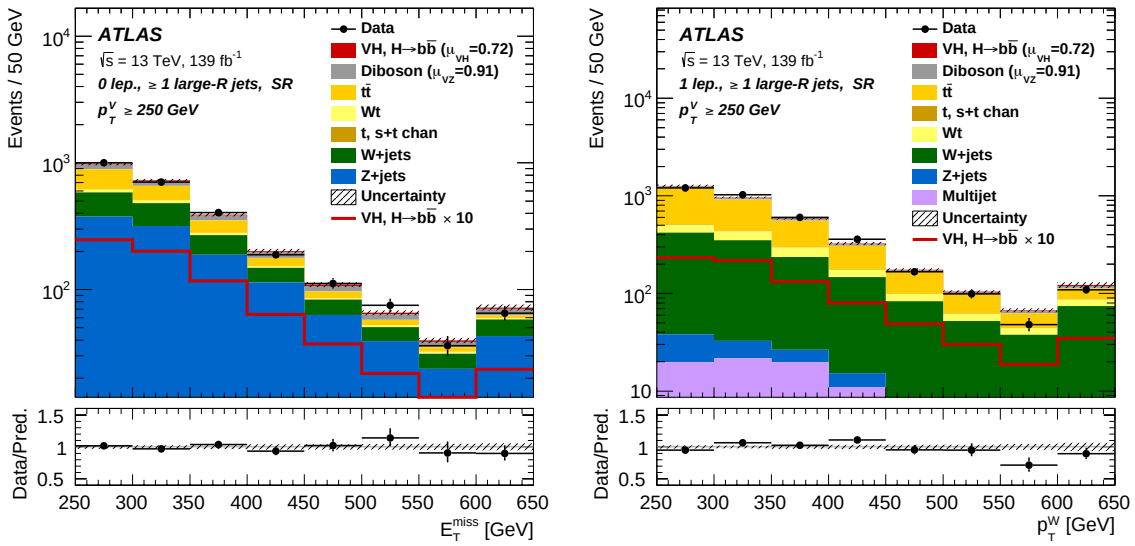
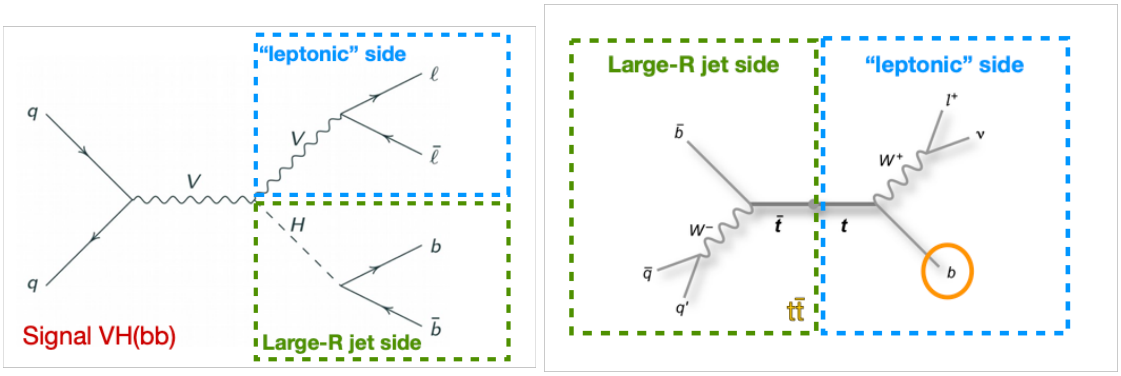


Figure 5.10: p_T^V distribution in the 0- (5.10a), 1- (5.10b) and 2-lepton (5.10c) channel signal regions. The size of the combined statistical and systematic uncertainty for the fitted signal and background is indicated by the hatched band. The ratio of the data to the sum of the fitted signal and background is shown in the lower panel.



(a) (b)

Figure 5.11: Graphical representation of the signal $VH(H \to b\bar{b})$ (5.11a) and $t\bar{t}$ background (5.11b) reconstruction in the $VH(H \to b\bar{b})$ boosted analysis. On the diagram of the $t\bar{t}$ process the b-quark that originates the b-tagged track jet not matched to the Higgs-candidate jet is highlighted in orange.

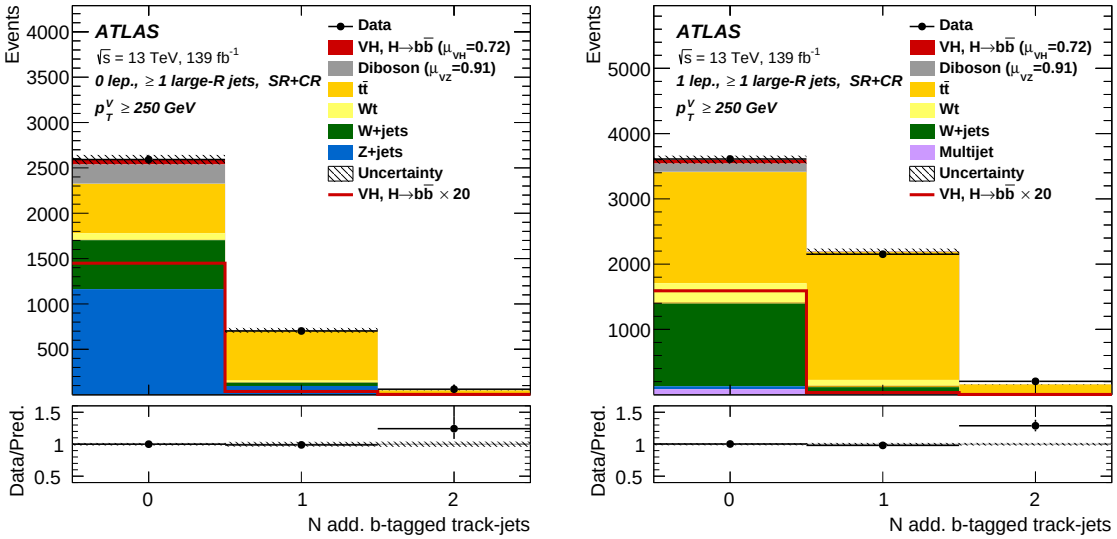
Splitting into Signal Region and Control Region

The $t\bar{t}$ process is one of the main backgrounds in the 0- and 1-lepton channels. $t\bar{t}$ events reconstructed in the analysis phase space have one top-quark decay reconstructed in the large-R jet, while the other is reconstructed on the "leptonic" side, giving an additional b-tagged jet (outside the large-R jet). In this case, the W boson decay, from the top quark decay chain, mimics the V boson decay, which also occurs in $VH(H \to b\bar{b})$. This is graphically explained in Figure 5.11.

Taking this into account, two types of regions are defined in the 0- and 1-lepton channels, based on the number of "additional" b-tagged VR track jets, not matched to the large-R (Higgs-candidate) jet (see Figure 5.12). More specifically, *signal regions* (SRs) are defined by vetoing b-tagged VR track-jets outside the large-R jet and top-quark-enriched *control regions* (CRs) are built from events failing this veto. As shown in Figure 5.13, the CRs are very pure in $t\bar{t}$ events, making them useful in constraining the $t\bar{t}$ background.

Splitting of the Signal Regions

Events in the 0- and 1-lepton signal region are further split according to the expected



(a) (b)

Figure 5.12: Number of additional b-tagged VR track-jets non-matched to the Higgs candidate large-R jet in the 0-lepton (5.12a) and the 1-lepton (5.12b) for $p_T^V > 250$ GeV. The size of the combined statistical and systematic uncertainty for the fitted signal and background is indicated by the hatched band. The ratio of the data to the sum of the fitted signal and background is shown in the lower panel.

topology of $VH(H \to b\bar{b})$ events at leading order. This categorisation is based on the absence or presence of additional small-R calorimeter jets, not matched to the large-R jet (see Figure 5.14). A *high purity* signal region (HP SR) is obtained by vetoing additional small-R calorimeter jets, otherwise events are categorised into a *low purity* signal region (LP SR). The large-R jet mass distributions for the HP SR and the LP SR are shown in Figures 5.15 and 5.16.

A summary of the ten signal regions and four control regions is given in Table 5.5.

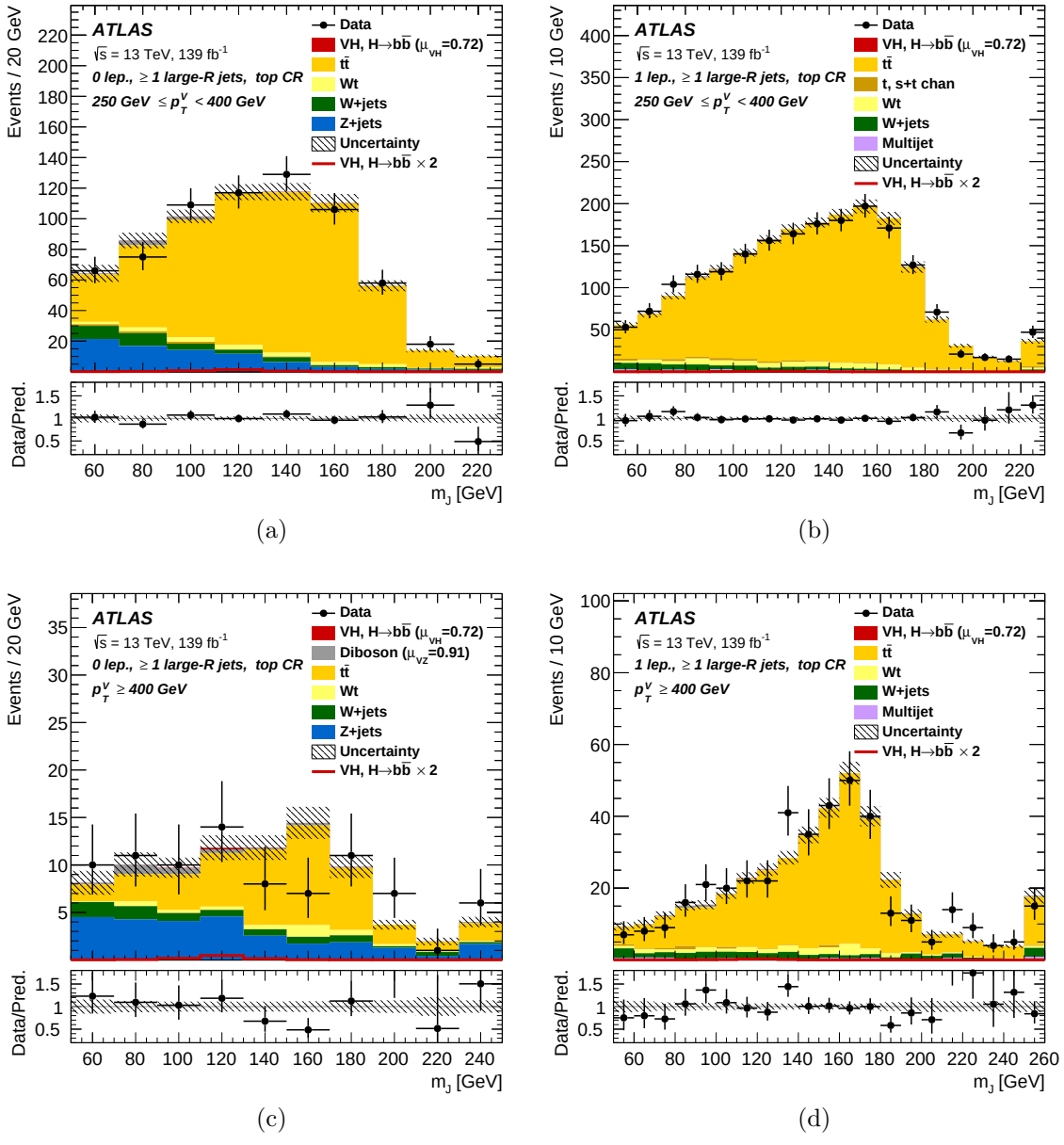
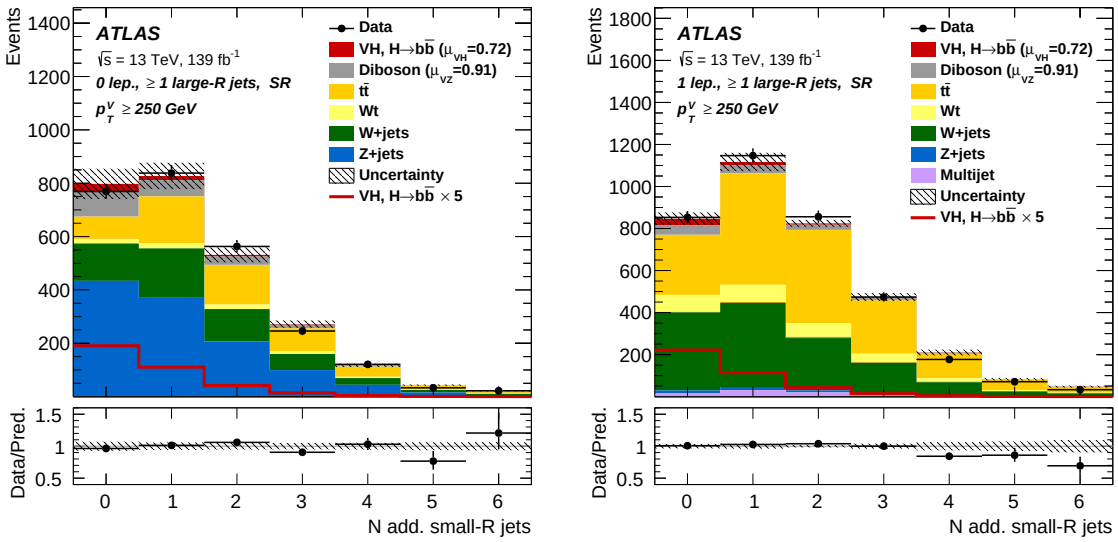


Figure 5.13: Large-R jet mass distribution in the $t\bar{t}$ control regions, in the 0-lepton (5.13a, 5.13c) and 1-lepton (5.13b, 5.13d) channel, for 250 GeV $\leq p_T^V <$ 400 GeV (5.13a, 5.13b) and $p_T^V \geq$ 400 GeV (5.13c, 5.13d). The size of the combined statistical and systematic uncertainty for the fitted signal and background is indicated by the hatched band. The ratio of the data to the sum of the fitted signal and background is shown in the lower panel.



(a) (b)

Figure 5.14: Number of additional small-R calorimeter jets non-matched to the Higgs candidate large-R jet in the 0-lepton (5.14a) and the 1-lepton (5.14b) for $p_T^V > 250$ GeV. The size of the combined statistical and systematic uncertainty for the fitted signal and background is indicated by the hatched band. The ratio of the data to the sum of the fitted signal and background is shown in the lower panel.

Channel	Categories					
	$250 < p_T^V < 400$ GeV			$p_T^V > 400$ GeV		
	0 add. b-track-jets		≥ 1 add. $b-$track-jets	0 add. b-track-jets		≥ 1 add. b-track-jets
	0 add. small-R jets	$\geq$ 1add. small-R jets		0 add. small-R jets	$\geq$ 1add. small-R jets	
0-lepton	SR	SR	CR	SR	SR	CR
1-lepton	SR	SR	CR	SR	SR	CR
2-lepton	SR			SR		

Table 5.5: Summary of the definition of the analysis regions for the $VH(H \to b\bar{b})$ boosted analysis. Regions with relatively large signal purity are marked with the label SR. The $t\bar{t}$ background-enriched regions are marked with the label CR.

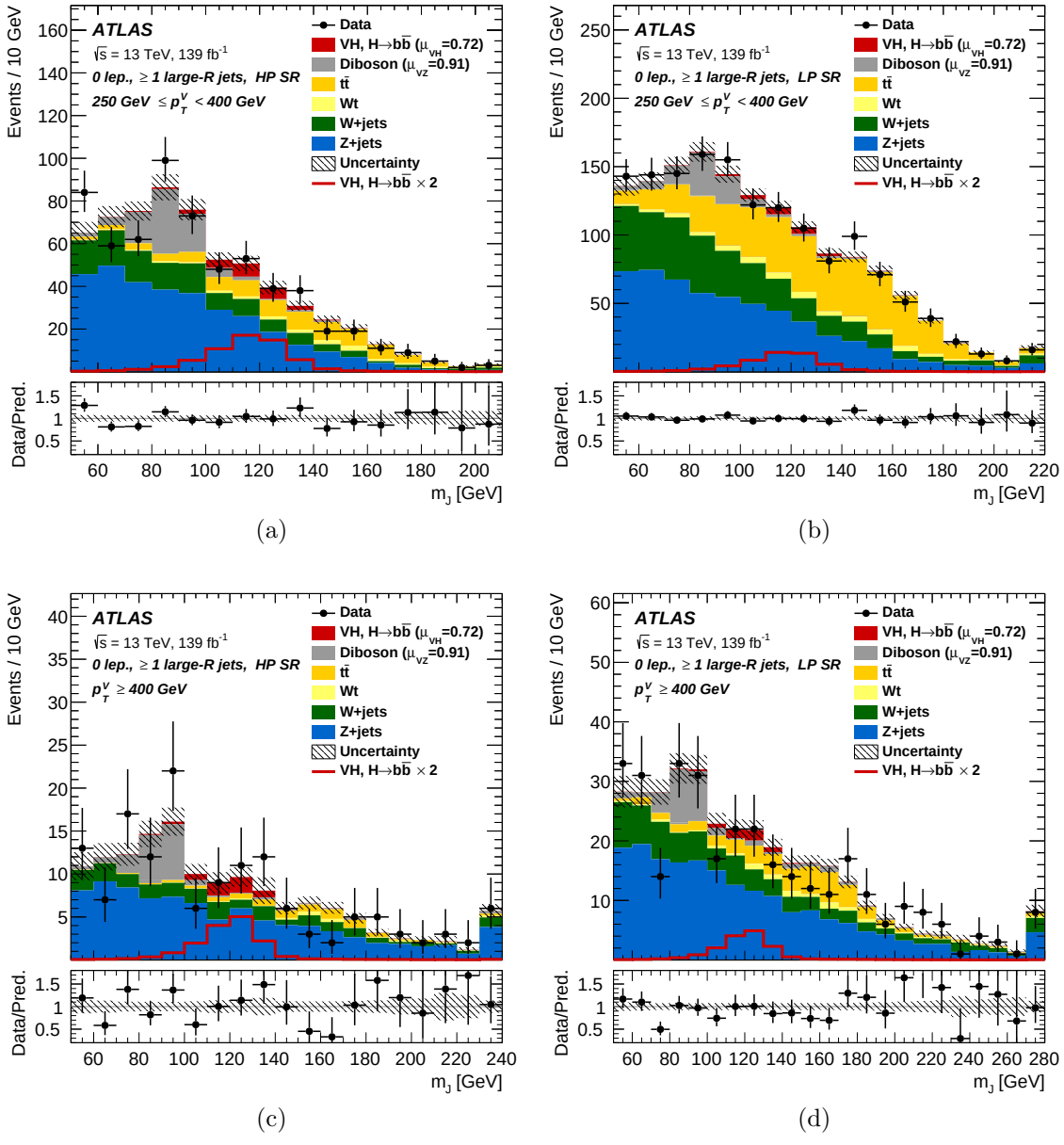


Figure 5.15: Large-R jet mass distribution in the high purity (5.15a, 5.15c) and low purity (5.15b, 5.15d) signal regions in the 0-lepton channel, for 250 GeV $\leq p_T^V <$ 400 GeV (5.15a, 5.15b) and $p_T^V \geq$ 400 GeV (5.15c, 5.15d). The size of the combined statistical and systematic uncertainty for the fitted signal and background is indicated by the hatched band. The ratio of the data to the sum of the fitted signal and background is shown in the lower panel.

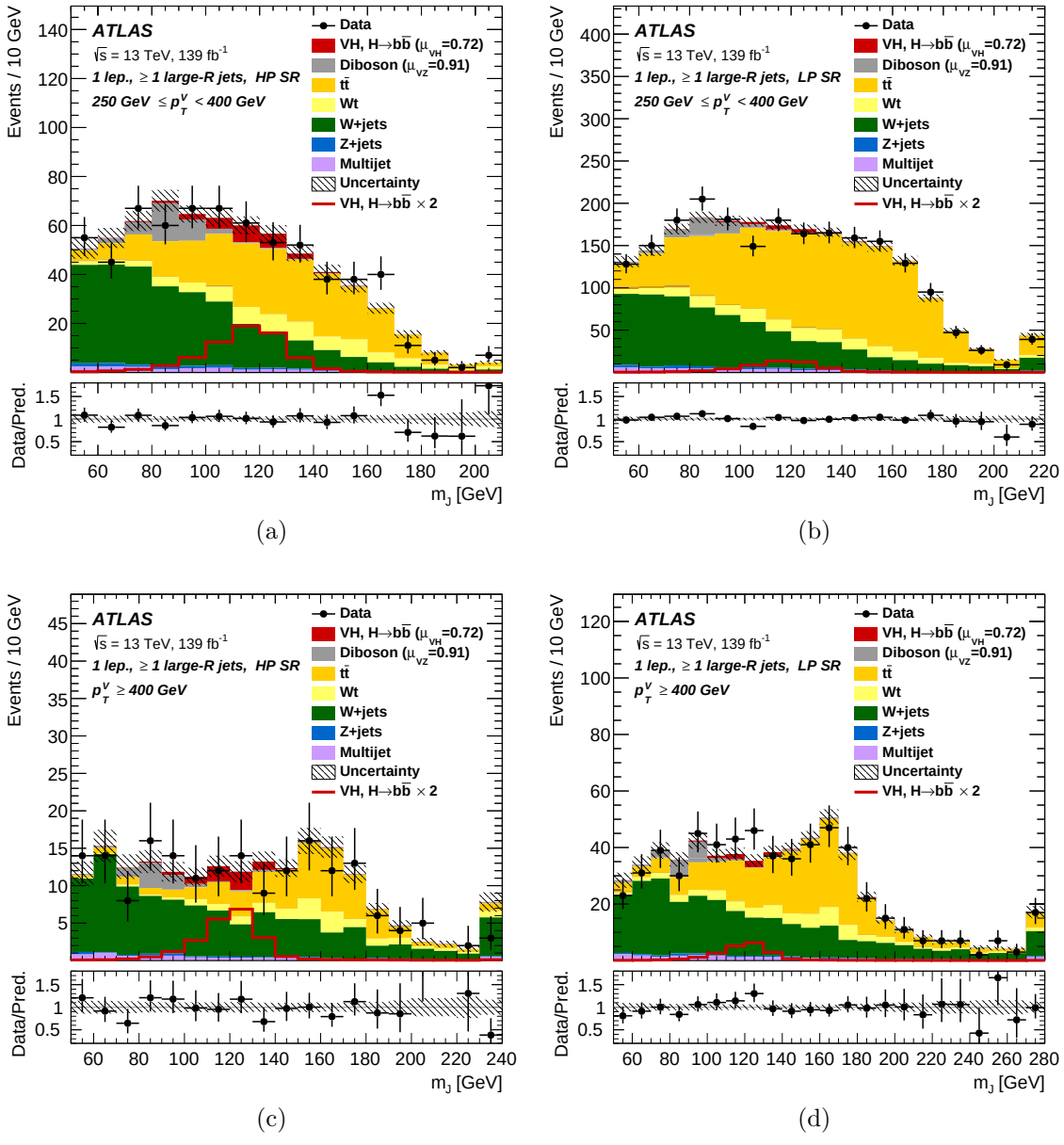


Figure 5.16: Large-R jet mass distribution in the high purity (5.16a, 5.16c) and low purity (5.16b, 5.16d) signal regions in the 1-lepton channel, for 250 GeV $\leq p_T^V <$ 400 GeV (5.16a, 5.16b) and $p_T^V \geq$ 400 GeV (5.16c, 5.16d). The size of the combined statistical and systematic uncertainty for the fitted signal and background is indicated by the hatched band. The ratio of the data to the sum of the fitted signal and background is shown in the lower panel.

5.4 Optimisation of the analysis selection criteria: studies on the $\Delta R(b_1, b_2)$ in the 0-lepton channel

At leading order, $VH(H \to b\bar{b})$ events in the boosted regime present a large-R jet with a two-object substructure, with no (or very little) p_T imbalance between the two sub-jets. Conversely, for background events more sub-jets and/or p_T imbalance and thus different angular separation between the two b-tagged VR track jets are expected.

For this reason, it is interesting to perform a "jet-substructure" study, analysing the angular separation $\Delta R(b_1, b_2)$ between the two b-tagged VR track jets (b_1 and b_2) associated to the large-R jet[3].

The study presented below is performed in the 0-lepton channel before splitting into the high purity and the low purity signal regions. All selection criteria for the reconstruction of the Higgs boson and V boson candidate (introduced in Section 5.3.2) are applied in this study.

The distributions of $\Delta R\,(b_1, b_2)$ for signal events and the most crucial background processes are shown in Figure 5.17, where differences among the different processes can be appreciated. In V+*jets* events, the two b-tagged VR track jets tend to be very close and the distributions peak at low $\Delta R\,(b_1, b_2)$ values. This is a consequence of the fact that the two b-jets originate from gluon splitting. On the other hand, signal $VH(H \to b\bar{b})$ and $t\bar{t}$ events show a very similar behaviour, with the majority of events concentrating at higher values of $\Delta R\,(b_1, b_2)$, as a consequence of the higher invariant mass of the two b-tagged track jets captured within the large-R jet. For diboson events $\Delta R\,(b_1, b_2)$ peaks around 0.6, at "intermediate" values.

[3]A reminder that:

$$\Delta R\,(b_1, b_2) = \sqrt{(\Delta\eta_{1,2})^2 + (\Delta\phi_{1,2})^2} = \sqrt{(\eta_{b_1} - \eta_{b_2})^2 + (\phi_{b_1} - \phi_{b_2})^2}$$

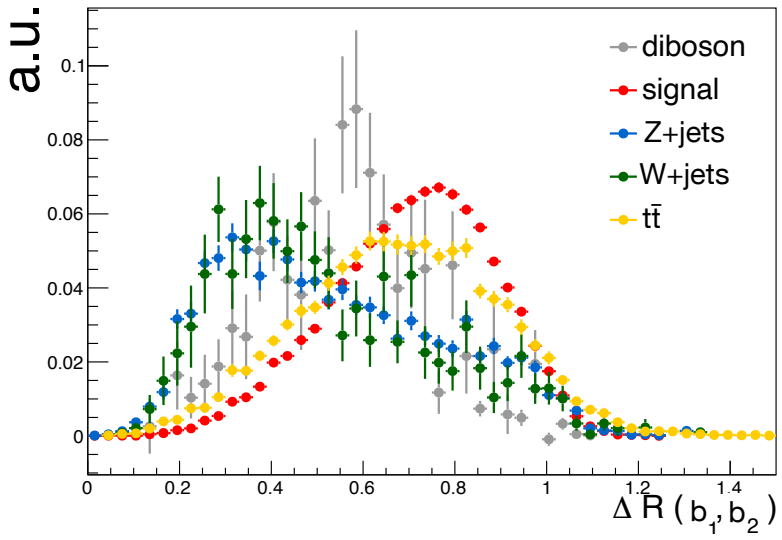


Figure 5.17: 0-lepton $\Delta R\,(b_1, b_2)$ distributions for signal $VH(H \to b\bar{b})$, V+*jets*, $t\bar{t}$ and diboson events in arbitrary units. All the curves are normalised to the same area.

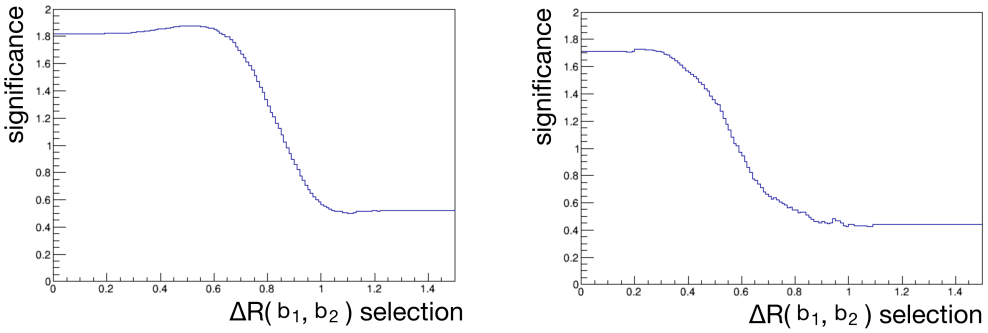
The goal of the study is to understand whether a selection on $\Delta R\,(b_1, b_2)$ improves the sensitivity of the analysis, by increasing the signal-to-background ratio. In order to do so, the significance of the analysis in the 0-lepton channel is calculated as a function of the selection applied on $\Delta R\,(b_1, b_2)$, specifically removing the events for which $\Delta R\,(b_1, b_2)$ is lower than a fixed threshold.

At a first stage, the 0-lepton channel analysis significance is calculated using the formula for a binned statistical significance:

$$S_{bin_i} = \sqrt{\sum_{i=0}^{bin_i} 2 \cdot \left((n_s^i + n_b^i) \cdot \ln\left(1 + \frac{n_s^i}{n_b^i}\right) - n_s^i \right)} \tag{5.2}$$

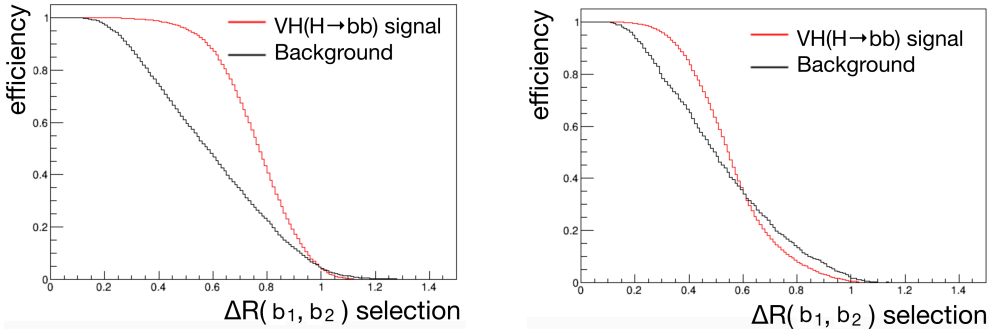
where n_s^i (n_b^i) is the number of signal (background) events in each bin i of the large-R jet mass distribution (m_J). The significance S_{bin_i} is evaluated as a function of the proposed $\Delta R\,(b_1, b_2)$ selection for both the medium- and high-p_T^V regions, as shown by Figure 5.18. Furthermore, the signal and background efficiencies are computed as a function of the $\Delta R\,(b_1, b_2)$ selection (Figure 5.19).

The two values for the $\Delta R\,(b_1, b_2) \geq \alpha$ selection that maximise the increase in sig-



(a) (b)

Figure 5.18: Binned statistical significance as a function of $\Delta R\,(b_1, b_2)$ selection in the medium- (5.18a) and high- (5.18b) p_T^V regions of the analysis. The significance in the 0-lepton channel is evaluated for $\Delta R\,(b_1, b_2) \geq \alpha$.



(a) (b)

Figure 5.19: Signal (red) and background (black) selection efficiency as a function of $\Delta R\,(j_1, j_2)$ selection in the medium- (5.19a) and high- (5.19b) p_T^V regions of the analysis 0-lepton channel. The selection efficiencies are evaluated for $\Delta R\,(b_1, b_2) \geq \alpha$.

	250 GeV $\leq p_T^V <$ 400 GeV	$p_T^V \geq$ 400 GeV
Selection	$\Delta R\,(b_1, b_2) > 0.5$	$\Delta R\,(b_1, b_2) > 0.2$
Signal Efficiency	96%	100%
Background Rejection	40%	5%
Significance gain	+3%	+2%

Table 5.6: Summary of the impact of the $\Delta R\,(b_1, b_2)$ selection on the binned statistical significance of the $VH(H \to b\bar{b})$ boosted analysis in the 0-lepton channel.

nificance are $\alpha_1 = 0.5$ and $\alpha_2 = 0.2$ in the medium- and high-p_T^V regions respectively, leading to an overall improvement of the analysis significance of $\sim 3\%$ in the 0-lepton channel (see Table 5.6).

The statistical binned significance formula has been used since it provides an immediate estimate of the statistical significance and it is therefore very useful for preliminary studies, when potential improvements need to be tested.

However, the binned statistical significance does not take into account uncertainties, neither of statistical nor of systematic nature. Therefore, in order to have a more realistic picture, a fit (see Section 5.5 for a more detailed description of the fit architecture) in the 0-lepton channel only and with a minimal set of uncertainties is performed. The fit takes into account statistical uncertainties and enables the normalisation of $Z+h.f.$[4] and $t\bar{t}$ background processes to be free-floating parameters. Figure 5.20 shows results of the fitting procedure on the large-R jet mass distributions in the 0-lepton channel, for the $\Delta R\,(b_1, b_2)$ selection. The median significances extracted by the conditional fit to data[5] is 1.77 pre-ΔR-selection and 1.66 post-ΔR-selection, with a decrease of 6%.

This decrease in significance is explained by the reduction of events at low values of m_J (as seen in Figure 5.20) after the selection on $\Delta R\,(b_1, b_2) > \alpha$. The low-mass

[4] $h.f.$ stands for "heavy-flavour" jets, originating by heavy flavoured quarks. In the $VH(H \to b\bar{b})$ boosted analysis bb, bc, bl and cc are considered "heavy-flavour" pairs.

[5] The fit is performed fixing the $VH(H \to b\bar{b})$ signal to the SM prediction.

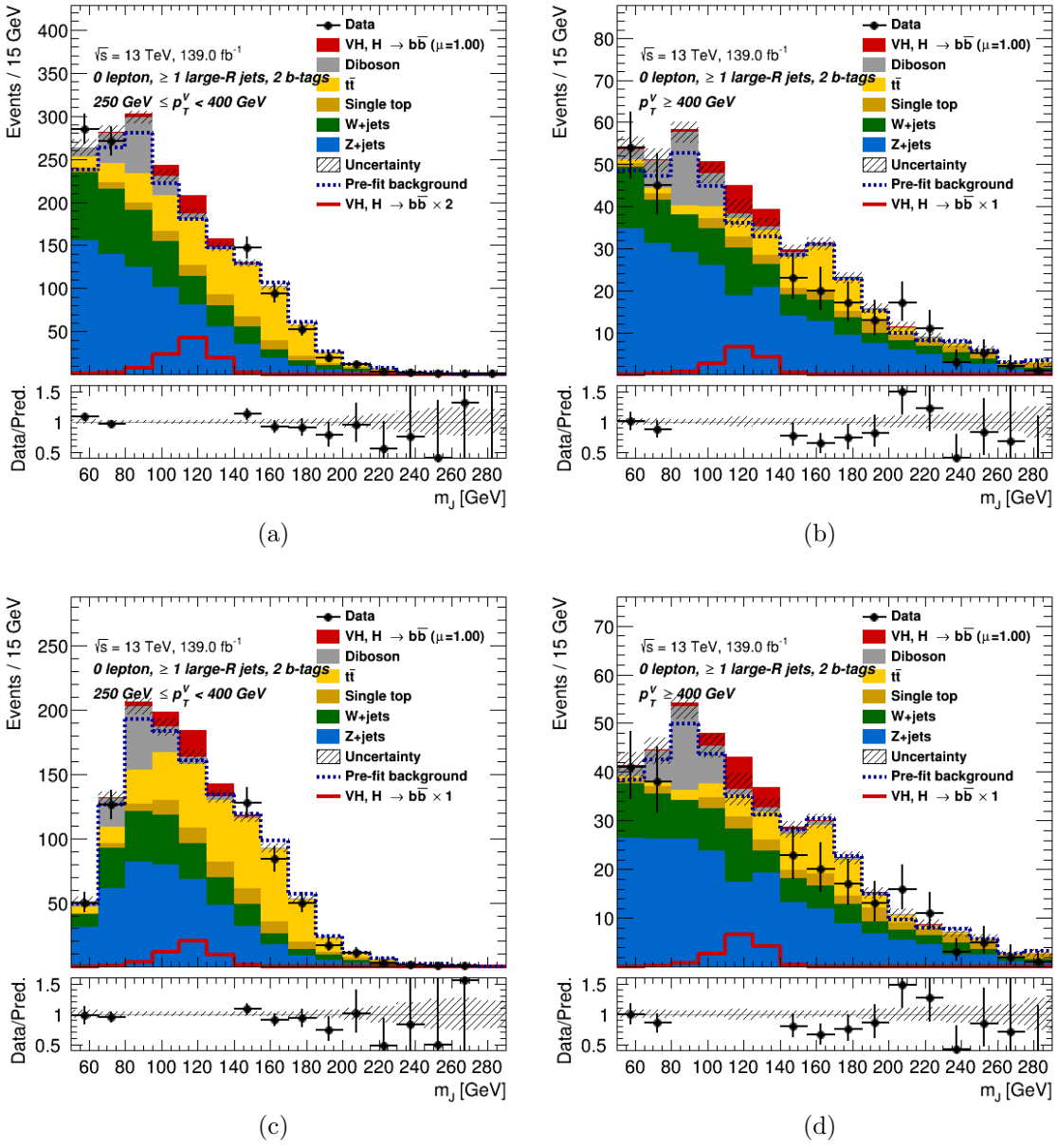


Figure 5.20: Large-R jet mass distribution in the 0-lepton channel signal region, for 250 GeV $\leq p_T^V <$ 400 GeV (5.20a, 5.20c) and $p_T^V \geq$ 400 GeV (5.20b, 5.20d), for no $\Delta R\,(b_1, b_2)$ selection (5.20a, 5.20b), and when the $\Delta R\,(b_1, b_2)$ selection is implemented (5.20c, 5.20d). The size of the combined statistical and systematic uncertainty for the fitted signal and background is indicated by the hatched band. The ratio of the data to the sum of the fitted signal and background is shown in the lower panel. Data is not displayed in the region around the Higgs boson mass, since this studies were conducted at an initial stage of the analysis, when the analysis was still "blinded". The fit is performed fixing the $VH(H \to b\bar{b})$ signal to the SM prediction.

region is used by the fit to constrain the values for the floating normalisation of the $Z + h.f.$ background, therefore the significant reduction of events in this critical region leads to an increase in the uncertainty on its floating normalisation, affecting the overall analysis significance, that decreases.

To conclude, even though $\Delta R\,(b_1, b_2)$ is a powerful signal-to-background discriminator, it has been demonstrated that performing a selection on it is not affordable with the current data sample size. However, it might be possible to use it when more statistics will be available in the future. In such conditions there will be enough events for the fit to fix the normalisation parameters even after the selection is applied and the gain on the signal-to-background ratio would be transferred to the significance of the analysis.

Finally, a viable alternative to exploit $\Delta R\,(b_1, b_2)$ could be using a likelihood-based approach, instead of performing a selection. Such approach could be effective, while not penalising the data sample size. Similar studies go beyond the scope of the $VH(H \to b\bar{b})$ boosted analysis, but can be explored for the future.

5.5 Statistical analysis

After the event selection and categorisation are applied, a statistical fitting procedure is employed, to extract the strength of the Higgs boson signal from data. The signal strength μ is defined as the measured product of the production cross section (σ) and branching ratio (BR) of the process of interest, in units of the corresponding SM values:

$$\mu = \frac{\sigma \times \mathrm{BR}}{\sigma_{SM} \times \mathrm{BR}_{SM}} \tag{5.3}$$

The statistical procedure that extracts the signal strength is based on a binned profile likelihood fit performed on the invariant mass of the Higgs candidate (large-R)

jet m_J. The fit is performed simultaneously in all analysis regions and all three lepton channels.

A novel feature of the $VH(H \to b\bar{b})$ boosted analysis, is that the fit extracts simultaneously the $VH(H \to b\bar{b})$ and $VZ(Z \to b\bar{b})$ signal strengths, called $\mu_{VH}^{b\bar{b}}$ and $\mu_{VZ}^{b\bar{b}}$ respectively. $\mu_{VH}^{b\bar{b}}$ and $\mu_{VZ}^{b\bar{b}}$ are therefore the fit *parameters of interest* (POIs). An insight on the two-POIs fit is provided in Section 5.5.1.

The likelihood function used in the $VH(H \to b\bar{b})$ boosted analysis can be schematically written as:

$$\mathcal{L}(\vec{\mu}, \vec{\theta}, \vec{\gamma}) = \prod_{i \in \text{bins}} \text{Poiss}(N_i | \vec{\mu} \cdot \vec{s_i}(\vec{\theta}) + \gamma_i b_i(\vec{\theta})) \times \prod_{\theta \in \vec{\theta}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(\theta-\theta_0)^2}{2\sigma^2}} \times \prod_{i \in \text{bins}} \text{Gauss}(\beta_i | \gamma_i \beta_i, \sqrt{\gamma_i \beta_i}) \tag{5.4}$$

The first factor describes a counting experiment in each bin using a set of Poissonian probability terms where N_i is the number of measured data events in each bin i and s_i and b_i are the expected signal and background counts from Monte Carlo simulation:

$$\prod_{i \in \text{bins}} \text{Poiss}(N_i | \vec{\mu} \cdot \vec{s_i}(\vec{\theta}) + \gamma_i b_i(\vec{\theta})) = \prod_{i \in \text{bins}} \frac{\left(\vec{\mu} \cdot \vec{s_i}(\vec{\theta}) + \gamma_i b_i(\vec{\theta})\right)^{N_i}}{N_i!} e^{-[\vec{\mu} \cdot \vec{s_i}(\vec{\theta}) + \gamma_i b_i(\vec{\theta})]} \tag{5.5}$$

The signal strength $\vec{\mu}$ is indicated as a vector, since there is no conceptual limit as to how many POIs are parametrised (two in the present case: μ_{VH} and μ_{VZ}). The POIs are common scale factors to the signal yields in all bins.

The second factor in Equation 5.4 describes the parametrisation of systematic uncertainties, as a set of Gaussian-distributed nuisance parameters (NPs, $\vec{\theta}$), where θ is the NP to be fitted, while θ_0 and σ are its central value and uncertainty assigned to the prior probability distribution, respectively. Some NPs, such as normalisation

factors, have no corresponding Gaussian constraint but are floating and their value is entirely determined in the fit. The inclusion of the NPs in the likelihood adds flexibility to the model but reduces the analysis sensitivity.

Finally, the third term in Equation 5.4 addresses the problem of limited simulation statistics for the background processes by introducing γ-parameters (γ_i) as multiplicative factors to Gaussian distributed background yields[6]. In Equation 5.4, $\beta_i = \frac{1}{\sigma_{rel}^2}$, with σ_{rel} being the relative statistical uncertainty on the expected total background yield in each bin. The signal VH is not affected by these γ NPs, since the statistical uncertainties on the predicted signal simulation are small with respect to the background processes [159].

In order to test a certain hypothesis for the measured value of $\vec{\mu}$, a profile likelihood ratio is considered as test statistic [160]:

$$\lambda(\vec{\mu}) = \frac{\mathcal{L}(\vec{\mu}, \hat{\hat{\vec{\theta}}}, \hat{\hat{\vec{\gamma}}})}{\mathcal{L}(\hat{\vec{\mu}}, \hat{\vec{\theta}}, \hat{\vec{\gamma}})} \tag{5.6}$$

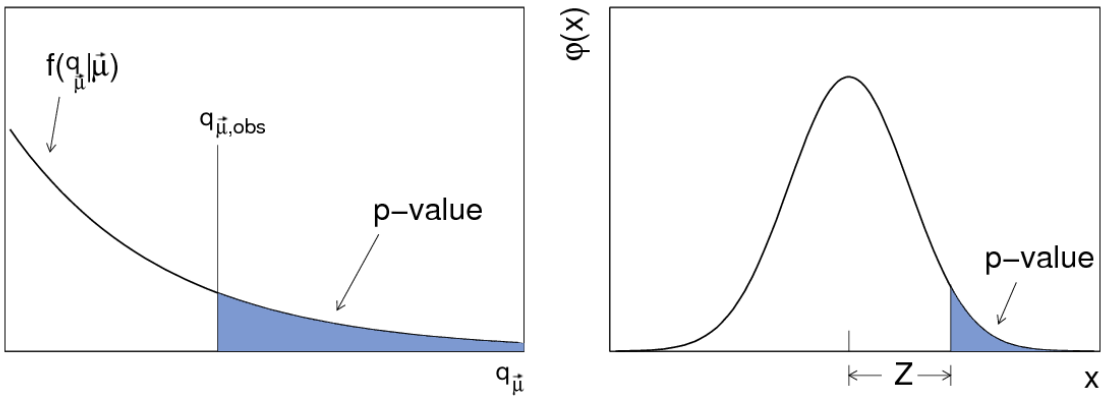
where $\{\hat{\vec{\mu}}, \hat{\vec{\theta}}, \hat{\vec{\gamma}}\}$ are the values that maximise the likelihood and $\{\hat{\hat{\vec{\theta}}}, \hat{\hat{\vec{\gamma}}}\}$ the ones that maximise $\mathcal{L}$ for a fixed value of $\vec{\mu}$.

The likelihood ratio is defined in the range $[0, 1]$, and describes a good agreement between the given hypothesis for $\vec{\mu}$ and data when $\lambda(\vec{\mu}) \sim 1$ and disagreement as $\lambda(\vec{\mu})$ moves to 0. It is convenient to construct the test statistic $q_{\vec{\mu}}$ [161]:

$$q_{\vec{\mu}} = -2 \ln \lambda(\vec{\mu}) \tag{5.7}$$

so that as the incompatibility between the data and the hypothesis $\vec{\mu}$ increases, the value of $q_{\vec{\mu}}$ gets larger. The degree of the agreement (or disagreement) can be quantified using *p-values*, computed from the test statistic as:

[6]The procedure is explained in detail in Section 2.2.1 of Reference [158].



(a) (b)

Figure 5.21: Illustration of the relation between the p-value obtained from an observed value of the test statistic $q_{\vec{\mu}}$ (5.21a). Gaussian distribution showing the relation between the significance Z and the p-value (5.21b). Figure adapted from [160].

$$p_{\vec{\mu}} = \int_{q_{\vec{\mu},obs}}^{\infty} f\left(q_{\vec{\mu}}|\vec{\mu}\right) dq_{\vec{\mu}} \tag{5.8}$$

where $q_{\vec{\mu},obs}$ is the observed value of $q_{\vec{\mu}}$ in data and $f\left(q_{\vec{\mu}}|\vec{\mu}\right)$ describes the probability density functions of $q_{\vec{\mu}}$ under the assumption $\vec{\mu}$. In this context, the p-value $p_{\vec{\mu}}$ can be interpreted as the probability, under the hypothesis $\vec{\mu}$, of finding data that are incompatible with the hypothesis (a graphical representation can be found in Figure 5.21a). In other terms, as the incompatibility between data and the hypothesis increases, $q_{\vec{\mu}}$ increases, and the p-value decreases.

p-values are often expressed in terms of the *significance* Z:

$$Z = \Phi^{-1}(1 - p_{\vec{\mu}}) \tag{5.9}$$

where Φ^{-1} is the inverse Gaussian cumulative distribution function. Z indicates the number of Gaussian standard deviations that quantify the deviation of the observed data from the hypothesis $\vec{\mu}$ (see Figure 5.21b). Here, smaller values of p-values correspond to higher values of Z.

Discovering a signal ($\vec{\mu} \geq 0$) can be thought of as rejecting the *background-only*

hypothesis ($\vec{\mu} = 0$). With this purpose, the test statistic $q_{\vec{\mu}}$ can be re-defined as:

$$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{\vec{\mu}} \geq 0 \\ 0 & \hat{\vec{\mu}} < 0 \end{cases} \tag{5.10}$$

Therefore, measuring an increased yield with respect to the expected background ($\hat{\vec{\mu}} > 0$) corresponds to finding a larger q_0. As q_0 gets larger, the incompatibility of data with the $\vec{\mu} = 0$ hypothesis increases, resulting in a small p-value and large values of the significance Z.

In recent times and within the accelerator-based particle physics community, fixed values of Z are used to define different types of scientific claims: $Z = 3$ is considered the threshold for the *evidence* of a new phenomenon, $Z = 4$ is regarded as *strong evidence*, while $Z = 5$ is required to claim an *observation*, or *discovery*. A significance of $Z = 5$ corresponds to a p-value of $\sim 3 \times 10^{-7}$.

The uncertainty on $\hat{\vec{\mu}}$ is obtained with a profile likelihood method by varying $q_0 = -2\ln\lambda(\hat{\vec{\mu}})$ of one unit. This corresponds to the $\pm 1\hat{\vec{\sigma}}$ variation on $\hat{\vec{\mu}}$ [7].

The technical implementation of the statistical analysis is done within `HistFactory` [158] together with `RooFit` [162] and `RooStats` [163].

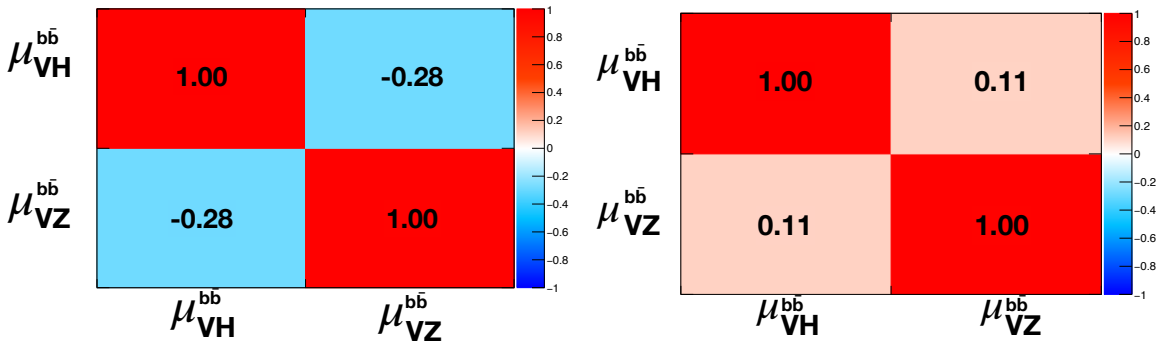
5.5.1 The VH-VZ fit

A novel feature of the $VH(H \to b\bar{b})$ boosted analysis is the simultaneous measurement of the $VH(H \to b\bar{b})$ and $VZ(Z \to b\bar{b})$ signal strengths: $\mu^{b\bar{b}}_{VH}$ and $\mu^{b\bar{b}}_{VZ}$.

In the history of the $VH(H \to b\bar{b})$ analysis [151, 164, 3], the measurement of the $VZ(Z \to b\bar{b})$ signal strength has always been performed as a cross-check for the $VH(H \to b\bar{b})$ result, given the similarities between the two processes, in terms of

[7] This is given by Taylor-expanding the log-likelihood around its maximum:

$$\ln\left(\mathcal{L}(\hat{\vec{\mu}} \pm \hat{\vec{\sigma}})\right) \sim \ln\mathcal{L}(\hat{\vec{\mu}}) - \frac{1}{2} \tag{5.11}$$



(a) (b)

Figure 5.22: Correlations between the two parameters of interest $\mu^{b\bar{b}}_{VH}$ and $\mu^{b\bar{b}}_{VZ}$ in the two-POI fit without (5.22a) and with (5.22b) systematic uncertainties.

reconstruction, as well as their proximity in the invariant mass spectrum.

The simultaneous $VH - VZ$ fit introduced in the $VH(H \to b\bar{b})$ boosted analysis is more interesting than conducting two separate one-POI fits, since it enables a simultaneous measurement of both the interesting signal strengths, without the need of imposing additional constraints to the non-fitted POI that would otherwise need to be specified to the best knowledge.

Figure 5.22a shows the correlation matrix between the two POIs $\mu^{b\bar{b}}_{VH}$ and $\mu^{b\bar{b}}_{VZ}$ in the two-POI statistical-only fit, where all the systematic NPs are dropped. This example is particularly instructive, since it shows that in a single simultaneous fit to both POIs, the two signal strengths are anti-correlated: an overestimation of the one of the two would lead to an underestimation of the other one. This is due to the nature of the underlying physics, with the two mass distributions of the Z boson the Higgs boson ($m_Z = 91$ GeV, $m_H = 125$ GeV) overlapping in the upper and lower tail respectively, and the VZ process acting effectively as a background to the signal VH and mimicking its features.

Once all the systematic uncertainties are inserted in the fit, a positive correlation between the two POIs is found (Figure 5.22b), due to the common large backgrounds.

5.6 Systematic Uncertainties

The $VH(H \to b\bar{b})$ boosted analysis is affected by several systematic uncertainties, that can be grouped into two main categories, according to their origin: *experimental uncertainties* and the uncertainties on the signal and background processes: *modelling uncertainties*.

5.6.1 Experimental Uncertainties

Experimental systematic uncertainties are associated with the physics objects used in the analysis and beam conditions measurements. In the ATLAS Collaboration, dedicated working groups provide standardised prescriptions, based on independent measurements, to the physics analyses groups.

The uncertainties affecting the physics objects have been discussed in the respective sections in Chapter 3. In the following, a brief summary of the experimental systematic uncertainties affecting the $VH(H \to b\bar{b})$ boosted analysis is presented.

Small-R jet uncertainties

The two main sources of uncertainties affecting small-R jets are the Jet Energy Scale (JES) and the Jet Energy Resolution (JER). The former is characterised by 27 independent components, for a total uncertainty of $\sim 4.5\%$ at jet $p_T = 20$ GeV and $\sim 1\%$ at jet $p_T = 200$ GeV [114]. The total JER uncertainty ranges between 10% and 20% at jet $p_T = 20$ GeV, depending on the η value, and is less than 5% for jets with $p_T > 200$ GeV. JER uncertainties are parametrised through 8 independent components [115]. Finally, a specific O(%) uncertainty on the efficiency of the JVT is included [116].

Large-R jet uncertainties

The uncertainty on the large-R jet mass resolution (JMR) is the most significant uncertainty component. This is due to the fact that the analysis extracts the

signal as a *peak* in the large-R jet mass, and it is susceptible to any uncertainties on the mass resolution. The total relative uncertainty for the JMR is 20%. The uncertainty is treated as five independent NPs, one for each process: signal $VH(H \to b\bar{b})$, V+*jets*, $t\bar{t}$, single top, diboson. An additional uncertainty on the JER of the large-R jet of 2% is also included. Finally, the scale uncertainties on the mass (JMS) and the energy (JES) of the large-R jet are of the order of 5% and implemented as 12 independent components [119].

E_T^{miss} uncertainties

Since the E_T^{miss} is calculated from the jets and leptons with the addition of a soft term, two types of systematic uncertainties are considered: one resulting from the propagation of the scale and resolution uncertainties of the physics objects mentioned above, and one covering the soft term [127]. No E_T^{miss} trigger uncertainties are considered, since the trigger is fully efficient in the analysis phase space.

Lepton uncertainties

The lepton experimental systematic uncertainties take into account the lepton energy/momentum scale and resolution. Additional uncertainties cover the efficiencies of lepton triggers, reconstruction, identification, and isolation. Electrons and muons are considered separately, see Sections 3.3.1 and 3.3.3.

b-tagging uncertainties

The b-tagging uncertainties (see Chapter 4) are assessed from the calibration data (or purely from simulation) in various kinematic regions, and separately for b-, c, and light-flavour jets. The uncertainties are then decomposed in each of the flavour categories into independent components. The approximate size of the uncertainty on the tagging efficiency is 5% for b-jets, 10% for c-jets and 30% for the light-flavour jets [1].

Pile-up uncertainty

Scale-factor variations are introduced in the simulations to cover the data pile-up profiles across the years of data-taking. A pile-up uncertainty of the order of 3% is applied on the correction [165].

Luminosity uncertainty

The total combined uncertainty on the integrated luminosity for the full LHC Run 2 is $\sim 2\%$ [62].

5.6.2 Uncertainties on the modelling of signal and background processes

The $VH(H \to b\bar{b})$ boosted analysis is affected by a diverse background composition (see Section 5.2), in terms of processes and m_J spectra. For this reason, the background predictions cannot be fixed using data, but rely crucially on the usage of MC simulations. *Modelling* uncertainties on the accuracy of the MC prediction are therefore assessed.

For each physics process, the main MC generator (as seen in Table 5.1) is referred to as *nominal*. Alternative settings of the nominal generator, expressed as internal weights, or different generators, are then compared to the nominal generator in order to compute modelling uncertainties. Each setting or alternative generator evaluates a specific variation, i.e. a specific source of systematic uncertainty.

Before explaining the modelling strategy of the $VH(H \to b\bar{b})$ boosted analysis, it is necessary to explain the *tagging* strategy used in the modelling studies. Section 4.3.2 explains how b-tagging is conventionally used in the physics analyses. This strategy, also known as *direct tagging* is applied to a jet according to whether or not it passes the specific selection (given by the chosen working point) on the `MV2` discriminant. However, while this approach must be followed to tag jets in data, for MC samples,

removing the events that do not pass the MV2 selection considerably reduces the statistics for the modelling studies and leads to unreliable modelling uncertainties in the analysis. In the specific case of the $VH(H \to b\bar{b})$ boosted analysis, two b-jets are required at 70% efficiency, yielding a total selection efficiency of 49% for b-jets. This reduction is even more drastic for c- and light-flavour jets, due to their low tagging efficiency (know as *mis-tag rate*).

In order to solve this issue, *truth tagging* is applied. The procedure consists of associating a weight to each MC event, based on the expected probability for a jet to be tagged as a b-jet. In this way, all the events of a sample are retained, maximising the statistical power, and conserving the correct event normalisation.

The modelling uncertainties considered in this analysis are:

Normalisation uncertainties

Normalisation uncertainties cover the uncertainty on the total cross-section and acceptance for a given process. This type of uncertainty is calculated from a ratio between the number of events in the alternative sample and that in the nominal sample in the analysis phase space as:

$$\sigma_{norm} = 1 - \frac{N_{alt}}{N_{nom}} \tag{5.12}$$

This is done for the so-called *subdominant* background processes of the $VH(H \to b\bar{b})$ boosted analysis. Conversely, the normalisations (and their uncertainties) for the dominant backgrounds of the analysis ($t\bar{t}$ and $V + h.f.$) are left floating in the fit and therefore are directly constrained from data.

Relative acceptance uncertainties

Relative *acceptance double ratios* (AR) are evaluated to account for the event migration between two different categories. The structure of such ratios can be

described as below:

$$AR_{A\to B} = \frac{N^A_{var}}{N^A_{nom}} / \frac{N^B_{var}}{N^B_{nom}} - 1 \tag{5.13}$$

where the relative migration of events from category A to category B of the analysis is described. $N^A_{nom}(N^A_{var})$ is the number of nominal (variated) events in category A and $N^B_{nom}(N^B_{var})$ are the corresponding yields for category B of the analysis.

Categories A and B can be defined in different ways: for example they can be represented by the different analysis regions determined by the event categorisation. In this case, three ratios are studied, defined in correspondence of the categorisation splittings on p_T^V, signal-to-control regions and finally high-to-low purity signal regions. Other types of acceptance uncertainties, corresponding to different categories but following the structure from Equation 5.13, are discussed below.

Different uncorrelated sources of systematic uncertainties concur to the calculation of the acceptance uncertainties: the total acceptance is derived considering all the sources and summing their contribution in quadrature.

Shape uncertainties

Shape uncertainties keep the number of events within an analysis region constant, while varying the distribution of a chosen variable. In the $VH(H \to b\bar{b})$ boosted analysis the discriminant variable is the large-R jet mass m_J. In principle, the uncertainty originating from the comparison of the nominal and alternative generator in a given distribution is given by the alternative-to-nominal distribution ratio, however different techniques can also be adopted. In the following, the individual cases are discussed.

The modelling uncertainties are evaluated for the signal and for each of the background processes separately.

Signal $VH(H \to b\bar{b})$

The signal samples are normalised to the best theoretical prediction of the cross section for the different processes, and the $H \to b\bar{b}$ BR. The systematic uncertainties on the calculations are assigned following the recommendations of the LHC Higgs Cross Section Working Group [166, 167, 32, 168, 169].

The systematic uncertainties on the overall VH production cross section are obtained by varying the renormalisation and factorisation scales (μ_R, μ_F) by factors $2^{\pm 1}$, as well as the PDF sets. Additional uncertainties originate from missing higher order EW corrections, PDF+α_S variations. The effects of these uncertainties on the large-R jet mass shape are found to be negligible, while acceptance ratio uncertainties are evaluated. Finally, different parton shower and underlying event models are evaluated in acceptance and shape, using alternative samples generated with HERWIG7.

Vector boson + jets production (V+*jets*)

More than 90% of the V+*jets* background is represented by the $V + h.f.$ component, i.e. V+bb, V+bc, V+bl and V+cc, where the vector boson is reconstructed through its leptonic decays and the additional jets represent the "Higgs boson-candidate". The V+*jets* background is treated with separate NPs for the W+*jets* and the Z+*jets* component. The normalisations of $W + h.f.$ and $Z + h.f.$ are derived directly from the fit, that extracts them from the "sidebands" in the large-R jet mass spectrum in the 1- and 2-lepton channel respectively. For this reason, two extra degrees of freedom (*extrapolation uncertainties*), one for $W + h.f.$ and one for $Z + h.f.$, are introduced in the fit, to allow some freedom on the normalisation of these two component in the 0-lepton channel. The $W/Z + cl$ and $W/Z + ll$ components are normalised to the MC predictions.

Acceptance uncertainties are then evaluated on the event migration between the high purity and the low purity SR, the medium and the high p_T^V regions, and on the

flavour composition for $V + h.f.$. The latter are calculated as acceptances for which category A in Equation 5.13 is represented by the $V+bb$ component, while in category B fall the $V + bc$, $V + bl$ and $V + cc$ component one at the time. The main sources of uncertainties evaluated in these ratios are the variations in the renormalisation and factorisation scales (μ_R, μ_F, by factors $2^{\pm 1}$), as well as a different generation and shower model, obtained using MadGraph [170] interfaced to Pythia8, instead of Sherpa. PDF uncertainties are negligible.

Finally large-R jet mass shape uncertainties are derived from the comparison with the variation in the renormalisation scale μ_R. The other contributions are negligible.

Top quark pair production $t\bar{t}$

$t\bar{t}$ events in the $VH(H \to b\bar{b})$ boosted analysis

Top quark pair production is one of the main background processes in the 0- and 1- lepton channels, and dedicated CRs have been constructed to evaluate this background. Slightly more than 50% of the $t\bar{t}$ events in the analysis fall in the CRs, while the remaining part affects the signal regions. Events for which the $t\bar{t}$ pairs decay semi-leptonically are selected in about 96% and 85% of the cases for the 0- and 1-lepton channels, respectively. For these events, the hadronically-decaying top quark is (partially or fully) reconstructed within the large-R jet, as shown in Figure 5.11b. Dileptonically decaying $t\bar{t}$ pairs represent the remaining $t\bar{t}$ background in the analysis, while no fully hadronically $t\bar{t}$ events are present. This is due to the p_T^V selection criterion, that targets the vector boson's leptonic decay in signal events.

Within the semi-leptonically decaying $t\bar{t}$ events, a prevalence of τ + jets events is observed in the 0-lepton channel, via $t\bar{t} \to bW^+ \, \bar{b}W^- \to bj_1j_2 \, \bar{b}\tau\bar{\nu}_\tau$, with the τ lepton decaying hadronically. In the 1-lepton channel the μ+*jets* and e+*jets* decays occur at the same rate.

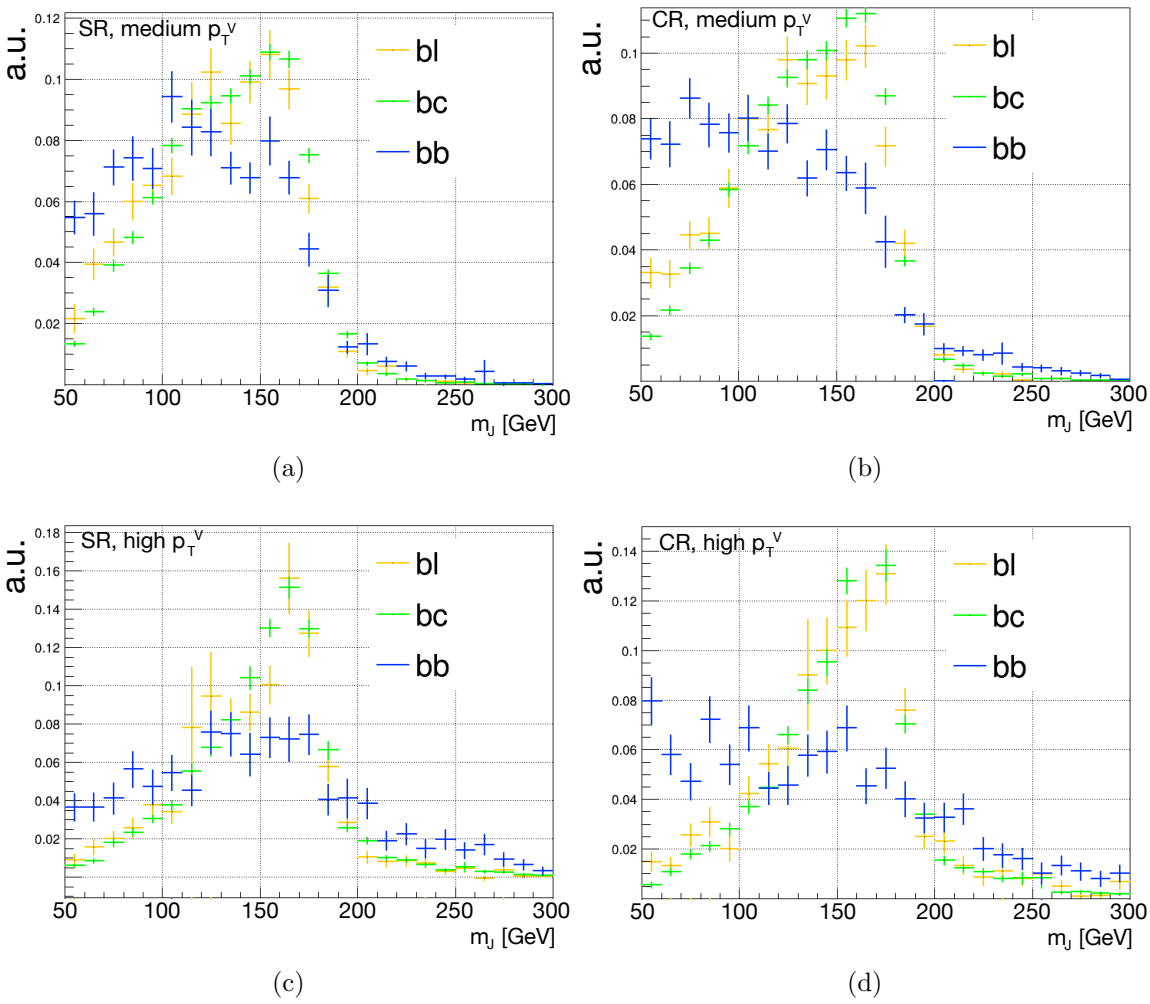


Figure 5.23: Large-R jet mass distributions for different flavour compositions of the large-R jet in $t\bar{t}$ events in the 0-lepton channel in the signal region (5.23a, 5.23c) and in the control region (5.23b, 5.23d), in the medium p_T^V region (5.23a, 5.23b) and in the high p_T^V region (5.23c, 5.23d). All the curves are normalised to the same area.

In order to further understand the topology of the $t\bar{t}$ events in the analysis, the flavour composition for the large-R jet in the 0-lepton and 1-lepton channel is studied. About 75% of the $t\bar{t}$ events that pass the event selection are those for which the leading two track-jets belong to the bc category, with the b-jet from the $t \to bW$ decay and the c-jet from the W boson decay. The bl and bb component are found to be sub-dominant. As shown in Figure 5.23, while the m_J mass distribution of the large-R jets containing bc and bl jets peaks at around the top quark mass, the bb distribution tends to be broader. Often, the mass of $t\bar{t}$-originated large-R jets is smaller than the top quark mass due to non-complete containment of the decay. As p_T^V increases, the probability of full containment increases, due to the larger boost. All three top quark decay products are contained in the large-R jets for 71% of the $t\bar{t}$ events in the high-p_T^V region. In contrast, 45% of the $t\bar{t}$ events in the medium-p_T^V region are those for which only two of the decay products of the top quark decay are contained in the large-R jet. As a result, the m_J distribution has a lower shoulder around 120 GeV in the medium-p_T^V region, both in the SR and the CR, as shown by bc and bl large-R jets in Figures 5.23a and 5.23b.

The $t\bar{t}$ background is negligible in the 2L channel, since dileptonic $t\bar{t}$ event are reduced by the analysis selection criteria.

Sources of systematic uncertainties for $t\bar{t}$ events

Several alternative $t\bar{t}$ samples have been studied to assess modelling systematic uncertainties for this process. The considered variations are:

- Powheg+Herwig7 sample, to provide a comparison with a different parton shower model with respect to the nominal prediction, while maintaining the same hard scattering simulation.
- Madgraph aMC@NLO+Pythia8 sample, to assess the systematic un-

certainties related to the use of a different model for the hard scattering generation, while maintaining the same parton shower model.

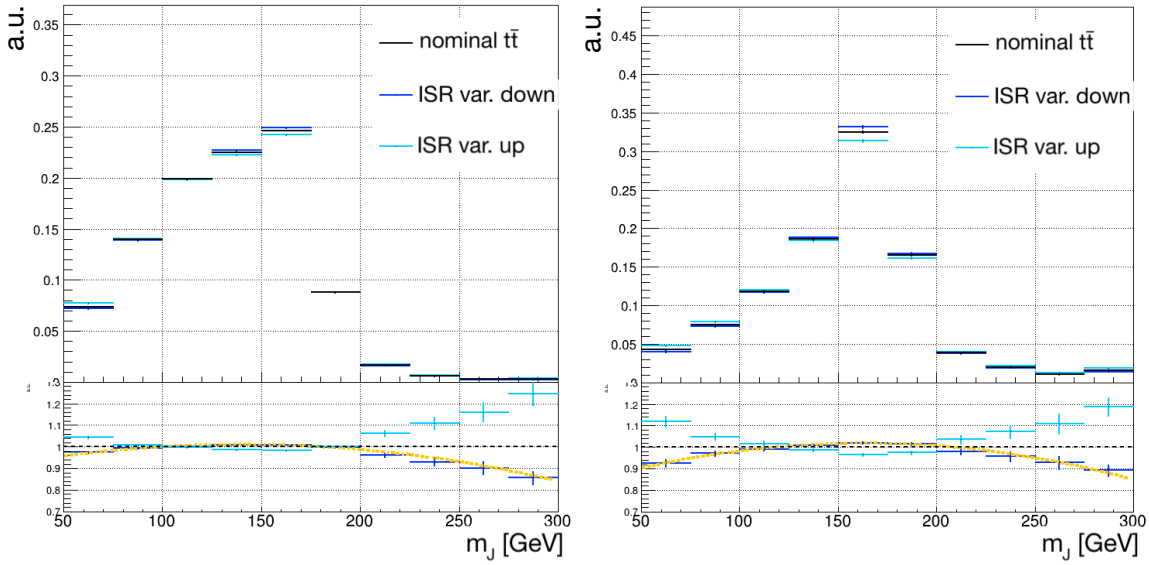
- *Initial state radiation* (ISR) and *final state radiation* (FSR) uncertainties are derived from the nominal sample, using different shower tuning parameters and renormalisation scales respectively.

$t\bar{t}$ modelling systematic uncertainties

The uncertainties are similar in the 0- and 1-lepton channels, indicating that the two channels probe a similar $t\bar{t}$ phase space. However, the $t\bar{t}$ modelling in the two channels is described using two sets of independent NPs: one for the 0-lepton channel and one for the 1-lepton channel, with the only exception of shape uncertainties that are in common between the two.

Two independent normalisation factors, for the two separate channels, are left floating in the fit, constrained in the CRs.

Four acceptance uncertainties are considered: three arise from the event categorisation and describe the event migration between the SR and the CR, the high purity and the low purity SR, and the medium- and high-p_T^V regions, while the fourth one accounts for the differences between partially and fully contained top quark decay within the large-R jet. In both channels and for all the acceptance double ratios, the total modelling uncertainty is $\leq 20\%$ and in particular is set to be 18% for the high-to-low purity signal region relative acceptance, 6% for the signal-to-control region relative acceptance and finally 20% for the medium-to-high p_T^V relative acceptance. A 20% value is chosen for the "containment uncertainty". The most significant source of systematic uncertainty on the acceptance ratios is the one from the hard scattering calculations, followed by the one on parton showers, while final state radiation (FSR) and initial state radiation (ISR) uncertainties are less significant.



(a) (b)

Figure 5.24: Large-R jet mass distributions in $t\bar{t}$ events in the 1-lepton channel for initial state radiation (ISR) variations, inclusively in the signal and control region, in the medium-p_T^V region (5.24a) and in the high-p_T^V region (5.24b). All the curves are normalised to the same area.

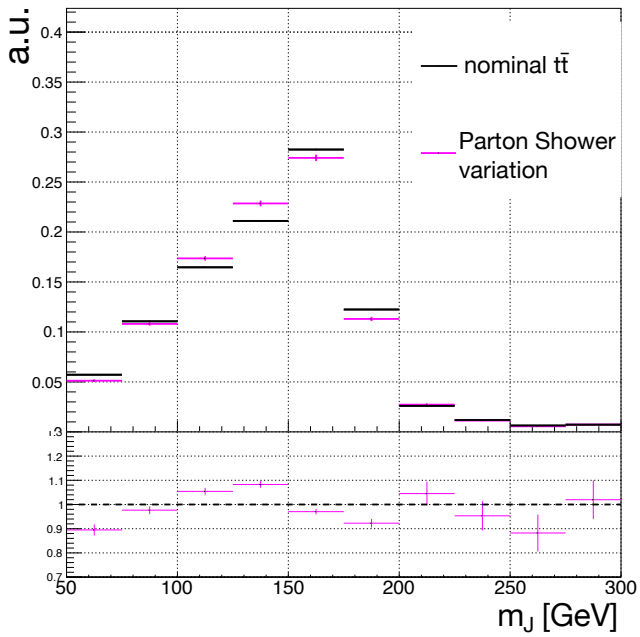


Figure 5.25: Large-R jet mass distributions in $t\bar{t}$ events for the parton shower variation (PS), obtained with Herwig7, inclusively in the signal and control region, in the medium- and high-p_T^V region, in the 0- and 1-lepton channel. All the curves are normalised to the same area.

In terms of shape uncertainties, there are two prominent contributions. The first one is given by the ISR variation, where a tail effect of the order up to 20% in both the 0- and 1- lepton channels can be seen. This shape variation is parametrised through a second order polynomial function, inclusively in the SR and the CR of the 1-lepton channel and separately for the medium- and high-p_T^V regions (see Figure 5.24). The variation is then applied both in the 0- and 1-lepton channel.

The second shape uncertainty on m_J is given by parton shower variations, obtained trough the comparison with the HERWIG7 sample. In this case an oscillating effect can be seen, in both the 0- and 1-lepton channels (Figure 5.25).

Single top

The single-top production contributes to the 0- and 1-lepton channels and arises from different production channels: s-, t- and Wt-channel. The latter accounts for $\sim 95\%$ of the single top production.

The normalisations of all single-top production processes are constrained to the MC predictions.

Additionally, acceptance and shape uncertainties are considered on the Wt-channel. The sources of uncertainties that are considered include ISR/FSR variations, parton shower variations (through the comparison with events generated with HERWIG7) and the hard scattering variations (from the comparison with MADGRAPH aMC@NLO). Finally an additional source of uncertainty is added, to account for the ambiguities in the simulation of the interference between $t\bar{t}$ and single-top production. This is done through a PYTHIA8 sample, using a different approach (*scheme*) for the treatment of this interference with respect to the nominal sample [171].

Diboson production

The diboson background process consists of final states arising mostly from WZ and

ZZ events, where a Z boson decays to a pair of b-quarks, and its normalisation is measured simultaneously with the VH signal in the two-POI fit. Similarly to what has been done for $V + h.f.$, three extra degrees of freedom (*extrapolation uncertainties*) are introduced in the fit, to allow the normalisation of ZZ in the 0- and 1-lepton channel, and of WZ in the 0-lepton channel to vary within the allowed range.

Variations in the renormalisation and factorisation scales, and in the hard scattering generation and parton showering given by the comparison with a POWHEG+PYTHIA8 sample are used in the medium-to-high p_T^V acceptance uncertainties, in the low-to-high purity acceptance uncertainty, and finally for shape systematic uncertainties.

Multijet background

As described in Section 5.2, the multijet background, due to semileptonic heavy-flavour-hadron decays or misidentified jets, is found to be negligible in the 0- and 2-lepton channels, and in the 1-lepton muon sub-channel, after applying the event selection criteria. In the 1-lepton electron sub-channel, the residual contribution is evaluated using a data-driven method, since MC simulation samples are statistically limited and are not expected to reproduce the multijet production in this region of the phase space.

The data-driven procedure consists of creating a dedicated multijet-enriched region, obtained by inverting the *signal* lepton isolation requirement, and by removing the requirement on $E_T^{miss} > 50$ GeV. At this point, the invariant transverse mass distribution $m_T^W = \sqrt{2p_T^e E_T^{miss}\,(1 - cos(\Delta\phi(e, E_T^{miss})))}$ is evaluated. The contributions from the other electroweak (EW) processes (V+*jets*, $t\bar{t}$, single top, diboson) determined from MC simulations are subtracted and the remaining distribution is fitted. The variable m_T^W is chosen because it offers an optimal discrimination between QCD- and EW-induced processes. The multijet contribution in the 1-lepton electron sub-channel is 2.3% of the total background, with an uncertainty of 55% estimated

mainly from the statistical uncertainty of the transverse mass fit. This uncertainty is set as the normalisation uncertainty for the multijet background in the final two-POI fit.

5.7 Analysis results

The results of the $VH(H \to b\bar{b})$ search at high transverse momentum, performed with the full Run 2 data by the ATLAS detector, corresponding to an integrated luminosity of 139 fb^{-1} at a centre of mass energy $\sqrt{s} = 13$ TeV, are reported in this section.

The global likelihood fit in the three lepton channels yields a signal strength parameter for the $VH(H \to b\bar{b})$ and $VZ(Z \to b\bar{b})$ processes of:

$$\mu_{VH}^{b\bar{b}} = 0.72^{+0.39}_{-0.36} = 0.72^{+0.29}_{-0.28}\,(\text{stat.})^{+0.26}_{-0.22}\,(\text{syst.})$$

$$\mu_{VZ}^{b\bar{b}} = 0.91^{+0.29}_{-0.23} = 0.91 \pm 0.15\,(\text{stat.})^{+0.25}_{-0.17}\,(\text{syst.})$$

corresponding to an observed (expected) significance $Z_{VH}^{b\bar{b}} = 2.1(2.7)$ and $Z_{VZ}^{b\bar{b}} = 5.4(5.7)$ standard deviations respectively. The total uncertainty is of $\sim 54\%$ and 32% on $\mu_{VH}^{b\bar{b}}$ and $\mu_{VZ}^{b\bar{b}}$ respectively, with the statistical uncertainty accounting for $\sim 60\%$ of the total. Both measurements are in agreement with the SM expectations, and the VZ measurement is compatible with the $W^{\pm}Z$ differential cross-section measurement performed by ATLAS at high transverse momentum of the Z boson ($p_T^Z > 220$ GeV) in the fully leptonic final state ($W^{\pm}Z \to \ell\nu\,\ell^+\ell^-$) [172].

Figure 5.27a shows the large-R jet mass distribution, summed over all channels and analysis regions, weighted by their respective values of the ratio of fitted Higgs boson signal-to-background yields and after the subtraction of all background processes, except for the WZ and ZZ diboson processes. In Figure 5.27b, the likelihood contours

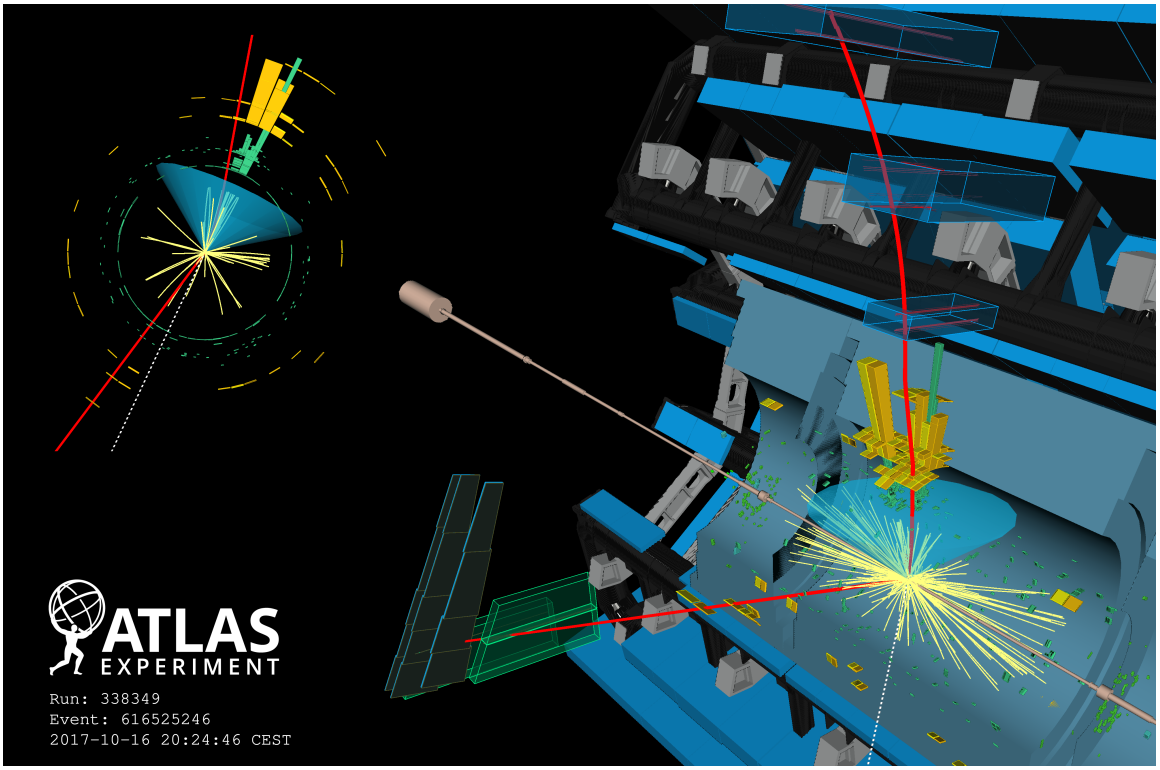
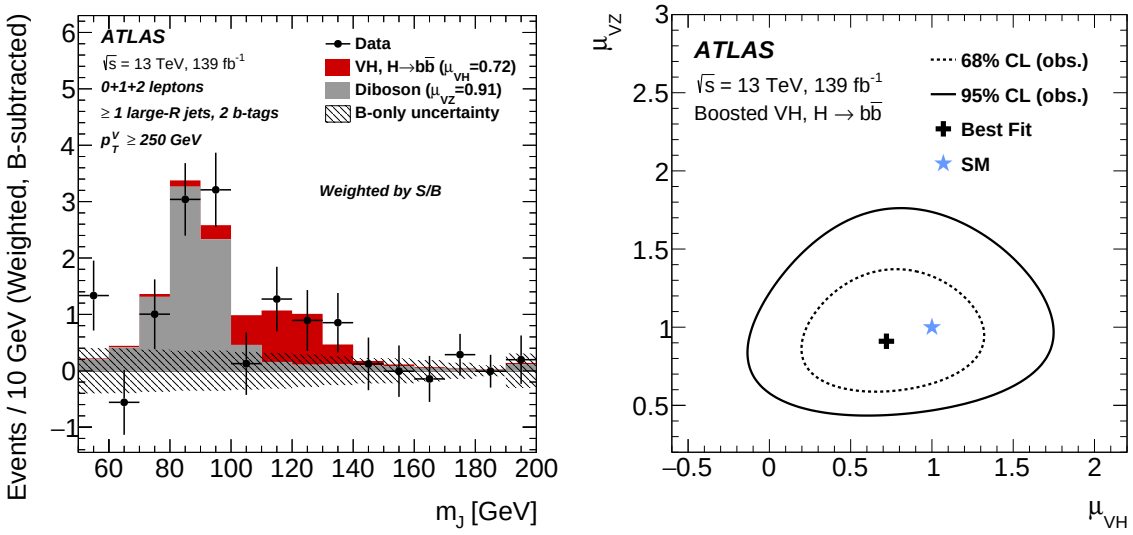


Figure 5.26: Event display of a candidate event for the process $WH \to b\bar{b}$ (Run 338349, Event 616525246). An isolated muon is shown as the lower red track producing hits in the endcap-muon chambers; it has a transverse momentum of 196 GeV and a pseudorapidity of 1.16. The missing transverse momentum is identified by a white dashed line and it has a magnitude of 832 GeV; the transverse momentum (mass) of the muon-E_T^{miss} system is 1.03 TeV (112 GeV). The Higgs-boson candidate is reconstructed as a single large-R jet (blue cone) containing two b-tagged sub-jets built from tracking information. The jet has a transverse momentum of 973 GeV and a mass of 116 GeV. A clear two-prong structure is visible as energy deposits in both the electromagnetic (green) and hadronic (yellow) calorimeters. A second muon (upper red track), shown with hits in the barrel muon chambers, is found close to the Higgs-boson candidate, being a potential "muon-in-jet".



(a) (b)

Figure 5.27: Large-R jet mass distribution in data after subtraction of all backgrounds except for the WZ and ZZ diboson processes. The contributions from all lepton channels and signal regions are summed and weighted by their respective values of the signal-to-background ratio. The size of the combined statistical and systematic uncertainty is indicated by the hatched band (5.27a). Observed contours at 68% and 95% confidence level (CL) in the ($\mu_{VH}^{b\bar{b}}$, $\mu_{VZ}^{b\bar{b}}$) plane. The cross indicates the best-fit value and the star the SM prediction. The correlation between $\mu_{VH}^{b\bar{b}}$ and $\mu_{VZ}^{b\bar{b}}$ is 11%.

for the Higgs boson and diboson signal strength are shown. The correlation among the two signal strength parameters is approximately 11%.

The normalisations of the largest backgrounds, $t\bar{t}$ (in the 0- and 1-lepton channels), $W+h.f.$ and $Z+h.f$, left unconstrained in the fit, are shown in Table 5.7. The scale factors deviate from unity, the difference has been attributed to a poor modelling in the simulation, already observed in other ATLAS analyses [3, 173].

A detailed breakdown of the uncertainties can be found in Table 5.8. In this result the largest uncertainties are the statistical uncertainties, that include the impact from the floating normalisations. The impact of systematic uncertainties, although lower, is of the same order as the total statistical uncertainty. The dominant source of

Process	Normalisation factor
$t\bar{t}$ 0-lepton	0.88 ± 0.10
$t\bar{t}$ 1-lepton	0.83 ± 0.09
$W + h.f.$	1.12 ± 0.14
$Z + h.f.$	1.32 ± 0.16

Table 5.7: Floating normalisation values for the main backgrounds of the $VH(H \to b\bar{b})$ boosted analysis.

Source of uncertainty		Avg. impact
Total		0.372
Statistical		0.283
Systematic		0.240
Experimental uncertainties		
small-R jets		0.038
large-R jets		0.133
E_T^{miss}		0.007
Leptons		0.010
b-tagging	b-jets	0.016
	c-jets	0.011
	light-flavour jets	0.008
	extrapolation	0.004
Pile-up		0.001
Luminosity		0.013
Theoretical and modelling uncertainties		
Signal		0.038
Backgrounds		0.100
$\hookrightarrow Z$ + jets		0.048
$\hookrightarrow W$ + jets		0.058
$\hookrightarrow t\bar{t}$		0.035
$\hookrightarrow$ Single top quark		0.027
$\hookrightarrow$ Diboson		0.032
$\hookrightarrow$ Multijet		0.009
MC statistical		0.092

Table 5.8: Breakdown of the absolute contributions to the uncertainty in μ_{VH}^{bb} inclusive in p_T^V. The sum in quadrature of the systematic uncertainties attached to the categories differs from the total systematic uncertainty due to correlations. The reported values represent the average between the positive and negative uncertainties on μ_{VH}^{bb}.

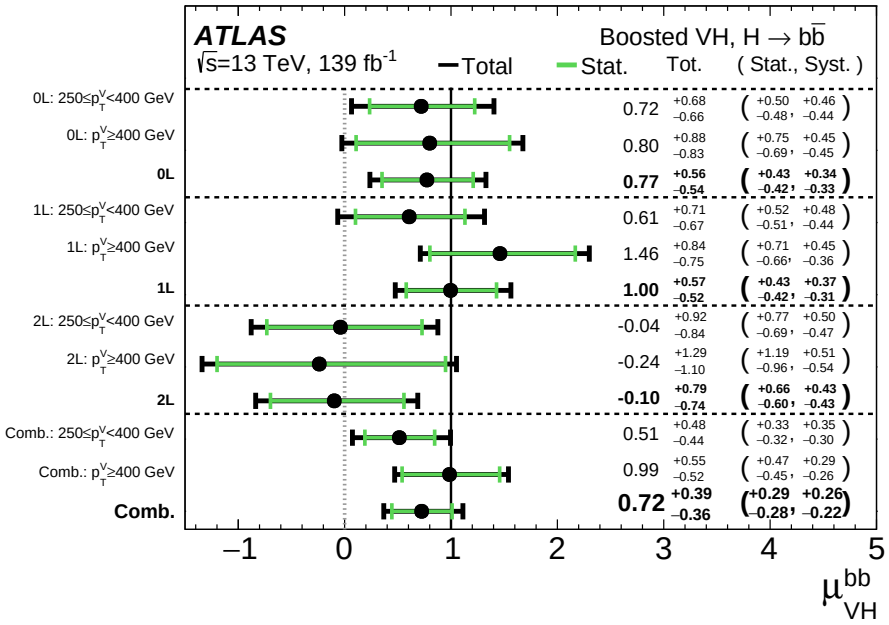
systematic uncertainty is the large-R jet mass resolution. The background modelling uncertainties follow, along with the statistical uncertainties due to the limited statistic of the MC samples. Further details on the systematic uncertainty NPs can be found in Appendix B.

Several cross-checks of the measurement were performed. Figure 5.28 shows the results of several multi-POI fits on the signal strength parameters $\mu_{VH}^{b\bar{b}}$ and $\mu_{VZ}^{b\bar{b}}$. Figure 5.28a shows the results of a six-POI fit for $VH(H \to b\bar{b})$ and a simultaneous one-POI fit for $\mu_{VZ}^{b\bar{b}}$, as well as a three-POI fit for $VH(H \to b\bar{b})$ and a simultaneous one-POI fit for $\mu_{VZ}^{b\bar{b}}$. The six $VH(H \to b\bar{b})$ POIs are the individual signal strengths parameters for $VH(H \to b\bar{b})$ in each of the three channels and two p_T^V regions, while the three POIs correspond to the signal strengths parameters in each individual lepton channels. Figure 5.28b shows the "mirrored" results of multi-POI fits for the signal strength parameter for $VZ(Z \to b\bar{b})$ and a simultaneous one-POI fit for $\mu_{VH}^{b\bar{b}}$. All results are compatible and in agreement with the SM predictions. A statistical under-fluctuation of data can be observed in the 2-lepton channel.

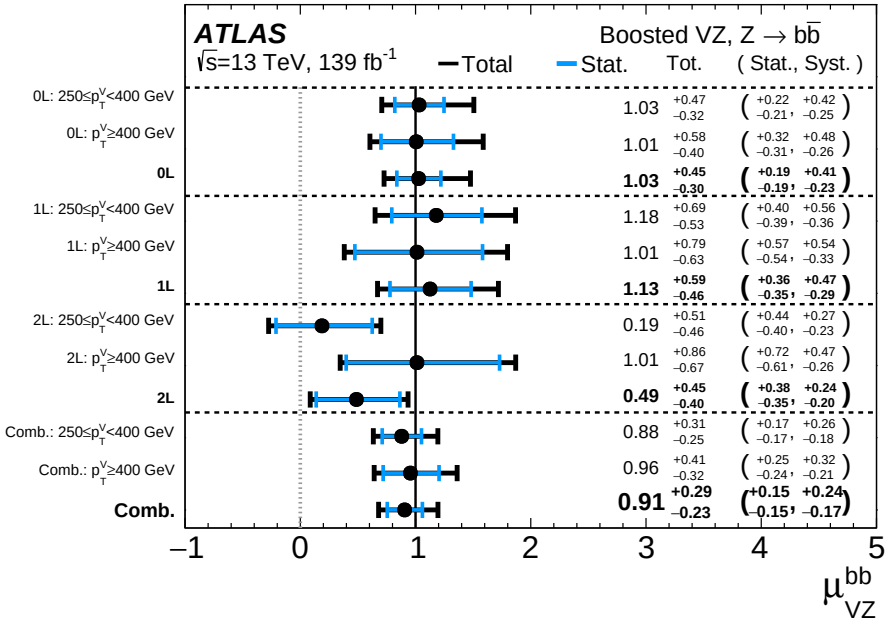
5.8 Simplified Template Cross Section and Effective Field Theory Interpretation

After the inclusive cross section measurement, a differential cross section measurement is provided, in *simplified fiducial regions*, using the Simplified Template Cross Section (STXS) approach [174].

In this case, a fiducial region is a portion of the phase-space defined at truth-level, that is populated by particles with a relatively long lifetime $> 10^{-8}$ s. At the detector-level, particles are reconstructed as physics objects within the detector acceptance, with a certain efficiency. The goal of a fully fiducial measurement would be to perform a measurement and thus define a selection for which the truth-level



(a)



(b)

Figure 5.28: Fitted values of the Higgs boson signal strength parameter $\mu_{VH}^{b\bar{b}}$ (5.28a) and the $VZ(Z \to b\bar{b})$ signal strength parameter $\mu_{VZ}^{b\bar{b}}$ (5.28b) for the 0-, 1- and 2-lepton channels in different p_T^V regions separately and for various combinations.

and the detector-level are as close as possible, in order to minimise uncertainties due to the extrapolation between the two.

A first step towards a fully fiducial measurement can be made by introducing the STXS framework [175, 176], by dividing the cross-sections into sub-cross sections, known as *bins*. Another advantage of STXS measurements is that the Branching Ratio is factorised out of the measurement, reducing theory-dependent factors originating from combining measurements.

The STXS measurement presented here is obtained in four bins, defined by the production channel WH or ZH and the transverse momentum of the vector boson, as seen in the event categorisation: $p_T^V \in [250; 400[$ GeV and $p_T^V \in [400; \infty[$. The expected signal distributions (as well as the acceptance and efficiencies) for each STXS bin are estimated from the signal samples, by selecting events using the generator *truth information*.

The global likelihood fit used to extract the STXS measurement is slightly different from the one described in Section 5.5: it has multiple POIs, corresponding to the cross sections times BR($V \to \ell\ell, H \to b\bar{b}$) in the four regions. The systematic uncertainty are defined in Section 5.6, however the theoretical cross-section and branching fraction uncertainties on the signal samples are not included in the STXS measurement, since they affect only the signal strength measurements.

Figure 5.29 shows the measured VH simplified template cross-sections times the $H \to b\bar{b}$ and $V \to$ leptons branching fractions, where the red error bands correspond to the theoretical uncertainty of the fiducial cross-section prediction in each bin.The measurements are in agreement with the SM expectations within the uncertainties, that span from $\sim 70\%$ to 90% depending on the STXS bin. The statistical under-fluctuation observed on the signal strength measurement in the 2-lepton channel is reflected in the two STXS bins showing the ZH contribution.

For the first time in ATLAS, a measurement of $VH(H \to b\bar{b})$ in a dedicated bin

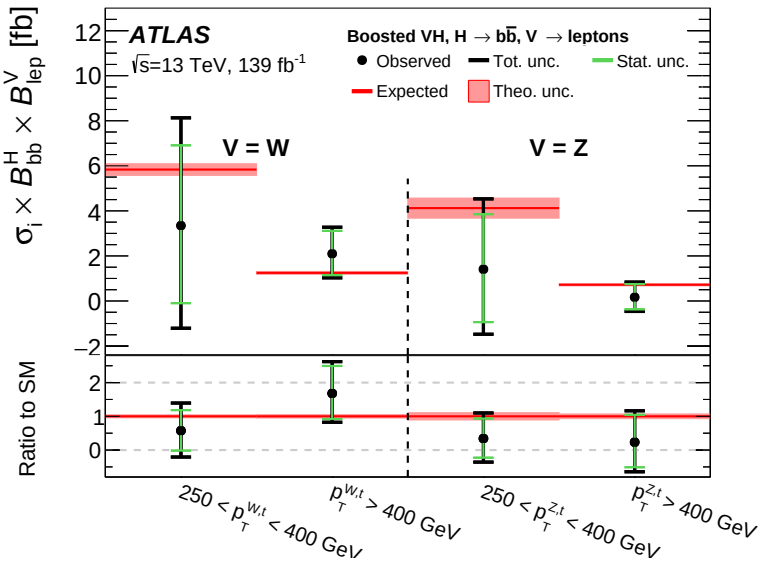


Figure 5.29: Measured VH simplified template cross-sections times the $H \to b\bar{b}$ and $V \to$ leptons branching fractions.

above 400 GeV in p_T^V is performed.

Finally, the STXS results are interpreted in an Effective Field Theory (EFT) approach (see Section 1.6.1), for which the SM Lagrangian is extended with higher-dimensional operators that depend on their Wilson coefficients (c_i), as shown in Table 5.9. The VH production cross section and the $H \to b\bar{b}$ decay are then parametrised as linear/quadratic functions of these coefficients and the 68% and 95% C.L. limits for the coefficients are extracted. The linear terms originate from the interference between the SM and non-SM amplitudes and are of order $1/\Lambda^2$ (see Equation 1.32), while the quadratic ones, of order $1/\Lambda^4$, come from the squared non-SM amplitudes. It is to be noted that in this picture the $1/\Lambda^4$ terms are incomplete since the interference between SM and dimension-8 operators is not considered. Parameter Λ is set to 1 TeV.

As shown in Figure 5.30, all limits are compatible with null values of the coefficients, and therefore with the SM predictions.

Finally, it is extremely interesting to appreciate the effect of the dedicated $p_T^V >$

Coefficient	Operator
$c_{H\Box}$	$(H^\dagger H)\Box(H^\dagger H)$
c_{HDD}	$(H^\dagger D^\mu H)^*(H^\dagger D_\mu H)$
c_{dH}	$(H^\dagger H)(\bar{q}_p d_r H)$
c_{HW}	$H^\dagger H W^I_{\mu\nu} W^{I\mu\nu}$
c_{HB}	$H^\dagger H B_{\mu\nu} B^{\mu\nu}$
c_{HWB}	$H^\dagger \tau^I H W^I_{\mu\nu} B^{\mu\nu}$
$c^{(1)}_{Hl}$	$H^\dagger i \overleftrightarrow{D}_\mu H(\bar{l}_p \gamma^\mu l_r)$
$c^{(3)}_{Hl}$	$H^\dagger i \overleftrightarrow{D}^I_\mu H(\bar{l}_p \tau^I \gamma^\mu l_r)$
$c^{(1)}_{He}$	$H^\dagger i \overleftrightarrow{D}_\mu H(\bar{e}_p \gamma^\mu e_r)$
$c^{(1)}_{Hq}$	$H^\dagger i \overleftrightarrow{D}_\mu H(\bar{q}_p \gamma^\mu q_r)$
$c^{(3)}_{Hq}$	$H^\dagger i \overleftrightarrow{D}^I_\mu H(\bar{q}_p \tau^I \gamma^\mu q_r)$
c_{Hu}	$H^\dagger i \overleftrightarrow{D}_\mu H(\bar{u}_p \gamma^\mu u_r)$
c_{Hd}	$H^\dagger i \overleftrightarrow{D}_\mu H(\bar{d}_p \gamma^\mu d_r)$
c'_{ll}	$(\bar{l}_p \gamma_\mu l_r)(\bar{l}_s \gamma^\mu l_t)$

Table 5.9: Wilson coefficients and their corresponding dimension-6 operators considered in this analysis [53, 177].

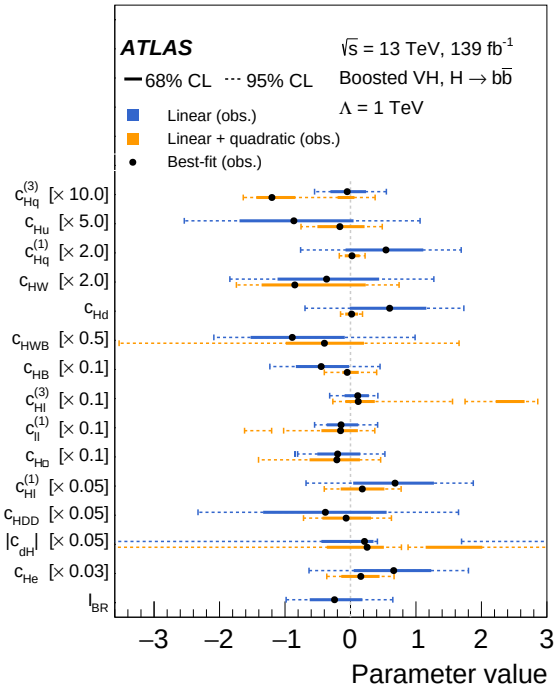


Figure 5.30: Summary of the observed 68% (solid lines) and 95% (dashed lines) CL limits for the Wilson coefficients of the EFT operators from a fit of the STXSs, using a linear-only parametrisation in blue and including quadratic terms in orange. In addition, the constraint on modifications of the $H \to b\bar{b}$ branching ratio $\Delta BR/BR_{SM}$ is included. Coefficients are measured in one-POI fits, where the Wilson coefficients for the other operators are fixed to the SM expectations (0).

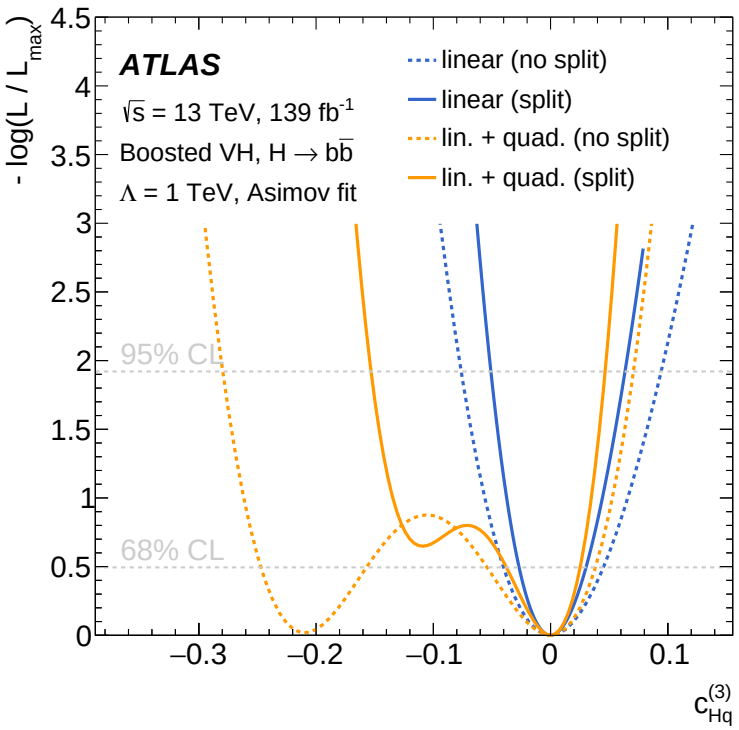


Figure 5.31: Expected results for the negative log-likelihood as a function of a single Wilson coefficient $c_{Hq}^{(3)}$ is shown using the two p_T^V bins in the analysis (solid lines) and using a single bin of $p_T^V > 250$ GeV (dashed lines). Wilson coefficients of all other operators are set to zero. The blue curves are valid when using a linear parametrisation of the VH production cross section and the Higgs decay branching fraction, while the orange lines also include quadratic terms in the parametrisation. The choice of coefficient $c_{Hq}^{(3)}$ lies in the fact that it is the coefficient for one of the BSM EFT operators to which the $VH(H \to b\bar{b})$ boosted analysis is most sensitive to.

400 GeV STXS bin on the constraints of the EFT operators, as shown in Figure 5.31. The plot shows how the dedicated high-p_T bin allows to set more stringent limits on coefficient $c_{Hq}^{(3)}$ (solid lines), with respect to the case where only one bin $p_T^V > 250$ GeV is present (dashed lines). This is true both in the linear (blue) and linear+quadratic approximation (orange).

Chapter 6

Future measurements of $VH(H \to b\bar{b})$

The previous chapter described the $VH(H \to b\bar{b})$ *boosted* analysis. On the same time scale, another analysis using the entire Run 2 data sample has been developed to study $VH(H \to b\bar{b})$ at lower p_T regimes [3]. This analysis, called $VH(H \to b\bar{b})$ *resolved* analysis, reconstructs the Higgs boson candidate through two small-R jets, as it was done "traditionally" for the $VH(H \to b\bar{b})$ analysis in ATLAS [151, 164].

The $VH(H \to b\bar{b})$ boosted analysis is optimised for boosted reconstruction techniques and is developed independently from the resolved analysis, from a common baseline. Since the resolved analysis results from the improvement of the "traditional" ATLAS $VH(H \to b\bar{b})$ analysis, it is highly refined and very sensitive to the signal $VH(H \to b\bar{b})$. On the other hand, the boosted analysis has a higher signal efficiency for events with $p_T^V \geq 350$ GeV, due to the performance of boosted reconstruction techniques. It is also expected to provide better high-p_T STXS measurements, and thus tighter constraints on the EFT parameters. It is, therefore, interesting to combine the two analyses to cover the full p_T spectrum and provide a more precise measurement of $VH(H \to b\bar{b})$. This combination requires understanding the significant *overlap*

between the two analyses.

In this Chapter, some of the possible developments for a combined *legacy* $VH(H \to b\bar{b})$ analysis will be discussed. Section 6.1 summarises briefly the resolved analysis, while overlap events are described in Section 6.2. Possible combination strategies are discussed in Section 6.3, and Section 6.4 presents a potential improvement for the combined analysis, given by the introduction of an effective $t\bar{t}$-enhanced region in the resolved spectrum. In the following, results with an **ATLAS** label are taken from the resolved publication [3].

6.1 The $VH(H \to b\bar{b})$ resolved analysis

There are many similarities between the $VH(H \to b\bar{b})$ resolved and the boosted analysis described in Chapter 5, including the analysis strategy, the data and simulation samples, the object and event selection, the statistical analysis and the estimation of systematic uncertainties. The main traits of the resolved analysis and the main differences with respect to the boosted analysis are highlighted in the following.

6.1.1 Analysis strategy

The main conceptual difference between the boosted and the resolved $VH(H \to b\bar{b})$ analyses is given by the reconstruction techniques used for the Higgs boson candidate. In the $VH(H \to b\bar{b})$ resolved analysis the Higgs boson candidate is reconstructed using two b-tagged small-R calorimeter jets.

The reconstruction of the V boson candidate is identical to the boosted analysis, and three analysis channels are considered: $ZH \to \nu\bar{\nu}\, b\bar{b}$ (0-lepton channel), $WH \to \ell\nu\, b\bar{b}$ (1-lepton channel) and $ZH \to \ell^-\ell^+\, b\bar{b}$ (2-lepton channel).

6.1.2 Signal and Background processes

Signal events are from WH and ZH production, via quark- and gluon-induced diagrams.

The main background processes of this analysis are the production of a vector boson in association with jets (V+*jets*) and the production of $t\bar{t}$ pairs. Single top production plays an important role in the 1-lepton channel and a sub-dominant role in the two other lepton channels. Diboson production has a similar experimental signatures to the signal $VH(H \to b\bar{b})$. Lastly, the multijet background, copiously produced at the LHC, plays a negligible role in the 0-lepton and 2-lepton channel, and a minor role in the 1-lepton channel, where it is estimated using data-driven techniques. The evaluation of all other processes relies on MC simulations, except for the estimation of $t\bar{t}$ production in the 2-lepton channel.

The data used in the $VH(H \to b\bar{b})$ resolved analysis is from the pp collisions collected by the ATLAS experiment in the full LHC Run 2, from 2015 to 2018, corresponding to an integrated luminosity of 139 fb^{-1} at a centre of mass energy $\sqrt{s} = 13$ TeV.

6.1.3 Event selection and categorisation

In all channels, events are required to have exactly two b-tagged jets from the decay of the Higgs boson candidate, and at least one of the b-tagged jets must have $p_T \geq 45$ GeV. Events are categorised as "2-jet", if no additional non-b-tagged jets are found in the event, or "3-jet", if such jets are present. In the 0-lepton and 1-lepton channels, only one further jet is permitted, while in the 2-lepton channel any number of jets is accepted in the 3-jet category.

The reconstructed vector boson transverse momentum p_T^V corresponds to E_T^{miss} in the 0-lepton channel, to the vectorial sum of $\vec{E}_T^{miss}$ and of the charged lepton

transverse momentum in the 1-lepton channel, and to the transverse momentum of the dilepton system in the 2-lepton channel. The analysis focuses on events with $p_T^V \geq 150$ GeV. In the 2-lepton channel, the sensitivity is increased by the addition of a lower-p_T^V category, 75 GeV $\leq p_T^V < 150$ GeV.

Additional requirements, optimised in order to maximise the background rejection, are applied in each specific channel. In particular, a series of angular cuts are applied in the 0-lepton channel to reduce the QCD background.

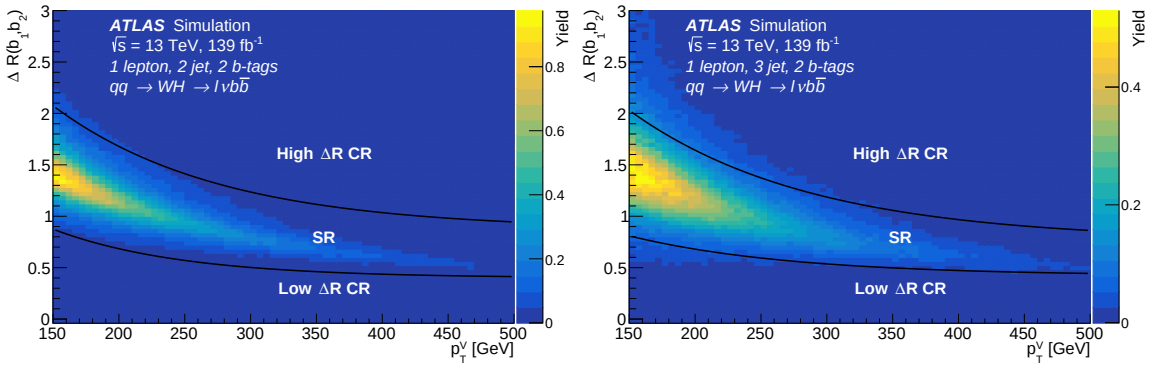
Events passing the analysis selection are classified according to multiple criteria in signal and control regions. The categorisation criteria are on the number of charged leptons (0-, 1-, and 2-lepton channels), on the number of jets in the event ("2-jet" and "3-jet"), on the transverse momentum of the reconstructed vector boson p_T^V (for which regions of $p_T^V \in [75; 150[$ GeV, $[150; 250[$ GeV and $[250; +\infty[$ are constructed) and finally on a continuous selection based on two p_T^V-dependent $\Delta R(b_1, b_2)$ criteria. The continuous selection is shown in Figure 6.1. The selections, initially designed for the 1-lepton channel, define control regions based on $\Delta R(b_1, b_2)$ in all three lepton channels. The regions with large $\Delta R(b_1, b_2)$ (High-ΔR-CR) are enriched in $t\bar{t}$ events, while the ones with small $\Delta R(b_1, b_2)$ (Low-ΔR-CR) are dominated by V+*jets* events. In addition, in the 2-lepton channel, six[1] very pure $t\bar{t}$ control regions ($e\mu$-CRs) are defined. In the $e\mu$-CRs, the same-flavour lepton requirement in the 2-lepton event selection is replaced by requiring exactly one electron and one muon.

6.1.4 Multi Variate Analysis (MVA)

In order to maximise the discrimination between signal and background, several Boosted Decision Trees (BDTs) are trained to cover the phase space of the different analysis categories.

BDTs are a multivariate technique widely used in High Energy Physics to dis-

[1]The six regions are given by categories: (2-jet, 3-jet)$\times$([75; 150[GeV, [150; 250[GeV, $[250; +\infty[)p_T^V$ categories.



(a) (b)

Figure 6.1: Signal $WH(H \to b\bar{b})$ yield distribution of the ΔR between the two b-tagged jets, $\Delta R(b_1, b_2)$, as a function of p_T^V in the 1-lepton channel for 2-b-tag events, in the 2-jet (6.1a) and exactly 3-jet (6.1b) categories for $p_T^V \geq 150$ GeV. The lines demonstrate the continuous lower and upper selection on $\Delta R(b_1, b_2)$ used to categorise the events into the signal and control regions.

criminate a signal process from its backgrounds [178]. In the $VH(H \to b\bar{b})$ resolved analysis, the Gradient Boosting algorithm is chosen for the training of the BDTs.

A set of observables, summarised in Table 6.1 is combined into the BDTs, designed to separate Higgs boson events from the expected backgrounds and yield the final discriminant variable used in the analysis. The most powerful variables for the discrimination are m_{bb}, $\Delta R(b_1, b_2)$ and p_T^V . The BDT is implemented with the TMVA (Toolkit for Multivariate Data Analysis) package in ROOT [142]. The BDT is trained on MC simulations, and it is evaluated on both data and MC simulations. Subsequently, the signal strength is extracted from the fit of the output score distribution in the signal regions. The yields of the dominant backgrounds are determined from the control regions, using a single bin of p_T^V. The global maximum likelihood fit used in the $VH(H \to b\bar{b})$ resolved analysis has the same structure as the one in the boosted analysis described in Section 5.5. Also the systematic uncertainties, due to experimental and modelling effects, are derived similarly to Section 5.6.

Variable	0-lepton	1-lepton	2-lepton
m_{bb}	×	×	×
$\Delta R(\vec{b_1}, \vec{b_2})$	×	×	×
$p_T^{b_1}$	×	×	×
$p_T^{b_2}$	×	×	×
p_T^V	$\equiv E_T^{miss}$	×	×
$\Delta\phi(\vec{V}, \vec{bb})$	×	×	×
MV2(b_1)	×	×	
MV2(b_2)	×	×	
$\lvert\Delta\eta(\vec{b_1}, \vec{b_2})\rvert$	×		
m_{eff}	×		
$p_{\text{T}}^{\text{miss,st}}$	×		
E_T^{miss}	×	×	
$\min[\Delta\phi(\vec{\ell}, \vec{b})]$		×	
m_T^W		×	
$\lvert\Delta y(\vec{V}, \vec{bb})\rvert$		×	
m_{top}		×	
$\lvert\Delta\eta(\vec{V}, \vec{bb})\rvert$			×
$E_T^{miss}/\sqrt{S_{\text{T}}}$			×
$m_{\ell\ell}$			×
$\cos\theta(\vec{\ell^-}, \vec{Z})$			×
	Only in 3-jet events		
$p_T^{\text{jet}_3}$	×	×	×
m_{bbj}	×	×	×

Table 6.1: Variables used as input for the multivariate discriminant in each of the channels, where the × symbol indicates the inclusion of a variable.

6.1.5 Results

The measured signal strength for the $VH(H \to b\bar{b})$ resolved analysis is:

$$\mu_{VH}^{b\bar{b}} = 1.02^{+0.18}_{-0.17} = 1.02^{+0.12}_{-0.11}\,(\text{stat.})^{+0.14}_{-0.13}\,(\text{syst.}) \tag{6.1}$$

well compatible with the Standard Model Higgs boson within the errors. The measured signal strength corresponds to a rejection of the background-only hypothesis with an observed (expected) significance of 6.7 (6.7) standard deviations. The combined fit has also been performed separately on the signal strengths for the WH and ZH productions, respectively observed (expected) at 4.0 (4.1) and 5.3 (5.1) standard deviations in significance.

Equation 6.1 shows that the experimental precision of the $VH(H \to b\bar{b})$ resolved analysis is limited by systematic uncertainties. For this reason, simply adding more data during the next period of data taking (Run 3) would not significantly improve the precision of the measurement. Therefore, analysis strategy improvements are necessary. One such improvement could be the introduction of an additional control region, as explained in Section 6.4.

Also in this case, a differential cross section measurement is provided, in simplified fiducial regions, using the Simplified Template Cross Section (STXS) approach. The STXS measurement is performed in five bins, divided according to the production channel (WH and ZH) and p_T^V ($[150; 250[$ GeV and $[250; +\infty[$). ZH production is also measured in an additional p_T^V bin: $[75; 150[$ GeV. Finally, the STXS results are interpreted using an Effective Field Theory (EFT) approach.

All the obtained results are fully compatible with the SM predictions.

Resolved $VH(H \to b\bar{b})$	Boosted $VH(H \to b\bar{b})$
Higgs boson candidate	
≥ 2 or 3 small-R jets, $p_T > 20$ GeV, $\lvert\eta\rvert < 2.5$	≥ 1 large-R jets, $p_T > 250$ GeV, $\lvert\eta\rvert < 2.0$, $m_J > 50$ GeV
	≥ 2 track jets, $p_T > 10$ GeV, $\lvert\eta\rvert < 2.5$, matched to large-R jet
2 b-tagged jets, $p_T^{b_1} > 45$ GeV	Leading 2 matched track jets b-tagged
Vector boson candidate	
$E_T^{miss} \geq 250$ GeV	$E_T^{miss} \geq 250$ GeV

Table 6.2: Summary of the event selection for the $VH(H \to b\bar{b})$ boosted-resolved overlap studies in the 0-lepton channel.

6.2 Studies of the overlap of the $VH(H \to b\bar{b})$ resolved and boosted analyses

It has already been mentioned that the $VH(H \to b\bar{b})$ boosted analysis presented in Chapter 5 and the $VH(H \to b\bar{b})$ resolved analysis described in the previous section significantly overlap. Therefore, the two results cannot be straightforwardly combined.

Before discussing possible combination strategies, it is interesting to understand the number of signal events selected by both analyses and their properties. The study summarised in the following is carried out in the 0-lepton channel and the applied event selection is summarised in Table 6.2. Only events with $E_T^{miss} \geq 250$ GeV are considered. Furthermore, the study is conducted inclusively in the SRs and CRs of both the analyses. In this region, $\sim 36\%$ of simulated signal events are selected by both analyses, while $\sim 37\%$ is selected exclusively by the resolved analysis and the remaining 27% is only reconstructed by the boosted analysis. These fractions depend on p_T^V, as shown in Figure 6.2. At 250 GeV almost 70% of events are reconstructed exclusively by the resolved analysis, while the remaining 30% is equally divided between events reconstructed only in the boosted analysis and events belonging to the overlap. However, as p_T^V increases, the fraction of events reconstructed only in the boosted analysis increases to above 60%. The percentage of signal events in the "over-

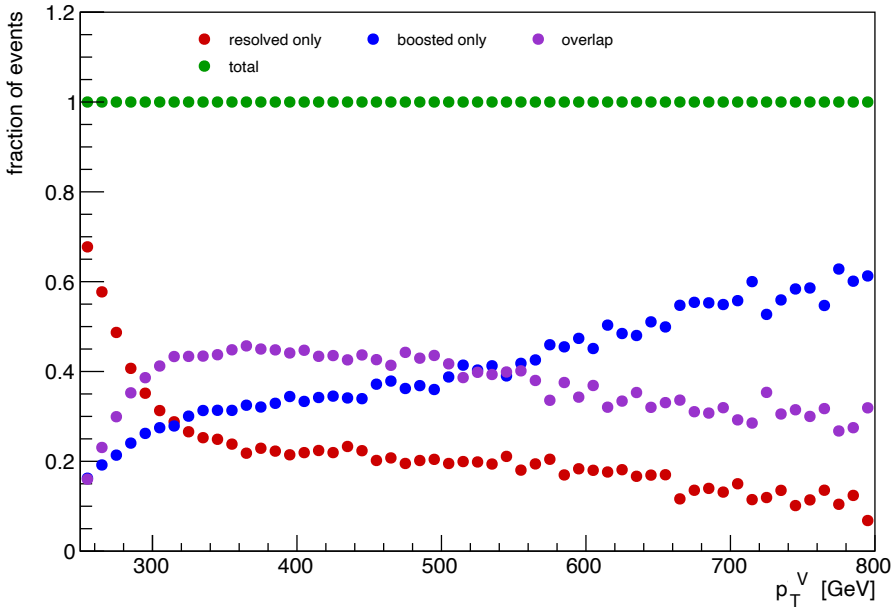
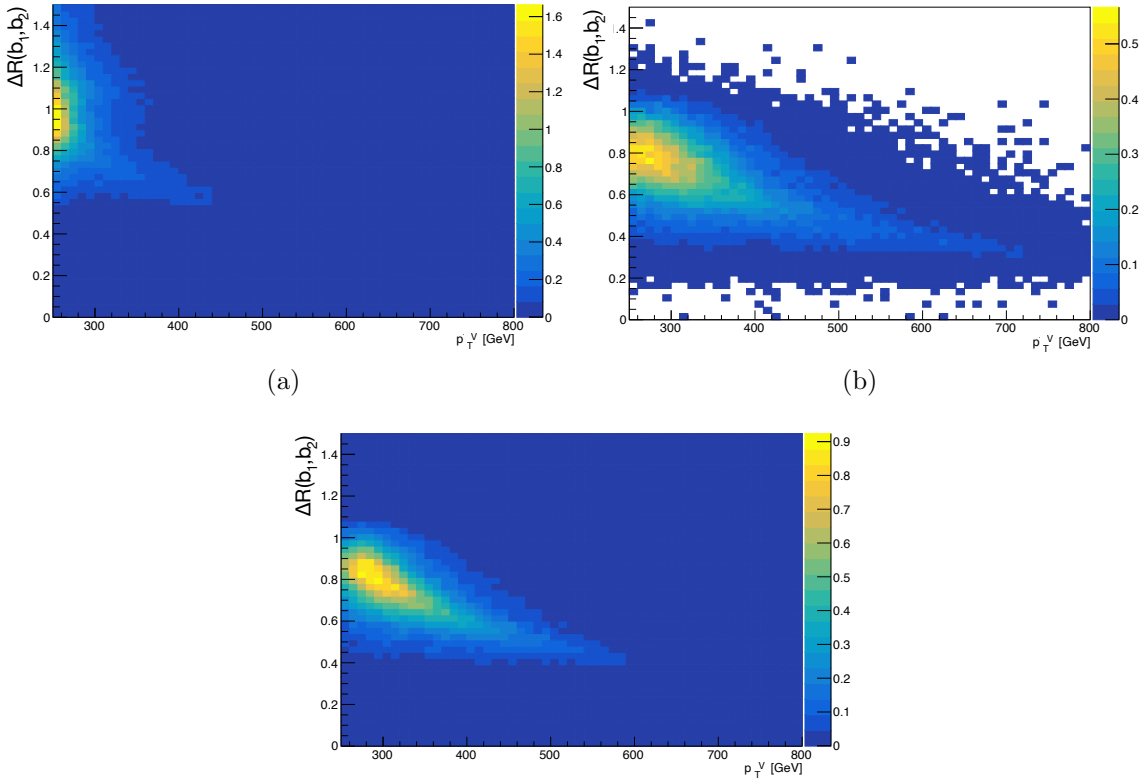


Figure 6.2: Fraction of signal $VH(H \to b\bar{b})$ events selected only by the resolved analysis (red), only by the boosted analysis (blue), or both (overlap events - purple), as a function of p_T^V.

lap" is $\geq 30\%$ across most of the p_T^V spectrum, rising to above 40% in the p_T^V range between 300 GeV and $\sim$ 500 GeV. Similar studies in the 1- and 2-lepton channels confirm this behaviour.

After understanding the classification of signal $VH(H \to b\bar{b})$ events in resolved-only, boosted-only or overlap, a more detailed study has been carried out. The goal of these studies is to acquire a better knowledge of overlap events. With this in mind, some of the key variables in the analyses have been studied in more detail.

Figure 6.3 shows the $\Delta R(b_1, b_2)$ distributions as a function of p_T^V for signal $VH(H \to b\bar{b})$ events that are reconstructed only by the resolved analysis (Figure 6.3a), only by the boosted analysis (Figure 6.3b), or both (Figure 6.3c). It can be seen that resolved-only events concentrate at values of $\Delta R(b_1, b_2)$ around 1.0, with a tail of events at lower $\Delta R(b_1, b_2)$ that extends up to $\sim$ 450 GeV in p_T^V. Conversely, boosted-only events populate the region below $\Delta R(b_1, b_2) = 1.0$, as a consequence of the large-R jet $R = 1.0$ requirement. The distributions of boosted events extends towards lower



(a) (b)

(c)

Figure 6.3: $\Delta R(b_1, b_2)$ distribution as a function of the vector boson transverse momentum, for signal $VH(H \to b\bar{b})$ events, with $p_T^V > 250$ GeV for resolved only events (6.3a), boosted only events (6.3b) and overlap events (6.3c).

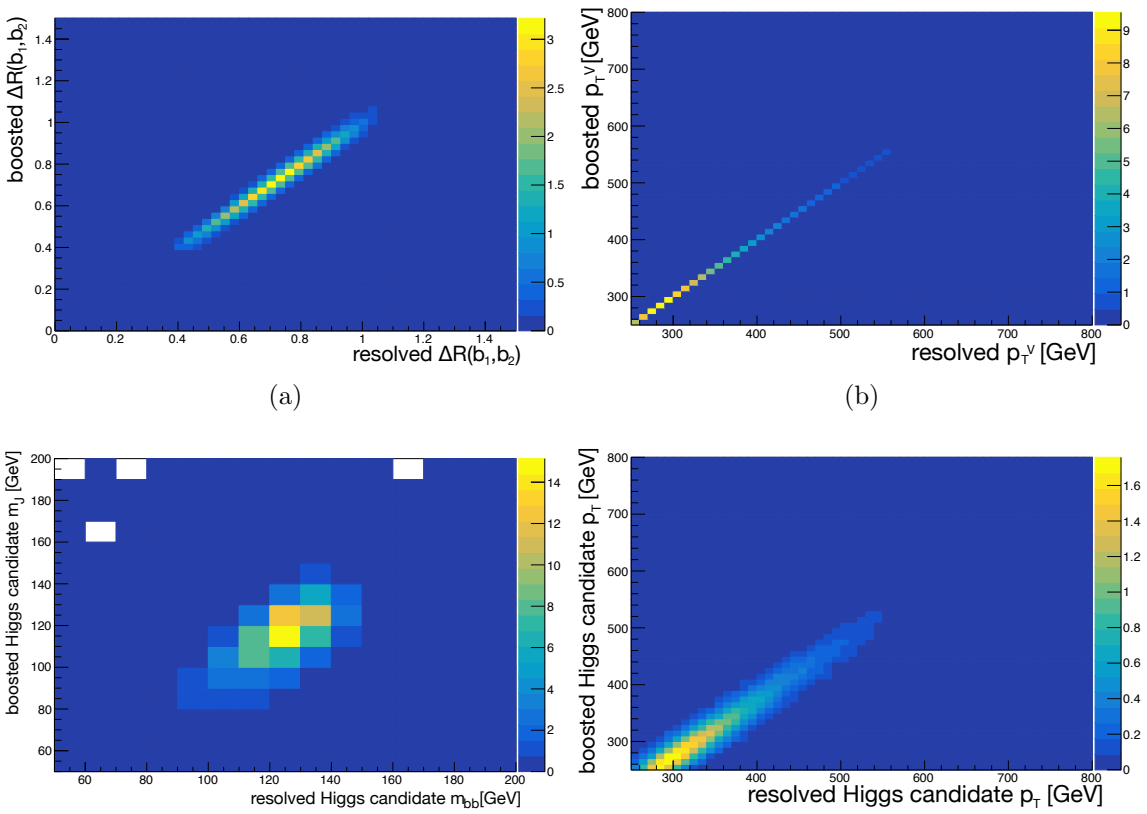


Figure 6.4: $\Delta R(b_1, b_2)$ (6.4a), p_T^V (6.4b), Higgs boson candidate mass (6.4c) and p_T (6.4d) distributions for signal $VH(H \to b\bar{b})$ overlap events. On the y-axis the variable is obtained through the boosted reconstruction, while on the x-axis the same variable is obtained through the resolved reconstruction.

$\Delta R(b_1, b_2)$ values, up to $\Delta R(b_1, b_2) = 0.2$ at $p_T^V = 700$ GeV. Overlap events show intermediate kinematic properties, peaking at $\Delta R(b_1, b_2) = 0.9$ and extending down to $\Delta R(b_1, b_2) = 0.4$ at p_T^V =600 GeV.

Figure 6.4 presents the kinematic distributions of four key-variables in the $VH(H \to b\bar{b})$ analyses for signal events in the overlap. For each plot, the "same" kinematic variable is shown, for the boosted reconstruction (y-axis) and for the resolved one (x-axis). Studying these distributions could unveil unexpected features due to differences in the reconstruction techniques of the two analyses. The $\Delta R(b_1, b_2)$ and p_T^V of overlap events follow a diagonal distribution (see Figures 6.4a and 6.4b), indicating that the boosted and resolved reconstruction techniques yield consistent results. This does

not occur for the Higgs boson candidate mass (m_J, m_{bb}) and p_T (p_T^J, p_T^{bb}), as shown in Figures 6.4c and 6.4d. The boosted reconstruction yields values of m_J (p_T^J) that are lower than m_{bb} (p_T^{bb}). Since the large-R jet mass (m_J) and p_T (p_T^J) are correlated, the p_T of the Higgs boson candidate has been studied. The two reconstructed Higgs boson candidate p_T (p_T^J and p_T^{bb}) show a discrepancy of $\sim 7\%$. The comparison of p_T^J and p_T^{bb} with the Higgs boson *truth* p_T in MC ($p_T^{H,truth}$) shows that, on average, the boosted p_T^J underestimates $p_T^{H,truth}$ by 5%, while the resolved p_T^{bb} overestimates it by 2%. These findings represent a starting point for potential improvements in the Higgs boson candidate reconstruction in the legacy analysis.

6.3 Studies of the combination strategy of the $VH(H \to b\bar{b})$ resolved and boosted analyses

There are several ways in which the $VH(H \to b\bar{b})$ resolved and boosted analyses can be combined. For example, one could think of taking the two independent analyses described above and combining them using a single multi-dimensional likelihood function. However, although this is possible in principle, the two analyses are significantly different in the strategy (i.e. only one uses multivariate techniques), and they are described using many nuisance parameters, that also present many differences. Such nuisance parameters are not always completely statistically independent, they can be highly correlated and the correlations are often not to be trivially determined. Therefore, the most practical approach is to consider the two analyses as separate for the events that are exclusively reconstructed by one of the two, and then decide on a strategy to treat the overlapping events discussed above.

The studies of overlap signal events presented in the previous section indicate that the combination of the resolved and boosted $VH(H \to b\bar{b})$ analyses cannot be performed by simply setting a p_T^V threshold below (above) which events would be

reconstructed using the resolved (boosted) reconstruction. Furthermore, the study of the kinematic properties of overlap events has not identified any feature to achieve the optimal performance in this region.

In addition, while the overlap studies provide useful information on the number of selected signal events, no information about the number of background events is provided. Therefore, the signal-only studies need to be completed with studies assessing which analysis strategy will give the best sensitivity for the measurements of signal strength and to constrain the EFT parameters.

A first step towards the estimation of the sensitivity of the analysis, without the necessity to perform a full profile likelihood fit, is given by the binned statistical significance formula, presented in Equation 5.2.

In the following, a first combination strategy is described. The study is carried out in the 0-lepton signal regions, with the event selection summarised in Table 6.2. The significances of the different strategies are obtained from the binned statistical significances of the resolved and boosted $VH(H \to b\bar{b})$ analyses. The "resolved" and "boosted" significance terms are obtained separately and then summed in quadrature into the final significance.

The binned statistical significance for the resolved terms is obtained separately for the 2-jet and 3-jet categories, using the BDT output score as the variable of interest. The final "resolved" significance term is then calculated as the sum in quadrature of the two partial terms obtained separately in the two 2-jet and 3-jet categories.

The statistical significance for the boosted term is obtained separately in the two p_T^V regions ($[250; 400[$ GeV, $[400; \infty[$ GeV) and the two high- and low purity signal regions, described in Section 5.3.3. The discriminant variable is the large-R jet mass m_J. The final "boosted" significance term is then calculated as the sum in quadrature of the terms obtained in the separate categories.

Figure 6.5 shows the different combination strategies analysed in the study. Events

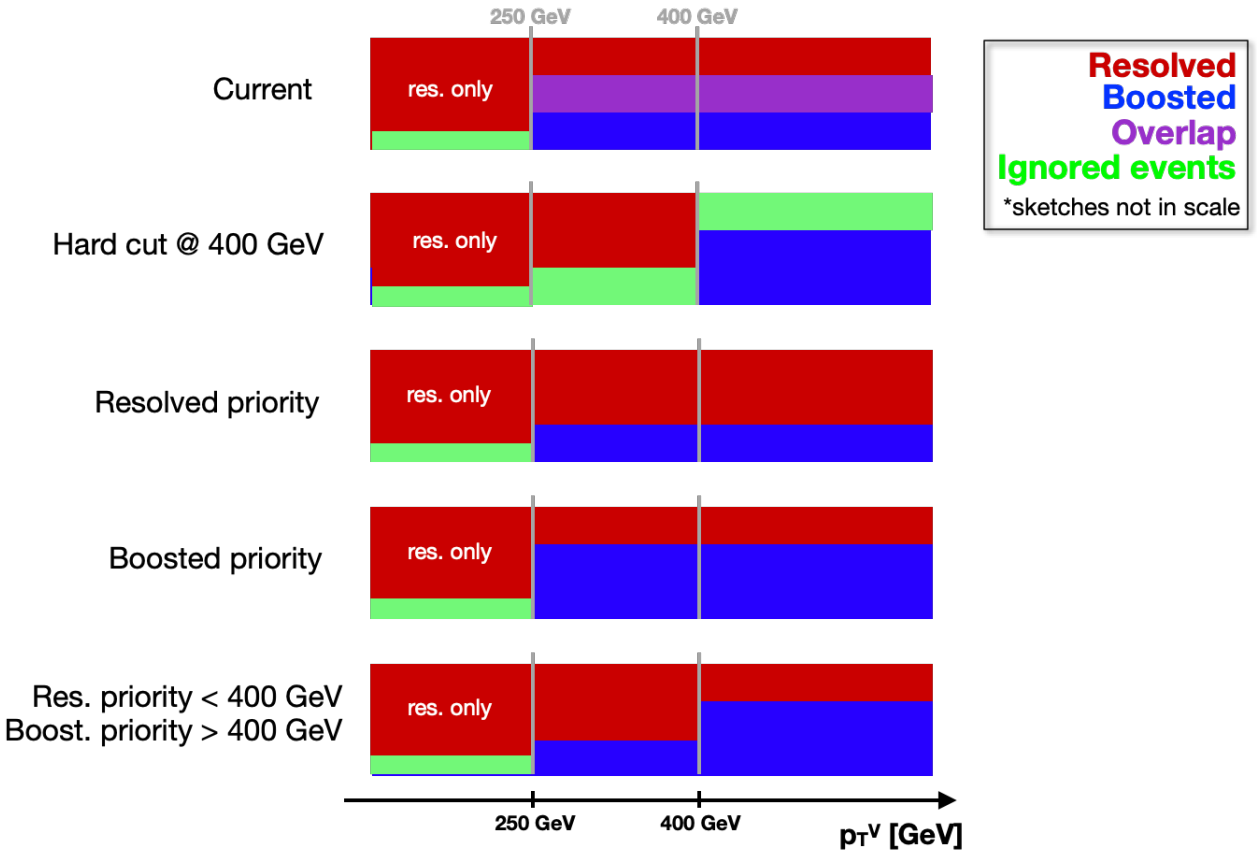


Figure 6.5: Sketch of the possible different combination strategies for the $VH(H \to b\bar{b})$ boosted and resolved analyses. Events reconstructed with different analysis techniques are represented using different colours: red for the resolved reconstruction, blue for the boosted reconstruction, purple for events in the overlap between the two, green for ignored events. The p_T^V dependence of the different combination criteria is also shown.

reconstructed with different analysis techniques are represented using different colours for the resolved reconstruction (red), for the boosted reconstruction (blue) and for the overlap (purple). Finally, *ignored events*, that would not be used in some of the combination strategies, are shown in green. The sketch provides also a schematic view of the p_T^V dependence of the different combination criteria. For $p_T^V < 250$ GeV, only the resolved analysis is used. A small fraction of signal events, that could be reconstructed using boosted techniques are ignored given the $p_T^V \geq 250$ GeV selection of the $VH(H \to b\bar{b})$ boosted analysis. The five combination possibilities will be now briefly described.

The *current* option shows what is currently done with both the resolved and the boosted $VH(H \to b\bar{b})$ analyses, using events with $p_T^V \geq 250$ GeV, yielding a significant overlap.

The second option, called *Hard cut @ 400 GeV*, is the simplest alternative and entails using the resolved reconstruction for events with $p_T^V < 400$ GeV and the boosted reconstruction above that threshold. In this case, all the "boosted only" events with $p_T^V < 400$ GeV, and the "resolved only" events with $p_T^V \geq 400$ GeV are not used and would be discarded in the combined analysis. This would imply a loss in signal $VH(H \to b\bar{b})$ events of $\sim 20\%$ of the total.

The third and fourth options *Resolved priority* and *Boosted priority* would instead make use of all the available signal events above $p_T^V = 250$ GeV. These would be obtained by reconstructing the overlap events using only the resolved and boosted reconstruction respectively.

Finally, the fifth option is a combination of the two above, also exploiting all the available signal $VH(H \to b\bar{b})$. It consists of assigning the overlap events to the resolved reconstruction in the range 250 GeV $\leq p_T^V < 400$ GeV, and to the boosted reconstruction above 400 GeV in p_T^V.

For all the possible combination strategies the combined statistical significance

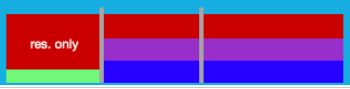
Strategy	Significance (stat. only), $p_T^V > 250$ GeV	Total Significance (stat. only)
Resolved (current) 	3.98	5.04
Boosted (current) 	2.83	4.20
Hard cut @ 400 GeV 	3.91	**4.99**
Resolved priority 	4.24	**5.25**
Boosted priority 	3.48	**4.66**
Res. Priority < 400 GeV Boost. Priority > 400 GeV 	4.11	**5.15**

Figure 6.6: Statistical significances in units of standard deviations for the possible different combination strategies for the $VH(H \to b\bar{b})$ boosted and resolved analyses in the 0-lepton channel. In the second column the significances obtained in region $p_T^V \geq 250$ GeV are reported. The last column shows the total statistical significance, obtained by summing the second column with the stat. only significance of 3.1 standard deviations, obtained in the resolved analysis for 0-lepton events with 150 GeV $\leq p_T^V < 250$ GeV.

of the $VH(H \to b\bar{b})$ boosted-resolved analysis has been calculated in the 0-lepton channel, as shown in Figure 6.6. In this Figure the significances in units of standard deviations obtained in region $p_T^V \geq 250$ GeV are reported in the second column. The last column, instead, shows the total statistical significance, obtained by summing the second column with the stat. only significance of 3.1 standard deviations, obtained in the resolved analysis for 0-lepton events with 150 GeV $\leq p_T^V < 250$ GeV[2].

The results in Figure 6.6 show that there is no clear optimal choice, since all strategies yield significances within a 13% band, which become 5% for the three options with the highest significance. The combination strategy that has the highest absolute statistical significance is the one where the resolved reconstruction is prioritised with respect to the boosted reconstruction. This can be easily explained considering that the resolved analysis is highly optimised and therefore more refined than the boosted analysis, only recently introduced. The determining factors are the presence of the BDT, and the usage of objects that are well understood. For example, the resolved analysis uses small-R jets that are well calibrated, while the calibration of the large-R jets utilised in the boosted analysis is not yet reaching the same precision. However, the introduction of a boosted multivariate discriminant, and the implementation of a more precise large-R jets calibration should improve the performance of the boosted analysis.

The strategy prioritising the resolved analysis below 400 GeV and the boosted analysis above that threshold achieves the second highest significance. The loss in significance is only 2% with respect to the highest-significance option.

The final combination strategy will be decided shortly and will be implemented in the legacy analysis.

[2]This value is obtained from Figure 28 in Reference [3], by approximating $Z_{stat.only} \sim \frac{\mu}{\Delta\mu_{stat.only}}$, and assuming the contribution to the total ZH significance equally split between the 0-lepton and the 2-lepton channel.

6.4 Studies of three *b*-jet events in the $VH(H \to b\bar{b})$ resolved analysis

A potential improvement for a future combined $VH(H \to b\bar{b})$ analysis is presented in this section. In Section 6.1.3 the event categorisation of the resolved $VH(H \to b\bar{b})$ analysis has been described, and the presence of a control region enriched in $t\bar{t}$ events in the 2-lepton channel (the $e\mu$-CR), and in the 1-lepton channel (the High-ΔR-CR) has been discussed. Conversely, in the 0-lepton channel an effective $t\bar{t}$ control region is not present, and the normalisation of the $t\bar{t}$ background and its shape are derived from the 1-lepton channel. Given the importance of the $t\bar{t}$ background in the $VH(H \to b\bar{b})$ analysis, the introduction of a $t\bar{t}$-enriched region in the 0-lepton channel of the resolved analysis has been studied.

The selection of this $t\bar{t}$ region is derived from the $VH(H \to b\bar{b})$ boosted analysis, where such control region (see Section 5.3.3) in the 0- and 1-lepton channel is defined by requiring at least one b-tagged track jet outside the large-R (Higgs boson candidate) jet. Essentially, this implies requiring events to have at least *three b-tagged jets*, two of which define the Higgs boson candidate. The same requirement can therefore be applied in the resolved analysis, by defining a *3b-tagged jet control region* (3b-CR) in the 3-jet category.

The boosted analysis $t\bar{t}$ control regions, pure in $t\bar{t}$ at 52%-80% level (depending on the lepton channel and p_T^V), have proven extremely useful in controlling the $t\bar{t}$ background normalisation and large-R jet mass shape. In Section 5.6.2 the topology of the $t\bar{t}$ events in the $VH(H \to b\bar{b})$ boosted analysis has been discussed and it has been explained how more than 75% of the selected $t\bar{t}$ events have a true b- and c-jet as the two jets in which the Higgs boson candidate is reconstructed. It is important to control the bc component because it is very "signal-like", considering that the bc invariant mass from $t \to bW \to b\,cj$ is close to the Higgs boson mass. The bc

Process	Yields
Signal $VH(H \to b\bar{b})$	3.232 ± 0.048
diboson	11.91 ± 0.63
Z+*jets*	164.2 ± 4.15
W+*jets*	41.7 ± 2.9
single top	45.3 ± 3.0
$t\bar{t}$	559.6 ± 6.1
Total background	822.8
Data	918

Table 6.3: Signal, background and data yields in the 3b-CR in the 0-lepton channel of the $VH(H \to b\bar{b})$ resolved analysis, for 150 GeV $\leq p_T^V <$ 400 GeV.

component is very "signal-like" also in the resolved analysis BDT, since m_{bb} has a high discriminating power. For this reason, a powerful $t\bar{t}$ (bc) control region is of great usefulness.

In the following, the introduction of the 3b-CR in the 0-lepton channel of the $VH(H \to b\bar{b})$ resolved analysis is discussed. The event selection is described in Section 6.1.3, and an additional requirement on $p_T^V(E_T^{miss}) < 400$ GeV is applied, given the high efficiency of the resolved reconstruction up to that threshold. Above 400 GeV the boosted analysis is used, exploiting the effective $t\bar{t}$ control regions previously discussed. Furthermore, this additional selection makes the studies directly comparable to the *Hard cut @ 400 GeV* combination strategy (from the previous section). In the following the topology of events in the 3b-CR will be described with a focus on $t\bar{t}$ events, then the addition of the 3b-CR to the resolved spectrum of the $VH(H \to b\bar{b})$ analysis will be studied.

6.4.1 Event topology in the 3b-CR

Table 6.3 presents the event yields in the 3b-CR, divided between the signal and background simulated processes, and also gives the number of events for the recorded data from pp collisions. Even if at this stage normalisation scale factors are not

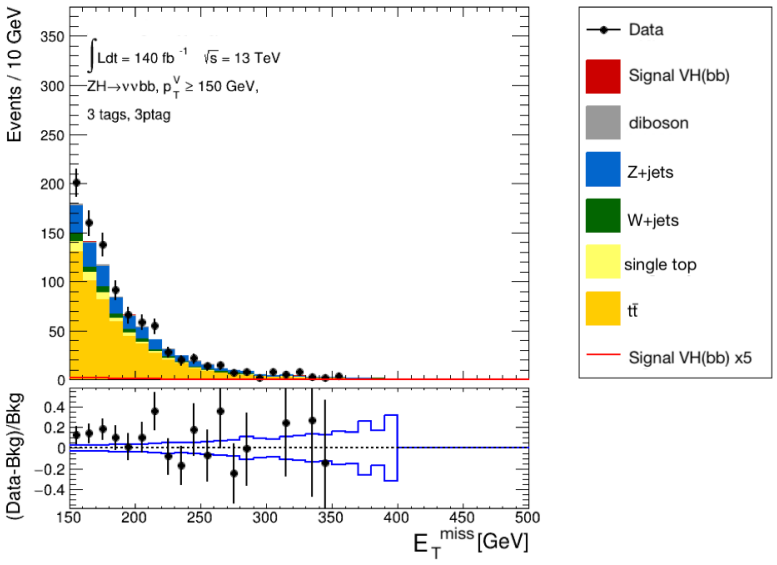


Figure 6.7: E_T^{miss} distribution in the 3b-CR in the 0-lepton channel of the $VH(H \to b\bar{b})$ resolved analysis, for 150 GeV $\leq p_T^V <$ 400 GeV. The stacked signal and background contributions are compared to data. In the lower panel, the ratio of the data minus the background component over background is shown. The blue bands represent the statistical uncertainty from the simulated background.

applied to MC simulations[3], they model the data well. This can be seen also from Figure 6.7, where the E_T^{miss} distribution in the 3b-CR is shown. The 3b-CR is pure in $t\bar{t}$ at the 68% level, with the second most important contribution (20% of the total yields) given by the V+*jets* background.

It is interesting to study the distributions of the key-variables in the analysis, such as the mass of the Higgs candidate m_{bb}, and the angular separation $\Delta R(b_1, b_2)$ between the two b-tagged jets representing the two Higgs boson decay products. When three b-jets are selected, however, the Higgs boson candidate is not uniquely determined and different pairing strategies can be followed. The two prominent strategies are pairing the two higher-p_T b-jets (*leading two b-jets strategy*), or the two spatially closest – in terms of $\Delta R(b_1, b_2)$ – b-jets (*closest two b-jets strategy*). Figure 6.8 shows the m_{bb} and $\Delta R(b_1, b_2)$ distributions for the two cases. All the distributions MC simulations

[3]This is different from the final results, where the major background processes: $t\bar{t}$, $Z + h.f.$ and $W + h.f.$, as well as the signal $VH(H \to b\bar{b})$ are parametrised as free-floating.

model data quite well, already at pre-fit stage. The m_{bb} distribution obtained with the *leading two b-jets strategy* shows a peak around the Higgs boson mass, from the reconstruction of the true b- and c-jets from the top quark decay, as explained above. On the other hand, in the *closest two b-jets strategy*, the peak at the Higgs boson mass values is complemented by a low-mass component in $t\bar{t}$ and V+*jets* events (the latter originating from gluon-splitting).

After analysing the main features of all the background processes in the 3b-CR, a direct comparison of $t\bar{t}$ events in the 3b-CR and in the 2b-tagged jet region (i.e.: the inclusive signal+control region) of the $VH(H \to b\bar{b})$ resolved analysis is given. In the resolved analysis[4], the Higgs boson candidate is reconstructed by pairing two true b-jets (one from the decay of each top quark in the pair) in $\geq 75\%$ of the cases, and less than 23% of the $t\bar{t}$ events are bc events. Conversely, in the 3b-CR, independent of the pairing strategy used for the reconstruction of the Higgs boson candidate (*leading two b-jets*, or *closest two b-jets*), more than 60% of $t\bar{t}$ events is bc, and less than 33% is bb, closer to the flavour composition of $t\bar{t}$ events in the $VH(H \to b\bar{b})$ boosted analysis.

Although the total number of $t\bar{t}$ events in the 3b-CR is roughly 10% of the number of $t\bar{t}$ events in the 2b-tagged jet region, the number of bc events in the two regions is comparable. Furthermore, in the 3b-CR the bc component is well separated from the bb and also other backgrounds, which is not the case in the 2b-tagged jet region. These features indicate that the 3b-CR could be the best option for the estimation of the $t\bar{t}$ background in the 0-lepton channel of the resolved analysis.

6.4.2 First studies for the addition of the 3b-CR

The introduction of the region has been studied in a standalone 0-lepton fit.

The 3b-CR is treated similarly to the other control regions in the $VH(H \to b\bar{b})$ resolved analysis, that are used only to estimate the normalisation of the background

[4]I.e. for events with 2b-tagged jets

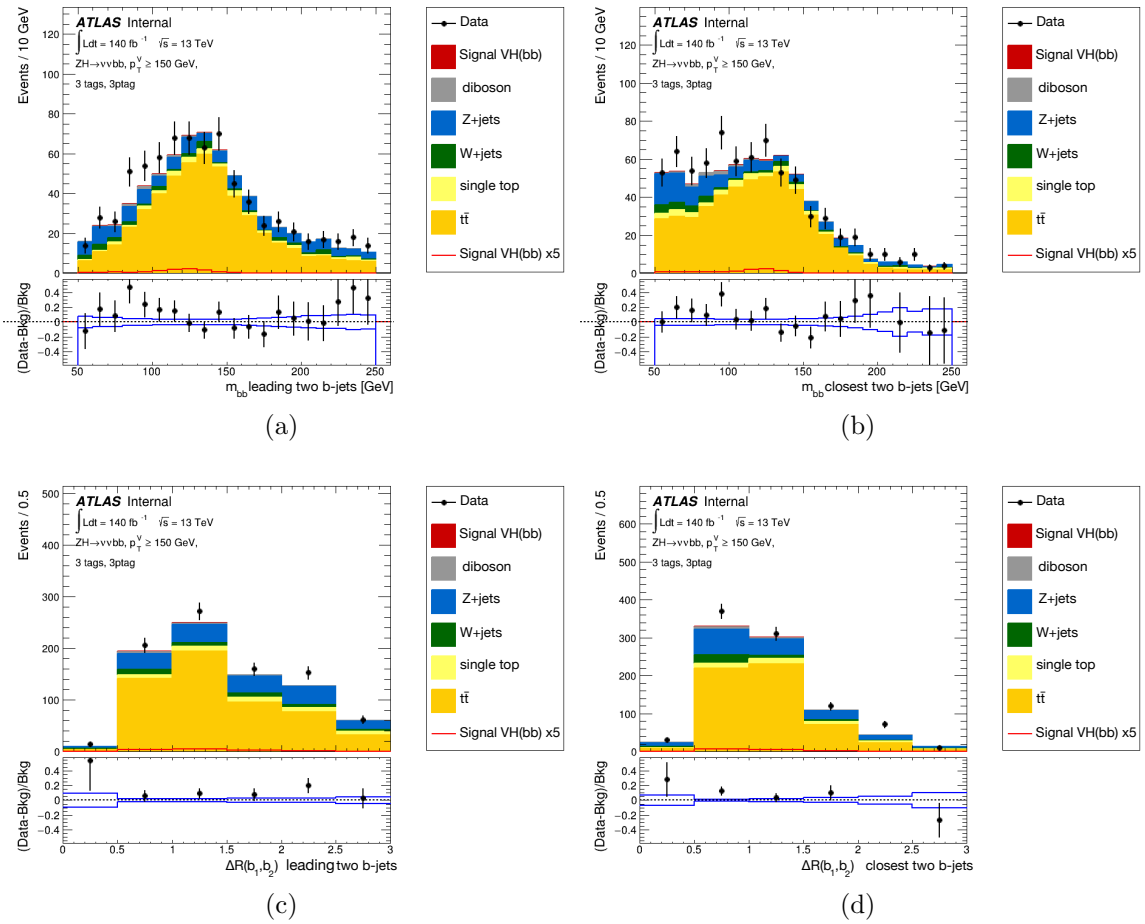


Figure 6.8: m_{bb} (6.8a, 6.8b) and $\Delta R(b_1, b_2)$ (6.8c, 6.8d) distribution in the 3b-CR in the 0-lepton channel of the $VH(H \to b\bar{b})$ resolved analysis, for 150 GeV $\leq p_T^V <$ 400 GeV for the *leading two b-jets* (6.8a, 6.8c) and *closest two b-jets* (6.8b, 6.8d) strategy. The stacked signal and background contributions are compared to data. In the lower panel, the ratio of the data minus the background component over background is shown. The blue bands represent the statistical uncertainty from the simulated background.

	Combination	Combination + 3b-CR
Significance	3.57	3.62 (+1.4%)

Table 6.4: Summary of the impact of the introduction of the 3b-CR in the resolved $VH(H \to b\bar{b})$ analysis in the 0-lepton channel. The fit significance in units of standard deviations is displayed.

processes ($t\bar{t}$, in this case). The 2b-tagged jet regions (signal region, Low-ΔR-CR, and High-ΔR-CR) are treated as in the fit of the $VH(H \to b\bar{b})$ resolved analysis described in Section 6.1.4, with the BDT score distribution used as variable of interest in the signal regions.

Events are split into the two 2- and 3-jet categories (only 3-jet for the 3b-CR), and two p_T^V categories: one for events with $p_T^V \in [150, 250[$ GeV and the other one for events with $p_T^V \in [250, 400[$. The selection $p_T^V(E_T^{miss}) < 400$ GeV has been added to these studies, to ensure compatibility with the combinations, where the *Hard cut @ 400 GeV* strategy has been used.

The results obtained by the fit with and without the inclusion of the 3b-CR are compared.

Table 6.4 summarises the significances, in units of standard deviations, for the resolved-only conditional fit to data ($\mu_{VH}^{b\bar{b}} = 1$) of the combination in the 0-lepton channel with and without the introduction of the 3b-CR. The significance gain provided by this additional region is $\sim 1.4\%$. The gain is rather small, being of the same order of magnitude as the impact of the uncertainties from the statistical fluctuations in the MC samples. However, the beneficial impact of this additional region on the analysis is also seen from other figures of merit.

For example, in the Asimov fit, it is interesting to study the nuisance parameter constraints and the mutual correlations between nuisance parameters. The Asimov fit is a likelihood fit in which an Asimov dataset is created, such that the estimators

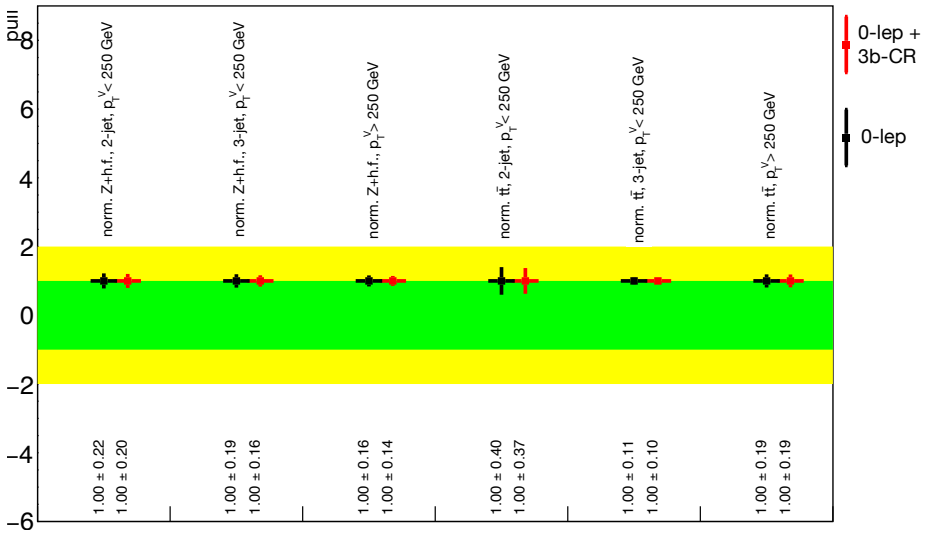
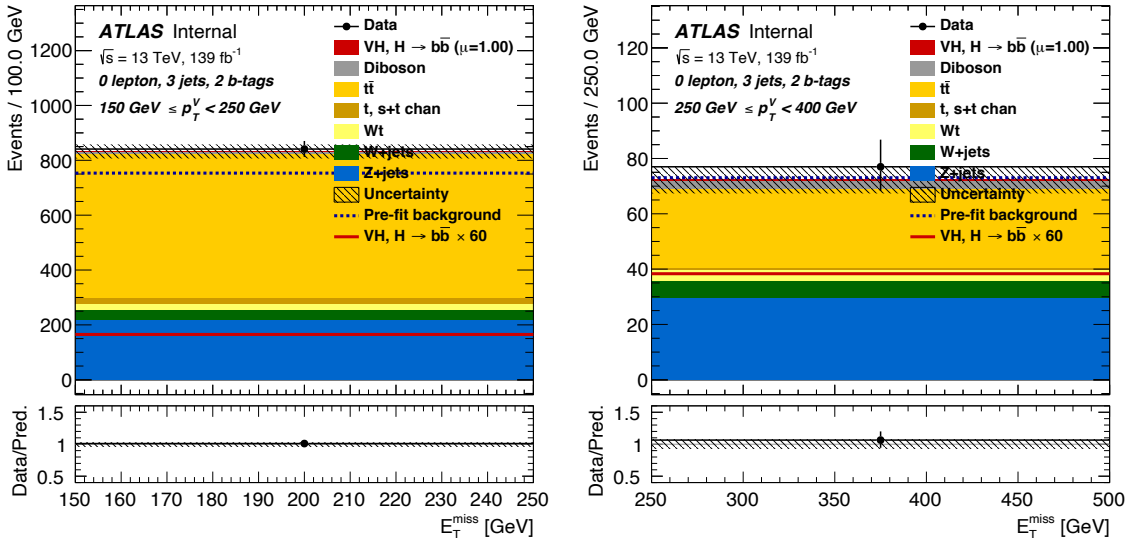


Figure 6.9: Comparison of the background floating normalisation nuisance parameters (NPs) for the 0-lepton Asimov fit with (red) and without (black) the introduction of the 3b-CR for the $VH(H \to b\bar{b})$ resolved analysis.

extracted by the fit correspond by construction to the simulated parameters, allowing correlations between nuisance parameters, as well as constraints, to be evaluated.

Figure 6.9 shows the direct comparison of the floating normalisation nuisance parameters (NPs) for the 0-lepton Asimov fit with and without the introduction of the 3b-CR. It can be clearly seen that the introduction of the 3b-CR constrains the floating normalisation NPs for $Z + h.f.$ and $t\bar{t}$, reducing their error bands by up to 13%. A similar effect is also seen on the top-related NPs. This confirms the importance of the role of the 3b-CR in controlling the background normalisations, and the $t\bar{t}$ background. Additionally, it can be shown that, although the correlations increase among the background floating normalisation NPs, as a result of the fact that the normalisations are constrained using the same region (the 3b-CR), they decrease with respect to the other NPs, as a result of the increase in the statistical power of the fit in constraining the NPs, thanks to the additional region.

Figure 6.10 shows the conditional post-fit output distributions in the new 3b-CR.



(a) (b)

Figure 6.10: Conditional post-fit output distributions of E_T^{miss} (p_T^V) in the new 3b-CR, in the region for 150 GeV $\leq p_T^V <$ 250 GeV (6.10a) and for 250 GeV $\leq p_T^V <$ 400 GeV (6.10b). The fit model is a standalone 0-lepton fit in the $VH(H \to b\bar{b})$ resolved spectrum.

It can be seen that the data is very well reproduced by the MC simulations. In the case of the lower-p_T^V region (Figure 6.10a) the original MC prediction, represented by the dashed dark blue line, is significantly moved as a result of the fit, to match the data, thanks to the introduction of the 3b-CR.

The results are very encouraging, since the 3b-CR is shown to improve the analysis sensitivity, help constrain several important analysis NPs and decrease their correlations. The study shown above represents the first step towards the introduction of this region in the legacy analysis.

An interesting future development would be to study the introduction of the 3b-CR not only as a normalisation, but also fitting the shapes of kinematic distributions. Additionally, the region could be used to constrain the different $t\bar{t}$ components in the analysis, bb and the very signal-like bc. The variable of interest for the resolved analysis is the BDT output score, therefore a training of the BDT in the dedicated

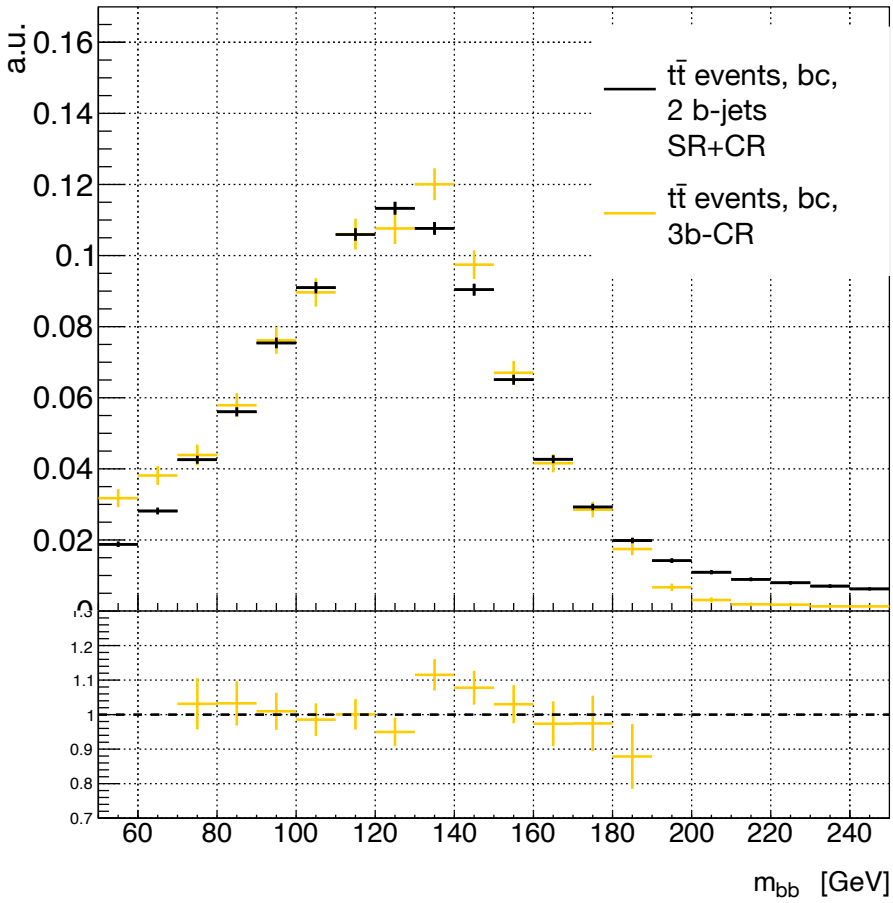


Figure 6.11: m_{bb} distributions of $t\bar{t}$ events for the bc component in the 2 b-jet region (black) and in the 3b-CR region using the *closest two b-jets strategy* (orange). The distributions are obtained in the 0-lepton channel of the $VH(H \to b\bar{b})$ resolved analysis.

3b-CR would be necessary. Under the assumption that m_{bb} is one of the variables with the highest discriminating power for the BDT, one could assume that if the m_{bb} shapes could be replicated in the 3b-CR, so would the BDT score distributions. A very preliminary study on the m_{bb} (see Figure 6.11) shows that this can be achieved, for the bc component using the *closest two b-jets strategy*.

Finally, these studies can be extended to the 1-lepton channel, where the 3b-CR could be even more effective than the High-ΔR-CR in the estimation of the $t\bar{t}$ background, especially for the bc component. After these elements will be added, a final decision on the introduction of the 3b-CR in the legacy analysis will be made.

Chapter 7

Conclusions

The discovery of the Higgs boson represents an outstanding success of particle physics, and opens a window for testing the Standard Model and its possible extensions. So far, all the measurements performed during the LHC Run 1 and Run 2 by the ATLAS and CMS collaborations are found to be compatible with the Standard Model predictions.

Among these measurements, the study of the Higgs boson decay to beauty quarks plays an important role. This decay has the highest branching ratio in the SM, and represents the only process for which it is possible to measure the Yukawa coupling to down-type quarks up to now. The channel is sensitive to many potential Beyond the SM (BSM) phenomena, especially at high-p_T regimes.

The $VH(H \rightarrow b\bar{b})$ boosted analysis described in this thesis is the first analysis probing this process at high p_T with the ATLAS experiment. The analysis is performed on the full LHC Run 2 dataset from pp collisions at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 139 fb^{-1}. Many studies were made in order to understand the topology and kinematic properties of signal and background events in this new phase space.

The analysis has measured the $VZ(Z \rightarrow b\bar{b})$ and $VH(H \rightarrow b\bar{b})$ signal strengths:

$$\mu_{VH}^{b\bar{b}} = 0.72^{+0.39}_{-0.36} = 0.72^{+0.29}_{-0.28}\,(\text{stat.})^{+0.26}_{-0.22}\,(\text{syst.}) \quad ,$$

$$\mu_{VZ}^{b\bar{b}} = 0.91^{+0.29}_{-0.23} = 0.91 \pm 0.15\,(\text{stat.})^{+0.25}_{-0.17}\,(\text{syst.}) \quad .$$

corresponding to an observed (expected) significance $Z_{VH}^{b\bar{b}} = 2.1(2.7)$ and $Z_{VZ}^{b\bar{b}} = 5.4(5.7)$ standard deviations, respectively. The total uncertainty is of $\sim 54\%$ and 32% on $\mu_{VH}^{b\bar{b}}$ and $\mu_{VZ}^{b\bar{b}}$ respectively, with the statistical uncertainty accounting for $\sim 60\%$ of the total. Both measurements are in agreement with the SM expectations.

Cross-sections of the associated production of a Higgs boson decaying into beauty quarks and a vector boson have been measured as a function of the vector-boson transverse momentum in four bins using the STXS framework. The total uncertainties on each bin ranged from 70% to 90%, depending on the bin. An effective Lagrangian, sensitive to dimension-six operators that could potentially modify the Higgs boson couplings to the electroweak gauge bosons, has been constrained.

The limited data statistics is the main challenge of the analysis. Therefore it will be interesting to perform the study when more data will be available in the future. Further possible developments of the analysis could result from the introduction of more advanced techniques in the reconstruction of the events, and in the signal-to-background discrimination.

Finally, it will be interesting to perform a single measurement of the $VH(H \to b\bar{b})$ process in the ATLAS experiment, using different techniques depending on the energy regime and thus combining the Run 2 $VH(H \to b\bar{b})$ boosted analysis with the $VH(H \to b\bar{b})$ analysis at lower p_T regimes (*resolved analysis*). For the combined analysis, a key point is the treatment of the overlap between the boosted and resolved analyses and thus the combination strategy to be adopted. Several options have been tested in the 0-lepton channel. The study of additional regions, such as the $t\bar{t}$-enriched control region with three b-tagged jets, is important to perfect the understanding of

the analysis event topology and phase space.

The successful Run 2 data collection period allowed the experiments at the LHC to improve the methodologies for the study of the Higgs boson properites. The start of the LHC Run 3 in 2022, doubling the data statistics of Run 2 at $\sqrt{s} = 13$ TeV, and possibly even pushing the energy frontier to $\sqrt{s} = 14$ TeV, will provide an exceptional opportunity to improve our current knowledge of the SM and the Higgs sector, as well as search for new phenomena.

Many projections on Run 3 and even at the HL-LHC predict very high precisions Higgs boson measurements, that will be dominated by theoretical uncertainties [179, 180]. Among these measurements, the studies performed by extrapolating the $VH(H \to b\bar{b})$ analysis to the 3000 fb^{-1} delivered by the HL-LHC predict a precision of a few % on the signal strength parameter and of $\sim 5\%$ on the coupling of the Higgs boson to b-quarks.

Appendix A

Future silicon pixel detectors: the ATLAS upgrade and beyond

This appendix briefly describes my contribution to the ITk pixel detector and the development of depleted monolithic active pixels for future applications.

A.1 Introduction on silicon detectors

The ATLAS Pixel Detector and SemiConductor Tracker (SCT), introduced in Section 2.2.3, are based on silicon technology. Silicon tracking detectors are widely used in particle physics, and represent the future of tracking detectors, given their high granularity and precision, and radiation hardness [181].

The detection of minimum ionising particles (MIPs) is based on ionisation or excitation of atoms in the medium caused by the passage of charged particles [182, 183]. For semiconductors the excitation occurs when an electron acquires enough energy to move from the *valence* band to the *conduction* band. This creates a net negative charge in the conduction band and a net positive charge in the valence band, called a *hole*. In silicon detectors, the energy required to create an electron-hole (e-h) pair is 3.6 eV, yielding about 80 e-h/μm.

In intrinsic silicon the number of thermal free charge carriers is $\sim$ five orders of magnitudes higher than those induced by a MIP traversing the average thickness of a silicon detector (a few hundreds of μm), meaning that a MIP signal would be lost.

This is overcome by *doping* the semiconductor with atoms of other (*dopant*) materials, since the addition of impurities can change the amount of charge carriers. A silicon crystal can be doped with *donor* atoms (elements of group V) or *acceptor* atoms (elements of group III), making the silicon n-type or p-type, respectively. The fundamental component of a silicon particle sensor is the p-n junction, which consists of n-type silicon in contact with p-type silicon. When n-type and p-type materials are put in contact, the different concentrations of majority carriers lead to a diffusion current, removing all free carriers from the junction region. This area is referred to as the *depletion region*, filled with ionised atoms of the dopant materials. The fixed charge of the ionised dopants generate a built-in potential V_{bi}, corresponding to an electric field $\vec{E}_{bi}$ going from the n-type material to the p-type material. A *reverse bias* voltage V can be applied to the junction in the same direction of V_{bi} making the depletion region larger than the one at equilibrium. In this region, where no free charge carries are found, the e-h pairs produced by an incoming MIP can be easily detected.

Position information can be achieved by finely segmenting the p-n junctions into strips, with a mono-dimensional resolution, or pixels, where the resolution is bi-dimensional. Pixels are typically used in the innermost layers of particle physics detectors, and are crucial for the reconstruction of primary and secondary vertices.

Pixel detectors can be of two main types: *hybrid* silicon pixels, for which the sensitive silicon bulk is *bump-bonded* to a front-end read-out chip, and *monolithic* silicon pixels, where the silicon bulk of the front-end chip is used as sensor and the bump-bonding is not required. All systems presently used at LHC are based on hybrid pixel detectors, but monolithic options are being studied and developed for the Phase

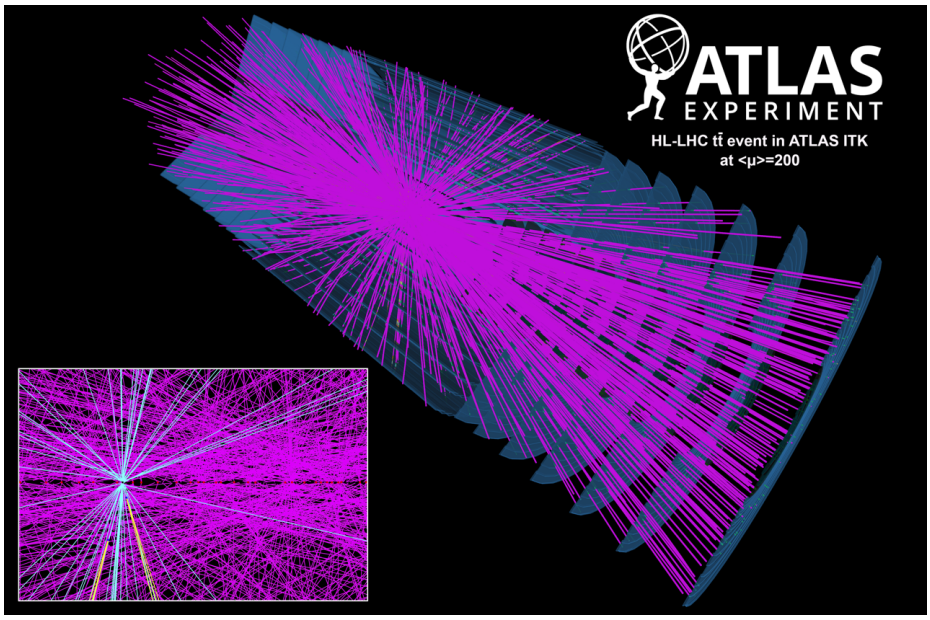


Figure A.1: Simulated event display of a $t\bar{t}$ event in the ATLAS ITk at the HL-LHC. The average number of inelastic interactions per bunch crossing is set to $\langle\mu\rangle = 200$. Image source: [184].

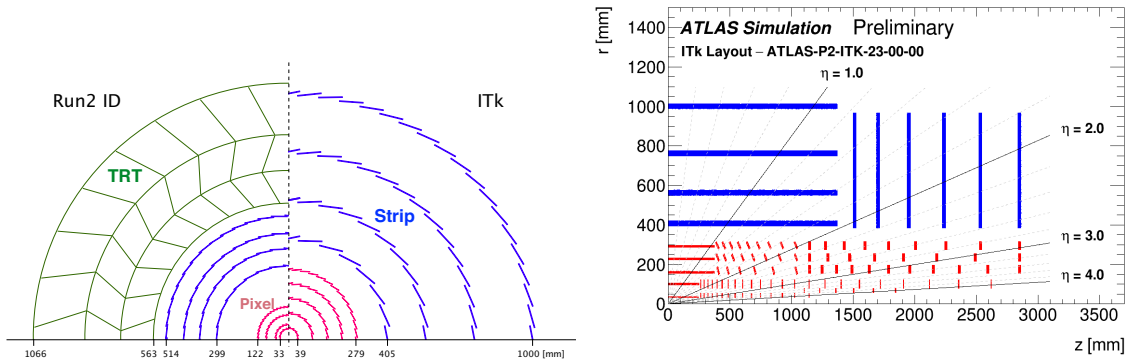
II upgrades of the LHC experiments[1] and new experiments at future colliders.

A.2 The ITk upgrade

In Section 2.1.2 the schedule of the LHC has been briefly discussed. After Run 3, starting in 2022 and expected to extend over three years, the LHC will undergo an upgrade, denoted as the High Luminosity LHC or HL-LHC, that will accumulate up to 3000 fb^{-1}, ten times the data collected at the LHC. This will be achieved thanks to an increase in instantaneous luminosity to about 7.5×10^{34} cm^{-2}s^{-1}, with ~ 200 inelastic pp collisions per beam crossing [185], as shown in Figure A.1.

In order to cope with the increase in luminosity, the current ATLAS ID will be completely replaced after Run 3, and a new tracker will be installed: the Inner Tracker, or ITk, as sketched in Figure A.2a.

[1]The new tracking detector of ALICE, which is currently under installation, is based on monolithic pixels.



(a) (b)

Figure A.2: The radial layouts of the current ATLAS tracker and the ITk upgrade (A.2a), and the schematic depiction of the ITk Layout, including the updated layout for the Pixel detector. The active elements of the strip detector are shown in blue, and those of the pixel detector are shown in red (A.2b). Image sources: [187, 188].

The design goal for the ITk is to maintain the tracking performance of the current ATLAS ID, while coping with the increase in instantaneous luminosity and in integrated radiation dose in the HL-LHC environment.

The ITk will be an all-silicon tracker, divided into two sub-subsystems, a Pixel Detector in the innermost part [5] and a Strip Detector in the outermost [186], with extended η coverage to $|\eta| \leq 4$, as shown in Figure A.2b.

For the Pixel Detector, *hybrid pixel modules* have been chosen. The *bare module* comprises the passive high resistivity silicon sensor and the front-end read-out chip fabricated in CMOS technology. The silicon sensor and front-end read-out chip are joined using a high-density connection technique called "flip-chip bump-bonding". A schematic of the elements of a bare pixel module is presented in Figure A.3.

In the outer barrel layers and the outer end-cap rings of the Pixel detector of the ITk, *quad modules* will be used. They consist of four chips bump-bonded to a single sensor, $\sim 4 \times 4$ cm^2 in area. The electrical connections to the modules are routed to the active elements via a flexible printed circuit (module flex) which is glued to the backside of the sensor.

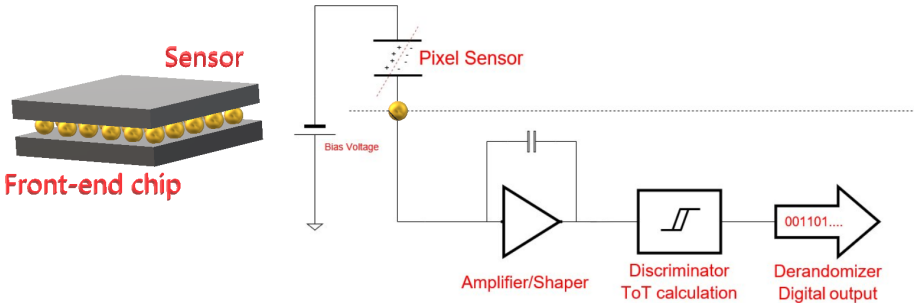


Figure A.3: Schematic view of am ITk bare pixel module. Image source: [5].

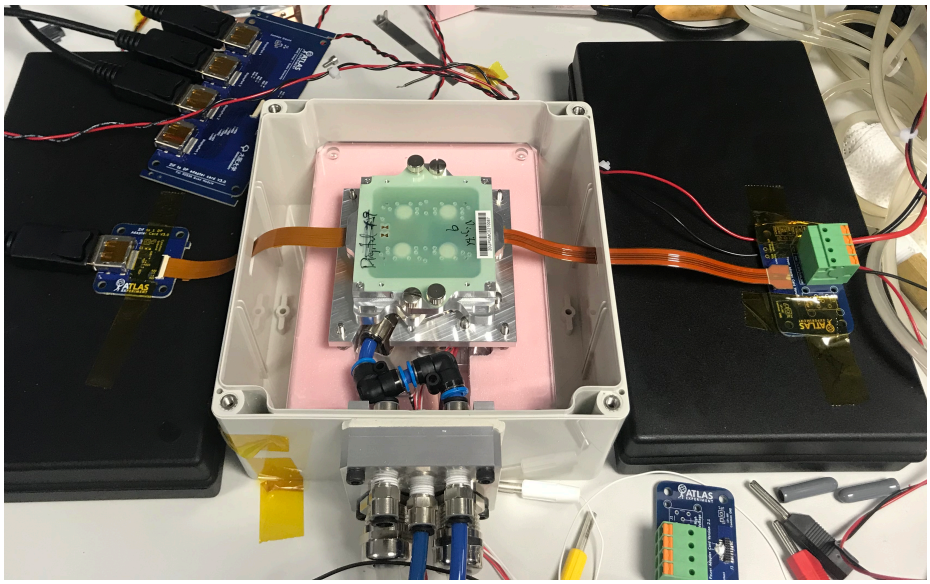


Figure A.4: Experimental setup for the ITk testing qualification in Oxford. November 2020, photo credits: MGF.

Testing of the ITk pixel digital quad modules

Extensive R&D and *testing* is being carried out by the ATLAS institutions taking part in the ITk, before starting module production. Testing assesses the quality of the modules that will be installed in the detector, and institutions are asked to *qualify* as testing sites, to guarantee that it is performed correctly.

Figure A.4 shows the experimental setup used at Oxford for the testing qualification of the ITk pixel quads. At this stage the focus is on the setup and the electrical testing, therefore only the readout chip of the quad module is tested, without the

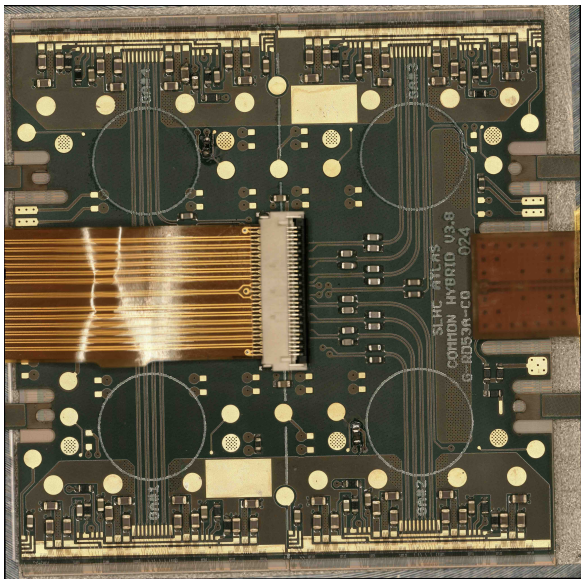


Figure A.5: 50× magnification stitched image of the module, produced with a Keyence VHX5000 digital microscope.

sensor. The qualification is considered completed if a series of set requirements is met. The three most important ones are explained in the following.

The first requirement is to perform an accurate visual inspection of the digital quad in use. At this stage institutions are required to demonstrate their ability to check (and document) for potential damage or anomalies in wirebonds and the electrical components of the flex. A 50× magnified picture of the digital quad used in Oxford is shown in Figure A.5.

The second requirement is to prove the ability to measure the module temperature via a Negative Temperature Coefficient (NTC) sensor [189] and to keep the temperature of the sensor below 55°C. This is achieved using a water chiller in combination with a Peltier cooling element, and is recorded in Figure A.6.

The final requirement consists in demonstrating the ability to complete a *tuning procedure* of the four read-out chips in the digital quad. First, the chip is set in a specific configuration, checked through dedicated scans verifying the writing and reading of global and pixel registers, followed by a checkout of the default configuration and chip trimming. The tuning consists of a series of scans that test both the digital and

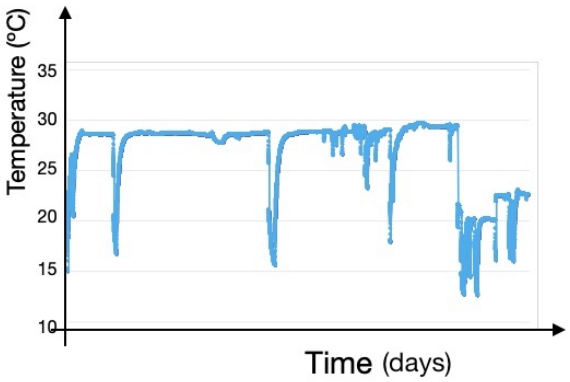


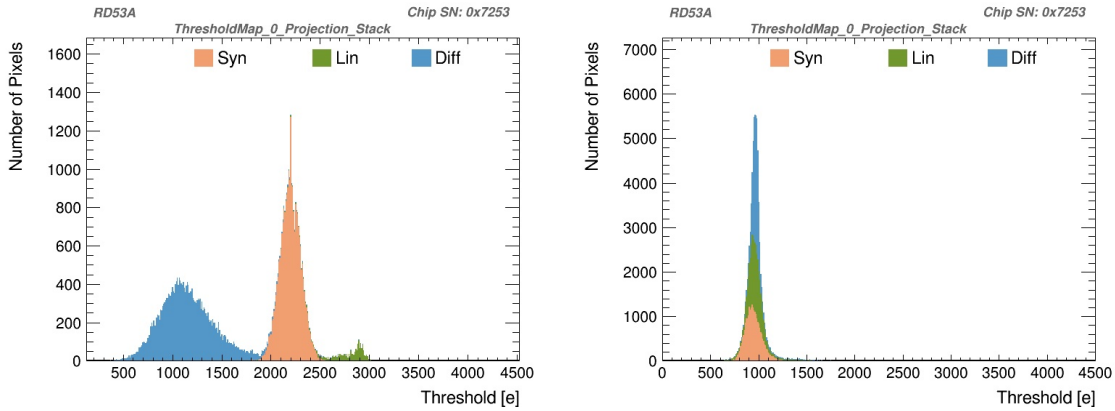
Figure A.6: Temperature of the ITk pixel digital quad during testing in Oxford. The dips in the distributions correspond to the times when the cooling system was ON, but the digital quad was still OFF. As soon as the digital quad was switched ON, the temperature began to rise, to plateau at an equilibrium temperature, always below 30° C.

analogue parts of the chip. An additional set of scans aims at tuning the threshold and time over threshold (ToT) of each pixel. The read-out chip in the current version of the digital quads is a prototype (RD-53, [7]), where three different types of amplification logic are implemented. Therefore the tuning procedure is performed on all three of them, as pictured in Figure A.7.

As it can be seen on Figure A.7a, before the tuning procedure is performed the pixel thresholds are distributed along different values and their spread is significant. After the tuning is performed, the pixel thresholds of the three front-ends are all aligned around the desired value with a significantly lower spread with respect to the pre-tuning configuration, as pictured in Figure A.7b.

A.3 The MALTA project

The Monolithic from ALice To Atlas (MALTA) project [4, 190] focuses on the study of Depleted Monolithic Active Pixel Sensors (DMAPS) produced in CMOS technology and was initially explored as an option for the outermost pixel layer of the ITk. The monolithic solution is discarded for the ITk, however R&D work is continuing, given



(a) (b)

Figure A.7: Threshold distributions for the synchronous (orange), linear (green) and differential (blue) front-end amplifiers for one of the chips on the RD-53 digital quad tested in Oxford. The distributions are shown before (A.7a) and after (A.7b) the tuning procedure.

the strengths of the MALTA pixel sensor, such as the high granularity, fast response and competitive price.

The MALTA sensor, pictured in Figure A.8, has a total area of about 2×2 cm^2, and is a matrix with 512×512 square pixels of 36.4 μm pitch. The chip operates in an asynchronous readout architecture.

Characterisation of the response of MALTA chips

The MALTA chip R&D is ongoing. Given the global scale of the MALTA effort, it is important to define a quality assurance (QA) procedure to assess the quality of the response of each chip. This series of tests is designed to be performed in one working day, to verify the correct functioning of each new chip, before they are used.

The procedure includes several steps, that will be briefly described in the following.

Powering

This step ensures that the chip switches ON and OFF correctly, and that draws the correct current (< 1 mA) when powered at 6 V.



Figure A.8: Picture of a MALTA chip. The pixel matrix is the 2×2 cm^2 squared sensor at the top of the board, protected with the yellow tape. August 2020, photo credits: MGF.

Read-out channels

Once the powering has been checked, it is necessary to verify that there is no malfunctioning in the transmission of the chip output. The output from each pixel on the MALTA sensors is sent out on 37 *channels* (or *bits*), identifying the pixel position in the 512×512 pixel matrix. Thanks to the electronics on the MALTA board, it is possible to *pulse* a specific pixel for a certain time interval Δt and read its output. The bits corresponding to the pulsed pixel are expected to switch ON and send a Δt-long signal. Figure A.9 reports the results of the read-out tests on two different MALTA chips: a perfectly working one, pictured in Figure A.9a and A.9c, and one with a problem in the read-out electronics, shown in Figure A.9b and A.9d. As it can be seen from Figures A.9c and A.9d, when using a spherical Sr^{90} β^- radioactive source, the functioning chip returns a clear silhouette of the source, as opposed to the problematic chip, that distorts it.

Pixel threshold and noise measurement

After ensuring the correct functioning of the read-out system of the chip, the procedure involves a threshold scan on a region of MALTA of $\sim$ 1000 pixels. Ideally, the full chip should be scanned, but the procedure would be impractically long. The scan is performed on four different configurations, chosen in order to cover the full threshold configuration space. In the threshold scan, both the thresholds and noise values for the pixels are measured. The typical RMS of the threshold is expected to be less than 10% of the mean value of the threshold distribution, however measurements indicate that this value is higher and $\sim$ 15%. The noise distribution has a long positive tail. The pixel threshold and noise distributions for the chosen configurations are shown in Figure A.10a

Chip noise rate measurement

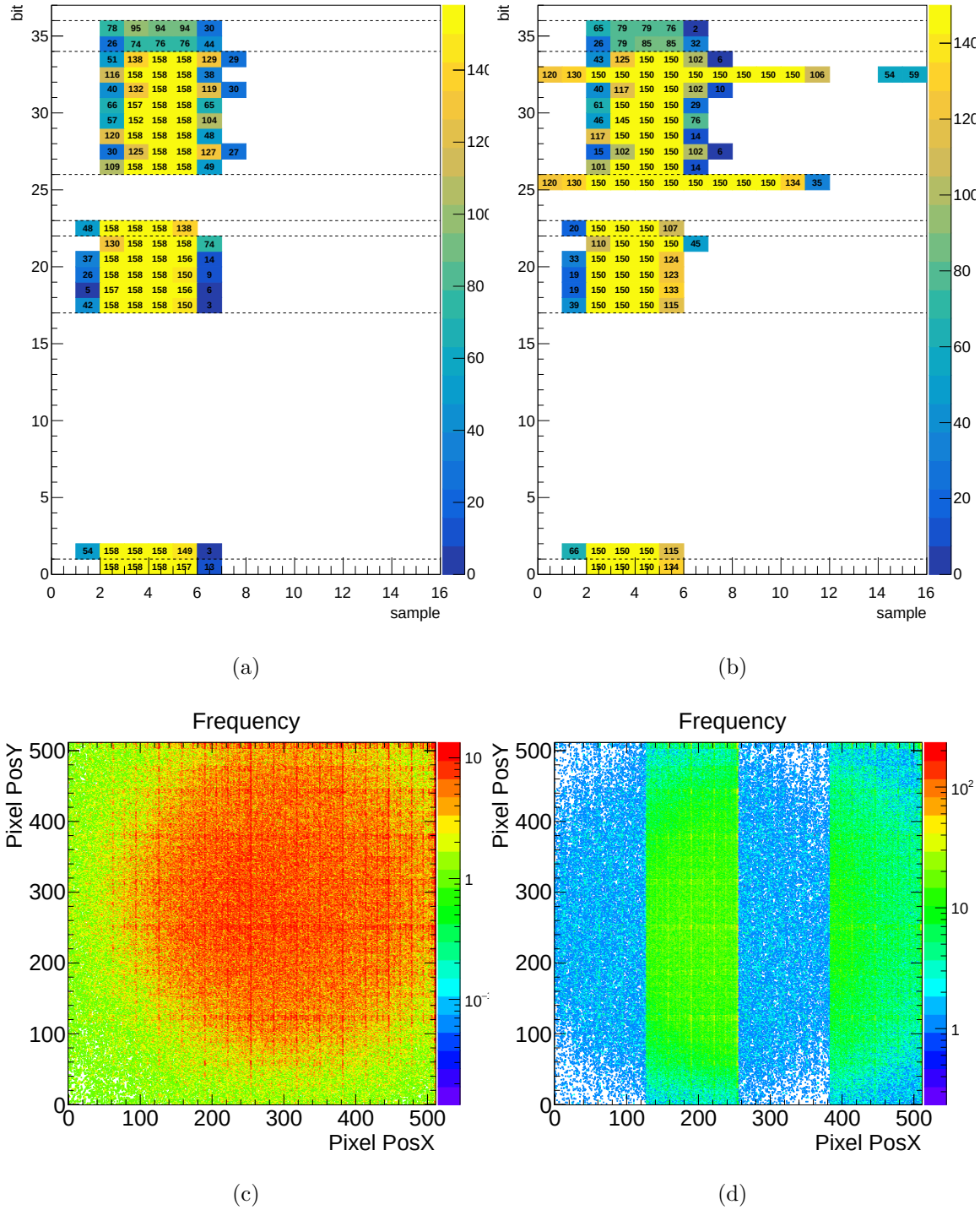
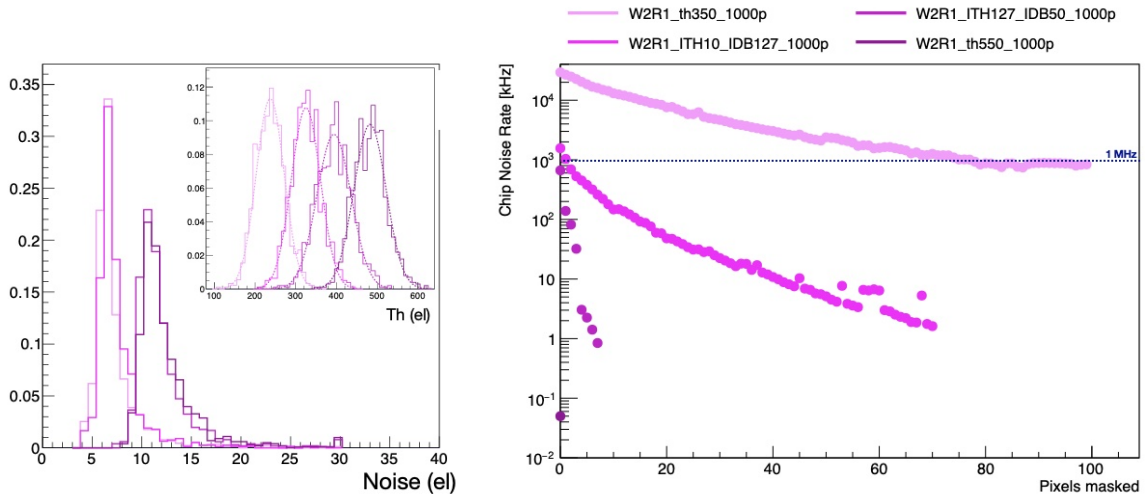


Figure A.9: Results of the read-out test on two MALTA chips. On the left, a perfectly functioning chip is shown (A.9a, A.9c), while the chip presented on the right has a read-out problem (A.9b, A.9d). In the top line the output of the 37 output bits are shown, with respect to the reference bit (bit 0). The reference bit represents the pulsing signal sent to the pixels. As it can be seen for the chip on the right, two bits (n. 25 and 32) are "stuck", i.e. always switched ON and therefore cannot replicate the input signal. The effect of this defect can be seen by comparing the two pictures on the bottom line. Figures A.9c and A.9d show two source scans performed using a Sr^{90} β^- source. The functioning chip returns a clear silhouette of the spherical source, as opposed to the problematic chip, that distorts it.



(a) (b)

Figure A.10: Results of the threshold and noise tests on MALTA. Figure A.10a shows the noise and threshold (Th) distributions for $\sim$ 1000 pixels on a MALTA chip in the four chosen configurations. As it can be seen, while the threshold values cover the full threshold spectrum, the noise distributions are grouped in pairs. Figure A.10b shows the total chip noise as a function of the number of masked pixels in the chip. Pixels are masked in decreasing order of noise.

When chips are operated, a fraction of the chip pixels may be too noisy and therefore affect the chip performance, since the signal-over-background ratio is reduced in a noisy device. The final step of the QA procedure is to identify and run an algorithm to measure the chip noise and reduce it up to a certain acceptable rate, $Rate_{max} = 1$ kHz. The algorithm, whose flow chart is depicted in Figure A.11, masks the noisiest pixel in the chip and is run iteratively until either $Rate_{max}$ is reached, or a set number of pixels, $N_{max} = 100$, is masked.

Figure A.10b shows the chip noise rate as a function of the number of masked pixels in the chip, for the four chosen configurations. It can be seen that the chip noise highly depends on the configuration, with the noisiest configurations corresponding to the lower thresholds.

The first formal MALTA QA procedure above has proven extremely useful in the characterisation of MALTA chips. It is now regularly used, and the output documen-

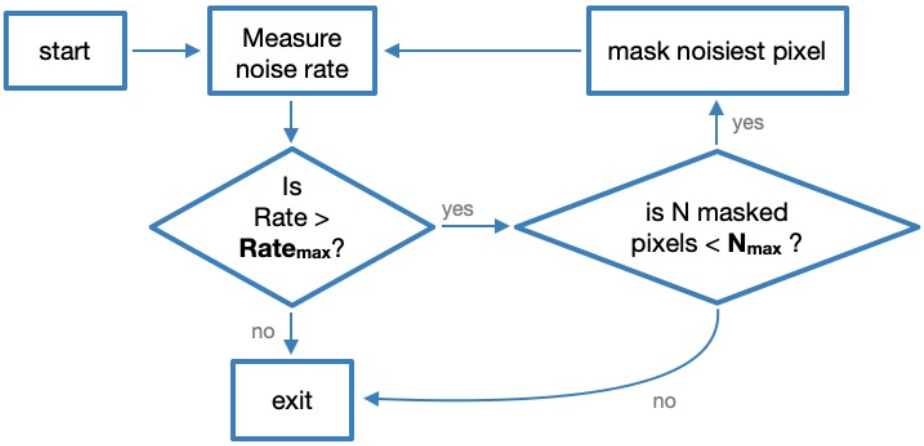


Figure A.11: Flow chart of the iterative masking procedure for a MALTA chip. Parameters $Rate_{max}$ and N_{max} are set to 1 kHz and 100 respectively.

tation is sent to collaborators using MALTA chips.

Appendix B

Pull Plots for the $VH(H \to b\bar{b})$ boosted analysis

A pull on a nuisance paramenter (NP) θ_i is the difference between the estimated $\hat{\theta}_i$ from the fit and its nominal value θ_i^0 (pull $= (\hat{\theta}_i - \theta_i^0)/\Delta\theta_i^0$). Pulls are expressed in units of standard deviations.

The pull plots shown in this appendix are obtained from the unconditional fit (see Section 5.5) to the Asimov dataset (red dots) and to the actual dataset (black dots).

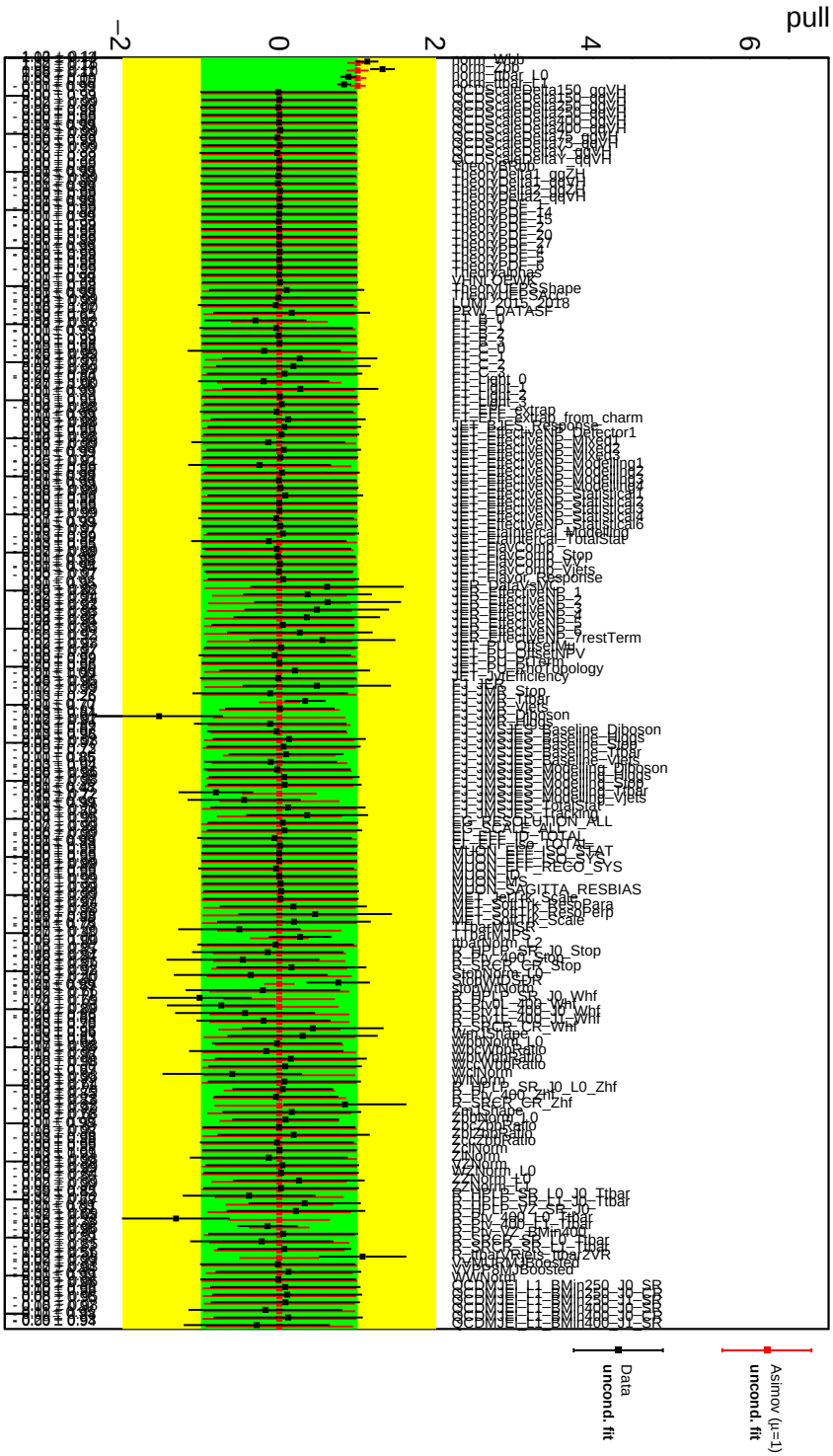
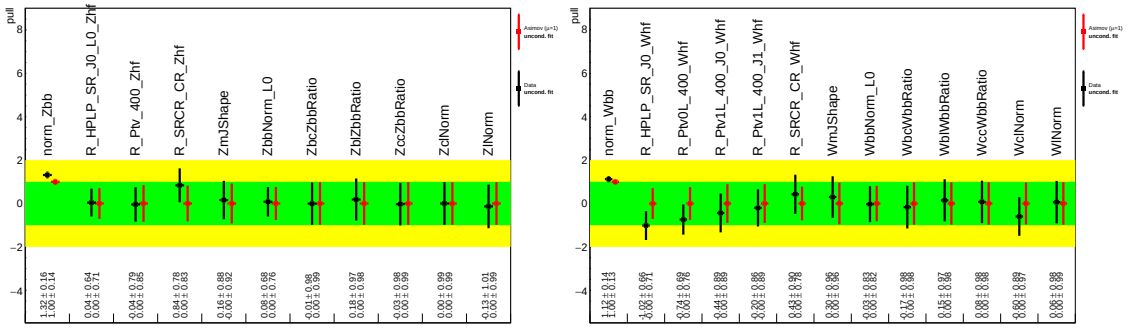
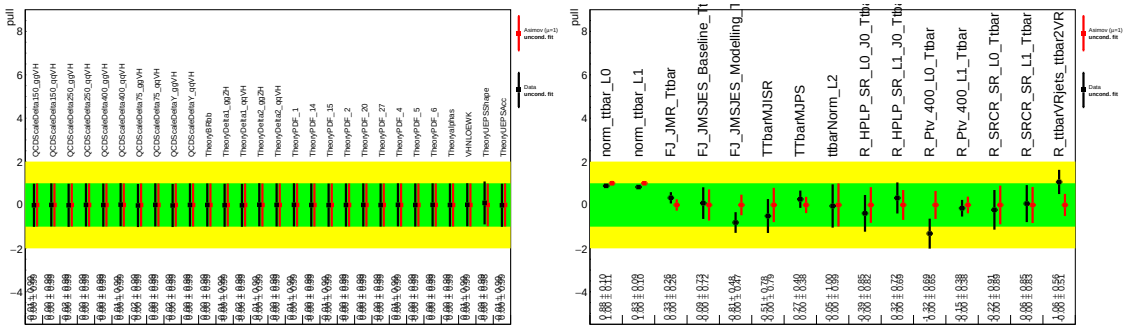


Figure B.1: Nuisance parameter pulls, constraints and scale factors from free floating parameters obtained from the combined unconditional fit in all three lepton channels to the Asimov- (red) and the actual dataset (black).



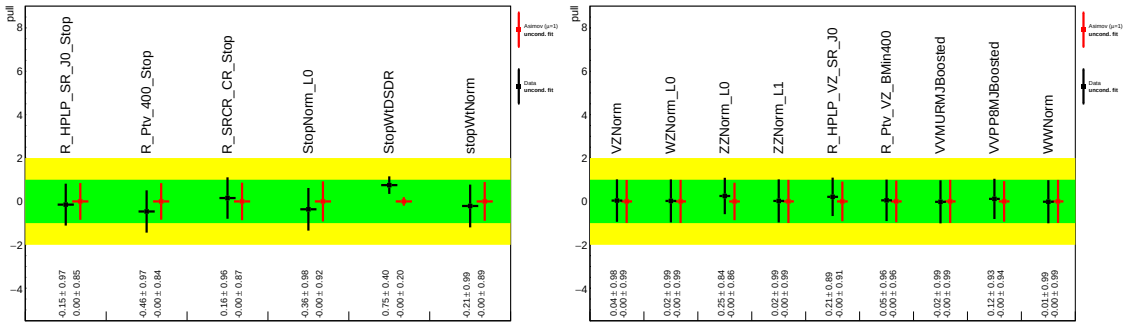
(a) Z+*jets* modelling

(b) W+*jets* modelling



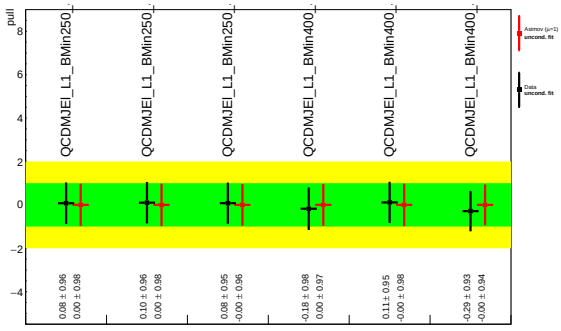
(c) VH signal modelling

(d) $t\bar{t}$ modelling



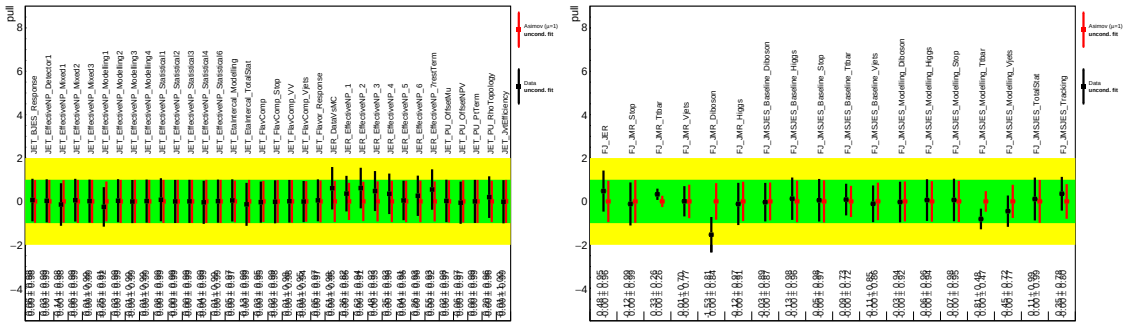
(e) single top modelling

(f) diboson modelling



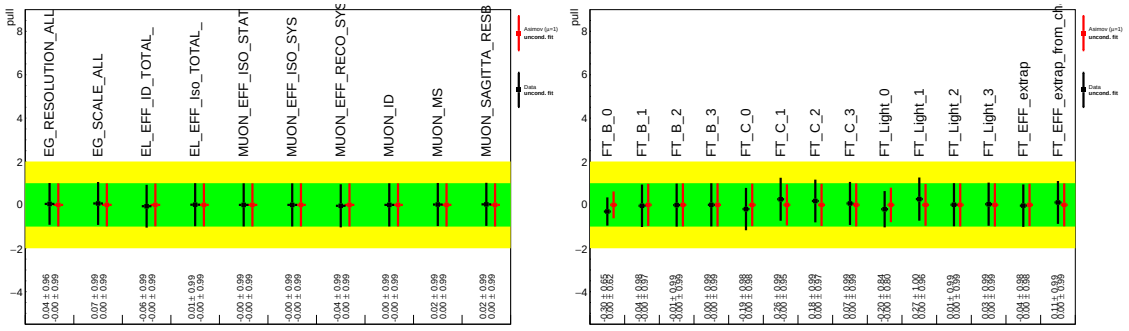
(g) QCD multijet

Figure B.2: Monte Carlo modelling-related and QCD multijet nuisance parameter pulls, constraints and post-fit normalisation scale factors for the freely floating parameters, obtained from the combined unconditional fit in all three lepton channels to the Asimov- (red) and the actual dataset (black).



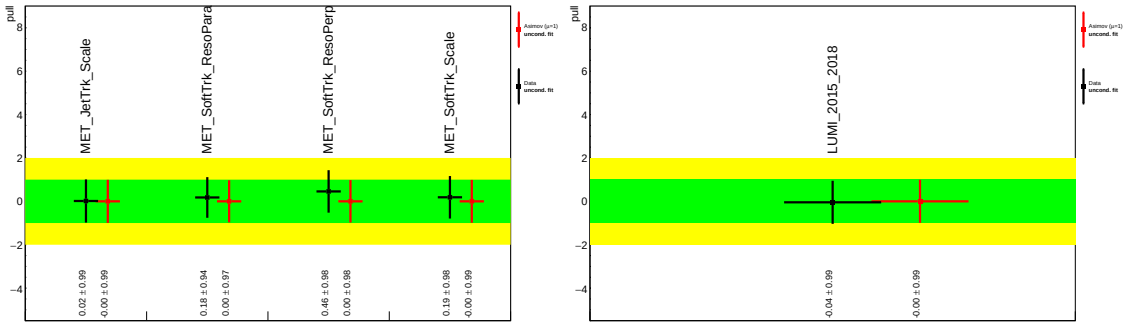
(a) Small-R Jets

(b) Large-R Jets



(c) Leptons

(d) b-tagging



(e) Missing transverse energy

(f) Luminosity

Figure B.3: Detector-related nuisance parameter pulls and constraints, obtained from the combined unconditional fit in all three lepton channels to the Asimov- (red) and the actual dataset (black).

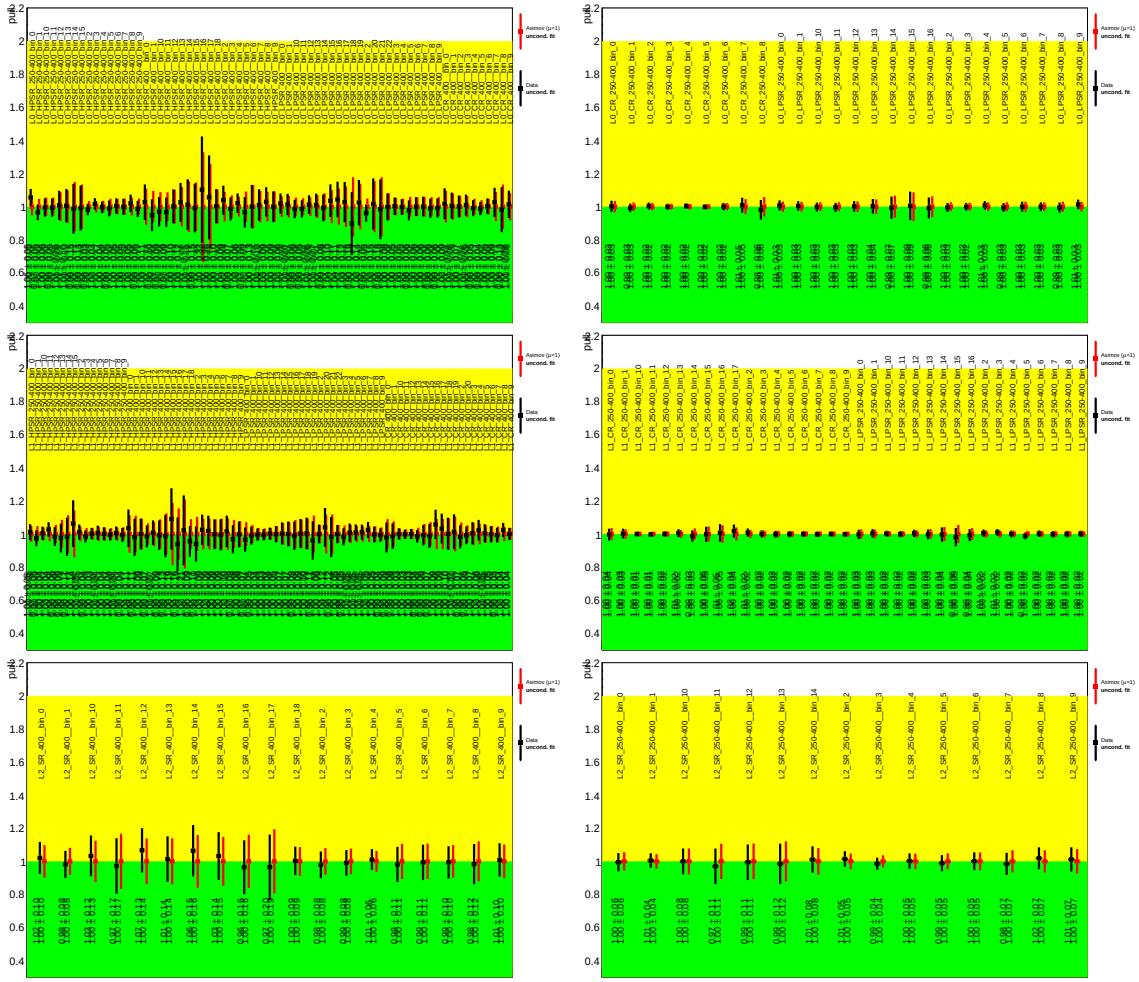


Figure B.4: Simulation statistics ("gamma") nuisance parameter pulls, obtained from the combined unconditional fit in all three lepton channels to the Asimov- (red) and the actual dataset (black). The top row shows the 0-lepton channel, the middle row the 1-lepton channel and the 2-lepton channel is depicted on the bottom row.

Bibliography

[1] ATLAS Collaboration, "ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J.*, vol. C79, no. 11, p. 970, 2019.

[2] ATLAS Collaboration, "Measurement of the associated production of a Higgs boson decaying into b-quarks with a vector boson at high transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Lett. B*, vol. 816, p. 136204, 2021.

[3] ATLAS Collaboration, "Measurements of WH and ZH production in the $H \to b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector", *Eur. Phys. J. C*, vol. 81, no. 2, p. 178, 2021.

[4] A. Sharma *et al.*, "The MALTA CMOS pixel detector prototype for the ATLAS Pixel ITk", *PoS*, vol. VERTEX2018, p. 014, 2019.

[5] ATLAS Collaboration, "Technical Design Report for the ATLAS Inner Tracker Pixel Detector", Tech. Rep. CERN-LHCC-2017-021, ATLAS-TDR-030, CERN, Geneva, Sep 2017. Available at: https://cds.cern.ch/record/2285585.

[6] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 1.* Geneva: CERN, 1997.

[7] RD-53 Collaboration, "RD-53 Collaboration Home", Available at: https://rd53.web.cern.ch.

[8] M. Thomson, *Modern Particle Physics*. Cambridge University Press, 2013.

[9] D. H. Perkins, *Introduction to high energy physics; 4th ed.* Cambridge: Cambridge Univ. Press, 2000.

[10] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B*, vol. 716, pp. 1–29, 2012.

[11] CMS Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC", *Phys. Lett. B*, vol. 716, pp. 30–61, 2012.

[12] C. Quigg, *Gauge Theories of the Strong, Weak and Electromagnetic Interactions*, vol. 56. 1983.

[13] A. Purcell, "Go on a particle quest at the first CERN webfest. Le premier webfest du CERN se lance à la conquête des particules", p. 10, Aug 2012. Available at: https://cds.cern.ch/record/1473657.

[14] J. Ellis, "The discovery of the gluon", *International Journal of Modern Physics A*, vol. 29, p. 1430072, Dec 2014.

[15] D. J. Gross, "The discovery of asymptotic freedom and the emergence of QCD", *Proc. Nat. Acad. Sci.*, vol. 102, pp. 9099–9108, 2005.

[16] M. Gell-Mann, "A Schematic Model of Baryons and Mesons", *Phys. Lett.*, vol. 8, pp. 214–215, 1964.

[17] G. Zweig, "An SU_3 model for strong interaction symmetry and its breaking; Version 2", p. 80 p, Feb 1964. Available at: https://cds.cern.ch/record/570209.

[18] L. Di Lella and C. Rubbia, "The Discovery of the W and Z Particles", *Adv. Ser. Dir. High Energy Phys.*, vol. 23, pp. 137–163, 2015.

[19] S. L. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.*, vol. 22, pp. 579–588, 1961.

[20] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, 1967.

[21] A. Salam, "Weak and Electromagnetic Interactions", *Conf. Proc. C*, vol. 680519, pp. 367–377, 1968.

[22] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Phisics*. 1984.

[23] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.*, vol. 13, pp. 321–323, 1964.

[24] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Lett.*, vol. 12, pp. 132–133, 1964.

[25] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.*, vol. 13, pp. 508–509, 1964.

[26] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.*, vol. 13, pp. 585–587, 1964.

[27] P. A. Zyla and others (Particle Data Group), "Review of Particle Physics", *PTEP*, vol. 2020, no. 8, p. 083C01, 2020.

[28] CMS Collaboration, "Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C*, vol. 79, p. 421. 67 p, Sep 2018. Submitted to Eur.Phys.J.

[29] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. CERN Yellow Reports: Monographs, Geneva: CERN, Oct 2016. Available at: https://cds.cern.ch/record/2227475.

[30] H. M. Georgi, S. L. Glashow, M. E. Machacek, and D. V. Nanopoulos, "Higgs Bosons from Two Gluon Annihilation in Proton Proton Collisions", *Phys. Rev. Lett.*, vol. 40, p. 692, 1978.

[31] P. Bolzoni, F. Maltoni, S.-O. Moch, and M. Zaro, "Vector boson fusion at next-to-next-to-leading order in qcd: Standard model higgs boson and beyond", *Phys. Rev. D*, vol. 85, p. 035002, Feb 2012.

[32] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*. CERN Yellow Reports: Monographs, Geneva: CERN, 2011.

[33] ATLAS Collaboration, "A combination of measurements of Higgs boson production and decay using up to 139 fb^{-1} of proton–proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment", Tech. Rep. ATLAS-CONF-2020-027, CERN, Geneva, Aug 2020. Available at: https://cds.cern.ch/record/2725733.

[34] M. L. Mangano, "Introduction to QCD", no. CERN-OPEN-2000-255, 1999. Available at: https://cds.cern.ch/record/454171.

[35] J. C. Collins, D. E. Soper, and G. F. Sterman, "Factorization of Hard Processes in QCD", *Adv. Ser. Direct. High Energy Phys.*, vol. 5, pp. 1–91, 1989.

[36] S. Dittmaier and M. Schumacher, "The higgs boson in the standard model—from lep to lhc: Expectations, searches, and discovery of a candidate", *Progress in Particle and Nuclear Physics*, vol. 70, p. 1–54, May 2013.

[37] D. E. Soper, "Parton distribution functions", *Nuclear Physics B - Proceedings Supplements*, vol. 53, p. 69–80, Feb 1997.

[38] F. Pop, "High Performance Numerical Computing for High Energy Physics: A New Challenge for Big Data Science", *Adv. High Energy Phys.*, vol. 2014, p. 507690, 2014.

[39] SLAC Theoretical Physics, "Simulations." Available at: https://theory.slac.stanford.edu/ourresearch/.

[40] P. Skands, "Introduction to QCD", in *Theoretical Advanced Study Institute in Elementary Particle Physics: Searching for New Physics at Small and Large Scales*, 7 2012.

[41] B. R. Webber, "Fragmentation and hadronization", *Int. J. Mod. Phys. A*, vol. 15S1, pp. 577–606, 2000.

[42] Z. Marshall, "Simulation of Pile-up in the ATLAS Experiment", *J. Phys. Conf. Ser.*, vol. 513, p. 022024, 2014.

[43] J. Ellis, "Physics beyond the standard model", *Nuclear Physics A*, vol. 827, p. 187c–198c, Aug 2009.

[44] Y. Fukuda *et al.*, "Evidence for oscillation of atmospheric neutrinos", *Phys. Rev. Lett.*, vol. 81, pp. 1562–1567, 1998.

[45] R. Ahmad, R. Allen, T. Andersen, J. Anglin, J. Barton, E. Beier, M. Bercovitch, J. Bigu, S. Biller, I. Blevis, R. Boardman, J. Boger, E. Bonvin, M. Boulay, M. Bowler, T. Bowles, S. Brice, M. Browne, and M. Yeh, "Direct evidence for neutrino flavor transformation from neutral-current interactions in the sudbury neutrino observatory", *Physical review letters*, vol. 89, p. 011301, 08 2002.

[46] L. Susskind, "Dynamics of Spontaneous Symmetry Breaking in the Weinberg-Salam Theory", *Phys. Rev. D*, vol. 20, pp. 2619–2625, 1979.

[47] R. Peccei and H. Quinn, "Cp conservation in the presence of pseudoparticles", *Physical Review Letters - PHYS REV LETT*, vol. 38, pp. 1440–1443, 06 1977.

[48] H. Georgi, "Effective field theory", *Ann. Rev. Nucl. Part. Sci.*, vol. 43, pp. 209–252, 1993.

[49] E. Fermi, "Tentativo di una teoria dell'emissione dei raggi beta", *Ric. Sci.*, vol. 4, pp. 491–495, 1933.

[50] A. V. Manohar, "Introduction to Effective Field Theories", 4 2018.

[51] S. Weinberg, "Baryon and Lepton Nonconserving Processes", *Phys. Rev. Lett.*, vol. 43, pp. 1566–1570, 1979.

[52] A. de Gouvea, J. Herrero-Garcia, and A. Kobach, "Neutrino Masses, Grand Unification, and Baryon Number Violation", *Phys. Rev. D*, vol. 90, no. 1, p. 016011, 2014.

[53] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, "Dimension-Six Terms in the Standard Model Lagrangian", *JHEP*, vol. 10, p. 085, 2010.

[54] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, "LHC Design Report", 2004.

[55] CERN, "LEP design report", 1984.

[56] UK Research and Innovation, "CERN Accelerator Complex", 2016.

[57] F. Marcastel, "CERN's Accelerator Complex. La chaîne des accélérateurs du CERN", Oct 2013. https://cds.cern.ch/record/1621583.

[58] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST*, vol. 3, p. S08002, 2008.

[59] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST*, vol. 3, p. S08003. 437 p, 2008.

[60] CMS Collaboration, "The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment", *JINST*, vol. 3, p. S08004. 361 p, 2008.

[61] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST*, vol. 3, p. S08005, 2008.

[62] ATLAS Collaboration, "Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC", Tech. Rep. ATLAS-CONF-2019-021, CERN, Geneva, Jun 2019. Available at: http://cds.cern.ch/record/2677054.

[63] ATLAS Collaboration, "Luminosity Public Results", 2013. Available at: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults.

[64] ATLAS Collaboration, "Luminosity Public Results Run 2", 2019. Available at: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[65] J. Pequenao, "Computer generated image of the whole ATLAS detector." Available at: https://cds.cern.ch/record/1095924, Mar 2008.

[66] M. Schott and M. Dunford, "Review of single vector boson production in pp collisions at $\sqrt{s} = 7$ TeV", *Eur. Phys. J. C*, vol. 74, p. 60 p, May 2014.

[67] ATLAS Collaboration, *ATLAS magnet system: Technical Design Report, 1.* Geneva: CERN, 1997.

[68] ATLAS Collaboration, "ATLAS Insertable B-Layer Technical Design Report", tech. rep., Sep 2010. Available at: https://cds.cern.ch/record/1291633.

[69] ATLAS Collaboration, "Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s} = 13$ TeV", tech. rep., CERN, Geneva, Jul 2015. Available at: https://cds.cern.ch/record/2037683.

[70] "IBL Efficiency and Single Point Resolution in Collision Events", tech. rep., CERN, Geneva, Aug 2016. Available at http://cds.cern.ch/record/2203893.

[71] H. Pernegger, "The Pixel Detector of the ATLAS Experiment for LHC Run-2", tech. rep., CERN, Geneva, Feb 2015. Available at: https://cds.cern.ch/record/1985432.

[72] ATLAS Collaboration, "The Silicon microstrip sensors of the ATLAS semiconductor tracker", *Nucl. Instrum. Meth. A*, vol. 578, pp. 98–118, 2007.

[73] ATLAS Collaboration, "Operation and performance of the ATLAS semiconductor tracker", *JINST*, vol. 9, p. P08009, 2014.

[74] ATLAS TRT Collaboration, "The ATLAS TRT barrel detector", *JINST*, vol. 3, p. P02014, 2008.

[75] C. W. Fabjan and F. Gianotti, "Calorimetry for Particle Physics", *Rev. Mod. Phys.*, vol. 75, pp. 1243–1286. 96 p, Oct 2003.

[76] J. Pequenao, "Computer Generated image of the ATLAS calorimeter." Available at: https://cds.cern.ch/record/1095927, Mar 2008.

[77] ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical design report. ATLAS, Geneva: CERN, 1996.

[78] ATLAS Collaboration, "Construction, assembly and tests of the ATLAS electromagnetic barrel calorimeter", *Nucl. Instrum. Methods Phys. Res., A*, vol. 558, pp. 388–418. 77 p, Jul 2005.

[79] ATLAS Collaboration, *ATLAS tile calorimeter: Technical Design Report*. Technical design report. ATLAS, Geneva: CERN, 1996.

[80] D. M. Gingrich *et al.*, "Construction, assembly and testing of the ATLAS hadronic end-cap calorimeter", tech. rep., CERN, Geneva, Apr 2007. Available at: https://cds.cern.ch/record/1026994.

[81] A. Artamonov *et al.*, "The ATLAS Forward Calorimeter", *JINST*, vol. 3, p. P02010, 2008.

[82] D. Gillberg, "Performance of the ATLAS Forward Calorimeters in First LHC Data", tech. rep., CERN, Geneva, Aug 2010. Available at: https://cds.cern.ch/record/1284260.

[83] J. Pequenao, "Computer generated image of the ATLAS Muons subsystem." Available at: https://cds.cern.ch/record/1095929, Mar 2008.

[84] ATLAS Collaboration, *ATLAS muon spectrometer: Technical Design Report*. Technical design report. ATLAS, Geneva: CERN, 1997.

[85] F. Bauer *et al.*, "Construction and test of mdt chambers for the atlas muon spectrometer", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 461, p. 17–20, Apr 2001.

[86] T. Argyropoulos *et al.*, "Cathode strip chambers in ATLAS: Installation, commissioning and in situ performance", *IEEE Trans. Nucl. Sci.*, vol. 56, pp. 1568–1574, 2009.

[87] G. Aielli *et al.*, "The RPC first level muon trigger in the barrel of the ATLAS experiment", *Nuclear Physics B - Proceedings Supplements*, vol. 158, pp. 11–15, 2006.

[88] S. R. Majewski, G. Charpak, A. Breskin, and G. Mikenberg, "A thin multiwire chamber operating in the high multiplication mode", *Nucl. Instrum. Methods Phys. Res.*, vol. 217, pp. 265–271, 1983.

[89] M. zur Nedden, "The Run-2 ATLAS Trigger System: Design, Performance and Plan", tech. rep., CERN, Geneva, Dec 2016. Available at: https://cds.cern.ch/record/2238679.

[90] J. Pequenao and P. Schaffner, "How ATLAS detects particles: diagram of particle paths in the detector." Available at: https://cds.cern.ch/record/1505342, Jan 2013.

[91] ATLAS Collaboration, "The ATLAS Simulation Infrastructure. The ATLAS Simulation Infrastructure", *Eur. Phys. J. C*, vol. 70, pp. 823–874. 53 p, May 2010.

[92] S. Agostinelli *et al.*, "GEANT4: A Simulation toolkit", *Nucl. Instrum. Meth.*, vol. A506, pp. 250–303, 2003.

[93] R. W. Jansky, "The ATLAS Fast Monte Carlo Production Chain Project", tech. rep., CERN, Geneva, May 2015. Available at: https://cds.cern.ch/record/2014717.

[94] CERN, *Computing in High Energy Physics and Nuclear Physics 2004. Computing in High Energy Physics and Nuclear Physics 2004*, (Geneva), CERN, 2005.

[95] ATLAS Collaboration, "Topological cell clustering in the atlas calorimeters and its performance in lhc run 1", *The European Physical Journal C*, vol. 77, Jul 2017.

[96] ATLAS Collaboration, "Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2", *The European Physical Journal C*, vol. 77, Oct 2017.

[97] T. Cornelissen, M. Elsing, I. Gavrilenko, W. Liebig, E. Moyse, and A. Salzburger, "The new ATLAS track reconstruction (NEWT)", *J. Phys. Conf. Ser.*, vol. 119, p. 032014, 2008.

[98] R. Frühwirth, "Application of Kalman filtering to track and vertex fitting", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 262, no. 2, pp. 444–450, 1987.

[99] ATLAS Collaboration, "Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC. Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton-proton collisions at the LHC", *Eur. Phys. J. C*, vol. 77, p. 332. 52 p, Nov 2016.

[100] ATLAS Collaboration, "Performance of primary vertex reconstruction in proton-proton collisions at $\sqrt{s}$ =7 TeV in the ATLAS experiment", tech. rep., CERN, Geneva, Jul 2010. Available at: https://cds.cern.ch/record/1281344.

[101] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, S. Menke, S. Rajagopalan, D. Rousseau, S. Snyder, and G. Unal, "Calorimeter Clustering Algorithms: Description and Performance", tech. rep., CERN, Geneva, Apr 2008. Available at: https://cds.cern.ch/record/1099735.

[102] ATLAS Collaboration, "Performance of electron and photon triggers in atlas during lhc run 2", *The European Physical Journal C*, vol. 80, Jan 2020.

[103] ATLAS Collaboration, "Electron and photon energy calibration with the atlas detector using 2015–2016 lhc proton-proton collision data", *Journal of Instrumentation*, vol. 14, p. P03017–P03017, Mar 2019.

[104] ATLAS Collaboration, "Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C*, vol. 79, p. 639. 40 p, Feb 2019.

[105] ATLAS Collaboration, "Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data", tech. rep., CERN, Geneva, Jun 2016. Available at: https://cds.cern.ch/record/2157687.

[106] ATLAS Collaboration, "Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run 2 data collected in 2015 and 2016", *Eur. Phys. J. C*, vol. 79, p. 205. 55 p, Oct 2018.

[107] ATLAS Collaboration, "Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV", *Eur. Phys. J. C*, vol. 76, no. 5, p. 292, 2016.

[108] G. P. Salam, "Towards jetography", *The European Physical Journal C*, vol. 67, p. 637–686, May 2010.

[109] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *Journal of High Energy Physics*, vol. 2008, p. 063–063, Apr 2008.

[110] ATLAS Collaboration, "Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Rev. D*, vol. 96, p. 072002. 36 p, Mar 2017.

[111] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Physics Letters B*, vol. 659, p. 119–126, Jan 2008.

[112] ATLAS Collaboration, "Jet energy resolution in proton-proton collisions at $\sqrt{s}$ = 7 TeV recorded in 2010 with the ATLAS detector", *Eur. Phys. J. C*, vol. 73, p. 2306. 13 p, Oct 2012.

[113] ATLAS Collaboration, "A measurement of the calorimeter response to single hadrons and determination of the jet energy scale uncertainty using LHC Run-1 *pp*-collision data with the ATLAS detector", *Eur. Phys. J. C*, vol. 77, no. 1, p. 26, 2017.

[114] ATLAS Collaboration, "Jet energy scale and uncertainties in 2015-2017 data and simulation", 2018. Available at: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/JETM-2018-006/.

[115] ATLAS Collaboration, "Jet energy resolution in 2017 data and simulation", 2018. Available at: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/JETM-2018-005/.

[116] ATLAS Collaboration, "Tagging and suppression of pileup jets with the ATLAS detector", tech. rep., CERN, Geneva, May 2014. Available at: https://cds.cern.ch/record/1700870.

[117] ATLAS Collaboration, "Jet mass reconstruction with the ATLAS Detector in early Run 2 data", tech. rep., CERN, Geneva, Jul 2016. Available at: https://cds.cern.ch/record/2200211.

[118] ATLAS Collaboration, "ATLAS Combined Mass resolution and W mass reconstruction performance", 2017. Available at: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/JETM-2017-002/.

[119] ATLAS Collaboration, "In situ calibration of large-radius jet energy and mass in 13 tev proton–proton collisions with the atlas detector", *The European Physical Journal C*, vol. 79, Feb 2019.

[120] ATLAS Collaboration, "Performance of jet substructure techniques for large-r jets in proton-proton collisions at $\sqrt{s} = 7$ tev using the atlas detector", *Journal of High Energy Physics*, vol. 2013, Sep 2013.

[121] D. Krohn, J. Thaler, and L.-T. Wang, "Jet trimming", *Journal of High Energy Physics*, vol. 2010, Feb 2010.

[122] D. Krohn, J. Thaler, and L.-T. Wang, "Jets with variable R", *Journal of High Energy Physics*, vol. 2009, p. 059–059, Jun 2009.

[123] ATLAS Collaboration, "Variable Radius, Exclusive-k$_T$, and Center-of-Mass Subjet Reconstruction for Higgs($\to b\bar{b}$) Tagging in ATLAS", tech. rep., CERN, Geneva, Jun 2017. Available at: https://cds.cern.ch/record/2268678.

[124] M. Cacciari, G. P. Salam, and G. Soyez, "The catchment area of jets", *Journal of High Energy Physics*, vol. 2008, p. 005–005, Apr 2008.

[125] ATLAS Collaboration, "Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC", tech. rep., CERN, Geneva, Nov 2015. Available at: https://cds.cern.ch/record/2064383.

[126] ATLAS Collaboration, "Reconstruction of hadronic decay products of tau leptons with the atlas experiment", *The European Physical Journal C*, vol. 76, May 2016.

[127] ATLAS Collaboration, "Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV", *The European Physical Journal C*, vol. 78, Nov 2018.

[128] D. Adams, C. Anastopoulos, A. Andreazza, *et al.*, "Recommendations of the Physics Objects and Analysis Harmonisation Study Groups 2014", Tech. Rep. ATL-COM-PHYS-2014-451, CERN, Geneva, May 2014. Available at: https://cds.cern.ch/record/1700874.

[129] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interaction", *Prog. Theor. Phys.*, vol. 49, pp. 652–657, 1973.

[130] ATLAS Collaboration, "Performance of the ATLAS Secondary Vertex b-tagging Algorithm in 900 GeV Collision Data", Tech. Rep. ATLAS-CONF-2010-004, CERN, Geneva, Jun 2010. available at: http://cds.cern.ch/record/1273194.

[131] ATLAS Collaboration, "Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data", Tech. Rep. ATLAS-CONF-2011-102, CERN, Geneva, Jul 2011. Available at: https://cds.cern.ch/record/1369219.

[132] ATLAS Collaboration, "Performance of b-Jet Identification in the ATLAS Experiment", *JINST*, vol. 11, no. 04, p. P04008, 2016.

[133] ATLAS Collaboration, "Expected performance of the ATLAS b-tagging algorithms in Run-2", Tech. Rep. ATL-PHYS-PUB-2015-022, CERN, Geneva, Jul 2015. Available at: https://cds.cern.ch/record/2037697.

[134] ATLAS Collaboration, "Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run", Tech. Rep. ATL-PHYS-PUB-2016-012, CERN, Geneva, Jun 2016. Available at: https://cds.cern.ch/record/2160731.

[135] ATLAS Collaboration, "Secondary vertex finding for jet flavour identification with the ATLAS detector", Tech. Rep. ATL-PHYS-PUB-2017-011, CERN, Geneva, Jun 2017. Available at: https://cds.cern.ch/record/2270366.

[136] P. Billoir, R. Fruhwirth, and M. Regler, "Track element merging strategy and vertex fitting in complex modular detectors", *Nucl. Instrum. Meth. A*, vol. 241, pp. 115–131, 1985.

[137] P. Billoir and S. Qian, "Fast vertex fitting with a local parametrization of tracks", *Nucl. Instrum. Meth. A*, vol. 311, pp. 139–150, 1992.

[138] ATLAS Collaboration, "Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run", Tech. Rep. ATL-PHYS-PUB-2017-013, CERN, Geneva, Jul 2017. Available at: https://cds.cern.ch/record/2273281.

[139] ATLAS Collaboration, "Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC", Tech. Rep. ATL-PHYS-PUB-2018-025, CERN, Geneva, Oct 2018. Available at: https://cds.cern.ch/record/2645405.

[140] G. Piacquadio and C. Weiser, "A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS", *J. Phys. Conf. Ser.*, vol. 119, p. 032032, 2008.

[141] ATLAS Collaboration, "Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV", *Journal of High Energy Physics*, vol. 2018, Aug 2018.

[142] A. Hocker *et al.*, "TMVA - Toolkit for Multivariate Data Analysis", Tech. Rep. CERN-OPEN-2007-007, March 2007. Available at: http://cds.cern.ch/record/1019880.

[143] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *The European Physical Journal C*, vol. 71, Feb 2011.

[144] T. Sjöstrand, S. Mrenna, and P. Skands, "A brief introduction to PYTHIA 8.1", *Computer Physics Communications*, vol. 178, no. 11, pp. 852–867, 2008.

[145] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP*, vol. 11, p. 070, 2007.

[146] J. Bellm *et al.*, "HERWIG 7.0/HERWIG++ 3.0 release note", *Eur. Phys. J.*, vol. C76, no. 4, p. 196, 2016.

[147] J. Mitrevski, "Preparing ATLAS reconstruction software for LHC's run 2", *Journal of Physics: Conference Series*, vol. 664, p. 072034, Dec 2015.

[148] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP*, vol. 04, p. 063, 2008.

[149] ATLAS Collaboration, "Impact of ATLAS quasi-stable particle simulation on quantities related to flavour tagging and tau reconstruction", 2020. Available at: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2020-002/.

[150] ATLAS Collaboration, "Modelling of Track Reconstruction Inside Jets with the 2016 ATLAS $\sqrt{s} = 13$ TeV pp dataset", Tech. Rep. ATL-PHYS-PUB-2017-016, CERN, Geneva, Jul 2017. Available at: http://cds.cern.ch/record/2275639.

[151] ATLAS Collaboration, "Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector", *Phys. Lett.*, vol. B786, pp. 59–86, 2018.

[152] K. Mimasu, V. Sanz, and C. Williams, "Higher Order QCD predictions for Associated Higgs production with anomalous couplings to gauge bosons", *JHEP*, vol. 08, p. 039, 2016.

[153] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP*, vol. 06, p. 043, 2010.

[154] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, "Event generation with SHERPA 1.1", *JHEP*, vol. 02, p. 007, 2009.

[155] D. Goncalves and J. Nakamura, "Role of the Z polarization in the $H \to b\bar{b}$ measurement", *Phys. Rev. D*, vol. 98, no. 9, p. 093005, 2018.

[156] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, "Jet substructure as a new Higgs search channel at the LHC", *Phys. Rev. Lett.*, vol. 100, p. 242001, 2008.

[157] ATLAS Collaboration, "ATLAS Sensitivity to the Standard Model Higgs in the HW and HZ Channels at High Transverse Momenta", tech. rep., CERN, Geneva, Aug 2009. Available at: https://cds.cern.ch/record/1201444.

[158] K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke, "HistFactory: A tool for creating statistical models for use with RooFit and RooStats", 6 2012.

[159] R. J. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples", *Comput. Phys. Commun.*, vol. 77, pp. 219–228, 1993.

[160] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C*, vol. 71, p. 1554, 2011. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[161] S. S. Wilks, "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses", *Annals Math. Statist.*, vol. 9, no. 1, pp. 60–62, 1938.

[162] W. Verkerke and D. P. Kirkby, "The RooFit toolkit for data modeling", *eConf*, vol. C0303241, p. MOLT007, 2003.

[163] L. Moneta, K. Belasco, K. S. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, and M. Wolf, "The RooStats Project", *PoS*, vol. ACAT2010, p. 057, 2010.

[164] ATLAS Collaboration, "Measurement of VH, $H \to b\overline{b}$ production as a function of the vector-boson transverse momentum in 13 TeV pp collisions with the ATLAS detector", *JHEP*, vol. 05, p. 141, 2019.

[165] ATLAS Collaboration, "Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC", *Phys. Rev. Lett.*, vol. 117, no. 18, p. 182002, 2016.

[166] R. V. Harlander, S. Liebler, and T. Zirke, "Higgs Strahlung at the Large Hadron Collider in the 2-Higgs-Doublet Model", *JHEP*, vol. 02, p. 023, 2014.

[167] O. Brein, R. V. Harlander, and T. J. E. Zirke, "vh@nnlo - Higgs Strahlung at hadron colliders", *Comput. Phys. Commun.*, vol. 184, pp. 998–1003, 2013.

[168] S. Dittmaier *et al.*, "Handbook of LHC Higgs Cross Sections: 2. Differential Distributions", 1 2012.

[169] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties", 7 2013.

[170] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, "MadGraph 5 : Going Beyond", *JHEP*, vol. 06, p. 128, 2011.

[171] S. Frixione, E. Laenen, P. Motylinski, B. R. Webber, and C. D. White, "Single-top hadroproduction in association with a W boson", *JHEP*, vol. 07, p. 029, 2008.

[172] ATLAS Collaboration, "Measurement of $W^{\pm}Z$ production cross sections and gauge boson polarisation in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Eur. Phys. J. C*, vol. 79, no. 6, p. 535, 2019.

[173] ATLAS Collaboration, "Measurements of the production cross-section for a Z boson in association with b-jets in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *JHEP*, vol. 07, p. 044, 2020.

[174] M. Boggia *et al.*, "The HiggsTools handbook: a beginners guide to decoding the Higgs sector", *J. Phys. G*, vol. 45, no. 6, p. 065004, 2018.

[175] J. R. Andersen *et al.*, "Les Houches 2015: Physics at TeV Colliders Standard Model Working Group Report", in *9th Les Houches Workshop on Physics at TeV Colliders*, 5 2016.

[176] N. Berger *et al.*, "Simplified Template Cross Sections - Stage 1.1", 6 2019.

[177] ATLAS Collaboration, "Methodology for EFT interpretation of Higgs boson Simplified Template Cross-section results in ATLAS", tech. rep., CERN, Geneva, Oct 2019. Available at: https://cds.cern.ch/record/2694284.

[178] K. Albertsson *et al.*, "Machine Learning in High Energy Physics Community White Paper", *J. Phys. Conf. Ser.*, vol. 1085, no. 2, p. 022008, 2018.

[179] R. K. Ellis *et al.*, "Physics Briefing Book: Input for the European Strategy for Particle Physics Update 2020", 10 2019.

[180] ATLAS Collaboration, "Projections for measurements of Higgs boson cross sections, branching ratios, coupling parameters and mass with the ATLAS detec-

tor at the HL-LHC", tech. rep., 2018. Available at: https://cds.cern.ch/record/2652762.

[181] D. Bortoletto, "How and why silicon sensors are becoming more and more intelligent?", *JINST*, vol. 10, no. 08, p. C08016, 2015.

[182] L. Rossi, P. Fischer, T. Rohe, and N. Wermes, *Pixel Detectors: From Fundamentals to Applications*. Particle Acceleration and Detection, Berlin: Springer-Verlag, 2006.

[183] D. A. Neamen, *Semiconductor physics and devices: basic principles*. 4th edition, McGraw-Hill, 2011.

[184] ATLAS Collaboration, "Exploring the scientific potential of the ATLAS experiment at the High-Luminosity LHC", May 2019. Available at: https://atlas.cern/updates/news/scientific-potential-high-luminosity-lhc.

[185] "High-Luminosity Large Hadron Collider (HL-LHC): Technical design report", vol. 10/2020, 12 2020.

[186] ATLAS Collaboration, "Technical Design Report for the ATLAS Inner Tracker Strip Detector", Tech. Rep. CERN-LHCC-2017-005, ATLAS-TDR-025, CERN, Geneva, Apr 2017.

[187] ATLAS Collaboration, "Sensitivity of the ATLAS experiment to long-lived particles with a displaced vertex and E_T^{miss} signature at the HL-LHC", Tech. Rep. ATL-PHYS-PUB-2018-033, CERN, Geneva, Nov 2018.

[188] ATLAS Collaboration, "ITk Pixel Layout Updates", April 2020. Available at: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/ITK-2020-002/.

[189] Temperature Protection Devices - TDK, "Chip NTC thermistor, NTCG series", 2020. Available at: https://product.tdk.com/en/system/files?file=dam/doc/product/sensor/ntc/chip-ntc-thermistor/catalog/tpd_automotive_ntc-thermistor_ntcg_en.pdf.

[190] H. Pernegger *et al.*, "Radiation hard monolithic CMOS sensors with small electrodes for High Luminosity LHC", *Nucl. Instrum. Meth. A*, vol. 986, p. 164381, 2021.